USC OMDL 1501

(2)

DEVICES AND SYSTEMS FOR NONLINEAR OPTICAL INFORMATION PROCESSING

AFOSR-83-0185

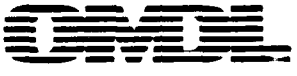
FINAL TECHNICAL REPORT
(RESEARCH PERIOD: 05/15/83 - 11/30/87)

Submitted To:

Air Force Office of Scientific Research
Electronics and Materials Science Division
AFOSR/NE
Bolling AFB, Bldg. 410
Washington, D.C. 20332
Attn: Dr. C. Lee Giles

Submitted By:

Dr. Armand R. Tanguay, Jr.
Optical Materials and Devices Laboratory
University of Southern California
University Park, MC-0483
Los Angeles, California 90089



OPTICAL MATERIALS AND DEVICES LABORATORY

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

ADA203034

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) USC OMDL-1501	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 88-1250
6a. NAME OF PERFORMING ORGANIZATION UNIVERSITY OF SOUTHERN CALIFORNIA 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
6c. ADDRESS (City, State and ZIP Code) UNIVERSITY OF SOUTHERN CALIFORNIA UNIVERSITY PARK, MC-0483 LOS ANGELES, CALIFORNIA 90089-0483	7b. ADDRESS (City, State and ZIP Code) AFOSR/NE BUILDING 410 BOLLING AFB, DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Same as 7a 8b. OFFICE SYMBOL (If applicable) PKD	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-83-0185
8c. ADDRESS (City, State and ZIP Code) Same As 7b	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2305; TASK NO. B1; WORK UNIT NO.
11. TITLE (Include Security Classification) DEVICES AND SYSTEMS FOR NONLINEAR OPTICAL INFORMATION PROCESSING	
12. PERSONAL AUTHOR(S) DR. ARMAND R. TANGUAY, JR.	
13a. TYPE OF REPORT FINAL TECHNICAL	13b. TIME COVERED FROM 05/15/83 TO 11/30/87
14. DATE OF REPORT (Yr., Mo., Day) 1988, NOVEMBER	15. PAGE COUNT 136
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	VARIABLE GRATING MODE LIQUID CRYSTAL DEVICE (VGM LCD), OPTICAL INFORMATION PROCESSING, OPTICAL COMPUTING, SPATIAL LIGHT MODULATORS, FUNDAMENTAL AND TECHNOLOGICAL LIMITATIONS

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

The principal thrust of this research effort is the development and evaluation of optical devices and associated systems architectures for the implementation of highly parallel nonlinear optical information processing functions such as thresholding, level slicing, logarithms, and power laws. Such devices include the Variable Grating Mode Liquid Crystal Device (VGM LCD), which is uniquely capable of performing a novel and powerful intensity-to-position encoding function.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ OTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Giles	22b. TELEPHONE NUMBER (Include Area Code) (202) 767-4931 22c. OFFICE SYMBOL NE

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

19. ABSTRACT (continued)

A related and evolving research thrust is the detailed analysis of both the fundamental and technological limitations that affect the performance of envisioned optical information processing and computing devices and systems, for both linear and nonlinear function implementation. The primary objective here is to compare and contrast electronic and optical (as well as hybrid electro-optical or optoelectronic) devices, circuits, and systems with respect to their current, near term projected, and ultimate computational capabilities. The demonstrable goal of such an effort is to be able (on the basis of the above analyses) to segregate classes of problems most appropriate for solution by electronic, optical, or hybrid processors, subject to specific and pre-defined performance metrics.

During the course of the research effort, the Variable Grating Mode Liquid Crystal Device was further developed and extensively characterized. A new dynamic model for grating formation and grating modulation was developed that explains for the first time both the observed linear relationship between applied voltage and grating spatial frequency, and the device temporal response characteristics. The experimentally observed polarization properties of diffraction from electrically addressed VGM liquid crystal cells were utilized to derive a model of the liquid crystal grating structure, which in turn provided strong evidence for the involvement of the converse flexoelectric effect in grating formation. For the first time, both thin (Raman-Nath-like) and thick (Bragg-like) diffraction characteristics in alternating diffraction orders originating from the same optical device were observed and explained. These results were shown to confirm predictions of a new model of the general polarization properties of birefringent phase gratings, developed under this contract. Finally, experimental implementations demonstrated for the first time the flexibility and computational power inherent in the intensity-to-position conversion process afforded by the VGM LCD for application to fully parallel implementation of nonlinear optical processing and computing functions such as level slicing, thresholding, logarithms, exponentiation, polynomial expansions, all sixteen primitive binary logic functions, and full binary addition.

Examination of the fundamental and technological limitations of optical information processing and computing allowed the fundamental relationships (and therefore the basic tradeoffs) among dynamic range, frame rate, resolution, and sensitivity for quantum-limited analog multi-dimensional arrays to be derived. The fundamental and technological limitations imposed by the choice of numerical representation, inherent computational complexity, and detection for both electronic and optical computational systems were compared, with particular emphasis on the relative advantages and disadvantages of analog, digital, and hybrid representations and detections. The maximum possible rate of irreducible binary operations per unit power was derived at the quantum/thermal limit, at the current device level, and at the current system level. This study showed that as many as 14 orders of magnitude in effective computational power are currently lost for semiconductor electronic implementations in the transition from the fundamental limits to the system integration level. Finally, the key factors contributing to the relative insensitivity of volume holographic grating formation in photorefractive materials for multi-dimensional weighted interconnection applications were determined, which in turn allowed for the determination of the maximum reconfiguration rate that can be achieved per unit power.

DEVICES AND SYSTEMS FOR NONLINEAR OPTICAL INFORMATION PROCESSING

TABLE OF CONTENTS

DEVICES AND SYSTEMS FOR NONLINEAR OPTICAL INFORMATION PROCESSING

ABSTRACT

The principal thrust of this research effort is the development and evaluation of optical devices and associated systems architectures for the implementation of highly parallel nonlinear optical information processing functions such as thresholding, level slicing, logarithms, and power laws. Such devices include the Variable Grating Mode Liquid Crystal Device (VGM LCD), which is uniquely capable of performing a novel and powerful intensity-to-position encoding function.

A related and evolving research thrust is the detailed analysis of both the fundamental and technological limitations that affect the performance of envisioned optical information processing and computing devices and systems, for both linear and nonlinear function implementation. The primary objective here is to compare and contrast electronic and optical (as well as hybrid electro-optical or optoelectronic) devices, circuits, and systems with respect to their current, near term projected, and ultimate computational capabilities. The demonstrable goal of such an effort is to be able (on the basis of the above analyses) to segregate classes of problems most appropriate for solution by electronic, optical, or hybrid processors, subject to specific and pre-defined performance metrics.

During the course of the research effort, the Variable Grating Mode Liquid Crystal Device was further developed and extensively characterized. A new dynamic model for grating formation and grating modulation was developed that explains for the first time both the observed linear relationship between applied voltage and grating spatial frequency, and the device temporal response characteristics. The experimentally observed polarization properties of diffraction from electrically addressed VGM liquid crystal cells were utilized to derive a model of the liquid crystal grating structure, which in turn provided strong evidence for the involvement of the converse flexoelectric effect in grating formation. For the first time, both thin (Raman-Nath-like) and thick (Bragg-like) diffraction characteristics in alternating diffraction orders originating from the same optical device were observed and explained. These results were shown to confirm predictions of a new model of the general polarization properties of birefringent phase gratings, developed under this contract. Finally, experimental implementations demonstrated for the first time the flexibility and computational power inherent in the intensity-to-position conversion process afforded by the VGM LCD for application to fully parallel implementation of nonlinear optical processing and computing functions such as level slicing, thresholding, logarithms, exponentiation, polynomial expansions, all sixteen primitive

binary logic functions, and full binary addition.

Examination of the fundamental and technological limitations of optical information processing and computing allowed the fundamental relationships (and therefore the basic tradeoffs) among dynamic range, frame rate, resolution, and sensitivity for quantum-limited analog multi-dimensional arrays to be derived. The fundamental and technological limitations imposed by the choice of numerical representation, inherent computational complexity, and detection for both electronic and optical computational systems were compared, with particular emphasis on the relative advantages and disadvantages of analog, digital, and hybrid representations and detections. The maximum possible rate of irreducible binary operations per unit power was derived at the quantum/thermal limit, at the current device level, and at the current system level. This study showed that as many as 14 orders of magnitude in effective computational power are currently lost for semiconductor electronic implementations in the transition from the fundamental limits to the system integration level. Finally, the key factors contributing to the relative insensitivity of volume holographic grating formation in photorefractive materials for multi-dimensional weighted interconnection applications were determined, which in turn allowed for the determination of the maximum reconfiguration rate that can be achieved per unit power.

DEVICES AND SYSTEMS FOR NONLINEAR OPTICAL INFORMATION PROCESSING

1. INTRODUCTION

The research effort described in this report was initiated as a follow-on to an earlier contract sponsored by the Air Force Office of Scientific Research, entitled "Nonlinear Real-Time Optical Signal Processing" [1]. The goal of this earlier program was to extend fast parallel nonlinear operations to optical processing systems with large time-bandwidth and space-bandwidth products. The current contract evolved naturally from the critical need identified in the earlier program for optical information processing components that were capable of implementing highly parallel nonlinear functions with essentially arbitrary programmability.

During the preceding research effort, a unique optical information processing component was invented and preliminarily investigated. This device, the Variable Grating Mode Liquid Crystal Device or VGM LCD [2-6], was based on an unusual physical effect in certain nematic liquid crystals that produced a variable period diffraction grating in response to an applied field. Optically addressed versions of the basic device structure were fabricated in collaboration with Hughes Research Laboratories, Malibu, CA, and were used in preliminary systems characterization experiments to demonstrate all sixteen Boolean logic functions, full binary addition, exponentiation, logarithms, power laws, thresholding, level slicing, and other nonlinear operations [2-3]. Functional programmability was demonstrated to require only a low contrast ratio, low resolution programmable Fourier plane filter.

The initial primary goal of the current researh effort was to further characterize the Variable Grating Mode Liquid Crystal Device, and of the variable grating mode liquid crystal effect itself, with a specific focus on the achievement of sufficient physical understanding of the mechanism and its limitations to allow advanced development of the VGM LCD device technology to proceed rapidly. Emphasis was placed throughout on the early determination of performance limiting factors that could be classified as fundamental in nature, in order that the full potential of the VGM LCD could be clearly defined, and on the basis of which further decisions on continued advanced development could be judiciously made.

During the course of the research effort, the consideration of fundamental and technological limitations was extended to devices other than the VGM LCD, and in fact to the very nature of optical information processing and computing as a potential burgeoning technology. These investigations addressed numerous important issues such as the energy

cost of numerical representation in analog, digital, and hybrid formats at the quantum limits, and at the current technological limits for both semiconductor (VLSI) electronics and for optical device implementations.

The long term goal of this effort is to clearly delineate if possible the computational advantages that accrue to particular choices of representation, computational paradigm, and detection schemes, all in the highly parallel context of optical information processing and computing systems. Such a delineation would then allow decisions to be made on the optimum implementation of certain classes of computational problems, both on a technology-independent and technology-dependent basis. In other words, systems design studies could then be broadened to include consideration of the maximum achievable computational power of a given paradigm, and of the likely achievable computational power of the same or other paradigms implemented with various currently available or forseeable technologies.

Both of these principal research thrusts are summarized in this report. Section 2 outlines the background of the research effort, with reference to significant preceding work. The technical objectives for the program are stated in Section 3, along with a description of the associated scientific approaches utilized. The principal results of the research effort are then summarized in Section 4, with details provided in the several appendices. Publications resulting from AFOSR sponsorship of this research are listed in Section 5, including a Ph.D. Thesis on the Variable Grating Mode Liquid Crystal Device. The scientific personnel that contributed to the program are listed in Section 6, and references to the technical sections of the report are provided in Section 7.

2. BACKGROUND OF THE RESEARCH PROGRAM

The Variable Grating Mode Liquid Crystal Device (VGM LCD) is comprised of an oriented thin layer of a certain nematic liquid crystal material [3] sandwiched between two transparent conducting electrodes. In electrically addressed VGM LCDs, the liquid crystal layer is placed in direct contact with the electrode surfaces. In optically addressed VGM LCDs, an additional photoconductive layer is placed in series with the liquid crystal layer to provide a local conversion of input intensity to voltage applied across the liquid crystal layer.

The variable grating mode liquid crystal effect manifests itself in the dramatic response of the liquid crystal material to an applied field. Above a given threshold voltage, the homogeneously aligned nematic liquid crystal phase exhibits an additional periodic orientational component, forming a birefringent phase grating characterized by a spatial frequency that is a linear function of the applied voltage. A wide range of spatial frequencies have been induced in a single VGM cell, typically from 200 - 1200 lp/mm.

Hence the VGM LCD essentially performs a voltage-to-spatial frequency conversion function, which can be transformed to a voltage-to-spatial position encoding mechanism by utilizing collimated light for readout in conjunction with a Fourier transform lens. Addition of the photoconductive layer provides an input-intensity-to-voltage conversion function, such that the optically addressed version of the VGM LCD becomes in effect an input-intensity-to-spatial-frequency or input-intensity-to-spatial-position converter.

This input-intensity-to-spatial-frequency transformation property is unique to the Variable Grating Mode Liquid Crystal Device, to the best of our knowledge, and provides an exceptionally flexible means of introducing arbitrary nonlinear optical information processing and computing functions. In operation, an input image spatially modulates the photoconductive layer, locally altering the spatial frequency of the VGM domains through variations in the local electric field within the liquid crystal layer. Readout at a nondestructive wavelength by a collimated beam then produces in the back focal plane of the Fourier transform lens a one-dimensional mapping of all areas in the input image that are characterized by the same input intensity.

The arbitrarily programmable nonlinearity is then introduced in the Fourier plane by means of a one-dimensional transmissive filter that implements the desired nonlinear transformation [1-3], such as a level slice, a variable offset threshold, or a logarithmic function. Dynamic programmability is exceptionally easy to implement, requiring no alteration to the VGM LCD and resetting of only the transmission characteristics of the one-dimensional filter.

It should be emphasized that the performance requirements implied for the Fourier

plane filter are quite modest (relatively low contrast and resolution), since it need only support in resolution the total number of resolvable grey levels in the input image, and in contrast ratio the total number of resolvable grey levels required for the output image. For the generation of sigmoid-like threshold functions as utilized, for example, in several proposed implementations of neural network models, Fourier plane filters with contrast ratios of 10:1 and a resolution spanning 100 pixels will likely be more than adequate.

Following the invention of the Variable Grating Mode Liquid Crystal Device at USC under the previous program and its subsequent patenting [6], extensive characterization efforts were undertaken. These efforts involved an attempt to understand the striking polarization properties of the VGM domains as evidenced in the diffraction pattern [4], parallel efforts at Hughes Research Laboratories to synthesize VGM variants with wider dynamic range and faster response times [3], and the characterization of optically addressed cells as utilized in a variety of potential applications ranging from the implementation of parallel optical logic functions [2] through full parallel addition with carries, to nonlinear thresholding, level slicing, and logarithmic filtering [3, 5, 6].

In summary, then, the unique intensity-to-spatial frequency conversion function implemented by the Variable Grating Mode Liquid Crystal Device provides a highly versatile method of introducing essentially arbitrary functional nonlinearities into parallel optical information processing and computing systems. The further characterization of the VGM LCD and of the VGM effect itself became the subject of the current research effort, as described in detail below.

The evolving effort on fundamental and technological limitations of optical information processing and computing derived from a perceived need to delineate on the basis of quantifiable metrics among the many possible strategies (whether electronic, optical, or a hybrid combination) for dramatically enhancing available computational power. An extensive parallel history of such efforts exists in the VLSI literature [7, 8, 9], in which basic physical principles have been invoked to both understand current VLSI performance and to project future operational capabilities. Over the past few years, initial efforts have been undertaken to analyze the comparative fundamental and technological limitations of digital optical computing [10, 11, 12, 13, 14] and of integrated optical devices [15]. On the other hand, the tradeoffs inherent in analog as opposed to digital approaches, and in particular in the use of mixed representations (as is inherent, for example, in neural network models), were not previously explored. This apparent void provided the specific point of departure for the studies initiated under this contract, with results and implications as described in Section 4.

3. STATEMENT OF TECHNICAL OBJECTIVES

The technical objectives of the concluded research program are summarized in this section, in conjunction with a description of the primary scientific approach utilized to achieve each objective.

3.1 Technical Objectives

To physically characterize and develop active components such as the Variable Grating Mode Liquid Crystal Device (VGM LCD) that are capable of implementing nonlinear optical information processing functions such as thresholding, level slicing, logarithms, and power laws.

To develop complementary systems concepts utilizing active components such as the VGM LCD that are capable of implementing nonlinear optical information processing and computing functions.

To analyze in detail both the fundamental and technological limitations that affect the performance of envisioned optical information processing and computing devices and systems, for both linear and nonlinear functions.

To compare and contrast electronic and optical (as well as hybrid electro-optical or optoelectronic) devices, circuits, and systems with respect to thier current, near term projected, and ultimate computational capabilities.

On the basis of the above analyses, to develop methods for segregating classes of problems most appropriate for solution by electronic, optical, or hybrid processors, subject to specific performance metrics.

3.2 Scientific Approach

Theoretically and experimentally analyze the fundamental mechanism of grating formation and modulation in the Variable Grating Mode Liquid Crystal Device, in order to optimize the device response time.

Develop a dynamic model of grating formation based on selective amplification of specific wave vector components of the (nonlinear) flexoelectric instability.

Theoretically and experimentally analyze the strikingly unique polarization properties of diffraction from variable grating mode liquid crystal behavior.

Utilize the analysis of polarization properties to experimentally determine the characteristic parameters of the VGM grating structure as a function of applied voltage (and grating spatial frequency).

Examine the fundamental and technological limitations imposed on both electronic and optical computational systems by the choice of numerical representation (binary, multi-valued logic, analog), computational complexity (architecture and algorithm dependent), and detection scheme (digital or analog).

4. SUMMARY OF RESEARCH PROGRAM

In this section, the principal accomplishments and results that have been developed under this research program are summarized. As this research has resulted in numerous journal publications, book chapters, and conference presentations, several such key publications are provided as appendices. These publications describe the specific research results in considerable depth.

Extensive characterization efforts of both electrically addressed and optically addressed VGM liquid crystal cells were undertaken in an effort to identify the underlying mechanism responsible for the observed variable grating mode phenomenon. This included detailed measurements of the device response time from below to above threshold, and of the relaxation time between various spatial frequency conditions. In addition, the diffraction efficiency was measured as a function of the applied voltage, and the polarization properties of the diffracted orders were examined. The results of these measurements are detailed in Appendix 1 [AFOSR Jour. Publ. 1]. In order to perform these experiments, significant improvements in cell fabrication technology were made, resulting in superbly uniform alignment and grating spatial frequency across wide aperture liquid crystal cells [AFOSR Thesis 1].

Strikingly unique polarization properties of diffraction from electrically addressed variable grating mode liquid crystal cells were observed during these experiments. For linearly polarized input beams, the diffraction orders exhibited completely different behavior in alternating diffracted orders. In the odd orders, the diffracted beams were essentially linearly polarized, but in a direction that rotated counterclockwise as the incident polarization was rotated clockwise (as is characteristic of a half-wave plate, for example). In the even orders, the diffracted beams were always linearly polarized parallel to the direction of liquid crystal quiescent alignment, with an intensity that varied as the projection of the incident polarization on the quiescent alignment direction.

These unusual experimentally observed polarization properties were utilized to derive a model of the liquid crystal grating structure, in direct analogy to the use of x-ray diffraction for the elucidation of atomic and molecular crystal structure [Appendices 1, 2; AFOSR Jour. Publs. 1, 2]. Observed along the direction of grating modulation, the motion of the liquid crystal director describes a cycloidal figure, with components in the plane of the electrodes as well as normal to it. By utilizing careful measurements of the diffracted intensities under a wide range of incident polarization and applied voltage conditions, the angular excursions of these components could be determined [Appendices 1, 2; AFOSR Jour. Publs. 1, 2].

With this experimental evidence in hand, the free energy for the nematic liquid crystal

structure including the flexoelectric contribution was minimized, resulting in analytic expressions for the various components of molecular motion, in complete agreement with the experimental results [Appendix 1; AFOSR Jour. Publ. 1]. This provided rather conclusive evidence of the importance of the flexoelectric mechanism to variable grating mode grating formation, as had been earlier predicted theoretically [16].

This accomplishment then allowed us to begin to examine the temporal response of grating formation and alteration, which resulted in the development of a new dynamic model for grating formation and grating modulation in variable grating mode liquid crystal structures [AFOSR Jour. Publ. 9]. This model is based on the selective amplification of specific wave vector components of the (nonlinear) flexoelectric instability. This model explains for the first time both the observed linear relationship between applied voltage and grating spatial frequency, and the device temporal response characteristics, neither of which evolve from the static free energy minimization approach.

Further research on the unusual polarization properties of the Variable Grating Mode Liquid Crystal Device produced results that could not be explained on the basis of the thin grating approximation utilized previously. In particular, it was observed that both thin grating (Raman-Nath-like) and thick grating (Bragg-like) diffraction characteristics occurred in alternating diffraction orders originating from the same optical device. The anisotropic optical beam propagation method [Appendix 4; AFOSR Jour. Publ. 5] was then developed in order to incorporate the effects of both large natural anisotropy and spatially varying birefringent axes in a model of birefringent phase grating diffraction. Such an analysis resulted in completely satisfactory agreement between the experimental observations and theoretical predictions [AFOSR Jour. Publ. 8].

These results on the unique polarization properties of the Variable Grating Mode Liquid Crystal Device were seen to have profound implications for a wide range of modern optical devices, in which such polarization properties could be utilized to advantage. Such devices included photorefractive volume holographic optical elements, electrooptic spatial light modulators such as the PROM, PRIZ, PICOC, and MSLM, and the Linear Array Total Internal Reflection Spatial Light Modulator. Therefore, a unified theory of the polarization properties of birefringent phase gratings was developed for the first time, that in fact accounts for a very wide range of observed but previously unexplained phenomena [Appendix 4; AFOSR Jour. Publ. 7, AFOSR Bk. Ch. 2].

With regard to the evolving investigation of the fundamental and technological limitations of optical information processing and computing, a key early result was the derivation of a fundamental relationship (and therefore a basic tradeoff) among dynamic range, frame rate, resolution, and sensitivity for quantum-limited analog multi-dimensional ar-

rays [Appendices 3, 5; AFOSR Jour. Publs. 3, 4, and 6]. This relationship applies equally well to spatial light modulators (as used, for example, as incoherent-to-coherent optical converters) and to array detection of output results. The implications are rather surprising, in that at the quantum limits, that is irrespective of envisioned technological implementation, the tradeoffs place strict limitations on the joint specification of these parameters. For example, to achieve a 1000:1 dynamic range at a frame rate of 1 kHz over a 1 square centimeter 1000 x 1000 array requires approximately 2 mW per square centimeter of absorbed incident power. This is an extremely large number in view of the strong push over the past decade to achieve ever more sensitive SLMs and detector arrays.

Utilizing this key result, the fundamental and technological limitations imposed by the choice of numerical representation, inherent computational complexity, and detection were compared for both electronic and analog computational systems. Taking the computation of a two-dimensional Fourier transform as an example, the fundamental quantum limitations on the total energy required to perform the computation were derived for both analog and digital representations (assuming that the analog computation was performed optically). At the quantum limits, the digital FFT approach is slightly superior to the analog approach from the energy metric perspective alone. However, when the technological limitations are considered, the digital approach as implemented in VLSI suffers strongly in comparison at the device, circuit, and particularly the system implementation levels [Appendix 5; AFOSR Jour. Publ. 6]. Hence the optimum computational system from an energy consumption point of view will tend to favor analog implementations for high computational complexity problems in the foreseeable future.

The results described above can be expressed in other equally useful ways. For example, it is possible to derive the maximum rate of irreducible binary operations per unit power possible at the quantum/thermal limit, at the current device level, at the current circuit level, and at the current system level. Surprising results accrue to such a calculation, including the fact that for the vast majority of current electronic computational systems ranging from personal computers to supercomputers, the number of irreducible operations performed per unit input power is about 14 orders of magnitude lower than that implied by the relevant fundamental limits [Appendix 5; AFOSR Jour. Publ. 6]. Not only does this result have powerful implications on future electronic computational systems design, but it also suggests a rather sizable gap in energy efficient computation that could conceivably be filled by appropriately configured optical information processors and computers.

Highly parallel interconnections that can be dynamically reconfigured have been recognized as key to the potential success of optical information processing and computing

systems, ranging from simple high speed Vander Lugt correlators to sophisticated neural-like networks. Hence, we undertook to determine the key factors that contribute to the seeming relative insensitivity of volume holographic grating formation in photorefractive materials for multi-dimensional weighted interconnection applications. By again starting at the quantum limited recording conditions, we were able to derive the derating factors that affect the recording sensitivity (or response time at a given incident power level) for a wide range of photorefractive materials operated under various experimental configurations (including the utilization of so-called enhanced recording techniques such as alternating applied fields and running gratings). Such considerations allow estimates to be made of the maximum reconfiguration rate of the holographic interconnection matrix that can be performed per unit incident power under a wide variety of circumstances. These results are described in detail in Appendix 6 [AFOSR Bk. Ch. 1], which has been written in such a manner as to hopefully prove accessible to researchers approaching optical information processing and computing from the systems as opposed to the materials and devices perspective.

5. PUBLICATIONS UNDER AFOSR SPONSORSHIP

The following technical publications and conference presentations describe research supported in part by this contract.

5.1 Journal Publications

1. A. R. Tanguay, Jr., P. Chavel, T. C. Strand, C. S. Wu, and B. H. Soffer, "Physical Characterization of the Variable Grating Mode Liquid Crystal Device", Optical Engineering (Special Issue on Spatial Light Modulators: Critical Issues), 22(6), 687-694, (1983).

2. A. R. Tanguay, Jr., P. Chavel, T. C. Strand, C. S. Wu, and B. H. Soffer, "Polarization Properties of the Variable Grating Mode Liquid Crystal Device", Optics Letters, 9, 174-176, (1984).

3. A. R. Tanguay, Jr., "Materials Requirements for Optical Processing and Computing Devices", Optical Engineering (Special Issue on Critical Reviews of Technology: Optical Computing), 24(1), 2-18, (1985); (Invited Review Paper). (Note: This paper was the basis for the award to the author of the 1985 Rudolph Kingslake Medal and Prize of the Society of Photo-Optical Instrumentation Engineers (SPIE)).

4. A. R. Tanguay, Jr., "Materials Requirements for Optical Processing and Computing Devices", Proceedings of the Society of Photo-Optical Instrumentation Engineers, 456 (Critical Reviews of Technology: Optical Computing), 130-158, (1984); (Invited Review Paper).

5. R. V. Johnson and A. R. Tanguay, Jr., "Optical Beam Propagation Method for Birefringent Phase Grating Diffraction", Optical Engineering, 25(2), 235-249, (1986).

6. A. R. Tanguay, Jr., "Physical and Technological Limitations of Optical Information Processing and Computing", Materials Research Society Bulletin, Special Issue on Photonic Materials, XIII(8), 36-40, (1988); (Invited Paper).

7. A. R. Tanguay, Jr., "Polarization Properties of Birefringent Phase Gratings", in preparation for Journal of the Optical Society of America, (1988).

8. C. S. Wu, R. V. Johnson, and A. R. Tanguay, Jr., "Optical Diffraction Characteristics of the Variable Grating Mode Liquid Crystal Device", in preparation for Journal of the Optical Society of America, (1988).

9. C. S. Wu, R. V. Johnson, and A. R. Tanguay, Jr., "Flexoelectric Instability of the Variable Grating Mode Liquid Crystal Device: Dynamic Model of Grating Formation", in preparation for Journal of Applied Physics, (1988).

5.2 Book Chapters

1. R. V. Johnson and A. R. Tanguay, Jr., "Fundamental Physical Limitations of the Photorefractive Grating Recording Sensitivity", in *Optical Processing and Computing*, H. Arsenault and T. Szoplik, Eds., Academic Press, New York, (1988).

2. A. R. Tanguay, Jr. and R. V. Johnson, "Polarization Properties of Birefringent Phase Gratings", in preparation for *Progress in Optics*, E. Wolf, Ed., (Invited Review Chapter), (1988).

5.3 Theses

1. C. S. Wu, "The Origin and Properties of the Variable Grating Mode Liquid Crystal Effect for Optical Processing Device Applications", Ph.D. Thesis, University of Southern California, (1987).

5.4 Conference Presentations

1. A. R. Tanguay, Jr., "Materials Requirements for Optical Processing and Computing Devices: Electrooptics", National Science Foundation Symposium on The Future of Lightwave Technology, Los Angeles, California, (1984); (Invited Paper).

2. A. R. Tanguay, Jr., "Materials Requirements for Optical Processing and Computing Devices", SPIE International Symposium, Los Angeles, California, (1984); (Invited Paper).

3. A. R. Tanguay, Jr., "Materials Requirements for Optical Processing and Computing", National Bureau of Standards, Gaithersburg, Maryland, (1984); (Invited Paper).

4. A. R. Tanguay, Jr., "Variable Grating Mode Liquid Crystal Device Research", Air Force Office of Scientific Research, Washington, D. C., (1984); (Invited Presentation).

5. A. R. Tanguay, Jr., "Optical Devices for Image Processing and Computing", Yale University, New Haven, Connecticut, (1984); (Invited Colloquium).

6. A. R. Tanguay, Jr., "Functional Requirements for Optical Processing and Computing", ITT Corporation Symposium on Ultra-high Bandwidth Communications, SEL Laboratories, Stuttgart, Federal Republic of Germany, (1985); (Invited Paper).

7. A. R. Tanguay, Jr., "Optical Materials Requirements for Optical Processing and Computing Devices", Symposium on Materials in Emerging Technologies, American Chemical Society, Chicago, Illinois, (1985); (Invited Paper).

8. A. R. Tanguay, Jr., "Optical Computing Components: Fundamental Issues", 1985 Annual Meeting of the Optical Society of America, Washington, D. C., (1985); (Invited Paper).

9. R. V. Johnson and A. R. Tanguay, Jr., "Optical Beam Propagation in Anisotropic Media", 1985 Annual Meeting of the Optical Society of America, Washington, D. C., (1985).

10. C. S. Wu, R. V. Johnson, and A. R. Tanguay, Jr., "Optical Diffraction Characteristics of the Variable Grating Mode Liquid Crystal Device", 1985 Annual Meeting of the Optical Society of America, Washington, D. C., (1985).

11. K. Rastani, Z. Karim, R. V. Johnson, and A. R. Tanguay, Jr., "Optical Information Processing Components", Gordon Research Conference on Optical Information Processing and Holography, Santa Barbara, California, (1987); (Invited Paper).

12. C. Kyriakakis, P. Asthana, R. V. Johnson, and A. R. Tanguay, Jr., "Fundamental Physical Limitations of Optical Information Processing and Computing", 1987 Optical Society of America Topical Meeting on Optical Computing, Incline Village, Nevada, (1987); (Invited Paper).

13. A. R. Tanguay, Jr., "Fundamental Physical Limitations of Optical Information Processing and Computing", NATO Collaborative Award Colloquium, University College London, London, England, (1987); (Invited Colloquium).

14. A. R. Tanguay, Jr., "Fundamental Physical Limitations of Optical Information Processing and Computing", British Telecom Research Laboratories, Ipswich, England, (1987); (Invited Colloquium).

15. A. R. Tanguay, Jr., "Fundamental Physical Limitations of Optical Information Processing and Computing", University of Rochester, Institute of Optics, Rochester, New York, (1987); (Invited Colloquium).

16. A. R. Tanguay, Jr., "Optical Processing and Computing Devices: The Materials Perspective", American Conference on Crystal Growth-7, Monterey, California, (1987); (Invited Paper).

17. A. R. Tanguay, Jr., "Fundamental Physical Limitations of Optical Information Processing and Computing", DARPA Panel on Neural Networks, California Institute of Technology, Pasadena, California, (1987); (Invited Presentation).

18. C. Kyriakakis, P. Asthana, R. V. Johnson, and A. R. Tanguay, Jr., "Spatial Light Modulators: Fundamental and Technological Issues", Optical Society of America Topical Meeting on Spatial Light Modulators, Lake Tahoe, Nevada, (1988); (Invited Paper).

19. E. J. Herbulock, M. H. Garrett, and A. R. Tanguay, Jr., "Electric Field Profile Effects on Photorefractive Grating Formation in Bismuth Silicon Oxide", 1988 Annual Meeting of the Optical Society of America, Santa Clara, California, (1988).

20. P. Asthana and A. R. Tanguay, Jr., "Charge-Screening-Induced Switching in Spatially Multiplexed Sub-holograms in $Bi_{12}SiO_{20}$", 1988 Annual Meeting of the Optical Society of America, Santa Clara, California, (1988).

6. SCIENTIFIC PERSONNEL

The following scientific personnel were directly involved with the research described in this report, and were supported in part either by contract funds or by cost-sharing.

Dr. Armand R. Tanguay, Jr.	(Principal Investigator)
Dr. Richard V. Johnson	(Senior Investigator)
Dr. Chiung-Sheng Wu	(Graduate Research Assistant; Postdoctoral Fellow)
Praveen Asthana	(Graduate Research Assistant)
Chris Kyriakakis	(Graduate Research Assistant)
Frank Lum	(Laboratory Manager and Senior Research Technician)
Howard Chin	(Undergraduate Research Assistant)

7. REFERENCES

1. A. A. Sawchuk, T. C. Strand, and A. R. Tanguay, Jr., "Nonlinear Real-Time Optical Signal Processing", AFOSR Annual Technical Report, USCIPI Report 1100, December, (1983).

2. P. Chavel, A. A. Sawchuk, T. C. Strand, A. R. Tanguay, Jr., and B. H. Soffer, "Optical Logic with Variable-Grating- Mode Liquid-Crystal-Devices", Optics Letters, 5, 398-400, (1980).

3. B. H. Soffer, J. D. Margerum, A. M. Lackner, D. Boswell, A. A. Sawchuk, T. C. Strand, A. R. Tanguay, Jr., and P. Chavel, "Variable Grating Mode Liquid Crystal Device for Optical Processing and Computing", Molecular Crystals and Liquid Crystals, 70, 145-161, (1981).

4. A. R. Tanguay, Jr., T. C. Strand, P. Chavel, A. A. Sawchuk, and B. H. Soffer, "Theoretical and Experimental Polarization Properties of the Variable Grating Mode Liquid Crystal Structure", Journal of the Optical Society of America, 71, 1630, (1981).

5. A. R. Tanguay, Jr., "Recent Progress in Spatial Light Modulators for Coherent Optical Processing Applications", 1982 Gordon Research Conference on Holography and Optical Information Processing, Plymouth, New Hampshire, (1982).

6. P. Chavel, A. R. Tanguay, Jr., T. C. Strand, A. A. Sawchuk, and B. H. Soffer, "Method and Apparatus for Optical Computing and Logic Processing by Mapping of Input Optical Intensities Into Position of an Optical Image", United States Patent 4,351,589, September 28, (1982).

7. C. Mead and L. Conway, Introduction to VLSI Systems, Addison-Wesley, Reading, MA, (1980), pp. 333-372.

8. R. W. Keyes, The Physics of VLSI Systems, Addison-Wesley, Reading, MA, (1987).

9. R. W. Keyes, "Physical Limits in Digital Electronics", Proceedings of the IEEE, 63(5), 740-767, (1985).

10. P. W. Smith, "On the Physical Limits of Digital Optical Switching and Logic Elements", Bell Systems Technical Journal, 61, 1975-1993, (1982).

11. R. L. Fork, "Physics of optical switching", Physical Review A, 26(4), 2049-2064, (1982).

12. R. W. Keyes, "Optical logic–in the light of computer technology", Optical Acta, 32(5), 525-535, (1985).

13. R. Landauer, "Optical Logic and Optically Accessed Digital Storage", in *Optical Information Processing*, edited by Yu. E. Nesterikhin, G. W. Stroke, and W. E. Kock, Plenum Press, New York, (1976), pp. 219-253.

14. R. W. Keyes and J. A. Armstrong, "Thermal Limitations in Optical Logic", Applied Optics, 8(12), 2549-2552, (1969).

15. H. Kogelnik, "Limits in Integrated Optics", Proceedings of the IEEE, 69(2), 232-238, (1981).

16. Yu. P. Bobylev and S. A. Pikin, "Threshold Piezoelectric Instability in a Liquid Crystal", Sov. Phys. - JETP, 45(1), 195-198, (1977).

APPENDIX 1

Physical characterization of the variable grating mode liquid crystal device

A. R. Tanguay, Jr.
C. S. Wu
P. Chavel*
T. C. Strand†
A. A. Sawchuk
University of Southern California
Image Processing Institute, and Departments of
Electrical Engineering and Materials Science
Los Angeles, California 90089-0483

B. H. Soffer
Hughes Research Laboratories
3011 Malibu Canyon Road
Malibu, California 90265

Abstract. The physical principles of operation of the variable grating mode (VGM) liquid crystal device are described. The VGM device is capable of performing a two-dimensional intensity-to-spatial frequency conversion, which in turn allows the implementation of a wide range of nonlinear optical processing and computing functions. The device utilizes certain nematic liquid crystal mixtures that are observed to form variable frequency diffraction gratings under the influence of an applied bias voltage. Both fundamental and technological limitations to device performance characteristics are discussed.

Keywords: spatial light modulator; variable grating mode; liquid crystal device; optical information processing; optical devices.

Optical Engineering 22(6), 687-694 (November/December 1983).

CONTENTS

1. INTRODUCTION

A wide variety of one- and two-dimensional operations are necessary for full-scale implementation of parallel optical processing and computing systems. Incoherent-to-coherent conversions are often required for algorithms involving spectrum analysis and modification, correlation, convolution, and holographic image formation, particularly when the information to be processed is available only in time-sequential or matrix-addressed raster format. A number of one- and two-dimensional spatial light modulators capable of this type of image transduction are described within this special issue of *Optical Engineering*,[1-4] as well as in several review articles.[5-8]

Other, equally important processing and computing functions, such as logic operations, programmable matrix addressing, binary addition, linearity compensation, and input-output nonlinearities (e.g., exponentials, logarithms, power laws, thresholds, level slices, and level restoration), have proven particularly difficult to implement. All of these functions, on the other hand, can be implemented by means of some form of intensity-to-position encoding in conjunction with either fixed (single function) or programmable (multifunction) masks. This general statement follows from the realization that all of the functions listed above are special cases of data-dependent multiplications, in which the input value (e.g., pixel intensity) selects the appropriate multiplier (e.g., mask location) to obtain the desired product (e.g., output intensity).

The variable grating mode liquid crystal device (VGM LCD)[9-12] transforms input intensities to spatial positions when used in conjunction with a Fourier transform lens. The nature of this image transformation can be realized in the following manner. The VGM LCD primarily consists of a photoconductive layer in series with a layer of nematic liquid crystal mixture. A dc bias voltage is applied across the device to provide a voltage division between the two layers. Within a given image pixel, the input intensity decays the voltage across the photoconductive layer and correspondingly enhances the voltage across the liquid crystal layer. The photoconductor thus implements an intensity-to-voltage conversion. The nematic liquid crystal mixture employed in the device has the unusual property that the alignment of the liquid crystal molecules, which is homogeneous in the quiescent state, exhibits spatially periodic modulation when a bias voltage is applied across the layer. This modulation results in a birefringent phase grating[13] characterized by a spatial frequency that depends linearly on the applied voltage. The effect of the liquid crystal layer is thus to implement a voltage-to-spatial frequency conversion. If both layers are considered together, the entire device is thus seen to perform an image-wise intensity-to-spatial frequency conversion, which can be modified to the more general intensity-to-position transformation by placing a Fourier transform lens behind the VGM LCD. Collimated readout illumination normally incident on the device (at a wavelength of photoconductive insensitivity) is angle encoded within each image pixel by diffraction from each induced phase grating and subsequently angle-to-position mapped by the Fourier transform lens into its focal plane.

This type of process is shown schematically in Fig. 1, in which the input image is assumed to consist of two separate regions of differing intensity. The VGM LCD encodes both regions with different spatial frequencies, resulting in separated diffraction orders in the filter (Fourier) plane. Insertion of an appropriate spatial filter or programmable mask (not shown in Fig. 1) into the Fourier plane allows the separated orders to be selectively modified to implement any desired data-dependent multiplication or point nonlinearity. In the reconstructed output image, all regions of equal input intensity are modified identically, irrespective of their location in the input image field. Thus, all of the data-dependent multiplications are performed in parallel. Functional programmability is achieved by replacement or reprogramming of the Fourier plane mask, which need only be a low resolution device with a total number of resolution elements equal to the number of gray levels required to be processed.

The overall input-output characteristic of nonlinear function

* *Permanent address: Institut d'Optique, Université de Paris sud, BP 43, 91406 Orsay Cedex, France.*

†Present address: IBM Corporation, 5600 Cottle Rd., San Jose, CA 95193.

Invited Paper LM-104 received July 15, 1983; revised manuscript received July 28, 1983; accepted for publication Aug. 7, 1983; received by Managing Editor Aug. 29, 1983. This paper is a revision of Paper 388-12 which was presented at the SPIE conference on Advances in Optical Information Processing, Jan. 20-21, 1983, Los Angeles, CA. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 388.

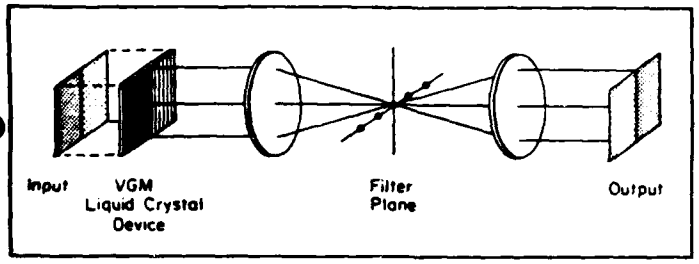


Fig. 1. Experimental arrangement for demonstration of intensity-to-position encoding by means of an intensity-to-spatial frequency conversion in a VGM LCD.

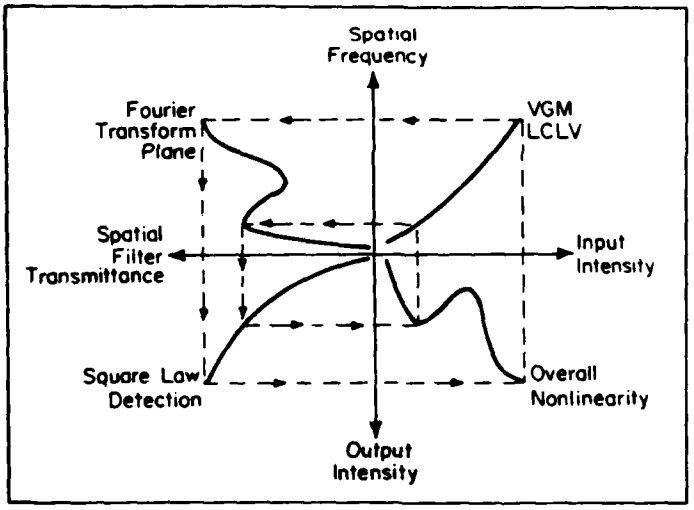


Fig. 2. VGM nonlinear processing. The overall input-output characteristic can be found by stepping through the successive nonlinear transformations, including (1) the intensity-to-spatial frequency conversion, (2) spatial filtering, and (3) intensity detection.

implementation utilizing the VGM LCD is shown schematically in Fig. 2 by specifying sequentially the nature of the transformations from input intensity to spatial frequency within the VGM LCD, from spatial frequency to spatial filter amplitude transmittance in the Fourier transform plane, and from output amplitude to output intensity (usually by means of square law detection). The overall nonlinearity achieved can be easily compensated for the functional dependences of the separate steps by adjustment of the selected spatial filter transmittance function.

The VGM LCD has thus far been utilized to perform a wide variety of parallel nonlinear point transformations, including level slicing,[9,10,12] binary logic functions (AND, OR, NOR, etc.),[10-12] and full binary addition (inputs: two addend bit planes and one carry bit plane; outputs: sum bit plane and carry bit plane).[11] The purpose of this paper is to describe the physical principles of operation of the variable grating mode liquid crystal device, identify areas of strength and weakness, and differentiate limitations to current device performance thought to be fundamental in origin from those that are seemingly technological. Section 2 consists of a more detailed description of the device, its operating mode, and its operational properties. The fundamental origins of these operational properties are examined in Sec. 3, in which the natural focus will be the physical mechanism of the variable grating mode effect in nematic liquid crystal mixtures. Experimental and theoretical efforts to elucidate the nature of this mechanism are described in Sec. 4, which concludes with several important but as yet unanswered questions.

2. DEVICE DESCRIPTION AND OPERATIONAL MODE

A number of important aspects of device construction and device operation are reviewed in this section. A more complete description

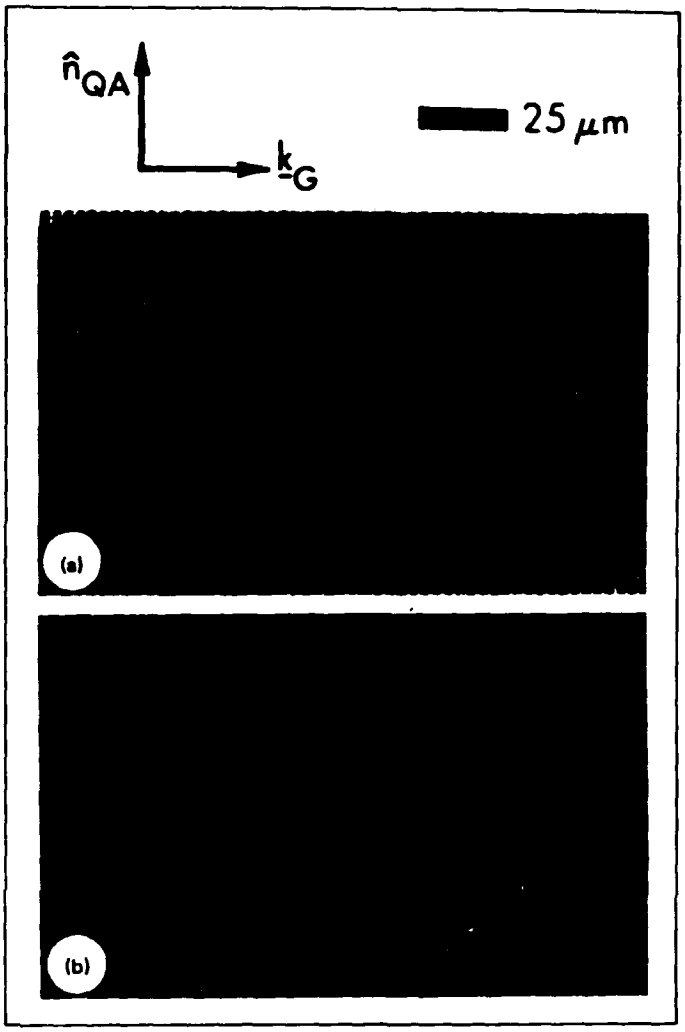


Fig. 3. Polarization photomicrographs of the liquid crystal domain pattern in an electrically activated cell. In (a), the polarizer was oriented at 90°, and the analyzer at 90°, with respect to the grating wave vector. In (b), the polarizer was oriented at 90°, and the analyzer at 10°, with respect to the grating wave vector. The unit vector $\hat{n}_{QA}$ denotes the direction of quiescent alignment, and $\underline{k}_G$ indicates the direction of the grating wave vector.

of these concepts has appeared previously.[9-12]

The critical element of the VGM LCD is a thin (4 to 12 μm) layer of nematic liquid crystal mixture[12] that exhibits a periodic modulation of the liquid crystal director, and hence of the index ellipsoid, under application of an electric field normal to the plane of the layer. By means of suitable preferential alignment techniques,[9,12] the quiescent state of the liquid crystal is homogeneous (parallel to the plane of the layer). As will be discussed in Sec. 4, this periodic variation of the principal axes of the dielectric tensor gives rise to a birefringent phase grating characterized by striking and unique optical properties.[14] The grating can be visualized in a polarizing microscope, as shown in Fig. 3, by utilizing the birefringence properties of the periodic perturbation. Distinct polarizer/analyzer combinations give rise to remarkably different grating images, as can be seen by comparison of Figs. 3(a) and 3(b) (see Sec. 4). Furthermore, the grating period is observed experimentally to be related inversely to the applied voltage across the layer. Above the threshold for domain formation, therefore, the spatial frequency of the grating is a linear function of the voltage across the layer, as shown for a variety of nematic liquid crystal mixtures in Fig. 4.

This voltage-to-spatial frequency transformation can be optically

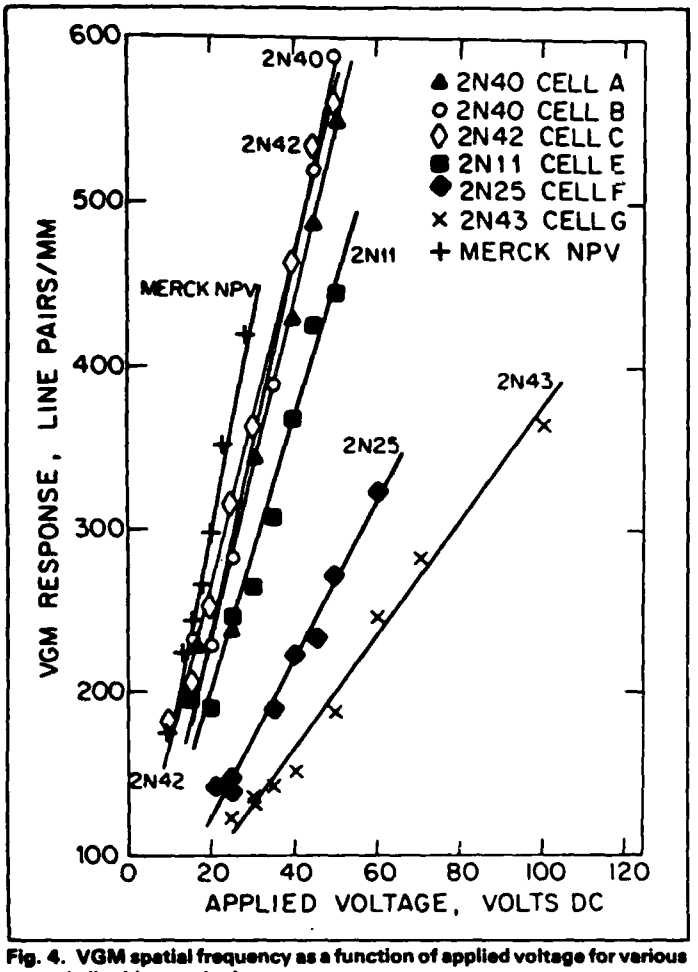


Fig. 4. VGM spatial frequency as a function of applied voltage for various nematic liquid crystal mixtures.

addressed by means of a photoconductor placed in series with the liquid crystal layer, as shown in Fig. 5 and described in the previous section. The photoconductive layer employed in devices constructed thus far is comprised of evaporated or ion-beam sputtered zinc sulfide (ZnS), chosen to optimize the impedance match with the liquid crystal layer ($\rho > 10^{10}\ \Omega \cdot$ cm). The layer thicknesses employed were of order 1.5 to 5 μm. As shown in Fig. 5, the photoconductive and liquid crystal layers are sandwiched between indium tin oxide (ITO)-coated 1.2 cm thick glass optical flats. The liquid crystal layer thickness is determined by a perimeter Mylar spacer.

In operation, a dc bias voltage is applied between the indium tin oxide electrodes, of order 40 to 150 V. The input image to be spatial frequency encoded is focused on the ZnS photoconductor, producing image-wise modulation of the local voltage across the liquid crystal layer, thus effecting a parallel intensity-to-spatial frequency conversion. The high lateral impedance of the thin film layers allows high resolution images to be processed with low pixel-to-pixel crosstalk. The device sensitivity is optimized for exposure at blue and near-ultraviolet wavelengths due to the peak photosensitivity of zinc sulfide in that spectral region. Quasi-nondestructive readout can be accomplished at wavelengths beyond the photoconductivity edge, such as that of the He-Ne laser (6328 Å). Image erasure occurs with removal of the input image, within the dielectric and liquid crystal relaxation times of the device (see Sec. 3). To date, all VGM LCDs that we have constructed have been designed for transmissive readout, although reflective readout is possible with incorporation of an appropriate dielectric mirror. Such a configuration would have the advantage of fully separating the reading and writing functions, allowing for increased effective optical gain.

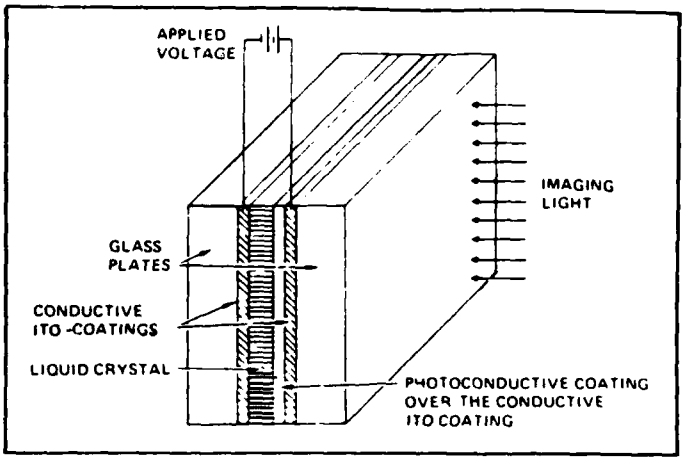


Fig. 5. Schematic diagram of the VGM liquid crystal device. Current devices are read out in transmission at a wavelength of photoconductive insensitivity.

3. FUNDAMENTAL ORIGINS OF THE OPERATIONAL PROPERTIES

As was mentioned in Sec. 2, the operational properties of the VGM LCD are primarily determined by the variable grating mode effect exhibited by the nematic mixture liquid crystal layer. The current state of knowledge concerning the physical origin of this unique effect is summarized in Sec. 4. In this section, we outline a number of key factors and considerations that affect several important device properties in order to provide both focus and a frame of reference for the succeeding section. These device properties include the accessible range of spatial frequencies, the number of accessible gray levels, the functional dependence of diffraction efficiency on applied voltage, maximum diffraction efficiency, response time, device uniformity, device input sensitivity, and device operational lifetime.

The accessible range of spatial frequencies extends from the threshold for grating formation at the low end to the onset of dynamic scattering induced by high electric fields at the high end, as shown in Fig. 4. For phenyl benzoate mixtures with slightly negative dielectric anisotropy (<-0.30) such as HRL 2N40,[15] this range extends from approximately 200 line pairs/mm to over 600 line pairs/mm. In order to avoid overlap of higher diffracted orders from lower spatial frequencies with lower diffracted orders from higher spatial frequencies, the maximum range that can be processed (uniquely assigned to specific gray levels) spans a factor of two in spatial frequency. For example, a usable range in HRL 2N40 extends from 300 line pairs/mm to 600 line pairs/mm without order overlap. This accessible range can be extended by an additional factor of two by utilizing the orthogonal polarization behavior of alternating diffracted orders, as described in Sec. 4. Hence, the accessible range of spatial frequencies observed in several of the nematic liquid crystal mixtures tested so far is sufficient for optimized gray scale processing. It should be noted that although the maximum number of resolution elements that can be processed is linearly proportional to the highest spatial frequency utilized for devices of a given size (see discussion below), use of significantly larger spatial frequencies begins to place stringent requirements on the Fourier transform lens due to f-number reduction. For example, to utilize a spatial frequency range of 600 line pairs/mm requires the output optics to have an f-number less than 1.2 (or less than 2.6 if the lens is displaced off-axis to accept only the positive diffracted orders). This is primarily a pragmatic limitation rather than a fundamental one, as VGM effects have been observed at spatial frequencies exceeding 1000 line pairs/mm.[16]

The number of accessible gray levels that lead to well-separated diffraction orders in the filter plane is limited by the ratio of the frequency range between VGM harmonics to the object spectrum bandwidth, as shown in Fig. 6. The object spectrum bandwidth is in turn limited primarily by two effects: spot size due to diffraction from

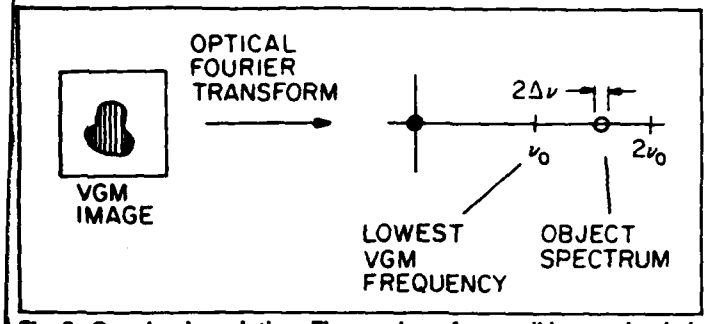


Fig. 6. Gray level resolution. The number of accessible gray levels is limited by the ratio of the separation between VGM harmonics to the object spectrum bandwidth.

finite-sized pixel apertures, and grating imperfections that cause local deviations from uniform spatial frequency. The first effect is fundamental and has been treated previously.[10] The principal result of this analysis is the inequality

$$b\nu_0 \geq 2N \ , \qquad (1)$$

in which b is the pixel width, ν_0 is the lowest usable VGM spatial frequency, and N is the desired number of distinguishable gray levels. This inequality requires that the pixel size contain at least 2N periods of the lowest grating frequency if N gray levels are to be processed. For example, a 256×256 pixel image could be processed with 32 distinguishable gray levels on a 50 mm square device with $\nu_0 = 300$ line pairs/mm. An additional restriction arises from scattering effects due to grating imperfections, which tend to further increase the size of the diffracted orders. The most common type of imperfection observed in these devices is the joining or splitting of grating lines, as shown in the photomicrograph in Fig. 7. The origin of these "disclinations" is not at present understood, although the density of occurrence of such imperfections is directly related to the quality of substrate preparation.

A typical measurement of the functional dependence of diffraction efficiency on the applied voltage across the VGM liquid crystal layer is shown in Fig. 8. Since the applied voltage is linearly related to the induced grating spatial frequency, this relationship is illustrative of the dependence of the diffraction efficiency on spatial frequency as well. To first order, the nature of this dependence is not important to the implementation of optical processing functions since any variation in diffraction efficiency with spatial frequency can be linearized by insertion of an appropriate multiplicative filter in the focal plane of the Fourier transform lens. In any case, the theoretical functional dependence can be derived only from knowledge of the relationship between the induced orientational angles of the liquid crystal director and the applied voltage across the layer. This relationship is discussed further in the subsequent section.

The maximum diffraction efficiency that can be achieved at a given spatial frequency and applied voltage depends fundamentally on the magnitude of the anisotropy in the index of refraction ($\Delta n = n_e - n_o$, with n_e the extraordinary refractive index for polarization parallel to the molecular axis and n_o the ordinary refractive index for polarization perpendicular to the molecular axis), the magnitude of the periodic angular reorientation of the liquid crystal director, and the thickness of the VGM liquid crystal layer. Full periodic reorientation of the index ellipsoid from homogeneous (parallel to the substrate) alignment to homeotropic (perpendicular) alignment for HRL 2N40 ($\Delta n = 0.15$) in a 6 μm cell read out at 6328 Å gives rise to an optical phase modulation of approximately 9 rad. Hence, the maximum diffraction efficiency is fundamentally limited to that expected for a pure sinusoidal phase grating.[17] In practice, full reorientation is typically not achieved before the onset of dynamic scattering, although reorientation angles of 45° are thought at present to be commonly reached (see Sec. 4). As can be seen from Fig. 8, typical second-order diffraction efficiencies are of order 20%. This

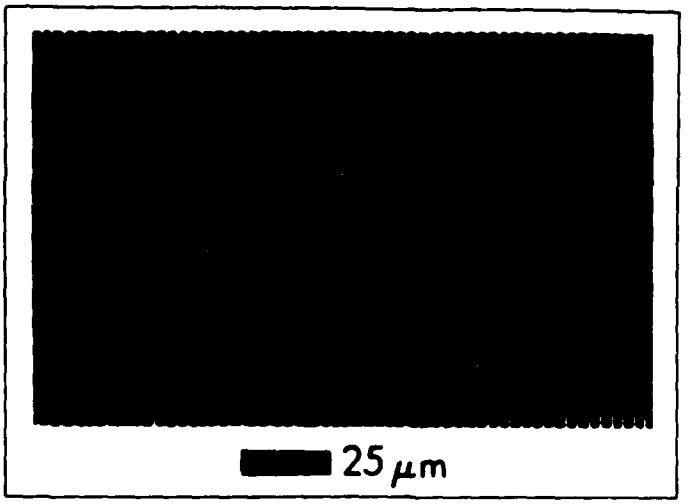


Fig. 7. Polarization photomicrograph of a VGM liquid crystal layer exhibiting a number of grating discontinuities (circled).

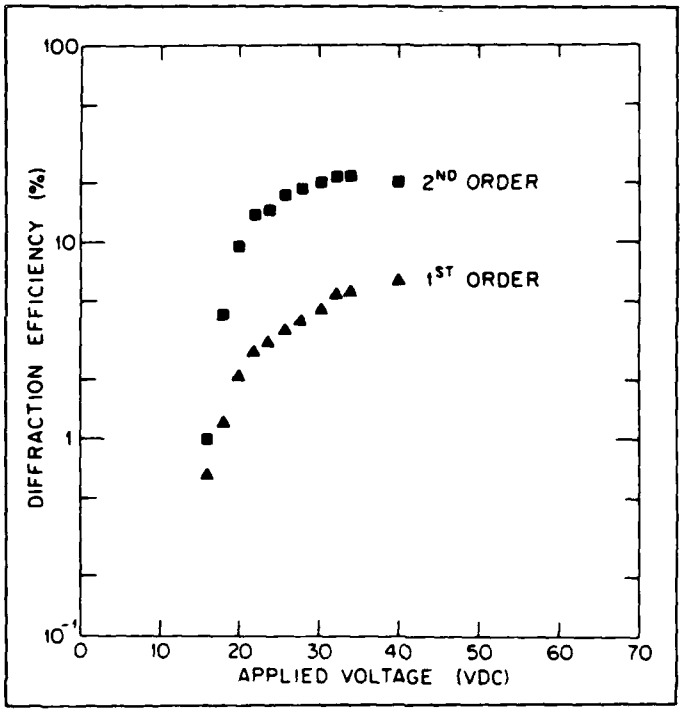


Fig. 8. Diffraction efficiency as a function of applied voltage across a nematic liquid crystal mixture of phenyl benzoates (HRL 2N40). The layer thickness was approximately 8 μm, as defined by a perimeter Mylar spacer.

order is larger than the first diffracted order due to the peculiar nature of the birefringent phase grating formed by the VGM distortion (details are given for this phenomenon in Sec. 4).

The response time of the cell is a critical parameter that directly affects the achievable overall processing throughput rate. At present it is the major factor inhibiting widespread incorporation of VGM devices in optical processing systems. The rise time for grating formation from below to above threshold varies from mixture to mixture, but is typically of order one second. The response time for grating change in response to a step increase in applied voltage (corresponding to a step increase in grating spatial frequency) is typically of order a fraction of a second.[18] The response time of the photoconductive voltage division across the liquid crystal layer is not a significant factor by orders of magnitude relative to the reorien-

tation response time, so that eventual improvements in VGM LCD response time will accrue only by advances in the state of understanding of the physical origin of the VGM effect and the dynamical nature of the grating reorientation process, followed by appropriate modification of the device operational mode to enhance the rate of molecular reorientation and/or a search for a nematic liquid crystal mixture with physical characteristics optimized for dynamic VGM effects.

The input sensitivity of the VGM LCD, defined as the input (writing) intensity per unit area per unit change in grating spatial frequency, is determined by a number of factors. These include the slope of induced grating spatial frequency as a function of applied voltage for the particular nematic liquid crystal mixture employed, the wavelength dependence of the photoconductive layer photosensitivity, and the cell switching ratio (fractional increase in voltage across the liquid crystal layer from illumination at the threshold for grating formation to saturation). The first factor (liquid crystal response slope) varies significantly from mixture to mixture (see Fig. 4). Of the VGM nematic liquid crystal mixtures investigated to date, HRL 2N40 has proved to be nearly optimum in this regard. It is not yet clear what fundamentally influences and eventually limits this parameter. The photosensitivity of the photoconductive layer is determined primarily by the choice of photoconductive material (limited to those that can be appropriately impedance-matched to the liquid crystal layer), method and quality of thin film deposition, layer thickness, spectral width and central wavelength of the exposure (writing) illumination, and the operational bias voltage employed. The cell switching ratio is a function of the series impedance of the liquid crystal layer, the impedance of the unilluminated photoconductive layer, and the impedance of the photoconductive layer under saturation illumination. In addition, the cell switching ratio will be altered by incorporation of surfactant layers to improve liquid crystal quiescent alignment, and of a dielectric mirror in the reflective readout device structure. At this stage of the device development, the input sensitivity of the VGM LCD has not been optimized. A typical value of 15 ($\mu W/cm^2$) · (mm/line pair) was obtained with a VGM LCD consisting of a 6 μm layer of HRL 2N40 in series with a 5 μm thick evaporated ZnS layer that had been polished and rubbed with surfactant polyvinyl alcohol, operated at 160 V dc, and illuminated in the passband 410 to 550 nm.[9]

The uniformity of VGM LCD response depends inherently on technological issues, including uniformity of layer thicknesses, homogeneous mixing of the liquid crystal material employed, and the as-deposited spatial dependence of photoconductive sensitivity. Whereas it is relatively straightforward to construct electrically activated VGM cells (see Sec. 4) that exhibit a high degree of spatial uniformity, deposition of a photoconductive layer with equivalent spatial homogeneity has proven more difficult. Nonuniformity of the device response characteristic can be a contributing factor in the establishment of the maximum number of accessible gray levels discussed previously. In experimental devices constructed thus far, device uniformity has not proven to be the limiting factor. In any case, it is expected that response nonuniformities can be minimized significantly by improvements in the photoconductive layer deposition process.

The lifetimes of experimentally constructed VGM LCDs have ranged from less than a week to over a year. The causes of VGM device failure have not yet been extensively studied, although several contributing factors can be identified. These factors include the purity and composition of the liquid crystal mixture employed, the nature of the liquid crystal/photoconductive layer interface, the integrity of the device sealing process, and the device operational history. Since the VGM effect requires a dc applied voltage, unidirectional ion poisoning may contribute to gradual device degradation.

4. PHYSICAL ORIGIN OF THE VARIABLE GRATING MODE EFFECT

As can be clearly understood from the discussion presented in Sec. 3, a vast majority of the important device operational properties depend critically on the detailed nature of the grating formation and dynamic

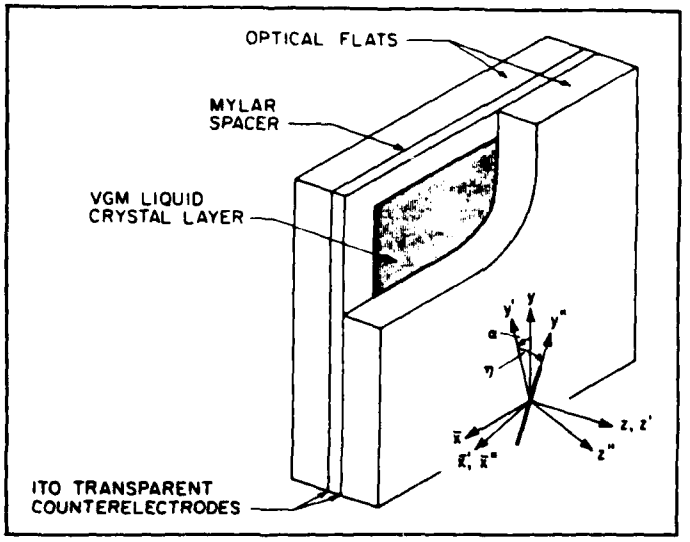


Fig. 9. Variable grating mode liquid crystal test geometry showing the Cartesian coordinate system referred to in the text as well as the molecular orientation angles α and η. This configuration was utilized in the polarized light diffraction efficiency and photomicroscopy experiments.

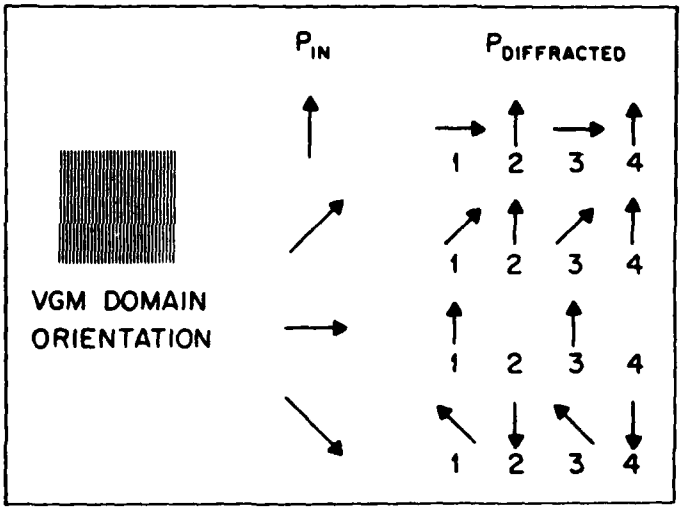


Fig. 10. The polarization behavior of VGM diffracted orders, with illumination normal to the plane of the liquid crystal layer. The left-hand column indicates the input polarization associated with each row of output polarizations. The inset shows the corresponding orientation of the VGM grating.

reorientation process. The elucidation of the fundamental nature of the grating (molecular orientation angles as a function of applied voltage across the liquid crystal layer) and of the physical mechanism that gives rise to the observed periodic instability and its reorientation dynamics has been a subject of considerable experimental and theoretical interest.[11,16,19–30] In this section, the present state of understanding of the VGM effect is described, and a number of important unresolved questions are presented.

The grating produced by the periodic spatial reorientation of the nematic liquid crystal molecules is quite unusual, giving rise to striking polarization-dependent properties.[11] These diffraction effects can be investigated in an electrically activated VGM cell with no intervening photoconductive layer, as shown in Fig. 9. The orientation of the grating is such that the grating wave vector is perpendicular to the direction of unperturbed alignment, which is homogeneous and induced by unidirectional rubbing or ion-beam milling. That is, the periodic modulation direction is perpendicular to the initial (zero applied bias) liquid crystal director (long molecular axis), as shown

in the polarization micrographs (Fig. 3).

For all linear input polarization angles, the even and odd diffraction orders are found to be essentially linearly polarized. In addition, the even diffraction orders are nearly linearly polarized parallel to the "domains" comprising the VGM grating, as shown in Fig. 10. For input polarization perpendicular to the domains, the even orders are found to be almost fully extinguished. On the other hand, the odd diffraction orders are nearly linearly polarized with a major axis that rotates counterclockwise at the same rate as the input polarization is rotated clockwise. This effect is the same as that produced by a half-wave plate oriented at 45° with respect to the grating wave vector. For input polarization at 45° to the wave vector, all orders are observed in the far-field diffraction pattern. An analyzer placed on the output side of the VGM device can be rotated to extinguish the even orders (when oriented parallel to the grating wave vector) or the odd orders (when oriented at −45° to the grating wave vector).

These unusual polarization properties have recently been utilized to determine the spatial distribution of the molecular orientation within the VGM liquid crystal layer.[11] The polarization dependence of the diffraction phenomena is directly related to the formation of a birefringent phase grating,[13] in which the principal axis of the index ellipsoid varies periodically both in the plane of the grating (characterized by an orientation angle α) and normal to the plane of the grating (characterized by an orientation angle η). The angular coordinates are as shown in Fig. 9.

The polarization properties of light diffracted by the liquid crystal birefringent phase grating can be summarized by means of a transfer matrix that connects the output polarization at each point (x,y) on the rear surface of the liquid crystal layer with the input polarization at the front surface of the liquid crystal layer. On the basis of the experimental observations, this matrix must be of the form

$$\begin{bmatrix} A_0 + A(x;p) & B(x;2p) \\ C(x;2p) & D_0 + D(x;p) \end{bmatrix}, \tag{2}$$

in which the notation A(x;p) indicates that the complex amplitude A varies in the x direction with periodic repetition distance p. For uniaxial liquid crystal molecules at an arbitrary orientation (α,η), assumed uniform throughout the layer thickness at a given coordinate in the x direction, the Jones matrix can be determined by appropriate rotations of the index ellipsoid, which yields

$$\begin{bmatrix} 1 - \sin^2\alpha(1 - e^{j\phi}) & \sin\alpha\cos\alpha(1 - e^{j\phi}) \\ \sin\alpha\cos\alpha(1 - e^{j\phi}) & 1 - \cos^2\alpha(1 - e^{j\phi}) \end{bmatrix}, \tag{3}$$

where

$$\phi = \frac{2\pi t}{\lambda}\left[\left(\frac{\sin^2\eta}{n_o^2} + \frac{\cos^2\eta}{n_e^2}\right)^{-1/2} - n_o\right],$$

in which t is the liquid crystal layer thickness, n_o is the ordinary index of refraction, n_e is the extraordinary index of refraction, and λ is the wavelength of readout illumination employed. The angles α and η are assumed to be periodic functions of x and independent of y and z. Measurement of the intensities in each diffraction order for a minimum set of polarizer/analyzer orientations uniquely determines the magnitudes of the Fourier components of the polarization transfer matrix. These experimentally derived values can then be compared with the theoretically calculated coefficients of Eq. (3) (by harmonic expansion) for different possible assumptions concerning the spatial distribution of the orientation angles α and η. An example of such a comparison between theory and experiment is shown in Fig. 11, under the assumption that the spatial dependences of α and η are given by

$$\alpha = \alpha_o \cos\frac{2\pi x}{\Lambda}; \qquad \eta = \pm\eta_o \sin\frac{2\pi x}{\Lambda}. \tag{4}$$

This procedure allows the extraction of the maximum orientational excursion angles $\alpha_o(V)$ and $\eta_o(V)$ as functions of the applied bias voltage above the threshold for grating formation, as shown in Fig. 12(α_o) and Fig. 13(η_o). In each case, subject to the assumed forms of α and η implicit in Eq. (4), it is observed that the maximum excursion angles both in and out of the plane of the grating seem to increase as the logarithm of the applied voltage.

The spatial distribution of the ends of the liquid crystal molecules described by Eq. (4) is approximately cycloidal. Such a dependence of the angles α and η on the spatial coordinate x has been predicted by direct minimization of the free energy in a similar nematic liquid crystal system.[21] This particular solution is obtained by incorporation of the converse flexoelectric effect in the expression for the free energy.[13,31–33] The flexoelectric effect describes a strain-induced polarization that arises due to molecular shape effects in conjunction with a nonzero dipole moment, as shown schematically in Fig. 14. The converse flexoelectric effect thus pertains to a polarization-induced strain within the liquid crystal layer, which can result in a periodic molecular reorientation characterized by a linear dispersion relation between the grating wave vector and the applied field, as is observed experimentally. Including the dielectric, distortion, and flexoelectric contributions to the free energy, yields an expression of the form

$$\begin{aligned} F_{VGM} &= \int \left(F_{dielectric} + F_{distortion} + F_{flexoelectric}\right) dV \\ &= F_o - \frac{1}{8\pi}\int (\epsilon_e - \epsilon_o)(\underline{E}\cdot\hat{n})^2 dV \\ &\quad + \frac{1}{2}\int \Big[K_1(\nabla\cdot\hat{n})^2 + K_2(\hat{n}\cdot\nabla\times\hat{n})^2 \\ &\qquad + K_3(\hat{n}\times\nabla\times\hat{n})^2\Big] dV \\ &\quad - \int \Big[e_1(\nabla\cdot\hat{n})(\hat{n}\cdot\underline{E}) \\ &\qquad + e_3\left\{\left[(\nabla\times\hat{n})\times\hat{n}\right]\cdot\underline{E}\right\}\Big] dV, \end{aligned} \tag{5}$$

in which K_1, K_2, and K_3 are the elastic constants for splay, twist, and bend deformations, respectively; $\hat{n}$ is the liquid crystal director; $\underline{E}$ is the applied electric field; ϵ_e and ϵ_o are principal components of the dielectric tensor of the liquid crystal; and e_1 and e_3 are flexoelectric coefficients.[31] Minimization of F_{VGM} with respect to the orientation angles α and η of the director, subject to fully pinned boundary conditions at both substrate surfaces, and with the simplifying assumption that $K_1 = K_2 = K$, yields[21]

$$\alpha = \alpha_o\cos(kx)\cos(\pi z/t); \qquad \eta = \eta_o\sin(kx)\cos(\pi z/t), \tag{6}$$

in which k is the grating wave vector, and t is the liquid crystal layer thickness. This solution generates a dispersion relation between E and k of the form

$$E^2 = \left(\frac{K}{e^*}\right)^2 \cdot \frac{\left[k^2 + (\pi/t)^2\right]^2}{k^2 + \mu\left[k^2 + (\pi/t)^2\right]}. \tag{7}$$

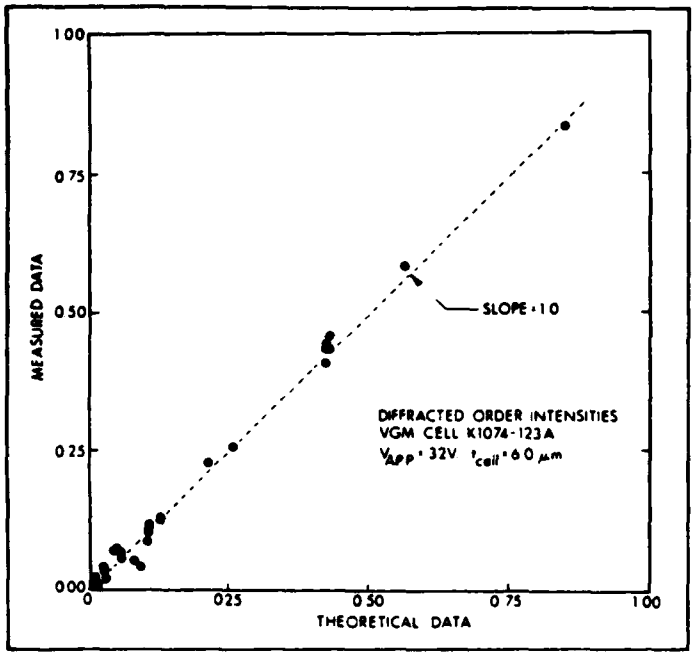


Fig. 11. Measured diffracted order intensities as a function of theoretical intensities calculated from the uniaxial VGM model described in the text.

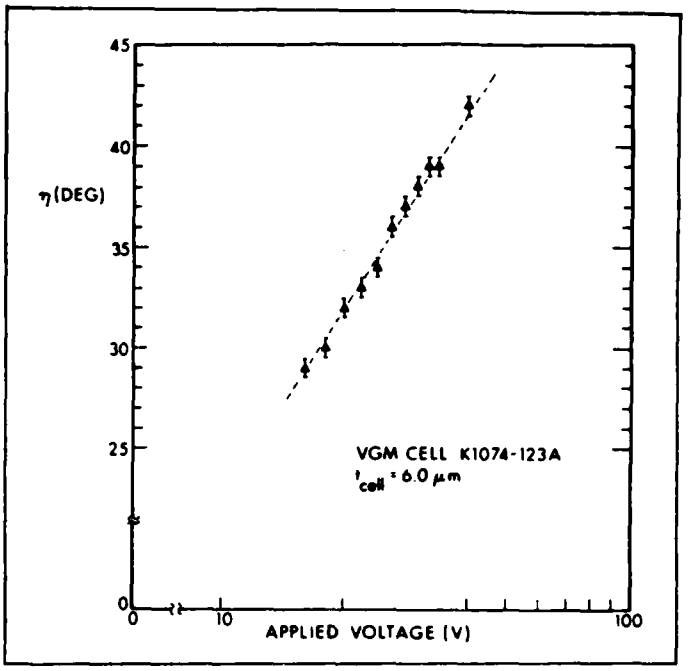


Fig. 13. The out-of-plane molecular orientation angle (η_0) as a function of the applied dc bias voltage across the cell (V) (see Fig. 5).

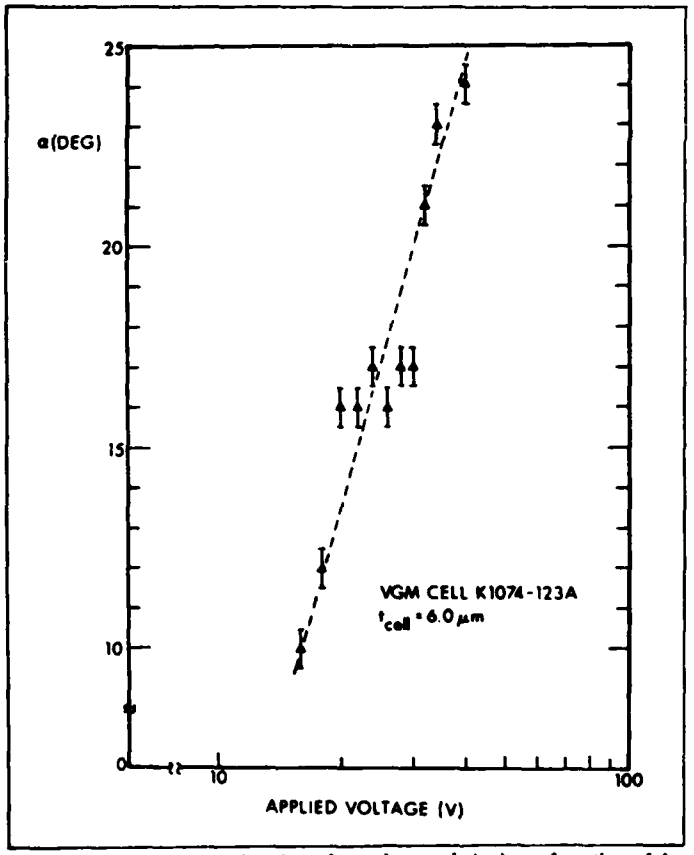


Fig. 12. The in-plane molecular orientation angle (α_0) as a function of the applied dc bias voltage across the cell (V) (see Fig. 5).

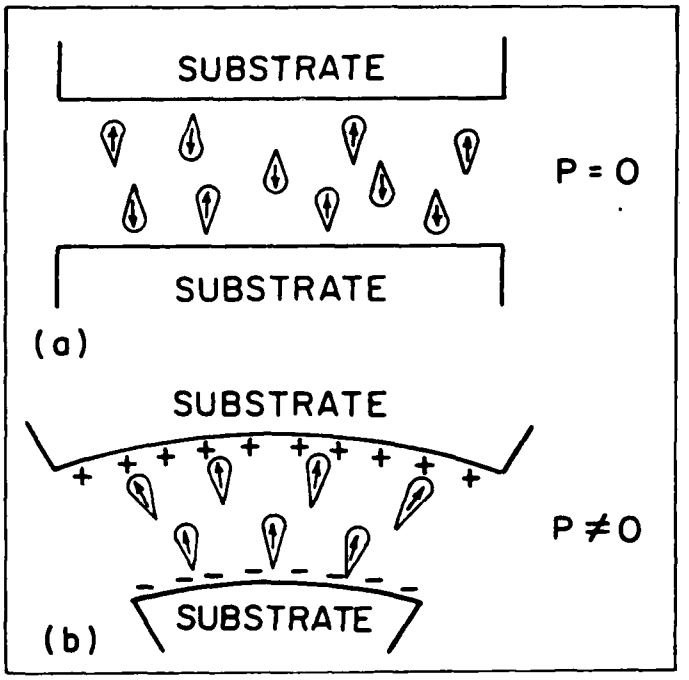


Fig. 14. Schematic diagram illustrating a possible mechanism for the occurrence of the flexoelectric effect due to a shape anisotropy in liquid crystal molecules with a permanent dipole moment. In (a), the molecular orientations are randomly distributed, resulting in zero net polarization. In (b), the applied distortion induces a shape-dependent molecular realignment, resulting in a net polarization. (After Ref. 31.)

in which $e^* = e_1 - e_3$, $\mu = (\epsilon_a K/4\pi e^{*2})$, and $\epsilon_a = \epsilon_e - \epsilon_o$. The dispersion relation is linear when $k \gg \pi/t$, as shown in Fig. 15 (compare with the experimental relationship shown in Fig. 4).

The origin of the periodic instability in VGM liquid crystals has not yet been established beyond doubt, although the accumulated evidence points strongly toward the converse flexoelectric effect. This assignment is also intuitively appealing since the VGM effect is observed only in liquid crystal mixtures with slightly negative dielectric anisotropy,[12] for which both the dielectric and distortion contributions to the free energy *increase* for deviations from uniform alignment. The addition of the flexoelectric term counteracts these

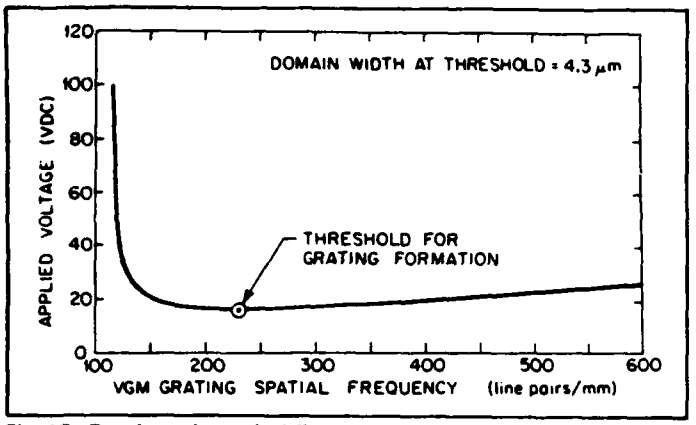


Fig. 15. Resultant theoretical dispersion relation for liquid crystal parameters corresponding to HRL 2N40 phenyl benzoate mixture [see Eq. (7)], subject to the assumption that $K_1 = K_2 = K$.

effects, producing a free-energy minimum at nonzero distortion angles. Furthermore, this model predicts a static periodic perturbation; no hydrodynamic or electrohydrodynamic effects are observed in our test cells.

The dynamics of grating formation, reorientation, and relaxation are subjects of ongoing experimental and theoretical investigation. Understanding of the basic physical principles underlying these effects is vital to the success of efforts to improve the VGM LCD response time.

A number of spatial discontinuities which are similar in appearance to crystallographic dislocations can be observed within the liquid crystal layer (see Fig. 7). The "dislocation" density increases with time as each VGM device deteriorates, and strongly depends on the manner in which the applied bias is brought from below to above the threshold of grating formation. In some cases, as the applied bias across an electrically activated cell is increased or decreased, the dislocations propagate past each other while the grating period decreases or increases, respectively. Dislocations with opposite orientations propagate in opposite directions until a new stable equilibrium situation is achieved. It is not yet clear whether these dislocations represent true disclinations, for which the liquid crystal director is discontinuous near the defect, or whether they are in fact alternative local continuous solutions to the free-energy minimization, perhaps induced by an as yet undiscovered perturbation.

A large number of research directions have been described in the foregoing paragraphs. In addition, several others are worthy of note. First, numerous experimental measurements of the off-diagonal elements in the Jones matrix describing polarized light propagation through the VGM cell have revealed a significant asymmetry not predicted by the uniaxial model. The origin of this "B/C" asymmetry effect has not yet been elucidated. Second, the thickness dependence of the molecular orientation angles α and η has not been fully established. The solution proposed[21] in Eq. (6) assumes full surface pinning at the substrate boundaries, and represents the lowest order z-dependent mode. The polarization-dependent diffracted order measurements described herein do not provide clear differentiation between a uniform z-dependence and the lowest order mode. Third, the nature of the transient solutions to the free-energy minimization incorporating molecular dynamic and viscosity effects has not been treated. Solution of this problem is key to response time optimization of the VGM liquid crystal device. Fourth, it is not yet clear whether a nematic liquid crystal mixture can be found that exhibits a useful VGM effect under ac bias. Such a mixture may provide significant immunity from long-term ion-poisoning effects. Finally, it should be emphasized that the variable grating mode liquid crystal effect provides only one possible means of achieving parallel intensity-to-position encoding. The processing potential of the intensity-to-position algorithm should provide more than adequate inducement to intensify the search for alternative implementations.

5. ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of D. Boswell, A. M. Lackner, and J. D. Margerum (Hughes Research Laboratories), and K. Sherman and G. Edwards (University of Southern California). This research was supported by the U.S. Air Force Office of Scientific Research, Electronics and Solid State Sciences Division, under grant AFOSR-81-0082 at the University of Southern California and contract F49620-81-C-0086 at Hughes Research Laboratories.

6. REFERENCES

1. R. V. Johnson, D. L. Hecht, R. A. Sprague, L. N. Flores, D. L. Steinmetz, and W. D. Turner, Opt. Eng. 22(6), 665 (1983).
2. D. L. Pape and L. J. Hornbeck, Opt. Eng. 22(6), 675 (1983).
3. U. Efron, P.O. Braatz, M. J. Little, R. N. Schwartz, and J. Grinberg, Opt. Eng. 22(6), 682 (1983).
4. C. Warde and J. Thackara, Opt. Eng. 22(6), 695 (1983).
5. D. Casasent, Proc. IEEE 65, 143(1977).
6. S. Lipson, "Recyclable Incoherent-to-Coherent Image Converters," in *Advances in Holography*, N. Farhat, Ed., Marcel Dekker, New York (1979).
7. G. Knight, "Interface Devices and Memory Materials," in *Optical Data Processing*, S. H. Lee, Ed., Springer-Verlag, Heidelberg (1976).
8. A. R. Tanguay, Jr., Proc. ARO Workshop on Future Directions for Optical Information Processing, Texas Tech University, Lubbock, Texas, 52(1980).
9. B. H. Soffer, D. Boswell, A. M. Lackner, A. R. Tanguay, Jr., T. C. Strand, and A. A. Sawchuk, in *Devices and Systems for Optical Signal Processing*, T. C. Strand and A. R. Tanguay, Jr., Eds., Proc. SPIE 218, 81(1980).
10. B. H. Soffer, D. Boswell, A. M. Lackner, P. Chavel, A. A. Sawchuk, T. C. Strand, and A. R. Tanguay, Jr., in *1980 International Optical Computing Conference (Book II)*, Proc. SPIE 232, 128(1980).
11. P. Chavel, A. A. Sawchuk, T. C. Strand, A. R. Tanguay, Jr., and B. H. Soffer, Opt. Lett. 5, 398(1980).
12. B. H. Soffer, J. D. Margerum, A. M. Lackner, D. Boswell, A. R. Tanguay, Jr., T. C. Strand, A. A. Sawchuk, and P. Chavel, Mol. Cryst. Liq. Cryst. 70, 145(1981).
13. A. R. Tanguay, Jr., submitted for publication in Opt. Lett.
14. A. R. Tanguay, Jr., P. Chavel, T. C. Strand, C. S. Wu, and B. H. Soffer, submitted for publication in Opt. Lett.
15. J. B. Margerum, J. E. Jensen, and A. M. Lackner, Mol. Cryst. Liq. Cryst. 68, 137(1981).
16. W. Greubel and U. Wolff, Appl. Phys. Lett. 19(7), 213(1971).
17. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York (1968).
18. B. H. Soffer, "Real-Time Implementation of Nonlinear Optical Processing Functions," Annual Technical Report on AFOSR F49620-81-C-0086, (1982).
19. R. B. Meyer, Phys. Rev. Lett. 22(18), 918(1969).
20. A. I. Derzhanski, A. G. Petrov, Chr. P. Khinov, and B. L. Markovski, Bulg. J. Phys. 2, 165(1974).
21. Yu. P. Bobylev and S. A. Pikin, Sov. Phys. JETP 45(1), 195(1977).
22. Yu. P. Bobylev, V. G. Chigrinov, and S. A. Pikin, J. de Phys. 40, Suppl. 4, C3-331 (1979).
23. S. A. Pikin, Mol. Cryst. Liq. Cryst. 63, 181(1981).
24. J. M. Pollack and J. B. Flannery, Society for Information Display 1976 Intern. Symp. Digest, 143(1976).
25. J. M. Pollack and J. B. Flannery, in *Liquid Crystals and Ordered Fluids*, J. F. Johnson and R. E. Porter, Eds., Vol. 2, pp. 557-571, Plenum Press, New York (1974).
26. J. M. Pollack and J. B. Flannery, in *Liquid Crystals and Ordered Fluids*, J. F. Johnson and R. E. Porter, Eds., Vol. 3, pp. 421-442, Plenum Press, New York (1978).
27. L. K. Vistin', Sov. Phys.-Dokl. 15(10), 908(1971).
28. L. K. Vistin', Sov. Phys.-Crystallogr. 15(3), 514(1970).
29. A. Derzhanski, A. G. Petrov, and M. D. Mitov, J. de Phys. 39, 273(1978).
30. M. I. Barnik, L. M. Blinov, A. N. Trufanov, and B. A. Umanski, J. de Phys. 39, 417(1978).
31. P. G. de Gennes, *The Physics of Liquid Crystals*, pp. 97-101, Oxford University Press, London (1974).
32. J. Prost and J. P. Marcerou, J. de Phys. 38, 315(1977).
33. J. Prost and P. S. Pershan, J. Appl. Phys. 47, 2298(1976).

APPENDIX 2

Polarization properties of the variable-grating-mode liquid-crystal device

A. R. Tanguay, Jr., P. Chavel,* T. C. Strand,† and C. S. Wu

Image Processing Institute and Departments of Electrical Engineering and Materials Science, MC-0483, University of Southern California, Los Angeles, California 90089

B. H. Soffer

Hughes Research Laboratories, 3011 Malibu Canyon Road, Malibu, California 90265

Received January 4, 1984; accepted February 17, 1984

The principal features of the liquid-crystal molecular orientation within the variable-grating-mode liquid-crystal device have been determined as a function of the applied voltage across the cell by measurement of the polarization properties of light diffracted by the liquid-crystal birefringent phase grating.

The variable-grating-mode (VGM) liquid-crystal device is a two-dimensional spatial light modulator that is capable of implementing an intensity-to-spatial-frequency conversion over an input image field.[1] In this process, the intensity variations in an input image distribution are converted to local spatial-frequency variations in a phase-grating structure within the liquid-crystal layer. As a direct result of this intensity-to-spatial-frequency transduction, programmable spatial filtering of the converted image results in selected modifications of the input intensities. By utilizing this device concept, a wide variety of optical processing and computing functions have been demonstrated, including linearity compensation and nonlinear function implementation, thresholding, level slicing, binary logic (AND, OR, NOR, etc.), full binary addition, and matrix addressing operations.[1-3] The operation of this photoactivated device was described in detail previously.[4]

An important area of research on the VGM liquid-crystal device is concerned with a fundamental understanding of the origin of the variable-grating effect, in which a thin layer of certain nematic liquid-crystal mixtures[1] sandwiched between transparent conductive coatings (Fig. 1) is observed to exhibit a periodic refractive-index anisotropy when a dc bias voltage above a threshold value is applied between the electrodes. This index modulation results in the formation of a birefringent phase grating characterized by a fundamental spatial frequency that depends on the magnitude of the voltage across the liquid-crystal layer. The orientation of the grating is such that the grating wave vector is perpendicular to the direction of unperturbed alignment, which is homogeneous and induced by unidirectional rubbing or ion-beam milling.

The polarization behavior of light diffracted from this birefringent phase grating is quite striking. For all linear input polarization angles, the even and odd diffraction orders are found to be essentially linearly polarized. In addition, the even diffraction orders are linearly polarized parallel to the domains comprising the VGM grating, as shown in Fig. 2. The even-order intensities vary continuously with the component of incident polarization parallel to the domains. For input polarization perpendicular to the domains, the even orders are found to be almost fully extinguished. On the other hand, the odd diffraction orders are linearly polarized with a major axis that rotates counterclockwise with essentially constant intensity at the same rate as the input polarization is rotated clockwise. This effect is the same as that produced by a half-wave plate oriented at 45° to the grating wave vector. For input polarization at 45° to the wave vector, all orders are observed in the far-field diffraction pattern. For this situation, an analyzer placed on the output side of the VGM device can be rotated to extinguish the even orders (when oriented parallel to the grating wave vector)

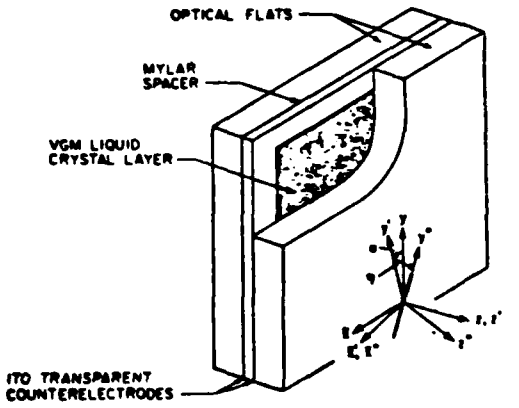


Fig. 1. VGM liquid-crystal test geometry showing the Cartesian coordinate system referred to in the text as well as the molecular orientation angles α and η. This configuration was utilized in the polarized-light diffraction efficiency and photomicroscopy experiments.

0146-9592/84/050174-03$2.00/0

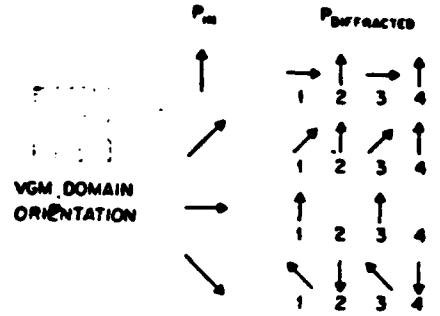


Fig. 2. The polarization behavior of VGM diffracted orders. The left-hand column indicates the input polarization associated with each row of output polarizations. The inset shows the direction of VGM-domain orientation.

or the odd orders (when oriented at −45° to the grating wave vector). Above the voltage threshold for grating formation, the intensities of the diffracted orders increase dramatically as a function of increasing applied voltage (increasing spatial frequency) and asymptotically saturate.

Observations of the VGM structure in the polarizing microscope provide correlating evidence for the diffraction phenomena described above. For input polarization perpendicular to the grating wave vector, the periodic modulation is observed through a parallel analyzer to have a principal grating period p, which corresponds to the measured diffraction angles of the even orders. For input polarization parallel to the grating wave vector, the periodic modulation is observed through a perpendicular analyzer to have a period $2p$, which corresponds to the measured diffraction angles of all orders. Other input polarization orientations produce apparent superposition of the p and $2p$ gratings, in agreement with the diffraction experiment results. These experiments are performed with the focus set at the upper surface of the liquid-crystal layer. The grating contrast can be altered, and in some cases reversed, by adjusting the focal plane to lie within or below the liquid-crystal layer.

In the thin-grating approximation, the polarization properties of light diffracted by the liquid-crystal birefringent phase grating can be summarized by means of a transfer matrix that connects the output polarization at the rear surface of the liquid-crystal layer with the input polarization at the front surface of the liquid-crystal layer. On the basis of the experimental observations, this matrix must be of the form

$$\begin{bmatrix} A_0 + A(x;p) & B(x;2p) \\ C(x;2p) & D_0 + D(x;p) \end{bmatrix}, \quad (1)$$

in which the notation $A(x;p)$ indicates that the complex amplitude A varies in the x direction with periodic repetition distance p. Fourier expansion of the transfer matrix yields the complex amplitudes A_n, B_n, C_n, and D_n, which completely specify the contributions of each component of the input polarization to the observed polarization of the nth diffracted order:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \sum_{n=-\infty}^{\infty} \exp(j2\pi x n/\Lambda) \begin{bmatrix} A_n & B_n \\ C_n & D_n \end{bmatrix}, \quad (2)$$

where $\Lambda = 2p$ is the fundamental (lowest-order) grating wavelength.

The polarization transfer matrix can be calculated by considering the uniaxial liquid-crystal molecules at a given spatial coordinate x to be characterized by a twist angle α in the plane of the grating, followed by a tilt angle η out of the plane of the grating. The Jones matrix[5] for this case can be determined by appropriate rotations of the index ellipsoid, which yield

$$\begin{bmatrix} 1 - \sin^2\alpha(1 - e^{j\phi}) & \sin\alpha\cos\alpha(1 - e^{j\phi}) \\ \sin\alpha\cos\alpha(1 - e^{j\phi}) & 1 - \cos^2\alpha(1 - e^{j\phi}) \end{bmatrix}, \quad (3)$$

where

$$\phi \equiv \frac{2\pi t}{\lambda}\left[\left(\frac{\sin^2\eta}{n_o^2} + \frac{\cos^2\eta}{n_e^2}\right)^{-1/2} - n_o\right],$$

in which t is the liquid-crystal layer thickness, n_0 is the ordinary refractive index, n_e is the extraordinary refractive index, and λ is the wavelength of the readout illumination employed. A uniform phase factor in expression (3) has been suppressed. The angles α and η are assumed to be periodic functions of x and independent of y and z; since the index ellipsoid will in general have its principal axes skewed relative to the original coordinate system, the polarization transfer matrix is not diagonal.

Measurement of the intensities in each diffraction order for a minimum set of polarizer–analyzer orientations uniquely determine the values of $|A_n|^2$, $|B_n|^2$, $|C_n|^2$, and $|D_n|^2$. These experimentally derived values can then be compared with the theoretically calculated coefficients of expression (3) [utilizing the harmonic expansion of Eq. (2)] for different possible assumptions concerning the spatial distribution of the orientation angles α and η. Simple harmonic variation of the out-of-plane angle η produces a simple phase grating that yields the polarization behavior of the even diffracted orders, while similar variation of the in-plane angle α produces a birefringent phase grating that yields the polarization behavior of the odd diffracted orders. This

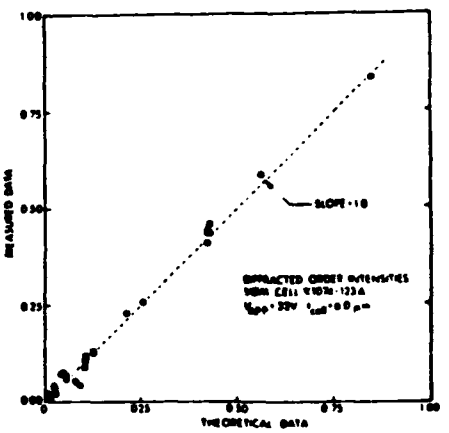


Fig. 3. Measured diffracted-order intensities for a set of polarizer–analyzer orientations as a function of theoretical intensities calculated from the uniaxial VGM model described in the text.

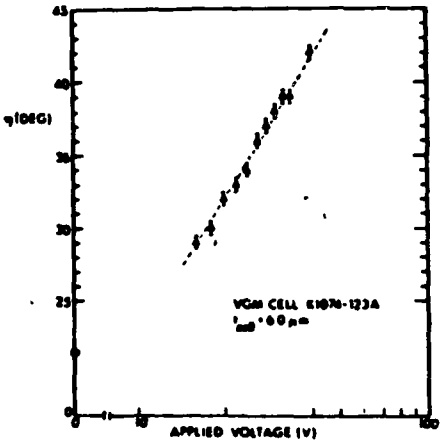


Fig. 4. The out-of-plane molecular orientation angle, η_{max}, as a function of the applied dc bias voltage across the cell, V.

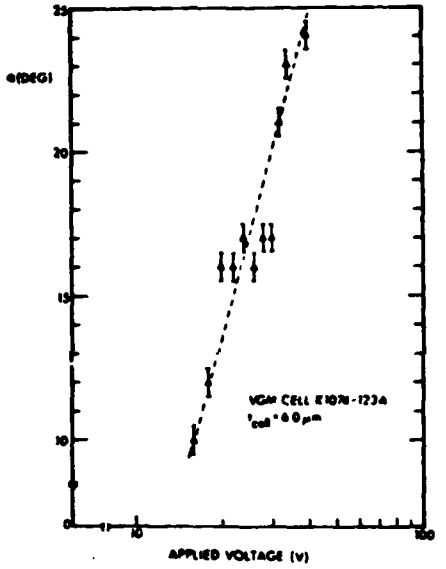


Fig. 5. The in-plane molecular orientation angle, α_{max}, as a function of the applied dc bias voltage across the cell, V.

spatial dependence can be represented by

$$\alpha = \alpha_{max} \cos \frac{2\pi x}{\Lambda}, \qquad \eta = \pm \eta_{max} \sin \frac{2\pi x}{\Lambda}. \quad (4)$$

The spatial distribution in the x direction of the ends of the liquid-crystal molecules at a given z coordinate as described by Eqs. (4) is cycloidal in nature.

An example of such a comparison between theory and experiment for all five measurable diffracted orders is shown in Fig. 3. In this experiment, the VGM liquid crystal utilized was a phenyl benzoate mixture (Hughes 2N40[1]), the cell thickness was 6.0 μm, and the applied voltage was 32 V dc. To obtain the data shown in this figure, α_{max} and η_{max} have been utilized as the only adjustable parameters. The assumed simple harmonic spatial dependence of the angles α and η is justified by the quality of fit evident in Fig. 3. In addition, such a dependence of the angles α and η on the spatial coordinate x has been predicted by direct minimization of the free energy in a similar nematic liquid-crystal system.[6]

The fitting procedure described above permits the extraction of the maximum orientational excursion angles $\alpha_{max}(V)$ and $\eta_{max}(V)$ as functions of the applied bias voltage above the threshold for grating formation, as shown in Fig. 4 [$\eta_{max}(V)$] and in Fig. 5 [$\alpha_{max}(V)$]. In each case, it is observed that the maximum excursion angles both in and out of the plane of the grating seem to increase as the logarithm of the applied voltage. It should be noted that the assumption of pinned boundary conditions with a resultant sinusoidal z dependence[6] results in an equivalently good fit, with altered values of the peak orientational angles in the layer center. Within experimental error, the fitting technique described above cannot be used to discriminate among possible z-axis variations.

Careful experimental measurements of the off-diagonal elements in the Jones matrix describing polarized light propagation through the VGM cell have revealed a statistically significant asymmetry not predicted by the uniaxial model. The origin of this "B/C" asymmetry effect is under continuing investigation.

In conclusion, we have utilized the polarization properties of the diffracted orders from a liquid-crystal birefringent phase grating to describe the principal features of the spatial distribution of the molecular orientation within the liquid-crystal layer.

The authors gratefully acknowledge the contributions of D. Boswell, A. M. Lackner, and J. D. Margerum (Hughes Research Laboratories) and of A. A. Sawchuk (University of Southern California). This research was supported by the U.S. Air Force Office of Scientific Research, Electronics and Solid State Sciences Division, under grants AFOSR-81-0082 and AFOSR-83-0185 at the University of Southern California and contract F49620-81-C-0086 at Hughes Research Laboratories.

* Permanent address, Institut d'Optique, Université de Paris-Sud, BP 43, 91406 Orsay Cedex, France.

† Present address, IBM Corporation, 5600 Cottle Road, San Jose, California 95193.

References

1. B. H. Soffer, J. D. Margerum, A. M. Lackner, D. Boswell, A. R. Tanguay, Jr., T. C. Strand, A. A. Sawchuk, and P. Chavel, Mol. Cryst. Liq. Cryst. 70, 145–161 (1981).
2. P. Chavel, A. A. Sawchuk, T. C. Strand, A. R. Tanguay, Jr., and B. H. Soffer, Opt. Lett. 5, 398–400 (1980).
3. B. H. Soffer, D. Boswell, A. M. Lackner, P. Chavel, A. A. Sawchuk, T. C. Strand, and A. R. Tanguay, Jr., Proc. Soc. Photo-Opt. Instrum. Eng. 232, 128–136 (1980).
4. B. H. Soffer, D. Boswell, A. M. Lackner, A. R. Tanguay, Jr., T. C. Strand, and A. A. Sawchuk, Proc. Soc. Photo-Opt. Instrum. Eng. 218, 81–87 (1980).
5. R. C. Jones, J. Opt. Soc. Am. 31, 488–503 (1941).
6. Yu. P. Bobylev and S. A. Pikin, Sov. Phys. JETP 45, 195–198 (1977).

Materials requirements for optical processing and computing devices

Armand R. Tanguay, Jr.
University of Southern California
Departments of Electrical Engineering
and Materials Science
Optical Materials and Devices Laboratory
University Park, MC-0483
Los Angeles, California 90089-0483

Abstract. Devices for optical processing and computing systems are discussed, with emphasis on the materials requirements imposed by functional constraints. Generalized optical processing and computing systems are described in order to identify principal categories of requisite components for complete system implementation. Three principal device categories are selected for analysis in some detail: spatial light modulators, volume holographic optical elements, and bistable optical devices. The implications for optical processing and computing systems of the materials requirements identified for these device categories are described, and directions for future research are proposed.

Subject terms: optical computing; optical information processing; spatial light modulators; electrooptic materials; volume holographic optical elements; bistable optical devices.

Optical Engineering 24(1), 002-018 (January/February 1985).

CONTENTS

1. INTRODUCTION: WHY MATERIALS?

The invention and subsequent development of the laser as a powerful source of both spatially and temporally coherent illumination has spawned a large number of complementary optical technologies. Not the least of these are the research areas of coherent optical processing and computing. These areas have been characterized by the rapid invention and development of architectures and algorithms that seek to take advantage of the highly multiplexed interconnectability provided by phase- or amplitude-modulated coherent wavefronts. In recent years, significant efforts have been made to extend these concepts to incoherently illuminated systems, in order to avoid some of the noise and stability characteristics of coherent images. The result of both parallel efforts has been the development of rather sophisticated software for a wide range of applications, including pattern recognition, earth resources data acquisition and analysis, texture discrimination, synthetic aperture radar (SAR) image formation, radar ambiguity function generation, spread spectrum identification and analysis, analog and digital computation, systolic array processing, phased array beam steering, and artificial (robotic) vision. These and other applications are treated in more detail in many of the accompanying papers in this special issue of *Optical Engineering*,[1] as well as in a recent special issue of *IEEE Proceedings* on optical computing.[2]

For the most part, these architectures and algorithms have been tested through the use of film-based data input fields, Fourier plane filters, and detector masks. This is a direct consequence of the continued lack of wide availability of appropriate two-dimensional real-time devices to perform these functions. In a sense this is quite surprising, since in fact the rapid development of architectures and algorithms has been accompanied by an equally rapid series of inventions of associated candidate device implementations. In the case of one-dimensional input transducer systems, for example, the acoustooptic modulator has proven to be capable, reliable, and relatively inexpensive.[3-6] On the other hand, the past 10 years have witnessed a far slower development of practical two-dimensional transducers characterized by excellent performance, operational reliability, ease of fabrication, and reasonable cost.

This seeming enigma of rapid invention followed by much less rapid development of candidate device technologies can be traced primarily to limitations in materials and materials technology. Each of the candidate technologies has placed rather severe demands on the state of the art of the requisite material or materials (in most cases, unfortunately, plural). One might observe that in general, regardless of the nature of the active optical effect utilized (e.g. electrooptic, magnetooptic, photochromic, photodichroic, photorefractive, elastomeric, electroabsorptive, etc.), the magnitude of the observable optical perturbation per unit excitation is not large enough with readily available materials to allow flexible device engineering.

Invited Paper OP-101 received Oct. 13, 1984; revised manuscript received Nov. 7, 1984; accepted for publication Nov. 8, 1984; received by Managing Editor Nov. 13, 1984. This paper is a revision of Paper 456-04 which was presented at the SPIE conference on Optical Computing: Critical Review of Technology, Jan. 24-25, 1984, Los Angeles, CA. The paper presented there appears (unrefereed) in SPIE Proceedings Vol. 456.

This is not in most cases due so much to fundamental physical limitations as it is to the paucity of readily available materials. Either few materials with the requisite interaction mechanism have been discovered or synthesized, or the known materials have not been developed into producible optical grades, or the multiple necessary characteristics of such materials have not been carefully characterized, understood, and optimized. In other words, "optical" silicon has yet to be invented, or at the very least labeled as such and fully developed.

Due to the importance of materials and associated materials technologies to the successful development of active optical devices for optical processing and computing functions, and to the importance of such devices for the implementation of powerful real-time (or much faster than real-time) systems, the decision as to whether optical processing and computing will come of age may ultimately rest on the capabilities and fortunate discoveries of materials scientists and process engineers.

In Sec. 2, the elements of a generalized optical processor or computer are described from the component or device perspective, with the purpose of identifying component types that satisfy a broad range of requisite performance functions. In succeeding sections, three principal categories of such components are discussed in more detail, including spatial light modulators (Sec. 3), volume holographic optical elements (Sec. 4), and bistable optical devices (Sec. 5). The perspective of these more detailed discussions is such as to provide a backdrop of fundamental principles of device operation against which the critical materials requirements for device optimization can be viewed. Two concluding sections contain remarks on the implications of the current status of optical material and device research for optical processing and computing (Sec. 6) and on a number of suggested directions for intensified research and development (Sec. 7).

2. ELEMENTS OF AN OPTICAL COMPUTER: THE DEVICE PERSPECTIVE

In order to provide an appropriate context for the discussion of material and device issues, it is useful to consider in some detail the spectrum of components that will likely comprise generalized optical processors and computers. Such fundamental elements of optical processing systems are diagrammed schematically in Fig. 1. This diagram has not been constructed with the intent of completeness in the sense that it represents all possible architectures or allows straightforward implementation of all possible algorithms. Rather, it is intended to convey the nature of specific functions required for a wide range of applications, as well as examples of specific device implementations of these functions. It is further understood that in some cases the distinctions between two or more processing functions may blur to the extent that they can be performed within a single component.

The principal constituent elements of a generalized optical processor or computer are a central processing unit (CPU) that performs the essential implementable function, a data management processor that orchestrates the flow of data and sequence of operations (usually considered part of the CPU in a traditional electronic computer), several types of memory elements for both short-term and long-term data storage and buffering, format devices to spatially organize input data fields, input devices to convert data input types to a form amenable to subsequent processing, output devices to convert processed results to detectable and interpretable forms, and detectors to produce externally addressable results. In Fig. 1, feedback interconnects are explicitly shown as separate components in anticipation of iterative architectures and algorithms. The solid-line boxes in Fig. 1 depict constituent components discussed in detail herein, whereas dashed-line boxes depict components beyond the scope of the current discussion. It is further assumed at the outset that some fraction of the components shown may best be implemented in electronic circuitry, such that the given diagram is applicable to hybrid optical/electronic processors as well.

The central processing unit is the functional heart of the optical processor/computer and contains the hardware implementation of the processor function. This implementation hardware may be as

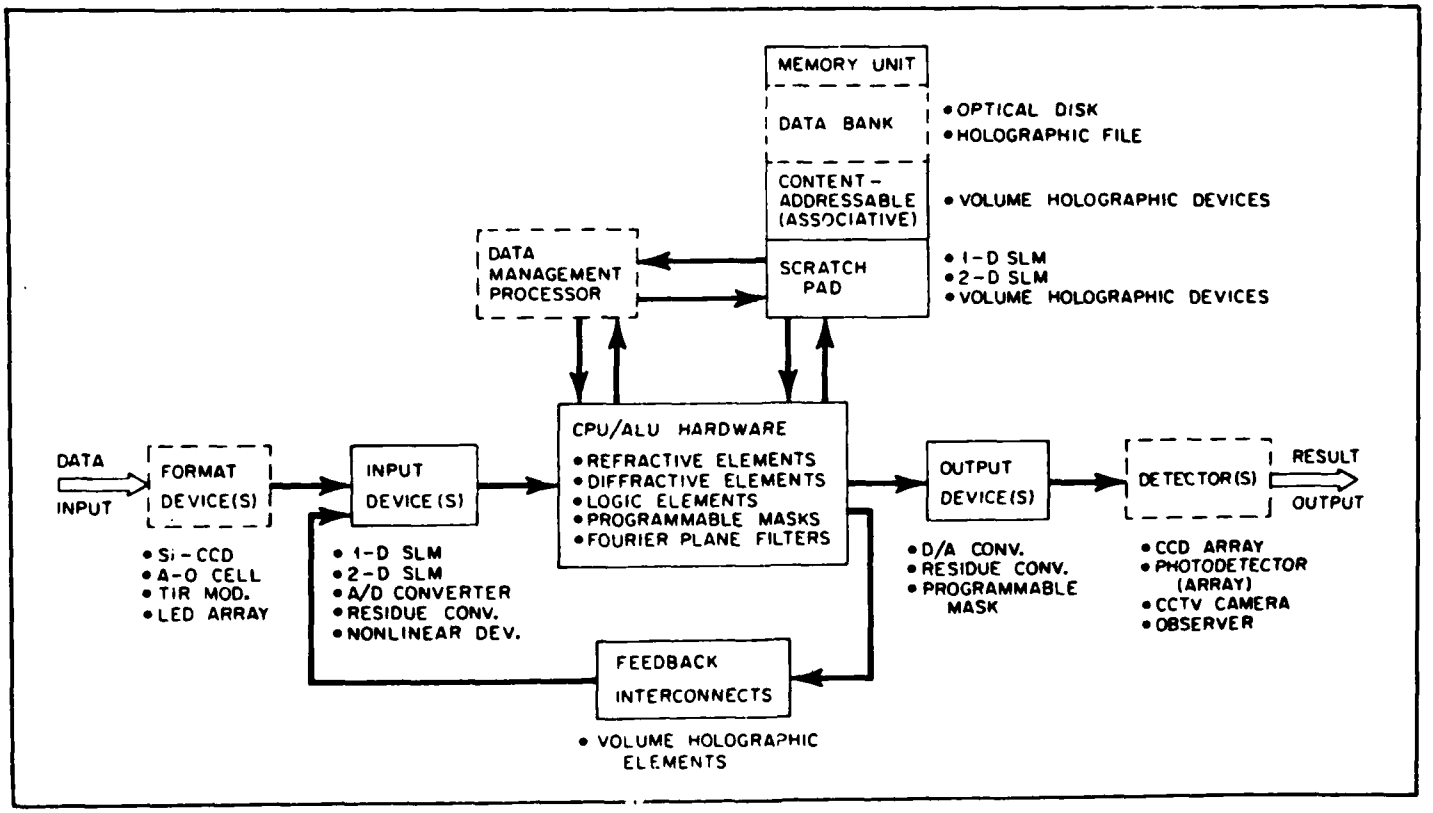


Fig. 1. Schematic diagram of the elements of a generalized optical processor or computer.

simple as a refractive element (lens) that performs a two-dimensional Fourier transformation, and may also include diffractive elements (such as ruled gratings), logic elements (e.g., two-dimensional bistable arrays or programmable Boolean logic arrays), programmable image plane masks, and programmable Fourier plane filters.

A memory unit is required for all but the least sophisticated optical processors, and can be segregated into a number of different functional types. A scratch pad memory provides intermediate result storage, buffering between the CPU and data management processor, and upward buffering into longer-term memories. This memory unit might consist of one- or two-dimensional spatial light modulators or of volume holographic optical elements. A data bank or archival storage unit may be employed to store algorithms, sequences of accessible look-up tables for algorithm implementation, and previous output results. This unit will likely consist of optical disk drives or massive holographically encoded files. It is also very likely amenable to replacement by semiconductor or magnetic tape memories, in cases for which the requisite access times and total storage capacity are not severely constrained. Such a memory replacement would entail a further component requirement for an intermediate parallel/optical-to-serial/electronic converter (such as a CCD array). A third type of memory unit is the content-addressable, or associative, memory. In this type of information storage, the output data plane from the CPU generates a direct memory access without resort to two-dimensional or multidimensional random access encoding. The associated memory content then enters the CPU for further processing or may give rise directly to the system output. An example of this type of function is an image plane correlator, which gives rise to a spatially encoded correlation intensity peak when addressed by an appropriate recognizable input pattern. Current candidates for this type of memory function include a number of volume holographic optical elements.[7,8]

Given the sense in which we have defined the CPU hardware herein, a separate data management processor is required for timing, sequencing, and organization of memory access. Given the speed and sophistication of already available electronic subsystems, it is most likely that this element will not be implemented with purely optical components, although it may in fact contain some optical-to-electronic or electronic-to-optical converters.

Since the input data stream to an optical processor can be envisioned to comprise a wide array of formats (one-, two-, or three-dimensional in various mixtures of temporal and spatial coordinates), devices will be required for format conversion functions. These functions may include, for example, serial/electronic-to-parallel/optical conversion (as required for synthetic aperture radar image formation), serial/optical-to-serial/electronic conversion, serial/optical-to-parallel/optical conversion, etc. Devices that implement these types of format functions include one- and two-dimensional Si-CCD detector arrays, acoustooptic modulators, linear array total internal reflection spatial light modulators,[9] and one- and two-dimensional light emitting diode arrays. Such format devices may also be configured as hybrid optical/electronic components to perform pixel-by-pixel preprocessing functions such as level restoration, thresholding, rescaling, amplification, and signal-to-noise reduction through integration in both time and space.

It is further convenient to identify a separate (but not always distinct) device category that performs the actual input function to the CPU. This category includes such devices as one- and two-dimensional spatial light modulators (that may or may not be required to perform an incoherent-to-coherent conversion function), analog-to-digital converters (or digital-to-analog converters), residue converters, and devices that implement point nonlinearities.

At this point it is worthwhile to carefully draw attention to the distinctions among format, input, and CPU devices. It is assumed that the processor will in general fully utilize the advantages of optics if nonlocal (as opposed to point) operations are introduced at some stage in the optical train. These functions are identified with the CPU. Format devices organize input information in space and/or time for appropriate subsequent processing, and input devices perform certain preprocessing but pointwise transformations on the formatted input data field. A specific example that brings out this architectural identification is that of synthetic aperture radar image formation by coherent optical reconstruction techniques.[10-12] Consider the case of the familiar film-based optical processor.[13] Each range-distributed return at a particular azimuth coordinate must be converted into an appropriately scaled optical line scan by a format device such as a laser scan converter. Each scan line is written onto an input device (photographic film) that performs a pointwise incoherent-to-coherent optical conversion. The CPU consists of (typically) three sequential lens elements that form an anamorphic imaging system. These lenses transform spatially distributed anisotropic Fresnel zone plates (the SAR return from a point target on the ground) into a point target with amplitude proportional to the reflectivity of the object at the illuminating radar frequency.

Output devices perform a subsequent pointwise transformation on the data field emergent from the CPU. Such devices might include digital-to-analog converters (or analog-to-digital converters), residue converters, or programmable detector masks.

Numerous types of detectors that in essence perform a "deformatting" or format transformation on the output data field can be envisioned. The dimensionality of the output detector is not necessarily given by the dimensionality of any previous element. For example, a two-dimensional processing function that detects the presence or absence of a certain texture class in an input image may be configured in such a way that the resultant detection represents the two-dimensional spatial integral of the texture amplitude as a temporally variant measure of the image texture content. Hence, detectors of importance to optical processors and computers span the range from point detectors through one-dimensional linear arrays to fully two-dimensional "cameras." Examples of such devices include CCD arrays, photodetectors and photodetector arrays, CCTV cameras, and human observers.

A crucial element in the implementation of all but the simplest processing algorithms is the incorporation of parallel interconnects in a feedback path.[14-16] The representative path drawn in Fig. 1 is but one of many possible configurational schemes for the incorporation of feedback to allow iterative and successive approximation solutions to multidimensional problems. Although the most straightforward feedback path might be from the detector output to the format device input, the cost in processing time and efficiency may obviate this approach in all cases except those in which a change in format at each iteration is essential to the solution algorithm. It is more likely that feedback will be incorporated from the output of the CPU to the input device(s), as shown in the diagram, with an intermediate content-addressable memory access. Volume holographic optical elements are prime candidates for such a feedback interconnect device.

Holographic interconnect devices also provide the operational flexibility of spatial reformatting between successive iterations. As pointed out by Sawchuk,[14,15] this type of configurational flexibility may allow the input/output, CPU, and memory functions to be located in a single device. Parallel access to individual I/O, processing, and storage functions is allowed sequentially with each iteration and spatially by means of a holographic optical interconnect in a feedback path.

The generalized configuration of the optical processor diagrammed in Fig. 1 allows for a very wide range of potential processor configurations, and in particular for various degrees of electronic/optical hybridization. Following Lohmann,[17] it is useful to differentiate among the various categories of signal processing architectures on the basis of the relative advantage to be gained compared with currently envisioned all-electronic processors. This comparison is shown in Fig. 2. The left-hand column lists potential input data formats, while the top row lists types of processing hardware. The anticipated advantage of combining a given input format with a given processing hardware is illustrated by the number of plus signs at the appropriate intersection. The particular assignment shown is certainly controversial, and is also highly dependent on the specific nature of the processing task identified for comparison. Rather than

INPUT SIGNAL	PROCESSING HARDWARE		
	ELECTRONIC (1-D)	OPTICAL (1-D)	OPTICAL (2-D)
ELECTRONIC (1-D)		+	+ +
OPTICAL (2-D)	+ +	+ + +	+ + + + +

Fig. 2. Schematic diagram of the categories of signal processing ranked according to the computational advantage of associated input signal/processing hardware pairings.

imply that this represents an average over all possible processing functions (which it certainly does not), we merely wish to note that numerous processing functions do exist for which the hierarchy shown is relatively defensible. In particular, the bottom row assignments can be described as follows. The complexity of problems associated with intrinsically two-dimensional optical data fields is well known, as are the benefits to be derived from high speed analysis of such data. Electronic processing algorithms for two-dimensional optical inputs already provide a certain richness of potential impact. Adding the capability of one-dimensional optical processing enhances the implementation of all operations that require extensive nonlocal interrelationships and that are in addition reducible to sequential one-dimensional operations. The full power of a parallel optical processor will be realized when two-dimensional optical hardware acts on intrinsically two-dimensional optical data (as opposed to artificially formatted, essentially unrelated data fields). A reduction in throughput with perhaps a concomitant gain in programming flexibility can be realized by processing one-dimensional (in time) electronic signals with the assistance of optical interconnects,[14-16,18] as represented by the upper right corner of Fig. 2.

The range of functional hardware needed for the development of currently envisioned optical processors was identified in Fig. 1. From the diagram it is possible to extract three principal device categories that impact a number of critical device functions. Spatial light modulators can be utilized as format, input, CPU, memory, and output devices. Various types of input, CPU, memory, and interconnect functions may be implemented by means of volume holographic optical elements. Finally, bistable optical devices may be used in certain input, CPU, output, and memory roles. These three principal device categories are described in more detail in the following sections.

3. SPATIAL LIGHT MODULATORS

The primary functions of an optically addressed spatial light modulator are to detect an incident intensity distribution that represents the data field to be processed, to utilize the detected intensity to modify a given material property, and finally to encode the amplitude and/or phase of the readout illumination with material state-dependent information. Examples of these functions are incoherent-to-coherent conversion, wavelength conversion, serial-to-parallel conversion (usually in conjunction with an appropriate formatting device), and intensity-to-spatial frequency conversion.[19,20]

In certain optical processing applications it is desirable to combine the format and input device functions in a single device to allow for serial-to-parallel conversion of electronic time-variant signals. In addition, such electrically addressable spatial light modulators can be used in conjunction with an input parallel-to-serial converter such as a CCD array to allow effective optical (2-D)-to-optical (2-D) operations to be implemented. Electrically addressable spatial light modulators are discussed in more detail below.

The potential applications of spatial light modulators in optical computers are numerous and varied. These applications include image amplification, time/space transformation (such as serial-to-parallel conversion or parallel-to-serial conversion), input transduction (with memory in some cases), scratch pad memory (for single two-dimensional data fields), programmable Fourier plane filtering and correlation, programmable detector masking, and page composition for holographic memories (in cases for which spatial or temporal reformatting is desirable).

TABLE I. Available Spatial Light Modulators

- Electrooptic spatial light modulators (PROM, PRIZ, MSLM, e-beam DKDP, photo-DKDP, TIR)
- Liquid crystal light valve (CdS, ZnS, Si, $Bi_{12}SiO_{20}$, CCD-addressed; hybrid twisted nematic, variable grating mode, multiple period)
- Deformable mirror devices (oil film, e-beam DMD, CCD-DMD, photo-DMD)
- Magnetooptic spatial light modulators (integrated magnetooptic arrays, Curie point)
- Photochromic spatial light modulators
- Electroabsorptive spatial light modulators (GaAs CCD)
- Photorefractive spatial light modulators (photorefractive incoherent-to-coherent optical converter, volume holographic optical elements)
- Photodichroic spatial light modulators
- Acoustooptic spatial light modulators

In the past few years, the list of spatial light modulators under development or in production has expanded considerably. The principal types are outlined in Table I, with specific examples and variations as noted. These spatial light modulators may additionally be characterized by the principal modulation effect utilized: reflective, refractive, diffractive, or absorptive.

Given the extensive variety of modulation categories (and of device types within each category) that are subjects of intensive current investigation, it is beyond the scope of this review to cover the specific attributes of each in depth. Rather, focus will be placed on a particular modulation effect and on device examples drawn from this category, with the purpose of identifying the key materials issues and their impact on expected device performance. Extension of these concepts to other modulation effects and associated device types is far from straightforward, but the essential conclusions drawn at the end of this review are independent of the examples chosen.

Consider the origin of refractive modulation effects in single crystal materials, as shown schematically in Fig. 3. The left-hand portion of the figure shows a schematic plot of the potential of a given ion or valence electron in a single crystal material as a function of the configuration coordinate, both in equilibrium in the lattice (solid curve) and following the application of a perturbation (dashed curve). The key features to recognize are that in general the influence of an externally applied perturbation will be such that the potential minimum will be displaced, the binding energy (depth of the potential well at the minimum) will be altered, and the shape of the potential well will be changed. These changes collectively alter the energy levels of the bound states of the ions and valence electrons in the system. These energy levels are responsible for the dispersion (dependence on wavelength) of both the refractive index and the absorption coefficient, as shown in the right-hand portion of the figure. Here the solid curves show the index (n) and absorption coefficient (α) as a function of the wavelength of probe irradiation (λ) for the ion or electron in equilibrium, and the dashed curves show the effects of the applied perturbation. The region of nominally pure refractive modulation occurs where the magnitude of the absorption is negligible. The presence of absorptive modulation is thus seen to be directly linked to that of refractive modulation, shifted only with respect to the wavelength region of interest, and in fact is *not* separable from the effects of refractive modulation in any case.

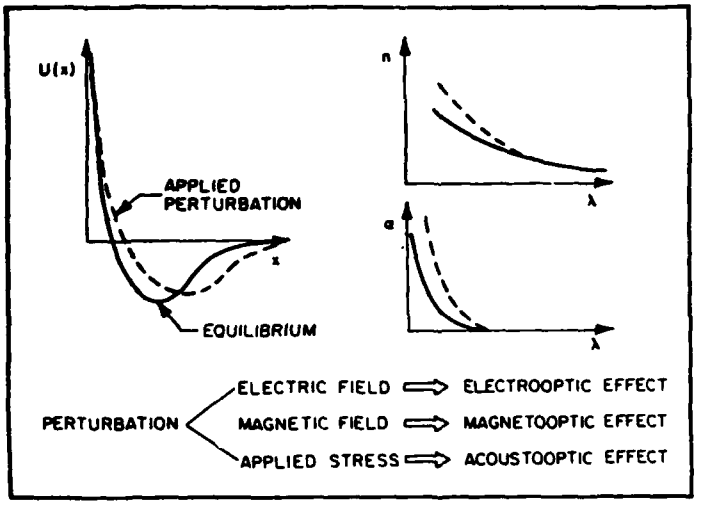


Fig. 3. Physical description of the origin of refractive (and birefringent) modulation. The potential energy of an ion or valence electron is shown as a function of the configuration coordinate, both for equilibrium and under the influence of an applied perturbation. Induced changes in the index of refraction and absorption coefficient as a function of wavelength are shown for both cases.

The effect of an applied perturbation on the ions gives rise to a distinct contribution to the refractive modulation, as do the perturbation effects on the electrons. In addition, lattice distortions alter the electron energy states by distorting the crystal field, giving rise to a joint ionic-electronic contribution.[21] The relative importance of these effects depends somewhat on the wavelength and also on the frequency of the applied perturbation. For example, if the applied perturbation oscillates at a sufficiently high frequency such that the lattice is incapable of responding with a concomitant distortion, the electronic contribution alone will be witnessed.

The principal types of refractive modulation effects are categorized by the nature of the applied perturbation. Applied electric fields give rise to the electrooptic effect, applied magnetic fields give rise to the magnetooptic effect, and high frequency applied stresses give rise to the acoustooptic effect. Certain material properties can be utilized to couple these effects so that, for example, an applied electric field can give rise to a strain through the converse piezoelectric effect, which in turn gives rise to a refractive modulation through the elastooptic effect. The selection of appropriate materials that exhibit large figures of merit for these effects thus depends on the identification of ionic and/or electronic distributions that are (1) highly polarizable (large induced distortion per unit applied perturbation) and (2) distributed with a lattice symmetry that allows for substantial net polarizabilities.[22]

The application to a refractive material of a uniaxial perturbation such as an applied electric field will in general lower the symmetry of the material. A natural consequence of this is that the perturbed refractive index will be dependent on the polarization of the probing illumination. For a given direction of light propagation in the material, two orthogonal eigenstates will exist, characterized by distinct refractive indices. Input polarization states that are superpositions of the two eigenstates will decompose into the appropriate components, which will accumulate a phase difference on propagation through the perturbed material. The emergent polarization will be a function of both the input polarization and the magnitude of the applied perturbation, giving rise to the phenomenon of induced birefringence (double refraction).

As a specific example of a spatial light modulator that utilizes refractive modulation, consider the class of electrooptic spatial light modulators (ESLMs).[23,24] This class includes the Pockels Readout Optical Modulator (PROM),[25,26] the PRIZ[27,28] (Soviet acronym for a crystallographically modified PROM), the microchannel spatial light modulator (MSLM),[29,30] the TITUS [electron-beam addressed potassium dideuterium phosphate (DKDP) device],[31,32] and the photo-TITUS[33,34] (optically addressed DKDP device).

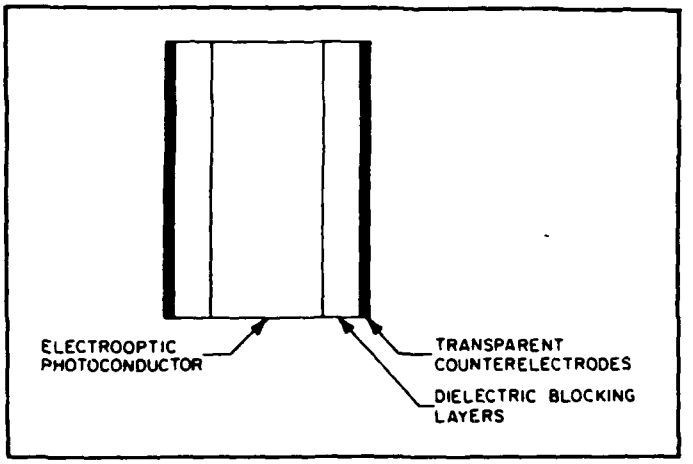


Fig. 4. Schematic diagram of an electrooptic spatial light modulator, as described in the text.

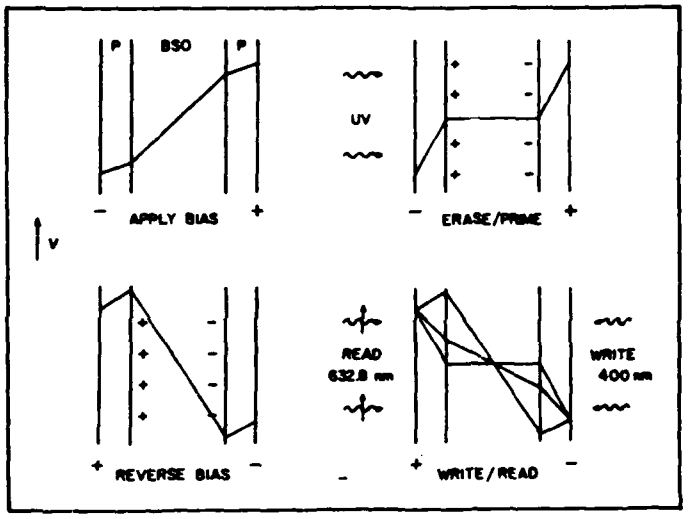


Fig. 5. Longitudinal potential distribution for the Pockels Readout Optical Modulator during an operational cycle. Three different write exposures are shown for the write/read function.

A schematic diagram of a generalized ESLM is shown in Fig. 4. A single crystal layer of a photoconductive electrooptic material is sandwiched between dielectric blocking layers overcoated with transparent counterelectrodes. In some ESLMs, the photoconductive and electrooptic modulation functions are accomplished in separate layers. The configuration shown in Fig. 4 is most representative of the PROM and certain of the PRIZ structures (with dielectric blocking layers).[23,24] In the following discussion, the PROM is used as a specific example to illustrate the basic principles of ESLM device operation.

Consider the schematic representation of the cross section of an ESLM, as shown in Fig. 5. A voltage is applied across the device and is divided among the three dielectric layers in inverse proportion to the capacitance of each layer. The device is then "primed" (and any previous stored image erased) by uniform exposure to illumination at a photosensitive wavelength. Electron-hole or electron-trap pairs are created by photoexcitation, and the mobile electrons drift under the influence of the applied field so as to effectively cancel the internal field in the electrooptic material and drop the entire applied potential difference across the dielectric blocking layers. The sense of the applied electric field is then reversed, approximately doubling the

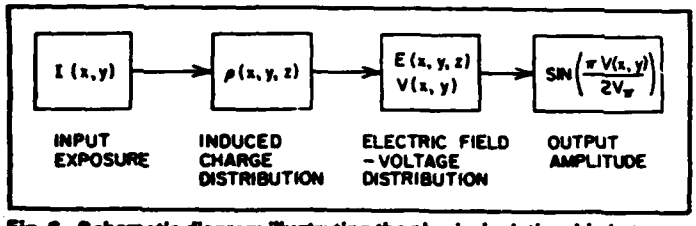


Fig. 6. Schematic diagram illustrating the physical relationship between the output amplitude and the input intensity as mediated by charge transport and the linear electrooptic effect for the Pockels Readout Optical Modulator.

applied potential across the electrooptic layer (which enhances the response sensitivity and saturation modulation or saturation diffraction efficiency). Imagewise-modulated illumination (the input data field) then causes space-variant carrier generation, which induces a corresponding variation in the internal space-charge field. Polarized readout light at a wavelength of photoinsensitivity is then encoded by the resultant refractive index modulation. If the index modulation alone is sensed (by appropriate choice of crystallographic orientation or input polarization), the device acts as a pure phase modulator. If the index modulation is anisotropic in the plane of the electrooptic crystal, proper choice of the readout polarization will induce a space-variant birefringence (differential refractive modulation for each of two orthogonal polarization components) that can be converted to pure amplitude modulation by means of a suitably oriented polarizing filter (analyzer).

The process of incoherent-to-coherent conversion in an ESLM can be summarized by the proportionality diagram given in Fig. 6. As shown in the figure, an input intensity distribution $I(x,y)$ gives rise to a net charge separation $\sigma(x,y,z)$, which induces a change in the space-charge field within the electrooptic crystal $E(x,y,z)$. The longitudinal integral of the space-charge field corresponds to a space-variant potential difference across the layer $V(x,y)$, which results in an output amplitude through the (crossed) analyzer that is proportional to the input intensity $I(x,y)$ (for appropriate values of the ratio of $V(x,y)$ to the electrooptic half-wave voltage $V_{\lambda/2}$).

More complicated structures with additional performance features can be constructed, as shown schematically in Fig. 7. In this diagram, the functions of the photoconductor and electrooptic modulator are spatially separated by means of an absorptive layer and a multilayer dielectric mirror. This configuration allows the photoconductor to be addressed without altering the properties of the electrooptic modulation layer, and the readout function to be performed in reflection (which doubles the refractive or birefringent modulation per unit applied perturbation). The significant advantage is the capability for optimizing each critical function independently. The significant disadvantage is the fact that severe constraints are placed on the choice of the photoconductive layer due to the large voltages required for efficient modulation with currently available electrooptic materials. The photoconductive layer must be capable of supporting the full applied voltage in regions that are unexposed. In addition, this structure adds the complication that high contrast ratios are obtainable only if a suitable combination of layer dielectric constants, dark resistivities, and photoconductive sensitivity can be found such that negligible voltage is applied across the electrooptic crystal in spatial locations corresponding to unexposed regions of the photoconductive layer.

With these remarks in mind, the materials requirements for ESLMs are summarized in Table II. The principal point here is that even conceptually simple devices such as those represented by the class of electrooptic spatial light modulators can produce a significant list of desired material types. At this point, the scope of the discussion will be reduced to examining in detail the materials issues involved in the two important categories: electrooptic crystals and dielectric blocking layers.

Electrooptic single crystal layer requirements for application in

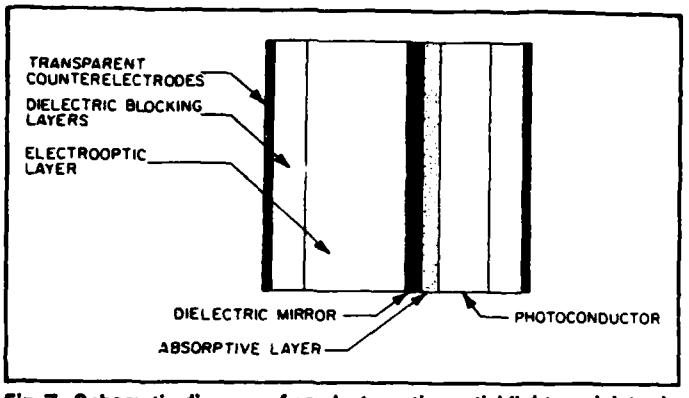


Fig. 7. Schematic diagram of an electrooptic spatial light modulator in which the photoconductive voltage division and electrooptic modulation functions are physically separated.

TABLE II. Electrooptic Spatial Light Modulators: Materials Requirements

- Bulk electrooptic single crystals
- Thin film electrooptic layers
- Dielectric blocking layers
- Transparent conductive coatings
- Antireflection coatings
- Dichroic mirrors
- Thin film photoconductive layers
- Absorptive (light-blocking) layers

TABLE III. Electrooptic Crystal Requirements for Electrooptic Spatial Light Modulators

- Cubic space group symmetry
- Large linear electrooptic effect (longitudinal or transverse)
- High dark resistivity
- High mobility-lifetime product
- Quantum efficient photoconductivity
- Optically transparent in readout window
- Low residual strain birefringence
- Uniformity of resistivity/absorption coefficient
- Not optically active (if possible)

ESLMs are summarized in Table III. Cubic space group symmetry is desirable for the enhancement of readout contrast ratio. Materials with large uniaxial anisotropy induce undesirable birefringent polarization effects for all readout propagation directions not parallel to the optic axis. A large linear electrooptic (Pockels effect) coefficient reduces the magnitude of the requisite applied voltage and relaxes the requirements on the thickness and dielectric breakdown strength of the dielectric blocking layers. High dark resistivity in combination with a large dielectric permittivity gives rise to long storage times, for applications requiring lengthy integrations of the input data or multiple interrogations. The mobility-lifetime product for the contributing charge carrier (usually either electrons or holes) provides a measure of the device sensitivity (change in voltage difference per unit optically excited charge carrier). The photoconductive quantum efficiency also affects the device sensitivity by determining the fraction of absorbed photons that generate photoexcited charge carriers. The spectral dispersion of the absorption is particularly important in providing a window of transparency for (relatively) nondestructive

readout. This influences the available readout gain as well as the number of multiple interrogations at a given signal-to-noise ratio that can be accommodated without significant image degradation. Optimally, the crystal should be able to be grown by a technique that is capable of producing optical quality material with a minimum of residual strain birefringence, rotation-induced striae, resistivity inhomogeneities (due to impurity and/or defect incorporation), and fluctuations in the absorption coefficient at both the writing and readout wavelengths. Finally, the utilization of crystallographic point groups that exclude natural optical activity (optical rotatory power) is preferred to eliminate the complication of elliptically polarized eigenstates and concomitant signal-dependent phase distortions.[35,36]

In the case of electrooptic materials with simultaneous photoconductivity, an additional and rather pressing requirement is for the complete characterization of the photoexcitation process, electronic transport characteristics, and defect/impurity content of the as-grown and as-doped material. This is of particular importance not only for application to electrooptic spatial light modulators, but also to volume holographic optical elements, as discussed in Sec. 4. Such characterization directly impacts the capability of producing single crystal materials with reproducible characteristics, as well as the potential modification of material properties for specific device needs by intentional incorporation and/or control of intrinsic defects and extrinsic incorporants.

After a period of explosive growth and creative intensity during the 1960s, the growth and characterization of electrooptic materials recently has emerged from a period of nearly complete dormancy during the 1970s. Several developments portend well for the advancement of the state of the art of electrooptic materials during the 1980s. Increased efforts to grow optical quality electrooptic materials such as bismuth silicon oxide ($Bi_{12}SiO_{20}$; BSO), barium titanate ($BaTiO_3$), strontium barium niobate ($Sr_xBa_{1-x}Nb_2O_6$; SBN), lithium niobate ($LiNbO_3$), and potassium tantalate niobate ($KTa_xNb_{1-x}O_3$; KTN) are under way at a number of research laboratories. Associated with these growth efforts are extensive characterization programs specifically formulated for optical device applications.[37] To a certain extent these efforts are amplified by a concurrent interest in essentially the same set of materials for photorefractive device applications, including phase conjugation and nonlinear multiwave mixing. In addition, interest has been spurred by the development of ultrahigh speed electrooptic modulators for integrated optics applications and by the necessity to understand and minimize optical damage in these materials.[38-41] Recently, predictions[42,43] and observations[44,45] of enhanced nonlinear optical coefficients in superlattice structures fabricated from ultrathin periodic multilayers of III-V semiconductors such as $GaAs/Al_{1-x}Ga_xAs$ have pointed toward the potential of engineering materials with specifically enhanced electrooptic properties. Finally, significant effort has been initiated toward the synthesis of organic polymers that exhibit remarkably large linear and nonlinear optical effects.[46-48] The availability of such materials in single crystal form with suitable optical quality may mark a revolution in the availability of inexpensive, high performance electrooptic light modulators, with the advanced development of ESLMs not far behind.

For ESLM applications, there are several materials requirements for the dielectric blocking layers, which provide the necessary voltage division to generate space-variant potential distributions across the electrooptic material. These requirements include high resistivity ($\rho > 10^{18}$ Ω·cm) to prevent charge transport across the layers even under extremely high electric field conditions ($E > 10^7$ V/cm), high dielectric permittivity ($\epsilon = 20$ to 500 ϵ_0) for proper impedance matching to electrooptic crystals with relatively high dielectric permittivities (which influences the resultant device resolution[23,24]), high dielectric breakdown strength ($\cong 10^7$ V/cm) to withstand the high fields applied across the electrooptic materials, optical transparency in the write and read spectral windows, and the capability of deposition with negligible interfacial strain on the electrooptic crystal layer. Other necessary properties include very low residual scattering from refractive index inhomogeneity (which contributes to reduced signal-to-noise ratio in the Fourier plane of the processor), the capability for deposition with excellent thickness uniformity over substrates of order 2 to 5 cm in size, and the capability for deposition at low substrate temperatures to avoid the buildup of residual birefringent strain. Finally, the dielectric blocking layers must exhibit very low concentrations of mobile ions that can migrate under high applied fields to produce significant history-dependent bias voltages (as is the case, for example, with SiO_2[49]). These requirements are summarized in Table IV.

TABLE IV. Dielectric Blocking Layer Requirements

- High resistivity ($\rho > 10^{18}$ Ω·cm)
- High dielectric constant (~20–500 ϵ_0)
- High dielectric breakdown strength (~10^7 V/cm)
- Optically transparent in write/read windows
- Low interfacial strain
- Low temperature deposition method available
- Low postdeposition mobile ion content

Progress in the development of appropriate materials has been less than spectacular during the past 10 years. The standard material employed in these devices is parylene (poly-dichloro-para-xylylene[50-52]), which is a vapor-deposited polymer with conformal coating properties, good optical transparency in the visible, and adequate dielectric breakdown strength. The volume resistivity of parylene can exceed the requirement, but these values are difficult to achieve in practice.[53] The dielectric constant is relatively low ($\epsilon = 3$), which significantly reduces the available device resolution. In addition, the material is sensitive to humidity and ultraviolet radiation, which can cause dielectric breakdown substantially below the specified threshold.[54] The potential for improved materials amenable to ESLM incorporation rests on the continued development of vapor-deposited organic polymers and on the advancement of the state of the art of inorganic thin films deposited by sputtering, ion-beam deposition, and chemical vapor deposition methods.

It should be noted here that successes in the programs designed to produce materials with enhanced electrooptic coefficients will have a direct impact on the availability of appropriate dielectric blocking layers. This interrelationship between the two materials development programs arises naturally from the anticipated significant reduction of constraints on the blocking layers that results from reductions in the device operating voltage.

The previous two examples of materials requirements for spatial light modulator applications amply illustrate the complexity of the materials invention, modification, growth, and characterization problem. The situation is similar for the range of other important spatial light modulator component materials, as summarized in Tables II and V.

Recently, a new class of spatial light modulators has emerged that may offer significant advantages for the parallel processing in one and two dimensions of sequentially formatted data fields. These hybrid spatial light modulators seek to couple the extensive signal conditioning capabilities of silicon (and gallium arsenide) based integrated circuit technologies with the parallel processing capabilities of optical modulator element arrays. The advantages of such a hybridization include the capability of on-chip data coding and decoding, level restoration, spatial reformatting, rescaling, and baseline subtraction. The possibility of utilizing the excellent photosensitivity (both quantum efficiency and speed) of Si and/or GaAs devices also makes hybrid optical-to-optical converters of considerable interest.

Hybrid spatial light modulators under development are summarized in Table VI. These currently include the linear array total internal reflection (TIR) spatial light modulator,[9] the CCD-addressed deformable mirror device,[55,56] the CCD-addressed liquid crystal light valve,[57,58] and the GaAs CCD-addressed elecroabsorptive spatial light modulator.[59,60] Optically addressed versions of these

TABLE V. Spatial Light Modulator Materials

• Electrooptic	• Magnetooptic
• Acoustooptic	• Piezooptic
• Liquid crystal	• Deformable membrane
• Photochromic	• Photodichroic
• Electroabsorptive	• Photoconductive

TABLE VI. Hybrid Spatial Light Modulators

- Total internal reflection spatial light modulator[9]
- CCD-addressed deformable membrane device[58, 59]
- CCD-addressed liquid crystal light valve[57, 58]
- GaAs-CCD electroabsorptive spatial light modulator[59, 60]
- Optically modulated total internal reflection spatial light modulator[61]
- Optically addressed membrane spatial light modulator[62, 63]
- Si-photoconductor liquid crystal light valve[64]

devices include the optically modulated total internal reflection spatial light modulator,[61] the optically addressed membrane spatial light modulator,[62,63] the Si liquid crystal light valve,[64] and the GaAs CCD-addressed electroabsorptive spatial light modulator.[60]

In order to illustrate the additional materials requirements imposed by the hybridization incorporated in these device concepts, two examples will be discussed in more detail: the linear array total internal reflection spatial light modulator and the CCD-addressed liquid crystal light valve. The former is an example of a one-dimensional serial-to-parallel converter, while the latter is two-dimensional.

The linear array total internal reflection spatial light modulator[9] is shown schematically in Fig. 8. A silicon driver chip containing data formatting and address encoding electronics is configured with over 5000 individually addressable parallel electrode stripes. The electroded surface of the chip is placed in intimate contact with an optically polished face of a suitably oriented single crystal of lithium niobate. Proximity coupling is achieved by application of uniaxial pressure normal to the interface between the two components. An input data stream is serial-to-parallel segmented in subunits of 16 electrode addresses by on-chip circuitry. The data are encoded such that the data value d(m) associated with index m is represented by the difference in applied voltage between electrode m and electrode m + 1. The applied voltage between two adjacent electrodes creates a fringing field pattern within the lithium niobate electrooptic crystal by proximity coupling. The anamorphic collimated readout beam is expanded in the plane of the electrodes and internally reflected from the interface. The phase of the coherent wavefront is modulated by the local electric field through the linear electrooptic effect such that individual electrode pairs with nonzero voltage differences induce diffraction around a Fourier plane stop in a Schlieren readout configuration. As shown in Fig. 8, the same lens used to generate the Fourier plane produces a one-to-one mapping between each electrode pair on the TIR device and a corresponding point in the output plane.

The overall function of the linear array TIR spatial light modulator is thus to accomplish a serial/electronic-to-parallel/optical conversion. Such a converter can serve as a format device for a two-dimensional spatial light modulator that is optically addressed or as a combination format/input device in one-dimensional processing architectures. Although current device implementations are binary (designed for reprographic printing applications), the extension to analog signal levels involves only redesign of the on-chip circuitry.

The CCD-addressed liquid crystal light valve[57,58] is an example of a two-dimensional hybrid spatial light modulator. As shown in Fig. 9, the device incorporates a Si CCD serial-to-parallel converter to format and convert an input array of data into a two-dimensional

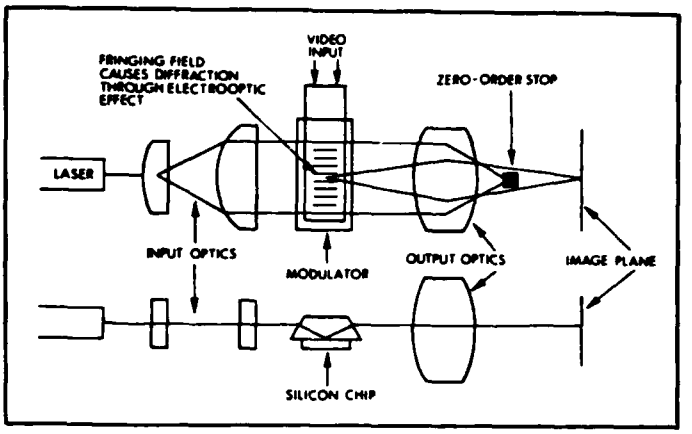


Fig. 8. Readout configuration for the linear array total internal reflection spatial light modulator (after Ref. 9). The top view shows the electrode structure at the interface between the silicon chip (with associated drive circuitry) and the proximity-coupled electrooptic crystal. The total internal reflection at the interface is shown in the side view (bottom half of the diagram).

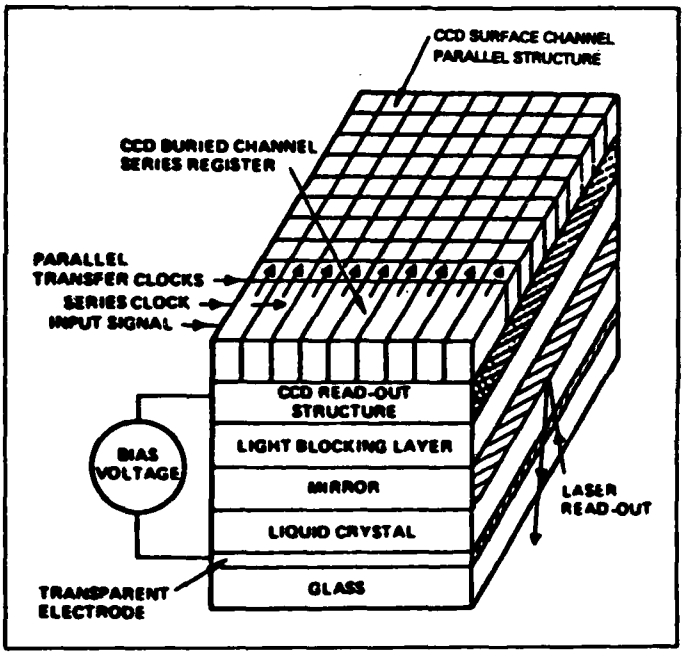


Fig. 9. Structure of the CCD-addressed liquid crystal light valve (after Refs. 57, 58), an example of a hybrid spatial light modulator.

stored charge pattern. The entire frame of space-variant charge is then transferred to the interface of the Si chip and an active liquid crystal layer. The two layers are separated by a light blocking layer and a dielectric multilayer mirror to separate the charge transfer and optical modulation functions. The local charge transfer process switches a fraction of the applied voltage from the Si layer to the liquid crystal layer, which modifies the spatial orientation of the liquid crystal molecules so as to induce a localized change in the optical activity and birefringence of the liquid crystal layer. This space-variant optical inhomogeneity can be sensed by polarization changes induced in a coherent readout beam.[57]

Materials issues common to hybrid spatial light modulators include the interfacial region between the electronic processing medium (silicon or gallium arsenide) and the optical modulation material, the optimization of the coupling between electronic device voltage modulation and the resultant induced optical modulation,

the optimum match between electronic and optical material operating modes, processing constraints imposed by each type of material on the other, overall device phase uniformity as limited by the ultimate polishing and assembly figure of the semiconductor wafer after final electronic processing, joint response time/format time limitations, the (usual) necessity for reflective readout geometries with concomitant doubling of figure-induced wavefront nonuniformities, and the necessity of operation at transistor-transistor logic (TTL)- or complementary metal-oxide semiconductor (CMOS)-compatible voltage levels (though the resultant electric field strengths can be quite high).

Among the many advantages of hybrid spatial light modulators are their natural interface with CCTV input devices, capability for high frame rate operation, potential for high resolution, geometric accuracy of interpixel spacing, and capability for on-device preprocessing of input data streams.

Some hybrid spatial light modulator configurations also allow for an additional format flexibility that has important consequences for highly parallel and pipelined systolic array processors.[65,66] As shown in Fig. 10, a much-needed device for application to such data-flow computational schemes is a two-dimensional scrolling spatial light modulator that accepts a parallel data stream input (or serial-to-parallel converts) and subsequently steps the data across the device by displacing each column data vector one column to the right with each processor iteration. Thus, the data element in row n, column m at time t is displaced to row n, column m + 1 at time t + 1 (in which time is measured in units of the iteration clock cycle period). Such scrolling spatial light modulators would find further immediate application in synthetic aperture radar image formation,[10-13] resulting in a scrolling output image at the pulse repetition frequency of the radar without the necessity of full-frame integration.

The multiplicity of applications outlined above for various categories of spatial light modulators has created a parallel multiplicity of parametric constraints on device performance requirements. The historical tendency has been to continually expand the requirement boundary for the "optimum" spatial light modulator to simultaneously include the maximum envisioned specifications from each potential application, without due regard to the fundamental interrelationships that necessarily exist among the various performance parameters. In the assessment of achievable device quality, on the other hand, it is often crucial to take fundamental physical limitations into account, and to carefully analyze the impact of such interrelated constraints as they may imply.

As a specific example, consider the dependence of the requisite device input sensitivity on resolution, dynamic range, and frame rate. Assume that it is desired to achieve a 1000×1000 array of resolution elements and that the spatial light modulator technology utilized will support a resolution of 50 lp/mm. This yields active device dimensions of 2 cm×2 cm, or an active device area of 4 cm^2. Assume further that a dynamic range of 1000:1 is required and that the device frame rate is to be 1 kHz. We now wish to examine the impact of these conjoint requirements on the device input sensitivity.

Consider next the statistical fluctuation of the integrated photon density that exposes a given device pixel during a single frame of operation. If we wish to discriminate among 1000 gray levels (which is equivalent to achieving a dynamic range of 1000:1), then we require the integrated photon density that participates in the effective analog decision for each pixel to be sufficient in magnitude to assure appropriate statistical accuracy. For iterative computations, in which errors tend to accumulate rather than average, we might require that the frame-to-frame fluctuation be less than 1 part in 1000 with a statistical certainty of 5 standard deviations or more. This implies immediately that the minimum number of photons per pixel per frame, N, required to participate in asserting the maximum intensity value must exceed 2.5×10^7 from statistical considerations alone. At a pixel density of 10^6 pixels per 4 cm^2, the photon density I must be greater than 6×10^{12} photons/cm^2 for each parallel decision plane. At a frame rate of 1 kHz, we then require that $I>6\times10^{15}$ photons/$cm^2\cdot$sec = 2 mW/cm^2 (at a wavelength of 632.8 nm). At a frame rate of 1 MHz, the required photon density will correspond to 2 W/cm^2.

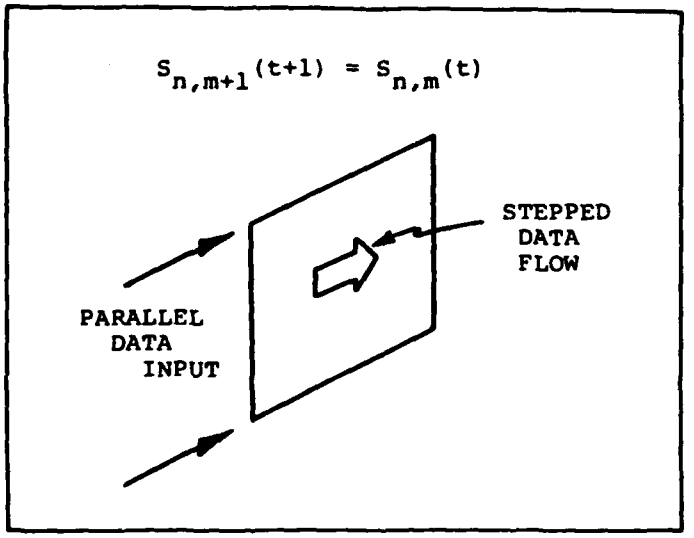


Fig. 10. Conceptual diagram of a scrolling spatial light modulator, which is useful for systolic array processing architectures and synthetic aperture radar image formation.

The required photon density at 1 kHz exceeds the input sensitivity of a wide variety of currently available spatial light modulators, in some cases by orders of magnitude. Hence, the achievement of greater input sensitivities in concert with the assumed constraints on resolution, dynamic range, and frame rate is counterproductive to optimized device performance. The intrinsic photon field statistical variation assures that it is in general not feasible to make rapid, highly precise optical decisions in low light level environments. It should be further noted that image amplification cannot relieve this fundamental limitation, as the signal-to-noise ratio of the amplified signal will be degraded in all cases relative to that characteristic of the input.

This is but one of a number of possible examples of potentially unrealizable combinations of device performance requirements. Such fundamental physical limitations affect not only achievable device development goals, but also directly impact the potential practicality of proposed system architectures and algorithms.

4. VOLUME HOLOGRAPHIC OPTICAL ELEMENTS

A second major device category of significant importance to the development of optical processing and computing systems is that of volume holographic optical elements. The primary function of this type of device is to detect the space-variant intensity pattern deriving from coherent object (information-bearing) and reference beams within the volume of an optically thick material (as depicted schematically in Fig. 11), to utilize the detected intensity to spatially modify a particular material property, and to generate sufficient modulation of the phase or amplitude of a readout beam to allow operation in the Bragg diffraction regime upon reconstruction of the object information. Features of this primary function include holographic storage of page-oriented information, Bragg-selective angle encoding of stored information, operating regimes that exhibit both linear and nonlinear response characteristics, and the capability for both parallel and sequential encoding.

Potential applications of volume holographic optical elements in optical processors and computers span a relatively wide range. Such devices can be configured as content-addressable memories[7,8] that allow fully parallel and simultaneous access to multiple planes of stored data driven by results of previous computations, without resort to intermediate address coding and decoding overhead. Several real-time optical processing and computing functions, such as correlation and convolution operations, have been demonstrated,[67]

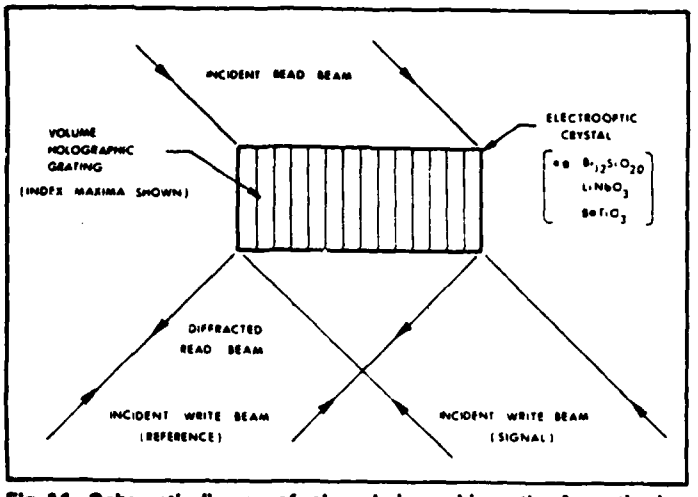


Fig. 11. Schematic diagram of volume holographic grating formation in a photorefractive material. The two incident write beams (signal and reference) interfere to form an index modulation that is sensed by the read beam, giving rise to a diffracted read beam that counterpropagates along the reference beam path.

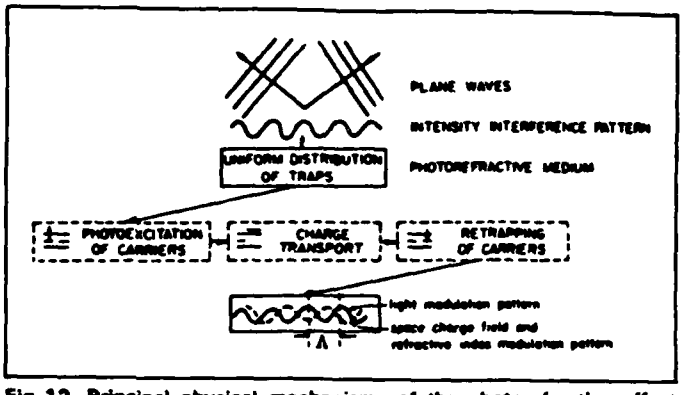


Fig. 12. Principal physical mechanisms of the photorefractive effect (after Ref. 72).

as have nonlinear operations such as image edge enhancement.[68] Volume holographic optical elements are capable of intermediate storage of computer-generated holograms[69] that can be sequentially written and reconstructed in parallel. Compensation for phase nonuniformities in an optical processor can be achieved by phase conjugation in degenerate four-wave mixing configurations,[70,71] which is of considerable importance when phase-sensitive optical algorithms must be implemented with imperfect optical elements. The Bragg selectivity of a thick volume hologram allows implementation of wavelength multiplexing and demultiplexing schemes for information storage and retrieval. Finally, volume holographic optical elements allow programmable optical interconnections to be envisioned with independent control of both of the orthogonal interconnection angles, and of the focal length of the interconnection, at each pixel location. The latter feature may allow multiplane interconnection architectures with threshold-response elements at each plane such that only the addressed element in a given plane is illuminated with above threshold switching energy. This concept allows great flexibility in optical processor and computer design, with concomitant compactness and reduction in the total number of required optical elements.

The utilization of volume holographic optical elements in processing and/or computing systems will be sharply limited unless these elements can be dynamically reprogrammed at high frame rates. Currently available materials for real-time volume holographic storage consist primarily of a class of photoconductive and electrooptic crystals (photorefractive materials) characterized by large electrooptic coefficients, quantum efficient photoconductivity, low carrier mobilities, high dark resistivities, wide band gaps, and moderate concentrations of deep traps (either intrinsic, extrinsic, or both).

The essential physical mechanism involved in volume holographic storage in electrooptic crystals is diagrammed schematically in Fig. 12.[72] The interference pattern created by two coherent writing beams coincident on the crystal gives rise to a periodic variation in the local optical intensity. Provided that the crystal absorbs at the incident wavelength and that the absorption mechanism is capable of creating mobile carriers, the local free charge distribution will exhibit the same periodicity as the interference pattern. The photogeneration of free charge can be induced by electron-hole pair excitation if the incident photon energy coincides with the energy gap between the valence and conduction bands, by free electron generation from filled deep donor states (leaving behind ionized donor levels with net positive charge), or by free hole generation from filled deep acceptor states (leaving behind ionized acceptor levels with net negative charge). A periodic variation in the net volume space charge distribution then results from diffusion-induced or drift-induced (in an applied electric field) separation of the photogenerated free carriers from either bound ionized states or free carriers of opposite charge, followed by retrapping in either shallow traps, ionized states of the type from which the photogenerated charge originated, or deep traps of other origin. This periodic space charge distribution exhibits the same spatial frequency as the original interference pattern, but will in general be shifted in phase as a function of the particular mechanism of charge separation.[73]

The space charge field that results from the stored charge distribution induces a periodic variation of the refractive index through the electrooptic effect. This index variation is in general sensitive to the polarization of readout illumination, and can comprise either simple volume phase gratings sensitive to a single component of the readout polarization, or birefringent phase gratings that couple to orthogonal components of the readout polarization.[74–76] The nature of the generated phase grating is in addition a function of the crystallographic space group symmetry of the photorefractive material, the orientation of the applied field (if any) with respect to the crystallographic axes, and the plane of incidence of the writing beams.

To date, research on photorefractive materials has concentrated on a number of electrooptic refractory oxides, including bismuth silicon oxide,[77,78] bismuth germanium oxide,[79] barium titanate,[80,81] strontium barium niobate,[82,83] potassium tantalate niobate,[84] and lithium niobate (both undoped and doped with iron).[85,86] These materials represent a wide range of saturation diffraction efficiencies, sensitivities, response times, and crystallographic perfection. The fundamental limitations of these performance characteristics have not yet been established. In particular, extensive characterization of these materials for application to photorefractive devices such as volume holographic optical elements has been the subject of much recent research interest, although the relationship between observed response characteristics and the growth and process parameters of the constituent single crystal materials is at the present time not at all clear. On the other hand, the potential of these materials for very high resolution storage of holographic information for time periods of very short to very long duration has been well documented.

An example of the promise held by such real-time photorefractive materials is shown in Fig. 13,[87] in which the square root of the holographic diffraction efficiency in bismuth silicon oxide (BSO) is plotted as a function of spatial frequency, with the applied electric field as a parameter. Note that at an applied field of 2 kV/cm, the characteristic curve is essentially flat from 0 to in excess of 2000 lp/mm. This broad bandwidth with concomitant adjustable modulation transfer function has not been observed in any other type of holographic or photographic recording medium.

Quite recently, significant photorefractive effects have been observed in GaAs:Cr,[88] in InP:Fe,[88] and in undoped GaAs.[89] These III-V semiconductor materials at present exhibit submicrosecond

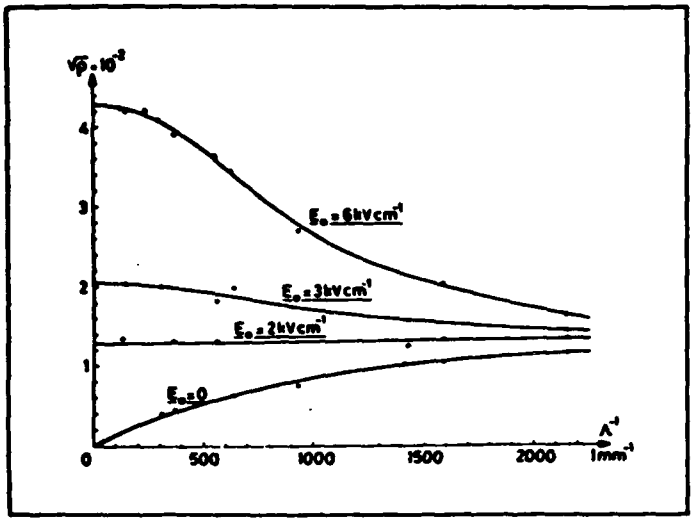


Fig. 13. Plot of the amplitude modulation transfer function for volume holographic storage in $Bi_{12}SiO_{20}$ (after Ref. 87).

response times with high sensitivity, primarily due to very large carrier mobilities, and may be capable of nanosecond time constants at modest power levels.[88] A drawback of the III-V and II-VI semiconductor materials is their relatively low values of the electrooptic coefficient, which delimits the saturation diffraction efficiency attainable without resort to large interaction volumes. The observation of enhanced nonlinearities in quantum well structures (such as gallium arsenide/aluminum gallium arsenide)[44,45] may indicate that appropriate electrooptic coefficients can be engineered in these materials to overcome this drawback, at the cost of severely limiting the potential interaction thickness of the device due to its inherent multilayer nature.

The range of currently investigated photorefractive materials for application to volume holographic optical elements is summarized in Table VII. A number of fundamental questions must be carefully analyzed and resolved before such devices can be practically incorporated in operational optical processing and computing systems. A number of these interrelated questions are discussed in the following paragraphs.

The maximum saturation diffraction efficiency achievable in a photorefractive material is of key importance, since it determines the overall efficiency of information storage and retrieval. The saturation diffraction efficiency is a function of a wide range of parameters (dependent on whether the charge transfer and storage process is engendered in the drift or diffusion regime), but for the most part observed diffraction efficiencies are of the order of 1 to 5%. A notable exception is the case of barium titanate, for which saturation diffraction efficiencies in excess of 30% have been reported.[90] This is due in part to extremely large values of the electrooptic coefficient characteristic of ferroelectric materials.[21]

To date, high saturation diffraction efficiency has appeared to be inversely correlated with device sensitivity (diffraction efficiency per unit incident grating writing energy). For example, bismuth silicon oxide (which has a moderate electrooptic coefficient[35]) has been reported as having diffraction efficiencies of order 4%[91] with high sensitivity.[92] Barium titanate, on the other hand, exhibits quite large diffraction efficiencies (as noted above) but is an order of magnitude less sensitive than BSO.[93] The fundamental origin of these differences is not yet clear, and indeed neither is the question as to whether or not these differences are in fact fundamental at all. The principal issue is the necessity for complete characterization of the as-grown single crystal materials to identify the nature, distribution, and time constants of both deep and shallow impurity levels so that correlations with the saturation space-charge field resulting from photoexcitation can be performed.

TABLE VII. Single Crystal Photorefractive Materials

- Bismuth silicon oxide ($Bi_{12}SiO_{20}$; BSO)
- Bismuth germanium oxide ($Bi_{12}GeO_{20}$; BGO)
- Lithium niobate ($LiNbO_3$, $LiNbO_3$:Fe)
- Barium titanate ($BaTiO_3$)
- Strontium barium niobate ($Sr_xBa_{1-x}Nb_2O_6$; SBN)
- Potassium tantalate niobate ($KTa_xNb_{1-x}O_3$; KTN)
- Gallium arsenide (GaAs, GaAs:Cr)
- Indium phosphide (InP, InP:Fe)
- Gallium arsenide multiple quantum well structures (e.g.,...$GaAs/Al_{1-x}Ga_xAs/GaAs$...)

A related issue is the identification of photorefractive device properties that can be significantly modified through appropriate changes in growth or process conditions. For example, optimization of the saturation diffraction efficiency in the drift-aided charge transport mode, sensitivity at a given wavelength, and response times for grating formation and erasure can most likely be achieved through control of growth-induced shallow traps and deep levels and through concomitant incorporation of selective dopants.

Further fundamental questions concern the maximum information capacity that can be stored per unit volume in a volume holographic optical memory. This issue is related to both the limiting resolution (highest spatial frequency for which significant modulation can be induced) and to practical considerations such as the allowable modulation transfer function of the optical system employed to write information into and retrieve stored information from the memory. Current materials seem capable of supporting very high capacities, as indicated in Fig. 13. The most likely limitation will derive from the corresponding optical system. In particular, the effective utilization of volume holographic optical elements as dynamically programmable optical interconnections depends on improvements in the available methods for recording computer-generated holograms at very high rates.

With regard to the ultimate information capacities achievable in volume holographic optical elements, a related question pertains to the minimum limits of interplane crosstalk in multiplane associative memories.[8] Crosstalk must be reduced to very low values for most applications, but even more so for iterative processing and computing functions performed in the analog domain.

A final comment on the status of real-time materials for volume holographic functions should be directed toward practical device design. The current research emphasis is correctly focused on the optimization of material properties for device applications, but significant effort must be devoted to the advanced development of implementable device configurations before such elements will be incorporated readily in advanced optical processing and computing systems.

An interesting coupling exists between the two types of devices discussed above, namely, spatial light modulators and volume holographic optical elements. The photorefractive incoherent-to-coherent optical converter[94-97] is a two-dimensional spatial light modulator that utilizes modified volume holographic recording techniques for information storage and reconstruction. The basic mechanism is illustrated in Fig. 14. Two coherent writing beams form a uniform carrier grating within the volume of a photorefractive single crystal material such as bismuth silicon oxide.[94] A spatially modulated incoherent erasure beam (input image) causes selective spatial erasure of the recorded carrier grating. Reconstruction of the modulated carrier grating then yields a coherent negative replica of the incoherent image intensity distribution. Operating modes that result in coherent positive replicas are also feasible.[96,97]

The carrier grating can be written before, during, or after the incoherent image exposure. Each operating mode exhibits distinct performance characteristics and operational advantages. In the

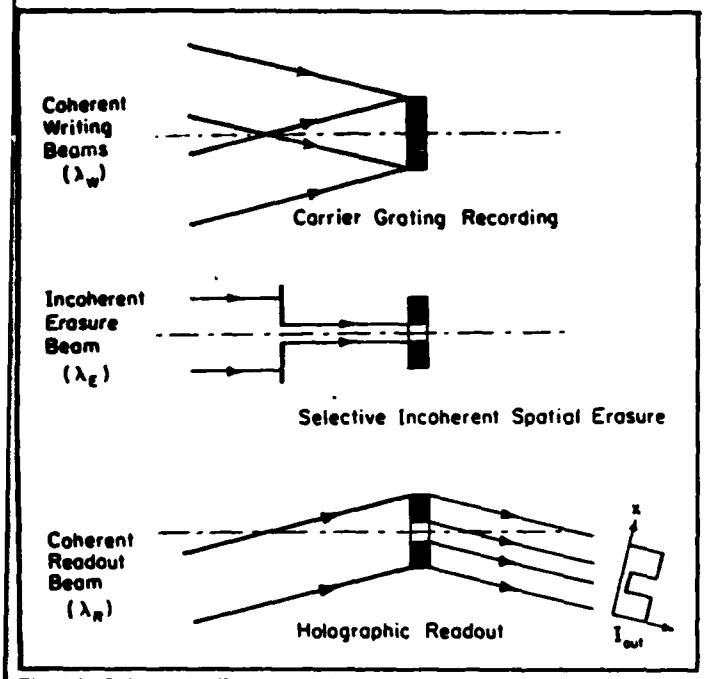


Fig. 14. Schematic diagram of the selective grating erasure process in the photorefractive incoherent-to-coherent optical converter (PICOC).

simultaneous erasure/writing mode (SEWM),[94-97] for example, the writing beams and incoherent erasure beam are coincident on the crystal, resulting in a steady-state variation in the effective modulation ratio of the carrier grating, as shown in Fig. 15. This configuration obviates the need for temporal sequencing of the writing and erasure beams and provides for simultaneous incoherent recording and coherent reconstruction of the input image. An experimental arrangement of the SEWM mode is shown schematically in Fig. 16, in which the direct analogy with degenerate four-wave mixing is immediately apparent.

From the materials perspective, it is of interest to note that the electrooptic crystal bismuth silicon oxide has been utilized as the active element in four separate types of electrooptic spatial light modulators. The relative directions of applied electric field, charge transport, sensed component of induced electric field, and writing/readout propagation directions are summarized for these four BSO-based ESLMs in Fig. 17. In the PROM,[23-26,35,36] the applied voltage (V_{app}) creates a longitudinal electric field divided between the electrooptic crystal and two dielectric blocking layers. The directions of the image-induced charge transport (J), sensed electric field component (E), and wave vectors of the writing and reading beams (K_i) are all parallel. In the PRIZ,[23,24,27,28] the crystallographic orientation of the electrooptic crystal is changed so that both longitudinal and transverse components of the image-modulated electric field are sensed (inducing both phase and amplitude modulation on the coherent readout beam). Volume holographic optical elements (VHOEs) utilizing BSO[72,73,89,92,93] are configured such that the applied field is transverse, as are the photoinduced charge transport and modulated space charge field directions. Information recording and reconstruction can be performed at arbitrary angles relative to the crystal normal, dependent on the desired spatial frequency of the carrier grating. Finally, in the case of the photorefractive incoherent-to-coherent optical converter (PICOC),[94-98] the incoherent information is in general incident at an angle distinct from that of the coherent writing beams. The optimum angle is in fact important in establishing the ultimate resolution of the device.[95-98]

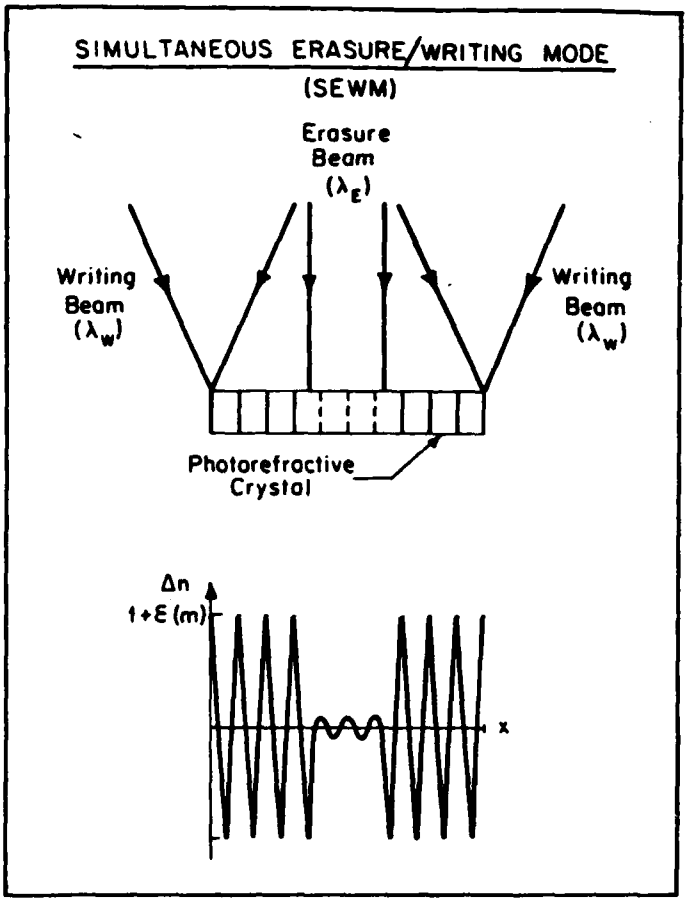


Fig. 15. Schematic diagram of the physical configuration and resultant index of refraction profile for the simultaneous erasure/writing mode of the photorefractive incoherent-to-coherent optical converter.

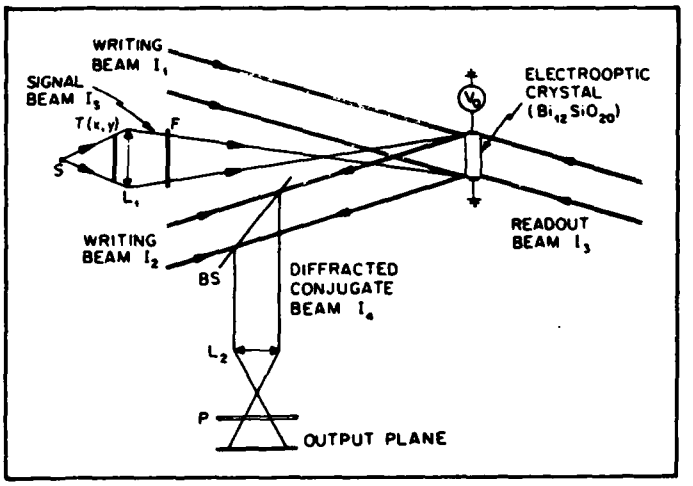


Fig. 16. Experimental arrangement for the demonstration of the simultaneous erasure/writing mode of the photorefractive incoherent-to-coherent optical converter (after Ref. 94). The transparency T(x,y) is imaged by lens L_1 and incoherent source S through spectral filter F onto the surface of the photorefractive device. A retroreflected readout beam derived from the incident writing beam I_1 generates a coherent replica of the transparency in the output plane through lens L_2 and polarizer P.

5. BISTABLE OPTICAL DEVICES

The primary function of a bistable optical device is to detect whether or not an incident intensity distribution exceeds a predetermined threshold at particular spatial locations, to utilize the detected intensity to spatially modify a local material property in such a fashion that distinct stable states correspond to regions illuminated below

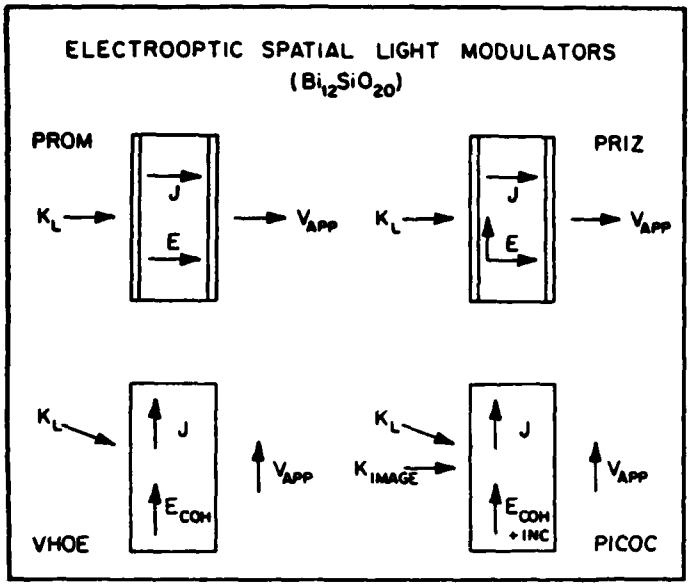


Fig. 17. Schematic diagram of the principal components of the applied voltage, charge transport, sensed electric field components, and input light wave vectors for four types of spatial light modulators that utilize single crystal bismuth silicon oxide ($Bi_{12}SiO_{20}$).

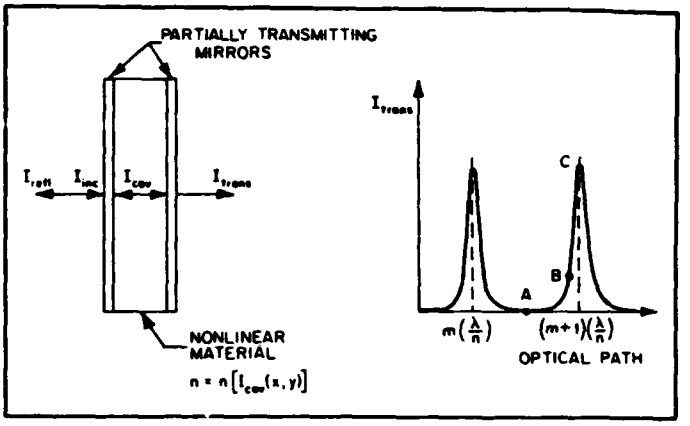


Fig. 18. Fundamental principles of optical bistability as observed in a nonlinear medium placed in a Fabry-Perot etalon.

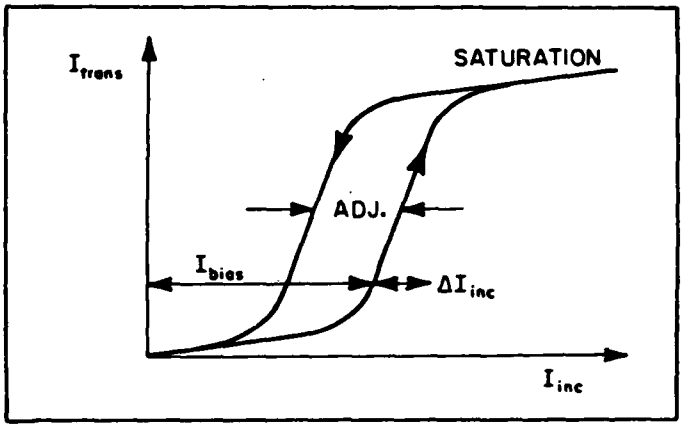


Fig. 19. Schematic illustration of the hysteresis and saturation effects characteristic of optical bistability.

and above threshold, and to encode the amplitude and/or phase of readout illumination with material state-dependent information.

Applications of bistable optical elements in optical processors and computers include the implementation of logic functions (AND, OR, NOR, etc.), level restoration, level amplification, bistable switching with momentary contact (requiring simultaneous input), bistable latching (requiring previous input), and variable thresholding. The use of bistable optical devices in these applications is envisioned in direct analogy with the use of semiconductor switching elements (transistors) in integrated electronic circuits. The fundamental difference is the availability of two-dimensional arrays of bistable optical elements that can be accessed in parallel by static or dynamically programmable optical interconnections.[14–16]

The parallel access feature is crucial to the eventual incorporability of bistable optical elements in optical processing and (especially) optical computing systems. The mere capability for constructing two-dimensional arrays of switching elements is not sufficient to provide the significant technological advantages required to warrant massive investments in additional primary (not to mention concomitant ancillary) switching technologies. After all, very large scale and wafer scale integrated circuits are in essence densely packed two-dimensional arrays of switching elements. If, however, reconfigurable interconnections can be implemented in the context of a given switching technology, distinct advantages accrue for the solution of extremely complex classes of computational and processing problems.[14–16] In this case, the distinction between the interconnections and the processor as separate functional elements blurs somewhat, allowing one to think of the combined entity as a reconfigurable machine. Thus, the excitement evident in the optics community as a result of recent significant advances in the capabilities of bistable optical devices derives partially from seemingly unrelated but very much interdependent advances in dynamically programmable optical interconnections.

In recent years, a number of fundamental physical effects have been utilized to produce bistable optical devices.[99–102] A large number of these devices depend for their operation on intensity-dependent modification of the effective optical path length of a Fabry-Perot etalon. The essential operational principles of this type of bistability can be illustrated with reference to Fig. 18. An optical material with intensity-dependent index of refraction is arranged between two partially transmitting mirrors to form a resonant cavity (as depicted graphically in the left-hand side of Fig. 18). Incident illumination causes a buildup of intensity within the cavity by means of multiple reflections between the partially reflective mirrors. The transmitted and reflected intensities from the device depend rather sensitively on the optical path length within the cavity, which determines whether net constructive or destructive interference occurs.

The dependence of transmitted intensity on the path length (plotted in units of the optical wavelength divided by the index of refraction) is depicted in the right-hand side of Fig. 18. Resonant behavior is observed such that peaks in the transmitted intensity occur whenever the path length is equal to an integral number of wavelengths (within the medium). For purposes of discussion, the relevant path length here is twice the thickness of the cavity.

Consider the initial optical thickness of the cavity to be such that the system resides at the point labeled "A" in Fig. 18. This point also corresponds to the origin of a plot of transmitted intensity as a function of incident intensity, as diagrammed in Fig. 19. As the incident intensity is increased, the intensity in the resonant cavity increases; thus, the transmitted intensity scales in proportion. Since we have assumed that the material is nonlinear in the sense that the index of refraction is a function of the intensity within the cavity, the increase in incident intensity alters the resonance condition. If the index of refraction is increased, the resonance will occur at a lower value of the path length. This is the equivalent of moving from "A" toward "B" in Fig. 18 (although in actuality the entire resonance curve plotted in the figure shifts to the left to achieve the same end).

Above a critical threshold in incident intensity, this situation is unstable. The intensity-induced shift in the resonance condition lowers the reflected intensity and increases both the transmitted

intensity and the intracavity intensity. The latter increase tends to increase the refractive index of the material further, which creates positive feedback and drives the system to a stable equilibrium condition, as represented by point "C" in Fig. 18. Further increases in incident intensity increase the cavity intensity, increasing the refractive index and hence *decreasing* the transmitted intensity. This results in negative feedback, which stabilizes the transmitted intensity in the saturation region of Fig. 19. Hysteresis occurs as a result of the fact that a relatively large decrease in the incident intensity is now required to significantly detune the cavity from the resonance condition and switch back to a region below the threshold for positive feedback. The existence of two distinct values of the transmitted intensity for a given value of the incident intensity is a clear indication of the presence of a bistable mechanism. As indicated in Fig. 19, the width of the bistable hysteresis characteristic is adjustable, and depends on the initial relationship between the actual path length of the cavity and the optical path length at zero input intensity.

The addition of a bias intensity level allows a number of different functions to be performed. For example, if the bias is set as indicated in Fig. 19, the device will amplify small additional signal intensities due to the steep slope of the characteristic curve. If binary input intensities are employed, the bias provides a significant increase in the optical sensitivity of the switch. The bias level can be set so that two units of coincident signal intensity are required to switch into saturation, which implements the Boolean logical AND function. Other logical operations can be performed with appropriate use of the bias input or of the phase delay between the bias input pulse and the signal inputs.[103]

Currently investigated materials that exhibit significant nonlinearities, and hence the capability for fabrication of parallel arrays of bistable optical elements, include gallium arsenide (GaAs),[104] indium antimonide (InSb),[105] copper chloride (CuCl),[106] cadmium sulfide (CdS),[107] zinc sulfide (ZnS),[108] zinc selenide (ZnSe),[108] and multiple quantum well structures of gallium arsenide/aluminum gallium arsenide ($GaAs/Al_{1-x}Ga_xAs$).[108–110] The physical mechanism utilized to create the necessary nonlinearity is free carrier dispersion in the case of indium antimonide and dispersion near exciton resonances or biexciton resonances in the other four cases.

Multiple quantum well structures are a special case of the general class of periodic multilayer materials, as shown schematically in Fig. 20. The layers are grown by molecular beam epitaxy (MBE), metal-organic chemical vapor deposition (MO-CVD), or liquid phase epitaxy (LPE) techniques. Layer thicknesses range from atomic monolayers (a few angstroms) to in excess of a thousand angstroms. Dependent on the nature of the structure, the total number of layers can range from of order 5 to of order 5000.

Progress in the development of bistable optical elements from these materials has been substantial in the past three years. Less than 1 ns response times have been observed for switching in CuCl (at 2 to 77 K), with peak intensities of order 10 MW/cm^2.[106] Room temperature switching has been reported for both $GaAs/Al_{1-x}Ga_xAs$ multiple quantum well structures[109] (20 to 40 ns at 100 kW/cm^2) and InSb[105] (30 ns at 100 kW/cm^2).

Recently, a novel type of optical bistability has been observed[111] in self-electro-optic-effect devices with the lowest bistable optical switching energies reported to date (of order 18 erg/cm^2). The device structure is a p-i-n photodiode with a $GaAs/Al_{1-x}Ga_xAs$ multiple quantum well structural element within the intrinsic region. The principle of operation involves the use of the photodetection capability of the structure in conjunction with the electric field dependence of the band edge (and associated exciton) absorption. The device is biased with a series resistor such that the incident light intensity causes photoconductive decay of the applied voltage across the device. The resultant reduction in the electric field shifts the band edge so as to induce additional photogeneration, further reducing the electric field and creating sufficient positive feedback to drive the device to saturation. The switching time constant of the device is limited by the RC time constant of the bias resistor in series with the device capacitance, and results in a trade-off between switching time (1.5 ms to 400 ns) and switching power (670 nW to 3.7 mW).

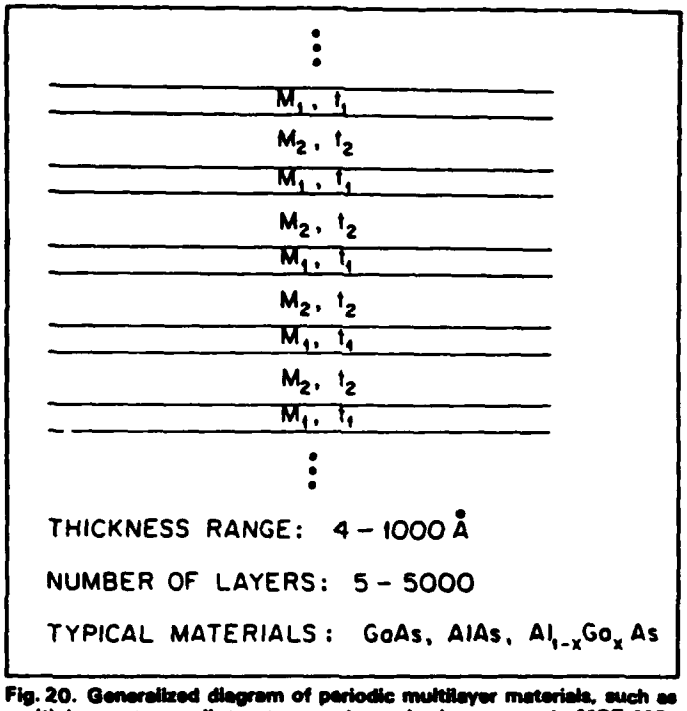


Fig. 20. Generalized diagram of periodic multilayer materials, such as multiple quantum well structures and superlattices, grown by MBE, MO-CVD, and LPE techniques.

A key factor in the rapid development of bistable optical switching elements has been the thoroughness with which the appropriate active materials have been characterized, in most cases for purposes other than optical bistability. Certainly in the case of the III-V semiconductors, and extending to MBE and MO-CVD multiple quantum well structures, knowledge of the underlying material parameters in combination with the availability of highly perfect single crystal materials has allowed research to progress quickly without the obscuration of irreproducible experimental results.

In addition to those mentioned in the beginning of this section, materials issues yet to be fully resolved in the area of bistable optical devices include the continued exploration and utilization of novel nonlinear interaction mechanisms, fabrication and evaluation of large arrays of devices with the uniformity and stability requisite of digital computation systems, and optimization of previously investigated nonlinear mechanisms by suitable tailoring of the material characteristics (particularly in the case of multiple quantum well structures). System integration of bistable optical device arrays will face significant difficulties, with maintenance of the extremely high average power levels required for continuous iterative switching algorithms. These high power levels require careful consideration of thermal dissipation, thermal nonuniformity with concomitant fluctuations in noise margin, beam disposition in cases where purely dispersive nonlinearities can be utilized, and the development of appropriate high power coherent or incoherent sources. As an example of the potential difficulty involved, it may be noted that the recent observation of room-temperature diode-laser-induced optical bistability in a $GaAs/Al_{1-x}Ga_xAs$ etalon[110] at 6 mW per switching element translates into an average continuous power requirement of approximately 6 kW for a 1000×1000 array operated at 100 MHz with unity duty cycle (assuming a 10 ns pulse duration).

6. IMPLICATIONS FOR OPTICAL PROCESSING AND COMPUTING

In this section, the implications of the current status of materials and devices for optical processing and computing applications are discussed.

With the possible exception of processors with one-dimensional architectures, present implementations of optical processing and computing algorithms and architectures are strongly device-limited. For the most part, algorithms have been developed independently of the current levels of device performance. The greatest fraction of existing algorithms for various optical processing and computing architectures have not been implemented with even experimental versions of devices; thus, performance trade-offs and algorithmic circumventions of hardware shortcomings have yet to be identified.

As pointed out in Sec. 1, several optical processor/computer component functions can be satisfied by a single device category, but not likely by a single device configuration within that category. Hence, a multiplicity of device configurations within each category will be required for complete systems. In addition, flexibility within a given device configuration to optimize device design with respect to the distinct figures of merit applicable to various applications is a highly desirable goal.

Device performance limitations are of three primary types: (1) fundamental, (2) design-related, and (3) materials-related. All three types of limitations are highly interrelated such that advances in each area are critical to advanced device development and device characterization. It is of particular importance when assessing the potential of a given device technology to determine whether reported experimental results derive primarily from fundamental physical limitations that cannot be circumvented, or in fact from technological limitations deriving from less than optimum choices (or availability) of active optical materials, device design parameters, device operating mode parameters, and/or performance characterization techniques.

Finally, it is important to note that significant developments in all of the device categories discussed above are required before the full promise of parallel optical processing and computing can be fulfilled. The phenomenal success of electronic processors and computers is due in no small part to the availability of and compatibility among a complete set of requisite components.

7. PROPOSED RESEARCH DIRECTIONS

In this final section, several recommended directions for future research effort in the area of materials and devices for optical information processing and computing systems are outlined.

Characterization and analysis of currently investigated optical information processing components should be accelerated. The focus of this effort should be placed on discriminating among those limitations to device performance characteristics that derive from fundamental physical origins, device configurations and operating modes, material properties, technological difficulties inherent in materials growth and processing, and technological difficulties inherent in device fabrication.

Extensive materials research programs are needed if advanced optical information processing components are to become a viable and economic part of the computational repertoire. Crystal growth and characterization, as well as thin film deposition and characterization, are research areas of proven merit in the development of contemporary microelectronics. Concomitant efforts in the area of optical materials will be required before devices with exceptional performance are likely to emerge. Two fundamental paths are worthy of significantly increased effort: the perturbation of known materials and the investigation of novel material systems. With regard to those materials that have demonstrated potential for a wide range of applicability to optical devices, the perturbation of material properties through growth-induced alterations in stoichiometry and/or selective incorporants is an extremely important and essentially unresearched avenue of opportunity. Concurrently, new materials categories (such as organic polymers with enhanced nonlinear optical properties, and periodic multilayer structures that allow for the engineering of certain optical properties and for inherent multifunction hybridization) offer untapped potential for a fundamental breakthrough. The availability of new or modified materials with selected optical properties that are in many cases enhanced only tenfold would have far-reaching implications on the viability of current optical device designs.

Significant advances will accrue to the design and implementation of algorithms and architectures with *existing* components. This interaction between the developers of systems and devices is crucial to the discrimination between critical and noncritical device shortcomings, and hence to focused efforts aimed at enhancing fundamentally important characteristics.

Finally, once critical device shortcomings are identified by actual incorporation in operational processing and computing subsystems, research effort should be directed to the circumvention of these limitations in the following order of priority. First, modify the algorithm if at all possible. Second, modify the subsystem architecture (with concomitant changes in the requisite algorithms). If both efforts fail to relieve the constraint, seek to enhance device performance characteristics through research efforts such as those outlined above. This order of priority is certainly familiar to users of electronic computational subsystems, for which chip and subsystem redesign even within an identifiably understood technology is viewed as both expensive and time-consuming.

8. ACKNOWLEDGMENTS

Research on materials and devices for optical processing and computing applications at the University of Southern California has been supported by the Defense Advanced Research Projects Agency (Office of Naval Research), the Air Force Office of Scientific Research, the Air Force Systems Command, the Army Research Office, the Naval Ocean Systems Center, the Joint Services Electronics Program, the National Science Foundation, the Jet Propulsion Laboratory (California Institute of Technology), ITT Corporation, Xerox Palo Alto Research Center, and Newport Corporation. The author wishes to express his gratitude for the gracious hospitality extended by the California Institute of Technology and by Newport Corporation during a sabbatical leave.

9. REFERENCES

1. John A. Neff, ed., *Optical Computing*, Proc. SPIE 456, (1984); John A. Neff, ed., Opt. Eng., Special Issue on Optical Computing, 24(1) (1985).
2. H. J. Caulfield, S. Horvitz, G. P. Tricoles, and W. A. Von Winkle, eds., Proc. IEEE, Special Issue on Optical Computing, 72(7) (1984).
3. W. T. Rhodes and P. S. Guilfoyle, "Acoustooptic Algebraic Processing Architectures," Proc. IEEE 72(7), 820 (1984).
4. D. Casasent, "Acoustooptic Linear Algebra Processors: Architectures, Algorithms, and Applications," Proc. IEEE 72(7), 831 (1984).
5. D. Psaltis, "Two-Dimensional Optical Processing Using One-Dimensional Input Devices," Proc. IEEE 72(7), 962 (1984).
6. D. Psaltis, "Input/output devices," in *Optical Computing*, John A. Neff, ed., Proc. SPIE 456, 66 (1984).
7. C. C. Guest and T. K. Gaylord, "Truth-Table Look-Up Optical Processing Utilizing Binary and Residue Arithmetic," Appl. Opt. 19, 1201 (1980).
8. M. M. Mirsalehi, C. C. Guest, and T. K. Gaylord, "Residue Number System Holographic Truth-Table Look-Up Processing: Detector Threshold Setting and Probability of Error Due to Amplitude and Phase Variations," Appl. Opt. 22(22), 3583 (1983).
9. R. V. Johnson, D. L. Hecht, R. A. Sprague, L. N. Flores, D. L. Steinmetz, and W. D. Turner, "Characteristics of the linear array total internal reflection (TIR) electrooptic spatial light modulator for optical information processing," Opt. Eng. 22(6), 665 (1983).
10. C. Elachi, T. Bicknell, R. L. Jordan, and C. Wu, "Spaceborne Synthetic-Aperture Imaging Radars: Applications, Techniques, and Technology," Proc. IEEE 70(10), 1174 (1982).
11. D. Psaltis and K. Wagner, "Real-time optical synthetic aperture radar (SAR) processor," Opt. Eng. 21(5), 822 (1982).
12. I. Abramov, Y. Owechko, A. R. Tanguay, Jr., and T. Bicknell, "Real Time Synthetic Aperture Radar Image Formation Utilizing an Electro-optic Spatial Light Modulator," Proc. NASA Spaceborne Imaging Radar Symposium, Jet Propulsion Laboratory Publication 83-11, 107 (1983).
13. E. N. Leith, "Synthetic Aperture Radar," in *Optical Data Processing: Applications*, Topics in Applied Physics, Vol. 23, D. Casasent, ed., pp. 89-117, Springer-Verlag, New York (1978).
14. A. A. Sawchuk and T. C. Strand, "Digital Optical Computing," Proc. IEEE 72(7), 758 (1984).
15. A. A. Sawchuk, "Digital logic and computing with optics," in *Optical Computing*, John A. Neff, ed., Proc. SPIE 456, 41 (1984).
16. J. W. Goodman, F. J. Leonberger, S.-Y. Kung, and R. A. Athale,

"Optical Interconnections for VLSI Systems," Proc. IEEE 72(7), 850 (1984).
17. A. W. Lohmann, "Chances for Optical Computing," Optik 65(1), 9 (1983).
18. J. W. Goodman, "Architectural Development of Optical Data Processing Systems," J. Electrical and Electronics Eng. Australia 2, 139 (1982).
19. B. H. Soffer, J. D. Margerum, A. M. Lackner, D. Boswell, A. R. Tanguay, Jr., A. A. Sawchuk, T. C. Strand, and P. Chavel, "Variable Grating Mode Liquid Crystal Device for Optical Processing and Computing," Mol. Cryst. Liq. Cryst. 70, 145 (1981).
20. P. Chavel, A. A. Sawchuk, T. C. Strand, A. R. Tanguay, Jr., and B. H. Soffer, "Optical Logic with Variable-Grating Mode Liquid Crystal Devices," Opt. Lett. 5, 398 (1980).
21. I. P. Kaminow, *An Introduction to Electrooptic Devices*, Academic Press, New York (1974).
22. C. C. Shih and A. Yariv, "A Theoretical Model of the Linear Electrooptic Effect," J. Phys. C. 15, 825 (1982).
23. Y. Owechko and A. R. Tanguay, Jr., "Theoretical Resolution Limitations of Electrooptic Spatial Light Modulators. I. Fundamental Considerations," *J. Opt. Soc. Am. A 1(6), 635 (1984).*
24. Y. Owechko and A. R. Tanguay, Jr., "Theoretical Resolution Limitations of Electrooptic Spatial Light Modulators. II. Effects of Crystallographic Orientation," J. Opt. Soc. Am. A 1(6), 644 (1984).
25. B. A. Horwitz and F. J. Corbett, "The PROM—theory and applications for the Pockels Readout Optical Modulator," Opt. Eng. 17(4), 353 (1978).
26. Y. Owechko and A. R. Tanguay, Jr., "Effects of Operating Mode on Electrooptic Spatial Light Modulator Resolution and Sensitivity," Opt. Lett. 7, 587 (1982).
27. M. P. Petrov, A. V. Khomenko, M. V. Krasin'kova, V. I. Marakhonov, and M. G. Shlyagin, "The PRIZ Image Converter and Its Use in Optical Data Processing Systems," Sov. Phys. Tech. Phys. 26, 816 (1981).
28. D. Casasent, F. Caimi, and A. Khomenko, "Test and Evaluation of the Soviet PROM and PRIZ Spatial Light Modulators," Appl. Opt. 20, 4215 (1981).
29. C. Warde and J. Thackara, "Operating modes of the microchannel spatial light modulator," Opt. Eng. 22(6), 695 (1983).
30. C. Warde, A. M. Weiss, A. D. Fisher, and J. I. Thackara, "Optical Information Processing Characteristics of the Microchannel Spatial Light Modulator," Appl. Opt. 20, 2066 (1981).
31. G. Marie, J. Donjon, and J.-P. Hazan, "Pockels-Effect Imaging Devices and Their Applications," in *Advances in Image Pickup and Display*, Vol. 1, B. Kazan, ed., Academic Press, New York (1974).
32. G. R. Knight, "Interface Devices and Memory Materials," in *Optical Information Processing: Fundamentals*, Topics in Applied Physics, Vol. 48, S. H. Lee, ed., pp. 119-121, Springer-Verlag, New York (1981).
33. J. Donjon, F. Dumont, M. Grenot, J.-P. Hazan, G. Marie, and J. Pergrale, "A Pockels-Effect Light Valve: PHOTOTITUS. Applications to Optical Image Processing," IEEE Trans. Electron Devices ED-20, 1037 (1973).
34. D. Casasent, S. Natu, T. Luu, G. Lebreton, and E. DeBazelaire, "New birefringence theory and uses of the photo-DKDP spatial light modulator in optical data processing," Proc. SPIE 202, 122 (1980).
35. A. R. Tanguay, Jr., "The Czochralski Growth and Optical Properties of Bismuth Silicon Oxide," Ph.D. Dissertation, Yale University (1977).
36. A. R. Tanguay, Jr., and R. C. Barker, "Implications of Concurrent Optical Activity and Electric Field Induced Birefringence for Pockels Readout Optical Modulator Performance," J. Opt. Soc. Am. 68, 1449(A) (1978).
37. Optical Processing Program, Electronic Sciences Division, Defense Sciences Office, Defense Advanced Research Projects Agency.
38. R. L. Holman and P. J. Cressman, "Optical damage resistance of lithium niobate waveguides," Opt. Eng. 21(6), 1025 (1982).
39. D. A. Bryan, R. Gerson, and H. E. Tomaschke, "Increased Optical Damage Resistance in Lithium Niobate," Appl. Phys. Lett. 44(9), 847 (1984).
40. D. A. Bryan, R. R. Rice, R. Gerson, H. E. Tomaschke, K. L. Sweeney, and L. E. Halliburton, "Magnesium-doped lithium niobate for higher optical power applications," Opt. Eng. 24(1), (1985).
41. R. A. Becker, "'Thermal Fixing' of Ti-Indiffused $LiNbO_3$ Channel Waveguides for Reduced Photorefractive Susceptibility," Appl. Phys. Lett. 45(2), 121 (1984).
42. G. Cooperman, L. Friedman, and W. L. Bloss, "Corrections to Enhanced Optical Nonlinearity of Superlattices," Appl. Phys. Lett. 44(10), 977 (1984).
43. W. L. Bloss and L. Friedman, "Theory of Optical Mixing by Mobile Carriers in Superlattices," Appl. Phys. Lett. 41(11), 1023 (1982).
44. D. A. B. Miller, D. S. Chemla, D. J. Eilenberger, P. W. Smith, A. C. Gossard, and W. T. Tsang, "Large Room-Temperature Optical Nonlinearity in $GaAs/Ga_{1-x}Al_xAs$ Multiple Quantum Well Structures," Appl. Phys. Lett. 41(8), 679 (1982).
45. D. A. B. Miller, D. S. Chemla, D. J. Eilenberger, P. W. Smith, A. C. Gossard, and W. Wiegmann, "Degenerate Four-Wave Mixing in Room Temperature GaAs/GaAlAs Multiple Quantum Well Structures," Appl. Phys. Lett. 42(11), 925 (1983).
46. K. Jain, G. W. Hewig, Y. Y. Cheng, and J. I. Crowley, "New Organic Materials with Large Optical Nonlinearities," IEEE J. Quantum Electron. QE-17(9), 1593 (1981).
47. G. F. Lipscomb, A. F. Garito, and R. S. Narang, "An Exceptionally Large Linear Electro-Optic Effect in the Organic Solid MNA," J. Chem. Phys. 75(3), 1509 (1981).
48. G. M. Carter, Y. J. Chen, and S. K. Tripathy, "Third-Order Nonlinear Susceptibility in Multilayers of Polydiacetylene," ACS Symposium Series No. 233, *Nonlinear Optical Properties of Organic and Polymer Materials*, D. J. Williams, ed., pp. 213-228, American Chemical Society (1983).
49. S. M. Sze, *Physics of Semiconductor Devices*, 2nd Edition, Wiley-Interscience, New York (1981).
50. W. F. Gorham, "A New General Synthetic Method for the Preparation of Linear Poly-P-Xylylenes," J. Polym. Sci., Pt. A-1, 4, 3027 (1966).
51. W. E. Loeb, "Encapsulation by Vacuum Deposition of Polymers," SPE J. 27, 46 (1971).
52. R. C. Taylor and B. Welber, "Laser-Monitored Deposition of Parylene Thin Films," Thin Solid Films 26, 221 (1975).
53. Y. Owechko, "Effects of Charge Transport and Crystallographic Orientation on Electrooptic Spatial Light Modulator Resolution and Sensitivity," Ph.D. Dissertation, University of Southern California (1983).
54. R. Aldrich, Itek Corporation, Optical Systems Division. Unpublished communication.
55. D. R. Pape and L. J. Hornbeck, "Characteristics of the deformable mirror device for optical information processing," Opt. Eng. 22(6), 675 (1983).
56. L. J. Hornbeck, "128×128 Deformable Mirror Device," IEEE Trans. Electron Devices ED-30(5), 539 (1983).
57. J. Grinberg, W. P. Bleha, P. O. Braatz, K. Chow, D. H. Close, A. D. Jacobson, M. J. Little, N. Masseti, R. J. Murphy, J. G. Nash, and M. Waldner, "Liquid-Crystal Electro-Optical Modulators for Optical Processing of Two-Dimensional Data," in *Effective Utilization of Optics in Radar Systems*, B. W. Vatz, ed., Proc. SPIE 128, 253 (1977).
58. U. Efron, P. O. Braatz, M. J. Little, R. N. Schwartz, and J. Grinberg, "Silicon liquid crystal light valves: status and issues," Opt. Eng. 22(6), 682 (1983).
59. R. H. Kingston and F. J. Leonberger, "Fourier Transformation Using an Electroabsorptive CCD Spatial Light Modulator," IEEE J. Quantum Electron. QE-19(9), 1443 (1983).
60. R. H. Kingston, B. E. Burke, K. B. Nichols, and F. J. Leonberger, "An electroabsorptive CCD spatial light modulator," in *Spatial Light Modulators and Applications*, Uzi Efron, ed., Proc. SPIE 465, 9 (1984).
61. S. McCahon, S. Kim, and A. R. Tanguay, Jr., "Optically Modulated Linear Array Total Internal Reflection Spatial Light Modulator," 1984 Annual Mtg. of OSA, FG-7; J. Opt. Soc. Am., summary to be published (1984).
62. D. R. Pape, "Optically addressed membrane spatial light modulator," Opt. Eng. 24(1), 107 (1985).
63. D. R. Pape, "An optically addressed membrane spatial light modulator," in *Spatial Light Modulators and Applications*, Uzi Efron, ed., Proc. SPIE 465, 17 (1984).
64. P. O. Braatz, K. Chow, U. Efron, J. Grinberg, and M. J. Little, "A Fast Silicon Photoconductor-Based Liquid Crystal Light Valve," Proc. Int. Electron Devices Mtg., 540 (1979).
65. H. J. Caulfield, W. T. Rhodes, M. J. Foster, and S. Horvitz, "Optical Implementation of Systolic Array Processing," Opt. Commun. 40, 86 (1981).
66. R. P. Bocker, "Optical digital RUBIC (rapid unbiased bipolar incoherent calculator) cube processor," Opt. Eng. 23(1), 026 (1984).
67. D. M. Pepper, J. AuYeung, D. Fekete, and A. Yariv, "Spatial Convolution and Correlation of Optical Fields Via Degenerate Four-Wave Mixing," Opt. Lett. 3(1), 7 (1978).
68. J. Feinberg, "Real-Time Edge Enhancement Using the Photorefractive Effect," Opt. Lett. 5(8), 330 (1980).
69. B. K. Jenkins and T. C. Strand, "Computer-generated holograms for space-variant interconnections in optical logic systems," in *International Conference on Computer-Generated Holography*, Sing H. Lee, ed., Proc. SPIE 437, 110 (1983).
70. D. M. Pepper, "Nonlinear optical phase conjugation," Opt. Eng. 21(2), 156 (1982).
71. J. P. Huignard, J. P. Herriau, P. Auborg, and E. Spitz, "Phase-Conjugate Wavefront Generation Via Real-Time Holography in $Bi_{12}SiO_{20}$ Crystals," Opt. Lett. 4(1), 21 (1979).
72. D. M. Kim, T. A. Rabson, R. R. Shah, and F. K. Tittel, "Photorefractive Materials for Optical Storage and Display," Opt. Eng. 16(2), 189 (1977).
73. R. J. Woods and L. Young, "Spatial Phase Shift During Hologram Writing in Lithium Niobate," Ferroelectrics 46, 275 (1983).
74. A. R. Tanguay, Jr., "Polarization Properties of Birefringent Phase Gratings," J. Opt. Soc. Am. 72(12), 1832 (1982).
75. A. R. Tanguay, Jr., "Polarization Properties of Birefringent Phase Gratings," to be published.
76. R. V. Johnson and A. R. Tanguay, Jr., "Anisotropic beam propagation analysis of birefringent phase gratings for optical device applications," to be published in Opt. Eng. 25(1), (1986).
77. M. Peltier and F. Micheron, "Volume Hologram Recording and

Charge Transfer Process in $Bi_{12}SiO_{20}$ and $Bi_{12}GeO_{20}$," J. Appl. Phys. 48(9), 3683 (1977).
78. A. Marrakchi, J. P. Huignard, and P. Gunter, "Diffraction Efficiency and Energy Transfer in Two-Wave Mixing Experiments with $Bi_{12}SiO_{20}$ Crystals," Appl. Phys. 24, 131 (1981).
79. Y. H. Ja, "Beam Coupling and Decoupling in Degenerate Two-Wave Mixing in a Reflection Geometry with Photorefractive $Bi_{12}GeO_{20}$ Crystals," Opt. and Quantum Electron. 16, 399 (1984).
80. J. Feinberg, "Self-Pumped, Continuous-Wave Phase Conjugator Using Internal Reflection," Opt. Lett. 7(10), 486 (1982).
81. S. Ducharme and J. Feinberg, "Speed of the Photorefractive Effect in a $BaTiO_3$ Single Crystal," J. Appl. Phys. 56(3), 839 (1984).
82. B. Fischer, M. Cronin-Golomb, J. O. White, A. Yariv, and R. Neurgaonkar, "Amplifying Continuous Wave Phase Conjugate Mirror with Strontium Barium Niobate," Appl. Phys. Lett. 40(10), 863 (1982).
83. K. Megumi, H. Kozuka, M. Kobayashi, and Y. Furuhata, "High-Sensitive Holographic Storage in Ce-Doped SBN," Appl. Phys. Lett. 30(12), 831 (1977).
84. D. von der Linde, A. M. Glass, and K. F. Rodgers, "High-Sensitivity Optical Recording in KTN by Two-Photon Absorption," Appl. Phys. Lett. 26(1), 22 (1975).
85. D. von der Linde, A. M. Glass, and K. F. Rodgers, "Multiphoton Photorefractive Processes for Optical Storage in $LiNbO_3$," Appl. Phys. Lett. 25(3), 155 (1974).
86. R. Grousson, M. Henry, S. Mallick, and S. L. Xu, "Measurement of Bulk Photovoltaic and Photorefractive Characteristics of Iron Doped $LiNbO_3$," J. Appl. Phys. 54(6), 3012 (1983).
87. J. P. Huignard, J. P. Herriau, G. Rivet, and P. Gunter, "Phase-Conjugation and Spatial Frequency Dependence of Wave-Front Reflectivity in $Bi_{12}SiO_{20}$ Crystals," Opt. Lett. 5(3), 102 (1980).
88. A. M. Glass, A. M. Johnson, D. H. Olson, W. Simpson, and A. A. Ballman, "Four-Wave Mixing in Semi-Insulating InP and GaAs Using the Photorefractive Effect," Appl. Phys. Lett. 44(10), 948 (1984).
89. M. B. Klein, "Beam Coupling in Undoped GaAs at 1.06 μm Using the Photorefractive Effect," Opt. Lett. 9(8), 350 (1984).
90. J. Feinberg, D. Heiman, A. R. Tanguay, Jr., and R. W. Hellwarth, "Photorefractive Effects and Light-Induced Charge Migration in Barium Titanate," J. Appl. Phys. 51(3), 1297 (1980).
91. J. P. Huignard, Thomson-CSF, Paris, France. Unpublished communication.
92. J. P. Huignard and F. Micheron, "High-Sensitivity Read-Write Volume Holographic Storage in $Bi_{12}SiO_{20}$ and $Bi_{12}GeO_{20}$ Crystals," Appl. Phys. Lett. 29(9), 591 (1976).
93. G. C. Valley and M. B. Klein, "Optimal properties of photorefractive materials for optical data processing," Opt. Eng. 22(6), 704 (1983).
94. Y. Shi, D. Psaltis, A. Marrakchi, and A. R. Tanguay, Jr., "Photorefractive Incoherent-to-Coherent Optical Converter," Appl. Opt. 22(23), 3665 (1983).
95. D. Psaltis, J. Yu, A. Marrakchi, and A. R. Tanguay, Jr., "Photorefractive incoherent-to-coherent optical conversion," in *Spatial Light Modulators and Applications*, Uzi Efron, ed., Proc. SPIE 465, 2 (1984).
96. A. Marrakchi, A. R. Tanguay, Jr., J. Yu, and D. Psaltis, "Photorefractive incoherent-to-coherent optical converter: physical and materials considerations," in *Spatial Light Modulators and Applications*, Uzi Efron, ed., Proc. SPIE 465, 82 (1984).
97. A. Marrakchi, A. R. Tanguay, Jr., J. Yu, and D. Psaltis, "Physical characterization of the photorefractive incoherent-to-coherent optical converter," Opt. Eng. 24(1), 124 (1985).
98. A. A. Kamshilin and M. P. Petrov, "Holographic Image Conversion in a $Bi_{12}SiO_{20}$ Crystal," Sov. Tech. Phys. Lett. 6(3), 144 (1980).
99. H. M. Gibbs, S. L. McCall, and T. N. C. Venkatesan, "Optical bistable devices: the basic components of all-optical systems?," Opt. Eng. 19(4), 463 (1980).
100. P. W. Smith and W. J. Tomlinson, "Bistable Optical Devices Promise Subpicosecond Switching," IEEE Spectrum, p. 26 (June 1981).
101. C. M. Bowden, M. Ciftan, and H. R. Robl, eds., *Optical Bistability*, Plenum Press, New York (1981).
102. E. Abraham, C. T. Seaton, and S. D. Smith, "The Optical Computer," Sci. Am. 248(2), 85 (1983).
103. J. L. Jewell, M. C. Rushford, and H. M. Gibbs, "Use of a Single Nonlinear Fabry-Perot Etalon as Optical Logic Gates," Appl. Phys. Lett. 44(2), 172 (1984).
104. A. Miller, D. A. B. Miller, and S. D. Smith, "Dynamic Non-Linear Optical Processes in Semiconductors," Adv. Phys. 30(6), 697 (1981).
105. A. K. Kar, J. G. H. Mathew, S. D. Smith, B. Davis, and W. Prettl, "Optical Bistability in InSb at Room Temperature with Two-Photon Excitation," Appl. Phys. Lett. 42(4), 334 (1983).
106. N. Peyghambarian, H. M. Gibbs, M. C. Rushford, and D. A. Weinberger, "Observation of Biexcitonic Optical Bistability and Optical Limiting in CuCl," Phys. Rev. Lett. 51(18), 1692 (1983).
107. M. Dagenais and W. F. Sharfin, "Cavityless Optical Bistability Due to Light-Induced Absorption in Cadmium Sulfide," Appl. Phys. Lett. 45(3), 210 (1984).
108. N. Peyghambarian, "Optical Bistability: A Novel Approach to Optical Signal Processing and Communications," Proc. NASA Optical Information Processing Conference II, (1984).
109. H. M. Gibbs, S. S. Tarng, J. L. Jewell, D. A. Weinberger, K. Tai, A. C. Gossard, S. L. McCall, A. Passner, and W. Wiegmann, "Room-Temperature Excitonic Optical Bistability in a GaAs-GaAlAs Superlattice Etalon," Appl. Phys. Lett. 41(3), 221 (1982).
110. S. S. Tarng, H. M. Gibbs, J. L. Jewell, N. Peyghambarian, A. C. Gossard, T. Venkatesan, and W. Wiegmann, "Use of a Diode Laser to Observe Room-Temperature, Low Power Optical Bistability in a GaAs-AlGaAs Etalon," Appl. Phys. Lett. 44(4), 360 (1984).
111. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, "Novel Hybrid Optically Bistable Switch: The Quantum Well Self-Electro-Optic-Effect Device," Appl. Phys. Lett. 45(1), 13 (1984).

Optical beam propagation method for birefringent phase grating diffraction

Richard V. Johnson, MEMBER SPIE
Armand R. Tanguay, Jr., MEMBER SPIE
University of Southern California
Optical Materials and Devices Laboratory
and
Center for Photonic Technology
Departments of Electrical Engineering
and Materials Science
Los Angeles, California 90089-0483

Abstract. *The analysis of volume diffraction gratings is important to the optimization of broad classes of devices, including acoustooptic Bragg cells, volume holograms, and spatial light modulators. A number of these devices involve gratings that exhibit striking polarization properties, which if used advantageously can offer significant opportunities for extended performance. Although the standard formalism for analyzing volume grating diffraction is the coupled wave approach, the optical beam propagation method has proven to be a powerful alternative formulation. For example, Thylen and Yevick recently utilized this method to study the coupling of polarized waveguide modes. In this paper, the optical beam propagation formalism of Thylen and Yevick for anisotropic media is further extended to the analysis of polarization effects in volume phase gratings. This method is both extremely intuitive (all intermediate mathematical steps have immediate physical significance) and also superior in cost of computation measures (execution time, computer memory required) for intricate modulation structures. Several broad classes of birefringent phase grating diffraction problems are identified, and sample solutions using the optical beam propagation method are exhibited. The cost of computation of the optical beam propagation method is shown to scale gracefully with increasing modulation complexity.*

Subject terms: optical information processing; gratings; birefringent phase gratings; polarization effects; spatial light modulators; optical activity; optical beam propagation method; coupled wave method; diffraction; anisotropic diffraction.

Optical Engineering 25(2), 235-249 (February 1986).

CONTENTS

1. INTRODUCTION

Volume diffraction gratings are fundamental to a large number of technologically important devices, including acoustooptic Bragg cells, spectrum analyzers, volume holographic optical elements, and spatial light modulators.[1] A number of these gratings have unusual optical polarization properties that can be advantageous for achieving significantly improved device performance.

An interesting and intricate example of a polarization diffraction problem is volume holography in photorefractive materials such as bismuth silicon oxide ($Bi_{12}SiO_{20}$, or BSO).[2-10] The analysis of the polarization states of the diffraction orders in bismuth silicon oxide is complicated by the presence of natural optical activity. In addition, an electric field is often applied to the crystal to increase the holographic photosensitivity. Because bismuth silicon oxide is electrooptic, an electric field induced linear birefringence appears that further complicates the polarization analysis.

In cases such as this, anisotropic Bragg diffraction can occur, in which the polarization state of the diffracted light beam is orthogonal to the polarization state of the incident light beam. Such orthogonality can enable improved signal-to-noise light ratio by the insertion of a polarization analyzer into the diffracted (signal) beam adjusted so as to block the scattered incident light without degrading the signal light intensity.[5-13] Anisotropic diffraction can also significantly extend the frequency bandwidth of acoustooptic Bragg cells by relieving phase matching constraints.[14-16]

An elegant formalism for analyzing the polarization properties of arbitrary media, including linear birefringence and natural optical activity, is the Jones matrix calculus.[17,18] The Jones matrix formalism has recently been applied to the

Invited Paper MD-106 received Oct. 18, 1985; revised manuscript received Nov. 18, 1985; accepted for publication Nov. 20, 1985; received by Managing Editor Dec. 5, 1985.

analysis of birefringent phase gratings, in which the polarization properties of the medium vary periodically from location to location.[19,20] This spatial variation can be Fourier decomposed into spatial harmonics, such that the resulting polarization matrix operator for each spatial harmonic governs the polarization state of a corresponding diffraction order. Such an analysis has been given by Tanguay et al. for the variable grating mode liquid crystal device.[19] In addition, a generic study of birefringent phase gratings in the thin grating limit using the Jones matrix formalism has been developed by one of the authors.[20]

The predictions of the Jones matrix analysis are most accurate for thin gratings. A thin grating enables the two principal processes of light propagation, namely, modulation and diffraction, to be clearly separated. The modulation process imposes a point-by-point local modification of the optical phase and/or polarization of the propagating wavefront, and thereby redistributes the energy in the angular spectrum of plane wave components of the light field. The intensity profile of the light field is redistributed following interaction with the grating because of the diffraction process, eventually breaking up into discrete diffraction orders. A key assumption of a thin grating analysis is that diffraction effects are so small that a ray optics model of light propagation inside the grating is sufficiently accurate to allow this separation of modulation and diffraction processes.

For thicker gratings, the diffraction and modulation processes become convolved, and the grating analysis becomes far more intricate. In particular, optical phase mismatch effects between the incident probe and the diffracted signal beams become important. Phase mismatch effects are exhibited, for example, in the angular alignment sensitivity of a grating, i.e., the variation of the diffraction efficiency with the angular alignment between the incident probe beam and the grating wave vector. Many of the birefringent phase gratings discussed in this paper exhibit distinctive angular alignment sensitivity curves.

A standard formalism for studying the diffraction behavior of thick volume gratings is coupled wave analysis, usually identified with Kogelnik.[21] Excellent reviews of coupled wave/coupled mode theory have been given by Russell[22] and by Gaylord and Moharam.[1] A coupled wave analysis of light propagation through homogeneous media with both linear birefringence and natural optical activity has been published by Yariv and Lotspeich.[23] Analyses of certain features of grating diffraction in bismuth silicon oxide volume holograms in the presence of optical activity and electric field induced linear birefringence have been presented by Kukhtarev et al.,[3] by Marrakchi et al.,[7,8] and by Vachss and Hesselink.[4] Rigorous coupled wave analyses for isotropic gratings have been published by Moharam and Gaylord.[24-26] A treatment of gratings with completely arbitrary optical anisotropies by means of rigorous coupled wave analysis has been given by Rokushima and Yamakita.[27]

Rigorous formulations of the coupled wave analysis can account for slanted gratings and/or large grating frequencies with concomitantly large diffraction angles. The various spatial harmonics of the light field include both propagating and evanescent diffraction orders. Although the evanescent orders do not participate in the transfer of optical power, they must be included in the analysis to satisfy the appropriate boundary conditions at the entrance and exit windows of the grating. The existence of backscattered light beams and the proper treatment of boundary conditions are critical issues in a rigorous analysis.

A considerably simplified grating analysis is possible when the probe beam remains essentially collimated. This occurs whenever the grating frequencies are sufficiently small, with low concomitant diffraction angles. Another constraint is that the birefringence of the grating medium, if any, must not cause a strong walkoff of orthogonally polarized laser beams. (See, however, Fleck and Feit for a treatment that includes such birefringent walkoff effects.[28]) In this approximate analysis the boundary conditions become trivial to satisfy, and the second order wave propagation equation can be well approximated by a first order equation in which backscattered orders are neglected. All higher diffraction orders with any measurable optical power are assumed to be propagating, not evanescent. Such assumptions are accurate only when the optical inhomogeneities defining the grating are sufficiently weak, as is the case for most of the device modeling reported in this paper.

When the above approximations are valid, alternative formalisms can be applied to grating analysis. One such formalism is the optical beam propagation method,[28-36] first proposed by Fleck et al.[29] The optical beam propagation method has been applied to a large variety of problems, including studies of Gaussian beam propagation through uniaxial birefringent media,[28] propagation of high energy laser beams through a turbulent atmosphere,[29] graded index fibers,[30] fiber optic couplers,[31] solitons,[32] grating lenses,[33] and optical waveguides with bends and tapers.[34] Reviews of the accuracy and applicability of the optical beam propagation method have been given by Van Roey et al.[34] and by Thylen.[35] Recently, Thylen and Yevick proposed an extension of the optical beam propagation method to the study of polarization effects induced by anisotropy, with application to polarization rotation in a double heterostructure pn-junction waveguide, and to TE-TM mode conversion in a lithium niobate slab waveguide.[36] In the present paper, the Thylen and Yevick formalism is further extended to the study of the polarization properties of birefringent phase gratings.

The optical beam propagation method is an extremely intuitive algorithm The essence of the method is to replace a computationally difficult distributed modulation/diffraction problem (as shown in Fig. 1) with a simpler "lumped element" system (in the terminology of transmission line analysis). As shown in Fig. 2, two distinct types of lumped elements are used in conjunction to model the interaction of the refractive and diffractive contributions to the overall optical wave propagation process in a volume phase grating. The first element is an infinitesimally thin plane that exhibits the same phase modulation and anisotropic polarization properties (space-variant birefringence) as a given finite thickness of the volume phase grating in the absence of diffraction effects. The second element comprises a *homogeneous* region of the same given finite thickness, within which the modulated wavefront is allowed to diffract without further space-variant modulation.

A volume phase grating cannot be adequately modeled by a single infinitesimally thin modulation/polarization plane paired with a single homogeneous region of diffraction. This is particularly true of the angular alignment sensitivity, which exhibits distinctly different behavior in the Raman-Nath and Bragg regimes. However, a finite number of such paired elements can in fact quite accurately approximate the full

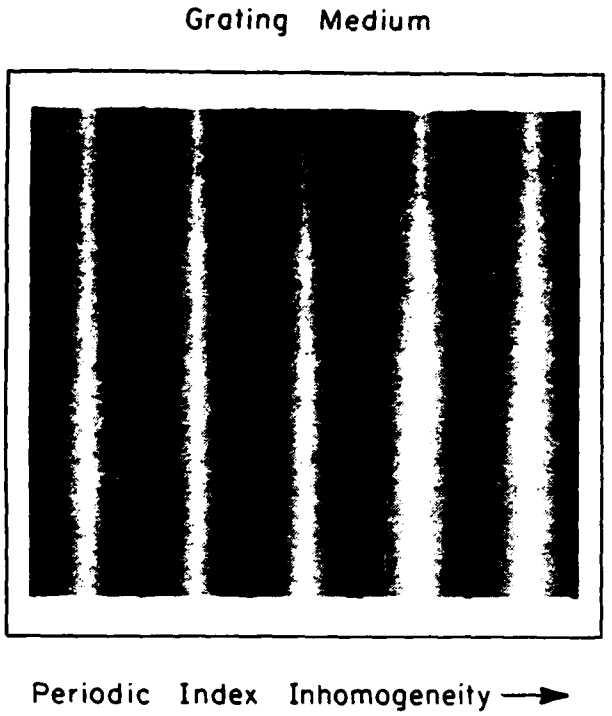


Fig. 1. A volume birefringent phase grating is shown, which consists of a spatially periodic variation in the birefringent properties of the grating medium. See also Fig. 6.

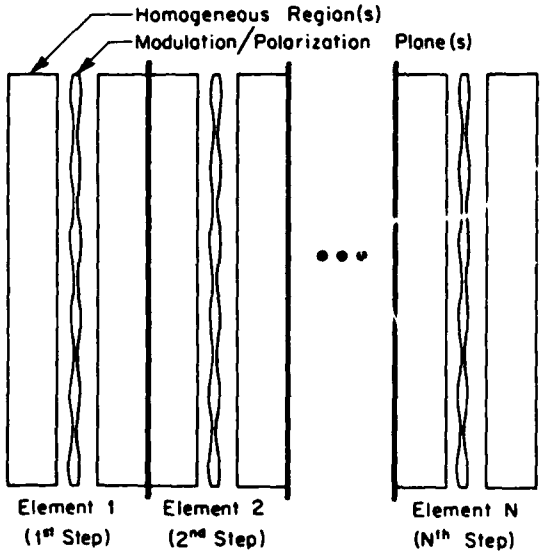


Fig. 2. Schematic diagram of the optical beam propagation method concept: replacement of a computationally difficult distributed modulation/polarization problem with a vastly simplified "lumped element" approximation consisting of optically homogeneous plates separated by infinitesimally thin phase and polarization modulation planes. The modulation planes are placed at the center of the homogeneous plates. The impact of each of these lumped elements on a coherent monochromatic light beam can be quickly calculated on a digital computer.

complexity of refraction, diffraction, and polarization phenomena observed in volume phase gratings. As will be shown in Sec. 4, the required number of paired elements is in fact surprisingly small, even for cases well into the Bragg diffraction regime.

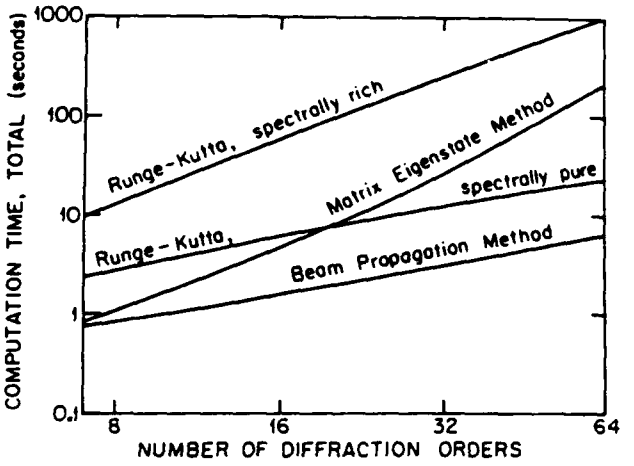


Fig. 3. Total computation time comparison of the various algorithms for a typical isotropic sinusoidal phase grating problem (Q = 20, v = 2, Bragg angle incidence, as defined in Sec. 2.3) to achieve better than a 0.1% error tolerance. The results shown here demonstrate the significant advantage of the optical beam propagation method for intricate spatial modulation problems involving more than seven diffraction orders. Details of the analysis presented in this figure are given in Sec. 3.

The computation of the light propagation through each associated pair of elements (i.e., the combination of a phase/polarization modulation plane and associated homogeneous diffraction/propagation region), hereinafter referred to as a computational step, is vastly simpler than the computation of the light propagation through the full distributed modulation system of Fig. 1. A comparison of the total computation time needed by the optical beam propagation method and by alternative coupled wave equation methods is given in Fig. 3 for a typical isotropic phase grating problem, as discussed more fully in Sec. 3. For this particular case, the optical beam propagation method gives the fastest computation times of any of the methods for problems involving more than seven diffraction orders. By dividing the volume phase grating into a large enough number of computational steps, the beam propagation model can approximate the actual distributed grating to any desired degree of accuracy. The numerical convergence to the asymptotic limit is quite rapid, as shown in Fig. 4 and discussed in Sec. 4.1.

All intermediate mathematical steps in the optical beam propagation method have a direct physical meaning, in contrast to alternative solution methods such as eigenstate analysis of the coupled wave equations. The method consists of a series of Fourier transforms (followed by phase/polarization modulation and inverse Fourier transforms), thus alternating between a near field and a far field description of the light beam; the coupled wave analysis, on the other hand, concentrates solely on a far field description of the light beam. The discrete diffraction order structure of the light beam is clearly resolved in the far field description, whereas the total light profile is propagated along a sequence of planes *within* the grating in the near field description. The near field description exhibits features such as a positive focusing effect in regions of locally maximum index of refraction, features that are not readily apparent with the coupled wave analysis. This dual view leads to a significantly heightened intuitive understanding of the modulation/diffraction process that makes the optical beam propagation method especially useful for teaching

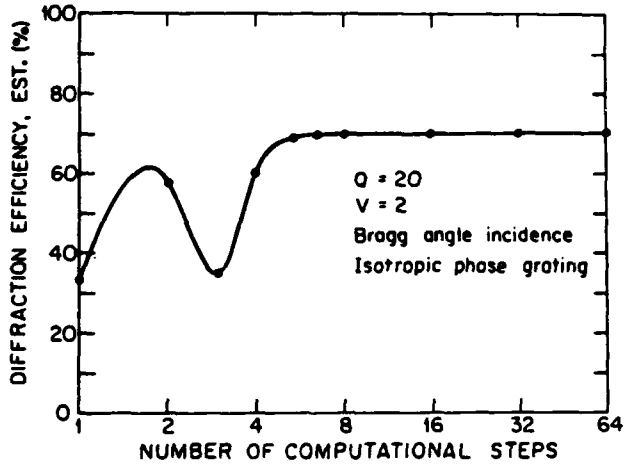


Fig. 4. Asymptotic convergence of the optical beam propagation method for one particular isotropic sinusoidal phase grating problem. As the number of computational steps increases, the diffraction efficiency calculated by the optical beam propagation method more closely approximates the desired analytical value.

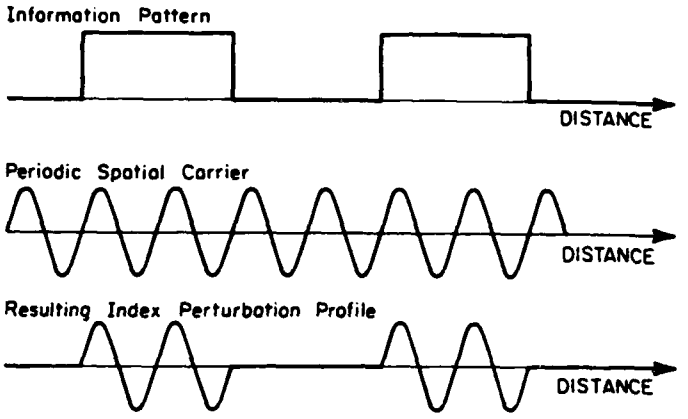


Fig. 5. Spatial modulation of a periodic spatial carrier. The mathematical algorithms for studying a simple grating can also be applied to the analysis of a more intricate spatial modulation problem, as shown here. The spatial modulation problem in general requires a finer numerical mesh (equivalent to more diffraction orders and suborders) to describe the problem adequately.

physical optics concepts to students, as well as for fundamental device invention and characterization.

The optical beam propagation method also proves to be superior to alternative methods in reduced execution time for the more intricate problems associated with modulated grating structures (as encountered, for example, in certain spatial light modulators). Several types of spatial light modulators involve a spatial carrier that is modulated by the information signal (as shown schematically in Fig. 5). Examples include acoustooptic Bragg cells,[37] the photorefractive incoherent-to-coherent optical converter,[38] the variable grating mode liquid crystal device,[19] the linear array total internal reflection (TIR) electrooptic spatial light modulator,[39] and the optically modulated version of the TIR spatial light modulator.[40] These modulated grating structures can be analyzed with the same computational methods used for simple phase gratings, but they require a larger number of numerical mesh points (equivalently, a larger number of diffraction orders) to describe adequately the more intricate modulation structure. Therefore, one important theme considered in this paper is the scaling of the computation time required by the various algorithms with increasing complexity of the spatial modulation profile, with results as shown in Fig. 3 and discussed in Sec. 3.

The mathematical basis for the approximate optical beam propagation analysis of volume birefringent phase gratings is presented in Sec. 2. The set of coupled wave equations, subject to this approximate treatment, is derived, and the optical beam propagation method is described. Numerical experiments performed on a Digital Equipment Corporation VAX 11/750 comparing the computation times required by the several algorithms are reported in Sec. 3. The scaling of the computation times for the various algorithms with increasing modulation complexity is also explored. The application of this extended anisotropic optical beam propagation method to several generic birefringent phase grating problems is demonstrated in Sec. 4.

2. DESCRIPTION OF THE OPTICAL BEAM PROPAGATION METHOD

A completely rigorous formulation of the grating problem requires a coupled wave analysis (or equivalently, a coupled mode analysis), in which the second order terms are retained in the differential wave equation and in which the boundary conditions at the entrance and exit windows of the grating are included.[24–26] However, many problems can be handled with perfectly acceptable accuracy by a simplified coupled wave formalism. All problems for which this simplified coupled wave analysis is applicable could equally well be treated by the optical beam propagation method. The necessary approximations are detailed in this section. The approximate treatment of the rigorous diffraction problem is formulated in Sec. 2.1, and typical birefringent phase gratings applicable to devices of interest are described in Sec. 2.2. The appropriate set of coupled wave equations, subject to this approximate formulation, is derived in Sec. 2.3, and various methods of numerical solution are discussed. The detailed implementation of the optical beam propagation method as applied to volume birefringent phase gratings is described in Sec. 2.4.

2.1. Approximate mathematical statement of the grating problem

An approximate wave propagation equation for inhomogeneous media [Eq. (4) below], which forms the basis for discussions of the coupled wave equation formalism and the optical beam propagation method, is derived in this section. The precise optical characteristics of a birefringent phase grating are specified by its constitutive relation. Broad aspects of constitutive equations for birefringent phase gratings are outlined in this section, and detailed examples are given in the next section.

The rigorous differential equation describing monochromatic wave propagation through a grating is derived from Maxwell's curl equations for a nonmagnetic medium. These equations combine to give the vector form of the wave equation:

$$\nabla \mathrm{x}(\nabla \mathrm{x}\mathbf{E}_\mathrm{T}) = \nabla(\nabla \cdot \mathbf{E}_\mathrm{T}) - \nabla^2 \mathbf{E}_\mathrm{T} = \mathrm{k}^2\, \mathbf{D}_\mathrm{T}\,, \tag{1}$$

in which $\mathbf{E}_\mathrm{T}$ is the electric field and $\mathbf{D}_\mathrm{T}$ is the electric displacement field associated with the light field, and k is the magnitude of the optical wave vector (defined *in vacuo*). The subscript T distinguishes the total light field from the slowly

varying envelope field, to be introduced in Eq. (3) below. The x coordinate is parallel to the grating wave vector; the z coordinate is perpendicular to the grating wave vector, lies in the plane of incidence, and is approximately parallel to the light propagation direction. For an optical beam that remains well collimated on diffraction (i.e., for low grating frequencies with concomitantly low diffraction angles), the term $\nabla(\nabla\cdot\mathbf{E}_T)$ is negligible compared with the $\nabla^2\mathbf{E}_T$ term. This yields the scalar form of the wave propagation equation; i.e.,

$$\nabla^2\mathbf{E}_T + k^2\mathbf{D}_T = 0 . \tag{2}$$

Although this equation is traditionally referred to as the scalar wave equation, it is nevertheless capable of describing the polarization effects of interest in this study.

An additional consequence of the collimation of the light beam (i.e., when the angular spectrum of plane wave components spans a small range of angles) is that the expression for the light field can be factored into the product of a slowly varying amplitude profile $\mathbf{E}(x,z)$ and a collimated and rapidly oscillating optical carrier:

$$\mathbf{E}_T(x,z) = \mathbf{E}(x,z)e^{in_0kx} . \tag{3}$$

This factorization assumes that all diffraction orders propagate approximately along the z axis. Small diffraction angle departures from the axis are expressed by the x dependence of the slowly varying amplitude.

In Eq. (3), n_0 denotes the geometrical average of the two indices of refraction associated with the birefringent phase grating medium. The geometrical average of the indices is selected because this decomposes the light field into an average optical phase evolution, which is quite rapid and exhibited by both polarization components, and birefringent deviations from this average phase evolution that are equal in magnitude and opposite in sign.

Substitution of this factored expression into the scalar wave equation [Eq. (2)] yields the following approximate wave propagation equation in terms of the slowly varying amplitude $\mathbf{E}(x,z)$:

$$\frac{\partial^2\mathbf{E}}{\partial x^2} + 2in_0k\frac{\partial\mathbf{E}}{\partial z} + k^2(\mathbf{D} - n_0^2\mathbf{E}) = 0 . \tag{4}$$

In this expression, a second order partial derivative $\partial^2\mathbf{E}/\partial z^2$ is neglected in favor of the first order derivative term $2in_0k(\partial\mathbf{E}/\partial z)$. This is valid only as long as the beam remains sufficiently well collimated, and for cases in which any slant angle of the grating with respect to the light propagation direction is small.

The description of the grating is embodied in the constitutive equation relating the light beam's D and E fields within the anisotropic and inhomogeneous medium:

$$\mathbf{D} = \epsilon(x,z)\mathbf{E} . \tag{5}$$

In typical optical media the permittivity ϵ is homogeneous, but in a grating the permittivity varies from point to point within the grating medium, and is periodic in the transverse coordinate x with spatial period Λ:

$$\epsilon(x,z) = \epsilon(x + \Lambda, z) \quad \forall x . \tag{6}$$

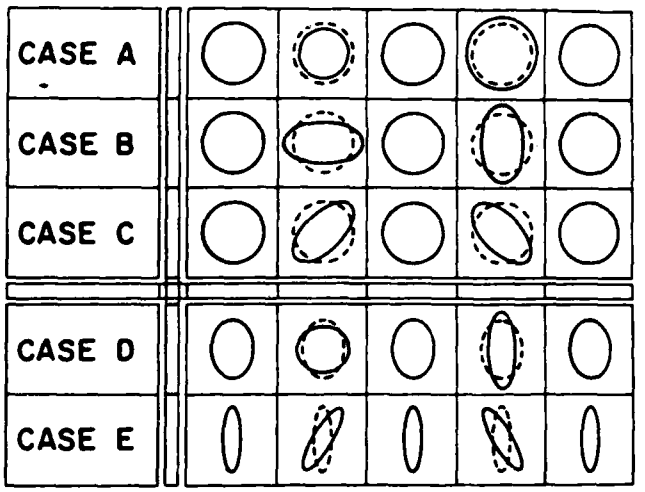


Fig. 6. Categories of birefringent phase gratings, in terms of the periodic spatial evolution of the index ellipsoid. The evolution is shown over one full period of the grating wavelength in the medium. Case A corresponds to an isotropic phase grating; cases B, C, and D correspond to different gratings with fixed principal axes and varying birefringence; case E corresponds to a grating with fixed birefringence and varying principal axes. Cases A, B, and C reside in an isotropic medium, while cases D and E reside in a medium with inherent optical anisotropy.

This periodicity implies that the permittivity can be Fourier decomposed into discrete spatial harmonics of the form

$$\epsilon(x,z) = \sum_{m=-\infty}^{+\infty} \epsilon_m(z)\exp(imK_Gx) \tag{7}$$

for all coordinates x and all integers m, in which K_G is the grating wave vector magnitude defined by

$$K_G = \frac{2\pi}{\Lambda} . \tag{8}$$

For the problems considered herein, the higher spatial frequency components ϵ_m are all significantly smaller than the zero frequency component $\epsilon_{m=0}$, which will be called the homogeneous component.

The permittivity $\epsilon(x,z)$ is a second rank tensor with an associated rank two matrix $\epsilon_{ij}(x,z)$, and the *mth* spatial harmonic is then represented by $\epsilon_{ij,m}(z)$. For a nonabsorbing medium in the absence of optical activity, the matrix ϵ_{ij} is symmetric with real components and can be represented by an abstract mathematical surface called the index ellipsoid.[41]

An isotropic medium is one for which the homogeneous permittivity tensor $\epsilon_{m=0}$ can be represented by the product of a scalar and a unit tensor, and its corresponding index surface is a sphere. A birefringent medium is one for which the homogeneous permittivity component $\epsilon_{m=0}$ cannot be so factored, and its corresponding index surface is an ellipsoid. Birefringent phase gratings can be classified by the spatial variation of the shape of this index surface, as discussed in the following section.

2.2. Typical birefringent phase gratings

The spatial variation in the index ellipsoid across one full grating period is exhibited in Fig. 6 for several categories of

gratings. Three broad categories are shown in Fig. 6: an isotropic phase grating [Fig. 6(A)], three versions of a grating with fixed principal axes and varying birefringence [Figs. 6(B), 6(C), and 6(D)], and a grating with varying axes and fixed birefringence [Fig. 6(E)]. The categories can be further classified into gratings embedded in isotropic media [Figs. 6(A), 6(B), and 6(C)] and gratings embedded in birefringent media [Figs. 6(D) and 6(E)]. The optical diffraction phenomena characteristic of these gratings are presented in Sec. 4. Note that Fig. 6 is not meant to contain an exhaustive list of grating classes, but rather to be representative of the rich variety of birefringent phase gratings characteristic of optical materials and devices.

An isotropic phase grating is defined herein as a grating that generates a diffracted signal beam with a polarization state identical to that of the incident probe beam, regardless of the incident polarization state. Figure 6(A) shows the spatial variation of the index ellipsoid for a typical isotropic grating, with a permittivity profile of

$$\epsilon(x) = \begin{pmatrix} n_0^2 & 0 \\ 0 & n_0^2 \end{pmatrix} + 2n_0\Delta n \cos(K_G x)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{9}$$

in which n_0 is the average index of refraction and Δn is the amplitude of the index modulation. The index surface in the absence of the grating modulation is a circle. The perturbation induced by the grating causes this surface to expand and contract, but always to remain a circle. Thus, no preferred principal axis system can be identified for this highly symmetric case. An acoustooptic interaction with a longitudinal sound wave in an isotropic medium such as glass approximates an isotropic grating (rigorously so in media for which the photoelastic tensor components p_{11} and p_{12} are equal).[42] Some of the diffraction properties of the isotropic phase grating are discussed in Sec. 4.1.

One of the most interesting birefringent phase gratings is that of fixed principal axes with spatially varying birefringence, superimposed on an isotropic medium. Two important cases can be identified, in which the principal axes are (1) aligned with the grating wave vector, as in Fig. 6(B), or (2) aligned at 45° with respect to the grating wave vector, as in Fig. 6(C).

The first case is typically characterized by a permittivity distribution of the form

$$\epsilon(x) = \begin{pmatrix} n_0^2 & 0 \\ 0 & n_0^2 \end{pmatrix} + 2n_0\Delta n \cos(K_G x)\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{10}$$

The corresponding spatial distribution of the index ellipsoid is shown in Fig. 6(B). The index surface in the absence of a grating is a circle. The grating perturbs the medium so that the index surface becomes elliptical, with the major axis of the ellipse alternately parallel and perpendicular to the grating wave vector.

The second case is as shown in Fig. 6(C), with a typical permittivity distribution of the form

$$\epsilon(x) = \begin{pmatrix} n_0^2 & 0 \\ 0 & n_0^2 \end{pmatrix} + 2n_0\Delta n \cos(K_G x)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{11}$$

The principal axes are rotated 45° with respect to the grating wave vector. One example of such a birefringent phase grating is an acoustooptic interaction with a transverse (shear) sound wave in an amorphous medium.[42] A second example is a volume hologram in a particular crystallographic orientation of cubic bismuth silicon oxide (for cases in which the naturally occurring optical activity can be neglected).[8]

The media in Figs. 6(A), 6(B), and 6(C) are all isotropic in the absence of the grating perturbation. Figure 6(D) shows the spatial distribution of the index ellipsoid for a birefringent phase grating in a medium that is birefringent. A volume hologram in a photorefractive electrooptic uniaxial crystal would have this distribution when the modulated space charge electric field $\mathcal{E}(x)$ is oriented along the crystallographic c axis.[5-9] For the case of lithium niobate, the permittivity profile would be

$$\epsilon(x) = \begin{pmatrix} n_o^2 & 0 \\ 0 & n_e^2 \end{pmatrix} + \mathcal{E}(x)\cos(K_G x)\begin{pmatrix} r_{13}n_o^4 & 0 \\ 0 & r_{33}n_e^4 \end{pmatrix}, \tag{12}$$

in which n_o is the ordinary index of refraction and n_e is the extraordinary index of refraction. The index surface for the medium in the absence of the grating perturbation is an ellipse. The action of the space charge field $\mathcal{E}(x)$ is to induce a slight change in the size and shape of this ellipsoid, but without rotating its principal axes.

Figures 6(B), 6(C), and 6(D) are all examples of birefringent phase gratings with fixed principal axes and varying birefringence. A grating for which the index ellipsoid exhibits varying principal axes and fixed birefringence is shown in Fig. 6(E). This type of behavior is observed, for example, in certain liquid crystal devices.[19] A typical permittivity profile for this case might be

$$\epsilon(x) = \begin{pmatrix} n_o^2 & 0 \\ 0 & n_e^2 \end{pmatrix} + (n_e^2 - n_o^2)\sin\theta(x)\begin{pmatrix} \sin\theta(x) & \cos\theta(x) \\ \cos\theta(x) & -\sin\theta(x) \end{pmatrix}. \tag{13}$$

The angle $\theta(x)$ that characterizes the orientational variation of the principal axes is assumed to be small, and is given by

$$\theta(x) = \theta_{max}\cos(K_G x). \tag{14}$$

For sufficiently small angles $\theta(x)$, Eq. (13) can be approximated by

$$\epsilon(x) = \begin{pmatrix} n_o^2 & 0 \\ 0 & n_e^2 \end{pmatrix} + (n_e^2 - n_o^2)\theta(x)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{15}$$

The diffraction properties of these classes of birefringent phase gratings are explored in Sec. 4.2.

Optical activity, which has not been considered in the birefringent phase grating cases presented thus far, can easily be included in the analysis by adding appropriate off-diagonal elements to the permittivity matrix.[3,4,7,8] These additional off-diagonal elements for a nonabsorbing grating medium must be imaginary and antisymmetric. Volume phase gratings in bismuth silicon oxide and bismuth germanium oxide include the effects of natural optical activity, and hence diffraction gratings in such photorefractive media exhibit quite striking polarization properties. The permittivity equation with optical activity included for the $\{\mathbf{K}_G \perp <001>\}$ orientation of bismuth silicon oxide is given by[8]

$$\epsilon(x) = \begin{pmatrix} n_0^2 & ig \\ -ig & n_0^2 \end{pmatrix} + 2n_0\Delta n \cos(K_G x)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} . \quad (16)$$

in which the parameter g is related to the optical rotatory power ρ (°/mm) by

$$g = \frac{\rho\lambda n_0}{\pi} . \quad (17)$$

Sample analyses of grating problems with optical activity are presented in Sec. 4.3.

2.3. Coupled wave formalism

The birefringent phase grating diffraction problem has now been completely formulated. Two methods for numerically solving this problem are discussed: the coupled wave formalism, as described in this section, and the anisotropic optical beam propagation method, as described in the next section.

The coupled wave formalism derives from a decomposition of the light field into discrete spatial harmonics. Two alternative decompositions of the field are presented below, leading to two distinct sets of coupled wave equations, depending upon the assumed wave vector matching condition. The first decomposition assumes that all diffraction orders exhibit the same $\exp(in_0kz)$ dependence, so that no explicit dephasing of adjacent diffraction orders is provided. In fact, however, each diffraction order can be represented as a plane wave-like field, such that the *m*th order propagates with a distinct angle θ_m and with a corresponding z dependence of $\exp(in_0kz\cos\theta_m)$. The z dependence of adjacent diffraction orders is therefore not identical, so that an optical dephasing of these orders can be explicitly included to form the second possible decomposition. Gaylord and Moharam have presented a more complete discussion of this point.[1]

Both diffraction order decompositions describe the same grating physics and hence must predict the same diffraction characteristics (within phase factors). Each, however, exhibits distinctly different properties for numerical analysis. Both formulations are needed in the following cost of numerical computation experiments, because only the first formulation is amenable to the matrix eigenstate analysis, while the second formulation exhibits much better convergence properties for numerical integration, as discussed in Sec. 3.

Consider a collimated incident light beam with propagation angle θ_{in} defined with respect to the z axis:

$$E_{in}(x,z) = E_0\exp[in_0k(x\sin\theta_{in} + z\cos\theta_{in})] . \quad (18)$$

This beam interacts with the grating to produce a light field that in the first formulation of the coupled wave equations is assumed to have the form

$$E(x,z) = \sum_{m=-\infty}^{+\infty} E_m(z)\exp[i(n_0k\sin\theta_{in} + mK_g)x] . \quad (19)$$

This can be alternatively expressed as

$$E(x,z) = \sum_{m=-\infty}^{+\infty} E_m(z)\exp(in_0kx\sin\theta_m) , \quad (20)$$

in which the angle θ_m is defined by

$$\sin\theta_m = \sin\theta_{in} + 2m\sin\theta_B . \quad (21)$$

and θ_B is the isotropic Bragg angle inside the grating, defined by

$$\sin\theta_B = \frac{K_G}{2n_0k} . \quad (22)$$

A given diffraction order, indexed by integer m and propagating with angle θ_m with respect to the z axis, is associated with each spatial harmonic in the transverse coordinate x.

The coupling among the diffraction orders is described by a set of linear first order differential equations of the form

$$\frac{dE_{i,m}(z)}{dz} = \sum_{j=1}^{2}\sum_{n=-\infty}^{+\infty} C_{ij,mn}(z)E_{j,n}(z) , \quad (23)$$

in which i and j indices specify the two polarization states while m and n indices specify the diffraction orders. Equation (23) results from substituting the diffraction order superposition Eq. (20), into the approximate light propagation Eq. (4), *subject to the constitutive* Eq. (5) *and the spatial* decomposition of the permittivity Eq. (7), and finally by collecting all terms with the same spatial harmonic dependence in the x coordinate.

The matrix element $C_{ij,mn}$ couples the light energy from the *n*th diffraction order, *j*th polarization state, into the m*th* diffraction order, i*th* polarization state. These coupling matrix elements are governed by the spatial frequency spectrum of the permittivity perturbation:

$$C_{ij,mn}(z) = \frac{ik}{2n_0}\epsilon_{ij,m-n}(z) - \frac{iQ}{2d}\frac{m + \sin\theta_{in}}{2\sin\theta_B}\delta_{ij}\delta_{mn} , \quad (24)$$

in which δ_{ij} is the Kronecker delta function (i.e., $\delta_{ij} = 1$ when $i = j$; $\delta_{ij} = 0$ otherwise). The dimensionless parameter Q scales the grating thickness d, defined by[43]

$$Q = \frac{K_G^2 d}{n_0 k} . \quad (25)$$

A dimensionless grating strength parameter v for sinusoidal isotropic gratings is defined for future reference as

$$v = k\Delta n d , \quad (26)$$

in which Δn is the maximum perturbation in the index of refraction.[43]

This first formulation of the coupled wave equations yields coupling coefficients $C_{ij,mn}$ that are independent of the z coordinate when the permittivity tensor is independent of the z coordinate. This is a necessary condition for applying the matrix eigenstate analysis formalism. The harmonic terms for the higher diffraction orders, however, are not solutions of the homogeneous wave equation, whereas we expect the various diffraction orders to approximate closely such homogeneous solutions. This forces the amplitude terms $E_m(z)$ to assume a phase evolution that rapidly becomes severe for higher diffraction orders.

The second formulation of the coupled wave equations avoids this problem by positing harmonic terms for each diffraction order that are solutions of the homogeneous wave equation. This formulation evolves from the following decomposition of the light field:

$$\mathbf{E}(x,z) = \sum_{m=-\infty}^{+\infty} \mathbf{E}_m(z)\exp\left[in_0k\left(x\theta_m - \frac{1}{2}z\theta_m^2\right)\right]. \quad (27)$$

This last equation derives from replacing the $\exp(in_0kz)$ term in Eq. (3) with an $\exp(in_0kz\cos\theta_m)$ term for each diffraction order, performing a Taylor series expansion on the sine and cosine terms in powers of the angle θ_m, and keeping terms only up to θ_m^2. This leads to a set of coupled wave equations of precisely the same form as Eq. (23), but with different coupling coefficients $C_{ij,mn}$, given by

$$C_{ij,mn} = \frac{ik}{2n_0}(\epsilon_{ij,m-n} - n_0^2\delta_{ij}\delta_{mn}) \times \exp\left[\frac{iQz}{2d}(m-n)\frac{m+n+\theta_{in}}{\theta_B}\right]. \quad (28)$$

The coupling coefficients $C_{ij,mn}$ defined by Eqs. (24) and (28) are cumbersome for numerical computations. Two modifications can greatly simplify the numerical implementation of various algorithms. The first modification involves elimination of unnecessary higher diffraction orders of negligible energy content, which can be implemented by truncating the infinite set of diffraction orders at some finite upper limit N. This upper limit is typically determined by computational experiment, i.e., by calculating the grating diffraction properties assuming a small value for N and then repeating the calculations for larger and larger values of N until the solutions reach an asymptotic limit.

The second modification to the coupling coefficients $C_{ij,mn}$ reduces the number of indices from four to two by combining indices i and m into a new index p, and similarly by combining indices j and n into a second index q, as explained in the next paragraph. Note that this is not a contraction of indices by symmetry arguments, but rather is a bookkeeping strategy to allow use of the large number of available matrix manipulation routines.

Recall that the subscript i of $C_{ij,mn}$ denotes the polarization state component of the *m*th diffraction order, and similarly the subscript j denotes the polarization state component of the *n*th diffraction order. Each diffraction order has two polarization state components, and these components can be indexed by integers 1 and 2. The diffraction order index m is truncated to lie in the range $-N \leq m \leq +N$, and similarly for order index n. One possible definition of the new index p in order to make it a positive integer might be $p = 2(m + N) + i$. Thus, the index p lies in the range $1 \leq p \leq 4N + 2$, and a unique pair of indices i and m is associated with each allowed value of p. Similar comments apply to index q.

The set of coupled wave equations can be solved by state variables methods, but only when the first decomposition of the diffraction orders is used, and then only when the permittivity tensor ϵ_{ij} is independent of coordinate z. Only for this restricted class of grating problems will the coupling coefficients $C_{ij,mn}$ be independent of coordinate z. In the state variables approach, the eigenvalues and eigenvectors of the coupling matrix $C_{ij,mn}$ are first calculated.[44] The incident light beam is decomposed into an equivalent set of eigensolutions of the coupled wave equation set. Each eigensolution is advanced in phase, corresponding to propagation through the grating, in proportion to its associated eigenvalue multiplied by the grating thickness. The eigensolutions are summed up at the exit window to define the exiting diffracted light beam.

The coupled wave equations can alternatively be solved by brute force numerical integration techniques, such as the Runge-Kutta method.[44] The computation time required for numerical integration is governed in part by which formulation of the coupled wave set is used, as discussed in Sec. 3, and in part by how densely packed the coupling matrix C_{pq} is. The packing density of the coupling matrix is driven in turn by the spectral purity or complexity of the permittivity perturbation profile [e.g., Eqs. (9) through (17)]. The usual profile assumed in most grating analyses is a simple sinusoid, as shown in Fig. 7(a). This profile has the purest possible spatial frequency spectrum, consisting of one positive frequency component $\exp(iK_Gx)$ and one negative frequency component $\exp(-iK_Gx)$ [Fig. 7(b)]. The corresponding coupling matrix is sparse, consisting only of three nonzero diagonal sets of elements: the main diagonal and the two adjacent diagonals (Fig. 8). Other perturbation profiles (such as, for example, a sawtooth spatial dependence) would have more intricate spectra, and hence their associated coupling matrices would be more densely packed with nonzero elements [Figs. 7(c) and 7(d), and Fig. 9]. Sparse coupling matrices have few interactions among the diffraction orders, typically involving only nearest neighbors, and so exhibit shorter computation times compared with densely packed coupling matrices.

2.4. The optical beam propagation method

A powerful tool for analyzing the polarization properties of an arbitrary medium is the Jones matrix calculus, which describes the propagation of collimated plane waves.[17,18] This is a most effective technique for calculating the polarization properties of gratings when they are sufficiently thin.[19,20]

For thicker gratings, an additional mechanism not considered in the Jones matrix calculus becomes important. That mechanism is the optical phase shift introduced into adjacent diffraction orders relative to each other as they propagate through the grating medium, caused by the differing propagation directions of these plane wave components. This phase shift introduces a dephasing of the coupling between the diffraction orders, which in turn causes the striking angular alignment sensitivity and diffraction order selectivity characteristic of thick phase gratings.

The essence of the optical beam propagation method is to replace a computationally difficult distributed modulation/diffraction/polarization problem (Fig. 1) with a much simpler approximate lumped element system (Fig. 2). The grating is conceptually divided into a series of plane parallel plates, such that each plate is thin enough to permit application of the Jones matrix calculus. The polarization and phase modulation characteristics of each plate, integrated over the thickness of the plate, are described by a rank two matrix applied on a point-by-point basis that represents the spatial inhomogeneity of the grating. An equivalent infinitesimally thin modulation/polarization plane can be defined that exhibits the same Jones matrix distribution, and which then comprises the first element of the optical beam propagation method. Standard practice in the optical beam propagation method literature is to

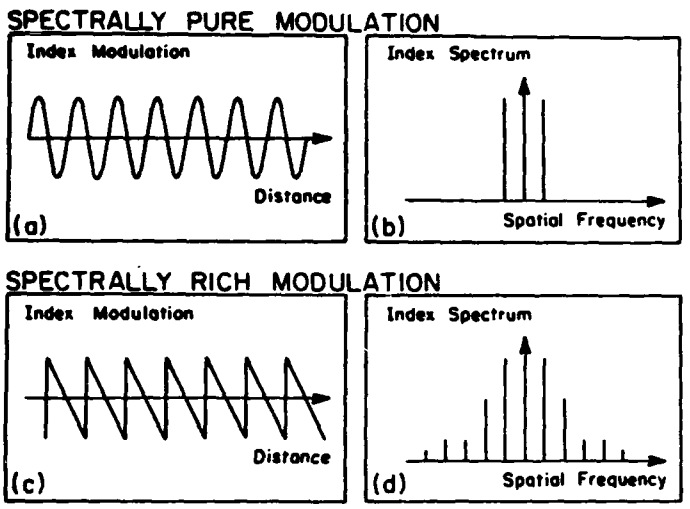


Fig. 7. Schematic diagram of spectrally pure and spectrally rich index perturbation profiles, and their concomitant spatial frequency spectra. Most phase grating analyses typically assume a sinusoidal perturbation. This leads to a simplified coupling of the diffraction orders in which only nearest neighboring orders interact. A more general index perturbation profile has a more intricate spatial frequency spectrum, with a consequently more intricate interaction of nonneighboring diffraction orders. This greater coupling complexity induces significantly greater computation times for numerical integration algorithms such as the Runge-Kutta.

locate this infinitesimally thin modulation plane at the center of the plate, as shown in Fig. 2.

The optical dephasing of the diffraction orders is accounted for in the optical beam propagation method model by introducing an additional class of element into the approximation. That element is a spatially homogeneous (albeit possibly birefringent) region. Since each plate is by convention assumed to be bisected by the modulation/polarization plane, the plate consists of the following sequence of three elements: a homogeneous region spanning the first half of the thickness of the plate under analysis, followed by the modulation/polarization plane, and completed by a second homogeneous region spanning the second half of the plate. In practice, the final homogeneous region of one plate is often computationally combined with the initial homogeneous region of the next contiguous plate, as shown in Fig. 2. The light propagation through each homogeneous region can be easily calculated by Fourier spectral techniques, as described below.

The combination of a modulation/polarization plane and a homogeneous plate is called hereinafter a "computational step." By dividing the grating into a sufficiently large number of steps, the calculated grating diffraction results can be brought arbitrarily close to the desired fully distributed modulation result.

The class of grating problems that can be treated by the optical beam propagation method is limited only by the class of problems that can be treated by the Jones matrix or equivalent calculus. Birefringent phase gratings, absorption gratings, and dichroic absorption gratings are all amenable to this formalism. Hereinafter, we restrict our attention to birefringent phase gratings of the type described in Sec. 2.2.

The full distributed modulation/polarization problem is described by the wave propagation equation

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} + 2in_0k\frac{\partial \mathbf{E}}{\partial z} + k^2[\epsilon(x,z) - n_0^2]\mathbf{E} = 0\,, \tag{29}$$

$$\begin{pmatrix} D & a & 0 & 0 & 0 & 0 & 0 \\ a & D & a & 0 & 0 & 0 & 0 \\ 0 & a & D & a & 0 & 0 & 0 \\ 0 & 0 & a & D & a & 0 & 0 \\ 0 & 0 & 0 & a & D & a & 0 \\ 0 & 0 & 0 & 0 & a & D & a \\ 0 & 0 & 0 & 0 & 0 & a & D \end{pmatrix}$$

Fig. 8. Form of the coupling matrix for a spectrally pure perturbation profile, as shown in Fig. 7(a).

$$\begin{pmatrix} D & a & b & c & e & f & g \\ a & D & a & b & c & e & f \\ b & a & D & a & b & c & e \\ c & b & a & D & a & b & c \\ e & c & b & a & D & a & b \\ f & e & c & b & a & D & a \\ g & f & e & c & b & a & D \end{pmatrix}$$

Fig. 9. Form of the coupling matrix for a spectrally rich perturbation profile, as shown in Fig. 7(c).

which results from combining Eqs. (4) and (5). The Jones matrix analysis derives from an approximation to this equation by assuming that all light beam components travel exactly parallel to the z axis. This assumption is equivalent to neglecting the $\partial^2\mathbf{E}/\partial x^2$ term. With this approximation, the resulting equation

$$2in_0k\frac{d\mathbf{E}}{dz} = k^2[n_0^2 - \epsilon(x,z)]\mathbf{E} \tag{30}$$

can be readily integrated over region Δz to give an expression of the form

$$E_i(\Delta z) = M_{ij}(x)E_j(z=0)\,, \tag{31}$$

in which $M_{ij}(x)$ is the Jones matrix associated with the grating region Δz. Evaluation of this expression comprises the first (modulation/polarization) element of the optical beam propagation method.

The second element of the beam propagation approximation is an optically homogeneous plate through which the light beam is propagated. The homogeneous propagation is described by

$$\frac{\partial^2 \mathbf{E}}{\partial x^2} + 2in_0k\frac{\partial \mathbf{E}}{\partial z} + k^2[\epsilon_{m=0} - n_0^2]\mathbf{E} = 0\,, \tag{32}$$

as observed from Eq. (29). This homogeneous propagation

problem is readily solved by decomposing the light beam into the polarization eigenstates of the permittivity matrix ϵ_{m-o} and by using Fourier spectral analysis techniques, i.e., by further decomposing the light field into plane wave components. The spectral decomposition can be accomplished very quickly and efficiently on a digital computer by the fast Fourier transform (FFT) algorithm.[45,46] Each spatial frequency corresponds to a plane wave component propagating at an angle θ_m with respect to the z axis. A beam of light traversing a homogeneous plate of thickness d and nominal index of refraction n_0 undergoes an optical phase shift of $\psi(\theta_m)$, where

$$\psi(\theta_m) = \frac{2\pi n_0 d}{\lambda \cos\theta_m}. \tag{33}$$

For a nearly collimated laser beam, the angles θ_m are small, so that the cosine can be well approximated by the first two terms in a Taylor series expansion. The first term of the Taylor series expansion of the optical phase shift then consists of a term $2\pi n_0 d/\lambda$, which is constant for all angles and hence does not contribute to the physics of the diffraction problem. The second term in the expansion consists of a quadratic phase term $-\pi n_0 d\theta_m^2/\lambda$, which contains the essential physics of the diffraction process.

The computational process within the second (homogeneous) element then is to Fourier transform the light beam profile at the entrance face of the homogeneous region. Each spatial frequency component is then phase shifted in proportion to the region thickness. Finally, an inverse Fourier transform is performed on the phase-modulated spectrum to obtain the spatial profile at the exit window of the homogeneous region.

The mathematical steps involved in the full implementation of the optical beam propagation method consist of alternate phase modulations and Fourier transforms. These alternating Fourier transforms allow the analyst to monitor the evolution of the laser beam simultaneously in the wavefront amplitude and diffraction order component pictures. The amplitude of the beam wavefront can be observed to concentrate into regions of locally maximum index of refraction because of positive lensing/waveguiding effects. Simultaneously, the interchange of optical power among the diffraction orders can be observed.

Further details on the optical beam propagation algorithm can be found in Ref. 29. Analyses of the applicability and accuracy of the optical beam propagation method can be found in Refs. 34 and 35.

3. COST OF COMPUTATION EXPERIMENTS

The various methods of grating analysis described in this paper have been programmed on Digital Equipment Corporation VAX 8600's, 780's, 750's, and several brands of microcomputers. A set of numerical experiments was performed on a VAX 11/750 to compare the execution times of alternative algorithms for solving the grating diffraction problem. For simplicity, only the execution times for the isotropic phase grating problem were tested, rather than the full birefringent phase grating problem. The algorithms considered in this study include coupled wave analysis (both eigenstate analysis and numerical integration) and the optical beam propagation method. The eigenstate analysis is based on a Jacobi iterative routine for diagonalizing the coupling matrix.[44] The numerical integration algorithm is a fourth order Runge-Kutta method.[44] The diffraction problem of the optical beam propagation method is solved by a fast Fourier transform spectral analysis.[45,46] All programs were written in FORTRAN and in some cases employ the International Mathematical and Statistical Libraries (IMSL) package of library routines.[47]

Two categories of permittivity perturbation profile are considered: a simple sinusoid and a square wave train of 50% duty cycle. The number of diffraction orders considered in the problem ranged from 7 (corresponding to a standard grating diffraction problem) up to 64 (corresponding to a typical modulated carrier problem).

Results of the numerical experiments with isotropic phase gratings are shown in Figs. 3, 10, and 11. Figure 3 presents the total computational time required for a typical isotropic grating analysis (Q = 20, v = 2, Bragg angle incidence, with the Q and v parameters as defined in Sec. 2.3) to achieve solutions accurate to within 0.1%. The total computation time is determined by multiplying the time per step (Fig. 10) by the total number of steps. The total number of steps is determined in turn by the numerical error tolerance specified for the computation, as shown in Fig. 11 and discussed below. The matrix eigenstate analysis is a one-step operation, so that the computation time shown in Fig. 10 for this one curve corresponds essentially to the total computation time. All other curves in Fig. 10 correspond to algorithms requiring multiple computational steps to approach the desired solution asymptotically.

The eigenstate analysis is most attractive for simple grating problems with low numbers of diffraction orders, but does not scale well with increasing problem complexity. If the number of diffraction orders considered in the analysis is N, then the computation time required for the eigenstate analysis is of order N^3, so that modest increases in the complexity of the problem introduce significant increases in computation time. The computation time for a densely packed coupling matrix (Fig. 9) proves to be comparable to the computation time for a sparsely packed matrix (Fig. 8) when using the eigenstate analysis.

The Runge-Kutta numerical integration methods scale more gracefully with increasing problem complexity, as shown in Fig. 10, but the scaling depends very much on how sparse the coupling matrix is (Figs. 8 and 9). For sparse matrices the computation time scales linearly with N, while for densely packed coupling matrices the computation time scales as N^2.

The computation time for the optical beam propagation method using the fast Fourier transform spectrum analysis algorithm scales approximately linearly with N, and the computation time is independent of the spectral richness of the permittivity perturbation.

The convergence properties of the Runge-Kutta numerical integration method and the optical beam propagation method are compared in Fig. 11 for a particular isotropic sinusoidal phase grating problem (Q = 20, v = 2, Bragg angle incidence, with the Q and v parameters as defined in Sec. 2.3). The convergence properties of the Runge-Kutta method prove to be quite sensitive to the formulation of the coupled wave equations. The results presented in Figs. 3 and 11 pertain specifically to the second set of coupled wave equations presented in Sec. 2.3. The first set of coupled wave equations presented in Sec. 2.3, which is the only set amenable to eigenstate analysis, gives significantly different convergence

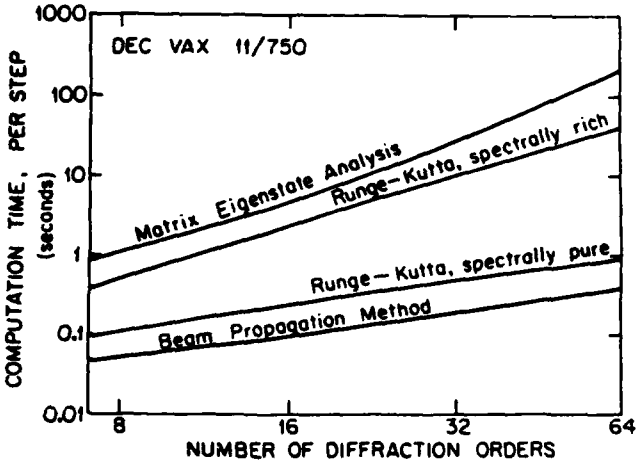


Fig. 10. Computation time per numerical step for several alternative algorithms, as performed on a Digital Equipment Corporation VAX 11/750 using FORTRAN. To determine the total computation time, shown in Fig. 3, multiply the computation time per step shown here by the number of steps needed to achieve a given error tolerance, as shown in Fig. 11. Note that the matrix eigenstate analysis is a one step process, so that for this case the computation time shown here is essentially the total computation time.

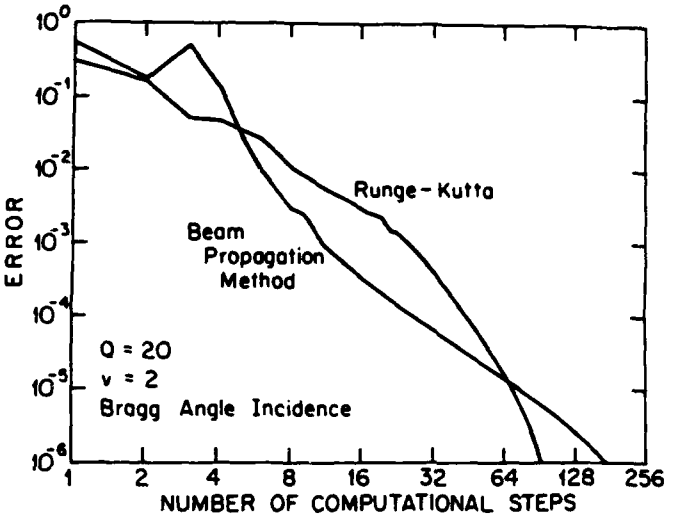


Fig. 11. Convergence of the Runge-Kutta and the optical beam propagation methods for a typical isotropic sinusoidal phase grating problem. To achieve better than a 0.1% error tolerance, about 25 steps are required by the Runge-Kutta routine, and 16 steps are required by the optical beam propagation method for this particular problem.

results, with better convergence for simple problems with N less than seven orders and remarkably poorer convergence for larger N.

Computationally, the difference in convergence behavior for the two alternative sets of equations is due to the numerical analysis concept of a stiff set of coupled equations.[48] Physically, the difference is due to the optical phase evolution required of the diffraction order amplitudes $\mathbf{E}_m(z)$ for the first set of coupled wave equations, but not for the second set. The optical phase evolution is most rapid for the highest diffraction orders, limiting the step size for the numerical integration, despite the fact that these highest orders receive a negligible fraction of the optical power.

Note that the results shown in Figs. 3, 10, and 11 are only intended to suggest general computational trends and should not be considered as absolute guidelines. Individual results depend heavily on which problem is being modeled, on which computer hardware is utilized, and on how efficiently the problem is formulated and encoded.

4. TYPICAL GRATING ANALYSES

Typical grating analyses that have been performed with the anisotropic optical beam propagation method are exhibited in this section. Isotropic gratings, in which the permittivity tensor is characterized by a scalar, are discussed in Sec. 4.1 [Fig. 6(A)], and birefringent phase gratings in the absence of optical activity are treated in Sec. 4.2 [Figs. 6(B) through 6(E)]. Finally, optical activity is added to the grating problem in Sec. 4.3.

4.1. Isotropic gratings

Diffraction from an isotropic single frequency sinusoidal phase grating forms a useful set of test cases for the optical beam propagation method, in particular because this problem has been well characterized by both analytic and approximate methods. Klein and Cook, for example, have presented numerous examples of numerical integrations of the coupled wave equations.[43] A partial list of the large number of devices exhibiting this grating behavior has been compiled by Gaylord and Moharam.[4]

The various sinusoidal phase grating problems are often scaled by the dimensionless parameters Q and v, which have been defined in Sec. 2.3. The Q parameter scales the thickness d of the diffraction grating, and the v parameter scales the strength of the grating. A thin grating (Raman-Nath regime) is often associated with $Q < 1$, while a thick grating (Bragg regime) is often associated with $Q > 10$. The v parameter corresponds to the optical phase shift induced in a light ray by the grating index perturbation. Further distinctions can be drawn between thick and thin gratings, and between Bragg and Raman-Nath regimes, in the case of strong grating modulation.[49]

Sample solutions of the sinusoidal phase grating problem are shown in Figs. 12 and 13. Figure 12 shows the fraction of the incident light beam power diffracted into the +1st diffraction order for a thick grating operating at or near the Bragg regime ($Q = 20$) as a function of the grating strength v up to a maximum optical phase shift of 5 rad. The convergence of the optical beam propagation method solutions to the desired asymptotic limit is evident for increasing numbers of steps. A similar figure for the Raman-Nath regime would show immediate convergence; i.e., only one step is required to obtain excellent correspondence with the asymptotic limit. The deeper the grating is into the Bragg regime, and the stronger the grating strength v, the more are the steps needed by the optical beam propagation method to reach reasonable convergence, as might be intuitively expected.

Figure 13 shows the angular alignment sensitivity of the isotropic grating. The diffraction efficiency for coupling into the +1st diffraction order is plotted as a function of the angle between the incident light wave vector and the grating wave vector. Peak coupling is observed for one particular angular alignment, corresponding to the Bragg angle. The grating in Fig. 13 is in the Bragg regime, with $Q = 20$ and $v = 2$. As in Fig. 12, several curves are shown, corresponding to successively closer approximations to the distributed grating prob-

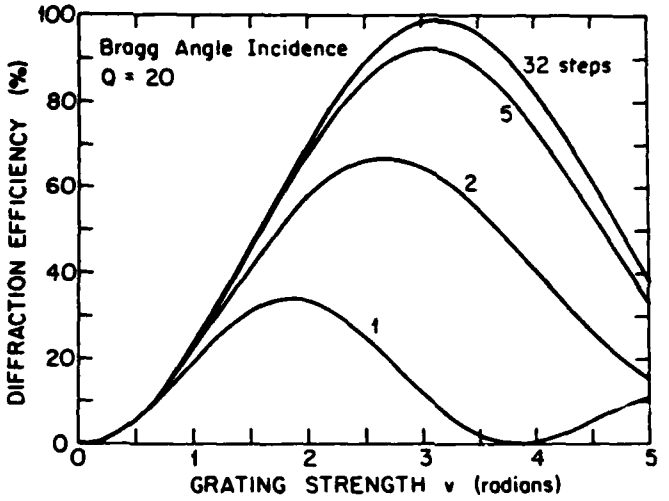


Fig. 12. Diffraction efficiency as a function of grating strength v (defined in Sec. 2.3) for a typical isotropic sinusoidal phase grating, showing the convergence characteristic of the optical beam propagation method for increasing numbers of computational steps.

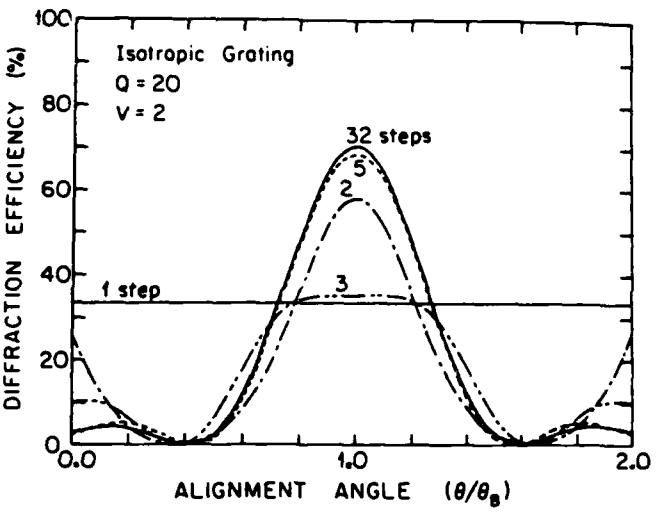


Fig. 13. Angular alignment sensitivity of a typical isotropic phase grating, showing the convergence of the optical beam propagation method to the desired solution for increasing numbers of computational steps.

lem. A one step approximation fails to account for any angular alignment sensitivity, since all of the modulation/polarization is collapsed to a single infinitesimally thin plane, and such a thin grating exhibits extremely broad angular alignment sensitivity. A two step approximation has roughly the correct profile within the range considered, deviating most noticeably in the tail region. A five step approximation gives a very good approximation within the range considered.

The asymptotic approach to the desired solution is shown in Fig. 4 for a Bragg regime thick grating (Q = 20, v = 2) with the incident light aligned at the optimum Bragg angle. The abscissa in Fig. 4 corresponds to the number of steps of the optical beam propagation method algorithm, spaced on a logarithmic scale, while the ordinate corresponds to the fraction of the incident light beam power coupled into the +1st diffraction order. As can be seen from Fig. 4, approximately six steps are required to converge to the desired solution for this particular combination of grating parameters. Figure 11 exhibits the convergence of the same grating problem, but with an ordinate corresponding to the logarithm of the error. In general, the thicker the grating (large Q) and the stronger the grating (large v), the more are the steps needed to approach asymptotically the desired solution within a given error tolerance.

4.2. Birefringent phase gratings

4.2.1. Gratings with fixed principal axes and varying birefringence

The gratings of Figs. 6(B) and 6(C) (birefringent phase gratings embedded in an isotropic medium) respond precisely like the isotropic phase grating of the previous section in terms of the dependence of the diffraction efficiency on the grating strength (Fig. 12) and on the angular alignment sensitivity (Fig. 13). These birefringent phase gratings are distinguished from the isotropic grating in that the polarization states of the odd diffraction orders are derived from the polarization state of the incident light beam as though the volume phase grating were a half-wave plate. If the incident light beam is linearly polarized at an angle ψ with respect to a principal axis of the grating, then the odd diffracted orders are also linearly polarized, but at an angle $-\psi$ with respect to the principal axes. The even diffraction orders of these gratings exactly reproduce the polarization state of the incident light beam. These results of the anisotropic optical beam propagation analysis confirm and extend similar results obtained analytically for thin birefringent phase gratings.[20]

The birefringent phase gratings of Figs. 6(B) and 6(C) are distinguished from each other by the orientation of the principal axes of the effective half-wave plate. The principal axes in Fig. 6(B) are parallel and perpendicular to the grating vector, whereas the principal axes for Fig. 6(C) are rotated 45° with respect to the grating wave vector.

The grating of Fig. 6(D) (fixed principal axes and varying birefringence embedded in a birefringent medium) exhibits much more intricate polarization behavior. The response is approximately that of an isotropic grating of Fig. 6(A), but the polarization states of the diffracted signal and transmitted probe beams are generally elliptical, even when the incident probe beam is linearly polarized. As the polarization of the incident probe beam is rotated by an angle ψ, the polarization of the diffracted signal beam rotates by approximately the same amount and in the same direction.

Computer studies of the polarization state of the diffracted signal beam, using the optical beam propagation method, show two polarization states of the incident probe beam that lead to strikingly simple polarization states of the diffracted signal beam for the grating of Fig. 6(D). These two input states are linear polarizations parallel and perpendicular to the grating wave vector. The resulting diffraction is isotropic; i.e., the diffracted signal beam exhibits precisely the same polarization state as the incident probe beam. These two configurations constitute polarization eigenmodes of the diffraction process, as presented by Petrov et al.[10]

A typical angular alignment sensitivity curve for a grating with fixed principal axes and varying birefringence closely approximates the isotropic response shown in Fig. 13, regardless of the polarization state of the incident probe beam. This is in marked contrast to the angular alignment sensitivity of the grating class discussed next.

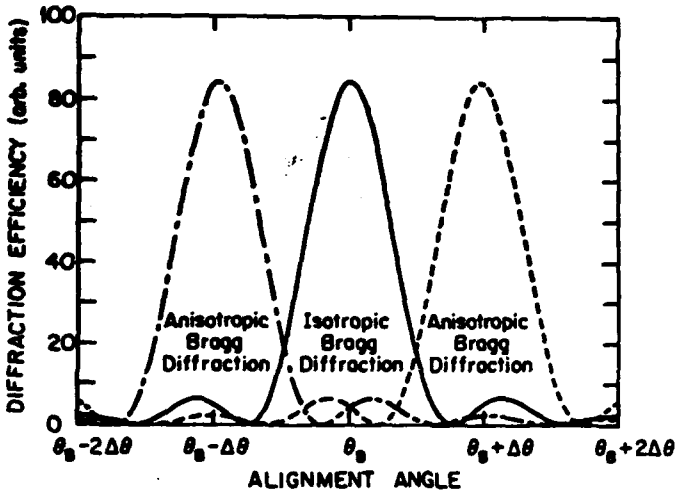


Fig. 14. Comparison of the angular alignment sensitivity of anisotropic Bragg diffraction [in a birefringent phase grating of the fixed birefringence/varying principal axes type, as shown in Fig. 6(E)] and isotropic Bragg diffraction [in a birefringent phase grating of the varying birefringence/fixed principal axes type, as shown in Fig. 6(D)].

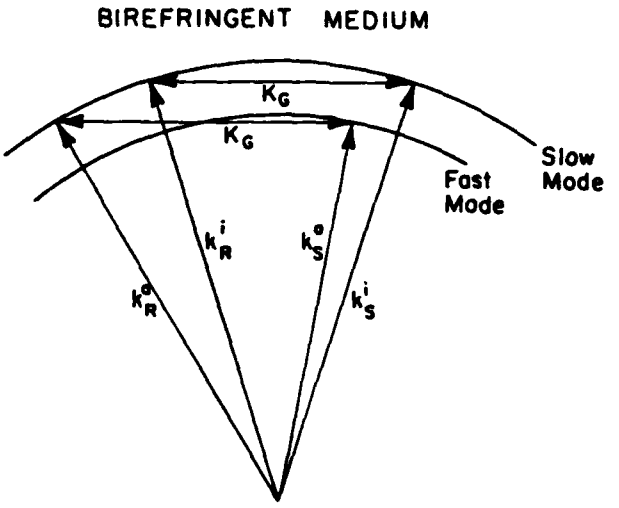


Fig. 15. Momentum phase-matching diagram for a birefringent phase grating in a birefringent medium, comparing isotropic Bragg diffraction (in which the polarization state of the diffracted signal beam is identical with the polarization state of the incident probe beam) with anisotropic Bragg diffraction (in which the polarization state of the diffracted signal beam is orthogonal to that of the incident probe beam). In the figure, K_G is the grating wave vector, k_R^i and k_S^i are the readout (probe) and signal beam wave vectors for the isotropic diffraction process, respectively, and k_R^o and k_S^o are the corresponding wave vectors for the anisotropic diffraction process.

4.2.2. Gratings with fixed birefringence and varying principal axes

The grating of Eig. 6(E) (fixed birefringence with varying principal axes) also exhibits intricate polarization behavior. The response is approximately that of the half-wave plate-like grating of Fig. 6(B), but the polarization states of the diffracted signal and transmitted probe beams are generally elliptical, even for a linearly polarized incident probe beam. As the polarization angle of the incident probe beam is rotated, the consequent polarization state of the diffracted signal light rotates with roughly a comparable angle, but in the opposite direction.

Two unique polarization eigenmodes can be identified with the grating of Fig. 6(E). Each polarization eigenmode exhibits anisotropic Bragg diffraction characteristics in which the polarization state of the diffracted light beam is orthogonal to the polarization state of the incident light beam when that state is linearly polarized *parallel* or *perpendicular* to the grating wave vector.

A grating with spatially varying axes is also distinguishable from a fixed axes grating in its angular alignment sensitivity, at least for sufficiently thick gratings. Compare the alignment sensitivity curves shown in Figs. 13 and 14. Each of the eigenmodes of the anisotropic Bragg diffraction process has associated with it an angular alignment sensitivity curve that is shifted in peak response angle from the comparable isotropic diffraction curve. This angular offset in peak response is proportional to the birefringence of the medium containing the grating. Since the grating media of Figs. 6(B) and 6(C) are isotropic (no birefringence), these birefringent phase gratings do not exhibit any resolvable fine structure or splitting of the angular alignment sensitivity curves. The angular resolution of this peak diffraction offset improves with increasing grating thickness d.

A momentum-matching diagram explaining the shift in peak response angle in the kinematic limit is shown in Fig. 15.[10] The birefringent medium has associated with it two wave vector surfaces, corresponding to the two possible eigenstates of polarization for propagation through the homogeneous medium in the absence of the grating perturbation. Isotropic diffraction processes couple light from a given wave vector surface to that same surface, whereas anisotropic diffraction couples the two neighboring wave vector surfaces. The separation between the two wave vector surfaces introduces a change in the optimum propagation angles (i.e., those giving rise to peak diffraction efficiency) for the anisotropic diffraction processes as compared with the isotropic diffraction processes.

The anisotropic Bragg diffraction process exhibits striking polarization and phase-matching characteristics that are crucial to the fundamental understanding and optimization of shear mode acoustooptic Bragg cells and variable grating mode liquid crystal devices. This anisotropic Bragg diffraction mechanism enables the relaxation of phase-matching constraints, and thereby extends the frequency bandwidth of acoustooptic Bragg cells.[14–16] The variable grating mode liquid crystal device exhibits remarkable angular alignment and thickness sensitivities of the various diffraction orders that can be fundamentally related to the physics of the molecular alignment distribution.[50]

4.3. Gratings with optical activity

The final class of polarization problem to be considered is that of a birefringent phase grating with concomitant natural optical activity. An example of such a grating might be a volume hologram in bismuth silicon oxide. The polarization properties of these gratings can be quite intricate, as suggested by Fig. 16, in which the evolution of the polarization states of the probe and diffracted signal beams is shown as a function of crystal thickness for each of the two principal orientations typically used for volume holography, for a typical input probe beam linearly polarized at 45° with respect to the grating wave vector. A uniform grating strength and negligible self-diffraction effects have been assumed in Fig. 16, so that

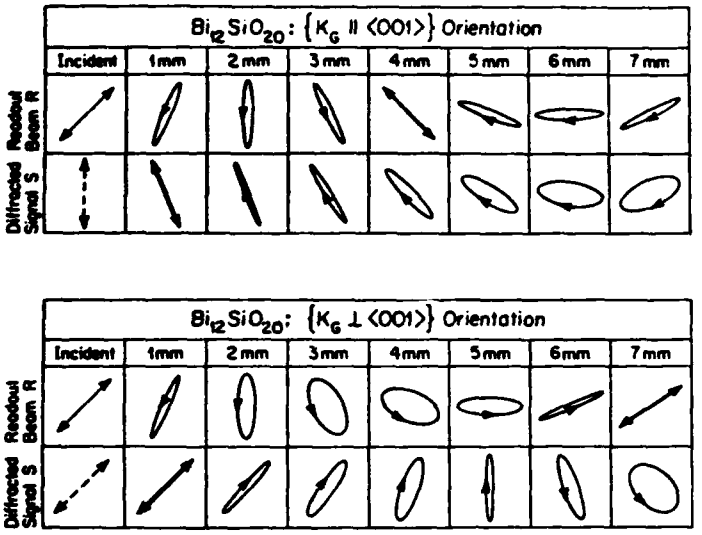


Fig. 16. Example of the rich polarization properties of a birefringent phase grating induced by the addition of natural optical activity. In this case, volume holograms in the two principal orientations of bismuth silicon oxide are assumed. The evolution of the polarization states within the crystal thickness of both the probe beam R and the diffracted signal beam S are shown for both principal crystallographic orientations.[8]

actual experimental results may deviate from these predictions. In the top sequence of Fig. 16, the probe and the diffracted beams start with a 45° angle between their polarization directions, but quickly evolve to a condition wherein they exhibit almost identical polarizations. In the bottom sequence, the probe and diffracted beams start with identical polarization states, but quickly evolve to a condition wherein they exhibit almost orthogonal polarization states. These results have significant implications for self-diffraction and for separating the diffracted signal light from the scattered probe light by polarization analysis techniques. Such results are not easily anticipated without computer analysis. A far more detailed discussion of the polarization properties of volume holograms in bismuth silicon oxide is given in Ref. 8. As shown therein, optical activity may be included in the optical beam propagation method with only a modest increase in the complexity of the Jones matrix analysis.

5. SUMMARY

Several classes of birefringent phase grating problems have been studied that exhibit striking polarization characteristics. The analysis is restricted to the limits of low grating frequencies and of small permittivity perturbations. In this limit an incident collimated laser beam remains reasonably well collimated, and backscattering of the incident light beam is negligible. Several alternative formalisms for solving the grating problem have been compared. The optical beam propagation method is shown to be particularly attractive. Intermediate mathematical steps are seen to have high intuitive content. The method exhibits exceptionally fast computational times and scales gracefully with increasing problem complexity, as is characteristic, for example, of the modulated spatial carrier problems encountered in spatial light modulator analyses. This desirable complexity scaling is independent of the spectral richness of the grating perturbation, unlike alternative analysis formalisms. Sample analyses of both isotropic and birefringent phase grating problems, corresponding to actual devices, have been presented.

6. ACKNOWLEDGMENTS

This research was supported in part by the Defense Advanced Research Projects Agency, the Air Force Systems Command (Rome Air Development Center), the Joint Services Electronics Program, and ITT Corporation. One of us (A. R. T.) gratefully acknowledges the hospitality extended by the California Institute of Technology and Newport Corporation during a sabbatical leave. One of the authors (R. V. J.) would like to thank David Hecht and David Yevick for introducing him to the optical beam propagation method and for many stimulating technical discussions. The authors gratefully acknowledge technical contributions on the modeling of volume holograms in bismuth silicon oxide by A. Marrakchi, and on the variable grating mode liquid crystal device by C. S. Wu.

7. REFERENCES

1. T. K. Gaylord and M. G. Moharam, "Analysis and applications of optical diffraction by gratings," Proc. IEEE 73(5), 894 (1985).
2. A. R. Tanguay, Jr., "The Czochralski growth and optical properties of bismuth silicon oxide," Ph.D. dissertation, Yale University (1977).
3. N. V. Kukhtarev, G. E. Dovgalenko, and V. N. Starkov, "Influence of the optical activity on hologram formation in photorefractive crystals," Appl. Phys. A 33, 227 (1984).
4. F. Vachss and L. Hesselink, "Holographic beam coupling in generally retarding media," J. Opt. Soc. Am. A 1(12), 1221 (1984).
5. M. P. Petrov, S. V. Miridonov, S. I. Stepanov, and V. V. Kulikov, "Light diffraction and nonlinear image processing in electrooptic $Bi_{12}SiO_{20}$ crystals," Opt. Commun. 31(3), 301 (1979).
6. J. P. Herriau, J. P. Huignard, and P. Aubourg, "Some polarization properties of volume holograms in $Bi_{12}SiO_{20}$ crystals and applications," Appl. Opt. 17(12), 1851 (1978).
7. A. Marrakchi and A. R. Tanguay, Jr., "Polarization properties of volume phase gratings in optically active materials," J. Opt. Soc. Am. A 1(12), 1313 (1984).
8. A. Marrakchi, R. V. Johnson, and A. R. Tanguay, Jr., "Polarization properties of photorefractive diffraction in electrooptic and optically active sillenite crystals (Bragg regime)," accepted for publication in J. Opt. Soc. Am. B, February 1986 (in press).
9. T. G. Pencheva, M. P. Petrov, and S. I. Stepanov, "Selective properties of volume phase holograms," Opt. Commun. 40(3), 175 (1982).
10. M. P. Petrov, T. G. Pencheva, and S. I. Stepanov, "Light diffraction from volume phase holograms in electrooptic photorefractive crystals," J. Optics (Paris) 12(5), 287 (1981).
11. M. P. Petrov, S. I. Stepanov, and A. A. Kamshilin, "Image polarization characteristics storage in birefringent crystals," Opt. Commun. 21(2), 297 (1977).
12. M. P. Petrov, S. I. Stepanov, and A. A. Kamshilin, "Light diffraction from the volume holograms in electrooptic birefringent crystals," Opt. Commun. 29(1), 44 (1979).
13. M. P. Petrov, S. I. Stepanov, and A. A. Kamshilin, "Light diffraction in photorefractive ferroelectrics," Ferroelectrics 21, 631 (1978).
14. R. W. Dixon, "Acoustic diffraction of light in anisotropic media," IEEE J. Quantum Electron. QE-3(2), 85 (1967).
15. N. J. Berg and J. N. Lee, eds., *Acousto-Optic Signal Processing*, Marcel Dekker, Inc., New York (1983).
16. M. Gottlieb, C. L. M. Ireland, and J. M. Ley, *Electro-Optic and Acousto-Optic Scanning and Deflection*, Sec. 6.4, Marcel Dekker, Inc., New York (1983).
17. R. C. Jones, "A new calculus for the treatment of optical systems I. Description and discussion of the calculus," J. Opt. Soc. Am. 31, 488 (1941); R. C. Jones, "A new calculus for the treatment of optical systems II. Proof of three general equivalence theorems," J. Opt. Soc. Am. 31, 493 (1941); R. C. Jones, "A new calculus for the treatment of optical systems III. The Sohncke theory of optical activity," J. Opt. Soc. Am. 31, 500 (1941); R. C. Jones, "A new calculus for the treatment of optical systems IV," J. Opt. Soc. Am. 32, 486 (1942); R. C. Jones, "A new calculus for the treatment of optical systems V. A more general formulation, and description of another calculus," J. Opt. Soc Am. 37, 107 (1947); R. C. Jones, "A new calculus for the treatment of optical systems VI. Experimental determination of the matrix," J. Opt. Soc. Am. 37, 110 (1947); R. C. Jones, "A new calculus for the treatment of optical systems VII. Properties of the N-matrices," J. Opt. Soc. Am. 38, 671

Physical and Technological Limitations of Optical Information Processing and Computing

Armand R. Tanguay, Jr.

Introduction

Over the past four decades, the growth of information processing and computational capacity has been truly remarkable, paced to a large extent by equally remarkable progress in the integration and ultra-miniaturization of semiconductor devices. And yet it is becoming increasingly apparent that currently envisioned electronic processors and computers are rapidly approaching technological barriers that delimit processing speed, computational sophistication, and throughput per unit dissipated power. This realization has in turn led to intensive efforts to circumvent such bottlenecks through appropriate advances in processor architecture, multiprocessor distributed tasking, and software-defined algorithms.

An alternative strategy that may yield significant computational enhancements for certain broad classes of problems involves the utilization of multidimensional optical components capable of modulating and/or redirecting information-carrying light wavefronts. Such an optical processing or computing approach relies for its competitive advantage principally on massive parallelism in conjunction with relative ease of implementation of complex (weighted) interconnections among many (perhaps simple) processing elements. A wide range of computational problems exist that lend themselves quite naturally to optical processing architectures, including pattern recognition, earth resources data acquisition and analysis, texture discrimination, synthetic aperture radar (SAR) image formation, radar ambiguity function generation, spread spectrum identification and analysis, systolic array processing, phased array beam steering, and artificial (robotic) vision. In addition, many neural network processes that inherently rely on intricate interconnection patterns have been or can be implemented optically. These and other applications are treated in more detail in accompanying articles in this special issue of the MRS Bulletin[1,2] and in special issues of *IEEE Proceedings*[3] and *Optical Engineering*[4] on optical computing.

A generalized optical processor or computer can be depicted schematically as shown in Figure 1. The physical constituent elements of such a system include a central processing unit (CPU) that performs the essential implementable function, a data management processor that orchestrates the flow of data and sequence of operations (usually considered part of the CPU in a traditional electronic computer), several types of memory elements for both short-term and long-term data storage and buffering, format devices to spatially organize input data fields, input devices to convert data input types to a form amenable to subsequent processing, output devices to convert processed results to detectable and interpretable forms, and detectors to produce externally addressable results. In Figure 1, feedback interconnects are explicitly shown as separate components to emphasize their crucial role in implementing parallel iterative algorithms and complex weighting functions.

A wide variety of optical components are required to implement processors based on the generalized architecture shown in Figure 1.[5] These include one- and two-dimensional spatial light modulators, volume holographic optical elements, threshold arrays, optical memory elements, sources, source arrays, detectors, and detector arrays. The state of the technology is such that while demonstration devices and prototypes are proliferating, with the exception of sources, detectors and detector arrays, few (if any) such components have as yet achieved significant commercial success or even demonstrated technological viability. In large part, this is due to the fact that each of the candidate technologies has placed rather severe demands on the state-of-the-art of the materials employed regardless of the nature of the optical effect utilized (e.g., electrooptic, magnetooptic, photorefractive, or electroabsorptive). In other words, the magnitudes of observable optical perturbations per unit excitation are just not large enough with readily available materials to allow flexible device engineering. As such, the answer to whether optical processing and computing will come of age may ultimately rest on the capabilities and fortunate discoveries of materials scientists and process engineers.

It is of considerable interest, nonetheless, to examine the physical as well as technological limitations that apply to optical information processing and computing systems in order to assess their potential performance advantages (if any) over comparable electronic counterparts, and to provide much-needed guidance for continued research efforts in optical materials, devices, algorithms, and system architectures. Although the study of fundamental physical limitations in the context of digital (binary) electronic systems (particularly VLSI) is well established,[6,7] it should be noted that comparable studies of optical processing and computing systems are rela-

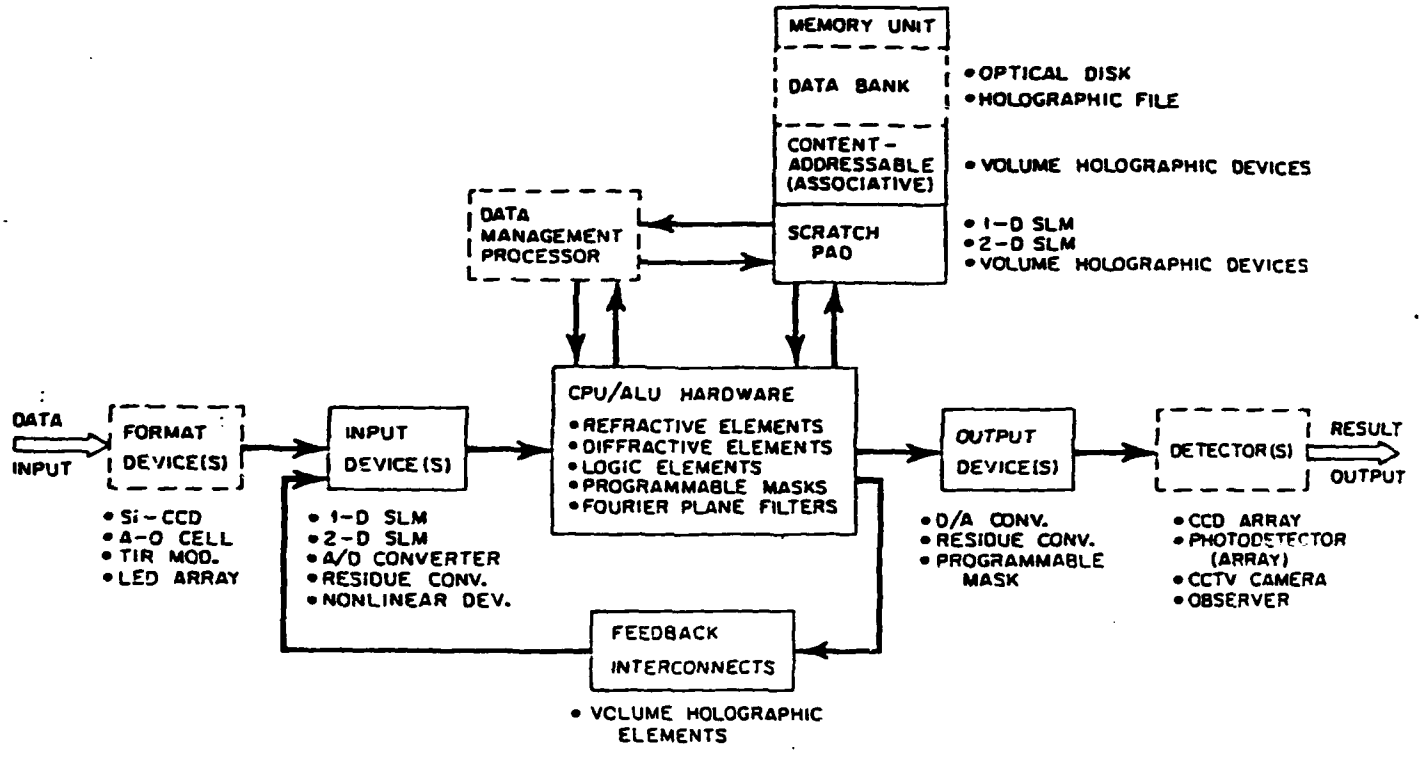


Figure 1. Schematic diagram of the elements of a generalized optical processor or computer (after Ref. 5).

tively recent, and have to date focused primarily on digital optical computing as opposed to analog optical processing.[8-11]

The remainder of this article will describe several such physical and technological limitations of both optical information processing and computing. The following section addresses the nature of computation from the perspective of identifying physical (*technology-independent*) limits. The next section will then provide an example of a *technology-dependent* limit in the context of photorefractive volume holographic optical interconnections. Conclusions are drawn in a final section.

The Nature of Computational Constraints

In order to implement any computational algorithm on a machine of given architecture and component hardware, we must first choose a *representation* for inputting data and executing computational steps. Such a representation might, for example, be digital, analog, or even symbolic. As we shall see, such a choice of representation has profound implications on the physical limits that apply to subsequent computation.

For a given representation, we can then assess the *computational complexity* of the chosen algorithms. We define the computational complexity of an operation as the equivalent number of irreducible binary switching operations required to perform the same operation in the most efficient manner possible. For example, the relatively simple operation of transferring 1,000 ten-bit words from the CPU to memory will require at the very least 10^4 binary switching operations (assuming that all of the data is transferred in parallel), and most likely considerably more if shift registers are employed to multiplex the data transfer. This definition of computational complexity provides a convenient means for comparing the overall efficiency or total minimum energy cost of a given computation performed by different algorithms on a machine of given architecture and technology, or by machines based on vastly different architectures and perhaps disparate technological hardware. It does not, however, take into account the interconnection complexity required to implement communications pathways at the device, circuit, and subsystem levels.

Finally, we must end each computation with a *detection* of the desired results, in order to allow the result to be used (for example, to implement a desired action). This separation of every computational process into the implementation of representation, computational complexity, and detection functions allows us to establish sets of interrelated limits that apply to various combinations of choices.

For purposes of discussion in this article, let us examine each of these functions from the point of view of energy cost. In so doing, we seek to identify physical limits on the maximum computational throughput achievable per watt of dissipated power. It should perhaps be noted here that the "requirement" of energy dissipation is not in fact a fundamental limit, and derives instead from a system design demand to *assert* each stored value as rapidly as possible in order to complete the entire computation deterministically and in minimum time. Thus adiabatic computational processes which do not require a minimum energy dissipation[12] are not applicable to the present discussion.

In the digital (binary) realm, the lowest possible energy cost of representing

a number requiring a given number of bits is equal to the number of bits times the minimum energy to store a single bit. For semiconductor electronic switches, the minimum energy in turn is equal to a few tens of k_BT per electron[6,7] times the number of electrons required to guarantee detection with a given bit error rate (BER) [a quantum rather than a thermal limit]. For a bit error rate of 10^{-9} or so, about ten electrons are required (assuming an "ideal" detector [or following switch] that can unambiguously differentiate between the presence and absence of a *single* electron). Therefore each switching event (or representation of each bit) requires the dissipation of approximately 200 k_BT. This places an immediate upper boundary on the maximum computational throughput of such an electronic digital (binary) computing engine of approximately 10^{18} transitions (irreducible binary switching operations) per second per watt at room temperature. Note that this bound does not include the assessment of any cost for internal communication of information, as would be required to execute an actual computation.

To place this number in proper perspective, we need only look at the switching energies representative of current semiconductor technology. For electronic devices, we might examine a typical CMOS capacitor with a 100 μm^2 cross-sectional area and a 1,000 Å oxide thickness, for which the charge/discharge cycle consumes CV^2 worth of energy. If we operate at the minimum switching voltage of a few tens of k_BT/q,[6,7] the switching energy is about 10^6 k_BT at room temperature, about four orders of magnitude above the fundamental limit. Current digital logic circuits operate at roughly 2.5×10^{-12} J/transition, which is about 10^9 k_BT.[7] This is also roughly the minimum energy required to operate even a local interconnection at GHz rates,[7,8] an unavoidable cost of communication between devices at the circuit level. At the system level, a vast amount of additional overhead comes into play. For example, the DEC VAX 11-750 dissipates approximately 3 kW running fully loaded at its maximum throughput rate of 750,000 instructions/second (200,000 operations/second). Thus the energy required to perform a single instruction is about 3 mJ or 10^{19} k_BT, which corresponds to a throughput rate of 250 instructions/second/watt.

By comparison, analog representations (as used extensively, for example, in optical processors) require far more energy than the binary equivalent. This is due to the necessity of utilizing a much higher particle count (electrons or photons) in order to minimize the effects of quantum statistical fluctuations on the BER. For example, if we wish an analog representation of the number 1,000, then we require a dynamic range of at least 1,000:1. For incoherent illumination, quantum fluctuations in the emission/detection process produce a photon number distribution with a relative standard deviation of $\sigma \cong \sqrt{N}/N$. The equivalent of a 10^{-9} BER for the digital case corresponds to roughly 12 standard deviations. Therefore, the number of photons required must be greater than 1.5×10^8 from statistical considerations alone. For a GaAs semiconductor laser characterized by a photon energy of ~1.5 eV, this corresponds to about 10^{10} k_BT. To represent 1,000 optically in binary requires approximately 14 bits (10 bits for the number plus 4 bits of overhead) at 15 eV each (10 photons at 1.5 eV each, assuming direct detection and an ideal detector), or about 10^4 k_BT.

Given the remarkably higher cost of analog representation as compared with digital, any competitive advantages for analog systems from the perspective of an energy dissipation metric due to physical limits must come from enhanced computational complexity. The various tradeoffs involved can perhaps best be described in terms of an example. Consider, then, a highly interconnected, nonlocal problem such as the Fourier transformation of a two-dimensional function. Assume that the function is sampled on a 1,000 × 1,000 element array at 10 bits.

If we operate in the binary regime, we can utilize a highly efficient discrete Fourier transform (DFT) routine such as the Cooley-Tukey algorithm. The number of complex operations (multiplies and adds) scales as $1.5N^2\log_2 N^2$ for an $N \times N$ input. If we assume an electronic machine implementation that requires 20 switching events/bit/complex operation, then approximately 10^{10} switching events are required to perform the computation, or 10 nJ for operation near the physical limits described above.

If, on the other hand, the input data field (two-dimensional function) is represented in analog form by means of a spatial light modulator, illumination by a coherent wavefront will produce the required transform in the back focal plane of a (following) lens. Here we see illustrated the notion of a computation as the *transformation* of information. The energy cost of this particular operation accrues only to the initial representation of the function (essentially, the optical or electronic addressing of the spatial light modulator) and to the subsequent detection of the final result. This is then equivalent to 2×10^6 individual pixel detection operations for the optically addressed case, or 70 μJ at the physical limits for 1.5 eV photons, still about four orders of magnitude larger than the binary equivalent. The computational complexity implemented by the analog optical processor is thus seen to partially offset the large initial difference in representation energy costs. For other classes of problems with even higher inherent computational complexity, the physical boundary on energy cost may thus favor an analog approach.

The situation looks quite a bit different if we examine current *technological* (rather than more fundamental) constraints. For current digital circuits that switch at around 2.5 pJ/bit, the DFT just described dissipates 25 mJ for the required number of switching operations alone. Current digital systems such as the VAX 11-750 consume about 10^{-4} J/bit, or 1 MJ for the DFT. With regard to optical systems, spatial light modulators are available with input sensitivities of about 200 pJ/pixel at high signal-to-noise ratio, and CCD detector arrays dissipate approximately 1 mJ/Mpixel on readout. Hence the analog optical Fourier transform can be performed for an energy cost of about 1.2 mJ. This apparent capability of analog optical systems to generate significantly enhanced computational throughput per unit input power is thus due largely to the fact that currently available analog optical components operate far closer to the relevant thermal and quantum limits than current digital electronic (VLSI) components. It should be noted that such a comparison begs the question of *overall* computational accuracy, which clearly favors the digital implementations due to nonlinearities and nonuniformities in currently available analog optical components.

The third principal component of the computational process, that of detection of the results (whether electronic or optical), is subject to the same thermal and quantum statistical limitations as the process of representation. This statement follows directly from an assumption inherent in assessing the minimum energy required for representation — that each number must be *capable* of detection at a given BER,

assuming an ideal detector. Technological constraints can then be added to the fundamental limit problem as before to assess realistic current capabilities. For example, available optical detectors typically require a mean photon number of 1,000 for a 10^{-9} BER in the direct detection mode, rather than the value of 10 assumed above.

Volume Holographic Optical Interconnections

As pointed out in the Introduction (and discussed in detail in Ref. 1), the ability to provide complex, multidimensional, programmable, weighted interconnections by means of volume holograms is an attractive feature of optical processing and computing systems. In addition, this ability perhaps more so than any other contributes significant computational complexity to optical architectures. To this end, it is of considerable importance to assess the fundamental limitations that apply to such volume holographic optical interconnections.

Although many types of photosensitive media can be utilized for the recording and readout of volume holograms, the dynamic reprogrammability offered by photorefractive crystals such as lithium niobate, barium titanate, bismuth silicon oxide, and gallium arsenide has made such materials the objects of intensive study. In these materials, the interference between two coherent optical wavefronts (the signal and reference beams) generates a space-variant photoexcitation distribution that produces in turn a related space charge redistribution and associated electric field pattern by carrier drift and diffusion. Many interesting limits apply to such a process, including the maximum number of independent interconnections that can be sequentially or simultaneously recorded at a given value of allowable crosstalk, the highest achievable diffraction efficiency of each independent interconnection at the maximum interconnection density, the absolute minimum number of photoevents per unit volume required to record an interconnection of given analog weight within the quantum fluctuation limits for statistical accuracy of recording and reconstruction (readout), and the maximum asymmetry possible between the recording and erasure processes.[15] An additional limit of considerable importance is that of the photorefractive sensitivity,[16-18] or the refractive index modulation obtained in recording a uniform grating of fixed spatial frequency per unit (incident or absorbed) energy density, as this parameter places an upper bound on the maximum rate of interconnection reprogrammings that can be accomplished per unit average power. In what follows, the fundamental physical limitations on the photorefractive sensitivity are examined in a bit more detail.

A number of factors contribute to the photorefractive sensitivities characteristic of photoconductive, electrooptic materials. One such factor is the photogeneration quantum efficiency, which represents the number of photogenerated mobile charge carriers per photon absorbed from the recording beam(s). A second factor is the charge transport efficiency, which is a measure of the degree to which the average photogenerated mobile charge carrier contributes to the forming space charge grating after separation from its original site by means of drift and/or diffusion and subsequent trapping. The magnitude of the space charge field generated by a given space charge grating is inversely proportional to the dielectric permittivity ε of the photorefractive material, which thus contributes a third factor to the grating recording sensitivity. A fourth factor describes the perturbation of the local index ellipsoid (dielectric tensor at optical frequencies) that results from a given space charge field through the electrooptic frequencies) that results given space charge field through the electrooptic (Pockels or Kerr) effect.

In addition, several other physical quantities factor into an evaluation of the photorefractive sensitivity, including the wavelength of the recording illumination (to convert the number of absorbed photons into an equivalent energy), the absorption coefficients of the material at the wavelengths of both the recording and readout beams (to correct for the fractional absorbance of the recording beams and the fractional transmittance of the readout beam), and the magnitude of the applied voltage (which significantly alters the sensitivity characteristics for certain materials by changing the nature of the dominant charge transport mechanism from the diffusion regime to the drift regime).

In order to provide a quantitative metric that does in fact have a fundamental physical limitation, we define the *grating recording efficiency*[19] of a photorefractive recording model or configuration as the magnitude of the space charge field produced by a fixed number of photogenerated mobile charge carriers at a given spatial frequency, normalized by the maximum quantum limited space charge field that can be produced by the same number of photogenerated carriers. This metric thus effectively combines the notions of a photogeneration efficiency and a charge transport efficiency, and provides an estimate of the fundamental quantum efficiency of the photorefractive grating recording process. For simplicity, we confine our attention here to a photorefractive model characterized by a single mobile charge species, and a single type of donor site with associated un-ionized donor and ionized donor (trap) states.

Let us compare the grating recording efficiencies of four idealized grating recording models, generated by combining two types of photogeneration profiles (a comb function and a sinusoid) with two types of charge transport processes (a translation of each photogenerated charge carrier by exactly half of a grating wavelength; and uniform redistribution, or randomization, of all photogenerated charge). The resulting four combinations are shown schematically in Figure 2.

The maximum quantum limited space charge field that can result from a fixed number of photogenerated mobile charges is given by the bipolar comb distribution, which generates a square wave electric field profile from alternating positive and negative sheets of charge spaced by half of a grating wavelength. It is in fact the first harmonic component of this electric field profile that we are interested in, as we assume a volume grating operated deep within the Bragg diffraction regime. The bipolar comb distribution can thus be assigned a grating recording efficiency of unity.

By comparison, the monopolar comb, transport-efficient sinusoid, and baseline sinusoid cases yield grating recording efficiencies of 1/2, 1/2, and 1/4, respectively. These results are summarized in Figure 3, which shows the evolution of the first harmonic component of the space charge field for all four cases as a function of grating recording time.

The saturation behavior of each case derives from assuming a fixed finite trap density prior to grating recording. Since the resultant diffraction efficiencies of such gratings scale as the square of the space charge field (for small modulation), these results imply that the baseline sinusoid distribution will exhibit a diffraction efficiency lower than the bipolar comb distribution by a factor of 16. This conclusion is of considerable

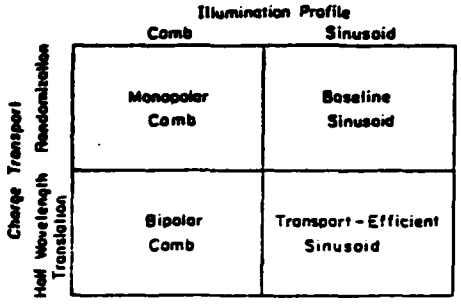


Figure 2. Schematic diagram depicting the illumination profile and charge transport assumptions underlying each of the four idealized photorefractive recording models discussed in the text.

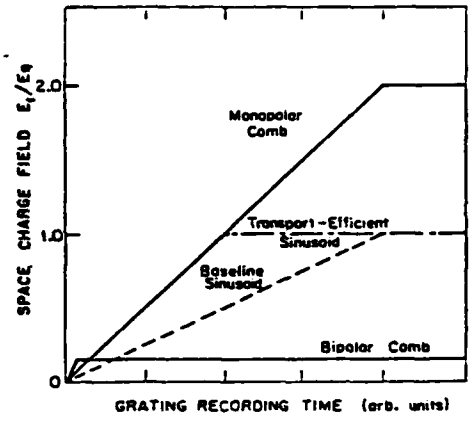


Figure 3. The evolution of the space charge field as a function of grating recording time for each of the four idealized photorefractive recording models. The normalization parameter $E_q = eN_A/\varepsilon\varepsilon_0 K_C$ in which N_A is the maximum available trap density and K_C is the grating wavevector.

interest because the baseline sinusoid case can be shown[19] to represent an asymptotic upper bound on the grating recording efficiency predicted by more realistic (rather than idealized) photorefractive grating recording models that assume a sinusoidal illumination profile. In fact, several photorefractive materials have been demonstrated to exhibit photorefractive sensitivities that approach the upper bound represented by the baseline sinusoid case.[18]

Nonetheless, the photorefractive grating recording process is from this perspective somewhat quantum inefficient, in that it does not make the best possible use of either input photons or photogenerated carriers. In view of the analysis presented above, therefore, it is of considerable interest to imagine novel engineered photorefractive materials with donor planes spatially separated from trap planes, illuminated by a comb-like intensity pattern (which can be generated, for example, by utilizing stratified volume holographic optical elements [SVHOEs][20]). Such materials could find applications in devices based on generation of a carrier grating of given spatial frequency, such as the Photorefractive Incoherent-to-Coherent Optical Converter (PICOC).[21]

Furthermore, it should be noted that the limits derived for interconnections based on photorefractive materials pertain to a particular physical mechanism involving photoexcitation, charge transport, and electrooptically induced index changes. The importance of multiplexed interconnections to overall schemes for optical processing and computing suggests continued emphasis on the search for alternative physical effects that can provide either enhanced sensitivity or more conveniently implementable desirable features such as selective erasure.[22]

Conclusions

An examination of both the fundamental physical and current technological limitations to computational performance shows that in terms of throughput per unit power, digital (binary) representations (whether electronic or optical) presently operate many orders of magnitude away from the relevant boundaries. Analog representations and detections are inherently much more power consumptive than digital representations and detections, and processing schemes based on analog algorithms must demonstrate considerably enhanced computational complexity prior to each intermediate detection plane in order to be energy competitive at the appropriate thermal and quantum limits.

Hybrid architectures and algorithms (such as those employed by numerous neural network models) that effectively combine analog representations as weights in highly multiplexed parallel interconnections with binary representations in the form of threshold arrays (decision planes) may ultimately prove to be a nearly optimum compromise. This is particularly applicable to optical processing and computing architectures, for which available analog components (spatial light modulators and photorefractive volume holographic optical elements) operate much closer to the relevant quantum statistical limits than the corresponding digital components (threshold arrays). In such hybrid architectures, the threshold array performs the equivalent of a parallel level restore function (saturated nonlinearity) that is essential to the proper convergence of many processing and computing algorithms.

Further study of the fundamental physical limitations that apply to optical processing and computing will provide important guidance for the continued development of active optical materials, primarily by differentiating between avenues of opportunity with large potential gain that depend on key materials parameters, and those for which current materials characteristics are sufficient to approach relevant performance boundaries.

Acknowledgments

Research on the fundamental limitations of optical information processing and computing at the University of Southern California is supported in part by the Air Force Office of Scientific Research, the Defense Advanced Research Projects Agency, and the Army Research Office.

References

1. D.Z. Anderson, *MRS Bull.* (this issue).
2. A.M. Glass, *MRS Bull.* (this issue).
3. J.A. Neff, ed., *Opt. Eng.*, Special Issue on Optical Computing 24 (1) (1985).
4. H.J. Caulfield, S. Horvitz, G.P. Tricoles, and W.A. Von Winkle, eds., *Proc. IEEE*, Special Issue on Optical Computing 72 (7) (1984).
5. A.R. Tanguay Jr., *Opt. Eng.* 24 (1) (1985) p. 2-18.
6. C. Mead and L. Conway, *Introduction to VLSI Systems*, (Addison-Wesley, Reading, MA, 1980) p. 333-372.
7. R.W. Keyes, *The Physics of VLSI Systems*, (Addison-Wesley, Reading, MA, 1987).
8. R.W. Keyes, *Proc. IEEE*, 63 (5) (1985) p. 740-767.
9. P.W. Smith, *Bell Systems Tech. J.* 61 (1982) p. 1975-1993.
10. R.L. Fork, *Phys. Rev. A.* 26 (4) (1982) p. 2049-2064.
11. R.W. Keyes, *Opt. Acta* 32 (5) (1985) p. 525-535.
12. R. Landauer, in *Optical Information Processing*, edited by Yu. E. Nesterikhin, G.W. Stroke, and W.E. Kock (Plenum Press, New York, 1976) p. 219-253.
13. C.H. Bennett, *Int. J. Theor. Phys.* 21 (12) (1982) p. 905-941.
14. C. Kyriakakis, P. Asthana, R.V. Johnson, and A.R. Tanguay Jr., *Proc. Opt. Soc. Am. Topical Meeting on Spatial Light Modulators*,

Lake Tahoe, Nevada, (1988).
15. See Reference 1 for a discussion of the usefulness of this asymmetry.
16. A.M. Glass and D. von der Linde, *Ferroelectrics* 10 (1976) p. 163-166.
17. F. Micheron, *Ferroelectrics* 18 (1978) p. 153-159.
18. A.M. Glass, *Opt. Eng.* 17 (5) (1978) p. 470-479.
19. R.V. Johnson and A.R. Tanguay Jr., in *Optical Processing and Computing*, edited by H. Arsenault and T. Szoplik (Academic Press, New York, 1988), (in press).
20. R.V. Johnson and A.R. Tanguay Jr., *Opt. Lett.* 13 (3) (1988) p. 189-191.
21. A. Marrakchi, A.R. Tanguay Jr., J. Yu, and D. Psaltis, *Opt. Eng.* 24 (1) (1985) p. 124-131.
22. D. von der Linde and A.M. Glass, *Appl. Phys.* 8 (1975) p. 85-100. □

APPENDIX 6

Fundamental Physical Limitations of the Photorefractive Grating Recording Sensitivity

R. V. Johnson and A. R. Tanguay, Jr.

Optical Materials and Devices Laboratory,
and Center for Photonic Technology
Departments of Electrical Engineering
and Materials Science
University of Southern California
University Park, MC-0483
Los Angeles, California 90089-0483

Abstract

A number of inherent inefficiencies exist in the photorefractive recording of grating structures due to the nature of the photoexcitation and charge transport processes. These inefficiencies are identified herein by postulating highly idealized photoexcitation and charge transport models that yield optimum (quantum limited) space charge field distributions, assuming a photoexcitation constraint of no more than one mobile charge per incident photon. By comparing such highly idealized photorefractive recording models with more realistic models, the fundamental origins of several such inherent inefficiency factors are identified, and their magnitudes estimated. In this manner, the grating recording efficiencies of photorefractive materials can be directly compared with the fundamental physical limitations imposed by quantum constraints.

Fundamental Physical Limitations of the Photorefractive Grating Recording Sensitivity

R. V. Johnson and A. R. Tanguay, Jr.

University of Southern California
Optical Materials and Devices Laboratory,
and Center for Photonic Technology
Departments of Electrical Engineering
and Materials Science
University Park, MC-0483
Los Angeles, California 90089-0483

I. Introduction

Of all of the parameters that seek to quantify the absolute or relative performance of photorefractive materials, one of the most important is the photorefractive sensitivity (von der Linde and Glass, 1975; Micheron, 1978; Glass, 1978; Gunter, 1982; Yeh, 1987a and 1987b; Glass *et al.*, 1987; Valley and Klein, 1983). This key parameter is typically defined in theoretical analyses either as the refractive index modulation obtained in writing a uniform grating of fixed spatial frequency per unit absorbed recording energy density (energy per unit volume) (von der Linde and Glass, 1975; Micheron, 1978; Glass, 1978; Gunter, 1982; Glass et al., 1987; Valley and Klein, 1983), or as the inverse of the recording energy density required to achieve a specified value of the diffraction efficiency for a uniform grating of fixed spatial frequency in a material of given thickness (Valley and Klein, 1983; Huignard and Micheron, 1976). Alternatively, for purposes of experimental measurement, the photorefractive sensitivity may be specified as the inverse of the recording energy density required to reach a given fraction of the saturation diffraction efficiency of a particular material (Gunter,

1982; Amodei and Staebler, 1972). The photorefractive sensitivity, and the fundamental physical limitations that apply to it, are of current significant interest because they establish the maximum reconfiguration rate of volume holographic optical elements (VHOEs) (von der Linde and Glass, 1975; Tanguay, 1985) at constant average optical input power. The maximum reconfiguration rate of VHOEs is in turn important for applications ranging from massively parallel interconnections in optical processing and computing systems (Tanguay, 1985) to the development of the photorefractive incoherent-to-coherent optical converter (PICOC) (Kamshilin and Petrov, 1980; Shi et al., 1983; Marrakchi et al., 1985).

A number of factors contribute to the photorefractive sensitivity characteristic of photoconductive, electrooptic materials. One such factor is the quantum efficiency, which represents the number of photogenerated mobile charge carriers per photon absorbed from the recording beam(s). A second factor is the charge transport efficiency, which is a measure of the degree to which the average photogenerated mobile charge carrier contributes to the forming space charge grating after separation from its original site by means of drift and/or diffusion and subsequent trapping. The magnitude of the space charge field generated by a given space charge grating is inversely proportional to the dielectric permittivity ϵ of the photorefractive material, and as such contributes a third factor to the grating recording sensitivity. A fourth factor describes the perturbation of the local index ellipsoid (dielectric tensor at optical frequencies) that results from a given space charge field through the electrooptic (Pockels or Kerr) effect. And finally a fifth factor pertains to the physical optics readout process, whereby the diffraction efficiency and polarization properties of the readout beam are derived directly from the index ellipsoid modulation.

In addition, several other physical quantities factor into an evaluation of the photorefractive sensitivity, including the wavelength of the recording illumination (to

convert the number of absorbed photons into an equivalent energy), the absorption coefficients of the material at the wavelengths of both the recording and readout beams (to correct for the fractional absorbance of the recording beams and the fractional transmittance of the readout beam), and the magnitude of the applied voltage (which significantly alters the sensitivity characteristics for certain materials by changing the nature of the dominant charge transport mechanism from the diffusion regime to the drift regime).

In previous treatments of the photorefractive sensitivity and its associated limits, all of the above mentioned factors have been addressed to some degree, though not necessarily within a single unified treatment, or even within a consistent set of constraints, assumptions, and approximations. Such a unification is also beyond the scope of the present work. In this paper, we present the results of a study in which we have approached the photorefractive sensitivity issue from a somewhat different perspective: that of the absolute *quantum efficiency* of the photorefractive grating recording process. As such, we attempt to answer an oft-stated but as yet unanswered question as to the origin of the apparent relative insensitivity of photorefractive grating recording.

In order to provide a quantitative metric that does in fact have a fundamental physical limitation, we define herein the *grating recording efficiency* of a photorefractive recording model or configuration as the magnitude of the space charge field produced by a fixed number of photogenerated mobile charge carriers at a given spatial frequency, normalized by the maximum quantum limited space charge field that can be produced by the same number of photogenerated carriers. This metric thus effectively combines the notions of a photogeneration efficiency and a charge transport efficiency, and provides an estimate of the fundamental quantum efficiency of the photorefractive grating recording process.

In order to fully utilize the concept of the grating recording efficiency, we first calculate the maximum quantum limited space charge field that can be produced by a fixed number of photogenerated mobile charge carriers, assuming optimum photoexcitation and redistribution (transport) functions. We then calculate the grating recording efficiencies that describe several important limiting cases with selected idealized photoexcitation and redistribution functions. This in turn allows for the definition of a baseline case against which more realistic charge transport models can be compared. Examination of a particular charge transport model as a test case then indicates several other generically applicable factors that act to further decrease the grating recording efficiency in various recording configurations. These results allow the above analysis to be conveniently applied to a wide range of applications of current technological interest.

The organization of the remainder of the paper is as follows. A brief summary of the origin, material dependence, and implications of the several parameters that affect the photorefractive grating recording sensitivity is provided in the following section (Sect. II). Alternative models of the photorefractive recording process are defined in Sect. III, and the appropriate grating recording efficiencies are presented therein for each case. Assumptions and limitations common to all of the models are discussed in Sect. III.A. Idealized photogeneration and charge transport models are defined and analyzed in Sect. III.B, and more realistic models based upon the analyses of Young *et al.* (1974) and Moharam *et al.* (1979) in the initial stages of recording before significant space charge field amplitudes evolve, and of Kukhtarev (1976) for temporal evolution with small modulation depths of the illumination profile are discussed and analyzed in Sect. III.C. Representative grating recording efficiency calculations are presented in Sect. IV for several common materials and applications in order to illustrate the effect of the quantum inefficiency factors on the overall quantum efficiency of photorefractive recording. Finally, conclusions drawn from the above analyses are discussed in Sect. V.

II. Factors Contributing to the Photorefractive Sensitivity

In this section, each of the five factors outlined in the Introduction that collectively determine the photorefractive sensitivity is briefly discussed, in order to provide a suitable context for the derivation of the quantum limited grating recording efficiency as presented in the following section. It should perhaps be emphasized at the outset that although each of the primary factors considered herein affects at least one of the aforementioned alternative photorefractive sensitivity parameters, not all of the factors enter into each defined parameter.

In the context of photorefractive grating recording, the photogeneration quantum efficiency is related to the fraction of the incident photon flux that generates mobile charge carriers free to participate in subsequent charge transport and trapping processes. For a given recording wavelength, several distinct photoexcitation processes can contribute to the total absorption coefficient. In most commonly considered models of the photorefractive effect, the dominant process is the photogeneration of free carriers from deep donor or acceptor states, such that only one sign mobile carrier (either an electron or a hole) is liberated for each photoevent. In wavelength regions of significant photoconductivity, a second important process is the creation of electron-hole pairs (as well as excitons in certain materials and material structures) by means of band-to-band transitions. Examples of photoinduced transitions that are not likely to contribute to the photorefractive effect, and hence tend to reduce the photogeneration quantum efficiency, are intersubband absorptions, intraionic level promotions, quantum well interlevel excitations, and photochromic charge transfer exchanges. Since each contributing process will in general be characterized by its own charge transport efficiency (discussed below), it is most appropriate to assign separate photogeneration quantum efficiencies not only to each charge carrier type, but also to each distinct photoproduction origin (or photoexcitation channel).

The inherent absorptive inefficiency implied by a finite thickness photorefractive medium also affects the overall photogeneration quantum efficiency. In the case of a thin crystal (such that $\alpha d << 1$, in which α is the absorption coefficient at the recording wavelength and d is the crystal thickness), only a small fraction ($\approx \alpha d$) of the recording beam intensity is absorbed and hence has an opportunity to participate in the photorefractive process. In a thicker crystal, for which the thickness may be optimized for maximum saturation diffraction efficiency, the absorbed fraction is $[1 - e^{-\alpha d}]$ if only the entrance surface of the photorefractive medium is allowed to achieve saturation. In this case the recorded grating will exhibit an exponential nonuniformity throughout the crystal thickness that will decrease the maximum achievable diffraction efficiency. If the entire crystal is exposed to saturation, the photogeneration quantum efficiency will be further reduced by a factor of order $e^{-\alpha d}$. Finally, the effects of reflection at both front and rear crystal surfaces reduce the fraction of incident photons that contribute to the formation of a given grating component, and hence also reduce the effective quantum efficiency. For a crystal thickness optimized for maximum saturation diffraction efficiency with equal write and read wavelengths, and for indices of refraction typical of common photorefractive materials, the combined effects of absorption of the recording beams, absorption of the readout beam, and reflection losses (assuming uncoated surfaces) on the photorefractive sensitivity is approximately an order of magnitude.

An additional effect that acts to reduce the photogeneration quantum efficiency is the constraint imposed indirectly by the nature of the photoexcitation distribution. As we shall demonstrate in the next section, the sinusoidal intensity interference pattern generated by two coherent recording beams is not the optimum (quantum limited) photoexcitation distribution function.

Perhaps the most critical factor that determines the photorefractive sensitivity is the charge transport efficiency, in that this quantity above all others exhibits the greatest degree of variation among commonly investigated photorefractive materials. The charge transport efficiency quantifies the degree to which the average photoproduced charge carrier contributes to the forming space charge grating following photoexcitation, charge transport, and subsequent recombination or trapping. Charge transport in the refractory oxides, as well as in the compound semiconductor family, is a statistical process in which the net result of a given photoinduced event may be to increase, decrease, or leave unchanged the magnitude of the space charge modulation at the fundamental grating frequency.

The statistical nature of the charge transport process can be subsumed in the standard band transport treatment (Young *et al.*, 1974; Moharam *et al.*, 1979; Kukhtarev *et al.*, 1976, 1979), or made explicit as in the hopping conduction model (Feinberg *et al.*, 1980); both approaches lead to essentially equivalent results (Feinberg *et al.*, 1980; Jaura *et al.*, 1986). The charge transport efficiency is strongly affected in photorefractive materials by the dominant conduction mechanism (diffusion and/or drift in an externally applied field), and by the ratio of the average displacement of a photoexcited carrier (before recombination or trapping) to the grating spacing. In both the drift and diffusion regimes, the average displacement depends primarily on the mobility-lifetime product of the photoexcited species. As such, separate charge transport efficiencies should be assigned to each carrier type. A significant net charge transport efficiency will be realized only if there is a net differential in the carrier displacement and trapping process.

It should be noted that there are several situations that can yield vanishing charge transport efficiencies even with large photogeneration quantum efficiencies. For example, in a single donor/single trap model, if the initial Fermi level is more than ten

k_BT or so above the un-ionized donor level, no substantial charge rearrangement is possible due to the unavailability of ionized donors (traps) outside the regions of significant photoexcitation. Likewise, if the initial Fermi level is more than ten k_BT or so below the un-ionized donor level, only band-to-band photoexcitations can contribute with non-vanishing photogeneration quantum efficiencies, which again is likely to yieid negligible charge transport efficiency unless the mobilities and/or lifetimes are significantly different, or operation in the drift regime is engendered by employing an externally applied electric field.

The space charge grating that results from the combination of photoexcitation and charge transport processes in turn gives rise to a modulation of the local electric field at the same spatial frequency through the first Maxwell equation:

$$\nabla\cdot[\underline{\underline{\epsilon}}\mathbf{E}(x)] = \rho(x)/\epsilon_0 \tag{1}$$

in which E is the total electric field at each point in space x, $\underline{\underline{\epsilon}}$ is the dielectric permeability tensor, ρ is the local space charge amplitude, and ϵ_0 is the dielectric permeability of free space. Note that the magnitude of the space charge field derived from a given space charge grating amplitude is inversely proportional to the grating wave vector, as implied by the differential relationship expressed in Eq. (1). The tensor character of $\underline{\underline{\epsilon}}$ is important to note, as many photorefractive materials (particularly the ferroelectric oxides) exhibit marked dielectric anisotropy. Hence, space charge gratings oriented in different directions within the same crystal can give rise to quite large variations in the resultant space charge field. Note further that the magnitude of the space charge field scales inversely with a diagonal component of the dielectric permeability for a given space charge grating oriented along a principal dielectric axis of the crystal. Thus, materials with large dielectric constants (such as $BaTiO_3$ and SBN)

require correspondingly large space charge amplitudes in order to produce an internal electric field modulation of given amplitude.

In the types of photorefractive materials considered herein, the index of refraction is a function of the local electric field. This dependence can arise from a number of electrorefractive effects, including among others the linear electrooptic (Pockels') effect, the quadratic electrooptic (Kerr) effect, the Franz-Keldysh effect, and the quantum confined Stark effect (Chemla *et al.*, 1985). In some materials, notably multiple quantum well structures in compound semiconductors, more than one such electrorefractive effect can contribute simultaneously to the establishment of the resultant index perturbation. For our purposes herein, we consider only the linear electrooptic effect, in which the change in the dielectric impermeability tensor $\underline{\mathbf{B}}$ (the inverse of the dielectric tensor $\underline{\epsilon}$) is linear in the electric field:

$$\Delta B_{ij} = \Delta(\underline{\epsilon}^{-1})_{ij} = r_{ijk}E_k \tag{2}$$

in which the Einstein summation rule is implied, and in which r_{ijk} is the third rank tensor representing the electrooptic coefficient (Kaminow, 1974). As shown in Eq. (2), the tensor nature of the electrooptic effect implies a dependence of the effective index of refraction on the orientation of the grating within the crystal, as well as on the direction of propagation of the readout beam and its polarization. Since in general one can derive an effective electrooptic coefficient for a given experimental configuration, Eq. (2) may be rewritten in the form:

$$\Delta n_{eff}(x) = -1/2n_o^3 r_{eff}\, E(x) \tag{3}$$

in which the x-coordinate is taken parallel to the grating wave vector, and in which $\Delta n_{eff}(x)$ is the effective index modulation resulting from the combination of the space

charge field and the readout configuration, and n_o is the corresponding unperturbed index at the wavelength of the readout beam.

In discussions of the photorefractive sensitivity, it is of considerable value to combine the effects of the previous two factors, since it has been shown that for a wide range of common photorefractive materials, the ratio $n_o^3 r_{eff}/\epsilon$ exhibits considerably reduced variation compared with that of each parameter separately (Glass *et al.*, 1984; Glass, 1984). This is indicative of the general observation that materials with large static polarizabilities typically also exhibit concomitantly large perturbations of the dielectric tensor at optical frequencies in response to low frequency applied (or internal) electric fields.

Once the magnitude of the index perturbation is established, the resultant diffraction efficiency can be directly determined from the thickness (and uniformity) of the grating, the grating wave vector, the readout wavelength, and the corresponding absorption coefficient. Provided that the grating structure is sufficiently thick to assure diffraction in the Bragg regime, and that the phase and amplitude distortions associated with self-diffraction effects can be neglected (Kukhtarev *et al.*, 1979; Marrakchi *et al.*, 1987), the diffraction efficiency is given by (Kogelnik, 1969):

$$\eta = \exp(-\alpha d/\cos\theta_B)\sin^2(\pi\Delta n d/\lambda \cos\theta_B) \qquad (4)$$

in which θ_B is the Bragg angle. The first term in this expression denotes the inherent inefficiency associated with finite absorption at the readout wavelength, while the second term derives from the grating modulation induced diffraction process. Since for suitably small values of the argument $\pi\Delta n d/\lambda \cos\theta_B$ the diffraction efficiency scales as the square of both the index modulation and the grating thickness, this photorefractive sensitivity factor is inherently nonlinear and must be utilized with considerable caution.

Until this point in the discussion, we have assumed implicitly that the grating recording exposures and spatial frequencies employed have been large enough and small enough, respectively, to keep the photorefractive recording process outside the additional limitations imposed by quantum statistical fluctuations. For the recording of photorefractive gratings at very high spatial frequencies and at very low grating recording exposures, several additional factors will come into play, including statistical fluctuations in the photogeneration process, corresponding fluctuations in the charge transport and trapping processes that yield a locally inhomogeneous charge distribution with spatial frequency components near that of the grating wave vector, and concomitant variations in the direction and magnitude of the local space charge field. These additional factors may act to further reduce the overall photorefractive sensitivity.

III. The Grating Recording Efficiency

As defined in the Introduction (Sect. I), the grating recording efficiency of a given photorefractive recording model or configuration is the ratio between the magnitude of the space charge field at a given spatial frequency produced by a fixed number of photogenerated mobile charge carriers, and the maximum quantum limited space charge field that can be produced by the same number of photogenerated carriers. The grating recording efficiency is introduced as a useful metric that effectively compares the photogeneration and charge transport efficiencies of any given model with the optimum quantum limited case, and as such provides an estimate of the overall fundamental quantum efficiency of the photorefractive grating recording process. Note that since the grating recording efficiency is normalized, any effect of the dielectric permeability tensor in establishing the magnitudes of the space charge fields cancels out. Hence the grating recording efficiency may be equivalently defined directly in terms of the space charge grating amplitudes, or in terms of the space charge fields for a dielectric constant of unity.

In this section, idealized photogeneration and charge transport models are postulated and analyzed in order to determine the maximum quantum limited space charge amplitude that can be produced by a fixed number of photogenerated mobile charge carriers at a given spatial frequency. The grating recording efficiencies of several such idealized models are then calculated, to form a set of baseline cases against which the corresponding efficiencies of more realistic photorefractive recording models can be directly compared. For one such model, that of a single mobile charge species transported between a single type of donor site and its associated (ionized) trap sites, several factors are identified that contribute to the grating recording efficiency, and that are particularly illustrative of the fundamental limitations on the photorefractive sensitivity inherent in the model.

III.A. Constraints Common to Alternative Models of Photorefractive Recording

The idealized photogeneration and charge transport models defined herein are abstractions of a more realistic model which has been studied by many authors and analyzed in considerable detail by Kukhtarev *et al.* (1976; 1979), and as such these models share key assumptions concerning quantum recording limitations and implications of the readout process. For simplicity, we confine our attention herein to a version of the model characterized by a single mobile charge species, and a single type of donor sites with associated un-ionized donor and ionized donor (trap) states, although more intricate models have been proposed for particular materials systems (e.g., the existence of mobile holes as well as electrons in lithium niobate (Orlowski and Kratzig, 1978) and barium titanate (Strohkendl *et al.*, 1986), or the existence of multiple trap levels in bismuth. E. Attard and Brown, 1986; Valley, 1986)).

Four principal material species are considered in the single mobile charge/single trap level model, as diagrammed in Fig. 1. The mobile charge species (usually electrons) has number density $n(x,t)$, and is represented by the symbol "e^-" in Fig. 1. It is assumed that negligible densities of these mobile charges exist under dark conditions; essentially all mobile charges are created by photoionization of donors. In this model, "donors" and "traps" are assumed to be different valence states of the same impurity atom (e.g., iron in lithium niobate) or lattice defects, as diagrammed in Fig. 2. The total number of such impurity ions or defects is distributed uniformly throughout the crystal at potential donor sites. A donor is converted into a trap (ionized donor), simultaneously with the creation of a mobile charge, by photogeneration; conversely, a mobile charge is removed from the conduction band, and a trap is converted into a donor, by recombination (Fig. 2). The sum of the number densities of donors and ionized donors is denoted by N_D, which is therefore the total density of potential donor sites, and is assumed to be

constant in space and time. If the number density of ionized donors is taken to be $N_D^+(x,t)$, then the number density of (un-ionized) donors is $[N_D - N_D^+(x,t)]$. In Fig. 1, the donors are represented by the symbol "O" and the traps are represented by the symbol "+". A fourth material species is needed in this model to assure charge conservation, as the density of ionized donors $N_D^+(x,t)$ frequently exceeds the density of mobile charges $n(x,t)$ by several orders of magnitude. This fourth species, with number density N_A, has traditionally been called an acceptor, but is also called a charge compensation site herein and is represented in Fig. 1 by the symbol "-". The charge compensation sites are electrically negative with respect to the donors, and are presumed to be negatively charged impurity ions or lattice defects incorporated during the crystal growth process. The density N_A is assumed to be constant in space and time, and is further assumed to be numerically equal to $N_D^+(x, t=0)$, the concentration of ionized donors in equilibrium. In point of fact, the only requirement for charge compensation is that the product of the charge compensation site density and the effective number of negative charges on each be equal to $N_D^+(x, t=0)$.

It is instructive to consider the relative densities of these various species. Consider for example a crystal of bismuth silicon oxide (Valley and Klein, 1983). Typical concentrations of donors, charge compensation sites, and mobile charges under reasonable optical intensities are of order 10^{19} cm^{-3}, 10^{16} cm^{-3}, and no more that 10^{12} cm^{-3}, respectively. (The number density of electrons under dark conditions for typical crystals of bismuth silicon oxide is entirely negligible.) Thus, $N_D >> N_D^+ \simeq N_A >> n$. A number of other photorefractive media exhibit similar proportionalities. For these materials, the total space charge density $\rho(x,t)$ is given by

$$\rho(x,t) = e\,[N_D^+(x,t) - n(x,t) - N_A] \approx e\,[N_D^+(x,t) - N_A] \qquad (5)$$

The local density of ionized donors N_D^+ can be spatially redistributed under the influence of inhomogeneous photogeneration, as shown by comparing Figs. 1 and 3.

This redistribution occurs as follows: a photon is absorbed by a donor, converting the donor into a trap (ionized donor) and generating a mobile charge carrier. The mobile charge carrier is transported some distance through the photorefractive medium due to drift and/or diffusion, and is subsequently captured by a trap thus generating another donor. If we choose to follow a particular electron, the entire process can be viewed as a simple exchange between equivalent donor sites of a donor state with an ionized donor (or trap) state.

As a final note, the space charge profile $\rho(x,t)$ might evolve into a highly distorted profile with respect to the distribution of incident illumination because of nonlinearities inherent in the recording process. However, since the grating readout is typically performed deep within the Bragg regime, at most a very limited range of spatial frequencies is effective in diffracting the readout light. Thus only one spatial harmonic of the space charge field is of interest, which is assumed herein to be the fundamental harmonic.

In summary, the photorefractive recording model studied by Kukhtarev and all idealized photogeneration and charge transport abstractions considered in the following analysis share these basic assumptions: (1) only a finite amount of space charge $[N_D^+(x, t=0) = N_A]$ exists for generating the space charge electric field; (2) this space charge can be spatially redistributed under the influence of an illumination pattern; (3) no more than one mobile charge is generated for each absorbed photon of the illumination beam; and (4) only the fundamental spatial harmonic of the space charge is effective in the holographic readout process. The photorefractive recording model and the idealized models differ, however, in the details of the photogeneration and charge transport processes, as discussed next.

III.B. Idealized Photogeneration and Charge Transport Models

The grating recording efficiency comprises three successive physical processes: photogeneration, charge transport, and trapping. In order to determine the maximum quantum limited space charge field (at unity dielectric constant) that can result from a fixed number of photogenerated mobile charges, we investigate four idealized photogeneration and charge transport models, as defined below.

The photogeneration process is controlled by the recording illumination profile. Two alternative profiles are considered herein, a periodic comb function and a sinusoidal function. The comb illumination profile $I_C(x)$, as shown in Fig. 4, is defined by

$$I_C(x) = I_{C0}\, \mathrm{comb}(x/\Lambda_G) \Rightarrow I_{C0} \sum \delta(p + x/\Lambda_G)\,, \tag{6}$$

in which the arrow => indicates that the desired function asymptotically approaches a sequence of Dirac delta functions, i.e., a series of intensity peaks with spatial extent small compared with the spatial period Λ_G, but unlike a mathematical delta function the intended comb peaks are large enough to overlap a reasonable number of donor sites. The sinusoid illumination profile $I_S(x)$ is defined by

$$I_S(x) = I_{S0}\,[1 + \cos(2\pi x/\Lambda_G)] \tag{7}$$

in which Λ_G is the spatial period of the illumination profile. Note that the periodic comb function does not correspond to any normal recording configuration. This is acceptable for purposes of this analysis because the maximum quantum limited space charge field at unity dielectric constant is intended to define an upper limit against which more realistic models might be compared. To compare these two illumination profiles, the same photon flux is assumed, i.e., the scaling parameters I_{S0} and I_{C0} are adjusted such that the following normalization integral is satisfied:

$$\int I_S(x)\, dx = \int I_C(x)\, dx\,, \tag{8}$$

from which we derive the relation that $I_{S0} = I_{C0} \equiv I_0$. In the idealized photogeneration

models, all photons in the illumination profile are assumed to be absorbed and generate mobile charge carriers.

Two alternative models of charge transport and trapping are also considered: half wavelength translation, and randomization (Figs. 5 and 6). In the half wavelength translation process shown schematically in Fig. 5, each electron is assumed to translate precisely half the grating period of the illumination profile before capture, without diffusive blooming of the electron cloud. In the analysis that follows, it will be shown that half wavelength translation corresponds to the quantum limited charge transport process, giving rise to the maximum grating recording efficiency. The second idealized transport and capture process to be considered herein is complete randomization, as shown schematically in Fig. 6, in which the photogenerated electrons are assumed to redistribute randomly until they exhibit no spatial variation in density, and only then are recaptured by local traps. The randomization model is intended to represent a more realistic charge transport model; even so, this model corresponds to an upper bound on realizable charge transport efficiencies, as shown in the next section.

Combination of the two photogeneration models and the two charge transport and trapping models produces four alternative idealized recording models for comparison (Fig. 7), which will hereinafter be identified as the bipolar comb, the monopolar comb, the transport-efficient sinusoid, and the baseline sinusoid. Each of the four combinations will be considered in turn. Grating recording efficiencies are calculated for these four combinations in Sect. III.B.1, and the corresponding field saturation behavior is considered in Sect. III.B.2.

III.B.1 Idealized Grating Recording Efficiencies

The bipolar comb results from a periodic comb illumination function combined with a half wavelength translation. The resulting space charge density $\rho(x)$ induced by

this illumination and transport combination is a periodic comb function superimposed on a periodic comb function of opposite sign shifted by half wavelength, as shown in Fig. 8. For a given peak photogenerated charge density ρ_0 (proportional to $I_0 t$, in which t is the exposure time), the resultant space charge distribution $\rho(x)$ is given by:

$$\rho(x) = \rho_0 \, [\mathrm{comb}(x/\Lambda_G) - \mathrm{comb}(1/2 + x/\Lambda_G)] \, . \tag{9}$$

This space charge profile $\rho(x)$ can be readily integrated (see Eq. (1)) to yield a space charge field E(x) which is a square pulse train (Fig. 8). The first spatial harmonic component E_1, defined by

$$E_1 = \; 2\Lambda_G^{-1} \int E(x) \sin(2\pi x/\Lambda_G) \, dx \, , \tag{10}$$

has a magnitude of

$$E_1 = \; 4 \, e \, \rho_0 / \epsilon \, \epsilon_0 \, K_G \, , \tag{11}$$

in which $K_G = 2\pi/\Lambda_G$ is the wave vector of the illumination profile.

The monopolar comb commands interest because it can support the highest space charge field before saturation due to limited ionized trap density, as discussed in the next section. The monopolar comb results from a periodic comb illumination function combined with electron randomization. The resulting space charge profile $\rho(x)$ induced by this illumination and transport combination is a periodic comb function superimposed on a uniform background of opposite charge, as shown in Fig. 9. The resulting charge distribution is represented by

$$\rho(x) = \rho_0 \, [\mathrm{comb}(x/\Lambda_G) - 1)]. \tag{12}$$

The corresponding space charge field exhibits a sawtooth profile (Fig. 9), and the magnitude E_1 of its first spatial harmonic is

$$E_1 = \quad 2\, e\, \rho_0/\epsilon\, \epsilon_0\, K_G\,. \tag{13}$$

The transport-efficient sinusoid results from a sinusoidal illumination profile combined with a half wavelength translation, as shown schematically in Fig. 10, leading to a charge density $\rho(x)$ given by:

$$\rho(x) = 2\, \rho_0 \cos(2\pi x/\Lambda_G)\,, \tag{14}$$

and a first spatial harmonic field component E_1 of

$$E_1 = \quad 2\, e\, \rho_0/\epsilon\, \epsilon_0\, K_G\,. \tag{15}$$

The transport-efficient sinusoid combination is included for completeness, but is not emphasized because it is neither realistic nor does it correspond to any upper bound of grating recording efficiency or saturation performance.

The *baseline sinusoid*, *although seemingly* artificial, is of very pronounced interest because it represents an asymptotic upper bound on the grating recording efficiency predicted by more realistic recording models, as discussed in Sect. III.C below. The baseline sinusoid results from a sinusoidal illumination profile combined with electron distribution randomization. The resulting space charge distribution exhibits a sinusoidal profile, as does the corresponding space charge field, as shown in Fig. 10. The charge density is given by:

$$\rho(x) = \rho_0 \cos(2\pi x/\Lambda_G) \tag{16}$$

corresponding to a first spatial harmonic field component E_1 of

$$E_1 = \quad e\, \rho_0/\epsilon\, \epsilon_0\, K_G\,. \tag{17}$$

As can be seen from the analysis above, the bipolar comb photogeneration/charge transport combination generates the maximum quantum limited space charge field at

unity dielectric constant, and hence provides the normalization constant required for evaluation of the grating recording efficiencies of both idealized and realistic photorefractive recording models. Physically, this optimum combination of photogeneration and charge transport results from the maximum possible average separation of the positive and negative charge distributions, as well as from the fact that the amplitude of the first harmonic of a square wave exceeds the amplitude of the square wave itself.

The grating recording efficiencies for the four idealized photorefractive grating recording combinations are shown in Table 1. The most quantum efficient recording occurs for the bipolar comb, which therefore is assigned a grating recording efficiency of unity. The next most efficient configurations are the monopolar comb and the transport-efficient sinusoid, which both yield a grating recording efficiency of 0.5. The least efficient configuration is the baseline sinusoid, with a grating recording efficiency of 0.25. As we shall show later, the efficiency of the baseline sinusoid is an asymptotic upper limit of more realistic recording models. Table 2 gives the relative photorefractive grating recording sensitivities for these four cases in terms of diffraction efficiency per unit incident photon flux, assuming low diffraction efficiencies such that the efficiency is proportional to the square of the space charge field component E_1. Note that by this measure the sensitivity of the baseline sinusoid is degraded by a factor of 16, over an order of magnitude, compared with the quantum limit represented by the bipolar comb charge distribution function.

III.B.2 Space Charge Saturation for the Idealized Models

Not only is the photorefractive recording sensitivity of concern, but also saturation limitations of the space charge field occurring because of limited ionized trap density. Consider for example the bipolar comb example shown in Fig. 8. The regions of positive space charge grow by the photoexcitation of neutral donors, which converts

them into positively ionized donors (traps) and generates mobile electrons. The electrons are removed from this region by the various transport processes, leaving behind the positively ionized donors. As will be shown later, it is these regions of net positive space charge that primarily contribute to the buildup of the space charge field. The regions of negative space charge grow by the reverse process, i.e., by capturing mobile electrons at local ionized donor sites to form neutral donors. In photorefractive crystals which initially are in quasi-thermodynamic equilibrium, the density of donors typically far exceeds the density of ionized donors, which implies that the regions of negative space charge will saturate first. This corresponds to the complete conversion of all ionized donors into neutral donors at the peaks of the negative space charge distribution, resulting in the local complete cancellation of $N_D^+(x, t=0) = N_A$, as shown schematically in Fig. 11.

The bipolar comb combination exhibits by far the most rapid charge saturation at the lowest space charge field of the four combinations considered herein, because for this combination the electrons after transport are concentrated into a very small volume, with a correspondingly small number of ionized donors available to capture these electrons. Conversely, the monopolar comb combination exhibits the slowest saturation at the highest space charge field because the electrons are essentially uniformly distributed throughout the volume of the photorefractive medium, and hence can be captured by almost all of the ionized donors. The saturation characteristics of the sinusoidal combinations are intermediate between the bipolar and the monopolar combinations.

The relative photorefractive grating recording sensitivities (grating recording efficiencies) and saturation characteristics of these four photogeneration/charge transport combinations are schematically diagrammed in Fig. 12. The photosensitivities are indicated by the inital linear slopes, and the saturation by the final space charge field

levels. The space charge field is plotted here in units of E_q, which is defined as $eN_A/\epsilon\epsilon_0 K_G$ (Amodei, 1971) (this expression is valid when the density of charge compensation sites N_A is much smaller than the total density of potential donor sites N_D). Note that the saturation level for the bipolar comb should really be much closer to the horizontal axis; it has been overstated for clarity of illustration.

Having calculated the grating recording efficiencies and saturation fields for these four highly idealized cases, we now proceed to compare these ideal results with more realistic photogeneration and charge transport models in the next two sections.

III.C Realistic Recording Models

The idealized photorefractive recording models discussed above allow the effects of the photogeneration profile and charge transport process on the grating recording efficiency to be assessed relative to the fundamental quantum limits. We now examine a more realistic model applicable to a wide range of commonly investigated photorefractive media, which exhibits an overall efficiency degraded from that of the baseline sinusoid case presented above by several additional factors. In this model, the photogeneration profile is assumed to be sinusoidal, and the effects of the photogeneration quantum efficiency, absorption coefficients, and reflections losses are assumed to be space-invariant and hence can be directly incorporated in any estimate of the photorefractive grating recording sensitivity. Another major factor is an inherent inefficiency in the charge transport process; this inefficiency is studied in Sect. III.C.1 using analytical solutions derived by Young *et al.* (1974; see also Moharam *et al.*, 1979) which are valid in the initial recording interval, before significant space charge fields have evolved. A related factor derives from the reduced recording sensitivity exhibited as the space charge field approaches its steady state limit. This is studied in Sect. III.C.2 using analytical solutions derived by Kukhtarev (1976) which describe the temporal evolution of the space charge field in the limit of very low illumination profile

modulation depths. Recording configurations using low modulation depths are inherently inefficient, as most of the photons in the illumination contribute a uniform background photocurrent, and only a fraction of the incident intensity contributes to the spatial structure of the image. The modulation depth, therefore, is also a factor that reduces the photorefractive grating recording sensitivity. However, low modulation depths are necessary in certain recording techniques for enhancing the space charge field in the steady state limit, such as the running grating process discussed in Sect. III.C.3.

III.C.1 Initial Recording Sensitivity

To determine the existence of degraded charge transport efficiency, the idealized recording models must be compared with more realistic charge transport solutions, such as those given by Young *et al.* (1974; see also Moharam *et al.*, 1979). Under normal recording conditions, the coupled photorefractive recording equations are nonlinear, making analytic solutions difficult or impossible to derive. However, a significantly simplified analysis can be utilized during the initial recording period, which enables analytic solutions to be derived at least for certain illumination profiles. These analytic solutions are well worth studying for the physical insight they furnish into the charge transport process.

The analytical simplification derives from a linearization of the recording equations, in which two recording parameters, the total electric field and the ionized donor (trap) density, remain essentially constant throughout space during the initial recording interval (Young *et al.*, 1974; Moharam *et al.*, 1979). This assumes that the photorefractive crystal is initially in quasi-thermodynamic equilibrium, i.e., with a spatially uniform distribution of ionized donors $N_D^+(x, t=0) = N_A$, and that the trap density remains essentially constant throughout this initial recording interval.

The analytic solutions which exist in this regime have typically emphasized recording with sinusoidal illumination profiles. While such a profile corresponds closely with typical experimental situations and simplifies the mathematics, the physics of the transport process is somewhat obscured in comparison with an alternative illumination profile, that of a very narrow spike, an approximation of a Dirac delta function. Typical electron density profiles obtained in response to a narrow spike illumination profile are shown in Fig. 13 for the cases of diffusion-only transport (top illustration), drift-only transport (middle illustration), and one particular combination of drift and diffusion processes (bottom illustration); the derivation of these figures is described below. These figures emphasize several important features of a more realistic transport analysis. The transport mechanism is inherently a random process, with a spread in characteristic transport lengths associated with a corresponding spread in charge carrier lifetimes. Useful parameters for characterizing the transport processes are the average transport lengths L_E for drift-induced transport and L_D for diffusive transport, defined as (Young *et al.*, 1974; Moharam *et al.*, 1979)

$$L_E = \mu\tau E_0 \,, \tag{18}$$

and

$$L_D = (D\tau)^{1/2} = (k_B T/e)^{1/2}(\mu\tau)^{1/2} \,, \tag{19}$$

in which μ is the mobility of the free carriers, E_0 is the applied bias electric field, τ is the free carrier lifetime, D is the diffusion coefficient for the free carriers, k_B is Boltzmann's constant, T is the crystal temperature, and e is the charge of an electron. Einstein's relation between the diffusion coefficient and the mobility has been used in Eq. (19). Note that in both cases, the transport lengths are functions of the $\mu\tau$ product.

For the diffusion-only case, the electron spread is symmetrical, which introduces no net phase shift in the photorefractive response to any arbitrary illumination profile. For the drift case, as well as for the combined drift/diffusion case, the electron

distribution is skewed to one side by the presence of an applied bias field, which does in fact introduce a phase shift when recording particular illumination profiles, as shown in Fig. 13.

Now consider an illumination profile I(x) which is sinusoidal and of the form

$$I(x) = I_0 [1 + m \cos(K_G x)] , \tag{20}$$

in which m is the modulation depth of the light profile, and K_G is the wave vector associated with the interference pattern. Young *et al.* (1974; see also Moharam *et al.*, 1979) have derived expressions for the growth of the space charge field for the initial recording interval. When transport is dominated by diffusion, the initial growth of the first harmonic component E_1 of the space charge field is expressed by

$$E_1 = m [teg_0/\epsilon\epsilon_0 K_G] [K_G^2 L_D^2/(1+K_G^2 L_D^2)] , \tag{21}$$

in which g_0 is the photogeneration rate and t is the time relative to the initiation of grating recording. The initial growth when drift transport dominates is expressed by

$$E_1 = m [teg_0/\epsilon\epsilon_0 K_G] [K_G L_E (1+K_G^2 L_E^2)^{-1/2} e^{i\phi}], \tag{22}$$

in which the phase shift ϕ is defined by

$$\tan \phi = K_G L_E . \tag{23}$$

The first bracketed term $[teg_0/\epsilon\epsilon_0 K_G]$ in Eqs. (21) and (22) corresponds to the grating recording efficiency predicted by the baseline sinusoid model (with $\rho_0 = teg_0$), as discussed in Sect. III.B. This represents the upper bound on achievable recording sensitivity. The second bracketed terms in Eqs. (21) and (22) correspond to an additional charge transport inefficiency inherent in more realistic models of photorefractive recording, the subject of this section. This transport inefficiency factor is plotted as a function of increasing transport length in Fig. 14 for diffusion-only

transport and in Fig. 15 for drift-only transport. Recall that these curves apply only during the initial recording interval, before significant space charge has accrued. Later recording will have less transport efficiency because of the presence of the space charge field. Note in Figs. 14 and 15 that the charge transport efficiency asymptotically approaches its maximum value in the limit of very long transport lengths, as intuitively expected, but even in this limit the maximum recording sensitivity is that of the baseline sinusoid, not that of the transport efficient sinusoid.

The reason why the recording sensitivity only reaches the baseline sinusoid level in this limit is best understood by considering the spatial modulation profile of the mobile charge density n(x). Analytical expressions for the modulation depth of the mobile charge density can be readily derived from the same analysis that led to Eqs. (21) and (22) (Young *et al.*, 1974; Moharam *et al.*, 1979), for times sufficiently long compared with the mobile carrier lifetime so that the mobile charge density represents a quasi-steady state distribution, and also sufficiently short so as to remain in the initial recording regime. In the diffusion only case, the mobile charge density is

$$n(x) = \tau g_0 \ [1 + m_e(K_G L_D)\cos(2\pi x/\Lambda_G)] \ , \tag{24}$$

in which τ is the mobile carrier lifetime, and in which the modulation depth $m_e(K_G L_D)$ is given by

$$m_e(K_G L_D) = m \ /(1+K_G^2 L_D^2) \ . \tag{25}$$

Note that this charge distribution has the least efficient possible phase for building up a space charge field. Recall from the discussion in Sect. III.B.1 that the optimum phase of the mobile charge profile is a 180^0 phase shift with respect to the incident illumination profile, allowing the space charge field contribution of the mobile charge after subsequent trapping to add to the contribution of the excess positively charged trap profile produced by the photogeneration process. Here, however, we find that the

mobile charge profile is aligned coincident with the illumination and with the excess positively charged trap profiles. As a consequence, when a mobile charge is captured, it removes one of the incremental photogenerated traps, thereby reducing the overall space charge profile. A net space charge field can then accrue in the diffusion-dominated transport case only by randomizing the mobile charge distribution, leading to a reduction in the modulation depth $m_e(K_G L_D)$. Note in Fig. 14 that the rise in charge transport efficiency with increasing transport length $K_G L_D$ is coincident with a reduction of the mobile charge modulation depth. Optimum charge transport efficiency occurs when the mobile charge profile has been completely randomized.

Similarly, for drift-dominated transport, the mobile charge profile has a modulation depth of

$$m_e(K_G L_E) = m / (1+K_G^2 L_E^2)^{1/2} , \tag{26}$$

and a phase shift of ϕ, as given by Eq. (23). The charge transport efficiency, mobile charge modulation depth, and phase shift of the mobile charge profile with respect to the incident illumination are plotted in Fig. 15 as a function of transport length $K_G L_E$ for drift-dominated transport, which is in turn proportional to the applied bias field E_0. For short transport lengths $K_G L_E$, the phase shift is close to 0^o, which as pointed out above is the least efficient phase for building a space charge field, and the modulation depth m_e exhibits its maximum value of m, the modulation depth of the illumination profile. For longer transport lengths, the mobile charge profile phase shifts away from 0^o, but is always less than 90^o, and hence always degrades the net space charge. The modulation depth m_e similarly decreases with increasing charge transport length. The most efficient charge transport occurs for very large transport lengths $K_G L_E$, for which the mobile charge profile has become completely randomized.

The reason why the phase of the mobile charge profile never exceeds 90^{o} for drift-dominated transport (and why the phase is always 0^{o} for diffusion-dominated transport) can perhaps best be appreciated from Fig. 13, which shows the mobile charge profile which is generated in response to a narrow spike (i.e., very tightly focused) illumination profile. The profiles shown in Fig. 13 can be derived by Fourier decomposing the Dirac delta function of the illumination profile into an equivalent set of spatial harmonics, applying Eqs. (25) and (26) to find the mobile charge profile in response to each spatial frequency, and then performing an inverse Fourier transform. But the profiles shown in Fig. 13 are also intuitively reasonable. Diffusion tends to broaden the mobile charge density symmetrically about the location of an illumination region, whereas drift tends to pull the mobile charges all to one side. Finally, note that the mobile charge distribution resulting from any arbitrary illumination profile can be derived by convolving the illumination profile with the distributions shown in Fig. 13 (which can be considered to be blur functions). Because of the monotonically decreasing shape of these blur functions, no phase shift in excess of 90^{o} is feasible.

Thus we find that the baseline sinusoid model represents an upper bound on the grating recording efficiency predicted by the single mobile charge species/single donor/single trap photorefractive recording model.

III.C.2 Temporal Approach to Steady State

In addition to the charge transport inefficiency factor just discussed, two additional factors must be considered when evaluating any realistic recording situation. One factor pertains to the saturation in temporal growth of the space charge field, resulting in a reduced growth rate as the field approaches its steady state limit. The second factor pertains to the modulation depth m of the illumination profile, which is often chosen to be small to enable enhanced recording techniques, as discussed in the next section.

The reduction in sensitivity of the recording process as the space charge field approaches its steady state limit can be assessed from analytic solutions which have been derived by Kukhtarev (1976). Typical solutions are shown in Fig. 16 for a variety of bias fields E_0, scaled to the maximum possible field E_q due to limited ionized trap density (discussed in Sect. 3.2.2). These curves were generated based upon typical charge mobility-lifetime product $\mu\tau$ parameters for bismuth silicon oxide ($Bi_{12}SiO_{20}$; BSO), as given in Table 3. (A bias field to saturation field ratio of 10:1 is unphysical for BSO, requiring exceptionally high bias fields and/or spatial frequencies, but is included for generality.) Note that in all cases shown the time needed to reach saturation is approximately a factor of two longer than that predicted by the baseline sinusoid idealized transport model.

Due to the fact that the analytical solutions for temporal evolution derived by Kukhtarev (1976) are valid only for small modulation depths m of the illumination profile, such solutions describe low recording efficiency situations in which the majority of the photons in the illumination beam contribute a uniform photocurrent and only a small fraction of the photons convey the spatial structure in the image profile. This point is emphasized in Fig. 17, in which numerical solutions of the photorefractive equations are presented showing the temporal evolution of the first spatial harmonic component of the space charge field for various modulation depths m. These solutions have been produced by the authors using numerical techniques discussed by Moharam *et al.* (1979), but applied to the full set of photorefractive equations proposed by Kukhtarev (1976).

In Fig. 17, a crystal of bismuth silicon oxide illuminated by a 300 cycle/mm sinusoid has been assumed. Note in this figure the slight temporal oscillations which can be observed in the temporal evolution. The strength of these oscillations is directly dependent on the charge mobility-lifetime product $\mu\tau$; the curves in Fig. 17 correspond

to the material and configuration parameters as listed in Table 3. The reader is cautioned that a wide distribution in mobility-lifetime products has been measured in crystals with nominally the same composition (Lesaux *et al.*, 1986).

Note also in Fig. 17 that the highest space charge fields are associated with the highest illumination profile modulation depths, a regime for which the analytical solutions derived by Kukhtarev are no longer applicable. Furthermore, to first order the space charge field E_1 both in the initial recording regime as well as in saturation is directly proportional to the modulation depth parameter m.

III.C.3 Enhanced Recording Techniques

The recording configuration just considered assumes a stationary illumination profile and a constant applied bias field (for brevity, called hereinafter the stationary illumination technique). Space charge fields with much higher steady state limits can be obtained by either of two alternative nonstationary recording configurations: one technique is to translate the illumination profile with respect to the photorefractive crystal (hereinafter referred to as the running grating technique) (Huignard and Marrakchi, 1981; Stepanov *et al.*, 1982; Valley, 1984; Refregier *et al.*, 1985), and the second technique is to periodically reverse the direction of the applied bias field (the alternating field technique) (Stepanov and Petrov, 1985). These techniques do not improve the rate at which space charge builds up with time under constant illumination intensity, compared with the stationary illumination/constant field recording configuration.

The running grating technique in particular was chosen for study herein, both to illustrate enhanced photorefractive recording concepts, and because it is commonly employed to provide significant amplification of weak images. This technique can be

studied by a combination of analytical solutions which are valid in the linearized regime of very small modulation depths, and numerical solutions for larger image modulations.

The relative advantage of the running grating technique is summarized in Fig. 18, which has been derived from the analytical solutions of Valley (1984) and Refregier *et al.* (1985) in the limit of small illumination profile modulation depths. A grating spatial frequency of 110 cycles/mm, and parameters typical of BSO have been assumed in the solutions shown in Fig. 18. The spatial frequency is chosen to be 110 cycles/mm, rather than 300 cycles/mm, because the enhancement of the steady state space charge field is maximized for this grating frequency, assuming material parameters for BSO as given in Table 3. Note that the steady state limit of the space charge field is in fact increased by the running grating technique relative to that obtained with a stationary grating, but that the initial recording sensitivities are asymptotically equal. The initial growth of the space charge field continues to be bounded by a combination of quantum limitations and charge transport inefficiencies, as described above. Furthermore, the additional factor of two reduction in sensitivity observed on approach to saturation obtains for both solutions.

The magnitude of the steady state space charge field is necessarily quite small in the linearized regime for which the analytical solutions apply, as the field is proportional to the modulation depth m of the illumination profile, which must be kept small to assure accurate analytical solutions. Larger space charge fields require larger illumination modulation depths, and eventual nonlinear saturation in the space charge field. Numerical solutions have been generated by the authors to explore the onset of this field saturation, with typical results as shown in Fig. 19, assuming typical BSO material parameters listed in Table 3. Note that the saturation of the steady state field occurs at quite modest values of the modulation depth, which is consistent with experimental observations reported by Refregier *et al.* (1985). Thus the running grating

technique proves to be most effective for amplifying weak images, but not for recording large space charge fields. For recording the largest fields, stationary illumination is preferred.

IV. Representative Grating Recording Efficiency Calculations

In order to illustrate the above concepts, we proceed in this section to consider the several factors that contribute to the overall grating recording efficiency for two different types of materials and for two different types of applications. The materials considered are bismuth silicon oxide ($Bi_{12}SiO_{20}$, or BSO) and barium titanate ($BaTiO_3$); the principal material parameters assumed in the estimates are listed in Table 3, based upon a set of values published by Valley and Klein (1983). The two applications considered are those of reconfigurable holographic interconnections and the amplification of weak images.

Let us first consider a reconfigurable holographic interconnection implemented in bismuth silicon oxide. For simplicity, we assume that only one grating with a spatial frequency of 300 cycles/mm is recorded in the crystal. To achieve maximum diffraction efficiency, the modulation depth of the recording beams should be as large as possible; ideally, m = 1. Also, a bias electric field is typically applied to crystals of BSO to enhance the photosensitivity; a field of 6 kV/cm is typical, implying a drift transport length L_E from Eq. (18) of 9 μm, assuming the mobility-lifetime product given in Table 3. For a 300 cycle/mm grating frequency, this implies an (L_E/Λ_G) ratio of about 2.7, and from Fig. 15 we see that this corresponds to essentially 100% charge transport efficiency in the initial stages of recording.

The recording efficiency for a BSO interconnect, compared with the ideal (i.e., bipolar comb) quantum efficiency limit, is listed in Table 4. Three factors are considered in this and subsequent tables. The first is the 25% efficiency factor that applies between the baseline sinusoid and bipolar comb cases, a factor which is common to all recording configurations considered in this section. The second factor is the charge transport efficiency, comparing actual recording performance to that of the baseline sinusoid case. The BSO interconnect is assigned a 50% charge transport

efficiency to account for the factor of 2 increase in recording energy (photon flux) estimated to reach saturation, as shown in Fig. 16 and discussed in Sect. III.C.2. The final factor is the modulation depth, which we have assumed to be unity for a reconfigurable interconnect. Thus the total grating recording efficiency, in terms of the magnitude of the space charge field generated per unit photon flux, compared with ideal quantum efficient recording, is only about 12.5% for a BSO interconnection. This gives a diffraction efficiency derating factor of only 1.6% (assuming a diffraction efficiency which is proportional to the square of the space charge field).

As a second example, consider image amplification in BSO using a running grating enhanced recording technique. We will again assume a 110 cycle/mm grating spatial frequency (see Sect. 3.3.3), and a bias field of 6 kV/cm applied to the BSO crystal. However, the modulation depth must be reduced from 100% to of order 10% to achieve the peak space charge enhancement in saturation provided by the running grating recording technique, as shown in Fig. 19 and as discussed in Sect. III.C.3. For our calculations, we'll assume a modulation depth of 10%, implying a corresponding reduction in the quantum efficiency of the recording process; i.e., most of the photons must supply the pump beam, with comparatively few photons in the signal beam containing signal information. Thus the total quantum efficiency for image amplification, shown in Table 5, is an order of magnitude lower than that given in Table 4 for the reconfigurable interconnection. Correspondingly, the diffraction efficiency derating factor is two orders of magnitude lower for this case than for the case of the reconfigurable interconnection, and nearly four orders of magnitude less efficient than the quantum limitation.

As a final example, consider a reconfigurable interconnection in barium titanate, with grating recording efficiency as shown in Table 6 for the case of recording by diffusion transport only. Because of the large electrooptic coefficient in barium titanate,

only very modest space charge fields, typically a fraction of the diffusion field for a 300 cycle/mm grating, are needed to achieve peak diffraction efficiencies in reasonable size crystals. Hence diffusion transport often proves to be sufficient in this material in order to generate experimentally useful diffraction efficiencies. Unfortunately, two factors work against the high electrooptic coefficient in barium titanate. One is the concomitantly high dielectric constant, which from Maxwell's first equation implies that a considerable amount of space charge must be moved to achieve a modest space charge field. As mentioned in the Introduction, the combined material parameter (n03reff/e), which is a measure of the amount of optical index modulation per unit space charge, proves to be surprisingly constant from material to material (Glass et al., 1984; Glass, 1984). Based upon this measure alone, barium titanate proves to be modestly superior to BSO as shown in Table 3.

The second, and far more serious, factor degrading grating recording efficiency for diffusion recording in barium titanate is its low mobility-lifetime product $\mu\tau$ which is almost two orders of magnitude lower than that for BSO, implying a seriously degraded charge transport efficiency. The diffusion transport length L_D predicted by Eq. (19) is of order 35 nm, based upon the mobility-lifetime product $\mu\tau$ given in Table 3. Assuming a 300 cycle/mm grating, this implies by Eq. (21) a charge transport inefficiency factor of order 0.004. We include an additional factor of 0.5 in Table 6 to account for the additional photon flux required to reach saturation, as indicated in Fig. 16 and the discussion in Sect. III.C.2. Thus we find that the grating recording efficiency for diffusion recording in barium titanate is some 250 times less than that for drift recording in BSO, in terms of space charge generated per unit photon.

Thus we have considered representative examples of several different materials, transport mechanisms and efficiencies, and recording applications. We find that, even in the most efficient grating recording configurations, a significant inefficiency still exists between the ideal quantum limitations and actual performance.

V. Conclusions

In this paper, we have considered a number of the fundamental physical limitations that delimit the potential performance of photorefractive materials. In particular, we have described several idealized photogeneration and charge transport models in terms of the grating recording efficiency, and have identified one such model (the bipolar comb) as the quantum limit against which other such models may be compared. The bipolar comb model generates the maximum possible fundamental harmonic of the space charge field at unity dielectric constant for a given number of photoexcited mobile charge carriers. A second idealized model, the baseline sinusoid, provides an upper bound for the grating recording efficiency of a more realistic photorefractive grating recording model involving a single mobile charge carrier and a single donor/single trap photorefractive center. This upper bound was shown to be a factor of 4 less efficient in generating a given space charge field than the quantum limitations imply for an optimum photogeneration distribution and perfectly efficient charge transport. An additional factor of 2 accrues from the nonlinearity of the grating recording process observed as the space charge field nears saturation. The combination of absorption and reflection losses in typical photorefractive recording configurations contributes approximately one order of magnitude to the inefficiency of grating recording relative to the incident photon flux. Hence overall, the usual photorefractive recording configuration exhibits a sensitivity that is approximately two orders of magnitude less than that achievable in the quantum limit. This results in roughly four orders of magnitude reduction in the corresponding diffraction efficiency per incident photon.

These considerations explain to a certain degree why photorefractive recording has proven to be relatively insensitive as compared with distinct but related mechanisms of spatial light modulation. Perhaps far more important, however, are the implications of the above analysis for the conceptual design and technological implementation of grating recording media that operate far closer to the quantum limits.

Acknowledgements

This research was supported in part by the Defense Advanced Research Projects Agency (through the Office of Naval Research), the Air Force Office of Scientific Research, and the Joint Services Electronics Program.

References:

1. Amodei, J. J. (1971). "Analysis of transport processes during holographic recording in insulators," RCA Review 32, 185-198.

2. Amodei, J. J., and Staebler, D. L. (1972). "Holographic recording in lithium niobate," RCA Review 33, 71-93.

3. Attard, A. E., and Brown, T. X. (1986). "Experimental observations of trapping levels in BSO," Appl. Opt. 25(18), 3253-3259.

4. Chemla, D. S., Miller, D. A. B., and Smith, P. W. (1985). "Nonlinear optical properties of GaAs-GaAlAs multiple quantum well material: phenomena and applications," Opt. Eng. 24(4), 556-564.

5. Feinberg, J., Heiman, D., Tanguay, Jr., A. R., and Hellwarth, R. W. (1980). "Photorefractive effects and light-induced charge migration in barium titanate," J. Appl. Phys. 51(3), 1297-1305.

6. Glass, A. M. (1978). "The photorefractive effect," Opt. Eng. 17(5), 470-479.

7. Glass, A. M. (1984). "Materials for optical information processing," Science 226, 657-662.

8. Glass, A. M., Johnson, A. M., Olson, D. H., Simpson, W., and Ballman, A. A. (1984). "Four-wave mixing in semi-insulating InP and GaAs using the photorefractive effect," Appl. Phys. Lett. 44(10), 948-950.

9. Glass, A. M., Klein, M. B., and Valley, G. C. (1987). "Fundamental limit of the speed of photorefractive effect and its impact on device applications and material research: comment," Appl. Opt. 26(16), 3189-3190.

10. Gunter, P. (1982). "Holography, coherent light amplification, and optical phase conjugation with photorefractive materials," Phys. Rep. 93, 199-299.

11. Huignard, J. P., and Marrakchi, A. (1981). "Coherent signal beam amplification in two-wave mixing experiments with photorefractive $Bi_{12}SiO_{20}$ crystals," Opt. Commun. 38(4), 249.

12. Huignard, J. P., and Micheron, F. (1976). "High-sensitivity read-write volume holographic storage in $Bi_{12}SiO_{20}$ and $Bi_{12}GeO_{20}$ crystals," Appl. Phys. Lett. 29(9), 591-593.

13. Jaura, R., Hall, T. J., and Foote, P. D. (1986). "Simplified band transport model of the photorefractive effect," Opt. Eng. 25(9), 1068-1074.

14. Kaminow, I. P. (1974). An Introduction to Electrooptic Devices (Academic Press, New York).

15. Kamshilin, A. A., and Petrov, M. P. (1980). "Holographic image conversion in a $Bi_{12}SiO_{20}$ crystal," Sov. Tech. Phys. Lett. 6(3), 144-145.

16. Kogelnik, H. (1969). "Coupled wave theory for thick hologram gratings," Bell Syst. Tech. J. 48(9), 2909-2947.

17. Kukhtarev, N. V. (1976). "Kinetics of hologram recording and erasure in electrooptic crystals," Sov. Tech. Phys. Lett. 2(12), 438-440.

18. Kukhtarev, N. V., Markov, V. B., Odulov, S. G., Soskin, M. S., and Vinetskii, V. L. (1979). "Holographic storage in electrooptic crystals," Ferroelectrics 22, 949-964.

19. Lesaux, G., Launay, J. C., and Brun, A. (1986). "Transient photocurrent induced by nanosecond light pulses in BSO and BGO," Opt. Commun. 57(3), 166-170.

20. von der Linde, D., and Glass, A. M. (1975). "Photorefractive effects for reversible holographic storage of information," Appl. Phys. 8, 85-100.

21. Marrakchi, A., Johnson, R. V., and Tanguay, Jr., A. R. (1987). "Polarization properties of enhanced self-diffraction in sillenite crystals," IEEE J. Quantum Electron. QE-23(12), 2142-2151.

22. Marrakchi, A., Tanguay, Jr., A. R., Yu, J., and Psaltis, D. (1985). "Physical characterization of the photorefractive incoherent-to-coherent optical converter," Opt. Eng. 24(1), 124-131.

23. Micheron, F. (1978). "Sensitivity of the photorefractive process," Ferroelectrics 18, 153-159.

24. Moharam, M. G., Gaylord, T. K., Magnusson, R., and Young, L. (1979). "Holographic grating formation in photorefractive crystals with arbitrary electron transport lengths," J. Appl. Phys. 50(9), 5642-5651.

25. Orlowski, R., and Kratzig, E. (1978). "Holographic method for the determination of photo-induced electron and hole transport in electro-optic crystals," Solid State Comm. 27(12), 1351-1354.

26. Refregier, Ph., Solymar, L., Rajbenbach, H., and Huignard, J. P. (1985). "Two-beam coupling in photorefractive $Bi_{12}SiO_{20}$ crystals with moving grating: Theory and experiments," J. Appl. Phys. 58(1), 45.

27. Shi, Y., Psaltis, D., Marrakchi, A., and Tanguay, Jr., A. R. (1983). "Photorefractive incoherent-to-coherent optical converter," Appl. Opt. 22(23), 3665-3667.

28. Stepanov, S. I. and Petrov, M. P. (1985). "Efficient unstationary holographic recording in photorefractive crystals under an external alternating electric field," Opt. Commun. 53(5), 292.

29. Stepanov, S. I., Kulikov, V. V., and Petrov, M. P. (1982). "Running holograms in photorefractive $Bi_{12}SiO_{20}$ crystals," Opt. Commun. 44(1), 19.

30. Strohkendl, F. P., Jonathan, J. M. C., and Hellwarth, R. W. (1986). "Hole-electron competition in photorefractive gratings," Opt. Lett. 11(5), 312-314.

31. Tanguay, Jr., A. R. (1985). "Materials requirements for optical processing and computing devices," Opt. Eng. 24(1), 2-18.

32. Valley, G. C. (1984). "Two-wave mixing with an applied field and a moving grating," J. Opt. Soc. Am. B 1(6), 868.

33. Valley, G. C. (1986). "Simultaneous electron/hole transport in photorefractive materials," J. Appl. Phys. 59(10), 3363-3366.

34. Valley, G. C., and Klein, M. B. (1983). "Optimal properties of photorefractive materials for optical data processing," Opt. Eng. 22(6), 704-711.

35. Yeh, P. (1987a). "Fundamental limit of the speed of photorefractive effect and its impact on device applications and material research," Appl. Opt. 26(4), 602-604.

36. Yeh, P. (1987b). "Fundamental limit of the speed of photorefractive effect and its impact on device applications and material research: author's reply to comment," Appl. Opt. 26(16), 3190-3191.

37. Young, L., Wong, W. K. Y., Thewalt, M. L. W., and Cornish, W. D. (1974). "Theory of formation of phase holograms in lithium niobate," Appl. Phys. Lett. 24(6), 264-265.

List of Figures

17. Increase in saturation space charge field with increasing modulation depth, showing transition from linear recording to nonlinear recording regimes.

18. Temporal evolution of the space charge field, comparing stationary with running gratings, low modulation limit.

19. Steady state space charge field as a function of modulation depth, comparing stationary with running gratings.

Table 1: Grating Recording Efficiencies of the Various Idealized Models

Bipolar Comb	1.0
Monopolar Comb	0.5
Transport-Efficient Sinusoid	0.5
Baseline Sinusoid	0.25

Table 2: Relative Diffraction Efficiencies of the Various Idealized Models

Bipolar Comb	1.0
Monopolar Comb	0.25
Transport-Efficient Sinusoid	0.25
Baseline Sinusoid	0.0625

Table 3: Material Parameters Assumed in the Calculations[8]

	$Bi_{12}SiO_{20}$	$BaTiO_3$	Units
Mobility-Lifetime Product $\mu\tau$	15	0.5	$\mu m^2/V$
Trap Density N_A	1×10^{16}	2×10^{16}	cm^{-3}
Dielectric Constant ϵ	56	4300 (ϵ_{11})	
Index of Refraction n_0	2.5	2.4	
Electrooptic coefficient r_{ij}	5 (r_{41})	1640 (r_{42})	pm/V
$n^3 r_{ij}/\epsilon$	1.4	5.3	pm/V

Table 4: Representative Grating Recording Efficiency Calculation:

Reconfigurable Interconnection

Drift Recording in Bismuth Silicon Oxide

Modulation Depth = 1.0

300 cycles/mm Grating Frequency

Space Charge Efficiency Factors:

Baseline Sinusoid / Bipolar Comb	0.25
Charge Transport (to saturation)	0.5
Modulation Depth Factor	1.0
Grating Recording Efficiency	0.125
Diffraction Efficiency Derating Factor	0.015

Table 5: Representative Grating Recording Efficiency Calculation:

Two-Wave Image Amplification

Running Gratings in Bismuth Silicon Oxide

Modulation Depth = 0.1

110 cycles/mm Grating Frequency

Space Charge Efficiency Factors:

Baseline Sinusoid / Bipolar Comb	0.25
Charge Transport (to saturation)	0.5
Modulation Depth Factor	0.1
Grating Recording Efficiency	0.0125
Diffraction Efficiency Derating Factor	0.00015

Table 6: Representative Grating Recording Efficiency Calculation:

Reconfigurable Interconnect

Diffusion Recording in Barium Titanate

Modulation Depth = 1.0

300 cycles/mm Grating Frequency

Space Charge Efficiency Factors:

Baseline Sinusoid / Bipolar Comb	0.25
Charge Transport (to saturation)	0.002
Modulation Depth Factor	1.0

Grating Recording Efficiency	0.0005
Diffraction Efficiency Derating Factor	2.5×10^{-7}

Photorefractive Recording Model

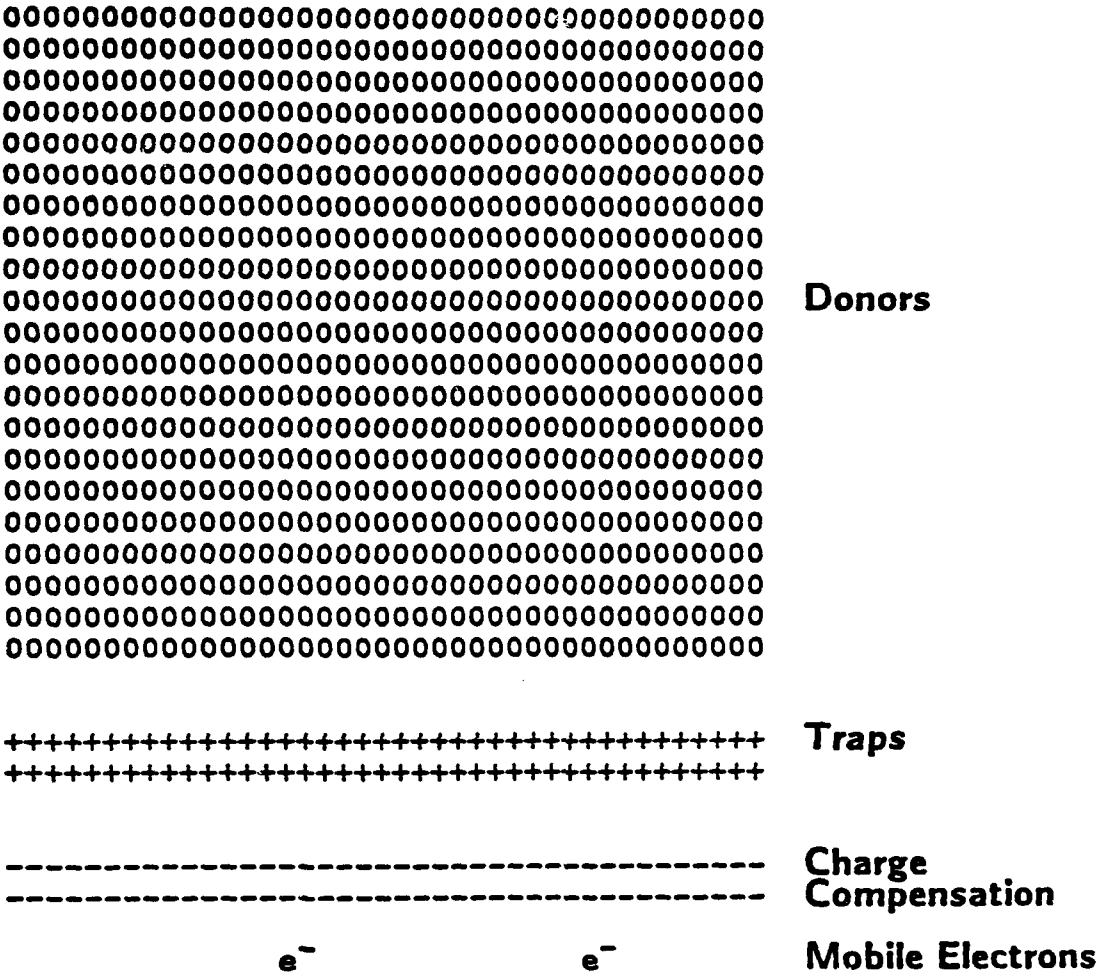


Figure 1

PHOTOREFRACTIVE RECORDING MODEL

Assume single active species:

Donor (e.g.: Fe^{2+}) Trap (e.g.: Fe^{3+})

Generation Process:

$h\nu$ + Donor ⇒ Trap + e^-

Photon Donor Trap Electron

Recombination Process:

Trap + e^- ⇒ Donor

Trap Electron Donor

Figure 2

Photorefractive Recording Model

Charge Redistribution

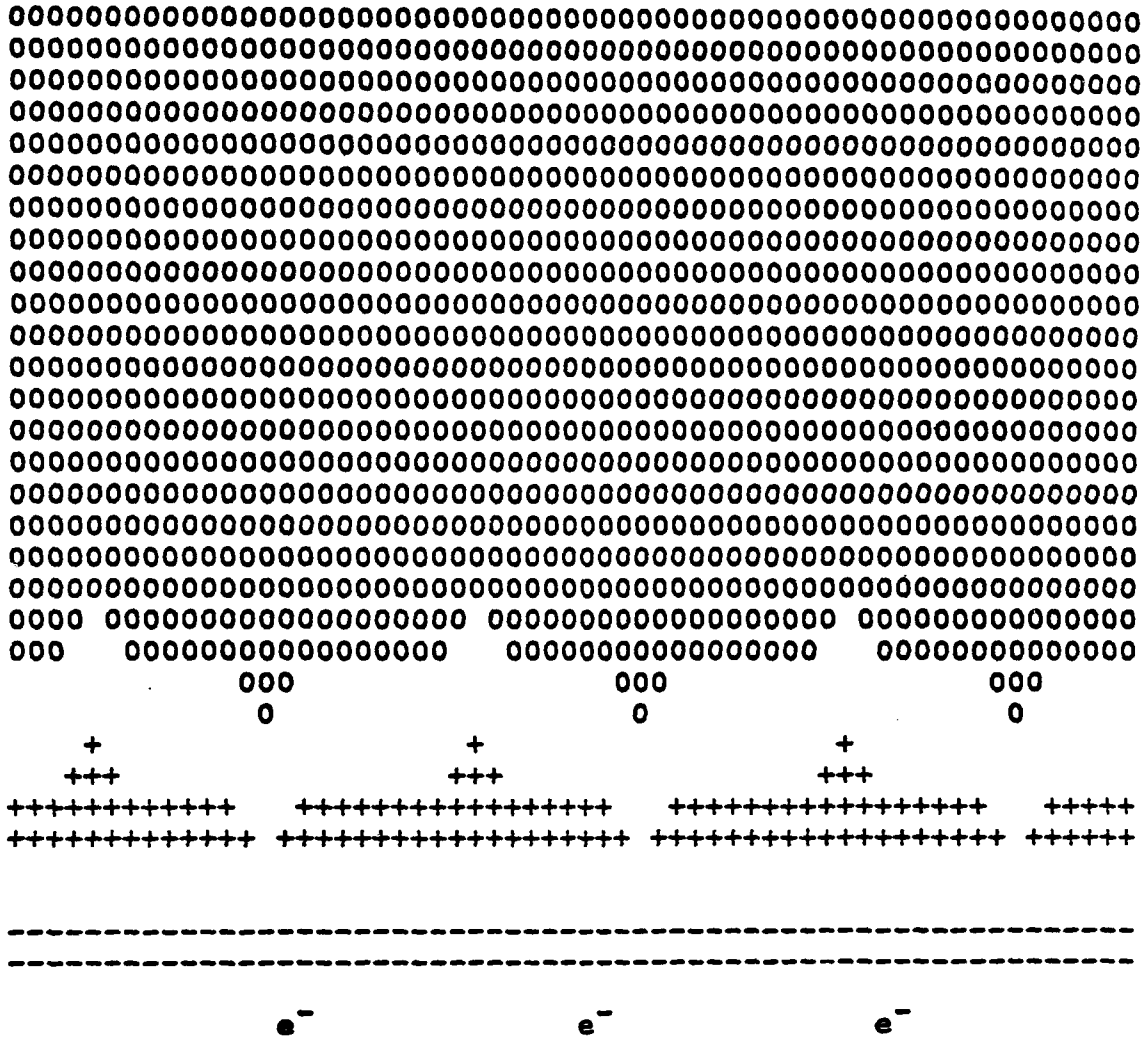


Figure 3

PHOTOSENSITIVITY ANALYSIS
Illumination Profiles

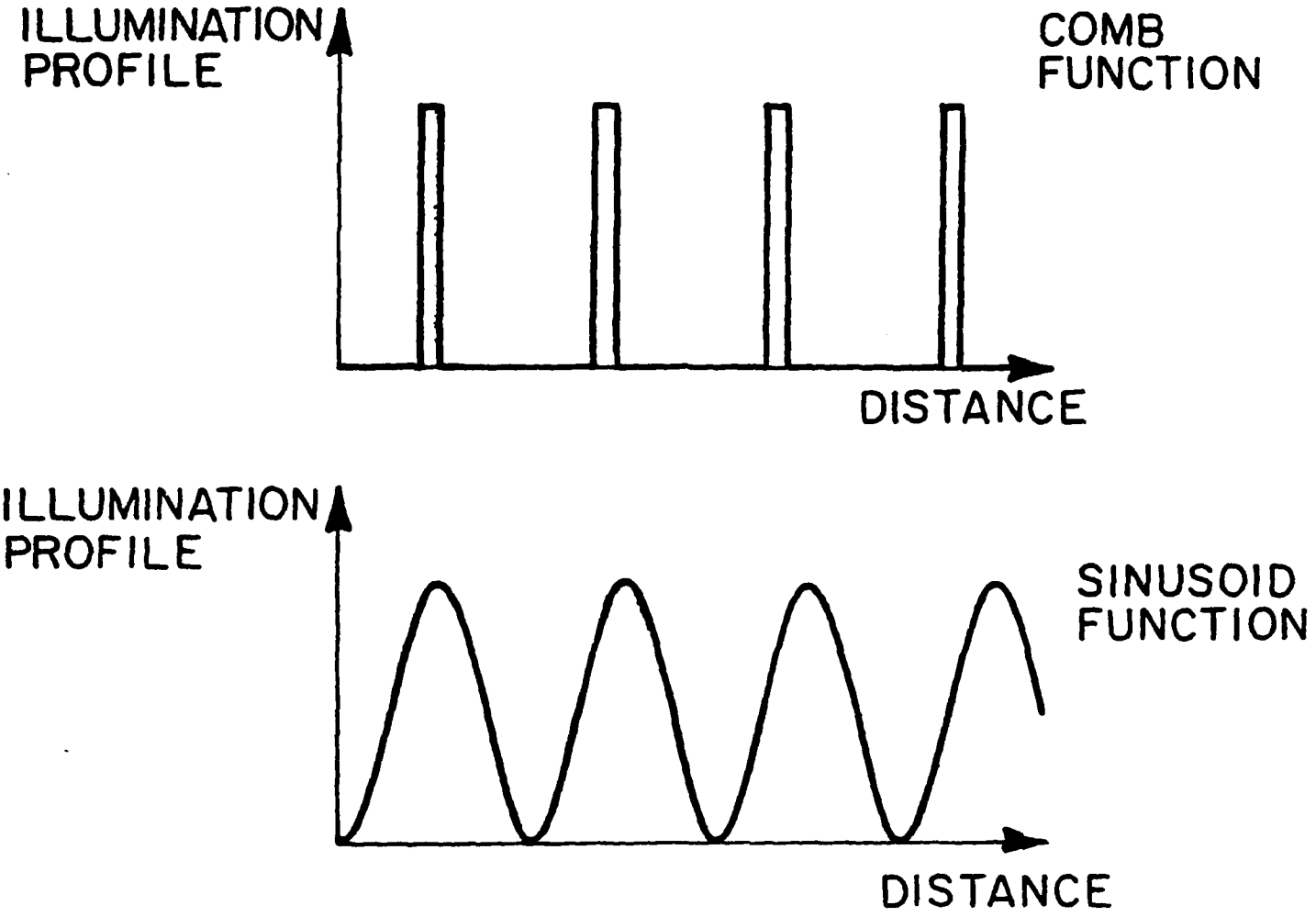


Figure 4

PHOTOSENSITIVITY ANALYSIS
Electron Transport

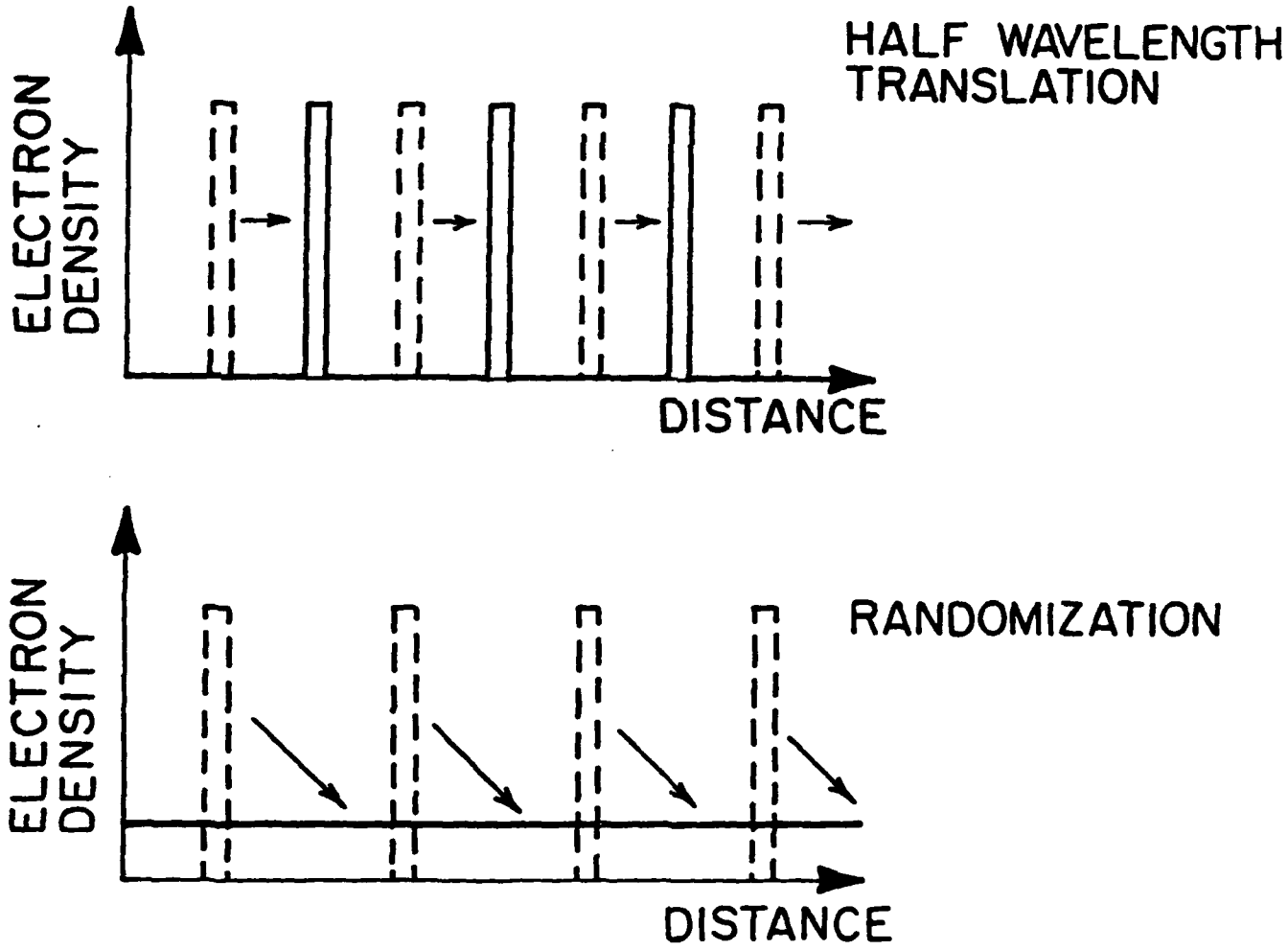


Figure 5

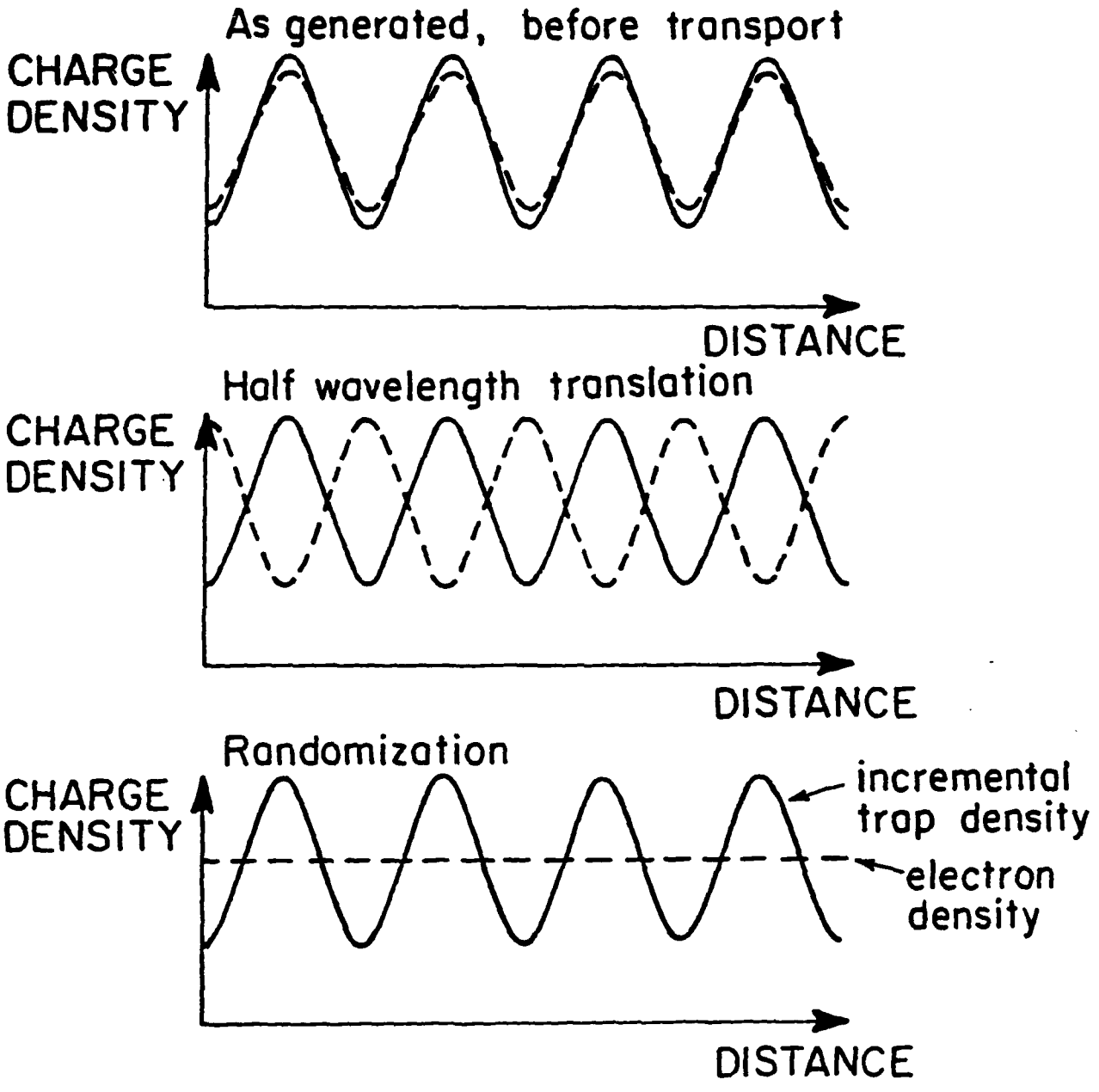

ELECTRON TRANSPORT ASSUMPTIONS
FOR SINUSOIDAL ILLUMINATION
As generated, before transport
CHARGE
DENSITY
DISTANCE
Half wavelength translation
CHARGE
DENSITY
DISTANCE
Randomization
CHARGE
DENSITY
incremental
trap density
electron
density
DISTANCE

Figure 6

PHOTOSENSITIVITY ANALYSIS CASES
Illumination Profile
Comb
Sinusoid
Monopolar Comb
Baseline Sinusoid
Bipolar Comb
Transport-Efficient Sinusoid
Charge Transport
Randomization
Half Wavelength
Translation

Figure 7

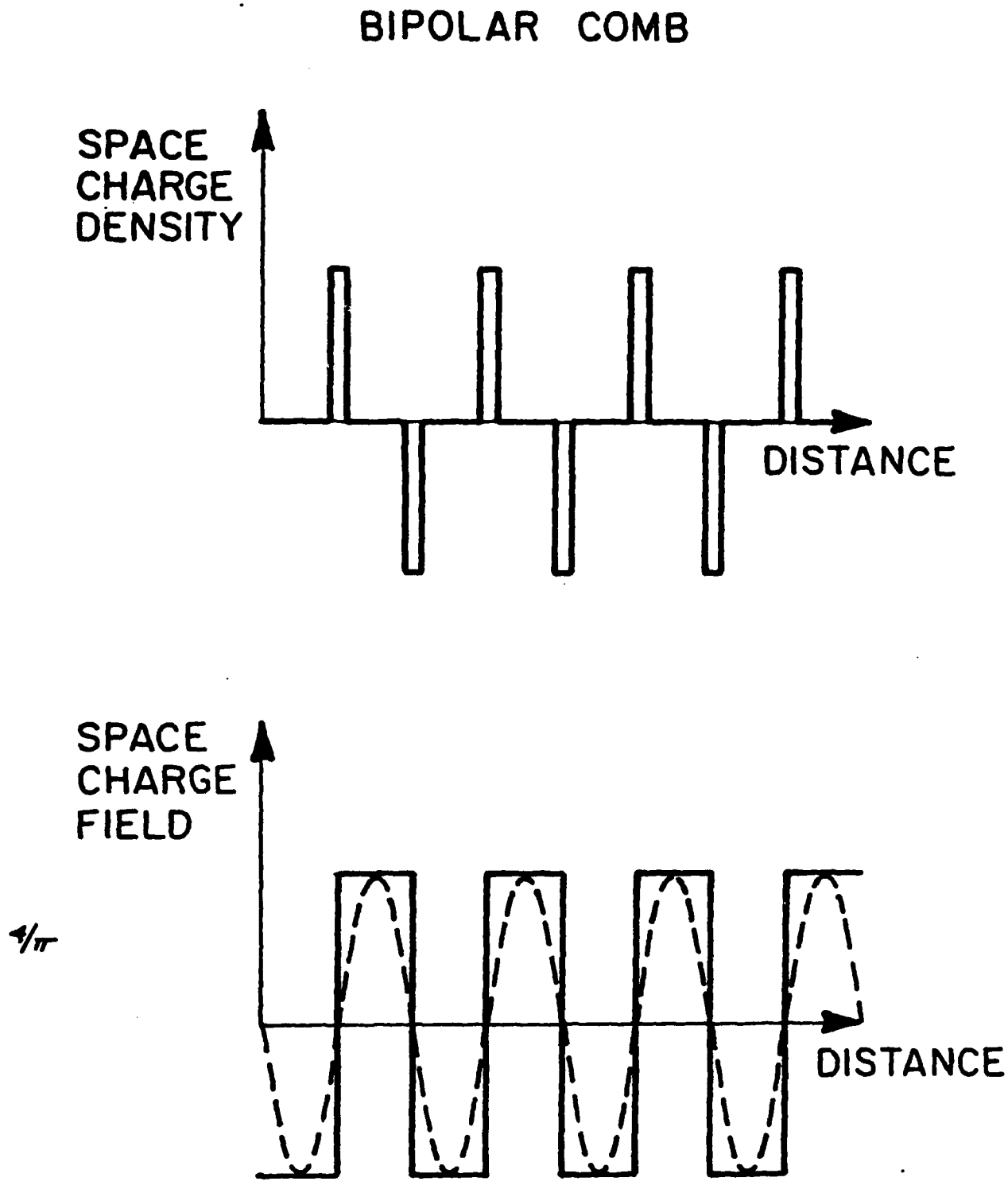

BIPOLAR COMB
SPACE CHARGE DENSITY
DISTANCE
SPACE CHARGE FIELD
4/π
DISTANCE

Figure 8

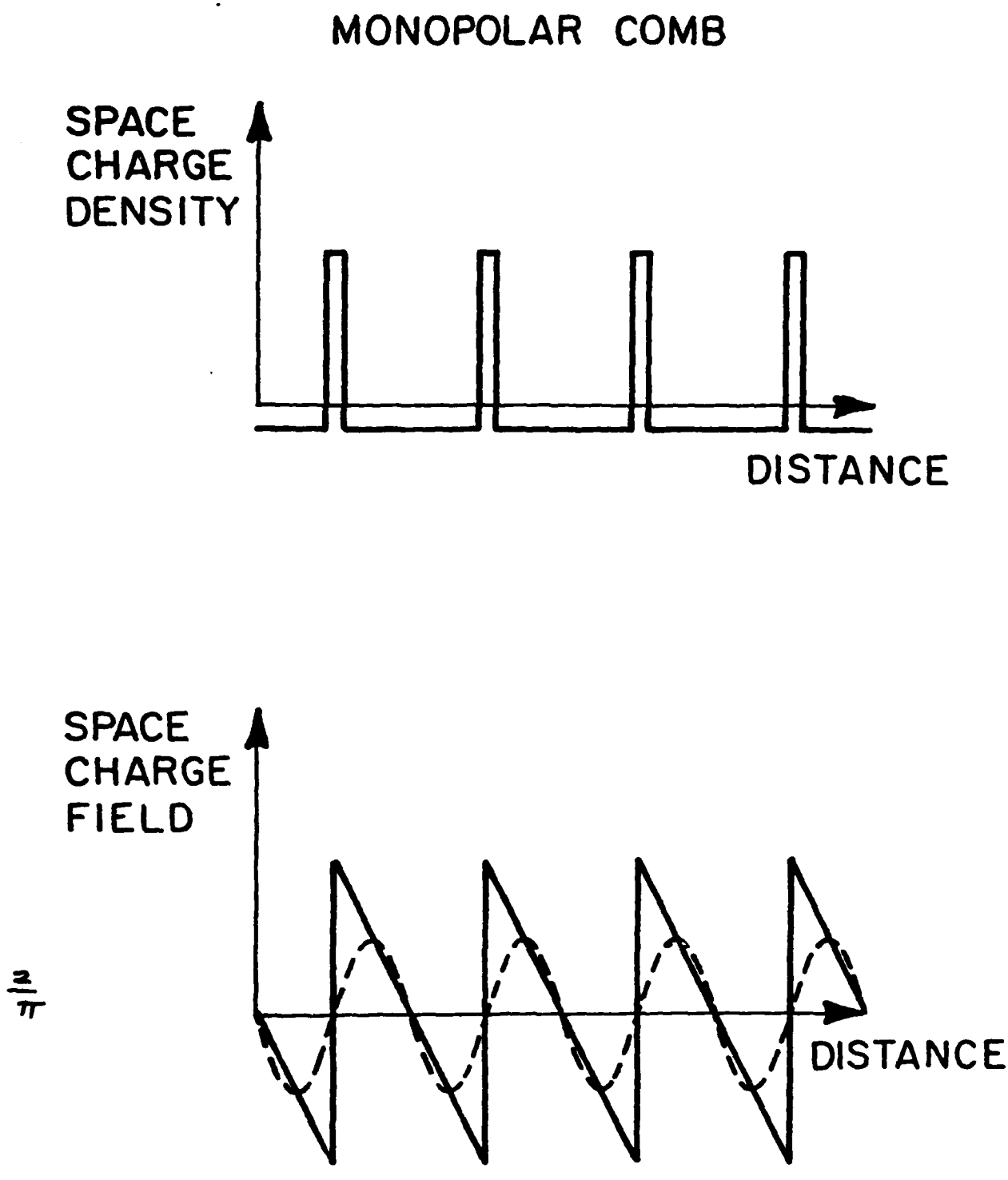

MONOPOLAR COMB
SPACE CHARGE DENSITY
DISTANCE
SPACE CHARGE FIELD
2/π
DISTANCE

Figure 9

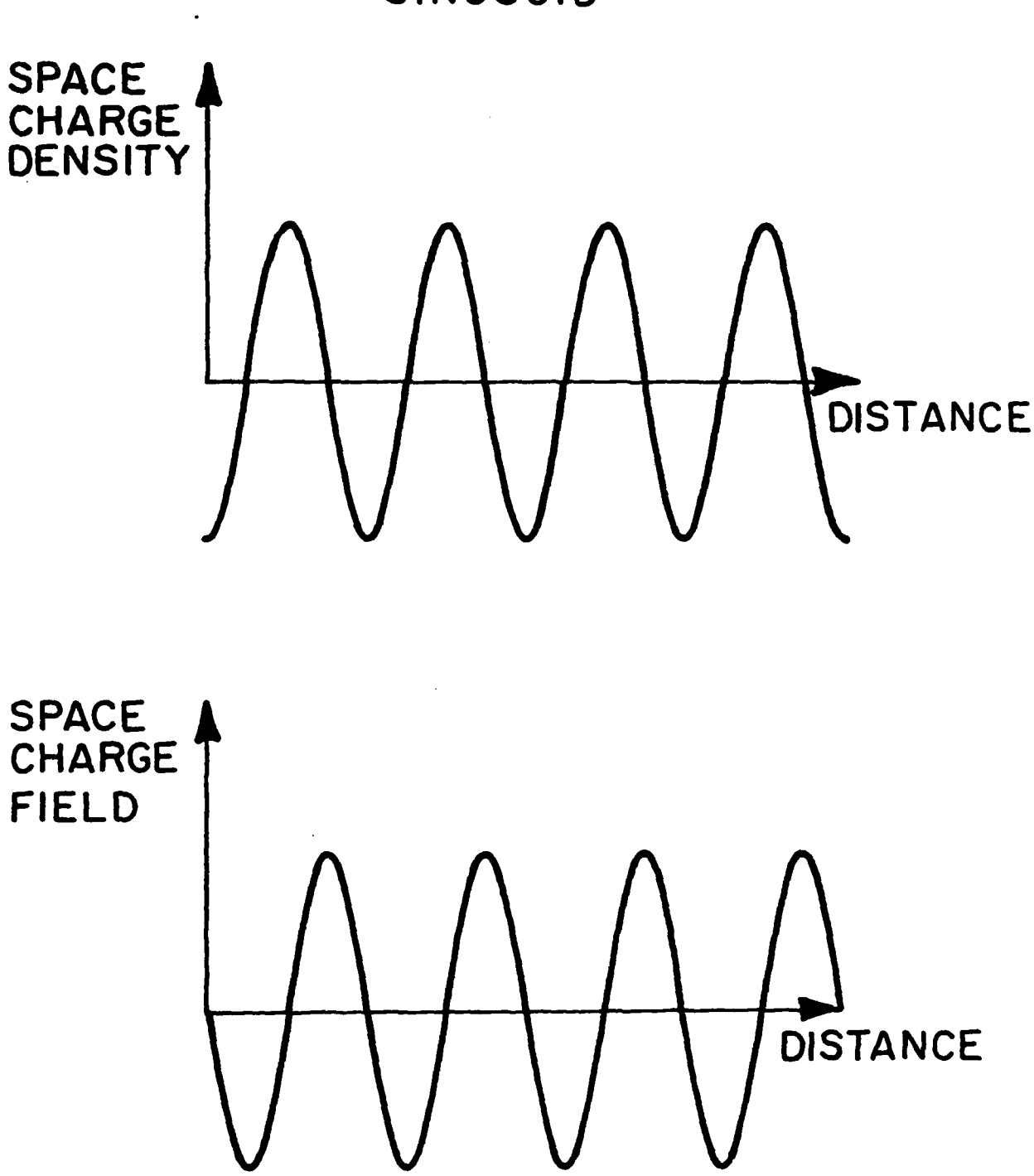

SINUSOID
SPACE CHARGE DENSITY
DISTANCE
SPACE CHARGE FIELD
DISTANCE

Figure 10

SATURATION OF THE SPACE CHARGE FIELD

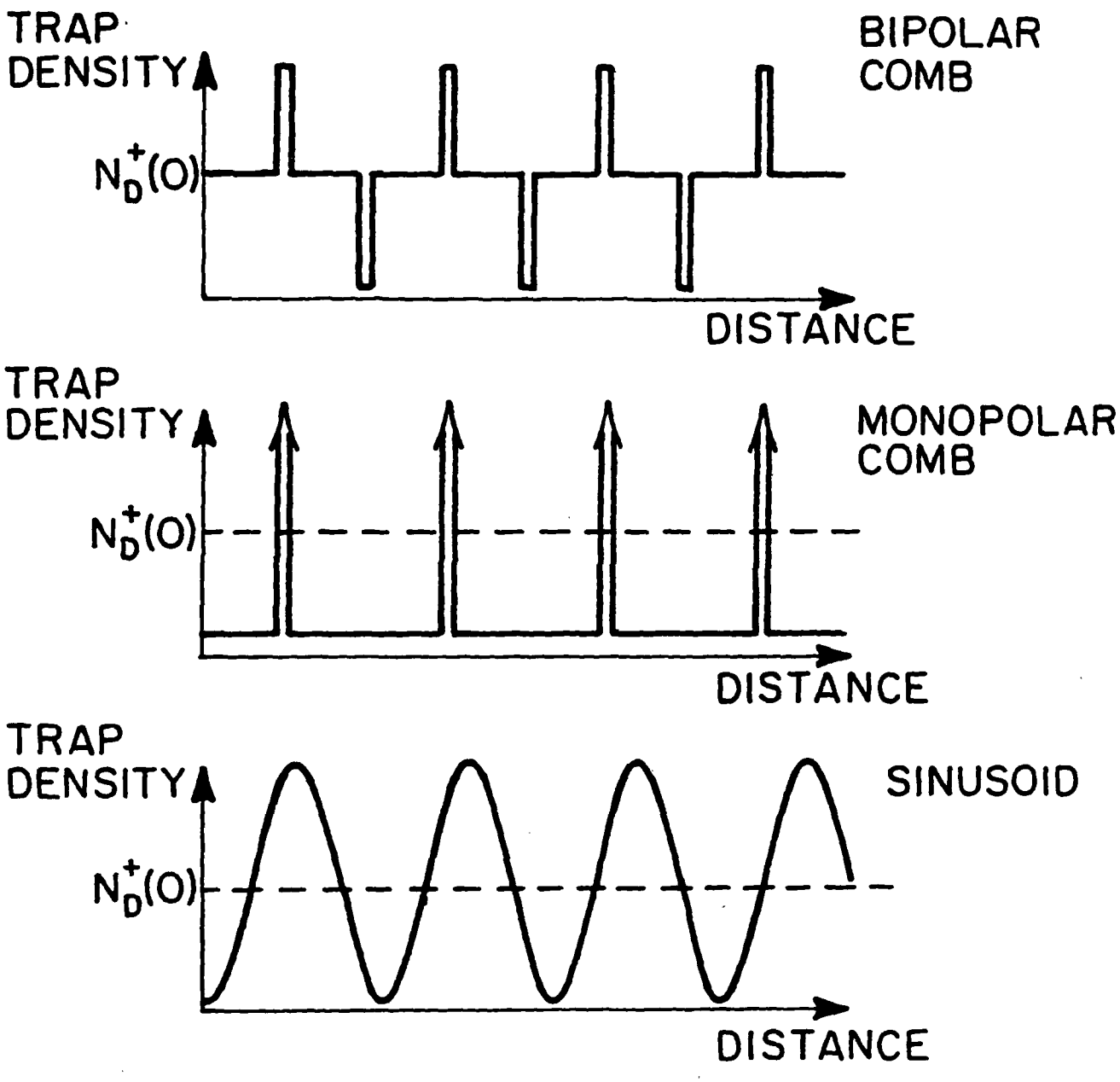


Figure 11

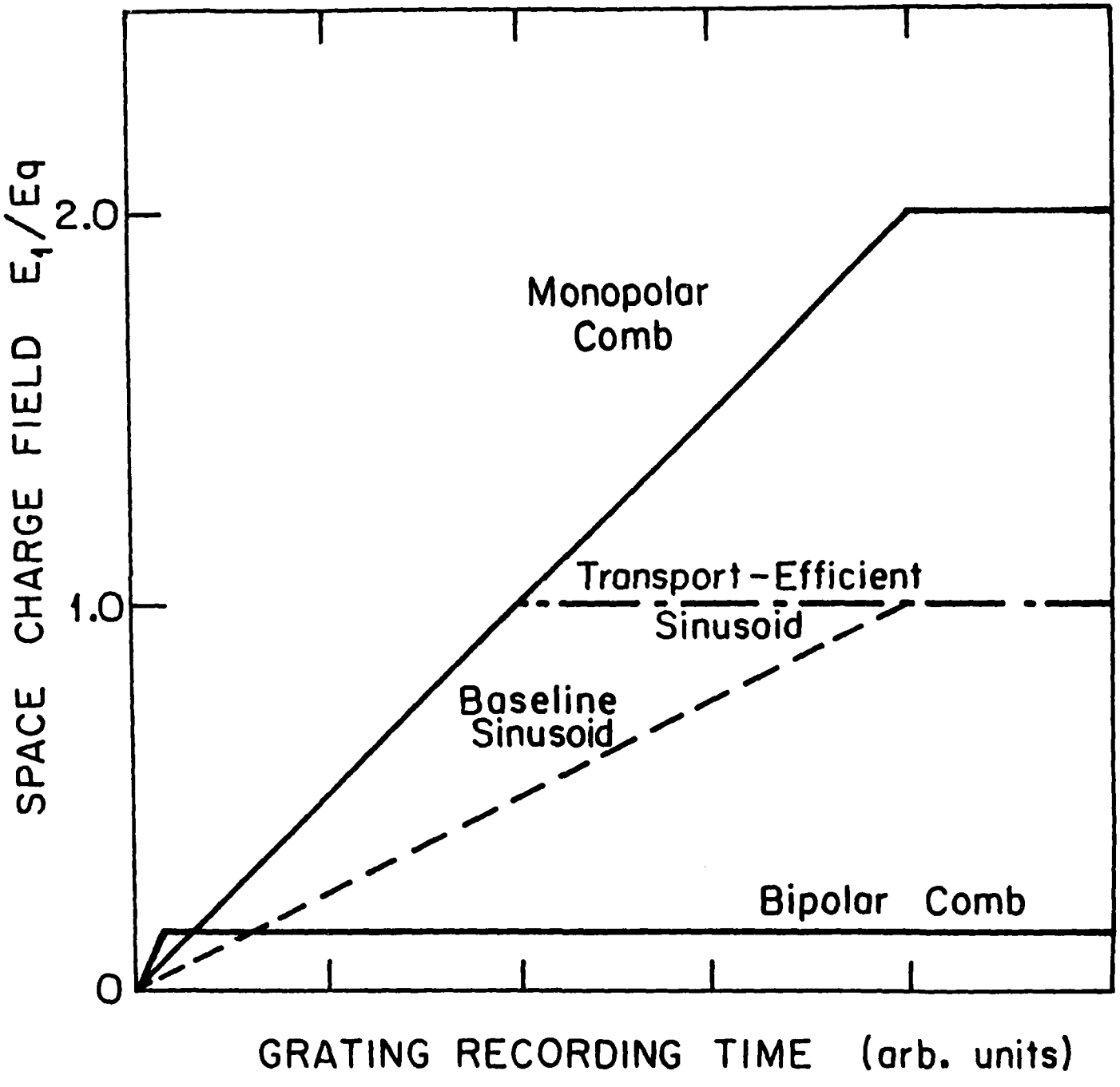

SPACE CHARGE FIELD E_1/E_q
2.0
1.0
0
Monopolar
Comb
Transport-Efficient
Sinusoid
Baseline
Sinusoid
Bipolar Comb
GRATING RECORDING TIME (arb. units)

Figure 12

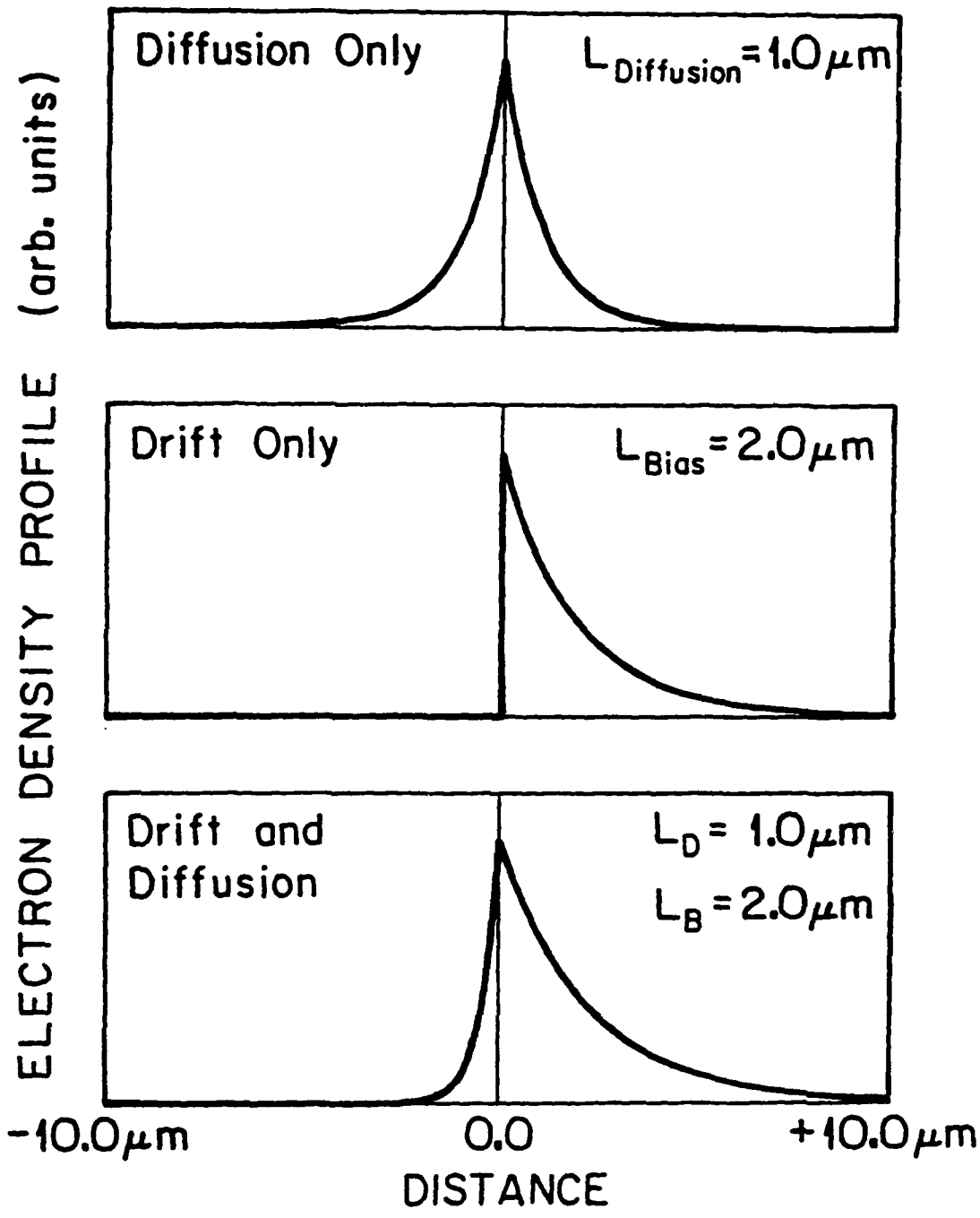

Diffusion Only
$L_{Diffusion}$=1.0μm
Drift Only
L_{Bias}=2.0μm
Drift and
Diffusion
L_D= 1.0μm
L_B=2.0μm
ELECTRON DENSITY PROFILE (arb. units)
-10.0μm
0.0
+10.0μm
DISTANCE

Figure 13

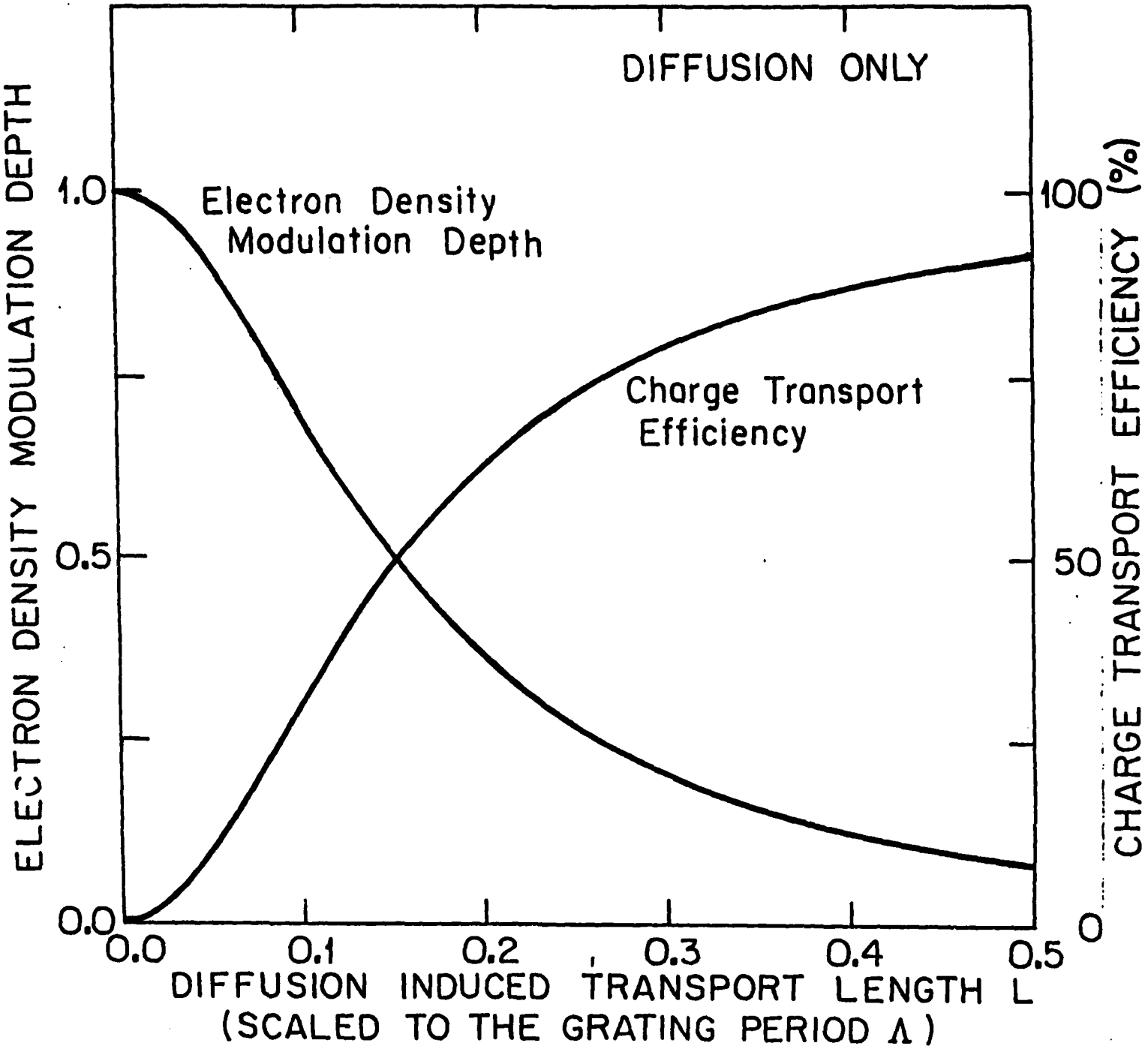

DIFFUSION ONLY
Electron Density
Modulation Depth
Charge Transport
Efficiency
ELECTRON DENSITY MODULATION DEPTH
CHARGE TRANSPORT EFFICIENCY (%)
1.0
0.5
0.0
100
50
0
0.0
0.1
0.2
0.3
0.4
0.5
DIFFUSION INDUCED TRANSPORT LENGTH L
(SCALED TO THE GRATING PERIOD Λ)

Figure 14

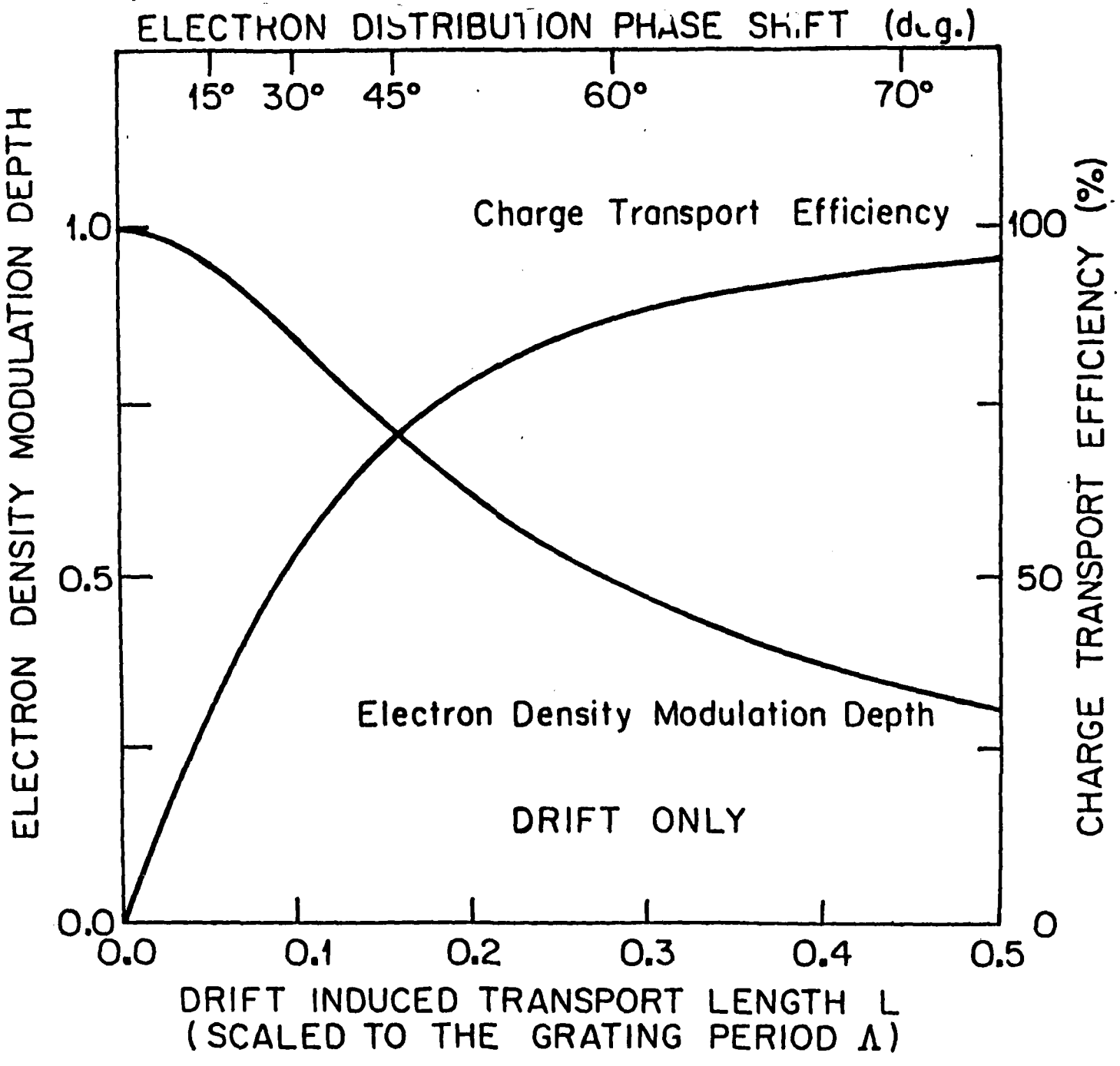

ELECTRON DISTRIBUTION PHASE SHIFT (deg.)
15°
30°
45°
60°
70°
Charge Transport Efficiency
Electron Density Modulation Depth
DRIFT ONLY
ELECTRON DENSITY MODULATION DEPTH
CHARGE TRANSPORT EFFICIENCY (%)
1.0
0.5
0.0
100
50
0
0.0
0.1
0.2
0.3
0.4
0.5
DRIFT INDUCED TRANSPORT LENGTH L
(SCALED TO THE GRATING PERIOD Λ)

Figure 15

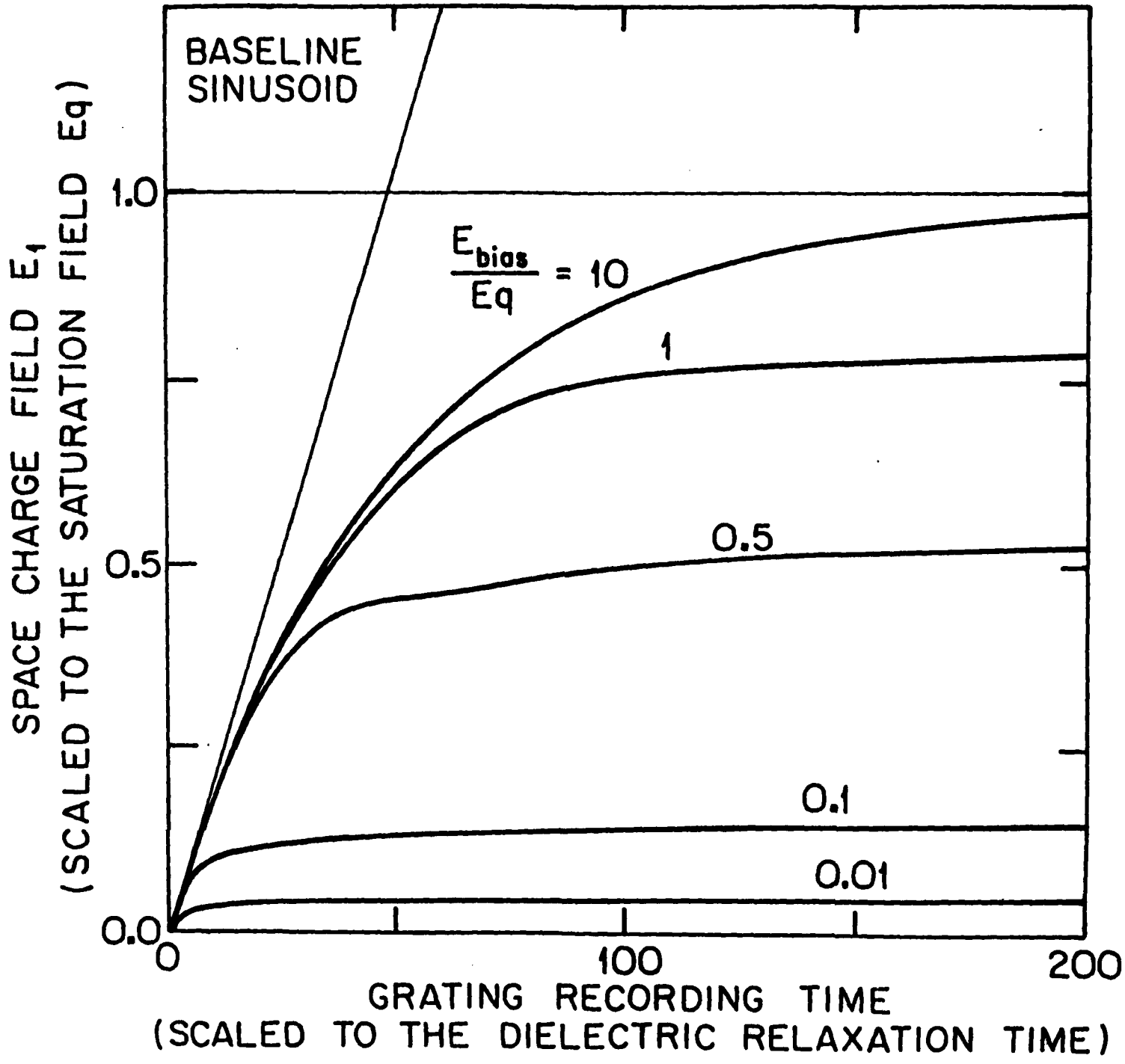

BASELINE
SINUSOID
$\frac{E_{bias}}{Eq} = 10$
1
0.5
0.1
0.01
1.0
0.5
0.0
0
100
200
SPACE CHARGE FIELD E_1
(SCALED TO THE SATURATION FIELD Eq)
GRATING RECORDING TIME
(SCALED TO THE DIELECTRIC RELAXATION TIME)

Figure 16

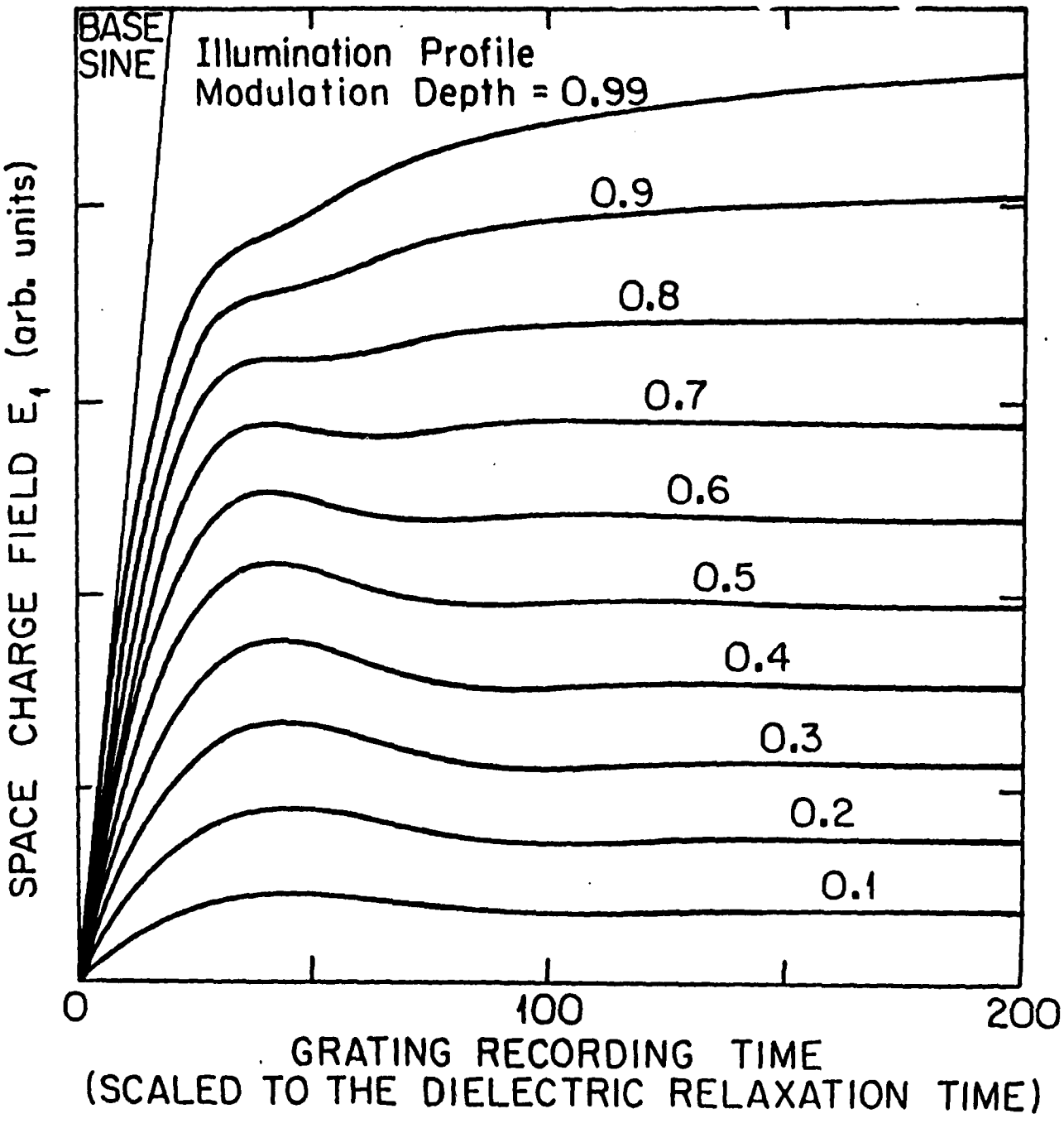

BASE
SINE
Illumination Profile
Modulation Depth = 0.99
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
SPACE CHARGE FIELD E_1 (arb. units)
0
100
200
GRATING RECORDING TIME
(SCALED TO THE DIELECTRIC RELAXATION TIME)

Figure 17

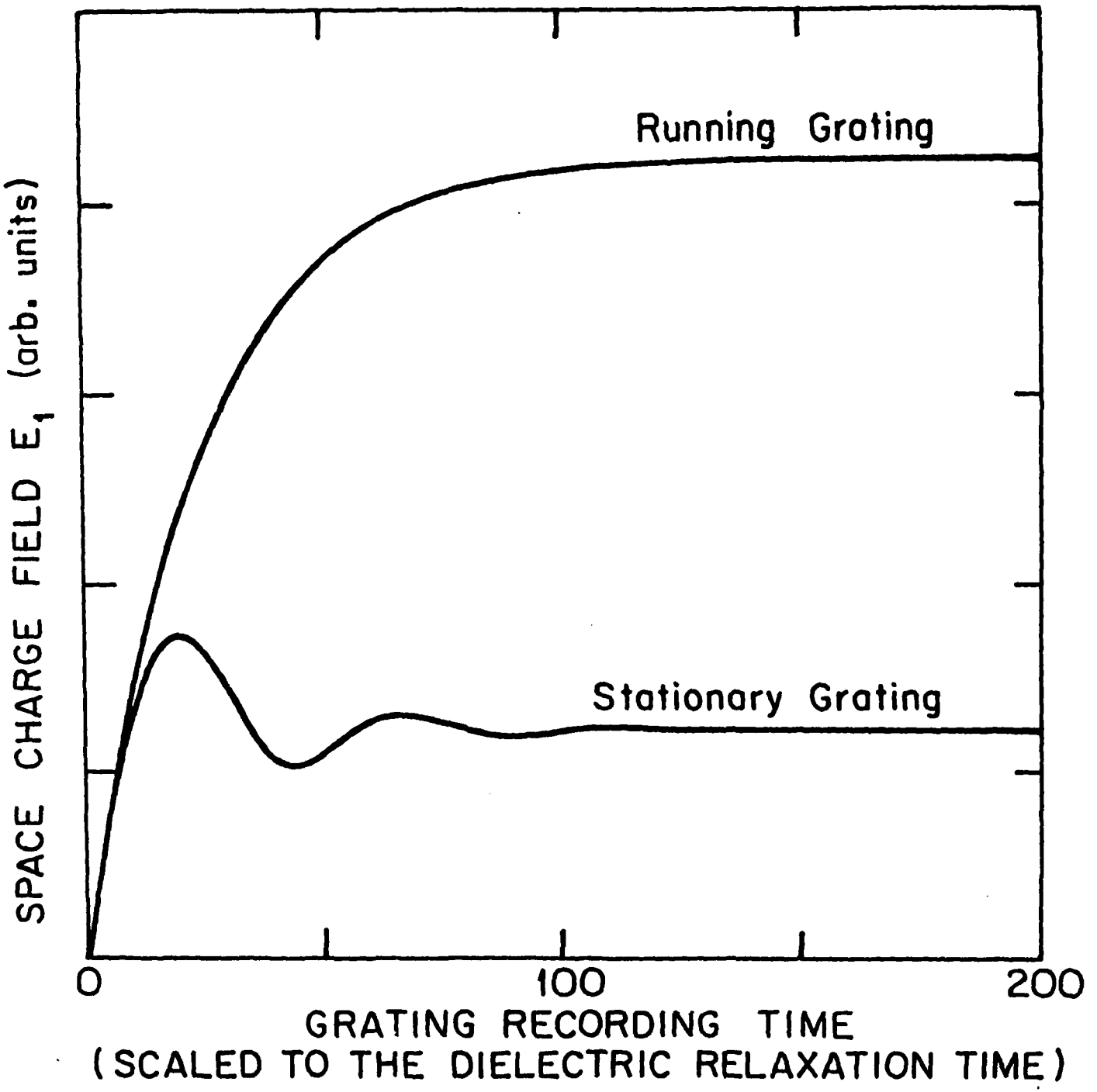

Running Grating
Stationary Grating
SPACE CHARGE FIELD E_1 (arb. units)
0
100
200
GRATING RECORDING TIME
(SCALED TO THE DIELECTRIC RELAXATION TIME)

Figure 18

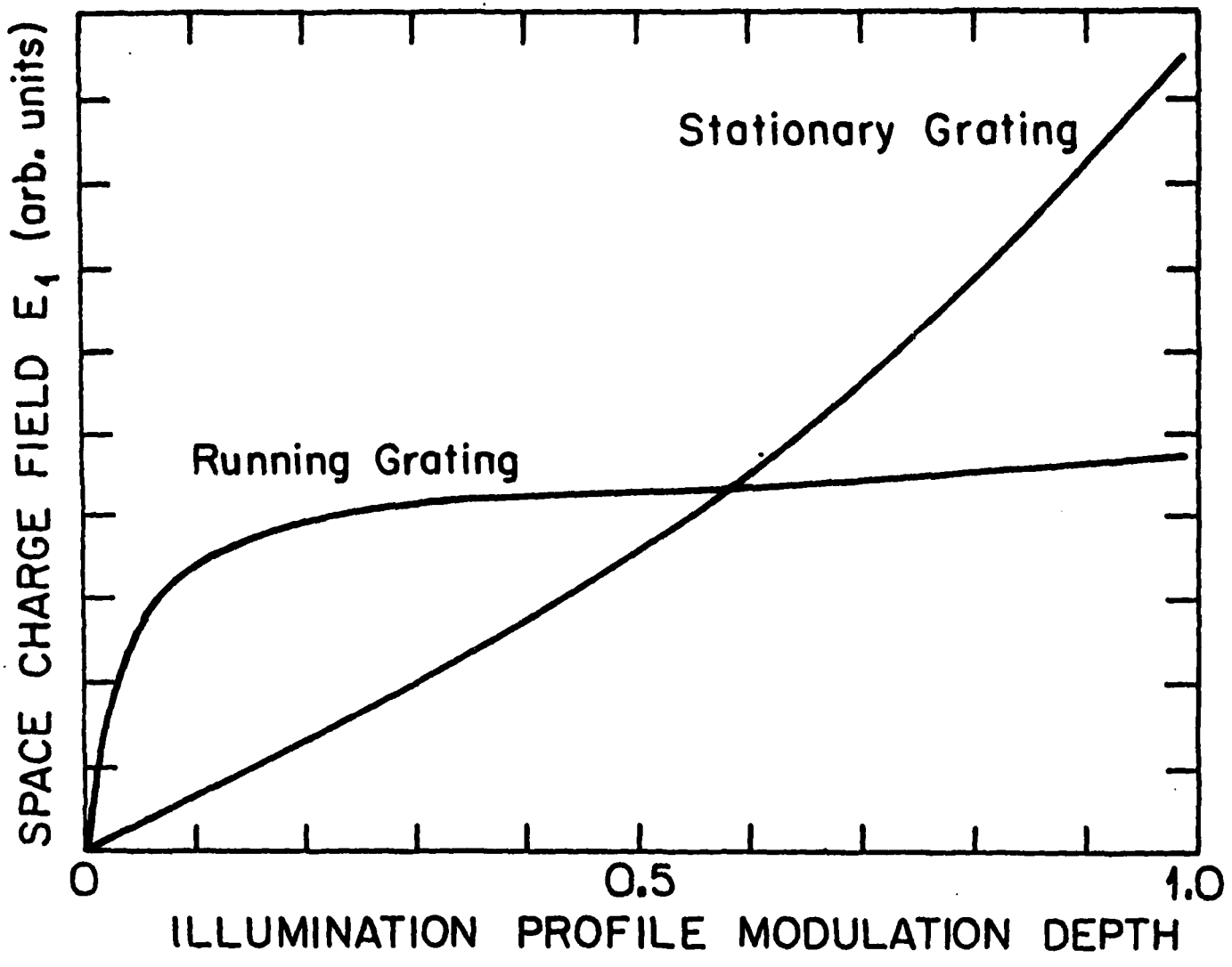

SPACE CHARGE FIELD E_1 (arb. units)
Stationary Grating
Running Grating
0
0.5
1.0
ILLUMINATION PROFILE MODULATION DEPTH

Figure 19