AD-A235 013

NTATION PAGE

Form Approved OMB No. 0704-0188

ted to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

RT DATE

3. REPORT TYPE AND DATES COVERED
FINAL 01 Sep 87 TO 30 Sep 90

4. TITLE AND SUBTITLE
Computational Complexity and Efficiency in Electro-Optical Computing Systems

5. FUNDING NUMBERS
AFOSR-87-0386

6. AUTHOR(S)
Professor John Reif

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Department of Computer Science
Duke University
Durham, NC 27706

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NE
Bldg 410
Bolling AFB Washington DC 20332-6448
Dr. Alan E. Craig

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
2305/B1

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

(1) To develop robust theoretical model for a wide class of electro-optical computing systems
(2) To extend the known capabilities, by design of new, more efficient algorithms for electro-otpical computing using less time, volume and energy. In particular, to develop efficient algorithms that use optimal combinations of time, volume and energy on electro-optical computing systems
(3) To determine the fundamental theoretical limitations and capabilities of electro-optical computing systems.
In particular, to determine lower bounds on tradeoffs between volume, time, and other resources (such as energy) of any electro-optical computing system to solve fundamental problems.

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500 Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std. Z39-18

Final Report
Feb, 1991

Contract Number:
AFOSR-87-0386

Title of Contract:
Computational Complexity and Efficiency
in Electro-Optical Computing Systems

PRINCIPAL INVESTIGATOR:
John H. Reif
Department of Computer Science
Duke University
Durham, North Carolina 27706
email: reif@cs.duke.edu
(919) 660-6568

CONTRACTING OFFICER:
Alan Craig
Department of the Air Force
Air Force Office of Scientific Research (AFOSR)/NE
Buiding 410
Bolling Air Force Base
Washington, DC 20332-6448
(202) 767-4931

A-1

Final Report

AFOSR Contract:

Computational Complexity and Efficiency in Electro-Optical Computing Systems

John Reif, Duke University

0 Abstract of RESEARCH APPROACH and Objectives:

(1)To develop robust theoretical model for a wide class of electro-optical computing systems

(2)To extend the known capabilities, by design of new, more efficient algorithms for electro-optical computing using less time, volume and energy. In particular, to develop efficient algorithms that use optimal combinations of time, volume and energy on electro-optical computing systems

(3)To determine the fundamental theoretical limitations and capabilities of electro-optical computing systems.

In particular, to determine lower bounds on tradeoffs between volume, time, and other resources (such as energy) of any electro-optical computing system to solve fundamental problems.

1 Summary of Technical Progress during this Contract:

Work by Reif in optical computing has been in five areas:

(A). Efficient Optical Algorithms

(A.1) the VLSIO model

(A.2) Efficient Electro-Optical Algorithms in the VLSIO model

(B) Lower bounds for Optical Computation

(B.1) Lower Bounds for the Volume of Electro-Optical Devices in the VLSIO model

(B.2) Lower Bounds for the energy consumption of Electro-Optical devices in the VLSIO model.

(C) The Ray Tracing Problem

(D) Optical Memory Storage and Computation Using Fiber Optic Delay Loops

(E) Holographic Based Computing

(E.1) Reif's Holographic Message Routing System

(E.2) Holographic Memory Storage

(E.3) Optical Expanders

We give the details in the following.

(A). Efficient Optical Algorithms
(A.1) the VLSIO model

Our goal is to determine the fundamental theoretical limitations and capabilities of optical computing systems. Our first step is to develop a robust theoretical model for a wide class of electro-optical computing systems. [Barakat and Reif,1987] developed a new model for Electro-Optical devices, known as VLSIO. The VLSIO model includes both electrical and also optical components; that is it allows combinations of 2D VLSI chips as well as optical devices such as lenses and holograms. The VLSIO model allows us to compare the time, volume and energy of a wide variety of distinct electro-optical systems.

No other model had been previously invented. The VLSIO model allows one to give a precise comparisons between proposed optical algorithms, using well defined metrics such as time, volume and energy.

This is a new model of computation and we expect that the growth in the optical technology during this decade would spur growth in algorithm research.

See appendix A.1 for more details.

(A.2) Efficient Electro-Optical Algorithms in the VLSIO model

Our goal here is to extend the known capabilities of electro-optical devices, by design of new, more efficient algorithms for electro-optical computing systems in the VLSIO model. This requires that we develop algorithms that make optimal tradeoffs between key resources of time, volume and energy. We used both known techniques from VLSI algorithms as well as the special 3D properties of optical devices in the VLSIO model.

[Barakat and Reif, 87] developed efficient new VLSIO algorithms using small volume and constant time for matrix multiplication and other matrix problems. Recently [Reif and Tyagi,90] they developed efficient optical algorithms for a much larger class of fundamental problems(including most problems found in standard algorithm texts), which occur frequently in practice.

Actually we consider the two models of computation—VLSIO and DFT-Circuit. We describe both algorithms for a set of direct applications of DFT, as well as algorithms that seem unrelated to the DFT; in particular two sorting algorithms, an algorithm for the element distinctness, and also both one dimensional and two-dimensional string matching algorithms. We compare the performance of DFT-VLSIO algorithms with the known VLSIO lower bounds. In many cases, these are near optimal and much more efficient than other optical algorithms previously proposed and in some cases our algorithms are optimal. See Appendix A.2.

(B) Lower bounds for Optical Computation

Our goal here is to determine lower bounds on volume, time, and other resources (such as energy) of any elecro-optical computing system in the VLSIO model to solve fundamental problems. We strive to get tradeoffs between resources. To do this, we extend techniques developed for obtaining lower bounds for VLSI.

(B.1) Lower Bounds for the Volume of Electro-Optical Devices in the VLSIO model

INITIAL THEORITICAL RESULTS: Previously, [Barakat and Reif,87] showed the first known lower bounds for any optical device to compute various functions of n inputs within time T and volume V in the VLSIO model. This was the first time anyone had given general lower bounds on the volume and time tradeoff of Electro-Optical devices. The lower bounds hold for a large class of problems (known as transitive problems) including sorting, routing, and most other standard combinatorial or algorithmic problems.

(B.2) Lower Bounds for the energy consumption of Electro-Optical devices in the VLSIO model.

[Tyagi and Reif, 1989] recently for the first time proved lower bounds on energy consumption, volume and time for a large class of problems using any possible Electro-Optical devices. This is the first time anyone has given general lower bounds on the energy consumption of Electro-Optical devices. In particular, they showed for time T and energy E, the Product ET is greater than a certain function of the input size and demonstrated matching upper bounds on the ET product for shifting. Again, these lower bounds hold for a large class of problems (known as transitive problems), including sorting, routing, and most other standard combinatorial or algorithmic problems. See Appendix B

(C) The Ray Tracing Problem

In a recent paper, [Reif, Tygar, Yoshida,90] we have investigated a problem that is fundamental for optical system design. In particular, we consider optical systems consisting of a set of refractive or reflective surfaces. The ray tracing problem is, given an optical system and the position and direction of an initial light ray, to decide if a light ray reaches some given final position. We assume the position and the tangent of the incident angle of the initial light ray is rational. For many years, ray tracing has been used for designing and analyzing optical systems. Ray tracing is now also extensively used in computer graphics to render scenes with complex curved objects.

The computability and complexity of various ray tracing problems are investigated. Our results are:

- Ray tracing in three dimensional optical systems which consist of a fixed finite set of curved reflective or refractive surfaces is undecidable, even if all the surfaces are represented by systems of rational quadratic inequalities. However, the problem is recursively enumerable.

- Ray tracing in three dimensional optical systems which consist of a fixed finite set of flat reflective or refractive surfaces is undecidable, if the coordinates of the endpoints of some of surfaces are irrational. However, the ray tracing system is

PSPACE-hard, if we restrict ourselves to surfaces with rational coordinates.

- For any $d \geq 2$, the ray tracing of d dimensional optical systems which consist of a fixed finite set of flat reflective surfaces is in PSPACE, if the positions of all the surfaces are rational, and are placed perpendicular to each other.

For details, see Appedix C.

(D) Optical Memory Storage and Computation Using Fiber Optic Delay Loops

The use of delay loops for memory is an old idea, dating back to the use of mercury storage tubes in the early digital computers of the 50's. Nevertheless it is an becoming an important now for optical computation, since it is one of very few known methods for doing storage completely in the optical domain. The key problem is that data can only be accessed with the delay for the propagation around the loop.

In very new research , Reif and Tyagi have developed efficient algorithms for bit serial optical computers using fiber optic delay lines for auxiliary storage. In particular, they have some very interesting new techniques for using a very small set of optical delay loops to manage the intermediate storage for a wide range of algorithms and computations on interconnect networks. The key new idea is a method for utilizing data just at the right time so that there is no delay for the propagation around the appropriate loop. This extends the work of [Jordan, 1989] at Boulder, who has implemented a delay loop memory system and discussed its use in simulating networks.

[Reif and Tyagi,to appear 90]

(E) Holographic Based Computing

(E.1) Reif's Holographic Message Routing System

This is a very interesting outgrowth of Reif's work in optical computing. See Appendix E for details.

Message routing in a parallel machine concerns providing arbitrary interconnections between its processors. The Connection Machine, for example, is a 65,536 processor bit serial *SIMD* parallel machine, requiring 65,536 messages to be routed to distinct addresses. There is a bottleneck in this information transfer mechanism: the routing time in these parallel machines is approximately a thousand times longer than the instruction time. Optical hardware provides the potential for high bandwidth, low crosstalk and power dissipation for connecting processors at the board-to-board level. It has also been shown that impedance matching requirements favor optics over electronics for fast data transfer.

Previous work on dynamic optical interconnects has employed spatial light modulators (SLMs) in optical crossbars, or volume holograms to re-configure connections in real-time. These two approaches have disadvantages: the former requires setting N^2 switches to achieve the interconnections, while the latter is limited by the slow response time of photorefractive recording materials.

Dynamic holographic architectures for connecting processors in parallel computers have been limited by the response time of the holographic recording media.

In [Reif,90] and [Maniloff, Johnson, and Reif,89] we present a novel optical interconnect architecture, involving spatial light modulators (SLMs) and volume holograms. which uses spatial light modulators to dynamically control the holographic routing of messages between originator and destination processors. This system is not limited by the response time of the volume holographic recording media, which stores the destination address: the routing is achieved as fast as the optical beam can be modulated by the SLM.

Multiple-exposure holograms are stored in a volume recording media, which associate the address of a destination processor on a spatial light modulator with a distinct reference beam. A destination address programmed on the spatial light modulator is then holographically steered to the correct destination processor.

A small prototype of the Holographic Message Routing System was constructed by Maniloff and Johnson at Boulder CO in a collaborative project with Reif. We in [Maniloff, Johnson, and Reif,89] present the design and experimental results of a holographic router for connecting four originator processors to four destination processors. Our first prototype holographic router used ferroelectric liquid crystal (FLC) SLMs to connect four originator processors to four destination processors at 10 kHz.

In [Reif,90] We also present results on reducing the number of switches in the SLM required to route N originator processors to N destination processors in a single time step.

(E.2) Holographic Memory Storage

The use of holography for memory storage is an old idea, but is becoming increasingly practical and exciting due to the use of $LiNbO_3$ crystals which can store from hundreds up to a thousand images, where each image can resolve a page of up to a few megabytes of storage. A key problem in the practical development of holographic memory storage is the use of orthogonal images to address the holographic memory, which is solved by the use of the optical expanders described in E.3 See appendix E.3 for a further discussion of holographic matching and holographic memory storage.

(E.3) Optical Expanders

An Optical Expander is a device that expands the dimension of a pattern space. This is a new idea due to Reif that was motivated by needs of the holographic message routing system but appears to be a very basic problem. An optical expander allows the Holographic Message Routing System to be scaled up to very large sizes using a small (logarithmic number) of address bits. Reif has worked with his student Akitoshi Yoshida and with Barakat on new methods for optical expanders. For more detail, see Appendix E.3

2 Summary of new Research in Spring, 1991

2.1 Optical Memory and Storage

One of the biggest challenges in the electro-optical field to to develop methods for fast memory storage and retrieval, for large amount of data.

2.2 Multi-frequency Optics

The use of multiple frequencies to aid in computation and in optical storage is very intriguing; Reif is just beginning to explore this idea.

2.2.1 Multi-frequency Storage

Using a single fiber optic delay loop of approx a kilometer on a single frequency, up to tens of kilobytes can be stored. It is possible that with the use of multiple frequency up to possibly a megabyte could be stored. Reif is investigating these possibilities.

2.2.2 Multi-frequency Computation

Reif is investigating the use of multi-frequency in general computation; this may decrease the volume required by electro-optical devices. Also, Reif is investigating the use of multi-frequency to allow numerical computations to be done in optics with much higher accuracy. There may be limitations to the use of multi-frequency; Reif is investigating lower bounds as well.

3 Publications:

(1)R. Barakat and J Reif, "Lower bounds on the computational efficiency of optical computing systems", Journal of Applied Optics, Vol 26, p 1015-1018, March 15, 1987.

(2)R. Barakat and J. Reif, "A Discrete Convolution Algorithm for Matrix Multiplication with Application in Optical Computers", Journal of Applied Optics, Vol 26, p 2707-2711, 1987.

(3)E.S. Maniloff, K.M. Johnson, and J. Reif "Holographic Routing Network for Parallel Processing Machines" EPS/Europtica/SPIE International Congress on Optical Science and Engineering, Paris, France, April 1989.

(4)J.H. Reif and A. Tyagi, Efficient Algorithms for Optical Computing with the DFT Primitive in *The 10th Conference on Foundations of Software Technology and Theoretical Computer Science* , Lecture Notes in Computer Science, Springer-Verlag, Bangalor, India, December 1990A.

(5)J.H. Reif and A. Tyagi, An Optical Delay Line Memory Model with Efficient Algorithms, *Advanced Research in VLSI Conference*, MIT Press,March 25-27, Santa Cruz, CA, 1991B.

(6)J. Reif, A. Yoshida, and D. Tygar. The Computability and Complexity of Optical Beam Tracing, *31st IEEE Symposium on Foundations of Computer Science,* Saint Louis, Missouri, October, 1990.

(7)A. Tyagi and J.H. Reif, Energy Complexity of Optical-Computations , appeared in *The 2nd IEEE Symposium on Parallel and Distributed Processing*. Dallis, TX, December 1990.

(8)J. Reif, Optical Expanders Give Constant Time Holographic Message Routing Using O(*N*log*N*) Switches, August 1989, submitted for publication.

(9) J. Reif and A. Yoshida. Optical Expanders with Holographic Memory and Routing Applications, May, 1990, submitted for publication.

4 Personnel

4.1 The Background of the PI:

Reif is a theoretical computer scientist and applied mathematician by training, but is known for working in diverse areas, including robotics and parallel computing, and has written over 80 papers in these areas. His research style is to work on newly developing area, and to contribute basic new models, new lower bound techniques and particularly new and novel algorithmic techniques which can be used in the particular domain.

To solve problems in a new emerging area, Reif has brought to bear to a large number of diverse techniques he has learnt in exploring other related areas (some time obviously related, sometime apparently unrelated). In some cases, Reif's work leads to results that may be practical and that have been implemented. Examples are

(1) the parallel nested dissection algorithm of [Pan and Reif] implemented in [Leiserson et. al, 86] and [Opsahl and Reif, 86]

(2). the massively parallel BLITZEN machine described in [Davis and Reif, 88] and [Blevins et. al, 90], and

(3) the parallel compression described in [Storer and Reif, 88]

(4) as well as the holographic routing system described herein.

Bibliography

Solving sparse systems of linear equations on the Connection Machine (with C.E. Leiserson, J.P. Mesirov, L. Nekludova, S.M. Omohundro and W. Taylor). *Annual SIAM Conference,* A51, Boston, MA, July 1986.

Solving sparse systems of linear equations on the Massive Parallel Machine (with T. Opsahl). *First Symposium on Frontiers of Scientific Computing,* NASA, Goddard Space Flight Center, Greenbelt, MD, 2241-248, Sept. 1986.

Real-time compression of video on a grid-connected parallel computer (with J.A. Storer). *3rd International Conference on Supercomputing,* Boston, MA, May 1988.

Architecture and Operation of the BLITZEN Processing Element (with E.W. Davis). *3rd International Conference on Computing on Supercomputing,* Boston, MA, May 1988.

BLITZEN: a highly integrated, massively parallel machine (with D.W. Blevins, E.W. Davis and R.A. Hector). *2nd Symposium on Frontiers of Massively Parallel Computation,* Fairfax, VA, Oct. 1988. Also in Journal of Parallel and Distributed Computing, Feb. 1990.

4.2 Other Faculty Supported

4.3 Graduate Student Support

Akitoshi Yoshida	(to receive masters)
Sandeep Sen	(received PhD)
Steve Tate	(received PhD)

5 Travel:

On March 15, 1989, visit to Boulder, Colorado to view 1st demonstration of prototype holographic router being constructed in collaboration between Reif and Johnson at University of Colorado at Boulder. (This work began under AFSOR support, and was in 1989 augmented by a DARPA/ARO contract to Reif which has now expired.)

On Aug, 1989 visited Barakat at Harvard to work on paper on Optical Expanders to improve holographic router. Begin computer simulations of optical expander system.

On Sept, 1989 visit to Boulder, Colorado to discuss with Johnson construction of a larger scale holographic router at University of Colorado at Boulder.

On Sept, 1989 gave a talk optical computation and holographic routing at Univ Saarbucken, West Germany on optical routing system. Possible collaboration discussed.

On Feb, 1990 gave an invited talk on optical computation and holographic routing at the Pen State.

On Feb, 1990 gave an invited talk on optical computation and holographic routing to a large audience at the Parallel Computation Workshop at Courant Inst, NYU.

On April, 1990 gave an invited talk on optical computation and holographic routing at the University of North Carolina

On May, 1990 gave a invited talk on optical computation and parallel algorithms at the Parallel Computation Workshop (run by Vishkin at Univ Maryland) Workshop at Annapolis, Maryland.

On June, 1990 gave an invited talk on optical computation and holographic routing at Brandeis Univ, MA.

On July,1990 gave invited talks on optical computation and holographic routing in Greece (at Crete) and at various locations in Israel (at Technion, at the University of Tel Aviv, and the University of Jerusalem)

6 Conference Talks:

J.H. Reif and A. Tyagi, Efficient Algorithms for Optical Computing with the DFT Primitive in *The 10th Conference on Foundations of Software Technology and Theoretical Computer Science* , Lecture Notes in Computer Science, Springer-Verlag, Bangalor, India, December 1990A.

J.H. Reif and A. Tyagi, An Optical Delay Line Memory Model with Efficient Algorithms, *Advanced Research in VLSI Conference*, MIT Press,March 25-27, Santa Cruz, CA, 1991B.

J. Reif, A. Yoshida, and D. Tygar. The Computability and Complexity of Optical Beam Tracing, *31st IEEE Symposium on Foundations of Computer Science,* Saint Louis, Missouri, October, 1990.

A. Tyagi and J.H. Reif, Energy Complexity of Optical-Computations , appeared in *The 2nd IEEE Symposium on Parallel and Distributed Processing.* Dallis, TX, December 1990.

Appendix A
VLSIO Algorithms
A.1 The VLSIO MODEL

DFT-VLSIO and DFT-Circuit Models

VLSI Model:

It has been observed many times that the conventional electronic devices are inherently constrained by 2-dimensional limitations. Indeed, this was the original motivation for the VLSI model of Thompson [Thompson 80] which has been successfully applied to model such circuits. The widely accepted VLSI model allowed us both to compare the properties of algorithms such as area and time, and also to determine the ultimate limitations of such devices.

Let us first summarize the 2-D VLSI model, which is essentially the same as the one described by Thompson [Thompson 79]. A computation is abstracted as a communication graph. A communication graph is very much like a flow graph with the primitives being some basic operators that are realizable as electrical devices. Two communicating nodes are adjacent in this graph. A layout can be viewed as a convex embedding of the communication graph in a Cartesian grid. Each grid point can either have a processor or a wire passing through. A wire cannot go through a grid point with a processor unless it is a terminal of the processor at that grid point. The number of layers is limited to some constant γ. Thus both the fanin and fanout are bounded by 4 γ. Wires have unit width and bandwidth and processors have unit area. The initial data values are localized to some constant area, to preclude an encoding of the results. The input words are read at the designated nodes called input ports. The input and subsequent computation are synchronous and each input bit is available only once. The input and output conventions are where-determinate but need not be when-determinate.

VLSIO Model:

The recent development of high speed electro-optical computing devices allows us to overcome the 2-D limitations of traditional VLSI. In particular, the optical computing devices allow computation to be done in 3 dimensions, with full resolution in all the dimensions.

A rather different model for 3-D electro-optical computation is described in [Barakat, Reif, 87], which combines use of optics and electronics components in ways that models currently feasible devices. This model is known as the VLSIO model, with the O standing for optics. In this model, the fundamental building block is the optical box, consisting of a rectilinear parallelpiped whose surface consists of electronic devices modeled by the 2-D VLSI model and whose interior consists of optical devices. Communication from the surface is assumed to be done via electrical-optical transducers on the surface. Given specified inputs on the surface of the optical box, it is assumed that the output to the surface is produced in 1 time unit. Note that we do not rule out the possibility of two wide optical beams crossing, while still transmitting distinct information. However, there is an assumption (justified by a theorem of Gabor [Gabor, 61]) that a beam of cross section A can transmit at most $O(A)$ bits per unit time. This is the only assumption made about the power of the optical boxes.

For the purposes of upper bounds, we would have to be more specific about the computational power of optical boxes. The use of electro-optical devices will certainly allow us to overcome the 2-D limitations. The VLSIO potentially has more advantages over 2-D VLSI than just 3-dimensional interconnections of 3-D VLSI. In particular, it is well known that a 2 dimensional Fourier transform or its inverse can be computed by an optical device in unit time. In our discrete model, we assume that an optical box of size $n^{1/2} \times n^{1/2} \times n^{1/2}$ with an input image of size $n^{1/2} \times n^{1/2}$ can compute a 2-D Discrete Fourier Transform (DFT) in unit time. We call this the DFT-VLSIO model.

This is consistent with the capabilities of the electro-optical components constructed in practice. In this case, the VLSIO model is clearly more powerful than the 3-D VLSI model, *e.g.* since in that model we cannot do a DFT in constant time. A VLSIO device consists of a convex volume with a packing of optical boxes whose interiors do not intersect, but may be connected by wires between their surfaces. This allows for communication between two optical boxes. Note that the VLSIO model encompasses the 3-D VLSI model as a subcase: the particular subcase where each optical box is just a 2-D surface with no volume.

A VLSIO circuit is an embedding of a communication graph with the nodes corresponding to optical boxes in a three dimensional grid. The volume of a VLSIO circuit is the volume of the smallest convex box enclosing it. Due to Gabor's theorem [Gabor, 61] establishing a finite bound on the bandwidth of an optical beam, without any loss of generality, we assume that only binary values are used in transmitting information.

The DFT-Circuit Model:

Let R be an ordered ring. A circuit over R consists of an acyclic graph with a distinguished set of input nodes, and a labeling of all the non-input nodes with a ring operation. In the DFT circuit model, we allow:

1. scalar operations such as x, /, + and comparison with 2 inputs, and

2. DFT gates with n inputs and n outputs.

The *size* of the DFT circuit is the sum of the number of edges and the number of nodes. Recall from Parberry, Schnitger [Parberry, Schnitger, 88] that a *threshold* circuit is a Boolean circuit of unbounded fanin, where each gate computes the threshold operation. Threshold circuits are shown in Reif and Tate [Reif, Tate, 87] to compute a large number of algebraic problems such as polynomial division, triangular Toeplitz inverse, integer division, sin, cosine etc. in $n^{O(1)}$ size and simultaneous $O(1)$ depth.

Since the first output of a DFT gate is the sum of the inputs, and since comparison operations are allowed, a DFT circuit clearly has at least the power of a threshold circuit of the same size and depth. The question we address in this section is the power of the DFT operation above and beyond its power to compute threshold. Note that no non-trivial lower bounds on a threshold circuit computing a DFT are known. But, just by its definition, at least n threshold gates are required for a DFT computation.

A2 Efficient Optical Algorithms Using The DFT Primitive

A2.0

The optical computing technology offers new challenges to the algorithm designers since it can perform an n-point DFT computation in only unit time. Note that DFT is a non-trivial computation in the PRAM model. We develop two new models, DFT-VLSIO and DFT-Circuit, to capture this characteristic of optical computing. We also provide two paradigms for developing parallel algorithms in these models. Efficient parallel algorithms for many problems including polynomial and matrix computations, sorting and string matching are presented. The sorting and string matching algorithms are particularly noteworthy. Almost all of these algorithms are within a polylog factor of the optical computing (VLSIO) lower bounds derived in [Barakat, Reif 87] and [Tygar, Reif 89].

A2.1

Over the last 15 years, VLSI has moved from being a theoretical abstraction to being a practical reality. As VLSI design tools and VLSI fabrication facilities such as MOSIS became widely available, the algorithm design paradigms such as systolic algorithms, that were thought to be of theoretical interest only, have been used in high performance VLSI hardware. Along the same lines, the theoretical limitations of VLSI predicted by area-time tradeoff lower bounds have been found to be important limitations in practice. The field of electro-optical computing is at its infancy, comparable to the state of VLSI technology, say, 10 years ago. Fabrication facilities are not widely available—instead, the crucial electro-optical devices must be specially made in the laboratories. However, a number of prototype electro-optical computing systems—perhaps most notably at Bell Laboratories under Wong, as well as optical message routing devices at Boulder, Stanford and USC, have been built recently. The technology for electro-optical computing is likely to advance rapidly in the 90s, just as VLSI technology advanced in the late 70s and 80s. Therefore, following our past experience with VLSI, it seems likely that the theoretical underpinnings for optical computing technology—namely the discovery of efficient algorithms and of resource lower bounds, are crucial to guide its development.

What are the specific capabilities of optical computing that offer room for new paradigms in algorithm design? It is well known that optical devices exist that can compute a two-dimensional Fourier transform or its inverse in unit time, see Goodman [Goodman, 82]. This is a natural characteristic of light. This opens up exciting opportunities for the algorithm designers. In the widely accepted model of parallel computation—PRAM, not many interesting problems can be solved in $O(1)$ time. In particular, the best known parallel algorithm for Discrete Fourier Transform—FFT, takes time $O(\log n)$ for an n-point DFT. Given this powerful technology, the question we address is, "which problems can use the DFT computation primitive gainfully?" It is not immediately clear that given a problem, apparently disparate from DFT, such as sorting, how one reduces it to several instances of DFT to derive an efficient algorithm. We identify two general techniques that benefit a host of problems. First, we show a way to compute 1-dimensional n-point DFT efficiently using a series of 2-dimensional DFTs. Note that the optical devices compute a 2-dimensional DFT. However, the 1-dimensional DFT seems to be the one which is more naturally usable in most of the problems. Secondly, we demonstrate an efficient way to perform a parallel-prefix computation with DFT primitives. Equipped with these two techniques, we propose constant time solutions for a variety of problems including sorting, several matrix computations and string matching.

We consider discrete models for optical computing with a DFT primitive. In particular, an n-point DFT operation or its inverse can be computed in unit time using n processors. The development of a new model of computation is a task full of trade-offs. Only the essential characteristics of the underlying computing medium should be reflected in the model. Any unnecessary characteristics only serve to undermine the usefulness of such a model. PRAM (parallel random access machine) has provided a much needed model for the development of parallel algorithms for some time now. The algorithm designers do not have to worry about underlying networks and the details of timing inherent in the VLSI technology used to implement the processors. In a similar vein, our objective is to develop a model that captures the essence of optical computing medium with respect to algorithm design. We believe that the most important characteristic that distinguishes the optical technology from the VLSI technology is the ability to compute a powerful primitive, DFT, in unit time. Not surprisingly then, this is the focus of our models. Our new models are:

- [DFT-Circuit Model:] where we allow an n-point DFT primitive gate along with the usual scalar operations of bounded fanin.

- [DFT-VLSIO:] which extends the standard VLSI model to 3-dimensional optical computing devices that compute the 2-D DFT as a primitive operation. We refer to an electro-optical computation as VLSIO, where O stands for *optics*.

Note that although we did not mention a PRAM-DFT model where a set of n processors can perform a DFT in unit time; all the algorithms in DFT-Circuit model work for such a PRAM-DFT model.

A PRAM-DFT can simulate a DFT-Circuit of size $s(n)$ and time $t(n)$ with $s(n)$ processors in time $O(t(n))$. Hence, a PRAM-DFT model is an equally acceptable choice for the development of parallel algorithms in optical computing.

Our main results are efficient parallel algorithms for solving a number of fundamental problems in these models.

The problems solved include:

1. prefix sum
2. shifting
3. polynomial multiplication and division
4. matrix multiplication, inversion and transitive closure.
5. Toeplitz matrix multiplication, polynomial GCD, interpolation and inversion.
6. sorting
7. 1 and 2 dimensional string matching

Note: The sorting and string matching algorithms were not at all obvious. Although, we don't have any lower bounds in the DFT-circuit model, many of these parallel algorithms are optimal with respect to the VLSIO model. The known lower bound results in VLSIO are as follows. Barakat and Reif Barakat, Reif 87] showed a lower bound of $\Omega(I_f^{3/2})$ on $V\ T^{3/2}$ of a VLSIO computation for a function f with information complexity I_f. V denotes the volume of the VLSIO system computing f. We [Tyagi, Reif 89] proved a lower bound of $\Omega(I_f\, f(I_f^{1/2})\ )$ on the energy-time product for a VLSIO model with the energy function $f(x)$. We compare our results with the best-known PRAM algorithms for the corresponding problems. All the bounds are in Big-Oh notation (O).

Appendix B.
Lower Bounds for the energy consumption of Electro-Optical devices in the VLSIO model.

Over the last 15 years, VLSI has moved from being a theoretical abstraction to being a practical reality. As VLSI design tools and VLSI fabrication facilities such as MOSIS became widely available, the algorithm design paradigms such as systolic algorithms, that were thought to be of theoretical interest only, have been used in high performance VLSI hardware. Along the same lines, the theoretical limitations of VLSI predicted by area-time tradeoff lower bounds have been found to be important limitations in practice. The field of electro-optical computing is at its infancy, comparable to the state of VLSI technology say 10 years ago. Fabrication facilities are not widely available—instead, the crucial electro-optical devices must be specially made in the laboratories. However, a number of prototype electro-optical computing systems—perhaps most notably at Bell Laboratories under Wong, as well as optical message routing devices at Boulder, Stanford and USC, have been built recently. The technology for electro-optical computing is likely to advance rapidly in the 90s, just as VLSI technology advanced in the late 70s and 80s. Therefore, following our past experience with VLSI, it seems likely that the theoretical underpinnings for optical technology—namely the discovery of efficient algorithms and of resource lower bounds, are crucial to guide its development.

Barakat and Reif [Barakat, Reif 87] developed a model for electro-optical computing systems. They refer to an electro-optical computation as VLSIO, where *O* stands for *optics*. Since we anticipate the number of VLSI components in optical computers to be large, the VLSI prefix in VLSIO can be reasonably used. The following two significant aspects distinguish VLSI from VLSIO. VLSIO has a 3 dimensional character. Secondly, the information in VLSIO is carried by optical beams rather than electrical currents.

Just as *area*, *energy* and *time* are three fundamental resources in a VLSI computation, *volume*, *energy* and *time* are the resources of interest in a 3-D VLSI circuit or an optical computing system. The *volume*, *time* lower bounds for optical computations have been established by Barakat and Reif [Barakat, Reif 87] along the lines of AT^2 VLSI bounds. But, a similar asymptotic analysis of energy bounds in VLSIO computations is missing. A study of energy requirements in 3-D VLSI has also not been undertaken. Energy has received increased attention recently because the power consumption largely determines the total cost of a high performance computer due to heat dissipation. The theoretical physicists have also considered the viability of characterizing the computational costs entirely in terms of energy. All of the recent research activity in energy complexity has been directed at the study of the energy requirements in 2-D VLSI computations. More specifically, the first formal result in switching energy was due to Lengauer, Mehlhorn [Lengauer, Melhorn 81], which shows that the switching energy of transitive functions, E, is $\Omega(n^2/P \log(AP^2/n^2))$, which is $\Omega(n^2)$ for $AP^2 = O(n^2)$. P is the period of a pipelined computation. Kissin [Kissin 82, 85] proposed a formal model for switching energy distinguishing between uniswitch and multiswitch models. When a wire is assumed to switch at most once during the course of computation, it is a *uniswitch* circuit. Most of the pipelined computations fall in this class. The more general model, that allows each wire to switch any number of times, is called the *multiswitch* model. Snyder, Tyagi [Snyder, Tyagi 86] and Leo [Leo 84] considered variations on Lengauer, Mehlhorn result. The first tight bound on uniswitch and multiswitch energy-period product [$\Omega(n^2)$] for shifting was obtained by Aggarwal et. al. [Aggarwal et. at, 88]. Tyagi [Tyagi 89] derived a tight bound on multiswitch energy, $\Omega(n^{1.5})$, and average case uniswitch and multiswitch energy. The 3-D VLSI model has been studied by Rosenberg [Rosenberg 81], Preparata [Preparata 83], and Leighton, Rosenberg [Leighton,Rosenberg 86] with respect to volume-time trade-offs. *We analyze the energy requirements in 3-D VLSI and VLSIO systems.*

The energy consumption model developed in Kissin [Kissing 82] applies to the 3-dimensional VLSI as well. But, as a first step, a consistent model of energy consumption in optical computing is needed. In this section, we propose two models for the energy consumption in an optical computer which are consistent with the VLSIO model described in [Barakat, Reif 87]. Within these models, we demonstrate tight bounds on both energy and energy-time product for the optical computation of several functions.

A key property which we have considered in this work is the energy consumed by an electro-optical device. This is determined by summing the energy consumed by each wire and by each optical beam. This energy consumption is assumed to be due to switching. In all the energy models considered to date—a wire of length d consumes switching energy $\Theta(d)$, which is consistent with the currently used CMOS technology. However, in an optical computation, an energy cost non-linear (even exponential) in the length of the switching wire is justifiable for some frequency range. This leads to a generalization of the energy model. In particular, we assume an energy function, $f(d)$, such that $f(d)$ energy is consumed by a wire/beam of length d switching between 0 and 1. Here $f(d)$ is a function that may or may not be nonlinear, but f and its first derivative must be continuous functions. We argue that $f(d)$ can, in theory, be an exponential function in d for optical beams. We also show why, in practice, $f(d)$ may be a polynomial or even a linear function. Our energy lower bounds encompass any such energy function $f(d)$. Note that the case of a nonlinear energy function has not been considered previously even for 2-D VLSI. The local cutting techniques used for the linear energy model consider the energy consumption of the unit-length wire segments incident on the cut. However, in such a local context, any non-linear energy function, at best, measures the same energy consumption at the cut as does the linear energy function. The unit length segments consume the same order of energy for all the energy functions. Hence a somewhat more global lower bound approach is needed in the generalized energy model.

Results: We derive the lower bounds, shown in the table below, on uniswitch and multiswitch energy E and energy-time product ET of a transitive function. The matching upper bounds are established for a transitive function: *shifting*.

Note that the objective of multiswitch circuits is to find a tight embedding for the devices under the premise that it leads to shorter links. The overall energy saving is derived from the observation that the repeated use of short links leads to a smaller ET product. On the other hand, a uniswitch circuit will have to make links long in order to propagate information far enough. But it will use every link only once. Hence, as shown in [Tyagi 89], in 2-D VLSI a multiswitch circuit always has a lower energy consumption than a uniswitch circuit. Interestingly, as we show, the only 3-D VLSI examples satisfying the multiswitch lower bound for $f(x) < x^{4/3}$ are uniswitch circuits. We believe that no 3-D circuits exist satisfying the lower bound in this energy function range. This says that for the 3-D case, there is a zone : $x < f(x) < x^{4/3}$, where long links leading to higher volume perform better than a circuit with short links, defying the conventional wisdom.

Appendix C
Complexity of Optical Ray Tracing

We examine ray tracing problems in [Reif, Akitoshi, and Tygar, 90]. The history of ray tracing goes back at least to Archimedes, who examined images formed by a mirror to understand the law of reflections. In the 15th to 18th centuries, many scientists and astronomers in Europe worked on geometrical optics and invented optical instruments such as telescopes. In 1730, Newton published his book "Opticks" in which he formally defined the reflective and refractive laws of optics, and first defined and investigated some ray tracing problems. These classical ray tracing problems are very important to the design of most optical systems which consists of a set of refractive or reflective surfaces, and involve tracing the path of rays to investigate the performance of the systems. Ray tracing also has important application in computer graphics, where ray tracing is used to render pictures which consist of objects with surfaces that reflect or refract light rays.

The *ray tracing problem* is a decision problem: given an optical system (namely, a finite set of reflective or refractive surfaces) and an initial position and direction of a light ray and some fixed point p, does the light ray eventually reach the point p.

Our optical systems consist of a finite set of optical objects that may be totally reflective (we call these *mirrors*), partially reflective (we call these *half-silvered mirrors*), or totally absorbent (we call these *lenses*). We restrict ourselves to optical systems constructed out of flat (e.g., line segments) mirrors and half-silvered mirrors; and out of lenses whose boundaries are quadratic curves. (We call these lenses *quadratic lenses*.) Do mirrors reflect if a light-beam is directed exactly at an endpoint? It will turn out that this matters for the case when we form a corner out of two mirrors. What should happen when the light beam is directed exactly at the corner? We shall allow mirrors (and half-silvered mirrors) to reflect entirely along the surface of either a closed, half-closed, or open line segment.

The positions of our mirrors, half-silvered mirrors, and lenses can be either *rational* or *irrational*. If the optical system consists only of mirrors or half-silvered mirrors with endpoints with rational coordinates, we say that the optical system is *rational*. If the optical system contains mirror or half-silvered mirrors with endpoints that have irrational coordinates then we say the optical system is *irrational.*

We are interested in if the light will reach a final certain position, and not in the intensity of the light at that position. Throughout this section, we assume that the path taken by light rays are determined by the classical laws of optics: *the law of reflection* and *the law of refraction.*

(The law of reflection states that the incident angle and the reflected angle are equal, and the law of refraction states that the angle of refraction depends on the incident angle and the index of refraction of the materials.) We always assume that the initial position of the light ray has rational coordinates and the tangent of the initial incident angle is rational, and the test point p has rational coordinates. (In general, in our lower bound proofs, it suffices to let the light rays initially enter perpendicular to a window of the optical systems.) Our surprising discovery is that if the optical system is rational it may have high complexity, or even be undecidable. We generally denote n to be the number of bits in binary encoding of the optical system.

Our results of the computational complexity for ray tracing in various optical systems may be summarized as follows:

1. Ray tracing in three dimensional optical systems which consist of a finite set of mirrors, half-silvered mirrors, and quadratic lenses is undecidable, even if the endpoints of the objects in the optical system all have rational coordinates. However, the problem is recursively enumerable.

2. Ray tracing in three dimensional optical systems which consist of a finite set of mirrors is undecidable, if the mirrors' endpoints are allowed to have irrational coordinates. However, the ray tracing problem is PSPACE-hard, if we restrict ourselves to mirrors with endpoints that are rational coordinates.

2. For any $d \geq 2$, ray tracing of d dimensional optical systems which consist of a finite set of mirrors surfaces lies in PSPACE, if the positions of all the surfaces are rational, and they lie perpendicular to each other. For $d \geq 3$, the problem is PSPACE-complete.

We consider three optical models in this section:

In optical model (1), each optical system consists of a finite set of quadratic lenses, mirrors, and half-silvered mirrors. A light ray travels through the system with reflections or refractions. We show that the problem of deciding if the light ray will reach a given final position in this system is undecidable. In order to show this, we simulate a universal Turing machine with this optical model. What is perhaps surprising, is that our optical system has a fixed number of optical lenses and mirrors, and yet the ray tracing problem for it simulates any recursive enumerable computation, where the input is given by the initial position of the light ray.

In optical model (2), each optical system consists of a finite set of mirrors and half-silvered mirrors in three dimensional space. We again show that the problem of deciding is undecidable. To show this, we simulate a 2-counter machine with this optical model. Next, we consider the computational complexity when we restrict ourselves to rational optical systems. In this case, we show that the problem is PSPACE-hard. To show this, we first define a certain augmented bounded 2-counter machine. Then, we simulate this augmented bounded 2-counter machine with this optical system. By showing the augmented bounded 2-counter machine can compute an arbitrary polynomial space problems, we conclude that the problem of deciding if the light ray reach a given final position in this system is in PSPACE-hard. (Although we show that the problem is PSPACE-hard, we do not even know if this restricted problem is decidable.)

Optical model (3) is a generalization of optical model (2). In optical model (3), each optical system occurs in a unit-sized d dimensional hypercube. The hypercube contains a rational optical system of mirrors. Each of the mirrors lies perpendicular to every other mirror. We show that the problem of deciding if the light ray will reach a given final position has a non-deterministic polynomial space algorithm, thus showing the problem is in PSPACE.

Theoretically, these optical systems can be viewed as general optical computing machines, if our constructions can be carried out with infinite precision, or perfect accuracy. However, these systems may not be practical, since the above assumption may not hold in physical world. The motivation for this work comes from an interest in investigating the problem complexities in ray tracing problems.

Apppendix D
Optical Memory Storage and Computation Using Fiber Optic Delay Loops
D.1 Data Storage: A Key Problem in Optical Computing

Optical computing technology can obtain extremely high data rates beyond which can be obtained by current semiconductor technology. Therefore, in order to sustain these extremely high data rates, the dynamic storage must be based on new technologies which will likely be wholly or partly optical. Jordan at the Colorado Optoelectronic Computing Systems Center and some other groups have proposed and used optical delay loops for dynamic storage. In these data storage systems, an optical fiber, whose characteristics match the operating wavelength, is used to form a delay line loop. In particular, the system sends a sequence of optically encoded bits down one end of the loop and after a certain delay (which depends on the length and optical characteristics of the loop), the optically encoded bits appear at the end of the loop, to be either utilized at that time and/or once again sent down the entrance of the loop.

This idea of using propagation delay for data storage dates back to the use of mercury delay loops in early electronic computing systems before the advent of large primary or secondary memory storage. Jordan at Boulder has achieved over 10^4 bits per fiber loop of approximately one kilometer. This was achieved in a small, low cost prototype system with a synchronous loop without very precise temperature control. Nevertheless, Jordan used such a delay loop system to build the second (after Wong's) known purely optical computer (which can simulate a counter). This does not represent the ultimate limitations of optical delay loops, which could in principal provide very large storage using higher performance electro-optical transducers and the use of multiple loops. Actually, the key problem with such a dynamic storage is that it is not a random-access memory. A delay line loop cannot be tapped at many points since a larger number of taps leads to excessive signal degradation. This implies that if an algorithm is not designed around this shortcoming of the dynamic storage, it might have to wait for the whole length of the loop for each data access. Systolic algorithms also exhibit such a tight inter-dependence between the dynamic storage and the data access pattern.

D.2. Our New Delay Loop Memory Model and Our Results

We have studied the repercussions of the use of memory loops on algorithm design. The use of delay loops as memories is necessitated by the required extremely high data rates .

In [Reif and Tyagi,90], we proposed the delay loop memory(DLM) model as a theoretical model of sequential electro-optical computing with dynamic storage using a fixed number of delay loops.

Our theoretical model contains the basic features that current delay loop systems use, as well as systems in the future are likely to use. It would seem that the restrictive discipline imposed on the data access patterns by a loop memory would degrade the performance of most algorithms, because the processor might have to idle waiting for data. We demonstrate that an important class of algorithms, ascend/descend algorithms, can be realized in the loop memory model without any loss of efficiency. In fact, the sequential realizations span a broad range for the number of loops required. A parallel implementation performing the optimal amount of work is also shown. Some matching lower bounds are illustrated, as well. of optical delay systems that exists and may be built in the feature.

We developed an optimal implementation of the ascend-descend class of algorithms on DLM model. Note that many problems including merging, sorting, FFT, matrix transposition and multiplication and data permutation are solvable with an ascend/descend algorithm which is a very general class of parallel algorithms described by [Ullman,87] text book on Computational Aspects of VLSI.

An ascend or descend phase takes time $O(n \log n)$ in DLM model using $\log n$ loops of geometrically increasing sizes.1, 2,4,...n. Note that a straight-forward emulation of a butterfly network with $O(n \log n)$ time performance requires $O(n)$ loops: n loops of size 1, $n/2$ loops of size 2, $n/4$ of size 4, .., 1 of size n. It can be implemented in time $n^{1.5}$ just with two loops of sizes $\sqrt{n}$ and n each. This can be generalized into an ascend-descend scheme with time $n\,k + n^{1.5}2^{-k/2}$ with $1 < k < 1+\log n$ loops. At this point in time, a loop is a precious resource in optical technology, and hence tailoring an algorithm around the number of available loops is an important capability. The k-loop adaptation of the ascend/descend algorithm provides just this capability.

A single loop processor takes n^2 time. A matching lower bound also exists for this case, which is derived from one tape Turing machine crossing sequence arguments. Matrix multiplication and matrix transposition can also be performed in DLM without any loss of time.

We also consider a butterfly network with $p \log p$ DLM processors, where $1 < p < 1+n$. The work (# of processors, time product) of this network for ascend-descend algorithms is shown to be $O(n \log n)$. Note that a butterfly network performs $n \log n$ work. This shows that the ascend-descend algorithms can be redesigned in such a way as not to incur any work loss due to the restrictive nature of the loop memories.

Appendix E
Holographic Based Computing
E.1 Holographic Message Routing

We describe an electro-optical message routing system for sending N messages between N processors in constant time using $2N \log N$ switches. A spatial light modulator (SLM) is used to holographically steer messages directly to their destination processor. The system is unique in that it uses fixed holograms to achieve free space dynamic routing. A small prototype implementation has been already constructed [Maniloff, Johnson and Reif,89]. (An appendix describes practical issues.)

We introduce a new optical technique which we call the optical expander. We discuss how an optical expander can be used to solve a key problem, namely the orthogonality of message patterns. In particular, the optical expander system is used to decrease the number of address bits used by the router and to improve separation of distinct address patterns matched by the holograms. We discuss the theory of the optical expander system and give for the first time a rigorous proof of its correctness and performance.

E.1.1 The Potential of Optical-Electronic Systems

The inherent high parallelism and connectivity of optical signal processing lends itself directly to such applications as optical interconnection. (See the recent text of [Feitelson,88]). The recent development of moderately high speed, high dynamic range spatial light modulators has lead to the prototype development of variety of optically based signal processing systems.

E.1.2 Our Holographic Routing System

Dynamic message switching is the problem of sending N messages between N processors, where the destination permutation is given dynamically. In this section we describe a novel holographic message routing system for dynamic message switching. We use a spatial light modulator (SLM) to holographically steer messages directly in free space to their destination processor. An important innovation of our holographic routing system is the use of fixed holographs to do the dynamic message switching. It uses $2N \log N$ boolean switches, which is optimal within a factor of 2.

In brief, our holographic message routing system is a unique architecture which uses N multiple-exposure holograms, each containing N images to connect N processors to N processors, via free space routing. The system uses N spatial light modulators (SLMs), each with $2\log N$ pixels. A column of light illuminates each processor's SLM which is programmed with an encoded address for a destination processor. This optically encoded address is routed directly to the correct processor by a hologram containing N images, each correlated with a particular destination processor. This optical interconnection network is a direct message router taking constant time as compared to conventional fixed interconnection networks which require time delay at least $\log N$. Our holographic message system can be applied to do very high speed message routing for massively parallel machines such as the CONNECTION machine.

E.1.3 An Implementation of the Holographic Routing System

There was a collaborative Optical Routing Project between theoretical computer scientist, John Reif, at the Computer Science Department, Duke University and optical engineers Kristina Johnson and Eric Maniloff at the Center for Optoelectronic Computing Systems at University of Colorado, Boulder. While Reif initially conceived of the theory of the system, the practical implementation was due to Johnson and Maniloff, who built a 4 by 4 prototype holographic routing system (for implementation details see [Maniloff, Johnson and Reif,89]) at the Center for Optoelectronic Computing Systems at University of Colorado, Boulder. This running prototype implementation was completed in April, 1989. Because of the small size of this prototype system, an optical expander system was not required. They have also developed in [Strasser, Maniloff, Johnson, Goggin,89] a, procedure for recording multiple-exposure holograms with equal diffraction efficiency in photorefractive media. Reif has also directed computer simulations of the message routing applications.(the availability of a device which can control light with a high spatial resolution and with a short cycle time is critical to the successful realization of a second generation our system; for this we acknowledge the technical assistance from Derek Lile, Colorado State University, on the development of III-V MQW/CCD SLMs.)

E.1.4 Comparison with other Routing Systems

Interconnection networks in parallel processing computers are very important subjects. There are many interconnection networks for different applications, since different algorithm requires different degree of globality of the interconnects. Because of the availability of non-linear devices as gates which is extensively used in the interconnection network, electrically implemented interconnections are widely seen among many computer organizations. However, the future of electric interconnections is not necessarily bright. The problem comes from its restricted dimension—the wiring is confined on a two dimensional plane—and from RC delay on interconnections.

These drawbacks which are found in electrical interconnections do not exist in optical interconnections. Light beams need not be confined in a wave guide such as an optical fiber, but can travel freely through space. In addition, light beams can have a great bandwidth, and the propagation of light traveling through space or in a fiber is not affected by resistance, capacitance, or inductance. Thus, optical interconnections offer a high data transfer rate in a simple architecture by a set of light beams freely traveling through space. The various papers discuss the potential of optical interconnections.

Among various message routing networks the highest level of interconnection is a crossbar network which uses N^2 interconnects to connect N source units and N destination units. The number of electrical interconnection wires required by each processing unit to communicate with the other processing unit on- and off-board will limit the feasible size of the network. The property of light beams which we briefly mentioned above may give a great potential for an alternative high-speed optical crossbar type of networks.

The property of light beams which we briefly mentioned above may give great potential for an inexpensive and high-speed optical crossbar network.

There are several optical interconnection networks which have already been proposed. One is *optical crossbar network*. The optical crossbar network typically uses an $N \times N$ spatial light modulator (SLM) to connect N source processors to N destination processors. Each source processor uses a column of the $N \times N$ SLM to address one of N distinct destination processors. The advantage of this optical crossbar is that once all the entries of the $N \times N$ SLM are set, the message can be transmitted at very high data rates, namely at optical pulse modulation rate. This matrix-vector multiplier based crossbar network has two drawbacks. One is that at most $1/N$ of the power incident on the SLM will reach the detector. The other is that it takes a long time to electrically set an $N \times N$ SLM.

E.2 Holographic Memory Storage

Holographic Matching

In this section, we describe the general idea of holograms and that of holographic associative matching.

Principle of Holograms

A photograph records the intensity distribution of the light wave scattered by an object. A hologram, however, records the intensity and phase distribution of the light scattered by an object. Since a hologram has the information about the intensity and the phase of the scattered light wave, we can reconstruct the image of the object from the hologram.

In order to record the phase information of the scattered light, we superimpose a reference wave to the light wave scattered by an object. Then, the resulting interference pattern can be recorded on a photographic plate.

Wave Front Recording and Associative Matching

For wave front recording and holographic associative matching, two coherent beams are used in the recording. Both the object beam, which we wish to record, and a reference beam illuminate the photographic medium. The photographic medium records the interference fringes which are produced as the interaction between the object beam and the reference beam. After recording, when the recorded fringes are illuminated by a reconstruction beam—typically a reproduction of the reference beam, the fringes diffract the reconstruction beam into three main beams; the zero order term which corresponds to the reconstruction beam, a first order diverging virtual image which corresponds to the reconstructed object beam, and the other first order converging real image which corresponds to the conjugate of the object beam. The arrangement of the recording must be carefully done so that these beams do not overlap each other. When the wave length or the position of a reconstruction beam differs from those of the reference beam, the reconstructed images are altered.

The geometry of hologram formation affects the diffraction properties of the hologram. The thickness of plane holograms is small compared to the spacing of the interference fringes recorded on the media. This type of the holograms can be considered as a plane diffraction grating. On the other hand, volume holograms are thick, and the interference fringes are recorded in three dimensions. Thus, the volume holograms can be considered as volume diffraction gratings where the diffracted beams obey Bragg's law. The reconstruction of the volume hologram is very sensitive to the direction of the reconstruction beam. If this direction is not identical to the direction obtained from Bragg's law, there will be no images reconstructed. This property offers a possibility in making multiple-exposure distinct holograms in a single piece of volume photographic medium. The distinct holograms may be recorded by using distinct reference beams. Later, each hologram can be reconstructed by using the corresponding reference beam as a reconstruction beam. Thus, illuminating a multiple-exposure volume hologram by a reconstruction beam can be viewed as addressing a stored image associated with the reconstruction beam.

Media for Volume Holograms

As a media for volume holograms, thick photographic emulsion has been used for many years. However, other mediums such as various types of photorefractive nonlinear optical crystals have received much attention for their flexibility in dynamic recording. The most widely used such media is Fe-doped lithium niobate ($LiNbO_3$). When this type of crystals is illuminated, the concentration of photocarriers in the crystal will be changed. These photocarriers will be trapped, and will produce the change in the refractive index of the crystal.

Many researchers have investigated multiple-exposure holograms on volume media. They showed hundreds of distinct holograms may be recorded, if the medium is thick enough, and the different reference beams has an angular displacement of a few minutes. Staebler et al. showed that as long as the distinct reference beams enter at angular displacements of at least $\pi/1000$, it is possible to record at least 512 multiple holographic exposures in a volume medium. Therefore, we can use a single volume hologram to store $N = 512$ images as long as we use N mutually orthogonal addressing beams. These N beams can be constructed by use of our optical expander.

Holographic Memory Storage

Holograms can be used to implement random access memory storage systems. The basic idea of holographic memory storage is that the data are arranged in blocks which are stored in holograms. A block of memory can be retrieved at time by using its corresponding reconstruction beam. This type of memory is particularly suited for read-only applications, since the holograms can be fixed. However, dynamically modifiable holograms such as photorefractive materials may give potential for active holographic memory storage systems. The work in the 70s promised the advantage of holographic memory over the other types of memory in terms of bit/volume ratio, size, and throughput. However, the lack of appropriate recording materials and fast addressing methods kept holographic memory behind the progress of MOS VLSI based memory. Recently, the advance in recording materials such as various photocrystals and, the success in fabricating an array of large number of micro lasers have provided a chance for holographic memory to be efficiently implemented. Several prototypes of such a memory storage system have been developed at Microelectronics and Computer Technology and Bellcore.

In a typical holographic memory storage system, the data are organized in blocks. Our proposed holographic memory storage system uses d light beams to retrieve N blocks of data, where $d \leq 2 \log N$. Without our optical expander, such systems require either a beam deflector to deflect a laser beam into one of N unique directions, or an electrically implemented line decoder which accepts log N bits of binary information and creates one of N unique laser beams. Both approaches have several disadvantages. We mention these disadvantages in E.3. Optical Expander.

Our optical expander will provide an alternative approach by utilizing its three dimensionality with flexibility and accuracy provided by digital operations.

E.3 Optical Expanders

An optical expander takes as an input a boolean pattern of size $d = c \log N$ bits, and expands it to a boolean pattern of size N bits, where c is a constant satisfying $1 \leq c \leq 2$. Each expanded boolean pattern is required to be mutually orthogonal to the others. Thus, an optical expander can be viewed either as an electrooptical line decoder which converts d bits of optically encoded binary information to up to N unique optical outputs, or as a digital beam deflector which uses a control signal encoded in d bits to deflect an input laser beam into one of N directions.

More precisely, an optical expander takes as input one of N distinct boolean vectors $p_1, p_2, \ldots p_N$ of length d. We call these vectors the *input patterns*. Each input pattern is optically encoded by using d pixels, each pixel being either ON (denoted by 1) or OFF (denoted by 0). We will require that each input pattern has exactly $d/2$ pixels ON. The optical expander produces a spatial output pattern r_i from given input pattern p_i. Each output pattern r_i is one of N distinct orthogonal boolean vectors of length N.

In addition to our standard optical expander, we define a generalized optical expander. A generalized optical expander is also an electrooptical system which takes as an input a boolean pattern of size d bits and expands it to a boolean pattern of size N bits. Here, unlike the standard optical expander, each expanded boolean pattern may have more than one ON (denoted by 1) in its elements. In other words, a generalized optical expander creates a boolean pattern of size N which is a bit wise OR product of some subset of the N mutually orthogonal boolean patterns. The advantage of this generalized optical expander becomes clear in certain applications. It can be used in broadcasting messages in a message switching network. It can also be applied to a holographic memory system with a multiple readout capability, where bit wise OR, AND, or XOR products of several images (data) can be directly obtained as a superimposed output on the detector array.

Our optical expander accepts an input pattern encoded in d bits, and expands it into a pattern encoded in N bits. We wish to have an exponential expansion, so d has to be represented by $d = c \log N$ for some constant c. First, we describe an optical expander with the constant $c = 2$. Later, we will look at an encoding scheme with the constant $c \approx 1$ for a large d. This allows us to produce a greater number of orthogonal patterns with the same number of input bits. However, setting $c = 2$ offers several advantages. First of all, it makes the coding scheme simple, since $d = 2 \log N$ offers a coding scheme where each p_i can be a concatenation of two binary strings: one representing i in binary format, and the other representing i in one's complement binary format. Thus, p_i can be easily produced from the binary-coded output from the electrical interface without any additional electrical mapping interfaces. Secondly, it also makes optical interconnection patterns from d optical inputs to the threshold array regular, thus resulting in a simple implementation. Finally, it can provide an addressing scheme for a generalized optical expander.

Optical Expanders require Non-linear optical systems

A linear optical system can not be used as an optical expander, since any linear mapping from an input of size d creates no more than d linear independent output patterns. Thus, it is impossible to create a set of $N > d$ mutually orthogonal patterns by any linear optical system on d linear independent patterns.

Non Linear Optical Filters

Non-linearity can be introduced to an optical system by two methods. One method is to use a non-linear device. Thresholding input intensity at a certain level to produce output is a non-linear operation. It can be implemented by optical non-linear devices such as optical logic etalon (OLE), or by electrooptical non-linear devices such as self electrooptic effect device (SEED). The other method is to translate input into spatial patterns, and then to use a linear filter on the fourier image plane. An example is Theta modulation, where data are encoded as a grating of different orientations. In our optical expanders, we use non-linear devices.

Disadvantages of Other Approaches

We review the disadvantages of previous approaches such as a beam deflector based on an acoustooptic effect or on Kerr cell, and a VLSI implementation.

Systems which require N distinct entry beams either to an N-superimposed hologram or to an array of N devices may use an optical expander to generate N beams from optical input of $c \log N$ bits. Without our optical expander, such systems require either a beam deflector to deflect a laser beam into one of N unique directions, or an electrically implemented line decoder which accepts $\log N$ bits of binary information and creates one of N unique laser beams.

Analog beam deflectors based on acoustooptic effect have several drawbacks. First of all, they are bulky and acoustooptic modulators require high drive power. Secondly, they are limited by capacity-speed product. A frequency band width Δf as high as 300MHz can be obtained by the acoustooptic material such as alpha-iodic acid. If we want to switch the deflector every 1μsec, then with a safety factor of 2, the number of resolvable points will be at most 150. In order to overcome the disadvantage of the acoustooptic beam deflector, a multistage digital beam deflector has been designed. They demonstrated a 20-stage deflector consisting of a series of nitrobenzene Kerr cells and birefringent calcite prisms. The laser beam was deflected into a two dimensional 1024 x 1024 plane in every 2μsec. This approach provided a great flexibility and accuracy in controlling the deflection angle. However, it required very high bias voltage and switching voltage of several kilovolts, and the power consumption was 400W.

Electrically implemented large line decoders are not practical in terms of speed and wiring areas for a large N. As we mentioned earlier, the I/O constraints limit the size of system which can be practically implemented. For a large N, the output may have to be serially transmitted from the chip.

Our optical expander will provide an alternative approach to these devices by utilizing its three dimensionality with flexibility and accuracy provided by digital operations.

Our Results

We designed two optical expanders, and investigated each model in terms of size, power requirement, and speed.

One approach was based on an idea of implementing a large line decoder by using optical interconnections. This was done by using optical matrix-vector multiplication followed by a thresholding operation. In this model, the optical signal emitted from a signle laser diode (LD) source is distributed to N threshold devices. Therefore, the maximum switching cycle B (cycle/sec) is proportional to the radiation power from a single LD source P_{ID} and inversely proportional to the output size N. The physical size is determined by the integration density of a $\sqrt{N} \times \sqrt{N}$ threshold device array.

The other approach used a set of small identical switching cells to implement a novel digital beam deflector. In this model, all the switches have a fan-out of 1, and are connected in series. Therefore, given the radiation power of an input laser and the required optical output power of the optical expander, the maximum output size N is determined by the loss at the switching cells.

Applications of optical expanders have been also discussed to motivate the design and construction of our optical expanders.

See [Reif and Yoshida, 90] for details.