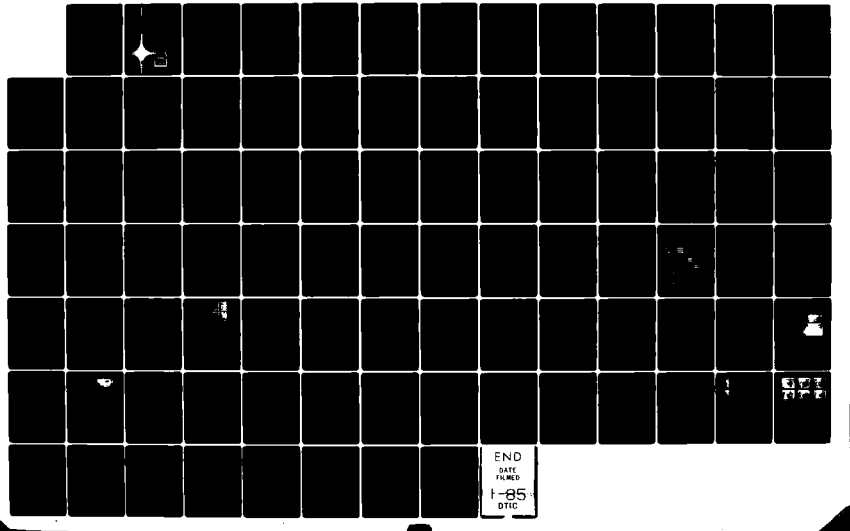
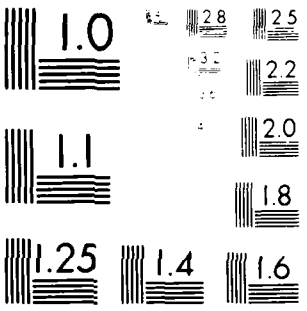
AD-A148 293 NONLINEAR REAL-TIME OPTICAL SIGNAL PROCESSING(U) 1/1
UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES IMAGE
PROCESSIN.. A A SAWCHUK ET AL. 01 OCT 84 USCIPI-1130
UNCLASSIFIED AFOSR-TR-84-1017 AFOSR-81-0182 F/G 20/6 NL




1.0
2.8
2.5
3.2
2.2
1.1
2.0
1.8
1.25
1.4
1.6

USCIPI Report 1130

(14)

UNIVERSITY OF SOUTHERN CALIFORNIA

FINAL TECHNICAL REPORT

April 15, 1981 - June 30, 1984

NONLINEAR REAL-TIME OPTICAL SIGNAL PROCESSING

A.A. Sawchuk, Principal Investigator
T.C. Strand and A.R. Tanguay, Jr.

October 1, 1984

Department of Electrical Engineering
Image Processing Institute
University of Southern California
University Park-MC 0272
Los Angeles, California 90089-0272

Research sponsored by the
Air Force Office of Scientific Research
Electronics and Materials Science Division
under Grant No. AFOSR-81-0182

IPI

IMAGE PROCESSING INSTITUTE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ~~USCIPI Report 1130~~	2. GOVT ACCESSION NO. AD-A145 763	3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)* NONLINEAR REAL-TIME OPTICAL SIGNAL PROCESSING		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report 4/15/81 to 6/30/84
		6. PERFORMING ORG. REPORT NUMBER USCIPI Report 1130
7. AUTHOR(s) A.A. Sawchuk, T.C. Strand, and A.R. Tanguay,Jr.		8. CONTRACT OR GRANT NUMBER(s) AFOSR-81-0182
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Electrical Enginering Image Processing Institute, MC 0272 University of Southern California, LA, CA 90089		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research Bldg. 410, Bolling AFB Washington, D.C. 20332		12. REPORT DATE October 1, 1984
		13. NUMBER OF PAGES 86
14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)* As above		15. SECURITY CLASS. *(of this report)* Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*
The United State Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation hereon.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*
Optical information processing
Nonlinear optical processing
Optical computing
Light valves
Optical digital computing

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This report summarizes the results of a research program in nonlinear real-time optical signal processing. The program began April 15, 1981 and ended June 30, 1984. The research effort has centered on optical sequential logic systems and their use in digital optical computers, and on variable grating mode (VGM) liquid crystal spatial light modulators. As part of this study, parallel and twisted nematic liquid crystal light valve (LCLV) devices have been used as a nonlinear element in a feedback arrangement to implement a binary sequential

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

logic system. A computer generated hologram fabricated on an e-beam system serves as a beamsteering interconnection element. A completely optical oscillator and frequency divider have been experimentally demonstrated, and various circuit interconnection techniques have been explored. Variable-grating mode (VGM) liquid crystal devices that perform local spatial frequency modulation as a function of the incident intensity have also been investigated. These devices can be used for nonlinear processing by selection and recombination of these spatial frequency components. These devices have many interesting physical effects with useful applications in both analog and numerical optical signal processing. Results on the physical modeling of VGM devices are given, with particular emphasis on experimental measurements of the Jones matrix describing polarized light propagation through the VGM cell. The program was performed with the cooperation of the Hughes Research Laboratories of Malibu, California.

FINAL TECHNICAL REPORT

April 15, 1981 - June 30, 1984

NONLINEAR REAL-TIME OPTICAL SIGNAL PROCESSING

A.A. Sawchuk, Principal Investigator
T.C. Strand and A.R. Tanguay, Jr.

October 1, 1984

Department of Electrical Engineering
Image Processing Institute
University of Southern California
University Park-MC 0272
Los Angeles, California 90089-0272

Research sponsored by the
Air Force Office of Scientific Research
Electronics and Materials Science Division
under Grant No. AFOSR-81-0182

USCIPI Report 1130

TABLE OF CONTENTS

ABSTRACT

This report summarizes the results of a research program in nonlinear real-time optical signal processing. The program began April 15, 1981 and ended June 30, 1984. The research effort has centered on optical sequential logic systems and their use in digital optical computers, and on variable grating mode (VGM) liquid crystal spatial light modulators. As part of this study, parallel and twisted nematic liquid crystal light valve (LCLV) devices have been used as a nonlinear element in a feedback arrangement to implement a binary sequential logic system. A computer generated hologram fabricated on an e-beam system serves as a beamsteering interconnection element. A completely optical oscillator and frequency divider have been experimentally demonstrated, and various circuit interconnection techniques have been explored. Variable-grating mode (VGM) liquid crystal devices that perform local spatial frequency modulation as a function of the incident intensity have also been investigated. These devices can be used for nonlinear processing by selection and recombination of these spatial frequency components. These devices have many interesting physical effects with useful applications in both analog and numerical optical signal processing. Results on the physical modeling of VGM devices are given, with particular emphasis on experimental measurements of the Jones matrix describing polarized light propagation through the VGM cell. The program was performed with the cooperation of the Hughes Research Laboratories of Malibu, California.

1. RESEARCH OBJECTIVES AND PROGRESS

1.1 Introduction and Project Overview

This report summarizes the final results of a research effort in performing nonlinear operations in optical signal processing and achieving operation in real time using various input transducers. The project began April 15, 1981 and ended June 30, 1984. This section contains an introduction, motivation for the work and an overview of the research program.

The research described in this report addresses the need for signal processing systems that can perform high throughput parallel multi-dimensional operations on signals with large time-bandwidth and space-bandwidth products. In many of these applications, digital hardware is inadequate. As part of this project, modifications have been made to existing liquid crystal light valve (LCLV) real-time spatial light modulators and new devices such as the variable grating mode (VGM) light valve have been developed. Another goal of this research has been to explore numerical optical computing using binary or residue arithmetic. In these systems, signals exist as discrete levels rather than as analog signals. This new approach holds much promise for the future if real-time processing speed, accuracy, and flexibility can be maintained.

A book chapter [1] written by research personnel supported under this grant has been recently completed summarizing the

state-of-the-art of all techniques of nonlinear optical processing. This chapter contains extensive references. Three major techniques of optically implementing nonlinear point functions have been developed. They are: halftoning; direct nonlinear processing using the inherent characteristics of image detectors and transducers; and intensity-to-spatial frequency conversion. Application examples and real-time implementation of these techniques are described in the review papers.

In the research phase summarized in this report we have concentrated on: optical sequential logic systems that directly rely on the input-output characteristics of LCLV devices; and on variable grating mode (VGM) devices and their applications.

Nonlinear optical functions can be achieved directly using the inherent transfer characteristics of an optical recording medium or real-time image transducer. With this type of nonlinear processing, there is no pulse-width modulation, intensity-to-spatial frequency conversion or other type of intermediate mechanism. Thus, these techniques offer the potential of simple systems that avoid the noise problems associated with many optical filtering techniques and have much less stringent space-bandwidth product requirements than systems which must modulate the input data. Such systems can implement parallel combinatorial logic and, with the addition of feedback, parallel sequential logic. Section 1.2 of this report describes recent results on these subjects and work on the general area of

digital optical computing.

Another convenient method of obtaining point nonlinearities is through intensity-to-spatial frequency conversion. The idea is to encode each resolution element of an image with a grating structure where the period and/or the orientation of the grating is a function of the image intensity at the point in question. Assuming certain sampling requirements are met, each intensity level of interest is uniquely assigned to a different point in Fourier space and all points with a given intensity in the image are assigned to the same point in Fourier space (assuming space-invariant operation is desired). Then a pure amplitude spatial filter can alter the relative intensity levels in an arbitrary way, and combination of the filtered components produces various nonlinear functions. Both continuous-level (analog) nonlinear functions and various numerical logic functions (binary or residue) are possible. This method relies on the behavior of variable-grating mode (VGM) liquid crystal real-time devices which have been developed under this AFOSR program. Section 1.3 of this report describes work on physical modeling and measurements of VGM liquid crystal devices. The goal of this work is to improve their temporal response, uniformity, lifetime, etc. Several new types of electrically and optically activated VGM devices have been constructed and evaluated.

This program has been very productive; a number of oral

presentations have been made and many written papers have been published describing research results. Some of the most significant of these papers are reprinted as part of this report.

The overall program has been a joint cooperative effort between the University of Southern California (USC) group and the Hughes Research Laboratories (HRL) in Malibu, California. Each group has participated in the project together and a separate progress report is being submitted by HRL as a comparison to this report. Both groups have worked closely together in their particular areas of expertise toward the project goals.

1.2 Digital Optical Computing

1.2.1 Introduction

There has been considerable work in recent years in developing optical systems that perform essentially digital processing functions. The reasons for this interest include extending the flexibility of optical processing systems and the possibility of using the parallel capabilities of optical systems for digital signal processing. The first steps in this digi optical computing research have included parallel A/D conversio and optical combinatorial logic implementation. Both of these have been demonstrated in real-time systems at USC [1]-[3].

The next step in this progression of experiments is to demonstrate the feasibility of optical sequential logic. Here the basic logic gates are interconnected in a circuit which generally includes some form of feedback. In this system the temporal response characteristics of the system become very important. We have developed an optical system which demonstrates the feasibility of optical sequential logic. In particular we have implemented a totally optical system which includes a clock driving a master-slave flip-flop. The basic elements of a sequential logic system are a nonlinear element that performs the desired logic function and an interconnection system to route the outputs of the nonlinear device to the appropriate inputs.

Four recent papers that contain the details of this work are reprinted here. The first paper is "Sequential Optical Logic Implementation", by B.K. Jenkins, A.A. Sawchuk, T.C. Strand, R. Forchheimer and B.H. Soffer. This paper recently appeared in Applied Optics and concentrates on experimental results from the sixteen gate clocked master-slave optical flip-flop.

A second paper "Architectural Implications of Digital Optical Processors" by P. Chavel, R. Forchheimer, B.K. Jenkins, A.A. Sawchuk and T.C. Strand, has just been published in Applied Optics. It describes various interconnection techniques for optical sequential logic systems, including space-variant, space-invariant and hybrid computer-generated holograms.

A third paper, "Computer-Generated Holograms for Space-Variant Interconnections in Optical Logic Systems," by B.K. Jenkins and T.C. Strand was published in SPIE Proceedings vol. 437 for the International Conference on Computer-Generated Holography held in August 1983 in San Diego. This paper describes details of space-variant computer hologram interconnection schemes for optical logic and computing systems.

Another paper, "Digital Optical Computing," by A.A. Sawchuk and T.C. Strand appeared in the July 1984 Proceedings IEEE Special Issue on Optical Computing. This paper is a comprehensive overview of binary combinatorial and sequential logic with individual devices and arrays of devices. The paper discusses communication, interconnection and input-output

problems of digital electronic computers at the gate, chip and processor level. Some architectural techniques for avoiding some of the interconnection problems of electronic VLSI are presented. The possibility of non Von Neumann parallel digital processors, limitations and future needs of optical logic devices and digital optical computing systems are discussed.

Sequential optical logic implementation

B. K. Jenkins, A. A. Sawchuk, T. C. Strand, R. Forchheimer, and B. H. Soffer

An optical system that performs sequential binary logic operations is described. The system consists of a spatial light modulator (SLM) used to provide a nonlinear threshold response and a computer-generated hologram to provide interconnections between logic gates. A 2-D array of logic gates with binary inputs and outputs is formed on the active surface of the SLM. These gates are interconnected by a 2-D array of subholograms, one for each gate. Arbitrary logic circuits consisting of NOR gates and inverters can be implemented, and the system can be reconfigured by changing a single holographic element. The system is demonstrated using a twisted-nematic liquid crystal light valve as the SLM. A test circuit is implemented that includes a synchronous master–slave flip-flop and an oscillator consisting of five inverters in a feedback loop. Experimental results of this test circuit are presented.

I. Introduction

The vast majority of optical processing systems to date have operated with analog signal levels. While many of these systems can perform specific operations with extremely high throughput rates, at the same time they suffer from two basic limitations: the variety of operations that can be performed and the accuracy of the results. These limitations preclude the use of optical processing systems in certain application areas that could otherwise benefit from some of the inherent advantages of optics. These advantages include a high degree of parallelism, both in processing and input/output, and a high density and number of interconnections. Two approaches have been taken to eliminate or at least substantially reduce these limitations. Both approaches utilize discrete, instead of analog, signal levels. One approach is based on residue arithmetic operations, and optical systems utilizing this principle have been studied.[1-7] The other approach is based on binary logic operations, such as used in conventional electronic computers.

The first step in the binary approach has been to demonstrate Boolean operations optically. Two-dimensional arrays of some of these operations such as OR, NOR, AND, NAND, XOR, and XNOR have been demonstrated using a variety of optically addressed 2-D spatial light modulators (SLMs). These include the Pockels readout optical modulator (PROM),[8,9] the microchannel spatial light modulator (MSLM),[10,11] the Hughes liquid crystal light valve (LCLV),[12-14] and a segmented liquid crystal light valve used as an optical parallel logic (OPAL) device.[15,16] These operations have also been demonstrated using light-emitting diodes with optical masks.[17] Other schemes, too numerous to mention, have also been used to implement logic gates optically. A review of optical computing systems, including these combinatorial operations, is given in Ref. 18. Recently, individual all-optical logic gates have been demonstrated in InSb (Ref. 19) and GaAs.[20]

Although these operations form the basic building blocks for combinatorial logic, to build a sequential circuit or an optical computer, memory is also needed. This can be achieved using optical feedback. An array of optical flip-flops has been demonstrated using an LCLV in an optical system with feedback,[21,22] and an array of optical latches has been demonstrated by using an OPAL device with optical feedback.[23] Fast individual bistable elements have also been demonstrated in many materials using feedback by means of a Fabry-Perot cavity.[24]

These elements, however, have not been combined to form an all-optical logic circuit. Such a circuit may have the potential of combining the high degree of parallelism and interconnection density found in optical processors with the flexibility and accuracy of digital electronic computers. A method for combining these binary elements using fiber optics has been described,[25] but a complete system has not been demonstrated ex-

B. H. Soffer is with Hughes Research Laboratories, Malibu, California 90265. *When this work was done the other authors were with University of Southern California, Department of Electrical Engineering, Los Angeles, California 90089; T. C. Strand is now with IBM Corp., San Jose, California 95193 and R. Forchheimer is now with Linkoping University, Linkoping, Sweden.

Received 6 January 1984.
0003-6935/84/193455-10$02.00/0.

perimentally. Other digital optical processing systems have also been described.[26,27] A system with an array of optical gates and some optical memories, under electronic control, has been demonstrated.[28] This system is particularly useful for operations requiring only local communication, as is the case for many image processing operations, e.g., cellular logic machines.

In this paper a system is presented which interconnects optical logic elements to form a sequential logic circuit. Every signal in this system is represented optically, and at the same time the system permits implementation of arbitrary connections between the individual logic gates. Because this system is digital, the accuracy limitation mentioned above can be overcome by selecting the number of bits per data element to yield the desired accuracy. In addition, the arbitrary interconnections of the system presented here permit implementation of a very large variety of processing operations. Finally, the advantages of parallelism and interconnection density are retained to a large degree.[29]

In the system presented here, an SLM is used as a 2-D array of independently acting logic gates. These gates are interconnected via an optical system that utilizes a computer-generated hologram. In this paper we discuss the SLM implementation of logic gates in Sec. II, the interconnection system in Sec. III, and some experimental results of the implementation of a test circuit to demonstrate the feasibility of the optical logic system in Secs. IV and V. Finally, in Sec. VI we point out some of the fundamental limitations involved in using optical devices as arrays of logic gates and discuss the relevance of these limitations to the system presented here.

II. Logic Gates

Binary logic gates may be implemented optically through the use of a point nonlinearity. The general scheme used for each gate in our system is depicted in Fig. 1. First, the two binary input lines to the gate are added to yield a single three-level signal. The value of this three-level signal [or $(N + 1)$ − level signal for the case of N-input gates] is then equal to the number of input lines that are true (i.e., that have a value of 1). This signal is then operated on by a nonlinear function with a binary-valued output. As will be shown below, any logic operation can be performed in this manner by choosing the appropriate nonlinear function. Figure 2 shows the extension of this method to N-input gates. To implement this scheme optically, the binary values are represented by intensity levels with a high intensity level representing a 1 and a low intensity level representing a 0. The addition is done merely by optically superimposing the input line signals. With a detector that integrates the signal over the input spot, this has the effect of adding the intensity levels regardless of the coherence properties of the light (see Sec. III). In this paper, the term gate input will refer to this superimposed signal (i.e., the input to the nonlinearity), and the term input lines will refer to the binary inputs before superposition.

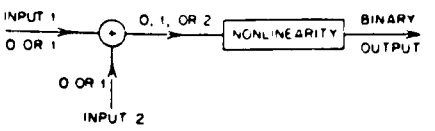


Fig. 1. Implementation of a logic gate.

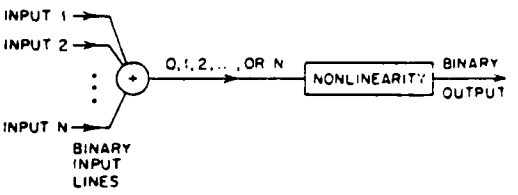


Fig. 2. Generalization of Fig. 1 to N-input gates.

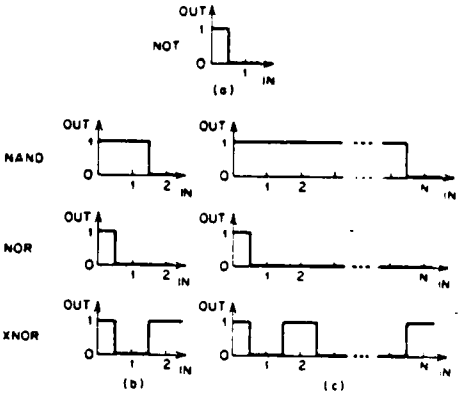


Fig. 3. Nonlinearities for implementation of logic operations: (a) one-input gates; (b) two-input gates; (c) N-input gates. (N is assumed to be even in the XNOR case.)

Possible choices of the nonlinear functions for some of the common logic operations are shown in Fig. 3. In the case of two-input gates, there are a total of sixteen possible operations. The operations AND, NAND, OR, NOR, XOR, XNOR, TRUE, and FALSE may be implemented with this scheme directly [Table I(a)]. The remaining logic functions require the ability to distinguish between the two input lines A and B. These operations are A, B, $\bar{A}$, $\bar{B}$, $\bar{A} \cdot B$, $A \cdot \bar{B}$, $\bar{A} + B$, and $A + \bar{B}$. These operations can be implemented with a single gate by doubling the signal level of one of the input lines, say, A [Table I(b)]; this is conceptually equivalent to using a three-input gate with A going into two of the input lines.

However, several subsets of the operations which are realizable with a single gate form logically complete sets, obviating the need for these asymmetric functions. For example, all logic operations can be built out of NOR gates.

The nonlinearity required for optical logic may be implemented with any of a variety of optical devices. Fast switching times ($\lesssim$1 nsec) may be obtained by using a bistable optical device.[19,20,24,30] This prospect is discussed in Sec. VI. While in principle these devices can be built as 2-D spatial light modulators, at present

Table I. Desired Values of the Nonlinearity for the 16 Possible Logic Operations on Two Binary Inputs

Input lines A	B	Gate input $(A+B)$	Outputs AND	NAND	OR	NOR	XOR	XNOR	TRUE	FALSE
0	0	0	0	1	0	1	0	1	1	0
0	1	1	0	1	1	0	1	0	1	0
1	0	1	0	1	1	0	1	0	1	0
1	1	2	1	0	1	0	0	1	1	0

(a)

Input lines A	B	Gate input $(2A+B)$	A	B	$\bar{A}$	$\bar{B}$	Outputs $A\cdot\bar{B}$	$\bar{A}\cdot B$	$A+\bar{B}$	$\bar{A}+B$
0	0	0	0	0	1	1	0	0	1	1
0	1	1	0	1	1	0	0	1	0	1
1	0	2	1	0	0	1	1	0	1	0
1	1	3	1	1	0	0	0	0	1	1

(b)

Note: Also listed are the gate inputs. The inputs to the adder are A and B in (a) and are $2A$ and B in (b).

they are not available. Other optically addressed SLMs are presently available. While their characteristically slow response times may appear to be a major drawback, they provide a practical means for demonstrating the system concept. The same sequential optical logic system described and demonstrated here may be used with other SLMs, including much faster ones, so long as they can provide a suitable nonlinearity, and satisfy the following requirements. First, the inputs and outputs of the gates must be the same wavelength. This eliminates such devices as the PROM,[9] with which the write illumination must be a different wavelength than the read illumination. Second, we assume the device has no memory, although in some cases devices with memory could also be used. Finally, we assume here that the gate inputs and outputs appear on opposite sides of the SLM, although it is worth noting that with a slight rearrangement of the system components, SLMs with gate inputs and outputs on the same side can also be accommodated.

Given these requirements, there are still a variety of SLMs that can be used. The microchannel spatial light modulator[11] could be used in this system, as could various types of LCLV.[12-14] For the demonstration of the system presented in this paper, a Hughes LCLV was used. An LCLV with liquid crystal molecules exhibiting a uniform parallel alignment in the off state may be used to implement nonmonotonic nonlinearities, which are needed for such operations as XOR.[14] In this system, NOR gates were used, and LCLVs with liquid crystal molecules in a variety of configurations can provide the appropriate nonlinearity. For example, Fig. 4 shows the steady-state response for an LCLV with a twisted nematic liquid crystal layer biased to implement the NOR operation. The horizontal and vertical axes in Fig. 4, respectively, represent relative writing and reading intensities, and these quantities are nonnegative. For clarity, physical units are omitted in Fig. 4. The 0 and 1 labels on the axes of Fig. 4 denote writing and reading signal levels corresponding to various binary logic levels.

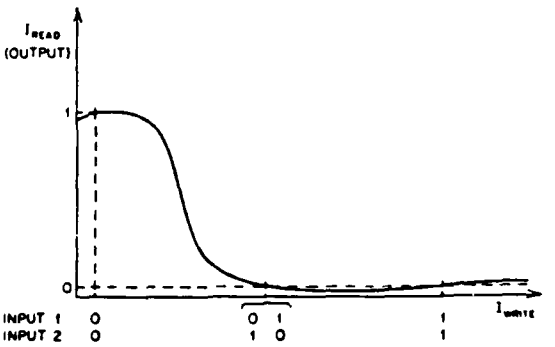


Fig. 4. LCLV imput–output characteristic used to implement the NOR operation, expressed as a function of relative writing and reading intensities. (Physical units are omitted for clarity.) The 0 and 1 labels denote signal levels corresponding to various binary logic levels.

The input/output characteristic of an SLM, when used in this system, essentially serves as an approximation to the corresponding ideal nonlinearity (e.g., Fig. 3). An important criterion in determining whether a particular nonlinearity is a sufficiently accurate approximation is the regeneration or restandardization of the signal level at each pass through a gate. The general requirement is that a signal should not degenerate in passing through a large number of gates in series. Assuming a transition time of zero, the signal level at each pass through a gate may be read off the SLM input/output curve. If $f(x)$ represents this curve and x is the signal level at the gate input, for the case of a simple inverter, the following requirement ensures that the signal will not degenerate (Fig. 5):

$$a_{1-i} < kf(x) < b_{1-i} \text{ for all } x\epsilon(a_i,b_i), \qquad i = 0,1,$$

where k is the gain from the output of one gate to the input of the next, and $a_i < x < b_i$ defines the range of gate input signal levels interpreted as the discrete level i. Thus a gate input $x_0\epsilon(a_1,b_1)$ will have an output $y_0 = f(x_0)$ in the shaded region, and the input to the next gate will be $x_1 = ky_0$ on the x axis and in the shaded

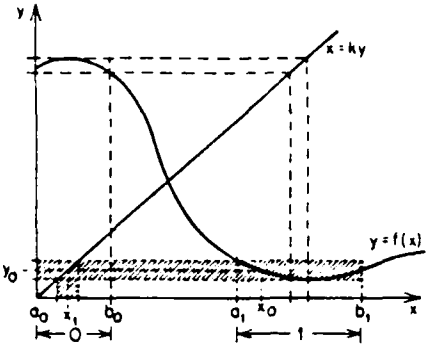


Fig. 5. Gate output y vs gate output x for an inverter that satisfies the regeneration criterion. The shaded region shows the criterion for a 1 input and a 0 output.

region. The above requirement states that x_1 will be interpreted as an input level of 0, i.e., that $x_1 \epsilon (a_0, b_0)$. This can be generalized to include the other Boolean operations by defining output intervals as the (union of the) mapping through f of the corresponding gate input intervals. One then requires appropriate combinations of output intervals (i.e., all possible sums of the elements) to map into the corresponding gate input intervals. The input/output curve of the LCLV used in the experiment does satisfy these criteria for the NOR and NOT operations.

III. Interconnections

The previous section dealt with optical implementation of logic gates, and in this section we address the problem of interconnecting these gates optically. The general problem is to be able to implement an arbitrary connection pattern between gate outputs and gate inputs. Since gate outputs correspond to inputs of the interconnection system and outputs of the interconnection system become gate inputs, for the remainder of this section the words input and output will refer to the interconnection system unless explicitly stated otherwise.

In direct analogy to the use of wires in an electrical circuit, optical fibers could be used for the interconnections. Although the idea is simple when there is a one-to-one correspondence between inputs and outputs, the method is less obvious when there is not. A possible scheme is described here. Assuming each input illuminates enough fibers, the fibers that are illuminated by an input j that addresses more than one output can be split so that an equal number of fibers P_j go to each output. Since all outputs may not have the same number of fibers, a mask is needed at the input plane with an intensity transmittance at each input that is inversely proportional to P_j. Alternatively, the same (albeit small, perhaps) number of fibers could be used to address all outputs making P_j independent of j. The problem of fabricating fiber optic bundles for large arbitrary interconnection patterns, however, is a serious drawback unless one has an automated system to do so.

The interconnections in the system presented here are implemented with a holographic element instead of

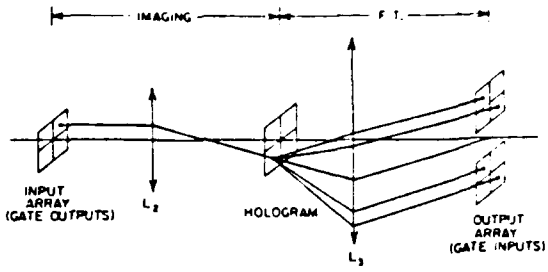


Fig. 6. System schematic for interconnecting gates. Each pixel of the input and output arrays corresponds to a gate. The input array is the SLM output (gate output array), and the output array is the SLM input (gate input array). In general the hologram produces multiple diffraction orders, only one of which is used. (The rays shown in this perspective sketch are a general indication of the path of propagation and are distorted for clarity. They should not be interpreted as an exact indication of ray paths.)

the fiber optic assembly. The holographic element consists of an array of subholograms in a one-to-one correspondence with the gates or pixels of the input array. The interconnection system is shown schematically in Fig. 6. This is only one of several holographic systems that could be used to interconnect the gates.[29] In this system, the input array is imaged onto the subhologram array. The hologram is encoded in the Fourier domain. A Fourier transform is then taken optically to obtain the output array. Each subhologram reconstructs a set of dots, one dot for each connection to a pixel of the output array. Because of the Fourier transform relationship, the coherence area of the illumination at the hologram must be larger than or equal to the subhologram size. If the coherence area is much larger, fringe patterns will appear in each pixel in the reconstruction plane. This occurs when different subholograms, whose separation is less than the coherence area, reconstruct dots at the same location in the output array. Spatial averaging over each element in the output array then results in an effective intensity summation, as desired for gate inputs (Figs. 1 and 2). Obviously, there must be at least one complete fringe cycle within the reconstructed pixel area for spatial average to be indicative of the sum of intensities. This will be the case so long as the subholograms do not overlap; defining the pixel area as the area of the main lobe of the reconstructed dot, it will include at least two fringe cycles. (The number of fringes per reconstructed pixel depends on the spacing of the contributing subholograms relative to the subhologram size but is independent of uniform scaling factors, lens focal lengths, and propagation distances.)

This interconnection system is essentially a space-variant filter providing a different point-spread function for each input pixel. Because of the Fourier transform relationship, each subhologram stores the absolute position of the output pixels it addresses. This absolute addressing scheme implies that the system does not distinguish between global and local interconnections but discriminates only on the basis of the position of the output array pixels relative to the origin of the output array.

To avoid a manufacturing problem similar to that of the fiber optic system, the hologram can be recorded optically using an automated system under computer control[31,32] or can be generated entirely by computer. In this experiment the hologram was generated by computer. Many types of computer-generated Fourier-transform hologram have been demonstrated,[33] and many of them would suffice for this application. To demonstrate the operation of the sequential logic system we used a binary version of the hologram proposed by Lee in 1970[34] for reasons of simplicity, diffraction efficiency, and SNR in the reconstruction.[35] Because of the coding process used in the hologram, the desired output array appears in the (1,0) diffracted order in the reconstruction plane leaving the spatially inverted (−1,0) order available for probing the system.

The effect of the hologram on the input array may be represented by the matrix equation

$$\mathbf{O} = \mathbf{MI}. \tag{2}$$

Here **I** is a vector representing the 2-D input array, lexicographically ordered. This input array is actually the gate output array augmented by the system inputs. Similarly, O is a vector representing the lexicographically ordered output array (which is the gate input array augmented by the system outputs). Each element of the vector **I** is binary valued and represents the signal of the corresponding pixel in the input array. Each element of O is a non-negative integer representing the signal level of the corresponding output array pixel. This integer is between 0 and N if the output pixel represents a gate with N input lines and is binary if the output pixel represents a system output. M is a matrix representation of the interconnect pattern—each matrix element m_{ij} is nonzero if and only if there is a connection between pixel j of the input array and pixel i of the output array. m_{ij} is an integer equal to the signal level created at output pixel i due to a signal level of 1 at input pixel j.

Using this notation, the fan-in to gate i (or number of input lines to gate i) is equal to the sum of the elements in row i of M. The fan-out of gate j (or number of gate input lines, plus the number of system outputs, that come from the output of gate j) is equal to the sum of the elements in column j of M. The fan-out is limited only by such parameters as the power of the illumination source, hologram efficiency, and SLM input sensitivity. The maximum fan-in is a function of the SLM intensity input/output characteristic.

With this interconnection technique, the hologram, or equivalently the elements of M, completely defines the circuit. Within the limit of the number of gates and the maximum power available, any operation that can be represented by a digital circuit can be implemented optically with this system by encoding the appropriate interconnection pattern into the hologram.

IV. Experimental Demonstration

For the experimental demonstration of this system, an LCLV with liquid crystal molecules in a 45° twisted nematic configuration was used as the SLM. The light valve is read out between crossed polarizers and is biased to implement a NOR operation. Its steady-state input/output relationship is shown in Fig. 7. This response, together with the attenuation of the interconnection system, satisfies the regeneration criteria stated in Sec. II.

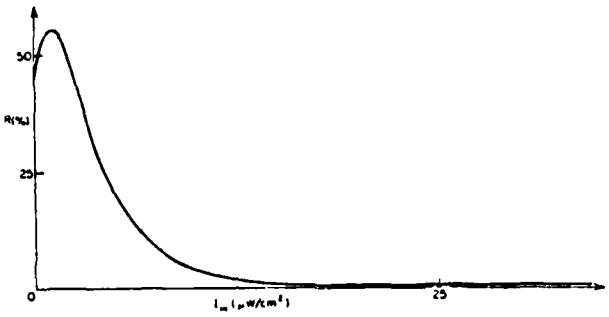


Fig. 7. Steady-state output vs input relationship for LCLV used to implement the NOR operation in the experimental system.

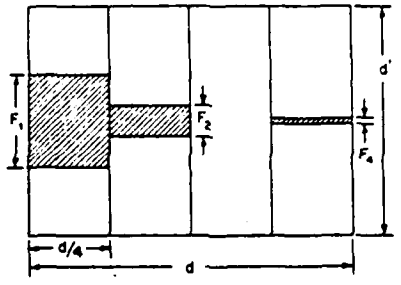


Fig. 8. Resolution cell of the 1970 Lee computer-generated hologram with binary-valued transmittance. Usually $d = d'$.

The gates are interconnected with a binary version of the Lee (1970) computer-generated hologram.[34] In a Lee hologram, each complex-valued sample is described by a linear combination of four real non-negative numbers (x_1,x_2,x_3,x_4), i.e., is decomposed into its components along each of the four half-axes in the complex plane. Each cell of the hologram is divided into four subcells, one subcell for each of these four components. One complex-valued sample is taken at the center of each subcell. Stored in each subcell is the corresponding component x_k of its complex-valued sample. Because of the locations of these subcells, upon taking the optical Fourier transform, these four components are added with the correct phases to obtain the reconstruction in the (1,0) diffracted order. If the transmittance of the hologram is binary valued, each subcell actually contains a rectangle whose width is equal to the subcell width and whose height F_k is proportional to the value of the corresponding sample component x_k (Fig. 8).

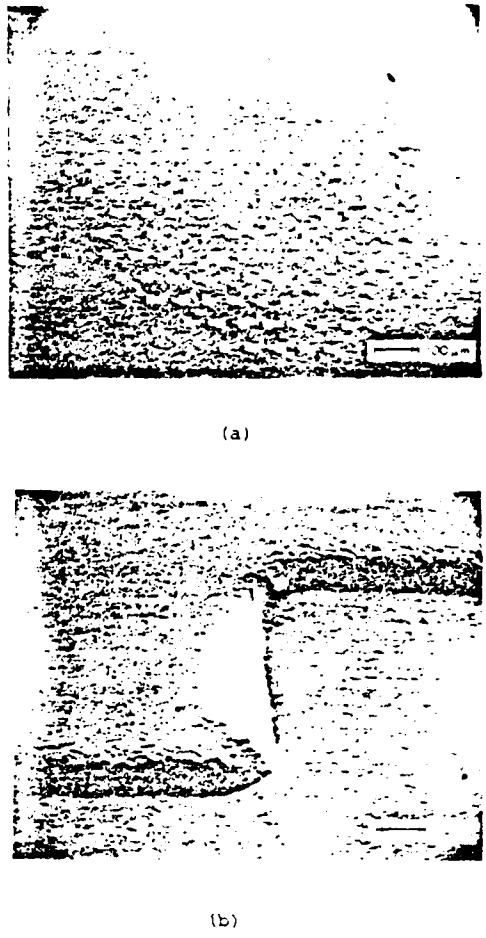
For the hologram used in this experiment, the transmittance is binary valued and these values are represented by different optical path lengths, i.e., a phase hologram. If the optical path lengths differ by a phase of π, the theoretical efficiency of the hologram is four times that of the equivalent absorption holo-

gram.[33] A possible trade-off is that the (0,0)-order intensity may increase by more than a factor of 4.

The hologram was written onto photoresist via electron-beam lithography. Surface relief of the photoresist provides the optical path length difference in the hologram. The electron-beam machine used has a step size of 0.125 μm and has written patterns with linewidths as small as 0.5 μm. It writes 1.024- × 1.024-mm fields and can stitch them together to cover a maximum area of 102 × 102 mm. The machine provides a far greater space–bandwidth product than was needed for our test circuit.

Our test circuit comprises 16 gates so the hologram comprises 16 subholograms, which are laid out in a 4 × 4 array. Each subhologram covers a circular area and has a diameter of ~1.04 mm. Each cell is a square 62.5 μm on a side, and there are 17 cells (68 subcells) along the horizontal diameter of each subhologram. Each subcell has a width of ~15.6 μm, or 125 steps of the electron-beam system, and has a height of 500 steps; both dimensions have more steps than were needed; 251 quantization levels were used for each subcell sample, keeping the apertures centered in each subcell.

Figure 9 shows pictures of one subhologram taken with a scanning electron microscope. Figure 9(a) shows nearly the entire subhologram. The rectangles are pits, the exteriors of which are photoresist, and the interiors of which are just glass substrate. Figure 9(b) shows a close-up of the top edges of two rectangles. The rectangle interiors (no photoresist) are located below these edges. The thickness of the photoresist is 1.25 μm, and the edges are inclined at ~32° with respect to the substrate normal. The pictures reveal that the photoresist is slightly rough near the edges but otherwise quite smooth (except for an occasional defect). Defects are apparent on the glass but are too small to affect the optical quality. Linewidths down to ~1 μm were obtained.



(a)

(b)

Fig. 9. Scanning electron microscope pictures of the interconnection hologram.

Defining efficiency as the power in the desired reconstruction pixels due to one subhologram divided by the power incident on that subhologram, the maximum efficiency over all subholograms was measured to be 5%. The other subholograms were intentionally generated to have lower efficiencies to normalize the intensities in the reconstruction plane. This measurement was taken using an illumination wavelength of 514.5 nm, the wavelength used in the sequential logic system. This is close to the optimum wavelength for this hologram. Aside from efficiency, we must consider the noise appearing in the desired reconstruction order, of which there are three sources: (1) the encoding process used to represent the complex-valued function on the hologram; (2) scattering from the photoresist and glass substrate; and (3) the tail of the (0,0) order. The pictures of Fig. 9 indicate that the contribution due to scattering should be small, and this is verified by experiment. The effect of the (0,0) order could be substantial with some phase computer-generated holograms. This effect can be prevented by an appropriate choice of the hologram encoding method or can be remedied by choosing the location of the limiting aperture appropriately and filtering out its diffraction effects spatially. This leaves the encoding process as the major source of noise. The effect of the encoding process on noise is discussed in Refs. 35 and 36. For the case of an interconnection hologram, we can define the SNR in each reconstruction pixel as the ratio of the power in the reconstruction pixel when it represents a maximum signal level to the maximum power in the same reconstruction pixel when it represents a signal level of 0. Measurements on our test hologram indicate a typical SNR of ~60.

A diagram of the main components of the sequential logic system is shown in Fig. 10. An expanded Ar laser beam is incident on the readout side of the LCLV (gate output plane). It is reflected off the internal mirror of the LCLV and is imaged from the liquid crystal (gate output) plane to the hologram via L_2. The liquid crystal plane is situated between crossed polarizers. The Fourier transform of the field transmitted by the hologram appears at the write side of the LCLV (gate input plane) via L_3. The phase of the illumination at the gate input plane is not correct, but only the intensity is of interest. Note that the Fourier transform rela-

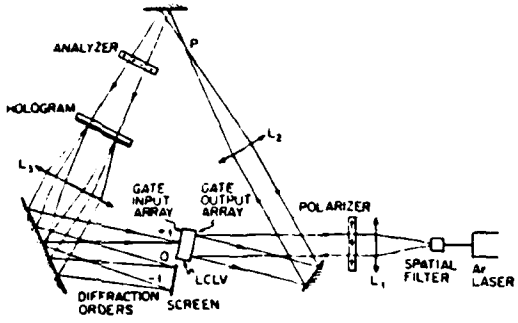


Fig. 10. Experimental system. Lens L_2 images from the LCLV gate output plane (liquid crystal plane) to the hologram plane. Lens L_3 provides a Fourier transform from the hologram plane to the LCLV gate input plane. The hologram comprises an array of subholograms.

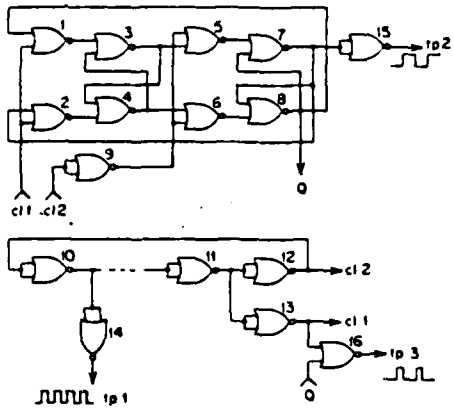


Fig. 11. Test circuit consisting of a synchronous master–slave flip-flop with driving clock.

tionship provides for complete regeneration of spot location during each pass through the feedback loop. In addition, since the subholograms are not contiguous, a mask is effectively incorporated into the hologram. This provides regeneration of the size and shape of each pixel and also facilitates alignment.

A diffuser is placed just in front of the LCLV gate input plane to average over the fringe patterns mentioned in Sec. III. This is not necessary when the pixels are small enough for the fringe patterns to be beyond the resolution limit of the device. Since a Lee phase hologram was used, and the effects of the (0,0) diffracted order were of concern, an aperture was used as a spatial filter at P, to filter out diffraction effects from the limiting aperture. This spatial filtering could be obviated by using a hologram encoding method that puts less power into the (0,0) order (e.g., Hseuh-Sawchuk hologram[37]). The (−1,0) diffracted order in the hologram reconstruction can be used to monitor the gate inputs during system operation. The gate outputs can also be probed by using a reflection off of the analyzer or the hologram.

For purposes of demonstration, a test circuit was selected. A gate may be used in a circuit with feedback, e.g. to achieve oscillation or memory, or in a circuit

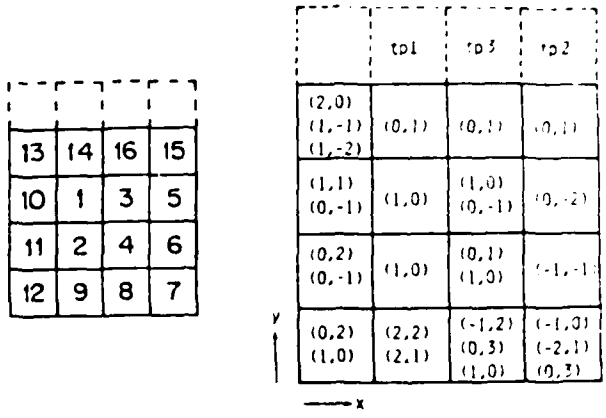


Fig. 12. (a) Layout of test-circuit gates on the hologram. (b) Interconnection patterns stored in each subhologram. Each ordered pair represents the relative x and y coordinates of a pixel, or gate input, that is addressed.

without feedback. A test circuit was chosen that includes both classes of circuit (Fig. 11), and a hologram with the appropriate interconnection pattern was generated (Fig. 12). The test circuit includes a synchronous master–slave flip-flop and a driving clock. The clock circuit is a ring oscillator consisting of an odd number of inverters. Clock circuits with three gates and with five gates have been implemented. The flip-flop functions as a frequency divider and outputs a signal whose frequency is half of that of the clock and whose duty cycle is close to 50%. The outputs of some of the gates in the test circuit are shown in Figs. 13 and 14. A discussion of them follows.

V. Discussion of Results

The clock circuit, consisting of five gates in a feedback loop, will be considered first. The phase delay of the signal in passing through one gate is expected to be 180° (for inverting gates) plus an additional $36° + 72°n$, n = integer, to insure that the total delay through all five gates is a multiple of 360°. The phase delay in passing through gate 12 is measured to be 216° (Fig. 13) within experimental error. An oscillator has also been constructed out of three gates, in which case the expected phase delay through each is $240° + 120°n$, and was measured to be close to 240°. Although nonuniformities across the spatial extent of the LCLV caused variations in characteristics from one gate to another yielding a slightly different phase delay through each gate, the sum of the three phase delays was 720° to within the accuracy of the measurement, as expected.

The frequency of oscillation of the clock circuit can be changed by changing the number of gates in the feedback loop. The five-gate clock circuit oscillated at a frequency of 2.64 ± 0.07 Hz (measurement error). (Speed is device-dependent and is discussed in Secs. II and VI.) If we assume the temporal behavior of the LCLV can be modeled by a simple RC circuit (consisting of a series resistor followed by a shunt capacitor), this measurement implies that the three-gate clock should oscillate at a frequency of 6.32 Hz. We observed

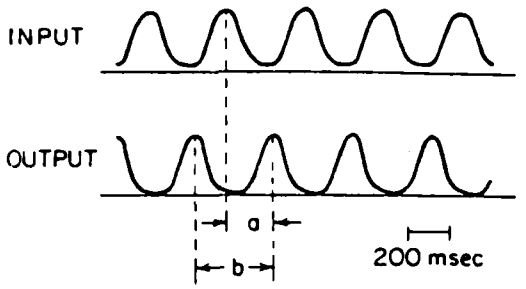


Fig. 13. Input and output of one of the five gates in the clock circuit (gate 12). Each trace represents intensity vs time. The period is b, and the phase delay is $360° \cdot a/b$.

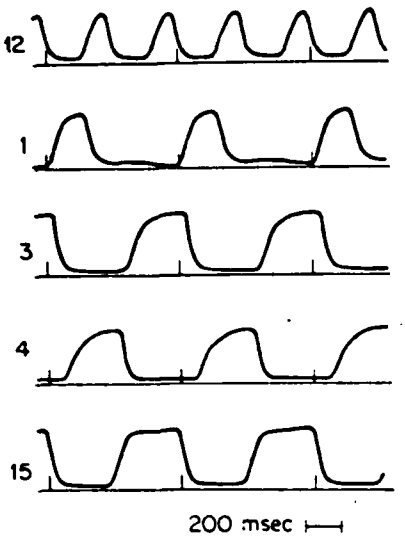


Fig. 14. Test circuit outputs.

6.40 ± 0.12 Hz. No attempt to measure the frequency stability was made. The system did, however, sustain oscillations for periods lasting 8 h. (The system was turned off after 8 h.) Given the number of gates in the clock circuit, the frequency of oscillation is determined solely by the temporal characteristics of the LCLV. For sufficiently fast SLMs, the optical path length of the interconnection system will also have an effect.

Measurements we have taken indicate that the LCLV response time is limited by the photoconductor and that the response time of the liquid crystal is much faster.[14] Improvements in LCLV photoconductor response times[38] should yield higher clock frequencies. Use of optical bistable devices could improve clock frequencies by many orders of magnitude (see Secs. II and VI).

The test circuit (which has five gates in its clock circuit) functions correctly. The output waveforms (Figs. 13 and 14) are not expected to be square. The gates are operating near their maximum speed, so their rise and fall times are substantial in comparison with their pulse widths. With one minor exception, all gates in the test circuit output the expected waveforms. The exception is that the pulse widths of the outputs of gates 5 and 6 are less than those of the outputs of gates 1 and 2. They were expected to be the same. This is a result of unequal interconnection losses among different gates in the clock circuit, due to an error in the hologram, which caused gate 11 to output larger pulse widths than gate 12.

VI. Device Limitations

We have mentioned in Sec. II the possibility of using much faster devices in place of the LCLV, for example, an optical bistable device. We must consider the fundamental potentials and limitations involved in using such devices as optical logic gates, as well as how they might comapre with their electrical counterparts. A number of people have studied this subject.[39–46]

A very important consideration is power dissipation of these devices. This takes the form of heat which must be removed from the device to keep its temperature within operating bounds. Earlier works have indicated that optical logic gates may suffer from a higher power dissipation than electronic gates and in particular noted the unfavorable trend on fundamental limits in optical logic of increasing power dissipation with decreasing delay time vs a fundamental limit of power dissipation that is independent of delay time in the case of semiconductor electronic devices.[40,41] More recently it has become evident that most of the power dissipation in a practical integrated circuit is due to the on-chip interconnections instead of the transistors themselves and that the lower limit on this power dissipation also increases with decreasing delay time.[44] This limit is essentially the same, when plotted on a power vs delay time graph, as the lower limit on power of an optical switching device using an absorptive nonlinearity given in Ref. 45. Furthermore, the limit on power in the case of an optical switch using a reactive nonlinearity increases more slowly (than the optical absorptive and electrical cases) with decreasing delay, although its power level is higher in the region of common delay times (>1 psec). While lowering the operating temperature of semiconductor logic will lower its power dissipation limit to a point,[44] use of an optical resonator can, in some cases, reduce the limits on the power dissipation of the optical switch.[45,46] Finally, we point out that in the optical case much of the power can potentially be dissipated external to the device, permitting the operation of switches at significantly higher power levels than would otherwise be possible.

While these fundamental limits on optical switches can be approached with known materials,[45] significant progress, some of a relatively fundamental nature, will have to be made for these switching devices to become competitive with electronics.[45,46] Optical gates will probably not replace electronic gates for use in general-purpose computers in the near future, but their use in an optical special-purpose computer could permit realization of a number of architectural advantages over semiconductor electronics.[29] These advantages include parallel input/output, global as well as local interconnections, and implementation of interconnection-intensive circuits and processors without reducing the active device area available for gates.

VII. Conclusions

We have presented an all-optical sequential logic system. It is all optical in that every signal is represented optically, and it is sequential in that it can include memory elements and clocks. We demonstrated the operation of the system using a test circuit consisting of a synchronous master–slave flip-flop and its driving clock. The circuit functioned properly, and the output of each gate was as expected for the given interconnection hologram.

On this system, any digital circuit can be implemented up to limitations in the total number of gates. The circuit is encoded in the hologram. Since the hologram represents a fixed interconnection pattern, the circuit or processor is not reconfigurable in real time. This does not eliminate the possibility of software control, however, as can be seen by noting that the interconnections between gates in a general-purpose electronic computer are also fixed. As in an electronic sequential circuit, software control is obtained by changing the inputs to appropriate control lines.

The speed of operation of the processor depends on the device. While an LCLV was used to demonstrate the system concept, much faster SLMs in the same system will yield much faster processors. Recent progress in optical bistability provides hope for a fast optical logic system.

The maximum number of gates that can be implemented is limited by the space–bandwidth product of the hologram.[29] However, this restriction can be alleviated by using a different interconnection technique.[29]

Aside from the question of speed and number of gates, this optical system has some architectural advantages over conventional digital electronic systems. First, parallel inputs (and outputs) can easily be incorporated into the system. This permits large amounts of parallel data to be input to and output from the system, alleviating the pin-out constraints found in semiconductor electronics. Second, communication intensive operations may be performed easily with the optical system. And finally, the optical system cannot tell the difference between global and local interconnections. It is the lack of these features that is becoming a substantial limiting factor in the design and development of state-of-the-art semiconductor electronic systems. These points are treated more fully in the subsequent paper.[29]

The authors thank P. Chavel for many helpful technical discussions. The holograms used in this work were fabricated at the Microelectronics Research and Development Center of Rockwell International by R. Imerson and A. B. Jones. The authors gratefully acknowledge their assistance. This work was supported by the Air Force Office of Scientific Research under grant AFOSR-81-0182. Portions of this paper were presented at the 1982 Annual Meeting of the Optical Society of America, Tucson, Ariz., Oct. 1982.[47]

References

1. A. Huang, "The Implementation of a Residue Arithmetic Unit via Optical and Other Physical Phenomena," in *Proceedings, 1975 International Optical Computing Conference* (1975), p. 14.
2. A. Huang, Y. Tsunoda, J. W. Goodman, and S. Ishihara, "Optical Computation Using Residue Arithmetic," Appl. Opt. 18, 149 (1979).
3. D. Psaltis and D. Casasent, "Optical Residue Arithmetic: A Correlation Approach," Appl. Opt. 18, 163 (1979).
4. A. Tai, I. Cindrich, J. R. Fienup, and C. C. Aleksoff, "Optical Residue Arithmetic Computer with Programmable Computation Modules," Appl. Opt. 18, 2812 (1979).
5. S. A. Collins, Jr., J. Ambuel, and E. K. Damon, "Numerical Optical Data Processing," *Proceedings 1978 International Optical Computing Conference*, IEEE cat. no. 79CH1305-2, 1974 (1978).
6. F. A. Horrigan and W. W. Stoner, "Residue-Based Optical Processor," Proc. Soc. Photo-Opt. Instrum. Eng. 185, 19 (1979).
7. C. C. Guest and T. K. Gaylord, "Truth-Table Look-up Optical Processing Utilizing Binary and Residue Arthimetic," Appl. Opt. 19, 1201 (1980).
8. B. Horwitz and F. Corbett, "The PROM-Theory and Applications for the Pockels Readout Optical Modulator," Opt. Eng. 17, 353 (1978).
9. S. Iwasa and J. Feinleib, "The Prom Device in Optical Processing Systems," Opt. Eng. 13, 235 (1974).
10. C. Warde, A. D. Fisher, J. I. Thackara, and A. M. Weiss, "Image Processing Operations Achievable with the Microchannel Spatial Light Modulator," Proc. Soc. Photo-Opt. Instrum. Eng. 252, 25 (1980).
11. 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).
12. S. A. Collins, Jr., M. T. Fatehi, and K. C. Wasmundt, "Optical Logic Gates Using a Hughes Liquid Crystal Light Valve," Proc. Soc. Photo-Opt. Instrum. Eng. 232, 168 (1980).
13. B. H. Soffer, D. Boswell, A. M. Lackner, P. Chavel, A. A. Sawchuk, T. C. Strand, and A. R. Tanguay, Jr., "Optical Computing with Variable Grating Mode Liquid Crystal Devices," Proc. Soc. Photo-Opt. Instrum. Eng. 232, 128 (1980).
14. A. A. Sawchuk, T. C. Strand, and A. R. Tanguay, Jr., "Nonlinear Real-Time Optical Signal Processing," USCIPI Report 1080, Image Processing Institute, U. Southern California (1982).
15. R. A. Athale and S. H. Lee, "Development of an Optical Parallel Logic Device and a Half-Adder Circuit for Digital Optical Processing," Opt. Eng. 18, 513 (1979).
16. R. A. Athale, H. S. Barr, S. H. Lee, and B. J. Bartholomew, "Digital Optical Processing," Proc. Soc. Photo-Opt. Instrum. Eng. 241, 149 (1980).
17. J. Tanida and Y. Ichioka, "Optical Logic Array Processor," in *Proceedings, Tenth International Optical Computing Conference*, IEEE cat. no. 83CH1880-4, 18 (1983).
18. A. A. Sawchuk and T. C. Strand, "Digital Optical Computing," Proc. IEEE 72, 58 (1984).
19. C. T. Seaton, F. A. P. Tooley, M. R. Taghizadeh, W. J. Firth, and S. D. Smith, "Optical Logic Operations in InSb Bistable Devices," in *Technical Digest, Conference on Lasers and Electrooptics* (Optical Society of America, Washington, D.C., 1983), paper WC5; D. A. B. Miller, S. D. Smith, and C. T. Seaton, IEEE J. Quantum Electron. QE-17, 313 (1982).
20. 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, 172 (1984); J. L. Jewell, M. C. Rushford, H. M. Gibbs, and N. Peyghambarian, "Single-Etalon Optical Logic Gates," in *Technical Digest, Conference on Lasers and Electrooptics* (Optical Society of America, Washington, D.C., 1984), paper THg2.

21. U. K. Sengupta, U. H. Gerlach, and S. A. Collins, "Bistable Optical Spatial Device Using Direct Optical Feedback," Opt. Lett. 3, 199 (1978).
22. U. H. Gerlach, S. A. Collins, and U. K. Sengupta, "Single-Spatial Light Modulator Bistable Optical Matrix Device Using Optical Feedback," Opt. Eng. 19, 452 (1980).
23. R. Athale and S. H. Lee, "Bistability and Thresholding by a New Photoconductor-Twisted Nematic Liquid Crystal Device with Optical Feedback," Appl. Opt. 20, 1424 (1981).
24. H. M. Gibbs, S. L. McCall, and T. N. C. Venkatesan, "Optical Bistable Devices: The Basic Components of All-Optical Systems?" Opt. Eng. 19, 463 (1980).
25. D. H. Schaefer and J. P. Strong III, "Tse Computers," Proc. IEEE 65, 129 (1977).
26. A. Huang, "Design for an Optical General Purpose Digital Computer," Proc. Soc. Photo-Opt. Instrum. Eng. 232, 119 (1980).
27. A. Huang, "Parallel Algorithms for Optical-Digital Computers," in *Proceedings, Tenth International Optical Computing Conference*, IEEE cat. no. 83CH1880-4, 13 (1983).
28. H. Barr and S. H. Lee, "A Digital Optical Processing System," in *Proceedings, Tenth International Optical Computing Conference*, IEEE cat. no. 83CH1880-4, 171 (1983).
29. B. K. Jenkins, P. Chavel, R. Forchheimer, A. A. Sawchuk, and T. C. Strand, "Architectural Implications of a Digital Optical Processor," Appl. Opt. 23, 3465 (1984).
30. P. W. Smith and W. J. Tomlinson, "Bistable Optical Devices Promise Subpicosecond Switching," IEEE Spectrum 26 (June 1981).
31. S. K. Case, P. R. Haugen, and O. J. Lokberg, "Multifacet Holographic Optical Elements for Wave Front Transformations," Appl. Opt. 20, 2670 (1981).
32. P. R. Haugen, H. Bartelt, and S. K. Case, "Image Formation by Multifacet Holograms," Appl. Opt. 22, 2822 (1983).
33. W. J. Dallas, "Computer-Generated Holograms," in *The Computer in Optical Research: Methods and Applications*, B. R. Frieden, Ed. (Springer, New York, 1980), p. 291.
34. W. H. Lee, "Sampled Fourier Transform Hologram Generated by Computer," Appl. Opt. 9, 639 (1970).
35. J. Bucklew and N. C. Gallagher, Jr., "Comprehensive Error Models and A Comparative Study of Some Detour-Phase Holograms," Appl. Opt. 18, 2861 (1979).
36. J. P. Allebach, "Representation-Related Errors in Binary Digital Holograms: A Unified Analysis," Appl. Opt. 20, 290 (1981).
37. C. K. Hsueh and A. A. Sawchuk, "Computer-Generated Double-Phase Holograms," Appl. Opt. 17, 3874 (1978).
38. U. Efron, P. O. Braatz, M. J. Little, R. N. Schwartz, and J. Grinberg, "Silicon Liquid Crystal Light Valves: Status and Issues," Proc. Soc. Photo-Opt. Instrum. Eng. 388, 75 (1983).
39. R. W. Keyes, "Physical Problems and Limits in Computer Logic," IEEE Spectrum 36 (May 1969).
40. R. W. Keyes and J. A. Armstrong, "Thermal Limitations in Optical Logic," Appl. Opt. 8, 2549 (1969).
41. R. W. Keyes, "Power Dissipation in Information Processing," Science 168, 796 (1970).
42. R. Landauer, "Optical Logic and Optically Accessed Digital Storage," in *Optical Information Processing*, Y. E. Nesterikhin, G. Stroke, and W. Kock, Eds. (Plenum, New York, 1975), p. 219.
43. R. W. Keyes, "Physical Limits in Digital Electronics," Proc. IEEE 63, 740 (1975).
44. R. W. Keyes, "Fundamental Limits in Digital Information Processing," Proc. IEEE 69, 267 (1981).
45. P. W. Smith, "On the Physical Limits of Digital Optical Switching and Logic Elements," Bell Syst. Tech. J. 761, 1975 (1982).
46. R. L. Fork, "Physics of Optical Switching," Phys. Rev. A 26, 2049 (1982).
47. B. K. Jenkins, A. A. Sawchuk, T. C. Strand, and B. H. Soffer, "Sequential Optical Logic Implementation," J. Opt. Soc. Am. 72, 1721A (1982).

Polar Duke Replaces Hero

The NSF has acquired the services of an ice-strengthened research ship to support the U. S. Antarctic Program for three years (with possible extension of an additional two years) beginning in Dec. 1984. The ship will be used exclusively in Antarctica each austral summer (Nov. through Apr.), but may be available for other programs during the austral winters (May through Sept.), beginning in May 1985.

Polar Duke, whose 219-foot hull is classified as strong as that of an ice-breaker, will replace the 125-foot wooden-hulled *Hero*, the NSF ship which has supported antarctic research since 1968. Built in 1983 in Norway, *Polar Duke* has a 43-foot beam, two main engines that give it a maximum speed of 15 knots, and a helicopter deck. It currently accommodates up to 27 science personnel. It was designed to support scientific and transport expeditions in arctic and antarctic waters.

Like the *Hero*, the *Polar Duke* will operate between the southern tip of South America and the Antarctic Peninsula where NSF funds research in biology, oceanography, and geology. The *Hero* will continue work in southern Chilean coastal waters until Sept. 1984 when it will return to the U.S.

Architectural implications of a digital optical processor

B. K. Jenkins, P. Chavel, R. Forchheimer, A. A. Sawchuk, and T. C. Strand

A general technique is described for implementing sequential logic circuits optically. In contrast with semiconductor integrated circuitry, optical logic systems allow very flexible interconnections between gates and between subsystems. Because of this, certain processing algorithms which do not map well onto semiconductor architectures can be implemented on the optical structure. The algorithms and processor architectures which can be implemented on the optical system depend on the interconnection technique. We describe three interconnection methods and analyze their advantages and limitations.

I. Introduction

There has recently been considerable research in optical systems for parallel computing with applications in signal processing. The advantages of optical and hybrid optical-electronic systems for high throughput parallel multidimensional processing on signals with large time-bandwidth and space-bandwidth products are well known. Nearly all these systems to date are basically analog and have severe limitations in accuracy, programmability, and flexibility compared with electronic digital systems.

Our recent research concentrated on optical combinatorial and sequential logic systems for parallel digital processing. Some of this work included parallel analog-to-digital conversion[1] and two different implementations of optical combinatorial logic.[2,3] Other implementations of parallel optical combinatorial logic have also been reported.[4–8] In addition, a few systems of varying degrees of flexibility using optical logic operations as the primary data processing element have been reported.[8–13] Recently, we implemented an optical sequential logic circuit including a clock and a master-slave flip-flop used as a frequency divider.[14] This sequential logic system provides an advantageous combination of a very high degree of flexibility along with all signals (including data and control signals) being optical. The main components of the sequential logic system are a nonlinear spatial light modulator (SLM) (ideally having a threshold response function) and a computer generated hologram (CGH) used as a beam-steering element for interconnections. The SLM functions as a 2-D array of independent logic gates, and the CGH (or set of them) generally consists of a 2-D array of subholograms that interconnect the gates to form a circuit. In the current system the nonlinear element is a Hughes liquid crystal light valve (LCLV) with a 45° twisted orientation of the nematic liquid crystal molecules.[15] Although a major limitation of this current SLM is its slow response time (10–100 msec), we feel that recent improvements in both LCLV technology[16] and the exploration of new technologies such as all-optical bistability[17–20] will significantly improve response times. We will not directly consider the question of device speed in this paper; it is discussed in Ref. 14.

The main emphasis in this paper is on processor architectures for optical sequential logic. Section II briefly reviews the fundamentals of optical sequential logic. Section III describes basic design considerations, which are considerably different from those used in very large scale integration (VLSI) in digital electronics. Sections IV–VI describe details of CGHs used as interconnection elements. Two basic interconnection methods, space-variant and space-invariant, are described. The main limitation on the number of gates is due to space–bandwidth limitations of the CGH and SLM. A hybrid interconnection system having both space-variant and space-invariant elements is described in Sec. VI, and various processor architectures that take advantage of each interconnection method are described.

II. Fundamentals of 2-D Optical Sequential Logic

To implement any logic system, we require two fundamental elements: a nonlinear device to provide the

P. Chavel is with Institut d'Optique, Orsay, France; R. Forchheimer is with Linkoping University, Linkoping, Sweden; when this work was done B. K. Jenkins, A. A. Sawchuk, and T. C. Strand were with University of Southern California, Department of Electrical Engineering, Los Angeles, California 90089; T. C. Strand is now with IBM Corporation, San Jose, California 95193.

Received 6 January 1984.
0003-6935/84/193465-10$02.00/0.

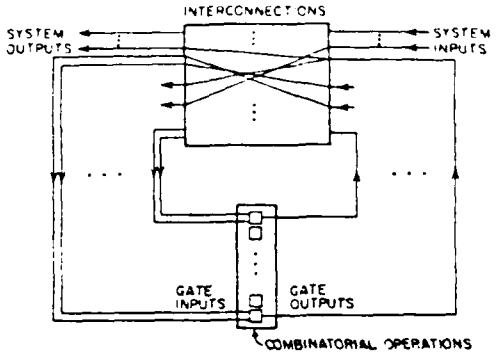


Fig. 1. Functional block diagram of sequential optical logic.

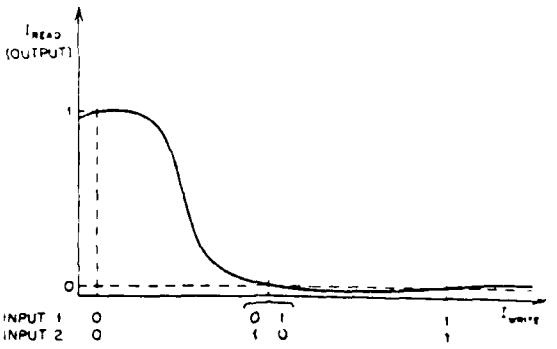


Fig. 2. LCLV input/output characteristic.

gate function or basic combinatorial operations and an interconnection element (Fig. 1). Furthermore, if we want to provide for sequential logic, the interconnection path must include feedback paths for generating clock signals and for obtaining memory elements. The introduction of optical feedback and optical timing signals makes this work significantly different from previous work, because the dynamic behavior of the nonlinear device now plays a critical role in operation of the circuit.

We use the Hughes LCLV as the nonlinear component, although other nonlinear devices could also be used. This device produces a pointwise nonlinear behavior, which to some extent can be modified and in particular can take a shape adequate for our present needs. For example, Fig. 2 depicts a response function for a 45° twisted nematic device operated in the backslope mode. We used the device in this mode to implement the NOR function. If we consider the total input to the device (gate input) as the sum of two binary inputs (input lines), the output will be a binary valued NOR of the input lines. Other binary operations can be performed by altering the characteristic curve of the device.[3] Possible input and output values are indicated in Fig. 2.

The nonlinearity is applied to all points on the device simultaneously. Thus each resolution element or pixel on the light valve acts as an independent gate. Using resolution figures quoted for current SLMs,[16] arrays of 10^5–10^6 pixels can be anticipated.

The remaining problem is how to interconnect the gates. Although several techniques are possible, CGHs seem to offer the best solution. By using CGHs in an optical feedback system, the output from any gate can be directed to the input of any other gate or combination of gates. Given that CGHs are to be used for interconnections, there are still a multitude of possible systems for achieving the desired circuit. In the following sections we describe three basic interconnection methods. Naturally, each method offers certain design trade-offs and limitations. It is the purpose of this paper to examine those trade-offs and describe how they affect system design.

III. Processor Design Considerations

The rapid evolution in semiconductor fabrication technology has brought with it significant changes in the design criteria for digital systems. The technology of optical logic further alters the ground rules of circuit design. Because the design limitations also determine the architectures appropriate for a given technology, we will briefly consider the evolution in these critical factors.

From the earliest digital electronic systems, up through the stages of small and medium scale integration, the gate count, or total number of gates required to implement a given circuit, has been the overriding design criterion. Circuit design was an exercise in Boolean algebra utilizing Karnaugh maps and related techniques to minimize the total gate count while still performing the desired function.

With the development of large scale integration (LSI) and very large scale integration (VLSI), the cost of individual gates plummeted to the point where it constitutes only a minor factor in the overall system cost function. The major concern in VLSI systems has become internal and external communications.[21,22] The internal wiring network affects the amount of active chip area available for gates and also imposes several timing restrictions on the systems. In current systems it is common for as much as 70% or more of the chip area to be devoted to interconnections.[23] The speed of the system is a function of propagation delay and gate delay. Except for very short wires, propagation is a diffusion process, and the propagation delay is determined by the length, resistivity, and parasitic capacitance associated with the wire.[24] The gate delay is also affected by parasitic capacitance of the wire. Although the propagation delays can be reduced by process technology (e.g., using thick metal layers), it appears they will still be a limiting factor in the system timing considerations.[25] The same propagation delay considerations lead us to the conclusion that any attempt to distribute synchronizing clock signals over an entire chip will inevitably introduce serious clock skew.[24,26] The wire length also has a direct effect on the power consumption within the chip. Another restriction in the interconnection network is that the VLSI circuit is essentially a 2-D structure. Although wires can cross over one another in a multilayer structure, the maximum feasible number of overlapping wires or conducting layers is very small. A final VLSI limitation related to wires and

communication is the problem of pin-outs. The number of pins available for communication between the chip and the rest of the system is severely limited. In VLSI systems there is a great deal of interest in systolic arrays and related architectures that achieve high processing throughput rates for certain algorithms by exploiting concurrency and pipelining.[21,26] However, highly parallel systems usually require a large number of input or output data channels, and highly pipelined systems frequently require a large number of input control lines. Thus the extent of concurrency and pipelining which can be achieved is directly limited by the number of pin connections to the chip.[27] Pin-out constraints will similarly limit the efficiency of other VLSI processors.

One effect of these communication limitations is the emphasis on highly structured and regular interconnection patterns. Modular building blocks such as programmable logic arrays (PLAs), read-only memories (ROMs), and standard cells are heavily relied upon to produce VLSI chips. Another effect is the close scrutiny of data flow by VLSI designers to find ways in which pipelining can be maximized and to find ways of eliminating synchronization requirements.[28] All this has led to a concentration of interest in special purpose processors such as wave front array processors and systolic arrays.[26,28,29]

For the optical logic system described here, the major design considerations are admittedly not so well defined as for VLSI at this point. However, it is clear that the cost function is much different than for the integrated circuit systems. In particular, most of the communication costs affecting VLSI design are not associated with the optical system. For the systems described here, all interconnections have the same length to first order due to their optical imaging or Fourier transform nature. Thus synchronization problems due to clock skew should be eliminated, making large synchonized systems more feasible. Being able to synchronize the circuits eliminates the need for handshaking or other asynchronous techniques which introduce various amounts of waiting time for individual circuit elements. Although an interconnection path length would still ultimately affect the speed of the optical system, since the path length is constant, it is not a design variable that affects architectural decisions. The other problems associated with wire length, namely, power consumption and device area utilization, are not relevant to the optical logic system. With the optical systems, none of the active device area is occupied by interconnections. Also the power consumption of the system is independent of the interconnection path length.

A striking feature of the optical system is the interconnection flexibility. As we pointed out, wire length is not a problem for an optical system. Thus, for example, a gate may be connected to the most distant gates on the device as easily as it is connected to its nearest neighbors. With VLSI, even if such interconnections were not impractical due to wire length restrictions, they would be impractical due to the inevitable problems of wires crossing over one another in conjunction with the reduction in device area utilization. This restriction makes many interesting architectures such as cellular pyramids very difficult to implement in VLSI.[30] The optical system, on the other hand, has no such restriction. Thus pyramid structures and other architectures requiring complex interconnections would merit consideration for an optical logic system.

Finally, the problem of pin-outs is also alleviated by the optical logic system described here. The optical system naturally can accept a large number of parallel inputs and at the same time generate a large number of parallel outputs. Thus the problem of communication between separate devices is eliminated as is the limitation on concurrent/pipelined processing mentioned above.

Having shown that the communication restrictions which dictate design criteria for VLSI do not apply to the optical logic system, what are the major limiting factors for the optical system? Given that we are ignoring the important question of gate switching times at this point, with the expectation that future developments will improve this parameter, the most important factor seems to be the space-bandwidth product (SBWP) of the holograms which form the interconnections. This is, we should point out, loosely connected to the question of wire lengths, although it is more appropriately associated with a measure of the interconnection complexity and regularity. In the following sections we describe three different interconnection schemes which differ in their interconnection complexity: a space-invariant system with a perfectly regular structure with minimal space–bandwidth requirements; a space-variant structure with severe space–bandwidth requirements; and a hybrid system. The constraints associated with these systems will determine the architectures which are most appropriate for implementation with the optical logic system.

Since there are significant differences between the factors which limit VLSI and our optical logic system, one might ask if there are application areas specifically well suited to one or the other. Due to the restrictions on communications, VLSI systems are being particularly considered for applications which involve very regular structures and simple data flow that can be handled with only local communications. This has led to the development of systolic array architectures.[26,29] These structures are well suited to many vector–matrix and matrix–matrix operations.

However, there are many common algorithms which inherently require global communications. These cannot be conveniently handled by VLSI but could, in principle, be implemented with the optical logic system.[31] For example, although the discrete Fourier transform (DFT) can be expressed as a problem involving only local communications, the fast Fourier transform (FFT) algorithms inherently require global communications due to their pyramidlike butterfly structure. Thus by going to an optical logic system, one might be able to take advantage of the FFT structure and speed at the hardware level. Other examples of communication-limited operations are value counting

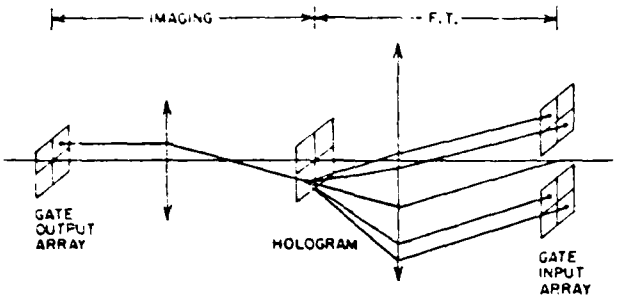


Fig. 3. Space-variant interconnection system. In general the hologram produces multiple diffraction orders, only one of which is used.

or histogramming and regional property computation, both of which require global summing operations.[30] Additional thoughts on parallel optical digital computers and applications to symbolic substitution are given in Ref. 11.

IV. Space-variant Interconnection Method

The most general interconnection system is one in which any gate output can be connected to the input of any gate or combination of gates. If we think of the interconnection system as imaging the gate output array plane onto the gate input array plane, this approach represents a space-variant imaging system. The image of a gate output consists of a collection of spots (the impulse response of the system for that particular point) that illuminate the appropriate gate inputs and form the circuit interconnections. Because each object point (gate output) sees a different impulse response (interconnection pattern) this represents a general space-variant system. A space-variant system has been built to demonstrate the concept of sequential optical logic. The demonstration circuit that was implemented comprises a ring oscillator which generates a clock signal and a master–slave flip-flop driven by the clock. This system is operational and described in another paper.[14]

A schematic diagram of the optical system used for the space-variant interconnections is shown in Fig. 3. First, the gate outputs are imaged onto the interconnection hologram. This CGH consists of an array of subholograms, one subhologram for each gate. When illuminated by its corresponding gate output, a subhologram will reconstruct an image on the write side, or gate input side, of the light valve. The reconstructed images are simple dot patterns, each bright dot illuminating a gate input. Because the reconstructed images can be designed to illuminate any combination of gate inputs, arbitrary interconnections are possible. As shown in Fig. 3, the desired interconnections are formed in one particular diffraction order. Typically, a conjugate image will also be produced, in which case it can be used to probe the system without affecting system operation or to access the system outputs.

While this interconnection scheme allows complete generality, a price is paid in terms of the space-bandwidth product requirements on the CGH. Let there be an $N \times N$ array of subholograms on the CGH and an $N \times N$ array of gates on the nonlinear device. Each subhologram must be capable of addressing any of the N^2 gate inputs. The number of addressable points is equal to the number of complex-valued samples in a subhologram assuming a Fourier hologram. Thus the actual SBWP of each subhologram is

$$S_S = p^2 q^2 N^2, \tag{1}$$

where p^2 is the number of resolution elements in the hologram used to represent one complex-valued sample, and q^2 is a factor representing the amount of oversampling in the hologram plane. Generally $p^2 > 1$ because the complex sample values must be encoded into the hologram, e.g., as real values. Also we generally have $q^2 > 1$ to avoid cross talk. These problems are discussed below.

The entire interconnection hologram consists of N^2 subholograms, one for each gate. Thus the total SBWP of the hologram is S_T where

$$S_T = N^2 S_S = p^2 q^2 N^4. \tag{2}$$

Because $S_T \propto N^4$, we expect that the hologram SBWP S_T will quickly become the limiting factor as N increases. We will verify that below, but first we need to study the cross talk in the gate-input plane to get a feeling for the expected values of q^2.

The cross talk can be represented by α, the ratio between gate inputs of the worst-case (largest) zero value I_0 and the worst-case (smallest) one value I_1:

$$\alpha = I_0/I_1. \tag{3}$$

We require $\alpha < 1$ to distinguish all possible zero and one states. Assume that the intensity profile of a single gate input reconstructed from a subhologram is $F(x,y)$. [The reconstruction of a subhologram can be represented by a set of Dirac delta functions (one for each addressed gate input) convolved with the Fourier transform $W(x,y)$ of the aperture function of the subhologram. Then $F(x,y) = |W(x,y)|^2$.] Thus the worst-case one value is the integral of F over the defined area a of the gate input:

$$I_1 = \iint_a F(x,y)\,dx\,dy. \tag{4}$$

The worst-case zero level occurs when all gates have their maximum input levels except for the gate in question which has a zero input level. If each gate has m input lines (a fan-in of m) and gates are contiguous, the worst-case zero can be shown to be

$$I_0 = m\left[\iint_A F(x,y)\,dx\,dy - \iint_a F(x,y)\,dx\,dy\right]. \tag{5}$$

where A is the area of the entire gate array. This assumes that the spatially integrated sum of gate input lines from different subholograms is effectively an incoherent sum. Note that such inputs actually add coherently but that they produce interference fringes. Spatial integration over these fringes results in the effective incoherent summation. Similarly, the above equation assumes that the gate inputs reconstructed from a single subhologram also can be modeled as

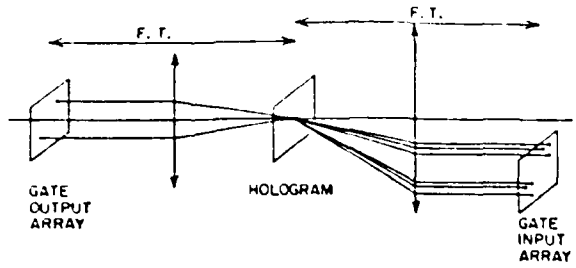


Fig. 4. Space-invariant interconnection system.

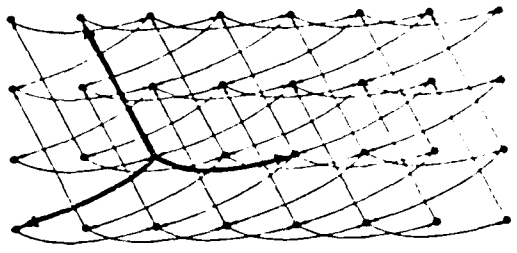


Fig. 5. Example of a space-invariant interconnection network. Nodes represent gates, and arrows are the (optical) interconnections.

adding incoherently. This will be essentially true if a pseudorandom phase is applied to the n distinct gate inputs, where n is the fan-out. Combining the above equations we get

$$\alpha = m(1/E_0 - 1), \tag{6}$$

where E_0 is the fraction of the single-gate input intensity profile which falls within the defined area of that gate:

$$E_0 = \iint_a F(x,y)dxdy \Big/ \iint_A F(x,y)dxdy. \tag{7}$$

We define the minimum gate area as $a_0 = 1/L_{SH}^2$, where L_{SH} is the width of each (square) subhologram. Allowing for N^2 contiguous gates implies a Nyquist frequency $f_{Nyq} = N/L_{SH}$. The actual gate area is

$$a = q^2 a_0, \tag{8}$$

where q is the oversampling factor defined by $f_{samp} = qf_{Nyq}$ and f_{samp} is the sampling rate. Obviously, as q increases E_0 approaches unity and α approaches zero.

We now consider an example. If we use a triangle function in x and y for the aperture (window) function of each subhologram, its Fourier transform is a 2-D sinc^2 function,

$$F(x,y) = \text{sinc}^4(x/2)\,\text{sinc}^4(y/2). \tag{9}$$

Choosing $q = 2$ yields a cross talk $\alpha = (0.11)m$, so that 3-input gates cause a cross talk of 0.33. Thus for NOR gates the thresholding of the nonlinear device could be performed anywhere between relative gate input levels of 0.33 and 1.0. In this case the sampling rate is twice the Nyquist rate. Increasing q permits a larger fan-in, e.g., $q = 3$ implies $\alpha = (0.0082)m$, and a cross talk of 0.33 permits 40-input gates to be used. Also note that use of a more appropriate aperture function could permit smaller values of q.

To estimate the SBWP that can be written onto a CGH, we assume the CGH is written using electron-beam lithography, as was the case for the experimental demonstration of the optical logic system.[14] This electron-beam system has written linewidths down to 0.5 μm and has a maximum file size of 1.024 mm on a side. Files can be stitched together to yield a maximum size of 10 cm on a side. If we minimize the effect of the stitch error by making the file boundaries coincident with subhologram boundaries, an effective SBWP of 4×10^{10} is attainable.

The hologram coding parameter p, defined in Eq. (1), for the case of a Burckhardt hologram,[32] has a minimum value of 3 assuming square cells. Having found that q will typically be in the range of 2–3, we conclude that the maximum feasible number of gates corresponds to a value of pq of the order of 10. From Eq. (2) and the above SBWP, we find that the gate array dimension is $N \times N$ where

$$N \approx 100\text{-}200$$

for space-variant interconnections. Because this is less than the SBWP capabilities of some spatial light modulators, the CGH is the limiting element.

Since the space-variant system allows arbitrary interconnections, the only other possible limitation on the circuits that can be implemented is the requirement that all gates must perform the same binary operation, e.g., NOR in this case. However, since all the Boolean operations may be constructed out of NOR gates, this does not limit the types of processing operations that can be performed. Another feature is that circuits with any degree of inherent parallelism, or lack thereof, can be implemented with approximately equal ease.

V. Space-invariant Interconnection Method

If one is willing to compromise on the arbitrariness of the gate interconnections, a substantial increase in the possible number of gates results. The extreme case is a totally space-invariant interconnection. This is the idea behind the interconnections of the processor suggested by Huang.[12] We implement this interconnection method optically via an imaging system with a space-invariant filter using one simple hologram for the entire circuit (Fig. 4). The filter has an impulse response consisting of a series of spots which illuminate the appropriate gate inputs as in the space-variant case. However, in this case, the impulse response (interconnection pattern) is the same for every gate output, and the gate inputs are addressed relative to the position of the gate output. The space-variant method worked on the basis of absolute addressing.

An example of a space-invariant interconnection pattern is shown in Fig. 5. Each dot in the figure represents a (NOR) gate, and each arrow represents an interconnection from the output of one gate (dot) to the input of another. Each gate is considered to have one additional unconnected input line for an enable/disable

signal. A particular circuit is implemented by disabling the appropriate gates. In the NOR case, a gate is disabled by projecting light onto it (i.e., putting a 1 onto the unconnected input line). With the illustrated interconnection pattern it is possible to transfer data in various directions without getting unintended feedback loops. The major limitation of this interconnection method is that the implementation of many circuits will require a large number of gates to be disabled. Obviously, some circuits can utilize the gates more efficiently than others.

As an example, a good utilization of gates can be achieved for some digital systems that perform parallel image processing operations over an array of image points. In this discussion, we define image point as a resolution element of an image to be processed, while we use the term pixel to refer to a resolution element on the nonlinear device that forms one gate. Parallel image processing systems generally require several gates per image point. We refer to the set of gates associated with an image point as a logic element. Here we consider the important case of identical logic elements. Let the number of gates per logic element be denoted by r.

When r is equal to three, each logic element has three gates, and we label these gates x, y, and z. A possible physical layout for all the individual gates in the processor is shown in Fig. 6, where each gate is denoted by a black dot. The active device area contains three identical subarrays of gates. The x subarray is a collection of every gate in the processor labeled x, one from each logic element. Similarly, the y and z subarrays are composed of all the y and z gates, respectively. Thus the number of gates in each subarray is equal to the total number of logic elements (which is equal to the total number of image points to be processed in parallel). The number of subarrays is equal to r. All gates are identical and communicate with each other in a space-invariant manner. With up to three gates per image point, interconnections within a logic element can be chosen arbitrarily, and interconnections to and from a logic element can be chosen almost arbitrarily. Once these interconnections are chosen, all logic elements are connected identically.

Figure 6 shows a (space-invariant) point spread function that connects the output of an x gate to the input of the corresponding z gate (within-logic element connection) and also to the input of the y gate of the logic element located two positions down in the array (between-logic element connection). Because of the locations of the gates on the device, this point spread function connects the outputs of the y and z gates to points that are either off the device or in an unused area of the device; thus these connections have no effect.

One restriction on the between-logic element interconnections is that the output of one gate cannot be connected to the same gate of a different logic element, e.g., x gate to x gate, without also connecting the y gate to y gate and z gate to z gate. The only other requirement is on the directions of long between-logic element interconnections. Because of boundary effects, there

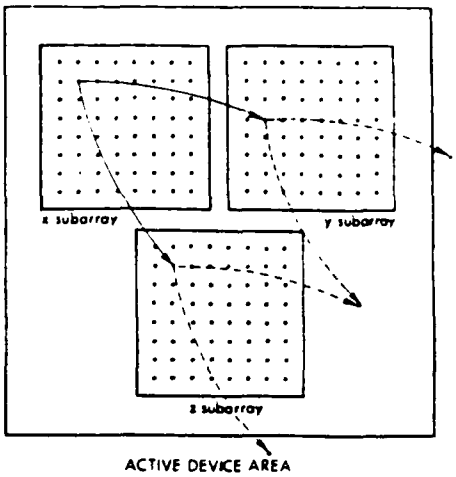


Fig. 6. Example of a system with space-invariant interconnections for parallel image processing operations. The system has three gates per image point: x, y, and z. These three gates are called a logic element. The gates are represented by dots in the figure, and their physical layout on the active device area is shown. The x subarray is composed of all the x gates in the processor with similar subarrays for the y and z gates. This configuration allows arbitrary interconnections within logic elements and also achieves a high gate utilization.

must be an unused area, or buffer zone, separating the x, y, and z subarrays. This buffer zone is small when connections between logic elements in certain directions are limited in length (one gate distance for nearest neighbor connections) but can become quite large when distant logic elements are connected in certain directions. Distant logic elements can be connected in other directions, however, without requiring large buffer zones. With four or more gates per logic element ($r \geq 4$) some restrictions apply on the interconnections both within and between logic elements. A set of interconnection rules can be enumerated for a given $r \geq 4$. We also note that the subarrays in Fig. 6 need not be square but can be other shapes to better fill the active device area.

These ideas are further illustrated with a simple example that uses three gates per image point as shown in Fig. 7. Each image point is labeled E, and each is connected to its four nearest neighbors: A, B, C, and D. This circuit implements a binary version of a Laplacian edge detector—if the value of image point E is 1, and at least one of its nearest neighbors is 0, 1 is output at image point E. Otherwise, zero is output. The point spread function (PSF) that implements this circuit is depicted in Fig. 8. As this PSF is convolved with the gate array, all interconnections are accounted for.

Since the holographic element used in this interconnection system is simple, a very large number of gates can be interconnected. Even allowing the PSF to address simultaneously any set of gates in the array, the space–bandwidth product required of the hologram is of the order of $p^2q^2N^2$ [see Eq. (1)]. Thus if the full SBWP available with the CGH could be exploited in this system and if $p^2q^2 \sim 100$, $\sim 4 \times 10^5$ gates could be

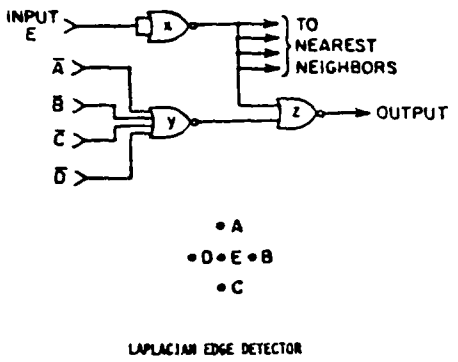


Fig. 7. Example of a circuit with three gates per image point which could be implemented with the space-invariant arrangement of Fig. 6. This circuit implements a binary Laplacian edge detector on an image. All image points are processed in parallel. Each image point is denoted by E, and each is connected to its four nearest neighbors: A, B, C, and D.

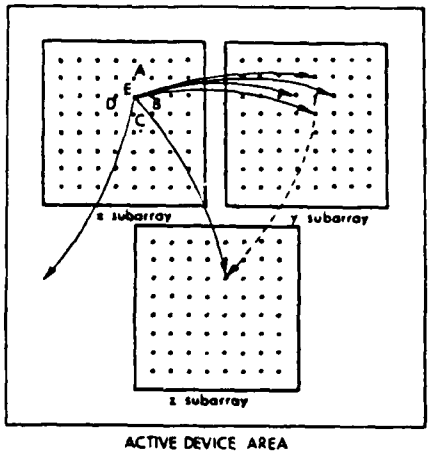


Fig. 8. Space-invariant interconnection pattern required for the binary Laplacian edge detector of Fig. 7.

interconnected. The hologram for this system could also be recorded optically. In either case, the number of gates with the space-invariant interconnection method is limited by the SBWP of the spatial light modulator.

As mentioned above, the method of disabling gates to implement circuits decreases the number of gates that are actually used and, therefore, severely restricts the variety of operations that can be performed efficiently. It also adds a degree of complexity to the system. However, this method of optically disabling gates also provides a potential advantage—it provides a means of easily rewiring the system in real time by changing the disable signals. This could offer considerable flexibility in making an adaptive system.

VI. Hybrid Interconnection Method

At this point we have seen two approaches to interconnecting gates. In the space-invariant case there is only one interconnection pattern which is applied to all

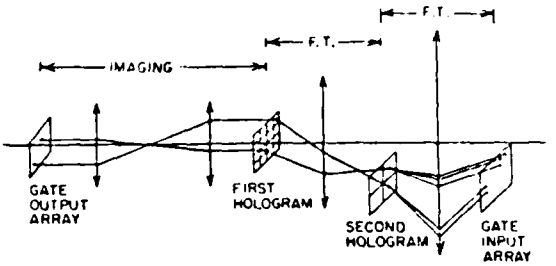


Fig. 9. Hybrid interconnection system. The first hologram is a space-variant element as in Fig. 3. The second hologram is an array of space-invariant filters.

gates, whereas in the space-variant case the number of distinct interconnection patterns is in general large and potentially equal to the number of gates. These two approaches represent the extreme cases in terms of the space–bandwidth requirements they place on the CGH. The trade-offs between these two is increased flexibility at the cost of increased hologram complexity. Since the space-invariant case generally suffers from inefficient gate utilization and the space-variant system is limited by the hologram to the number of gates it can address, it is worthwhile considering if there is a combination of techniques which can achieve high gate utilization efficiency and at the same time be limited in gate count only by the space–bandwidth product limitations of the spatial light modulator.

Our approach to this has been to consider a hybrid system which combines space-variant and space-invariant interconnections. The idea is to define a finite number M of distinct interconnection patterns. We then assemble our circuit using only these M interconnection patterns. If the total number of gates is N^2, we assume

$$1 \ll M \ll N^2 \tag{10}$$

so that this system is truly intermediate between the space-variant and space-invariant cases. If M is large, we anticipate that we have almost complete flexibility in designing our circuit.

The optical implementation of this system is schematically diagrammed in Fig. 9. Here the gate output array is imaged onto a space-variant filter element as in Fig. 3. The purpose of this element is to deflect the light from each gate output through one of M subholograms in the second CGH (Fig. 9). These subholograms act as space-invariant filter elements which produce the M different interconnection patterns in the gate input plane.

Although the space-variant element would appear to have the same space–bandwidth limitations as in the simple space-variant case, we note that the SBWP of each subhologram in this plane is now of the order of M rather than of the order of N^2. Thus the total SBWP requirement in this element is much less than in the previous space-variant case. The subholograms in the space-invariant element generally have a relatively low SBWP.

The SBWP S_{S_1} of a subhologram in the first holo-

gram H_1 is

$$S_{S_1} = p_1^2 q_1^2 M, \tag{11}$$

where M is the number of subholograms in the second hologram H_2, and p_1^2 and q_1^2 represent coding and oversampling factors, respectively, as in the space-variant interconnection section. Similarly, the SBWP of a subhologram of H_2 is in the worst case

$$S_{S_2} = p_2^2 q_2^2 N^2, \tag{12}$$

where N^2 = number of gates. This worst case allows the H_2 subhologram to address any n gates in the array. If the gates it addresses are localized, i.e., are all contained in the same portion of the array, its SBWP can be significantly reduced by introduction of a carrier frequency.[33] Since H_1 consists of N^2 subholograms and H_2 consists of M subholograms, their total SBWPs are given by

$$S_{T_1} = p_1^2 q_1^2 M N^2 \qquad S_{T_2} = p_2^2 q_2^2 M N^2, \tag{13}$$

and here again S_{T_2} is a worst-case estimate. If we assume both holograms are written in the same manner, $S_{T_1} = S_{T_2} = S_T$ and $p_1 = p_2 = p$, from which it follows that $q_1 = q_2 = q$; thus

$$S_T = p^2 q^2 M N^2. \tag{14}$$

As in the space-variant case, we need to analyze the cross talk to estimate q.

For the hybrid interconnection scheme, two sources of cross talk exist. Interpixel cross talk occurs between pixels in the gate-input plane and is analogous to the cross talk treated in the space-variant interconnections section. Intrahologram cross talk occurs between subholograms in the second hologram and also contributes to noise in the gate-input plane. (We assume negligible cross talk at the first hologram because it is in the image plane of the gate-output array.)

The interpixel cross talk is completely analogous to the cross talk in the space-variant interconnection case when applied to H_2 and the gate-input array, and the same equations apply.

To analyze the intrahologram cross talk, we have to find the effect of this cross talk in the gate-input plane. We assume that each H_1 subhologram addresses only one H_2 subhologram and that all H_2 subholograms have the same fan-out n. Through a given H_2 subhologram k, there are only n subholograms of H_1 that can address a given gate input ρ. Any unintentional illumination of k from one of these H_1 subholograms will contribute cross talk to gate ρ. The worst-case (maximum) zero input to ρ then occurs when all n of these H_1 subholograms illuminate the nearest-neighbor subholograms of k. We then sum this result over all H_2 subholograms k to obtain the worst-case zero-level input,

$$I_0^{(H)} = \sum_{k=1}^{M} n \iint_{\sigma_{NN}} g(x,y)\,dx\,dy, \tag{15}$$

where $g(x,y)$ is the intensity profile divided by n in the H_2 plane due to the illumination of one H_2 subhologram by one H_1 subhologram and is analogous to $F(x,y)$ in our previous derivation. σ_{NN} is a nearest-neighbor subhologram to the illuminated H_2 subhologram. All additions here are incoherent because we are effectively averaging over fringe patterns again. The worst case one-level gate input is simply the integral of $g(x,y)$ over the illuminated subhologram σ, thus the intrahologram cross talk α_H is

$$\alpha_H = \frac{I_0^{(H)}}{I_1^{(H)}} = \frac{nM \iint_{\sigma_{NN}} g(x,y)\,dx\,dy}{\iint_{\sigma} g(x,y)\,dx\,dy}. \tag{16}$$

The total cross talk α is the sum of contributions to the worst-case zero-level gate input from both sources of cross talk divided by the worst-case one-level gate input. To first order in α_P and α_H it is given by

$$\alpha = \alpha_P + \alpha_H, \tag{17}$$

where α_P is the interpixel cross talk.

Now we look at an example. Again taking a 2-D triangle function as the subhologram aperture function in the H_1 plane, $g(x,y)$ in the H_2 plane is given by $F(x,y)$ of Eq. (9). Now the intensity profile $F(x,y)$ in the gate input plane is the squared modulus of the convolution of this triangle aperture function with the sinc function resulting from the rect aperture function in the H_2 plane. Taking, for example, $q = 3$ and calculating α_P from Eq. (6) and α_H from Eq. (16), Eq. (17) yields

$$\alpha = m(0.0025) + nM(0.0018), \tag{18}$$

where m is the fan-in to each gate. For example, if eight-input gates are used and we allow $\alpha = 0.50$, we can have fifty-three different interconnection patterns each with a fan-out of five. These numbers are strongly dependent on the aperture function. Since no attempt to optimize, or even improve on, the aperture function was made, one can expect a significant improvement in these numbers by using a more appropriate aperture function.

Thus, as in the space-variant case, we conclude that the maximum number of gates corresponds to a value of pq of the order of 10. Again using $S_T = 4 \times 10^{10}$ (see Sec. IV), from Eq. (14) we get $MN^2 \approx 4 \times 10^3$, and thus for $M \approx 50$ different interconnection patterns, we get of the order of 10^7 gates or an $N \times N$ array with $N \approx$ 2000–3000, which is above the SBWP capabilities of presently available spatial light modulators.

To implement an arbitrary circuit with the hybrid interconnection method, the M interconnection patterns may be considered a basis set from which one constructs the desired interconnections. With a large enough M, any circuit can be implemented. However, the potential of this architecture can be exploited more fully by implementing circuits with a high degree of regularity or symmetry. Examples include parallel arrays, such as systolic arrays and cellular automata. As mentioned earlier, the principal motivation behind the systolic array concept is the adaption to the limitations of VLSI structures.

A more appropriate application of our hybrid interconnection scheme is that of 2-D, or more generally N-dimensional, processor arrays. A conceptual di-

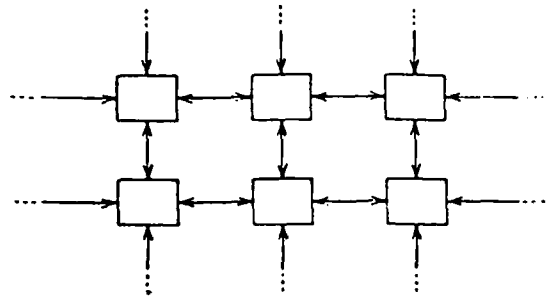


Fig. 10. Two-dimensional processor array as an example of the use of the hybrid interconnection system of Fig. 9. Connections within each block are space-variant, and connections between blocks are space-invariant.

agram of a 2-D array is shown in Fig. 10. Each block represents a processing element made up of a number of gates connected in some arbitrary manner. Each block may be identical to the others or different, although the more similarities there are between blocks the fewer interconnection patterns are needed. The connections between blocks will typically be space-invariant. Although the figure depicts only nearest-neighbor connections between blocks, much more complicated interconnection patterns, without regard to the physical distance between connected blocks, may be made almost as easily, permitting implementation of N-dimensional arrays. Thus we have a processor array with global as well as local interconnections.

This concept may also be applied to computational algorithms that require global interconnections, which are often difficult or inefficient to implement using VLSI. In this case, the blocks in the figure are often identical, but the interconnections between the blocks are space-variant. Frequently, however, as in the case of the FFT mentioned earlier, there is still a high degree of symmetry among the connections between blocks. Thus this class of algorithms is also particularly well suited to the hybrid interconnection scheme.

VII. Conclusion

We have described some of the architecture and design considerations which affect the implementation of sequential logic in an optical system.

Whereas in VLSI systems, communication costs are the major limiting design factor, in the optical system, the communication costs are much less severe. Thus many architectures which may be impractical in VLSI due to communication problems may be implemented easily in an optical system.

The interconnection network can be produced in several different ways. We discussed three methods: a space-variant technique which allows arbitrary interconnections but is limited by the SBWP of the CGH; a space-invariant system which requires only simple CGHs but suffers from low gate-utilization rates; and a hybrid of these two techniques which provides large gate counts with high gate-utilization rates.

The major limitations of the system at this point relate to the device used as a gate array. In particular, the speed of current devices is a major bottleneck. Current work in optical bistability may alleviate that restriction, although significant progress will have to be made before that technology develops into functional devices.

We thank B. H. Soffer and S.-Y. Kung for many useful discussions in connection with this research. This work was supported by the Air Force Office of Scientific Research under grant AFOSR-81-0182.

Portions of this paper were presented at the Tenth International Optical Computing Conference.[34]

References

1. A. Armand, A. A. Sawchuk, T. C. Strand, D. Boswell, and B. H. Soffer, "Real-time Parallel Analog-to-Digital Conversion," Opt. Lett. 5, 129 (1980).
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," Opt. Lett. 5, 398 (1980).
3. A. A. Sawchuk, T. C. Strand, and A. R. Tanguay, Jr., "Nonlinear Real-time Optical Signal Processing," USCIPI Report 1080, USC Image Processing Institute (1982).
4. S. H. Lee, Ed., *Optical Information Processing* (Springer, New York, 1981), Chap. 7.
5. S. A. Collins, Jr., M. T. Fateh, and K. C. Wasmundt, "Optical Logic Gates Using a Hughes Liquid Crystal Light Valve," Proc. Soc. Photo-Opt. Instrum. Eng. 232, 168 (1980).
6. B. Horwitz and F. Corbett, "The PROM-Theory and Applications for the Pockels Readout Optical Modulator," Opt. Eng. 17, 353 (1978).
7. 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).
8. J. Tanida and Y. Ichioka, "Optical Logic Array Processing," *Proceedings, Tenth International Optical Computing Conference*, IEEE Catalog. 83CH1880-4 (1983), p. 18.
9. R. A. Athale, H. S. Barr, S. H. Lee, and B. J. Bartholomew, "Digital Optical Processing," Proc. Soc. Photo-Opt. Instrum. Eng. 241, 149 (1980).
10. D. H. Schaefer and J. P. Strong III, "TSE Computers," Proc. IEEE 65, 129 (1977).
11. A. Huang, "Parallel Algorithms for Optical-Digital Computers," in *Proceedings, Tenth International Optical Computing Conference*, IEEE Catalog 83CH1880-4 (1983), p. 13.
12. A. Huang, "Design for an Optical General Purpose Digital Computer," Proc. Soc. Photo-Opt. Instrum. Eng. 232, 119 (1980).
13. H. Barr and S. H. Lee, "A Digital Optical Processing System," in *Proceedings, Tenth International Optical Computing Conference*, IEEE Catalog. 83CH1880-4 (1983), p. 171.
14. B. K. Jenkins, A. A. Sawchuk, T. C. Strand, and B. H. Soffer, "Sequential Optical Logic Implementation," J. Opt. Soc. Am. 72, 1721A (1982); B. K. Jenkins, A. A. Sawchuk, T. C. Strand, R. Forchheimer, and B. H. Soffer, "Sequential Optical Logic Implementation," Appl. Opt. 23, 3455 (1984).
15. J. Grinberg, A. Jacobson, W. Bleha, L. Miller, L. Fraas, D. Boswell, and D. Myer, "A New Real-Time Non-Coherent to Coherent Light Image Converter: The Hybrid Field Effect Liquid Crystal Light Valve," Opt. Eng. 14, 217 (1975).
16. U. Efron, B. H. Soffer, M. J. Little, J. Grinberg, and M. A. Monahan, "Applications of Silicon Liquid Crystal Light Modulators for Optical Data Processing," Proc. Soc. Photo-Opt. Instrum. Eng. 388, 152 (1983).

17. H. M. Gibbs, S. L. McCall, and T. N. C. Venkatesan, "Optical Bistable Devices: The Basic Components of All-Optical Systems," Opt. Eng. 19, 463 (1980).
18. E. Abraham, C. T. Seaton, and S. D. Smith, "The Optical Computer," Sci. Am. 248, 85 (Feb. 1983).
19. P. W. Smith and W. J. Tomlinson, "Bistable Optical Devices Promise Subpicosecond Switching," IEEE Spectrum 18, 26 (June 1981).
20. C. R. Bowden, M. Ciftan, and H. R. Robl, Eds., *Optical Bistability* (Plenum, New York, 1981).
21. C. A. Mead and L. A. Conway, Eds., *Introduction to VLSI Systems* (Addison-Wesley, Reading, Mass., 1980), Chap. 8.
22. L. S. Haynes, R. L. Lau, D. P. Siewiorek, and D. W. Mizell, "A Survey of Highly Parallel Computing," Computer 15, 9 (Jan. 1982).
23. J. W. Tompkins and S. Hollock, "The Impact of Multi-Level Metalization on Semi-Custom LSI," in *Proceedings, Custom Integrated Circuits Conference* (1982), p. 276.
24. Ref. 21, C. L. Seitz, "System Timing," Chap. 7.
25. C. L. Seitz, "Ensemble Architectures for VLSI—A Survey and Taxonomy," in *Proceedings, 1982 Conference on Advances in Research in VLSI* MIT, Cambridge, 1982, p. 130.
26. H. T. Kung, "Why Systolic Architectures?" Computer 15, 37 (Jan. 1982).
27. D. M. DeRuyck, L. Snyder, and J. D. Unruh, "Processor Displacement: An Area-Tme Tradeoff Method for VLSI Design," in *Proceedings, Conference on Advances in Research in VLSI*, MIT, Cambridge, Mass. (1982), p. 182.
28. S.-Y. Kung, K. S. Arun, R. J. Gal-Ezer, and D. V. Bhaskar Rao, "Wavefront Array Processor: Language, Architecture, and Applications," IEEE Trans. Comput C-31, 1054 (1982).
29. Ref. 21, H. T. Kung and C. E. Leiserson, "Algorithms for VLSI Processor Arrays," pp. 271–292.
30. A. Rosenfeld, "Parallel Image Processing Using Cellular Arrays," Computer 16, 14 (Jan. 1983).
31. S.-Y. Kung, Electrical Engineering Department, University of Southern California; private communication.
32. C. B. Burckhardt, "A Simplication of Lee's Method of Generating Holograms by Computer," Appl. Opt. 9, 1949 (1970).
33. S. Lowenthal and P. Chavel, "Reduction of the Number of Samples in Computer Holograms for Image Processing," Appl. Opt. 13, 718 (1974).
34. P. Chavel *et al.*, "Architectures for a Sequential Optical Logic Processor," in *Proceedings, Tenth International Optical Computing Conference*, IEEE Catlog 83CH1880-4 (1983), p. 6.

Computer-generated holograms for space-variant interconnections in optical logic systems

B.K. Jenkins and T.C. Strand*

Image Processing Institute, University of Southern California,
Los Angeles, California 90089-0272

Abstract

Interest in optical logic systems has been revived recently for several reasons. One major reason is that optics provides a means for overcoming the communication bottlenecks which are beginning to limit the development of electronic systems both at the chip level and at higher levels. This stems from the fact that an optical system in effect processes a large number of independent parallel channels. However, in attempting to implement optical interconnection networks, one is quickly faced with the need to realize space-variant interconnections. In this paper we describe various means of achieving space-variant interconnections for optical logic systems. All of the techniques use computer-generated holograms (CGH) as combination beam-splitting and beam directing elements. The space-variance can be obtained with either a direct approach where a separate CGH element forms the interconnections for each logic element, or with a basis set approach which provides significant reductions in the CGH storage requirements. Examples of the different approaches are discussed.

Introduction

There is a recurring interest in the subject of space-variant optical processing. There are two major questions to be asked on this subject: a) How can you implement a space-variant system? and, b) What do you do with a space-variant system? In this paper we address both questions and indicate how computer-generated holograms play a role in this problem.

We will address the second question first by describing an example where space-variant imaging is required. An increasingly important problem today is making interconnections in digital logic systems. This is a particularly acute problem at the gate level of digital logic systems. Interconnection costs comprise the major factor in the design of very-large-scale integrated (VLSI) circuits [1]. Interconnections between chips and between subsystems are also an increasing problem due to the need for large numbers of high-bandwidth lines. Optics offers some potential advantages for making such interconnections. Fiber optic systems are useful where relatively small numbers of very high-bandwidth lines are required, such as between computers. When a large number of interconnections must be made, for example on a VLSI chip, other techniques must be found. Since an optical imaging system can be thought of as consisting of a large number of parallel communication channels, we might consider if it could be used for such interconnections. In fact if we want a fixed interconnection network connecting one array of nodes to another array in a one-to-one fashion, this can be accomplished by a simple imaging system. Gate level interconnections on a chip are not one-to-one however since typically one gate output will feed many other gates (fan-out) and a gate will generally have several inputs (fan-in). Furthermore, each gate will have a different interconnection pattern than its neighbors. If we consider the array of gate output nodes as object points and the gate input nodes as image points, then an optical interconnection system can be implemented via a space-variant imaging system where the point spread-function (psf) associated with a given object point is simply the desired interconnection pattern for the corresponding gate (see Fig. 1). Thus, any techniques developed for space-variant imaging are potentially applicable to the interconnection problem.

Before we discuss space-variant imaging techniques, it is worth noting the special features of the interconnection problem which may impact the choice of implementation for the space-variant system. The most striking feature of the interconnection problem is that the object and the image planes comprise discrete arrays of isolated points, the gate input and output nodes. Thus, we need only implement a discrete and finite set of psfs or interconnection patterns. Obviously, each psf also simply consists of a discrete set of points. As we will explain later, one implication of this restriction is the psfs can be conveniently encoded in a multi-facet holographic element [2], i.e., a discrete array of small holograms. Another important fact is that each psf must be exactly reproduced. This eliminates from consideration space-variant techniques which, in order to minimize storage requirements, approximate the psf.

*Present address: IBM Corp., K-46/282, 5600 Cottle Road, San Jose, CA 95193.

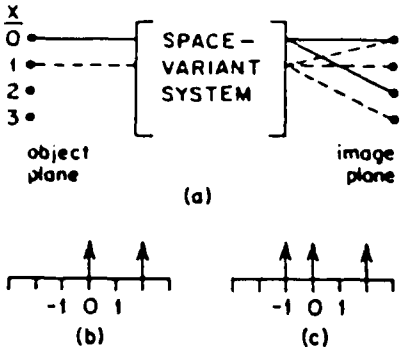


Fig. 1. (a) Optical interconnections implemented via a space-variant imaging system; (b) and (c) interconnection patterns, or psfs h(u-x;x) vs. u, (b) for x=0, and (c) for x=1.

The storage question is a major aspect of any space-variant system. In essence, one needs to store a separate psf for every point in the object. Thus for objects with a reasonably large space-bandwidth product (SBWP), the filter storage problems are severe. Most of the work in space-variant imaging consists of exploring methods of reducing the information storage requirements. In general there are three different problems to solve. First and foremost, one must reduce the information storage requirements to a minimum. Secondly, since there will still generally be a large number of filters to store, one needs an effective multiplexing scheme. Finally, a related problem is that of demultiplexing, or addressing of the stored information. Several different techniques have been developed to address these problems.

Past work in space-variant imaging has been described in two excellent reviews [3,4] to which the reader is referred for full details and extensive bibliographies. In the context of digital image processing, space variant imaging has been treated by Sawchuk [5], who used a coordinate transformation technique to turn certain space-variant problems into space-invariant problems, and by Robbins and Huang [6]. An early discussion of the problem in optical processing terms was given by Lohmann and Paris [7]. A considerable body of work has evolved concerning the use of holographic techniques for filter storage (e.g., references [8]-[10]). More recently, some new approaches have been suggested for one-dimensional or separable space-variance [11] and for two-dimensional space-variant processing [12].

In this paper we describe three different approaches to space-variant imaging for our special problem of digital logic interconnections. The first approach is a brute force technique, where a separate filter is stored for each point in the image. This straightforward approach is identical in concept to earlier work [e.g. [8]] although the implementation is different and the interconnection problem is a new context. Then we introduce the concept of a basis function approach which allows us to considerably reduce the storage requirements for the system [13]. Two basis set approaches are discussed. The first is very straightforward to implement but yields a suboptimal solution in terms of reducing the information storage requirements. We finally discuss an approach that is optimal in terms of storage requirements.

Space-variant imaging: direct implementation

As described above we will be discussing space-variant imaging in the context of interconnections in a digital logic systems. Although the logic system could be electronic, it could also be optical. Since we have been developing optical logic systems [14], we will use those systems as our basis for discussion.

Since again we are dealing with a discrete and finite number of object points, it is a simple matter to determine the psf for each point and store a separate filter for each point. This we refer to as the direct implementation.

In this paper, when used in the context of an interconnection network, the terms gate outputs, object points, and (interconnection network) input nodes are synonymous. Similarly, the terms gate inputs, image points, and output nodes are synonymous. The direct implementation of a space-variant interconnection network utilizes one multi-faceted hologram. Each facet, or subhologram, is used to store one psf. A schematic diagram of the optical system used for the direct implementation is shown in Fig. 2. First, the gate outputs are imaged onto the interconnection hologram. This hologram consists of an array of subholograms, one subhologram for each gate output. Since the hologram is encoded in the

Fourier transform plane, a Fourier transform is taken optically to obtain the gate inputs in the image plane. When illuminated by its corresponding gate output, a subhologram will reconstruct an image consisting of a simple dot pattern, each bright dot illuminating a gate input. Because the reconstructed images can be designed to illuminate any combination of gate inputs, arbitrary interconnections are possible. If these interconnection networks are to be used as part of an optical logic system, then the gate inputs can be used as inputs to a spatial light modulator, and the outputs of the spatial light modulator become the gate outputs. As shown in Fig. 2, the desired interconnections are formed in one particular diffraction order. Typically, a conjugate image will also be produced, in which case it can be used to probe the system without affecting system operation or to access the system outputs. Note that because we have a discrete array of gates we were able to form the connections with a discrete array of subholograms, or what is sometimes referred to as a multi-facet hologram [2].

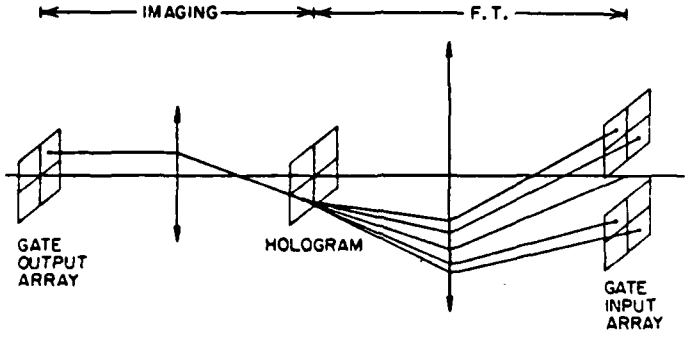


Fig. 2. Optical system for the direct implementation of space-variant interconnections. In general, the hologram produces multiple diffraction orders, only one of which is used.

As mentioned above, the storage requirements for the direct implementation are severe, as is revealed by the SBWP requirements on the hologram. Assuming an array of N gate outputs and an array of N gate inputs, there are N subholograms, and the total SBWP of the hologram is [14]

$$S_T = p^2 q^2 N^2 \tag{1}$$

where, in the case of a computer-generated hologram, p^2 is the number of resolution elements in the hologram used to represent one complex-valued sample and q^2 is a factor representing the amount of oversampling in the hologram plane. Generally $p^2>1$ because the complex sample values must be encoded into the hologram, e.g., as real values. Also we generally have $q^2>1$ to avoid crosstalk. It has also been shown [14] that the minimum feasible value of p^2q^2 is on the order of 100. The severity of the storage requirements can be seen just by noting that the SBWP is proportional to the square of the number of elements.

We will now illustrate the direct implementation of space-variant interconnections through the use of a simple example. A circuit consisting of NOR gates that form a master-slave flip-flop and a driving clock is shown in Fig. 3. A similar circuit has been implemented optically to demonstrate the operation of an optical sequential logic system [15]. Since this circuit has 16 gates, it can be laid out as a 4 x 4 array. A layout of the psfs onto the hologram is shown in Fig. 4(a), with each number representing the gate output that is imaged onto that subhologram. The array of gate outputs (object points) is the spatial inversion of this, because of the imaging relationship. The psf stored in each subhologram is shown in Fig. 4(b), with the ordered pairs representing the locations of the image points that each subhologram addresses. In this example, we have a 16-gate circuit and therefore must have 16 psfs and 16 subholograms.

We have generated a hologram, by computer, that implements the interconnection pattern and layout of Figs. 3 and 4. Many types of computer-generated Fourier transform holograms have been demonstrated [16] and many of them would suffice for this application. Here we used a hologram of the type proposed by Lee [17] for reasons of simplicity, diffraction efficiency, and signal-to-noise ratio in the reconstruction.

The hologram may be written onto film using a Dicomed film recorder or similar device, or may be written onto photoresist using electron-beam lithography. The latter method can write holograms with a much larger space-bandwidth product. An electron-beam system was used to create the hologram for this demonstration circuit. The device used has a step size

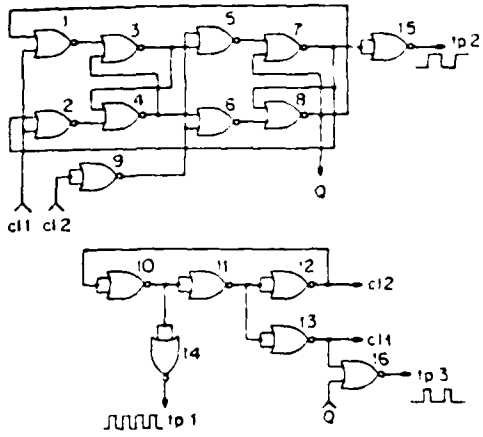


Fig. 3. Master-slave flip-flop and driving clock as an example of space-variant interconnections.

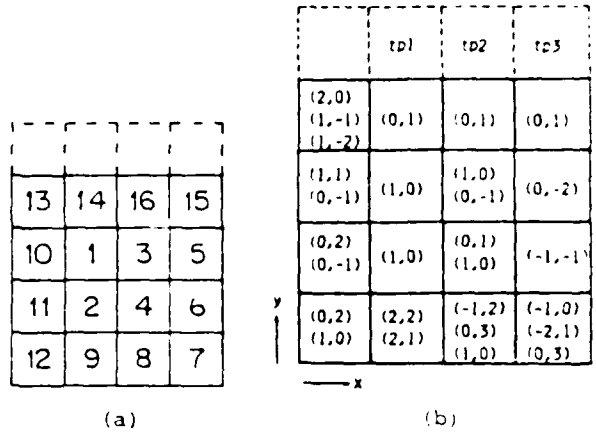
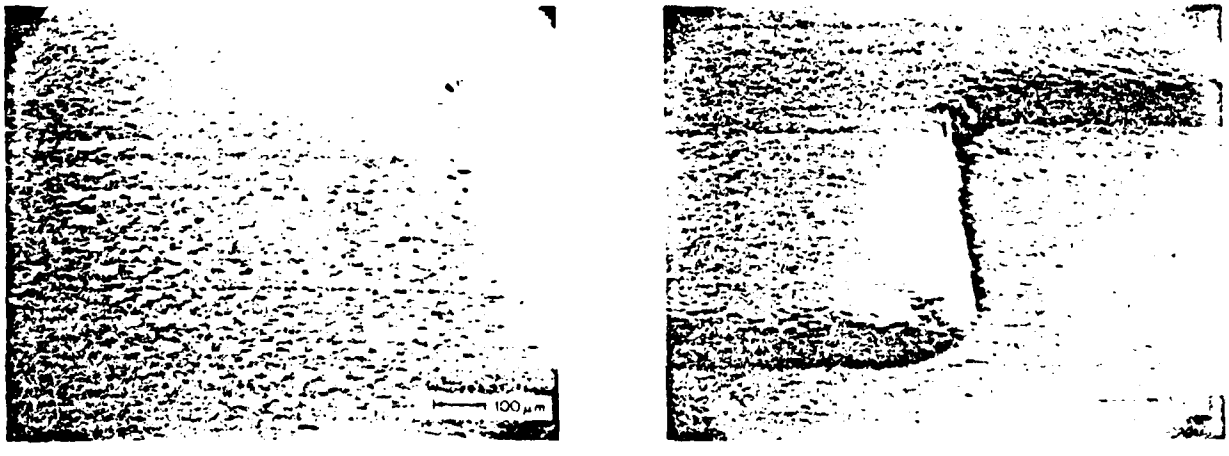


Fig. 4. Direct implementation of the interconnections of Fig. 3. (a) Layout of the psfs onto the hologram. (b) psf stored in each subhologram. The ordered pairs represent the locations of the image points addressed (relative to the subhologram location).

of 0.125 μm and has written patterns with linewidths as small as 0.5μm. It writes 1.024x1.024mm fields and can stitch them together to cover a maximum area of 102x102mm. The device provides a far greater space-bandwidth product than was needed for this demonstration circuit.

One absorption and two phase holograms of our demonstration circuit were made. This discussion will focus on the phase holograms, which use surface relief of the photoresist to obtain the optical path length difference. Figure 5 shows pictures of one subhologram, taken with a scanning electron microscope (SEM). Figure 5(a) shows an entire subhologram, whose diameter is 1.04mm, and Fig. 5(b) shows a close-up of a rectangle edge. The lower and right portions of Fig. 5(b) are rectangle interiors, which are just glass substrate. The rectangle exteriors consist of a layer of photoresist on top of the substrate. The thickness of the photoresist is 1.25μm and the edges are inclines at approximately 32 degrees with respect to the substrate normal. The pictures reveal that the photoresist is slightly rough near the edges but is otherwise quite smooth (except for an occasional defect). Defects are apparent on the glass but are too small to affect the optical quality. Linewidths down to approximately 1 micron were obtained.



(a) (b)

Fig. 5. Scanning electron microscope pictures of a phase CGH that directly implements the interconnections of Figs. 3 and 4.

Defining efficiency as the power in the desired reconstruction pixels due to one subhologram divided by the power incident on that subhologram, the maximum efficiency over all subholograms was measured to be 5%. The efficiencies of the other subholograms were intentionally reduced in order to normalize the intensities in the reconstruction plane. This measurement was taken using an illumination wavelength of 514.5 nm which is close to the optimum wavelength for this hologram. Besides efficiency, we must consider the noise appearing in the desired reconstruction order, of which there are three sources: (1) the encoding process used to represent the complex-valued function on the hologram, (2) scattering from the photoresist and glass substrate, and (3) the tail of the (0,0) order. The pictures of Fig. 5 indicate that the contribution due to scattering should be small, and this is verified by experiment. The effect of the (0,0) order could be substantial with a phase hologram, but can be filtered out spatially if the location of the limiting aperture is chosen appropriately. This leaves the encoding process as the major source of noise. The effect of the encoding process on noise is discussed in Ref.[18]. Measurement on our test hologram indicate a signal-to-noise ratio of approximately 60.

Basis set approach

The concept of basis functions for space-variant processing has been put forth by Marks and Krile [9] where a basis set decomposition was applied to the input signal. In a related approach, we chose to look at the space of point-spread functions which our system must implement, and then find a basis set for that space. We assume that the basis set will be finite, e.g., for systems with finite SBWP and limited variation such that a sampling theorem can be applied [19], or that all elements of interest can be adequately approximated with a finite basis set. The impulse response at any point is a weighted sum of the basis functions. The entire imaging system output is effectively a sum of space-invariant outputs, one for each basis element. The space-variance comes in through the spatially varying basis function coefficients. Thus, a space-variant system can be implemented using a limited number of filters without requiring any isoplanaticity [13]. The problems at hand involve storing the basis filters and encoding the basis function weighting coefficients in an object plane mask such that each coefficient effectively sees only the appropriate filter. These are essentially multiplexing/demultiplexing problems. Although there are several different multiplexing schemes available, we have chosen to use spatial multiplexing of the basis filters and spatial frequency encoding of the coefficients. These techniques will be described later.

At this point the question arises of how to obtain the basis set. For the logic circuit interconnections there is a simple technique which consists of enumerating the unique psfs or interconnection patterns and taking this as the basis set. This generally results in a suboptimal set. A minimum basis set can also be found. We discuss both techniques in the following subsections.

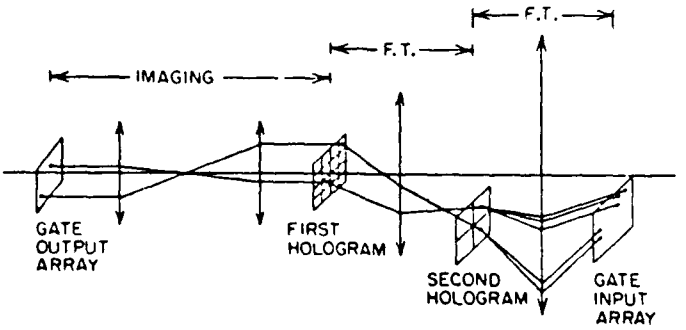


Fig. 6. Optical implementation of the basis set approach to space-variant interconnections. The second hologram stores the basis filters and the first hologram stores their coefficients for each gate output.

The optical implementation of each of these basis set techniques uses two holograms. Both techniques are implemented via the same optical system, which is diagrammed schematically in Fig. 6. The basis functions are spatially multiplexed onto the second hologram. Each facet, or subhologram, of the second hologram stores one basis element and functions as a space-invariant filter. The first hologram is also divided into an array of subholograms, one for each gate output. Each of these subholograms stores the weighting coefficients of the basis elements for the corresponding gate output.

Suboptimal basis set

As stated above, a convenient suboptimal basis set is obtained by enumerating the unique gate interconnection patterns. Thus, the first basis element is the interconnection pattern for the first gate. If the second gate has a different interconnection pattern, that pattern becomes the second basis element and so on through all the gates.

In this simple suboptimal case, all the weighting coefficients for a given gate output are 0 except for one coefficient, which has a value of 1. Thus, each subhologram in the first hologram directs the light from a gate output through one of the basis element filters.

In this case, the maximum total SBWP of each hologram is given by [14]

$$S_T = p^2 q^2 MN \tag{2}$$

where N is again the number of gates that are being interconnected and M is the number of basis elements. p and q are defined above. Thus, the SBWP is directly proportional to the number of elements in the basis set. In the limiting case of the number of basis elements being equal to the number of gates, the SBWP using the basis set approach is equal to the SBWP using the direct implementation.

The use of the suboptimal basis set approach can be illustrated using the same example that was used in discussing the direct implementation (shown in Figs. 3 and 4). As shown in Fig. 4(b), there are 16 psfs using the direct implementation. Some of these, however, are redundant. The psf that addresses (0,-1) appears in three places, and the psf that addresses (1,0) appears in two places. (All other psfs are unique.) In the suboptimal basis set approach we use a different subhologram for each different psf, thus requiring 13 subholograms (basis elements) instead of 16. In this case, three different subholograms in the first hologram would address the same (-1,0) element in the second hologram, for example. In many large circuits we expect this suboptimal basis set approach to provide a significant reduction in the required SBWP of the holograms when compared to the direct implementation.

As an example of the application of this approach to the interconnections in a large circuit, we consider the case of the fast Fourier Transform (FFT) algorithm. As one might expect, the butterfly structure seen in FFT signal flow graphs implies a high degree of interconnection redundancy. Here, we consider implementing the base 2 Cooley-Tukey algorithm using the suboptimal basis set approach. We lay out the circuit as the signal flow graph is typically drawn (e.g. Ref. [20]), only in place of each node we have a block of gates which perform a multiply on one of the inputs to the block and add the result to the other input. In this way, all the blocks can be identical. Let r be the number of different psfs inside one of these blocks. In the worst case, i.e., no redundancy of interconnections within each block, r will be equal to the number of gates per block minus one. Otherwise, r will be smaller. Thus, we need r psfs to represent the interconnections within all blocks. In order to connect the blocks (nodes) to each other, we note that each stage of the FFT requires only two different psfs. Since an n-point base 2 FFT algorithm has $\log_2 n$ stages, we need $2\log_2 n$ psfs to connect all the blocks to each other, and thus the total number of psfs required is

$$B = r + 2 \log_2 n \tag{3}$$

where B is the number of elements in this basis set. In this example, the computation of different stages is pipelined rather than done "in place." Thus, all nodes (blocks) in the signal flow graph are computed in parallel. For the purpose of comparison, we can calculate the number of psfs that would be required if we were to use the direct implementation. In this case we need one psf for each gate output. Let s be the number of gates per block. Since there are $n \log_2 n$ blocks (nodes), the number of psfs required for the direct implementation is

$$B = s\,(n \log_2 n), \quad s > r \tag{4}$$

In the case of the FFT algorithm, we thus see a major reduction in the number of psfs, and therefore in the hologram SBWP required, by going to a basis set approach.

Optimal basis set

The dimension of a space spanned by a set of vectors (in our case the vectors being the desired psfs), will be equal to, or less than the number of vectors in the original set. A basis set must exist where the number of basis elements is equal to the dimension of the space. Such a basis set represents a minimum spanning set for the space and as such is an

optimal basis set from the standpoint of having the fewest elements to store. Given a space spanned by a finite set of desired psfs, it is possible to find a (orthonormal) minimum spanning set by applying a Gram-Schmidt orthonormalization procedure to the initial set [21]. If, however, one wishes to utilize an incoherent system, one has to add a positivity constraint when searching for the spanning set. A positivity constraint generally precludes one from obtaining the minimum dimension spanning set but it still should be possible to reduce the number of required basis elements from the original number of psfs. Thus the dimension of the interconnection space should always be determined to ascertain the reduction one might obtain by searching for a (constrained) optimal basis set.

The optical implementation of the optimal basis set approach uses the same system as in the suboptimal case (Fig. 6), only the psfs are, in general, different. The SBWP requirements are given by Eq. (2), and thus any reduction in the number of psfs used as compared to the suboptimal case is accompanied by a corresponding reduction in the SBWP requirements on the holograms.

Table 1. Elements a_i of the optimal basis set for the interconnections of Figs. 3 and 4. Notation for the ordered pairs is the same as in Fig. 4.

a_1: (2,0),(1,-1),(1,-2)

a_2: (0,1)

a_3: (1,1)

a_4: (1,0)

a_5: (0,-1)

a_6: (0,-2)

a_7: (0,2)

a_8: (-1,-1)

a_9: (2,1),(2,2)

a_{10}: (0,3),(-1,2)

a_{11}: (0,3),(-1,0),(-2,1)

	tp1	tp2	tp3
a_1	a_2	a_2	a_2
a_3 a_5	a_4	a_4 a_5	a_6
a_7 a_5	a_4	a_2 a_4	a_8
a_7 a_4	a_9	a_{10} a_4	a_{11}

y ↑ , → x

Fig. 7. Optimal basis set implementation of the interconnections of Fig. 3. The layout is given in Fig. 4(a) and the desired psfs are given in Fig. 4(b). The basis elements (enumerated in Table 1) used for each subhologram are shown.

We can illustrate the application of the optimal basis set approach with, again, the example of the circuit of Fig. 3. Instead of merely eliminating the redundant psfs of Fig. 4(b), we define a new set of psfs that is the smallest set that can be used to implement all the interconnections. The resultant set of psfs is enumerated in Table 1, and the combination of these basis set elements that is used to implement each of the interconnection patterns is shown in Fig. 7. We see that 11 psfs are required using the optimal approach. In this particular example, the number of elements of the minimum spanning set under the positivity constraint is the same as that of the minimum spanning set without any constraints. (The set of psfs given satisfy the positivity constraint, since the interconnection patterns are formed by adding, but never subtracting, psfs.)

Discussion

We have discussed three methods of implementing space-variant interconnections using computer-generated holograms. As a means of comparing these techniques, we consider some examples. One example, that of the circuit of Fig. 3, has been studied above. The circuit consists of 16 gates. We saw that a direct space-variant implementation of the interconnections requires 16 psfs, whereas the suboptimal basis set implementation requires 13 psfs, and the optimal basis set implementation further reduces the number of psfs required to 11. As an example of a large circuit, we discussed the case of the Cooley-Tukey FFT algorithm, and found a major reduction in the number of psfs required in going from the direct implementation to the suboptimal basis set approach. In the optimal case, no further reduction in the number of psfs required for the internode connections will be achieved. For the interconnections within each node (block), the amount of reduction obtained, if any, will depend on the particular circuit and layout that is used to implement the multiply and add operations.

Since the gain obtained in each successively more complex implementation technique is completely circuit (and layout) dependent, we compare the results of one more example. This time we use a practical, medium-size circuit - that of a 4-bit arithmetic logic unit (ALU), manufactured in integrated circuit form by Texas Instruments, type number SN74S381 [22].

This example uses the circuit schematic given in the reference except for two minor changes: first, the circuit is implemented solely with NOR gates, and second, the fan-out is limited to 10. The circuit consists of 112 gates laid out in a 4x28 array. 105 of the gates have outputs that are connected to other gates. Since the other 7 gates furnish the circuit outputs, their interconnection patterns are unknown and are therefore omitted in this discussion. Thus, the direct implementation would require the use of 105 psfs. Making use of the interconnection redundancies , wers this requirement to 38 psfs for the suboptimal basis set implementation. Optimizing this basis set reduces the number of basis elements by one, requiring 37 psfs. In this example, this is the minimum basis set both with and without positivity constraints.

For a given circuit and a given layout, the reduction in the SBWP obtained in going from a direct implementation to the suboptimal basis set approach can be ascertained fairly easily by enumerating the psfs and counting the redundancies. We expect that for large circuits, the reduction in many cases will be significant (as in the ALU example), and in some cases will be dramatic (as in the FFT example). Whether the optimal basis set approach will provide a significant reduction over the suboptimal approach is not so easily answered. Examples can be concocted in which the optimal basis set is considerably smaller than the suboptimal basis set, but in the examples considered in this paper only minor reductions were obtained.

Conclusion

In conclusion, we have discussed three methods of implementing space-variant interconnections in optical logic processors via computer-generated holograms. The first method is a direct implementation that is conceptually simple but inefficient in terms of SBWP utilization of the hologram. The second method is a suboptimal basis set approach that utilizes redundancies in the interconnection patterns to reduce the SBWP requirements. It is relatively easy to implement and, depending on the circuit and its layout, can make significantly better use of the hologram SBWP. Finally, the suboptimal basis set can be optimized. This technique is significantly more difficult to implement at this point but does minimize the SBWP of the holograms given the basis set approach.

Acknowledgments

The authors wish to thank G.L. Best for many useful discussions on the subject of space-variant systems. The CGHs used in this work were fabricated at Rockwell International by R. Imerson and A.B. Jones. The authors gratefully acknowledge their assistance. This work has been supported by the Air Force Office of Scientific Research under grant AFOSR-81-0082 and by the National Science Foundation under grant ECS-7919691.

References

1. C.A. Mead and L.A. Conway, eds., Introduction to VLSI Systems, Addison-Wesley, Reading, Mass. (1980), Ch. 8.

2. S.K. Case, P.R. Haugen, and O.J. Lokberg, "Multi-Facet Holographic Optical Elements for Wavefront Transformations," Appl. Opt., vol. 20, pp. 2670-2675 (1981).

3. J.W. Goodman, "Linear Space-Variant Optical Data Processing," in Optical Information Processing, S.H. Lee, Springer-Verlag, Berlin, pp. 235-260 (1981).

4. W.T. Rhodes, "Space-Variant Optical Systems and Processing," in Applications of Optical Fourier Transforms, H. Stark, ed., Academic Press, N.Y., pp. 333-370 (1982).

5. A.A. Sawchuk, "Space-Variant Image Motion Degradation and Restoration," Proc. IEEE, vol. 60, pp. 854-861 (1972).

6. G.M. Robbins and T.S. Huang, "Inverse Filtering for Linear Shift-variant Imaging Systems," Proc. IEEE, vol. 60, pp. 862-872 (1972).

7. A.W. Lohmann and D.P. Paris, "Space-Variant Image Formation", J. Opt. Soc. Am., vol. 55, 1007 (1965).

8. L.M. Deen, J.F. Walkup, and M.O. Hagler, "Representations of Space-Variant Optical Systems Using Volume Holograms," Appl. Opt., vol. 14, pp. 2438-2446 (1975).

9. R.J. Marks II and T.F. Krile, "Holographic Representation of Space-variant Systems: System Theory," Appl. Opt., vol. 15, pp. 2241-2245 (1976).

10. R.J. Marks II, "Two-dimensional Coherent Space-variant Processing Using Temporal Holography: Processor Theory," App. Opt, Vol. 18, p. 3670 (1979).

11. A.W. Lohmann and N. Streibl, "Map Transformations by Optical Anamorphic Processing," Appl. Opt., vol. 22, pp. 780-783 (1983).

12. R. Bamler and J. Hofer-Alfeis, "2D Linear Space-Variant Processing by Coherent Optics: A Sequence Convolution Approach," Opt. Commun., vol. 43, pp. 97-102 (1982).

13. G.L. Best and T.C. Strand, "Space-Variant Image Processing," to be published.

14. P. Chavel, et al., "Architectures for Sequential Optical Logic," Proc. Tenth Intl. Optical Computing Conf., IEEE Catalog No. 83CH1880-4, pp. 6-12 (1983).

15. B.K. Jenkins, A.A. Sawchuk, T.C. Strand and B.H. Soffer, "Sequential Optical Logic Implementation," J Opt. Soc. Am., vol. 72, 1721A (1982); "Sequential Optical Logic Implementation," to be submitted to Appl. Opt..

16. W.J. Dallas, "Computer-Generated Holograms," in The Computer in Optical Research: Methods and Applications, B.R. Frieden, ed., Springer-Verlag, New York, pp. 291-363 (1980).

17. W.H. Lee, "Sampled Fourier Transform Hologram Generated by Computer," Appl. Opt., vol. 9, p. 639 (1970).

18. J. Bucklew, N. Gallagher, Jr., "Comprehensive Error Models and a Comparative Study of Some Detour-Phase Holograms," Appl. Opt., vol. 18, p. 2861 (1979).

19. R.J. Marks II, J.F. Walkup, and M.O. Hagler, "A Sampling Theorem for Space-Variant Systems," J. Opt. Soc. Am., Vol. 66, pp. 918-921 (1976).

20. E.O. Brigham, The Fast Fourier Transform, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1974), p. 175.

21. T.M. Apostol, Calculus, Vol. II, Second Edition, Xerox College Publishing, Waltham, Mass. (1969), pp. 22-26.

22. Texas Instruments, Inc., The TTL Data Book for Design Engineers, Second Edition, Texas Instruments, Inc., Dallas, Texas (1977), p. 7-485.

Digital Optical Computing

ALEXANDER A. SAWCHUK, SENIOR MEMBER, IEEE, AND TIMOTHY C. STRAND, MEMBER, IEEE

Invited Paper

This paper concerns binary digital computing systems in which the information-carrying medium consists entirely or primarily of photons. The paper begins with a review of analog, discrete, and binary methods of representing information in a computer, followed by a survey of many techniques for implementing binary combinational and sequential logic functions with individual optical devices and arrays of devices. Next is a discussion of communication, interconnection, and input–output problems of digital electronic and optical computers at the gate, chip, and processor level. A particular architecture for implementing general sequential optical logic systems including digital optical processors is described. This architecture avoids some of the interconnection problems of electronic integrated circuits and VLSI systems, and offers the potential of non von Neumann parallel digital processors. Finally, the current limitations and future needs of optical logic devices and digital optical computing systems are outlined.

I. Introduction

Over the last twenty years there has been considerable research effort in the general area of information processing by optical techniques [1]–[8]. More recently, the term *optical computing* has been applied to parts of this work. We define optical computing as "the use of optical systems to perform numerical computations on one-dimensional or multidimensional data that are generally not images." In a broad sense, optical computing could include systems that use optics to form images (the most common application of optical components and systems) or even the use of a computer to design, simulate, or analyze optical systems.

There are many potential advantages of optical computing systems, and also many disadvantages. Optical systems can have high space–bandwidth and time–bandwidth products; hence, many independent channels (degrees of freedom) that could be exploited for demanding computations. Optical processors are inherently two-dimensional and parallel. Optical signals can propagate through each other in separate channels with essentially no interaction, and can propagate in parallel channels without interference and crosstalk. Optical signals can also interact on a subpicosecond time scale via an intermediary medium, offering the potential for high throughput. In addition, optical systems may be cheaper than equivalent digital systems for certain signal processing applications. As described later in this paper, these advantages have very promising implications for many traditional signal processing problems in which digital electronic hardware is inadequate. Some areas of digital computing that currently overburden existing technologies include image processing, radar signal processing, image analysis, machine vision, and artificial intelligence. This paper discusses these advantages in detail along with disadvantages and limitations of various optical computing methods.

The eventual goal of the work described here is to perform binary digital computing with optical systems, using photons as the primary information-carrying medium, avoiding electronic logic, and having as few photon–electron and electron–photon conversions as possible. We refer to such systems as digital optical processors (DOPs). Although DOPs may never be as universal in their application as electronic digital computers, we will describe research results and design concepts for DOPs that are programmable and reconfigurable to be usable in a large variety of optical computing applications.

This paper is intended as a review of the current state of the art in digital optical computing, with the emphasis on recent work at USC. Section II reviews previous work on optical computing with analog and multilevel processors. Some advantages and limitations of these systems are given. Section III discusses individual binary logic devices (gates) and arrays of devices needed for DOPs. We make a comparison with electronic digital devices and integrated optics. We describe several techniques for achieving optical combinatorial binary logic functions. Section IV is concerned with communications and input–output problems of digital processors, both within the processor (on-chip) and among various processors (off-chip). Recent research on a sequential optical digital processor with a non von Neumann free-interconnection architecture is described. Section V describes new types of computing architectures and algorithms made possible with optical processors. Section VI describes physical and technological limitations of electronic and optical computers. Section VII briefly discusses future needs of optical computers.

II. Optical Computing for Analog and Multilevel Techniques

In this section, we briefly review analog and multilevel optical computing. This short review is not intended to be an exhaustive account of past research and the reader is referred to the literature for more complete surveys [3]–[10]. The paper by Jahns [10] has particularly good examples that compare analog, multilevel, and binary arithmetic and also

Manuscript received March 8, 1984. This work was supported by the Air Force Office of Scientific Research under Grants AFOSR-81-0182 and AFOSR-83-0185.

A. A. Sawchuk is with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA.

T. C. Strand was with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA. He is now with IBM Corporation, San Jose Research Laboratory, San Jose, CA 95193, USA.

0018-9219/84/0700-0758$01.00 ©1984 IEEE

discusses tradeoffs in computational accuracy versus parallelism and other factors.

A. Optical Analog Computing

We can think of a general computation as an operation, θ, applied to a set of input functions, $f_i, i = 1, \cdots, N$ to produce a set of output functions, $g_j, j = 1, \cdots, M$ in the form

$$\theta[f_1(x), \cdots, f_i(x), \cdots, f_N(x)] = [g_1(y), \cdots, g_j(y) \cdots, g_M(y)]. \quad (1)$$

Here x and y are the sets of independent variables associated with the input and output functions, respectively.

We define an analog system as one where the dependent variables f_i and g_j can assume any of a continuum of values. If the dependent variables are restricted to a specific set of discrete values, the system will be called digital. The independent variables x and y can, of course, also be analog in that they may take on any one of a continuum of values, in which case the system will be referred to as continuous. If only discrete values of the independent variables are allowed, we say that the system is sampled.

Historically, most of the work in optical computing has been in the analog, continuous domain. Typically, in optical systems, the dependent variables are either complex amplitudes or intensities. In order to utilize the parallel processing capabilities of optics, the independent variables are generally two orthogonal spatial coordinates. The simplest and most common example of this type of common operation is optical imaging where a two-dimensional intensity (or amplitude) pattern is in essence transmitted from one location in space to another. Operations of the analog, continuous variety have been widely developed over the last twenty years. There exists a fairly well-defined repertoire of operations which can be performed, including such important operations as addition, subtraction, and multiplication of images, two-dimensional Fourier transformation, correlation, convolution, and other operations derivable from these. For a further exposition of these operations the interested reader is referred to the literature [1], [3]–[6], [9], [10].

There are four major drawbacks to optical analog processing. These drawbacks are limited flexibility, noise accumulation, deterministic noise, and input–output device limitations. The first three problems are peculiar to analog systems and can be circumvented by going to a digital system. The problem of devices is shared with both analog and digital systems although the constraints are very different in the two cases. In the following, we will briefly expand upon these problems of analog systems to emphasize the interest in developing DOPs. The first difficulty is the limited flexibility provided by the available optical analog computing software [6], [9], [11]. As mentioned above, the available operations are primarily some simple arithmetic operations and Fourier transformation. These are certainly powerful operations, and particularly in the case of the Fourier transformation, have much higher throughputs in optical implementations than is possible electronically. However, these operations are inadequate for many desirable computations. Although considerable work is being done to extend the range of available operations (see for example [5]) the extensions will not lead to the general processing capabilities available with digital systems. The flexibility of digital systems is a major factor behind the interest in developing digital optical computing techniques.

The second major drawback associated with optical analog systems is their noise susceptibility. Although the noise advantages of a digital system over an analog system are perhaps intuitively apparent, the following discussion provides a more quantitative comparison. It also points out that the important consideration is noise propagation through a cascade of operations rather than noise in a single operation.

We wish to compare two equivalent (sampled) systems, one analog and one digital. The basic unit of the analog system is shown in Fig. 1. It consists of an amplifier, additive noise n, and a quantizer Q. The basic unit can be cascaded into a multistage unit as shown in Fig. 2. Here the

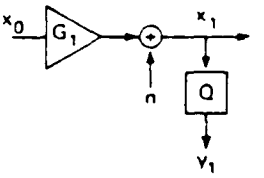


Fig. 1. Basic unit of an analog system. G_i is a gain and Q is a quantizer.

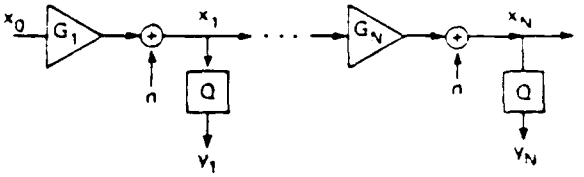


Fig. 2. Cascade of basic analog system units shown in Fig. 1.

x_i's and y_i's represent signal amplitudes. The amplification operation was chosen because it is the simplest operation one might utilize. In terms of optical systems, this could be a model of an ideal imaging system. We want to study the effects of noise in this cascaded system. We assume that the gain G_i is unity for all amplifiers. All the noise contributions are assumed identically distributed Gaussian random variables which are mutually independent. Finally, the output of the quantizer y_i is defined as

$$y_i = x_0, \quad \text{if } |x_i - x_0| < \delta x, 1, \cdots, i, \cdots, N. \quad (2)$$

We will not explicitly define Q outside the region defined by the quantization step δx, but merely state that Q acts like a quantizer function. Without loss of generality, the initial input is assumed to be quantized to a positive value x_0 which is $> \delta x$. Note that in the system of Figs. 1 and 2, Q essentially acts as a measurement device which does not affect the value transmitted from one stage to the next. With this arrangement

$$x_i = x_0 + n_i \quad (3)$$

where

$$n_i = \sum_{k=1}^{i} n = n \cdot i. \quad (4)$$

Since the n_i are assumed mutually independent, the probability density function (PDF) of n_i, denoted by $p(n_i)$, is the

ith-order convolution of the PDF of n in the form

$$p(n_i) = p(n) \odot_1 p(n) \odot_2 \cdots \odot_i p(n) \tag{5}$$

$$\triangleq p(n) \odot_i \tag{6}$$

where $\odot$ denotes convolution. Given that $p(n)$ is Gaussian with variance σ_0^2, then $p(n_i)$ is also Gaussian with variance

$$\sigma_i^2 = i\sigma_0^2 \tag{7}$$

In this case, the probability of one or more of the y_i outputs in the analog system being incorrect ($\neq x_0$) is given by

$$P_A = 1 - \prod_{i=1}^{N} [2\,\mathrm{erf}\,(z_i)] \tag{8}$$

where

$$z_i = \delta x/\sigma_i = \delta x/(i^{1/2}\sigma_0) \tag{9}$$

and

$$\mathrm{erf}\,(t) = (2\pi)^{-1/2}\int_0^t e^{-r^2/2}\,dr. \tag{10}$$

The function erf (t) varies from zero to 0.5 as t varies from zero to infinity. Thus as N increases, P_A approaches unity, implying that an error becomes a near certainty.

For the digital case, the system is very similar—the only difference being that the quantizer Q acts on the value being transmitted from one stage to the next. This is shown in Figs. 3 and 4. Where the quantizer funtion output $y_i, i = 1, \cdots, N$ is both measured or observed and passed on to the next stage. In order to calculate the probability of getting

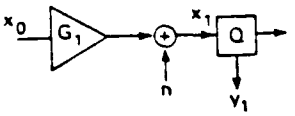


Fig. 3. Basic unit of a digital system. G_1 is a gain and Q is a quantizer

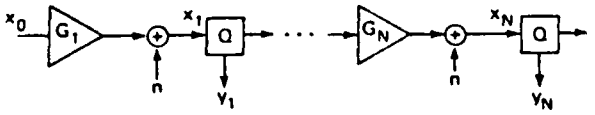


Fig. 4. Cascade of basic digital system units shown in Fig. 3

one or more errors as above, we first find the complementary probability of no errors. In computing this we note that when there are no errors the value x_i at the input of the ith quantizer stage is x_0, the correct value. Thus the ith stage looks just like the first stage and all stages individually have the same error probability. This is another way of saying that errors do not accumulate in a cascaded digital system as they do in a cascaded analog system. From this it can be shown that the probability of making one or more errors in the digital system is

$$P_D = 1 - \prod_{i=1}^{N} [2\,\mathrm{erf}\,(z_i)] = 1 - [2\,\mathrm{erf}\,(z_i)]^N \tag{11}$$

with

$$z_i = \delta x/\sigma_0. \tag{12}$$

Comparing (11) with (8) and noting that

$$0 \leqslant \mathrm{erf}\,(z_i) \leqslant \mathrm{erf}\,(z_i), \qquad 1, \cdots, i, \cdots, N \tag{13}$$

we conclude that the probability of making an error is much greater for the analog case than for the digital case.

The above result for digital systems can be easily extended to general computations which might be characterized as having M parallel channels, each with N stages. Since errors do not accumulate in cascaded digital systems, the probability of no errors occurring depends only upon the total number of stages involved, i.e., the product $N \times M$, not on the relative sizes of N and M (assuming the quantizer and noise properties are independent of the relative sizes of N and M).

For the sampled analog case, the tradeoff between the number of channels and the number of stages per channel is not so clear. However, the fact that errors accumulate in a cascaded analog system would seem to favor a system that reduces the cascading.

The third problem with analog systems is one of deterministic noise or approximation error. In order to perform a given operation it is necessary to configure an analog system with the appropriate response function. This response function is invariably an approximation to the desired function. For example, the simple process of amplification is obtained by a device whose response can only be (approximately) linear over a finite range. A digital system may also only approximate a desired function, but in this case the accuracy and the valid range of the approximation can be set at any desired level.

The final problem with analog systems is the lack of adequate devices. The primary goals in developing analog devices, either optical or electronic, are high speed, high accuracy, and good repeatability. For optical systems one generally wants to exploit parallel processing. In this case, additional important device parameters are space–bandwidth product (SBWP), which determines how many channels can be processed in parallel, and uniformity. Single-channel devices are generally difficult to develop with the desired characteristics, especially accuracy and repeatability, and the difficulty in general increases rapidly with the SBWP for multichannel devices.

As mentioned at the beginning of this section, the problems of flexibility, of stochastic noise, and of deterministic noise can be circumvented by going to digital systems. Also as will be discussed in Section III, the device requirements are easier to meet for binary digital devices than for analog devices. These are the basic reasons for the interest in optical digital computing.

One might include limited accuracy or dynamic range in a list of problems associated with analog systems and point out that digital systems are not limited in this respect. However, one should realize that the high accuracy or large dynamic range of a digital system is achieved by encoding the problem onto multiple channels. This concept can also be applied to analog systems although the methodology is not as well developed as it is for digital systems. One simple example of an analog system which uses multiple channels to increase the usable dynamic range is described by Goodman and Ströbin [12]. Another example is an optical Fourier transform system where space–bandwidth product in one domain can be traded for dynamic range in the transform domain.

Before proceeding to a discussion of digital optical

processing work, there is one area of optical analog computing which deserves mention here. There has been considerable interest in the past decade in optical systems which incorporate feedback (see, for example, [13]). This work is of interest for two reasons. First of all, it is utilized to overcome some of the drawbacks of analog systems, particularly dynamic range and accuracy limitations and also some of the flexibility limitations. Secondly, feedback has been applied to sampled analog systems to improve their accuracy in vector–matrix computations [14]–[17]. Finally, feedback has been utilized to achieve binary optical switching devices and bistable devices. Some details of feedback systems for these applications are given in Section III.

B. Optical Discrete Multilevel Computing

Another approach to optical computing is to perform computations with data stored as a set of two or more discrete numerical values. The most significant example of this technique is in optical residue arithmetic processors.

Residue optical processors rely on encoding the sampled and quantized data and performing arithmetic operations in a residue number system. The residue number system is based on a K-tuple of integers $(m_1, \cdots, m_K)$, each of which is called a modulus. The moduli are integers of any size and must have no common factors (i.e., they must be relatively prime). A number in a residue processor must be scaled to be a nonnegative integer x satisfying

$$0 \leqslant x \leqslant \left(\prod_{k=1}^{K} m_k \right) - 1 \tag{14}$$

to be uniquely represented; thus (14) defines the effective dynamic range of the processor. The number x is actually stored as a set of residues (or remainders) $(r_1, \cdots, r_K)$ defined by

$$r_k = [x]_{m_k} = x \bmod m_k, \qquad k = 1, \cdots, K \tag{15}$$

where r_k denotes the least positive integer remainder of the division of x by m_k. Thus each of the r_k also satisfies

$$0 \leqslant r_k \leqslant m_k - 1. \tag{16}$$

The basic integer arithmetic operations of addition, subtraction, and multiplication of two numbers x and y are performed according to

$$[x \bullet y]_{m_k} = \left[[x]_{m_k} \bullet [y]_{m_k} \right]_{m_k} \tag{17}$$

for each of the K moduli, where $\bullet$ represents either $+$, $-$, or $\times$. Thus K parallel operators in modular arithmetic are performed over the K different moduli.

Residue arithmetic can be illustrated with the following example for the addition of 78 and 35. Choose three moduli as 5, 7, and 9. Their product is 315, thus the calculations are unique only between 0 and 314. We arrange the residues as follows

	5	7	9
78	3	1	6
+ 35	0	0	8
113	3	1	5

Here 78 is converted to the 3-tuple (3, 1, 6) in the residue system, and 35 is converted to (0, 0, 8). Each column of moduli are independently added in modular fashion. As a check, the result of 113 can be expressed by the unique 3-tuple (3, 1, 5). The conversion of a set of moduli back to a decimal representation is somewhat involved and will not be treated in detail. There are two basic procedures, the Chinese Remainder Theorem and the Mixed Radix technique [18]–[20].

An important feature of residue arithmetic in comparison to binary or decimal arithmetic is that no carries occur between computations in the various bases. Thus all modular computations can take place in parallel without waiting for previous computations and their inherent propagation delay. The drawback of residue processing is that the input data (analog or binary digital) must be converted to residue form, and that the output residue must also be decoded. Division of two numbers is also a problem because the quotient may not be an integer and thus may not have a residue representation. All input data and arithmetic results must be scaled to be in the dynamic range of the processor. Because the residue representation is cyclic, any number that is generated which lies outside the dynamic range will "wrap around" and produce errors. Other difficulties are that sign testing or relative magnitude comparisons cannot be easily accomplished without conversion to decimal or binary form. Thus conditional tests on intermediate results are difficult. Although these problems are not easy to solve, several theoretical and experimental residue-based processors have been described which we now review.

Several different types of physical phenomena have been proposed for the implementation of residue processors. These phenomena must have a relatively small number of discrete repeatable levels equal to the maximum of the moduli chosen. Two major techniques are used: table lookup of results based on input operands; and the use of phase, polarization, and spatial position in circumstances where they provide an inherent cyclic response characteristic. Particular hardware realizations include coherent optical, electrooptical, and integrated-optical methods.

Huang [21] proposed several electrooptical techniques for residue encoding, processing, and decoding based on cyclic optical polarization and phase phenomena. Collins [22] described a residue processor based on a liquid-crystal spatial light modulator (SLM). In this system, arithmetic is performed by repeated accumulation of polarization rotations (differential phase delays) in the birefringent SLM. Additional related work in this area was described by Collins *et al.* [23], [24]. Psaltis and Casasent [25] described systems for decimal-to-residue conversion, residue processing, and residue-to-decimal conversion in which numerical values are represented by spatial position. The processors contain multiple diffraction gratings and rely on the cyclic nature of the coherent diffraction patterns of the gratings.

Horrigan and Stoner [26] described a number of cyclic and lookup table residue processors based on various electrooptical technologies. The comprehensive paper by Huang *et al.* [27] concentrates on table lookup techniques for residue processing and describes details of an optical residue matrix–vector multiplier. In this paper, the primary method of numerical encoding is spatial position. Both electrooptics and integrated-optics technologies are considered. The paper also contains estimates of processing speed and hardware complexity. Additional details of the processor and descriptions of a system for binary input and output are contained in [28]. Tai *et al.* [29] described a residue processor based on integrated-optics directional coupler waveguide switches. The programmable system relies on a spatial posi-

tion representation of numerical values and table lookup processing. Guest and Gaylord [30] have described a coherent optical lookup table numerical processor for binary or residue computing that uses content addressable optical holographic memories.

It is interesting to note that residue computations have been used in electronic digital systems for fast numerical computations. Hughes Research Laboratories has built a system called RADIUS (residue arithmetic digital image understanding system) that performs a 5 × 5 pixel generalized convolution operation of the form

$$y = \sum_{i=1}^{25} f_i(x_i) \tag{18}$$

on raster image data [31], [32]. Here y is an 8-bit output pixel, x_i represents 8-bit input pixels in the 5 × 5 kernel, and f_i represents a polynomial function of the corresponding x_i. Each f_i is arbitrary and independently programmable to produce one- or two-dimensional convolution operations or transforms. The system operates at rates to 10 Mbits/s and interfaces to standard computer buses such as UNIBUS™. All numerical computations are performed by residue arithmetic using the bases 31, 29, 23, and 19, which are the four largest primes less than $2^5 = 32$. This allows the additions and multiplications to be done by table lookup of residues for each base in random-access memories with 32 locations and 5 stored bits per location. Conversion from binary-to-residue and from residue-to-binary (via the mixed-radix method) is also accomplished using table lookup. The high speed of the processor is possible because of the table lookup arithmetic.

Accuracy and noise immunity remain an important issue in multilevel systems, since the signals must maintain an adequate signal-to-noise ratio (SNR) to be able to unambiguously distinguish the different allowable levels [10]. Thus the transducers in a multilevel system must have good SNR characteristics and perhaps more importantly, must be immune to variation in the bias and gain parameters of the input–output characteristics.

III. Binary Digital Logic

Although the multilevel residue systems described in the preceding sections offer several advantages over analog systems, they still suffer from some of the same shortcomings. In this section, we concentrate on various techniques for binary optical logic that have been proposed over the past few years and describe current work at USC and elsewhere on developing and extending this technology to arrays of devices and systems. Section III-A begins with a discussion of the advantages of binary optical logic and the requirements necessary for any binary logic technology to be effective in computing systems. We also refer to problems in interconnection, communication, and architecture as discussed later in this paper. Section III-B briefly describes current technology in integrated optics binary logic and switching. Section III-C discusses binary optical combinatorial logic devices and arrays. Sections III-D and III-E summarize USC work on combinatorial logic with liquid-crystal light valves (LCLVs) and variable grating mode (VGM) liquid-crystal devices, respectively, as the nonlinear thresholding element. Finally, Section III-F describes devices that combine a thresholding element with electronic or optical feedback to achieve bistability (memory).

A. Advantages and Needs of Binary Processors

A wide variety of physical phenomena have been proposed and utilized to implement binary logic, memory, and computational systems. At the fundamental device level, all these phenomena have similar characteristic curves that describe the input–output relationships of the device. In this subsection, we examine these fundamental properties and describe important parameters to be considered in binary computing technology.

Binary digital combinatorial logic functions can be implemented optically using a point nonlinear function. The general techniques used for each logic gate are illustrated in Figs. 5 and 6. The two binary input lines to the gate are

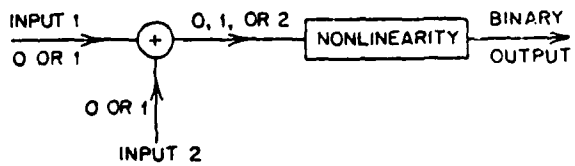


Fig. 5. Implementation of a two-input binary logic gate

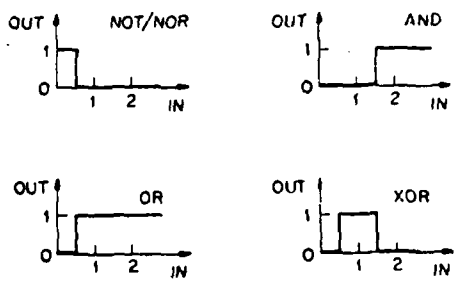


Fig. 6. Nonlinearities for implementation of two-input binary logic operations.

summed to yield a single three-level signal whose value is equal to the number of input lines with binary value 1. This signal then goes to the nonlinear function which produces a binary output. In an optical implementation, the binary value 0 is represented by a low intensity and the value 1 is represented by a high intensity. The addition can be accomplished by optically superimposing the signals on the input lines. With a detector that integrates the signal over the input spot, this can have the effect of adding the intensity levels regardless of the coherence properties of the light.

There are 16 different possible binary functions of two binary inputs, 8 of these are commutative functions. Possible choices of the nonlinear functions for some common commutative logic operations are shown in Fig. 6. Note that the nonlinear functions consist ideally of thresholds located at various input values. The remaining 8 binary logic functions that are not commutative require the ability to distinguish between the two binary inputs. Several methods of implementing these noncommutative functions are given in [33].

This basic principle for implementing logic functions can be extended to N-input gates as shown in Figs. 7 and 8. Here the signal on the input line to the nonlinearity can take on one of $N + 1$ possible values. Physical quantities other than light intensity have been exploited for optical logic. Optical logic systems based on complex amplitude,

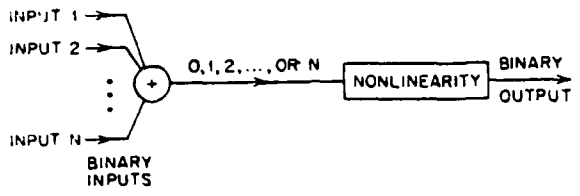


Fig. 7. Implementation of an N-input binary logic gate.

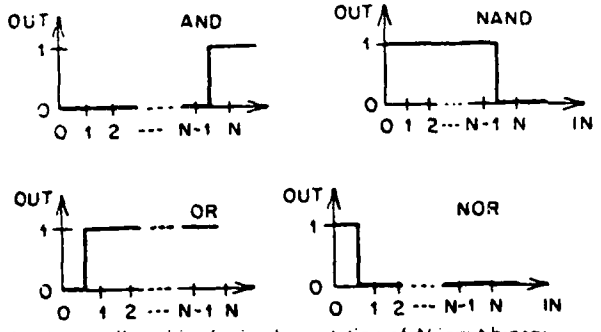


Fig. 8. Nonlinearities for implementation of N-input binary logic operations.

phase, and polarization have been suggested; some of these techniques are reviewed here.

Although no physical mechanism will achieve the ideal threshold switching characteristics of Figs. 6 and 8, useful optical logic can be implemented with a physical device having an input–output characteristic of the type shown in Fig. 9. This curve can be roughly divided into three regions as indicated. In the subthreshold region (A) the device produces little or no response to an input. In the analog operation region (B), the output is typically a linear function of the input. In the saturation region (C) the device output is only weakly dependent upon the input. For multilevel systems, the inherent deterministic and stochastic noise properties of the device are such that the analog operation region is divided up into N distinguishable levels. Changes in the characteristic curves with environment and errors due to noise produce an output uncertainty indicated by the dotted lines above and below the solid line representing the nominal characteristic curve in Fig. 9.

Although these devices may be utilized in both multilevel and binary numerical processing systems, for stability

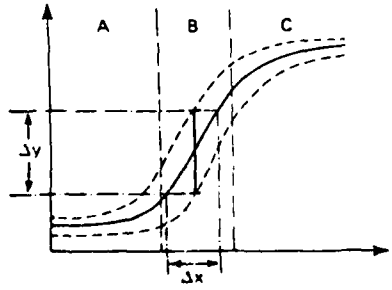


Fig. 9. Typical transducer input–output characteristic. Region A is the subthreshold region and region C is the saturation region. In both these regions input variations cause little or no output variation. Region B is the operating region for analog devices. The dotted curves indicate the range of variability in the curve due either to stochastic or deterministic variations. This variation determines the number of distinguishable levels in the operating region.

reasons, they are generally more suited to binary operation. Because the outputs in the subthreshold and saturation regions are insensitive to changes in the input, they are generally also only slightly affected by variations in the device operating parameters. However, the analog operation region is often very sensitive to device parameters. This is true for electronic devices as well as optical devices. Thus while the gain of an electronic amplifier circuit may be very sensitive to temperature variations, the threshold and saturation voltage levels will be relatively unaffected by temperature variations. Furthermore, the inherent noise in a transducer is generally signal dependent and will often be most severe in the operating range. This is related to the fact that the subthreshold and the saturation levels are physically well-defined states. For example, in photographic film, below threshold no grains are exposed so that the only noise contribution comes from background fog. When the film is exposed into saturation essentially all grains are developed and the only noise is the variation in the number of grains present in the field. In between these extremes one has the added noise contribution that the number of grains exposed is a stochastic function.

In searching for better logic devices there are several parameters to be optimized, including slope of the device input–output curve, and shape of the device subthreshold and saturation regions. Current research is aimed at determining necessary and sufficient conditions for optical devices to serve as useful switching elements. Generally, the device input–output curve should have a large slope in the region of analog operation, and a threshold saturation characteristic. The threshold should be repeatable and well-defined. As described later in Section III-B, this is the fundamental characteristic necessary to construct gates and various logic functions. We emphasize that devices with a physical *bistable* characteristic having two memory states are not needed. As discussed further in Section III-F, bistability results from feedback combined with a device having a general characteristic curve like that shown in Fig. 9. At the end of this paper, we summarize additional desirable properties for digital optical logic systems.

Some of the most important device parameters are switching time, delay time, and power consumption. Ideally, all of these parameters should be minimized. For many families of logic devices, the product of the power consumption and the delay time is approximately constant within the family, i.e., increased speed can be obtained at the expense of increased power consumption [34], [35]. Thus this power–delay product is often used to characterize logic devices. The power–delay product is equivalent to the device switching energy.

It is desirable to have input and output signals in binary logic represented by the same information-carrying medium so that signals may be propagated and cascaded without too many changes of state. Thus logic based on electronic signals generally needs electronic output; logic based on optical signals of a certain wavelength should have an optical output at the same wavelength. Logic systems containing a hybrid of media (i.e., electronic control of optical signals or *vice versa*) are generally undesirable. However, this hybrid of media is important for input or output transducers that must interface with external sensors and traditional electronic digital computing systems. Within a given logic gate, it is also important to avoid internal changes of state, such as photon–electron–photon conversions present

in many real-time electrooptical devices. It is these conversions that generally reduce speed and energy efficiency, and also increase noise.

A set of elementary logic functions is said to be complete if any combinatorial circuit can be constructed from them. For the binary number system, either the combination of logical NOT (an inverter) with logical OR, or the combination of logical NOT with logical AND is a complete set. These operations can be combined in one device which performs either the logical NOR or logical NAND. From these combined functions the NOT is achieved by fixing one input in a 0 or 1 state. In fact, the high-speed CRAY-1 processor is constructed almost entirely of NOR gates [36]. Another important element of any computer system is a binary memory element. These can be constructed from simple logic elements or bistable devices can be used to implement memory elements directly.

As we noted earlier, residue multilevel processors cannot do logical comparisons of numbers without conversion to binary, decimal, or some more traditional number system. It might be possible to develop and implement a complete set of multivalued logic functions in order to make a general-purpose processor, however, this could require many different fundamental functions and would lead to very complicated, clumsy design rules.

Another important consideration is the fabrication of parallel arrays of logic devices. Ideally, these devices should be easily and cheaply fabricated in large arrays with a high packing density so that the overall physical dimensions of the system are small.

A final goal is to choose devices that allow easy interconnections within the device array (on-chip) and easy communications from the device arrays to external systems (off-chip). These factors greatly influence system architecture and software. Some of these considerations are discussed in Sections IV and V of this paper.

Obviously, it is impossible to improve on all of these factors simultaneously. One typically finds that a device or technology which is particularly strong in one area will be weak in other areas.

The standards by which any new technology is judged are set by current and projected electronic systems. Current systems have switching energies in the picojoule range [35]. Minimum feature sizes on current integrated circuitry is in the micrometer range with the average linear dimensions of logic gates being a factor of ten larger. The gate density is limited by interconnection requirements as will be discussed later. For memories, where the gate density is highest, gates occupy about half the chip area whereas for more general circuitry, gates may occupy one-tenth the chip area. Projected electronic systems may have femtojoule switching energies, minimum feature size of a few tenths of a micrometer, and clock periods of a few nanoseconds [37].

Optics research has afforded several approaches to implementing binary logic elements and logic circuits. These range from integrated optics, which is a planar optical analog of electronic integrated circuits, to three-dimensional systems in which a two-dimensional array of logic elements are interconnected using the third dimension. Section III-B discusses current binary integrated optical switching and logic technology. Section III-C reviews other approaches to optical logic technology, particularly those producing arrays of gates.

3 Integrated Optics

Integrated optics technology is capable of fabricating guided-wave structures, electrooptic devices, sources, and detectors in a planar array [38]. Integrated optics circuits have many of the same logic and signal processing capabilities as semiconductor chips. Integrated optics is attractive due to its potential for high switching speeds and compatibility with fiber-optic communication links.

Taylor has proposed a system for implementing logic functions with integrated optics [39]. As he points out, the limiting element in such systems is the photodetector. Systems can be built where several processing steps can be executed between detection steps, these systems can be very fast, e.g., a predicted 4 ns for a complete 16-bit addition. Goldberg and Lee have demonstrated an integrated optical half-adder [40]. Integrated optical interconnection networks have also been discussed [41], [42]. Verber [43] has described many integrated optical systems for logic and arithmetic operations.

C. Optical Combinatorial Logic Devices and Arrays

Interest in developing optical binary logic and computing systems goes back to at least the 1950s. A patent granted to von Neumann in 1957 proposed the use of signals at microwave wavelengths. He described how switching, amplifying, and memory functions could be achieved when the signals are combined using a nonlinear element. A similar technique was described by Goto [45]. In the late 1950s Diemer and Van Santen [46] described techniques of integrated optical logic utilizing photoconductors combined with neon bulbs and electroluminescent devices.

Work on optical logic continued into the early 1960s. The general approach was to use optical nonlinearities such as gain quenching in injection lasers and saturable absorption to obtain optical logic circuit elements such as gates and flip-flops [47]–[51]. The goal was to obtain ultra fast switching by circumventing the *RC* time constants and carrier delay times of semiconductor electronic logic. Several review articles summarize the results of this work [52]–[55]. This goal was attained in some cases (switching times on the order of 0.1 ns) but at the expense of very low energy efficiency [34], [35], [53], [56]. The energy efficiency of these techniques suffered due to photon–electron–photon conversions, and it was difficult to envision such system with a large ($> 10^3$) number of gates in a two-dimensional array.

Although interest in optical logic elements waned in the late 1960s, interest in optical memory and storage techniques remained strong [52], [53], [55], [57], [58], thus maintaining the connection between optics and binary computing systems. Beginning in the mid-1970s interest in developing arrays of high-speed optical logic gates was revived as new materials and techniques provided more energy efficient solutions. At the same time, the study of optical logic expanded from the level of single gates to the level of circuits and systems. These systems require large parallel two-dimensional arrays of gates (with minimum size on the order of 10^2–10^3 gates) and several efforts were concerned with increasing the size of gate arrays from a few to several hundred [8], [57], [58]. Also there were continuing efforts to improve the switching speed.

The throughput of a system can be increased either by increasing device speed or by introducing parallelism. Several systems that we describe here have been proposed which exploit parallelism to obtain a high performance optical logic system. Most of these systems would be classified as optoelectronic in that they have electronic controls and they generally require a photodetection to mediate the logic operation. These systems differ significantly, each having its own particular advantages. We summarize them in the remainder of Section III-C and describe our own work in this area in Sections III-D and III-E.

Seko and Nishikata [59] have described an optical parallel threshold array. The system consists of an extremely thin (5-μm) GaAs crystal wafer that is pumped at 591.0 nm by a dye laser. An image (consisting of many resolution cells or gates) is impressed on the wafer and a single-mode laser oscillation at 834.0 nm is observed over the output array. The input–output curve of the system shows a threshold at 1 MW/cm^2. One shortcoming of this system is the change in wavelength from input to output.

Seko and Sasamori [60] and Seko [61] have reported an optical logic array that demonstrates thresholding, logical AND, and logical OR operations. The system consists of a Fabry–Perot laser resonator plate with an Nd: glass core and an input array of 2500 fibers. Pump images corresponding to an input logic array are superimposed on the fiber plate. The input–output curve of the device exhibits a sharp threshold at the onset of lasing, thus producing various logical operations by thresholding. Unfortunately, this system also produces a change of wavelength from input to output.

Seko, Kobayashi, and Shimizu [62] have described the use of a channel plate image intensifier as parallel logic device. By sequentially applying two logical array inputs and controlling the time sequence of the final acceleration voltage to the phosphor screen, the two inputs can be made to sum or subtract. Multiplication can be achieved with a cascade of two systems and binary logical operations can be derived from these algebraic functions. Processing times as low as 20 ns were reported, but the technique has the drawbacks of requiring high voltage (10 kV), relatively low spatial resolution (50 μm/gate), and temporal memory characteristics. This technology also fundamentally relies on photon–electron–photon conversions.

Watrasiewicz [63] described optical logic systems that encode binary data as spatial regions with orthogonal polarizations. He showed that various binary logic operations could be achieved with this type of encoding and considered the application of various real-time spatial light modulators for experimental implementation.

Tsvetkov, Morozov, and Elinson [64] envisioned a logic scheme very similar to that of Watrasiewicz. They described several types of electronically and optically controlled liquid-crystal (LC) cells that contain arrays of binary elements. They show that many different Boolean logic functions and pixel shifting can be implemented by cascading cells and altering polarizers placed between them. A shortcoming of this system as with many other LC systems is the slow response time. For twisted nematic LC systems the turn-on time is estimated at 0.2–10 ms, with a turn-off time of 15–200 ms.

Schaefer and Strong [65] have described a binary optical computer architecture called the TSE computer consisting of electrooptical binary logic arrays combined with fiber optic bundles and components for interconnections. Several devices were considered to provide the logic gates, among them thin-film field-effect transistor arrays. An array of 128 × 128 devices for the optical NOT function was constructed, and some of the fiber-optics components such as array combiners, interleavers, fan-in and fan-out devices were constructed. Two shortcomings of the TSE computer are the need for photon–electron conversion and reconversion at each stage and the difficulty of fabricating large parallel fiber-optics interconnections.

Lee [66] and Athale and Lee [67] described the construction of a planar electron-optical light modulation device (OPAL—for Optical Parallel Logic) that can perform thresholding and logic on data arrays. The device has a segmented structure consisting of opaque CdS photoconductor cells paired with transmissive cells containing a twisted nematic LC material as the electrooptical medium. The device is arranged so that input light striking the photoconductor alters the electric field across the LC material, in turn altering the polarization and transmission properties of each cell. The device can perform logical AND, OR, and NOT, and an 8 × 8 array of devices has been constructed. They have described an interconnection of two OPAL devices to produce logical sum and carry outputs in a binary half-adder. The switching time for one gate in the device is approximately 68 ms. More recently, Athale and Lee [68] have reported an improved version of this device with a sharper threshold and an improved ON:OFF transmission ratio of 50:1. Placing an optical feedback loop around the device gives it bistable characteristics, as described in Section III-F. Michaelson [69] described and experimentally verified the use of an external optical positive feedback loop placed around a Hughes LC light valve (LCLV) to increase the slope of the threshold for optical logic and image processing applications. The LCLV device is described in detail in the next section.

Warde *et al.* [70] have described an optically addressed microchannel spatial light modulator (MSLM) that can perform contrast reversal, image addition and subtraction, thresholding, and the AND, OR, XOR, and NOR binary logic functions. The device uses a photocathode and microchannel plate to detect and amplify illuminating signals and impress them on a plate of electrooptic material such as $LiTaO_3$ and $LiNbO_3$. The threshold characteristics of the device are enhanced by coating the material to form an array of high finesse Fabry–Perot cavities.

Fatehi, Wasmundt, and Collins [71] have described systems that perform all 16 binary functions of two binary input variables. Their system used Hughes LCLV devices combined with various orientations of polarizers to change the operating curves of the system. The system is complicated by the need to replicate some signals and the need to make two passes through a light valve to achieve some functions. As discussed earlier, most binary processors are designed using no more than three or four basic logic functions, so that in practical applications the system could be simplified.

Recently, Bartelt and Lohmann [72] described a parallel optical logic scheme involving spatial encoding and coherent optical spatial filtering. The input data arrays are encoded with a transmittance grating whose angular orientation varies with level. This encoding is known as theta

modulation [73]. Several ways to achieve the encoding are available, but this necessary preprocessing is generally slow by current technologies. A system for coherent spatial filtering with two input arrays is shown to produce various logic functions.

Tanida and Ichioka [74]–[76] have constructed a parallel optical combinatorial logic array processor that can also perform all 16 binary functions of two-input binary variables. The system uses a lensless shadow-casting system with an incoherent LED light source. The various functions are selected by changing the illumination through the LEDs. They demonstrate various binary picture processing operations such as shifting, binary differentiation, etc., and describe extensions to parallel gray-level picture processing. One limitation of the system is that it requires an encoding of input binary arrays before they can be processed. A recent paper [76] describes an off-line coding scheme using a computer generated holographic filter.

D. Combinatorial Logic with Liquid-Crystal Light Valves

As described earlier, an optical device or array of devices with a nonlinear threshold characteristic is necessary to implement logic functions. We have constructed a combinatorial logic system that uses a Hughes LCLV as the nonlinear element [33], [77], [78].

The Hughes LCLV is an optical image transducer that accepts a low-intensity input spatial image and converts it to an output image with readout light from another source [79]–[81]. The device operates at room temperature at approximately television frame rates (30 ms), with resolution similar to that of photographic film (approximately 40 cycles/mm). The device has aperture sizes up to 50 mm by 50 mm, and requires only a single 5–10 V rms audio frequency power supply.

The LCLV device (Fig. 10) consists of a sandwich of several thin-film layers on either side of a cell filled with LC

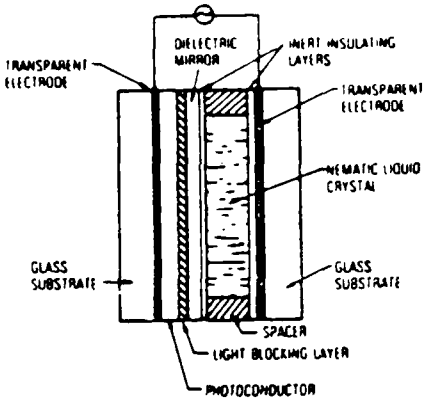


Fig. 10. Cross section of a Hughes photoactivated liquid-crystal light valve (LCLV).

material. The device has a CdS photoconducting layer, a CdTe light blocking layer, a dielectric mirror, and a biphenyl LC layer placed between indium–tin–oxide transparent conducting electrodes placed on optically flat glass substrates. An ac bias voltage is impressed across the electrodes to activate the device. An input image spatially modulates the input impedance of the photoconductor, thus altering the voltage drop across the LC material and varies the electrooptical properties. The alignment of the LC molecules with the substrate can be arranged in several ways to take advantage of polarization rotation effects (twist) or optical birefringence.

The particular LCLV used in our combinatorial logic experiments is called a 45 degree twisted nematic device. The LC molecules at the electrodes are aligned with their optical axes parallel to the electrode surface and oriented with a fixed preferred direction. The orientation is accomplished by mechanical rubbing or ion-beam etching.

The LCLV device is operated between crossed *polarizers* as shown in Fig. 11. P_1 and P_2 in Fig. 11 are polarizer and analyzer, respectively, while BS_1 is a beamsplitter. A typical input–output characteristic curve is shown in Fig. 12. The

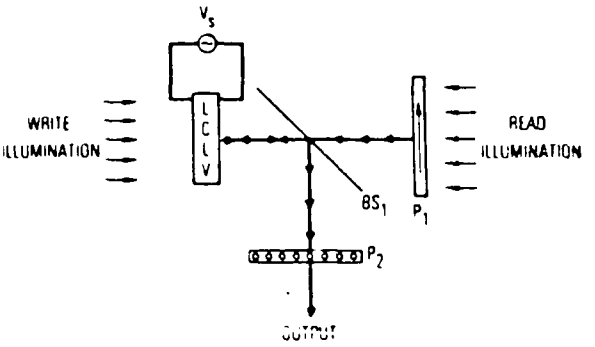


Fig. 11. System for measuring the input–output response of the LCLV device in Fig. 10. P_1 is a polarizer, P_2 is an analyzer, and BS_1 is a beamsplitter.

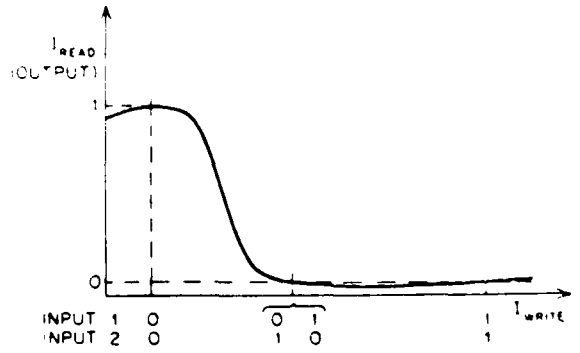


Fig. 12. LCLV input–output characteristic used to implement the NOR operation.

binary states 0 or 1 corresponding to logic gate outputs are represented by the two intensity levels marked on the vertical (output) axis. The inputs to each logic gate are optically superimposed to yield 0, 1, or 2 in the case of two input gates, and these inputs can produce various binary logic operations as shown in Fig. 6. Fig. 6 shows ideal input–output characteristic curves needed to achieve NOT, NOR, AND, OR, and XOR functions. By appropriate choice of bias and scaling, the curve of Fig. 12 can be utilized for the NOR and NAND operations, even though it does not have a sharp threshold.

An LCLV with a nematic LC material and zero degrees of twist is called a parallel-aligned LCLV. These devices have a

quasi-periodic oscillatory output as a function of the input intensity [33], [80]. Typical parallel aligned LCLVs can also be used for NOR and XOR functions as shown in Fig. 6. Note that an oscillatory LCLV characteristic is needed for the XOR operation and XNOR operation (not shown in Fig. 6), while a monotonic behavior over the selected dynamic range is needed for the other operations. Either a twisted-nematic or parallel-aligned LCLV with an appropriate LC thickness may be used to obtain the monotonic characteristic. The accuracy necessary to approximate the ideal logic function curves is discussed in [33], dynamic effects that must be considered are discussed in Section IV-B where the combinatorial logic circuit is combined with feedback to form a sequential processor.

Several combinatorial logic operations have been demonstrated using a parallel-aligned LCLV device. In this test, incoherent light sources were used for both input and output beams. The test inputs are shown in Fig. 13. In this

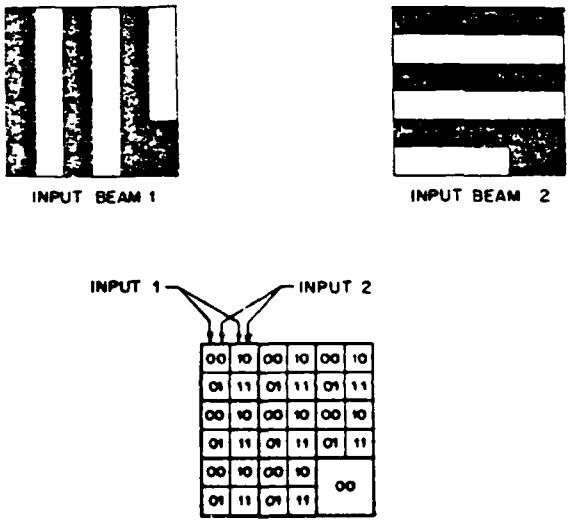


Fig. 13. Combinatorial logic inputs to generate truth tables.

case, one input was a vertical grating and the other a horizontal grating so that a replicated two-dimensional grid of all four possible input pairs is impressed on the photoconductor. Fig. 14 shows the expected output patterns for each logic operation. The particular LCLV used in this experiment had a poor response uniformity, thus each logic operation could be accurately performed only over a limited spatial region. Some of the sample outputs of the system are shown in Fig. 15. The grid size shown is approximately 2 lines/mm at the LCLV surface. Because this was a preliminary experiment, no attempt to determine the LCLV resolution limit was made. All the operations AND, NAND, OR, NOR, and XNOR were performed by adjustment of the LCLV operating parameters and searching for a region of uniform response on the LCLV.

As described earlier, only a few elementary logic functions such as NOR are needed to form a logically complete set. An important practical consideration is the uniformity and repeatability of this and any other parallel logic array. In Section IV-B we discuss the use of this combinatorial logic scheme as part of a sequential logic processor.

E. Combinatorial Logic with Variable Grating Mode (VGM) LC Devices

A completely different LC image transducer called a variable-grating mode (VGM) device has also been used for implementing a parallel array of combinatorial logic gates [82], [83]. Certain nematic LC mixtures are observed to form a laterally periodic optical phase grating when combined in a thin cell with transparent electrodes as shown in Fig. 16. Fig. 17 shows the typical phase structure when observed

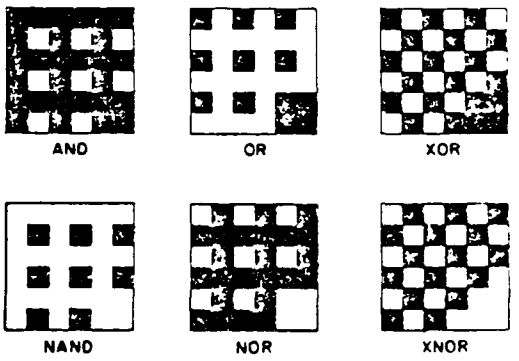


Fig. 14. Expected output patterns for each logic operation.

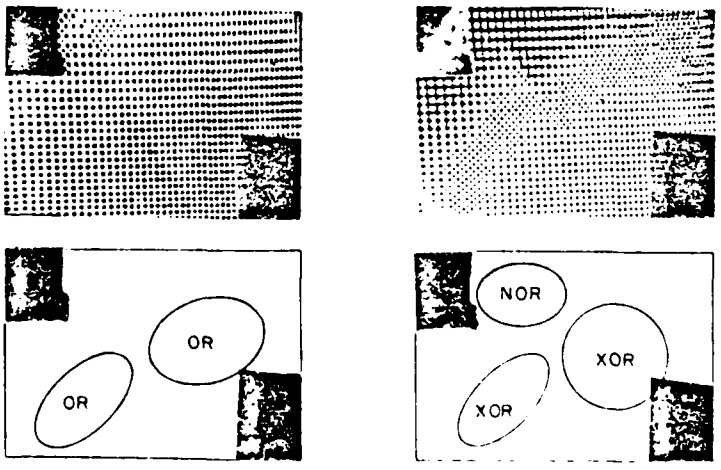


Fig. 15. Light valve outputs for combinatorial logic experiment.

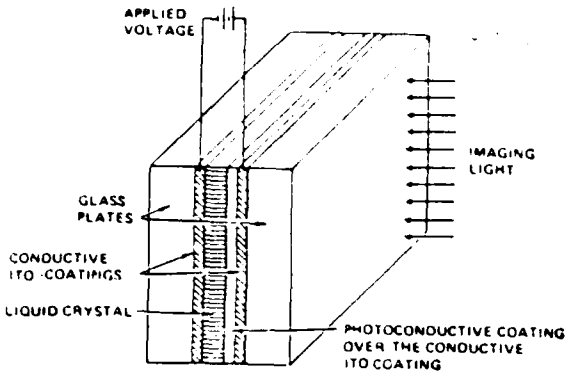


Fig. 16. Schematic diagram of the variable-grating mode (VGM) device construction. At present, devices are read out in transmission at a wavelength at which the photoconductor is insensitive.

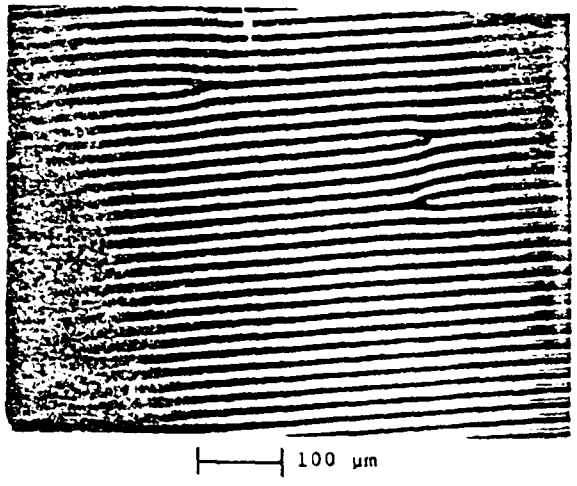


Fig. 17. Phase structure of the VGM device viewed through a polarizing microscope. Typical spatial frequencies range from 100 to 500 cycles/mm.

through a polarizing microscope. The local spatial frequency of the grating is quite high (typically 100 to 500 cycles/mm) and varies linearly with the magnitude of the local voltage across the LC layer. The grating period can be optically controlled by placing a two-dimensional photoconductive layer in series with the layer of LC. In the particular device used in our experiments, the photoconductor is sputter-deposited ZnS and the LC material is placed in a cell between indium–tin–oxide (ITO) transparent electrodes deposited on glass substrates. The device is biased with dc voltage impressed across the electrodes. A spatial illumination pattern incident on the photoconductor locally modifies the voltage on the LC material. Thus the device converts an input intensity distribution into a local variation of the phase-grating spatial frequency. The high lateral impedance of the thin photoconductive film prevents significant spreading of the photoconductivity and LC VGM effect; thus the device can retain high-resolution images. Fig. 18 schematically illustrates this mapping of intensity of spatial frequency. Note that a spatial region of the input image with a constant gray level is converted to a region with constant spatial modulation. When illuminated coherently and placed in an optical Fourier transform system shown in

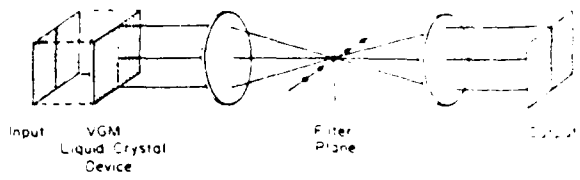


Fig. 18. Schematic arrangement of VGM device for optical processing indicating the mapping of intensity to spatial frequency in the Fourier filter plane.

Fig. 18, different spatial frequency components of the encoded image, corresponding to different input intensities, appear in different locations in the Fourier plane. Within the dynamic range of the device, intensities can be mapped monotonically into positions along a line in Fourier space. If the VGM grating carrier frequency is much higher than the spatial frequency of the input gray-level distribution [e.g., then it is possible to obtain various nonlinear transformations of the input intensity in the output plane by placing filters in the Fourier filter plane of Fig. 18.

The VGM device can be used to implement logic functions by inserting appropriate filters in the Fourier plane to achieve the nonlinear input–output curves of Fig. 6. Since the nonlinearities associated with optical logic are binary, they can be implemented with simple slit apertures having 0 or 1 transmittance values. The Fourier plane filter has a transmittance variation in the axis perpendicular to the variable grating which is essentially a plot of the desired nonlinearity. In addition, the logic function can be altered by changing the aperture in the spatial frequency plane.

A series of experiments was conducted to demonstrate the fundamental logic functions. The experimental setup of Fig. 19 was used. Two input fields were superimposed at the VGM plane along with a bias illumination. The total illumination intensity on the photoconductor of the VGM device was thus the sum of the two input intensities and the bias intensity. The input illumination was a filtered high-pressure mercury-arc lamp. The bias illumination was provided by a tungsten-bulb source. The VGM device was read out in transmission, using a He–Ne laser. Filters were placed in the Fourier plane to select the diffraction orders required in each case. For these experiments, the inputs consisted of two rectangular apertures, one vertical and one horizontal. When these were superimposed along with the bias, a square image was formed with the four quadrants having the intensity levels shown in Fig. 20. This image corresponds to the logic truth table shown in the figure. Thus the output images have intensity levels determined by the truth table associated with the desired logic function. The logic functions AND, OR, XOR, and their complements were implemented sequentially, as shown in Fig. 20, by altering only the Fourier plane filter. Imperfections visible in the output-plane data arise from defects in the cell structure of the VGM device employed in these experiments.

The input light to the VGM logic array need not be coherent, and the spatial coherence on the output side can be relaxed to only distinguish a small number of distinct spatial frequencies. Other particular characteristics of the VGM device make it interesting for use in residue processing systems [77], [82] and for more complex binary combinatorial logic circuits such as full adders. Its drawbacks are

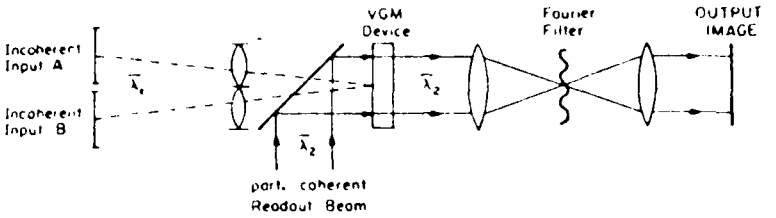


Fig. 19. Experimental arrangement for performing logical operations on two-dimensional binary inputs with a VGM device. The two input images are superimposed on the photoconductor. The device is read out in transmission. Simple slit apertures can be used to achieve the desired logic operations. $\bar{\lambda}_1$ and $\bar{\lambda}_2$ are the mean input and output wavelengths.

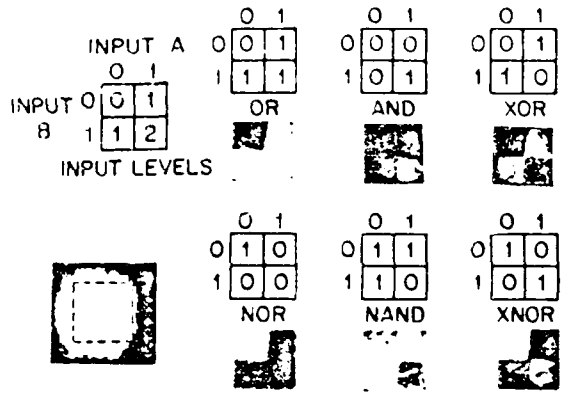


Fig. 20. VGM logic results. At left, two binary images were superimposed to produce the input intensities on the photoconductor as shown. Without any filter, the output is ideally a uniform field (logic 1). The actual output is shown below the indicated input levels. Right, ideal outputs and the actual outputs for the logical operations OR, NOR, AND, NAND, XOR, and XNOR, respectively.

that current devices are not spatially uniform and are quite slow (turn-on times on the order of 100–200 ms). Research on improvements in device speed and pixel uniformity is in progress.

F. Optical Bistable Devices and Arrays

A bistable device can be thought of as an optical nonlinearity (to provide a gate or switching function) combined with feedback (to provide two stable states) [84]–[92]. A bistable device may function as an electronic flip-flop because it can act as a logical memory unit. Bistable devices have also demonstrated switching, limiting, ac amplification, and modulation [87], [88]. The strong recent interest in optical bistability has been motivated by the possibility of very fast switching times (as low as 10^{-12} s) [87], [88] combined with switching energies competitive with electronic systems [87]. In this brief discussion of optical bistability we consider single devices or small arrays of devices first; then we discuss work on larger arrays of bistable elements.

Many different types of materials and techniques have been used to demonstrate optical bistability. The first bistable devices were called "hybrid" because light interacting with a nonlinear optical material was detected by a photodetector and electrically fed back to alter the transmission of the material [84]–[86]. For optical computing, this procedure is inherently limited by the speed of photon–electron conversion and is low in energy efficiency. The need for individual electrical feedback on each element also eliminates the possibility of large parallel bistable arrays.

Another class of bistable systems are called all-optical or intrinsic bistable systems [87]–[91]. Many of these bistable devices are Fabry–Perot etalons containing a material whose refractive index changes with applied light intensity. These devices are called intrinsic dispersive bistable devices because the feedback occurs internally within the Fabry–Perot cavity and the optical path length changes with intensity rather than the absorption. For optical computing, dispersive bistability is inherently more interesting because of its higher energy efficiency. Intrinsic bistable systems have been demonstrated with GaAs–GaAlAs at room temperature (300 K) with switching power of 5×10^{-3} W and switching times of 10^{-9} s.

Several groups have constructed arrays of bistable elements. These systems primarily involve external optical feedback around an LC spatial light modulator or similar device. O'Brien [93] patented a system in which two spatial light modulators (a Hughes LCLV or other device) are connected as bistable flip-flop, so that the output of one controls the input of the other and *vice versa*. Sengupta, Gerlach, and Collins [94]–[96] described a similar bistable system with a Hughes LCLV using only one SLM. Here the individual logic elements are spatially separated regions on the light valve and one output converts to a spatially separate input region of the same SLM. Sengupta *et al.* have performed a stability analysis of the system and determined numerical criteria for equilibrium [94], [95].

As described earlier in Section III-C, Athale and Lee [33] [68] have constructed an 8×8 array of logic gates as a segmented LC cell combined with a photoconductor such as CdS. By placing an optical feedback loop around the device so that a portion of the output signal is summed at the input side, the device shows a sharper threshold and bistable behavior.

IV. Communications

As is evident from work in very large scale integration (VLSI) of electronic circuits, it is not sufficient to produce fast, reliable gates and logic functions in order to produce efficient computing hardware [37], [97]. There are many other aspects of the total system that are important in determining the overall system performance. One critical area is that of communication. Communication problems arise at three distinct levels: the chip level where gates must be connected to one another, the subsystem level where

chips and discrete components must be interconnected, and the system level where different functional units such as processors, memory, and input–output devices must communicate [98]. In a progression from the chip level to the system level, the number of interconnections generally decreases. This is partly due to the difficulty of effectively controlling large networks of high-level systems.

A second major concern is system architecture. Historically, most computer systems have been designed in the classic von Neumann architecture in which memory elements are accessed one at a time using an address describing their location. This structure was well-suited to early hardware restrictions but is inappropriate given modern hardware constraints. In the remainder of Section IV, we describe some of these communication difficulties in more detail and consider the impact of optics on certain aspects of the problem. We describe in detail a digital optical sequential logic processor with a very general interconnection structure. In Section V, we describe new types of computing architectures and algorithms which are made possible by optical interconnections.

A. Gate-to-Gate Communications

In electronic VLSI systems, communications from gate to gate on a chip are a major expense in terms of device area required and design time. The costs associated with interconnection networks has led to an emphasis on highly regular structures where the number and lengths of interconnection lines is minimized. Systolic processors [99], [100] and wavefront array processors [101] are examples of numerical computing systems constructed under these constraints. Although these systems are finding a wide range of applicability, they sometimes impose a suboptimal structure on the processor design. Thus, for example, although it is possible to express a discrete Fourier transform (DFT) as a matrix–vector multiplication which is easily mapped onto a systolic-array architecture, a fast Fourier transform (FFT) algorithm inherently requires global interconnections and is not conveniently implemented in VLSI. Algorithms such as sorting and routing are also inherently global, generally requiring connections to widely separated gates on a chip.

The problem of input–output is also significant at the chip level. This is a problem that has been exacerbated with the development of VLSI where it is possible to put a large number of gates ($\sim 10^5$–10^6) on a chip but the number of pins to connect the chip to the outside world is severely limited ($\sim 10^2$). Considerable overhead in time and hardware is often spent incorporating multiplexers/demultiplexers in the system to reduce the number of input–output lines. With optical systems it is possible to have a large number of parallel inputs and outputs.

An example of a system developed for parallel input to a digital system is the parallel optical analog-to-digital converter [102]. This system was designed to take a gray-level image as an input and produce N binary images as outputs, each image being one bit plane of the binary representation of the image. These bit planes can then go directly into an optical logic processor. Although outputs can also be parallel with optical systems, the need for parallelism is not so great at the output as at the input, because processing tends to reduce information.

Goodman has recently described a number of optical techniques for interconnection of electronic components at the chip and subsystem level [98]. The techniques described include guided waves (the use of integrated optics and fiber optics) and free-space propagation (using focused or imaged interconnections and broadcast or unfocused interconnections). One particular proposed method has a reflection hologram positioned above a VLSI chip serving as an imaging element connecting a large number of integrated optoelectronic sources and detectors. Jain and Snyder [103] have experimentally demonstrated the distribution of picosecond pulse optical clock and test signals on GaAs FET logic chips as a method of avoiding differential timing delay problems in very high speed electronic processors.

The optical flip-flops of O'Brien [93] and Sengupta, Gerlach, and Collins [94]–[96] make use of optical interconnections (of optical gates on an LCLV) although the number of gates is small and the interconnections are arrangements of multiple mirrors. Some more complex extensions of these ideas to the design of digital optical shift registers and accumulators have also been proposed by Fatehi *et al.* [104].

Ferrano and Häusler described an optical feedback system in which a TV camera views a monitor displaying the output of the camera [105], [106]. They describe several techniques of parallel analog and digital optical signal processing that can be implemented. For digital processing, the TV frame is divided in half and a lenslet array is used to define an array of spatially separated pixels. Using simple optics and modifications to the TV scanning system, an array of independent binary flip-flops is constructed.

A digital sequential logic system consists of a combinatorial logic circuit combined with feedback. Thus the overall system has memory capabilities and its output will generally depend on past and current inputs. Barr and Lee [107] have described a sequential logic system with an optical array of gates and optical interconnections. The system consists of an optical logic unit, a set of optical memory units, electrooptic shutters, beam-steering mirrors and a microprocessor controller. The logic and memory units are constructed as 8×8 segmented LC electrooptic logic arrays, similar to those described in [67], [68]. Those units used as memories have their own feedback loops. The system functions as a cellular logic processor and performs parallel binary picture processing operations.

In the next section, we describe an optical sequential logic system constructed at USC that allows arbitrary interconnections between gates.

B. Free-Interconnection Sequential Optical Processor

We described in Section III-D how spatially parallel optical threshold functions could be used to implement binary logical operations. In this section we describe how such an array has been combined with a general optical interconnection scheme to produce an optical binary sequential logic system [33], [108], [109].

A schematic diagram of the system is shown in Fig. 21. The two main components are a spatially parallel array of optically implemented independent binary gates, and an optical interconnection unit that directs some of the signals on the output side of the gate array back to the input side of the gate array. In the particular experimental system built at USC, the gate array is the active area of a Hughes 45 degree twisted nematic LCLV. The light valve is biased so that each gate implements a NOR operation, which is a complete logical set for the construction of any combina-

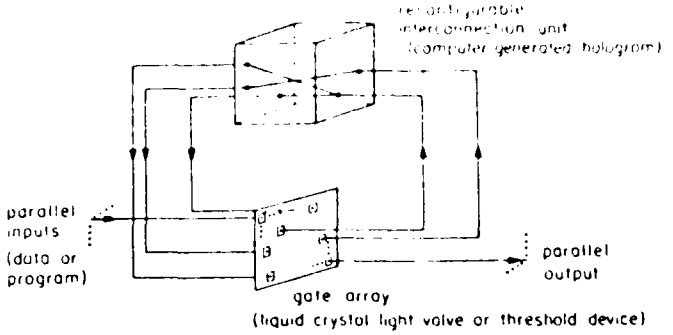


Fig. 21. Functional block diagram of sequential optical logic system.

torial logic function. The operation of this device for combinatorial logic is essentially identical to the system described in Section III-D. It is noteworthy that the same wavelength of illumination is used at the input and output of this LCLV device, in this case 514.5 nm.

The optical interconnection unit in Fig. 21 is a computer-generated hologram that serves as a beam-steering element to arbitrarily interconnect gate outputs to gate inputs. The hologram is effectively the wiring pattern interconnecting gates on the optical chip (the LCLV gate array). Use of the hologram allows both local and global interconnections on the chip because light beams corresponding to different wires can propagate through each other without interference and come to a focus on the input side of the LCLV. This lack of interference is a distinct advantage over interconnections with metallic wiring for electronic chips and optical-fiber wiring for optical chips. Another advantage of the system in Fig. 21 is that parallel input–output to the processor is possible by controlling inputs to the optical chip and intercepting outputs of the chip. In principle, a very large array of optical gates (more than 10^6) could be implemented on the chip, subject to space–bandwidth constraints of the gate array and interconnection hologram. It is interesting to note that the entire system can be rewired or reconfigured by physically changing the hologram or by writing the hologram in an optically or electronically alterable material that can be modified at higher speeds. As discussed in Section V of this paper, this free-interconnection flexibility leads to new types of computer architectures that avoid some existing processing bottlenecks.

Experimental System: An experimental sequential optical system like that in Fig. 21 was constructed to demonstrate the concept. The system has 16 gates arranged on a Hughes 45 degree twisted nematic LCLV. The light valve is read out between crossed polarizers and has a steady-state input–output relationship shown in Fig. 12. This response implements a parallel NOR operation and satisfies certain signal-regeneration criteria discussed in detail elsewhere [33].

The gates are interconnected with a binary version of the Lee computer-generated hologram [110]. In a Lee hologram, each complex-valued sample is described by a linear combination of four real nonnegative numbers, i.e., is decomposed into its components along each of the four half-axes in the complex plane. Each cell of the hologram is divided into four subcells, one subcell for each of these four components. One complex-valued sample is taken at the center of each subcell. Stored in each subcell is the corresponding component of its complex-valued sample. Because of the locations of these subcells, upon taking the optical Fourier transform, these four components are added with the correct phases to obtain the reconstruction in the (1,0) diffracted order. If the transmittance of the hologram is binary valued, each subcell actually contains a rectangle whose width is equal to the subcell width and whose height is proportional to the value of the corresponding sample component (Fig. 22).

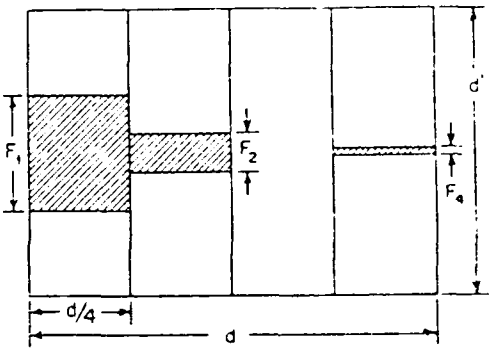


Fig. 22. Resolution cell of the Lee computer-generated hologram with binary-valued transmittance. Usually $d = d'$.

For the hologram used in this experiment, the transmittance is binary valued and these values are represented by different optical path lengths, i.e., a phase hologram. If the optical path lengths differ by a phase of π, the theoretical efficiency of the hologram is four times that of the equivalent absorption hologram. A possible tradeoff is that the (0,0)-order intensity may increase by more than a factor of four.

The hologram was written onto photoresist via electron-beam lithography. Surface relief of the photoresist provides the optical path length difference in the hologram. The electron-beam machine used has a step size of 0.125 μm and has written patterns with linewidths as small as 0.5 μm. It writes 1.024 × 1.024 mm fields and can stitch them together to cover a maximum area of 102 × 102 mm. The machine provides a far greater space–bandwidth product than was needed for our test circuit, although its capabilities would eventually be required for a very large-scale system.

Our test circuit comprises 16 gates so the hologram comprises 16 subholograms, which are laid out in a 4 by 4 array. Each subhologram covers a circular area and has

diameter of 1.04 mm. Each cell is a square 62.5 μm on a side, so there are a maximum of 17 cells (68 subcells) across each subhologram in the horizontal direction. Each subcell has a width of approximately 15.6 μm, or 125 steps of the electron-beam system, and has a height of 500 steps; both dimensions have more steps than were needed. 251 quantization levels were used for each subcell sample, keeping the apertures centered in each subcell. References [33] and [77] contain scanning-electron micrographs of the holograms and details on their theoretical and experimental SNR and efficiency.

A diagram of the main components of the sequential logic system is shown in Fig. 23. An expanded Ar-ion laser

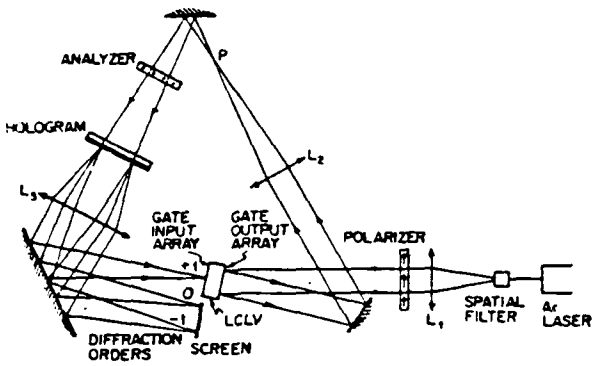


Fig. 23. Experimental system. Lens L_2 images from the LCLV gate output plane (LC plane) to the hologram plane. Lens L_3 provides a Fourier transform from the hologram plane to the LCLV gate input plane. The hologram comprises an array of subholograms.

beam is incident on the readout side of the LCLV (gate output plane). It is reflected from the internal mirror of the LCLV and is imaged from the LC (gate output) plane to the hologram via L_2. The LC plane is situated between crossed polarizers. The Fourier transform of the field transmitted by the hologram appears at the write side of the LCLV (gate input plane) via L_3. The phase of the illumination at the gate input plane is not correct, but only the intensity is of interest. Note that the Fourier transform relationship provides for complete regeneration of spot location during each pass through the feedback loop. In addition, since the subholograms are not contiguous, a mask is effectively incorporated into the hologram. This provides regeneration of the size and shape of each pixel, and also facilitates alignment.

A diffuser is placed just in front of the LCLV gate input plane in order to average over the fringe patterns that may result from coherent interference. This is not necessary when the pixels are small enough for the fringe patterns to be beyond the resolution limit of the device. Since a phase hologram was used and the effects of the (0,0) diffracted order were of concern, an aperture was used as a spatial filter at P to filter out diffraction effects from the limiting aperture. The $(-1,0)$ diffracted order in the hologram reconstruction can be used to monitor the gate inputs during system operation. The gate outputs can also be probed by using a reflection from the analyzer or the hologram.

For purposes of demonstration, a test circuit (shown in Fig. 24) was chosen and a hologram with the appropriate

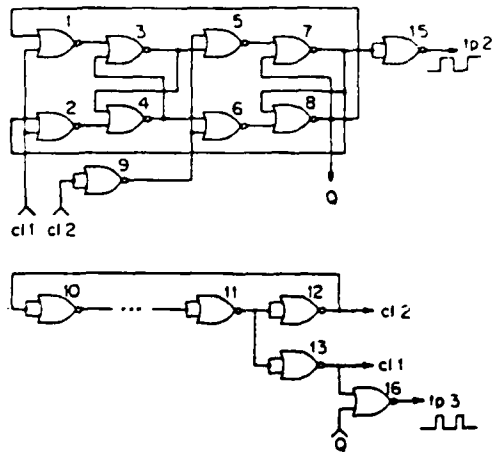


Fig. 24. Test circuit consisting of a synchronous master–slave flip-flop with driving clock.

interconnection pattern was then generated. The test circuit includes a synchronous master–slave flip-flop and a driving clock. The clock circuit is a ring oscillator consisting of an odd number of inverters. Clock circuits with three gates and with five gates have been implemented. The flip-flop functions as a frequency divider and outputs a signal whose frequency is half that of the clock and whose duty cycle is close to 50 percent.

Fig. 25 shows the spatial arrangement of the 16 numbered gate outputs corresponding to the circuit diagram of Fig. 24. The circuit cycles among four distinct states which are shown in Fig. 25. This figure shows the optical binary signals at the output side of the LCLV. In this system the frequency of oscillation of the clock circuit with five gates was extremely slow—2.65 Hz. The speed is entirely limited by the temporal characteristics of the LCLV. The particular LCLV used here was extremely slow, having a characteristic response time of more than 300 ms, thus accounting for the very slow oscillation. A similar 3-gate clock oscillated at 6.3 Hz. In principle, the circuit should perform the same logical operation at a much higher speed by replacing the LCLV with a faster device based on LC or other technology. For very fast SLMs, the optical path length of the interconnection system eventually must be considered. Details of output waveforms and other experimental measurements are described in [33]. In the next few sections, we discuss details of particular holographic interconnection methods that can be used in the free-interconnection sequential processor.

Space-Variant Interconnection: The sequential optical logic processor as described previously allows completely arbitrary interconnections. Each gate has a separate subhologram associated with it, and each subhologram encodes a different interconnection pattern. If the interconnection pattern associated with a given point or gate is considered to be an impulse response or point-spread function (PSF) then the system can be characterized as a general space-variant imaging system shown schematically in Fig. 26. The reconstructed images are simple dot patterns, each illuminating a gate input. As shown in Fig. 26, the desired interconnections are formed in one particular diffraction order. Typically, a conjugate image will also be produced, and this can be used to probe the system without affecting its operation or to access the system outputs.

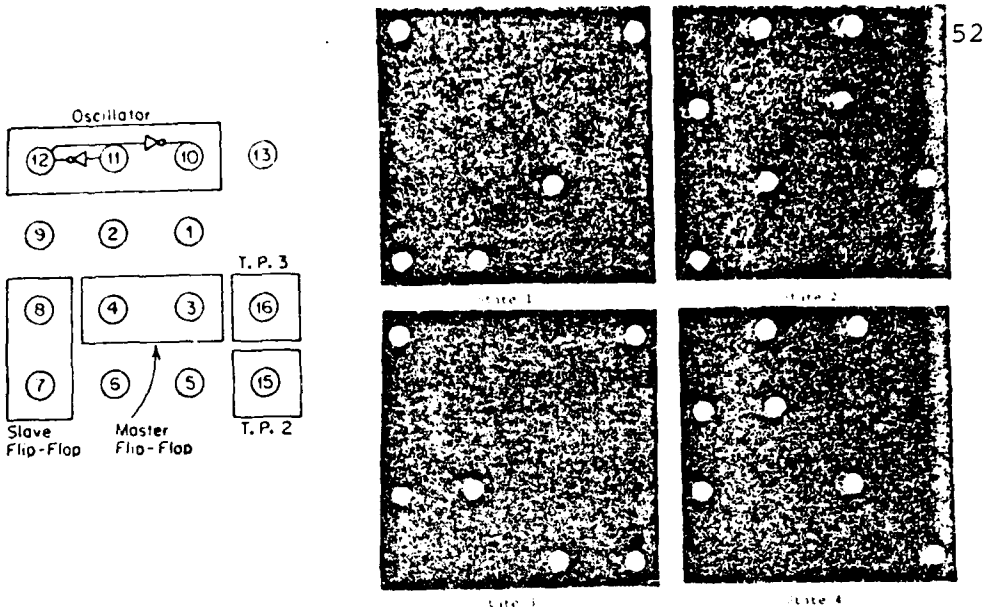


Fig. 25. Optical sequential logic output The spatial arrangement of the 16 numbered gate outputs are shown on the left.

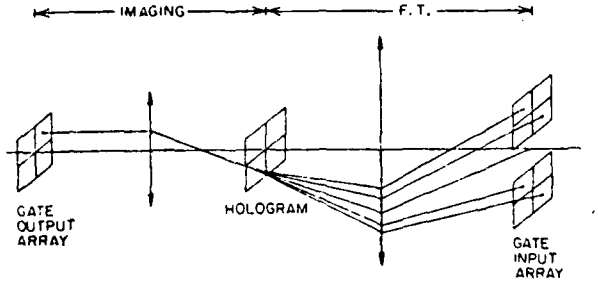


Fig. 26. Space-variant interconnection system. In general the hologram produces multiple diffraction orders, only one of which is used.

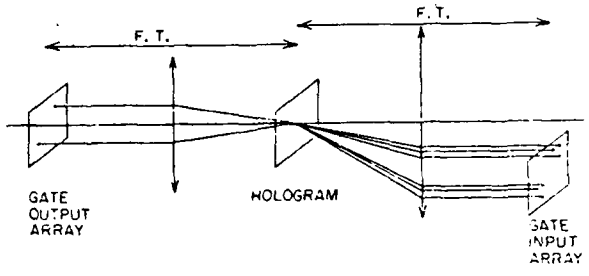


Fig. 27. Space-invariant interconnection system.

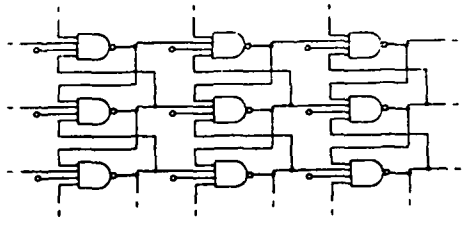


Fig. 28. Example of a gate array with space-invariant interconnections. The free inputs are enable inputs.

With the space-variant interconnection approach, the designer has essentially unlimited freedom in choosing an architecture to suit his needs. The limitation with this technique is in the space-variant element. The system described here uses a multi-element computer-generated hologram. The number of subholograms is equal to N, the number of gates, and each subhologram must have an SBWP of order N to allow arbitrary addressing. Thus the SBWP of the hologram is proportional to N^2. Since the SBWP of the hologram is restricted by available recording devices, this limits the number of gates which can be implemented. Calculations indicate that with current plotting capabilities, the limit on N is on the order of 10^4 gates [108], [109]. Additional details on crosstalk and design considerations are described in [111].

Space-Invariant Interconnections: A large increase in the number of gates is possible at the expense of a fixed, highly regular interconnection pattern. The extreme case is a totally space-invariant interconnection shown schematically in Fig. 27. This interconnection is optically implemented with one simple Fourier plane hologram for the entire circuit. The PSF of the system is a fixed interconnection pattern which is the same for every gate output. Thus the gate inputs are addressed relative to the position of the gate output, as compared to the absolute addressing of space-invariant interconnections.

Fig. 28 shows an example of a space-invariant array of NAND gates. Each gate is shown with an additional input that serves as an enable/disable signal, so that a particular circuit is implemented by disabling the appropriate gates. For NAND gates, a gate is enabled by projecting an optical signal onto the enable input. An obvious limitation is that circuits with irregular structure will require many gates to be disabled. An advantage is that for interconnecting N gates, only one interconnection pattern is needed, so that SBWP of the hologram required is roughly N, or 10^8 gates

with current hologram plotting systems. With this type of interconnection, the SBWP of the parallel-gate array is generally the limiting factor. References [108] and [109] describe cellular logic [112] and other parallel processing operations that can be implemented with space-invariant interconnections.

Hybrid (Basis-Set) Interconnection: In many instances, neither the direct space-variant nor the space-invariant approaches may be well suited for interconnections. For example, a system that requires a very large number of gates which cannot be interconnected in a regular fashion is difficult to implement with either approach described above. In these cases, a hybrid or basis-set approach that combines the space-variant and space-invariant techniques can be used [108], [109], [111]. The idea is to define a finite number M of distinct interconnection patterns as PSFs, and then assemble the circuit using only these M patterns. We generally expect that M will be much less than N, the number of gates. If the necessary set of PSFs is defined, it may be possible to determine a basis set of lower dimensionality that achieves all interconnections. If so, it is easier to store the set of basis elements rather than store the PSFs directly. Of course, in this case, the coefficients for each gate must be stored along with the basis elements.

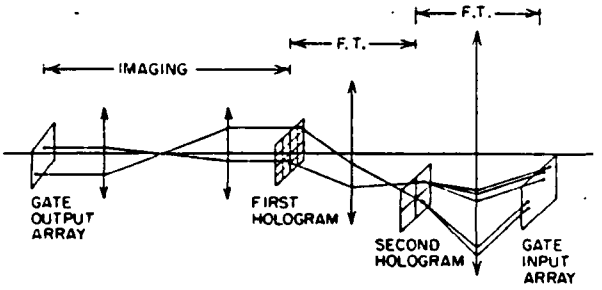


Fig. 29. Hybrid (basis-set) interconnection system. The first hologram is a space-variant element as in Fig. 26 that stores the coefficients for each gate output. The second hologram is an array of space-invariant filters that stores the interconnection basis functions.

Fig. 29 schematically depicts an optical system which uses the basis function approach. Two holograms are used. The second hologram is a composite hologram consisting of one subhologram for each basis element. The first hologram is also a composite hologram with one subhologram for each gate in the array. Each subhologram records the basis function coefficients for the corresponding gates. The subhologram essentially diffracts the light from the gate to all of the appropriate basis elements with the proper relative strengths. For this case, if there are approximately $M = 50$ basis elements required, then on the order of $N = 10^5$ gates can be implemented with current hologram generation techniques. As before, the limiting factor on the number of gates is the device implementing the gates rather than the interconnection elements [108], [109]. Thus with this approach the designer has some minor limitations on the interconnections which can be used, but he has a potentially large number of gates at his disposal.

Of course, other interconnection schemes are possible and each will present the designer with its own set of constraints and freedoms. In the end, one must pick an approach which is best matched to the problem at hand.

C. Chip-to-Chip Communications

At the intermediate level of the communications hierarchy is the need for high-bandwidth chip-to-chip interconnections. With advances in VLSI electronic chips, inexpensive processors are now available. The demand for increased processing power has led to the desire to interconnect large numbers of processors which can work concurrently. In these systems, the interconnection network is a critical element.

There are currently significant limitations on the number of pins available for input–output on electronic chips. To avoid these difficulties intensive on-chip multiplexers and demultiplexers are required to compress parallel signals into a serial form. This greatly reduces the effective input–output bandwidth and takes up valuable gate area on the chip. In effect, a single chip can interconnect with only a small number of parallel lines even though it may be capable of highly parallel processing internally.

Optics offers some parallel techniques of chip-to-chip interconnection for both optical and electronic logic systems. One possible approach is illustrated in Fig. 30. Here a

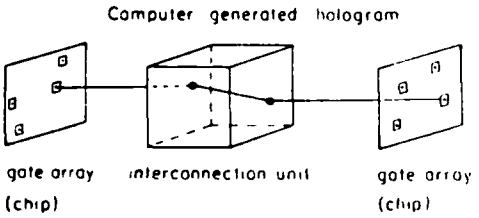


Fig. 30. Chip-to-chip interconnection of an optical gate array with a computer-generated hologram.

hologram serves as a chip-to-chip interconnection unit similar to its use in the optical sequential logic processor described in the previous section. The individual gates on the chips can be optical, or they can be electronic gates connected by integrated optical sources. The hologram directs the parallel set of optical signals to another chip where they are utilized directly in optical logic or detected and converted to electronic form for VLSI chips. Here the input–output is completely parallel, allowing the transfer of entire two-dimensional data arrays such as images. Various interconnection holograms such as space-variant, space-invariant, and hybrid methods can be adapted to this problem.

Goodman has described a related set of optical chip-to-chip interconnection techniques [98] and we also note that the parallel-gate array cellular logic processing system of Barr and Lee [107] can be utilized in a similar way.

In the next section, we describe some techniques of interconnection at the processor-to-processor level. Many VLSI electronic chips are actually self-contained processors themselves, so that the distinction between chip and

processor is no longer well defined. Many of the processor-to-processor interconnection schemes described would also be valuable at the chip-to-chip level.

D. Processor-to-Processor Communications

At this highest level, a relatively smaller number of interconnections are required, but they should have a very large bandwidth. Fiber-optics communication links are already finding use in connecting computers with each other and with peripheral equipment. They provide high-bandwidth links with good interference immunity, although they eventually suffer from the limitations of electronic buses in that only a small number of devices actively communicate at any time and that processing overhead is required to implement protocols.

In many instances where the processors perform a single function the interconnection network itself defines the operation. Thus a tree network [113] is used for sorting problems and systolic networks [99], [100] are used for many matrix operations. The tree and systolic networks are generally fixed structures which perform a fixed operation. Ideally, one would like to have a dynamic structure that could be reconfigured to meet the users' needs. Such networks are the subject of considerable research currently.

A general-purpose dynamic network is the crossbar switch shown in Fig. 31. This switch allows any of N inputs to be

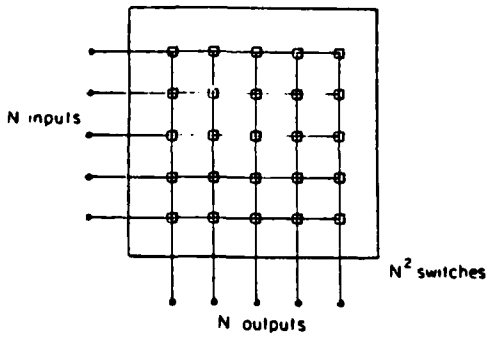


Fig. 31. Crossbar interconnection network.

connected to any of N outputs. Although this is the most general switch one would normally need, it is seldom implemented because of the large hardware requirements (N^2 switches). Instead, considerable research has been devoted to devising simpler networks such as the Baseline network [114], the Omega network [115], the binary n-Cube network [116], the Banyan network [117], and the shuffle exchange network [118], [119]. These networks generally require less hardware than the crossbar network and also provide less performance. Since the crossbar and other networks are fixed for a given computation and are only changed infrequently, the emphasis on implementing these systems is on parallelism or number of channels, rather than on speed. Optical approaches to implementing the crossbar would appear to be a viable solution to the interconnection problem. In fact, optical systems which have been developed for matrix–vector multiplication [120]–[123] are readily adapted to function as crossbar switches. The "matrix" for a crossbar switch contains only binary entries, a 1 representing a connection and a 0 implies no connection. The N inputs comprise the multiplying vector and the N outputs constitute the resultant vector. With a real-time device representing the matrix these systems become dynamic crossbar switches.

An optical interconnection network has been implemented by Tajima *et al.* [124]. They use an optical system to connect a large number of processing units to a large number of memory units. They broadcast the signals optically so that no wiring is required.

V. Architectures and Algorithms

The use of optical systems, in particular highly parallel systems, leads to certain implications in the design of computing hardware. Some general comments about these implications follow in this section. A similar but somewhat different perspective is given by Huang [125].

In the early days of computer development, the hardware for both gates and for the gate interconnections was expensive. Thus a computer design evolved which minimized the hardware requirements. These requirements have led to the classical von Neumann computing system architecture shown in Fig. 32. In a von Neumann machine, all of the processing logic is contained in the central processing unit (CPU) and the memory is located almost entirely in a

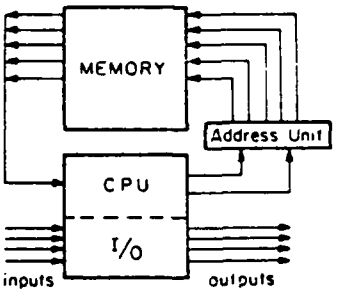


Fig. 32. Von Neumann computer.

separate unit. Input–output is generally performed by the CPU or by a system closely coupled to the CPU. The CPU accesses the memory through a binary addressing unit and memory contents are returned to the CPU through a single or small number of lines. This multiplexing scheme reduces the communications requirements and minimizes the number of lines. This serial addressing of memory results in a tradeoff of time for interconnection complexity. The eventual limitation on computing speed is known as the von Neumann bottleneck. Current high-performance electronic computers operate with very short clock cycles, this means that coaxial electronics with very short physical wire lengths are needed to avoid electromagnetic interference (EMI), crosstalk, and clock skew problems. The coaxial electronics for interconnections requires terminating resistors on the ends of cables, further reducing system energy efficiency.

With the development of integrated circuitry and VLSI, the situation has changed considerably. In VLSI, gates are relatively cheap while communication lines are relatively

expensive. Even with this modified cost function, the von Neumann structure is still reasonable; however, alternative architectures also become viable if they minimize communication at the expense of adding more gates. An example of a non von Neumann architecture is the systolic array [99], [100]. Systolic arrays have a large number of processors with very little memory at each. The signals flow through a highly regular two-dimensional array of processors in a pipelined fashion. An opposite approach is to have numerous local memories, say one for each processor. Here the total system memory has a large amount of redundancy in order to minimize communications at the expense of extra hardware.

With optical systems, the situation is again changed. Here communications can be considered cheap. Depending upon the exact implementation, gates can be considered either cheap or moderately expensive. In this case, the computer designer would consider architectures that minimize hardware redundancy by utilizing the freedom to make arbitrary connections. Such a processor could consist of a system of several processors and memory elements interconnected with a fixed network or dynamic crossbar switch.

Another type of digital optical processor is the non von Neumann system shown in Fig. 33. In this computer, all memory elements are accessible in parallel. With the availability of relatively cheap interconnections, the input–output shown as part of the CPU in Fig. 33 can be separated, as shown in Fig. 34, allowing direct parallel communication among all three parts of the computing system. The memory, CPU, and input–output systems in Fig. 34 themselves consists of gates and internal interconnections; we redraw this block diagram to separately emphasize the gates and interconnections to produce the free-interconnection architecture realization shown in Fig. 35. Here memory, CPU,

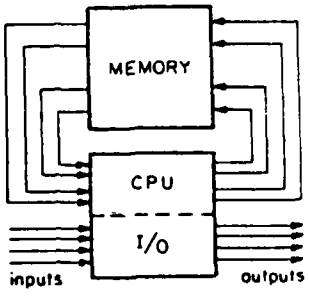


Fig. 33. A non von Neumann computer

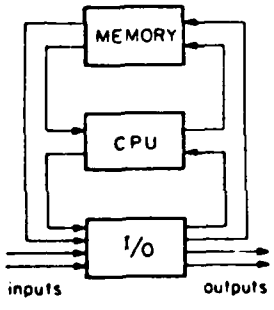


Fig. 34. A free-interconnection non von Neumann computer

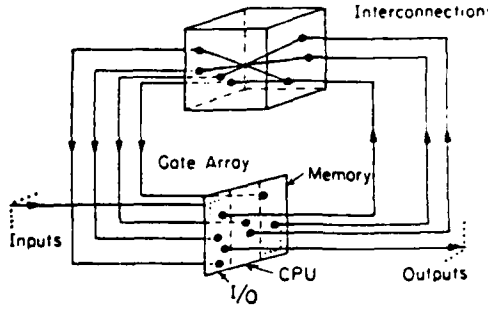


Fig. 35. The free-interconnection non von Neumann computer of Fig. 34 reconfigured as a sequential optical processor.

and input–output subsystems can be considered essentially indistinguishable portions of a large optical gate array (chip), while the interconnections provide communication with and among subsystems. The processor shown in Fig. 35 is essentially the sequential optical logic system described in Section IV-B.

Parallel optical computing architectures will strongly influence algorithms for computationally intensive tasks. Some particularly interesting applications are in image analysis and image understanding, where cellular logic arrays could be implemented in an optical parallel processor. Another important application is in artificial intelligence, where computational demands are extreme. Optical computers could be used to rapidly implement associative memory and symbolic substitution operations important for production rule based algorithms.

VI. Limitations

To complete this paper on optical computing, we briefly discuss limitations on the technology. These limitations can be addressed at several levels from the fundamental physical limitations to practical limitations imposed by current devices or system configurations [34], [35], [55], [56], [87], [88], [126], [127].

The ultimate limits are those imposed by quantum mechanics and thermodynamics [88], [126]. These limits are fixed and unalterable. From a quantum-mechanical standpoint we can say that information about a given logic state is carried by discrete elements, photons in our case, each having energy $h\nu$, where h is Planck's constant and ν is the optical temporal frequency. Because of the statistical nature of these information carriers, there is a minimum number required on the average to reliably define a given state or state transition. The exact number required depends upon the reliability criterion and the statistical model, however a minimum of 100 photons would be a typical number to establish a state reliably. Another quantum limit is that any stable switching operation necessarily dissipates a minimum energy h/τ, where τ is the switching time. The thermodynamic limit is set by thermal noise, which will cause unwanted state transitions unless the transition energy is significantly greater than kT, where k is Boltzmann's constant and T is the absolute temperature. In optical systems, the quantum limit dominates since the photon energy $h\nu$ is $\sim 100kT$ for visible photons.

No systems currently approach the quantum limits and will not in the foreseeable future. In practice we can point

to other factors which currently set the limits on the performance of optical logic systems. These limits are closely related to current device characteristics. A primary concern is heat dissipation, which is a severe problem because energy consumption in practical systems far exceeds the quantum limits. Heat dissipation becomes a problem when combined with the requirement of high gate packing densities. In turn, the need for high gate densities is ultimately driven by time constraints on the communication interconnections. For optical logic systems based on absorption nonlinearities, heat dissipation is probably the limiting factor and optical systems compare unfavorably with electronic systems [34], [55], [56]. However, the situation may be different for systems based on reactive nonlinearities and resonant structures where absorption in the device is low and heat can be dissipated external to the logic elements [35], [88], [126], [127]. It has been shown in the past few years that limitations in this area are not fundamental, but rather are technology dependent, thus they can change dramatically with innovations in the field.

There are also practical limitations which are tied to the overall system design. These are primarily communication limits. The lengths of communication lines and the corresponding signal transit times determine the minimum cycle time. As discussed previously, communication limitations for optical systems differ significantly from electronic systems, if one utilizes three-dimensional interconnection networks between planar gate arrays. However, the effective distance between planes still determines the minimum cycle time and should be kept as short as possible. For a given communication distance, however, the optical system has a faster minimum cycle time because communicating over electronic lines involves charging and discharging capacitances.

VII. Future Needs and Conclusions

In this paper we have reviewed optical logic elements and digital optical computing systems. It is evident that much more work is needed on basic research and development efforts on both the devices and systems.

We summarize here a few thoughts about desirable properties for optical logic elements. Devices with a fast switching time and low power consumption per bit are needed to be competitive with electronic devices. It is desirable to have input and output signals represented by the same optical wavelengths to allow cascading of switching stages, and it is desirable to avoid photon–electron–photon conversions to improve switching speed and energy efficiency. A complete logical set can be achieved with NOR, NAND, and other types of logical elements [112]; all 16 two-variable binary logic functions are probably not needed or wanted. Easy fabrication, high packing density, and room-temperature operation of arrays of devices is very desirable; these arrays must also allow easy nonplanar interconnections on-chip and off-chip. A useful practical system that might be constructed as an optical computer is a parallel cellular logic processor for an image with 500×500 pixels [112]; such a processor might have 40 gates per pixel, or a total of 10^7 gates. This number is not an unreasonable goal for a device array.

At the systems level we have seen that optical computing systems differ fundamentally from electronic systems. One important principle is that optical computers should not be built by merely replacing electronic devices with optical devices. Designing an optical logic system requires a new set of cost functions for trading off gate count and interconnection complexity. These different design criteria lead to new architectures for logic systems and make optical implementations attractive where electronic implementations would be impractical. Designs which are interconnection intensive can often be easily mapped onto relatively simple optical systems, whereas interconnection limitations are already severely restricting VLSI design. Optical computing offers the possibility of non von Neumann parallel architectures with different mechanisms for interconnections and communications; it remains a challenge to develop optical devices, architectures, and algorithms that make use of this potential. The next few years will be a critical and exciting period for these technologies.

Acknowledgment

The authors wish to thank B. K. Jenkins, P. Chavel, B. H. Soffer, A. R. Tanguay, Jr., and R. Forchheimer for their contributions to the research described in this paper.

References

[1] J. W. Goodman, *Introduction To Fourier Optics*. New York: McGraw-Hill, 1968.
[2] K. Preston, Jr., *Coherent Optical Computers*. New York: McGraw-Hill, 1972, Ch. 8.
[3] D. Casasent, Ed., *Optical Data Processing-Applications* (*Topics in Applied Physics, Vol. 23*). Berlin: Springer-Verlag, 1978.
[4] S. H. Lee, Ed., *Optical Information Processing-Fundamentals* (*Topics in Applied Physics, Vol. 48*). Berlin: Springer-Verlag, 1981.
[5] H. Stark, Ed., *Applications of Optical Fourier Transforms* New York: Academic Press, 1982.
[6] J. W. Goodman, "Operations achievable with coherent optical information processing systems," *Proc. IEEE*, vol. 65, pp. 29–38, 1977.
[7] D. Casasent, "Coherent optical pattern recognition," *Proc IEEE*, vol. 67, pp. 813–825, 1979.
[8] E. A. Mnatsakanyan, V. N. Morozov, and Yu. M. Popov, "Digital data processing in optoelectronic devices (review)," *Sov. J. Quantum Electron.*, vol. 9, pp. 665–677, 1979.
[9] G. Häusler, "Optical software survey," *Opt. Acta.*, vol. 24, pp. 965–977, 1977.
[10] I. Jahns, "Concepts of optical digital computing—A survey," *Optik*, vol. 57, pp. 429–449, 1980.
[11] J. W. Goodman, "Architectural development of optical data processing systems," *J. Electrical Electron. Eng., Australia*, vol. 2, pp. 139–149, 1982.
[12] J. W. Goodman and H. B. Strübin, "Increasing the dynamic range of coherent optical filters by means of modulating gratings," *J. Opt. Soc. Amer.*, vol. 63, pp. 50–58, 1973.
[13] R. P. Akins, R. A. Athale, and S. H. Lee, "Feedback in analog and digital optical image processing: A review," *Opt. Eng.*, vol. 19, pp. 347–358, 1980.
[14] D. Psaltis *et al.*, "Iterative color-multiplexed, electro-optical processing," *Opt. Lett.*, vol. 4, pp. 348–350, 1979.
[15] H. J. Caulfield *et al.*, "Eigenvector determination by noncoherent optical methods," *Appl. Opt.*, vol. 20, pp. 2263–2265, 1981.
[16] J. W. Goodman and M. Song, "Performance limitations of an analog method for solving simultaneous linear equations," *Appl. Opt.*, vol. 21, pp. 502–506, 1982
[17] W. K. Cheng and H. J. Caulfield, "Fully-parallel relaxation algebraic operations for optical computers," *Opt. Commun.*,

vol. 43, pp. 251–254, 1982.
[18] H. L. Garner, "The residue number system," *IRE Trans. Elect. Comp.*, vol. EC-8, pp. 140–147, 1959.
[19] A. Sasaki, "Addition and subtraction in the residue number system," *IEEE Trans. Elect. Comp.*, vol. EC-16, pp. 157–164, 1967.
[20] N. S. Szabo and R. I. Tanaka, *Residue Arithmetic and Its Applications to Computer Technology*. New York: McGraw-Hill, 1967.
[21] A. Huang, "The implementation of a residue arithmetic unit via optical and other physical phenomena," in *Proc. 1975 Int. Opt. Comp. Conf.* (Washington, DC, Apr. 1975), pp. 14–18.
[22] S. A. Collins, Jr., "Numerical optical data processor," *Proc. Soc. Photo-Opt. Inst. Eng.*, vol. 128, pp. 313–319, 1977.
[23] S. A. Collins, Jr., J. Ambuel, and E. K. Damon, "Numerical optical data processing," in *Proc. 1978 Int. Opt. Comp. Conf.* (London, England, Sept. 1978), pp. 194–197.
[24] ——, "Optics for numerical calculation," in *Proc. ICO-11 Conf.* (Madrid, Spain, 1978), pp. 311–314.
[25] D. Psaltis and D. Casasent, "Optical residue arithmetic: A correlation approach," *Appl. Opt.*, vol. 18, pp. 163–171, 1979.
[26] F. A. Horrigan and W. W. Stoner, "Residue-based optical processor," *Proc. Soc. Photo-Opt. Inst. Eng.*, vol. 185, pp. 19–27, 1979.
[27] A. Huang, Y. Tsunoda, J. W. Goodman, and S. Ishihara, "Optical computation using residue arithmetic," *Appl. Opt.*, vol. 18, pp. 149–162, 1979.
[28] A. Huang and J. W. Goodman, "Number theoretic processors, optical and electronic," *Proc. Soc. Photo-Opt. Inst. Eng.*, vol. 185, pp. 28–35, 1979.
[29] A. Tai, I. Cindrich, J. R. Fienup, and C. C. Aleksoff, "Optical residue arithmetic computer with programmable computation modules," *Appl. Opt.*, vol. 18, pp. 2812–2923, 1979.
[30] C. C. Guest and T. K. Gaylord, "Truth-table look-up optical processing utilizing binary and residue arithmetic," *Appl. Opt.*, vol. 19, pp. 1201–1207, 1980.
[31] S. D. Fouse *et al.*, "A residue-based image processor for VLSI implementation," in *Proc. Image Understanding Workshop*, L. Baumann, Ed., DARPA/Science Applications, Inc., Rep. SAI-82-391-WA, pp. 188–198, Apr. 1981.
[32] S. D. Fouse, G. R. Nudd, and A. D. Cummings, "A VLSI architecture for pattern recognition using residue arithmetic," in *Proc. 6th Int. Conf. on Pattern Recognition* (Munich, Germany, Oct. 19–22, 1982), IEEE 82CH1801-0, IEEE Comp. Soc., pp. 262–269.
[33] B. K. Jenkins *et al.*, "Sequential optical logic implementation," submitted to *Appl. Opt.*
[34] R. W. Keyes and J. A. Armstrong, "Thermal limitations in optical logic," *Appl. Opt.*, vol. 8, pp. 2549–2552, 1969.
[35] ——, "Fundamental limits in digital information processing," *Proc. IEEE*, vol. 69, pp. 267–278, 1981.
[36] L. C. West, "Construction of a general purpose digital computer based on optical logic elements," in *CLEO Tech. Dig.*, paper TUG5, Opt. Soc. of America, p. 44, May 1983.
[37] C. Mead and L. Conway, *Introduction to VLSI Systems*. Reading, MA: Addison-Wesley, 1980.
[38] T. Tamir, Ed., *Integrated Optics, 2nd Ed.* (*Topics in Applied Physics, Vol. 7*). Berlin: Springer-Verlag, 1981.
[39] H. F. Taylor, "Guided wave electro-optic devices for logic and computation," *Appl. Opt.*, vol. 17, pp. 1493–1498, 1978.
[40] L. Goldberg and S. H. Lee, "Integrated optical half adder circuit," *Appl. Opt.*, vol. 18, pp. 2045–2051, 1979.
[41] C. H. Bulmer, W. K. Burns, and T. G. Giallorenzi, "Performance criteria and limitations of electrooptic waveguide array deflectors," *Appl. Opt.*, vol. 18, pp. 3282–3295, 1979.
[42] R. A. Becker and W. S. C. Chang, "Electrooptical switching in thin film waveguides for a computer communications bus," *Appl. Opt.*, vol. 18, pp. 3296–3300, 1979.
[43] C. Verber, "Integrated Optical Architectures and Implementation," *Proc. SPIE*, vol. 456, 1984.
[44] J. von Neumann, "Non-linear capacitance or inductance switching, amplifying and memory organs," US Patent 2 815 488, 1954.
[45] E. Goto, "The parametron, a digital computing element which utilizes parametric oscillations," *Proc. IRE*, vol. 47, pp. 1304–1316, 1959.
[46] G. Diemer and J. G. Van Santen, "Power amplifiers based on electro-optical effects," *Philips Res. Rep.*, vol. 15, pp. 368–389, 1960.
[47] A. B. Fowler, "Quenching of gallium-arsenide injection lasers," *Appl. Phys. Lett.*, vol. 3, p. 1, 1963.
[48] ——, "Cooperative effect in GaAs lasers," *J. Appl. Phys.*, vol. 35, pp. 2275–2276, 1964.
[49] G. J. Lasher and A. B. Fowler, "Mutually quenched injection lasers as bistable devices," *IBM J. Res. Develop.*, vol. 8, pp. 471–475, 1964.
[50] G. J. Lasher, "Analysis of a proposed bistable injection laser," *Solid-State Electron.*, vol. 7, pp. 707–716, 1964.
[51] M. I. Nathan *et al.*, "GaAS injection laser with novel mode control and switching properties," *J. Appl. Phys.*, vol. 36, pp. 473–480, 1965.
[52] O. A. Riemann and W. F. Kosonocky, "Progress in optical computer research," *IEEE Spectrum*, vol. 2, pp. 181–195, Mar. 1965.
[53] W. V. Smith, "Computer applications of lasers," *Proc. IEEE*, vol. 54, pp. 1295–1300, 1966.
[54] N. G. Basov, W. H. Culver, and B. Shah, in *Laser Handbook*, F. T. Arecchi and E. O. Schulz-DuBois, Eds. Amsterdam: North-Holland, 1972.
[55] R. Landauer, "Optical logic and optically accessed digital storage," in *Optical Information Processing*, Yu. E. Nesterikhin, *et al.*, Eds. New York: Plenum, 1976, pp. 219–253.
[56] R. W. Keyes, "Power dissipation in information processing," *Science*, vol. 168, pp. 796–801, 1970.
[57] N. G. Basov *et al.*, "Methods of realization of an optical processor with variable operators," *Sov. J. Quantum Electron.*, vol. 8, pp. 307–312, 1978.
[58] C. C. Guest and T. K. Gaylord, "Two proposed holographic numerical optical processors," *Proc. Soc. Photo-Opt. Inst. Eng.*, vol. 185, 1979.
[59] A. Seko and M. Nishikata, "Image emission platelet laser for optical information processing," *Appl. Opt.*, vol. 16, pp. 1272–1274, 1977.
[60] A. Seko and A. Sasamori, "Fiber laser plate," *Appl. Opt.*, vol. 18, pp. 2052–2055, 1979.
[61] A. Seko, "All-optical parallel logic operation using fiber laser plate for digital image processing," *Appl. Phys. Lett.*, vol. 37, pp. 260–262, 1980.
[62] A. Seko, H. Kobayashi, and K. Shimizu, "A high-speed image information preprocessor using a channel plate image intensifier," *Opt. Eng.*, vol. 19, pp. 229–232, 1980.
[63] B. M. Watrasiewicz, "Optical digital computers," *Optics Laser Technol.*, vol. 7, pp. 213–215, 1975.
[64] V. A. Tsvetkov, N. A. Morozov, and M. I. Elinson, "Picture logic and liquid crystals," *Sov. J. Quant. Electron.*, vol. 3, pp. 989–993, 1975.
[65] D. H. Schaefer and J. P. Strong, III, "Tse computers," *Proc. IEEE*, vol. 65, pp. 129–138, 1977.
[66] S. H. Lee, "Review of coherent optical processing," *Appl. Phys.*, vol. 10, pp. 203–217, 1976.
[67] R. A. Athale and S. H. Lee, "Development of an optical parallel logic device and a half-adder circuit for digital optical processing," *Opt. Eng.*, vol. 18, pp. 513–517, 1979.
[68] ——, "Bistability and thresholding by a new photoconductor-twisted nematic liquid crystal device with optical feedback," *Appl. Opt.*, vol. 20, pp. 1424–1432, 1981.
[69] J. Michaelson, "Characterization of liquid crystal light valves and their applications to real-time nonlinear optical processing," Ph.D. dissertation, Dep. Elec. Eng., Univ. of So. Calif., Los Angeles, Oct. 1979; USCIPI Rep. 930, USC Image Processing Inst., Los Angeles, CA 90089, 1979.
[70] C. Warde, A. M. Weiss, A. D. Fisher, and J. I. Thackara, "Optical information processing characteristics of the microchannel spatial light modulator," *Appl. Opt.*, vol. 20, pp. 2066–2074, 1981.
[71] M. T. Fatehi, K. C. Wasmundt, and S. A. Collins, Jr., "Optical logic gates using liquid crystal light valve: Implementation and application example," *Appl. Opt.*, vol. 20, pp. 2250–2256, 1981.
[72] H. Bartelt and A. W. Lohmann, "Hybrid logic processing in parallel," *Appl. Opt.*, vol. 22, pp. 2519–2520, 1983.
[73] J. D. Armitage and A. W. Lohmann, "Theta modulation in optics," *Appl. Opt.*, vol. 4, p. 399, 1965.
[74] J. Tanida and Y. Ichioka, "Optical logic array processor," in *Proc. IEEE 1983 10th Int. Opt. Comp. Conf.*, pp. 18–23, Apr. 1983.

[75] J. Tanida and Y. Ichioka, "Optical logic array processor using shadowgrams," *J. Opt. Soc. Amer.*, vol. 73, pp. 800–809, 1983.

[76] ——, "Image encoding by a computer generated holographic filter," *Proc. SPIE*, vol. 432, 1983.

[77] A. A. Sawchuk, T. C. Strand, and A. R. Tanguay, Jr., "Nonlinear real-time optical signal processing," USCIPI Rep. 1080, USC Image Processing Inst., Los Angeles, CA 90089, 1982.

[78] ——, "Nonlinear real-time optical signal processing," USCIPI Rep. 1100, USC Image Processing Inst., Los Angeles, CA 90089, 1983.

[79] I. Grinberg *et al.*, "A new real-time noncoherent to coherent optical image converter," *Opt. Eng.*, vol. 14, pp. 217–275, 1975.

[80] I. Grinberg *et al.*, "Photoactivated birefringent liquid-crystal light valve for color symbology display," *IEEE Trans. Electron Devices*, vol. ED-22, pp. 775–783, 1975.

[81] W. P. Bleha *et al.*, "Application of the liquid crystal light valve to real-time optical data processing," *Opt. Eng.*, vol. 17, pp. 371–384, 1978.

[82] B. H. Soffer *et al.*, "Variable grating mode liquid crystal device for optical processing and computing," *Mol. Cryst. Liq. Cryst.*, vol. 70, pp. 145–161, 1981.

[83] 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.*, vol. 5, pp. 398–400, 1980.

[84] P. W. Smith, E. H. Turner, and P. J. Maloney, "Electrooptic nonlinear Fabry-Perot devices," *IEEE J. Quant. Electron.*, vol. QE-14, pp. 207–212, 1978.

[85] E. Garmire, S. D. Allen, and J. Marburger, "Bistable optical devices for integrated optics and fiber optics applications," *Opt. Eng.*, vol. 18, pp. 194–197, 1979.

[86] P. W. Smith, "Hybrid bistable optical devices," *Opt. Eng.*, vol. 9, pp. 456–462, 1980.

[87] H. M. Gibbs, S. L. McCall, and T. N. C. Venkatesan, "Optical bistable devices: The basic components of all-optical systems?" *Opt. Eng.*, vol. 19, pp. 463–468, 1980.

[88] P. W. Smith and W. J. Tomlinson, "Bistable optical devices promise subpicosecond switching," *IEEE Spectrum*, vol. 18, no. 6, pp. 26–33, June 1981.

[89] C. M. Bowden, M. Ciftan, and H. R. Robl, Eds., *Optical Bistability*. New York: Plenum, 1981.

[90] E. Abraham, C. T. Seaton, and S. D. Smith, "The optical computer," *Scientif. Amer.*, vol. 248, no. 2, pp. 85–93, Feb. 1983.

[91] 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.*, vol. 44, pp. 172–174, 1984.

[92] *Digest of Papers* (Topical Meeting on Optical Bistability, Rochester, NY, June 1983), Opt. Soc. of America, Washington, DC, 1983.

[93] J. A. O'Brien, "Bi-stable image storage apparatus," US Patent 3 941 455, Mar. 2, 1976.

[94] U. K. Sengupta, U. H. Gerlach, and S. A. Collins, Jr., "Bistable optical spatial device using direct optical feedback," *Opt. Lett.*, vol. 3, pp. 199–201, 1978.

[95] U. H. Gerlach, U. K. Sengupta, and S. A. Collins, Jr., "Single-spatial light modulator bistable optical matrix device using optical feedback," *Opt. Eng.*, vol. 19, pp. 452–455, 1980.

[96] S. A. Collins, Jr. *et al.*, "Multi-stable optical system," US Patent 4 272 157, June 9, 1981.

[97] D. F. Barbe, Ed., *VLSI: Fundamentals and Applications*. Berlin: Springer-Verlag, 1982.

[98] J. W. Goodman, "Optical interconnections in microelectronics," *Proc. SPIE*, vol. 456, Jan. 1984.

[99] H. T. Kung and C. E. Leiserson, "Systolic arrays (for VLSI)," in *Sparse Matrix Proc. 1978* (Soc. for Industrial and Applied Mathematics), pp. 256–282, 1979.

[100] H. T. Kung, "Why systolic architectures, *IEEE Computer*, vol. 15, pp. 37–46, 1982.

[101] S. Y. Kung *et al.*, "Wavefront array processor: Language, architecture, and applications," *IEEE Trans. Comput.*, vol. C-31, pp. 1054–1066, 1982.

[102] A. Armand *et al.*, "Real-time parallel optical analog-to-digital conversion," *Opt. Lett.*, vol. 5, pp. 129–131, 1980.

[103] R. K. Jain and D. E. Snyder, "Addressing and control of high-speed logic circuits with picosecond light pulses," *Opt. Lett.*, vol. 8, pp. 85–87, 1983.

[104] M. T. Fatehi *et al.*, "Sequential logic element arrays for use in optical digital processors: Applications and limitations," *Proc. SPIE*, vol. 241, pp. 139–148, 1980.

[105] G. Ferrano and G. Häusler, "A bistable TV-optical feedback system," *Opt. Commun.*, vol. 32, pp. 375–379, 1980.

[106] G. Ferrano and G. Häusler, "TV optical feedback systems," *Opt. Eng.*, vol. 19, pp. 442–451, 1980.

[107] H. Barr and S. H. Lee, "A digital optical processing system," in *Proc. 10th Int. Optical Computing Conf.*, pp. 171–177, IEEE Cat. 83CH1880-4, IEEE Computer Soc., Silver Spring, MD, 1983.

[108] P. Chavel *et al.*, "Architectures for a sequential optical logic processor," in *Proc. 10th Int. Optical Computing Conf.*, pp. 6–12, IEEE Cat. 83CH1880-4, IEEE Computer Soc., Silver Spring, MD, 1983.

[109] B. K. Jenkins *et al.*, "Architectural implications of digital optical processors," submitted to *Appl. Opt.*

[110] W. H. Lee, "Sampled Fourier transform hologram generated by computer," *Appl. Opt.*, vol. 9, p. 639, 1970.

[111] B. K. Jenkins and T. C. Strand, "Computer generated holograms for space-variant interconnections in optical logic systems," in *Proc. Int. Conf. on Computer Generated Holography, SPIE Proc.*, vol. 437, pp. 110–118, 1983.

[112] K. Preston *et al.*, "Basis of cellular logic with some applications in medical image processing," *Proc. IEEE*, vol. 67, pp. 826–856, 1979.

[113] L. Snyder, "Introduction to the configurable, highly parallel computer," *IEEE Computer*, vol. 15, no. 1, pp. 47–58, 1982.

[114] C. Wu and T. Feng, "On a class of multi-stage interconnection networks," *IEEE Trans. Comput.*, vol. C-29, pp. 694–702, 1980.

[115] D. H. Lawrie, "Access and alignment of data in an array processor," *IEEE Trans. Comput.*, vol. C-24, pp. 1145–1155, 1975.

[116] M. C. Pease, "The indirect binary n-cube microprocessor array," *IEEE Trans. Comput.*, vol. C-26, pp. 458–473, 1977.

[117] L. R. Goke and G. J. Lipovski, "Banyan networks for partitioning multiprocessor systems," in *Proc. 1st Annu. Symp. on Computer Architecture*, pp. 21–28, Dec. 1973.

[118] T. Lang and H. S. Stone, "A shuffle-exchange network with simplified control," *IEEE Trans. Comput.*, vol. C-25, pp. 55–65, 1975.

[119] H. S. Stone, "Parallel processing with the perfect shuffle," *IEEE Trans. Comput.*, vol. C-20, pp. 153–161, 1971.

[120] R. A. Heinz *et al.*, "Matrix multiplication by optical methods," *Appl. Opt.*, vol. 9, pp. 2161–2168, 1970.

[121] R. P. Bocker, "Matrix multiplication using incoherent optical techniques," *Appl. Opt.*, vol. 13, pp. 1670–1676, 1974.

[122] J. W. Goodman *et al.*, "High-speed incoherent optical method for performing discrete Fourier transforms," *Opt. Lett.*, vol. 4, pp. 1–3, 1978.

[123] R. A. Athale, "Optical matrix algebraic processors: A survey," in *Proc. 10th Int. Opt. Comput. Conf.*, IEEE Computer Soc. Press, pp. 24–31, 1983.

[124] H. Tajima, Y. Okada, and K. Tamura, "A high speed optical common bus for a multi-processor systems," *Trans. Inst. Electron. and Commun. Eng. of Japan*, vol. E66, pp. 47–48, 1983.

[125] A. Huang, "parallel algorithms for optical digital computers," in *Proc. 10th Int. Opt. Comput. Conf.*, IEEE Computer Soc. Press, pp. 13–17, 1983.

[126] P. W. Smith, "On the physical limits of digital optical switching and logic elements," *Bell Syst. Tech. J.*, vol. 61, pp. 1975–1993, 1982.

[127] R. L. Fork, "Physics of optical switching," *Phys. Rev. A*, vol. 26, pp. 2049–2064, 1982.

1.3 Variable Grating Mode Liquid Crystal Devices

The variable grating mode (VGM) liquid crystal device has been further studied experimentally [4-16]. This work has been in conjunction with the device development work being done at Hughes Research Laboratories.

An important area of research on the Variable Grating Mode Liquid Crystal Light Valve is concerned with a fundamental understanding of the origin of the "variable grating" effect. A wide variety of studies have been performed on electrically addressed VGM cells (no intervening photoconductive layer), in which diffraction order polarization and intensity were measured as a function of input polarization and applied cell voltage and the results correlated with observations of the VGM structure under polarized illumination in a polarizing microscope, as well as the with various theoretical models of the VGM domain structure.

Three recent papers summarizing these results are included here. The first paper "Physical Characterization of the Variable Grating Mode Liquid Crystal Device" by A.R. Tanguay, Jr., C.S. Wu, P. Chavel, T.C. Strand, A.A. Sawchuk, and B.H. Soffer, has been published in Optical Engineering, special issue on Spatial Light Modulators: Fundamental Characteristics, November/December 1983. This paper summarizes the physical principles of operation of VGM devices and describes work in: a) experimental measurements of the Jones matrix describing

polarized light propagation through the VGM cell; b) the thickness dependence of the molecular orientation angles; c) understanding of the physical constraints on the VGM response time; and d.) the possibility of a VGM device with ac bias.

A second paper "Polarization Properties of the Variable Grating Mode Liquid Crystal Device" by A.R. Tanguay, Jr., P. Chavel, T.C. Strand, C.S. Wu, and B.H. Soffer, has been published in Optics Letters and is also included here. This paper contains detailed results on the spatial distribution of the orientation of liquid crystal molecules in a VGM device. Experimental measurements of the polarization properties of light diffracted by the liquid crystal birefringent phase grating have been made as a function of the applied voltage across the cell.

A final paper, "Intensity-to-Spatial Frequency Transformations in Optical Signal Processing," by A.A. Sawchuk, T.C. Strand and B.H. Soffer describes applications of VGM devices to perform optical nonlinear transformations and logic operations.

Physical characterization of the variable grating mode liquid crystal device

A. R. Tanguay, Jr.
C. S. Wu
P. Chavel*
T. C. Strandt
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. 2, 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.

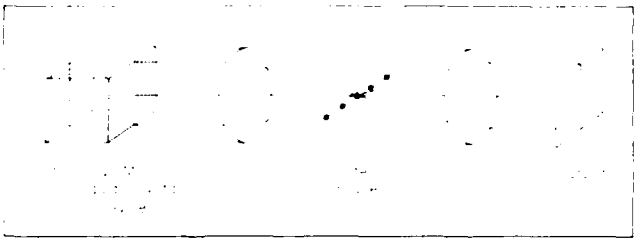


Fig. 1. Experimental arrangement for demonstration of intensity-to-position encoding by means of an intensity-to-spatial frequency conversion in a VGM LCD.

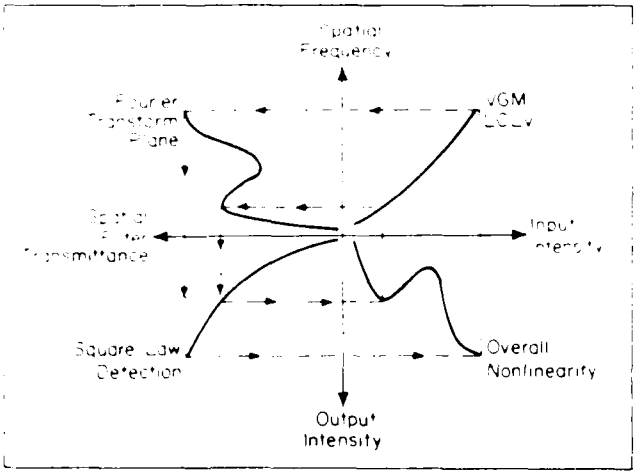


Fig. 2. VGM nonlinear processing. The overall input-output characteristic can be found by stepping through the successive nonlinear transformations, including (1) the intensity-to-spatial frequency conversion, (2) spatial filtering, and (3) intensity detection.

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,[7,8] 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

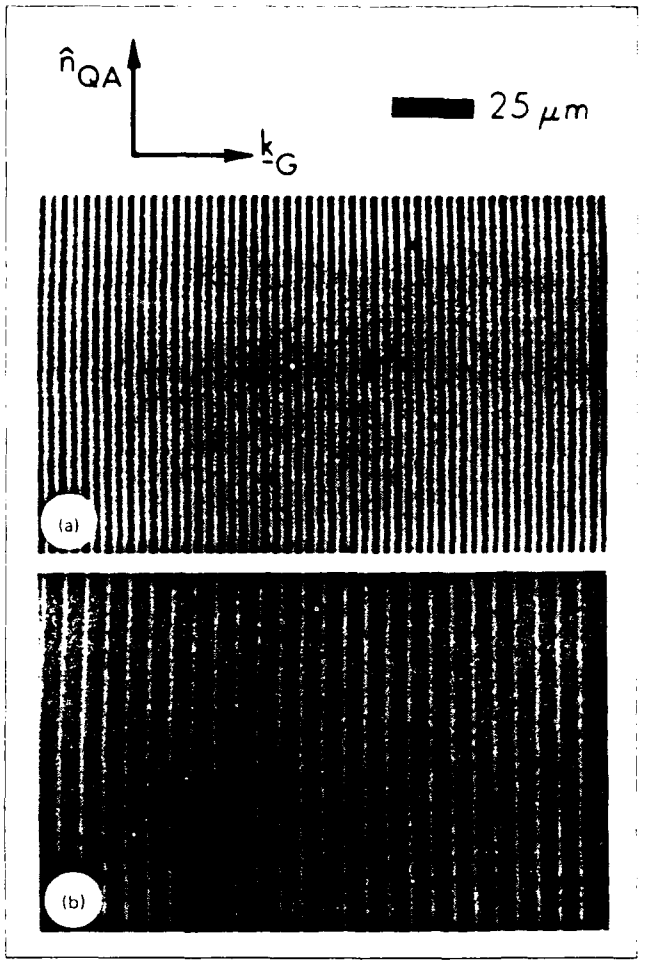


Fig. 3. Polarization photomicrographs of the liquid crystal domain pattern in an electrically activated cell. In (a), the polarizer was oriented at 90°, and the analyzer at 90°, with respect to the grating wave vector. In (b), the polarizer was oriented at 90°, and the analyzer at 10°, with respect to 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.[4] 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

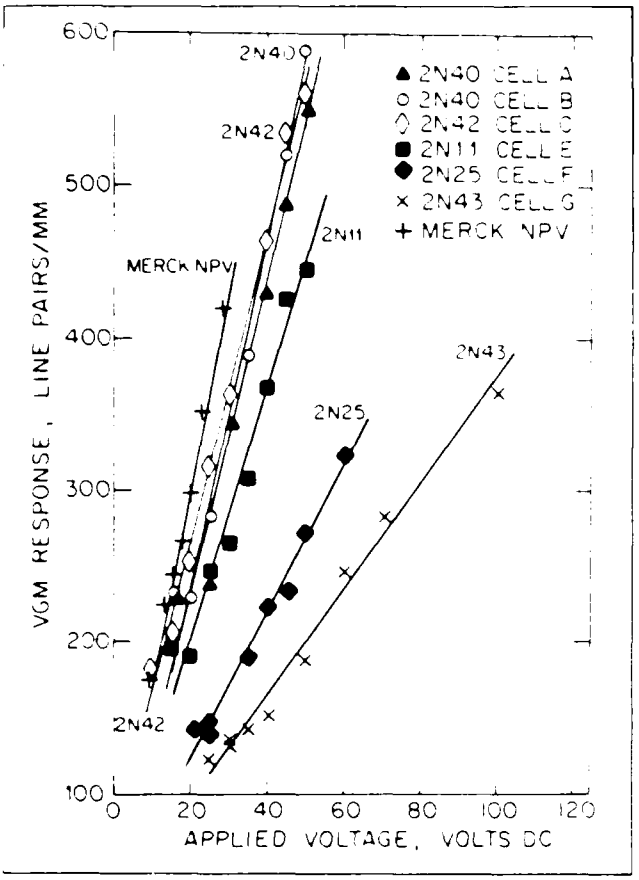


Fig. 4. VGM spatial frequency as a function of applied voltage for various nematic liquid crystal mixtures.

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 ($\rho \geq 10^{10}\ \Omega \cdot$ 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.

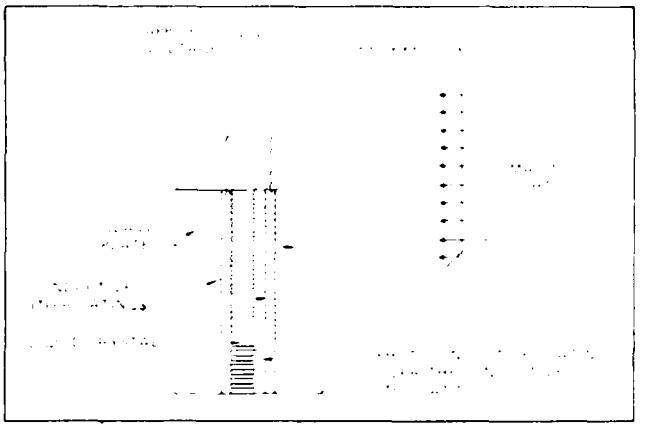


Fig. 5. Schematic diagram of the VGM liquid crystal device. Current devices are read out in transmission at a wavelength of photoconductive insensitivity.

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

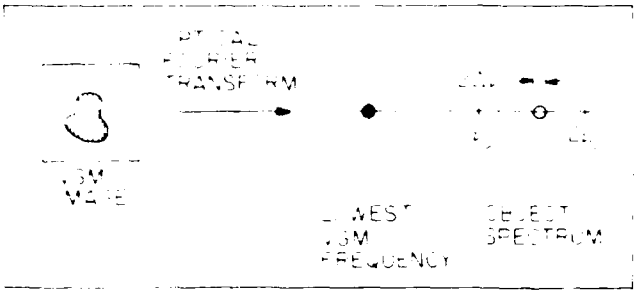


Fig. 6. Gray level resolution. The number of accessible gray levels is limited by the ratio of the separation between VGM harmonics to the object spectrum bandwidth.

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

$$b\nu_0 \geq 2N, \tag{1}$$

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

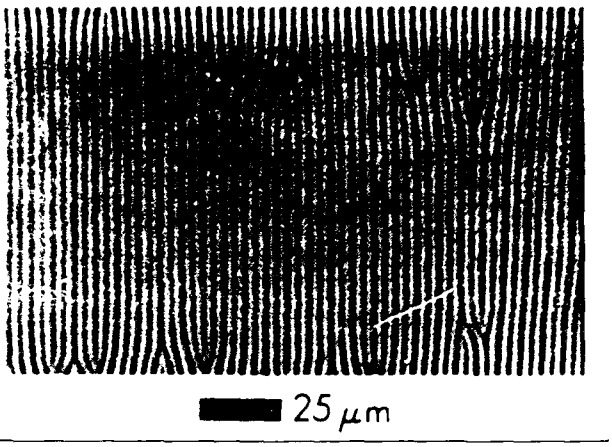


Fig. 7. Polarization photomicrograph of a VGM liquid crystal layer exhibiting a number of grating discontinuities (circled).

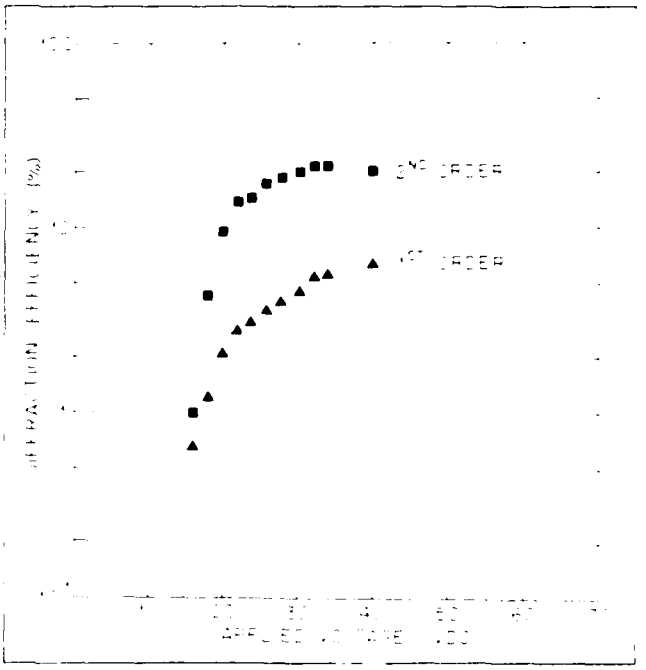


Fig. 8. Diffraction efficiency as a function of applied voltage across a nematic liquid crystal mixture of phenyl benzoates (HRL 2N40). The layer thickness was approximately 6 μm, as defined by a perimeter Mylar spacer.

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 (μW 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 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

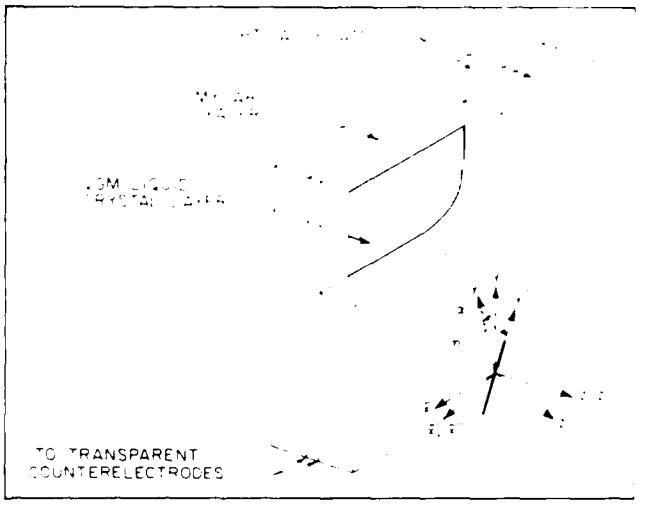


Fig. 9. Variable grating mode liquid crystal test geometry showing the Cartesian coordinate system referred to in the text as well as the molecular orientation angles α and η. This configuration was utilized in the polarized light diffraction efficiency and photomicroscopy experiments.

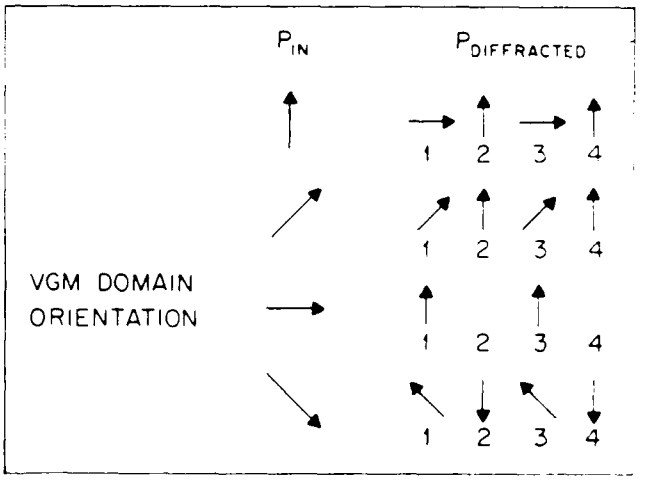


Fig. 10. The polarization behavior of VGM diffracted orders, with illumination normal to the plane of the liquid crystal layer. The left-hand column indicates the input polarization associated with each row of output polarizations. The inset shows the corresponding orientation of the VGM grating.

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

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

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

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

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

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

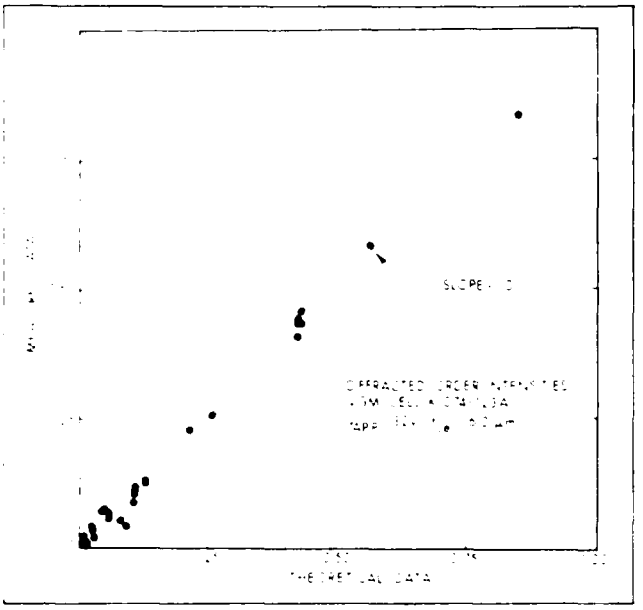


Fig. 11. Measured diffracted order intensities as a function of theoretical intensities calculated from the uniaxial VGM model described in the text.

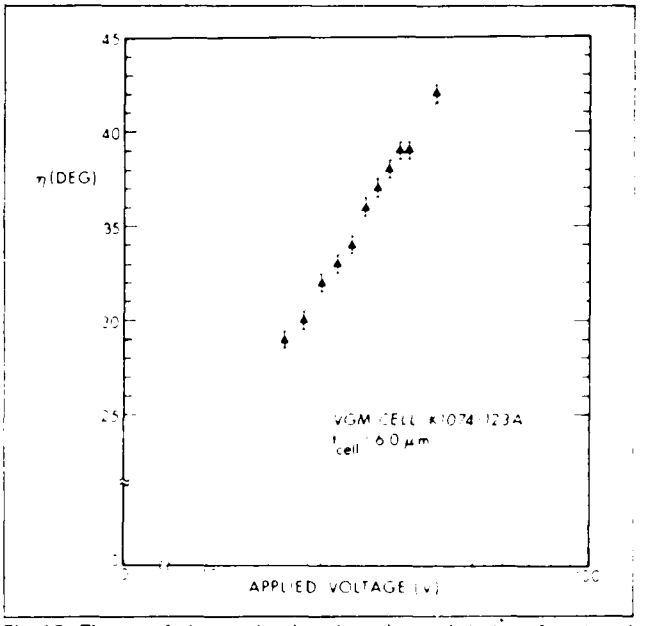


Fig. 13. The out-of-plane molecular orientation angle (η_o) as a function of the applied dc bias voltage across the cell (V) (see Fig. 5).

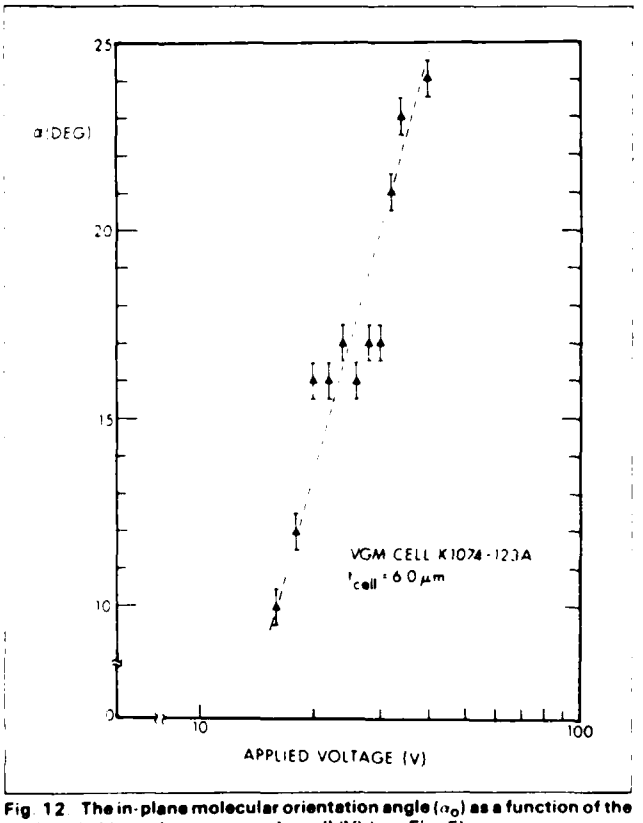


Fig. 12. The in-plane molecular orientation angle (α_o) as a function of the applied dc bias voltage across the cell (V) (see Fig. 5).

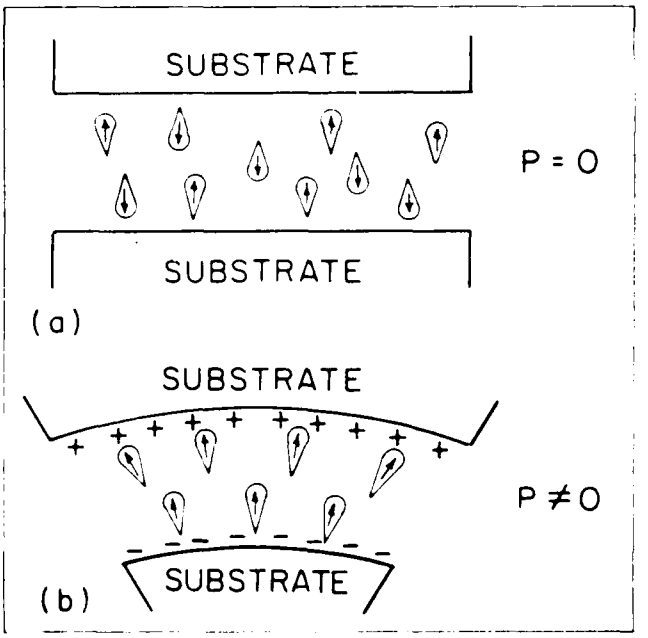


Fig. 14. Schematic diagram illustrating a possible mechanism for the occurrence of the flexoelectric effect due to a shape anisotropy in liquid crystal molecules with a permanent dipole moment. In (a), the molecular orientations are randomly distributed, resulting in zero net polarization. In (b), the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

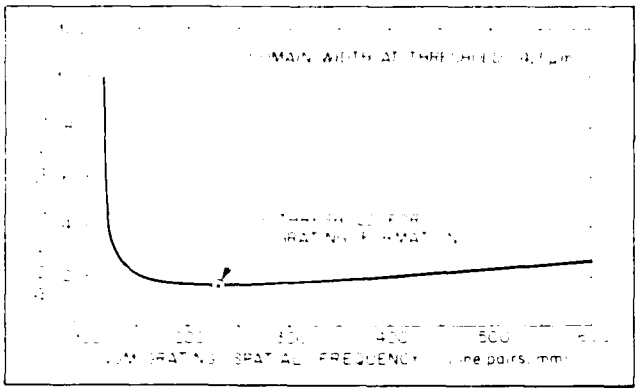


Fig. 15. Resultant theoretical dispersion relation for liquid crystal parameters corresponding to HRL 2N40 phenyl benzoate mixture [see Eq. (7)], subject to the assumption that $K_1 = K_2 = K$.

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., J. 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, J. 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. I. 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. I. K. Vistin', Sov. Phys.-Dokl. 15(10), 908(1971).
28. I. 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).

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)

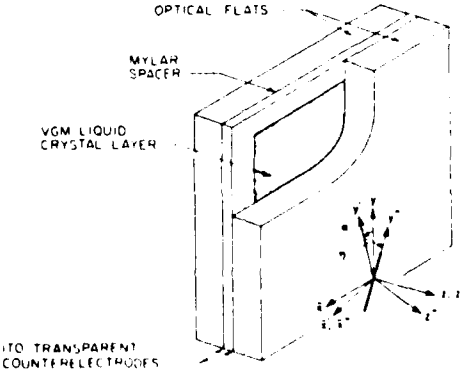


Fig. 1. VGM liquid crystal test geometry showing the Cartesian coordinate system referred to in the text as well as the molecular orientation angles σ and η. This configuration was utilized in the polarized light diffraction efficiency and photomicroscopy experiments.

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

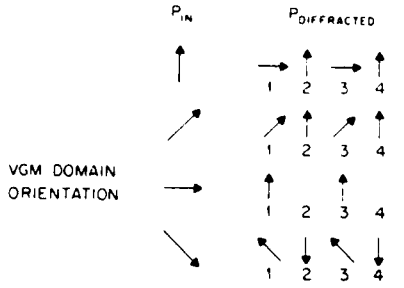


Fig. 2. The polarization behavior of VGM diffracted orders. The left-hand column indicates the input polarization associated with each row of output polarizations. The inset shows the direction of VGM-domain orientation.

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

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

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:

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

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

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

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 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

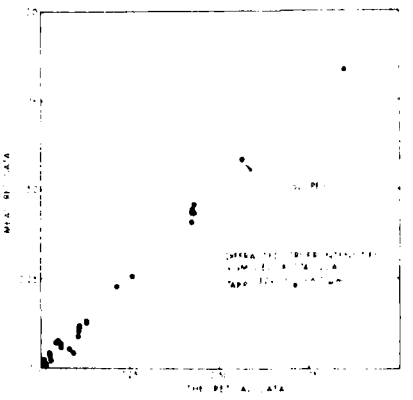


Fig. 3. Measured diffracted-order intensities for a set of polarizer analyzer orientations as a function of theoretical intensities calculated from the uniaxial VGM model described in the text.

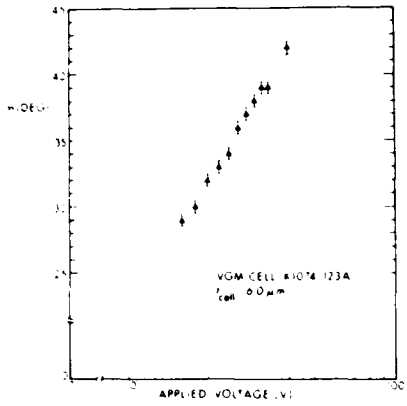


Fig. 4. The out-of-plane molecular orientation angle, η_{max}, as a function of the applied dc bias voltage across the cell, V.

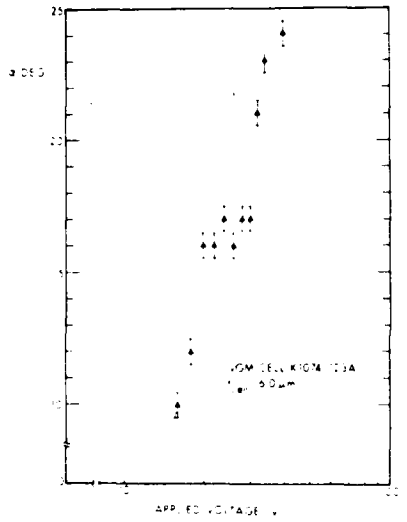


Fig. 5. The in-plane molecular orientation angle, α_{max}, as a function of the applied dc bias voltage across the cell, V.

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[†]), 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).

Intensity-to-spatial frequency transformations in optical signal processing

A. A. Sawchuk
T. C. Strand
University of Southern California
Image Processing Institute
Department of Electrical Engineering
Los Angeles, California 90089

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

Abstract. Intensity-to-spatial frequency transformations can be exploited for a variety of useful optical information processing operations. Resolution elements of an image are encoded with a grating structure whose spatial frequency and/or orientation is a function of the local image intensity. Assuming that certain sampling requirements are met, each intensity level is assigned to a different point in Fourier space. Various schemes of spatial filtering can then be used to alter the relative intensity levels. The procedure can be used for nonlinear analog point transformations and for numerical processing using binary or residue arithmetic. Theta modulation and frequency modulation are special cases of the technique. Off-line methods for implementing intensity-to-spatial frequency transformations include halftoning and grating techniques; real-time methods include new electro-optical systems such as the variable-grating-mode (VGM) liquid crystal devices. Details of these devices and experimental results are presented.

1. INTRODUCTION

In recent years, there has been great interest in expanding the range of operations that can be performed by optical information processing systems. Linear operations such as correlation, convolution, and Fourier spatial filtering can be performed easily because of the inherent linear nature of the system.[1] Three major classes of techniques for nonlinear processing involving both analog and discrete-level signals have been studied, and various electro-optical input devices have been developed that allow real-time operation.[2,3] The first class of techniques utilizes special halftone screens and an optical threshold device to convert continuous optical signals into sampled pulse-width modulated signals. Selective spatial filtering of the pulse-width modulated signal produces a variety of point nonlinear mappings. The second class of techniques uses real-time input devices having an inherent nonlinear input-output characteristic function. The choice of operating point on the device may provide the desired nonlinearity, or the device may be combined in a system with threshold devices.

In this paper, we concentrate on the third class of nonlinear processing techniques. These techniques use intensity-to-spatial frequency transformations in a general sense to achieve both analog and numerical processing. Devices for real-time implementation such as the variable-grating mode (VGM) liquid crystal light valve are described.

2. PROCESSING TECHNIQUES

The basic idea of intensity-to-spatial frequency conversion is to encode resolution elements or regions of an image with a grating modulation in which the spatial frequency and/or orientation of the grating is a function of the local image intensity. Assuming that certain sampling requirements are met, each intensity level in the input is assigned to different points in the Fourier domain; all points with the same intensity in the image are assigned to the same point in Fourier space (assuming space-invariant operation is desired). A purely absorbing spatial filter placed in the Fourier plane can then alter the relative intensity levels in an arbitrary way.

Figure 1 illustrates the basic principle by an optical level-slice example. The quantity S in Fig. 1 is used as a generalized spatial frequency coordinate. A device is used to convert the local input intensity I_{IN} to spatial frequency S in Fig. 1(a). Figure 1(b) shows a Fourier plane filter that selectively alters different components as a function of S. Figure 1(c) shows the overall I_{IN} versus I_{OUT} characteristic found by graphically tracing through the two characteristics of Fig. 1(a) and 1(b).

Figure 2 shows several examples of intensity-to-spatial frequency mappings. In the most general case of intensity-to-spatial frequency conversion, I_{IN} is mapped to an arbitrary curve S in the frequency plane, as shown in Fig. 2(a). Locations along this curve can be specified in polar coordinates as (ϱ,θ). A special case, shown in Fig. 2(b), is called theta (θ) modulation. In θ modulation S varies along an arc of constant radius ϱ. Figure 2(c) shows the special case of frequency modulation, corresponding to paths along a line at constant angle θ.

Intensity-to-spatial frequency conversion systems may be easily programmed to change the functional nonlinearity by placing different spatial frequency filters in the Fourier plane. This programmability is an important advantage of such systems.

3. SAMPLING CONSIDERATIONS

In all the intensity-to-spatial frequency transformations shown in Fig. 2, there is a fundamental relationship between the number of intensity levels to be distinguihed, the spatial-frequency bandwidth

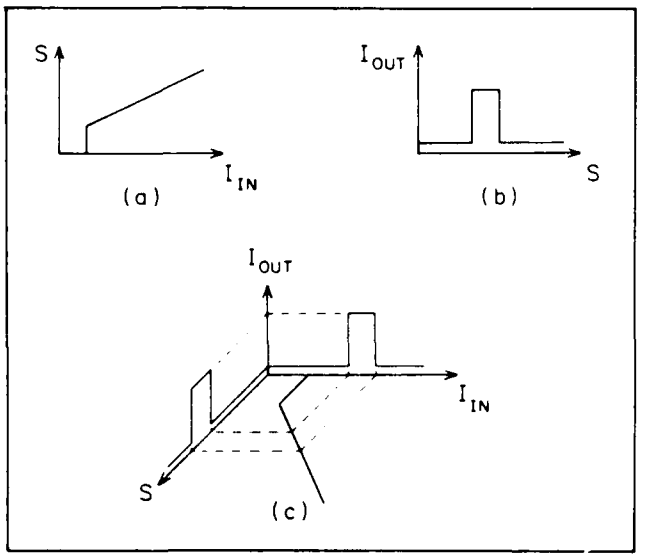


Fig. 1. Intensity-to-spatial frequency conversion for nonlinear processing. (a) A device is used to convert input intensity into a varying spatial frequency S. (b) A Fourier plane filter selectively attenuates the spatial frequency component S. (c) The I_{OUT} vs. I_{IN} characteristic of the overall filtering system can be found graphically by tracing through the two characteristics of (a) and (b).

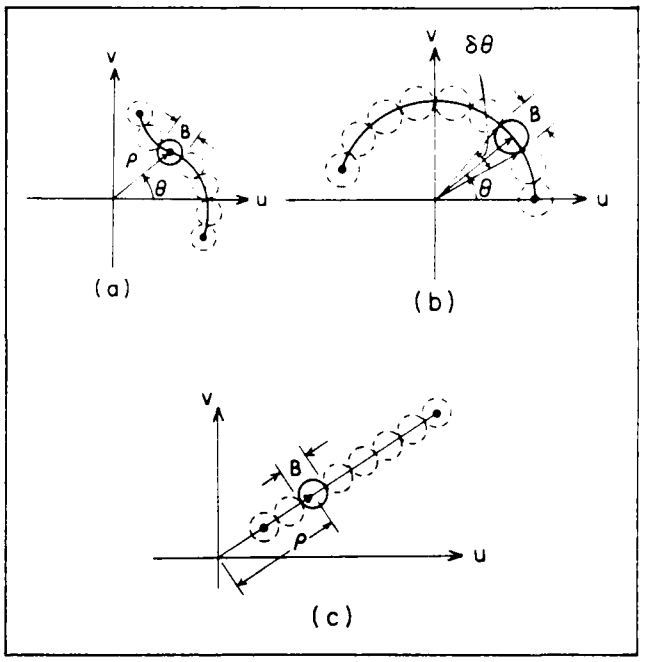


Fig. 2. Intensity-to-spatial frequency modulation: (a) general case, (b) theta modulation, (c) frequency modulation.

of the input function I_{IN}, and the spatial frequency required for the modulated grating. A grating with a given fundamental frequency (ϱ,θ) in polar coordinates, corresponding to a specific input intensity, produces a diffraction order at (ϱ,θ) surrounded by a spectral island whose dimension is roughly equal to the bandwidth B of the input. If N distinct intensity levels are to be uniquely processed by the system, the curve S must be chosen so that N object spectra can be placed along it without overlap, as shown in Fig. 2. With θ modulation and the modulated input assumed a real function, ϱ is a constant (ϱ_0), and the curve S is restricted to an arc of π radians because of conjugate symmetry of the transform. If the number of intensity levels N is large, we can assume that the angular width $\delta\theta$ of the object spectrum is

$$\delta\theta \cong B/\varrho_0 . \tag{1}$$

If N distinct levels are needed, then we must satisfy the condition

$$N\delta\theta = \frac{NB}{\varrho_0} \le \pi \tag{2}$$

or

$$\varrho_0 = \frac{NB}{\pi} . \tag{3}$$

For frequency modulation, the useful modulation range is limited because the grating generally produces multiple diffraction orders. If the lowest fundamental grating frequency is ϱ_0 and its first significant harmonic is ϱ_1, the usable frequency modulation range $\Delta\varrho$ is

$$\Delta\varrho = \varrho_1 - \varrho_0 . \tag{4}$$

If we assume there are no missing orders in the grating spectrum, then

$$\varrho_1 = 2\varrho_0 \tag{5}$$

and we have

$$\Delta\varrho = \varrho_0 . \tag{6}$$

The input spectrum has width B; to place N of them in the Fourier domain without overlap we must have

$$NB \le \Delta\varrho = \varrho_0 \tag{7}$$

or

$$\varrho_0 \ge NB . \tag{8}$$

Because the bandwidth B is equal to twice the highest spatial frequency in the input, Eqs. (3) and (8) provide the relationship between the required sampling frequency ϱ_0 and the highest input spatial frequency. If N is large, the very high spatial frequency gratings required may be the limiting factor. Fortunately, the optical system does not need to image these gratings because only the modulation of the grating is important. The second part of the system is a spatial filtering of the spectrum in the neighborhood of a single fundamental diffraction order. It is this limited spectrum that determines the required bandwidth, or resolution, of the optical system.

4. IMPLEMENTATION

Of the intensity-to-spatial frequency conversion techniques described, θ modulation was the first to be discussed and demonstrated.[4,5] The actual implementation of the spatial frequency encoding in both θ modulation and frequency modulation still must be considered. One solution to the θ modulation problem is to use special halftone screens.[6] The screen profile combined with a thresholding copy step produces halftone dots whose edges are oriented as a function of the original input intensity.

The θ modulation method is particularly well suited to processing with white light because the spectra at various orientations are more

separated than those in a pure frequency modulation scheme. Indebetouw[7] used halftone θ modulation techniques to achieve pseudocoloring of a monochromatic image. A Ronchi ruling (one-dimensional constant period grating) of low contrast is used in contact with the continuous tone input to sequentially expose a high contrast film. For each exposure there is a different angular orientation of the Ronchi ruling and a different exposure time. Assume for the moment an intermediate exposure time. For low transmittance areas of the input, the light passing through the sandwich of input transparency and Ronchi ruling will be too weak to expose the high contrast film. For high transmittance areas of the input, on the other hand, light passing through the sandwich will completely expose the film. For a certain narrow range of input transmittances (determined by the Ronchi ruling contrast), the light passing through the sandwich impresses a local modulated grating structure on the high contrast film. By changing the exposure time and rotating the screen between exposures, different level-sliced versions of the input are encoded as directional information on the high contrast film. The resultant transparency is then placed in a spatially coherent optical spatial filtering system and illuminated with collimated light from a white light point source. In the spatial frequency plane, level slice information recorded for different ranges of input image intensity values appears as lines of diffraction spots oriented at different angles. This pattern of spots can be modified by an angular array of color filters. After inverse transforming, the desampled continuous-space image has various colors associated with different bands of gray levels. The pseudocolored image resulting from this method can be formed, either in parallel as described, or as a sequence of reconstructions from a set of spectrally narrowband light sources.

Lee has recently described a system that can achieve the general modulation of Fig. 2(a) or the special cases of θ or frequency modulation.[3] A scanning interferometric system is used to encode each spatial resolution element of an image as a small grating segment whose frequency and orientation can be changed with input intensity levels. The system does not currently operate at television rates (30 frames per second or faster) but can utilize the two-dimensional nature of the Fourier plane better than either the θ or frequency modulation techniques.

5. VARIABLE-GRATING MODE (VGM) DEVICES

Fast implementation of intensity-to-spatial frequency transformations by frequency modulation has recently become possible with the development of variable-grating mode (VGM) liquid crystal (LC) electro-optical devices.[8-10] The VGM device is shown schematically in Fig. 3. It consists of a thin (thickness < 12 μm) nematic LC layer sandwiched with a photoconductor between two transparent ITO (indium tin oxide) electrodes. If a dc voltage is applied across the liquid crystal cell, a periodic phase grating structure is produced in the LC material. As the voltage is varied, the period of the grating changes as shown in Fig. 4. The fundamental VGM spatial frequency is a roughly linear function of applied voltage for several different LC mixtures. Typically the spatial frequency varies from 200 to 600 line pairs/mm. Figure 5 shows the LC material in a cell as viewed through a polarizing microscope.

The voltage across the LC cell can be locally controlled by projecting an input image on the VGM device at a wavelength at which the photoconductor is sensitive. Thus the VGM device serves as an intensity-to-spatial frequency converter. The photoconducting layer of the cell is designed to accept most of the bias voltage with no illumination; that portion of the voltage falling across the LC is below the activation threshold of the VGM effect. When light falls on the photoconductor, its impedance drops, thereby switching the voltage from the photoconductor onto the LC material and activating the VGM effect. The high lateral impedance of the thin film layers allows very little spreading of the photo-induced conductivity pattern or of the associated LC electro-optic effect. As a result, the light activation has inherently high resolution, and the device accepts photographic quality images. Several VGM devices

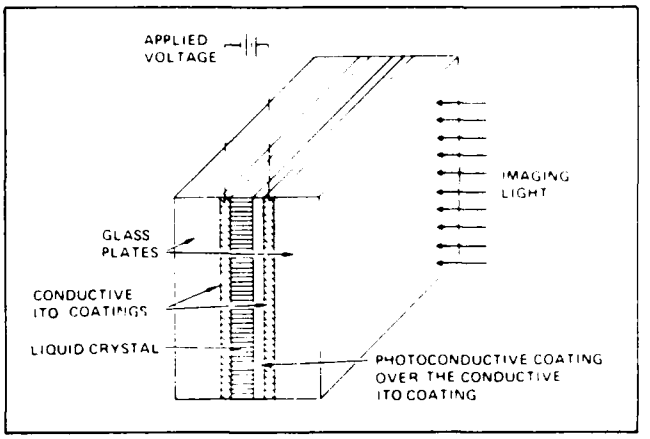


Fig. 3. Schematic diagram of the VGM device construction. Current devices are read out in transmission at a wavelength at which the photoconductor is insensitive. ITO (indium tin oxide) is a transparent conducting electrode material.

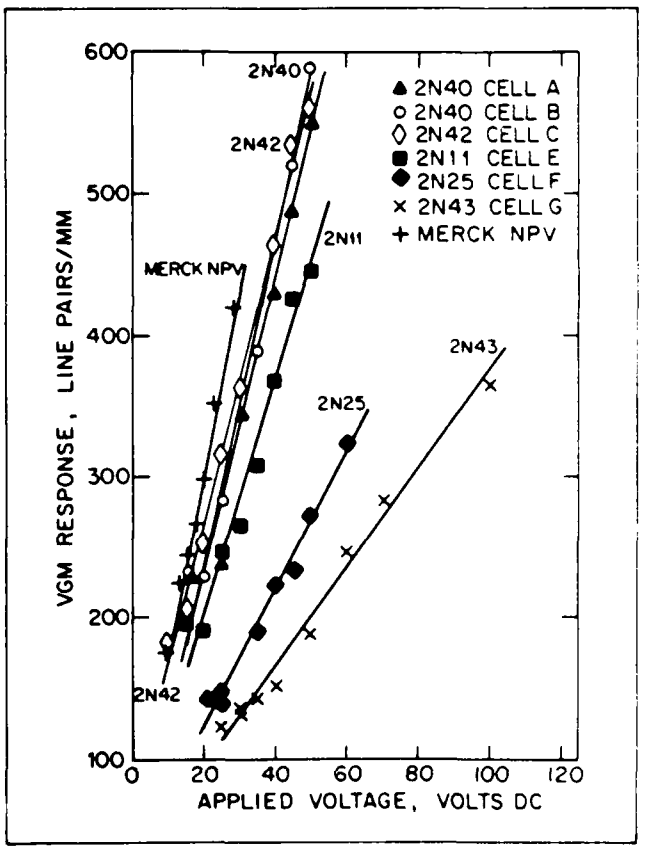


Fig. 4. VGM spatial frequency as a function of applied voltage

have been fabricated by Hughes Research Laboratories, and extensive modeling and experimental work is underway to physically characterize the VGM mechanism. Typical devices have a ZnS photoconductor. Construction details of the VGM device and a discussion of LC physics are given in several references.[8-10]

6. VGM IMPLEMENTATION OF A LEVEL-SLICE

An experimental implementation of a level slice function using a

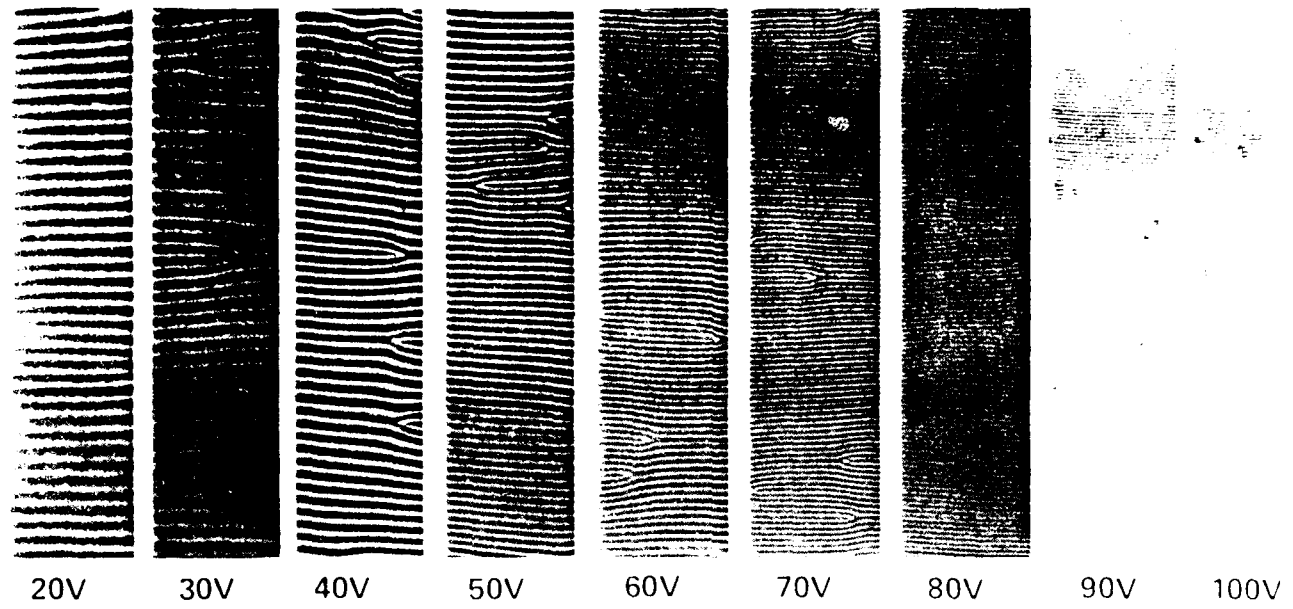


Fig. 5. VGM viewed through polarizing microscope as a function of applied voltage.

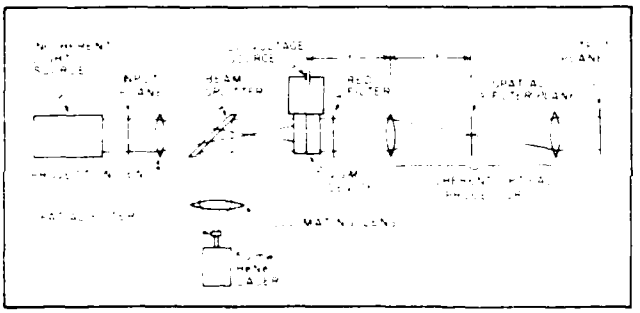


Fig. 6. Experimental setup used to perform the level slice experiments. The spatial filter is a variable annular aperture.

VGM device has been performed. The experimental arrangement is shown in Fig. 6. A continuous tone input picture is illuminated by an arc lamp source and imaged onto the photoconductor surface of a VGM device that initially exhibits a uniform phase grating structure due to a dc bias voltage. The grating period is locally modulated by the input picture intensity, and this modulation is mapped into a position along a line in the spatial filter plane. A red filter ensures that only the readout laser beam enters the coherent optical processor. Sectors of small circular annuli of varying radii are used to pass certain spatial frequency bands. This allows only prescribed input intensity ranges to appear in the output. Circular rather than straight slits are used to capture the weak light which in small part is diffracted into circular arcs because of the grating imperfections (see Fig. 4). Figure 7 shows both the input and level-sliced output pictures. Figure 7(a) shows a positive print of the original image through the system as it appears on the imaging screen. A negative of the original was used in the experiments. Figure 7(b) shows a low intensity level slice corresponding to a VGM spatial frequency of 100 line pairs/mm with approximately 15% fractional bandwidth. Here fractional bandwidth is defined as the slit width in the Fourier plane divided by distance between first and second diffraction orders. Figures 7(c) through 7(f) are a sequence of level slice outputs at successively higher input intensity levels. The Fourier plane slit having the same 15% fractional bandwidth is moved closer to the second diffraction order to obtain this sequence. Figure 7(f) shows a level slice result very close to the second diffraction order at 200 line pairs/mm. Some of the noise visible in Fig. 7(b) and 7(c) is due to surface imperfections on the VGM device. Also, interference fringes visible in Fig. 7(d) and 7(e) are due to reflections in the external apparatus.

7. VGM IMPLEMENTATION OF OPTICAL LOGIC

It is possible to use the VGM device to implement logic operations. The key to such operation lies in noting that a logic gate can be represented as a simple binary nonlinearity [10]. The input-output characteristics of four basic logic functions are shown in Fig. 8. The input in these figures is assumed to be the simple arithmetic sum of the two input logic levels. Thus NOT is simply a hard-clipping inverter, AND and OR are nonlinear threshold functions, and XOR is a level slice function. These functions can be realized by light modulators or by optical processing systems that can produce these nonlinearities, and can be used for implementing combinational logic. Furthermore, if the system allows feedback to be readily introduced, sequential or latching logic can also be implemented.

The VGM device is well suited to implementing these nonlinearities. With the VGM approach, nonlinearities are obtained using a Fourier plane filter whose transmittance variation in one dimension is essentially a plot of the desired nonlinearity. Thus arbitrary nonlinearities are easily produced and changed. For logic operations the situation is quite straightforward. Since the nonlinearities associated with logic operations are binary functions, they can be implemented with simple slit apertures, i.e., 0 or 1 transmittance values. A feature of the VGM technique that makes it especially well suited for logic processing is that the input and output are physically separate beams. The input beam modulates a photoconductor; concurrently, the image is read out with a second beam. This separation of input and output provides for the possibility of restoring the output levels to the 0 and 1 values even if the input levels are not exactly correct. This feature is essential to the production of a reliable logic system that is immune to noise and systematic errors in the levels. Electronic logic elements possess such level restoring capability, but other currently proposed optical logic schemes lack this essential characteristic.

Two further aspects of logic operations make them advantageous

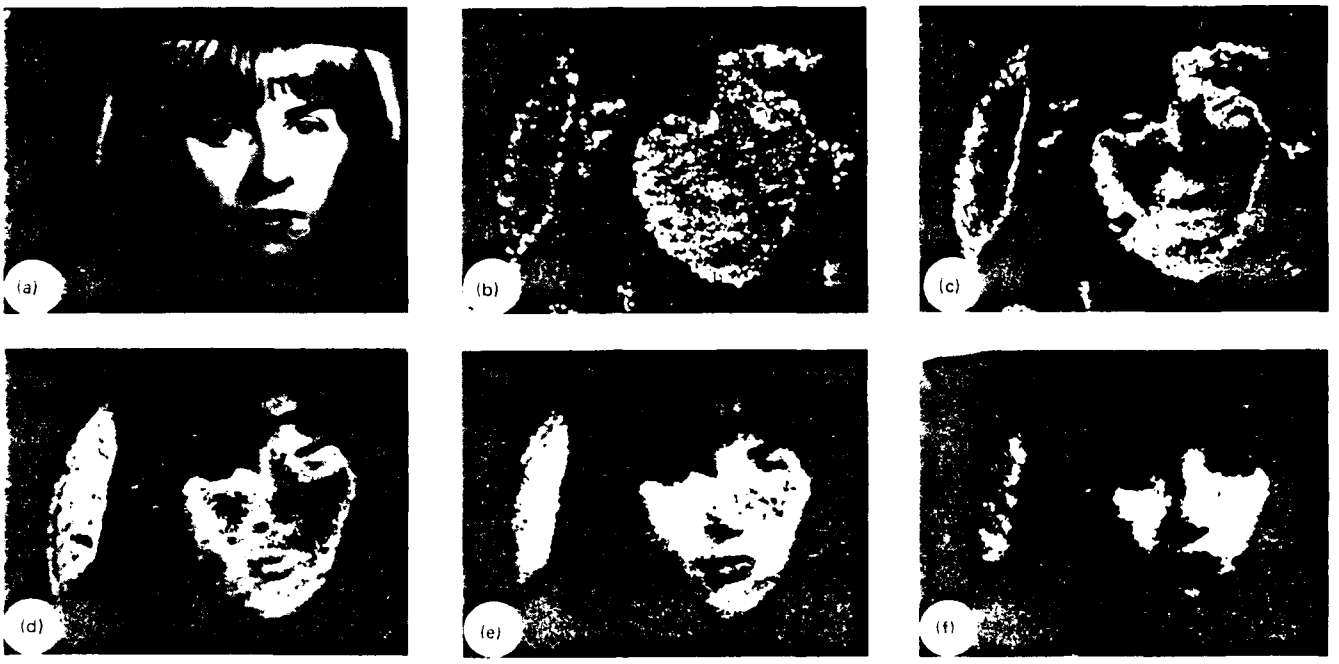


Fig. 7. VGM level slice results.

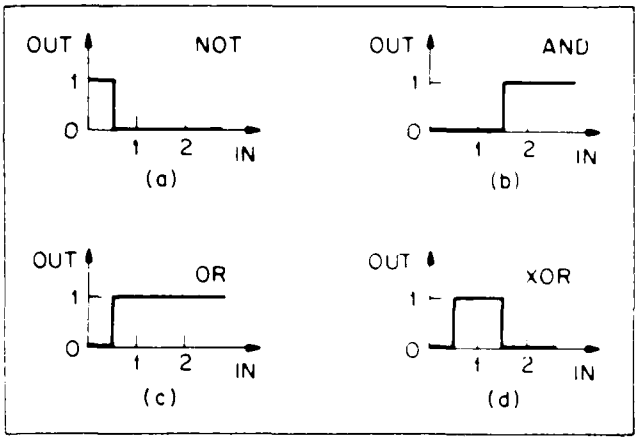


Fig 8. Logic functions as simple nonlinearities. Given an input consisting of the sum of two binary inputs, different logical operations can be effected on those inputs by means of the depicted nonlinear characteristics: (a) NOT, (b) AND, (c) OR, and (d) XOR.

for optical implementation purposes: (a) logic operation input levels are discrete, and (b) only a small number (2-4) of distinguishable levels are required. The existence of discrete input levels implies that the nonlinearities need not have the exact forms shown in Fig. 8. In particular the transition regions need not be perfectly sharp thresholds, since the input values are presumed not to occur within the transition regions in any event. The fact that only a few levels need be distinguished implies that the gray-level resolution requirements for the system are minimal.

The operations discussed above are the basic combinatorial logic functions. However, sequential logic may also be implemented with appropriate feedback, i.e., imaging the output plane onto the input plane (which can be accomplished with incoherent illumination). However, with the present VGM transmission devices there is a problem of separating the read and write functions in a feedback system since their wavelengths must be identical and current devices work in transmission.

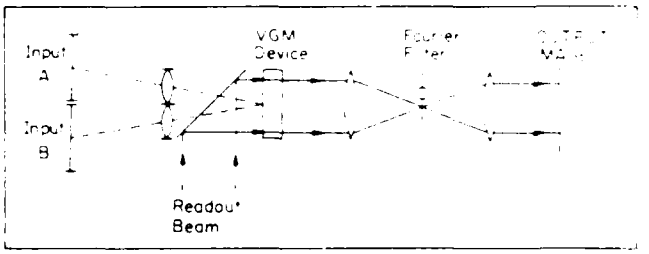


Fig. 9. Experimental arrangement for performing logical operations on two-dimensional binary inputs with a VGM device. Input image arrays A and B are superimposed on the photoconductor. The device is read out in transmission at a wavelength different from the input. Simple slit apertures are used to achieve the desired logic functions.

A series of experiments was conducted to demonstrate the fundamental logic functions. The experimental setup of Fig. 9 was used. The total illumination intensity on the photoconductor of the VGM device was the sum of the two input intensities and a bias intensity. The input illumination was provided by a filtered high-pressure mercury arc lamp, the bias illumination by a collimated tungsten bulb source. The VGM device was read out in transmission using a HeNe laser. A filter was placed in the Fourier plane to select the desired diffraction orders for each logic function.

For these experiments, the inputs consisted of one vertical rectangular aperture and one horizontal aperture. When these were superimposed along with the bias, a square image was formed with the four quadrants having the intensity levels shown in Fig. 10. This image corresponds to the logic truth table shown. Thus the output images have intensity levels determined by the truth table associated with the desired logic function. The logic functions AND, OR, XOR, and their complements were implemented sequentially as shown in Fig. 10 by altering only the Fourier plane filter. Imperfections visible in the output plane data arise from defects in the cell structure of the VGM device employed in these experiments. Investigations on VGM device improvements, limitations, processing speed, and pixel uniformity are in progress.

By adding feedback to optical combinatorial logic, memory

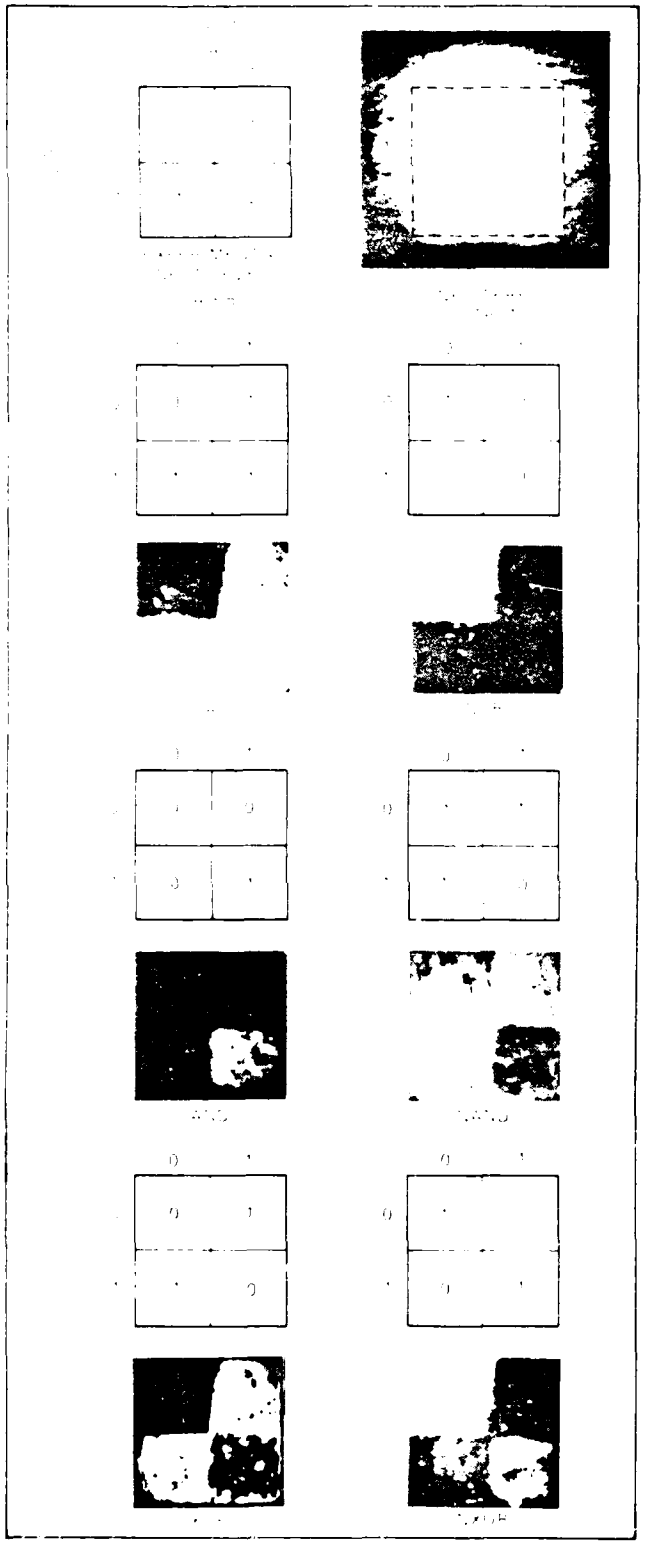


Fig. 10. VGM logic results. The right-hand column indicates ideal output levels for an image consisting of four quadrants corresponding to truth table values. The left-hand column shows the corresponding experimental results. The first row shows the input for all experiments, consisting of a superposition of two binary images. Succeeding rows show results for the logic operations OR, NOR, AND, NAND, XOR, and NXOR, respectively.

elements such as latches or flip-flops can be implemented. Other types of numerical processing such as residue arithmetic based systems can also be implemented with intensity-to-spatial frequency conversion and VGM devices.

8. OTHER APPLICATIONS OF VGM DEVICES

Variable grating cells have a number of other possible applications in both electrically activated and photoconductor activated systems. Electrically controllable variable gratings could be used for modulation transfer function (MTF) tests in optical systems, as beam deflectors, or as optical spatial demodulators or reference masks for correlators. By introducing a nonuniform cell thickness a spatially variable electrically activated grating such as a chirp function could be formed. Photoconductor activated structures could also serve as programmable spatially variable reference gratings.

9. CONCLUSIONS

We have described several general techniques for intensity-to-spatial frequency transformations and an electro-optical variable-grating-mode (VGM) transducer capable of real-time implementation. We have shown theoretically and experimentally that the device can be used for implementing arbitrary nonlinearities on continuous tone images and can be used for combinatorial logic operations on binary data arrays. An advantage of the VGM approach is that it may easily be programmed by low resolution optical spatial filtering. Other applications for intensity-to-spatial frequency conversion have been described.

ACKNOWLEDGMENTS

We wish to acknowledge the support of the Air Force Office of Scientific Research, Electronics and Materials Sciences Division, under Grants AFOSR-77-3285 and AFOSR-81-0182 at USC, and Contracts F49620-77-C-0080 and F49620-81-C-0086 at Hughes Research Laboratories. We also acknowledge support from the National Science Foundation, under Grant ENG78-11368, and the Joint Services Electronics Program (AFSC) at USC.

The authors are grateful to A. Armand, P. Chavel, and A. R. Tanguay, Jr. for their contributions to the research described in this paper.

REFERENCES

1. J. W. Goodman, *Introduction to Fourier Optics*, McGraw-Hill, New York, 1968.
2. A. A. Sawchuk and T. C. Strand, "Fourier optics and digital optical processing," in *Applications of Optical Fourier Transforms*, H. Stark, ed. (Academic Press, New York, 1982), Ch. 9.
3. S. H. Lee, "Nonlinear optical processing," in *Optical Information Processing Fundamentals*, S. H. Lee, ed. (Springer-Verlag, Berlin, 1981), 261.
4. B. Morgenstern and A. W. Lohmann, "Analysis of nonlinear properties," Optik 20, 450 (1963).
5. J. D. Armitage and A. W. Lohmann, "Data processing by nonlinear techniques," Appl. Opt. 4, 399 (1965).
6. W. Schneider, "Spatial pulse modulation for optical signal processing and stereometry," Optica Acta 21, 563 (1974).
7. G. Indebetouw, "Production of color-coded transparencies by nonlinear filtering," Appl. Opt. 16, 1951 (1977).
8. B. H. Soffer, D. Boswell, A. M. Lackner, A. R. Tanguay, Jr., T. C. Strand, and A. A. Sawchuk, "Variable grating mode liquid crystal device for optical processing," in *Digital Image Processing and Analysis/Optical Signal Processing*, T. C. Strand and A. R. Tanguay, Jr., eds., Proc. SPIE 218, 81 (1980).
9. B. H. Soffer, D. Boswell, A. M. Lackner, P. Chavel, A. A. Sawchuk, T. C. Strand and A. R. Tanguay, Jr., "Optical computing with variable grating mode liquid crystal devices," in *1980 International Optical Computing Conference*, Book II, W. T. Rhodes, ed., Proc. SPIE 232, 128 (1980).
10. B. H. Soffer, J. D. Margerum, A. M. Lackner, D. Boswell, A. R. Tanguay, Jr., T. C. Strand, A. A. Sawchuk, and P. Chavel, "Variable grating mode liquid crystal device for optical processing and computing," Mol. Cryst. Liq. Cryst. 70, 145 (1981).
11. 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).

1.5 References

1. A.A. Sawchuk and T.C. Strand, "Fourier Optics in Nonlinear Signal Processing," Ch 9 of Applications of Optical Fourier Transform, H. Stark, ed., Academic, New York, 1981.

2. A. Armand, A.A. Sawchuk, T.C. Strand, B.H. Soffer and D. Boswell, "Real-time Parallel Optical Analog-to-Digital Conversion," Optics Letters, vol. 5, pp. 129-131, 1980.

3. 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, vol. 5, pp. 398-400, 1980.

4. 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, vol. 70, pp. 145-161, 1981.

5. A.A. Sawchuk, "Intensity-to-spatial Frequency Transformations," Proceedings Society of Photo-Optical Instrumentation Engineers Advanced Institute of Transformations in Optical Signal Processing, Seattle, (February 1981).

6. A.A. Sawchuk, "Recent Developments in Optical and Digital Processing," Proceedings Workshop on Optical Information Processing, Centro de Investigaciones en Optica, Cuernavaca, Mexico, (January 1982).

7. A.R. Tanguay, Jr., P. Chavel, T.C. Strand, J. Wu, and B.H. Soffer, "Physical Characterization of the Variable Grating Model Liquid Crystal Devices," SPIE Proc. Advances in Optical Information Processing, vol. 388, (January 1983).

8. 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, vol. 9, pp. 174-176, 1984.

9. 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," presented at 1981 Annual Meeting, Optical Society of America, Orlando, Florida, October 1981, Journal of the Optical Society of America, vol. 71, p. 1630, (December 1981).

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

11. A.R. Tanguay, Jr., "Polarization Properties of Birefringent Phase Gratings," presented at 1982 Annual Meeting, Optical Society of America, Tucson, Arizona, October 1982: Journal of the Optical Society of America, vol. 72, (1982).

12. A.R. Tanguay, Jr., "Spatial Light Modulators for Real Time Optical Processing," Proc. ARO Workshop on Future Directions for Optical Information Processing, pp. 52-76, Lubbock, Texas, May, 1980.

13. P.G. de Gennes, The Physics of Liquid Crystals Claredon Press, Oxford, (1975).

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

15. R.B. Meyer, Phys. Rev. Lett., vol. 22, 918, (1969).

16. P. Auborg, J.P. Huignard, N. Hareng, and R.A. Mullen, "Liquid Crystal Light Valve Using Bulk Monocrystalline Bi SiO as the Photoconductive Material," Appl. Opt., vol. 21, 3706-3711, (1982).

2. PROFESSIONAL PERSONNEL

The following individuals contributed to the research effort supported by this grant during the period April 15, 1981 to June 30, 1984.

1. Alexander A. Sawchuk, Professor, Department of Electrical Engineering, Director - Image Processing Institute, Principal Investigator.

2. Timothy C. Strand, Research Assistant Professor, Image Processing Institute, Senior Investigator.

3. Armand R. Tanguay, Jr., Associate Professor, Departments of Electrical Engineering and Materials Science, and Image Processing Institute, Senior Investigator.

4. Pierre H. Chavel, Visiting Research Scientist, Image Processing Institute (on leave from Institut d'Optique, Orsay, France).

5. Robert Forchheimer, Visiting Research Scientist, Image Processing Institute (on leave from Linkoping University, Linkoping, Sweden).

6. B. Keith Jenkins, Research Assistant Professor, Department of Electrical Engineering.

7. John Wu, Research Assistant, Ph.D. Candidate, Department of Electrical Engineering.

The following individuals have received degrees from the University of Southern California while participating in the research effort:

1. B. Keith Jenkins, Ph.D., Electrical Engineering, May 1984, thesis title: "Optical Logic Systems: Implementation and Architectural Implications," (published as USCIPI Report 1110).

3. PUBLICATIONS

This section lists written publications resulting from AFOSR support of this project during the period April 15, 1981 to June 30, 1984.

1. A.A. Sawchuk and T.C. Strand, "Fourier Optics in Nonlinear Signal Processing," Ch. 9 of Applications of Optical Fourier Transforms, H. Stark, ed., Academic, New York, (1981).

2. A. Armand, A.A. Sawchuk and T.C. Strand, "Nonlinear Optical Processing with Halftones: Accurate Predictions for Degradation and Compensation," in preparation for Applied Optics.

3. A. Armand, A.A. Sawchuk, T.C. Strand, and B.H. Soffer, "Real-Time Parallel Logarithmic Filtering," Optics Letters, vol. 7, pp. 451-453, (September 1982).

4. B.H. Sotfer, J.D. Margerum, A.M. Lackner, D. Boswell, A.A. Sawchuk, A.R. Tanguay, Jr., T.C. Strand, and P. Chavel, "Variable Grating Mode Liqui' Crystal Device for Optical Processing and Computing," Molecular Crystals and Liquid Crystals, vol. 70, pp. 145-161, (1981).

5. A.A. Sawchuk, "Intensity-to-Spatial Frequency Transformations," Proceedings Society of Photo-Optical Instrumentation Engineers Advanced Institute on Transformations in Optical Signal Processing, vol. 373, pp. 69-75, (February 1981).

6. A.A. Sawchuk, "Recent Developments in Optical and Digital Processing," Kinam, vol. 5, Serie C, pp. 105-112, (1983).

7. A.R. Tanguay, Jr., C.S. Wu, P. Chavel, T.C. Strand, A.A. Sawchuk, and B.H. Soffer, "Physical Characterization of the Variable Grating Mode Liquid Crystal Device," SPIE Proc. Advances in Optical Information Processing, vol. 388, pp. 85-95, (January 1983).

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

9. P. Chavel, R. Forchheimer, B.K. Jenkins, A.A. Sawchuk, and T.C. Strand, "Architectures for a Sequential Optical Logic Processor," Proceedings 10th International Optical Computing Conference, IEEE Cat. No. 83CH1880-4, pp. 6-12, (April 1983).

10. A.R.Tanguay, Jr., C.S. Wu, P. Chavel, T.C. Strand, and A.A. Sawchuk, "Physical Characterization of the Variable Grating Mode Liquid Crystal Device", Optical Engineering, Vol. 22, pp. 687-694 (November/December 1983).

11. B.K. Jenkins, A.A. Sawchuk, T.C. Strand, R. Forchheimer, and B.H. Soffer, "Sequential Optical Logic Implementation," Applied Optics, vol. 23, pp. 3455-3464, (1 October 1984).

12. B.K. Jenkins and T.C. Strand, "Computer-Generated Holograms for Space-Variant Interconnections in Optical Logic Systems, "SPIE Proc., vol. 437, pp. 110-118, (1983).

13. B.K. Jenkins, P. Chavel, R. Forchheimer, A.A. Sawchuk and T.C. Strand," Architectural Implications of Digital Optical Processors", Applied Optics, vol. 23, pp. 3465-3474, (1 October 1984).

14. A.A. Sawchuk, "Digital Logic and Computing with Optics," SPIE Proc., vol. 456 (1984).

15. A.A. Sawchuk and T.C. Strand, "Digital Optical Computing," Proc. IEEE, vol. 72, pp. 758-770 (1984).

16. B.K. Jenkins and A.A. Sawchuk, "Characteristics of Digital Optical Processors," Proceedings of the IEEE Global Telecommunications Conference, Atlanta, Georgia (November 1984).

17. B.K. Jenkins, A.A. Sawchuk, and T.C. Strand, "Spatial Light Modulator Requirements for Sequential Optical Logic," in preparation for Applied Optics.

18. A.A. Sawchuk, "Numerical Optical Computing Techniques," Proc. 13th Congress of the International Commission for Optics, ICO-13, pp. 6-9, (August 1984).

4. ORAL PRESENTATIONS

This section lists oral presentations at meetings and conferences describing research supported by this grant during the period April 15, 1981 to June 30, 1984.

1. 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," presented at 1981 Annual Meeting, Optical Society of America, Orlando, Florida, October 1981, Journal of the Optical Society of America, vol. 71, p. 1630, (December 1981).

2. 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), (Invited Paper).

3. A.R. Tanguay, Jr., "Polarization Properties of Birefringent Phase Gratings," 1982 Gordon Research Conference on Holography and Optical Information Processing, Plymouth, New Hampshire, (1982), (Invited Paper).

4. B.K. Jenkins, A.A. Sawchuk, T.C. Strand, and B.H. Soffer, "Sequential Optical Logic Implementation," presented at 1982 Annual Meeting, Optical Society of America, Tucson, Arizona, October 1982; Journal of the Optical Society of America, vol. 72, p. 1721, (1982).

5. A.R. Tanguay, Jr., "Polarization Properties of Birefringent Phase Gratings," presented at 1982 Annual Meeting, Optical Society of America, Tucson, Arizona, October 1982; Journal of the Optical Society of America, Vol. 72, p. 1832, (1982).

6. B.K. Jenkins and T.C. Strand, "Computer-Generated Holograms for Space-Variant Interconnections in Optical Logic Systems," Annual Meeting of the SPIE, San Diego, California (August 1983).

7. B.K. Jenkins, A.A. Sawchuk, T.C. Strand, "Spatial Light Modulator Requirements for Sequential Optical Logic," presented at 1983 Annual Meeting, Optical Society of America, New Orleans, October 1983; Journal of the Optical Society of America, Vol. 73, (1983).

8. A.A. Sawchuk and T.C. Strand, "Digital Optical Computing Techniques", presented at 1983 Annual Meeting, Optical Society of America, New Orleans, October 1983; Journal of the Optical Society of America, Vol. 73, (1983).

9. A.A. Sawchuk, "Digital Logic and Computing with Optics," presented at SPIE Technical Symposium, Los Angeles (January 1984).

10. A.A. Sawchuk, "Numerical Optical Computing Techniques," presented at 13th Congress of the International Commission for Optics, ICO-13, Sapporo, Japan, (August 1984).