DTIC FILE COPY   AFOSR·TR· 88-0402

TOPICAL MEETING ON
OPTICAL COMPUTING

Approved for public release

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFSC)
NOTICE OF
This technical report has been reviewed and is
approved for public release IAW AFR 190-12.
Distribution is unlimited.
MATTHEW J. KEPPER
Chief, Technical Information Division

DTIC
S ELECTED
MAY 0 2 1988
H
D

**TECHNICAL
DIGEST SERIES
VOLUME 11**

MARCH 16-18. 1987
INCLINE VILLAGE. NEVADA

88   5   02   285

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION Unclassified		1b. RESTRICTIVE MARKINGS n/a			
2a. SECURITY CLASSIFICATION AUTHORITY n/a		3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release- distribution unlimited			
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE n/a					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) AFOSR-87-0209		5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR.TR. 88-0402			
6a. NAME OF PERFORMING ORGANIZATION Optical Society of America	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Department of the Air Force- Office of Scientific Research			
6c. ADDRESS (City, State, and ZIP Code) 1816 Jefferson Place, N.W. Washington, D.C. 20036		7b. ADDRESS (City, State, and ZIP Code) Bolling Air Force Base Washington, D.C. 20332-6448			
8a. NAME OF FUNDING / SPONSORING ORGANIZATION Same as 7a	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-87-0209			
8c. ADDRESS (City, State, and ZIP Code) Same as 7b		10. SOURCE OF FUNDING NUMBERS			
		PROGRAM ELEMENT NO. 61102F	PROJECT NO. 2305	TASK NO. B4	WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Organization of a Topical Meeting on Optical Computing

12. PERSONAL AUTHOR(S)
Jarus W. Quinn

13a. TYPE OF REPORT Final	13b. TIME COVERED FROM 03/15/87 TO 03/31/88	14. DATE OF REPORT (Year, Month, Day) 88/03/31	15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17.	COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

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

A combination of materials research, device design and fabrication, algorithm/
architecture development, and applications systems study forms the body of optical
computing research. This topical meeting is designed to bring together researchers
from these widely different disciplines. The main aim of the meeting was not only to
report on the current state of the art in optical computing, but to look into possible
future directions that are likely to emerge from an interaction between new devices,
novel architectures, and nontraditional applications for optical computing.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Jarus W. Quinn, Executive Director Giles	22b. TELEPHONE (Include Area Code) 202-223-8130 767-4933	22c. OFFICE SYMBOL NE

DD Form 1473, JUN 86   *Previous editions are obsolete.*   SECURITY CLASSIFICATION OF THIS PAGE

TOPICAL MEETING ON

OPTICAL COMPUTING

**Summaries of papers presented at the
Optical Computing Topical Meeting**

March 16–18, 1987

Incline Village, Nevada

TABLE OF CONTENTS

LOWER LOBBY

6:00 PM-9:00 PM   REGISTRATION/REFRESHMENTS

PROSPECTOR/RUBICON ROOM

10:30 AM-12:00 M
MB   SESSION 2
Alan Huang, *AT&T Bell Laboratories, Presider*

MONDAY, MARCH 16, 1987

LOWER LOBBY

7:30 AM-5:30 PM   REGISTRATION/SPEAKER CHECKIN

PROSPECTOR/RUBICON ROOM

8:30 AM-10:00 AM
MA   SESSION 1
Adolf W. Lohmann, *Erlangen University, F. R. Germany, Presider*

8:30 AM   (Invited Paper)
MA1   Optical Computing—an Overview, Joseph W. Goodman, *Stanford U.* The field of optical computing finds itself at a convergence of two different technological streams. One stream flows from the field of analog optical information processing, while the other flows from the field of nonlinear optical devices. Both streams are attempting to reach the same destination, namely, a useful computer based on optical technology. (p. 2)

9:00 AM   (Invited Paper)
MA2   Systolic Array Machines can be both Fast and Programmable, H. T. Kung, *Carnegi Mellon University.* This talk describes some latest developments in the area of high-performance, programmable systolic arrays. (p. 3)

9:30 AM   (Invited Paper)
MA3   Programming on Optical Computer, Y. Abu-Moustafa, *California Institute of Technology. (p. 4)*

SIERRA ROOM

10:00 AM-10:30 AM   COFFEE BREAK

10:30 AM
MB1   Programmable Optical Processor Based on Symbolic Substitution, Karl-Heinz Brenner, G. Stucke, *U. Erlangen-Nuremberg, F. R. Germany.* A new architecture for a programmable optical processor is proposed. It is based on a few simple substitution rules and offers general computing capability. (p. 6)

10:45 AM
MB2   Digital Design Technique for Optical Computing, M. J. Murdocca, N. Streibl, *AT&T Bell Laboratories.* A digital optical computer design technique is presented for cascadable optically nonlinear arrays. The technique is efficient despite a regular free-space interconnection topology. (p. 9)

11:00 AM
MB3   Optical Systems for Symbolic Substitution, Joseph N. Mait, Karl-Heinz Brenner, *U. Erlangen-Nuremberg, F. R. Germany.* Optical systems for both recognition and substitution in a symbolic substitution system are presented. The systems use only classical optical and phase-only holographic elements. Methods for designing the holograms are discussed. (p. 12)

11:15 AM
MB4   Strengths and Weaknesses of Optical Architectures Based on Symbolic Substitution, Thomas J. Cloonan, *AT&T Bell Laboratories.* Four architectures that implement symbolic substitution are presented. The strengths and weaknesses of the different architectures are compared for a typical application (binary addition). (p. 16)

11:30 AM
MB5   Binary Image Algebra and Digital Optical Cellular Image Processors, K. S. Huang, B. K. Jenkins, A. A. Sawchuk, *U. Southern California.* We summarize binary image algebra for image processing and its implementation on digital optical cellular image processors (DOCIPs). Two promising architectures, DOCIP-array and DOCIP-hypercube, are discussed. (p. 20)

11:45 AM
MB6   Bit Serial Optical Computer, Harry F. Jordan, *U. Colorado at Boulder.* Current technology allows the immediate construction of a completely optical, stored program computer. The key is to use the techniques of bit serial processing and pipelining. (p. 24)

12:00 M-1:00 PM   LUNCH BREAK

PROSPECTOR/RUBICON ROOM

1:00 PM-2:30 PM
MC SESSION 3
Joseph W. Goodman, *Stanford University, Presider*

1:00 PM
MC1 Four-Dimensional Optical Crossbar, Adolf W. Lohmann, Wilhelm Stork, *U. Erlangen-Nuremberg, F. R. Germany.* A 4-D crossbar supports simultaneous dialogues among channels belonging to a 2-D array. Polarization optics is well suited to implement this concept. (p. 30)

1:15 PM
MC2 Cellular Optical Processor Architecture with Modulable Holographic Interconnections, J. Taboury, J. M. Wang, P. Chavel, F. Devos, *U. Paris-Sud, France.* We describe an interconnection architecture comprising one image plane hologram and one Fourier hologram whereby the connection scheme can be modified during the algorithm. (p. 31)

1:30 PM
MC3 Parallel Interfacing of Integrated Optics with Free Space Optics, Adolf W. Lohmann, *U. Erlangen-Nuremberg, F. R. Germany.* Free space optics is good for transporting signals in parallel. Integrated optics is good for switching. We have designed parallel interfaces to combine these two technologies. (p. 35)

1:45 PM
MC4 Scattering from Small Structures for Optical Beam Shaping and Interconnects, M. T. Lightbody, M. A. Fiddy, *King's College London, U.K.* We examine the importance of scattering from 3-D wavelength and subwavelength structures for optimal optical beam shaping and switching between fixed arrays. (p. 36)

2:00 PM
MC5 Engineering Limits to Optical Interconnects, Davis H. Hartman, *Bell Communications Research, Inc.* Optical interconnects offer a means to overcome electronics interconnect problems common in high speed computers. Fundamental engineering limits to optical interconnects are identified and discussed. (p. 40)

2:15 PM
MC6 Comparison of Encoding Schemes for E-Beam Fabrication of Computer Generated Holograms, H. Farhoosh, Michael R. Feldman, Sing H. Lee, Clark C. Guest, Y. Fainman, *UC-San Diego.* A set of criteria is established according to which various encoding methods of computer generated holograms are systematically evaluated for electron beam recording. These criteria are based on the computing resource limitations and the desired wavefront properties. (p. 44)

SIERRA ROOM

2:30 PM-3:00 PM   COFFEE BREAK

PROSPECTOR/RUBICON ROOM

3:00 PM-5:30 PM
MD SESSION 4
Ravindra A. Athale, *BDM Corporation, Presider*

3:00 PM   (Invited Paper)
MD1 Optical Computer Architecture: What is the Ideal? W. Daniel Hillis, *Thinking Machines Corporation.* We define the ideal computer as one which can execute any calculation as fast as any other computer, within a multiplicative constant. (p. 50)

3:30 PM
MD2 Globally Folding Combinatorial Logic Cells in Digital Optical Systolic Computing Arrays, P. S. Guilfoyle, W. J. Wiley, *OptiComp Corporation.* Higher order computation often requires substantial combinatorial interaction. This places a severe load on the input structure of an optical computer. By folding the data in time and space this load is considerably reduced. This paper applies combinatorial folding to $n \times n$ bit digital optical linear systolic multiplication arrays for matrix linear algebra. (p. 54)

3:45 PM
MD3 Residue Position-Coded Look-Up Table Processing, A. P. Goutzoulis, D. K. Davies, *Westinghouse R&D Center;* E. C. Malarkey, J. C. Bradley, P. R. Beaudet, *Westinghouse Advanced Technology Division.* Residue position-coded look-up table processing is discussed. The types, complexity, and performance of look-up tables are considered along with initial experimental results. (p. 58)

4:00 PM
MD4 Optical Arithmetic/Logic Unit Based on Residue Number Theory and Symbolic Substitution, C. David Capps, R. Aaron Falk, Theodore L. Houk, *Boeing Aerospace Company.* The concept for a GHz-rate, digital adder, multiplier, or logic unit that requires no spatial light modulators or optically nonlinear materials is presented. (p. 62)

4:15 PM
MD5 Digital Optical Matrix-Vector Multiplier using a Holographic Look-Up Table and Residue Arithmetic, S. F. Habiby, Stuart A. Collins, Jr., *Ohio State U.* The demonstration of a digital optical matrix-vector multiplier is reported. It uses position coding, a residue arithmetic representation, a holographic memory, and a look-up table approach, reducing effective computation time to one Hughes liquid crystal light valve response time. (p. 66)

4:30 PM
MD6 Limitatins to Optical Fredkin Circuits, Robert Cuykendall, Debra McMillin, *U. Iowa.* Severe computing limitations exist for recently proposed optical Fredkin gates. Sequential addition and shuffle cascades computing arbitrary switching functions are possible, but not sequential multiplication. (p. 70)

4:45 PM
MD7 Matrix-Vector Multiplication Using Polarization Rotators, L. Scharf, W. Thomas Cathey, K. M. Johnson, *U. Colorado at Boulder.* A new approach to optical matrix-vector multiplication is described which matches signal processing algorithms and architectures to optical primitives that directly perform rotation operations. (p. 73)

5:00 PM
MD8 Monte Carlo Matrix Inversion Using an Optical Random Number Generator, Anthony J. Martino, G. Michael Morris, *U. Rochester.* Monte Carlo matrix inversion is performed using an optical random number generator and an electronic computer. Experimental results, including the speed-accuracy tradeoff, are presented. (p. 77)

5:15 PM
MD9 Monte Carlo Processor Arrays Using Optical Random Number Generators, F. Devos, K. Madani, P. Chavel, J. Taboury, *U. Paris-Sud, France.* To allow Monte Carlo algorithm implementation in one-chip electronic massively parallel processor arrays, 2-D arrays of random numbers are generated optically using single photoevent amplification of speckle. (p. 81)

5:30 PM BREAK

LAKESIDE ROOM

6:00 PM-8:00 PM CONFERENCE RECEPTION

SIERRA ROOM

8:00 PM-9:30 PM
ME POSTERS: SESSION 5

ME1 Free Space Optical Interconnects by Cascaded Holographic Elements, W. J. Hossack, *King's College London, U.K.* A cascaded holographic system for optical coordinate transformations is used in a free space optical interconnect. The design criteria for afocal systems performing conformal mappings is presented. (p. 86)

ME2 Optical Interconnect Complexity Limitations for Holograms Fabricated by Electron Beam Lithography, Michael R. Feldman, Clark C. Guest, *UC-San Diego.* The ability of optical communication systems, employing computer generated holograms, to perform complex interconnections for electronic integrated circuits is analyzed. (p. 90)

ME3 Design of Computer Generated Holograms for E-Beam Fabrication by a Computer Aided Design System, H. Farhoosh, Sing H. Lee, *UC-San Diego.* A procedure for designing computer generated holograms for electron beam fabrication using a computer aided design system is described, and design considerations are discussed. (p. 94)

ME4 Two-Dimensional Clos Optical Interconnection Network, Shing-Hong Lin, Thomas F. Krile, John F. Walkup, *Texas Tech U.* A 2-D Clos three-stage optical interconnection network is proposed. Applications include constructing large size 2-D crossbar interconnection networks. (p. 98)

ME5 Optical Interconnects Using Resonated Holograms, Stuart A. Collins, Jr., *Ohio State U.* We discuss optical interconnects formed by the use of thick holograms with resonant mirrors to achieve high efficiency and large information density. (p. 102)

ME6 Comparison of Optical and Electrical Interconnections Based on Power and Speed Considerations, Michael R. Feldman, Sadik C. Esener, Clark C. Guest, Sing H. Lee, *UC-San Diego.* Interconnect delay time limitations as a function of power dissipation are analyzed for both electronic integrated circuit transmission lines and optical communication paths. (p. 105)

ME7 Optical Implementation of Minimum and Maximum Operation, Hedong Yang, Clark C. Guest, *UC-San Diego.* An optical approach is proposed to implement the direct bitwise maximum and minimum operation on two data-pages. (p. 109)

ME8 Optical MSD Adder Using Polarization Coded Symbolic Substitution, P. A. Ramamoorthy, S. Antony, *U. Cincinnati.* The design of a parallel optical adder based on modified signed-digit number representation using symbolic substitution and polarization coding is shown. (p. 111)

ME9 Digital Optical Processor Based on Symbolic Substitution Using Matched Filtering, Ho-In Jeon, *U. Southern California.* A parallel digital optical processor that utilizes matched filtering and performs symbolic substitution is proposed. Its use for the example of binary addition is described. (p. 115)

ME10 Optical Parallel Image Processing Using CCD Image Sensors, J. Tokumitsu, H. Matsuoka, K. Iijima, *Canon Research Center, Japan.* An optical system consisting of CCDs and an image-shifting mechanism has been built. It performs the convolution operation on an input image at video rate. (p. 119)

LOWER LOBBY

7:30 AM-5:30 PM   REGISTRATION/SPEAKER CHECKIN

PROSPECTOR/RUBICON ROOM

8:00 AM-10:00 AM
TuA   SESSION 6
H. John Caulfield, *University of Alabama in Huntsville,*
Presider

8:00 AM   (Invited Paper)
TuA1   Advances in Brain-Style Computation, David E.
Rumelhart, *University of California, San Diego.* A sketch of
current work on brain-style computation is provided.
Emphasis is on applications for building content-addressable memories and learning machines. (p. 124)

8:30 AM
TuA2   Architectures for Optoelectronic Analogs of Self-Organizing Neural Networks, Nabil H. Farhat, *U. Pennsylvania.* Architectures for partitioning optoelectronic analogs
of neural nets into input/output and internal units to enable
self organization and learning, where a net can form its
own internal representations of the environment, are
described. (p. 125)

8:45 AM
**TuA3   Optical Neural Nets Implemented with Volume
Holograms,** Demetri Psaltis, Jeffrey Yu, Xiang Guang Gu,
California Institute of Technology; Hyuk Lee, *Polytechnic
Institute of New York.* We examine the advantages of using
volume holograms as opposed to planar media for storing
an interconnect pattern in a neural network. We present
methods for achieving different types of arbitrary global interconnections and we present experimental results using
a photorefractive crystal (LiNbO₃) as the volume element.
(p. 129)

9:00 AM
TuA4   Multilayer Optical Learning Networks, Kelvin
Wagner, Demetri Psaltis, *California Institute of Technology.* We present a trainable, self aligning, multilayer perceptron pattern transformation processor that uses backwards error propogation to modify volume holographic interconnections between nonlinear Fabry-Perot etalons.
(p. 133)

9:15 AM
TuA5   Optical Associative Processing Elements with Versatile Adaptive Learning Capabilities, Arthur D. Fisher,
John N. Lee, *U.S. Naval Research Laboratory.* Optical
associative-processing architectures are presented for implementing four types of versatile adaptive learning dynamics which are applicable to parallel symbolic-processing problems. Both electrooptic and holographic configurations are presented. (p. 137)

9:30 AM
TuA6   Optical Symbolic Computing: Architectural Considerations, M. W. Derstine, P. R. Haugeo, A. Husain,
Honeywell Physical Science Center; A. Guha, R. Ramnarayan, *Honeywell Corporate Systems Development Division;* A. Vaid, *U. Southern California.* Examination of computational models for current symbolic processing languages reveals that manipulation of data structures is a
critical function. Optical approaches to these operations
are discussed. (p. 141)

9:45 AM
**TuA7   Comparison on Adaptive Pattern Recognition and
Image Restoration with Hetero-associative and Auto-associative Memories,** Jack Y. Jau, Y. Fainman, Sing H.
Lee, *UC-San Diego.* We discuss the close relationships between adaptive pattern recognition and hetero-associative
memory, and between iterative image restoration and auto-associative memory based on the algorithm they use. We
present a hybrid architecture for the implementation of
adaptive pattern recognition and iterative image restoration and compare it with those in the existing literature.
(p. 145)

SIERRA ROOM

10:00 AM-10:30 AM   COFFEE BREAK

PROSPECTOR/RUBICON ROOM

10:30 AM-12:00 M
TuB   SESSION 7
John N. Lee, *U.S. Naval Research Laboratory, Presider*

10:30 AM   (Invited Paper)
TuB1   Analog Complexity Theory, Kenneth Steiglitz,
Princeton U. Analyzing computational complexity is more
difficult in the analog than in the digital case because of
the modeling problem, and theory and technical are at an
earlier stage of development. We discuss this theory, and
give some examples of its application. (p. 150)

11:00 AM
**TuB2   Unified Approach to Analyzing Optical Computing
Systems,** Ravindra A. Athale, Charles W. Stirk, Michael W.
Haney, *BDM Corporation.* Optical computing systems can
be analyzed in terms of their algorithms, architectures or
hardware. A formal method is presented that interrelates
these three aspects to provide a uniform basis for comparing different approaches and to suggest new directions for
research. (p. 151)

viii

11:15 AM
TuB3 Rule-Based Probabilistic Symbolic Target Classification by Object Segmentation, David Casasent, Abhijit Mahalanobis, *Carnegie Mellon U.* Correlation outputs of segmented object regions are considered as symbolic object descriptions. These are processed by rule-based probabilistic techniques. Initial experimental results are included. (p. 155)

11:30 AM
TuB4 Real-Time Acoustooptic Spotlight Mode SAR Processor, Michael W. Haney, *BDM Corporation;* Demetri Psaltis, *California Institute of Technology.* Novel hybrid optical/electronic techniques are employed in the application of the programmable acoustooptic SAR architecture to spotlight mode geometries. (p. 159)

11:45 AM
TuB5 Processing of Synthetic Aperture Radar Data Using the PRIMO Optical Outer-Product Processor, Y. Owechko, J. Grinberg, E. Marom, B. H. Soffer, *Hughes Research Laboratories.* An optical real-time processor of SAR data is described. The processor is based on outer-product decomposition and is compact, rugged, lensless, and utilizes incoherent light. (p. 163)

12:00 M–1:00 PM   LUNCH BREAK

PROSPECTOR/RUBICON ROOM

1:00 PM–2:30 PM
TuC   SESSION 8
Satoshi Ishihara, *Electrotechnical Laboratory, Japan, Presider*

1:00 PM   (Invited Paper)
TuC1 Two-Dimensional Arrays of Semiconductor Optical Gates for Optical Computing, Hyatt M. Gibbs, *U. Arizona.* ZnS interference filters and GaAs etalons are attractive nonlinear optical logic devices and are used to perform simple symbolic-substitution experiments. (p. 168)

1:30 PM
TuC2 Restoring Optical Logic: Demonstration of Extensible All-Optical Digital Systems, S. D. Smith, F. A. P. Tooley, A. C. Walker, N. C. Craft, B. S. Wherrett, *Heriot-Watt U., U.K.* We show experimentally how all-optical circuit elements are used to create restoring logic. A loop circuit demonstrated simulates an optical classical finite state machine. (p. 172)

1:45 PM
TuC3 Highly Cascadable Optically Bistable Device for Large Fan-Out Optical Computing Applications, N. C. Craft, S. D. Smith, *Heriot-Watt U., U.K.* Bistable devices consisting of two thermally coupled nonlinear etalons are shown to exhibit large changes in output power when used in a hold-and-switch configuration. (p. 175)

2:00 PM
TuC4 Polarization-Based Optical Parallel Logic Gates Using Ferroelectric Liquid Crystal Spatial Light Modulators, K. M. Johnson, M. Handschy, W. Thomas Cathey, N. A. Clark, D. Walba, *U. Colorado at Boulder.* We report on progress made in engineering faster switching ferroelectric liquid crystal compounds and their use as optical parallel logic gates. (p. 179)

2:15 PM
TuC5 Optimum Control Beam Angle for a Biased Fabry-Perot Bistable Device, R. Cush, I. Bennion, *Plessey Research Caswell, Ltd., U.K.* The switch-on power of the control beam in a multibeam, cascadable Fabry-Perot bistable device may be minimized by varying the beam incident angle. (p. 183)

SIERRA ROOM

2:30 PM–3:00 PM   COFFEE BREAK

PROSPECTOR/RUBICON ROOM

3:00 PM–5:30 PM
TuD   SESSION 9
Carl M. Verber, *Georgia Institute of Technology, Presider*

3:00 PM   (Invited Paper)
TuD1 Materials and Devices for Optical Computing, Armand R. Tanguay, Jr., *U. Southern California.* (p. 188)

3:30 PM
TuD2 Variable-Gamma Spatial Light Modulator, Suzanne Lau, Cardinal Warde, *MIT Department of Electrical Engineering & Computer Science.* We describe how standard and Fabry-Perot MSLMs can be operated to generate a variety of real-time variable-gamma characteristics and present preliminary image processing results. (p. 189)

3:45 PM
TuD3 Integrated Electrooptic Bragg Modulator Modules for Optical Computing, D. Y. Zang, P. Le, C. S. Tsai, *UC-Irvine.* A variety of titanium-indiffused proton-exchanged microlens-based integrated electrooptic Bragg modulator modules has been constructed in a LiNbO$_3$ substrate size $0.2 \times 1.0 \times 1.8$ cm^3. These modules have been used to perform matrix–vector and matrix–matrix multiplications. (p. 193)

4:00 PM
TuD4 Toward an Optical/Electronic Hybrid Image Processor, M. G. Nicholson, G. G. Gibbons, S. Mayo, C. R. Petts, *GEC Research, Ltd., U.K.;* B. Loiseaux, J. P. Huignard, *Thomson-CSF, France;* F. Dubois, J. Ebbeni, *Free U. Brussels, Belgium.* This paper describes recent work on the optoelectronic components required to build a practical hybrid processor based on dynamic holography in a photorefractive material. (p. 197)

4:15 PM
TuD5 Index Grating Lifetime in Photorefractive Semi-Insulating Cr-Doped GaAs, Li-Jen Cheng, Afshin Partovi, *Jet Propulsion Laboratory;* Elsa Garmire, *U. Southern California.* Results demonstrate that information storage time in volume holographic gratings in GaAs:Cr under illumination of a weak reading beam can be as long as 2.5 s. (p. 201)

4:30 PM
TuD6 Optical Crossbar Arithmetic/Logic Unit, R. Aaron Falk, C. David Capps, Theodore L. Houk, *Boeing Aerospace Company.* Optical crossing interconnects coupled with nonlinear thresholds at the intersections produces a multilevel logic device. A residue adder/multiplier using this concept has been demonstrated. (p. 205)

4:45 PM
TuD7 Passive Single-Mode Optical Networks for Coherent Processing, M. E. Marhic, *Northwestern U.* Passive architectures, using interference in single-mode networks, can perform discrete spatial Fourier or Hadamard transform. Phase stability and single polarization are required. (p. 209)

5:00 PM (Invited Paper)
TuD8 Review of Some Current Optical Computing Research in the Soviet Union, William T. Rhodes, *Georgia Institute of Technology.* The author, along with three other scientists from North America, attended a July 1986 meeting on optical computing held in Novosibirsk in the USSR. A number of current Soviet research efforts on optical computing and optical signal processing were reported at that meeting. These efforts, which include nonlinear optical switching, pipeline optoelectronic processing, pattern recognition, and optical computing generally, are reviewed and summarized. (p. 212)

5:30 PM BREAK

SIERRA ROOM

8:00 PM-9:30 PM
TuE POSTERS: SESSION 10

TuE1 Switch Power Drift in Optically Bistable ZnSe Interference Devices, R. J. Campbell, J. G. H. Mathew, S. D. Smith, A. C. Walker, *Heriot-Watt U., U.K.* Investigations of switch power drift in optically bistable thin-film interference devices are presented. Ways of minimizing this drift through device design are discussed. (p. 214)

TuE2 Thermooptical Beam Guide and Switching Experiments, Lauren M. Peterson, *Environmental Research Institute of Michigan.* Laser radiation (0.01 mJ) focused into an absorbing liquid generates a graded-index waveguide which switches (in 10 ns) the direction of a second laser beam. (p. 217)

TuE3 Ferroelectric Liquid Crystal Spatial Light Modulators, D. Armitage, J. I. Thackara, *Lockheed Missiles & Space Company, Inc.;* N. A. Clark, M. A. Handschy, *Displaytech.* A photoaddressed ferroelectric liquid crystal (FLC) spatial light modulator is described with experimental results. Current developments in FLC technology applied to optical processing are discussed. (p. 221)

TuE4 Theory of All-Optical GaAs Logic Devices, M. E. Warren, S. W. Koch, Hyatt M. Gibbs, *U. Arizona.* All-optical semiconductor devices are numerically modeled and optimized using a microscopic theory for the optical nonlinearities of room-temperature GaAs. Single-frequency NOR-gate operation in reflection is predicted. (p. 225)

TuE5 Optical NOR Gate Using Diode Laser Sources, Masahiro Ojima, *Hitachi, Japan;* Arturo Chavez-Pirson, Yong H. Lee, Jean F. Morhange, Hyatt M. Gibbs, Nasser Peyghambarian, *U. Arizona;* Feng-Yu Juang, Pallab K. Bhattacharya, Doreen A. Weinberger, *U. Michigan.* An optical NOR gate has been successfully demonstrated using two diode lasers and a GaAs/AlGaAs multiple quantum-well etalon. (p. 229)

TuE6 Multiple Polarization State Threshold Logic and Processor, Shudong Wu, Xiang Zhang, Zhijiang Wang, *Shanghai Institute of Optics & Fine Mechanics, China.* Based on using multiple polarization states, a novel technique for implementing different logic operations in parallel is described. Full optical A/D converter, look-ahead adder, and multiplier are proposed. (p. 233)

TuE7 Interferometric Pattern Encoding for Parallel Logic Operation, Makoto Ikeda, Toyohiko Yatagai, *U. Tsukuba, Japan;* Satoshi Ishihara, Yoshinobu Mitsuhashi, *Tsukuba Electrotechnical Laboratory, Japan;* Junichi Kaya, *Nippon Institute of Technology, Japan.* An interferometric technique of spatial encoding for optical parallel pattern logic operations is proposed and its use in space-variant logic-gate arrays is discussed. (p. 237)

TuE8 Infrared Predetection Dynamic Range Compression via Photorefractive Crystals, Hua-Kuang Liu, Li-Jen Cheng, *Jet Propulsion Laboratory.* The predetection infrared dynamic range compression concept via the nonlinear photorefractive two-wave mixing in GaAs crystals is discussed. Some experimental results are presented to support this idea. (p. 241)

TuE9   Fingerprint Enhancement by Fourier Domain Optical Processing, D. M. Monro, B. G. Sherlock, *Imperial College, U.K.:* C. R. Petts, *GEC Research, Ltd., U.K.* Fourier domain directional filtering of fingerprint images controlled by local ridge orientation gives effective enhancement. Implementation in optical hardware provides advantages over digital computer processing. (p. 245)

WEDNESDAY, MARCH 18, 1987

PROSPECTOR/RUBICON ROOM

1:30 PM–5:20 PM
WB   Joint Photonic Switching and Optical Computing Plenary Session, T. Kenneth Gustafson, *National Science Foundation, Presider*

1:30 PM   (Plenary Paper)
WB1   Photonic Switching Components: Current Status and Future Possibilities, John E. Midwinter, *University College London, U.K.* The range of components becoming available for routing signals in optical networks is vast and varied. We review their character and typical performance and point to the network characteristics they support. (p. 8)

2:20 PM   (Plenary Paper)
WB2   Optical Digital Computers, Alan Huang, *AT&T Bell Laboratories.* (p. 9)

MONDAY, MARCH 16, 1987

PROSPECTOR/RUBICON ROOM

8:30 AM–10:00 AM

MA1–3

SESSION 1

Adolf W. Lohmann, Erlangen University, F. R. Germany, *Presider*

1

SUMMARY

MA1-1

OPTICAL COMPUTING -- AN OVERVIEW

J.W. Goodman
Stanford University

The field of optical computing finds itself at a convergence of two different technological streams. One stream flows from the field of analog optical information processing, while the other flows from the field of nonlinear optical devices. Both streams are attempting to reach the same destination, namely a useful computer based on optical technology.

Analog optical information processing systems have reached a high state of development in certain limited and specialized applications, such as image formation from synthetic aperture radar data, and Bragg cell spectrum analysis. Much effort has been spent and to some extent continues to be spent on attempts to extract high numerical accuracy from analog systems, using various forms of number representation, in hopes of making a fast optical arithmetic unit. To date these efforts have not proved successful. The problem is not that these schemes fail to work, but rather that they fail to be competitive, in cost and/or performance, with electronic approaches to the same problem.

What then will be the role of optics in numerical computing of the future? Our hypothesis is that both optical interconnections and arrays of nonlinear optical elements will prove to have an important role to play. Optical interconnections will gradually filter down the heirarchy of interconnects in electronic computers, from inter-machine, to backplanes, to inter-chip communications. The likelihood of optics playing a significant role at the intrachip role is small. Arrays of nonlinear optical elements will ultimately be important in switching, multiplexing, and demultiplexing optical streams of data. However, the probability that a performance-competitive all-optical computer will emerge in this century is not regarded as very high..

Systolic array machines can be both fast and programmable.

H.T. Kung
Department of Computer Science
Carnegie Mellon University

Warp is a programmable systolic array machine developed by Carnegie Mellon. Currently two 10-cell machines are operational at Carnegie Mellon, with each cell being a 10 MFLOPS programmable processor. These machines have been used in a diverse range of applications, including navigation for robot vehicles, signal processing, and medical image processing, and as a tool for vision research. For these applications, Warp is typically several hundred times faster than the VAX 11/780. General Electric, which is Carnegie Mellon's industrial partner for the Warp project, is building at least eight additional Warp machines.

Warp has become a useful machine not only because of its high-performance but also because of its high degree of programmability. The simplicity and regularity of the systolic array architecture, which helped Warp achieve high-performance, have also helped the successful development of an optimized compiler capable of generating efficient code for the machine. With this compiler, programming the machine for a variety of applications becomes practical.

Carnegie Mellon has started working with Intel on the design of a custom VLSI Warp implementation, called iWarp. The iWarp chip is a high-performance floating-point microprocessor, using on the order of 600K transistors. With the iWarp chip, Warp machines having hundreds or even thousands of programmable cells configured in one- or two-dimensional arrays are possible.

This talk will describe these latest developments in the area of high-performance, programmable systolic arrays.

Programming an Optical Computer

Y. Abu-Moustafa
California Institute of Technology

MONDAY, MARCH 16, 1987

PROSPECTOR/RUBICON ROOM

10:30 AM–12:00 M

MB1–6

SESSION 2

Alan Huang, AT&T Bell Laboratories, *Presider*

A PROGRAMMABLE OPTICAL PROCESSOR BASED ON SYMBOLIC SUBSTITUTION

K.-H. Brenner, G. Stucke

Physikalisches Institut der Universität
Erlangen-Nürnberg, West-Germany

Abstract

A new architecture for a programmable optical processor is proposed. It is based on a few simple substitution rules and offers general computing capability.

1.   Symbolic substitution

The concept of symbolic substitution was recently introduced by Huang and Brenner /1,2/. It is a powerful method to perform optical logic in a two step process. On a rectangular array of light sources, the first step is to recognize all the locations of a certain spatial pattern within the array. In the case of polarization logic the single pixels of the array can be distinguished by orthogonal polarization states /3/. The second step is to substitute a new pattern wherever the search-pattern was recognized. This two step process is called a substitution rule. Special rules can be applied to perform logic, arithmetic and also communication. Thus a Turing machine can be realized, which means that symbolic substitution is able to solve any computable problem. In a practical system, efficiency with respect to speed and hardware is an important consideration.

2.   A new architecture

This paper introduces a new architecture (fig. 1) which offers both generality and simplicity. Only three types of modules are necessary:
    Shift modules
    Switch modules
    Logic module.
Each module consists of several recognition and substitution units to perform a special set of rules. The modules are arranged in a feedback loop. Normally an algorithm needs several loops through the processor to produce the result. Nevertheless input to and output from the processor can take place every cycle. Before a new input enters a module, control information is added to the data. This information represents the program that controls the processor.

3.   Data and control plane

The input plane for a module is divided into two parts. A column of data bits is followed by two columns of control bits (fig. 2). The meaning of the data bits is not restricted to a special case and can be adapted to the actual problem.

4. Shift modules

There are five different types of shift modules (fig. 3). Four of these modules perform programmable horizontal shifts. Every row has its own shift control. Therefore it is possible e.g. to shift the second row to the left and the third row to the right. The fifth module can shift the whole input plane vertically by one pixel.
To achieve faster global interconnections the horizontal shifts cover a range from -15 to +15 pixels.

5. Logic module

The logic module connects two vertically adjacent data bits and generates four result bits. If the data bits are called 'a' and 'b', the following logic operations are carried out:

    a AND b
    a OR b
    a XOR b
    NOT a

These operations offer enough generality for flexible computing. The advantages of this module are the small sized recognition rules and the capability to select the required logic operation. In addition to that, the logic operation can be changed by changing the substitution rule - not the hardware.

6. Switch module

This module selects, which result bit from the logic module output is used as the actual result. The result bit can assume the state of bit 'a' or 'b' of the logic module, depending on the state of the control information entering the module.

7. Summary

We have proposed a new architecture for a programmable digital optical processor. Shift modules, switch modules, and logic modules in a feedback loop, together with external control information, constitute a general purpose programmable digital optical processor. All the modules are based on symbolic substitution.

8. References

/1/   A. Huang: "Parallel Algorithms for Optical Digital Computers" in Technical Digest, IEEE Tenth International Optical Computing Conference, (1983), p. 13-17.

/2/   K.-H. Brenner, A. Huang, N. Streibl : "Digital optical computing with Symbolic Substitution", Appl. Optics 25, 3054 (1986).

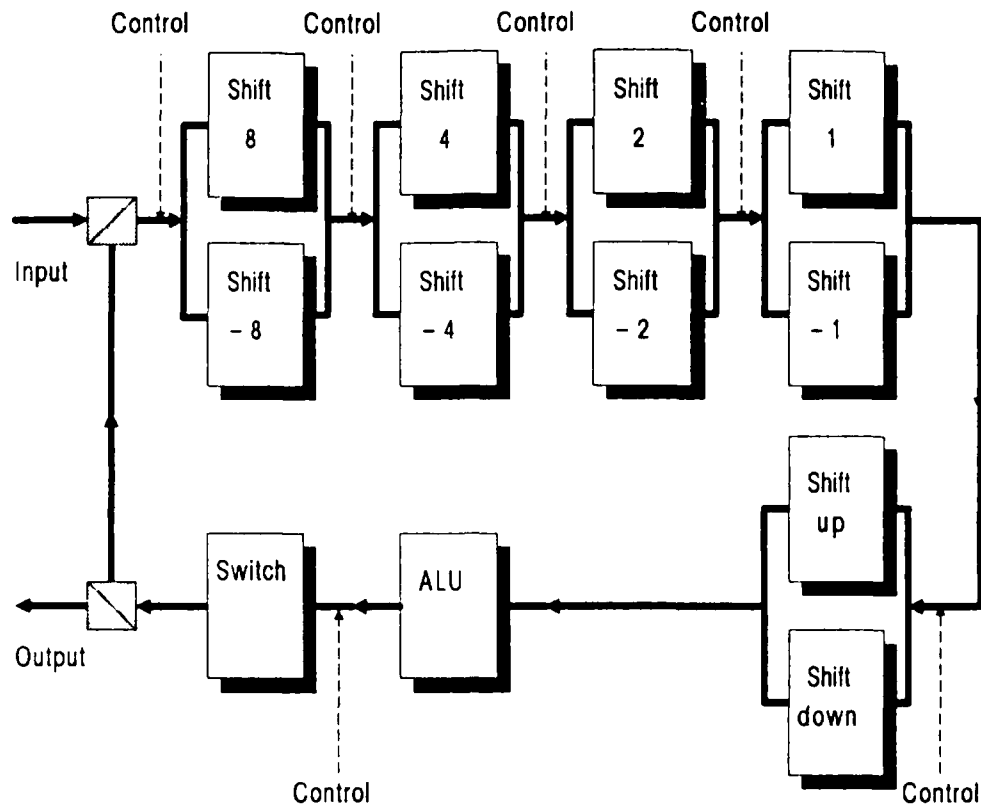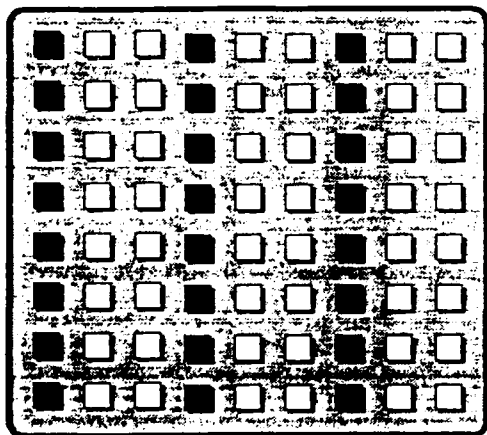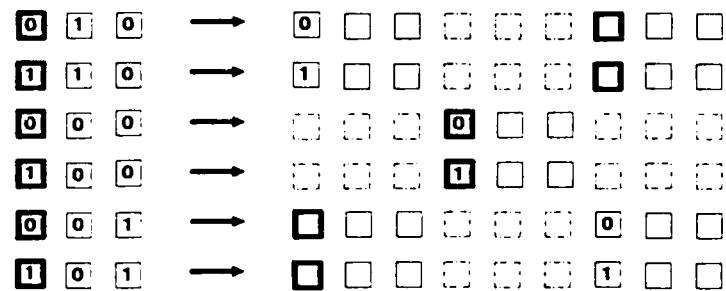/3/ K.-H. Brenner: "New implementation of symbolic substitution", Appl. Optics 25, 3060 (1986).

Fig.1 Programmable optical processor architecture

■ data pixel

☐ control pixel

Fig.2 Layout for the data plane.
Every data bit is associated
with two control bits.

for example: **SHIFT 2**

Fig.3 Substitution rules for the
shift module. Conditioned
by the control bits the
datum is shifted +2,0 or -2
positions.

8

A Digital Design Technique for Optical Computing

M. J. Murdocca
N. Streibl

AT&T Bell Laboratories
Holmdel, New Jersey 07733

1. INTRODUCTION

Optically nonlinear arrays have been studied experimentally in the last few years. Results encourage the development of computer architectures suitable for optics. We present a computer design technique that makes use of optically nonlinear arrays and free-space interconnects. In order to take advantage of the natural parallelism of this architecture without suffering from limitations due to the regularity of the interconnects, novel computing techniques are needed [1]. Pattern transformations and regular interconnects are the basic architectural building blocks of the system presented here. The focus is on computational aspects of the architecture.

2. THE ARCHITECTURE

We propose an architecture that consists of four pattern transformation rules and a regular interconnect (Figure 1).
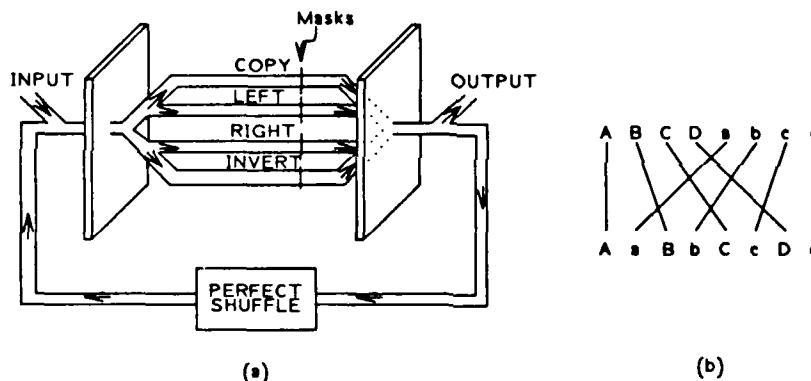
Figure 1. Schematic of a digital optical computer (a), and 8-bit 1-dimensional perfect shuffle.

A two-dimensional input pattern is split into four identical images. Each of these images is transformed by one of the operations: COPY, LEFT, RIGHT, or INVERT and is passed through a mask before being combined with the other images on the target plane. The target plane is fed through an optical perfect shuffle [2] and is imaged back onto the input plane. The machine communicates with the outside world through the input and output planes.

We present four transformation rules based on Huang's *symbolic substitution* [1] that provide sufficient design flexibility at a relatively small hardware cost:

$$\text{COPY} \qquad a_{x,y} \rightarrow a_{x,y}$$

$$\text{LEFT} \qquad a_{x,y} \rightarrow a_{x-1,y}$$

$$\text{RIGHT} \qquad a_{x,y} \rightarrow a_{x+1,y}$$

$$\text{INVERT} \qquad a_{x,y} \rightarrow \overline{a_{x,y}}$$

The boundary bits $a_{0,y}$, $a_{N+1,y}$, $a_{x,0}$ and $a_{x,N+1}$ are assumed to be zero. Bits that are imaged off of the array do not take part in the computation. Each operation can be locally enabled or disabled by making corresponding mask bits transparent or opaque. In the text that follows we show how to generate the masks to implement arbitrary logic functions.

9

3. AN EXAMPLE

Figure 2 shows a serial adder that we will use to illustrate the technique. There are two input lines where two binary numbers enter, least significant bit first. The result appears on the output line, least significant bit first. For two N-bit numbers N time steps are needed to complete the addition.
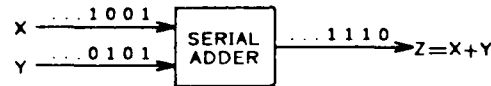

Figure 2. Serial adder.

The adder can be characterized by two logic functions. One function computes the current output, while the other function computes the carry. Boolean equations for the carry function c_{t+1} and the output function z_{t+1} are given by:

$$c_{t+1} = \overline{\overline{(x+y)} + \overline{(x+c_t)} + \overline{(y+c_t)}} \tag{1}$$

$$z_{t+1} = \overline{\overline{(x+y+c_t)} + (\overline{(x+\bar{y}+\bar{c_t})} + \overline{(\bar{x}+\bar{y}+c_t)} + \overline{(\bar{x}+y+\bar{c_t})})} \tag{2}$$

where a logical OR is denoted by $+$. Subscripts have been dropped from x and y for clarity.

Figure 3 illustrates a circuit corresponding to these equations that can be directly implemented on the computer shown in Figure 1. Masks are shown in Figure 4.
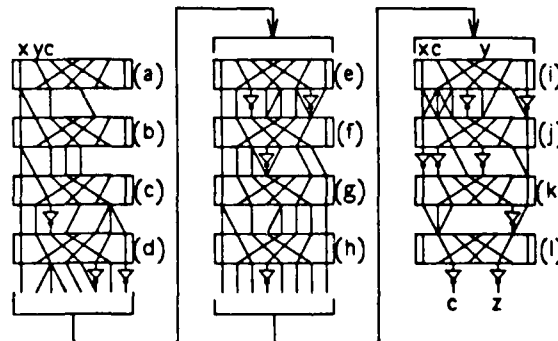

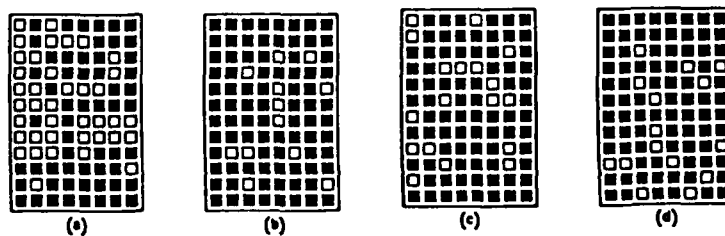**Figure 3.** Circuit layout of a serial adder.

Figure 4a-d. Masks for a serial adder. COPY mask (a), LEFT mask (b), RIGHT mask (c), INVERT mask (d). Note that the two transparent tiles in the bottom row of (d) correspond to output bits c and z in Figure 3.1.

The circuit in Figure 3 was generated by tracing the system backwards through the perfect shuffles and picking COPY, LEFT, RIGHT, and INVERT primitives as necessary to implement the functions. The reason for going backwards is that there is only one variable on the left hand side of the equations while there are many variables on the right hand sides. It is easier to

expand the left hand side until it looks like the right hand side rather than searching for an arrangement of the perfect shuffle interconnect that combines all the right hand side variables to produce the left hand side variable. We will implement both the next state function c_{i+1} and the output function z_{i+1} for a network that is 8 bits wide.

Starting with the next state function

$$c_{i+1} = \overline{\overline{(x+y)} + \overline{(x+c_i)} + \overline{(y+c_i)}}$$

we see that the outermost operation is the negation that covers the entire construct. The negation is shown in Figure 4.l as c at an arbitrarily chosen output point. The operation to be mapped now (with subscripts removed from c, for clarity) is:

$$\overline{(x+y)} + \overline{(x+c)} + \overline{(y+c)}$$

The outermost operation now is the three-input OR of the maxterms. If we trace Figure 4.l back through the perfect shuffle, we can place the three-input OR at that point using the RIGHT, COPY, and LEFT transformations as shown in Figures 4.k-4.l. The operations to be mapped now are:

$$\overline{(x+y)} \quad \overline{(x+c)} \quad \overline{(y+c)}$$

The negation of each of the maxterms can be realized by the INVERT transformation as shown in Figures 4.j-4.k.

$$(x+y) \quad (x+c) \quad (y+c)$$

Each of the 2-input OR operations can be implemented as shown in Figures 4.i-4.j. The terms x, y, and c have been placed at the inputs to the maxterms as shown in Figure 4.i, and are carried back through the system as shown in Figures 4a-4i. The additional levels sort the variables into position for cascading and pad out the circuit for z_{i+1} which needs more than 4 levels.

Next we consider the output variable z_{i+1} whose Boolean equation is mapped onto the network analogously. We start with the outermost inversion at an arbitrarily chosen output point (Figure 4.l). Since we do not have 4-input OR joins available in our model, we must break the function up and implement it a few maxterms at a time as shown in the parenthesized grouping in Equation (2) and in Figures 4.i-4.l. The rest of the equation is implemented in the remaining levels shown in Figures 4.a-4.i. Note that the carry c at the bottom of Figure 4.l directly lines up with c in Figure 4.a for easy cascading.

4. COMMENTS

The serial adder shown here is 8 bits wide and 12 pixels deep. The depth of the circuit can be reduced by increasing the width of the circuit or increasing fan-in and fan-out. All path delays through the system can be made equal to within a few femtoseconds, so the whole system can be pipelined at the gate level. This means that the throughput in the optical implementation of this architecture can be greater than an electronic implementation, which would typically have 4 or 5 gate delays per clock step.

The perfect shuffle (or a similar global interconnect, such as the banyan or the hypercube) is not necessary to create a logically correct circuit, but it is necessary to create a shallow circuit. The perfect shuffle provides the ability to permute space into an otherwise topologically dense area. Through the use of the perfect shuffle, there is little need for random interconnections between logic gates.

[1] A. Huang, "Parallel Algorithms for Optical Digital Computers," *IEEE 1983 10th International Optical Computing Conference*, 13, (1983).

[2] A. W. Lohmann, "Optical Perfect Shuffle", *Applied Optics*, 25, No. 10, 1530, (May 15 1986).

Optical Systems for Symbolic Substitution

Joseph N. Mait* and Karl-Heinz Brenner
Physikalisches Institut der Universität Erlangen-Nürnberg
Erwin-Rommel-Str. 1
D 8520 Erlangen
Federal Republic of Germany

I. Introduction

Symbolic substitution has been proposed as one means of performing digital computations optically [1,2]. In such a system, information is distributed throughout an assemblage of spatial patterns and is then processed via the transformation of these patterns. For example, the logical AND operator is represented by four transformations: 0 for 00, 0 for 01, 0 for 10, and 1 for 11. In addition, when the data are seen as patterns and the operator as a pattern transformation, spatial coding of patterns follows naturally.

An optical system for implementing symbolic substitution must consist of a pattern recognition system, an optical NOR gate, and a pattern substitution system [1]. Using only classical optical and phase-only holographical elements, optical systems for both the recognition and substitution components have been designed. Phase-only elements were the only holographical elements considered so as to insure maximum light throughput from input plane to output plane.

II. Review of Symbolic Substitution

The necessary steps for implementing a given pattern transformation are recognition of the search pattern in the input (e.g., 00) followed by substitution of the scribing pattern (0 for 00). Since each transformation rule requires its own recognition/substitution system, the total number of such systems is equal to the total number of transformation rules. It is therefore necessary that the input be replicated this same number of times and, similarly, it is necessary to combine the individual outputs from each system to realize the final output.

To recognize a particular search pattern the input must be further replicated according to the number of logical zeroes present in the pattern [1]. Each replica is then shifted and overlayed such that if the search pattern is present in the input all the zeroes of the search pattern are aligned in one reference location. Performing a logical NOR operation on this location produces a logical one only if the pattern is present and a zero if it is not, which completes the recognition.

Replications and shifts can also be used to substitute the scribing pattern once the search pattern has been found. Since the scribing pattern is simply a pattern of ones and zeroes it can be generated by replicating the output of the recognizer according to the number of ones present in the scribing pattern and then shifting the replicas to place the ones in their proper positions.

III. Optical Systems for Recognition and Substitution Operations in Symbolic Substitution

When considering optical implementations of symbolic substitution, data can be coded using either intensity [1] or polarization [2]. To avoid ambiguities in processing when intensity coding is used, it is necessary to code logical ones and zeroes as patterns of both high and low intensity, not simply one or the other. Since the complement of a result is always present, intensity coding is also referred to as dual-rail logic [1].

For polarization-based logic systems it is unnecessary during recognition to replicate and shift according to the zeroes of the search pattern [2]. Instead, polarization rotators and a polarized filter produce a null response if the search pattern is present. The operation of the NOR gate and the substituter are unchanged.

In Ref. 1 an optical system is presented that employs a Michelson interferometer to produce two shifted replicas for a simple dual-rail logic system. The operation of the system, however, is based on geometrical optics and not diffraction. With diffraction-based systems, though, it is possible to achieve higher order replications and shifts. To this end, both single-channel and dual-channel systems have been designed for performing symbolic substitution; the number of channels indicates the number of holographic elements utilized.

In a single-channel system a single hologram first produces multiple replicas of the input data. For dual-rail logic the number of replicas is determined by the total number of logical zeroes in all the patterns to be recognized; for polarization based logic, the number of replicas is equal to the number of substitution rules. The necessary shifts or polarization rotations are then accomplished separately using prisms or rotators.

A dual-channel approach uses two holograms to produce the replications and shifts simultaneously. The dual-channel system, however, can only be used for dual-rail logic since it does not allow for rotation of polarization.

A. Single-Channel Systems

In a single-channel system, all operations (replication, shift, and polarization rotation) are performed separately and, therefore, follow each other sequentially as represented in Fig. 1. Figure 1a represents schematically a recognition system using dual-rail logic, wherein replication of the input is performed by the hologram and the necessary shifts are performed using prisms. The second set of prisms images each of the shifted replicas on top of each other. Figure 1b is a polarization-based logic system. Replication of the input is again performed by a hologram, the necessary changes in polarization are achieved using polarization rotators, and the prisms allow the images to be overlapped. The final polarizing filter completes the recognition system by producing a logical zero if the search pattern is present.

Substitution systems for dual-rail and polarization logic can be constructed by reversing the order of operations in the recognition systems. The role of holographic combiners and splitters must be interchanged. Several procedures for designing the splitting and combining holograms are possible, including iterative techniques [3,4] and the solution of nonlinear equations [5,6].

B. Dual-Channel Systems for Dual-Rail Logic

As mentioned above, a Michelson interfermeter can be used to produce two shifted replicas of an input. However, using two phase-only holograms in conjunction with the interferometer it is possible to realize multiple shifted replicas. It is necessary, though, that the holograms be placed in the Fourier plane of the system as represented in Fig. 2. The form of the holograms depends on whether the lenses in the system are cylindrical or spherical, as is described below.

1. Cylindrical Lenses

In a single-channel approach to dual-rail symbolic substitution the replications and shifts are performed separately. In a dual-channel approach holograms are used to both replicate and shift. Since the impulse response of the system is two-dimensional, it can be expressed in terms of several one-dimensional responses and one-dimensional methods can still be used for hologram design. An optical system

13

for realizing such an implementation must be capable of imaging in one dimension and Fourier transformation in the other. This can be accomplished using cylindrical lenses: the holograms being designed on a row by row basis, each row producing a one-dimensional sequence of replicas. The necessity for two holograms follows from the need to produce separate holograms corresponding to the real and imaginary parts of the transfer function [7].

2. Use of Spherical Lenses--Dual-Phase Method

With spherical lenses the transfer function of the system in Fig. 2 can be written

$$P(u,v) = |P(u,v)| \exp\{j\Theta(u,v)\} \tag{1}$$

$$= (1/2) \{\exp[j\Theta_+(u,v)] \exp(j\phi) + \exp[j\Theta_-(u,v)]\}.$$

where

$$\Theta_+(u,v) = \Theta(u,v) + \cos^{-1}|P(u,v)| - \phi. \tag{2a}$$

$$\Theta_-(u,v) = \Theta(u,v) - \cos^{-1}|P(u,v)|. \tag{2b}$$

Equations (1) and (2) will be refered to as the dual-phase representation of P(u,v). The dual-phase decomposition is neither new nor novel, being used first as a method for designing single phase-only holograms [8]. To construct a single phase-only hologram the effects of the cosine, or parity, term must be reduced. However, using two pupil functions Eq. (1) can be realized exactly to within the limits allowed by quantization. To this end the algorithm presented in Ref. 9 has been improved to further reduce the effects of quantization error [7].

IV. Discussion and Concluding Remarks

Symbolic substitution represents a new and powerful logic wherein spatial location of data is as important to function realization as is the data itself. The two-dimensional nature of symbolic substitution logic is therefoe ideally suited to an optical implementation. As has been presented here, the simplicity of the logic philosophy is augmented by the simplicity of the optical systems necessary to construct a symbolic substitution system.

A symbolic substitution system requires only a pattern recognizer, an optical NOR gate, and a pattern substituter. Since the recognizer and substituter perform similar operations, similar optical systems can be used to realize them. Optical systems using only classical optical elements and phase-only holographical elements have been presented for realizing these operations.

*Permanent address: Department of Electrical Engineering, Thornton Hall, University of Virginia, Charlottesville, Virginia 22901.

References

1. K.-H. Brenner, A. Huang, and N. Streibl, "Digital optical computing with symbolic substitution," *Appl. Opt.*, vol. 25, pp. 3054-3060 (1986).
2. K.-H. Brenner, "New implementation of symbolic substitution logic," *Appl. Opt.*, vol. 25, pp. 3061-3064 (1986).
3. N. C. Gallagher and B. Liu, "Methods for computing kinoforms that reduce image reconstruction error," *Appl. Opt.*, vol. 12, pp. 2328-2335 (1973).
4. J. R. Fienup, "Iterative method applied to image reconstruction and to computer-generated holograms," *Opt. Eng.*, vol. 19, pp. 297-305 (1980).

5. H. Dammann and E. Klotz, "Coherent optical generation and inspection of two-dimensional periodic structures," *Opt. Acta*, vol. 24, pp. 505-515 (1977).

6. W. B. Veldkamp, J. R. Leger, and G. J. Swanson, "Coherent summation of laser beams using binary phase gratings," *Opt. Letters*, vol. 11, pp. 303-305 (1986).

7. "High efficiency systems for symbolic substitution," Final report to the Deutsche Forschungsgemeinschaft.

8. D. C. Chu and J. W. Goodman, "Spectrum shaping with parity sequences," *Appl. Opt.*, vol. 11, pp. 1716-1724 (1972).

9. C.-K. Hsueh and A. A. Sawchuk, "Computer-generated double-phase holograms," *Appl. Opt.*, vol. 17, pp. 3874-3883 (1978).
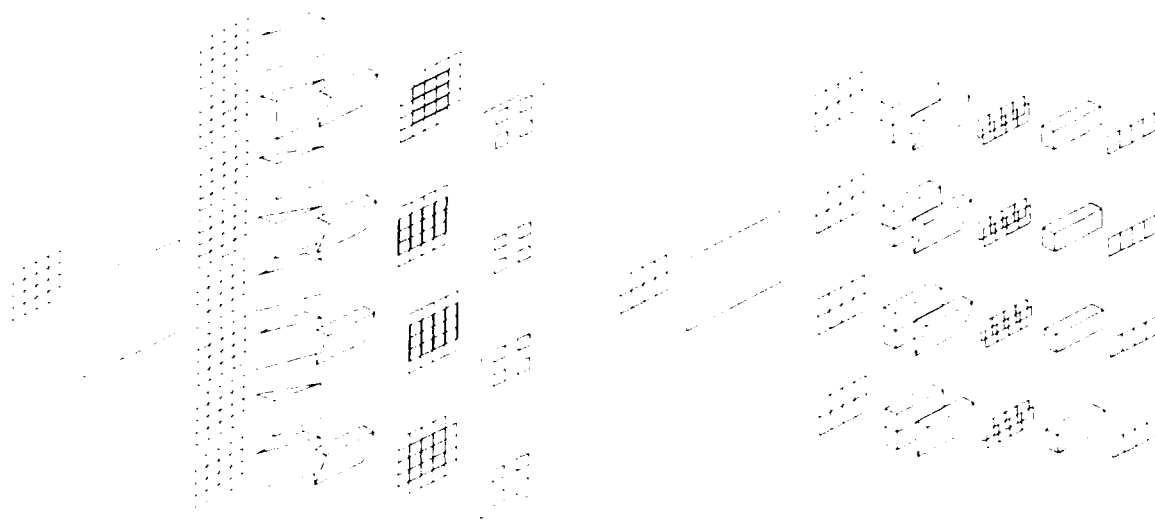
Figure 1. Single-channel recognition system using (a) dual-rail logic and (b) polarization-based logic.
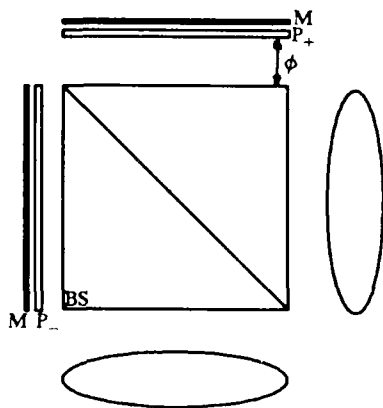
Figure 2. Dual-channel Michelson interferometer.

THE STRENGTHS AND WEAKNESSES OF OPTICAL ARCHITECTURES BASED ON SYMBOLIC SUBSTITUTION

Thomas J. Cloonan
AT&T Bell Laboratories
Naperville Rd., Naperville, IL. 60566

1. Introduction– Symbolic substitution (S.S.) is a parallel technique for pattern replacement within a binary array. S.S. can be used in optical architectures to perform many different operations[1]. There are several different ways to implement hardware which performs S.S. operations. This paper presents four different optical architectures based on S.S., and the strengths and weaknesses of each architecture are analyzed. This paper will deal only with architectural issues, leaving the detailed issues of implementation to future studies.

2. Architecture A– Architecture A (Fig. 1) implements "bit-by-bit S.S." using bistable latches as storage elements. Bit-by-bit S.S. is defined to be S.S. in which the matching phase (identification of the Left-Hand Side (LHS) pattern) is performed by repeatedly shifting array A and exposing it on array B. The scribing phase (writing of the Right-Hand Side (RHS) pattern wherever the LHS pattern was found) is performed by repeatedly shifting array B and exposing it on array C. The results in array C are then inverted by the inverting array and written back into array A (in preparation for the execution of the next set of rules).

The bistable optical NOR latch used in Architecture A is "powered up" in the set state, and any logic "1" input resets the device to the reset state. Thus, the device can record the occurrence of any logic "1" input. Unfortunately, the latch can only be returned to the set state by re-initializing the entire array. If the destination array is re-initialized prior to data movements, then the shuttering spatial light modulators (SLM's) can control the flow of data. The dynamic beam-steering elements can provide five global space-invariant connections corresponding to North, South, East, West, and Straight data movements in the latch arrays. Any two of these data movements can be implemented with a single pass around the small array loops, because each loop contains two beam-steering elements. Combinations of these data movements can be used to shift arrays any distance in any direction.

Architecture A has two main strengths. Its primary strength is that any Boolean function can theoretically be implemented by executing a different sequence of LHS→RHS rules. Thus, the hardware is capable of performing AND, NAND, OR, and NOT functions even though the basic logic element is a bistable NOR latch. Another strength of Architecture A is realized if the hardware performs parallel processing operations using Single-Instruction Multiple-Data (SIMD) concepts. For example, many pairs of numbers can be added in exactly the same amount of time as a single pair of numbers, because the same set of LHS→RHS rules are used in either case. Thus, the effective processing power of the simple architecture can be greatly increased by merely enlarging the latch arrays to allow for more data storage.

Architecture A has several distinct weaknesses. The most obvious weakness is the extensive processing time required to implement relatively simple functions. This is a direct result of the serial execution of many shift and expose instructions. The processing time is also increased due to the use of bistable latches in a dual-rail system. In order to maintain integrity on the dual-rail data, extra S.S. rules must be implemented to match on all possible input bit combinations in array A so that all of the bits in array C will be set to valid dual-rail values. The processing time can also suffer from the fact that array A must be cleared prior to the feedback of data from array C. As a result, data in array A can be saved only by copying it to array C. This will greatly add to the processing time. Another weakness is that Architecture A requires a large number of bistable latches. This hardware complexity is due to the use of dual-rail logic, and can also be due to the use of distinguishing symbols around operands.

3. Architecture B– Architecture B (Fig. 2) is very similar to Architecture A, because it also implements S.S. on a bit-by-bit basis. The primary difference is that Architecture B uses R-S

flip-flops as storage elements instead of bistable latches. The R-S flip-flop is a logic element with two optical inputs (set and reset) and two optical outputs (uncomplemented data and complemented data). Inputs to the flip-flop must exceed a specific threshold intensity to set or reset the device, so the flow of data can be controlled using enable pulses instead of SLM's. The elimination of SLM's simplifies the hardware, but it also complicates the dynamic beam-steering elements, because the beam-steering elements must perform the multiplexing function. For the architecture in Fig. 2, six different connections are provided by the beam-steering elements. Five of them are the North, South, East, West, and Straight data movements used in Architecture A. The sixth connection is the Copy connection which multiplexes data from array A to array B (or B to C). In addition to providing these connections, the hardware must also provide selective enables on the set and reset inputs to the flip-flops. The operation of Architecture B requires data from array A to be copied into array B. Matching of the LHS pattern is achieved by shifting and exposing array B onto array C. Scribing of the RHS pattern is achieved by shifting and exposing array C back onto array A.

Due to the similarities between Architecture A and Architecture B, all of the strengths of Architecture A are also found in Architecture B. Thus, Architecture B can support SIMD processing, and it can also perform (in theory) any Boolean function. Architecture B offers several other strengths due to its use of R-S flip-flops. One of the biggest advantages is that flip-flops are automatically set up in a valid state (i.e., complementary signals always have complementary values). This can eliminate the need for matching on all possible input combinations and greatly improve the overall processing speed. Another strength is due to the fact that flip-flop arrays do not need initialization prior to scribing. Since scribing of array A occurs only where matches occurred, unmatched regions of the array are left unaltered, and copying of data is no longer necessary. The elimination of initializations and selective copying can help improve the processing speed of the architecture.

Since Architecture A and Architecture B share many similarities, they also share many of the same weaknesses. One of the weaknesses that they share is their slow processing speeds (which result from numerous bit-by-bit processing steps). Both of the architectures also require a large number of gates, and both require the inefficient use of distinguishing symbols. One disadvantage found in Architecture B (and not in Architecture A) is the need for selective enables on the flip-flop arrays. These can add to the hardware complexity.

4. Architecture C– Architecture C (Fig. 3) is slightly different from the previous architectures, because it employs "parallel S.S." Parallel S.S. is defined to be S.S. in which the matching phase and scribing phase take place simultaneously. In addition, all of the bits in the LHS pattern are scanned in parallel and all of the bits in the RHS pattern are written in parallel. This eliminates the need for shifting the data arrays, which greatly simplifies the system hardware. The hardware requires only two storage arrays (X and Y). These storage arrays can be built with either bistable latches or R-S flip-flops, but the design presented here will utilize flip-flops. The processing logic block consists of a dynamic beam-steering element (to select the set of rules to be executed), a beam splitter (to split the source array up into N identical copies for the N LHS→RHS rules to be executed), a static input beam-steering element (to direct the beams to the appropriate AND gates for matching), an array of AND gates (to detect the desired input combinations), and a static output beam-steering element (to direct the beams to the appropriate set or reset inputs on the destination array of flip-flops).

A typical instruction is executed by first copying the data from array X to array Y. The processing logic then matches on array Y and scribes the appropriate results back into array X. Data flow is controlled using enable pulses as in Architecture B. The dynamic beam-steering element allows the user to choose between several different functions offered by the static beam-steering elements. For example, a typical system might provide static elements to do LHS→RHS rules for both addition and multiplication. The configuration of the dynamic beam-steering element would determine which of these functions is executed.

Architecture C has several unique strengths. The biggest advantage is its increased processing speed, which results from the execution of multiple rules in parallel and from the elimination of array shifts. Its speed can also be increased since the need for data copying is eliminated (results are written directly back into the appropriate flip-flops of array X). Architecture C can also benefit from SIMD processing (as did the previous two architectures). Another strength of Architecture C is its reduced gate count. Only the two storage arrays (X and Y) and the AND gate array are required for the entire system.

There are several weaknesses associated with Architecture C. The biggest weakness is that parallel S.S. is not flexible, because it is not capable of implementing all Boolean functions (like bit-by-bit implementations can). The only Boolean functions it can implement are those which are supplied by the static beam-steering elements in the processing logic. There are also physical limits to the number of times a light beam can be split, and this limits the number of substitution rules which can be executed in parallel. Architecture C still requires the use of distinguishing cells (as did the previous two architectures), so this is another weakness.

5. Architecture D– Even though Architecture D (Fig. 4) implements parallel S.S., it is very different from the other architectures because it employs "one-rule S.S." Murdocca presented one-rule S.S. as a means of implementing cellular automata[2]. Using one-rule S.S., any Boolean function and any data shift can be implemented through repetitive execution of only a single LHS→RHS rule. Instead of the Boolean function being defined by substitution rules in an external control unit (as in the previous three implementations), the Boolean function is defined by bit patterns held in the storage array along with the regular data bits.

A single processing cycle starts with the instruction patterns and "input" data bits stored at specified locations in the X array. These instruction patterns and data bits are then copied to array Y. The substitution rule would then be applied to the entire Y array with the results selectively scribing array X. This processing cycle would have to be repeated many times before the "output" data bits would eventually appear at specified locations in array X.

The primary strength of Architecture D is that any Boolean function can be implemented using only a single LHS→RHS rule. As a result, the hardware in the processing logic is simplified. In addition, the elimination of dynamic beam-steering elements greatly reduces the overall system complexity. Like Architecture C, Architecture D allows for SIMD processing and eliminates the need for data copying (which can improve processing speeds).

The two biggest weaknesses of Architecture D are the large storage array sizes and the slow processing speeds. Both of these problems can be seen as trade-offs for the simplicity of the single substitution rule. The system requires large storage arrays, because the instruction patterns consume a large number of flip-flops in the storage arrays. Slow processing speeds result from the many cycles which must be repeated to produce final results.

6. Discussion– Several general observations can be made about architectures based on S.S. First, as a result of the two-dimensional nature of pattern matching, the systems do take advantage of the inherent parallelism of optics. In addition, they are well-suited for SIMD applications (without requiring extensive software modifications). However, most S.S. architectures also share the inherent problems of slow processing speeds and the need for data boundary identification (forcing the use of inefficient distinguishing cells).

In order to compare processing speeds, a typical application (4-bit binary addition) was simulated for each of the architectures. If T_L = latch (or flip-flop) switching time, T_D = dynamic beam-steering element switching time, and T_S = SLM switching time, then the total processing times were given by:

1) Architecture A = $232 {}^* T_L + 136 {}^* T_D + 204 {}^* T_S$ 3) Architecture C = $12 {}^* T_L + 4 {}^* T_D$

2) Architecture B = $228 {}^* T_L + 180 {}^* T_D$ 4) Architecture D = $450 {}^* T_L$.

Some generalizations regarding S.S. architectures can be deduced from these results. First, architectures that employ R-S flip-flops tend to be faster than those that use bistable latches.

Secondly, as hardware complexity increases, processing speed also tends to increase. However, as hardware complexity increases, system flexibility often decreases (i.e, the system is unable to perform all Boolean functions). System designers must weigh these trade-offs carefully before deciding on the type of architecture to use in an optical processing system.

7. Acknowledgements– The author would like to thank Scott Hinton, Tony Lentine, Rick McCormick, and Miles Murdocca for their invaluable contributions to this work.

8. References–

[1] Huang, A., "Parallel Algorithms for Optical Digital Computers," 10*th International Optical Computing Conference*, *Proc.*, IEEE Computer Society, April 6, 1983, pp. 13-17.

[2] Murdocca, M. J., "Digital Optical Computing with One-Rule Cellular Automata", submitted to *Applied Optics*.
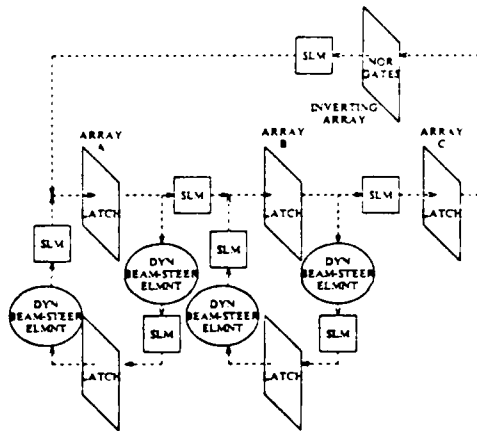
FIG. 1- ARCHITECTURE A
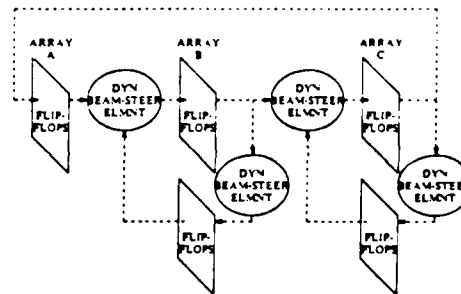
FIG. 2- ARCHITECTURE B
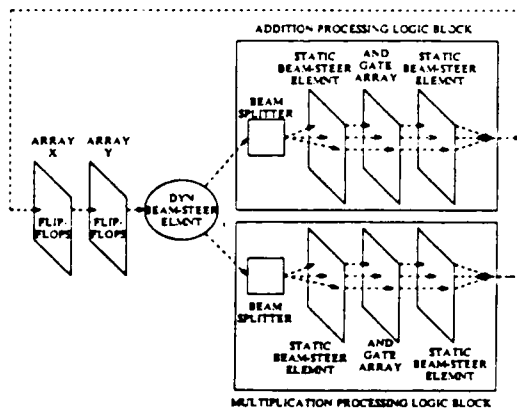
FIG. 3- ARCHITECTURE C
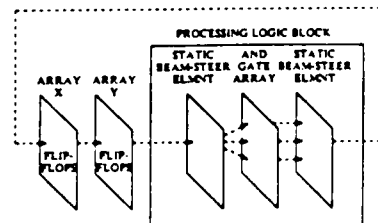
FIG. 4- ARCHITECTURE D

Binary Image Algebra and
Digital Optical Cellular Image Processors

K. S. Huang, B. K. Jenkins, A. A. Sawchuk
Signal and Image Processing Institute
Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90089-0272

Image processing and image analysis tasks have large data processing requirements and inherent parallelism and are well suited to implementation on digital optical processors because of the parallelism and free interconnection capabilities of optical systems [1][2]. Recently, several techniques for constructing optical cellular logic processors for image processing have been proposed [2]-[5]. Through parallel studies of architectures, algorithms, mathematical structures, and optics we have found that: 1) cellular automata are appropriate models for parallel image processing machines [6]; 2) an image algebra extending from mathematical morphology [7] [8] can lead to a formal parallel language approach to the design of image processing algorithms; 3) the algebraic structure serves as a framework for both algorithms and architectures of parallel image processing; and 4) optical computing techniques are able to efficiently implement image algebra based on cellular logic architectures (e.g. cellular array, cellular hypercube etc.). Here we will first discuss image algebra and then architectures for its implementation.

An axiomatic image algebraic structure has been developed to provide a standardized, unified, efficient, and simple mathematical structure for image processing. Two special cases of this are "binary image algebra(BIA)" which deals with 2-D binary digital images and "spatial image algebra(SIA)" which is a generalization of BIA and deals with gray-level and complex-valued images. In these algebraic structures, images are vectors in a space, and image description or information extraction is done by using reference images to model or transform the original image to a final state in which the desired property can easily be measured. Thus, the art of designing image processing algorithms becomes how to choose "good" reference images and transformations, which play the same role as the reference axes in describing a vector in a space.

In BIA, an image X is defined as an element of $P(W)$: the power set of the universal image W ($W = \{(a,b) \mid a \in Z, b \in Z\}$, where $Z = \{0, \pm1, \pm2, ..., \pm n\}$ and n is an integer), and an image transformation is a function $T : P(W) \rightarrow P(W)$. We have shown that two fundamental principles can serve as the basis of BIA:

Principle 1. Fundamental Principle of Image Transformations
Any image transformation T can be implemented by using appropriate reference images R and the three fundamental operations: (1) Complement $\overline{X}$ of an image X, (2) Union $\cup$ (or Intersection) of two images, and (3) Dilation $\oplus$ (or Erosion) of two images;

Principle 2. Fundamental Principle of Reference Images
Any reference image R can be generated from a basis set of elementary images E_i that includes a pixel at the origin (0,0) and its four nearest neighbors, by using the three fundamental operations.

In practical applications, a reference image R can be generated from a set of elementary image(s) E_i by a "sequential dilation". Symbolically, if $R = E_1 \oplus E_2 \oplus ... \oplus E_k$, then

$$X \oplus R = (...((X \oplus E_1) \oplus E_2) \oplus ... \oplus E_k).$$

Thus, a small programmable neighborhood configuration mask with a simple gate array and an interconnection network can be used to carry out any operation which employs an arbitrary reference image. Figure 1 shows a block diagram of a digital optical cellular image processor (DOCIP) which implements the two fundamental principles in parallel.

This system is a finite state machine based on cellular logic which minimizes the optical hardware complexity and can easily implement BIA algorithms. To avoid the well-known drawbacks of conventional computers based on von Neumann principles [1]-[5], the machine in Fig. 1 has one instruction which implements the three fundamental operations of Principle 1 along with fetch and store. This design usees the parallelism of optics to simultaneously execute instructions involving all N^2 picture elements.

Basically, the proposed DOCIP as shown in Fig. 1 is a cellular SIMD machine and consists of an array of cells or processing elements (PEs) under the supervision of a control unit. The control unit includes a clock, a program counter, a test and branch module for feedback control, and an instruction decoder for storing instructions and decoding them to supervise cells. The array of cells includes a destination selector, three memory elements for storing images, a memory selector, and a dilation unit.

The entire system can be realized by an optical gate array with optical 3-D interconnections [1] [2] [9] [10] . Alternatively, control of the DOCIP can easily be realized by using an electronic host instead of the optical control unit, since control of SIMD systems is primarily a serial process. The tradeoff is a possible inefficiency in the interfaces between electronic and optical units. Because of this, the all-optical approach may be preferable in the long term.

The DOCIP shown in Fig. 1 operates as follows: (1) a binary image ($N \times N$ matrix) is selected by the destination selector and then stored in any memory as the instruction specifies; (2) after storing the images (1 to 3 $N \times N$ matrices), these images and their complemented versions are piped into the next stage, which forms the union of any combination of images; (3) the result is sent to a dilation where the reference image specified by the instruction is used to control the type of dilation; (4) finally, the dilated image can be output, tested for program control, or fed back to step (1) by the address field of the instruction. The allowed configuration of the reference images E_i at a cycle actually define the interconnection network of DOCIP. Therefore, the system of Fig. 1 can implement a conventional nearest-neighbor connected cellular array (DOCIP-array), and can be extended to a cellular hypercube (a two dimensional DOCIP-hypercube is shown in Fig. 2) which is very difficult to realize on a planar VLSI chip [11].

We have performed computer simulations and some preliminary gate-level design work on the cells for both the DOCIP-array and DOCIP-hypercube. To efficiently utilize optical gates, they can be interconnected them with a 2-D optical multiplexing technique in which a common controllabe mask is used for all cells. The optical multiplexing technique has following advantages: 1) the DOCIP will no longer require the broadcasting of instructions from the control unit; 2) it will reduce the number of gates; and 3) each cell has a simple structure — essentially containing only a 3-bit memory with inverting and non-inverting outputs, and a multiple-input OR gate for dilation.

In the DOCIP-array, each cell is connected with its 8 nearest neighbors (this is called a Moore neighborhood) or its 4 nearest neighbors (this is called a von Neumann neighborhood). Our preliminary design work indicates that each cell will require $O(1)$ gates ($\sim$ 43 3-input NOR gates for the 8-neighborhood case, $\sim$ 37 3-input NOR gates for the 4-neighborhood case). By further applying the optical multiplexing technique as stated above, it can be reduced to $\sim$ 22 3-input NOR gates per cell for the 8-neighborhood case and $\sim$ 20 3-input NOR gates per cell for the 4-neighborhood case. The DOCIP-array performs global operations (employing reference images R of size $O(N) \times O(N)$) in $O(N)$ time, but requires only $O(1)$ to carry out input/output and local operations (employing reference images R of size $O(1) \times O(1)$).

In the DOCIP-hypercube, each cell has $O(logN)$ connections (at most $4\lceil log((N+1)/2)\rceil + 1$ connections for extending the 4-neighborhood and at most $8\lceil log((N+1)/2)\rceil + 1$ connections for extending the 8-neighborhood) for an $N \times N$ array. The complexity (the number of gates required) of each cell will increase to $O(logN)$ which is proportional to the number of connections

for each cell. For example, considering a 127×127 array: 1) when extending the 4-neighborhood, each cell in the DOCIP-hypercube has at most 25 connections and requires ~ 47 3-input NOR gates or ~ 30 gates with the optical multiplexing technique; 2) for extending the 8-neighborhood, each cell in the DOCIP-hypercube has at most 49 connections and requires ~ 59 3-input NOR gates or ~ 38 gates with the optical multiplexing technique.

In contrast with the DOCIP-array, the DOCIP-hypercube increases the interconnection complexity to $O(logN)$ and cell complexity to $O(logN)$, but is able to perform global operations in $O(logN)$ time. Comparing with the conventional electronic array processors having serial or N-parallel input/output, the DOCIP-array will have the same order of performance in local and global operations but will be improved in input/output performance. The DOCIP-hypercube will not only be improved in input/output performances but also in global operations. One important feature in the design of the DOCIP-array and DOCIP-hypercube is that optical 3-D free interconnection capabilities can be used to reduce the cell hardware requirements as well as solve the global connection and I/O problems which are difficult to solve by the planar VLSI technology.

Another interesting question is: "Can we also build an *analog* optical computer to do morphological image processing?" The answer is "yes", because the dilation and erosion can be achieved by adding thresholding to the convolution and correlation operations of Fourier optics. However, analog optical morphological processors will face analog drawbacks such as dynamic range, accuracy limitations, and flexibility limitations etc. as do other analog systems. On the other hand, DOCIP not only offers the advantages of digital systems and optical signal processing, but also can be implemented with hardware of low complexity.

References

[1] A. A. Sawchuk and T. C. Strand, "Digital Optical Computing," *Proc. IEEE*, Vol.72, pp. 758-779,1984.

[2] B. K. Jenkins and A. A. Sawchuk, "Optical Cellular Logic Architectures for Image Processing," *IEEE Computer Society Workshop on Computer Architecture for Pattern Analysis and Image Database Management*, Miami, Florida, pp. 61-65, Nov. 1985.

[3] T. Yatagai, "Cellular Logic Architectures for Optical Computers," *Applied Optics*, Vol. 25, pp. 1571-1577, 1986.

[4] J. Tanida and Y. Ichioka, "Optical-logic-array Processor Using Shadowgrams. II. Optical Parallel Digital Image Processing," *J. Opt. Soc. Am. A*, Vol. 2, No. 8, pp. 1237-1244, 1985.

[5] H. J. Caulfield, "Systolic Optical Cellular Array Processors," *Optical Engineering*, Vol. 25, No. 7, pp. 825-827, 1986.

[6] K. Preston and M. J. B. Duff, *Modern Cellular Automata — Theory and Applications*, Plenum Press, New York, 1984.

[7] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, Inc., New York, 1982.

[8] R. M. Longhead, D. L. McCubbrey, and S. R. Sternberg, " Cytocomputers: Architectures for Parallel Image Processing," *Proc. IEEE Workshop Picture Data Description and Management*, Pacific Grove, CA, pp. 281-286, 1980.

[9] B. K. Jenkins, et al., "Sequential Optical Logic Implementation," *Appl. Opt.*, Vol. 23, pp. 3455-3464, 1984.

[10] B. K. Jenkins, et al., "Architectural Implications of a Digital Optical Processor," *Appl. Opt.*, Vol. 23, pp. 3465-3474, 1984.

[11] A. Rosenfeld, "Parallel Image Processing Using Cellular Arrays," *Computer*, Vol. 16, No. 1, pp. 14-20, 1983.
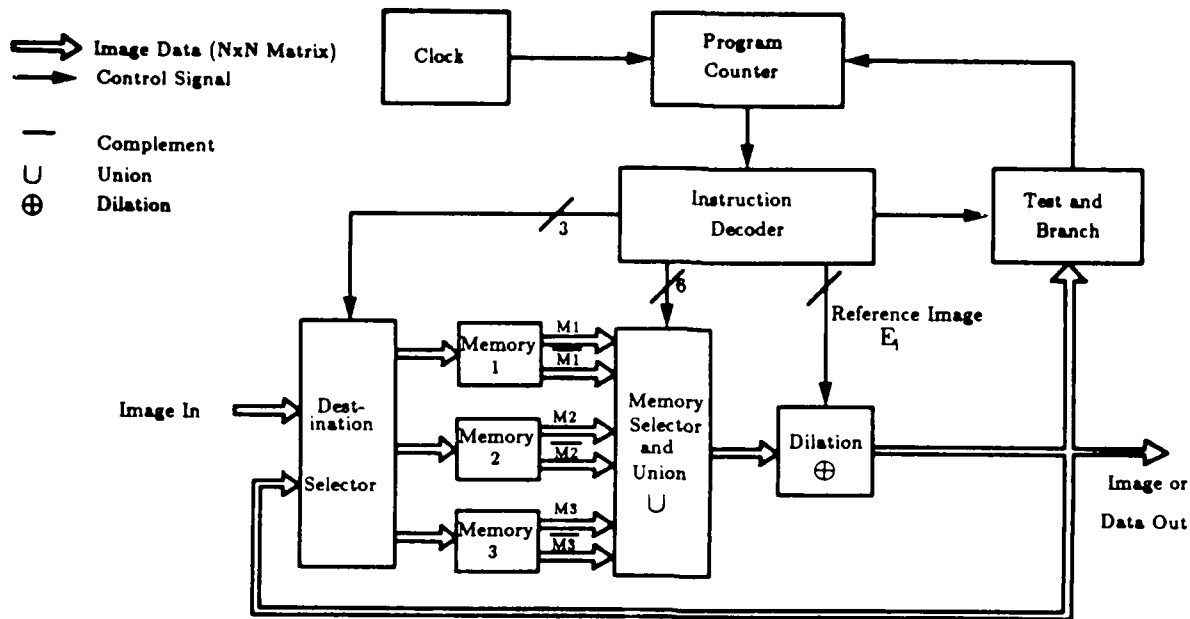
Figure 1. A digital optical cellular image processor (DOCIP) architecture — one implementation of binary image algebra (BIA). The DOCIP-array requires 9 (or 5) control bits for reference image E_i. The DOCIP-hypercube requires $O(\log N)$ control bits for reference image E_i.
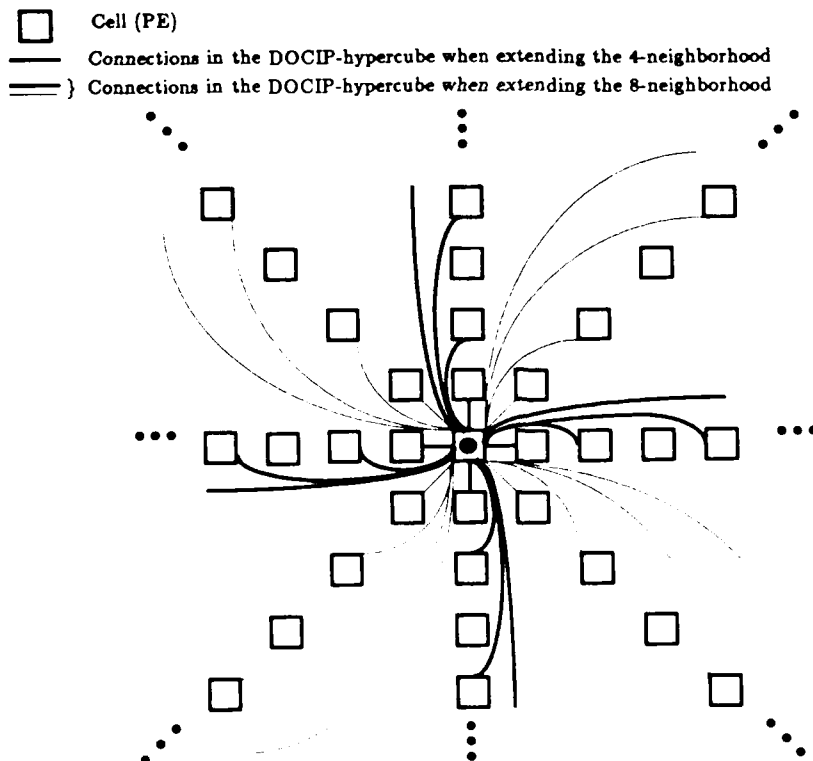

Figure 2. A two-dimensional cellular hypercube — DOCIP-hypercube. Each cell connects with cells in the 4 or 8 directions (depending on extending the 4- or 8-neighborhood) at distances $1, 2, 4, 8, ..., 2^k$ from it. Here, only the connections of one cell with those at distances 1, 2, and 4 are shown.

A Bit Serial Optical Computer

Harry F. Jordan
Department of Electrical and Computer Engineering
Universi'y of Colorado
Boulder, CO 80309-0425

Summary

Optical techniques have a number of potential benefits for information processing. They include a high degree of parallelism, high speeds, short pulses and non-interference of signals. Many of the attempts to use the non-linear optical technology which is beginning to emerge in the device physics area has concentrated on the spatial parallelism available in optics. Problems which have arisen with this approach have included design and fabrication of arrays of optical elements, accuracy of data representation, controlled permutation of data, synchronization with a master clock and optical to electronic interfacing. But spatial parallelism is not the only advantage of optical computing. The extremely high speeds and short pulses which are possible allow the exploitation of the time domain to obtain significant processing power. Serial computer designs are based on the time domain. Exploiting the time domain is not an alternative to spatial parallelism in optical computing but is complementary to it. Since information in an optical computer is represented by pulses propagating at the speed of light there is a homogeneity between time and space which is not present in an electronic computer. Difficult serial design issues must be addressed in optical computer architectures even in predominantly parallel designs. Finally, there are reasons why a serial architecture may lead more rapidly to functional optical computers than one relying heavily on spatial parallelism.

The immature state of optical logic and switching elements from the perspective of computer architecture promises to delay the study of optical computer architectures unless a way can be found to produce interesting architectures using only a few logic elements. The situation with optical devices today is much the same as was the situation in the 1950's with electronic devices; there are few working devices and they are of a rudimentary nature. One of the prime architectures of the 1950's was the bit serial one; an architecture which required few devices, yet achieved its complexity through the speed of those devices. In this type of machine, logic for a single bit suffices to handle all bits of a word. Useful machines were built with as few as a dozen active electronic logic elements and delay line type memory. In optical computing, bit-serial design is not merely a way of implementing a significant system with few components. The duality of time and space which underlies a system using photons to represent information makes an understanding of serial operation essential to the design of any optical architecture. As a simple illustration of the impact of this duality, consider the addition of numbers represented digitally by bits presented at equally spaced positions along a line in space at an instant of time. The fundamental problem in digital addition is the carry, which allows information from the low order bits to propagate as far as the high order ones. If the bits are represented electronically,

the carry propagation can be done with logarithmic delay using the standard carry lookahead circuit, as sketched in Fig. 1. But, the logarithmic nature of the result depends on the assumption that all delays are lumped in active devices and that none occur along interconnecting wires. If the bits are represented optically, the interconnection time is as important as the logic delay, and the carry propagation time becomes linear in the number of bits. In fact, the fastest way to propagate the carry is to have the photon packets representing the bits move transversly along their line of presentation past an active optical element which will compute and propagate the carry. This is a bit serial approach.

Data storage in early bit-serial computers was usually supplied by some delay mechanism. Optical fiber loops seem to offer the best method for delay line data storage in the proposed system. The fibers would also supply interconnection between system components. The fact that interconnection is supplied by the same delay lines used for data storage naturally introduces pipelining at the logic design level and gives a significant geometric component to the architectural design. To preserve information for long periods, signal level restoration is essential. To locate information in a storage loop and access it reliably, temporal synchronization is required. Level restoration
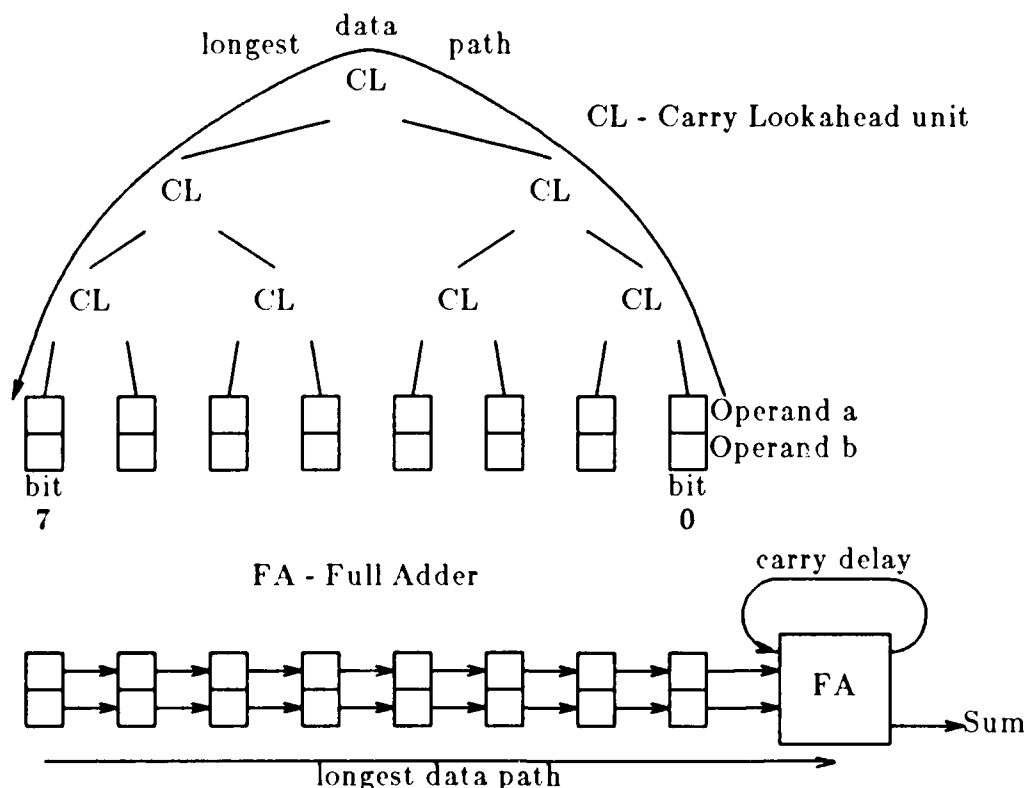
Figure 1: Optical Carry Path in Addition

and synchronization are basic to all parts of the computer architecture, but they are perhaps most easily discussed in relationship to the memory loops. Amplification can be provided by discrete excitation locked lasers, mimicing the use of discrete transistor amplifiers in electronic systems. A promising, but more remote, possibility is offered by externally pumped, doped optical fiber, in which distributed amplification of propagating pulses has been demonstrated. Several alternatives for temporal resynchronization in the memory loops are possible. Self-synchronizing bit streams containing timing, address and data bits are used in single track recording devices such as flexible disks and are perhaps the most robust mechanism for optical data storage. The key factor to be studied in this connection is the extent of the optical logic required to extract the information encoded in such a stream. Other alternatives are electronically controlled optical phase shifters to initially align and correct for differential drift in multiple synchronized, delay line storage loops. If discrete optical amplifiers are used for level restoration, these can be clocked to provide simultaneous temporal resynchronization.

Another method used to process multiple data items with the same hardware is that of pipelining. This general technique appears in the form of overlap in sequential computers, vector arithmetic units in supercomputers and systolic arrays in VLSI design. Shared memory multiprocessors have also been built by pipelining instructions, as well as data streams. An architecture drawing ideas from both bit serial computers and multiple stream pipelining promises to yield a high degree of computational richness using only a few optical switching elements, leading to an actual hardware implementation in a relatively short time frame. An implementation would not only lead to a better understanding of optical computer architecture but would also stimulate and interact with optical device technology. In the bit serial domain, pipelining ideas lead naturally to the time multiplexing of independent information streams. This gives a natural interface between optical processing at high speeds but with limited parallelism and electronic devices where spatial parallelism is well developed but processing speed is limited. In the early stages of development, a moderate sized electronic computer will be required to supply data, control operation and record outputs from the optical computer.

Bit serial computation requires delay lines of several lengths. A common unit is the one word delay line used for an accumulator or other working register. The longest delay lines are those used for multiple word memory loops while the shortest is the one bit loop used, for example, for the carry bit in addition. Practical construction issues place a lower limit on the size of the smallest loop, but it is possible and desirable to multiplex many, noninteracting bits within such a loop. For example, at a 10 GHz bit rate a one bit delay would have a length of 2 cm (in glass), but, depending on the speed of the optical switching devices, 10 to 100 noninteracting bits could be multiplexed within the 2 cm loop. The technique of processing noninteracting information units in adjacent time intervals is the essence of pipelining. Since there is no distinction between the limitations on processing control or data bits optically, a powerful approach is to time multiplex complete instruction streams to produce a pipelined multiprocessor. This has the

advantage that spatial parallelism is easily incorporated to increase the degree of multiprocessing as arrays of optical logic elements become available. The switching rate of active optical devices can be kept much lower than the bit rate of the switched stream by properly engineering the granularity of the time multiplexed information stream. By granularity we mean a minimum time, and hence space, separation between adjacent bits of the same information packet. Bits which follow each other more closely in an optical stream need not interact and are passed through logic and switches in a pipelined manner.

The most promising technology for future optical computers does not necessarily correspond to that available for immediate application. Without a pilot program using available components, computer architects will be unable to contribute to this important, evolving area until late in its growth cycle. An initial system can be built using $Ti:LiNbO_3$ directional couplers in various hybrid configurations as the passive and active elements of the system and optical fiber as the "between" element transmission medium. A more mature system would use monolithically integrated GaAs circuit elements, connected by optical fiber off chip and by integrated optical waveguide on chip.

High serial bit rates are presently attainable with optical devices. Mode locked semiconductor lasers can, for example, produce 1 psec pulses separated by 100 psec. Pipelining techniques can be used to produce a much faster clock by optically fanning out the low duty cycle clock described above using a passive star coupler, cutting fiber lengths to provide interchannel delays, and recombining the shifted pulse trains with an $N \times 1$ coupler. Such delay line shift multiplexing can increase the clock frequency by one to two orders of magnitude. The generation of input data streams for the optical computer can be done in a similar way. One can provide N electronic data streams, each at a clock rate equal to the optical clock rate divided by N, using a highly parallel electronic computer. Using $Ti:LiNbO_3$ directional couplers, each electronic bit can be strobed with the correctly delayed clock channel, and the information can be time multiplexed in a single fiber using optical fan-in as described above for the clock.

Conclusions

There is a twofold advantage to studying bit-serial optical computers. On the practical side, a hardware implementation is realistic enough to excite the interest of computer architects as well as device designers. Promising ideas and pitfalls can be more rapidly determined with an actual implementation to guide the research. On the theoretical side, there is really no parallel operation in an optical computer since data items separated in space require time to interact. One could imagine that the architecture of an optical computer is limited to the surface of a relativistic light cone. This interchangeability of time and space will lead to some hard problems, even in the most parallel architectures, that can be addressed in their purest form by starting with a bit-serial design.

MONDAY, MARCH 16, 1987

PROSPECTOR / RUBICON ROOM

1:00 PM–2:30 PM

MC1–6

SESSION 3

Joseph W. Goodman, Stanford University, *Presider*

Summary:

Four-dimensional optical crossbar

Adolf W. Lohmann and Wilhelm Stork

Physikalisches Institut der Universität

8520 Erlangen, FRG

An ordinary crossbar is a two-dimensional array of N^2 binary switches. N input channels communicate in any configuration with N output channels. It has been suggested before to implement the N channels as an one-dimensional array of parallel light rays, and to implement the binary switches by polarising beam-splitters and electro-optical half-wave plates. Each of the rays may be subdivided into several pixel channels in order to exploit fully the inherent parallelism of free-space optics.

Such a two-dimensional crossbar would be "planar" in the sense of a typical optical setup with many components (lenses, prisms, mirrors...) mounted at the same height above a table. The next obvious generalisation of such a crossbar would be a cubical volume, filled with beam-splitters, switches etc. A typical group of components (2 beam splitters and two polarisers) is called a "knot". These knots are arranged in three-dimensional cartesian fashion. Each knot would have three input-channels, coming from three orthogonal directions. Three output channels leave the knot at faces opposite to the inputs.

This concept may look attractive from a geometrical point of view. But such a three-dimensional bus would require ternary (three-way) switches as parts of the knots. A three-way switch is not natural for polarisation optics. But a four-way switch can be implemented out of binary (two-way) switches. Hence, a crossbar with conceptual four-dimensional topology is well suited both for accepting two-dimensional arrays of input channels and for the implementation of the switches by optical polarisation components.

CELLULAR OPTICAL PROCESSOR ARCHITECTURE WITH MODULABLE HOLOGRAPHIC INTERCONNECTIONS

J. TABOURY, J.M. WANG, P. CHAVEL
Institut d'Optique, Laboratoire associé au CNRS,
Université de Paris-Sud, B.P. 43, 91406 ORSAY Cedex, France

F. DEVOS
Institut d'Electronique Fondamentale, Laboratoire associé au CNRS, Université de Paris-Sud, 91405 ORSAY Cedex, France

I - INTRODUCTION

The recent rise in interest in optical computing has already led to considerable conceptual progress in optical processing architectures such as multiple matrix product, associative memories, digital optical computing. Since 3-D optics naturally provides a parallel environment with high connectivity, optical computing becomes an attractive field of reflexion. Nevertheless, in optical cellular logic systems using non linear elements, the practical implementation problems of gate interconnections must be solved.

In the present work, we propose a parallel architecture realizing space shift invariant connections between each pixel of an object plane. Such invariance is readily obtained by use of a Fourier plane hologram. However, a second image-plane hologram is added to allow for easy modification of the connection network during algorithm execution. In the next section, we present the principle used to achieve such interconnection using holographic elements. In the last section, we discuss the advantages and limitations of such a network.

II - PRINCIPLE

So as to realize a connection between a pixel $P(i,j)$ and different pixels $P(i+n, j+m)$, we have adopted a double diffraction configuration. A first interconnection hologram H_1 is placed in the object plane. Its first diffraction order illuminates a second hologram H_2 in the Fourier plane ; the first diffraction order of H_2 reconstructs shifted versions of the original object in the image plane (Figure 1) :

Figure 1 : Interconnection architecture comprising one image plane hologram and one Fourier hologram. Optical feedback and nonlinear element are not shown.
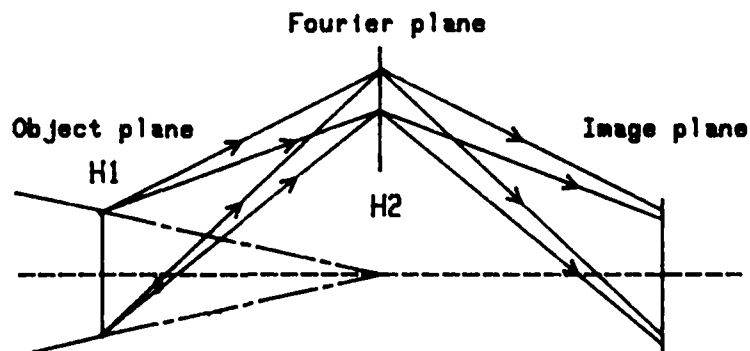
The connection between pixel P(i+n, j+m) and pixel P(i,j) for all (i,j) leads to shift the image of the former onto the latter. The translation vector $\vec{T}(n,m)$ must be invariant over the whole image.

This property is realized by an appropriate hologram H_2 in the Fourier plane. So as to connect s pixels on to pixels P(i,j) s different holograms are recorded ; each one is individually associated to a unique translation vector $\vec{T}(n,m)$. H_2 is therefore subdivised into s subholograms. Any pixel interconnection scheme (see figure 2) out of the 2^s possible schemes can be implement by selecting the appropriate subholograms with a (programmable) shutter mask.

Object   Shutter   Image
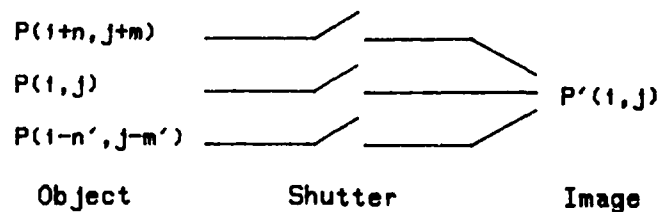
Figure 2 : *The s connections are selected by appropriate shutter set out of 2^s connection schemes.*

This configuration is suitable for any connection determined by the computing algorithm in a cellular feedback architecture.

Such flexibility, not shown by previous space invariant digital optical setups [1-2], is possible thanks to the addition of hologram H_1 in the object plane (or a conjugated plane) to divide the wavefront behind the object into s convergent beams (Figure 1). s shutters set the required connections during each step of algorithm. This is a convenient optical solution to the problem of the s x N x N leads and switches which would be required in an electronic implementation of the same machine (N x N is the number of pixels). This type of interconnections is suitable for addressing in a feedback loop a 2-D array of nonlinear optical elements such as a NOR gate plane (or, in general, a spatial light modulator, SLM, with non linear input-output characteristic) therely realizing a N x N "massively parallel" array of 1-bit cellular processors (the SLM and feedback loop are not shown in figure 1 for simplicity).

III - DISCUSSION

In spite of its potential advantages, the setup described above suffers a number of limitations inherent to holographic optical elements, some of which have already been stressed in the litterature [3 - 5]. We concentrate here on two aspects : on the one hand, hologram H_1, located in the object plane, must produce in its first diffracted order s beams of equal brightness, and nevertheless no intermodulation

terms can be tolerated. On the other hand, geometrical distorsion in the image plane is a problem :

a) We have found that the first difficulty can be solved simply by successive, rather than simultaneous, exposure of the s beams of hologram H_1. The diffraction efficiency in the absence of all intermodulation is theoretically limited to $1/s$. We have reached 70 % of that value for s = 10 with a relative efficiency dispersion of less than 10 % and virtually no parasitic diffracted light.

b) Distorsion due to hologram H_2 must be traded off with diffraction efficiency. Our holograms H_1 and H_2 are recorded on dichromated Kodak 649 F plates. The Bragg angle is choosen so as to concentrate the incident energy in the first diffraction order.

For a theoretical 100 % efficiency, implying good rejection of all higher-orders, it can be shown that a 15 μm thick gelatin plate must have a Bragg angle of more than 20°. However, distorsion increases with Bragg angle. If no solution to this problem is found, distorsion can have a dramatic effect on the processing : poor superposition of the pixels imaged through the feedback loop on the 2-D nonlinear optical element can lead to faulty results. To avoid this implies a minimum tolerable pixel size and spacing ; this minimum increases with the length of the translation vector $\vec{T}(m,n)$ and can easily become quite large (several mm for example).

To keep the advantage of massive parallelism, distorsion must therefore be compensated for. To this end, we propose to follow the principle sketched on figure 3 : two holographic double diffraction setups are cascaded ;
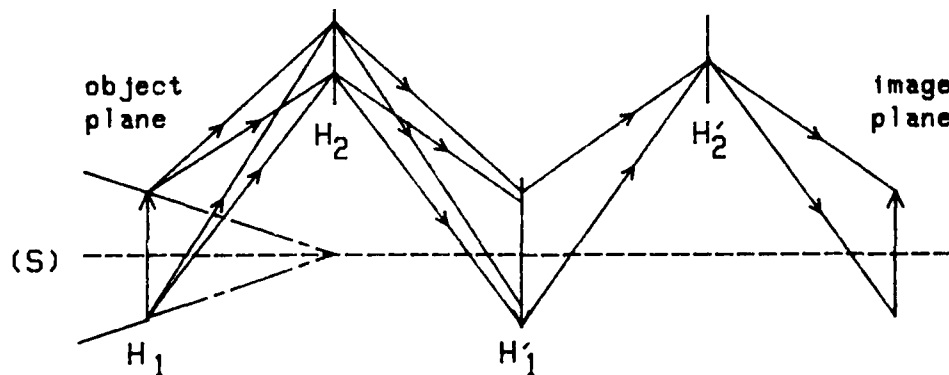
Figure 3 : Two holographic double diffraction setups are cascaded to compensate distorsion.

the second H_2 hologram, labelled H_2' , is in fact a single holographic lens, with a position and a Bragg angle matched to the average characteristics of the s holograms in H_2. We hope to obtain satisfactory results from such a setup with at least 1000 pixels and s larger than 10 for a 1 cm^2 object field ; distorsion error is then reduced to less than 100 μm.

IV - CONCLUSION

Holographic connectors can be very useful for implementing cellular optical processors taking advantage of the potential advantages of optical processing. However, further investigations on holographic elements in various architectures is needed before the expected performance is reached.

Applications of the setup described in the present communication with two interconnection holograms cover a wide range of operations and algorithms on two-dimensional binary or analog data. The connections may be binary or analog in nature, and they can be modulated by an on-off shutter as shown in figure 2 or by a grey-level programmable mask. Insertion of a SLM as the nonlinear processing element is influential on the nature of data processed. For example, a thresholding SLM provides binary output from a binary or analog input : a NOR-gate array SLM allows to realize an array of cellular automata ; even then, the connections may be binary or analog, each NOR-gate operating as a threshold on an analog sum of its inputs.

References

1. B.K. Jenkins et al., Appl. Opt., 23, 3455 (1984)
2. B.K. Jenkins et al., Appl. Opt., 23, 3465 (1984)
3. K.M. Johnson, L. Hesselink and J.W. Goodman, Appl. Opt., 23, 218 (1984)
4. R.K. Kostuk, J.W. Goodman and L. Hesselink, Appl. Opt., 24, 2851 (1985)
5. J.W. Goodman, Opt. Acta, 32, 1489 (1985).

Summary:

Parallel Interfacing
of Integrated Optics with Free-Space Optics

Adolf W. Lohmann

Physikalisches Institut der Universität

8520 Erlangen, FRG

There exist two distinct optical technologies with potential for data processing: integrated optics (IO) and free-space optics (FSO). The IO technology is well suited for performing optical nonlinear interactions and electro-optical control operations. As a planar technology, IO can implement one-dimensional parallelism quite naturally, but not so well two-dimensional parallelism. The FSO technology on the other hand, is very well suited for two-dimensional parallelism. But in terms of nonlinear interactions FSO is lagging behind IO.

Expressed in other terms, to confine the light in waveguides is good for interactions but not so good for data transport due to the planar topology. Without confinement, in free space, it is the other way around. Due to this supplementary situation it is desirable to use both technologies in an optical parallel processor, where many light interactions but also many light transport operations have to be performed. For using both technologies together one needs interfaces.

The design of IO/FSO interfaces is a problem of overcoming the mismatch in dimensionalities. To arrange IO devices in two dimensional fashion we propose to put IO chips on top of each other in staircase fashion. The wave guides emit light into (or receive from) the free space at the edges of the stair case. The lateral extent of a two-dimensional array of light emitting wave guide endings may be quite ordinary compared to the usual flat objects of FSO setups. But the large depth of the staircase requires careful attention in the design of the FSO setup.

A study of scattering from small structures for optical beam shaping and interconnects

M.T. Lightbody and M.A. Fiddy
Physics Department, King's College London, Strand, London WC2R 2LS, UK

Introduction

Optics is attractive for computing because of its distinct advantages over electronics, in particular the inherent non-interaction of multiple beams passing through, or near, each other and the potentially massive parallelism. In order to compete with electronics these two properties should be fully exploited. Therefore it is of importance to consider what fundamental limitations will apply to the density of optical sources, or secondary sources, in terms of the acceptable levels of crosstalk that will occur in an adjacent array of detecting elements. In order to beat electronics the dimensions of each decision making plane must be kept as small as possible. The physical consequences of this necessity must be well understood before such questions as optimal architectural structure can be addressed.

We consider here a very simple geometry consisting of two adjacent arrays of elements or pixels, which may, for example, be computer generated optical elements. There is much talk of the advantages, in a variety of optical computing architectures, of being able to direct light beams from individual sources to individual detectors [1], [2]. A computer generated optical element, or a coherent optical holographic element copied from one, offers a potentially compact and efficient means for doing this. Of course, it would be desirable to have an adaptive element which will switch from specific sources to specific detectors on command. As our understanding of various materials such as photorefractives improves, reprogrammable holographic elements should play a key role in reconfiguring interconnect patterns but their optimal dimensions will be limited by the same problems as those described below for a fixed element. Similarly, one may consider using a bank of holographic elements to map beams, which are used in conjunction with a single real-time mask which selects the appropriate interconnect pattern. [3].

Beam profiles

In the case of relatively thin (on the scale of the wavelength) computer generated masks, which have encoded structures that are relatively large, a simple interaction model, based on Kirchoff or physical optics theory, is valid. Very crudely approximating a single element as a square aperture of dimensions a x a, the emerging beam pattern will have a sinc profile whose main lobe has a half width tending to $\lambda L/a$ as the wave propagates a distance L towards the far field. We could assume that beyond a hundred wavelengths or so this may be a good estimate for the beam width. Clearly there is considerable energy in the side lobes which will combine to fall on elements adjacent to the specified one. Roughly speaking, however, if a $\sim 10\lambda$ or 5μm then the main lobe of the pattern is also $\sim$ 5μm at the adjacent plane.

It is possible that this is in some sense an optimal geometry but it is interesting to speculate how the cell or element dimensions in each plane could be reduced. One could envisage using a small array of elements to specifically beam-shape the output but the overall dimensional gains are not obvious. We are also aware of the fundamental limitations on fan-in and fan-out of interconnections arising from the constant radiance theorem [4]. This states that the product of the cross-sectional area and the square of the numerical aperture of an optical beam (in the geometrical optics limit) must remain constant under any lossless linear transformation of that beam. The co-ordinates on the second plane are given by $\alpha = L\sin\theta$ where $\sin\theta$ is the numerical aperture [5]. The consequences of this are important for fan-in, as occurs when several elements must simultaneously address one element in the adjacent decision plane. Thus the beam pattern of the receiving element (viz. its numerical aperture) or its cross-sectional area must be able to be made correspondingly larger to avoid power loss.

Diffraction by small apertures

In making a computer generated element, for example, with an e-beam plotter, one may achieve written structures having a scale considerably smaller than the wavelength or sampled representations of wavelength scale structures. The question arises whether an array of sub-

wavelength structures can be exploited to beam shape while retaining or reducing the overall dimensions of an effective element. Of importance with structures having these lateral dimensions is the fact that their thickness (typically 5 to 30μm emulsion thickness) becomes significant and the Kirchoff approximation no longer holds.

The diffraction of an electromagnetic wave by an aperture in a thick screen, whose linear dimensions are of the order of a wavelength has been studied by only a few groups [6], [7], [8]. We wished to establish whether there are any fundamental limitations on the use of subwavelength three dimensional structures for shaping the outgoing beam pattern in both the near and far field. It is known from the inverse problem, namely the super resolution of subwavelength structure, that evanescent fields make an important contribution to the scattering even though they decay exponentially[9]. For example, the sinc beam width referred to earlier can be bettered by a two-dimensional array of subwavelength-spaced point sources, at the expense of throwing more power into the subsiduary sidelobes. However, the authors' study of the corresponding 3-D direct problem has led to the concept of a thickness dependent mode transfer function that will allow only certain specified spatial frequencies to propagate through the optical element unattenuated. Therefore one can envisage a situation where a series of sub-wavelength spaced 3-D secondary sources, ultimately computer generated in real time, could be utilised in order to generate the necessary reduced beamwidth whilst at the same time suppressing the unwanted sidelobes by selective filtering. Such structures could perhaps then be fabricated as a solid compact unit.

Acknowledgement

References

[1]   Applied Optics 25 (1986) issue number 18.

[2]   H.M. Berenyi and M.A. Fiddy, Application of homometric sets for beam manipulation, J.O.S.A. A3 (1986) 373-375.

[3]   J.W. Goodman, F.J. Leonberger, S-Y Kung and R.A. Athale, Optical interconnections for VLSI systems, Proc. IEEE 72 (1984) 850-866.

[4]   J.W. Goodman, Fan-in and fan-out with optical interconnections, Opt. Acta 32 (1985) 1489-1496.

[5]   F.T.S. Yu, Optical Information Processing, John Wiley & Sons, (1983), chapter 3.

[6]   S.C. Kashyap and M.A.K. Hamid, Diffraction characteristics of a slit in a thick conducting screen, IEEE AP-19 (1971) 499-507.

[7]   K. Hongo and G. Ishii, Diffraction of an electromagnetic plane wave by a thick slit, IEEE AP-26 (1978) 494-499.

[8]   O.M. Medez, M. Cadilhac and R. Petit, Diffraction of a two-dimensional electromagnetic beam wave by a thick slit pierced in a perfectly conducting screen, J.O.S.A. 13 (1983) 328-331.

[9]   H.G. Schmidt-Weinmar, Superresolution with coherent light, Proceedings of the International Conference on Lasers '79, December 17-21 (1979), 798-806.

ENGINEERING LIMITS FOR OPTICAL INTERCONNECTIONS

by

Davis H. Hartman
Bell Communications Research
435 South Street
Morristown, New Jersey

The all-optical computer has been lauded by many as offering the potential for increasing computation power over electronic computers by many orders of magnitude. However, to date, attempts to develop such a machine have been limited by technology and by the electronics these computers must interface with. An alternative approach, that will aid in the development of the needed technology for optical computers is the hybrid electrical/optical computer. In such a machine, electronics is used where it functions best, i.e., as a switching mechanism, and photonics is used where it functions best, i.e., as an interconnect medium.

Optical interconnections can be used to allay the deleterious effects of signal distortion, cross-talk, package parasitics and frequency dependant loss caused by conventional microstrip type interconnection media. When system complexity (i.e. parallelism) or board size are combined with high switching speed requirements, conventional metalic interconnection media face all of these deleterious effects simultaneously. This condition is depicted in figure 1 for data distribution and in figure 2 for clock distribution. Figure 1 plots system switching rate vs. system parallelism (e.g. eight parallel 100 Mb/s channels compose a system with M=8, B=100 Mb/s, with high data throughput, Q, is 800 Mb/s). Figure 2 plots clock rate vs. board size, in centimeters.

There are many optical interconnection genre for computer applications. These genre range in distance from 100 feet down to fractions of inches, and from fiber optic media to reconfigurable holographic interconnects. However, common to all these genre is the need for high sensitivity, high speed, high dynamic range receivers.

Driven by the need for packaged O-E transducer sub-miniaturization within the physical shapes typical of computers, many of the steps commonly taken to yield high sensitivity (pulse shaping, filtering, equalization, clock recovery) are not feasible. Monolithic optoelectronic design becomes vital particularly of receiver interfaces, tempered by the requirement for ultra-high reliability.

This paper discusses the design for optical interconnection applications. In particular, a transmitter receiver analysis package, that numer-

ically solves the link equations, is described. All of the limitations associated with the requirement for package sub-miniaturization are included, as well as effects of laser diode noise on the link performance. Link power budgets are derived for laser diode links as well as LED based links. Fundamental engineering limits to attainable data fanout, link switch rate and circuit packing density are given.
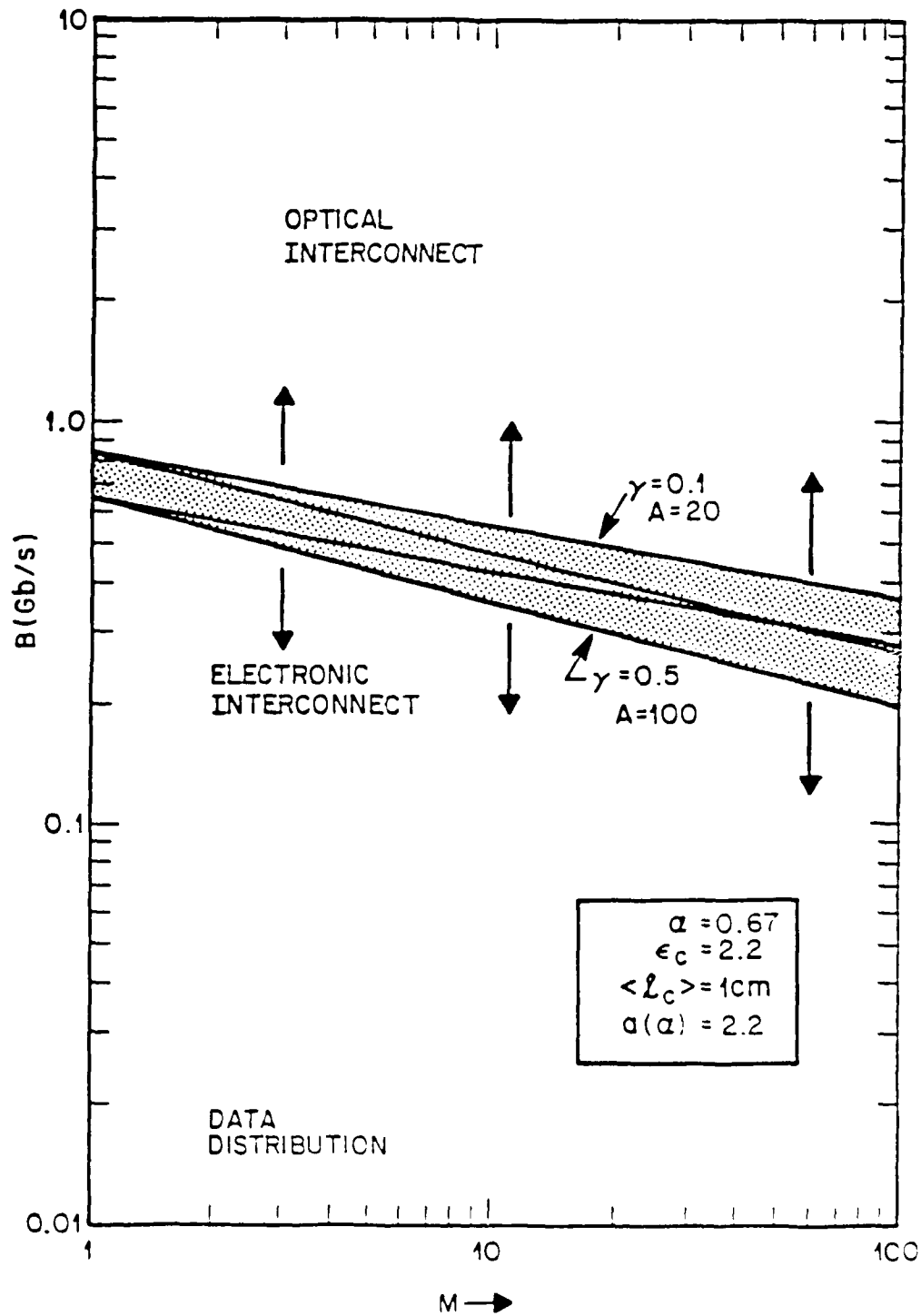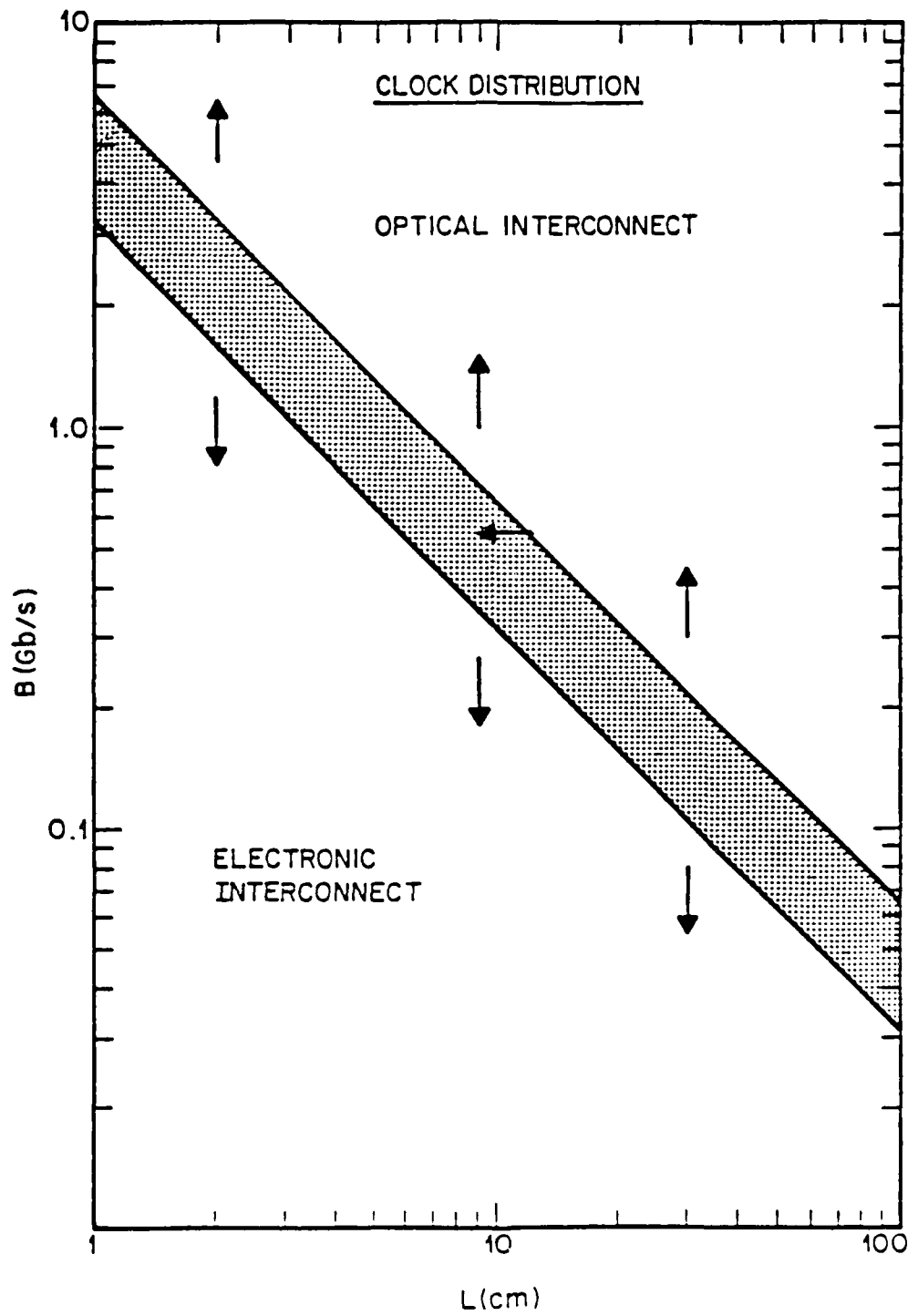
FIGURE 1

FIGURE 2

A Comparison of Encoding Schemes for E-beam
Fabrication of Computer Generated Holograms

H. Farhoosh, M. R. Feldman, S. H. Lee, C. C. Guest, and Y. Fainman
Department of Electrical Engineering and Computer Science
University of California, San Diego
La Jolla, CA 92093

I - INTRODUCTION :

Motivated by many attractive features and capabilities of electron beam lithography systems, which include large space band-width product (SBWP), direct writing at submicron resolution, and small distortion errors, we carried out an investigation on the suitability of various encoding methods of computer generated holograms (CGH) for e-beam fabrication.

E-Beam fabrication of CGH has been investigated in depth by S. M. Arnold [1,2], and used in a few applications by others [3-6]. The purpose of this paper is to systematically evaluate the suitability of various encoding methods for e-beam recording of CGHs.

For most types of CGH fabrication procedures, the SBWP is limited by the capabilities of the recording device. A general comparison of CGH encoding methods have been performed by many authors [7-8]. These comparisons are inherently based on the limited SBWP of the recording device. However, when e-beam lithography is employed, the SBWP of the hologram is limited, not by the recording device, but by computer memory, computation time, and data storage capabilities. These constraints impose a different basis of comparison between encoding methods. The evaluation presented in this paper is based on the ability of encoding methods to achieve high quality holograms (e.g. high diffraction efficiency and signal-to-noise ratio) while subjected to above limitations. In this way the suitability of these methods for e-beam recording of CGH can be compared.

II. EVALUATION CRITERIA :

In this section we shall present and discuss a set of criteria according to which the encoding schemes can be evaluated. These criteria are divided into two groups as follows :
A. Hologram Qualities: which include 1) the size and bandwidth of the reconstructed wave, 2) signal-to-noise ratio (SNR), and 3) diffraction efficiency.
B. Computer Limitations: which include 1) computation toll, and 2) amount of graphical data.

The first group of criteria determines the quality of the CGH, while the second group are constraints imposed by limited capability of digital computers and data storage media. These constraints determine practical limitations in hologram synthesis.
A. Hologram Qualities

1) The size and bandwidth of the reconstructed wave

Parameters that determine the quality of the reconstructed wavefront are hologram size(X', Y'), and the bandwidth of the reconstructed wavefront at the hologram plane (BW'_x, BW'_y). For Fourier transform holograms, these parameters determine the size (X, Y) and bandwidth (BW_x, BW_y) of the reconstructed image according to

$$X = \lambda F BW'_x, \qquad Y = \lambda F BW'_y \tag{1a}$$

$$BW_x = \frac{X'}{\lambda F} \qquad BW_y = \frac{Y'}{\lambda F} \tag{1b}$$

where F is the focal length of the Fourier transform lens and λ is the wavelength of the light. It follows that

$$SBWP = BW'_x BW'_y X' Y' = BW_x BW_y XY = N_o^2 \tag{2}$$

where N_o^2 is the SBWP of the image or, in other words, the number of resolution elements in the reconstructed image, provided that the Nyquist sampling is observed. The SBWP of the CGH, defined as the number of wavefront samples represented by the hologram, can be larger than N_o^2 (e.g. due to the addition

of a carrier wave or by interpolating the complex wavefront resulting from the FFT operation). If an interpolation scheme is used, the SBWP of the CGH (denoted by $N_x N_y$) would become

$$N_x N_y = M_x M_y N_0^2. \tag{3}$$

where M_x and M_y are interpolation factors that relate the number of points in the hologram plane to the number of points in the image plane. In this case the size of the hologram can be written as

$$X' = \delta x' M_x N_o \tag{4}$$

where $\delta x'$ is the CGH sampling period.

Although the CGH SBWP is determined primarily by the computing resources employed, the maximum bandwidth will vary with the encoding method. The maximum bandwidth, obtained by setting $\delta x'$ in Eq.(4) to its minimum value, depend on the number of amplitude and phase quantization levels (n_a, n_p) used by the encoding method.

2. SNR :

The definition we assume for the SNR is

$$SNR = \frac{<intensity\ in\ the\ desired\ image>}{<error\ intensity>} = \frac{\frac{1}{XY} \int \int |g(x,y)|^2 dx dy}{\frac{1}{XY} \int \int |h(x,y)-g(x,y)|^2 dx dy} \tag{5}$$

where g(x,y) is the desired image obtained if a hologram had recorded the complex wavefront in an ideal manner (without any errors), and h(x,y) is the actual reconstructed image.

The SNR as well as the diffraction efficiency of a CGH are reduced by the amount of error introduced in various stages of hologram production. Five sources of error are considered here. These errors are i) Sampling errors (spatial discretization), ii) Errors due to the finite size of the hologram, iii) Quantization error introduced because of digital representation of an analog function (modulation discretization), iv) Representation related errors, v) Distortion errors caused by the non-ideal behavior of the recording device. The first three errors in this list are inherent to all CGH. The other two depend on the encoding scheme and the recording device, respectively.

Assuming that sampling is done properly (according to the Nyquist criterion), the first two sources of error do not have any significant effect on the SNR. Using Eq. (5) we have evaluated the SNR due to quantization, representation, and distortion, the results of which are tabulated in Table 1.

3. Diffraction efficiency

Diffraction efficiency, η is defined as:

$$\eta = \frac{P_s}{P_{inc}} \tag{6}$$

where P_s is the total light power in the reconstructed image in the presense of error and P_{inc} is the light power incident on the hologram. The diffraction efficiency of a hologram is determined by the manner in which the object wavefront is encoded and it is affected by errors mentioned above. Although different sources of error have different effects on the reconstructed image, in general the presence of error in hologram encoding reduces the power diffracted into the desired image. Therefore, the actual diffraction efficiency can be written as

$$\eta = r \eta_t \tag{7}$$

where η_t is the theoretical diffraction efficiency in the absence of any error except the representation related error (η_t depends on the encoding method), and r is the power reduction factor introduced by other sources of error. The results of our evaluation of the diffraction efficiency are included in Table 1.

B. Computer Limitations

1. Computation toll:

In theory a typical e-beam system is capable of writing a CGH with a SBWP of over 10^{10}. Yet if a computer such as a multi-user VAX is employed, the largest two dimensional FFT that can be performed

in a reasonable amount of time is about 2048 by 2048 $= 4 \times 10^6$ points. When considering the computation limitations, it is noted that there are basically two classes of CGH. The first class of CGH includes wavefronts that are known in analytic form, e.g. holographic optical elements (HOE's) for optical testing. Wavefronts of the second class are obtained by sampling an image and then computing a Fresnel or Fourier transform. Since no FFT is necessary for computing the wavefronts of the first class (the analytic wavefront need only be sampled), the SBWP of this type of hologram can be larger than that of the second class.

2. Size of graphical data

Another constraint on the encoded wavefront is the amount of graphical data needed to generate the hologram. It is generally desired to have hologram patterns that generate the least possible number of primitive shapes because the amount of graphical data representing the pattern is directly proportional to the number of primitive shapes (NPS) that comprise the hologram pattern. Orientation of these patterns is important as well, e.g. pattens consisting of rectangular apertures oriented along the x and y axes generate much less graphical data than patterns consisting of curved lines in arbitrary directions.

For all of the conventional encoding schemes the NPS is given by

$$NPS = (number\ of\ hologram\ cells)\ N_c \tag{8}$$

where N_c is the number of shapes per cell. The NPS for each encoding method is presented in Table 1.

III. COMPARISON OF ENCODING SCHEMES

Using the criteria of the previous section, nine different CGH encoding methods are compared. Table 1 shows the results of this comparison and it will be discussed in the conference.

IV. CONCLUSIONS

Electron-beam lithography systems appear to be a prime candidate for recording of computer generated holograms because of their ability to record patterns of submicron resolution on large substrates. However, there are three obstecles in the utilization of their full potential. These are, 1) Limited processing power of commonly available computers, 2) Large amount of graphical data needed to specify geometrical patterns, and 3) High cost of fabrication.

In some CGH applications, e.g. HOE's for aspheric testing, the hologram function is known in analytic form, and the computer processing power is less of a problem. For other types of CGH dedicated hardware, such as array processors, can provide a partial solution to this problem.

The size of graphical data is a problem that can be alleviated by means of efficient graphical coding. A CAD system solves this problem significantly. Undoubtedly other, more efficient, graphical coding methods can also be devised.

The high cost of e-beam fabrication of CGH can perhaps be justified by the quality of the hologram. Also an increase in the production volume would bring the cost down.

Finally, we would like to point out that the comparisons that were carried out in the previous section can serve as a guide to choosing of the most suitable encoding scheme for a particular application. Different applications of CGH impose different requirements on the quality of the CGH. In general it is desirable to have high diffraction efficiency, high SNR, and large size images with high resolution for all types of applications. However, as it can be seen from the Table 1, achievement of all of these requirements at once is not possible, and one should be willing to compromise one requirement for another. For example if the application of a CGH is in pattern recognition, one is concerned with the noise in the correlation pattern. Therefore, it is desired to have a high SNR. From Table 1 it can be seen that Burch's method would be a good candidate for this type of application. When the CGH pattern is known in analytic form, such as in HOE's, Arnold's method generates the least NPS and high enough SNR to be considered suitable for this type of application.

REFERENCES :
1. S. M. Arnold, "Electron Beam Fabrication of Computer Generated Holograms," Opt. Eng., vol. 24, 803, (1985).

2. S. M. Arnold, "E-Beam Written Computer Generated Holograms," Final Technical Report NO. E47069, Honeywell Corporation Technology Center, Bloomington, Minnesota, (1983).

3. K. M. Leung, J. C. Lindquist, and L. T. Shepherd, "E-beam Computer- Generated Holograms For Aspheric Testing," Proc. of SPIE, vol. 215, 70, (1980).

4. J. L. Freyer, R. J. Perlmutter, and J. W. Goodman, "Digital Holography : Algorithms, E-Beam Lithography, and 3-D display," Proc. of SPIE, vol. 437, 38, (1983).

5. R. A. Athale, C. L. Giles, and J. A. Blodgett, "Use of E-beam Written CGH in Optical Pattern Recognition," Proc. of SPIE, vol. 437, 48, (1983).

6. M. R. Feldman and C. C. Guest, "Automated Design of Holographic Optical Elements for Interconnection of Electronic Circuits," OSA annual meeting, Seattle, Washington, Oct. 19-25, 1986.

7. R. A. Gabel, "Reconstruction Errors in Computer Generated Binary Holograms : A Comparative Study," App. Opt., vol. 14, 2252, (1975).

8. R. A. Gabel and B. Liu, "Minimization of Reconstruction Errors in Computer Generated Binary Holograms,"App. Opt., vol. 9, 1180, (1970).

$M_x M_y = 8$, $N_s = 1024$										
Encoding Method	η (%)	NPS $\times 10^6$	BW' 1/mm	BW' 1/mm	SNR				Hologram size (mm)	
					Q	R	D	TOT	X'	Y'
Lohmann $n_p = 8$ $M_x = M_y = \sqrt{8}$	9.4	8.4	147	177	15	2.4	75	2.0	7.0	5.8
Lee(1970) with interp. $M_x = 4, M_y = 2$	4.7	4.2	500	125	128	221	53	32	2.0	8.2
Lee(1970) w/out interp. $M_x = M_y = \sqrt{8}$	4.8	17	177	177	32	2.4	53	2.1	5.8	5.8
Lee(1979) $M_x = 8, M_y = 1$	7.3	1	417	500	15	14	57	6.4	2.5	2.0
Burkhardt $M_x = 2.99$ $M_y = 2.67$	5.4	5.6	667	137	32	247	30	14.5	1.5	5.5
Burch $M_x = 4, M_y = 2$	4.7	8.4	167	333	324	101	243	58.5	6.1	3.1
$M_x M_y N_a^2 = (4096)^2 = 1.7 \times 10^7$										
Arnold (staircase) $M_x = 14$ $M_y = 10$	9.6	8.5	238	200	60	122	232	34	1.4	1.7
Lee(1974) $M_x = 14$ $M_y = 2$	9.0	1.2	238	1000	60	-	51	27	3.3	0.78
Kinoform $M_x = M_y = 2$	8.8	17	1000	1000	1.7	-	31	1.6	2.1	2.1

Table 1. Comparison of 9 encoding schemes in terms of diffraction efficiency, SNR, bandwidth, NPS, and area of the hologram. It is assumed that the first six methods are used to encode CGH's requiring FFT operations of 1024x1024. For the last three methods an analytic form for the CGH is assumed. Q, R, and D refer to the quantization, representation, and distortion errors, respectively.

MONDAY, MARCH 16, 1987

PROSPECTOR / RUBICON ROOM

3:00 PM–5:30 PM

MD1–9

SESSION 4

Ravindra A. Athale, BDM Corporation, *Presider*

Optical Computer Architecture: What is the Ideal?

By W. Daniel Hillis

Abstract

We shall define the ideal computer as one which can execute any calculation as fast as any other computer, within a multiplicative constant.

Summary

Our current conception of what a computing machine is, is distorted by the form of existing computers. These current forms are largely a consequence of the limitations of available electronic switching and memory components, rather than of the computational requirements. Optical components will no doubt offer their own set of compromises, but the standard against which they should be measured is not their ability to reimplement the architecture of electronic machines, but rather their ability to implement an "ideal" computer whose form is determined by the requirements of the computations, and not by the components.

So what is the ideal? We know that in some sense all computer architec-

tures are equally powerful. Turing universality assures us that a machine with a few states and a very long tape can do anything that can be done with the most sophisticated supercomputers. Yet this form of universality ignores an important distinction: the time complexity of the algorithm. Multiplying two n-bit numbers, for example, may require n^2 steps on one machine, n on another and a single step on a yet different type of machine. This advantage of one type of machine over another cannot be compensated for, by measuring decreasing time required for a single operation. since for some size of number. n^2 operations will take longer than n operation, no matter what the relative time per operation. This difference in time-complexity gives us a standard by which we can define our "ideal" computer architecture.

We shall define the ideal computer as one which can execute any calculation as fast as any other computer, within a multiplicative constant. In other words. it can be made at least as fast as any other computer by simply adjusting the time per operation. We shall restrict ourselves, for the moment. to considering only those computers which can be, in principle. defined in terms of standard switching functions such as logic gates and

memory elements. This avoids comparisons with hypothetical computers that contain, for instance. oracles that can predict the future. Even in this limited sense of an ideal, von Neumann machines do not reach the mark, since they can be surpassed by many types of parallel machines. These parallel machiness can in turn often surpass one another for particular tasks.

One model of a machine that cannot be surpassed by more than a constant factor, and that is therefore ideal in the sense defined above, is an architecture similar to the Connection Machine, but with an infinite number of processors and a perfect communication system. Since this machine can simulate any other machine that can be built of gates and memory elements in constant time, it is guaranteed to have the desired property. It is therefore a plausible choice for an ideal architecture which a new component technology should strive to implement. It is. of course. not the only such choice.

It may be that the restriction to machines that can be implemented in terms of gates and memory is too narrow. For example. a machine with a true random number generation is strictly more powerful than one without. It is possible that optical computers can be constructed to go beyond

the confines of turing universality. Even if they do not, the architects of optical machines should feel no compulsion to squeeze their designs into the Procrustean couch of conventional computer architecture.

Globally Folding Combinatorial Logic Cells in Digital Optical Systolic Computing Arrays

P.S. Guilfoyle and W.J. Wiley
OptiComp Corporation
PO Box 10779
Zephyr Cove, Lake Tahoe
Nevada 89448

This paper is the third in series of publications describing various alternative combinatorial logic based optical computing architectures.[1,2] Within the first paper titled "Combinatorial Logic Based Optical Computing," justification for combinatorial logic is initially debated through the coupled use of **extensive** optical interconnects with the natural "and-or-invert" capability of most every optical system. Figure 1 depicts the interconnect concept. Optical systems are capable of connecting points between planes in any 3 dimensional configuration by using, for example, Fourier transform holography, global fibers, or even simple lenses, depending on the interconnect complexity desired. The ability for an optical system to interconnect in three dimensions, is, in the opinion of the authors, the absolute greatest asset of an optical computer. As shown in figure 1 and explained in more detail in reference 1, should an optical computer completely exploit it's fully global interconnect capability between a set of spatial light modulators, where each has a space bandwidth product of 256 by 256, then the total interconnect gate density reaches 4×10^9. The problem plaguing silicon integrated circuit designers is the *inability* to interconnect various processing elements. This inability limits the chip's ultimate performance in terms of operations per square centimeter of silicon.

Where silicon fails, optics prevails. Optical systems need not obey the "nearest neighbor interconnect" law. In addition to the massive interconnect capability of optics, the laws of physics have also granted an inherent "and-or-invert" capability. This capability, described as well in more detail in reference 1, has been and continues to be a fundamental digital logic primitive with which most circuits are designed. A wide range of digital optical architectures capable of performing numerous, if not myriad, sets of digital functions other than mathmatical, are possible when the designer starts with the "and-or-invert" digital logic primitive.

Figure 1: 3-D Globally interconnected optical gates

Input plane 1
Input plane 2
Output plane

In reference 1 many examples are shown which exploit both of these inherent features. Both are coupled and modeled as an array of parallel programmable logic arrays (PLA) of *sum-of-products* (ORs of ANDs). With such a capability, an array of sequential logic functions can be ultimately realized. From there, as described by Brayton[3], "Sequential logic functions can be represented as Finite State Machines (FSMs) and implemented by a combinational and a storage component[4]. In particular, PLA based Finite State Machines can be designed efficiently, because the properties of two level combinational functions are well understood. PLAs and memory elements can be seen as primitives of a general digital design methodology." Within the construct of two level combinational functions, the original paper[1] describes a simple example of an optical "text" processor, a word and phrase comparitor, which can be used for massive text look-up and search. The first step of the combinatorial process is, in general, developed outside of the optical regime using silicon devices, for this level is needed only once. The first level could be generated optically but perhaps not as efficiently with respect to the second level. Once the Boolean combinational terms are generated the second level interactions are computed several hundreds of times or more using the optical PLA systolic arrays. The original paper continues to describe a digital optical full adder enhanced by the full global broadcast capability of optics. Finally, a 2 x 2 bit systolic multiplier is described for matrix processing.

The second paper[2] follows the original by describing a 3 x 3 bit systolic multiplier array based on combinatorial logic such that the outputs of each multiplication region is a full 6 bit binary weighted answer. First the Boolean

Table 1: Karnough equations for 2x2 multiplication:

$$O_1 = A_0B_0$$
$$\text{ch1}$$

$$O_2 = A_0A_1B_1 + A_0B_0B_1 + A_0A_1B_0$$
$$\text{ch2} \qquad \text{ch3} \qquad \text{ch4}$$

$$+ A_1B_0B_1$$
$$\text{ch5}$$

$$O_3 = A_1B_0B_1 + A_0A_1B_1$$
$$\text{ch6} \qquad \text{ch7}$$

$$O_4 = A_0A_1B_0B_1$$
$$\text{ch8}$$

Table 2: Drive channel assignment of combinatorial terms:

Combinations	Drive "A"	Drive Ch.	"B"	Drive Ch.
1	A_0	1,3	B_0	1,4
2	A_1	5,6	B_1	2,7
3	A_0A_1	2	B_0B_1	5
4	$\overline{A_0}A_1$	4,7	B_0B_1	3,6
5	A_0A_1	8	B_0B_1	6

expressions are derived. The same method of division of the combinatorial terms is then shown, followed by a parallel optical implementation. A full global broadcast methodology is then described.

This paper reviews the 2 x 2 bit combinatorial multiplier and subsequently shows the planar systolic global interconnect topology. This topology is then *folded* into three dimensions thus reducing the number of input ports or *pins*, and consequently exploiting the three dimensional interconnect capability of optics. A discussion follows with respect to the extent, i.e., the number of bits, to which this global folding can be efficiently implemented for systolic multiplication arrays. Finally, the hardware implementation is shown.

Global Folding of 2 x 2 Bit Combinatorial Multiplication

As explained in the first reference, 2 x 2 bit multiplication can be reduced to 4 equations, one for each binary weighted desired resultant bit. These equations are shown for review in Table 1. Notice that although 8 Boolean products are required only 5 Boolean first level combinations are required as shown in Table 2. Figure 2 depicts the "parallel only" optical implementation where two 8-channel acousto-optic devices are used to a.) multiply the two sets of 5 Boolean expressions, accordingly, and b.) generate the AND products. The subsequent focusing by the output optics generates the appropriate "fan-in" for each respective OR gate detector. Each detector must only operate as a threshold device for light or no light rather than sum or suffer from the need to detect at an intermediate threshold point as in optical threshold logic schemes. In figure 3, the 8 inputs are reduced to 5 for the second phase of the combinatorial multiplication. Here, only 5 channels are used in each acousto-optic spatial light modulator. However, notice the interconnect is no longer parallel as in figure 2. Rather, the interconnect is "global". Any pixel in plane 1 may address any pixel in plane two, of course under the rules of the combinatorial interconnect, in this case the second level of the 2 x 2 multiplication.

Figure 2: 2 x 2 bit 3 by 1 parallel systolic multiplication vector
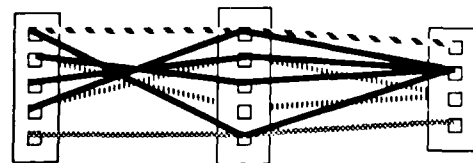
Figure 3: Planar Global interconnect configuration for each PLA plane of the 2 x 2 systolic multiplication vector of figure 2.

As shown in reference 2, however, the number of combinatorial terms grows rapidly for higher order word lengths. In particular for a 3 x 3 bit multiplier, a straight parallel implementation requires a total of 35 second level combinatorial sum of products, with a minimum number of 15 combinations used. Using a planar PLA optical systolic implementation as shown in reference 2, the straight parallel implementation requires a telecentrically imaged pair of 35-channel acousto-optic devices. Although this is certainly achievable with today's technology, the planar global interconnect implementation reduces this number to two 15-channel acousto-optic devices, a greater than 2 to 1 reduction.

These numbers grow dramatically as higher order word lengths are desired. The following equation calculates the maximum number of combinations required for level one combination generation for n x n bit multiplication:

$$C = \sum_{i=1}^{n} \binom{n}{i} (2^i - 1)$$

Table 3 shows the result of this equation for n x n multiplication up to 12 bits. Notice that for 3 x 3 bit multiplication, the equation yields 19 combinational terms, although in reference 2 we were capable of reducing this number to 15, with great difficulty. In general, the number of combinations shown should act as a maximum, although further reduction is possible.

Table 3: Combination table for n x n bit multiplication

number of bits	number of combinations	Square root
1	1	1.00
2	5	2.24
3	19	4.36
4	65	8.06
5	211	14.53
6	665	25.79
7	2059	45.38
8	6305	79.40
9	19171	138.46
10	58025	240.88
11	175099	418.45
12	527345	726.19

Assuming that the design was to be a planar global interconnect systolic PLA, then the number of channels required for a multiplication array would correspond to the numbers in the second column of table 2. Unfortunately beyond 4 to 5 bits the number of channels would become far to high to represent a realistic hardware design. For example, if an 8 x 8 multiplier was desired, a planar global topology would require two 6,305 multi-channel acousto-optic devices. The solution is to *fold* the problem into the three dimensions that optics affords.

Much work has been performed in the analog regime on *folded spectrum* signal processing.[5,6] The same type of concept may be applied here to reduce the number of channels. Figure 4 depicts the folding of the simple 2 x 2 bit multiplier. Rather than have 8 parallel channels, or 5 planar globally interconnected acousto-optic devices or SLM, only two channels are used. Even 1 channel would be sufficient if the user was willing to pay the price in the other dimension.
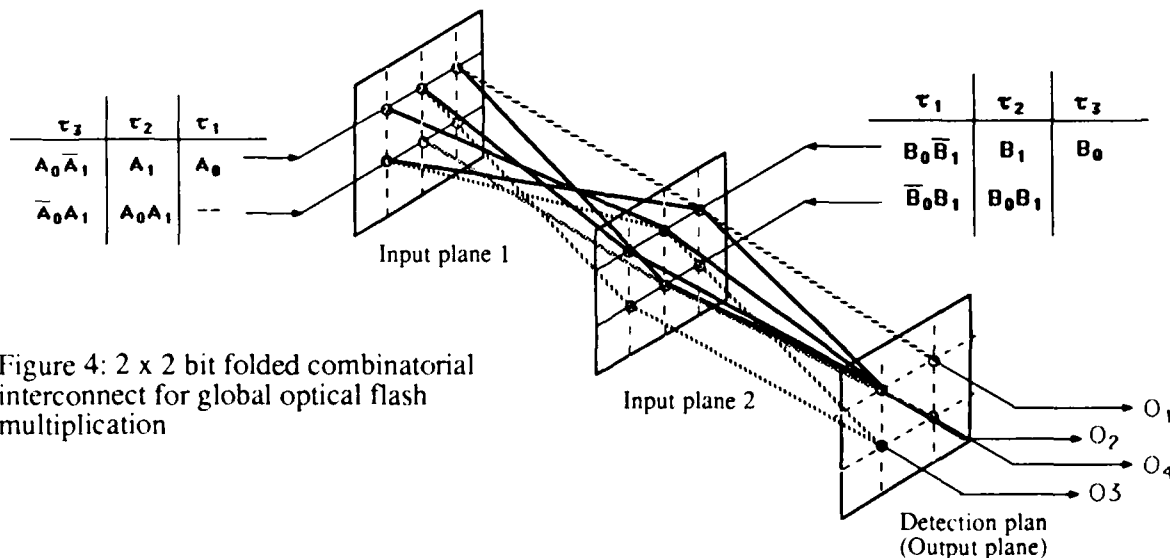
Figure 4: 2 x 2 bit folded combinatorial interconnect for global optical flash multiplication
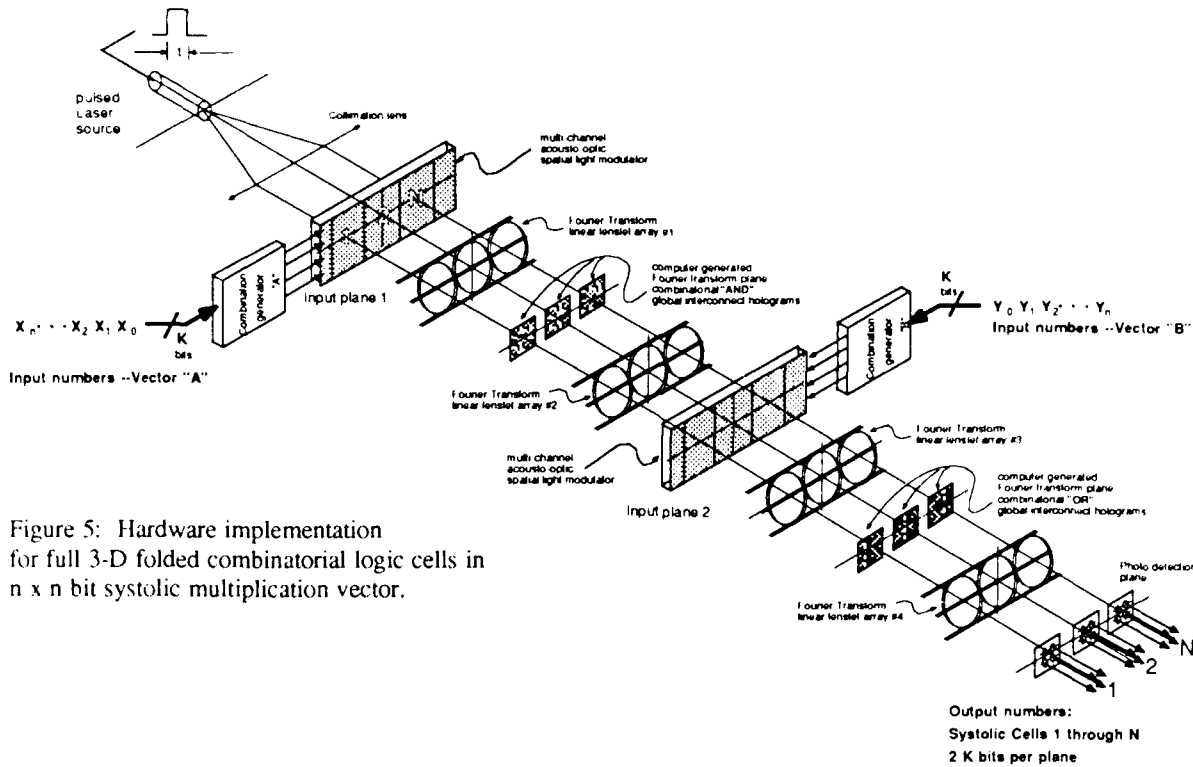
Figure 5: Hardware implementation
for full 3-D folded combinatorial logic cells in
n x n bit systolic multiplication vector.

As shown in figure 4, the 5 combinatorial terms are *time sequenced* into two telecentrically imaged acousto-optic devices. There the five combinatorial terms are sequenced in time with three clock cycles. After three clock cycles, the source pulses. The correct binary weighted answer is then at the detection plane. Notice that the interconnect is now 3-D global where the appropriate shading of figure 4 corresponds identically to the shading of the 2-D global interconnect of figure 3. Notice from table 3 that the 8 x 8 bit multiply would require only 45-channel devices given that the user desired to input a square array. A single channel device could be used if it had a time-bandwidth product of 6,305. This table also suggests that, for multiplications above 10 x 10 bits, the efficiency of global broadcast for these higher word length multiplications is far from the capability of optics. Multiplication is almost a "parallel" problem.

Finally, figure 5 shows some hardware advantages to optical global folding. Here two n-channel acousto-optic devices are driven from two combination generators. The combinations are thus used repeatedly through the acousto-optic device. The system shown is a length "N" systolic multiplication array, were the outputs are the correct binary weighted answers. However, rather than have one Fourier transform hologram performing the global interconnect for the entire array, N Fourier transform holograms are used for both the AND functional interconnect and the OR functional interconnect. This relaxes the requirements by N on each hologram, although 10^9 interconnects are certainly feasible with one hologram.

References

1. P.S. Guilfoyle, W.J. Wiley, "Digital Optical Linear 3 x 3 Bit Combinatorial Systolic Multiplication Array," PROCEEDINGS OF THE SPIE, Real Time Signal Processing IX, Vol. 698-30, August, 1986.

2. P.S. Guilfoyle, W.J. Wiley, "Combinatorial Logic Based Optical Computing," PROCEEDINGS OF THE SPIE, Vol. 639-17, April, 1986.

3. R.K. Brayton, G.D. Hachtel, C.T. McMullen, A.L. Sangiovanni-Vincentelli, Logic Minimization Algorithms for VLSI Synthesis, Kluwer Academic Pub., 1984.

4. F. Hill and G. Peterson, Introduction to Switching Theory and Logical Design, Wiley, 1981.

5. G. Lebreton, "Power Spectrum of Raster-Scanned Signals," Optical Acta, Vol. 29, No. 4, pp. 413-418, 1982.

6. T. Turpin, Proc. I.E.E.E., 61, 79, 1981.

RESIDUE POSITION-CODED LOOK-UP TABLE PROCESSING

A. P. Goutzoulis, D. K. Davies
Westinghouse R&D Center
1310 Beulah Rd., Pittsburgh, PA 15235

E. C. Malarkey, J. C. Bradley, P. R. Beaudet
Westinghouse Advanced Technology Division
P. O. Box 1521, Baltimore, MD 21203

1. **Introduction**. Residue arithmetic(1) has some very desirable features(2,3) which include: lack of carries, bounded input/output dynamic range and the ability to decompose a calculation into many parallel subcalculations of lesser complexity. Such features, when combined with high-speed position-coded optoelectronic look-up tables (LUT), result in high-speed, power efficient, low complexity processors.

2. **Residue Position-Coded LUTs**. The objective of a LUT is to create, in modulo m_i, the product (Figure 1) or sum of two input residue numbers X and Y without performing an actual arithmetic operation. For different values of the inputs (X and Y take one of the m_i possible values) we obtain a correspondingly different value at the output, this value also being one of the m_i possible values. The output values are pre-calculated and stored (by means of position-coding) and are read out upon interrogation of the LUT by the inputs X and Y.

There are various possible implementations of position-coded LUTs(2-5), one of which is based on the utilization of small-size, high speed 1- or 2-D arrays of LEDs or LDs, in conjunction with fiber-optic combiners or holograms(6,7). Depending on the arrangement, three possible classes of LUTs are possible having the different complexities m_i^2, $2m_i$, $4\sqrt{m_i}$, where complexity (C) is defined as the number of LEDs or LDs necessary to implement a modulo m_i LUT.

In the m_i^2 class, we use an interlaced 2-D grid of electrodes in conjunction with LEDs or LDs at the intersection points (Figure 2). The simultaneous application of current pulses (each pulse is ~ 70% of the threshold) to intersection lines causes only one of the diodes on these lines, the one at the intersection point, to emit strongly. The emited light is then fed, by means of a hologram or a fiber-optic bundle, to a detector that is encoded for the number corresponding to that table location. The advantage of this approach is that well-established, low-cost technology can be used for the fabrication of the LUTs. The disadvantage is that the number of LEDs required grows as m_i^2, and thus large moduli (e.g., >19) cannot be used.

In the $2m_i$ class (Figure 3), two 1-D arrays of LEDs are arranged in a cross-configuration. Each LED emits a stripe-like optical beam. This is achieved through the use of a hologram not shown in Figure 3. At any time, only two perpendicular stripe beams will be present. These beams are incident on a nonlinear film which performs a thresholding operation, i.e., it allows light only at the intersection of the two beams to propagate. Subsequently, through the use of holograms or fibers, the light is directed to the proper detector. The advantages of this class of LUTs are: the number of LEDs grows proportional to $2m_i$, rather than m_i^2, and lack of electronic interconnections. The disadvantages are practical since there is, currently, no available optical thresholding film of adequate sensitivity.

In the $4\sqrt{m_i}$ class, one employs two sets (one for the X and one for the Y inputs) of one 1-D LD array and one 1-D array of aperture-type optical

switches (e.g., SEEDs). Figure 4 shows such a set with each array having 3 elements. Light from each LD illuminates all 3 switches. At any time, only one LD and one switch per set are "on," and thus there will be only one output light stripe. The location of this stripe depends on the specific pair of LD/switch that is "on" and since we have 3 LDs/switches, 9 different locations are possible. To create a LUT, the two sets of LDs/switches are arranged in a cross-configuration with a nonlinear film (like the $2m_i$ type). The advantage of this type of LUT is the highly reduced complexity, proportional to $4\sqrt{m_i}$, as compared to either the m_i^2 or $2m_i$ types. The disadvantages are the non-availability of thresholding film and the very complicated optical arrangement.

3. <u>Expected LUT Performance</u>. The performance of the m_i^2-type LUTs (these are currently the only practical LUTs), can be examined with respect to the multiplication speed (MS) and the system efficiency (SE) performance measures. Because LDs with subnanosecond switching times exist, and LEDs capable of operation at >1 GHz have been reported, MS of the order 1-3 GHz can be expected with currently available technology. In fact, using bulk commercially available LDs, we have fabricated[7] an m_i^2-type 7x7 LUT (Figure 5) and we have demonstrated MS of the order of 500 MHz (NRZ data). Such MS figures are well above those projected with GaAs multipliers. The SE figure has been estimated previously[6], and was found to be superior, by about an order of magnitude, to any current or projected electronic technology, including GaAs.

4. <u>LUT Matrix Multiplier Complexity</u>. To evaluate the benefits of the various LUTs and compare the residue and conventional electronic digital approaches, we should consider the relative complexity not only of the LUTs but also of complete systems that consist of the processing LUTs as well as the necessary residue/binary converters. We have performed such an analysis for a square matrix-matrix multiplication array of dimension N_2 which can be implemented with the 3 types of LUTs as well as a factored m_i^2-type LUT and compared with conventional and pipelined electronic digital multiplier-accumulators (MAU). Prior to describing some of the results, we describe the factored m_i^2-type approach for the case of a multiplier (the adder LUTs can be handled similarly).

With reference to Figure 1, observe that if either input of the multiplier is 0, the result is 0. Thus, if an input 0 can be detected, one needs an $(m_i-1) \times (m_i-1)$ LUT for operating in modulo m_i. If m_i is a prime number, m_i-1 is an even number and can be expressed as the product of various submoduli m_{ij}, e.g., $m_i=13$ and $m_i-1=12=3\times4$ ($m_{i1}=3$ and $m_{i2}=4$). One can show that employment of this technique allows the realization of m_i^2-type LUTs of much reduced complexity, e.g., $m_i=41$ C=1600 and factored $2\times2\times2\times5=40$, ($m_{i1}=2$, $m_{i2}=2$, $m_{i3}=2$, $m_{i4}=5$) and C=171. Note that not all prime numbers are conveniently represented. If we restrict our moduli to ≤73 and the factored LUTs to no greater than 7x7, then the following prime numbers can be used: 3(2), 5(2,2), 7(2,3), 11(2,5), 13(2,2,3), 19(2,3,3), 29(2,2,7), 31(2,3,5), 37(2,2,3,3), 41(2,2,2,5), 43(2,3,7), 61(2,2,3,5), 71(2,5,7) and 73(2,2,2,3,3). Note that 3x5x7x11x13x19x29x31x37x41x43x61x71x73 = 5.29×10^{18} or about 64 bits, which is enough dynamic range for a variety of computationally demanding applications.

Figure 6 shows the number of gates (for the 6 implementations) as a function of N for 16 input bits per MAU. We see that the m_i^2-class requires the largest number of gates; twice as much as the conventional digital approach and equal to that of the pipelined digital approach. Thus, this type of LUTs offers no significant complexity reduction. The factored m_i^2-type

approach requires about an order of magnitude less gates than those of the m_i^2-class, about half the gates of the conventional digital and about 12% of the gates of the pipelined digital approach. This demonstrates that significant complexity reduction can be accomplished even at the processor level. Note, that the complexity reduction can be improved even further if the $2m_i$ or the $4\sqrt{m_i}$ types of LUTs are employed.

5. **Conclusions.** Analyses and experimental results suggest that residue LUT processing can yield significant advantages in all three areas of speed, system efficiency and processor complexity. Such conclusions, coupled with the fact that virtually none of the 3 types of LUTs can be implemented by digital electronics (simply because of the formidable interconnect and fan-in/fan-out requirements) make residue LUT processing a promising optoelectronic computing approach.

References

1. N. Szabo and R. Tanaka,"Residue Arithmetic and Its Applications to Computer Technology," McGraw-Hill, New York, 1967.
2. A. Huang et al., Appl. Optics, 18:149 (1979).
3. S. Collins Jr., Proc. Soc. Phot. Opt. Ins. Eng., 128:313 (1977).
4. D. Miller et al., Appl. Phys. Letters, 45 (1):13 (1984).
5. A. Tai et al., Appl. Optics, 18 (16):2812 (1979).
6. P. Beaudet et al., Appl. Optics, 25(18):3097 (1986).
7. A. Goutzoulis et al., to be published in Opt. Communications (1987).

Figure 1. Look-up table for multiplication in modulo 5 residue arithmetic.
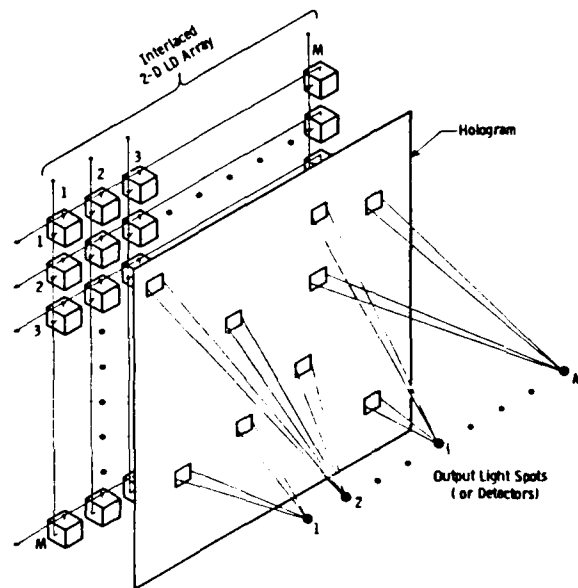
Figure 2. m_i^2-type LUT.
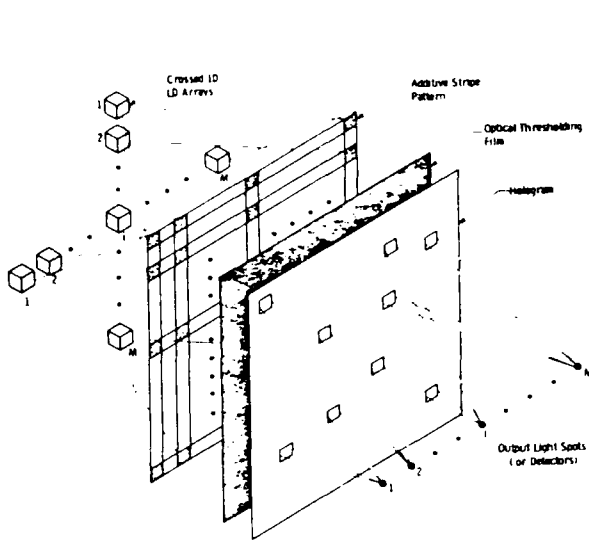
Figure 3. $2m_i$-type LUT.
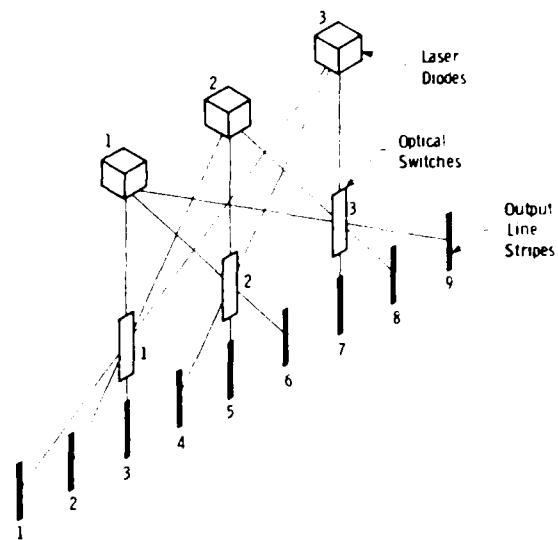
Figure 4. $4\sqrt{m_i}$-type LUT.
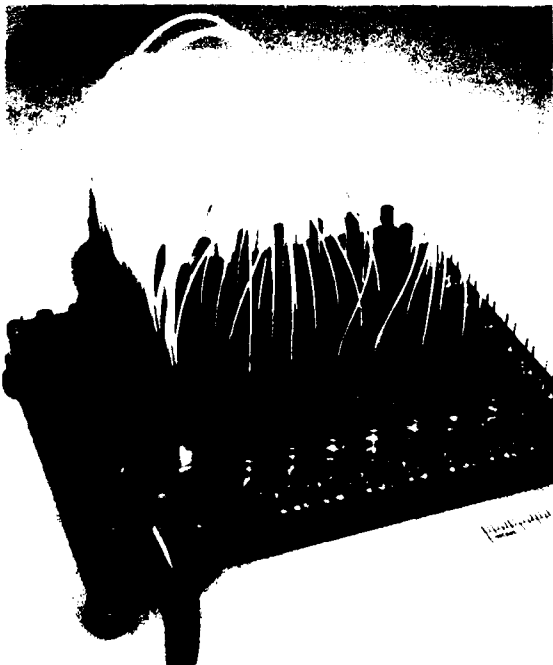
Figure 5. Interlaced electrode grid LUT prototype.
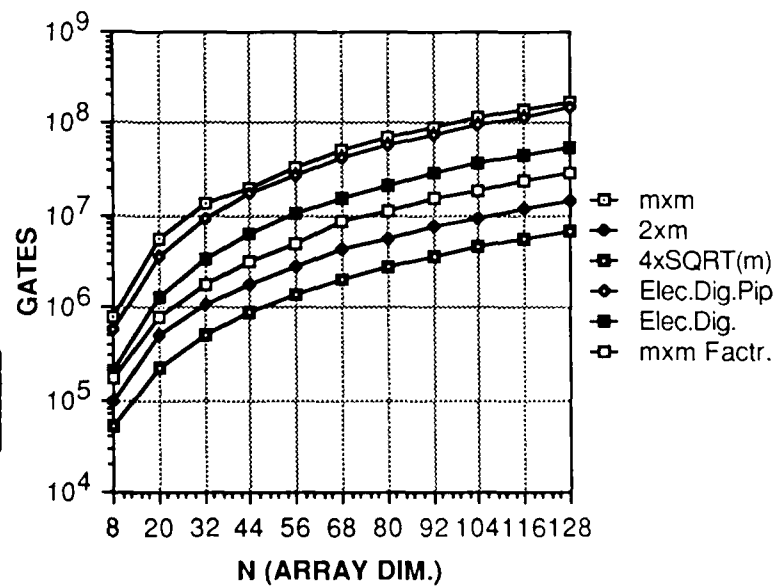
Figure 6. Gate complexity as a function of N for various LUT and digital approaches.

An Optical Arithmetic/Logic Unit Based on Residue
Number Theory and Symbolic Substitution

C. David Capps, R. Aaron Falk, and Theodore L. Houk
Boeing Aerospace Company
P. O. Box 3999, MS 87-50
Seattle, WA 98124

Introduction

In this paper we shall show how the concepts of residue number theory[1] and symbolic substitution[2] can be combined with current technology to create a device to perform arithmetic or logic operations at gigahertz rates. Since residue arithmetic involves no "carry" operations between different positions in the representation of a number, several of these devices, each based on a different radix, may be operated in parallel to perform digital operations at a rate that is independent of word size. The present device exploits optical processing for the pattern recognition portion of its function while using electronic elements for detection and thresholding and electro-optic devices as source modulators. As nonlinear optical devices mature they can be incorporated into this arithmetic/logic unit concept to achieve an all optical device.

Arithmetic/Logic Unit Concept

Figure 1 shows a radix 5 residue arithmetic table. The goal is to recognize the combination of input numbers and replace it with the correct answer. At first it appears there are 5^2 possible states that must be recognized. However, an examination of the matrix reveals that all the elements of an antidiagonal, one of which is shaded, are identical. Thus, if one can identify the appropriate antidiagonal given the inputs, then there are only 9 possible states. One method of doing this is to count the number of elements along the edge of the matrix between the two inputs. There is a unique correspondence between this "distance" and the operation result. This generalizes to radix N so that for NxN inputs there are only 2N-1 states that must be recognized to determine the answer. Thus, a recognition problem of quadratic complexity can be converted to one of linear complexity. An examination of the radix 5 multiplication table in Figure 2 a) seems to indicate that this approach will not work for multiplication. However, a theorem in residue arithmetic guarantees that if the radix is prime then the order of the inputs can be permuted so that the multiplication table, with zero inputs omitted, is antidiagonal like the addition table, Figure 2b). Thus, a device that can perform residue addition can, with slight modification, also do multiplication.

The method of determining the antidiagonal immediately suggests a positional scheme such as the one shown in Figure 3, for encoding the data. In this scheme a point source is turned on at the position corresponding to the input number. The algorithm for determining the appropriate antidiagonal then corresponds to finding the distance between the point sources. In the example shown all possible combinations with a distance of 5X between the sources, i.e. 4 + 0, 3 + 1, 2 + 2, 1 + 3, and 0 + 4, have the same answer, 4. A physical means of performing a radix 2 arithmetic or logic operation according to this scheme is sketched in Figure 4. A coherent source is

injected into an optical fiber and the power divided into four channels. The fibers go to four individual modulators which serve to turn on the appropriate combination of inputs and on to a linear fiber array. A spherical lens is used to take the Fourier transform of the source array resulting in a fringe pattern in the Fourier plane where an array of filters matched for 1X, 2X, and 3X source spacing is placed. A cylindrical lens then retransforms the light in one dimension for collection by a set of detectors, one for each possible spacing. As the spatial frequency in the Fourier plane will match one and only one filter, the detector in that channel will have a stronger signal than the others. Our calculations indicate that the margin, given amplitude only filters, will be at least two to one. Thus, by thresholding the detector outputs, a parallel determination of the correct source spacing can be achieved. Connecting the outputs of the appropriate channels and identifying them with the appropriate answer then leads to the positional coding of the output which can be cascaded to the next computational operation. Extension of this architecture to higher radices is straightforward.

In summary, we have presented a conceptual scheme for optically performing rapid arithmetic and logic operations that can be implemented with current technology. We are at present building this device to demonstrate its technical feasibility.

References

1) N. S. Szabo and R. I. Tanaka, "Residue Arithmetic and Its Application to Computer Technology", McGraw Hill, New York, 1967.

2) K. Brenner and A. Huang, "An Optical Processor Based on Symbolic Substitution", Conference Proceedings of the Topical Meeting on Optical Computing, Optical Society of America, 1985.

Input 1

	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Figure 1. Radix 5 Residue Addition Table

Input 1

	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

a. Normal Order

Input 1

	1	2	4	3
1	1	2	4	3
2	2	4	3	1
4	4	3	1	2
3	3	1	2	4

b. Permuted Order With Zero Inputs Removed

Figure 2. Radix 5 Residue Multiplication Table

● Source on
○ Source off

Point sources array

Spacing Nx Corresponds to unique operation result

Example: $3 + 1 \Rightarrow 5X$ spacing $\Rightarrow 4$ answer
also $2 + 2 \Rightarrow 5X \Rightarrow 4$
$0 + 4 \Rightarrow 5X \Rightarrow 4$

Input 1
4 ○
3 ●
2 ○
1 ○
0 ○

Input 2
0 ○
1 ●
2 ○
3 ○ — Uniform
4 ○ — x spacing of sources

Output
4 ●
3 ○
2 ○
1 ○
0 ○
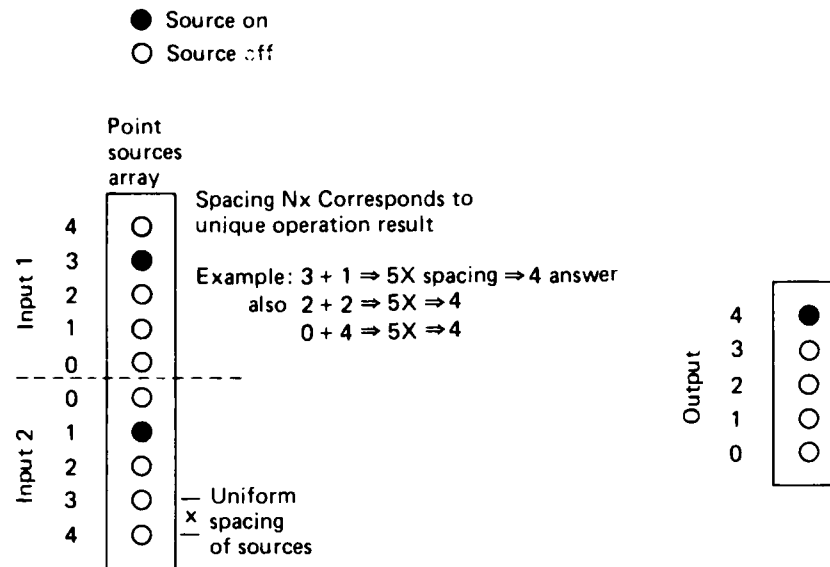
Figure 3. Positional Coding of Information for a Radix 5 Adder

Input 1

Mach-Zehnder modulators

Optical fibers

Fiber array

Input 2

Power divider

≋○ Coherent source

Spherical F. T. lens

Matched filter array

Cylindrical F. T. lens

Detectors

Output

Thresholders
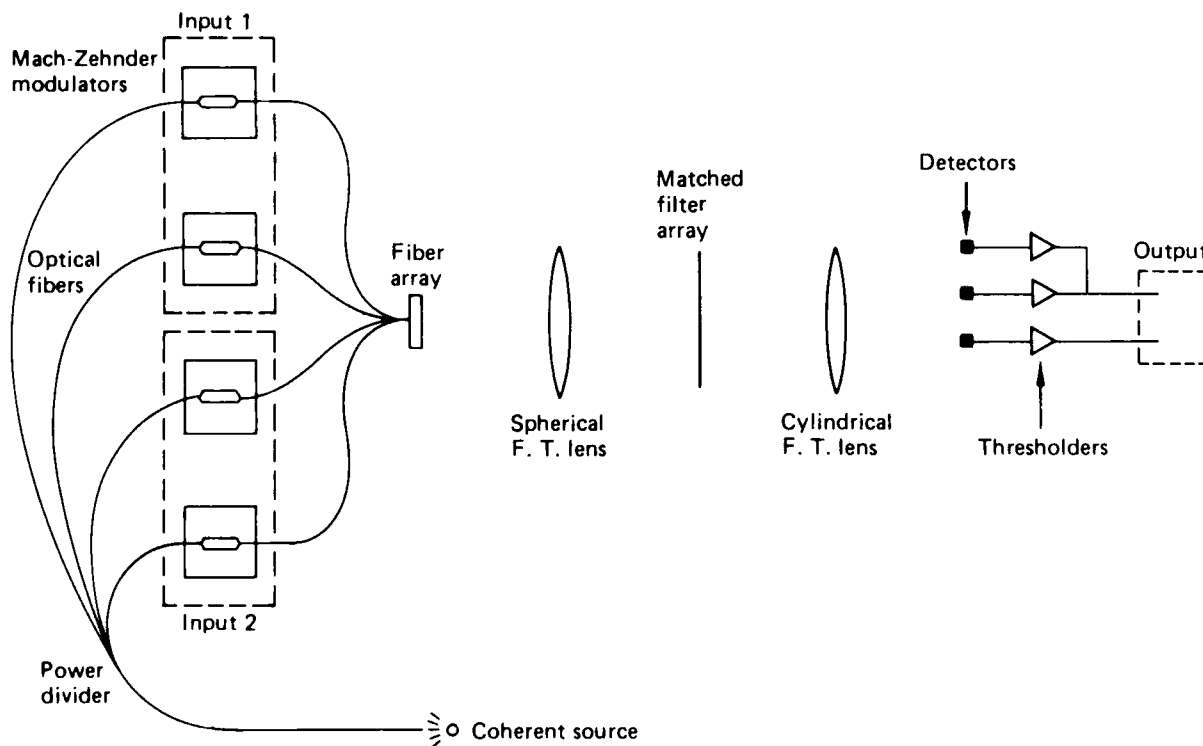
Figure 4. Device Configuration (Radix 2 Adder or Logic Unit)

65

Demonstration of a Digital Optical Matrix-Vector Multiplier

Using a Holographic Look-up Table and Residue Arithmetic

S. F. Habiby and S. A. Collins Jr.

Ohio State University
ElectroScience Laboratory
1320 Kinnear Road, Columbus, OH 43212

Summary

The optical matrix-vector multiplier described in this paper uses a Hughes liquid crystal light valve as the active element, the residue arithmetic number system and a holographic table lookup. It is designed with the premise that only one light valve response time is to be used to optimize throughput and that there are a plethora of image elements available on the light valve.

In the paper we will first describe the mapping approach to residue arithmetic and then describe how it is implemented in principle and in actuality including the role of a holographic lookup table. Finally results of operation are shown.

Residue arithmetic is well known. Given a modulus, m, and a number, N, the residue of N with respect to m is the remainder after dividing N by m. A number is represented by a set of residues $(r_1, r_2, r_3, \ldots)$ with respect to a set of moduli, $(m_1, m_2, m_3 \ldots)$. For example with modulus set (2,3,5) the residue representation for 17 is (1,2,2). A residue representation is unique through any consecutive range equal to the product of the moduli as long as the moduli are relatively prime. Numbers can be added, subtracted, and multiplied by adding, subtracting, and multiplying the individual moduli without the necessity of resorting to carries.

In this talk arithmetic operations of addition and multiplication are implemented by mappings such as those shown in Figure 1 where we see at the top mappings for times 1 and times 4 using a simple representation where the input is on the right side and output is on the top. To make connection one shifts left from the number on the input to the black square and then moves up to the output. For example, in the times four modulo five table 4x4=16 =1modulo5 so an input of four is connected to an output of one. Other tables for multiplication and addition are shown.

In this talk we will use the moduli 3,4, and 5 for illustration. In practice the modulus set (9,10,11) would be more appropriate, having a ten bit dynamic range.

The process is implemented optically as shown in Figure 2 where we see a liquid crystal light valve with the input beams on the left hand or input side. The light valve is configured so that a bright input rotates by ninety degrees the polarization plane of light

reflected off it, while dark input lets light reflect off it with unchanged polarization. Also shown in Figure 2 is a hypothetical polarizing mirror. That is a device which passes one polarization and reflects the other. In operation the input light passes through the polarizing mirror, has its polarization rotated by the light valve, is reflected a desired number of bounces, has polarization rotated again and passes out, thus representing one row of a typical map.

All the rows of a full map would be represented as shown in Figure 3 where we see the full set of input locations on the right and the output locations on the bottom right. The numerical value of an input number is represented by position coding so that only one of the input positions is illuminated at a time. Once the spots on the input side of the light valve have been illuminated then the particular output is obtained in the transit time of the light.

To implement the mapping operations without the hypothetical polarizing mirror, the loop configuration shown in Figure 4 is used. There we see the input spot array positioned vertically at the upper right and output array positioned horizontally on the bottom right. Light in an illuminated input spot passes through the polarizing prism on the right, and is imaged onto a spot on the light valve at upper center. It has polarization rotated so that it will pass through the polarizing prism on the left, is imaged again onto the mirror at lower center, and reimaged onto the light valve. One of the mirrors is tipped so that the spot is reimaged to a location adjacent to its original position, as desired. After the desired number of reflections off the light valve, the polarization is changed and the light passes through the left hand polarizing prism and is imaged onto the output line as shown. In the full matrix multiplier configuration this mapping is used many times for both multiplication and addition.

A schematic representation of the full matrix-vector multiplier is shown in Figure 5. This is configured to perform the matrix multiplication operation $c_{ij} = \sum a_{ij} b_j$. In Figure 5 we see two light valves. The one on the left is intended to perform the multiplications and has illuminating it on its write (left) side mappings representing the matrix elements a_{ij}. These patterns are generated by the microprocessor- controlled CRT at the left. The light valve on the right is intended to preform the final addition operation. A holographic lookup table memory converts the output from the individual $a_{ij} b_j$ products into maps which are then used to illuminate the input of the summation light valve.

The operation of the holographic lookup table memory is shown in Figure 6 where we see the reconstruction process illustrated. For a given operation and modulus, the hologram has superimposed as separate holographic exposures in the same area all the mappings for that operation. Thus addition modulo five will have mappings for +0, +1,+2, +3, and +4 all stored in that area. The exposures are differentiated by angle multiplexing, so that light from the different spots in the input plane, one focal length away from the lens. all reconstruct different mappings. The mappings reconstructed from the holographic lookup table memory then serve as input to the write side of the light valve for the addition loop.

The actual system implementing Figure 5 will be pictured

in the talk. One light valve is used for both multiplication and addition operations, both operations being performed in the same loop, and a hologram is used to connect the loops.

For practical reasons the multiple bounce lines of spots on the light valve go in a forty-five degree line. This was chosen to more effectively utilize our light valve which has liquid crystal in a parallel off-state configuration rather than the more usual twisted nematic hybrid field configuration, and in addition allow light to be vertically or horizontally polarized when passing through the polarizing prisms.

Only enough of the matrix-vector multiplier was set up to demonstrate operation. This includes one multiplication for one modulus, and one addition which responds to the particular product from the multiplication. The portion demonstrated are highlighted in Figure 5. To perform a numerical operation with full dynamic range three parallel mappings would be required, one for each modulus.

For indication of operation the optical patterns produced in various planes are shown in Figure 7. Figure 7a shows representative patterns from the holographic input to the addition modulo five operation. These are photograms made by placing photographic contact paper in an image plane. The grid was dubbed in during the printing process. The patterns are identical with those shown in Figure 1 except that they are rotated forty-five degrees to conform with the desired light valve operation. Similar patterns were generated by the CRT input for the multiplication operation. Figure 7b shows simultaneous outputs from the multiplication and addition operations. The system is configured to add three from the addition loop to the output from the multiplication loop using modulus five. Thus the part to the top in 7b shows a spot representing a value of one for the output from the multiplication loop and four from the addition loop as expected and the part to the right shows an output of four from the multiplication loop and two, (4+3)modulo5 as the output of the addition loop.

To summarize, we have designed, built, and demonstrated proof of concept for a residue-based matrix vector multiplier using a hologaphic lookup table and performing a complete matrix vector multiplication operation in one light valve response time.
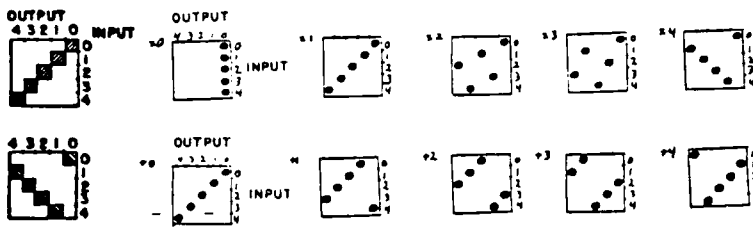
Figure 1.
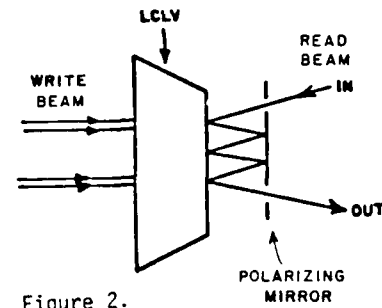
Figure 2.
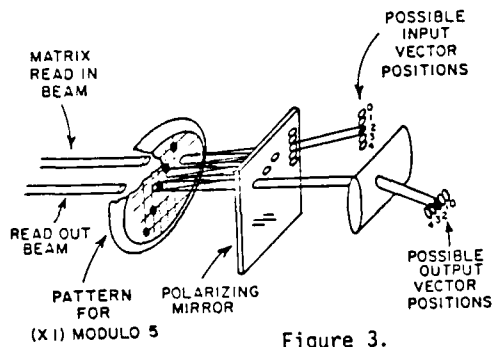
Figure 3.
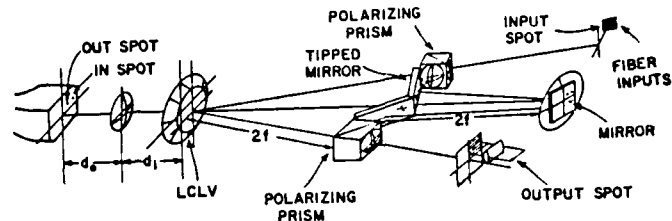
Figure 4.
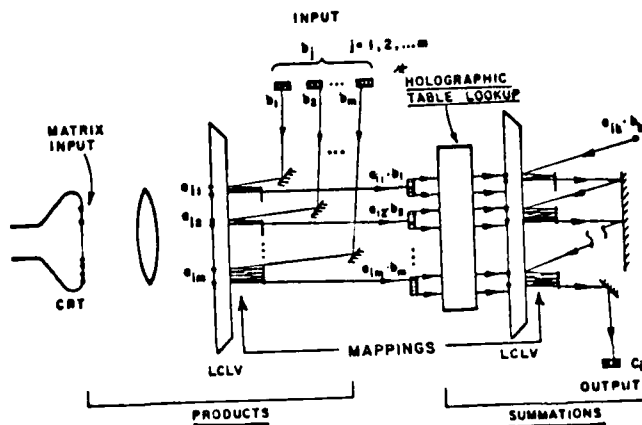
Figure 5.
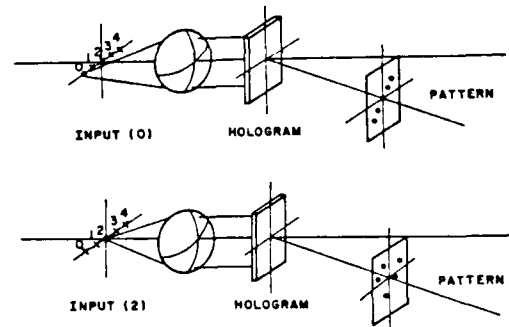
Figure 6.
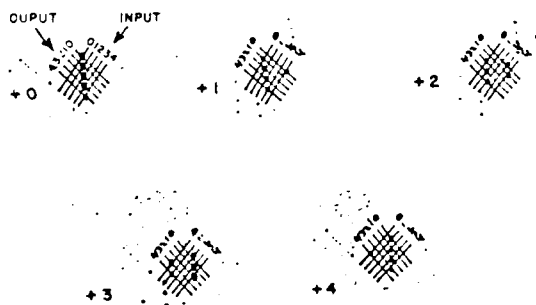
(a)
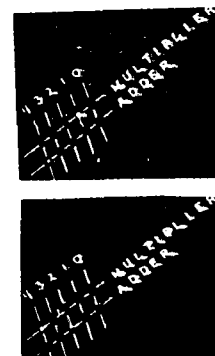
(b)

Figure 7.

LIMITATIONS TO OPTICAL FREDKIN CIRCUITS

Robert Cuykendall and Debra McMillin

Department of Electrical and Computer Engineering
University of Iowa , Iowa City, IA 52242

Several optical Fredkin gate [Fig. 1] implementations have recently been proposed[1] as building blocks for optical computers. It is shown that such conditional-routing devices can in fact be cascaded to compute the function $f(x_1,...,x_n,y_1,...,y_n)=[(x_1 \cdots x_n)+(y_1 \cdots y_n)]mod2$. This is accomplished through the design of a Fredkin-based minimal full adder circuit [Fig. 2] with carry-out feedback possessing the property that its control signals never interchange with data signals. Computation is performed by inserting the variable arguments initially into various control lines, while constant values are inserted into the initial data lines. The circuit is thus 'programmed' by scratchpad constants entered into its first-level data lines.

Utilizing the proposed building blocks, additional circuits can be designed having some computational properties beyond their obvious interconnection properties. For example, a 1-line to n-line demultiplexer [Fig. 3] can be implemented without intermixing control and data lines. Such a circuit can then be programmed to compute all minterms for an arbitrary switching function $f(x_1,...,x_n)$. However, these circuits cannot sum their outputs, thus cannot compute f, nor can they be cascaded in any way which would be computationally productive. Moreover, the perfect shuffle on n inputs can be implemented [Fig. 4] and programmed to compute not only specific minterms, but any minimal sum of products for an arbitrary switching function $f(x_1,...,x_n)$. It can thus compute f, but again these circuits cannot themselves be cascaded in order to compute functions in a composite (or sequential) form, or to compute functions having a dynamic computational dependence.
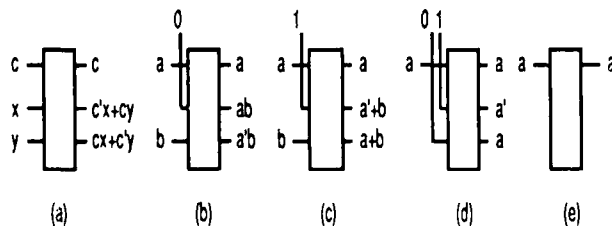
Fig. 1. (a) Fredkin gate realization of: (b) AND, (c) OR, (d) NOT-FANOUT and (e) DELAY.
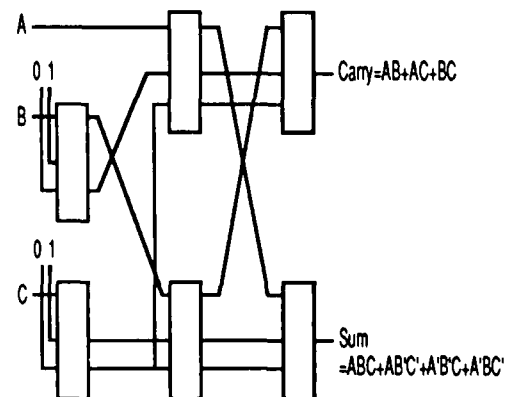
Fig. 2. Minimal restricted Fredkin adder [6 sink, 4 sink delays].
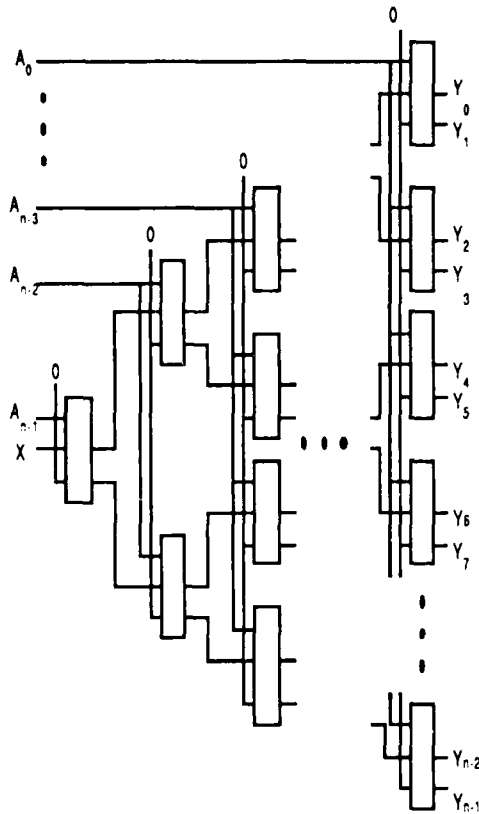
Fig. 3. 1-line to n-line demultiplexer [address
inputs $A_0,...,A_{n-1}$ specify to which of
the n outputs $Y_0,...,Y_{n-1}$ the data signal
X is to be routed].
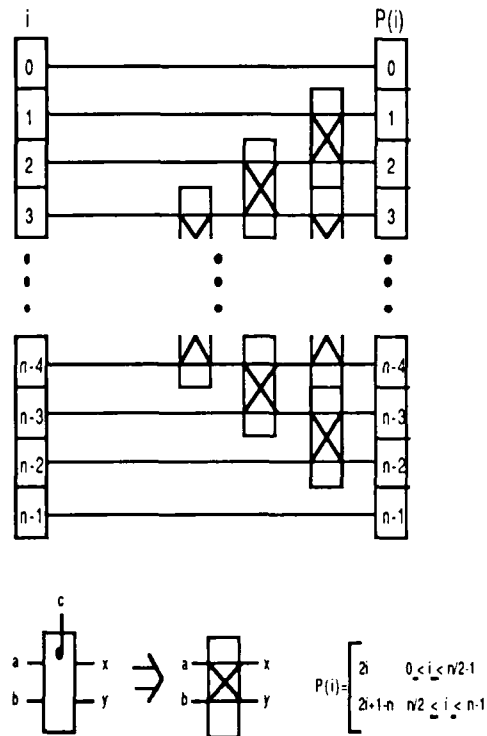
Fig. 4. n-point perfect shuffle.

General circuits for implementing higher-order logic functions (such as n-bit binary multiplication) may not be possible due to the fundamentally different nature of the control signals in the devices proposed, and the observation that every computing primitive requires the control line be used to input an argument value. This dynamic dependence of the control signal on the computation remains when the Fredkin gate is constructed by cascading elementary Priese switch gates[2] [Figs. 5,6].
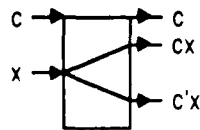
Fig. 5. Priese switch gate [input signal x routed
to one of two output paths depending on
the value of the control signal c].
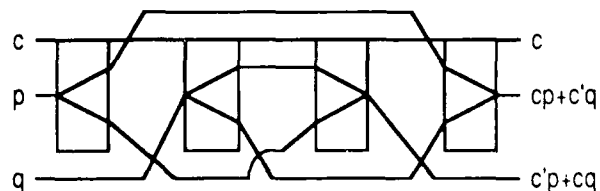
Fig. 6. Fredkin gate realization [bridge symbol
indicates nontrivial crossover; all
other crossovers are trivial].

The "interaction" gate [Fig. 7], on the other hand, is a reversible two-input universal logic gate which utilizes no control-specific signals. It is well known[3] that a Fredkin gate can also be realized [Fig. 8] by cascading interaction gates. Thus, any cascadable optical implementation of an interaction gate would *a fortiori* allow an optical realization of a Fredkin gate in which the control line is not basically different in nature than the other two lines. While such a cascade may not be practical when compared to simply using the interaction gate itself in computing circuits, the resulting composite Fredkin gates could be cascaded into arbitrary reversible sequential circuits[4].
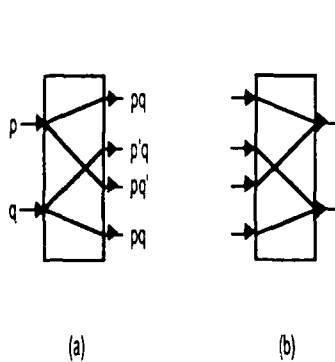
Fig. 7. (a) The interaction gate and (b) its inverse.
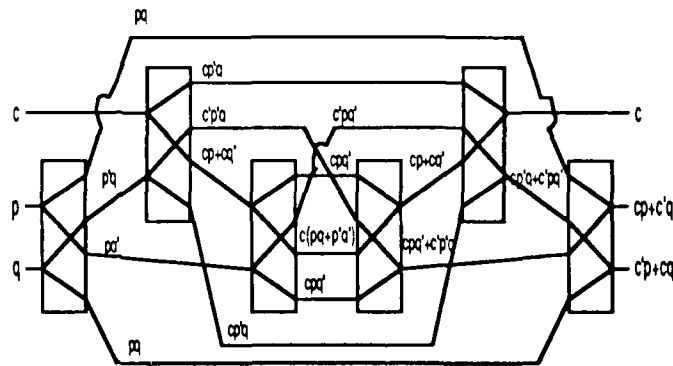
Fig. 8. Fredkin gate realization [bridge symbol indicates nontrivial crossover; all other crossovers are trivial].

References

1. J. Shamir, H. Caulfield, W. Micelli and R. Seymour, "Optical Computing and the Fredkin Gates", Applied Optics **25**, 1604 (1986).

2. L. Priese, "On a Simple Combinatorial Structure Sufficient for Sublying Nontrivial Self-Reproduction", J. Cybernetics **6**, 101 (1976).

3. E. Fredkin and T. Toffoli, "Conservative Logic", Int. J. Theor. Phys. **21**, 219 (1982).

4. T. Toffoli, "Reversible Computing", Tech. memo **151**, Lab. for Comp. Sci, MIT (1980).

MATRIX-VECTOR MULTIPLICATION
USING POLARIZATION ROTATORS

L. Scharf, W. T. Cathey, K. M. Johnson

Center for Optoelectronic Computing Systems
Electrical and Computer Engineering Department
University of Colorado, Boulder
Boulder, CO 80309-0425

Introduction

The potential application of optical systems to perform high speed, low cost signal processing with large parallelism has attracted the attention of researchers for many years. General optical processors have been developed that compute matrix-vector multiplications and other linear algebraic operations using incoherent light. One example is the Optical Matrix-Vector Multiplier (OMVM), which calculates the discrete operation of a matrix-vector product, rather than the continuous correlation and convolution more commonly associated with optical processing [1]. The OMVM can be used to compute discrete Fourier transforms (DFT's), and for performing linear algebraic operations, including matrix-matrix multiplications. It has been suggested as a method for implementing associative memory [3-5] and optical crossbars [4]. The first OMVM had several disadvantages, including low accuracy, low speed, and a nonprogrammable matrix mask. Recent implementations use real-time spatial light modulators (SLM) [5-7] and acousto-optic cells [8]. The two-dimensional spatial light modulators used in many of these optical processors operate at millisecond speeds, are expensive and have low resolution [5, 7]. One-dimensional modulators such as acousto-optic cells are faster, but the major drawback of computing matrix-matrix operating using one-dimensional devices is that to calculate two-dimensional matrix-matrix operations, data from the rows and columns of matrices must be loaded serially. The cycle time through the processors increases with the order of the matrix, and the natural parallelism of optics is lost.

Objective

The goal of our research is to achieve 100 x 100 matrix-matrix multiplications in a microsecond, with 10 bit or greater accuracy. To achieve this goal, a new approach is needed. We describe a two-dimensional optical systolic processor with new algorithms, architectures, and devices which we believe will result in the evolution of an optical processor capable of meeting this goal. In this paper we outline our design principles for high-speed, high precision optical implementations of linear algebraic computations.

One can view the matrix-matrix multiplications problem with the frame work of an i/o problem and a realization problem.

(i)   I/O problem : multiply matrices **A** and **B**.

For this I/O problem there are an infinite number of *realizations*
or *algorithms* that one can use to perform the multiplications.

We can use this freedom to optimize criteria associated with the
computation. For example, in some digital processing problems we
choose an algorithm to minimize the number of computations. In this
particular application we wish to design algorithms which use low
accuracy primitives to obtain a high accuracy result. We also wish to
pipeline computations, develop highly regular and locally connected
geometrices, and to use simple optical primitives as the basis of the
algorithms.

(ii)   Realization problem.

The realization problem consists of finding architectures that consist
of simple optical primitives, connected in modular geometries, to
produce high-accuracy results by pipelining the computations through low
accuracy cells. This goal involves:

   (a) low accuracy primitives for high accuracy results
   (b) modular geometrices
   (c) pipeline computations
   (d) simple, optical primitives

Algorithms and Architectures

   The algorithms being used for this processor break-up matrices into repetitive operations
on a smaller set of orthogonal rotation matrices. The algorithms are low loss and the archi-
tectures used to implement the algorithms are cellular, as shown in Fig. 1, and based on opti-
cal operations [9].
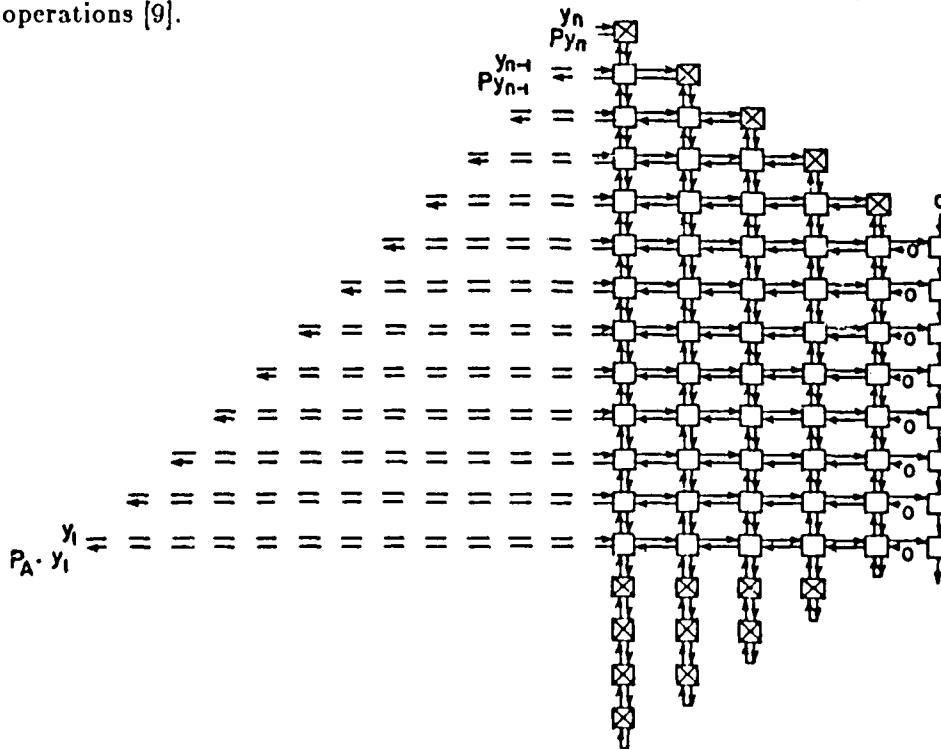
Figure 1. Cellular Implementation of a Vector Pipelined Projection Operator.

74

Optical Implementations

Figure 2 illustrates the rotation operation on incoming signals as a (2 x 2) matrix map. This same operation can be implemented optically using devices that rotate the polarization of the input vector. One optical implementation of the rotator-combiner is shown in Fig. 3, where the first element is a polarizing beamsplitter which separates the x and y components. The second polarizing beamsplitter acts as a combiner of the appropriate components, and a polarization rotator then imparts the desired rotation onto the resulting vector. For hard-wired applications, quartz, which gives a rotation fo 21.7° /mm, could be used. The thickness can be controlled to yield the desired rotation. Electrically controlled rotators would give programmability and an array of liquid crystals could provide discrete rotations.

$$\begin{vmatrix} V_1 \\ V_2 \end{vmatrix} = \begin{vmatrix} \cos\Theta & \sin\Theta \\ \cos\Theta & \sin\Theta \end{vmatrix} \begin{vmatrix} U_1 \\ U_2 \end{vmatrix}$$
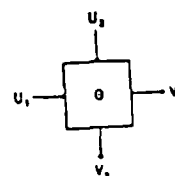
Figure 2. Signal Rotation Operation.

Figure 4 shows that, with the development of a rotator-combiner cell, the general problem of implementing matrix-vector and matrix-matrix multipliers in numerically stable machines can be implemented in a regular cellular array of such rotator-combiner cells.
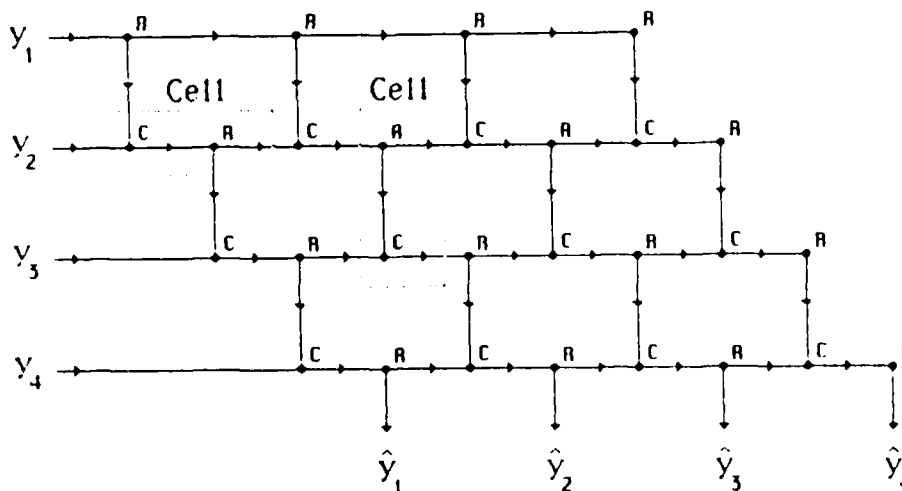
Figure 3. Cellular Architecture for Implementing a Sequence of Rotations.

We will discuss implementing the rotator-combiner cell using polarizing beamsplitters, and ferroelectric liquid crystal (FLC's) which can switch the polarization of incident light in less than a microsecond [10, 11]. These crystals, developed at the University of Colorado, Boulder, in the Physics Department have already been fabricated successfully in 32 x 32

matrix arrays [12]. By making 256 x 256 matrix arrays, a trade-off between array size and accuracy can be achieved. In addition, since these FLC's are capable of submicrosecond switching speeds, a trade-off between speed and accuracy can now be made for the first time.
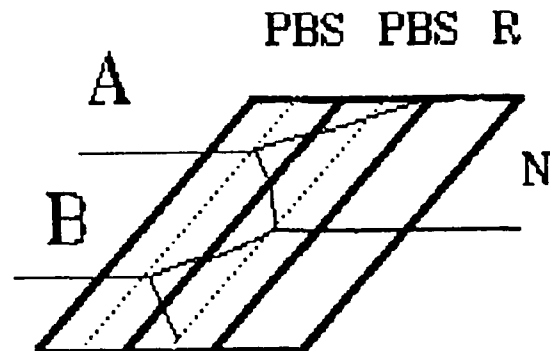
Figure 4. Integrated Rotator-Combiner Using Polarizing Beamsplitters.

References

(1) Goodman, J. W., Dias, A. R., Johnson, K. M., and Peri, D., "Parallel Incoherent Optical Matrix-Vector Multipliers," Proc. Workshop on Optical Signal Processing, Texas Tech. University, Lubbock, Texas, 116-128 (1980).

(2) Goodman, J. W., Johnson K. M., "The Incoherent Matrix-Vector Multiplier," Final Report (L7-1-1, April (1981).

(3) Psaltis, D. and Farhat, N., "Optical information Processing Based on an Associative Memory Model of Neural Nets with Thresholding and Feedback," Opt. Lett.10, 98 (1985).

(4) Bell, T.L A., "Optical Computing: A Field in Flux," IEEE Spectrum 23, 34 (1986).

(5) Athale, R., Lee, J. N., "Optical Processing Using Outer-Product Concepts," Proc. IEEE 72, 931 (19084).

(6) Minemoto, T., Okamoto, K., Miyamoto, K., "Optical Parallel Logic Gate Using Spatial Light Modulators With the Pocket Effect," Appl. Opt.24, 2055 (1985).

(7) Grinberg, J., Soffer, B. H., Owechko, Y., and Marom, E., "Programmable Real-Time Incoherent Matrix Multiplier for Optical Processing," in Tech. Digest on Optical Computing, Paper TuD-1, OSA (1985).

(8) Verber, C. M., "Integrated-Optical Approaches to Numerical Optical Processing," Proc. IEEE 72, 942 (1984).

(9) Rialan, C. P. and Scharf, L. L., "Cellular Architectures for Implementing Projection Operators," to be published in IEEE Trans. ASSP.

(10) Clark, N. A., Lagerwall, S. T., "Submicrosecond Bistable Electro-Optic Switching in Liquid Crystals," Appl. Phys. Letters 36, 899 (1980).

(11) Clark, N. A., Lagerwall, S. T., "Physics of Ferroelectric Fluids: The Discovery of a High-Speed Electro-Optic Switching Process in Liquid Crystals," Rec. Dev. in Cond. matter Physics 4, 309 (1981).

(12) Pagano-Stauffer, L. A., Johnson, K. M., Masterson, H., Handschy, M., Clark, N., "Optical Parallel Processing Using Ferroelectric Liquid Crystal Arrays," SPIE Conference, August (1986).

Monte Carlo Matrix Inversion
Using an Optical Random Number Generator

Anthony J. Martino and G. Michael Morris
Institute of Optics
University of Rochester
Rochester, New York 14627

Introduction

Matrix inversion and other linear algebra problems are a frequent subject of interest in the field of optical computing. In this paper we describe laboratory experiments involving Monte Carlo matrix inversion using an opto-electronic hybrid system. The hybrid system is composed of an optical random number generator coupled to an electronic digital processor.

The deterministic "pseudo-random" number routines usually found on digital computers sometimes are unsuitable for extensive calculations because of repetitions in the sequence and other departures from randomness, and because they are inflexible with respect to the distribution of the random number output. The advantages of such routines are that they are simple and can generate the same sequence repeatedly for testing purposes.

In the past, devices that employ naturally stochastic physical processes have been used to produce true random numbers. Processes involved have included time sequences of noise voltage in electronic circuits and emission of particles from radioactive substances. These devices have suffered from difficulty of calibration and inflexibility of distribution.

Recently, devices have been reported that produce true random numbers by using the spatial variation of stochastic optical processes such as photon-limited images[1,2] and laser speckle[3]. The use of spatial rather than temporal randomness reduces problems in calibration and allows one to produce random numbers with any bounded two-dimensional distribution.

The random number device used in the experiments reported herein is based on photon-limited imaging.

The Optical Random Number Generator

The optical random number generator works as follows. An object (transparency), which can be used to control the distribution, is projected by a lens/filter combination onto the cathode of a detector at an irradiance level such that photon counting is possible. The detector is a two-dimensional, position-sensitive

photon counter; its output is the spatial coordinates of the location of each photoelectron as it is ejected from the cathode. These coordinates, digitized to 8 bits in each direction, are used by the electronic computer as random numbers. The maximum rate is 100 thousand random numbers per second with the present detector and its support electronics; this is sufficiently fast that the electronic processor speed and not the random number rate limits the speed of computation.

The Matrix Inversion Algorithm

The following algorithm is described in detail in the literature. (See, for example, Ref. 4.)

An $n \times n$ matrix A is to be inverted. Compute the matrix $B = I - X_0 A$ in which I is the identity matrix and X_0 is an initial estimate of the inverse matrix. Assume that the initial estimate X_0 is equal to the identity matrix I, so that $B = I - A$. For the algorithm to work, it is necessary that the norm of B, denoted $\|B\|$, is less than unity.

Next, form a discrete Markov process with $n + 1$ states, in which one state, state $m = n + 1$, is an absorbing state. The transition probabilities are governed by

$$p_{ij} = \frac{B_{ij}}{v_{ij}} \quad ; \qquad 1 \leq i, j \leq n \ , \tag{1}$$

$$p_{im} = 1 - \sum_{j=1}^{n} p_{ij} \quad ; \qquad 1 \leq i \leq n \ , \tag{2}$$

$$p_{mj} = \delta_{mj} \ , \tag{3}$$

in which the quantities v_{ij} are called value factors. Note that the transition probabilities and the value factors are somewhat arbitrary and can be optimized subject to the constraints of Eq. (1)-(3) and the requirement that all probabilities be positive and less than or equal to unity.

The random variable G_{ii} is defined as

$$G_{ij} = 0 \quad ; \qquad s_k \neq j \ , \tag{4}$$

$$G_{ij} = \frac{v_{is_1} v_{s_2 s_3} \dots v_{s_{k-1} s_k}}{p_{jm}} \quad ; \qquad s_k = j \ , \tag{5}$$

in which s_1 through s_k represent the sequence of states in one realization of the Markov process. It can be shown that the expected value of G_{ij} is equal to the element $(A^{-1})_{ij}$ of the inverse matrix.

Theoretical Predictions

The performance of the matrix inversion algorithm can be affected significantly by the choice of transition probabilities, numbers of chains, and other variable parameters. To predict such effects, we use methods similar to those given in Ref. 5.

For example, we wish to know the effect of varying the "stop" probability p_m. The probability that a Markov chain has length l is given by

$$P(l) = (1 - p_m)^{l-1} p_m \quad , \qquad (6)$$

and therefore the mean chain length is given by

$$\langle l \rangle = \frac{1}{p_m} \quad , \qquad (7)$$

If d is a number such that the absolute error of an element of the Monte Carlo solution is less than d with probability 0.95, then it is approximately true that

$$d = \frac{2\sigma}{v^{1/2}} = \frac{2\sigma}{(p_m N_{row})^{1/2}} \quad , \qquad (8)$$

where v is the number of Markov chains completed while calculating a row, N_{row} is the random number count for the row, and σ is the maximum standard deviation of any element in the row.

The variance of an element is given by

$$\sigma_{ij}^2 = \frac{1}{p_m} ((I - C)^{-1})_{ij} - ((A^{-1})_{ij})^2 \quad , \qquad (9)$$

where $C_{ij} = B_{ij} v_{ij}$.

Finally, the computation time necessary to invert a matrix is given by

$$T = n^2 T_{setup} + n \left| \frac{2\sigma}{d} \left[\left(\frac{1}{p_m} - 1 \right) T_1 + T_2 \right] \right| \quad , \qquad (10)$$

in which T_{setup} is the time to calculate each value factor, T_1 is the time for each non-terminal step of a chain, and T_2 is the time spent at the end of each chain.

Experimental Results

Experiments were performed using matrices of different orders and norms. Measurements were made of chain-length statistics, total random numbers used, execution times, and error rates while varying the stop probability p_m and the order n of the matrix.

Figure 1 contains a plot of error rate r vs. p_m for a 50×50 matrix that was generated randomly and normalized to have $\|B\| = 0.5$. The three curves show results for three (fixed) random number counts. Note that r decreases with increasing p_m and with increasing N_{row} as predicted by Eq. (8).
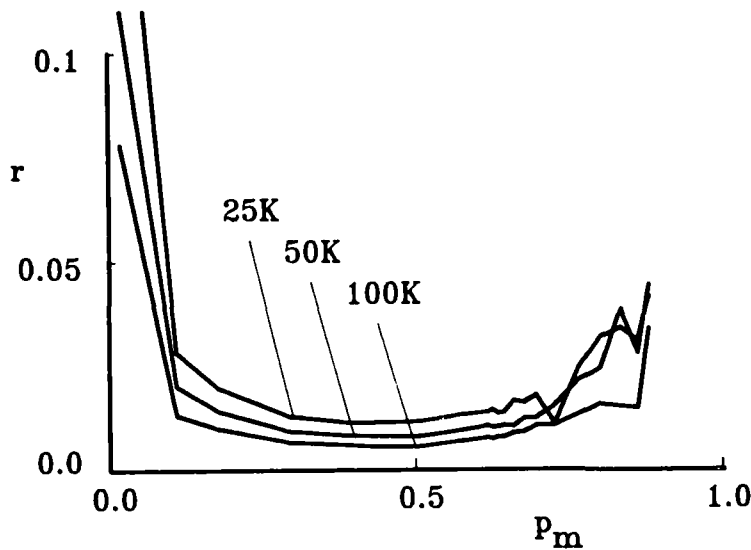
Figure 1. Plot of error rate $r = d / \|B\|$ against stop probability p_m for a 50×50 matrix. Curves are for three different values of N_{row}, the random number count per row.

This and other experimental results that will be presented clearly show the tradeoff between speed and accuracy that exists for Monte Carlo matrix inversion. The tradeoff is greatly affected by the choice of transition probabilities and other parameters, such as the order of the matrix and the initial estimate. For example, if p_m is too small, error will increase; if p_m is too large, the number of chains (and hence the time) will increase.

The execution time for a 50×50 matrix inversion by this Monte Carlo method is much longer than the time for the same inversion by Gauss Elimination. However, the n^2 dependence of W is of a lower order than the $\sim n^{2.8}$ dependence of the more refined Gauss Elimination routines. Estimates will be given for the "breakeven value" for n, above which the Monte Carlo method will be more efficient.

References

1. G.M. Morris, Optical Engineering 24, 86-90 (1985).

2. A.J. Martino and G.M. Morris, OSA Topical Meeting on Optical Computing, 1985.

3. J. Marron, A.J. Martino, and G.M. Morris, Applied Optics 25, 26-30 (1986).

4. R.Y. Rubinstein, *Simulation and the Monte Carlo Method.* (Wiley, New York, 1981).

5. J.H. Curtiss, in *Symposium on Monte Carlo Methods*, H.A. Meyer, ed. (Wiley, New York, 1956) 191-233.

MONTE-CARLO PROCESSOR ARRAYS USING OPTICAL RANDOM NUMBER GENERATORS.

F. DEVOS, K. MADANI

Institut d'Electronique Fondamentale,laboratoire associé au C.N.R.S., Université de Paris-Sud , 91405 ORSAY cedex, France.

P. CHAVEL, J. TABOURY

Institut d'Optique, laboratoire associé au C.N.R.S., Université de Paris-Sud. B.P. 43. 91406 ORSAY cedex, France.

I INTRODUCTION :

The use of optical phenomena for the generation of random number arrays has been introduced by Morris and his coworkers in recent publications [1-3].

In the present work. we elaborate on the usefulness of optical binary random 2-D array generation for massively parallel electronic Monte Carlo computing. The principles proposed are described in section II and relevant numerical orders of magnitude are given in section III. In section IV, we discuss the respective advantages of speckle and of single photoevent detection for this purpose and present a preliminary experimental illustration.

II PRINCIPLES :

The importance of Monte-Carlo algorithms for solving computation problems in high dimensionality spaces is well known. For example, the simulated annealing procedure, a powerful method for obtaining near optimal solution to NP complex optimization problems, derived by Kirkpatrick [4] from the Metropolis algorithm [5], has recently attracted considerable attention. The design of parallel processors dedicaced to such algorithms is an sensible issue, since the multidimensional problems suitable for stochastic computing also require a large number of iterations of computing cycles based on random trials, but can usually accomodate a high degree of parallelism. Massive parallelism, involving typically 10^4 to 10^5 parallel processing elements (P.E..s), is compatible with present on-chip technology only if the P.E.s are limited to a few tens of transistors. This precludes the use of pseudo-random number generation techniques by the P.E.s themselves to obtain the required random numbers ; instead, physical random phenomena have to be used.

There is therefore a need for physical means of iteratively generating 2-D random number arrays at a rate compatible with a desirable computing cycle rate of typically 1 μs, with each drawing independant from the others. Specifically, since a PE of the kind considered can only operate on one-bit words, binary random arrays of the type $[a_{ij}{}^t]$ are required. with $a_{ij}{}^t$ equal to 0 with probability $P(i,j,t)$ and to 1 with probability $1 - P(i,j,t)$, where i,j are the PE line and column numbers,i.e. pixel number

and t denotes a discrete time sequence. Depending upon each particular algorithm, the probability P(i,j,t) may actually depend on time only or on space only, or on all three variables.

The random array generation process must be compatible with the parallelism of the processor, i.e. require no additional calculation or out-of-chip component. We propose to produce optical random phenomena on a photosensitive input located on each P.E.. Standard NMOS and CMOS technologies are readily suitable for diffusion of photodiodes on silicon.

III RELEVANT ORDERS OF MAGNITUDE :

The following numerical values are only orders of magnitude intended to investigate the requirements and feasibility of the proposed procedure. Each PE may be provided with a 10 x 10 μm photosensitive area with a typical capacitance of 0.01 pF. A voltage drop of about 0.1 V is required for "low" state detection and therefore the drawing of, for example, value 1 of $a_{ij}{}^t$. This means that about 5 000 photoelectrons must be extracted during one computation cycle time of duration 1 μs. For a photoelectric yield of 10 %, 50 000 photons must reach the photosensitive area of each PE i,j during each cycle t for which $a_{ij}{}^t = 1$. With the maximum sensitivity of silicon in the red and near infrared part of the spectrum, this corresponds to an energy of 10^{-14} J. Therefore, if unfocussed light illuminates the whole chip of area round 2cm^2, about 1 mW is necessary.

It must be kept in mind that the probability p(i,j,t) of $a_{ij}{}^t$ being equal to zero must be modulable from 0 to 1 with a good accuracy by external control of the illumination level. If the random array is derived from a randomly varying physical quantity q assuming a large number of values, then some threshold q_t has to be set such that $a_{ij}{}^t = 0$ if q does not reach the threshold q_t, 1 otherwise. Any fluctuation or inaccuracy in the threshold affects the accuracy on p(i,j,t). As a consequence, the photon arrival statistics related to a pulse height modulated bunch of about 50 000 photons cannot be used : such a pulse shows a dispersion of only 230 photons : if the average light flux is modulated and a detector threshold fixed at about 50000 photons, a flux variation of only about one percent would change the detection probability from almost zero to almost unity. This means that a fluctuation of only a few percent in the photodetector characteristic from one PE to the next would make the required control completely impossible. It is therefore necessary to rely on a different random phenomenon, showing a large dispersion. Two possibilities will be considered here : the amplification of single photoevents, and speckle.

IV DISCUSSION OF SUITABLE OPTICAL RANDOM PHENOMENA :

i - Microchannel plates image intensifiers are suitable for easy adaptation to a chip. They provide resolution in the tens of micrometers range, with a typical gain in photons of roughly 1 000. Two of these in cascade can therefore, out of single arriving photons produce adequate light levels for the desired operation. The experiment then consists of a low, controlled light level image projected on the photocathode of the first image intensifier. The random arrays produced by this

method show independance in space and in time. For a preliminary experimental illustration, we have cascaded two image intensifiers. Due to low gain of the particular devices used, we had to feed back the output signal to the output window with an optical fiber. The saturation gain was then reached and random signals with uniform spatial density were generated out of the dark current of the first electron multiplier. The outcoming pulse of light was proximity coupled onto a fiber and then used to illuminate one photosensitive area on a PE out of aN-MOS test circuit made of a 8x10 PE array {6}. The oscilloscope trace on the figure below shows an arriving bunch of photons clearly triggering low -state detection : the capacitance voltage vs time curve shows a sudden slope discontinuity upon arrival of the light pulse. This experiment needs further refinement for several reasons : in particular, a higher gain is required to generate the light pulse by an incoming low light level image and ultra-high response time phosphors will be needed to provide the desired iteration rate. If these problems can be solved, then the method shows potential for generating the low-light level image driving probability $P(i,j,t)$ in situ on the cellular processor chip by incorporating a low power photodiode in addition to the photodetector (using an adequate technology, obviously not silicon alone) : this would be very useful in the case of simulated annealing algorithm where the value of $P(i,j,t)$ results from local energy estimates.
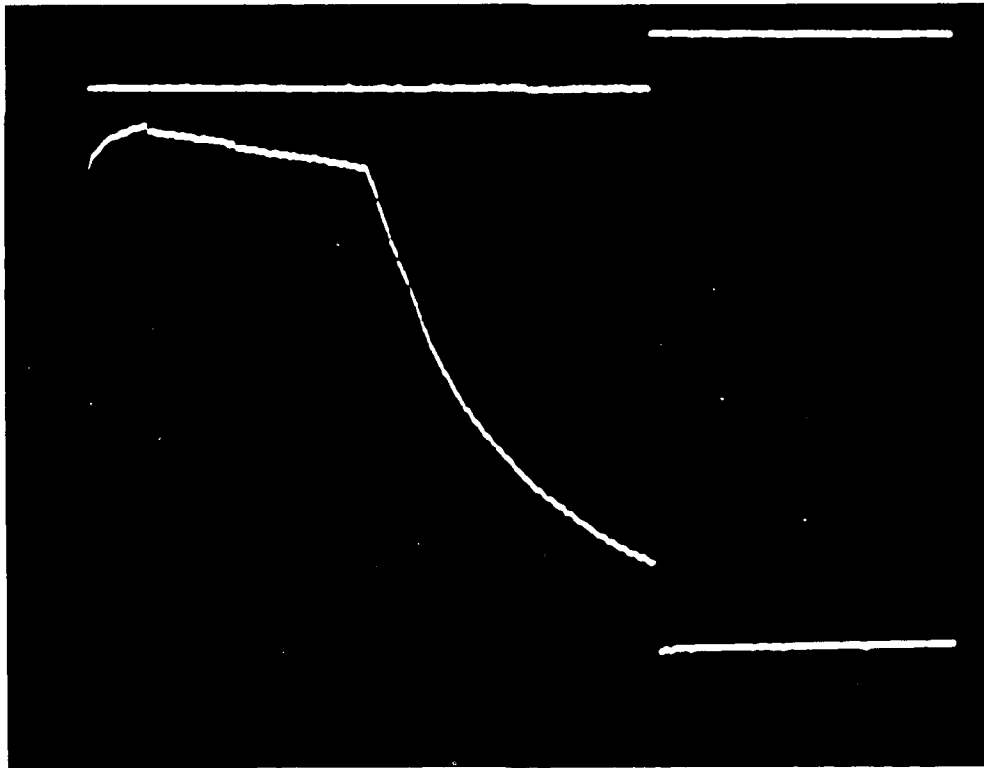
Figure : oscilloscope trace showing detection of a bunch of photons by the photodetector of one processor of a PE array. The bunch of photons is obtained from one single event amplified by a double microchannel plate image intensifier. The upper trace is a clock signal.

ii - Speckle is also an adequate phenomenon for our goal. The energy considerations in section III show that the power required in the speckle pattern is quite reasonable. If it is possible to use a 1 W near infrared diode laser, it is not even necessary to exercise particular care at saving energy in the diffusion process used to generate speckle. The way of obtaining the required spatial and temporal independance between the successive drawings at the various PEs by a suitable use of the speckle statistics is described elsewhere {7}.

V CONCLUSION :

A monochip 2-D array of processing elements can be provided with a spatially and temporally modulated random number generator by insertion in each processing element of a photosensitive area exposed to a randomly variable optical phenomenon. The case of individual photoevents amplified by microchannel plate image intensifiers and the case of speckle have been considered. Applications include in particular massively parallel implementation of simulated annealing algorithms on 2-D arrays of typically 256 x 256 binary images. The device is effectively a binary retina with randomly variable input and local processing power.

References :

1. G.M. Morris, Opt. Engin. 24, 86 (1985).
2. G.M. Morris, Appl. Opt. 23, 3152 (1984).
3. J. Marron, A.J. Martino and G.M. Morris, Appl. Opt. 25, 26 (1986).
4. S. Kirkpatrick, C.D. Gelatt, jr. and M.P. Vecchi, Science 220, 671(1983).
5. M. Metropolis et al., J. Chem. Phys. 21, 1087(1953).
6. P. Garda, B. Zavidovique and F. Devos, Proc. 11th ESSIRC, Toulouse (France), 1985.
7. F. Devos, P. Garda and P. Chavel, submitted for publication.

MONDAY, MARCH 16, 1987

PROSPECTOR/RUBICON ROOM

8:00 PM–9:30 PM

ME1–10

POSTERS: SESSION 5

Free Space Optical Interconnects by Cascaded Holographic Elements

W.J. Hossack

Wheatstone Laboratory, King's College London,

The Strand, London WC2R 2LS, UK.

Abstract

Cascaded holographic system for optical coordinate transformations is applied to free space optical interconnect. The design criteria for afocal systems, performing conformal mappings is presented.

Introduction

In parallel architecture optical computers the wiring interconnections of digital electronics must be replaced by optical routing components, typically in free space. This free space interconnect problem has been approached by aspheric glass optical elements, [Lohmann 85], for implementation of the perfect shuffle, and multi-facet computer generated holographic elements for a 16 gate optical sequential logic system, [Jenkins 84]. Gate array technology is approaching 10^6 gate per device, [Sawchuck 86], as a 2-D array, presenting considerable interconnect problems. Interconnection by aspheric optical elements are limited to specific interconnection patterns, are technically difficult to produce, and frequently require complex optical systems. The multi-facet hologram technique is very flexible, allowing essentially arbitrary interconnections, but requires one computer plotted "facet", or lens elements per gate. Due to diffraction effects and the resolution limit of CGH plotting equipment, is unlikely that each facet can be reduced significantly below 100μm by 100μm. This results in prohibitively large optical elements, (10cm by 10cm for 10^6 gates), and extensive production times, sepically for production by e-beam lithorgaphy systems. However if totally arbitrary interconnection of gate is not required, the techniques of optical coordinate transformation, and in particular holographic phase plates, [Bryngdhal 74], may be utilized, where either the whole array, or a sub-array of gates may be treated as input distribution or "image" which is to be mapped to a output plane under going a single or series of coordinate transformations.

This paper presents an afocal extension of the simple coordinate transformation system, initially developed by Bryngdhal, which allows cascading of transformation elements. The conditions for the production of twin element afocal transformations and the applications to optical interconnects.

Simple Afocal System for Optical Coordinate Transformation

Coordinate transformation involves the mapping of any point (x_1, y_1), in plane P_1, to a point (x_2, y_2) in plane P_2, separated by a distance $d_{1,2}$, where the transformation is described by,

$$x_2 = X_{1,2}(x_1, y_1)$$
$$y_2 = Y_{1,2}(x_1, y_1)$$

where $X_{m,n}(x_m, y_m)$ and $Y_{m,n}(x_m, y_m)$ describes the desired coordinate transformation between planes P_m and P_n. To perform afocal coordinate transformation, two holographic elements are required, [Hossack 86]. The elements are designed by a simple geometric ray optics model, which can be shown to be equivalent to the stationary phase approximation.

If we consider a plane wave incident on a phase functions, $W_{1,2}(x_1, y_1)$, the direction of the propagated wave will be given by its directional cosines, $(\alpha_1, \beta_1, \gamma_1)$, which for the small angle approximation will give,

$$W_{1,2}(x_1, y_1) + kz = k(\alpha_1 x_1 + \beta_1 y_1 + \gamma_1 z)$$

where $k = 2\pi/\lambda$. This gives the phase function as a pair of differential equations, as,

$$\frac{\partial W_{1,2}(x_1, y_1)}{\partial x_1} = k\alpha_1$$
$$\frac{\partial W_{1,2}(x_1, y_1)}{\partial y_1} = k\beta_1$$

Now by substituting for α_1 and β_1, in terms of (x_1, y_1) and (x_2, y_2), and again applying the small angle approximation, we get expressions for the phase function, in terms of the transformation, given by,

$$\frac{\partial W_{1,2}(x_1, y_1)}{\partial x_1} = [X_{1,2}(x_1, y_1) \quad x_1] \cdot \frac{k}{d_{1,2}}$$
$$\frac{\partial W_{1,2}(x_1, y_1)}{\partial y_1} = [Y_{1,2}(x_1, y_1) - y_1] \cdot \frac{k}{d_{1,2}}$$

To form a afocal system, the out rays must be parallel, requiring a second holographic filter, denoted as $W_{2,1}(x_2, y_2)$. Since in any linear optical system the ray paths are reversible, the second filter must implement the inverse coordinate transformation, $(x_2, y_2) \rightarrow (x_1, y_1)$, given by,

$$x_1 = X_{2,1}(x_2, y_2)$$
$$y_1 = Y_{2,1}(x_2, y_2)$$

Therefor by analogy with the first filter, the phase function of the second, phase correction filter, is given by,

$$\frac{\partial W_{2,1}(x_2, y_2)}{\partial x_2} = [X_{2,1}(x_2, y_2) \quad x_2] \cdot \frac{k}{d_{1,2}}$$
$$\frac{\partial W_{2,1}(x_2, y_2)}{\partial y_2} = [Y_{2,1}(x_2, y_2) \quad y_2] \cdot \frac{k}{d_{1,2}}$$
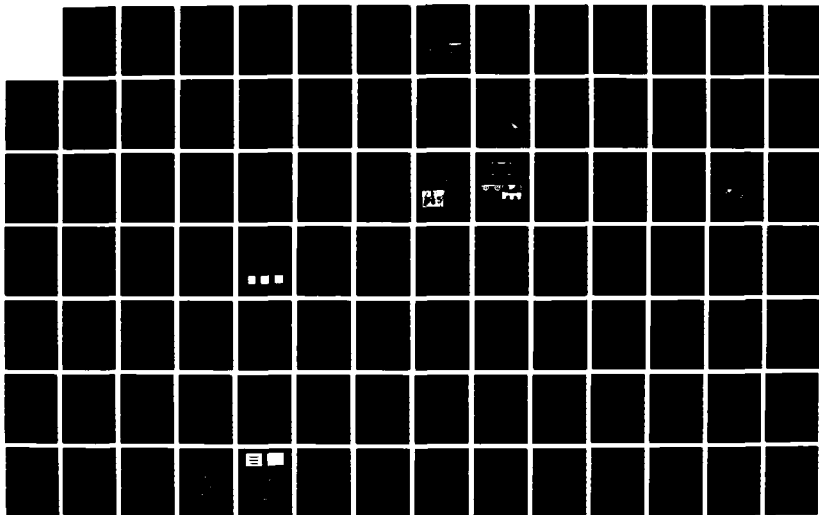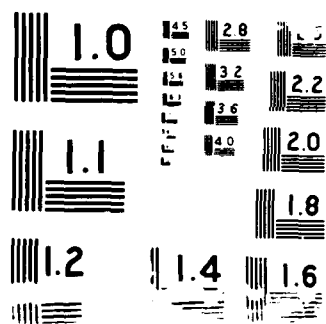
Conditions of the Existence of the Phase Filters

There are two conditions on the existence of the phase filters:

Firstly for the inverse to exist, the transformation must be point to point, ie. conformal. Secondly the phase function $W_{1,2}$ and $W_{2,1}$ are two dimensional continuous functions, so that,

$$\frac{\partial^2 W_{1,2}}{\partial x_1 \partial y_1} = \frac{\partial^2 W_{1,2}}{\partial y_1 \partial x_1}$$

So that the allowable transformations for which the phase function can be found are limited by the relation,

$$\frac{\partial X_{1,2}}{\partial y_1} = \frac{\partial Y_{1,2}}{\partial x_1}$$

Therefore if both of these conditions are valid, then an afocal two hologram system can be produced. It should be noted that if the conformal requirement in not valid, then a single hologram system may still be produced, but the second hologram to from the afocal property does not exist.

Since the transformation system is afocal, then multiple systems may be cascaded together, (also producing an afocal system), to perform transformations that do not obey the above existence criteria, and in particular any conformal transformation can be implemented, provided it can be decomposed into a series of conformal transformations, each of which obey the above criteria.

Applications to Optical Interconnections

By considering a 2-D array of logic gates, (or a sub-section of), as an "image", a single hologram, or cascaded sequence of holograms can be used to interconnect the gates provided the transformation can be implemented under the above conditions. It should be noted that the requirement for the system to be conformal, a sever restriction for interconnects, may be relaxed on the last holographic transformation element of a cascaded system, provided there is no requirement for the whole system to be afocal.

Conventional amplitude CGH filters have very low diffraction efficiency, typically a few percent. However the original CGH may be copied, holographically, [Fairchild 82], to form a COHOE, which with the use of Dicromated Gelatine, has the potential for diffraction efficiency approaching 100 %.

This method of interconnections does not allow totally arbitrary routing of beams, thus imposing restrictions on the logical interconnections. In particular the usual "fan-in" and "fan-out" systems are difficult to implement. This may be tolerated at the considerable reduction in the CGH plotting overhead, and possible size reductions.

Acknowledgement

References

Bryngdhal, O.; 1974   Optics Comm. **10** 164- 168

Fairchild, R.C., & Fienup, J.R.; 1982   Opt. Eng. **21** 133   140

Hossack, W.J., Darling, A.M., & Dahdouh, A.; 1986   (Submitted to Optica Acta)

Jenkins, B.K., & Sawchuck, A.A.; 1984   App. Opt. **23** 3455- 3464

Sawchuck, A.A., & Jenkins, B.K.; 1986   Proc. SPIE **625**

Lohmann, A., Stoke, W., & Stucke, G.; 1985   OSA Optical Computing, Tahoe, March 18 20

Optical Interconnect Complexity Limitations for Holograms Fabricated with Electron Beam Lithography

Michael R. Feldman and Clark C. Guest

Department of Electrical Engineering and Computer Sciences
University of California, San Diego
La Jolla, Ca 92093

Introduction

As VLSI chip sizes and device densities increase, signal communication limitations begin to dominate system perfomance[1,2]. By replacing particular electronic transmission lines with optical signal paths, perfomance can be improved in terms of speed, alleviation of clock skew and a reduction in silicon area devoted to interconnects. These improvements can be achieved for on chip or chip-to-chip communication, but become dramatically evident for wafer scale communication. At the wafer scale level, global interconnections are infeasible to perform electronically due to the line lengths and complexity involved. In addition, arrays of parallel processing elements are utilized in order to alleviate problems of yield. The use of optical interconnections would allow for the interconnection of these processing elements in a global highly parallel manner (such as hypercube and butterfly machines) that are difficult to achieve electronically[3]. However, in this case the optical system must be able to handle a large number of sources, each requiring a large fanout. Highly complex interconnection schemes can be achieved with the use of a holographic optical element (HOE) for free space optical interconnections.

The use of computer generated HOE's provides a hologram fabrication process compatible with integrated circuit processing procedures. The fabrication method employed consists of etching a 1/4 wavelength thick silicon dioxide layer on a silicon wafer substrate and overcoating the wafer with aluminum[4]. In this way a reflectiive surface relief hologram can be fabricated in an I.C. compatible procedure.

In this paper a particular design method for computer generated HOE's fabricated with electron beam lithography is first described. The interconnect complexity limitations of these HOE's for two types of architectures are then discussed. Finally some experimental HOE's are described.

System Design

An important parameter in determining interconnect complexity is the distance a signal beam can be deflected by the HOE. The communication distance, C, is defined as the distance between a diffracted HOE output spot and the center of the undiffracted light in the VLSI plane (Fig. 1). C_x and C_y, the x and y components of C, can be evaluated by using the thin grating equation and by considering the

largest angles involved in the imaging system,

$$C_x = h \tan\{ \sin^{-1}(0.5 \, BW \, \lambda - 0.5 \, \theta_x) \} - h \tan(0.5 \, \theta_x) \qquad (1)$$

where h is the distance between the HOE and the VLSI plane, BW is the spatial bandwidth of the wavefront diffracted by the HOE, λ is the optical wavelength and θ_x is the source divergence angle in the x-direction. (Equation (1) was derived for source beams with normal incidence on the HOE and is an approximation for sources incident at large angles.) For a wavelength of 0.8 μm and source divergence angles of 15° and 30°, a hologram bandwidth of 2000 lines/mm is needed to allow communication distances of $C_x = 0.37$ h and $C_y = 0.77$ h .

In order to produce high bandwidth HOE's, electron beam lithography is used for fabrication and the binary kinoform method for encoding. This encoding method enables holograms to achieve bandwidths of up to $1/X_{min}$ where X_{min} is the minimum feature size of the hologram fabrication process. A multi-element design in which the HOE is divided into subholograms (or elements) such that each source illuminates only a single element can be employed to provide space-variant imaging between sources. Each element is further divided into facets such that each facet diverts light to a particular detector. This multi-element multifacet scheme minimizes amplitude quantization effects which would be severe for complex images encoded with the binary kinoform method.

Interconnect Complexity Limitations

Interconnect complexity can be analyzed in terms of the maximum number of laser sources, N_s , each with a specified fanout, F, that can be handled by a holographic interconnect scheme. Two architectures (Fig. 1, Fig.2) will be evaluated for the binary kinoform, multi-element, multifacet HOE design method.

The first architecture utilizes a single reflective HOE. The maximum number of space-variant connections is limited by the hologram size divided by the size of the subholograms. For laser beams directed at optimum HOE incident angles, communication distance components of one half of the corresponding chip dimensions are needed for each source to be able to address any chip location. Combining this information with equation (1), restricting HOE dimensions to twice the chip dimensions and assuming typical values for wavelength and source divergence angles yields:

$$N_s < 33 \text{ for } X_{min} = 0.5 \, \mu m .$$

The maximum source fanout is limited by the size of the smallest facet diverting light to a particular detector. The diffraction limited spot produced by this facet must be smaller than the detector size. Diffraction limited analysis reveals that a maximum fanout of 30 can be achieved for 15 μm by 15 μm detectors.

If space invariant imaging is utilized, the interconnection limitations are determined by the minimum separation of adjacent semiconductor lasers and

detectors.   The HOE consists of a single element with F facets.   Assuming a minimum source separation of 100 μm, and a minimum detector spacing of 30 μm, a typical system for a 1.5 cm by 1.5 cm substrate with normally incident laser source beams could accommodate over 5,000 sources, each with a fanout of 30.

The second architecture employs a single transmissive HOE which, as in the previous case, is composed of many subholograms[5].   However, in addition to the multifacet subholograms placed over each source, the HOE also contains subholograms placed directly above each detector.   Since each signal beam passes through the HOE twice, the system can accommodate larger source numbers (by keeping h_1 small) and still achieve large communication distances.

An analysis of this system, taking into account diffraction limits and equation (1), reveals that N_s is directly proportional to the chip dimensions.   For typical laser source properties, a HOE minimum feature size of 0.5 μm, detector sizes of 15 μm x 15 μm and chip dimensions of 4 cm. x 2 cm., this architecture can provide space-variant interconnections for 300 normally incident sources each with a 25:1 optical fanout.

Both the design method and the architectures discussed can be modified to allow for more complex interconnections.   For example an off-axis encoding method would allow for larger fanout but would reduce hologram bandwidth thereby reducing the maximum communication distance for the same HOE-to-VLSI separation.   A multilevel kinoform (as opposed to the current binary process) would produce similar results, but with higher diffraction efficiency.

Experimental Work

An experimental HOE prototype was designed on a Calma CAD station and fabricated with an electron beam lithography system.   Designed for the architecture of Fig. 1, it performs an optical fanout from one source to five detectors.   When the HOE is placed 2 cm. above the VLSI plane, the five output spot sizes are each less than 15 μm x 15 μm. The minimum feature size of the hologram is 0.75 μm and the HOE SBWP is 1.5×10^9.

Several other HOE's are currently being designed.   One is designed to optically interconnect 64 sources each with 4 detectors in a spatially invariant pattern.   Also multilevel kinoform HOE's are planned.

Summary

The integrated circuit industry has a need for a communication technology that can perform highly complex interconnections.   In this paper we have discussed the ability of free space optical interconnects to meet this need for a particular HOE design method.   Experimental HOE's are being fabricated in order to verify these results.

References

1. J. W. Goodman, F. I. Leonberger, S. Y. Kung and R. A. Athale, "Optical Interconnection for VLSI Systems," Proc. IEEE, vol. 72, pp. 850-866, 1984.

2. K. C. Saraswat, F. Mohammadi, "Effect of Scaling of Interconnections on the Time Delay of VLSI Circuits," IEEE Trans. Electron Devices, vol. ED-29, no. 4, pp. 645-650, 1982.

3. L. Bergman, et. al., "Applications and Design Considerations for Optical Interconnects in VLSI," Proc. SPIE, vol. 625, pp. 117-126, 1986.

4. N. C. Gallagher, Jr., J. C. Augus, F. E. Caffield, R. V. Edwards and J. A. Mann, Jr., "Binary Phase Digital Reflection Holograms: Fabrication and Potential Applications," Appl. Opt., vol. 16, pp.413-417, 1977.

5. This version of this architecture was brought to my attention by Eric Bradley, EECS Department, University of California, San Diego.
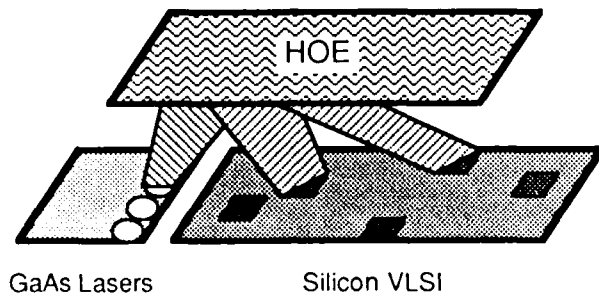
Figure 1) Single Reflective HOE architecture.
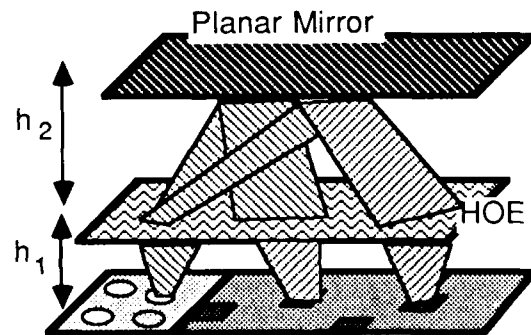
Figure 2) Architecture allowing more complex space-variant imaging

Design of Computer Generated Holograms for E-Beam Fabrication

by Means of a Computer Aided Design System

H. Farhoosh and S. H. Lee

Department of Electrical Engineering and Computer Science
University of California, San Diego
La Jolla, CA 92093

I. INTRODUCTION

Computer generated holograms (CGH) have many potential applications, such as holographic optical elements (HOE), pattern recognition, optical interconnect, and display of abstract objects[1-5]. In order to use these potentials to a full extent, a CGH must be recorded at very high resolution with precision on a large area. High resolution and large area require recording devices that have an extremely large space bandwidth product (SBWP). Electron beam lithography systems can satisfy these requirements. E-beam systems have a SBWP of order of 10^{10} and they are capable of direct writing patterns of submicron features with very high accuracy.

In order to record a CGH by means of an e-beam system one needs to prepare a large amount of data that is compatible with the input requirements of the e-beam system. The purpose of this paper is to show how the capabilities of a computer aided design (CAD) system can be used for e-beam fabrication of computer generated holograms.

The material in this paper is organized as follows. In section II general characteristics of CAD systems are discussed. In section III we present a procedure for the CGH encoding process on a CAD system. Section IV contains some design considerations. Finally, section V contains the concluding remarks.

II GENERAL CHARACTERISTICS OF CAD SYSTEMS

CAD systems have become a valuable and necessary design tool in many areas of engineering. CAD systems are used today to design electrical, mechanical, electronic, optical, and many other components and systems. Sophisticated CAD systems, such as CALMA, are equipped with the hardware, firmware, and software resources necessary for handling complicated designs such as VLSI. Fast CPU's, easy to use input

devices, high resolution output devices, and storage media of large capacity and fast access time are examples of their hardware resources. Their firmware resources include dedicated ROMs to handle many design and display tasks, while on the software part they provide menu-driven interface, graphical programming languages, and a large library of very usefull functions. In the next section we shall describe how these resources can be used to design a CGH for e-beam fabrication.

III. DESIGN PROCEDURE

Steps required for designing a CGH depend, to some extent, on the type of the CGH. There are basically two classes of CGH. The first class of CGH includes wavefronts that are known in analytic form, e.g. HOE's for optical testing. Wavefronts of the second class need to be computed numerically and they are obtained by sampling an object and computing a Fresnel or Fourier transform of the sampled pattern. Designing of the first class of CGH involves evaluation of an analytic function describing the hologram at discrete points and conversion of the numerical values to geometrical data. Usually evaluation of the analytic function is not computationally intensive and it can be performed on the CAD system directly. The second class of CGH, on the other hand, require computationally intensive FFT operations which are best carried out on a number crunching computer. Conversion of complex numbers into real values can be done on this computer as well. The results from these computations are then transferred to the CAD system.

On the CAD system the numerical data serve as input to a program that encodes the data in the form of geometrical patterns. The algorithm that converts numerical data into geometrical patterns depends on the particular encoding scheme. For example, if a CGH is to be encoded by Lohmann's method[6], the phase and amplitude values of the wavefront at the hologram plane can be computed on a number crunching computer. The phase and amplitude data are then transferred to the CAD system where they are encoded in the form of geometrical patterns. The encoding algorithm reads these data a row at a time and calls system provided functions to place rectangular apertures of fixed width and a height proportional to the amplitude values at designated locations determined by the phase values. Therefore, the program only needs to keep track of the position and size of the rectangles; the CAD system would take care of the actual placement of rectangles as geometrical objects. The geometrical information is stored in a database that can be displayed or plotted at any desired magnification for checking and verification. These

graphical data are then converted into a format compatible with e-beam input requirements. This conversion, called data fracturing, consists of decomposition of geometrical patterns into graphical primitives such as trapezoids. A CAD system, by means of its dedicated firmware and tailored software performs this task automatically and efficiently, which is otherwise very cumbersome to do on a general purpose computer. The fractured data are then fed to the e-beam system which writes them on a resist and chrome coated glass or quarts substrate. The substrate is then chemically processed and the chrome under the exposed regions of the resist is etched to form the final hologram.

The design procedure described above has been implemented for six encoding schemes on a CALMA design station at UCSD. As a result of this implementation we have arrived at some design cosiderations and trade offs, which are discussed in the next section.

IV. DESIGN CONSIDERATIONS

Because of the enormous information content of holograms, when designing a CGH one should be careful to keep the size of data to manageable proportions. The amount of graphical data describing a CGH depends on the encoding method used in the design algorithm. A comparison of CGH encoding schemes as a function of the CGH quality (size and bandwidth of the reconstructed image, SNR, and diffraction efficiency) and computation requirements is given in another paper which is submitted for this conference[7]. Here we shall point out general rules that must be observed in order to keep the amount of data in manageable proportions.

It is generally desired to have hologram patterns that generate the least number of primitive shapes (NPS) because the amount of graphical data is directly proportional to the NPS. Orientation of these patterns is important as well, e.g. patterns consisting of rectangular shapes oriented along the x and y axes generate much less data than patterns consisting of curved lines in arbitrary directions.

The amount of graphical data is also a function of the application program (on the CAD system) that generates the data. Normally a graphical database is organized heirarchically such that graphical primitives are clusterred into small structures, and small structures are grouped into larger ones. This type of "referencing" helps to reduce the size of graphical data considerably. However, this heirarchical structure

elongates data access and processing time significantly. Therefore, there is a trade off between processing speed and amount of data.

V. CONCLUSIONS :

Electron beam lithography systems make the recording of CGH's with large SBWP and submicron resolution possible. CAD systems can be used in the design and data preparation of holograms with remarkable ease, provided that the design considerations discussed in the previous section are observed.

Finally, we would like to emphasize that a significant improvement in CGH design and fabrication can be achieved if the CAD system and the e-beam machine are both available at the same location and connected together. In this case CGH design and recording can be carried out in parallel, thus reducing graphical data storage requirements as well as the total fabrication time by a significant amount. Furthermore, and perhaps more importantly, more complex CGH's can be fabricated because the graphical data can be generated in segments, i.e. as the e-beam is writing the previously generated segment, the CAD system can generate the next patch of data.

REFERENCES

1. S. M. Arnold, "Electron Beam Fabrication of Computer Generated Holograms," Opt. Eng., vol. 24, 803, (1985).

2. K. M. Leung, J. C. Lindquist, and L. T. Shepherd, "E-beam Computer Generated Holograms for Aspheric Testing," Proc. of SPIE, vol. 215, 70, (1980).

3. R. A. Athale, C. L. Giles, and J. A. Boldgett, "Use of E-beam Written CGH in Optical Pattern Recognition," Proc. of SPIE, vol. 437, 48, (1983).

4. J. L. Freyer, R. J. Perlmuter, and J. W. Goodman, "Digital Holography : Algorithms, E-beam Lithography, and 3-D Display," Proc. of SPIE, vol. 437, 38, (1983).

5. M. R. Feldman and C. C. Guest, "Automated Design of Holographic Optical Elements for Interconnection of Electronic Circuits'" OSA annual meeting, Seattle, Washington, Oct. 19-25, (1986).

6. B. R. Brown and A. W. Lohmann, "Complex Spatial Filtering with Binary Masks," App. Opt., vol. 5, 967, (1967).

7. H. Farhoosh, M. R. Feldman, S. H. Lee, C. C. Guest, and Y. Fainman, "A Comparison of Encoding Schemes For E-beam Fabrication of Computer Generated Holograms," Submitted to the OSA topical meeting on optical computing, Lake Tahoe, Nevada, March 16-18, (1987).

Introduction

ME4-L

An important aspect in the design of a 2-D parallel processor is efficient interconnections. Crossbar networks(CBN) are considered to be the most desirable candidates because every processor can communicate with every other processor without conflicts. Implementing a large 2-D CBN, however, is very difficult[1]. A Clos network(CN)[2], Clos(p,q,r), as shown in Fig. 1 is proposed to replace the CBN. In Fig. 1, each block represents a feasible subcrossbar of medium size. The CN has the same characteristics as the CBN. The CN is a nonblocking network with which it is always possible to connect together an idle input-output pair of processors without disturbing existing connections, if $q \geq 2p-1$ is satisfied. Both one-to-one and one-to-many connections are available in the CN. The most important factor is that the number of switching elements has been reduced in the CN($\sim$10:1 for a 1000x1000 network, as compared with the CBN). The difficulty of determining connections, however, is a major problem to be overcome[3,4].
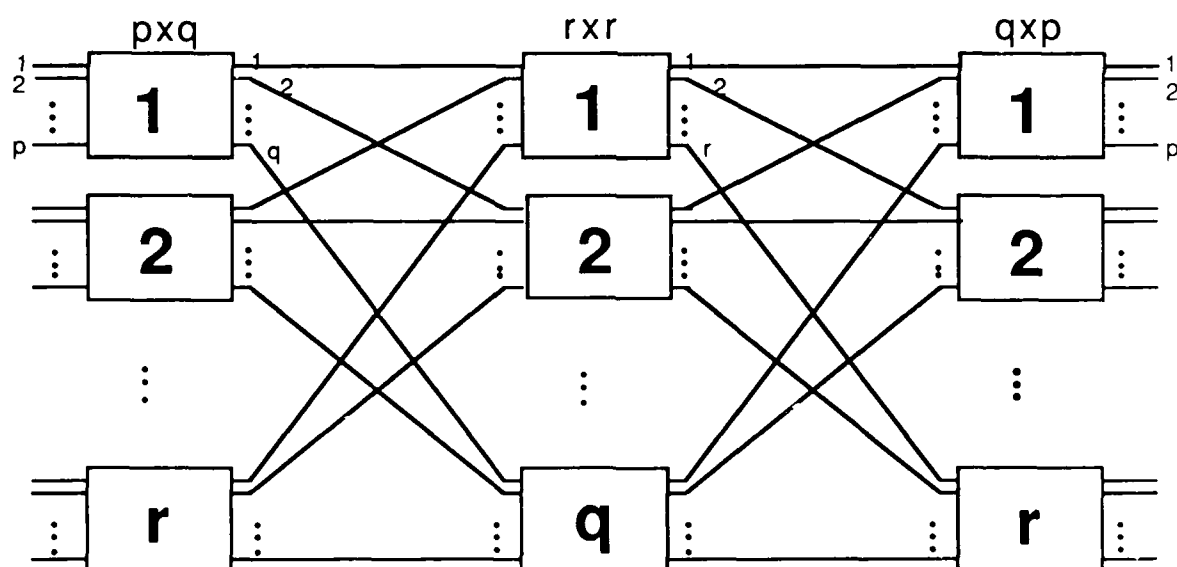
Fig. 1   A Clos(p,q,r) network

98

Routing Algorithm

We have come up with a straightforward algorithm to route the CN connections. This algorithm can be performed either by a uniprocessor or by parallel processors. An example is shown in Fig. 2, where the rows of numbers at the bottom of the figure are elements of the same output group(OGi, i.e., output elements in each subcrossbar of the last stage). The arrows indicate that we choose one element from each row(i.e., from each output group) such that these selected elements which determine the output of the second stage(I2) are in different input groups(IGi). For example, 2,7, and 5 are chosen from the rows because they are in different input groups(IG1, IG3, and IG2, respectively). Once the elements in I2 have been chosen, the input elements of the second stage(I1) also have been determined, since the input elements of subcrossbars in the second stage come from different input groups. Therefore, the switching functions for the subcrossbars of the three stages(IGi to I1, I1 to I2, I2 to OGi) are determined. The way to select elements in I2 is not unique. In the example we could choose 2, 8, and 5 instead of 2, 7, and 5. This possibility for alternative choices gives the routing algorithm an ability to tolerate faults.
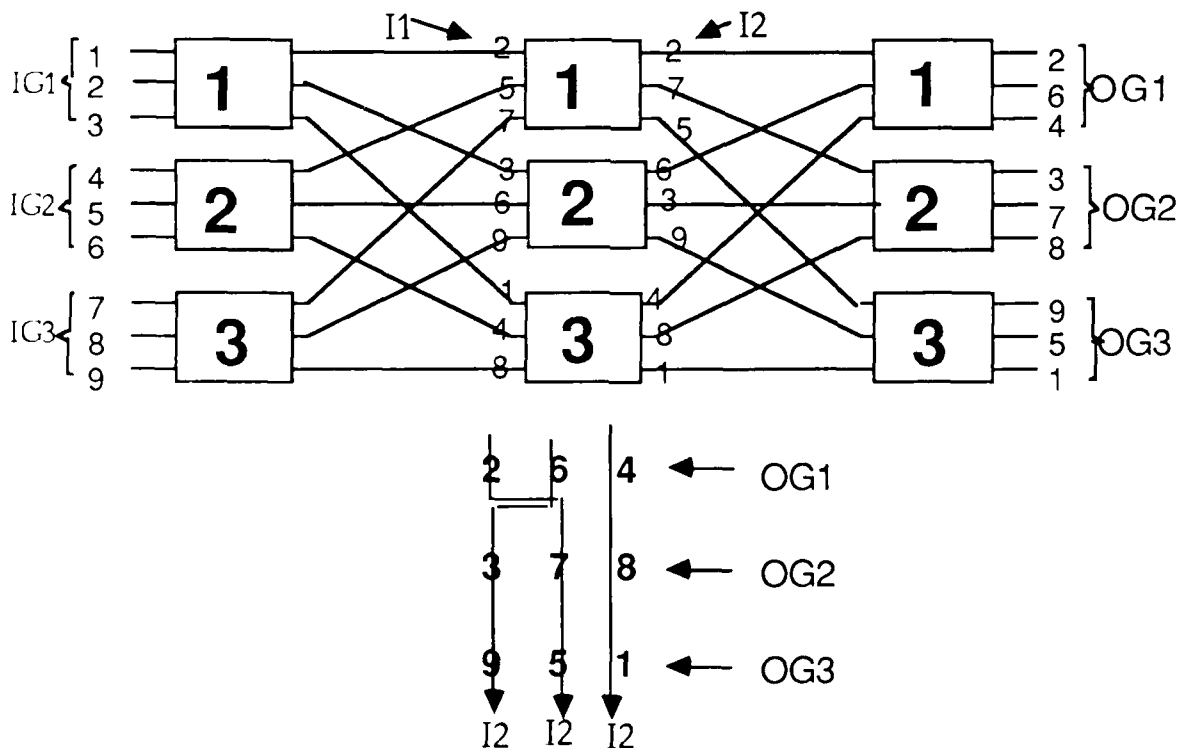
Fig 2 Routing scheme for a Clos(3,3,3) network

Optical Implementation

     Fig. 3 shows the optical set-up for one stage of a 2-D Clos(4,4,4) network where a 4x4 input array needs to be replicated 2x2 times, an LCLV acts as a switching medium with switching elements on the WRITE side, and a 4x4 lenslet array is used to collect the desired intermediate output elements. 2x2 elements are selected from each subquadrant(with 2x2 size) in a quadrant of the replicated input array(4x4 size) and these 2x2 elements go to the same subcrossbar of the second stage. The beauty of this scheme is that connections between stages turn out to be straight lines due to the replication of the input, so that Fig. 3 represents one stage of a cascadable system.
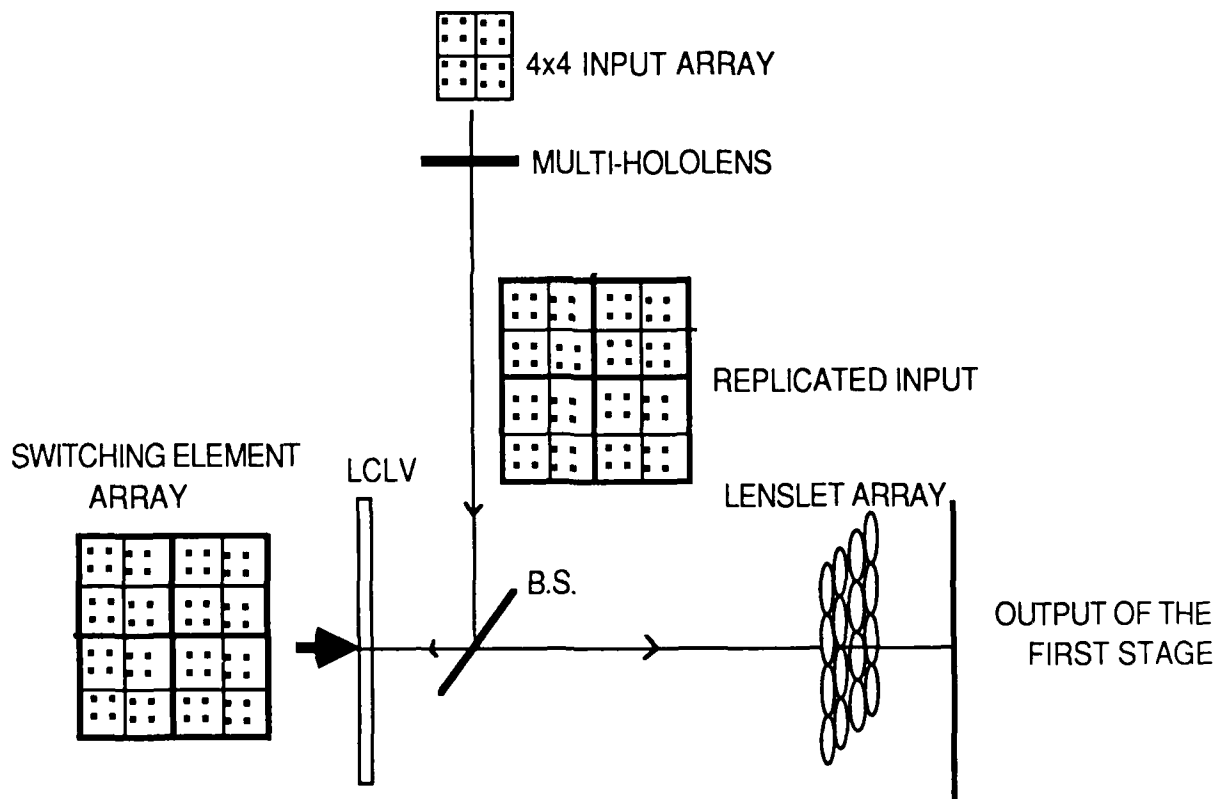
Fig. 3 Optical implementation of one stage of a Clos(4,4,4) network

2-D Neural Networks

For nxn neural cells, interconnections having n^4 degrees of freedom are needed. In other words a 2-D nxn CBN can provide an optimal solution for 2-D neural networks. The corresponding CN(i.e., with $q \geq 2p-1$) can also implement 2-D neural networks. For example, the Hopfield model could be implemented with a CN by replacing the subcrossbars in the last two stages with weighted subcrossbars followed by a threshold operator.

Summary

We have shown an optical implementation of the 2-D Clos network for which a routing algorithm has been found. A 2-D CN can also be used to realize 2-D neural networks, and furthermore, the neural connections are programmable. This will give neural processors the flexibility to perform more complex computations.

Acknowledgement

References

1. S. H. Lin, T. F. Krile and J. F. Walkup, "A 2-D Optical Crossbar Based on Triple Product Processing,"(presented at the 1986 OSA Annual Meeting, Seattle,Washington, WQ4).

2. C. Clos, "A Study of Non-Blocking Switching Networks," The Bell System Technical Journal, Volume XXXII, 406(1953).

3. V. E. Benes, Mathematical Theory of Connecting Networks and Telephone Traffic, Academic Press, NY(1968)

4. A. A. Sawchuk and B. K. Jenkins, "Dynamic Optical Interconnections for Parallel Processors," SPIE 625, Los Angeles, January(1986).

Optical Interconnects using Resonated Holograms

Stuart A. Collins Jr.

The Ohio State University
ElectroScience Laboratory
1320 Kinnear Road
Columbus, OH 43212

This paper deals with optical interconnects formed by thick holograms with mirrors to resonate them. A calculation for a design is submitted which predicts good efficiency, small angular bandwidth, and large numbers of interconnects with modest holographic capability.

The physical situation under consideration is shown in Figure 1 where we see input and output planes separated by the interconnecting mechanism which images individual spots in the input plane onto individual spots, randomly selected, in the output plane. It is desired that there be no overlap between light from adjacent, or other spots, and that the efficiency be high.

The motivation for this type of optical interconnect comes from numerical optical computing where the output of one calculation stage must be fed back to the optical CPU for continued calculation. This approach would also be interesting in image restoration or image transformation since it allows more efficient use of a hologram to convert one optical image into another.

Holograms by themselves are poor as optical interconnects. The thickest emulsion, Kodak 649F does not have a sufficiently large index variation to have more than one exposure approaching one hundred percent efficiency. Photochromic materials have greater thickness but the index variation is still too small to be effective.

The approach here is to put mirrors around the hologram as shown in Figure 2. Mirrors 1 and 2 reflect the light back and forth through the film increasing the number of passes and therefore the efficiency. However there is now a beam passing opposite to the incident beam which gives rise to a second diffracted beam. Mirror 4 is placed to contain that light. Mirror 3 is also added to match the output phase. The reflectivity of mirrors 2 and 4 is set as close as possible to 100%

A calculation assuming an energy conserving bleached hologram and based on a plane wave model and boundary condition matching shows that the reflectivity of mirrors 1 and 3 is related to the hologram diffraction efficiency, S, by the expression

$$R = \frac{(1-S^{1/2})}{(1+S^{1/2})}. \tag{1}$$

The angular half width can also be estimated. An approximate expression shows the angular half-width of the resonated hologram with optimum transmission is given by the expression

$$= (kL)^{-1/2}(R^{-1/4} - R^{1/4}) \tag{2}$$

where the optical wavelength is $= 2/k$ and L is the distance from the hologram to the mirrors.

The derivation also deals with requirements for maintaining resonance for all exposures at different angles and other practical considerations.
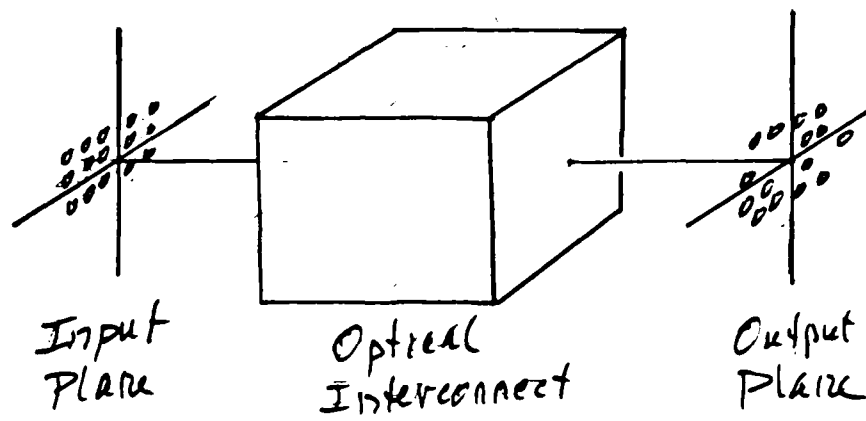
Figure 1
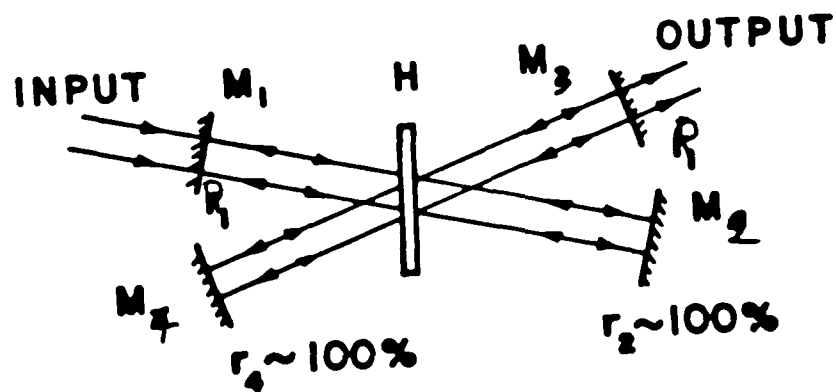
Figure 2

A Comparison Between Optical and Electrical Interconnections Based on Power and Speed Considerations

Michael R. Feldman, Sadik C. Esener, Clark C. Guest and Sing H. Lee

Department of Electrical Engineering and Computer Sciences
University of California, San Diego, La Jolla, CA 92093

Introduction

By replacing electronic transmission lines in very large scale integrated circuits with optical communication systems, increased system performance can be achieved.[1,2] Electronic transmission lines suffer from long signal propagation time and large power dissipation as lengths and fanout grow. On the other hand, optical systems are limited by the power dissipation of the injection laser sources and the optical detection speed, fundamentally governed by the supplied optical energy. In this paper, power versus speed tradeoffs are examined for both electrical and optical interconnections. Equations are derived that can be used to determine specific conditions for which optical systems become advantageous. These results are then applied to concurrent wafer scale integrated circuits (WSI) where power, speed, throughput, and cost advantages can be achieved by performing interconnections optically.

Electrical Interconnections

A model for CMOS VLSI interconnections is illustrated in Fig. 1[3]. R_L and C_L are the resistance per square and capacitance per unit area of the transmission line. C_o and C_{in} are the output and input capacitances of a minimum size CMOS gate. The switching energy of such an interconnection, defined as the energy needed to switch the state of the receiving inverter from one state to another and back, is given by,

$$E_{sw} = (M C_o + F C_{in} + L W C_L) V^2 \qquad (1)$$

where M is the gate width of the driving inverter expressed as a multiple of a minimum size inverter's gate width, L and W are the length and width of the line. F is the fanout and V is the supply voltage.

The interconnect delay time (defined as the time from $V_1 = 10\%$ V to $V_2 = 90\%$ V) was also estimated from the model of Fig. 1. By approximating the dynamic resistance of a CMOS inverter as the linear range of resistance of the two transistors in parallel, the total transmission line rise time can be calculated as,

$$0.89 R_L C_L L^2 + (1.1 V/(M I_o))(M C_o + F C_{in} + C_L L W) + 2.2 F C_{in} R_L L/W \qquad (2)$$

where I_o is the maximum current that can be sourced by a minimum size CMOS inverter gate. Each term was found by multiplying each capacitance with the sum of the resistances that occur before it on the transmission line[4]. The first term is due to the distributed RC of the line[5]. There are three additional terms associated with the driving gate charging the three capacitances in Fig. 1. The remaining term is due to the line resistance and the receiving gate's input capacitance.

A computer simulation program was developed in order to determine the validity of this interconnection model. Employing the SPICE[6] circuit simulation program as a subroutine, the computer program can determine rise time and switching energy from user inputs of fanout, linelength and SPICE process parameters. The simulation results agree to within 20% of the analytic estimations over a wide range of transmission line properties.

Note that increasing the driving gate size, M, in equation (2), increases the current sourcing capabilities of the driving gate, thereby decreasing the delay time. (The delay time and switching

small line width, W, the delay time can be further reduced by decreasing the line resistance with a larger W. As W increases, though, the term due to the driving gate charging the line capacitance causes the total delay time to increase. The minimum delay time is reached when both M and W in equation (2) approach infinity with M much larger than W,

$$\tau_{min} = 0.89\, R_L\, C_L\, L^2 + VC_o/I_o \tag{3}$$

Of course this case is impractical since it requires infinite area and infinite energy, as described by equation (1). For a fixed energy, M and W are related by equation (1), and the minimum delay time occurs for specific values of M and W. (Fig. 2 is a plot of time versus M and W for $E_{sw} = 30$ pJ.) Based on these values, speed performance limitations can be found for given energy constraints. These speed limitations were found by setting $d\tau/dM$ to zero in equation(2) and solving the resulting polynomial numerically. Results of this analysis were used to plot minimum delay time as a function of power dissipation for specific values of line length and fanout.

In order to compare these limitations with those of optical interconnections, optical communication systems must be evaluated.

Optical Interconnections

We will consider free space optical communication systems consisting of semiconductor laser sources, a holographic optical element and silicon photodetectors. The detectors to be analyzed, illustrated schematically in Fig. 3, are CMOS compatible optical gates.[2] The required optical switching energy is given by

$$E_{sw} = F\,(\,C_{pd} + C_{in}\,)\, V\, h\, \nu\, /\, (\eta\, q) \tag{4}$$

where C_{pd} is the photodiode capacitance, h is Plank's constant, ν is the optical frequency, q is the electronic charge and η, the energy conversion efficiency, is the optical power absorbed by the photodiode divided by the electrical power input to the laser source. Thus,

$$\eta = \eta_l\, \eta_h\, (1 - e^{-\alpha d}) \tag{5}$$

where η_l is the efficiency of the lasers, η_h is the efficiency of the hologram, α is the absorption constant and d is the thickness of the detector active region. Ideally, the optical delay, τ_o, is determined by the optical energy provided to the detector (since laser diode speeds are typically much faster) and is given by,

$$\tau_o = E_{sw}\, /\, (2\, P\,) \tag{6}$$

where P is the power supplied to the laser sources.

Comparison

Using the equations derived above, optical and electrical communication links can be compared in terms of both power and speed. This comparison was based on typical present day parameters. Values for C_{in}, C_o, R_L and C_L were obtained from SPICE parameters for 3 µm CMOS provided by the MOS Implementation Service (MOSIS)[7] process. A laser efficiency of 30%, a hologram efficiency of 30%, a wavelength of 0.8 µm and a 4 µm thick, 9 µm square detector size was assumed. From equation (4), an optical switching energy of 118 pJ is required.

Figure 4 is a plot of power versus speed for both a polysilicon CMOS transmission line and an optical interconnection system. Both links are for a 3 mm communication distance with a 1:10 fanout. Note that for low speeds, electronic lines consume less power when operated at the same speed as an optical system. At higher speeds (delay times < 21 nsec in this case), optical systems consume less power. In accordance with equation (3), electrical lines cannot transmit data with delay times of less than 16 nsec at any power. A lower bound on the optical interconnect delay time is determined from equation (6) by the maximum electrical power that can be supplied to the laser source. This time limit is less than 1 nsec for a maximum input power of 100 mW.

Additional results for metal interconnection lines will be presented at the conference.

Application to WSI

From the above analysis, it is clear that as line length and interconnect complexity increase, optical communication systems become advantageous in terms of power and speed considerations. These advantages are dramatically evident in wafer scale integrated systems where both large fanout and long communication links are required.

In concurrent VLSI systems, composed of N processing elements, the throughput or perfomance, r, is given by,

$$r = N / (\tau + \tau_{pe}) \tag{7}$$

where τ_{pe} is the latency of the slowest processing element and τ is the interconnect delay time. The cost per performance, C/P, is given by,

$$C/P = (A \$ + A_I \$_I) (\tau_{pe} + \tau) \tag{8}$$

where A and $ are the area and cost per minimum feature size of each processing element and A_I and $\$_I$ are the average area and cost per minimum feature size of each communication link. For small values of N, throughput can be increased by increasing N, thereby decreasing τ_{pe} , A$ (assuming a constant wafer area) and C/P. As N increases though, τ and $A_I\$_I$ increase at increasing rates resulting in higher C/P. However, by performing interprocessor communication optically, while maintaining local interconnections as electrical transmission lines, N can be increased with little effect on $A_I\$_I$ and τ. Therefore increasing N can result in increased throughput and decreased C/P. The processing element size should be chosen such that τ is approximately equal to τ_{pe} for the same power. Since the average maximum line length and fanout of a single processor can be expressed as a function of area, the above analysis can be performed to determine the processor grain size for which $\tau = \tau_{pe}$.

Conclusions

Tradeoffs between power and speed have been analyzed for both electrical and optical interconnections. In general, for sufficient values of interconnect length and fanout, optical systems consume less power when operated above a particular speed. The value of this speed is determined by line length, fanout and integrated circuit process parameters. Additional advantages of throughput and cost per performance can be achieved by applying optical interconnects to concurrent wafer scale integration.

References

1. J. W. Goodman, F. I. Leonberger, S. Y. Kung and R. A. Athale, "Optical Interconnections for VLSI Systems," Proc. IEEE, vol. 72, pp. 850-866, 1984.
2. L. A. Bergman, et. al., "Holographic optical interconnects in VLSI," Optical Engineering, vol. 25, pp. 1109-1118, 1986.
3. C. Mead and L. Conway, Introduction to VLSI Systems, Menlo Park, Calif., Addison-Wesley, 1980.
4. L. A. Glasser and D. W. Dobberpuhl, The Design and Analysis of VLSI Circuits, Menlo Park, Calif., Addison-Wesley, 1985.
5. K. C. Saraswat and F. Mohammadi, "Effect of Scaling of Interconnections on the time delay of VLSI Circuits," IEEE Trans. Electron Devices, vol. ED-29, pp.645-650, 1982.
6. T. Quarles, A.R. Newton, D.O. Pederson, A. Sangiovanni-Vincentelli, SPICE Version 3A7 User's Guide, University of California, Berkeley, September 23, 1986.
7. The MOSIS System (what it is and how to use it), Report No. ISI/TM-84-128, Information Sciences Institute, University of Southern California, Marina del Rey, CA. 90292, March 1984.
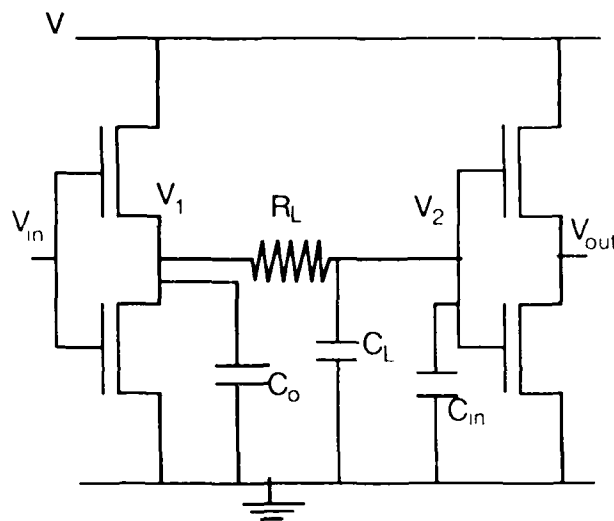
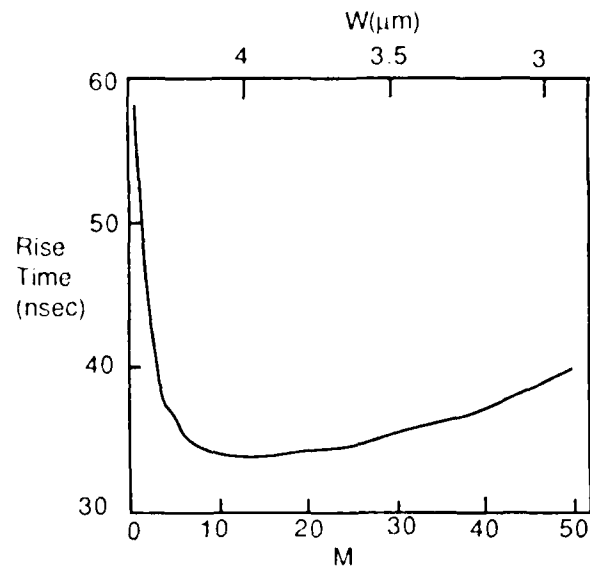Fig. 1. Schematic diagram of electrical interconnection of two CMOS gates.

Fig. 2. Interconnect delay time as a function of driving gate size, M, and line width, W, for constant energy loss. $E_{sw} = 30$ pJ, L = 3 mm, F=10 .
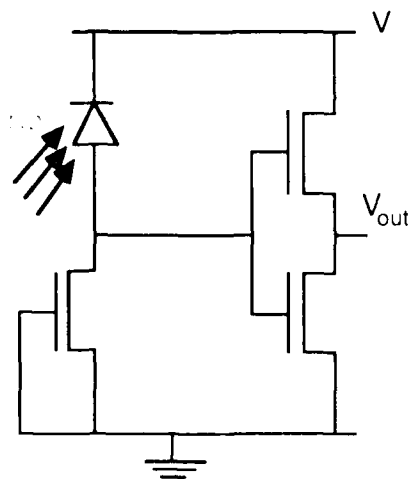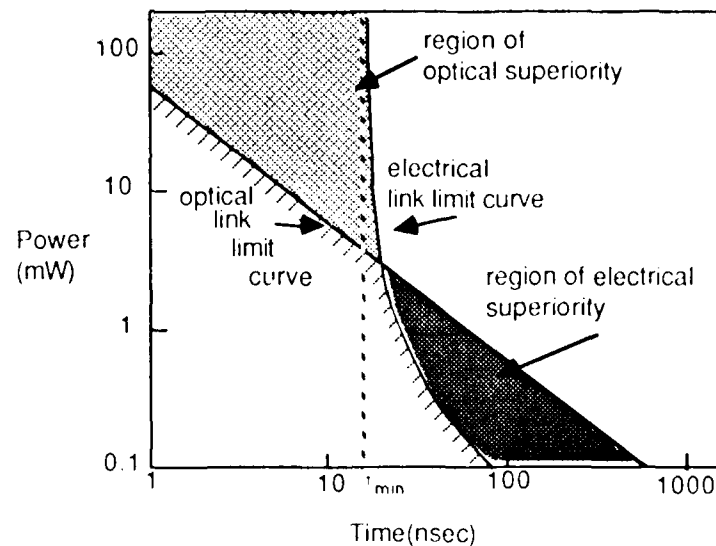

Fig. 3. Schematic diagram of photodetector circuit.

Fig. 4. Power versus delay time for optical interconnects and electrical polysilicon interconnects. L= 3 mm, F = 10 .

Optical Implementation of Min and Max Operation

Hedong Yang and Clark C. Guest

Department of Electrical Engineering and Computer Sciences
University of California, San Diego
La Jolla, CA 92093

I. Introduction

In fuzzy logic [1,2] and fuzzy cognitive maps (FCM) [3-5], the operations of maximum and minimum play roles that are parallel to those played by addition and multiplication in ordinary matrix algebra. For example, in a fuzzy associative memory (FAM) [6], the connection matrix is constructed by taking the outer product of the vector to be memorized with min operation substituting for product of elements, i.e, $t_{ij} = min(a_i, a_j)$.

In order to perform the optical implementation of fuzzy logic, FCM or FAM, it is necessary to implement the min and max operation optically. An approach for this task is proposed using coherent subtraction.

II. Approach

The approach described here is based on following identities:

$$max(a,b) = (a + b + |a - b|)/2$$
$$min(a,b) = (a + b - |a - b|)/2$$

The optical implementation is shown in Figure 1. Transparencies A and B represent two datapages. They are illuminated by coherent beams from a laser. In the left path, a halfwave plate is used to introduce a 180° phase delay between two data patterns such that the light amplitude incident on the LCLV [7] is proportional to the difference of the amplitudes of A and B. Obviously, on the reading side of the LCLV which works in its linear range, another beam is modulated and its intensity is proportional to |A-B|. In the right path, a beam is formed by adding A and B together. At beamsplitter 4, two beams, one proportional to |A-B| and the other to (A+B), are brought together. Thus the bitwise max operation on A and B is achieved. The min operation can be done by inserting a halfwave plate after the LCLV.

References

1. L. A. Zadeh, "Fuzzy Sets," Information and Control, Vol. 8, pp. 338-353, 1969.

2. L. A. Zadeh, "The Concept of a Linguistic Variable and its Application to Approximate Reasoning," Information Sciences, Vol. 8, pp. 199-249, 1975.

3. B. Kosko and J. S. Limm, "Vision as Causal Activation and Association," Proc. SPIE Intelligent Robots and Computer Vision, vol. 579, pp. 104-109, 1985.

4. B. Kosko, "Counting with Fuzzy Sets," IEEE Trans. Pattern Analysis and Machine Intelligence, Accepted for publication. 1986.

5. B. Kosko, "Adaptive Inference," IEEE Trans. on Systems, Man and Cybernetics. Accepted for publication. 1986.

6. B. Kosko, "Fuzzy Associative Memory," To be Published.

7. G. R. Knight, "Interface Devices and Memory Materials," Optical Information Processing, S. H. Lee, ed., pp. 114-119, Berlin: Springer-Verlag 1981.
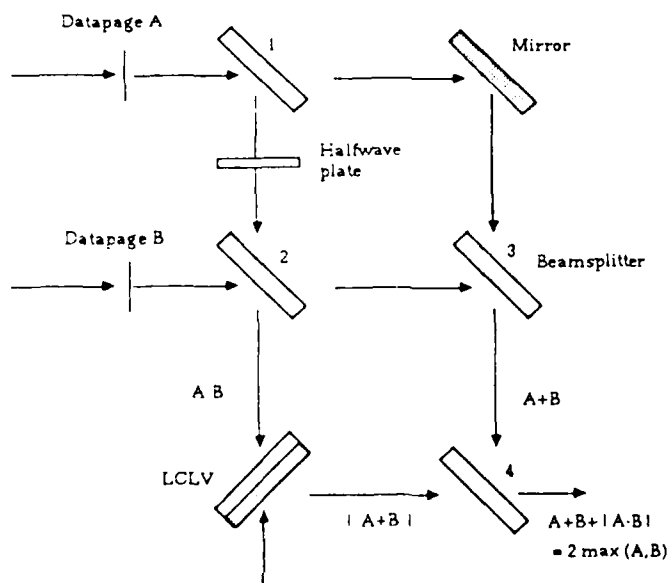
Figure 1

Optical MSD Adder Using Polarization Coded Symbolic Substitution
P. A. Ramamoorthy, S. Antony
Mail Location # 30, Department Of Electrical and Computer Engineering,
University of Cincinnati, Cincinnati, Ohio 45221.

1. Introduction

The optical adder using MSD combines the flexibility and accuracy of digital systems with the parallel information processing capability of optics. Hence the addition of two MSD numbers can be performed in three steps regardless of the number of digits in the MSD numbers. The MSD representation is a subset of the signed-digit representation where radix r equals two. Requirements of fully parallel addition and subtraction and of a unique representation for the zero value are fulfilled by signed-digit representations. A MSD number is represented by three digits $x_i (i = \bar{1}, 0, 1)$. For a precision of b bits, a given decimal number can be represented in MSD number system as follows[1,2].

$$X = \left[1, 0, 1\right].2^{b-1} + \cdots + \left[\bar{1}, 0, 1\right].2^1 + \left[\bar{1}, 0, 1\right].2^0$$

where one of the digit from the set $[\bar{1}, 0, 1]$ is selected for each term to give the appropriate representation. It has to be noted that any number can have more than one representation in MSD. The implementation of the MSD adder using symbolic substitution and light polarization for data coding is dicussed in this paper.

2. MSD Addition/Subtraction

The addition of two MSD numbers is performed in three successive stages by generating transfer and weight digits. The MSD adder for two 4-digit numbers is depicted in Fig. 1. As indicated here we use three functional blocks, block A, B and C for stages 1, 2 and 3 respectively. The input/output relationship for the blocks A,B and C for all input combinations is shown in Fig. 2. The transfer and weight digits generated by stage 1 are used to generate a second set of transfer and weight digits by stage 2. These are then summed in stage 3 to form the final output. Addition of two large numbers is carried out by simply adding the required number of identical functional blocks.

3. Polarization Coding and Symbolic Substitution Logic(SSL)

The optical implementation of the three functional blocks A, B and C is considered here. We need three different states to represent the three possible values of a MSD digit. This can be taken care of using polarization of light. Thus 1 can be represented by vertically polarized light (denoted by a vertical arrow), 0 by horizontally polarized light (denoted by a horizontal arrow) and $\bar{1}$ by light polarized at 45^0 (denoted by an arrow inclined at 45^0). Thus a MSD number will be represented by a spatial distribution of properly polarized light.

Symbolic substitution is a method that is highly suited for spatially distributed or two dimensional arrays. An architecture based on symbolic substitution works by recognizing one symbol and replacing it with another symbol. By a symbol we mean a collection of bits or digit patterns. The rules of substitution depend on the exact function, the SSL implements.

A limitation of symbolic substitution is that it can only be used in a spce-invariant interconnected architecture. That is, there has to be the same number of inputs to each gate or module, the same number of outputs and all the connections are exactly the same between the modules in a stage. Mathematically this implies that symbolic substitution method can be used to implement input/output relationships that are invariant in the spatial domain. For MSD addition using substitution logic, we can take care of the above constraint by modifying the MSD adder structure of Fig. 1 as shown in Fig. 3. In Fig. 3, we have inserted two zero valued inputs and added two more blocks in stages 2 and 3. This ensures that we perform identical operations within each operations. To each block of stage 1, we input two digits and obtain two outputs. The process continues in stage 2 and stage 3.

The input/output relationship for block A, B and C, when the inputs and outputs are polarization coded, is shown in Fig. 4 for two of the nine possible input combinations. The above relationship for the remaining input combinations can be derived in a similar way using Fig. 2. Thus the architecture based on SSL for MSD addition should be able to recognize the patterns shown in the LHS of Fig. 3 and substitute the recognized patterns by the patterns corresponding to the functional block or stage that is being implemented. The recognition of the patterns involves basically four steps as explained below. For a complete description and implementation of the SSL, the readers are referred to elsewhere[3].

Corresponding to each specified cell in the LHS pattern (the LHS pattern is always defined with respect to a reference cell) a copy of the input data plane is produced. The generated copy for a specified cell is then passed through a halfwave plate oriented at 45° for a zero, a quarterwave plate oriented at 45° for a $\bar{1}$ and no plates for a 1 in the cell considered. The resulting copy is then shfted in suc! a direction that the cell associated with the copy is moved to the position of the reference cell. Finally the shifted copies are superimposed. The cells in the superimposed plane containing two vertically polarized beams uniqely represent the presence of the search patterns. Cells containing any combinations other than two vertically polarized beams are considered as unwanted. These are removed from the recognition output plane by intensity thresholding so that only those cells containing the search pattern are bright and the light thus generated is polarized vertically.

Substitution of the recognized patterns by the output patterns corresponding to the functional block that is being implemented also involves the same four steps. It should be noted that for each stage of the MSD adder, we have nine possible input search patterns and hence nine possible substitution rules. Also each pattern has two distinct cells that have to be recognized. Thus the implementation involves nine pattern transformations taking place in parallel.

4. Conclusion

The MSD representation eliminates carry propagation chains in addition and subtraction and provides an arithmetic that is fully parallel. The MSD adder takes full advantage offered by the MSD number representation along with the massive parallelism of optics providing the result in a fixed time that is independent of the number of digits involved.

5. References

1. A. Avizienis, "Signed-digit number representation for fast parallel arithmetic," IRE Trans. Electronic Computers EC-10, 389-400 (1961).

2. B. L. Drake, R. P. Bocker, M. E. Lasher, R. H. Patterson and W. J. Miceli, "Photonic computing using the modified signed-digit number representation," Opt. Eng. 25 (1), 038-043 (1986).

3. K. H. Brenner, "New implementation of symbolic substitution logic," Appl. Opt. 25 (18), 3061-3064 (1986).
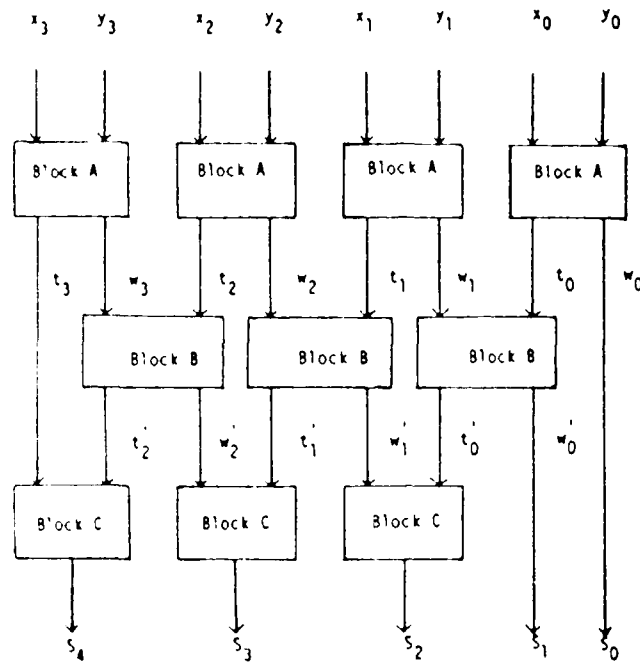
Fig. 1 MSD adder for two 4-digit numbers.

Fig. 2 Outputs of the functional blocks for two given digits

(a) Outputs t_i and w_i for Block A for inputs x_i and y_i

(b) Outputs t_i' and w_i' for Block B for inputs t_i and w_i

(c) Outputs S_i for Block C for inputs t_i' and w_i'
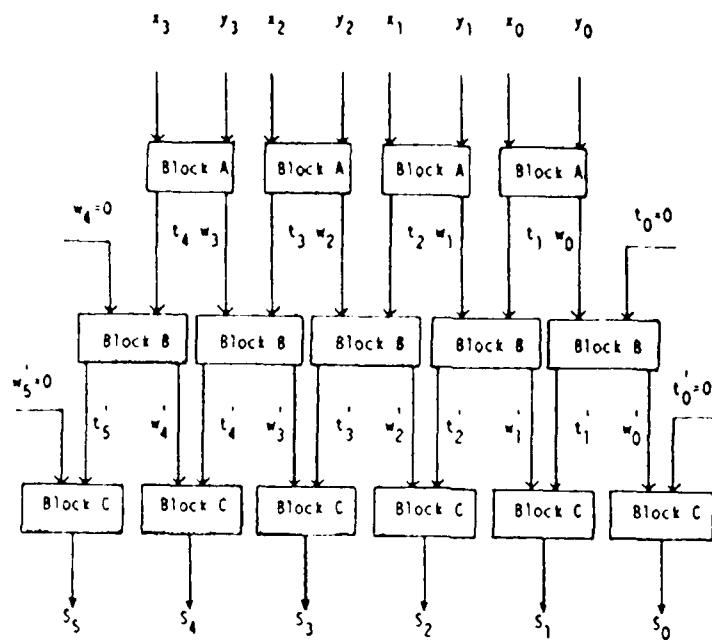
Fig. 3 Modified MSD adder for two 4-digit numbers.

(a)   (b)   (c)   (d)
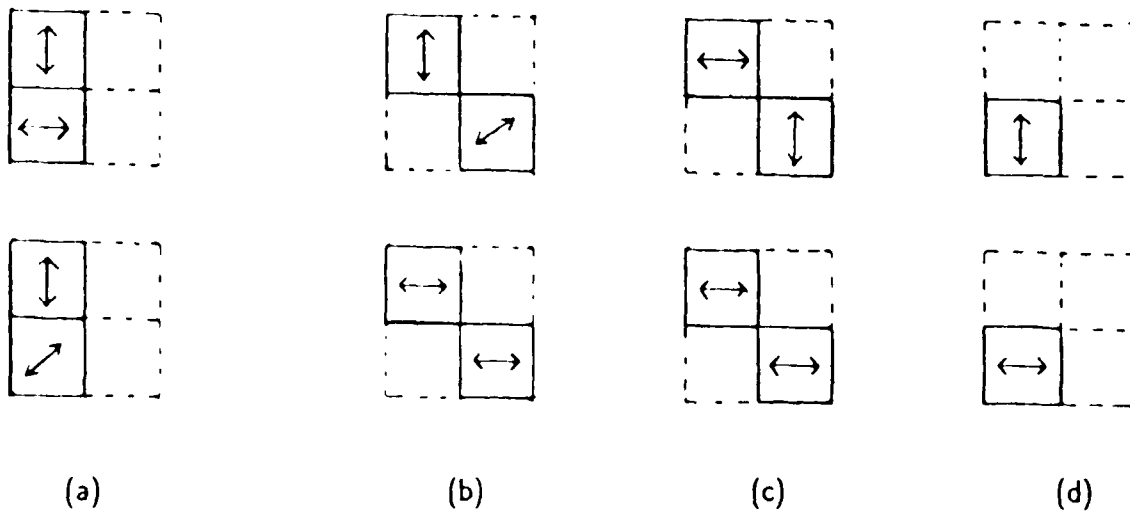
Fig. 4   Substitution rules for the functional blocks for two of the nine possible input combinations
(a)   LHS patterns representing the possible combinations of the input digits
(b)   Substitution rule for Block A
(c)   Substitution rule for Block B
(d)   Substitution rule for Block C

Digital Optical Processor Based on Symbolic Substitution Using Matched Filtering

Ho-In Jeon

Department of Electrical Engineering-Systems
University of Southern California
2801 S. Orchard Ave. #2
Los Angeles, California 90007

Abstract

In this paper, we propose a digital optical processor design based on symbolic substitution using holographic matched filtering and space-invariant interconnections. The proposed system performs binary addition in a highly parallel manner, i.e., the processing time depends on the word size but not on the array size. Crosstalk in symbolic substitution is described and new symbols which can prevent crosstalk in binary addition are introduced.

1. Introduction

The symbolic substitution technique for digital optical computing is a rule by which the symbols associated with two operands are replaced by new symbols associated with the results of the operation. The new symbols are fed back into the input of the system and this process continues until the desired output is obtained. A technique based on shifting and overlapping copies of the input is presented in Reference [1]. Recently, other techniques for implementing digital optical processors based on symbolic substitution have been described in References [2, 3, 4].

In this paper, a digital optical processor design based on symbolic substitution using holographic matched filtering and space-invariant interconnections is proposed. Crosstalk due to neighboring symbols, which can occur in the pattern recognition process, is described and new symbols which can prevent the crosstalk in binary addition are introduced.

2. Crosstalk and Data Representation for Binary Addition [5]

In symbolic substitution, symbols are used to represent input data and patterns are formed based on these symbols. Each pattern is first recognized and then replaced by new patterns according to given substitution rules. However, pixels bordering a given pattern can in some cases cause additional patterns to be recognized in unintended locations. This is what we call crosstalk; it can be prevented by suitable choice of the symbols used to encode the data or by incorporating masks in the optical system. Here we will consider the former approach and will introduce symbols that prevent crosstalk in binary addition. The symbols that represent logical zero and one must have equal amounts of optical power for correlation to provide correct recognition. Furthermore,

the spatial arrangement of pixels must be the same for a logical one as for a logical zero. Given this, the use of two (dark and bright) pixels to represent each logical value will necessarily cause crosstalk to occur. By extending the dimension of the signal space, new symbols can be defined as shown in Fig. 1 which eliminate crosstalk during the addition operation. Also, whether these new symbols or masks are used, isolation pixels must be inserted between adjacent patterns. Fig. 2 shows the relevant substitution rules for binary addition (from the symbols of Fig. 1 and the rules of Reference [2]).

In symbolic substitution for binary addition, the two operands are placed in two rows. Each pair of bits in the operands is replaced by a sum and carry bit, with the upper of two rows representing carries and the lower row representing sum bits. Since at each successive iteration each carry bit must be added to next most significant bit, a skew is introduced as shown in Fig. 2. However, this causes another problem: when the first iteration is performed, the rightmost symbol of the upper row and the leftmost symbol of the lower row should be the symbol which represents a logical zero. Since the replacement step is space-invariant and fixed, and since the positions of the zeros to be inserted vary from one iteration to the next, it is impossible to insert zeros only into those positions. Failure to insert zeros will cause the rightmost symbol of the lower row and the leftmost symbol of the upper row to die out in the recognition process of next iteration, and, consequently, we cannot obtain a desired output in the final iteration. This problem can be solved by padding N zeros on both the right and left hand sides of the two operands, where N is the word size.

3. Operating Principles

Symbolic substitution can be implemented with pattern recognition, replacement, and feedback. The pattern recognition can be performed by using holographic matched filtering and thresholding, and the replacement can be done by utilizing space-invariant interconnections. The schematic diagram of the proposed system is shown in Fig. 3.

The encoded binary input pattern formed by using the symbols shown in Fig. 1 is fed into shutter S1 and split into four identical portions using a "binary tree structure." Since we have four possible combinations of symbols for binary addition ($0 + 0$, $0 + 1$, $1 + 0$, and $1 + 1$), we need four different holographic matched filters, one to recognize each of them. Each of these filters has a transfer function which is the complex conjugate of the Fourier transform of one of the four different symbols shown in the left hand sides of Fig. 2. Through these filters and Fourier transform lenses, autocorrelations are produced in the output planes of the matched filters. With threshold elements placed in these planes, the autocorrelation peaks occur at all positions where the four different patterns are matched. The recognized patterns are used to generate new patterns based on the substitution rules shown in Fig. 2. Since a hologram can be used as a beam-steering element, any new pattern can be generated using a computer generated or optically recorded hologram placed between Fourier transform lenses. The replaced patterns are combined through the beam splitters and mirrors. The combined output pattern is again split into two parts and stored in the optical memory M1, where S2 is opened and S4 is closed. After the processed output is stored, S1 and S2 are closed and S3 is opened. We will then get an intermediate result of the first iteration, and it is fed back into the input through the beam splitter for the second iteration. In the second iteration, S2 is still closed and S4 is opened to store the result of the second iteration in the optical

memory M2. After the result of the second iteration is stored, we close S3 and S4, erase M1, and open S5 to feed the output back into the input. The final result for addition of N bit numbers is obtained after $(N+1)$ iterations.

The optical processor proposed in this paper requires a non-inverting threshold instead of a NOR gate array because correlative pattern recognition is used. Optical bistable devices could be used for the threshold elements and optical memories.

4. Conclusion

The symbolic substitution technique is quite different but general and powerful compared to current approaches towards computation in the sense that symbols are used and it operates in a highly parallel manner. This property provides some useful applications to the implementation of array processors, switching networks, and simulation of physical processes [1]. It also provides for direct implementation of parallel searching.

In this paper, a system design of a digital optical processor based on symbolic substitution using holographic matched filtering and space-invariant interconnections was proposed. The proposed system performs binary addition in a highly parallel manner, i.e., the processing time depends on the word size but not on the array size. Crosstalk in symbolic substitution was described and new symbols which can prevent it were introduced for the case of binary addition.

Acknowledgments

References

[1] Alan Huang, "Parallel Algorithms For Optical Digital Computers," *Proc. Tenth International Optical Computing Conference, IEEE Catalog Number 83CH1880-4*, pp. 13-17, 1983.

[2] K.H. Brenner, and Alan Huang, "An Optical Processor Based on Symbolic Substitution," *OSA Topical Meeting on Optical Computing*, Incline Village, Nevada, pp. WA4-1 - WA4-3, March 1985.

[3] K.H. Brenner, Alan Huang, and Norbert Streibl, "Digital Optical Computing with Symbolic Substitution," *Applied Optics*, Vol. 25, No. 18, pp. 3054-3060, September 1986.

[4] K.H. Brenner, "New Implementation of Symbolic Substitution Logic," *Applied Optics*, Vol. 25, No. 18, pp. 3061-3064, September 1986.

[5] Ho-In Jeon, "Optical Processor Design Using Symbolic Substitution Architecture Based on Matched Filtering," *Final Report, EE689 Term Project*, University of Southern California, May 1986.

Fig. 1 Logical values zero and one represented by symbols
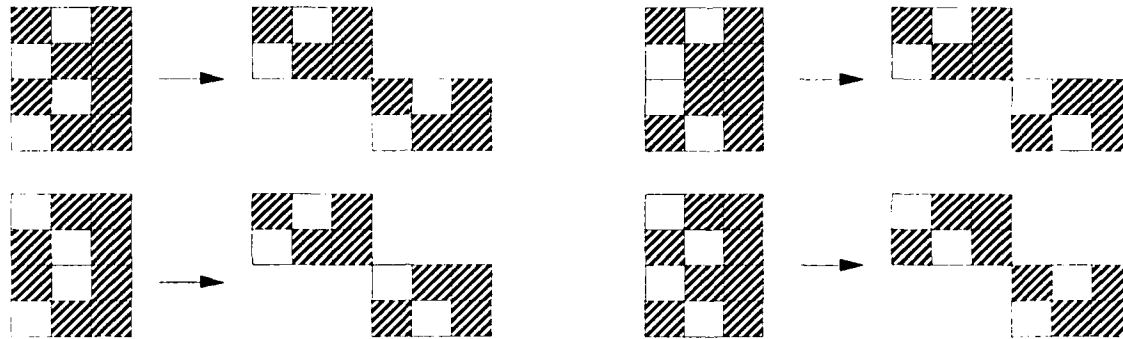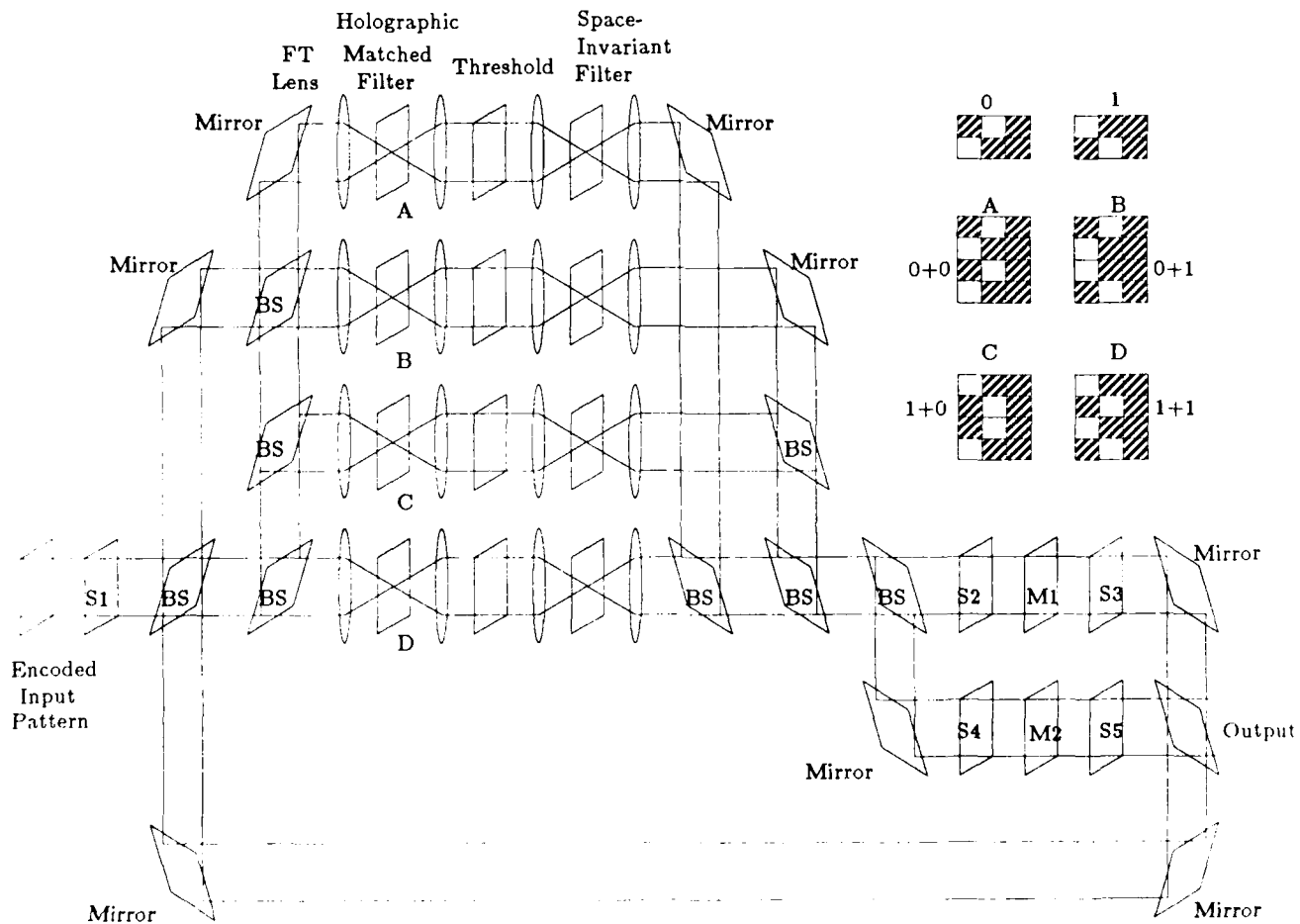

Fig. 2 Substitution rules for binary addition

Fig. 3 Schematic Diagram of Digital Optical Processor Based on Symbolic Substitution Using Matched Filtering

Optical Parallel Image Processing Using CCD Image Sensor

J.Tokumitsu,H.Matsuoka and K.Iijima

Canon Research Center
5-1,Morinosato-Wakamiya,Atsugi 243-01,Japan

1.Introduction

Image processing is one of the most promising application fields of optical computing in which parallelism inherent in optics well matches two dimensional nature of images. Convolution is a basic operation in preprocessing images, and various kinds of methods for optically implementing it have been proposed[1-3].

In this paper we present a simple optical system which can perform convolution operation at video rate by taking advantage of parallelism in optics. Modulation of an image during image sensing is also employed as in the literature[3], but our system is capable of real time operation.

2.Principles of Operation

Convolution can be expressed in a discrete form as

$$g(x,y)=\sum_{i}\sum_{j} a_{ij}f(x-x_i,y-y_j) \quad , \quad (1)$$

where f, g and a are an original image, a processed image and a kernel function, respectively. Each pixel is sequentially processed according to Eq.(1) in most electronic image processing systems.

Equation (1), however, suggests that convolution can be performed to all the pixels in parallel, if we can shift the original image horizontally and vertically and have a weighted sum of those shifted images.

In our system the shift of the image is achieved by horizontally oscillating the film used as an input image by means of a bimorph and by vertically transferring charges on CCD image sensors. Multiplication of f by a is done by illuminating the film with intensity-modulated light. The summation of Eq.(1) is done by accumlating charges on CCD image sensors at every step of the shift. The final processed image is displayed on TV monitor in real time.

The problem that the kernel function a often has a negative value has been overcome as follows: Two independent pairs of CCD image sensor and light source are used to represent positive and negative components of a, respectively. Then output electronic currents from two CCDs corresponding to two separate images are fed into an electronic differential amplifier in order to have the difference between two images.

3.Experimental System

The system configuration is schematically shown in Fig.1. Figure 2 is a photograph of the optical system together with a sensor to monitor the displacement of the bimorphs.

LED 1 emits light with the intensity determined by positive components of the kernel function. The linearly polarized light through PBS(Polarization Beam Splitter) 1 illuminates a film and the film is imaged onto CCD 1 through PBS 2. Likewise, the film illuminated by the orthogonally polarized light from LED 2 is imaged onto CCD 2.

The film was attatched to the bimorphs oscillating at the frequency of 180 Hz. This frequency was so chosen that it realizes the kernel function of moderate size of 5x5 pixels, i.e. LEDs are pulsed 25 times during two and a half periods of oscillation with charges on CCD being transferred 4 times by one line (Fig.3) and video signal is read out in the subsequent half period, which enables one TV field to be completed in 1/60 second.

We used CCD image sensors operating at frame-transfer mode[4] and drivers of Canon Ci-10 Compact Color Video Camera Module, without color filters on the CCD chips.

The whole system is controled by a personal computer and the additional electronics which supply timing signals and the intensity singals of the LEDs.

4.Experimental Results

Figures 4 (a)-(c) are photographs of the original and the processed images displayed on the TV screen. A bar chart was used as an original image(Fig.4(a)). Edge extraction (Fig.4(b)) and smoothing(Fig.4(c)) were performed by setting values to the kernel function according to the operations. Since these processing are executed at video rate, we can obtain the processed image in real time when the kernel functions are changed through the computer terminal.

5.Conclusions

We have demonstrated the system which performs convolution on an image at video rate. This is a practical hybrid system, where optics and electronics work complementarily providing each inherent advantage. Optics offers paralellism, while electronics provides image sensing means and flexibility of selecting the kernel function.

Applications to preprocessing for pattern recognition and processing of moving images are envisioned. This system will find more applications using other type of operations which may be possible by making use of recent progress in optical array logic and cellular logic[5,6].

References

1. T. Honda, et al., Opt. Acta, 21, 653(1974).
2. A.Vander Lugt, Proc. IEEE, 62, 1300(1974).
3. T. Ikuta, Appl. Opt., 24, 2907(1985).
4. T. Sasaki, First ICO Topical Meeting - Image Science '85, 46(1985).
5. J. Tanida and Y. Ichioka, Appl. Opt., 25, 1565(1986).
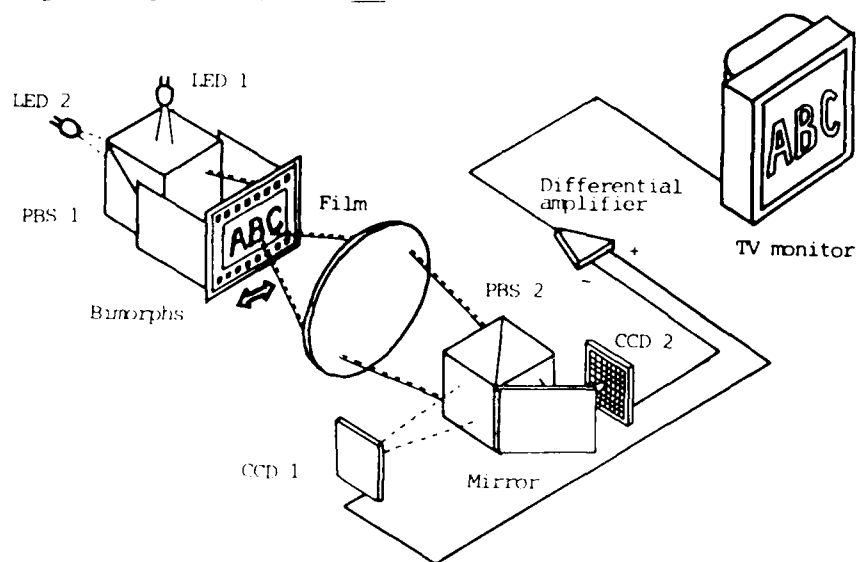6. T. Yatagai, Appl. Opt., 25, 1571(1986).

Fig. 1. Configuration of the experimental system

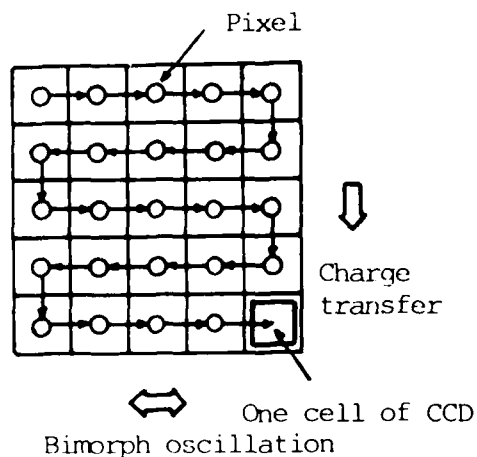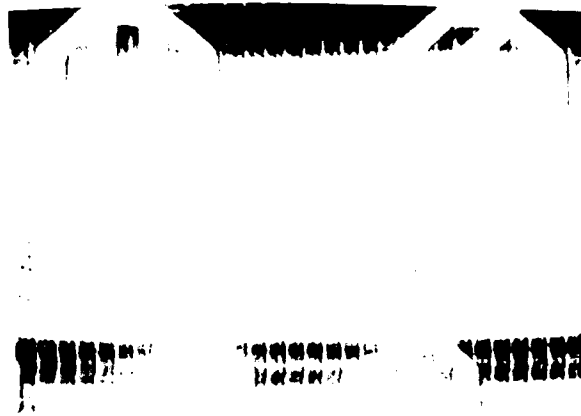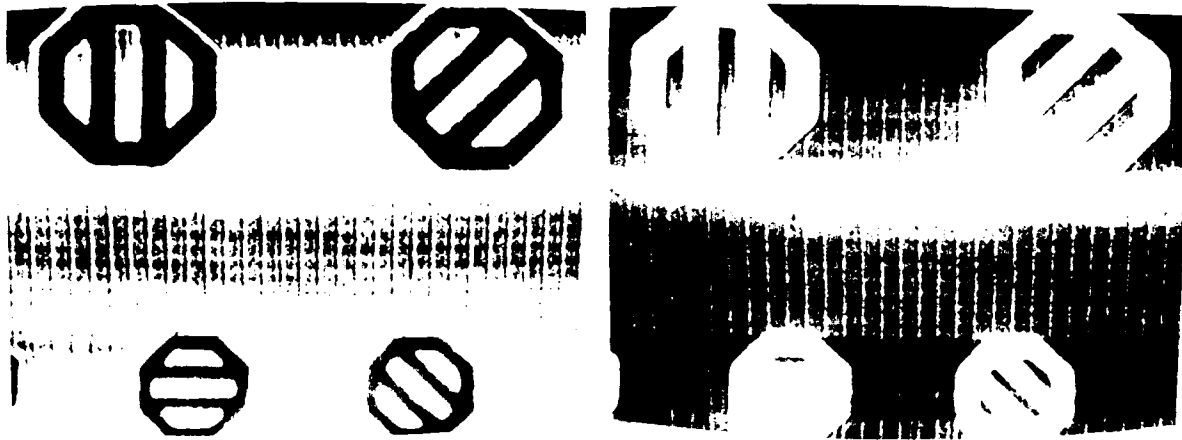Fig. 2. Photograph of the optical system

Fig. 3. Illustration of processing principle

Fig. 4. Experimental results
   (a) original image
   (b)edge extraction
   (c) smoothing

TUESDAY, MARCH 17, 1987

PROSPECTOR/RUBICON ROOM

8:00 AM–10:00 AM

TuA1–7

SESSION 6

H. John Caulfield, University of Alabama in Huntsville, *Presider*

Advances in Brain-Style Computation
David E. Rumelhart
University of California, San Diego

A sketch of current work on brain-style computation is provided.
Emphasis is on applications for building content-addressable
memories and learning machines.

ARCHITECTURES FOR OPTO-ELECTRONIC ANALOGS OF SELF-ORGANIZING NEURAL NETWORKS

Nabil H. Farhat
University of Pennsylvania
Electrical Engineering Department
Electro-Optics and Microwave-Optics Laboratory
Philadelphia, PA 19104-6390

Abstract

Architectures for partitioning opto-electronic analogs of neural nets into input/output and internal units to enable self-organization and learning where a net can form its own internal representations of the "environment" are described.

1. INTRODUCTION: In our preceeding work on optical analogs of neural nets, [1],[2], the nets described were programmed to do a specific computational task, namely a nearest neighbor search by finding the stored entity that is closest to the address in the Hamming sense. As such the net acted as a content addressable associative memory. The programming was done by computing first the interconnectivity matrix using an outer-product recipe given the entities we wished the net to store and become familiar with followed by setting the weights of synaptic interconnections or links between neurons accordingly.

In this paper we are concerned with architectures for opto-electronic implementation of neural nets that are able to program or organize themselves under supervised conditions, i.e., of nets that are capable of (a) computing the interconnectivity matrix for the associations they are to learn, and (b) of changing the weights of the links between their neurons accordingly. Such self-organizing networks have therefore the ability to form and store their own internal representations of the entities or associations they are presented with.

Multi-layered self-programming nets have been described recently [3]-[5] where the net is partitioned into three groups. Two are groups of visible or external input/output units or neurons that interface with the outside world i.e., with the net environment. The third is a group of hidden or internal units that separates the input and output units and participates in the process of forming internal representations of the associations the net is presented with, as for example by "clamping" or fixing the states of the input and output neurons to the desired associations and letting the net run through its learning algorithm to arrive ultimately at a specific set of synaptic weights or links between the neurons that capture the underlying structure of all the associations presented to the net. The hidden units or neurons prevent the input and output units from communicating with each other directly. In other words no neuron or unit in the input group is linked directly to a neuron in the output group and vice-versa. Any such communication must be carried out via the hidden units. Neurons within the input group can communicate with each other and with hidden units and the same is true for neurons in the output group. Neurons in the hidden group can not communicate with each other. They can only communicate with neurons in the input and output groups as stated earlier.

Two adaptive learning procedures in such partitioned nets have attracted considerable attention. One is stochastic involving a **simulated annealing** process [6],[7] and the other is deterministic involving an **error back-propogation** process [8]. There is general agreement however; that because of their iterative nature, serial digital computation of the links with these algorithms is very time consuming. A faster means for carrying out the required computations is needed. Never-the-less the work mentioned represents a milestone in that it opens the way for powerful collective computations in multilayered neural nets and in that it dispels earlier reservations [8] about the capabilities of early models of neural nets such as the Perceptron [9] when the partitioning concept is introduced. What is most significant and noteworthy, in our opinion, is the ability to now define buffered input and output groups with unequal number of neurons in a net which was not possible with earlier nets where all neurons participate in defining the initial (input) and final (output) states of the net.

2. ANALOG IMPLEMENTATIONS: Optics and opto-electronic architectures and techniques can play an important role in the study and implementation of self-programming networks and in speeding-up the execution of learning algorithms. We have done some exploratory work in this regard to see how the neurons in an opto-electronic analog of a neural net can be partitioned into groups with specific interconnection patterns. Here, for example, a method for partitioning an opto-electronic analog of a neural net into input, output, and internal units with the selective communication pattern described earlier to enable, stochastic learning, i.e., carrying out a simulated annealing learning algorithm in the context of a **Boltzmann machine** formalism is described. (see Fig. 1(a)). The arrangement shown in Fig. 1(a) derives from the neural network analogs we described earlier [2]. The
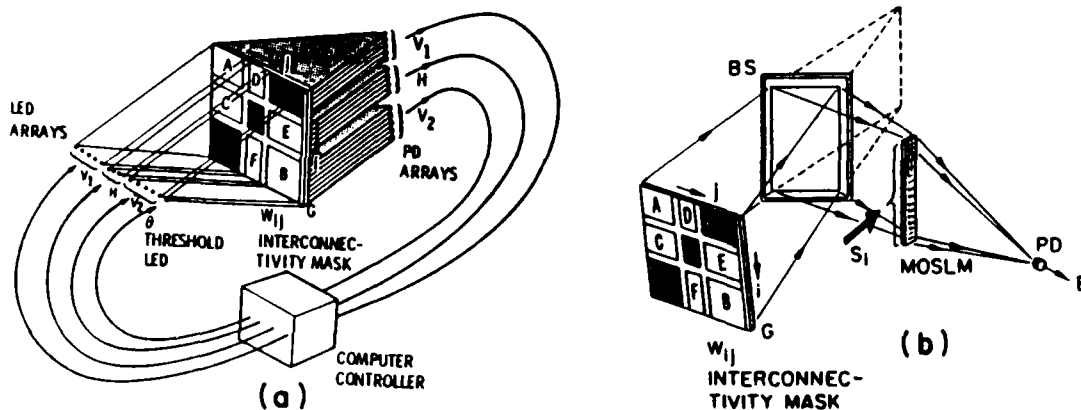
Fig. 1. Partitioning concept (a) and method for rapid determination of the net's energy E.

network, consisting of say N neurons, is partitioned into three groups. Two groups, V_1 and V_2, represent visible or exterior units that can be used as input and output units respectively. The third group H are hidden or internal units. The partition is such that $N_1 + N_2 + N_3 = N$ where subscripts 1,2,3 on N refer to the number of neurons in the V_1, V_2 and H groups respectively. The interconnectivity matrix, designated here as W_{ij}, is partitioned into nine submatrices, A,B,C,D,E, and F plus three zero matrices shown as blackened or opaque regions of the W_{ij} mask. The LED array

represents the state of the neurons, assumed to be unipolar binary LED on = neuron firing, LED off = neuron not-firing. The W_{ij} mask represents the strengths of interconnection between neurons in a manner similar to earlier arrangements [2]. Light from the LEDs is smeared vertically over the W_{ij} mask with the aid of an anamorphic lens system (not shown in Fig. 1(a)) and light emerging from rows of the mask is focused with the aid of another anamorphic lens system (also not shown) onto elements of the photodetector (PD) array. Also we assume the same scheme utilized in [2] for realizing bipolar values of W_{ij} in incoherent light is adopted here, namely by separating each row of the W_{ij} mask into two subrows and assigning positive values W_{ij}^{+} to one subrow and negative values W_{ij}^{-} to the other, then focusing light emerging from the two subrows separately onto pairs of adjacent photosite connected in opposition in the V_1, V_2 and H segment of the photodetector array. Submatrix A with $N_1 \times N_1$ elements, provides the interconnection weights of units or neurons within group V_1. Submatrix B with $N_2 \times N_2$ elements, provides the interconnection weights of units within V_2. Submatrices C (of $N_1 \times N_3$ elements) and D (of $N_3 \times N_1$ elements) provide the interconnection weights between units of V_1 and H and submatrices E (of $N_2 \times N_3$ elements) and F (of $N_3 \times N_2$) provide the interconnection weights of units of V_2 and H. Units in V_1 and V_2 can not communicate with each other directly because locations of their interconnectivity weights in the W_{ij} matrix or mask are blocked out (blackened lower left and top right portion of W_{ij}). Similarly units within H do not communicate with each other because locations of their interconnectivity weights in the W_{ij} mask are also blocked out (center blackened square of W_{ij}). The LED element θ is always on to provide a fixed or adaptive threshold level to all other units by contributing to the light focused onto only negative photosites of the photodetector (PD) arrays.

By using a computer controlled nonvolatile spatial light modulator to implement the W_{ij} mask in Fig. 1(a) and including a computer/controller as shown the scheme can be made self-programming with ability to modify the weights of synaptic links between its neurons to form internal representations of the associations or patterns presented to it. This is done by fixing or clamping the states of the V_1 (input) and V_2 (output) groups to each of the associations we want the net to learn and by repeated application of the simulated annealing procedure with Boltzmann, or other, stochastic state update rule and collection of statistics on the states of the neurons at the end of each run when the net reaches thermodynamic equilibrium.

For each clamping of the V_1 and V_2 units to one of the associations, annealing is applied, starting from an arbitrary W_{ij}, with switching states of units in H until thermodynamic equilibrium is reached. The state vector of the entire net, which represents a state of global energy minimum, is then stored by the computer. This procedure is repeated for each

association several times recording the final state vectors every time. The probabilities P_{ij} of finding the i-th and j-th neurons in the same state are then obtained. Next with the output units V_2 unclamped to let them free run like the H units the above procedure is repeated for the same number of annealings as before and the probabilities P'_{ij} are obtained. The weights W_{ij} are then incremented by $\Delta W_{ij} = \eta(P_{ij}-P'_{ij})$ where η is a constant that controls the speed and efficacy of learning. Starting from the new W_{ij} the above procedure is repeated until a steady W_{ij} is reached at which time the learning procedure is complete. Learning by simulated annealing requires calculating the energy E of the net [3],[5]. A simplified version of a rapid scheme for obtaining E opto-electronically is shown in Fig. 1(b). A slight variation of this scheme that can deal with the bipolar nature of W_{ij} would actually be utilized. This is not detailed here because of space limitation.

3. REMARKS: The partitioning architecture described is extendable to multilayered nets of more than three layers and to 2-D arrangement of neurons. Learning algorithms in such layered nets lead to multivalued W_{ij}. Therefore high-speed computer controlled SLMs with graded pixel response are called for. Methods of reducing the dynamic range of W_{ij} or for allowing the use of W_{ij} with ternary weights are however under study to enable the use of commercially available nonvolatile SLM devices that are mostly binary e.g., Litton's MOSLM.

4. ACKNOWLEDGEMENT: The work reported was supported by grants from DARPA/NRL, The Army Research Office, and the University of Pennsylvania Laboratory for Research on the Structure of Matter.

REFERENCES:

1. Psaltis, D. and N. Farhat, Digest of the 13-th Congress of Intern. Commiss. on Optics, ICO-13, Saporro, Japan, 1984. Also, Opt. Lett., Vol. 10, pp. 98-100, 1985.
2. Farhat, N.H., et. al., App. Optics, Vol. 24, pp. 1469-1475, 1985.
3. G.F. Hinton, et. al., Carnegie-Mellon University Report No. CMU-CS-84-119, May (1984).
4. D.E. Rumelhart, G.E. Hinton and R.J. Williams, Institute for Cognitive Science Report 8506, Univ. of California, San Diego, Sept. 1985.
5. T.J. Sejnowski and C.R. Rosenberg, Johns Hopkins University, Electrical Engineering and Computer Science Technical Report No. JHU/EECS-96/01 (1986).
6. N. Metropolis, et. al., J. Chem. Phys., Vol. 21, 6, pp. 1087-1092, June, 1953.
7. S. Kirkpatric, et. al., Science, Vol. 220, pp. 671-680, May 1983.
8. M.L. Minsky and S. Papert, Perceptrons, MIT Press, Cambridge, MA, (1969).
9. F. Rosenblatt, Principles of Neuro-Dynamics: Perceptions and the Theory of Brain Mechanisms, Spartan Books, Wash. D.C., 1962.

Optical Neural Nets Implemented with Volume Holograms

Demetri Psaltis, Jeffrey Yu, Xiang Guang Gu, and Hyuk Lee*

California Institute of Technology
Department of Electrical Engineering
Pasadena, California, 91125

An optical implementation of a neural net computer consists of two basic components: neurons and connections. The neurons are simple nonlinear processing elements (e.g. thresholding units) that accept inputs from other neurons and produce a single output that is broadcast to many other neurons. Typically we think of each neuron being connected to thousands others. Hence the number of connections in a network is much larger than the number of neurons. This simple fact is the principal motivation for considering optical implementations of neural nets [1]. The basic approach we have adopted is shown in Fig. 1. The neurons are arranged in a planar configuration and interconnected with optical elements (holograms or masks). Several emerging optical technologies can be considered for the implementation of the two basic components. We have come to the conclusion that the most promising technology for the implementation of neurons is optoelectronics: a two dimensional array of LEDs, a detector adjacent to each LED and a saturating amplifier connecting the two [2,3]. We are considering two possiblities for performing the connections: optical memory disks and volume holograms [2]. In this paper we examine the advantages of using volume holograms as opposed to planar media for storing the interconnect pattern, we present methods for achieving different types of arbitary global interconnections, and we present experimental results using a photorefractive crystal (LiNbO$_3$) to implement modifiable synapses.

The motivation for using volume holograms comes from their ability to store information in three dimensions. The potential for a dramatic increase in storage density that results was recognized early on by Van Heerden [4] and more recently volume holographic memories have been developed by Gaylord and co-workers [5]. In the optical implementation of a neural net (Fig. 1) the use of volume holograms provides a much better match between the area of the neural planes and the transverse area that is required by the device that performs the interconnections. Let A_1 and A_2 be the areas of the input and output processing planes. Then the maximum number of neurons that can be packed at the input and output planes is A_1/λ^2 and A_2/λ^2 respectively, where λ is the wavelength. The total number of arbitary connections that need to be specified is therefore $A_1 A_2/\lambda^4$. If the interconnections are specified by a planar optical transparency (e.g. a hologram, a memory disk) with area A_H, then the number of degrees of freedom on the hologram is A_H/λ^2. This leads to the following relationship:

$$A_H \cdot \frac{A_1 A_2}{\lambda^2} \tag{1}$$

The above tells us that the area of the hologram needs to be much larger than the areas of the neural planes. For instance A_H needs to be at least as large as A_1 multiplied by the maximum number of output points (A_2/λ^2) in order to have arbitary interconnect capability. The use of volume holograms allows us to have a better match between the transverse dimensions of the neural and interconnections devices by making effective utilization of the third dimension. The maximum number of connections that can be specified if a volume hologram is used in Fig. 1 is upper bounded by the degrees of freedom V_H/λ^3, available

in the volume of the crystal:

$$V_H > \frac{A_1 A_2}{\lambda} \tag{2}$$

where V_H is the volume of the hologram. The above relationship can be interpreted in a number of different ways. For instance suppose that $V_H = A_t \times L$ where A_t and L are the transverse area and the thickness of the volume hologram respectively. Let us also assume that $A_t = A_1$. Then Eq. (2) tells us that $A_2 < \lambda L$, which is typically much smaller than A_1. If we attempt to perform this same interconnection with a planar hologram its transverse area would have to satisfy $A_H > A_1(L/\lambda)$. L/λ is the number distinct samples in the longitudinal dimension of the volume hologram and A_1 is its transverse area in this example. Thus A_H must be at least as large as the total area of all the equivalent planes that are "stacked" in the three dimensions of the crystal.

The above discussion is based on bounds. We still need to find specific interconnection schemes that will yield the increase in connecting capability that volume holograms can in principle provide. Perhaps the simplest demonstration of the increase in storage capacity that results from using volume holograms is the familiar VanderLugt correlator. The VanderLugt system with a planar hologram can be thought of as an arbitary interconnect between the N^2 pixels of the input and a single point at the output. When a volume hologram is used in the VanderLugt system [6], multiple holograms are recorded by interfering in a photorefractive crystal the transformed field of a set of images with a plane wave reference each incident at a different angle. When an input is presented to the system, each of the stored filters is correlated with the input and the outputs are produced in parallel spatially separated. Because of Bragg discrimination, crosstalk between the various correlations is suppressed. Thus we can in this case arbitarily connect the N^2 points at the input with a 1-D array of L points at the output. Typically $L \approx N$, hence the volume VanderLugt system can be thought of as an arbitrary $N^2 \mapsto N$ connection network. The use of the volume hologram in this case has increased the number of connections by a factor of N.

A counterpart to the $N^2 \mapsto N$ interconnect scheme is the performance of an $N \mapsto N^2$ interconnect. The same volume hologram that is used to perform a specified $N^2 \mapsto N$ connectivity pattern can be now used to do the corresponding $N \mapsto N^2$ mapping simply by reversing the direction in which light propagates.

The $N^2 \mapsto N$ and $N \mapsto N^2$ mappings are both useful for the implementation of neural net models. The $N \mapsto N^2$ is useful for instance in the early stages of a multilayered network where it is desiarable to increase the dimensionality of the data (introduce a large number of "hidden units") in order to make more classification assignments of the input patterns. Similarly the fan-in mapping $N^2 \mapsto N$ is useful for reducing the dimensionality of the feature space once classification has been made possible. However there is also need for dimensionality preserving mappings that connect a roughly equal numbers of input and output neurons. For instance this type of mapping is required for implementing a Hopfield style network. Let us assume that the volume hologram is a cube, each of its faces having an area A_t. Let us assume as an example that $A = A_t = A_1 = A_2$. Then from Eq. (2) we conclude that $1 > (\sqrt{A}/\lambda)$ which implies that we can only connect one input spot to one spot at the output with a λ^3 crystal. In order to do more meaningful mappings we need to have $A_1, A_2 < A_t$ or equivalently have the neurons in the input and output planes arranged sparsely. Let N be the maximum number of samples that can be stored in one dimension either at the neural planes or the volume hologram. Then the maximum number of connections that the volume hologram can specify is N^3 whereas the total number of connections that can be possibly made between the input and output planes is N^4. Therefore we are essentially one spatial dimension short if we attempt to fully and

arbitarily interconnect two 2-D surfaces with a 3-D hologram. The maximal dimensionality preserving mapping we can do with a volume hologram having $V_H \lambda^3 \quad N^3$ degrees of freedom is $N^{3/2} \cdot N^{3/2}$.

There are several ways to do a $N^{3/2} \cdot N^{3/2}$ mapping. One possibility is illustrated in Fig.2. This example is for the case $N = 16$. This means that the number of resolvable spots in the input and also the output planes is $N^2 \quad 256$. From our discussion above, we know we can only use $N^{3/2} \quad 64$ of these at either plane. How do we decide which of the $N^{3/2}$ out of N^2 points we are allowed to use? The answer is that the spots must be arranged such that no two input-output pairs are connected with an identical grating in the volume hologram.This is necessary for the implementation of arbitary interconnections. To satisfy this requirement we must break all symmetries and this leads to systematic ways for selecting the $N^{3/2}$ points. The arrangement in Fig. 2 is an example of this. The top diagram in Fig. 2 corresponds to the arrangement of neurons at the input plane and the bottom part of the figure is the output. Each plane is partitioned in $\sqrt{N} \cdot \sqrt{N}$ blocks each containing $\sqrt{N}$ points. In the input plane all blocks have 4 points along the diagonal. At the output planes the location of the points is permuted within each block. We have shown that this arrangement allows us to perform an arbitary $N^{3/2} \rightarrow N^{3/2}$ mapping.

The above method has been experimentally verified. The experimental arrangement is shown in Fig.3. A set of optical transparencies are created by computer. Two examples for $N = 4$ are shown in Fig. 4. The dot on the left in each case is positioned at one of the allowable input positions. The spots on the right are the points at the output to which we wish to connect each input point. All the output points must of course be placed at allowable output locations. A separate transparency is made for each of the input points and each transparency is placed at the input plane in Fig.3. The system is a joint transform arrangement that records in an iron doped $LiNbO_3$ crystal a hologram of the desired connectivity for each point. Once a hologram for each input point has been recorded, we test the system by placing an input transparency in which all the allowable input points are "on", and observe the output plane on a CCD camera. Fig. 5a shows the input pattern used in the experiment. The hologram was exposed to the two interconnect p ttern shown in Fig. 4. The light diffracted at the output plane is shown in Fig. 5b. It contains a total of 5 spots rather than the expected 3. Two of these spots however are at output locations that are not allowable. Since neurons are never placed in the prohibited locations, light that is diffracted there will be inconsequential. To demonstrate this idea, a mask was placed at the output plane transmitting light only at the allowable output locations. This is shown in Fig. 5c showing only 3 outputs ports receiving light from the 8 input points, as expected.

References

[1] Psaltis, D. and N. Farhat, *Optics Letters*, vol. 10, p. 98, 1985.

[2] Psaltis, D., in *Optical Bistability III*, Gibbs et. al. eds., Springer-Verlag, p. 1, Berlin Heidelberg, 1985.

[3] Psaltis, D., *Proceedings of the FJCC IEEE Conf.*, IEEE Cat. No. 86ch2345-7, Dallas, Texas, November 1986.

[3] Van Heerden, P., *Applied Optics*, vol. 2, p. 387, 1963.

[4] Guest, C. and T. K. Gaylord, *Applied Optics*, vol. 19, p. 1201, 19
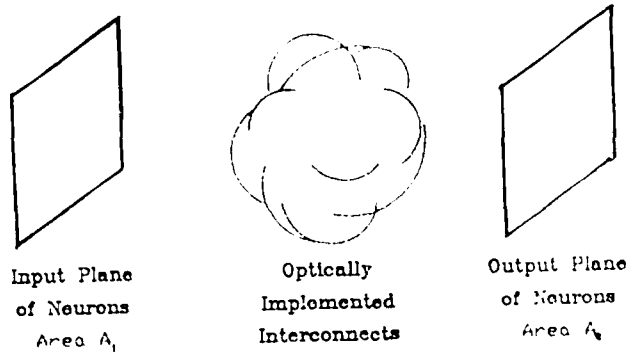
[5] Yu, J, F. Mok, and D. Psaltis, Submitted to *Applied Optics*.
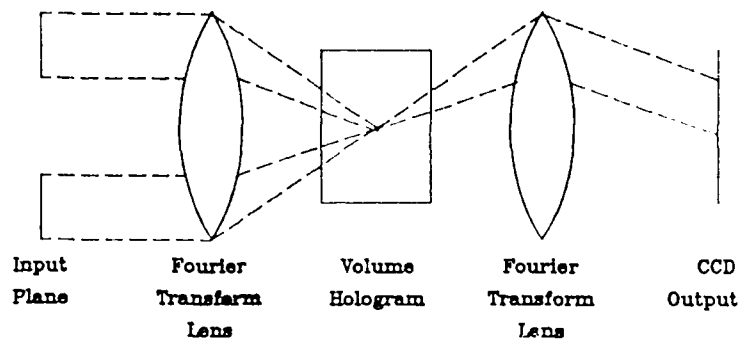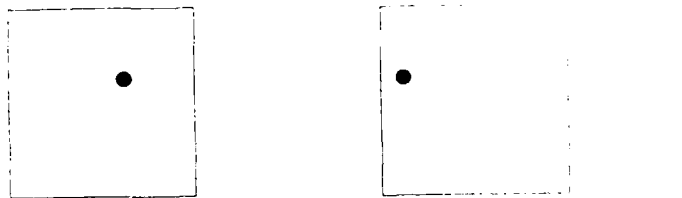
Figure 1

Figure 3
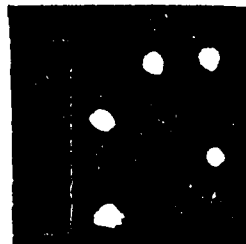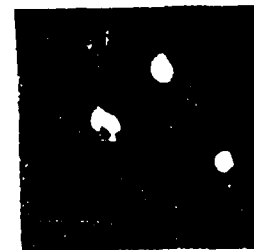

Figure 4a

Figure 4b

Figure 5a

Figure 5b

Figure 5c

Multilayer Optical Learning Networks

Kelvin Wagner and Demetri Psaltis
California Institute of Technology
Pasadena Ca. 91125

In this paper we present a new approach to multilayer neural network learning which is based on holographically interconnected nonlinear Fabry-Perot etalons. The network can learn the interconnections that form a distributed representation of a desired pattern transformation operation. The interconnections are formed in an adaptive and self aligning fashion, as volume holographic gratings in photorefractive crystals. Parallel arrays of globally space integrated inner products diffracted by the interconnecting hologram illuminate arrays of nonlinear Fabry-Perot etalons for fast thresholding of the transformed patterns. A phase conjugated reference wave interferes with a backwards propagating error signal to form holographic interference patterns which are time integrated in the volume of the photorefractive crystal in order to slowly modify and learn the appropriate self aligning interconnections. A holographic implementation of a single layer perceptron learning procedure is presented that can be extended to a multilayer learning network through an optical implementation of the backward error propagation (BEP) algorithm[1].

The learning algorithm in a single layer optical perceptron begins with the repetitive presentation to the network of the set of training patterns in random sequence. Initially, the holographic interconnection and output nonlinearity gives rise to a sequence of output patterns which are different from the desired target response sequence. An error pattern is formed, either electronically or optically, by taking the difference between the actual output pattern and the targeted response. The difference pattern is sent backwards through the network with a different polarization, or a slightly different wavelength, or pulsed at a slightly jittered time, than the forward propagating pattern, in order to avoid interference between the forward and backward waves. Meanwhile, the undiffracted portion of the input pattern is phase conjugated by an auxiliary phase conjugate mirror, which retroreflects each component of the input wavefront back towards the position at the input from which it originated. The phase conjugate beam has the polarization rotated or the wavelength shifted in order to act as a self aligning reference beam for the backwards propagating error wavefront. A volume hologram is recorded within the photorefractive crystal between the phase conjugated input pattern and the backwards propagating error signal. This is mathematically equivalent to changing the holographic connectivity matrix by the outer product of signal and error pattern vectors[2]. The next time that this particular input pattern is presented to the network, it will produce a diffraction pattern that more closely resembles the desired output pattern. Eventually the hologram will learn the correspondence between a set of input patterns and the associated responses as long as the set of input patterns are linearly separable, which implies that a holographic interconnection can be found that will produce the desired pattern transformation. Since the holographic reference wave is generated by a phase conjugate mirror, as the network learns it will also self align as well as correct for some of the optical imperfections present in the system components.

When the desired pattern transformation is not linearly separable, as in most difficult problems of interest, it is necessary to adaptively implement more complex nonlinear decision surfaces[3]. This can be accomplished by backwards propagating the error signal through a trainable multilayer network of holographically interconnected nonlinear devices. When the error pattern strikes the hologram part of it is diffracted towards the previous layer of hidden units by the exact same interconnection matrix seen by the forward propagating patterns. The BEP algorithm requires that the transmission function of the hidden units to backwards propagating signals be the derivative of the forward mode sigmoid transfer function evaluated at the current operating level of each device. The derivative is peaked where the nonlinear sigmoid transfer characteristic has a large differential gain, so that if the hidden unit is operating in this region, the connections leading to it will be strongly modified by the efficiently transmitted error signal. The interconnections will be continuously modified until all the patterns within the training set produce outputs very near the flat upper or lower levels of the nonlinear device sigmoid response, so that the error signals are not allowed to back propagate through the network. When convergence is reached the error signals that are generated at the final layer become very small for all members of the training set.

Optical systems cannot implement the idealized derivative transmission, but a similar peaked response can be obtained by operating the nonlinear etalons in the probe mode[4] for the backwards propagating error signal. In this mode the Fabry-Perot resonance is scanned by the nonlinear dependance of the index on the intracavity intensity, which varies in response to the high power forward beam intensity. The weak backwards propagating probe beam is modulated by the current state of the cavity transmission function. The probe mode transmission is peaked at the resonance of the Fabry Perot, which occurs when the sigmoid response to the forward beam reaches the upper level. The peak maximum is not exactly at the region of the highest slope of the forward beams nonlinear sigmoid response, but since the forward and backward beams are different polarizations, or different wavelengths, the resonance function can be offset in order to achieve a properly positioned probe beam resonance peak. In the polarization multiplexed case this shift

TuA4-2

can be induced by including a thin birefringent sheet in the cavity, or perhaps a tunable birefringence can be caused by applying a static external field on the cavity. This type of birefringent nonlinear Fabry-Perot etalon and a simulation of the forward mode sigmoid transfer function is shown in Figure 1, along with its derivative and the shifted probe mode response. This device can implement the desired sigmoid nonlinearity of the high intensity forward propagating signals with a differential gain greater than one, although the actual gain in transmission is usually much less than one. The probe mode response is not symmetric about the peak because the Airy function resonance is scanned by the intracavity intensity which is equal to the transmitted sigmoid response divided by one minus the back mirror reflectivity. This asymmetry will continue to allow signals that are above threshold to build up interconnection gratings in the previous layer, corresponding to correlated inputs, thereby compensating for the slow forgetting of gratings by the volume hologram. By decreasing the Q of the cavity to the forward propagating beam a tradeoff can be made between the peak width of the probe mode response, and the switching energy for the forward propagating nonlinear device characteristic. Another possibility, illustrated in Figure 2, would be to use two closely spaced cavities, both addressed by the same forward and backwards propagating resolution spots. In this case one cavity is optimized to produce a sigmoid response of the forward propagating beam while blocking the backwards propagating error signal, while the other cavity is resonant to the backwards propagating beam and the Fabry-Perot resonance is linearly scanned by the 100% reflected forward propagating mode intensity. We expect that convergence can be achieved with the forward and backwards response that can be obtained from these scanned resonance devices, even though the responses do not precisely match the nominal responses of the BEP algorithm, because of the robustness of this learning procedure.

An architecture that can perform this type of multilayer perceptron learning procedure, using polarization multiplexing of the forward propagating processing beam and backwards propagating teaching beam is shown in Figure 3. The illustrated architecture is one implementation of this class of backwards error propagating holographic learning machines that serves to illustrate the principles involved. Notice that no lenses are shown in this diagram because the volume hologram can perform the desired weighted interconnection imaging by exposing it with the proper expanding image and focussing reference beam to form a thick volume hologram. If Fourier lenses are inserted between the etalon arrays and the volume holographic crystal then the exposed hologram will be a Fourier hologram with planar fringes, and the k-space analysis will be simplified, but the processor learning and self aligning operations will be similar. The interference of a backwards propagating error signal emerging from a particular etalon at the output with the phase conjugated forward propagating beam emerging from a particular etalon from the input, will produce a volume Fresnel holographic interference pattern that will connect these two etalons for both forward and backwards propagating beams, due to the reciprocity of linear electromagnetic systems. The indicated nonreciprocal polarization filtering will remove the unwanted reflections from the nonlinear etalons, as well as the phase conjugated reference after it has been used to expose the volume hologram. Each layer is completely compatible with the previous and the following layers. Therefore this type of learning network can be stacked up to form a complex multilayer learning machine.

This implementation is based on a polarization switching diffraction mechanism that takes place in some electrooptic volume holographic materials, such as $Bi_{12}SiO_{20}$, $LiNbO_3$, and $BaTiO_3$. The polarization switching diffraction efficiency, and the holographic storage capacity can be simultaneously maximized by having the input and output beams propagating at large angles. The unwanted polarization switch grating exposures due to the simultaneous presence of multiple reference (or object) beams will produce a cross talk of the undiffracted forward propagating beam, which can be eliminated with the appropriate polarization filtering. The learning operation must occur slowly for the algorithm to converge properly and this is well matched with ferroelectric photorefractive crystal volume holograms, in which the crystal response times are slow and the perturbation of an existing space-charge grating by a single outer product exposure is small. It is necessary to be able to both selectively erase holographic gratings, thus decreasing the connection strength between particular etalons, as well as to strengthen individual gratings, thus increasing the corresponding elements of the interconnection matrix. Selective erasure can be accomplished by using a phase encoded backwards propagating error signal, where a phase angle of 0 is used to represent all positive error signals, and a phase angle of π is used to represent all negative error signals. Gratings that are built up with a phase angle of 0, can have the corresponding interconnection decreased by shifting the recording interference profile by π. Alternatively selective interconnection erasure can be accomplished by strengthening interconnection gratings when the applied bias field is in one direction, causing the resulting space-charge grating to shift away from the optical intensity profile in the direction of the field by approximately $\pi/2$, while decreasing interconnection gratings when the bias field is reversed, producing a cancelling space charge grating with a phase shift of $-\pi/2$. Another approach to decreasing interconnection strength would be to rely on the simultaneous erasure of all the gratings by the readout mechanism or thermal effects, thereby inserting a "forgetting" term in the dynamical equation for the holographically represented interconnection matrix. This approach requires continuous reinforcement to avoid forgetting everything that has been learned. A scheme must be devised to implement negative interconnection strengths, or else all the signals must be placed on a bias. Another possibility is to use the phase shift of each hologram to represent its sign, and count on destructive interference within each nonlinear etalon to subtract the

134

positively and negatively weighted diffracted components. The storage capacity of the volume hologram will enforce limits on the number of nonlinear devices that can be interconnected and upon their topology. A sparse array of etalons will have to be utilized in order to implement a fully global interconnection without unwanted cross talk, which will also facilitate the dissipation of heat generated in the nonlinear etalons.

A complete system will require a high speed method of entering data for pattern transformation processing, and another means of introducing backwards propagating error signals for the learning phase. Probably the best approach to high speed data entry at the back end of the system would be to use a sparse, parallel laser diode array or a fiber optic input array, demagnified onto the first layer bistable nonlinear etalon array, in order to modulate the coherent bias beams transmitted by each addressed device, thereby using the nonlinear Fabry-Perot etalon array as a high speed incoherent to coherent converter with memory. At the final layer of the system error signals need to be computed, and injected back into the system with the appropriate polarization or wavelength, and the phase shift or timing needed to represent the sign of the error. When the number of outputs of the pattern transformation procedure is less than 1000, they can be arrayed in a linear format which allows the utilization of high speed linear detector arrays for output, and the utilization of linear spatial light modulators, in order to introduce the backwards propagating error signals. The fan-out capability of each layer is determined by the nonlinear device gain and the holographic diffraction efficiency, and it may dictate an information collapsing network. For example if the product of nonlinear device gain times holographic diffraction efficiency is only 1:30, then a network with 30,000 bit input patterns might be processed by 1000 hidden units that communicate with 30 output devices, simplifying the error generation process. The ability of the system to process large amounts of data in parallel at a very high speed is limited by the electronic addressing of the input array, and the output photodetector array readout time, and not by the intervening optical system, because of the extremely fast response achievable with nonlinear etalons.

The forward propagating signal can be a narrow pulse since the response of GaAs nonlinear Fabry-Perot etalon is determined by the peak power incident. In this case the backwards propagating error signal can be either pulsed or CW. In the pulsed mode the PCM would need to have practically instantaneous response, such as a nonlinear optical semiconductor might provide, and the forward and backward propagating pulses could be time jittered so they do not overlap in the volume hologram, but the phase conjugate reference and the backwards propagating error pulse would overlap within the crystal, thereby exposing a hologram. Alternatively, the backwards propagating error signal could be a low power CW beam that would not nonlinearly modify the index within the Fabry-Perot etalons, and the forward propagating pulse could be turned into a quasi CW phase conjugated reference by using a ferroelectric crystal based PCM which has a slow, integrated response. The Fabry-Perot etalons would need to have a slow relaxation time of the nonlinearly shifted index, so the probe beam would have the appropriate response for most of the interval between pulses of the forward beams. In this case the holographic exposure would be due to the time integral of the CW waves in the volume hologram, and the orthogonally polarized pulsed forward propagating beam would not contribute to the hologram exposure.

The non ideal optical implementation may actually have improved performance over that of an idealized digital simulation because noise will always be present in the system, helping it to avoid shallow local minima, and pushing the interconnection matrix away from solution boundaries. Imperfections of the holographic interconnection will help the system perform symmetry breaking, which the idealized model cannot perform spontaneously. The simultaneous self aligning and learning of the optical system make this approach to multilayer optical neural processing experimentally feasible, and allow the implementation of complicated systems that could not be completely specified a priori, but can be learned and modified as the desired processing operation slowly changes. The slow learning of the holographic crystals combined with the extremely high speed processing of the nonlinear etalons gives this system an enormous throughput potential and the capability for solving complicated but learnable problems. The added possibility of feedback between layers would result in a dynamical processing system reminiscent of Hopfield's neural networks[6,7], but the dynamic interconnections give the trainable network an additional adaptive problem solving capability.

References.

1. D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel Distributed Processing, Vol. 1, Ch. 8*, D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning internal representations by error propagation, Cambridge, Ma., MIT Press, (1986)

2. M. Cohen, Design of a new medium for volume holographic information processing, Appl. Opt., vol. 25(14), p. 2288 (1986).

3. D. Psaltis and C. Park, Nonlinear discriminant functions and associative memories, Proc. Snowbird conference on neural networks for computing, APS (1986).

4. H. M. Gibbs, et al., Optical modulation by optical tuning of a cavity, Appl. Phys. Lett., Vol. 34(8), p. 511 (1979).

5. A. W. Lohman, Polarization and optical logic, Appl. Opt., Vol. 25(10), p. 1594 (1986).

6. J. J. Hopfield, Neurons with graded response have collective computational properties like those of two-state neurons, Proc. Natl. Acad. Sci. USA, Vol. 81, p. 3088 (1984).

7. D. Psaltis and N. H. Farhat, Optical information processing base on an associative memory model of neural nets with thresholding and feedback, Opt. Lett., vol. 10(2), p. 98 (1985).

COUNTERPROPOGATING PUMP-PROBE NONLINEAR FABRY-PEROT
WITH INTRACAVITY BIREFRINGENT PLATE

BEP DUAL NONLINEAR FABRY-PEROT CAVITY CAVITY

NLFP Transmission and Derivative with Birefringent

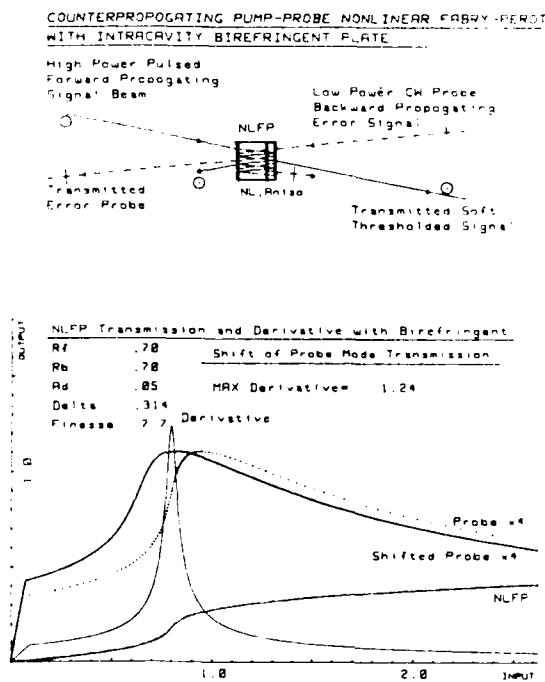NLFP Transmission and Derivative and Aux

Figure 1. Nonlinear Fabry-Perot etalon sigmoid response, its derivative and the probe mode response for the two polarizations, with an auxiliary intracavity birefringence.
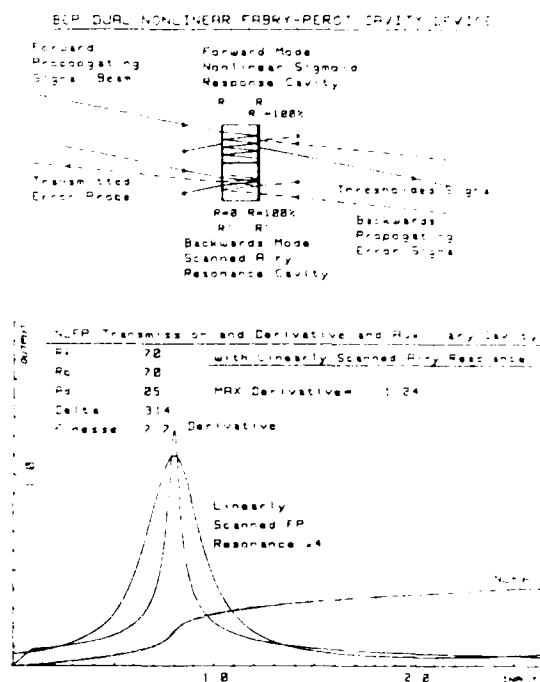
Figure 2. Dual cavity nonlinear Fabry-Perot etalon with its forward propagating nonlinear sigmoid response and backwards propagating linearly scanned Airy resonance probe mode response.

SINGLE LAYER OF BEP OPTICAL LEARNING NETWORK
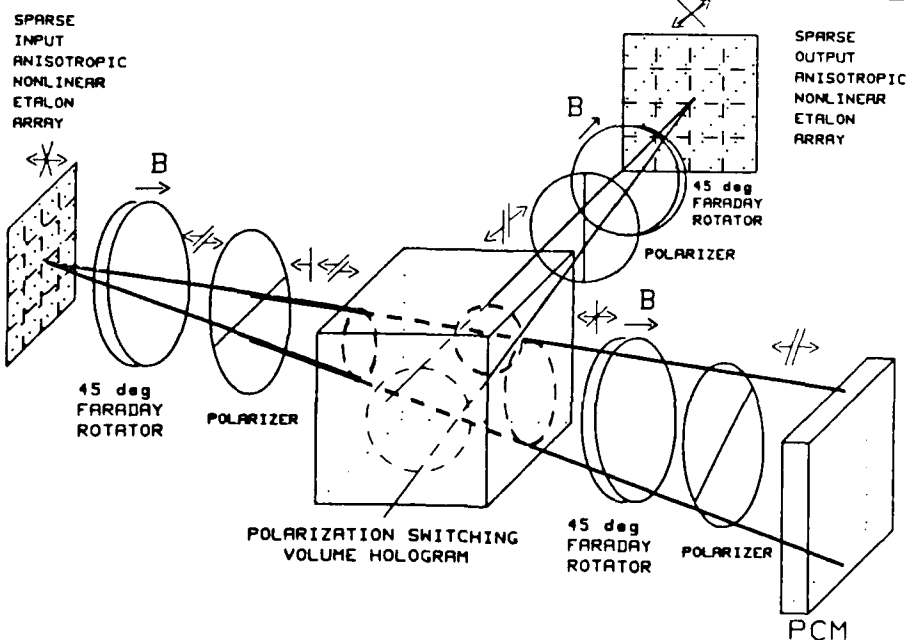
Figure 3. Optical backward error propagation architecture with polarization multiplexed forward and backward waves, nonreciprocal polarization filtering, and self-aligning polarization-switching volume hologram.

Optical associative processing elements with versatile adaptive learning capabilities

Arthur D. Fisher and John N. Lee

Optical Sciences Division, Naval Research Lab, Washington, DC 20375.

Introduction

We are investigating associative processing architectures for tackling the massive parallel symbolic processing demands of problems from image understanding, robotic manipulation and locomotion, expert-system problem solving, and other difficult artificial intelligence domains. A modular approach is taken, where a number of smaller adaptive, associative "modules" are nonlinearly interconnected and cascaded under the guidance of a variety of "organizational principles" to construct larger architectures for solving specific problems.[1,2] Each module is a complete associative memory which adapts as it is exposed to associated information patterns (u,v), (e.g., feature vectors, encoded symbols, images,...), so that subsequent presentation of one pattern u results in recall of its paired pattern v.

Adaptivity of the individual modules assumes a central role in this approach. The information required for successful performance in real-world applications, such as image understanding, is often too extensive and detailed to be prespecified. Adaptive learning provides a more practical means for selecting, acquiring, and structuring the relevant knowledge, as well as fine-tuning and extending the underlying procedures or "algorithms". The multi-module architecture becomes, in effect, a "knowledge filter", ideally only gathering information relevant to its designed task(s) and tending to avoid saturation with inappropriate information.

Its inherent parallel processing and interconnection capabilities makes optics an attractive medium for expressing the innate parallelism of these associative concepts. Furthermore, these systems are often low-precision or binary and hence are compatible with the limited dynamic range capabilites of optics. Four optical adaptive, associative module implementations, including both electrooptic and holographic configurations, are briefly outlined. All these modules are optically cascadable, with all inputs and outputs in the form of 1-D or 2-D image beams or intensity arrays.

Widrow-Hoff Learning-Rule Module

The electrooptic associative implementation of Fig. 1 performs real-time learning of m pairs of associated n-element vectors (u^k, v^k) by the Widrow-Hoff[3] (or least mean square) dynamic equation-of-learning, $dM/dt = g'(v^k - Mu^k)u^{kT}$. Here g' is a gain factor, k designates a particular associated vector pair, and T is the matrix transpose operator. Associations are retrieved by the equation-of-recall, $v = Mu$. Note that changes in M are driven by the difference between the desired output v^k and the actual output v. The current state of the memory matix M is stored as an electronic charge distribution in the microchannel spatial light modulator[4] $MSLM_1$.

The operation of the configuration in Fig. 1 has been described previously.[2] Briefly, v^k is stretched in one dimension (along z) onto the detector side (D) of $MSLM_2$, and u^k is perpendicularly stretched (along y) and reflected from the modulator side (M) of $MSLM_2$ to produce the outer product $v^k u^{kT}$ (or vu^{kT}, with feedback of v by rotating m'_4 to m_4). In recall, u^k is stretched along x , multiplied by M in reflecting from $MSLM_1$, and then compressed along z to form $v = Mu^k$. The difference equation form, $M_{n+1} = M_n + gv^k u^{kT} - gvu^{kT}$, is implemented by utilizing the capability[4] of $MSLM_1$ to add or subtract charge to/from its stored image.

Aside from being heteroassociative and adaptive, this architecture exhibits <u>incremental learning</u>, whereby the learning of an associated pair improves[2] on successive encounters (need not be sequential). It is usually operated in a <u>gated-learning</u> mode, where it spends most of its time performing recall-only, without adaptation; learning is gated on only when another part of the system signals that a significant event has occurred. The negative feedback term tends to: 1) prevent saturation of the dynamic range of M, by allowing m_{ij} to decrease as well as increase , 2) implement controlled forgetting, with newer associations replacing older obsolete associations, when the information capacity ($\leq n$ pairs) is exceeded, and 3) correct for aberrations of the optical system by storing compensating modifications in M.

Hebbian Learning-Rule Module

With removal of the v feedback path , the associative module of Fig 1 implements the simpler Hebbian learning rule, $M_{n+1} = M_n + gv^k u^{kT}$, or equivalently $\Delta m_{ij} = gv_i^k u_j^k$. Unlike the full Widrow-Hoff configuration, this simpler formulation 1) requires orthogonality between the u^k vectors for perfect recall; and 2) easily saturates the dynamic range of M , since Δm_{ij} is always positive and never negative.

Differential Learning–Rule Module

The optical module in Fig. 2 implements differential dynamic learning equations of the form $d\mathbf{M}/dt = g'(d\mathbf{v}/dt)(d\mathbf{u}/dt)^T$. The advantages of differential learning rules of this type have been discussed by Kosko[5], Klopf[6], Barto and Sutton[7], and others. Instead of learning $(\mathbf{u},\mathbf{v})$ pairs which happen to be large, pairs are reinforced in which a change in $\mathbf{u}$ caused a change in $\mathbf{v}$. Among other properties, these rules may have enhanced "credit assignment" capabilities for 1) "back-propagating" the correct weight changes, Δm_{ij}, to intermediate layers which have contributed to a correct output result in a multi-module configuration, and 2) learning intermediate steps in a time sequence of events. Other features of the implementation of Fig. 2 include: incremental learning, gated learning, and resistance to saturation by allowing both negative and positive changes in $\mathbf{M}$. This module is shown in a two-port ($\mathbf{u}$-$\mathbf{v}$) flow-through configuration, which is the required geometry for application in some parts of a multi-module architecture. Alternatively, it can be operated with three-ports, ($\mathbf{u}^k$-$\mathbf{v}^k$-$\mathbf{v}$), when the training input $\mathbf{v}^k$ in Fig. 2 is available.

The MSLMs in Fig. 2 can be sequenced to generate a variety of specific learning rules, for example, the adaptive difference equation, $\mathbf{M}_{n+1} = \mathbf{M}_n + g(\mathbf{v}_{n+1} - \mathbf{v}_n)(\mathbf{u}_{n+1} - \mathbf{u}_n)^T$, or the "lagged conjunction"[5] relation, $\mathbf{M}_{n+1} = \mathbf{M}_{n-1} + g(\mathbf{v}^k_{n+1} - \mathbf{v}^k_n)(\mathbf{u}_n - \mathbf{u}_{n-1})^T$. The differences, e.g., $(\mathbf{u}_{n+1} - \mathbf{u}_n)$ and $(\mathbf{v}_{n+1} - \mathbf{v}_n)$, are computed by switching MSLM$_1$ and MSLM$_2$ between their addition and subtraction modes. These $\Delta\mathbf{u}$ and $\Delta\mathbf{v}$ difference vectors, which are stretched in perpendicular directions, are multiplied by MSLM$_2$ to form the required outer-product. The resulting $\mathbf{M}$ matrix is accumulated and stored in MSLM$_3$. MSLM$_3$ also multiplies $\mathbf{M}$ by $\mathbf{u}$ to form the associative output $\mathbf{v} = \mathbf{M}\mathbf{u}$, with anamorphic imaging and stretching optics similar to those of Fig. 1. In gated-learning operation, MSLM$_1$ and MSLM$_2$ are continuously updated, but MSLM$_3$ is only activated when a learning cycle is desired.

Holographic Modules

Holographic configurations of the form depicted in Fig. 3 are also under investigation and may result in reduced module complexity. To be useful this module must be capable of adaptively learning a large number of associated 2-D or 1-D images (u^k,v^k). The H element is a real-time, reusable holographic storage medium, such as a thermoplastic film or a volume bulk photorefractive material. This module is operated in a gated-learning mode, where it spends most of its time performing nonadaptive, nondecaying holographic readout; learning is turned on only in short bursts to capture significant events. Depending on the specific materials employed, learning can be activated by a pulse of increased light intensity in u^k and v^k, heat, flood illumination (e.g., for heating or a two-photon material), and/or bias voltage. This is designed to be an incremental-learning process, where each exposure to a new associated pair superimposes a weak component to the holographic gratings, and multiple exposures to a given pair increase its strength. Some materials will tend to avoid dynamic-range saturation by eventually replacing old associations with new associations through a process of conservative redistribution of charges (photorefractive) or material (thermoplastic).

With Fourier transforming optics on the u^k, v^k and v paths in Fig. 3, and no R element, presentation of the pattern v^p recalls the output distribution of Eq.(1a), where the $\bullet$ and O represent convolution and correlation, respectively[8,9]. Alternatively, with imaging of u^k, v^k and v to and from the hologram, the output is given by Eq. (1b), with the $*$ indicating complex conjugation.[9]

$$v = \Sigma_k v^k \bullet (u^k O u^p) \qquad \text{(Fourier)} \quad (1a), \qquad\qquad v = |u^p|^2 v^p + \Sigma_{k \neq p} (u^{k*} u^p) v^k \qquad \text{(imaging)} \quad (1b)$$

The R element in Fig. 3 serves the critical role of recoding the u^k patterns to obtain a usefully large information capacity; there are a variety of possibilities for its implementation. Angular encoding, which corresponds to each u^k being a uniquely-angled, plane-wave in the imaging configuration of Eq.(1b), or a displaced point (delta function) in the Fourier configuration of Eq.(1a), gives the desired recall of $v \approx v^p$. By utilizing the excellent Bragg angular selectivity of volume-photorefractive holograms, very large information capacities can be realized.[10] Unfortunately, assigning a unique angle to each u^k pattern, and the same angle to repeated or nearly identical u^k patterns is problematical. One hypothetical possiblity is to make R a "hash table" which produces a unique angle for nearly any possible u^k pattern (given a finite resolution and dynamic range in u^k). An approximation to this is to use apriori information about the problem domain to prerecord a holographic lookup table in R assigning reference beam angles to expected u^k patterns,

A more general approach is to create an R hologram which maps an orthogonal decomposition, such as Walsh or Fourier components, into unique reference beam directions. A thresholding operation is then required to clean up u^k.

A quite different approach, also applicable to "thin" holograms, e.g., thermoplastic, is to employ R to add unique high frequency structure to u^k (in the Fourier configuration of Eq.(1a)), for example through edge enchancement by a high-pass spatial filter. An MSLM could be used[4] to convert u^k to an edge-enhanced, phase-only image; which is also applicable to the imaging configuration of Eq.(1b), particularly when the output is followed by a low-pass filter and thresholding. (It should be noted that multiplying every u^k pattern by the same fixed random-phase mask at R does not solve the encoding problem.) More esoteric approaches are also under consideration, such as extracting from H a summation of all previoius u^ks, and subtracting this (e.g., with an MSLM at R) to produce a novel or "orthogonalized" reference beam from the current u^k.

Still another approach is to limit u^k and v^k to one-dimensional (1-D) patterns and use one dimension for encoding. The two 1-D images $u^k(y)$ and $v^k(x)$ are stretched in perpendicular directions to record a 2-D hologram, which contains the term $\Sigma_k v^k(x) u^{k*}(y)$. The holographic output beam is Fourier transformed along only y (the $u^k(y)$ decoding direction), and passed through a slit along the x direction located at the zero-order of the Fourier plane. Presentation of the pattern $v^p(x)$ then recalls the output distribution $\Sigma_k v^k(x)[\int u^{k*}(y) u^p(y) dy]$. For uncorrelated u^k patterns, the overlap integral, which is the peak of the crosscorrelation, is small and $v^k(x)$ is recalled. This is essentially identical to the outer-product formulation ($M = \Sigma_k v^k u^{kT}$ and $v = Mu$) discussed above, except that it takes a continuous, rather than dicrete-matrix form! It should be noted that since the SLMs and other optical components in Figs. 1 and 2 are continuous resolution devices, those architectures are not limited to matrices, but can also process continuous 1-D images! In fact, continuous images will result in an enhanced information capacity, which is ultimately proportional to the number of resolvable pixels.

Concluding Remarks

Although only the learning dynamics have been emphasized here, the overall performance of an associative architecture also depends on the choice of recall dynamics (which can generally be expressed as a differential or difference equation in v). The recall formulation determines, for example, whether response to distorted or partial inputs is exact[11] recall of, similar to, or a superposition of stored patterns; or is even random patterns or a null result. Most of the optical modules mentioned above directly implement simple static recall, e.g., $v = Mu$; however these modules can be configured in larger systems to implement a variety of recall dynamics. In multi-module and recursive configurations, it is assumed that the module output is followed by a nonlinear operation. For example, a follow-on MSLM can implement[4] versatile image thresholding operations. In some instances it is also desirable to insert a spatial light modulator to implement short-term-memory decay dynamics in v. Other useful interconnections include feedback of v to parts of u^k and or v^k, or optical Fourier transforms to allow shift-invariant patterns to be stored and recalled. These adaptive associative modules can also be employed in nonlinear-resonator[12] recall configurations.

References

1. A. D. Fisher, R. C. Fukuda, and J. N. Lee, Proc. SPIE 625, p.196, (1986). and A. D. Fisher, C. L. Giles, and J. N. Lee, J. Opt. Soc. Am. A 1, p. 1337 (1984).

2. A. D. Fisher, "On applying associative architectures to the solution of interesting problems", in preparation.

3. B. Widrow and M. E. Hoff, IRE WESCON Convention Record, part IV, p.99 (1960).

4. C. Warde, A. M. Weiss, A. D. Fisher, and J. I. Thackara, Appl. Opt. 20, p. 2066 (1981). and J. A. McEwan, A. D. Fisher, and J. N. Lee, Conference on Lasers and Electrooptics (CLEO), p. PD-1 (1985).

5. B. Kosko and J. S. Limm, Proc. SPIE 579 (1985).

6. A. H. Klopf, Proc. of the Neural Networks for Computing Conference, Snowbird, UT (1986).

7. A. G. Barto and R. S. Sutton, in Synaptic Modification, Neuron Selectivity, and Nervous System Organization (W. B. Levy et al., eds.) Lawrence Earlbaum, Hillsdale, NJ (1985)

8. D. Gabor, "Associative holographic memories", IBM J. Res.Devel., p.156 (1969).

9. G. R. Knight, "Page-oriented associative holographic memory", Appl. Opt. 13, p.904 (1974).

10. H. Kogelnick, Bell Sys. Tech. J. 48, p.2209 (1969).

11. J. J. Hopfield, Proc. Natl. Acad. Sci. 79, p.2554 (1982).
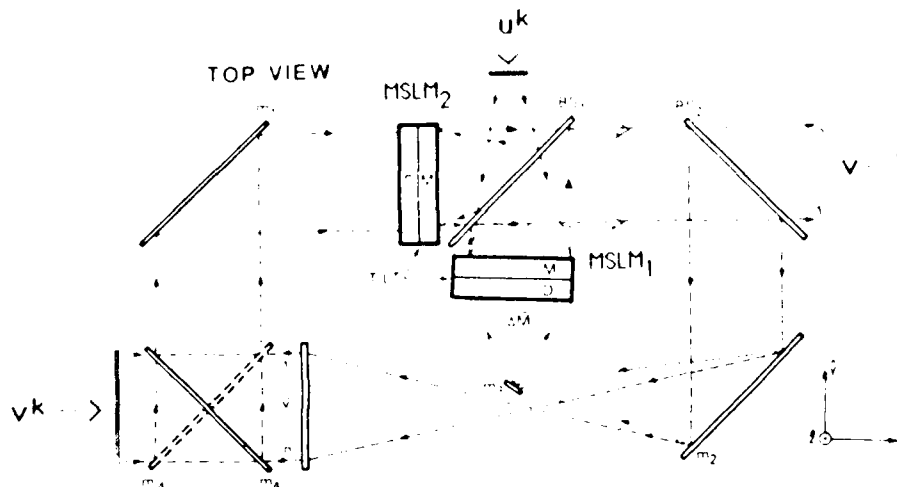
12. D. Z. Anderson, Opt. Lett. 11, p. 56 (1986).

Fig. 1   Associative module implementing Widrow-Hoff adaptive learning rule. (Top View). BS_n's are beamsplitters, m_n's are mirrors, D is the input detector side and M is the output reflective-modulator side of the MSLM's (microchannel spatial light modulators).
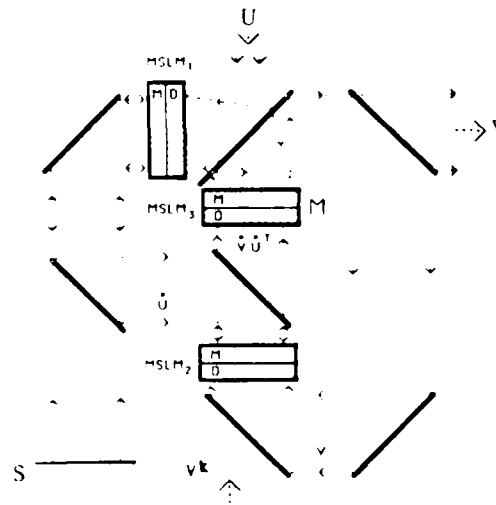

Fig. 2   Associative module implementing learning rules of the form $dM/dt = g'(dv/dt)(du/dt)^T$. (Top View). D is the input detector side and M is the output reflective-modulator side of the MSLM's (microchannel spatial light modulators), S is a uniform two-dimensional light source.
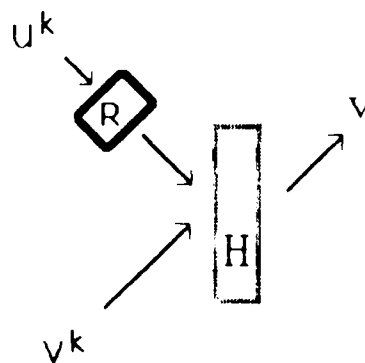
Fig. 3   Holographic adaptive, associative module.   H is a real-time, reusable holographic medium and R implements recoding operations to increase information capacity.

Optical Symbolic Computing: Architectural Considerations

M. Derstine
*Honeywell Physical Science Center
10701 Lyndale Ave. S
Bloomington, MN 55420

A. Guha, R. Ramnarayan
Honeywell Corporate Systems Development Division
1000 Boone Ave N
Golden Valley, MN 55427

A. Vaid
Department of Electrical Engineering
University of Southern California
Los Angeles, CA 90086

P.R. Haugen*, A. Husain*

Many researchers have suggested that the parallelism of optics might be exploited for symbolic processing applications.[1-8] Optics can perform functions needed for symbolic computing such as searching (using correlations) and high bandwidth data transfer (imaging). It is, however, an open question as to the applicability of optics to an overall system which does general and nontrivial symbolic computing. This paper examines how optics could be used within the framework of implementing currently specified computer languages.

LANGUAGES

Languages for symbolic computing can be divided into three categories: imperative, logic, and functional. Imperative languages, particularly LISP (as it is used today), form the basis of current expert system shells. Due to the assignment operation, the execution of an imperative language program can be viewed as a series of changes to a large state space. To implement a language which executes in this way, important primitive operations would include memory access and compare. Imperative languages are inherently serial because of the need to operate upon a well defined state space.

The best known logic language is Prolog, which has received wide-spread attention as a result of the Japanese Fifth Generation effort. Concurrent logic programming languages, such as Concurrent Prolog and PARLOG, alleviate problems posed by the sequential semantics of Prolog. These languages use complex data structures such as graphs and trees to represent the program and the data. Computation can be viewed as various types of unification between different data structures.[9]

Programs in functional languages are essentially definitions and applications of functions. Pure functional languages,[10] such as pure LISP, compute by value and not by effect, and functions are used to compute new

values from old. There are basically two computational models for func-
tional languages: dataflow and reduction. Both of these are amenable to
parallel computation and, like imperative and logic languages, require
the maintenance and manipulation of complicated data structures.

In summary, one common feature of these types of AI languages is that
manipulations of data structures are critical computational primitives.
Moreover, the execution of concurrent logic languages and functional
languages can be described as the reduction of a graph which represents
the program.[9],[10] We can conclude that data structure representation in
optical computers must be done efficiently and expose some parallelism.

DATA STRUCTURE REPRESENTATION

Representing data by graphs and trees implies that some mechanism must
be used to express the connections. Traditional computer designs handle
this problem through the use of pointers. Pointers are typically
addresses of locations where other data items are stored. This approach
to the representation of complex data structures is attractive because it
allows complicated relationships to be efficiently stored and modified.
It does require, however, that the machine possess addressable memory and
a separate processor. This separation between memory and processor is
needed to allow the processor to have some "knowledge" of the way data is
stored. This knowledge is required because the computational models
require the processor to explictly store and retrieve the data struc-
tures. If the memory were merged with the processor, the combination
would have to have some explicit "knowledge" about its organization, a
topic for a much deeper discussion than possible here.[11]

Another problem is that these data structures must be represented
exactly, again because of the explicit and exact nature in the computa-
tional model. The use of complex data structures to represent the pro-
gram and the data for current languages implies that the representations
must be exact. Errors in the data structure representations could have
such extreme consequences as "forgetting" portions of the program, losing
track of where the program is executing, or corrupting the working
memory. Analog representation could be employed if the probability of
error was sufficiently low, but in practice, digital systems are the pre-
ferred choice. This does not imply that all of the computation must nec-
essarily be digital. However, as most operations involve changes and
comparison of data structures, the use of analog optical processors for
the matching may do little to improve overall system performance.

OPTICS

All approaches to provide the primitive operations required must take
into account the overall nature of the task. Since data structure mani-
pulations have been identified as critical and difficult, we examine opti-
cal approaches to complex data representation and manipulation.

Operations like searching and matching of digital data items could still
be performed using correlations. However, since the functions are all
manipulations on data structures, correlations cannot be employed unless

the data structure can be represented as an entity rather than as items connected together. At present, this type of representation is difficult to achieve in a optical computer because the data structures change, requiring a means for selecting, adding, deleting, splitting and joining.

One solution to addressable memory is to actually construct memory which has binary addresses. The problem with this approach has been the difficulty in generating the decoding addresses. We have developed a possible approach for constructing an address decoder which employs the inherent parallelism of optics to reduce the number of devices required as compared to electronics.

A totally different approach would be to develop a computing structure which does not require addressable memory. The optical finite state machine (OFSM)[1],[2] is such an architecture. Unlike conventional electronic computers, this architecture does not separate the memory from the processor. The conventional way to design a finite state machine is to enumerate all the possible inputs, outputs and next states, and then develop some combinatorial logic to perform that function. However, design of a system with over 10^{12} states (assuming a 1000 x 1000 array of optical gates) is practically impossible when done in this manner. Such an effort would be tantamount to specifying all of the possible data structures, all the values of the data items, and the answer to the computation at the time the machine is designed.

The other approach to developing a finite state machine would be to specify the transition rules for the states in such a way as to avoid specifying all of them explicitly. Symbolic substitution is such a method.[2] It has the disadvantage that the machine is no longer massively interconnected because only pixels within a certain neighborhood can communicate directly. However, symbolic substitution does might be easily implemented[2] and may be able to employ high-speed (gigabit) optical components.[8] The use of this type of architecture will require the development of algorithms which provide addressable storage.

Another method for representing data structures is the use of adjacency matrices.[4],[7] Graph structures can be represented in a matrix structure by assigning nodes of the graph to rows and columns. When there is a connection between nodes an entry is made at the intersections of rows and columns of the two elements. A directed graph may be represented by using the rows to indicate the node the connection is from and the columns to indicate the node that is the destination. This scheme has the disadvantage that memory is used inefficiently; only a few connections are made between nodes, while there is memory allocated for any of the possible connections.

No addressing is required to check interconnections between data items; it is all present in the matrix. To set up the connections, however, some means is required to address and set/reset the elements of the matrix. This is made even more difficult when the elements to be added to the existing matrix make up another graph. To be added as rows and columns to the existing graph, the new subgraph must be rearranged. If elements were to be removed from the graph, some means would be needed either to

keep track of the empty rows and columns or to rearrange the graph so that the empty rows and columns are no longer in the interior of the data structure. Both of these methods require other data structures, such as linked lists, to keep track of the altered data. Thus to perform nontrivial operations on data stored in matrix format, some form of addressing must be used at some point.

CONCLUSIONS

Examination of the computation models of current computing languages shows that some way to perform addressable memory is required to represent and implement essential data structure manipulations. Optical devices and architectures may be able to provide the required functions, but development of better addressable memory architectures would greatly expand the number of computing applications where optics can play an significant role.

REFERENCES

1. A.A. Sawchuk and T.C. Strand, Proc. of IEEE, **72**, 758 (1984).
2. Alan Huang, Proc. of IEEE, **72**, 780 (1984).; K.-H. Brenner, A. Huang, and N. Streibl, Appl. Opt., **25**, 3054 (1986).
3. K.B. Jenkins and C.L. Giles, SPIE **625**, 22 (1986).
4. R.A. Schmidt and W.T. Cathey, SPIE **625**, 226 (1986).
5. J. Tanida and Y. Ichioka, JOSA **73**, 800(1983).
6. T. K. Gaylord et al, Opt. Eng. (1985).
7. C. Warde and J. Kottas, Appl. Opt., **25**, 940 (1986).
8. P.W. Smith and W.J. Tomlinson, IEEE Spectrum, **18** 26(1981).
9. R. Ramnarayan et al, 'Interim Report on Very Large Parallel Dataflow Program', Honeywell Corporate Systems Development Division, May, 1986; K. Clark and S. Gregory, 'PARLOG: Parallel Programming in Logic', Research Report DOC 84/4, June 1985, Department of Computing, Imperial College of Science and Technology, University of London.
10. P. Henderson, **Functional Programming Application and Implementation**, Prentice-Hall International, 1980; P. Hudak and B. Goldberg, IEEE Trans. on Comp., **C-34** 881 (1985).
11. D. R. Hofstadter, **Godel, Escher, Bach: An Eternal Golden Braid**, (Vintage Books, New York, 1980).

Comparison of Adaptive Pattern Recognition and Image Restoration with Hetero-associative and Auto-associative Memories

Jack Y. Jau, Y. Fainman and Sing H. Lee

Department of Electrical Engineering and Computer Sciences
University of California at San Diego
La Jolla, California 92093

I. Introduction

Our research group has recently begun an investigation on the relationship between the field of digital image processing and the field of artificial neural network. So far, we found much commonalities between the two fields in the algorithms they develop, although they frequently use different terminologies. For example, we found that image restoration is closely related to auto-associative memory, while pattern recognition is closely related to hetero-associative memory. As another example, background suppression in image processing is quite similar to *attentive associative memory*, while match-filtering process can be performed by *accretive associative memory*. In this paper we concentrate on the comparison of adaptive pattern recognition and iterative image restoration algorithms with *associative mappings*.

II. Comparison of Pattern Recognition Process with Hetero-associative Mappings

In a pattern recognition system, feature extraction is considered a process of mapping the observed samples to a set of more effective characteristics. That is

$$y = A x \tag{1}$$

where the sample of the i-th class is given by a $n \cdot 1$ vector x, and its corresponding features are defined by a $p \cdot 1$ vector y, and A is a $p \cdot n$ matrix. If A satisfies certain criterion, y could be the feature of x, and the linear transformation A can be seen as the feature extractor. In an *adaptive* pattern recognition system, the transformation A is calculated by successive adjustment. It follows that

$$A_{+1} = A - 2\rho A x - [(x - x^l). \tag{2}$$

where x is the l-th given input, $\Gamma(x)$ is the desired output of the mapping, and l is the index in discrete time domain, and $2\rho x^l$ is the gain factor Γ.

Kohonen [2] has described the relationship between the discriminant functions of the linear classifier and the optimal linear *associative mapping*. Consider a linear system described by

$$y = M x, \quad x \in R^n, \quad y \in R^p. \tag{3}$$

We may regard the p-dimensional output vector y as the memorized *data* and the n-dimensional input vector x as the *key pattern* by which y is encoded and retrieved. In *hetero-associative mapping* arbitrary *key patterns* can be paired with arbitrary output data via the $p \cdot n$ matrix M. If an adaptive *hetero-associative mapping* is considered, the new value M_k of M is a function of the previous M_{k-1} and of the new observation pair (x_k, y_k). It follows that

$$M_k \quad f(M_{k-1}, x_k, y_k) \quad M_{k-1} \cdot (y_k \quad M_{k-1} x_k) G_k . \tag{4}$$

where G_k is the gain vector. Therefore, we can consider that the system has two modes, *store* and *retrieve*. In *store* mode, the observation pair (x_k, y_k) must be given and Eq.(4) updates the *memory* M. In *retrieve* mode, only the *key pattern*, x_k, is needed in Eq.(3) to retrieve y_k. As a matter of fact, this kind of asymptotic transfer properties of adaptive system is very much equivalent to the orthogonal projection operations.

The feature extraction process given by Eq.(1) may be put in terms of neural network systems as pre-stored *hetero-associative mapping*, in which classes of patterns are directly mapped onto a set of discrete features. For example, the linear mapping-based [3] and the eigenvector based [4] algorithms may be considered as special cases of such mapping for pattern classification. This approach requires that the transformation (or mapping) A be calculated off-line; or, in other words, the *store* mode can not be performed on the pre-stored *hetero-associative memory*, A.

By comparing Eq.(2) with Eq.(4), we conclude that the adaptive pattern recognition system may be seen as an adaptive *hetero-associative memory*. The *store* mode of the *hetero-associative memory* is equivalent to the supervised on-line training of the adaptive pattern recognition system; and the *retrieve* mode of the *hetero-associative memory* is equivalent to extracting features from the adaptive pattern recognition system.

III. Comparison of Iterative Image Restoration with Auto-associative memory

The problem of image restoration concerns with the reconstruction of an image, f, from its *incomplete* or *partial information*, g. The problem can be formulated in terms of finding a filter which will reconstruct the desired image approximately as $\hat{f}$. It should be noted that g (the *incomplete information* of f) can be a space-truncated image, or a band-limited image [5,6], or the phase of f in the spatial frequency domain [7], or a noise-added image of f. In order to solve the image restoration problem, there exist various approaches to design the filter : (i) a prior knowledge design (e.g. Wiener filter) and (ii) iterative or recursive design (e.g. generalized alternating orthogonal projections [8], Kalman filtering [9], simulated annealing [10], phase retrieval algorithm [11], etc.).

In an iterative image restoration system, an image f can be restored by recursive computation of f.

$$f_{k+1} \quad g \quad Q_i P_i f_k, \quad \hat{f}_1 \quad g . \tag{5.a}$$

where g is the *incomplete information* of f. Q_i and P_i are orthogonal projection operators. It has been proved that $\hat{f}_k$ will converge to f as $k \to \infty$; i.e.

$$\lim \hat{f}_k \quad f . \tag{5.b}$$

For example, in Ref. 6 Q_a is designed as space-truncated operator, while P_b is designed as band limited operator.

In a neural network system the *accretive auto-associative memory* is implemented in a recursive fashion; that is

$$\hat{x}_{k+1} \quad \Phi x_0 \cdot \Lambda \Omega \hat{x}_k, \quad \hat{x}_1 \quad x_0 . \tag{6.a}$$

where the input *key pattern*, x_0, is a *fraction* (or *incomplete pattern*) of the expected output *data* x. and Φ, Λ and Ω are three operators. In each iteration $\hat{x}_k$ is updated by two operators, Λ and Ω. If

those operators are appropriately designed. $\dot{x}_k$ will converge to x as $k \to \infty$ [12]; i.e.

$$\lim_{k \to \infty} \dot{x}_k = x. \tag{6.b}$$

For example, in Hopfield's model [13] of neural networks Φ is chosen to be 0 (zero operator), Λ is a thresholding operator and Ω is a vector-matrix multiplication operator which can be implemented optically.

By comparing Eqs.(5.a) and (5.b) with Eqs.(6.a) and (6.b), it can be easily seen that the iterative image restoration process is mathematically identical to *accretive auto-associative memory*. If the input of *incomplete information* to an iterative image restoration system is considered as the *key pattern* to the *auto-associative memory*, the image restoration processes can also be regarded in terms of neural network as *auto-associative memory*.

VI. Optical Implementations

To implement adaptive and iterative processing, we are currently investigating a hybrid approach (see Figure 1) by combining an optical analog processor with a microcomputer. The optical analog processor performs the time-consuming operations (e.g. the inner products, etc.) on 2D data array, and is updatable in real time. Since the adaptive algorithms have usually reasonable amount of built-in tolerance on the accuracy of the processor, the analog nature of the optical processor should not cause major concern. The controls, thresholding, and the memory requirements of the hybrid processor are provided by a microcomputer(IBM - PC-AT with a video board memory). Such a hybrid processor is capable of performing adaptive iteration in quasi-real time. This hybrid architecture will be compared during the conference with other architectures discussed in the literature [14,15,16].

V. Conclusions and Discussions

We studied some of the adaptive processing algorithms for pattern recognition and image restoration, which in terms of neural network systems can be seen as adaptive *hetero-associative* and *auto-associative mappings*, respectively. The hybrid optical electronic processor under study is capable of implementing adaptive pattern recognition and image restoration algorithms operating on large size images. In the next paper, we shall compare other digital image processing algorithms with the corresponding ones in artificial neural network system, e.g. background suppression with *attentive associative memory*.

REFERENCES

1   K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1972, Chapter 7.

2   T. Kohonen, *Self-Organization and Associative Memory*, Springer-Verlag, 1983, Chapter 7.

3   Z-H Gu, J. R. Leger and S. H. Lee, "Optical Implementation of the Least-Square Linear Mapping Technique for Image Classification." *J. Opt. Soc. Am*, Vol. 72, No. 6, June 1982, pp. 787-793.

4   J. R. Leger and S. H. Lee, "Image Classification by An Optical Implementation of the Fukunaga-Koontz Transformation." *J. Opt. Soc. Am.*, Vol. 72, No. 5, May 1982, pp. 556 - 564.

5   A. Papoulis, "A New Algorithm in Spectral Analysis and Band-Limited Extrapolation," *IEEE Trans. Circuit and System*, Vol. CAS-22, No. 9, September 1976, pp. 735 - 742.

6. J. A. Cadzow, "An Extrapolation Procedure for Band-Limited Signal," *IEEE Trans. Acou. Speech and Sig. Proc.* Vol. ASSP-27, No. 1, February 1979, pp. 4 - 12.

7. A. V. Oppenheim, M. H. Hayes, J. S. Lim, "Iterative Procedures for Signal Reconstruction from Phase," *SPIE* Vol. 231, 1980, pp. 121 - 129.

8. D. Youla, "Generalized Image Restoration by the Method of Alterating Orhtogonal Projections," *IEEE Trans. Circuit and System.* Vol. CAS-25, No. 9, September 1978, pp. 694-702.

9. R. E. Kalman, *Fundamental Study of Adaptive Control System,* ASTIA AD S 28-73, 1962.

10. W. E. Smith, H. Barrett and R. G. Parman, "Reconstruction of Objects from Coded Image by Simulated Annealing," *Optics Letters,* Vol. 8, No. 4, April 1983, pp. 199 - 201.

11. J. R. Fienup, "Phase retrival Algorithms: A Comparison," *Applied Optics,* Vol. 21, August 1982, pp.2758 - 2769.

12. M. Cohen and S. Grossberg, "Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks," *IEEE Trans. Sys. Man and Cyber.,* Vol. SMC-13, No. 5, September October 1983, pp. 815 - 826.

13. N. H. Farhat, D. Psaltis, A. Prata and E. Pack, "Optical implementation of the Hophed Model," *Applied Optics,* Vol. 24, No. 10, May 1985, pp. 1469 - 1475.

14. D. Psaltis and N. Farhat, "Optical implementation processing based on an associative memory model of neural nets with thresholding and feedback," *Optics Letters,* Vol. 10, No. 2, February 1985, pp. 98 - 100.

15. B. H. Soffer, G. J. Dunning, Y. Owechko and E. Marom, "Associative holographic memory with feedback using phase conjugate mirrors," *Optics Letters,* Vol. 11, February 1986, pp. 118 - 120.

16. A. D. Fisher and C. L. Giles, "Optical adaptive associative computer architecture," *Proc. IEEE COMPCON* Spring 1985, pp. 342 - 344.
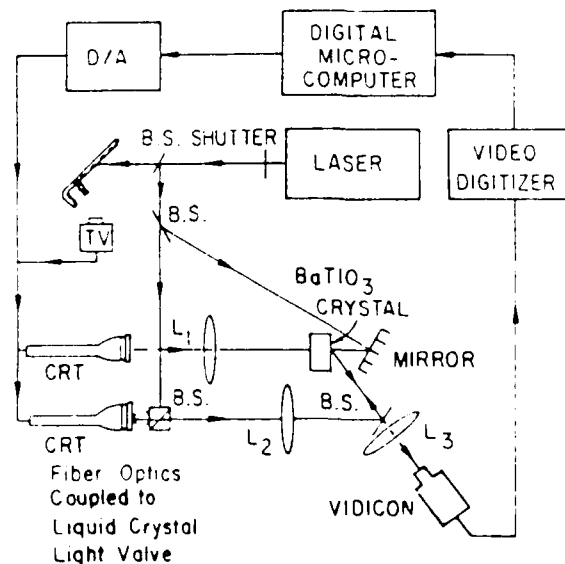
Figure 1. A schematic of a hybrid optical electronic processor for adaptive pattern recognition and iterative image restoration.

TUESDAY, MARCH 17, 1987

PROSPECTOR / RUBICON ROOM

10:30 AM–12:00 M

TuB1–5

SESSION 7

John N. Lee, Naval Research Laboratory, *Presider*

Analog Complexity Theory †

Kenneth Steiglitz
Dept. of Computer Science
Princeton University
Princeton NJ 08544

Summary

Digital computing algorithms are analyzed using a simple and abstract model for computation: the Turing Machine, or close equivalents. The simplicity of the model makes it possible to measure complexity in terms of only two resources -- time and space, and allows us to use asymptotics without concern for noise or the breakdown of physical laws. Analyzing the complexity of analog computation is more difficult because of the modeling problem, and the theory and technique are at an earlier stage of development. In this talk we will discuss this theory, and give some examples of its application. Much of the discussion is based on [1].

We will begin by discussing the definitions of *digital* and *analog* systems, by no means a trivial issue. We will argue that the major distinction between the two stems from the fact that a digital computer can use any number of physical quantities (registers) to represent a problem variable, while an analog computer can use only a fixed number.

Next we take up the important differences between measuring the complexity of analog and digital computation. While the Turing Machine is taken as a valid model for any digital computation, and many other discrete models have been shown to be equivalent to it, there is an endless variety of essentially different models for analog systems. This creates an important and interesting difficulty: A particular analog model is usually valid over only a limited range of problem sizes, and therefore asymptotic results may be meaningless.

Noise is an important limiting factor in the performance of analog systems, while it is modeled away in digital systems. The arbitrary precision of digital computation is realized by the use of indefinite storage for one variable, its distinguishing characteristic. In certain cases it is possible to trade time for precision in analog computation, in effect re-using the analog variables, and creating a hybrid. This idea goes back to Lord Kelvin, and is discussed in more detail in [2].

Finally, we address briefly the open question of whether analog computers are in any sense more general than digital. The central issue here revolves around a stronger than usual version of Church's thesis.

1. A. Vergis, K. Steiglitz, B. D. Dickinson, "The Complexity of Analog Computation," *Mathematics and Computers in Simulation*, vol. 28, pp. 91-113, 1986.

2. H. J. Caulfield, J. H. Gruninger, J. E. Ludman, K. Steiglitz, H. Rabitz, J. Gelfand, E. Tsoni, "Bimodal Optical Computers," *Applied Optics*, vol. 25, no. 18, pp. 3128-3131, Sept. 15, 1986.

A Unified Approach to Analyzing Optical Computing Systems

Ravindra A. Athale, Charles W. Stirk, and Michael W. Haney
The BDM Corporation
Mail Stop D321
7915 Jones Branch Drive
McLean, VA 22102

Many problems exist that available methods of computation, including those implemented by advanced parallel electronic computers, cannot solve within the bounds established by the immutable requirements of numerous significant applications. The bounds for some of these applications are generally described in terms of environmental stresses, input format, desired output, computational throughput, accuracy, size and power. In recent years, the potential capabilities of many optically implementable algorithms, architectures and technologies have been considerably enhanced. What must be done now is to construct composite system performance metrics in terms of individual algorithmic, architectural and technological capabilities that can predict whether or not a given optical system can satisfy the requirements imposed by these applications. (Clearly, a given application can be solved by more than one algorithm, each implemented on several architectures, using a wide variety of technology.) The method used to construct these composite metrics should reveal what individual performance gains must be achieved before an optical system can be applied to a given problem. Thus, a methodology that unifies heretofore unrelated aspects of algorithms, architectures, and technology will serve as a tool to point out fruitful avenues of research to the optical computing community.

While portions of each of these catagories have been considered separately in the past, such an approach has led to performance metrics that are by themselves meaningless. Only with the context originating from the other descriptive structures can meaning be extracted from a performance gain in an isolated area. Moreover, the lack of an explicit formalism to express the interdependence of mathematical formulations, architectural organizations, and hardware realizations in optical computing has led to poor communication between the communities exploring these three aspects of research. In such a vacuum, a fundamental advance can be made at some level without its implications being understood for several years. This type of research environment leads to research that is at best inefficient, and at its worst ineffective.

The descriptive structures that we have chosen can be further broken down into subcategories. For instance, a

problem in linear algebra can usually be solved by more than one algorithm. Similarly, an algorithm such as eigen-value/eigen-vector decompostion can be described as being composed of one of several distinct organized applications of a finite number of lesser elements labeled higher order operations (Figure 1). Each of these higher order operations can be further reduced to a set of elementary operations. Finally all elementary operations can be described in terms of the ordered application of the members of a finite set of computational primitives.

Every algorithm, operation and primitive can be implemented on a variety of different optical architectures. Any architecture can be described as having some of the specific characteristics listed in figure 2. The value that each of these characteristic parameters assumes along with the context provided by the application, algorithm and technology offers the means by which different architectures can be compared. Similarly, a given architecture can be realized by potentially many different technologies (Figure 3). These technologies can be organized in a similar manner and their effects on specific architectures and algorithms can be quantified.

More specifically, since each algorithm, architecture, and technology can be specified in terms of performance metrics that are determined by its intrinsic properties and the properties of its constituents, the overall performance of a given optical system for an application will be a function of these individual performance metrics (Figure 4). Furthermore, the constraints imposed by an application will limit the available algorithms, architectures and technologies. Direct limitations are imposed by application requirements directly on each of the aspects, while indirect limits are propagated through the composite performance metrics from other aspects.

In addition, by using this formal construct, the global effect of an advance in research at an isolated location can be immediately quantified since the performance of one aspect of an optical computing system is determined by the performance of its constituent aspects in all the descriptive spaces. Moreover, the advantages offered by a new combination of constituent elements can be rapidly determined. Hence, the communication pathways between different research areas pertinent to optical computing can be greatly enhanced and new directions of research can be revealed that may provide significant performance gains on specific applications.

Examples, such as SAR and pattern recognition, will be examined in detail to illustrate the utility of this formal description. A similar analysis of more complex systems like associative memories will also be attempted.

Figure 1   **ALGORITHMIC VIEWPOINT:**
EXAMPLE

ALGORITHM

EIGEN-VALUE/EIGEN-VECTOR COMPUTATION

HIGH-ORDER OPERATIONS

Q-R FACTORIZATION, GRAM-SCHMIDT, GIVENS ROTATIONS,
HOUSEHOLDER PLANE REFLECTIONS

ELEMENTARY OPERATIONS

SCALER-VECTOR PRODUCT, VECTOR-VECTOR INNER/OUTER PRODUCT,
VECTOR-MATRIX PRODUCT, MATRIX-MATRIX PRODUCT

COMPUTATIONAL PRIMITIVES

MULTIPLICATION, ADDITION/SUBTRACTION, SQUARE-ROOT, DIVISION

Figure 2   **ARCHITECTURAL VIEWPOINT**

- **CONTROL/DATA**
  - HARDWIRED, PROGRAMMABLE, FEED-FORWARD, FEED BACK
- **PARALLELISM**
  - N^0, N^1, N^2, N^3, N^4
- **INTERCONNECTS**
  - 1-1, 1-M, M-1, M-M
  - SPACE-INVARIANT, SPACE-VARIANT
- **CLOCKING**
  - SYNCHRONOUS, ASYNCHRONOUS
- **MULTIPLEXING**
  - SPACE, SPATIAL FREQUENCY, TIME, TEMPORAL FREQUENCY, POLARIZATION, COLOR, ...
- **INTEGRATION**
  - SPATIAL, TEMPORAL

  •
  •
  •

Figure 3   **TECHNOLOGICAL VIEWPOINT**

- ACTIVE, PASSIVE

- LINEAR, NONLINEAR

- BULK OPTICS, GUIDED-WAVE OPTICS

- ACOUSTO OPTICAL, ELECTRO OPTICAL, MECHANICAL OPTICAL, NONLINEAR OPTICAL

- INTERFACES (OPTICAL-OPTICAL, ELECTRONIC-OPTICAL, OPTICAL-ELECTRONIC)

Figure 4   **EXAMPLE FOR EXERCISING
THE FRAMEWORK**

ALGORITHMIC	ARCHITECTURAL	TECHNOLOGICAL
•	CONTROL/DATA	ACTIVE, PASSIVE
•	HANDWIRED, FEED-FORWARD	
		LINEAR, NONLINEAR
	PARALLELISM	
ELEMENTARY OPERATIONS	N^0, N^1, N^2	BULK OPTICS, GUIDED
VECTOR-MATRIX PRODUCT		WAVE OPTICS
	INTERCONNECTS	
	1-1, 1-M, M-1	A-O, E-O, MECH.-O, NLO
•	CLOCKING	INTERFACES
	SYNCHRONOUS, ASYNCHRONOUS	
	MULTIPLEXING	
	SPACE, SPATIAL FREQUENCY, TIME	
	INTEGRATION	
	SPATIAL, TEMPORAL	

RULE-BASED, PROBABILISTIC, SYMBOLIC TARGET CLASSIFICATION BY OBJECT SEGMENTATION

David Casasent and Abhijit Mahalanobis
Carnegie Mellon University
Department of Electrical and Computer Engineering
Pittsburgh, PA 15213

1. INTRODUCTION

Optical symbolic processing applications in pattern recognition rather than logic operations [1,2] are considered in this paper. The database we employ is summarized in Section 2. Optical correlators represent one of the most powerful functions possible and preferable for realization on optical systems. We thus retain this architecture as the fundamental level-one symbolic processor to be used [2,3]. We utilize the attractive aspects of distortion-invariant iconic optical matched spatial filter (MSF) filters [4] in this work. We increase the flexibility, capacity and performance of such filters by using segments of the input object as separate filters (Section 3). The correlation outputs for these object sections represent the symbolic description of the input to be processed. A hierarchical set of rules for symbolic output processing and substitution is then employed (Section 4). The symbolic substitution used, the interactive expert system nature of the rule design, and the confidence of each rule are then detailed (Section 4). Tests of the system are then presented (Section 5).

2. DATABASE

The database used consisted of ATR tank and APC objects with 32 x 32 pixel resolution. For each object, 36 views at 10° increments in aspect from a fixed 30° depression angle are available. Earlier work by us discussed the use of a model-based 3-D target description [5]. Such ATR objects are more complex and more different between classes than are the aircraft objects initially used. Thus, we centered each object in the database and segmented each image into a 4 x 4 grid of 16 regions or sectors. Filters for each of these object sections were formed and the resultant 16-dimension output vector is viewed as the symbolic description of the object. Table 1 lists the parameters and values used for each input to our synthesis procedure (Sections 3 and 4) and our tests (Section 5).

3. SYMBOLIC OBJECT DESCRIPTION

A MACE (minimum average correlation energy) iconic filter is formed for each of the $M = 16$ object sectors from training set images (12 images per class, 2 classes, $N = 24$ total images). Thus, $N = 24$ sub-images per sector are used to form $M = 16$ filters $\underline{f}_i$. We form $s = 3$ sets of these filters with different ideal output symbolic filters $\underline{y}$ per set s and object class c. Figure 1 shows the $16 = 4 \times 4$ output symbolic vectors chosen for the class 1 object (a T62 tank). The output symbolic vectors for class 2 objects (the APC model M113) are the complement of the corresponding class 1 vectors. These outputs are the values of a set of 16 multiplexed space or frequency-multiplexed MSFs or linear discriminant functions (LDFs) used in a multiplexed optical correlator.

4. SYMBOLIC, RULE-BASED, EXPERT SYSTEMS, AND CONFIDENCE

We used 12 of the 36 images per class ($N = 24$) to design $s = 3$ sets of $M = 16$ filters each, given by $\underline{f}$. The outputs expected for the 3 filters for class 1 objects are given in Figure 1a.

TABLE 1: Parameters and Values Used

PARAMETER	VALUE	REMARKS
dxd	d=32	Input image resolution
M,m	M=16=4x4	Number of object sectors or symbols
kxk	k=8	Resolution per image symbol sector
N,n	n=72 (2 classes)	Number of images (total)
f_m	m=16 filters	Symbolic filter for sector m
v	m=16 element vector	Output symbolic vector
g_{mn}	Each of n training images has m sectors	Sector m of training image n
o_m	Sum of g_{mn} over all n for one m	Training set for sector filter m
$\underline{v}_1, \underline{v}_2$	2 output 16-element vectors	Output symbolic vector for class 1, 2
$\underline{v}^{(s)}$	s=3 filters used 16 element output vector	Output symbolic vector for filter set s of sector filters
$\underline{v}_c^{(s)}$	$\underline{v}_1^{(s)} = \overline{\underline{v}}_2^{(s)}$ c=2=number classes	Output for filter s for input in class c
$\underline{f}_m^{(s)}$	s=1 to 3 filter sets m=16 filters/set	Symbolic filter set s

(a) Output symbol vector $\underline{v}_1^{(1)}$

(b) Output symbol vector $\underline{v}_1^{(2)}$

(c) Output symbol vector $\underline{v}_1^{(3)}$

FIGURE 1: Output symbolic vectors $\underline{v}_c^{(s)}$ for filter set s for object class c = 1.

4.1 SYMBOLIC DESCRIPTIONS

Each symbolic output $\underline{v}^{(s)}$ is thresholded to yield "1" or "0" elements. Rules are then applied to it. Rule 1 is now summarized. In this rule, we assign symbol A to all one-valued output symbols and B to all zero-valued output symbols. We then compare the 16 measured output symbols to the $\underline{v}_1^{(s)}$ patterns. If all 16 symbols match in all patterns, we declare the input object to be a T62. If this rule is not successful, we assign symbol A to all zero-valued outputs and B to all one-valued outputs and apply the same matching algorithm. The use of the expected output and its complement is necessary in the optical realization of this step. The complement rule and the true rule are thus tested together to achieve matching [2,6]. We describe this processor for a symbolic two-state machine. Obvious extensions allow the use of multi-levels or several elements per output bit. Standard rules are used for the choice of the symbolic outputs in Figure 1.

4.2 HIERARCHICAL RULES AND EXPERT SYSTEMS

If rule 1 tests fail, rule two is invoked. This rule does not look at approximately 6 of the 16 x 3 output symbols from the 3 filter sets. Failing rule 2, rule three (which omits approximately 7 of the 48 output symbols) is used, then rule 4 (which omits 13 output symbols), and finally rule 5 (which omits over 20 output symbols). Each rule has true and complement parts with different values for A and B symbols used. The supervising expert determines the number of rules to be used and the confidence of each. This choice can be guided by the confidence or probability of each rule as we now detail.

4.3 INCREASED KNOWLEDGE AND CONFIDENCE LEVELS

We define rule 2 and subsequent rules by testing the rule 1 system (just as a person learns). When rule 1 was applied to the outputs from all 36 tank images, we found the number of errors obtained and which sectors or output symbols were generally in error. These are the digits removed in rule 2 (different symbol output digits are removed from the outputs of the 3 different filter sets). For tank and APC images, if rule 1 is passed, the confidence of the output decision is 1.0. For tank (APC) inputs to rule 2, the confidence is 0.86 (0.80). Different confidences are expected for the APC data, since the symbols omitted were chosen from tank data inputs only. In practice, the experts should state the confidence of each rule and utilize this with the above probabilities to determine the confidence in the class estimate rather than the conifdence of satisfying the rule.

4.4 SYMBOLIC SUBSTITUTION

The use of symbols A and B allows the same logic system/algorithm to be used for comparisons. We use the information in one output vector (the elements in error) to rectify errors in other output vectors. This symbolic-substitution rule module reverses the symbols for those elements expected to be in error and then checks for a match. Our rule-based recognition system allows us to identify elements that may be in error.

4.5 ASSOCIATIVE MEMORY PROCESSOR

If no rules perform well with high confidence (see Section 5), then the input data is fed to an associative memory, several elements of the input vector are corrected by this processor, and the new input is fed to the symbolic processor. This has been successfully demonstrated as our data in Section 5 will show.

5. TEST RESULTS

The results of tests on 10 images in 2 classes at different aspect views for the case when two sectors of each image were dead (zero outputs) were fed to a five-rule symbolic processor described above. The class estimates, the rule satisfied and the confidence of the rule satisfied are given in Table 2. For the low confidence output (test 4), the class estimate was wrong. For this, we used an associative memory followed by the symbolic processor as described in Section 4.5. This hierarchical system gave the final correct object recognition and classification. Table 2 shows the data results obtained.

These initial results are most attractive. The system described includes many facets of advanced AI: expert use, a rule-based system, use of probabilistic processor information, symbolic pattern recognition and symbolic substitution as well as classification techniques, plus an associative memory element and a symbolic substitution rule module.

Advanced obvious extensions to this system include: selecting symbols to omit from tests on both object classes, the use of joint probabilities, attention to the more reliable object sectors in each case study, and class confidence versus confidence of satisfying the rules.

TABLE 2: Initial Test Results

TEST NUMBER	ROTATION (DEGREES)	ACTUAL CLASS	DETERMINED CLASS	RULE NUMBER	CONFIDENCE
1	0	tank	tank	2	0.86
2	20	tank	tank	2	0.86
3	50	tank	tank	2	0.86
4	90	tank	APC	4	0.62
5	110	tank	tank	3	0.76
6	0	APC	APC	3	0.77
7	20	APC	APC	3	0.77
8	50	APC	APC	2	0.80
9	90	APC	APC	3	0.77
10	110	APC	APC	3	0.77

ACKNOWLEDGMENTS

REFERENCES

1. K.H. Brenner and A. Huang, "An Optical Processor Based on Symbolic Substitution", *Topical Meeting on Optical Computing*, Optical Society of America, pp. WA4.1-WA4.3, March 1985.

2. C. Warde and J. Kottas, "Hybrid Optical Inference Machines: Design Considerations", Applied Optics, Vol. 25, pp. 940-947, March 1986.

3. D. Casasent, "Optical AI Symbolic Correlators: Architecture and Filter Considerations", Proc. SPIE, Vol. 625, pp. 220-225, January 1986.

4. A. Mahalanobis and D. Casasent, "Large Class Iconic Pattern Recognition: An OCR Case Study", Proc. SPIE, Vol. 726, October 1986.

5. D. Casasent and S.A. Liebowitz, "Model-Based Knowledge-Based Optical Processors", Applied Optics, Vol. 26, *Special Issue on Artificial Intelligence and Non-Numerical Processing*, submitted (publication January 1987).

6. D. Casasent and E. Botha, "Knowledge in Optical Symbolic Pattern Recognition Processors", Optical Engineering, *Special Issue on Optical Computing*, Accepted (publication January 1987).

REAL-TIME ACOUSTO-OPTIC SPOT-LIGHT MODE SAR PROCESSOR

Michael Haney
The BDM Corporation
7915 Jones Branch Drive
McLean, Virginia, 22102

Demetri Psaltis
California Institute of Technology
Department of Electrical Engineering
Pasadena, California, 91125

INTRODUCTION

High-resolution imaging with the Synthetic Aperture Radar (SAR) technique is widely recognized as the most successful application of optical information processing to date. Optical signal processing (OSP) techniques have been applied to the collection and processing of SAR data since the introduction of the technique over thirty years ago. Despite this success the use of optical techniques for real-time SAR applications has generally been precluded by the need for chemical processing of the film on which the radar data are recorded. Consequently, and in concert with the dramatic successes in integrated circuit technology, most existing real-time SAR processors are based on electronic signal processing techniques. Recent advances in optical transducer technology, however, have given rise to renewed interest in real-time optical SAR. Specifically, a time-and-space integrating (TSI) architecture has been developed that uses acousto-optic (AO) Bragg cells and charge-coupled device (CCD) detector arrays to generate SAR images at real-time rates [1]. This architecture compares favorably with the all-electronic approaches in the areas of speed, size, power consumption, and EMI/EMP immunity. Recent developments in the TSI architecture have incorporated electronic programmability and flexibility to expand the realm of practical application of the approach [2,3].

In the development of the TSI architecture the emphasis thusfar has been on its application to strip-map mode SAR, in which an arbitrarilly long swath of the ground is imaged in a scrolled manner by generating a rastered sequence of lines in the image as the radar flys by the target area. In this paper the application of the TSI architecture is extended to spot-light mode SAR, in which data are collected only from a specific section of the target scene and the resulting image is generated in a framed mode.

SAR DATA COLLECTION GEOMETRIES

The typical SAR data collection geometry is depicted in Figure 1. The radar platform moves parallel to the ground at a constant velocity. The radar beam illuminates a portion of the target scene to one side of the flight path. A periodic pulse train is transmitted and the radar echoes associated with the pulse train are received by the radar at
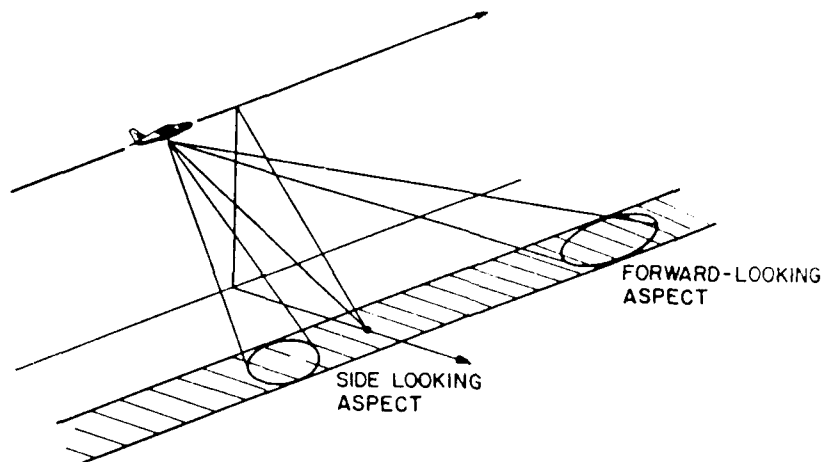
Figure 1. SAR Data Collection Geometry

different locations on the trajectory (along a "synthesized" aperture). The received signals are stored and processed, using correlation techniques, to generate an image. The coordinate axes of this image are the range and azimuth positions of the point scatterers referenced to the radar's location at a particular point in time.

In strip-map SAR the antenna is fixed to the body of the platform such that a long swath of the target scene is scanned as the platform moves. In spot-light mode SAR the antenna is steered to maintain illumination on a specific region. Higher theoretical azimuth is therefore achievable because the integration time is not determined by the size of antenna footprint, as in strip-map SAR. However the longer integration times of spot-light mode SAR cause the effects of range migration to be more pronounced, thereby imposing more requirements on the real-time processor.

In Figure 1 a further characterization is made regarding the radar/target aspect. In the side-looking aspect the signal processing requirements are simpler because the scene being imaged is located directly abeam of the radar causing the average range to the target to be approximately zero during the data collection period. In the forward looking aspect, however, the range decreases monotonically during the data collection. The signal processor must compensate for this range walk effect to produce sharp images.

SPOT-LIGHT MODE ARCHITECTURE DESCRIPTION
The reader is referred to previous publications [1-3] for a detailed description of the real-time TSI SAR architecture. The operation of the processor is summarized here to illustrate the new features that apply to a spot-light mode implementation. With reasonable approximations the SAR signal processing problem is: linear, shift-invariant except that the azimuth integration variable is scaled by the range, and separable in the two variables of integration. Furthermore the two dimensional (2-D)
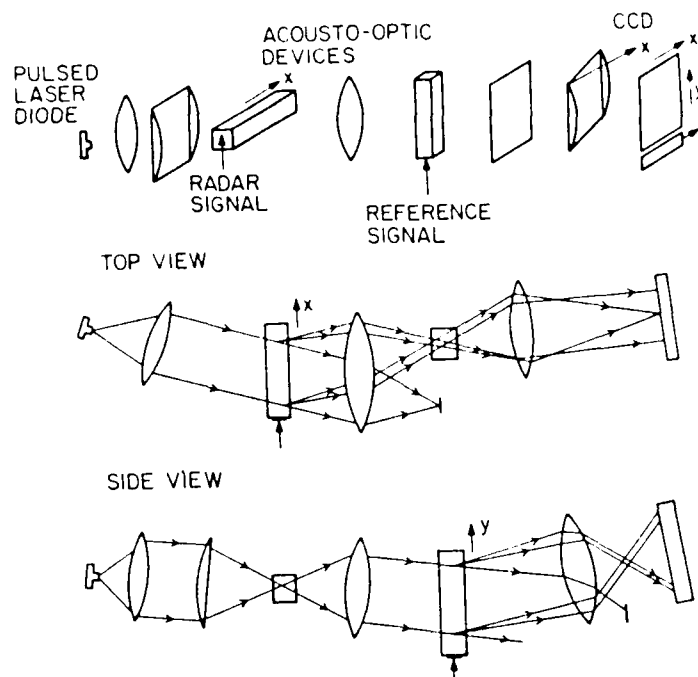
Figure 2. Programmable Real-time SAR Processor

unfocused data in SAR is received in a one dimensional (1-D) format. These features permit the 2-D processing problem to be decomposed into a cascade of two 1-D integrations that can be implemented in real-time with 1-D optical transducers as shown in the schematic diagram of the programmable acousto-optic TSI architecture in Figure 2. The principal elements are a laser diode, two orthogonally oriented AO cells, and a CCD detector array. The top view of the processor depicts the range compression operation for one return pulse from a single scatterer. The radar return signal is summed to a sinusoidal reference and applied to the first AO cell. The laser diode, pulsed in synchronism with the radar, illuminates the first AO cell and freezes the moving diffraction patterns to perform range focusing by a spatial integration of light (as indicated by the focusing rays at the output plane). The reference signal diffracts a collimated beam that mixes interferometrically with the range focused beam at the output plane to detect its phase. The resulting intensity is modulated by the real part of the phase function of the radar return. All of the light rays diffracted by the first AO cell are directed by lenses to pass through the second AO cell as well. Azimuth compression is performed with a temporal integration of light over the sequence of radar returns. This is accomplished by correlating the detected phase function with the known phase history for the given gemometry. This known phase history is stored in electronic memory and loaded into the processor via the second AO cell. The azimuth focusing is performed by incrementally shifting the azimuth reference function in the AO cell to perform the shift and sum operation of a correlation. If the stored phase function

matches the received phase function then the output is an autocorrelation with a peak corresponding to the azimuth location of the scatterer. The resulting charge pattern that builds up on the CCD thus corresponds to the range and azimuth focused image of the point scatterer.

In the spotlight mode implementation the TSI processor must simultaneously compensate for range migration, range/azimuth coupling, and dynamic changes in the data collection geometry. The electronic programmability of the architecture makes this possible. Range walk is compensated by electronically adjusting the timing of the laser pulse to keep the range focused data within the same range bin throughout the integration period. The gross doppler shift due to range walk is also compensated electronically with the aid of a programmable frequency synthesizer that can generate the appropriate compensating frequency to mix with the radar return. To compensate for range/azimuth coupling a cylindrical lens can be tilted [2] as shown in Figure 2. However this can also be accomplished by modulating the reference function in the first AO cell by a suitable phase function that is derived from the parameters of the radar/target geometry. This programmable solution to range/azimuth coupling gives the processor the flexibilty to adjust for changes in geometry without having to change the tilt of a lens, which may be impractical. When long integration periods are used to achieve high azimuth resolution, range curvature (which is a component of range migration that is approximately quadratic in time) becomes significant and must be dealt with. The TSI architecture offers a unique approach to range curvature correction in which the output CCD array is rotated about the optical axis during the integration period, through a small angle, at a constant rate. The rotation rate is determined by the parameters of the radar/target geometry and the optics of the processor. An interesting feature of this technique is that if the optical magnfications in the range and azimuth directions are such that the range-to-azimuth scale factor on the CCD array is 1:1 then the rotation of the CCD during the intetgration period simultaneously corrects for range/azimuth coupling and range curvature, and no electronic compensation for range/azimuth coupling is necessary. Furthermore, the rotation rate of the CCD equals the rotation rate of the radar antenna in the spot-light mode. This feature reveals the strength of the match between the real-time SAR processing problem and the TSI architecture. The real-time TSI spot-light mode imager is an excellent example of a hybrid optical/electronic signal processor that effectively combines the best features of both worlds.

REFERENCES

[1] Psaltis and Wagner, Opt. Eng., Vol. 21, No. 5, p. 822.
[2] Psaltis, Haney, & Wagner, NASA OIP Conf., 1983, p. 199.
[3] Haney and Psaltis, SPIE, Vol. 545, April 1985.

Processing of Synthetic Aperture Radar Data
Using the PRIMO Optical Outer-Product Processor

Y. Owechko, J. Grinberg, E. Marom, B.H. Soffer
Hughes Research Laboratories
Malibu, California 90265

The processing of SAR data is one of the great success stories of optical data processing. It is a problem which readily lends itself to the parallelism of optics. Optical solutions have, however, lacked two desireable characteristics: small size and real-time operation. Recently, novel optical architectures which overcome these drawbacks have been developed by Psaltis and Tanguay. (Some of which are reviewed in (1).) In this paper, we describe a new compact lensless real-time optical processor of SAR data which is unique in that it utilizes incoherent light and readily available components and materials.

The form of the received SAR return from a point target after demodulation, assuming a linearly chirped radar pulse and no range curvature, is given by the expression (1):

$$E(t',nT) = A(t'-2r_1/c)\cos[(a/2)(t'-2r_1/c)^2 + 2\pi(VnT-x_1)^2/\lambda r_1]$$

where T is the pulse repetition period, λ is the radar wavelength, x_1 is the azimuth of the target, r_1 is the slant range of the target, a is the radar chirp rate, V is the velocity of the SAR platform, c is the speed of light, and n is an integer. The radar emits a pulse with fixed phase every T seconds. The inter-pulse delay time, t', is measured with respect to the time of origin, nT, of the n-th pulse. The total elapsed time is given by $t=nT+t'$. A plot of $E(t',nT)$ as a two-dimensional function of t' and nT (which is the natural format for unprocessed SAR data) shows that the SAR radar return from a point target is in general chirped in both azimuth and range, resulting in a deformed Fresnel zone plate distribution. If one neglects the effects of Doppler offset due to the Earth's rotation or motion of targets on the ground, range walk, and range curvature (the nature of these effects and methods for compensating them will be discussed below), then the return will be a simple elliptical zone plate. This approximation is normally a good one for SAR's in aircraft.

It is clear from the above equation that the azimuth correlation required to reconstruct the point target is range-dependent, necessitating a space-variant correlation operation. The PRIMO optical processor can implement this space-variant correlation by utilizing a combination of time- and space-integrating architectures. (The acronym PRIMO stands for

Programmable Real-time Incoherent Matrix-multiplier for Optical processing (2)). The architecture of the PRIMO SAR processor is shown in Fig. 1. The return data E(t) modulates the intensity of an LED which illuminates the processor. The processor consists of two crossed one-dimensional edge-addressed electrooptic modulators, each consisting of linear stripe electrodes on a slab of electrooptic material. Polarizers are situated between the layers so that the intensity transmittance of the processor is given by the outer product of the addressing voltages. The transmitted light is incident on a two-dimensional CCD detector which is capable of both in-place and shift-and-add integration of the incident light.

Referring to Fig. 1, the range correlation is performed in time by applying range correlation functions with successive delays to successive electrodes in the top modulator and using in-place integration on the CCD detector. The range correlation peaks are then separated spatially along the range dimension of the CCD detector. The range-variant azimuth correlation is performed spatially by applying the azimuth correlation functions to the bottom modulator and shifting the data in the CCD detector to the right by one data cell every pulse repetition period, T. In the azimuth processing, the space variable is used for correlation and the time variable is used to incorporate the dependence of the azimuth processing on the range of the target. The output of the processor is in a convenient scrolling format, amenable to real-time processing.

The correlation function implemented by PRIMO is formed by crossing two one-dimensional modulators, forming an outer-product matrix. Thus, the PRIMO correlation function (or matched filter in terms of Fourier analysis) is separable. The signal-to-noise ratio (SNR) of the correlation between a separable zone plate and an elliptical zone plate is, however, only slightly less than for the correlation between two elliptical zone plates. The fact that the PRIMO SAR processor is based on outer product multiplication, therefore, only slightly degrades the SNR. This degradation in the SNR due to the separable nature of the PRIMO filter function can, however, be completely eliminated by using two PRIMO processors in tandem, as shown in Fig. 2. By maintaining a 90° phase shift between the two processors, it can be shown that a nonseparable, range-variant SAR correlation function can be implemented.

As discussed in (1), the effects of Doppler offset, range walk, and range curvature become important for spaceborne SAR such as in satellites or the Space Shuttle and need to be compensated. Compensation techniques suggested by Psaltis and discussed in (1) can also be used here. Doppler offset is due to relative motion between the point target and the radar due to the

Earth's rotation or nonstationary point targets. This effect can be simply compensated in PRIMO by adjusting the azimuth correlation function and making it asymmetric. Range curvature is caused by the non-negligible change in range of the point target as it passes through the radar beam. This effect can be compensated by physically rotating the range correlation electrooptic modulator layer relative to the azimuth and detector layers. Range walk is the difference in range of a point target when it leaves the radar beam relative to when it enters the beam. It can be compensated by rotating the detector relative to the range and azimuth electrooptic layers.

A bias-based method for multiplication of bipolar numbers is shown in Fig. 3. By segregating each data cell into positive and negative parts and utilizing the data sequencing shown, it can be shown that the output of the differential amplifier is a bipolar voltage representing the product of two bipolar numbers. Furthermore, the output is devoid of bias levels to the extent that the bias levels do not differ between adjacent cells. Most importantly, however, the output is linear for voltages much less than the electrooptic half-wave voltage. In other words, this method also eliminates the quadratic nonlinearity between the voltages applied to the modulators and the detector output.

1. C. Elachi, T. Bicknell, R.L. Jordan, and C. Wu, Proc. IEEE, Vol. 70, No. 10, Oct. 1982.
2. B. H. Soffer, Y. Owechko, E. Marom, and J. Grinberg, Appl. Opt., Vol. 25, p. 2295, July 15, 1986.

Fig. 1. PRIMO SAR processor
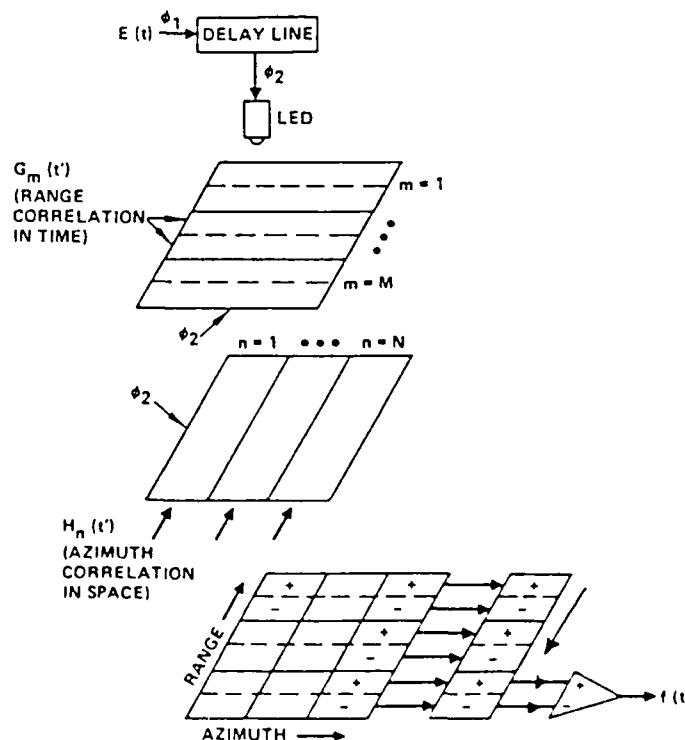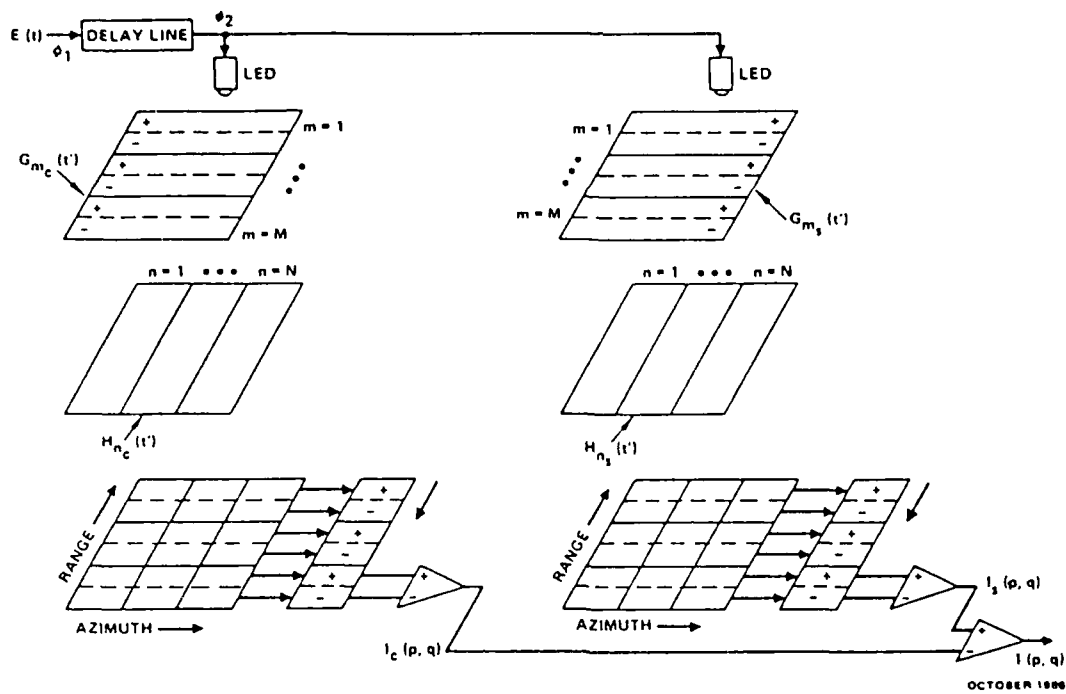
Fig. 2. Dual processor configuration.

Fig. 3. Bias-based
method for bipolar
multiplication.
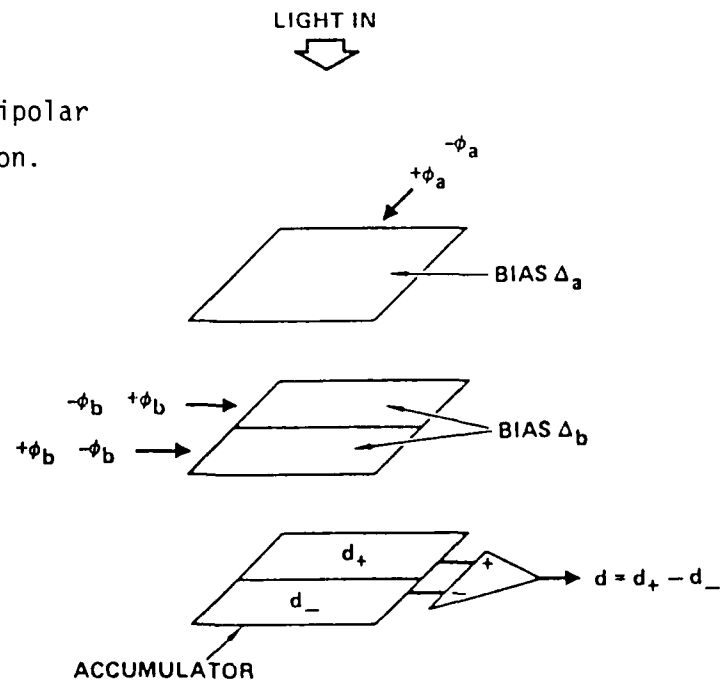
TUESDAY, MARCH 17, 1987

PROSPECTOR / RUBICON ROOM

1:00 PM-2:30 PM

TuC1-5

SESSION 8

Satoshi Ishihara, Electrotechnical Laboratory, Japan, *Presider*

Two-Dimensional Arrays of Semiconductor
Optical Gates for Optical Computing

Hyatt M. Gibbs, Optical Sciences Center, University of Arizona

Tucson, AZ 85721

(602) 621-2941

The emphasis here is to assess the potential role of nonlinear thin-film etalons in optical computing. Nonlinear optics can contribute *decisions* to optical signal processing and computing.[1] The optical nonlinearity makes the device's transmission intensity dependent, so one can obtain the thresholding needed for logic decision making. Nonlinear decision-making devices can be constructed as waveguides in which the light is guided in the plane of the nonlinear thin film or as etalons in which the light is imaged from one nonlinear thin film to the next in such a way that its intensity is highest as it interacts with each film. Guided-wave devices are most likely to find application where data are handled in a pipeline manner, for example, in optical-fiber communication and interconnect systems, data encryption, etc. However, waveguides are much like wires except for their higher bandwidth. Etalons permit massive parallelism and global interconnectivity, i.e., one can perform many operations simultaneously and interconnect in the next plane two or more pixels far apart in the present plane. Consequently we anticipate the use of guided-wave devices in the near term and increased introduction of etalons in the long term.

ZnS interference filters are relatively simple and inexpensive to grow with reasonable uniformity. An optical nonlinearity arises from the shift of the band edge with heating, so it is only weakly resonant. Operation is good at 514.5 nm, which permits the use of many-watt Ar lasers to produce multiple beams. Visible light is also very convenient for

learning to work with many beams in parallel and is more impressive for demonstrations. There are also undesirable features of ZnS filters. The thermal response (0.01 to 1 ms) is much slower than that of GaAs, but the power per pixel ($\cong$ 10 mW) is about the same. Research is being continued to improve uniformity and long-term stability and to reduce the power required per pixel.

GaAs etalons are more likely candidates for commercial implementation. Shift of the etalon peak in a few picoseconds has been demonstrated and interpreted as an ultra-fast NOR gate.[2] This time should also be that for bistability switch on. Recovery of the gate requires removal of the carriers produced by the logic operation. This recovery takes more than 10 ns in the usual GaAs and multiple-quantum-well etalons. Recently recovery as short as 30 ps has been achieved using a thin GaAs etalon with no AlGaAs outside layers, normally used to stop etching of the GaAs substrate as well as to stop surface recombination.[3] Two AND-gate operations were performed, separated by only 70 ps.[3] Thermal considerations may be greater limitations than the recovery time. When many gates operate in parallel, a specific array of gates may not need to be revisited but once each nanosecond or more, allowing cooling. In addition to high speed, GaAs boasts compatibility with electronics: diode lasers can be used as light sources, silicon detectors can be used at the peak of their sensitivity, and electronic circuitry can be constructed using GaAs. The latter is more of an advantage for waveguide integrated systems than for etalon systems.

Without concern for architectural and interconnection problems, one can imagine operating 10^6 spots or pixels on a 5 cm x 5 cm bistable ZnS filter. Assuming 10 mW per pixel and 25% absorption, the heat load would be 100 W/cm² which is challenge enough to remove. One could operate at a 10-kHz rate, resulting in 10^{10} bit operations per second. Of course, 10-kW of laser power would be required.

More promising for implementation are GaAs NOR-gate arrays. Assume 10^4 pixels on 1 cm² requiring 10 pJ per bit operation and operating once every nanosecond. This

requires 100 W of laser power and results in 100 W/cm² heat load. It yields 10^{13} bit operations per second, 10 to 100 times faster than a CRAY. Of course, much work is required to convert the array's capability for many operations in parallel into a programmed or programmable system able to make useful computations, to recognize patterns, or to learn.

For digital optical computing one can imagine a system such as Figure 1. An optical pulse output from one pixel of one etalon is redirected and possibly divided by a holographic lens and is absorbed in the next etalon. Each etalon has input pulses which are gated through or not according to the results of the logic operations. Those input pulses may be gated on or off by a spatial light modulator on their way into the system, providing input communication. In spite of the ring appearance of the schematic, it is not an interferometer; a given pulse never circulates around the ring.

It has been shown that digital computations can be performed by reorganizing simple patterns and replacing them with other simple patterns.[4] A very simple, but complete, symbolic substitution has been achieved using ZnS interference filters operated with a fanout of about 4. The desired pattern is the simultaneous occurrence of bright spots in the lower lefthand and upper righthand corners of an arbitrary 2 x 2 array. When that pattern occurs, the symbol-scription part generates an output pattern consisting of a bright top row and a dark bottom row. This is accomplished by an AND-gate operation of the output of the recognition stage with the strong holding beams. If the desired pattern is not present, the output is completely dark. The addition of two one-bit numbers is underway. These experiments illustrate pattern recognition, cascading, and pattern generation. They require considerable expansion of single-beam techniques: beam division, multiple-beam focusing and reimaging, nonlinear etalon uniformity and stability, etc.
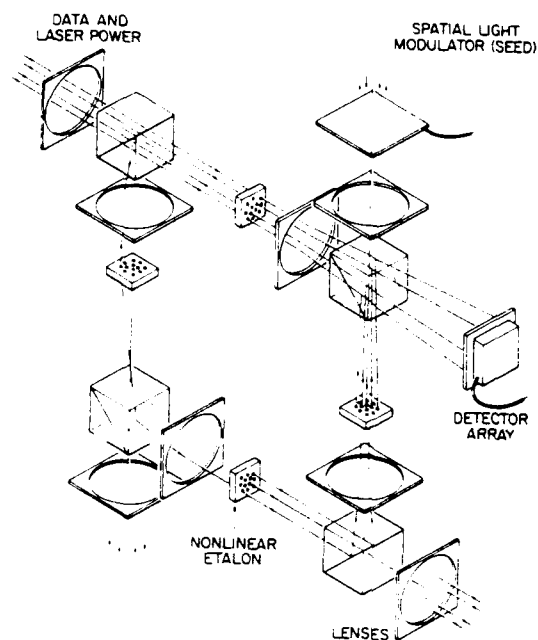
Figure 1. Greatly simplified sketch of "all-optical" computer. The quotes around "all-optical" emphasize the extensive use of electronics in input lasers and spatial light modulators, output detectors, and associated computers.

References

1. Gibbs, H. M., *Optical Bistability: Controlling Light with Light*, (Academic, New York, 1985).

2. Migus, A., A. Antonetti, D. Hulin, A. Mysyrowicz, H. M. Gibbs, N. Peyghambarian, and J. L. Jewell, Appl. Phys. Lett. **46**, 70 (1985).

3. Lee, Y. H., H. M. Gibbs, J. L. Jewell, J. F. Duffy, T. Venkatesan, A. C. Gossard, W. Wiegmann, and J. H. English, Appl. Phys. Lett. **49**, 486 (1986).

4. Brenner, K.-H., A. Huang, and N. Streibl, Appl. Opt. **25**, 3054 (1986).

Restoring Optical Logic: The Demonstration of Extensible All-Optical

Digital Systems

S.D. Smith, F.A.P. Tooley, A.C. Walker, N. Craft and B.S. Wherrett

Dept. of Physics, Heriot-Watt University, Edinburgh EH14 4AS, U.K.

We have recently shown that by the use of a "lock and clock"
architecture and an off-axis configuration of the power and signal beams
indefinitely extensible optical logic is possible.

The demonstration optical circuits which have been constructed use
optical logic gates based on nonlinear interference filters (NLIF). These
NLIF operate over a range of wavelengths extending throughout the visible
to the NIR. This large range is due to the origin of the nonlinearity
which is a thermally-induced change in refractive index. Normally,
operation at a wavelength corresponding to the most powerful Argon ion
laser line, 514.5 nm is convenient. Successful operation of a number of
circuits has been achieved. Figure 1a shows one of the first circuits
which was designed to demonstrate the basic principle of restoring logic.
Figure 1b shows the successful operation of this circuit. The details of
operation will be outlined in this presentation. The operation of an
optical classical finite state machine has also been demonstrated by
expanding the number of information channels in the circuit shown in figure
1a. The circuit shown in figure 2 is a schematic representation of a
circuit constructed which contained three information channels. Six beams
are incident on each optical gate array, a set of three direct from the
laser and overlapping these, on the array only, a set of three beams which
are the reflected output from the previous gate array. Details of
operation of this and similar parallel circuits will be presented.

Further Information

[1] S.D. Smith, J.G.H. Mathew, M.R. Taghizadeh, A.C. Walker, B.S. Wherrett

    and A. Hendry,

    Optics Commun. 51 (5) 357 (1984).

[2] S.D. Smith, I. Janossy, H.A. MacKenzie, J.G.H. Mathew, J.J.E. Reid,

    M.R. Taghizadeh, F.A.P. Tooley and A.C. Walker,

    Opt. Eng. 24 (4) 569 (1985).

[3] S.D. Smith, A.C. Walker, B.S. Wherrett and F.A.P. Tooley

    To be published in Nature 1987.
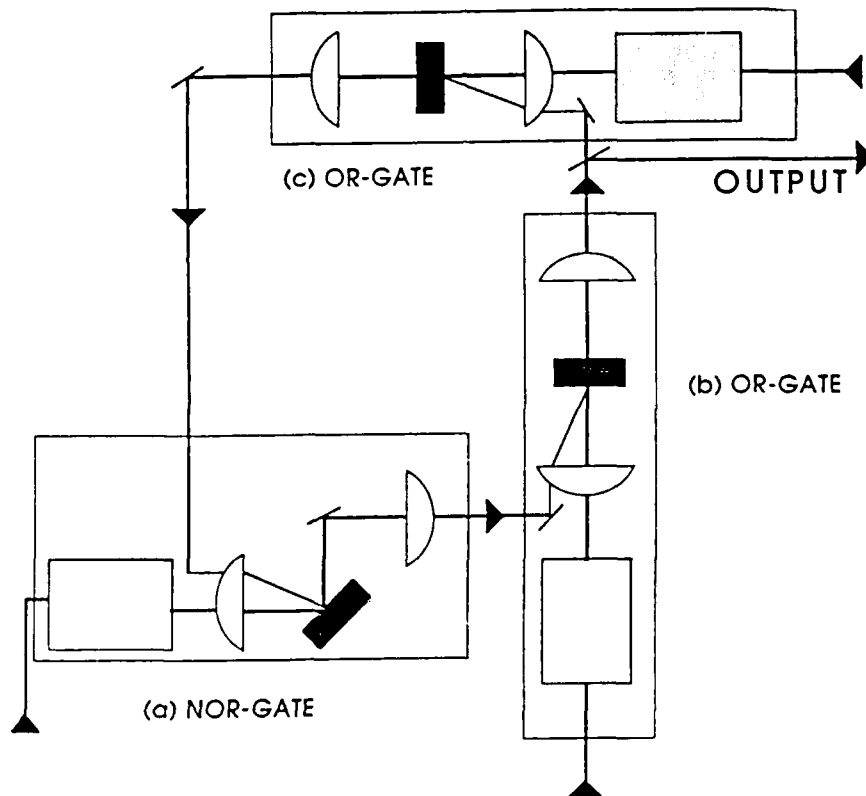
Figure 1(a) Schematic of the optical configuration used in the

demonstration of a looped optical circuit.
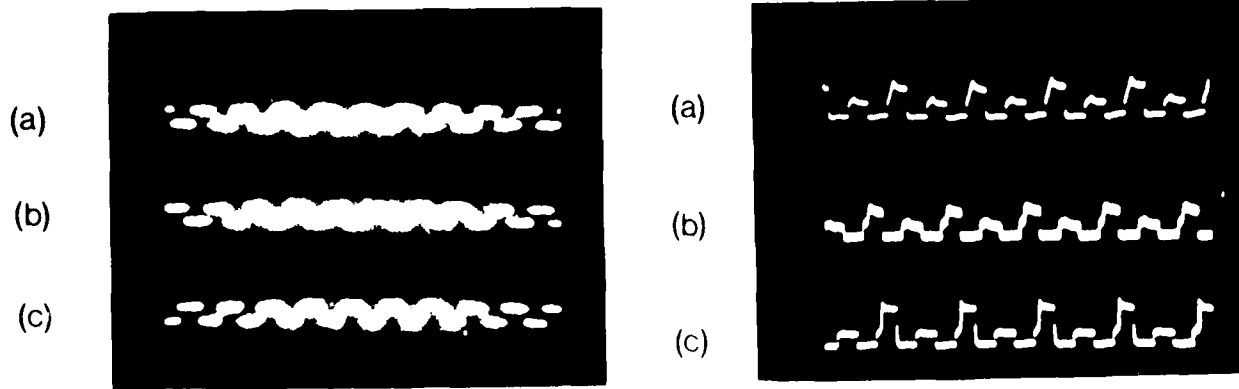
Figure 1(b) Left: Input to the three optical gates of figure 1(a).
Right: Transmitted outputs from the three optical gates.
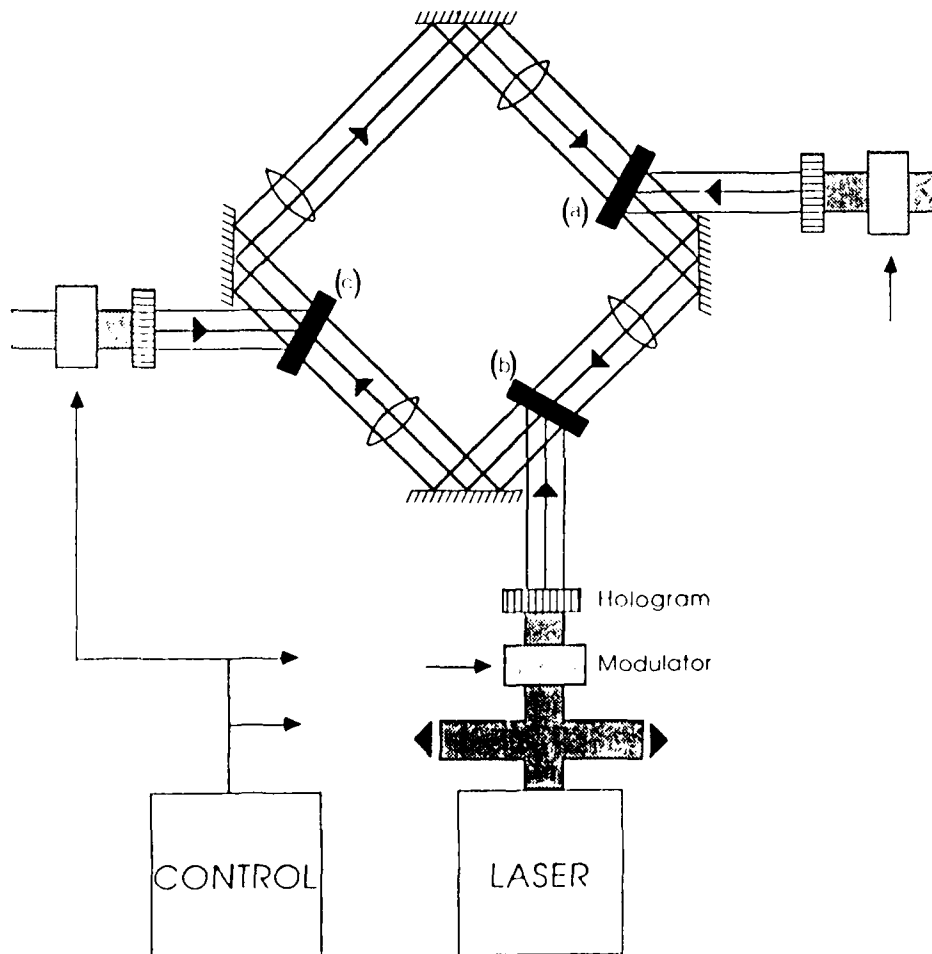
Figure 2. Schematic representation of circuit constructed to simulate an optical classical finite state machine.

A Highly Cascadable Optically Bistable Device for Large Fan-out

Optical Computing Applications

N.C. Craft and S.D. Smith

Dept. of Physics, Heriot-Watt University, Edinburgh EH14 4AS, U.K.

Introduction

Smith et al have shown (1) that in order to ensure that the change in output from one bistable device is sufficient to switch the succeeding one, the technique know as 'hold and switch' is desirable. This involves holding a device as close as possible to its switch point with one laser beam, and using the change in output from a previous similar device to induce switching, thus providing fully restoring logic. Due to the fact that the holding power cannot be made arbitrarily close to the switching power, and that a substantial "over switch" is required to avoid the effects of critical slowing down, there exists a fundamental limit upon the cascadability and fan out potential of conventional bistable devices. This is because their change in output power can only be less than or equal to their switching power. A better device would be one in which the change in output power could be significantly larger than the switching power. The twin cavity device (TCD) described here will be shown to possess this quality.

Principle of Operation

The TCD consists of two thermally nonlinear bistable etalons, seprated by a thin, heat conducting, optically opaque layer as shown in figure 1. The devices currently under investigation employ zinc selenide and zinc sulphide interference filters (2) separated by a metallic layer, constructed by thermal evaporation.

By selecting particular values for parameters such as thickness, absorption coefficient and detuning, the two bistable etalons are made to

175

have significantly different switching powers (3). Both etalons are illuminated with separate holding beams so that each is held just below its switching point. The lower power cavity is used as the input side of the device: because this has a low holding power, it can be caused to switch on by a small "signal", or change in input intensity. The temperature rise in the input cavity due to its switching onto resonance is transferred via the conducting layer to the high power cavity. The associated change in detuning causes this output cavity to move onto resonance, thus causing a reduction in the power reflected from the output side of the device. Now because the power in the holding beam of the output cavity is much greater than the total power incident on the input side needed to induce switching, it is possible for the change in output power as the device switches to be significantly greater than the switching power of the device. The limited cascadability of conventional, single cavity bistable etalons is thus avoided.

Figure 2 shows a typical theoretical TCD characteristic obtained from a one dimensional computer model, further theoretical work to optimise TCD characteristics is underway, and prototype devices have been constructed and used to demonstrate high changes in output power.

Conclusions

It has been shown experimentally and computationally that the limitations of conventional bistable devices for optical circuitry applications can be avoided by the use of two beam, twin cavity devices. It should be possible to use these devices to demonstrate large fan out and rapid switching of cascaded devices.

References

[1] S.D. Smith, A.C. Walker, B.S. Wherrett, F.A.P. Tooley, N. Craft,
    J.G.H. Mathew, M.R. Taghizadeh, J. Redmond, R.J. Campbell,
    "Restoring Optical Logic: The Demonstration of Extensible All-Optical
    Digital Systems" Opt. Eng. In press.

[2] I. Janossy, M.R. Taghizadeh, J.G.H. Mathew and S.D. Smith,
    "Thermally Induced Optical Bistability in Thin Film Devices",
    IEEE J. of Quant. Elect. Special Issue on Optical Bistability
    Vol. QE-21 No. 9 (Sept. 85).

[3] B.S. Wherrett, D. Hutchings and D. Russell,
    "Otically Bistable Interference filters - Optimisation Considerations"
    J. Opt. Soc. Am. B Vol. 3 p. 351 (1986).

FIG 1 : COUPLED CAVITY BISTABLE DEVICE

THERMALLY CONDUCTIVE
OPAQUE LAYER

DICHROIC BEAM
SPLITTER

1/4 PLATE

INPUT

HOLDING
BEAM

INPUT
CAVITY

OUTPUT
CAVITY

REFLECTED
OUTPUT
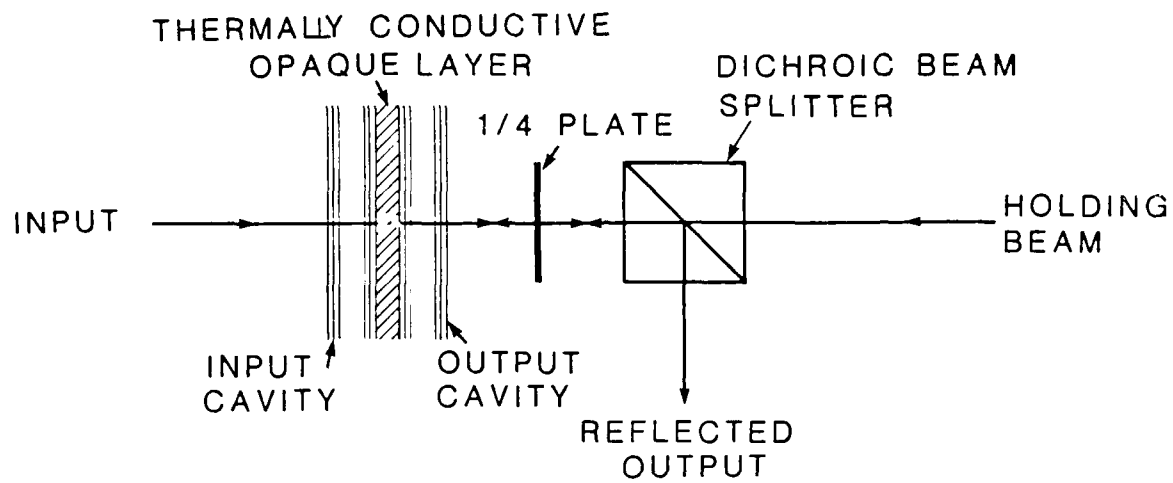
FIG. 2 : TYPICAL DEVICE CHARACTERISTICS
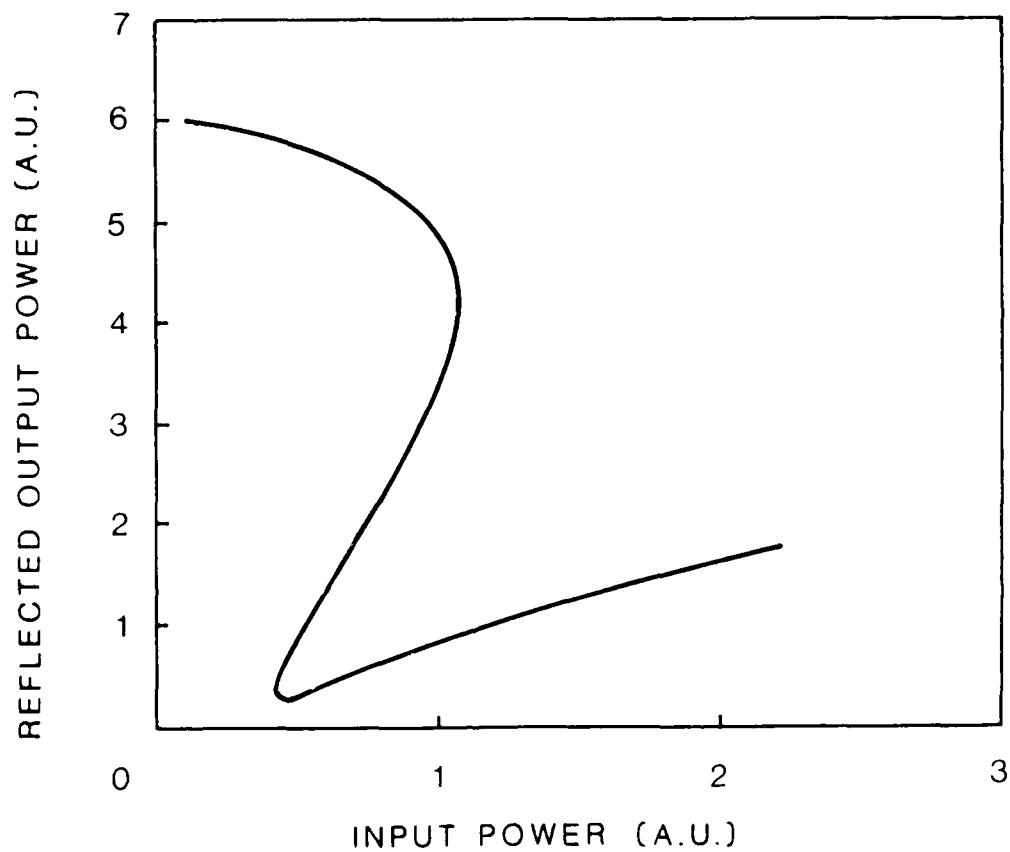
POLARIZATION-BASED OPTICAL PARALLEL LOGIC GATES USING FERROELECTRIC LIQUID CRYSTAL SPATIAL LIGHT MODULATORS

K. M. Johnson, M. Handschy, W. T. Cathey, N. Clark, D. Walba

Center for Optoelectronic Computing Systems
University of Colorado
Boulder, Colorado 80309-0425

Introduction

Optical computing systems offer an increased information processing rate by facilitating parallel computing architectures. Previous experience with electronic computers indicates that desired accuracy can be achieved only with digital computation. Since the simplest digital arithmetic is binary, most recent work on optical computing is focused on the construction of binary optical logic gates. Many practical implementations of such logic gates have been suggested; a recent review is given by Sawchuck and Strand [1]. Most previous schemes operate on light intensity, much in the way that electronic systems operate on voltage or current. Another natural optical scheme represents the two binary states with two orthogonal polarizations of light. The optical element necessary to implement this scheme is a device with two states, one of which passes light of a chosen polarization unchanged, and the other of which converts light of the chosen polarization to its orthogonal complement. Tsvetkov et al. [1] have described a practical implementation of this logic using the now common twisted nematic (TN) liquid crystal device, which has two voltage-selected states, one of which rotates the polarization direction of appropriately oriented linearly polarized light by 90° and the other of which has no rotary power. Another implementation would use any of the variable retardation effects such as the Pockels effect. One state of the device would be chosen to have zero retardation, and the other to have half-wave retardation. In addition to either passing unchanged or imparting 90° rotation to linearly polarized light, this scheme could also work by either passing unchanged or reversing the handedness of circularly polarized light. An advantage pointed out by Lohmann [3] that any implementation of polarization-based logic has over logics based on intensity is that no light is lost in the logical operation of inversion. In intensity-based logics, it is difficult to invert an already dark input, since light has to be "recreated"; polarization-based elements, as described above, can convert the light representing either logical state to the other, making easy the realization of any desired Boolean function.

We describe below a third implementation, in which the optical element is a ferroelectric liquid crystal device that functions as a half-wave plate whose axis can be electrically toggled between two orientations that make a 45° angle to each other. These elements have extremely useful operating characteristics for optical parallel processing, including fast response time (submicrosecond), low-power, low-voltage switching (tens of Volts), and bistability [4]. FLC elements have already been used in an intensity-based logic scheme, where their high contrast (up to 1500) has been exploited to advantage [5]. The polarization-based gate can perform all 16 Boolean logic functions possible with two binary inputs, without the need to manually remove or change any of the optical elements. In particular, we show especially simple implementations of the XOR and XNOR logical operations.

FLC Electrooptics

Ferroelectric liquid crystals possess properties especially attractive for optical logic applications when used in the so-called surface-stabilized geometry, which has been described extensively elsewhere [6, 7, 8]. Briefly, the FLC is disposed between two closely spaced glass plates, coated on their inner surfaces with a transparent electrical conductor. The FLC material itself is optically uniaxial (we ignore a weak biaxiality), with the uniaxis coupled to the ferroelectric polarization $\vec{P}$ so that when $\vec{P}$ is perpendicular to the glass plates, the uniaxis is parallel to them. Two such orientations of $\vec{P}$ are easily selected by voltages applied across the transparent electrodes; $\vec{P}$ prefers to be parallel to the resulting electric field $\vec{E}$. The optic axis states selected by applied voltages of opposite sign, while both parallel to the plates, differ in orientation by an angle $2\psi_0$, where the "tilt angle " ψ_0 is a material property determined by the thermodynamic characteristics of the FLC. Many FLC materials have ψ_0 close to 22° over large temperature ranges, allowing the optic axis to be electrically rotated through approximately 45°. If the thickness d of the FLC layer is chosen so that $\Delta n = \lambda/2$, where Δn is the FLC's birefringence and λ is the vacuum wavelength of the incident light, the FLC becomes a half-wave plate. If the polarization of normally incident light is chosen either parallel or perpendicular to one of the voltage-selected optic axis states, it will be transmitted through the FLC unaffected. The optic axis state selected by the opposite applied voltage is then 45° to either incident polarization, so that both the ordinary and extraordinary modes will be excited. For correct FLC elements thickness d at total phase shift of π will accumulate between these two modes, and the incident light's polarization will be rotated by 90°.

Beside the previously mentioned switching speed, the surface-stabilized FLC geometry offers another feature useful in optical logic systems: bistability. After either applied voltage brings the optic axis to one of its preferred orientations, that voltage may be removed without the optic axis returning to its previous state. This allows a two-dimensional array of FLC elements to be matrix addressed. For instance, if the conductors are divided on one plate into column electrodes and on the other plate into row electrodes, appropriate waveforms applied to the rows and columns would allow a selected element where a given pair of row and column electrodes overlap to be changed without disturbing any of the other elements in the array. A practical scheme for accomplishing this has been demonstrated by Wahl et al. [9], who achieved 1000:1 multiplexing. Thus, a large number of FLC elements ($1000 \times 1000 = 10^6$) can be simply fabricated on a single substrate, and driven with an economical number of electrical connections.

Ferroelectric Liquid Crystal Logic Gate

The XOR (AB' + A'B) and XNOR (AB + A'B') Boolean functions are the most difficult to implement optically using bright and dark logic. This is because light is irretrievably lost when creating not A (A') and not B (B'). Logic gates using bright and true logic, therefore, require four separate inputs; A, B, A', and B'.

With polarization logic, these functions are easily implemented using two FLC arrays, an optical controller, and an analyzer as shown in Fig. 1. In this gate light is not absorbed, and does not require regeneration.

For the XOR operation, the controller is in a non-switched state, and vertical light illuminates FLC array A. This array is a programmable matrix made up on transparent pixel elements which either rotate or do not rotate incident light (switched or not switched pixels). When vertically polarized laser light illuminates the switched pixels, the light is rotated to the horizontal polarized state. When the incident laser light illuminates non-switched pixels, no rotation occurs and vertical light is transmitted. A pattern made up of horizontal and vertical polarized light illuminates FLC array B. If either vertical or horizontal light illuminates a switched pixel in FLC B, the polarization is rotated by 90°; vertical rotates to horizontal and horizontal rotates to vertical. If light is incident on a non-switched pixel, the

180

transmitted light retains its polarization. The truth table in Fig. 2 summarizes the logical function. An analyzer at the output provides visual inspection of the XOR function.

To realize the XNOR, the FLC optical controller is switched which rotates the incident vertical laser light to horizontal light. The truth table for the XNOR function is also shown in Fig. 2.

Conclusions

We describe a new optical parallel logic gate implemented with spatial light modulators made of arrays of ferroelectric liquid crystals (FLC) electrooptic elements. The unique optical properties of the FLC elements make particularly simple a logic where two orthogonal polarizations of transmitted light represent the two binary states. A feature of this logic is that light need never be absorbed, allowing all 16 Boolean functions of two binary inputs to be implemented in a single gate; additionally, cascaded gates are equally feasible. FLC's also confer the advantages of submicrosecond switching speed and intrinsic two-state memory.

We will also discuss progress in synthesizing new FLC materials with faster switching speed, improved contrast ratio and temperature stability. Scattering and insertion losses, and switching energy measurements will be presented. A comparison of the FLC spatial light modulator with the deformable mirror device, the silicon PZLT, and the magneto-optic spatial light modulators will be made.
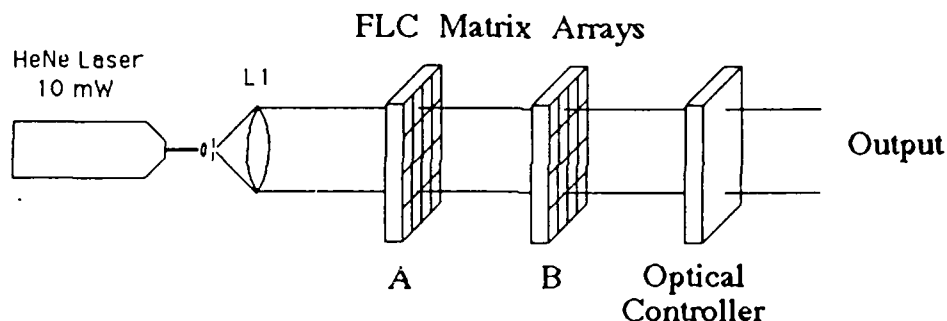
Figure 1. FLC XOR and XNOR Optical Logic Gate.

References

(1) A. A. Sawchuck and T. C. Strand, *Proc. IEEE* **72**, 758 (1984).

(2) V. A. Tsvetkov, N. A. Morozov, and M. I. Ellinson, *Sov. J. Quant. Electron*, **4**, 989 (1975).

(3) A. W. Lohmann, *OSA Annual Meeting, (1985)*.

(4) N. A. Clark and s. T. Lagerwall, *Appl. Phys. Lett.* **36**, 899 (1980).

(5) L. A. Pagano-Stauffer, K. M. Johnson, N. A. Clark, and M. A. Handschy, to appear *Proc. SPIE* (1986).

(6) N. A. Clark, M. A. Handschy, and S. T. Lagerwall, *Mol. Cryst. Liq. Cryst.* **94**, 213 (1983).

(7) M. A. Handschy and N. A. Clark, *Ferroelectrics* **59**, 69 (1984).

(8) S. T. Lagerwall and J. Wahl, to appear *Mol. Cryst. Liq. Cryst.* (1986).
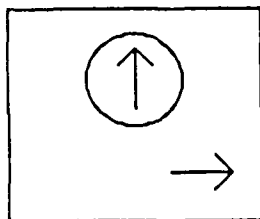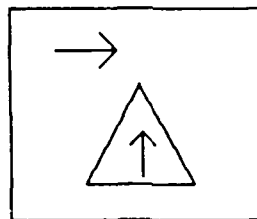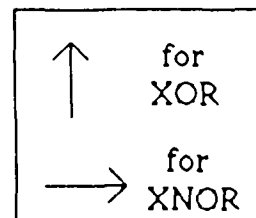
Figure 2   XOR and XNOR Polarization Truth Table

OPTIMUM CONTROL BEAM ANGLE FOR A BIASED FABRY-PEROT BISTABLE DEVICE

R. Cush, I. Bennion

Plessey Research Caswell Limited
Allen Clark Research Centre,
Caswell, Towcester, Northants

1. INTRODUCTION

The use of bistable Fabry-Perot devices as multi-port cascadable devices has been frequently discussed in the literature (1-3). It has been envisaged that such devices could be addressed by a number of bias and control beams each approaching the device at a different angle (fig. 1). However, if these devices are to be cascadable, the same wavelength of light must be used for all control and bias beams. In this situation the control beam intensities inside the devices will be determined by the amplification effects of the cavity. Their behaviour will differ from that of the bias beam as the different incident angles produce additional phase changes. It has been shown that these additional phase changes give rise to bistable switching as the incident angle of the beam is changed (4) and this behaviour is demonstrated experimentally.

In order to cascade devices and achieve maximum fan-out it is necessary to minimize the amount of light required for switch-on. By using the variable amplification effect of the cavity with incident angle, an optimum angle for addressing the device is identified, together with angles at which the switch-on energy is greater.

2. BISTABILITY WITH ANGLE

When considering non-normal incidence angles the Airy function describing cavity behaviour is modified (4, 5):

$$J_t = J_0 . K \left[1 + F\sin^2 \left(G + (\phi_0 - J_t)^2 - (\phi_0 \sin\theta/n_0)^2 \right)^{1/2} \right]^{-1} \qquad (1)$$

where J_0, J_t are the normalized incident and transmitted intensities,:

$J = A (1 + R\alpha)2\pi Ln_2 I/\alpha L(1 - R)(1 - A)\lambda$ $K = (1 - R)^2(1 - A)/(1 - R\alpha)^2$

I is the intensity R is the cavity reflectivity

$F = 4R\alpha/(1 - R\alpha)$ $R\alpha = (1 - A)R$

$A = 1 - e^{-\alpha L}$

α is the coefficient of linear absorption $\phi_0 = 2\pi n_0 L/\lambda$

θ is the incident angle $G = 2\pi t\cos\theta/\lambda$

t is the width of the cavity air gap

L is the length of the nonlinear material

$n_0 - n_2 I$ is the refractive index of the nonlinear material λ is the incident wavelength

The nonlinear refractive index occurs in both the cavity path length term and also in the term determining the angle of the beam inside the cavity via Snells Law. It gives rise to multivalued solutions to the above equation and hence bistability for certain values of incident angle .

The values of the loss terms K. $R\alpha$, etc. will vary with angle but over the angles considered here the variation is small and has been neglected in order to simplify the calculations.
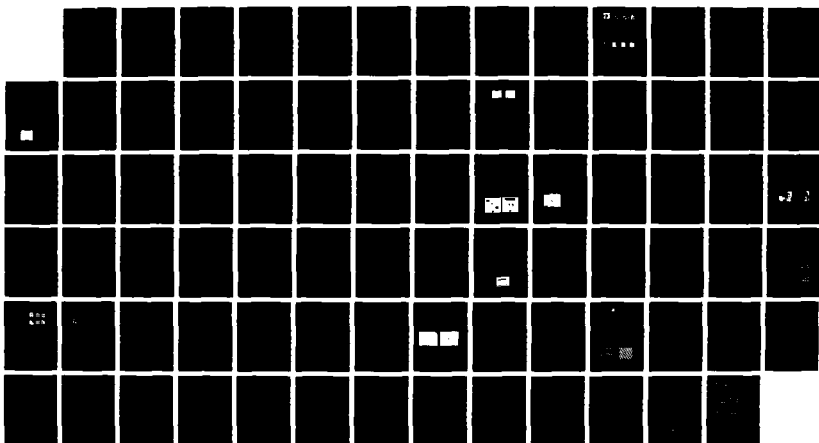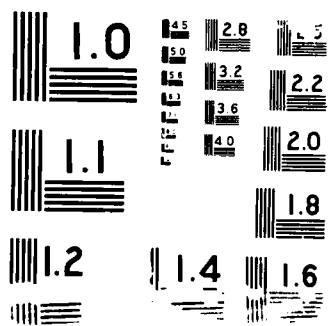
3. EXPERIMENTAL VERIFICATION

The system used to verify the above behaviour experimentally comprised a Fabry-Perot cavity with mirrors of reflectivity R = 0.861 containing a nonlinear material layer and an air gap. The nonlinear material consisted of an organic material, 2-xanthylidine indan-1-3-dione combined with a polymer matrix of diethylene glycol bis(allyl carbonate), 100µm thick, produced by the process of solvent assisted indiffusion (6). This material exhibits a thermal nonlinearity when exposed to light at 514.5nm wavelength. The resulting cavity had a finesse of 6 (F = 15).

The experimental arrangement is shown in fig. 2. The incident beam was set at a level of 22Wmm^{-2} and its angle of incidence varied by lateral translation of the beam before the focussing lens. The experimental results are shown in fig. 3.

The values of F, n_2 (= 9 x 10^{-6}Wmm^{-2}), I_0 and L for the experimental arrangement were used in equation 1 and the initial tuning position varied to match the experiment. The results are shown in fig. 4 and correspond to an air gap of t = 180µm. This figure was confirmed by measurements on the cavity which gave a value for t of around 170µm. The agreement between the experiment and theory confirms the model as a description of this cavity arrangement.

4. MULTIPLE BEAM ADDRESSED CAVITY

For a cavity addressed by two beams, a bias beam at normal incidence and a control beam incident at an angle θ, two coupled equations exist which describe cavity behaviour (4):

bias beam.
$$J_{t_1} = J_{0_1}.K.(1 + F\sin^2(\theta_0 - \gamma + G))^{-1} \qquad (2)$$

control beam:
$$J_{t_2} = J_{0_2}.K.(1 + F\sin^2(G + ([\theta_0 - \gamma]^2 - [\theta_0\sin\theta/n_0]^2)^{1/2}))^{-1} \qquad (3)$$

where $\gamma = (J_{t_1} + J_{t_2})$

In order to create a switch the bias beam must be set at a level within the bistable loop of the cavity. The initial detuning of the cavity was set and the corresponding bistable loop with changing incident intensity was calculated (fig. 5). From this a value of 16 for the normalized bias beam intensity was chosen.

To determine the optimum angle, the intensity of the control beam required for switch-on was calculated by first setting an angle and then gradually increasing the control beam intensity from zero until switching occured. This was done for a number of angles and the results are shown in fig. 6.

5. OPTIMUM CONTROL BEAM ANGLE

Examining the graph shown in fig. 6 it can be seen that for zero incident angle the intensity required for switch on is that expected from fig. 5 for the bias beam alone. However, as the incident angle of the control beam is increased the amplification effects of the cavity result in the switching intensity being lower by up to a factor of four. This would allow four times as many devices to be switched by a transmitted beam compared to the normal incidence system.

However, what is perhaps more significant is the next region where the incident angle results in a far greater control beam intensity being required for switching. At an angle of 0.037 radians, a four times greater intensity than the normal incidence case is required. The use of these angles is obviously detrimental to the system and so limits the number of suitable angles which can be used.

Experiments to measure the variation in switch-on intensity of the system described in section 3 have been carried out and do indeed show changing switch-on energy with angle. These results will be discussed in detail and composed with the theory.

The parameters discussed here correspond to the experimental system described. This cavity has a much greater separation than many of the other bistable Fabry-Perot devices used as logic gates. However, even much thinner cavities behave as described here when larger ranges of incidence angle (up to 45^0) are considered, angles which may well be used when cascading devices.

6.   SUMMARY

The changes in the cavity path length which occur with non-normal incidence angles give rise to the observation of bistable switching when varying the incident angle.

This behaviour is also significant when the non-normal incidence beam is used as a control beam to switch a second beam. At certain angles the amplification effects of the cavity result in a reduction in the intensity required for switching, which is of great importance in order to minimize energy requirements and maximize device cascading.

However, for certain incident angles the effect of the cavity is to greatly increase the switching energy required so limiting the angles at which the device may be addressed.

7.   ACKNOWLEDGEMENTS

8.   REFERENCES

1.   S. D. Smith, I. Janossy, H. A. MacKenzie, J. G. H. Mathew, J. J. E. Reid, M. R. Taghizadeh, F. A. P. Tooley, A. C. Walker, Opt. Eng. 24 (1985) 569

2.   E. Abraham, C. T. Seaton, S. D. Smith, Sci. Am. (March 1983) 63

3.   A. Huang, Proc. IEEE 72 (1984) 780

4.   C. H. Golaire, P. Mandel, P. Laws, Opt. Commun. 57 (1986) 297

5.   D. A. B. Miller, IEEE J. Quant. Elec. QE-17 (1981) 306

6.   M. J. Goodwin, R. Glenn, I. Bennion, Elec. Letts, 22 (1986) 789

Fig. 1
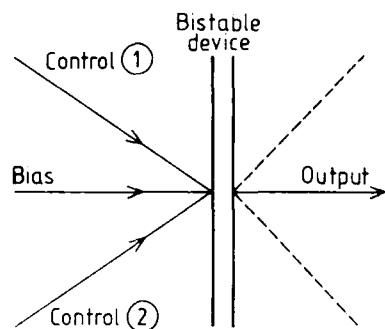
1. Multiport bistable device.
2. Experimental arrangement for measuring bistability with changing incident angle.
3. Experimental results showing bistability with changing incident angle.
4. Theoretically predicted bistability corresponding to experimental conditions.
5. Bistable behaviour of bias beam.
6. Variation in control beam switch-on intensity with angle.
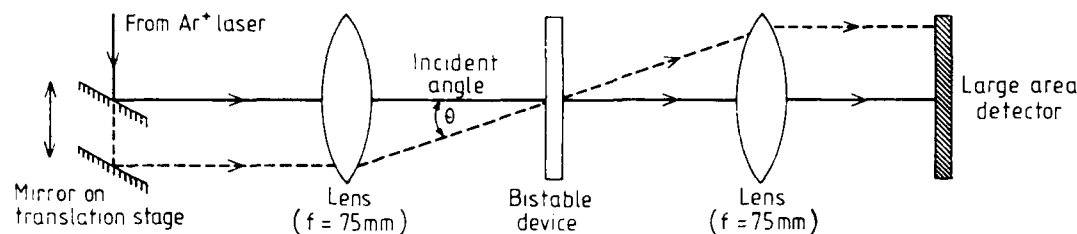
Fig. 2
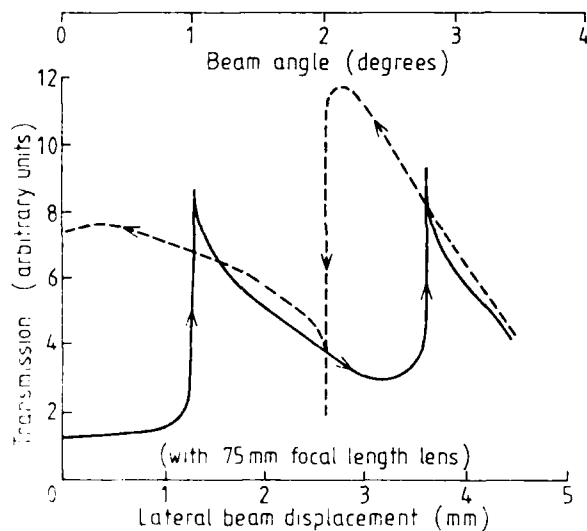
Fig. 3
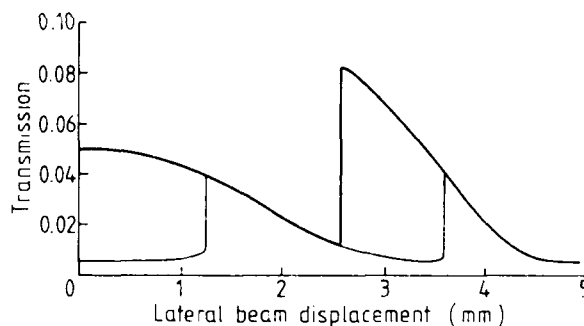
Fig 4
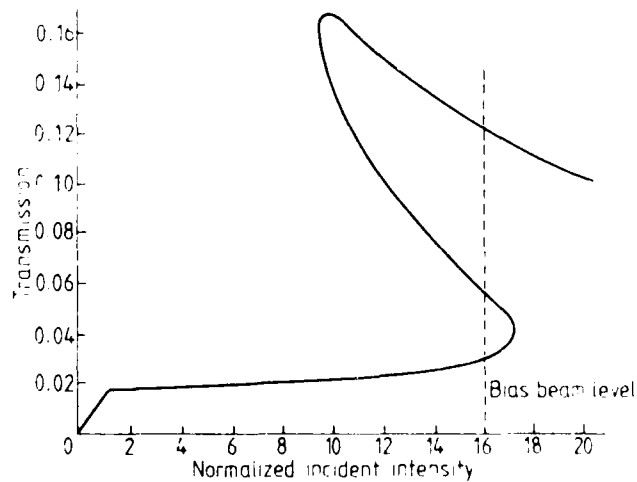
Fig. 5
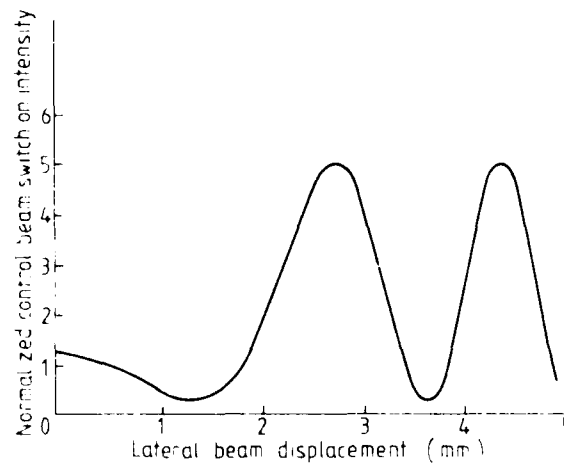
Fig. 6

TUESDAY, MARCH 17, 1987

PROSPECTOR / RUBICON ROOM

3:00 PM–5:30 PM

TuD1–7

SESSION 9

Carl M. Verber, Georgia Institute of Technology, *Presider*

Materials and Devices for Optical Computing

Armand R. Tanguay, Jr.
University of Southern California

VARIABLE-GAMMA SPATIAL LIGHT MODULATOR

Suzanne Lau and Cardinal Warde

Department of Electrical Engineering and Computer Science
Massachusetts Institute of Technology, Cambridge, MA 02139

INTRODUCTION

A high-resolution, optically-addressed, variable-gamma spatial light modulator would be very useful for general-purpose, real-time optical processing. In this paper, we describe how the standard and Fabry-Perot Microchannel Spatial Light Modulators (MSLM)[1-3] can be modified and operated to achieve a wide range of gamma similar to that for phototgraphic emulsions (see Fig. 1). Such a variable-gamma device will be applicable to both linear (low gamma) and nonlinear (high gamma) optical processing. Preliminary image processing results are presented.
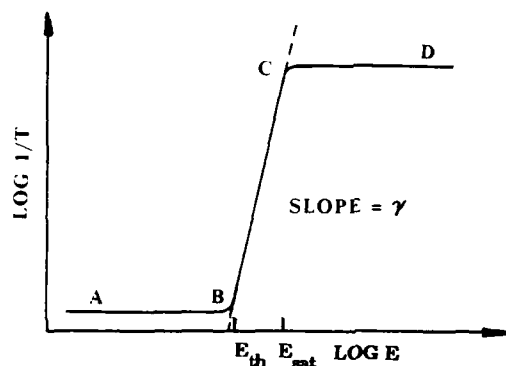
Fig. 1. Gamma characteristic

GAMMA CHARACTERISTIC

The gamma characteristic is a plot of log $1/T$ vs. log E_m, where T is the modulator readout transmission and E_m is the corresponding modulator input exposure. The input exposure is the product of the incident intensity I_w and the exposure time t_w. For a negative gamma characteristic, the readout light remains in the ON state for input exposures below threshold, E_{th}, and is OFF for exposures above saturation, E_{sat}. The gamma γ of the device is the slope of the linear region of the curve, and is given by

$$\gamma = \log(T_{th}/T_{sat})/\log(E_{sat}/E_{th}) \tag{1}$$

The MSLM is capable of both positive and negative variable-gamma characteristics. The techniques presented for manipulating the gamma of an MSLM include: (1) converting the standard device to a Fabry-Perot MSLM, (2) operation in the real-time nonlinear hard-clipped thresholding mode, and (3) stored-image analog thresholding.

MICROCHANNEL SPATIAL LIGHT MODULATORS

Standard MSLM

The standard MSLM is a versatile, real-time image processing device which exhibits high optical sensitivity and a high framing speed. It consists of a photocathode, a

microchannel plate (MCP), a planar acceleration grid and an electro-optic crystal plate (see Fig. 2). The crystal has a high-resistivity dielectric mirror on the side that faces the grid, and a transparent conducting electrode on the other.

In the electron-deposition mode, the write beam (coherent or incoherent light) incident on the photocathode creates an electron image which is amplified by the MCP and proximity focused onto the dielectric mirror. The resulting spatially varying electric field modulates the refractive index of the crystal. Thus, the readout light which makes a double pass through the crystal is phase or amplitude modulated, depending on the crystal cut and readout scheme (polarization or interferometric) employed.

The image is erased by flooding the photocathode with light so that the electrons are removed from the mirror by secondary electron emission. Alternatively, the device can be operated in the reverse mode, in which the image is written by removing charge from the dielectric mirror surface by secondary electron emission and erased by adding charge to the mirror.

In the linear operating regions, the incremental surface charge density $\sigma(E)$ deposited on the crystal is proportional to the exposure E_m. For Pockel's effect crystals, the induced phase change $\Delta\phi_x{}'$, $\Delta\phi_y{}'$ in the crystal is proportional to $\sigma(E)$, and for read out between crossed polarizers, the transmittance of the crystal is given by

$$T = I_r/I_i = \sin^2(\Gamma/2) \tag{2}$$

where $\Gamma = \Delta\phi_x{}' - \Delta\phi_y{}'$ is the phase retardation.
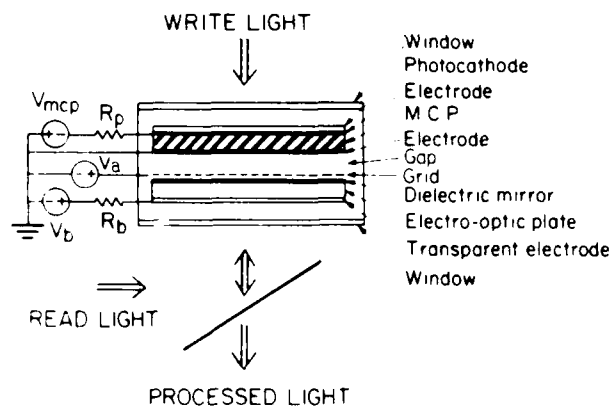
Fig. 2.   Microchannel Spatial Light Modulator
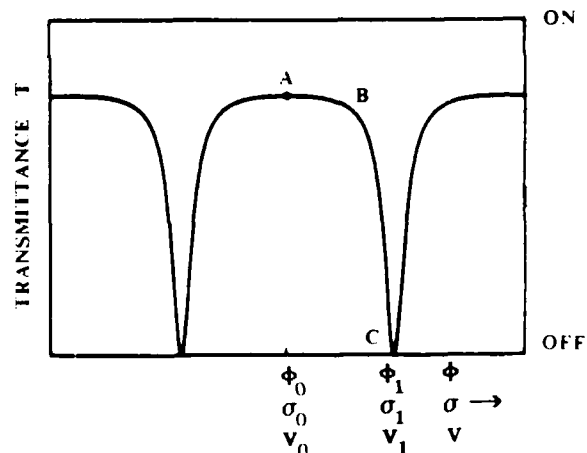
Fig. 3.   Readout characteristic of a Fabry-Perot crystal

Fabry-Perot MSLM

The Fabry-Perot version of the MSLM employs a crystal that functions as an electro-optically tunable, Fabry-Perot etalon when electrons are deposited or removed from its surface. To fabricate such a crystal standard dielectric mirrors are deposited on both surfaces of the crystal. For a Fabry-Perot etalon with surfaces of reflectivity R, it is well known that the ratio T, of the total reflected intensity to the incident readout intensity is given by

$$T = I_r/I_i = \frac{4R \sin^2\left[\frac{\phi}{2}\right]}{(1-R)^2 + 4R \sin^2\left[\frac{\phi}{2}\right]} \tag{3}$$

Fig. 3 is a plot of T vs ϕ as given by Eq. (3). Note that the reflected readout transmission of the MSLM approaches zero when ϕ takes on integer multiples of 2π radians.

Note that the region A-B-C of the Fabry-Perot characteristic closely approximates the corresponding part of the desired gamma characteristic of Fig. 1. It is then only necessary to avoid multiple-valued output, by ensuring that no points in the write image can be driven past the point C in Fig. 3. This is best accomplished by operating the device in the real-time hard-clipped thresholding mode. Gammas as high as 10 are expected from Fabry-Perot MSLMs.[3]

VARIABLE-GAMMA OPERATION

The intrinsic gamma of the MSLM, defined under linear operating conditions, depends on the specific characteristics of the modulating element. For an electro-optic crystal, the parameter which influences the gamma is the halfwave surface charge density σ_π.[3] In the case of the Fabry-Perot device gamma also depends on the mirror reflectivity, R. However, this intrinsic device gamma can be varied by altering the operating conditions as described below.

Real-Time Hard-Clipped Thresholding Mode

Both the standard and the Fabry-Perot MSLMs can be operated in the hard-clipped thresholding mode to vary their gamma. To operate the MSLM with a negative gamma characteristic, the device is first biased in the ON state with a charge density of σ_0 (point A in Fig. 3 or at a peak of the $\sin^2\Gamma/2$ characteristic of a standard device). The grid voltage is set to V_1, corresponding to the OFF state (point C), and then, with the optical input image incident on the photocathode, the crystal voltage V_b is ramped downward from V_0 at some preselected rate, $\dot{V}_b$, to a terminal voltage which is usually selected to be the grid voltage. The rate $\dot{V}_b$ establishes both the threshold exposure E_{th}, and the gamma. All exposures below E_{th} will be barely recorded and remain in the ON state; those above E_{sat} will saturate at the grid voltage (OFF state); and the exposures in between will be determined by the gamma.

Preliminary results were obtained using a Hamamatsu vacuum-sealed standard device[5] employing electron optics for focussing the image onto the MCP and a 50-μm-thick, oblique-cut lithium niobate crystal. Due to voltage limitations for this particular device, gamma values were measured in an operating region around the $\sigma_{\pi/4}$ point. Variable-gamma and variable-threshold operation was achieved by varying the ramp rate (i.e., the write time t_w from V_0 to V_1). Unlike photographic film, the results demonstrate that both the threshold exposure and the gamma depend on the specific values of the write time t_w and the write intensity I_w. We have found that gamma increases and the threshold decreases when either (1) the write intensity is increased with fixed write time (fixed ramp rate), or (2) the write time is increased (decreased ramp rate) with fixed write intensity. The above experiments have yielded gammas ranging from less than 0.4 to greater than 3.3; these values should be compared with the measured intrinsic gamma of the device (about 1.4 at $\sigma_{\pi/4}$) and to nominal values for high-gamma photographic film, around 2-3.[4] Figure 4 shows the results after thresholding a grayscale image of an M.I.T. student with increasing write times. Higher gammas would have been achieved for operation in the neighborhood of $\sigma_{\pi/2}$.
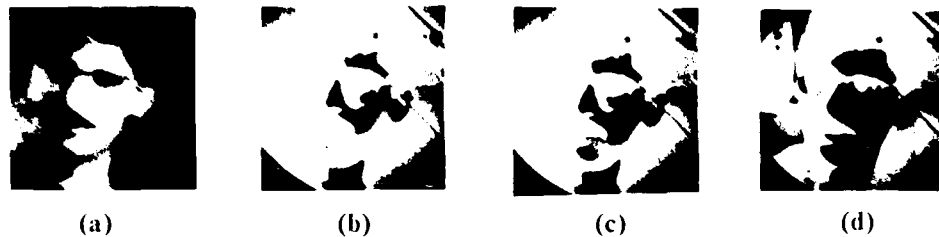
(a)   (b)   (c)   (d)

Fig. 4.   Hardclipped Thresholding: (a) unthresholded electron deposition image; (b)-(d) thresholding with increasing write times

Stored-Image Analog Thresholding Mode

Analog thresholding, performed on stored images, can also lead to an effective increase in gamma. Consider two areas of the crystal A' and A'', with electron densities σ' and σ'' repectively, within the electron distribution on the surface of the dielectric mirror. If σ'' is sufficiently greater than σ', then, when the photocathode is uniformly illuminated, and V_b slowly ramped downward from a preset threshold voltage V_{th}, primary electrons from the MCP will be repelled from A'' but not from A'. Under these conditions A' and all areas with electron density less than σ' will be erased, while A'' and all other areas for which $\sigma > \sigma''$ will be unaffected. This operation (analog thresholding) corresponds to erasing all parts of an image with exposure less than the threshold exposure E_{th}, and leaving the remainder unaffected. Thus, the threshold exposure is determined by V_{th}. Note that the gamma of the thresholded region is effectively increased. Selective thresholding may be attained by erasing using spatially nonuniform erase light. Fig. 5 shows the results of analog thresholding a grayscale bar chart with successively increasing V_{th}.

(a)   (b)   (c)   (d)

Fig. 5.   Stored-Image Analog Thresholding: (a) unthresholded electron deposition image; (b)-(d) thresholding with increasing V_{th}

SUMMARY

Preliminary results using a standard MSLM show that specific gamma characteristics can be achieved with reasonable accuracy. Operation in a low gamma mode would allow the device to replace photographic film in conventional linear processing (e.g., in pattern recognition, Fourier plane filtering and real-time holography). Possible high-gamma applications include logarithm, exponentiation, intensity level slicing, thresholding, analog-to-digital conversion, logic and bistability.

REFERENCES

1. C. Warde, A. M. Weiss, A. D. Fisher and J. I. Thackara, Appl. Opt. 20, 2066 (1981).
2. C. Warde and J. I. Thackara, Opt. Lett. 7, 344 (1982).
3. C. Warde, H. Lamela-Rivera and H. K. Liu, Proc. SPIE, 613, 255 (1986).
4. J. W. Goodman: "Introduction to Fourier Optics," McGraw-Hill (1968).
5. T. Hara, Y. Ooi, T. Kato and Y. Suzuki, Proc. SPIE, 613, 153 (1986).

INTEGRATED ELECTROOPTIC BRAGG MODULATOR MODULES FOR OPTICAL COMPUTING*

D.Y. Zang, P. Le, and C. S. Tsai

Department of Electrical Engineering
University of California
Irvine, CA 92717

SUMMARY

The prospects for realization of Lithium Niobate ($LiNbO_3$) -based integrated optic (IO) devices and modules for applications in RF signal processing and communications have already been firmly established. For example, the planar waveguide acoustooptic (AO) Bragg cells are now widely used in the development of IO modules for spectral analysis and correlation of wideband RF signals. A variety of electrooptic (EO) devices that utilize parallel channel waveguides and crossed channel waveguides have also been developed into compact modules for wideband communications and switching as well as high-speed analog to digital conversion of RF signals. While advancements on the IO device modules for communications and signal processing are continuing, serious attemps on realization of device modules for computation of both analog and digital data have also been started recently (1-4).

A significant advancement toward eventual realization of one form of hybrid integrated-optic computers has been made through fabrication of single-mode microlenses and microlens arrays in planar $LiNbO_3$ waveguides (5) using a simple technique entitled titanium-indiffused proton-exchanged (TIPE) (6). Through this TIPE technique a variety of lens combinations may be fabricated using a single masking step. These microlenses and microlens arrays have recently been integrated with channel waveguide arrays and AO and/or EO Bragg diffraction arrays to form a variety of IO modules in a substrate size as small as $0.2 \times 1.0 \times 1.8$ cm^3. Most recently, these IO modules have been utilized successfully to perform optical systolic array processing and computing as well as typical applications in RF signal processing and communications. In this paper, realization and measurement of such TIPE microlens-based integrated EO Bragg modulator modules and their applications in computing, e.g., matrix-vector and matrix-matrix multiplications (7-10) are reported.

Fig. 1 shows the architecture of one of the integrated EO Bragg modulator modules that have been realized. The fabrication steps involved are similar to those for integrated AO Bragg modulator module (2). Each channel waveguide is followed by a TIPE-microlens and an interdigital finger electrode pattern in the planar waveguide. Thus, an array of EO Bragg diffraction gratings are created by applying voltages across the array of interdigital finger

electrodes. Note that in this first version each electrode array was of the conventional type which consisted of a single array of parallel interdigital finger electrodes. Efficient and wideband Bragg diffraction have been achieved using the electrode arrays with 13 μm periodicity and 2.0mm aperture. Specifically, 95% diffraction at a drive voltage of 6.0 volt and 750 MHz rf bandwidth were measured. A second version as shown in Fig. 2 in which each diffraction grating consisted of a Herringbone electrode array (1) was also realized most recently. Again, efficient and wideband Bragg diffraction was readily obtained. It is important to note that the two separate electrode arrays of the Herringbone type facilitate application and thus multiplication of two independent sets of data. Thus, in contrast to their AO counterparts (2), these two integrated EO Bragg modulator modules can accept multiple sets of data at a much higher rate. An array of up to 12 individual modulators has been realized thus far in both versions of the EO modulator modules.

The two integrated EO Bragg modulator modules just described have been used to perform algebraic manipulations. In such application "Multiplication" is facilitated by EO Bragg diffraction, and "Addition" by the integrating lens. Specifically, the first version of the EO Bragg modulator module was used to perform matrix-vector multiplicaton. The components of the vector were separately applied to the element EO modulators while the column elements of the matrix were used to pulse-modulate separately the diode lasers (at 0.792 μm wavelength) prior to coupling into the channel waveguide array. As in the AO computing experiment (2), master pulse generator was employed to provide the synchronization between the matrix elements and the vector components. The compnents of the matrix-vector product were obtained from the output of a photodetector located at the focal plane of the integrating lens. The output of the photodetector readily produced the correct results of the matrix-vector product.

The second version of the EO Bragg modultor module was readily used to perform matrix-matrix multiplication. In this case the column elements of two matrices A and B were used to activate, respectively, the first and the second segments of the Herringbone electrode. Fig. 3 shows the correct result of multiplication involving the matrices $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, namely, $\begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$.

In summary, successful fabrication of high performance microlenses and microlens arrays using the TIPE technique has enabled realization of a variety of integrated EO Bragg modulator modules in the $LiNbO_3$ channel-planar composite waveguides of $0.2 \times 1.0 \times 1.8$ cm^3 substrate size. Through the channel-waveguide and the TIPE microlens arrays, the very large channel capacity that is inherent in the diode laser and the optical fiber as well as the photodetector arrays may be conveniently exploited. The encouraging results that have been demonstrted with a variety of experiments suggest that such integrated EO Bragg modulator modules can be utilized in future multi-channel optical computing as well as RF signal processing and communication systems.

REFERENCES

1. C. V. Verber, Proc. IEEE., 72, 949 (1984).
2. C. S. Tsai, D. Y. Zang, and P. Le, Appl. Phys. Lett., 47, 549 (1985).
3. R. Arrathoon, E. R. Schroeder, and L. D. Hutcheson, Opt. Lett., 10, 244 (1985).

4. T. J. Bicknell, D. Psaltis, and A. R. Tanguay, Jr., J. Opt. Soc. Am., 2, (1985).
5. D. Y. Zang and C. S. Tsai, Appl. Phys. Lett., 46, 703 (1985).
6. M. DeMicheli, J. Botineau, P. Sibillot, D. B. Ostrowsky, and M. Papuchon, Opt. Commun., 42, 101 (1982).
7. H. J. Caulfield, W. J. Rhodes, M. J. Foster, and S. Horvita, Opt. Comm., 40, 86 (1981).
8. M. Tur, J. W. Goodman, B. Moslchi, J. E. Bowers, and J. H. Shaw, Opt. Lett., 7, 463 (1982).
9. D. Casasent, Appl. Opt., 21, 1859 (1982).
10. R. A. Athale, W. C. Collins, P. D. Stilwell, Appl. Opt., 22, 368 (1983).
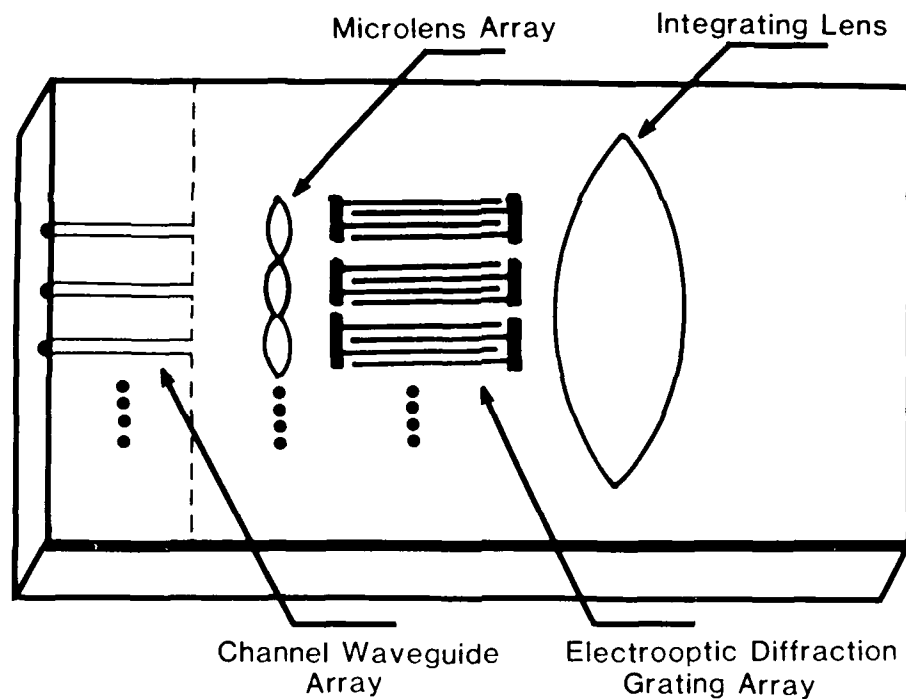
Fig. 1 Integrated Electrooptic Bragg Modulator Module In Y-Cut LiNbO$_3$ For Matrix-Vector Multiplication
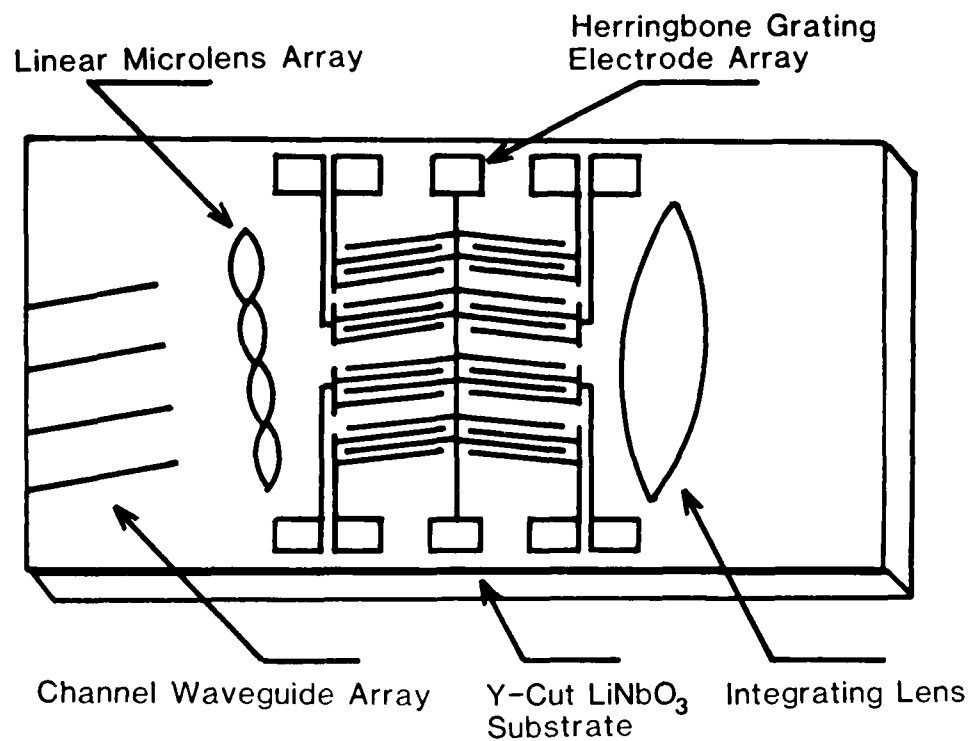
Fig. 2 Integrated Electrooptic Bragg Modulator In Y-Cut LiNbO$_3$
For Matrix-Matrix Multiplication
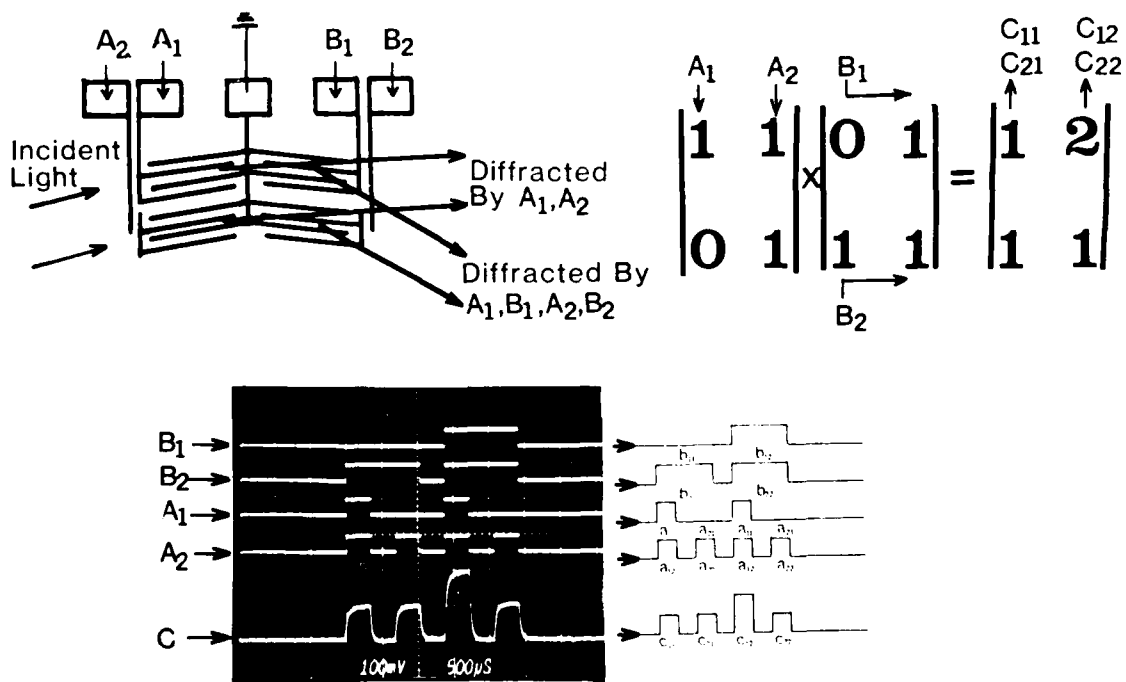
Fig.3 2X2 Matrix-Matrix Multiplication Using Integrated Electrooptic Bragg Module

Towards An Optical/Electronic Hybrid Image Processor

M G Nicholson, G G Gibbons, S Mayo and C R Petts

GEC Research Ltd
Hirst Research Centre
East Lane
Wembley, Middlesex
HA9 7PP, United Kingdom

B Loiseaux and J P Huignard

Laboratoire Central de Recherches
Thomson-CSF
Domaine de Corbeville
BP 10 91401
Orsay, France

F Dubois and J Ebbeni

Universite Libre de Bruxelles
Service des Milieux Continus
Avenue Adolphe Buyl 87
B1050 Bruxelles
Belgium

SUMMARY

1 Introduction

This paper describes work directed at utilising the parallel processing
power of relatively simple analogue optical systems, to improve the
capability, compactness and cost of machine vision systems. We are
investigating methods which allow the combination of the best features of
both electronic and optical systems in such a way that the control and
decision-making are performed by the digital electronic system, whilst the
parallel computations and data reduction are performed by analogue optical
means. The resulting machine is termed a hybrid processor. Figure 1 shows a
generalised block diagram of the Optical Functional Unit of such a system.

In this contribution, we will concentrate on an Optical Processing Unit
(OPU) that performs fast image correlation with a variable spatial frequency
weighting.

Many optical systems that are capable of performing image correlation have
been demonstrated since the original proposal by Vander Lugt [1]. These
systems have used pre-recorded holographic filters as the main processing
elements, which although effective for small numbers of reference-image
sets, does not lend itself to the iterative search procedure required for
identification of large numbers of reference images, unconstrained
recognition of three dimensional objects or syntactic pattern recognition.

For applications involving iterative schemes it is useful to be able to alter the filtering function of the processing element under control of the decision-making system. The use of dynamic holography in photorefractive materials have been proposed and demonstrated for this purpose [2,3,4]. Recently a photorefractive optical correlator has been shown to be operable at video frame rates [5] and a related system using a pulsed Nd:YAG laser has been shown to offer major practical advantages in terms of immunity to vibrational problems and overall power consumption, as well as forming the correlation products of two input images in under 200 nanoseconds [6].

2 The Hybrid processor

In this contribution, recent work on components required to build a practial hybrid processor will be described together with predictions of the overall system performance. The component parts are considered below.

The image correlator and input interfaces

A particular implementation of the generalised optical functional unit of Figure 1 is shown in Figure 2. The correlator uses a pulsed laser system to give a 'TEST' image framing rate of up to 50 frames/second. The correlation product is formed instantaneously by illumination of the reference image, the framing rate being limited only by the control system and the control interface (the reference-channel spatial light modulator). In practise the input (TEST) scene would be derived from a TV addressed spatial light modulator. The reference scene is provided by an electrically addressed spatial light modulator via a frame buffer. The form of the correlation function can be altered to incorporate different spatial frequency weighting [5] via an electrooptic modulator under the control of the decision-maker.

Optically addressed spatial light modulator

The optically addressed spatial light modulator (SLM) [9] consists of an association of a bulk photoconductor crystal with a thin electrooptic layer. The photoconductive material which is a $Bi_{12} SiO_{20}$ monocrystal has an optimum sensitivity in the spectral range of the data writing source. Under local illumination the photoconductor impedance decreases so that the electrical voltage applied to the cell is transferred to the liquid crystal layer thus modifying its optical properties. The resulting spatial distribution of birefringence allows the encoding of the amplitude and/or phase of the readout beam according to the spatial distribution of the data projected onto the device.

This transducer is used for the coherent conversion of the image displayed onto a CRT screen and projected on the BSO - Liquid Crystal SLM. The characteristics of this device will be reviewed as well as the problems posed by coupling to the CRT. The main performance of actual SLMs produced to date are as follows:

Static spatial band pass at 50% from the maximum : 10 $lp.mm^1$

Writing light intensity : 300 $\mu W.cm^{-2}$ at $\lambda = 400$ nm

Light value aperture 30 x 3u mm^2

Synthetic discriminant function

The access time of the reference image framestore, together with the
response time of the optoelectronic interface devices and the processing
time of the electronic controller will all limit the data throughput of the
hybrid system. In order to increase the processing speed, synthetic
discriminant functions [7] are used such that a number of reference images
can be inputted in parallel. In the updatable image correlator, it is
likely that the SDF will be loaded from a framestore in the image plane,
consequently the SDF must be real and positive everywhere. We will show
that this requirement effectively means that each member of the SDF training
set must contain the same amount of energy, i.e. that the diagonal
elements of the correlation matrix must be greater than any off-diagonal
element. This will limit the utility of the SDF to providing rotational
invariance. In our contribution we will demonstrate the output from an
updatable correlator that uses an SDF reference image. The output will be
compared with theoretical predictions on performance.

Correlation plane analysis

The hybrid processor must use fast optical correlation combined with
flexible electronic image processing. We will discuss various digital
techniques for reference image display and correlation plane analysis; these
will include grey-level slope histograms and Freeman chaincodes [8]. The
prospects for real-time implementation will also be considered.

ACKNOWLEDGEMENTS

REFERENCES

1   L A VanderLugt, IEEE Trans. IT-10 (2), 139, (1964)
2   D M Pepper et al, Opt. Letts., 3, 7, (1978)
3   L Pichon and J-P Huignard, Opt. Commun., 36, 277, (1981)
4   L Connors et al, Proc. SPIE 492, 361, (1984)
5   I R Cooper et al, IEE(J), 133, 70, (1986)
6   M G Nicholson, G G Gibbons, L C Laycock and C R Petts, to be published
    in Electronics Letters
7   D Casasent, Appl. Opt., 23, 1620, (1984)
8   F Merkle and T Lorch, Applied Optics, 23 (10), 1509, (1984)
9   P Aubourg et al, Appl. Opt., 21, 3706, (1982)
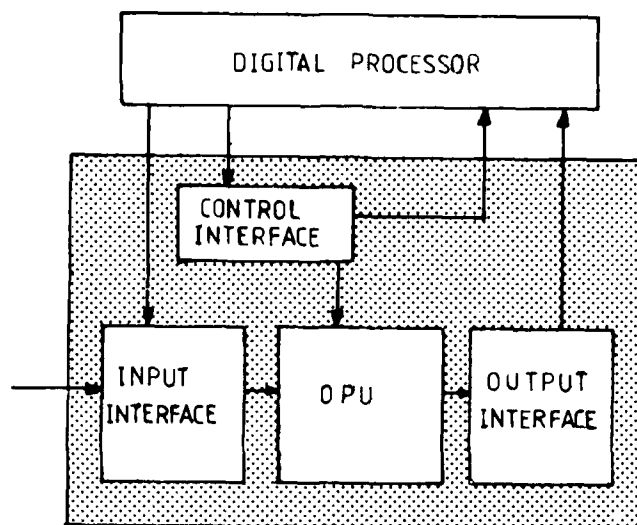
Figure 1   Block diagram of an optical functional unit
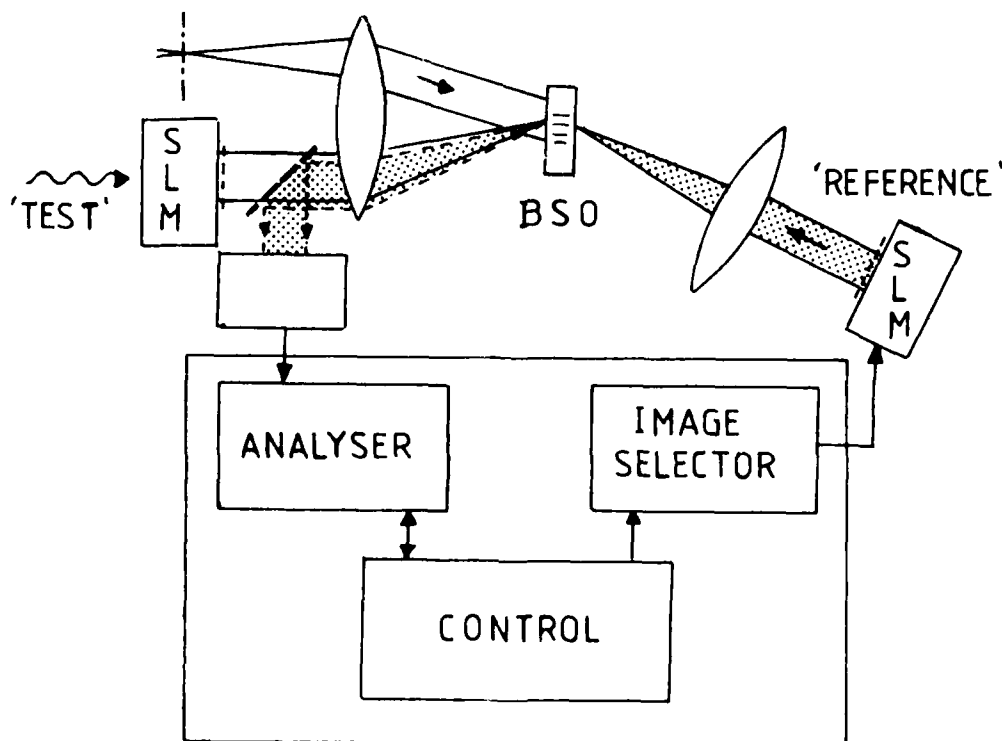
Figure 2   The Hybrid image correlator

INDEX GRATING LIFETIME IN PHOTOREFRACTIVE, SEMI-INSULATING, Cr-DOPED GaAs*

Li-Jen Cheng and Afshin Partovi
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

and

Elsa Garmire
Center of Laser Studies
University of Southern California
Los Angeles, California 90089-1102

Compound semiconductors, such as GaAs and InP, are potential photo-refractive materials for real-time volume holographic elements in optical computing applications. The advantages of these photorefractive compound semiconductors include fast response[1,2] (now down to picoseconds[3]), high sensitivity[1,4], reconfigurability, high degree of parallelism, and operation at infrared wavelengths compatible with semiconductor lasers[5] as well as VLSI technology. Recently, Cheng and Partovi[4] investigated temperature and intensity dependence of photorefractive effect in semi-insulating, Cr-doped GaAs, revealing some operation characteristics for the semiconductor as a practical device. For example, at room temperature, minimum beam intensities of about 10 mW/cm^2 are needed to form index gratings of near-saturation amplitudes, but the requirement increases to about 100 mW/cm^2 at 50°C. This is due to the competing effects of the dark and light-induced electrical conductivities.

In this paper, we report lifetime of index grating in semi-insulating Cr-doped GaAs as a function of reading beam intensity and grating spacing at room temperature. The result is critical for realistic evaluation of the GaAs:Cr potential as practical volume holographic elements for optical computing applications.

The experimental technique used was beam coupling (two-wave mixing) with a 1.7 mW, 1.15 micron He-Ne laser. Samples used were semi-insulating Cr-doped GaAs crystals, similar to those described in Reference 4. The laser beam was split into two beams of the same intensity. One of the beam (pump beam, I_p) was incident on an electronically controlled shutter. The other beam (signal beam, I_s) passed through a variable neutral density filter for intensity varia-tion. When the shutter was open, the two beams interfere in the crystal to form an index grating. The intensity of the second beam (signal beam, I_s) was increased due to beam coupling in the crystal. When the shutter was closed,

---

the energy transfer from I_p to I_s was stopped immediately. However, the grating was still there because of its finite lifetime. Its amplitude started to decrease with a characteristic lifetime, depending mainly on I_s, which was still illuminating the sample and simulating a reading beam. If closed time of the shutter was shorter than the grating lifetime under the illumination, a significant portion of the grating still remained in the crystal when the shutter was opened again. A sharp rise of I_s was observed due to the diffraction of the remaining grating, as illustrated by oscilloscope traces of I_s in two photos of Figure 1. Then, there was a slow increase of I_s following the sharp one. This was due to the increase of the grating amplitude when two beams interfered again in the crystal. When shutter closed time increased, the magnitude of the sharp rise decreased as expected (see difference between the two photos). Analysis indicated the amplitude of the sharp rise decreased exponentially with shutter closed time. Namely, the diffraction efficiency decays exponentially with time under illumination of I_s. Therefore, the diffraction efficiency decay time, defined as the time for the diffraction to decrease to the 1/e value of the original, can be measured. By changing the neutral density filter value, the grating lifetime can be measured as a function of I_s. Since the diffraction efficiency is proportional to the square of the amplitude of the index grating for small amplitudes (as under our experimental condition), the index grating lifetime is twice the measured diffraction efficiency decay time.

Figure 2 gives the measured grating lifetime as functions of I_s with four different values of grating spacing. The experimental result shows clearly the existence of two important features of the grating lifetime in GaAs:Cr. Firstly, the grating lifetime increases with the decrease of I_s and the rate of the increase becomes slower as I_s becomes lower than 1 mW/cm^2. This feature is consistent with reported anomalous observations of grating erasure rates in photorefractive oxides, such as $BaTi_3$, being proportional to optical intensity to a fractional power of between 0.5 to 1.0.[6] The second feature is that the grating lifetime increases with the decrease of grating spacing. A similar phenomenon was observed in InP.[7] The basic mechanism governing the complicated relationship among grating lifetime, reading beam intensity, and grating spacing in GaAs is currently under study. In this paper, only those experimental observations important to the operation of GaAs volume holographic elements are discussed.

The data in Figure 2 demonstrate that the lifetime of gratings with grating spacing being about 0.66 microns can be as long as 2.5 seconds when I_s is about 0.1 mW/cm^2. The lifetime of the same grating spacing reduces to about 0.4 seconds when I_s increases to about 10 mW/cm^2. The lifetime can be further reduced if larger grating spacing is created using smaller incident angle between two beams.

Our results demonstrate that information stored in volume holographic elements of GaAs:Cr can vary from 20 milliseconds to a few seconds, depending on I_s and grating spacing. Conceptually, the lifetime can be reduced further into the microsecond range using higher intensities. The availability of a large range of grating lifetime in GaAs:Cr provides excellent opportunities for using GaAs as real-time spatial light modulators, reconfigurable beam-steering devices, and dynamic memory elements in optical computing applications.

As illustrated in Figure 2, the rate of the increase becomes slower when intensity becomes lower. This could be due to that the photo-ionization rate gradually becomes comparable with the thermal emission rate of trapped carriers from the Cr level. The thermal emission rate is the ultimate physical phenomenon dictating the longest information storage time, namely the storage time in the dark.

In our experiments, the 1.15 micron beam was used to simulate the reading beam. If a beam of a 1.3 or 1.5 micron injection semiconductor laser is used as reading beam, both storage time and reading beam intensity can be increased. This is due to fact that the cross section for the photo-ionization at these wavelengths is much smaller. In addition, it is known that an application of electric field on the sample can increase the lifetime. These are among the subjects under study.

It is interesting to note that the index grating lifetime in photorefractive Fe-doped InP was reported to be only about several hundred microseconds.[7] The short index grating lifetime observed can be attributed to the fact that the energy bandgap of InP is smaller than that of GaAs.

References:

1. A.M. Glass, A.M. Johnson, D.H. Olson, W. Simpson, and A.A. Ballman, Appl. Phys. Lett., 44, 948 (1984).

2. M.B. Klien, Opt. Lett., 9, 350 (1984).

3. G.C. Valley, A.L. Smirl, M.B. Klien, K. Bohnert, T.F. Boggess, Opt. Lett., 10, 647 (1986).

4. L.J. Cheng and A. Partovi, Appl. Phys. Lett. (issue of November 24, 1986).

5. J. Strait and A.M. Glass, Appl. Opt., 25, 338 (1986).

6. G.V. Valley, J. Appl. Phys., 59, 3363 (1986).

7. J. Strait and A.M. Glass, J. Opt. Soc. Am. B, 3, 342 (1986).
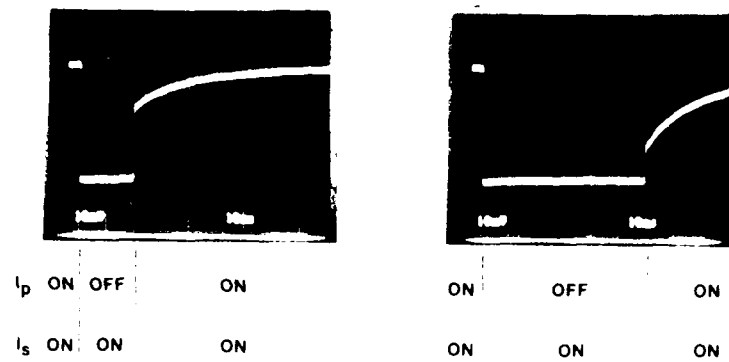
Figure 1. Two oscilloscope traces of Is with different closed
times of Ip. The Is signals were biased with a
DC voltage for clear demonstration of the effect of I_p
on I_s .
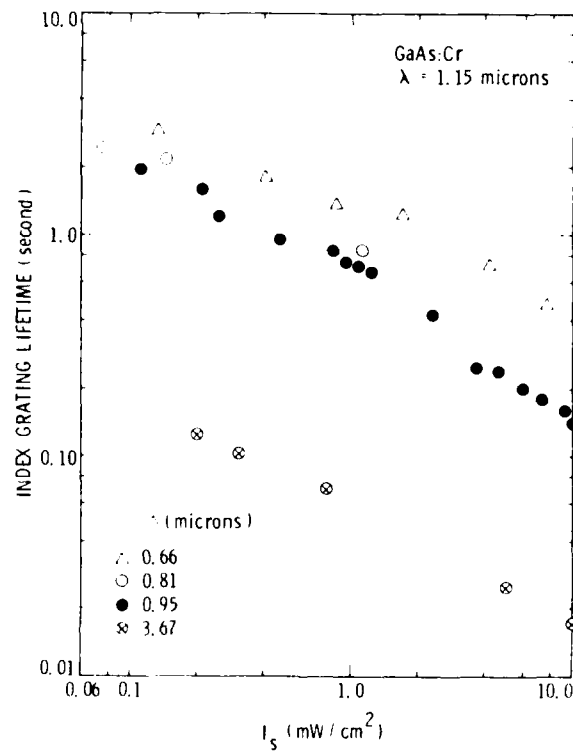
Figure 2. Measured grating lifetime in GaAs:Cr as
function of I_s intensity. ∧ is the grating
spacing in microns.

An Optical, Cross-Bar Arithmetic/Logic Unit

R. Aaron Falk, C. David Capps and Theodore L. Houk
Boeing Aerospace Company
P. O. Box 3999, MS 87-50
Seattle, WA 98124

As indicated in an earlier paper presented at this conference, [1] use of parallel Fourier optical pattern recognition techniques in conjunction with a final non-linear threshold allows rapid computation of sums and products in residue arithmetic. The coherence properties of the architecture reduce the number of non-linear elements to $2n-1$ where n is the size of the radix. Incoherent point sources could also be used with the grating filters performing a holographic interconnect function, but, at the expense of requiring n^2 non-linear elements. This expense is offset by allowing performance of any multi-level logic function, residue arithmetic being only one example of a multi-level logic function. This paper will describe an alternative technique to performing the desired n^2 interconnect pattern which requires no lenses or filters, thereby significantly reducing the fabrication and alignment difficulties.

Figure 1 illustrates the basic interconnect concept. The two inputs are arranged as in a truth table, with one set of inputs being the rows and the other being the columns. The desired inputs are turned on by filling the corresponding row and column with an equal intensity of light. At each intersection point either 0, 1, or 2 units of light will be present with only the single correct intersection having 2 units. A simple non-linear threshold device, which performs the equivalent of an AND operation, placed at each intersection will select the correct answer. Equivalent outputs would then be OR'd to form the final output. An important result of the multi-line (one-of-many) representation used in this architecture is the elimination of the INVERT operator for performing the logic.

There are many potential ways to construct the cross-bar architecture including the use of electronic AND and OR gates. The use of optical interconnects eliminates the dispersion difficulties associated with electronic wires and allows the possibility of high speed opto-electronic or all optical AND's along with simple "hardwired" OR's. As an initial demonstration of the concept, a fiber optic device, as shown in Figure 2, was constructed. Off-the-shelf LED communications modules were used for the inputs. One-to-three fiber optic couplers were used to route the light to the appropriate one of nine detector-threshold devices which perform the AND operation. Electronic OR's were used to combine the equivalent channels for performing residue 3 addition (as shown) and simultaneously residue 3 multiplication (not shown). Although it was not done in the actual device, the OR'd outputs could then be used to drive another set of LED's to form the input for the next stage.

The fiber optic device was tested by using two RAMS, one containing a pseudo-random sequence of inputs and the other containing the correct sums and products for residue 3 arithmetic. This latter set of data was used to check the operation of the optical channel. The device was operated at 50 Mop (limited by electronics) with a better than 10^{-11} error rate. In order to display the inputs and outputs the positional notation was converted to a

time notation using electronic parallel-to-serial converters. A typical output is shown in Figure 3.

Although the current demonstration of this computing architecture was performed in fiber optics, future examples will be based on integrated optics. One possibility is indicated in Figure 4. Because the logic is based on optical interconnects with a single, nonlinear threshold as the final step, scaling to higher speeds only involves improving the threshold process. Thresholds using current electronics technology (GaAs) can support several GHz operation and all-optical means (opto-electronic hybrids or optical bistable devices) can increase the speed to 10's and possibly 100's of GHz.

In conclusion, an optical arithmetic/logic unit based on optical interconnects has been demonstrated. Use of existing technology would allow the device to compete directly with the best electronic circuitry available to date and future, all-optical means should allow several orders of magnitude improvement.

(1) "An Optical Arithmetic/Logic Unit Based on Residue Number Theory and Symbolic Substitution", C. D. Capps, R. A. Falk and T. L. Houk, this conference.
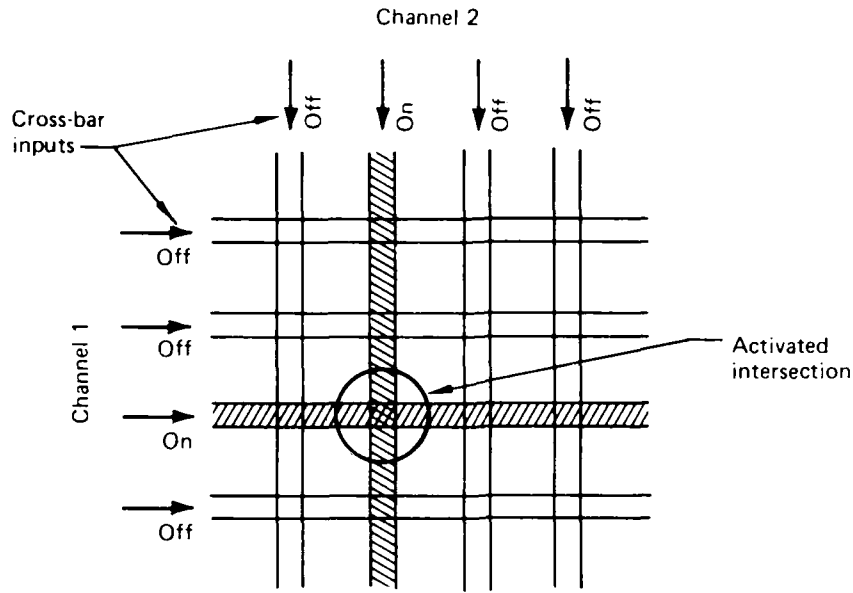
Figure 1. Optical Cross-Bar Concept
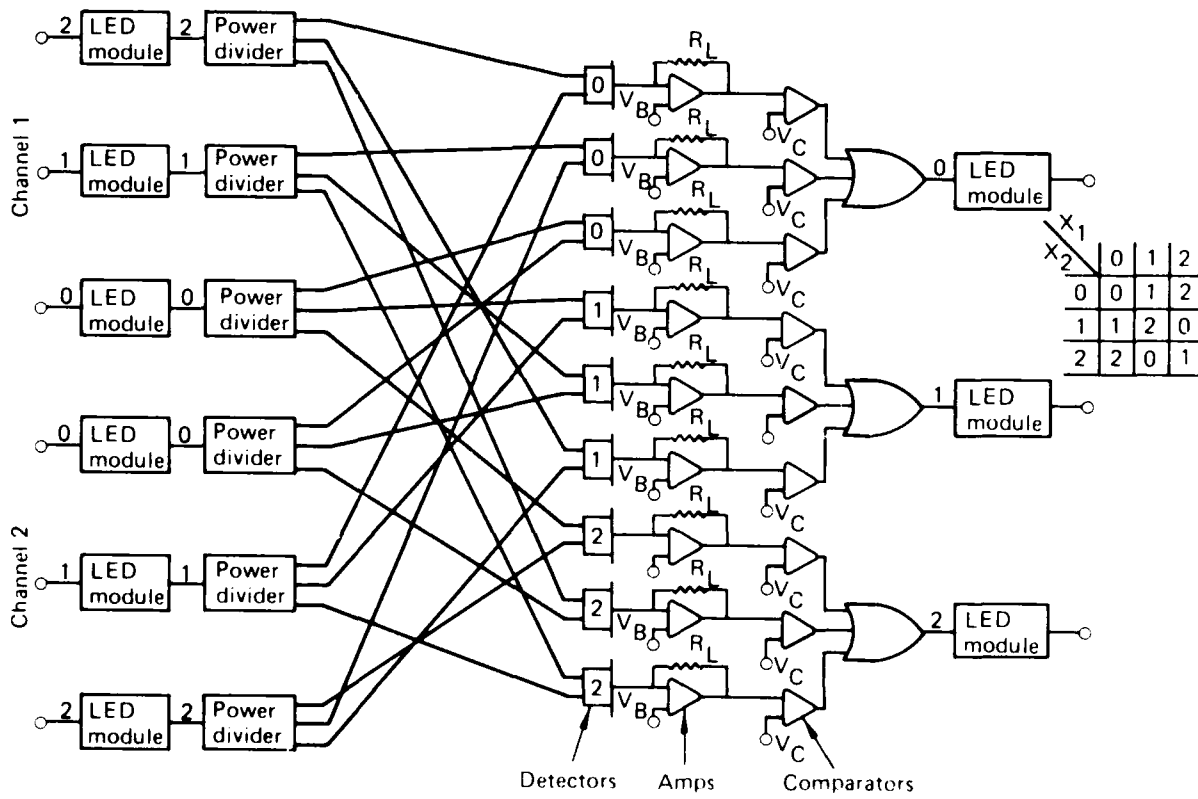
Figure 2. Fiber Optic Version of Optical Cross-Bar Residue 3, Adder
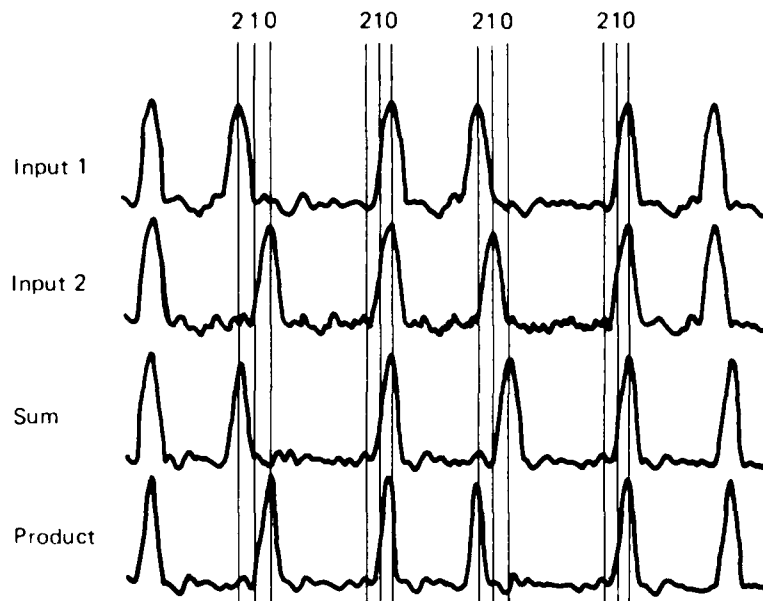
Figure 3. Tracing of Oscilloscope Output Showing Example Inputs to, and Sum and Product From, Device Shown in Figure 2. Operations are Performed on a 20 nsec Cycle.
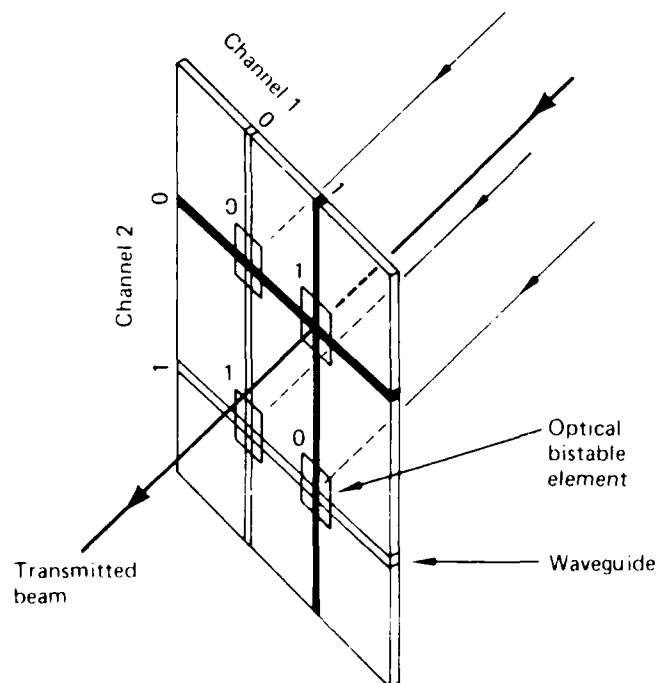
Figure 4. Example of Possible Cross-Bar Architecture in Integrated Optics. Function Shown is XOR.

Passive Single-Mode Optical Networks for Coherent Processing

M. E. Marhic

Department of Electrical Engineering and Computer Science
Northwestern University, Evanston, IL 60201

Introduction

We consider in this paper optical networks made entirely of passive single-mode interconnection components. These can be lengths of fiber, or integrated optic waveguides, in which the same kind of single mode propagates. These individual modes can also be added/divided by means of directional couplers interconnecting the waveguides. [Similar networks could also be considered, making use of free-space Gaussian beams and beam-splitters; this however, would introduce difficulties due to alignment and diffraction.] We refer to the resulting architectures as passive single-mode optical networks.

Some of these networks have been studied by others, under the assumption that the various signals are incoherent, i.e. add on a power basis when combined, and thus do not lead to observable interference phenomena[1,2]. From a practical standpoint incoherent operation is desirable because it avoids phase noise; however it does restrict the kind of operations which can be performed.

The object of this paper is to study some of the benefits which can accrue when single-mode networks are operated in a coherent manner. This leads for example to networks which can perform discrete Fourier/Hadamard transforms without any active components. These benefits, however, can only be obtained if precise phase control can be achieved; some practical implications of this fact will be examined.

Matrices associated with $N \times N$ directional networks

Consider the unidirectional single-mode network of Fig.1. It is assumed to work only with a single frequency and state of polarization, so that a complex scalar can be used to describe the field amplitude and phase at any location. There are N input ports and N output ports, so that the network constitutes an $N \times N$ directional coupler. The fields injected at the inputs, a_i, are all coherent, being derived from the same monochromatic source, and so the output fields, b_j, result from the interference between the input signals as they are split and superimposed by the elements making up the $N \times N$ coupler. Since all operations performed within the network are linear, we have a discrete linear transform between inputs and outputs, i.e.

$$b_j = \sum_{i=1}^{N} C_{ji} a_i, \quad j = 1, \ldots, N, \tag{1}$$

where the C_{ji}'s are constants determined by the internal structure of the network. Since total output power cannot exceed total input power, any given set of C_{ji}'s must satisfy

$$\sum_{k=1}^{N} \sum_{l=1}^{N} a_k a_l^* \sum_{j=1}^{N} C_{jk} C_{jl}^* \leq \sum_{i=1}^{N} a_i a_i^* \tag{2}$$

for all possible choices of a_i's. The equal sign holds only for lossless networks.

Let $[C]$ denote the matrix of elements C_{ji}. The eigenvalues of $[C]$ must have magnitudes smaller than or equal to unity. This implies that

$$|det[C]| \leq 1, \tag{3}$$

the equal sign holding only for lossless networks.

For lossless $N \times N$ networks, Eq.(1) implies furthermore that $[C]$ is unitary, i.e. that

$$\sum_{k=1}^{N} C_{kl} C_{km}^{*} = \delta_{lm}, \quad \text{or} \quad [C]^{-1} = [C]^{\dagger}, \quad \text{or} \quad ([C]^{-1})_{kl} = C_{lk}^{*}. \tag{4}$$

If a lossless network is reciprocal, then $[C]^{-1}$ describes the physical operation of the network in reverse, and as such it too must satisfy the above relations. This imposes the additional condition

$$\sum_{l=1}^{N} C_{kl} C_{jl}^{*} = \delta_{kj}. \tag{5}$$

There are many types of matrices which satisfy the above conditions, and hence many types of discrete transforms performed by unidirectional coupling networks. In the following we study some particularly interesting transforms.

Discrete spatial Fourier or Hadamard transforms by single-mode star networks

Single-mode star networks were recently introduced to provide even distribution of the power of any one input among all outputs. These networks are made from suitably interconnected 2X2 couplers.[3,4] They have been implemented in both fiber optic and integrated optic form. The interconnection patterns used are similar to those used in VLSI arrays to perform operations such as the fast Fourier transform, and this analogy led to the question of whether these optical networks themselves could perform discrete Fourier transforms (DFTs), without requiring any active processors. It has recently been shown that this is indeed the case.[5] To see this, consider an ideal lossless star network with $N = 2^n$; it is an evenly-dividing $N \times N$ network, i.e. it is such that

$$C_{kl} = \frac{1}{\sqrt{N}} e^{i\phi_{kl}}, \tag{6}$$

where the ϕ_{kl}'s are arbitrary (real) phase angles. Potentially interesting forms for Eq. (1) result if the ϕ's are chosen in particular ways. For instance if

$$\phi_{kl} = \frac{2\pi kl}{N}, \tag{7}$$

Eq. (1) corresponds to a discrete Fourier transform. It can be shown that the corresponding C_{kl}'s satisfy Eqs.(3)-(5), and that they can be physically realized, simply by appropriately adjusting the optical path lengths between the various components in the network.[5] The proof relies upon the fact that an elementary 2X2 coupler itself performs an elementary discrete Fourier transform; by combining such elements with appropriate lengths of waveguides (phases), and suitable interconnection patterns (such as the perfect shuffle), it is possible to build up a large Fourier-transform network in a hierarchic manner.

Star networks can also perform the Hadamard/Walsh transform, which is used extensively in image processing. In that case the ϕ_{kl}'s, and hence the phase shifts introduced by the interconnecting paths, need only be equal to 0 or π.[5]

The preceding shows that single-mode interconnection networks, although entirely passive, can be made to perform sophisticated signal-processing operations when used in a coherent fashion. This suggests that they might play a direct role in some high-speed operations, such as DFTs, thus assuming a primary fast processing role. This function would go beyond the currently-envisioned ancillary role for optical networks, namely that of providing fast interconnections among digital processors.[6]

Practical considerations

The major obstacle to the implementation of the above schemes is in setting, and maintaining, phases to their desired values. The degree of difficulty will increase as the size of the envisioned networks increases, since each phase will have to be set more accurately. Networks made from segments of fibers will be more prone to phase noise than integrated optic networks. Phase setting/adjusting can be achieved by suitable means with these various technologies; some solutions, however, may turn out to be expensive. [A technique has been proposed to manufacture monolithic $N \times N$ stars from thin, uniform slab waveguides.[7] Such devices should exhibit excellent phase stability. They implement some discrete transform, however it is neither Fourier nor Hadamard, and they cannot be adjusted to perform desired transforms.]

Phase noise will be less of a problem at longer wavelengths, and it may thus be desirable to consider implementations at infrared wavelengths, or beyond, particularly in the initial stages.

Conclusion

Single-mode networks operating in a coherent manner can perform a number of functions not achievable with incoherent networks. Among these are discrete Fourier or Hadamard transforms by star networks. The practical utilization of these features will require the ability to accurately set and maintain phase. This will initially be best accomplished with small-scale networks, and/or at long wavelengths.

References

1. B. Moslehi, J. W. Goodman, M. Tur, and H. J. Shaw, Proceedings of IEEE **72**, 909 (1984).

2. K. P. Jackson, S. A. Newton, B. Moslehi, M. Tur, C. C. Cutler, J. W. Goodman, and H. J. Shaw, IEEE Transactions on Microwave Theory and Techniques **MTT-33**, 193 (1985).

3. M. Marhic, Optics Letters **9**, 368 (1984).

4. M. Marhic, in Proceedings of FOC/LAN'84, Las Vegas, NV (Information Gatekeepers, Boston, 1984).

5. M. Marhic, Optics Letters, to be published.

6. J. W. Goodman, F. J. Leonberger, S.-Y. Kung, and R. A. Athale, Proceedings of IEEE **72**, 850 (1984).

7. Th. Niemeier and R. Ulrich, Optics Letters **11**, 677 (1986).

Figures
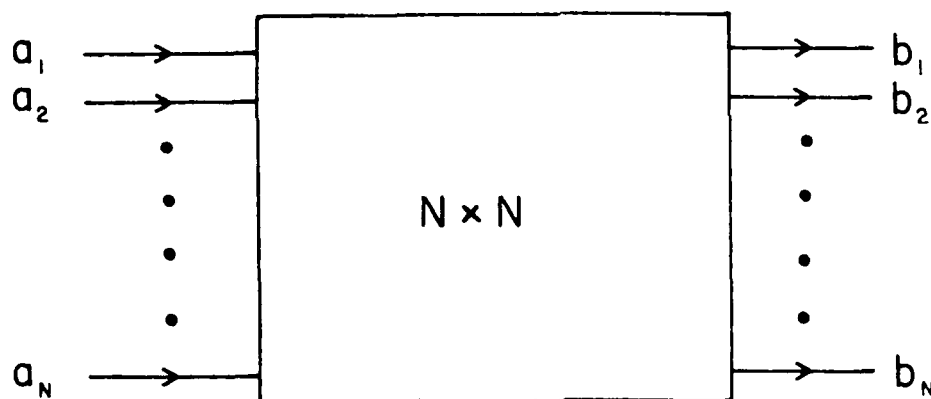
Fig.1: An $N \times N$ unidirectional coupling network.

Review of Some Current Optical Computing Research
in the Soviet Union

William T. Rhodes
Georgia Institute of Technology
School of Electrical Engineering
Atlanta, Georgia 30332

This past July, along with eight other scientists from western countries, I
attended a meeting on optical computing held in Novosibirsk. The meeting was
sponsored and hosted by the Institute of Automation and Electrometry, USSR
Academy of Sciences, Siberian Branch, under the direction of Academician Yu. E.
Nesterikhin.

A number of interesting papers on current efforts in optical signal processing
and optical computing were presented by Soviet attendees at the meeting. The
purpose of my talk is to summarize selected papers from that collection that
are particularly appropriate for the OSA Topical Meeting on Optical Computing.
This is done only in the absence of any Soviet representation at the Topical
Meeting.

Included will be discussions of research on self-switching of light in
tunneling-coupled optical waveguides, pipe-line opto-electronic processors,
digital computers based on integrated optics, pattern recognition using
symmetry features, and some thoughts presented on optical computing generally.

TUESDAY, MARCH 17, 1987

PROSPECTOR/RUBICON ROOM

8:00 PM–9:30 PM

TuE1–9

POSTERS: SESSION 10

Switch Power Drift in Optically Bistable ZnSe Interference Devices

R.J. Campbell, J.G.H. Mathew, S.D. Smith and A.C. Walker

Dept. of Physics, Heriot-Watt University, Edinburgh EH14 4AS, U.K.

Optically bistable devices based on nonlinear thin-film interference filters have been reported by Smith et al [1] and Weinbe.ger et al [2] and have been shown to be suitable for use as logic elements in digital optical circuits [3] and pattern recognition applications [4]. These devices use ZnSe or ZnS as the active material and rely on thermally induced changes in the refractive index to provide the required nonlinear response.

Early experiments with optically bistable interference filters showed highly irreproducible input-output characteristics [5,6]. In particular switch-up powers increased rapidly with time and the bistable region was seen to expand on each scan of input power. We have observed similar effects with nonlinear filters grown on low temperature substrates and consequently made up of less dense thin-film layers. The changes in operating characteristic are consistent with a decrease in refractive index of the spacer layer and a resultant increase in detuning between the operating wavelength and the low-power band-pass peak. A likely explanation is that heating induces desorption of water previously taken up within the pores of the low density material, thus reducing the average refractive index for the layer.

This effect can be avoided, or at least greatly reduced, by using an elevated substrate temp. (150-200°C) when growing the films. However under certain conditions an irreversible drift of the device characteristics occurs in the opposite manner, such that the switch-up power and the width of the bistable region are reduced. We describe experimental measurements of this latter phenomenon; the conditions under which it occurs and the techniques by which it may be minimised.
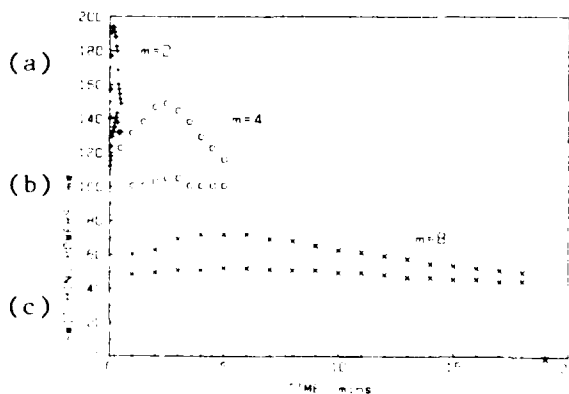
Fig. 1 Variation of switch powers with time for a 13 layer ZnSe spacer filter with spacer thicknesses of order (a) M=2 (b) M=4 and (c) M=B. Upper traces are switch up powers lower traces are switch down powers, a spot diameter of 60 nm was used in each case.
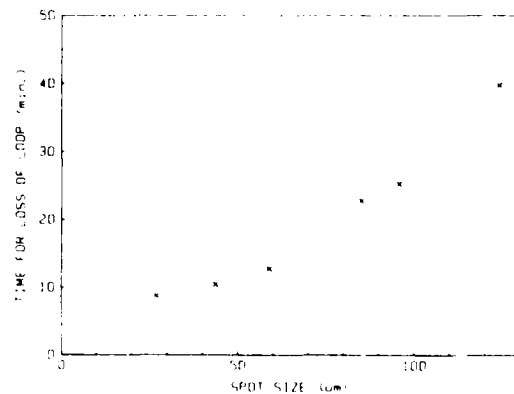
Fig. 2. Dependence of time for loss of bistability with incident spot diameter.

Figure (1b) shows how the switch powers can vary with time for a 13 layer ZnSe-spacer filter. The filter was held in its high transmission state near the switch-up power and the switching powers measured at regular intervals by ramping the incident power down and up in 10 seconds. As can be seen, the variation in the switch-up power was larger than the change in switch-down power. The initial rise in the switch up power increased the width of the bistable loop and is consistent with a slight increase in the filter detuning as described above. This trend then reversed so that eventually the switch-up and switch-down powers became equal and bistability was not observed. This induced change appeared permanent and no recovery was observed even after leaving the filter unilluminated for many hours. However the bistable characteristic was recoverable by increasing the detuning. This is consistent with the refractive index of the spacer layer having increased and could be explained by some form of structural change occurring, rather than the desorption of volatile inclusions.

The rate of switch power drift was a maximum when high input irradiances were being used. In addition, drift effects were much more apparent when the devices were held long-term in their high transmission (switched-up) state. This result suggests that the high temperature and/or the internal irradiance when in the high transmission state is responsible for the drift of the characteristics of the device. To distinguish these effects we have investigated how the time for loss of bistability varies with incident spot size for a fixed detuning, using the 13 layer ZnSe filter. For the same detuning the operating temperature at the held (switched-up) level must be the same for each spot size. However, as we have shown [7], the switch power is directly proportional to the spot diameter and thus the switch-up irradiance increases as the spot size is reduced. Figure (2) shows the results of this experimental study. As can be seen the device stability decreases with the spot size. It follows that the mechanism underlying the changes in device characteristics is related to the internal irradiance and may be associated with, for example, photo-structural effects.

It is clear that the problem of switch-power drift could be solved by employing more structurally stable materials for their fabrication. However, it is important to determine to what extent the very convenient growth technology of thermal evaporation can continue to be exploited. To maximise the stability of the present devices it is apparent, from the above, that the operating internal irradiances must be minimised. One method of doing this is to increase the thickness of the spacer. This has the effect of reducing the temperature rise required for switching and therefore lowers the internal irradiance necessary. Figure (1) shows the effect on stability of such an increase in spacer thickness. The three filters were identically structured but of different order where m is the spacer optical thickness in half wavelengths. As the spacer thickness is increased it can be seen that the drift effects are dramatically reduced.

Further details of these measurements will be presented together with some initial results of using alternative growth techniques to fabricate these nonlinear multilayer structures.

References

[1] S.D. Smith, J.G.H. Mathew, M.R. Taghizadeh, A.C. Walker, B.S. Wherrett and A. Hendry,
Opt. Commun. 51, 357 (1984).

[2] D.A. Weinberger, H.M. Gibbs, C.F. Li and M.C. Rushford,
J. Opt. Soc. Am. 72, 1769 (1982).

[3] S.D. Smith, A.C. Walker, F.A.P. Tooley, J.G.H. Mathew and M.R. Taghizadeh,
"Optical Bistability III" Ed. H.M. Gibbs, P. Mandel, N. Peyghambarian (Springer-Verlag, Berlin, 1986) p8.

[4] R. Jin, L. Wang, R.W. Sprague, H.M. Gibbs, G.C. Gigioli, H. Kulcke, H.A. Macleod, N. Peyghambarian, G.R. Olbright and M. Warren,
In Opt. Bistability III, eds. H.M. Gibbs, P. Mandel, N. Peyghambarian and S.D. Smith p61 (1986).

[5] G.R. Olbright, N. Peyghambarian, H.M. Gibbs, H.A. MacLeod,
Appl. Phys. Lett. 45, 10 (1984)

[6] F.V. Kaspusko and G.V. Sinitsyn,
J. Appl. Spectroscopy (USSR) 29, 1323 (1978).

[7] I. Janossy, M.R. Taghizadeh, J.G.H. Mathew and S.D. Smith,
IEEE J. Quant. Electr. QE-21, 1447 (1985).

THERMO-OPTICAL BEAM GUIDE AND
SWITCHING EXPERIMENTS*

Lauren M. Peterson
Environmental Research Institute of Michigan
Advanced Concepts Division
P.O. Box 8618
Ann Arbor, MI 48107

Radiation passing through a nonlinear optical medium can lead to a change in the refractive index of that medium. If the radiation is in the form of a focused Gaussian beam, the focal region where the radiation field is the strongest will have the greatest index change and the transverse spatial profile of this index change will be bell-shaped since the transverse radiation profile is Gaussian. For a large f-number focused beam, the focal region is a long cylinder with a transverse graded-index (GRIN) profile quite analogous to a graded index waveguide or fiber. We have used this real-time refractive index waveguide due to one beam of pulsed laser radiation to rapidly redirect or switch a second beam of laser radiation.

The material media in which we have observed real-time beam guiding are liquids to which dyes have been added to make the liquid optically dense. The refractive index changes have been induced by absorption of the radiation followed by rapid ($\sim 10^{-10}$ sec [1]) thermalization of the energy. The change in temperature of the liquid leads to the change in its refractive index (presumeably due to local changes in density).

---

[1] K.F. Herzfeld and T. A. Litovitz, Absorption and Dispersion of Ultrasonic Waves, Academic Press, N.Y. (1959).
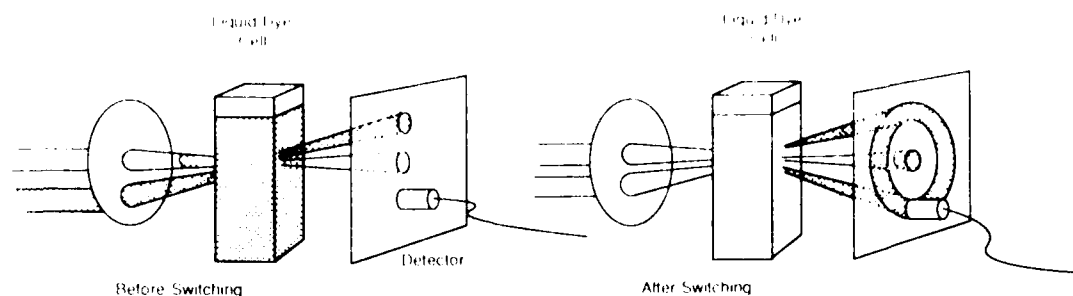
Figure 1. Experimental arrangement

Figure 1 shows the experimental arrangement for observing real-time beam guiding. A pulsed nitrogen laser pumped dye laser using coumarin 500 dye provides 500 nm pumping radiation for inducing the refractive index waveguide. This radiation is focused by a lens to produce a focal region approximately $500\mu m$ long and $12\mu m$ in diameter (full-width at half intensity). HeNe laser radiation at 633 nm serves as the probing beam and is also focused by the lens. In the absence of any absorber in the liquid sample, the two beams simply cross in the focal region. A screen placed behind the sample cell displays the two unaltered beams as shown in Fig. 2a. In the presence of significant absorption (> 0.1 cm^{-1} at 500 nm), the HeNe beam is guided by the refractive index induced by the green laser pulse. Guiding of the HeNe radiation in the cylindrical focal volume redistributes the radiation into a cone such that an annulus of light is displayed on the screen as shown in Fig. 2b. Separation of the switched radiation from the switching radiation is easily accomplished with a spectral filter.
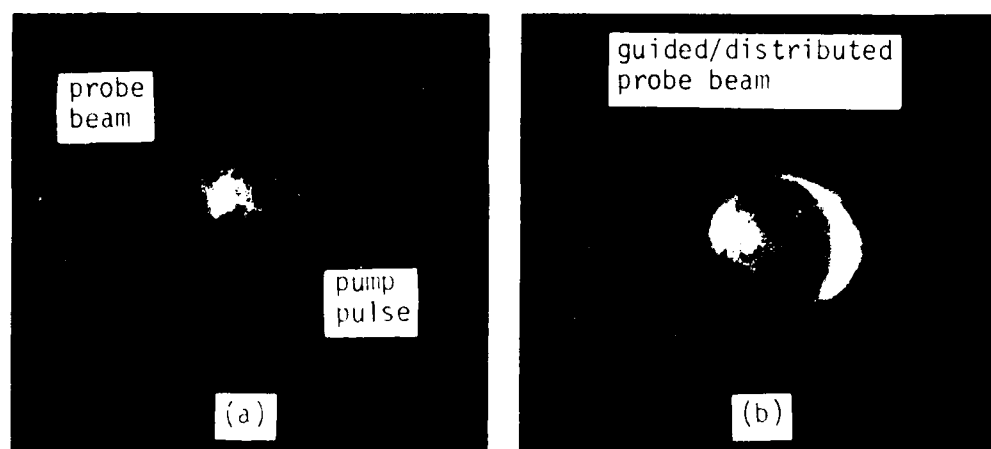
Figure 2. (a) No beam guiding in the absence of absorber.   (b) Beam guiding in absorbing liquid.

The optically induced beam guiding has been demonstrated in our laboratory using carbon disulfide, carbon tetrachloride, acetone and methanol as solvents with iodine, eosine and cobalt nitrate as dyes. The dye laser pump pulses were 7 nsec in duration with energies on the order of 10μJ or less. Switch-on time for the guided HeNe probe beam was on the order of 10 nsec. It is not known at this time whether this switching time is characteristic of the medium and the thermalization process, or if it simply represents the deposition time of the laser pulse and therefore would decrease if the laser pulse were shorter.

Once the probe beam has been switched, it remains in the switched state long after the pump radiation is gone. The amount of probe radiation which is guided or, the efficiency of the interaction decreases exponentially with time following the pump pulse and has a time constant on the order of a millisec as shown in Figure 3. This is on the order of the thermal diffusion time for liquids [2] and represents the equilibration of the temperature gradient in the focal region.
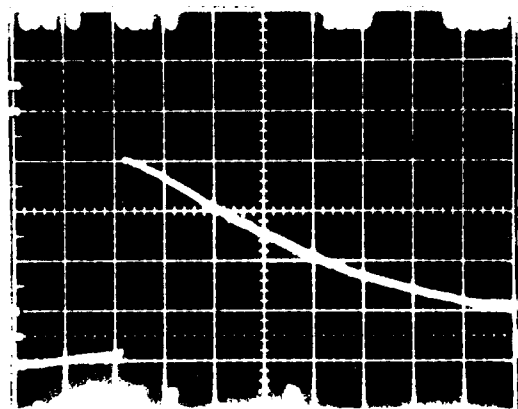
Figure 3. Beam guide temporal response. 10 nsec switch-on; exponential decay; 500μs/div.

[2]   R. L. Carman, P. L. Kelley, Appl. Phys. Lett. 12 (1968) 241 and H. Eichler, G. Salje, H. Stahl, J. Appl Phys. 44 (1973) 5383.

The efficiency of the beam guiding was observed to be quite high. The guided probe radiation which was distributed into an annulus of light was gathered and focused by a lens onto a fast Si PIN detector. Pump radiation was blocked using a red filter and unguided light was blocked using a spatial stop. The efficiency was found to be the greatest (93%) for CS_2 with an absorption coefficient of about 3 cm⁻1. The efficiencies for the other solvents and dyes were in the 10% to 70% range, being lowest for the lowest laser energies.

In summary, we have observed optically induced beam guiding or switching at energy levels in the microjoule range. Switch-on times are on the order of 10^{-8} sec or shorter and the switched beam can persist (i.e. memory) for 10^{-3} sec or more. It is anticipated that a second pump or control beam at a different angle could be used to switch-off the signal or probe beam in times comparable to switch-on. Although our experiments utilize pump (control) and probe (signal) beams at different wavelengths, the interaction is expected to be identical for degenerate wavelengths leaving open the likelihood of cascadable switches. Not only does the reported interaction lead to rapid switching of an optical beam, but also to its redirection. This may prove to be of value in all-optical beam control, or in optical interconnects for all-optical and hybrid optical/electronic computers.

FERROELECTRIC LIQUID CRYSTAL SPATIAL LIGHT MODULATORS

D. Armitage, J. I. Thackara
Research & Development Division

LOCKHEED MISSILES & SPACE COMPANY, INC.
3251 Hanover Street, Palo Alto, California 94304

N. A. Clark, M. A. Handschy
Displaytech, Boulder, CO 80306

The ferroelectric liquid-crystal (FLC) phase is a relatively new development in liquid crystals. The basic symmetry restrictions demanded by ferroelectricity are met by the chiral smectic-C liquid-crystal phase. The chirality or lack of mirror symmetry at molecular level gives rise to a spiral polarity or helioelectric structure. The helix can be unwound by an applied electric field to create a uniform FLC. However, this is not the most favorable device arrangement.

The optimum device configuration maximizes polarization advantage, while minimizing the viscoelastic retardation. This is achieved in the surface-stabilized ferroelectric liquid-crystal (SSFLC) configuration illustrated in Fig. 1. The surfaces in contact with the FLC are treated to promote uniform parallel alignment. For a sufficiently thin cell, the surface forces suppress the natural helix of the FLC. Polarity switching is accompanied by molecular rotation in the smectic plane, where viscoelastic retardation is minimum. The molecular tilt angle is engineered at the materials synthesis stage to give an optic axis rotation approaching 45 deg. In a crossed polarizer configuration, the cell switches between dark and bright according to the polarity of the applied voltage. With proper surface preparation, bistable operation is achieved and either state can be stored indefinitely at zero applied voltage. The SSFLC device retains the low voltage and power advantage of the nematic liquid crystal (NLC), but the first-order interaction of the net dipole moment with the applied electric field greatly increases the bidirectional switching speed.[1]

The storage property of the SSFLC device and well-defined threshold voltage favor passive large-scale X-Y matrix-addressing. Devices have been fabricated which demonstrate 640 x 400 addressable pixels.[2] The obvious applications in the display industry have fueled a rapid expansion in FLC material and device work in recent years. Room-temperature materials are now available with switching speeds of order 10 μs.[3] Continuing materials development is now pushing the switching speed towards 1 μs.

The evolving FLC display technology is clearly relevant to the development of optical-processing devices. The high resolution inherent to the SSFLC device is of particular interest. Pixel scale of 17 μm is readily demonstrated and the resolution limit is determined by ferroelectric domain wall thickness of
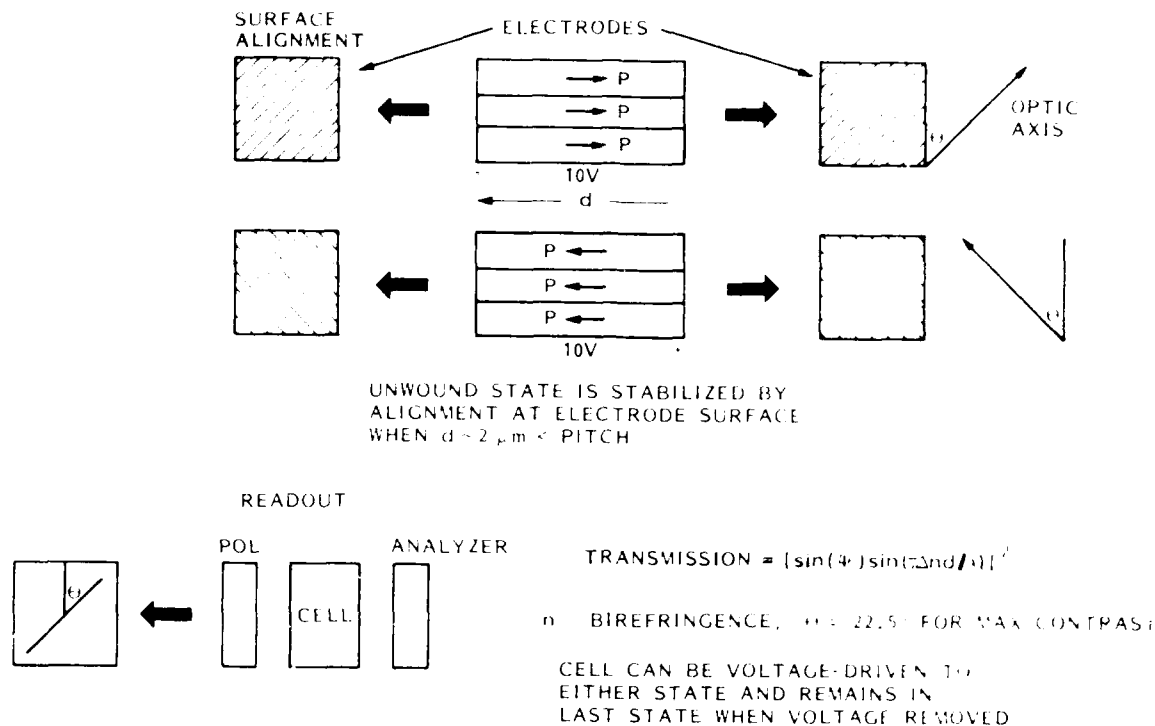
SURFACE
ALIGNMENT   ELECTRODES

OPTIC
AXIS

10V

d

10V

UNWOUND STATE IS STABILIZED BY
ALIGNMENT AT ELECTRODE SURFACE
WHEN d ~2 μm < PITCH

READOUT

POL   ANALYZER

CELL

$TRANSMISSION = [\sin(4\theta)\sin(\pi \Delta n d/\lambda)]^2$

n   BIREFRINGENCE, θ = 22.5° FOR MAX CONTRAST

CELL CAN BE VOLTAGE-DRIVEN TO
EITHER STATE AND REMAINS IN
LAST STATE WHEN VOLTAGE REMOVED

Fig. 1   Surface stabilized ferroelectric liquid crystal device.

1 μm.[4]   However, matrix structures are inherently line-by-line-addressed
devices, which is a severe restriction on frame rate.

Photoaddressed FLC devices have been neglected in comparison with matrix
electrode addressing. In the photoaddressed FLC spatial light modulator
(SLM), the addressing is inherently parallel and the full speed of the SSFLC
device is attainable. We have provided the first demonstration of a
photoaddressed SSFLC device.[5] Figure 2 shows the device configuration
employing bismuth silicon oxide (BSO) as the photoconductive addressing
medium. FLC alignment was achieved with the rubbed-nylon technique.[5]

Figure 3 compares a refreshed Air Force resolution chart image with the same
image stored for 15 h. Flaws in the refreshed image are associated with
alignment imperfections. The stored image is seen to have deteriorated over
the 15 h period. This is again associated with the current limitations in our
alignment technology.

The BSO addressing was used in the initial demonstration because of the
simplicity in device structure. The response speed is limited by detrapping
times in the BSO to frame rates of order 10 Hz. Further development employs
single-crystal silicon photodiode addressing which will enable full assessment
of the FLC response time. Preliminary experiments show that the SSFLC can be
addressed with a single-crystal silicon structure.
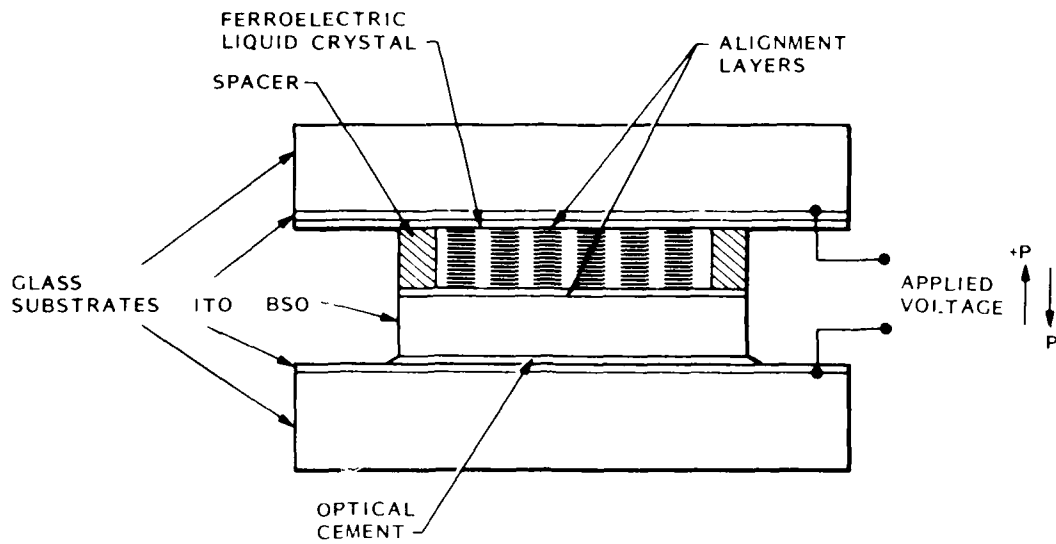
Fig. 2.   Bismuth silicon oxide photoaddressed SSFLC SLM.
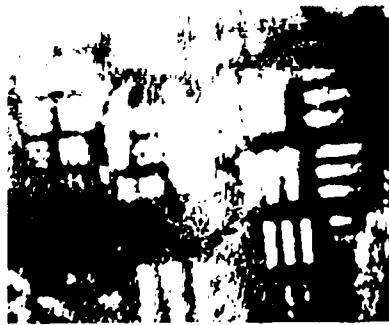
REFRESH

OVERNIGHT STORAGE

Fig. 3.   SSFLC SLM showing refreshed image and stored image.

The performance of the silicon SSFLC SLM using a revised alignment approach will be presented.[5] Current developments in FLC materials and devices will be discussed.

REFERENCES

1. N. A. Clark. S. T. Lagerwall, App. Phys. Lett., Vol. 36, p. 899, 1980.

2. T. Harada, M. Taguchi, K. Iwasa, M. Kai, SID Digest, p. 131, 1985.

3. Catalogue, Displaytech Inc., P. O. Box 7246, Boulder, CO 80306.

4. N. A. Clark, M. A. Handschy, SPIE, Vol. 625, p. 60, 1986.

5. J. S. Patel, T. M. Leslie, J. W. Goodby, Ferroelectrics, Vol. 59, p. 137, 1984.

6. T. Uemura, N. Ohba, N. Wakita, H. Ohnishi, I. Ota, SID Digest (Japan), p. 464, 1986.

Theory of All-Optical GaAs Logic Devices

M.E. Warren, S.W. Koch and H.M. Gibbs

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

All-optical semiconductor devices are numerically modeled and optimized using a microscopic theory for the optical nonlinearities of room-temperature GaAs. Single-frequency NOR-gate operation in reflection is predicted.

Numerical simulations of all-optical room-temperature GaAs devices are presented and the operation of a single-frequency NOR-gate in reflection is predicted. The computations are done using the equation for the transmission of a nonlinear Fabry-Perot resonator of length L, in which the space between the mirrors is filled with the semiconductor material. The transmitted intensity is

$$I_t = \frac{I_0 (1-R)^2}{(e^{\alpha(\omega,N)L/2} - R\, e^{-\alpha(\omega,N)L/2})^2 + 4R\,\sin^2(\delta + \omega\Delta n(\omega,N)L/c)} , \qquad (1)$$

where I_0 is the input intensity, δ is the linear phase shift (detuning), and R is the mirror reflectivity. The transmitted light intensity is coupled to the electron-hole-pair density N in the semiconductor via the absorption coefficient $\alpha(\omega,N)$ and via the nonlinear part of the refractive index $\Delta n(\omega,N)$. The density N in turn is related to the intensity I of the light inside the resonator by the equation

$$\frac{dN}{dt} = -\frac{N}{\tau} + \frac{\alpha(\omega,N)}{\hbar\omega} I \qquad (2)$$

where τ is the carrier relaxation time. A microscopic plasma theory[1] is used to consistently describe the nonlinear absorption and dispersion of room-temperature GaAs. Spectroscopic studies have shown the calculated values of absorption α and refractive index changes Δn to be in close agreement with experiment[2].

Equations (1) and (2) are solved for pulsed and steady-state excitation. The numerical results for the transmitted intensity as a function of the input intensity are plotted in Figs. 1a - 1c for triangular input pulses at a frequency ω well below the exciton resonance and pulse widths of 0.1, 1, and 10 μs, respectively, together with the corresponding cw characteristics (Fig. 1d). The different curves (1-4) in Figures

1a-1d show hysteresis curves obtained for slightly different resonator lengths, corresponding to different detunings of the excitation frequency with respect to the near-
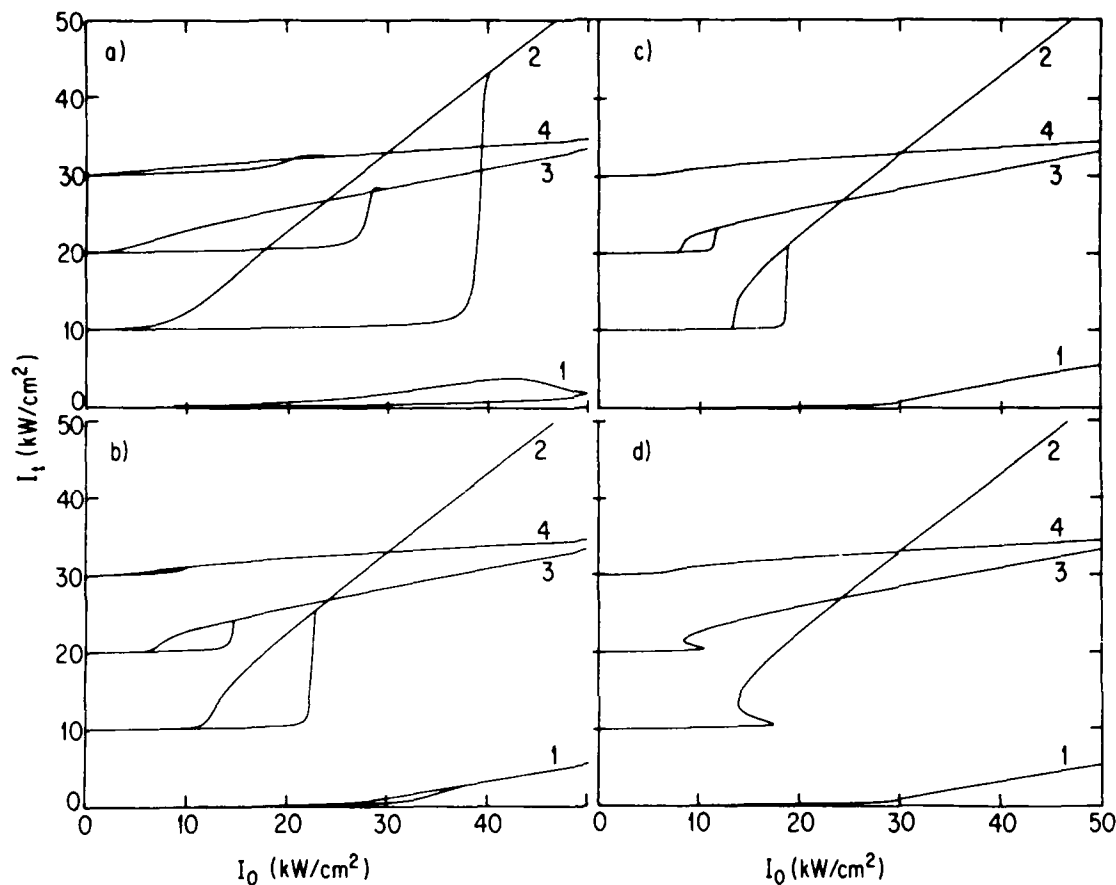
Fig. 1 Transmitted intensity versus input intensity computed for GaAs at room temperature at an excitation energy $\hbar\omega = 1.4032$ eV well below the exciton resonance at 1.420 eV. Figs. 1a - 1c are obtained assuming pulsed excitation with a triangular pulse of full width 0.1 μs (a), 1.0 μs (b), and 10. μs (c). Fig. 1d shows the steady state results. The different curves 1 - 4 in each figure are for the respective resonator lengths $L = 2.046 \mu m$, $2.042 \mu m$, $2.038 \mu m$, and $2.034 \mu m$, causing different resonator eigenfrequencies ω_R which give rise to the detunings $\Delta\hbar\omega = \hbar\omega_R - \hbar\omega = -0.0170$ eV, -0.0142 eV, -0.0115 eV, and -0.008 eV, respectively. The mirror reflectivity $R = 0.9$ and the carrier relaxation time has been taken as $\tau = 10$ ns. The baseline for the transmitted intensity in curves 2, 3, and 4 has been shifted by 10, 20, 30 kW/cm^2, respectively.

est resonator eigenfrequency ω_R, where $\omega_R < \omega$. The transmission characteristics 2 and 3 display well developed bistable loops similar to those observed in experiments under the assumed operating conditions. These loops get wider for shorter pulses

due to dynamical hysteresis, and they show the interplay between the dispersive tuning of the resonator and the saturation of the absorbing medium. In curve 2, the initial detuning of the resonator is such that the peak transmission of the Fabry Perot resonator nearly coincides with the excitation frequency at the same intensity I that saturates the semiconductor absorption. This mixed dispersive and absorptive behavior leads to a much higher transmission of the device with the price of some-what increased switch-on power. As is well-known[3] the dynamical hysteresis effect can even produce seemingly bistable behavior, as in curves 1 and 4, which vanishes for longer pulses.

The steady-state switch-on intensity for bistable operation is recorded together with the transmitted intensity just after switch on for different resonator lengths (Fig. 2). Within each bistable regime there is a definite length for which the ratio of the transmitted intensity to the switch-on intensity is a maximum, indicating opti-mum contrast. This optimum occurs for bistable characteristics of the type shown in
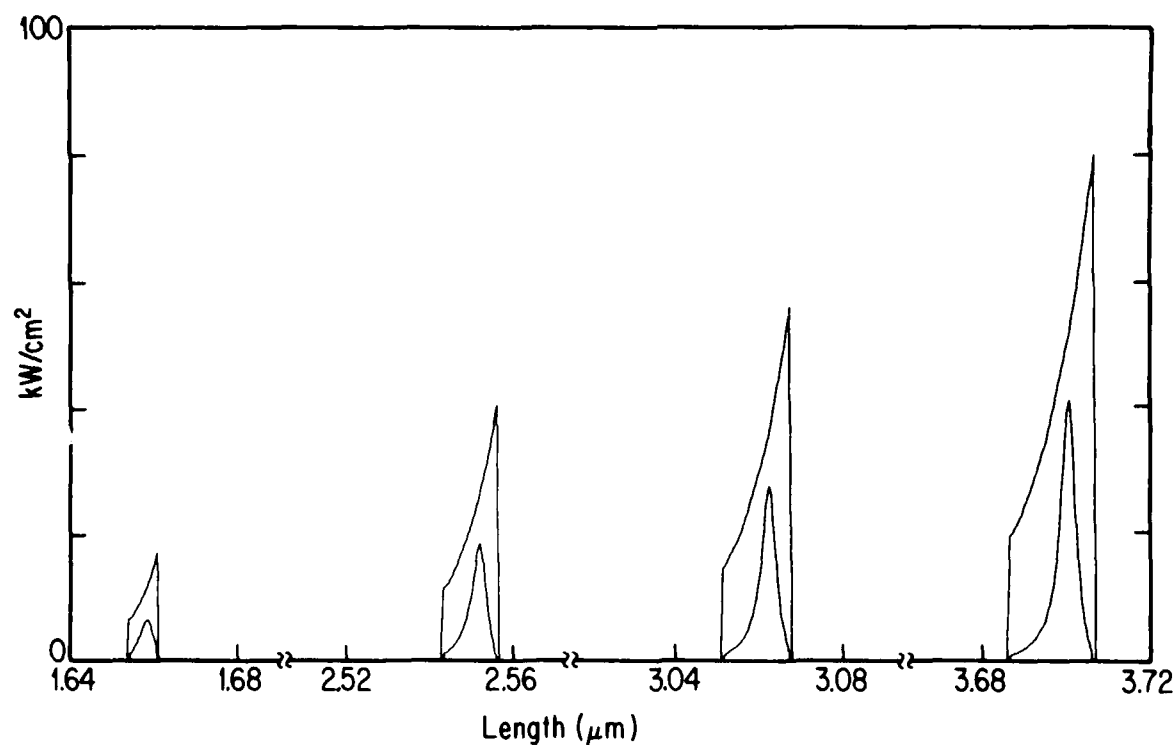
Fig. 2   Steady-state results for bistable transmission characteristics: The input intensity at switch on to high transmission (upper curves) and the transmitted intensity just after switch on (lower curves) are plotted versus resonator length for the same parameters specified in Fig. 1.

curves 2 of Fig. 1. In Fig. 3 the reflection hysteresis loop is plotted which corresponds to the transmisson characteristics 2 in Fig. 1. The highly transmitting upper branch causes very low reflection for high input intensities and thus makes single-frequency NOR-gate operation possible. A probe beam of an intensity up to 18 kW/cm^2 is largely reflected in the absence of other input, but the presence of one or two additional beams induces switch down to low reflection.
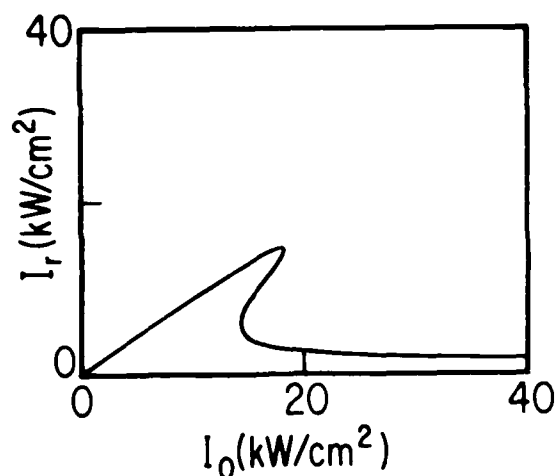
Fig. 3 Reflected intensity versus input intensity for the same parameters as curve 2 in Fig. 1d.

It is evident that the presented simulations are useful for modeling and optimizing devices for specific operating requirements. Work is continuing to study dynamic operation as well as waveguide design.

1. L. Banyai and S.W. Koch, Z. Physik. **B63**, 283 (1986)
2. Y.H. Lee, A. Chavez-Pirson, S.W. Koch, H.M. Gibbs, S.H. Park, J. Morhange, N. Peyghambarian, L. Banyai, A.C. Gossard, and W. Wiegmann, Phys. Rev. Lett. , accepted for publication
3. H.M. Gibbs, Optical Bistability - Controlling Light with Light, Academic, New York (1985)

Optical NOR Gate Using Diode Laser Sources

Masahiro Ojima

Hitachi, Central Research Laboratory

Arturo Chavez Pirson, Yong H.Lee, Jean F.Morhange,
Hyatt M.Gibbs, Nasser Peyghambarian
Univ.of Arizona, Optical Sciences Center

Feng Yuh Juang, Pallab K.Bhattacharya,Doreen A.Weinberger
Univ.of Michigan, Dep.of Elect.Eng.& Comp.Science

1.Introduction

Optical logic elements, especially in the form of 2 D arrays, can possibly be applied to massively parallel signal processing. GaAs/AlGaAs multiple quantum well etralons have previously shown optical bistability[1] and various types of logic gate operation[2] at room temperature. Most of these experiments have used a dye laser, which is large in size and expensive. One attractive feature of these devices is that a GaAs/AlGaAs diode laser can be used as a light source making the total system more practical. The present paper demonstrates a stable optical NOR gate using two diode lasers.

2.Experiments & Results

Figure 1 shows the two optical configurations that were tested to demonstrate an optical NOR gate using two diode laser sources. The main difference between the two is that a Faraday isolator is used in Fig.1(a), while a quarterwave plate (plus a polarization beam splitter)

serves as an isolator in Fig.1(b). The optical gate consists of a GaAs/AlGaAs multiple quantum well crystal sandwitched between two dielectric mirrors. The free exciton absorption peak was observed at 825 nm at room temperature, and a diode laser with 825 nm peak wavelength was selected for the pump.

Figure 2 is a typical output signal of the optical NOR gate depicted in Fig.1(a). The operating principle has been explained as follows.[3] The transmission peak wavelength of the nonlinear Fabry Perot etalon is initially set to the probe laser wavelength so that transmission is high without the pump. When the pump light pulse (input) is absorbed, the gate's transmission (output) is driven low. This is because the free exciton absorption is saturated and the index of refraction is decreased at the probe wavelength.

The essential component in the aparatus is a Faraday isolator. When a quarterwave plate plus a polarization beam splitter was used instead of the Faraday isolator, the spectrum of the probe laser was found to change due to the pump laser as shown in Figure 2. The spectrum change in the probe laser can cause erroneous gate operation. Even with the pump laser off, mode hopping noise can be induced by feedback of probe laser light reflected from the gate. This situation is similar to the feedback induced noise in an optical disk system.[4] Such noise is often enhanced by the wavelength dependence of the gate's transmission. The Faraday isolator eliminates both the pump and feedback effects on the probe laser frequency.

This demonstration is far from an actual signal processing system in terms of cascadability, power dissipation, and 2 D array. However, it obtains the answer to a logic question using an all optical gate and diode lasers. As such, it is one step toward optical signal processing.

3.Conclusion

Optical NOR gate operation has been successfully demonstrated using two diode lasers and a GaAs/AlGaAs multiple quantum well etalon. The essential component is a Faraday isolator that eliminates probe laser frequency shifts caused by unintentional pump beam injection and probe beam feedback. The fact that diode lasers can be used as light sources makes GaAs optical logic gates much more attractive for optical signal processing.

References

1. H.M.Gibbs et al.,Appl.Phys.Lett.$\underline{41}$,221,(1982).
2. J.L.Jewell et al.,Appl.Phys.Lett.$\underline{46}$,918,(1985).
3. H.M.Gibbs et al.,Appl.Phys.Lett.$\underline{34}$,511,(1979).
4. M.Ojima et al.,Appl.Opt.$\underline{25}$,1404,(1986).

(a)

(b)
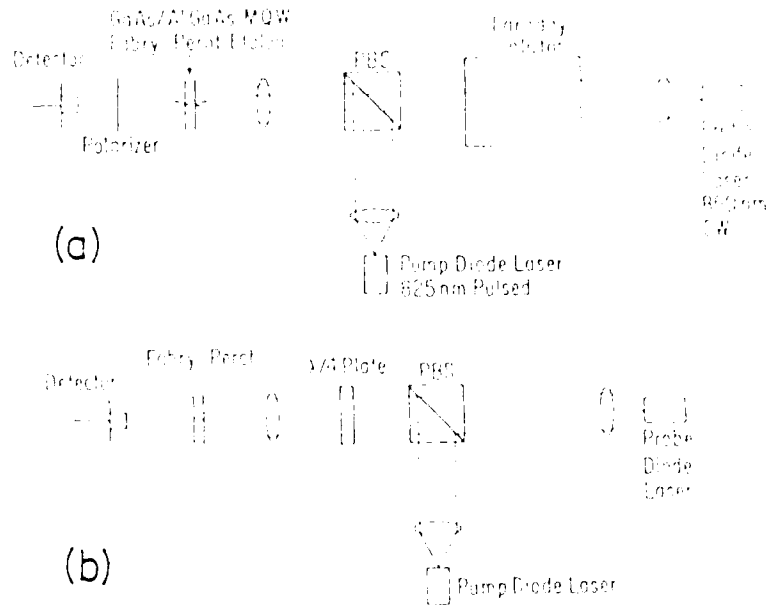
Fig. 1. Configurations for optical NOR gate operation with two diode laser sources. The optical isolator is either (a) a Faraday rotator or (b) a quarterwave plate plus polarization beam splitter(PBS)
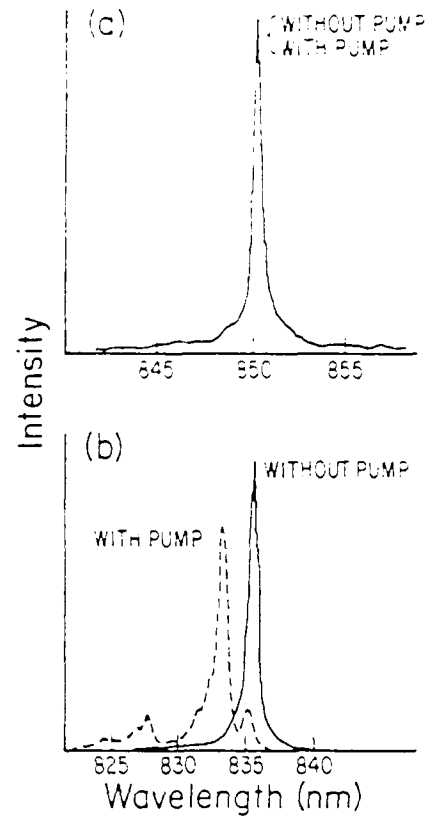
Fig. 3. Typical probe diode laser spectra: (a)for configuration of Fig.1(a) and (b) for configuration of Fig.1(b).

Input

0 1 0 1
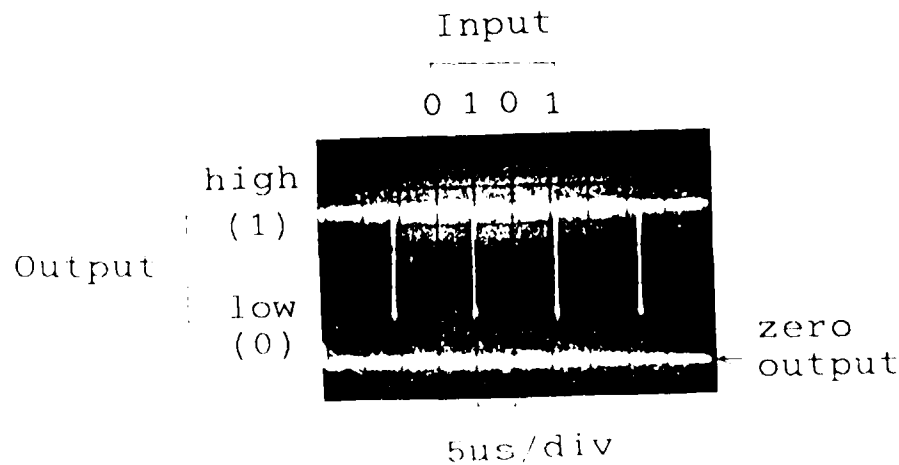
Fig. 2. Optical NOR gate output signal

247

Multiple Polarization State Threshold Logic and Processor

Shudong Wu   Xiang Zhang   Zhijiang Wang

Shanghai Institute of Optics and Fine Mechanics,
Academia Sinica, P.O. Box 8216, Shanghai, China

This paper suggests a novel technique of using multiple polarization states to implement different logic operations in parallel. Sophisticated full optical A/D converter, look-ahead adder and multiplier are proposed.

Intensity-Polarization Encoding

We proposed a phase-polarization encoding principle[1], based on which phase distribution may be converted to polarization distribution. With the same principle, light intensity in photorefractive material (PRM) in an optical logic may also be encoded to a polariration state. See Fig.1. Suppose the reading beam having no effect on PRM. The polarization state of the output beam depends on the phase variation in PRM, which is induced by the writing beam I_1. Consequently intensity I_1 is encoded to a polarization state.

Multiple Polarization State Threshold Logic

In Fig.1, the intensity behind the analyzer P_2 is $I_2 = \frac{1}{2} I_0 \left[1 + \cos(\Delta\phi \cdot I_1 + 2\theta) \right]$, where $\Delta\phi$ is the phase variation induced by unit intensity of writing beam I_1, θ is the orientation angle of P_2, which is chosen in such way that when $I_1 = 0$, $\theta = 0$, I_2 is maximum.

For a fixed $\Delta\phi$, I_2 may be regarded as the output of an optical logic with input I_1. With proper thresholding detection, different logic functions may be realized by setting different angle θ. For $\Delta\phi = \frac{\pi}{2}$ and threshold of 0.7, the logic functions are listed in table 1. Therefore one logic element can simultaneously implement different operations, provided the output beam is splitted into several beams and each beam has an analyzer in corresponding orientation. If the input A and B are not symmetric, asymmetric logic functions may also be performed. Consequently any of 16 logic func-

tions can be realized by adjusting $\Delta\phi$ and θ. With an electro-optic plate or a biasing writing power the logic function may be programmable. This technique is more flexible and reliable than Imai's fringe shifting processor[2].

Full Optical A/D Converter

Fig.2 shows a full optical A/D converter. The basic element is the same as in Fig.1. The light beam, whose intensity is to be digitized, is multiply-reflected by the two surfaces of a reflecting plate with reflectance of 50% and 100%. A series of beamlets is generated by multiple reflection. They have an intensity decreasing ratio of 2. See Fig.3. Each beamlet serves as the writing beam for one digit. The output intensity behind the analyzer varies periodically with the intensity of the writing beam. An intensity increment in the writing beam, which causes one period variation in the output, is taken as unit intensity. With proper threshold, the output will be the binary digits representing the intensity of the input beam. The first beamlet generates the least significant bit. The final beamlet, whose intensity is greater than half of unit intensity, generates the most significant bit. By using another direction, multichannel A/D converter (vector A/D converter) can be realized.

Compact Optical Look-Ahead Adder

The speed of a ripple adder is limited by carry propagation. Chandran[3] et al proposed an optical look-ahead adder. Unfortunitely the proposed system needs 6 nonlinear elements and too complicated to apply. Based on the principle of A/D converter described above, we may construct an compact optical look-ahead adder as shown in Fig.4. The light beams of two numbers are first combined digit by digit, then reflected by the multiply-reflecting plate. The incident angle is chosen in such way that each reflected beamlet by the upper surface coincides with its adjacent last reflected beamlet by the lower surface. In this adder the carries are automatically generated by multiple reflection. Each beamlet input to the nonlinear element is the analog combination of A, B and the corresponding carry. The final output is the result of $A \oplus B \oplus C$, where $\oplus$ represents XOR operation.

This adder is highly parallel. The carries are generated in

light speed. Only 2 nonlinear elements (including thresholding) are required. Addition of vectors may also be implemented in this processor.

Full Optical Binary Multiplier

To obtain final binary data of multiplication by means of convolution, summation and digitization operations are required. Usually they are performed electronically, such as in acousto-optic approach[4]. Here we propose to implement them all optically.

As is shown in Fig.5, the two binary numbers are arranged orthogonally with each other. Each digit is extended to a column and a row, respectively. They are combined and impinge upon a 2-D AND logic array. At the output plane of the AND array, a 2-D array of data results. After an optical combination along the diagonal direction of the data array with a cylindrical lens, we get a mixed binary number. This number is taken as the multiple beamlet input to the reflecting plate of the A/D converter shown in Fig.2. The carries addition and digitization can all be implemented automatically. The final binary data of multiplication result at the output plane of the A/D converter.

Discussion and Conclusion

1) Polarization encoding may be a promising approach for forming optical logic. The logic elements are highly parallel, flexible and not sensitive to turbulence. The encoding scheme was experimentally demonstrated.

2) The logic using multiple polarization states may be regarded as analog threshold logic. From the given examples it is shown that for realizing certain logic functions, threshold logic may greatly reduce the number of required gates.

3) Thresholding detection is the key problem for the performance. An analysis of the accuracy requirement will be given.

4) The proposed A/D converter, look-ahead adder and multiplier have a mumber of advantages over other schemes. They are of great potential applications for optical computing, signal and image processing.

Reference
1. S.T. Wu, C.L. Xu, Z.J. Wang, ICO-13 Conference Digest,
       p. 458 (1984) Sapporo, Japan.

2. Y. Imai, Y. Ohtsuka, Opt. Eng., 25, 1, 98(1986).

3. V. Chandran, T.F. Krile, J.F. Walkup, Appl. Opt., 25, 14,2272(1986)

4. P.S. Guifoyle, Opt. Eng., 23, 1, 20(1984).

INPUT		2 x ANALYZER ANGLE θ					
A	B	0	$\frac{1}{2}\pi$	π	$\frac{5}{4}\pi$	$\frac{3}{2}\pi$	$\frac{7}{4}\pi$
0	0	1	0	0	0	0	1
0	1	0	0	0	1	1	1
1	0	0	0	0	1	1	1
1	1	0	0	1	1	0	0
LOGIC		NOR	F	AND	OR	XOR	NAND

Table 1 Logic Functions
with
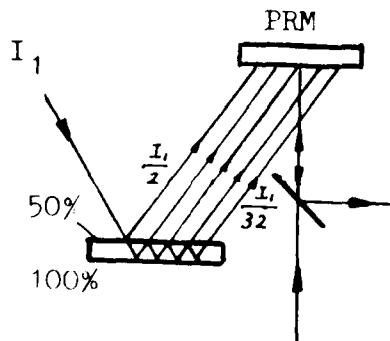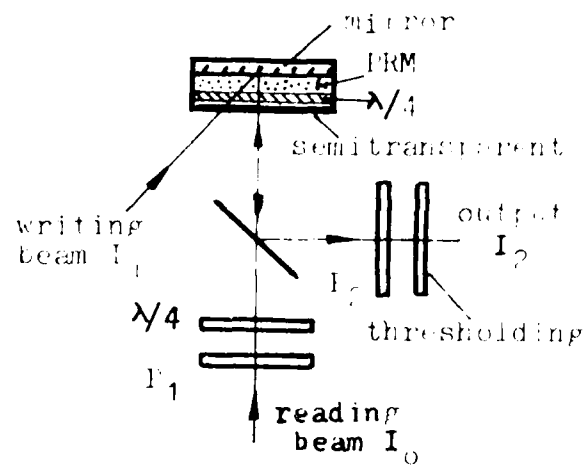
Fig.1 Intensity-Polarization
Encoding

Fig.2 Full Optical A/D
Converter
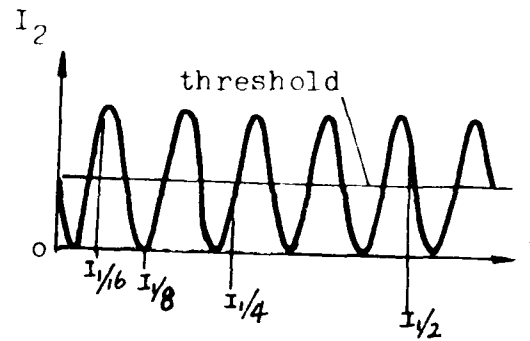
Fig.3 Input and Output
Intensity, Thresholding
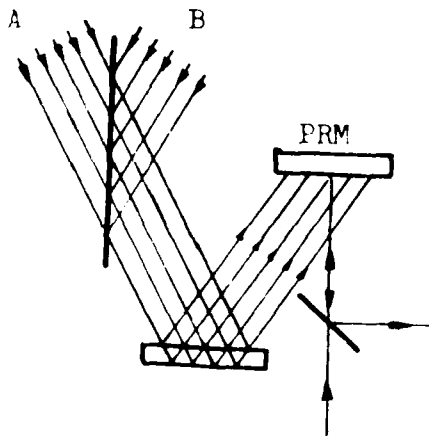
Fig.4 Look-Ahead Adder
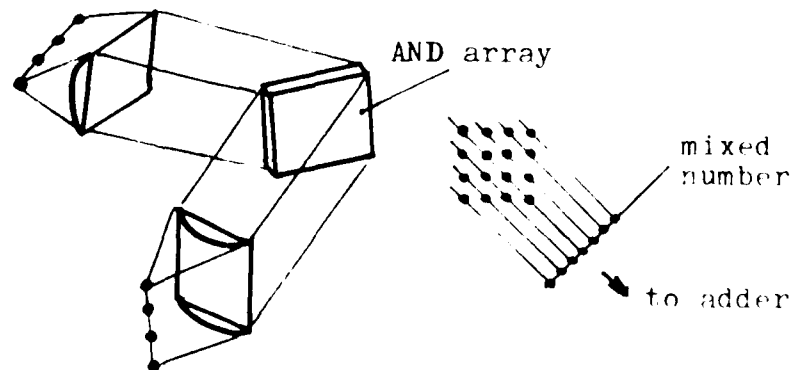
Fig.5 Full Optical Multiplier

INTERFEROMETRIC PATTERN ENCODING FOR PARALLEL LOGIC OPERATION

Makoto IKEDA and Toyohiko YATAGAI

University of Tsukuba, Institute of Applied Physics
Tsukuba Science City 305 JAPAN

Satoshi ISHIHARA and Yoshinobu MITSUHASHI

Electrotechnical Laboratory, Opto-Electronics Section
Tsukuba Science City 305 JAPAN

Junichi KAYA

Nippon Institute of Technology
Miyashiro, Minamisaitama-gun, Saitama 345 JAPAN

I. Introduction

Recently, some optical parallel pattern logic operations have been proposed[1-2]. In order to implement these techniques, binary input data represented in transparent or opaque pixels must be spatially encoded. Tanida and Ichioka have described the method of spatial encoding technique[3] based on a holographic EXCLUSIVE-OR (XOR) operation for their parallel logic operation.

In this paper, we propose a method of spatial encoding by a simple and practical interferometric technique. Its application to the space-variant logic-gate array technique [2] which is one of the parallel pattern logic operation is also discussed.

Fig. 1

Spatial encoding table.
a_{ij}, b_{ij}: input data
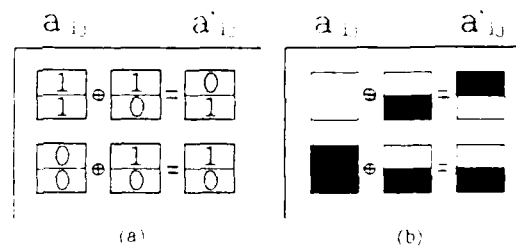a'_{ij}, b'_{ij}: encoded output

(a)   (b)

Fig. 2

Spatial encoding method.
a_{ij}: input data
a'_{ij}: encoded output
$\oplus$: XOR operation

II. Spatial encoding method[1]

Fig. 1 shows the spatial encoding table for input data a_{ij} and b_{ij}, which represent ij-pixel of the input pattern A and B, respectively. XOR operation of a_{ij} and the code $\boxed{\begin{smallmatrix}1\\0\end{smallmatrix}}$ enables us to obtain the desirable encoded output a'_{ij}, as shown in Fig. 2(a) or Fig. 2(b). Of course, the encoded output b'_{ij} can be obtained from input data b_{ij} when we choose the code $\boxed{1\,0}$.

With code pattern G_A or G_B as shown in Fig. 3, this XOR encoding operation can be applied to the whole pixels in a pattern.

$$A \oplus G_A = A'$$

$$(a)$$

$$B \oplus G_B = B'$$

$$(b)$$
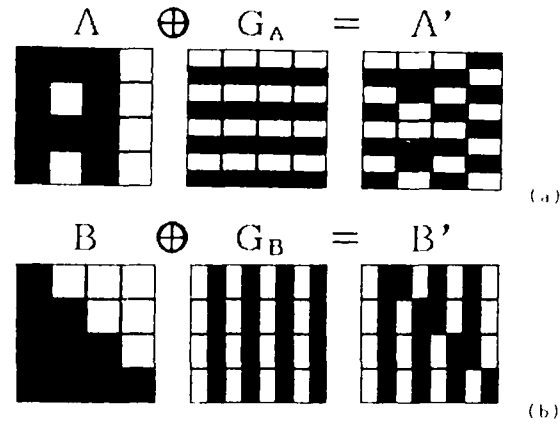
Fig. 3

Input pattern encoding process.
A , B :input pattern
G_A , G_B :code pattern
A', B':encoded pattern
$\oplus$:XOR operation

III. Spatial encoding with an interferometer

The XOR operation can be performed with an interferometer, in which the phase difference between two arms is π. Let us consider an interferometer as shown in Fig. 4, in which the input plane P1 and P2 are simultaneously imaged on to the output plane P3. Therefore transparent objects in planes P1 and P2 make the plane P3 dark by means of interference phenomenon. Plane P3 is also dark when both of planes P1 and P2 have opaque object. Opaque object on either plane of P1 or P2 causes plane P3 bright.

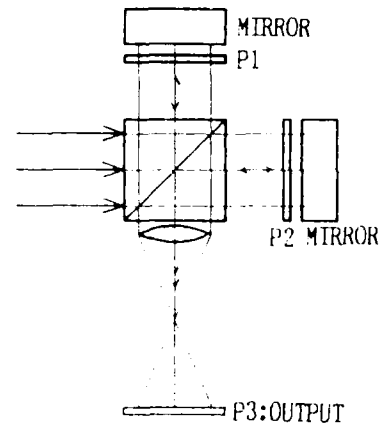If we regard transparent and opaque object as logical 1 and 0, respectively,

Fig. 4

Interferometric pattern encoding system.
P1 and P2 are imaged in P3. P2 has phase
delay π with respect to P1.

then XOR operational output can be observed. Consequently, input pattern A and code pattern G$_A$ in the plane P1 and P2, respectively, cause the encoded pattern A' in plane P3.

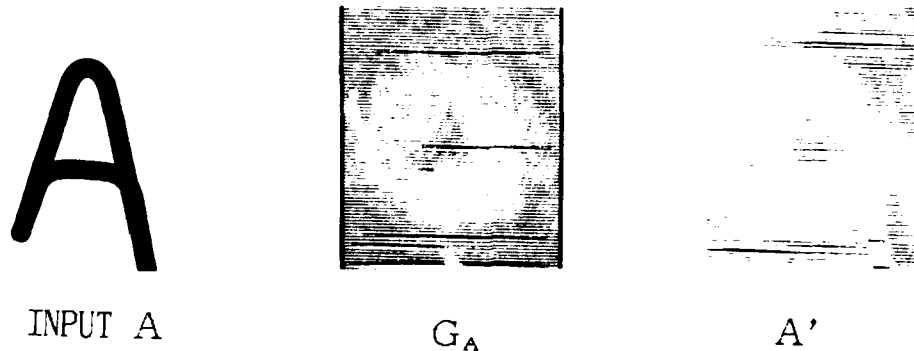Fig. 5 is an experimental result of this method. 80x65 (=5200) pixels are encoded in parallel in 12x10mm² area.

INPUT A   G$_A$ A'

Fig. 5
Experimental result. 80x65=5200 pixels
are spatially encoded in 12x10mm² area.
A : input pattern
G$_A$: code pattern
A' : encoded output

IV. Superimposing technique

We present here a serial connection of the interferometric encoding system as shown in Fig. 6. Setting an input pattern A in the plane P1 and a code pattern G$_A$ in the plane P2 makes an encoded pattern A' in the planes P3 and P4. We also set an input pattern B in plane P3 and a code pattern G$_B$ in plane P4 that makes an encoded pattern B' superimposed with A' in the plane P5.

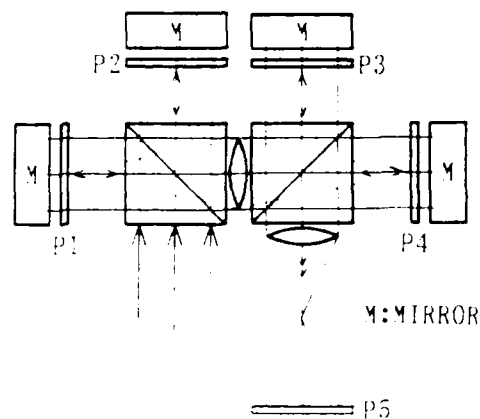A detailed explanation of this procedure is shown in Table 1.

M:MIRROR

Fig. 6
Optical system for logic-gate array. The
spatially encoded and superimposed
pattern. A'∧B' can be obtained on P5
plane.
∧:AND operation

Therefore we can observe the spatially encoded and superimposed pattern $A' \wedge B'$ on the P5 plane. Where $\wedge$ represents the AND operation.

PLANE	INCIDENT	PATTERN	OPERATION
P1	1	A	A
P2	1	$G_A \cdot \exp(+i\pi)$	$G_A \cdot \exp(+i\pi)$
P3	P1+P2	B	$(A \oplus G_A) \wedge B$
P4	P1+P2	$G_B \cdot \exp(+i\pi)$	$(A \oplus G_A) \wedge (G_B \cdot \exp(+i\pi))$
P5	P3+P4	—	$(A \oplus G_A) \wedge (B \oplus G_B)$ $= A' \wedge B'$

Table 1

Spatial encoding and superimposing process.

V. Implementation of a logic-gate array

By using a decoding mask which is presented in the space-variant logic-gate array technique[2] on plane P5 we can implement this technique. Application of a spatial light modulator (SLM) to the planes P1, P3 and P5 is effective for data input and logical operation.

The logic-gate array performs a variety of space-variant logical operations. Therefore it permits the design of more flexible and powerful parallel pattern logic operation system.

VI. Conclusion

We proposed the spatial encoding method using an interferometric technique. The application of this method to the space-variant logic-gate array was also discussed.

References

1) Y. Ichioka and J. Tanida, Proc. IEEE, Vol. 72, No. 7, 787 (1984)

2) T. Yatagai, Opt. Lett., Vol. 11, No. 4, 260 (1986)

3) J. Tanida and Y. Ichioka, Proc. SPIE, Vol. 437, 119 (1983)

INFRARED PRE-DETECTION DYNAMIC RANGE COMPRESSION VIA PHOTOREFRACTIVE CRYSTALS

Hua-Kuang Liu and Li-Jen Cheng

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91109

ABSTRACT

Pre-detection infrared dynamic range compression concept via the nonlinear photorefractive two-wave mixing in GaAs crystals will be discussed. Some experimental results will also be presented to support this idea.

Pre-detection dynamic range compression is of importance for solving the problem where the input image has such a high dynamic range that no detectors or sensors can record the complete intensity range of the image without saturation. An example is in the detection of a scene with a shiny automobile or an aluminum building under strong solar illumination in a background of low reflectance. Similar infrared scenes of military or industrial interests also pose a problem for image recording and/or data acquisition.

Recently, photorefractive properties of materials such as BSO, BGO, $BaTiO_3$, $LiNbO_3$, and the compound semiconductors such as [2-6] GaAs, CdTe, and InP have been studied. The compound semiconductors are of special interest because of their infrared wavelength of operation and high speeds due to their large electron mobility.

In addition, photorefractive crystals have also been used in many real-time image processing operations. For example, photorefractivity has been applied for convolution and correlation[7], edge enhancement[8], division[9], differentiation[10], inversion[11], subtraction[12], and, optical digital logic operation[13]. In this paper, we shall describe a new concept of using the photorefractive crystals for real-time pre-detection dynamic range compression. Experimental results using chrome-doped GaAs crystals[6] are presented to demonstrate the feasibility of the idea.

First we shall present the basic idea of pre-detection dynamic range compression. The dynamic range (D.R.) of an input image may be defined as follows:

$$D.R. = \frac{I_{max}}{I_{min}} \qquad (1)$$

where I_{max} and I_{min} represent the maximum and minimum intensities

of the input image.

Before the image is received at the detector, a device may be
used to map the input image into an output image in the following
functional form:

$$I_{out} = f(I_{in}) \tag{2}$$

where f is a function or mapping, I_{out} and I_{in} are the output and
input intensities respectively. All the intensities may be in
one- or two-dimensional form. The function f may be linear or
nonlinear. For example, we may write

or

$$I_{out} = A\left(I_{in}\right)^x,$$

$$I_{out} = B \log I_{in},$$

where A and B are constants and x is an index number which should
be less than one for D.R. Compression. The essence of D.R.
compression is to find a device with a characteristic f such that
f will cause

$$D.R._{output} < D.R._{input} \tag{3}$$

The idea and its usefulness can be illustrated with the
assistance of Fig. 1. The upper part of the figure shows the
characteristic of a typical detector such as a photographic film.
If the input intensity range is too large for the detector's D.R.,
saturation occurs as shown. After pre-detection D.R.
compression, this saturation problem can then be resolved. Since
there is still a one-to-one correspondence in terms of the input
details, the information capacity is kept intact in this process.

Next we describe how the two-wave mixing beam coupling scheme can
be used to compress the D.R. range of an input image or datum.

It is well-known that during the two-wave mixing process in a
photorefractive crystal, one beam can gain energy at the expense
of the other. This effect is illustrated in Fig. 2. The input
beams are represented by I_{in1} and I_{in2} and the output beams are
represented by I_{out1} and I_{out2} respectively. The shaded part
indicates the zone of intereaction between the two beams and z is
the thickness of the effective interaction zone of the crystal.
The gain of the beam No. 2 over beam No. 1 due to the coupling
may be written as

$$\Gamma = \frac{1}{z} \ln \frac{I_{in1} \quad I_{out2}}{I_{in2} \quad I_{out1}} \tag{4}$$

We now present a special case to illustrate the feasibility of
the idea. Referring again to Fig. 3, if we assume that $I_{in1} = I_{in2} = 1/2\, I_{in}$, then according to Eq. (4), we have

$$I_{out1} = \frac{I_{in}\, e^{-\Gamma z}}{1 + e^{-\Gamma z}} \tag{5}$$

We further assume that $\Gamma z \ll 1$, then Eq. (5) may be approximately

written as

$$I_{out1} \cong (1 - \sqrt{\vphantom{z}\mathbf{z}}) \, I_{in1} \qquad\qquad (6)$$

Equation (6) clearly shows that I_{out1} is smaller when I_{in1} is larger provided Γ increases with I_{in1}. Recently, Cheng and Partovi[6] have reported that Γ indeed increases with I_{in1} in Cr:GaAs, supporting the idea of the D.R. compression described above. The result is shown in Fig. 3.

It can be seen that when the material has a maximum gain $\Gamma_{max} = 0.21$ cm^{-1}, the D.R. compression is not very effective with $\Gamma z = 1$ or the thickness of the crystal to be even approximately 5 cm. Therefore, higher Γ_{max} should be sought after for D.R. compression. In some cases, a maximum gain of 10 or higher can be achieved with applied electric field. With such a high gain, referring to Fig. 3, a crystal thickness of $z = 0.5$ cm will offer us a significant D.R. compression.

Based on the above discussions, we may conclude that by using the nonlinear gain effect of photorefractive crystals, we can achieve real-time pre-detection dynamic range compression in both 1-D and 2-D cases. Problems to be investigated in the future are optical architecture for testing 2-D operation in both coherent and the more difficult incoherent cases, material selection for high gain operation, and operation time requirements for real-time implementations. In addition, engineering packaging will be needed for industrial applications.

References
1. P. N. Gunter, Phys. Rev. 93, 199 (1982).
2. M. B. Klein, Opt. Lett., 9, 350 (1984).
3. A. M. Glass, A. M. Johnson, D. H. Olson, W. Simpson, and A. A. Ballman, Appl. Phys. Lett., 44, 948 (1984).
4. J. Strait and A. M. Glass, JOSA(B), 3, 342 (1985).
5. G. Albanese, J. Kumar, and W. H. Steier, Opt. Lett., 11, 650 (1986).
6. L. J. Cheng and A. Partovi, Appl. Phys. Lett., to be published in Nov., 1986.
7. J. O. White, A. Yariv, Appl. Phys. Lett., 37, 5 (1980).
8. J. Feinberg, Opt. Lett., 5, 330 (1980).
9. Y. H. Ya, Opt. Comm., 44, 24 (1982).
10. Y. H. Ya, Appl. Phys., B, 36, 21 (1985).
11. E. Ochoa, E. Henselink, and J. W. Goodman, Appl. Opt., 24, 1826 (1925).
12. A. E. Chiou, Pochi Yeh, and M. Khoshnevisan, SPIE Proc., 613, 201 (1986).
13. Y. Fainman, C. C. Guest, and S. H. Lee, Appl. Opt. 25, 1599 (1986).
14. N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin, and V. L. Vinetskii, Ferroelectrics, 22, 949, 961 (1979).
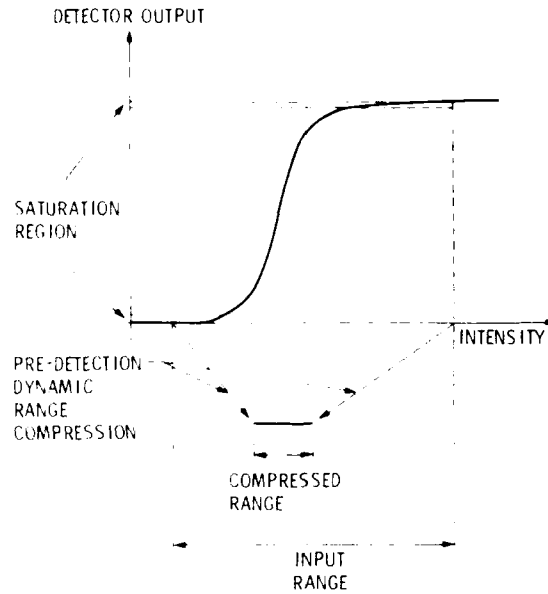
Figure 1. Illustration of pre-detection dynamic range compression. The upper part of the figure shows the characteristic of a detector or a sensor with original range of inputs and its saturation effect. The compressed range of the output is also shown.
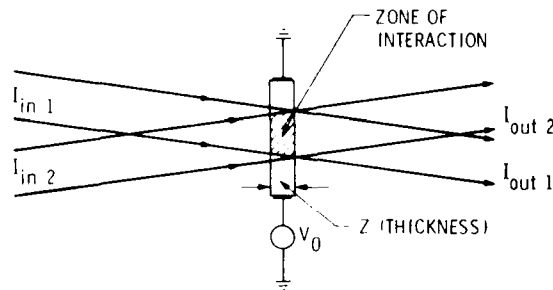

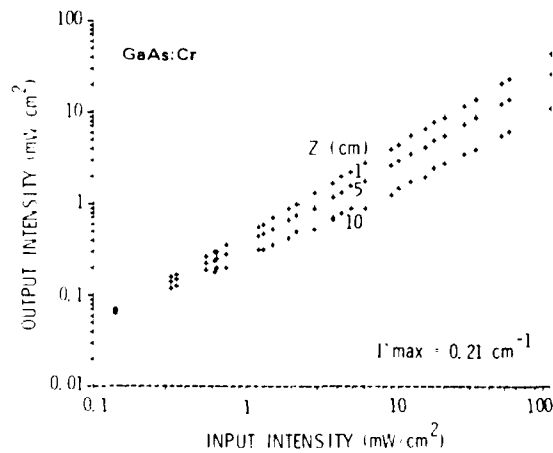Figure 2. Two-beam coupling in a photorefractive crystal.

Figure 3. Dynamic range compression using photorefractive effect

FINGERPRINT ENHANCEMENT BY FOURIER DOMAIN OPTICAL PROCESSING

D M Monro, B G Sherlock and C R Petts [*]

Department of Computing, Imperial College, 180 Queen's Gate London, SW7 2BZ
and [*] GEC Research Ltd, Hirst Research Centre, East Lane, Wembley, Middlesex

INTRODUCTION

Fingerprint enhancement is an essential stage in the computer processing of
fingerprint images. A technique based on directional Fourier domain filtering
which is capable of producing excellent results is discussed in this paper.
Implementation of this algorithm on a general-purpose digital computer is
slow, but using optical hardware removes the computational overhead of Fourier
domain filtering.

FINGERPRINTS IN THE FOURIER DOMAIN

Figure 1 shows a typical fingerprint image together with its amplitude Fourier
transform. Characteristically, a fingerprint image consists of parallel
curved ridges of roughly equal spacing, which at each point are orientated at
an angle known as the local ridge orientation (LRO).

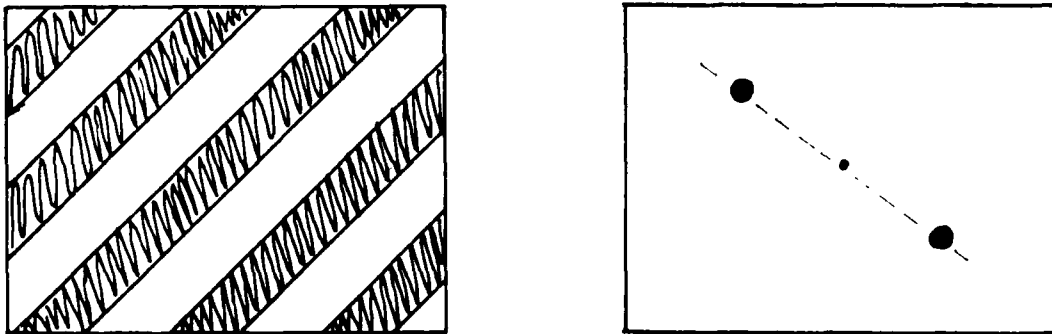Figure 1 (a) Rolled fingerprint image and (b) its amplitude Fourier transform

Figure 2 (a) Locally windowed fingerprint (b) its amplitude Fourier transform

The equal spacing of the ridges in the spatial domain concentrates the Fourier components within a well-defined annulus of frequency components centred on frequency (0,0). Components are present in all directions perpendicular to the LRO values that are present in the spatial image.

Figure 2 shows diagrammatically a small windowed section of a fingerprint image, together with its amplitude Fourier transform. Such a window, if its size is chosen correctly, selects equally spaced parallel straight lines, which corresponds to a pair of diagonally-opposite components, aligned perpendicularly to the selected lines in the spatial domain (i.e. the LRO).

THE FILTERS USED

Since each point on the fingerprint has a neighbourhood resembling Fig 2 (with the exception of a small number of singularities - the cores and delta points), it is clear that at each point we can heavily filter the image without damaging local ridge information, by passing only components within a small neighbourhood surrounding the components illustrated in Figure 2. The image is bandpass filtered radially to allow through only components of wavelength close to the ridge spacing, while at the same time being filtered angularly to allow through only components of direction close to the LRO.

The radial filter was chosen to be a second-order Butterworth bandpass filter and the directional filter was chosen to have a cosine shape and bandwidth π/n radians ($n=8$ worked well in our filters).

Thus,
$$H(r,\theta) = H_{radial}(r) \cdot H_{angle}(\theta - \theta_o)$$
where
$$H_{radial}(r) = \text{2nd order butterworth bandpass filter}$$

$$H_{angle}(\theta) = \cos^2(1/n(\theta - \theta_o))$$

θ_o is perpendicular to the LRO.

Therefore, provided that one knows the LRO at each point, one can determine θ and hence $H(r,\theta)$ for each point in the image.

OPTICAL IMPLEMENTATION

A possible optical implementation of the process is illustrated in Figure 3. In this case the raw image (in the form of a transparency) is illuminated with a uniform plane beam derived from a low power laser. A small aperture is used to define the region of the transparency to be interrogated. The aperture is arranged to be sufficiently large to contain a number of ridge wavelengths, but small enough so that the LRO does not vary significantly.

In order to enhance an image, the transparency is raster-scanned through the plane so that the local region surrounding a point x_1,y_1 is illuminated. The light from this region is Fourier transformed by the lens L_1 such that the image corresponds with the filter plane P_2. A proportion of the light is reflected from a beam splitter onto a sectorized photo-detector. The LRO of this region results in a Fourier transform with a well-defined direction, this direction can be determined by the sector-detector.
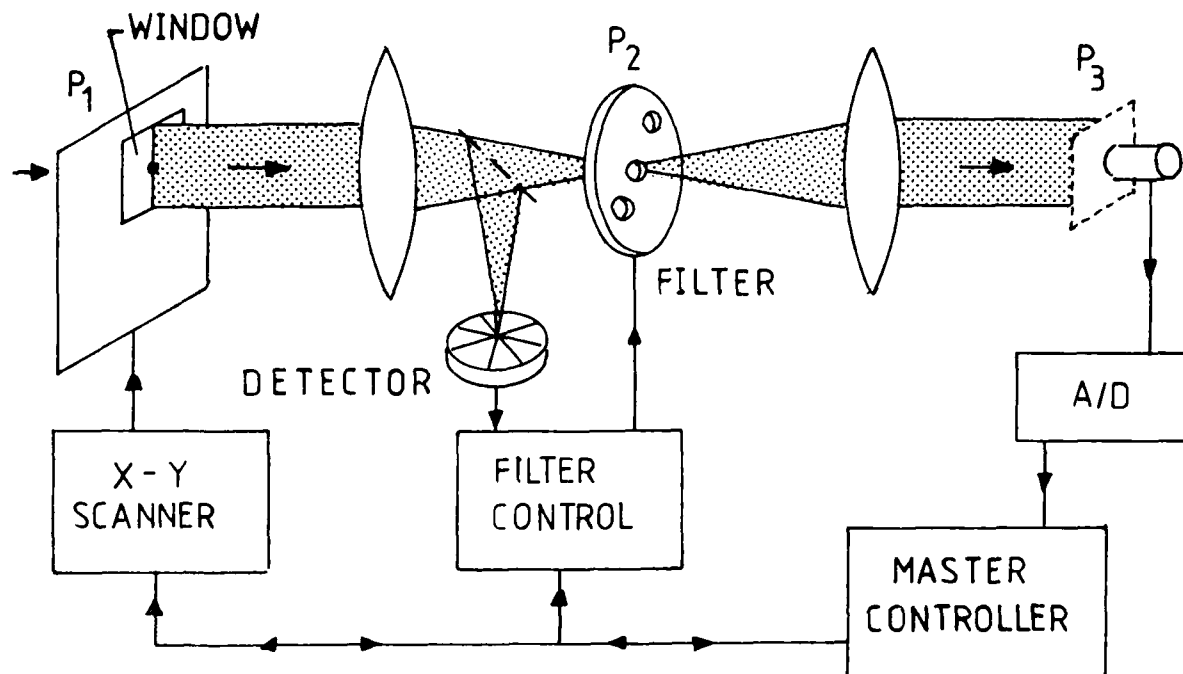
Figure 3. The optical fingerprint enhancement apparatus

The Fourier plane filter is arranged to have the transmittance described above, with a zero spatial frequency component added to make all amplitudes positive. The axial position θ_0 of the filter is controlled from the sector-detector via a servo-system in order to line up with the LRO of the region.

Once the transform of the local region has been appropriately filtered, the distribution is retransformed to give an enhanced image at the plane P_3. The intensity at the point x_1' , y_1' , is then recorded using a single photo-detector whose aperture corresponds to the original pixel size. An A/D converter is then used to record the intensity $I(x_1'$, y_1' $)$ for each scan position x,y. The full enhanced image is therefore built up point by point during the mechanical scan and may be displayed using a frame-store.

A parallel alternative to the above approach is possible, but would involve making a simplifying assumption – that it is sufficient to filter the image in n (say n=8) possible equally-spaced directions $\theta_i = (i-1)\pi/n$, $i=0..n-1$, and that the use of the filter corresponding to the θ_i closest to the true value determined by the LRO will be sufficiently accurate.

In this alternative approach one would use n+1 (say 9) separate optical channels after the transforming lens L_1. Eight of these channels would be used to illuminate fixed filters of different orientations, the ninth being used to determine the LRO using a sector-detector. Such a system has edvantages in terms of processing speed since the switching between the outputs of the eight channels can be accomplished much faster than if a mechanical servo-system is used to align the filter. The disadvantage is the need to replicate the optical components.
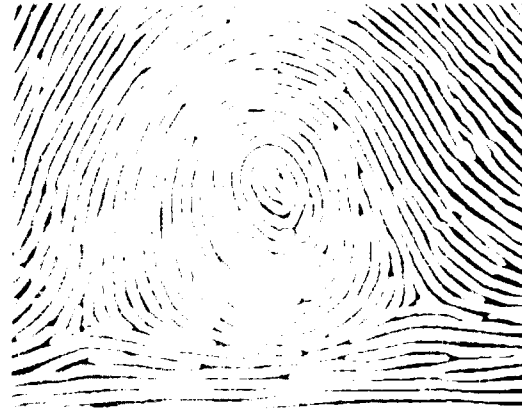
Figure 4 (a) Enhanced fingerprint produced   (b) Thresholded version of (a)
by computer simulation.

RESULTS

A computer simulation of the above apparatus produced the results shown in
Figure 4. However, the enhancement shown in Figure 4 is better in one respect
than would be expected from the optical system as shown in Figure 3. In the
neighbourhood of the cores and deltas, the curvature of the ridges becomes
very great and the angular bandwidth of the filter has to be increased the
nearest one gets to these points. Although this was implemented in the
simulation, we have not yet determined how to include this widening of angular
bandwidth in the optical hardware. A method is needed to determine local ridge
curvature, which could probably be done using the same sector-detector as was
used to determine LRD.

Figures 5a and 5b show results obtained from a simple optical filtering
experiment. In this experiment a periodic image with defects has been
processed using a form of the filter described above. Although this is only a
preliminary result, the effect of cleaning-up the image is quite apparent.
This can be regarded as the computation of a single x,y point of an enhanced
image. In presenting our paper we will present initial results obtained from
equipment of the form illustrated in Figure 3, along with a comparison with
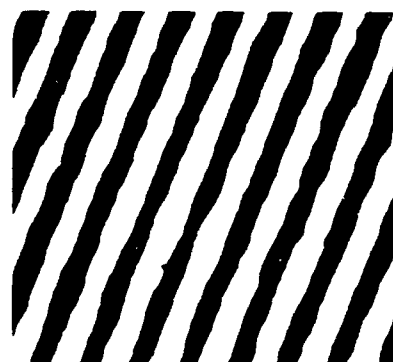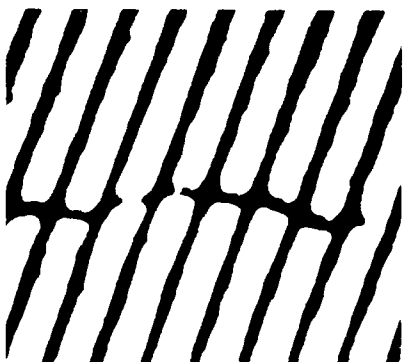results obtained in a general purpose computer.

Figure 5 (a) Noisy fingerprint fragment   (b) The effect of optical filtering

WEDNESDAY, MARCH 18, 1987

PROSPECTOR / RUBICON ROOM

1:30 PM–5:20 PM

WB1–4

JOINT PHOTONIC SWITCHING AND OPTICAL COMPUTING PLENARY SESSION

PHOTONIC SWITCHING COMPONENTS: CURRENT STATUS AND FUTURE POSSIBILITIES

J. E. Midwinter
University College London
Torrington Place
London WC1E 7JE

I. INTRODUCTION

The interest in optical switching seems to stem from several factors. One might well be cost, avoiding the need for two opto-electronic interfaces at the switching node. Further to that, in a local or campus network, the low insertion loss and bandwidth of single mode fibre offers the possibility of truly "transparent" networks, in which the format and data rate are set by the communicating terminals and the routing is achieved independently in a way that does not require synchronous or rigidly formatted data streams. Such switches may also support bidirectional operation both within the fibre and the switch. Another attribute of such technology is likely to be the provision of very large communication bandwidths as well as the usual optical advantages of low cross talk and distortion.

These factors are applied most readily in networks in which an incoming fibre carries data from a single user to a single user so that wideband circuit switching without multiplexing is required. However, often some form of multiplexing will be needed, with the result that the incoming and outgoing fibres will be carrying multiple services simultaneously, travelling to or derived from diverse locations. In these circumstances, the switch must perform a more complex operation. At present, two approaches appear to be under study, based upon WDM and TDM. The former case reduces to the circuit switched situation having dispersed wavelengths while the latter requires a much more sophisticated accurately timed switch structure, placing a very high premium on the timing accuracy.

II. ELECTRICALLY CONTROLLED EXCHANGE-BYPASS UNITS.

The basic building block for most switches is some form of four port element having the characteristic that two input ports, A & B can be connected directly (bypass) or crossed (exchange) to the two output ports C & D. Key operating parameters are the insertion loss, the cross talk level, the wavelength response, the switching time and precision. In single mode switches, the polarisation properties can also be important. One family of switches is based upon optical fibres. The simplest involves optical fibres connected via micro-optic elements (lenses, beam splitters etc) that are moved electro-mechanically or fibres themselves that are physically moved. Such devices lend themselves more to multimode than single mode operation, will generally have low insertion loss (particularly for multimode) and good cross talk characteristics but are inevitably rather slow in operation (order milliseconds), imprecisely timed, relatively insensitive to wavelength or polarisation but are physically bulky. They lend themselves to use within simple wideband video security networks. They do not appear to be more generally applicable.

Another class of fibre based devices is based upon fibre interferometers or "waveguide directional couplers", almost certainly using single mode fibre. For example, one might form a Mach-Zehnder Interferometer with two parallel single mode fibres fused at two points to form two 3dB directional couplers.

This leads to a four port coupler whose transfer characteristics depend upon the relative phase length of the two fibre paths between the two couplers. Changing one path relative to the other by a half wavelength switches the device from exchange to bypass or vice-versa. Such a change is readily induced by electrical heating of one arm. By the same token, thermal drift is likely to be problem. As with all fibre based devices, elements of this type can exhibit extremely low insertion loss and are potentially broad band. However, if the arms are unbalanced (unequal length), then the transfer characteristic becomes wavelength sensitive. The device may also be polarisations sensitive. Response time is expected to be slow, typically milliseconds and with relatively large dimensions (mm. to cm.). Once again, the devices seem most suited to use in circuit switched video surveillance or studio networks where switch set-up time is unimportant.

A totally different class of electrically controlled exchange bypass units emerges from the use of the electro-optic effect in guided wave integrated-optic form and commonly made by Titanium diffusion into Lithium Niobate crystal substrates. The exchange bypass unit so formed has two diffused waveguides that for part of their length run parallel and sufficiently close for their evanescent fields to interact. Switching is achieved by means of suitably placed electrodes that change the relative refractive indices in the two guides. These devices can have relatively low insertion loss, perhaps 1dB or better, although low loss coupling to fibre is difficult. Cross talk can easily be a problem. Fabricating large arrays poses major technological problems, since the devices are long (typically mm.) and confined to a single wiring plane. It appears that up to 16x16 may be possible on a single chip. Most devices are polarisation sensitive but clever design can overcome this. They can switch very fast, well into the sub-nanosecond regime, but present substantial capacitance to the drive circuit. Given a matrix occupying a large area (sq.cms), exploiting this speed to synchronously reset/set is almost certainly more difficult to achieve than switching a monolithically integrated electronic cross point. However, once the optical path is established, it offers "infinite" and bi-directional data bandwidth. Such elements are capable of very fast multiplexing or demultiplexing given electrical synchronisation. Wavelength response is also limited, typically to a few percent of the centre wavelength.

III. WAVELENGTH ROUTING
The growing availability of tunable sources and receivers opens possibilities of using a passive guided-wave network in a communication mode equivalent to that of free space radio communications, with terminals indentified by means of optical frequency. Some semiconductor lasers can be made to tune by the use of an external cavity over a range of 50-100nm, corresponding to more than 10,000 GHz, and in the case of those developed for coherent communication systems, can exhibit stabilities of better than 1MHz. Assuming channel spacings of 2GHz, one might speculate that as many as 5000 parallel and simultaneous channels could become accessible. Using less sophisticated lasers of the monolithically integrated type, a few tens to a few hundreds of wavelengths are accessible by electrical control.

At the receiver end, the coherent receiver selects its wavelengths primarily by means of the tunability of its local oscillator, a similar laser to the remote transmitter, although it is likely in a network using very many

wavelengths that some previous band-filtering might be desirable. A wide variety of techniques exist ranging from tunable filters based upon electro-optic directional coupler designs to those fabricated in fibres by means of grating structures impressed on the waveguide to simpler bulk dispersive elements interposed in the light path. Non-linear interactions in the fibre may also seriously limit perfromance. In any case, the network is transparent and passive and could be bi-directional (single fibre per terminal) and probably of a star format with central power splitting node. Alternatively, a dispersive element could be included at the central node although how this would be done for more than a small numbers of outgoing fibres is unclear. In principle, such wavelength switched networks looks extremely powerful given good tunable sources and receivers at competitive prices.

IV. OPTICALLY ACTIVATED SWITCHES

Studies in non-linear optics have led to a variety of optically activated switches, mainly based on optical bistability, that have led to much speculation on optical computing but have generated little interest for switching. Optical activation opens up interesting new opportunities in control and interconnection, particularly when coupled with free space (2 dimensional) optical "wiring" and the possibility of normally addressed planar arrays of devices, since the wiring rapidly becomes a limiting factor in other optical matrix concepts. Optical logic gates span a huge range of switching speeds from ps to ms but tend to require similar switching energies, at present in the 1E-9 to 1E-6 Joule range (ie 1000 gates at 1Gbit/s = 1kW to 1MW power!). It is expected that pico-joule sensitivities will be possible in time with bistable laser and other hybrid devices reaching femto-joule levels. Bistable gates are operated as threshold logic elements and thus incur all its normally associated problems. Hence there is growing interest in alternative approaches involving hard limiting opto-electronic circuits, preferably with good I/O isolation. However, the bistable element does imply a memory capability and this has been demonstrated in simple time-slot interchange switches although the engineering problems involved in constructing a large switch look formidable.

An alternative approach is to explore hybrid opto-electronic matrices, combining optical wiring and control, electronic logic and photodetectors and electro-absorption modulators monolithically integrated as I and O elements. This approach promises to combine the interconnect strengths of optics with the high speed logic strength of small electronic logic circuits and hence to lead to large ultra-fast optically controlled arrays. In time, all optical logic solutions may emerge and some novel approaches to their logical design have already been postulated that draw heavily on ideas generated in studies of optical computing. The technology for optically activated switches is still at an early stage of development although a variety of very interesting ideas can be discerned already. The major gain arise from optical wiring which provides very accurate timing control at high data rates (in excess of 1Gbit/s per port), thus opening up the possibility of switching packet or TDM data from ultra wideband highways.

WB2

Optical Digital Computers

Alan Huang
AT&T Bell Laboratories
Holmdel, NJ   07733

.

OPTICAL NEURAL COMPUTERS

Demetri Psaltis

California Institute of Technology
Department of Electrical Engineering
Pasadena, California, 91125

The development of optical computers of any type is based on the notion that semiconductor technology imposes limitations in the performance of current computers which prevent them from being effectively used for the solution of a class of interesting computational problems. If optics is used instead, these limitations will be lifted and we will therefore be able to now solve these interesting problems. Global connectivity is perhaps the most distinctive feature of optics vis-a-vis semiconductor technology, and the development of optical neural computers can be viewed as an attempt to exploit this feature. In a neural network each elementary computational unit, the neuron, directly communicates to thousands of others, while in electronic computers each gate is typically connected to only two or three gates. With optics it is feasible to realize the dense connectivity that is evident in neural networks. This provides the impetus for examining neural network models of computation to get ideas about how to build optical computers whose performance is clearly better than their electronic counterparts.

A neural network consists of two basic components: a large collection of neurons and a dense network of connections. Neurons are typically modeled as thresholding elements and information is stored in the strength of the connections largely through error driven learning. If during a learning phase the response of the network is correct then the connections remain unaffected. Otherwise they are modified to eventually produce a desired response. Within this basic framework (large number of neurons, dense connections, and learning by modifying the connections) numerous models have been developed that attempt to explain different aspects of natural neural systems. These models have attracted the attention of computer scientists and engineers and have served as a source of ideas for building computers that are well suited for solving the types of problems that humans are good at. A prime example of such a problem is pattern recognition: we do it extremely well but current computers do it poorly. The hope is that even a partial understanding of how pattern recognition is done in a neural network, will prove helpful in designing computers that solve the problem. Neural computers derive their potential advantages largely from the fact that they are specialized. A particular neural computer will be a machine that is tuned to solve a specific set of problems. Experimental versions of such neural circuits have been built using either both optics and electronics. Optics is a technology that is particularly well suited for building neural computers because of the extensive connectivity it can provide. A neural optical computer can be built by arranging the neurons in a planar geometry and using the third dimension to globally interconnect the neural planes with light. It is the relative easy access to the third dimension that we have in an optical system that gives this technology an edge over electronic neural nets.

Overview of Switching System/Network Architectural Possibilities

Tadahiko Yasui and Katsuaki Kikuchi

NTT, Electrical Communications Laboratories,
3-9-11, Midori-cho, Musashino-shi, Tokyo, 180, Japan

1. Need for optical switching systems

To ensure the provision of high-speed, broadband services, new telecommunications networks should be constructed using both high-speed, broadband switching systems as well as high-speed transmission systems.

Optical technology has become a practical reality for high-speed transmission systems, and optical fiber transmission systems are rapidly replacing metallic-cable in long-haul transmission systems. As the bit-rate of optical fiber transmission systems increases beyond 1Gbps and the repeater-spacing continues to increase, it becomes more and more necessary to develop high-speed switching systems. It is difficult, however, to fabricate high-speed switching facilities using conventional electrical technologies, because the bandwidth and cross-talk problems inherent in electrical technologies do not allow high-speed operations. A very promising candidate for solving these problems is a high-speed optical time-division switching system in which it is not necessary to convert the high-speed optical signals to low-speed electrical signals.

It is also expected that optical fibers will penetrate into the present copper-wire subscriber loop networks. Therefore, optical fibers are essential for the provision of enhanced high-speed, broadband services of the future. In optical subscriber networks, two different switching functions must be performed; one is line concentration for broadband bidirectional services, and the other signal distribution for CATV services. Optical technologies have made space-division switching networks almost independent of transmission bit-rate. Such networks are suitable for realizing the two functions mentioned above. Optical line-concentration will contribute very much to the construction of cost-effective optical subscriber loop networks.

An optical communications network can be realized by combining an optical switching system with optical fiber transmission lines without the need for electonic-optical and optical-electronic converters. Such optical communications networks are expected to offer both enhanced functions and high performance in the provision of high-speed, broadband services of the future.

2. Switching Network Architectural Possibility

To fully realize high-speed time division switching networks, it is necessary to clarify the potential advantages of optical switching networks over electronic switching networks.

The relevant Optical and electrical technologies are shown together in relation to device fabrication and interconnection techniques in Fig. 1.

Optical signal transmission features low loss, wide bandwidth and non-inductiveness, while electrical signal transmission has a speed limit due to the product of the resistance and capacitance in the electrical line. Optical interconnection exhibits excellent broadband transmission characteristics, and is essential for a high-speed operation. Optically controlled optical devices (OCOD) are capable of attaining a very high switching speed on the order of picoseconds [1]. In contrast, electrically controlled optical devices (ECOD) cannot achieve such high-speed operation, because the switching speed in such devices is limited by electrical control signals. Therefore, an optical time division switching system with OCODs and optical interconnection will be able to realize a very high-speed operation.

In such synchronous systems as time division switching networks, one of the most difficult problems arising from high-speed operation is clock skew caused by differences in propagation delay time among optical signals through different optical paths. One solution to this problem is to use a two-dimensional beam steering technique [2] which exploits the a non-interaction feature of photons. Two-dimensional arrays of optical memories are incorporated in this novel switching networks.

Electronic devices can realize virtually the same speed of operation as ECODs. Therefore, for the time being, time division switching networks are likely to be constructed with opto-electronic integrated circuits (OEICs), which incorporated the best advantages of both optical and electrical technologies – electronic logic circuits and optical interconnections. Wavelenghth division technologies, however, make it possible to extend the throughput of a switching network without increasing the operating speed. Thus, optical switching systems with ECODs also have their own merits over conventional electronic switching systems with OEICs.

Optical technology has the great advantage of being able to transfer two-dimensional images using fiber bundles or graded index fibers without the need to convert into electrical signals. In this application, image switching networks are required to exchange images transmitted through such transmission media.

3. Optical Switching Technological Possibility

Eight-by-eight optical matrix switches, which are the basic components in the construction of an optical space division switching system, have already been demonstrated. These switches appear to have the capability required by small-size optical switching systems. Recently, some small-size system experiments using such optical matrix switches have been reported.[3],[4] To realize large-size optical switching systems, such as telecommunications switching networks, however, it is necessary to clarify that the problems of loss accumulation and cross talk

can be overcome.

The key devices supporting the development of optical time division switching networks are high-speed optical switches and memories. Many studies of optical bistable devices are currently being conducted in a number of countries. Recently, semiconductor optical bistable devices based on either laser diodes or multiple quantum well structures have been attracting special interest. Optical time switches using either fiber delay lines or bistable laser diodes as optical memories have been demonstrated. [5],[6],[7] Furthermore, optical retiming and regenerating techniques are essential to synchronizing high-speed optical signals.

References
[1] P.W.Smith, "On the Physical Limit of Digital Optical Switching and Logic Elements", Bell Syst. Tech. Jour., Vol.61, 1982

[2] P.Gravey, et al., "Optical Switching Techniques for High Capacity Exchanges", ISS'84, 1984

[3] S.Suzuki, et al., "An Experiment on 32-lines Optical Space-division Switching", Natl. Conv. IECE Japan, 156, 1986.

[4] T.Matsunaga, et al., "Optical Space-division Switching System", ISS'87, 1987. (To be published)

[5] T.Matsunaga, et al., "Experimental Application of LD Switch Modules to 256Mb/s Optical Time Switching", Electron. Lett., Vol.21, No.20, 1985.

[6] R.A.Thompson, et al., "Experimental Photonic Time-slot Interchanger Using Optical Fibers as Reentrant Delay-line Memories", OFC'86, 1986.

[7] S.Suzuki, et al., "An Experiment on High-speed Optical Time-division Switching", IEEE J.Lightwave Tech., Vol.LT-4, No.7, 1986.
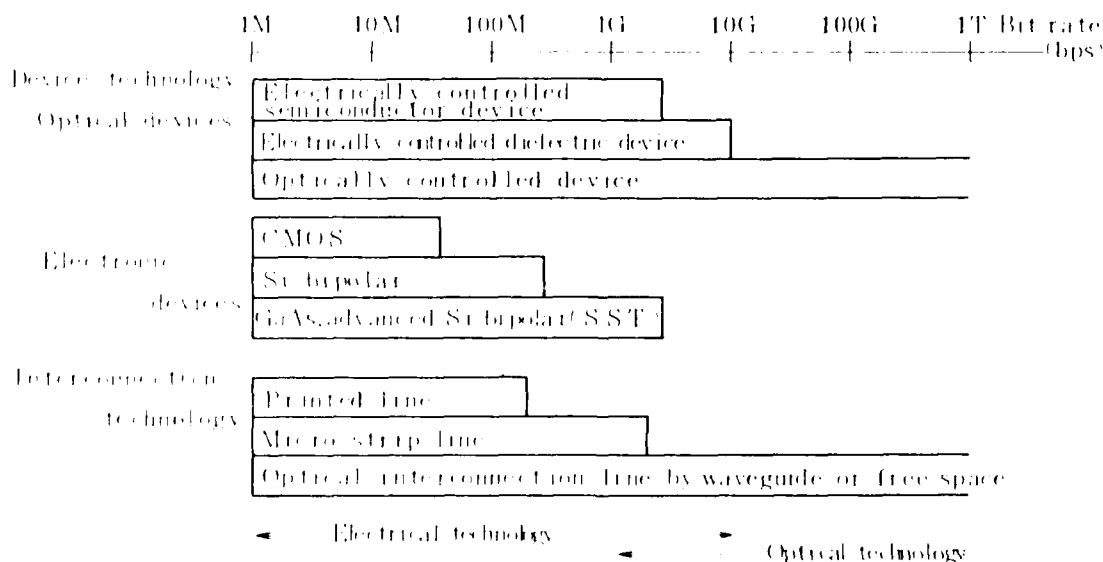
Fig. 1. Comparison of optical and electrical technologies

Abul-Moustafa, Y — MA3
Antony, S — ME8
Armitage, D — TuE3
Athale, Ravindra A — TuB2

Beaudet, P R — MD3
Bennion, I — TuC5
Bhattacharya, Pallab K — TuE5
Bradley, J C — MD3
Brenner, Karl-Heinz — MB1, MB3

Campbell, R J — TuE1
Capps, C David — MD4, TuD6
Casasent, David — TuB3
Cathey, W Thomas — MD7, TuC4
Chavel, P — MC2, MD9
Chavez-Pirson, Arturo — TuE5
Cheng, Li-Jen — TuD5, TuE8
Clark, N A — TuC4, TuE3
Cloonan, Thomas J — MB4
Collins, Jr , Stuart A — MD5, ME5
Craft, N C — TuC2, TuC3
Cush, R — TuC5
Cuykendall, Robert — MD6

Davies, D K — MD3
Derstine, M W — TuA6
Devos, F — MC2, MD9
Dubois, F — TuD4

Ebbeni, J — TuD4
Esener, Sadik C — ME6
Fainman, Y — MC6, TuA6
Falk, R Aaron — MD4, TuD6
Farhat, Nabil H — TuA2
Farhoosh, H — MC6, ME3
Feldman, Michael R — MC6, ME2, ME6
Fiddy, M A — MC4
Fisher, Arthur D — TuA5

Garmire, Elsa — TuD5
Gibbons, G G — TuD4
Gibbs, Hyatt M — TuC1, TuE4, TuE5
Goodman, J W — MA1
Goutzoulis, A P — MD3
Grinberg, J — TuB5
Gu, Xiang Guang — TuA3
Guest, Clark C — MC6, ME2, ME6, ME7
Guha, A — TuA6
Guilfoyle, P S — MD2

Habiby, S F — MD5
Handschy, M A — TuC4, TuE3
Haney, Michael W — TuB2, TuB4
Hartman, Davis H — MC5
Haugen, P R — TuA6

Hillis, W Daniel — MD1
Hossack, W J — ME1
Houk, Theodore L — MD4, TuD6
Huang, K S — MB5
Huignard, J P — TuD4
Husain, A — TuA6

Iijima, K — ME10
Ikeda, Makoto — TuE7
Ishihara, Satoshi — TuE7

Jau, Jack Y — TuA6
Jenkins, B K — MB5
Jeon, Ho-In — ME9
Johnson, K M — MD7, TuC4
Jordan, Harry F — MB6
Juang, Feng Yuh — TuE5

Kaya, Junichi — TuE7
Koch, S W — TuE4
Krile, Thomas F — ME4
Kung, H T — MA2

Lau, Suzanne — TuD2
Le, P — TuD3
Lee, Hyuk — TuA3
Lee, John N — TuA5
Lee, Sing H — MC6, ME3, ME6, TuA6
Lee, Yong H — TuE5
Lightbody, M T — MC4
Lin, Shing-Hong — ME4
Liu, Hua-Kuang — TuE8
Lohmann, Adolf W — MC1, MC3
Loiseaux, B — TuD4

Madani, K — MD9
Mahalanobis, Abhijit — TuB3
Mait, Joseph N — MB3
Malarkey, E C — MD3
Marhic, M E — TuD7
Marom, E — TuB5
Martino, Anthony J — MD8
Mathew, J G H — TuE1
Matsuoka, H — ME10
Mayo, S — TuD4
McMillin, Debra — MD6
Mitsuhashi, Yoshinobu — TuE7
Monro, D M — TuE9
Morhange, Jean F — TuE5
Morris, G Michael — MD8
Murdocca, M J — MB2

Nicholson, M G — TuD4

Ojima, Masahiro — TuE5
Owechko, Y — TuB5

258

FACTORED LOOK-UP TABLES FOR OPTICAL
RESIDUE NUMBER SYSTEM (RNS) COMPUTATIONS[**]

by

E.C. Malarkey, F.R. Beaudet, J.C. Bradley
Westinghouse Electric Corporation
Advanced Technology Division
Box 1521
Baltimore, Maryland 21203

and J.L. Davis
Rome Air Development Center, DCLW
Griffiss Air Force Base
New York 13441-5700

ABSTRACT

The details of factored tables and their uses in RNS multiplication and addition are explained. Optical circuitry necessary for their implementation is given.

INTRODUCTION

Residue number systems offer reasonable factorizations into parallel computations. This feature has two distinct advantages; it provides a means of exact calculations and a reduction in parts count in the associated circuitry. In this paper we discuss a second level of factorization -- factored look-up tables. The reduction in hardware that results from the use of these tables is significant.

The development presented in this paper begins with a brief description of optical look-up tables; there is a more extensive description of these tables given in reference 1. Second, the basic notions of factorization of tables is given. The remaining entries left in a multiplication table after deleting zero is a group of elements that can be factored into smaller multiplication tables: a zero entry must be handled separately. With some added complexity this idea is extended to addition.

Although table factorization is a general concept, it is best explained by way of an example. We do this here in some detail for modulo 7 addition and multiplication. The paper concludes with a parts count necessary for constructing factored modulo 31 tables. Estimates on the amount of hardware needed to solve a twelfth order linear system are reported.

LOOK-UP TABLES (LUTs)

Numerical operations can be performed in residue arithmetic simply by causing a light pulse to reach a detector that has been

encoded for the number resulting from each operation. The idea is illustrated in figure 1 where the LUT for modulo 5 multiplication is shown, and inputs of 3 and 2 are depicted by arrows. At the intersection of the inputs, light produced by one or another means excites a properly encoded detector: in the figure, the product 3×2(mod 5) = 1 sends a pulse to a detector labeled 1. Similarly, the sum 3+2 modulo 5 would illuminate a detector encoded for 0 in an addition LUT.

Any of a variety of approaches to implementation of the LUTs can be envisioned, one of the simplest being that illustrated for a modulo 5 multiplication table in figure 1. It uses an interlaced electrode grid with high speed LEDs (or LDs) at the intersection points. A voltage pulse applied to each input line, with voltages selected so that neither alone exceeds the diode junction voltage but the sum exceeds it by a considerable margin, causes the diode at the intersection point to emit strongly. The emitted light goes to a detector that is encoded for the number to be produced at that table location, as indicated by the number in the lower left corner of each grid box. To minimize the number of detectors required and to promote flexibility in LUT geometry, we use fibers to conduct light from each LED to the proper detector and use a single detector every time a given digit appears in the table.

LUT FACTORIZATION AND IMPLEMENTATION

Given a prime p, an LUT for multiplication modulo p is a p×p table with entries $0,1,\ldots,p-1$. If 0 can be detected by an independent means, then the remaining non-zero elements in the table forms the cyclic group labeled $Z_p^* = \{1,2,\ldots,p-1\}$. The number p-1 is composite and has a prime factorization $q_1^{n_1}\cdots q_s^{n_s}$; the group Z_p^* can be factored into a set of cyclic subgroups, one of order $q_1^{n_1}$, a second of order $q_2^{n_2}$, etc. Consequently, the complexity in the p×p table can be reduced because it can be replaced by a set of smaller tables that involve significantly fewer LEDs and detectors. A similar, but somewhat more complicated method can be used for addition tables.

In carrying out the reduction for addition two basic approaches are used. One, the direct method, involves the use of factored tables together with an auxilliary 2×2 table that is used somewhat like a limited carry to execute the addition calculation directly in modulo p arithmetic. The second method uses logarithmic and exponential functions as well as several wired maps (1×p tables) to complete the modulo p additions. Multiplication is similar for the two methods. They will be illustrated with an example of mod 7 arithmetic.

We dicuss multiplication first. The direct method uses Z_7^* = GxH which is genrated by 3 where the subgroups G = {1,6} and H = {1,2,4} are generated by 6 and 2, respctively. The logarithmic method uses replicas of these groups which are additive. In particular, Z_p^* is isomorphic to $Z_6 = Z_2 \times Z_3$. The logarithm $\log_3$

maps Z_7^* onto Z_σ where $3^\sigma = 3^0 = 1$. In terms of the factored groups, $(\log_\sigma, \log_2)$ maps $G \times H$ onto $Z_2 \times Z_3$. These relations are used to develop encoding for multiplication:

Table I. Encoding for Multiplication

Number	Direct p, p'	log ℓ, ℓ'
$1 = 3^\sigma$	1, 1	0, 0
$2 = 3^2$	1, 4	0, 2
$3 = 3^1$	6, 2	1, 1
$4 = 3^4$	1, 2	0, 1
$5 = 3^5$	6, 4	1, 2
$6 = 3^3$	6, 1	1, 0

The direct method for addition uses the additive subgroups $\{0, 3\}$ and $\{0, 2, 4\}$ and the auxilliary table

	ℓ	u
ℓ	ℓ	b
u	b	u

In terms of these quantities the following encoding is used for addition:

Table II. Encoding for Direct Addition

1	$(\ell, 3, 2)$
2	$(\ell, 0, 4)$
3	$(\ell, 3, 0)$
4	$(u, 0, 2)$
5	$(u, 3, 4)$
6	$(u, 0, 0)$

The rules for decoding in direct addition are as follows: (i) The output $(b, 3, 2)$ is a zero flag (note that $3, 2$ never occurs with ℓ or u; (ii) If $(\ell, -, -)$ occurs, then the numerical characters are correct; (iii) If $(u, -, -)$ occurs, then overflow has occurred and the numerical characters must be shifted down by 1; i.e., (iiia) $3 \to 0$, $0 \to 3$, and (iiib) $4 \to 2$, $2 \to 0$, $0 \to 4$; (iv) If $(b, -, -)$ appears, then overflow has occurred if the output corresponds to the lower half range, but not if the upper half range results. Digits corresponding to 4, $(0, 2)$, do not occur with b, and those for 3, $(u, 3, 0)$, and 4, $(u, 0, 2)$, do not occur after B decoding. The following sequential multiply - add sequence is carried in modulo 7 arithmetic using the direct method: We compute $5 \times 4 + 6 \times 3 + 2 \times 5$ modulo 7. First,

$$5 \times 4 \longrightarrow (6, 4) \cdot (1, 2) = (6, 1) \longrightarrow (u, 0, 0)$$
$$6 \times 3 \longrightarrow (6, 1) \cdot (6, 2) = (1, 2) \longrightarrow (u, 0, 2)$$

Then,

$$(u, 0, 0) + (u, 0, 2) = (u, 0, 2) \longrightarrow (\ell, 3, 0)$$

and,

$$2 \times 5 \longrightarrow (1, 4) \cdot (6, 4) = (6, 2) \longrightarrow (\ell, 3, 0)$$

Adding the last two expressions gives $(\ell, 0, 0)$ or, the answer, 6.

Before we give details of the mod 7 computation using the logarithmic method we explain general addition modulo p by this method. In addition formulas, addition modulo p occurs on the base line while that in an exponent of b (a generator of Z_p^*) is computed modulo p-1. In this case there are unique numbers α and β allowing us to write

$$x + y = b^\alpha + b^\beta = b^\alpha \cdot (1 + b^{\alpha - \beta}) = b^{\alpha + \gamma} = z$$

where γ is found using a wired map defining the relation $1 + b^{\alpha-\beta}$
$= b^{\gamma}$. Resuming the mod 7 calculations where $b = 3$ we get

$$5\times4 \longrightarrow (1,2) + (0,1) = (1,0)$$
$$6\times3 \longrightarrow (1,0) + (1,1) = (0,1)$$
$$2\times5 \longrightarrow (0,2) + (1,2) = (1,1)$$

Then,

$$(1,0) + (0,1) = (1,0)\cdot(1 + (0,1)-(1,0))$$
$$= (1,0)\cdot(1 + (1,1))$$
$$= (1,0) + (0,1) \quad \text{(adding exponents)}$$
$$= (1,1).$$

Finally, this calculation concludes with
$$(1,1) + (1,1) = (1,1)\cdot(1 + (1,1)-(1,1))$$
$$= (1,1)\cdot(1 + (0,0))$$
$$= (1,1) + (0,2) \quad \text{(adding exponents)}$$
$$= 6.$$

The parts count for both methods is given for modulo 31 since it is typical of the primes that will be used in this work. For the direct method 66 detectors, 168 LEDs and 456 optical interconnects are required. This compares with 72 detectors, 159 LEDs and 220 optical interconnects for the logarithmic calculation. The number of time cycles necessary to carry through an add-multiply for the direct method is 3 using mod 31 arithmetic; the time cycles for the logarithmic calculation is 6. The number of "gates" (i.e., detectors or LEDs) used in optical factored tables with an RNS is roughly 14% of the number of gates used in a digital electronic system to solve a twelfth order linear algebraic system using Gauss' method.

REFERENCE

1. P.R. Beaudet, A.P. Goutzoulis, E.C. Malarkey and J.C. Bradley, "Residue Arithmetic Techniques for Optical Processing of Adaptive Phased Array Radars," Appl. Opt. 21,3097 (1986).
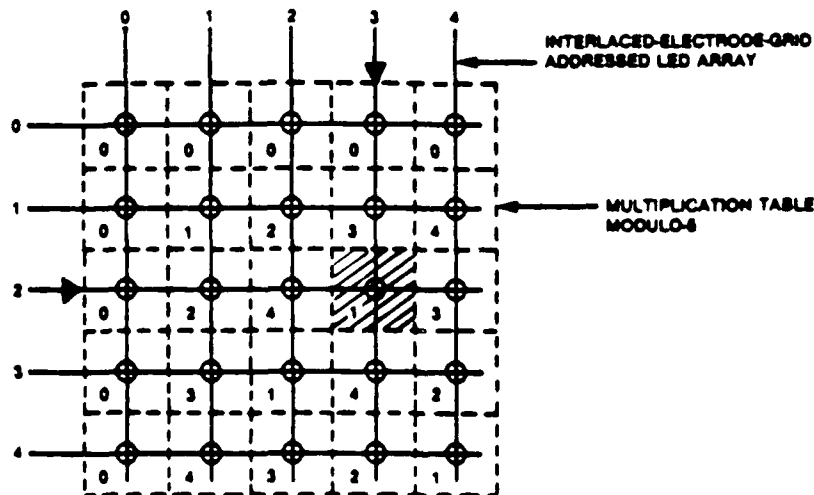

Figure 1. Schematic representation of a modulo 5 multiplication table with LED light sources as

END

DATE

FILMED

8-88

DTIC