

LOAN DOCUMENT

PHOTOGRAPH THIS SHEET

AD-A261 267



DTIC ACCESSION NUMBER

LEVEL

INVENTORY

AFOSR-TR-92-0517

DOCUMENT IDENTIFICATION

22 May 92

DISTRIBUTION STATEMENT
Approved for public release
Distribution Unlimited

DISTRIBUTION STATEMENT

ACCESSION FOR	
NTIS	GRA&I <input checked="" type="checkbox"/>
DTIC	TRAC <input type="checkbox"/>
UNANNOUNCED	
JUSTIFICATION	
BY	
DISTRIBUTION/	
AVAILABILITY CODES	
DISTRIBUTION	AVAILABILITY AND/OR SPECIAL
A-1	

DISTRIBUTION STAMP

DTIC QUALITY INSPECTED 3

DTIC
ELECTE
S FEB 5 1993 **D**
C

DATE ACCESSIONED

DATE RETURNED

DEFENSE TECHNICAL INFORMATION CENTER



9301943

327P8

DATE RECEIVED IN DTIC

REGISTERED OR CERTIFIED NUMBER

PHOTOGRAPH THIS SHEET AND RETURN TO DTIC-FDAC

H
A
N
D
L
E

W
I
T
H

C
A
R
E

93 22 039

Sponsored by
Optical Society of America

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank) 2. REPORT DATE
May 22, 1992 3. REPORT TYPE AND DATES COVERED
Final 1/1/91-12/31/91

4. TITLE AND SUBTITLE
Organization of the 1991 Optical Society of America Photonic Science Topical Meeting Series Vol 6 5. FUNDING NUMBERS
G - AFOSR-91-0176

6. AUTHOR(S)
Jarus W. Quinn

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Optical Society of America
2010 Massachusetts Ave. NW
Washington, DC 20036
AFOSR-TR-92-0017 8. PERFORMING ORGANIZATION
REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Air Force Office of Scientific Research
Department of the Air Force
Bolling Air Force Base
Washington, DC 20332-6448
NE Schlossberg 10. SPONSORING/MONITORING
AGENCY REPORT NUMBER
2305/A1

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release
Distribution unlimited 12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)
Attach list of reports supported by Optical Society of America
Photorefractive Materials, Effects, and Devices
Integrated Photonics Research
Nonlinear Guided Wave Phenomena
Optical Amplifiers and Their Applications
Optical computing
Picosecond Electronics and Optoelectronics
Quantum Optoelectronics
Photonic Switching
Microphysics of Surfaces: Beam Induced Processes
Soft X-ray Projection Lithography
Short Wavelength Coherent Radiation, Generation & Applications
Persistent Spectral Hole-Burning: Science & Applications

14. SUBJECT TERMS 15. NUMBER OF PAGES
16. PRICE CODE

17. SECURITY CLASSIFICATION 18. SECURITY CLASSIFICATION 19. SECURITY CLASSIFICATION 20. LIMITATION OF ABSTRACT

AEOSR-TR- 92 05 17

Summaries of papers presented at the
Optical Computing Topical Meeting

Optical Computing

1991 Technical Digest Series
Volume 6

CONFERENCE EDITION
March 4-6, 1991 — Salt Lake City, Utah



AIR FORCE OF SCIENCE (AFOSR)
NOTICE: This document is the property of the Air Force Office of Scientific Research (AFOSR) and is loaned to you for your use only. It is not to be distributed outside your organization without the prior written permission of AFOSR. This document is to be returned to AFOSR when requested. This document is not to be used for any purpose other than the one for which it was loaned to you. This document is not to be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage and retrieval system, without the prior written permission of AFOSR. This document is not to be used for any purpose other than the one for which it was loaned to you. This document is not to be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage and retrieval system, without the prior written permission of AFOSR.

Sponsored by
Optical Society of America

Optical Society of America
2010 Massachusetts Avenue, NW
Washington, DC 20036

93 2 2 039

Articles in this publication may be cited in other publications. In order to facilitate access to the original publication source, the following form for the citation is suggested:

Name of Author(s), Title of Paper, in Optical Computing, 1991,
Technical Digest Series, Vol. 6 (Optical Society of America,
Washington, DC 1988), pp. xx-xx.

ISBN Number

Conference Edition	1-55752-170-0 (softcover)
Postconference Edition	1-55752-171-9 (hardcover)
(Note: Postconference Edition includes postdeadline papers.)	
1991 Technical Digest Series	1-55752-192-1 (hardcover)

Library of Congress Catalog Card Number

Conference Edition	90-63029
Postconference Edition	90-63028

Copyright © 1991, Optical Society of America

Individual readers of this digest and libraries acting for them are permitted to make fair use of the material in it, such as to copy an article for use in teaching or research, without payment of fee, provided that such copies are not sold. Copying for sale is subject to payment of copying fees. The code 1-55752-192-1/91/\$2.00 gives the per-article copying fee for each copy of the article made beyond the free copying permitted under Sections 107 and 108 of the U.S. Copyright Law. The fee should be paid through the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970.

Permission is granted to quote excerpts from articles in this digest in scientific works with the customary acknowledgment of the source, including the author's name and the name of the digest, page, year, and name of the Society. Reproduction of figures and tables is likewise permitted in other articles and books provided that the same information is printed with them, permission of one of the original authors is obtained, and notification is given to the Optical Society of America. Republication or systematic or multiple reproduction of any material in this digest is permitted only under license from the Optical Society of America; in addition, the Optical Society may require that permission also be obtained from one of the authors. Address inquiries and notices to the Director of Publications, Optical Society of America, 2010 Massachusetts Avenue, NW, Washington, DC 20036. In the case of articles whose authors are employees of the United States Government or its contractors or grantees, the Optical Society of America recognizes the right of the United States Government to retain a nonexclusive, royalty-free license to use the author's copyrighted article for United States Government purposes.

Printed in the U.S.A.

CONTENTS

Advance Program	v
MA Optoelectronic Components	1
MB Micro-Optics.....	23
MC Optical Interconnections.....	37
MD Spatial Light Modulators.....	53
ME Poster Preview	59
TuA Digital Systems	189
TuB Fuzzy and Cellular Systems.....	207
TuC Memory Issues.....	227
TuD Architectures and Signal Processors.....	243
WA Optical Learning Systems.....	265
WB Neural Network Components.....	285
WC Joint Plenary Session on Optical Computing and Photonic Switching: I	303
WD Joint Plenary Session on Optical Computing and Photonic Switching: II	309

Optical Computing
TECHNICAL PROGRAM COMMITTEE

C. Lee Giles, *General Cochair*
NEC Research Institute

Sing H. Lee, *General Cochair*
University of California, San Diego

B. Keith Jenkins, *Program Chair*
University of Southern California

H. Arsenault
University Laval, Canada

R. A. Athale
George Mason University

L. A. Coldren
University of California, Santa Barbara

M. Dagenais
University of Maryland

U. Efron
Hughes Research Laboratory

I. Glaser
Tel Aviv University, Israel

L. Hesselink
Stanford University

H. S. Hinton
AT&T Bell Laboratories

A. Husain
Honeywell Systems and Research Center

Y. Ichioka
Osaka University, Japan

S. Ishihara
Electrotechnical Laboratory, Japan

R. K. Kostuk
University of Arizona

K. Kyuma
Mitsubishi Electric Corporation, Japan

F. J. Leonberger
United Technologies Research Center

A. W. Lohmann
NEC Research Institute

W. J. Miceli
Office of Naval Research

J. Midwinter
University College London, UK

J. A. Neff
Dupont Corporation

M. E. Prise
AT&T Bell Laboratories

G. Roosen
Institut D'Optique, France

C. M. Verber
Georgia Institute of Technology

K. H. Wagner
University of Colorado

C. Warde
Massachusetts Institute of Technology

P. Yeh
University of California, Santa Barbara

Monday, March 4, 1991

Salon D

7:00 am–8:30 am Buffet Breakfast

Grand Ballroom Foyer

7:00 am–5:30 pm Registration/Speaker and Presider Check-in

Salon F

8:15 am–8:30 am

Opening Remarks

C. Lee Giles, *NEC Research Institute, General Cochair*;

Sing H. Lee, *University of California, San Diego, General Cochair*

8:30 am–10:20 am

MA, Optoelectronic Components

Mario Dagenais, *University of Maryland, Presider*

MA1

8:30 am (Invited)

Progress in arrays of optoelectronic bistable devices and sources, K. Kasahara, I. Ogura, Y. Yamanaka, *NEC Corporation, Japan*. Recent progress in vertical-to-surface transmission electro-phonic devices and the resultant optical functional interconnections will be presented. . . . 2

MA2

9:00 am

Integrated array of self-electro-optic effect device logic gates, A. L. Lentine, L. M. F. Chirovsky, M. W. Focht, J. M. Freund, G. D. Guth, R. E. Leibenguth, G. J. Przybylek, L. E. Smith, L. A. D'Asaro, D. A. B. Miller, *AT&T Bell Laboratories*. We demonstrate a 16×16 array of batch-fabricated SEED CMOS-like logic gates and discuss the advantages of this type of optical logic gate. . . . 6

MA3

9:20 am

Binary arithmetic using optical symbolic substitution and cascable surface-emitting laser logic devices, Julian Cheng, G. R. Olbright, R. P. Bryan, *University of New Mexico, Sandia National Laboratories*. Cascable optical logic based on heterojunction phototransistors and vertical-cavity surface-emitting lasers is demonstrated. We discuss a scheme for implementing binary arithmetic by using optical symbolic substitution. . . . 10

MA4

9:40 am

Reliability of optical logic, Charles W. Stirk, Demetri Psaltis, *California Institute of Technology*. The reliability of optical logic depends on fan-in, contrast ratio, and noise. We calculate the fundamental and practical BER for optical devices and multilayer circuits. . . . 14

MA5

10:00 am

Optical binary multiplication based on a non-holographic content-addressable memory, Andrew Kostrzewski, George Eichmann, Dai Hyun Kim, Yao Li, *Physical Optics Corporation, City College of the City University of New York*. A new fast binary multiplication scheme based on a non-holographic optical content-addressable memory and a sign/logarithm number system is presented. The design and experimental demonstration of a 7-bit multiplier are presented. . . . 18

Salon D

10:20 am

Coffee Break

Salon F

10:50 am–12:20 pm

MB, Micro-Optics

Adolf W. Lohmann, *NEC Research Institute, Presider*

MB1

10:50 am (Invited)

Binary optics and applications, Wilfrid B. Veldkamp, *Massachusetts Institute of Technology*. In a classic example of technology transfer, binary optics is allowing optical designers to create innovative optical components which promise to solve key problems in optical sensors, communication, and optical processors. . . . 24

MB2**11:20 am**

Three-dimensional integration of digital optical systems, K.-H. Brenner, *Universität Erlangen-Nürnberg, Federal Republic of Germany*.

Complex digital optical systems require methods for integration. Planar integrated optics excludes many of the advantages of optics. Concepts and technologies for a three-dimensional optical integration are proposed.25

MB3**11:40 am**

Integrated free-space optical permutation network, Jürgen Jahns, Walter Däschner, *AT&T Bell Laboratories*. An optical implementation of a permutation network is demonstrated that uses free-space optical light propagation inside a single glass substrate. Diffractive lenses are etched into the substrate to provide beam steering of the light beams.29

MB4**12:00 m**

Optical bus interconnection system using SELFOC lenses and planar microlens arrays, Kenjiro Hamanaka, *Nippon Sheet Glass Company, Ltd., Japan*. A novel optical bus interconnection system using SELFOC lenses and planar microlens arrays has been proposed. Features and possible applications are discussed with experimental results.32

12:20 pm—2:00 pm Lunch Break**Salon F****2:00 pm—3:20 pm****MC, Optical Interconnections**

Ravindra A. Athale, *George Mason University, Presider*

MC1**2:00 pm**

Spatial noise reduction in array illuminators, Adolf W. Lohmann, Stefan O. Sinzinger, *NEC Research Institute, Inc., Physikalisches Institut der Universität, Federal Republic of Germany*. An array illuminator provides an array of optical gates or smart pixels with photon power. Reducing the coherence improves the homogeneity of the beamlet array.38

MC2**2:20 pm**

Cellular hypercube interconnections for optical processor arrays, C. B. Kuznia, A. A. Sawchuk, *University of Southern California*. We discuss communication times versus detectors per cell for cellular hypercube interconnections in optoelectronic fine-grain cellular arrays and their implementation with binary-phase gratings. . . . 41

MC3**2:40 pm**

Multiplexed hybrid interconnection architectures, Haldun M. Ozaktas, Joseph W. Goodman, *Stanford University*. We discuss methods of organizing information flow in computation in a manner that permits the multiplexing of signal paths with distinct sources and destinations. 45

MC4**3:00 pm**

Two-dimensional spatially variant optical interconnects, E. J. Restall, B. Robertson, M. R. Taghizadeh, A. C. Walker, *Heriot-Watt University, UK*. A volume holographic approach to such two-dimensional optical interconnects as the banyan, butterfly, half cross over, and perfect shuffle is described. Prototype networks that are compatible with current demonstration optical circuits are also discussed. 49

Salon D**3:20 pm—3:50 pm Coffee Break****Salon F****3:50 pm—4:40 pm****MD, Spatial Light Modulators**

Uzi Efron, *Hughes Research Laboratory, Presider*

MD1**3:50 pm (Invited)**

Some practical issues in design and fabrication of high-contrast quantum-well modulator arrays, G. Parry, M. Whitehead, E. Zouganeli, A. Rivers, K. Woodbridge, J. S. Roberts[†], *University College London, UK*, [†]*University of California, Santa Barbara*, [†]*University of Sheffield, UK*. Asymmetric Fabry-Perot modulators offer the prospect of high contrast (>20 dB) and low voltage (<5 V) as well as useful optical bandwidths. This paper will discuss the practical problems of designing and fabricating arrays of devices to these specifications. 54

MD2

4:20 pm

Design and fabrication of VLSI ferroelectric liquid-crystal spatial light modulators, David A. Jared, Richard Turner, Kristina M. Johnson, *University of Colorado, Boulder*. Issues surrounding the design and fabrication of a 64×64 DRAM spatial light modulator (SLM) and three 32×32 optically addressed SLMs are presented. . . .55

Salon F

4:40 pm–5:40 pm

ME, Poster Preview

Lee Giles, *NEC Research Institute, President*

Salon D

6:30 pm–8:00 pm

Conference Reception

Salon F

Salon D

7:30 pm–9:00 pm

ME Poster Session

ME1

Huge optical amplification by applying pulsed electric fields to photorefractive crystals, P. Mathey, G. Pauliat, J. C. Launay, G. Roosen, *Centre National de la Recherche Scientifique, France*. The wave-mixing gain is considerably enlarged by applying a pulsed electric field. Unlike when other enhancement techniques are used, the photorefractive gain is no more limited by crystal trap densities. . . .60

ME2

Enhanced photorefractive effects with a dc field and moving grating in GaP at 633 nm, Jian Ma, Yoshinao Taketomi, Yeshaiah Fainman, Joseph E. Ford, Sing H. Lee, Ken'ichi Chino, *University of California, San Diego, Sumitomo Metal & Mining Company, Ltd., Japan*. We demonstrate that the photorefractive effect in GaP crystals at 633 nm can be enhanced by using an externally applied dc field and a moving grating. Two-beam coupling gain of 1.9 cm^{-1} and phase-conjugate reflectivity of 4.5% were obtained. . . .64

ME3

Optical thresholding and Max operation, Claire Gu, Pochi Yeh, *Rockwell International Science Center, University of California, Santa Barbara*. Self-oscillations in nonlinear-optical four-wave mixing and resonators are considered. Some unique properties of these oscillations can be used to implement parallel optical thresholding, comparing, and Max operation. . . . 68

ME4

Gray-scale controllable ferroelectric liquid-crystal spatial light modulator, Cleber M. Gomes, Susumu Tsujikawa, Hiroki Maeda, Hiroyuki Sekine, Takashi Yamazaki, Mikio Sakamoto, Fujio Okumura, Shunsuke Kobayashi, *Tokyo University of Agriculture and Technology, Japan, NEC Corporation, Japan*. Memorized gray-scale capability has been demonstrated in a ferroelectric liquid-crystal spatial light modulator prepared by using polyimide Langmuir-Blodgett films to orient the liquid crystal. . . . 72

ME5

Optoelectronic neuron, Anton Rohlev, Christian Radehaus, Jacques I. Pankove, R. F. Carson, G. Borghs[†], *University of Colorado, Boulder, Sandia National Laboratories, †IMEC, Belgium*. This integratable semiconductor optoelectronic neuron has electrical and optical input/output, is endowed with memory, exhibits inhibition and enhancement of sensitivity, and can have weighted synapses. . . . 76

ME6

Design and demonstration of an optoelectronic neural network using fixed planar holographic interconnects, Paul E. Keller, Arthur F. Gmitro, *University of Arizona*. Implementation of an optoelectronic Hopfield-style associative memory neural network is discussed with emphasis on the construction of an experimental system that uses binary amplitude holograms. . . . 80

ME7

Custom-designed electro-optic components for optically implemented, multilayer neural networks, M. G. Robinson, K. M. Johnson, D. Jared, D. Doroski, S. Wichart, G. Moddel, *University of Colorado, Boulder*. Presented is a novel amorphous silicon/ferroelectric liquid-crystal device for an optically implemented two-layer connectionist architecture. Results of device and system performance are described. . . . 84

ME8

Optical matrix-vector implementation of binary-valued backpropagation, Stephen A. Brodsky, Clark C. Guest, *University of California, San Diego*. An operational optoelectronic neural network based on the binary-valued backpropagation training algorithm was constructed. This adaptive system uses optical interconnectivity for associative recall.88

ME9

Experimental comparison of different associative memory techniques implemented optically by the same system architecture, K. J. Weible, N. Collings, W. Xue, G. Pedrini, R. Dändliker, *University of Neuchâtel, Switzerland*. The same optical architecture is used to compare the experimental performance of two different inhibitory neural systems (binary and gray scale) and a discrete binary correlator.92

ME10

Optical modular architectures for multilayer BAM with two-dimensional patterns, Soo-Young Lee, Hyuek-Jae Lee, Sang Yung Shin, *Korea Advanced Institute of Science and Technology, Korea*. Optical modular architectures based on both inner-product matrix formation and outer-product recall schemes for two-dimensional patterns are presented for multilayer BAM.96

ME11

Optical processing unit for relational data base operations, Pericles A. Mitkas, P. Bruce Berra, *Colorado State University, Syracuse University*. An optical processing unit based on spatial light modulators is used to perform a rich set of relational data base operations including projections, selections, and joins.100

ME12

Fault-tolerant computing on POEM, Dau-Tsuong Lu, Ting-Ting Y. Lin, Fouad E. Kiamilev, Sadik C. Esener, Sing H. Lee, *University of California, San Diego*. POEM is shown to be fault tolerant when using reprogrammed or reconfigurable optical interconnections. This is demonstrated by algorithmic testing, recovery, and reconfiguration using VHDL simulation.104

ME13

Optoelectronic full adder using a beam-scanning laser diode, Hideo Itoh, Seiji Mukai, Masahiko Mori, Masanobu Watanabe, Hiroyoshi Yajima, *Electrotechnical Laboratory, Japan*. A novel optoelectronic full adder with a simple configuration and fast operation has been implemented by using a single beam-scanning logic gate.108

ME14

Implementation of a fiber-optic delay line memory, Todd J. Soukup, Vincent P. Heuring, *University of Colorado, Boulder*. We describe the circuitry and design parameters for a 1024-bit fiber-optic delay line memory. It will be used in a 50-MHz bit serial optical computer currently under construction.112

ME15

Fan-out analysis of a low-skew clock distribution network with optical amplifiers, C.-S. Li, F. Tong, D. G. Messerschmitt, *IBM T. J. Watson Research Center, University of California, Berkeley*. The skew of a tree-structured optical clock distribution when using optical amplifiers is analyzed and is shown to have a large fan-out increase as compared with that of a single-stage distribution.116

ME16

Reconfigurable interconnects using computer-generated holograms and spatial light modulators, James E. Morris, Michael R. Feldman, *University of North Carolina at Charlotte*. A new method of implementing reconfigurable interconnect systems has been developed that involves combining high-frequency fixed computer-generated holograms with a small number of spatial light modulators.120

ME17

Demonstration of colored optical interconnects and implementation of a 2 x 2 optical crossbar switch with bistable diode laser amplifiers, Zeqi Pan, Mario Dagenais, *University of Maryland, College Park*. Wavelength-division demultiplexing is demonstrated by using bistable diode laser amplifiers. The performance of a generalized non-blocking 2 x 2 optical crossbar switch based on bistable diode laser amplifiers is studied.124

ME18

Compact crossbar switch for optical interconnects, Freddie Lin, Eva M. Strzelecki, William Liu, *Physical Optics Corporation*. A compact optical crossbar switch based on the concept of vector-matrix multiplication is realized by using waveguide gratingcoupler arrays and a spatial light modulator. 128

ME19

Techniques for implementation of high-speed free-space optical interconnections, Dean Z. Tsang, *MIT Lincoln Laboratory*. Differential current efficiencies of 8% have been achieved in free-space optical interconnections between 3-GHz transmitter and 1.5-GHz receiver modules on opposite boards without using micro-positioners. 132

ME20

Two-dimensional optical buses for massively parallel processing, Shigeru Kawai, Masanori Mizoguchi, *NEC Corporation, Japan*. Optical buses using two-dimensional waveguides for interconnections between processors are presented. Their principles are successfully demonstrated by using glass plates with concave lenses. 136

ME21

Hologram recording by a pulsed method on phase change media, Koutarou Nonaka, Yasuhide Nishida, Susumu Fujimori, *NTT Applied Electronics Laboratories, Japan*. A novel real-time hologram recording is presented. It gives media parameters for high-diffraction efficiency. A two-dimensional character pattern is successfully reconstructed by using GeTe alloy film. 140

ME22

Progress in diffractive phase gratings used for spot array generation, Rick L. Morrison, Sonya L. Walker, *AT&T Bell Laboratories*. Diffraction gratings have been incorporated into prototype digital optical logic systems to generate spot arrays. We review improvements that enhance the grating's performance. 144

ME23

Passive optical array generators, Mohammad R. Taghizadeh, Jari Turunen, Brian Robertson, Antti Vasara, Jan Westerholm, *Heriot-Watt University, UK, Helsinki University of Technology, Finland*. We demonstrate the largest and most efficient space-invariant array generators realized so far and show that a thin hologram can reconstruct different images at different wavelengths or angles of incidence. 148

ME24

Design rules for fan-out elements recorded as volume holograms, H. P. Herzig, D. Prongué, P. Ehbets, R. Dändliker, *University of Neuchatel, Switzerland*. The recording of efficient fan-out elements as thick volume holograms is analyzed by using coupled-wave theory. Criteria for high efficiency and uniformity of the fan out have been determined. 152

ME25

Polarization metrology for optical interconnects that use polarization beam combining, J. Larry Pezzaniti, Russell A. Chipman, *University of Alabama, Huntsville*. The imaging polarimeter is introduced as a polarization metrology tool to study optical interconnects by using polarization beam combining. A method to align the optical interconnects is summarized. 156

ME26

Photorefractive parallel matrix-matrix multiplier using a mutually incoherent laser array, John Hong, Pochi Yeh, *Rockwell International Science Center*. We demonstrate the operation of a parallel matrix-matrix multiplier by using a photorefractive crystal in conjunction with an array of mutually incoherent laser sources. 160

ME27

Figure of merit for pattern recognition filters, Ph. Refregier, *Thomson-CSF, France*. Optimal tradeoffs among Horner efficiency, correlation-peak sharpness, and noise robustness with explicit solutions provide a rigorous way to evaluate different filters. 165

ME28

Digital approach for pattern scale measurements, Joseph Rosen, Lior Dezialoshinski, Ehud Nahtomi, Joseph Shamir, *Technion—Israel Institute of Technology, Israel*. A shift-invariant scale detection system is composed of k optical correlators operating in parallel to create a digital word of $2^k - 1$ scale levels. Independent of the number of objects and their locations in the input plane, we can measure the locations and the sizes of all the input patterns simultaneously. The system was implemented experimentally on an optical correlator, and the results are presented. 169

ME29

Image correlation using photorefractive GaAs, Li-Jen Cheng, Duncan T. H. Liu, Keung L. Luke, Norman S. Z. Kwong, *California Institute of Technology, Ortel Corporation*. The characteristics of image correlation when using photorefractive GaAs as a real-time dynamic holographic matched filter-medium are discussed. 173

ME30

Filter generation in hybrid electro-optical correlators by using genetic algorithm, Uri Mahlab, Joseph Shamir, *Technion—Israel Institute of Technology, Israel*. A genetic algorithm is used to generate spatial filters for optical pattern recognition and classification. The procedure is implemented directly on a hybrid electro-optical system by using commercial liquid-crystal television as a binary transparency. Experimental results demonstrate the efficiency of this novel approach. 208

ME31

Hardware and software system design for hybrid optical-electronic signal processing, R. D. Griffin, J. N. Lee, *U.S. Naval Research Laboratory*. The performance of high-speed optical processors within host digital processing systems depends critically on hardware and software design. We describe the design of a matched-filter system. 231

ME32

Multichannel Bragg cells for optical computing applications, Dennis R. Pape, *Photonic Systems Inc.* Multichannel Bragg cells designed for optical computing applications will be discussed. Optical computing systems using multichannel Bragg cells for switching and processing will be described. 185

Tuesday, March 5, 1991

Salon D

7:00 am—8:30 am Buffet Breakfast

Grand Ballroom Foyer

7:30 am—8:00 pm Registration/Speaker and
Presider Check-in

Salon F

8:30 am—10:20 am

TuA, Digital Systems

Michael Prise, *AT&T Bell Laboratories, Presider*

TuA1

8:30 am

Paper to be announced. 190

TuA2

9:00 am

Hardware compiler for digital optical computing, Miles Murdocca, Vipul Gupta, Masoud Majidi, *Rutgers University*. A hardware compiler is under development at Rutgers University. This compiler translates descriptions of digital circuits into gate-level layouts for optical logic arrays interconnected in free space. 191

TuA3

9:20 am

Shared memory optical/electronic computer: architecture design, Clare Waterson, B. Keith Jenkins, *The Aerospace Corporation, University of Southern California*. An optical/electronic MIMD digital computer architecture is presented. It comprises a passive optical shuffle exchange network interconnecting electronic processing elements, shared memory modules, and network control. 195

TuA4

9:40 am

Design and construction of a programmable optical 16×16 array processor, A. C. Walker, R. G. A. Craig, D. J. McKnight, I. R. Redmond, J. F. Snowdon, G. S. Buller, E. J. Restall, R. A. Wilson, S. Wakelin, N. McArdle, P. Meredith, J. M. Miller, G. MacKinnon, M. R. Taghizadeh, S. D. Smith, B. S. Wherrett, *Heriot-Watt University, UK*. A digital optical processor with 256 channels and a nearest-neighbor interconnect has been designed and constructed. Results obtained with the processor will be discussed. 199

TuA5**10:00 am**

Digital optical computer II: performance specifications, Peter S. Guilfoyle, Ronald S. Rudokas, Richard V. Stone, Edward V. Roos, *OptiComp Corporation*. System performance and key specifications of subassembly hardware will be introduced for a general purpose 32-bit digital optical computer. 203

Salon D**10:20 am****Coffee Break****Salon F****10:50 am–12:40 am****TuB, Fuzzy and Cellular Systems**

Kelvin Wagner, *University of Colorado, Boulder*, *Presider*

TuB1**10:50 am (Invited)**

Adaptive fuzzy systems, Bart Kosko, *University of Southern California*. Fuzzy systems estimate sampled functions without a mathematical model of how outputs depend on inputs. Expert advice and engineering judgment generate fuzzy systems. Sample data generates adaptive fuzzy systems. 208

TuB2**11:20 am**

Optoelectronic fuzzy logic system, Gary C. Marsden, Brita Olson, Sadik Esener, Sing H. Lee, *University of California, San Diego*. We present a digital application of the dual-scale topology optoelectronic processor to a parallel modus Ponens algorithm, which is generalized to include fuzzy values. 212

TuB3**11:40 am**

Optical morphological image processing with acousto-optic devices, Ravindra A. Athale, Joseph N. Mait, Dennis W. Prather, *George Mason University, U.S. Army Harry Diamond Laboratories*. An optical morphological image processor based on Fourier-plane filtering with an acousto-optic device is described and demonstrated. The size and shape of the structuring element can be changed easily and rapidly by changing the drive signal to the acousto-optic device. 216

TuB4**12:00 m**

Optical morphological image processor, Gary E. Lohman, K.-H. Brenner, *Universität Erlangen-Nürnberg, Federal Republic of Germany*. Mathematical morphology and present spatial light modulator technology enables nonlinear binary image processing at greater than 10 kHz. Preliminary experimental results and digital simulations of the unit under construction are presented. 220

TuB5**12:20 pm**

Cellular automata through multikernel incoherent holographic convolution, I. Glaser, *Tel Aviv University, Israel*. An optical implementation of cellular machines by using incoherent holographic convolution multiple kernels and nonlinear point transformations is described. Cellular machines with very few pixels per PE are possible. 223

12:40 pm–2:00 pm**Lunch Break****Salon F****2:00 pm–3:10 pm****TuC, Memory Issues**

Cardinal Warde, *Massachusetts Institute of Technology*, *Presider*

TuC1**2:00 pm (Invited)**

Optical data-base machines, P. Bruce Berra, *Syracuse University*. In this paper we discuss the application of optical technology to relational data-base management, full text processing, and multimedia information systems. 228

TuC2**2:30 pm**

Optical implementation of SELECTION operation in data base machines, Ravindra A. Athale, Michael W. Haney, *George Mason University, BDM International, Inc.* Two designs for numerical inequality detection optical circuits are described. These circuits work with parallel access optical memory (disks or holographic) for a critical component of an optical data base machine. 231

TuC3**2:50 pm**

Demonstration of an all-optical addressing circuit, Donald M. Chiarulli, Steven P. Levitan, Rami G. Melhem, *University of Pittsburgh*. A demonstration is presented of both single and parallel selection in a one of four addressing circuit when using coincident pulse addressing. Scalability issues of synchronization and power distribution are also addressed. 235

TuC4**3:10 pm**

Optical respite from the Von Neumann bottleneck, Alex Dickinson, *AT&T Bell Laboratories*. The high instruction bandwidth required by new computer architectures is straining the processor/memory communication bottleneck. Here we describe how a wide cache together with a wide optical link between processor and memory can significantly reduce average memory access times. 239

Salon D**3:30 pm–4:00 pm Coffee Break****Salon F****4:00 pm–5:40 pm**

TuD, Architectures and Signal Processors
Pochi Yeh, *University of California, Santa Barbara, Presider*

TuD1**4:00 pm**

Dual-scale topology optoelectronic processor: comparative analysis and technological feasibility, A. V. Krishnamoorthy, J. Ford, G. C. Marsden, G. Yayla, S. C. Esener, S. H. Lee, *University of California, San Diego*. We analyze the fully connected dual-scale topology optoelectronic processor system, compare it with existing electronic implementations, and discuss its technological feasibility and applicability to neural networks. 244

TuD2**4:20 pm**

Ring-array processor distribution topology for optical processing and interconnect, Yao Li, Berlin Ha, *City College of New York*. A ring-array processor distribution topology for optical SIMD processing and interconnect is proposed. Experimental demonstrations and discussions are presented. 248

TuD3**4:40 pm**

Guided-wave acousto-optic matrix algebra processor module, A. Kar-Roy, C. S. Tsai, *University of California, Irvine*. A high-speed integrated optic analog matrix algebra processor module, which utilizes a new architecture involving guided-wave multifrequency acousto-optic Bragg diffractions, has been realized in a Y-cut LiNbO₃ channel-planar-channel composite waveguide 1.0 mm × 10.0 mm × 28.0 mm in size. . . . 252

TuD4**5:00 pm**

4 × 4 photorefractive reconfigurable interconnect using laser diodes, Arthur E. Chiou, Pochi Yeh, *Rockwell International Science Center*. We report the demonstration and characterization of a 4 × 4 optical reconfigurable interconnect using laser diodes at 780 nm and a ferroelectric liquid-crystal spatial light modulator in conjunction with a photorefractive barium titanate crystal. The photorefractive hologram improves the energy efficiency by 9 dB over the conventional approach 256

TuD5**5:20 pm**

A compact photorefractive joint-transform correlator for industrial recognition tasks, H. Rajbenbach, S. Bann, J. P. Huignard, *Thomson-CSF, France*. A compact multichannel joint-Fourier-transform correlator using a nonlinear updatable holographic BSO crystal operates with a diode-pumped mini-YAG laser and performs multiobject recognition with high signal-to-noise ratios. 260

Salon F**7:30 pm****Postdeadline Papers**

John A. Neff, *DuPont Corporation, Presider*

Wednesday, March 6, 1991**Salon D****7:00 am–8:30 am****Buffet Breakfast****Grand Ballroom Foyer****7:00 am–5:30 pm****Registration/Speaker and Presider Check-in**

Salon F

8:30 am–10:20 am

WA, Optical Learning Systems

Bernard H. Soffer, Hughes Research Laboratories, President

WA1

8:30 am (Invited)

Learning in optical neural networks, Demetri Psaltis, California Institute of Technology.

Methods for learning in optical neural networks are reviewed, and challenges we face before we can make such systems practical are discussed. 266

WA2

9:00 am

Hologram multiplexing using orthogonal phase codes and incremental recording, Yoshinao Taketomi, Joseph E. Ford, Hironori Sasaki, Jian Ma, Yeshayahu Fainman, Sing H. Lee, Jack Feinberg, University of California, San Diego, University of Southern California. We present an approach that will efficiently store many holograms in a single photorefractive crystal by using phase-coded reference beams and incrementally recorded images. 268

WA3

9:20 am

Generalization in an optical on-line learning machine, John R. Wullert II, Eung Gi Paek, J. S. Patel, Bellcore. We report the demonstration of generalization in an optical on-line learning machine as well as the ability to map multiple inputs to each output category. 272

WA4

9:40 am

Closed-loop optical-disk-based associative memory, Mark A. Neifeld, Demetri Psaltis, California Institute of Technology. We describe and experimentally demonstrate a self-locking shift-variant optical-disk-based associative memory with a capacity of 10^4 images and a 10 ms access time. 276

WA5

10:00 am

Competitive optical learning with winner-take-all modulators, Kelvin Wagner, Tim Slagle, University of Colorado, Boulder. A competitive optoelectronic modulator-detector array using VLSI and liquid crystals has been fabricated. A self-aligning, unsupervised optical learning architecture based on this device is presented. . . . 280

Salon D

10:20 am–10:50 am Coffee Break

Salon F

10:50 am–12:20 pm

WB, Neural Network Components

Kristina Johnson, University of Colorado, President

WB1

10:50 am (Invited)

Parallel implementations of neural networks: electronics, optics, biology, Joshua Alspector, Bellcore. We consider the implementation aspects of neural networks in a variety of physical embodiments. Using physics as a guide to what is possible, we attempt to predict where each technology will be most useful for neural systems. Important differences emerge when we consider the relative roles of computation, communications, power, synapse density, adaptivity, flexibility, and input representation. 286

WB2

11:20 am

First demonstration of an optical learning chip, Kazuo Kyuma, Yoshikazu Nitta, Jun Ohta, Shuichi Tai, Masanobu Takahashi, Mitsubishi Electric Corporation, Japan. A GaAs optical learning chip is reported for the first time. A novel type of variable-sensitivity photodiode is developed as a synaptic device. The learning speed exceeding 640 MCUPS is obtained for the 8-neuron chip. The application to the pattern classification is demonstrated. 291

WB3**11:40 am**

GaAs-based optoelectronic neurons, Steven Lin, Francis Ho, Jae Kim,[†] Demetri Psaltis, *California Institute of Technology*, Stanford University, [†]*Jet Propulsion Laboratory*. An optoelectronic thresholding device, consisting of a light-emitting diode monolithically integrated with a double-heterojunction bipolar phototransistor and two MESFETs on a single GaAs substrate, is reported. This device exhibits a differential optical gain of 6 and an optical switching energy of 10 pJ. 295

WB4**12:00 m**

Nonlinear thresholding characteristic for optical computing of optically addressable GaAs/FLC spatial light modulators, Masashi Hashimoto, Yukio Fukuda, Shigeki Ishibashi, Ken-ichi Kitayama, *NTT, Japan*. The accumulative thresholding characteristics for inputting newly developed optically addressable transmission-type GaAs/FLC spatial light modulation (SLM) are experimentally investigated. A processor utilizing SLM, performing OR, and storing results is proposed. 299

12:20 pm–1:30 pm Lunch Break**Salon F****1:30 pm–3:00 pm**

WC, Joint Plenary Session on Optical Computing and Photonic Switching: I
B. Keith Jenkins, *University of Southern California*, *Presider*

WC1**1:30 pm Plenary**

Progress in optical interconnection technologies and demonstrators under the ESPRIT II OLIVES program, J. W. Parker, *STC Technology Ltd., UK*. OLIVES combines the skills of nine organizations to develop optical interconnect technologies and demonstrators. A review will be given of the aims and progress to date. 304

WC2**2:15 pm Plenary**

Recent advances in communication network theory, F. T. Leighton, *Massachusetts Institute of Technology*. In this survey of recent advances in communication network design and algorithms for message routing, emphasis is placed on a novel class of randomly-connected networks known as multibutterflies. 308

Salon D**3:00 pm–3:30 pm Coffee Break****Salon F****3:30 pm–5:00 pm**

WD, Joint Plenary Session on Optical Computing and Photonic Switching: II
Joseph Goodman, *Stanford University*, *Presider*

WD1**3:30 pm Plenary**

Flexible and fast switching, H. T. Kung, *Carnegie Mellon University*. 310

WD2**4:15 pm Plenary**

Ultrafast all-optical fiber soliton logic gates, M. N. Islam, C. J. Chen, C. E. Socolich, *AT&T Bell Laboratories*. We demonstrate a cascable soliton-dragging NOR gate with a switching energy of 5.8 pJ, a fan out of 6 and a potential speed of 0.2 TBit/s. This is one example of a novel time-domain chirp switch architecture in which digital logic is based on time-shift keying. 311

5:00 pm–5:15 pm**Closing Remarks**

B. Keith Jenkins, *University of Southern California*, *Program Chair*

Monday, March 4, 1991

Optoelectronic Components

MA 8:30am-10:20am
Salon F

Mario Dagenais, *Presider*
University of Maryland



Progress in Arrays of Opto-Electronic Bistable Devices and Sources

K. Kasahara, I. Ogura and Y. Yamanaka
Opto-Electronics Research Labs., NEC Corporation
34 Miyukigaoka, Tsukuba, Ibaraki 305, Japan

1 Motivations for the VSTEP

With recent progress in semiconductor arrays of opto-electronic bistable devices and sources, highly parallel optical interconnections and information processing has gradually become a reality. The Vertical to Surface Transmission Electro-Photonic device, or VSTEP, is a concept proposed to meet these requirements¹⁾. The essential ideas for the VSTEP are electrophotonic interfusion at the device level and the resulting performance efficiency improvements in power consumption and uniformity for two-dimensional matrix integration.

Another motivation for the existence of the VSTEP is in realizing an optical interconnection device with functions which allow for compact configuration resulting from an absence of electric circuits for controlling the state of optical interconnections(Fig. 1). For realizing quick reconfiguration, routing and level regeneration in a compact configuration, it becomes necessary to develop an optical functional interconnection device which not only has light emission/absorption but also has such functions as thresholding, latch and optical amplification²⁾. Based on this point, VSTEPS have been fabricated.

2 Concrete Examples of a VSTEP

A pnpn-device is an example of a VSTEP. In this device, the pnpn doped structure with dual gate electrodes exhibits thyristor-like electronic nonlinearity, which is necessary in realizing the latch function. The ON or OFF state can be determined and memorized either optically or electrically. Switching-off is performed by the application of a negative reset pulse to the anode³⁾.

Low power consumption is realized through the electro-photonic operational mode. During the retention period, low power consumption as small as a few μW is attained through

operation of the optical dynamic memory³⁾. The Electro-Photonic concept has also been extended to the task of lowering the optical switching energy by using the electronically assisted switching scheme⁴⁾.

The feasibility of larger-scale integration has also been confirmed by successful fabrication of a 1K bit VSTEP matrix, where 32 by 32 pnpn elements are integrated on a Si-GaAs substrate of about 1mmx1mm⁴⁾.

In the LED-mode VSTEP, poor light output conversion efficiency compels the VSTEP to be driven with low optical switching speed in a cascaded connection. Based on this, a laser-mode VSTEP has been fabricated (Fig. 2)⁵⁾. The device has a pnpn structure with three inserted undoped InGaAs strained quantum wells. Two distributed Bragg reflector (DBR) mirrors with alternating $\lambda/4$ AlAs/GaAs layers are formed at both ends. These active layers serve as absorption layers in the OFF state. To achieve high absorption efficiency with thin absorption layers, absorption enhancement using multiple reflection mirrors for lasing is utilized.

The switching voltage is 5V, and the holding voltage is 2.5 V. The threshold current (I_{th}) was as low as 1.2mA for a 10 μm -square device. The oscillation wavelength is 955 nm. Using two 30 μm -square VSTEPs (I_{th} :18mA), one as a laser light source and the other as an optical switch, cascability as fast as 10ns write-in time has been successfully demonstrated. This is a two to three orders of magnitude improvement over the experiments with an LED mode VSTEP being used as a light source.

3 Optical functional interconnection

Reconfigurable optical interconnection becomes significant when a 2-dimensional VSTEP matrix is applied, for example, to an optical crossbar switch. The resultant problem, however, is a high-speed reconfiguration scheme. Figure 3 reveals how to realize the reconfiguration in a compact configuration when a latch function is located within an optical interconnection⁶⁾. The ON/OFF state is decided and memorized here according to the electrical write-in to individual devices.

Using this driving scheme, N^2 optical interconnections can be reconfigured in N time slots through 2N electrical control lines. This greatly alleviates the reconfiguration procedure, particularly, when N increases. After the ON/OFF states write-in to individual devices is

completed throughout the VSTEP elements, the VSTEP matrix acts as an active spatial light modulator. Using the LED-mode VSTEP matrix, actual write-in time for one row has been confirmed to be reduced to as short as a few ns. This driving scheme is also used in the module, which is fabricated for application as a feed-forward type neural network.

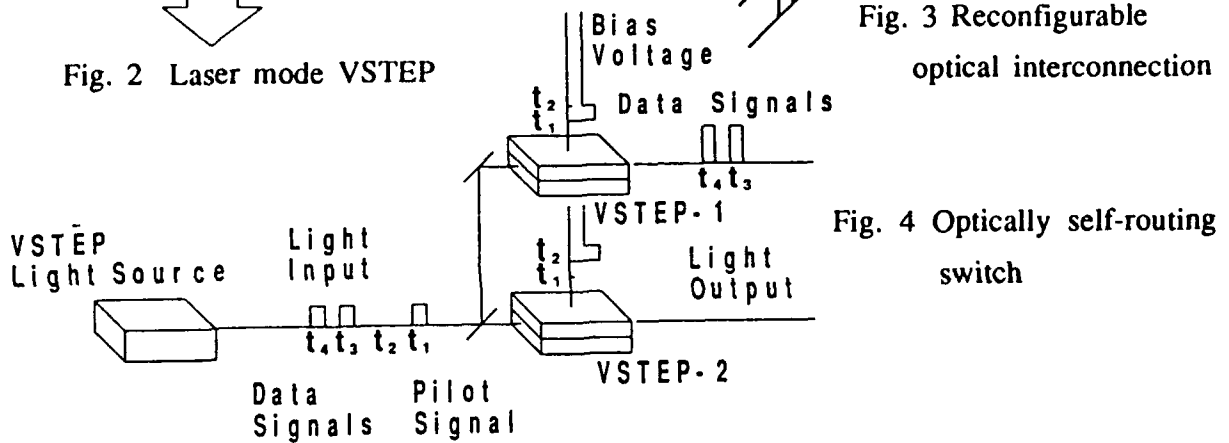
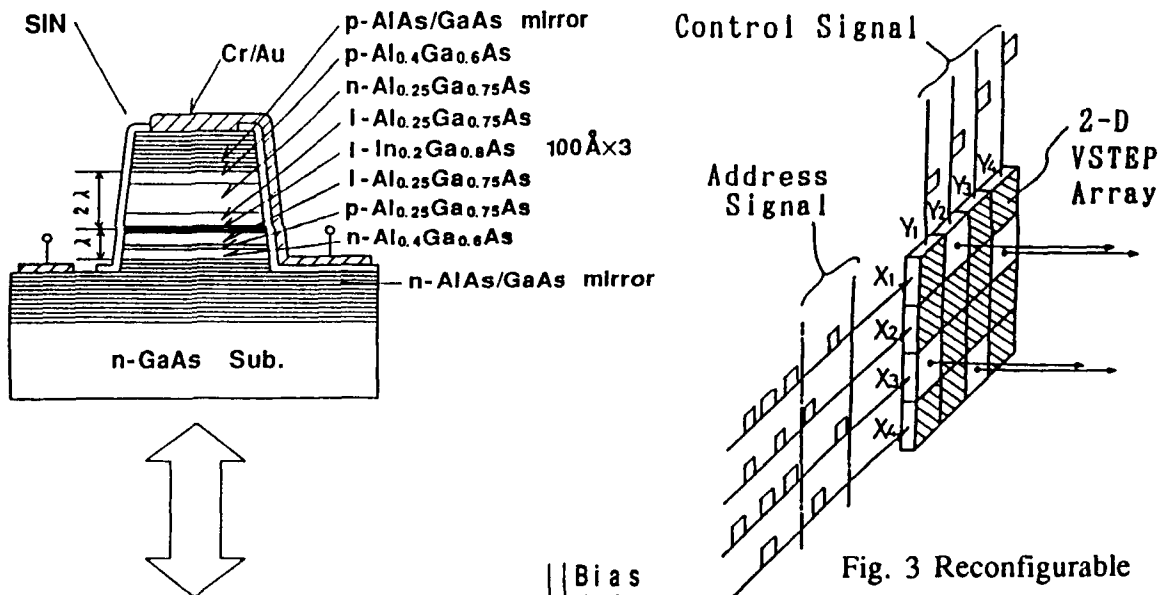
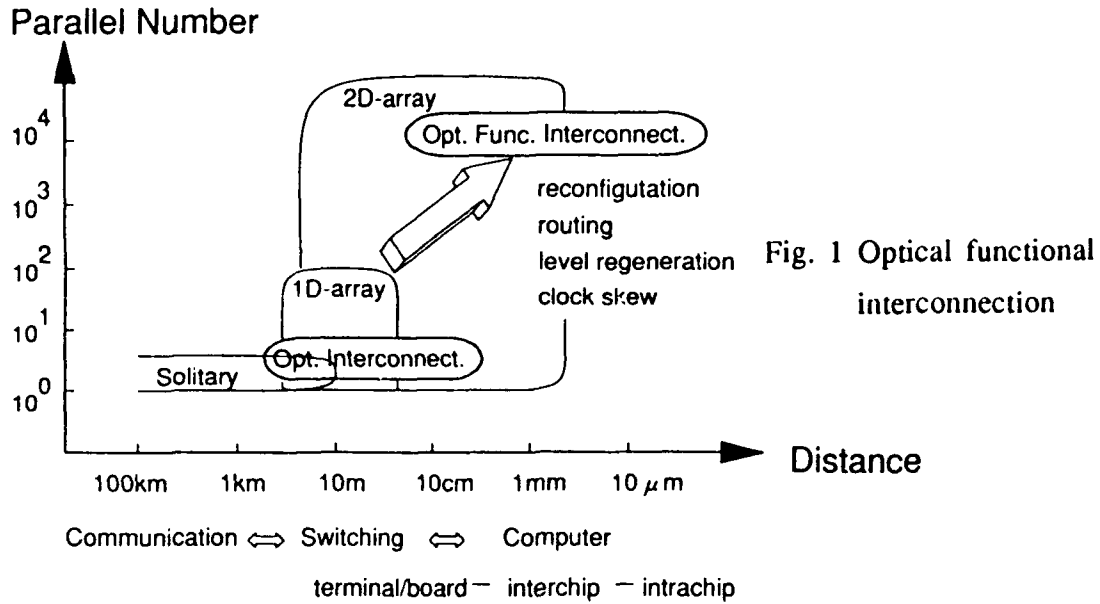
As another optical functional interconnection, an optically self-routing switch has also been successfully demonstrated (Fig. 4)⁷⁾. Normally, if the number of optical interconnections is increased, it becomes difficult to provide the electric cables necessary for routing. This self-routing switch was realized by using laser mode VSTEPs. The light input signal, which illuminates the VSTEPs, consists of a pilot signal and data signals. The pilot signal controls the ON/OFF state of the VSTEPs and seeks an output port for the data signal. The pilot signal at t_1 or t_2 , determines which VSTEP is turned on or not. In this case, only VSTEP-1 is turned on, and acts as an optical amplifier. As a result, optical data signals are transmitted out of the output port of VSTEP-1.

Acknowledgements

The authors are very grateful to M. Sakaguchi, R. Lang, K. Kobayashi, N. Nishida, K. Asakawa and K. Hotta for valuable comments and encouragements. The authors are also grateful to R. A. Linke at NEC Research Institute for VSTEP experiments and useful discussions, and their colleagues who carried out much of the experimental work on VSTEP described in this paper. We also thank C. L. Giles at NEC Research Institute.

References

- 1) Y. Tashiro et al., Appl. Phys. Lett. 54, 329 (1989).
- 2) K. Kasahara et al., in Tech. Digest, OEC'90 (1990), 12B4-4.
- 3) K. Kasahara et al., Appl. Phys. Lett. 52, 679 (1988).
- 4) K. Kasahara et al., in Tech. Digest, IOOC'89 (1989), 20C3-1.
- 5) S. Sugimoto et al., in Tech. Digest, LEOS'90 (1990), PD8.
- 6) I. Ogura et al., Appl. Phys. Lett. 57, 540 (1990).
- 7) I. Ogura et al., in Tech. Digest, 22st SSDM (1990), D-3-5.



Integrated array of self electro-optic effect device logic gates

A. L. Lentine

AT&T Bell Laboratories, Naperville IL 60566

L. M. F. Chirovsky, M. W. Focht, J. M. Freund, G. D. Guth, R. E. Leibenguth,
G. J. Przybylek, and L. E. Smith

AT&T Bell Laboratories, Breinigsville, PA. 18031

L. A. D'Asaro

AT&T Bell Laboratories, Murray Hill NJ 07974

D. A. B. Miller,

AT&T Bell Laboratories, Holmdel NJ 07733

Arrays of symmetric self electro-optic effect devices (S-SEEDs) have been made with low operating energies and fast switching speeds [1,2]. The device has the characteristics of a set-reset latch, although it can be made to do logic functions such as a NOR gate by presetting the state of the device before the application of the data inputs [3]. Logic gates that can perform more complex functions without preset beams may be realized by using electrically connected detectors configured like transistors in NMOS or CMOS circuits together with an output S-SEED to provide the output beams [4]. In this paper, we describe the first integrated arrays of these logic gates, each of which can perform the four basic logic functions without the use of preset beams. Each logic gate in the array consists of six quantum well p-i-n diodes, four input diodes configured similar to transistors in a CMOS NOR gate, and two output diodes (i. e. a S-SEED) that provide a set of complementary output beams. Like the S-SEEDs, this device has time sequential gain, in which the low power input beams set the state of the device and a set of equal higher power clock beams subsequently read the state. This device retains many desirable qualities of the S-SEED such as signal regeneration and retiming, wavefront restoration, and operation over several decades in power levels due to its differential nature. Because the logic gate contains only quantum well diodes, the same batch fabrication procedures [1] used for S-SEED arrays were used to make the arrays of these devices.

A photograph of part of the array is shown in Fig. 1 and a schematic diagram of each device is shown in Fig. 2. In making systems with this device, signals are routed in pairs with their logic state determined by the ratio of the optical powers in the two beams. We can define a logic "one" when $A > \bar{A}$, $B > \bar{B}$, and $C > \bar{C}$. The uncomplemented signals, A and B , are incident on parallel connected diodes and the complemented signals, \bar{A} and \bar{B} , are incident on serially connected diodes. The two groups of diodes are connected in series and the center node between them is connected to a S-SEED that provides the output beams. In operation, the signals are applied first, setting the node voltage, V_n , and then the clock beams are applied to read the device state with the same time sequential gain mechanism as the S-SEEDs. If input A , input B , or both inputs are "high", either or both of inputs \bar{A} and \bar{B} must be low. Initially more current will flow through the parallel connected diodes than the serially connected ones, so the node voltage, V_n , will tend switch toward V_0 . If we operate at a wavelength where the absorption is less at high voltages, then, when the clock beams are applied, output C will be low. However, if both inputs A and B are low, the serially connected diodes will generate more initial current and the device will switch toward essentially 0 volts, and C will be high when the clock beams are applied. This is the characteristic of a NOR gate, and any differential NOR gate can perform AND, OR, and NAND functions by redefining the logic state of the inputs and/or outputs. For example, if we define a logic "one" when $A < \bar{A}$, $B < \bar{B}$, and $C < \bar{C}$, then the logic gate is a NAND gate.

The chip contains a 16×16 array of these optical logic gates. Each optical input and output window is $5\mu\text{m} \times 5\mu\text{m}$. The inputs and outputs (e. g. A and \bar{A}) are on $20\mu\text{m}$ centers, and the different inputs (e. g. A and B) are on $10\mu\text{m}$ centers. Including the power leads, the unit cell size is $35\mu\text{m} \times 40\mu\text{m}$ for a total array size of $560\mu\text{m} \times 640\mu\text{m}$. The unit cell size for S-SEEDs with the same size windows was $20 \times 40\mu\text{m}$ [2]. The devices were reflection mode devices [5] with 72 periods of 100\AA GaAs quantum wells and 35\AA $\text{Ga}_{0.7}\text{Al}_{0.3}\text{As}$ barriers. Bistability data on S-SEEDs from the same wafer showed a contrast ratio of 4:1 at 6 volts increasing to 7:1 at 15 volts.

The devices were tested by generating data using two differential quantum well modulators. The contrast ratio of the inputs was roughly 2:1 at 15 volts, because the drive circuit used did not give the full voltage swing and because there was some saturation of the quantum well material as these devices had thicker (65\AA) barriers [6]. For the gate to operate properly, the input contrast ratio divided by the bistable loop width must be greater than two, so that two logic "zeros" on the parallel connected diodes generates sufficiently less current than a logic "one" on the serially connected diodes. By attenuating the inputs to the parallel connected diodes by 50%, optimum operation is achieved with low contrast inputs. A diagram of the experimental set-up is shown in Fig. 3. The data inputs from the differential modulators are reflected from patterned mirrors onto the respective optical windows of the device. The clock input and output pass through the transparent portion of the mirrors. We used patterned mirrors with alternating chromium and gold mirrors and reflected the inputs incident on the parallel connected diodes from the chromium mirrors, and the inputs incident on the serially connected diodes from gold mirrors. Since the reflectivity of the chromium mirrors was 56% of that of the gold mirrors, almost ideal attenuation was provided. It has also been shown that using this selective attenuation for inputs to a S-SEED operating as a logic gate increases the allowed variations in optical signal levels [7].

The oscilloscope photo in Fig. 4 showing the A and B inputs and C output demonstrated that the device has the correct NOR functionality. The data input powers on the device were $2\mu\text{W}$ and $4\mu\text{W}$ for the attenuated inputs and $\sim 4\mu\text{W}$ and $8\mu\text{W}$ for unattenuated inputs. By defining the differential energy as the difference in power level of the two inputs multiplied by the switching time, these energies correspond to 4 pJ per data input (using the unattenuated powers) for the device at 13 volts bias. This energy is about a factor of two higher than that of a comparable S-SEED [2]. One reason for this is that the capacitance is larger in these devices because there are 6 diodes as opposed to the equivalent area of 4 diodes in the S-SEED. A second reason is that when the inputs are attenuated, the equivalent difference in power in terms of how much difference in photocurrent is generated is only $2\mu\text{W}$, thus giving another factor of two increase in the required energy. However, for higher input contrast ratios, the optical energy required for this gate should be comparable to a S-SEED logic gate, because the selective attenuators will no longer be required for this logic gate.

In conclusion, we have built and tested 16×16 arrays of SEED logic gates using the same batch fabrication procedures presently used for S-SEED arrays. These logic gates have the same desirable characteristics as the S-SEED set-reset latches, but do not require a pre-set beam for logic functions. The optical energy required for operation is slightly higher than that of S-SEEDs, but we would expect the energies to be comparable as the contrast ratios of the input signals improve.

The authors would like to thank S. L. Walker for fabricating the binary phase gratings and asymmetric patterned mirrors, and S. J. Hinterlong for designing the laser diode mount and mounting the laser diode used in our experiment.

References

1. L. M. F. Chirovsky, L. A. D'Asaro, C. W. Tu, A. L. Lentine, G. D. Boyd, and D. A. B. Miller, "Batch Fabricated Symmetric SEEDs," OSA Proceedings on Photonic Switching, (Optical Society of America, Washington D. C. 1989) pp. 2-6.
2. A. L. Lentine, F. B. McCormick, R. A. Novotny, L. M. F. Chirovsky, L. A. D'Asaro, R. F. Kopf, J. M. Kuo, and G. D. Boyd, "A 2 kbit array of symmetric self electro-optic effect devices," IEEE Photon. Technol. Lett., vol 2, pp. 51-53 (1990)
3. A. L. Lentine, H. S. Hinton, D. A. B. Miller, J. E. Henry, J. E. Cunningham, and L. M. F. Chirovsky, "Symmetric self-electro-optic effect device: Optical set-reset latch, differential logic gate, and differential modulator/detector," IEEE J. Quantum Electron. QE-25 pp. 1928-1936 (1989)
4. A. L. Lentine, D. A. B. Miller, J. E. Henry, J. E. Cunningham, L. M. F. Chirovsky, L. A. D'Asaro, "Optical logic using electrically connected quantum well p-i-n diode modulators and detectors," Applied Optics, vol. 29, pp. 2153-2163 (1990)
5. G. D. Boyd, D. A. B. Miller, D. S. Chemla, S. L. McCall, A. C. Gossard, and J. H. English, "Multiple quantum well reflection modulator," Appl. Phys. Lett. 50 pp. 1119-1121 (1987)
6. A. M. Fox, D. A. B. Miller, J. E. Cunningham, J. E. Henry, W. Y. Jan, "Excitonic saturation intensity in GaAs/AlGaAs quantum well optical modulators," OSA Annual Meeting, 1989 Technical digest series vol. 18 (Optical Society of America, Washington, DC, 1989), paper MB2
7. N. C. Craft and M. E. Prise "Optical systems tolerances for symmetric self electro-optic effect devices in optical computers," Topical Meeting on Optical Computing, Technical Digest Series 1989, Volume 9, (Optical Society of America, Washington DC 1989) pp. 334-337

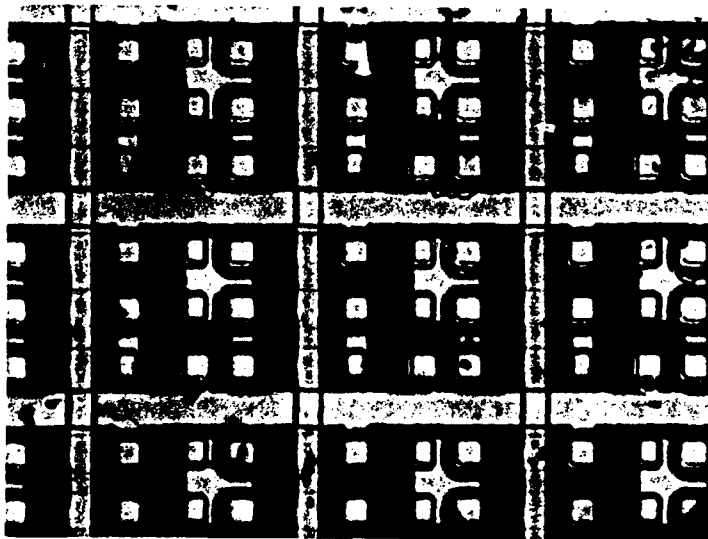


Fig. 1 Photograph of a section of the logic gate SEED array

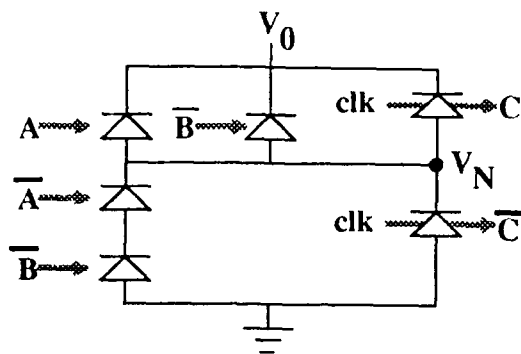


Fig. 2 NOR gate Schematic diagram

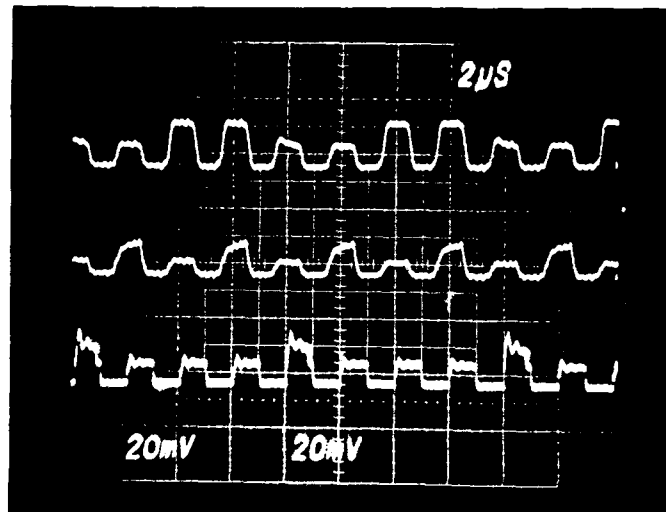


Fig. 4 Experimental Results of NOR gate demonstration. Top trace is input A, middle trace is input B, and bottom trace is output C.

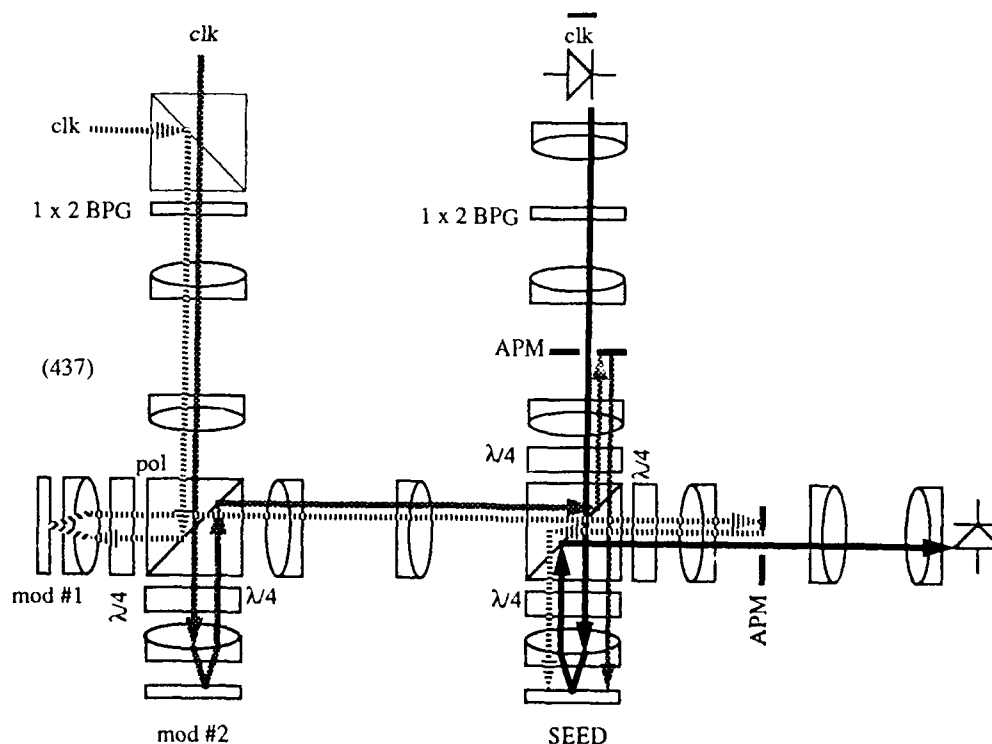


Fig. 3 Experimental set - up for the NOR gate demonstration. BPG - binary phase grating , APM - Asymmetric patterned mirror. Differential modulators #1 and #2 provide input data. Arrows show the optical beam paths - they are not ray traces.

Binary Arithmetic Using Optical Symbolic Substitution and Cascadable Surface-Emitting Laser Logic Devices

Julian Cheng

University of New Mexico[†], Center for High Technology Materials, Albuquerque, NM 87131
Voice: (505) 277-5605, FAX: (505) 277-6433

G. R. Olbright and R. P. Bryan

Sandia National Laboratories^{††}, Albuquerque, NM 87185

In this paper, we describe the design and operation of optical logic gates based on heterojunction phototransistor (HPT) and vertical-cavity surface-emitting laser (VCSEL) structures. We call the HPT/VCSEL structure a surface-emitting laser logic device. These structures will find use in optical communication systems as well as in parallel optical computing architectures. We illustrate complete sets of optical logic functions, upon which arithmetical logic units (ALU) are based, and provide specific examples of binary arithmetic operations based on optical symbolic substitution.

Two-dimensional surface-normal optical switching architectures represent potentially very compact, high throughput, parallel processors that are free from the effects of electromagnetic interference. These systems require the development of high speed photonic switches that are compatible with a surface-normal architecture, and which can provide high optical gain and contrast, and operate with low optical input energies. Direct optical addressing is particularly desirable since the electrical addressing of large matrices of active devices necessarily entails the added complexity of matrix-scanning and time-multiplexing. The switching elements should not be excessively sensitive to temperature variations, external optical feedback, or polarization diversity. Optical switches based on the integration of a HPT or pnpn devices with a VCSEL meet many of these requirements. Here, a single device provides electronic amplification, optical gain, switching, control and logic with little or no electronic intermediary. It eliminates the need for an additional external optical source or an optical bias beam. By varying the degree of positive optical or electrical feedback between the VCSEL and HPT, these structures can function alternatively as an optical amplifier, an optical switch, or a bistable logic or memory device.

Photodiode-controlled switching of electroluminescence has been demonstrated by NEC¹ and Mitsubishi,² using arrays of integrated AlGaAs/GaAs p-n-p-n HPT/LED structures called the VSTEP (vertical to surface transmission electrophotonic array), which demonstrated optical switching of LED-like power levels at data rates in excess of 100 MB/s. However, LED-based structures are inefficient devices with high drive current, low optical output, and little or no optical gain. Moreover, the LED electroluminescence is not collimated but Lambertian, which gives rise to serious optical crosstalk problems that become intractable for a densely-packed array in a free-space optical system. A VCSEL, on the other hand, has very low beam divergence, much higher radiative efficiency, and is capable of providing high optical gain and contrast³. VCSELs exhibiting low-threshold current and high differential quantum efficiency have been demonstrated⁴, using proton-implant current isolation and planar vertical-injection device structures.

Complete optical logic functions such as inversion, AND, NAND, OR, NOR, and exclusive XOR, etc., can be realized using simple combinations of phototransistors or photodiodes and lasers. The principle of operation of the optical OR and AND gates are shown in Fig. 1, while actual demonstrations of these logic operations are shown in Fig. 2, which displays the input and output optical pulse sequences. In the dark, the phototransistors are in the OFF-state, which exhibits a high bias voltage and low collector current. When the optical input is sufficiently strong, and the HPT gain is sufficiently high, the collector current exceeds the threshold of the VCSEL. The HPT goes into a low bias voltage, high conductance ON-state and switches on the VCSEL. If multiple optical inputs, each of sufficient intensity to switch on the VCSEL, are incident on the HPT, then an

optical OR gate is obtained. To operate as an AND-gate, the intensity of each optical input must be such that they can collectively, but not individually, produce enough current gain to switch on the VCSEL. The AND and OR gates are sufficient to carry out binary addition, but all the other logic functions, including the exclusive OR (XOR), can be implemented using only a single logic level without cascading. Each logic gate contains a single phototransistor and a single VCSEL, except for the exclusive-OR gate, which is based on a symmetrical differential drive configuration.

Boolean logic recognizes a combination of input bits and outputs one bit. Symbolic substitution,⁵ which is based on optical pattern transformations, recognizes not only a combination of bits but also their relative spatial configuration. Thus it recognizes an input symbol, i.e. an optical pattern of bits, and outputs another symbol, i.e. a new optical pattern of bits. Because of the added degree of freedom represented by the configurational information, it is well suited for the high speed, massively parallel processing of optical data. Not only are multiple patterns processed in parallel, the logical functions can sometimes be repeatedly sequenced in parallel. We will illustrate this with the example of a two-dimensional binary half-adder, using a two-dimensional array of phototransistor/VCSEL or photothyristor/VCSEL logic gates.

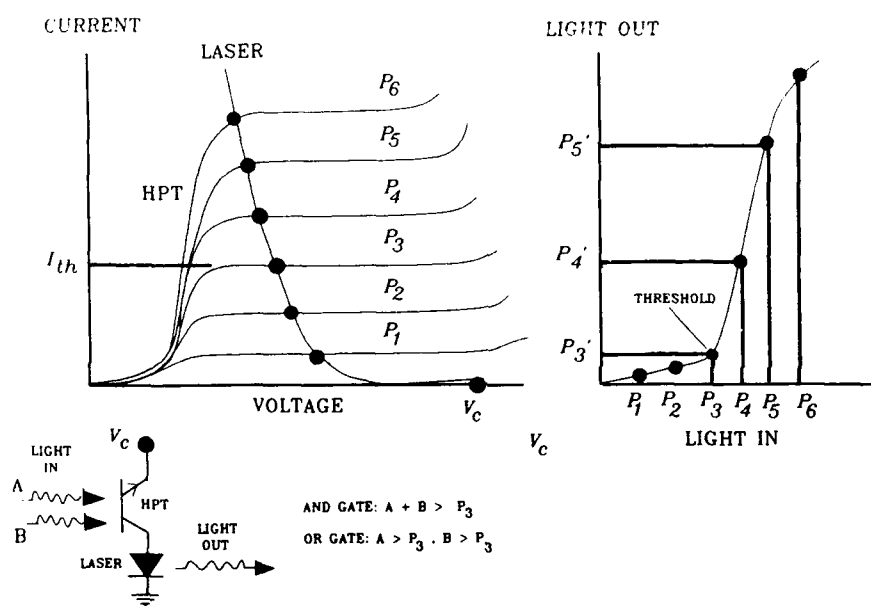


Fig. 1. Non-latching optical AND (a) and OR (b) logic gates based on a phototransistor and a vertical-cavity surface-emitting laser, and their operating principles.

Binary addition involves the SUM and CARRY operations and can be simulated using only AND- and OR-gates (but this would require complementary optical inputs). It is simpler to simulate these optically using the exclusive OR (XOR) logic function. The SUM is implemented with an XOR-gate, and the CARRY with an AND-gate. Using symbolic substitution, the states 0 and 1 are represented by symbols, i.e. by a VCSEL in the ON-state or OFF-state. To implement binary addition, the inputs consist of two N-bit words, i.e. two linear arrays of N symbols (optical inputs A and B) arranged as parallel rows of optical bits (Fig. 3). The addition algorithm consists of a set of rules, which prescribe the pattern shifts and transformations that simulate the SUM (XOR) and CARRY (AND) operations (see Fig. 4). The result of adding bits A and B is to produce new symbols, in which the top half contains a left-shifted symbol representing the CARRY-bit, while the bottom contains a right-shifted symbol representing the SUM-bit. The SUM bit is 1 only if A or B is 1 (i.e. $A \text{ XOR } B$), while the CARRY-bit is 1 only if A and B are both 1 (i.e. $A \text{ AND } B$).

The optical "HALF-ADDER" hardware contains a two-dimensional array of optical switches enabled by an array of input optical signals incident on columns of photodetectors, which in turn are interleaved with columns of VCSELs that generate the optical output pattern. Each position in the array consists of two optical-logic gates, AND and XOR, each of which contains one or more

HBTs and VCSELs. The schematic layout for each element (bit) in a row of this N-bit ADDER is depicted in Fig. 3, which also illustrates the lateral spatial shift in the symbolic substitution scheme. The switched VCSEL outputs are shifted diagonally as shown to simulate the symbolic substitution.

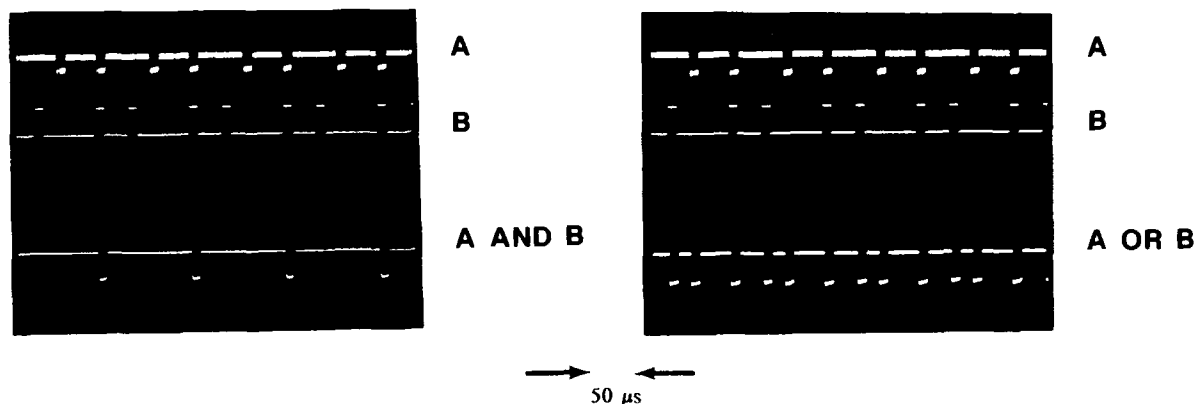


Fig. 2. Optical input and output pulse sequences for an optical AND- and an OR-gate. The VCSEL used in this experiment has an overall optical gain of > 20 and an on/off contrast of 34 dB.

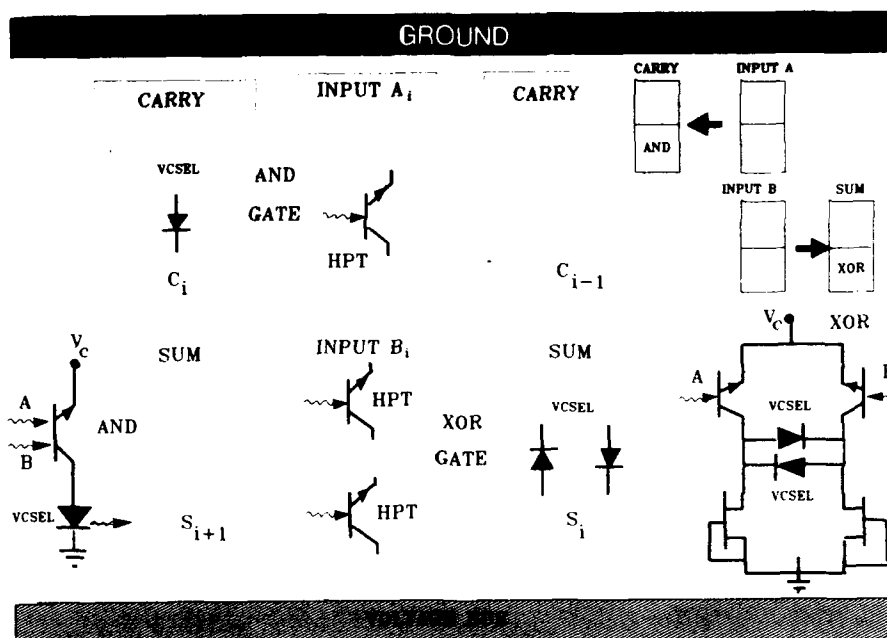


Fig. 3. The binary half-adder based on optical symbolic substitution and its implementation using HPT/VCSEL logic. The rows represent bits in a N-bit binary word. Each bit position contains an AND gate and an XOR gate, with two diagonally-staggered optical output VCSELs (SUM and CARRY).

We illustrate binary addition in Fig. 4. Each bit in the sum of $A + B$ is replaced by the corresponding left and right-shifted CARRY and SUM bits (VCSEL outputs), thus replacing the original rows of A and B with new, spatially-shifted symbols representing rows of CARRY and SUM bits. These are fed into the next logic array to undergo another symbolic substitution cycle, thereby producing a new row of CARRY and SUM bits. These steps are repeated until there are no 1 bits left in the CARRY row.

The N-step symbolic substitution procedure can be achieved by cascading N HPT/VCSEL logic arrays, but it can also be done by cycling the output through the the logic array N times during a

complete arithmetic operation. The cycling of the 2-D optical signals is achieved with the optical scheme depicted in Fig. 4. However, this requires that the logic array be reset after each pass through the half-adder, while preserving the previously generated optical outputs as the inputs for the next pass. This sequence requires a latching pnpn/VCSEL array and an optical memory buffer array (see Fig. 4). The latter consists of a simple array of latching identity switches, i.e., photodetector/laser switches, in which optical input logic level of 1 (light on) switches the photodetectors and thus the VCSELs on. The outputs of the optical logic processor (S1) trigger the memory buffer array (S2), whose latched optical outputs "store" the switched optical data from the previous pass. Switching the bias voltage on and off clocks the logic unit, and initiates new passes through the processor, while S2 is then erased (reset) to store the next set of outputs from S1. A maximum of N passes are needed to complete an N-bit binary addition. Thus, using a 128 x 128 array cycling at a 10 ns clock rate, 128 pairs of 64-bit words can be added in parallel in less than 640 ns.

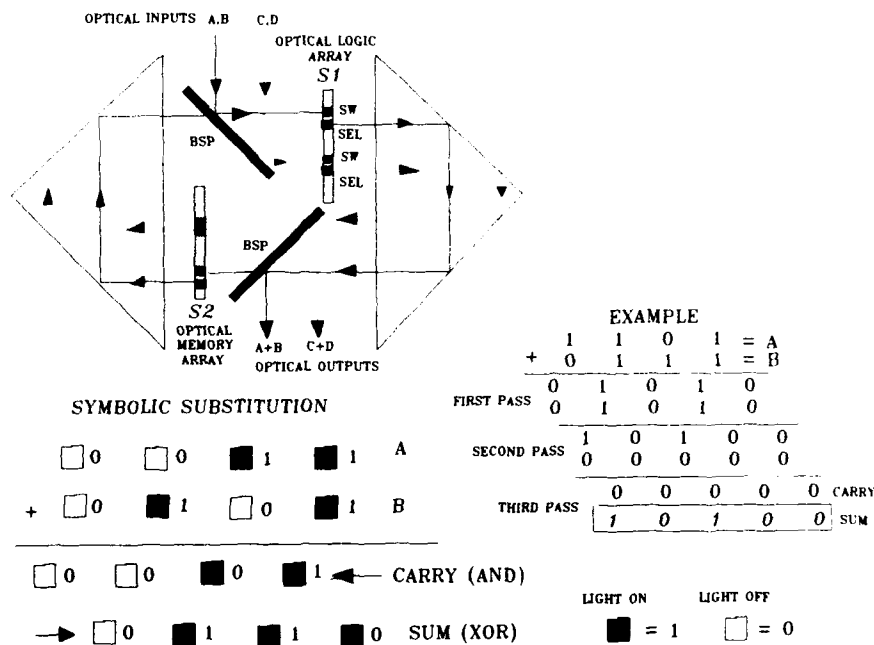


Fig. 4. The optical ALU hardware for a multi-pass binary half-adder, which includes a logic array and a memory array. Also shown are the rules for addition using symbolic substitution and an example of binary addition using the half-adder.

Here we have described binary addition using symbolic substitution and surface-emitting laser logic devices. The half-adder described above is simple, compact, and has a relatively low component count. Full adders, which would simultaneously take into account all CARRY operations, would speed up the ALU process time by a factor of N, (a single pass or clock period is required).

[†] UNM research is supported by AFOSR contract No. F49620-89-C-0028.

^{††} SNL research is supported by DOE contract No. DE-ACO4-76DP00789.

1. Y. Tashiro, K. Kasahara, H. Hamao, M. Sugimoto, T. Yanase, Jpn. J. Appl. Phys. 26, L1014 (1987).

2. K. Hara, K. Mitsunaga, K. Kyuma, IEEE Photonic Technol. Letters 1, 370 (1989).

3. G. R. Olbright, K. Lear, J. L. Jewell and P. Esherick, Invention disclosure filed with Sandia Corporation, SD-4896, S-71 738 (Jan. 1990).

4. Y. H. Lee, B. Tell, K. Brown-Goebeler, J. L. Jewell, R. E. Leibenguth, M. T. Asom, G. Livescu, L. Luther, V. D. Matterna, Electron. Lett. 26, 1308 (1990).

The Reliability of Optical Logic

Charles W. Stirk and Demetri Psaltis
Caltech 116-81
Pasadena, CA 91125

One of the potential niches for optical logic is very high speed digital circuits. Conventional lithographic manufacturing techniques decrease the individual logic device cost when the device density per unit area increases. Thermal power dissipation limitations, however, restrict the device density at a given duty cycle and switching speed. Thus, we desire optical logic devices with small switching energies for high speed systems. Since switching energy usually decreases with decreasing device area, small devices decrease thermal dissipation problems and increase manufacturing density.

On the other hand, small switching energy has some significant drawbacks. The main drawback is that since the number of photons required to switch the device is quite small, statistical fluctuations in the detected number of photons can make the device switch when it is not supposed to, or not switch when it is. The focus of this paper is to analyze the effect of the contrast ratio and fan-in of quantum noise limited optical logic devices on the reliability of their circuits as measured by the bit error rate (BER), which represents how often on average the circuit will give an erroneous output. The device models we use approximate the behaviour and parameters of several devices reported in the literature: the self-electrooptic effect device (SEED); vertical-cavity surface-emitting lasers (VCSEL), and bistable laser diodes (BLD).

The maximum desired BER for a device depends on the characteristics of the system. Suppose we have one million independent and identical devices each switching every nanosecond. If we want to have a low probability of any device error occurring during the systems' ten year lifetime, the BER of each device must be much less than 10^{-23} .

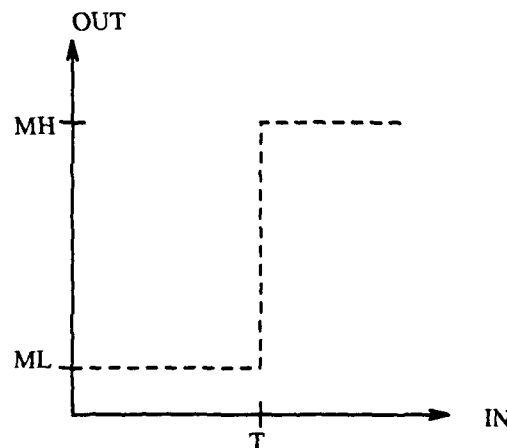


Figure 1: Input-Output Characteristic of Ideal Device

The ideal optical logic device has a switching characteristic that is a step function of the form shown in figure 1. One non-ideality that we consider is finite contrast ratio of the output, defined as the ratio of the mean number of photons for a logic 1, m_H , and the mean number of photons for a logic zero, m_L . The only other non-ideality is the shot noise of the detected

light, which we assume is Poisson distributed.

$$\Pr(k\text{-events in } n\text{-tries when probability is } \lambda) = P_{\lambda n}(k) = \frac{(\lambda n)^k}{k!} e^{-\lambda n} \quad (1)$$

For the following analysis we also assume that the inputs to the circuit are independent and that logic 1's and 0's are equiprobable.

The first logic family that we consider is optical logic, where separate optical inputs are fanned-in and summed on a single detector. Figure 2 depicts how the threshold or switching energy of the device determines the BER for an optical detector. The fraction of the area that is shaded is proportional to the BER.

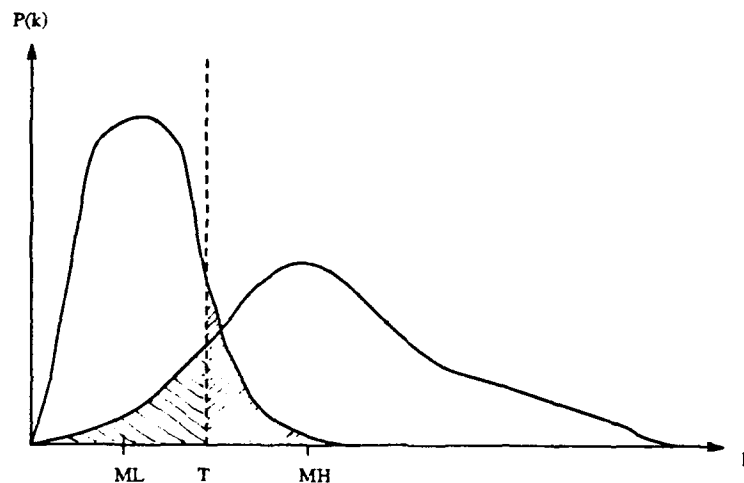


Figure 2: Calculation of BER

Equation 2 and figure 3 illustrate the dependence of BER on the threshold, T (in units of photons), for the optical OR, where N is the fan-in.

$$BER_{O-OR}(T) = \sum_{i=0}^{N-1} \binom{N}{i} \left(\frac{1}{2}\right)^N \int_0^T P_{im_L + (N-i)m_H}(k) dk + \left(\frac{1}{2}\right)^N \int_T^\infty P_{Nm_L}(k) dk \quad (2)$$

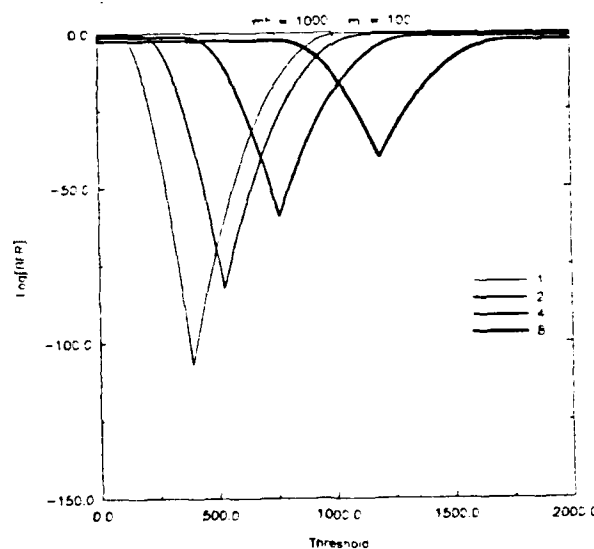


Figure 3: BER vs Threshold for Optical OR

The optimum threshold—that which produces the lowest BER—shifts to higher values with increasing fan-in due to the increasing mean number of photons for the all logic low case. Hence, the second term of equation 2 gets larger with increasing N . As expected, the BER at the threshold optimum decreases with increasing fan-in due to higher noise levels. Another feature is the broadening of the curves with increasing fan-in, as measured by the FWHM. This is a consequence of the broadening of the Poisson distribution with increasing mean, which also occurs when the fan-in is high. The reason that the BER does not go to its highest value of 0.5 at threshold equal to zero or infinity is due to the poor approximation of the Poisson to the Binomial distribution far from the mean.

The optical AND is the same device as an optical OR with different assignments of inputs to outputs. The corresponding reliability relationship is shown in equation 3 and BER vs threshold is plotted in figure 4.

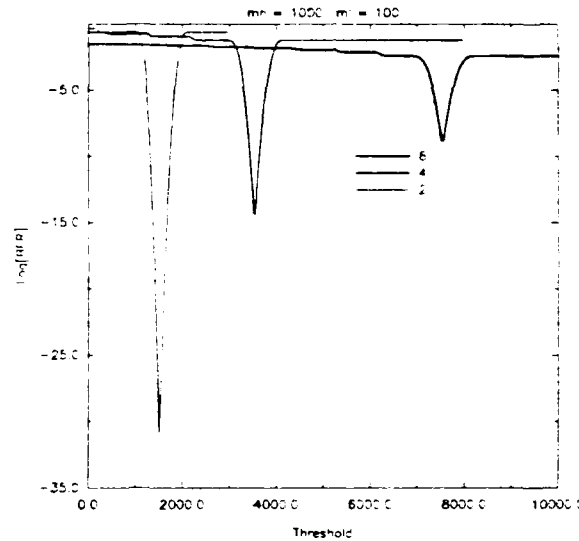


Figure 4: BER vs Threshold for Optical AND

$$BER_{O-AND}(T) = \sum_{i=0}^{N-1} \binom{N}{i} \left(\frac{1}{2}\right)^N \int_T^{\infty} P_{im_H+(N-i)m_L}(k) dk + \left(\frac{1}{2}\right)^N \int_0^T P_{N m_H}(k) dk \quad (3)$$

Like the optical OR, the optimum threshold for the optical AND decreases and shifts to higher values when the fan-in is increased. The BER at the threshold optimum for a given fan-in, however, is much less for the optical AND. Note that the threshold optimum lies just below the expected number of photons for a logic high, and gets closer as fan-in increases. This is due to the exponential increase in the number of terms in the sum of equation 3, where logic low conditions erroneously produce a logic high output. The inadequacy of the Poisson approximation to the Binomial far from the mean is evident in the low BER at zero and infinity.

Another reliability problem arises from the sharpness of the BER peaks around the threshold optimum. Due to random nonuniformities in manufacturing, the actual threshold of a device is a random variable. We assume that the distribution of the devices' threshold is Gaussian.

$$\Pr(t) = \frac{1}{\sqrt{2\pi}\sigma_T} e^{-\frac{(t-\mu_T)^2}{2\sigma_T^2}} \quad (4)$$

where μ_T is the threshold optimum and σ_T^2 is its variance. The average BER is

$$BER = \sum_{t=0}^{\infty} \Pr(t) BER(t) \quad (5)$$

Tables 1 and 2 list the weighted average BER for the optical OR and AND, respectively, when the standard deviation of the threshold is a fixed percentage of the mean and $m_H = 1000$, $m_L \approx 100$.

Table 1: Optical OR				
Fan-in	σ [% μ]	log[BER]	σ [% μ]	log[BER]
1	1	-101.08	10	-12.77
2	1	-77.26	10	-9.58
4	1	-54.39	10	-6.88
8	1	-35.92	10	-5.56

Table 2: Optical AND				
Fan-in	σ [% μ]	log[BER]	σ [% μ]	log[BER]
1	1	-101.08	10	-12.77
2	1	-26.84	10	-2.89
4	1	-11.29	10	-1.88
8	1	-6.44	10	-2.63

The Gaussian distribution of the thresholds decreases the average BER from the BER at the threshold optimum. The situation is more severe for the optical AND due to the narrowness of the peaks in its BER, and its lower overall optimum BER. The non-zero Poisson tails in the plot of BER vs threshold for the optical AND when weighted by a wide Gaussian make the average BER for the high fan-in situation appear greater than it actually is.

Along with variation in threshold, variation in contrast is a real effect in these devices. In the oral presentation we will show the effect this has on the BER. We have also considered the family of differential logic devices using this method of analysis. For multilayer circuits the calculation is a little more difficult due to multiple errors. For an optical combinational circuit, one can compute what is the largest fan-in device that achieves a desired BER. One can also ask the question, given an infinite contrast device, i.e. no photons for logic low, what is the quantum limit on the fan-in for a given BER and mean logic high?

In conclusion, low switching energy optical logic devices have quantum limits placed on their reliability. This is because shot noise can cause a configuration of the input values to be erroneously classified. The limit on the BER is most severe when the fan-in and non-uniformity of the devices are high. Thus, for high-reliability, low-power circuits it is imperative that the devices be very uniform and the fan-in be restricted to rather low values.

OPTICAL BINARY MULTIPLICATION BASED ON A NON-HOLOGRAPHIC CONTENT ADDRESSABLE MEMORY

Andrew Kostrzewski
Physical Optics Corporation,
20600 Gramercy Pl., Torrance, CA 90501

George Eichmann, Dai Hyun Kim and Yao Li
Department of Electrical Engineering,
City College of The City University of New York,
New York, NY 10031

ABSTRACT

A new fast binary multiplication scheme based on a non-holographic optical content addressable memory (CAM) and a sign/logarithm number (SLN) system is presented. The design and experimental demonstration of a 7-bit multiplier are presented.

Recently, the application of a non-holographic CAM to optical computing was proposed [1]. A CAM based processor compares an input pattern with all previously stored reference patterns and, when a match is found, it generates an output. Optical binary multiplication has attracted increasing attention in the optical information processing community. A number of digital multiplication schemes have been developed and experimentally implemented [2-4]. For a binary multiplication, the direct implementation of a truth table (e.g., in reference [3]) for large dynamic range calculations leads to an increasing hardware complexity. Here, the sign/logarithm number (SLN) system, especially suitable for both multiplication and division operations is applied. Due to the mantissa truncation, the calculation error is inherently "built-in" the proposed scheme and SLN system should be applied for large dynamic range calculations where the accuracy is not the main objective e.g., data preprocessing. To multiply (divide) two binary numbers (BNs) a and b , the numbers are converted to their SLN equivalents. The multiplication (division) is then obtained by adding (subtracting) appropriate logarithms. As a final multiplication step,

$$ab = \text{antilog}_2(\log_2 a \pm \log_2 b), \quad (1)$$

a conversion from SLNs to BNs is needed. To add/subtract logarithms, optical carry look-ahead (CLA) adder can be employed [1]. The addition result is obtained by setting input carry of the CLA adder equal to zero. To obtain subtraction result, the subtrahend is complemented and the input carry is set to one. The major advantage of the SLN system is that the product/quotient calculation is performed via fixed-point binary addition/subtraction. Because addition can be performed in a considerably shorter time than multiplication, the SLN based algorithm can achieve a higher multiplication speed than those achievable with any other equivalent length fixed-point number system [6].

To perform BN to a SLN system conversion a CAM is used [7]. This approach takes advantages of optical parallelism and high CAM storage capacity. The conversion accuracy is determined by the number of mantissa digits. The largest error is for the linear approximation and for the rounding approximation the error decreases by a factor of two with each additional mantissa bit.

A binary coded SLN is represented as

$$S \ A_{n-1} A_{n-2} \dots A_0 . A_{-1} \dots A_{1-m} A_{-m} \quad (2)$$

where S is the sign of the BN number. The n digits A_{n-1} to A_0 represent the characteristic, while the m digits A_{-1} to A_{-m} represent the mantissa. The corresponding decimal number N is expressed as

$$N = S \sum_{j=-m}^{n-1} A_j 2^j \quad (3)$$

For an integer, both the characteristic and the mantissa are positive numbers. To represent fractions, both the characteristic and the mantissa can be regarded as a two's complement. In such a case, a representation of negative powers is possible.

To implement a SLN multiplier, a 3-stage CAM is required. The first CAM performs the conversion from BNs to SLNs. To add the two logarithms, a second CAM performs a carry look-ahead (CLA) addition [1]. Finally, a third CAM does the conversion from the SLNs to BNs. The storage capacity needed for each CAM depends on the range of input numbers and the calculation accuracy. For example, for a 7-bit binary multiplication, to implement the first stage CAM, a 3-bit characteristic and a 5-bit mantissa is used. For each output bit, a separate truth table is needed. The truth tables for the output bits $A_2, A_1, A_0, A_{-1}, A_{-2}, A_{-3}, A_{-4}, A_{-5}$ contain 112, 76, 42, 71, 65, 63, 62, 114 product terms, respectively, for a total number of 605 product terms. Using Quine-McCluskey minimization method [8], the number of product terms was reduced to 111 (3, 3, 3, 10, 18, 21, 28, 25, respectively). To add two 8-bit logarithmic numbers, an 8-bit CLA requires 2519 product terms to be stored in the CAM [1]. A final, a third, CAM does the conversion from a SLNs to a BNs. For a 7-bit multiplier, the final result is a 14 bit word ($F_{13}, F_{12}, F_{11}, F_{10}, F_9, F_8, F_7, F_6, F_5, F_4, F_3, F_2, F_1, F_0$). To implement the conversions for each output bit, the number of product terms to be stored is 16, 22, 29, 36, 43, 53, 59, 68, 74, 81, 91, 97, 106, 116, respectively. Using Quine-McCluskey minimization method, the total number of product terms was reduced from 891 to 329.

To represent an optical CAM's pixel value, dual rail spatial logic is used [1]. Each product term is encoded as one column of the CAM mask. The CAM mask is illuminated by an array of line sources corresponding to a particular input configuration. The intensity transmitted through each column is summed and detected by a detector array. To establish the output result, the detected signal is thresholded.

To implement a binary coded SLN multiplier, three fixed binary masks are required. The first mask represents BN to SLN conversion, second, a CLA and third, a SLN to BN conversion. In Fig. 1, a schematic of an optical CAM processor is shown. An array of LDs is associated with each CAM mask. Each LD array illuminates only one CAM mask, which corresponds to implementation of a one CAM stage. To activate one LD array an active high decoder, where only one output line is at logic "one" is used. This logic state provides the enable signal for one array of LDs. The operation select lines are connected to an operation sequencer which contains a set of instructions. To store this program, a fast 2-bit word memory is used. Each pair of LDs is controlled by the status of the data out bus. If a particular bit in the data word is 1 (0), the lower (upper) LD from a corresponding pair is activated. Two AND gates, associated with each pair of LDs, constitute a simple 2 : 1 demultiplexer. To integrate the transmitted light intensities for each CAM mask column, an output anamorphic optical stage is employed. While intensity integration applies to all the masks, only a single CAM mask is illuminated at a time. To obtain the final result, the detected electronic signal is thresholded. The threshold level is set at 1/2 (between level 0 and level 1). If a single column output intensity is below the threshold, the output is at logic one, otherwise it is at logic zero.

In our experiment, the data was entered via an array of red light emitting diodes (LEDs - Panasonic P421). Using a 40 mm focal length cylindrical lens and a 375 mm focal length spherical lens, an anamorphic stage was built. The CLA masks were printed on a laser jet printer and, using an optical demagnification process, were reduced by a factor of twenty. As an example, the multiplication of two BNs $a=0111101$ and $b=1101110$ was performed. First, the two BNs were optically converted to their corresponding SLNs. In Fig. 2a, the eight conversion (from a BN to a SLN system) masks $M_2, M_1, M_0, M_{-1}, M_{-2}, M_{-3}, M_{-4}$ and M_{-5} corresponding to the output functions $A_2, A_1, A_0, A_{-1}, A_{-2}, A_{-3}, A_{-4}$ and A_{-5} , respectively, are shown. In Fig. 2b, the result of eight masks illumination and the corresponding average intensity distribution along each column are shown. In the detector plane, the transmitted intensity was acquired by a CCD camera (JVC TK 870V). For the BN a , the intensities below the threshold level were detected for $A_2, A_0, A_{-1}, A_{-2}, A_{-3}$, and A_{-5} . This corresponds to a SLN $A=101.11101$, while for the BN b , the intensities below the threshold were detected for

B_2, B_1, B_{-1}, B_{-2} , and B_{-5} , corresponding to a SLN $B=110.11001$. For the addition operation, a 8-bit CLA can be employed. Although it is feasible to implement this size of CAM, in our experiment two 4-bit CLAs operating in a ripple carry mode were used [1]. This approach decreases the overall processing speed, but on the other hand the hardware complexity is substantially reduced. This solution is not justified if the high processing speed is the major objective. For this case, one should try to use as large CLAs as possible without subjecting it to a ripple mode. The addition result d was 1100.10110. This result was converted back to a BN using CAM masks (see Fig. 3a). The output result was represented by the 14 masks F_{13}, \dots, F_0 . In Fig. 3b, the mask illumination together with the intensity distribution (integrated along the masks' columns) are shown. The intensities below the threshold level were detected for $F_{12}, F_{11}, F_8, F_7, F_6, F_2, F_0$ which corresponds to a BN $F=01100111000101$ (6597). Because of the mantissa approximation, the error was equal to 1.684%.

With this non-holographic CAM architecture, the multiplication speed is limited by the LD/LED maximum switching speed, detector response time and the electronic comparator's propagation delay. Using the fastest LDs with the switching rate of 14 GHz, optical detectors and comparators with response time as fast as 200 ps and 500 ps, respectively, the processing time for a single CAM stage of about 600ps is anticipated. Thus, with the state-of-the-art opto-electronic technology, for a 16-bit multiplier, a total multiplication time of about 2 ns is possible.

REFERENCES

1. A. Kostrzewski, Y. Li, G. Eichmann and D. H. Kim, "Fast Optical Digital Arithmetic Processor," Proc. SPIE 1296, in print, (1990).
2. F. T. S. Yu and T. Lu, Proc. SPIE 963, 446 (1988).
3. A. K. Cherri and M. A. Karim, Opt. Eng., 28, 638 (1989).
4. D. Psaltis, D. Casasent, D. Neft and M. Carlotto, Proc. SPIE 232, 251 (1980).
5. H. Henkel, IEEE Trans. ASSP, 37, 301 (1989).
6. G. L. Sicuranza, IEEE Trans. ASSP, 31, 877 (1983).
7. T. A. Brubaker and J.C. Becker, IEEE Trans. Comput., 24, 761, (1974).
8. R. K. Brayton, G. D. Hachtel, C. T. McMullen and A. L. Sangiovanni-Vincentelli, Logic Minimization Algorithms for VLSI Synthesis, Boston: Kluwer Academic Publishers, 1984.

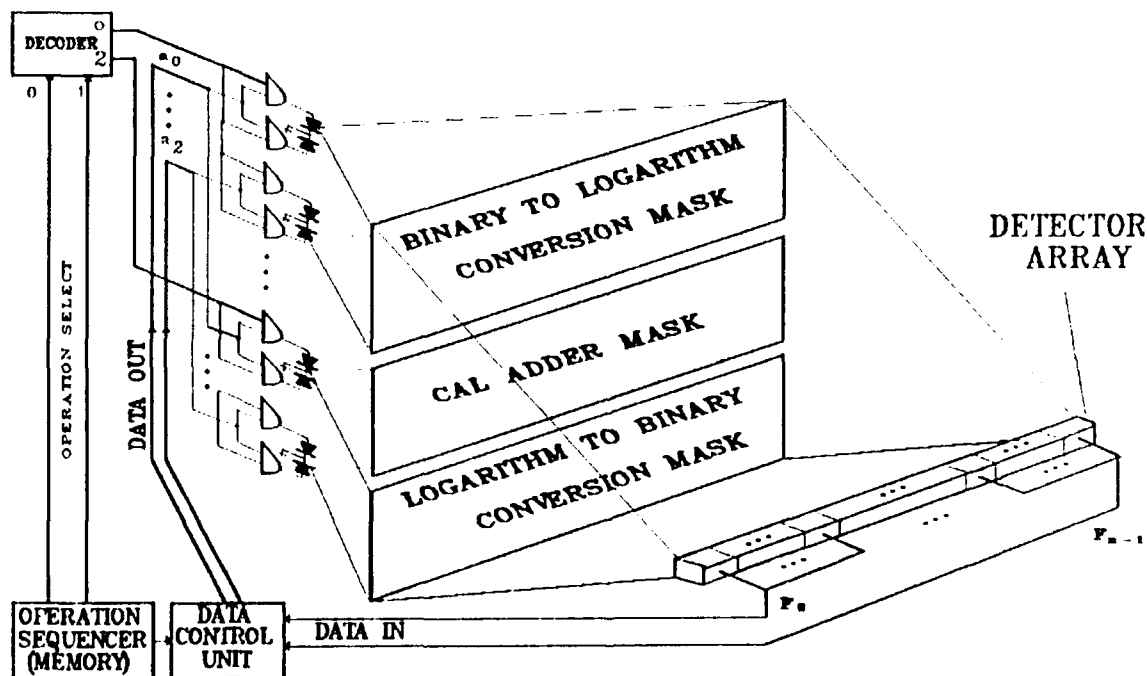


Fig. 1. Optical implementation of a n-bit CAM multiplier. For each operation, a separate CAM mask and a LD array is associated.

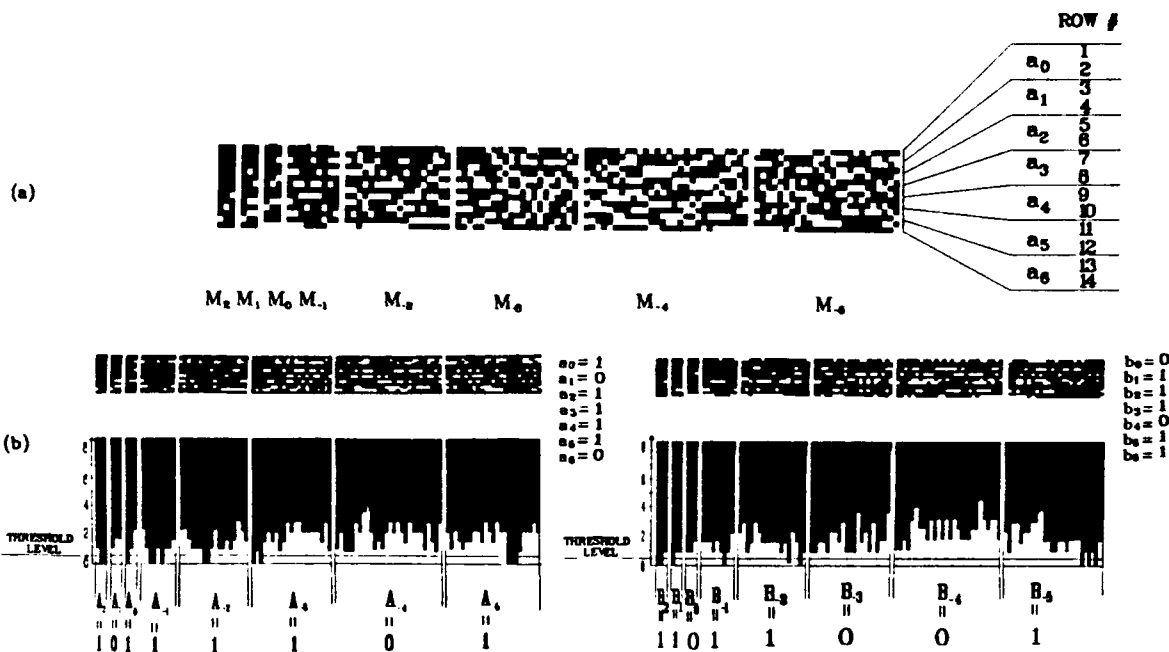


Fig. 2.(a) BNS to SLN conversion, eight CAM masks $M_2, M_1, M_0, M_{-1}, M_{-2}, M_{-3}, M_{-4}, M_{-5}$ corresponding to the output bits $A_2, A_1, A_0, A_{-1}, A_{-2}, A_{-3}, A_{-4}, A_{-5}$ are shown. (b) Transillumination result representing BN $a=1101110$ and intensity distribution which corresponds to SLN $B=110.11001$.

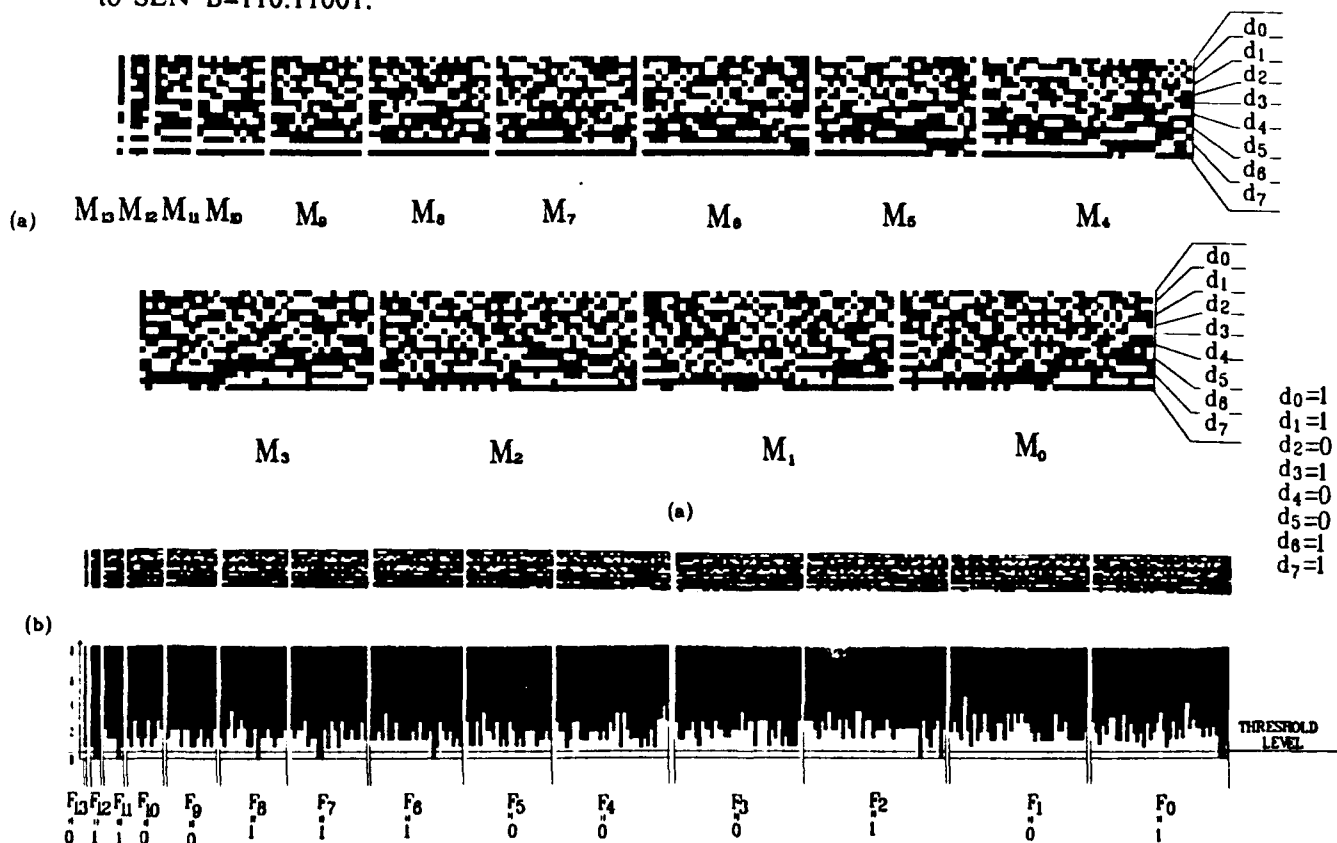


Fig 3. (a) A CAM mask for a conversion from SLN system to a BN system.

(b) The mask illumination together with the intensity distribution along mask columns are shown. The binary multiplication result is $F=01100111000101$ (6597).

Monday, March 4, 1991

Micro-Optics

MB 10:50am-12:20pm
Salon F

Adolf W. Lohmann, *Presider*
NEC Research Institute



Binary Optics and Applications

Wilfrid B. Veldkamp
Massachusetts Institute of Technology
Cambridge, MA 02139

In a classic example of technology transfer, binary optics is allowing optical designers to create innovative optical components which promise to solve key problems in optical sensors, communication, and optical processors.

3D-Integration of Digital Optical Systems

K.-H. Brenner

Angewandte Optik, Physikalisches Institut
der Universität Erlangen-Nürnberg
D-8520 Erlangen, Federal Republic of Germany

1. Introduction

The general goal of integration is to provide ease of fabrication, enhanced stability and compactness of complex systems by reducing the number of degrees of freedom in the assembly. Currently the term 'Integrated Optics' represents two-dimensional planar integration. With a variety of techniques passive optical components have been integrated on a single substrate, guiding the light along predetermined paths /1/. By including electro-optic switching mechanisms and also, more recently, by integrating semiconductors planar integration of monolithic integrated optical circuits (OEICs) has become possible. Planar integration permits only a propagation of zero-dimensional optical signals. Furthermore the need of coupling into and out of wave guides generates interface problems lessening the advantages provided by integration.

One of the main potentials of optics arises from the fact that the wavefield is three-dimensional, allowing interconnection of a large number of information channels simultaneously through space with a high bandwidth and with low crosstalk. A first approach to an integrated three-dimensional structure utilized diffractive-reflective components fabricated in dichromated gelatine to connect arrays of optical devices through free space /2/. Jahns and Huang /3/ suggested etching diffractive elements into a glass substrate to provide three-dimensional integrated optical systems. Here we want to propose an alternate concept for 3D-integrating regularly structured digital optical systems /4/.

2. Integration of optical functions

The basic functions in any optical system involve light collimation, beam splitting and combining and beam deflection. For image processing applications, the large space-bandwidth product of macro-optics is indispensable. Thus lenses, mirrors and beam splitters in the centimeter range must be used. For digital optical processing however a connectivity of 1000 channels is sufficient for competitiveness with electronic interconnect technology. For channel numbers in this range lens diameters of $200\mu\text{m}$ are sufficient according to space-bandwidth considerations and microintegration is thus a primary goal for these applications.

Typical digital optical systems (fig. 1) show certain regularities. They mostly consist of a regular sequence of a light source array, a Fourier-transformer, a filter, another Fourier transformer and a light detector array. Even in cases where a filter is not needed, imaging is performed by a sequence of two Fourier

transformers in order to have telecentric imaging properties.

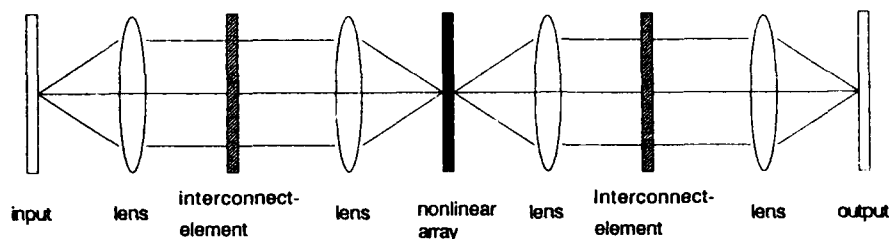


Fig. 1 General structure of optical processing systems

The schemes used for Fourier transformers most commonly are the well known 2f-system and the light pipe. The light pipe (fig. 2) requires two lenses but has the advantage of reduced vignetting and a larger aperture as compared to the 2f-system. The filter in digital optical systems is mostly required for nearest neighbor interconnections. To this end a structure is needed that splits one collimated beam into several collimated beams of different selected directions but equal intensity.

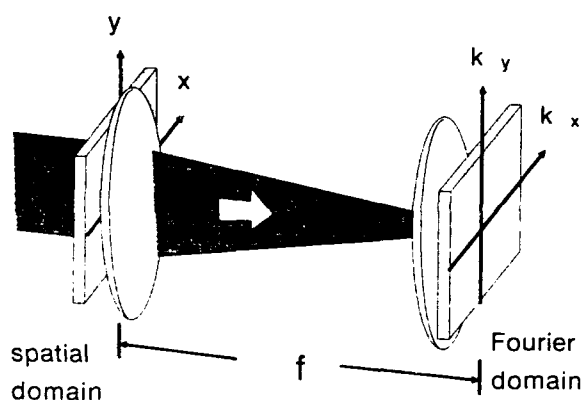
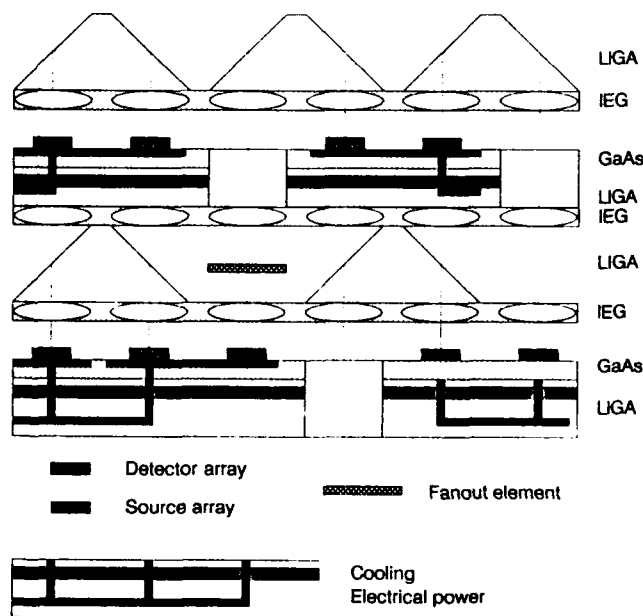


Fig. 2 A light pipe as a Fourier transformer

In order to integrate such systems suitable technologies are necessary. These technologies should allow for optical accuracy and should provide enough free parameters. Thus lithographic techniques are favourable.



LIGA: Deep Synchrotron Lithography
IEG: Ion exchange in Glass
PMMA: Photopolymerisation in PMMA

Micro lenses can be fabricated by many different techniques. A very accurate and flexible method is the fabrication of micro lenses by ionexchange in glass /5/. Beam shaping and beam splitting can be done by etching in glass or by photopolymerisation. Holographic techniques usually do not show the accuracy of lithographic techniques are therefore not considered here. Beam deflection together with mechanic packaging can be achieved by deep lithography in PMMA. PMMA-targets are irradiated by synchrotron radiation /6/ or by protons with kinetic energies between 5 MeV and 10 MeV /7/ through a metal mask which is

Fig. 3 General optical system in a regularized form

either transparent or opaque for the radiation. The high energy radiation splits the long molecular chains to shorter molecular chains. Penetration depths of up to $1000\text{ }\mu\text{m}$ can be achieved with this procedure. With a special developer the exposed areas can be removed from the substrate leaving a 3D-structure in PMMA.

From the above one can conclude that technologies for 3D-integration are available, however different materials and technologies are necessary. Thus a monolithic integration as in Si-electronics is not possible. In order to accomodate that it is necessary to rearrange the optical system in such a way that different types of functions are separated in different layers so that one technology serves for one layer. Fig. 3 shows an approach to this. The active devices are located in one layer. A spacer is provided as PMMA by deep lithography. The lenses, fabricated as lens arrays are all located in another layer. The optical path is folded by reflective structures realized again by volume PMMA-structures. The filter components are also arranged all in one plane. Thus a 3D-integrated system can be fabricated by stacking different layers of prefabricated structures. As an additional benefit the electrical wiring and cooling can be performed by the 3D-PMMA-structure. For alignment now only three positioning degrees of freedom are left (Rotation, Shift-x and Shift-y). The alignment problem can be solved either by alignment marks or by grooves for assembling the system.

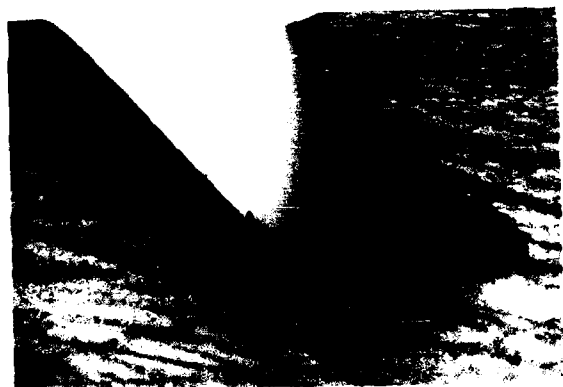


Fig. 4 A 45-degree slit produced in PMMA by Proton-lithography. The depth is $300\text{ }\mu\text{m}$.

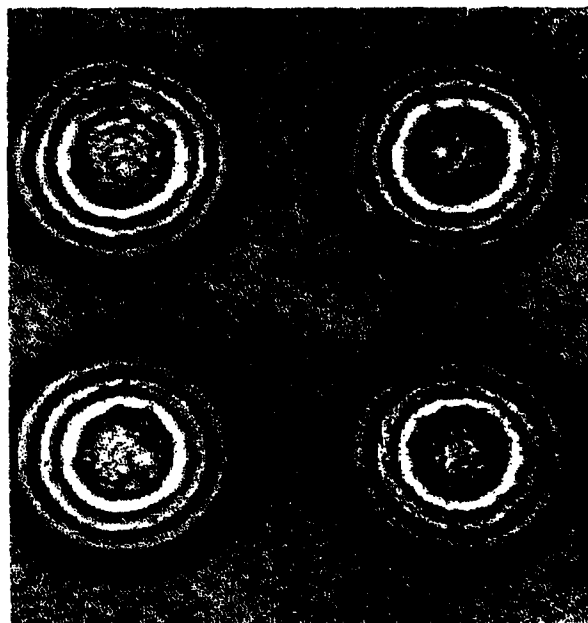


Fig. 5 Interferogram of a microlens array

3. Experiments

Using Proton-lithography we have generated a slit in a PMMA-target with a width of $300\text{ }\mu\text{m}$ and a depth of $500\text{ }\mu\text{m}$ (fig. 4). The orientation of the slit is 45° relative to the target surface. Fig. 5 shows an interferogram of a microlens array which we have fabricated using the Na-Ag exchange in glass. The diameter of the lenses is approx. $100\text{ }\mu\text{m}$ and the focal length approx. $500\text{ }\mu\text{m}$.

Fig. 6 shows an interferogram of a phase Fresnel zone lens which we have fabricated by photopolymerisation. The surface relief is determined by the amount of UV-exposure and can reach values of up to 6π .

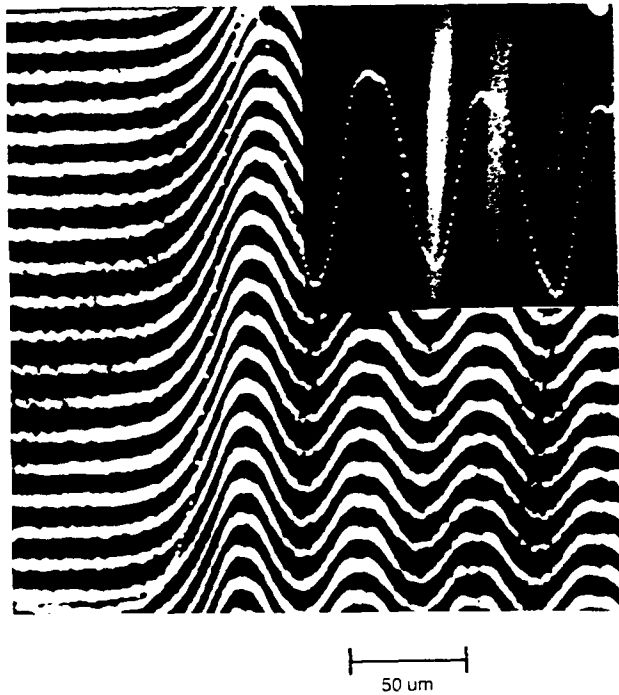


Fig. 6 Interferogram of the edge of a Fresnel zone lens.

4. Conclusion

In conclusion the paper proposes a concept for integrating miniaturized three-dimensional optical structures in millimeter to submillimeter range using existing technologies for structuring PMMA and glass. These structures can be combined with other microoptical components to fabricate three-dimensional integrated optical systems.

References :

1. H. Nishihara, M. Haruna, T. Suhara, "Optical integrated circuits", McGraw-Hill, New York 1989.
2. K.-H. Brenner, F. Sauer, "Diffractive-reflective optical interconnects", Appl. Opt. 27 (20) 4251, 1988.
3. J. Jahns, A. Huang, "Planar integration of free-space optical components", Appl. Opt. 28 (9) 1602-1605, 1989.
4. K.-H. Brenner, "Implementations and Integration of Regular Optical Logic", to appear in SPIE proceedings of the OE/LASE '90 Conf. 1214, Digital Optical Computing (Critical reviews), Los Angeles, Jan. 1990.
5. K. Iga, Y. Kokubun, M. Oikawa, "Fundamentals of Microoptics", Academic Press, Orlando 1984.
6. E.W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, D. Münchmeyer, Microelectronic Engineering 4 (1986) 35.
7. K.-H. Brenner, M. Frank, M. Kufner, S. Kufner, "H⁺-Lithography for 3D-Integration of Optical Circuits", Appl. Opt. Sept. 10, 1990

Integrated Free Space-Optical Permutation Network

Jürgen Jahns and Walter Däschner

AT&T Bell Laboratories

Rm. 4G-524

Crawfords Corner Road

Holmdel, NJ 07733

SUMMARY

Permutation networks such as the Perfect Shuffle, the Banyan, and the Crossover network can be used in optical computing or photonic switching to implement parallel algorithms efficiently [1]. Many different implementations for these various networks have been proposed recently; see for example [2-4]. A very flexible way of implementing space-variant permutation networks is by using diffractive lenslet arrays [5, 6]. The basic concept is to give each optical channel its own miniaturized optical system, consisting typically of two diffractive off-axis lenslets. By controlling the angle under which the light beams travel, it is possible to realize arbitrary interconnect schemes. The optical setup for this is shown in Figure 1. Using lithographic techniques, all components in an array can be fabricated at the same time with high alignment precision. In order to achieve high efficiencies diffractive optical elements can be implemented as phase structures with multiple discrete phase levels [7, 8]. A 2-D cyclic shifter was demonstrated recently using lithographically fabricated lenslet arrays [9]. An experimental result is shown in Fig. 2.

One problem that occurs when two physically separated lenslet arrays are used is the difficulty of alignment. A lateral or longitudinal misalignment between the input plane and the first array or between the array 1 and array 2 may result in crosstalk where light from one input pixel couples over to the wrong output position. Our goal is to eliminate this alignment problem by integrating lenslet arrays on single substrates and by using the concept of planar integrated optics for realizing the system [10]. Fig. 3 visualizes this idea. Two 1-D arrays of diffractive micro-lenses are placed on one side of a glass substrate. The input and output positions are on the other side of the substrate opposite the lenses. The light paths are folded inside the substrate. The typical path for a beam of light is shown in Fig. 3 b. Since each input pixel is located on the optical axis, it is imaged through the pair of off-axis lenses without spatial aberrations. Furthermore, the symmetry of the system makes the optical signal very immune to wavelength shifts of the input beam. An integrated version of an optical permutation network can be built on small substrates that reduce mechanical and thermal problems.

References:

- [1] M. J. Murdocca, A. Huang, J. Jahns, and N. Streibl, "Optical design of programmable logic arrays," *Appl. Opt.* 27 (1988) 1651.
- [2] A. W. Lohmann, W. Storck, and G. Stucke, "Optical perfect shuffle," *Appl. Opt.* 25 (1986) 1530.
- [3] J. Jahns and M. J. Murdocca, "Crossover networks and their optical implementation," *Appl. Opt.* 27 (1988) 3155.
- [4] J. Jahns, "Optical implementation of the Banyan network," *Opt. Comm.* 76 (1990) 321.
- [5] A. W. Lohmann and F. Sauer, "Holographic telescope arrays," *Appl. Opt.* 27 (1988) 3003.
- [6] B. Robertson, E. J. Restall, M. R. Taghizadeh, and A. C. Walker, "Space-variant interconnection networks in dichromated gelatin," *Proc. SPIE* 1319 (1990) 128.
- [7] H. Dammann, "Blazed synthetic phase-only holograms," *Optik* 31 (1970) 95.
- [8] G. J. Swanson and W. B. Veldkamp, "Diffractive optical elements for use in infrared systems," *Opt. Eng.* 28 (1989) 605.
- [9] J. Jahns and W. Däschner, "Optical cyclic shifter using diffractive lenslet arrays," to appear in *Opt. Comm.*
- [10] J. Jahns and A. Huang, "Planar integration of optical components," *Appl. Opt.* 28 (1989) 1602.

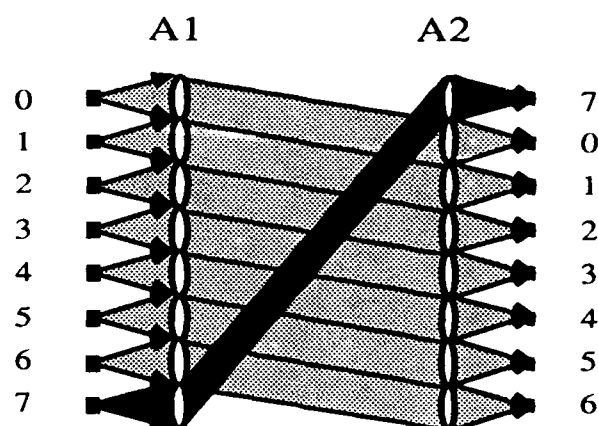


Fig. 1: Space-variant optical interconnects using lenslet arrays. Here a cyclic shifter is shown. A1 and A2 are the lenslet arrays.

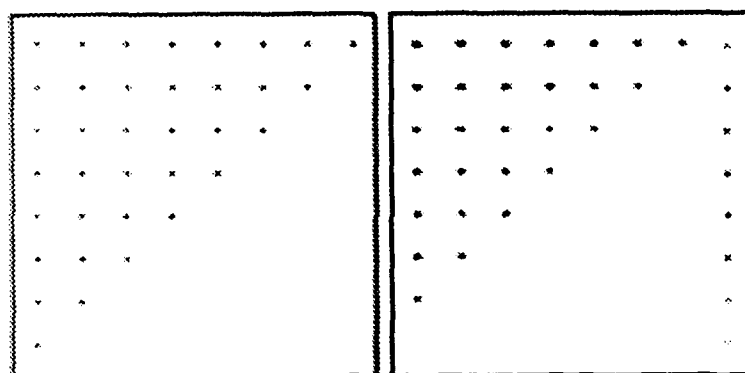


Fig. 2: Experimental result for a cyclic shifter (left: input, right: output).

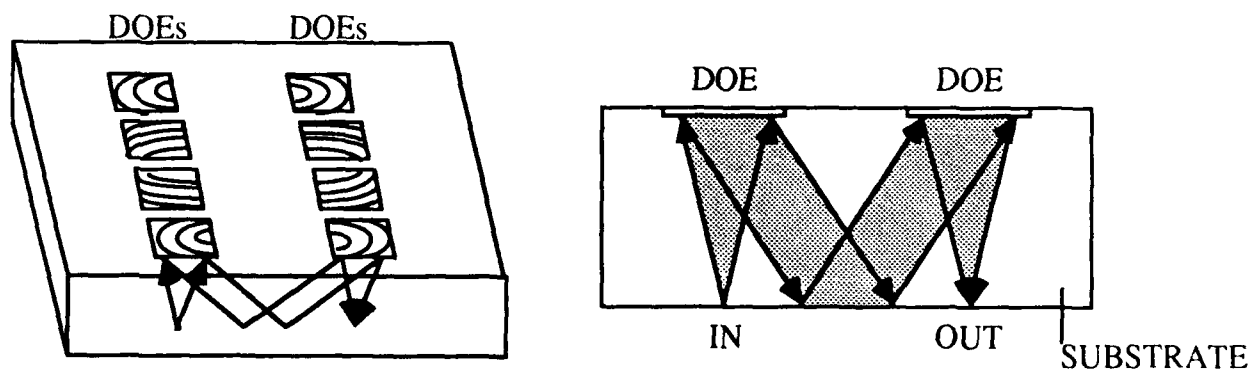


Fig. 3: left: Integrated-optical interconnections on single substrate; right: single optical channel; side view (DOE: diffractive optical element).

OPTICAL BUS INTERCONNECTION SYSTEM BY USING SELFOC LENSES AND PLANAR MICROLENS ARRAYS

Kenjiro HAMANAKA

Tsukuba Research Laboratory, Nippon Sheet Glass Co. Ltd.
5-4, Tokodai, Tsukuba-City, Ibaraki-Pref., 300-26, Japan
(Phone No.+81-298-47-8681, Fax No.+81-298-47-8693)

1. Introduction

Optical interconnection in optical processing system or in electronic computer system has many potential advantages in terms of channel capacity, transfer rate and so on. Recently, many types of optical interconnection system have been proposed and examined[1,2]. Free-space (three-dimensional) optical system can deal with the large amount of information, however, it is difficult to achieve precise assembling and high durability.

Here, we have proposed an optical bus interconnection system by using SELFOC lenses[3] and planar microlens arrays[4,5], which has possibility to overcome the problems associated with the free-space optical system. A way for fabrication, features, optical properties and possible applications are discussed with some experimental results.

2. Configuration and features

Figure 1 shows the configuration of the system which is applied to board-to-board interconnection. SELFOC rods are aligned and fixed on a substrate with grooves (Fig.1(b)). Then, perpendicular gaps are fabricated by using slicing machine. The position of the individual gap is suitably determined so that the conjugate image planes of unit magnification are located at the same positions in all the gaps (Fig.1(c)). In practice, the SELFOC rod is divided into many collimated lenses and they form a telecentric optical system which is suitable for cascade interconnection (Fig.1(d)).

When a LED matrix array is placed at one end of the each SELFOC rod and a mirror is fixed at another end, the image of the LED pattern can be transmitted to all the conjugate planes. Consequently, a signal generated at the LEDs can be led into many electronics circuit boards, if the boards with transparent type photodetector arrays are inserted in the gaps. This signal

transmission is applicable to clock distribution, for instance. The word "transparent" means a small absorption ratio (several percent or so) as well as a large transmission ratio. A thin layer of a-Si, works as photodiode or photoconductor, is an example of the transparent detector.

On the other hand, one of the circuit boards can work as a "talker" while a light pattern is displayed on the board by using a transmission type SLM, e.g. liquid crystal SLM, fabricated on it. The image of the displayed pattern is transferred to any other conjugate image planes, then the information of the pattern can be detected at the individual transparent photodetector arrays. If the individual conjugate plane is divided into two half parts, one is for the SLMs and another is for the photodetector array, or if the LED array is put at the both ends of the SELFOC rod, bi-directional optical interconnection is accomplished. As the result, the signal from the arbitrary board can be transferred into all the other boards, i.e. the optical bus interconnection can be completed.

Since the both lenses are small gradient index lenses fabricated by an ion exchange technique, all the input/output plane in the optical system are plane. Therefore, many kinds of optical or opt-electronic component such as SLMs, light source arrays, photo-detector arrays, spatial filters, etc. can be easily assembled by contacting the plane surfaces each other. The in-plane alignment is only essential in this case. Anti-reflection coating is not required while using liquid or resin for index matching at the boundaries. Coaxial alignment of the many SELFOC lenses in the optical channel is realized perfectly because it is fabricated from one SELFOC rod by the slicing technique. It allows us to transmit the high resolution image through long distance. Moreover, since it is possible to fill glass material through the optical pass, good reliability associated with the temperature variation, etc. is achievable.

3. Neural interconnection

It is obvious that the applications of the optical bus system are not restricted to the board-to-board interconnection. Here, we discuss an application for neural network briefly. In the most case, neural interconnection is based on matrix-vector multiplication which is easily implemented by using multiple imaging system. If the planar microlens array is inserted at one of the gaps, then the multiple imaging system is composed of the

SELFOC lens and the microlens array (Fig.2(a))[6,7]. While an output signal of neurons is displayed by the LED array and a weighted matrix of the neural interconnection is represented at a SLM, a resultant neural signal can be obtained at a photodetector array (Fig.2(b)). One-to-many and many-to-one interconnections are also applicable to many kinds of parallel processing such as pattern recognition.

The resolving power of the multiple images are evaluated experimentally. A SELFOC microlens ($\phi 3\text{mm}$, $f=3\text{mm}$) and a planar microlens array ($\phi 0.2\text{mm}$, $f=1\text{mm}$, pitch= 0.4mm) are utilized in the experiment. The MTF of the multiple imaging system is measured from the multiple images of a resolution test pattern. The modulation depth is about 50% at 120lp/mm in average, as shown in Fig.3. The result indicates that 21 optical channels which contain 100×100 resolving elements in each, can be utilized in the small image plane of 3mm in diameter.

4. Conclusion

A novel optical bus interconnection system by using SELFOC lenses and planar microlens array has been proposed. The system has the advantages of precise alignment, easy assembling and high durability. It is applicable for board-to-board interconnection, neural processing and so on.

The author appreciates K.Koizumi, K.Nishizawa and T. Kishimoto, Nippon Sheet Glass Co. Ltd., for their supports.

References

- 1) I.Glaser, et al., Opt.Lett. 11 (1986) 53.
- 2) J.Jahns, et al., Appl.Opt. 28 (1989) 1602.
- 3) K.Matsushita, et al., Appl.Opt. 19 (1980) 1070.
- 4) K.Iga, et al., Appl.Opt. 21 (1982) 3456.
- 5) M.Oikawa, et al., Proc.of IEEE 898 (1988) 3.
- 6) K.Hamanaka, et al., Appl.Opt. 29 (1990).
- 7) K.Hamanaka, et al., J.J.A.P.Letters 29 (1990) 1277.

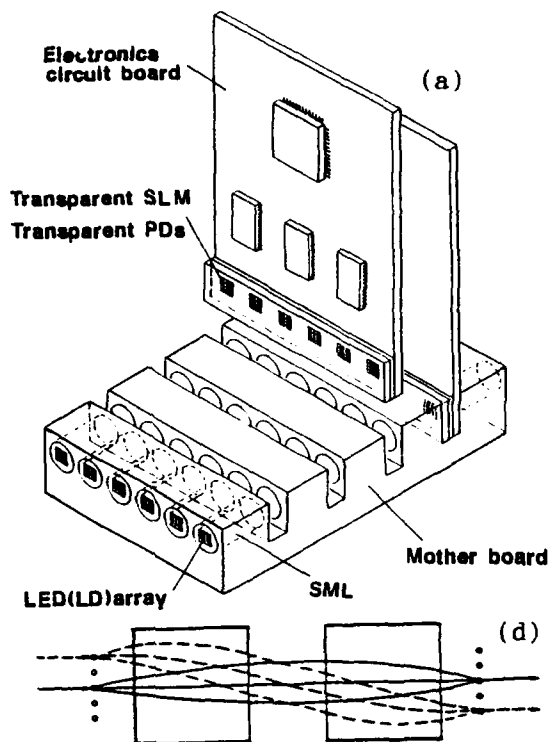


Fig.1
Optical bus interconnection system.
(a),(c) System configurations.
(b) A method for fabrication.
(d) A pair of collimated lens
for cascade interconnection.

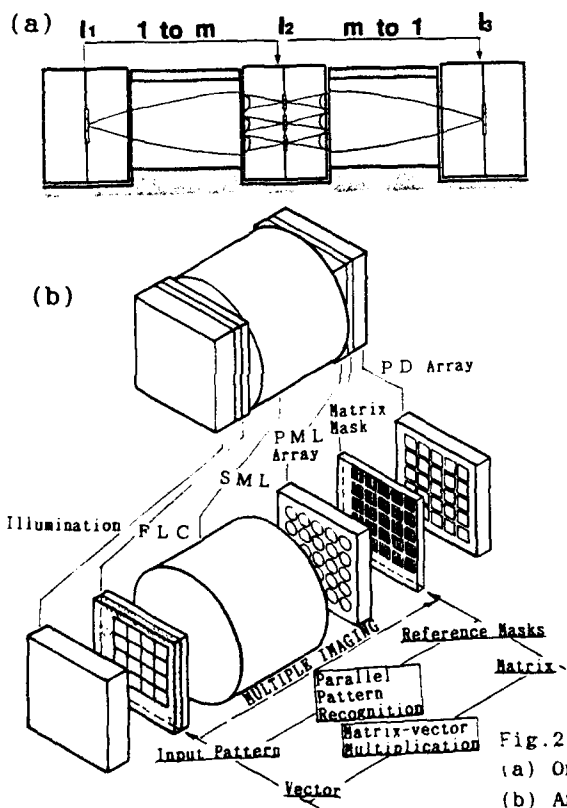
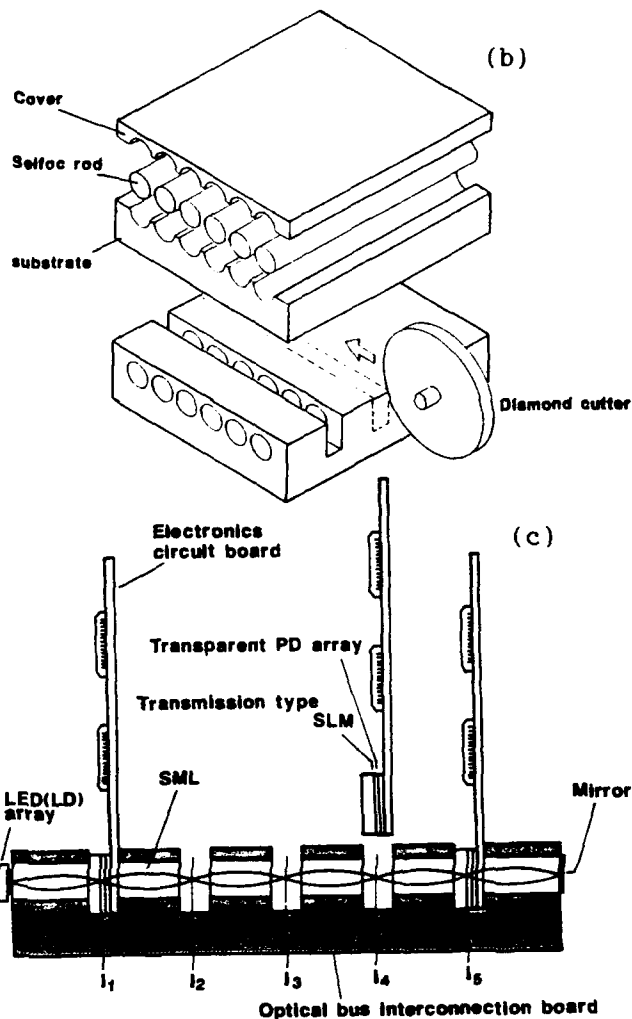


Fig. 2

(a) One-to-many/many-to-one interconnection.
(b) An application of the 1-to-m interconnection.

Fig. 3

The MTF measured by the 1-to-m experimental system.

Monday, March 4, 1991

Optical Interconnections

MC 2:00pm-3:20pm
Salon F

Ravindra A. Athale, *Presider*
George Mason University



Spatial Noise Reduction in Array Illuminators

Adolf W. Lohmann
NEC Research Institute, Inc.
4 Independence Way
Princeton, NJ 08540
TEL: 609-951-2653
FAX: 609-951-2482

Stefan O. Sinzinger
Physikalisches Institut der Universitat
Staudt Str. 7
8520 Erlangen
Germany
TEL: 49-9131-858376
FAX: 49-9131-13508

1. Motivation, Plan

An ideal array illuminator would provide equal amounts of light power to all elements of an array of gates or smart pixels. Existing array illuminators (abbreviated: AIL) achieve a homogeneity of 5 to 10%. That may seem to be good enough, if the signals are binary. However, it is desirable to achieve the best possible homogeneity, since there might be other causes for inhomogeneous behavior of the array system. Having a good AIL relieves the burden of tolerances for the other components of the overall system. Furthermore, when the signals are analog, as in some neural systems, the homogeneity of the power supply becomes even more important.

One of the main causes for any inhomogeneity is due to the coherence of the laser light. Hence, we want to reduce the degree of coherence. We do this by using more than one laser. These lasers should be manufactured identically, but operated without mutual phase coupling. The output of these lasers are intermingled in the array illuminator system. The more lasers participate the lower will be the temporal coherence. The lateral configuration of these lasers will cause a reduction of the spatial coherence. The lasers may be arranged for example as a grid of size 3x3 or larger.

In our proposed systems it will not be essential that all lasers operate at the same power level. The homogeneity of AIL output will not suffer even if one of the lasers breaks down completely. Hence, our systems will be favorable in terms of life time and manufacturing yield.

The existing and proposed AIL systems can be categorized into 5 groups⁽¹⁾. We will consider here three of these five types, beginning with "phase contrast"⁽²⁾ in section 2, followed by Dammann gratings⁽³⁾ in section 3, and by the Talbot-AIL⁽⁴⁾. Holo telescope arrays⁽⁵⁾ and grating coupler arrays⁽⁶⁾ have their own

merits, but they are probably a bit more vulnerable in terms of homogeneity. We will not hide short comings of our three types, such as loss of homogeneity at the edges of the array and clock skew.

2. The Phase Contrast AIL

The basic setup consists of a point source, which illuminates a phase grating by means of a collimating lens. The phase grating acts as an object which is converted into an amplitude image by means of a phase contrast imaging system. A phase shifting structure in the Fourier plane is responsible for the conversion of a uniform-intensity input into an array of bright dots in the image plane⁽²⁾.

The intensity distribution in the Fourier plane will consist of a few isolated bright spots if the object is periodic as is the case here. Now let us place a second laser into the source plane, shifted sideways by a certain amount. Light from this second laser will produce in the Fourier plane the same configuration of bright spots as the first laser, however shifted sideways. The location of the two sets of diffraction spots will be completely disjointed if the geometry of the setup is designed accordingly. Hence, we may provide another phase shifting structure for the light from the second laser. In other words, several phase contrast operations are interlaced.

The congestion of diffraction spots in the Fourier plane will increase while increasing the array of laser sources. The relative size of any diffraction spot is inversely proportional to the number of grating periods and hence inversely proportional to the number of elements of the AIL. As a consequence the number of admissible laser sources is directly proportional to the number of AIL elements. For a quadratic array the upper limit of the number of sources will be one ninth of the number of AIL elements. Hence, the power of the individual lasers may be quite low. The inhomogeneity of the source array, or even a few dead lasers, would not affect the homogeneity in the output plane of the AIL.

Two drawbacks ought to be mentioned. A local defect in the phase grating will cause a local defect in the output of the AIL. If all laser sources are triggered simultaneously their contributions to the central AIL element will be synchronized. However, at the edges of an $N \times N$ AIL there might be mutual delays in the order of up to $N\lambda/c$. In other words the temporal blur would be less than a picosecond, if N is less than one thousand.

3. The Dammann AIL

The basic setup consists of a point source, which illuminates a phase grating by means of a collimating lens. A second lens produces the Fraunhofer diffraction orders, which ought to be equally bright, ideally⁽³⁾. This AIL is very robust in terms of

local grating defects. The relative distances between diffraction spots can be very large. However, it is difficult to achieve array formats larger than 64×64 (see (1)).

Basically, in this setup one generates multiple images of the source. If the total source consists of an array of individual sources one will get an incoherent superposition of shifted diffraction pattern. The homogeneity of the output will benefit from the shifted superposition, except for the edge regions. Suppose there are $M \times M$ sources and $N \times N$ diffraction orders, then only the inner $(N-M) \times (N-M)$ dots will be more or less uniform in brightness. The remainder of the $(N+M) \times (N+M)$ spots will fall off in a trapezoidal manner.

4. The Talbot AIL

This AIL relies on the fractional Talbot effect which occurs when a grating is illuminated by a monochromatic plane wave⁽⁴⁾. The Talbot effect is also known by the name of "self imaging" since there are no lenses or any other components needed in the space between object and image plane.

If the illuminating plane wave is tilted the image will be shifted laterally. If the shift matches the grating period the image will remain where it was before. That remains true, if we use a set of tilted plane waves, each coming from a different laser. In terms of clock skew and edge effect the Talbot AIL is roughly equivalent to the two other kinds.

References

- (1) N. Streibl, "Array Generators", J. Mod. Opt. (1990).
- (2) A. W. Lohmann, J. Schwider, N. Streibl, J. Thomas, "An Array Illuminator Based on Phase Contrast", Appl. Opt. 27 (1988) 2915.
- (3) H. Dammann, W. Gortler, "High-Efficiency, In-Line Multiple Imaging by Means of Multiple Phase Holograms", Opt. Comm. 3, (1971) 312.
- (4) A. W. Lohmann, J. A. Thomas, "Realization of an Array Illuminator Based on the Talbot Effect", Appl. Opt. 29, (1990).
- (5) A. W. Lohmann, F. Sauer, "Holographic Telescope Arrays", Appl. Opt. 27, (1988) 3003.
- (6) T. Kubota, M. Takeda, "An Array Illuminator Using Grating Couplers", Opt. Lett. (1989).

Cellular Hypercube Interconnections for Optical Processor Arrays

by C. B. Kuznia and A. A. Sawchuk

University of Southern California

Signal and Image Processing Institute, MC 0272

Los Angeles, CA 90089-0272

I. Digital Optical Array With Cellular Hypercube Interconnections

Digital optical cellular arrays are single-instruction-multiple-data (SIMD) arrays of many low complexity (fine-grain) processing elements (PEs). The PEs themselves can be implemented by electronic or optoelectronic methods. These arrays have many general applications in numerical processing and symbolic substitution computing. They are particularly suited to bit plane images (images in which each PE represents a pixel, and each pixel takes on the value 0 or 1). In this application, each PE is referred to as a cell and is responsible for computing the output of one image pixel according to a single instruction broadcast to all PEs from a central control unit.

A *direct connection* is defined as an interconnection from one PE to another which is not routed through an intermediate PE. The *neighborhood* of a PE is defined as the set of PEs to which it is directly connected, which is determined by the inter-PE connection network of the cellular array. In this paper we concentrate on the cellular hypercube interconnection network [1, 2], which can be fully or partially implemented by optoelectronics. In the cellular hypercube of size $N \times N$ ($N = 2^l$), each PE is directly connected to other PEs in the up, down, left and right directions spaced at distances 2^n , $n = 0, 1, \dots, l - 1$. With this numbering convention, the maximum number of PEs in the neighborhood is $4 \log_2 N - 2$. The time needed to send one data bit from a PE to the directly connected PEs in its neighborhood is defined as one clock cycle.

By implementing binary image algebra (BIA), the cellular array can perform general image processing and data manipulation algorithms [1, 2]. Using BIA, any sequence of operations can be decomposed into three fundamental operations: 1. complement - complement each bit (pixel) stored in the PE array; 2. union - the Boolean union (OR) function is performed on two binary images cell by cell; 3. dilation - data from one binary image is replicated under control of a second image in a manner similar to spatial convolution. The complement and union operations are local operations. The third, dilation, is global and hence its execution is dependent on the connection network of the array. To perform dilation the cellular array must be able to shift its data any arbitrary direction and distance.

II. Communication Time in Cellular Hypercube Connected Arrays

We assume that each PE has its own integrated light source (LED, diode laser or light modulator) for transmitting data, and one or more detectors for receiving data from the $4 \log_2 N - 2$ total processors in its directly connected neighborhood. The interconnection paths are made via a shift-invariant optical fan-out system which can simultaneously image the output of each PE onto the detectors of directly connected PEs. A 1-D example for $N = 16$ is shown in Fig. 1, in which two PEs (0 and 11) can transmit simultaneously without superposing their outputs on the same detector. Here we consider the trade-off between the execution time for various communication operations and the number of detectors per PE in the cellular hypercube interconnected PE array. The execution time for a communication step is defined as the number of clock cycles needed for each PE in the array to shift one data bit to another PE. In general, several clock cycles are needed for each communication step because the outputs of several light sources

operated simultaneously cannot be allowed to hit the same detector.

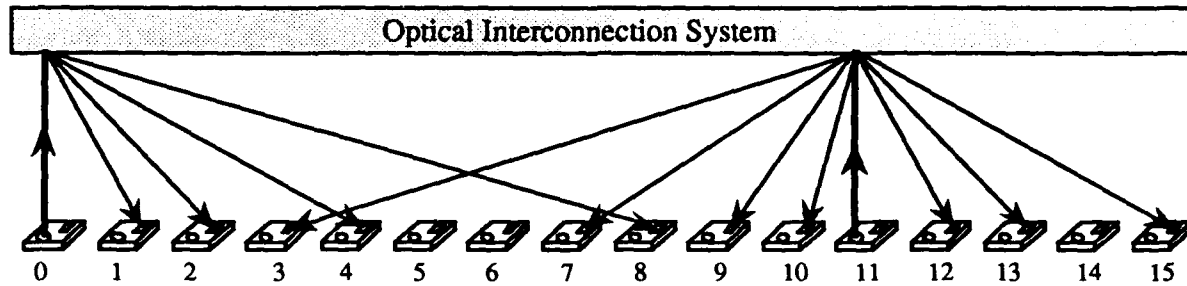


Figure 1: One dimensional cellular hypercube for $N = 16$.

Since only one bit is being sent per PE and the imaging system fans the light out to the proper detectors in all of the connected PEs, only one source is needed per PE. However, there are advantages to having more than one detector per PE because the PE could receive more than one input at a time. If each PE in the array has $4 \log_2 N - 2$ detectors (one for each connected PE), then all data bits can be received at once and a communication step requires only one clock cycle. In this situation, all PEs can transmit data at the same time. However, because of the limited resolution of the imaging system and need to minimize the physical PE size, it might not be possible to integrate $4 \log_2 N - 2$ detectors into each PE. Reducing the number of detectors results in more clock cycles needed to perform a communication step. In the example of Fig. 1, each PE contains only one detector. In order that no detector receives more than one data bit during a single clock cycle, eleven clock cycles are needed.

In Table 1 a comparison is made between the number of clock cycles per communication step and the number of detectors per PE for a one dimensional array. The left half labeled 'UNIT DISTANCE ELECTRONIC' shows the results obtained when the nearest neighbor PEs (unit distance) are interconnected electronically rather than optically. This stems from the idea that short distance connections are best made with electronics and the inter-cell distances may be small [3]. The reduction in clock cycles is due to the availability of one additional detector.

The table is determined by finding a value M so that every M th PE (PEs numbered $0, M, 2M, \dots$) transmits at once. Since the connection pattern is spatially-invariant, any set of PEs numbered $0 + m, M + m, 2M + m, \dots$ can transmit simultaneously. To allow for each PE in the array to transmit its data, the value m is incremented during each clock cycle from 0 to $M - 1$, and then the corresponding set of PEs transmits its data. The value M is then the number of clock cycles needed for each communication step. To implement an arbitrary shift of data in the array, $O(\log_2 N)$ communication steps are required in general, each of which consists of one or more clock cycles as shown in Table 1.

The generalization to a two-dimensional array is straight forward if the number of detectors per PE is doubled and the detectors are dedicated to one of the dimensions. The number of clock cycles needed for a communication step is not increased from the one-dimensional case.

III. Design of Binary Phase Gratings to Implement Cellular Hypercube

The cellular hypercube can be implemented by using a phase grating and a Fourier transforming optical system. The idea is to use a computer to design a phase grating, $f(x, y)$, such that its power spectrum, $|FT[f(x, y)]|^2$, is the cellular hypercube pattern.

The Damman grating is one such computer generated binary phase grating that could be

used to generate the cellular hypercube function. Dammann gratings typically exhibit $\sim 60-70\%$ diffraction efficiency with less than 10% variation of intensity between orders and have a 10 dB on/off ratio for arbitrary patterns [4]. The grating can be solved numerically in a variety of techniques for array sizes $N < 50$, however, for larger N the computing complexity becomes extreme [5].

This section describes a computationally easy way to create binary phase gratings for the cellular hypercube interconnection pattern. For now, we look at the one-dimensional cellular hypercube pattern which can be generated by taking the power spectrum of a summation of sine functions at frequencies $f_n = 2^n, n = 0, \dots, \log_2 N - 1$. In the following, $Y(f)$ is the Fourier transform of $y(x)$ and the phase terms r_n are lost when computing the power spectrum.

$$y(x) = \sum_{n=0}^{\log N-1} \beta_n \sin(f_n 2\pi x + r_n) \xrightarrow{\text{Power Spectrum}} |Y(f)|^2 = \sum_{n=-(\log N-1)}^{\log N-1} \frac{\beta_n^2}{4} \delta(f - f_n). \quad (1)$$

However, the function $y(x)$ cannot be put into phase grating form due to areas of negative light transmission. A binary phase grating can be constructed by hard-limiting $y(x)$ to create a new function $yt(x)$ defined as follows:

$$yt(x) = \begin{cases} +1 & \text{if } y(x) \geq 0 \\ -1 & \text{otherwise} \end{cases} \quad (2)$$

The Fourier transform of $yt(x)$ is the cellular hypercube pattern with additional noise added due to the hard-limiting process. The noise creates unwanted variations in the frequency components f_n by spreading the power into other unwanted frequency components. By adjusting the weights β_n and phase values r_n in the original function $y(x)$, the variations in the frequency components f_n can be minimized.

Taking $y(x)$ to be a single sinusoid at frequency f_o , then $yt(x)$ is a square grating at frequency f_o . The power spectrum exhibits the odd harmonics of f_o with the first sidelobe down 9.5dB, a value which is the maximum on/off ratio of this design. As more sinusoids are added to $y(x)$ the frequency components will now depend on the relationships between the phases r_n , the frequencies f_n and the frequency weights β_n .

The function $yt(x)$ with 1024 sample points has been solved several times for a binary phase grating giving a hypercube connection pattern of dimension 8. The computed FFT output has a typical dynamic range of 1.07 and on/off ratio of 9db. The diffraction efficiencies as computed from the power spectrum ranged from 60 – 62%. The power spectrum of a typical $yt(x)$ is shown in Fig. 2a. No attempt has been made to lower the background noise. Two-dimensional cellular hypercube patterns have also been solved for a 256×256 array, and a portion is shown in Fig. 2b.

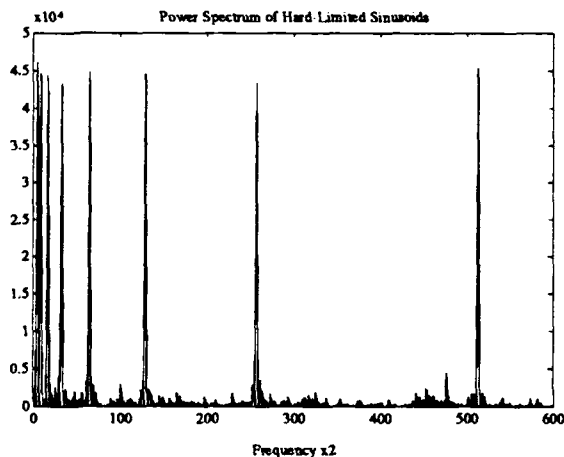
References

- [1] K.S. Huang, B. K. Jenkins, and A. A. Sawchuk, "Binary Image Algebra and Optical Cellular Logic Design", *Computer Vision, Graphics and Image Processing*, vol. 45, pp. 295-345, (1989).
- [2] K. S. Huang, B. K. Jenkins and A. A. Sawchuk, "Image Algebra Representation of Parallel Optical Binary Arithmetic", *Applied Optics*, vol. 28, pp. 1263-1278 (1989).
- [3] D. A. B. Miller "Optics for low-energy communication inside digital processors: quantum detectors, sources, and modulators as efficient impedance converters", *Opt. Lett.*, 14, 146-148 (1989).

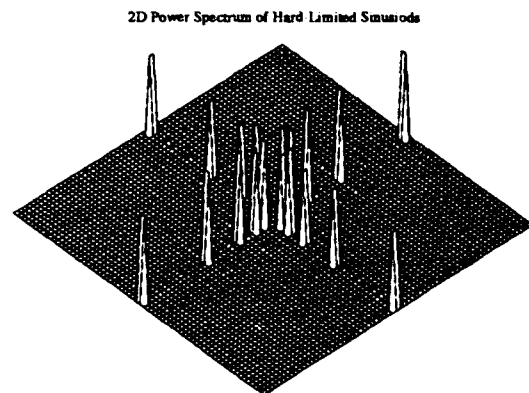
- [4] J. Turunen, A. Vasara, J. Westerholm, "Stripe-Geometry Two-Dimensional Dammann Gratings", *Opt. Comm.*, 74, 245-252 (1989).
- [5] J. Jahns, M. M. Downs, M. E. Prise, N. Streibl and S. J. Walker, "Dammann Gratings for Laser Beam Shaping", *Opt. Eng.*, 28, 1267-1275 (1989).

UNIT DISTANCE OPTICAL				UNIT DISTANCE ELECTRONIC		
N	#of det/PE	#of clk cys/comm. step	% of PEs on	#of det/PE	#of clk cys/comm. step	% of PEs on
32	1	11	9	1	11	9
	2	9	11	2,3	5	20
	3,4	5	20	4-6	3	33
	5,6	3	33	7	1	100
	7,8	2	50			
	9	1	100			
128	1	19	5	1	13	8
	2	11	9	2	9	11
	3	7	14	3-5	5	20
	4-6	5	20	6-10	3	33
	7-10	3	33	11	1	100
	11,12	2	50			
	13	1	100			
512	1	19	5	1	19	5
	2	11	9	2	9	11
	3,4	7	14	3	7	14
	5-8	5	20	4-7	5	20
	9-14	3	33	8-14	3	33
	15,16	2	50	15	1	100
	17	1	100			

Table 1. Number of clock cycles per communication step vs. no. of detectors per PE for array size N .



(a): 1D connection distances 2^n , $n = 0, 1, 2, 3, 4, 5, 6, 7$



(b): 2D connection distances 2^n , $n = 0, 1, 2, 3, 4$

Figure 2: One and two dimensional cellular hypercube interconnection patterns computed from hard-limited sinusoids.

Multiplexed Hybrid Interconnection Architectures

Haldun M. Ozaktas
Joseph W. Goodman
Information Systems Laboratory, Durand Building
Department of Electrical Engineering
Stanford University
Stanford, California 94305

1 Introduction

A major advantage of optical and superconducting interconnections is their ability to transfer large amounts of information per unit cross section over long distances. Let the maximum information flux a given communication medium can support be denoted by \mathcal{I} and be measured in *bits/m²sec*. For the length scales involved in a computing system (< 10 m), it is possible to reduce the effects of dispersion and attenuation to the extent that \mathcal{I} may be assumed to be independent of length for optical and superconducting interconnections. On the other hand, \mathcal{I} is a decreasing function of communication length for resistive interconnections, making them disadvantageous over longer distances. However, for distances less than about the order of a centimeter, they can provide greater information flux than optical or superconducting interconnections.

Let T denote the minimum pulse repetition interval for a single physical optical communication channel (i.e. corresponding to a single spatial degree of freedom). Since we are ignoring dispersion, T will probably be set by the speed of the switching devices or electrooptic transducers. If wavelength division multiplexing is employed, an appropriate effective value of T should be used.

We assume that we would like to establish a pre-specified pattern of $n/2$ pairwise connections among a collection of $n \gg 1$ points. For simplicity the extension to fan-out and fan-in is not considered. Although we restrict ourselves to a fixed connection pattern, the extension to reconfigurable or message routing systems is possible. We also limit ourselves to single layer 2 dimensional layouts, the extension to multi-layer and 3 dimensional layouts being straightforward. B will denote the rate at which binary digi-

tal pulses are emitted into each connection. Our purpose is to implement the given pattern of connections in a manner that results in smallest possible system area, which we assume is dominated by the space required for establishing communication.

The number of binary pulses in transit at any given time in an optical communication network occupying area A may not exceed $\sim A/(f\lambda cT)$, where c and λ denote the speed of light and wavelength of radiation respectively [1]. f is a dimensionless constant factor which in principle can approach the order of unity, but may be quite larger in practice. Starting from this relation, it is possible to derive an approximate lower bound on the linear extent L of our system

$$L = A^{\frac{1}{2}} \geq \kappa n^q (BT) f \lambda \quad (1)$$

where κ is a constant coefficient and $1/2 \leq q \leq 1$ is a measure of the connectivity of the system [2] [3] [4]. This bound represents the intrinsic information carrying capacity of optical wavefields and applies to any architecture or implementation. Notice the tradeoff between system size and B .

One way of implementing the desired pattern of connections is simply to allocate $\lceil BT \rceil \simeq \max(BT, 1)$ parallel channels between every pair of points to be connected. When $BT \geq 1$, such an implementation is as efficient as any other in terms of making maximum usefulness of the available capacity of the optical channels. In this case, the above lower bound may be approached, for instance, by use of waveguides with effective line to line spacing of $\sim f\lambda$. However, if B is less than $1/T$, the channels are underutilized and the bound of equation 1 cannot be approached, since no matter how small B is, a channel with capacity $1/T$ is allocated for every pairwise connection. Thus when $B < 1/T$, the layout area is not any less than when $B = 1/T$, so that L can at best approach

the bound

$$L \geq \kappa n^q f \lambda. \quad (2)$$

In this paper we concern ourselves with methods of restoring the broken tradeoff between system size and B when $BT < 1$.

To achieve our objective, we would like to multiplex $1/BT > 1$ independent signal paths into the same physical channel, so as to saturate its capacity. However, this is not straightforward when the many signal paths have distinct source and destination localities. In the next sections we describe 3 architectures which enable information flow to be organized in a manner enabling overlap between such signal paths, allowing them to be multiplexed. The reduction in the number of physical channels thus possible results in a decrease in system size and propagation delay for communication limited layouts.

2 The multiplexed grid architecture

The multiplexed grid architecture is based on the family of k -ary m -dimensional meshes (grids) of $k^m = n$ nodes [5]. The hypercube is a special case with $k = 2$ and $m = \log_2 n$. For sake of illustration, we consider the case $m = 2$ and $k = n^{1/2}$, which corresponds to the familiar planar mesh with $n^{1/2}$ nodes on an edge. An arbitrary connection is established in several nearest neighbor (in m -space) 'hops', and multiplexed together with other connections with which it overlaps, as illustrated in figure 1. If at least $1/BT$ connections can be overlapped along each edge of the mesh, then complete utilization of the available capacity $1/T$ of the physical channels may be achieved. Finally, the multiplexed m dimensional mesh is laid out in 2 dimensions, as described in [5]. Of course, this is a trivial task when $m = 2$.

The price that must be paid in return for efficient utilization of the high capacity optical channels is the additional area cost and delays associated with demultiplexing and remultiplexing of independent signal paths. Low dimensional meshes allow a larger number of connections to be overlapped, but increase the number of hops, and hence the number of device delays a signal must go through. High dimensional meshes decrease the number of hops but do not enable as many signal paths to be overlapped and multiplexed, possibly resulting in less than complete utilization of the capacity of the channels and

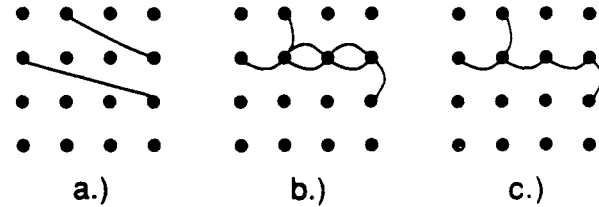


Figure 1: The multiplexed grid architecture with $m = 2$, $k = 4$ and $n = 16$. Part a.) shows two of many to-be-established connections. Part b.) shows each connection established in several hops. Part c.) shows overlapping portions of these connections multiplexed into high capacity channels, reducing the total number of physical channels and thus layout area.

thus larger layout area and propagation delays. The optimal value of m minimizing overall signal delay (propagation plus device) is found to decrease with increasing n and asymptotically approaches 2 for 2 dimensional layouts. In this case $\propto n^{1/2}$ device delays are suffered in the worst case [6].

3 The multiplexed global interconnection architecture

We now turn our attention to another architecture, illustrated in figure 2. The n points among which connections are to be established are partitioned into n/n_1 'modules' of n_1 points each. All connections between points in one particular module to another particular module are bundled together and multiplexed into the smallest possible number of physical channels. The relatively short connections between points in the same module are made directly and would probably be implemented with conductive wiring, because of the greater density they offer over short distances.

The larger the value of n_1 , the larger the number of connections between each module pair, so that a greater number of independent signal paths may be bundled (overlapped) and multiplexed together, resulting in a reduction of the area consumed by global communication channels. On the other hand, increasing n_1 increases the area required by the internal con-

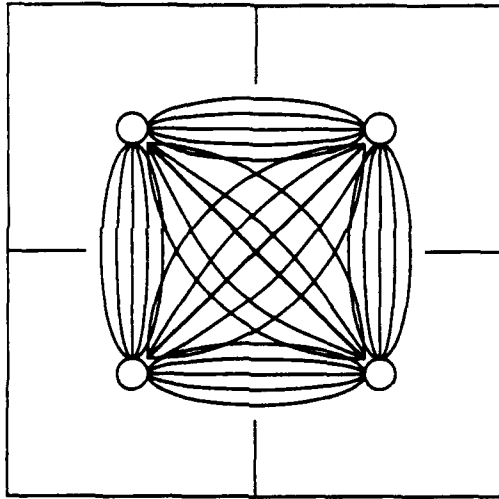


Figure 2: *The multiplexed global interconnection architecture with $n/n_1 = 4$.* Connections internal to a module (not shown) are made directly, probably with conductive wiring. A connection to a destination in another module is first wired to a common locality with other connections destined to the same target module and multiplexed together. Demultiplexing takes place at the destination module, followed by wiring to the individual destinations. Thus 2 device delays are involved for global connections.

nections. Thus there is an optimal value of n_1 resulting in minimum system area.

The multiplexed global interconnection architecture is not very useful for applications exhibiting a great degree of locality. In such systems there will not be enough connections between distant module pairs to saturate the capacity $1/T$ of a single physical channel. It may be useful, for instance, for the implementation of fine grain parallel random access machine models [6] or connectionist systems.

4 The multiplexed fat-tree architecture

The fat-tree architecture, illustrated in figure 3, was first advocated by Leiserson [7] in a multiprocessor interconnection context. We define the fat-tree to have $[n^q] \simeq n^q$ connections emanating from sub-trees containing n' points. This rate of growth of capacity as we climb the tree is consistent with a layout with measure of connectivity q as introduced

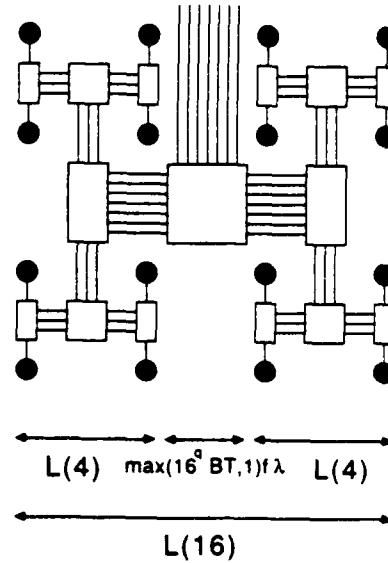


Figure 3: *The multiplexed fat-tree architecture.* The points to be connected are located at the leaves, and the internal nodes provide demultiplexing and remultiplexing functions. Each connection is established in several hops, $2 \log_2 n$ in the worst case. The number of connections emanating from the sub-trees increase as we go up the tree. The overlapping portions of the connections are multiplexed into the smallest possible number of physical channels.

in the first section [2].

For concreteness, let us assume that waveguides of effective line to line spacing of $f\lambda$ are used. Noticing that the n^q connections emanating from a sub-tree of n' points can be multiplexed into $\max(n^q BT, 1)$ physical channels, and assuming the area required for the multiplexing functions not to be the limiting factor, it is possible to show that the linear extent of the fat-tree approximately satisfies

$$\begin{aligned} \max(n^q(BT), n^{\frac{1}{2}})f\lambda &\leq \\ L &\leq \max(n^q(\log_2 n^{\frac{1}{2}})(BT), n^{\frac{1}{2}})f\lambda \end{aligned} \quad (3)$$

The second term is unavoidable for any 2 dimensional layout. The first term corresponds to the communication area and is what we are interested in. Upon comparison with equation 1, we observe that the multiplexed fat-tree allows the smallest possible system size to be approached within a logarithmic factor. (Of course, if BT is not small enough to satisfy $BT < \kappa/\log_4 n$, the use of a fat-tree may not prove advantageous.) What essentially happens is that the total communication area is dominated by

the longer higher level connections, which we succeed in multiplexing to the greatest possible extent.

Once again the price paid is the cost of multiplexing functions and the additional device delays incurred. It would probably be preferable to implement the shorter lower level connections without demultiplexing and remultiplexing at every step up the tree, and with conductive wires. This would enable reduction of the number of device delays incurred and the multiplexing circuitry. A detailed simulation would reveal the level beyond which multiplexing and optical interconnections should be utilized.

5 Conclusion

We have discussed the importance of organizing information flow in a manner enabling maximum multiplexing of independent signal paths, leading to a reduction in the number of area consuming longest interconnections, which results in smaller communication area and propagation delays. Among the architectures discussed, the fat-tree is near optimal in this respect.

The latter two of the presented architectures provide a natural environment for the joint use of optical and conducting interconnections so as to bring out the best in both and may prove more promising than simple replacement of individual long wires with optics. Optical interconnections are used to provide high density/bandwidth multiplexed information transfer over long distances. Submicron scaled normal conductors are used to provide communication at a density unachievable with optics over shorter distances. This is also consistent with the energetic properties of the interconnection media. Optical interconnections consume less energy per transmitted bit over longer distances compared to normal conductors [8] [9] [10].

Both the multiplexed global interconnection architecture and the fat-tree architecture are especially suited for high density (i.e. f close to unity) free-space optical implementations because of the regular pattern of interconnections.

Detailed quantitative analysis and simulation of these architectures will be the subject of subsequent expositions.

This work was supported by the Air Force Office of Scientific Research, Grant No. AFOSR-88-0024.

References

- [1] Haldun M. Ozaktas and Joseph W. Goodman. Lower bound for the communication volume required for an optically interconnected array of points. *Journal of the Optical Society of America A*, 7:2100-2106, 1990.
- [2] Michael Feuer. Connectivity of random logic. *IEEE Transactions on Computers*, 31:29-33, 1982.
- [3] W.E. Donath. Wire length distribution for placements of computer logic. *IBM Journal of Research and Development*, 25:152-155, 1981.
- [4] W. E. Donath. Placement and average interconnection lengths of computer logic. *IEEE Transactions on Circuits and Systems*, 26:272-277, 1979.
- [5] William J. Dally. *A VLSI Architecture for concurrent data structures*. Kluwer Academic Publishers, Norwell, Massachusetts, 1987.
- [6] Alfred C. Hartmann and Jeffrey D. Ullman. Model categories for theories of parallel systems. In G.J. Lipovski and M. Malek, editors, *Parallel Computing: Theory and Experience*. John Wiley and Sons, 1986.
- [7] Charles E. Leiserson. Fat-trees: Universal networks for hardware-efficient supercomputing. *IEEE Transactions on Computers*, 34:892-901, 1985.
- [8] D.A.B. Miller. Optics for low-energy communication inside digital processors: Quantum detectors, sources and modulators as efficient impedance converters. *Optics Letters*, 14:146-148, 1989.
- [9] Michael R. Feldman, Sadik C. Esener, Clark C. Guest, and Sing H. Lee. Comparison between optical and electrical interconnects based on power and speed considerations. *Applied Optics*, 27:1742-1751, 1988.
- [10] Raymond K. Kostuk, Joseph W. Goodman, and Lambertus Hesselink. Optical imaging applied to microelectronic chip-to-chip interconnections. *Applied Optics*, 24:2851-2858, 1985.

Two Dimensional Spatially Variant Optical Interconnects

E. J. Restall, B. Robertson, M. R. Taghizadeh and A. C. Walker
Holography Group, Physics Department, Heriot-Watt University,
Riccarton, Edinburgh EH14 4AS. Scotland.

Introduction.

Spatially variant interconnects (SVIs) show great potential in the fields of optical computing and optical communications. Two dimensional forms of these interconnects offer even more power over their stacked one dimensional¹ or wrap around counterparts².

Several ingenious classical optical approaches³ exist to implement these complicated optical routing patterns, however all exhibit scalability and efficiency problems which make them unsuitable for use in a practical connection scheme. The interconnects we describe are recorded in dichromated gelatin (DCG), utilising its high space band width product, efficiency and good uniformity. They offer point to point on-axis interconnection with a high packing density (compatible with current demonstration optical circuits^{4,5}) and have been generated for a range of wavelengths.

Furthermore, we have coupled several interconnection stages together (in both transmission and reflection modes of operation) to demonstrate prototype networks.

Recording the Interconnects.

The interconnects are double or quadruple element holograms depending upon the nature of the interconnection pattern. An asymmetric interconnect pattern, such as the perfect shuffle (figure 7 a.), requires four elements whereas a symmetric pattern, such as the Banyan, requires only a double element. The structure of such a doublet is shown in figure 3, a collimating element takes the cones of light from an array of point sources and couples them into the next element at a common angle. This redirecting element generates the interconnect pattern and this is then either coupled back into the first doublet (the symmetric case) or passes through another doublet performing the inverse interconnect (the asymmetric case) and is then focussed back to a point. Similar paired single element interconnects, which collimate and redirect in one stage, have been demonstrated¹ but they can only be replayed at the recording wavelength, this new approach facilitates replay at any wavelength specified at the design stage. Typical recording arrangements used for the collimating and redirecting elements are shown in figures 1 and 2, respectively. We have already demonstrated the recording of high quality lenslet arrays for the visible and near infra-red⁶ and similar quality planar re-directing gratings are comparatively straight forward.

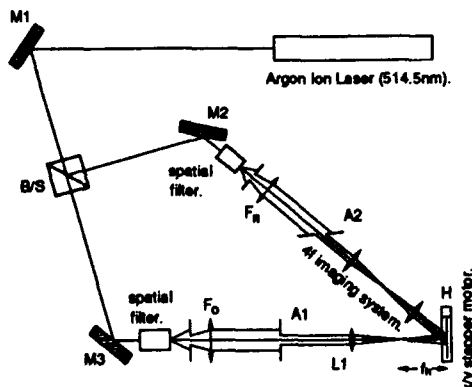


Figure 1. Holographic lens recording geometry.

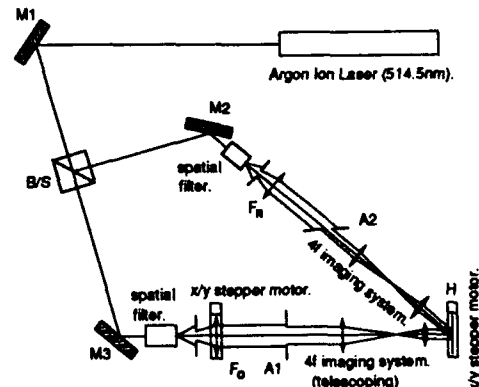


Figure 2. Holographic SVI recording geometry.

The required redirection of beams is achieved during recording by translating a collimating lens, figure 2, with a stepper motor system¹. The full 2D required interconnection pattern results from a step and repeat process using two independent stepper motor systems controlled by the same computer system.

To compensate for diffraction spreading over the interconnection distance and to minimise cross-talk the redirection elements are designed to have some small amount of focussing power. Using these recording schemes we have produced several stages of a variety of 16x16 2D networks based on symmetric interconnection patterns and both the perfect shuffle and its inverse in full 2D. Efficiencies of 95% per element have been achieved, making the doublet efficiency (once cemented together) of the order of 90%.

Replay of Interconnects.

The two types of interconnect replay in different manners. The asymmetric interconnect is transmissive in nature and its mode of use is shown (schematically) in figure 4. The symmetric type interconnect, on the other hand, is conducive to a reflective orientation in which a doublet SVI is used twice, in conjunction with a mirror, to achieve the point to point mapping (figure 3). A symmetric interconnect can be used in a similar fashion to the asymmetric arrangement, although it is not as compact and it may complicate its use unnecessarily.

A demonstration of the replay quality of an asymmetric interconnect is shown in figures 6 and 8. The interconnect used was the first stage of a 2D Banyan network. To illustrate its operation the square input image was split diagonally into two regions of 100 and 50 percent relative intensities (figure 6 a).). The interconnect has the effect of swapping diagonally opposed quadrants of the image, as shown in figure 6 b).

The experimental arrangement to demonstrate the interconnect working is shown in figure 5. At this stage the doublets were not cemented together and the first element was merely a planar grating which coupled collimated input beams into the redirecting element. A quarter wave plate was used after the SVI to eliminate stray reflections due to the elements being air spaced and not being anti-reflection coated. The lens on the output side of the beam splitter was used to image the interconnect element in the output plane.

The actual interconnect chosen for this proof of principle experiment was designed for use at 514nm (because of the ready availability of polarising optics) and had a facet spacing of 200μ . The total efficiency of the doublet interconnect, for two passes through it (an input and an output pass) was measured to be 60%, which once reflections have been eliminated would be about 80%. Figure 8 shows the output from the experiment described, the quality of the interconnect would be improved if used with the full imaging geometry discussed in the next section.

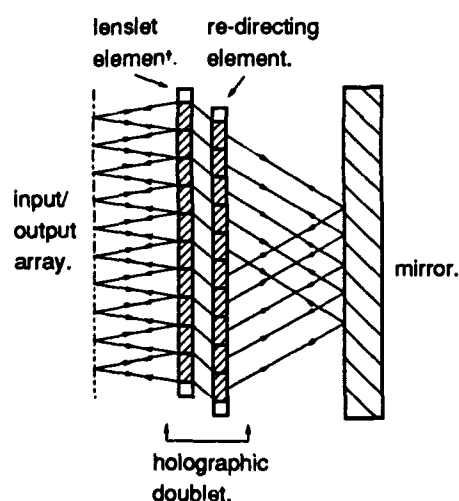


Figure 3. Banyan (symmetric interconnect) implementation.

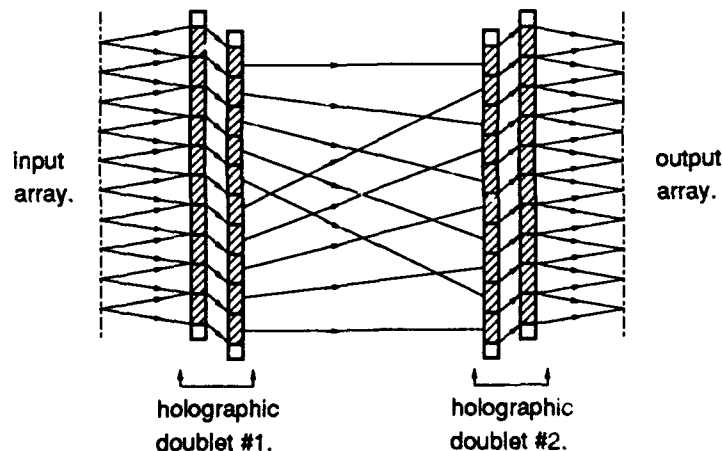


Figure 4. Perfect Shuffle (asymmetric interconnect) implementation.

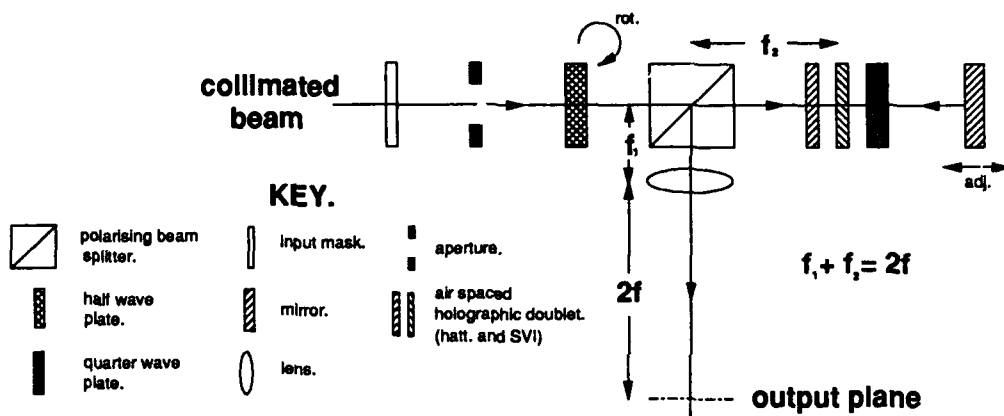


Figure 5. Experimental arrangement to test an on-axis SVI (no focussing power).

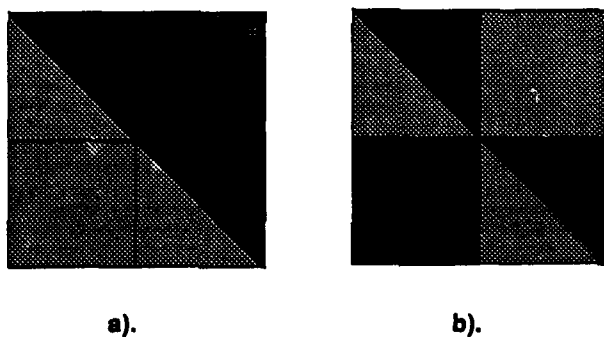


Figure 6. 2D Banyan a). Input; b). Output images.

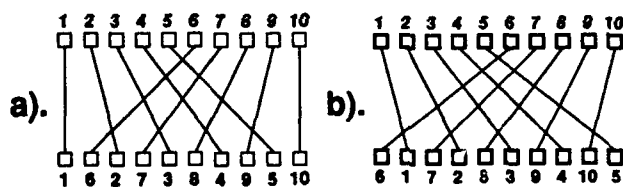


Figure 7. Schematic a). Perfect shuffle; b). Stacked deck interconnect.

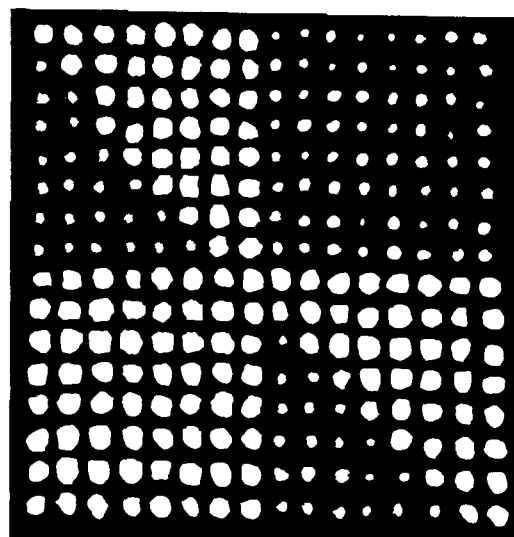


Figure 8. Experimental result.

Networks of Interconnects.

The use of space-variant networks (SVNETs) in optical communication systems has been analysed intensively^{7,8,9,10}. Several optical circuits have been demonstrated in which SVIs could play an important role.

In any SVNET, logic planes are required to control the flow of information through the network and the most notable devices used to date which are being utilised in optical circuitry are S-SEEDs¹¹ and NLIFs¹². Figure 9 shows a modular optical arrangement for the use of symmetric SVIs in conjunction with S-SEED or NLIF devices. These modular units can be linked together to form the stages of a 2D logarithmic optical network with full interconnection capability.

The perfect shuffle is an asymmetric interconnect which replicates itself at each stage of a logarithmic network. Consequently, separate modules are not strictly necessary at each stage of interconnection. The output of one stage can be routed back as the input to another, by use of some controlled feedback, to form a perfect shuffling machine (PSM). The implementation of the interconnection section of a PSM using SVIs already described is shown in figure 10.

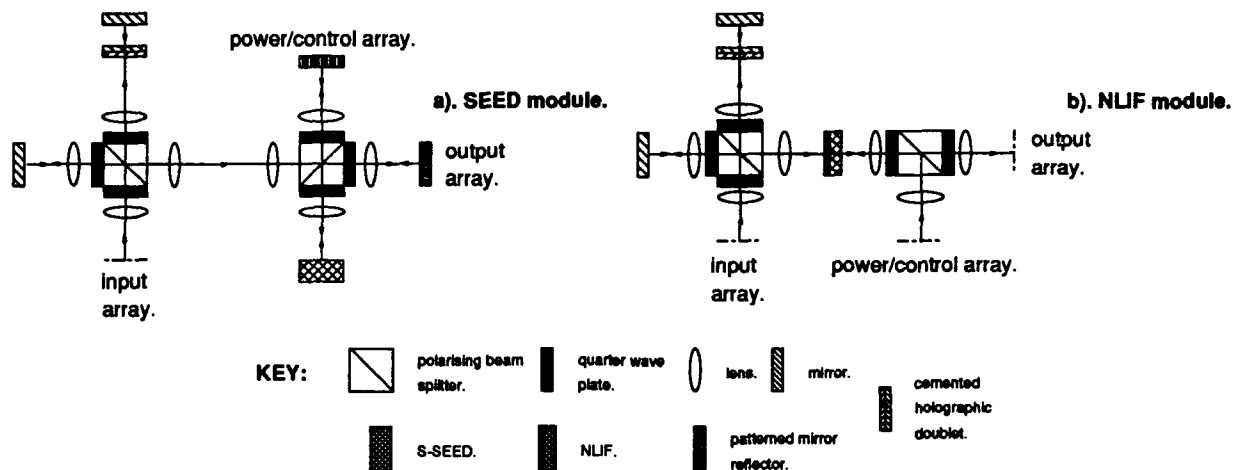


Figure 9. Modular optics for S-SEED and NLIF based communications networks.

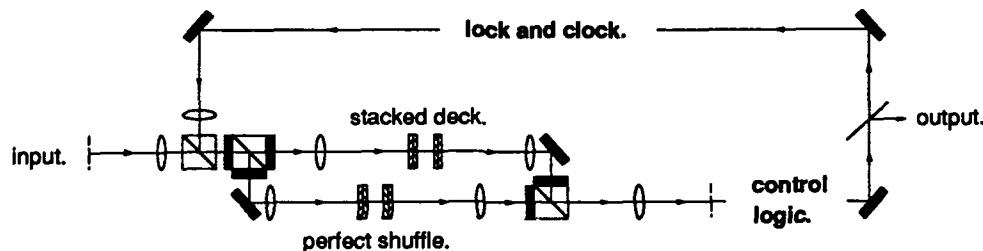


Figure 10. Realisation of a perfect shuffling machine using DCG SVIs.

Conclusions.

SVIs of the type described form a compact and efficient way of implementing optical networks and they are compatible with current demonstration circuitry. Scalability is the primary restriction of such a method of producing these SVIs. If a large array is required to be interconnected, the interconnection angles will become large, and the size of the angle that this method is capable of recording is then limited by the finite aperture of any lens being stepped in the recording process. We anticipate solving this problem by moving to a fibre/aperture based recording system, instead of relying purely on translatable imaging optics.

- 1 B. Robertson et al., Proc. Optics in Complex Systems, Garmisch-Partenkirchen, Germany, SPIE Conf. Vol. **1319**, 128-129 (1990)
- 2 C.W. Stirk et al., Appl. Opt. **27**, 202-203 (1988)
- 3 A.W. Lohmann, Appl. Opt. **25**, 1543-1549 (May 1986)
- 4 B.S. Wherrett et al., Digital Optical Computing II, Proc. SPIE Vol. **1215**, 264-273 (1990)
- 5 M.E. Prise et al., Appl. Opt. **29**, 2164-2170 (1990)
- 6 I.R. Redmond et al., Proc. IEE, Conf. Vol. **311**, 190-194 (1989)
- 7 H.S. Hinton, Digital Optical Computing II, Proc. SPIE Vol. **1215**, 132-142 (1990)
- 8 M.J. Murdocca, PhD Thesis, Rutgers University, New Jersey (1989)
- 9 J.E. Midwinter, Proc. IEE, Vol **134**, Part J, 261-268 (1987)
- 10 F.B. McCormick et al., Proc. CLEO, CPDP1 (1990)
- 11 F.B. McCormick et al., Appl. Opt. **29**, 2013-2018 (1990)
- 12 S.D. Smith et al., Nature **325**, 27-31 (1987)

Monday, March 4, 1991

Poster Preview

ME 4:40pm–5:40pm
Salon D

Lee Giles, *Presider*
NEC Research Institute



Some practical issues in design and fabrication
of high-contrast quantum-well modulator arrays

G. Parry, M. Whitehead,^{*} E. Zouganeli, A. Rivers,
K. Woodbridge, J. S. Roberts,[†]
University College, London, UK

^{*}University of California, Santa Barbara, CA 93106

[†]University of Sheffield, UK

Asymmetric Fabry–Perot modulators offer the prospect of high contrast (>20 dB) and low voltage (<5 V) as well as useful optical bandwidths. This paper will discuss the practical problems of designing and fabricating arrays of devices to these specifications.

Design and fabrication of VLSI ferroelectric liquid crystal spatial light modulators

David A. Jared
Richard Turner
Kristina M. Johnson

University of Colorado
Department of Electrical and Computer Engineering
Optoelectronic Computing Systems Center
Boulder, CO 80309-0525

1. Introduction.

This paper discusses several design and fabrication issues surrounding VLSI, ferroelectric liquid crystal (FLC) spatial light modulators (SLMs). These SLMs consist of a VLSI CMOS backplane and FLC modulators as shown in Fig. (1). The FLC is sandwiched between the CMOS backplane and a sheet of glass coated with a transparent conductor. The design and fabrication issues that are described include: FLC material selection, alignment of the FLC, installation of the glass cover, and the design of photodetectors, amplifiers, and pad drivers. An electrically addressed dynamic RAM SLM with 64×64 pixels and three optically addressed SLMs with 32×32 pixels are described to discuss these issues.

2. Device Design and Fabrication.

Electrically Addressed SLMs. Several electrically addressed SLMs (EASLMs) using this technology have been developed. The first electrically addressed device was built in 1985 by Underwood at the University of Edinburgh [1]. This device consist of a 16×16 array with $200 \mu\text{m}$ square pixels using an nMOS silicon backplane, static RAM addressing, and a guest-host nematic liquid crystal.

A 1×128 linear array with pixels on $20 \mu\text{m}$ centers using shift register addressing, and a 64×64 two-dimensional array with pixels on $60 \mu\text{m}$ centers using static RAM addressing have been demonstrated by Displaytech and Drabik [2,3]. The static RAM device has achieved a frame rate of 4.5 kHz [2,3]. A similar 50×50 two-dimensional SLM using a nematic liquid crystal atop an nMOS static RAM has been demonstrated by McKnight at the University of Edinburgh [4].

The device developed at the University of Colorado consists of a 64×64 two-dimensional SLM using an FLC modulator and a dynamic RAM[5]. The pixels are located on $40 \mu\text{m}$ centers. The layout of the chip is shown in Fig. (2). The rows are sequentially addressed using a dynamic shift register that can be seen along the right side of the array. The pixel data is loaded along the columns using 16 parallel lines. A 2-bit multiplexer is used to address all 64 columns. Each pixel in the array consists of a CMOS pass gate and a CMOS inverter with the appropriate row and column select lines as shown in Fig. (3). The device was fabricated using the $2 \mu\text{m}$, n-well CMOS process provided by MOSIS. A smectic C* FLC material, SCE13 from British Drug House, is used in the EASLM. A photo of the EASLM displaying an image in shown in Fig. (4).

Optically Addressed SLMs. The optically addressed SLMs consist of a CMOS backplane containing photodetectors, analog/digital processors, and metal pads to modulate a liquid crystal material. The photodetector and liquid crystal serve as the input and output for the processors on the backplane. By designing different analog/digital processors on the CMOS backplane, different types of specialized OASLMs can be fabricated that will perform specific computations on two-dimensional intensity data. Since the input and output is optical, the optoelectronic computing architecture can be modularized by cascading the output of one SLM to the input of another SLM. Thus, complex computational systems can be constructed by combining SLMs that perform different functions. The ability to design modular optoelectronic computing systems is completely new.

The OASLMs designed at the University of Colorado consist of a 32×32 array of phototransistors, amplifiers, and modulating pads^[6]. Three different optically addressed SLMs were fabricated: two variable thresholding SLMs and a logarithmic SLM. The schematic of a pixel for each of these SLMs is shown in Fig. (5). The pixels are on $100 \mu\text{m} \times 50 \mu\text{m}$ centers. The backplane was fabricated using the $2 \mu\text{m}$, n-well, low-noise analog CMOS process provided by MOSIS. Two types of FLC materials were used in the OASLM: SCE13, and a distorted helix FLC, 6304 from Hoffman-LaRoche. A parasitic bipolar transistor with a floating base was used as a photodetector. Several types of photodetectors can be used as depicted in Fig. (6). Figure (7) shows the output of the OASLM when the device is shadowed. The FLC is thresholded to different states on either side of the shadow.

Installation of the glass cover and filling. The glass cover was $6 \text{ mm} \times 7 \text{ mm} \times 4 \text{ mm}$ cut from optical flats. Indium-tin-oxide was evaporated on one-side of the glass to form the transparent electrode. A layer of chromium was evaporated along one edge of the glass marginally overlapping onto the indium-tin-oxide. This allowed for a wire to be easily attached to the glass cover using conductive epoxy. The glass was placed on the chip using a specially designed jig. The jig consists of an X-Y translation stage for positioning the chip, and three micrometers for lowering the glass cover onto the chip. The thickness of the air gap between the chip and the glass surface was measured using a Pohl interferrometer. White light fringes and capacitance measurements were also used to measure the thickness of the air gap. Once the desired thickness was achieved, the glass cover was glued to the chip and removed from the jig. The chips were filled with the FLC in a vacuum.

FLC Selection and Alignment. Several different alignments were tested with FLC materials SCE13 and CS1014 in order to determine the best match which optimized device speed and contrast ratio. In particular, we tried single-side alignment using nylon 66, polyvinyl alcohol (PVA), poly-1,4-butylene terephthalate (PBT), and oblique evaporation of SiO_2 . Our experiments show that SCE13 with the single-sided PVA alignment layer produced the best contrast ratio (800:1 in an FLC cell made with two glass plates) and switching speed with $\pm 2.5 \text{ V}$ applied (500 μsec turn-on and 400 μsec turn-off times).

Acknowledgements. Discussions with Tim Slagle and Kelvin Wagner were gratefully appreciated in developing this project. David Doroski and Chongchang Mao helped substantially in filling the chips with the FLCs. This work was supported by the NFS ERC for OCS, CDR862228. Graduate Fellowship support by NASA-JSC for David Jared is gratefully acknowledged.

References

- [1] I. Underwood, D. G. Vass, and R. M. Sillitto. Evaluation of an nMOS VLSI array for an adaptive liquid-crystal spatial light modulator. *IEE Proceedings*, 133(1):77-82, 1986.
- [2] L. K. Cotter, T. J. Drabik, R. J. Dillon, and M. A. Handschy. Ferroelectric-liquid-crystal/silicon-integrated-circuit spatial light modulator. *Opt. Lett.*, 15(5):291-293, 1990.
- [3] T. J. Drabik, L. K. Cotter, and M. A. Handschy. Ferroelectric liquid crystal/silicon VLSI spatial light modulator. *OSA Tech. Dig. Series*, 18:183, 1989.
- [4] D. J. McKnight, D. G. Vass, and R. M. Sillitto. Development of a spatial light modulator: a randomly addressed liquid-crystal-over-nMOS array. *Appl. Opt.*, 28(22):4757-4762, 1989.
- [5] R. Turner, D. Jared, K. Johnson, B. Crossland, M. Birch, D. Vass, and I. Underwood. High-speed compact correlator using VLSI liquid-crystal spatial light modulators. in Technical Digest of the Annual Meeting of the Optical Society of America, Boston, 1990, paper FT4.
- [6] D. Jared, T. Slagle, K. Johnson, and K. Wagner. Optically addressed CMOS VLSI liquid-crystal spatial light modulators. in Technical Digest of the Annual Meeting of the Optical Society of America, Boston, 1990, paper FV2.

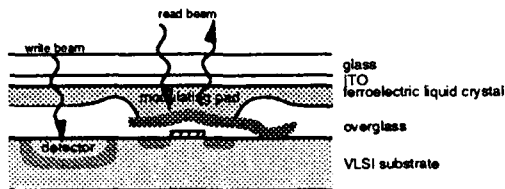


Figure 1: Schematic of a VLSI/FLC SLM. The devices consist of FLC layer sandwiched between a VLSI chip and a piece of glass coated with a transparent electrode. The semiconductor backplane can contain combinations of photodetectors, analog/digital electronics, and metal pads to modulate the FLC.

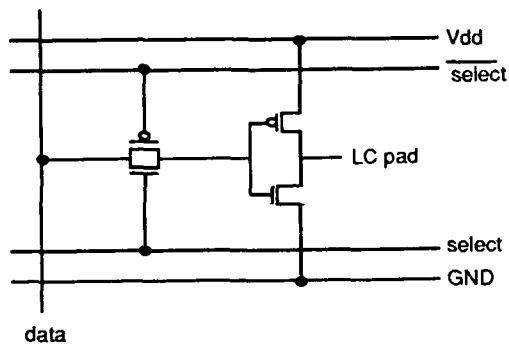


Figure 3: Schematic of the DRAM pixel.

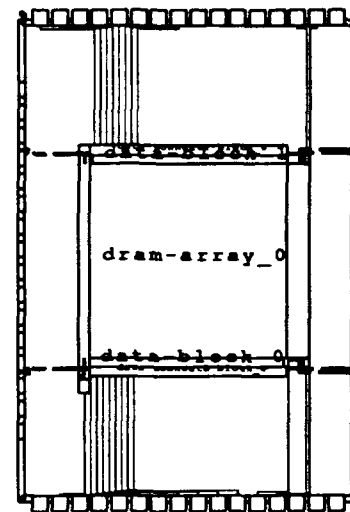


Figure 2: Layout of the 64 x 64 DRAM SLM.

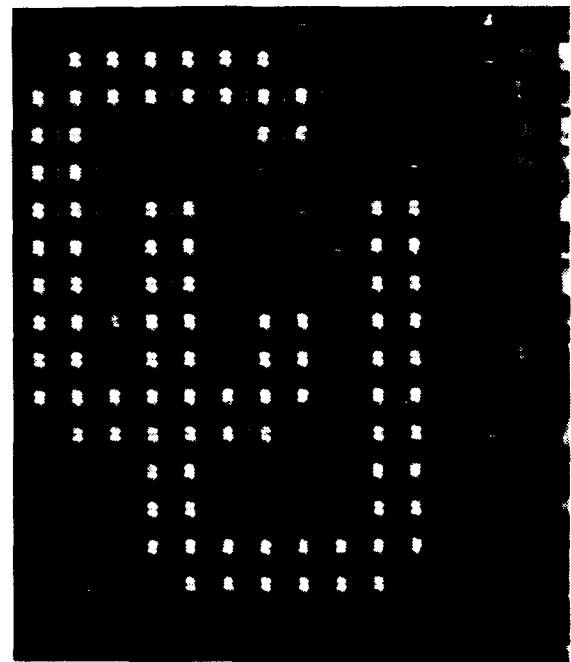


Figure 4: Photo of the DRAM SLM displaying an image.

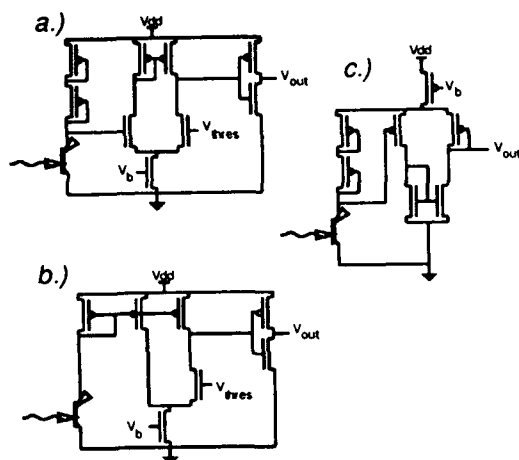


Figure 5: Schematics of the optically addressed SLMs: a.) thresholding SLM #1, b.) thresholding SLM #2, and c.) logarithmic SLM.

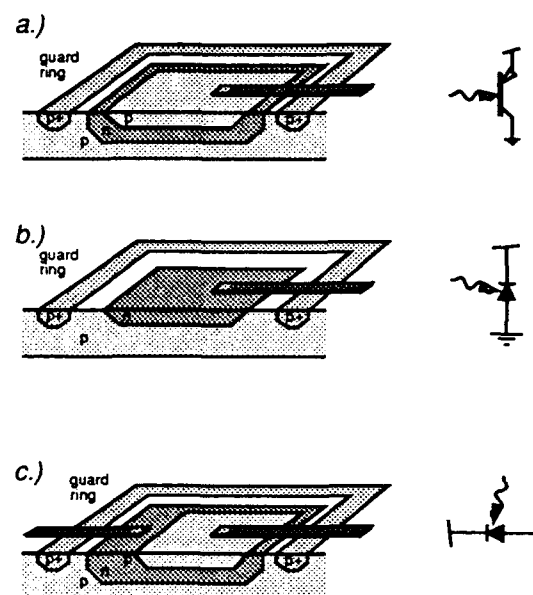


Figure 6: Types of photodetectors that can be easily fabricated using a CMOS process. a.) a PNP phototransistor can be formed by placing an p-diffusion region in an n-well. b.) a PN photodiode can be form by simply placing an n-well or n-diffusion region in the substrate. c.) an isolated PN photodiode can be formed by placing a well contact on detector a.). A substrate contact surrounds the photodetectors to isolate them from other devices on the chip.

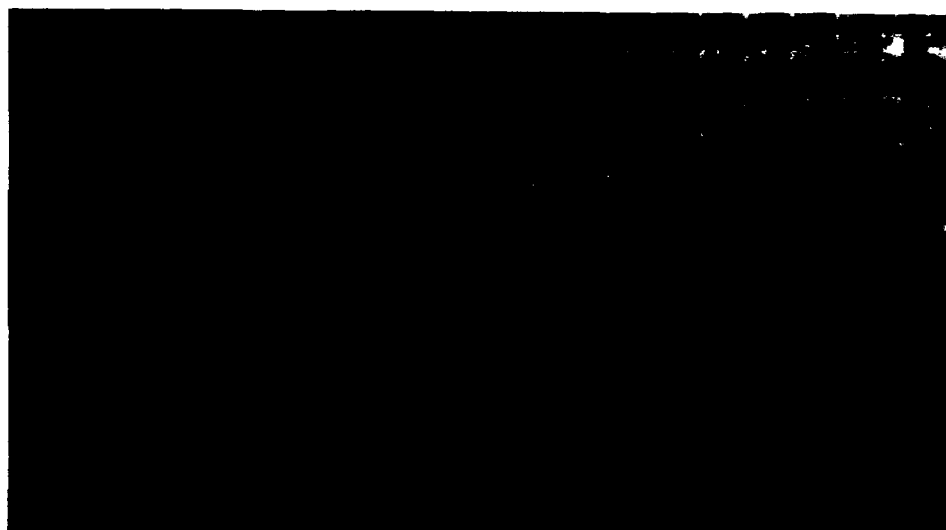


Figure 7: Photo of the response of the thresholding OASLM when an input intensity difference is incident on the device.

Monday, March 4, 1991

Spatial Light Modulators

MD 3:50pm-4:40pm
Salon F

Uzi Efron, *Presider*
Hughes Research Laboratories



Huge optical amplification by applying pulsed electric fields to photorefractive crystals.

P. MATHEY, G. PAULIAT, J.C. LAUNAY* and G. ROOSEN
*Institut d'Optique Théorique et Appliquée, Unité de Recherche Associée au
 Centre National de la Recherche Scientifique,
 Centre Scientifique d'Orsay, Bât. 503, B.P. 147, 91403 ORSAY Cedex, France*

* *Laboratoire de Chimie du Centre National de la Recherche Scientifique,
 351 Cours de la Libération, 33405 Talence Cedex, France*

A large number of proposed or demonstrated architectures for optical computing takes advantage of the unique properties of photorefractive crystals.

Efficient operation requires large optical nonlinearities. BaTiO₃ exhibits these strong refractive index changes due to its large electro-optic coefficient. Unfortunately it has a low sensitivity and thus a slow response time. BSO is much more sensitive but its electro-optic coefficient is smaller. It is consequently necessary to induce large internal electric fields. Different methods were already proposed and demonstrated to enhance the space charge field in photorefractive crystals:

- the "moving grating" technique [1] in which the interference pattern of the two incoming beams is moved in presence of a D.C. applied electric field;
- the "alternating field" technique [2] in which an alternating electric field is applied to the crystal without moving the interference pattern;
- the "resonant intensity" technique [3] in which the optical irradiance is adjusted to equalize the emission rates for electrons and holes.

Unfortunately, with all these methods, the maximum two beam coupling gain is limited by crystal parameters. The gain is indeed always proportional to the ratio E_1/m where E_1 is the photo-induced space charge electric field and m is the modulation ratio of the interference pattern. E_1 is at most equal to E_Q that is the maximum space charge field which could be induced in a photorefractive material if all the charges were completely redistributed in the trap sites:

$$E_Q = e \Lambda N_A / (2\pi\epsilon) \quad (1)$$

with e the absolute value of the charge of electron, Λ the fringe spacing of the induced grating, N_A the density of trap sites and ϵ the static dielectric constant. Therefore, we could think that for a low enough modulation ratio m , the two wave mixing gain could be as large as wanted. This assumption is wrong for all the above mentioned techniques because the maximum field E_1 we can induce is proportional to the product m times a function depending on material and experimental parameters. Therefore, the maximum field we can get with these three techniques is about $m.E_Q$. It is the same as what can be obtained with a very large applied D.C. field only [4]. The gain is thus limited by the density of trap sites N_A although only a small part of them, $m.N_A$, is used.

Here we propose a new method to overcome this limitation. In a previous paper [5] we demonstrated that applying a sinusoidal electric field of a given frequency to photorefractive samples leads to a resonance of the two wave mixing gain. This enhancement can be explained as follows. When the applied field is close to zero, a charge-carrier grating is excited. It is then shifted by a length d during half a period of the sinusoidal field. If the drift length d is equal to half a fringe spacing, then the charges photoexcited from the bright fringes of the interference pattern mainly recombine in the dark ones. The two wave mixing gain is thus enlarged. From the analysis conducted in this previous paper we concluded that the resonance appears if the following conditions are fulfilled. First the period T of the applied field must be so that the travel length d during $T/2$ is about half the fringe spacing. Second, the charge-carrier grating must not relax ($T < \tau_d$; the dielectric relaxation time constant) nor recombine ($T < \tau_R$ the recombination time of the charges) and nor diffuse ($T < \tau_D$ the diffusion time constant) before the travel length is reached.

In this conference we will demonstrate how to optimize the temporal shape of the applied field and get huge amplification factors.

We will first discuss the theoretical approach. It is based on the band transport model and accounts for the modulated density of charge-carriers in the expression of the space charge field. Considering that only one kind of charge-carriers is involved in the photorefractive effect and an applied electric field E_0 that is time independent, then the kinetics of the induced space charge field is governed by this second time derivative equation:

$$\frac{\partial^2 E_1}{\partial t^2} + \left(\frac{1}{\tau_a} + \frac{1}{\tau_b} \right) \frac{\partial E_1}{\partial t} + \frac{1}{\tau_a \tau_b} (E_1 - E_{sc}) = 0 \quad (2)$$

where E_{sc} is the steady state space charge field and τ_a and τ_b are two complex constants whose expressions are given in Ref. [6]. Here we use the expressions for τ_a , τ_b and E_{sc} in the case of low optical irradiance [5].

We must remark that Eq. (2) is valid for a time independent applied field E_0 only. If the field is time dependent, the expressions for τ_a and τ_b depend on the time derivation $\partial E_0 / \partial t$ which can therefore not be neglected. However because the solution of Eq. (2) is very simple for time independent fields, we restricts our analysis to the cases where we can decompose the time in time intervals T_i during which the applied field is constant. We study the temporal shape of the applied field depicted in Fig. 1. During T_0 the applied field is zero and during T_1 it is equal to $\pm \bar{E}_0$ so that Eq. (2) can be used to predict the kinetics of E_1 during T_0 and T_1 .

Numerical computations were conducted for $\text{Bi}_{12}\text{GeO}_{20}$ crystals using typical material parameters and with $\tau_{di} = 5$ ms. In fig. 2 we plotted the two wave mixing gain G versus the fringe spacing of the induced grating. G is defined by:

$$G = \frac{\pi}{\lambda} n^3 r_{41} \frac{E_1}{m} \quad (3)$$

with n the refractive index for the optical wavelength λ , and r_{41} the electro-optic coefficient.

The straight mixed line represents the maximum gain obtainable with conventional enhancement techniques and with a very large applied electric field ($E_0 \gg E_q$). The corresponding space charge field is $E_1 = mE_q$. The curve in full line represents the gain calculated using a symmetric pulsed field with $\bar{E}_0 = \pm 7.5 \text{ kV/cm}$, $T_0 = 1.051 \mu\text{s}$ and $T_1 = 50 \mu\text{s}$.

At the resonance peak (around $\Lambda_r = 55 \mu\text{m}$) the gain rises much above the limitation given by $E_1 = mE_q$. For comparison, in dashed line is plotted the gain using either the usual "square wave AC field" or moving grating enhancement techniques for $\bar{E}_0 = \pm 7.5 \text{ kV/cm}$. Although the peak applied field (7.5 kV/cm) is the same than previously, the maximum gain is smaller. Furthermore, it is clearly visible that for low fringe spacings, the maximum gain with the "AC field" technique is limited by the trap density.

We will then describe the experimental technique developed for achieving extremely high amplification gains.

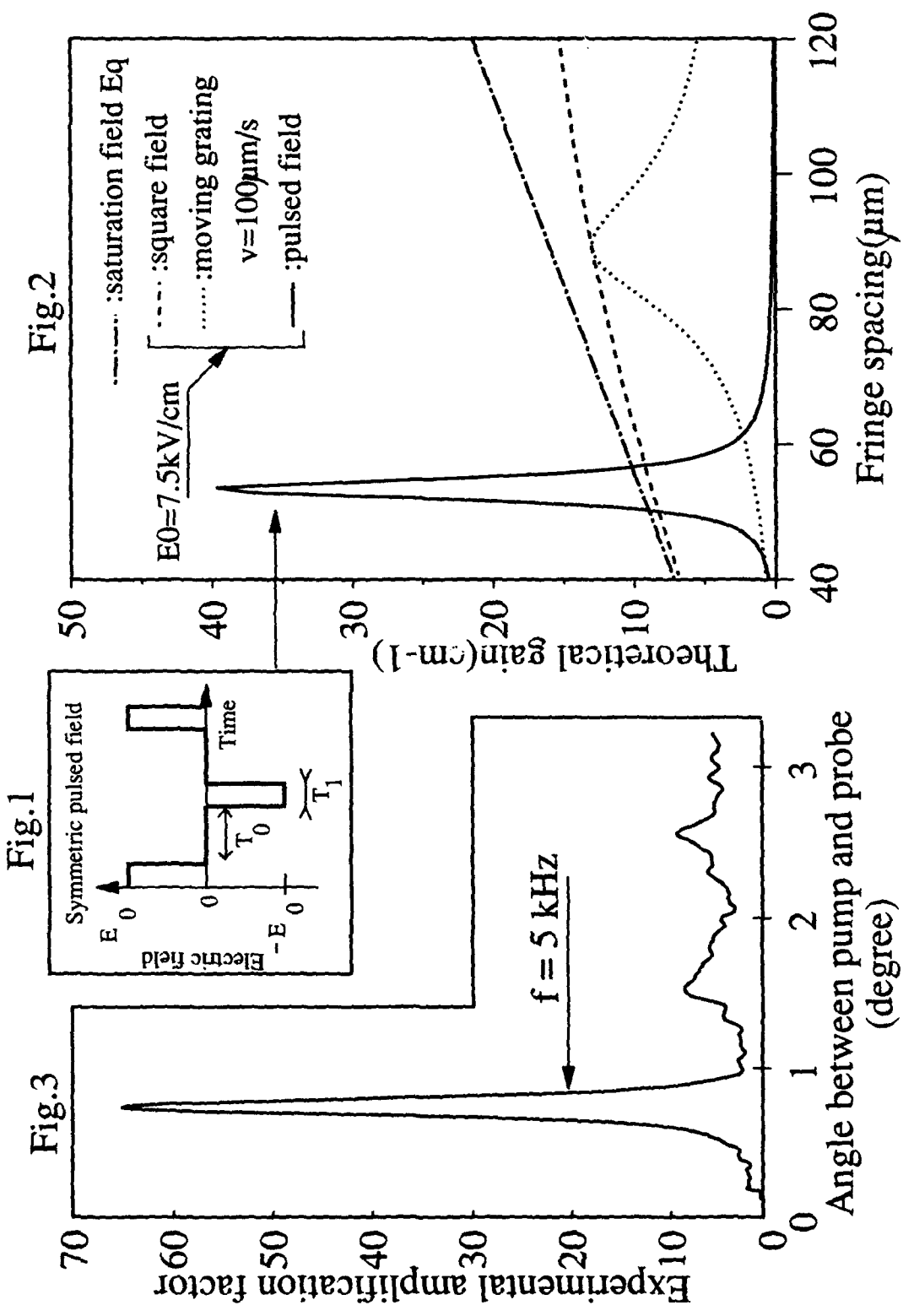
Because of the relatively slow rise time of the power supply, the temporal shape of the applied field was not the one depicted in Fig. 1. Thus, for sake of simplicity we preferred to work with a sinusoidal applied electric field which is a crude approximation for a pulsed field.

The full line plotted in Fig. 3 represents the amplification factor obtained with a 5 kHz sinusoidal applied field. The peak field is $\bar{E}_0 = \pm 7.5 \text{ kV/cm}$. The maximum gains obtained with a sinusoidal field are weaker than what must be expected using a pulsed field. Nevertheless, within the constraints imposed by our power supply, by modifying the shape of the applied field we reached an amplification factor of 10^3 corresponding to a gain up to 10 cm^{-1} (with the same peak applied field of 7.5 kV/cm). This value for the amplification factor must be compared to the amplification of about 5 we obtained with the same sample using the usual "AC field" enhancement technique with a square wave ($\bar{E}_0 = \pm 7.5 \text{ kV/cm}$).

Ultimate performances and limitations will be discussed at the conference.

References:

- [1] J.P. Huignard and A. Marrakchi,
Opt. Comm. **38** (1981) 249.
- [2] S.I. Stepanov, V.V. Kulikov and M.P. Petrov
Opt. Comm. **44** (1981) 19.
- [3] G. Picoli, P. Gravey, C. Ozkul and V. Vieux
J. Appl. Phys. **66** (1989) 3789.
- [4] Duncan T.H. Liu, Li-Jen Cheng, Arthur E. Chiou and Pochi Yeh
Opt. Comm. **72** (1989) 384.
- [5] G. Pauliat, A. Villing, J.C. Launay and G. Roosen
J. Opt. Soc. Am. **B7** (1990) 1481.
- [6] G.C. Valley
IEEE J. Quantum Electron. **QE - 19** (1983) 1637.



Enhanced photorefractive effects with a dc field and moving grating in GaP at 633 nm

Jian Ma, Yoshinao Taketomi, Yeshaiahu Fainman, Joseph E. Ford and Sing H. Lee
University of California, San Diego,
Dept. of Electrical and Computer Engineering, La Jolla, CA 92093

Ken'ichi Chino
Sumitomo Metal & Mining Co., Ltd., Japan

Photorefractive (PR) devices have found applications in optical computing, image processing and pattern recognition^[1-3], because PR materials provide unique features such as real time operation, optical gain, storage, nonlinear operations, phase conjugation and correlation. New PR materials are being investigated in order to meet the device and system requirements of sensitivity, speed, and operation wavelength (e.g., response to the near infrared spectral range for systems operated with semiconductor lasers). Compound semiconductors may satisfy these requirements. For example, optical signal amplification by two-beam coupling and amplified phase-conjugate beam reflection by four-wave mixing have been reported in GaAs^[4] and InP^[5] at the wavelength of 1.06 μm . Recently, GaP^[6-7] was shown to possess a relatively weak PR effect in the spectral range of 0.6 to 0.9 μm . In this manuscript we report enhancement of the PR effect in GaP using an externally applied electric field and moving grating. In particular, two- and four-wave mixing experiments were used to demonstrate a gain coefficient of $\Gamma = 1.9 \text{ cm}^{-1}$ and a phase conjugate reflectivity, $R = 4.5\%$. In addition, several figures of merit of GaP, i.e., steady-state index change, absorption coefficient, response time and PR sensitivity were characterized.

First, we examine numerically the effects of the external field and moving grating on the space-charge field in GaP. Let two beams with intensity I_1 and I_2 interfere with a modulation $m = 2\sqrt{I_1 I_2} / (I_1 + I_2)$ in the volume of the crystal with an applied field E_0 . Beam 1 is frequency shifted to cause the fringes to travel with a velocity v . From Kukhtarev's equations, the imaginary part of the equilibrium space charge field which contributes to energy coupling is given by

$$\text{Im}(E_{sc}) = m \frac{E_T + E_T^2/E_q + E_0^2/E_q - vk_g \tau_d E_0}{[vk_g \tau_d (1 + E_T/E_M) - E_0/E_q]^2 + [1 + E_T/E_q + vk_g \tau_d E_0/E_M]^2}, \quad (1)$$

where E_T is the diffusion field, E_q is the maximum value of the space charge field, k_g is the value of grating vector, τ_d is the dielectric relaxation time, $E_M = \gamma_R N_A / \mu k_g$, γ_R is the recombination rate coefficient, μ is the carrier mobility, and N_A is the acceptor density. The computer simulation of $\text{Im}(E_{sc})$ versus grating moving velocity for different fringe spacings Λ_g and an applied dc field $E_0 = 20 \text{ kV/cm}$ for GaP is shown in Fig.1. This result demonstrates significant increase in the $\text{Im}(E_{sc})$ and therefore we expect enhanced beam coupling gain coefficient and phase-conjugate reflectivities.

To verify this prediction we first measured two-beam coupling (see Fig.2). A collimated He-Ne laser was split into two ordinary polarized beams (reference and signal), with intensity beam ratio $I_1/I_2 = 10^{-2}$. A single undoped GaP grown by Sumitomo Metal & Mining Co. was cut along (001), (110) and (110) crystallographic planes. The crystal thickness (the interaction length) was 3 mm. The electric field E_0 and the grating wavevector k_g were both in the (001) direction. The motion of the fringes was achieved by reflecting the reference beam from a moving piezo-mirror driven by a saw-tooth waveform voltage. The velocity of the piezo-mirror (v_m) was detected in real-time using a Mach-Zehnder interferometer formed by BS₂, BS₃, M₁ and the piezo-mirror. The fringe velocity v can be calculated from

$$v = 2 \frac{\Lambda_g}{\lambda_0} v_m \cos 45^\circ. \quad (2)$$

This measurement is more accurate and more convenient than the indirect method based on equation $\delta\omega = f(\chi, U_m, t, \theta_0)$ ^[8], since it includes the nonlinearities of the piezo mirror response. The experimental results for the two-beam coupling exponential gain coefficient Γ as a function of moving fringe velocity for an applied electric field of $E_0 = 20$ kV/cm and different fringe spacings Λ_g , are shown in Fig.3. Exponential gain coefficient values as high as $\Gamma = 1.9$ cm⁻¹ were reached. This value is 5.8 times larger than that previously reported [6].

Two-wave mixing is strongest when the imaginary component of E_{sc} is maximized. In contrast, the phase-conjugate reflectivity for four-wave mixing is optimized by maximizing the magnitude of the space charge field, $|E_{sc}|$. $|E_{sc}|$ may in turn be increased under dc fields and a stationary grating. However, when the fringes are moving at the optimum velocity $|E_{sc}|$ can be further enhanced. The experimental arrangement for four-wave mixing differs somewhat from the set-up for two-beam mixing (see Fig.2). Beam intensity ratios were set at $I_1/I_2 = 0.5$ and $I_2/I_3 = 1$ respectively. The experimental result of phase-conjugate reflectivity, R , versus applied dc field without moving grating is shown in Fig.4. A significant increase in the conjugate beam intensity (factor 60x) was obtained. The phase-conjugate reflectivity of GaP can be further enhanced by optimizing the fringe velocity. The experimental result of R as a function of fringe velocity for a field of $E_0 = 20$ kV/cm and fringe spacing of $\Lambda_g = 4.5$ μ m are shown in Fig.5. We have characterized several figures of merit of GaP at 633 nm: steady-state index change $\Delta n_{ss} = 1.2 \times 10^{-5}$, absorption coefficient $\alpha = 1$ cm⁻¹, response time (for 1 W/cm²) $\tau = 0.8$ ms, PR sensitivity $S = 10^{-2}$ cm³/J.

In summary, we have shown that the photorefractive effects of GaP crystal can be enhanced by using a dc field and moving grating at the laser wavelength of 633 nm. A increased two-beam coupling gain of 1.9 cm⁻¹ and phase conjugate reflectivity of 4.5% were obtained. The investigation of the PR effect in GaP at 0.85 μ m is in process. We believe that this material shows great promise for a number of optical computing applications.

- [1] J. E. Ford, S. H. Lee and S. Fainman, Opt. Lett. **19**(1990), 1088.
- [2] E. G. Peak and D. Psaltis, Opt. Eng. **26**(1987), 428

- [3] S. Fainman, Proc. SPIE **1150**(1989), 120.
- [4] H. Rajenbach, B. Imbert, J. P. Huignard and S. Mallick, Opt. Lett. **14**(1989), 78.
- [5] C. Ozkul, G. Picoli, P. Gravey and N. Wolffer, Appl. Opt. **29**(1990), 2711.
- [6] K. Kuroda, Y. Okazaki, T. Shimura, H. Okamara, M. Chihara, M. Itoch and I. Ogura, Opt. Lett. **15**(1990), 1197
- [7] M. Itoch, K. Kuroda, T. Shimura and I. Ogura, Conf. Record of 1990 International Topical Mtg. on Opt. Computing, paper 10C2, p.175.
- [8] Ph. Refregier, L. Solymar, H. Rajbenbach and J.P. Huignard, J. Appl. Phys. **58**(1985), 45.

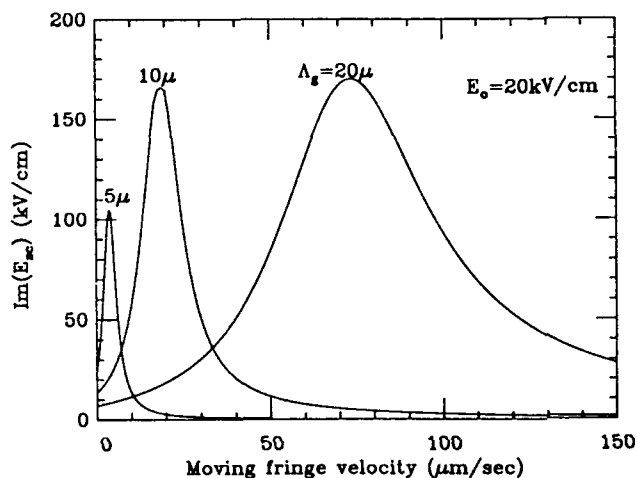


Fig.1, Computer simulation results of the imaginary part of the space charge field vs moving fringe velocity for different grating spacing.

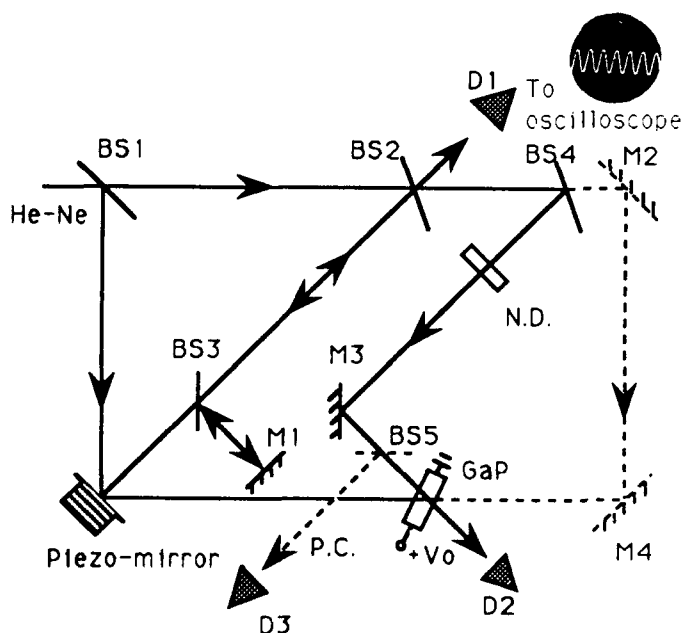


Fig.2. Experiment set-up for two- and four-wave mixing. The components represented by solid-lines are for two-wave mixing; the components represented by dotted are for four wave-mixing.
 BS1-BS5: beam splitters
 M1-M4: mirrors
 D1-D3: detectors
 N.D.: neutral density filter
 P.C.: phase-conjugate

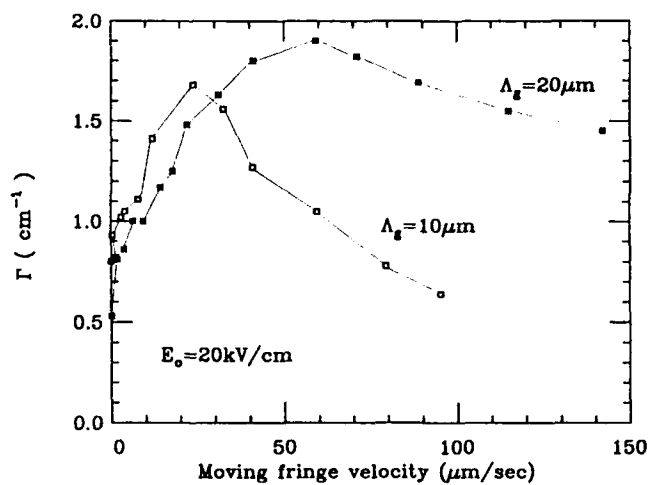


Fig.3. Experimental results of coupling coefficient vs fringe velocity.

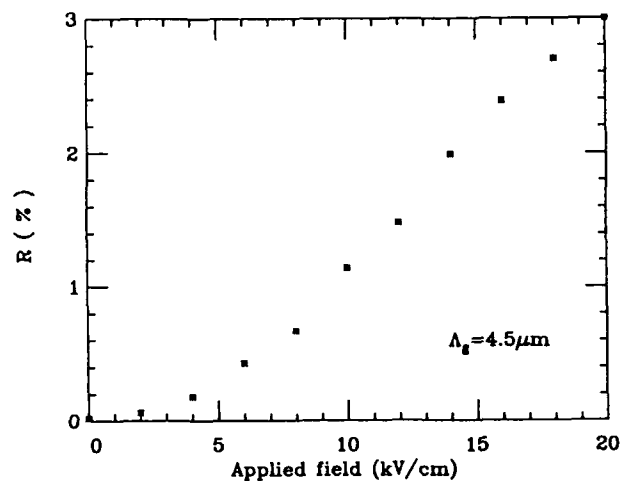


Fig.4. Experimental result of phase-conjugate reflectivity vs applied dc field with stationary grating.

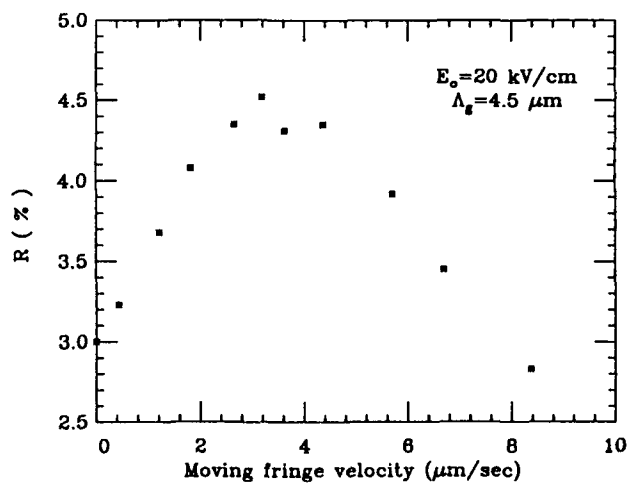


Fig.5. Experimental result of phase-conjugate reflectivity vs moving grating velocity for a fringe spacing of 4.5 μm .

Optical Thresholding and Max Operation

Claire Gu

Rockwell International Science Center
1049 Camino Dos Rios, Thousand Oaks, CA 91360
(805) 373-4180

and

Pochi Yeh*

Department of Electrical and Computer Engineering
University of California, Santa Barbara, California 93106
(805) 893-3981

SUMMARY

Thresholding and Max operations are essential elements in the implementation of neural networks. Although there have been several optical implementations of neural networks, the thresholding functions are performed electronically [1-3]. Optical thresholding and Max operations have the advantages of parallelism and cascability without resorting to opto-electronic conversion. Unfortunately, there has been very limited work in this area. In this paper, we propose and study the properties of self-oscillation in nonlinear optical (NLO) four-wave mixing (FWM) and NLO resonators for parallel optical thresholding and Max operations.

Referring to Fig. 1, consider a NLO medium pumped by two counter-propagating plane waves with amplitudes A_2 and A_3 . In the configuration of phase conjugation via FWM, a probe beam A_1 is incident on the nonlinear medium, and a phase conjugate beam A_4 is generated. Self-oscillation occurs if the two counter-propagating beams A_1 and A_4 are generated without an incident probe, i.e., $A_{10} = 0$. This corresponds to an infinite phase conjugate reflectivity at an infinitesimal probe intensity. Using coupled mode analysis, we obtain the following condition for self-oscillation

$$s + (I_{20} - I_{3L}) \tanh \frac{s\Gamma^*}{4I_0} L = 0, \quad (1)$$

with

$$s^2 = (I_{20} - I_{3L})^2 + 4 I_{20} I_{30} = (I_{20} - I_{3L})^2 + 4 I_{2L} I_{3L}, \quad (2)$$

where Γ is the complex coupling constant and $I_0 = I_1 + I_2 + I_3 + I_4$ is the total intensity. In addition, it can be shown that the oscillation frequency shift, $\Omega = \omega_1 - \omega_2 = \omega_3 - \omega_4$, is given by

$$\frac{\sigma}{\tanh\left(\frac{\pi}{2\Omega\tau}\right)} \frac{\Gamma_0 L}{4I_0} \frac{\Omega\tau}{1 + (\Omega\tau)^2} = -\frac{\pi}{2}. \quad (3)$$

where $\sigma = I_{20} - I_{3L}$, Γ_0 (assumed to be real) is the coupling constant for degenerate FWM and τ is the photorefractive time constant. We note that if $\Gamma_0 < 0$, self-oscillation can only occur when $\sigma > 0$, i.e., $I_{20} > I_{3L}$. Given I_{20} , I_{3L} and $\Gamma_0 L$, Eq. (3) can be solved numerically for $\Omega\tau$. Once $\Omega\tau$ is obtained, we can calculate the self-oscillation intensity I_{1L} which is given by

$$I_{1L} = I_{20} - \frac{\sigma^2}{4I_{3L}} \left[\frac{1}{\tanh^2\left(\frac{\pi}{2\Omega\tau}\right)} - 1 \right]. \quad (4)$$

Fig. 2 shows the intensity I_{1L} as a function of I_{20} when $I_{3L} = 1$ is fixed and $\Gamma_0 L = -15$. It is interesting to note that when I_{20} is less than some threshold value, there is no self-oscillation. When I_{20} is above the threshold value, I_{1L} increases monotonically with I_{20} .

The unique property shown in Fig. 2 can be used to implement optical thresholding. In a thresholding operation, the intensity I_{3L} is used as a reference and I_{20} is the input signal. The oscillating beam intensity I_{1L} is the thresholded output, as shown in Fig. 2. We note that the threshold intensity depends on the reference I_{3L} . Therefore, we can adjust the threshold by varying the reference intensity I_{3L} . In this thresholding operation, the signals being thresholded remain in the optical domain for further processing purposes.

We can also use a bidirectional ring resonator pumped by two counter-propagating beams to implement optical thresholding. Referring to Fig. 3, self-oscillation occurs in the ring resonator for the same reason as in ordinary FWM. Fig. 4 shows the response of such a ring resonator. With the ring resonator, the oscillating beams form a feedback loop. Therefore, it requires a smaller coupling constant $\Gamma_0 L$ to establish steady state oscillation in the ring resonator than that is needed in FWM.

Furthermore, self-oscillations can be employed to implement parallel thresholding and Max operation. Referring to Fig. 5, we consider an array of input light beams with different intensities (represented by different line types) and an array of reference beams with equal intensities. When these pairs of input-reference beams interact at different locations inside the crystal, we can do thresholding in parallel. In addition, if we adjust the intensity of the reference beam, we can identify the beam with maximum intensity. This is done by increasing (or decreasing) the intensity of the reference beam. In the case of decreasing the reference intensity, oscillation occurs when the intensity of the brightest input reaches the regime which allows self-oscillation. At this point in time, the brightest beam is selected and located. With this technique, the comparison can be done in parallel and the maximum can be found without measuring electronically the intensities of all the light beams. This approach is extremely useful when the number of input beams becomes very large.

In summary, we have proposed and studied the properties of self-oscillation in FWM and oscillations in a bidirectional ring resonator pumped by two counter-propagating beams. These properties can be used to implement optical thresholding and Max operation. Similar properties also exist in other NLO media besides photorefractive media. Our further investigation will also include the demonstration of these operations experimentally.

* Pochi Yeh is also a Principal Technical Adviser at Rockwell International Science Center.

REFERENCES

1. J. Hong, S. Campbell, P. Yeh, Appl. Opt. **29**, 3019 (1990).
2. D. Psaltis, D. Brady, and K. Wagner, Appl. Opt., **27**, 1752 (1988).
3. E. G. Paek, J. Wullert, J. S. Patel, Opt. Lett., **14**, 1303 (1989).
4. See, for example, M. Cronin-Golomb, B. Fischer, J. O. White, and A. Yariv, IEEE J. Quantum Electron. **QE-20**, 12 (1984).

FIGURES

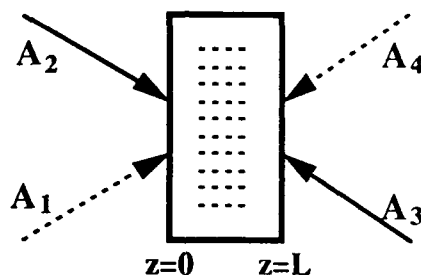


Fig. 1 FWM via transmission grating.

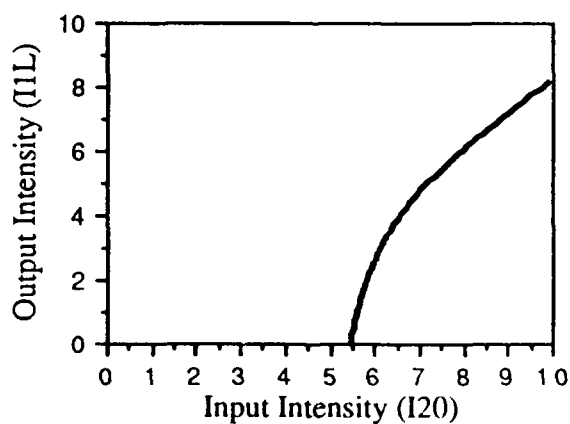


Fig. 2 Output intensity I_{1L} as a function of the input intensity I_{20} when the reference intensity is fixed as $I_{3L}=1$ and the coupling constant is $\Gamma_0 L=-15$.

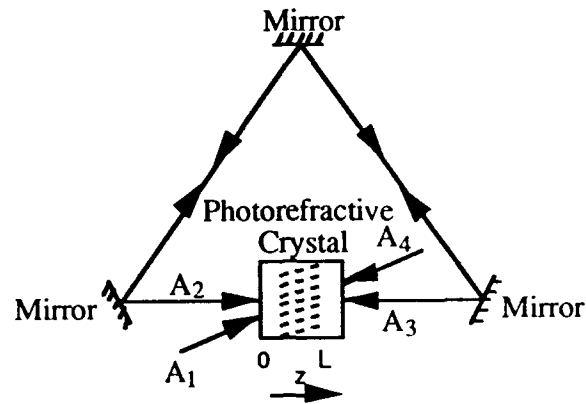


Fig. 3 Bidirectional ring resonator.

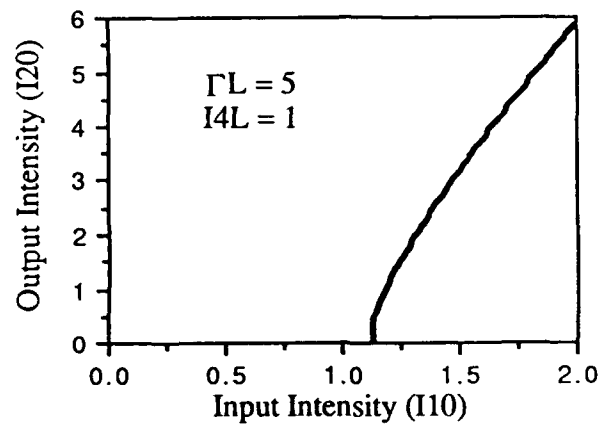
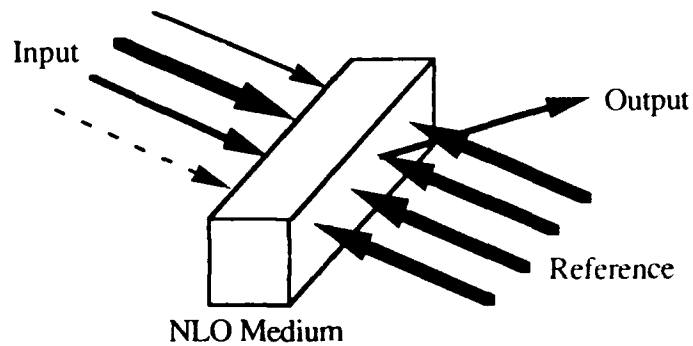
Fig. 4 Output intensity I_{20} as a function of the input intensity I_{10} when the reference intensity is fixed as $I_{4L} = 1$ and the coupling constant is $\Gamma L = 5$.

Fig. 5 Parallel thresholding and Max operation.

Gray-Scale Controllable Ferroelectric Liquid Crystal Spatial Light Modulator

Cleber M. Gomes, Susumu Tsujikawa*, Hiroki Maeda, Hiroyuki
Sekine, Takashi Yamazaki, Mikio Sakamoto*, Fujio Okumura**
and Shunsuke Kobayashi

Division of Electronic and Information Engineering, Graduate
School of Technology, Tokyo University of Agriculture and
Technology, Koganei, Tokyo 184, Japan.

* Opto-Electronics Research Laboratories, ** Functional
Devices Research Laboratories NEC Corp., Kawasaki, Kanagawa 216,
Japan.

Summary

A new optically addressed spatial light modulator (OASLM) with memorized gray-scale capability has been developed, which has potential application in analog optical computing. This OASLM exploits the occurrence of a gray-scale effect in the memory state of surface stabilized ferroelectric liquid crystal (SSFLC) cells when the SSFLC is oriented by ultra-thin polyimide Langmuir-Blodgett (LB) films.¹⁾ In the case of LB orientation films, the gray-scale can be attributed to the topography of the substrate ITO films which produces spatial fluctuation in the spontaneous polarization (multidomain).²⁾ The OASLM has many pixels, each corresponding to the smallest unit of the multidomain gray-scale. Because it is possible to add switching devices such as thin film transistors to the pixels,³⁾ the OASLM has the potential of particularly designed functions.

Fig. 1 shows the structure of two pixels of the OASLM. The device consists of an hydrogenated amorphous silicon (a-Si:H) layer 1.2 μm thick deposited onto a glass plate coated with a line patterned indium tin oxide (ITO) electrode. The lines of ITO meet at the edge of the device to allow for their simultaneous driving. A chrome (Cr) layer, 0.1 μm thick is sputtered onto the a-Si and patterned to define the pixels of the device. Each pixel of the system ITO/a-Si/Cr defines a Schottky barrier photodiode to address the corresponding portion of the SSFLC over the Cr pixel. The line patterned ITO electrode is necessary to reduce the ratio between the capacitance of the photodiode and that of the corresponding FLC pixel, C_{PD}/C_{FLC} , to allow for the utilization of the bipolar driving voltage pulse shown in Fig. 1. Five molecular layers of the Y type polyimide LB (PI-LB) film are deposited onto the photodiode system and onto the top ITO-coated glass plate. The FLC material ZLI-3654 (Merck) is injected between the substrates to complete the device.

The final device had 100X100 360 μm square pixels driven by 100 ITO lines 60 μm wide. The measured thickness of the FLC layer

was $2.3 \mu\text{m}$. The capacitance ratio C_{PD}/C_{FLC} was 0.53.

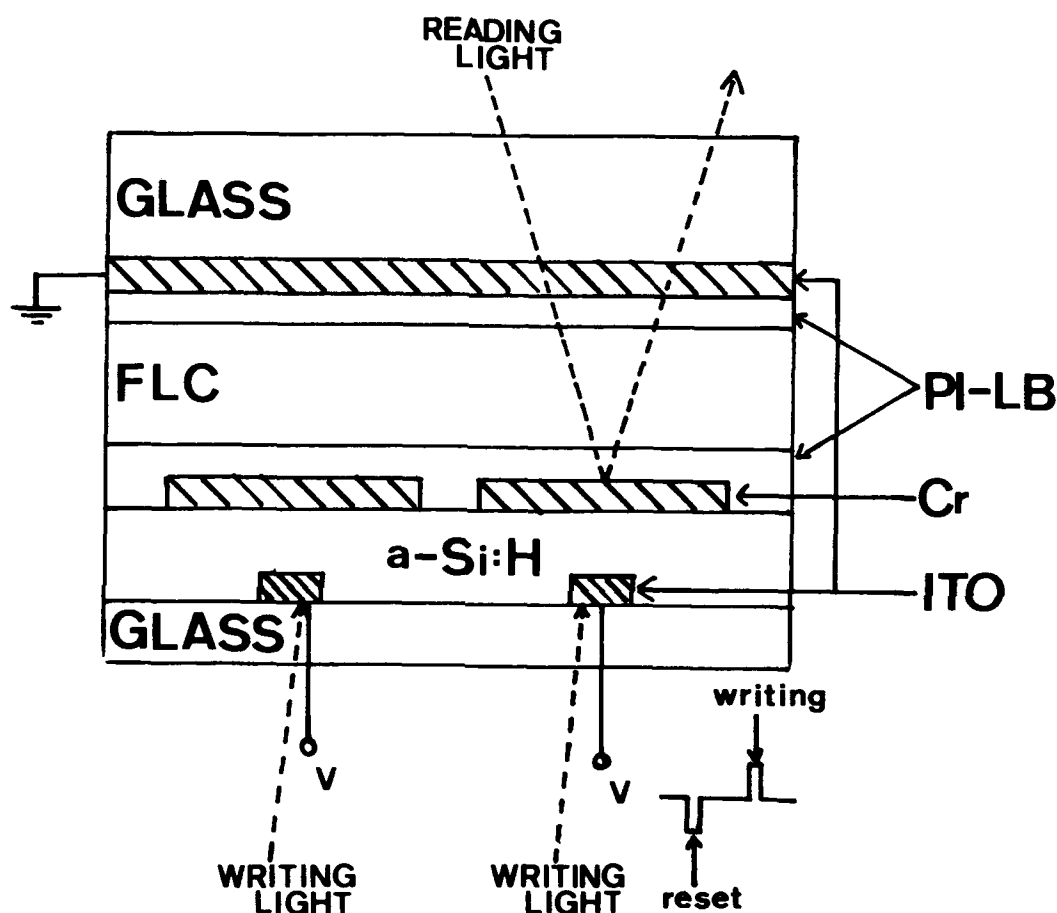


Fig.1: Cross section depicting two pixels of the fabricated SLM.

We illuminated the a-Si surface of the device with the writing light from an incandescent lamp of variable intensity. The reading light was obtained by illuminating the FLC surface with another incandescent lamp, projected through a polarizing beam splitter (PBS). The reading light reflected by one Cr pixel was detected by an avalanche photo-detector whose response is shown in Fig. 2 for three different writing light intensities. The gray-scale capability in the memory state is demonstrated.

The driving signal consisted of bipolar pulses of $300 \mu\text{s}$ of width and $\pm 16\text{V}$ of amplitude. The negative pulse acted as a reset

by forward-biasing the photodiode layer while during the positive pulse each FLC pixel was charged to a voltage proportional to the light intensity reaching the corresponding area on the a-Si:H layer.

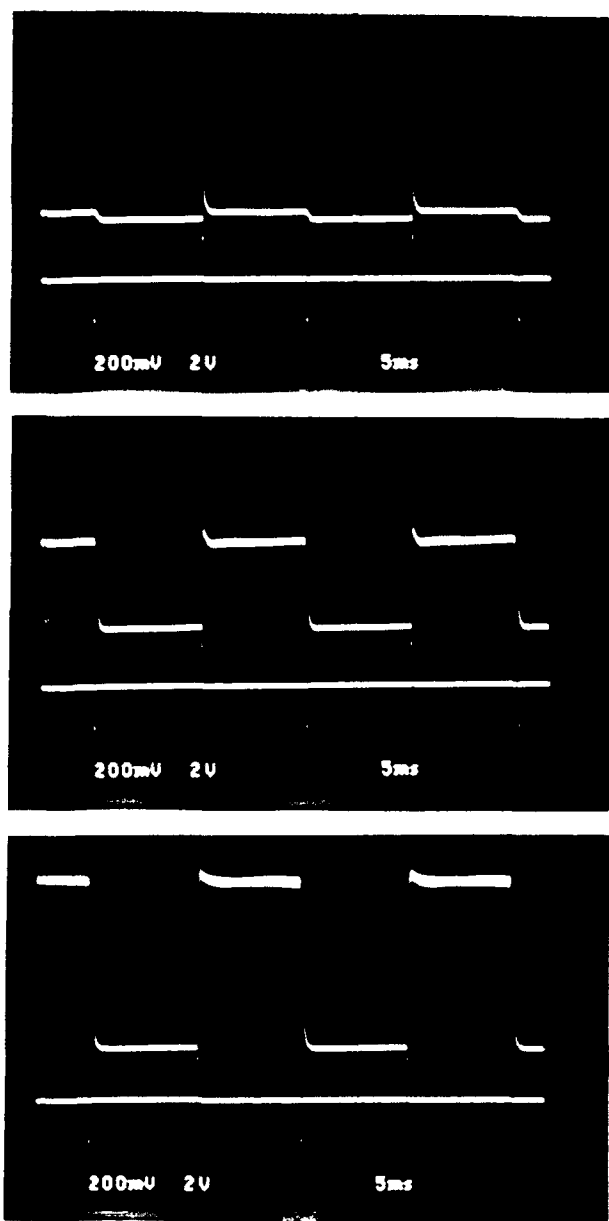


Fig.2: Optical response of one pixel at different writing light intensities: from top, no writing light, 7mW/cm^2 , and 11mW/cm^2 .

A fast rise time of $200\text{ }\mu\text{s}$ and a contrast ratio of 20:1, enough for optical computing, was obtained.

Fig. 3 shows the plot of the normalized reading light intensity, calculated from the memory voltage level of the photo-detector (level just before the next reset pulse), against the writing light intensity. The gray-scale was obtained between 0 and about 20 mW/cm^2 of writing light intensity.

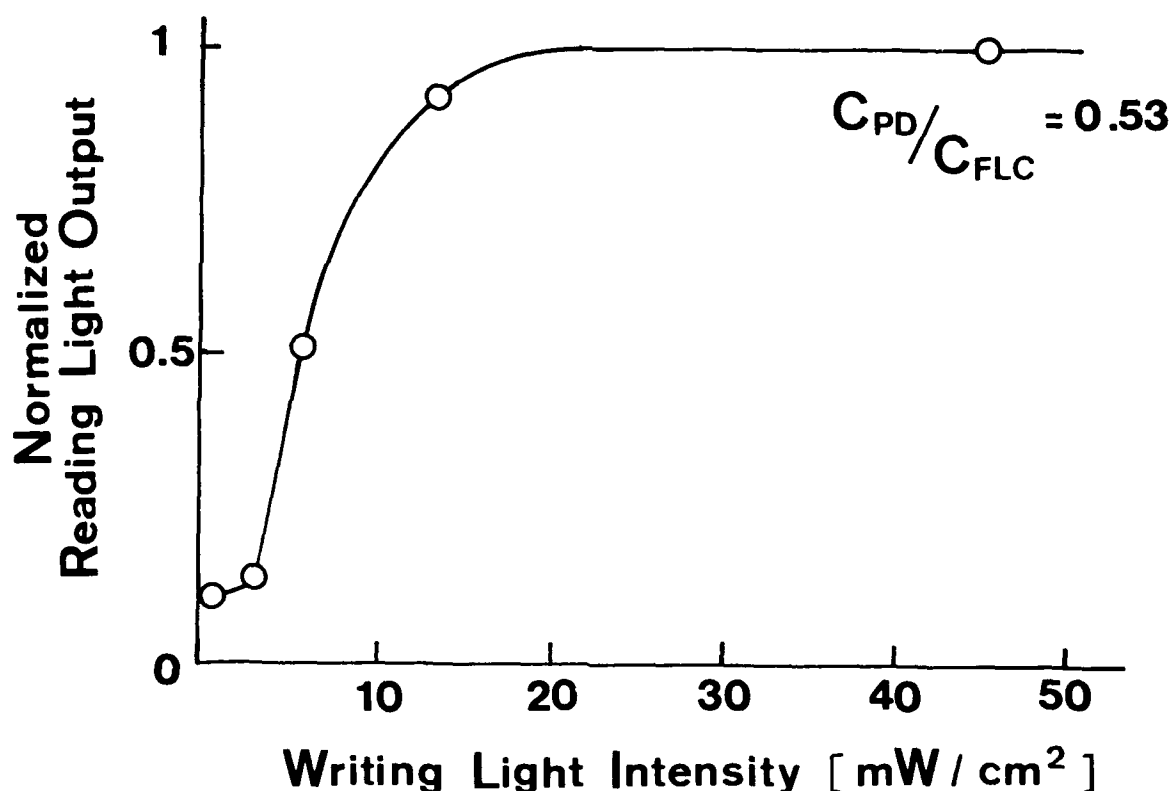


Fig.3: Plot of the normalized read-out light intensity in the memory level of one pixel against the writing light intensity.

Because this OASLM has the gray-scale memory capability, it can be used in analog optical computing, for example to implement the graded interconnections matrix of optical neural networks.

The authors gratefully thank Drs. F. Saito, M. Sakaguchi, N. Nishida, K. Hotta, S. Esho and C. Tani of NEC Corp. for their suggestions and encouragement.

References:

- 1) H. Maeda, C.M. Gomes, M. Yoshida, B.Y. Zhang, M. Kimura, H. Sekine and S. Kobayashi: 6th Int'l Display Res. Conf. (1989) 344.
- 2) M. Kimura, H. Maeda, C.M. Gomes, M. Yoshida, B.Y. Zhang, H. Sekine and S. Kobayashi: to be published in Proc. SID.
- 3) S. Tsujikawa, F. Okumura, M. Sakamoto, H. Ichinose, M. Imai, K. Sera, H. Asada and K. Kubota: Proc. 1990 International Topical Meeting on Optical Computing Kobe 9C3 (1990) 23.

OPTOELECTRONIC NEURON

Anton Rohlev, Christian Radehaus, and Jacques I. Pankove

NSF ERC Center for Optoelectronic Computing Systems

Department of Electrical and Computer Engineering

University of Colorado, Boulder, Colorado 80309-0425

303/492-5470, FAX 303/492-2758

R. F. Carson

Sandia National Laboratories

Albuquerque, New Mexico 87185

G. Borghs

IMEC

Leuven B-3030, Belgium

1. INTRODUCTION

In analogy to biological neurons that interact with other neurons both electrically and chemically, the optoelectronic neurons interact with other optoelectronic neurons electrically and/or optically. Like a biological neuron, the optoelectronic neuron can have multiple input and thresholding. While the chemical interaction between biological neurons can employ different neurotransmitters, the optical interaction between optoelectronic neurons can employ photons of different wavelengths.

2. PRINCIPLE AND STRUCTURE OF OPTOELECTRONIC NEURONS

We have used a pnpn photodiode made of III-V semiconductors^{1,2}. The inner two layers are made of GaAs. In operation, the inner homojunction is reverse-biased and the outer heterojunctions are slightly forward-biased. The $I(V)$ characteristic of this device is shown in Figure 1. Depending on the design, the breakdown in the dark is due to Zener breakdown or to punch-through of the depletion layer to one of the heterojunctions. The effect of illumination on the $I(V)$ characteristic is to reduce the breakdown voltage. With the loadline as shown, illumination causes the device to switch from point A (high voltage, low current) to point B (low voltage, high current). After breakdown, efficient double injection occurs at the heterojunctions, flooding the inner two layers with carriers, annulling the reverse bias and causing radiative recombination of the injected carriers. With a suitable optical cavity or distributed feedback, the device will lase, resulting in a large optical gain, i.e. a large ratio of optical output to input^{3,4}.

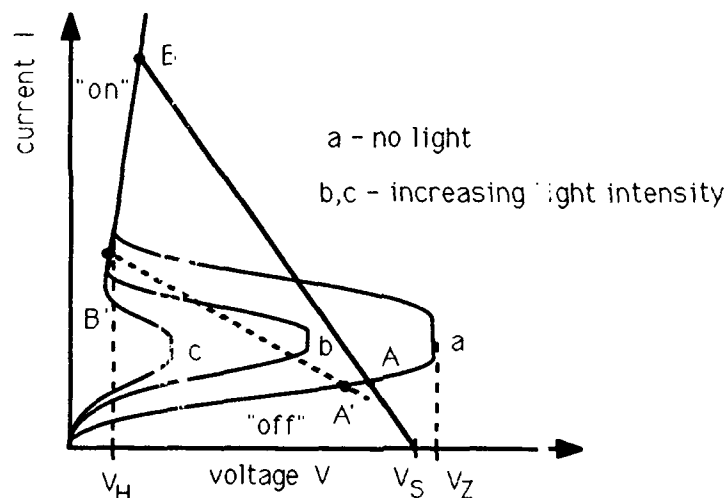


Figure 1: $I(V)$ characteristics of pnpn device without illumination (a) and with illumination (b,c).

Loadline AB for optoelectronic bistability; loadline $A'B'$ for all optically driven bistability.

This work was supported in part by the NSF under grant #CDR 8622236 and in part by CATL, an agency of the State of Colorado. Dr. Radehaus is a Max Kade Fellow.

3. OPERATIONAL PROPERTIES OF OPTOELECTRONIC NEURONS

3.1 Memory

Once the device has switched along the loadline AB , it will stay on as long as sufficient electrical bias is applied, even if the input light is turned off. This is a power consuming mode of operation. To turn off the device, one must lower the electrical bias such that the operating point B goes below the holding voltage, V_H .

To operate the memory in a low power consumption mode, the applied bias consists of pulses such that the operating point falls between V_H and the voltage at point B while the device is in the ON state, then the device will pulsate its light emission between a negligible value at V_H and the value corresponding to point B . This mode of operation will interrogate the device to find if it is in the switched mode. At biases that put the operating point slightly above V_H , the power dissipation is very low—in the tens of nanowatts range⁵. If the device is ON, the voltage across the device varies a little from V_H to V_B and the light pulsates. If the device is in the OFF state, the voltage across the device varies between V_H and V_A and no light comes out. To turn off the device, the applied voltage must bring the internal bias below V_H .

3.2 Thresholding

For the device to turn on, the exciting light must exceed a value determined by the load resistor and the applied bias, V_S . The exciting light intensity may be the sum of several incident lights, possibly at different wavelengths. Then one can obtain AND (coincidence or correlation) and OR logics.

The device can be switched both ON and OFF, all optically. To do this, one has to choose a load line as shown by the dashed line $A'B'$ of Figure 1. This provides an optical bias that defines both ON and OFF optical threshold values.

3.3 All optical nonlinearity and bistability

It should be possible to use the device as an optical switch that modulates the transmission or reflection of a light beam in response to the beam's intensity. As shown in Figure 2(a), the device could operate in a mode similar to that of a self electro-optic device (SEED).⁶ At low intensity, the device is absorbing because of the Franz-Keldysh effect in the reverse-biased central pn junction. When the intensity reaches threshold, the device turns ON, flooding the inner layers with free carriers and eliminating the high electric field. The inner layers become more transparent by virtue of the band filling (or Burstein-Moss effect). These effects are illustrated in Figure 2(b). Note that, although the device emits its own light when it is ON, it is capable of transmitting and modulating the incident light beam. We are currently testing this mode of operation.

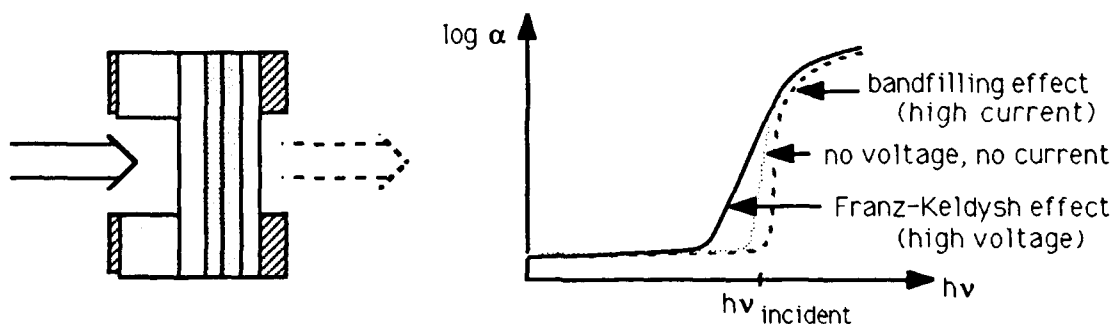


Figure 2: (a) pnpn device operated as a transmission modulator, and (b) absorption characteristic of pnpn device before switching (solid line, off-state) and after switching (dashed line, on-state).

4. INTERACTIVE PROPERTIES OF A NETWORK OF OPTOELECTRONIC NEURONS

4.1 Inhibition

When several devices are connected in parallel and biased through a common resistor (Figure 3), and one of the illuminated devices switches, the voltage across all the devices drops to V_B . The winning device is ON and emits light, while the other devices do not have enough voltage to be able to switch. This is a winner-takes-all network that provides global inhibition.⁷

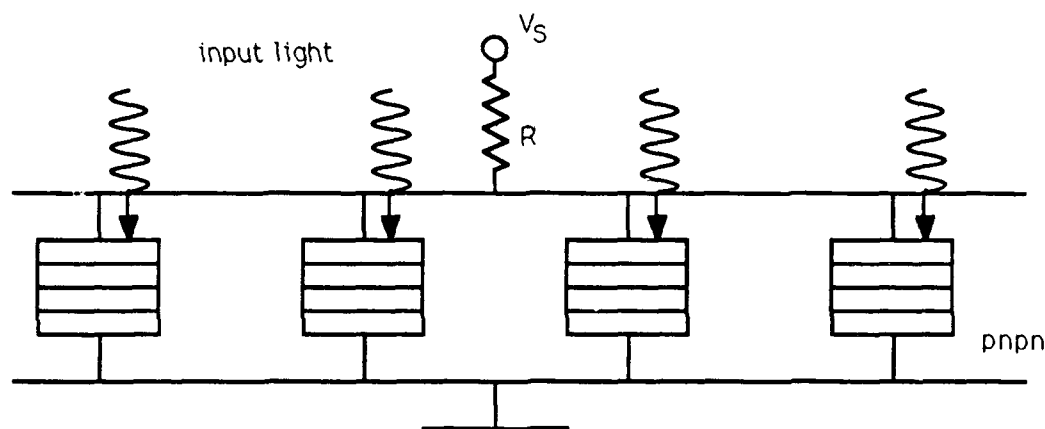


Figure 3: Winner-takes-all neural network.

To obtain local inhibition rather than global inhibition, a resistive network is inserted between adjacent devices in the array as shown in Figure 4, when one device is switched, the voltage across the other devices is reduced, which means that more light is needed to cause the other devices to switch. Hence, electrical cross-coupling results in local inhibition.

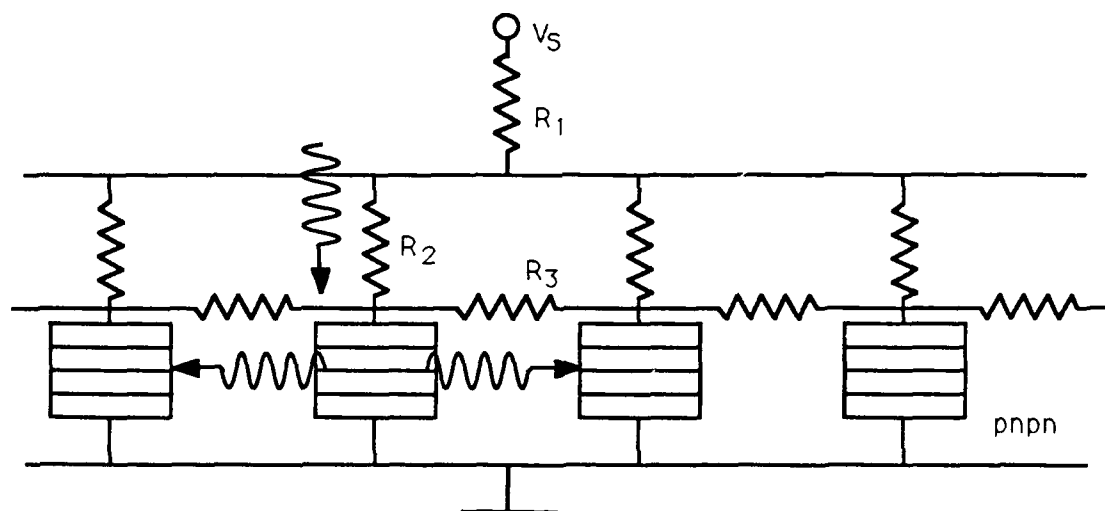


Figure 4: Network of pnpn devices interconnected for local electrical inhibition and optical enhancement.

4.2 Enhancement

In the presence of biasing light, a smaller optical input signal will suffice to reach the switching threshold than in the absence of optical bias. When a device switches, it emits light that may be directed to adjacent devices (local enhancement) or to more distant devices. Because the light bias reduces the intensity needed to trigger other devices, we call this mode of operation "enhancement of sensitivity." The use of optical fiber or optical waveguides determines whether the enhancement is local or nonlocal.

4.4 Learning

If the resistors in the array of Figure 4 are replaced by photoconductors or photo-transistors, one can optically adjust the weights of all the interconnections. One could interconnect devices and networks with optical fibers or optical waveguides which may be adjustable either electrically or optically to control the operation of the network. The adjustment of interconnection weights is a requirement for making adaptation

and learning possible. The change in weights can be self-adjusting via an algorithm or it can be supervised by a "teacher."

5. CONCLUSION

As an optoelectronic neuron, a pnpn structure has versatile switching properties, including thresholding and memory. The technology of these devices lends itself to the fabrication of integrated arrays, forming neural networks. Hence, in the future we can expect optoelectronic semiconductor chips that will provide dedicated neural functions such as pattern recognition. Since the devices are a combination of receiver and transmitter (detector and emitter), it should be possible to cascade arrays of these devices. One application that we are considering is to make a crossbar vector multiplier where the devices are in the shape of parallel strips, so that when one detector strip is turned on, it also forms an emitting strip that illuminates a set of detector strips through a weight matrix. This would be a variation of the technique proposed in [2]. The winner-takes-all array is a decision-making component.

6. REFERENCES

1. J. I. Pankove, R. Hayes, A. Majerfeld, M. Hanna, E. G. Oh, D. M. Szmyd, D. Suda, S. Asher, R. Matson, D. J. Arent, G. Borghs, and M. G. Harvey, *Optical Computing 88*, J. W. Goodman, P. Chavel, and G. Roblin, eds., *Proc SPIE* **963**, 191 (1989).
2. K. Kubota, Y. Tashiro, K. Kasahara, and S. Kawai, *Optical Computing 88*, J. W. Goodman, P. Chavel, and G. Roblin, eds., *Proc SPIE* **963**, 255 (1989).
3. Zh. I. Alferov, F. A. Andreev, V. I. Korol'kov, V. G. Nikitin, E. L. Portnoi, and A. A. Yakovenko, *Soviet Physics-Semiconductors* **6**, 637 (1972) and **8**, 1741 (1975).
4. W. F. Kosonocky, R. H. Cornely, and I. J. Hegyi, *IEEE J Quantum Electronics* **QE-4**, 176 (1968).
5. M. Kuijk, R. Pereira, M. Van Hove, R. P. Mertens, G. Borghs, and J. I. Pankove, *Electronics Letters* **26**, 281 (1990).
6. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus, *Phys Rev Lett* **53**, 2173 (1984).
7. J. I. Pankove, C. Radehaus, and K. Wagner, *Electronics Letters* **26**, 350 (1990).
8. C. Radehaus, to be published.
9. H.-G. Purwins and C. Radehaus, *Neural and Synergetic Computers*, H. Haken, ed. (Berlin: Springer-Verlag, 1988), 137.

Design and Demonstration of an Opto-Electronic Neural Network using Fixed Planar Holographic Interconnects

Paul E. Keller and Arthur F. Gmitro
Optical Sciences Center
University of Arizona
Tucson, Arizona 85721

1. Introduction

A key element of most neural network systems is the massive number of weighted interconnections used to tie relatively simple processing nodes together in a useful architecture. The inherent parallelism and interconnection capability of optics make it a likely candidate for the implementation of the neural network interconnection process. While there are several optical technologies worth exploring, we are looking at the capabilities and limitations of using fixed planar holographic interconnects in a neural network system and have implemented an initial test system using planar holograms and opto-electronic nodes.

2. System

All neural network systems consist of nodes (simple non-linear elements crudely imitating biological neurons) and weighted interconnections (synapses) between nodes. The basic system we have looked at employs optical interconnects and electronic nodes in a feedback architecture. A prototype is shown in Figure 1.

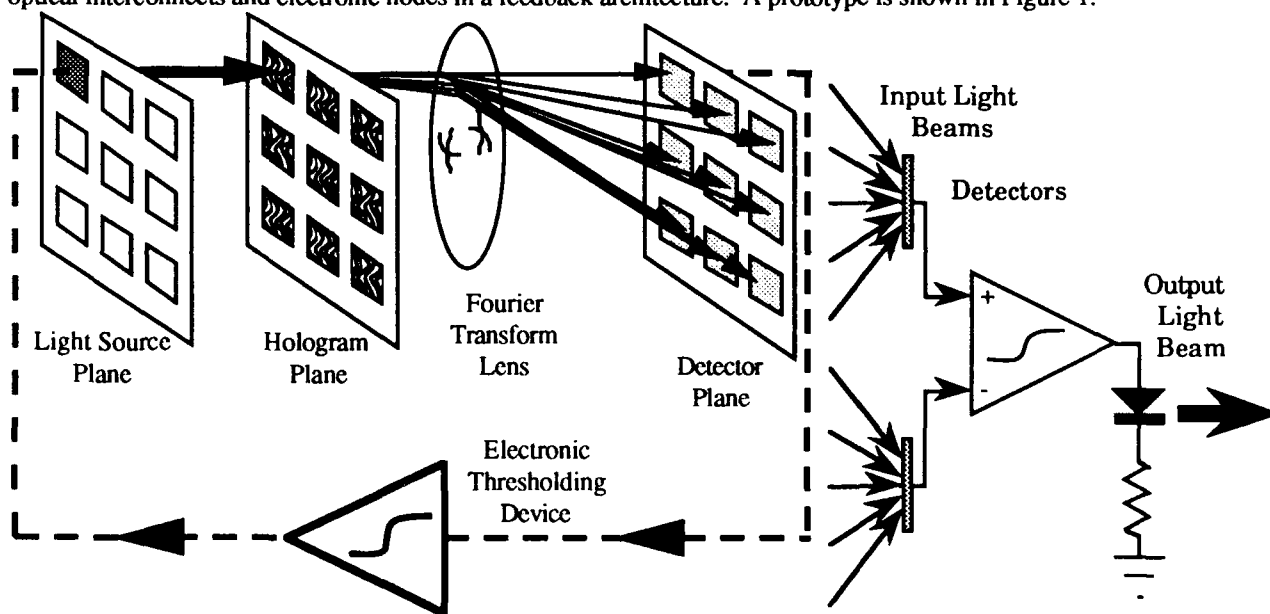


Figure 1: Prototype opto-electronic neural network system.

Figure 2: Single node in the system.

Each node is composed of an input summing port, non-linear transfer device, and an output port. In an opto-electronic system, a differential pair of detectors is operated as an input to the node; signals with positive (excitatory) weights arrive at one detector, and signals with negative (inhibitory) weights arrive at another detector. These detectors sum up the intensity of each optical signal arriving at the node. A threshold operation is electronically applied to the detected signal to produce an output signal. The output signal of the node modulates an optical source. Figure 2 illustrates an idealized node.

An individual node drives an optical beam that illuminates a single subhologram. Each subhologram stores the connection weights between that node and all other nodes. A subhologram is designed as a Fourier transform hologram and used in a coherent optical system so that the diffracted connection pattern is independent of subhologram position.

3. Design

The Hopfield¹ auto-associative memory model was chosen as a means to test the interconnect capability of planar holographic optical interconnects in the experimental opto-electronic neural network. This neural network

tries to associate each pattern presented to it with a pattern that it was trained on during an initial batch training process. Using the Hopfield outer product formulation, a training set of patterns was used to construct the fixed interconnection weights. The Hopfield model is a globally interconnected neural network; all nodes are connected to all other nodes with a deterministic strength or weight. These weights were then encoded in an array of binary amplitude subholograms.

Much effort went into the construction of the holograms used in the interconnect process. After examining a variety of computer generated hologram (CGH) techniques for accuracy of reconstructed interconnect weights, computation time, required space bandwidth product, and diffraction efficiency, two techniques, error diffusion and random search, were found to satisfy many of these criteria.

Both techniques were used to produce binary amplitude holograms. The general design techniques are as follows. The weights connecting a single node output to all node inputs are represented as an intensity pattern in the detector plane. The optical amplitude at the detector plane is given by the square root of the intensity. To reduce the dynamic range required to encode a hologram, a random phase function is added. Since amplitude holograms must produce a hermitian diffraction pattern, the mapped weights are shifted off the optic axis and a hermitian conjugate is added. To compensate for the sinc function roll-off in the connection pattern due to finite sized hologram pixels, the connection weights are multiplied by an inverse sinc function weighting. This predetermined diffraction pattern is then inverse Fourier transformed, and the transformed data values are normalized using the extreme amplitude values for the entire set of subholograms. Lexicographically scanning these sampled values, each sample is binarized. Since the data is continuous valued, an error is produced by the binarization. This error is propagated to the adjacent unbinarized pixels of the hologram; this is the error diffusion process.² The net effect of the error diffusion process is to reduce the total quantization error across the entire hologram. What remains is a high frequency binarization error that manifests itself as diffracted light far off the optic axis in the detector plane. The location of this diffracted light is controlled by the method used to distribute the binarization error in the hologram. To improve interconnection weight accuracy and to confine the diffracted spots of light to the center of each detector cell, each hologram is replicated 4-times vertically and horizontally (16 replicas). The error diffusion algorithm has shown the best performance for a non-iterative CGH design process.

Random search is an iterative process used to improve the connection accuracy of the error diffusion holograms. Starting with an error diffusion hologram, this method determines whether a perturbation (flipping a pixel in the hologram from opaque to transparent or vice versa) improves the accuracy. A perturbation is kept only if the accuracy is improved. This process is repeated until convergence. The main disadvantage of the random search process is the massive computation required. This process is related to the simulated annealing process except that no annealing takes place. The simulated annealing process allows accuracy degrading perturbations of the hologram to be kept with a probability modeled by the Maxwell-Boltzmann distribution.³ Simulated annealing, in theory, is able to find the globally optimum solution; in practice, limited computation requires compromises that may or may not produce good results. We have found that the random search algorithm produces holograms with almost the same performance as simulated annealing but requiring far less computation time.

For a large scale problem, electron beam fabrication would be required to produce the array of subholograms. For the small scale problem that was implemented, a photolithography process was used; the hologram mask was printed onto a sheet of film using a laser film writer and photographically reduced onto a holographic plate.

4. Experiment

The experimental opto-electronic neural network is illustrated in Figure 3. An initial pattern of 8 by 8 pixels is fed into the system by a computer; this pattern represents the initial state of the neural network. The pattern is written onto a Hughes Liquid Crystal Light Valve SLM using a high-intensity projection television. This binary pattern is polarization encoded onto the coherent optical laser beam by the SLM. The polarization beam splitting cube reflects only the vertical component of this polarized signal so that a binary amplitude pattern illuminates the hologram array. Each pixel of the pattern illuminates an individual subhologram. There are 64 nodes with 4096 bipolar interconnections in the experimental system.

The Fourier transform (Fraunhofer diffraction pattern) of the hologram array is produced at the back focal plane of the lens. To reduce scatter, the low frequency information of the diffraction pattern is filtered out. A relay lens is used to image the filtered Fourier plane onto a video camera. The light beams (diffraction from the hologram plane) arriving at the detector plane constitute the input to the node plane. In a practical opto-electronic neural network, each electronic node would take the difference between the signal on its positive-weight detector and its negative-weight detector, threshold the result, and drive an optical source such as a laser diode to be either on or off. For our experimental test system, a video camera is used to detect the optical input signals. The video signal is fed into the computer where it is digitized by a video frame buffer. The computer splits up the video frame into a grid and sums up the intensity in each cell to simulate a detector array. The difference and thresholding operations are performed digitally and the output stored in a video frame buffer, where the video output represents

the next iteration of the network. This forms the new network state, which illuminates the hologram plane, and the process continues until the network converges to a stable state.

As this experimental system was described, a node can take on two values; a value of 0 is represented by a dark pixel, and a value of 1 is represented by a light pixel. The performance of a Hopfield style neural network is significantly improved by using bipolar node values instead of unipolar node values. As an experimental test of bipolar nodes, a two step process was used. During the first step, a pattern was projected onto the SLM, and the detected pattern on the video camera was stored. During the second step, the inverse of the pattern was projected onto the SLM, and its detected pattern on the video camera was subtracted from the first detected pattern. Polarization encoding is another method for constructing bipolar nodes. A node value of +1 is encoded as horizontally polarized light, and a node value of -1 is encoded as vertically polarized light. With bipolar weights and bipolar state values, four detectors and two polarizers are used in the input summing port of the node.

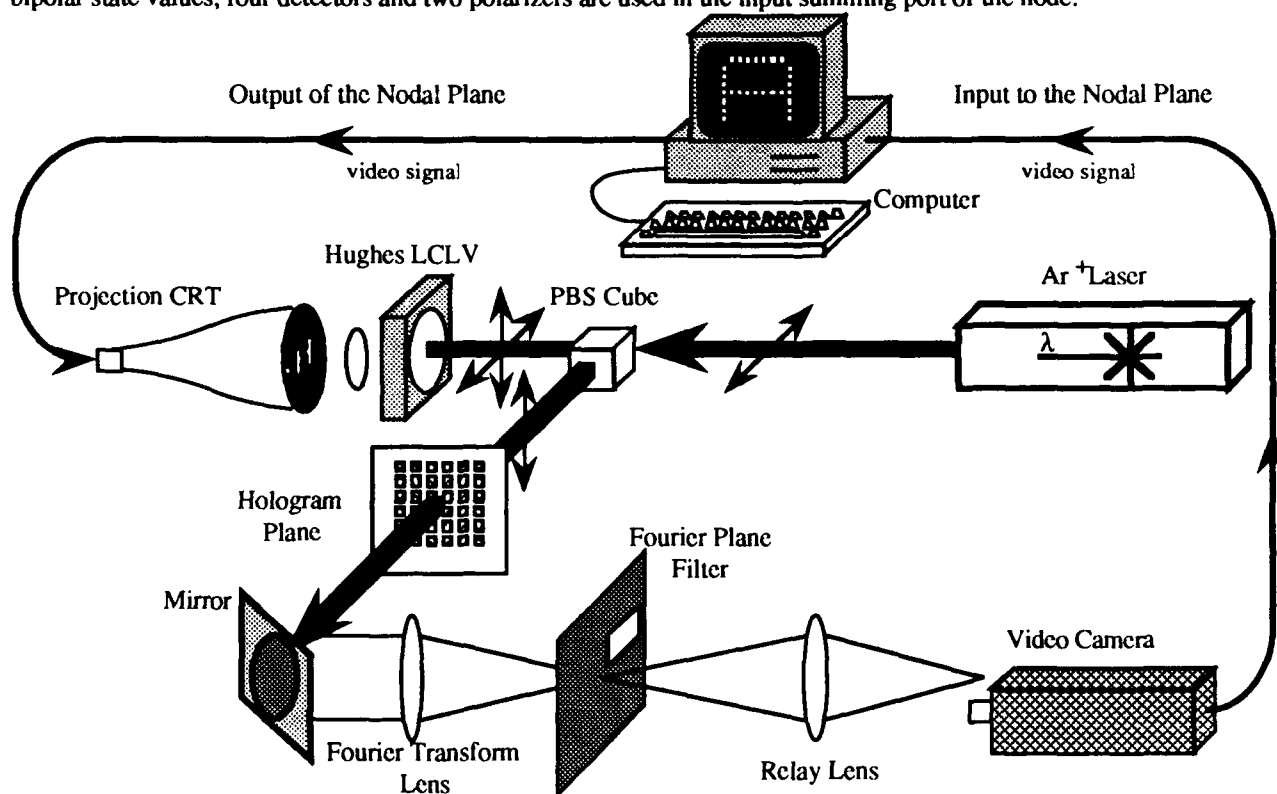


Figure 3: Experimental opto-electronic neural network used to test and evaluate the performance of planar holographic interconnects.

5. Results

The best performance of the associative memory neural network would come from a network storing randomly generated patterns, but since patterns of distinct structure (vertical lines, horizontal lines, diagonal lines) are generally encountered in vision and pattern recognition tasks, it was decided to use a set of ordinary typewriter characters (letters, numbers, symbols) to construct the test network. Using the Hopfield outer product formulation, a training set of three patterns, ABX , was used to determine the interconnection weights.

A prime feature of auto-associative memory neural networks is the convergence of the network to the ideal stored pattern when the input pattern is corrupted. By randomly flipping the pixels of the training set, a test set of corrupted patterns was generated. These patterns were presented to the experimental opto-electronic neural network, a computer simulation of the opto-electronic neural network, and a computer simulation of the ideal neural network. From the simulation, it was found that the auto-associative neural network constructed with random search holograms performed almost identically to the same neural network with ideal interconnect weights. The experiment, while not performing quite as well as the simulation, did come close for both unipolar and bipolar state values. The results with error diffusion holograms were not as good as with the random search holograms but show that error diffusion based holographic interconnects are a good trade-off between system performance and CGH computation time for bipolar state values. Figure 4 illustrates the performance of the experimental opto-electronic neural network with a test set composed of corrupted versions of the letter B. This figure is a graph of

the probability that the network converges to the original letter **B** as a function of the number of corrupted pixels in the input pattern. The total number of pixels in the input is 64. From this graph, it is apparent that when the number of incorrect pixels in the input pattern becomes too large, the network does not converge to the ideal pattern. Similar responses were found with the other stored patterns.

The small differences between the experimental results and the simulation results were caused by aberrations in the Fourier transform lens and relay lens, non-uniformity of the video camera, high frequency roll off in the holograms due to loss of resolution during the hologram fabrication process, and RF interference in the electronics produced by the argon ion laser's plasma discharge tube.

The experimental opto-electronic neural network system along with its computer simulation shows that a planar hologram can be used to implement the interconnect weights of a neural network. The results we have found with the experiment agree well our analytic calculations of neural network performance.⁴

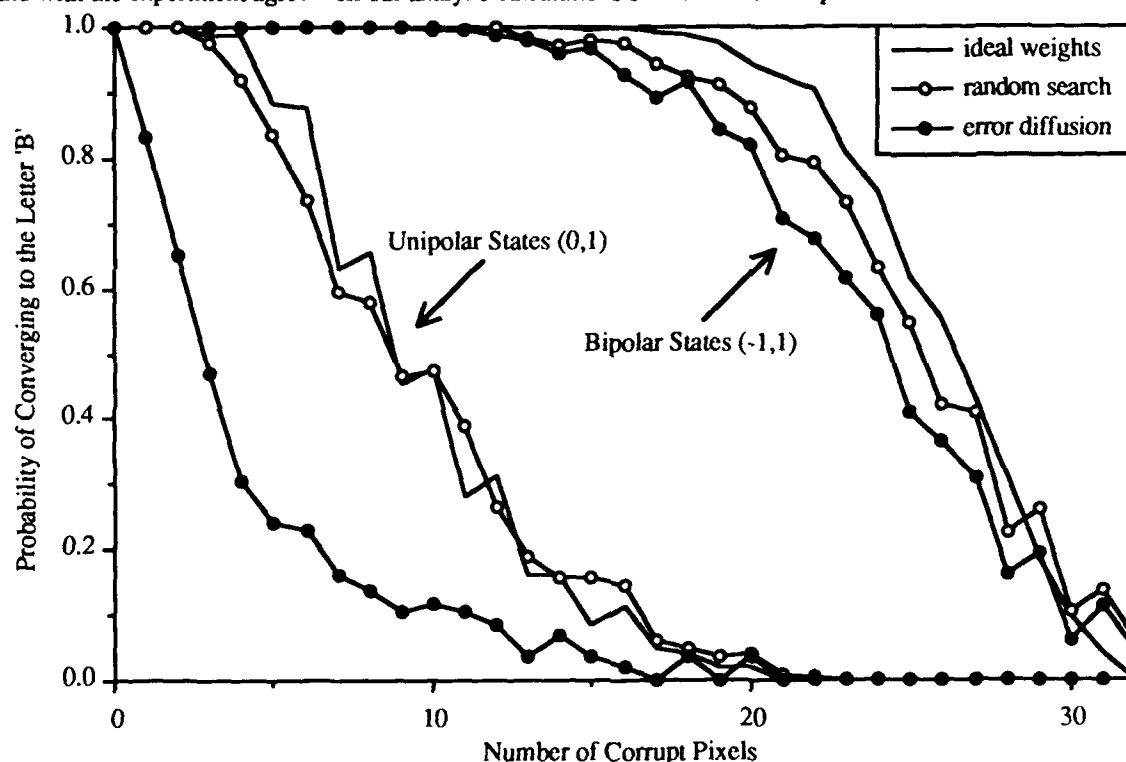


Figure 4: Performance of the ideal associative memory and the opto-electronic implementations. The graph plots the probability of convergence of the network to the correct state versus the number of corrupted pixels.

6. Conclusions

We have demonstrated that a system employing planar holographic optical interconnects can be used to implement a neural network architecture and that the performance of an optically implemented Hopfield style network comes close to that of an arbitrary system employing ideal interconnect weights.

This research is supported by the Optical Circuitry Cooperative at the University of Arizona.

7. References

- 1 J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proc. Natl. Acad. Sci. USA* **79**, 2554-2558 (1982).
- 2 S. Weissbach, F. Wyrowski, and O. Bryngdahl, "Quantization Noise in Pulse Density Modulated Holograms," *Opt. Comm.* **67**, 167-171 (1988).
- 3 M. Kim, M. Feldman, and C. Guest, "Optimum encoding of binary phase-only filters with a simulated annealing algorithm," *Optics Letters* **14**, 545-547 (1989).
- 4 A. Gmitro, P. Keller, and G. Gindi, "Statistical Performance of Outer-Product Associative Memory Models," *Appl. Opt.* **28**, 1940-1948 (1989).

Custom designed electro-optic components for optically implemented, multi-layer neural networks.

M.G.Robinson, K.M.Johnson,
D.Jared, D.Doroski, S.Wichart and G.Model.

Center for Optoelectronic Computing Systems,
Department of Electrical and Computer Engineering,
University of Colorado, Boulder, CO. 80309.
Tel: (303) 492 3260; Fax: (303) 492 3674.

Optical implementations of one-layer, *perceptron-like* neural networks have been shown to be very successful at associating pattern/target sets despite large system errors [1,2]. It has also been shown that large systems can be realized with such architectures ($\geq 4 \times 10^4$ interconnections [2,3]), and appreciable processing speeds have been demonstrated ($>10^8$ interconnections/sec [4]). However, single layer networks are limited due to their inability to associate patterns that are not linearly separable. A more general network is the two layer network, which is able to model arbitrary functions, and create any decision boundary within the input vector pattern space [5]. In order to implement such a network, it is necessary to perform a nonlinearity at the hidden layer before performing a subsequent matrix multiplication. In general, optical materials performing fast nonlinear processing require high optical powers. Hybrid opto-electronic devices can perform nonlinear operations at moderate speeds and low optical powers [6].

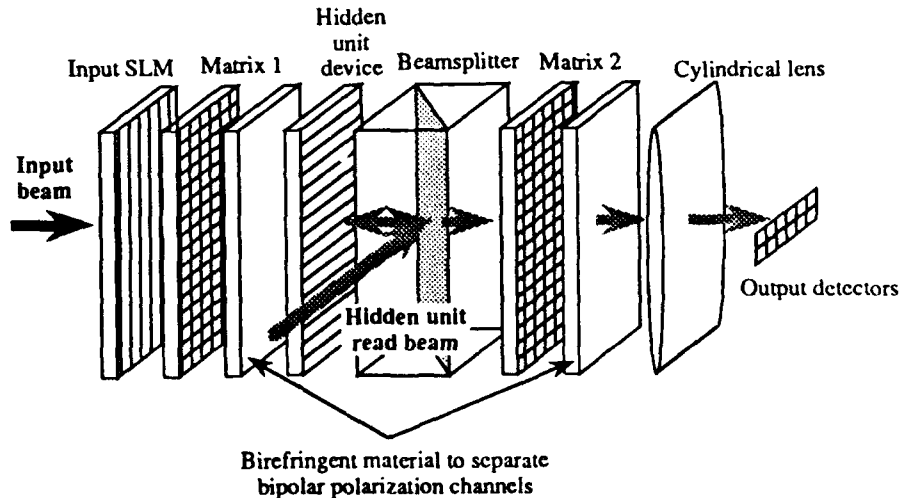


Figure 1 - Schematic of an SLM based, 2-layer, optically implemented neural network

In this paper, we present a electro-optic, nonlinear hidden unit device fabricated from an amorphous silicon p-i-n structure which is used to address a ferroelectric liquid crystal modulator. Designed specifically for spatial light modulator (SLM) architectures of the type used in previous one layer implementations (see Fig 1), the device has four important attributes. First, it has a striped format for summation in one dimension and striped optical

output for subsequent matrix multiplications; particularly suitable for compact systems. Second, it isolates the optical input from the optical output and offers optical gain. Third, it has a sigmoidal type nonlinear response required by hidden processing units, and finally it is bipolar allowing the output to be a function of the difference of positive and negative input optical channels. This last feature is particularly attractive as it allows the representation of negative weight values.

The detailed structure of the device is given in Figure 2.

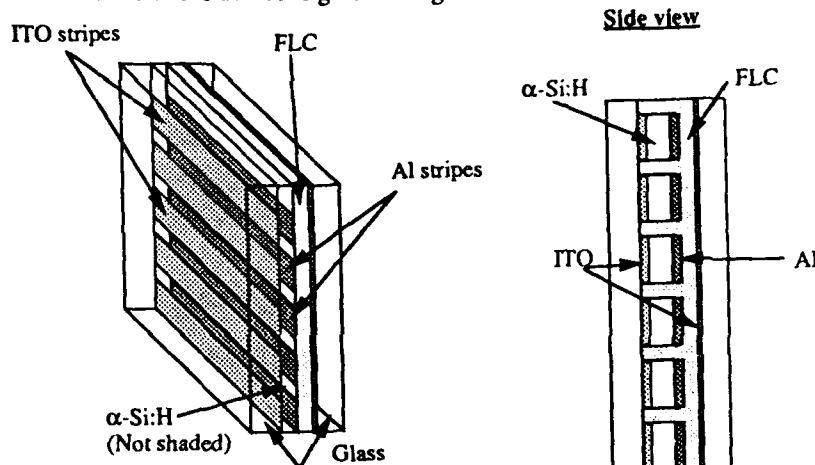


Figure 2 - a-Si:H hidden unit device schematic

It consists of stripes of a transparent conductor layer (ITO), a p-i-n amorphous silicon (a-Si:H) photodiode layer, and an opaque chromium (Cr) metal layer, coated onto an optically flat glass substrate. Between this striped structure and a further ITO coated substrate is a surface stabilized ferroelectric liquid crystal (FLC) material layer. By etching the Cr and a-Si:H at one end of the substrate, electrical contact via flex connectors can be made separately to both sides of each striped diode. By connecting the diodes in series via these external connections, and placing $\pm V$ across them the equivalent circuit shown in Figure 3 is obtained.

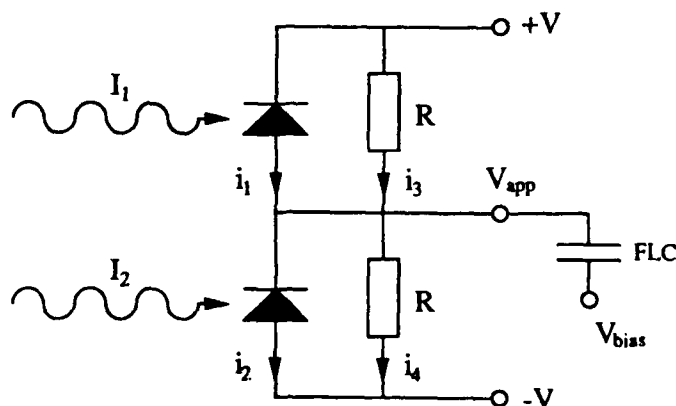


Figure 3 - Electronic device configuration

By choosing suitable external resistances R , the diodes are forced to operate in their linear region, i.e. their induced photo-current, i , is strictly proportional to the incident light intensity, $i = \alpha I$. Taking i_1 , i_2 , i_3 and i_4 as the currents through the two photodiodes and the external resistances respectively, and I_1 and I_2 as the incident optical intensity (see Fig 2), and applying Kirchoff's law yields an expression for the applied Voltage, V_{app} , on the FLC modulator,

$$\alpha I_1 + \frac{(V - V_{app})}{R} = \alpha I_2 + \frac{(V + V_{app})}{R} \Rightarrow V_{app} = 2R\alpha(I_1 - I_2)$$

The response of the FLC introduces the required nonlinearity, which can be chosen by altering FLC material used, the alignment treatment, and/or the operating temperature. The bias voltage, V_{bias} , on the common ITO electrode can also be adjusted to compensate for the non-zero FLC switching threshold voltage. The device operated in this manner demonstrates all four features necessary for a hidden unit layer in our multi-layer connectionist network architecture.

Experiments have been carried out on a device of this type containing 16, 1.32 mm x 21.12 mm, bipolar hidden units. Applying $\pm 15V$ to the device, $-5V$ to the common electrode, and $R = 20 K\Omega$, the optical output together with V_{app} was obtained as a function of the differential optical input, $I_1 - I_2$. The results of this experiment are shown in Figure 4 showing the required nonlinear response of the device.

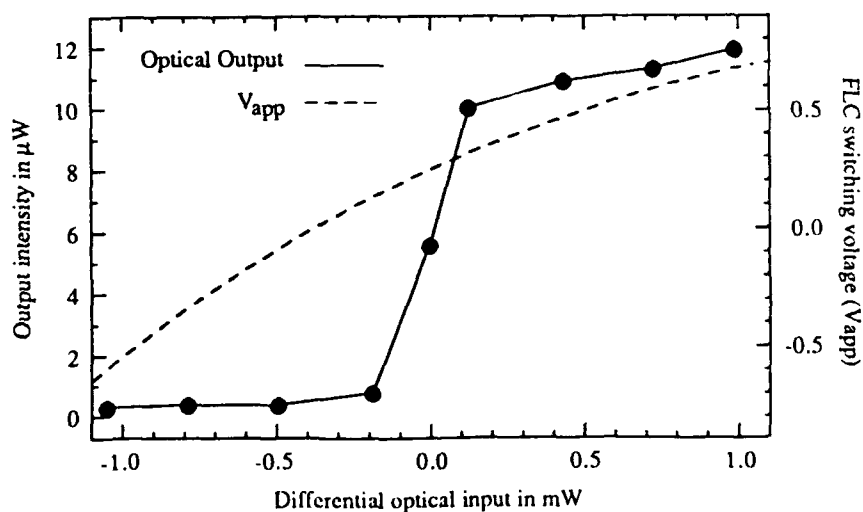


Figure 4 - Experimental output of the a-Si:H device

For this particular device the FLC material, CS1014, was bistable and has a dynamic switching response for small applied voltages. For this reason, the output was measured at a fixed period, 500μs, after the input optical beam is incident on the device. Further devices are being fabricated with alternative alignments, which

allow d.c. operation with a sigmoidal nonlinear response. Fig. 4 also shows the applied voltage to be slightly nonlinear and is attributed to a mismatch between the diode responses. Within a network incorporating such a device this can be considered a systematic error, and therefore compensated for during training [2]. The switching speed is dependent on the FLC material for optical power densities above $100 \mu\text{W}/\text{cm}^2$ at $\lambda = 514 \text{ nm}$ and is expected to be as low as $10 \mu\text{s}$ for variety of room temperature switching FLC mixtures.

Although this device satisfies the needs of hidden units in an SLM based 2 layer neural network, there are alternative technologies that are more appropriate for systems using holographic interconnects. Such an alternative technology is analog VLSI IC devices with FLC modulators fabricated on top as a means of obtaining two-dimensional optical outputs. This highly flexible technology has been used to produce three 32×32 arrays of thresholding optical elements in the form of an optically addressed SLM [7]. In the context of neural networks this can be considered as an array of neurons. Switching at a speed of up to $100 \mu\text{s}$, this device represents the largest array of smart SLM pixels to date and systems based on these devices will also be described.

In conclusion, we have described in detail a novel amorphous silicon/ferroelectric liquid crystal device designed specifically for implementation in a two layer optical connectionist neural network. Results from individual elements of the device have been described, and its operation within the optical connectionist machine will be presented.

References:

1. M.G.Robinson, K.M.Johnson, L.Zhang and B.J.Bigner, SPIE Vol. 1215 389 (1990).
2. M.G.Robinson, K.M.Johnson, "An Optical Connectionist Architecture: A System Analysis", Submitted to Appl. Opt.
3. E.G.Paek, J.R.Wullert II and J.S.Patel, Opt. Lett. 14 1303 (1989).
4. M.G.Robinson, L.Zhang and K.M.Johnson, Invited Talk, SPIE Conference on applications of neural networks, Orlando, Florida 1-5 April, 1991.
5. J.Baum Journal of Complexity 4 193 (1988).
6. K.M.Johnson, C.C.Mao, G. Moddel, M.A.Handschy and K.Arnett, Opt. Lett. 15 1114 (1990).
7. D.A.Jared, T.M.Slagle, K.M.Johnson and K.Wagner, OSA Technical Digest '90, p.255, FV4.

Optical Matrix-Vector Implementation of Binary Valued Backpropagation

Stephen A. Brodsky and Clark C. Guest

Department of Electrical and Computer Engineering
University of California at San Diego, R-007
La Jolla, CA 92037

Optical implementations of neural networks can combine advantages of neural network adaptive parallel processing and optical free-space connectivity. Binary valued Backpropagation¹, a supervised learning algorithm related to standard Backpropagation², significantly reduces interconnection storage and computation requirements. This implementation of binary valued Backpropagation used optical matrix-vector multiplication³ to represent the forward information flow between network layers. Previous analog optical network memory systems have been described⁴.

Binary valued Backpropagation (BVBP) is a bit-oriented network involving binary weights and computations. Connection networks trained with BVBP perform heteroassociative memory functions, using input patterns as keys for recalling output class representatives. The associative classes are stored as a distributed representation within interconnection strength matrices and thresholds. BVBP associations are performed by a cascade of thresholded matrix-vector products.

The elemental computation in BVBP is summation of Boolean Exclusive-Or operations on matrix and vector elements. The matrix and vector values are displayed on spatial light modulators (SLMs) allowing real-time system input and weight updates. Each channel optically performs an Exclusive-Or operation between one input vector element A and one weight matrix element B, described in Figure 1. Channels operate on corresponding dual rail pixel pairs of the matrix and vector SLMs. The Exclusive-Or can be expressed as a sum of And products, $AB' + A'B$, since the complemented variables A' and B' are available in the dual rail representations. Figures 1a, b, and c show the multiplication And, dual rail encoding, and Exclusive-Or computation, respectively. In Figure 1a, an intensity And is performed by shadow casting. Each channel's dual rail encoding is shown in Figure 1b. Matrix and vector elements are expressed with complementary representations. Figure 1c shows the Exclusive-Or function as separate left and right side And operations, followed by summation. The Exclusive-Or computations are performed in parallel as part of the optical summation.

BVBP uses this basic Exclusive-Or summation to combine net input vector information with stored matrix connection weights. A BVBP system is described by its connection layers, and the information flow through those layers. The implemented system contains two fully connected layers, consisting of a 12x12 array of bit connection weights, the Exclusive-Or connection function, discrete 12-valued thresholds, summation, and a nonlinear step output function. The two layers interconnect three vectors: system input, a hidden vector containing an indirectly constrained internal representation, and system output.

BVBP associative recall occurs with the forward information flow through the connection layers. Each layer can be viewed as row components operating in unison to perform the Exclusive-Or summation. Connection matrix row values indicate connection strengths between column vector inputs and row-oriented idealized neurons. Each binary input element is combined with the corresponding connection weight through the Exclusive-Or connection function. The results of the Exclusive-Or operations are discretely summed along the neuron row to produce the neuron's net input. Every neuron has an electronically implemented adaptive digital threshold, with range equal to the net input range. The neuron output is the binary step decision of the net input over the threshold, considered high when the net input is greater than or equal to the threshold. The parallel net input to all row neurons is equivalent to matrix vector multiplication. The parallel output is a thresholded matrix-vector product.

The matrix-vector multiplication system is based on two CRT-driven Hughes liquid crystal light valve (LCLV) spatial light modulators separated by a polarizing beam splitter (PBS). The CRTs control the LCLVs by defining regions of polarization rotation over the LCLV surface. Binary data values are encoded by dual-rail orthogonal polarization states within aligned 12x24 binary pixel arrays. The encoding of matrix and vector pixels is chosen to allow direct optical computation of the inner product by shadow-casting and summation. A photograph of the system is shown in Figure 2. The compact array operations are performed within the PBS cube and adjacent SLM surfaces.

The optical path, shown in Figure 3, begins with vertically polarized light reflected from the PBS into the vector LCLV. The PBS separates horizontal and vertical polarizations by selectively reflecting the vertical mode. Vector SLM pixels in a "one" state rotate light to horizontal polarization, which is reflected through the PBS into the matrix LCLV. Set matrix pixels rotate the remaining light to vertical polarization for reflection out of the PBS. Rows are summed horizontally with a cylindrical lens onto the CCD detector array. Significant intensity is output only when both corresponding pixels on the matrix and vector SLMs are set for rotation. If either pixel is not set for polarization rotation, minimal light exits for that pixel pair. This is equivalent to an intensity And function.

The Exclusive-Or connection function is the sum of two And products. By representing signals in dual-rail logic, each pixel V_i in the vector and the corresponding pixel M_{ji} in the matrix are adjacent to their complements, V_i' and M_{ji}' . The Exclusive-Or function is performed by arranging V_i opposite M_{ji}' , V_i' opposite M_{ji} , and summing the results, $V_i M_{ji}' + V_i' M_{ji}$, as part of the row fan-in. The row summation is focused by a cylindrical lens onto the CCD detector. Because the system uses dual-rail representation and optical fan-in, the Exclusive-Or computation is performed without additional cost. The convention for the matrix and vector display allows local representation for matrix and vector values.

The BVBP learning algorithm provides a fault-tolerant heteroassociative memory. By training directly on the optical system, the network gains a measure of fault tolerance by learning

associative solutions to imperfect hardware behavior. The network learns vector associations and system defects concurrently since hardware imperfections appear to the algorithm as incorrectly adjusted internal network values. The erroneous connection weights are adjusted to bypass imperfect configurations, providing associative corrections to system errors. Knowledge of specific system flaws is not required.

The network was trained on pattern pairs of characters described by binary vectors; each pair consists of an input and target pattern vector. The system was presented with an input pattern vector example and corrected by the BVBP algorithm until the system output vector and target pattern vector were identical for all patterns in the training set. The network was trained using the optical network for the forward interconnections and electronic software for the backward weight and threshold adjustment. Personal computers store connection weights, thresholds, and training patterns, and generate CRT signals containing spatial display information.

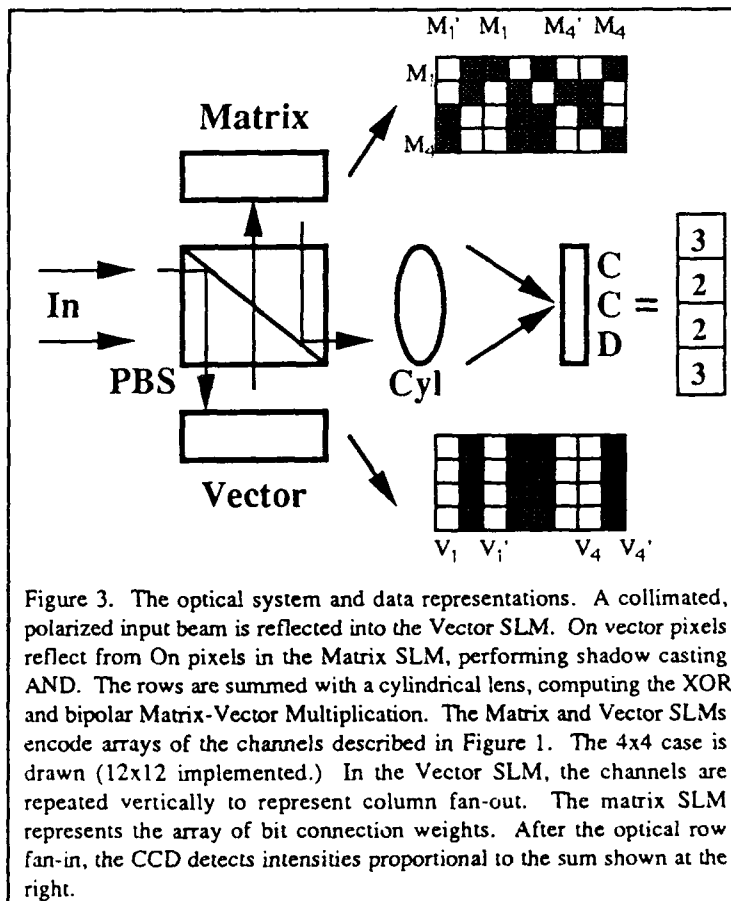
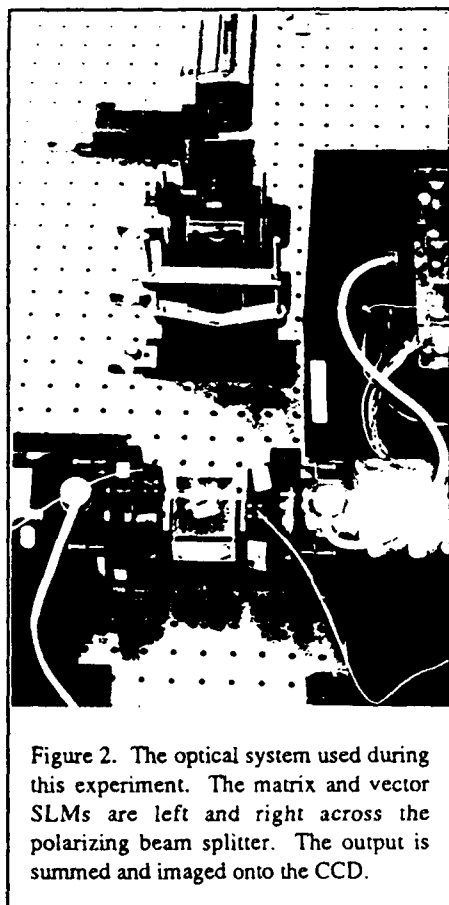
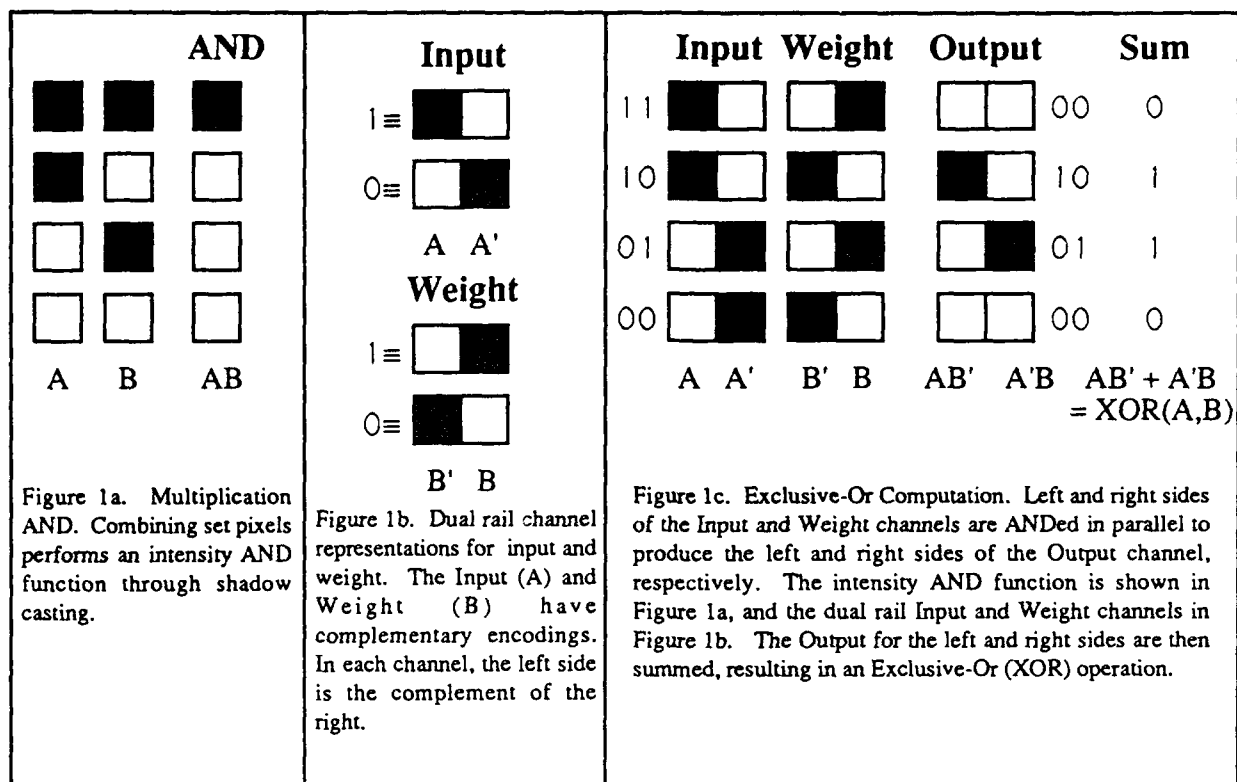
The system starts with a self-characterization and calibration procedure, and the network connections are initialized to random values. The learning procedure begins with a feedforward pass a thresholded optical matrix-vector multiplication for each connection layer. The input vector is presented on the vector SLM and the first connection weight matrix is displayed on the matrix SLM. The optical sum is collected on the CCD detector array, and thresholded electronically to produce the hidden layer input vector. The hidden vector and second connection matrix are then displayed on the SLMs, producing the system output vector. The errors are corrected electronically using the BVBP algorithm. The updated connection matrices are displayed in future iterations.

The network learned pattern pairs representing character symbols. An example of learning is the training of the four patterns in a training cycle of 74 pattern set presentations, resulting in 400 weight or threshold corrections. The maximum theoretical memory capacity of the network is one learned bit per weight stored, or 12 patterns. The system learned 1/3 of the maximum theoretical capacity. Some portion of the memory capacity learned associative corrections to possible faults in the optical system. The fault tolerance of the system is demonstrated by its ability to learn on non-ideal SLM devices.

The authors would like to acknowledge Gary C. Marsden and Dr. Sing H. Lee for their technical comments and the loan of the LCLVs.

References

1. S. Brodsky and C. C. Guest, "Binary Backpropagation in Content Addressable Memory," International Joint Conference on Neural Networks, (1990).
2. D. E. Rumelhart, J. L. McClelland and the PDP Research Group, "Learning Internal Representations by Error Propagation," *Parallel Distributed Processing, Explorations in the Microstructure of Cognition, Volume 1*, MIT Press, (Cambridge, Massachusetts, 1986).
3. J. W. Goodman, A. R. Dias, and L. M. Woody, "Fully parallel, high speed incoherent optical method for performing discrete Fourier transforms," *Opt. Lett.* 2(1) 1-3 (1978).
4. M. Kranzdorf, B. J. Bigner, L. Zhang, and K. M. Johnson, "Optical connectionist machine with polarization-based bipolar weight values," *Optical Engineering*, Vol. 28, 844-848 (Aug. 1989).



Experimental comparison of different associative memory techniques implemented optically by the same system architecture

K.J. Weible, N. Collings, W. Xue, G. Pedrini and R. Dändliker
Institute of Microtechnology, University of Neuchâtel
CH-2000 Neuchâtel, Switzerland

In recent years, much work has been going on in the optical implementation of artificial neural network systems. The parallel and crosstalk free interconnection characteristics of optical systems are well suited to exploit fully the desired parallel characteristics of artificial neural networks. The application of these systems as associative memories has been explored in many cases.¹⁻⁴ To facilitate optical implementation of these neural systems various modifications to the original Hopfield model have been proposed.⁵ It has been shown both theoretically and experimentally, that the storage and recall capacity of neural systems based on the Hopfield model are not significantly reduced when only the inhibitory (negative) interconnections are used.^{6,7}

In yet a further simplification, the binarization of the inhibitory interconnections has been proposed.⁸ Here also, the theoretical justification claiming negligible effects on the efficiency of the network has been presented.⁶ The aim of this paper is two-fold: First, to present an experimental analysis of the effects on the performance of the inhibitory model after binarization of the interconnections. Second, to compare the experimental performance of these inhibitory models to that of a discrete binary correlator implemented using the same system architecture.

Binary Inhibitory Neural Network

The binarization of the interconnection weights simplifies greatly the ease of implementation, both electrically and optically. The optical generation of the interconnections between neurons is achieved using components such as, holographic optical elements (HOE's), diffraction gratings, or microlenslet arrays. For the diffractive elements, binary weighting of the interconnections is much more uniform than gray level weighting. Using microlenslet arrays it is necessary to use some type of masking technique to apply the interconnection weights. In this case also, binary weighting is substantially simpler to implement with good uniformity requirements.

In the binarization or clipping of the inhibitory interconnections of the Hopfield model a large decrease in the recall capability of the system is seen. This is due to the lack of a uniform threshold point for the output neural plane. In our preliminary experimental results, even the steady-state input patterns would tend to deteriorate towards the set of neurons within the pattern that had the lowest threshold point. In a recent paper, another model using only binary, inhibitory interconnections has been presented.⁸ The Inverted Neural Network (INN) model simply sets all the interconnection weights between neurons that exist within a stored pattern to zero. The result is that in the steady-state, there is no light contribution from the other neurons within the same state. A consequence of this model is that the number of "on" neurons within the stored patterns must be limited so that a high input level to the other neurons not within the memory state is insured. Although it has been claimed that theoretically this model will perform to the same memory capacity as the inhibitory Hopfield model, it is not clear how the recall capacity of the system will be affected using a binary model. In the section on experimental results, a comparison of the recall capacity of the two models using the same system architecture will be compared.

Discrete Binary Correlator

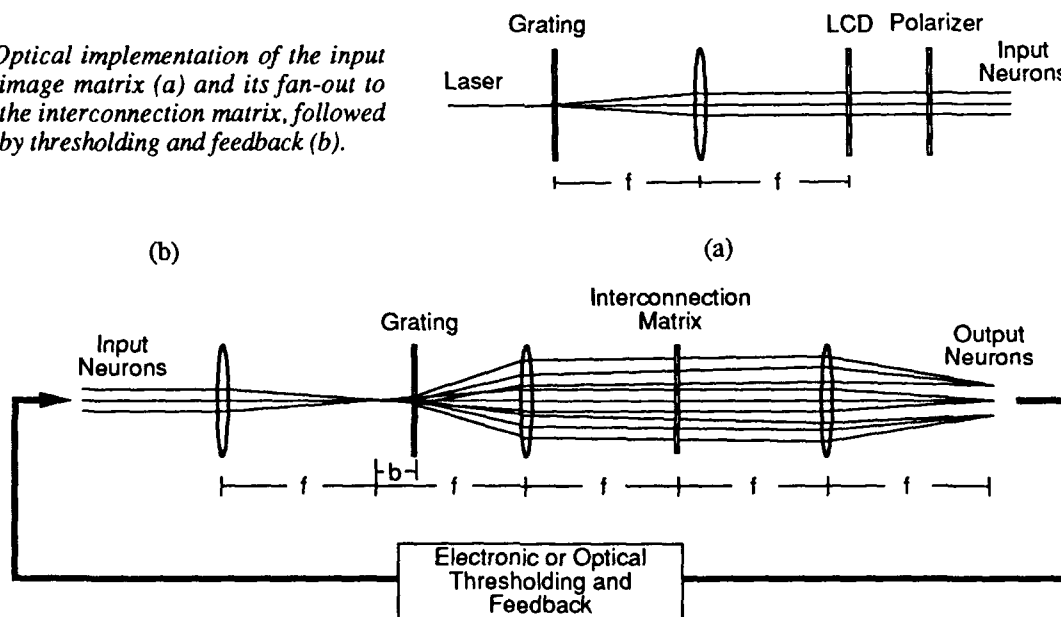
Although the emphasis of this paper is on the effects of binarizing the interconnections in an inhibitory neural system, it serves as a worthwhile performance measure to compare the results of these systems to that of a discrete binary correlator. The output results of the correlator differ from that of the associative memory neural systems in that its output is not a stored pattern as in a Content Addressable Memory (CAM), but a decision as to which pattern the input image most closely resembles. To truly compare the neural systems with a correlator it is necessary to place a pattern substitution system after the correlator to regenerate the memory pattern chosen by the correlator. A system of this type has already been studied in many symbolic substitution applications.⁹ In this case, the recognition part of the symbolic substitution system is simply a discrete binary correlator and the subsequent

substitution part reproduces the desired memory pattern. The difference between the system's performance as a CAM compared to that of symbolic substitution, is that now the thresholding operation of the CAM must be modified so that only the strongest pattern recognized produces an output. Whereas in the case of symbolic substitution, the thresholding operation produces many outputs. Since we are dealing with binary intensity patterns and not with polarization encoded data, it is necessary to use dual-rail recognition of the patterns, i.e. both the pattern and its inverse are recognized within the correlator. As a result of this dual-rail recognition, the intensity resulting from correlation with the inverse pattern must be subtracted from the intensity resulting from the pattern's correlation before the decision process is made. Although this intensity subtraction was part of the initial motivation to inhibitory only neural systems, it is implemented here so that the performance of the different techniques may be compared.

Optical Implementation

The three techniques for pattern recognition discussed above can all be optically implemented with the same system architecture. The binarized inhibitory system and the binary correlator system can both be realized using the architecture designed for the all inhibitory system.⁷ The architecture is based on the fan-out of a HeNe laser source using binary Dammann phase gratings.¹⁰ A pair of crossed phase gratings placed in the front focal plane of a lens produces the input image matrix, which consists of $N \times N$ diffraction orders of approximately equal intensity, Fig. 1(a). Each pixel within the input matrix is electrically addressed by means of a Liquid Crystal Display (LCD) placed in the Fourier plane of the lens. The LCD is used to rotate the polarization of the desired "on" pixels. A polarizer placed after the LCD converts the polarization encoded data into intensity. The desired fan-out of the input image matrix to the $N^2 \times N^2$ interconnection matrix is achieved using another set of crossed phase gratings, Fig. 1(b). These gratings are displaced a distance b from the focal plane of the lens following the LCD to provide a separation of the output neurons equal to that of the input neurons.⁹ In the output plane the intensity contributions from all the neurons within the input image multiplied by their synaptic weights are summed and an intensity threshold is performed. The thresholded results are fed back into the system as the new input image matrix to complete the cycle and the successive iterations continue.

Figure 1. Optical implementation of the input image matrix (a) and its fan-out to the interconnection matrix, followed by thresholding and feedback (b).



Experimental Results

Inhibitory Interconnections

The experimental results of the inhibitory neural system have been presented earlier and are presented here only for discussion purposes with respect to the results achieved using the two binary systems. Using only inhibitory

interconnections 5 patterns of 7×7 neurons were stored, Fig. 2(a). The grayscale intensity mask used to store these patterns is recorded on a millimask plate, Fig. 2(b). The convergence statistics represented in Fig. 5, are based on a sampling of 50 measurements (10 trials for each of the 5 stored patterns), except for the cases of 10, 11, and 12 errors, where the sampling was 100, 100, 70 measurements, respectively.

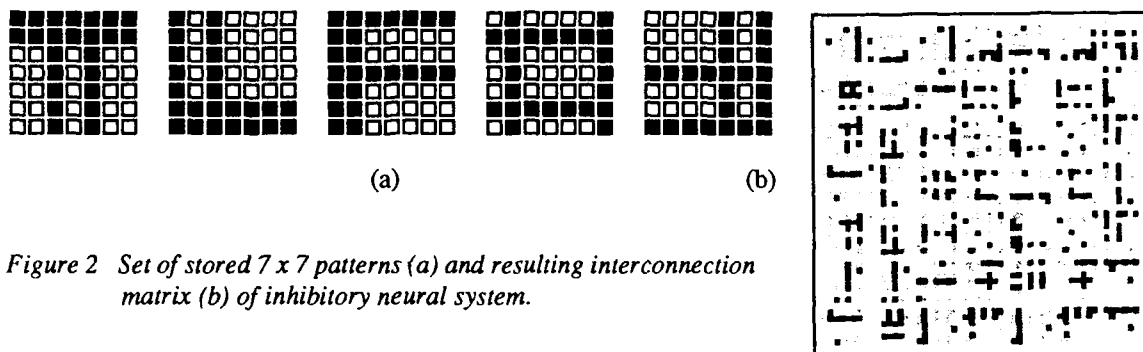


Figure 2 Set of stored 7×7 patterns (a) and resulting interconnection matrix (b) of inhibitory neural system.

Binary Inhibitory Interconnections

The experimental results of the binary inhibitory neural system were performed on a different set of stored patterns. The binary system performs better on patterns containing a fraction of "on" neurons considerably less than half, whereas 50% "on", 50% "off" is the optimum in the case of the inhibitory system. For this reason the patterns stored in the binary neural system were changed to those in Fig. 3(a). The binary intensity mask used to store these patterns is presented in Fig. 3(b). The convergence statistics represented in Fig. 5, are again based on a sampling of 50 measurements (10 trials for each of the 5 stored patterns).

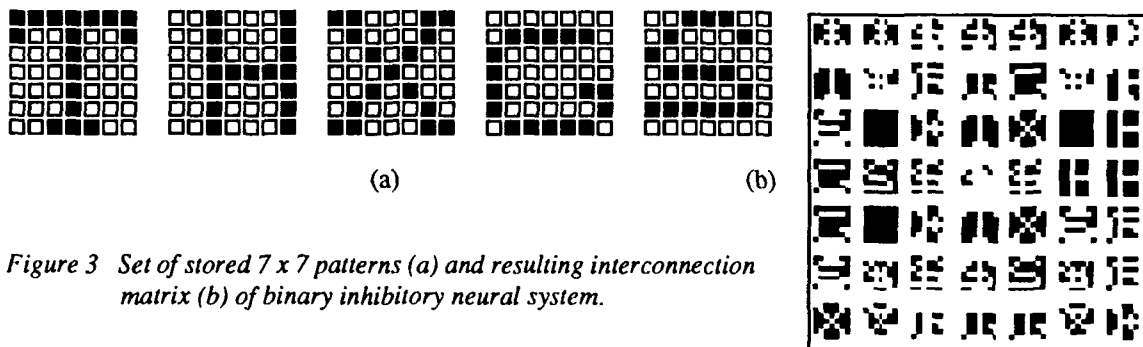


Figure 3 Set of stored 7×7 patterns (a) and resulting interconnection matrix (b) of binary inhibitory neural system.

Discrete Binary Correlator

The experimental results of the discrete binary correlator system were performed on the set of stored patterns used in the inhibitory neural system, Fig. 2(a). For the correlator system the number of "on" pixels within the stored patterns is not a factor. It is only important that all the stored patterns have approximately the same number of "on" neurons to account for a just comparison of the their correlated intensities. Therefore, the results obtained are characteristic for either of the sets of stored patterns. The binary intensity mask used to store the patterns is presented in Fig. 4. As in the neural systems, the curve representing the correlation statistics is based on a sampling of 50 measurements (10 trials for each of the 5 stored patterns), Fig. 5.

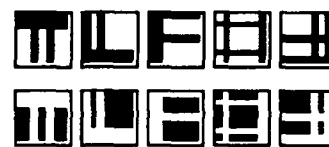


Figure 4. Correlator matrix for stored patterns displayed in Fig. 2(a).

Discussion of Results

The results of the optical implementation of each of the above systems are presented in Fig. 5. It can be seen in the statistical results that the performance of the two inhibitory neural systems are comparable up to 7 input errors (14% errors). As expected, the performance of the binary neural system deteriorates more rapidly than that of the inhibitory system, due to the information lost from the binarization of the interconnection weights.

However, the experimental performance of the discrete binary correlator significantly outperformed the inhibitory neural systems. The correlator system doesn't deteriorate to 50% convergence until a Hamming distance of 21 errors. The superior performance of the correlator suggests that it is much better suited for the recognition of discrete binary patterns. This same conclusion has been reached, elsewhere, based on simulation results.¹¹ Not only is the performance of the correlator much better, but the restrictions regarding the selection of stored patterns is less than is the case for the two neural systems. Although the correlator technique lacks robustness with respect to interconnection errors, it has a much greater Space-Bandwidth-Product (SBWP). As a result, it is possible to compensate for the robustness with the increased SBWP.

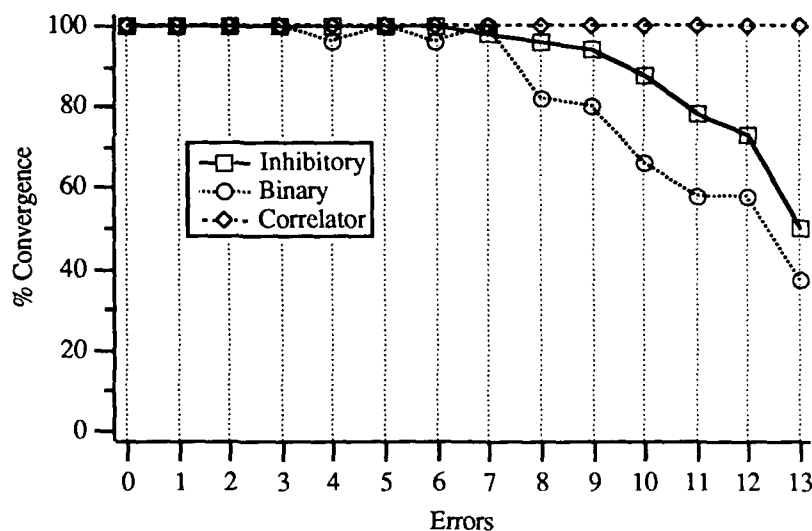


Figure 5. Statistical results of 7x7 pattern recognition

References

- 1) N.H. Farhat, D. Psaltis, A. Prata, and E. Pack, "Optical Implementation of the Hopfield Model", *Appl. Opt.* **24**, No. 10, 1469 (1985).
- 2) Ju-Seog Jang, Su-Won Jung, Soo-Young Lee, and Sang-Yung Shin, "Optical implementation of the Hopfield model for two-dimensional associative memory", *Optics Letters* **13**, No. 3, 248 (1988).
- 3) H.J. White, "Experimental results from an optical implementation of a simple neural network", *SPIE 963 Optical Computing* **88**, 570 (1988).
- 4) I. Shariv and A.A. Friesem, "All-optical neural network with inhibitory neurons", *Optics Letters* **14**, No. 10, 485 (1989).
- 5) J.J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- 6) H. Shouval, I. Shariv, T. Grossman, A.A. Friesem, and E. Domany, "An All-Optical Hopfield Network: Theory and Experiment", To be published in *Neural Networks*.
- 7) K.J. Weible, G. Pedrini, W. Xue, and R. Thalmann, "Optical Implementation of a Neural Network Associative Memory Using Diffraction Gratings", *Japanese Journal of Applied Physics* **29**, No. 7, L1301 (1990).
- 8) I. Shariv, T. Grossman, E. Domany, and A.A. Friesem, "All-optical implementation of the inverted neural network model", *SPIE 1319 Optics in Complex Systems*, 194 (1990).
- 9) R. Thalmann, G. Pedrini and K.J. Weible, "Optical symbolic substitution using diffraction gratings", *Appl. Opt.* **29**, No. 14, 2126 (1990).
- 10) The binary phase gratings have been kindly fabricated for our use by M.T. Gale et al. at the Paul Scherrer Institute (PSI), c/o RCA Laboratories Ltd., Zürich, Switzerland.
- 11) L. Tarassenko, B.G. Seifert, J.N. Tombs, J.H. Reynolds and A.F. Murray, "Neural Network Architectures For Associative Memory", First IEE International Conference on Artificial Neural Networks, Publication No. 313, 17 (1989).

OPTICAL MODULAR ARCHITECTURES FOR MULTI-LAYER BAM WITH 2-DIMENSIONAL PATTERNS

Soo-Young Lee, Hyuek-Jae Lee, and Sang-Yung Shin
Department of Electrical Engineering
Korea Advanced Institute of Science and Technology
P.O.Box 150 Chongryangni, Seoul, Korea

INTRODUCTION

After the first demonstration of optically-implemented Hopfield model [1] many neural network models have been investigated for large-scale optical implementation [2-8]. The 1-dimensional Hopfield model had been extended for 2-dimensional patterns [2], and optical implementation of bidirectional associative memory (BAM) [3-5] and quadratic associative memory [6,7] had been investigated. Adaptive neural network models such as multi-layer perceptron [8] had also been demonstrated. However performance of the simple Hopfield model and BAM is very limited, and many adaptive learning algorithms are too complicated to be implemented efficiently by optics. Also, when a new pattern need be added to the existing system, the correlation matrix learning rule of both the Hopfield model and BAM requires simple addition to existing interconnection weights, while error back-propagation learning rule for multi-layer perceptron requires to bring over all the previously stored patterns. Recently we had extended the BAM into multi-layer architecture, of which performance is quite comparable to that of multi-layer perceptron [9]. This multi-layer BAM (MBAM) still utilizes correlation matrices for easy optical implementation with outer-product matrix formation or inner-product recall. In this paper optical system architectures for the MBAM are presented for 2-dimensional patterns, and several implementation issues are discussed.

MULTI-LAYER BIDIRECTIONAL ASSOCIATIVE MEMORY

Let's consider a multi-layer neural network. Although our model is quite general, we just present 3-layer (2 hidden layers) network here for simplicity. The input layer, first hidden layer, second hidden layer, and output layer are represented by \mathbf{x} , $\mathbf{h1}$, $\mathbf{h2}$, and \mathbf{y} , respectively. In general different node numbers may be assigned to each layer.

Suppose M sets of input \mathbf{x}^s and output \mathbf{y}^s ($s = 1, 2, \dots, M$) need be learned. Provided corresponding hidden layer activations $\mathbf{h1}^s$ and $\mathbf{h2}^s$ were known, one might define interconnection weights as correlation matrices, i.e.

$$Z_{\mu} = \sum_{s=1}^M h1_j^s x_i^s, \quad Z_{kf} = \sum_{s=1}^M h2_k^s h1_j^s, \quad Z_{ik} = \sum_{s=1}^M y_i^s h2_k^s. \quad (1)$$

It may be understood as a multi-layer perceptron with correlation matrix interconnections between adjacent layers. Instead of interconnection weights the hidden layer activations $\mathbf{h1}^s$ and $\mathbf{h2}^s$ ($s = 1, 2, \dots, M$) are selected to minimize output global error defined as

$$E = \frac{1}{2} \sum_{s=1}^M \sum_l |y_l^s - y_l(\mathbf{x}^s)|^2, \quad (2)$$

where $y_l(\mathbf{x}^s)$ denotes the l th element of output vector corresponding to input \mathbf{x}^s and may be represented as

$$y_l(\mathbf{x}^s) = S_3(\sum_k Z_{ik} h2_k), \quad h2_k(\mathbf{x}^s) = S_2(\sum_j Z_{kf} h1_j), \quad h1_j(\mathbf{x}^s) = S_1(\sum_{\mu} Z_{\mu} x_i^s) \quad (3)$$

with proper nonlinear Sigmoid functions S_1 , S_2 , and S_3 for the first hidden layer, second hidden layer, and output layer, respectively [9]. Also these $\mathbf{h1}^s$ and $\mathbf{h2}^s$ may be selected from pre-

defined orthogonal sets with less performance [10].

During recall process the input signal propagates forward to the output layer, and then propagates backward to the input layer. This process goes on until no change is made during iteration. This bidirectional nature greatly increases error correction performance, especially for very noisy input.

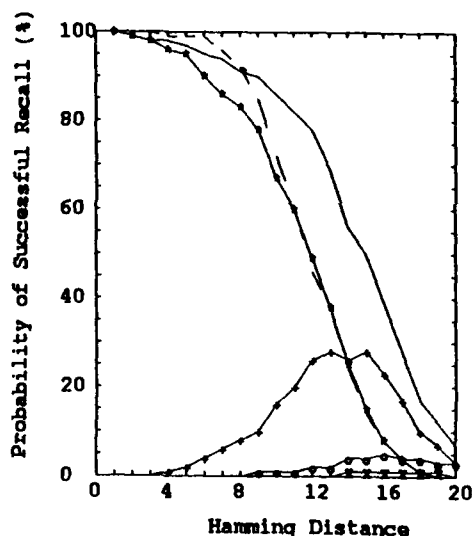


Fig.1 Error correction performances of 3-layer BAM (—) and 3-layer perceptron (---). Each training set consists of pseudo-random input/output binary patterns with 48 bits long. Symbols "*", "+", "o", and "x" denote error correction probability of the 1st, 2nd, 3rd, and 4th iterations, respectively, for MBAM.

Also, comparing the number of trained pairs to that of a single-layer BAM, i.e. about $N/(4 \log_2 N)$ and less than 3 for $N = 48$ in this case, MBAM demonstrates much higher storage capacity.

OPTICAL ARCHITECTURES

For optical implementation of the MBAM we adopted modular approach. Optical system architectures for single-layer BAM are first developed, and MBAM is organized as a cascade of these modules. Therefore optical architectures of these modules should be designed in consideration of cascability. Also they should not require excessive demand on spatial light modulator (SLM) performance, which is regarded as major limiting factor for large-scale optical implementation.

Optical architectures of single-layer BAM for 1-dimensional patterns had been reported with both vector-vector outer-product matrix formation scheme [3,4] and vector-vector inner-product recall scheme [7]. Both architectures utilize line-shape photo-diode/laser-diode (PD/LD) arrays for summation and non-linear operation. However, for large number of neurons, the line-shape PD/LD elements become very long, of which uniformity may limit implemented system performance.

Performances of MBAM and multi-layer perceptron models are also compared in Fig.1. Ten binary patterns with 48 bits long are trained in 3-layer hetero-associative memory neural networks by MBAM and perceptron learning algorithms, and probabilities of successful recall are plotted as functions of input Hamming distance, i.e. number of different bits with a stored pattern. In the figures 100 input patterns are randomly generated to satisfy required Hamming distance with each of the stored patterns, fed to MBAM and 3-layer perceptron models, and their overall convergence characteristics are collected. For small Hamming distances the perceptron model has slightly higher error correction performance which may be resulted from higher degree-of-freedom of the perceptron model, i.e. whole elements of the interconnection matrices instead of hidden layer activations. However, for large Hamming distances, the bidirectional nature of MBAM greatly increase error correction performance and this new model works much better than the other. Although successful recall of feed-forward network only, denoted as "*" in Fig.1, is less than or equal to that of perceptron, error correction by multiple iteration becomes important for heavily corrupted inputs.

In Fig.2 optical architectures of the two basic implementation schemes are shown for 2-dimensional patterns. Instead of line-shape PD/LD arrays 2-dimensional PD/LD arrays are used. A multi-focus hologram (MFH) and a 2-dimensional lenslet array (LA) with a spherical lens perform the required rank-4 optical interconnections [6]. Unlike 1-dimensional architectures using cylindrical lenses, only one directional path is allowed, and cascade of two modules is required for a single-layer BAM.

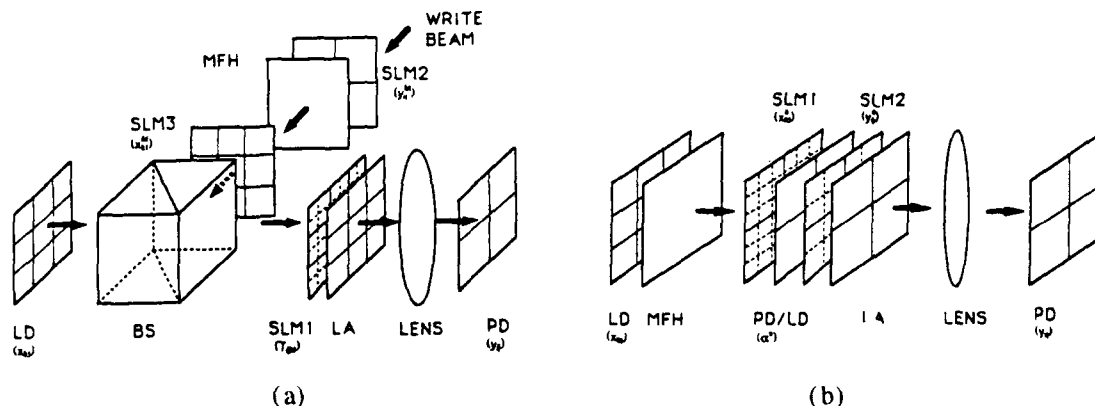


Fig.2 Optical BAM architectures for 2-dimensional patterns
(a) outer-product scheme, (b) inner-product scheme

For the outer-product scheme in Fig.2(a) the SLM stores analog interconnection weights, which may be modified by outer-products of new input (x) and output (y) matrices. This matrix-matrix outer-products are implemented by a 2-dimensional MFH, and matrix-tensor multiplications are done by a 2-dimensional lenslet array (LA) with a spherical lens. Because the interconnection weights are represented as sum of matrix-matrix correlations, as shown in Fig.2(b), the BAM may also be implemented by first calculating inner-products (α^s) between recall input matrix (x) and all stored input matrices (x^s , $s=1, \dots, M$) and later calculating another inner-products between the α^s and all stored output matrices (y^s , $s=1, \dots, M$). These inner-products are also implemented by a MFH and a LA with a spherical lens, respectively. The PD/LD arrays in Fig.2(b) receive lights from the left, sum the intensities, and emit lights to the right side. For higher order associative memories square or exponential operation may also be applied to the summed intensity values [7]. Unlike the outer-product scheme this inner-product scheme utilizes binary SLMs, which may be commercially available in high resolution. Also, in many practical applications, number of stored patterns is less than number of matrix elements, and this inner-product scheme requires less number of SLM elements.

An optical architecture for MBAM is shown in Fig.3. Four of the modules, 2 for each layer for bidirectionality, are cascaded for 2-layer (1 hidden-layer) BAM. For cascade operation the photo-detector (PD) should activate laser diode (LD) of the following module, and both operations are merged in a PD/LD array. Although both outer-product and inner-product architectures may be used, only inner-product modules are used here for their efficiency on SLM usage. This modular architecture is quite general, and applied to MBAM of as many layers required.

CONCLUSION

In this paper we presented optical architectures for MBAM. Both outer-product matrix scheme and inner-product recall scheme are investigated for 2-dimensional patterns. The latter requires binary SLMs, which is advantageous for practical large-scale implementation. Also the

module may become available in a solid compact form for many practical applications. Both architectures are cascadable for general n -layer BAM.

Acknowledgement: This research was supported by the Korea Science and Engineering Foundation.

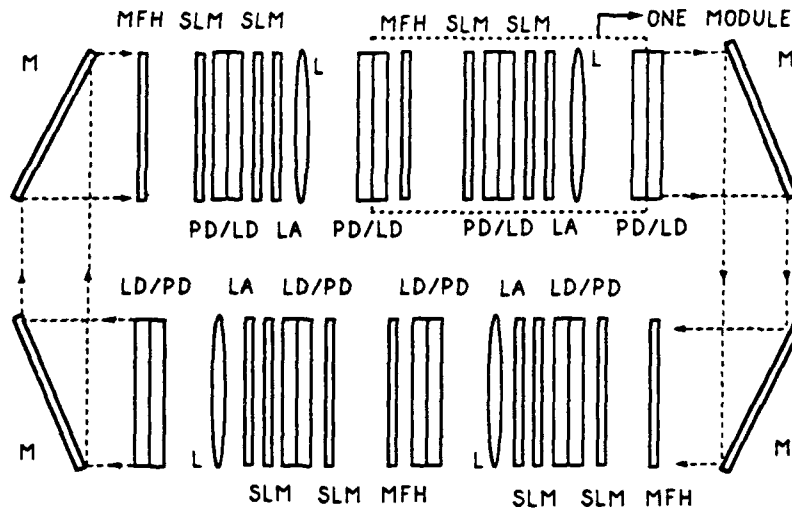


Fig.3 Optical inner-product architecture of 2-layer BAM for 2-dimensional patterns

REFERENCES

- [1] N. Farhat, D. Psaltis, A. Prata, and E. Pack, "Optical implementation of the Hopfield model," *Appl. Opt.* **24**, 1469 (1985).
- [2] J.S. Jang, S.W. Jung, S.Y. Lee, and S.Y. Shin, "Optical implementation of the Hopfield model for two-dimensional associative memory," *Opt. Lett.* **13**, 248 (1988).
- [3] B. Kosko and C. Guest, "Optical bidirectional associative memory," *Proc. SPIE* **758** (Jan. 1987).
- [4] R. Athale, H. Szu, and C. Friedlander, "Optical implementation of associative memory with controlled nonlinearity in correlation domain," *Opt. Lett.* **11**, 482 (1986).
- [5] J.M. Kinser, H.J. Caulfield, and J. Shamir, "Design of a massive all-optical bidirectional associative memory: the big BAM," *Appl. Opt.* **27**, 3442 (1988).
- [6] J.S. Jang, S.Y. Shin, and S.Y. Lee, "Programmable quadratic associative memory using holographic lenslet arrays," *Opt. Lett.* **14**, 838 (1989).
- [7] A. VonLehman, E.G. Pack, L.C. Carrion, J.S. Patel, and A. Marrakchi, "Optical implementation of a high order associative memory," *Proc. IJCN*, II-473 (June 1989).
- [8] N. Kasama, Y. Hayasaki, T. Yatagai, M. Mori, and S. Ishihara, "Experimental demonstration of optical three-layer neural network," *Japanese J. of Appl. Phys.* **29**, L1565 (1990).
- [9] S.S. Min and S.Y. Lee, "Multi-layer bidirectional associative memory," *Proc. Inter. Joint Conf. Fuzzy Logic and Neural Networks*, 251 (July 1990).
- [10] T.Y.D. Lam and J.E. Carroll, "Double layer associative memories," *Opt. Comm.*, vol.70, no.4, pp.293-298, March 15th, 1989.

An Optical Processing Unit for Relational Database Operations

Pericles A. Mitkas
Dept of Electrical Engineering
Colorado State University
Fort Collins, CO 80523

P. Bruce Berra
Electrical and Computer Engineering Dept
Syracuse University
Syracuse, NY 13244

1. INTRODUCTION

The management of very large databases (order of hundreds of gigabytes), combined with the real-time response requirement, poses a formidable task even for today's powerful computers. Special-purpose computers dedicated to database management, known as database machines, must provide adequate secondary storage to accommodate the database, high transfer rates to the processing units, and a large degree of parallelism.

Recent advances in optical technology have yielded optical memories, such as modified optical disks and holograms, that offer both large storage capacity and massive transfer rates. That, coupled with the development of fast and highly parallel optical processing elements, renders the application of optical techniques to database processing worth investigating. Berra et al, in a series of papers [1,2,3], introduced the potential for using parallel optical disks and optical processing in very large data and knowledge bases. They emphasize the superior capabilities of optics in terms of capacity and parallelism, and argue that optical processing will become practical with the development of the appropriate devices.

The elementary operations required by database management applications are often limited to comparisons and textual pattern matching. Another common characteristic is that a large amount of data have to be retrieved and processed in order to produce the query result, which is usually a small fraction of the database. These characteristics suggest that an optical database machine will need very high bandwidth but will not require extremely complex processing capabilities.

We have designed an opto-electronic processing system capable of performing a rich set of relational database operations (i.e., union, intersection, set difference, projection, selection and join). The relational database model is adopted because of its popularity and its tabular representation of data which dovetails nicely with the array processing capabilities of optics. In this paper we first describe the system and then show how the various relational operations are performed. We conclude with an initial performance analysis of the system.

2. THE OPTICAL PROCESSING SYSTEM

A block diagram of the system operating as a database machine connected to a front-end host is shown in Fig. 1. When a request for a transaction is issued by the host, it is passed to the control unit where it is compiled. The relations involved in the query are located in the secondary memory and their contents are retrieved into the buffer. Tuples from the buffer are loaded into the optical unit in a tuple- or page-oriented mode and the results of the processing are recorded in the bit array. At the end of an operation, the contents of the bit array indicate the tuples that satisfy the query. Only those tuples are retrieved from the buffer and transported to the front-end computer.

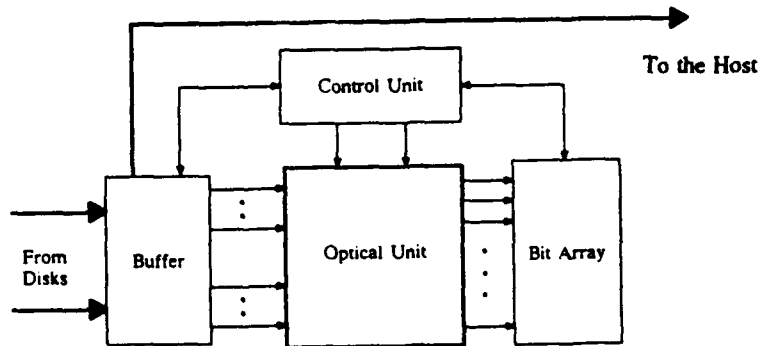


Fig. 1.
Block diagram
of the opto-electronic system.

The optical database processing unit (ODPU) performs parallel digital word comparisons according to the polarization-based method and is shown in Fig. 2. It consists of two spatial light modulators (A and B), a photodetector array, and a set of cylindrical lenses. The SLM A is one-dimensional with n pixels. The SLM B is two-dimensional and its size, $m \times n$, will be in the order of 10,000 pixels. Its vertical size, n , will be that of SLM A. Each character occupies 8 pixels. A suitable SLM is the SIGHTMOD magneto-optic modulator [4], developed by Semetex.

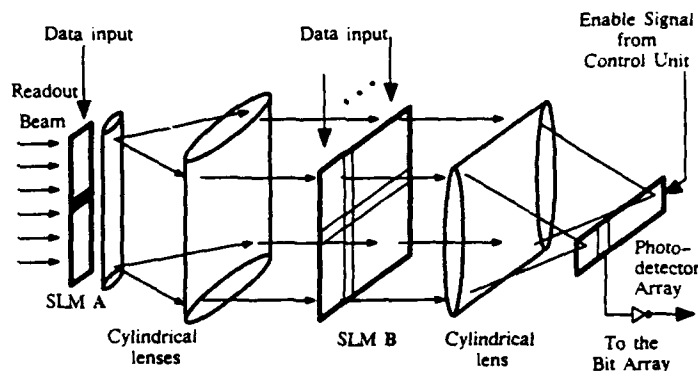


Fig. 2.
The optical database processing unit.

The search argument is loaded into the SLM A. The readout beam is modulated according to the information in A and then expanded horizontally (that is, it is replicated m times) by the combination of the two cylindrical lenses. The expanded beam is flashed onto SLM B which holds a page of a relation, one tuple per column. The cylindrical lens positioned after SLM B collects the output of an entire column into a single photodetector cell in the comparator. The information from the comparator is inverted and transferred in parallel to the electronic bit array. During the next cycle, a new search argument is loaded into A or a new page into B and the process is repeated. In this way, m comparisons of n -bit words take place in each cycle.

The results of the optical comparisons are recorded in a large 2-D bit array shown in Fig. 3. The size of the array ($x \times y$) depends on the cardinalities (r_1 and r_2) of the two largest relations in the database. The rows of the bit array are divided into segments, each m rows wide. A segment corresponds to a page (m tuples) of a relation. At any given instance, the m inverted outputs of the photodetector cells are transferred to m bit positions in the array. Three pointers are employed to point to these m positions: Page Pointer (PP), Row Pointer (RP), and Column Pointer (CP).

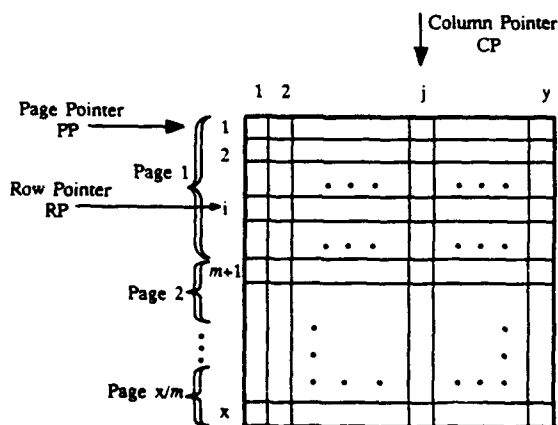


Fig. 3. The bit array.

3. IMPLEMENTATION OF RELATIONAL OPERATIONS

In the following discussion, two relations, R and S , are assumed with cardinalities r and s , respectively. The result relation will be denoted as Z and its cardinality z may vary from 0 to rs .

3.1 Union

Each tuple of the first relation has to be compared against all the tuples of the second relation. Pages from R are loaded, one at a time, into B while tuples from S are input, one by one, into A . A page stays in B until all the tuples of S are exhausted. If a match is detected, the tuple in A already exists in R and it is marked in the bit array, because it may not appear in Z twice. When the records of S are exhausted, a new page of R is loaded into B and the process is repeated with only the unmarked tuples of S . When all the pages of R are checked, Z is formed by adding the unmarked tuples of R to all the tuples of S .

3.2 Intersection

The intersection of two relations includes only the tuples that belong to both relations. The operation is executed much like the union of two relations. Pages of the smaller relation are loaded into B and compared to all the tuples of the other relation that are stepped through A . When a match is detected, the tuple in A belongs to the intersection and can be transferred directly to Z and then to the host. Thus, the formation and the output of Z can be overlapped with the execution process.

3.3 Set Difference

The set difference of relations R and S is the set of tuples in R but not in S . This set can be formed as a by-product of the union operation because at the end of a union the first column of the bit array indicates which tuples of R are in S (the ones for which the corresponding pixels have a value of 1) and which are not. Therefore, Z will contain the tuples of R that correspond to pixels with a value of 0.

Note that the result for all three operations will be available after a single execution of the process.

3.4 Selection

This operation selects from a relation the tuples for which the entries in the given data field(s) are equal to the selection argument. The selection argument is loaded into A and flashed onto B . B holds a page of the relation, one tuple per column. The contents of A do not change during the entire operation. The tuples of one page are checked in a single step. The corresponding bits in the first column of the bit array are set to 1 for those tuples that have an entry equal to the selection argument. When all the pages of the relation are exhausted, the result relation is formed by retrieving the tuples that are marked in the bit array. The basic algorithm can be slightly modified to accommodate selection based on multiple arguments (data fields).

3.5 Projection

Projection retrieves only the entries in selected data fields of a relation's records. The unwanted data fields can be dropped either during the record retrieval or their transport to the host. If, however, the projected data fields do not include the primary key(s) of the relation, the result may contain multiple entries of the same tuples. Therefore, the problem is focused on deriving a method for efficiently removing the duplicates. The projected data fields are input into A , one per cycle, and compared to the contents of B . If there is a match, the tuple in A is discarded because it already exists once in B . If there is no match, the new tuple is added into the next available column of B . At the end, the target relation is stored in B . The basic algorithm can be modified to allow removal of duplicates from long relations. The response time of this algorithm depends largely on the duplication factor (df) which is defined as the ratio of the cardinality of the relation over the number of distinct values in the projected data fields.

3.6 Join and Semijoin

Both the join and semijoin of two relations can be implemented with the optical database processing unit by using the parallel version of the nested-loop algorithm. Pages from R are loaded into B while the tuples of S are shifted through A . In each step, the join attribute(s) of a new S -tuple are compared to the join attribute(s) of an entire page of R and the hits are recorded once again in the bit array. Care must be taken so that the join attribute(s) occupy corresponding pixels in the two arrays. A page remains in B until all the tuples of S pass through A . Then, a new page is loaded and the process is repeated. Semijoin is easier to implement because it retains only those tuples of R whose join attributes are equal to the join attributes of at least one tuple of S . The al-

gorithm for the full equi-join is more complicated because the contents of both relations must be monitored for possible matches and the qualifying tuples must be concatenated.

4. DISCUSSION

We have shown how an optical processing unit can perform relational database operations by taking advantage of the highly parallel nature of optics. Two-dimensional processing allows the manipulation of data in a tuple-oriented mode which is the most appropriate in a relational database environment. Table 1 summarizes the execution times and the computational complexity for several operations. T_{re} and T_{pa} are the times required to load a record in A and a page in B, respectively.

TABLE 1.

Operation	Execution Time	Complexity
Union	$\left\lceil \frac{r}{m} \right\rceil T_{pa} + \left\lceil \frac{r}{m} \right\rceil s T_{re}$	$O\left\{ \frac{r \times s}{m} \right\}$
Intersection	"	$O\left\{ \frac{r \times s}{m} \right\}$
Set Difference	"	$O\left\{ \frac{r \times s}{m} \right\}$
Selection	$\left\lceil \frac{r}{m} \right\rceil T_{pa}$	$O\left\{ \frac{r}{m} \right\}$
k-argument selection	$\left\lceil \frac{r}{m} \right\rceil [T_{pa} + (k-1) T_{re}]$	$O\left\{ \frac{r}{m} \right\}$
Projection	$r T_{re}$	$O\{r\}$
with duplicate removal	$\frac{r}{m \times df} \left(\frac{r}{2} - 1 \right) T_{re}$	$O\left\{ \frac{r^2}{m \times df} \right\}$
Semijoin	$\left\lceil \frac{r}{m} \right\rceil (T_{pa} + s T_{re})$	$O\left\{ \frac{r \times s}{m} \right\}$
Join	$\left\lceil \frac{r}{m} \right\rceil (T_{pa} + s T_{re})$	$O\left\{ \frac{r \times s}{m} \right\}$

Binary operations require $O\{(r \times s)/m\}$ computations while unary operations are concluded in $O\{r/m\}$ or $O\{r\}$ steps (except projection with removal of duplicates). Although these numbers do not indicate a significant reduction in computational steps, the simplicity of the elementary operations, combined with the high speed at which they can be executed, results in an impressive throughput.

With $T_{re} = 10^{-8}$ seconds, $T_{pa} = 10^{-6}$ seconds, and $m = 1000$, the ODPU can perform a selection operation on an 1 million tuple relation in a few milliseconds and execute the join of two relations with 1 million tuples each in about ten seconds.

This corresponds to an effective throughput on the order of 10^{11} tuple comparisons per second, far better than any electronic database machine can achieve.

The ODPU can be classified as a *single-instruction multiple-data* (SIMD) processor with the additional functionality of associative processing. Since searching is performed on the basis of content and not physical address, the requirement for maintaining large database indices is eliminated thus, reducing overhead storage and processing.

REFERENCES

1. P.B. Berra and N.B. Troullos, "Optical Techniques and Data/Knowledge Base Machines," IEEE Computer, Vol. 20, Oct. 1987, pp 59-70.
2. P.B. Berra et al., "The Impact of Optics on Data and Knowledge Base Systems," IEEE Trans. on Knowledge and Data Engineering, Vol. 1, No. 1, March 1989, pp 111-132.
3. P.B. Berra et al., "Optical Database/Knowledgebase Machines," Applied Optics, Vol. 29, No. 2, 10 Jan. 1990, pp 195-205.
4. W.E. Ross et al., "Two-Dimensional Magneto-Optic Spatial Light Modulator for Signal Processing," Optical Engineering, Vol. 22, July/Aug. 1983, pp 485-490.

Fault-Tolerant Computing on POEM

Dau-Tsuong Lu, Ting-Ting Y. Lin, Fouad E. Kiamilev, Sadik C. Esener, and Sing H. Lee

University of California, San Diego
Dept. of Electrical and Computer Engineering, R-007
La Jolla, CA 92093-0407

1. Motivation:

1.1. Need for Fault-Tolerance

Wafer scale integration (WSI) promises to realize a complete multiprocessing system on the same wafer and eliminates the expensive steps required to dice and bond. The fundamental belief is that the internal connection between chips on the same wafer are more reliable and have a smaller propagation delay than external connections¹. However, achieving a high yield has proven to be a major challenge. Rather than aiming for 100% yield, the realistic solution is to determine the defective components on the wafer and replace them with spares. Which means, the design should be tolerant to faults developed during the manufacturing process. Moreover, faults occur during system operation, be it component failure, improper operation, or environmental factors. Therefore, a mean to detect these unexpected faults and recover from them is necessary to minimize down time and unavailability. Long and periodic system downs are a luxury that cannot be afforded for computers used in critical applications. In this paper, we show that the introduction of optical interconnection techniques into a multiprocessor environment (e.g. the Programmable Optoelectronic Multiprocessor, POEM) enables efficient implementation of fault-tolerant techniques.

1.2. Fault-Tolerant Computing and POEM

Fault tolerance in a computer system is achieved through redundancy in hardware, software, information, and/or computation. Redundancy techniques need to be accompanied by fault detection and fault recovery. For fault testing, one needs to test the system frequently to detect possible faults. Upon detection, such fault should be corrected before it affects the subsequent computation. Depending on the nature of timing, faults can be classified as transient and permanent. Transient faults may be caused by environmental factors, thus retrying the failed operation from a previously known correct point could lead to successful completion. Permanent faults that are irreversible refer to physical changes in the hardware where reconfiguration of the interconnection to make use of spare hardware to carry out further computation is necessary. Therefore the major steps necessary to achieve fault tolerance are fault testing, recovery and reconfiguration. They can be accomplished by a variety of hardware or software or combination of both techniques.

The most unreliable aspect of any system design is the physical connection². The connection between VLSI circuits are now implemented in patterns of thin film wires on the chip and are subject to problems with mechanical vibration, incomplete insertion, and dirty contacts that plague the circuit at the card packaging level of the system. However, the gain in reliability is brought by the increasing complexity on the chip, bringing a corresponding increase in the cost of failure detection and correction. In highly parallel systems, the processing elements (PE) are often the lowest level field-replacement units (FRU). Systems using electronic interconnection require redundant interconnections to route around faulty processors. Often, there is a trade off between spare utilization efficiency and interconnection complexity. This trade off exists because VLSI interconnection is effectively constrained in a planar surface. Both processing logic and interconnection competes for the same silicon resource.

The POEM architecture³ consists of two optically interconnected processing planes (Fig. 1). The electronic PEs on the planes are fabricated using conventional VLSI technology and later bonded with the PLZT modulators. Each PE has three detectors and a modulator for optical I/O, in addition to electronic interconnection with the four adjacent PEs. The optoelectronic PEs and the optical interconnection among PEs are separated. All global communication channels are established through the holographic optical interconnection, using either computer generated holograms or photorefractive crystals⁴. Since the interconnections have been moved into the third dimension, it does not compete for valuable silicon area with the processing circuits. Consequently, PEs physically far apart may be interconnected just like two neighboring PEs. This freedom allows spare processors located anywhere on a processor plane (or wafer) to be utilized. As faults may be randomly distributed on the plane, regular interconnections are less effective in terms of spare utilization. We believe efficient fault-tolerant computing requires a technology that supports irregular and preferably reconfigurable interconnections. This can be achieved with holographic optical interconnection. It is with this freedom in interconnection that we consider POEM as an efficient fault-tolerant architecture.

An additional benefit for using holographic interconnection is the distributed nature of holography. The stored

interconnection information is distributed across the whole storage medium and the quality of communication does not suffer significantly from local material defects as is the case in electrical and guided wave interconnections.

In the following section, we describe our approach to fault-tolerant computing on POEM and show how we use these techniques for the POEM prototype system.

2. Implementation of Fault Tolerance on POEM:

2.1. Testing

Fault detection is critical for fault tolerance. A set of algorithms have been designed and are being implemented to test the functional correctness of the optoelectronic processor arrays. There are three modules in each optoelectronic PE that require testing: RAM, ALU, and optical I/O devices (detectors and PLZT modulators). The stuck-at fault model is used for the optical I/O devices and the ALU. For the local RAM, the coupling fault model is also considered for detecting idempotent coupling faults (a transition in one RAM cell forces the content of the neighboring cell to have a certain logic value) and inversion coupling faults (the transition causes an inversion in the content of the neighboring cell)⁵.

For the current prototype, the test programs for ALU and RAM are implemented with a Tektronix LV500 ASIC tester to verify the CMOS PEs before bonding with the PLZT modulators. During system testing, the test programs are fed to POEM by the host computer. Outputs from the processor array are fed back to the host computer via the output detector array and later used to detect and isolate faults.

Memory Test: A checkerboard test, which only assumes the stuck-at model, has been implemented to test the local memory of the PEs. In this test, alternating 1's and 0's are written into the memory cells. The memory content is read and verified. This process is repeated for a complemented checkerboard. A memory cell stuck at either 1 or 0 will put the PE into "sleep" mode, ignoring further instructions. Though simple to implement, this process does not consider the possible decoder faults. The marching test, in addition to stuck-at faults, also detects coupling faults. It scans the memory in ascending and then descending order to detect coupling between a cell in lower address and another in higher address and detects decoder errors. Besides greater fault coverage, the marching test does not assume a two-dimensional memory architecture, giving us greater flexibility in applying this test to different memory designs. Both checkerboard and marching test will erase over the current content of the memory cells. For on-line testing when the memory contains user data, the memory cells are read, inverted, written, read, inverted, written during scanning. Since this approach does not test for decoder defects, the marching test must be used during off-line testing to cover decoder faults.

ALU Test: The ALU of the POEM prototype is responsible for the AND, OR, ADD, and NOT operations. NOT takes its operand from the R register while the other operations reference a memory location as the second operand. In the case of ADD, the carry register is also used as the third operand, allowing carry propagation by serial addition from the least significant bit. The test program creates all the possible input combinations to each operation, essentially constructing the truth table entries for these binary operators. The results are transmitted to the host bit-serially through the detector array and compared to isolate any erroneous operation. If a PE has a faulty ALU, the host would have an option to remove it from active use, or continue to use it for operations not involving the particular faulty operation.

Modulator/Detector Test: The two processor planes in the prototype are used to test each other's optical I/O devices. While PEs on one plane output 0 through their modulator, PEs on the other plane compare the values they receive on the three optical detectors. The modulators are then inverted to output a 1 to check for stuck-at faults. If a detected value is not the complement of the previous, either the sending modulator or the receiving detector is stuck-at-1 or stuck-at-0. Since the three detectors share the same logic, any fault developed in this shared portion could cause fault in all three detectors. On the other hand, modulators on different PEs are independent. Therefore, we would conclude a detector fault when all three detectors uniformly detect 1-1 or 0-0 sequence. When the detectors are not in unison, we can trace the error back to the modulators on the other plane with the interconnection graph stored in the host. The roles are reversed to test optical communication in the other direction.

2.2. Host Software

The POEM host broadcasts the clock and instruction sequences to the PE array and monitors the output detector array. At this level, checkpoints can be explicitly defined in the user program or inserted by the host to schedule diagnostic tests for memory, ALU, and I/O devices. If a processor passes the tests successfully, its corresponding memory contents are output from the modulator to be stored in host memory. This is a backup so that the processor can be restored on another spare processor when this processor fails at the subsequent checkpoints.

We are going to implement the host software for the current POEM prototype.

2.3. Reconfiguration

While current POEM prototype uses a computer generated hologram (CGH) for interconnecting the two processor planes, future designs could use photorefractive crystals (PRC) for higher storage capacity and reconfigurability. CGH provides fixed interconnections. However, its fabrication is decoupled from the fabrication of the optoelectronic chips. One can therefore incorporate into the CGH design the *a priori* knowledge of the faulty PEs. Thus CGH can be used to remove faults that occur during manufacturing. PRC can be used in two ways: as a set of preprogrammed interconnections or dynamically reconfigurable interconnections. In either scheme, a number of interconnection patterns (storage capacity depends on the material used) are stored in the crystal and selected by their unique phase codes using a spatial phase modulator. After the interconnection is established, the information transfer rate will be matched by the modulators on the PEs. The amount of time required to switch to another interconnection pattern is limited by the phase code SLM⁶.

Preprogrammed Interconnections: After the interconnection patterns have been recorded onto the PRC, the patterns can be "frozen" through ion redistribution by applying an electric field. Recording is done off-line after the wafer has been tested, so only the defect-free PEs are utilized. Different interconnection patterns can be recorded to support various parallel algorithms, e.g. butterfly, perfect shuffle etc., where the number of patterns is limited only by the storage capacity of the crystal. We could also record "backup" interconnections to anticipate PE failures during operation. However, these backup patterns consume valuable storage which could be used for efficient algorithms.

Reconfigurable Interconnections: When the PRC is used for reconfigurable interconnection, the content is not frozen after the initial recording. During recording, the interconnection patterns are taken from an SLM, which is controlled by an electronic computer to achieve reconfiguration. Assuming the computer can update the SLM faster than the frame rate of the phase code SLM, the time overhead for having reconfigurability is entirely on refreshing. However, the reconfigurability also gives us the freedom to utilize spare processors located anywhere on the processor plane(or wafer). The host can keep a linked list of spare PEs that can be assigned to replace a faulty PE regardless of its location. If PE *i* is detected as faulty and replaced by PE *j*, the host simply remaps the connection for PE *i* to PE *j* in the interconnections, restores the contents of PE *i* in PE *j* and continues operation. This reconfiguration approach works equally well for any interconnection topology.

The most distinct feature of POEM is the physical separation of processing and communication. Global optical links are used when local electronic links are not possible due to defects or algorithm requirement. As a result, there is no need for redundant wiring on the wafer in order to bypass faulty PEs. Reconfiguration may change the physical path length of a particular link, but such increase does not come with increased delay due to capacitive effects. That is, POEM does not suffer performance degradation after reconfiguration. Moreover, there is no restriction on the location of a spare relative to a faulty PE. In fact, for a system using reconfigurable interconnect, the host simply maintains a list of fault-free PEs that can be assigned to active use without regard to the logical topology. Such allocation removes the need for complex algorithms to map the desired topology onto the active processors efficiently⁷.

3. **VHDL simulation:**

A VHDL (VHSIC Hardware Description Language) model of the POEM prototype has been developed at UCSD⁸. It includes behavioral models of the optical components (e.g. polarizing beam splitters, computer generated holograms, etc.) and optoelectronic devices (e.g. PLZT modulators and detectors). This model functions as both design verification and fault analysis tools. It allows us to develop parallel algorithms before running on actual hardware. It can be scaled up easily to a larger array size to accommodate larger problems. Most importantly it allows the designer to inject faults into the virtual machine to test the diagnosis algorithms and the system performance under faults. The test algorithms described previously have been implemented to verify the correctness of this VHDL model. When faults are inserted during initialization, the algorithms correctly identified the faulty PEs and removed them from active use. We will implement the host software and demonstrate system operation in the presence of faults. We also expect to perform the simulation for larger array sizes.

4. **Granularity and Scalability Considerations**

The current POEM prototype has a 2x2 array of 1-bit PEs, each with only 64 bits of RAM. Due to the relatively small size of the PEs, it would not be economical to include built-in logic to perform on-line testing. Instead, the functional testing is implemented completely in software and the entire PE is considered as the lowest

level field-replacement unit (FRU). The sequential host computer presents a reliability bottleneck. For designs that require a larger grain size⁹, the hardware overhead to accommodate built-in test logic may be small compared to external testing. This will simplify the task of locating faulty PEs. The host is still needed to perform the necessary interconnection reconfiguration. Thus fault tolerance of host needs to be addressed as well.

The testing programs are carried out by all the PEs in parallel, therefore the time required is independent of array size. A bottleneck may form at the sequential host if it is not able to verify the test results quick enough. A solution is to have the PEs on one plane verified by the other plane, and vice versa. A combination of two space-invariant interconnection pattern can be used to connect a PE in a plane with two PEs under test in the opposite plane. This approach guarantees to detect the presence of a faulty PE. It is able to detect multiple PE faults provided they are not tested by the same PE.

5. Conclusion:

In this paper, we establish POEM as an efficient fault-tolerant architecture. This is accomplished by algorithmic testing, recovery, and reconfiguration. We also demonstrate the operation of POEM in the presence of faults using VHDL model. Progress in optical interconnection technology will allow us the implementation of fault-tolerant architecture with less performance overhead and greater flexibility than purely electronic implementations. In the future, we plan to develop fault models of optoelectronic components and incorporate them into the VHDL simulation. These models will allow us to develop testing strategies for parametric faults and help uncover more faults resulting in better fault coverage of the testing algorithms.

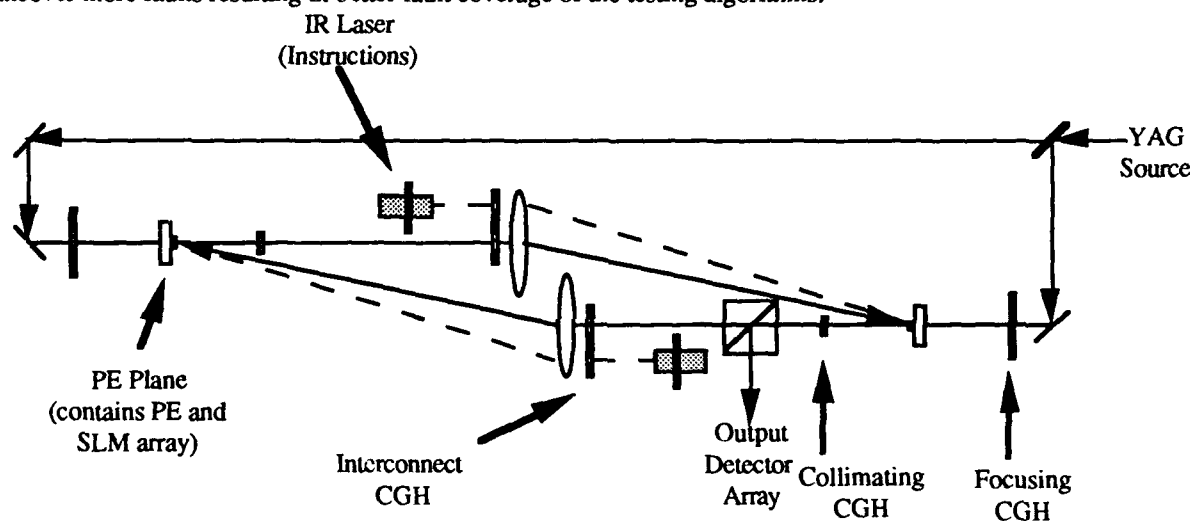


Figure 1. Setup of the POEM Prototype

- ¹ Fault-Tolerant Computing, Theory and Techniques Vol. I and II, D. K. Pradhan Editor, Prentice-Hall, Englewood Cliffs, New Jersey (1986).
- ² VLSI Engineering, chapter 1, Thomas E. Dillinger, Prentice-Hall, Englewood Cliffs, New Jersey (1988).
- ³ "Programmable Optoelectronic Multiprocessor Systems," F. E. Kiamilev, S. C. Esener, V. H. Ozguz, S. H. Lee, published in Digital Optical Computing, SPIE Vol. CR35 (1990).
- ⁴ "E-Beam Fabricated Holograms for Optical Interconnects," K. Urquhart, H. Farhoosh, and S. H. Lee, OSA Annual Meeting, Orlando, Oct. 1989.
- ⁵ "Built-in Self-Testing of Random-Access Memories", Manoj Franklin and Kewal K. Saluja, IEEE Computer, Vol. 23, #10, pp. 45-56 (1990).
- ⁶ "Application of Photorefractive Crystals to Optical Interconnection," J. E. Ford, S. H. Lee, Y. Fainman, Proceedings of SPIE, Vol. 1215, Paper 16, Jan. 1990.
- ⁷ "Fault-Tolerant Wafer-Scale Architecture for VLSI," 9th Annual Symp. Computer Architecture, pp. 190-198, (1982).
- ⁸ "VHDL for Simulation of Optoelectronic Computers," F. Kiamilev, D.-T. Lu, S. Esener, S. H. Lee, OSA Annual Meeting, Boston, Nov. 1990.
- ⁹ "Grain-Size Considerations for Programmable Optoelectronic Multiprocessor Multistage Interconnection Networks," F. Kiamilev, A. Krishnamoorthy, P. Marchand, K. S. Urquhart, S. Esener, S. H. Lee, OSA Annual Meeting, Boston, Nov. 1990.

An optoelectronic full adder using a beam scanning laser diode

Hideo Itoh, Seiji Mukai, Masahiko Mori, Masanobu Watanabe,
and Hiroyoshi Yajima

Electrotechnical Laboratory, Optical Information Section
1-1-4, Umezono, Tsukuba, Ibaraki, 305 Japan

1. Introduction

The present stage of the integration technology for optoelectronic digital computing systems is primitive, while this technology is essential for the systems to overcome electronic digital computers in the future. For example, an optical full adder is usually composed of several fundamental optical logic gates, and the number of these gates easily amounts to a prohibitive level when such adders are integrated. This difficulty will be relieved if the number of the gates composing an adder is reduced. The reduction of the gates also results in smaller size and higher operation speed of the system.

With this view, we have developed an optoelectronic digital computing system, the Beam Scanning Binary Logic system [1][2] in which all 2-input logic gates (including a half adder) operate in a single gate delay.

This paper reports an extension of our logic gates. A novel optoelectronic full adder (we call this a Beam Scanning Full Adder: BSFA), whose configurations and operational speed are almost the same as the Beam Scanning Binary Logic system[1], is proposed and the first experimental results are reported. It has the advantages of simple configuration, faster operational speed, smaller unit size, and easier monolithic integration than those conventional full adders composed of fundamental logic gates.

2. Principles

Figure 1 shows the schematic setup of a full adder of the

Beam Scanning Binary Logic system. The adder is composed of two photodetectors (D_a , D_b), and two amplifiers (A_a , A_b) on the input side of the Beam Scanning Laser Diode. D_s and D_c are photodetectors on the output side. Figure 2 shows a schematic structure of a Beam Scanning Laser Diode. The laser has two parallel p-electrodes on a rib structure and scans or switches the output beam by controlling the injection currents to the p-electrodes[3]. Figure 3 shows schematic far-field-patterns of a Beam Scanning Laser Diode for several injection currents into p-electrodes. Output detectors (D_s , D_c) are positioned at angles $\theta_4 < \theta_s < \theta_5$ and $\theta_6 < \theta_c < \theta_7$, respectively.

Figure 1 indicates an example in the case of $(X, Y, C_0) = (1, 0, 1)$. After two input data $X (=1)$ and $Y (=0)$ have been spatially encoded to $(X_a, X_b) = (1, 0)$ and $(Y_a, Y_b) = (0, 1)$, $X_a + Y_a + C_0$, $X_b + Y_b$ are sent to the input detectors D_a and D_b , respectively. The signal currents are amplified by A_a and A_b , respectively, and injected into the laser. In this example $(X_a + Y_a + C_0, X_b + Y_b) = (2, 1)$, and the output beam pattern is that of Fig. 3(e). Since the photodetector D_c receives much larger power than the D_s , the output (Z_s, Z_c) is recognized as $(0, 1)$, representing the result of Sum=0 and Carry=1.

3. Experimental results

To demonstrate full adder operations, two currents equivalent to the optical inputs were injected into the laser, and deflected output beams were detected by a Si-photodiode array whose pitch was 1 mm. The distance from the laser to the detector array was 40 mm. Figure 4 shows typical experimental results of a Beam Scanning Full Adder with all combinations of the inputs. The four lines indicate the injection currents to the right (i_a) and left (i_b) electrode, and optical outputs Z_c and Z_s , respectively. Bias currents for each electrode are 25mA. The width of the input signal pulses is 1 μ s.

4. Discussions

Both half and full adders based on the Beam Scanning Binary Logic system can be realized with the same configuration using 5 active elements and one gate delay. For comparison, a standard electronic half adder using 2-input resistor-transistor logic gates requires 3 gate-delays with 5 transistors, and a full adder requires 8 gate-delays with 27 transistors.

The compactness of our adder is due to the twofold representation of the data by the light intensity and its position.

The reduction of the number of gate-delays and active elements are desirable for high speed operation and small dimensions of the unit. Since a full adder circuit including the spatial interconnection can be easily made in $500 \times 775 \mu\text{m}^2$ area on a GaAs substrate, more than 6600 units on a $2 \times 2 \text{ inch}^2$ substrate can be integrated. This chip is equivalent to a TTL IC with about 400,000 transistors. The gate delay of the Beam Scanning Full Adder at present is around $1 \mu\text{s}$ which is limited by the delay of detectors and electronic circuits. The delay is expected to be reduced to less than 10ns by improving the circuit. This speed is much faster than the speed of a full adder of TTL IC.

5. Conclusions

A novel full adder operation using a single gate of the Beam Scanning Binary Logic system has been demonstrated. This full adder has the advantage of faster operational speed and smaller number of active elements required than one which is composed of fundamental logic gates.

References

- [1] H. Itoh, S. Mukai, M. Watanabe, M. Mori and H. Yajima, Jpn. J. Appl. Phys, 29, pp. L1268-L1269 (1990).
- [2] H. Itoh, S. Mukai, M. Watanabe, M. Mori, and H. Yajima, Technical Digest of 1990 Int'l Topical Meeting on PHOTONIC SWITCHING, Kobe Japan, 13C27, pp. 128 (1990).
- [3] S. Mukai, M. Watanabe, H. Itoh, and H. Yajima: Appl. Phys. Lett., 51 pp. 2091 (1987).

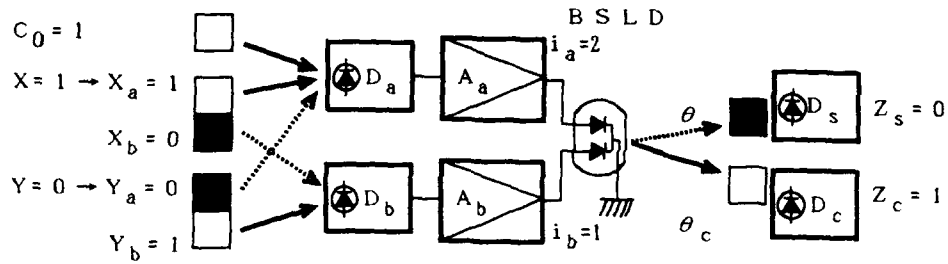


Fig.1 Schematic setup of a Beam Scanning Full Adder.

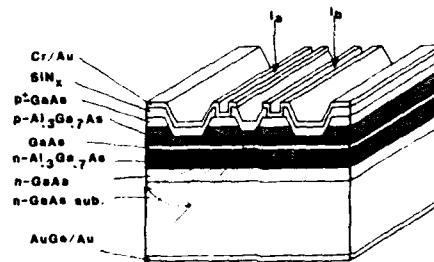


Fig.2 Schematic structure of a Beam Scanning Laser Diode.

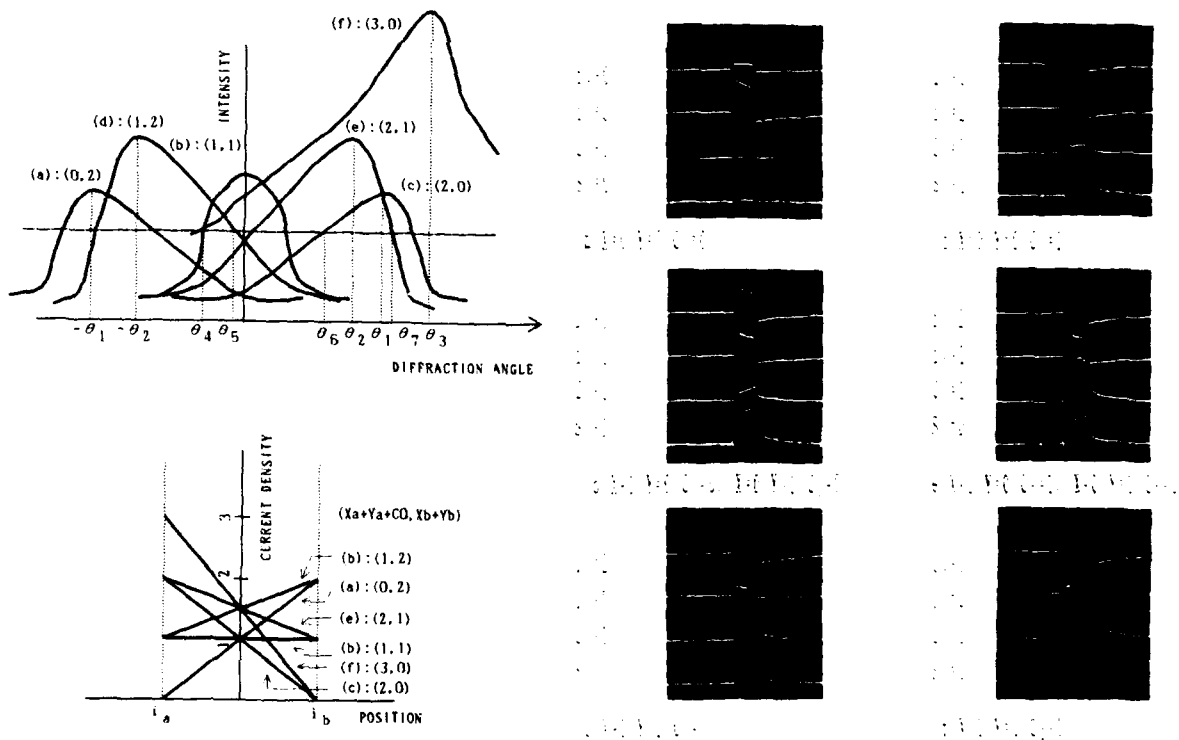


Fig.3 Schematic far-field patterns of a Beam Scanning Laser Diode for several injection currents.

Fig.4 Experimental results of full adder operations.

Implementation of a Fiber Optic Delay Line Memory

Todd J. Soukup and Vincent P. Heuring

*Optoelectronic Computing Systems Center
University of Colorado - Boulder
Campus Box 525
Boulder, CO 80309-0525*

Introduction

The objective of the Digital Optical Computer (DOC) group at the University of Colorado at Boulder is to implement a general purpose computer using the speed advantages of light.[1] The machine is being implemented using lithium niobate directional couplers as logic elements and optical fiber loops for memory. Figure 1 shows the logic functionality of the directional coupler. Terminal C, normally an electronic input, has been converted to an optical input by the addition of a sensitive detector, amplifier, and thresholder.[2] This paper describes the implementation of the fiber optic delay line memory.

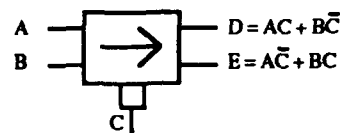


Figure 1
Logical description of the lithium niobate directional coupler.

Many of the basic problems of building a fiber optic machine have been solved including accounting for non-zero device delays [3], accommodating slight deviations in signal phases [4], verifying the synchronization of signals in the actual construction of these machines [2], and determining the behavior of a fiber combiner when used as a passive logic "OR" device [5].

The primary issue anticipated in the operation of very long fiber optic delay line memories is maintaining synchronization of the signals emerging from the fiber loop with the master clock in the face of clock drift and changes in the effective length of the loop due to thermal expansion and refractive index changes.[4]

Functional Description

The general memory subsystem operation is shown in Figure 2. The memory has three functional inputs, Write, Addr, and DEST, two functional outputs, MEM and MEMF, and two synchronization inputs, CLK and WCK. The CLK input is the master clock signal, which provides both optical power and bit synchronization to the memory. The WCK input is a word clock which provides a word synchronization reference by emitting a pulse once every word-time, 16 bits in the present case.

Both the read and write protocols require the external circuit to repeatedly send the desired data address on the Addr input line until a single "memory found" pulse is emitted on the MEMF output, indicating that the desired data item is available for reading or writing. During a memory write cycle, the Write input must be immediately raised high for the period of one word (16 bits), and the data to be written to the memory is presented at the DEST input during this period. A memory read operation responds to the MEMF signal by immediately reading the data from the MEM output.

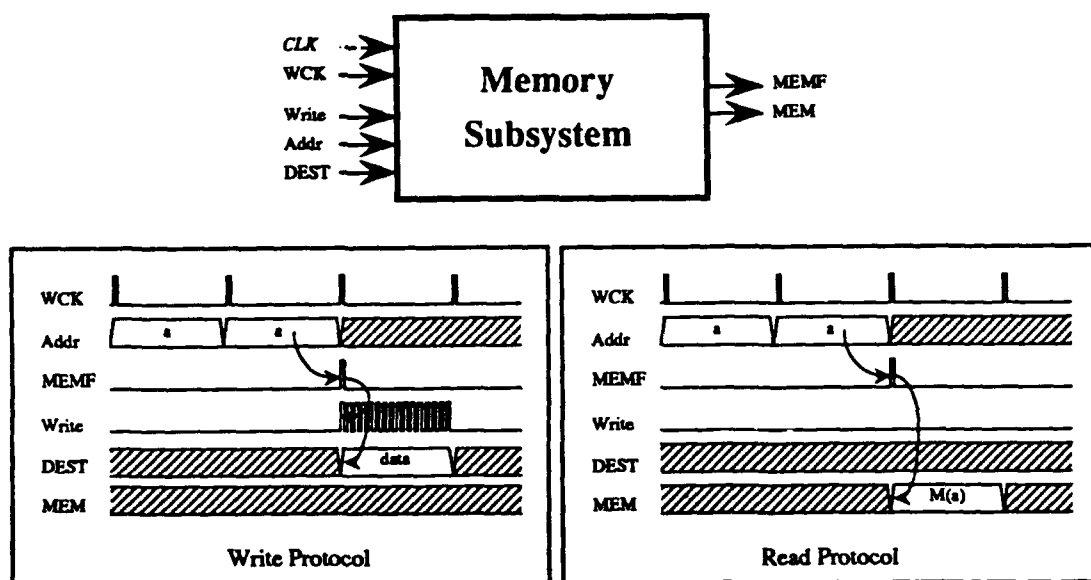


Figure 2
A functional view of the memory subsystem.

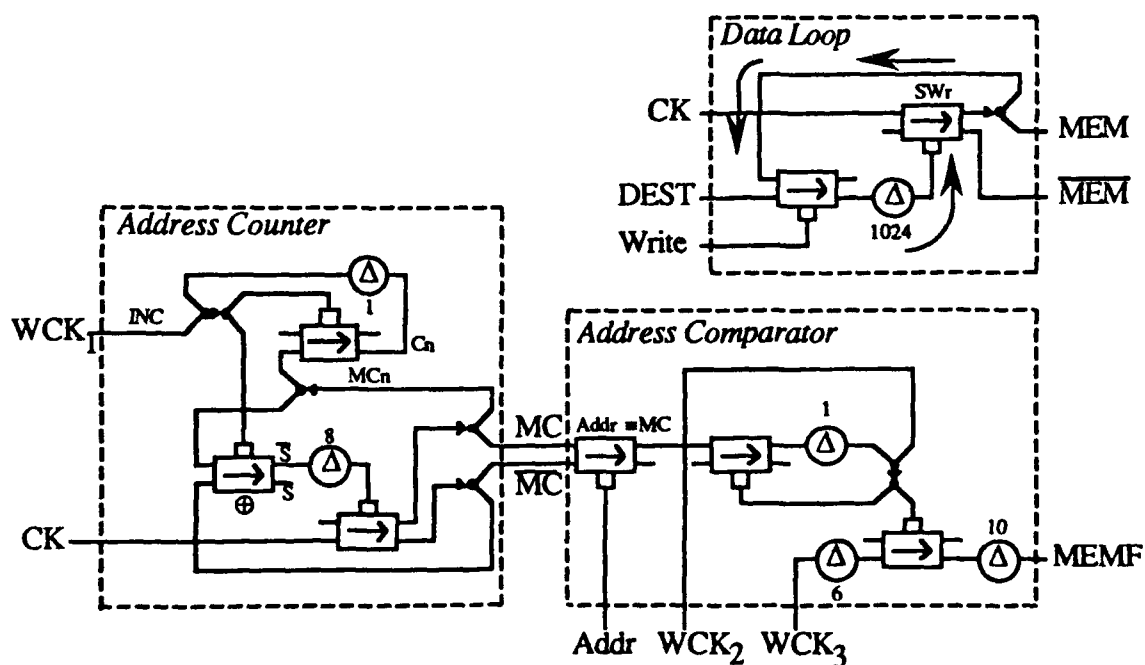


Figure 3
The 64 16-bit word optical delay line memory implementation using directional couplers.

Figure 3 shows the memory subsystem design using directional couplers, fiber interconnects, and splitter/combiners. The memory itself is divided into three subsections: the delay line memory loop which actually stores the data by continuously circulating it through the delay line, the address counter, which keeps track of which data item is available for read or

write at the data output, and the address comparator, which outputs a single pulse on MEMF when the address counter value matches the input address.

Memory Loop Design Parameters

The memory capacity of the present loop is 64 16-bit words, clocked at 50MHz. Thus, the capacity of the data loop is 1024 bits or clock periods, resulting in a total loop delay of 20.48 μ s and an average access time of 10.28 μ s. The refractive index of the fiber is approximately 1.47, resulting in a loop size of 4.2 km. We are investigating methods of reducing average access time by architectural means.

Memory System Reliability

This serial machine relies on precise temporal and spatial synchronization of signals rather than latching to ensure correct system operation.[3] Thus, in Figure 3, at switch SWr, signals must emerge from the delay line at terminal C still properly synchronized with the clock at terminal A. Since the control terminal of the optical switch has a tolerance in pulse arrival time of ± 4 ns due to pulse stretching[4], variance in the "effective" delay line length can be as high as ± 80 cm for correct system operation.

Assuming the system parameters and limits described above, a clock jitter or drift of as little as ± 0.0098 MHz would result in mis-synchronization of the system. Likewise, initial calculations indicate that the temperature may vary $\pm 24^\circ\text{C}$ without causing mis-synchronization in a bare-fiber delay line.

All three of these factors -- error in initial delay line length, frequency accuracy and stability, and temperature stability -- must not in sum cause the pulse arrival time to vary by more than ± 4 ns otherwise mis-synchronization will occur and the data in the loop will be corrupted.

The reliability of the memory should be tested by allowing it to run several days under computer monitoring. With the above synchronization effects within tolerance, the bit error rate of the memory should equal that of the switch terminal C electronics.

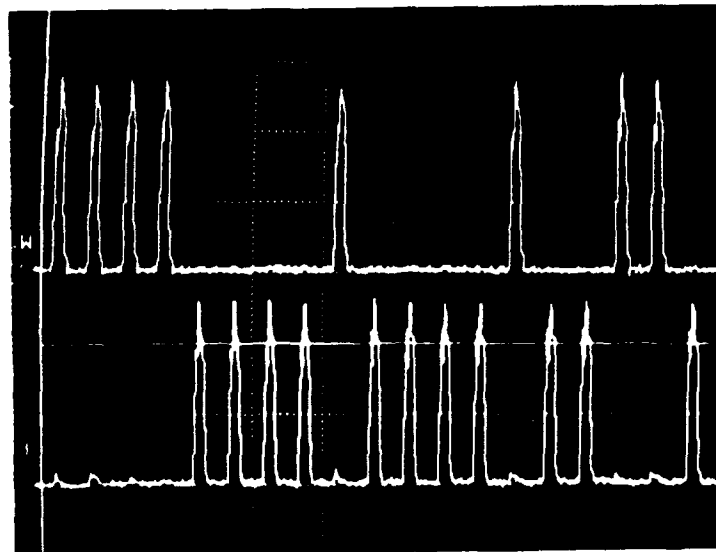


Figure 4
A typical bit stream emerging from a fiber loop.

Experimental Results

Experiments found that delay of the fiber used for the data loop changed less than $\pm 0.001\%/^{\circ}\text{C}$ and the frequency source did not vary by more than ± 0.0013 MHz thus allowing the memory to be run at room temperature without any special temperature control other than standard building heating.

The memory data loop was run under computer monitoring and control for several days without a data bit error. Room temperatures varied by about $\pm 4^{\circ}\text{C}$ over that period. Figure 4 shows a typical bit stream emerging from a fiber loop.

Conclusions

The 64 word memory subsystem described in this paper runs reliably enough for use in the bit serial optical computer we are now constructing. Under these conditions, a memory loop of considerably larger size could be built.

Acknowledgements

This work was supported in part by the National Science Foundation Engineering Research Center under Grant No. CDR 8622236, and by the Colorado Advanced Technology Institute (CATI), an agency of the State of Colorado.

References

- [1] V. P. Heuring, H. F. Jordan, and J. P. Pratt "A Bit Serial Architecture for Optical Computing," OCS Technical Report 88-01a.† Submitted to *Applied Optics*.
- [2] A. F. Benner, J. Bowman, T. Erkkila, R. J. Feuerstein, V. P. Heuring, H. F. Jordan, J. Sauer, and T. Soukup, "Digital Optical Counter Using Directional Coupler Switches," OCS Technical Report 90-31.† Submitted to *Applied Optics*.
- [3] Jonathan P. Pratt and Vincent P. Heuring, "Synchronization, Crosstalk, and Loss in Optical Systems," OCS Technical Report 90-30.† Submitted to *Applied Optics*.
- [4] David B. Sarrazin, Harry F. Jordan, and Vincent P. Heuring, "Fiber-Optic Delay Line Memory," *Applied Optics*, Vol. 29 No. 5, pp. 627-637, 1990.
- [5] Francis Ngai, "Study of Interference Effects in Passive OR-ing of Optical Signals," OCS Technical Report 90-21.†

† OCS Technical Reports are available without charge by writing to Optoelectronic Computing Systems Center, University of Colorado, Boulder, Campus Box 525 Boulder CO 80309-0525.

FANOUT ANALYSIS OF A LOW-SKEW CLOCK DISTRIBUTION NETWORK WITH OPTICAL AMPLIFIERS

C.-S. Li†, F. Tong†, and D. G. Messerschmitt‡

†IBM T. J. Watson Research Center, P. O. Box 704, Yorktown Heights, NY 10598, USA

‡Department of EECS, University of California, Berkeley, CA 94720, USA

I. INTRODUCTION

Optimization of the performance of a high-speed synchronous digital system requires tight control of the timing skew within a clock distribution network (CDN). Techniques for distributing the optical clock signals in a single-stage CDN have been suggested in [1,2] to reduce the timing skew, and the fanout is found to be much larger than that obtained from an electronic CDN. In this paper, we investigate the enhancement of fanout when optical amplifiers, which could either be semiconductor laser amplifiers [3] or fiber amplifiers [4], are introduced in a CDN.

The clock distribution architecture under consideration is shown in Fig. 1, displaying a tree with depth N , and d_j branches from each node at level j . The master clock signal is generated at the root of the distribution tree by on-off modulating a laser, while an optical receiver is connected to each leaf of the distribution tree. An optical amplifier is located at each nonleaf branch to boost the optical signals. From this architecture, the total number of optical amplifier required is $\prod_{j=1}^{N-1} d_j$ while the total number of leaves is $\prod_{j=1}^N d_j$.

The results of this paper can be summarized as follows:

- Instead of being limited by the splitting loss as in a single-stage CDN, the maximum fanout of a multi-stage CDN is limited by the noise introduced by the optical amplifiers.
- Compared to a single-stage CDN, the maximum fanout of a multi-stage CDN with optical amplifiers can be increased by a large factor while maintaining the same clock skew.
- Optimal operating conditions exist for the optical amplifiers and the photodetector to minimize the timing skew of the clock waveform.

- Fanout of the distribution network is very sensitive to the timing skew tolerance of each stage. A five-fold increase in the tolerance will decrease the maximum fanout by many orders of magnitude.

II. SKEW MODELING

In this section, a clock skew model is developed to analyze the fanout limitation of both a single-stage and a multi-stage CDN. The total clock skew, $\tau_{skew,total}$ for a sampling event at the output of a CDN equals the sum of three independent random variables:

$$\tau_{skew,total} = \tau_{dist} + \tau_{rx,s} + \tau_{rx,r} \quad (1)$$

where τ_{dist} , $\tau_{rx,s}$, and $\tau_{rx,r}$ are the distribution skew, the receiver static skew, and the receiver random skew, respectively.

Distribution skews are caused by the variation of the propagation length Δl and the refractive index Δn , both of which are due to the tolerance in fabrication in each stage of CDN. For a CDN of N stages, the τ_{dist} is given by

$$\tau_{dist} = \sum_{j=1}^N \tau_j \left(\frac{\Delta l_j}{l_j} + \frac{\Delta n_j}{n_j} \right) \quad (2)$$

where l_j , n_j , and τ_j represent, respectively, the nominal length, the nominal refractive index, and the nominal delay of the j^{th} segment of the distribution net, while Δl_j , Δn_j , and $\Delta \tau_j$ are the standard deviations of these parameters.

The propagation delay through a lightwave receiver can be modeled in a similar way to that of the gate delay in digital circuitry, in which the delay can be expressed as a linear combination of various RC time constants in the circuit [5]. The receiver static skew can then be derived as

$$\tau_{rx,s} = K_{00}\tau_F \left(\frac{\Delta\tau_F}{\tau_F} \right) + \sum_{i=1}^I \sum_{j=1}^J K_{ij} R_i C_j \left(\frac{\Delta R_i}{R_i} + \frac{\Delta C_j}{C_j} \right) \quad (3)$$

where τ_F is the transistor forward transit time, R_i and C_j are various transistor parasitic resistors and capacitors, $\Delta\tau_F$, ΔR_i , and ΔC_j are the standard deviation of these parameters, and K_{ij} is the weighting factor of the contribution to the total delay from each parasitic RC time constant. The coefficient K_{ij} can be determined from the circuit simulation of a lightwave receiver. Note that for a given input optical power, the static skew is independent of the number of stages in a CDN.

The receiver random skew at the output of the optical receiver is caused by the circuit noise as well as the noise in the clock signal, as shown in [1,2]

$$\tau_{rx,r} = \frac{t_r}{\sqrt{SNR}} \quad (4)$$

where SNR is the signal-to-noise ratio of the clock signal at the instant of sampling and t_r is the rise time of the clock. The clock rise time is related to the electrical bandwidth B_e through $B_e = b_f/t_r$, where b_f is a waveshape-dependent factor ranging from 0.1 to 1.

III. SIMULATION RESULTS

(A) Single-Stage Clock Distribution

In a single-stage CDN, the maximum fanout is determined by the minimum required optical power arriving at the receiver, which is set by the maximum allowable clock skew $\sigma_{skew,max}$. Assuming a transimpedance preamplifier is used in the receiver design, we can use a technique from [6] to show that the minimum required power at the receiver, $P_{rx,min}$, has to satisfy the following quadratic equation:

$$AP_{rx,min}^2 + BP_{rx,min} + D = 0 \quad (5)$$

where

$$A = B_e \frac{\sigma_{skew,max}^2}{b_f^2} \left(\frac{r-1}{r+1} \right)^2 \left(\frac{2Me}{h\nu} \right)^2 \quad (6a)$$

$$B = \frac{2e^2 M^2 F}{h\nu} \quad (6b)$$

$$D = \left(\left(\frac{1}{R_F} + \frac{1}{R} \right)^2 + \frac{4}{3} \pi^2 B_e^2 C^2 \right) V_A^2 + 4kT \left(\frac{1}{R_F} + \frac{1}{R} \right) + I_A^2 \quad (6c)$$

The parameters in (6a)-(6c) are defined as:

M	avalanche gain of an APD
F	excess noise factor of an APD
e	electron charge
$h\nu$	optical energy
kT	thermal energy
r	extinction ratio
R_F	feedback resistance
R	receiver input resistance
C	receiver input capacitance
V_A	amplifier noise voltage
I_A	amplifier noise current

The maximum fanout in a single-stage CDN thus equals

$$d_{max} = \frac{P_{out}L}{P_{rx,min}} \quad (7)$$

where P_{out} is the average output optical power, and L is the insertion loss of each branch. The maximum fanout is plotted in Fig. 2 as a function of clock rise-time for various maximum allowable skew. A shorter rise-time requires larger electrical bandwidth, and thus allows more noise at the receiver. On the other hand, a longer rise-time introduces more ambiguity in the decision region and is more vulnerable to noises. These conditions yield an optimal rise-time for the clock signal in which the fanout is maximized. Figure 3 shows the maximum fanout as a function of the avalanche gain for various ionization factors. Note that there is an avalanche gain that produces maximum fanout.

(B) Multi-Stage Clock Distribution

The total received power includes both the signal power P_s and the spontaneous emission power (noise power) P_{sp} . The signal power at the receiver is given by

$$P_s = P_{out} \frac{L_0}{d_0} \prod_{j=1}^N \frac{\eta_{in} G_j \eta_{out} L_j}{d_j} \quad (8)$$

where L_j is the loss of the j^{th} distribution stage in addition to the splitting loss $1/d_j$. η_{in} and η_{out} are the amplifier input and output coupling efficiency, respectively. The amplifier gain G_j at the j^{th} stage depends on the input power, $P_{in,j}$ [8]:

$$G_j = \frac{P_{sat,j}}{P_{in,j}} \ln \frac{G_{0,j}}{G_j} + 1 \quad (9)$$

where $P_{sat,j}$ is the internal saturation power and $G_{0,j}$ is the small signal gain of the amplifier. The

total spontaneous emission power generated by all optical amplifiers is

$$P_{sp} = \sum_{i=1}^N \frac{P_{sp,i}}{G_i \eta_{in}} \prod_{j=i}^N \frac{L_j \eta_{in} G_j \eta_{out}}{d_j} \quad (10)$$

where $P_{sp,i}$ is the spontaneous emission generated at the i^{th} optical amplifier, given by [3]:

$$P_{sp,i} = (G_i - 1) N_{sp} B_o h \nu \quad (11)$$

In this equation, N_{sp} is the spontaneous emission factor of the optical amplifier and B_o is the optical bandwidth. Note that the accumulated amplifier noise can be reduced by limiting the optical bandwidth of each amplifier. Inserting Eq.(11) into Eq.(10), the total spontaneous power becomes

$$P_{sp} = N_{sp} B_o h \nu \sum_{i=1}^N \left(1 - \frac{1}{G_i}\right) \frac{1}{\eta_{in}} \prod_{j=i}^N \frac{L_j \eta_{in} G_j \eta_{out}}{d_j} \quad (12)$$

The spontaneous emission noise accumulated by the optical amplifiers in cascade eventually saturate the gain of the following amplifier stage. In general, the small signal gain of the amplifier at the j^{th} stage, $G_{o,j}$, and the splitting d_j can be optimized to obtain a maximum fanout. However, a global optimization of all of these parameters is not possible due to the number of parameters involved. In this paper, we only consider a homogeneous case in which the $G_{o,j}$ and d_j are identical for all j levels. In our simulations, η_{in} and η_{out} equal 0.25 and 0.31, respectively. L_j , $G_{o,j}$, and P_{sat} are 2dB, 25dB, and 8.59dBm, respectively.

Figure 4 shows the maximum fanout as a function of rise time of the clock signal for various values of distribution skew. The maximum fanout is much larger than that obtained from the single-stage (cf. Fig.2). In contrast to the single-stage case, the fanout is very sensitive to the distribution skew. A five-fold increase in the distribution skew reduces the maximum fanout by many orders of magnitude. The optical amplifier approach is no longer attractive if the distribution skew reaches 20% of the maximum allowable skew. The optical bandwidth at each amplifier stage can significantly affect the fanout, as shown in Fig.5. Smaller optical bandwidth is thus desirable to prevent the accumulation of spontaneous emission noise. Maximum fanout vs. avalanche gain is shown in Fig. 6. As shown in this figure, an APD is inferior in performance as compared to a PIN detector in a multi-stage CDN since the spontaneous noise is

amplified in an APD ($M^2F > 1$) but not in a PIN ($M = 1, F = 1$). The maximum fanout can vary significantly with the amplifier gain, as shown in Fig.7. In order for the maximum fanout of a multi-stage CDN to surpass that of a single-stage CDN, the small-signal gain of an amplifier has to be at least 15 dB to compensate the coupling loss through the optical amplifier as well as the propagation loss of each stage. At higher amplifier gain, gain saturation effect begins to dominate and the maximum fanout levels off. An optimal small-signal gain thus exists to maximize the fanout.

REFERENCES

- [1] Khalil, R. L., "Clock Skew Analysis for Si and GaAs Receivers in Optical Clock Distribution Systems," *SPIE*, Vol. 1178, Optical Interconnects in the Computer Environment, pp. 171-176, 1989.
- [2] Khalil, R. L., L. R. McAdams, and J. W. Goodman, "Optical Clock Distribution for High-Speed Computers," *SPIE*, Vol. 991, Fiber Optic Datacom and Computer Networks, pp. 32-41, 1988.
- [3] For example, Olsson, N. A., "Lightwave Systems with Optical Amplifiers," *IEEE J. of Lightwave Technology*, Vol. 7, No. 7, pp. 1071-1082, 1989.
- [4] For example, Dakss, M. L. and W. J. Miniscalco, "Fundamental Limits on Nd^{3+} -Doped Fiber Amplifier Performance at 1.3 μm ," *Photonics Technology Letters*, Vol. 2, No. 9, pp. 647-649, Sept. 1990.
- [5] Fang, W., "Accurate Analytical Delay Expressions for ECL and CML Circuits and Their Applications to Optimizing High-Speed Bipolar Circuits," *IEEE J. of Solid-State Circuits*, Vol. 25, No. 2, pp. 572-583, April 1990.
- [6] Gowar, J., "Chapter 14: The Receiver Amplifier," in *Optical Communication Systems*, Prentice-Hall: NJ, 1984.
- [7] Gowar, J., "Chapter 13: Avalanche Photodetectors," in *Optical Communication Systems*, Prentice-Hall: NJ, 1984.
- [8] Saitoh, T. et al, "Pulse Energy Gain Saturation in Subpicosecond and Picosecond Pulse Amplification by a Traveling-Wave Semiconductor Laser Amplifier," *Photonics Technology Letters*, Vol. 1, No. 10, pp.297-299, Oct. 1989.

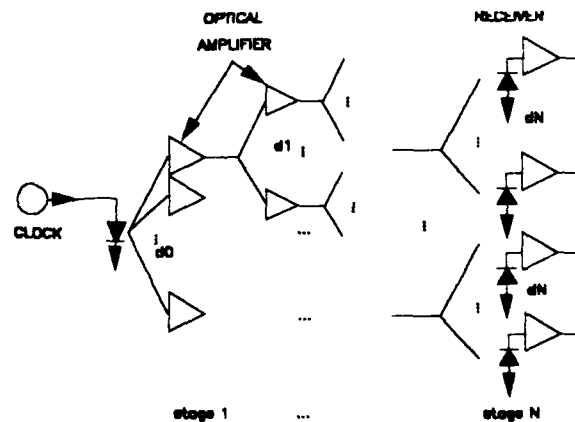


Fig. 1. The architecture of a multi-stage clock distribution network with depth N and d_j branches from each node.

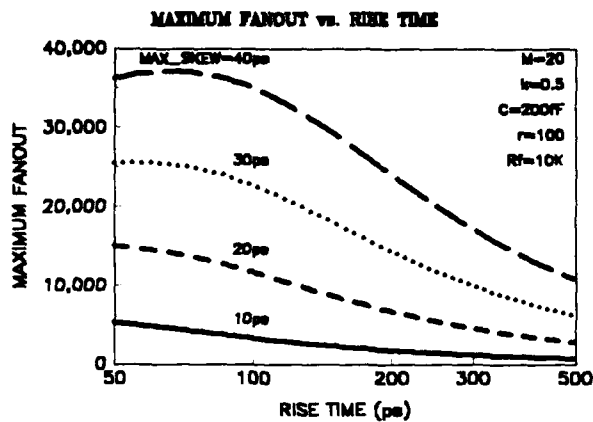


Fig. 2. Maximum fanout is plotted as a function of the rise time of the clock waveform.

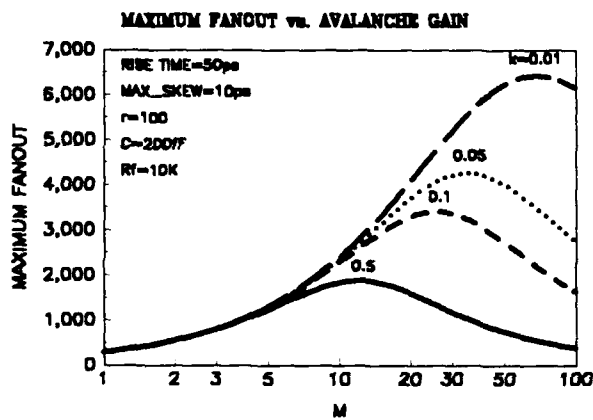


Fig. 3. Maximum fanout is plotted as a function of the avalanche gain of an APD for ionization factor equals 0.01, 0.05, 0.1, and 0.5. The ionization factor k determines the relationship between F and M [7].

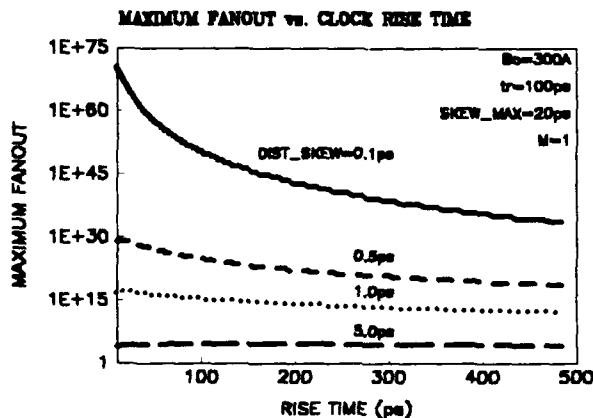


Fig. 4. Maximum fanout is plotted as a function of the rise time for distribution skew equals to 0.1, 0.5, 1, and 5ps stage while the maximum allowable skew is set at 20ps.

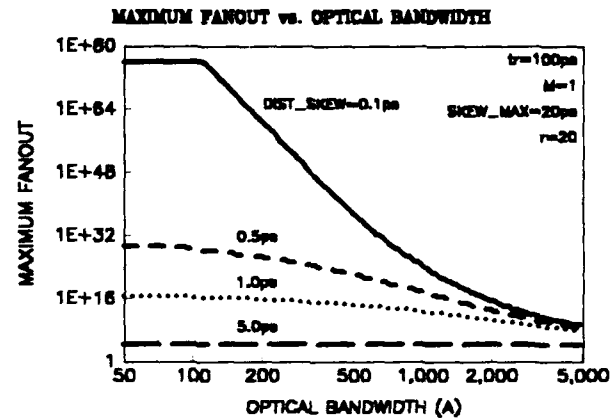


Fig. 5. Maximum fanout is plotted as a function of the optical bandwidth of the amplifier. The operating condition is similar to that of Fig. 4.

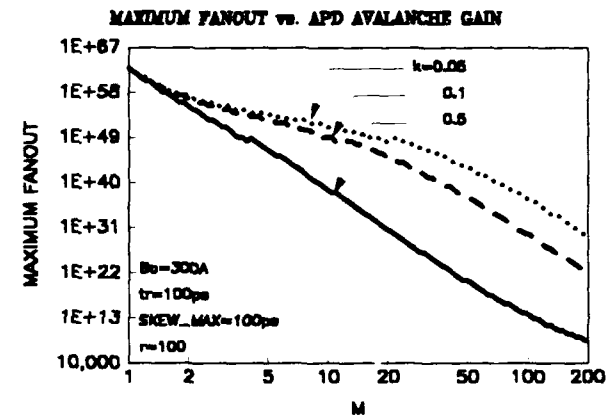


Fig. 6. Maximum fanout is plotted as a function of the avalanche gain M for various ionization factor.

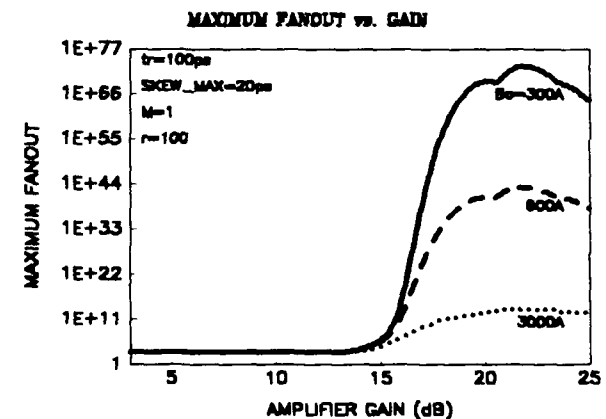


Fig. 7. Maximum fanout is plotted as a function of the optical amplifier small signal gain G_o .

Reconfigurable Interconnects Using Computer Generated Holograms and Spatial Light Modulators

James E. Morris, Michael R. Feldman
Department of Electrical Engineering
University of North Carolina at Charlotte
Charlotte, N.C. 28223

An efficient method of implementing programmable optical interconnects is needed for communication between processors in optically interconnected VLSI processor arrays[1,2], between optical logic gates in optical computers[3], and between chips, modules and boards in general purpose VLSI systems[4-6].

Previously proposed methods of implementing programmable connections with Spatial Light Modulators (SLM's) suffer from high power dissipation and/or long reconfiguration times. The use of SLM's to directly encode a hologram would result in a low efficiency hologram with a large switching energy. This poor performance is due to the relatively low spatial frequencies of SLM's when compared to the spatial frequencies needed for high performance holograms. (A hologram implementing a single connection requires a large array of pixels with pixel dimensions on the order of a wavelength.) Another approach is to implement all possible connections with fixed media (e.g. lenses, fiber optic connectors, or holograms) and use amplitude SLM's to mask off the undesired connections[4-6]. This approach suffers from high power dissipation and large area requirements. To form a crossbar switch between N transmitters and N receivers with this method, requires $O(N^2)$ power and $O(N^2)$ modulators.

A new method of implementing such a programmable interconnect system (that does not suffer from the above mentioned limitations) has recently been proposed [8]. This method involves combining high frequency fixed Computer Generated Holograms(CGH's)[9] with a small number of binary (phase or amplitude) light modulators (e.g., GaAs MQW devices[10], PLZT modulators[11], liquid crystal devices, deformable mirror devices[12]). By activating different subsets of the SLM's, the overall complex transmittance of the SLM-CGH structure is changed (effectively producing a different hologram) resulting in a different connection pattern. The modulators provide the switching capability and the CGH provides the high spatial frequency necessary to produce highly efficient, high performance connections.

This approach has fundamental advantages over previously proposed methods. Although high SBWP holograms are needed to implement a single connection, different connections can be implemented by changing only a small number of pixels. Thus, not all of the pixels in the hologram need to be programmable. With the proposed approach modulators are used to allow programmability of only a small number of critical pixels. The rest of the pixels are implemented with a fixed CGH. Since only a small number of modulators with large pixel dimensions are required, the switching energy, reconfiguration time, power dissipation and SLM cost complexity are greatly reduced.

It will be shown that with this combined CGH-SLM programmable interconnect method an $N \times N$ optical crossbar switch can be implemented with a constant power dissipation ($O(1)$) per node, independent of the number of nodes. The number of modulators per transmitting node can potentially be reduced to $O(\log N)$.

Two categories of interconnections will be investigated: (1) externally controlled and (2) locally controlled programmable connections. In externally controlled connections the programming of the interconnects is performed by an external controller. Examples include fiber

optic crossbar switch networks[8], and crossover and butterfly networks for interconnecting logic gates [3]. With locally controlled programmable connections, the particular connections to be implemented are determined locally by each processor. Examples include hypercube connection networks and connections between processors in computers based on Parallel Random Access Machine (P-RAM) computer models.

Figures 1(a) and 1(b) illustrate the proposed method for achieving programmable connections with external control and local control, respectively. In Fig. 1(a), the light from a transmitter (i.e., a laser, light modulator, or L.E.D.) illuminates a holographic structure consisting of two layers - a binary SLM layer followed by a thin phase CGH layer. An external controller is used to address the SLM plane. By changing the transmittance of the SLM's, a different effective holographic structure can be produced resulting in different connection patterns. In a locally controlled interconnect scheme (Fig. 1(b)), light modulators are incorporated into the individual processors in the input plane. Each processor (or transmitting node) activates the appropriate modulators to produce the desired connection pattern. Activation of different combinations of modulators will produce different wavefronts in the plane immediately in front of the hologram that will combine with the CGH to elicit different connection patterns. Thus, the modulators associated with a given node in Fig. 1(b) act as both a signal transmitter (emitting a logical 1 or a logical 0 signal) and as part of the programmable connection network (determining which detector is to receive the signal).

Note that if collimating optics are employed or if the divergence angles of the beams are small then the two holographic systems are equivalent in the sense that they will produce the same connection patterns for the same modulator transmittances and CGH's. The major difference between the two systems is that in Fig. 1 only 1 optical signal transmitter need be incorporated into each transmitting node, but an additional SLM plane is required. In Fig. 1(b) although no external SLM plane is needed, several modulators must be incorporated into each transmitting node.

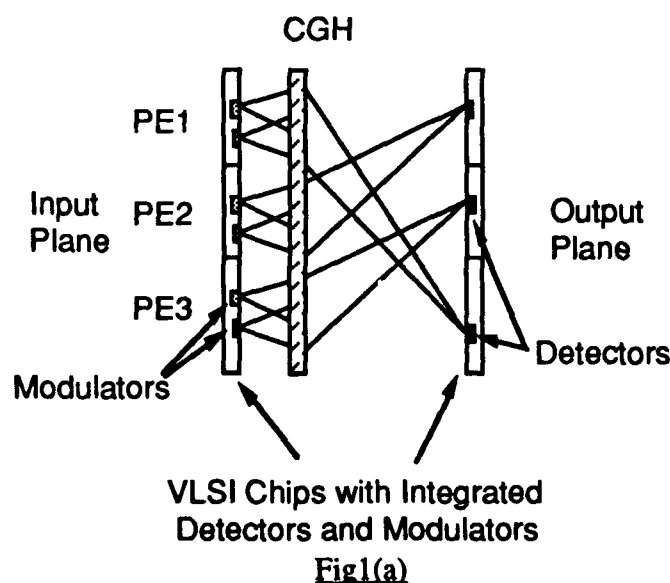


Fig1(a)

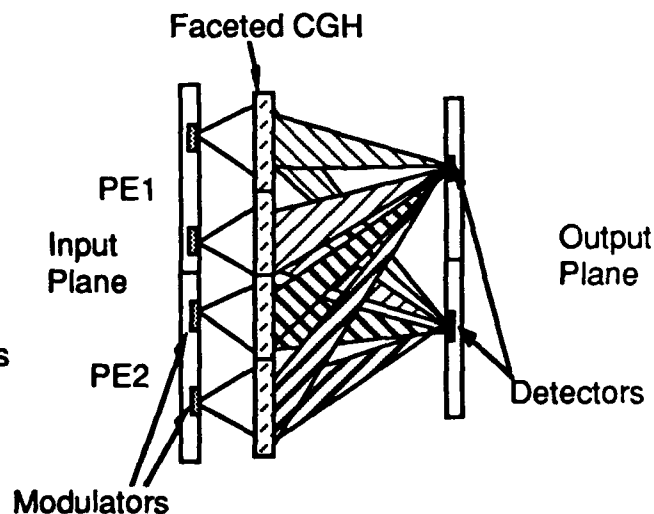


Fig1(b)

Figure 1.(a) Combined SLM-CGH programmable interconnect system with external control. The combined effects of the modulators in the SLM plane and of the CGH serve to connect each

transmitter to a particular detector. Different connection patterns are selected by changing the subset of modulators in the SLM plane that are activated. One-to-many and many-to-one connections can also be implemented. (b) Combined SLM-CGH programmable interconnect system with local control. Light from an external laser source is focussed onto all of the binary (phase or amplitude) optical modulators in the input plane. The input plane is divided into transmitting nodes or Processing Elements (PE's). A PE in the input plane can be connected to any one of the detectors in the output plane by activating the appropriate subset of modulators within the PE. In general, both planes will act as both input and output planes, or a reflective version will be employed.

Faceted CGH Approach

Initial investigations have led to the discovery of a particular combined CGH-SLM method, utilizing a faceted CGH, illustrated in Fig. 2 [8]. The CGH is divided into facets so that light passing through each modulator illuminates a distinct CGH facet. With this method, each transmitting node can be connected to any 1 of N receiving nodes with 0 dB insertion loss and $O(N)$ modulators per transmitting node.

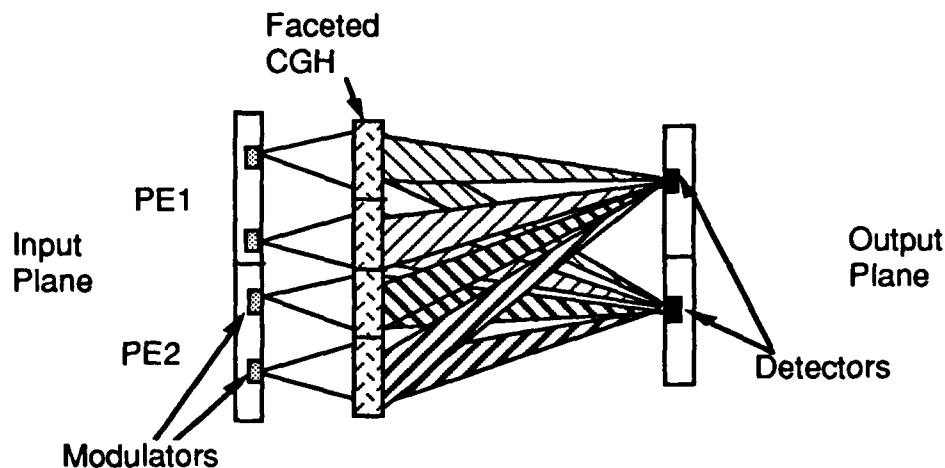


Fig. 2. Combined CGH-SLM programmable interconnect system with a faceted CGH (with local control). Each modulator illuminates a different CGH facet.

Consider first the particular case of connecting 1 transmitting node to any of 4 receiving nodes by incorporating 4 binary phase modulators into each transmitting node in a locally controlled programmable connection system. Each modulator illuminates a separate CGH facet. Each CGH facet splits the incident wavefront into 4 beams, focussing each beam onto a different detector. Hence, 4 beams are focussed onto each detector with each beam incident on a particular detector originating from a distinct modulator. The CGH facets are designed so that when no modulators are activated (i.e. they have the same phase transmittance), the 4 beams add coherently at detector #1 and incoherently at all the other detectors, illustrated in. Therefore, if 1 unit of light power illuminates each modulator, detector #1 will receive 4 units of optical power while the other detectors will receive no incident optical power. Activation of the first two modulators will change the phase delay of the beams resulting in coherent addition at detector #2 and incoherent addition at all other detectors. By choosing to activate different combinations of 2 detectors all of the light can be effectively focussed onto any one of the 4 receiving detectors. If only 1 modulator is activated,

each detector receives only 1/4 of the incident optical power. This mode can be used for, depending on the threshold settings of the detectors, either (a) transmitting a logical 0 to all detectors or (b) broadcasting a signal to all detectors. (If a broadcast mode is chosen, then a fifth modulator would be needed to allow a logical 0 to be sent to all detectors.)

In this manner the faceted CGH-SLM method can be used to implement a $N \times N$ crossbar switch, each of N communication ports can communicate simultaneously with any other port with (1) a total power dissipation of $O(N)$, (2) total number of modulators of $O(N^2)$, and (3) $O(1)$ communication time steps and negligible crosstalk. (This should be compared to a conventional amplitude SLM crossbar that requires $O(N^2)$ power dissipation and $O(N^2)$ modulators).

More complicated arrangements have been developed that can achieve this same function with less modulators. For example, a multipass system requires $O(N^{1/2})$ modulators per node. Both computer simulations and theoretical analysis will be presented.

References

1. M. R. Feldman, C. C. Guest, T. J. Drabik and S. C. Esener, "Comparison Between Optical and Electrical Interconnects for Fine Grain Processor Arrays based on Interconnect Density Capabilities, *Appl. Opt.*, 28, pp. 3820-3829, 1989.
2. F. Kiamilev, S. C. Esener, R. Paturi, Y. Fainman, P. Mercier, C. C. Guest and S. H. Lee, "Programmable Optoelectronic Multiprocessors and Their Comparison with Symbolic Substitution for Digital Optical Computing," *Opt. Eng.*, 28, pp. 396-409, 1989.
3. J. Jahns, M. J. Murdocca, "Crossover networks and their Optical Implementation," *Appl. Opt.*, 27, pp. 3155-3160, 1988.
4. J. Bristow, A. Guha, C. Sullivan and A. Husain, "Optical Implementations of Interconnection Networks for Massively Parallel Architectures," Optical Computing, 1989 Technical Digest Series, 9, (Optical Society of America, Washington, D.C.) pp. 118-121, 1989.
5. J. W. Goodman, F. I. Leonberger, S. Y. Kung and R. A. Athale, "Optical Interconnections for VLSI Systems," *Proc. IEEE*, 72, pp.850-866, 1984.
6. M. R. Feldman, S. C. Esener, C. C. Guest and S. H. Lee, "Comparison between Optical and Electrical Interconnects Based on Power and Speed Considerations," *Appl. Opt.*, 27, pp. 1742-1751, 1988.
7. C. Warde and A. D. Fisher, "Spatial Light modulators: Applications and Functional Capabilities," in *Optical Signal Processing*, J. L. Horner, ed., Academic Press, Inc., p. 486, 1987.
8. M. R. Feldman, "Programmable Interconnects Using Combined Effects of Computer Generated Holograms and Optoelectronic Modulators," to be submitted to *Applied Optics*.
9. W. B. Veldkamp, "Binary Optics: Out of the Laboratory, Into the Market," OSA Annual Meeting, 1989 Technical Digest Series, 18 (Optical Society of America, Washington, D.C.), p.73, 1989.
10. D. A. B. Miller, D. S. Chemla, T. C. Damen, T. H. Wood, C. A. Burrus, Jr., A. C. Gossard and W. Wiegmann, "The Quantum Well Self-Electrooptic Effect Device: Optoelectronic Bistability and Oscillation, and Self-Linearized Modulation," *IEEE Journal of Quantum Electronics*, QE-21, pp. 1462-1476, 1985.
11. S. H. Lee, S. C. Esener, M. A. Title and T. J. Drabik, "Two-Dimensional Silicon/PLZT Spatial Light Modulators: Design Considerations and Technology," *Optical Engineering*, 25, pp.250-260, 1986.
12. J. M. Florence, M. O. Poley, "Real Time Deformable Mirror Joint Transform Correlator," OSA Annual Meeting, 1989 Technical Digest Series, 18 (Optical Society of America, Washington, D.C.), p.177, 1989.

Demonstration of Colored Optical Interconnects and Implementation
of a 2×2 Optical Crossbar Switch with Bistable Diode Laser Amplifiers

Zeqi Pan and M. Dagenais

Department of Electrical Engineering

and

Joint Program for Advanced Electronic Materials

University of Maryland

College Park, MD 20742

Bistable diode laser amplifiers have the lowest switching energy of any optical switching devices. Single and double beam usable gains in excess of 50 and 250 respectively have been demonstrated, which implies very large fan out capabilities. These devices have also been shown to be cascable and can be operated at rates approaching 1 Gbit/s. We have used four of these devices to implement a 2×2 generalized nonblocking, optical crossbar switch. Any of the inputs can be connected with any of the outputs. In addition, any of the input channels can be recovered from any particular output channel using the concept of colored optical interconnects.

A 2×2 array of bistable diode laser amplifiers was configured with sources and detectors to form an optical vector-matrix multiplier. The light from two input sources operated at slightly different wavelengths was spread on two different rows of the matrix. The result of the vector-matrix multiplication is obtained by collecting the light from the elements representing a column of the matrix onto two detectors. There are no restrictions on the number of ones and zeros in the matrix. For instance, if a row of the matrix is composed of only ones, this corresponds to broadcasting. If the matrix has a column full of ones, this corresponds to multiplexing since different wavelength sources are used. Any intermediate situation is possible. Each diode laser on a particular row of the two-dimensional array has a similar frequency that matches the corresponding diode frequency in the input plane. On the detector side, the information can be demultiplexed and regenerated by a 1×2 array of bistable diode laser amplifiers, each operating at a wavelength corresponding

to one of the input wavelengths. We use here the wavelength rejection capability of the bistable diode laser amplifier when the input wavelength is not tuned near a Fabry-Perot transmission peak. We refer to this concept as high capacity communication using colored optical interconnects. The experimental set-up used to demonstrate the concept of optical interconnect is shown in Fig. 1. Some of the results are shown in Fig. 2. These results were obtained at a slow modulation rate, since a mechanical chopper was used to modulate the beams. We are presently extending these results at rates in excess of 100 Mbit/s. We have already obtained -41dbm sensitivity at 140 Mbit/s when the bistable diode laser was operated as a high sensitivity receiver.

The results presented in this work indicate the interesting potential of this technology. If each of the bistable diode laser in an $n \times n$ matrix is made to operate at rates of order $R=1$ Gbit/s, very large overall throughput ($n \times n \times R$) can be obtained with relatively modest size matrices. In addition, the matrix can be optically addressed very rapidly. Further work is required to demonstrate the full potential of optical crossbar switches based on bistable diode laser amplifiers.

We are pleased to acknowledge the financial support from NSF (contract number: ECS-8818797) and DARPA (contract number: DAAH01-89-C-0067).

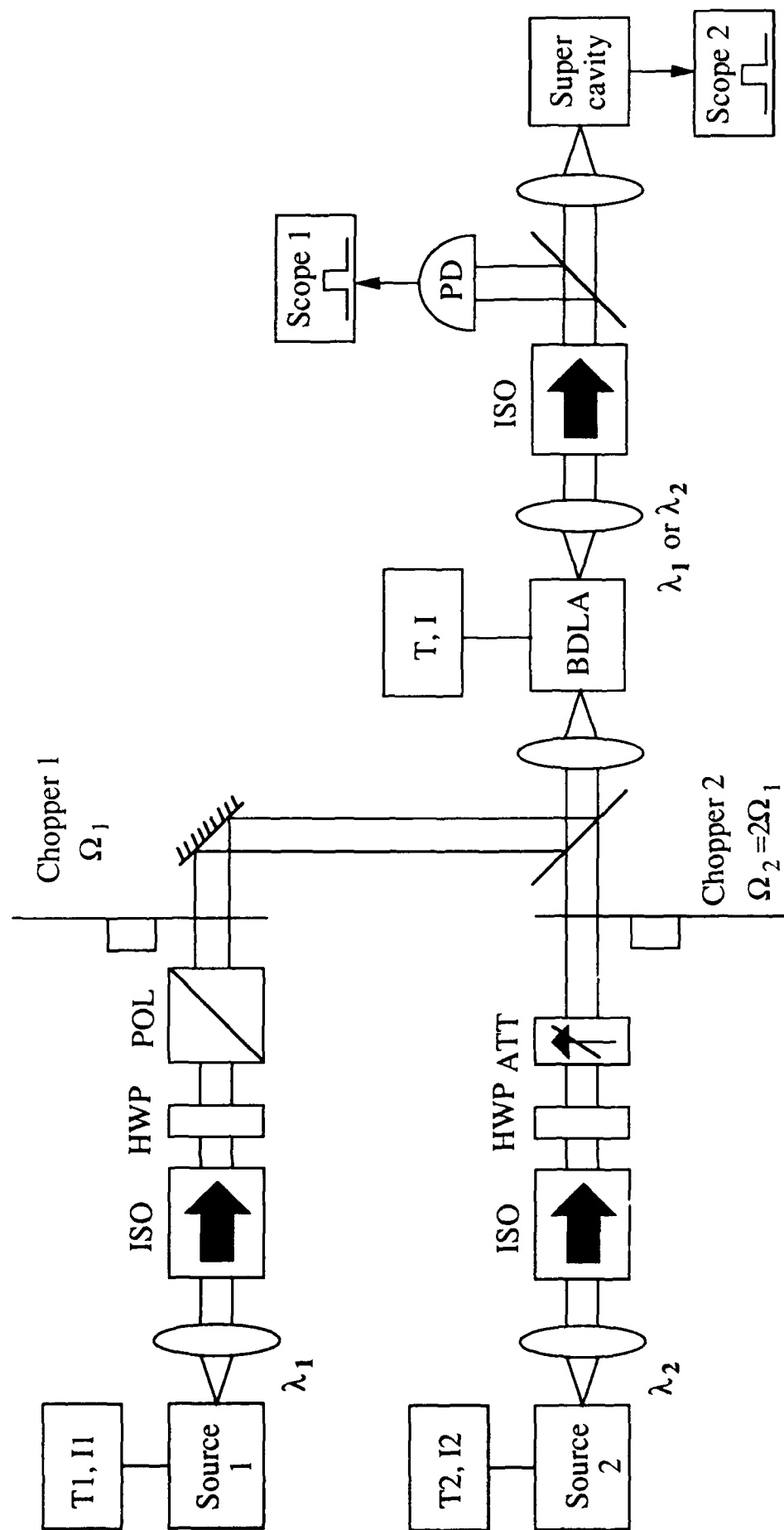
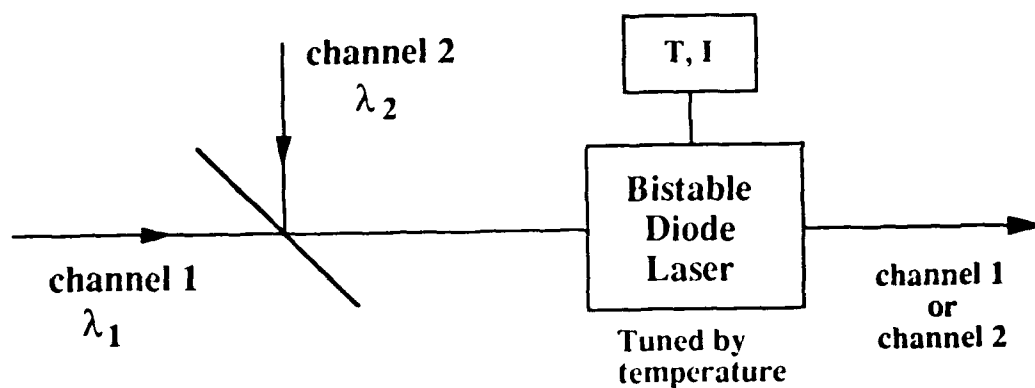
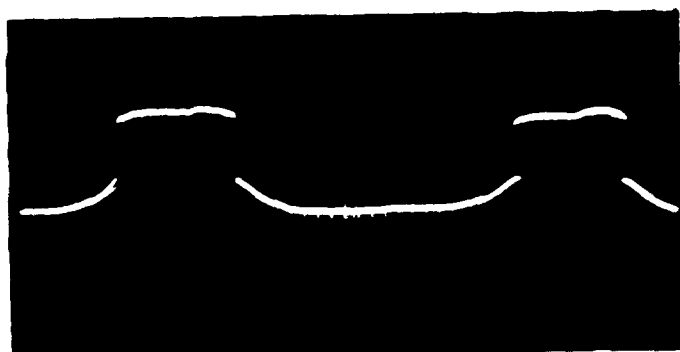


Figure 1

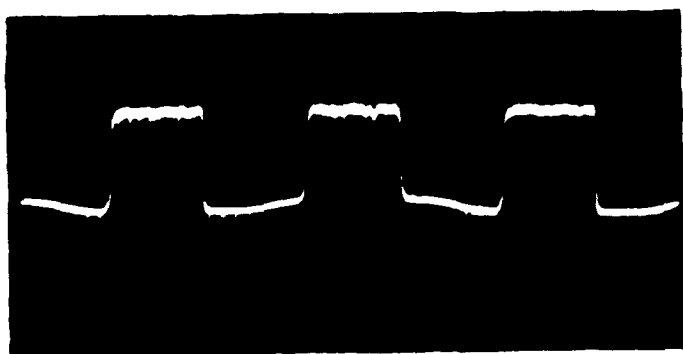
setup to demonstrate the concept of colored optical interconnects



$$\Delta\lambda = \lambda_2 - \lambda_1 = 0.45 \text{ \AA} (19 \text{ GHz})$$



output channel 1
with BDLA tuned close to λ_1



output channel 2
with BDLA tuned close to λ_2

Figure 2

Compact Crossbar Switch For Optical Interconnects

Freddie Lin, Eva M. Strzelecki and William Liu
Physical Optics Corporation
20600 Gramercy Place, Suite 101
Torrance, California 90501

1. Introduction

Optical crossbar switches are used in a variety of applications: in optical computing, optical communications, and optical interconnects in computers. High speed optical crossbars have been demonstrated for the use in communications such as waveguide electro-optic switches in LiNbO_3 [1], semiconductor quantum well modulators [2]. For applications in optical computing it is important to have very large switching arrays to utilize the massively parallel capability of optical signal processing. This has been achieved using spatial light modulators (SLM) which are available in large arrays, such as liquid crystal TVs and ferroelectric liquid crystal devices [3], and high speed, such as PLZT [4] or quantum well modulators [2]. Many systems based on SLMs utilize the vector-matrix multiplication configuration to realize crossbar networks, linear algebra operations, iterative vector-matrix multiplication, and optical neural networks. Crossbars based on this configuration, though suffer from fan-out losses, are very versatile, offer broadcasting capability needed in optical interconnects, and can easily form large array sizes. Most SLM based systems, which use bulk optics, lenslet arrays, and fiber optic couplers [5], are bulky and require tedious alignment. In this paper we describe a compact vector-matrix multiplier in which waveguides with arrays of grating couplers are used to distribute and collect light signals.

2. Compact Vector-Matrix Multiplier

The crossbar switch is shown schematically in Figure 1. It is composed of three planes. The first plane converts the input optical guided signals of a vector A into collimated beams of equal intensity. Uniform optical fan-out is provided by grating couplers with graded coupling efficiency. These signals are "multiplied" by a matrix, M , which is realized by a conventional 2-D spatial light modulator located in the second plane. The beams exiting the SLM are collected in the third plane, converted into guided waves, and combined (i.e., "summed") as elements of a resultant vector B , $B = M \cdot A$. The entire system can be integrated into a compact, flat device with an overall thickness of less than 5 mm. The first and third plane are composed of optical channel waveguides and arrays of grating couplers. These planes in some cases could be used to replace the front and back cover of a SLM. Optical waveguides with substrate mode operation [6] can also be used as the input and output planes, if the size and separation of the SLM pixels in the second plane are dimensionally compatible with the incident beam width and substrate thickness, respectively. Fibers or laser diodes can be directly attached to the channel waveguides to supply the inputs.

In the fan-out operation of the first plane, the important factors are uniform beam angular distribution and uniform beam intensity distribution. Uniform beam angular distribution of all output grating couplers is achieved by simultaneous holographic recording. Uniform beam intensity distribution is achieved by varying the efficiency of grating couplers along the same channel waveguide according to the number of grating couplers, waveguide loss, and the maximum achievable grating coupler efficiency. The grating coupler efficiency can be controlled by the exposure time of the holographic recording. In terms of holographic materials, holographic phase grating couplers [7] fit the fabrication of the graded-efficiency grating coupler array better than surface relief grating couplers because of their large dynamic range of refractive index modulation. In the fan-in operation of the third plane, special channel waveguides are designed to reduce the fan-in loss. Horn waveguide couplers are used to tailor the size of the input grating couplers to the width of the channel waveguide (e.g., a horn coupler can be used to convert a 1 mm-wide beam to a 20 μm -wide channel waveguide). In fact, no horn waveguide coupler is required if the thickness of the second plane (i.e., the SLM) is much less than the diffraction limited size. Curved channel waveguides are designed in order to combine channel waveguides from different rows of the matrix to form elements of the vector B . The combined channel waveguide will have a larger width than the two individual channel waveguides. This increase in the channel waveguide width reduces fan-in loss.

The use of grating couplers in both the input and output plane offers the possibility of dispersion free operation because their dispersions can compensate each other. The reason for this dispersion-free waveguide coupling is that the input and diffracted beams for both grating couplers in the first and third section of the crossbar are in the same plane. Another advantage of this waveguide-based optical crossbar is that the grating couplers are polarization sensitive, and additional polarizers can be eliminated if polarization-based SLMs are employed. In addition, this integrated optics approach is compatible with a number of SLMs because the dimensions of the gratings and waveguides are determined by photolithographic masks.

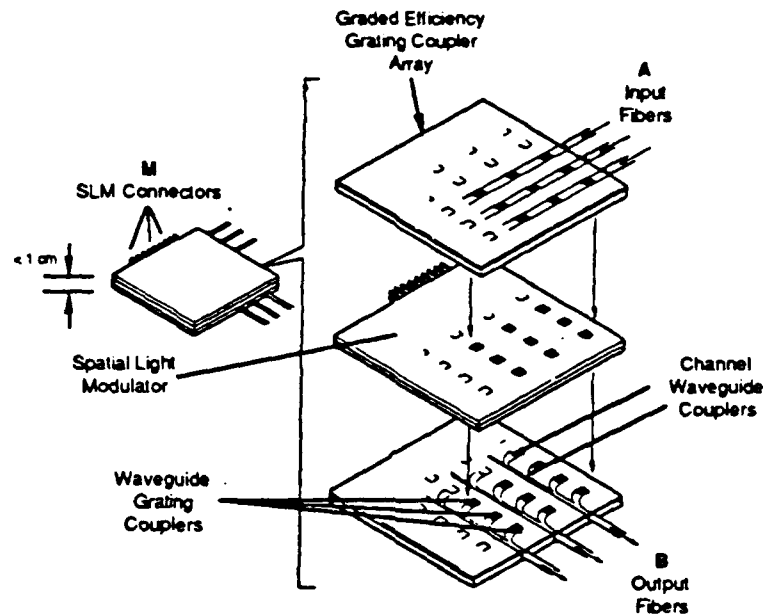


Figure 1. Optical crossbar switch packaging based on waveguide grating coupler arrays and a large array spatial light modulator.

3. Experimental Results

In the experiments we demonstrate the capability of fabricating arrays of grating coupler with equal output power distribution for the use in the first plane of the compact crossbar switch. We have used Na^+ ion exchanged single mode slab glass waveguides coated with a thin layer of dichromated gelatin (DCG) holographic material. The DCG layer acted as cladding for the guided wave launched into the glass waveguide. The interaction of the evanescent field with the DCG layer was strong enough for a grating recorded in the holographic material to diffract the guided wave in or out of the waveguide. Four grating couplers were recorded in the holographic material layer through a mask by two interfering Argon laser plane waves from free space incident on the plate. The grating area size was 1 mm wide and a few centimeters long allowing several parallel guided waves to interact with the gratings and resulting in an array of outcoupled beams of ~ 1 mm size. The separation between the gratings was made ~ 1 cm to match the separation of the pixels of a liquid crystal light modulator array used to modulate the outcoupled beams. The grating period was designed for a coupling angle of -20° for the 633 nm wavelength. Each of the gratings was recorded separately with different recording energies in order to obtain different diffraction efficiencies. The differences in the diffraction efficiencies had to account for the propagation loss in the waveguide, which was ~ 2 dB/cm, and the power outcoupled at each consecutive grating coupler. The fabricated structure is shown in Figure 2a. Light from a HeNe laser was coupled into the waveguide through a high index prism in an area uncovered by the holographic film and protection layers. The four outcoupled beams were parallel, suggesting uniformity of the recording and processing procedures. Output powers of the outcoupled beams varied by less than 5%. The maximum efficiency of the grating coupler

was 7%. Figure 2b is photograph of the waveguide grating coupler array with four outcoupled beams. The gratings had vertical Bragg planes and can be considered thin gratings for the guided waves. Therefore, their efficiency was not very high, even when large outcoupling angles were used. We have also obtained considerably higher efficiencies with slanted gratings (maximum efficiency ~50%) designed for coupling out in the direction normal to the plane of the waveguide. Arrays of slanted grating couplers in single mode waveguides are being developed.

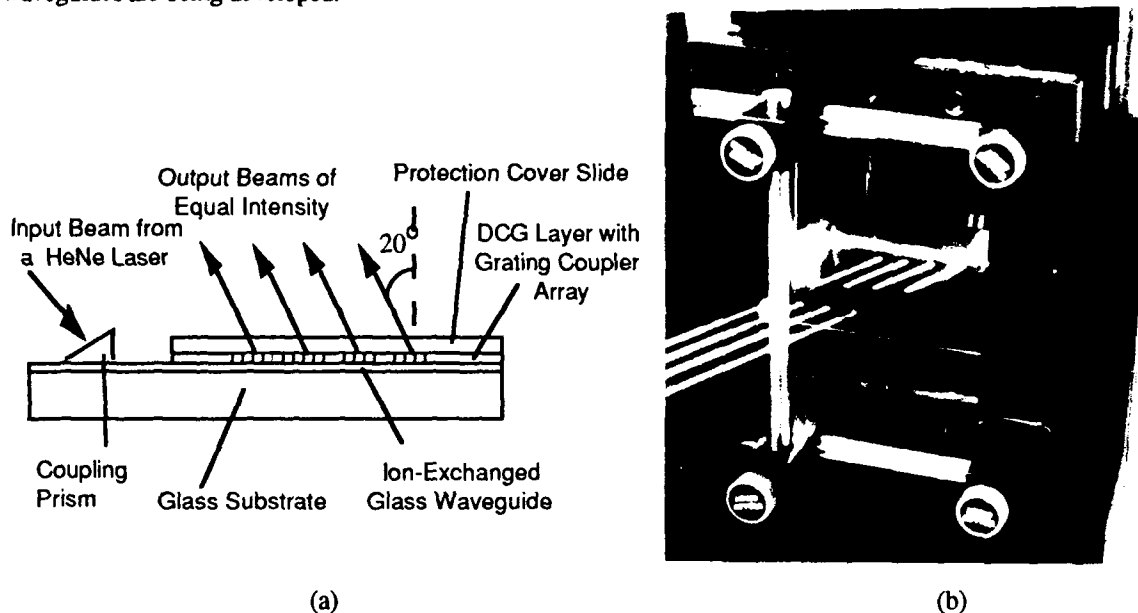


Figure 2. (a) Schematic of the structure of the grating coupler array in a single-mode waveguide. (b) Photograph of the grating coupler array in a single-mode waveguide with four outcoupled parallel free-space beams. The plate was mounted on a prism coupler.

We also demonstrate a compact matrix-vector multiplier using beams guided in the substrate mode of glass waveguides rather than single mode waveguides. The guided waves were distributed by arrays of slanted grating couplers into parallel free-space beams as shown in Figure 4. The slanted gratings were recorded using two Argon laser plane waves incident on a holographic plate through a prism. As in the previous case, each of the gratings had different diffraction efficiency. However, because the Bragg planes are slanted, their orientation and spacing change due to swelling of the holographic film during processing. To prevent uneven swelling, all grating regions received equal doses of total energy: part of the energy was used to create a grating using two interfering beams, and the remaining portion came from an additional one-beam exposure. Thus, all the gratings had the same swelling rate and therefore the same diffraction angle. A schematic of the assembled compact matrix-vector multiplier is shown in Figure 5a. For demonstration purposes, the input row of four beams was generated by another plate with an array of grating couplers. Two plates with arrays of four gratings were placed in orthogonal orientations and assembled using UV curing adhesive to create a 4×4 array of beams. A cylindrical lens was attached at the output with a space left in between the plates and the lens for the fixed mask. Four detectors in the focal plane of the lens collect and sum the output vector elements. The assembled device is shown in Figure 5b. Output power variations of one plate were less than 25%. Because two plates were used, the powers of the 16 output beams varied by ~50%. The large variations arise due to nonuniformities of the slanted gratings.

4. Conclusions

We have proposed using waveguide grating coupler arrays for optical crossbar devices based on an optical vector-matrix multiplier scheme. This approach offers compact size, large array size, is compatible with any of the state-of-the-art SLMs for the addressable matrix, and it offers dispersion free operation. Experiments demonstrating uniform output grating coupler arrays operated in single mode or substrate mode waveguides have also been performed.

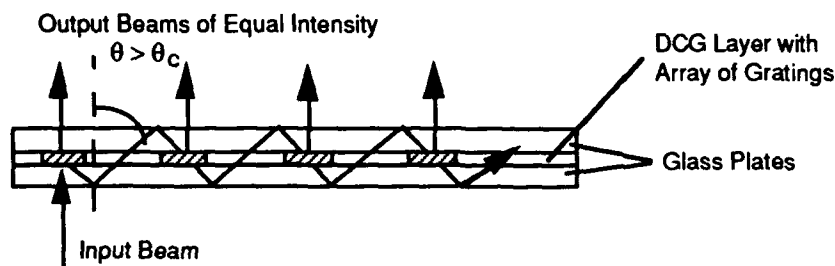


Figure 3. Schematic of the holographic grating array used to distribute a substrate guided beam.

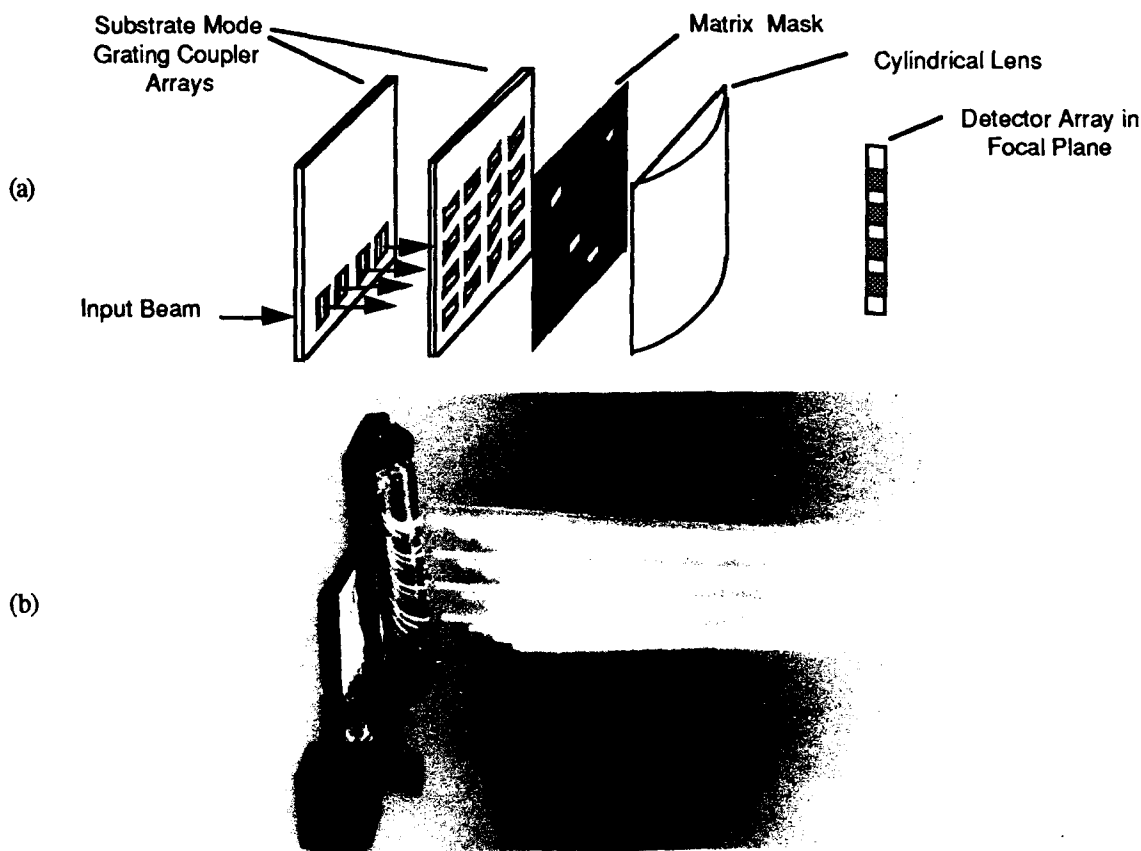


Figure 5. (a) Schematic of the assembled compact matrix-vector multiplier operating in substrate mode. (b) Photograph of the assembled matrix-vector multiplier with an array of 16 output beams which can be summed as a 4-element column vector at the focal plane of the cylindrical lens.

5. References

1. R. C. Alfemess, "Waveguide electrooptic switch array," IEEE J. Selected Areas in Communications, 6, 1117-1130 (1988)
2. D. A. B. Miller, "Quantum wells for optical information processing," Opt. Eng., 26, 368-372 (1987)
3. M. A. Handschy, K. M. Johnson, G. Model, and L. A. Pagano-Stauffer, "Electro-optic applications of ferroelectric liquid crystals to optical computing," Ferroelectrics, 85, 279 (1988)
4. S. H. Lee, S. C. Esener, M. A. Title, and T. J. Drabik, "Two-dimensional silicon/PLZT spatial light modulators: design considerations and technology," Opt. Eng., 25, 250-260 (1986)
5. A. R. Dias, R. F. Kalman, J. W. Goodman, and A. A. Sawchuck, "Fiber optic crossbar switch with broadcast capability," Opt. Eng., 27, 955-960 (1988)
6. T. Kubota and M. Takeda, "Array illuminator using grating couplers," Opt. Lett., 14, 651-652 (1989)
7. H. Kogelnik, and T. P. Sosnowski, "Holographic thin film couplers," Bell Syst. Tech J., 49, 1602-1608 (1970)

Techniques for Implementation of High-Speed Free-Space Optical Interconnections

Dean Z. Tsang

Lincoln Laboratory, Massachusetts Institute of Technology
Lexington, MA 02173-9108

A free-space optical interconnection has zero insertion force and no ground planes. Low propagation delays are possible in board-to-board applications because of the direct path between interconnected points compared to the circuitous path often required by electrical interconnections. The optical channel has very high bandwidth and low dispersion. Furthermore, very high interconnection densities are possible because light beams can cross in free space. However, high-speed optoelectronic components are typically small and difficult to align in a computer environment. Experimental free-space optical interconnections have required the use of micropositioners for alignment and are not readily adapted for conventional computers.¹ In this work, techniques for the practical implementation of high-speed board-to-board free-space optical interconnections without micropositioners have been developed. A free-space transmitter and receiver module combined with a card cage board enclosure designed for the application have made possible a differential electrical current efficiency (detector photocurrent divided by laser drive current) as high as 8%.

The use of optical modules with a source or detector and prealigned optics (Fig. 1) can greatly simplify the alignment of free-space optical interconnections for computer applications.² The miniature lens in the transmitter module converts the highly divergent laser output to a collimated beam with an expanded diameter. Optics in the receiver module focus the expanded beam onto a detector. The modules need only be aligned to the approximately millimeter tolerances of the expanded beam input/output optics. All critical alignments between the lens and the source or detector are performed during the assembly of the module. The modules can be mounted anywhere on the board to form board-to-board interconnections with three-dimensional layout

flexibility, while traditional electrical board-to-board interconnections are confined to the edges of the boards.

A photograph of a transmitter module is shown in Fig. 2. The package is a modified high-speed commercially available flatpack. An edge-emitting 1.3- μm diode laser was soldered to a submount for light emission vertical to the package. A drilled hole in the package lid was centered over the diode laser. A 1.8- or 2-mm-diameter 0.23-pitch graded-index (GRIN) lens was mounted in a lens collar, inserted into the hole, and aligned with the laser on. The position of the lens was adjusted for a collimated beam. Measurements of the transmitter module with an optical lightwave analyzer show that the transmitter module has a flat frequency response up to the 3-GHz limit of the analyzer.

A receiver module with an exterior appearance identical to that of a transmitter module has also been constructed. A 1.8- or 2-mm-diameter 0.23-pitch GRIN lens was positioned over a commercially available 100- μm -diameter InGaAs detector mounted on a ceramic subcarrier. The lens was aligned with a collimated input beam and current monitoring instrumentation. Measurements of the receiver module on the optical lightwave analyzer show that the receiver module response is limited by the detector to 1.5 GHz.

The modules were positioned on conventional electronic circuit boards. The circuit boards were inserted into a card cage with precision-milled slots (Fig. 3). The slots, which are made to ordinary machine shop tolerances, position the boards and reduce the effects of board flex. A differential electrical current efficiency (with the laser above threshold) as high as 8% has been measured. No micropositioners are necessary. The efficiency can be maintained even after removal and reinsertion of the boards. Even higher efficiency should be possible with more efficient lasers and detectors or with different lenses.

These results demonstrate that high-speed free-space optical interconnections can be implemented by the incorporation of prealigned lenses into transmitter and receiver modules.

These modules make possible board-to-board interconnections with card-cage enclosures for the boards constructed to ordinary machine shop tolerances.

This work was supported by the Defense Advanced Research Projects Agency.

1. See, for example, D. Z. Tsang, Proc SPIE, vol. 994, pp. 73-76, 1988.
2. D. Z. Tsang, Technical Digest, Topical Meeting on Optical Computing (Optical Society of America, Washington, DC, 1989), pp. 146-149.

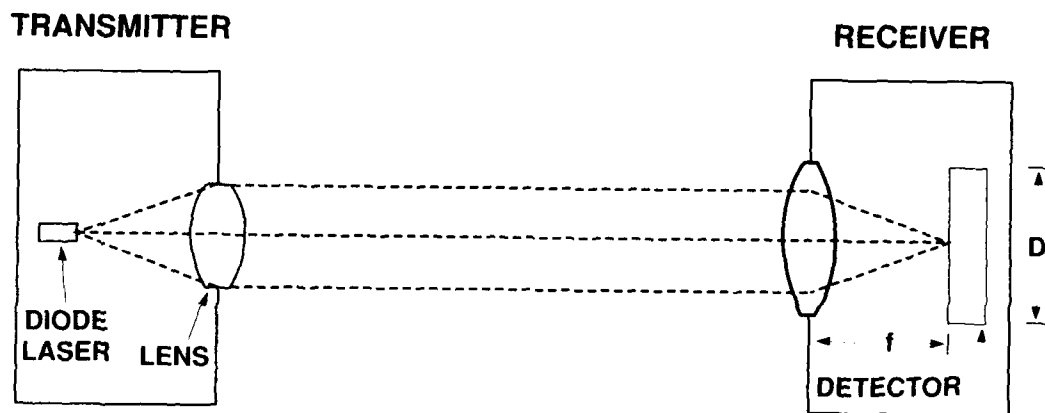


Figure 1. Transmitter and receiver modules for board-to-board optical interconnections. Each module has a lens that is prealigned to its respective laser or detector.

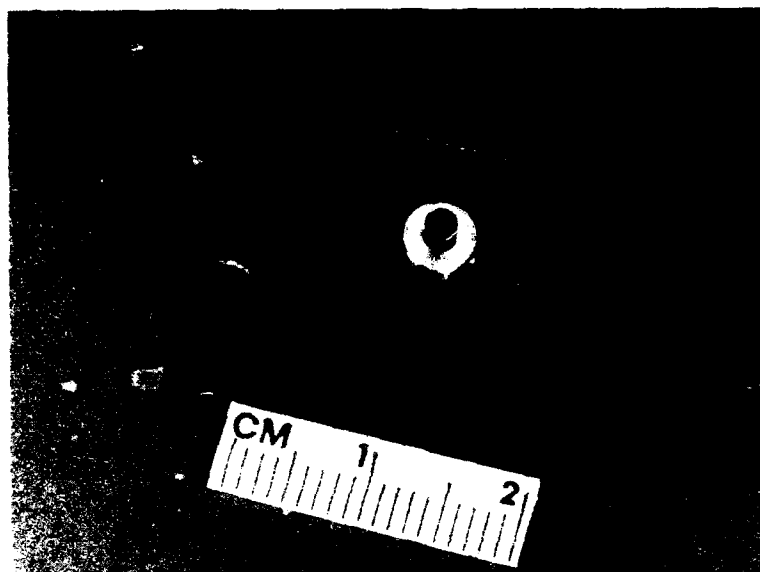


Figure 2. Photograph of a high-speed free-space transmitter or receiver module with GRIN lens optical input/output.

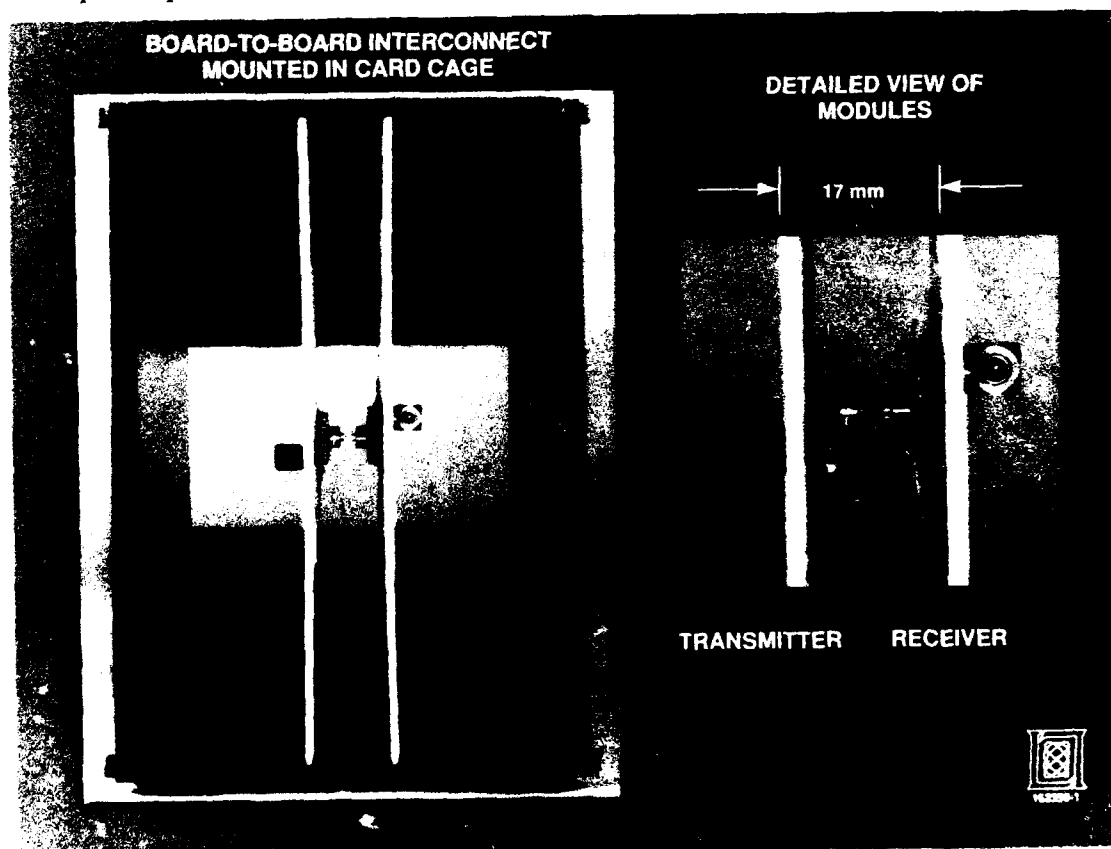


Figure 3. Photograph of modules mounted on boards that are aligned by insertion into a board enclosure without micropositioners.

Two-Dimensional Optical Buses for Massively Parallel Processing

Shigeru KAWAI and Masanori MIZOGUCHI

C&C Information Technology Research Laboratories

NEC Corporation

1-1, Miyazaki 4-chome, Miyamae-ku, Kawasaki, 216, Japan

INTRODUCTION

Massively parallel processing systems have the serious problems of interconnection pathways between processors. Light beams are particularly appropriate for this purpose, because they can cross each other with no mutual interference.[1] Optical interconnections which satisfy this condition can be classified into two techniques, which involve two-dimensional waveguide[2]-[4] and free-space interconnections. These interconnections have no physical pathways. Taking into account pathway density and power efficiency, waveguide techniques are suitable for bus-connections between in-plane processors, while free-space techniques are suitable for interconnections between plane-to-plane processors. One of the authors has already proposed free-space interconnections using micro lens arrays for this purpose.[5] This paper describes an attempted implementation of the new optical buses using two-dimensional waveguides for parallel processing initially proposed by R. A. Linke.[6] Their principles are successfully demonstrated, using glass plates with concave lenses.

TWO-DIMENSIONAL OPTICAL BUSES

Figure 1 shows a parallel processing system with the proposed optical bus. It consists of a two-dimensional waveguide with concave lenses, and electric processors with light sources and photo-detectors. The outgoing light beam from the light source is incident to the waveguide from the lens. Light beams, whose incident angles are greater than the critical angle for the waveguide, can propagate in the waveguide. They can be incident to the waveguide and outgoing from it at the concave lens, because the incident angle at the point is smaller than the critical angle. When processors with light sources and photo-detectors are arranged near the lenses, the output signal from a processor can drive the light source, the light beam from it can propagate in the waveguide, and the light beam can be detected by the photo-detectors at the other processors. The detected signal can drive circuits in the processor. The waveguide just works as an optical bus.

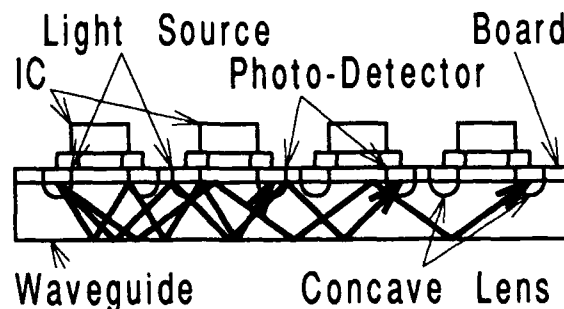


Fig. 1 Parallel processing system with proposed optical bus.

Figure 2 shows other applications for the optical bus. A light source or a photo-detector is set at the side of the waveguide. For example, when a light source is set, a system clock, or control signals and common data can be distributed to the processors. When a photo-detector is set, the status in the processor can be monitored.

Concave lenses on the waveguide surface have other variations, as shown in Fig. 3. One variation involves using a concave micro lens array, made by changing their refractive indices. This waveguide is easily integrated with electric circuits, because this type of micro lens array has completely flat planes. Using a waveguide with wedge shape ditches is another technique. The ratio of the critical angle to the incident angle at the part of the ditch is smaller than at the flat plane. This waveguide is easily fabricated.

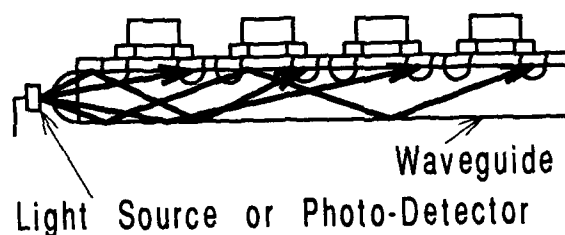


Fig. 2 Other optical bus applications.

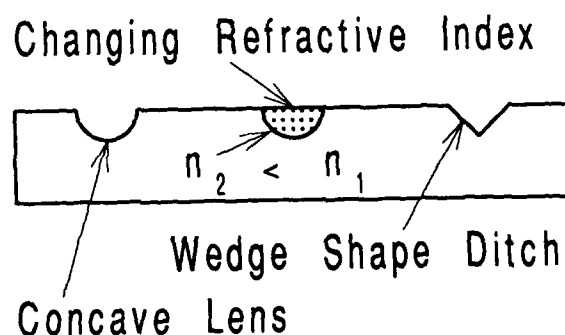


Fig. 3 Other kinds of waveguide variation.

PROPAGATION LOSS ESTIMATION

Propagation losses are estimated for various glass plates, which are the most appropriate for the two-dimensional waveguide. Figure 4 shows the calculated propagation losses in glass plates, when a light beam is incident from the side of the plate. The incident light beam almost propagates in the plate. The loss is dependent on the glass thickness and the detector size. When a 250 μ m thick glass plate is used upon which 1mm x 1mm size detectors are placed, the light beam can propagate several hundred milli-meters. Arranging lenses on the plate, one to several hundreds or several hundreds to one interconnections may be accomplished.

Figure 5 shows the calculated propagation losses, when a light beam is incident from one of the lens. It is difficult for beams to propagate in the plate using total reflection, because almost incident light beams pass through the glass plate. Coating the reflective materials on the surface of the plate assists the light beam to propagate. Shifting the optical axis for the light source, from the axis for the lens, brings more light beams for the plate. In this case, the loss is dependent on the beam divergence angle and is independent from the plate thickness. By appropriately controlling the angle, several hundred interconnections also may be set up in this case.

EXPERIMENTAL RESULTS

In order to confirm the principles involved for the optical bus, a 5mm thick glass plate

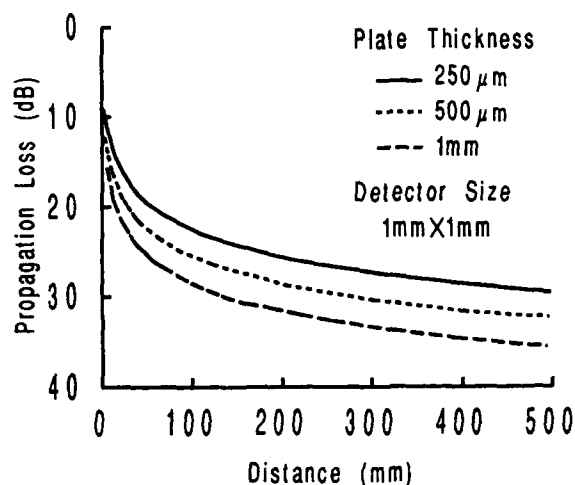


Fig. 4 Propagation losses, when a light beam is incident from the side of the plate.

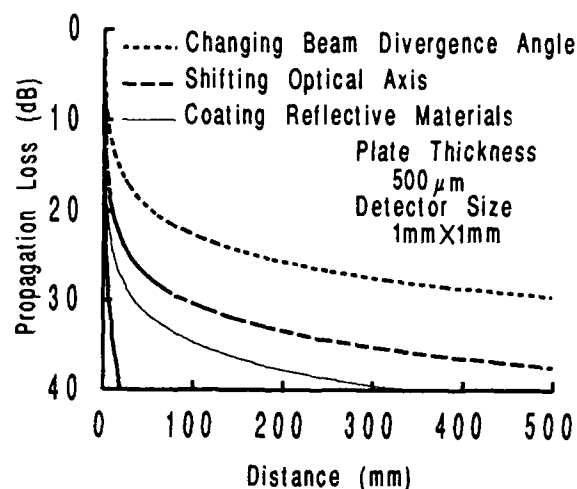


Fig. 5 Propagation losses, when a light beam is incident from one of the lenses.

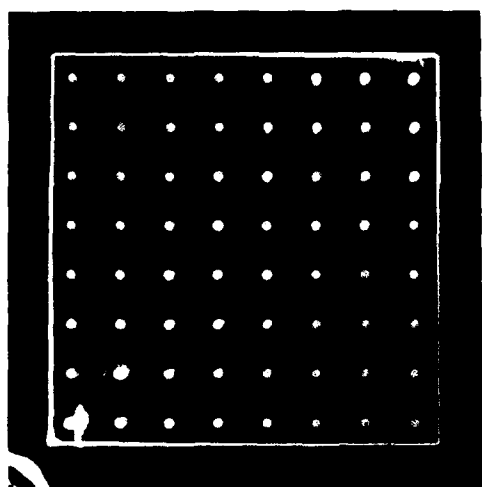


Fig. 6 Experimental result, when a light beam is incident from the side of the plate.



Fig. 7 Experimental result, when a light beam is incident from one of the lenses.

was made up with 2mm diameter concave lenses on its surface. Figure 6 shows the outgoing light beams from the lenses, when a laser beam was incident from the side of the plate. Figure 7 shows the outgoing light beams, when a laser beam was incident to one of the lenses.

INTERCONNECTIONS FOR PARALLEL PROCESSING

Figure 8 shows the concepts used in implementing optical interconnections for massively parallel processing. Many boards with electric circuits, or wafer scale integration circuits (WSI) are appropriately arranged. Interconnections between boards or WSIs are performed optically by free-space techniques while interconnections between processors in boards or WSIs are carried out by these waveguide techniques.

Figure 9 shows a multiple bus interconnection example. Several glass plates are stacked in layers. The position of individual lenses on a glass plate is different from the position of lenses on other plates. For example, a light beam passes through the first plate and is incident to the lens in the second plate. The light beam is diverged and propagates in the second plate. By changing the relative positions for the lens and photo-detector, a light beam can be caused to propagate in the desired glass plate.

Glass plates are used as the waveguides mentioned above. However, Si or GaAs plates may be used, because infrared light beams can propagate in these materials. If such a waveguide were to be realized, both waveguides and electric circuits would be made by the same materials. It may be easier to integrate with them.

SUMMARY

Optical buses with two-dimensional waveguides were presented. Their principles are successfully demonstrated, using glass plates with concave lenses. Combining this interconnection techniques with free-space optical interconnection techniques, interconnection pathway problems in parallel processing systems may be solved.

The authors gratefully thank Drs. K. Iinuma, T. Temma and J. Tajima for their suggestions and encouragement.

REFERENCES

- [1] J. W. Goodman et al., Proc. IEEE **72**, 850-866 (1984).
- [2] T. Kubota and M. Takeda, OC'90 Digest, Kobe, paper 10H10 (1990).
- [3] H. J. Haumann et al., OC'90 Digest, Kobe, paper 10B3 (1990).
- [4] J. Jahns et al., OC'90 Digest, Kobe, paper 10B4 (1990).
- [5] S. Kawai, PS'90 Digest, Kobe, paper 14D-1(PD) (1990).
- [6] R. A. Linke, private communications.

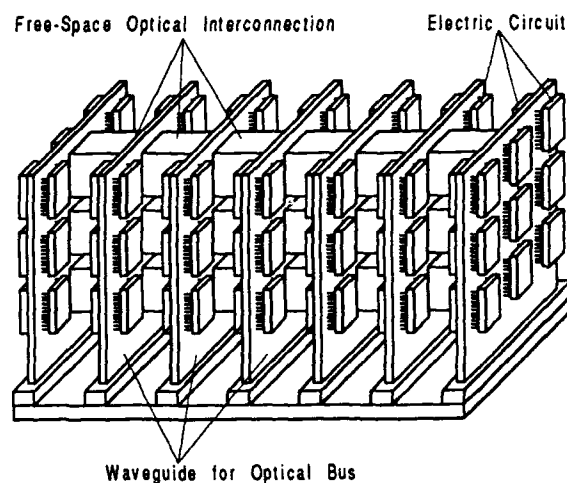


Fig. 8 Optical interconnections for massively parallel processing.

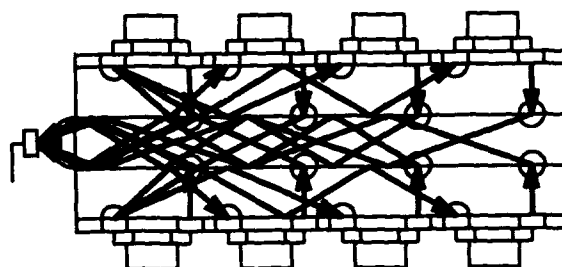


Fig. 9 Example for multiple bus interconnection.

Hologram Recording by a Pulsed Method on Phase Change Media

Koutarou NONAKA, Yasuhide NISHIDA, and Susumu FUJIMORI
NTT Applied Electronics Laboratories
3-9-11 Midori-Cho, Musashino-Shi, Tokyo 180, JAPAN

I. Introduction

This paper presents a novel real-time hologram recording method and derives a formula for calculating the diffraction efficiency of gratings recorded in phase change media. The formula gives media parameters for high diffraction efficiency. In addition, a two-dimensional character pattern is successfully reconstructed using GeTe alloy film.

A holographic memory is one of essential devices in optical computing systems utilizing massive parallelism of optics. The evaluation data of typical holographic recording media for information storage are listed in Table I.(1) Some of these media are used for demonstrations of optical computing systems, such as associative memories and optical neural networks. However, they do not fulfill the following requirements for recording media in practical optical computing: development-free, high recording sensitivity, high diffraction efficiency, nonvolatility, erasability, and low cost.

Phase-change media has been extensively studied as an optical recording media through pit formation employing reversible amorphous to crystalline phase transition.(2) This media comes closer to fulfilling the above requirements. Furthermore, because of a large difference in refraction index between the amorphous and crystal states, high diffraction efficiency is expected as a phase hologram. Pit formation has been achieved using amorphisation induced by a short laser pulse. Because this pulsed method reduces transverse heat flow,(3) it is also indispensable for high resolution holographic recording.

This report presents our theoretical and experimental results of hologram recording on phase-change media using the pulsed method.

II. Diffraction Efficiency Analysis

A. Model and analysis method

The media structure to be discussed is shown in Fig. 1. This multilayer structure results in a

high recording sensitivity and a high reconstruction SNR(4). To find media parameters for high diffraction efficiency, we first derive a formula for calculating diffraction efficiency with variables of transmittance, reflectance, and phase difference. Next, we quantify the variables of the

Table I. Evaluation data of various materials for holographic information storage.

Materials	Silver Halide	Ferroelectric Crystal (BSO)	Thermoplastic	Magneto Optic	Phase Change
Development	Chemical	Free	Charging and Heating	Free	Free
Sensitivity on Recording (mJ/cm ²)	~ 0.1	1~ 10	~ 10	~ 5	5 ~ 10
Diffraction Efficiency (%)	< 10	< 30	10 ~ 30	< 0.01	< 5
Resolution (lp/mm)	2000~5000	~ 1000	~ 1000	~ 2000	~ 2000
Non-volatility	Good	Poor	Good	Good	Good
Write-erase Cycle Repeatability	No	Good	Poor	Good	Good
Cost/bit	Low	High	Low	Low	Low

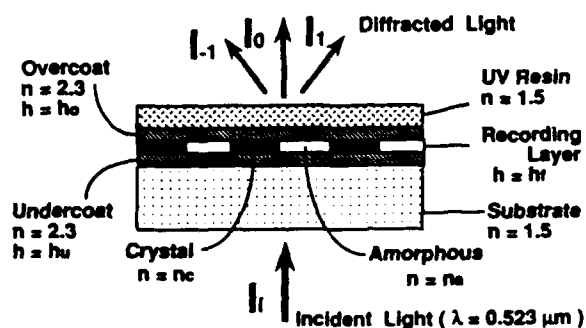


FIG. 1 Media structure for analysis and experiment.

formula in terms of the media parameters, that is, thicknesses of the overcoat, undercoat, and recording film, and index of refraction of the recording film.

We derive the formula using the model shown in Fig. 2. Figure 2(a) shows a cross-sectional view of the recording film. Infinitely long ribbon-like amorphous and crystal regions are aligned periodically in the x -direction. The phase of transmitted light $P(x)$, transmittance $T(x)$, and reflectance $R(x)$ become square-wave functions of x , as shown respectively in Figs. 2(b), (c), and (d). In these figures, $\Delta\phi$ is the phase difference between transmitted light of an amorphous region and that of a crystal region. T_a and R_a are the respective transmittance and reflectance of the amorphous region, and T_c and R_c are those of the crystal region.

B. Formula Derivation

The diffraction efficiency is the ratio of intensity of first -order diffracted light I_1 to that of incident light I_i . The I_1 is calculated by the Fourier-type integral with pupil functions of $P(x)$, $T(x)$, and $R(x)$.

Consequently, diffraction efficiency η is

$$\eta = \{ (t_a - t_c)^2 / \pi^2 \cos^2(\Delta\phi/2) + 4/\pi^2 [t_c + (t_a - t_c)(1 - s/p)]^2 \sin^2(\Delta\phi/2) \} \times \sin^2(\pi s/p) [1 - (R_a + R_c)/2], \quad (1)$$

where

$$t_a = \sqrt{T_a}, \quad t_c = \sqrt{T_c}.$$

III. Results of the Analysis

A. Evaluation of the diffraction efficiency formula

Phase difference $\Delta\phi$ versus diffraction efficiency η is shown in Fig. 3 with parameters T_a and T_c . η increases as $\Delta\phi$ increases in this region of $\Delta\phi$. The high transmittance of T_a and T_c also leads to high diffraction efficiency. $\Delta\phi$, T_a , and T_c are related to n_a and n_c which are complex refractive index of amorphous region and that of crystal region, respectively. Namely, $\Delta\phi$ increases as the difference between $\text{Re}(n_a)$ and $\text{Re}(n_c)$ increases. T_a and T_c also increase as $\text{Im}(n_a)$ and $\text{Im}(n_c)$ decrease. Therefore, high diffraction efficiency is obtained by decreasing $\text{Im}(n_a)$ and $\text{Im}(n_c)$ and by

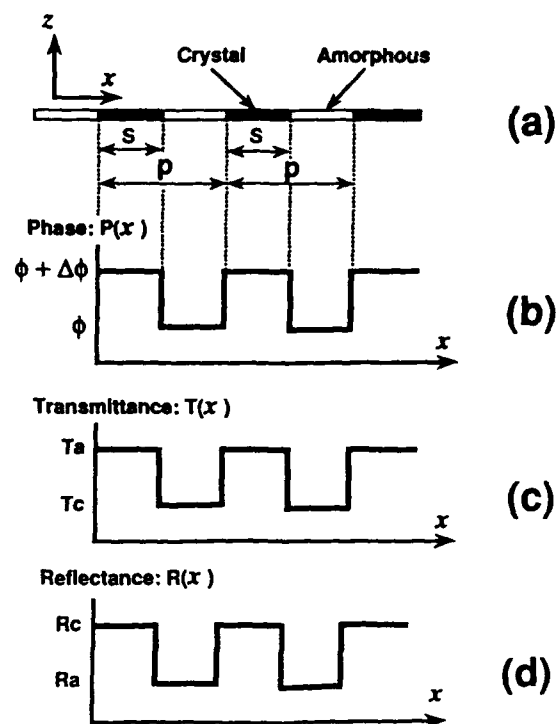


FIG. 2 Model for formula derivation for calculating diffraction efficiency. (a) Recorded gratings in the phase change film. (b) Phase of transmitted light. (c) Transmittance. (d) Reflectance.

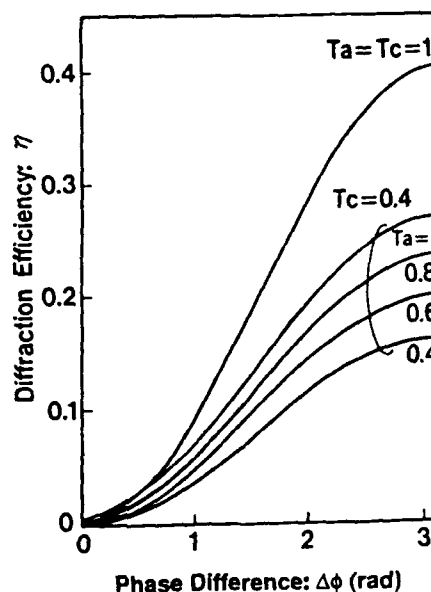


FIG. 3 Phase difference $\Delta\phi$ versus diffraction efficiency η . Parameters are transmittance T_a and T_c . ($p \approx 2s$, $R_a = R_c = 0$)

increasing $\text{Re}(n_c) - \text{Re}(n_a)$. When $T_a = T_c = 1$ and $\Delta\phi = \pi$, η is the maximum value of 0.4. This is a well-known result.

B. Transmittance, reflectance, and phase difference of the multilayer structure

By the matrix technique,⁽⁵⁾ T_a , T_c , R_a , R_c , and $\Delta\phi$ for the multilayer structure were calculated. An example of the calculated results is shown in Fig. 4. The diffraction efficiency can be calculated by combining formula (1) with the data shown in Fig. 4.

In this figure, we show that phase difference $\Delta\phi$ in the multilayer structure differs from that in monolayer case δ which is plotted by the dashed line. We write δ as

$$\delta = 2\pi [\text{Re}(n_c) - \text{Re}(n_a)] h_f / \lambda.$$

δ is proportional to h_f , as stated in this equation. However, when the recording film is very thin, the dependence of $\Delta\phi$ on h_f differs from that of δ on h_f . This difference between $\Delta\phi$ and δ is caused by the interference effects of the multilayer structure and by the absorption of light corresponding to $\text{Im}(n_a)$ and $\text{Im}(n_c)$.

C. Film thicknesses Design

The diffraction efficiency η as a function of overcoat film thickness h_o and undercoat film thickness h_u is shown in Fig. 5. Figures by the contour curves show the value of diffraction efficiency. η is the periodic functions of h_o and h_u . The maximum value of η occurs at $h_o = 0.04, 0.16 \mu\text{m}$, and $h_u = 0.04, 0.16 \mu\text{m}$. When the film becomes very thin, film uniformity and strength become poor. Hence, we found the optimum thickness to be $h_o = h_u = 0.16 \mu\text{m}$.

Recording film thickness h_f versus diffraction efficiency η is shown in Fig. 6 with parameters h_o and h_u . Maximum diffraction efficiency occurs at a particular film thickness of $0.04 \mu\text{m}$. This thickness is independent of h_o and h_u .

IV. Experiment

To confirm our analysis, we recorded gratings on phase change media and measured the diffraction efficiency of the grating. This

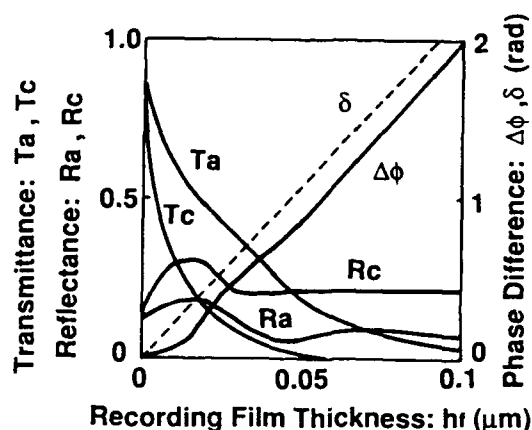


FIG. 4 Recording film thickness h_f versus transmittance T_a and T_c , reflectance R_a and R_c , and phase difference $\Delta\phi$, all for the multilayer structure shown in FIG. 1. δ is the phase difference for the monolayer case. ($n_a = 4.7 - i1.4$, $n_c = 6.5 - i3.5$, $h_o = h_u = 0.16 \mu\text{m}$)

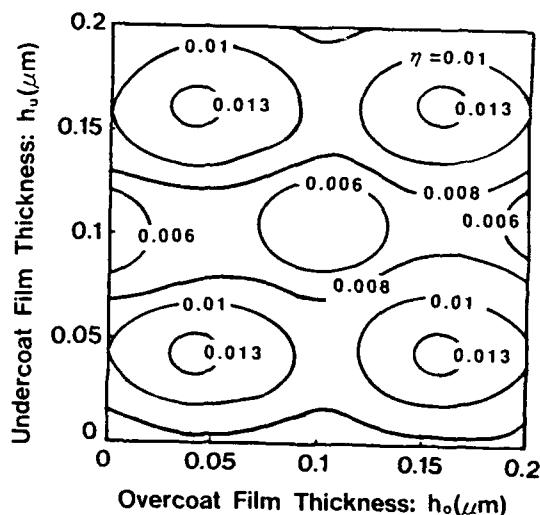


FIG. 5 Overcoat film thickness h_o and undercoat film thickness h_u versus diffraction efficiency η . ($p = 2s$, $h_f = 0.04 \mu\text{m}$, n_a and n_c are the same as that in FIG. 4)

experiment used a GeTe alloy as the recording film. The media structure was the same as that in Fig. 1. The grating was recorded by pulsed laser.(3)

The measured values of diffraction efficiency are plotted in Fig. 6 by closed circles. The maximum diffraction efficiency is 0.85% at $h_f = 0.04 \mu\text{m}$. The experimental results agree well with the calculated result plotted by the solid line. These data confirm experimentally that maximum diffraction efficiency is obtained at $h_f = 0.04 \mu\text{m}$.

An example of a GeTe phase change hologram together with the reconstructed image is shown in Fig. 7. The fairly good reconstructed image was obtained using phase change media. The diffraction efficiency of this hologram was 0.4%. The diffraction efficiency was decreased by modulation of the grating.

V. Conclusion

A novel real-time hologram recording using phase-change media was presented. A formula was derived for calculating the diffraction efficiency of gratings recorded in the phase change media. By combining the derived formula with the matrix technique, we obtained media parameters for high diffraction efficiency for multilayer recording media. These results were confirmed by the experiment. The two-dimensional character pattern was successfully reconstructed using GeTe alloy film. This hologram recording presented here fulfills the requirements for information storage in optical computing.

References

- (1) H. M. Smith ed., *Holographic Recording Materials* (Springer, Berlin, Heidelberg, New York, 1977).
- (2) S. Fujimori, S. Yagi, H. Yamazaki, and H. Funakoshi: *J. Appl. Phys.*, **64**, 1000(1983).
- (3) K. Nonaka, Y. Nishida, and Y. Koshimoto: *Jpn. J. Appl. Phys., Suppl.* **28-3**, 275 (1989).
- (4) A. E. Bell and F. W. Spong: *IEEE J. Q. E.*, **QE-14**, 487 (1978).
- (5) F. Abeles: *Ann. de Physique*, **5**, 596 (1950).

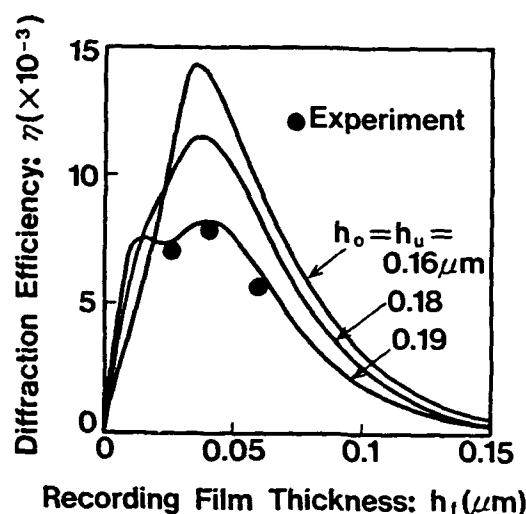


FIG. 6 Recording film thickness h_f versus diffraction efficiency η . Parameters are overcoat and undercoat film thicknesses. ($p = 2s$, n_a and n_c are the same as that in FIG. 4)

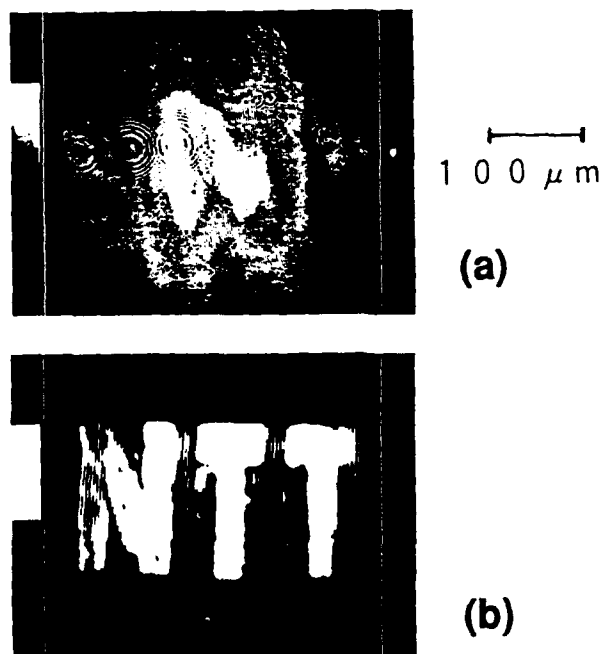


FIG. 7 Recorded hologram on GeTe phase change media (a) and reconstructed image (b).

Progress in diffractive phase gratings used for spot array generation

Rick L. Morrison

Sonya L. Walker

AT&T Bell Laboratories

Naperville, Illinois 60566

1. INTRODUCTION

Prototype free-space photonic switching^[1] and optical computing systems^[2] rely on spot array generating systems to produce illumination needed to transfer information between arrays of optical processing elements. In these systems, the light from a single laser is split into a set of beams that are focussed onto an array of optical logic devices, such as S-SEEDs^[3]. Although several methods are available for generating spot arrays^[4], diffraction gratings (also referred to as Dammann gratings^{[5] [6]}) were chosen to generate the spot arrays. Their advantages can be traced to the ease with which they are incorporated into an optical system. As illustrated in figure 1, the grating is simply inserted into the collimated beam with the appropriate imaging optics. Their operation is relatively insensitive to alignment. Provided the collimated beam illuminates a suitable size area of the grating, the performance is determined by the design and fabrication process.

The criteria upon which the diffraction gratings are judged include: the suitability of the spot array design to be used with the format of the optical logic array, the efficiency of coupling the laser light into the spots of interest, and the uniformity of the spot intensities within the array. The first concern mandates an array that is sufficiently large and matches the configuration of the devices. Efficiency becomes critically important whenever spots are generated for a complicated system which suffers severe power losses and where only limited laser power is available. Both topics are design issues. Finally, the mutual uniformity of the intensities must match the operational tolerances of the optical processing devices. The spot uniformity is linked primarily to limitations of the fabrication process.

In this paper, we will examine the improvements that have been incorporated into the diffraction grating design, fabrication, and characterization process that make them excellent elements for use in digital free-space optical systems.

2. GRATING DESIGN

Several advances have occurred since Dammann originally proposed the use of binary phase gratings for creating spot arrays. Higher efficiency designs were calculated for both binary^[7] and multi-level configurations^[8]. However, even with the widespread availability of moderately powerful microcomputer systems, a simplifying parametrization of the grating pattern can significantly improve the complex optimization process.

Even numbered spot arrays^[9] were demonstrated that naturally match the current S-SEED array system design configuration. A comparison of the standard odd numbered spot array design and the even-numbered design is shown in figure 2. The even numbered design produces a configuration of spots that consists of bright odd orders and dark or suppressed even orders. This configuration is highly desirable since it eliminates the central order spot from the regular array. This central order spot is primarily responsible for the array intensity non-uniformity due to its critical dependence on the accuracy of the phase levels.

The even-numbered design would at first seem to be more complicated than the standard design, since twice as many orders must be specified. Fortunately, symmetries can be incorporated in the design that significantly reduce this complexity. For the case of binary phase gratings, the symmetry amounts to a translation of the first half period phase transition locations into the second half period. In addition, the phase level values (either 0 or π) of the second half are exchanged with those found in the first half.

This design process for even-numbered spot arrays is extendible to discrete multi-level and continuous phase level patterns^[10]. For either case, the first half of the pattern period is translated into the second half with an additional phase offset of π added. Also, the second (fourth) quarter of a period pattern is a reflection of the first (third) period about their common boundary. Both the translation and the reflection symmetry are illustrated in figures 3 and 4. By extending the design using multi-level and continuous parametrizations, solutions with efficiencies greater than 90% have been calculated.

3. FABRICATION

Many of the fabrication processes that are essential for the production of diffraction gratings were originally developed for the semiconductor industry. The primary fabrication steps consist of mask production, microlithography, etching and/or material deposition. Limitations in the fabrication process ultimately determine the performance that can be obtained from diffraction gratings. As spot arrays become larger and designs necessarily become more complex, the size and granularity of the pattern generation step begin to stress one or more of the processes associated with fabrication. Figure 5 shows several of the problems that can occur.

One of the more severe problems that occurs during the lithography step is the dilation of features during the lithography process. During exposure, the features may dilate by more than half a micron. Although this may not appear to be significant, it can drastically change the performance of complex gratings with small periods. Fortunately, this effect can be quantified and corrected, leading to improved performance of the grating.

By carefully monitoring the fabrication process it is possible to routinely fabricate gratings with spots arrays having 128 spots with the standard deviation of the intensity variation limited to 1.5%. In addition, manufacturing economy occurs as a result of the process' ability to produce several identical devices simultaneously.

4. CHARACTERIZATION

Diffraction phase gratings will remain valuable as spot array generators only if they continue to match or exceed the performance criteria required by the digital optical system. New designs and enhanced fabrication processes will ultimately lead to higher efficiency gratings. Uniformity is especially critical in a system when a logic operation between input signals from two separated S-SEEDs depends on the size of the receiving S-SEED's bistable loop. In order to insure that these criteria are being met, it is necessary to characterize the grating performance. By doing so, the characterization can serve to identify limitations in the fabrication process.

Figure 6 shows the configuration used to measure a moderate size spot array. The system consists of a laser diode and collimating optics to create a monochromatic plane wave, the grating and objective lens to form the spot array, and a pinhole and optical power meter that are moved via computer controlled micro-positioning stages to measure each spot. This system is able to measure the relative intensities of each spot with a precision of about 1%.

The pin-hole/optical power meter combination works adequately for moderate array sizes; however, it is far less suited to large spot arrays due to the time required for measurement. An alternate approach is to utilize the image acquisition capabilities of charge coupled device (CCD) cameras to capture the information and then analyze the data. Although most CCD video cameras do not exhibit performance that matches the capabilities of the optical power meter system, CCD systems developed for astronomy are ideally suited for this task. These cameras can be configured with high spatial resolution (more than a million pixels), good intensity resolution (16 bit A/D systems), and good signal to noise ratios. Figure 7 shows an image captured by such a system developed to characterize grating performance.

An additional benefit of these image acquisition systems is their ability to serve as a diagnostic tool for optical computing and photonic switching systems. Such systems have been used to investigate the bistable loops of an array of S-SEEDs. Also, image acquisition systems have been used to track the uniformities of a spot array as it is relayed through elements in a prototype photonic system.

5. SUMMARY

Diffraction gratings will serve as the preferred method of generating spot arrays for the foreseeable future. The even numbered grating design, with the extension to multi-level and continuous patterns, assures that high efficiency spot production will keep pace with the demands of free-space optical systems. Eventually, limitations of the fabrication process could hinder further progress; however, it is not clear whether the operational simplicity would be sacrificed or if the logic system will be designed to be more tolerant. Also, by carefully characterizing the grating performance, the design and fabrication process have been improved. Furthermore, these image analysis systems, developed to characterize the grating performance, will serve as valuable diagnostic tools in digital optical systems.

REFERENCES

1. T. J. Cloonan, F. B. McCormick, M. J. Herron, F. A. P. Tooley, G. W. Richards, E. Kerbis, J. L. Brubaker, and A. L. Lentine, "A 3D Crossover Switching Network based on S-SEED Arrays," Proceeding of the 1990 International Topical Meeting on Photonic Switching, post-deadline paper 14D-2, Kobe, Japan, April, 1990.
2. M. E. Prise, N. C. Craft, R. E. LaMarche, M. M. Downs, S. J. Walker, L. A. D'Asaro, and L. M. F. Chirovsky, "A Module for Optical Logic Circuits using Symmetric Self Electrooptic Effect Devices," *Appl. Opt.* **29**, pp.2164-2170 (1990).
3. A. L. Lentine, H. S. Hinton, D. A. B. Miller, J. E. Henry, J. E. Cunningham, and L. M. F. Chirovsky, "Symmetric self-electro-optic effect device: Optical set-reset latch, differential logic gate, and differential modulator/detector," *IEEE Journal of Quantum Electronics* **25**, 1928-1936 (1988).
4. N. Streibl, "Beam shaping with optical array generators," *J. Mod. Opt.* **36**, 1559-1573 (1989).
5. H. Dammann and K. Gortler, "High-efficiency in-line multiple imaging by multiple phase holograms", *Opt. Commun.* **3**, 312-315 (1971).
6. H. Dammann and E. Klotz, "Coherent-optical generation and inspection of two-dimensional periodic structures," *Opt. Acta* **24**, 505-515 (1977).
7. U. Killat, G. Rabe, and W. Rave, "Binary phase gratings for star Couplers with high splitting ratios," *Fiber and Integrated Optics* **4**, 159-164 (1982).
8. S.J. Walker and J. Jahns, "Array generation with multilevel phase gratings," *OSA Annual Meeting, 1988 Technical Digest Series*, Vol. 11.
9. R. L. Morrison and S. L. Walker, "Binary phase gratings generating even numbered spot arrays," *Optical Society of America Annual Meeting, 1989 Technical Digest Series*, Vol. 18, submitted to *Applied Optics*.
10. R.L. Morrison, "Symmetries that simplify spot array phase grating design," to be published.

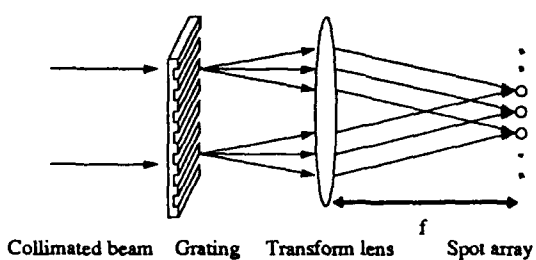


Figure 1- Spot array generated by a diffraction grating from a monochromatic plane wave.

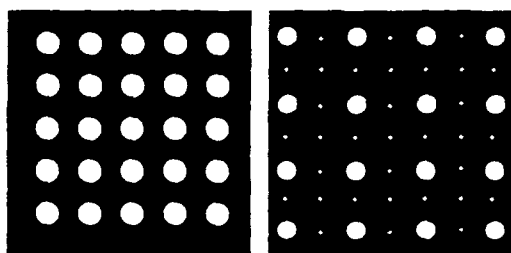


Figure 2 - Examples of standard odd numbered (left) and even numbered (right) designs. Smaller circles represent suppressed orders.

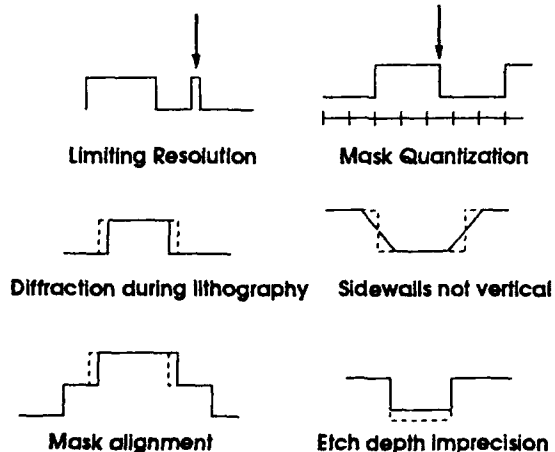


Figure 5 - Defects that can occur during grating fabrication.

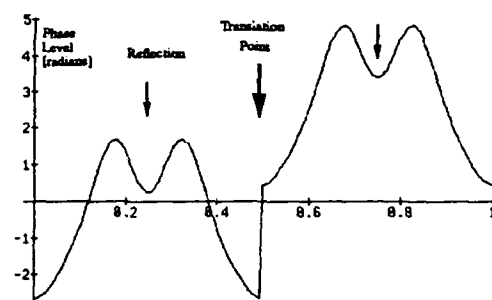


Figure 3 - Legendre parametrization 1x8 design showing translation and reflection symmetry.

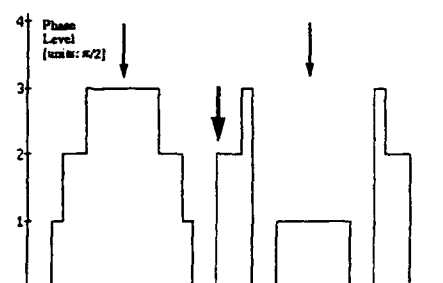


Figure 4 - 1x8 multilevel design showing both the translation and reflection symmetry. The phase is restricted modulus 2π .

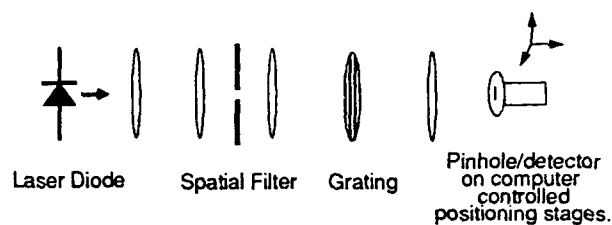


Figure 6 - Setup used to measure the intensities of individual spot in an array.

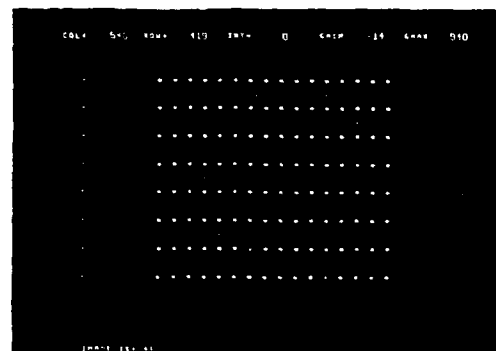


Figure 7 - Spot array image acquired by a high resolution CCD camera system.

PASSIVE OPTICAL ARRAY GENERATORS

Mohammad R. Taghizadeh, Jari Turunen, and Brian Robertson

Heriot-Watt University, Department of Physics, Riccarton, Edinburgh EH14 4AS, UK

Antti Vasara and Jan Westerholm

Helsinki University of Technology, Department of Technical Physics, 02150 Espoo, Finland

1. Introduction

Recent progress in designing and manufacturing space-invariant optical array generators is described. We begin by demonstrating Dammann gratings [1] that generate even-numbered arrays as large as 128×128 , and odd-numbered arrays of up to 201×201 spots. The concept of a hybrid hologram [2] is applied to the fabrication of array generators, and extremely high-efficiency (close to 90%) components are obtained. Several novel types of array generators with multiple phase levels are introduced. These can e.g. reconstruct arrays with different fan-out at different angles of incidence. The application of rigorous diffraction theory to design highly efficient and compact array generators is also discussed.

2. Fourier-domain array generators

Consider a grating with period d , located between planes $z = -h$ and $z = 0$. This grating is illuminated by an obliquely incident linearly polarized plane wave

$$U_0(x, z = -h; \theta) = \exp(i 2\pi \sin \theta x / \lambda), \quad (1)$$

where θ is the angle of incidence, λ is the wavelength, and U denotes the y -component of the electric or the magnetic field, depending on the state of the polarization of the input beam.

Field distribution of Eqn. (1) gives rise to reflected and transmitted diffraction orders with amplitudes R_m and T_m , respectively, that propagate in directions predicted by the usual grating equation. The total reflected and transmitted fields U_R and U_T in planes $z = -h$ and $z = 0$, respectively, are obtained from the Rayleigh expansions (see e.g. Ref. [3]), and they are of the form

$$U_R(x, z = -h; \theta) = r(x; \theta) U_0(x, z = -h; \theta); \quad U_T(x, z = 0; \theta) = t(x; \theta) U_0(x, z = -h; \theta). \quad (2)$$

The reflectance and transmittance functions $r(x; \theta)$ and $t(x; \theta)$ are given by

$$r(x; \theta) = \sum_{m=-\infty}^{\infty} R_m(\theta) \exp(i 2\pi m x / d); \quad t(x; \theta) = \sum_{m=-\infty}^{\infty} T_m(\theta) \exp(i 2\pi m x / d). \quad (3)$$

In general, the amplitudes R_m and T_m must be evaluated using rigorous diffraction theory. This involves solving the Helmholtz equation inside the grating for the given permittivity profile and matching the solution to the total fields $U_0 + U_R$ at $z = -h$, and U_T at $z = 0$, using the appropriate boundary conditions. [3]

If the grating can be considered optically thin (no volume effects) and if it does not contain feature sizes comparable with the wavelength of the incident wave, approximate methods can be applied to determine $r(x; \theta)$ and $t(x; \theta)$. We are interested in cellular surface relief profiles with relief depths $h(x) = h_l$, $x \in [x_{l-1}, x_l]$, where $l = 1, \dots, L$, $x_0 = 0$, $x_L = d$. Then, by assuming that light rays pass through the grating without deflections other than those predicted by Snell's law, we obtain a scalar-theoretic prediction for the transmittance function

$$t(x, \theta) = \exp\{ik h_l[(n^2 - \sin^2 \theta)^{1/2} - \cos \theta]\}, \quad (4)$$

and for the reflectance function we have

$$r(x; \theta) = \exp(i 2k h_l \cos \theta). \quad (5)$$

Here we have neglected, for simplicity, Fresnel reflection losses of dielectric gratings and the finite reflectivity of metallic gratings.

By a suitable choice of the surface-relief profile, it is e.g. possible to equalize the intensities

$$P_m(\theta) = \frac{\sqrt{1 - (\sin \theta - m \lambda / T)^2}}{\cos \theta} |T_m(\theta)|^2 \quad (6)$$

of M central transmitted orders $m = M_1, \dots, M_2$ (reflected orders can, of course be used if the grating is metallic); the resulting component is called array generator. The design of array generators is typically performed using nonlinear optimization methods based on Fourier-optical expressions

$$T_m(\theta) = \frac{1}{d} \int_0^d t(x; \theta) \exp[-i 2\pi m x / d] dx \quad (7)$$

for the amplitudes of the transmitted (or reflected) beams.

3. Binary array generators

Separable two-dimensional binary array generators, often called Dammann gratings [1] with a relatively low fan-out (up to 16×16) are already being employed in many prototype optical computing systems. As a result of the demonstration of large arrays of optically bistable devices, a need has emerged to develop space-invariant array generators with much higher fan-out.

The problem of optimizing almost arbitrarily large Dammann grating array generators has been solved by applying the method of simulated annealing [4]: gratings with fan-out capabilities exceeding 1000×1000 have been designed. Several of these designs have been realized using an electron-beam writer capable of recording with $0.1 \mu\text{m}$ accuracy, and by using the reactive ion beam etching technique to convert the mask into a phase grating. With 2.5 mm grating period, the uniformity of the 64×64 pattern shown in Fig. 1 was measured to be $\pm 7\%$. Dammann gratings for larger arrays (128×128 and 201×201 beams) were also fabricated and comparable performance was observed.

4. Hybrid kinoforms

To achieve diffraction efficiencies higher than some 70% obtainable with binary gratings, more than two phase levels must be used. A great deal of interest has recently been paid on the design and fabrication of such multilevel or continuous-profile array generators, which usually are called kinoforms [5].

Multistep lithographic methods are capable of fabricating array generators of this type and these may prove successful in the future, although only small arrays have been demonstrated so far [6,7]. An alternative approach to be investigated here is to use the concept of a hybrid hologram [2]. This is an optically recorded copy of a spatially filtered wavefront generated using a binary-amplitude mask that contains the desired phase information in the form of pulse width and density modulations.

We have adopted the hybrid recording technique to realize a 1×8 array generator in dichromated gelatin. The mask in which the continuous phase profile of a 96% efficient kinoform was stored in coded form was fabricated using an electron-beam writer. By copying, a hybrid element with 88% total efficiency (including reflection and absorption losses) was obtained; 95% of the transmitted light ended up in the desired eight orders. The uniformity of this hologram was $\pm 4\%$, which is only slightly inferior to the uniformity of the array generated by the amplitude mask ($\pm 3.2\%$). The method can be straightforwardly applied to generate larger (and two-dimensional) fan-out.

An additional merit of this technique is the possibility to reduce the grating period by down-imaging in the optical copying set-up. It appears that in this way smaller grating periods can be achieved than by using the lithographic alternative alone. Finally, it appears possible to employ reflective hybrid holograms in integrated planar-optical systems [8].

5. Novel types of array generators

A wide range of optical components for array generation and more arbitrary interconnections can be designed by going beyond the approximations of Fourier optics. According to Eqn. (4), the phase delay profile caused by a thin kinoform (and therefore the image it reconstructs) depends on the angle of incidence, which makes it possible to reconstruct different images from a single hologram by illuminating it at different angles of incidence. A number of solutions of this type have been calculated using nonlinear optimization methods. For example, three arrays with 3, 5, and 7 equal-intensity beams can be reconstructed from a kinoform with 25 cells by illuminating it at normal incidence, at 45° incidence, and at 60° incidence, respectively. The efficiencies of the images are 58%, 72% and 85%, neglecting the reflection losses. Similarly, we have calculated gratings that reconstruct different images if illuminated by lasers with different wavelengths. Finally, it is possible to design beamsplitters that create different images in transmitted and reflected light.

Going beyond scalar diffraction theory, binary gratings that generate arrays with close to 100% diffraction efficiency can be designed. The period is chosen small enough to ensure that all but the desired diffraction orders are evanescent, and numerical methods of rigorous diffraction theory are used to equalize the intensities. So far a three-beam lamellar (Dammann) grating has been designed with an efficiency exceeding 99%, including reflection losses at the grating-air boundary. Unfortunately, the fabrication of array generators of this type is extremely difficult for visible or near-infrared wavelengths, since the minimum feature size is of the order of the wavelength of the illuminating beam. These problems could be solved in the future by the development of projection X-ray lithography.

Acknowledgments

We wish to thank J. Huttunen, H. Ichikawa, J. M. Miller, E. Noponen, and E. Restall for their contributions to the experimental and theoretical aspects of this paper.

References

- [1] H. Dammann and K. Gortler, *Opt. Commun.* **3**, 312 (1971).
- [2] H. Bartelt and S. K. Case, *Appl. Opt.* **21**, 2886 (1982).
- [3] D. Maystre, in *Progress in Optics XXI*, ed. E. Wolf (North Holland, Amsterdam), 1 (1984).
- [4] J. Turunen, A. Vasara, J. Westerholm, G. Jin, and A. Salin, *J. Phys. D: Appl. Phys.* **21**, S102 (1988).
- [5] L. P. Lesem, P. M. Hirsch, and J. A. Jordan, *IBM J. Res. Dev.* **13**, 150 (1969).
- [6] S. J. Walker and J. Jahns, *J. Opt. Soc. Am. A* **7**, 1509 (1990).
- [7] H.-P. Herzig, D. Prongue, and R. Dandliker, *Jap. J. Appl. Phys.* **29**, 1307 (1990).
- [8] J. Jahns and A. Huang, *Appl. Opt.* **28**, 1602 (1990).

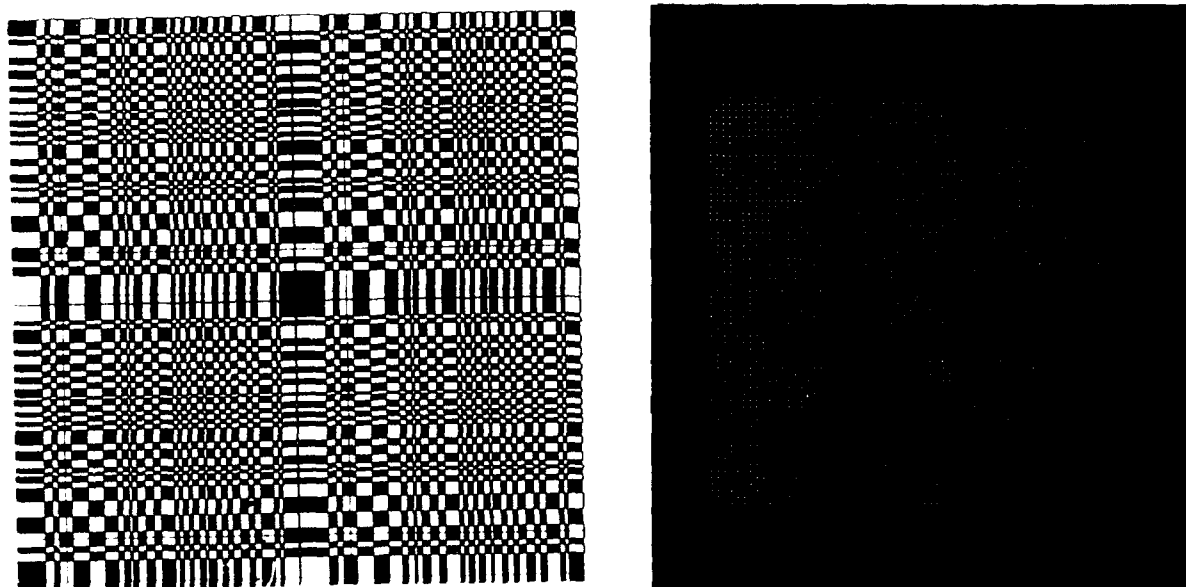


Figure 1. The structure of one period of a Dammann grating designed to generate an array of 64×64 equal-intensity beams, and the array reconstructed by this grating.

Design rules for fanout elements recorded as volume holograms

H. P. Herzig, D. Prongué, P. Ehbets, R. Dändliker

Institute of Microtechnology, University of Neuchâtel, Rue A.-L. Breguet 2,
CH-2000 Neuchâtel, Switzerland

Introduction

Optical fanout elements split a single laser beam into a regular array of equally intense light spots in one- or two-dimensions. They are used in many applications of modern optics, such as parallel optical processing and fiber optic communication. This paper deals with the recording of efficient fanout elements as volume holograms. We have applied coupled wave theory to determine how efficiency and uniformity of the fanout depend on the recording conditions.

Volume holograms as fanout elements

A fanout element can be fabricated by recording a hologram of N object waves with a reference wave. The waves are characterized by the wavevectors \mathbf{k}_i , the amplitudes A_i and the phases ϕ_i , where $i = 0$ denotes the reference beam (Fig. 1). Besides the desired N gratings $\mathbf{K}_{i0} = \mathbf{k}_i - \mathbf{k}_0$, unwanted intermodulation gratings $\mathbf{K}_{ij} = \mathbf{k}_i - \mathbf{k}_j$ ($i, j \neq 0$), between the object beams are also recorded. At readout, they generate intermodulation waves, which are coupled with the desired reconstructed beams through the primary gratings \mathbf{K}_{i0} . For the arrangement shown in Fig. 1 these off-Bragg interactions are important for a hologram thickness t smaller than

$$t < \lambda / (n \tan \beta \Delta \alpha), \quad (1)$$

where λ is the wavelength, β the reference beam angle and $\Delta \alpha = \alpha / (N - 1)$ the interbeam angle. As a consequence the diffraction efficiency and the uniformity of the fanout are reduced. In addition, at recording the intermodulations produce a spatially variable irradiance of the object wave field,

which requires a large dynamic response of the recording material. For regular fanouts, the intermodulation gratings, and also the energy exchange between intermodulation beams and reconstructed object beams, depend strongly on the relative phases of the N object waves.

High efficiency is predicted in the literature for high enough reference-to-object beam ratios (> 5) [1]. For strongly periodic fanout elements, this is only true if the fanout angle is sufficiently large and the emulsion is sufficiently thick (e.g. $\alpha > 5^\circ$ for $t = 15 \mu\text{m}$), then the losses due to the intermodulation gratings become negligible. Recording with a high reference-to-object beam ratio requires a large dynamic response of the holographic material, which is difficult to get. The intermodulations can be reduced by optimizing the relation of the phases ϕ_i between the N object waves [2]. High efficiency can now be obtained with much

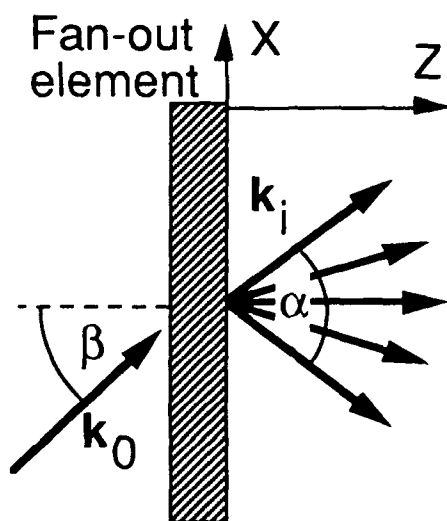


Fig. 1. Volume hologram as fanout element. Angles are defined inside the medium of refractive index n .

lower reference-to-object beam ratios, yielding higher diffraction efficiency in the same material. In the following, we will summarize the important parameters for successful recording of fanout elements as volume holograms.

Dynamic range

A holographic emulsion has a limited dynamic range, which depends on the material and the thickness. For an optimum hologram recording the exposure energy has to be within the dynamic range. If the exposure is increased above saturation, the emulsion will not generate a higher index modulation. For a high reference-to-object beam ratio and for large fanouts the limited dynamic range becomes important. We have compared the recording of one object wave with the case of N object waves. For the same efficiency, the optimized case requires the same total exposure energy. In the non-optimized case, however, the same efficiency would require an exposure energy which is \sqrt{N} times higher. The maximum fanout will then be limited by the saturation level.

Depth of the optimum plane

We have shown that the intermodulation gratings can be suppressed nearly perfectly [2]. However, this is only true for specific planes in the z -direction (Fig. 1). Nevertheless, the intermodulation remains small within a depth of

$$h < \lambda / [5n (1 - \cos(\alpha/2))], \quad (2)$$

where λ is the wavelength and α the full angle ($\mathbf{k}_1, \mathbf{k}_N$) of the fanout (Fig. 1). The consequence of this limited depth is that the hologram plane must be normal to the z -axis, within the tolerances given by Eq. (2). This restricts the permitted recording geometry. In the following, we will present examples of favorable recording geometries to fabricate highly efficient on- and off-axis fanout holograms. For very thick holograms [$t > \lambda / (n \tan \beta \Delta \alpha)$, Eq. (1)], the off-Bragg interactions become negligible and therefore the phases of the object waves become irrelevant.

Optimized fanout elements

Figure 2 shows the recording set-up for on-axis fanout elements. The object is an array of coherent sources with optimized phases ϕ_i . Such an array can be obtained by different techniques; e.g. using phase plates, CGH, kinoforms [2]. The hologram is placed in the far-field, i.e. in the Fourier plane of a lens. If the sources are in the front focal plane of the lens ($d = f$), the recorded element becomes non-focussing, and for larger object distances ($d > f$), it becomes focussing. Figure 3 shows the off-axis equivalent. Diffraction efficiencies over 90 % can be obtained with good uniformity. We will demonstrate experimental results for fanout elements recorded in dichromated gelatine. The results for optimized and non-optimized elements will be compared.

Recording without lens

If the lens is removed in Fig. 2, we get a focussing fanout element. But the optimized phases will generate a uniform illumination only if the hologram is in the far-field, i.e. at a distance $d > (Ns)^2/\lambda$ (Fig. 4). Another method to fabricate focussing fanout elements without a lens uses the self-imaging properties of large periodic structures [3]. In this case the object is a regular array of coherent sources with identical phases $\phi_i = 0$. Considering the beam propagation in free space, planes of reduced intermodulations are found, which are suitable for recording efficient holo-

grams. These planes are parallel to the object plane (Fig. 4), as in the case of optimized phases. Thus, if we incline the object, we have also to incline the hologram (Fig. 5). The self-imaging (Talbot) distance is proportional to s^2 and for inclined objects proportional to $(s \cos\theta)^2$. If we incline a regularly spaced 2-D array (same spacing s in x and y direction) with respect to the y -axis, we get two different distances for the optimum planes, depending on s^2 and $(s \cos\theta)^2$ respectively. However, a common minimum plane can be determined. This problem can be avoided if the initial array has two different periods Λ in the x and the y direction, namely $\Lambda_x = s$ and $\Lambda_y = s/\cos\theta$.

Due to the self-imaging properties, this method is only suitable for very large arrays. Note that depending on the position of the recording plane, this element becomes either a fanout (overlapping) or a lenslet array (non-overlapping beams).

Copying of fanout elements

Fanout elements in volume holograms can also be fabricated by copying the phase structure of already existing fanout element as e.g. Dammann gratings. But one has to take care of the fact that a simple image formation with a single lens would destroy the phase structure and therefore the properties of the fanout. This can be avoided by using a 4-f imaging system as shown in Fig. 6, which applies twice a Fourier transform thereby conserving the phase distribution [4]. Analog to Fig. 2 and 3, there exists also an off-axis arrangement of Fig. 6.

Conclusions

We have investigated the recording of efficient fanout elements as volume holograms by using coupled wave theory. In opposition to the results published in the standard literature, we have found that the efficiency and uniformity of regularly fanout elements depend strongly on the relative phases of the object waves, if the thickness t of the holographic emulsion is smaller than $t < \lambda/(\gamma \tan\beta \Delta\alpha)$, [Eq. (1)], e.g. $t < 32 \mu\text{m}$ for $\lambda = 488 \text{ nm}$, $n = 1.5$, $\beta = 30^\circ$ and $\Delta\alpha = 1^\circ$. High efficiency and uniformity can be achieved by optimized phases of the object beams, thereby requiring a low dynamic range of the holographic material.

The recording conditions are optimum if the irradiance of the object beam is uniform in the hologram plane. This can be achieved only in specific planes, which are parallel to the object plane. As a consequence only specific recording geometries are allowed. Several successful recording techniques to fabricate efficient and uniform fanout elements are presented.

References

- [1] R. K. Kostuk, "Comparison of models for multiplexed holograms," *Appl. Opt.* **28**, 771-777 (1989).
- [2] H. P. Herzig, D. Prongué, R. Dändliker, "Design and fabrication of highly efficient fanout elements," *Jpn. J. Appl. Phys.* **29**, L 1307-L 1309 (1990).
- [3] I. Seyd-Darwish, P. Chavel, J. Taboury, Y. Malet, "Array illuminator hologram based on the Talbot effect," *Conference Record of 1990 International Topical Meeting on Optical Computing*, Kobe, 294-296 (April, 1990).
- [4] B. Robertson, M. R. Tagizadeh, J. Turunen, A. Vasara, "High-efficiency, wide-bandwidth optical fanout elements in dichromated gelatin," *Opt. Lett.* **15**, 694-696 (1990).

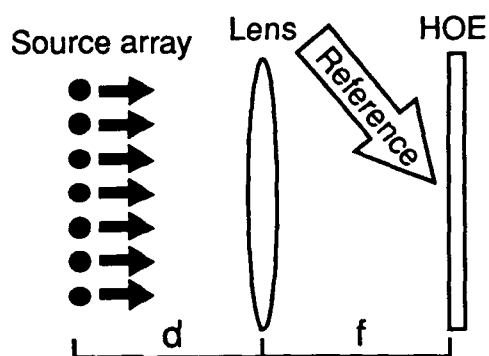


Fig. 2. Object waves on-axis, the HOE becomes focussing for $d > f$ and non-focussing for $d = f$.

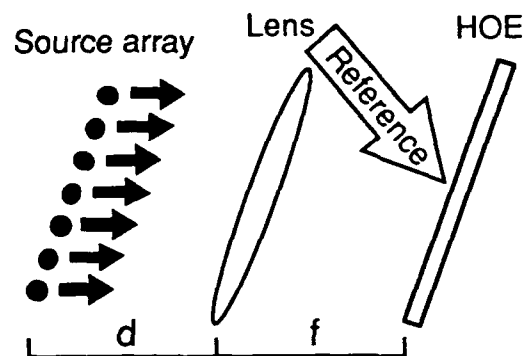


Fig. 3. Object waves off-axis.

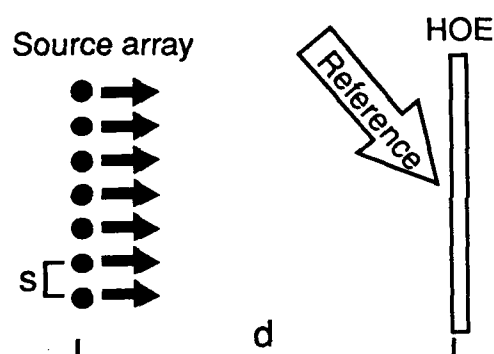


Fig. 4. Spherical object waves on-axis.

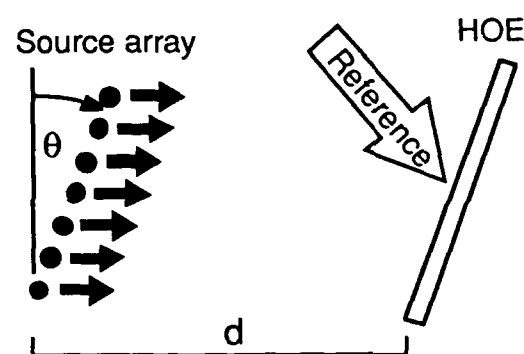


Fig. 5. Spherical object waves off-axis.

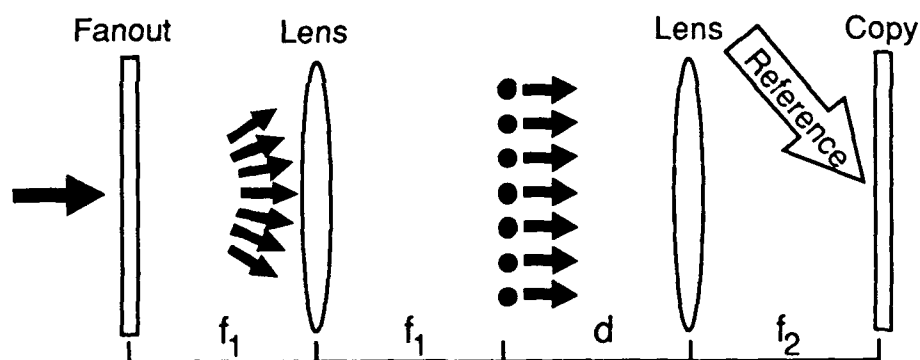


Fig. 6. 4f-system for copying fanout elements with magnification $m = f_2/f_1$, where $d = f_2$. Focussing power can be included if $d > f_2$.

A polarization metrology for optical interconnects which use polarization beam combining

J. Larry Pezzaniti and Russell A. Chipman
University of Alabama in Huntsville
Dept. of Physics RI D-4
Huntsville, AL 35899

1. Introduction

Several free space optical interconnects for digital optical computing which use polarization beam combining are currently being implemented.¹⁻⁵ These architectures interconnect 2-D arrays of optical logic devices by imaging arrays of spots, generated by binary phase gratings, from one logic device to the next. Polarization beam combining addresses the need to combine input beams and separate output beams by using space-variant mirrors in conjunction with polarizing beam splitters and waveplates. The throughput of the interconnect is limited primarily by the polarizing beam splitters and the waveplates.

Several problems inherent to these polarizing elements contribute to the loss of throughput. First, system designs require that the beams focus through the waveplates and polarizing beam splitters. This means that for optimal performance the optical characteristics of the polarizing elements should remain constant over a large angular range (about 10°). Typically, however, the polarizing elements display angular dependence. Second, since the beams pass through a large cross-section of the polarizing element, the element must be uniform. Third, problems such as crystal axis misorientation in the waveplates or thin film thickness variations associated with element fabrication may exist. Finally careful alignment of the polarizing elements in the system must be performed in order to optimize throughput.

The work presented here is motivated by the lack of comparative data on the needed polarizing elements over the required angular range, and the need for an instrument to align a large field of view, polarization critical optical system. The authors in conjunction with AT&T Bell Labs and Air Force Office of Scientific Research have built an imaging polarimeter which is able to address these issues.⁶ After a short summary of polarization beam combining, we present measurements on a polarizing beam splitter cube performed with the imaging polarimeter. Following this, we briefly outline a technique to align a polarization optical interconnect using the imaging polarimeter.

2. Polarization Beam Combining

Figure 1 shows a polarization beam combining system which allows four-port access to the s-SEED. From the left, two signal beams in orthogonal polarization states reflect and pass through the polarizing beam splitter cube according to their polarization state. The beams are imaged onto the reflective portions of the patterned reflectors after passing through the $\lambda/4$ plates which transforms them into left and right handed circular states. After reflection, their handedness is inverted. Passing back through the wave plates converts their polarization to linear so that the beams from mirror array 1 pass through the polarizing beam splitter and the beams from mirror array 2 reflect. Thus the signal inputs are imaged onto the s-SEED in orthogonal polarization states. The power beams are imaged between the mirror portions of mirror array 1 onto the s-SEED. The output power beams are reflected off the s-SEED, and the polarization is transformed again by the waveplate so that it reflects off the polarizing beam splitter and the beams exit the module through the spaces in mirror array 2. Note that the patterned reflectors and s-SEED array are in image planes so that the system must focus the spot arrays through the polarizing elements. Accordingly, the behavior of the polarizing elements over 0.174 NA (10° field of

view) should be determined. Furthermore the polarization behavior of many polarizing elements in sequence, such as a complete logic module consisting of ten polarizing beam splitters and twelve quarter wave retarders, should be understood.

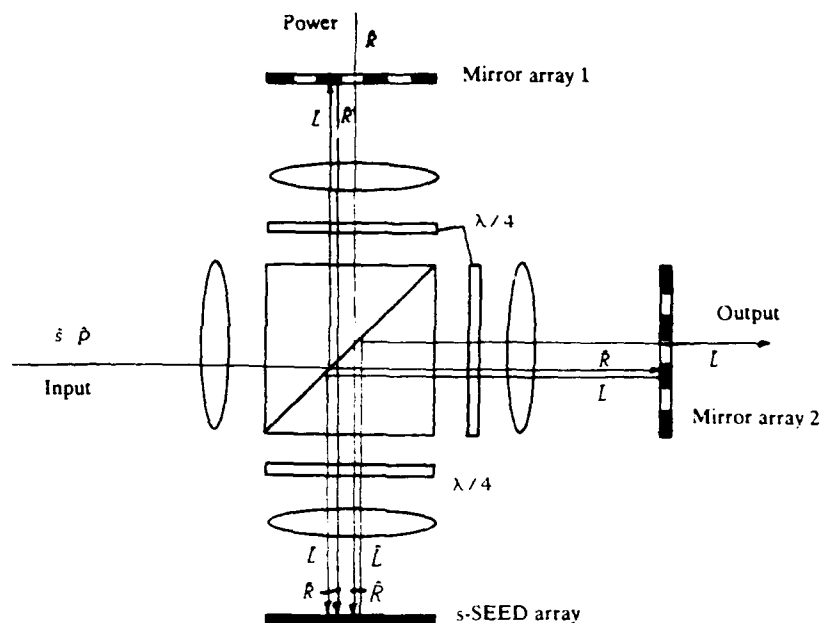


Figure 1 Polarization beam combining using polarizing beam splitter cubes and quarter wave plates.

3. Beam splitter cube analysis

Typically a polarizing beam splitter cube provides a large spectral range but a small angular range. Digital optical computer architectures need just the opposite. Research has been done to determine the conditions necessary to spread out the angular field to $\pm 5^\circ$ while keeping good optical characteristics.⁷ The behavior of this cube over a $\pm 10^\circ$ field of view is shown in fig. 2. Figure 2a is a pupil map of a polarizing beam splitter cube in transmission with incident cone of converging light in p polarization. The outermost edge of the map represents ray angles of 10° and the center represents an axial ray. The pupil map gives the percent transmission for each ray of light that remains in p polarization. This is the amount of light that will continue to propagate in the correct direction and polarization. The remaining light will follow an incorrect path the next time it encounters the polarizing beam splitter cube. Figure 2b is a pupil map showing percent reflection of the same cube with s light incident. Again the percents shown correspond to the percent of each beam that will continue to take the correct path at the next beam splitter.

The imaging polarimeter is presently measuring the performance parameters including throughput in reflection and transmission of a polarizing beam splitter. The block diagram in fig. 3 shows the system design of the imaging polarimeter used to measure a polarizing beam splitter. In this configuration a spherical wave with a set polarization state propagates through the beam splitter and the resultant intensity/ polarization is analyzed. The intensity/polarization is analyzed by the imaging polarimeter consisting in this instance of a rotating polarizer in front of a ccd camera. Each pixel of the ccd corresponds to a ray path through the beam splitter. By capturing a set of images for different incident states of polarization and different orientations of the rotating polarizer, a complete characterization of the polarizing beam splitter is obtained. Our plan is to

measure the performance of a number of beam splitters from different manufacturers to compare the performance of each in terms of field of view and uniformity to determine which ones will work best in the digital optical computer.

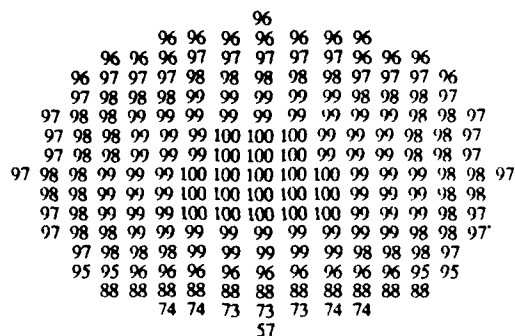


Figure 2a Pupil map is of a polarizing beam splitter cube in transmission with incident 10° cone of light in p polarization state done with polarization ray trace on Code V lens design software. Plane of incidence is vertical in above map and multilayer interface is tilted so that the bottom of plot is closest to the first surface of the cube.

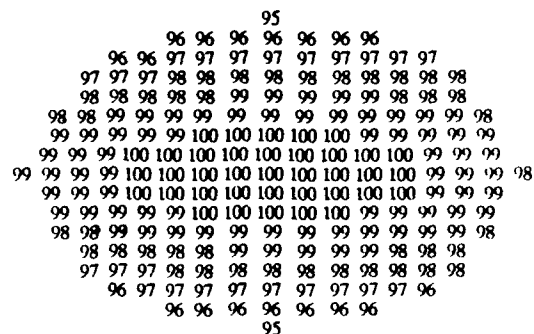


Figure 2b Pupil map is of a polarizing beam splitter cube in reflection with incident 10° cone of light in p polarization state done with Code V lens design software. Plane of incidence is vertical in above map and multi-layer interface is tilted so that the bottom of plot is closest to the first surface of the cube.

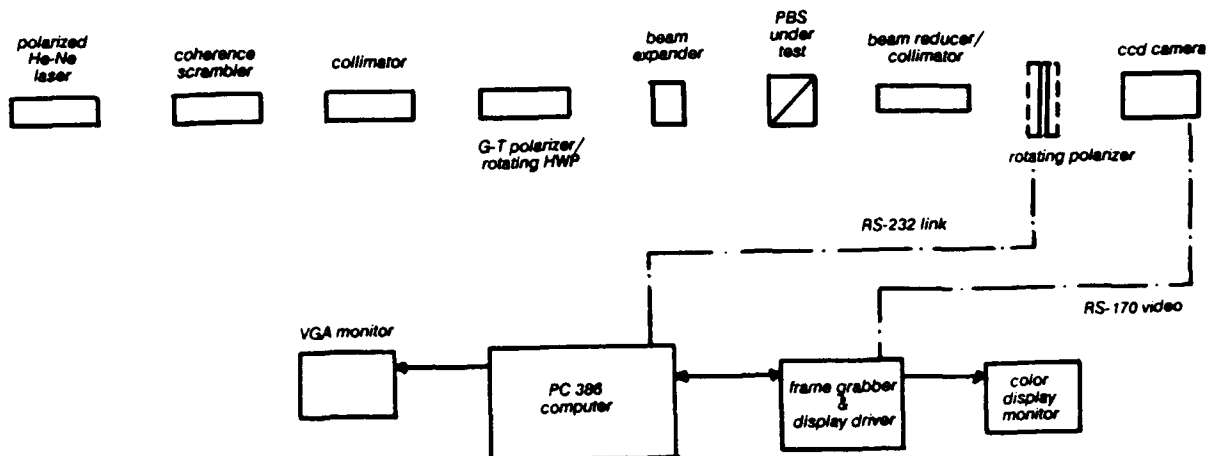


Figure 3 Polarizing beam splitter test block diagram.

4. Alignment of digital optical computer using the imaging polarimeter

Careful alignment of the logic modules must be performed so that the spots hit the correct places on the patterned reflectors and the s-SEEDs. Furthermore the azimuthal orientation of the waveplates must be adjusted to insure that the system is operating with the greatest possible throughput. The imaging polarimeter can be used to maximize the amount of light that is in the correct polarization state at each stage of the system. The procedure, in short, is to aim the imaging polarimeter directly into the output of a certain stage of the logic module and set the polarization analyzer orthogonal to the state of polarization transmitted from that part of the stage. Next adjust the orientations of the wave plates until optimal extinction is observed over the entire pupil. This maximizes the amount of light that is in the correct polarization state. By doing this at each stage of the logic module the entire system can be tweaked to arrive at the maximum output.

5. Summary

The throughput of optical interconnects using polarization beam combining is limited by the quality of the polarizing beam splitters and wave plates used in its design. The imaging polarimeter is a polarimetric metrology tool that will acquire comparative data on polarization components to determine which ones will perform best in the digital optical computer. The imaging polarimeter is needed to align and to understand the propagation of polarized light through systems with large numbers of polarizing beam splitters and retarders.

7. Acknowledgements

We would like to acknowledge the Air Force Office of Scientific Research AFOSR (contract #AFOSR-89-0542) for their support in this work. We would also like to thank AT&T Bell Labs and Michael Prise for their assistance.

6. References

1. F.B. McCormick and Michael E. Prise, "Optical circuitry for free-space interconnections," *Appl. Opt.* **29**, 2013-2018, (1990).
2. M.E. Prise, M.M. Downs, F.B. McCormick, S.J. Walker, and N. Steibl, "Design of an optical digital computer," in *Optical Bistability IV*, W.J. Firth et al. Eds. (Les Editions de Physique, Paris 1988), pp. C2-15.
3. M.E. Prise, N.C. Craft, M.M. Downs, S.J. Walker, L.A. D'Asaro, and L.M.F. Chirovsky, "A cascable optical logic module using Symmetric Self-Electrooptic Effect Devices," in *Technical Digest, Topical Meeting on Photonic Switching* (OSA, Washington, D.C., 1989), postdeadline paper PD5-1.
4. M.E. Prise, R.E. LaMarche, N.C. Craft, M.M. Downs, S.J. Walker, L.M.F. Chirovsky, and L.A. D'Asaro, "Optical systems using arrays of Symmetric Self-Electrooptic devices," in *Technical Digest, OSA Annual Meeting* (Optical Society of America, Washington, D.C., 1989), paper MQ3.
5. M.E. Prise, N. Streibl, and M.M. Downs, "Optical considerations in the design of a digital computer," *Opt. Quantum Electron.* **20**, 49-77 (1988).
6. J.L. Pezzaniti, R.A. Chipman, "Imaging polarimeters for optical system metrology," *Proc. SPIE* **1317**, (1990).
7. Jacques Mouchart, Jacqueline Begel and Eugene Duda, "Modified Macneille cube polarizer for a wide angular field," *Appl. Opt.* **28**, 2847-2853, (1989).

Photorefractive Parallel Matrix-Matrix Multiplier Using a Mutually Incoherent Laser Array

John Hong and Pochi Yeh
Rockwell International Science Center
1049 Camino dos Rios, A25A
Thousand Oaks, CA 91360

Summary

Wave mixing and holographic recording in photorefractive media have been used to perform parallel matrix-vector multiplication [1]. Although the technique can be extended to perform parallel matrix-matrix multiplication, the implementation requires critical alignment of the matrix elements. In this paper, we propose to demonstrate a novel technique to perform parallel matrix-matrix multiplication which uses the simultaneous formation of multiple gratings in a photorefractive crystal. The concept is shown in Fig.1 where a simple example is used. The matrices to be multiplied are given by a_{mn} and b_{kl} . While all the light sources shown in the figure are of the same nominal wavelength, each source differs from its neighboring source by some frequency $\delta\omega$ which is chosen to satisfy $\delta\omega \gg 1/\tau$ where τ is the photorefractive response time. Denoting each source frequency by ω_m , the optical amplitude distribution immediately following SLM1 which contains the matrix a_{mn} is given by

$$E_{mn} = a_{mn} \exp(i\omega_m t), \quad (1)$$

where (m,n) denote discrete (pixel) spatial variables. The same source array is rotated by 90° and is directed through SLM2 in like fashion to yield $b_{kl} \exp(i\omega_k t)$ (note the orientations of the two matrices). By imaging each distribution through a slit as shown, we obtain the distributions

$$\sum_m a_{mn} \exp(i\omega_m t), \text{ and } \sum_k b_{kl} \exp(i\omega_k t) \quad (2)$$

at the crystal plane where (l,n) are the two output coordinate variables. The steady state amplitude of the grating formed in the crystal is proportional to the time average of the product of these two distributions [2]:

$$\Delta n \propto \left\langle \sum_m a_{mn} \exp(i\omega_m t) \sum_k b_{kl}^* \exp(-i\omega_k t) \right\rangle_\tau = \sum_{m,k} a_{mn} b_{kl}^* \langle \exp[i(\omega_m - \omega_k)t] \rangle_\tau = \sum_m b_{ml}^* a_{mn}, \quad (3)$$

where $\langle \rangle_\tau$ indicates a time averaging operation with integration time τ . This is the desired matrix-matrix product which can be accessed holographically. By simply increasing the intensity and/or using fast crystals such as GaAs with the condition $\delta\omega \gg 1/\tau$ satisfied, this parallel method can be potentially very fast but most importantly, the latency and throughput rate of the system are no longer functions of the matrix size.

Although the simplest and perhaps the most elegant implementation of a mutually incoherent laser array is an integrated array of laser diodes (such as the emerging surface emitting laser diode arrays [3]), we have opted to use a novel scheme involving an acoustooptic device to achieve the desired result. The implementation is shown in Fig.2 where an acoustooptic device of sufficient length is illuminated with a collimated sheet laser beam. The device is driven with a frequency swept sinusoid (FM-chirp) in such a way that each point along the length of the device modulates the light with a different frequency at any particular instant. By carefully controlling the bandwidth and rate of the FM chirp, one can easily implement a large array limited by the number of resolvable deflection spots of the acoustooptic device which can be as high as 1000.

A simple experiment was performed to validate the acoustooptic concept just described using the apparatus shown in Fig. 3a. As shown, a collimated laser beam is modulated spatio-temporally with an acoustooptic device and the diffracted beam is split into two paths. One beam is directed into a photorefractive crystal (SBN: Strontium Barium Niobate) while the other is rotated spatially by 90° with a dove prism before falling on the crystal. A third beam is used to read the grating in a four wave mixing geometry. When the acoustooptic device is driven with a CW tone ($f_0=70\text{MHz}$), the diffracted beam from the acoustooptic device is spatially coherent so that a uniform grating is written. The diffracted beam pattern is shown in Fig. 3b for this uniform case. When the drive signal to the acoustooptic device is modulated by a frequency chirp signal ($\delta f=10\text{KHz}$), however, the diffracted beam from the acoustooptic device is no longer spatially coherent with each point oscillating at a slightly different frequency. The hologram written by the

two beams are non-uniform with only the diagonal portions being nonzero in the steady state since the frequencies of the two beams are equal in those areas as shown in Fig. 3c.

This work is sponsored, in part, by DARPA under contract F49620-87-C-0015.

References

1. P. Yeh and A. E. T. Chiou, *Opt. Lett.* 12, p.138 (1987).
2. N. V. Kukhtarev, V. B. Markov, S. G. Odulov, M. S. Soskin and V. L. Vinetskii, *Ferroelectrics*, 22, p.949 (1979).
3. J. L. Jewell, A. Scherer, S. L. McCall, Y. H. Lee, S. Walker, J. P. Harbison, and L. T. Florez, *Electronics Lett.* 25, p.1123 (1989).

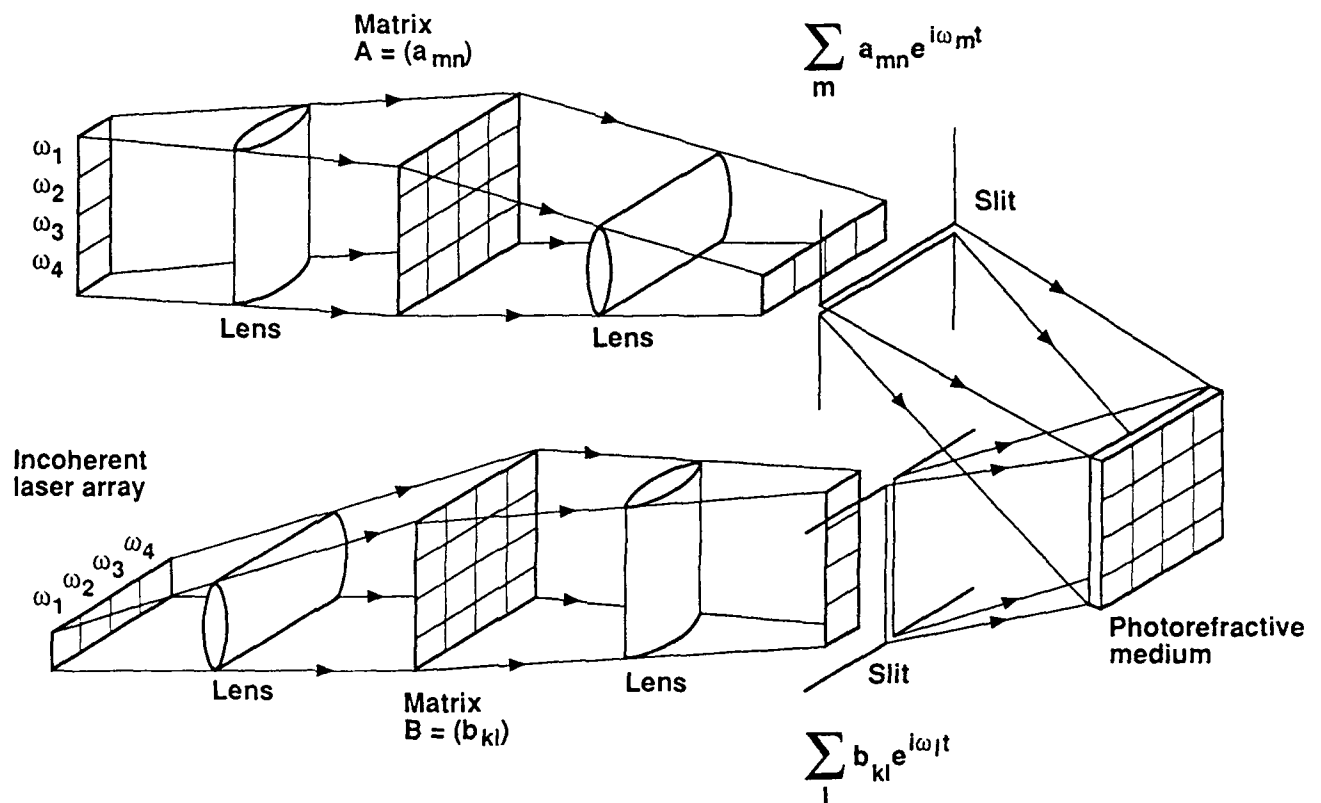


Fig. 1 Parallel Matrix Matrix Multiplier
(the matrix product is written in the photorefractive crystal and readout using a coherent plane wave not shown)

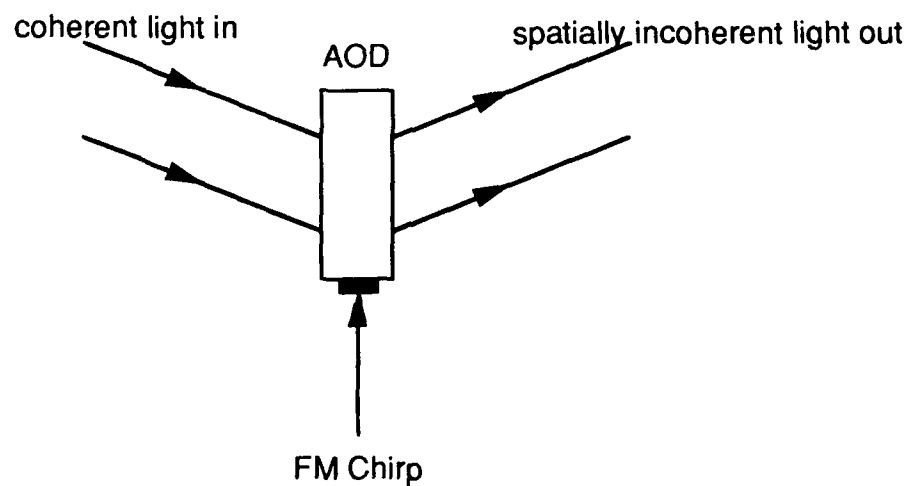
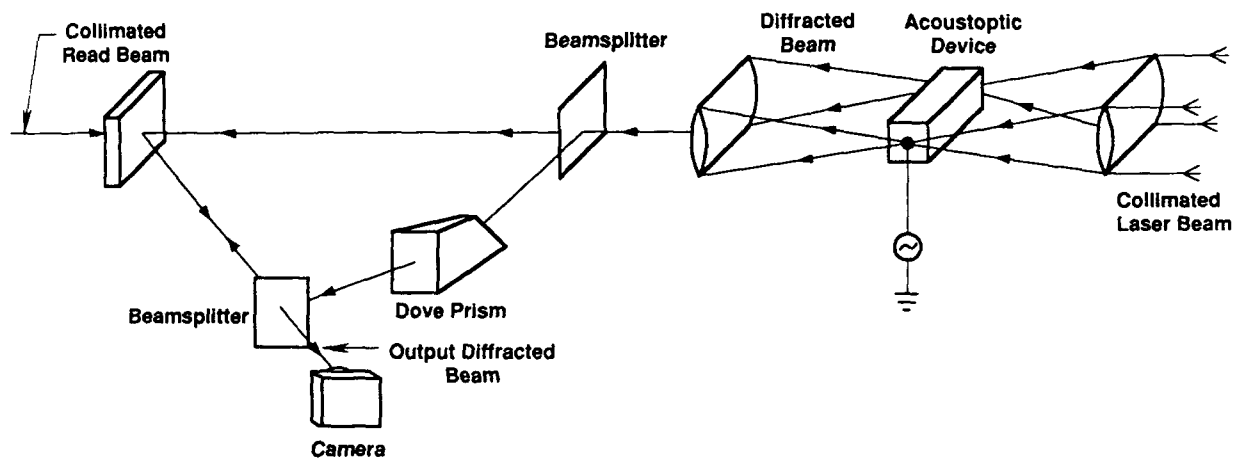
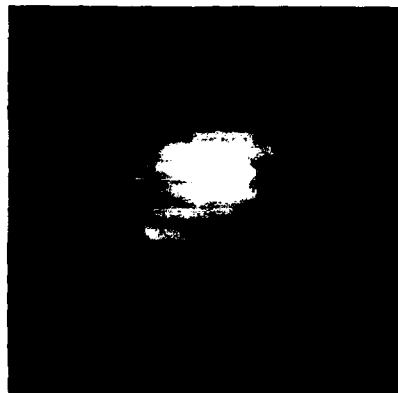


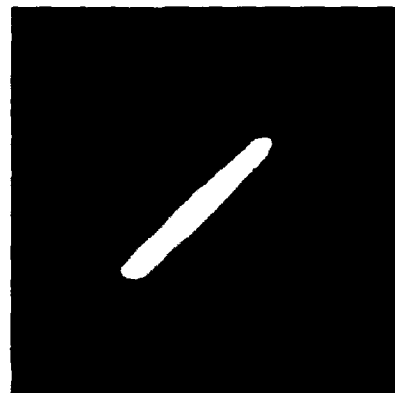
Fig. 2 Acoustooptic Implementation of Mutually Incoherent Laser Source Array



a)



b)



c)

Fig. 3
Mutually Incoherent Source Array Experiment
a) experimental set-up
b) uniform grating written by coherent beams
c) gratings written by incoherent source array from
acousto-optic device driven by FM chirp signal

A FIGURE OF MERIT FOR PATTERN RECOGNITION FILTERS

Ph. REFREGIER

Laboratoire Central de Recherches

Thomson-CSF 91404 ORSAY (cedex) France

TEL : 33 1 60 19 71 19, FAX : 33 1 60 19 74 16.

The design of filters for optical pattern recognition is extensively studied for more than ten years [1] [2] [3] [4] [5]. Indeed, optical correlation implementation leads to specific constraints in comparisons with classical signal processing technics [6]. In particular, although the spatial matched filter is optimal for noise robustness, its limitations such as broad correlation peaks and low diffraction efficiency [3] are well known. Many different approaches have improved some of these characteristics and a lot of work is still devoted to find trade-offs between them [7] [8] [9] [10] [11] [12].

It is therefore important to enable fair comparisons among the various filters, and then to define precise performance criteria and to find mathematical optimal filters as a standard for figure of merit. A very interesting attempt in this way has been recently performed, in which it is [13] proposed to analyze the possible trade-offs between the different criteria emphasizing that some of the most interesting are related to the Horner efficiency, the sharpness of the correlation function and the noise robustness of the filter.

Independently, this problem has been investigated in the context of Synthetic Discriminant Function filters [14] and optimal trade-off filters (OTF) for two of these criteria have been found. However, the optimization of the Horner efficiency leads to a very non-linear problem and optimal trade-offs including this criterion have only recently been found [15] for the detection of only one object. These OTF between noise robustness, sharpness of the correlation peak and Horner

efficiency, allow a rigorous characterization of filter performances which will be illustrated.

Let us note x_i the value of the i^{th} pixel in image x we want to detect, \hat{x}_k the value at frequency k of its Fourier transform, \hat{h}_k the value for the filter at the same frequency and N the total number of pixels of x .

The first criterion considered is the **Signal-to-Noise-Ratio (SNR)** :

$$SNR = |C_0|^2 / MSE \quad (1)$$

where the Mean Square Error (MSE) is defined by : $MSE = \hat{h}^\dagger \hat{S} \hat{h}$, \hat{S} being the spectral density of the noise with zero mean, and where $C_0 = h^\dagger \cdot x$, i.e. the central value of the correlation function (denoted C_i or \hat{C}_k in the Fourier domain). Maximizing the SNR is equivalent to optimizing the noise robustness when the input image is corrupted by noise. Optimization of this criterion leads to the well known matched filter. However, in this case, only the central value of the correlation function is considered which can results in large correlation peak (i.e., low peak-to-sidelobe ratio) and then in false detections.

Optimization of the **Peak to Correlation Energy (PCE)** [13] allows to overcome this problem :

$$PCE = |C_0|^2 / CPE \quad (2)$$

where the Correlation Plane Energy (CPE) is defined as : $CPE = \sum_{i=1}^N |C_i|^2$. It is easy to demonstrate that the inverse filter ($\hat{h}_k = \hat{x}_k / |\hat{x}_k|^2$) optimizes this criterion.

For optical implementation, it is necessary to consider a third criterion [3] which characterizes the relative amount of the input

is not a very efficient trade-off for the considered criteria, although it could have other interesting properties (simple amplitude coding).

Other interesting nonlinear transformations to the matched filter have been recently proposed [13] [8]. In these articles, the authors proposed to apply a power law nonlinearity to the filter :

$$\hat{h}_k = \begin{cases} |\hat{x}_k|^p \hat{x}_k / |\hat{x}_k| & \text{if } |\hat{x}_k| \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

We adopt the notations of [13] for this fractional power filters (FPF) since in this work the goal was also to provide trade-offs in filter design. As noted by Kumar *et al.* the matched filter, the POF and the inverse filter are obtained with $p = 1, 0, -1$.

The values of the preceding criteria for these trade-offs filters are shown in Fig.2. We see that whatever the value of p , the results are suboptimal for the considered criteria. In the insert of Fig.2, a zoom is presented in order to examine precisely the different filters for a fixed value of the correlation peak (the value of C_0 is 0.66 the one obtained with the POF). We see that OTF allow to obtain better SNR for the same PCE values for the filters or better PCE for the same SNR values.

In conclusion, we have proposed optimal trade-off filters for pattern recognition with explicit solutions. They provide a rigorous way for the evaluation of different filters by comparison with the figure of merit drawn by the OTF. We have successfully illustrated these results with examples and others will be presented. We have shown that classical filters are in general overspecialized. We believe that these results are general but more simulations are necessary to confirm this point.

The author acknowledges J-P. Huignard for his support in this work and H. Rajbenbach and S. Maze for their enlightening discussions and suggestions.

References

- [1] H.J. Caulfield and W.T. Maloney. *Applied Optics*, 8(11):2354-2355, 1969.
- [2] D. Casasent and D. Psaltis. *Appl. Opt.*, 15:1795-1799, 1976.
- [3] J.L. Horner. *Appl. Opt.*, 21(24):4511-4514, 1982.
- [4] Y.N. Hsu and H.H. Arsenault. *Appl. Opt.*, 21(22):4016-4019, 1982.
- [5] J.L. Horner and P.D. Gianino. *Appl. Opt.*, 23(6):812-816, 1984.
- [6] D.F. Elliott, editor. *Handbook of digital signal processing*. Academic Press, Inc., San Diego, 1987.
- [7] M.W. Farn and J.W. Goodman. *Appl. Opt.*, 28(15):3362-3366, 1989.
- [8] B. Javidi. *App. Opt.*, 29(8):1215-1224, 1990.
- [9] M.F. Dickey, K.T. Stalker and J.J. Mason. *Appl. Opt.*, 27(18):3811-3818, 1988.
- [10] B.V.K. Vijaya Kumar and K. Bahri. *Appl. Opt.*, 28(2):250-257, 1989.
- [11] M.F. Dickey, B.V.K. Vijaya Kumar, A.L. Romero and M.J. Connelly. *Opt. Eng.*, 29(9):994 - 1001, 1990.
- [12] S.A. Awwal, A.M. Karim, and R.S. Jahan. *Appl. Opt.*, 29:233 -236, 1990.
- [13] B.V.K. Vijaya Kumar and L. Hassebrook. *Appl. Opt.*, 29(20):2997-3006, 1990.
- [14] Ph. Réfrégier. *Opt. Lett.*, 15(15):854-856, 1990.
- [15] Ph. Réfrégier. submitted to *Opt. Lett.* 1990.
- [16] Ph. Réfrégier and J-P. Huignard. *Appl. Opt.*, 1990. (to be published).
- [17] B.V.K. Vijaya Kumar and K. Bahri. *Appl. Opt.*, 28(10):1919-1925, 1989.

light which will be detected in the correlation plane. This is quantitatively characterized by the **Horner efficiency** η_H :

$$\eta_H = \frac{\sum_{i=1}^N |C_i|^2}{\sum_{i=1}^N |x_i|^2}, \quad (3)$$

where the filter is constrained to $|\hat{h}_k| \leq 1$. This criterion is optimized with Phase-Only-Filters (POF) among which the choice $\hat{h}_k = \hat{x}_k/|\hat{x}_k|$ provides the highest intensity correlation peak.

An OTF is defined by the fact that for each possible values of two of the preceding criteria it is not possible to find a filter with a better value for the third criterion. Maximization of the SNR, the PCE and η_H is equivalent to the minimization of the MSE, the CPE and maximization of the central value of the correlation C_0 . This last optimization problem is easier to solve mathematically and is consider in the following. We will not detail the mathematical resolution of this non-linear optimization problem, but we only give the solution for the OTF [15] :

$$\hat{h}_k = \sigma_\lambda \left[\frac{\hat{x}_k}{\mu \hat{S}_k + (1 - \mu) |\hat{x}_k|^2} \right], \quad (4)$$

where $\sigma_\lambda[y] = y$ if $|y| \leq 1/\lambda$, $\sigma_\lambda[y] = \exp(i\psi)$ otherwise (ψ being the phase of y). Two parameters μ, λ are necessary to specify these OTF. They measure the relative amount of the three considered criteria (they appear as Lagrange multipliers in the multicriteria optimization problem). It is easy to verify that the matched filter, the inverse filter or the POF are special cases of these OTF with respectively the following values for the parameters (μ, λ) : $(1, 0)$; $(0, 0)$; $(\mu, +\infty)$. Furthermore, if the Horner efficiency is not optimized the filter is equal to the well known Wiener filter for the detection of a deterministic pattern in random noise [6] (i.e. $\hat{h}_k = \hat{x}_k/[\mu \hat{S}_k + (1 - \mu) |\hat{x}_k|^2]$). These filters optimize the pure signal processing capacities of the correlation operation without consideration of its optical implementation. However, if we want to improve the optical energy balance sheet by increasing the Horner efficiency, we see that the OTF are thresholded versions

of the Wiener filter (which is attractive for optical implementation).

We illustrate the performances of the OTF with numerical simulation experiments performed with an image of a truck of 256 x 256 pixels with 256 grey levels and with white stationary noise of variance equal to 1 (the same results would be obtained with another value). The size of the truck (identical to the one used in [16] - profil view) was approximately 1/10 of the total image (in pixels). The mean value of the intensity of the image was subtracted by setting the zero frequency value of the Fourier transforms equal to zero before processing in order to obtain a better energy repartition in the Fourier domain (if this preprocessing is not performed, only the PCE for $\mu = 1$ is smaller).

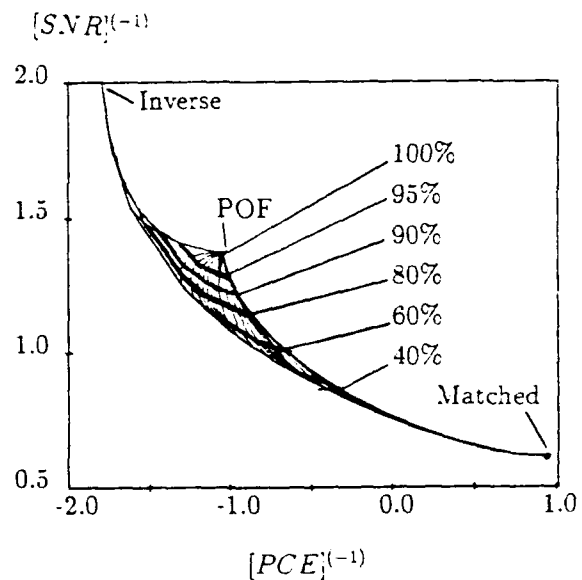


Figure 1: Curves in logarithmic scales of possible OTF for a truck. $MSE/[C_0]^2$ is drawn as a function of $CPE/[C_0]^2$ for different values of μ and λ . Thick curves correspond to filters with fixed values of C_0 (indicated in comparison with the one of the POF - i.e. $[C_0]_{POF} = 100\%$) while thin curves correspond to fixed values of μ and different values of λ . For representation convenience, the point for $\mu = 0.0$ is only shown in Fig.2, here the maximum value of μ is 0.01.

The results are presented in Fig.1. We see that both the matched and the inverse filters are over-specialized since optimization of

one criterion seriously deteriorate the performances from the point of view of the others criteria. The matched filter has low intensity and wide correlation peak. On the other hand the inverse filter has high intensity and sharp correlation peak but bad SNR. Finally the POF seems to be a not so bad trade-off, but if one considers only the pure signal processing capacities of the correlation operation (i.e. the SNR and the PCE), one can see that more interesting trade-offs than the POF can be found (see Fig.1). Examples of possible trade-off are summerized in table 1, where the values of the criteria are given as a ratio to their optimal values denoted by the subscript *opt*.

Filter	ρ_{SNR}	ρ_{PCE}	ρ_{C_0}
Matched	1.0	0.002	0.04
Inverse	0.04	1.0	0.3
POF	0.17	0.18	1.0
OTF $n^{\circ}1$	0.5	0.05	0.5
OTF $n^{\circ}2$	0.32	0.18	0.5

Table 1: Examples of possible optimal trade-offs for a truck. They are points of Fig.1. $\rho_{SNR} = SNR/SNR_{opt}$, $\rho_{PCE} = PCE/PCE_{opt}$, $\rho_{C_0} = C_0/[C_0]_{opt}$.

It has often been noted that, in general, the SNR of POF is not sufficient and different authors [9] [10] [17] [11] have proposed solution to improve it. Assume we only allow a SNR equal to the half of the matched filter one. In this case, the trade-off we obtain (shown in table 1 as OTF $n^{\circ}1$) improves the PCE by a factor $\simeq 25$ in comparison with the matched filter, and the amplitude of the correlation peak by a factor $\simeq 12$ (which is now only the half the one of the POF). Another example is given in this table (OTF $n^{\circ}2$): keeping the same PCE than for the POF, one can increase the SNR by a factor 2 if a decrease of C_0 with the same factor is accepted. For a very different object (a binary triangle) the ratio between the SNR of the POF and of the matched filter was 17 and approximately the same behavior for OTF has been found.

We propose now to show how these OTF provide a rigorous way to compare filter performances.

Recently a lot of interest has grown in ternary valued filters [9] [10] [17] [11], the reasons being mainly to improve the SNR of the POF without great complexity since only three value are permitted (-1,0,1). If we do not consider the optical implementation easiness, the three previous criteria are objective figure of merit. However, the correlation capabilities of the ternary valued filters can be function of the clipping procedure of the filters phase. So we analyze a generalized version of it, called the binary amplitude phase only filter (BAPOF) defined as follows :

$$\hat{h}_k = \begin{cases} \hat{x}_k/|\hat{x}_k| & \text{if } |\hat{x}_k| \geq 1/\lambda \\ 0 & \text{otherwise} \end{cases}$$

One can easily show that this binarization procedure is indeed optimal for BAPOF and noise robustness.

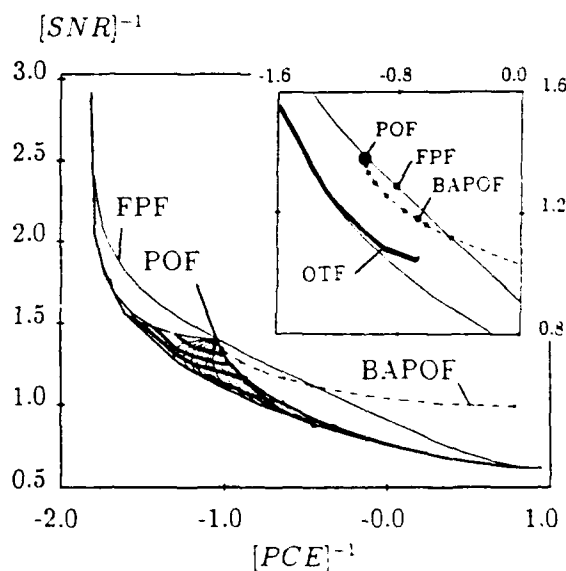


Figure 2: The same as Fig.1 with the criteria values of the BAPOF and the FPF. Insert : Zoom of a part of preceding curves. The thick curve corresponds to OTF with fixed values of C_0 equal to 0.66 the one of the POF. The BAPOF and the FPF are also shown with this same value of C_0 .

The values of the criteria for the BAPOF for different thresholds are shown in Fig.2. They show that for this example the BAPOF

DIGITAL APPROACH FOR PATTERN SCALE MEASUREMENT

Joseph Rosen, Lior Dezialoshinski, Ehud Nahtomi and Joseph Shamir

Department of Electrical Engineering

Technion - Israel Institute of Technology, Haifa 32000, Israel

Pattern size measurement is important for applications such as industrial classification and ranging. Optical systems offer fast and parallel processing of detailed pictures.

A recently proposed method [1], based on an optical correlator, measures pattern size using a specially designed spatial filter. This method is not shift invariant and is quite sensitive to noise owing to its analog operation.

In this work we extend the system to k parallel correlation channels, as shown in Fig. 1. The field of view of a TV camera is displayed on an SLM, and processed by k parallel correlators. The correlation peaks collected from the k channels at one arbitrary point, indicate the detection of a known objects. The binary word of k bits, which is created by these peaks, determine the scale of every object in the field of view simultaneously.

Every correlation channel is responsible for one bit of the binary word. Therefore every channel is equipped with a different spatial filter which produces a scale dependent binary correlation function. Examples of the correlation peak intensity versus the scaling factor, a , for a three channels system are given in Fig. 2. Each value of a yields a binary word coded in gray code. This code minimizes, both, the error measurement and the rate of intensity changes in every channel.

The method to achieve the above mentioned response functions in the correlator is by the superposition of basis functions of the form, $\frac{e^{js \ln r}}{r}$, where r is the radial variable of the image plane [2]. This method failed in our experiments since the correlation peak which indicates a logical '1' was too weak to be detectable. Therefore, instead of creating the exact square signal, we produce a sine wave with the frequency of the square wave and a thresholding operation creates the desired square wave.

Let $f(x, y)$ be the function of the detected object where $F(u, v)$ is its Fourier transform. Let $h(x, y)$ be the system impulse response for one channel, and $H(u, v)$ its Fourier transform. The correlation signal at the origin versus the scaling factor, a , is:

$$c(a) = \iint f\left(\frac{x}{a}, \frac{y}{a}\right) h^*(x, y) dx dy = a^2 \iint F(au, av) H(u, v) du dv = a^2 \int_0^{2\pi} \int_1^R F(a\rho, \phi) H(\rho, \phi) \rho d\rho d\phi \quad (1)$$

where (ρ, ϕ) are the polar variables of the spatial frequency plane. If we choose the filter function to be [3]: $H(\rho, \phi) = e^{-j\mu \ln \rho}$ then the correlation function becomes:

$$c(a) = a^2 \int_0^{2\pi} \int_1^R F(a\rho, \phi) e^{-j\mu \ln \rho} \rho d\rho d\phi = e^{-j\mu \ln a} \int_0^{2\pi} \int_a^{aR} F(\rho, \phi) e^{-j\mu \ln \rho} \rho d\rho d\phi \quad (2)$$

As we see, $c(a)$ depends on the scale a both in the range of integration and the harmonic function $e^{j\mu \ln a}$. We limit ourselves to a range of a where

$$\int_0^{2\pi} \int_a^{aR} F(\rho, \phi) e^{-j\mu \ln \rho} \rho d\rho d\phi \geq \frac{1}{2} \int_0^{2\pi} \int_1^R F(\rho, \phi) e^{-j\mu \ln \rho} \rho d\rho d\phi \quad (3)$$

In this range the rapid variations of $c(a)$ correspond to the harmonic oscillation of $e^{-j\mu \ln a}$ while the envelope, the integral, varies much slower.

To achieve a strong detectable correlation peak for a logical '1' we wish the filter to be as transparent as possible, and its phase distribution to match the phase of the object Fourier transform. With this kept in mind we search for a filter that produces a sine wave variation of the correlation peak as a function of $\ln a$. Under the assumption of Eq. (3), two phase-only harmonics will suffice to produce the desired wave. If the desired intensity wave in an arbitrary channel is:

$$|c(a)|^2 = \frac{1}{2} [1 + \cos(\mu \ln a + \psi)] \quad (4)$$

then the orders μ and 0 can be superposed, as well as every two orders with a frequency difference of μ . The choice between the above possibilities depends on the energy of those orders.

Obviously, the μ values are different from one channel to the other. The formula to calculate μ , for the i -th channel in the gray code, is given by:

$$\mu_i = \frac{(2^n - 1)\pi}{2^{i+2} \ln a_0} \quad i = 1..n \quad (5)$$

where n is number of channels and we assume that the measurement range is: $\frac{1}{a_0} \geq a \geq a_0$. The channel number 1 corresponds to the least significant bit (LSB), where the most significant bit, (MSB) is in channel n .

Up till now we concentrated only on the radial property of the filter. The angular dependence of the filter satisfies the matching condition, which guarantees largest sine wave amplitudes. Every harmonic term is given by:

$$H_\mu(\rho, \phi) = e^{j(\mu \ln \rho + \psi + \alpha_\mu(\phi))} \quad (6)$$

where

$$\alpha_\mu(\phi) = \frac{\int_1^R F(\rho, \phi) e^{-j\mu \ln \rho} \rho d\rho}{\left| \int_1^R F(\rho, \phi) e^{-j\mu \ln \rho} \rho d\rho \right|} \quad (7)$$

In other words, to obtain a sine wave as a response to scale variations we synthesize two phase only filters (POF) as given by Eq. (6), where μ dictates the wave cycle, and $\alpha(\phi)$ contains the angel information of the object. ψ is a constant for controlling the wave appearing on the a axis.

In our experiment we used three correlator channels to measure the size of a cross. For every channel we calculated the desired μ . We found it more efficient to superpose, in every channel, two POF's with frequencies of $\frac{\mu}{2}$ and $-\frac{\mu}{2}$, to obtain a sine wave with frequency μ . The scaling range in our experiment is $\frac{1}{2} < a < 2$. The three spatial filters in the three channels are shown in Fig. 3.

In one example we presented four crosses with different sizes, arranged along a line. The correlation results, before thresholding, for the least significant bit channel are shown Fig. 4. In the upper part of the figure the zero diffraction orders reconstruct the input pattern. The first diffraction order yields the desired correlation as shown in the lower part of the figure, with only two peaks present. Between the two orders a cross section of the correlation peaks, is displayed.

Comparison between theoretical, simulation and experimental results for the output correlation signals of the three channels, is shown Fig. 5. Finally, in Fig. 6 we demonstrate the output thresholded results of the three channels with the four crosses in the input plane. For every size we see different peak arrangements which are interpreted as a digital word - corresponding to the scale of the crosses.

References

- [1] B.V.K.V. Kumar, A.J. Lee and J.M. Connelly, *Opt. Eng.* **28**, 474, (1989).
- [2] D. Mendlovic, E. Marom and N. Konforti, *Opt. Comm.* **67**, 172, (1988).
- [3] J. Rosen and J. Shamir, *Appl. Opt.* **28**, 240, (1989).

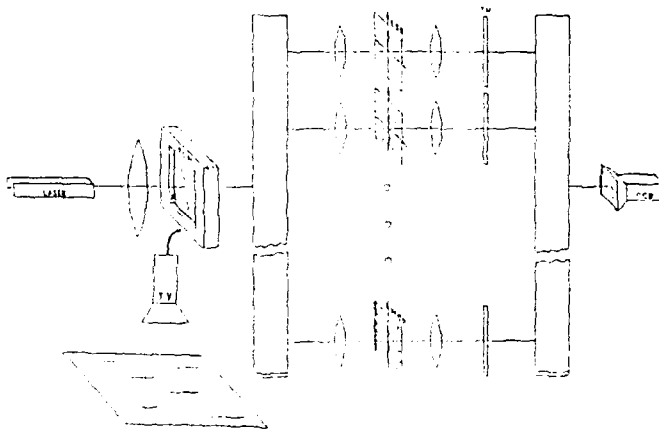


Fig.1. Experimental system for scale detecting.

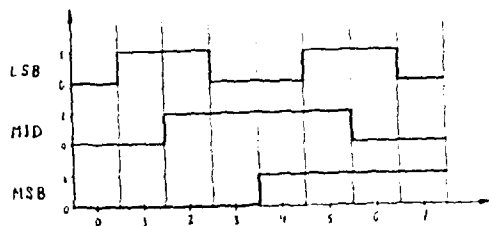


Fig. 2. Correlation peak intensity versus a scaling factor, in three correlation channels.

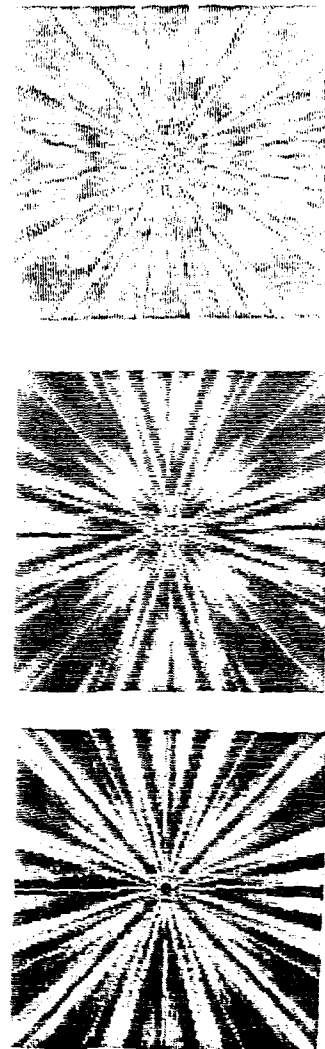


Fig. 3. Spatial filters in (a) the LSB, (b) the middle bit and (c) the MSB channels.

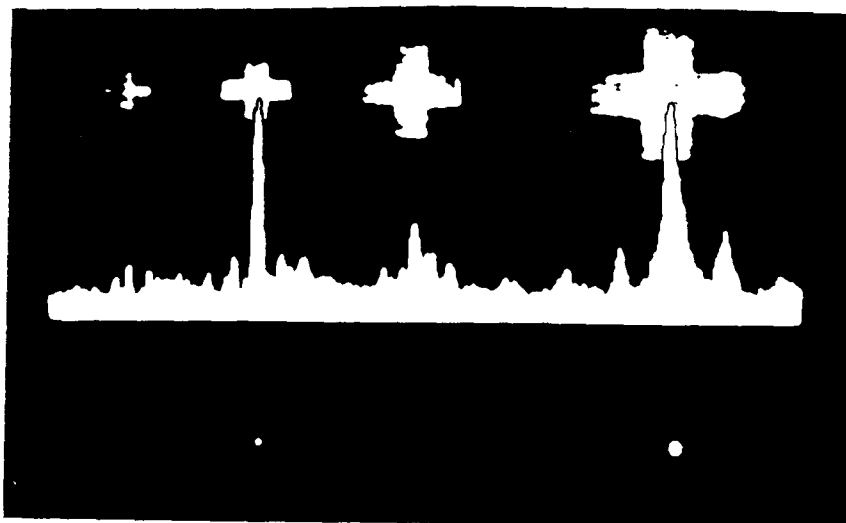


Fig. 4. The correlation results in LSB channel for the four crosses in the input (see the text).

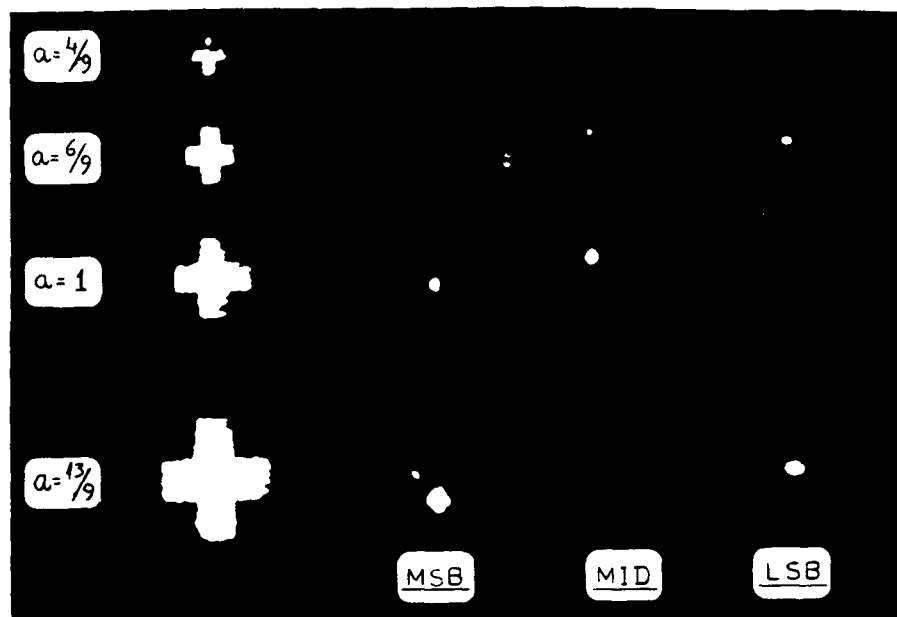


Fig. 5. The correlation signals versus scaling, a , for (a) the LSB (b) the middle bit and (c) the MSB channels.

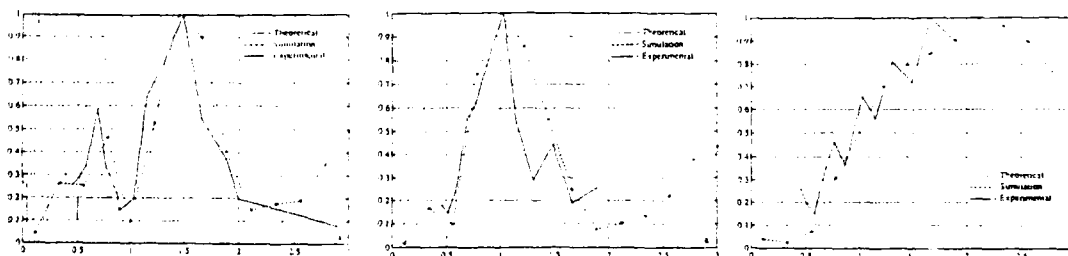


Fig. 6. The output correlation peaks, of the three channels, when the four crosses are in the input.

Image Correlation Using Photorefractive GaAs

Li-Jen Cheng, Duncan T.H. Liu, and Keung L. Luke
Center for Space Microelectronics Technology
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California 91109

Norman S.Z. Kwong
Ortel Corporation
Alhambra, California 91803

Image correlation can be implemented optically, which takes full advantages of light, namely parallel operation and global interconnection with the Fourier transform provided by lens. Photorefractive compound semiconductors can provide this type of implementation¹. This paper presents results from a detailed investigation on potentials of the photorefractive GaAs correlator for practical applications. The results illustrate that the matched filter formation rate in photorefractive GaAs crystal can be higher than 1000 frames per second. The filter contains complex values, leading to high quality correlation as demonstrated. Other advantages verified by experiments include real object image input with no need for preprocessing Fourier transform; edge enhancement automatically processed in the correlation process; dynamic spatial invariance; substantial enhancement of the signal by using a DC electric field providing high dynamic range; and easy alignment. In addition, this paper also presents the result of an experiment on imaging by phase conjugation in GaAs with 1.3 micron semiconductor injection lasers. This result provides realistic potentials to develop compact correlation modules using photorefractive semiconductors with semiconductor lasers. These modules could be building blocks for future "intelligent" automatic pattern recognition systems.

In the correlation experiment, two liquid crystal television spatial light modulators were used as input device. Figure 1 shows correlation images of a gray-scale car and its edge-enhanced pattern by a computer. The photographs in the left and the middle show autocorrelation images and line scans through the peak of the grey-scale car and the edge-enhanced car, revealing that the correlation signal intensities are the same. The signal is strong and the image quality is good. The background noise is very low, because a polarization switching configuration was used². The result illustrates an important feature that the edge enhancement is automatically performed during the correlation process, which can be attributed to the saturation of the DC component in the Fourier domain in the photorefractive crystal. The photograph in the right is

the correlation image and scan through the peak between the gray-scale car and its edge-enhanced pattern, showing that the correlation peak size is much smaller than those of the autocorrelation. This is due to the slight difference in edge enhanced patterns obtained by the optical process and the digital computing, as confirmed by results from further experiments.

The response time of the system, i.e. the recording time of the matched filter, was measured as a function of the total beam intensity before the entering into the LCTV SLMs when the system was performing an autocorrelation of a circle with the probe beam chopped. Only less than 7% of the total beam intensity could participate in the correlation process. The major loss of the beam intensity was the poor transmission of the LCTV SLM at 1.06 micron, about 8%. The oscilloscope scan in the right of Figure 2 gives a time dependence of the autocorrelation signal with the rise time of 0.8 milliseconds, which is equivalent to a processing rate of 1200 frames per second. In the left of the figure gives a plot of the response time of the system as a function of the total beam intensity, illustrating that the system speed is inversely proportional to the beam intensity. Because of the high filter recording speed, the correlation signal can follow the change in the incoming scene. This leads to dynamic spatial invariance as demonstrated experimentally. The dynamic spatial invariance provides capabilities for tracking moving objects.

It is known that application of an electric field can enhance the photorefractive effort drastically³. An experiment was done to evaluate the effect of DC electric field on the correlation signal in the GaAs system. A factor of 100 was observed by applying a 4 kV DC voltage on the GaAs crystal. This enhancement will substantially increase the dynamic range of the system. However, the voltage also reduce the speed of the system.

The effect of cluttered environment on the correlation signal is another important factor for system evaluation. Experiments were carried out to investigate the effect of random binary noise on the correlation signal as a simple simulation for the operation in a cluttered environment. The results have revealed that the correlation signal is still observable when about 60% of the image area was covered by the random noise.

Figure 3 gives a phase conjugate image obtained from a four-wave mixing in GaAs using two DFB single-mode InGaAsP/InP lasers. The phase conjugate signal is

strong and the quality of the image is reasonably good. The next step is to do correlation experiments with semiconductor lasers. The result will be presented.

The work described in this paper was performed by the Center for Space Microelectronics Technology, Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Defense Advanced Research Projects Agency, the Strategic Defense Initiative Organization/Innovative Science and Technology Office, and the U.S. Army Missile Command through an agreement with the National Aeronautics and Space Administration.

Reference:

1. G. Gheen and L.J. Cheng, Applied Optics, 27, 2756 (1988).
2. L.J. Cheng and P. Yeh, Optics Letters, 13, 50 (1988).
3. D.T.H. Liu, L.J. Cheng, M.F. Rau, and F.C. Wang, Applied Physics Letters, 53, 1369 (1988).

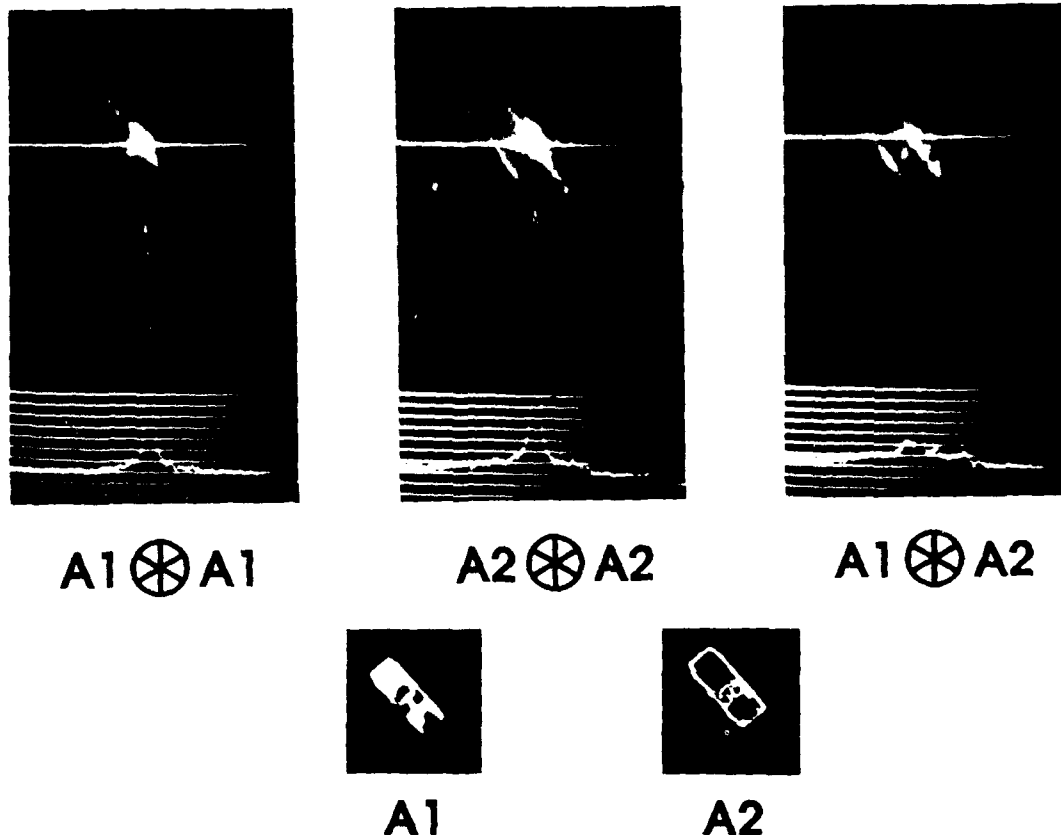


Figure 1. Autocorrelation images of a gray-scale car and an edge-enhanced car, illustrating that the photorefractive correlator can perform edge enhancement automatically. However, the amplitude of the cross correlation peak of the gray-scale car and edge-enhanced car is only a third of that obtained from the autocorrelation, indicating that the result from optical edge enhancement is slightly different from that made by a digital computer.

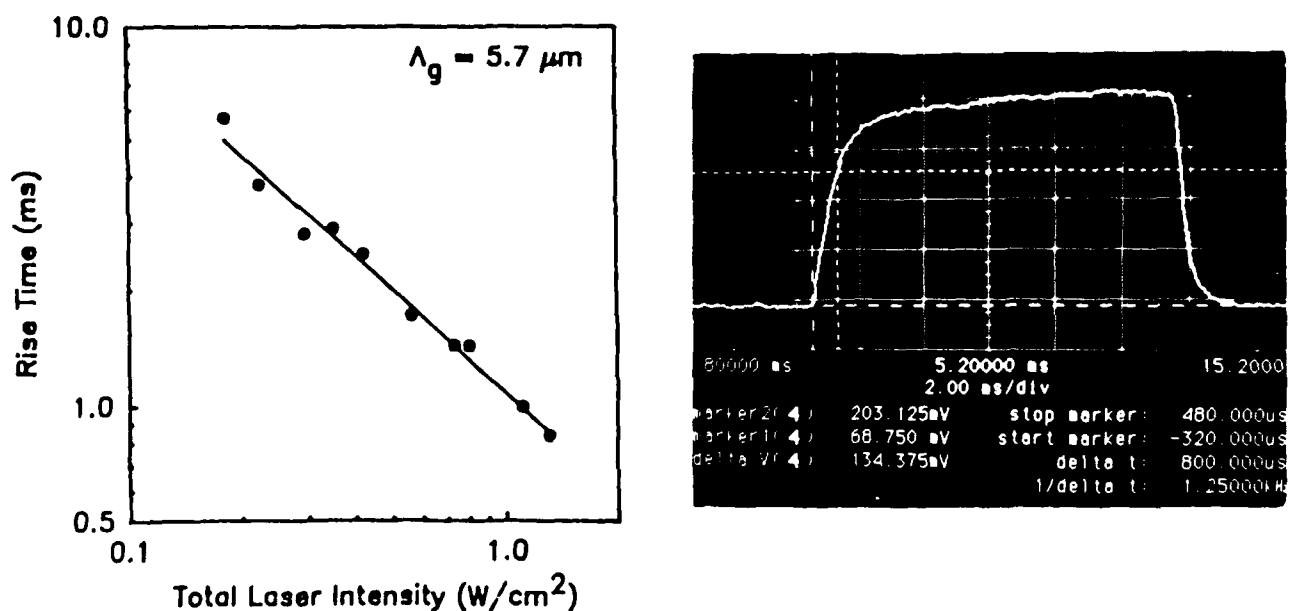


Figure 2 Response time of the correlation system as a function of the total intensity before the LCTV SLMs (left) and an oscilloscope scan (right), showing that the matched filter formation rate can be as high as 1200 frames per second (right).

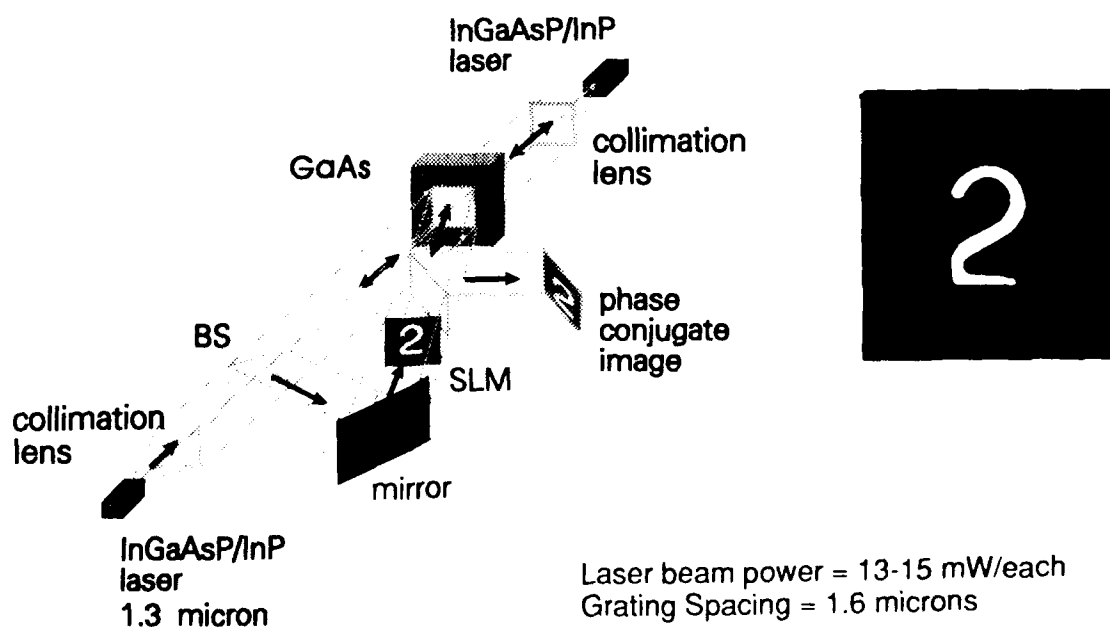


Figure 3. Sketch of the experimental setup for imaging by phase conjugation in GaAs using two InGaAsP/InP lasers and a phase conjugate image.

FILTER GENERATION IN HYBRID ELECTRO-OPTICAL CORRELATORS USING GENETIC ALGORITHM

Uri Mahlab and Joseph Shamir

Department of Electrical Engineering

Technion – Israel Institute of Technology, Haifa 32000, Israel

The purpose of this work is to introduce into the field of optical pattern recognition the parallel approach of genetic algorithm (GA) [1, 2] replacing conventional serial procedures. We start with a short review of the procedure for iterative spatial filter generation in a 4-f correlator [3], proceed with the adaptation of GA and present experimental results.

The complex amplitude distribution over the output plane of a coherent optical correlator is given by,

$$c(x_0, y_0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) h^*(x + x_0, y + y_0) dx dy \quad (1)$$

where $h(x, y)$ is the spatial filter function and $f(x, y)$ is the input function.

Starting from a training set $\{f_n(x, y)\}$ we define our goal as the detection of the presence of patterns out of a subset $\{f_n^D(x, y)\}$ while rejecting all other patterns denoted by the subset $\{f_n^R(x, y)\}$. Our criterion for detection is the appearance of a strong and narrow peak for a match between the input and the filter function as contrasted with a uniform distribution for a pattern to be rejected.

In a most general sense we may define a distribution function over the output plane by the relation

$$\Phi(x, y) = \frac{\mathcal{L}[c(x, y)]}{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathcal{L}[c(x, y)] dx dy} \quad (2)$$

where \mathcal{L} is a nonlinear operator over $c(x, y)$ such that $\mathcal{L}[c(x, y)]$ is a nonnegative quantity on (x, y) . For the present we choose \mathcal{L} to be the absolute value operator suitable for intensity detection. The distribution Φ has all the properties of a probability density for which one may define a *general entropy function* given by,

$$S = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi[\Phi(x, y)] dx dy \quad (3)$$

where Ψ is a strictly convex function [5].

Our criterion states that for a rejected pattern (the R subset) we require a uniform Φ over the whole output plane, maximizing the general entropy function (Eq. 3). At the same time, a strong and narrow peak for a pattern from the D subset results in the minimal value of the general entropy. Converting to a digitized form, $x \rightarrow m, y \rightarrow n$, we represent the various functions as two dimensional matrices of $N \times N$ pixels. A single steep peak over the correlation plane at some point, denoted by (k, l) , is represented by the ideal distribution,

$$\Phi^D(m, n) = \begin{cases} 1 & \text{at } (m = k, n = l) \in (\text{Domain of } \Phi) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

while a uniform distribution due to a rejected pattern has the form,

$$\Phi^R(m, n) = \frac{1}{(2N-1)^2} \quad \forall(m, n) \quad (5)$$

For each proper convex function Ψ the general entropy function has its upper and lower bounds. Taking for the present example [4]

$$\Psi(x) = x \log x \quad (6)$$

we obtain

$$S^D_{min} = 0 \quad \text{and} \quad S^R_{max} = \frac{1}{(2N-1)^2} \quad (7)$$

where S^D and S^R denote the entropy due to patterns from the D and R subsets, respectively.

A cost function defined by the relation,

$$M = \sum_{f_n^D} S^D - \sum_{f_n^R} S^R \quad (8)$$

has its ideal minimum determined by the ideal values given in Eq. (7). The first term of Eq. (8) contains the entropy measured for all the patterns from the subset to be rejected while the second term is a summation over all the patterns to be detected. (A complete description of this analysis is introduced in Ref. [4]).

Considering M to be a functional of a specific filter function we seek a generalized entropy function $h_{GEF}(i, j)$ which will minimize the cost function:

$$M_{min} = M[h_{GEF}(i, j)] \quad (9)$$

Since presently available spatial light modulators (SLMs) operate best in a binary mode we shall restrict our actual filter function to this mode. This simple representation together with the potentially high parallelism of optical processors indicate the usefulness of an iterative approach based on genetic algorithms.

Regarding the cost function in Eq. 8 as a fitness value for a given spatial filter function, our process lends itself quite readily for implementation by GA where each binary filter function constitutes a member of the population. The algorithm used is summarized as follows:

1) Start:

Select at random a population of m members (binary functions) $\{h_1, h_2, \dots, h_m\}$ and evaluate the values of the cost functions, M_i $\{i = 1, 2, \dots, m\}$. Compute the average value of the cost function $\theta = \frac{1}{m} \sum_{i=1}^m M_i$. Set a discrete time parameter t to zero. Define a probability P for a mutation to occur and set it to some P_{max} .

2) Crossover/mutate:

Select the function h_l which corresponds to the minimal cost function, M_l . Pick from the population a function h_j at random. The two functions, h_l and h_j are the parents to be used

for generating an offspring function. Select a random integer k between 0 and n , where n is the dimension of the vectors h representing the filter functions. Create the offspring function, h_c , by taking the first k elements from one of the parents, randomly, and the remaining $n - k$ elements from the other parent. Induce a mutation (inverting the elements) with probability P on each element of the offspring vector h_c . Evaluate M_c .

3) Reproduce:

Pick at random a function h_d from the population subject to the constraint: $M_d \geq \theta$. Replace h_d in the population with h_c and update the average value of the cost function, $\theta \rightarrow \theta + \frac{1}{m}(M_c - M_d)$.

4) Setting parameters:

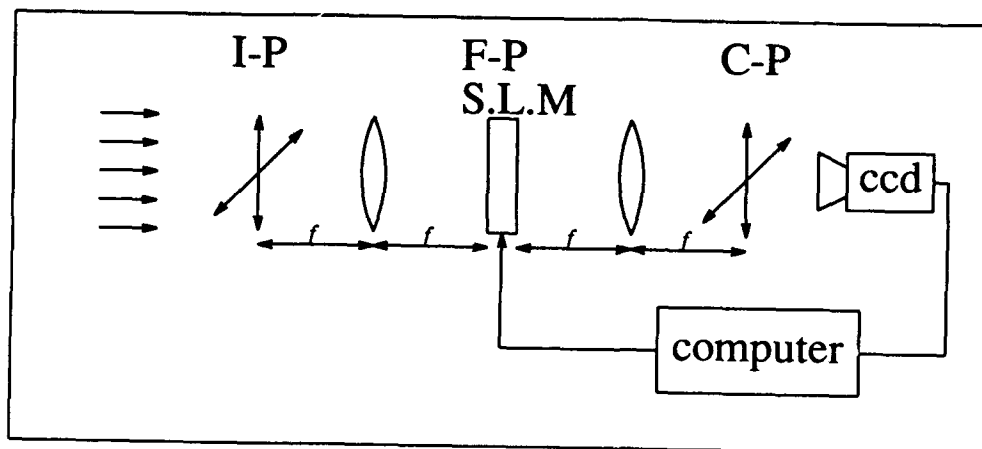
Set the new parameters, $t \rightarrow t + 1$ and $P \rightarrow P_{max} \left(\frac{1}{t}\right)^r$. If $P > P_{min}$ go to 2 otherwise go to 1. Selection of the parameters r , P_{min} , P_{max} depends on the particular problem at hand.

In this experiment the optical architecture is shown in Fig. 1. The input pattern is shown in Fig. 2. A filter $h(x, y)$ was generated in the frequency domain to detect the letter "T" and reject the letter "L". The Fourier transform of the filter $h(x, y)$ is a binary function and it was implemented directly on the LCTV. The size of the filter matrix is 64×64 elements. The output correlation plane is sampled by the CCD camera and fed into the computer to evaluate the cost function of Eq. (8) and perform the next iteration.

With the complete system controlled by an XT computer a discrimination ratio of 3:1 (Fig. 3) was obtained in 25 minutes. Since the whole process is implemented within the actual correlator system distortions were automatically compensated.

References

- [1] Lawrence Davis, *Genetic algorithm and simulated annealing*, Morgan Kaufmann, Los Altos, CA.
- [2] David E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*, Addison-Wesley Pub. (1987)
- [3] U. Mahlab and J. Shamir "Phase only entropy optimized filter by simulated annealing" *Opt. Lett.* **14**, 146, (1989).
- [4] U. Mahlab and J. Shamir "General iterative approach for filter generation for optical pattern recognition" (submitted for publication).
- [5] A. W. Robert and D. E. Veberg, *Convex Function*, Academic Press N.Y and London (1973).

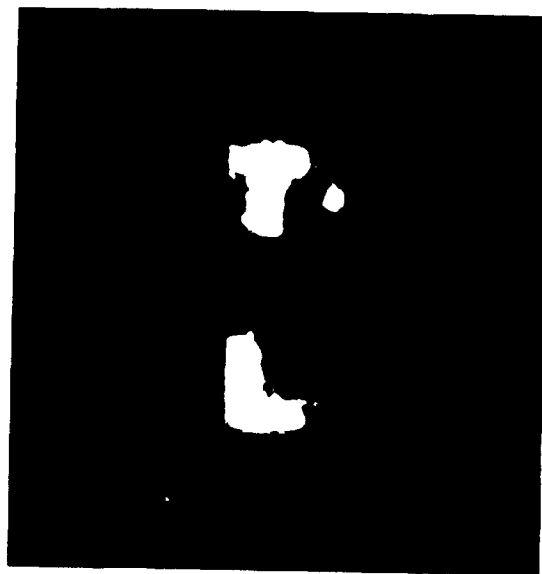


1. 4-F architecture for electro-optical implementation of learning algorithm. The filters functions are presented on the SLM sequentially and the control computer analyzes the output signal detected by the CCD camera.

(a) **T**

(b) **L**

2. Input training set: (a) Pattern to be detected (the letter "T"). (b) Pattern to be rejected (the letter "L").



3. Output correlation intensity produced by GA (discrimination ratio of 1:3)

Hardware and Software System Design for Hybrid Optical-Electronic Signal Processing

R. D. Griffin and J. N. Lee
 U.S. Naval Research Laboratory
 Optical Sciences Division, Code 6523
 Washington, D.C. 20375-5000

Many workers have demonstrated the potential of optical techniques to process high bandwidth data at very high computation rates.^{1,2} Optical processors have been proposed for such diverse applications as pattern recognition, neural networks, switching, digital computing, filtering, transformations, and matrix algebra. Since most of these methods implement specialized processors, they must be integrated into presently available digital electronic systems to obtain the necessary degree of control and flexibility for practical applications. Few systems exist today that realize the potential of these optical techniques, perhaps because of the amount of engineering and development effort that separates a successful laboratory demonstration from a useful and practical system. The engineering effort is complicated by the high bandwidth of the optical system: the input and output requirements of most optical systems can easily swamp traditional digital systems. Other complicating factors are data transduction between the electronic and optical domains, and dynamic range and signal-to-noise ratio requirements. In addition to the physical interface issues, the logical interface must be well-designed and easy to use. We present here some results of our effort to integrate a self-contained, "digital-in, digital-out" space-integrating one-dimensional matched filter system into a conventional digital processing system. This system can cross-correlate a 4000 point reference waveform with a 7000 point search waveform in about 100 μ sec.

The requirement for transduction of data between the electronic and optical domains has been the undoing of many proposals. *Even if an acceptable solution to the data transduction problem exists, one must still harness the power of the optical engine.* This requires that the optical engine be fed data at a rate commensurate with its processing capability and that the information produced by the engine be efficiently consumed by the downstream electronic system. Sometimes, the problem is less severe at either the input or the output end of the system. For example, a 2-D matched filter might be designed to provide only a present or absent indicator, in which case the output of the matched filter could be compared with a 2-D threshold. This reduces a potentially troublesome output bandwidth to a much lower data rate. We describe a 1-D matched-filter system with a low input data bandwidth but with an output bandwidth that would normally overload present-day digital systems.

A block diagram of the system is shown in Figure 1. Sensors provide the raw input data which are preprocessed to form search waveforms. After buffering, the search waveforms are transferred to the opto-electronic correlation system via a high-speed bus where they are cross-correlated with reference waveforms. Each correlation function can be digitized and returned to the host, or it can be compared with a threshold function and any threshold crossings reported to the host for further processing and display. The opto-electronic system is controlled by a digital signal processing microprocessor (DSP). Data transfer with the host system is via a first-in first-out buffer controlled by the DSP chip. The system is designed for interfacing to popular buses, for example, the PC bus and the VAX UNIBUS or BI bus. The VAX BI bus is capable of transferring several megabytes of data per second, which is a good match to the current correlator input-output system.

Figure 2 shows top and side views of the optical system. A diode laser and beam-shaping optics produce a collimated beam of light with rectangular cross section and large aspect ratio. This beam is doubly diffracted by the acoustical wavetrains in the pair of acousto-optic Bragg cells, and then the doubly diffracted and undiffracted beams are focused onto a photodiode by a lens. The signal from the photodiode is heterodyne-detected and log-amplified and then enters the post-processor.

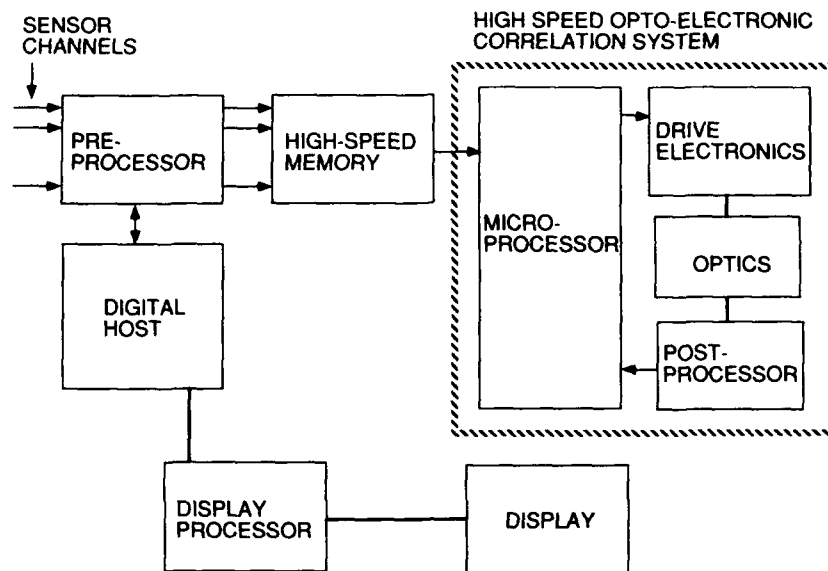


Fig. 1 System block diagram

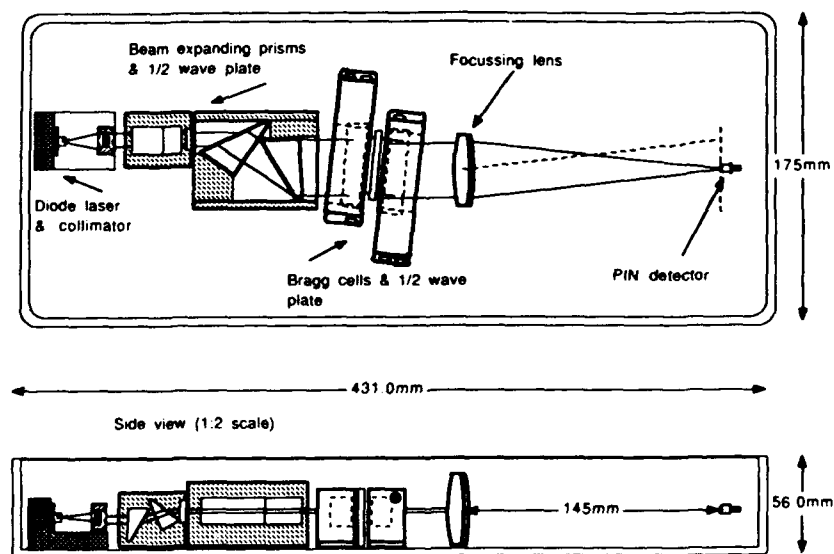


Fig. 2 Optical system

We designed our system hardware and software to satisfy the following requirements:

- 1) Provide all functionality of the correlator system via subroutines callable by the host from a variety of languages.
- 2) Make only modest demands on the host I/O system, but use the potential of the optical system as fully as possible.
- 3) Accept input data in a variety of formats, including 32-bit floating point and two's complement integers.
- 4) Accept search waveform data in blocks, typically containing several thousand samples, at an update period of about 50 ms.
- 5) Provide for overlap processing of the search waveforms.
- 6) Produce sampled and digitized correlation functions in integer or floating point format.
- 7) Produce threshold crossing reports in a digital format.

Since users of high-speed digital processors prefer to view these devices as functionally equivalent to a subroutine library, all the functionality of the correlator system is accessed via subroutines callable from a variety of high-level languages. Once the correlator hardware and software are installed in the host system, the user may essentially forget about the unique nature of the hardware.

The input data rate is kept low and the internal processing rate kept high in the following manner. Search waveforms are downloaded approximately every 50 ms. Reference waveforms are downloaded to the optical system only occasionally, and from these original waveforms the correlator generates a library of several hundred broadband Doppler distorted reference waveforms. A particular distorted reference is produced on demand for cross-correlation with the search waveform. Since a typical correlation requires about 100 μ s, the correlator can perform approximately 500 correlations during the 50 ms period before a new search waveform is received. Because a typical correlation can contain several thousand samples, the output rate can easily be 10 Msamples/sec or more. Most digital systems are not designed to handle such data rates for extended periods of time.

The reference and search waveform signals that drive the Bragg cells are generated from digital samples by eight-bit digital-to-analog converters operating at a nominal conversion rate of 80 MHz. The waveforms are in complex format, and the real and imaginary components are connected to the I and Q inputs of quadrature modulators that modulate a 75 MHz carrier wave. The Bragg cells have a nominal bandwidth of 40 MHz. For best performance the input waveforms should be preprocessed so that the input bandwidth when time-compressed by the correlator will match the bandwidth of the Bragg cells.

Unless the input waveforms are already in the desired format, the input samples must be converted to eight-bit integer data. There are two concerns when converting data formats. First, conversion from floating point to integer format can be a slow operation: one must insure that the conversion process does not become a bottleneck. Second, if the dynamic range of the data exceeds eight bits, one must preserve as much of the signal information as possible. One way to approach this problem is called block normalization. In this method, a segment or block of the search waveform is normalized via one of several possible methods and then correlated against the reference waveforms. The resulting correlation function segments are thus all associated with the same normalization parameter or parameters, and this information can be passed to the postprocessing and display system, or the correlator system can restore the dynamic range of the correlation functions before sending them to the host system.

There are many useful normalization methods. One of the simplest is to normalize all the data in a segment to the peak value in the segment. Another possible method is to normalize to the peak or average power in the segment. These methods have the disadvantage that information can be lost by truncation or clipping. A modified mu-law³ method that we have developed avoids this by compressing the amplitude logarithmically while maintaining the phase information. This method works very well for the waveforms of interest to us and allows us to map the dynamic range of the input waveform to that of

the opto-electronic system. We have measured autocorrelation peaks over an input dynamic range of 60 dB, i.e., the correlation peak was still detectable when the reference and search waveform rf input powers were both reduced 30 dB.

Accepting data in blocks requires that the correlator system have adequate buffers for storing the data. The primary concern here is the amount of power and volume consumed by the memory. We avoided the use of high-speed static RAM's by multiplexing slower but more efficient CMOS memories.

When the search waveforms are long they must be processed in segments. Special processing must be employed to insure that no returns are missed around the segment boundaries.⁴ This feature is provided via software and requires that the correlator system inform the user of the necessary amount of overlap from one waveform segment to the next.

The output signal from the log-amplifier can be digitized and returned to the host, but at a very high data rate. Since most signal processing systems do not typically process data at average throughputs of 10 Msamples/sec or more, we included as part of our system a postprocessor that would report only the information that is typically of interest. This postprocessor compares a correlation function with a threshold function. When the correlation function crosses the threshold, the correlator system samples the amplitude of the correlation function and assembles a threshold crossing message consisting of the correlation amplitude, the threshold amplitude, and the time of the crossing. These messages are then returned to the host system. For typical threshold settings, this method can reduce the output data bandwidth by one or two orders of magnitude compared to reporting digitized correlation functions. By including this postprocessing procedure (typically the next step performed in the processing chain, anyway) in the optical correlator system, the bandwidth requirements for the host system are reduced tremendously.

With a well-designed input and output system, the optical system could cross-correlate on average a 4000 point reference waveform with a 7000 point search waveform about every 100 μ sec. Because of design constraints on the I/O system, the current system can perform a correlation on the average about once every 250 μ s. This rate is reduced even further if the input data format conversion process is a lengthy one.

We have described the design issues that affect the performance of a high-speed optical processor inserted into a conventional digital processing system, and we have shown how these issues were addressed for the case of a matched-filter or correlator module. This 1-D correlator can be used for applications such as speech and Doppler processing. There is potential in this design to achieve very high system throughput by relatively modest improvements in the I/O system. The same design concepts might be applied in future digital multiprocessor architectures to produce a very high capability system.

References

1. *Optical Processing and Computing*, edited by H.H. Arsenault, T.Szoplik, and B. Macukow (Academic Press, San Diego, 1989).
2. *Acousto-Optic Signal Processing*, edited by N.J. Berg and J.N. Lee (Marcel Dekker, New York, 1983).
3. B. Smith, "Instantaneous companding of quantized signals," *Bell Syst. Tech. J.* **36**, 653-709 (1957).
4. E.O. Brigham, *The Fast Fourier Transform* (Prentice Hall, Englewood Cliffs, NJ, 1974), pp. 206-217.

Multichannel Bragg Cells for Optical Computing Applications

Dennis R. Pape
 Photonic Systems Incorporated
 1800 Penn Street, Suite 4B, Melbourne, Florida 32901

1. INTRODUCTION

Acousto-optic Bragg cell devices currently provide the most effective way of imparting electrical information in real-time onto a light beam. Multichannel Bragg cells, with individually addressable electrodes on the same transducer substrate, extend the power of optical processing in a compact package to optical computing applications such as two-dimensional optical switching and matrix-vector processing. The design of the cell, with the placement of multiple electrodes in close proximity on a common acoustic substrate, is constrained by crosstalk and thermal requirements not found in the design of a single-channel device. These constraints arise from the desire to place the electrodes as close as possible to maximize spatial duty cycle without creating unacceptable adjacent channel crosstalk and optical beam distortion. In this paper we discuss multichannel Bragg cell design principles and the use of RF stripline techniques and acoustically anisotropic acousto-optic materials with high thermal conductivity to achieve high multichannel cell performance. We describe several high performance Gallium Phosphide multichannel Bragg cells which employ these design techniques. We contrast the performance of these cells with devices manufactured commercially in the United States as well in the Soviet Union. Finally, optical computing systems using multichannel Bragg cells for switching and processing are discussed.

2. MULTICHANNEL BRAGG CELL

A perspective view of a multichannel Bragg cell is shown in Fig. 1. A piezoelectric transducer substrate is mechanically bonded, with metallic thin films, to an acousto-optic crystal. These metallic bonding layers also serve as the bottom electrode for the transducer assembly. The top electrode metallic layer contains multiple electrodes which are defined

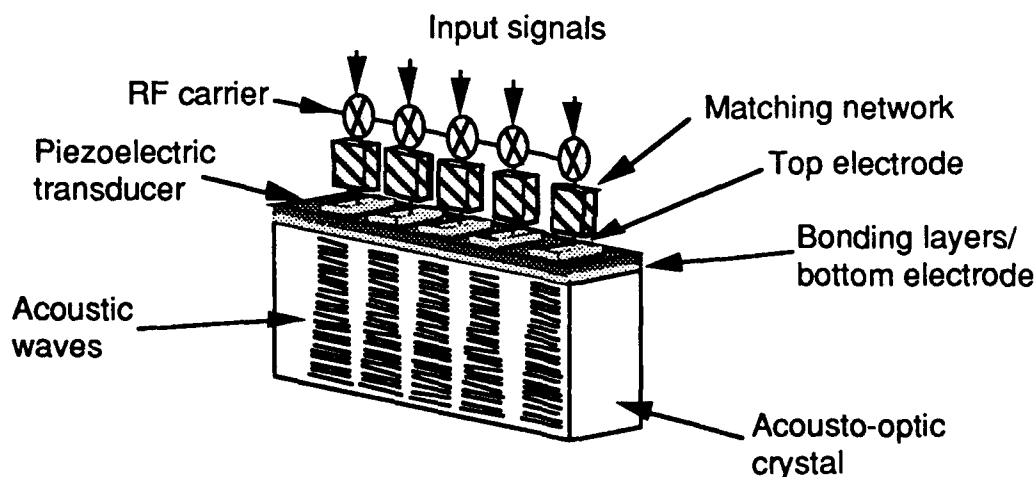


Fig. 1. Multichannel acousto-optic Bragg cell

photolithographically. Each electrode defines a single channel of the device. Each channel of the device operates in the same manner as a single-channel Bragg cell.

The multichannel Bragg cell has a number of attractive performance attributes relative to other types of spatial light modulators. These include: 1) capability for both analog and digital addressing, 2) transmissive operation with high transmission, 3) high diffraction efficiency, 4) high contrast ratio, 5) fast (ns) response time, and 6) amplitude light modulation. Unlike most two-dimensional spatial light modulators, the multichannel Bragg cell utilizes mature fabrication processes and is commercially available. Both narrowband quartz [1,2], and, more recently, wideband Lithium Niobate (LiNbO_3) [3], and Tellurium Dioxide (TeO_2) [4,5] multichannel Bragg cells have been described.

3. DESIGN ISSUES

The design of an individual channel of a multichannel Bragg cell is similar to that of a single channel device. Multiple electrodes on a common acoustic substrate, however, impose additional design constraints not found in the design of the single channel device. These constraints include minimization of electrode spacing, minimization of input RF power, minimization of thermal gradients, and minimization of adjacent channel crosstalk (both electrical and acoustic) [4]. Acoustic crosstalk is of particular concern when the time-aperture of the device is large.

3.1 Acoustic crosstalk design considerations

Acoustic crosstalk arises from diffraction spreading of the acoustic beam from one channel into neighboring channels. A high level of acoustic crosstalk can severely limit the utility of the device for many optical information processing applications. The most effective way of minimizing acoustic crosstalk is through the use of acoustic anisotropy. Acoustic modes exist in some anisotropic materials where the acoustic beam divergence is minimal, a so-called self-collimating mode. One important acousto-optic mode that exhibits self-collimation with an adequate figure of merit is a shear mode in Gallium Phosphide (GaP) with acoustic propagation in the $[1,-1,0]$ direction and optical propagation in the $[111]$ direction. Here the energy is almost perfectly collimated.

3.2 Electrical crosstalk design considerations

Electrical crosstalk in a multichannel Bragg cell arises primarily from coupling between the individual electrode matching networks and/or the transmission lines connected to each of the multiple transducers. A multichannel Bragg cell, with closely spaced electrodes, requires multiple matching networks which must be placed in close proximity to one another. Stripline transmission lines, where the conductors are embedded in a dielectric sandwiched between two ground planes, provide the ability to control crosstalk through the design of the dielectric sandwich. A scheme to interconnect a multichannel Bragg cell to a stripline structure is shown in Fig. 2. We have demonstrated 40 dB electrical crosstalk isolation between adjacent channels at a center frequency of 400 MHz and a channel spacing of 250 μm using this technique.

4. PERFORMANCE

Using the design principles described above, we have designed several GaP multichannel cells for optical computing applications [6,7]. The performance of these devices is shown in Table I. Quantitative measurements of device crosstalk were made by scanning a small

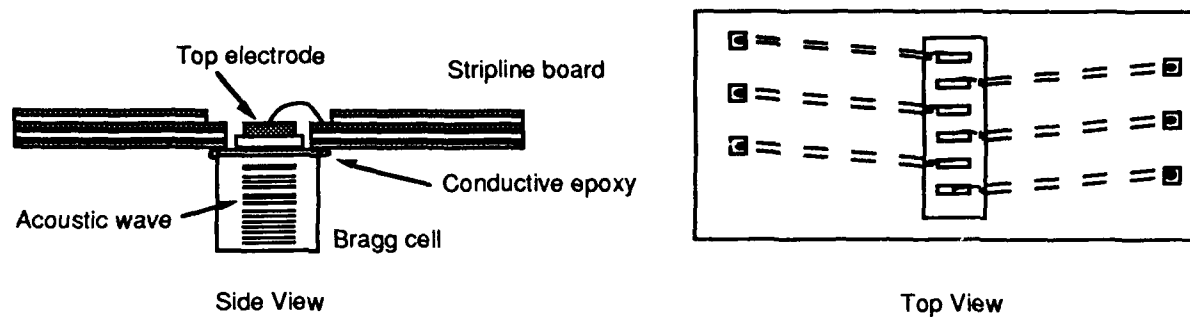


Fig. 2. Multichannel Bragg cell sculptured stripline interconnection structure

aperture detector across the schlieren image (in the direction orthogonal to the acoustic propagation) of the device. The experimentally measured crosstalk was less than -30 dB throughout the 8 mm aperture of this device.

Table I
GaP Multichannel Bragg Cell Performance

	Demonstrated [6]	Under Development [7]
Wavelength	632.8	830
Number of channels	64	64
Center frequency (MHz)	400	800
Bandwidth (MHz)	200	340
Time-aperture (μ s)	1.0	2.56
Time-bandwidth product/channel	200	870
Crosstalk (across full time-aperture) (dB)	< -30	< -30
Electrode height (μ m)	125	50
Electrode length (μ m)	1800	383
Electrode spacing (μ m)	250	250
Diffraction efficiency (% at 200 mW RF)	22	4
Diffraction efficiency channel uniformity (dB)	+/- 0.75	+/- 0.75
Signal time error (ns)	< +/- 1	< +/- 1

5. APPLICATIONS

The characteristics and performance of multichannel Bragg cells make them particularly attractive as two-dimensional spatial light modulators for optical computing applications including matrix-vector processing [7] and switching [8].

5.1 Digital optical computer

A general purpose 32 bit digital optical computer is currently under development at OptiComp Corporation [7]. The optical central processing unit of the machine performs Boolean logic matrix/vector multiplication using a laser diode array light source, two multichannel Bragg cells, and a silicon avalanche photodiode array. A digital data vector is input to the first multichannel device, a 64 channel modulator using the longitudinal [110] mode of GaP for maximum diffraction efficiency (here, with a device time-bandwidth product

of one, acoustic crosstalk is not of concern). This device is designed to have an efficiency in excess of 35% with 200 mW of input RF power and a risetime of 2 ns. A control matrix is input to the second multichannel device, a 64 channel cell using the shear [1,-1,0] mode of GaP described above. The design performance of this device is shown in Table I above.

5.2 Optical Switch

A nonblocking space-division optical switch with $O(N)$ complexity implemented with a multichannel acousto-optic Bragg cell has been recently described [8]. The light at each input fiber port illuminates one channel, respectively, of the multichannel device. Each channel of the Bragg cell is driven by an RF frequency synthesizer. The light from a particular channel is deflected at an angle proportional to the frequency of the RF signal input to the channel. The deflected beam is focused onto an array of output fiber ports. A switch with a single input channel and 4 output channel has been experimentally demonstrated with an insertion loss ranging from 4.6 to 5.6 dB, a worst case signal-to-cross-talk ratio of better than 30 dB, and a reconfiguration time of 1.4 μ sec.

6. SUMMARY

Multichannel Bragg cells are important components in many two-dimensional optical information processing systems. Minimization of crosstalk and thermal effects are key goals in the design of a multichannel cell. We have demonstrated that the use of a self-collimating shear mode in Gallium Phosphide substantially reduces acoustic crosstalk from that found in TeO_2 cells. We have also shown that the use of stripline transmission lines substantially reduces electrical crosstalk over that obtained using the more conventional microstrip techniques. The performance of two different GaP multichannel Bragg cells using the design principles outlined in this paper was described. Finally, several optical computing applications using multichannel Bragg cells were presented.

7. REFERENCES

1. O.B. Gusev, S.V. Kulakov, V.A. Mel'nikov, N.S. Mirgorodskii, V.P. Pikarnikov, V.V. Soroka, and S.P. Fadeev, "Multichannel acoustooptic modulator for real-time recording and optical processing," *Sov. Phys. Tech. Phys.*, Vol. 23, p.99, 1978.
2. L.M. Vasil'eva, O.B. Gusev, V.V. Kludzin, S.V. Kulakov, and V.P. Pikarnikov, "Multichannel optoacoustic modulator," *Sov. J. Opt. Technol.*, Vol 48, p. 275, 1981.
3. I.C. Chang and R. Cadieux, "Multichannel acousto-optic Bragg cells," *Proc. IEEE 1982 Ultrasonics Symposium*, p. 413, 1982.
4. D. R. Pape, "Multichannel Bragg Cells for Optical Systolic Matrix Processing," Topical Meeting on Optical Computing, March 18-20, 1985, Incline Village, NV., Tech. Digest, Paper TuC6-1, 1985.
5. M. Amano and E. Roos, "32 Channel Acousto-Optic Bragg Cell for Optical Computing," *Proc. SPIE*, Vol 753, *Acousto-optic, Electro-optic, and Magneto-optic Devices and Applications*, p. 37, 1987.
6. W.R. Beaudet, M. Popek, and D.R. Pape, "Advances in multichannel Bragg cell technology," *Proc. SPIE*, Vol. 639, *Optical Information Processing II*, p. 28, 1986.
7. P.S. Guilfoyle, D.R. Pape, J.A. Carter, R.V. Stone, "32 Bit digital optical computer: A hardware update," *Proc. SPIE*, Vol. 1296, *Optical Information Processing IV*, p. 20, 1990.
8. D.O. Harris and A. VanderLugt, "Acousto-Optic Photonic Switch," *Optics Letters*, Vol. 14, p. 1177, 1989.

Tuesday, March 5, 1991

Digital Systems

TuA 8:30am–10:20am
Salon F

Michael Prise, *Presider*
AT&T Bell Laboratories



Paper to be announced

A HARDWARE COMPILER FOR DIGITAL OPTICAL COMPUTING

Miles Murdocca, Vipul Gupta and Masoud Majidi
 Department of Computer Science
 Rutgers University, Hill Center
 New Brunswick, NJ 08903
 murdocca@aramis.rutgers.edu

Abstract

A hardware compiler for translating descriptions of digital circuits from a hardware description language (HDL) into gate-level layouts is under development at Rutgers University. The layouts are customized for optical processors that make use of arrays of optical logic gates interconnected in free-space with regular interconnection patterns such as perfect shuffles, crossovers, or global interconnects. Specific processors that the hardware compiler supports include the S-SEED based all-optical processor developed at AT&T Bell Labs, the S-SEED based all-optical processor under development at the Photonics Center at RADC / Griffiss AFB, and the acousto-optic modulator based RISC processor under development at OptiComp Corporation.

Introduction

Hardware compilation is the process of translating high-level descriptions of computer circuits into actual designs. The computer designer is relieved of managing the low-level layout details of the target computer and focuses instead on the development of the functional behavior of the target computer. Hardware compilers are common in the electronics community, although they are not used everywhere since the automated translation can limit performance, sometimes so severely that there is little motivation for using a hardware compiler at all. In this case manual design is generally better than automated design. Performance limitations due to automated design are frequently the result of simplifications in wiring topology. For example, a common simplification is to decompose a large circuit into a number of smaller modules, where inputs to the modules are at the left and outputs are to the right. When the natural flow of computation is not left-to-right, however, many compilers will not factor this into the design of the target machine.

A model of a digital optical computer that is supported here consists of arrays of optical logic devices interconnected in free-space with regular interconnection patterns such as perfect shuffles, crossovers, or global interconnects (expand and collect). This model restricts connectivity to predefined patterns and takes some of the connection burden off of the hardware compiler, so that performance limitations due to connections are not nearly as severe as they are for electronics, although the overall difficulty of the gate-level interconnection problem is increased for low fan-out interconnects.

There are a number of motivations for using a hardware compiler for optical computing. For example, the primary design concern for the all-optical computing model demonstrated at AT&T [1] is to manage design complexity. Normally in an electronic technology, the computer designer considers the functional behavior of a digital circuit separately from the physical layout. For the AT&T model, that is not the case since the functional behavior and the physical layout are tightly coupled. Design complexity can be managed, however, as evidenced by efforts in design of circuits for this model [2,3]. A second motivation for developing a hardware compiler for optical computing is that a simulator can be created for the target architecture, and then the same HDL description can be used for both the simulation and for the physical design, thus reducing the hazard of generating a physical design whose functional behavior differs from the simulation. Finally, there is a need to study how changes in the optical architecture affect performance. In order to study case examples, it is more productive to give the design task to an automated program than it is to generate designs manually. Although manual designs tend to be more efficient, the difficulty of design is such that only a few architectures can be created for a performance study if manual design is used exclusively.

Example Architectures

One model of a digital optical computer that is supported by the hardware compiler project is the all-optical processor based on arrays of S-SEED optical logic gates [4] developed at AT&T, as shown in Figure 1. The model consists of alternating arrays of optical logic gates and free-space regular interconnects such as crossovers [5]. Masks in the image planes block light at selected locations so that the interconnects are customized to perform specific logic functions such as addition and sorting. The system is fed back onto

itself and an input channel and an output channel are provided. Feedback is imaged with a single row vertical shift so that data spirals through the system, allowing a different section of each mask to be used on each pass. Information travels orthogonal to the device substrates.

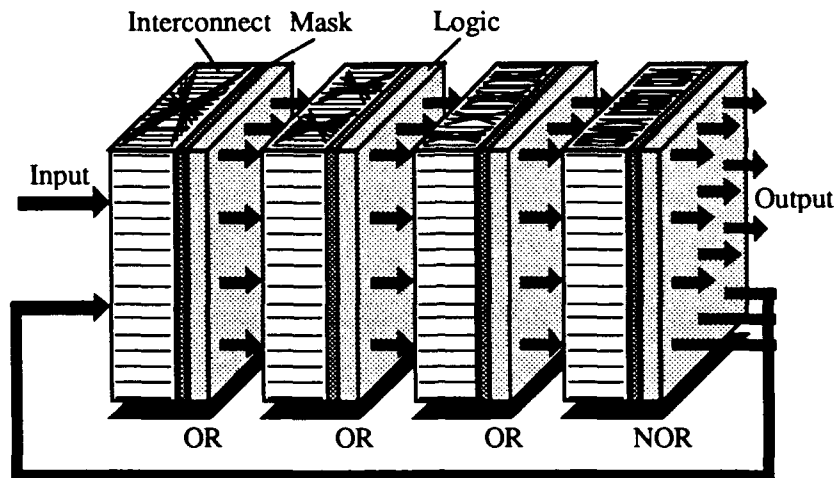


Figure 1: Arrays of optical logic gates are interconnected with optical crossovers. Masks in the image planes block light at selected locations which customizes the system for specific logic functions.

An alternative model that the compiler is being developed to support is the acoustooptic modulator based OptiComp digital processor [6] which is illustrated in Figure 2. A one-dimensional input vector is fanned out to the width of a two-dimensional spatial light modulator (SLM). Control patterns are set up on the SLM to enable or disable inputs from reaching the target output detectors. A single pass through this system realizes the AND stage or the OR stage of a programmable logic array (PLA), and several functions can be implemented in parallel. A significant difference between this model and the model shown in Figure 1 is the use of a global interconnect which simplifies the interconnection problem.

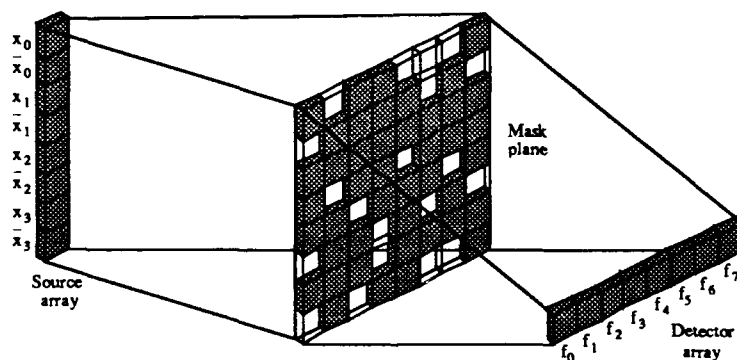


Figure 2: Model of an acoustooptic modulator based digital optical computer [6].

The Compilation Process

The functional behavior of a digital circuit can be described in terms of a hardware description language such as A Hardware Programming Language (AHPL) [7]. Lines 1-4 below describe a sequential eight's complementer and are repeated from the second edition of Ref. [7]. Each AHPL statement consists of two parts. The first part consists of a data transfer, marked by an assignment operator such as a left arrow or an equals sign as shown in statement 1. The second part of an AHPL statement consists of a transfer, indicated by a right arrow, and describes under what conditions control is transferred to another statement. For example, in statement 1 control is transferred to statement 3 if the *start* input is zero, otherwise control is transferred to statement 1.

1. $Y \leftarrow 0, x; z = 0;$
 $\rightarrow (start, \sim start) / (1,3).$
2. $Y \leftarrow Y_1, x; z = Y_0;$
 $\rightarrow start / (1).$
3. $Y \leftarrow Y_1, x; z = Y_0;$
 $\rightarrow start / (1).$
4. $z = COMP_0(Y, x);$
 $Y \leftarrow COMP_{1:2}(Y, x);$
 $\rightarrow (start, \sim start) / (1,2).$

The eight's complement of an octal digit is computed by subtracting the number from eight, modulo eight. For example, the eight's complement of 0 is $\text{mod}8(8 - 0) = 8 = 0$. The eight's complement of 5 is $\text{mod}8(8 - 5 = 3) = 3$. The AHPL example shown above has an input line x , an output line z , and an input line $start$. When the $start$ line goes from 1 to 0, the resulting circuit starts grouping the input stream on line x into three bits per octal digit, and outputs on line z the eight's complement of the octal digits.

The hardware compiler reported here takes the AHPL code shown above and translates the HDL description into a design of a circuit. Only the translation for the COMP unit is described here, which performs the eight's complement of three-bit octal digits. The diagram shown in Figure 3 is created by the compiler. The reader can verify that the relationship between inputs (top) and outputs (bottom) corresponds to the truth table for the eight's complementer shown to the left of Figure 3. Additional circuitry is needed for control sequencing. As of this writing, compilation of the control sequencer is partially completed, and no major obstacles are anticipated for its completion.

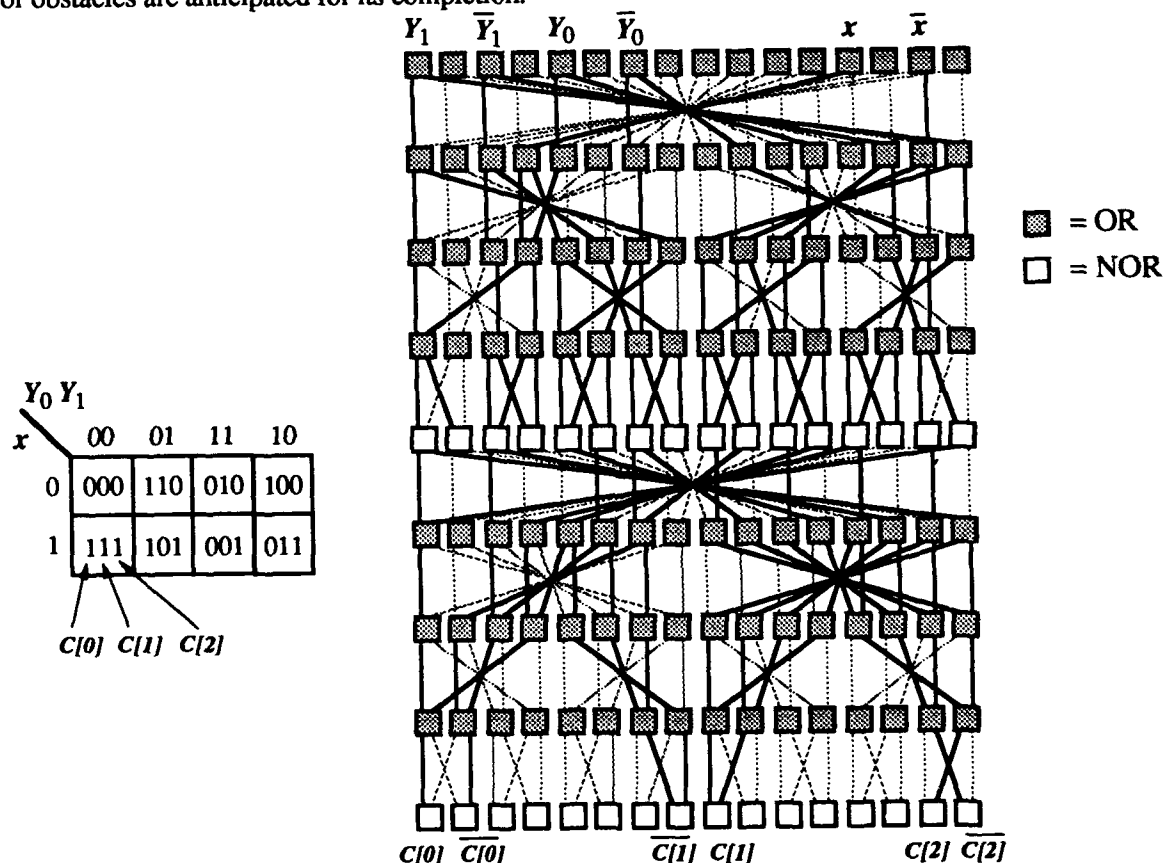


Figure 3: Truth table (left) and circuit layout (right) for eight's complementer following the model shown in Figure 1.

For the OptiComp model shown in Figure 2, the compiler when completed will produce a two-dimensional control pattern that implements the truth table shown in Figure 3. This is a relatively easy task for the compiler since routing conflicts are not a significant issue for the global interconnect as they are for low fanout interconnects like the crossover.

The compiler is contained in a single C program that takes as input descriptions of circuits in AHPL and produces intermediate dataflow lists that partition the circuits into topologically similar structures. The existing compiler does not generate complete gate-level circuits of an entire processor, although a full hardware compiler under development will take a similar form and will work with a more pervasive HDL such as VHDL. The existing suite of tools for automated PLA generation for the model shown in Figure 1 can be obtained from the authors, as well as a simulator.

Conclusion

A hardware compiler under development at Rutgers University for free-space digital optical computing is described. Although S-SEED and Bragg cell acoustooptic modulator based systems are used as example architectures, the compilation process is independent of the device technology and any suitable devices will suffice, for example, devices described in Refs. [8-10]. The AHPL hardware description language is used for the existing compiler, although plans are being made to extend the compiler to a more widely used language such as VHDL. The compiler is being extended to incorporate a small amount of electronics in a "smart pixel" approach such as described in Ref. [11].

This work was jointly supported by the Air Force Office of Scientific Research and the Office of Naval Research under grant N00014-90-J-4018. The Optical Computing Research Department at AT&T Bell Labs is acknowledged for its support and collaboration in the development of the AHPL compiler.

References

- [1] Prise, M. E., "Optical computing using self-electro-optic effect devices," *Digital Optical Computing*, Proc. of O-E/Lase '90, SPIE Press, CR35, pp. 3-27, (1990).
- [2] Murdocca, M. J., A. Huang, J. Jahns, and N. Streibl, "Optical design of programmable logic arrays," *Appl. Opt.*, **27**, 1651, (May 1, 1988).
- [3] Murdocca, M., *A Digital Design Methodology for Optical Computing*, The MIT Press, (1990).
- [4] Lentine, A. L., H. S. Hinton, D. A. B. Miller, J. E. Henry, J. E. Cunningham and L. M. F. Chirovsky, "The symmetric self electro-optic effect device," in *Conference on Lasers and Electro-optics*, Technical Digest Series 1987, vol. 14, (Optical Society of America, Washington, D.C., 1987, 249), postdeadline paper.
- [5] Jahns, J. and M. J. Murdocca, "Crossover networks and their optical implementation," *Appl. Opt.*, **28**, 182, (Aug. 1, 1988).
- [6] Guilfoyle, P. S., "Digital optical computer fundamentals, implementation, and ultimate limits," *Digital Optical Computing*, Proc. of O-E/Lase '90, SPIE Press, CR35, pp. 288-309, (1990).
- [7] Hill, F. J. and Peterson, G. R., *Digital Systems: Hardware Organization and Design*, 3rd ed., John Wiley & Sons, (1987).
- [8] Jewell, J. L., A. Scherer, S. L. McCall, Y. H. Lee, S. J. Walker, J. P. Harbison and L. T. Florez, "Low threshold electrically-pumped vertical cavity surface-emitting microlasers," *Electron. Lett.* **25**, 1123-1124, (1989).
- [9] Lang, R., K. Kasahara, and M. Sakaguchi, "VSTEP and its applications," *Proc. 1990 International Topical Meeting on Optical Computing*, April 8-12, Kobe, Japan, 365, (1990).
- [10] Lin, T.-H., A. Ersen, J. H. Wang, S. Dasgupta, S. Esener, and S. H. Lee, "Two-dimensional spatial light modulators fabricated in Si/PLZT," *Appl. Opt.*, **29**, pp. 1595-1603, (Apr. 10, 1990).
- [11] Kiamilev, F., S. Esener, R. Paturi, Y. Fainman, P. Mercier, and S. H. Lee, "Programmable optoelectronic multiprocessors and their comparison with symbolic substitution for digital optical computing," *Opt. Eng.*, **28**, no. 4, pp. 396-409, (1989).

Shared Memory Optical/Electronic Computer: Architecture Design

Clare Waterson[†] and B. Keith Jenkins

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

Several abstract models of parallel computation have been developed and studied by the computer science and parallel processing communities [1, 2]. The shared memory models are among the most computationally powerful of these models. They benefit from substantial theoretical foundations, and many algorithms have been mapped onto these models in order to characterize theoretically optimum parallel performance. A number of attempts have been made to develop electronic parallel architectures based on the shared memory model. Most of them have been unsuccessful, primarily due to the complexity of the interconnection network hardware and its associated control.

In this paper the design of a hybrid optical/electronic parallel digital computer, the shared memory optical/electronic computer (SMOEC), and the associated control algorithms are presented. The computer design is derived from the shared memory model of computation, and comprises an integration of electronic processing elements and memories with a reconfigurable optical interconnection network. The use of an optical interconnection network and a revised control strategy that incorporates the capabilities and constraints of optical hardware provides the potential of high computational throughput without the substantial drawbacks and bottlenecks of electronic implementations. Our design simultaneously focuses on three different areas: architectural desires, hardware options, and control algorithm requirements; we find that a decision made in one area can have a strong impact on the other two areas.

The shared memory model of computation [1, 2, 3] consists of a set of processing elements (PEs) that can all communicate (read and/or write) simultaneously with a shared memory. The shared memory comprises a set of cells. In one time step, any PE can communicate with any memory cell; furthermore, all PEs can communicate with different cells of the same memory simultaneously. Several compromises of the shared memory model ideal must be made in order to allow a physical realization. The interconnection network is a critical element of, and usually

the limiting factor in, parallel architectures based on a shared-memory computation model. Therefore, reducing the amount of compromise necessary in the design of this element is essential. A key feature of a shared memory machine is its parallel access to memory. By reducing addressing bottlenecks compared to a conventional von Neumann architecture, a substantial improvement in performance can be achieved.

The functional architecture of the SMOEC is shown in Figure 1. It consists of a bank of N processing elements (PEs) connected to a bidirectional optical interconnection network (OIN), which is connected at its other end to a bank of N memory modules (MMs). The OIN consists of a sequence of S shuffle-exchange stages, where $S = \mathcal{O}[\log N]$. The bank of PEs is also connected to the input of a single electronic shuffle-exchange stage. This is the address computer, which processes address bits from the PEs to compute the control settings for the OIN switches. The control signal distribution & interface buffers the control bits until it is time to switch the OIN. The optical switches are then all set at the same time, opening the OIN for bidirectional communication between connected PEs and MMs. A passive OIN is used to permit high bandwidth data transfer using near-term optical technology; the architectural implications of this choice are substantial.

The SMOEC is designed to be a fine-grained computer ($N \sim 10^2 - 10^5$) although it is more coarse grained than most proposed optical computing architectures. The complexity of each PE is on the level of a microprocessor, and the MM size is expected to be at least 1 Kbyte. Depending on the sizes of N and S , an occasional active "repeater" may be added to the OIN.

The interconnection network has a critical impact on the parallel computer performance. In the OIN, all communication algorithms are completed within $\mathcal{O}[1]$ passes through the network; no buffering of data is needed. This eliminates the formation of "hot spots" [4] within the network. Hot spots are formed when simultaneous references are made to the same memory location. In a buffered network, requests may stack up and overflow the buffer sizes, with backups propagating backward through the interconnection network in a tree pattern, until the backup affects the flow through most or all of the

[†] Also at The Aerospace Corporation, M/S: M4/904, P.O. Box 92957, Los Angeles, CA 90009-2957

Presented at: *Optical Society of America Topical Meeting on Optical Computing*, Salt Lake City, Utah, March 1991.

network.

A shuffle-exchange interconnection network topology [5, 6] provides hardware simplicity (ease of optical implementation) while still exhibiting sufficient topological generality. The shuffle connection is illustrated in Figure 2, and the bypass/exchange switch settings in Figure 3. The extended switch settings are not often used by other shuffle-exchange network designs, but here they are used to allow data to be combined and broadcast. Since passive optical switching was selected for high speed data throughput, the interconnection network is circuit switched. Since switching decisions must be carried out physically separate from the data passing through the passive switches, the electronic address computer is essential for network control. This design allows the control algorithm to be more complex. The interfaces between the optical and electronic signals are designed to be fully parallel (no addressing schemes) to avoid bottlenecks at these conversion points. Optical fibers are used to route signals to and from the PEs and MMs, and to format the signals into pixel arrays for entry and exit to the OIN.

Communication in the SMOEC is handled in separate phases, each consisting of only one type of request (such as read or write) to facilitate conflict resolution. Consider the case of simultaneous read requests; here we provide a simplified overview of this phase of operation. A read phase consists of up to N read requests that must be simultaneously satisfied in $\mathcal{O}[1]$ passes of the OIN. Each PE that needs to read information from a MM forms a read request consisting of the desired MM index (address). These addresses are fed directly into the address computer, from each PE to the corresponding node in the address computer. Using the routing algorithm, the addresses are repeatedly cycled through the electronic single-stage shuffle-exchange, enabling computation of the appropriate switch settings for the OIN. These bits are then buffered (in the control signal distribution and interface) until all control bits are ready to be sent simultaneously to their destinations in the OIN switches. Multiple read requests to the same MM are handled by using the upper and lower combine switch settings. Once the switches are set, bidirectional communication between the desired PEs and MMs is established. The address computer then sends a 1 to each requested MM, requesting data transmission. Each requested MM then generates an optical signal which is serially modulated by its entire memory content, and subsequently routed back to the desired requesting PE(s), with possible distribution (using the upper and lower broadcast switch settings in the OIN) to multiple PEs on this return

trip.

The serialization principle [7, 8] is a basic idea that allows data to be routed very efficiently and permits simultaneous write requests to be combined. Simply put, it is the requirement that if two requests are being sent to a common memory location simultaneously, the result must be the same as if the two requests had occurred in some unspecified order. This principle is used in the SMOEC to enable write requests to be completed within $\mathcal{O}[1]$ passes of the network.

Fetch-and-add is an inseparable combination of a read and a write operation which has been cited in the literature [7, 8] as "an important coordination primitive." The idea is to retrieve a value from a memory location X , and then add a predetermined integer increment value v to it. It is used in parallel algorithms (for example) to allow separate processing elements to have an indication of when a procedure is complete by requiring each processor to do a fetch-and-add operation with $v = 1$ when it completes its processing duties. Such fetch-and-add requests may be handled in a similar manner to the write requests, except that the small amount of data involved is processed solely by the address computer (the OIN is not used in this case), and special memory locations are used.

Several communication algorithms have been developed: read, write, generalized fetch-and-add, data sort, and data broadcast, each operating in its own "phase" as mentioned previously. It is important to note that since the contents of a MM are utilized as a complete package in the read and write algorithms, it is preferred that data be logically grouped together so that each MM contains correlated data. This is in sharp contrast to that of electronic implementations. A hashing function for memory addressing is not needed; in fact, it would only serve to decrease the performance of the SMOEC.

A spatial light modulator (SLM) is used to perform the switching and controlled loss functions within the network. A ferroelectric liquid crystal SLM [9, 10] is used; it provides the ability to switch the polarization of the incident optical signals. An optical bypass/exchange switch is easily implemented using a single SLM per stage if the two adjacent channels are superimposed with orthogonal linear polarizations. The SLM is used as an array of switchable half-wave plates that can either "exchange" the two orthogonal polarizations, or leave them unchanged ("bypass"). However, to additionally implement the extended switch operations of combine (PE to MM) and broadcast (MM to PE) a new modified SLM is proposed: a tri-state SLM. Each pixel of the tri-state

SLM has three states: NULL (a clear pixel), MIX, and SWITCH (a half-wave plate pixel). The MIX state sends light from one linearly polarized channel to both of the other channels. Two tri-state SLMs are required per stage to implement the required bypass/exchange switch states.

The MIX operation may be implemented via a quarter-wave plate or a half-wave plate with axes oriented at 22.5° with respect to the horizontal and vertical axes. The tri-state SLM can be implemented using ferroelectric liquid crystal technology, and one method of fabrication is described in [11]. It may be constructed of either two sandwiched quarter-wave plate layers, or a sandwich of one quarter-wave plate layer and one half-wave plate layer, or two sandwiched layers of half-wave plates oriented at 22.5° and 45° (this last option is used in Figure 4).

The essential switching operation of a full shuffle-exchange stage is illustrated in Figure 4. Figure 4a shows the combine operation. Signals a and b in channels A and B are both routed to channel A . When performed in reverse, this results in a broadcast operation (shown as right-to-left propagation) illustrated in Figure 4b. The signal a entering in channel A is broadcast to both channels A and B . Note that the optical hardware is identical for combine and broadcast.

In the PE to MM direction where combining is taking place, although the output is listed as " $a+b$ ", there is never any physical addition of optical signals. Combining of write signals is performed temporally using the address computer to arbitrate the write process. This allows multiple PEs to write to the same MM in a single pass through the OIN. The address computer ensures that at any given time step within a single write phase, for each MM there is at most one PE writing data to it.

The "broadcast" and "combine" operations (Figure 4) mandate a 50% optical loss per stage. For consistency of signal level, a 50% loss is also incorporated for "bypass" and "exchange" by setting TSLM2 to the MIX state for both cases (causing the 50% loss), and of course setting TSLM1 to NULL or SWITCH, respectively.

A single optical shuffle-exchange stage (Figure 5) includes lenses, lenslet arrays, Wollaston prisms, polarizers, fixed pixellated waveplates (PP), and two tri-state SLMs. One tri-state SLM (listed as Pol/TSLM2/Pol) is sandwiched between two polarizers to let it operate as a controllable 0%/50%/100% loss element. An analysis of the angular dependence of the Wollaston prisms indicates that a constant angular offset (as required in this system) will be provided for incident angles which satisfy a small-angle

approximation. The ray trace (Figure 5) illustrates that pixel size, spacing, and angles are all consistent at the input and output of the stage. Such consistency is essential for cascaded passive optical stages.

In conclusion, the initial design of the SMOEC architecture has resulted in a novel machine that balances the interrelated requirements of architectural desires, hardware options, and control techniques. We have found, for example, that the control capability of multiple PEs writing to the same MM in a single computation phase results in a need for bypass/exchange switch hardware that can combine two inputs to a common output. To the authors' knowledge the SMOEC represents the first application of a general-purpose MIMD shared memory paradigm to optical computing.

References

- [1] S. Fortune and J. Wyllie, "Parallelism in random access machines," in *Proc. 10th Ann. ACM Symposium on Theory of Computing*, pp. 114-118, 1978.
- [2] J. T. Schwartz, "Ultracomputers," *ACM Trans. Prog. Lang. Syst.*, Vol. 2, No. 4, pp. 484-521, Oct. 1980.
- [3] B. K. Jenkins and C. L. Giles, "Parallel processing paradigms and optical computing," in *Proc SPIE Vol. 625 Optical Computing*, pp. 22-29, Jan. 1986.
- [4] G. F. Pfister and V. A. Norton, "'Hot spot' contention and combining in multistage interconnection networks," in *Proc. 1985 Int. Conf. on Parallel Processing*, pp. 790-797, Aug. 1985.
- [5] H. S. Stone, "Parallel processing with the perfect shuffle," *IEEE Trans. Comput.*, Vol. C-20, No. 2, pp. 153-161, Feb. 1971.
- [6] D. S. Parker, Jr., "Notes on shuffle/exchange-type networks," *IEEE Trans. Comput.*, Vol. C-29, pp. 213-222, March 1980.
- [7] A. Gottlieb, R. Grishman, C. P. Kruskal, K. P. McAuliffe, L. Rudolph, and Mark Snir, "The NYU Ultracomputer - designing an MIMD shared memory parallel computer," *IEEE Trans. Comput.*, Vol. C-32, No. 2, pp. 179-183, Feb. 1983.
- [8] A. Gottlieb, B. D. Lubachevsky, and L. Rudolph, "Basic techniques for the efficient coordination of very large numbers of cooperating sequential processors," *ACM Trans. Prog. Lang. Syst.*, Vol. 5 No. 2, pp. 165-189, April 1983.
- [9] J. S. Patel and J. W. Goodby, "Properties and applications of ferroelectric liquid crystals," *Opt. Eng.*, Vol. 26, No. 5, pp. 373-384, May 1987.
- [10] K. M. Johnson, M. A. Handschy, and L. A. Pagano-Stauffer, "Optical computing and image processing with ferroelectric liquid crystals," *Opt. Eng.*, Vol. 26, No. 5, pp. 385-392, May 1987.
- [11] M. O. Freeman, T. A. Brown, D. M. Walba, "Quantized complex spatial light modulators," *submitted to Appl. Opt.*

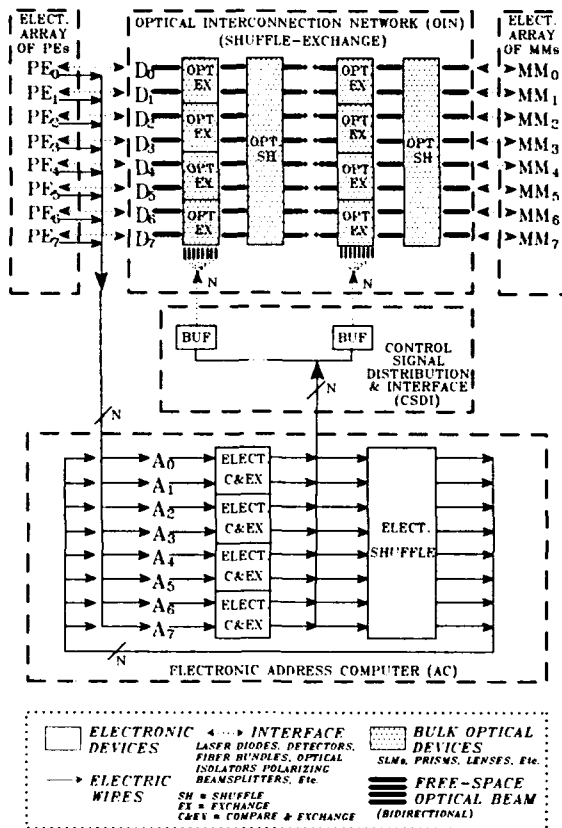


Figure 1: SMOEC: The Functional Architecture

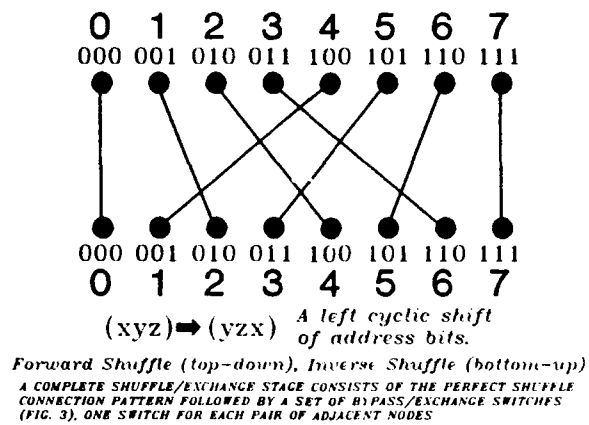


Figure 2: The Perfect Shuffle

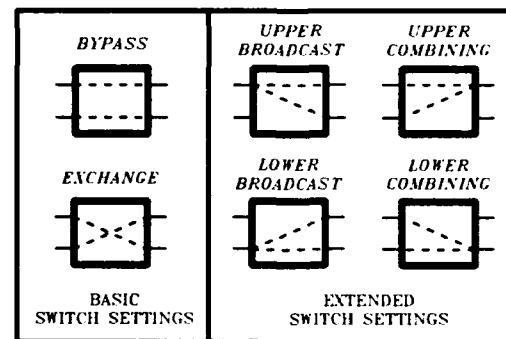


Figure 3: Bypass/Exchange Switch Settings

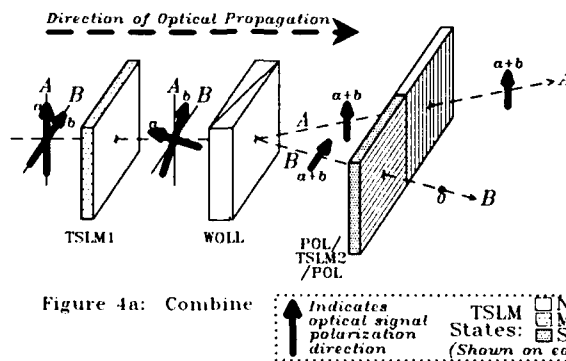


Figure 4a: Combine

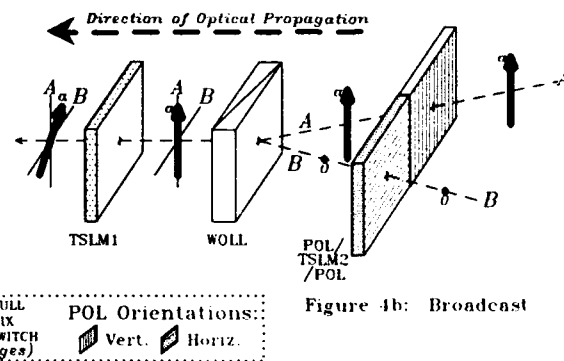


Figure 4b: Broadcast

Figure 4: Overview of Bypass/Exchange Switch Operation

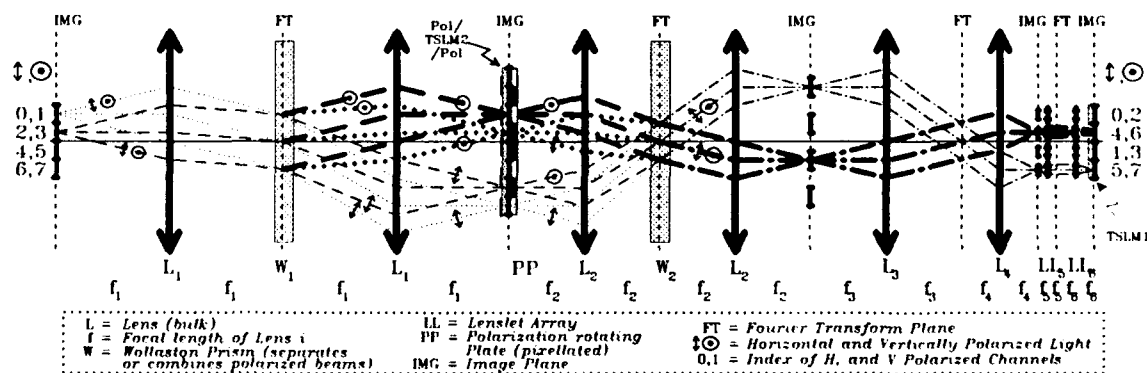


Figure 5: Schematic Diagram and Raytrace of the Optical Inverse Shuffle/Exchange Stage

Design and Construction of a Programmable Optical 16x16 Array Processor

A.C. Walker, R.G.A. Craig, D.J. McKnight, I.R. Redmond, J.F. Snowdon, G.S. Buller,
E.J. Restall, R.A. Wilson, S. Wakelin, N. McArdle, P. Meredith, J.M. Miller, G. MacKinnon,
M.R. Taghizadeh, S.D. Smith, and B.S. Wherrett

Physics Department, Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS, Scotland.

Introduction

There have been a number of significant advances in digital optical computing research over recent years. Experimental demonstrations of optical restoring logic¹, the lock-and clock control of data flow², a programmable optical logic unit³, optical switching networks⁴, and parallel logic modules⁵ have shown that the basic building blocks for a parallel digital optical computing system now exist. This paper describes recent work carried out at Heriot-Watt University in which such a demonstration optical processor has been constructed.

The architecture chosen for the demonstrator project corresponds to the Cellular Logic Image Processor^{6,7} (CLIP). In this implementation, the CLIP may be regarded as an optical co-processor residing within a host electronic computer which provides all the necessary control and clock signals. As the CLIP architecture is of the single instruction multiple data-stream (SIMD) type, the electronic machine can be used to supply program instructions to the optical modules in such a way that its overall computational power is increased by a factor equal to the parallelism of the optical processor. In addition to the high degree of parallelism available from optical systems, such a processor can also exploit the considerable flexibility of optical interconnects between logic plane arrays.

The basic blocks that make up the optical CLIP, shown in figure 1, are as follows. Firstly, a 2-D binary data-field is input via a spatial light modulator (SLM) which, in this case, is electrically addressed. This input plus the result from the previous machine cycle are incident on a processing plane which implements the Boolean logic operation that is being specified by the host computer for that cycle. The output from this stage goes through an interconnect which fans out individual signals to other positions in the array. A programmable NOR/NAND gate array is then used to convert the signals fanning together at the output of the interconnect module into binary outputs, according to the (programmable) chosen threshold level. Finally, the output from this stage is passed to a latching NOR-gate array which acts as a temporary memory, holding the result of that cycle until required as an input for the next. As all the logic arrays are made up of bistable, latching gates, the data can be circulated around the processing loop in the standard lock-and-clock fashion². The final output, after the requisite number of iterations, is available on a CCD camera.

A variety of primitive image processing algorithms can be implemented with such a machine. Operating on binary images and using a nearest-neighbour interconnect, these include image compression/decompression, noise removal, edge detection, path finding, etc. Use of 2-D perfect shuffle interconnects permits efficient implementation of more powerful algorithms such as fast sorts and discrete Fourier transforms.

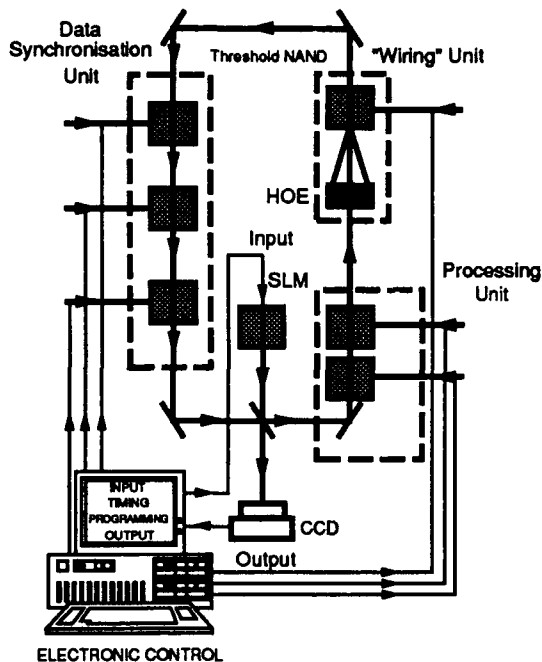


Figure 1. Block diagram of the O-CLIP.

Preliminary Experiments

An iterative digital optical processor, corresponding to a single channel of the optical CLIP, has already been successfully demonstrated at Heriot-Watt in the past year⁸. This system has demonstrated a range of serial processing operations on an optical input data stream including word recognition, number comparison, full addition and subtraction⁹.

In addition, two of the minimum of three optical logic planes required for the CLIP were tested in preliminary experiments⁸. These were both 15x15 logic arrays with a simple one-to-one imaging interconnect between them. The parallel transfer and latching of arbitrary data patterns from one array to the next was shown to be achievable. This logic array demonstration, together with the single channel processor and the SEED based systems constructed by AT&T, have been seen as significant milestones in the advance of digital optical processors. The full optical CLIP demonstrator described here represents an important next-step in this developmental sequence.

Experimental Details of the Optical CLIP.

The module upon which the processor is based is shown in figure 2. The power level of the beam that provides the main input to the array is first set by an acousto-optic modulator. The beam is then expanded and is incident on a Damman grating^{10,11} which converts the input into a two-dimensional array of equally intense beams. The components directly after the grating are a polarising beam splitter (PBS), quarter wave plate (QWP) and a triplet lens which Fourier transforms the output from the grating to produce the required focal spot array on the optical logic plane. Directly after the triplet is a monitoring beam-splitter which splits off a small fraction (~ 5%) of the input image for quality analysis. The reflected (output) beams are separated from the input at the PBS. This output is re-imaged as the input signal array onto the next logic plane. The monitoring beam-splitter is also used to sample the output from the device in order to read the states of all of the separate elements in the logic plane (via a CCD camera). The monitoring beam-splitter on the signal side of the logic plane receives the small amount of the power beams transmitted by the device and some of the reflected input signal, hence allowing correct registration of the two arrays.

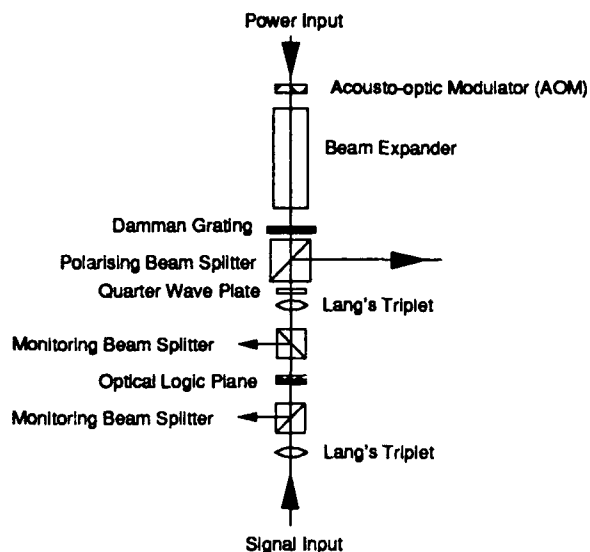


Figure 2. Circuit schematic for powered optical logic plane.

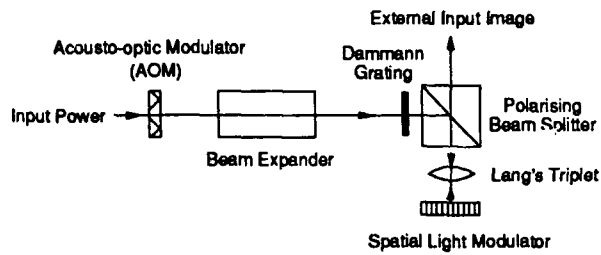


Figure 3. External input schematic.

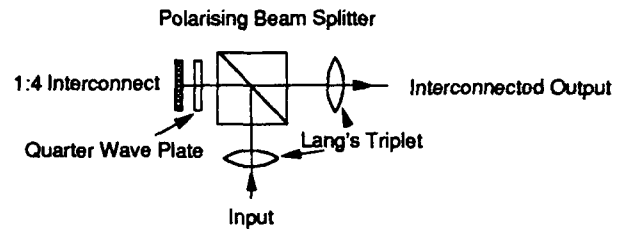


Figure 4. Reflection interconnect schematic.

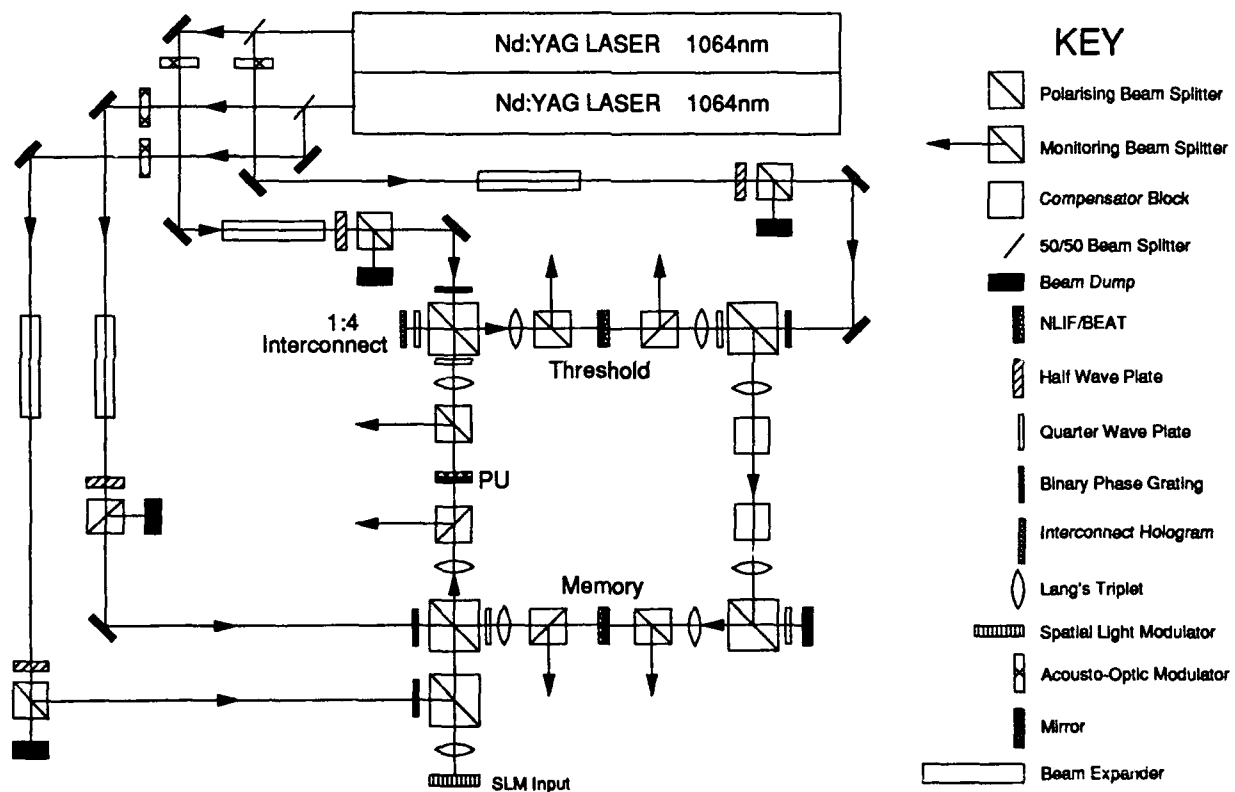


Figure 5. Full circuit implementation of the 16x16 O-CLIP.

The optical logic planes being used in this processor are ZnSe BEAT devices operating at a wavelength $\lambda = 1064\text{nm}$ and utilising a non-resonant thermal refractive response^{12,13}. These devices consist of a sapphire substrate with a polyimide coating, upon which the interference filter structure is deposited. On top of the filter structure, a thin absorbing layer of germanium is deposited to provide the absorptive mechanism for switching and efficient input coupling. The hold power beam array is incident on the filter side of the device and the input array is incident on the absorbing layer. This form of operation produces an inverting gate, i.e. a switching from high to low reflection output as the signal is increased, with the hard-limiting, latching response required for the digital circuitry. Switching times are in the range 10-100 μs .

For the external input to the optical CLIP, an additional array of power beams is reflected from the active elements of the SLM (see figure 3). At present, two forms of electrically addressed SLM's are being studied, one being a silicon back plane addressed liquid crystal device as developed at Edinburgh University¹⁴ and another based on ZnSe interference filter technology and known as the Electron-beam Tunable Interference Filter¹⁵ (ETIF).

The interconnect optics are shown in figure 4. The input image is Fourier transformed and reflected from a PBS, through a QWP on to the interconnecting element. The reflected beam is transmitted through the PBS and the inverse transform produces the interconnected spot array. At present, various nearest-neighbour interconnects are being developed including angled quadrant mirrors (both fixed and adjustable) and reflection holograms.

The circuit schematic for the full processor is shown in figure 5. Three optical logic planes are required for the implementation. The first plane in the system acts as the processing unit (PU) and can be programmed to produce any one of the four Boolean logic functions ON, OFF, NAND or NOR by correct setting of the input power beam level. The second logic plane acts as the thresholding unit, its output being passed on to the memory plane which completes the circuit. Clocking the three power beams in the correct manner produces a lock-and-clock synchronisation of the binary images circulating in the processor loop.

By using the appropriate data acquisition, analogue output and image grabbing systems, the host computer can be programmed to provide all of the timing signals required, determine the required holding levels for each function in the PU and threshold units, provide input data and sample various outputs from the processor.

Conclusions

The design and implementation of a 256 channel digital optical processor has been described. Results from this demonstrator system will prove invaluable in assessing the potential of this form of parallel optical processor and the demands placed on the associated technologies when expanding to a higher degree of parallelism.

Acknowledgements

This work has been supported by the Boeing Company, Seattle, USA and the UK Science and Engineering Research Council.

1. A.C. Walker *et al.*, Phil. Trans. Roy. Soc. London A **313**, p. 357-360 (1984).
2. S.D. Smith *et al.*, Nature **325**, 27-21 (1987).
3. R.G.A. Craig *et al.*, Appl. Opt. **29**, pp. 2148-2152 (1990).
4. F.A.B. McCormick *et al.*, CLEO, CPDP1 (1990).
5. M.E. Prise *et al.*, Appl. Opt. **29**, pp. 2164-2170, (1990).
6. M.J.B. Duff and T.J. Fountain, "Cellular Logic Image Processing", Academic Press (1986).
7. B.S. Wherrett, SPIE **769**, pp. 7-26 (1987).
8. B.S. Wherrett *et al.* Proc. SPIE **1215**, p. 264-273 (1990).
9. R.G.A. Craig *et al.*, Submitted to Applied Optics, September 1990.
10. H. Dammann and K. Götler, Opt. Commun. **3**, p. 312 (1971).
11. J. Turunen, A. Vasara, J. Westerholm, G. Jin, A. Salin, J. Appl. Phys. **21** pp. 102-105 (1988).
12. A.C. Walker, Opt. Commun. **59**, 145 (1986).
13. G.S. Buller, C.R. Paton, S.D. Smith and A.C. Walker, Opt. Commun. **70**, 522 (1989).
14. D. J. McKnight, D.G. Vass, and R.M. Sillito, Appl. Opt., **28**, 4757 (1989).
15. A. C. Walker, S. D. Smith, R. J. Campbell, and J. G. H. Mathew, Opt. Letts., **13**, p345 (1988).

Digital Optical Computer II (DOC-II): "Performance Specifications"
Peter S.Guilfoyle, Ronald S.Rudokas, Richard V.Stone, Edward V.Roos
OptiComp Corporation
P.O. Box 10779, Zephyr Cove, NV 89448
702-588-4176

Introduction:

Currently OptiComp® Corporation is completing construction of the first 32-bit general purpose digital optical computer. This effort is sponsored jointly by the Office of Naval Research (ONR), Strategic Defense Initiative (SDI), NASA space station and the Rome Air Development Center (RADC/USAF).

OptiComp has contracted Harris Corporation of Melbourne, FL to build and deliver the 64 channel acousto-optic spatial light modulators, Spectra Diode Laboratories of San Jose, CA to build and deliver the 8 element index guided laser diode bars capable of producing in excess of one watt TE00 at 830nm, Optical Research Associates (ORA) in Pasadena, CA to optimize the optical interconnects, and GE of Vaudreuil, Quebec to develop a 128 element high speed APD array.

This paper describes how these components are integrated to construct a general purpose 32-bit Digital Optical Computer (DOC II). The architecture allows the emulation of any digital instruction set (SUN RISC was chosen because of its wide utilization). For a more detailed discussion on the architecture and logic implementation please refer to references 1-4. The system has the following performance parameters.

System clock rate:	100 MHz
Input data rate:	12.8 G bits per second
Energy per effective gate:	< 600 attojoules (<2500 photons)
Peak compute rate:	>1 tera binary operations per second
Logical primitive:	64 bit minterm AND-OR

The architecture is shown in Fig. 1. The optical CPU consists of 3 planes of I/O. Plane 1 is a point source acousto-optic light modulator used to digitally amplitude modulate the laser. Plane 2 is a multi-channel spatial light modulator used to input the control operator microcode. Plane 3 is the 128 element APD which performs the minterm functional evaluation..

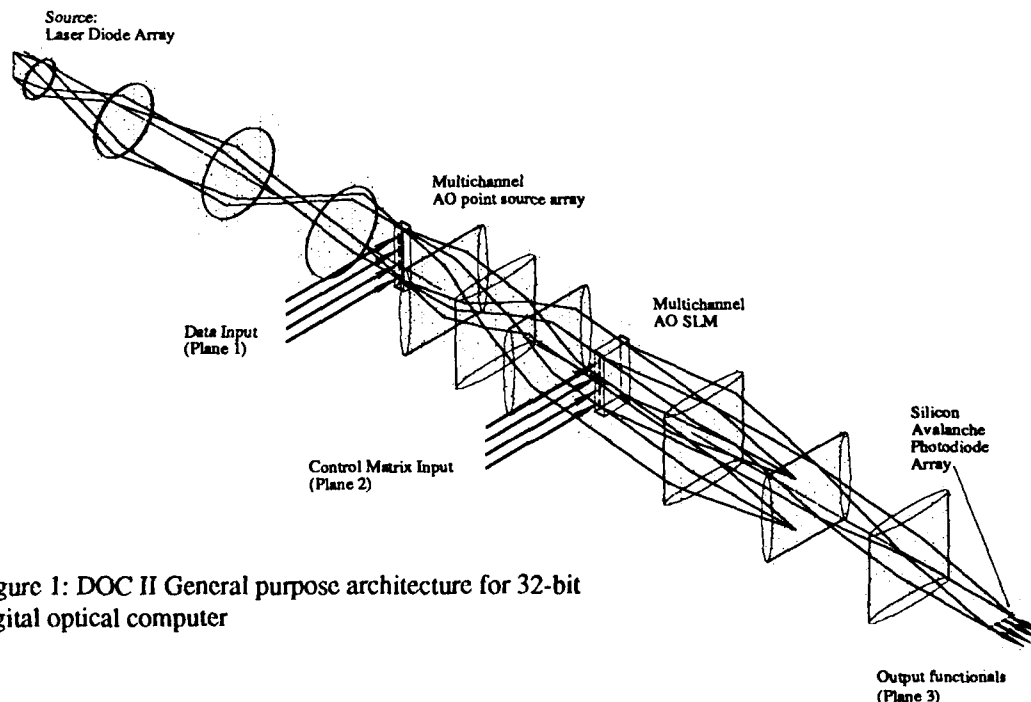


Figure 1: DOC II General purpose architecture for 32-bit digital optical computer

Optical Signal Path

The system consists of two main sections: the illumination assembly and the modulation relay assembly. The illumination assembly includes the laser diode collimators (LDC), 10X imager, angular aerial image combiner (AAIM), and the 1X relay. The LDC and the 10X imager, when taken together comprise the 10X relay assembly. These sections are shown in figures 2 and 3.

Laser Diodes:

The illumination source consists of eight individual arrays of eight emitter laser diode bars. These arrays consist of individual index guided GaAlAs semi-conductor lasers. Each emitter is single mode, both longitudinally and laterally. Spot size at the facets is 1.3 by 4.2 microns. Astigmatism is $< 1\mu\text{m}$. The active emitters are spaced at 100 microns giving an over-all active source length of 700 microns.

Table 1: Laser Diode Specifications

Wavelength	837nm
Optical distribution	Elliptical Gaussian
Divergence	$28^\circ \times 10^\circ$
Mode	TE00
Optical power	30mw per emitter
Drive current	40ma / element
Emitter spacing	100 μm
Wavelength uniformity	$\pm 1\text{nm}$
Power uniformity	$\pm 10\%$

Laser Diode Collimators / 10X Imager:

The design of the laser diode collimators consists of six optical elements in a configuration similar to an Amici microscope objective. A low level of back reflection coupling is required to prevent beat frequency interference between emitters. The 10X imager is similar to a Tessar lens. The 10X relay lens is packaged, assembled and tested as a module. These modules are integrated into the system without further adjustment of their internal sub-assemblies. If a module fails, it can be replaced with minimum impact to the system.

Table 2: Laser Diode Collimator/Imager

Magnification	10X $\pm 2\%$
Object height	0.8 mm x 0.8mm
Numerical aperture	$x=0.16, y=0.653$
Wavefront error(RMS)	$< .066$
Telecentricity	$< 0.1^\circ$
Spot size output	$x=12.5, y=42.5(\mu\text{m})$
Back reflection	70 db

Angular Aerial Image Combiner:

A special mirror array is required to angularly multiplex the aerial images produced by the the laser array. Figure 3 shows the use of a conventional Wiley Angular Aerial Image Multiplexer (AAIM) used to fan in the eight laser array bars into 64 collinear beams.

Table 3: Angular Aerial Image Combiner

Facet angles	8 (4 symmetric)
Facet width	250 μm
Facet length	$> 5\text{mm}$
Number of facets	64

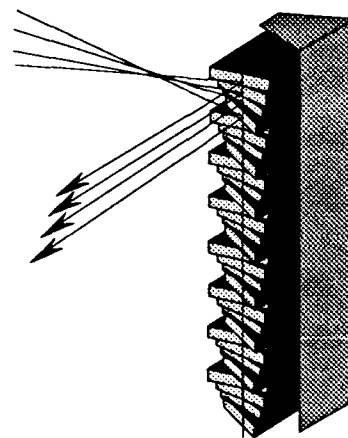
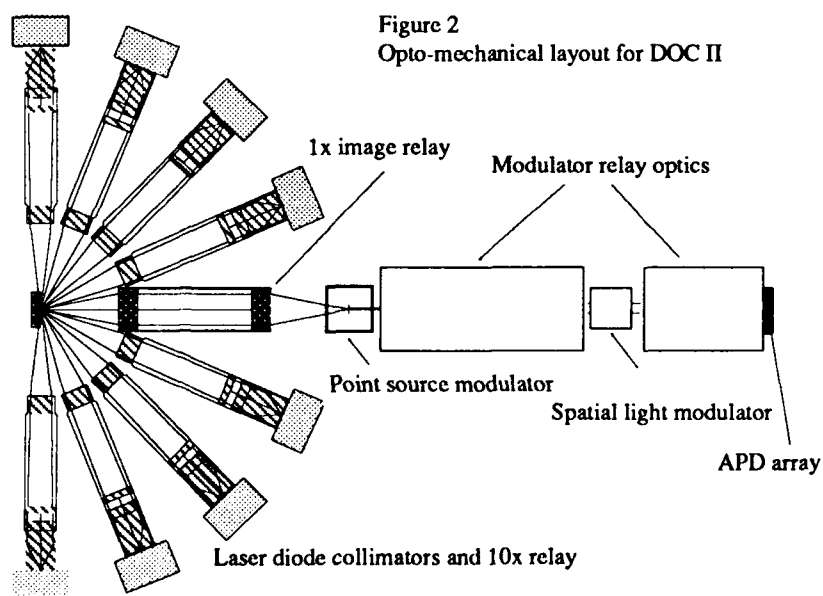


Figure 3
Wiley Angular Aerial Image Multiplexer (AAIM).

1X Relay:

The 1X relay is a simple telecentric symmetric relay consisting of six elements and a field stop. It relays the image of the emitter array at the image combiner to the point source array at a one to one magnification. In addition this assembly contains a half wave plate for polarization orientation.

Table 4:	1X Relay
Magnification	1X $\pm 2\%$
Object height	16mm
Numerical aperture	x = 0.016, y = 0.065
Spot size	x = 42.5, y = 12.5 (μm)

Point Source Array (PSA):

The PSA contains 64 individually addressable modulator channels which modulate the optical beams with the input digital data vector. A longitudinal mode GaP Bragg cell design was implemented because of high diffraction efficiency, high acoustic velocity and low acoustic attenuation. An apodized transducer height of 50 μm and a transducer center to center spacing of 250 μm was selected for optimum compatibility with the geometry of optical characteristics of the laser diode array. In addition the crystal is mounted to a flexible buried strip line printed circuit board to allow for fan out and minimum crosstalk.

Table 5:	Point Source Array
Drive power	200mw
Crystal & Mode	GaP L [110]
Diffraction eff.	>37% @ 200mW
Contrast ratio	>30 dB
Channel crosstalk	>30 dBc
Rise time	<2ns
Data rate	6.4 G bits / sec

First Anamorphic Relay:

The first anamorphic relay illuminates the second Bragg cell with the modulated laser energy from the PSA. In the y-z plane, the light diffracted by the PSA must be collimated to fill the 2.56 μs time aperture of the SLMA.

Table 6:	First Anamorphic Relay
Spot size	x = 48 μm y = 10.6mm

Spatial Light Modulator Array (SLMA):

The SLMA also contains 64 individually addressable channels. A shear mode [1,-1,0] GaP Bragg cell design is used because of its acoustic self collimation property, high diffraction efficiency and low attenuation. Unlike the PSA transducer apodization is not required because the beam is within the Fresnel region. The transducers are 50 μm in height with a center to center spacing of 250 μm .

Table 7:	Spatial Light Modulator Array
Drive power	200mw
Material	GaP (shear mode)
Diffraction eff.	>4% @ 200mW
Time aperture	>2.56 μs
Bandwidth	>400 MHz
Channel crosstalk	>30 dBc
Data rate	<6.4 G bits / s

Second Anamorphic Relay:

The second anamorphic relay illuminates the APD array with the laser energy from the SLMA. In the x-z plane the lens must focus the energy from each of the 64 transducers to the APD array. The design is configured for front to back focal plane (Fourier transform) operation.

Table 8:	Second Anamorphic Relay
Spot size x	2.0mm
y	150 μm

Avalanche Photodiode Detector (APD):

The APD is a monolithic array of 128 avalanche photodiodes with a common anode. The APD is fabricated utilizing an integrated array of 128 lens elements to reduce the dead space between channels and improve gain uniformity across the array.

Table 9:	Avalanche photodiode array
Element size	0.145mm X 2.0mm
Element spacing	150 μm
Gain	50X
Rise time	<1.8ns
Element crosstalk	<0.5%

Electronics:

In DOC II the electronics plays a supporting role of providing RF signals to drive the Bragg cells and converting the optical result into an ECL signal for evaluation or recirculation into the optical processor. Although straight forward the

high speeds required to support DOC II push the electrical design to the limits of ECL memory and GaAs logic performance. The three major electronic subsystems are the Bragg cell drivers, the APD amplifiers and the op-code and resultant data arrays.

Bragg Cell Drivers:

The Bragg cells require a minimum of 0.2W of 800MHz RF energy, on/off modulated with the digital signal, to modulate the laser energy. GaAs logic is used to combine the 800MHz carrier with the logic signal. A processor controlled variable output amplifier drives the Bragg cell.

Table 10:	Bragg Cell Drivers
Center frequency	800 MHz
Bandwidth	500 MHz
Output power	0.5W to 0.1W

APD Amplifiers:

A low noise wide band transimpedance amplifier is used to convert the optical energy to a useful signal. A 289MHz low pass filter is used for pulse shaping to improve SNR and reject crosstalk. A threshold decision is made using a high speed comparator with processor controlled threshold.

Table 11:	APD Amplifiers
Bandwidth	289 MHz
Rise time	<2.0 ns
RMS noise	1.8pA/ $\sqrt{\text{Hz}}$
BER	10 ⁻¹⁰
Crosstalk	<30dB

Memory Arrays and Logic:

To support the storage of op-codes and data, arrays of high speed ECL memory are used. The control and sequencing of DOC II is accomplished using ECL PAL's and GaAs discrete logic.

Table 12:	Memory Array and Logic
Memory data rate	>100MHz

Conclusions:

Designing and constructing a flexible 32-bit, high speed general purpose digital optical computer can be achieved. We have previously demonstrated that through the use of attainable state of the art technology a real-world digital optical computer can be constructed. Shannon's theorem, Morozov's control operator method, combinatorial arithmetic, and DeMorgan's law point to optical freespace interconnects as a reasonable next step that can sidestep the limitations of planar semiconductor technology.

Acknowledgements:

OptiComp corporation wishes to acknowledge the financial support of the Office of Naval Research, the National Aeronautics and Space Administration, the Rome Air Development Center, and the Strategic Defense Initiative Office for the funding of this program under contracts N00014-87-C-0077, N00014-89-C-0266, and N00014-89-C-0225.

References:

- [1.] P.S. Guilfoyle, "Digital Optical Computer Fundamentals, Implementation, And Ultimate Limits," Proceedings of the SPIE, Vol. 1214-15, and in the Critical Review Series, Digital Optical Computing, Volume CR 35, Los Angeles CA, 16 January. 1990 (invited)
- [2.] P.S. Guilfoyle, DR Pape, J.A. Carter, R.V. Stone, "32 Bit digital optical computer: A hardware update," Proceedings of the SPIE (invited), Vol 1296-02, Advances in Optical Information Processing IV, April 1990
- [3.] P.S. Guilfoyle, F.F. Zeise, "Reconfigurable Programmable Optical Digital Computer," Proceedings of the 1989 Third Topical Meeting on Optical Computing, Salt Lake City, Optical Society of America, Feb. 1989.
- [4.] P.S. Guilfoyle, "General Purpose Optical Digital Computer." Proceedings of the topical meeting of the International Commission for Optics (ICO)-- Optical Computing '88, Toulon, France, Sept. 1, 1988. (invited)
- [5.] P.S. Guilfoyle, "Programmable Optical Digital Computing," Proc. of the 21st Annual Asilomar Conference on Signals, Systems and Computers, Nov. 3, 1987.

Tuesday, March 5, 1991

Fuzzy and Cellular Systems

TuB 10:50am–12:40pm
Salon F

Kelvin Wagner, *Presider*
University of Colorado, Boulder



ADAPTIVE FUZZY SYSTEMS

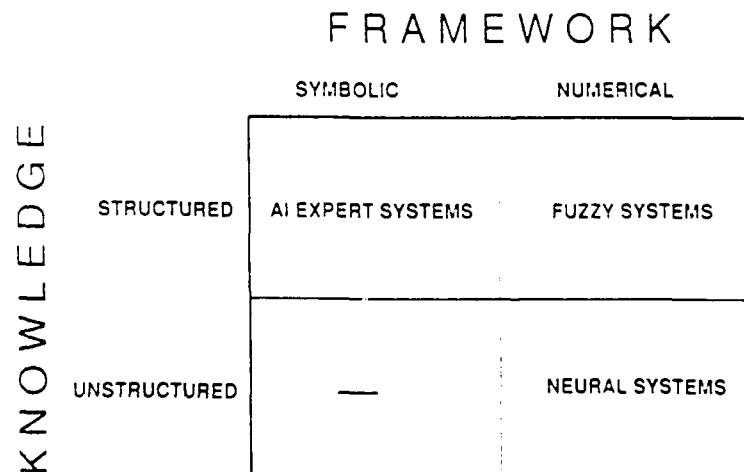
Bart Kosko
Electrical Engineering Department
Signal and Image Processing Institute
University of Southern California
Los Angeles, California 90089

Fuzziness is multivaluedness. Truth, set membership, and subset containment take on degrees in $[0, 1]$ instead of taking on only the limiting bivalent extremes in $\{0, 1\}$. The degrees in $[0, 1]$ define fuzzy units, or fits. Statements are true to some degree. An element belongs to, or fits in, a fuzzy set to some degree. One fuzzy set contains another fuzzy set to some degree.

Fuzzy subsets of set $X = \{x_1, \dots, x_n\}$ define points in the unit hypercube $I^n = [0, 1]^n$. The 2^n vertices of I^n , the Boolean n -cube lattice, define the power set 2^X , which contains the 2^n nonfuzzy subsets of X . The cube midpoint P is maximally fuzzy, equidistant to all vertices, and maximally breaks the "laws" of noncontradiction and excluded middle since $P = P^C$. Midpoint phenomena, such as Cretans who say that all Cretans lie and half-empty glasses, generate "paradoxes" in bivalent systems.

Fuzzy systems estimate functions without a mathematical model of how outputs depend on inputs. Mathematically fuzzy systems define mappings between unit hypercubes. These fuzzy associative memories (FAMs) associate output fuzzy-set descriptions with input fuzzy-set descriptions. Expert advice and engineering judgement generate fuzzy systems. Sample data generates adaptive fuzzy systems, time-varying mappings between fuzzy cubes.

Fuzzy systems resemble AI expert systems and neural network systems. All three behave as model-free estimators, mapping inputs to outputs without an assumed transfer function. The following figure shows a taxonomy of model-free estimators:



AI systems encode structured knowledge as propositional rules and process this information in a symbolic framework. Symbol

processing prohibits direct numerical mathematical analysis and hardware implementation.

Neural systems process information numerically but encode unstructured knowledge as input-output data samples fed to a black box we cannot examine. In general we do not know what the neural network has learned or what it will forget when it learns new samples. We can check only the output responses of the black box for all combinations of input variables, in general a prohibitive task. The neural system is unreliable if we do not check all cases and unnecessary if we do, since then we can store and use the input-output pairs in a lookup table.

Fuzzy systems encode structured knowledge but process it numerically. Fuzzy-associative-memory rules resemble if-then propositional rules: "IF the traffic is HEAVY, then keep the light green LONGER." A traffic engineer may state the FAM rule (HEAVY, LONGER) in this linguistic form without a numerical specification of the fuzzy subset HEAVY of traffic density and the fuzzy subset LONGER of green-light durations.

Beneath the words and symbols lies a numerical representation. FAM rules define large fuzzy outer-product matrices. In practice they define continuously infinite matrices. We cannot write these infinite matrices down and do not need to. Instead we use a virtual-representation scheme when input data initiates a fuzzy inference. A road sensor measures a traffic density value of 63 cars in a queue. According to our fuzzy-set definition, 63 cars may equal a heavy traffic measurement to degree .9, an exact numerical value that passes through the numerical system as a scaled bit vector or a delta pulse.

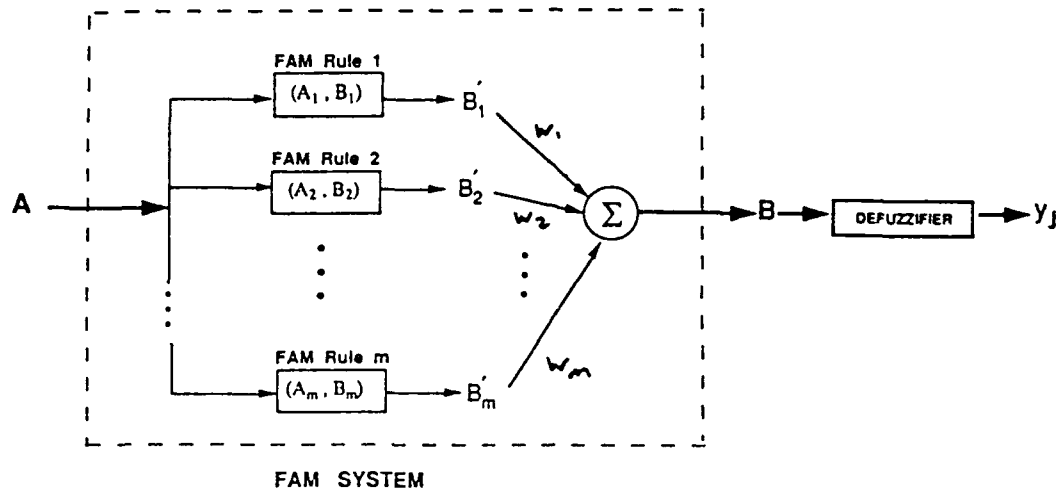
Fuzzy variables assume fuzzy-set values. In the traffic-control example the fuzzy variable TRAFFIC DENSITY takes on the fuzzy-set values LIGHT, MEDIUM, and HEAVY. In a control or mechanical system the fuzzy variable ANGULAR VELOCITY might assume the fuzzy-set values NEGATIVE MEDIUM, NEGATIVE SMALL, ZERO, POSITIVE SMALL, and POSITIVE MEDIUM, each defined with a symmetric trapezoid or triangle centered over values in, say, the angular-velocity interval $[-100, 100]$.

We can create fuzzy systems by entering FAM rules in a FAM-rule matrix. To control an inverted planar pendulum, principles of symmetry and error-nulling may lead to this band of FAM rules:

$$\theta$$

		NM	NS	Z	PS	PM
$\Delta\theta$	NM			PM		
	NS			PS	Z	
	Z	PM	PS	Z	NS	NM
	PS		Z	NS		
	PM			NM		

Fuzzy systems process several FAM rules in parallel. Every input fires every FAM rule to some degree. The following figure shows a minimal FAM system architecture:



The FAM system stores each FAM rule separately. This consumes space but avoids crosstalk and preserves a modular structure. In contrast, a neural network would add together or superimpose the FAM rules (or FAM-rule matrices), which saves space but ensures crosstalk and eliminates the modular structure. Neural systems sum throughputs. Fuzzy systems sum outputs.

The FAM system sums weighted output fuzzy sets. In nonadaptive fuzzy systems the fuzzy engineer implicitly chooses the FAM-rule weights as 1s or 0s when he includes or omits a FAM rule. Neural or statistical adaptation schemes can select the FAM-rule weights w_i as a function of system sample data. Adding fuzzy sets tends to invoke the fuzzy version of the Central Limit Theorem, producing a symmetric unimodal output fuzzy set. A centroid computation defuzzifies the output fuzzy set and generates an exact numerical output value. Analog and digital fuzzy VLSI chips execute hundreds of thousands of these FLIPS, or fuzzy logical inferences per second.

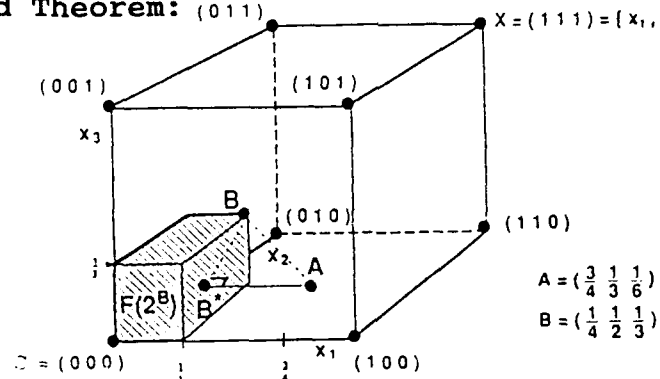
Fuzzy system theory depends on fuzzy set theory, and this reduces to the new concept of subsethood or degree of set containment. The quantity $S(A, B)$ denotes the degree to which A is a subset of B . In general $0 < S(A, B) < 1$ holds. If A equals a singleton set, if $A = \{x_i\}$, then subsethood reduces to multivalued elementhood: $S(\{x_i\}, B) = b_i$, the degree to which x_i belongs to fuzzy set B .

Subsethood reduces probability to set theory. The reduction depends on the unique l^p -norm extension of the Pythagorean Theorem:

$$||A - B||^p = ||A - B^*||^p + ||B^* - B||^p,$$

for all $p > 1$ and $p = 1$, not just for $p = 2$. In the unit cube I^n these "orthogonal" relations hold for exactly 2^n fuzzy sets B^* .

Each fit value b_i^* must equal fit value a_i or b_i . This leads to the Subsethood Theorem: (011)



$$S(A, B) = \text{DEGREE}(A \subset B) = 1 - \frac{\text{DISTANCE}(A, B^*)}{M(A)}$$

$$= \frac{M(A \cap B)}{M(A)} = \frac{3/4}{5/4} = \frac{3}{5} = 60\%$$

The shaded hyperrectangle defines the fuzzy power set $F(2^B)$, all fuzzy subsets of B . B^* equals the fuzzy subset of B closest to A . The fuzzy count $M(A)$ equals the l^1 or fuzzy Hamming distance from the origin (empty set) to A and hence equals the sum of A 's fit values: $M(A) = a_1 + \dots + a_n$.

Subsethood depends on distance. If A lies in B 's hyperrectangle, then A is a full subset of B , and $S(A, B) = 1$. The closer A is to B 's hyperrectangle, the more B contains A , and the larger the value $S(A, B)$. The distance $d(A, B^*)$ drives the subsethood measure. Boundary conditions lead to the choice $M(A)$ as normalization value. Then $S(A, B) = 1 - d(A, B^*)/M(A)$. The Subsethood Theorem shows that this equals a cardinality ratio that has the same form, and implies the axioms of, the conditional probability $P(B|A)$. The Subsethood Theorem derives what axiomatic probability defines.

The general subsethood thesis equates the probability $P(A)$ with the degree to which event A contains its own sample space X :

$$S(X, A) = P(A),$$

the degree to which the part contains the whole, an absurd relationship outside of multivalued theory. Note that in general $P(A) = P(A|X)$, which has the same form as the subsethood thesis.

Relative frequency reduces directly to subsethood. Let A denote the set of n_A successful trials out of n trials. A defines a bit vector of n_A 1s and $n - n_A$ 0s. Sample space X defines the bit vector of n 1s, the set of all trials. A intersected with X still gives A , with count $M(A) = n_A$. X has count $M(X) = n$. Then the Subsethood Theorem gives the relative frequency n_A/n as the subsethood value $S(X, A)$:

$$S(X, A) = n_A / n.$$

Kosko, B., Neural Networks and Fuzzy Systems: A Dynamical Systems Approach to Machine Intelligence, volumes I and III, Prentice-Hall, 1991.

Optoelectronic Fuzzy Logic System

Gary C. Marsden, Brita Olson, Sadik Esener, Sing H. Lee
University of California, San Diego
Department of Electrical and Computer Engineering
La Jolla, California 92093

It is often the case in reasoning problems that propositions are neither entirely true nor entirely false. In fuzzy logic,^{1,2} the truth values of propositions are not restricted to true or false, but rather may range between zero (absolutely false) and one (absolutely true), allowing a quantitative representation and evaluation of vague propositions. For example, the proposition, "Marsden is a boring speaker" is neither totally true nor totally false, but might have a value 0.30.³ Many existing Boolean reasoning methods can be extended to include fuzzy truth values. However, since Boolean operators such as AND and OR are undefined on non-Boolean data, analogous fuzzy operators must be defined for these algorithms to be useful. It has been shown that MIN and MAX have desirable properties when used as extensions of AND and OR, respectively.¹

In this paper we are concerned with the parallel implementation of the logic function Modus Ponens. In Modus Ponens, a proposition y_i is inferred to be true if both x_j and $x_j \rightarrow y_i$ are true. For simplicity of discussion we shall assume that the value, y_i is initially zero. Thus, the truth value for y_i is given by,

$$y_i = x_j \text{ AND } x_j \rightarrow y_i \quad [1]$$

With the appropriate substitutions of MIN for AND we can extend Modus Ponens to fuzzy logic,

$$y_i = \text{MIN}[x_j, x_j \rightarrow y_i] \quad [2]$$

A parallel algorithm for Boolean Modus Ponens inference was developed for use on an optical matrix-vector multiplier with binary thresholding on the output vector.⁴ In this algorithm, truth values are encoded as either zeros or ones. The matrix element M_{ij} represents the truth value of the implication $x_j \rightarrow y_i$. The product $M_{ij} * x_j$, which is equivalent to an AND, determines if y_i is true due to implication from x_j . If the sum of these products over index j is greater than zero, that is, if at least one of the AND operations is true, then y_i is implied from the input vector x . Boolean encoding is maintained by thresholding the output of the matrix-vector multiplication,

$$y_i = T\left(\sum_j M_{ij} * x_j\right) \quad [3]$$

The summation/threshold is equivalent to a global OR. Therefore the use of an optical matrix-vector multiplier allows many truth values, represented by the output vector y , to be inferred in parallel from the set of input values in the vector x . This algorithm can be extended to fuzzy inference by substituting MIN for the local (AND) multiplication and MAX for the global (OR) summation/threshold operation, with data ranging between [0,1]. That is,

$$y_i = \text{MAX}_j[\text{MIN}[M_{ij}, x_j]] \quad [4]$$

For the Boolean Modus Ponens algorithm, a standard transmissive optical matrix-vector multiplier is sufficient. Unfortunately, the local MIN and global MAX operations of the fuzzy algorithm are difficult to implement with such architectures. Nonlinear optical components might offer a solution but would be subject to dynamic range and response time limitations. Optoelectronic architectures, on the other hand, offer both the desired parallelism, through optical communication, and functionality, through tailored electronic circuitry.

An array of binary tree structures can be used to perform the necessary generalized matrix-vector multiplication. Figure 1 shows an abstract model of one processing element (PE) in this architecture. Each PE is dedicated to one element of the output vector. Elements of the input vector are transmitted optically to the electronic leaf units of the tree. These leaf units have local memory, which store the appropriate matrix elements, and logic circuitry to perform the necessary MIN operation. The results are passed down the tree, where at each intermediate fan-in unit, a MAX operation is performed. It is easily seen that the necessary combination of local MIN and global MAX operations is performed.

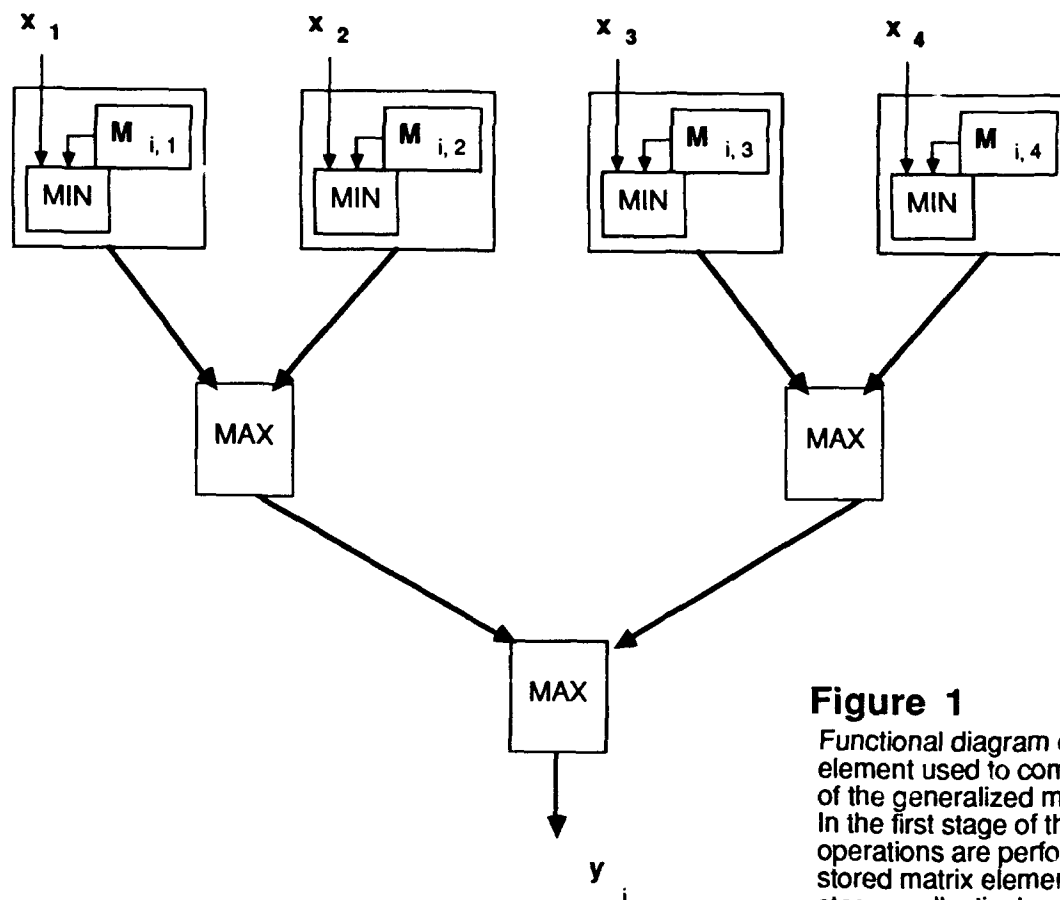
Although the results of these operations must traverse $\log_2 N$ stages of fan-in units, the proper choice of data representation allows a fully pipelined system. Fuzzy values are transmitted serially, most-significant-bit first. Figure 2 shows the operation of a MIN circuit. After reset, the circuit performs successive bitwise comparisons. As long as no difference is detected, the most significant bits which are common to both values are passed to the next stage of the tree. The smaller value is determined by a zero at the first bit level where the two values differ. Once the smaller value has been determined, the circuit passes the remaining bits of this value to the next stage of the tree. MAX circuits operate in a similar fashion, passing the larger value. A significant advantage of this methodology is that the length of the digital fuzzy value can be set to any desired accuracy.

The D-STOP architecture⁵ allows an efficient implementation of this binary tree structure. The system consists of an array of N processing elements (PEs). Each PE, as shown in Figure 3, consists of N processing sub-units having detectors and local memory. These detector units, which correspond to the leaf units of the abstract architecture of Figure 1, are connected by an H-tree interconnection. The fan-in units exist at intermediate nodes of the tree, as in the abstract model. The resulting output vector element is transmitted via an optical modulator.

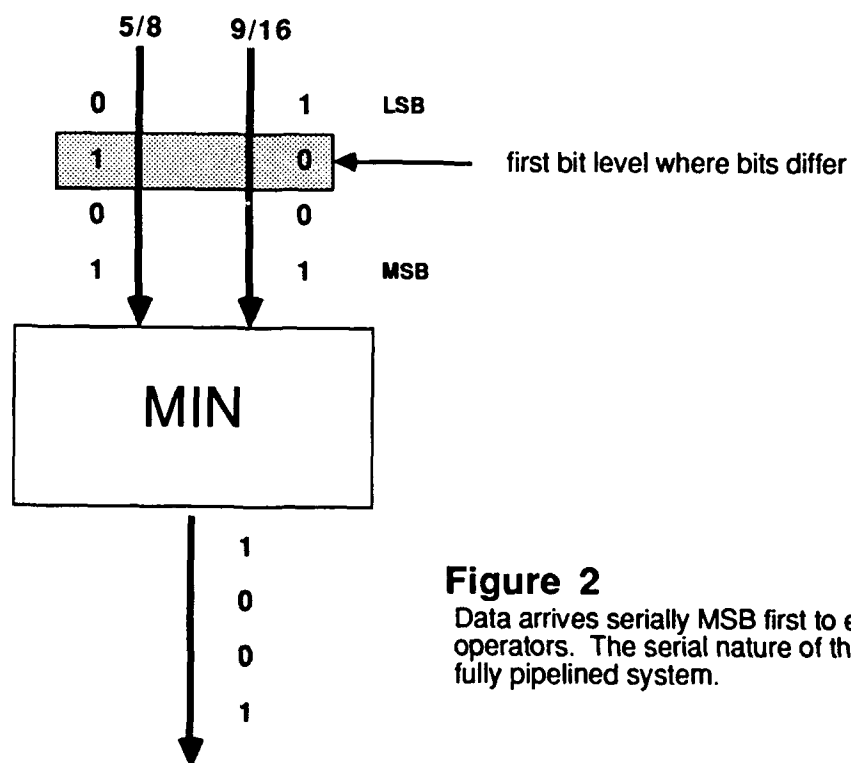
The MAX and MIN fuzzy operators in this system are implemented using bit serial comparators. A gate level description of the comparator used to realize the MAX operation is depicted in Figure 4. It is easily adapted to perform the MIN operation by inverting the inputs. The bit serial comparator is compatible with the serial arrival of the data, and therefore additional latches and control circuitry are not needed. It is smaller in size than most comparators and in particular to a parallel comparator. This combined with the regularity of the individual processing elements makes the system well suited for VLSI implementation.

REFERENCES:

- [1] G.J. Klir and T.A. Folger, Fuzzy Sets, Uncertainty, and Information, Prentice Hall, 1988
- [2] H.J. Zimmermann, Fuzzy Set Theory - And Its Applications, Kluwer-Nijhoff, 1985
- [3] S. Krishnakumar, private communication,
- [4] H.J. Caulfield, "Optical Inference Machines," *Optics Comm.* **55** (1985) pp.259-260
- [5] G.C. Marsden, A. Krishnamoorthy, S. Esener, and S.H. Lee, "Dual-Scale Topology Optoelectronic Processor (D-STOP)," OSA 1990 Annual Meeting, Boston (1990)

**Figure 1**

Functional diagram of the processing element used to compute the i th element of the generalized matrix-vector product. In the first stage of the tree the MIN operations are performed with the locally stored matrix elements. The following stages collectively perform the global MAX.

**Figure 2**

Data arrives serially MSB first to each of the bit-serial fuzzy operators. The serial nature of the logic units allows for a fully pipelined system.

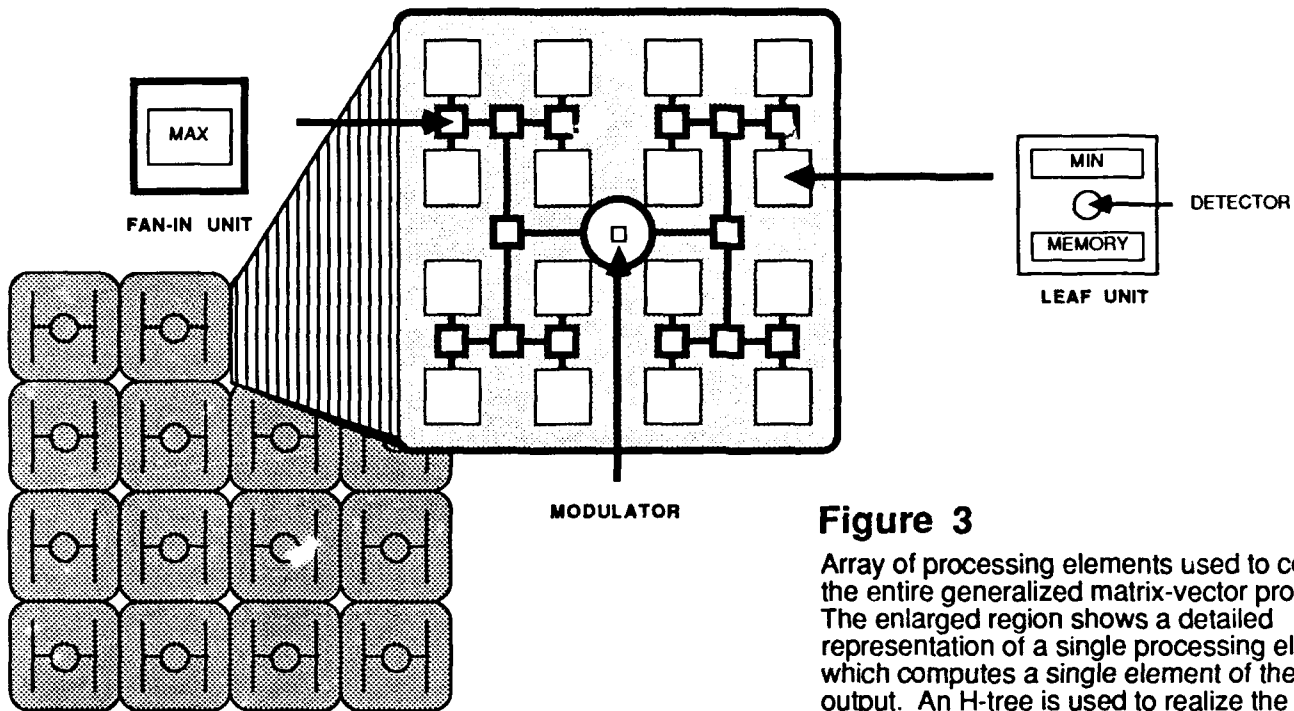


Figure 3

Array of processing elements used to compute the entire generalized matrix-vector product. The enlarged region shows a detailed representation of a single processing element which computes a single element of the vector output. An H-tree is used to realize the binary tree depicted in Figure 1 in an area efficient manner.

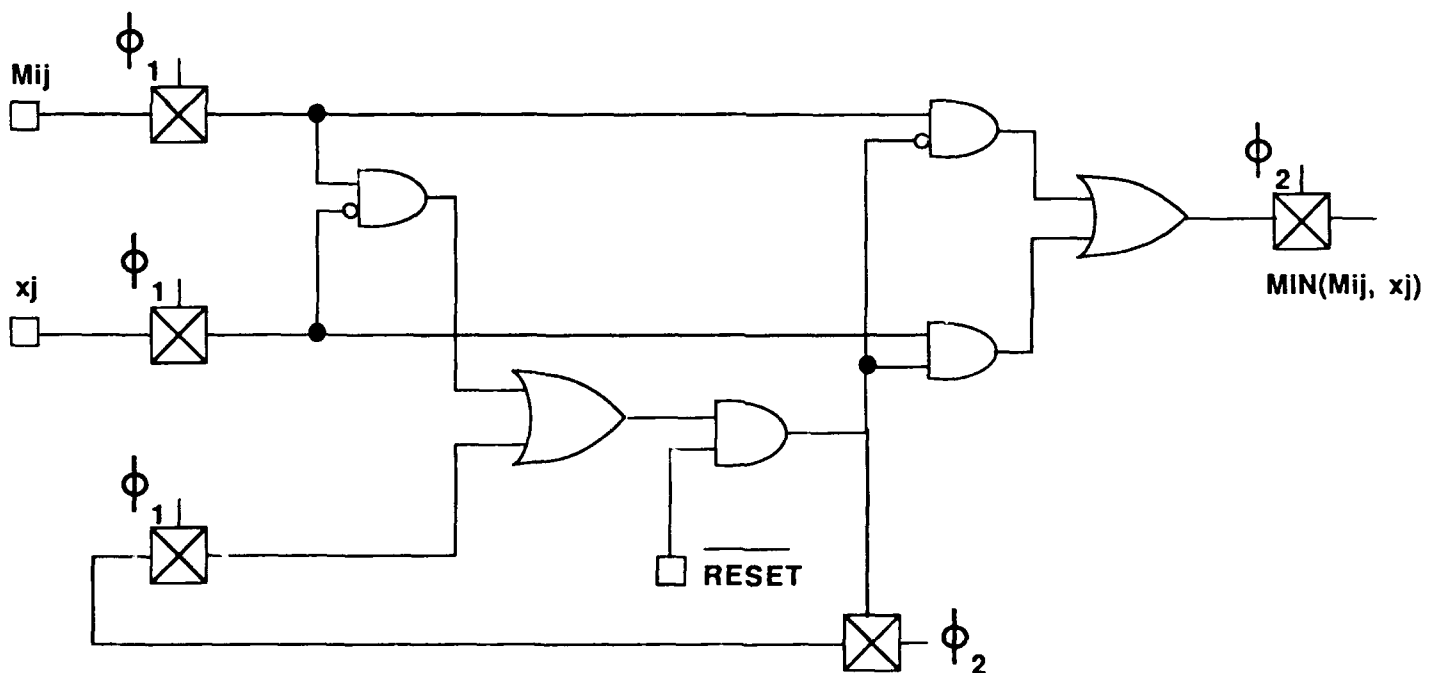


Figure 4

Logic diagram of the bit-serial comparator used to perform the MIN operation. A MAX operator can be achieved by inverting the inputs. The simplicity and compact size of the circuitry in conjunction with the regularity of the processing elements makes the system well suited for VLSI implementation.

Optical Morphological Image Processing with Acoustooptic Devices

Ravindra A. Athale
ECE Department
George Mason University
Fairfax, VA. 22030

Joseph N. Mait and Dennis W. Prather
US Army Harry Diamond Labs
Adelphi, MD.

Background

Morphological image processing based on binary set representation of an image has been receiving increasing attention as a viable alternative to the linear image processing based on Fourier domain filtering [1]. The fundamental morphological filtering operations of erosion and dilation are nonlinear operations from which more complex operations (opening, closing, pattern spectra) suitable for shape extraction and analysis can be synthesized. The morphological operations are defined between a working image and a much smaller image called the structuring element. In most cases the structuring element is binary while the working image can be binary, analog with binary threshold decomposition representation or analog with weighted binary representation. For the first two representations, the dilation and erosion operations consist of a superposition of several shifted replicas of the working image followed by a point-wise threshold operation at different levels; 1 for dilation and $(m-1)$ for erosion, where "m" is the number of images superposed. With the third representation, the minimum/maximum value (for dilation/erosion, respectively) for the superposed pixels is detected and assigned as the pixel value in the output image. In either case, the processing operations are of low complexity. The number of shifted replicas of the working image and the amount of shift corresponds to the number and position, respectively, of bright pixels in the structuring element. For large and irregularly shaped structuring elements the morphological processing operation complexity is dominated by data communication.

The parallelism and connectivity offered by analog optical systems have long been utilized to perform linear Fourier domain filtering operations on images. Recently modification of optical correlator architectures by incorporating a point nonlinearity on the output image has been proposed for implementing morphological operations [2-5]. The systems described in references 2 and 3 utilize the point spread function of a defocused

imaging system to implement a circularly symmetric structuring element. The system proposed in reference 4 utilizes a complex, holographically-recorded Fourier plane filter to realize the structuring element. The processor demonstrated in reference 6 is based on shadow casting in which a 2-D array of individually modulable light sources implement the structuring element. The first approach does not have the flexibility to realize arbitrary structuring elements. The lack of suitable real-time holographic recording material limits the programmability of the second approach. The third approach is based on geometrical optics formulation of shadow casting and hence will be limited to small size images and structuring elements.

Acoustooptic Fourier-plane Filtering System

An acoustooptic (AO) device driven by a programmable arbitrary waveform generator has been proposed and demonstrated for spot array generation [6]. In the optical morphological processor presented in this paper, the AO device is placed in the Fourier plane of a coherent optical correlator (Figure 1). The correlator output will contain a superposition of multiple shifted replicas of the input image. The input function to the AO device is an RF carrier modulated by another function capable of producing the desired array of spots. Techniques for designing kinoform array generators [8] were used to determine a Fourier function capable of generating a high diffraction efficiency spot array. Figure 2 shows the results of a simple proof-of-principle experiment performed with a 1-D AO device in the Fourier plane. The thresholding of the output image was performed by an electronic post processor. The property of erosion operation to eliminate features narrower than the structuring element can be clearly seen. A simple change of the subcarrier frequency can affect the scale of the structuring elements without changing its form, a property useful in calculating pattern spectra. A 2-D structuring element can be implemented time-sequentially by placing two orthogonally oriented AO devices in the Fourier plane. Such devices are commercially available in a single crystal configuration with transducers mounted on two orthogonal faces of the crystal [7]. Driving both transducers simultaneously generates a point-spread-function that is given by an outer product between the point-spread-functions of signals driving the individual transducers. An arbitrary structuring element can then be synthesized as a superposition of multiple outer products. Since the frame time of the AO device is on the order of microseconds, the morphological operations can be implemented at tens of kHz frame rate, which will be ultimately limited by the frame rate of the 2-D detector array with on-chip processing. The space-bandwidth product of the working image will be limited by the Bragg conditions for different Fourier components of the image as well as different subcarrier frequencies in the

AO cell for generating the structuring element. The results of analysis identifying the trade-offs between the image and structuring element size will be reported. The potential for improving the system performance via special transducer design will also be explored.

Summary

Morphological operations on images can be performed by combining optical correlators with simple electronic nonlinear operations on the output image. The simple binary nature of the structuring elements suggests that an acoustoptic device driven by an arbitrary waveform generator can generate the desired Fourier plane filters with easy and rapid programmability. Results of a simple proof-of-concept experiment are presented. The analysis of performance limits of the acoustoptic morphological image processor due to Bragg matching limitations will be presented.

Acknowledgement

This research was supported by Army Research Office under Summer Faculty Research in Engineering Program is gratefully acknowledged. Several useful discussions with Dr. John Pellegrino are gratefully acknowledged.

References

1. P. Maragos, "A Unified Theory of Translation-invariant Systems with Applications to Morphological Analysis and Coding of Images", Ph. D. Thesis, Georgia Institute of Technology, Atlanta, GA., 1985.
2. K. S. O'Neil and W. T. Rhodes, "Morphological Transformation by Hybrid Optical-electronic Methods", Proceedings of SPIE, Vol. 638, pp. 41-44, 1986.
3. J. M. Hereford and W. T. Rhodes, "Nonlinear Optical Image filtering by Time-sequential Threshold Decomposition", Opt. Eng., Vol. 27, pp. 274-279.
4. E. C. Botha and D. Casasent, "Applications of Optical Morphological Transformations", Opt. Eng., **28**, pp. 501-505, 1989.
5. Y. Li, A. Kostrzewski, D. H. Kim and G. Eichmann, "Compact, Parallel, Real-time Programmable Optical Morphological Image Processor", Opt. Lett., Vol. 14, pp. 981-983, 1989.
6. E. Tervonen, J. Turunen and A. Friberg, "Synthetic Acoustoptic Holograms", Optics in Complex Systems, Proc. SPIE, Vol. 1319, p. 288 (1990).
7. H. P. Herzig, D. Prongue and R. Dandliker, "Design and fabrication of highly efficient fan-out elements", Submitted to Japanese Journal of Applied Physics, 1990.
8. Available from Brimrose Corp., Baltimore MD.

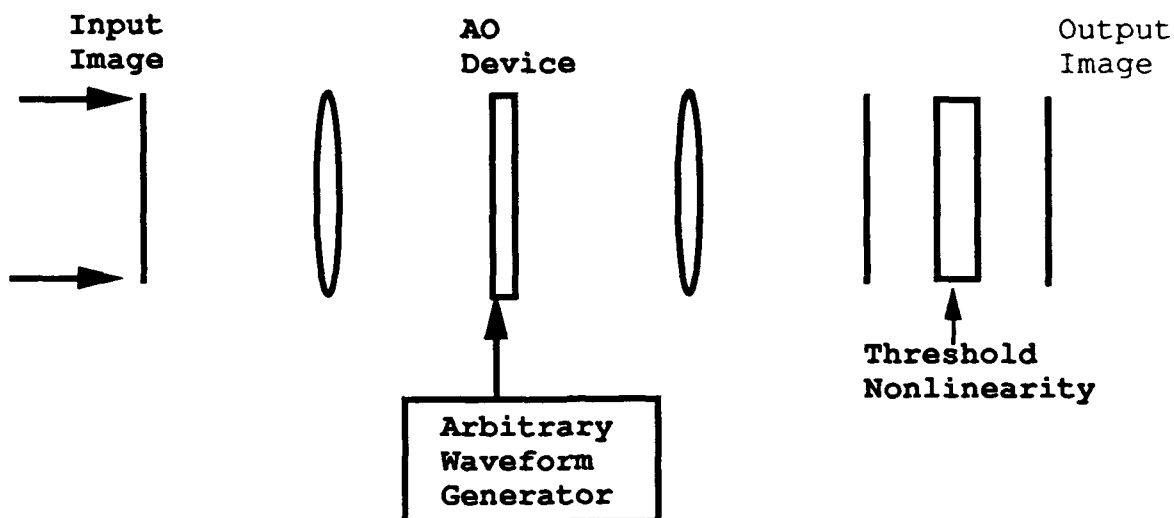


Figure 1: Schematic diagram of an AO-based morphological processor

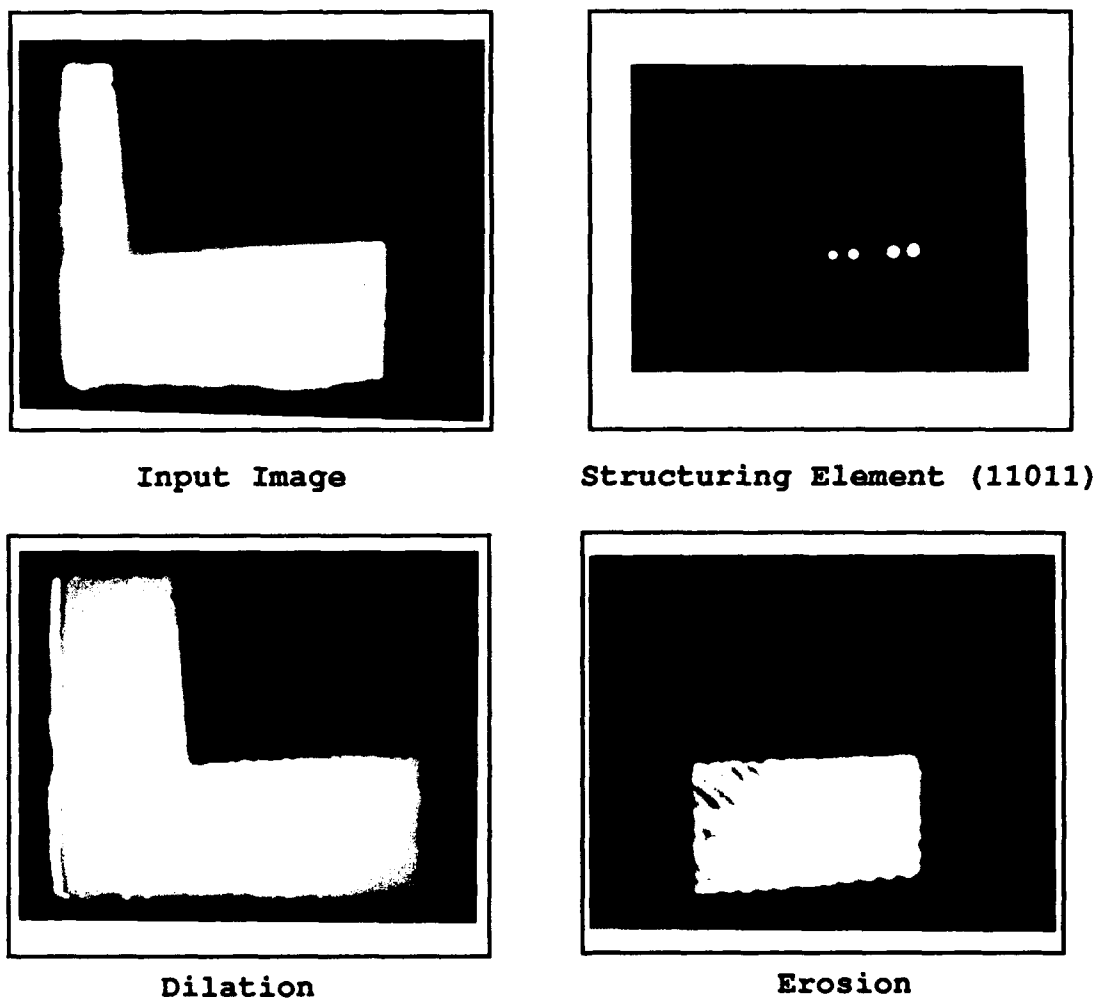


Figure 2: Preliminary Experimental results

Optical Morphological Image Processor

Gary E. Lohman, K.-H. Brenner

Angewandte Optik, Physikalisches Institut

der Universität Erlangen-Nürnberg

D-8520 Erlangen, Federal Republic of Germany

1. Introduction

Images in the natural sciences often possess distinctive topologies, thus rendering order statistics better suited for image processing than more traditional linear filtering. A useful subclass of order statistics based on binary images is mathematical morphology.^{/1/} Mathematical morphology is also well suited to an optical implementation.^{/2-5/} Optical mathematical morphology can be performed at a frame rate of 10-100 kHz., thus permitting real-time non-linear image processing in many applications. Our proposed optical architecture also allows for programmable parallel processing of very large images, under control of a small electronic micro-processor.

2. Mathematical Morphology

Mathematical morphology is based on the notions of image dilation and erosion. A useful notation here is borrowed from the Minkowski set operations of addition and subtraction, Fig. 1. Given an image A and a structuring kernel B , the sum $A \oplus B$ is found by the superposition of shifted copies of A as prescribed by B . This can also be imagined as the extent of a binary convolution. The dual to this is the Minkowski subtraction, $A \ominus B$.

The combination of an erosion with a dilation by means of an image difference is commonly referred to as the Hit-Or-Miss (HMT) transformation.

$$\text{HMT of } A \Rightarrow (A \oplus B^s) - (A \ominus C^s)$$

Where B^s is the set symmetric to B . Besides being by definition a symbolic substitution with B and C forming the recognition and replacement patterns, the HMT forms the basis of all morphological transformations.

Minkowski Addition: set union of shifted images



Minkowski Subtraction: set intersection of shifted images



Fig. 1. Minkowski addition and subtraction.

3. Optical Implementation

The optical implementation follows directly from the Minkowski addition and subtraction. Here we see that a superposition of shifted copies of an image must be formed. This must then be thresholded, and for constructing the dual to any morphological transformation, complementation of the image should also be possible. The first step is achieved using passive optics, Fig. 2a. The system in Fig. 2 differs from traditional symbolic substitution systems in that the rule is not fixed but is rather programmed by the activation of LEDs in an array. In digital processing, a few number of fixed rules are sufficient. Image processing on the other hand requires greater programmability. Fig. 2 also shows an experimental result from the optical dilation unit. The second step in the process, namely thresholding, latching and complementation is satisfied by a bistable optically addressable ferro-electric liquid crystal (BOAFLC). Such devices operating at approximately 10kHz. are presently available. Including these along with a feedback loop for iterative processing and input/output latches yields the basis for

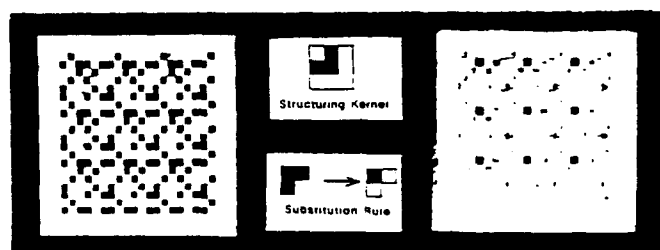
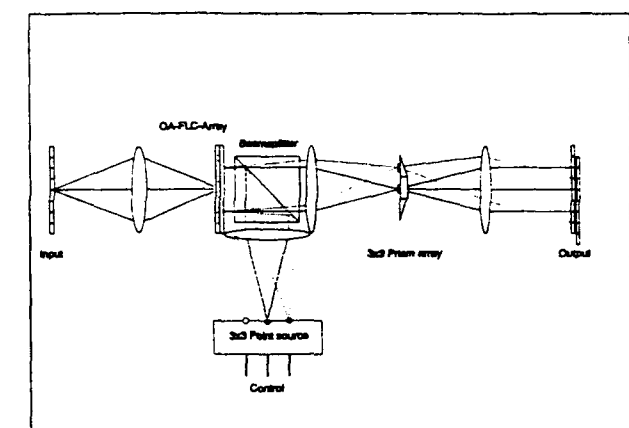


Fig. 2. Optical dilation via convolution and threshold.

a very powerful non-linear all optical processor. The unit under construction is depicted in Fig. 3a. Fig. 3b shows the more general processor architecture. Notice that not only is the processing performed in parallel, but via the BOAFLCs even the input/output addressing occurs in parallel. Furthermore, the programming is achieved via a small electronic micro-processor controlling the LED and BOAFLC addressing, thus rendering the actual optics transparent to the programmer. Aside from I/O and processing speeds well above video frame rates, very large images can be processed with this architecture. Denoting the cardinality of the image and the structuring kernel by $|A|$ and $|B|$ respectively, we obtain

$$|A||B| \leq SW.$$

Where SW is the space-bandwidth product permitted by the optics. Since $|B|$ is typically small, e.g. 3×3 in our case, $|A|$ can be very large. Furthermore since the computational power is determined by the number of iterations and not the filter complexity, a small $|B|$ is preferable for greater flexibility without loss of generality. Expandability of the optics is readily achieved by cascading multiple systems such as that in Fig. 3 together with possibly global permutation networks.

4. Application

The present application involves front-end processing for an industrial vision system, Fig. 4. Noise reduction, edge enhancement and object isolation improve detection reliability. Morphological skeletonization is useful for enhancing differences in similar objects. Direction dependent edge enhancement can improve vision quality of moving objects. Fig. 5 shows successive results of four morphological transformations on an image along with the optical Fourier transforms of the initial and processed images.

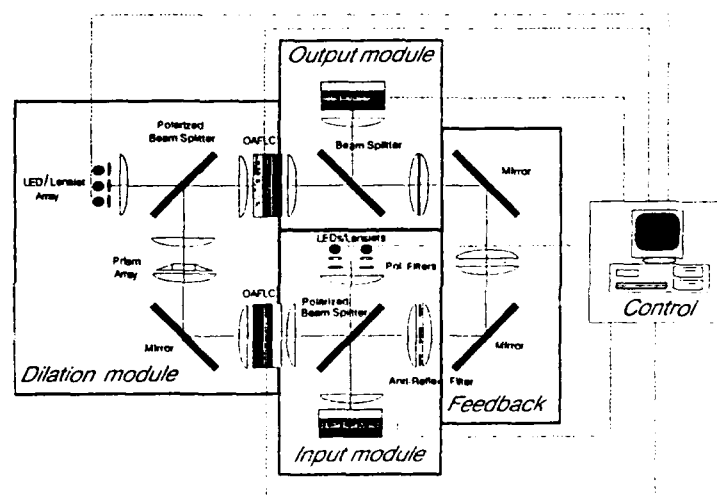


Fig. 3a. Optical morphological demonstration unit.

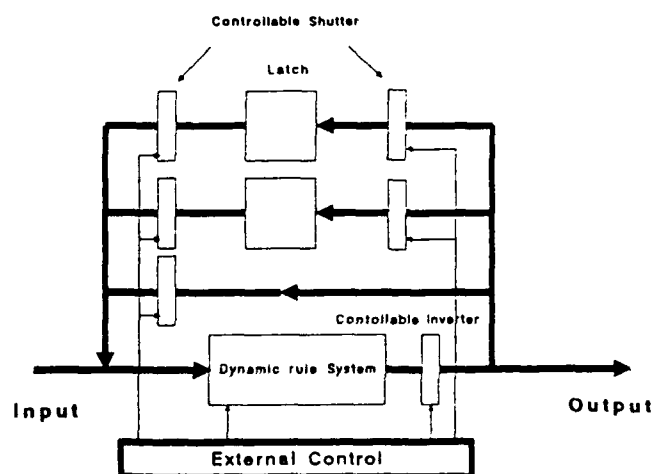


Fig. 3b. Generalized optical morphological processing system

5. Conclusion

A very fast and programmable optical image processing unit capable of many real-time vision tasks is under construction. Its expandability and programming versatility attest also to its future potential.

6. Literature

1. J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, New York/London, 1982.
2. K.-H. Brenner, G. Stucke, "Architectures for Digital Optical Image Processing using Morphological Filters," *Optical Computing '88, Toulon*, SPIE vol. 963, p. 657, 1988.
3. D. Casasent, E. Botha, "Optical symbolic substitution for morphological transformations", *Appl. Opt.*, 27, 3806 (1988).
4. G. Lohman, K.-H. Brenner, "Morphological optical image processor," *Optics in Complex Systems, ICO-15*, SPIE vol. 1319, p. 161, 1990.
5. G. Pedrini, *Discrete incoherent optical correlator using diffraction gratings*, doctoral thesis, Université de Neuchâtel

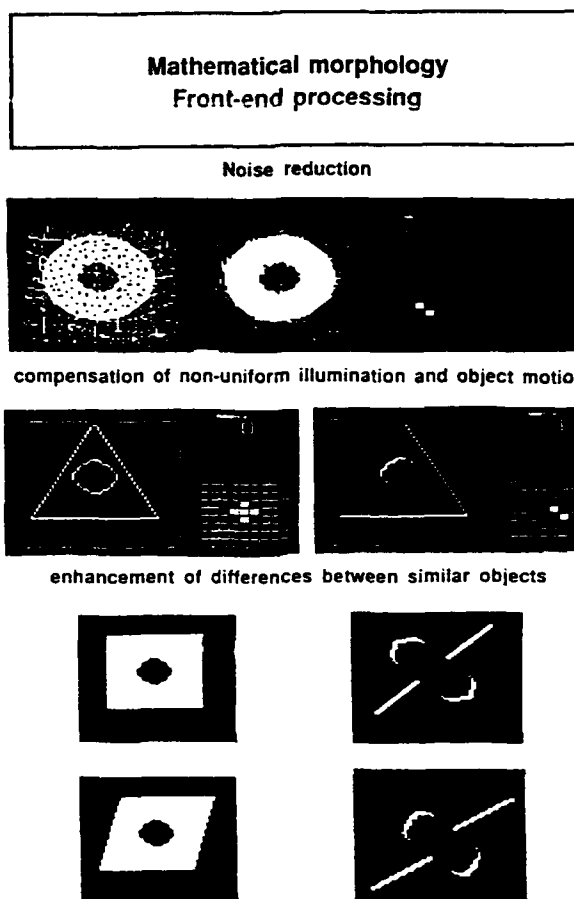


Fig. 4. Morphological pre-processing.

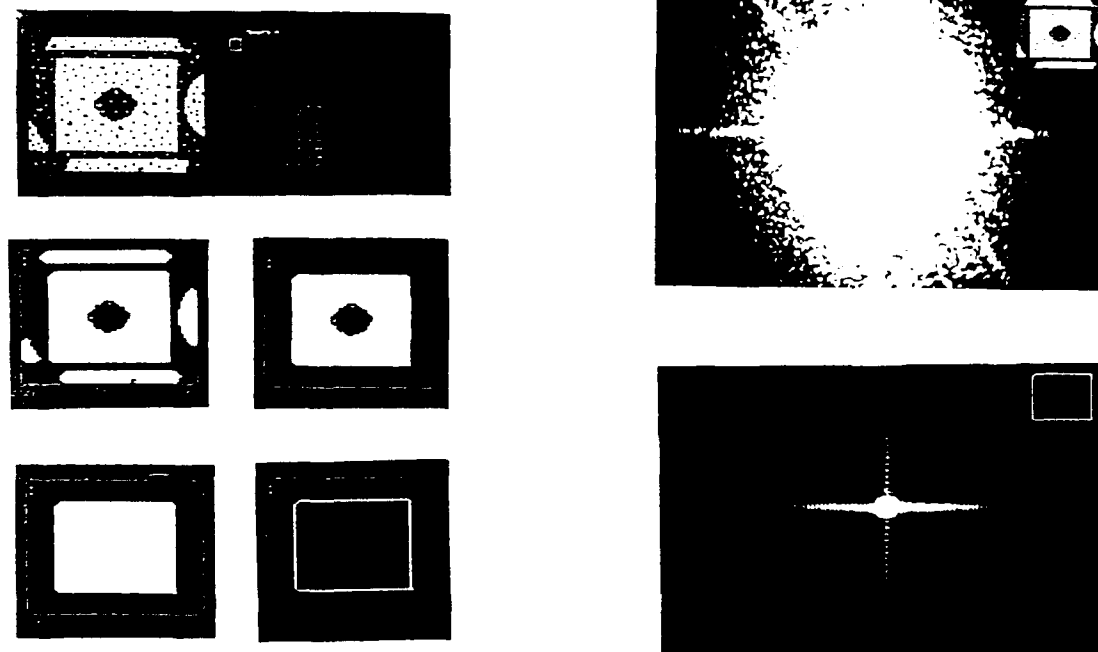


Fig. 5. Morphological processing example and Fourier images of initial and processed image.

Cellular Automata via Multi-Kernel Incoherent Holographic Convolution

I. Glaser

Tel Aviv University

Department of Electrical Engineering — Systems

Tel Aviv, Israel.

E-Mail: `glas@taunivm.bitnet`

1 Introduction

Cellular machines [1,2,3,4] are two dimensional arrays of rather primitive processing elements (PE-s) that are controlled by the internal state of their neighborhood PE-s and (usually) by some external instruction stream in a SIMD style. Such machines offer high computational speed for a useful class of problems, particularly in *image processing*. While some cellular machines were implemented in silicon, newer versions with enhanced connectivity, such as the cellular hypercube [5], are probably too complex for electronics. Furthermore, the 'almost-shift-invariant' geometry of cellular machine interconnections is a good match to optical convolution, which is easy to carried out.

There were several suggestions on implementing cellular automata optically [6]. Most notable are variations on *optical gate arrays*, or *sequential optical logic* [8,9,6,7] and the *convolve-and-point-transform* method [10]. Both approaches rely on an optical linear transformation and some (possibly electro-optical or electronic) point-non-linear transformation. The difficulty of each method can be expressed by two numbers: the number of optical resolution elements (ORE) that must be allocated to each cellular processing element (PE), $\nu \equiv \text{ORE-s/PE}$ and the number of levels, ℓ , that the point-non-linear transformation must discriminate. In terms of the optics of the system, the first number, ν , is related to the space-bandwidth-product of the (linear) optics while ℓ gives the required intensity precision and uniformity of both the optics and the point-non-linear sub-system. In other words, ν limits the number of PE-s a given combination of approach and technology can attain, while ℓ tells us if that combination can work at all.

Optical gate array require one optical resolution element per gate; for example, DOCIP of ref. [8]) uses $\nu \approx 50$. However, the number of intensity levels that the point-non-linear part of this system has to deal with there is quite small ($\ell \approx 4$ for DOCIP [8]). The convolve-and-point-transform method uses $\nu \approx 1$ ($\nu = 1$ for one bit per pixel machine [10]) but $\nu = O(10^2)$. This very large ν makes optical convolve-and-point-transform machines unlikely, because they are implementable only for machines with very small neighborhood, where electronics is adequate.

In this work we discuss an alternative method for implementing optical cellular machines. This method, like the Gerritsen [10] convolve-and-point-transform one, uses convolution followed by point-non-linearity. However, control (SIMD machine instruction) is given through both the convolution kernel and the point-non-linearity function. It requires an optical system that offers dynamically reconfigurable convolution kernel, but only modest ℓ or intensity precision.

2 The Proposed method

The proposed method (a brief description of which was presented in Ref. [11]) is a modification of the convolve-and-point-transform approach. In our method, specific convolution kernel is designed for each machine instruction, or small group of machine instruction. Since each instruction needs to access only a subset of the maximum neighborhood of the machine, its convolution kernel will be far simpler than that of the Gerritsen method. Even when several instructions share the same kernel (they are distinguished through use of different point-non-linearities) the grouping can be done in a way that will keep each kernel simple enough. We note that, in a sense, the Gerritsen method is an 'overkill': it allows for the full set of 2^μ operations (μ is the number of cells in the maximum neighborhood) though a much smaller subset of instructions suffices.

2.1 Multi-Kernel Incoherent Holographic OTF Synthesis

Incoherent holographic OTF synthesis is a technique where holograms, recorded with diffused coherent light, are used to obtain convolutions and/or correlations of input spatial signals with spatially non-coherent quasi-monochromatic illumination. With proper attention to optical design issues [12], this technique offers performance similar to that of coherent spatial filtering, but without the later's sensitivity to mechanical errors, minor optical aberrations and dust, or its need for optically flat SLMs. In this section we describe a method for producing multi-kernel convolutions with this approach.

Our approach uses *area segmentation* of the hologram plane, as shown in Fig. 1, which sits at the pupil of the optical system. The total area (which, for simplicity, is assumed here to be square) is divided into N^2 sub-pupils. We can put M holograms into this area by assigning N^2/M sub-pupils to each. If the selection is done at random, it can be shown that (for large N^2/M) the point spread function (PSF) of each hologram would be the PSF of the data recorded on it, convolved, in the average, with function of the form

$$h(x, y) \approx \frac{1}{\lambda^2 F^2} \left[\frac{1}{N^2} \text{sinc}^2 \left(\frac{x}{N\lambda F}, \frac{y}{N\lambda F} \right) + \frac{1}{M} \text{sinc}^2 \left(\frac{x}{\lambda F}, \frac{y}{\lambda F} \right) \right]$$

where λ is the average wavelength, F is the focal length of the system, (x, y) are coordinates at the output plane and $\text{sinc}(x, y) \equiv [\sin(\pi x) \sin(\pi y) / (\pi^2 xy)]$.

The first term above is the PSF of a single sub-pupil and the second term is the PSF of the entire aperture: it is larger than the first term by a factor of N^2/M , the number of sub-pupils allocated to each hologram. Thus, if $N^2/M > 2\ell$ the hologram would be as good as one made on the entire aperture. Fig. 2 shows $h(x, y)$ (in logarithmic scale) for $N = 64$ and $M = 16$. For the case of this graph $N^2/M = 256 \approx 10^{2.41}$.

To implement this approach we record the holograms sequentially, each with a proper mask to select its sub-pupils; no sub-pupil is used for more than one hologram. Because the holograms do not overlap, no compensation for reciprocity or memory effects in the recording material is necessary [13]. Once all holograms are recorded, the plate is sandwiched to some electro-optical shutter (an SLM) which selects the sub-pupils that will receive light. We note that we can use a binary (no gray scale) SLM even when the desired convolution kernels are not. At present, binary SLMs such as those using FLC are much faster than continuous scale ones.

Of course, the actual selection of the sub-pupils is *quasi-random* at best; a practical value for N^2 will probably not be large enough to satisfy the statistical assumptions behind the equation for $h(x, y)$, above, anyhow. Also, we are more interested with actual individual PSFs than with a statistical average one. It is possible, nevertheless, to get help from a chiseled dice. With an iterative algorithm it is possible to assign sub-pupils so that the actual PSF will be closest to the statistical average of the equation above. The outline of one such algorithm is shown in Fig. 3.

2.2 Selection of Kernels and Non-Linear Functions

As noted above, the approach presented here is a modification of Gerritsen's [10] convolve-and-point-transform method. In the Gerritsen method we start defining a maximum neighborhood, which is a union of all neighbor-

hoods used by all the instructions in our set. Assuming that there are μ cells in this maximum neighborhood, we define a convolution kernel where the i -th cell gets a weight of 2^i ($i = 1 \dots \mu - 1$). Thus any possible instruction can be defined as a point-non-linearity which takes numbers in the range $[0 \dots 2^\mu - 1]$ and convert them to a binary digit of 1 or 0. The use of $\ell = 2^\mu$ is dictated from the desire to cover all possible instructions, including non-commutative ones, with just one convolution kernel.

In practice, no instruction may need to address the *maximum* neighborhood. Most instructions will address only a subset, and they would be commutative over their subset. If μ_{subset} is the number of cells in the largest neighborhood addressed by a *single* instruction, and that instruction is commutative, we can get by with $\ell = \mu_{\text{subset}}$ if that instruction has its own convolution kernel.

3 Concluding Remarks

We have shown how, by using multi-kernel convolver, we can significantly simplify the optical implementation of cellular machines, and presented a method of obtaining multi-kernel convolutions with incoherent holographic OTF synthesis.

References

- [1] E. F. Cobb: *Cellular Automata* (Academic Press, New York, 1968).
- [2] S. Wolfram: *Theory and Application of Cellular Automata* (World Scientific, Singapore, 1986).
- [3] T. Toffoli and N. Margolus: *Cellular Automata Machines* (MIT Press, Cambridge Massachusetts, 1987).
- [4] J. Serra: *Image Analysis and Mathematical Morphology* (Academic Press, New York, 1982).
- [5] K. S. Huang, B. K. Jenkins and A. A. Sawchuk: "A Cellular HyperCube Architecture for Image Processing," *Applications of Digital Image Processing X, Proc. SPIE* **829**, 331-338 (1987).
- [6] J. Taboury, J. M. Wang, P. Chavel, F. Devos, P. Garda: "Optical Cellular Processing Architecture, 1: Principles," *Applied Optics* **27**, 1643-1650 (1988).
- [7] J. Taboury, J. M. Wang, P. Chavel, F. Devos, P. Garda: "Optical Cellular Processing Architecture, 2: Illustration and System Considerations," *Applied Optics* **28**, 3138-3147 (1989).
- [8] K. S. Huang, A. A. Sawchuk, B. K. Jenkins, P. Chavel, J. M. Wang, A. G. Weber, C. H. Wang and I. Glaser: "Implementation of a prototype Digital Optical Cellular Image Processor (DOCIP)," *Optical Computing '88 conference (IOCC)*, Toulon, France, August 29 to September 2, 1988. Summary in: *SPIE* **963**, 687-694 (1988).
- [9] K. S. Huang: *A Digital Optical Cellular Image Processor (DOCIP): Theory, Architecture and Implementation*, USC-SIPI Report No. 133 (University of Southern California, Los Angeles, 1988).
- [10] F. A. Gerritsen and P. W. Verber: "Implementation of Cellular-Logic Operations using 3×3 Convolution and Table Look-Up Hardware," *Computer Vision, Graphics and Image Processing* **27**, 115-123 (1984).
- [11] I. Glaser: "Dynamically Reconfigurable Interconnections for Digital and Neural Computing," French-Israeli Topical Meeting on Optical Computing, Paris, June 6-7 1990; summary in: Groupe d'Optoinformatique (Societe Française d'Optique) Bulletin N° 11, 21-36 (September 1990).
- [12] I. Glaser: "Holographic Incoherent Optical Transfer Function Synthesis: Analysis and Optimization," *Journal of the Optical Society of America A* **3**, 681-697 (1986).
- [13] K. M. Johnson: "Multiple Multiple-Exposure Hologram," *Applied Optics* **24**, 4467-4472 (1985).

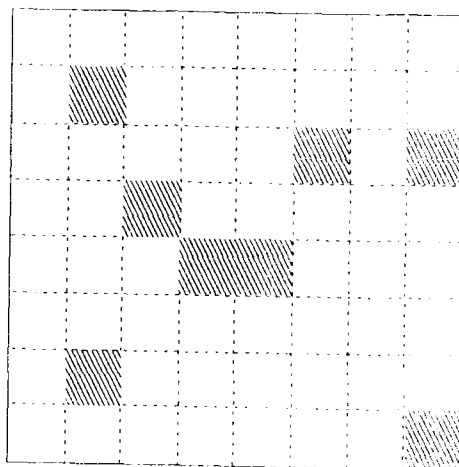


Figure 1. Segmented pupil plane with 8 sub-pupils selected.

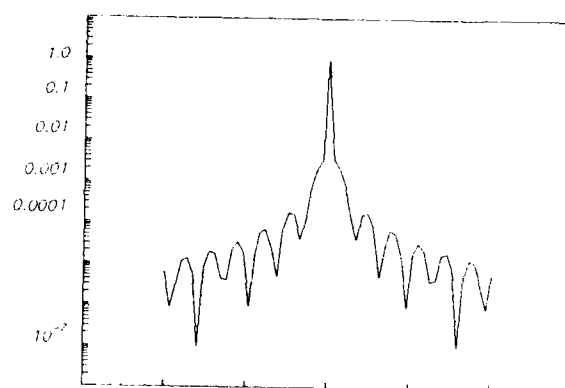


Figure 2. Typical $h(x,y)$, for $y=0$, h in logarithmic scale, for $N=64$, $M=16$. This is the statistical average PSF

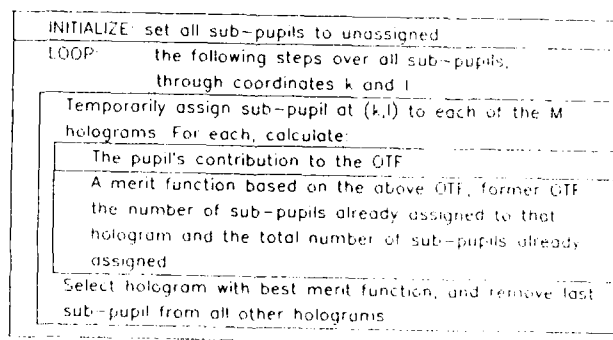


Figure 3. Sub-pupil assignment algorithm

Tuesday, March 5, 1991

Memory Issues

TuC 2:00pm–3:10pm
Salon F

Cardinal Warde, *Presider*
Massachusetts Institute of Technology



OPTICAL DATABASE MACHINES

P. Bruce Berra¹

Databases have become an important aspect of our daily lives. We encounter them in such diverse fields as airline reservations, stock quotation systems, medical information systems, entertainment, sports and a host of other areas. Database management systems (DBMS) place considerable demands on current computing systems primarily because of the large size of the databases, the general functionality of the DBMS and the stringent time requirements for the retrieval of the data. The large size of the database dictates that secondary storage such as optical and magnetic disks be used and this leads to input/output data accessing difficulties since these memory types are on the order of one million times slower in access time than main memory technology. The diverse functionality of DBMS leads to systems with millions of lines of code that consume an enormous number of machine cycles. Add to this the near real time requirement for many applications and one has a system that is both I/O and compute bound.

The tremendous advances in electronic technology have contributed immeasurably to the advancement of database systems but each new advance has been met with greater application requirements. One of the approaches that has been taken to improve the performance of these systems is that of developing database machines [Su88]. This approach takes advantage of the technology while utilizing the principles of parallelism, pipelining, decomposition and caching. Since much of the improvement in database systems has come from advances in electronic technology, it is important to consider other technologies. And optics with its inherent speed, high bandwidth and natural parallelism offers some interesting possibilities.

¹ Department of Electrical & Computer Engineering, 121 Link Hall, Syracuse University, Syracuse, NY 13244-1240.

Optics can impact database management in storage, interconnection/communication and processing. At least three possible interfaces between optics and electronics can be envisioned. First, we can imagine data in optical form being received from modified optical disks [Psa89] or holographic memory at rates two to three times higher than current magnetic disks. The data could be converted from photons to electrons but these data rates would be too high for most electronic computers unless suitable modifications were made. At the next level we can envision distributing data, in optical form, to several sites for conversion and subsequent use. A third level involves performing processing functions prior to conversion. This approach seems most appropriate since the data will have been reduced considerably and the data rate to the electronic computer would be more manageable and considerably richer in content.

In relational database management, a number of operations such as union, difference, intersection, cartesian product, selection, projection and join are typically performed. We have developed optical architectures for the execution of these operations and have evaluated their performance [Ber87, Ber89, Mit90]. Our results indicate that there is potential for significant performance improvement if suitable hardware devices were available. Some of our other work includes optical content addressable memories [Ber88] and data/knowledge base machines [Ber90a].

Full text search differs from database management in that at some level the entire database of documents must be searched word for word. The types of operations performed include counting the number of occurrences of a word in a document; finding a particular word in a document, page, paragraph or sentence; finding multiple different words in the same document, page, paragraph or sentence and finding words in proximity to each other. In addition, various operations need to be performed on character strings such as prefix, suffix and embedded don't care both for fixed and variable length search arguments [Mit89]. We feel that this application has even greater potential for performance gains than database

management.

Finally, an area that will have far reaching effects on database systems is multimedia. In the future DBMS will have to manage vast amounts of structured data, text, images, audio and video. These data types will place considerable constraints on the performance of these systems and thus optical processing may offer some interesting possibilities for improvement [Ber90b]. Our current research thrust is in this direction.

REFERENCES

- [Ber87] P.B. Berra, N.B. Troullos, "Optical Techniques and Data/Knowledge Base Machines," IEEE Computer, pp. 59-70, October 1987.
- [Ber88] P.B. Berra, S. Marcinkowski, "Optical Content Addressable Memories for Managing an Index to a Very Large Data/Knowledge Base," IEEE-CS Data Engineering Bulletin, 1988.
- [Ber89] P.B. Berra, A. Ghafoor, P. Mitkas, S. Marcinkowski, and M. Guizani, "The Impact of Optics on Data and Knowledge Base Systems," IEEE Transactions on Knowledge and Data Engineering, Vol. 1, No. 1, March 1989.
- [Ber90a] P.B. Berra, K.H. Brenner, W.T. Cathey, H.J. Caulfield, S.H. Lee, and H. Szu, "Optical Database/Knowledge Base Machines," Applied Optics, Vol. 29, January 1990.
- [Ber90b] P.B. Berra, C.Y.R. Chen, A. Ghafoor, C.C. Liu, T.D.C. Little, and D.H. Shin, "An Architecture for Distributed Multimedia Database Systems," Computer Communications, Vol 13, No. 4, pp. 217-231, May 1990.
- [Mit89] P. Mitkas, P.B. Berra, and P. Guilfoyle, "An Optical System for Full Text Search," 12th Annual Conference on Research in Information Retrieval, ACM, SIGIR, June 1989.
- [Mit90] P. Mitkas, "On Relational Data Base Operations Implemented in Optics," Ph.D. dissertation, Syracuse University, 1990.
- [Psa89] D. Psaltis, A.A. Yamamura, M.A. Neifeld, and S. Kobayashi, "Parallel Readout of Optical Disks," in Proc. OSA Top. Meet. Opt. Comput., Vol. 9, Salt Lake City, UT, pp. 58-61, February 1989.
- [Su88] S.Y.W. Su, Database Computers Principles, Architecture, and Techniques, McGraw-Hill, 1988.

Optical Implementation of SELECTION Operation in Database Machines

Ravindra A. Athale and Michael W. Haney*

Electrical and Computer Engineering Department
George Mason University
Fairfax, VA 22030

*BDM International, Inc.
7915 Jones Branch Drive
McLean, VA 22102

Introduction

In its most elemental form, a database can be viewed as a computer-based record keeping system. The database organization is optimized for efficient addition, deletion and updating of the records. Accurate, flexible and efficient techniques for retrieving and organizing data is the objective of a database management system. Applications of such database management systems range from banking and libraries in the commercial world to on-board electronic warfare systems for airplanes and logistics databases for weapons readiness management in the military world. In either domain, the size of the database is constantly growing, while the desired data retrieval time is simultaneously decreasing. In addition, on-board systems may have volume, power and weight limitations while maintaining ruggedness. The recent developments in optical storage, interconnects and switching technologies has initiated investigation into the use of optical technology to enhance the performance of conventional database machines [1].

The primary motivation for using optical storage is to exploit the 3-D interconnects provided by free space optical systems to increase the data transfer rate between secondary storage and the host processor/main memory of a database machine. Parallel readout techniques can be employed with optical disks as well as with holographic optical storage to retrieve 10's kbits simultaneously, effectively providing a data transfer rate of Gbits/sec for a frame transfer rate of 100 kHz [2,3]. In addition, volume holographic optical storage has the potential of providing microsecond random access time to any one of the frames (vs 10 milisecond for most disk media, optical or magnetic) [3]. This high throughput rate can cause a bottleneck at the host processor interface since a typical electronic host processor is capable of accepting data at 10-100 Mbits/sec. To ameliorate this bottleneck, it is essential to perform some preprocessing operations on the data while it is in parallel optical form. One basic preprocessing operation in database machines that reduces the data throughput is the SELECTION operation employed by database query languages. When the data retrieved from the secondary storage is filtered through the SELECTION processor, only that subset that meets the desired criteria are forwarded to the host processor for further processing. This SELECTION operation involves comparing the retrieved data items with the selection criteria and therefore involves binary string matching operations as well as alpha-numerical equality/inequality detection operations. For example, consider the following SELECTION criteria for retrieval of records from an employee database: (1) *NAME = ROBERT JONES*, (2) *SALARY < \$40,000 AND AGE > 55 OR GRADE = 12*. The first retrieval can be implemented with a simple binary string matching processor. The second retrieval, however, is based on a logical combination of numerical inequality detection operations.

The use of binary optical correlators followed by post-detection thresholding has been previously proposed to implement the binary string matching operation for database machines [4]. These processors can also perform the numerical equality detection operation. The main subject of this paper is optical processors for alpha-numerical inequality detection. Two distinct approaches are described: (1) time integrating bit-serial, word-parallel approach, based

on optical logic, and (2) space-integrating bit and word parallel approach, based on analog optical D/A conversion.

Bit-serial and Word Parallel, Time Integrating Approach

The operation of numerical inequality detection on a binary, fixed-point representation can be accomplished in a bit-sequential fashion. The two numbers are compared starting from their most significant bit. The logic diagram of the inequality detection circuit is shown in Figure 1. G1 and G2 are AND gates that generate " $a_i \text{ AND } (\text{NOT } b_i)$ " and " $(\text{NOT } a_i) \text{ AND } b_i$ ", respectively. L1 and L2 are crosscoupled latches that store the results from the first gate. Outputs of G1 or G2 will be "1" only when the two input bit streams are unequal. If the output of G1 becomes "1" before G2, $a > b$ and conversely if output of G2 becomes "1" before G1, $b > a$. The output of G1 (G2) causes L1 (L2) to be latched to "1". This in turn disables L2 (L1) due to the crosscoupling connection, thus preventing comparisons of lower significance bits from interfering with the word comparison results. The states of L1 and L2 (c_1, c_2) at the end of the bit streams representing numbers a and b encode the comparison as follows: $(1,0) \Rightarrow a > b$, $(0,1) \Rightarrow b > a$, $(0,0) \Rightarrow a = b$, $(1,1) \Rightarrow \text{ERROR}$. It should be noted that the same circuit can perform binary string matching and numerical comparison regardless of the length of the string/word. A partial comparison/matching can be accomplished by disabling the gates and latches during a specific part of the binary string.

In one possible optical implementation of this circuit the logical AND is performed by analog multiplication. A dual-rail encoding scheme obviates the need for an optical inverter. The dual-rail encoded reference bit pattern (MSB first) modulates a light source (LED, Laser Diode) in time. This time modulated light source is used to read data from an optical disk or a holographic memory. The light intensity on the output detector array, thus corresponds to the terms " $a_i \text{ AND } (\text{NOT } b_i)$ " and " $(\text{NOT } a_i) \text{ AND } b_i$ ". Each detector element has associated with it an electronic latch with the desired crosscoupled connections. The schematic diagram of the resulting system is shown in Figure 2. The light source is broadcast over the entire 2-D data array. The SELECTION operation is therefore performed over several data records in parallel. The mechanical motion of the disk in synchronism with the light source modulation naturally forms the desired bit products in the correct time sequence. The output detector array will therefore need to be only 1-D. If the data is retrieved from a holographic storage system, a 2-D array with scrolled readout (modified CCD scheme) is needed to generate the desired terms. The status of the latches are read at the end of each record to identify those that meet the SELECTION criterion. During the retrieval cycle, only these records are transferred to the host processor, resulting in the desired data reduction.

This bit-serial approach uses identical optical systems as those required for parallel readout optical disks or holographic storage systems. The main modification required is in the time modulation of light sources and detector arrays with simple on-chip processing. The dynamic range and contrast requirements for this scheme will be the same as for the systems without the SELECTION pre-processor. The high modulation rates possible with light sources imply that the desired operations can be performed without slowing down the retrieval rates. The data rate limitations will be primarily imposed by the available readout rates for the detector arrays and will be identical for systems with or without the SELECTION processors.

Bit and Word Parallel, Space Integrating Approach

As described earlier, the fixed-point binary representation of numerical data precludes the use of conventional optical correlation or pattern matching techniques for inequality detection. However, if an optical D/A conversion of the data is performed before comparison, a fast parallel comparison can be implemented. Figure 3 is a schematic depiction of a space-integrating pattern matching system with optical D/A conversion to render the intensity output linearly related to the degree of match to a SELECTION criterion. A mask with exponentially varying transmittance multiplies the binary optical array to perform the D/A conversion. The output of the multiplication is then spatially integrated and collected by a photodetector. The

detector output is fed to an electronic comparator with a reference voltage encoding the SELECTION criterion. This approach thus performs bit-parallel comparison.

Unlike the time-integration approach, the space-integrating approach is limited in accuracy by the analog accuracy of the optical system. Realistically then, only 4-6 bits of accuracy can be expected out of such a processor, and the processor would be setup to operate only on the 4-6 most significant bits of the data word. This approach, therefore, will be appropriate as a pre-filter to reduce the size of the search space. If just 4 bits accuracy are achievable, then such a prefilter will, on average, reduce the size of the database needed for the digital processor, by a factor of 16. Thus, the amount of data to be transferred from secondary memory to main memory for the SELECTION operation is reduced by over an order of magnitude.

The organization of the data into tables, in which values to be compared are located side-by-side, provides a straight forward method to word parallel operation. Figure 4 is a schematic depiction of this approach. The numerical data are assumed to be organized as, say, 32 bit words. The D/A transparency covers only the most significant 4-6 bits of a block of data. All other bits are covered by opaque areas of the transparency. Using an SLM for this transparency will provide programmability to correct for variations in optical components, etc. The outputs of the detector/comparator array thus serve as flags which tell the digital processor which words to down-load from the secondary memory for higher accuracy scrutiny. The flag outputs of such a module can be combined electronically with those of other modules to perform composite preselection operations which potentially reduce the secondary storage transfer rates even more.

The operation of the space-integrating SELECTION processor will be similar for parallel access disk-based and page-oriented holographic memory. The only operational difference will be that the movement of the disk will require the light source to be pulsed for a period less than the dwell time of the pixel under the transparency mask. If the data words are arrayed in parallel, across adjacent tracks of the disk, then the SELECTION processor would operate in a strobed mode, with valid outputs of the comparator only when the MSBs are under the transparency. The processor data rate would therefore be limited to about the time it takes a page of data words to move completely under the transparency -- several microseconds. Interestingly enough, the page comparison data rate for a holographic storage-based approach is also on the order of several microseconds. The difference is that the holographic storage approach also has a projected random access time of about several microseconds, while a disk-based system's access time is on the order of milliseconds.

Summary

Optical SELECTION processors based on numerical inequality/equality detection serve to reduce the amount of data that is transferred from high throughput optical memory to the host processor. This preprocessing operation thereby reduces a potential bottleneck while still exploiting the parallel data access capabilities of optical memory. Two designs for optical SELECTION processor based on bit-serial digital approach and bit-parallel analog approach are described. Both designs are fully compatible with parallel access optical storage based on disks or holograms.

References

1. P. B. Berra et al, "Optical database/knowledgebase machines", Appl. Opt., 29, p. 195, 1990.
2. D. Psaltis et al, "Optical memory disks in optical information processing", Appl. Opt., 29, p. 2038, 1990.
3. H. J. Caulfield, "Massively parallel optical database management", Proc. SPIE, 938, p. 52, 1988.
4. P. Guilfoyle and W. J. Wiley, "Combinatorial logic based digital optical computing architectures", Appl. Opt., 27, p. 1661, 1988.

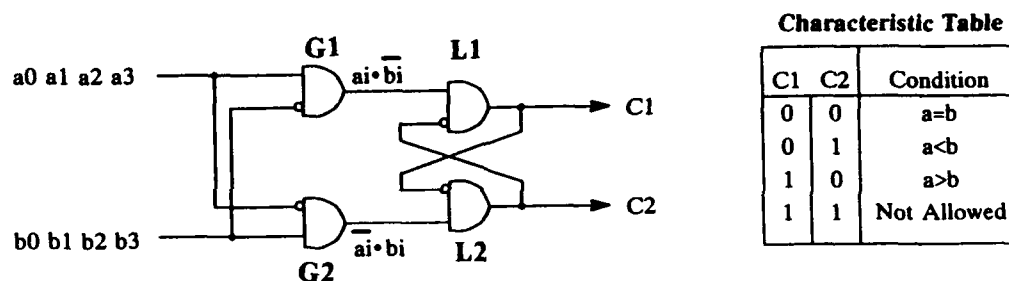


Figure 1: Bit-serial, time-integrating inequality detection circuit

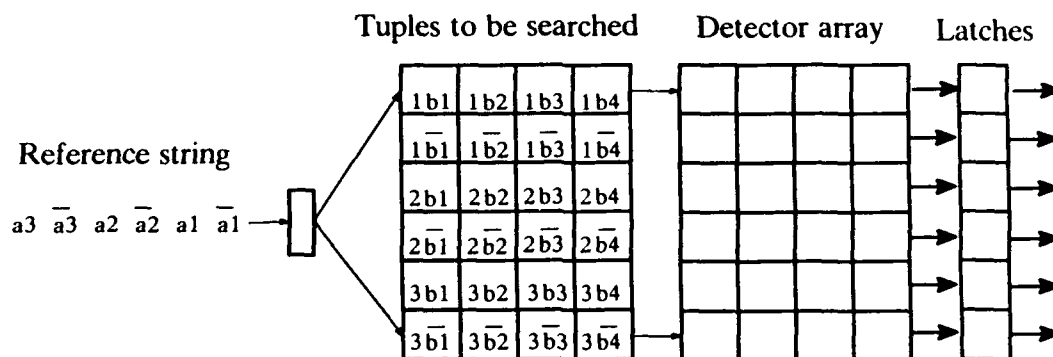


Figure 2: Schematic diagram of a bit-serial inequality detector system processing multiple records (tuples) in parallel. First index is the record index and the second one the bit index.

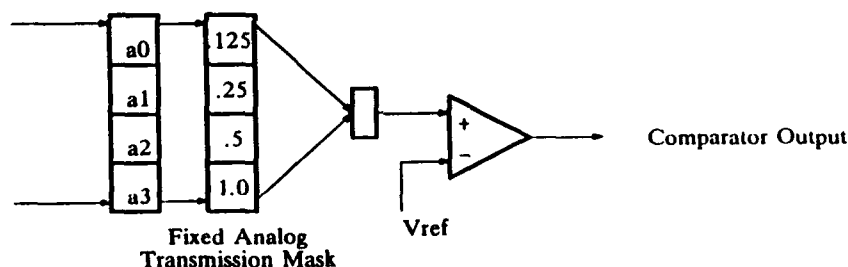


Figure 3: Space-integrating inequality detector based on analog optical D/A conversion.

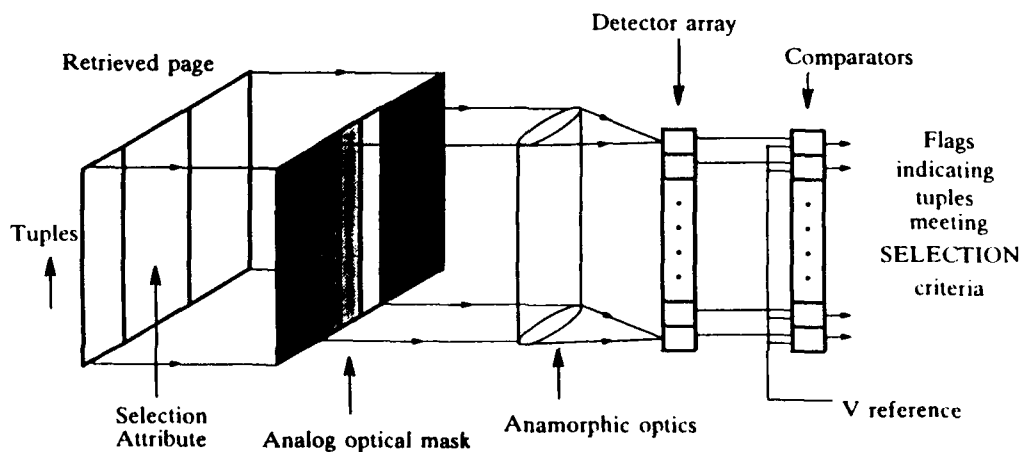


Figure 4: Analog optical inequality detector operating on multiple records (tuples) in parallel.

Demonstration of an All Optical Addressing Circuit

Donald M. Chiarulli, Steven P. Levitan, Rami G. Melhem
Departments of Computer Science and Electrical Engineering
University of Pittsburgh

1. Introduction

This experiment is based on two properties of optical signals, unidirectional propagation and predicable path delay. Using these properties, logic systems can be devised in which information is encoded as the relative timing of two optical signals. Coincident pulse addressing is an example of such a system. In this case, the address of a detector is encoded as the delay between two optical pulses which traverse independent optical paths to a detector. The delay is encoded to correspond exactly to the difference between the two optical path lengths. Thus, pulse coincidence, a single pulse with power equal to the sum of the two addressing pulses, is seen at the selected detector site. Other detectors along the two optical paths for which the delay did not equal the difference in path length, see both pulses independently, separated in time.

Stated more formally, consider a fiber of length L with two optical pulse sources, P_1 and P_2 coupled to each end. Each source generates pulses of width τ and height h . Define $l = \tau c_g$ where c_g is the speed of light in the fiber. In other words l is the length of fiber corresponding to the pulse width. Using 2×2 passive couplers, n detectors, labeled D_0 through D_n , are placed in the fiber with the two tap fibers from each coupler cut to equal length and joined at the detector site. The location of each coupler/detector is carefully measured so that the k th detector is located at $(L - nl)/2 + (k - 1)l$. The optical bus in the center of figure 1 shows such an arrangement for $n = 4$. To uniquely address any detector, a specific delay between the pulses generated by P_1 and P_2 is chosen. If this delay corresponds to $t_1 - t_2$, then when $t_1 - t_2 = [n - 1 - 2(k - 1)]\tau$ the two pulses will be coincident at detector D_k .

The same technique can be generalized to support parallel selections. If one of the sources is allowed to generate a series of pulses with each t_k timed relative to t_1 to select a specific detector k , then according to the addressing equation t_k will be in the range $-(n - 1)\tau \leq t_1 - t_k \leq (n - 1)\tau$, for $k = 1..n$. In other words, any or all of the k detectors can be uniquely addressed by a positionally distinguishable pulse from source P_2 . For convenience, this pulse train is referred to as the select pulse train and the single pulse emanating from P_1 is called the reference pulse. Since the length of the select pulse train is n , and each pulse in the train is separated by 2τ it follows that the system latency, $\sigma = 2n\tau$. Since up to n locations may be selected in parallel within a single latency period, the system throughput is thus $v = 1/2\tau$. Readers who are interested in the general application of coincident pulse techniques are referred to the references [1, 3].

2. Experimental Results

Figure 1 is a diagram of the prototype structure. The fiber bus consists of a length of multimode fiber tapped four times using Gould 10 dB fiber couplers. Select and reference bit patterns are generated by modulating the 4ns pulse output of a Tektronix PG502 pulse generator, shown in the diagram as clock, with the output of two ECL shift registers, one for select, one for reference, at gates G2 and G3. Gates G1 and G4 simultaneously hold the diode current for laser diodes P1 and P2 respectively at threshold while the outputs of G2 and G4 generate modulation current. The result is two, 4-bit, return to zero bit streams which encode the information in each of the shift registers.

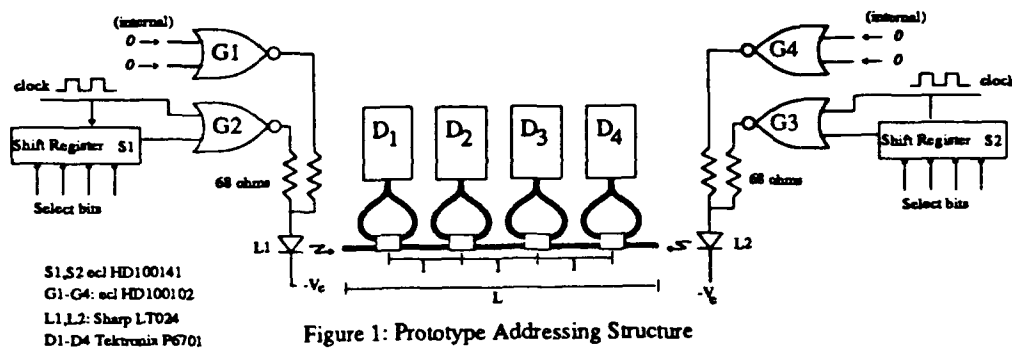


Figure 1: Prototype Addressing Structure

Figure 2 shows the output waveforms for detectors D1 and D3 for various selection patterns. Figures 2a and 2b show coincident and non-coincident waveforms at detectors D1 and D3 respectively. Note that in both cases, the non-coincident waveforms shown on the right are of unequal power. This is due to the fact that each pulse has passed through a different number of couplers and has hence become attenuated to different levels. Thus the relative power between coincident and non-coincident pulses is a function of the detector location. The amount of additional power in the coincident pulse relative to the largest non-coincident pulse is called the power margin, m , and is defined as a fraction of the maximum non-coincident pulse power by $m = [p_1 + p_2 - \max(p_1, p_2)] / \max(p_1, p_2)$. For both of the single selection experiments shown in figures 2a and 2b, the power margin is in excess of .5. This is true even for D_1 which is leftmost on the bus.

Figures 2c and 2d are examples of parallel selections. The left waveform in figure 2c shows a parallel selection waveform at detector site D_3 for the selection of three detectors, including D_3 . This coincident waveform peak compares to the non-coincident waveform on the right in which D_3 has been removed from the set of selected locations. Similarly figure 2d shows parallel selection of all four detectors at sites D_1 and D_3 .

3. Pulse Synchronization

In a second experiment, measurements were made to characterize the effect of synchronization error between the reference and select pulses on the power margin of the coincident pulse. Since clearly this error is characterized as a percentage of the pulse width, synchronization precision has a direct bearing on the absolute width and height of an addressing pulse that can be effectively detected. The apparatus used was identical to the previous experiment except that the number of detectors was reduced from four to three. This allowed detector D_2 to be located in the center of the bus resulting in exactly equal non-coincident pulse heights as shown in figure 3a. The reference and select pulse trains were configured to select D_2 . In each step of the experiment synchronization error was introduced by adding successively longer lengths of fiber to the bus. Length was added first on the reference pulse end of the bus, and then on the select pulse end of the bus.

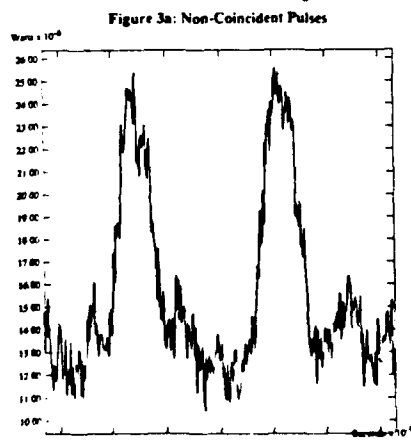


Figure 3a: Non-Coincident Pulses

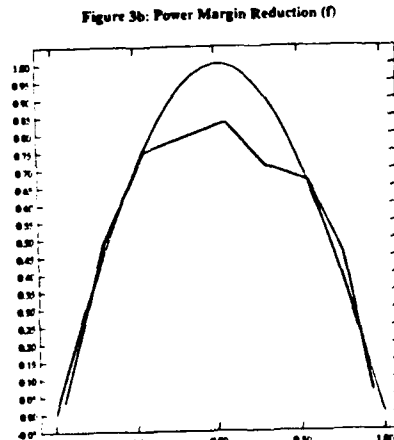


Figure 3b: Power Margin Reduction (f)

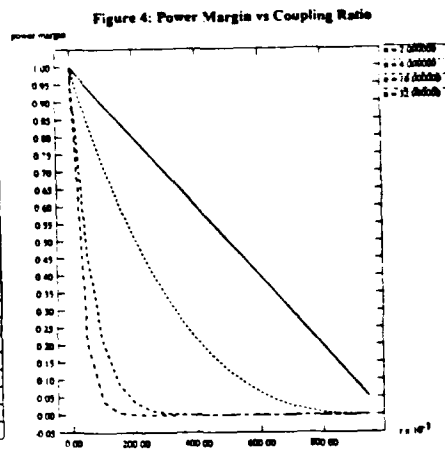


Figure 4: Power Margin vs Coupling Ratio

Figure 3b shows the reduction factor, f , of the power margin as a function of percent synchronization error. Percent synchronization error is the error, in time units, introduced by each length of fiber divided by the pulse width. In other words pulses at perfect coincidence (synchronization error = 0) yield a reduction factor of $f=1.0$ which is, by definition, the power margin. Synchronization error in either the select pulse, shown as positive error, or the reference pulse, shown as negative error, reduces the power margin by the factors shown. The solid line in figure 3b is the experimental result. The dotted line is a simulated result generated from the coincidence of sinusoidal pulse waveforms. In both cases power margin falls off in roughly the shape of the coincident waveforms. Thus the "flatness" of the experimental pulses results in a flattening of the power margin curve, while the sinusoids fall off somewhat more smoothly. These waveforms characterize the temporal limits on scalability. That is to say, the limit on pulse width, latency, and throughput.

4. Power Distribution

Since the bus configuration chosen for this experiment requires bidirectional propagation, we are constrained to use a single tapping ratio, r , for all couplers. Therefore, assuming a unit height pulse from each direction, the optical power p_1 and p_2 at detector D_k are given by the equations

$$p_1 = r^{(k-1)}(1-r), \quad p_2 = r^{(n-k)}(1-r)$$

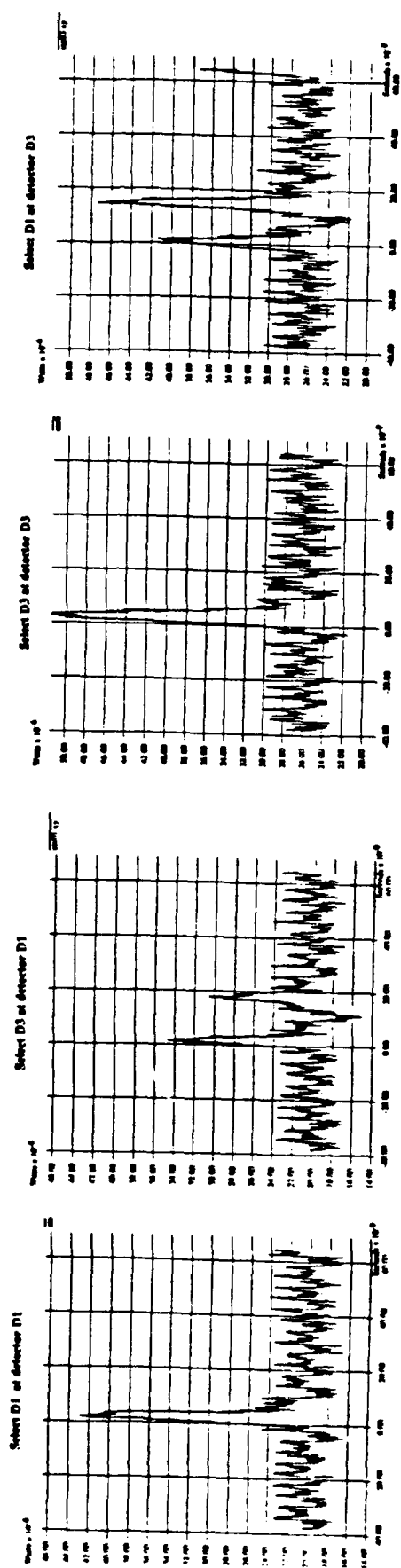
Since the absolute power falls off geometrically with increasing n , and power margin essentially bounds scalability, the size of the system is highly sensitive to the value of r . In figure 4 we have plotted worst case power margin versus coupling ratio for various bus sizes n . To determine an overall bound on system scale, the effects of synchronization error and power distribution limits must be considered jointly. The following procedure can be used. First, a minimum power margin m_d is selected such that a reasonable threshold can be established based on signal to noise ratio. Next, synchronization error, based on the pulse width and the accuracy of the fiber lengths, is used to determine worst case reduction in power margin, f . The actual power is calculated as m_d/f . Finally, the maximum number of detector sites can be determined based on figure 4 and the power equations above.

5. Discussion

Clearly, three factors, threshold power margin, synchronization error, and coupling ratio determine system scale. Based on current and near term technology, our experiments show that synchronization error does not contribute significantly to the bounds calculated above. Rather, power distribution effects dominate. However, we believe that near term technologies such as fiber amplifiers as well as alternate bus structures [2] will alleviate this problem. The fact that temporal scalability limits show that significantly shorter pulses can be supported, is very encouraging for the long term application of this technique.

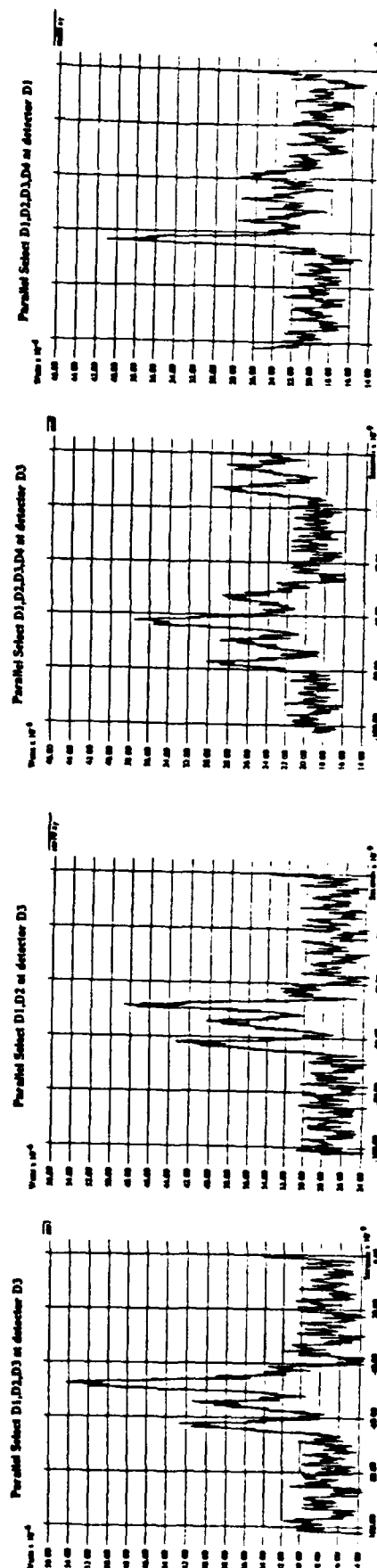
References

1. D. Chiarulli, R. Melhem, and S. Levitan, "Using Coincident Optical Pulses for Parallel Memory Addressing," *IEEE Computer*, vol. 20, no. 12, pp. 48-58, 1987.
2. R. Dittmore, D. Chiarulli, R. Melhem, and S. Levitan, "An Analysis of Power Distribution in Optical Buses," *this conference*.
3. S. Levitan, D. Chiarulli, and R. Melhem, "Coincident Pulse Techniques for Multiprocessor Interconnection Structures," *Applied Optics*, vol. 29, no. 14, pp. 2024-2033, May 1990.



(a)

(b)



(c)

(d)

FIGURE 2. D1 and D3 output; various tests

An Optical Respite from the Von Neumann Bottleneck

Alex Dickinson

AT&T Bell Laboratories
4E-514 Crawfords Corner Rd
Holmdel NJ 07733
(201) 949 1083

1 Introduction

The high end of microprocessor performance is currently dominated by Reduced Instruction Set Computer (RISC) architectures. These machines execute one or more instructions per clock cycle. A processor such as the i860¹ [1] runs with a 40MHz clock – requiring that on average an instruction must be delivered to the CPU every 25nS. With DRAM access times currently at around 100nS, timely instruction delivery has become a critical constraint on processor speed.

The primary tool for dealing with this problem is the use of fast cache memories local to the processor. These caches make use of both temporal locality (if the processor just accessed a location, it will probably do it again soon) and spatial locality (if the processor just accessed a location, it will probably access a nearby one soon). The caches are implemented in fast static RAM on the processor die. If an item is in cache (a 'hit') it may typically be retrieved within a single processor cycle (the *hit time*). If an item is not in cache (a 'miss') it must be retrieved from the off-chip main memory at a considerable cost in time. This later time is referred to as the *miss penalty* and so we may write [2]:

$$\text{AverageMemoryAccessTime} = \text{HitTime} + \text{MissRate} \times \text{MissPenalty} \quad (1)$$

—where all times are in processor cycles. Hit time and miss rate are dependent on a number of factors: the cache organization (direct mapped or associative), the number of blocks in the cache (blocks are the atomic units of storage in the cache), and the size of each block (a block may be any number of bytes wide).

From the above equation we can see that for a given hit time (typically a single cycle), we can only reduce memory access time by lowering the miss rate and/or the miss penalty. The miss penalty may be defined as [3]:

$$\text{MissPenalty} = \text{DRAMlatency} + \text{BlockSize} / \text{TransferSize} \quad (2)$$

That is to say that the miss penalty is comprised firstly by the time required to get an address to the DRAM and decode the row, and secondly by the number of cycles required to fill a cache block given the (typically smaller) size of each transfer from the DRAM (Figure 1). One cannot simply hope to see a dramatic speed-up in DRAM latency: their high density comes at the price of inherent low speed.

Simulations show that miss rates continue to reduce as block size is increased up to large blocks of 256 bytes [3]. But larger blocks increase the *BlockSize/TransferSize* ratio in Equation 1, and so the resulting increase in miss penalty outweighs the improvements in miss rate. This results in present machines using relatively small block sizes of less than 32 bytes.

How then can we harness the benefits of increased block size without paying the price of increased miss penalty? The solution can be found by examining the packaging and interconnection of the processor and memory. Internal to the DRAM, memory is accessed in wide rows that are time multiplexed out through the package pins, across the bus and into the processor. The inherent parallelism of the memory is lost because

¹i860 is a trademark of Intel Corp

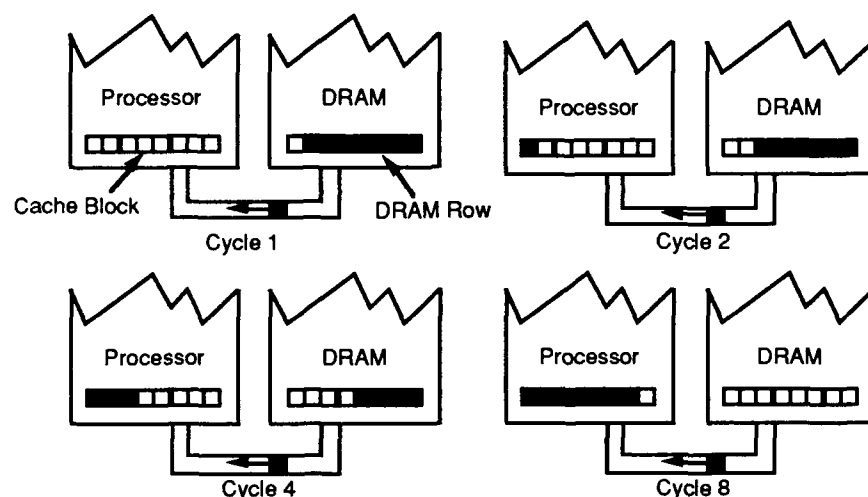


Figure 1: In this example, a cache miss has instigated a 32 byte block transfer from the DRAM to the cache. The processor/memory bus is only 4 bytes wide, thus incurring a penalty of 8 bus transfers for the miss.

of limited memory and processor package pin counts and the narrowness of the electrical bus: the physical incarnation of the Von Neumann bottleneck.

If we could provide a very large number of channels for communication between the memory and processor, large segments of memory could be transferred into the processor in a single access, allowing us to use wider cache blocks with very little increase in miss penalty over that required for small cache blocks. This would result in reducing the *BlockSize/TransferSize* ratio to unity, its lowest useful value.

In this paper we describe how free space optical interconnection technology may be applied to this problem. Some implementation issues are discussed, and a simple performance analysis presented.

2 Architecture

The scheme is illustrated in Figure 2. Addresses are passed from the processor chip to the memory chips over a conventional electrical bus. Each memory chip is read out via an array of k microlasers [4] or SEEDs [5]. The resulting array of k points is imaged onto CMOS photodetectors on the processor chip using free space optical techniques [6].

The number of channels (k) and hence the size of the array is determined by the cache block width, a parameter that can only be selected by extensive simulation of a particular architecture. Generalized results show however that block sizes of 256 bytes in an instruction cache result in optimal miss rate performance [3]. This corresponds to $k = 2048$, considerably in excess of the size arrays that we believe are currently practical. A more reasonable value of $k = 512$ can still be of considerable benefit, and such a link might in fact be time multiplexed to form a 2048 channel link (given the slow 100nS access times of the DRAMs, speed constraints are not tight).

When a cache miss occurs, the address of the required block is placed on the address bus and used to access a row in a DRAM chip. A block of k bits in that row are transferred over the optical link to the cache in a single cycle. Writes from the processor back to memory may be performed over an electrical bus. This is

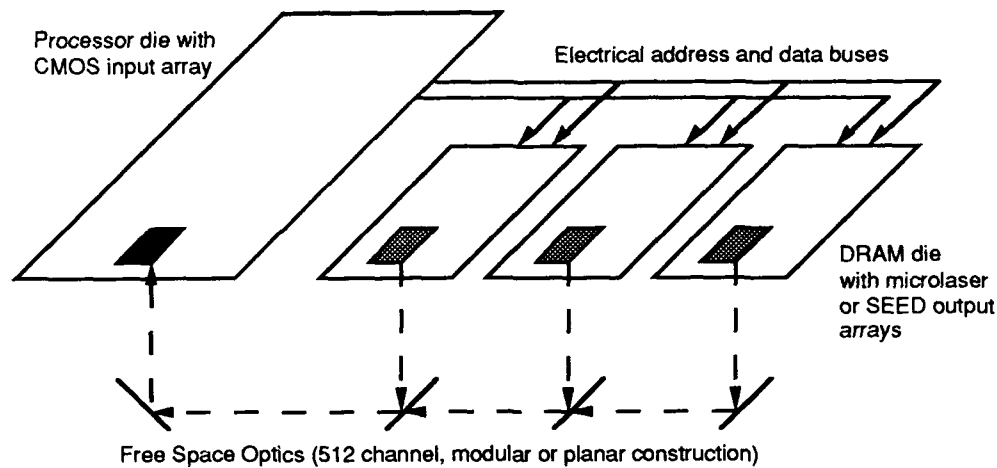


Figure 2: An overview of the scheme. In this configuration only the memory read bus is optical; the less critical address and write buses are electrical.

possible because the ratio of reads to writes is typically 10:1, and also writes are amenable to being buffered [2]. Similarly the address bus may be electrical because of the relatively small size of each address transfer.

The simple topology of the scheme makes it an attractive application for planar optics [7]. Such an implementation would offer the low component count, high stability and resulting low cost required for use in manufacturable systems. The optical path would be provided inside the planar glass substrate, and the chips could be flipped over and bump bonded onto electrical interconnect lithographically defined across the glass surface.

A primary constraint in a cost effective implementation would be the integration of optical devices (SEEDS or lasers) with the DRAM circuitry. This could be done either by the creation of Si/GaAs hybrids, or (more desirably) as the result of the ongoing development of GaAs on silicon.

3 Performance

For a performance estimate of this scheme we evaluate its first order effect on the memory access time of a recent commercial RISC chip, the Intel i860. The i860 incorporates a 4 Kbyte instruction cache with 32 byte blocks organized with 2-way set associativity [1]. The published simulation results indicate that this organization produces a cache miss rate of 0.064. The i860 contains a 64 bit data bus, requiring 4 memory accesses to transfer a 32 byte cache block. This results in a 24 processor cycle miss penalty [8] with no external bus pipelining. From Equation 1:

$$\begin{aligned}
 \text{Average Memory Access Time} &= \text{HitTime} + \text{MissRate} \times \text{MissPenalty} \\
 &= 1 + 0.064 \times 24 \\
 &= 2.5
 \end{aligned}$$

Now, with a 512 channel optical bus, it would be possible to replace a 64 byte cache block in a single memory access (6 processor cycles). The increased block size results in a reduced miss rate of 0.046 [1] which combines with the reduced miss penalty to give:

$$\begin{aligned}
 \text{AverageMemoryAccessTime} &= \text{HitTime} + \text{MissRate} \times \text{MissPenalty} \\
 &= 1 + 0.046 \times 6 \\
 &= 1.3
 \end{aligned}$$

Thus in this case the proposed scheme would result in an almost 2-fold improvement in average memory access time.

4 Conclusions

The scheme presented here reduces average memory access time by providing a wide optical data path between processor and memory. This wide path may be used to fill a wide cache block in a single access, reducing the high cost normally associated with such a configuration.

Simple calculations indicate that average memory access time may be reduced by a factor of two. This noteworthy given that DRAM access times have improved by less than a factor of two in the last 10 years.

From the optical implementation point of view the scheme has a number of advantages: the topology is simple point-to-point, a basic scheme would only require unidirectional communication, and the data rate required is quite low. The scheme makes use of the high interconnect density, low power and regularity that are the hallmarks of free space optical interconnect.

Further work involves investigating the use of flat optics for implementation of the scheme, and carrying out instruction set simulations to evaluate in more detail the efficacy of such wide caches.

References

- [1] P. Patel and D. Douglass, "Architectural Features of the i860 Microprocessor RISC Core and On-Chip Caches", *Proceedings of the IEEE International Conference on Computer Design*, Cambridge, Mass, October 1989.
- [2] John L. Hennessy and David A. Patterson, "Computer Architecture: A Quantitative Approach", Morgan Kaufmann, 1990.
- [3] "Performance Tradeoffs in Cache Design", S. Przybylski, M. Horowitz, and J. Hennessy, *Proceedings of the 15th Annual Symposium on Computer Architecture*, Honolulu, May, 1988.
- [4] J. L. Jewell, A. Scherer, S. L. McCall, Y. H. Lee, S. J. Walker, J. P. Harbison and L. T. Florez, "Low Threshold Electrically-Pumped Vertical-Cavity Surface-Emitting Micro-Lasers", *Appl. Phys.*, 31st July 1989
- [5] L. M. F. Chirovsky, L. A. D'Asaro, C. W. Tu, A. L. Lentine, G. D. Boyd, David A. B. Miller, "Batch Fabricated Symmetric Self Electro-optic effect Devices", *Proceedings of the Optical Switching Topical Meeting*, Salt Lake City, Utah, March 1989.
- [6] A. Dickinson and M. E. Prise, "A Free Space Optical Interconnection Scheme", *Proceedings of the Optical Computing Topical Meeting*, Salt Lake City, Utah, March 1989.
- [7] J. Jahns, "Planar Packaging of Complex Optical Systems", *Proceedings of the 15th Congress of the International Commission for Optics*, Garmisch-Partenkirchen, FRG, August 1990
- [8] M. Rhodehamel, "The Bus Interface and Paging Units of the i860 Microprocessor", *Proceedings of the IEEE International Conference on Computer Design*, Cambridge, Mass, October 1989.

Tuesday, March 5, 1991

Architectures and Signal Processors

TuD 4:00pm–5:40pm
Salon F

Pochi Yeh, *Presider*
University of California, Santa Barbara



Dual Scale Topology Opto-Electronic Processor (D-STOP): Comparative Analysis and Technological Feasibility

*A. V. Krishnamoorthy, J. E. Ford, G. C. Marsden, G. Yayla,
S. C. Esener, and S. H. Lee*

University of California, San Diego
Department of Electrical and Computer Engineering
9500 Gilman Drive
La Jolla, CA 92093-0407

Introduction:

A variety of applications in artificial neural networks, interconnection networks, artificial intelligence, relational databases, and numerical processing require parallel, large scale implementations of matrix-algebraic architectures. Existing VLSI implementations of these architectures are restricted in terms of their parallelism and bandwidth due to their inherent connectivity, pin-out, power dissipation, and crosstalk limitations.^[1,2] On the other hand, existing optical matrix-vector architectures suffer from limited SLM throughput and accuracy as well as limited functional flexibility. In the following sections we describe and analyze the Dual-Scale Topology Opto-Electronic Processor (D-STOP)^[3] which alleviates these limitations, and discuss its feasibility for a near-term implementation.

D-STOP Architecture/System Description:

D-STOP is a parallel, fully connected opto-electronic computing architecture designed for matrix-algebraic data processing, with the essential operations being generalized matrix-vector multiplication and vector outer-product. The D-STOP system consists of arrays of N opto-electronic Processing Elements (PEs) with modulators arranged in a 2-D topology. These PE arrays are fully connected via space invariant, free space optical interconnections. Each PE in an array consists of N electronic detector sub-units which have optical input and electronic output (Fig. 1). These detector sub-units are placed in a pattern similar to the PE layout in the array, but at a smaller scale. The outputs of the detector units are electronically summed via an area-efficient H-tree structure. At each node of the H-tree are additional fan-in processing sub-units. At the center of the H-tree is a single sub-unit that processes the electronically collected output of the H-tree. The output of this central unit is optically broadcasted to the corresponding detector sub-units of other PEs using an optical transmitter.

The dual-scale invariant layout of the PEs and their corresponding detector sub-units allows full connectivity to be achieved via demagnification and replication (Fig. 2). The transfer function of the optical system is space-invariant leading to a simple, scalable optical system. Several optical systems can provide the full broadcast interconnections desired. For example, Farhat et al.^[4] used a microlens array to replicate the input. Each lenslet forms an image of the entire input array onto one output processor. However, the aperture and resolution of each lenslet necessarily limits the resolution of the entire system. An alternative is to use holographic beamsplitting in a common-path system. The simplest of these uses a single demagnifying lens in contact with a holographic 1 to N beamsplitter and results in a system whose length scales as $O(N^{3/2})$ for a fixed F-number. The system shown in Figure 3 uses additional optical components to achieve better scaling behavior. The first two lenses form a demagnified image of the input array of modulators. The third lens transfers this image to the output plane. Finally, a holographic beamsplitter in contact with the third lens performs the replication and can also provide aberration correction. Because the light shares a common path, there is no small aperture bottleneck, and the system's diffraction-limited resolution is high. The telecentric demagnifying stage maintains high throughput efficiency, and separates the holographic beamsplitter from the short focal length demagnifying lens, allowing a fixed maximum diffraction angle.

Technology Considerations:

The D-STOP system has been designed to take full advantage of both free-space optical interconnections and electronic VLSI systems on a hybrid OEIC technology base. The system achieves full connectivity between PEs using space-invariant optical interconnections that can be efficiently implemented with existing refractive optical elements and rapidly developing multi-level phase diffractive optical elements. Since a thin CGH beamsplitter is used, mutually incoherent optical sources such as laser diodes or even narrow linewidth LEDs can be used instead of modulators. The system also minimizes the number of required modulators compared to existing opto-electronic

matrix-vector architectures^[5] thereby allowing the silicon ICs and the light modulators to be fabricated on separate chips (or wafers) and later bonded face-to-face using available electronic packaging technologies (Fig. 4). The electrical connections between the output of the ICs and the electrodes of the modulators are realized through Indium bonds. Since the density of modulators needed is low, the flip-chip bonding process can provide a near term OEIC implementation with relatively high yield. PLZT light modulators are best suited for such a D-STOP implementation since they can provide large fan-out (up to 1,000) with acceptable power dissipation due to their non-absorptive nature. Furthermore, they can be operated at high speeds with relatively large contrast ratios, which allows simple detector designs and high system bandwidth. The electronic H-tree fan-in structure is advantageous because it reduces signal skew and allows pipelined operations.

System Analysis:

The scaling of the system is well behaved since both the opto-electronic chip and the optical system have identical growth rates. The H-tree fan-in structure^[6] allows an $O(N)$ area layout for the detector sub-units of one PE. The total area (SBP) required by the optical system is also $O(N^2)$ since a space invariant optical system is used. Because the holographic beamsplitter is functionally separate from the demagnifying stage, the maximum diffraction angle does not increase with array size. As a result, the system length can be shown to scale as $O(N)$ while maintaining a constant F-number and CGH minimum feature size. In addition, the system size is not limited by the power dissipation of optical source/modulator, even at high switching speeds, since individual transmitters are placed far apart on the opto-electronic chip. The yield of the electronic circuitry does not limit the system size, since no inter-PE electronic communications are necessary. The PEs can therefore be implemented in a modular fashion on separate chips, which are then placed on a multi-chip carrier that can house several hundred such chips.^[7] Finally, total optical power requirements indicate that a system with 10^6 detector units can be achieved.

D-STOP Applications:

Since all the mathematical operations associated with the matrix-vector and outer-product procedures are performed electronically, these can be generalized to symbolic or nonlinear numeric operations. Additional processing is available during fan-in, generalizing the conventional summation of inner-products. The architecture can thus be tailored to suit a variety of algorithms and applications including multi-layer feed forward neural networks, back-propagation networks, crossbar interconnection networks, database systems, etc. A critical issue for D-STOP implementations is the method of data representation, which should be chosen to minimize silicon area and on-chip power dissipation while providing the precision necessitated by the application in question. For neural networks in particular, a combination of pulse width modulating optical neurons and pulse amplitude modulating electronic synapses provides the highest hardware efficiency.^[8] Hybrid analog/digital electronic circuits have been designed that allow system precision to be continuously traded-in for silicon area. Based on this design, a 1,000 neuron system with $> 10^6$ weighted interconnections and $> 10^{12}$ interconnections/sec can be implemented with feasible chip area, power dissipation, optical SBP, and power requirements [Table 1]. The memory capacity of the system can be increased to $> 10^8$ interconnections using parallel-accessed memory devices such as the motionless-head parallel readout optical disk.^[9]

Conclusions:

The D-STOP system uses an optimal combination of space-invariant, free-space optical interconnection and electronic interconnection/processing to achieve parallel implementations of generalized matrix-vector and vector-outer product operations. Using state of the art VLSI and opto-electronic technology, a system with greater than 1,000 fully connected processing elements can be achieved in the near-term for applications including neural networks crossbar multiprocessor systems, etc. In our presentation, we will provide a detailed system analysis and present an experimental demonstration of the optical system.

References

1. J. W. Goodman, F. I. Leonberger, S. Y. Kung, and R. A. Athale *Proceedings IEEE* 72, 850, 1989.
2. M. R. Feldman, S. C. Esener, C. C. Guest, and S. H. Lee *Appl. Opt.* 27 (9) May 1, 1989.
3. G. Marsden, A. V. Krishnamoorthy, S. C. Esener, and S. H. Lee, Paper TuJJ2, *OSA Annual Meeting* Boston, 1990.
4. N. Farhat and D. Psaltis, Optical Signal Processing, Chapter 2.3, Academic Press 1987.

5. J. W. Goodman, A. R. Dias, and L. M. Woody *Optic Letters* 2 (1) January 1978.
6. C. Mead and M. Rem, *IEEE JSSC* SC-14 (2) April 1979.
7. H. B. Bakoglu *Circuits Interconnections and Packaging for VLSI*, Addison Wesley, 1990.
8. A. V. Krishnamoorthy, G. Yayla, S. C. Esener, submitted to IJCNN, Seattle 1991.
9. P. J. Marchand, A. V. Krishnamoorthy, P. Ambs, and S. C. Esener, *SPIE Proceedings* 1347, San Diego, 1990.

Total Power	Area / neuron	Power diss.	CGH Area	System length
≈ 50 mW/neuron	≈ 0.1 cm ²	≈ 0.5 W/cm ²	10x10 cm ²	≈ 60 cm

Table 1: Application of D-STOP to a 1,000 neuron neural network implementation.

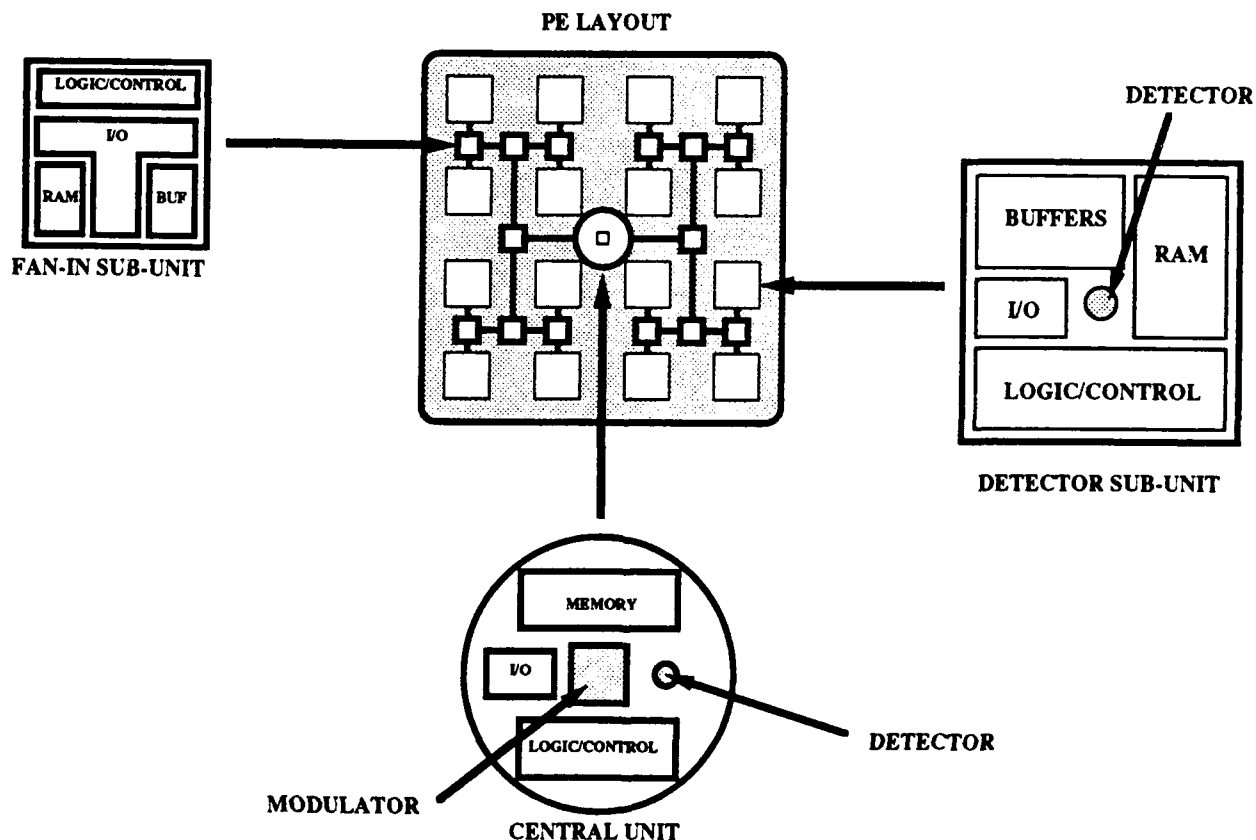


Fig 1: PE layout showing detector sub-units, fan-in sub-units, and central unit

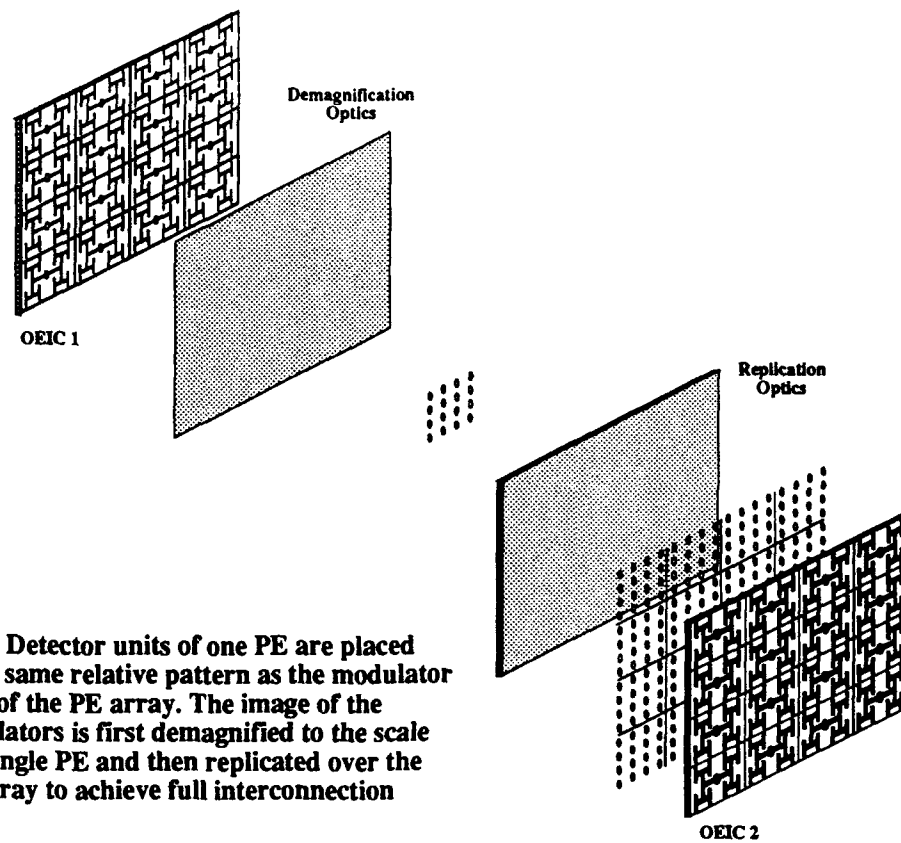


Fig 2: Detector units of one PE are placed in the same relative pattern as the modulator units of the PE array. The image of the modulators is first demagnified to the scale of a single PE and then replicated over the PE array to achieve full interconnection

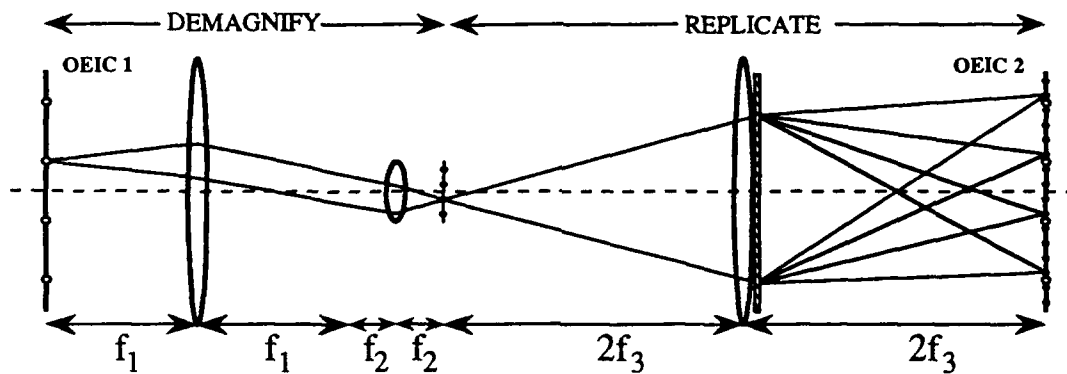


Figure 3: The DSTOP optical system using separate demagnification and replication.

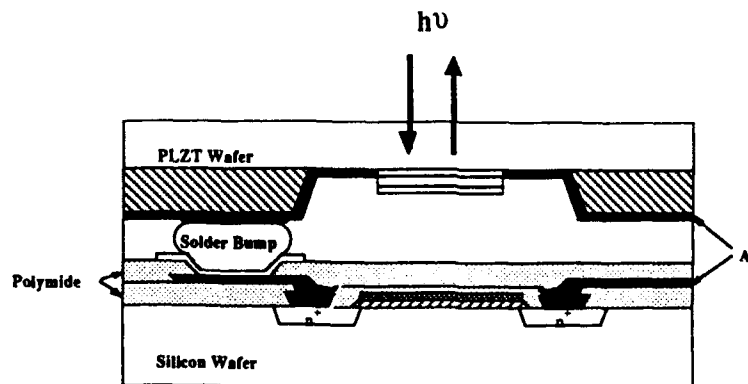


Fig 4: Flip-chip bonded PLZT on Silicon using Indium bumps

Ring Array Processor Distribution Topology For Optical Processing and Interconnect

Yao Li and Berlin Ha

Department of Electrical Engineering,
City College of New York, New York, NY 10031.

I. Introduction

In this paper, a study of a ring array processor distribution topology for optical digital processing and interconnect is presented. The work was motivated by the facts that (1) conventional optical imaging elements such as lenses are circularly symmetric about optic axes, and (2) the existing linear/rectangular array distribution topology is sometimes inefficient in terms of optical implementation and synchronization. The proposed new free-space optical ring array topology based processing and interconnect schemes can solve various existing problems in optical processing and interconnects.

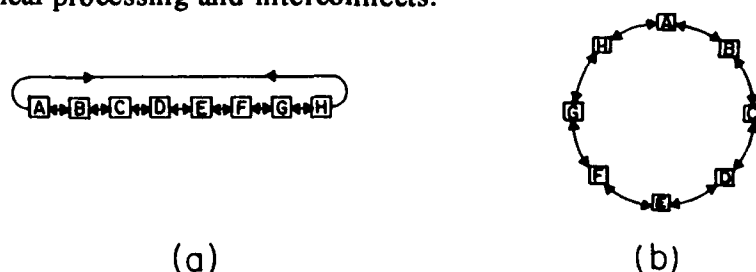


Fig.1. (a) A cyclic shift register. (b) A redrawn of (a) along a ring.

II. Optical Cyclic Shift

For digital processing, one of the gate-level fundamental operations is the cyclic shift operation handled by a register or a register array [1]. Such an operation is essential to digital counting, synchronization as well as to cyclic-convolution/correlation. Using a rectangular array, a unit shift of two adjacent elements and of two end elements (see Fig.1(a)) physically consumes different delay times, thereby reducing clock rate. On the other hand, when such an array is distributed along a ring (see Fig.1(b)), all elements are spaced uniformly. In this case, a clockwise or counterclockwise unit shift consumes minimum time needed for the signal to travel across.

To optically implement a circular shift, one possible scheme is to use a Dove prism pair configured in the way shown in Fig.2. Here, the reflection planes of the two Dove prisms are mutually tilted by an angle α . After two consecutive reflections, the output signal along the ring represents a cyclic shifted (by α) version of the input.

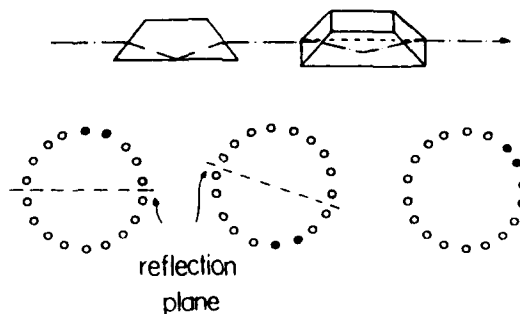


Fig.2. A Dove prism pair configured for cyclic image rotation.

III. Optical Interconnect for SIMD Array Processors

The proposed optical ring processor can also handle angular shifts of more than one units. When a rotation of K units is needed for every element along a ring where $N > K$ elements are uniformly distributed, the angle between the reflection planes of the two prisms needs to be tilted by an angle

$$\alpha = \frac{2\pi K}{N} \quad (1)$$

This rotation flexibility makes it possible for optical interconnect of various SIMD array processors [2]. What follows briefly summarizes the possible applications of optical interconnects using the ring distribution topology.

3-1. Nearest-Neighbor Interconnect

The nearest-neighbor (NN) interconnect provides for each of its N processing elements (PEs), four routing configurations [2]

$$NN_{\pm 1}(i) = (i \pm 1) \bmod N \quad (2a)$$

$$NN_{\pm r}(i) = (i \pm r) \bmod N \quad (2b)$$

where $r = \sqrt{N}$ is a positive integer, and $0 \leq i \leq N-1$. For the case of $N=16$, when these PEs are distributed in a rectangular array (see Fig.3(a)), the implementation of this interconnect requires the use of different, both space invariant and variant optical elements for handling the array's center and edge PEs. On the other hand, when the N PEs are distributed along a ring (see Fig.3(b)), the use of two routing paths each containing a Dove prism pair can accomplish this task.

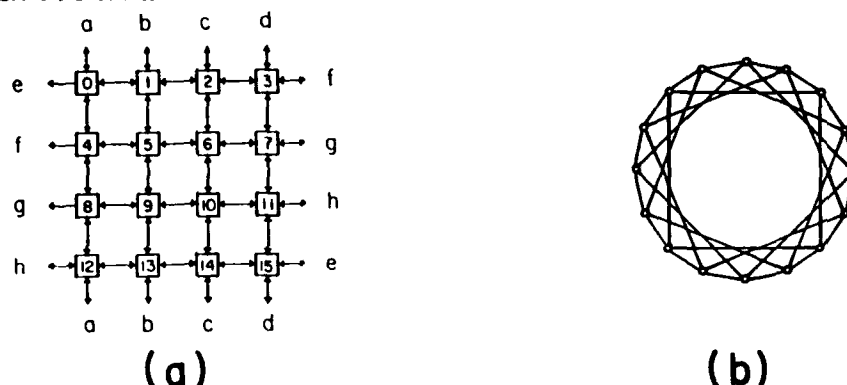


Fig.3. (a) A NN mesh for 16 PEs. (b) A redrawn of (a) along a ring.

3-2. Cross Over Interconnect

The use of ring array topology can also help implement various other SIMD interconnect schemes. A cross-over (CO) interconnect is topological equivalent to a perfect shuffle and has many applications to data permutation and sorting [3]. One form of CO interconnect is defined as

$$CO_S(i) = (i) \quad (3a)$$

$$CO_C(i) = (N - i - 1) \quad (3b)$$

where subscripts S and C denote a straightforward and a cross-over operations, respectively. For a ring array implementation, the PE sequence is first divided into two equal parts from the middle. The two parts are distributed one clockwise and other counterclockwise along the ring. The straightforward path does not perform permutation while cross-over path exchanges signals from the two opposite PEs (see Fig.4(a)). This exchange operation can easily

be implemented optically with a lens.

3-3. PM2I Interconnect

The plus-minus 2^j (PM2I) interconnect is an extension of NN interconnect. Unlike the NN network where only four routing paths are used, the PM2I employs $M = \log_2 N$ interconnect links defined as [2]

$$PM_{+j}(i) = (i + 2^j) \bmod N \quad (4a)$$

$$PM_{-j}(i) = (i - 2^j) \bmod N \quad (4b)$$

where $0 \leq j \leq M$. The PM2I interconnect for a ring array PEs is shown in Fig.4(b).

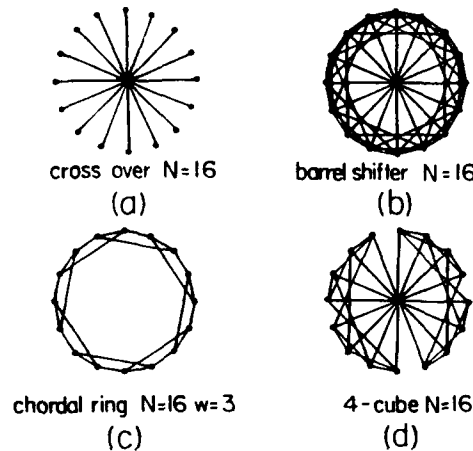


Fig.4. Four ring-array based SIMD interconnect networks.

3-4. Chordal Ring Interconnect

A chordal ring (CR) is described by the following two routing functions [2]:

$$CR_{ODD}(i) = (i + r) \bmod N \quad (5a)$$

$$CR_{EVEN}(i) = (i - r) \bmod N \quad (5b)$$

where $r \leq N/2$ is a positive and odd integer. In Fig.4(c), using a ring array topology, the routing paths of the CR network is shown which can easily be implemented optically by a screening followed by a rotation operations.

3-5. Hyper Cube Interconnect

While the PM2I is based on a modulo N addition/subtraction neighbor operation, the hyper cube (HC) is configured on a logical nearest neighbor base. The HC of N PEs is defined as [2]

$$HC_i(p_{N-1} \cdots p_{i+1} p_i p_{i-1} \cdots p_0) = p_{N-1} \cdots p_{i+1} \bar{p}_i p_{i-1} \cdots p_0 \quad (6)$$

where an output is different from its input by one bit if they are represented by binary addresses. In Fig.4(d), the routing paths of HC network of a ring array of PEs is shown.

IV. Network Architecture and Parameters

To perform a reconfigurable interconnect using a ring array of PEs, a general purpose optical architecture is shown in Fig.5. A ring cavity is used for device synchronization, in which K optical spatial light modulators (SLMs) are inserted at the middle image planes.

These SLMs can be used to select either an optical paths or points (PEs) in a path. To calculate the maximum allowable PEs using this network, the following parameters are defined: f , D , d , and λ denote the focal length, the diameter of the imaging lens, the diameter of the PE distribution ring, and the wavelength of optical source, respectively. Now, if the cross-talk-free practical minimum resolvable distance is assumed to be $p = (5\lambda f)/D$ where $\lambda = 0.6 \mu\text{m}$, $D = d = 0.5f = 1 \text{ cm}$, as many as $M = 5000$ PEs can be distributed along the ring. The use of ring cavity and imaging geometry inside not only lends itself for the use of point source (such as micro-lasers) but also provides a constant latency among all the PEs. Here, despite of different routing paths, all the data reach their destinations within the system's optical aberration time limit. Thus, even for an ultrahigh clock rate, say over 500 GHz, clock skew is not a problem.

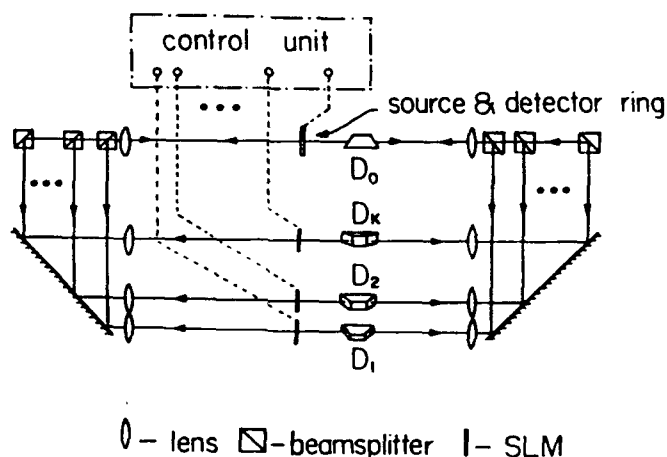


Fig.5. A reconfigurable ring-array based optical interconnect network.

V. Experimentals and Future Directions

To verify the proposed concepts, various proof-of-principle experiments were performed. As optical PE rings, masks were fabricated with 16 - 64 pixels uniform distributed along a ring with a diameter $d = 1 \text{ cm}$. Four cavity paths were implemented each employing two Dove prisms. Various SIMD routing operations including NN, CR, PM2I and HC were experimentally simulated. The results will be presented at the meeting.

The future direction of this work could be to study (1) multiple-ring configuration to increase the PE density and (2) optical fabrication using planar-integration of 3D optics approaches currently being investigated at AT&T Bell Lab [4].

The work is supported in part by a grant from the Air Force Office of Scientific Research (AFOSR88-0260).

VI. References

- [1] K.-H. Brenner and A. W. Lohmann, *Appl. Opt.* 27, 434 (1988).
- [2] K. Hwang and F. A. Briggs, *Computer Architecture and Parallel Processing*, (McGraw-Hill, New York, 1984).
- [3] J. Jahns and M. J. Murdocca, *Appl. Opt.* 27, 3155 (1988).
- [4] J. Jahn and A. Huang, *Appl. Opt.* 28, 1602 (1989).

A GUIDED-WAVE ACOUSTOOPTIC MATRIX ALGEBRA PROCESSOR MODULE*

A. Kar-Roy and C. S. Tsai

Department of Electrical and Computer Engineering
and Institute for Surface and Interface Science
University of California, Irvine, CA 92717.

Summary

Various architectures for implementing matrix algebra processors (MAPs) have been proposed and developed in bulk-wave optical systems⁽¹⁻⁷⁾ as well as in integrated optic (IO) devices.⁽⁸⁻¹⁰⁾ Bulk-wave optical systems have the advantage of an added dimension in implementing 3-D architectures over planar IO devices in which only 2-D architectures are possible. But the IO devices have the potential advantages in terms of drive power requirement, size, robustness, stability and planar technology for mass production. In this paper, we report on a guided-wave acoustooptic (AO) analog MAP module that is capable of implementing high-speed matrix-vector and matrix matrix multiplications. The architecture, the working principle, and a preliminary MAP constructed on a 1.0 X 10.0 X 28.0 mm³ Y-cut LiNbO₃ substrate to demonstrate the multiplication of a 4 X 4 matrix with a 4-element vector, are presented.

AO MAPs incorporate two fundamental properties of AO Bragg diffraction.⁽¹¹⁾ The first property is the optical beam steering into different directions by frequency-multiplexed acoustic waves. The second property is the modulation of the diffracted light intensity by the power of the RF signal that excites the acoustic wave. The first property was recently used to implement a 4 X 4 guided-wave AO space switch module using the architecture shown in Fig. 1.⁽¹²⁾ The working principle of this architecture can be briefly explained as follows. A light beam coupled into an input channel waveguide expands due to diffraction at the channel-planar waveguide interface and is then collimated by a large-aperture titanium-indiffused proton-exchanged (TIPE) lens.⁽¹³⁾ By placing the input channel waveguide off the lens axis, the resulting collimated beam is tilted from the lens axis in the propagation direction, incident and Bragg diffracted, and steered to different directions by varying the driving frequency of the surface acoustic wave (SAW) from a properly placed interdigital transducer (IDT). The input apertures of the output channel waveguides are placed in the back focal plane of a second large-aperture TIPE lens which collects and focuses the steered light beams. The output aperture of another input channel waveguide is placed further away from the lens axis than the first input channel waveguide, so that the resulting collimated and tilted light beam can be steered only by the SAW excited by another IDT aligned at an appropriate angle. The SAW from the second IDT steers the light beam from the second input channel waveguide to the same focal spots as the first IDT. In this way, the array size of the switch may be increased by adding more input and output channel waveguides. More recently, this guided-wave space switch architecture was evolved into a symmetric architecture and integrated with a hybrid beam expanding/collimating lens to provide improved performances.⁽¹⁴⁾

By applying the second property to the above AO space switch architecture, matrix-vector as well as matrix-matrix multiplications may be performed. The analog MAP operations are to be performed within the dynamic range where the relationship between the diffracted light intensity and the RF drive power to the IDT is linear. For convenience sake, we shall limit the explanations to matrix-vector operations of order 4. An example of such a multiplication of a matrix **A** with vector **B** to obtain the product vector **C** is expressed below,

* This work was supported in part by the NSF.

$$\begin{bmatrix} A_{11}(f_{11}) & A_{12}(f_{12}) & A_{13}(f_{13}) & A_{14}(f_{14}) \\ A_{21}(f_{21}) & A_{22}(f_{22}) & A_{23}(f_{23}) & A_{24}(f_{24}) \\ A_{31}(f_{31}) & A_{32}(f_{32}) & A_{33}(f_{33}) & A_{34}(f_{34}) \\ A_{41}(f_{41}) & A_{42}(f_{42}) & A_{43}(f_{43}) & A_{44}(f_{44}) \end{bmatrix} \times \begin{bmatrix} B_1(I_1) \\ B_2(I_2) \\ B_3(I_3) \\ B_4(I_4) \end{bmatrix} = \begin{bmatrix} C_1(O_1) \\ C_2(O_2) \\ C_3(O_3) \\ C_4(O_4) \end{bmatrix} \quad (1)$$

where f_{ij} designates the driving frequency for switching the light beam from input channel waveguide I_j to output channel waveguide O_i (Fig. 1). In the above expression, the f_{ij} terms in the brackets denote the corresponding RF frequencies of the SAWs which the matrix coefficients A_{ij} modulate, the I_j terms in the brackets denote the corresponding input channel waveguides through which the modulated optical vector elements are fed in parallel, and the O_i terms in the brackets denote the corresponding output channel waveguides through which the product vector elements C_i come out in parallel. All the sixteen frequencies in the above example are multiplexed and applied to their respective IDTs at the same time. For example, transducer S_2 which is dedicated to switch the light beam from I_2 , is multiplexed by four frequencies f_{i2} ($i = 1, 2, 3$, and 4). A schematic representation of the operation is illustrated in Fig. 2. Let τ_{AO} be the reconfiguration time of the integrated AO MAP module, which is the time taken by all the frequency-multiplexed SAWs fed at the same time to overlap all the incoming light beams in the common AO interaction region, and let τ_L be the modulation speed of the input light beams. All the product vector elements C_i are obtained instantaneously, after instantaneous N^2 multiplications and $N(N-1)$ summations for an $N \times N$ matrix, after time τ_{AO} . This module performs n such matrix-vector multiplications in time $\tau_{AO} + n\tau_L$ when the matrix is a constant and the vector varies, and in time $n(\tau_{AO} + \tau_L)$ when the matrix also varies. It is to be noted that since τ_{AO} is in the order of one microsecond, τ_L can be much smaller than τ_{AO} . Also, in comparison to other high-speed AO architectures capable of computing in the reconfiguration time τ_{AO} of their respective modules,⁽⁴⁾ the τ_{AO} of this integrated AO MAP module is significantly smaller as the optical beams overlap instead of being space-multiplexed, and hence reducing significantly the equivalent light beam aperture. This integrated AO MAP module can also facilitate iterative computations.

The matrix-matrix multiplication is a simple extension of the matrix-vector multiplication just described. It can be shown that the entire matrix multiplication of two $N \times N$ matrices is completed after a time of $\tau_{AO} + N\tau_L$. Since $\tau_L \ll \tau_{AO}$, this operation is quite fast and the speed of this IO module is comparable to other proposed high-speed bulk optic 3-D processors which can compute in time τ_{AO} of their respective modules.⁽⁷⁾

The integrated AO MAP architecture just described has been implemented on a $1.0 \times 10.0 \times 28.0 \text{ mm}^3$ Y-cut LiNbO_3 substrate. This IO module, consisting of titanium-indiffused (TI) channel-planar-channel composite waveguides,⁽⁹⁾ large aperture TIPE planar waveguide lens pair,⁽¹³⁾ and multiple tilted SAW transducers, was fabricated using established techniques.⁽¹⁰⁾ The two pairs of SAW (IDTs) had the center frequencies of 320 and 504 MHz. Planar micro-Fresnel lens arrays were used to facilitate efficient edge-coupling of light beams into the input channel waveguides.⁽¹⁵⁾ Photoresist phase-shift gratings were used to realize these planar lens arrays. Different combinations of the micro-Fresnel lenses were used to obtain different combinations of the elements of the vector. A more desirable way would have been to butt-couple the light beams from high

modulation speed diode laser arrays. Fig. 3 shows an example of the multiplication operation of a (4 X 4) matrix with a 4-element vector performed at a wavelength of 6328Å. The light beam intensities at the four output channel waveguides, obtained after the matrix and vector elements as illustrated in Fig. 3 were fed in the module, were imaged on a CCD array.

In summary, a new guided-wave AO MAP architecture is proposed. To the best of our knowledge, the resulting analog MAP module is capable of performing the fastest AO matrix-vector and matrix-matrix multiplications in an IO module. Experimental verification was done in an IO module realized on a Y-cut lithium niobate substrate consisting of channel-planar-channel composite waveguide, TIPE planar waveguide lens pair, and multiple tilted-SAW transducers.

References.

- [1]. J. W. Goodman, A. R. Dias and L. M. Woody, *Optics Letters*, **2**, 1-3, (1978).
- [2]. H. J. Caulfield, W. T. Rhodes, M. J. Foster and S. Horovitz, *Optics Communications*, **40**, 86-90, (1981); H. J. Caulfield, J.A. Neff, and W. T. Rhodes, *Laser Focus*, **19**, 100, (1983).
- [3]. R. A. Athale and W. C. Collins, *Applied Optics*, **21**, 2089-2090, (1982).
- [4]. D. Casasent, J. Jackson and C. Newman, *Applied Optics*, **22**, 115-124, (1983).
- [5]. W. T. Rhodes and P. S. Guilfoyle, *Proc. IEEE*, **72**, 820-830, (1984).
- [6]. K. Wagner and D. Psaltis, *Optics Communications*, **52**, 173-177, (1984).
- [7]. D. S. Kalivas, G. Albanese and A. A. Sawchuk, *Optics Letters*, **13**, 291-293, (1988).
- [8]. C. M. Verber, *Optical Engineering*, **24**, 19-25, (1985).
- [9]. C. S. Tsai, D. Y. Zang and P. Le, *Appl. Phys. Lett.*, **47**, 549-551, (1985).
- [10]. P. Le, D. Y. Zang and C. S. Tsai, *Applied Optics*, **27**, 1780-85, (1988).
- [11]. C. S. Tsai, *IEEE Trans. Circuits Syst.*, **CAS-26**, pp. 1072-1098, (1979).
- [12]. C. S. Tsai and P. Le, To be published.
- [13]. D. Y. Zang and C. S. Tsai, *Appl. Phys. Lett.*, **46**, 703-705, (1985).
- [14]. A. Kar-Roy and C. S. Tsai, To be published.
- [15]. P. Le, D. Y. Zang and C. S. Tsai, *Optics Letters*, **14**, 889-891, (1989).

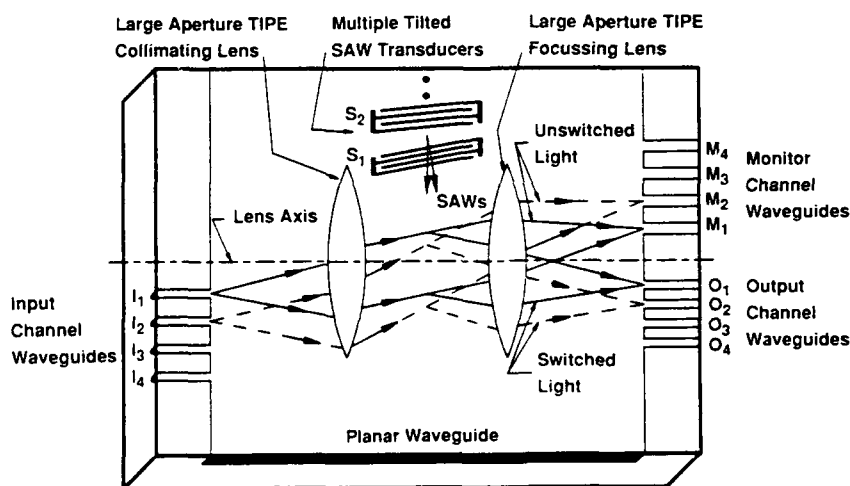


FIG. 1: ARCHITECTURE OF THE ACOUSTOOPTIC MATRIX ALGEBRA PROCESSOR MODULE USING A 4 X 4 NONBLOCKING INTEGRATED ACOUSTOOPTIC SPACE SWITCH (Note : Not To Scale).

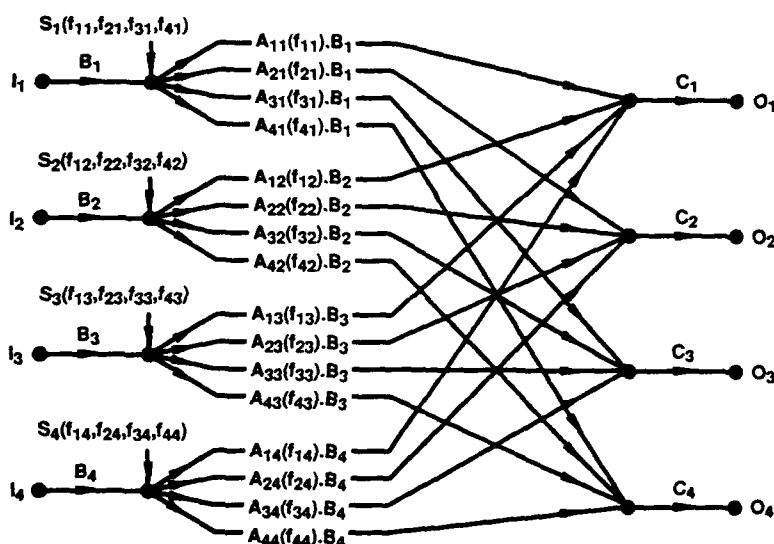


FIG. 2 : SCHEMATIC REPRESENTATION OF MATRIX-VECTOR MULTIPLICATION BY GUIDED-WAVE MULTI-FREQUENCY ACOUSTOOPTIC BRAGG DIFFRACTIONS IN THE INTEGRATED OPTIC MATRIX ALGEBRA PROCESSOR MODULE.

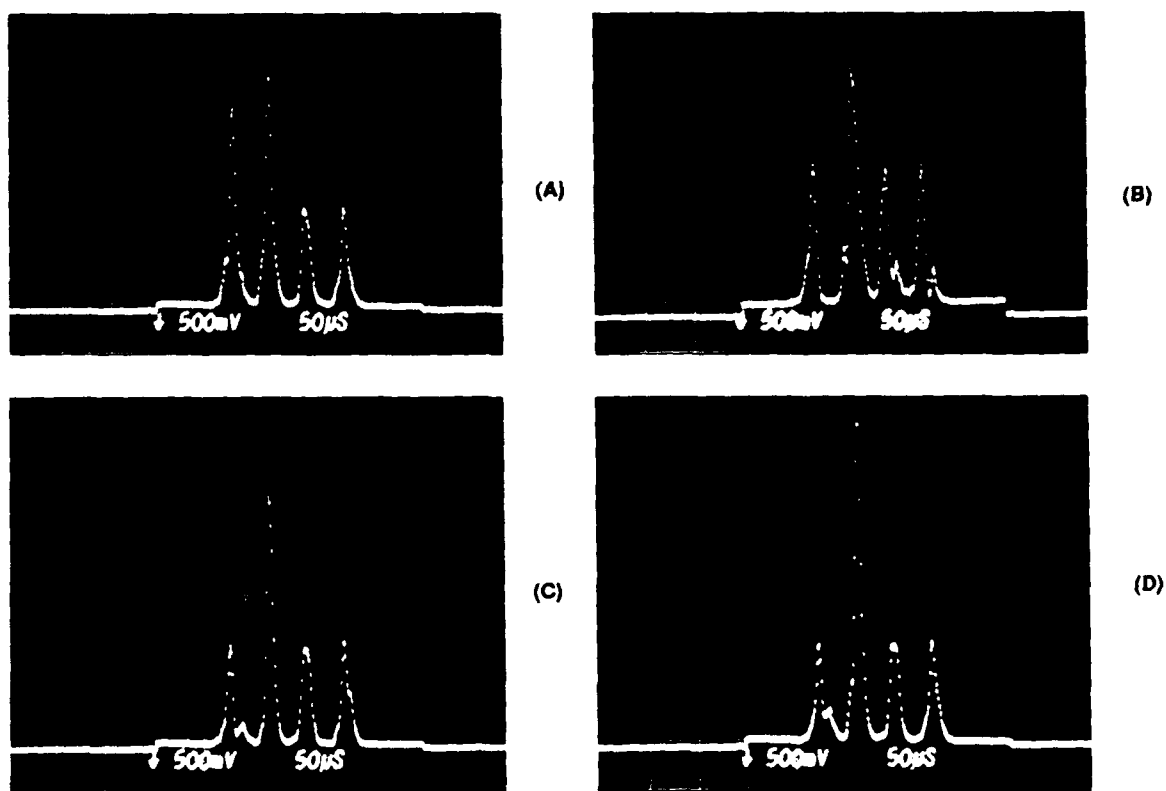


FIG. 3 : EXAMPLES OF PRODUCT VECTOR OF MATRIX-VECTOR MULTIPLICATIONS OBTAINED IN THE OUTPUT CHANNEL WAVEGUIDES. IN THE ABOVE PICTURES, THE CHANNEL WAVEGUIDE OUTPUTS FROM LEFT TO RIGHT DENOTE THE CORRESPONDING PRODUCT VECTOR ELEMENTS FROM TOP TO BOTTOM OF THE ILLUSTRATIONS BELOW. (26 μm Per Division)

$$\begin{aligned}
 \text{(A)} \quad \begin{bmatrix} 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 5 \\ 2 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} 4 \\ 5 \\ 2 \\ 2 \end{bmatrix} & \text{(B)} \quad \begin{bmatrix} 0 & 0 & 3 & 0 \\ 0 & 0 & 5 & 0 \\ 3 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} &= \begin{bmatrix} 3 \\ 5 \\ 3 \\ 3 \end{bmatrix} & \text{(C)} \quad \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} 2 \\ 5 \\ 2 \\ 2 \end{bmatrix} & \text{(D)} \quad \begin{bmatrix} 0 & 2 & 0 & 0 \\ 0 & 5 & 0 & 2 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 \end{bmatrix} \times \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} &= \begin{bmatrix} 2 \\ 7 \\ 2 \\ 2 \end{bmatrix}
 \end{aligned}$$

A 4x4 Photorefractive Reconfigurable Interconnect Using Laser Diodes

Arthur E. Chiou and Pochi Yeh
 Rockwell International Science Center
 1049 Camino Dos Rios
 Thousand Oaks, CA 91360
 Phone: (805)373-4464
 FAX: (805)373-4775

SUMMARY

Reconfigurable optical interconnection capable of partial or full broadcasting plays a key role in optical computing and optical neural network. The interconnection can be implemented using optical matrix-vector multiplication [1]. The reconfigurability is achieved by changing the interconnection matrix written on a spatial light modulator (SLM). For a one-to-one permutation link of an array of N sources to an array of N detectors, such an approach suffers a $1/N$ intrinsic fanout loss [1, 2].

Recently, we have proposed a new technique based on energy transfer in photorefractive dynamic holograms to achieve reconfigurable optical interconnection with a very high energy-efficiency [3, 4]. Using an argon ion laser, a photorefractive barium titanate crystal, and a 10×10 mask with a variable aperture, we have demonstrated a 1-to-100 selective broadcasting with energy efficiency of about 10% independent of the number of connected channels [4].

In this paper, we report the demonstration and characterization of a 4×4 reconfigurable interconnection using two laser diodes (780nm) and a ferroelectric liquid crystal spatial light modulator (FLCSLM) in conjunction with a photorefractive barium titanate crystal. Specifically, we compare the energy efficiency and crosstalk of this new approach with those of the conventional approach.

Fig.1 shows the experimental arrangement which uses two laser diodes to generate two sets of beams, designated as the pump and the signal beams, each consists of four columns of beam stripes. The collimated output from each laser diode (Liconix, Diolite 800-780), which has an elongated oval intensity profile, is sampled by a mask with a rectangular aperture. The size of the aperture ($1\text{mm} \times 4\text{mm}$) is chosen to match the pixel of a FLCSLM (Model $10 \times 10\text{PM}/10 \times 10\text{P}$ from DisplayTech). Each beam is split into four components by beam splitters BS1 and BS2 to form the pump and the signal beams with intensity profiles as shown in the lower right.

The experimental layout for a 4×4 photorefractive reconfigurable interconnect is shown in Fig.2. The signal beam is transmitted through the SLM which carries the desired binary interconnection pattern prescribed by a personal computer. Both the signal and the pump beams are Fourier transformed by identical lenses (focal length = 50cm) and the two transformed beams meet inside a photorefractive barium titanate crystal located at the back focal plane of the lenses. Diffraction of the beams from the photorefractive dynamic hologram resulting in an efficient energy transfer from the pump to the signal beams [5]. Energy loss (such as fanout loss and SLM insertion loss) suffered by the signal beam is thus compensated by the photorefractive gain which can be much higher than the loss. Shift invariance property of Fourier transform ensures maximum overlap of the two beams inside the crystal, and hence an efficient energy transfer independent of the interconnection pattern [4]. Examples of the signal beam carrying different interconnection patterns are shown in the upper right. The amplified signal beam passes through a second Fourier transform lens in series with a

line-image represent a spatial integration of all the signal beams through the corresponding row of windows on the SLM to a specific output channel.

With a detector positioned at each output channel, we measure the signal and the crosstalk by turning on each individual SLM-window (pixel), one at a time, while keeping all the other pixels at the "OFF" state. The experimental results are illustrated in Fig.3. In this specific example, the detector is positioned at the second channel (from the top), and the output from the detector is recorded as each one (and only one) of the 4x4 windows is turned on. The lower trace on the oscillogram represents the crosstalk (i.e. when none of the window in the second row is "ON") due to the poor contrast ratio (~15:1) of the SLM at 780nm. The upper trace represents the sum of the signal (through one of the window in the second row which is on) and the crosstalk from the leakage through all the other "OFF" channels.

The energy efficiency E_{ij} and average signal-to-crosstalk ratio X_{ij} are defined as

$$E_{ij} = \frac{\text{optical power received by detector "i" from source "j" through window "ij"}}{\text{total power transmitted by source "j"}}$$

$$X_{ij} = \frac{\text{optical power received by detector "i" from source "j" through window "ij"}}{\text{average optical power received by detector "i" when all the windows in the "i"th row are off}}$$

The experimental results are given in Table 1. In Table 2, the energy loss/gain in each optical element of a 4x4 reconfigurable interconnect are compared for the convention approach and the new approach using photorefractive holograms.

For the experimental configuration described above, the energy efficiency is mainly limited by the extremely inefficient way of matching the laser output to the SLM. Other factors that reduce the energy efficiency include non-optimum intensity ratio (of the pump and the signal beams) and relatively weak coupling strength ($\Gamma \sim 3$) of the photorefractive crystal at 780nm. The signal-to-crosstalk ratio is limited by the poor contrast (~ 15:1) of the SLM at 780nm. Approaches to improve both the energy efficiency and signal-to-crosstalk ratio as well as the scalability and limitation of this technique will be discussed.

This work is supported by DARPA/AFOSR under contract F49620-90-C-0006

REFERENCES:

- [1] A.A. Sawchuk, B.K. Jenkins, C.S. Raghavendra, and A. Varma, "Optical Crossbar Networks," IEEE Trans. Comput. C-20, No.6, 50-60 (1987).
- [2] J.W. Goodman, "Fan-in and Fan-out with Optical Interconnections," Optica Acta, 32(12), 1489-1496 (1985).
- [3] P. Yeh, A. Chiou, and J. Hong, "Optical Interconnections Using Photorefractive Dynamic Holograms," Appl. Opt. 27(11), 2093-2096 (1988).
- [4] A. Chiou and P. Yeh, "Energy Efficiency of Optical Interconnections Using Photorefractive Holograms," Appl. Opt. 29(8), 1111-1117 (1990).
- [5] P. Yeh, A. Chiou, J. Hong, P. Beckwith, T. Chang, and M. Khoshnevisan, "Photorefractive Nonlinear Optics and Optical Computing," Opt. Eng. 28(4), 328-343 (1989).

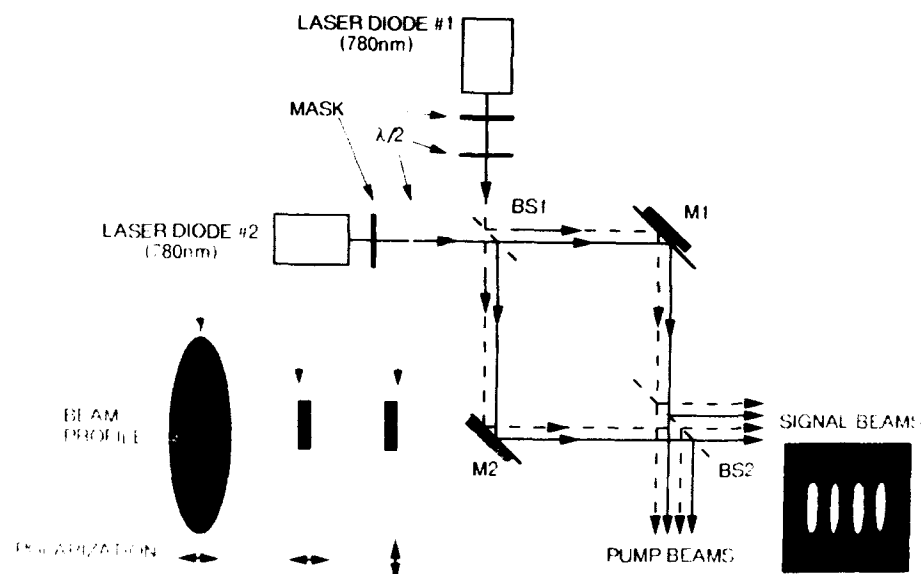


Fig.1 Experimental configuration for the generation of the pump and signal beams to match the spatial light modulator.

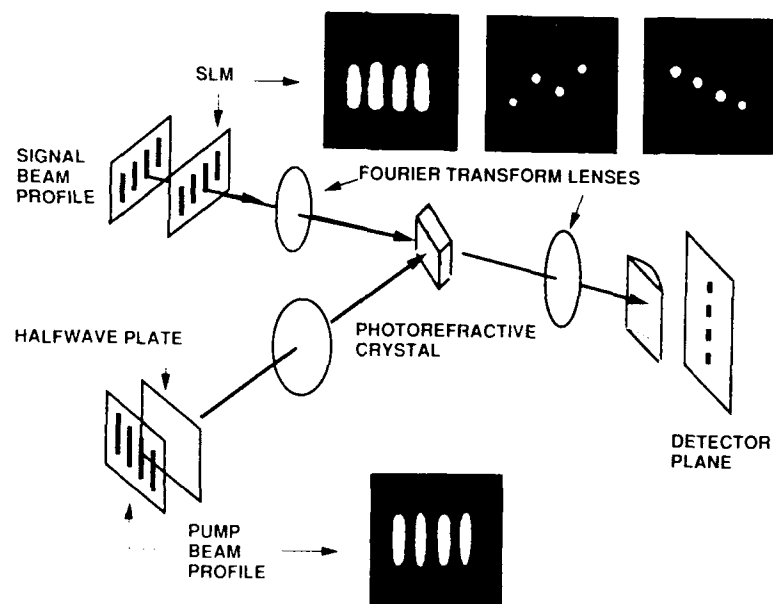


Fig.2 Experimental layout for a 4x4 photorefractive reconfigurable interconnections

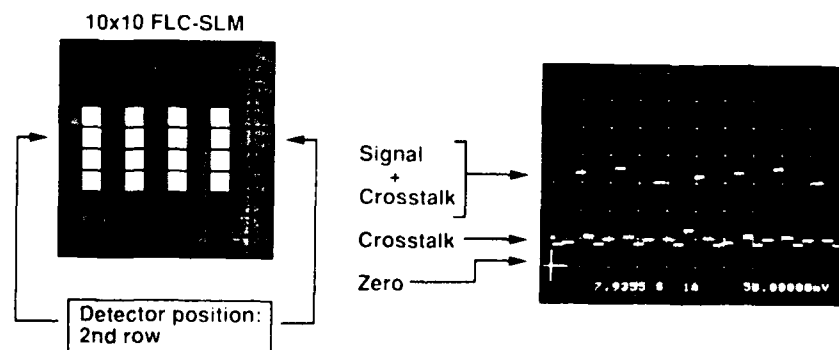


Fig.3 Signal and crosstalk measurement and the experimental result for 4x4 interconnections

ENERGY EFFICIENCY: E_{ij}

-22.0dB	-20.3dB	-22.4dB	-21.3dB
-22.6dB	-21.6dB	-21.2dB	-20.6dB
-22.7dB	-21.0dB	-22.1dB	-19.9dB
-25.8dB	-20.5dB	-22.5dB	-20.1dB

AVERAGE SIGNAL-TO-CROSSTALK RATIO: X_{ij}

1.8dB	3.5dB	1.4dB	1.8dB
-0.4dB	0.6dB	1.0dB	1.5dB
-0.4dB	1.3dB	0.2dB	2.4dB
-0.6dB	4.7dB	2.7dB	5.2dB

TABLE 1 Energy efficiency and average signal-to-crosstalk ratio: experimental results.

OPTICAL ELEMENT	ENERGY LOSS/GAIN	
	Conventional	Photorefractive
* FANOUT + MASK (to match SLM)	-9.6dB	-9.6dB
* BEAM SPLITTER (for pump beams)	<i>none</i>	<i>-3.0dB</i>
* NEUTRAL DENSITY FILTER (to protect SLM)	-8.5dB	-8.5dB
* SLM INSERTION LOSS	-5.5dB	-5.5dB
* FANOUT LOSS	-6.0dB	-6.0dB
* PHOTOREFRACTIVE GAIN	<i>none</i>	<i>+12dB</i>
* COUPLING INTO DETECTOR (aperture to reduce crosstalk)	-1.2dB	-1.2dB
NET ENERGY EFFICIENCY	-30.8dB	-21.8dB

TABLE 2 Comparison of energy loss/gain in each optical element of a 4x4 the optical interconnect using conventional approach and one using photorefractive hologram.

A compact photorefractive joint transform correlator for industrial recognition tasks

H.Rajbenbach, S.Bann and J.P.Huignard

Thomson-CSF, Laboratoire Central de Recherches
Domaine de Corbeville, 91404 ORSAY Cédex, France

Technology advances in solid state lasers, spatial light modulators and nonlinear optical materials are centrally important for the construction of optoelectronics processors that combine the massive interconnectivity and parallelism of optics with the accuracy and flexibility of digital electronics. In pattern recognition applications, hybrid optical-digital approaches in which optics performs correlation operations and electronics processes the output correlation plane for classification have already been demonstrated⁽¹⁻²⁾. Today, the performances of semiconductor lasers, diode-pumped YAG lasers, two dimensional liquid crystal light modulators and photorefractive materials allow the introduction of compact and more flexible optical hardware in optoelectronic processors. In this paper, we present a compact and reconfigurable multichannel joint transform optical correlator designed and constructed for industrial recognition applications. The principle of operation is shown in Fig.1. The object to be identified $S(x,y)$ is display on one half of the input scene. The other half of the input, allocated to the reference $R(x,y)$ is split in N subarrays, or channels, each containing a reference object or a calculated version of reference object. The sum $R(x,y) + S(x,y)$ is Fourier transformed and the spectrum is recorded in a dynamic holographic medium. The complex light field produced by reading out the joint-transform power spectrum contains the cross-correlation component $R(x,y) \otimes S(x-2a, y)$, where $2a$ is the separation between signal and reference and \otimes denotes the correlation operation⁽³⁾. The identification is performed by detecting the position and relative intensities of the correlation peaks in the corresponding subarrays of the output plane.

A schematic diagram of the optical implementation for a limited number of channels is shown in Fig.2. The input scene, loaded from a CCD video camera to a spatial light modulator contains four reference images and one unknown object. The spatial light modulator is a 320×264 pixel liquid crystal device with $80 \mu\text{m}$ pixel pitch. It modulates in polarisation a mini - 90 mW - CW - intracavity-doubled 532 nm wavelength diode pumped YAG laser. The Fourier transform is recorded in a 1 mm thick $\text{Bi}_{12}\text{SiO}_{20}$ (BSO) photorefractive crystal. This crystal allows the recording of an index modulation proportional to the incident intensity pattern of the data spectrum⁽⁴⁻⁵⁾. It operates in a high diffraction efficiency regime, with an externally applied transverse electric field $E_0 \sim 3 \text{ kV/cm}$. The average fringe spacing

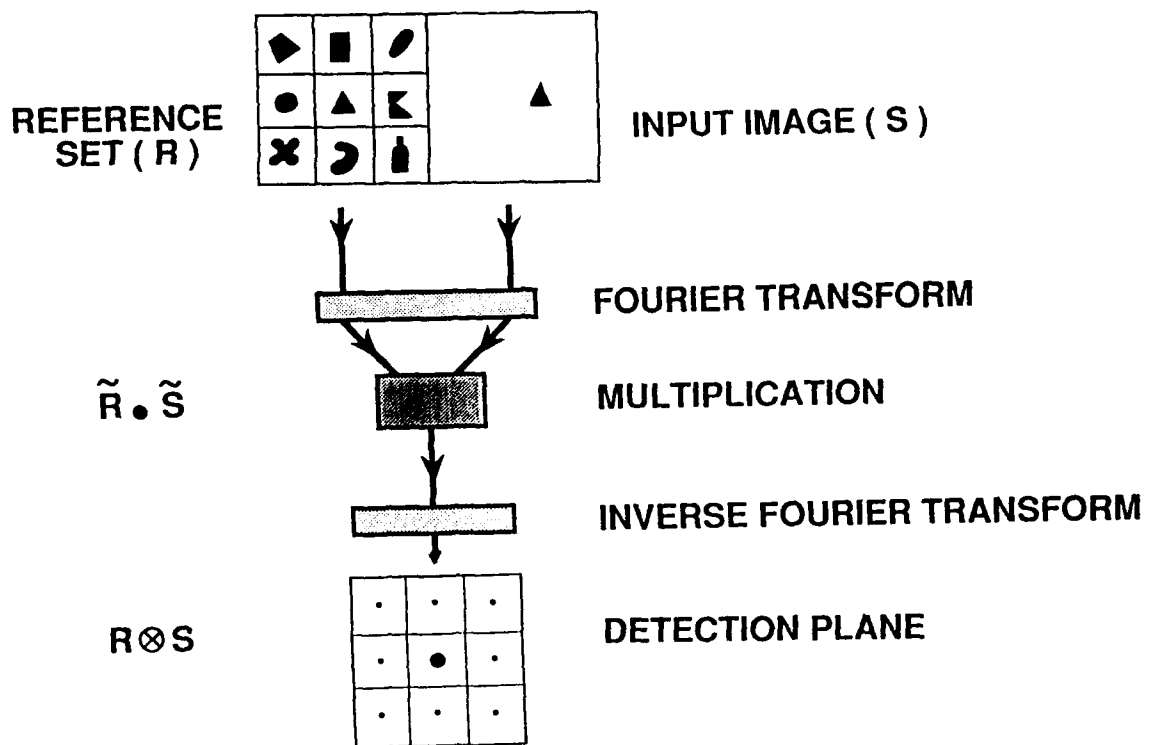


Figure 1 : Principle of operation of a Joint transform correlator

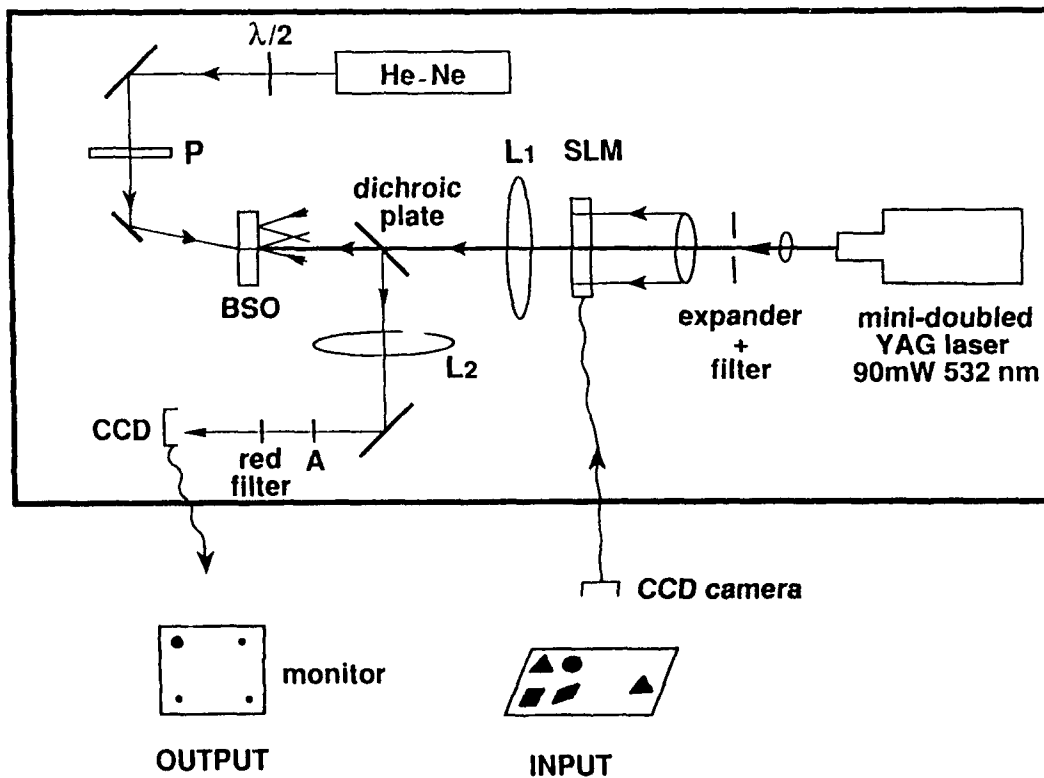


Figure 2 : The layout of the multichannel correlator

associated with the interference between the reference $R(x,y)$ and the input $S(x,y)$ is adjusted for optimum diffraction efficiency $\Lambda = 20 \mu\text{m}$ ⁽⁶⁾. The intensity in the BSO crystal is 2.25 mW and yields a response time shorter than the video period (30 msec). A HeNe laser beam ($\lambda = 633 \text{ nm}$) is used for readout of the filter. Its direction corresponds to the correct Bragg incidence for the average fringe spacing Λ . Furthermore, the crystal orientation (input face 1 10) leads to a diffracted beam whose polarization is rotated 90° with respect to the incident polarization⁽⁴⁾. The dc term is filtered out and high signal-to-noise ratio in the detection plane is obtained by proper orientation of polarizer P and analyser A. Finally a CCD sensor in the Fourier plane of the second lens L_2 records the correlation plane displayed on the output monitor. The whole optical system is only about 1 meter long, 0.5 meter wide and 0.2 m high. Its size can be further reduced with the use of low power semiconductor laser diodes in place of the HeNe laser.

Typical experimental results are shown in Fig. 3. The reference set consists of four objects, typically 15-20 mm across. It is displayed on the left side of the SLM. The right side of the SLM contains the input image. The output CCD sensor is divided into four subarrays, each allocated to the cross-correlation detection of one of the reference objects. The subarray containing the brightest correlation peak determines the object class. For the square and rhombus (look alike objects), note the presence of a weak correlation peaks in two channels simultaneously.

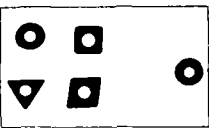

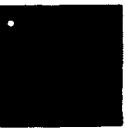
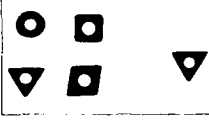


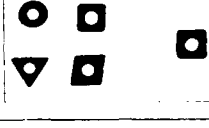


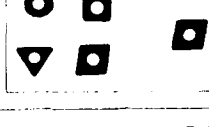


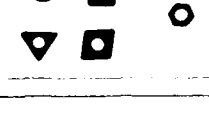


Reference	Input	Output
		
		
		
		
		

Figure 3 : The output correlation plane for different input objects

The position of the unknown object is given by the location of the correlation peaks in the output CCD sensor subarrays ; when the object moves, the average fringe spacing (Λ) in the BSO is modified in real time, which in turn produces a change in diffraction direction of the readout laser.

The multichannel correlator is flexible and extensible. New objects can be introduced in real time by modifying the reference subarrays ; there is no need to modify the optical hardware. When an unknown object is presented to the camera however, correlation peaks arise in few channels simultaneously, according to the similarity of this unknown object features with the reference objects. In the example shown by the lower picture in Fig.3, a screw bolt input generates weak correlation peaks in three of the four channels.

For practical applications, electronic pre-and post processing are needed to improve the performances of the multichannel optical correlator. Preprocessing consists of replacing the reference set of objects by computer-calculated invariant filters⁽⁷⁻⁸⁾ When optimum filters⁽⁸⁾ are displayed as the input references, the input image distortions (rotation, scale, background noise) do not affect the correlation peak intensities substantially. Electronic post processing can be fast because it operates on a limited amount of data, as low as 3 per channel (two for the position, one for the intensity of the correlation peak). In such conditions it is attractive to consider the practicability of unsupervised learning or neural-like techniques to further improve the performances of this processor.

In conclusion, we have designed and constructed a compact multichannel updatable joint transform optical correlator for use in an hybrid optoelectronic recognition system.

The authors acknowledge B.Loiseaux and Ph.Refrégier for enlightening discussions. This work was partly funded by the Commission of the European Communities under ESPRIT project 2288.

REFERENCES

- (1) D.Casasent, Proc Soc Photo-Opt Instrum. Eng. 528, 64 (1985)
- (2) F.Merkle and T.Lörch, Appl. Opt. 23, 1509 (1984)
- (3) C.S.Weaver and J.W.Goodman, Appl. Opt. 5, 1248 (1966)
- (4) J.White and A.Yariv, Appl. Phys. Lett. 37, 5 (1980)
- (5) B.Loiseaux, G.Illiaquer, JP.Huignard, Opt. Eng. 24, 144 (1985)
- (6) JP.Huignard, H.Rajbenbach, Ph.Refrégier and L.Solymar, Opt. Eng. 24, 586 (1985)
- (7) D.Casasent and D.Psaltis, Appl. Opt. 15, 1795 (1970)
- (8) Ph.Refrégier, Opt. Lett. 15, 854 (1990)

Optical Learning Systems

WA 8:30am-10:20am
Salon F

Bernard H. Soffer, *Presider*
Hughes Research Laboratories



LEARNING IN OPTICAL NEURAL NETWORKS

Demetri Psaltis

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

SUMMARY

In this paper we will review recent advances in training optical neural networks. We will focus on holographic implementations using photorefractive crystals [1]. The vast majority of learning algorithms in neural networks are based on some form of generalized "Hebbian Learning". With Hebbian learning the strength of the connection between two neurons is modified in proportion to the product (or possibly some other simple function) of the activation functions of the two neurons. These activation functions are typically the neuron response and error signals. The multiplicative Hebbian rule can be implemented if the hologram that connects two neurons is formed as the interference of two light beams generated by the two neurons. This simple and elegant method for training an individual connection can also form the basis for training large optical networks. There are several issues that need to be addressed however before such networks can be constructed. The following is a partial list of these issues, assuming photorefractives are selected as the synapse medium :

1. Architectures for Multiple Holographic Interconnections with 2-D and 3-D Media.
2. Recording Dynamics and Hologram Dynamic Range.
3. Suitable Devices for Neuron Implementation.

Clearly this is a partial list; issues such as accuracy, stability and alignment of large optical systems, hologram fixing, packaging, etc. must also be addressed before practical, large scale systems can be constructed. The issues listed are the minimum that needs to be solved before we can put together in the laboratory large scale adaptive optical networks. Work has been done in all three areas. In general, item 1 is the area that has been studied the most and item 2 is the area where most of the open questions remain.

It is relatively easy to design a holographic system to interconnect two points in space. It is considerably more complicated to design a system in which many pairs of points are simultaneously interconnected by the same hologram and the strength of each interconnection are independently set by recording the appropriate hologram (item 1). Many schemes have been proposed in the last few years to accomplish this and given the requirements of the problem at hand, one can select a suitable method. I have a bias towards using volume holograms because I believe that it will be much easier to demonstrate competitive advantages over analog VLSI if we can build optical networks that use 3-D holograms for specifying the interconnections.

The most difficult issue that we are faced with when we contemplate constructing a large optical system that can adapt to learn a difficult problem is the large number of examples that must be used and the huge number of training cycles that are typically required. In the optical implementation each training cycle is a new exposure on the crystal. When we are faced with the prospect of millions of learning cycles for some

of the problems we might want to learn with the optical system, it is clear that simply superimposing millions of holograms on the same crystal is not the answer. The finite dynamic range of the crystal will not allow us to do that. And yet, it is very likely that there is a set of weights that can be represented with sufficient accuracy by holographic gratings recorded in photorefractive crystals, that can implement the function we are interested in learning. The question then becomes:

Can we find optical learning algorithms that converge to such solution weight vectors that do not require a crystal dynamic range equal to the number of training cycles?

I believe that the answer to this question will turn out to be yes, however at this point we do not yet know the answer. If we can address this issue successfully then it should be possible to construct in the laboratory large scale optical learning machines using existing or emerging devices that work at least a thousand times faster than either current supercomputers or custom VLSI circuits programmed to learn the same task. I should point out that analog VLSI has a similar dynamic range problem to solve before it can be used for learning.

Finally, we need to consider the devices that simulate the neurons (item 3). These are likely to be more complex in an adaptive optical network than the neurons that are used in networks with fixed connections which are usually simulated by spatial light modulators. The reason for this is the need to produce more complex neuron activation functions than simple soft thresholding, including things such as adaptive thresholding, error function calculation, bidirectional capability, separate inhibitory and excitatory inputs, etc. Most of these capabilities can be implemented with conventional SLMs and additional optical interconnects, but in most cases this ends up being an exceedingly cumbersome solution when compared to an optoelectronic solution which can provide the same functionality with simple circuitry incorporated at each neuron site. Several approaches are being pursued for the implementation of these optoelectronic neuron arrays, including monolithic integration of circuits, detectors, sources and modulators in GaAs and mating silicon circuits and detectors with liquid crystal light modulators.

This research is funded by DARPA and AFOSR.

- [1] D. Psaltis, D. Brady, C. Gu, S. Lin, "Holography in Artificial Neural Networks", Nature, Vol. 343, Jan. 25 1990.

Hologram multiplexing using orthogonal phase codes and incremental recording

Yoshinao Taketomi, Joseph E. Ford, Hironori Sasaki, Jian Ma, Yeshayahu Fainman, Sing H. Lee

University of California, San Diego
Electrical and Computer Engineering Dept.
La Jolla, CA 92093

Jack Feinberg

University of Southern California
Physics Department
Los Angeles, CA 90007

Introduction

Photorefractive volume holography may prove useful for optical interconnection and data storage applications. However, the process of recording a set of uniform, high quality superimposed holograms normally involves a complicated recording procedure using a schedule calculated from the detailed material characteristics^{1,2}. A small error in material characterization (or change in the material characteristics) can result in highly nonuniform diffraction efficiencies. In this paper, we present a new incremental recording approach that relies only on an approximate knowledge of the materials characteristics. By avoiding long exposures, we avoid the high gain and fanning which tend to disrupt photorefractive performance. To achieve the highly repeatable recording necessary for this approach we use a set of orthogonal phase images for the reference beams. This choice minimizes readout of any unwanted images. The phase only reference images will be more reproducible when generated by a stationary phase spatial light modulator, compared to angular multiplexing. Compared to the simple sequential schedule of recording, the use of phase-coded reference beams and incremental recording of the holograms should produce brighter images with an improved signal-to-noise ratio.

Incremental recording approach

The object in superimposing photorefractive holograms is to end with a set of high quality equal diffraction efficiency holograms. Because each exposure partially erases all preceeding exposures, the recording schedule developed previously^{1,2} uses a long first exposure, followed by shorter and shorter exposures. The schedule depends on the precise material properties. In particular, it depends on the material's response times and the maximum attainable index modulation. If the values used in the schedule computation differ even slightly from the actual values, the final diffraction efficiencies will be highly nonuniform. In addition, making the long initial exposures introduces problems from the photorefractive gain. Coupling between the recording beams and fanning both tend to limit the maximum attainable index modulation to a value well below the theoretically calculated maximum. This becomes a dominant effect in crystals whose gain-length product is much greater than one.

In our approach, each of the N holograms are recorded with a series of incremental exposures, each very short compared to the material's response time. For low index modulations, the

slope of the writing curve is much steeper than that of the erasing curve (see Figure 1). The result is that some of the hologram written with the first increment remains after all $N-1$ other holograms are incremented. During recording, each image and reference pair is sequentially displayed, repetitively cycling through all N images. The holograms gradually increase in diffraction efficiency as each cycle is completed. The recording process will reach saturation when the growth rate equals the erasure rate.

The writing process can be approximated by an exponential rise with time constant τ_w to the maximum index modulation Δn_{\max} , and the erasing process by an exponential decay with time constant τ_e

$$\text{writing: } \Delta n(t) = \Delta n_{\max} (1 - e^{-t/\tau_w}) \quad (1)$$

$$\text{erasing: } \Delta n(t) = \Delta n e^{-t/\tau_e}$$

The requirement for equilibrium is that at the end of a recording cycle each hologram returns to its diffraction efficiency at the beginning of the cycle. This is true when the slope of the writing curve is $N-1$ times as steep as the slope of the erasing curve, or

$$\Delta n_o = \frac{d\Delta n_w}{dt} = (N-1) \frac{d\Delta n_e}{dt} \quad (2)$$

This requirement leads to a simple expression for the number of holograms which can be stored with a specified index modulation

$$N \approx 1 + \left\{ \frac{\Delta n_{\max}}{\Delta n_{\min}} - 1 \right\} \frac{\tau_e}{\tau_w} \quad (3)$$

where Δn_{\min} is the index modulation required to produce the minimum diffraction efficiency. An identical result was obtained from a more accurate computation which did not use the exponential rise approximation and which included energy coupling between recording and reference beams (but not fanning).

The result of Eq. 3 is the same as that predicted by the single scheduled exposure recording method in low gain media. However, because the incremental recording method avoids sustained exposures, it avoids the problems of low intensity modulation and noise generated by coupling and beam fanning. In addition, despite the multiple exposures and erasures, the total recording time can be shown to be almost half that for the scheduled recording approach. This is a result of avoiding the regions of the recording curve near saturation, where the effective sensitivity is much lower than at the foot of the curve. The exact value of the recording increment does not affect the result, providing that it is small compared to the writing response time. As it is increased to an appreciable fraction, the final hologram diffraction efficiencies will show some variation between the first and last of the N holograms.

The difficulty in implementing this recording approach is that the fringe pattern of the recording beams must be reproduced exactly in every cycle. A difference of more than a fraction of a fringe will prevent the recorded increments from reinforcing each other, disrupting the recording process. If angular multiplexing is used, then the crystal must be rotated to within a few milliradians of its previous position; a difficult tolerance to achieve. However, a fixed spatial phase modulator can provide a phase-coded reference beam³ with fast switching and a high degree of reproducibility. Among the several phase coding methods which we have investigated, we have found random phase coding useful for increasing the number of holograms which can be stored with a given space-bandwidth product of the phase coder. Crosstalk using random

phase coding becomes significant when the number of stored holograms grows too large. Alternately, deterministic phase codes⁴ can be used to improve the signal to noise ratio. In the following, we show that using a simple orthogonal phase code causes complete extinction (to first order) of any unwanted images.

Orthogonal phase coding

We want to store N images in the crystal. Let A_i be the electric field amplitude of the i^{th} image (with $i=1$ to N) at the crystal surface. The corresponding reference beam for the i -th image will be generated by Fourier transforming a point source array into the crystal. The point sources are produced by illuminating a lenslet array with light from a one-dimensional spatial light modulator, so the the reference image field amplitude at the crystal surface is given by a set of phased plane waves

$$\sum_{j=1}^M R_{ij} = \sum_{j=1}^M e^{j(k_j \cdot r + \phi_{ij})} \quad (4)$$

where \mathbf{k} is the propagation vector and ϕ_{ij} is a phase code given to the j -th pixel of the SLM with a total of M pixels. A one-dimensional reference array is used to avoid degeneracy arising from the cone of reference beam angles which satisfy the hologram's Bragg condition. The final refractive index variation Δn recorded in the crystal is proportional to $\sum_{i=1}^N A_i \sum_{j=1}^M R_{ij}^*$. Readout with the k^{th} reference image produces a reconstruction proportional to

$$R_{kj} \sum_{i=1}^N A_i \sum_{j=1}^M R_{ij}^* \quad (5)$$

where we assumed that the reference image pixel separation was large enough to satisfy the Bragg condition determined by the angular selectivity of the volume hologram and $F\#$ of the recording optics. If all the cross terms in Eq. (5) vanish, that is, if

$$A_i \sum_{j=1}^M R_{ij}^* R_{kj} = 0 \quad (i=1-N, i \neq k) \quad (6)$$

then the reconstructions of the undesired images destructively interfere to produce zero intensity, and a noiseless reconstruction of the k -th image will be obtained.

Equation 6 can be treated using matrix algebra to reduce the problem of finding optimum phase-codes to finding matrices U_{ij} which satisfy the relation $U_{ij} U_{ij}^* = E$, where E is the unit matrix. All row vectors of the matrix U_{ij} are then orthogonal with each other, and these vector sets can be used as phase code sets which will make all the cross terms in eq. (6) zero. In general, for an $M \times M$ matrix, M sets of orthogonal vectors with the length M exist. Therefore, an SLM with M continuous phase pixels can display M orthogonal phase codes.

Of course, with M amplitude-modulated pixels separated by more than the Bragg angle distinct reconstruction are trivially possible. However, there are several reasons to investigate phase-only addressing. A phase-only modulated reference image is necessarily light efficient. More importantly, phase-code addressing can provide simpler, quickly generated reference images. If some crosstalk is allowable, a closely spaced random phase-code can provide a simple reference mechanism. We are interested in developing deterministic phase-codes which allow a gradual trade-off between reconstruction signal-to-noise ratio and reference image complexity. Demonstrating that a simple phase-only reference image can produce minimal crosstalk

reconstructions is the first step towards that goal.

Conclusions

Experiments using orthogonal phase coding are now in progress. Preliminary results for images recorded in Lithium Niobate using a 16 bit binary code have shown good extinction between reconstructed images. More detailed results for both incremental recording and orthogonal phase codes will be presented at the conference. We expect to demonstrate significant improvements in superimposed hologram storage capacity of crystals of SBN:60 whose fanning behavior currently prevents optimum operation.

References

- [1] D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals", *Appl. Opt.* **27**, pp. 1752-1759 (1988).
- [2] J. E. Ford, Y. Taketomi, S. H. Lee, D. Bize, R. R. Neurgaonkar and S. Fainman, "Effects of applied voltage on holographic storage in SBN:60", *Proc. SPIE* **1148** (1989).
- [3] T. F. Krile, R. J. Marks II, J. F. Walkup, and M. O. Hagler, "Holographic representations of space variant systems using phase-coded reference beams", *Appl. Opt.* **16**, pp. 3131 (1977).
- [4] D. Z. Anderson and D. M. Lininger, "Dynamic optical interconnects: volume holograms as two-port operators," *Appl. Opt.* **26**, pp.5031-5038, (1987).

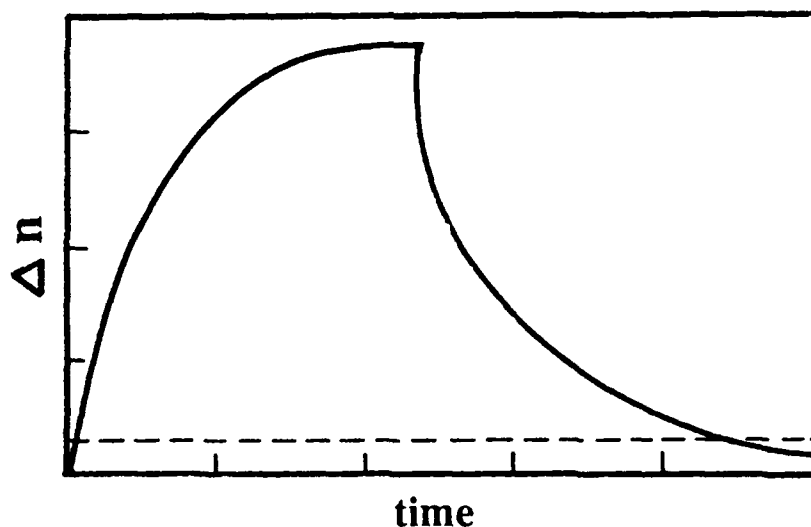


Figure 1: The writing and erasing curve for photorefractive hologram recording. For low index modulations (dashed line), the effective writing sensitivity is much higher than the effective erasing sensitivity.

Generalization in an Optical On-Line Learning Machine

John R. Wullert II, Eung Gi Paek and J.S. Patel
Bellcore
331 Newman Springs Rd
Red Bank, NJ 07701

Neural networks, characterized as a large number of highly interconnected simple processors, can be trained by varying the strength (weight) of the interconnections (synapses) between the simple processors (neurons). Several holographic optical systems have physically demonstrated this capability previously.^{[1] [2] [3] [4]} Since neural networks are trained by example rather than programmed with specific rules, they are likely to be able to generalize, or recognize patterns that do not exactly match those used for training. Such generalization is important in real world pattern-recognition problems where the size, orientation, position and background cannot be determined in advance.

In our Optical Multicategory Perceptron,^[2] we have demonstrated learning in an optical system. In this system, the interconnection weights are recorded as a volume hologram inside a photorefractive crystal. The volume nature of the crystal provides tremendous potential storage capacity, and this capacity has been exploited in allowing the system to learn to be invariant to modifications of the input patterns.

In these experiments, the system used was an optical implementation of the multicategory Perceptron algorithm, as described in ^[2]. The system is shown schematically in Fig. 1. The input to the system, in the form of a two-dimensional image, comes from the optical disc recorder as a video signal and is impressed upon a laser beam by the optically-addressed liquid crystal light valve. The resulting coherent image is spatially Fourier transformed by the lens L1 and then enters the photorefractive crystal. The image is then diffracted by the holograms stored in the crystal, which is equivalent to multiplying the input by the weight matrix, and the diffracted light is measured by the one-dimensional array of detectors. After thresholding, the outputs of the detectors are compared with the desired target values and the differences are the error signals. The error signals are used to activate individual elements in the one-dimensional liquid crystal modulator array, which generates beams that

interfere in the photorefractive crystal with the input image, modifying the stored gratings and thus the updating the weights. Then the next image is presented. This process is continued until all the images in the training set are identified correctly.

One method of allowing a system to deal with variations in the input patterns is to actually teach the system all the possible inputs it might see. This is not to be considered generalization, however, since the system would only be recognizing patterns that it had been taught, not correctly identifying previously unknown variations. This concept can also be viewed as the mapping of multiple input patterns into each output category, a capability that is of importance in learning systems that are intended to categorize rather than strictly identify. With this in mind, we investigated the ability of our optical learning system to learn multiple variations of the input patterns as being in a single category.

In these experiments, the system was taught to recognize up to three versions of each of up to eight characters. A sample of the input patterns and variations is shown in the inset of Fig. 2, while the rest of the figure shows the learning curve, i.e. the number of errors versus cycles through the data. The variations that were used included $\pm 25\%$ in scale of the characters and ± 90 degree rotation. The learning proceeded as described above, cycling through all three sets of the characters I-P and updating the weights until the zero-error condition was reached. In this case, each of the three variations of the characters was mapped to one particular output state (i.e. all I's = 10000000, all J's = 01000000, etc.).

As shown in the learning curve in Fig. 2, the system successfully learned to map the twenty-four characters into the eight output categories. Thus simple training offers one method of giving the system the ability to deal with input variations.

While the results presented above indicate it is possible to train the system to recognize

variations of input patterns, it is not desirable to have to do so, as it is impossible to predict all of the variations that might occur, and training any large number of variations would be tremendously time intensive. It is more desirable to have the system do some real generalization, where it can recognize patterns that are not used for training. We have demonstrated true generalization with our system, using methods similar to those discussed above.

In order to understand how our system would respond to variations in the input, we tested its ability, after learning characters of one orientation, to recognize those rotated at other angles. This experiment gives a measure of the inherent generalization of the system and provides a basis on which to compare our other results. For this experiment, the system was taught to recognize eight upright characters, and then tested with characters rotated 10 and 20 degrees away from the original. Copies of the characters used for training and testing as well as the results of this experiment (marked with diamonds) are shown in Fig. 3. The experiment involved teaching the system the upright characters and, after the system converged to the zero error state, testing with the rotated versions. The recognition rate was determined by averaging the number of characters incorrectly identified over 250 cycles through the data. In order to insure that the weights had not changed significantly, the original upright characters were tested for 100% recognition accuracy before and after each cycle through the rotated versions. The results show that the experimental recognition rate falls off rather quickly as the characters are rotated, with only 50% recognition accuracy after a rotation of only 10 degrees and, based on a linear extrapolation of the data, no recognition at angles greater than 25 degrees.

Additional data in Fig. 3, marked with plus signs, provide some further information about the generalization abilities of the system. The figure shows the average overlap over the eight characters, or average percentage of pixels that each of the rotated images has in common with the original, as a function of rotation angle from the training set orientation. These values were obtained by digitizing the input images and then dividing the number of pixels that each rotated image has in common with the original by the number of pixels lit in the original image. The results presented are the average of the percentage overlap for the eight characters used

in these experiments. At first it seems contradictory that the recognition rate actually falls off more quickly with rotation angle than the image overlap. This is, however, simply an indication that the overlap of an image with the correct original does not represent the entire recognition process. The amount the rotated image overlaps with all of the other images is important as well, since as the character is rotated, there is not only reduction in the overlap with the correct character, but also a potential increase in its overlap with the incorrect ones.

The final trace on Fig. 3, marked with the squares, shows the results of a digital computer simulation of the Perceptron algorithm on the digitized versions of the images used in the experiment. Again, the fall off in recognition with rotation angle is rather rapid. The agreement between this simulation and the experimental results is surprisingly good, given that the simulation simply modeled the Perceptron algorithm and did not attempt to take any of the system operational details into account. These results make it clear that rotation through relatively small angles causes a dramatic reduction in the recognition rate.

By combining the results of the two experiments described so far, however, it is possible to create a scheme to train the system with the information it needs to generalize. We accomplished this by training the system with both the original characters and the same characters rotated by relatively large angles, then testing it with angles in between. Specifically, the system was taught to recognize eight upright characters and eight characters rotated clockwise by 20 degrees. Then, the system was tested with characters rotated by 10 degrees. In this case, each iteration involved testing and training with the 0 and 20 degree characters and testing with the 10 degree versions. Despite not having been trained with the 10 degree characters, the system recognized them flawlessly, as shown by the learning curve in Fig. 4. The zero error condition indicates that all 24 characters are properly identified even though only sixteen of them were used for training. Digital computer simulations of the Perceptron algorithm produced identical recognition results. This generalization appears even more impressive when compared to the recognition results in Fig. 3, which indicate a 50% recognition at 10 degree rotation, making it clear that this training method produces a substantial improvement in recognition rates for

rotated images.

We have shown that it is possible, in fact relatively easy, to train our Optical Learning Machine to map multiple inputs into each output state, so that the system could correctly map each of twenty-four input patterns into one of eight output categories. This capability alone, however, does not solve the problem of generalization, since it is undesirable to have to teach the system all of the variations in advance. By training the system with a few selected variations of each input, however, we were able to demonstrate generalization, where the system could correctly identify images that had not been specifically included in the training set.

REFERENCES

1. D. Psaltis, D. Brady, and K. Wagner, *App. Opt.* 27, p. 1752 (1988).
2. E.G. Paek, J.R. Wullert II, and J.S. Patel, *Opt. Lett.* 14, p. 1303 (1989).
3. J. Hong and P. Yeh, *Digest of Topical Meeting on Optical Computing* (Optical Society of America, Washington, D.C., 1989), p. 307.
4. N.H. Farhat, *Proceedings of the IEEE International Conference on Neural Networks* (Institute of Electrical and Electronic Engineers, New York, 1988), Vol. 2, p. 365.

- 4 -

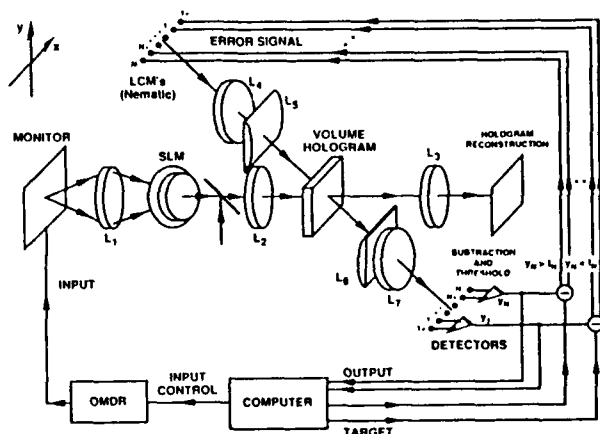


Figure 1. Schematic of the holographic Optical Learning Machine

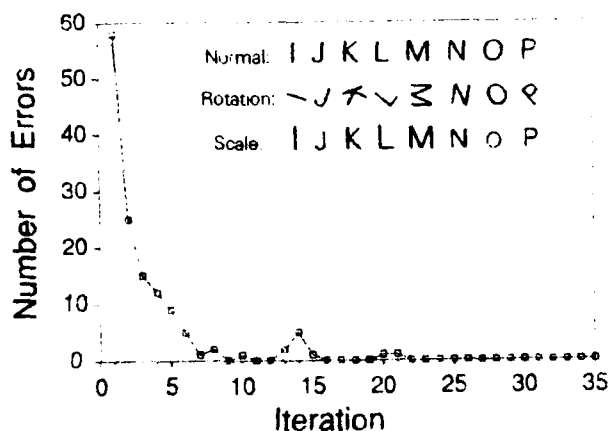


Figure 2. Input patterns (inset) and resulting learning curve when the system is taught to map multiple (24) inputs into eight output categories.

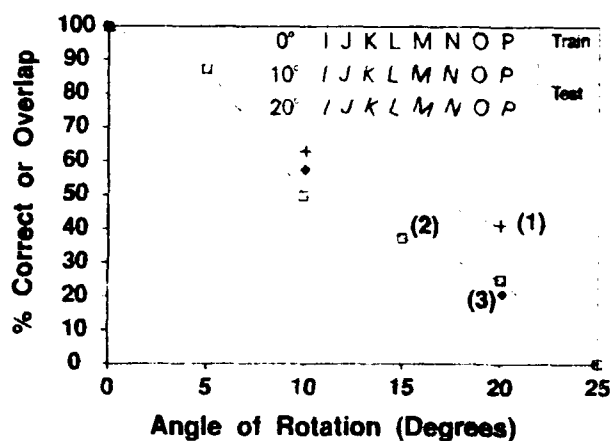


Figure 3. Three ways of measuring the inherent ability of the system to deal with input rotation: 1) The percentage of overlap between the upright and rotated characters; 2) Performance of a digital simulation of the Perceptron algorithm trained to recognize upright (zero degrees) characters as a function of input rotation angle; 3) Experimental recognition rate of the Optical Learning Machine as a function of rotation angle after being taught to recognize upright characters.

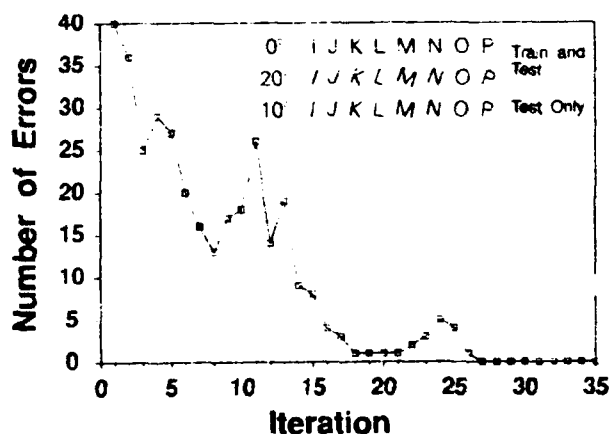


Figure 4. Data used for training (0 and 20 degrees) and testing (10 degrees) and the resulting learning curve demonstrating generalization, where the characters rotated by 10 degrees are properly recognized without having been taught.

Closed Loop Optical Disk Based Associative Memory

Mark A. Neifeld and Demetri Psaltis
Department of Electrical Engineering
California Institute of Technology 116-81
Pasadena, California 91125

Introduction

The optical disk is a simple computer addressable binary storage medium with very high capacity.^[1] More than 10^{10} bits of information can be recorded on a 12cm diameter optical disk. The natural two dimensional format of the data recorded on optical disk makes this media particularly attractive for the storage of images and holograms, while parallel access provides a convenient mechanism through which such data may be retrieved. Parallel access to data stored on optical disk has been shown to provide interesting solutions to problems in neural networks, database retrieval and pattern recognition.^[2] In this paper we will discuss a closed loop associative optical memory based on the optical disk. When presented with a partial or noisy version of one of the images stored on the optical disk, the optical system evolves to a stable state in which those stored images which best match the input are temporally locked in the loop.

System Description

The optical disk based associative loop is shown schematically in Figure 1. The system comprises an optical disk on which a number of stored images reside, a photorefractive crystal serving as a real time holographic storage medium, an input SLM for presentation of the association key and a one dimensional detector array followed by some simple electronics which generate the loop feedback as shown. The system operation is as follows. First, the input and reference illumination is "on" and the disk illumination is "off". A Fourier transform hologram of the input image is formed in the crystal with the reference beam as shown. The readout phase is initiated by first turning the reference and input illumination "off" and then illuminating the transmissive disk from below for one rotation. The output plane will contain the correlation between the input image and the illuminated portion of the optical disk. Since the photorefractive hologram is thick, it exhibits Bragg selectivity in the direction parallel to the disk tracks. This effect will cause only a single column of the 2D correlation pattern to be obtained in the output plane thereby resulting in the loss of horizontal shift invariance in the system. Fortunately, since images stored on the disk will appear to shift past the hologram field of view, the disk rotation can be used to recover horizontal shift invariance.^[3]

In order to measure the sequence of correlation columns which appear in the output plane as the disk rotates, a 1D detector array is used. By choosing the largest element of

the detected signal at any time and then further choosing the maximum such largest signal over one full disk rotation, the identity of that image on the disk which best matches the input is obtained. In order to realize the required Winner Take All (WTA) function over the 1D detector array, a custom analog VLSI chip is used. The detector array is shown in Figure 2. In Figure 2b we show an array of bipolar phototransistors which serve as variable current source inputs to the WTA circuit described by Mead.^[4,5] In our case, the current input at a given node is proportional to the intensity of the light falling on the corresponding phototransistor. The output of the WTA circuit is a voltage proportional to the logarithm of the largest such intensity. The output of the optical WTA detector is thresholded and delayed for one disk rotation so that stored in the feedback electronics, is a pulse whose temporal position indicates the location on the disk of the best match with the input image. Once such a pulse is in the loop, the readout phase is completed by closing the feedback path. In this phase, both the disk *and* reference illumination are controlled by the feedback signal so that the disk is read out once per rotation thereby retrieving the proper stored image and at the same time reinforcing the correct hologram in the crystal.

One attraction of the present system is its full 2D shift invariance. The Fourier transform hologram in the crystal provides vertical shift invariance while disk rotation provides shift invariance in the horizontal dimension. Since vertical position of the input image can be detected as the location of the "winner" in the optical WTA detector, this information can be stored for later retrieval. Horizontal position is represented as a temporal delay which in turn determines *when* the closed loop disk illumination will be pulsed. In this way a shifted version of one of the stored images can be retrieved from the system. Another attraction of this system is the locking or stable nature of the closed loop operation. Since both the disk illumination and the reference beam are pulsed together, the hologram stored in the crystal is reinforced during each disk rotation. This insures that the readout diffraction efficiency will remain stable and guarantees that correlation plane SNR does not degrade. This locking has a further advantage. If an incomplete or corrupted input is presented to the system, the loop will lock to a complete, uncorrupted association which in turn will reinforce the hologram in the crystal. In this way, the hologram of the distorted input is slowly corrected through the locking action of the loop.

Conclusions

In this paper we have described an optical disk based associative memory which takes advantage of the parallel access capabilities afforded by optical storage media. The capacity of this associative memory is given by the capacity of the optical disk image library and can exceed 10^4 , 1000X1000 pixel images. The retrieval time for recovery of a single image association is given by the disk rotation time and is approximately 10ms for conventional disk drives. A particularly attractive feature of this system is the full 2D shift invariance which occurs as a result of the marriage between the optical technology and analog VLSI based focal plane processing.

References

- [1] R. A. Bartolini, *et. al.*, "Review and analysis of optical recording media," *Optical Engineering*, Vol. 15, No. 2, March 1976.
- [2] D. Psaltis, *et. al.*, "Optical memory disks in optical information processing," *Applied Optics*, Vol. 29, No. 14, May 1990.
- [3] D. Psaltis, *et. al.*, "Image correlators using optical memory disks," *Optics Letters*, Vol. 14, No. 9, May 1989.
- [4] J. Lazzaro, *et. al.*, "Winner-take-all networks of $O(N)$ complexity," from *Advances in Neural Information Processing I*, pp.703-711, David S. Touretzky, ed. Morgan Kaufmann Publishers, San Mateo, Ca., 1989.
- [5] C. Mead, "Analog VLSI and Neural Systems," Addison-Wesley, Reading Ma., 1989.

Figures

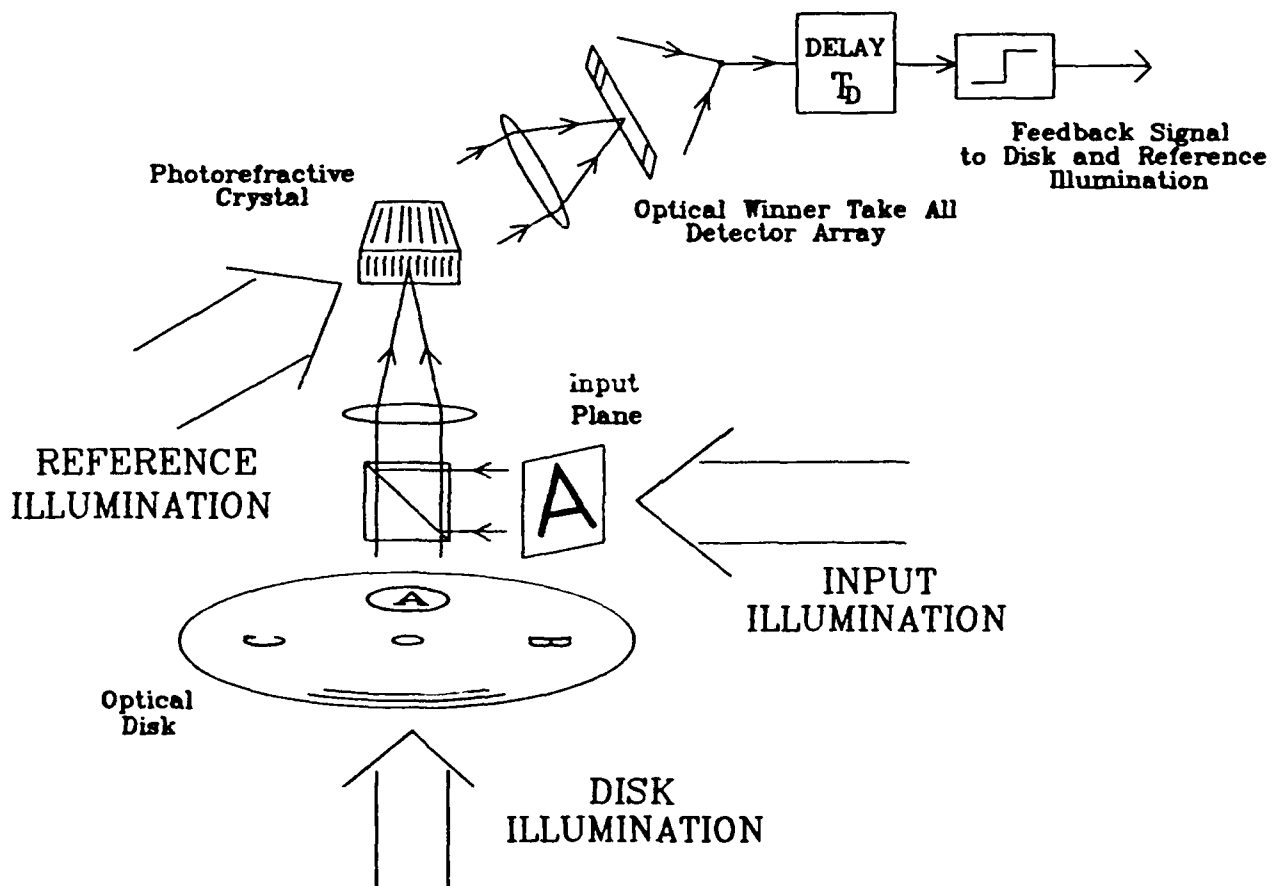
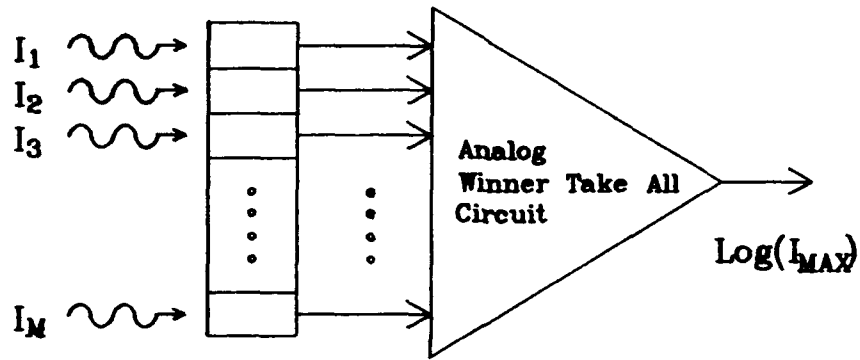
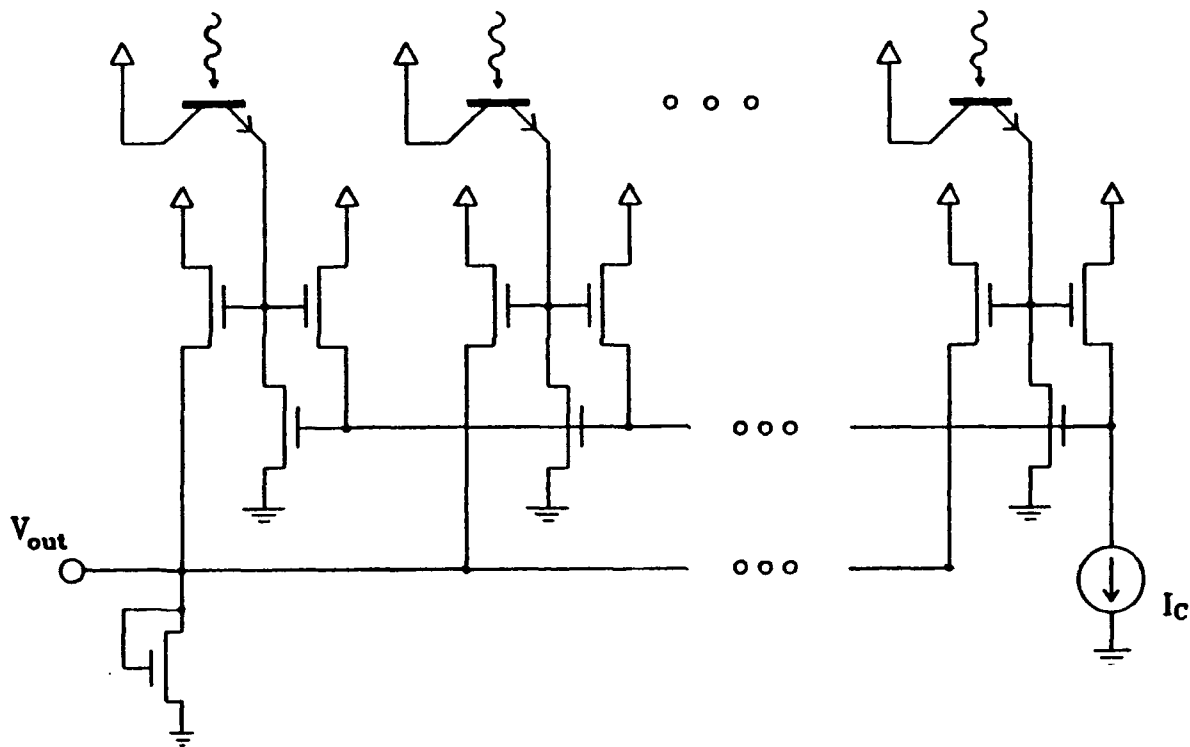


Figure 1 : Optical Disk Based Associative Loop.



(a)



(b)

Figure 2 : Optical Winner Take All Circuit
 (a) 1D detector array followed by electronic WTA circuit.
 (b) Circuit diagram for (a).

Competitive Optical Learning with Winner-Take-All Modulators

Kelvin Wagner and Tim Slagle

Optoelectronic Computing Systems Center, Campus Box 525
Department of Electrical and Computer Engineering
University of Colorado at Boulder
Boulder, CO 80309-0525

Modern neural network learning models such as competitive learning networks^[2,4], resonance correlation networks, and back propagation networks^[1] require a wider range of neuron behavior than a simple saturating threshold non-linearity. However, optical implementation of neurons that incorporate non-local, non-linear functions such as shunting inhibition, winner-take-all, and history-dependent behavior is beyond the capability of conventional optical devices. A new class of light modulator has been developed that combines the flexibility of analog and digital electronic VLSI circuits, optical detectors, and the switchable electro-optic capabilities of liquid crystal materials. In this paper we will show how these liquid crystal/VLSI modulators can be used in optical implementations of these learning networks. We discuss in detail a competitive optical learning network which uses LC/VLSI winner-take-all neurons on fractal grids to program adaptive volume holographic interconnections. We will present results from tests of the LC/VLSI winner-take-all modulator arrays, and in addition will show preliminary results from an optical competitive learning system that uses the LC/VLSI modulators as neurons.

Competitive Optical Learning Architecture

Many of the existing optical neural network implementations perform non-linear transfer functions and feedback with computers or discrete electronic circuits. Our prior work has emphasized using the full three dimensional nature of volume holography for interconnections, and using device physics to embody the neuron functionality^[1,2]. The LC/VLSI devices combine the advantages of both approaches. They have the flexibility and efficiency of electronic circuits, and have a high density and compactness (due to the advanced state of VLSI technology) which makes them suitable for large 2-D arrays. Arrays of LC/VLSI neurons can be interconnected by self aligning adaptive volume holographic weights which are formed by sequences of outer product exposures between sparse arrays of neurons^[1,2,3] in fractal geometries. The fractal array of neuron modulators, which utilizes only $N^{3/2}$ out of the N^2 chip area, is a good match for the LC/VLSI neurons since the remaining chip area can conveniently be used for the VLSI neural circuitry. This approach compares favorably with pure electronic VLSI neural networks because the chip area is used entirely for the neurons, and the interconnections are folded up into the third dimension available to the optics. This allows the implementation of much larger networks, and networks with higher neuron densities.

The optical competitive learning architecture is based on the competitive learning neural paradigm^[4]. A single layer of the self-aligning competitive optical learning network^[2] is schematically illustrated in Figure 1. The key components include the input spatial light modulator (SLM), the competitive winner-take-all modulator chip, the polarization switching photorefractive volume hologram, and a high speed phase conjugate mirror. A statistically clustered set of input patterns are applied one at a time to the input SLM. Some of the light is diffracted and polarization switched by the volume hologram through the polarizer and towards the array of LC/VLSI reflective modulators. The winner-take-all competition circuit causes the pixel in each competitive patch which is receiving the largest input to switch the liquid crystal above that pixel to the non-polarization

rotating state. The remaining pixels in each patch rotate the polarization of the reflected light by 90 degrees. The reflected light from the losing pixels is therefore blocked by the polarizer, while the light reflected from the winning pixels passes back through the polarizer and reilluminates the volume hologram. Meanwhile, the undiffracted light that has passed through the volume hologram records a dynamic grating in the phase conjugate mirror, which is read out by a strong orthogonally polarized counter propagating pump beam. The diffracted pump beam produces a phase conjugate wave focused back through the volume hologram towards the input SLM pixels from which the light originated. The phase conjugate wave interferes with the reflections from the winning pixels in the volume of the photorefractive crystal, thereby adding an outer product perturbation to the existing hologram. This hologram strengthens the interconnections between the input pattern and each of the winning pixels, while simultaneously decreasing all of the other interconnections with the losing pixels through incoherent erasure. The next time that a similar input pattern is presented to the network, it is even more likely to produce the largest diffraction towards the same winning node locations. After cycling through a statistically clustered set of input patterns many times, it is likely that individual nodes in the different competitive patches will become tuned to different clusters or groups of clusters of the inputs. Since the number of clusters in the input statistical distribution is unknown, different sizes of competitive patches are included in the fractal array, varying from 2 to 64 nodes in the initial chip. These different competitive patches will result in a clustering of the input pattern space with different levels of detail, resulting in the emergence of "feature detectors" sensitive to different topologically salient statistical features in the input pattern. Upon presentation of a particular input, the pattern of winning nodes will produce a sparse, partially distributed representation of the input pattern, that can itself be used as the input for subsequent levels of supervised or unsupervised processing.

Winner-Take-All LC/VLSI Modulator

The competitive optical learning architecture requires an optoelectronic means to select the pixel with the largest input, in order to reinforce that pixel's holographic connections. We have designed and fabricated a liquid crystal/VLSI device which detects optical inputs, selects the largest input with an electronic winner-take-all circuit, and switches the liquid crystal material so that all but the winning beam are blocked. An illustration of the operation of the liquid crystal/VLSI competitive modulator detector is shown in figure 2. The inputs to the device are optical beams focused onto photodetectors, which are formed as parasitic bipolar phototransistors in the CMOS VLSI process. The phototransistors are placed underneath a liquid crystal modulating pad structure, which is formed by an overglass cut to a metal electrode. Instead of using this structure as a bonding pad, it is filled with a liquid crystal, and is covered by an optical flat with an indium tin oxide (ITO) electrode and a polymer alignment layer coating. The thickness of the liquid crystal layer in the modulating pad is adjusted to give a half wave retardation upon reflection from the metal electrode in one state of the liquid crystal, and no retardation in the other state. Surface stabilized ferroelectric liquid crystals are being used in the initial devices, but homeotropic alignment could also be used as indicated in the figure. The metal electrode is thin enough to transmit a small fraction of the incident light into the photodetector underneath. The current produced in the photodetector is proportional to the incident optical flux, and this current is the input to a nonspecific global inhibition (winner-take-all) circuit^[5].

The operation of the circuit can be understood by considering the schematic diagram portion of Figure 2. When the currents I_1 and I_2 produced by the photodetectors are equal, the voltage of the competition bus, V_c , floats to a potential that allows each competitive node to contribute an equal current into the bus through transistors T_B . The sum of these currents is equal to the control current I_c . When the input intensities are equal, the competition voltages are the same:

$V_1 = V_2$. If the current I_2 increases, the current through T_{A2} must increase, which can only happen if V_c increases. However, the current through T_{A1} has not increased, so V_1 must decrease, while V_2 stays high. The final voltage output for the winning node in the two-unit WTA circuit is $V_{win} \propto [\ln(\frac{I}{I_o}) + \ln(\frac{I}{2I_o})]$, where I_o is a fabrication parameter. The output voltage of the losing nodes is suppressed close to zero. The output drivers threshold these voltages, resulting in 0 V being applied to the losing modulating pads, and 5 V to the winning pad. The threshold level is shifted by changing the I_c control current. The electric field between the ITO electrode and the pad causes the liquid crystal molecules to reorient, so that the reflection from the losing pad has its polarization rotated by 90 degrees and is blocked by the polarizer, while the reflection from the winning pad experiences no polarization change and passes back through the polarizer.

Photomicrographs of the fabricated chip are shown in Figure 3. The sparse fractal array is shown in Figure 3a, where long stripes of modulator detector structures are shown interconnected by the competitive circuitry. Competitive patches of various sizes are delineated by breaking the competitive bus between the different regions of competition, and including the global control current in a distributed fashion with a subthreshold transistor attached to each node. The large area between the rows of competitive modulator-detectors can not be used because of the Bragg degeneracy, and in fact there is significantly more room available for the implementation of more complicated neural functions. A close up of an array is shown in Figure 3b, showing the modulator pad/detector structure at top, and the winner-take-all and buffer circuitry at the bottom of the array.

Conclusion

A new type of competitive optoelectronic neuron array using VLSI circuitry with photodetectors and liquid crystal modulators has been fabricated. A self-aligning optical learning architecture that uses photorefractive crystals for the adaptive weights, and the competitive modulator array in a fractal topology has been presented.

The authors gratefully acknowledge the assistance of David Jared, Kristina Johnson, Dave Doroski, and John Lazzaro. T. Slagle acknowledges the support of an NSF Creativity in Engineering Fellowship. This work was supported by the NSF ERC in Optoelectronic Computing Systems CDR862228.

References

- [1] K. Wagner and D. Psaltis, Multilayer optical learning networks, *Appl. Opt.*, vol. 26(23), p. 5061 (1987).
- [2] K. Wagner and R. Feinleib, Competitive optoelectronic learning networks, *Proc SPIE* vol. 882-14 (1988).
- [3] D. Psaltis et al., Holography in artificial neural networks, *Nature*, vol. 343, p. 325 (1990).
- [4] D. Rumelhart and D. Zipser, Feature discovery by competitive learning, *Cog. Sci.* vol. 9, p. 75 (1985).
- [5] J. Lazzaro et al., Winner-take-all networks of $O(N)$ complexity, in *Advances in Neural Information Processing Systems 1*, Ed. D. Touretzky, p. 703, Morgan Kaufmann (1989).

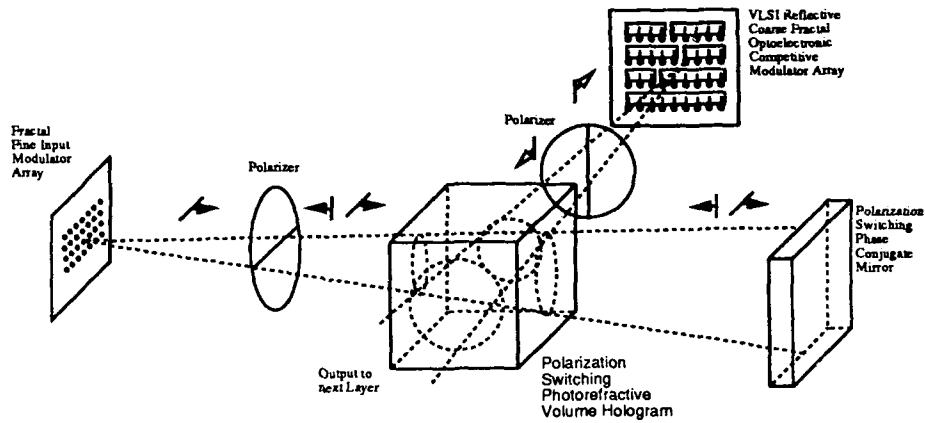


Figure 1. Self-aligning competitive optical learning architecture

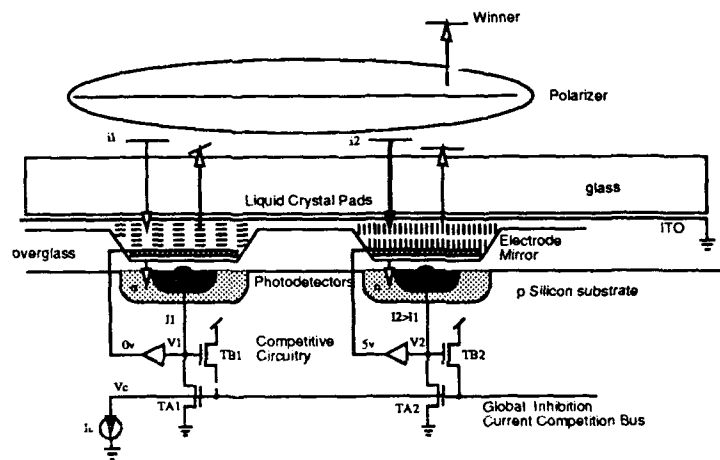


Figure 2. Schematic illustration of the competitive winner-take-all VLSI liquid crystal modulator-detector structure.

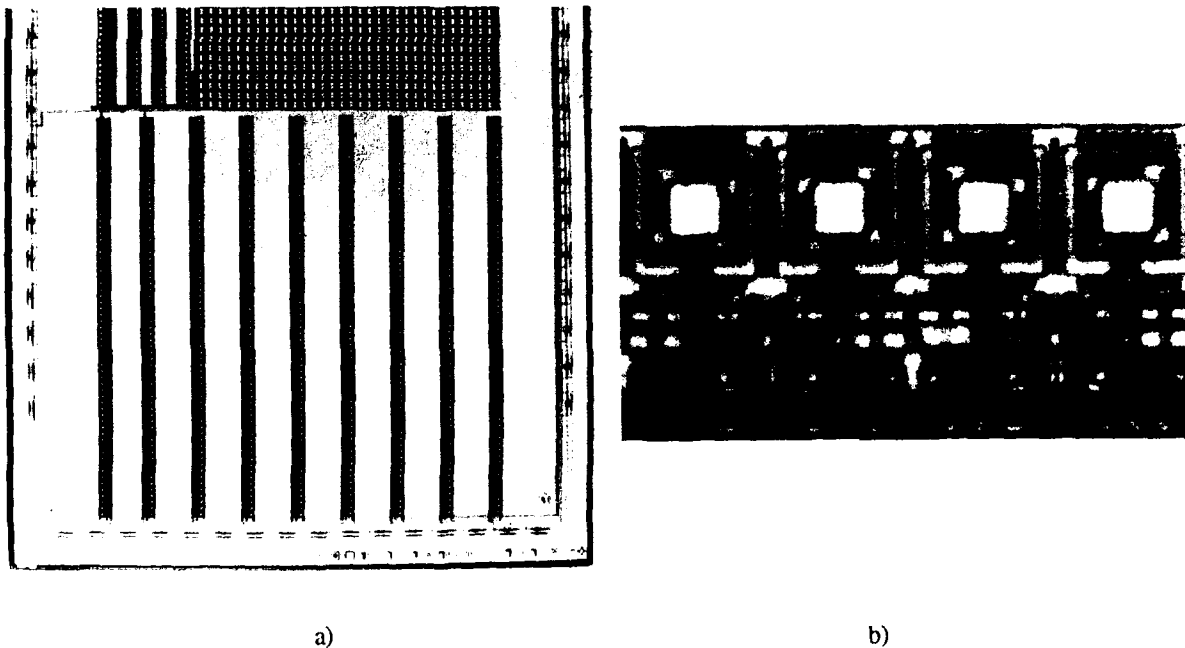


Figure 3. a) Photomicrograph of the fractal array of competitive modulators, b) a closeup of a few pixels.

Wednesday, March 6, 1991

Neural Network Components

WB 10:50am-12:20pm
Salon F

Kristina Johnson, Presider
University of Colorado



Parallel Implementations of Neural Networks: Electronics, Optics, Biology

Joshua Alspector
Bellcore, Morristown, NJ

1. COMPUTATIONAL REQUIREMENTS

The DARPA report^[1] notes that the human cerebral cortex is comprised of 100 billion neurons, each having about 1,000 dendrites that form about 100 trillion synapses. If you multiply that by its operating frequency of about 100 Hz, you get 10,000 trillion interconnections per second. These last two numbers (10^{14} connections and 10^{16} connections per second) form the basis for the graph in the DARPA report labeled Computational Requirements. The cortex not only has this prodigious capacity but dissipates about 10 watts, weighs 3 pounds, covers 0.15 square meters and is about 2 millimeters thick. To say it is something substantial for designers of neural systems to shoot for is a considerable understatement.

Fig. 1, adapted from the report, uses roughly this two-dimensional methodology to measure the capacity of various biological organisms and requirements for certain applications such as robotics or speech. Many important dimensions are missing such as ability to learn in order to endow those connections with knowledge, the speed at which those connections can be updated, the density of connections, the power dissipated, and scalability at a system level.

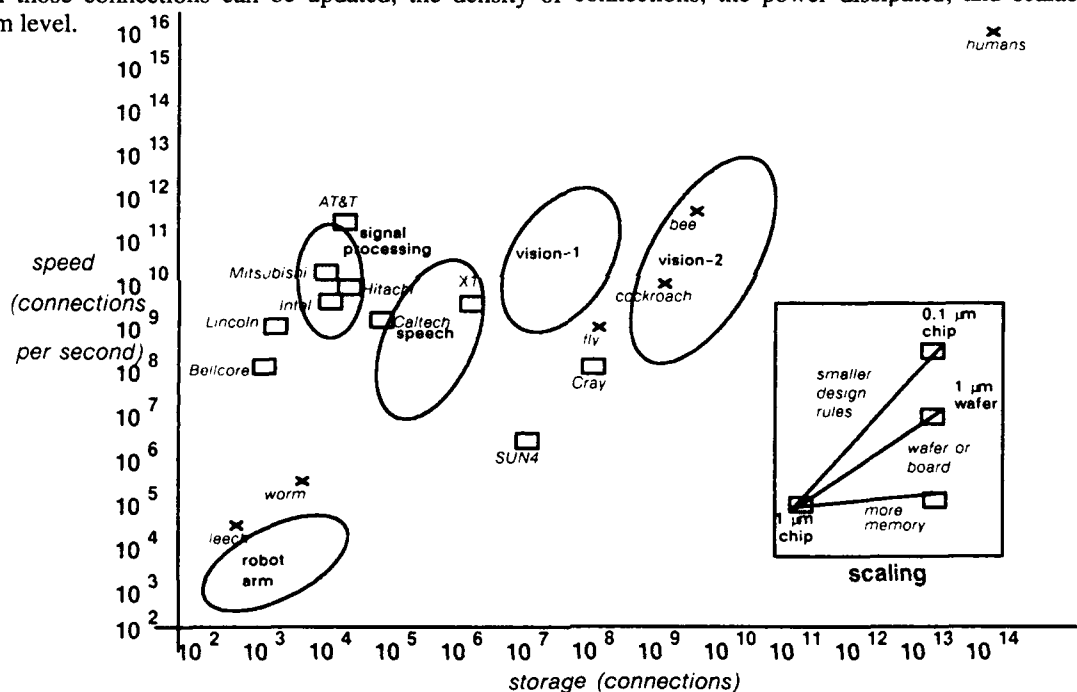


Figure 1. Computational Requirements

There are several chip designs on this chart. The largest number of connections per second (3×10^{11}) is from an AT&T chip^[2] with 1 bit weights and no learning. Adaptive Solutions' X1 chip^[3] is in fab and is a programmable digital parallel SIMD machine with 64 processors on a wafer and a capability for a large number of connections in digital memory. Hitachi^[4] also has a digital wafer-scale system. Bellcore^[5] and Mitsubishi^[6] have analog learning chips based on a Boltzmann architecture.^[7] The Bellcore chip is undergoing tests now and has a 32 channel uncorrelated noise source^[8] for stochastic learning as well as variable gain neurons for mean-field learning. The Mitsubishi design achieves a high synapse density through a special process for capacitive analog storage with refresh. The Intel chip^[9] achieves non-volatile analog storage through a special floating gate process. The Lincoln Lab^[10] and Caltech^[11] chips also use charge domain processing with CCDs.

The inset to Fig. 1 has some lines representing technology evolution for some paper designs.^[12] The line moving upward at the steepest angle represents an evolution from 1 micron technology to 0.1 micron technology and shows that you can not only pack more synapses on a single chip (storage) but also get them to run faster. The next steepest line represents sticking with 1 micron technology but moving from chip scale to wafer scale so you still get more connections. It is the kind of parallel speedup you can get by putting more chips on a board as well. Finally,

there is a horizontal line for a digital design method with off-chip RAM which simply represents adding more memory. Note that, for non-learning chips especially, connection updates per second (CUPS) can be dramatically lower than CPS. Note also, that biology, in general, has more storage and less speed than silicon.

2. PHYSICS OF COMPUTATIONAL DEVICES

2.1 Electrons and Photons

The two main categories that implementation of artificial neural networks fall into are electronic and optical. It therefore seems sensible to consider the interactions of electrons and photons from a fundamental viewpoint. Fig. 2 (a & b) shows the Feynman diagrams for electron-electron scattering and photon-photon scattering at the lowest orders in quantum electrodynamics.

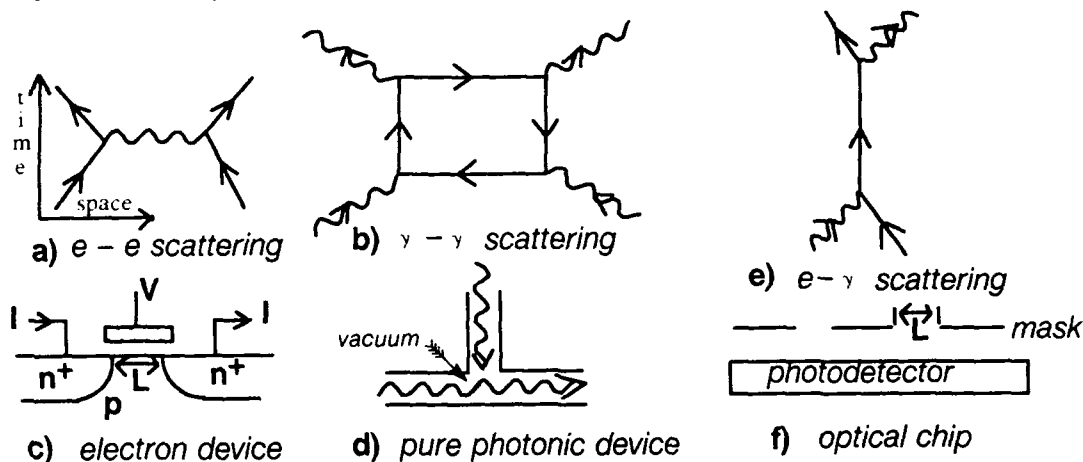


Figure 2. Basic electron and photon interactions.

Since there are two vertices in $e - e$ scattering while there are four in $\gamma - \gamma$ scattering, photon - photon scattering is greatly suppressed. For all practical purposes then, photons do not interact.

Figs. 2c & 2d show some devices that may be constructed using these basic interactions. Fig. 2c shows a MOSFET whereby charge on a gate controls the flow of electrons from source to drain. This is a well-known and highly successful electronic device found by the millions on integrated circuits. Fig. 2d shows an attempt at a purely photonic "transistor" where one light beam modulates another in the absence of electrons (vacuum). While there is a finite theoretical value for such modulation, in effect it does not exist and the search for a purely photonic equivalent of the transistor is hopeless.

What, then, is an optical device? Fig. 2e shows the Feynman diagram for electron - photon scattering which occurs through a single electron exchange with fairly high probability (cross-section). Therefore, the interaction of photons with the electrons in matter forms the basis for all so-called "optical" devices.

2.2 De Broglie Wavelengths

An important prelude to determining the density of optical or electronic devices attainable is the calculation of the de Broglie wavelength of the particles in question. Whether we use electrons or photons, the relevant kinetic energy is about 1 eV. This is about the bandgap of solid-state materials, the lowest conceivable voltage for operation of integrated circuits, and much greater than (about 40 times) the thermal energy of electrons at room temperature so that junctions behave properly. This is also roughly the energy of the atomic transitions that create photons in lasers and LEDs.

The de Broglie wavelength of a particle is $\lambda = h/p$ where h is Planck's constant and p is the momentum. We obtain for the wavelength of the 1 eV electron, $\lambda_e(1 \text{ eV}) = 10^{-3} \mu\text{m}$, while the photon has wavelength $\lambda_\gamma(1 \text{ eV}) = 1 \mu\text{m}$.

2.3 Device Density

Virtually all electronic devices depend for their operation on spatially localizing a transported particle. For example, the FET in Fig. 2c separates source and drain by a channel of length L . Electronic circuits work by sensing whether electrons moved across this channel. If, however, the dimensions of the channel were to approach the de Broglie wavelength of electrons, there would be no source-drain separation and the device would not work. Fig. 2f depicts an "optical chip" that some people have proposed for neural network computation. A mask (film,

LCD, etc.) localizes photons in such a way that the photodetector performs the sum of products operation in neural networks. The features on this mask cannot be smaller than the de Broglie wavelength of the photons or there would be blurring from adjacent features.

Table 1 describes the minimum characteristic length L that can be expected from both optical and electronic devices that depend on particle localization during the coming decade based on known technology and physics.

Year	Electronic device			Optical device		
	L	density	example	L	density	example
1990	$1\mu m$	$10^8/cm^2$	10 Mbit dRAM	$1\mu m$	$10^8/cm^2$	film, LCD, thin hologram
2000	$0.1\mu m$	$10^{10}/cm^2$	1 Gbit dRAM	$1\mu m$	$10^8/cm^2$	optical disk?

Table 1. Two dimensional density for particle devices.

We see that the conceivable improvement in MOS technology is not limited by the wavelength of electrons whereas no conceivable improvement in optical technology will reduce the space between optical particle devices below $1\mu m$. Therefore, we can expect the two-dimensional density of electronic devices to be two orders of magnitude greater than that of optical devices.

Optical devices, however, have an advantage when wave properties, imaging, and three dimensions are considered. By using holographic techniques in thick optical devices, one can store information in three dimensions. This degree of information storage can approach an analog quantity of a few bits precision every μm^3 , limited by λ_v . This is possible because of the polarizability of materials (temporary movement of electrons) on the scale of $1\mu m$, while being transparent to light. This leads to a density of $10^{12}/cm^3$, about two orders of magnitude greater than the equivalent (but 2-d) density for electronics.

3. NEURAL NETWORKS

3.1 Operations

Neural networks usually are characterized by several types of computation. The net input to each neuron (index i) is a multiply-accumulate operation, $net_i = \sum_j w_{ij} s_j$. Each neuron performs the computation, $s_i = f(\text{gain} * (net_i + \text{noise}))$ where f is a monotonic non-linear function such as \tanh . The learning rule which adjusts the synaptic weights w_{ij} from neuron j to neuron i is something like $\Delta w_{ij} = \eta[(s_i s_j)^{\text{desired}} - (s_i s_j)^{\text{actual}}]$.

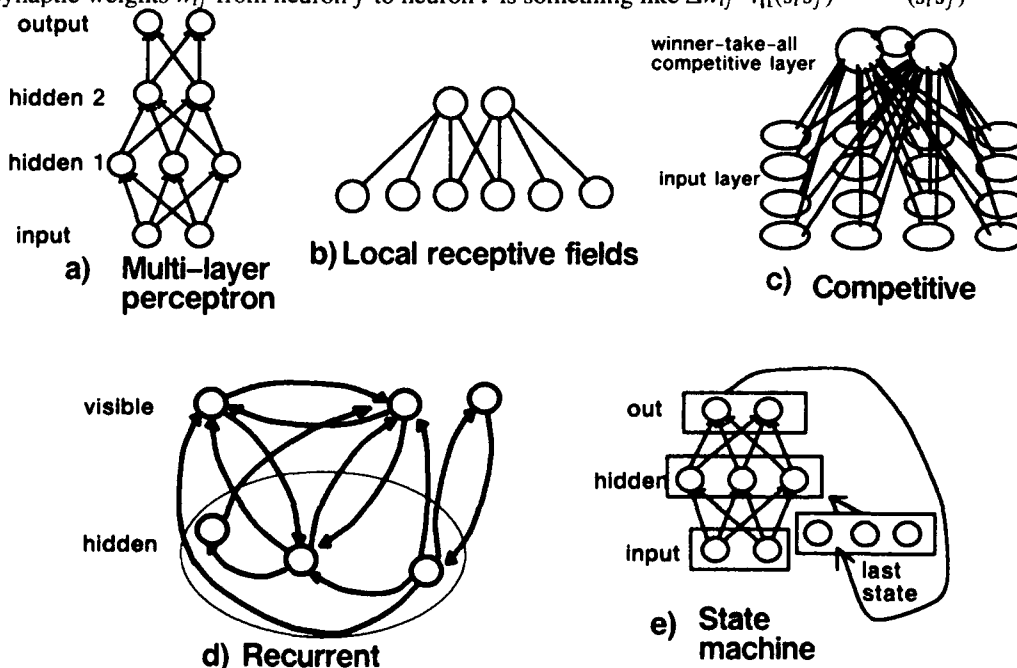


Figure 3. Neural network structures.

These operations are easily performed with electronic circuits, as has been shown,^[13] using conductances for multiplication and electron summing on wires for addition. Multiply-accumulate operations can also be done analogously in optics by masks and photon-summing but differences and non-linearities are more troublesome.

However, one strong point of optics involves high fan-in and fan-out operations, which, because of the non-interactivity of free-space photons, are relatively easy. This is not true for electronics, since each additional connection requires a wire. This means that spatially local structures are of great advantage for electronic implementations but are not a strong requirement in optics.

3.2 Structure

The question of locality raises the issue of the type of neural architectures that may be useful. Fig. 3 displays a selection of neural architectures that have received some attention. Multiple layer perceptrons (3a) which learn by back-propagation have the general architecture that the layers are fully connected from one layer to the next. This highly connected architecture has been found suitable for learning classification of static patterns of a single type. However, the introduction of modularity by means of local receptive fields (3b) and weight sharing has been shown to improve learning and generalization and reduce computation time for problems like speech recognition^[14] and optical character recognition.^[15] Local processing in combination with a competitive layer (3c) has^[16] been shown to be a powerful unsupervised learning method. When temporal patterns are considered, recurrent networks (3d) are required to take advantage of internal dynamical behavior. Supervised learning, where a teacher signal at the output is fed back to the network interior, requires feedback connections, at least for the learning phase. For implementation of local learning rules, full time feedback including feedback during processing has been found to be advantageous. By introducing loops (3e) in such nets, one can capture some of the sequential information in a state machine. These loops are a way of introducing locality and modularity in time analogous to the way that local receptive fields introduce locality and modularity in space. Convolutional nets are an extreme form of modularity and are widespread in biological sensory systems such as the retina.

4. ENERGY

The brain dissipates about 10 watts (10 Joules/sec) and potentially evaluates about 10^{16} synapses/sec. If we divide these two numbers, we get 10^{-15} Joules/connection. This should be considered a "holding" energy since only a small fraction of synapses are active at a given time. Von Neumann estimated that the brain dissipates about 300 pJ per binary act in a per neuron calculation. If one assumes 3000 synapses per neuron, we get about 10^{-13} Joules per active synapse.

We can get some idea of the limits of power dissipation in electronic devices by considering dRAM. A recent 64 Mbit dRAM technology^[17] can be characterized as having a "holding" energy of 10^{-18} J/bit. The "active" energy is about $2 \cdot 10^{-9}$ J/bit mostly due to system level considerations. If one looks only at the energy stored in the cell representing $4 \cdot 10^5$ electrons, we get about $5 \cdot 10^{-14}$ J/bit. We can imagine a technology which displaces 6000 electrons at about 1 volt for a switching energy of 10^{-15} J or 1 fJ. One can access a bit selectively in electronics by activating just the proper bit and word lines so that it isn't necessary to expend power for every bit in the chip. Actually, in neural networks, if one uses local learning rules, it isn't necessary to waste energy for bit and word line capacitances, since one needn't have global access to synapses to make use of their properties, so the 1 fJ figure is more realistic than for dRAM.

Because optics still must store information using electrons, the energy required to switch a bit must be at least the energy required for electronics plus conversion losses and any increases in sensitivity needed at the system level. In optics, estimates for the minimum power to switch a bit depend on materials but for discrete devices range from 1 to 100 fJ.^[18] The situation becomes worse when standby power is considered. In a holographic crystal, one usually bathes the entire crystal in light when writing the grating. It's not possible to illuminate only 1 of the 10^{12} possible bit locations in a 1 cm^3 crystal. Thus the selective update of certain synapses seems much more costly in energy than for electronics.

5. ROLE OF OPTICS

Optics has a tremendous advantage in being able to create two-dimensional images easily using a lens. Its use for connections is less clear but it may be that free space, coherent optics using three-dimensional holograms would lead to a density advantage over electronics. The incoherent optical chip seems not to have any advantage over an electronic chip. Just as optics has become the method of choice for telecommunication, so it may also prove useful in communication between chips using guided wave propagation. Optics uses the physics of atoms to advantage for high speed communications since atoms naturally cause electrons to vibrate at optical frequencies. Communication via electronics is less efficient because one must build a device to make electrons vibrate at high frequencies. In addition, electronic information transmission is lossy because of the need for charging and discharging a wire and terminating fast signals. A 10 Gbit/s optical bus may be a useful way to interconnect chips in an artificial neural system. Finally, a potential problem with wafer scale neural systems is power supply shorts. This would ruin the natural fault tolerance of a neural network. A graceful way to power such a silicon system may be by using solar cells and bathing pn junctions with light.

6. CONCLUSION

Optics will show its usefulness in image processing and communication for neural networks. Electronics has advantages in computation density, selective wiring, and power. However, for overall flexibility, robustness of design and creation, integrated functionality, density, and energy efficiency, biology is by far the best technology.

REFERENCES

1. "DARPA Neural Network Study", (AFCEA International Press, Nov., 1988).
2. H.P. Graf & D. Henderson, "A Reconfigurable CMOS Neural Network", Digest of Technical Papers, 1990 IEEE International Solid-State Circuits Conference, (San Francisco, CA, Feb. 14-16, 1990), p. 144.
3. M. Griffin, G. Tahara, K. Knorpp, R. Pinkham, R. Riley, D. Hammerstrom, & E. Means, Digest of Technical Papers, 1991 IEEE International Solid-State Circuits Conference, (San Francisco, CA, Feb. 13-15, 1991), to appear.
4. M. Yasunaga et. al., "Design, Fabrication and Evaluation of a 5-inch Wafer Scale Neural Network LSI Composed of 576 Digital Neurons", Proceedings of the International Joint Conference on Neural Networks (San Diego, CA, June, 1990) p. II-527.
5. J. Alspector, R.B. Allen, A. Jayakumar, T. Zeppenfeld, & R. Meir, "Relaxation Networks for Large Supervised Learning Problems", in *Advances in Neural Information Processing Systems* (Denver, CO, Nov. 1990), edited by D. Touretzky (Morgan-Kaufmann, Palo Alto, CA), to appear.
6. Y. Arima et. al., "A Self-Learning Neural Network Chip with 125 Neurons and 10K Self-Organization Synapses", 1990 Symposium on VLSI Circuits, Digest of Technical Papers, June 7-9, 1990, Honolulu, HI, (IEEE 90CH2885-2), p. 63. (1990). See also: Y. Arima et. al., "336-Neuron 28k-Synapse Self-Learning Neural Network Chip with Branch-Neuron-Unit Architecture", Digest of Technical Papers, 1991 IEEE International Solid-State Circuits Conference, (San Francisco, CA, Feb. 13-15, 1991), to appear.
7. D.H. Ackley, G.E. Hinton, & T.J. Sejnowski, "A Learning Algorithm for Boltzmann Machines", *Cognitive Science* **9** (1985) pp. 147-169.
8. J. Alspector, J.W. Gannett, S. Haber, M.B. Parker, & R. Chu, "A VLSI-Efficient Technique for Generating Multiple Uncorrelated Noise Sources and Its Application to Stochastic Neural Networks", *IEEE Trans. Circuits & Systems*, **38**, 109, (Jan., 1991).
9. M. Holler, S. Tam, H. Castro, & R. Benson, "An Electrically Trainable Artificial Neural Network (ETANN) with 10240 'Floating Gate' Synapses", Proceedings of the International Joint Conference on Neural Networks (June, 1989), p. II-191.
10. A. Chiang, R. Mountain, J. Reinhold, J. LaFranchise, J. Gregory, & G. Lincoln, "A Programmable CCD Signal Processor", Digest of Technical Papers, 1990 IEEE International Solid-State Circuits Conference, (San Francisco, CA, Feb. 14-16, 1990), p. 146.
11. A.J. Agranat, C.F. Neugebauer, R.D. Nelson, and A. Yariv, "The CCD Neural Processor: A Neural Network Integrated Circuit with 65536 Programmable Analog Synapses", *IEEE Trans. Circuits & Systems*, **37**, 1073 (1990).
12. Jim Burr, presentation at VLSI workshop, Neural Information Processing Systems conference, Keystone, CO, Dec., 1989.
13. J. Alspector, B. Gupta, & R.B. Allen, "Performance of a Stochastic Learning Microchip" in *Advances in Neural Information Processing Systems* (Denver, CO, Nov. 1988), edited by D. Touretzky (Morgan-Kaufmann, Palo Alto, Feb., 1989), pp. 748-760.
14. A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, & K. Lang, "Phoneme Recognition Using Time-Delay Neural Networks", *IEEE Trans. Acoust. Speech & Sig. Proc.*, **37**, pp. 328-339, (March, 1989).
15. Y. Le Cun, et. al. "Handwritten Digit Recognition with a Back-Propagation Network", in *Advances in Neural Information Processing Systems 2* (Denver, CO, Nov. 1989), edited by D. Touretzky (Morgan-Kaufmann, Palo Alto, 1990), pp. 396-404.
16. T. Kohonen, "Self-organized Formation of Topologically Correct Feature Maps", *Biol. Cybern.* **43**, 59-69, (1982).
17. Y. Nakagome et. al., "A 1.5V Circuit Technology for 64Mb DRAMs", 1990 Symposium on VLSI Circuits, Digest of Technical Papers, June 7-9, 1990, Honolulu, HI, (IEEE 90CH2885-2), p. 17, (1990).
18. P.W. Smith, "On the Physical Limits of Digital Optical Switching and Logic Elements", *The Bell System Technical Journal*, **61**, 1975, (Oct., 1982).

THE FIRST DEMONSTRATION OF AN OPTICAL LEARNING CHIP

Kazuo Kyuma, Yoshikazu Nitta, Jun Ohta,
Shuichi Tai, and Masanobu Takahashi

Mitsubishi Electric Corporation, Central Research Laboratory
8-1-1 Tsukaguchi-Honmachi, Amagasaki-city, Hyogo, 661 JAPAN

1. INTRODUCTION

Recently, there has been a strong interest in artificial neural networks for real time applications. Among several approaches, opto-electronic neural networks¹⁾ are quite attractive because of a dense-interconnection, a parallel-processing, and a large-scale integration capabilities using the advanced GaAs semiconductor technologies.

We previously reported several GaAs optical neurochips^{2, 3)} which consist of a light-emitting-diode(LED) array, a static(fixed) interconnection synaptic mask and a photodiode array in a 3-D layered structure. However, in order to utilize the learning capability which is one of the most important features of neural networks, the static interconnection mask must be replaced with a dynamic (variable) one. Until today, the most serious problem to realize such an optical learning neurochip has been a lack of the analogue spatial light modulators (SLM) which is suitable for 3-D integration as the dynamic interconnection device.

In this paper, we report on the optical learning chip for the first time, which acquires knowledge from the external circumstances in real time. The key point to succeed in it is the development of a fast-operating and variable-sensitivity photodiode(VSPD) that has the combined functions of the analogue SLM and the photodiode. The learning speed exceeding 640 MCUPS, which is 500 to 1000 times higher than that of the present engineering work stations, was obtained for the 8 neuron- and 64 synapse-optical neurochip. We also demonstrate the experimental results of the pattern classification with 12 training signals using this chip and the back-propagation(BP) learning algorithm.

2. OPTICAL LEARNING CHIP

2-1 Variable Sensitivity Photodiode

The schematic diagram of the VSPD is shown in Fig. 1. It is a photodiode having a metal-semiconductor-metal (MSM) structure. The principle of operation is that the photocurrent is proportional to the transverse electric field which is applied between the interdigital electrodes. The MSM-VSPD was fabricated by evaporating Al Schottky contacts on the GaAs substrate. The whole size of the photosensitive area was $100 \times 100 \mu\text{m}^2$. The gap width w_g and the finger width w_f

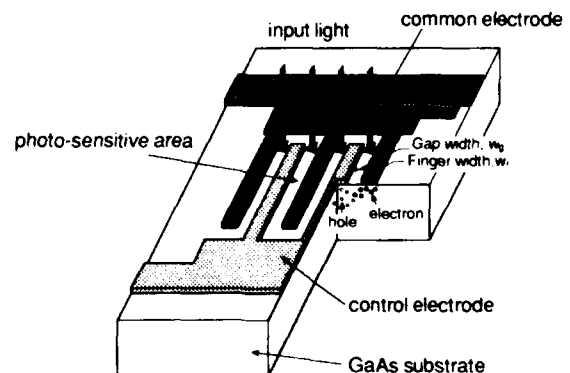


Fig. 1 Schematic diagram of the VSPD.

of the interdigital electrodes were, respectively, $10\mu\text{m}$ and $5\mu\text{m}$. These parameters were optimized so that the higher sensitivity and the wider dynamic range were obtained.

Figure 2 shows a typical experimental relationship between the photocurrent and the applied bias voltage under constant illumination power of $30\mu\text{W}$ for several VSPD devices. It is found that the photocurrent, which is proportional to the detection-sensitivity, is varied with the bias voltage. The sensitivity of 0.3 A/W was obtained at the bias voltage of 10V . It is also found that the relation between the photocurrent and the bias voltage is symmetric about the origin of the coordinate axes, and the direction of the photocurrent is reversed by changing the polarity of the voltage. This is because the MSM-VSPD has a symmetric structure about the photosensitive semiconductor area. As described in the next section, this feature is very useful for implementing optical neural networks. The response time was faster than $0.1\mu\text{s}$. The dark current was measured to be less than 1nA because the Schottky barrier is high enough to reduce it. And the breakdown voltage was higher than 15V .

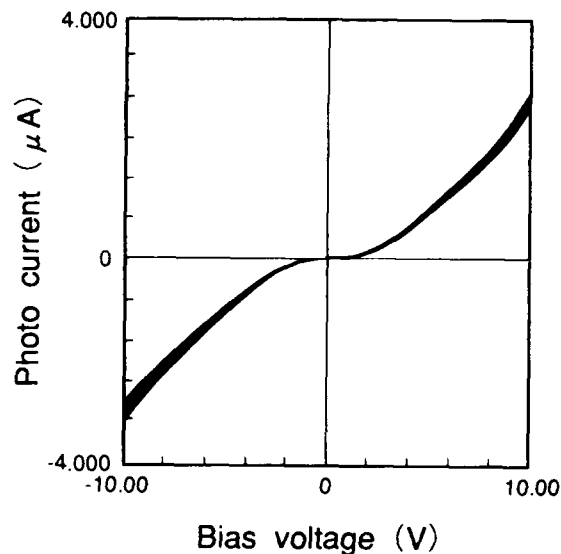


Fig. 2 Photocurrent as a function of the bias voltage for several VSPDs.

2-2 Optical learning chip

The MSM-VSPD is very useful for the optical implementation of neural networks in the following points:

- (1) The detection-sensitivity is monotonously increased with the bias voltage. Supposing that the synaptic weight corresponds to the sensitivity, the analogue synaptic weights essential to the neural networks can be implemented.
- (2) The direction of the photocurrent is reversed by changing the polarity of the bias voltage. This unique property permits us to implement both excitatory (positive) and inhibitory (negative) synapses in one VSPD device. Though the conventional optical architectures have required two optical modulators.
- (3) Since the structure of the MSM-VSPD is very simple and made of GaAs, it is quite suitable for optical integration in the form of the optical neurochips.

With these benefits, we have succeeded in fabricating optical learning chip with variable synaptic weights. The optical learning chip is consisting of a 2-D VSPD array with 8×8 elements and a line-shaped LED array with 8 elements. The epitaxial wafers for these arrays were grown by the molecular beam epitaxy. These two chips were integrated in a layered structure by using the modified flip-chip bonding technique as shown in Fig. 3. The chip size was $6 \times 6\text{ mm}^2$. Figure 4 shows the pictorial view of the fabricated optical neurochip mounted in

a LSI package.

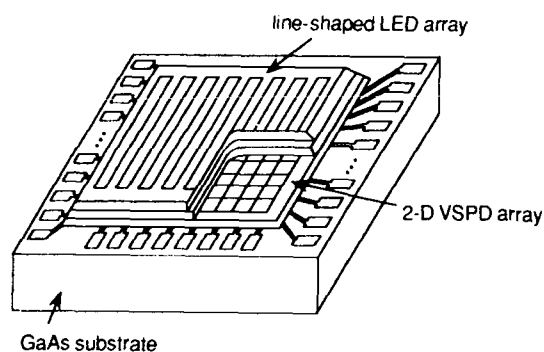


Fig. 3 Schematic Diagram of the optical learning chip.

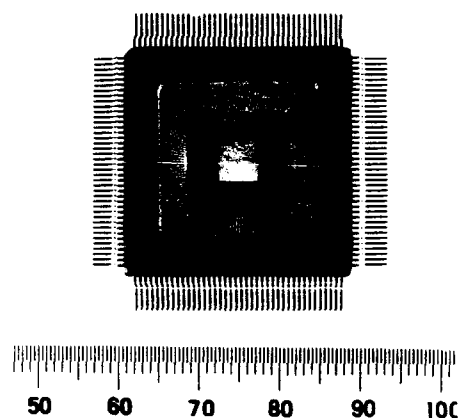


Fig. 4 Pictorial view of the optical learning chip.

The function of the neurochip is to perform the vector-matrix multiplication required for the neural processing in parallel. The LED array represents the input state vector V of neurons with analogue values. The one-side of the VSPD interdigital electrodes are biased by addressing the external analogue signals in parallel in order to yield the detection-sensitivity which is corresponding to the synaptic weights W . On the other hand, the other-side of the electrodes are mutually connected within every row in order to produce the matrix-vector product $u = Wv$. Then the matrix-vector product is obtained in parallel from the output of the SVPD array.

Figure 5 shows the measured photocurrent as a function of the number of the VSPD devices on which the bias voltage V_0 is applied, while the uniform light was emitted from all the 8 LEDs. It is shown that the photocurrent is proportional to the number of the on-state VSPD for every positive- and negative-bias voltage. These experimental results indicate that the fabricated device performs a good vector-matrix multiplication. The response time of the LED and VSPD is higher than $0.1\mu s$. Then the corresponding learning and retrieval speed exceeds 640 MCUPS and 640 MCPS, respectively.

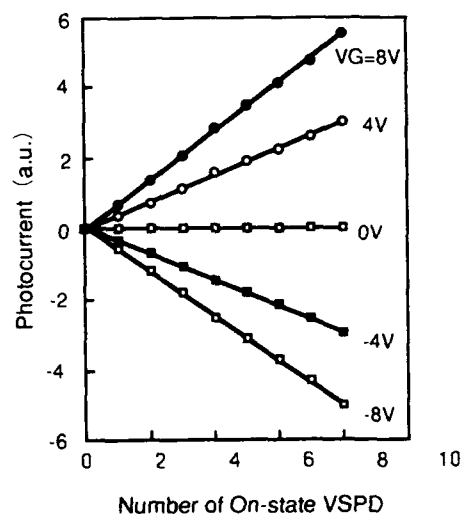


Fig. 5 Photocurrent as a function of the number of the on-state VSPDs with the parameter of the bias voltage.

3. APPLICATION TO THE PATTERN CLASSIFICATION PROBLEM

The optical neurochips were used to implement the error-driven BP learning algorithm. The schematic diagram of the learning system is shown in Fig. 6. The numbers of neurons in the input, hidden and output layers are, respectively,

8, 8 and 3. In order to implement the three layered network, the time-division multiplexing technique⁴⁾ was employed. As an example, we have applied the optical neurochip to the problem in which 12 patterns with binary codes are classified into 3 classes through learning. In the learning process, the 12 training patterns are successively presented at the input layer. If the output is not correct, the synaptic connections are modified according to the BP algorithm. The learning curve for the experimental system is shown in Fig. 7. The average recognition rate is plotted as a function of the number of the learning cycles. All the 12 patterns were correctly classified after 500 presentations.

4. CONCLUSION

We have developed a GaAs optical learning chip for the first time, by using the sensitivity variable photodiode as the synaptic interconnection device. The principle of operation and the fundamental characteristics of the VSPD and the optical learning chip were described. It was shown that the very fast learning speed exceeding 640 MCUPS can be achieved. We have also succeeded in the experimental demonstration of the pattern classification using the fabricated optical learning chip.

(REFERENCES)

- 1) D.Psaltis and N.Farhat, Opt. Lett., 10, 98(1985).
- 2) J.Ohta, et al., Opt. Lett., 14, 844(1989).
- 3) K.Kyuma, et al., SPIE Vol.1281, 124(1990).
- 4) M.Oita, et al., Opt. Lett., 15, 227(1990).

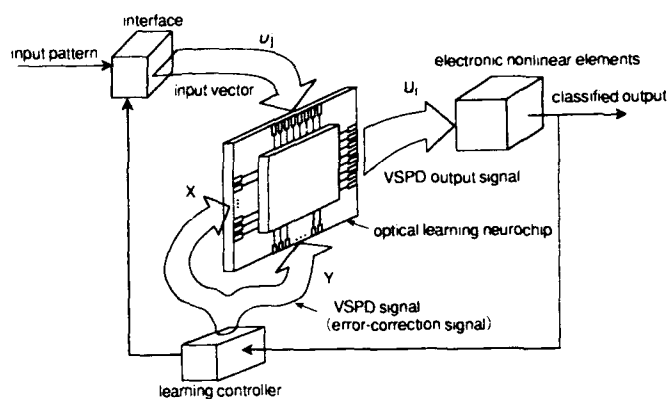


Fig. 6 Schematic diagram of the learning system.

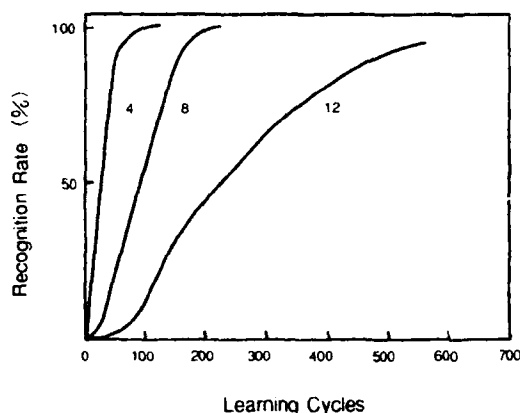


Fig. 7 Learning curve for the experimental system.

GaAs-BASED OPTOELECTRONIC NEURONS

Steven Lin, Francis Ho^{a)}, Jae Kim^{b)}, and Demetri Psaltis

California Institute of Technology
Department of Electrical Engineering, 116-81
Pasadena, California 91125
TEL : 818-356-3888
FAX : 818-568-8437

a) Stanford University, Stanford, CA 94305

b) Jet Propulsion Laboratory, California Institute of Technology, Pasadena CA 91109

The optical implementation of a neural network consists of two basic components : a 2-D array of neurons and interconnections. Each neuron is a nonlinear processing element that, in its simplest form, produces an output which is the thresholded version of the input. Liquid crystal spatial light modulators are candidates for such 2-D array of neurons. However, they are not flexible in their use. Optoelectronic integrated circuits (OEIC's), either hybrid, such as liquid crystal on silicon, Si-PLZT, and flip-chip devices, or monolithic integration in III-V compounds, is another solution. In order for these devices to be used as neurons in a practical experiment, they must be large in number ($10^4/cm^2 - 10^6/cm^2$) and exhibit high gain. This puts a stringent requirement on the electrical power dissipation. Thus, these devices have to be operated at low enough current levels so that the power dissipation on the chip does not exceed the heat-sinking capability , and yet the current levels need to be large enough to be able to produce high gain. This means sensitive input devices are a must. To achieve these goals, the speed requirement of the devices must be relaxed as the operation of neural network does not have to be too fast.

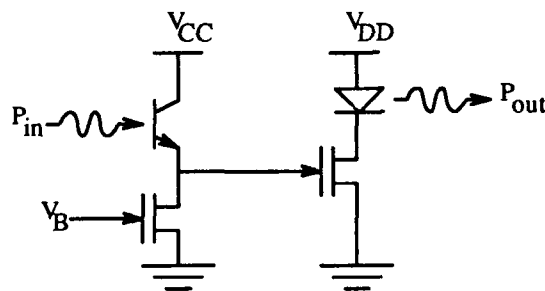


Fig. 1 Schematic Circuit Diagram of an Optoelectronic Neuron

In this paper, we present an optoelectronic neuron that monolithically integrates a

detector, 2 transistor amplifiers, and a light source on a single GaAs substrate. LED's have been chosen as the light source, as opposed to lasers, because no threshold currents are needed to drive the LED's so that a large array of neurons at low currents is possible and LED's are inherently simpler to fabricate. The circuit diagram of the optoelectronic neuron we describe in this paper is shown schematically in Fig. 1. A switching circuit at the input is formed by connecting a double heterojunction bipolar phototransistor in series with a biasing MESFET. Upon detecting enough incoming light, the phototransistor becomes saturated, thus pulling up the source-drain voltage across the biasing MESFET. This voltage turns on the other MESFET, which, in turn, drives the LED to emit light. The input thresholding characteristics are controlled by the gate voltage, V_B , of the biasing MESFET. The larger the V_B is, the larger the threshold is because the photocurrent generated by the phototransistor has to satisfy the current drawn by the biasing MESFET before the excess current can flow to the gate of the LED-driving MESFET and charge up its gate. The output saturation is provided by the finite swing of the gate voltage in the driving MESFET. The differential gain of the neuron before becoming saturated is determined by the slopes in the I-V curves of the phototransistor and the biasing MESFET. If the slopes for these two transistors are zero, the differential gain in the neuron would be infinite. Thus, by minimizing these slopes, such an integrated optoelectronic neuron is capable of turning on the neuron at very low input light levels. This is essential for systems, such as neural network, that require large gains, large number of neurons, and yet low enough power dissipation on the chip.

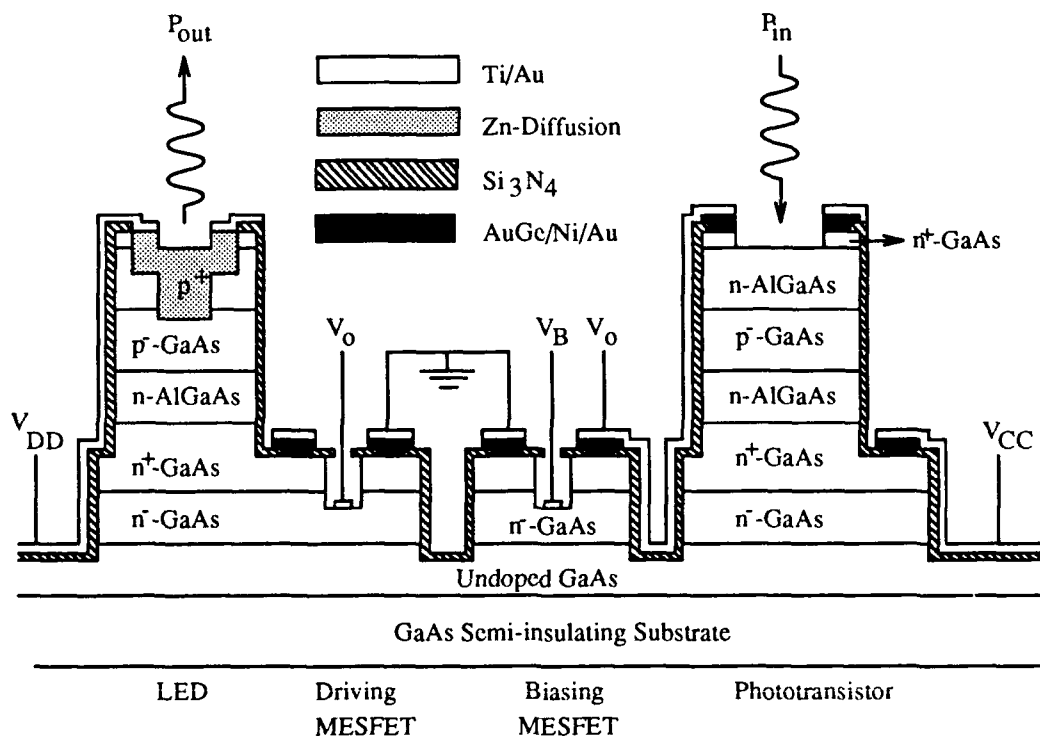


Fig. 2 Cross Sectional View of an Optoelectronic Neuron

The cross section of the optoelectronic neuron is shown in Fig. 2. The epitaxial

layers are grown by MOCVD. Upon standard substrate cleaning processes, the substrate is subjected to two chemical wet etchings in defining each device in a neuron and isolating the adjacent neurons. A Zn-diffusion down to the active p-GaAs layer through a 1000 Å-thick Si_3N_4 mask is then followed to form a double heterojunction LED. Another shallower, yet wider Zn-diffusion is performed to aid the current flow through the LED so that the emitted light is not under the evaporated metals of the LED. Appropriate windows are subsequently opened for all AuGe/Ni/Au n-type contact evaporations, and are followed by proper alloying. The gates of the MESFET's are recessed from the surface and are defined by etching and measuring the source-drain currents at the same time. Once the proper recessed depths for the gates are determined, Ti/Au are evaporated to form the gates and also to interconnect the devices. The size of a fabricated optoelectronic neuron is about $200 \times 200 \mu\text{m}^2$. The gates for the biasing and driving MESFET's are measured to be $6 \times 70 \mu\text{m}^2$ and $6 \times 100 \mu\text{m}^2$. And the LED and the phototransistor light-sensitive areas are $40 \times 40 \mu\text{m}^2$ and $80 \times 60 \mu\text{m}^2$, respectively.

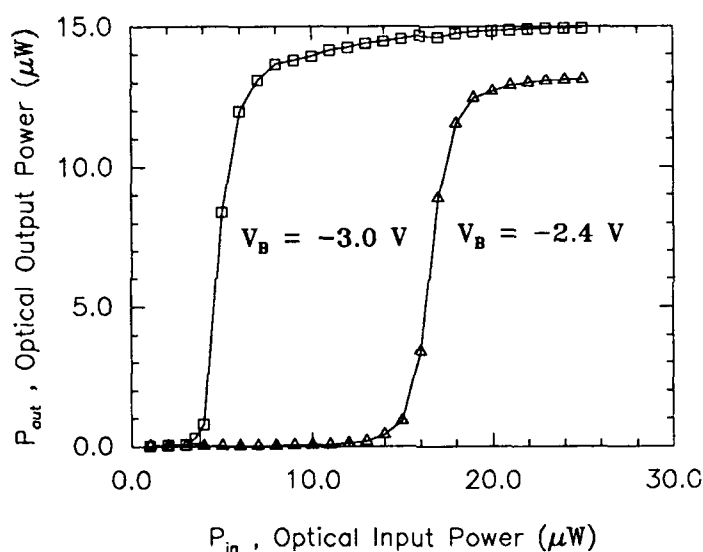


Fig. 3 Measured Input-Output Characteristics of an Optoelectronic Neuron

Fig. 3 shows the measured input-output characteristics of an optoelectronic neuron. A variable threshold controlled by the gate voltage of the biasing MESFET, V_B , is clearly evident in the plot. For the curve at $V_B = -3\text{V}$, the output initially remains close to zero for input up to $3 \mu\text{W}$, then rises to $12 \mu\text{W}$ within $2 \mu\text{W}$ of input light power. This implies a differential optical gain of 6 in the neuron. The output of the neuron continues to rise gradually as the input increases further. The differential optical gain of 6 is limited by the leakage currents across the gate-drain schottky diodes in both MESFET's as well as the finite slopes in the I-V curves of the phototransistor and the biasing MESFET. With further reduction in the doping concentration in the MESFET's conduction n^- layer and an increase in the doping concentration in the phototransistor's base layer, the optical differential gain can be further improved. It is noted that the output saturation levels for $V_B = -3\text{V}$ and $V_B = -2.4\text{V}$ curves are different owing to a higher common-emitter satu-

ration voltage for the phototransistor, $V_{CE,SAT}$, and thus a smaller swing in the switching circuit for the $V_B = -2.4V$ curve. When characterized individually, the LED and the phototransistor are measured to exhibit efficiencies of 0.01 W/A and 1 A/W, respectively, and the transconductance of the MESFET's, g_m , is measured to be 20 mS/mm. The efficiencies in the LED and the phototransistor are limited by the thick p-GaAs layer in both devices, which causes self-absorption in the LED and the degradation in the current gain, β , of the phototransistor. It is expected that much improvement can be obtained by reducing the thickness of this layer. The current through the LED is about 1.2 mA, which implies, with $V_{DD} = 2V$, the electrical power consumption per neuron is about 2.4 mW. The response of the neuron is measured to be 5 μsec as shown in Fig. 4, and is found to be limited by the charging of the capacitors in the circuits. With these results, the optical switching energy per neuron is thus calculated to be $(2 \mu\text{W}) \times (5 \mu\text{sec}) = 10 \text{ pJ}$.

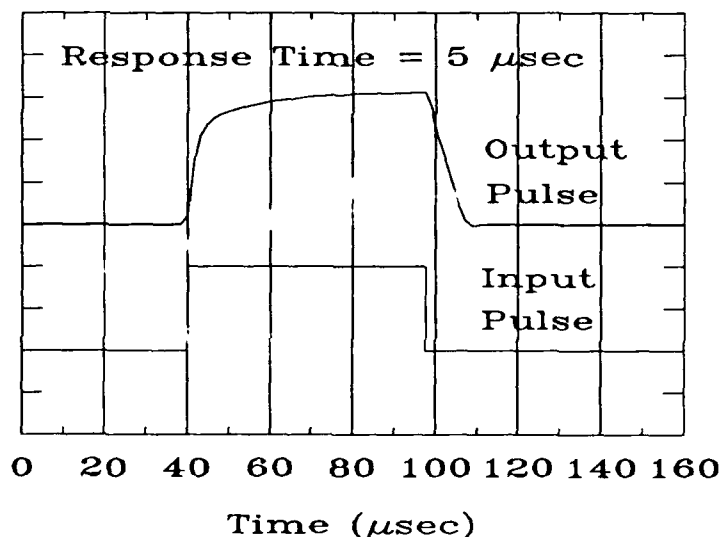


Fig. 4 Time Response Measurements of an Optoelectronic Neuron

In conclusion, a GaAs-based monolithically integrated optoelectronic neuron with variable thresholding characteristics is demonstrated. The differential optical gain of 6 is obtained. The threshold of the neuron can be controlled by the gate voltage in the biasing MESFET. With the measured 5 μsec in the response time of the neuron, the optical switching energy is 10 pJ per neuron. By optimizing epitaxial layer parameters, the performance of such an optoelectronic neuron is expected to improve by at least an order of magnitude.

Nonlinear thresholding characteristic for optical computing of

optically addressable GaAs/FLC-SLM

by

Masashi HASHIMOTO, Yukio FUKUDA*, Shigeki ISHIBASHI**
and Ken-ichi KITAYAMA

NTT Transmission Systems Laboratories,
1-2356 Take, Yokosuka-shi, Kanagawa 238-03, Japan

*NTT Opto-Electronics Laboratories,

**NTT Applied Electronics Laboratories.

1. Introduction

In a wide variety of optical parallel processings, spatial light modulators (SLMs)⁽¹⁾, especially optically addressable SLMs (OASLMs), are versatile devices. SLMs have been used as an optical encoder and latch memory for optical computing⁽²⁾ and synaptic weighting mask for optical neural network.⁽¹⁾

In this paper, accumulative thresholding (AT) characteristic, originated from the nonlinear transfer function, of newly developed GaAs/FLC-SLM that is a special class of OASLM structured with GaAs p-i-n diode and ferroelectric liquid crystal (FLC) is investigated. A monotonous sigmoid-like thresholding against optical energy of input pulses sequence is experimentally obtained. Furthermore, the application of the nonlinear characteristic of the SLM to both OR logic gate and latch memory is proposed.

2. Optically addressable GaAs/FLC-SLM

The structure of the transmission-type GaAs/FLC-SLM is shown in Fig. 1(a). The GaAs p-i-n photodiode is constructed on a p-GaAs substrate which is coated with indium tin oxide (ITO). ITO is also deposited on n-GaAs to form pixelized Schottky electrode.

The FLC, applied with forward bias voltage, acts as capacitance (C_{FLC}) and switches to a stable state, resulting in an erasure of stored information. (Phase 1 in Fig. 1(b)) Under a reversed bias voltage, the photodiode becomes highly resistive and is considered as capacitance ($C_{G.A.S.}$). When the write light is absent, a voltage ($= (\text{bias voltage}) * C_{FLC} / (C_{FLC} + C_{G.A.S.})$) is applied to FLC, which remains in its OFF state. (Phase 2) When the write light is turned on, photocurrent from the photodiode causes the increase of the applied voltage to FLC. (Phase 3) The read light from laser diode ($1.3 \mu m$) rotates the polarization of the light by passing through FLC, resulting in the on-state. Figure 1(c) shows an example of response of GaAs/FLC-SLM. The write light is $150 \mu sec$ pulse of $1.8 mW/cm^2$ from He-Ne (633 nm). The rise time (10-90%) and fall time are $86 \mu sec$ and $98 \mu sec$, respectively.

In Figs. 2 (a) and (b), optical output versus bias voltage and pulse width are shown, respectively. It is found that there is thresholding characteristic on not only voltage when write light is given, but also pulse width that corresponds to input optical energy. It is clear from (a) that thresholding can be controlled by bias voltage under a constant illumination. From the Fig. 2(b) and phase 3 model in Fig. 1(b), the GaAs/FLC-SLM is expected to show AT characteristic.

In Fig. 3 (a), the AT characteristic is schematically shown against accumulative energy of input pulse sequence. Assume that the energy of a single input optical pulse is too low to obtain the output "1" but the accumulation of energy of several pulses leads to value "1". If input pulse is set to be over the

threshold, optical logic operation OR between sequential inputs is executed. In Fig.3 (b), transient responses for one and two input optical pulses are shown. In the case of single input pulse whose energy is below the threshold, the output returns to 0. In the case of two input pulses, the input energy exceeds the threshold. Then, the output becomes "1" and keeps its state by memory characteristics of FLC in a readout time. In Fig.3 (c), the result of AND operation between sequential optical inputs on different timing is shown. Input light from He-Ne laser is 2mW/cm^2 having pulse width of one input pulse $20\mu\text{sec}$.

3. Application of AT-SLM to sequential logic operation

The AT-SLM can be applied to 2-D optical latch memory as well as AND and OR optical logic gates. The use of AT-SLM will provide simpler optical implementation and processing algorithm for optical processor than previous optical systems having optical feedback bus or time sequential control. In Fig.4(a), an existing architecture of sequential operation based upon finite state machine is shown.⁽²⁾ For sequential logic operations such as the execution of functional form of sum of products parallel feedback loop between memory and logic array is needed. In Fig.4(b), a new architecture for optical parallel processing using AT function is shown. The optical logic array performs Boolean logic operation for two binary inputs.⁽³⁾ AT section executes OR operation and store the result to add the product terms from logic array which executes AND operation. As a result, sum of products operation can be obtained from AT-SLM. It is very difficult to construct precise optical feedback path for sequential operation. Then, simpler configuration by introduction of AT-SLM will be practically useful because optical feedback path is not needed. In Fig.5, the processing algorithm for sum of products is schematically shown. As shown in Fig.5, this shows another merit that simple processing algorithm is available because latching is not necessary in this architecture.

4. Conclusion

Accumulative thresholding function a unique characteristic of GaAs/FLC-SLM has been investigated. The device shows the thresholding characteristics for time sequential inputs. Its experimental results of AND operation results are given.

The architecture and processing algorithm for optical parallel processor using AT function as both the latch and OR operation has been presented. This versatile SLM would also be applicable to optical synaptic weighting in optical neural networks.

The authors would like to thank Dr. S. Shimada, Dr. H. Ishio and Dr. T. Matsumoto for their encouragement.

References

- (1) For example, K.M. Johnson et al. Appl. Opt. 28, 4888(1989).
- (2) M.Hashimoto et al. OC'89 WD5(Salt Lake, Mar. 1989).
- (3) For example, M.Hashimoto et al. Appl. Opt. 28, 4305(1989).

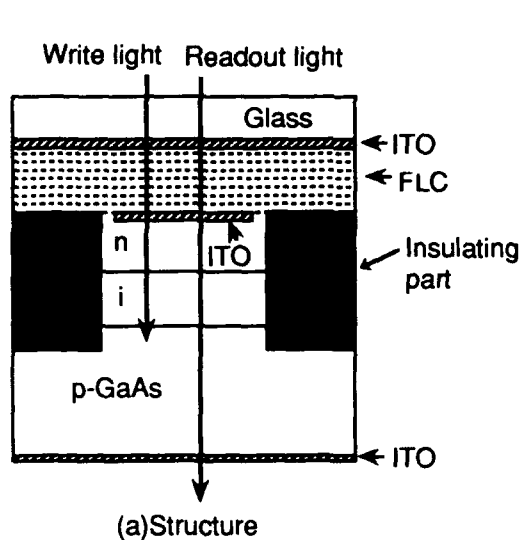


Fig.1 GaAs/FLC-SLM

Optically addressable transmission-type SLM. (a) is a cross section of one pixel. (b) Circuit model of GaAs/FLC-SLM. Under reverse bias write light induces photocurrent. (c) Response signal of the SLM. The rise time (10-90%) and fall time were 86 μ sec and 98 μ sec, respectively.

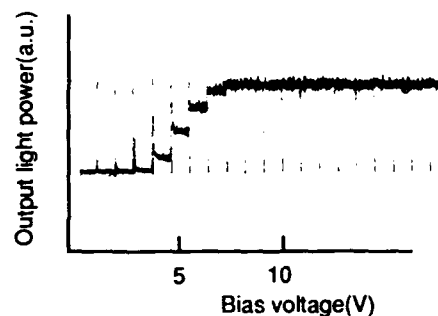
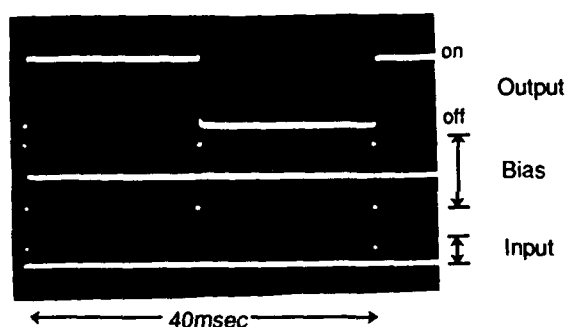
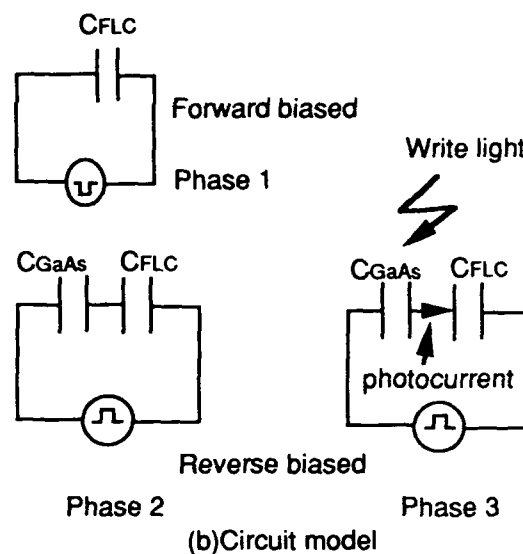
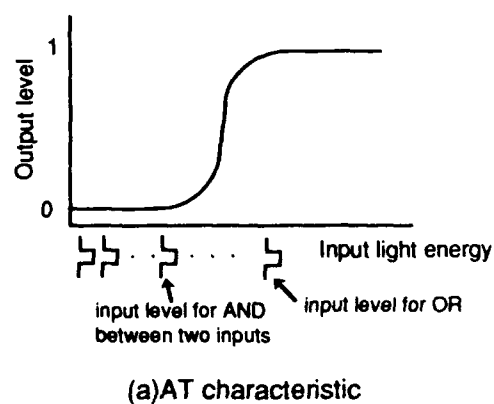
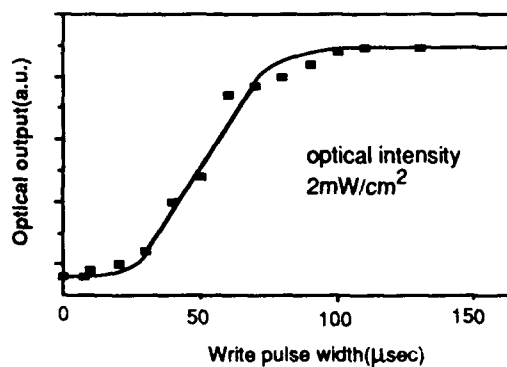


Fig.2 Thresholding characteristics

Fig.3 AT characteristic of GaAs/FLC-SLM
(to be continued)

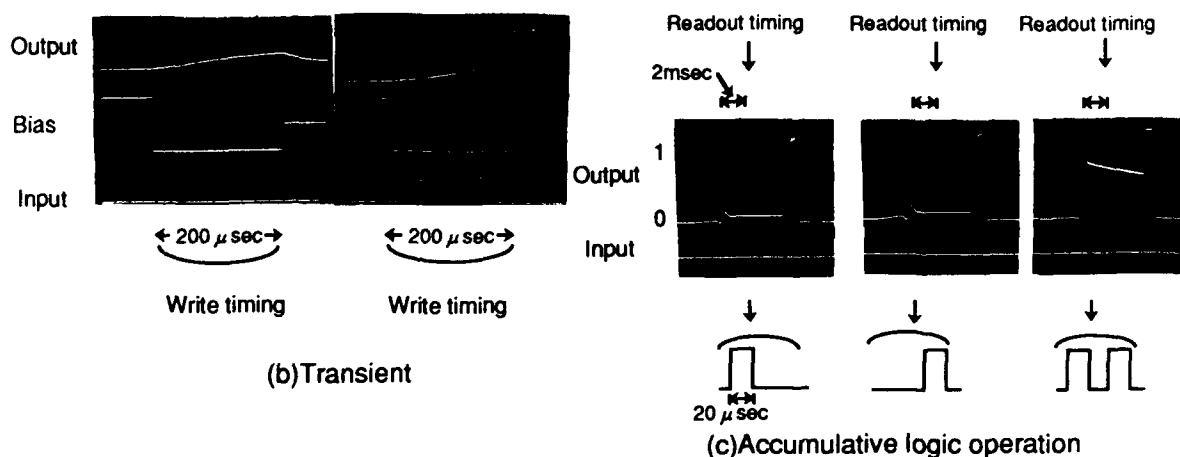


Fig.3 AT characteristic of GaAs/FLC-SLM (continue)

(b) and (c) show experimental AT characteristic. In (c), logic AND is executed between two sequential inputs.

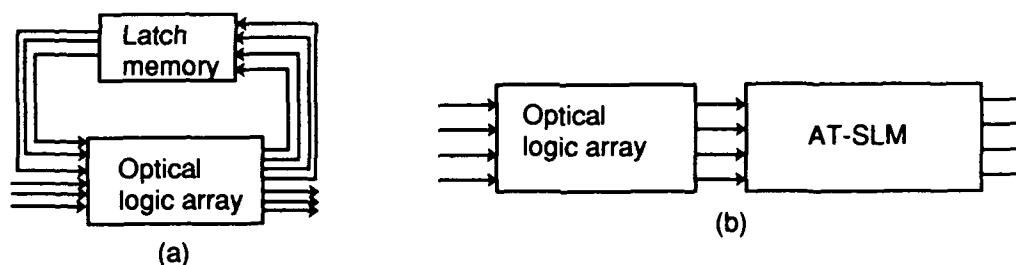


Fig.4 Architecture for sequential logic operation

(a) Existing architecture for sequential logic operation.

(b) Optical parallel processor using AT-SLM.

Optical latch memory and feedback loop are replaced with AT-SLM.

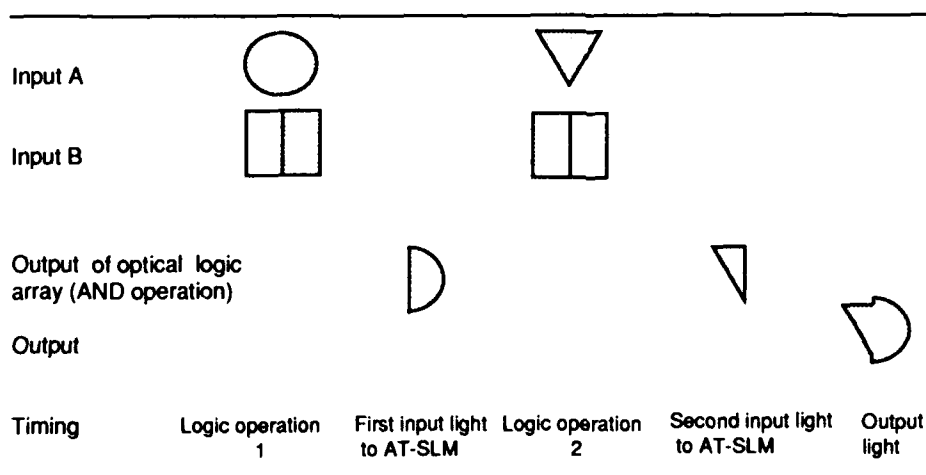


Fig.5 Execution of sum of products with AT-SLM

First, logic AND between two inputs A and B is executed by optical logic array. Second, output light of AND is introduced to AT-SLM. After repeating this step, sum of products, in this case two products, is performed.

Wednesday, March 6, 1991

Joint Plenary Session on Optical Computing and Photonic Switching: I

WC 1:30pm–3:00pm
Salon F

B. Keith Jenkins, *Presider*
University of Southern California



Progress in optical interconnection technologies and demonstrators under the ESPRIT II OLIVES programme

J.W. Parker
OLIVES Consortium, c/o STC Technology Ltd., London Road,
Harlow, Essex, England, CM17 9NA.

Introduction

OLIVES (Optical Interconnections for VLSI and Electronic Systems) is a three year collaborative project which commenced in January 1989 and combines the complementary skills of four major electronics companies (STC, Siemens, Plessey and Thomson-CSF), a chemical company (Akzo) and five academic institutions (University College London (UCL), Foundation for Research and Technology, Hellas/RCC, Centro Nacional de Microelectronica (CNM), Interuniversitair Microelectronisch Centrum (IMEC), and Eidgenössische Technische Hochschule Zurich (ETH)). Some of the key aims and achievements to date in this project are described.

Subsystem Demonstrators

Four major demonstrators of optical interconnect subsystems are under construction, each at a different level within the hierarchy of system construction. These are supplemented by major technology demonstrators of low-power high-density optical interfaces, described in the following section, and of GaAs/Si technology.

1. Module/Subsystem Interconnects - The Optical Bus Demonstrator

Figure 1 shows an optical realisation of a conventional electrical time division multiplexed bus. A number of nodes (eight in this case) separated by 0.5 - 5 m are connected by multimode ribbon fibre through an array of passive star couplers. The demonstrator will have an aggregate bit rate of 6 Gbits/s but the same technology is capable of total rates up to 32 Gbits/s, which exceeds the projected performance of even the most ambitious electrical busses. Multiple instances of the basic unit shown could be combined to achieve a total rate of 100s of Gbits/s.

At each node is a compact array transmitter and receiver module based on silicon motherboard opto-hybrid technology¹. This uses silicon v-grooves to align the fibres and provide reflective structures, solder bump self alignment of the laser and receiver arrays, and a high density interconnect on the silicon substrate to make electrical connections to the hybridized driver chips and passive components. This is the key to achieving the integration density required to minimize the size and power dissipation of the modules. The simulated dissipation of a transmitter array hybrid operating at 32 Gbits/s is 15W, and that of the receiver is similar.

2. Backplane Interconnects - The Mastercard Demonstrator

Figure 2 shows the concept of the mastercard demonstrator² for backplane interconnects. The 'mastercard', which is fabricated from a conventional borosilicate mask plate as used in the micro-electronics industry, is provided with computer generated holographic elements. These holograms, which have a grating constant of 1.2 μ m, deflect the beams to direct them to the required electronic daughterboard and split the power to provide fanout. Collimation optics, contained in the packages of the emitters and receivers, eliminates the requirement to realise this function holographically and improves the overall performance. Early mastercards, assembled

for clock distribution with a fanout of 4, gave a measured optical clock skew of 100ps, an excess loss of 7 dB and a non-uniformity between the 'receiving' elements of 0.9 dB.

The principal advantage of this scheme is the reduction of the volume required for the interconnection. For example, in an 8 board system, the clock distribution network occupies some 120 cm³ in conventional electronic technology (using co-axial cables), while the optical mastercard occupies only 5 cm³. The gains when more boards are required, or where multiple data paths are to be provided, are even more spectacular.

3. Board/MCM Interconnects - The Waveguide Array Demonstrator

Figure 3 shows a schematic of an optical overlay to a silicon multichip module³. Conventional electrical interconnects are used for the short distance interconnections, while arrays of silica-on-silicon waveguides in the overlay provide the long distance parallel data connections. The demonstrator will comprise an eight channel parallel link using a single mode waveguide array, laser diode arrays and photo-receiver arrays. The pitch of these arrays is 125 μ m. The waveguides are fabricated by flame-hydrolysis. Silicon micro-etching is used to produce submounts and alignment features for the assembly of the demonstrator.

4. Chip Interconnects - The Chip Level Clock Distribution Demonstrator

Within a single chip the delay (typically 0.5 ns or more with state of the art technology) caused by the conversion of electrical signals to optical signals and back again makes the use of optical interconnection for data unattractive in most instances. Clock signals, however, are distinguished by a requirement to minimise differential delay. The superior fanout capability of optics allows electrical buffer stages to be eliminated and path length differences minimised, the main sources of chip-level skew. Figure 4 is a schematic of the chip level clock distribution demonstrator⁴. A laser diode adjacent to the chip is reflected onto a multiplexed computer generated holographic element, realised as a relief structure etched in silicon with an SF₆ plasma and metallised with Ti/Au to improve the reflectivity. The light is focused by the hologram onto four photodiodes on the chip. A diffraction efficiency into the first order of 39% has been measured with the binary holograms realised to date, close to the theoretical maximum of 40%. Calculations of the improvement in clock skew in a typical chip give an estimated reduction from 2 ns to 350 ps with a fanout of 17 which is easily achievable, corresponding to an increase in maximum speed from 50 MHz to nearly 300 MHz.

Technology Development

The programme includes a significant effort devoted to optimisation of the optical pathways (i.e. holographic elements and waveguides), parts of which are described above. In addition, there are tasks to develop specific optoelectronic components and component hybridization techniques.

1. Optoelectronic Interfaces

A key component of several of the demonstrators is a receiver array. A monolithic 8-element array has been designed and fabricated⁵ on a commercial ECL process. One variant of this is designed for solder-bump mounting of a photodetector array which was also fabricated within the programme. The entire 8-element array is about 2.3 mm square and gives an ECL-compatible output. The first samples of this device have a measured total power consumption of 280 μ W, including the output buffers, a delay of 1.4 ns, and a minimum input level for the '1' state of 7 μ A. No measurable inter-channel crosstalk has been detected. The area of a single channel is equivalent to that of 4 ECL gates. This device (or its wirebond variant) will be used

in several of the demonstrators. In addition, CMOS receivers have been designed operating at 50 MHz with a power consumption of only 1.17 mW per channel and a 4 μ A sensitivity⁶.

Figure 5 shows an array of 64 reflective MQW modulators based on the asymmetric Fabry-Perot design^{6,7}. These are substrate entry devices designed for flip-chip mounting. With a 5V drive signal, a contrast ratio of up to 3 dB with a loss of 2 dB has been achieved. This was with a device of 100 μ m diameter having 47 quantum wells of 110Å. A free space interconnection of adjacent VLSI chips will be assembled, based on this type of modulator, to demonstrate the potential of these devices for optical interconnects with very low power.

2. Component Hybridization

Several methods for the precision mounting of optoelectronic components are under development within the project. The most flexible of these is solder bump mounting, where the surface tension of molten solder is used to pull the components into precise alignment. Figure 6 shows a 3-layer assembly made with this technique. This comprises a modulator array, similar to that described above, flip chip bonded onto a silicon mount together with a (simulated) diffused glass array, using a combination of high melting point (300°) and low melting point (180°) SnPb solders⁶. The alignment accuracy between the top and bottom layers was assessed using verniers and found to be better than 2 μ m. A flux-less technique for the mounting of lasers and laser arrays based on AuSn eutectic solder has also been developed¹, and arrays have been mounted using this process with no observable performance degradation.

Other Activities

Other activities include a critical assessment of the demonstrators against the system requirements of the industrial partners and the investigation of certain other possibilities, notably direct, high density, free space interconnects between adjacent parallel boards and optical backplane busses. In addition there is an ambitious task to demonstrate the technology for monolithic integration of MQW modulators with CMOS circuitry. A key achievement of this activity is the demonstration of both growth and pre-growth substrate preparation at a temperature of less than 400°C.

Acknowledgements

The author would like to acknowledge the assistance of all the members of the OLIVES team. The work was supported by the Commission for the European Communities under project 2289.

List of References

1. Peall R.G. et al., International Conference on High Performance Electronic Packaging, London, Nov. 1990.
2. Sebillotte, C., International Symposium on Advances in Interconnects and Packaging, Boston, Nov 1990.
3. Karstensen, H. et al., PROC SPIE 1281.
4. Zarschisky, H. et al., International Symposium on Advances in Interconnects and Packaging, Boston, Nov 1990.
5. Wieland, J and Melchior, H., ibid.
6. Goodwin, M.J. et al., ibid.
7. Whitehead, M and Parry, G., Electronic Letters, 25, p567, 1989.

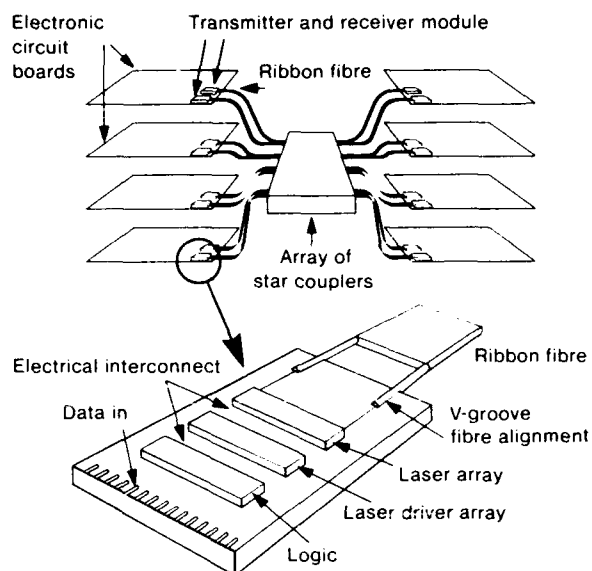


Fig. 1 Module interconnects - optical bus

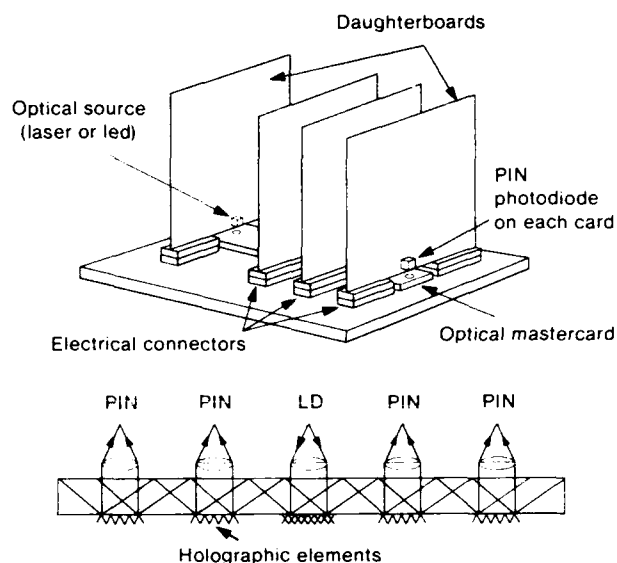


Fig. 2 Backplane interconnects - mastercard

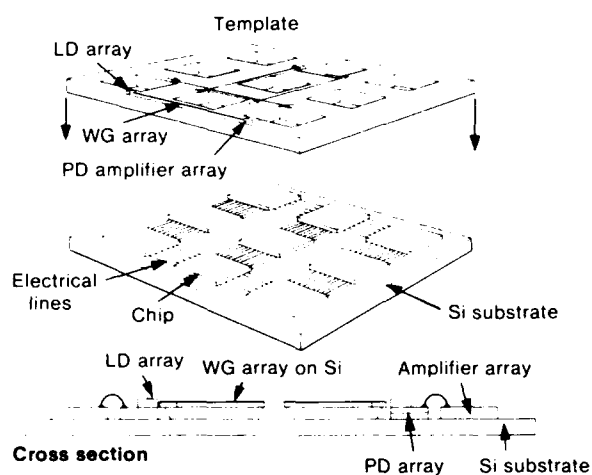


Fig. 3 MCM interconnects - waveguide array

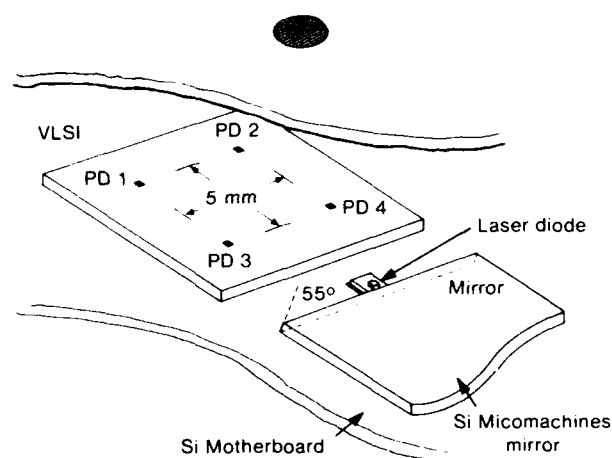


Fig. 4 Chip level clock distribution

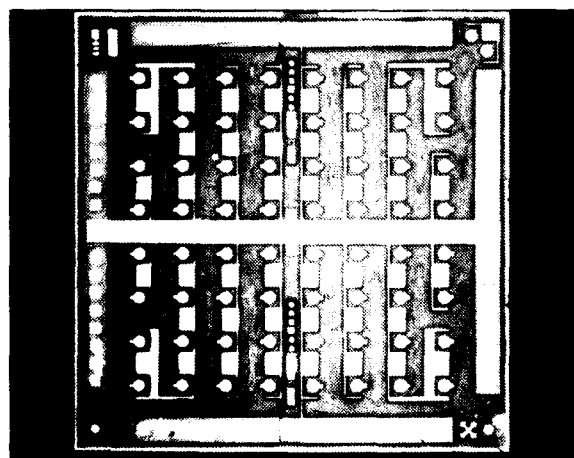


Fig. 5 64 element MQW modulator array

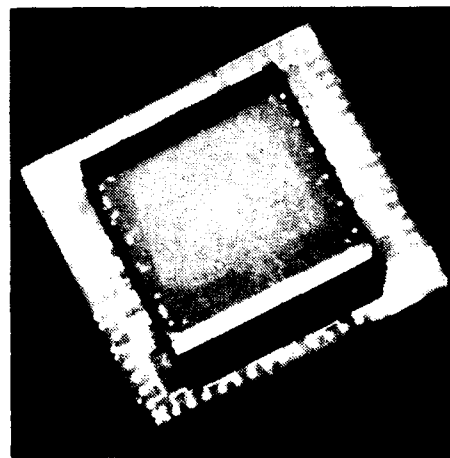


Fig. 6 Three layer solder - bumped assembly

Recent advances in communication network theory

F. T. Leighton

Massachusetts Institute of Technology

Cambridge, MA 02139

In this survey of recent advances in communication network design and algorithms for message routing, emphasis is placed on a novel class of randomly-connected networks known as *multibutterflies*.

Wednesday, March 6, 1991

Joint Plenary Session on Optical Computing and Photonic Switching: II

WD 3:30pm-5:00pm
Salon F

Joseph Goodman, *Presider*
Stanford University



Flexible and fast switching

H. T. Kung
Carnegie Mellon University
Pittsburgh, PA 15213

Ultrafast All-Optical Fiber Soliton Logic Gates

M.N. Islam, C.J. Chen and C.E. Socolich
AT&T Bell Laboratories, Room 4D-411
Holmdel, N.J. 07733

We demonstrate a 5.8pJ switching energy all-optical NOR-gate with a fanout of six that is based on timing shifts from soliton dragging [1] in a fiber. This three-terminal, cascable gate satisfies all requirements for a clocked digital optical processor. Furthermore, we show that soliton dragging logic gates are one embodiment in fiber form of a novel switch architecture of time domain chirp switches (TDCS). Although TDCS have a long latency, for high bit-rate applications TDCS lead to switching energies approaching one Picojoule.

The logic gate operates based on time shifts from soliton dragging in a clocked digital system. In time shift keying a "1" corresponds to a pulse that arrives within the clock window and a "0" either to no pulse or an improperly timed pulse. In soliton dragging two temporally coincident, orthogonally polarized pulses interact in the fiber through cross-phase modulation [2] and shift each others velocities. The velocity shift converts into a time shift after propagating some distance in the fiber. For the NOR-gate the fiber length is trimmed so that in the absence of any signal the power supply or control pulse C arrives within the clock window and corresponds to a "1". When either or both signals are incident, they interact with the control pulse through soliton dragging and pull C out of the clock time window.

The insert in Fig. 1 shows a schematic of the NOR-gate that consists of two birefringent fibers connected through a polarizing beam splitter with the output filtered by a polarizer. The

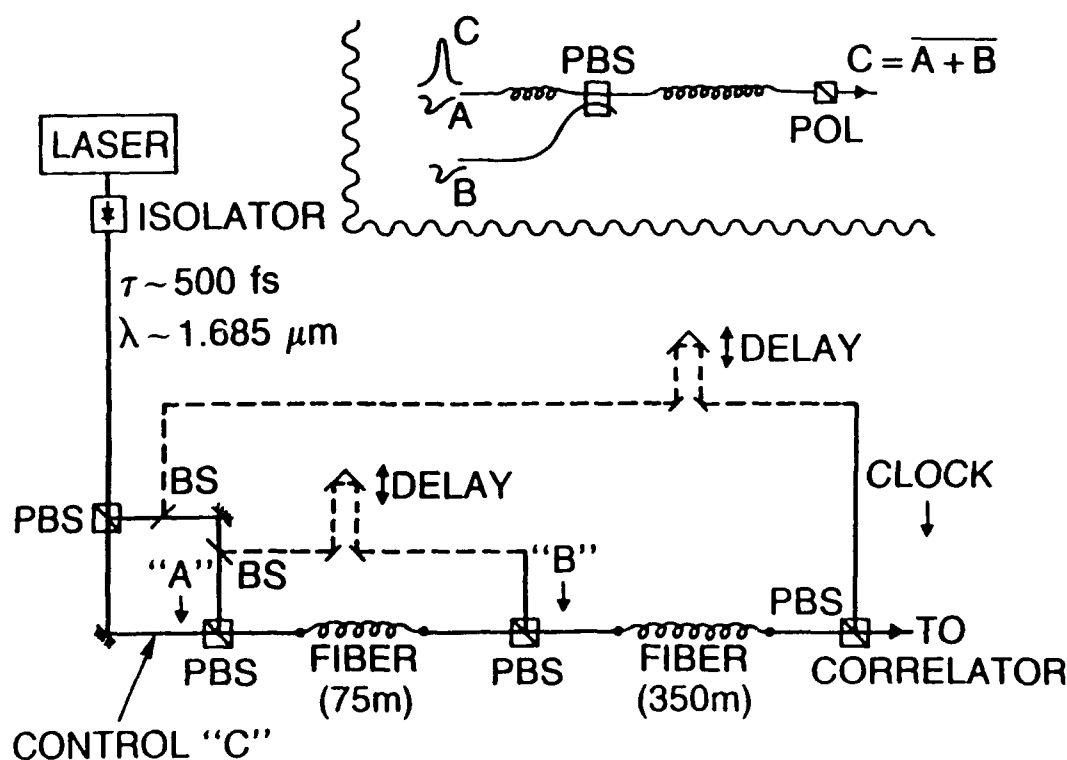


Fig. 1. Experimental configuration for testing an all-optical NOR-gate. The insert shows a simplified schematic of the NOR-gate.

control pulse C provides gain and logic level restoration, propagates along one principal axis in both fibers and corresponds to $A \text{ NOR } B$ at the output. The two signal pulses A and B are polarized orthogonal to C and are blocked by the polarizer at the output. The signals are timed so that A and C coincide at the input to the first fiber and B and C coincide (in the absence of A) at the input to the second fiber.

Figure 1 shows the experimental apparatus for testing a single NOR-gate. We obtain $\tau \sim 500\text{fsec}$ pulses near $1.685\mu\text{m}$ from a passively mode locked NaCl color center laser in which a 2mm thick quartz birefringent plate limits the bandwidth and, thus, intentionally broadens the pulses [3]. The input stage separates the control C, signals A and B, and clock beams, and stepper motor delay stages are used to time properly signal B and the clock. The two fibers are 75m and 350m long, have a polarization dispersion of about 80psec/km , and exhibit a polarization extinction ratio better than 14:1. The control pulse output and the clock are directed to a correlator to measure the time shifts.

The correlation of the clock with the NOR-gate output is illustrated in Fig. 2. The dotted box corresponds to the clock window, and we see that C arrives within this window when no signal is present. When $A=1$ or $B=1$, C shifts between 2 to 3 psec out of the clock window; the shift from A is larger since C can time shift in both fibers. When $A=B=1$, C shifts by about 4psec. In this example the signal energies are 5.8pJ each and the fanout or gain (control out / signal in) is six. The control pulse energy in the first fiber is 54pJ and is reduced to 35pJ in the second fiber because of coupling losses.

To prove the cascability and fan-out of the logic gate, we implemented an all-optical multivibrator or ring oscillator by connecting the NOR-gate as an inverter and feeding the output back to the input ($A = 0$, $B = \text{previous output from gate}$). We placed a 50:50 beam splitter at the output and sent half of the output through a delay line to the B input. The correlator was set to the center of the clock time window. As Fig. 3 shows, with the feedback blocked the

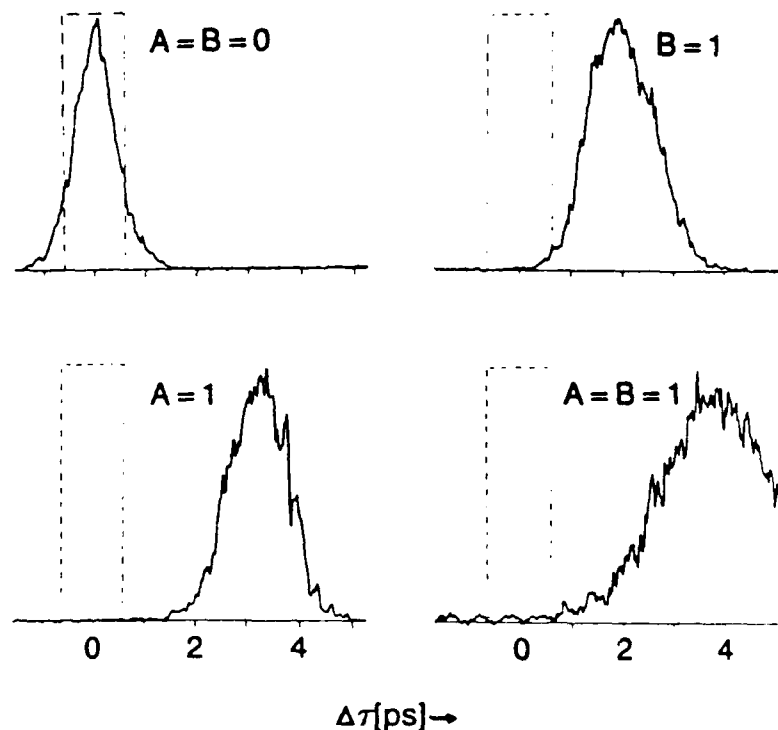


Fig. 2. Correlation of clock with NOR-Gate output.

output is a string of 1's. When the feedback is added, the output becomes an alternating train of 1's and 0's whose period is twice the fiber latency (1.75 μsec).

Soliton dragging logic gates are one example of a more general switch architecture of TDCS that is applicable to materials other than fibers. As shown in Fig. 4, the TDCS consists of a nonlinear chirper followed by a soliton dispersive delay line and has two orthogonally polarized inputs. In the absence of a signal pulse, the control pulse propagates through both sections and arrives at the output within the clock window. For a cascable switch the self-induced chirps on the control in both sections must balance, and the output pulse must resemble the input. Adding the signal pulse creates a time varying index change that chirps the control pulse and shifts its center frequency [2]. Then, as the control pulse propagates through the soliton dispersive delay line, the frequency shift is translated into a time change. Since a fundamental soliton acts as a particle, even a slight shift in the center frequency can cause the complete soliton to shift in time, which results in good contrast within the clock window. Furthermore, since the chirps from group velocity dispersion and nonlinearity are balance for a soliton, cascability for the control pulse can be satisfied using solitons.

The key feature of TDCS is that for high-bit rate, short pulse applications the TDCS requires less nonlinear interaction and, consequently, less switching energy than other all-optical switches such as Mach-Zehnder interferometers. For example, the rule of thumb for Mach-Zehnder interferometers is that a π -phase shift must be achieved through the interaction between two pulses in less than an absorption length. However, by using solitons in a TDCS we find that the nonlinear interaction in our demonstrated switch is less than $\pi/20$ [4]. Because solitons shift as a unit, solitons permit the effect in the nonlinear chirper to be accumulated through the entire length of the dispersive delay line. The trade-off is that TDCS have a long latency, which restricts their usage to feed-forward applications.

In summary, we have presented a TDCS that performs logic using time shift keying in a clocked digital optical processor. Soliton dragging is one example of a TDCS and has yielded the lowest switching energy of any all-optical gate because of the separation between the nonlinear interaction and the soliton dispersive delay line. The three-terminal NOR-gate has a switching energy of 5.8pJ, fanout of six and restores logic levels and timing at the output.

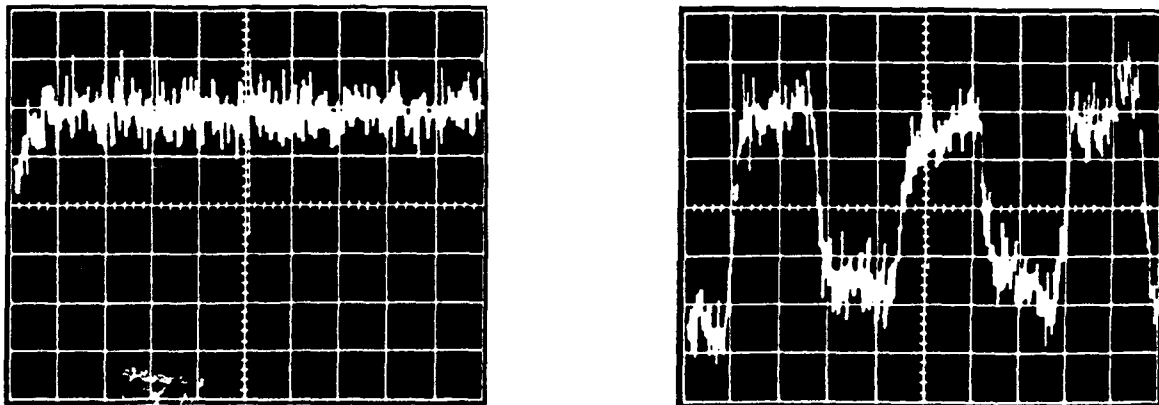


Fig. 3 Output from an all-optical multivibrator in which a NOR-gate is configured as an inverter and the output is fed back to the input. Feedback blocked on the left and feedback added on the right.

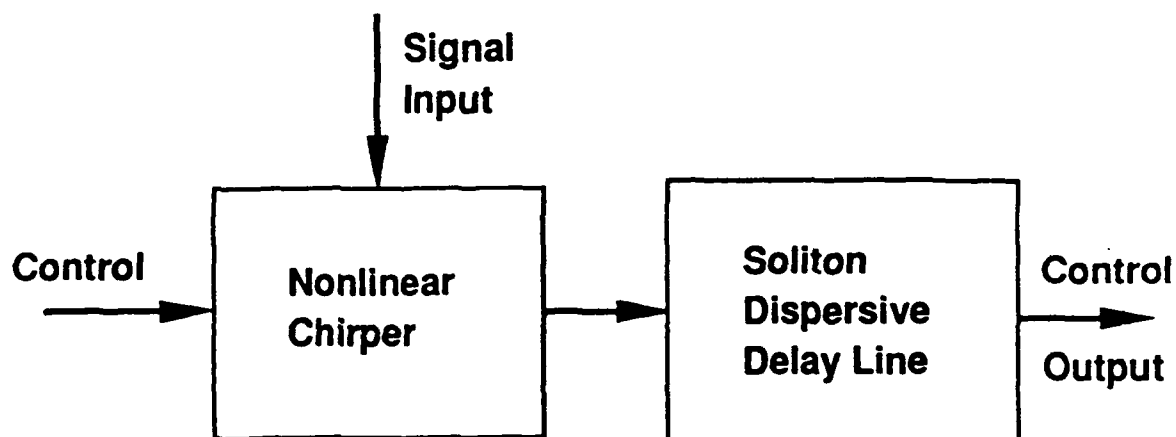


Fig. 4. General architecture for an all-optical time domain chirp switch (TDCS).

References

- [1] M.N. Islam, Opt. Lett. **15** , 417 (1990); M.N. Islam, C.E. Socolich and D.A.B. Miller, Opt. Lett. **15** , 909 (1990).
- [2] M.N. Islam, L.F. Mollenauer, R.H. Stolen, J.R. Simpson, and H.T. Shang, Opt. Lett. **12** , 625 (1987).
- [3] M.N. Islam, E.R. Suderman, C.E. Socolich, I. Bar-Joseph, N. Sauer, T.Y. Chang and B.I. Miller, IEEE J. Quantum Electron. **25** , 2454 (1989).
- [4] M.N. Islam, C.J. Chen and C.E. Socolich, "All-Optical Time-Domain Chirp Switches," (submitted to Opt. Lett.).

Alspector, Joshua – WB1
Athale, Ravindra A. – MC, TuB3, TuC2

Bann, S. – TuD5
Berra, P. Bruce – ME11, TuC1
Borghs, G. – ME5
Brenner, K.-H. – MB2, TuB4
Brodsky, Stephen A. – ME8
Bryan, R. P. – MA3
Buller, G. S. – TuA4

Carson, R. F. – ME5
Chen, C. J. – WD2
Cheng, Julian – MA3
Cheng, Li-Jen – ME29
Chiarulli, Donald M. – TuC3
Chino, Ken'ichi – ME2
Chiou, Arthur E. – TuD4
Chipman, Russell A. – ME25
Chirovsky, L. M. F. – MA2
Collings, N. – ME9
Craig, R. G. A. – TuA4

Dagenais, Mario – MA, ME17
Dändliker, R. – ME9, ME24
D'Asaro, L. A. – MA2
Däschner, Walter – MB3
Dezialoshinski, Lior – ME28
Dickinson, Alex – TuC4
Doroski, D. – ME7

Efron, Uzi – MD
Ehbets, P. – ME24
Eichmann, George – MA5
Esener, Sadik C. – ME12, TuB2, TuD1

Fainman, Yeshayahu – ME2, WA2
Feinberg, Jack – WA2
Feldman, Michael R. – ME16
Focht, M. W. – MA2
Ford, Joseph E. – ME2, TuD1, WA2
Freund, J. M. – MA2
Fujimori, Susumu – ME21
Fukuda, Yukio – WB4

Giles, Lee – ME
Glaser, I. – TuB5
Gmitro, A. – ME6
Gomes, Cleber M. – ME4
Goodman, Joseph W. – MC3, WD
Griffin, R. D. – ME31
Gu, Claire – ME3
Guest, Clark C. – ME8
Guilfoyle, Peter S. – TuA5
Gupta, Vipul – TuA2
Guth, G. D. – MA2

Ha, Berlin – TuD2
Hamanaka, Kenjiro – MB4
Haney, Michael W. – TuC2
Hashimoto, Masashi – WB4
Herzig, H. P. – ME24
Heuring, Vincent P. – ME14
Ho, Francis – WB3
Hong, John – ME26
Huignard, J. P. – TuD5

Ishibashi, Shigeki – WB4
Islam, M. N. – WD2
Itoh, Hideo – ME13

Jahns, Jürgen – MB3
Jared, David A. – MD2, ME7
Jenkins, B. Keith – TuA3, WC
Johnson, Kristina M. – MD2, ME7, WB

Kar-Roy, A. – TuD3
Kasahara, K. – MA1
Kawai, Shigeru – ME20
Keller, Paul – ME6
Kiamilev, Fouad E. – ME12
Kim, Dai Hyun – MA5
Kim, Jae – WB3
Kitayama, Ken-ichi – WB4
Kobayashi, Shunsuke – ME4
Kosko, Bart – TuB1
Kostrzewski, Andrew – MA5
Krishnamoorthy, A. V. – TuD1
Kung, H. T. – WD1
Kuznia, C. B. – MC2
Kwong, Norman S. Z. – ME29
Kyuma, Kazuo – WB2

Launay, J. C. – ME1
Lee, Hyuek-Jae – ME10
Lee, J. N. – ME31
Lee, Sing H. – ME2, ME12, TuB2, TuD1, WA2
Lee, Soo-Young – ME10
Leibenguth, R. E. – MA2
Leighton, F. T. – WC2
Lentine, A. L. – MA2
Levitan, Steven P. – TuC3
Li, C. -S. – ME15
Li, Yao – MA5, TuD2
Lin, Freddie – ME18
Lin, Steven – WB3
Lin, Ting-Ting Y. – ME12
Liu, Duncan T. H. – ME29
Liu, William – ME18
Lohman, Gary E. – TuB4
Lohmann, Adolf W. – MB, MC1
Lu, Dau-Tsuong – ME12
Luke, Keung L. – ME29

Ma, Jian – ME2, WA2
 MacKinnon, G. – TuA4
 Maeda, H. – ME4
 Mahlab, Uri – ME30
 Mait, Joseph N. – TuB3
 Majidi, Masoud – TuA2
 Marsden, Gary C. – TuB2, TuD1
 Mathey, P. – ME1
 McArdle, N. – TuA4
 McKnight, D. J. – TuA4
 Melhem, Rami G. – TuC3
 Meredith, P. – TuA4
 Messerschmitt, D. G. – ME15
 Miller, D. A. B. – MA2
 Miller, J. M. – TuA4
 Mitkas, Pericles A. – ME11
 Mizoguchi, Masanori – ME20
 Moddel, G. – ME7
 Mori, Masahiko – ME13
 Morris, James E. – ME16
 Morrison, Rick L. – ME22
 Mukai, Seiji – ME13
 Murdocca, Miles – TuA2

Nahtomi, Ehud – ME28
 Neff, John A. – PdP
 Neifeld, Mark A. – WA4
 Nishida, Yasuhide – ME21
 Nitta, Yoshikazu – WB2
 Nonaka, Koutarou – ME21

Ogura, I. – MA1
 Ohta, Jun – WB2
 Okumura, Fujio – ME4
 Olbright, G. R. – MA3
 Olson, Brita – TuB2
 Ozaktas, Haldun M. – MC3

Paek, Eung Gi – WA3
 Pan, Zeqi – ME17
 Pankove, Jacques – ME5
 Pape, Dennis R. – ME32
 Parker, J. W. – WC1
 Parry, G. – MD1
 Patel, J. S. – WA3
 Pauliat, G. – ME1
 Pedrini, G. – ME9
 Pezzaniti, J. Larry – ME25
 Prather, Dennis W. – TuB3
 Prise, Michael E. – TuA
 Prongué, D. – ME24
 Przybyłek, G. J. – MA2
 Psaltis, Demetri – MA4, WA1, WA4, WB3

Radehaus, C. – ME5
 Rajbenbach, H. – TuD5
 Redmond, I. R. – TuA4
 Refregier, Ph. – ME27

Restall, E. J. – MC4, TuA4
 Rivers, A. – MD1
 Rohlev, Anton – ME5
 Roberts, J. S. – MD1
 Robertson, Brian – MC4, ME23
 Robinson, M. G. – ME7
 Roos, Edward V. – TuA5
 Roosen, G. – ME1
 Rosen, Joseph – ME28
 Rudokas, Ronald S. – TuA5

Sakamoto, M. – ME4
 Sasaki, Hironori – WA2
 Sawchuk, A. A. – MC2
 Sekine, Hiroyuki – ME4
 Shamir, Joseph – ME28, ME30
 Shin, Sang-Yung – ME10
 Sinzinger, Stefan O. – MC1
 Slagle, Tim – WA5
 Smith, L. E. – MA2
 Smith, S. D. – TuA4
 Snowden, J. F. – TuA4
 Socolich, C. E. – WD2
 Soffer, Bernard – WA
 Soukup, Todd J. – ME14
 Stirk, Charles W. – MA4
 Stone, Richard V. – TuA5
 Strzelecki, Eva M. – ME18

Taghizadeh, M. R. – MC4, ME23, TuA4
 Tai, Shuichi – WB2
 Takahashi, Masanobu – WB2
 Taketomi, Yoshinao – ME2, WA2
 Tong, F. – ME15
 Tsai, C. S. – TuD3
 Tsujikawa, Susumu – ME4
 Tsang, Dean Z. – ME19
 Turner, Richard – MD2
 Turunen, Jari – ME23

Vasara, Antti – ME23
 Veldkamp, Wilfrid B. – MB1

Wagner, Kelvin – TuB, WA5
 Wakelin, S. – TuA4
 Welker, A. C. – MC4, TuA4
 Walker, Sonya L. – ME22
 Warde, Cardinal – TuC
 Watanabe, Masanobu – ME13
 Waterson, Clare – TuA3
 Weible, K. J. – ME9
 Westerholm, Jan – ME23
 Wherrett, B. S. – TuA4
 Whitehead, M. – MD1
 Wichart, S. – ME7
 Wilson, R. A. – TuA4
 Woodbridge, K. – MD1
 Wullert, John R. II – WA3

Xue, W. – ME9

Yajima, Hiroyoshi – ME13

Yamanaka, Y. – MA1

Yamazaki, Takashi – ME4

Yayla, G. – TuD1

Yeh, Pochi – ME3, ME26, TuD, TuD4

Zouganeli, E. – MD1

**Optical Computing
TECHNICAL PROGRAM COMMITTEE**

C. Lee Giles, *General Cochair*
NEC Research Institute

Sing H. Lee, *General Cochair*
University of California, San Diego

B. Keith Jenkins, *Program Chair*
University of Southern California

H. Arsenault
University Laval, Canada

R. A. Athale
George Mason University

L. A. Coldren
University of California, Santa Barbara

M. Dagenais
University of Maryland

U. Efron
Hughes Research Laboratory

I. Glaser
Tel Aviv University, Israel

L. Hesselink
Stanford University

H. S. Hinton
AT&T Bell Laboratories

A. Husain
Honeywell Systems and Research Center

Y. Ichioka
Osaka University, Japan

S. Ishihara
Electrotechnical Laboratory, Japan

R. K. Kostuk
University of Arizona

K. Kyuma
Mitsubishi Electric Corporation, Japan

F. J. Leonberger
United Technologies Research Center

A. W. Lohmann
NEC Research Institute

W. J. Miceli
Office of Naval Research

J. Midwinter
University College London, UK

J. A. Neff
Dupont Corporation

M. E. Prise
AT&T Bell Laboratories

G. Roosen
Institut D'Optique, France

C. M. Verber
Georgia Institute of Technology

K. H. Wagner
University of Colorado

C. Warde
Massachusetts Institute of Technology

P. Yeh
University of California, Santa Barbara