

RL-TR-97-229, Volume I (of two)
Final Technical Report
February 1998



ELECTRO-OPTIC COMPUTING ARCHITECTURES

Hughes Research Laboratories

Sponsored by
Advanced Research Projects Agency
ARPA Order No. A275

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

19980415 092

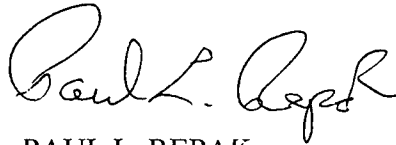
DTIC QUALITY INSPECTED 3

AIR FORCE RESEARCH LABORATORY
ROME RESEARCH SITE
ROME, NEW YORK

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

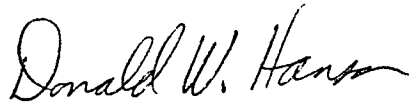
RL-TR-97-229, Volume I has been reviewed and is approved for publication.

APPROVED:



PAUL L. REPAK
Project Engineer

FOR THE DIRECTOR:



DONALD W. HANSON, Director
Surveillance & Photonics

If your address has changed or if you wish to be removed from the Air Force Research Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/SNDP, 25 Electronic Pky, Rome, NY 13441-4515. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

ALTHOUGH THIS REPORT IS BEING PUBLISHED BY AFRL, THE RESEARCH WAS ACCOMPLISHED BY THE FORMER ROME LABORATORY AND, AS SUCH, APPROVAL SIGNATURES/TITLES REFLECT APPROPRIATE AUTHORITY FOR PUBLICATION AT THAT TIME.

ELECTRO-OPTIC COMPUTING ARCHITECTURES

U. Efron, S. Esener, C. S. Wu,
P. J. Marchand, K. Sayyah, M. Yung,
M. J. Little, R. A. Forber, J. A. Neff, C. Stirk

Contractor: Hughes Research Laboratories
Contract Number: F30602-93-C-0173
Effective Date of Contract: 12 July 1993
Contract Expiration Date: 11 July 1997
Program Code Number: 3D10
Short Title of Work: Electro-Optic Computing Architectures
Period of Work Covered: Jul 93 - Jul 97

Principal Investigator: Uzi Efron
Phone: (310) 317-5806
RL Project Engineer: Paul L. Repak
Phone: (315) 330-3146

Approved for public release; distribution unlimited.

This research was supported by the Advanced Research Projects
Agency of the Department of Defense and was monitored by
Paul L. Repak, AFRL/SNDP, 25 Electronic Pky, Rome, NY.

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 instructions, searching 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 February 1998	3. REPORT TYPE AND DATES COVERED Final Jul 93 - Jul 97	
4. TITLE AND SUBTITLE ELECTRO-OPTIC COMPUTING ARCHITECTURES			5. FUNDING NUMBERS C - F30602-93-C-0173 PE - 61101E/62712E PR - A275 TA - 00 WU - 01	
6. AUTHOR(S) U. Efron, S. Esener, C. S. Wu, P. J. Marchand, K. Sayyah, M. Yung, M. J. Little, R. A. Forber, J. A. Neff, C. Stirk				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Hughes Research Laboratories 3011 Malibu Canyon Road Malibu CA 90265			8. PERFORMING ORGANIZATION REPORT NUMBER N/A	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Advanced Research Projects Agency Air Force Research Laboratory/SNDP 3701 North Fairfax Drive 25 Electronic Pky Arlington VA 22203-1714 Rome NY 13441-4515			10. SPONSORING/MONITORING AGENCY REPORT NUMBER RL-TR-97-229, Volume I (of two)	
11. SUPPLEMENTARY NOTES Air Force Research Laboratory Project Engineer: Paul L. Repack/SNDP/(315) 330-3146				
12a. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; distribution unlimited.			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The objective of the Electro--Optic Computing Architecture (EOCA) program was to develop multi-function electro-optic interfaces and optical interconnect units to enhance the performance of parallel processor systems and form the building blocks for future electro-optic computing architectures. Specifically, three multi-function interface modules were targeted for development - an Electro-Optic Interface (EOI), an Optical Interconnection Unit (OIU), and a Space-Time Compander (STC). Electro-optic 3-D interconnection module stacks were assembled, allowing for system global communication and fast efficient data routing and sorting. The achieved goal of the system study, to identify and analyze all the architectural implications due to the addition of optical based free-space interconnects in locally connected processor arrays, is leading to new highly optimized 3-D electro-optic computer processing networks.				
14. SUBJECT TERMS Optical Processing, Optical Computer Architecture, Optical Switching, Micro Optical Circuits			15. NUMBER OF PAGES 170	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT UNCLASSIFIED	18. SECURITY CLASSIFICATION OF THIS PAGE UNCLASSIFIED	19. SECURITY CLASSIFICATION OF ABSTRACT UNCLASSIFIED	20. LIMITATION OF ABSTRACT UL	

Standard Form 298 (Rev. 2-89) (EG)
Prescribed by ANSI Std. Z39.18
Designed using Perform Pro, WHS/DIOR, Oct 94

DTIC QUALITY INSPECTED 3

CONTENTS

	Page
1 INTRODUCTION	1
2 ELECTRO-OPTICAL INTERFACE	4
2.1 Optical Transpose Interconnection Network	4
2.2 Flip-chip Bonded Si/PLZT Smart Pixel.....	6
2.3 Electronic Switches.....	7
2.3.1 Overview.....	7
2.3.2 Designs of Half-Switches.....	9
2.4 Modeling.....	22
2.4.1 Area.....	22
2.4.2 Speed.....	23
2.4.3 Power Consumption.....	25
2.5 Discussion	30
3 OPTICAL INTERCONNECTION UNIT	33
3.1 Overview of OTIS Optical System.....	33
3.1.1 Geometry.....	33
3.1.2 Symmetry	35
3.1.3 Illumination.....	36
3.2 Design and Optimization	37
3.2.1 Systems	37
3.2.2 Approach.....	38
3.2.3 Code V [®] Simulation Results.....	42
3.3 OTIS Optical System Improvements	43
3.3.1 Birefringent Computer-Generated Hologram.....	43
3.3.2 Photorefractive Beamsplitter.....	45

CONTENTS (Continued)

	Page
4	SPACE-TIME COMPANDER..... 51
4.1	CCD-Based Liquid Crystal Imager/Modulator..... 52
4.2	CCD Array for STC Application..... 52
4.2.1	Electronic Driver for CCD-Based Modulator/Imager 54
4.2.2	CCD Wafer Testing 54
4.3	STC Packaging..... 59
4.3.1	Z-Axis Chip Bonding..... 60
4.4	Space-Time Compander Optical Interfaces..... 63
5	EOCA SYSTEM ANALYSIS 65
5.1	OTIS Architecture Studies..... 66
5.1.1	Optical Transpose Interconnection System..... 66
5.1.2	Terminology..... 69
5.1.3	OTIS-Mesh..... 70
5.1.4	OTIS-Hypercube..... 71
5.1.5	OTIS-Expander 73
5.1.6	Scalable Multibutterfly Construction..... 76
5.1.7	Bit-Parallel Crossbar..... 79
5.2	Comparison of Electrical and Optical Interconnection..... 80
5.2.1	Assumptions..... 82
5.2.2	Definition of Interconnection and Estimation of Energy..... 83
5.2.3	Speed and Energy of Off-Chip Electrical Interconnections..... 88
5.2.4	Speed and Energy of On-Chip Electrical Interconnections 96
5.2.5	Optical Interconnections 99
5.2.6	Effects of Technology Scaling.....108
5.2.7	Conclusions.....110
5.3	Manufacturing Cost Modeling.....111
5.3.1	Yield Models.....113
5.3.2	Shuffle Comparison117
5.3.3	Conclusions.....122

CONTENTS (Continued)

	Page
6 REFERENCES.....	123
APPENDICES	
A	128
B SUPERBUFFER DESIGN	129
C DETECTOR REQUIREMENTS.....	131
D TRANSMITTER AND RECEIVER STEADY-STATE CURRENTS DUE TO AMPLIFICATION	134
E ESTIMATION OF OPTICAL TIME-OF-FLIGHT DELAY.....	135
F MQW MODULATOR DRIVER DESIGN.....	136
G VCSEL DRIVER DESIGN	139
H MATHEMATICA CODE.....	142

ILLUSTRATIONS

	Page
1-1 Basic Building Block of Electro-Optic Computer Architectures.....	1
2-1 Optimization of OTIS	5
2-2 Sideview of OTIS.....	6
2-3 Cross Section of a Flip-Chip Bonded Si/PLZT Smart Pixel.....	7
2-4 Block Diagram of a 16-Channel Switch	9
2-5 Layout of a 16-Channel Switch	10
2-6 Half-Switch (Design A).....	12
2-7 Block Diagram of Half-Switch (Design B).....	14
2-8 Output Circuit of Design B	15
2-9 (a) Schematics of Control Signal c0	17
(b) Schematics of Control Signal c1	17
(c) Truth Table to c0 and c1.....	17
2-10 Contention Circuits of Design B.....	20
2-11 Schematics of the Set-Up Used to Evaluate the Optoelectronic OTIS Chip and the Three-Level Logic Transmission.....	31
3-1 2-D Cross-Section of a Symmetrical 4096 Channel OTIS With Photos of Experimental Input and Output Illustrating Transpose Operation.....	34
3-2 Bi-Directional System Showing Transmitter/Receiver Offset and Resulting Crosstalk	35
3-3 Lenslet Symmetry in OTIS	36
3-4 Off-Axis Area-Multiplexed Illumination System With Details of Area Overlap (Multiplexing) and Folded Geometry	37
3-5 Code V [®] Perspective View (VIE;VPT) of a 256 Channel OTIS Showing Input and Output Planes, Four Transmitter Lenslets, Ten Receiver Lenslets, Lens Substrates, and Ten Objects	39
3-6 Illumination Lenslets and Interconnect Lenslets May be Superimposed in the Same Plane by Utilizing Birefringent Computer-Generated Holograms (BCGH).....	43

ILLUSTRATIONS (Continued)

		Page
3-7	Birefringent computer-Generated Hologram Implementing Both the Illumination Lenslets and the OTIS Interconnection Lenslets	44
3-8	Modulator Input Generated by BCGH Illumination Lenslets.....	44
3-9	Spots in the Intermediate Image Plane (the Focal Plane of the Interconnected Elements) of the OTIS System.....	44
3-10	Beam Scans of Spots in the Modulator Plane and Intermediate Image Plane of the OTIS System.....	45
3-11	Overall System Schematic for a Free-Space Optical Interconnection System	45
3-12	Experimental and Theoretical Comparison of the Angular Selectivity of the Diffraction Grating.....	49
3-13	Transmission Efficiency of Both PRBS and PBS Over the Operational Range Needed for f/4 Interconnection System	50
4-1	Schematics of One Superpixel of the Space-Time Comander.....	51
4-2	Floorplan of Space-Time Comander Wafer.....	53
4-3	Floorplan of Space-Time Comander Chip.....	55
4-4	Space-Time Comander Superpixel (8×8 CCD Array) Layout	55
4-5	Electrical Test Setup for the CCD Superpixel Array.....	56
4-6	Optical Setup for Testing the Image Function of the CCD-Based STC.....	58
4-7	Cross Section of the STC with Conductive Feedthroughs in the Supporting Glass.....	60
4-8	Typical Microsphere Distribution of Z-Axis Adhesive Under the Contact Pad on the CCD Chip.....	62
4-9	The Concept of Using Microlens Array for Mapping Optical Images to Physically Separated CCD Arrays.....	64
5-1	The OTIS System.....	67
5-2	OTIS Interconnections	68
5-3	OTIS-Mesh Simulation.....	71
5-4	OTIS-Hypercube Simulation	73
5-5	OTIS-Expander - Small Groups Case.....	75
5-6	OTIS-Expander - Large Groups Case.....	75
5-7	Multibutterfly Construction	78
5-8	Definition of Interconnection Discussed in the Context of this Section.....	84
5-9	Electrical Model of Interconnection for the calculation of Energy Requirement.....	84

ILLUSTRATIONS (Continued)

		Page
5-10	Average 4-bit Transmission Through the Channel, (a) Non-Return to Zero, (b) Return to Zero	87
5-11	Model of Off-Chip Electrical Interconnection.....	88
5-12	Speed Performance and Energy Requirements of Off-Chip Electrical Interconnections as a Function of Interconnection Length for Serial and Parallel Terminated Lines in the case of One-to-One Connections.....	95
5-13	Model of On-Chip Interconnection.....	96
5-14	Speed Performance and Energy Requirement of On-Chip Electrical Interconnections as a Function of Interconnection Length for Different Loading Conditions.....	98
5-15	Model of Interconnection Using Free-Space Optics.....	99
5-16	Speed and Energy Comparison Between Off-Chip Electrical and MQW-Based Optical Interconnects for One-to-One Connections.....	104
5-17	Speed and Energy Comparison Between Wafer-Scale Electrical and MQW-Based Optical Interconnects for One-to-One Connections.....	105
5-18	Speed and Energy Comparison Between Off-Chip Electrical and VCSEL-Based Optical Interconnects for One-to-One-Connection.....	107
5-19	Speed and Energy Comparison Between Wafer-Scale Electrical and VCSEL-Based Optical Interconnects for One-to-One Connection	108
5-20	Comparison of Negative Binomial and Poisson Yield Models for a 1994 CMOS Process	113
5-21	The Cost of 1994 High Volume CMOS as a Function of Chip Area Including Yield	114
5-22	Cost of a 250 μm Pitch Array of 50 μm^2 VCSELs on a Single Chip as a Function of the Array Size	115
5-23	Cost of a 256-Element VCSEL Array as a Function of the Number of Chips Solder Bumped in the Array	116
5-24	Yield of Solder Bumps, Silicon and MQW Modulators.....	116
5-25	Cost Comparison of Monolithic HFET-SEED GaAs vs. Hybrid CMOS-SEED as the Chip Size Increases.....	117
5-26	The Architecture of the MCM-D Layout.....	118
5-27	Process Flow for a Module for CMOS Chips Flip-Chip Solder-Bumped to MCM-D....	119
5-28	Architecture of Optoelectronic MCM.....	119
5-29	Architecture of Free-Space Optical Shuffle Network.....	120

ILLUSTRATIONS (Continued)

	Page
5-30 Process Flow for Optoelectronic MCM Manufacture.....	121
5-31 Comparison of Manufacturing Cost of VCSEL/CMOS (Dashed Line) and CMOS Chips (Solid Line) Interconnected by the Shuffle-Exchange Network with Optics/MCM for an MCM, Respectively	122
B-1 An N-Stage Superbuffer Used to Drive Large Capacitive Loads.....	129
C-1 Photodetector Input/Output Waveform Used in the Calculations	131
C-2 Plot of Eq. (C-2) and Eq. (C-4).....	133
E-1 Geometrical Assumptions of the Optical Interconnect Scheme	135
F-1 CMOS MQW Modulator Driver Circuit.....	136
G-1 CMOS VCSEL Driver Circuit.....	139

TABLES

	Page
2-1 Calculations for Speed and Area.....	23
2-2 Calculations for Power Consumption and Density of the Electronic Switches.....	27
3-1 Refractive Lenslet Listing Showing NSS Range, Radius of Curvature, Glass, and Decentration Terms.....	40
3-2 Diffractive Lenslet Listing Showing NSS Range, HOE Coefficients, and Decentration Terms	41
3-3 Spot Size Simulation Results with Code V [®]	43
5-1 OTIS Modeling	69
5-2 VLSI and Electrical Packaging Constants	94
5-3 Optical Routing/Power Supply Constants.....	100
5-4 Photodetector Constants.....	102
5-5 MQW Modulator Technology Constants.....	103
5-6 VCSEL Constants	107
5-7 Effect of Scaling the VLSI on the Various Interconnection Technologies.....	110
5-8 Cost, Yield and Architecture Parameters.....	112

Section 1

INTRODUCTION

Future high performance parallel computing systems must rely on the development of a high throughput three-dimensional interconnection system. To maximize the throughput while minimizing crosstalk and power requirements, the Electro-Optic Computing Architecture (EOCA) program seeks to add global inter-wafer optical interconnection capability to locally connected parallel processors (e.g., the Hughes 3-D computer and Multiple-Chip-Module processors). This would enable us to (a) free the processor from its present I/O limitations, allowing efficient parallel communication with optical memories and sensors; (b) allow an efficient coupling of optical co-processors to handle fine grain image processing and global 2-D operations at throughput rates exceeding terabits/sec; (c) allow efficient sorting operations to be carried out through the use of optical shuffling, with expected throughput enhancements of about 100:1.

The system under investigation combines 3-D VLSI technology with free-space optoelectronic interconnection modules. The 3-D stacks containing the processing capabilities of the system are assembled with the optoelectronic modules for global communication allowing fast and efficient data routing and/or sorting. The objective of the EOCA program is to develop multi-function electro-optic interfaces and optical interconnect units to enhance the performance of the parallel processor system and form the building blocks for future electro-optic computing architecture. Specifically, three multi-function interface modules — an Electro-Optic Interface (EOI), an Optical Interconnection Unit (OIU), and a Space-Time Compander (STC) — will be developed. A conceptual schematic of the EOCA system is depicted in Fig. 1-1.

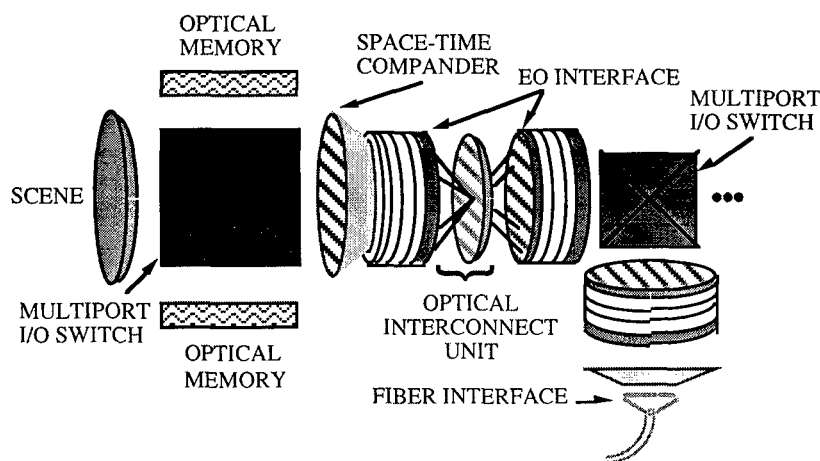


Figure 1-1. Basic building block of electro-optic computer architectures.

The electro-optical interface consists of an array of light modulator's flip-chip bonded to a chip that contains the drivers for the modulators, as well as the light detectors and their associated amplifiers. Silicon electronic switches are included to provide local connectivity and, thus, data packets can arbitrarily route between its inputs and outputs. To couple the 3-D computer system with optical interconnection, it is critical to build opto-electronic interface devices by integrating light detectors and light transmitters with silicon chips or wafers. The EOI allows bi-directional communication between an electronic processor through a parallel optical interconnect.

A 4×4 array of PLZT modulators has been flip-chip bonded to silicon driver chips. The drivers are fabricated in standard CMOS and the high driving voltages required for PLZT modulators are provided through a voltage amplification by the drivers using only a 5 V voltage supply. A 16×16 array of MQW modulators has also been fabricated and flip chip bonded to a standard CMOS silicon driver chip. We also built a 16×16 array of bypass and exchange switches. The switches have the unique ability that they can detect contention at any point in the system and propagate a signal back (using an intermediate voltage level of 2.5 V) to the input for which a packet had to be dropped in the network. In addition, the chip also contained 3-level test circuits for modulator drivers and receivers that have been shown operational at 100 MHz. A three-level transmission was shown operational with optical I/O using MQW diodes as light modulators and detectors.

The optical interconnection unit provides global connectivity between the different switches. The Optical Transpose Interconnection System (OTIS) is used since it provides simple, efficient, and scaleable full routing capability when used in conjunction with the appropriate electronic switches. The usefulness of the transpose interconnection has previously been shown for three architecture classes, namely, shuffle-based multistage interconnection networks, mesh of trees matrix processors, and hypercube interconnections. An experiment using off the shelf lenslets has been performed showing that 64×64 arrays can be interconnected through OTIS. We also introduced a novel modulator/illumination system consisting of an off-axis area-multiplexed lenslet array that can be combined, via Birefringent Computer Generated Hologram (BCGH) technology, into the same optical element as the interconnect optics. BCGH's have been built and demonstrated that the system packaging can be greatly simplified without compromising performance. A new device concept has also been developed and implemented where a single volume grating (PRBS: Photorefractive Beam Splitter) is used as the power coupling component in place of the usual Polarizing beamsplitter improving greatly the uniformity of the light transmission in the system. In addition, a detailed analysis of the OTIS optical system has been conducted as part of design analysis work. The analysis includes geometrical relations for the optical system (lens positions, system length, and system volume) as well as Code V simulations of the entire systems.

The concept of the space-time compander is developed to match fine-grain (e.g., 1024×1024) images with the coarse-grain (e.g., 128×128) processor array. The matching is performed by grouping every set of 8×8 pixels in the fine-grain image into a superpixel. Each superpixel is then

registered with the corresponding processor in the processor array. By either compacting 8×8 pixels into a superpixel or expanding a superpixel into 8×8 pixels, the STC provides a bi-directional communication between a fine-grain image and a coarse-grain processor array. A charge coupled device (CCD) array is used to function as a serial \leftrightarrow parallel buffer array to convert the 2-D spatial (parallel) information into 1-D time (serial) information. Under this program we modify the Hughes CCD-LCLV design, CCD array and LCLV structure, to accommodate the necessary matching function. This special CCD-based compander was designed to perform combined modulation/detection functionality. We designed and fabricated the CCD array and developed the related packaging technologies. The basic CCD design was verified by the operational functions of superpixel cells. In addition, light sensitivity was observed when the STC was operated as an imager. However, no definite resolution pattern has been unambiguously observed in the CCD output signal.

The goal of the system study is to identify and analyze all the architectural implications due to the addition of the OTIS based free-space optical interconnects in locally connected processor arrays. We have developed models for electronic interconnects on the Hughes 3D wafers and free-space optical interconnect technologies that allowed us to assess the unique advantages of the optically augmented 3-D computer approach. These models have been expanded to generalized wafer scale technology, MCM (transmission lines) technology, and free-space interconnects using VCSELs, MQW modulators, and PLZT modulators.

We have optimized the architecture of the one stage-shuffle interconnection network for permutation traffic in the 3-D computer and have developed the concept of the time-dilated network. We have also evaluated the applicability of OTIS for various network topologies and architectures and found that OTIS can simulate most existing networks with only a constant algorithmic slowdown.

The cost of a computer architecture depends on the costs and yields of the underlying technology along with the systems' physical organization. To compare architectures constructed with optics and electronics, we first built cost models for the underlying technology for active devices: CMOS and VCSELs. As an example, we showed that a multichip 256 element VCSEL array is lower cost than the monolithic alternative. Then these active device models were combined with models for passive elements such as solder bumps, MCMs, and optomechanics to produce process flows for complex systems. The cost models also contain architectural features such as the shuffle-exchange wire layout area, the number of parallel channels, and the number of computational nodes. Using these models, we also showed that an optoelectronic VCSEL/CMOS/MCM/Optics implementation of a shuffle-exchange network is lower cost than the all-electronic CMOS/MCM for greater than 20 nodes.

Section 2

ELECTRO-OPTICAL INTERFACE

To effectively couple the locally connected processors with optical interconnection, it is critical to realize opto-electronic interface by integrating light detectors and light transmitters with silicon wafers. The EOI allows bi-directional communication between a coarse-grain electronic processor array and a parallel optical interconnection unit.

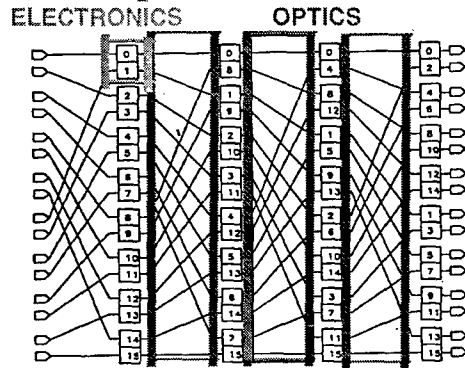
Two separate circuits, that use an optimized 2-D layout and are compatible with the Optical Transpose Interconnection System (OTIS), have been designed, implemented, and analyzed. The first design (Design A) was used as a proof of concept for the optimized 2-D layout, the second design (Design B) is a bi-directional self-routing concept that uses 3 level logic. Note that the optoelectronic implementation of this second design based on the CMOS/SEED technology⁽²⁻¹⁾ is under way.

2.1 OPTICAL TRANSPOSE INTERCONNECTION NETWORK

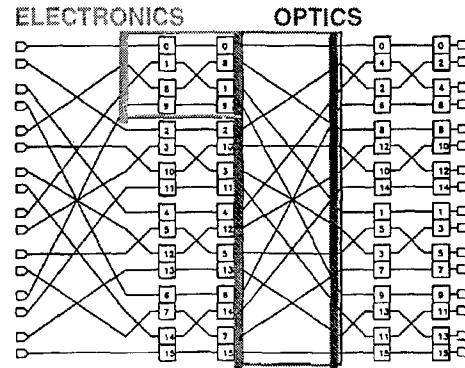
In a free-space optoelectronic multistage interconnection network (MIN), as described in [2-2], local routing is done with electronic fully connected bypass-and-exchange switches, and longer/global connections are achieved with optical links. For a MIN with N inputs and N outputs (i.e. N channels), the bandwidth and the total power consumption, both electrical and optical, of the overall system have been shown to be optimized if the N channels are partitioned into \sqrt{N} switches with \sqrt{N} channels each⁽²⁻²⁾. In Fig. 2-1, K is the number of channels per electronic switch, that is, each switch has K inputs and K outputs. To build a switching network with N channels, one can choose to have an all electronic network with no optical links at all. In this case, the MIN will consist of a single switch ($K = N$) with optical input/output. On the other extreme, one can choose to implement the same network with as much optics as possible. Then, the network will have $\log_2(N)$ stages of electronic switches with $K = 2$, and $\log_2(N)-1$ stages of optics. A third option is to choose a middle ground by setting K to be greater than 2 and smaller than N . The number of electrical and optical stages will then be adjusted to achieve a fully connected network. As shown in [2-2], as K grows beyond a certain point compared to N , that is, as more and more electronics is utilized, the bandwidth of the overall system drops dramatically. The reason for this drop is that as the electronic switch gets bigger, the maximum electronic delay inside the switch gets bigger, so the clock speed needs to be reduced, which in turn reduces the bandwidth of the network. For the range, where the bandwidth is constant, the total system power, including both the electrical and the optical power consumption, is minimized when $K \cong \sqrt{N}$. Note that if $K = \sqrt{N}$, only $\log_{\sqrt{N}}(N) = 2$ stages of electronics are required to achieve a fully connected network.

In addition, since optics is used to connect the electronic stages, only 1 stage of optics is needed for the 2 stages of electronics. This optimized point of $K \cong \sqrt{N}$, is only valid for given technology assumptions, however, further scaling investigations in⁽²⁻²⁾ showed that the trend remains the same for other technologies. It is based on this optimization that the Optical Transpose Interconnection System (OTIS) was conceived.⁽²⁻³⁾

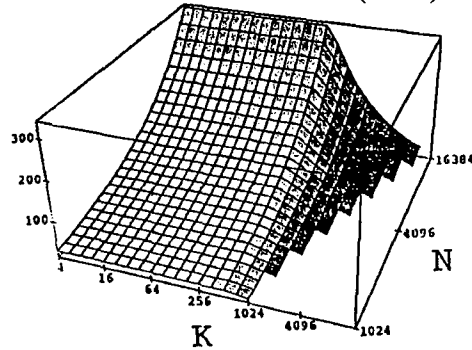
More Optics (K=2)



More Electronics (K=4)



SYSTEM BANDWIDTH (Gb/s)



SYSTEM POWER (W)

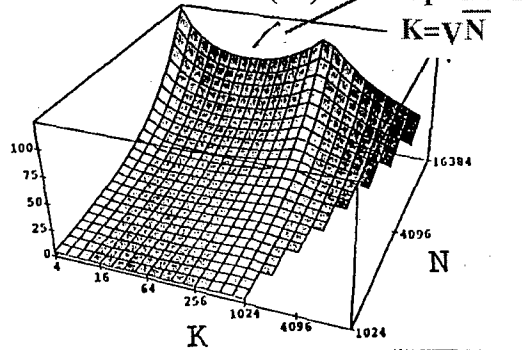


Figure 2-1. Optimization of OTIS.

Figure 2-2 shows a side view of OTIS with $N = 256$ in a 16×16 array configuration. The 256 channels are partitioned into 16 switches, with 16 inputs and 16 outputs each (i.e. $K = \sqrt{N} = 16$). There are two stages of electronic switches on either side of the single optical stage. The 16 outputs of each switch in plane 1 (i.e. arbitrarily chosen to be the left plane in Fig. 2-2) have a one-to-one optical link with one of the inputs of the 16 switches in plane 2 (i.e. the right plane in Fig. 2-2). Routing of a data packet is illustrated in Fig. 2-2. The incoming data enters one of the 16 switches. It is then routed within the electronic switch to the specific output that has the optical

link to the switch on the other side which contains the final destination of the data packet. The data is routed one more time, inside the electronic switch in plane 2, to arrive at the desired output node. Both the electronics and the optics are bi-directional so every input also acts as an output depending on the desired direction of the data flow.

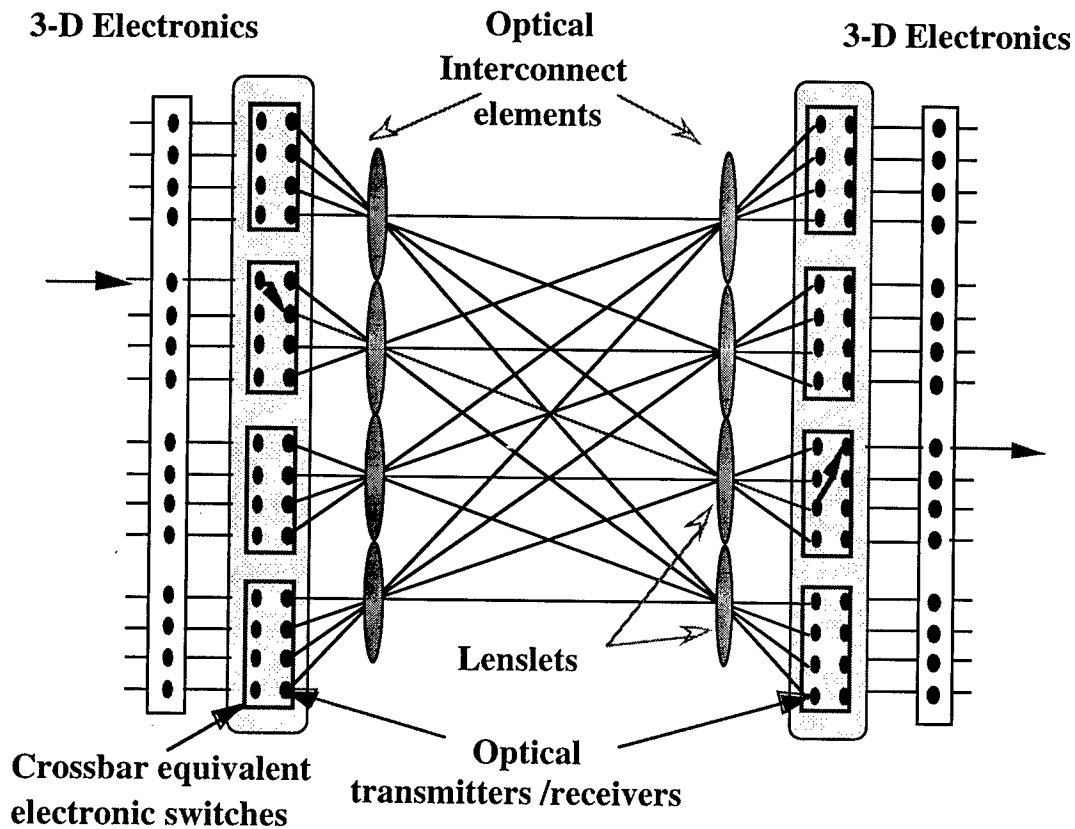


Figure 2-2. Sideview of OTIS.

2.2 FLIP-CHIP BONDED Si/PLZT SMART PIXEL

In the hybrid Si/PLZT optoelectronic technology a PLZT wafer is used both as a support substrate and for light modulation. The electronic driver circuitry is built on the silicon chips and connected to the PLZT modulator through metal bumps (as schematically shown in Fig. 2-3). The silicon chips can be tested separately before placement on PLZT to insure a high yield process. The flip-chip process then mechanically aligns the silicon wafers to the corresponding PLZT modulators. To achieve a large dynamic range PLZT generally requires 20-40 V to modulate an external light. High voltage bipolar and MOS processes are presently incapable of supporting VLSI circuit densities. On the other hand, transistor breakdown voltages in VLSI chips are too low to provide high voltage outputs directly. We have designed a special circuit capable of delivering a driving voltage swing up to 35 V from a standard 5 V power supply. This driver circuit is

fabricated using standard MOSIS 2 μm CMOS technology. High breakdown voltage of the circuit is accomplished using series connected transistors and a current-mirror like structure.

We have designed and integrated a 4×4 array of reflective PLZT modulators with the silicon driver circuit. The combination of the driver and the reflective PLZT modulator produces light modulation with a dynamic range of up to 600:1. Studies of speed response of PLZT 9.5/65/35 showed the rise and fall times to be less than 10 ns each, fundamentally limited by the driver circuitry. The Si/PLZT smart pixel is capable of building an optical link with a bit-error rate (BER) better than 10^{-14} under the following experimental conditions:

- 5 Mbits/sec data rate
- 4x4 array of 40x40 μm modulators
- Modulator bias at 60 V
- Modulator voltage swing at 25 V
- 300 μW optical input per modulator with 10 μm spot

The output power swing was measured at 100 to 200 μW per modulator over the array.

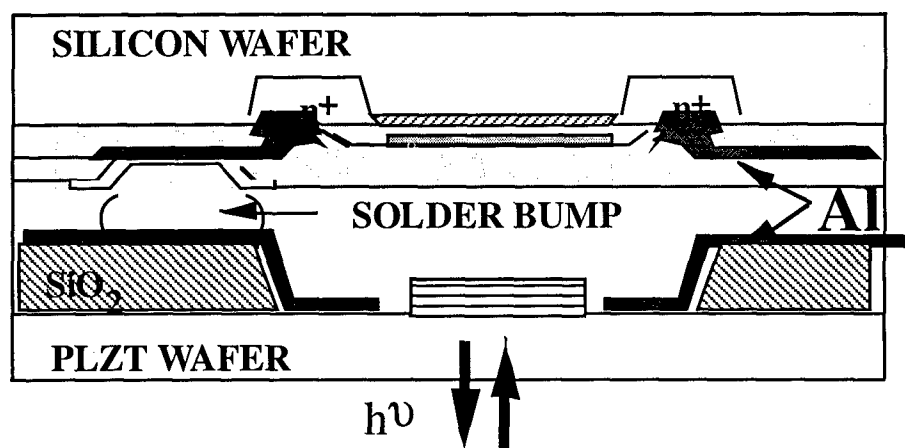


Figure 2-3. Cross section of a flip-chip bonded Si/PLZT smart pixel.

2.3 ELECTRONIC SWITCHES

Throughout this section, the design and experiments of a switch with 16 channels (i.e., $K = 16$, $N = 256$) are described. However, every derivation and the modeling analysis in Section 2-4 apply to switches with an arbitrary number of channels.

2.3.1 Overview

A switch is implemented by cascading 2×2 bypass-and-exchange switches, partitioned into two 2-to-1 multiplexers, called half-switches. A pair of such half-switches is called partner half-switches. Every half-switch gets an input of its own, plus the input of its partner half-switch as its second input. Depending on the control signal, each half-switch transmits one of the two possible inputs to its single output. A complete switch consists of $\log_2 N = 4$ stages of $N = 16$ half-

switches each, for a total of $N \log_2 N = 64$ half-switches. The block diagram of a 16 channel switch is shown in Fig. 2-4.

As an example in Fig. 2-4, at the third stage, the 8th and the 12th half-switches are partner half-switches. Assuming the direction of data flow is from left to right, number 8 in the middle (i.e. in the third stage) gets two inputs, one from number 8 of the previous stage (i.e. the second stage) labeled A, and one from its partner labeled B, which is the output of number 12 of the previous stage. Equivalently, this is the first input of number 12 of the third stage. Then it outputs one of these inputs to node C, where D is a floating node, that is, no transistor is pulling node D up or down. On the other hand, when the direction of data flow is reversed so that it goes from right to left, C and D become the two inputs to number 8, with A as its single output, and B is floating.

Laying out the switch as it is shown in the block diagram of Fig. 2-4, with all of the 16 inputs on one side, and the 16 outputs on the other side, gives a 1-D layout, suitable for VLSI implementations. However, if all the inputs and outputs are distributed in a 4×4 array of constant pitch, one can achieve a 2-D layout, suitable for optoelectronic implementations. The $\log_2 K$ half-switches from each stage with the same number are grouped together in the layout as in Fig. 2-5.

In Fig. 2-5, each triangle represents an input/output pair since the switch is bi-directional. Every rectangle in the figure contains $\log_2 K = 4$ half-switches, which form a channel. In addition, every connection between two half-switches of different channels imply that those are partner half-switches. Therefore, their inputs and outputs are connected to each other as shown in Fig. 2-4.

When 1-D and 2-D layout strategies are compared, the maximum wire length in the whole switch is much shorter in the 2-D layout. Thus, in terms of RC-limited maximum operation frequency, the 2-D layout has an advantage over the 1-D layout. In addition, if the 1-D layout is used for the optoelectronic implementation, extra routing is required between the actual inputs/outputs of the half-switches, and the transmitters/receivers of the system, since the latter are likely to be laid out on a 2-D array with a constant pitch. As a result, the 2-D layout is advantageous in terms of total area as well as maximum operation frequency.⁽²⁻²⁾ This advantage is magnified when bigger size networks are employed.

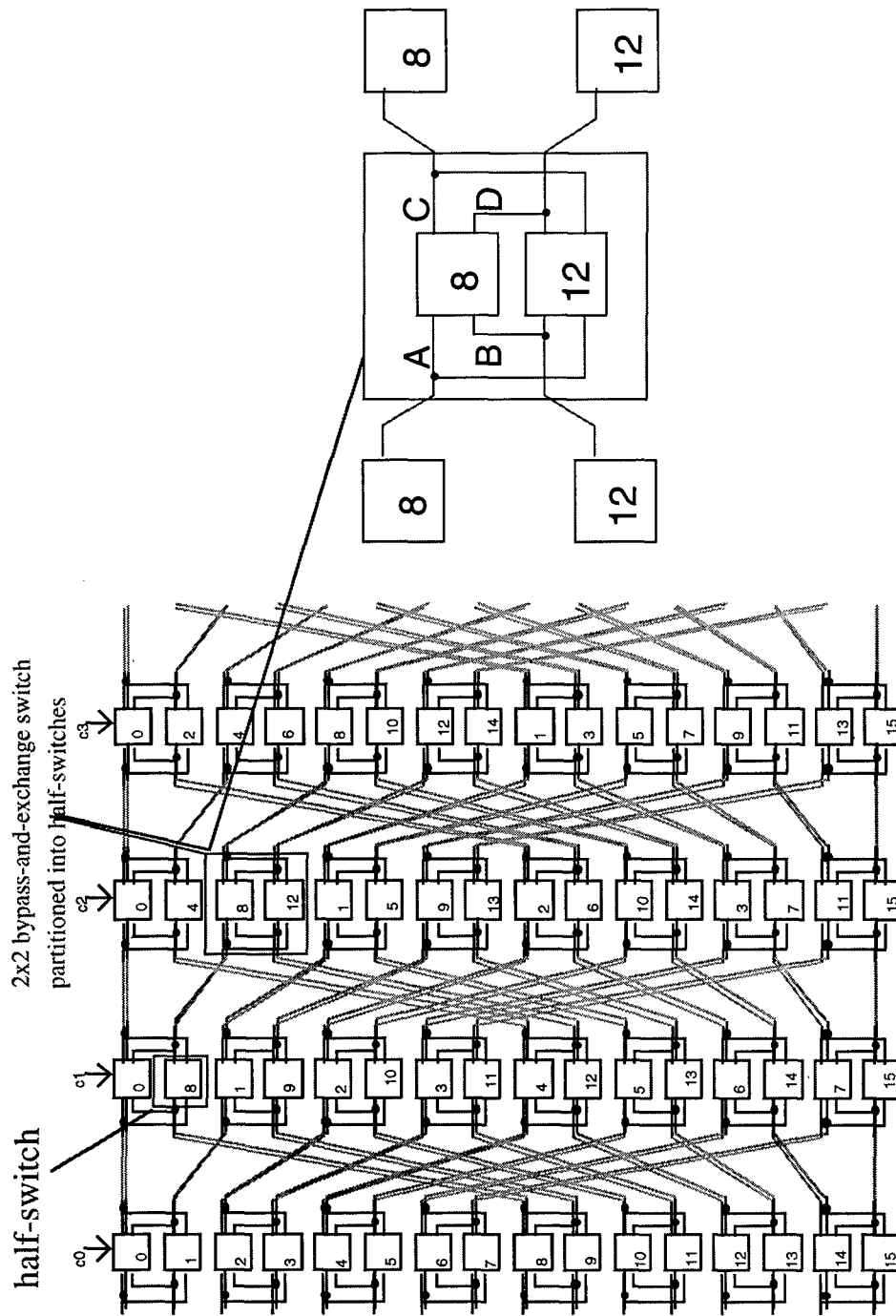


Figure 2-4. Block diagram of a 16-channel switch.

2.3.2 Designs of Half-Switches

Two different half-switches have been designed for the OTIS. The first one, design A, is a simple, bi-directional 2-to-1 multiplexer, that has no additional features that could be desired for a more powerful system. This was built as a proof of concept for the 2-D layout, and the operation of a half-switch as the building block of the complete switch. The second design, design B, is a novel self-routing half-switch, that can detect contention, and drop-and-resend data packets.

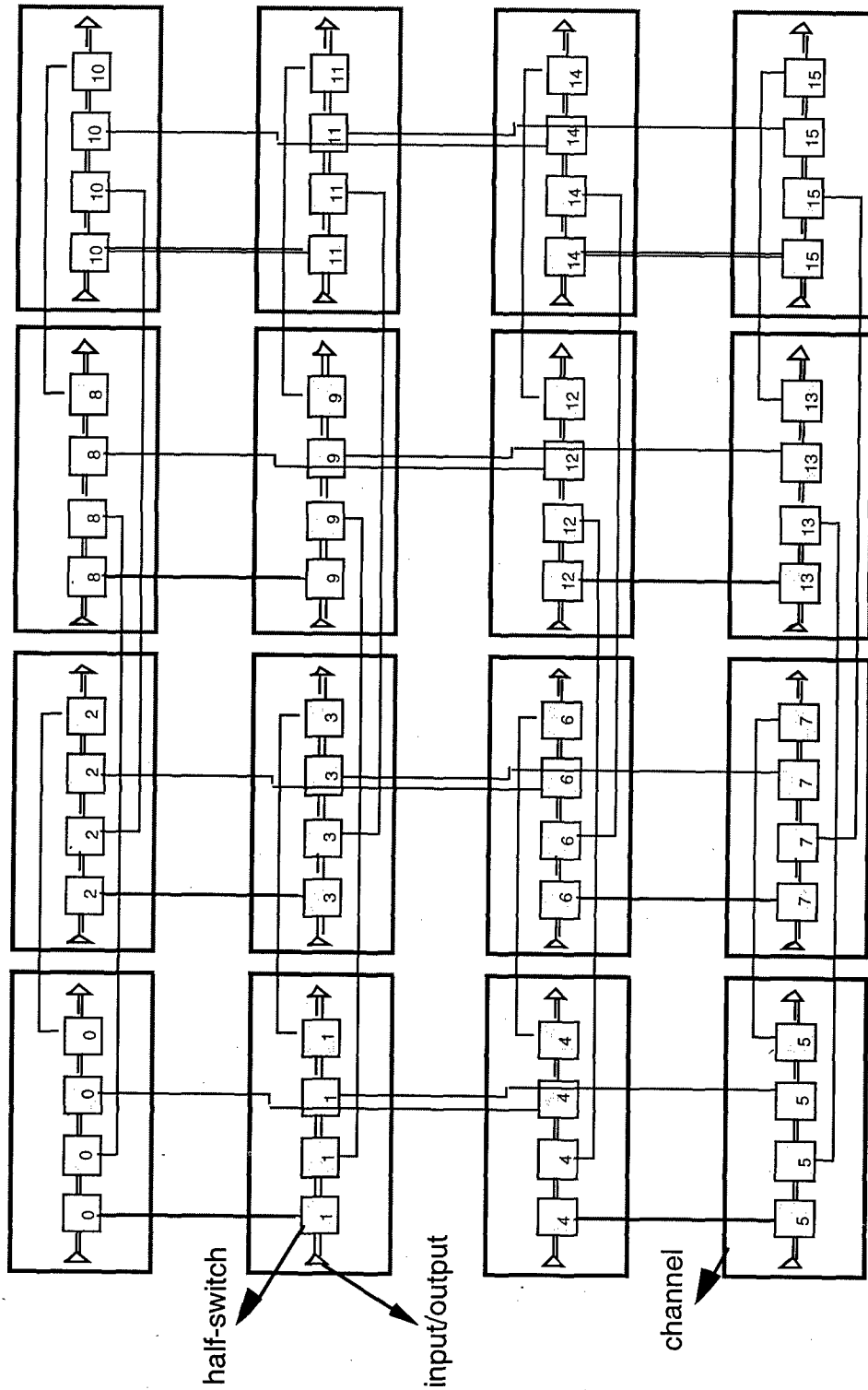


Figure 2-5. Layout of a 16-channel switch.

2.3.2.1 Design A

The block diagram of design A is shown in Fig. 2-6.⁽²⁻⁴⁾ The half-switch uses an external direction signal that is also broadcast to every other half-switch in the entire switch. This signal determines the direction of the data flow. The direction signal is arbitrarily chosen to be 1 (dir is the direction signal) for a left to right data flow. In this case, x0 and x1 are the two inputs and y0 is the output, while y1 is floating. Another external control signal is sent to the half-switch, controlling which input channel it should transmit to its output. Again arbitrarily, c is chosen to be 1 (c is the control signal) when x1 is to be transmitted, and similarly, c = 0 causes x0 to go through. If direction is reversed (i.e. dir = 0), then c = 0 causes y0 to be sent to x0, and c = 1 causes y1 to be sent to x0.

In this design, only four control bits are used for a four stage switch, labeled c0 through c3 (refer to Fig. 2-4). The same control bit is sent to all the half-switches on the same stage. As a result, only the final output destination of a single input can be determined, whereas the remaining 15 inputs go to the other 15 outputs without contention.

To speed up the overall system and to make it scaleable, the control signals, c0-c3, are fed into the switch in a pipelined fashion. In other words, the control bit, c1, which belongs to the second stage is delayed externally by the same amount of time it takes for a signal to propagate through the first stage of half-switches. Similarly, the control bit to the third stage, c2, is delayed twice that amount, and so on. This method ensures that the control signal and the inputs of a given stage arrive at the same time at the desired half-switches. Then, the speed of the overall system is directly equal to the speed of a single half-switch. As the switch size increases, the number of stages and the total number of half-switches increase but the overall speed stays constant since the propagation delay of a single half-switch is constant.

Figure 2-6(b) shows the circuit schematics for design A. The numbers next to the transistors give width/length in units of λ . The half-switch is implemented with only 24 transistors. For a given direction signal, half of the transistors are not used. As an example, if dir = 1, transistors M13 - M24 are not used since M13 and M16 block their final output. In addition, for that given direction signal, the control signal determines which input will be transmitted. For example, if c = 1, M5 and M8 block x0, and x1 is sent through. Similarly, if c = 0, M1 and M4 block x1, and x0 is transmitted. Effectively, for a given direction and control signal, the input is inverted twice to reach the output. Due to this simple structure, high-speed operations are achievable. For a 2.0 μm CMOS technology, simulations showed a maximum speed of 250 Mbits/s.

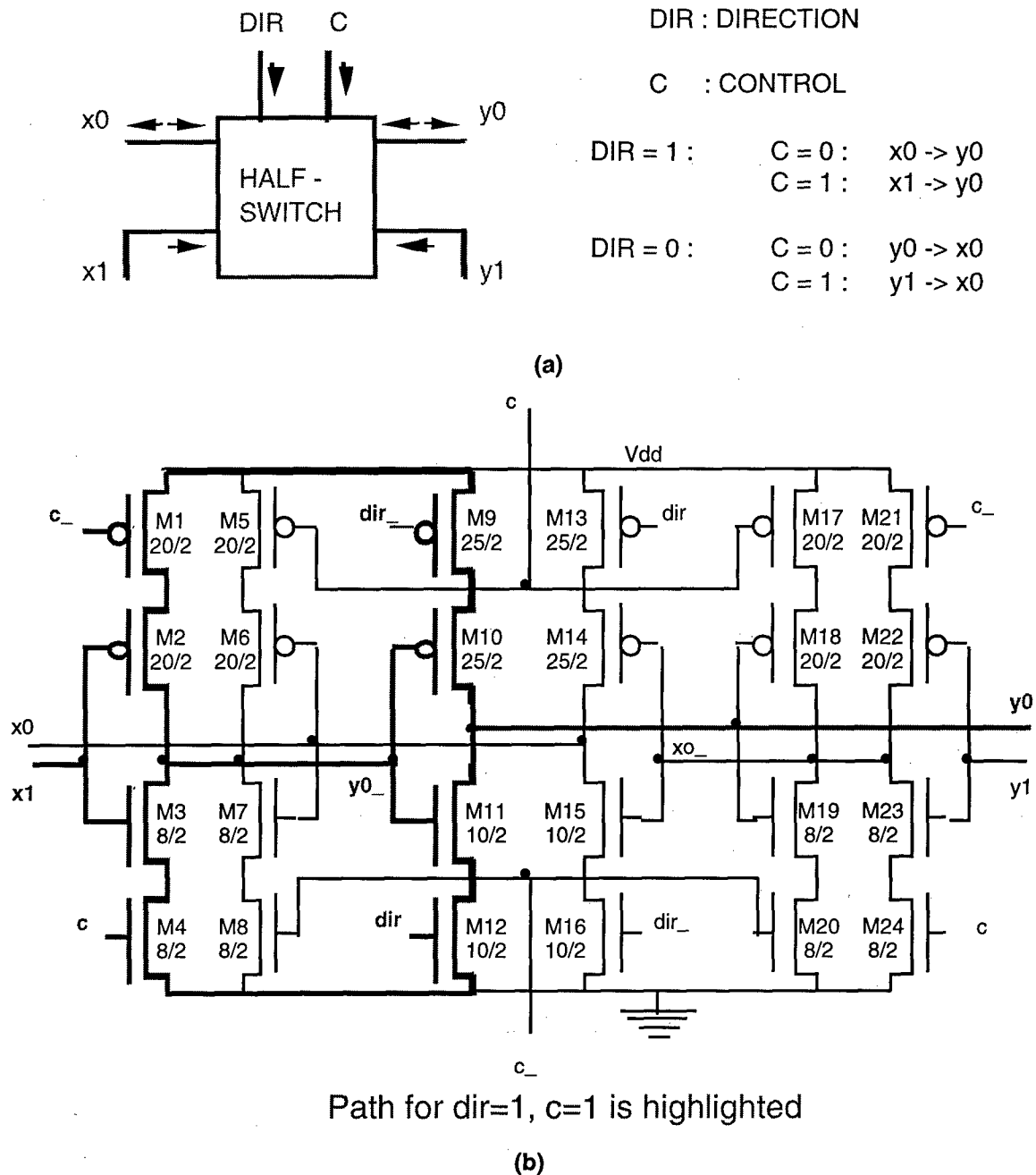


Figure 2-6. Half-Switch (design A). (a) Block diagram and truth table. (b) Circuit schematics.

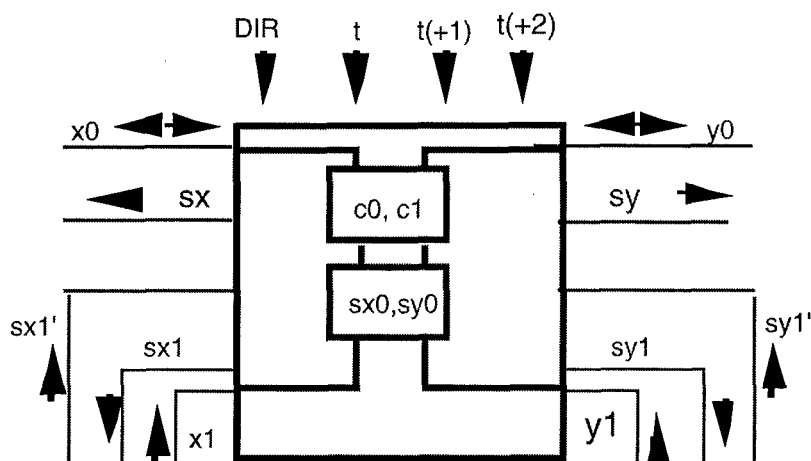
This circuit has been implemented through Mosis and its operation has been verified at 90 Mbits/s experimentally. The difference between the experimental and the simulated results is believed to arise from the experimental setup rather than the circuits themselves. The main problem may be due to the probes being used, that are not suitable for higher-speed measurements.

2.3.2.2 Design B

Design B is built upon design A, but it adds functionality to the switch operation. The block diagram is shown in Fig. 2-7. It still acts as a 2-to-1 multiplexer. However, it has built-in self-routing capability, that is, the control bit for each half-switch is computed internally. Every input packet contains as a header, the address of its desired output destination (i.e., for $N = 256$ channels, $\log_2 N = 8$ bits of address are needed). As data packets are presented, the half-switches in the first stage process the first bit of each of their inputs, and decide on their control signal. The remaining 23 bits are then transmitted untouched. The same processing is done in the next stages until the data packet arrive at their output destination.

As packets are transmitted through the switches, two of them may have to use the same half-switch to arrive at their output, and thus, there is contention (hot spot). In this case, the half-switch transmits one of the inputs in a deterministic way, and drops its other input. At the same time, to ensure that the dropped data is not lost, a contention signal is generated within the half-switch, where the blocking happened. This contention signal propagates in the direction opposite to the data flow, and follows backwards, the path that the dropped packet of data had followed up to that point. Once it reaches the dropped packet's input buffer, it sets the input buffer to resend the same packet, so that all the information is eventually routed through the network.

In this design, the direction of data flow is again determined by an external direction signal supplied to all the half-switches. In addition, an external transmission signal (t is the transmission signal) is provided to inform each half-switch that data transmission is occurring. This signal is set to 1 if the incoming bit is a data bit, and is set to 0 if it is an address bit. Therefore, for all the half-switches at a given stage, $t = 0$ during the first cycle of a data packet, and $= 1$ for the remaining 23 cycles. Just like in design A, the transmission signal is pipelined, that is, delayed by the same amount of time that the input takes to reach that stage. As a result, a transmission signal of 0 for 1 cycle, and 1 for 23 cycles, propagates from stage to stage at the same speed that the data propagates, with the 0 bit arriving at a stage when the control signals are to be computed at that stage (i.e. the incoming bits are address bits). This way, a single pulse of $t = 0$ at the input buffer stage enables all the half-switches in the entire switch to know exactly when to process the incoming bits as their address bits rather than data bits. As each data packet is introduced into the pipeline, first the address bits are processed. This processing of the header of a data packet takes exactly $2\log_2 N$ cycles, where a cycle is equal to the duration of a single bit. This is the time it



DIR : DIRECTION

t : TRANSMISSION (= 1 FOR DATA BIT,
= 0 FOR ADDRESS BIT)

t(+1) : t DELAYED BY ONE CYCLE

s : CONTENTION DETECTOR

c0 : CORRESPONDS TO " C " OF DESIGN A

c1 : =1 IF c0 IS A DON'T CARE

Figure 2-7. Block diagram of Half-Switch (design B).

takes for the last address bit (i.e. the $(\log_2 N)^{\text{th}}$ bit) to be processed by the last stage (i.e. the $(\log_2 N)^{\text{th}}$ stage). After that, the speed of that channel's throughput is equal to the speed of a single half-switch, since the pipeline is completely filled up at this point.

The design of the half-switch consists of three separate circuits, namely, the output, the control, and the contention circuits.

2.3.2.2.1 Output Circuit

The circuit schematics of design B's output is given in Fig. 2-8. The numbers next to the transistors indicate their width/length ratio in units of λ . Because of the contention possibility between data packets, the output circuit of design A is modified. In design B, the signal c0 is equivalent to the signal c of design A, that is, it determines which one of the inputs will be transmitted to the output. However, there is the possibility that neither of the two inputs will be active. This is the same thing as if both those inputs wanted to use its partner half-switch to reach their final destinations. Then c0 is a "don't care", that is, the half-switch does not care which input

is transmitted, since neither of the inputs will be using that path. In that case, c_1 is set to 1. In other words, c_1 is a 1 if c_0 is a "don't care", and is a 0 otherwise.

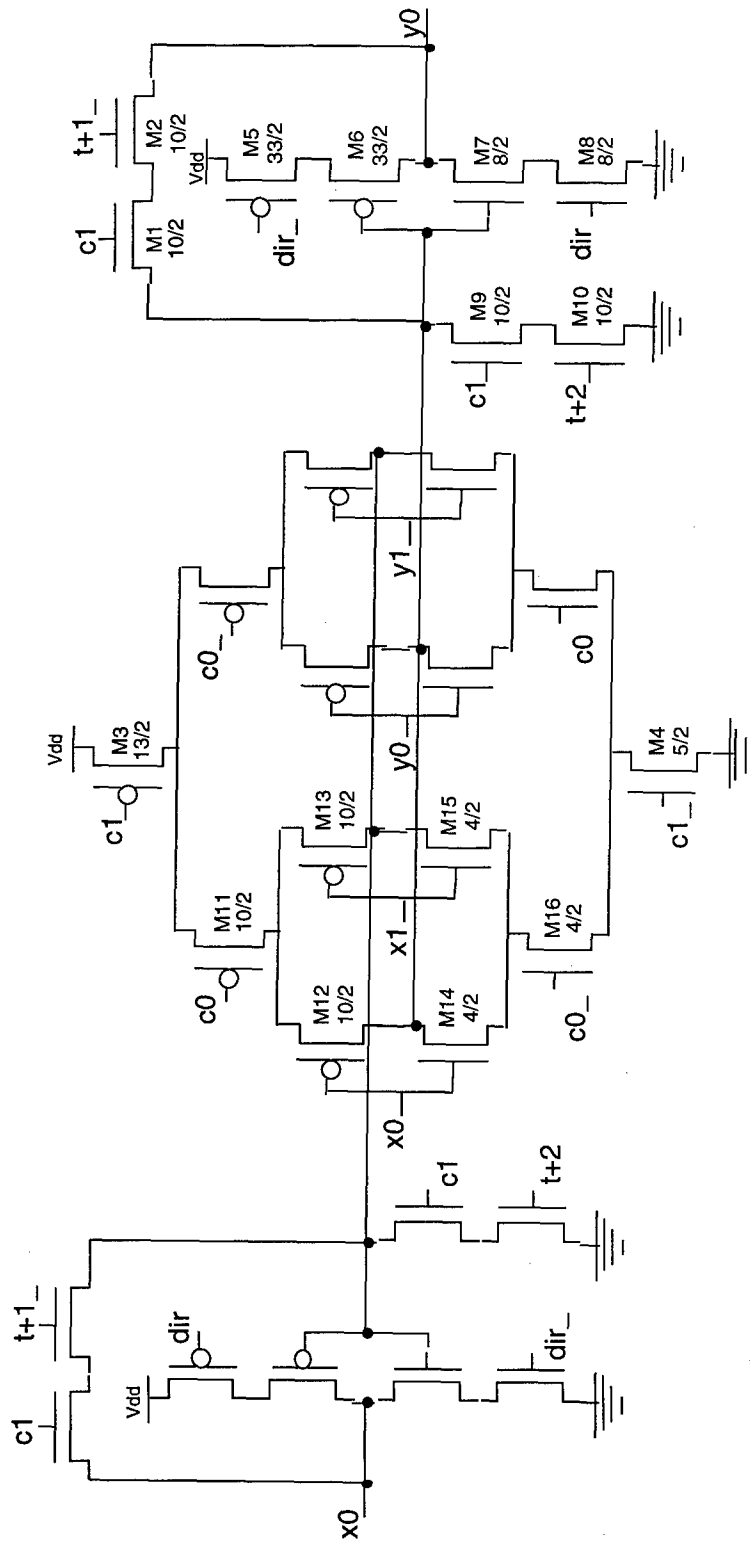


Figure 2-8. Output circuit of design B.

To increase the bandwidth of the overall system, if the half-switch, say on the second stage, finds out that none of its inputs will be occupying that path, then somehow, it should send a signal to the next stage (i.e. the third stage), indicating that the half-switches on the third stage should transmit their other inputs no matter what. Because of this third possible output state of a half-switch, the communication between stages have to be modified from design A. One way is to add a second wire between all the stages so that the three levels of output can be transmitted with two separate wires. However, as the system size grows, the global wiring becomes very difficult, especially in terms of device area.

Instead, a novel technique is employed. The single output wire between consecutive stages is designed to be able to carry three levels of logic. For a $5V V_{dd}$, these levels would be $0V$, $5V$, and $2.5V$ as the extra third level. Continuing our example from the previous paragraph, if a given half-switch computes its c_0 to be a "don't care" when $t = 0$ (i.e. it is the cycle to compute its address bit), then in the very next cycle (i.e. when $t(+1) = 0$, where $t(+1)$ is simply the signal t delayed by one cycle), the transistors M1 and M2 will be turned on (refer to Fig. 2-8). At the same time, transistors M3 and M4 will be off, and disable the remaining of the circuit except for transistors M5 through M8. This will provide a direct feedback path between the input and the output of the final inverter (i.e., M5 through M8). By appropriately sizing these transistors, the only stable voltage level, when M1 and M2 are on, can be set at $2.5V$. Note that the only possible cycle that a half-switch will set its output to $2.5V$ is when the half-switches on the next stage are computing their own control signals.

At this point, there is a direct path from V_{dd} to GND, so to reduce the power consumption, after another cycle (i.e. when $t(+2)=0$ and $t(+1)$ goes back to 1), M2 turns off. At the same time, the transistors M9 and M10 pull the inverter's (i.e., M5-M8) input to $0V$ and cause its output to go to $5V$. If this happens, then the power consumption of that half-switch for the remaining 22 cycles is exactly zero. Therefore, on average, a half-switch that has to output a $2.5V$ signal indeed consumes less power than an average half-switch that is active for the entire 24 cycles of a data packet.

The functionality of the direction signal and the remaining transistors are exactly the same as in design A. The direction signal disables half of the circuit, whose output is not necessary (i.e. in the wrong direction). The above example assumes that $dir = 1$ so that the transistors, that are not numbered, have no effect on the output of the half-switch.

In addition, since the whole design is asynchronous, the timing is crucial. Therefore, the transistors in Fig. 2-8 are designed such that the propagation delay of the half-switch's output circuit will be the same whether it is transmitting $0V$ or $5V$ through the two inverters (i.e., M11-M16 and M5-M8) or it is outputting $2.5V$ through M1 and M2.

2.3.2.2.2 Control Circuit

In this part of the half-switch, the address bit is computed for routing the data packets. As explained before, each half-switch knows that the incoming inputs are address bits when the transmission signal is set to zero (i.e. $t = 0$). The two transistors, M1 and M2 (refer to Fig. 2-9(a) and 2-9(b)), ensure that the determination of the control signals, c_0 and c_1 , occur only when $t = 0$. Otherwise, the whole circuit is disabled, and the two control signals float at their previously computed values for 23 cycles until t is 0 again (i.e. until it is the beginning of a new data packet for that half-switch).

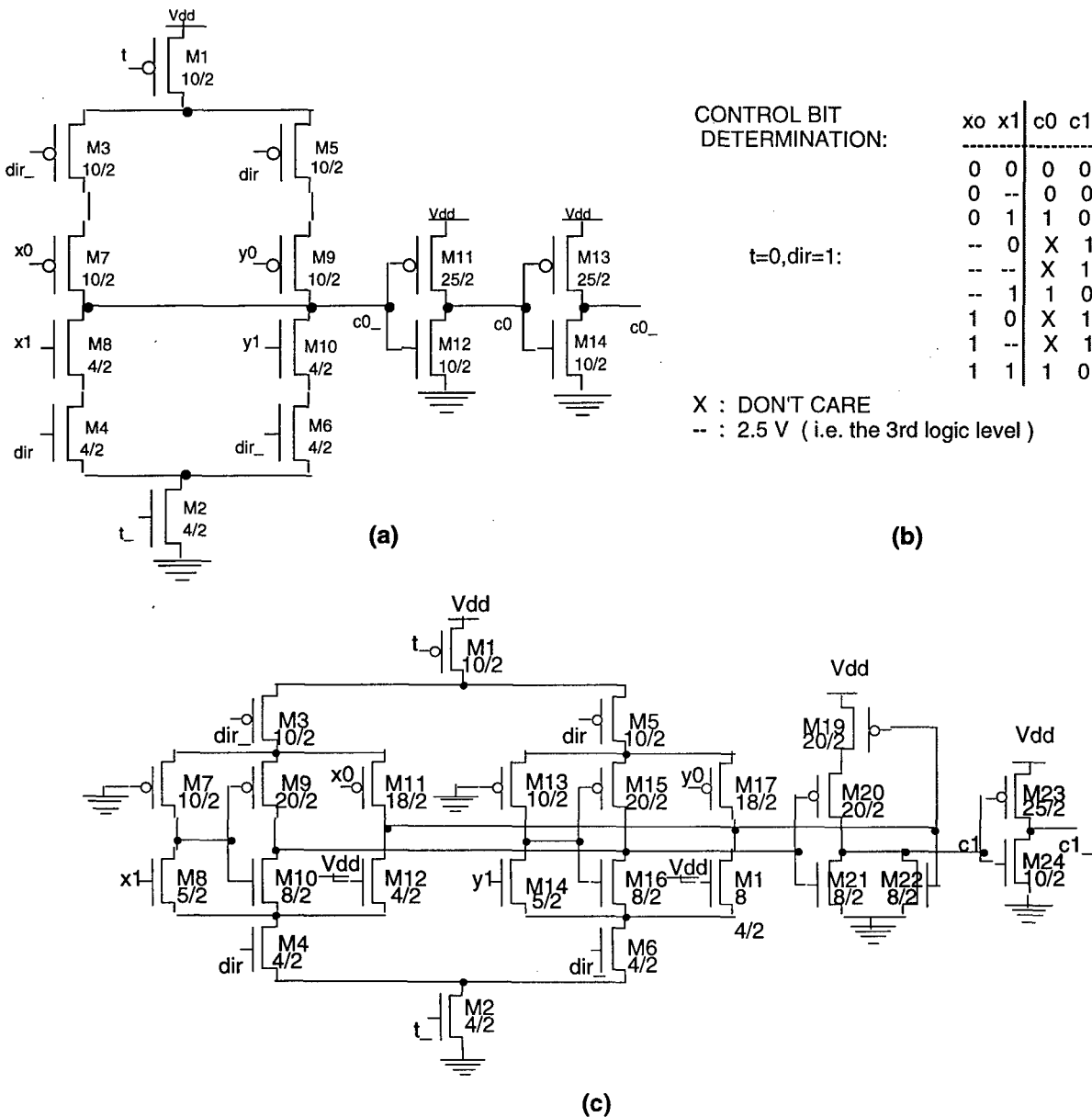


Figure 2-9. (a) Schematics of control Signal c_0 . (b) Schematics of Control Signal c_1 . (c) Truth table to c_0 and c_1 .

In addition, for both of the control signals, the transistors M3-M6 provide that the output will be determined by the inputs coming from the right direction. In other words, if $dir = 1$, x_0 and x_1 determine c_0 and c_1 , and otherwise, y_0 and y_1 determine c_0 and c_1 . c_0 , refer to Fig. 2-9(a), is the control bit that tells the half-switch which input will be transmitted to the single output during the next 23 cycles. If $c_0 = 0$, then x_0 is sent to y_0 , and if $c_0 = 1$, then x_1 is sent to y_0 (assuming $dir = 1$). During $t = 0$, if $x_0 = 0$ (arbitrarily chosen), then that input uses the half-switch for transmission. On the other hand, if $x_1 = 1$, then the second input uses that half-switch. If either one of them is 2.5V, then that input is neglected, since it means that there will not be any data coming from that channel. Again arbitrarily, if both $x_0 = 0$ and $x_1 = 1$ at the same time, that is both inputs want to use that half-switch, then x_0 is dropped and x_1 is transmitted in a deterministic way. The complete truth table for c_0 is given in Fig. 2-9(c).

In the schematics, assuming that $t = 0$ (so that M1 and M2 are on) and $dir = 1$ (so that M3 and M4 are on, M5 and M6 are off), c_0 is determined by the competition between M7 and M8 (i.e., x_0 and x_1). The two transistors are sized so that if either one of them is 2.5V and the other one is completely on, then the transistor, that is completely on, wins. In other words, c_0 is set so that the half-switch will transmit the data packet carried on the input channel, which is connected to the completely turned-on transistor. If both are on completely, then M8 wins over M7 (or M10 wins over M9 when $dir = 0$). The speed of this competition is greatly enhanced with the addition of the two inverters, M11-M14, at the output.

In addition to c_0 , there is a second control bit, c_1 , which tells the half-switch whether c_0 is a "don't care" or not [refer to Fig.2-9(b)]. If c_0 is a "don't care", then the output circuitry will produce a 2.5 V output in the next cycle (i.e., when $t(+1) = 0$). The truth table for c_1 is given in Fig. 2-9(c) as well. Notice that $c_1 = 1$ when c_0 is a X (i.e., "don't care"), and is 0 otherwise.

The implementation of this signal is somewhat complex. The signals, t and dir , have the same functionality for c_1 as they did for c_0 . When $t = 0$, M1 and M2 are on. Also, when $dir = 1$, M3 and M4 are on and M5 and M6 are off so that x_0 and x_1 determine c_1 . There are two sets of transistors, M11 and M12 for x_0 , and M7-M10 for x_1 . These transistors are sized so the if x_0 is 0 or if x_1 is 1, then c_1 is 0 since at least one of the inputs want to use that half-switch and thus, c_0 is not a "don't care". On the other hand, if $x_0 = 2.5$ V or 5 V and if $x_1 = 2.5$ V or 0 V, then c_1 is set to 1. The transistors M19-M24 are again added to improve the speed performance, as well as to take care of the necessary logic calculations.

The key in this implementation is that both x_0 and x_1 are directly fed into the competition transistors (i.e. M7-M12) without being inverted or buffered at any point before the competitions. The competition transistors refer to those that can provide a path to Vdd and Ground at the same time (i.e., they will try to pull the output up and down simultaneously) so that the outcome will depend on their input voltages. The reason is that the state of 2.5V is not stable at all once it is

produced at the previous stage's output circuit. If the input is 2.5 V into an inverter, a 0.1 V variation in the input level corresponds to almost a 1V variation at the output. As a result, the noise margins would be greatly reduced. However, with our implementation methods, a 1V noise margin is achieved for all cases. In other words, the three logic levels were 0-1 V, 1.5-3.5 V, and 4-5 V. If the output circuit was stabilizing the third logic level at anywhere between 1.5 V and 3.5 V, instead of exactly at 2.5 V, due to variations in the fabrication process, the right results would still be obtained. This choice of a 1 V noise margin is enough to cover the parameter variations, but still gives us a comfortable margin to distinguish the three levels from each other, as well as maintain the high speed of the network. If the margin was lower, the variations, that can change the speed, or the output current, of a transistor as much as 40%, could lead to a third level outside the expected and/or acceptable levels. On the other hand, if the margin was higher, then the transistors in competition would have to be sized closer to each other, and the net current that drives the load would be reduced, which in turn reduces the network's speed.

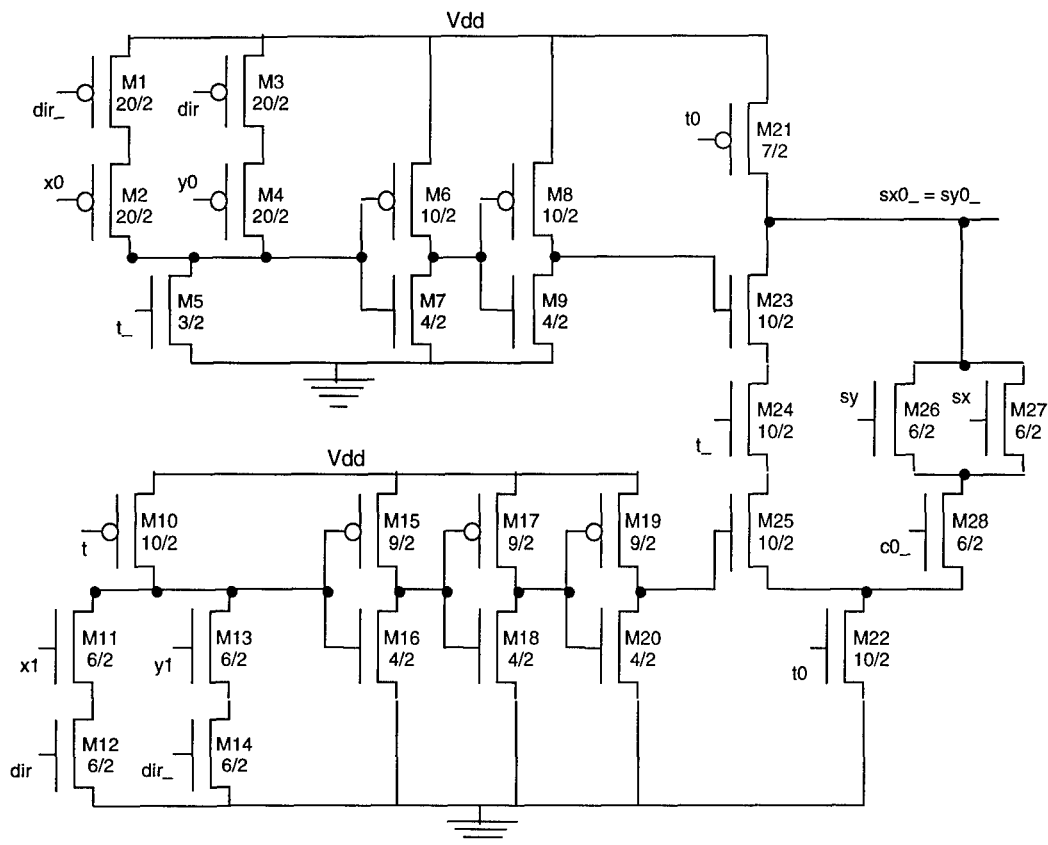
2.3.2.2.3 Contention Circuit

All the signals that relate to dropping-and-resending data packets are computed by the contention circuitry. All signals starting with the letter "s" are in this category (refer to Fig. 2-6). These signals are sx_0 , sx_1 , sx , sy_0 , sy_1 , and sy .

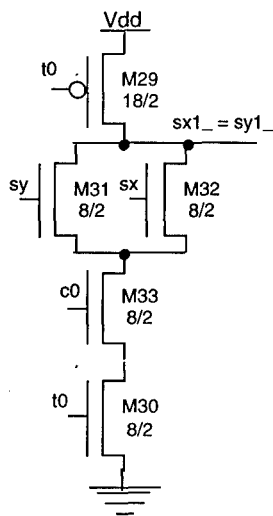
The two signals, sx and sy , are the final signals that are transmitted from one stage to the next to carry the information if a packet was dropped due to contention or not. If $dir = 1$, then the direction of data flow within a half-switch is from x to y . If a data packet is dropped, then $sx = 1$, and in this case, the contention signal flows from sy to sx . The reverse is true for sy when $dir = 0$.

The other four signals, sx_0 , sx_1 , sy_0 , and sy_1 , are intermediate signals, with sx_1 and sy_1 being transmitted between partner half-switches within the same stage. For a given direction, say $dir = 1$, only sx_0 and sx_1 are active, and the other two are floating. Within a half-switch, both sx_0 and sx_1 are computed. sx_1 is then sent to the partner half-switch, and the partner half-switch's sx_1 , which is called sx_1' , is received. Then, sx_0 and sx_1' are processed to find out what sx needs to be. If either one of them is a 1, then $sx = 1$, which means that the data packet, that used that specific half-switch, was dropped either at that stage or at some following stages before it was able to arrive at its final output destination. If it was dropped at a following stage, then that information would be carried to the half-switch through sy (i.e. for $dir = 1$, sy is an input).

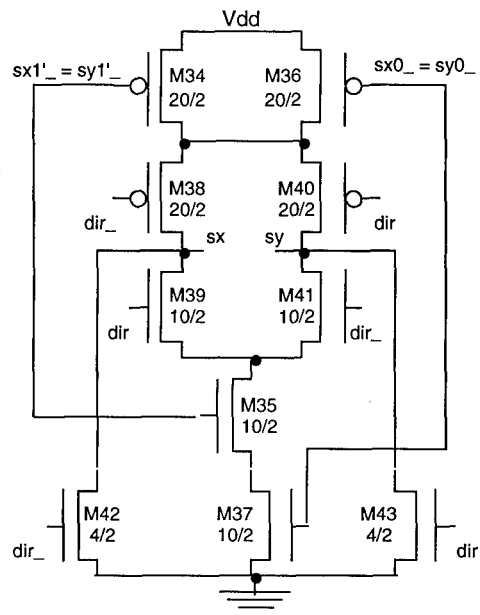
Figures 2-10(a) and 2-10(b) are the schematics for sx_0 and sx_1 , respectively (the signals' inverses are actually calculated to achieve the final necessary logic with as few transistors as possible). The signal t_0 is the transmission signal of the first stage, and it is used to reset all the contention circuits of all stages when a new data packet is introduced to the system. When $t_0 = 0$, the transistors, M21, M22, M29, and M30, ensure that $sx_0 = sx_1 = sx = 0$ (i.e. $sx_0 = sx_1 = 1$).



(a)



(b)



(c)

Figure 2-10. Contention circuits of design B. (a) Schematics of contention signals, $sx0_ = sy0_$. (b) Schematics of Contention signals, $sx1_ = sy1_$. (c) Schematics of contention signals, sx and sy .

Note that sx_0 and sy_0 cannot be on (i.e. arbitrarily chosen to be 1) at the same time, and similarly, sx_1 and sy_1 cannot be set to 1 at the same time. The reason is that the data should flow in one of two directions, and the contention signals will have to flow in the opposite direction. As a result, these two pairs of signals are treated as one (i.e. $sx_0 = sy_0$, $sx_1 = sy_1$), and then which direction they should flow is determined at the final stage when sx and sy are computed, with the help of the direction signal. This will allow us to reduce the necessary number of transistors, as well as to reduce the number of links between partner half-switches by one (i.e. $sx_1 = sy_1$, instead of two separate signals, sx_1 and sy_1).

The output of each half-switch goes to two partner half-switches on the next stage as input. As an example, please refer to Fig. 2-4. Number 8 on the left sends its output to number 8 and 12 in the middle as their x_0 and x_1 , respectively. If that data packet is dropped at some point in the switch, and it was using no. 8 in the middle, then sx_0 of no. 8 would be set to 1, which in return will set sx of no. 8 to 1. On the other hand, if the same packet was using no. 12 in the middle and was dropped, then sx_1 of no. 12 would be set to 1. This signal would be sent to no. 8 in the middle, which in return would set sx of no. 8 to 1. In either case, the intermediate signals (i.e. sx_0 of no. 8 and sx_1 of no. 12) determine whether sx of no. 8 is 1 or 0. In other words, they determine whether the primary input of no. 8 (i.e. x_0 of no. 8 in the middle, or y_0 of no. 8 on the left) was dropped or not.

Looking at Fig. 2-10(a), $sx_0_$ (or $sy_0_$) can be set to 0 in one of two ways (i.e. $sx_0 = sy_0 = 1$) after it is reset to 1 when t_0 turns 0 and back to 1. One possibility is that there is a contention in that very half-switch. For this to happen, assuming $dir = 1$, x_0 needs to be 0, which would turn on M23, and x_1 needs to be 1, which would turn on M25. At the same time, it is required that it is the half-switch's time to compute its control signals (i.e. $t = 0$, and M24 is on), so that the incoming x_0 and x_1 are address bits. In this case, x_0 would be dropped in a deterministic way, as explained before, and so $sx_0_$ would be set to 0 (i.e. $sx_0 = 1$).

The second possibility for sx_0 to be 1 is that if the data packet is dropped at some later stage (i.e. $sy = 1$), and x_0 was using that half-switch for transmission (i.e. $c_0 = 0$). In this case, M26 and M28 would be turned on, and $sx_0_$ would be 0 again.

On the other hand, when there is contention, always x_0 is dropped and x_1 is transmitted, so when sx_1 is calculated, the first possibility mentioned above does not exist for sx_1 [refer to Fig. 2-10(b)]. Then $sx_1_ = 0$ (i.e. $sx_1 = 1$) only when the data packet was dropped at a later stage (i.e. $sy = 1$), and the half-switch was transmitting its second input, x_1 , at that time (i.e. $c_0 = 1$). Then, M31 and M33 are on, and $sx_1_ = 0$. Of course, this is true when the half-switch is not resetting (i.e. when t_0 is not 0).

Figure 2-10(c) shows the final processing of $sx_0_$ and $sx_1'_$, or equivalently $sy_0_$ and $sy_1'_$, to compute sx and sy . Note that the $sx_1'_$ that is used to compute sx and sy comes from the

partner half-switch, where that half-switch's $sx1_$ is sent to its partner half-switch, so that its partner can compute its own sx and sy . If either $sx0_$ or $sx1'_$ is zero, then the output is set to 1. M38-M41 determine which output needs to be computed. M42 and M43 ensure that if the system is being operated in unidirectional mode, the disabled (i.e. the unused) output, which is sy if $dir = 1$, and is sx if $dir = 0$, does not float up to 5 V as time passes.

2.4 MODELING

The switch is implemented with the HP 0.8 μm CMOS technology where $\lambda = 0.5 \mu\text{m}$ but during fabrication, the minimum gate length is reduced from 1.0 μm to 0.8 μm . This process allows a faster maximum speed for the transistors, while not increasing the contact or metal resistance. The following analysis applies to the above mentioned CMOS process for various system sizes. It could be extended to other technologies by adjusting the transistors' input gate capacitance and the layout area. The design of a switch is evaluated in terms of its speed, area, total power consumption, and power density.

2.4.1 Area

First, the area of the switches will be discussed. The system is implemented using the CMOS-SEED [i] technology, so the pads are located above the CMOS circuits, and thus, they are not included in the area analysis. These pads make contact to the transmitters and receivers through the third metal level (i.e. metal3 layer). The area of a switch is determined by the number of half-switches per channel (i.e. $\log_2 K$, where $K = \sqrt{N}$ with N being the total number of channels of the system), the size of a single half-switch, the area for routing the wires, and the area for the transmitters and receivers. The area of each channel is:

$$A(\text{channel}) = \log_2 K * [A(\text{half-switch}) + \text{routing}] + 2 * [A(\text{trans.}) + A(\text{rec.})] \quad (2-1)$$

The area of a half-switch is about $230 \mu\text{m} \times 110 \mu\text{m}$. For each stage, the routing requires 6 horizontal and 6 vertical wires with 2 μm thickness and 2 μm spacing (i.e. $x0, x1, y0, y1, sx1_ , sx1'_$) between each pair of half-switches for a total of 24 μm in each direction. Then the expression in the first parentheses becomes:

$$A(\text{half-switch}) + \text{routing} = [(230 + 24) * (110 + 24)] \mu\text{m}^2 \quad (2-2)$$

For each channel, 2 transmitters and 2 receivers are used, with their sizes equal to approximately $50 \mu\text{m} \times 50 \mu\text{m}$, giving us an area of:

$$2 * [A(\text{trans.}) + A(\text{rec.})] = 4 * (50*50) \mu\text{m}^2 = 10^4 \mu\text{m}^2 \quad (2-3)$$

Substituting Eqs. (2-2) and (2-3) into (2-1) :

$$A(\text{channel}) = [\log_2 K * [(230 + 24) * (110 + 24)] + 10^4] \mu\text{m}^2 \quad (2-4)$$

Once the total area of a channel is known, the minimum required constant pitch between adjacent channels can be calculated, as well as the total area for a switch plane:

$$\Delta = \text{constant pitch} = \sqrt{A(\text{channel})} \quad (2-5)$$

$$A(\text{switch-plane}) = N * A(\text{channel}) \quad (2-6)$$

Thus, for a system, that is partitioned into K switches with K channels for a total of N channels, the length of a side of the complete electronic plane is:

$$L_{\text{plane}} = \Delta * K \quad (2-7)$$

Table 2-1 tabulates parameters, K, A(channel), Δ , A(switch-plane), and L_{plane} , for cases of N = 256, and N = 4096.

TABLE 2-1. Calculations for Speed and Area.

	N = 256	N = 4096
K	16	64
A(channel)	$1.46 * 10^5 \mu\text{m}^2$	$2.14 * 10^5 \mu\text{m}^2$
D	382 mm	463 mm
A(switch-plane)	0.37 cm ²	8.77 cm ²
L_{plane}	0.61 cm	2.96 cm
L_{max}	917 μm	2222 μm
$T_{90\%}$	$3.37 * 10^{10}$ s	$4.38 * 10^5$ s
Freq. max (RC)	2.97 Gb/s	2.29 Gb/s
Freq. max (h-s) (simulated)	250 Mb/s	250 Mb/s

2.4.2 Speed

The second variable is the speed of the system. Since the switch is pipelined by delaying the external signals the same amount of time it takes for the initial input to reach a given stage of half-switches, the overall system speed is equal to the speed of a single half-switch. From the simulations, this is 250 Mb/s. As the technology goes to smaller feature sizes, this speed will increase up to the point where the speed is limited by the RC time constant of the maximum length of wire, existing inside the switch.

For a system with N channels, each switch has $K = \sqrt{N}$ channels, distributed in a $\sqrt{K} \times \sqrt{K}$ array configuration. If Δ is the constant pitch between adjacent channels (Eq. (2-5)), then the maximum wire length in a switch is:

$$L_{\max} = \Delta * (\sqrt{K}/2), \text{ (routing not included)} \quad (2-8)$$

which is equal to half the length of the side of a switch. Because of possible extra routing that may be needed, another 20% is added to get:

$$L_{\max} = 1.2 * \Delta * (\sqrt{K}/2) = 0.6 * \Delta * \sqrt{K}, \text{ (routing included)} \quad (2-9)$$

To estimate the limiting frequency, the 0-90% rise/fall time of the circuit is calculated, where a driver transistor is driving the maximum length wire with a capacitive load at the end. The capacitive load is the input capacitance of all the transistors connected to that wire at the next stage. For a distributed RC load, it is weighed with 1.0, and for a lumped one, it is weighed with 2.3 so the resulting time delay is given by [2-5]:

$$T_{90\%} = 1.0 R_{\text{int}} C_{\text{int}} + 2.3 (R_{\text{tr}} C_{\text{int}} + R_{\text{tr}} C_{\text{load}} + R_{\text{int}} C_{\text{load}}) \quad (2-10)$$

where R_{int} and C_{int} are the resistance and capacitance of the wire (i.e. of the interconnect), R_{tr} is the on-resistance of the driving transistor, and C_{load} is the total load capacitance due to transistors' input gate capacitance of the next stage. R_{int} and C_{int} are equal to $(L_{\max} * R_{\text{wire}})$ and $(L_{\max} * C_{\text{wire}})$, respectively, where R_{wire} and C_{wire} are fabrication dependent parameters, and are given per unit length. For the 0.8 μm process with $\lambda = 0.5 \mu\text{m}$, and for a 2.0 μm thick wire (i.e. width of the wire = 4λ) in metal2, $R_{\text{wire}} = 0.03 \Omega/\mu\text{m}$, and $C_{\text{wire}} = 44 \text{ aF}/\mu\text{m}$.

In [2-5], it is shown that the on resistance of a MOS transistor can be approximated as: $R_{\text{tr}} = (L/W) / [\mu C_{\text{gox}}(V_{\text{dd}} - V_{\text{t}})]$. Since the driver PMOS and NMOS transistors are sized so that their output currents would be equal (i.e. the rise and fall times would be equal), they have the same on resistance. For the specific chosen run the on-resistance turns out to be $R_{\text{tr}} = 585 \Omega$.

To calculate C_{load} , note that each driver drives all the transistors labeled x0 in one half-switch on the next stage, and the ones that are labeled x1 in its partner half-switch on the next stage. In addition, it drives all the transistors with y0 as input within its own half-switch, and the ones that have y1 as input in its own partner half-switch on the same stage. The sum of all the input capacitance of all these transistors inside all the circuits (i.e. output, control, and contention circuits) give a total capacitance of 200 fF, so $C_{\text{load}} = 200 \text{ fF}$. Substituting the values for R_{wire} , C_{wire} , R_{tr} , and C_{load} into Eq. (2-10) :

$$T_{90\%} = [1.32 * 10^{-18} * (L_{\max})^2] + 2.3 [(2.57 * 10^{-14} * L_{\max}) + (1.17 * 10^{-10}) + (6 * 10^{-15} * L_{\max})] \quad (2-11)$$

where L_{\max} is given in μm , and the resulting $T_{90\%}$ is in seconds. Since $T_{90\%}$ is the rise/fall time of the maximum length wire, the maximum frequency in Mb/s (note that maximum frequency is not in MHz) will be:

$$\text{Freq}_{\max} (\text{RC}) = (T_{90\%})^{-1} \quad (2-12)$$

The above maximum frequency is labeled as RC because this is the frequency limited by the RC time constant of the longest wire. However, the maximum frequency of the circuits, from the simulations, is 250 Mb/s. Therefore, $\text{Freq}_{\max} (\text{RC})$ only sets an upper limit to the actual maximum frequency, called $\text{Freq}_{\max} (\text{h-s})$ (i.e. due to the speed of the half-switch's circuits). Since the switch is scaleable, $\text{Freq}_{\max} (\text{h-s})$ stays constant at 250 Mb/s as N grows. From Table 2-1, $\text{Freq}_{\max} (\text{h-s})$ still has a long way to go before $\text{Freq}_{\max} (\text{RC})$ becomes a limiting factor. L_{\max} , $T_{90\%}$, $\text{Freq}_{\max} (\text{RC})$, and $\text{Freq}_{\max} (\text{h-s})$ are also given in Table 2-1.

2.4.3 Power Consumption

In the power consumption calculations, the consumption of the wires inside the half-switches is neglected since these wires are quite short. Only the transistors in the half-switches, and the long wires between stages and between partner half-switches are taken into account. For all cases, the power consumption is given as :

$$P = C * V^2 * (f/2) * p , \quad (2-13)$$

where C is the capacitive load (i.e. input gate capacitance for a transistor, or the substrate to metal capacitance for a metal wire), V is the voltage swing (i.e. either 2.5V or 5V, depending on whether the third logic level is used or not, respectively), f is the operating frequency in Mb/s (i.e. divide by 2 to convert from Mb/s to MHz), and p is the probability of the previous situation changing (i.e., data switching, or the information on the wire switching).

First, the probability equations are examined. For a packet-switching network made up of $k \times k$ interchange boxes, if the initial probability for data to appear at the first stage is p_0 , then the probability of data arriving at the end of i^{th} stage can be approximated as:

$$P_i = \frac{1}{\frac{(k-1) \cdot i}{2k} + \frac{1}{p_0}} \quad (2-14)$$

but in our case, $p_0 = 1$ (i.e. every input buffer gets a data packet so the calculations are for a fully loaded network), and $k = 2$ (i.e. the switch is made up of 2×2 bypass-and-exchange switches partitioned into two half-switches), so Eq. (2-14) simplifies to:

$$p_i = \frac{4}{4+i} \quad (2-15)$$

The values of p_i for $i = 0, \dots, 12$ are given in Appendix A.

In addition, let $S = \log_2 N$ be the number of stages in the whole network, and note that there are only $S/2$ stages of half-switches in one switch plane, and the whole network consists of two switch planes. When the power consumption per channel is calculated, the calculations are done for both switch planes together, and the total power consumption is the sum of all the switches on both the switch planes.

The total power consumption per channel is grouped into three parts, namely, the half-switches, the output wires, and the contention wires. The half-switches consist of the transistors, since the short wires within the half-switches are neglected. To simplify the expressions, the input gate capacitance of an NMOS transistor with a given width, W , in terms of λ will be written as $NM(W)$, and similarly for a PMOS, the gate capacitance will be given as $PM(W)$. The gate length is not considered since the length of all the transistors are equal to 2λ , which is equal to $0.8 \mu\text{m}$ after the fabrication. As an example, the capacitance of an NMOS with width = $10 \mu\text{m} = 20\lambda$ is $NM(20)$.

The fact that the output signals, $x_0, x_1, y_0,$ and y_1 , will switch or not, does not depend on the direction of data flow, but on the probability of data arriving at that stage. Then, all the transistors in the three circuits of a half-switch are counted, and the total capacitance for these signals in one half-switch is found. The equations for the transistor capacitance (i.e. for the specific chosen run of the $0.8\mu\text{m}$ technology) are $NM(W) = (825 * W) \text{ aF}$, and $PM(W) = [(1160 * W) + 611] \text{ aF}$. The inverters that get switched directly because of these four signals are included in the total capacitance as well:

$$\text{Cap.}(x_0, x_1, y_0, y_1) = 6NM(4) + 2NM(5) + 2NM(6) + 4NM(8) + 6PM(10) + 2PM(18) + 4PM(20) + 2PM(33) = 354 \text{ fF}$$

If a data packet arrives at a stage, then these signals have a voltage swing of $V_{dd} = 5V$, and the probability of the signal changing is 0.5. If not, then the previous stage would output $2.5V$ due to contention for one cycle out of "d" cycles. Each data packet is preceded by a header of $\log_2 N$ bits of address so for a packet with 16 bits of information, $d = 16 + \log_2 N$. Then the power consumption per channel due to $x_0, x_1, y_0,$ and y_1 is:

$$P_{\text{channel}}(x_0, x_1, y_0, y_1) =$$

$$\sum_{i=0}^{S-1} \left[\left(\frac{1}{2} p_i (V_{dd})^2 + \frac{1}{d} (1 - p_i) \left(\frac{V_{dd}}{2} \right)^2 \right) \cdot \text{Cap}(x_0, x_1, y_0, y_1) \cdot \frac{f}{2} + \left(\frac{1}{d} \cdot (1 - p_i) \cdot i_{\text{sat}} \cdot V_{dd} \right) \right] \quad (2-16)$$

where the first term in the first parentheses is for an active channel, which is transmitting data. The 1/2 term is due to the fact that data has a 1/2 probability of switching from its previous value. The second term in the first parentheses is for a disabled channel. Then it will output 2.5V (i.e. $V_{dd}/2$) for one cycle out of d cycles (i.e. $1/d$), and the probability for this happening is $1 - p_i$, that is, when there is no input for that half-switch. When this happens, there is a direct path from V_{dd} to Ground, and the last term in Eq. (2-16) accounts for the saturation current flowing through the driving transistors. In the calculations, $i_{\text{sat}} = 3.6$ mA. The results of Eq. (2-16) are calculated in Table 2-2 for $N = 256$, and $N = 4096$.

TABLE 2-2. Calculations for Power Consumption and Density of the Electronic Switches (Excluding the Optical Transmitter/Receiver Circuits and Assuming $f = 100$ Mb/s).

	N = 256	N = 4096
K	16	64
S (# stages)	8	12
d (for 16-bit data)	24	28
$P_{\text{channel}}(x_0, x_1, y_0, y_1)$	3.51 mW	5.23 mW
$P_{\text{channel}}(c_0, c_1)$	0.05 mW	0.10 mW
$P_{\text{channel}}(t, \text{dir})$	0.31 mW	0.41 mW
$P_{\text{channel}}(\text{contention}, s?)$	0.03 mW	0.03 mW
$P_{\text{channel}}(\text{output wires})$	0.13 mW	0.33 mW
$P_{\text{channel}}(\text{contention wires})$	0.9 mW	2.4 mW
P_{channel}	4.03 mW	6.10 mW
P_{total}	1.03 W	25.0 W
P_{density}	1.38 W/cm ²	1.42 W/cm ²

Similarly, the gate capacitance of the transistors that have c_0 and c_1 as input are:

$$\text{Cap.}(c_0) = 2\text{NM}(4) + 1\text{NM}(6) + 1\text{NM}(8) + 2\text{NM}(10) + 2\text{PM}(10) + 2\text{PM}(25) = 118 \text{ fF}$$

$$\text{Cap.}(c_1) = 1\text{NM}(5) + 2\text{NM}(8) + 5\text{NM}(10) + 1\text{PM}(13) + 2\text{PM}(20) + 1\text{PM}(25) = 152 \text{ fF}$$

The power consumption per channel due to these two signals is:

$$P_{\text{channel}}(c0, c1) = \sum_{i=0}^{S-1} \left[\text{Cap.}(c0) \cdot \frac{1}{2} + \text{Cap.}(c1) \cdot p_{i+1} \cdot (1 - p_{i+1}) \cdot 2 \right] \cdot \frac{1}{d} \cdot (V_{dd})^2 \cdot \frac{f}{2} \quad (2-17)$$

The first term inside the bracket says that $c0$ has a $1/2$ probability of switching at every new data packet. The second term says that $c1$ will switch if there was an input in the previous cycle, and there is not one in the following cycle (i.e. $p_{i+1} * (1 - p_{i+1})$), or there was not one in the previous cycle, and there is one now (i.e. multiply that term by 2). The $1/d$ factor is for the fact that this switching occurs at every $1/d$ bits (i.e. once per data packet). Again, the results are in Table 2-2.

The capacitance due to the external signals, t and dir , is:

$$\text{Cap.}(t) = 1\text{NM}(3) + 1\text{NM}(8) + 6\text{NM}(10) + 2\text{NM}(21) + 1\text{PM}(7) + 1\text{PM}(10) + 1\text{PM}(18) + 2\text{PM}(62) = 281 \text{ fF},$$

$$\text{Cap.}(dir) = 2\text{NM}(8) + 2\text{NM}(4) + 2\text{NM}(6) + 2\text{NM}(10) + 4\text{PM}(20) + 2\text{PM}(33) + 2\text{PM}(62) = 399 \text{ fF}.$$

The power consumption due to t and dir is:

$$P_{\text{channel}}(t, dir) = \left(2 \cdot \text{Cap.}(t) + \frac{1}{2} \cdot \text{Cap.}(dir) \right) \cdot \frac{1}{d} \cdot S \cdot (V_{dd})^2 \cdot \frac{f}{2}, \quad (2-18)$$

$\text{Cap.}(t)$ is multiplied by 2 since t gets reset to 0 for one cycle and then equals 1 for the remaining $d-1$ cycles (i.e., switches twice every d bits). On the other hand, the direction bit has a $1/2$ probability of switching at every data packet, assuming a bi-directional switch operation. In addition, it is divided by d because these signals switch once at every data packet, and is multiplied by S (i.e., the number of stages) because these signals do not depend on which stage they are on but only on the number of stages. Whether there is an input arriving at a given stage or not, does not affect these signals, or their power consumption.

The last term for the half-switches is from the contention signals. The individual capacitance for the various signals are:

$$\text{Cap.}(sx0_, sy0_) = 5\text{NM}(4) + 1\text{NM}(6) + 2\text{NM}(10) + 3\text{PM}(9) + 2\text{PM}(10) = 96 \text{ fF},$$

$$\text{Cap.}(sx1_, sy1_) = 1\text{NM}(8) = 7 \text{ fF},$$

$$\text{Cap.}(sx, sy) = 2\text{NM}(10) + 2\text{PM}(20) = 64 \text{ fF}.$$

If there is a contention, then sx or sy switches, but only one of the other two pairs will switch with it. In other words, the contention comes from one of two channels, so the effective total capacitance is obtained by adding $\text{Cap.}(sx, sy)$ with $1/2$ of the other two capacitance (i.e. there is a $1/2$ probability for either of the channels). Then, the resulting power consumption is:

$$P_{\text{channel}}(sx, sy, sx0, sx1, sy0, sy1) =$$

$$\sum_{i=0}^{S-1} \left(Cap.(sx, sy) + \frac{1}{2} Cap.(sx0, sy0) + \frac{1}{2} Cap.(sx1, sy1) \right) \cdot (p_i - p_S) \cdot 2 \cdot \frac{1}{d} \cdot (V_{dd})^2 \cdot \frac{f}{2} \quad (2-19)$$

The total power per channel from the half-switches is simply equal to the sum of Eqs. (2-16)-(2-19), and is given in Table 2-2 for N = 256, and N = 4096.

In addition to the half-switches, the power is consumed by the long wires between stages and between partner half-switches. These wires can be categorized into two groups, output wires, and contention wires. The length of wires at each stage is not a constant. Looking at the 2-D layout (refer to Fig. 2-5), a formula is derived that will give us the length of wire between partner half-switches at each stage in terms of Δ , the constant pitch. Then, the length of wire at stage i , would be $f_i \cdot \Delta$, where f_i is the multiplicative factor, and is given as:

$$f_i = 2 \left(\frac{S-1-Mod(S/4)}{4} \left[Int \left(\frac{1}{2} Mod_S(i-1) \right) \right] \right) \quad (2-20)$$

f_i is given for $i = 0, \dots, S-1$ in Appendix A, for $S = 8$, and $S = 12$ (i.e. N = 256, and N = 4096, respectively).

From the fabrication parameters, the capacitance of metal2 wires, $Cap.(wires)$, is 44 aF/ μm , where $4\lambda = 2 \mu m$ wide metal2 wires are assumed to be used for all the long wires between partner half-switches. Then, the power consumption for the wires is given as:

$$P_{channel}(\text{output wires}) = \Delta \cdot \sum_{i=0}^{S-1} (f_i + f_{i+1}) \left[\frac{1}{2} p_{i+1} (V_{dd})^2 + \frac{1}{d} (1 - p_{i+1}) \left(\frac{V_{dd}}{2} \right)^2 \right] \cdot \frac{f}{2} \cdot Cap.(wire) \quad (2-21)$$

The whole summation is pre-multiplied by Δ since f_i is the length of wires in terms of Δ . The first parentheses tells us that the output of a half-switch is connected to y_0 and y_1 of half-switches at the same stage (i.e. f_i), and x_0 and x_1 of half-switches at the next stage (i.e. f_{i+1}) assuming $dir = 1$. The next bracket covers the two possibilities that the output level on the wire can either be 0 or 1 for d cycles, or it can be 2.5V for one cycle and 0 for the rest of the $d-1$ cycles.

Similarly, the capacitance for the contention wires is the same, and the power consumption is given as:

$$P_{channel}(\text{contention wires}) = \Delta \cdot \sum_{i=0}^{S-1} \frac{1}{d} \cdot \frac{1}{2} \cdot (p_{i+1} - p_S) \cdot f_i \cdot (V_{dd})^2 \cdot \frac{f}{2} \cdot Cap.(wire) \quad (2-22)$$

Again, the summation is pre-multiplied by Δ . In addition, there is a $1/d$ term because the contention circuitry switches once at every data packet. The $1/2$ term is due to the fact that only $sx1$ signal drives a long wire between partner half-switches, and $sx0$ is an internal signal to the half-switch. As a result, only if $c0 = 1$, the long wire switches. The $(p_{i+1} - P_s)$ term is the probability that there will be a contention signal propagating. This signal is generated only if the data packet is dropped before it reaches its final destination but after it goes through that given stage of half-switches.

Summing Eqs. (2-16)-(2-22), gives the total power consumption per channel. Looking at the power consumption for the different components in Table 2-2, the output signals dominate the overall power consumption. This is due to the fact that the consumption during when there is a direct path from Vdd to Ground (i.e. output is set at 2.5V) is included in this term. In addition, these transistors turn on and off at every bit, compared to the others that get switched every d cycles.

The total power consumption for the whole system (only the electronic power) is:

$$P_{total} = P_{channel} * N \quad (2-23)$$

The power density is given as:

$$P_{density} = \frac{1}{2} * \frac{P_{channel}}{\Delta^2} \quad (2-24)$$

In Eq. (24), Δ^2 is the area per channel per switch plane, and the $1/2$ term is added in because each channel occupies that much area on each switch plane. Effectively, the total area per channel is $(2 * \Delta^2)$. From Table 2-2, the electronic chip will have no trouble handling the heat dissipation in these circuits, despite the fact that there is a direct path from Vdd to GND at certain times.

2.5 DISCUSSION

We have optically tested the functions of the OTIS electronic switches. As reported previously, the switches are electrically operational. In this experiment (Fig. 2-11) we connected one modulator and one detector chip via a reflective system. Both the modulator and detector have three-level logic receiver and driver circuits. It has been shown that a three level transmission can be performed. This matches the three level logic switches implemented on the chip where the third logic level provides contention arbitration in the switch.

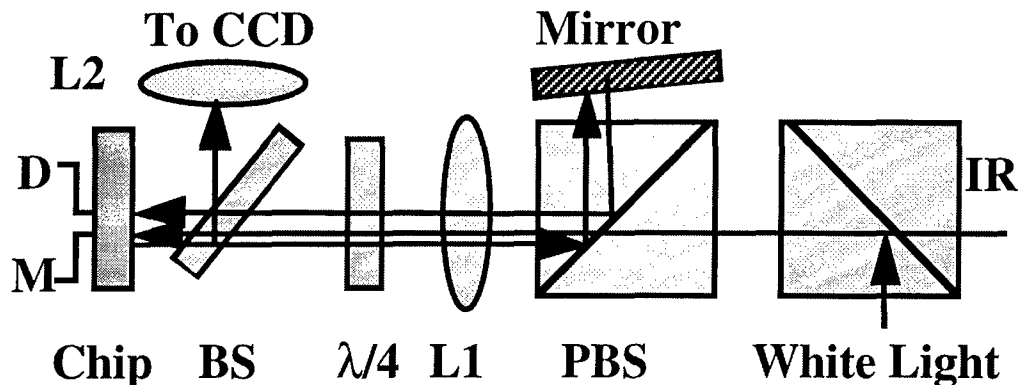


Figure 2-11. Schematics of the set-up used to evaluate the optoelectronic OTIS chip and the three-level logic transmission.

Looking at the results from Tables 2-1 and 2-2, a network with 4096 channels seems feasible. The side length of a switch plane for a 4096 channel network is found to be approximately 3 cm. Note that a network with N channels is constructed with $\sqrt{N} = K$ switches, and that these switches do not share any internal wires but only the external signals of direction and transmission. As a result, one can easily implement a network of this size by tiling together small switch chips, thus reducing potential fabrication problems on the electronic chips. Such a system has a $\sqrt{K} \times \sqrt{K} = 8 \times 8$ array of switches per switch plane, with each switch handling 64 channels. Thus, each switch only takes up a chip of side length $3/8 \text{ cm} = 3.75 \text{ mm}$, which is readily available in $0.8 \mu\text{m}$ CMOS technology at commercial silicon foundries.

A maximum speed of 250 Mb/s has been simulated for the electronic switches and knowing that the probability of acceptance of a 4096 channel system is $1/4$ (refer to Appendix A with $i = \log_2 N = 12$ stages), this yields a total throughput of $(250 \text{ Mb/s}) \times (1/4) \times (4096) = 256 \text{ Gb/s}$. In a related work,⁽²⁻⁶⁾ the modeling presented in section 4 of this report has been extended to include all the various components of the OTIS system including the free-space optical system required to implement the network. For the network size and speed mentioned above, the power consumption of the entire optoelectronic system is found to be approximately 90 W. This yields 22 mW per channel for the entire optoelectronic system, which is very competitive with available networks. The power consumption of the switches is modeled to be about 40 W, while the optical power requirement is only 1 W, which is readily available with solid-state lasers. The rest of the power is consumed by the transmitter circuits (2 W) and mostly by the receiver circuits (50 W). It is interesting to note that most of the power consumption comes from the receiver circuits and mainly from their DC power consumption, which is why these circuits are now receiving a lot of attention in the community in order to improve their performance.

The on-chip power density at 250 Mb/s is calculated to be 5 W/cm². Electronic chips do not require any special heat dissipating schemes until the power density approaches 10 W/cm², so there should not be any thermal dissipation problems for a 4096 channel network.

As smaller CMOS technologies are employed, the speed of the switches will increase. Adding to this the possibility of increasing the network size in terms of number of channels, one could reach a throughput in the Terabit regime in the near future. Note that as the system size is increased, the system speed will not be reduced due to the pipeline structure of the switches, and one can still maintain a relatively high yield on the electronic chips due to the independence of switches from one another.

Finally, an optoelectronic switch chip is now being built based on the AT&T flip-chip bonded CMOS-SEED technology that combines 0.8 μm CMOS chips (as modeled in the previous section) with GaAs MQW optical transmitters and receivers where operation of the receivers and transmitters at over 600 Mb/s⁽²⁻⁷⁾ has been demonstrated.

Section 3

OPTICAL INTERCONNECTION UNIT

The optical interconnection unit (OIU) allows a parallel optical interconnect to another EOI unit, thus enabling global, parallel interconnection between two electronic processor arrays. The optical system includes the actual interconnection lenslets, based on the Optical Transpose Interconnection System (OTIS), and the necessary elements required to power-up the modulators. We presented several designs for a 256 channel bi-directional free-space optical interconnection system; the system accomplishes a global transpose interconnection with only 2 planes of lenslet arrays. The overall system length could be less than 60 mm. Here the interconnection lenslets must furnish the following requirements: high light efficiency and bi-directionality. Systems composed of refractive, diffractive, and aspheric diffractive elements were designed and optimized. We also introduced a novel modulator illumination system consisting of an off-axis area-multiplexed lenslet array which can be combined, via using Birefringent Computer Generated Hologram (BCGH) technology, into the same optical element as the interconnect optics. To further improve the system performance, a Photorefractive Beam Splitter (PRBS) has been studied in an attempt to replace the Polarizing Beam Splitter (PBS) originally designed in the system. This PRBS utilizes Bragg selectivity inherent to volume holograms to efficiently differentiate optical paths between optical-power and data-interconnection.

In this section we discuss the optical design and optimization of the optical transpose interconnection system (OTIS). We have used Code V[®] optical system design software package⁽³⁻¹⁾ to design and optimize several different systems based on both refractive and diffractive micro-optic technologies. An overview of OTIS optical system is first given, including a discussion of the system symmetry and an illumination system which improves optical power efficiency. The next section describes the design and optimization procedures that we have implemented with Code V[®] and results are then given.

3.1 OVERVIEW OF OTIS OPTICAL SYSTEM

3.1.1 Geometry

The OTIS is a simple means of providing a transpose interconnection using only a pair of lenslet arrays. This system has been shown useful for implementing shuffle based multi-stage interconnection networks and mesh-of-trees matrix processors.^(3-2,3-3) The transpose interconnection is a one-to-one interconnection between L transmitters and L receivers, where L is the product of two integers, M and N . An $M \times N$ transpose is equivalent to a k -shuffle,⁽³⁻⁴⁾ where k equals N . To implement the interconnection a $\sqrt{N} \times \sqrt{N}$ array of lenslets is placed in front of the

input (source) plane, and an $\sqrt{M} \times \sqrt{M}$ array of lenslets is located before the output (detector) plane.

An interesting application occurs when $k = \sqrt{L}$ (i.e. $M = N$). The number of stages required for routing any arbitrary permutation through OTIS is $\log_k(L)$. In this case this number becomes a constant since $\log_{\sqrt{L}}(L) = 2$. This results in a symmetric transpose where both planes of lenslets are identical. Thus, such a system achieves full routing between the input and output planes with a minimal amount of hardware. For example, a 4096 channel ($M = N = 64$) interconnection can be implemented with two 8×8 lenslet arrays. Figure 3-1 shows a two-dimensional cross-section, and experimental input and output patterns for such a system.

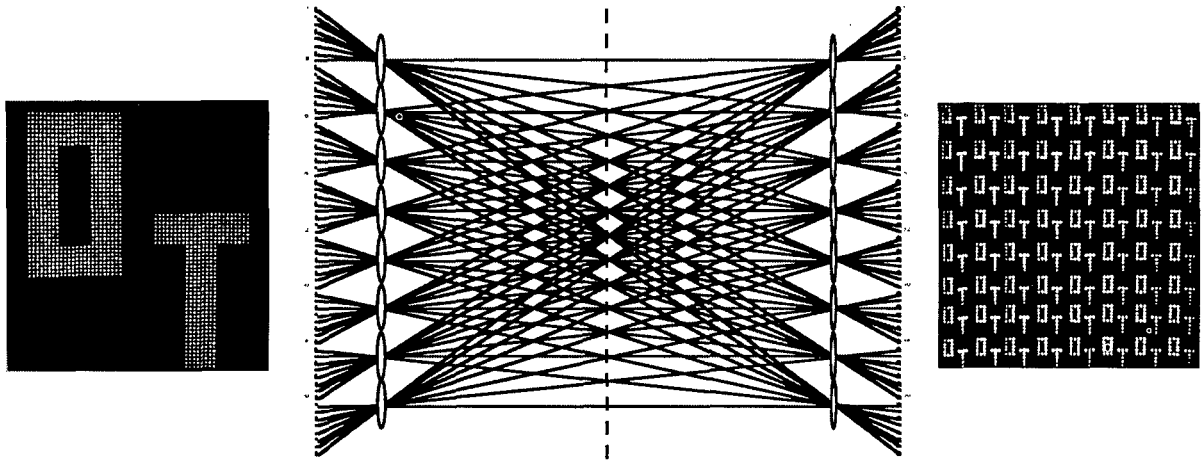


Figure 3-1. 2-D cross-section of a symmetrical 4096 channel OTIS with photos of experimental input and output illustrating transpose operation.

A symmetrical OTIS can easily be made bi-directional to accommodate systems where optoelectronic chips located on both planes have a transmitter/receiver pair per channel. As shown in Figure 3-2, the two planes are rotated 180° with respect to each other, so that a transmitter in plane 1 faces a receiver in plane 2. Unfortunately, this system configuration results in crosstalk: the light strikes the second lenslet plane off-center, it will partially fill an adjacent lenslet and part of energy will be focused to a different receiver. This optical cross effect is illustrated in Figure 3-2. There are two approaches to avoid this problem. The first solution is to place opaque areas on the edge of each lenslet that block the light going to the wrong receiver. However, this would result in lowered overall system efficiency. The second approach is to tailor the illumination system so only the proper portion of the lenslet will be illuminated.

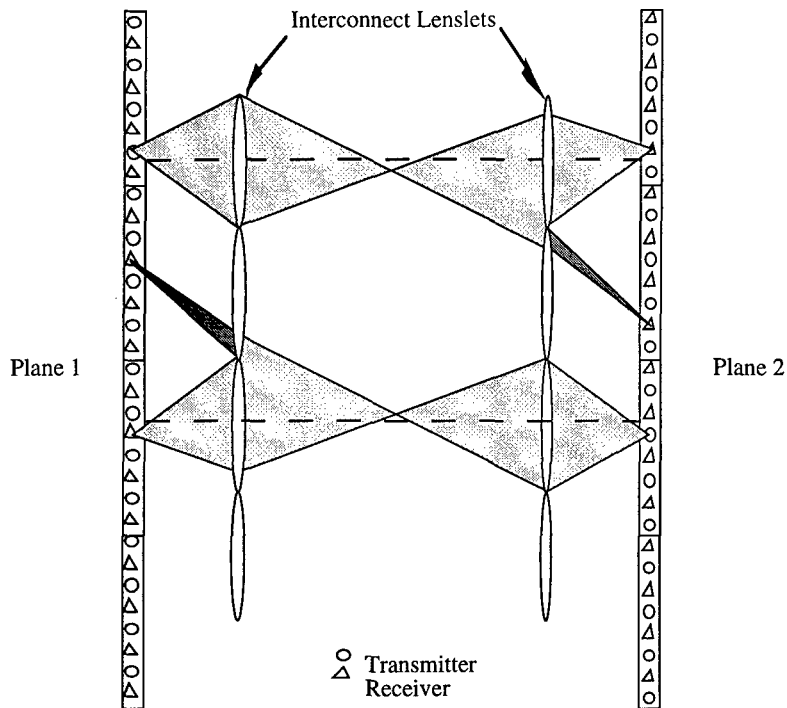


Figure 3-2. Bi-directional system showing transmitter/receiver offset and resulting crosstalk (dark gray).

3.1.2 Symmetry

The symmetry of OTIS significantly limits the number of lenslets within an array which perform unique functions. For example, as can be seen in Figure 3-1, all corner lenslets are functionally equivalent. Examination of the symmetry in OTIS reveals that the number of unique lens functions is given by:

$$\sum_{i=1}^{\sqrt{M}/2} i + \sum_{j=1}^{\sqrt{N}/2} j \quad (3-1)$$

This number is further reduced for a symmetric transpose ($M = N$):

$$\sum_{i=1}^{\sqrt{M}/2} i \quad (3-2)$$

For example, a 256 channel symmetric transpose system ($M = N = 16$) has only three unique lens functions, while a 4096 channel system ($M = N = 64$) would have ten. Figure 3-3 shows a 4×4 lenslet plane for a 256 channel system with functionally equivalent lenses shaded similarly. As will be discussed further, without careful examination of the symmetry, the system would be too complex to be efficiently modeled by current optical design software.

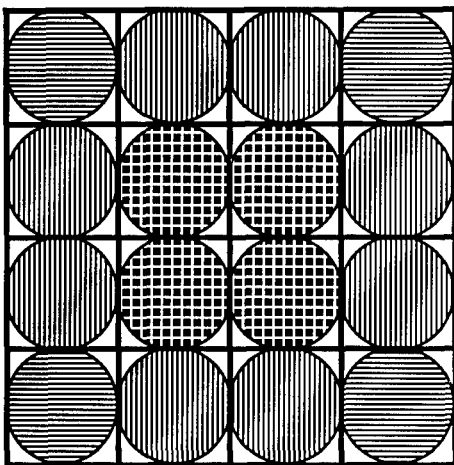


Figure 3-3. Lenslet symmetry in OTIS; equivalent lens functions are shaded similarly.

3.1.3 Illumination

In the case where the optical transmitters are modulators, we studied different illumination methods. To maximize light coupling into the interconnect optics, the modulators should be illuminated off normal. The most promising approach to achieve this uses an area-multiplexed off-axis diffractive lenslet array. Such a system provides the necessary directed illumination. Figure 3-4 shows a cross-section of the illumination system for a 256 channel OTIS. The upper detailed figure shows the off-axis illumination with two overlapped illumination lenslets. The overlap between the two lenslets is necessary for a normal plane wave is used in the system. For reflection-mode modulators the system must be 'folded' onto itself; the lower figure inset shows the same two interconnection lenslet in the same plane as the interconnection lenslets.

Both the illumination and interconnection optics may be combined into the same optical element using the technology of birefringent computer generated holography (BCGH).⁽³⁻⁵⁾ A BCGH is a computer generated hologram (CGH) with two different phase functions, one for each state of linear polarization. If multiple quantum well (MQW) electro-absorptive modulators are used, then a quarter-wave retardation plate should be inserted in between the BCGH and the lenslet plane. No waveplate is needed if Lead Lanthanum Zirconate Titanate (PLZT) electro-optic modulators are used, as these modulators provide the necessary polarization rotation. With such an illumination system the off-axis performance of the modulators may be an issue. MQW's have been shown to have a suitably wide (10°-15°) angular acceptance range.⁽³⁻⁶⁾ We are presently evaluating the off-axis performance of PLZT modulators; preliminary results show acceptable performance $\pm 15^\circ$ from the normal. If emitters (such as VCSEL arrays) are used rather than modulators, a surface mounted micro-prism or diffraction grating would be needed to deflect the illumination light, providing the required directivity to effectively couple light into the interconnect lenslets.

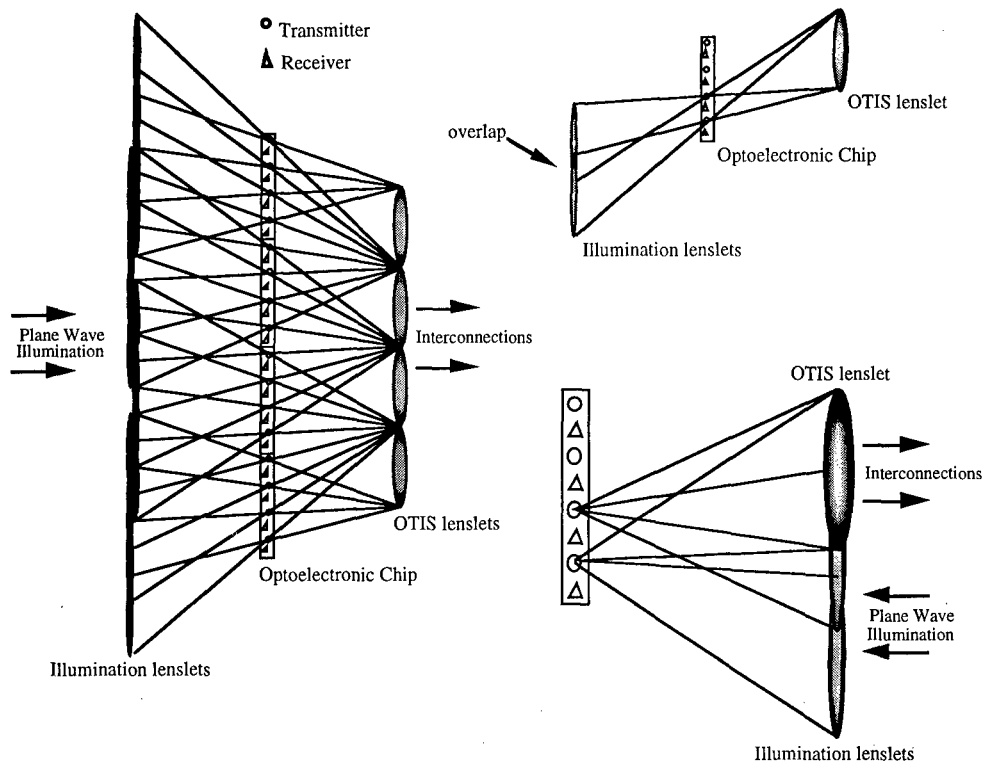


Figure 3-4. Off-axis area-multiplexed illumination system with details of area overlap (multiplexing) and folded geometry.

3.2 DESIGN AND OPTIMIZATION

We have modeled, using Code V® software, various 256 channel ($M = N = 16$) OTIS systems. As mentioned before the OTIS has a great deal of inherent symmetry; and this allows us to fully model the system with far fewer field points and surfaces than would otherwise be required. Without this reduction in scale, the computational task would present a nearly intractable problem. First order geometrical approximations determine the initial design of each system, with given fixed parameters such as $500\ \mu\text{m}$ source spacing, $f/4$ optics, and unit system magnification. The optimization goal is to maximize the amount of light captured by a small (typically $20\ \mu\text{m} \times 20\ \mu\text{m}$) aperture on the output plane, representing a detector element on an opto-electronic chip.

3.2.1 Systems

Various models have been considered for systems consisting of refractive lenslets as well as spherical and aspheric diffractive lenslets. Here lenslet array systems are defined as those in which all lenslets within an array are identical. Individual lenslet systems would be much more complex, for each lenslet would be independently optimized for the particular interconnect paths it is required to support. The four systems chosen for study represent a progression in increasing design complexity and fabrication cost as well as expected performance.

3.2.1.1 Refractive Lenslet Array

This type of system consists of a uniform array of refractive lenslets. OTIS has been experimentally verified using epoxy on glass refractive lenslet arrays.⁽³⁻⁷⁾ The results can be seen in Figure 3-1. This type of lenslet is close to plano-convex in shape, which is not optimal for the reduction ratio of the system. Refractive lenslet arrays are relatively inexpensive, but are not perfectly spherical. In general, they suffer focal length non-uniformity among the elements and limited available $f\#$'s. The refractive system was not modeled using individually optimized refractive lenslets as such components are difficult to fabricate and generally not commercially available, or prohibitively expensive. For the same reasons we did not consider aspheric lenslet arrays.

3.2.1.2 Diffractive Lenslet Array

A diffractive lenslet array system is also a uniform array of lenslets, except that the elements are now multi-level phase computer generated holograms. The CGH's offer more design flexibility and uniform fabrication over refractive lenslets. However, the minimum feature size and number of phase levels sets a lower bound to the achievable $f\#$. Also, like all diffractive optics, they suffer from large chromatic dispersion.

3.2.1.3 Individual Diffractive Lenslets

This system is identical to the diffractive lenslet array, except the focal length of each lenslets is individually optimized. Note that the outer most lenslets tend to support the longer (diagonal) paths. The individual lenslet approach lets the focal length of these elements compensate for this path difference. Those lenslets which have been identified as being functionally equivalent are linked together during optimization, while unique functions are allowed to take their optimal form.

3.2.1.4 Individual Aspheric Diffractive Lenslets

This represents the highest performance system which we can design and fabricate; each unique lenslet function is designed independently and its phase function is allowed to take an arbitrary shape. Aspheric design provides the extra degrees of freedom necessary for reduction of third-order aberrations without additional optical surfaces. The additional performance of the aspheric components is only realized when each lenslet is allowed to be individually optimized; therefore an array of identical aspheric diffractive lenses was not modeled.

3.2.2 Approach

Traditional optical system design software is not well suited to modeling a highly parallel 256 channel free-space optical interconnection system. To define defining the individually optimized lenslets, an enormous computational task is required to optimize large non-sequential surface

(NSS) ranges. In addition, the limits on the total number of field points also prohibit the implementation of a complete system modeling. To accurately model OTIS we take advantage of the large degree of symmetry inherent in the system and reduce the simulation complexity.

The sources, representing either optical modulator or vertical cavity surface emitting laser (VCSEL) arrays, are modeled as object field points. To increase efficiency and reduce cross-talk, the chief rays of each source is aimed at its respective interconnect lenslet. Again, complete modeling of all interconnect paths is impossible because of the limited number of field points allowed by the design software. A sub-set of interconnection paths must therefore be selected. For a 256 channel system, there are three unique lenslet functions on each lenslet plane. The paths are chosen such that each pairing of the lenslets functions between the two planes is represented; only unique paths are included. Each field point is weighted according to the total number of paths it represents. Figure 3-5, a typical system view from Code V[®], shows ten of the interconnect paths associated with an edge lenslet.

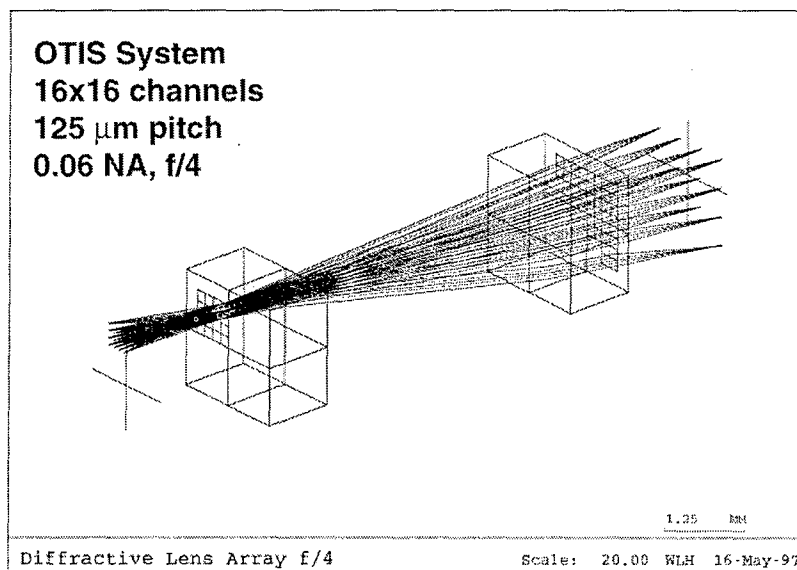


Figure 3-5. Code V[®] perspective view (VIE;VPT) of a 256 channel OTIS showing input and output planes, four transmitter lenslets, ten receiver lenslets, lens substrates, and ten objects.

The directed illumination of the modulators is formed by uniform illumination of an array of lenslets having square apertures. As a result the modulated light will have a sinc^2 profile. We have approximated this as a Gaussian, as this type of apodization is much easier to implement in Code V[®]. This approach is also valid if VCSELs are used rather than modulators.

All systems have two glass plates included which represent the lenslet substrates. We have included 1.2mm thick low-expansion (LE) glass plates as this is the substrate used for the in-house fabrication of CGH's.

Each interconnect lenslet plane is defined as a non-sequential surface range. Our initial designs were accomplished by specifying a single element and using the ARR command to replicate it into an array. For the individual lenslet systems we needed to set up a plane of non-identical lenslets. To insure that all rays intersect the proper lenslet we had to use non-sequential surfaces (NSS). Both NSS ranges have zero thickness; they are located on the surface of the substrates. Within each range the x-y location (decentration) and size (clear aperture and edge location) of all the lenslets are defined.

The refractive lenslets were modeled as plano-convex. Normal methods, such as specifying radius of curvature, thickness, and index of refraction for each surface were used to define the lenslets. A surface listing of one of the lenslets in the transmitter lenslet plane is given in the table below.

TABLE 3-1. Refractive lenslet listing showing NSS range, radius of curvature, glass, and decentration terms.

	RDY	THI	RMD	GLA	CCY		THC
6:	5.09137	0.000000		AIR	1		100
	INSS						
	GL2:	BK7_SCHOTT					
	XDE:	2.250000	YDE:	2.250000	ZDE:	0.000000	DAR
	XDC:	100	YDC:	100	ZDC:	100	
	ADE:	0.000000	BDE:	0.000000	CDE:	0.000000	
	ADC:	100	BDC:	100	CDC:	100	

All of the CGH's are modeled as diffractive optical elements (DOE's). We do not define the element in the normal manner by specifying the optical recording geometry, but rather suppress it by setting object and reference points to infinity. We then directly model the phase function, i.e. the optical path difference (OPD) introduced by the element. In Code V[®], the phase function can be defined by the coefficients of a polynomial expansion. The expansion may be tenth order in X and Y, in which case the number of the coefficient is given by Eq. (3-3):

$$j = \{(m+n)^2 + m + 3n\} / 2; j \leq 65 \quad (3-3)$$

where m,n are the powers of X,Y respectively. For a rotationally symmetric element, the expansion may also be in R. Equation (3-4) gives the coefficient number in this case:

$$j = m/2; j \leq 10 \quad (3-4)$$

where m is an even power of R, through R^{20} . For example, a spherical diffractive lens has only the first coefficient, C1, that of R^2 ; in the x-y expansion, there are two identical coefficients, C3 and C5, for X^2 and Y^2 . In both cases, the coefficients are equal to $-1/2f$.

In the aspheric model C3 and C5 are not required to be equal; C1 and C2 have been include to introduce linear phase terms ('tilt' the lens) which should aid the off-axis imaging of the corner lenslets. Other higher order terms, such as C10 and C14 for X^4 and Y^4 are included for aberration correction. When the final design is selected, the coefficients are used to generate the mask patterns for the fabrication of the CGH. An example listing of a diffractive lenslet is given below.

TABLE 3-2. Diffractive lenslet listing showing NSS range, HOE coefficients, and decentration terms.

	RDY	THI	RMD	GLA	CCY		THC
6:	INFINITY	0.000000			100		100
	INSS						
	GL2:						
	HOE:						
	HV1:	REA	HV2:	REA	HOR:		
	HX1:	0.000000E+00	HY1:	0.000000E+00	HZ1:	0.100000E+16	
	CX1:	100	CY1:	00	CZ1:	100	
	HX2:	0.000000E+00	HY2:	0.000000E+00	HZ2:	-.100000E+16	
	CX2:	100	CY2:	100	CZ2:	100	
	HWL:	514.50	HTO:	SPH	HCT:	R	
	HCO/HCC						
	C1:	-6.5036E-02					
	C1:	1					
	XDE:	2.250000	YDE:	2.250000	ZDE:	0.000000	DAR
	XDC:	100	YDC:	100	ZDC:	100	
	ADE:	0.000000	BDE:	0.000000	CDE:	0.000000	
	ADC:	100	BDC:	100	CDC:	100	

We used the automatic design (AUT) with the default error function to optimize the lens functions. The default error function is a center weighted RMS spot size based on weighted transverse ray aberrations. For the refractive system we let the radius of curvature, and therefore the focal length, of the lenslets vary. For the diffractive systems, the coefficients of the phase function were allowed to vary freely. Surfaces which were identified as being identical, or mirror images, had coupling codes established to maintain the relationship. In order to maintain the proper system geometry for the transpose operation, the reduction ratio of the system was fixed to one (unit magnification), and the distance from the object and image plane to the lenslet planes were linked. The refractive lenslet array system converged very quickly, usually within two or three cycles. The diffractive lenslet array system also converged within two or three cycles, however each cycle took considerably much longer, usually several minutes. The individual lenslet systems took more cycles to converge, usually between five and ten, and each cycle took longer, usually at least five minutes.

3.2.3 Code V[®] Simulation Results

The key figure of merit used to judge the performance of the system is the spot size on the output (detector) plane. We evaluated the diameter of the circle which captures 80% of the energy of the point spread function (PSF). The PSF option verified that the centroid of the spot lies very close (on the order of microns) to the chief ray. While the point spread function alone is not necessarily a good judge of the performance of an imaging system, we are concerned only with point-like sources. The spot size is critical as it determines how large the detector elements must be to achieve good efficiency. Smaller detectors are preferable as they are in general faster and result in receivers with lower power consumption; our goal is spot sizes, and detector areas, on the order of (20 microns)². The composite spot sizes of various interconnect paths for the different systems are summarized in Table 3-3.

The refractive system performs well on-axis, but poorly at even relatively small field angles. The diffractive system performed significantly better; it is interesting to note that at the optimal composite focal plane, the on-axis performance was inferior to the average of the intermediate off-axis fields. The individual diffractive lenses didn't improve the extreme off-axis performance as much as was hoped, but did offer the benefit offering more consistent performance through the on-axis and intermediate off-axis fields. Surprisingly, the aspheric system did not improve the results. This is possibly due to the default error function failing to accurately represent the system performance. We are optimistic that a custom error function would allow the aspheric system to perform significantly better.

Table 3-3. Spot size simulation results with Code V®.

80% Encircled Energy

	Array	Indiv. Spher.	Indiv. Asph.
Average	17.37	17.05	16.20
Minimum	14.78	14.69	14.05
Maximum	35.00	29.58	25.41
Std. Dev.	3.43	2.83	2.52

Strehl Ratio

	Array	Indiv. Spher.	Indiv. Asph.
Average	0.95	0.95	0.98
Minimum	0.60	0.59	0.85

3.3 OTIS OPTICAL SYSTEM IMPROVEMENTS

3.3.1 Birefringent Computer-Generated Hologram

For high efficiency and power uniformity, the OTIS optical system requires individually directed transmitters. To achieve directed signal beams in a system utilizing reflective modulators, an additional array of illumination lenslets is required (Figure 3-6). The interconnect lenslets and illumination lenslets may be implemented as a single optical element if birefringent computer-generated hologram (BCGH) technology⁽³⁻⁵⁾ is utilized. We have designed, built, and characterized such holograms. One of the BCGH elements is shown in Figure 3-7. The modulator input spots generated by this element are shown in Figure 3-8, and the spots produced by the same element at the intermediate image plane of the OTIS interconnection system are shown in Figure 3-9. Beam scans through these planes are shown in Figure 3-10. The spot size measured using the beam scan instrument tends to be less accurate (larger) than that measured by imaging the spots onto a CCD camera because of the difficulty in aligning the BeamScan's slit with a row of spots.

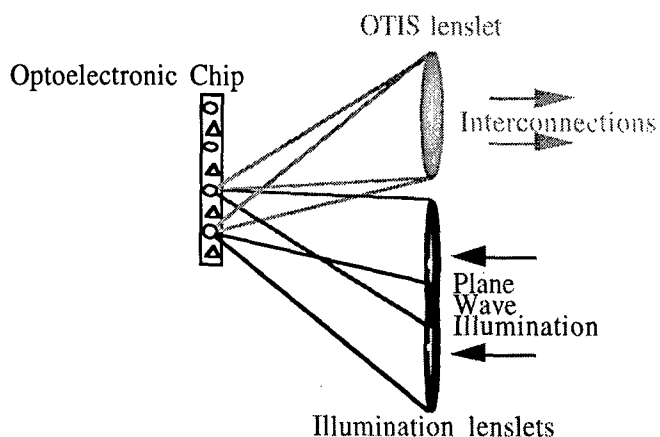


Figure 3-6. Illumination lenslets and interconnect lenslets may be superimposed in the same plane by utilizing birefringent computer-generated holograms (BCGH).

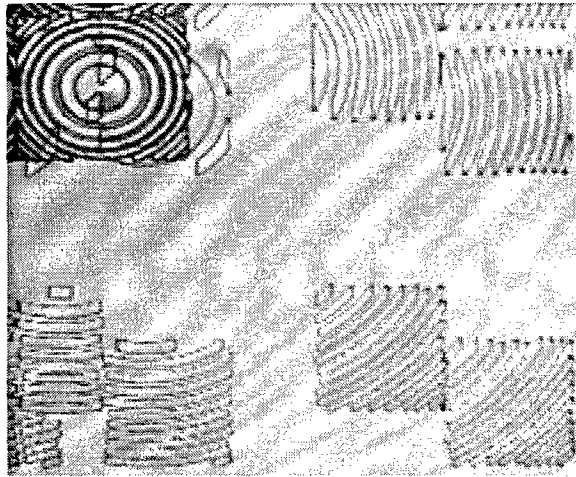


Figure 3-7. Birefringent computer generated hologram implementing both the illumination lenslets and the OTIS interconnection lenslets. *The minimum feature size of the hologram is $5\mu\text{m}$.*



Figure 3-8. Modulator input generated by BCGH illumination lenslets. *The full width at half-maximum (FWHM) spot diameters in the modulator plane range from $25\mu\text{m}$ to $37\mu\text{m}$.*

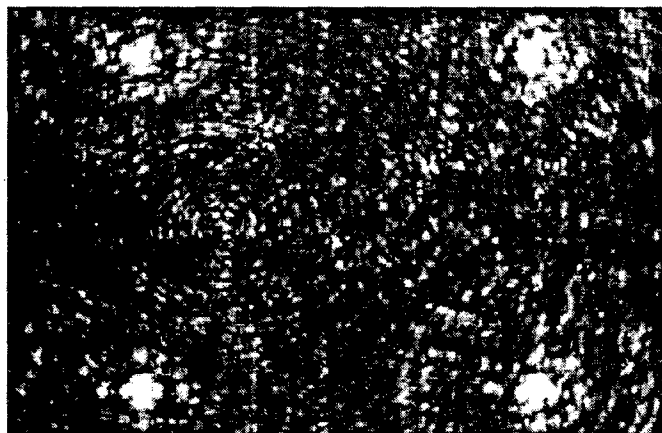


Figure 3-9. Spots in the intermediate image plane (the focal plane of the interconnect elements) of the OTIS system. *The light recorded here has passed through the BCGH elements twice, experiencing a different lens function on each of the two passes.*

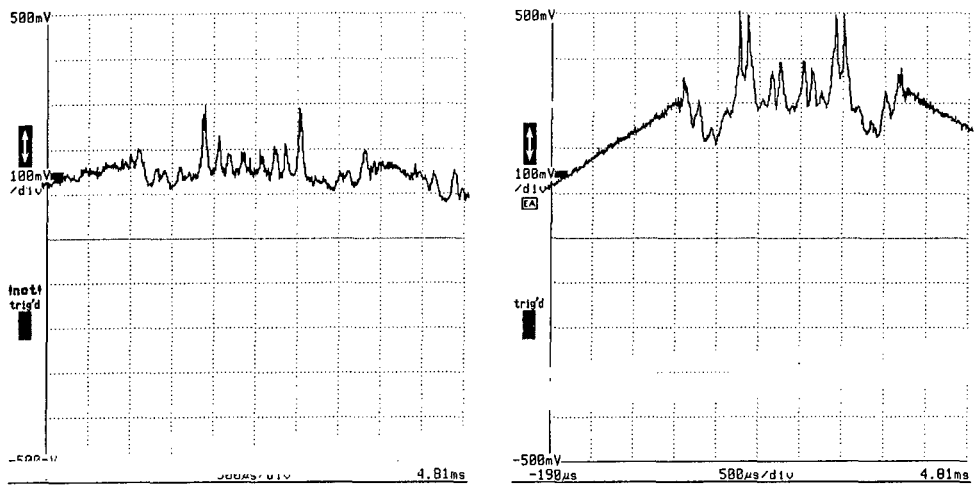


Figure 3-10. Beam scans of spots in the modulator plane (left) and intermediate image plane (right) of the OTIS system.

3.3.2 Photorefractive Beamsplitter

Free space optical interconnection systems utilizing reflective modulators as transmitters require an optical device which can couple optical energy from a transverse direction into the modulators.⁽³⁻⁸⁾ As illustrated in Figure 3-11, this same device needs to allow the optical energy reflected from modulators to transmit through to detectors on the opposite side. Not only to direct illumination light to the modulators, this element should also provide a low insertion loss on the interconnect path.

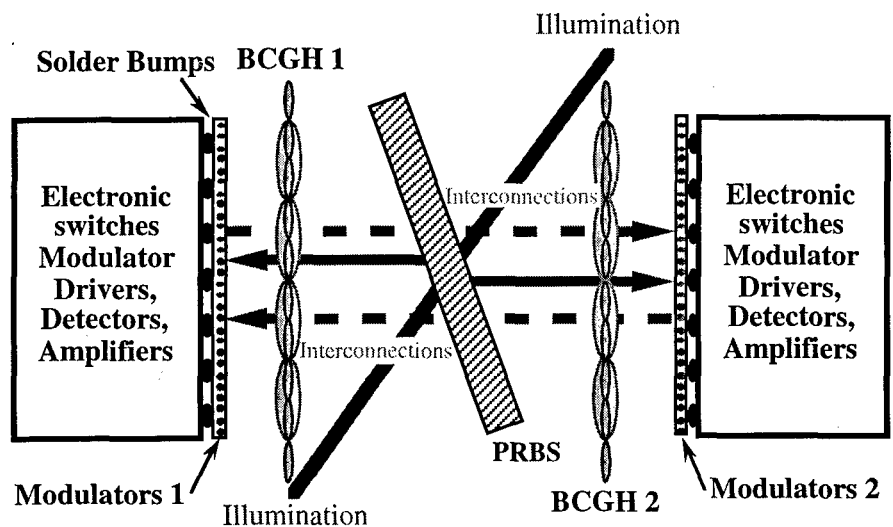


Figure 3-11. Overall system schematic for a free-space optical interconnection system. The beamsplitter is needed to bring light into the modulators, but also to remain transparent to the transmitted light.

Traditional approach uses a Polarizing Beam Splitter (PBS) in combination with a quarter-wave retardation plate. The PBS has a high reflectivity for the vertical polarization and a high transparency for horizontal polarization. A quarter wave plate placed between the PBS and modulators will perform the necessary polarization rotation. However, due to the thin film characteristics of the polarizing beamsplitter, the transmission percentage is highly dependent on the incident angle. If the interconnects are arranged in a shuffle fashion where a modulator in one corner of the system reflects its light to a receiver in the opposite corner of the system, off-axis interconnect angles are required. For incident light more than 5 degrees off-axis, the transmission percentage of a PBS decreases significantly.⁽³⁻⁹⁾ To maintain an efficient and compact free space optical interconnect system using a PBS, the distance between the modulators and detectors must be increased to match the longer $f\#$ required by the PBS through the following equation:⁽³⁻¹⁰⁾

$$f_{\#} \geq \frac{1}{\sqrt{2}} \left(\frac{\sqrt{M}}{\sqrt{M+1}} + \frac{\sqrt{N}}{\sqrt{N+1}} \right) \cdot \frac{1}{\tan \theta}, \quad (3-5)$$

M and N represent the number of transmitters and receivers and θ is the maximum acceptance angle. For a system where $M = N = 64$ and θ is 5 degrees, the $f\#$ would be 14.1. To achieve a compact interconnect system with a $f\#$ of 4, the incident angle requirement would have to be 17.5 degrees; three times the limitation of a PBS. So to create compact efficient interconnect systems other technologies must be examined for replacing the polarizing beamsplitter.

In this program we examined the feasibility of replacing the conventional PBS with a volume diffraction hologram recorded in a photorefractive crystal. This photorefractive beamsplitter (PRBS) was designed to have both high redirection efficiency and angular independent transmission. The PRBS works on the principle "splitting" the light into the zero and first order diffraction modes. The diffraction grating is recorded onto the PRBS by using the photorefractive effect to locally modulate the index of refraction inside the crystal.⁽³⁻¹¹⁾ The intersection of two laser beams inside the crystal create a sinusoidal interference pattern. Electrons residing in bright bands of this pattern become ionized and diffuse towards the dark bands of the interference pattern where they recombine. The displacement of these electrons creates an electric field inside the crystal, which via the Pockel's effect, modulates the index of refraction and creates the grating. The thickness of the crystal allows the diffraction grating to operate in the Bragg regime; where only an incident beam with a the required k vector will match the necessary Bragg condition,

$$\vec{k}' = \vec{k} + \Lambda \quad (3-6)$$

where Λ is the grating vector of the diffraction grating. The redirected light, given by k' , only appears when the incident light k matches this equation. Any other incident beam will not match the necessary conditions and therefore will pass through the crystal with no coupling loss. The

PRBS utilizes this volume hologram recording of a diffraction grating to redirect light with high efficiency while still allowing light to pass through in any other direction without coupling.

The photorefractive crystal used for the PRBS is LiNbO_3 with an iron doping at a concentration of 0.01%. The iron doping increases the number of donor atoms providing electrons to the grating formation and thus improves the diffraction efficiency of the grating. However, doping the crystal with too much iron will lead to higher absorption of optical energy and lower transmission. A doping concentration of 0.01% compromises between the increase in both diffraction efficiency and absorption loss. The LiNbO_3 crystal is an uniaxial crystal with refractive indices of 2.33 and 2.29 for the ordinary and extraordinary components respectively. LiNbO_3 is also classified as a tetragonal 3m crystal with its electro-optic coefficient r_{33} having a value of 32.6×10^{-12} m/V. Utilizing this electro-optic coefficient will provide a good modulation of its extraordinary refractive index. The r_{33} electro-optic coefficient requires that the z crystal axis be perpendicular to the propagation direction of the optical energy and parallel to the light's polarization. The Fe: LiNbO_3 crystal used was an x-cut crystal and was ordered to have dimensions of $1 \times 20 \times 20$ mm \pm 0.5 mm for the x, y and z axis. The x axis of the crystal was measured to be 1.45 mm, allowing the effective grating depth to be \sim 1.28 mm. The thickness of the grating is not 1.45 mm exactly because the intersection of the two recording beams does not occur over the entire 1.45 mm distance. This thickness does provide for a diffraction grating with a high angular selectivity because of a more stringent Bragg matching condition. The crystal also has an anti-reflective coating for 0.05 % reflectivity for 514.5 nm light incident at 35 degrees from the normal of the crystal. This coating provides better transmission for off axis incident light.

The recording of the crystal was setup to create a transmission diffraction grating. The recording beams from an Argon laser at a wavelength of 514.5 nm were co-directional and the redirected beam diffracts from the opposite side of the crystal as the pump beam. The angular selectivity increases when the diffraction grating has a shorter grating period. A theoretical analysis of the crystal showed that the two recording beams at 45 degrees with respect to the surface normal create a grating period of 0.36 μm . This small grating period requires that the recording setup be insensitive to all vibrations. Since the electrons in the photorefractive crystal ionize in the bright bands of the interference pattern and recombine in the dark bands, any lateral shift in the location of these bands deteriorate the diffraction grating and decrease its efficiency. The part of the recording equipment consisting of a beamsplitter, mirrors and photorefractive crystal were each placed on steel mounts and then all four mounts were placed on a single steel baseplate. The steel provides better insensitivity to thermal expansion than normal aluminum and the baseplate reduces the number of independent directional modes. Finally, the entire setup was covered with a box to reduce the influence of the surrounding atmospheric conditions. By replacing the photorefractive crystal with a second beamsplitter, a Michaelson interferometer was

created to measure the stability of the recording setup. Examining the phase shift of the interference pattern over time showed that the average phase of the pattern did not shift over an hour of measurement; indicating that the steel mounts and baseplate would provide adequate stability for the duration of the grating recording.

A second problem with the recording was determining when to stop the grating recording. Formation of transmission gratings varies sinusoidally over time and so the recording needs to be stopped at a peak in the grating formation when the diffraction efficiency is at a maximum. A He-Ne laser probe was used to measure the grating's diffraction efficiency over time. The He-Ne light was incident on the grating at an angle of 60.4 degrees; necessary to meet the Bragg criteria. A photodetector with a 514 nm line interference filter was placed on the opposite side of the crystal and the diffracted light from the He-Ne was measured and plotted over time to indicate when the grating had reached a maximum.

The two recording beams at 514 nm were about 1 cm in diameter and each had about 75 mW of power incident on the crystal. The recording time for the grating was about 26 minutes, emphasizing the need for recording stability. The highest diffraction efficiency to date is about 68%. Measurement of the Bragg angular selectivity shows a FWHM of 0.04 degrees. This proves that the thick diffraction grating can provide good redirection at a single angle and transmit anything else not matching the strict angle selectivity. This experimental Bragg selectivity was compared against the theoretical prediction for diffraction efficiency,

$$\eta = \frac{\sin\left(\sqrt{\chi(\theta_b, \theta_i)^2 + \phi(n', \theta_b, \theta_i)^2}\right)^2}{1 + \frac{\chi(\theta_b, \theta_i)^2}{\phi(n', \theta_b, \theta_i)^2}}. \quad (3-7)$$

The χ in this equation is the decoupling parameter and uses the Bragg and incident angles,

$$\chi(\theta_b, \theta_i) = \frac{\pi(\theta_i - \theta_b)}{\Lambda(\theta_b)}, \quad (3-8)$$

where t is the grating thickness, Λ is the grating period and the b and i subscripts correspond to the Bragg and incident angles. The ϕ in the efficiency equation is the modulation parameter,

$$\phi(n', \theta_b, \theta_i) = \frac{\pi n'}{\lambda \cos(\theta_i)}, \quad (3-9)$$

where n' is the index modulation. For an index modulation of 1.053×10^{-4} , an effective grating thickness of 1.28 mm and the Bragg angle inside the crystal at 17.998 degrees, the experimental results match the theoretical predictions, as shown in Fig. 3-12.

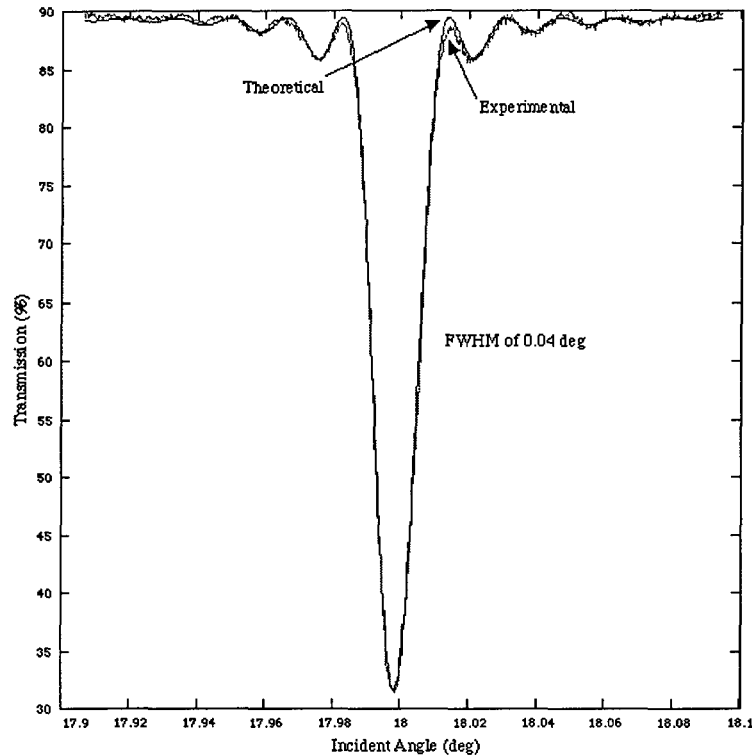


Figure 3-12. Experimental and theoretical comparison of the angular selectivity of the diffraction grating. The experimental results show the transmission loss as the light becomes diffracted into the first order. The theoretical predictions show almost a perfect match for an effective grating thickness of 1.28 mm and a modulation index of 1.053×10^{-4} .

Experimental results show that the PRBS has good diffraction efficiency and narrow angular selectivity. Another experimental test was performed to determine how well light would transmit through the crystal from a wide range of incident angles. The photorefractive crystal was placed on a motorized rotation stage and slowly rotated while a photodetector measured the transmitted light from a single beam. Results show that the PRBS has a transmission percentage range of 96% to 76% for incident angles of 25 to 65 degrees. The same transmission test was also performed on a polarizing beamsplitter to provide a simple comparison (shown in Figure 3-13) between the two devices. Within the operational range of 25 to 65 degrees, the PBS showed a transmission percentage range of 100% to 46%. These results show how angular dependence of the PBS affects the transmission. For an incident beam 17.5 degrees off-axis (corresponding to an $f\#$ of 4), the PBS transmits as low as 49% of the light while the PRBS transmits 80%. Since the PBS redirects 100% of the optical energy while the PRBS currently redirects 68%, the total transmission efficiency for the previous example is about 58% for both optical devices. The comparison shows that the PBS has a higher efficiency than the PRBS, but also has a larger deviation.

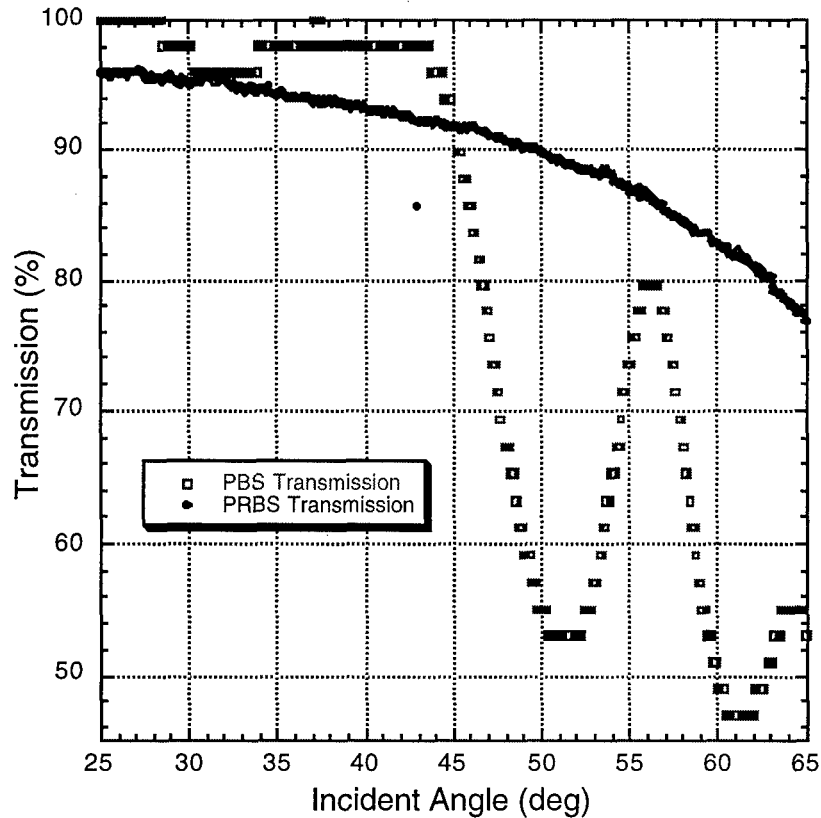


Figure 3-13. Transmission efficiency of both PRBS and PBS over the operational range needed for a $f/4$ interconnection system.

An initial simple comparison between the PBS and PRBS showed roughly equal performance for the two devices. More detailed measurements are needed for full characterization of the PBS and PRBS transmissions, including multiple beam transmissions from a lenslet array. The performance of the PRBS may be further optimized by various approaches, such as using a different crystal with higher iron doping, recording at a different geometry and an antireflective coating at a higher angle from the normal, where the transmission curve decreases. These improvements will make the PRBS's redirection and transmission efficiency comparable to the PBS, but will include the currently smaller deviation in efficiency. A study into other possible technologies, such as a wire grid polarizer,⁽³⁻¹²⁾ needs to be done to see how well the PRBS compares. Overall, this experimental study shows that the PRBS has the potential to provide a good replacement for the PBS in free space optical interconnect systems and enables the systems to be both compact and efficient.

Section 4

SPACE-TIME COMPANDER

One of the key functions of the space-time compander is to match fine-grain (e.g., 1024×1024) images with the coarse-grain (e.g., 128×128) processor array. The size matching is performed by grouping every set of 8×8 pixels in the fine-grain image into a superpixel. Each superpixel is then registered with the corresponding processor in the processor array. By either compacting 8×8 pixels into a superpixel or expanding a superpixel into 8×8 pixels, the STC provides a bi-directional communication between a fine-grain image and a coarse-grain processor array. The most straight forward approach is to have a buffer array to convert the 2-D spatial (parallel) information into 1-D time (serial) information. This serial \leftrightarrow parallel buffer array structure can be accomplished by the use of the charge coupled device (CCD) technology. This approach calls for the development of a special Compander with serial \leftrightarrow parallel CCD circuits on one side and an LCD spatial light modulating layer on the other (as shown in Figure 4-1).

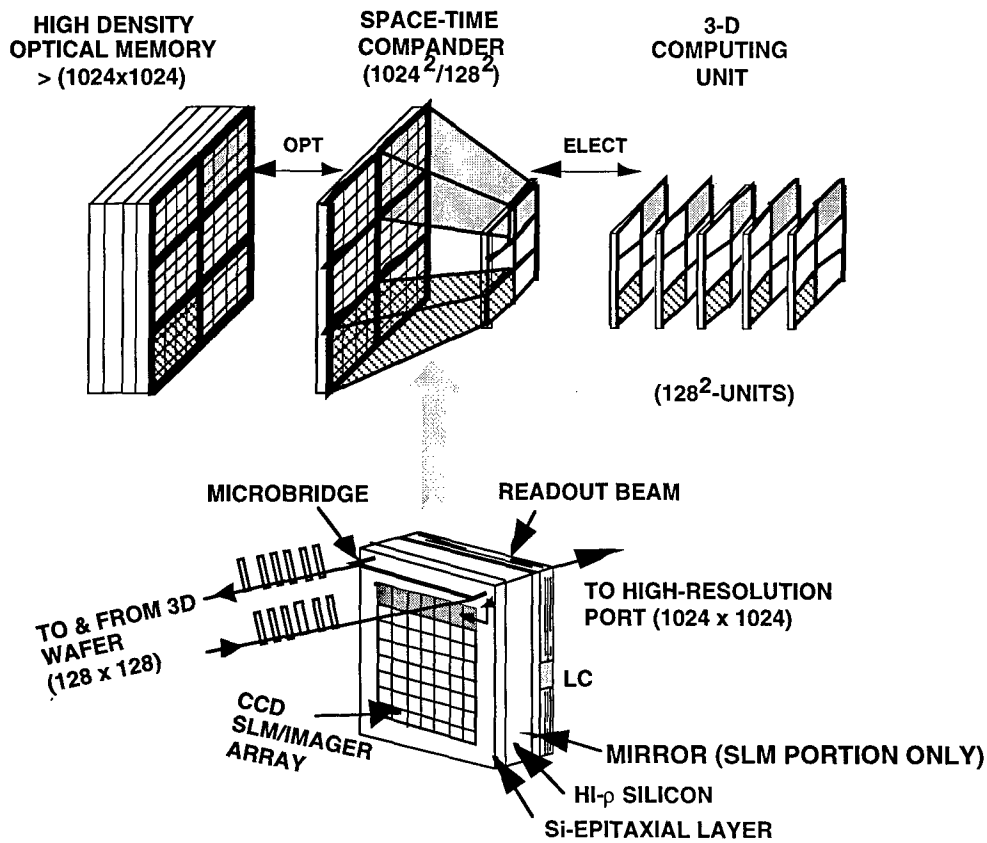


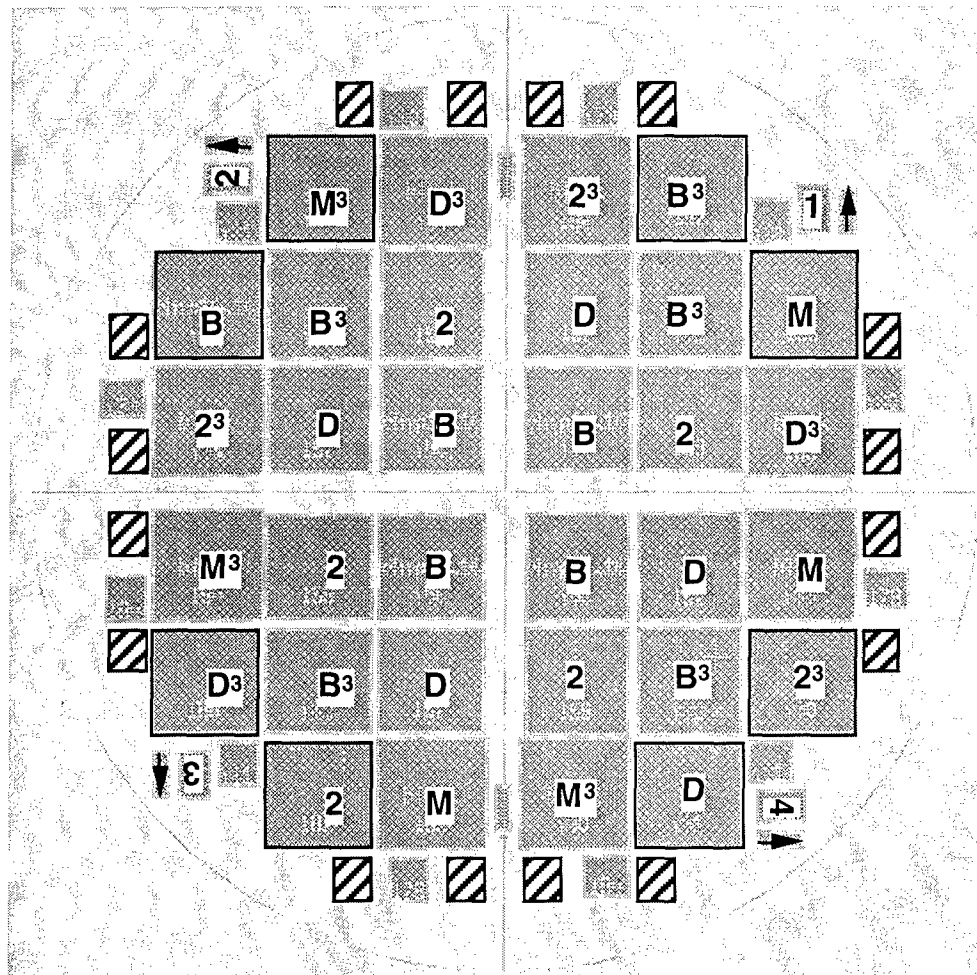
Figure 4-1. Schematics of one superpixel of the Space-Time Compander.

4.1 CCD-BASED LIQUID CRYSTAL IMAGER/MODULATOR

Traditionally, Hughes CCD-LCLV relies on a CCD array to distribute two-dimensional image information to the LC modulation layer.^(4-1, 4-2) In principle, the same CCD array can also be used to sense an input image, with the help of the through-wafer depletion operation scheme. This operation scheme allows the conventional CCD-LCLV to be used as an aerial imager. To achieve the combined modulation/detection functionality, the device should be operated at two different wavelengths for writing into and reading from a high resolution optical component (e.g., optical memory, etc.). In the write-in phase, the optical beam is reflected from the mirror (for the reading wavelength) and is modulated by the two-dimensional charge pattern information in the CCD array for addressing the high resolution component. In the readout phase, the optical information from the high resolution component can transmit through a specially designed leaky mirror (for the writing wavelength). This will result in a two-dimensional charge pattern that can then be readout by the same CCD array. The combined modulator/detector device permits a simpler optical design to be used in the system. Although performance of both modulation and detection may have to be compromised, this approach is preferred for the binary information operation because of its inherent simplicity. In addition, this approach permits a larger pixel size and thus supports a better modulation transfer function (MTF).

4.2 CCD ARRAY FOR STC APPLICATION

Analysis indicates that we can obtain a combined modulator/detector device with a slight modification of the current Hughes CCD-LCLV structure. To facilitate deep through-wafer depletion, the 5 μm epilayer used on the p-Si substrate for fabricating the CCD gate structures is replaced with a thin (about 0.2 μm) p-type sheet implant layer. We established the basic architecture of the CCD circuits and together with the silicon foundry, defined the fabrication process suitable for our high voltage requirement. The floorplan for the CCD circuit, as shown in Figure 4-2, consisted of four quadrants, each of which had eight circuit chips and seven process control monitors (PCM's). Two of the eight circuit chips and four of the seven PCM's would have metal patterns fabricated by the foundry for pre-thinning circuit characterization purpose. These two circuit chips would not be available as product chips. The other six circuit chips were product chips which together with the three remaining PCM's would receive their metal patterns after the wafer was scribed into quadrants and thinned down. To safeguard against process and design uncertainties, two clock schemes (two-phase and three-phase) for the serial-CCD circuit had been included on the wafers. The two-phase structure is desirable because it requires one less clock driver and the clock line width is slightly larger. However, it requires an extra implant and is more critical regarding clock drive parameters.



KEY	
2Ø serial CCD	3Ø serial CCD
B : Baseline design	B ³ : Baseline design
D : Double drive out. amplifier	D ³ : Double drive out. amplifier
2 : 2-stage out. amplifier	2 ³ : 2-stage out. amplifier
M: Mixed array ('B' in row 1-4, 'H' in row 5-8, 'D' in row 9-12, '2' in row 13-16; 'H' is 1.5 drive)	M ³ : Mixed array ('B ³ ' in row 1-4, 'H ³ ' in row 5-8, 'D ³ ' in row 9-12, '2 ³ ' in row 13-16; 'H ³ ' is 1.5 drive)

Figure 4-2. Floorplan of Space-Time Compander wafer.

The other variations of the circuit chips were the output amplifier drive capabilities. Four versions of chips were incorporated in the layout: the all-low drive, the all-medium drive, the all-high drive and the mixed drive. Totally there were eight different versions of the circuit chips incorporated in the floorplan of the whole wafer and one of each was available for pre-thinning characterization. In each of the circuit chips, besides the 16×16 superpixels, there were test circuits that could be used to characterize line resolution and transfer efficiency of the front-back charge transfer process. There were also test circuits for testing the operation of individual superpixel, a shorter serial-CCD chain and a shorter parallel-CCD chain (the product had an 8×8 CCD array). Figure 4-3 shows the floorplan of a typical Compander chip with test cells populating to the right and below the 16×16 super-pixel array. Figure 4-4 shows the layout of the superpixel containing an 8×8 CCD array. The masks of the layout design were fabricated and visually checked to be free of gross feature errors.

4.2.1 Electronic Driver for CCD-Based Modulator/Imager

For the compander to be used as an imager it will be necessary to view the output data visually to demonstrate proper operation. A computer program was written to display the active elements of the superpixel array on a computer screen with the proper aspect ratio and spacing. The data in selected superpixels could be dynamically updated on the screen as the CCD was read out. A commercial 24 channel logic analyzer module was used to sample superpixels simultaneously and to transmit data over a high speed serial link to the host computer. The update rate was somewhat less than the actual CCD frame rate, but the visual image would give the desired results.

A system level design was completed for driving the compander in both the imager and light modulator modes. The logic analyzer module was used to accept data in the imager mode and output it over the serial interface. The same signal lines were used to input data in the modulator mode. A circuit board was added to the CCD clock timing generator to allow data patterns to be generated for stripes and squares of various pixel sizes. The patterns were static but could be changed through wiring options for each superpixel.

4.2.2 CCD Wafer Testing

Initially, only a very small sample of parametric data was collected by the foundry. These data indicated good devices on the wafers tested but more tests were needed to cover all wafers and to establish device yield confidence. The foundry had completed these tests that consist of measuring the transistor threshold voltages, current factor, breakdown voltage, field threshold voltage, diffusion and poly resistance, contact resistance and inter-layer shorts. These parameters were collected from 18 of the 19 fabricated wafers. The device yields were found to be very high, except for poly-3 to N+ diffusion shorts, which was only 50% for four wafers and 75% for other four

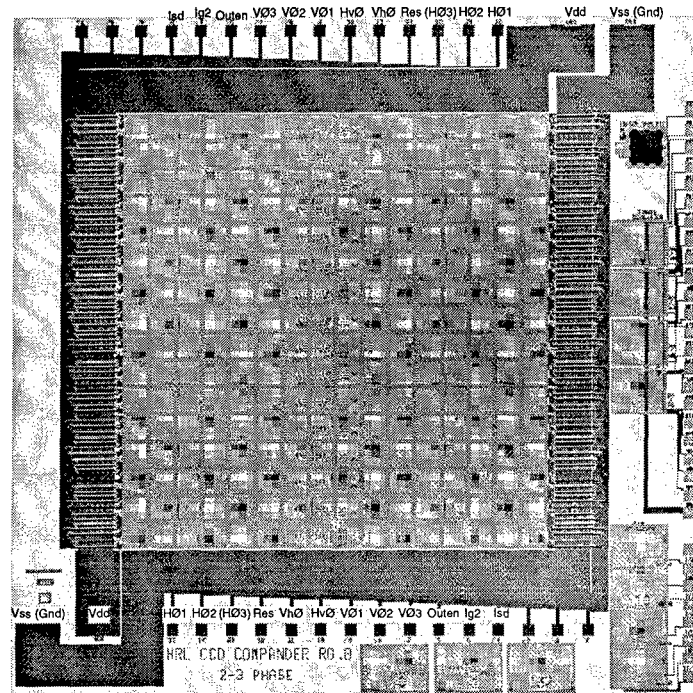


Figure 4-3. Floorplan of Space-Time Compander chip.

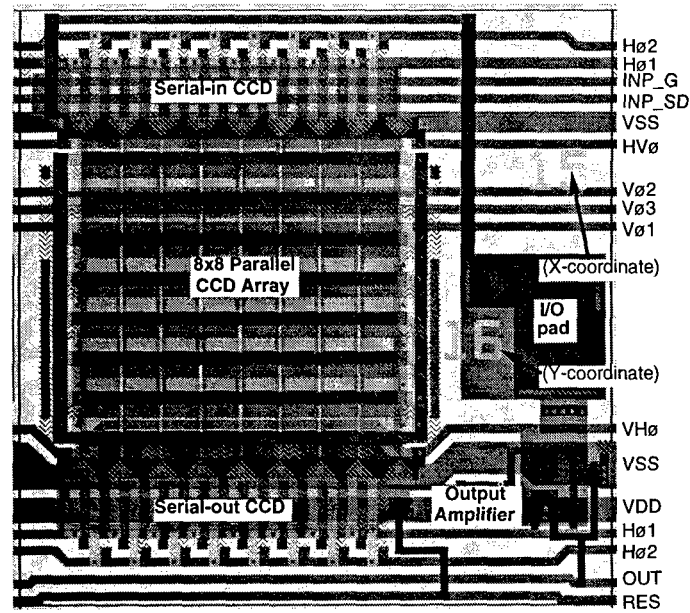


Figure 4-4. Space-Time Compander superpixel (8x8 CCD array) layout.

wafers. No single wafer had gross inter-layer short problem. If process non-uniformity was the cause of these shorts, we might expect yield impact on the CCD circuits.

We conducted functional testing of the circuits and found both the two-phase and three-phase clock versions of superpixel test cells operational, albeit with somewhat different clock voltages. This verified that the basic CCD design was correct. We also found functional superpixels in these full arrays, verifying that the interconnects used in these full arrays were correct. In addition, light sensitivity was observed when the STC was operated as an imager. Clearly, photo-generated charges were collected and transferred through the silicon wafer. However, no definite resolution pattern has been unambiguously observed in the CCD output signal.

4.2.2.1 Electrical Testing

Testing of the CCD chips was done at the wafer level using a standard probe station with a binocular microscope (shown in Figure 4-5). Two separate probe cards were designed to match the bonding pad layouts on the chips, one for the test devices and one for the main array. The probe cards could be interchanged in the probe station and used the same drive signals. The test devices were small scale subsets of the whole array that allowed specific parts of the CCD structure to be tested separately. The electronics consisted of a bank of CCD pulse amplifiers controlled by a 16-channel master clock generator. The pulse amplifiers allow continuous adjustment of the voltage levels and rise times.

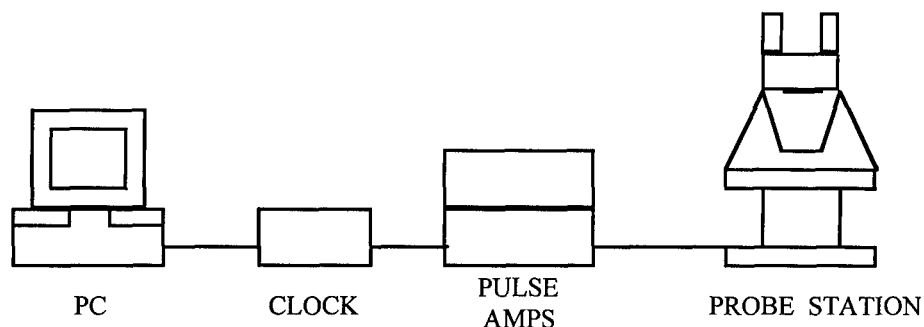


Figure 4-5. Electrical test setup for the CCD superpixel array.

Testing began with special test devices. The first device to show full CCD operation was a three-phase 8×8 pixel array. A pulse on the input gate would appear on the output after a delay of 8 lines. Varying the width of the input pulse caused a corresponding variation in the output pulse on a pixel by pixel basis. This indicated that both the horizontal and vertical CCD structures were working properly with no charge smearing. The pulse amplitude could also be modulated to show that the charge was continuously variable with no unusual thresholds or clipping.

Testing of other 3-phase CCD test devices was also performed. Both the serial and parallel test devices were functional. Wafer testing was performed to get an idea of the yield. It was found that nearly every die worked except for those with the 2-stage output amplifiers. The bias voltage levels required for the 2-stage amplifier generally led to breakdowns before they could bias properly.

The 2-phase CCD array was also found to be functional with a different set of timing and drive voltages. The signal handling capacity seemed to be less and it was more sensitive to clock timing. Wafer testing was also done on the 2-phase die, with similar results to the 3-phase.

After verifying that the special test devices were working, the main superpixel arrays were tested. The probe card for the main array was rewired to make it compatible with the test device probe card pinout. A separate single probe with X-Y-Z translation capability was used to contact the I/O pads on individual superpixels. Since the main array has 16x16 superpixels (256 total), the capacitive loading on the clock signals went up considerably. This caused distortion in the form of overshoots and level shifts, and adjustment of signals was necessary. The first array tested (baseline three-phase) was found to be working. Testing of other arrays on the wafer, however, showed no operation. Since there are 256 superpixels sharing the same signals, a defect in any superpixel will wipe out the entire array. Only one superpixel array was found to be functional, and it was subsequently damaged when a probe scratched the aluminum clock lines. Essentially at this stage we are facing a serious yield issue.

Wafer testing was continued to determine what the overall superpixel yield was. With about half the wafer lot tested, the yield appeared to be close to 33%. The defects were usually associated with the parallel clock lines, as expected. The breakdown voltages of the gate oxides were measured and appear to be right at the expected theoretical limit of 150 volts.

So far, all the wafer testing was performed on thick wafers processed by Orbit Semiconductor. In reality, the wafers must be thinned down to a thickness of about 5 mils before they can be used for the CCD-LCLV based compander. The thinning process consisted of a grinding and polishing operation similar to optical polishing. The CCD side of the wafer was protected with a layer of polysilicon prior to the thinning operation. After the back side of the wafer has been thinned and polished, a processing cycle was performed to place an array of diodes on the back surface. These act as charge collection points and keep the image from losing resolution. The polysilicon layer over the CCD was then stripped off after the final polish. Noticeable degradation was found between thinned wafers and thick wafers. This indicated that the thinning or back side processing was damaging the silicon and affecting electrical properties.

None of the chips tested initially exhibited characteristics that were good enough for assembly as a final compander. Some modifications were introduced to the back side processing in an attempt to cut down on the possibility of damage. Different metalization materials were used at

lower operation temperatures. Four more wafer quadrants were tested after the thinning process. Some of these had a different type of appearance due to the metalization changes. Two of the wafer quads were accidentally broken and had only a few die available for testing. Most of the die had problems that prevented them from working, but two were found that had good responses. One of them was excellent and was a good candidate for final assembly.

4.2.2.2 Optical Testing

Optical testing was also performed to see if the CCD arrays could perform as imagers. Shining bright light on the CCD structure showed that the chips were light sensitive. The question was whether the sensitivity could be maintained at resolutions down to the pixel level. This is necessary if the device is to be used as an imager. A single mode fiber, 4 μm core size, with He-Ne laser input was used as a light-emitting optical probe to illuminate isolated pixels. No discernible pattern could be detected, partly because the light intensity could not be controlled accurately enough to prevent the CCD from saturating.

Further optical testing was continued using wafer quadrants on the probe station with an optical beamsplitter underneath the wafer. . A test structure was designed to closely resemble the final STC structure. Wafers were first thinned to the typical 5 mils. A counterelectrode was prepared that had a continuous conductive indium-oxide coating together with an optical test pattern of 100 micron stripes. The optical test pattern would allow light to be transferred through the wafer and to selectively activate pixels in the CCD structure. The counterelectrode was bonded to the CCD substrate by the surface tension of a liquid crystal thin film. The output of the CCD would hopefully show a modulation pattern corresponding to the optical test pattern. This assembly was mounted in the probe station, as shown in Figure 4-6. A small beam-splitter cube was employed so light could be projected, through the optical test pattern on the counterelectrode, onto the back side of the wafer. The CCD array could then be probed from the top in the conventional manner.

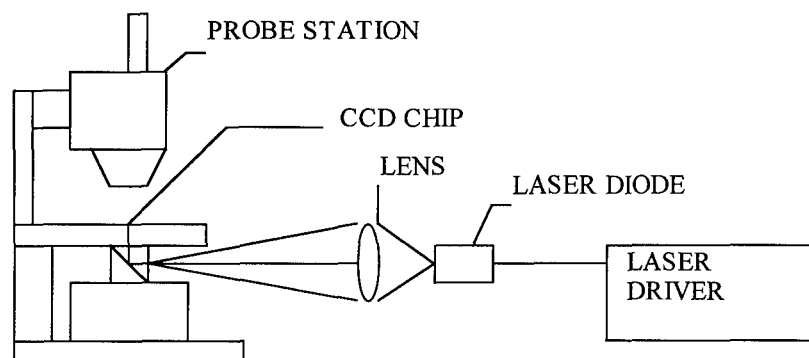


Figure 4-6. Optical setup for testing the image function of the CCD-based STC.

The CCD was driven using a modified clocking scheme that had a "dead" period where no clocking occurred. This period was intended to let photo-generated charge accumulate and be captured by the CCD array. The CCD array was then clocked out in the usual fashion, and the signal observed. The back side of the wafer was illuminated by a pulsed 839-nm laser diode with a focusing lens. The counterelectrode could be pulsed with a signal that was synchronized to the light pulse to aid in transferring the charge through the wafer. Light sensitivity was observed in this manner, indicating that photo-generated charge was transferring through the wafer, but no definite pattern was seen in the CCD output signal. The effect of driving the back surface counterelectrode with a voltage was quite significant on the CCD array and not entirely understood. Floating the CCD substrate electrically during charge injection was also attempted with no apparent improvement. No modulation pattern was ever observed on the CCD output.

4.3 STC PACKAGING

One of the important issues in developing the CCD-LCLV based STC involves device packaging. To directly communicate with processors, it is necessary to make electrically contact from STC superpixels to corresponding nodes on the processor array. Ideally, an optically flat Si substrate is desired in the CCD-LCLV to obtain a uniform liquid crystal layer and hence a uniform output light intensity. At present, the Si substrate in the CCD-LCLV is flattened using our transfer bonding technique.⁽⁴⁻³⁾ The Si wafer is first contact bonded temporarily to an optically flat glass and the exposed side of the Si wafer is then epoxy bonded to a supporting glass substrate. After curing the epoxy, the temporary optical flat is then removed and a Si surface as flat as the optical flat is exposed. Consequently, any nonuniformity in the thickness of the Si substrate is embedded in the epoxy. Unfortunately, individual conductive channels from superpixels to the outside processor array are inevitably blocked by the presence of this supporting glass.

The necessary through-glass contact can be accomplished by using a special glass substrate with high-density conductive feedthroughs. Conductive feedthroughs can be made by first forming an array of holes in the glass and then filled those through-substrate holes with a silver fret glass. The design concept of the freedthrough packaging is illustrated in Figure 4-7. The silver fret is initially in a paste form and mechanically pushed through the holes. It is then fired at about 300°C to form a solid conductive glass feedthrough. The holes in quartz substrate can be made by laser drilling using a high power CO₂ laser. Holes as small as 8-10 mils in diameter, which are within the predetermined 448 μm superpixel size, can be made using this technique. It is also possible to use fiber optic capillary arrays to form necessary through-substrate holes. The making of conductive feedthroughs significantly complicates the device manufacturing process.

A simpler approach is to treat the CCD-based STC as a single 3-D wafer, thus it can communicate with the 3-D computer through microbridges or metal bump technologies. This

approach will be simpler to implement, but it lacks a precise control in the thickness uniformity of the liquid crystal layer. Although this approach may suffer nonuniformity in output light modulation, it is well suited for binary data modulation. And this is the approach we adopted for the program.

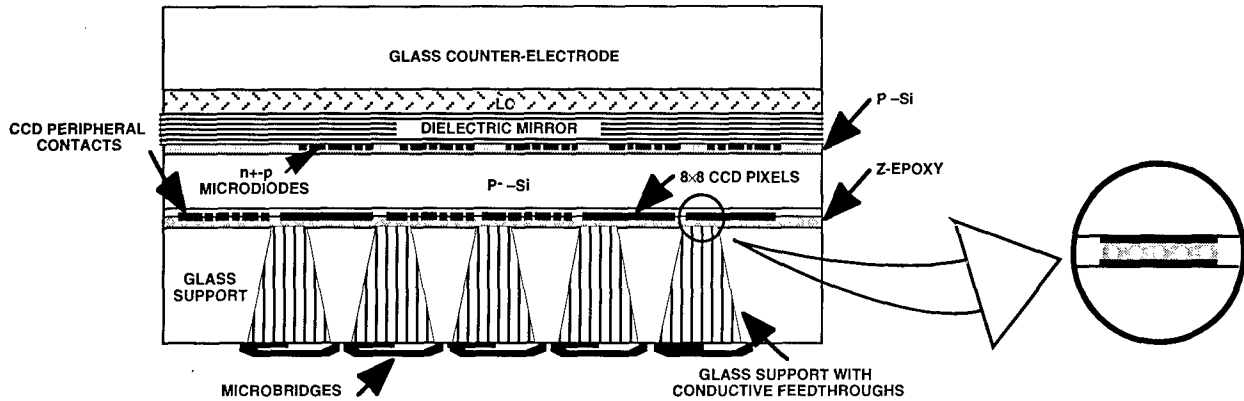


Figure 4-7. Cross section of the STC with conductive feedthroughs in the supporting glass.

4.3.1 Z-axis Chip Bonding

To contact between the processor array and the corresponding pads of the STC superpixels, we used a special adhesive that is conductive only in one direction. As a result of embedded gold-coated particles that are compressed along this direction during curing, this adhesive, which is referred to as the z-axis adhesive, conducts only in the direction along its thickness.⁽⁴⁻⁴⁾ There is no conduction in lateral directions due to the low density of these particles.

To facilitate the Z-axis bonding of the CCD chip to the substrate, a special modified probe station was developed. The probe card was replaced by a special clamping plate that was made for holding the substrate while aligning and gluing it to the CCD chip. This required precise alignment while applying high pressure and access for UV curing. Test substrates were fabricated with a fanout signal line pattern to test the Z-axis gluing process and measure the electrical characteristics.

Experiments were conducted using the Z-axis adhesive to provide contact between the CCD chip and the substrate. A mock-CCD chip was made that had the correct bonding pad layout but had all pads shorted together on a common metal layer. The substrate had a fanout pattern with one line to each bonding pad. If proper contact was made between the substrate lines and the bonding pads, all the substrate lines should be shorted together through the CCD chip metalization. Continuity testing showed most of the lines to be shorted together, which was the expected result, but it did not rule out lateral contact between lines. A second experiment was conducted using a

plain silicon chip with no shorting layer. In this case all lines measured open, indicating that no lateral shorting was occurring.

Wafer flatness and pad alignment are also an issue concerning the Z-axis bonding process. The first two chips bonded had very poor flatness and a possible wedge across the surface. The probe station used for aligning during the bonding operation was redesigned using a flexible rubber pad under the stage that was also heated. This allowed the CCD chip to conform to the substrate even if they were not parallel. The heat lowered the viscosity of the Z-axis epoxy and allowed it to flow easier, creating a thinner bonding layer and improving contact reliability.

Test results using plain silicon chips showed an improvement in flatness using the new stage. The surface of the CCD chip was still not optically flat and it was possible to see an imprint from the metal lines of the substrate coming through the silicon. It was also possible that the conductive balls in the adhesive bunched up in areas and caused bulges.

A new mask pattern was developed to test for both Z-axis conduction and lateral insulation by bridging every other bonding pad. The first attempt with the new mask showed open contacts in the center area of the chip. It was assumed that the problem was probably related to particles on the chip surface that were larger than the conductive balls in the Z-axis adhesive. A microscopic examination of remaining chips showed some particles large enough to cause the problem. It was concluded that the problems with flatness and bad contacts were probably the result of particle contamination during the assembly process. Removing these particles proved to be very difficult and frequently had to be done manually using a probe or "hair stick." The bonding operation was subsequently moved to the assembly area of a clean room environment and careful cleaning of each chip was done before bonding. This effort finally resulted in a chip that had all contacts intact and reasonable flatness.

There was still concern as to the number and distribution of the gold contact balls between the contact pads. To observe this distribution it was decided to use a substrate coated with indium-tin oxide instead of metal. This would be transparent and would allow the distribution of the gold contact balls to be visible. Continuity could still be measured to verify contact integrity. New substrates were fabricated and tested for this experiment. Successful contacts were made and about 12 balls could be seen in the contact areas (as shown in Figure 4-8). This increased the confidence level of the Z-axis adhesive and the process was ready to be tried on an actual CCD chip instead of a test device.

Curve tracer testing was done on a CCD chip that had been successfully bonded to the substrate using the Z-axis adhesive. The contact to the superpixels appeared to be good, but most of the drive signals showed a 200 ohm short to the CCD substrate. No cause could be found in the bonding process so the possibility of a chip defect had to be considered. An unbonded chip from the same wafer was tested and did not exhibit the same problem. It was decided that wafer

testing should be conducted before assembly, at the risk of scratching the bonding pads with the probe. Two new wafer quadrants were tested and no major defects were found. Neither wafer exhibited good dynamic characteristics, however, so they were not considered as good candidates for Z-axis bonding.

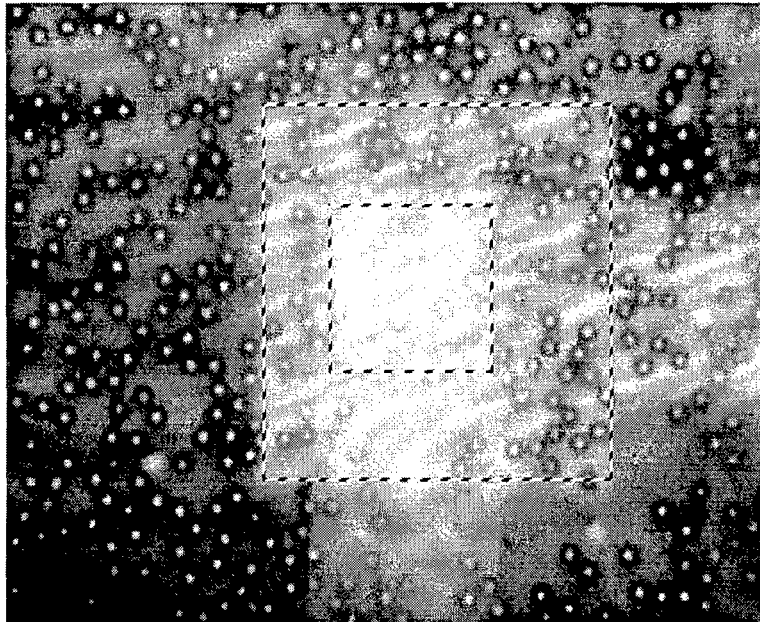


Figure 4-8. Typical microsphere distribution of Z-axis adhesive under the contact pad on the CCD chip.

The probe station used for aligning and gluing the CCD chips showed a problem with the CCD chips moving during the bonding process. The stage did not have a mechanism for holding the chip in place. The conventional vacuum chuck would pull the chip down in the center and distort it since the wafer had been thinned. Some micro-channel plates were obtained from Litton and used as a flat porous surface over the vacuum chuck to keep the chip flat. A sheet of special porous paper was used over the channel plate to prevent scratching the back of the chip. The first CCD chip was bonded with this technique.

The results were still inconsistent regarding contact reliability. Only a small percentage of the pads made contact, usually in certain areas near the edges. The cause of the problem had not been identified but was probably related to flatness variations over the surface. The use of the vacuum with the microchannel plate precluded the use of the flexible rubber pressure block that was used previously. The rubber block allowed the glass substrate to tilt slightly and conform to the chip if they were non-parallel. This may account for the decline in yield from the earlier test devices to the current chips.

Further tests were done using the Z-axis adhesive on working CCD chips to see if the process was affecting the performance of the CCD. One chip was successfully bonded with nearly all pads making contact. Curve tracer testing showed proper characteristics for the drive signals and superpixels, indicating that no damage had occurred as the result of the Z-axis bonding. It was finally possible to get reliable contacts on good CCD chips.

4.4 SPACE-TIME COMPANDER OPTICAL INTERFACES

Limited by the physical presence of microbridges in the 3-D computers and the nature of the CCD structures, STC superpixels must be physically separated from each other with a fairly large gap. This prevents us from using a simple magnify/demagnify optical system to bridge between fine-grain images and the coarse-grain processors. The optical system in the STC must be capable of grouping superpixels from continuous fine-grain images and matching those superpixels to isolated coarse-grain processors. This function may be achieved with appropriate lens arrays. The lens arrays were used to group the superpixels and physically separate them from a continuous image. The concept is illustrated in Figure 4-9. Each lens in the array will image a portion of the fine-grain image to the corresponding CCD sub-array. The design constraint is to prevent multiple images from neighboring lenses overlapped on a single CCD imager. On the other hand, the images are allowed to be overlapped in the area between the CCD arrays. Thus, a practical match can be achieved by balancing the image magnification ratio (MR) and the coverage ratio (CR). Here the coverage ratio is defined by the ratio of the size of input image "viewed" by a single lens to the size of the supercell. The lens design must follow the constraint:

$$CR \leq \frac{2 - MR}{MR} \quad (4-1)$$

here the magnification ratio is less than one; demagnification is required for the operation. Once the lens array satisfies the imager requirement, it will automatically meet the need for the modulator function. This is based on the fact that there is no information in the adjacent area between the CCD arrays.

We also looked into the practical issues of implementing optical compander for interfacing the 3-D computers with various optical memory systems. Again as discussed previously for the combined CCD modulator/imager approach (Section 4.1), the wavelengths for the write and read operation must be fairly different to ensure proper STC functions. Since the dielectric mirror can be tuned for close read/write wavelengths, the combined modulator/imager approach can properly interface with high density optical data storage.⁽⁴⁻¹⁾

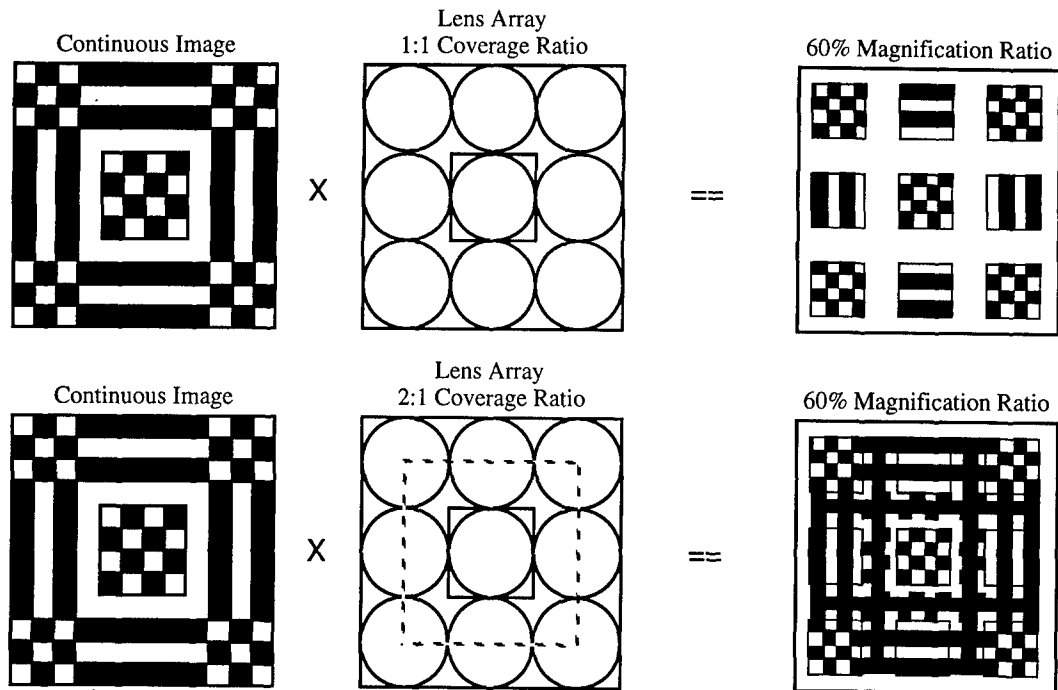


Figure 4-9. The concept of using microlens array for mapping optical images to physically separated CCD arrays. The ideal case is to demagnify the necessary area in the optical image to the CCD array. Image overlapping is allowed if the lens imaging constraint is followed.

Section 5

EOCA SYSTEM ANALYSIS

The goal of the system-analysis study is to identify and analyze all the architectural implications the Optical Transpose Interconnection System (OTIS) based architecture for locally connected parallel electronic processors. The Optical Transpose Interconnection System provides an efficient means of providing a particular set of global interconnections using optical communications. The use of free-space optics allows efficient, high-density, high-speed long distance communication. Using these interconnections, small groups with only local connections can simulate networks with powerful global connectivity. In this study, we demonstrated that a combination of optical and electrical technologies can simultaneously take the advantages of optics connectivity as well as the routability of electronics.

By using the optical transpose global interconnections, large systems can be build up from smaller subsystems. The OTIS-Mesh connects groups with 2-D mesh topology into a 4-D mesh. This technique can reduce wire lengths in large systems. By connecting groups with hypercube topologies with the OTIS, one produces the OTIS-Hypercube, which is capable of simulating a hypercube connecting all nodes in all groups with fewer wires. This technique also allows large expander graphs to be constructed in a scalable fashion by connecting many copies of a small expander into the OTIS-Expander network. The same techniques used to construct these scalable expanders can be used to construct large splitter graphs, for use in multibutterfly routing networks.

We have also been developing models for both electronic and free-space optical interconnect technologies that will allow us to regions of superiority between electrical and optical interconnects, and determine some of important trade-offs. . In this report, we compare the speed performance and energy cost of a class of electrical and optical interconnections for use in large-scale digital computing systems. On-chip, off-chip, and free-space digital interconnections based on Multiple-Quantum-Well (MQW) or Vertical Cavity Surface Emitting Lasers (VCSEL) have been evaluated. The study shows that free-space optical interconnects using MQW modulators or VCSELs as transmitters offer a significant speed advantage over both off-chip and wafer-scale on-chip electrical interconnects.

The cost of a computer architecture depends on the costs and yields of the underlying technology along with the systems' physical organization. To compare architectures constructed with optics and electronics, we first builded cost models for the underlying technology for active devices: CMOS, MQW modulators and VCSELs. As an example, we found that a multichip 256 element VCSEL array is lower cost than the monolithic alternative. For MQW modulators, the

cost models predict that the flip chip hybrid CMOS-SEED technology is lower cost than the monolithic FET-SEED integration. Then the active device models are combined with models for passive elements such as MCMs and optomechanics to produce process flows for complex systems. The cost models also contain architectural features such as the shuffle-exchange wire layout area, the number of parallel channels, and the number of computational nodes. Using these models, we show that an optoelectronic VCSEL/CMOS/MCM/Optics implementation of a shuffle-exchange network is lower cost than the all-electronic CMOS/MCM for greater than 20 nodes.

5.1 OTIS ARCHITECTURE STUDIES

In principle, optical interconnect technologies offer several advantages over electrical systems. Connections can be made at higher speeds with less crosstalk and less power consumption than electrical channels. The power required is nearly independent of the length of the connection, at least over the lengths of connections involved within a parallel machine. While some routing of optical links is possible using lenses and computer-generated holograms (CGH), arbitrary connections are more difficult to implement as space-variant optics than as wires on a VLSI circuit, multi-chip module or printed circuit board. In this study report, we attempt to demonstrate how a combination of optical and electrical technologies can achieve many of the advantages of optics as well as the routability of electronics.

The studies focus on fine-grained, massively parallel systems, consisting of many chips with many processing elements (PEs) per chip. Using the optical transpose global interconnections, we show how large systems can be built from smaller subsystems. These smaller pieces could then be implemented on individual chips, with connections made with on-chip electrical lines. Beginning with several copies of simple 2-D mesh topology, we show that the addition of the OTIS connections actually help create a more powerful 4-D mesh topology. In a similar fashion, the optical connections allow many small hypercube networks to simulate a large hypercube. The same techniques can be used to construct large expanders, randomly wired graphs useful in many routing applications. Normally, the random connections of these graphs would prohibit large-scale implementations. Finally, the same techniques can be applied to construct crossbars which switch entire words in parallel by connecting bit-serial crossbar chips using the optical transpose.

5.1.1 OPTICAL TRANSPOSE INTERCONNECTION SYSTEM

The Optical Transpose Interconnection System (OTIS) is an optoelectronic Multistage Interconnection Network (MIN) developed for parallel processing systems.⁽⁵⁻¹⁾ In an OTIS based free-space optoelectronic MIN, electronic bypass-and-exchange switches are required to perform the local routing. It has been shown that for an optoelectronic MIN with N^2 inputs and N^2

outputs, the bandwidth and the power consumption of the network are optimized if the electronic switch planes are partitioned into N switches.⁽⁵⁻²⁾ Thus, in an OTIS-based parallel system, N^2 processor nodes are logically divided into groups of N nodes each. In practice, these groups can be thought of as being implemented by a single chip, or perhaps a small number of densely-connected chips. Connections between groups are achieved via free-space optics: each processor node has an optical transmitter/receiver pair with which it sends and receives optical signals. Transmitters and receivers are connected via two planes of lens arrays each consisting of N lenses (See Fig. 5-1). These optical links connect the p^{th} processor of the g^{th} group to the g^{th} processor of the p^{th} group: a transpose of group and position coordinates. These optical connections also can be thought of as higher-order generalizations of the shuffle edges in the perfect shuffle (See Fig. 5-2).

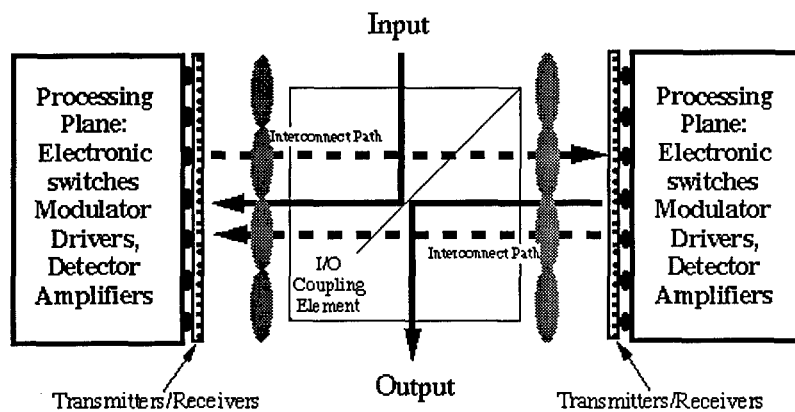


Figure 5-1. The OTIS System.

The OTIS implementation offers several advantages for large scale interconnection networks. In general, optical technology offer the advantage that the bandwidth and power requirement of a link are independent of the length of the link. For example, it has been shown^(5-3,5-4,5-5) that for line lengths greater than a few millimeters free-space optical interconnections require less power for a given bandwidth. In addition, free space optoelectronic systems also facilitate electronic layout due to the fact that the I/O is performed over the 2-dimensional area of the chip and is not limited to the 1-dimensional boundary of the chip.

The OTIS topology offers some additional advantages: the system can be folded to establish connections from a chip onto itself, it can be made bi-directional to provide interconnections between two processing planes, or it can be cascaded to accommodate successive processing planes. It can also accommodate any arbitrary optoelectronic layout as long as the layout of all the groups are identical. Finally, the optics in OTIS can be designed to allow bit-serial and/or bit-parallel communications between node (g, p) and node (p, g) in the system.

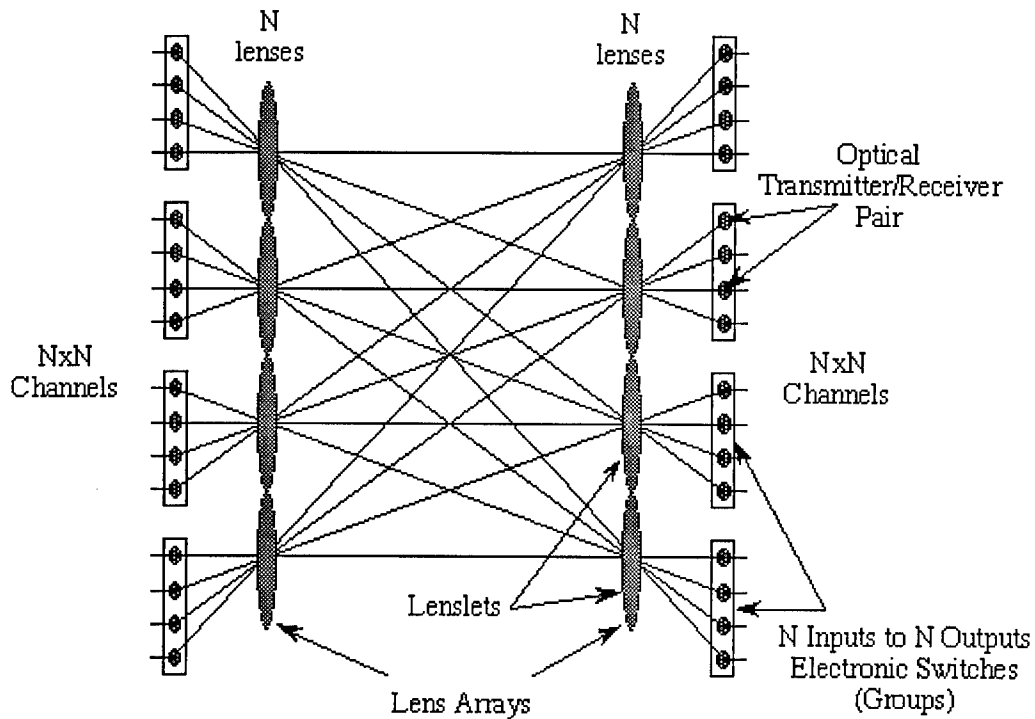


Figure 5-2. OTIS Interconnections.

OTIS relies on optoelectronic flip-chip bonding technology and free-space optics. The integration of 8,000 optical transmitters and detectors on a single 0.8 micron silicon chip using flip-chip bonding has been demonstrated.⁽⁵⁻⁶⁾ A simple free-space optical system for which 4096 bi-directional channel connections has also been demonstrated.⁽⁵⁻¹⁾

Extensive modeling of an OTIS based switching system using the OTIS-Hypercube topology has also been performed.⁽⁵⁻⁷⁾ Performance of the system in terms of throughput and cost in terms of system power consumption, area, volume, and maximum power dissipation per unit area have been computed (see Table 5-1). The modeling includes the VLSI switches, the optoelectronic receivers and transmitters, the optical interconnection system, and the main laser required to power-up the modulators and its associated optics. Note that in this modeling, it is assumed that the VLSI switches contain circuits to detect hot spots and allow the OTIS network to resend data packets that have been dropped due to contention.⁽⁵⁻⁸⁾ This modeling also assumes a single switch plane and the required optics to fold the interconnections back onto the chips.

Table 5-1. OTIS Modeling.

Total Throughput	1 Tbits/sec
Total Power Consumption	55 W
Optical Power at the plug	10 W
Electrical Switch Power	& 10 W
Total Silicon Area	8.8 cm ²
Longest electrical wire	2.2 mm
Power/channel	13 mW
Area/channel	460x460 microns
Power Density	5.6 W/cm ²

The modeling results of the OTIS-based switching system are very encouraging in terms of the feasibility of a large scale implementation (4096 channels) of the system. Although the required silicon area is quite large (8.8 cm²), it can be tiled into smaller chips since the longest wire is only 2.2 mm long which makes Multi Chip Module (MCM) implementation relatively easy. In addition, a total power requirement of 55 W is low for such a large system and the power density projections (below 10 W/cm²) remain within the limit of air cooling. If packaging issues related to integrating free-space optics with optoelectronic chips can be resolved at a reasonable cost, this system would prove competitive with electronic alternatives.

5.1.2 Terminology

In order to keep the notation simple, N refers to the size of a group in OTIS (i.e., OTIS networks have size N²). Also, OTIS processors will be referred to by pairs (g, p), where g represents the group the processor belongs to and p represents the position of the processor within the group. The basic topologies used (mesh, hypercube, expander) refer to the connections within each group. The optical transpose links, which connect nodes (g, p) and (p, g), provide the only connections between nodes in different groups. We also make a few additional assumptions for carrying out the emulations. All links are assumed to be bi-directional. The mesh and hypercube networks are assumed to run in Single Instruction Multiple Data (SIMD) fashion; that is, each node is allowed to send along only one of its edges at any time, and this choice must be uniform for all processors. Since such a definition would not make sense for the OTIS-expander, we assume a stronger property that each node can process a message per edge. If this assumption is made in general, the other two networks can be used in Multiple Instruction Multiple Data (MIMD) fashion. If the optical links run at the same speed as the electronic links, this MIMD simulation will be slowed by contention for the optical links. If

the optical links can run D -times faster for a degree D network, no additional slowdown will be incurred. Faster optical communication speeds are certainly supported by the current and projected technologies; the difference between the two speeds in an integrated system is more difficult to ascertain.

5.1.3 OTIS-MESH

Consider each group of OTIS has having a very simply 2-D mesh topology: each PE is connected to PEs which lie to the north, south, east, and west of it within the same group. Since each PE has degree at most four, and every connection is short, meshes are simple and can be implemented with a single level of electronic wiring. However, for the same reasons, the mesh topology is quite weak, having large diameter and (relatively) small bisection width. The OTIS-Mesh architecture consists of N groups, each of which is an N -node (i.e., $\sqrt{N} \times \sqrt{N}$) 2D-mesh. The OTIS interconnections provide communication between groups, connecting processors (g, p) and (p, g) for all $1 \leq p, g \leq N$.

Theorem 1: *OTIS-Mesh can simulate a 4-dimensional $(\sqrt{N} \times \sqrt{N} \times \sqrt{N} \times \sqrt{N})$ mesh with a slowdown of at most a factor of 3.*

Proof: For each PE on the OTIS-Mesh, interpret its address (g, p) as (g_x, g_y, p_x, p_y) by dividing the bits representing the group and position into two equal pieces. With this interpretation, the mesh within the group g connects (g_x, g_y, p_x, p_y) to the four PEs $(g_x, g_y, p_x \pm 1, p_y)$ and $(g_x, g_y, p_x, p_y \pm 1)$, except at the boundaries. Now, each of $g_x, g_y, p_x,$ and p_y will be a coordinate of the 4-D mesh address. Using the mesh connections within each group as described above, we can simulate the two last dimensions of the 4-D mesh. To simulate communication across the remaining two dimensions, we will need three steps. First, send the data across the optical transpose links of OTIS. The information initially stored in PE (g_x, g_y, p_x, p_y) is now stored in PE (p_x, p_y, g_x, g_y) . Then, using the mesh connections, this data can be moved to PE $(p_x, p_y, g_x \pm 1, g_y)$ or $(p_x, p_y, g_x, g_y \pm 1)$, depending on the intended destination. Finally, use the optical transpose again, bringing the data to $(g_x \pm 1, g_y, p_x, p_y)$ or $(g_x, g_y \pm 1, p_x, p_y)$, which is precisely the desired connectivity.

Figure 5-3 shows how connections across three of the four dimensions is accomplished. Node (g_x, g_y, p_x, p_y) communicates with its four neighbors within the same group using the mesh connections within the group. To communicate with $(g_x, g_y + 1, p_x, p_y)$ or $(g_x, g_y - 1, p_x, p_y)$, the transpose connections are used to simulate the higher-dimensional links of the 4-D mesh.

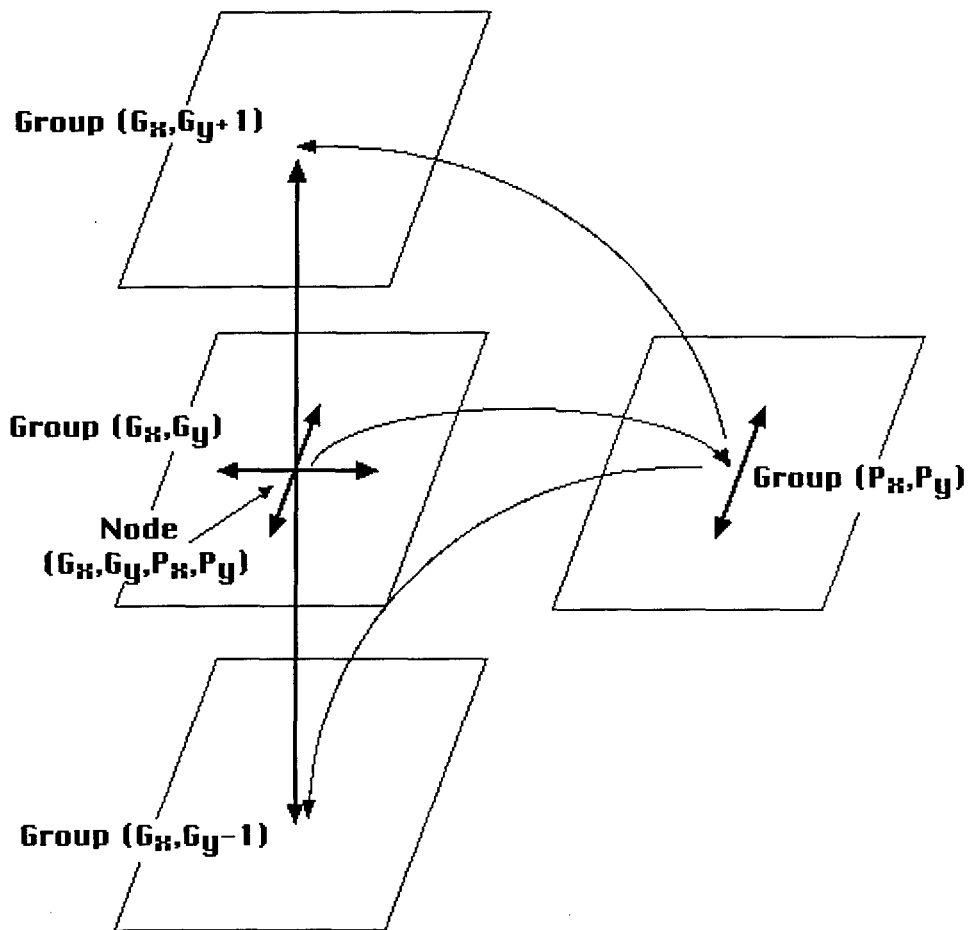


Figure 5-3. OTIS-Mesh Simulation.

This simulation allows more efficient solutions to problems which have faster algorithms on higher-dimensional meshes, like routing and sorting. However, it can also be viewed as a way to minimize wire lengths in large systems. By using a folded OTIS system, any two points on different chips can be connected with only wires across two chips. Using the same ideas as in the proof above, a signal is routed across the first chip to the location of the optical link to the destination chip. Then the signal is sent across the optical link, and routed across the destination chip to the desired location. With this technique, a long off-chip line across a PCB or MCM is replaced by an optical link plus on-chip wires to route the signal from the source to the transmitter and from the receiver to the destination.

5.1.4 OTIS-Hypercube

The hypercube or N-cube is a versatile network for multiprocessor architectures. The hypercube architecture can simulate many other important topologies, such as meshes, meshes-of-trees, butterflies, and even PRAMs. Also many problems such as sorting and routing have efficient hypercube algorithms.⁽⁵⁻⁹⁾ Several practical implementations of the hypercube are

available as commercial products. However, it is difficult to construct larger-dimensional hypercubes using electronic technology. Since the degree of a node is $\log N$, and not constant, the number of wires leaving each chip or group of chips must grow as the size of the network increases.

The OTIS-Hypercube consists of small hypercubes linked by an interconnection pattern which is significantly sparser than that of a hypercube. Each node in an N^2 -node OTIS has degree $(\log N) + 1$: $\log N$ connections to other nodes in its group and one optical link. An N^2 -node hypercube would have degree $2 \log N$. In a manner similar to the shuffle-exchange graph, OTIS can be used to simulate the connectivity of a full hypercube. The OTIS-Hypercube network is also closely related to the hierarchical cubic network (HCN) proposed by Ghose and Desai.⁽⁵⁻¹⁰⁾ The OTIS-Hypercube lacks the 'diameter' links of the HCN, but this does not affect its ability to simulate the hypercube.

Let $n = \log N$. Label the OTIS node (g, p) by $g_1 \dots g_n p_1 \dots p_n$, where $g_1 \dots g_n$ is the binary representation of g and $p_1 \dots p_n$ is the binary representation of p . We will show how to connect (g, p) to every node which differs from it in exactly one bit position. If the bit position in question lies in the second half (i.e., the bits corresponding to the position within the group), then there is direct connection since all the processors $(g, *)$ form a hypercube group. If the bit position lies in the first half (i.e., the bits of the group address), then we first perform an optical transpose. This will interchange the bits of the group and the bits of the position. Now, if the bit to be changed lies in the second half, so by the same argument as before, the nodes to be connected are now linked by an electrical wire. Finally, another optical transpose restores the nodes to their original position.

An example showing that this allows routing to all the $\log N$ hypercube neighbors is given in Fig. 5-4. More formally:

Theorem 2: *An N^2 -node OTIS-Hypercube network can simulate an N^2 -node hypercube with a slowdown factor of at most 3.*

Proof: Label the OTIS nodes by $2n$ -bit strings $x_1 \dots x_{2n}$ as above. The hypercube edges we wish to simulate are of the form:

$$x_1 \dots x_i \dots x_{2n} \rightarrow x_1 \dots \bar{x}_i \dots x_{2n}$$

For $i > n$, these are the connections provided by the electrical hypercubes.

For $i \leq n$, we use the following routing path:

- From $x_1 \dots x_i \dots x_{2n}$
- to $x_{n+1} \dots x_{2n} x_1 \dots x_i \dots x_n$ via optical transpose
- to $x_{n+1} \dots x_{2n} x_1 \dots \bar{x}_i \dots x_n$ via hypercube edges
- to $x_1 \dots \bar{x}_i \dots x_{2n}$ via optical transpose.

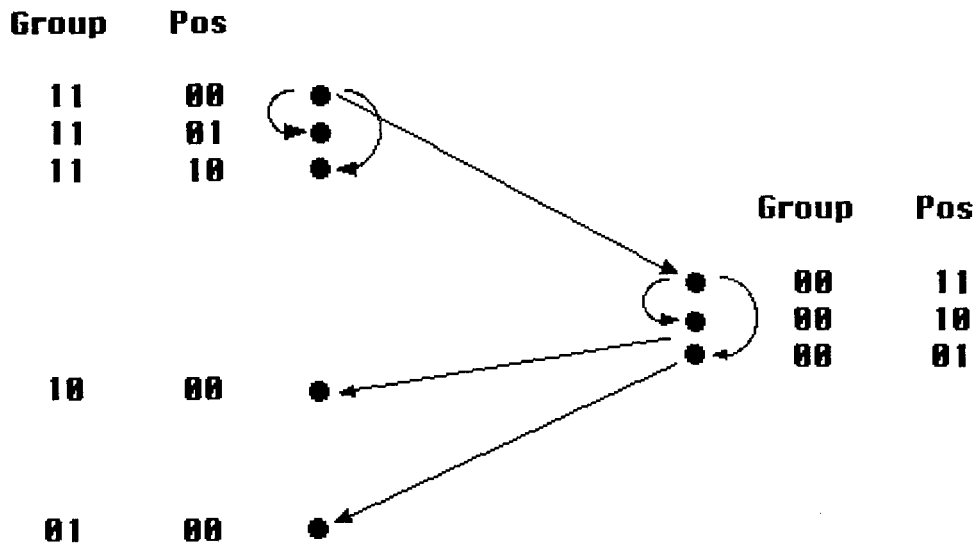


Figure 5-4. OTIS-Hypercube Simulation.

This factor of three slowdown in the proof is needed to ensure that the right 'parity' is preserved; that is, that the group and position coordinates have returned to their original roles. In some cases, this is not necessary, and the simulation can run correspondingly faster. In particular, the optical transpose is necessary only when the communication switches between the low (1, ..., n) and the high (n+1, ..., 2n) dimensional edges of the hypercube. Thus, problems in which this happens seldom, like routing, can be done quite efficiently. For example, butterfly routing (subject to blocking) using an OTIS system with two processing planes experiences some additional latency (compared to routing on a full hypercube), but no additional loss of throughput, by the use of pipelining. The first log N coordinates are routed exactly as before. Then the optical transpose takes place (which is much simpler and faster than the routing stage), while a new set of inputs are sent to the first plane. The last log N coordinates are then routed on the second plane and output, while the new inputs are routed by the first plane and transposed. The technique can also be applied to algorithms which make more than one pass over the edges of the hypercube. For example, sorting using Batcher's sorting network⁽⁵⁻¹¹⁾ requires $\frac{1}{2} \log^2 N$ steps. However, only 2log N switches between high and low dimensional edges are required, so the overhead in simulating this algorithm on an OTIS-Hypercube is relatively small. Finally, using known network emulations (see Ref. 5-9 for details), such a network can be used to simulate meshes of any even dimension, meshes-of-trees, and butterflies.

5.1.5 OTIS-Expander

In a similar fashion to the OTIS-Mesh and OTIS-Hypercube constructions, we will show how to construction large randomly-wired graphs known as expanders in a hierarchical fashion.

This will make the construction of large expanders more feasible; normally the random wiring of expanders does not scale as the size of the graph becomes large.

Unlike the mesh and the hypercube, expander graphs do not refer to an explicit set of connections. Rather, they refer to any graph which has the property that any for any set of nodes with $|S| \leq \alpha N$, the neighborhood of S has size at least $c|S|$, for some constant $c > 1$. An expander graph G with N nodes is one with constant degree, d , which has the property just described. We will use the notation (N, α, c) -expander to denote such a graph. The constant-degree restriction is intended to capture some notion of efficiency; without it, the complete graph on graph on N nodes would be an $(N, \alpha, 1/\alpha)$ -expander. While there are explicit constructions of graphs of this kind, the graphs with the strongest form of this property (i.e., large values of c) are constructed by choosing graphs at random. For large enough values of N , one can show that a random choice of G is likely to have c close to d , for values of α satisfying $c\alpha < 1$.

The OTIS-Expander is constructed from N identical copies of a fixed (N, α, c) expander, which are connected to one another using the transpose connections of OTIS. With this definition, we can show:

Theorem 3: *Let G be an N^2 OTIS-Expander constructed from N copies of an (N, α, c) expander. Then G can simulate an $(N^2, \alpha^2, c/2)$ expander with a slowdown of a factor of two.*

Proof: To show that a graph has expansion, we need to show that any sufficiently small set of nodes expands to a much larger set by following the edges of the graph. The intuition behind this theorem is that expansion will happen as long as the sets involved are not too large, and that the sets before and after the optical transpose cannot both be large.

Let S be a set of nodes of size at most $\alpha^2 N^2$. We will divide the nodes into two classes: those which begin in 'big' groups, (groups with more than αN elements of S) and those which begin in 'small' groups (groups with at most αN such elements). Since every node is in exactly one of these categories, one category contains at least half the nodes in S . The proof considers these two cases, illustrated in Fig. 5-5 and Fig. 5-5, individually.

If the 'small' groups have more nodes, ignore all nodes in large groups. By doing so, we only decrease the size of the set of neighbors of S and underestimate the expansion of the graph, because the size of the neighborhood of S is a monotonic function of the nodes of S . Consider the nodes of S group by group, and let S_i denote the number of nodes of S in group i , not counting the nodes in large groups that we ignored. Because we only consider nodes in small groups, for all i , $|S_i|$ is at most αN . Therefore, the neighborhood of each S_i has size at least $c|S_i|$, and the neighborhood of S has size at least $\sum |S_i|$. Since at least half of the nodes in S were in small groups, $\sum |S_i| \geq |S|/2$. This implies that the neighborhood of S has size at least $c|S|/2$, and thus the graph has expansion at least $c/2$.

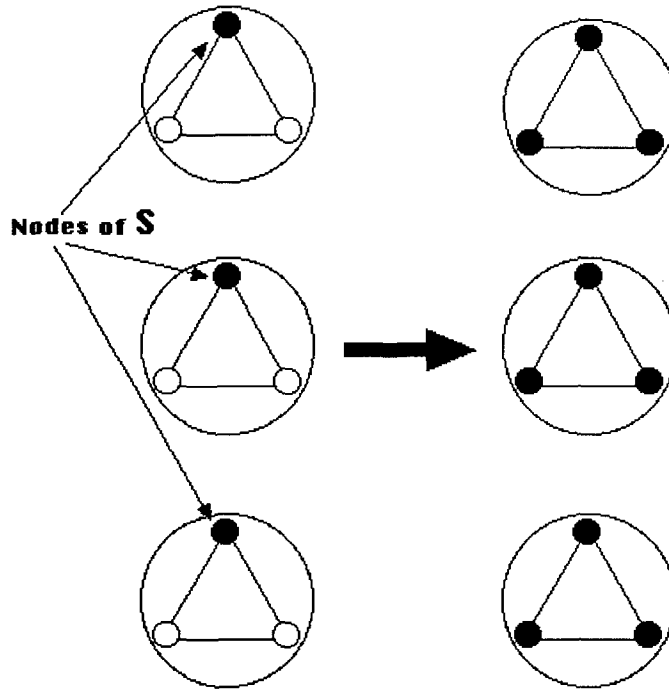


Figure 5-5. OTIS-Expander - Small Groups Case.

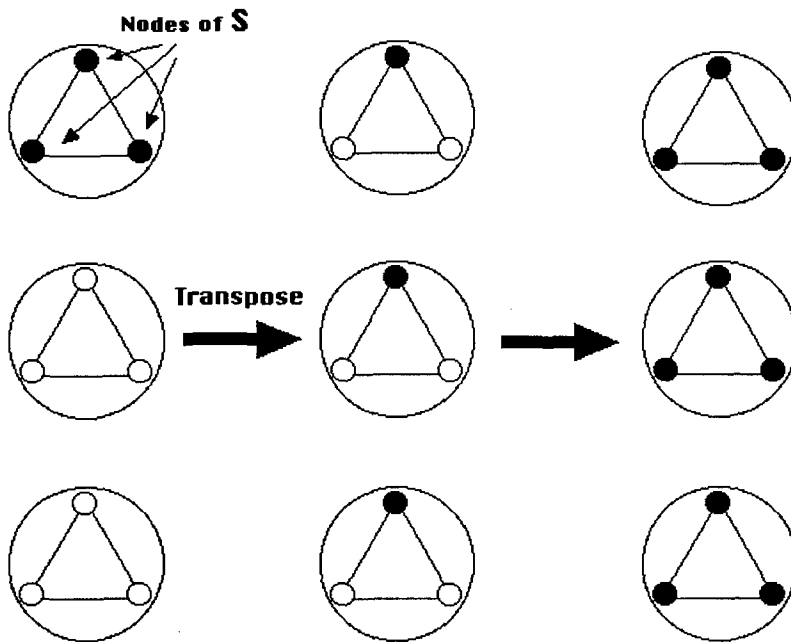


Figure 5-6. OTIS-Expander - Large Groups Case.

If the 'large' groups have more nodes, ignore all nodes in small groups. There are at most αN large groups, because each group we did not ignore has at least αN nodes of S , and S has size at most $\alpha^2 N^2$. From each such node, follow the optical link. The OTIS connections provide only one link between each pair of groups, so each group can only receive at most one node from each other group. However, since before the transpose there were at most αN groups containing nodes, during the transpose each group receives at most one node from each non-empty group, or at most αN nodes. Then, we can a similar argument to the one used in the small groups case above: each groups has at most αN nodes, so within each group, the neighborhood is larger by a factor of at least c . Since the number of nodes we did not ignore is at least $|S|/2$, the size of the neighborhood across all the groups has size at least $c|S|/2$, and the graph has expansion at least $c/2$.

5.1.6 Scalable Multibutterfly Construction

It may seem counterintuitive that random or near-random wiring could be helpful, there are several results which demonstrate that randomness or expander graphs are useful in many types of routing or sorting problems. For example, Valiant and Brebner demonstrated that choosing random intermediate destinations prevent worst-case behavior in butterfly routing. Instead of using randomness on-line, results involving expanders utilize the random-like connections of the network to obtain good worst-case performance from deterministic routing algorithms. The AKS sorting network⁽⁵⁻¹²⁾ used expander graphs to demonstrate that sorting could be done in parallel in $c \log N$ steps, for a sufficiently large value of c . More practical work on using expanders for routing has centered on the *multibutterfly* network studied in [5-13] and [5-14]. In a butterfly network, the each bit of the destination address is used to divide the packets into two classes, and so each node has an up wire and a down wire to nodes in the next stage corresponding to these two possibilities. These connections are made in a regular fashion: at the i th stage node $x_1 \dots x_n$ is connected to the same node of the next stage as well as node $x_1 \dots \bar{x}_i \dots x_n$. In a multibutterfly, the address bits are used to partition packets into two classes, and each node has wires to each class in the next stage. However, each node will have several such connections, and they will not form a regular pattern as in the butterfly. At level i , the nodes can be naturally divided into 2^i partitions based on the address bits followed to that point. In a multibutterfly, the bipartite graphs formed between each partition and either of the partitions connected to it in the next level are expander graphs. Bipartite graphs with this property are called *splitters* because of their applications to routing. This two-way expansion property guarantees that in order to cause a few nodes at level i to be blocked, many nodes at level $i+1$ would need to be blocked. By using this reasoning level-by-level, one can show that it is difficult to cause the inputs to become blocked, even in the presence of faults.

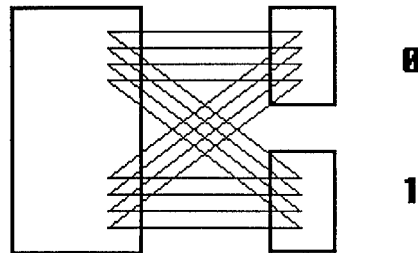
However, circuits which provide the connectivity of expander graphs are quite difficult to implement. While explicit constructions of expander graphs are known, they produce small values of c relative to d , which limits their usefulness. As mentioned above, random graphs are likely to have good expansion properties. However, if we attempt to partition a large random graph across many chips, we expect that nearly every edge of the graph will connect nodes on different chips. As the size of the required expander becomes larger, each chip requires more pins. The OTIS-Expander addresses this problem by producing an expander which can be partitioned into smaller pieces with manageable communications between the pieces.

Graphs of this type, referred to as hierarchical expanders, were first considered in [5-15]. There, the authors show how one can build large expander graphs without increasing the number of different wires required. Essentially, each chip or module has a small number of cables, and each cable is made 'thicker' as the number of nodes is increased. Here, however, the optical transpose provides wires from every group of nodes to every other group of nodes without the explosion in wiring complexity that the construction in [5-15] sought to avoid. At the same time, that construction relies on many different, independent random choices of wiring. For the systems of boards connected by cables that the authors envisioned, this can be accomplished simply by connecting cables to boards in a random fashion. However, at the finer scale of parallelism we envision, this would require the fabrication of many different chips, each with its own random wiring, greatly increasing the cost of such a system. However, the OTIS-Expander allows hierarchical expanders to be constructed using many copies of the same randomly-wired chip.

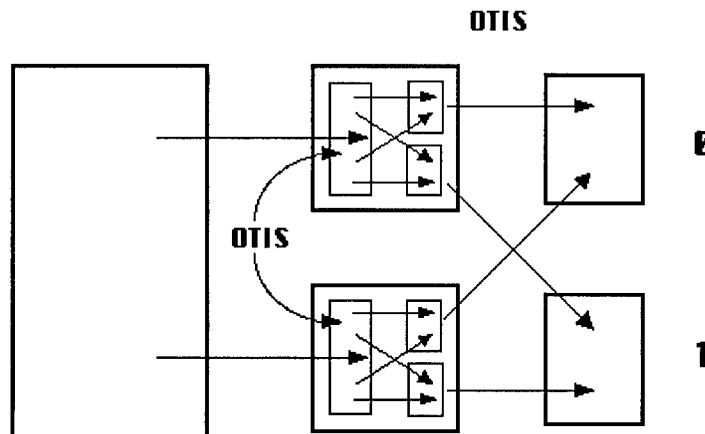
To construct a multi-butterfly using the OTIS-Expander, we use a similar construction. Here we describe the construction of the first stage, later stages use similar designs on smaller scales. As in the OTIS-Expander, each group is a copy of a single expander graph. However, each group will have N inputs, $N/2$ up outputs, and $N/2$ down outputs. The connections between the inputs and the outputs form a splitter; that is, any set S of at most αN inputs is connected to at least $c |S|$ up outputs and $c |S|$ down outputs. We refer to such a graph as an $(N, \alpha, c/2)$ -splitter. Obviously, this implies that $c \alpha$ must be smaller than $1/2$. The OTIS connections will be used to connect the input nodes of different groups, rather than to connect input nodes to output nodes. That is, the p^{th} input of the g^{th} group is connected to the g^{th} input of the p^{th} group via OTIS.

In order to show that this network can simulate a multibutterfly, it is necessary to show that any small set S of input nodes is connected to at least $c' |S|$ output nodes, for some constant $c' > 1$. The simulation of a stage of the multibutterfly will require three steps. In the first step, each input node communicates with every output node to which it is connected, and also to the node to which it is connected by the OTIS transpose link. In the second step, every input which received a message from the optical link in step one communicates that message to the outputs to

which it is connected. Finally, in the third step, the packets at the output nodes are sent via OTIS connections to the next stage. This simulation, shown together with the butterfly stage it is related to, is illustrated in Fig. 5-7.



Butterfly Stage



OTIS Multibutterfly Stage

Figure 5-7. Multibutterfly Construction.

Theorem 4: *Let G be an OTIS-Expander constructed from N copies of an (N, α, c) -splitter. Then G can simulate an $(N^2, \alpha, c/2)$ -splitter with a slowdown of a factor of three.*

Proof: The same argument used to show that the OTIS-Expander has expansion will show that both the up and down edges have expansion, or equivalently, that the graph is a splitter. Since in either stage, packets with different destinations never contend for the same edge, we consider only the packets traveling upwards. We then show that for any set S of inputs nodes which begin with packets traveling upwards. As in the OTIS-Expander proof, we examine two cases, depending on the initial distribution of packets or input nodes.

In the first case, the majority of these packets are in groups with less than αN other packets traveling upwards. Ignoring the groups with more than αN nodes, the number of nodes in each group expands by a factor of c during the first stage, as the inputs communicate with the outputs via the expander connections. Since at least half of the numbers were in small groups, this demonstrates expansion $c/2$. In the second case, the majority of these packets are in groups with at least αN such packets. We ignore the nodes in groups with less than αN nodes. Each group will receive at most one packet from each other group while communicating across the OTIS transpose edges. Thus, at the beginning of stage two, each group will have at most αN packets that have neither been ignored or been transmitted to the outputs. This ensures that during stage two, each group will expand by a factor of c as it communicates with the outputs. Since we ignored at most half of the packets initially, this implies that graph has expansion at least $c/2$.

At this point, packets traveling upwards have gone through some expander edge have now been sent to some of the first $N/2$ outputs of some groups. Likewise, packets traveling downwards have been sent to the last $N/2$ outputs of each group. Now, we use the OTIS connections once more, this time to connect the outputs of the different groups together rather than the inputs. Passing the data though these OTIS connections on the output nodes, packets traveling upwards reach the nodes in groups 1 through $N/2$, while packets traveling downwards finish in groups $N/2+1$ through N .

At the end of the first stage, the nodes traveling up are sent to one set of groups, and the nodes traveling down are sent to a different set of groups. To construct a complete multibutterfly, the same construction would then be applied separately to each of these sets of groups, continuing in a recursive fashion until each set contains only one group. At this point, no communication between different chips is necessary, and a single-chip multibutterfly can be used to perform the routing in the final stages.

5.1.7 Bit-Parallel Crossbar

In the OTIS applications considered thus far, we have only discussed implementing networks which correspond to connected graphs; that is, networks in which it is possible for any node to communicate with any other node, possibly by routing the messages through intermediate destinations. However, there are applications in which this is not necessarily required. If each processor wishes to communicate an N -bit word, it not necessary for the first pin on any processor to be able to reach the second pin of any other processor. For this application, we have N sources, each of which begins with an N bit word and the address of another source. The goal will be to route the data so that the first bit of the word destined for source q arrives at the first pin of q , the second bit at the second pin, and so on. We will assume that the desired routing is a

permutation to avoid discussing the effects of blocking; the same techniques will apply to the blocking case.

In this section, we will show how this routing can be facilitated by the use of transpose connections. The Bit-Parallel Crossbar consists of two planes, each with N groups of N optical I/O pins. The first plane corresponds to the N I/O pins of each of the N chips (referred to as sources to distinguish them from the single-bit PEs considered earlier) to be connected. There are no connections between any of the pins in this plane. Each of the N groups in the second plane is an identical copy of a graph (possibly with more than N nodes) capable of realizing all permutations of its N inputs. In practice, a chip with crossbar connectivity would likely be used. We assume that these chips are self-routing; that is, each node must start with the destination address to which it should route its information.

Theorem 5: *Let G be the topology of the routing chips, and let T_G be the number of bit-steps G requires to route an arbitrary permutation, including the time necessary to load the message bit and destination address for each input. Then the Bit Parallel Crossbar constructed from G can route any permutation of the input words in $T_G + 1$ bit-steps.*

Proof: Initially, the OTIS node (i, j) of the first plane holds the value of the j th bit of the i th source's word. The destination of the word held by the i th source is $\pi(i)$. In the first step, each OTIS node (i, j) of the first plane sends its bit across the optical link, followed by the bits needed to specify the destination $\pi(i)$ (note that these steps are included in T_G). In the process, the j th group of the second plane receives the j bit of each of the words, as well as the information needed to specify the routing π . By assumption, the routing chips are capable of performing this routing, along with the loading of values and addresses, in T_G steps. At the end of this routing, the bit which began in source i , position j is now located in OTIS node $(j, \pi(i))$ of the second plane. Another optical transpose moves this to node $(\pi(i), j)$ of the first plane, as desired, causing one additional step of overhead.

Here we have described a packet-switched mode of operation; each source sends address information along with one word, the word is routed, and arrives at its destination. However, the same technique could be used to establish a circuit-switched path. After the first bit is sent as described above, no new address bits are sent, and all subsequent message bits follow the same path as the first one.

5.2 COMPARISON OF ELECTRICAL AND OPTICAL INTERCONNECTION

The scaling of VLSI technology has dramatically increased microelectronic device densities and speeds. However, the interconnection technology between these devices did not advance proportionally. One of the main reasons is the limited availability of interconnection materials that are compatible with VLSI and electronic packaging technologies. The increased wire

resistance as a result of smaller feature size, the residual wire capacitance due to fringing fields, and the fields between interconnect wires are among other factors that prohibit more significant improvements in electrical interconnect performance. Consequently, the overall performance of VLSI systems become increasingly dominated by the performance of long interconnects. To overcome this limitation, free-space optical interconnections have been suggested, where long electrical interconnects are replaced by an optical link consisting of a light transmitter, interconnection optics, and a photodetector.^(5-16 to 5-23) This scheme, although devoid of electrical interconnection parasitics, has its own difficulties. The unavailability of monolithically integrated optical transmitters on silicon imposes hybrid integration schemes with large parasitic capacitance and increased cost. In addition, the transformation of information from electrical to optical domain and vice versa introduces inefficiencies into the energy budget.

Therefore, it is essential to identify the regions of superiority between electrical and optical interconnects, and determine some of important trade-offs. This would help system designers choose the proper interconnect technology for a given system application. In this report, we compare the speed performance and energy cost of a class of electrical and optical interconnections for use in large-scale digital computing systems. A similar analysis, comparing free-space optics and electrical interconnects was presented by Feldman *et al.*⁽⁵⁻²⁴⁾ about ten years ago. The study presented here expands that analysis in many ways. It includes more comprehensive interconnection models and, in addition to on-chip interconnections, it also evaluates off-chip electrical interconnects.

In this study on-chip, off-chip, and free-space digital interconnections based on Multiple-Quantum-Well (MQW) or Vertical Cavity Surface Emitting Lasers (VCSEL) are evaluated. Different interconnect technologies are compared after they are analyzed in detail and optimized for minimal delay. In the free-space case, the following considerations are included: the MQW modulator saturation phenomena, the dependence of the VCSEL output power on speed, the dependence of the VCSEL threshold on output power, and the effects of parasitics due to the hybrid integration of silicon devices with optical transmitters. In the case of off-chip electrical interconnects, the effects of both series and parallel termination schemes are discussed. In all cases, superbuffers are designed to minimize the propagation delay between minimum logic and off-chip line drivers. An optimum repeater design is adopted to maximize the speed of distributed on-chip interconnections. Finally, both return-to-zero (RZ) and non-return-to-zero (NRZ) transmission schemes are used where appropriate in order to minimize energy dissipation.

5.2.1 Assumptions

Throughout the following sections, we use the term "interconnection" to refer to the physical medium used for digital communications between electronic sub-systems. Below are the assumptions under which we analyze such interconnections here (A1 through A12):

- A1. The application range that we are considering can be defined as "large-scale digital computing systems using dense interconnections." This leads to the assumption that in the electrical domain, only silicon CMOS VLSI technology (chip or wafer scale) and Multi-Chip Module (MCM) technologies are considered. This restricts our analysis to a certain class of on-chip and off-chip digital interconnections. Off-chip interconnections are those that interconnect chips on an MCM substrate, whereas on-chip connections start and end on silicon within a chip or a wafer. Since the interconnection line length is an independent parameter in our analysis, increasing the on-chip line length automatically extends the analysis to the wafer scale integration (WSI) domain.
- A2. We do not consider a certain type of computing algorithm or architecture to calculate the required interconnection line lengths or fanout in a particular system. We consider interconnection length (L_{int}) and as an independent variable. Although the derivations are carried out for both one-to-one and fanout cases, we only consider one-to-one connections in the comparative results, as they are representative of the respective merits of the technologies. More comprehensive results that include fanout considerations as well as clock distribution networks, and also include Lead Lanthanum Zirconium Titanate (PLZT) modulators as an alternative free-space transmitter technology can be found in [5-25].
- A3. Analog fan-in is not considered due to the assumption of digital communication.
- A4. We analyze on-chip and off-chip electrical interconnections separately. The performance of systems that use both types of interconnections in the same channel can be estimated by combining the results of the independent studies.
- A5. To ease the communication protocol, we assume synchronous communication where data is forced to the transmitter end of the interconnection by the rising edge of a global clock signal. It is sensed at the receiver end of the interconnection by the falling edge of the clock. Thus, the sum of the delays of the various channel components as well as the maximum clock skew determines the minimum clock period (maximum frequency of synchronous operation.)
- A6. We assume static CMOS logic design with rail-to-rail voltage swings. However, the same analysis methodology could be applied to dynamic or reduced voltage swing logic designs by proper adjustments of voltage swings and currents in the analysis.
- A7. We do not include the scaling analysis of VLSI technology. We use 0.5-micron CMOS technology parameters for numerical illustrations. However, in Section 5.2.6 we discuss

the first order effects of technology scaling on both electrical and optical interconnect performance.

- A8. We only consider free-space optics in our optical interconnection analysis. However, we do not carryout any in-depth design of the optical routing sub-system, rather, we model it with an optical time-of-flight delay and an optical power transfer efficiency. The time-of-flight delay varies as a function of the interconnection length while the power efficiency is assumed independent of the interconnection length.
- A9. For the optical interconnections, we assume that the light transmitters (MQW modulators or VCSELs) are flip-chip bonded to silicon, and that integrated reverse-biased silicon p-n junctions are used as photodiodes.
- A10. For the series terminated off-chip electrical interconnection as well as for the modulator-based optical interconnections, we assume "Non-Return-to-Zero" (NRZ) communication. In this case, the channel logic level is not altered unless a new data bit to be transmitted is different from the previously transmitted data bit. For the parallel terminated electrical interconnection and the VCSEL-based optical interconnection cases, we assume "Return-to-Zero" (RZ) transmission scheme. This is because the DC power consumption due to the parallel termination resistor or the laser current is much higher than the power consumption of the interconnection due to switching.
- A11. In the optical interconnection channel, we assume that a required bit-error-rate can be achieved by requiring a certain voltage swing at the photodiode output, which, in turn, requires a certain input optical power. In our calculations, we assume a photodiode output voltage swing of 330 mV, which is approximately equal to the transition width of a CMOS inverter transfer characteristic for 3.3 V power supply voltage in 0.5 μm CMOS.
- A12. In the off-chip interconnection case, we assume that the interconnection conductor is lossless: this approximation holds within the limits of the independent parameters used.

5.2.2 Definition Of Interconnection And Estimation Of Energy

Based on assumptions A1 and A2, we define an interconnection in our scope as follows: "an interconnection is the physical implementation of a 1-bit wide digital communication channel within or between digital VLSI subsystems (chips, wafers), involving parasitics of the medium as well as active and passive design components used to force, restore, enhance, route (optical) and sense the data in the channel." This definition is illustrated in Fig. 5-8. Note that we require the interconnection to connect minimum geometry gates due to the large-scale integration requirement.

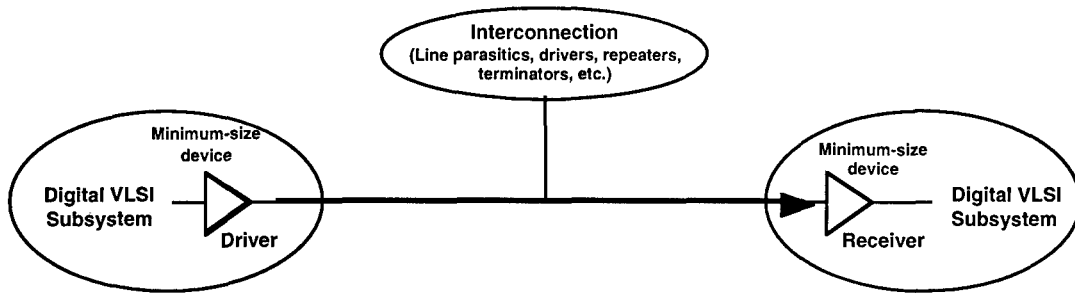


Figure 5-8. Definition of interconnection discussed in the context of this section.

Let us now consider the average energy requirement of a 1-bit data transmission through the interconnection. Since, in CMOS design, any logic gate (or a combination of gates) can be represented electrically with an equivalent inverter; we will consider the simple inverter circuit shown in Fig. 5-9. This inverter represents all the logic devices in the interconnection, whereas the capacitance C_{tot} represents the total capacitance switched during data transmission. As the input to the inverter switches from V_{sup} to ground, a current flows from power supply through the PMOS transistor. Some of the average value of this current is used to charge C_{tot} (capacitive component), while the remaining part flows to ground through the NMOS transistor, which is partly ON during switching (short-circuit component). The instantaneous capacitive current is given by:

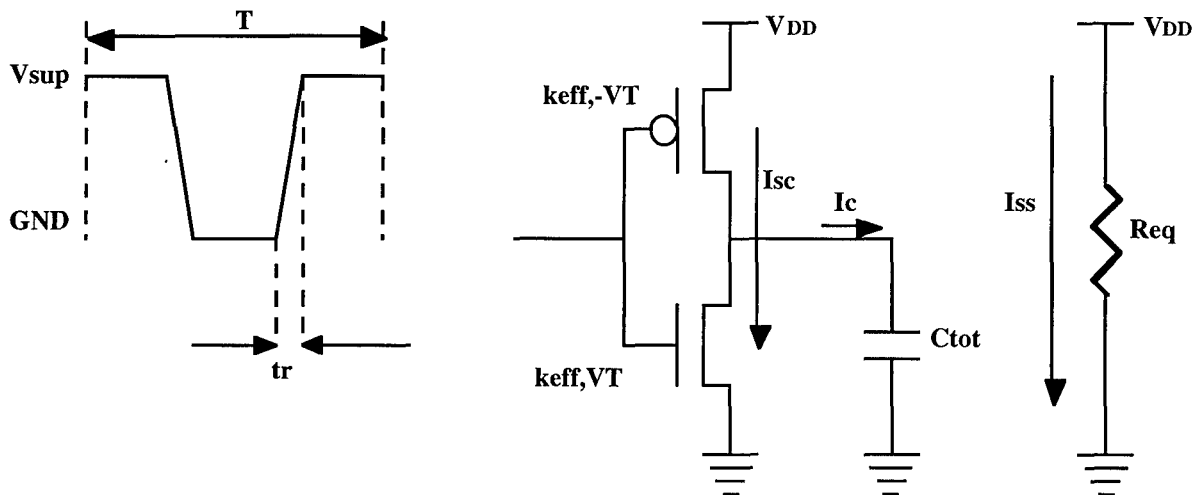


Figure 5-9. Electrical model of interconnection for the calculation of energy requirement. An inverter models all the logic devices in the interconnection, and a resistor models all the devices that require steady-state power. C_{tot} is the total interconnection capacitance that is switched during the transmission of a digital bit.

$$i_c = C_{tot} \frac{dv}{dt} \quad (5-1)$$

The supply power associated with this current is:

$$p_c = i_c V_{sup} = C_{tot} \frac{dv}{dt} V_{sup} \quad (5-2)$$

The energy is the integral of the power over the period of time that the power is dissipated:

$$E = \int_t p \cdot dt \quad (5-3)$$

Using Eq. (5-2) in Eq. (5-3) with the integration range over the full voltage swing (based on assumption A6), the capacitive component of the energy drawn from the power supply is given as:

$$E_C = C_{tot} V_{sup}^2 \quad (5-4)$$

From the electrostatic theory, we know that half of this energy is stored on C_{tot} while the other half is dissipated as heat over the PMOS transistor's resistance during the charging of C_{tot} . During the switching of the input from zero to V_{sup} , the inverter does not require any capacitive energy from the power supply. This is because the capacitive discharge currents originate from the energy stored on C_{tot} and not from the power supply. Therefore, Eq. (5-4) represents the total capacitive energy requirement from the power supply, during a period of the input signal involving two opposite transitions.

On the other hand, the average short-circuit current of a CMOS inverter (caused by the two switching transitions in one period; assuming equal rise and fall times) is given as: ⁽⁵⁻²⁶⁾

$$I_{sc} = \frac{1}{12} k_{eff} \frac{(V_{sup} - 2V_T)^3}{V_{sup}} \frac{t_r}{t} \quad (5-5)$$

In Eq. (5-5), k_{eff} is the effective transconductance parameter (modeling the equivalent transconductance of all the gates in the interconnection). V_T is defined as the transistor threshold voltage and t_r and T are the rise time and the period of the input signal respectively. Multiplying Eq. (5-5) by V_{sup} to calculate the power, and applying Eq. (5-3) over the period yields the short-circuit component of the energy drawn from the power supply as a result of two opposite switching transitions of the input signal:

$$E_{sc} = \frac{1}{12} k_{eff} t_r (V_{sup} - 2V_T)^3 \quad (5-6)$$

So far, we have considered the energy due to the capacitive and short-circuit currents. Some circuitry in the interconnection may also consume considerable steady-state current due to termination, biasing or high leakage. Covering such cases, we can express the total energy (per period of the input signal) as:

$$E_{IT} = E_C + E_{SC} + E_{SS} \quad (5-7)$$

where E_{SS} represents the energy consumed due to the steady-state currents:

$$E_{SS} = V_{sup} (I_H T_H + I_L T_L) \quad (5-8)$$

In Eq. (5-8), I_H and I_L are the average steady-state currents from power supply to ground during the steady state high and low levels of the input signal. T_H and T_L are the durations of the high and low logic levels.

Let us now consider Fig. 5-10, which shows the average scenario of a 4-bit serial data transmission, based on assumptions A4 and A10. From Fig. 5-10a, we observe that, in the non-return to zero case, the channel experiences only one pair of opposite switching (low-to-high, high-to-low) per 4 bits of data transmission. Thus, T_H and T_L are each equal to 2 clock periods. The average energy per bit can then be expressed as:

$$E_{nrz / bit} = \frac{E_C}{4} + \frac{E_{SC}}{4} + \frac{E_{SSnrz}}{4} \quad (5-9)$$

where:

$$E_{SSnrz} = 2V_{sup} T (I_H + I_L) \quad (5-10)$$

In the return-to-zero case (Fig. 5-10b), the channel performs two pairs of switching per 4 bits of data transmission, and T_H and T_L are equal to 1 and 3 clock periods respectively. In this case, the average energy per bit is then:

$$E_{rz / bit} = \frac{E_C}{2} + \frac{E_{SC}}{2} + \frac{E_{ssrz}}{4} \quad (5-11)$$

where:

$$E_{ssrz} = V_{sup} T (I_H + 3I_L) \quad (5-12)$$

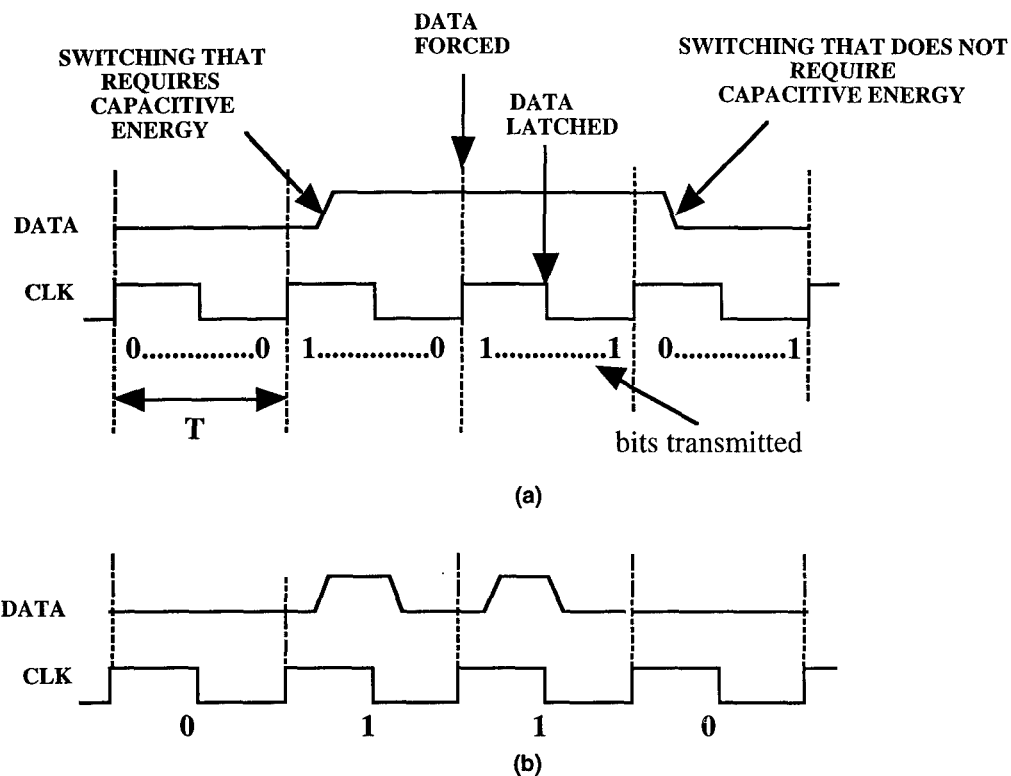


Figure 5-10. Average 4-bit transmission through the channel, (a) non-return to zero, (b) return to zero.

For a parallel terminated electrical interconnection, where the termination impedance is located at the end of the line, one must take into account the finite time-of-flight delay (t_f) of the electrical waveform in the energy calculation. Specifically, t_f should be subtracted from the high-level duration of the data T_H at the driver chip output. This is because it is only after a t_f delay of the waveform that the signal reaches the termination and creates a current through it. When the channel switches back to zero at the transmitter output, the parallel termination resistor continues to conduct current. However, the source of this energy does not come from the supply (since the transmitter driver output is disconnected from the power supply), but it comes from the energy stored in the transmission line.

Equation (5-4), Eq. (5-6), Eq. (5-9), and Eq. (5-11) are necessary and sufficient to calculate the energy requirement of the interconnection. Besides the technology parameters V_{sup} and V_T , the evaluation of these equations require that the following five parameters be known:

1. C_{tot} , total capacitance switched during transmission,
2. k_{eff} , effective transconductance of all active devices in the interconnection,
3. I_H, I_L , high and low level steady-state currents during data transmission,
4. t_r , rise time of the signal in the interconnection,
5. T , period of the synchronous system clock.

In order to calculate the above five parameters, we will apply well-established circuit techniques used to minimize the propagation delay through interconnections, such as the use of superbuffers, optimum repeaters, or transmission line terminations. Once the interconnection is designed using these techniques to operate at the highest possible speed, we will estimate the energy requirement of a 1-bit data transmission through the channel.

5.2.3 Speed And Energy Of Off-Chip Electrical Interconnections

Figure 5-11 illustrates the off-chip interconnection scheme. The transmitter chip involves an electronic sub-block composed of dense minimum-size devices necessary for large-scale computation or storage. This block is followed by a superbuffer to drive a large size line driver where a global off-chip interconnection is needed. The line driver forces the data into the off-chip conductor via an output pin. The data propagates along the conductor and is received by the various receiver chips. In the one-to-one connection case, there is only one chip at the end of the conductor. Each receiver chip is connected to the off-chip conductor via an input pin, which is then connected to a minimum-size inverter to receive and restore the data for use in the following electronic block. If a parallel termination scheme is used, then the off-chip conductor is terminated with a parallel termination resistor R_T . If a series termination is used, then the line driver output resistance is matched to the off-chip conductor's impedance.

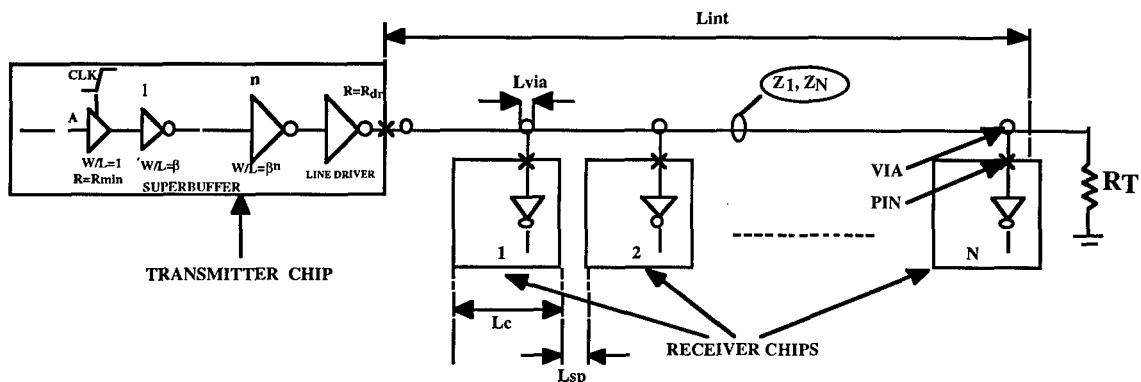


Figure 5-11. Model of off-chip electrical interconnection.

Because of the large width and thickness of off-chip conductors and the low resistivity of the materials used, such off-chip connections offer very low unit-length resistance (<1 ohm/cm). For practical purposes, they can be considered lossless. For short interconnection lengths, the transmission line behavior can be neglected, and the line acts as a lumped capacitor. In this regime, the optimum superbuffer (to minimize the propagation delay) can be designed based on the total load capacitance of the driver chip.⁽⁵⁻²⁷⁾ The interconnection can be treated as a lumped capacitor when:^(5-28,5-29)

$$t_r > 5t_f \quad (5-13)$$

where t_r is the rise time of the data signal in the interconnection and t_f is the time-of-flight delay of the signal propagation through the interconnection conductor. For long interconnection lengths, such that $t_r < 2.5t_f$, the transmission line phenomena becomes dominant.⁽⁵⁻²⁸⁾ To enable the continuity of the analysis, we will use a stricter criteria in this study, and assume that the transmission line phenomena needs to be considered when:

$$t_r < 5t_f \quad (5-14)$$

5.2.3.1 Interconnection in the lumped capacitor regime ($t_r > 5t_f$)

In this regime, a superbuffer is necessary and sufficient to drive the total load capacitance of the driver chip with minimal delay.^(5-27,5-28) The load capacitance of the driver chip can be estimated (Fig. 5-11) as:

$$C_L = C_{dr} + L_{int} C_{int\ off} + NC_{rc} \quad (5-15)$$

In Eq. (5-15), L_{int} and $C_{int\ off}$ are the off-chip interconnection length and interconnection capacitance per unit length, and C_{dr} and C_{rc} are the driver chip output and receiver chip input capacitances:

$$C_{dr} = C_{pin} + C_{sb,o} \quad (5-16)$$

$$C_{rc} = C_{pin} + C_{min,i} \quad (5-17)$$

where C_{pin} is the chip package pin capacitance given by the packaging technology, $C_{min,i}$ is the minimum inverter input capacitance given by the on-chip integration technology, and $C_{sb,o}$ is the last superbuffer stage output capacitance given by Eq. (B-5) in Appendix B. Using Eq. (5-15) in Eq. (B-3) (see Appendix B), solving n , the number of superbuffer stages from Eq. (B-5) in Appendix B, and equating to Eq. (B-3) in Appendix B provides C_L as a function of the technology parameters and independent variables:

$$C_L = \frac{C_{pin} + L_{int} C_{int\ off} + NC_{rc}}{1 - \frac{C_{min,o}}{\beta C_{min,i}}} \quad (5-18)$$

where $C_{min,o}$ is the minimum inverter output capacitance and β is the superbuffer tapering factor (see Appendix B for details). From Fig. 5-11, we see that the interconnection length as a function of the number of chips (N) is:

$$L_{int} = L_{eff}(N + 1) \quad (5-19)$$

where:

$$L_{eff} = L_c + L_{sp} \quad (5-20)$$

In Eq. (5-20), L_c is the side length of a chip and L_{sp} is the spacing between subsequent chips. Using Eq. (5-19) in Eq. (5-18) gives the total load capacitance of the driver chip:

$$C_L^N = \frac{C_{pin} + L_{eff}C_{int\ off} + N(L_{eff}C_{int\ off} + C_{rc})}{1 - \frac{C_{min, o}}{\beta C_{min, i}}} \quad (5-21)$$

For one-to-one connection, i.e. $N = 1$, Eq. (5-18) reduces to:

$$C_L^1 = \frac{C_{pin} + L_{int}C_{int\ off} + C_{rc}}{1 - \frac{C_{min, o}}{\beta C_{min, i}}} \quad (5-22)$$

Substituting Eq. (5-21) or Eq. (5-22) for load capacitance in Eq. (B-3) (see Appendix B) gives the required number $n_{1,N}$ of stages in the superbuffer. Using this result in Eq. (B-6) (see Appendix B) provides the total parasitic superbuffer capacitance C_{sb} . The total capacitance of the interconnection is then:

$$C_{tot}^{1,N} = C_{sb}^{1,N} + C_{pin}L_{int}C_{int\ off} + NC_{rc} \quad (5-23)$$

The effective transconductance k_{eff} of the interconnection is equal to the sum of the transconductances of the superbuffer stages, and is calculated by substituting $n_{1,N}$ in Eq. (B-3) (see Appendix B). Because there is no biasing or termination resistor, there is no steady-state current consumption in this regime of operation and:

$$I_H = I_L = 0 \quad (5-24)$$

The rise time of the signals in the superbuffer is estimated by Eq. (B-4) (see Appendix B). The superbuffer propagation delay is found by substituting $n_{1,N}$ in Eq. (B-1) (see Appendix B). Finally, based on the synchronous operation assumption, the minimum clock period T_{CLK} is required to be as long as the superbuffer propagation delay:

$$T_{CLK} = t_{sb, p} \quad (5-25)$$

This concludes the calculation of the five parameters of the interconnection (listed at the end of Section 5.2.2) necessary to estimate the speed performance and energy requirement of the interconnection in the lumped capacitor regime. Next, we extend the analysis to the case of long off-chip interconnections which behave as transmission lines.

5.2.3.2 Interconnection in the transmission line regime ($t_r < 5 t_f$)

In this regime, a series or parallel termination scheme is used to minimize reflections and spurious transitions. Figure 5-11 illustrates these termination schemes. In the case of series termination, the driver output resistance is matched to the impedance of the interconnection conductor. Due to the equal impedance of the conductor and the driver, the initial voltage transfer to the line is only half the supply level. To achieve full supply level across the conductor, the signal has to bounce from the receiver end, and propagate back to the driver site, thus requiring a round-trip propagation of the signal on the conductor. In the parallel termination scheme, the receiver end is terminated with a resistor equal to the line impedance. This way, no reflections occur from the receiver end, but the driver has to initially provide a high (or low) enough logic voltage to the line. Because only a one way trip of the signal is needed, a parallel termination scheme provides faster data transmission than the series termination, but it also requires more energy due to the steady-state current consumption through the parallel termination resistor. Since both termination schemes require larger than minimum geometry drivers, a superbuffer is needed to connect the minimum size logic to the line drivers and minimize the propagation delay.

In the case of one-to-one connection, the interconnection line is unloaded, and the characteristic line impedance can be calculated by:⁽⁵⁻²⁸⁾

$$Z_1 = \frac{1}{vC_{int\ off}} \quad (5-26)$$

where v is the propagation speed in the medium and C_{intoff} is the parasitic off-chip line capacitance per unit length. In the case of fanout, the line can be treated as distributed if the spacing between the receivers is short. Under this assumption the total loaded line impedance can be calculated as:⁽⁵⁻²⁸⁾

$$Z_N = \frac{Z_1}{\sqrt{1 + \frac{C_N}{C_{int\ off}}}} \quad (5-27)$$

where C_N is the fanout capacitance per unit length, which from Fig. 5-11, is seen to be:

$$C_N = \frac{C_{rc}}{L_{eff}} \quad (5-28)$$

Similarly, the time-of-flight delays for one-to-one and fanout connections are estimated as:

$$t_f^1 = \frac{L_{int}}{v} \quad (5-29)$$

$$t_f^N = \frac{L_{int}}{v} \sqrt{1 + \frac{C_{rc}}{L_{eff} C_{int\ off}}} \quad (5-30)$$

Note that in Eq. 5-27 and Eq. 5-30, we have neglected the increasingly smaller effect of the output capacitance of the driver chip as the interconnection length increases. Now, we are in a position to estimate the boundary of the lumped capacitor and transmission line regimes. Using Eq. 5-29 and Eq. 5-30 in Eq. 5-13 and Eq. 5-14, and solving for L_{int} provides the region of transmission line operation:

$$L_{int,1} > \frac{t_r v}{2.5} \quad (5-31)$$

$$L_{int,N} > \frac{t_r v}{2.5 \sqrt{1 + \frac{C_{rc}}{L_{eff} C_{int\ off}}}} \quad (5-32)$$

In the case of series termination, the output resistance of the last superbuffers stage in the driver chip should match the characteristic line impedance (Z_1 in one-to-one and Z_N in the fanout case):

$$R_{dr,1,N}^{ser} = Z_{1,N} \quad (5-33)$$

In the parallel termination case, while the buffer impedance does not have to match the line impedance, it should be low enough to provide the necessary high level voltage across the parallel termination resistor:

$$R_{dr,1,N}^{par} = R_T \left(\frac{V_{DD}}{V_H} - 1 \right) = Z_{1,N} \left(\frac{V_{DD}}{V_H} - 1 \right) \quad (5-34)$$

where R_T is the termination resistor whose value is matched to the line impedance, and V_H is the minimum acceptable logic high level voltage.

After the resistance of the last superbuffer stage is calculated, the entire superbuffer can be designed. The size S of the n^{th} (the largest) buffer stage is: $S = \beta^{n-1} = R_{\min}/R_{dr}$ where R_{\min} is the minimum size inverter output resistance. This allows the calculation of the total number of stages in the superbuffer:

$$n_{1,N}^{ser,par} = \frac{1}{\ln \beta} \ln \left\{ \frac{R_{\min}}{R_{dr_{1,N}^{ser,par}}} \right\} \quad (5-35)$$

and, from the biggest to the smallest, the stages diminish in size by a factor of β . After the number of stages is determined, the effective transconductance k_{eff} of the line is calculated from Eq. (B-7) (see Appendix B). The total capacitance of the superbuffer (C_{sb}) is obtained by substituting 5-35 for n in Eq. (B-6) (see Appendix B). As in the lumped capacitor regime, the total interconnection capacitance is calculated as:

$$C_{tot_{ser,par}}^{1,N} = C_{sb_{ser,par}}^{1,N} + C_{pin} + L_{\text{int}} C_{\text{int off}} + NC_{rc} \quad (5-36)$$

In the series termination case, there is no steady state current consumption. In the parallel termination case, there is a steady state current as long as the logic level of the interconnection is high:

$$I_{H,par}^{1,N} = \frac{V_H}{Z_{1,N}}; I_{H,ser}^{1,N} = 0 \quad (5-37)$$

$$I_{L,ser}^{1,N} = 0; I_{L,par}^{1,N} = 0 \quad (5-38)$$

Since the interconnection is driven by the superbuffer, the rise time of the data signals in the interconnection is calculated by Eq. (B-4) (see Appendix B). The minimum clock period T_{CLK} is equal to the sum of the superbuffer propagation delay given by Eq. (B-1) (see Appendix B), and the one-way or round-trip time-of-flight delay for parallel and series termination, respectively:

$$T_{CLK} = t_{sb,p}^{1,N} + m \cdot t_f^{1,N} \quad (5-39)$$

where $m = 1$ for parallel, and $m = 2$ for series termination. In Eq. (5-39), we assumed for simplicity that the last superbuffer stage propagation delay is equal to the propagation delay of a previous stage, although its size is determined by the matching condition rather than the output capacitance. In Eq. (5-39), we also neglected the rise time of the signal at the receiver end, which is quite small due to the small flip-chip bond and receiver input capacitance. In Eq. (5-39), the superbuffer propagation delay is different for series and parallel termination cases. However,

this difference is small. This is because the minimum high level voltage V_h is generally around half the supply level, requiring a buffer size as big as a buffer designed for the series termination case.

For short interconnection lengths, estimations of the interconnection speed with the lumped capacitor or transmission line models provide close results. Thus, we will only plot the speed performance estimated by the transmission line theory for all interconnection lengths. The technology constants used in the numerical illustrations are presented in Table 5-2. Figure 5-12 illustrates the maximum clock speed and energy per bit transmitted as a function of interconnection length for one-to-one connection. Because of Eq. (5-29) and Eq. (5-30), the speed decreases with interconnection length, from above 1 GHz to around 300 MHz (series terminated 20 cm long line) in the slowest case. The speed of parallel-terminated lines is higher mostly due to the one-way trip delay of the signal.

Table 5-2. VLSI and electrical packaging constants.

Symbol	Description	Value
VDD	VLSI power supply voltage level	3.3 V
VT	Transistor threshold voltage	0.5 V
VH	Minimum acceptable logic high-level voltage	$V_{DD}/2$
Rmin	Minimum-size transistor average resistance	8700 ohm
RCmin	Minimum-size inverter internal propagation delay	100 ps
kmin	Minimum-size transistor transconductance parameter	$80 \mu A/V^2$
Cmin,i	Minimum-size inverter input capacitance	6 fF
Cmin,o	Minimum-size inverter output capacitance	6 fF
	Optimum superbuffers tapering factor	5
Cpin, Cbond	Flip-chip bond capacitance (also used as pin capacitance in the text)	20 fF
Lc	Side length of a chip in MCM packaging	1 cm
Lsp	Spacing between MCM chips	0.2 cm
v	Speed of wave propagation on MCM substrate	15×10^9 cm/s
Cintoff	Off-chip interconnection capacitance per unit length	1 pF/cm
Rintoff	Off-chip interconnection resistance per unit length	0.8 ohm/cm
Cinton	On-chip interconnection capacitance per unit length	1.4 pF/cm
Rinton	On-chip interconnection resistance per unit length	90 ohm/cm
CminN,i	Minimum NMOS transistor input capacitance	3 fF
CminN,o	Minimum NMOS transistor output capacitance	2 fF
Wmin	Off-chip conductor minimum width	25 μ m
Cff	Off-chip conductor fringing field capacitance per cm	1 pF/cm
d _{on}	Side length of an electronic block (on a wafer) to which clock is routed	1cm
d _{off}	Side length of a chip (on an MCM) to which clock is routed	1cm

1-1 CONNECTION, SPEED AND ENERGY

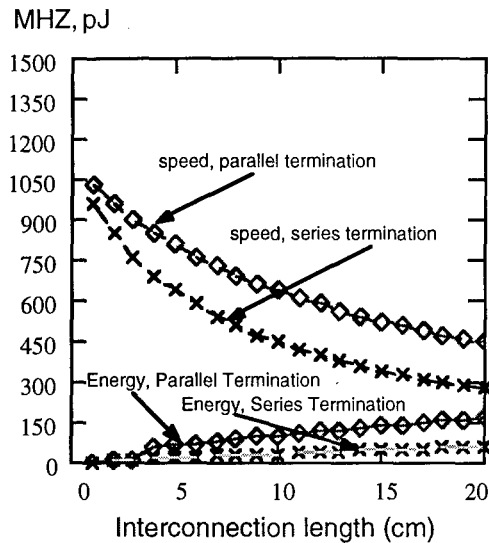


Figure 5-12. Speed performance and energy requirements of off-chip electrical interconnections as a function of interconnection length for serial and parallel terminated lines in the case of one-to-one connections.

The strongest dependence of the delay on the interconnection length comes from the time-of-flight delay, which increases linearly with line length. For short interconnections, as suggested by Eq. (5-31) and Eq. (5-32), the line is in the lumped capacitor regime. In this regime, the energy requirement increases linearly with the interconnection length due to the linear dependence of both C_{tot} and k_{eff} on L_{int} . As interconnections get longer, the transmission line phenomena becomes dominant. This yields a sudden increase in energy at the boundary of this regime. This increase is small in the series termination case, since this scheme requires nothing more than a slight increase of the line driver size. In the parallel termination case, however, the jump is drastic since this scheme requires a parallel termination resistor that consumes high steady-state power. Note that the boundary between lumped-capacitor and transmission line regimes is not precisely defined due to the approximate separation of the two regions by Eq. (5-13), and Eq. (5-14).

For short interconnections, the biggest contribution to the total energy comes from the termination resistor. As line lengths get longer, the capacitive component becomes quickly the dominant component, constituting up to 70% of the total energy. For short interconnections, the short-circuit component of the energy is about 20% of the other components. As line lengths get longer, its effect reduces to about 10%.

Because the line impedance is independent of line length, once the line impedance is matched (in the series termination case) the superbuffer size remains constant as interconnections get longer. However, as the line length increases, the total interconnection capacitance increases

linearly, resulting in an overall linear increase of energy. In the parallel termination case, the slope of the energy increase is higher than in the series termination case. This is because, in addition to the linear increase of the interconnection capacitance with line length, the energy requirement of the parallel termination resistor also increases with line length. Indeed, longer interconnections reduce the transmission speed of a bit, which results in power dissipation at the termination resistor over a longer period of time.

5.2.4 Speed And Energy Of On-Chip Electrical Interconnections

Figure 5-13 illustrates a typical on-chip (wafer) interconnection configuration. In this case, the distributed line behavior is dominant, since an on-chip conductor is lossy ($R_{int} \approx 100$ ohm/cm) Thus, for short interconnection lengths, a superbuffer is sufficient to drive the interconnection with small delays.

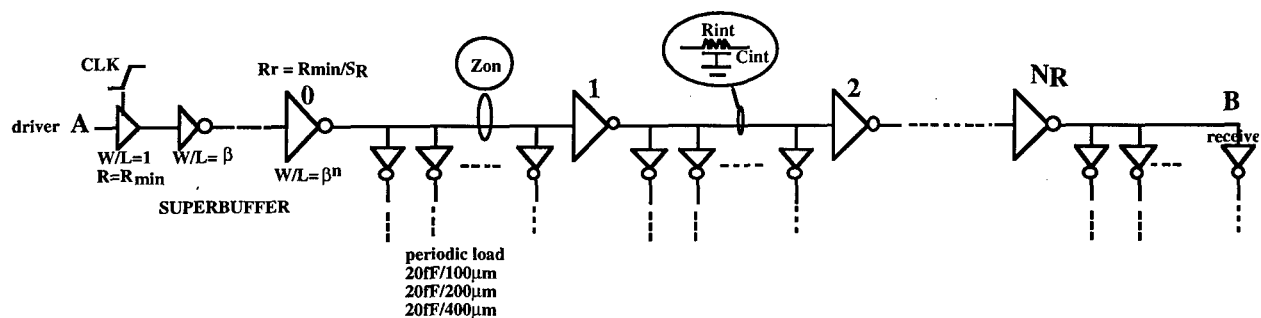


Figure 5-13. Model of on-chip interconnection.

As interconnection length gets longer, the line resistance becomes non-negligible. The propagation delay through the interconnection can only be minimized by using optimally sized and spaced repeaters.⁽⁵⁻²⁸⁾ Since a repeater is large, a superbuffer is still needed to drive the first repeater. We consider three different fanout loading conditions: a minimum inverter input load every 100, 200, and 400 μm . Again, assuming distributed load behavior, the extra loading due to fanout can be absorbed in the parasitic line capacitance. Given the line parasitics per unit length and the minimum inverter parameters, the optimum number (N_R) and size (S_R) of the repeaters can be calculated as:⁽⁵⁻²⁸⁾

$$N_R = \sqrt{\frac{0.4R_{inton}(C_{inton} + C_N)}{0.7R_{min}C_{min,o}}} \quad (5-40)$$

$$S_R = \sqrt{\frac{R_{min}(C_{inton} + C_N)}{R_{inton}C_{min,o}}} \quad (5-41)$$

where R_{inton} and C_{inton} are the on-chip interconnection resistance and capacitance per unit length. R_{min} and $C_{min,o}$ are the output resistance and input capacitance of a minimum size inverter, and C_N is the extra capacitance per unit length due to fanout. The resulting propagation delay through the interconnection (including the delay of the repeaters) is found as: ⁽⁵⁻²⁸⁾

$$t_{p,RP} = 2.5 \sqrt{R_{min}(C_{inton} + C_N)R_{inton}C_{min,o}} \quad (5-42)$$

Based on typical 0.5 μm technology parameters, there has to be one repeater (150 times larger than a minimum geometry inverter) approximately every centimeter of on-chip interconnection length in order to minimize the propagation delay (in the one-to-one connection case). Since the repeater size is much larger than a minimum size inverter, the first repeater has to be driven by a superbuffer. The input capacitance of a repeater is calculated as:

$$C_{R,in} = S_R C_{min,i} \quad (5-43)$$

The superbuffer load capacitance is equal to the sum of the input capacitance of a repeater given by Eq. 5-43 and the output capacitance of the last superbuffer stage:

$$C_{sb,L} = C_{R,in} + C_{sb,o} \quad (5-44)$$

Substituting Eq. 5-44 in Eq. (B-3) (see Appendix B) for the load capacitance and solving for n with the help of Eq. (B-5) (see Appendix B) yields the number of superbuffer stages $n_{sb,on}$. The total superbuffer capacitance $C_{sb,on}$ and the propagation delay $t_{sbp,on}$ are calculated by Eq. (B-1) and Eq. (B-6) (see Appendix B). Generally, a two-stage superbuffer is sufficient to drive the repeater. The superbuffer effective transconductance is calculated by Eq. (B-7) (see Appendix B). The effective transconductance of the interconnection is then estimated as:

$$k_{eff} = k_{eff}^{sb} + N_R S_R k_{min} \quad (5-45)$$

The total capacitance of the interconnection is then:

$$C_{tot} = \frac{k_{eff}}{k_{min}} (C_{mini} + C_{min,o}) + L_{int}(C_{inton} + C_N) \quad (5-46)$$

Since there is no termination or biasing requirement, there is no steady state current consumption.

Because of the use of a superbuffer cascaded with repeaters in the interconnection, the rise time of the signals varies slightly throughout the interconnection. The biggest contribution to the energy (for long connections) comes from the repeaters. Thus, for simplicity, we will use the signal rise time at the input of a typical repeater for the entire interconnection. This rise time can

be estimated by weighing the distributed RC terms (between two successive repeaters, i.e. about a centimeter) by 1 and lumped ones by 2.3: ⁽⁵⁻²⁸⁾

$$t_r = R_{inton}C_{inton} + 2.3 \left\{ \frac{R_{min}}{S_R} (C_{inton} + C_N + C_{R,in}) + R_{inton}C_{R,in} \right\} \quad (5-47)$$

Finally, the minimum clock period is calculated to be:

$$T_{CLK} = t_{sbp,on} + T_{p,RP} \quad (5-48)$$

Unlike in the off-chip interconnection case (i.e. in transmission line regime), k_{eff} is a function of the interconnection length, since longer interconnections involve more repeaters. This results in a faster increase of the energy requirement as a function of L_{int} . Figure 5-14 illustrates the results of the analysis. As can be seen in Fig. 5-14, the effect of loading on the speed is small. The energy requirement increases linearly due to the linear dependence of the number of repeaters as well as the line capacitance on L_{int} . The biggest contribution to the energy comes from the capacitive component, while the short circuit component constitutes only 25% of the overall energy. The energy requirement of the on-chip interconnection is comparable to that of an off-chip interconnection. While there is no need for terminations, on-chip interconnects require periodic repeaters because of the lossy nature of the interconnection conductor. For the same reason, on-chip interconnections are also slower than their off-chip counterparts.

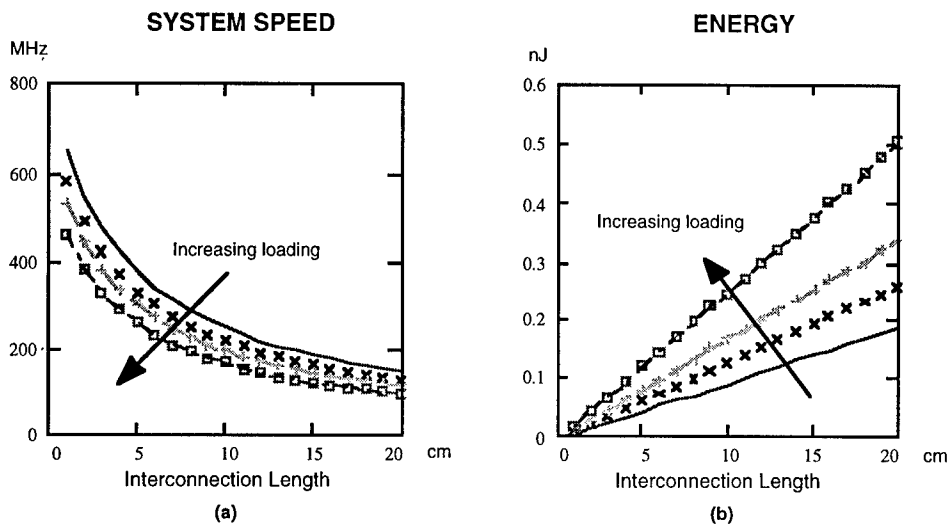


Figure 5-14. Speed performance and energy requirement of on-chip electrical interconnections as a function of interconnection length for different loading conditions (no load, 50fF/mm, 100fF/mm, and 200fF/mm), a) speed, b) energy.

5.2.5 Optical Interconnections

Figure 5-15 shows the typical optical interconnection model considered in this study. As in the electrical interconnection case, a superbuffer amplifies the minimum logic to drive the optical transmitter driver, which in turn switches the optical transmitter device. The transmitter is modeled by a current source and a capacitor. For a VCSEL, the current source models the laser current needed to produce the required laser output power. For an MQW modulator, the current models the current due to absorption. The transmitter capacitance includes the transmitter driver output capacitance, the transmitter device capacitance, and the flip-chip bond capacitance (based on Assumption A9).

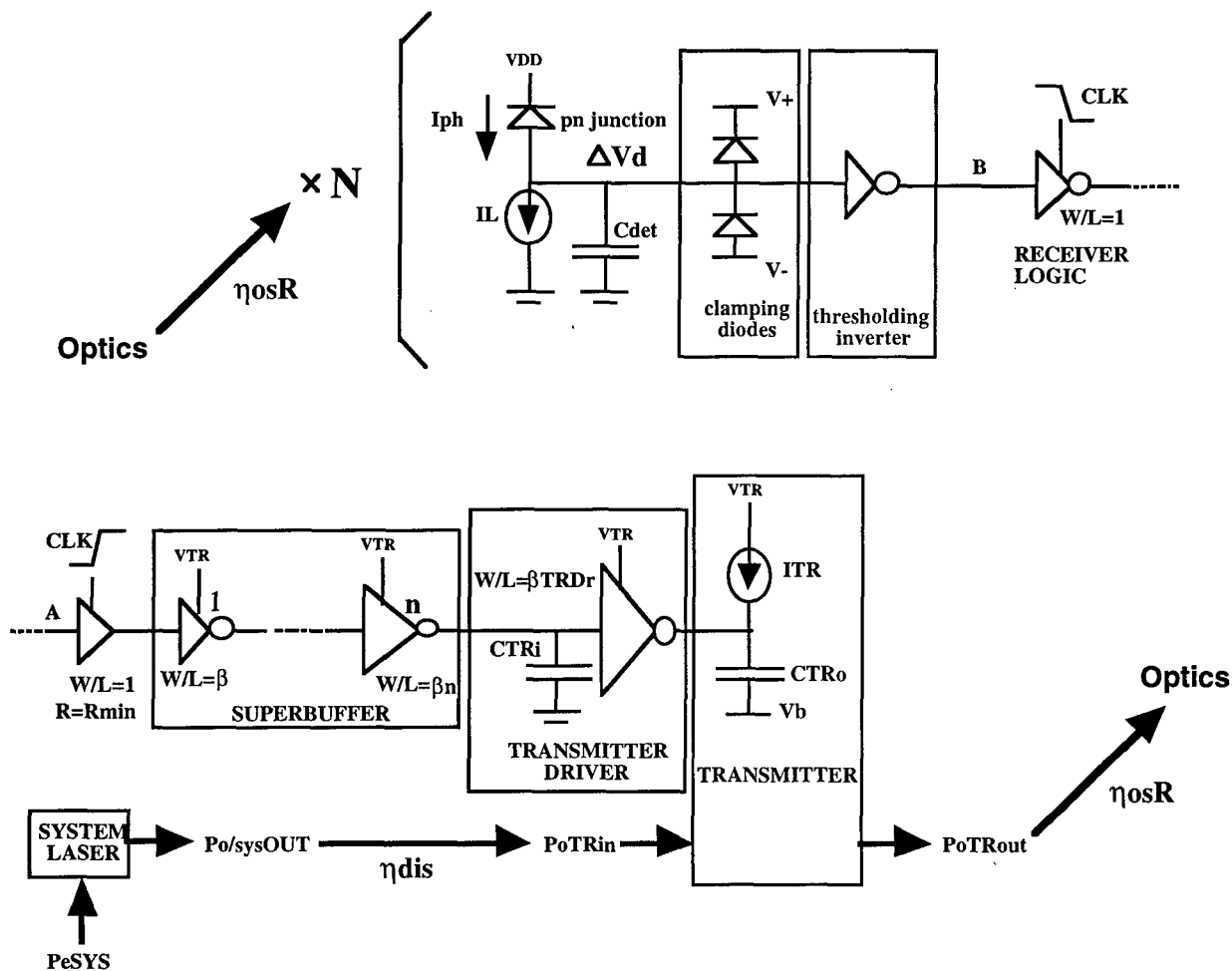


Figure 5-15. Model of interconnection using free-space optics. On the transmitter site, the interconnection includes a transmitter, a transmitter driver, and a superbuffer in the case where the transmitter is too large to be driven by minimum logic. The receiver site includes a photodiode with a thresholding current source, clamping diodes to limit the voltage swing, and a minimum-size inverter to amplify the photodiode output signal and restore the logic levels.

The minimum logic that drives the interconnection uses the logic-level power supply voltage V_{DD} . Due to the transmitter device requirements, the transmitter itself, the transmitter driver, and the superbuffer all use a separate supply level V_{TR} , which is generally larger than V_{DD} . We assume that V_{TR} is less than the breakdown voltage on the chip, such that no special circuitry is needed in these stages. Furthermore, if V_{TR} is only a couple of times larger than V_{DD} , an inverter in the logic is capable to drive a properly-biased, higher voltage inverter (first inverter in the superbuffer), since this inverter would provide the necessary gain. Note that the first inverter of the superbuffer never turns off completely due to the limited voltage at its input, resulting in steady-state power dissipation. Alternatively, the superbuffer could be operated at the VLSI supply level V_{DD} , and the amplification could be performed by the transmitter driver. In this "late amplification" scheme, the superbuffer consumes less energy due to the reduced supply level, but the transmitter driver consumes more energy because the transmitter driver does not switch off completely. Since the transmitter driver is generally larger than minimum, the energy requirement is higher than that of the first superbuffer stage. In the "early amplification" case, where the amplification is performed by the first inverter of the superbuffer, the superbuffer consumes more and the transmitter driver consumes less energy. A detailed analysis shows that, for transmitter voltages below 10V, the "early amplification" scheme is more energy efficient. In a dynamic design, or in cases where the transmitter driver does not have to be large, "late amplification" would be more beneficial.

In some cases, the transmitter may need to be biased at a certain voltage for optimum operation: a separate supply voltage V_b is used for this purpose (see Fig. 5-15). If a modulator is used as transmitter, the optical power needs to be routed to each modulator from an optical power supply, as is depicted in Fig. 5-15.

The free-space optical interconnections route a transmitter output to N receivers. They are modeled by a time-of-flight delay and a power transfer efficiency (see Table 5-3) At the receiver site, a current source models the absorbed photocurrent in a reverse-biased p-n junction. The current source (I_p) models the photodiode load current. Two clamping diodes are used to limit the photodiode output voltage to about 330mV (based on Assumption A11) and obtain fast switching. A minimum size inverter is used to threshold the photodiode output and restore the logic levels.

Table 5-3. Optical routing/power supply constants.

Symbol	Description	Value
$\eta_{OSR_{las}}$	VCSEL-based optical interconnection power routing efficiency	0.7
$\eta_{OSR_{mod}}$	Modulator-based optical interconnection power routing efficiency	0.5
η_{dis}	Optical power distribution efficiency	0.9
$\eta_{L,sys}$	System laser, current-to-optical power conversion efficiency	0.3 W/A

Independent of the type of transmitters used, the design of the interconnection always starts with the estimation of the required photocurrent dynamic range at the receivers. This photocurrent depends on the technology parameters, the operation speed, and the photodiode voltage swing. After the detector photocurrent dynamic range is determined, the transmitter output power dynamic range can be estimated based on the detector responsivity and the optical link power transfer efficiency. This is followed by the design of the particular transmitter used.

5.2.5.1 MQW modulator as transmitter

For optical interconnection applications, optical intensity modulation can be directly achieved in a Multiple Quantum Well (MQW) structure through electrically modulating the excitonic absorption. This effect is commonly referred to as the quantum-confined Stark effect.^(5-30,5-31) The MQW modulators are, potentially, high-speed devices and their use in free-space optoelectronic interconnection systems has been demonstrated.⁽⁵⁻³²⁾ The modulator contrast ratio, insertion loss, and absorption saturation are important issues, which constrain the applications of these modulators.⁽⁵⁻³³⁾ It has been observed experimentally that the contrast ratio and the insertion loss of an MQW modulator saturate at high optical intensities.⁽⁵⁻³⁴⁾ This saturation has been attributed to carrier screening in the material and analyzed in the past.⁽⁵⁻³⁵⁾ In this section, we will neglect the saturation effect by sizing the devices appropriately as a function of optical power requirement.

5.2.5.1.1 Analysis of MQW modulator interconnect

In the case of optical interconnection using MQW modulators, the driver design should take into the absorbed modulator current account. In addition, the size of the MQW modulator should be a function of the input optical power to avoid the saturation phenomena. Higher speed requires more modulator power, thus resulting in a larger modulator with bigger capacitance and current, affecting the design of the driver. The transmitter itself also contributes to the total steady-state current due to absorption. Appendix F presents the details of the MQW modulator driver design. Based on the input capacitance of the driver given in Eq. (F-12) (see Appendix F), a superbuffer can be designed as before if needed. The total capacitance C_{tot} is then;

$$C_{tot} = [C_{sb} + C_{TR,i} + C_{TR}]A_{TR} + NC_{rc} \quad (5-49)$$

where C_{sb} is the superbuffer total parasitic capacitance, $C_{TR,i}$ and C_{TR} are the transmitter driver input and transmitter capacitances given by Eq. (F-12) and Eq. (F-10) (see Appendix F), respectively, and C_{rc} is the receiver capacitance. The effective interconnection transconductance is;

$$k_{eff} = k_{eff}^{sb} + k_{dr} \quad (5-50)$$

where k_{dr} is the transmitter driver transconductance given in Appendix F. The high and low level steady-state currents are calculated as;

$$I_H = I_{TR} + I_{MQW,H} + I_{Load} + I_{RC} \quad (5-51)$$

$$I_L = I_{TR} + I_{ph,L} + I_{RC} \quad (5-52)$$

where I_{TR} and I_{RC} are the transmitter and receiver site steady-state currents due to amplification (estimated in Appendix D), $I_{MQW,H}$ is the MQW driver high level current given by Eq. (F-6) (see Appendix F), and the photodetector currents are the same as in the previous section. The minimum clock period of the interconnection can be estimated as;

$$T = t_{sb,p} + t_{dr,p} + t_{p,MQW} + t_{fopt} + t_{p,ph} + t_{p,det} + RC_{min} \quad (5-53)$$

where $t_{p,MQW}$ is the MQW modulator internal propagation delay, $t_{dr,p}$ is the modulator driver propagation delay which can be approximated as half the signal rise time t_r given by Eq. (F-9) (see Appendix F).

There is an additional energy requirement for an external system laser to optically drive the modulators. The electrical power requirement in the system laser due to a single MQW modulator is calculated by:

$$P_{opt,in} = (\eta_{MQW,H} - \eta_{MQW,L})^{-1} \cdot DR_{opt} \quad (5-54)$$

The technology parameters used for numerical illustrations in the next sections are presented in Table 5-4 and Table 5-5.

Table 5-4. Photodetector constants.

Symbol	Description	Value
ΔV_d	Photodiode output voltage swing	330 mV
A_{ph}	Photodiode area	50 μm^2
A_{diode}	Clamping diode area	10 μm^2
C_{ph}	Photodiode device capacitance per unit area	0.2 fF/ μm^2
C_{clamp}	Clamping diode capacitance per unit area	0.2 fF/ μm^2
R_{ph}	Photodiode responsivity	0.3 A/W
$t_{p,ph}$	Photodiode internal propagation delay	100 ps

Table 5-5. MQW modulator technology constants.

Symbol	Description	Value
VTRMQW	MQW modulator supply voltage	10 V
rMQW	MQW modulator responsivity	0.53 A/W
V _L	MQW modulator low level logic voltage	0.5V
I _s (V _m)	MQW modulator saturation intensity at the modulation voltage	800 W/cm ²
I _s (0)	MQW modulator saturation intensity at zero voltage	244 W/cm ²
K _m	MQW modulator absorption slope ratio	4
k(0)	MQW modulator absorption slope at zero voltage	0.2
C _{MQW}	MQW modulator device capacitance per unit area	0.12 fF/μm ²
t _{p,MQW}	MQW modulator internal propagation delay	30 ps

5.2.5.1.2 Comparison To Off-Chip Electrical Interconnects

Figure 5-16 illustrates the comparison between MQW-based optical and off-chip electrical one-to-one interconnections. The behavior of the curves indicates a superior optical speed performance. This can be attributed to the need of smaller buffer requirement for the MQW has a smaller device capacitance. The break-even line length for equal energy requirement is around 3 cm compared to parallel or series terminated electrical interconnects. For both system energy and processing plane energy requirement, MQW-based optical interconnects offer a simultaneous speed/energy advantage for one-to-one connection. For long interconnects, the MQW optical interconnects require almost 6 times less processing plane energy and yet operate faster than the fastest electrical interconnect. In terms of overall system energy, the energy efficiency of the optical interconnects drops to less than twice better than their electrical counterparts. Note that by operating the optical light detectors 10 times slower than the speed of minimum inverter, optical interconnects could be more energy efficient than even series terminated electrical interconnect for long interconnections, and yet operate faster. The biggest contribution to the system energy requirement comes from the external light source. The biggest contribution to the processing plane energy requirement comes from the steady-state modulator currents due to absorption.

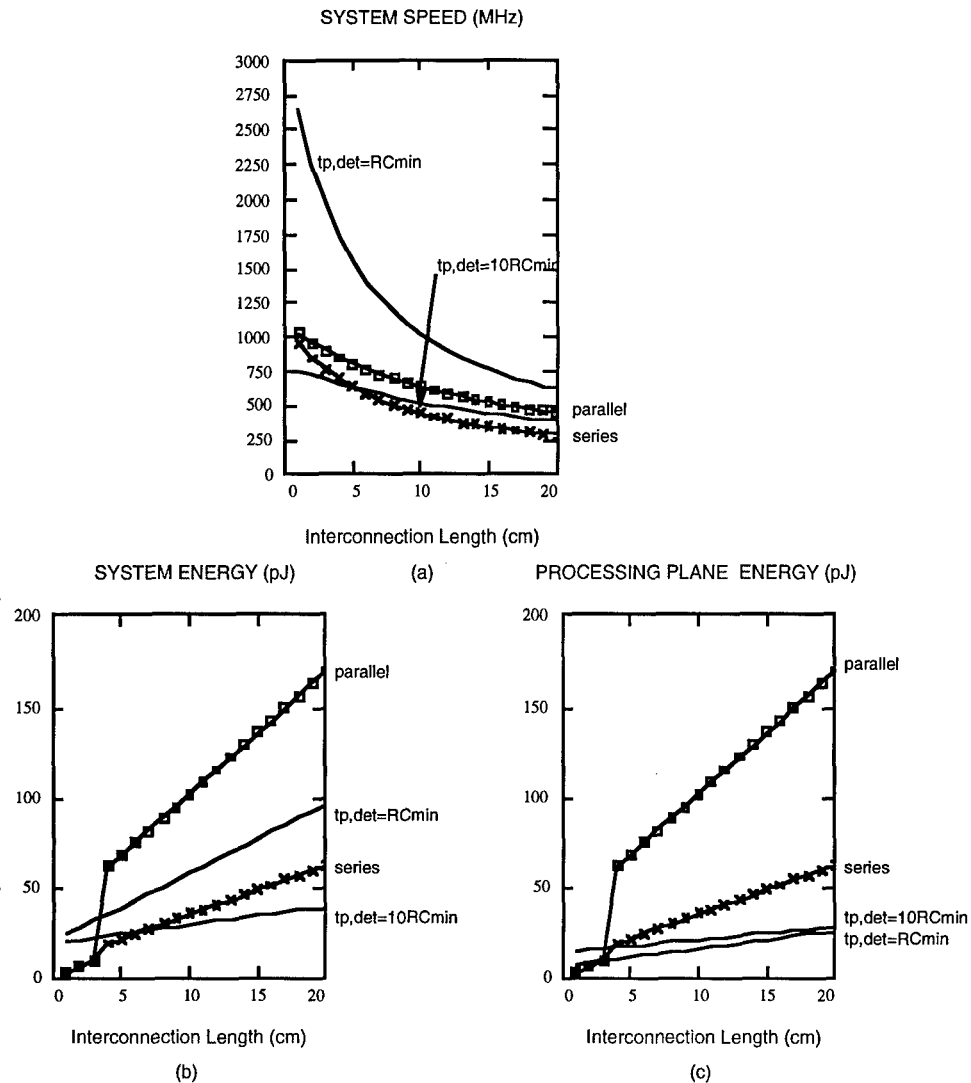


Figure 5-16. Speed and energy comparison between off-chip electrical and MQW-based optical interconnects for one-to-one connections: a) speed, b) total system energy requirement, c) processing plane energy requirement. Curves related to electrical interconnect are illustrated with symbols. The performance of optical interconnects is illustrated at different detection speeds ($t_{p,det}$) to allow comparison to electrical interconnects running at the same speed.

5.2.5.1.3 Comparison To On-Chip Electrical Interconnects

Figure 5-17 illustrates the comparison between MQW-based optical and on-chip electrical one-to-one interconnections. There is a break-even line length of about 4 cm for equal system energy. It is interesting to note that, faster optical interconnects require less on-chip energy per bit. This is due to the smaller contribution of the steady-state components. In this case, the break-even line length is only about 2 cm. For long wafer-scale interconnects, the simultaneous speed/energy advantage of the MQW-based optical interconnect exceeds an order of magnitude.

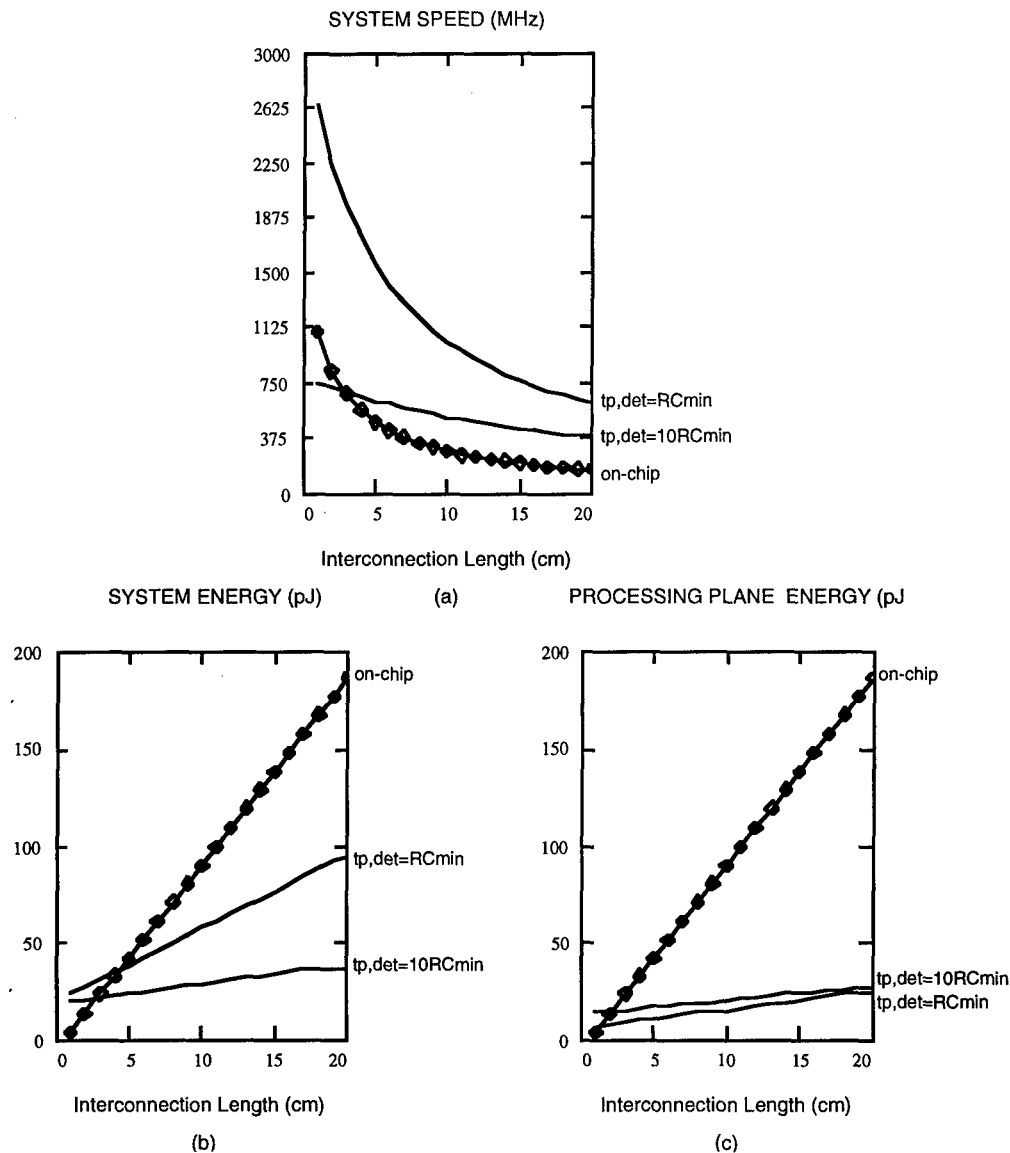


Figure 5-17. Speed and energy comparison between wafer-scale electrical and MQW-based optical interconnects for one-to-one connections: a) speed, b) total system energy, c) processing plane energy. Curves related to electrical interconnect are illustrated with symbols. The performance of optical interconnects is illustrated at different detection speeds ($t_{p,det}$) to allow comparison to electrical interconnects running at the same speed.

5.2.5.2 VCSEL as Transmitter

Compared to light modulators, the use of Surface-Emitting Lasers (SELs) as the optical transmitters in a free-space optical interconnection system significantly simplifies the optical system design by eliminating the requirement for an external laser source and associated optics. Thus, SELs have the potential for improving the optical link efficiency and the system stability and robustness. In particular, vertical-cavity surface-emitting lasers (VCSELs) are very promising for 2-D array applications.^(5-36,5-37) A great amount of work has been performed lately

to improve the uniformity of laser arrays, reduce the threshold currents, and increase the maximum power output.^(5-38,5-39,5-40) The threshold voltage and electrical to optical power conversion efficiency, along with the series resistance, are the main characteristics of lasers that influence the device performance in the applications of digital interconnection.^(5-41,5-42)

5.2.5.2.1 Analysis of VCSEL Interconnect

The details of the VCSEL transmitter driver design is presented in Appendix G. Based on the input capacitance of the driver given in Eq. (G-12) (see Appendix G), a superbuffer can be designed as before if needed. The total capacitance C_{tot} is then;

$$C_{tot,VCSEL} = [C_{sb} + C_{TR,i} + C_{TR}]A_{TR} + NC_{rc} \quad (5-55)$$

where C_{sb} is the superbuffer total parasitic capacitance, $C_{TR,i}$ and C_{TR} are the transmitter driver input and transmitter capacitances given by Eq. (G-12) and Eq. (G-10) (see Appendix G), respectively, and C_{rc} is the receiver capacitance. The effective interconnection transconductance is:

$$k_{eff} = k_{eff}^{sb} + k_n \quad (5-56)$$

where k_n is the transmitter driver transconductance given by Eq. (G-8) (see Appendix G). The high and low level steady-state currents are calculated as:

$$I_H = I_{TR} + I_{VCSEL,H} + I_{Load} + I_{RC} \quad (5-57)$$

$$I_L = I_{TR} + I_{th} + I_{ph,L} + I_{RC} \quad (5-58)$$

where the VCSEL transmitter high level current is given by Eq. (G-7) (see Appendix G), VCSEL threshold current I_{th} is given by Eq. (G-6) (see Appendix G), and the receiver currents are the same as in the previous section. The minimum clock period of the interconnection can be estimated as:

$$T = t_{sb,p} + t_{dr,p} + t_{p,VCSEL} + t_{fopt} + t_{p,ph} + t_{p,det} + RC_{min} \quad (5-59)$$

where $t_{p,VCSEL}$ is the VCSEL device propagation delay, t_{dr} is the laser driver propagation delay which can be approximated as half the signal rise time t_r given by Eq. (G-9) (see Appendix G).

Unlike in the modulator cases, there is no need for an external light source since each VCSEL is a light source itself. VCSEL parameters are shown in Table 5-6.

Table 5-6. VCSEL constants.

Symbol	Description	Value
V_{TRlas}	VCSEL supply voltage	10 V
ϕ	VCSEL, slope of threshold current/laser diameter characteristic	0.7 mA/ μm
γ	VCSEL, slope of output power/laser diameter characteristic	0.5 mW/ μm
η_{LI}	VCSEL, slope of output power/laser current characteristic	0.3 W/A
V_{th}	VCSEL threshold voltage	2 V
$t_{p,VCSEL}$	VCSEL internal propagation delay	30 ps
C_{las}	VCSEL device capacitance per unit area	0.2 fF/ μm^2

5.2.5.2.2 Comparison To Off-Chip Electrical Interconnects

Figure 5-18 illustrates the comparison between VCSEL-based and off-chip one-to-one interconnections. The speed of the VCSEL-based interconnection is slightly lower than in the MQW case due to the larger driver requirement of the VCSEL because of the high laser current. The break-even line lengths for equal energy is around 1 cm. For 20 cm long interconnects, VCSEL-based interconnect requires an order of magnitude less system energy and operates two to four times faster. Note that with improvements in the VCSEL technology, the numbers reported in this report are expected to improve further.⁽⁵⁻⁴²⁾ Also note that compared to the MQW modulator cases, the fastest VCSEL-based optical interconnect requires less system energy.

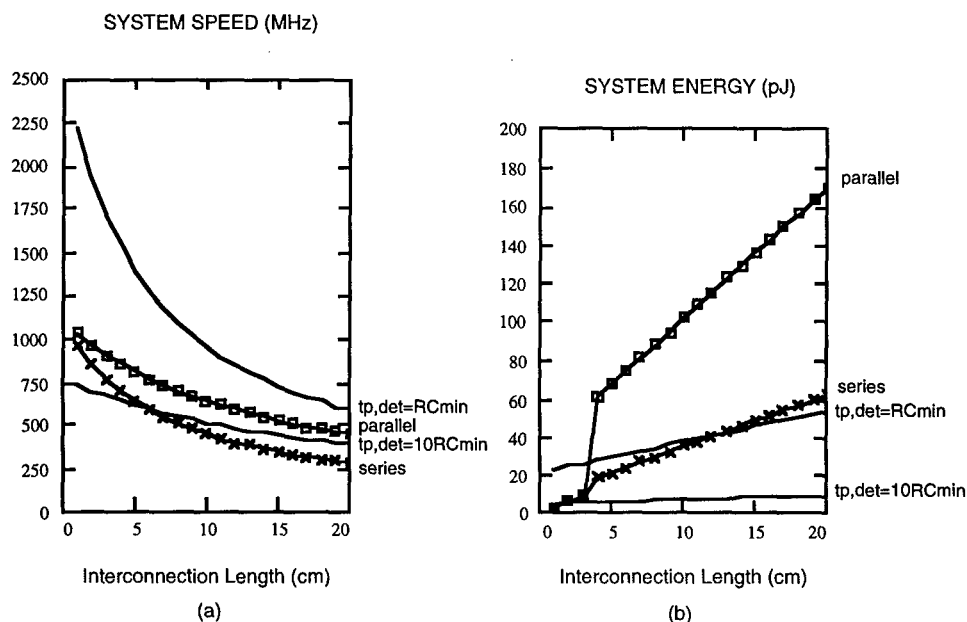


Figure 5-18. Speed and energy comparison between off-chip electrical and VCSEL-based optical interconnects for one-to-one connection: a) speed, b) system (processing plane) energy. Curves related to electrical interconnect are illustrated with symbols. The performance of optical interconnects is illustrated at different detection speeds ($t_{p,det}$) to allow comparison to electrical interconnects running at the same speed.

5.2.5.2.3 Comparison To On-Chip Electrical Interconnects

It can be observed from Fig. 5-19 that the break-even line length is less than a centimeter. Compared to the wafer-scale electrical connections, VCSEL-based optical interconnects yield a drastic speed and energy advantage: In the case of long interconnects, an optical interconnect is 4 times faster and yet 10 times more energy efficient. Note that even for very short interconnects, optics still provides faster operation.

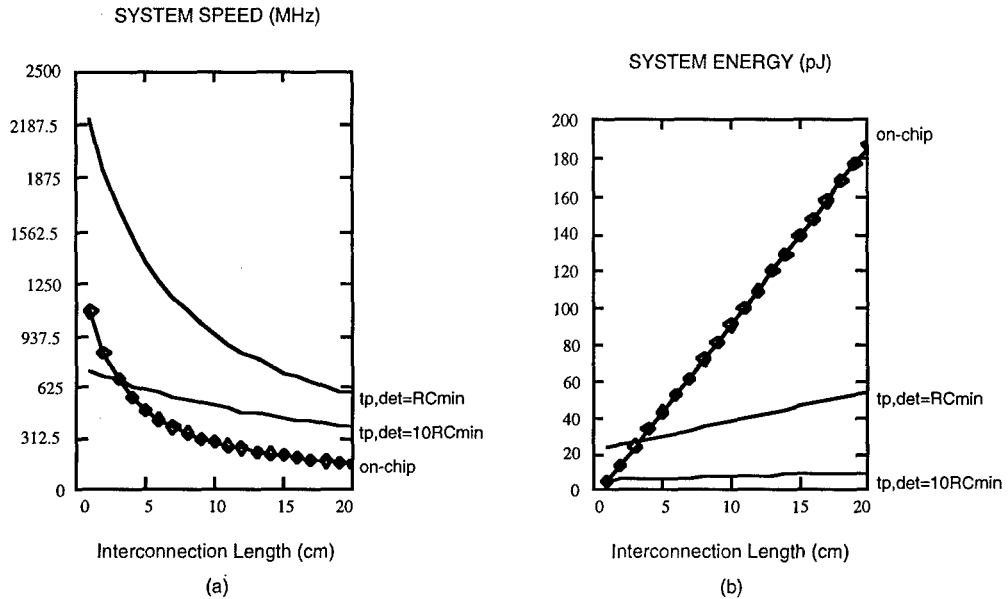


Figure 5-19. Speed and energy comparison between wafer-scale electrical and VCSEL-based optical interconnects for one-to-one connection, a) speed, b) system (processing plane) energy. Curves related to electrical interconnect are illustrated with symbols. The performance of optical interconnects is illustrated at different detection speeds ($t_{p,det}$) to allow comparison to electrical interconnects running at the same speed.

5.2.6 Effects Of Technology Scaling

Here we discuss, from a qualitative point of view, the potential effects of VLSI scaling (reduced minimum feature size) on the different interconnect technologies. From a general point of view, scaling down the VLSI technology usually yields faster circuits that consume less power. It yields reduced parasitics as the transistor areas get smaller. However, it also yields increased resistance of the metal wires on a VLSI chip.

5.2.6.1 Electrical Interconnects

Assume two VLSI chips interconnected via an off-chip interconnection line of impedance Z_0 . As the technology scales down (assuming ideal scaling), the transistor resistance remains constant to a first order. This is because the transconductance increases linearly with the scaling parameter s and the supply voltage V_{DD} decreases linearly with s . Thus, the minimum gate

delay decreases with s , because the resistance remains constant and the parasitic capacitances decrease with s . Therefore, the interconnect technology scaling allows the off-chip interconnection line width to be decreased to a first order by s . If this is done along with VLSI scaling, then the line impedance Z_0 increases by s . An increased line impedance makes a line easier to drive, but because the number of stages in a superbuffer is a logarithmic function of the output load, the effect of increased line impedance on the superbuffer driver performance is minimal. However, because the minimum gate delay scales down as s , the overall superbuffer propagation delay decreases as s . Scaling Z_0 up, however, may increase the propagation delay through the line if the line connects to large lumped capacitances, since the charging capability of the line is proportional to its impedance. If the line is practically unloaded, then the propagation delay is equal to the inherent time of flight, which is independent of the line impedance. Therefore, to a first order approximation, we can say that technology scaling (s) decreases the interconnect propagation delay by s if the line is unloaded, whereas the propagation delay remains roughly constant (or decreases less than s) if the line is loaded.

On the other hand, the energy of data transmission decreases significantly with scaling. The superbuffer and the interconnect capacitances decrease as s . Because the supply voltage also decreases by s and because the dominant capacitive component of the energy is proportional to the square of the supply voltage, the energy requirement scales down as s^3 (Note that the short-circuit component scales as s^3 , and the parallel termination component scales between s and s^3).

In the case of on-chip interconnects, the scaling of a repeater-based interconnect needs to be considered. When an interconnect is scaled down, the line capacitance decreases but the line resistance increases. Thus, the RC time constant of the line remains constant (unless other measures are taken to keep the resistance low). The repeater propagation delay however, scales down with s due to the decreased gate capacitances. The overall interconnection propagation delay, which is a function of both line RC delay and the repeater propagation delay, decreases only as $s^{0.5}$. If, while the line width is scaled to reduce the capacitance, the line thickness is increased to keep line resistance constant, then, the propagation delay decreases with s . This approach, however, is in contradiction with the VLSI trends to increase the number of interconnect layers. Increased line thickness also increases line to line as well as fringe capacitances and results decrease in line capacitance with line width that is less than s .

As in the off-chip interconnect case, the energy requirement of on-chip interconnects scales down as s^3 , due to the decrease in both supply voltage and capacitances. Therefore, the energy-delay product of electrical data transmission scales down between s^3 and s^4 .

5.2.6.2 Optical Interconnects

The energy requirement of optical interconnect is mostly dominated by static currents. As the VLSI technology scales down, the parasitic capacitance driven by a photodiode decreases by s . The gain of the amplifier following the photodiode increases with s because of the increased transconductance and the reduced current. Since the supply voltage also reduces by s , the voltage swing requirement at the output of the photodiode reduces as s^2 . Therefore, for the same speed, the required photocurrent reduces as s^3 because of the reduced capacitances along with a reduced voltage swing. Following the photocurrent, all the transmitter optical power and currents scale as s^3 . Because the voltage supply reduces by s , the energy requirement of optical interconnect scales down by s^4 (better than in the electrical interconnect case) without assuming any improvement in the efficiency of transmitters and interconnect optics. Because we assumed constant speed, the energy-delay product scales as s^4 (as good as the best electrical case). The comparison results of the scaling effects are summarized in Table 5-7.

Table 5-7. Effect of scaling the VLSI on the various interconnection technologies.

Interconnect	Energy/Delay Improvement
Off-Chip	s^3
On-Chip	$< s^4$
Optical	s^4

5.2.7 Conclusions

We have compared both electrical and free-space optical interconnections in terms of speed performance and energy cost for digital transmission in large-scale systems. Free-space optical interconnects using MQW modulators or VCSELs as transmitters offer a significant speed advantage over both off-chip and wafer-scale on-chip electrical interconnects. Compared to the fastest off-chip electrical interconnects (parallel-terminated lines) and for lengths up to a few centimeters, optical interconnects provide as much as twice better speed performance. For medium length off-chip interconnections (5 to 15 cm), the speed advantage of the optical interconnect is about 50%. Finally, compared to long off-chip interconnections, the speed advantage reduces to around 20-40%. Compared to series terminated off-chip electrical interconnects which require much less power than their parallel terminated counterparts, optical interconnects provide at least twice better transmission speed even for very long interconnects (up to 20 cm). MQW-based optical interconnect provides the best one-to-one speed performance while VCSEL-based interconnects follow with small differences. This is due to the smaller driver requirement of the MQW modulators owing to their much smaller transmitter current requirement compared to the VCSEL.

In comparison to wafer-scale VLSI connections, optical interconnects provide increasingly better speed performance as the line length gets longer, reaching 4 times faster transmission speed. In case of long on-chip connections of length up to 2 cm, optical interconnect provides twice faster interconnect speed.

The processing plane (on-chip) energy requirement of one-to-one optical interconnects is generally less than 50 pJ per bit transmitted. For electrical interconnects, it is on the order of several hundred picojoules. The break-even line length between optical and electrical interconnects for equal energy is on the order of a few centimeters. For applications where a large processing plane is needed, optical interconnects provide simultaneous speed and energy advantage over electrical interconnects reaching a combined factor of 20.

Modulator-based optical interconnects require more overall system energy than VCSEL-based optical interconnect due to the requirement for an external system light power source. VCSEL-based optical interconnects offer the best energy requirement, but also require the highest on-chip energy dissipation.

Our analysis in this report did not include the area requirement of the devices in the interconnection. Our main emphasis in this report was to design the interconnects to operate at the fastest speed they can, and then compare the energy requirements. This approach results in large area requirement by the electrical interconnects compared to the optical ones. In the off-chip electrical interconnect case, line drivers as well as line terminators, and in the wafer-scale connection case, the repeaters occupy a substantial amount of area on the processing plane. For comparable area, there is enough room on the optical side, to, for example, improve the light detector circuitry and reduce the input optical power requirement which considerably reduces the overall energy requirement of an optical interconnect. While more complicated detector circuitry may increase the detector propagation delay, the overall link delay may benefit since this may decrease the transmitter propagation delay because of reduced transmitter power requirements.⁵⁻⁴² Also, in some systems, slower than maximum speed optical interconnects can be designed to provide much less energy requirement than their electrical counterparts, while still operating faster than them. In such cases, there is room for increased detector delay, to obtain even better energy efficiency. Improved light detection, possibly with more complicated circuitry is therefore an important issue that should be addressed in the future.⁽⁵⁻⁴³⁾

5.3 MANUFACTURING COST MODELING

The optically enhanced computer system is composed of a variety of technologies that contribute to system cost and yield. This study is to investigate the impact that cost and yield have on the organization of optically interconnected computer architectures. The method used was building analytic abstractions of the economics of the manufacturing process – cost models

– parameterized in terms of the individual fabrication step costs and yield, and the physical and logical architecture characteristics. For example, a physical characteristic is the semiconductor integrated circuit size and an architectural parameter is the number of computational nodes in a parallel processor.

The critical features of the analytic abstraction of the fabrication process shown in Table 5-8 – costs, yield models, and defect rates – come from sources ranging in reliability from measurements, industry standards, and scientific papers, to educated guesses. Thus, the numerical results are not necessarily representative of what occurs in practice, but are specific to the choice of model parameters for the hypothetical processes. For each cost model plot in this report, the Mathematica code that includes the models and parameters is given in the Appendix H.

Table 5-8. Cost, yield and architecture parameters.

PARAMETER	VALUE
Channel width	w = 64 bits
Si size	1 cm ²
Si defect rate	0.3 /cm ²
Wafer process	\$1,250/wafer
GaAs cost	\$ 247/cm ²
GaAs defect	1000/cm ²
VCSEL size	50x10 ⁻⁸ cm ²
GaAs size	1.28x10 ⁻³ /cm ²
CMOS cost	\$ 3.85/cm ²
MCM wire pitch	0.0125 cm
MCM-D cost	\$ 4.80/cm ²
Solder cost	\$ 35/wafer
Solder yield	0.99998/bump
Flip-chip bond	\$ 5/chip
CGH cost	\$ 4/cm ²
Optomechanics	\$10/alignment
Die attach	\$ 0.25
Test cost	\$ 1

We begin with a discussion of semiconductor yield models, and then use them in cost models of active components including CMOS and VCSELs. Next we add cost and yield models for MQW modulators and solder integration to compare monolithic vs hybrid implementations. Finally, we compare VCSEL-based optically interconnected shuffle-exchange networks with the all-electronic alternative. The choice of VCSELs simplifies the optics which makes complete optoelectronic system cost analysis more tractable.

5.3.1 Yield Models

Yield is the probability that a device meets its specifications. In this report, we will deal with fatal defects that prevent the device from operating at all, as opposed to parametric defects, which measure the suitability of the device parameters for a particular application. The former fatal defects are detected by tests during the manufacturing or burn-in procedures.

Once a process has been refined, typically the remaining fatal defects occur probabilistically based on underlying fundamental physical processes. In CMOS silicon processing, the fundamental defect mechanism is particulate contamination. For instance, dirt causes regions to not be exposed during lithography, creating opens or shorts. Dirt also causes material to not be implanted to the proper doping density, or materials to adhere properly.

The two main semiconductor yield models are Poisson and negative binomial in the critical area. Poisson yield occurs when the point defects are independently and identically distributed over the surface of the wafer. Negative binomial yield occurs when the defects are clustered.⁽⁵⁻⁴⁴⁾

Figure 5-20 compares the yield predicted by these two models as a function of chip area. The clustering parameter typical for CMOS processes is from [5-45]. The defect rate is for 1994 from [5-46]. The fill factor of 80% is the fraction of chip area that is susceptible to defects: the critical area.

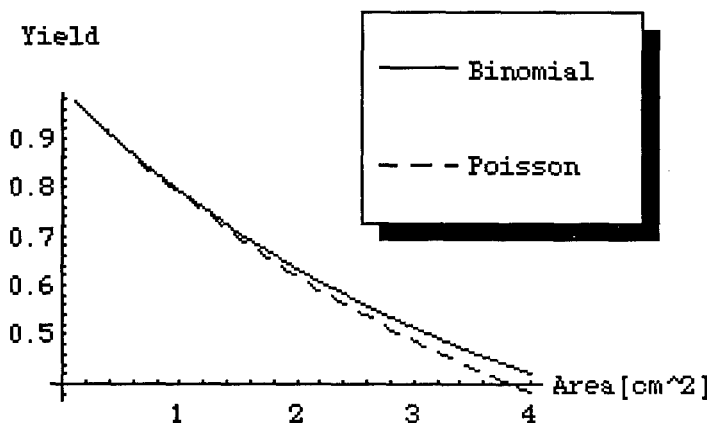


Figure 5-20. Comparison of negative binomial and Poisson yield models for a 1994 CMOS process.

The figure shows that the Poisson yield model predicts a lower a yield than the negative binomial for large chips. Since particulate distribution in MOS processes have been shown to be clustered, we will use the negative binomial for CMOS yield. The Poisson yield model will be used when there is no evidence of clustering, like with VCSELs.

5.3.1.1 CMOS

CMOS production is well-characterized in terms of costs and yields. Besides defect rates used in the previous section, the SIA roadmap also contains the cost of \$4/cm² for industry standard silicon production at high volume. Combining this cost with the negative binomial yield model produces a CMOS cost model that determines cost as a function of chip area as shown in Fig. 5-21.

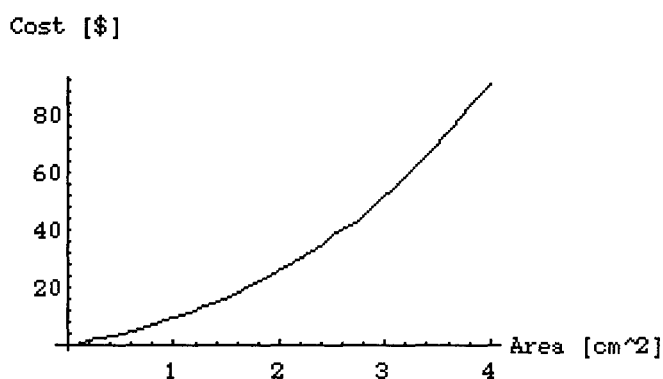


Figure 5-21. The cost of 1994 high volume CMOS as a function of chip area including yield.

The optoelectronic vs MCM system comparison at the end of this report uses the CMOS model along with models for solder bumps and MCMs. In addition, the optoelectronic hardware uses cost models VCSELs, which is addressed in the next section, and for optomechanics.

5.3.1.2 VCSELs

The cost model for VCSELs has a similar form to that for CMOS. However, the epitaxially grown VCSEL wafers are very expensive (\$ 5000 for a 2 inch wafer) and the yield is dependent on the rate of dislocation defects in the crystal lattice.

In a laser, dislocation defects are centers for non-radiative recombination of electrons and holes. Such recombination decreases the cavity gain, which in a highly tuned structure like a VCSEL can be disastrous. The net effect is that the threshold current is greatly increased or the gain may be so low that the device may not lase at all; it will be an LED.

The dislocation density of commercial epitaxial grown wafers is in the 2000 defects per square centimeter range. The hypothesis that these defects lead to device failure has been supported by dark current failure of pin detectors.⁽⁵⁻⁴⁷⁾

The VCSEL cost and yield forms a cost model that can predict the cost of a VCSEL array as a function of the array size as shown in Fig. 5-22. Just as in CMOS, the yield produces a nonlinear increase in cost as a function of the chip size. Note that these chips are much smaller than the previous CMOS chips; the 256 element VCSEL array occupies an area of 0.16 cm².

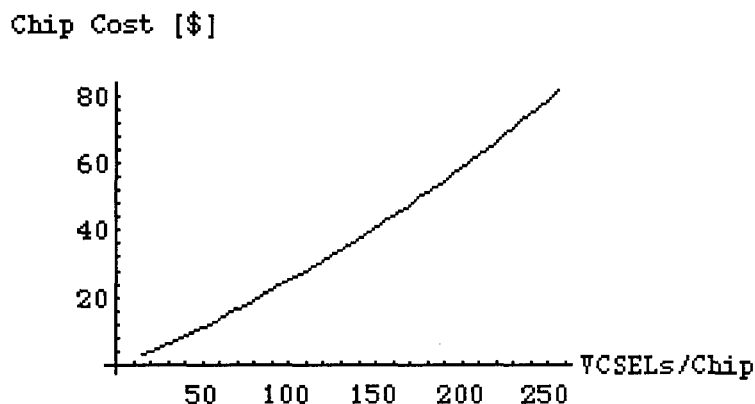


Figure 5-22. Cost of a 250 μm pitch array of 50 μm^2 VCSELS on a single chip as a function of the array size (or number of VCSELS in the array).

A large VCSEL array can be produced using a single chip or multiple chips bonded to an MCM substrate. Finding the optimum chip size is known as the partitioning problem, which requires adding cost models for the MCM and chip-to-MCM connector.

Figure 5-23 depicts the case when equal-size VCSEL chips are first tested and the good ones flip-chip solder-bumped to an MCM to produce a 256-element VCSEL array. There is a separate electrical connection for every VCSEL. The solder bump yield is very high, 0.99998 per bump, but the accurate alignment of the chips to the module with the flip-chip bonder is relatively expensive (\$ 5).

The result is that a 256-element VCSEL array is lower cost to manufacture as a multichip array than monolithically. Recall that the partitioning problem definition requires that the number of chips per module must be an integer. Mathematica interpolates between a set of points to produce the curve in Fig. 5-23. In this case, a 256-element VCSEL array is cheapest to fabricate as 2 or 3 separate chips bonded to a common substrate.

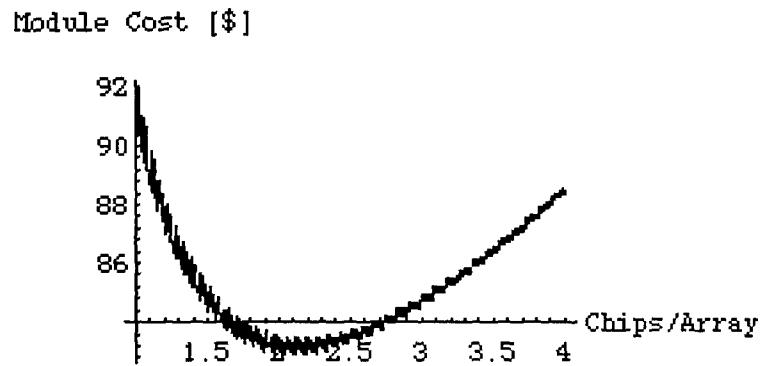


Figure 5-23. Cost of a 256-element VCSEL array as a function of the number of chips solder bumped in the array.

5.3.1.3 Multiple Quantum Well Modulators

This section compares the costs and yields of the various parts of the hybrid process that Lucent uses to solder bumps and flip chip attach GaAs multiple quantum well modulators (MQW) to silicon integrated circuits. The wafer costs are roughly the same as for VCSELs. However, MQW modulator yield is much higher because they are not significantly degraded by dislocation defects; only MBE spitting defects with a density of 100 defects/cm².

Figure 5-24 compares the yield of the silicon, solder bumps, and modulators. The modulators are on a 125 by 62.5 micron pitch grid. The graph shows that the modulators are the yield limiting technology.

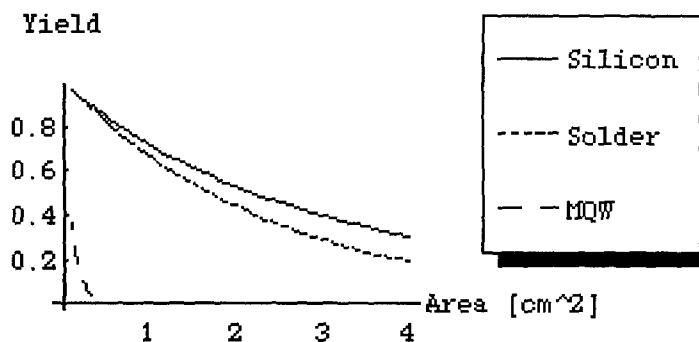


Figure 5-24. Yield of solder bumps, silicon and MQW modulators.

Next we compare the monolithic FET-SEED and the hybrid CMOS-SEED processes in terms of cost. The main yield difference between them is that the GaAs MBE spitting defects will damage the HFETs, while the CMOS circuits have a lower defect rate. The material cost differences are also significant due to the high cost per unit area of the epitaxial GaAs wafers. Figure 5-25 shows that for all reasonable chip sizes, the hybrid integration approach is lowest cost. Note that the low modulator yield means the optimum chip size is quite small. And the high GaAs cost makes the chip cost higher than simple CMOS.

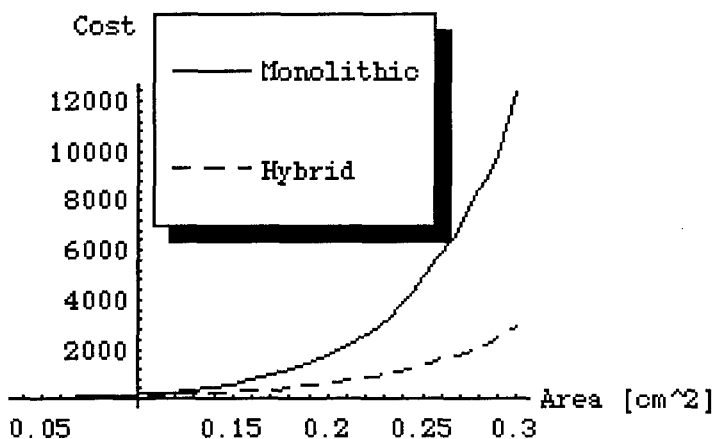


Figure 5-25. Cost comparison of monolithic HFET-SEED GaAs vs. hybrid CMOS-SEED as the chip size increases.

5.3.2 Shuffle Comparison

The shuffle exchange is the underlying data flow graph for a number of important algorithms including switching, bitonic sorting, FFT, Viterbi trellis decoders, and data encryption standard decoders. Thus, low-cost hardware implementations of shuffle exchange networks will have a significant impact on signal processing, communications, and computer systems.

The perfect shuffle possesses many properties that motivate an optical implementation. The regular structure makes the optical system simple; spreading the cost of a lens over many parallel channels. On the other hand, the globality makes the 2-dimensional MCM layout occupy a large area. Because MCM yield is inversely exponential in the area, large layout areas are prohibitively expensive.

The assumptions of the technology comparison were the following. The chips were 1 square centimeter with cost and yield parameters from the SIA Roadmap. The shuffle and exchange channels were 64 bits wide. As shown in Fig. 5-26, the electrical implementation used chips that were solder bumped to a MCM-D substrate that contained the shuffle and exchange channels.

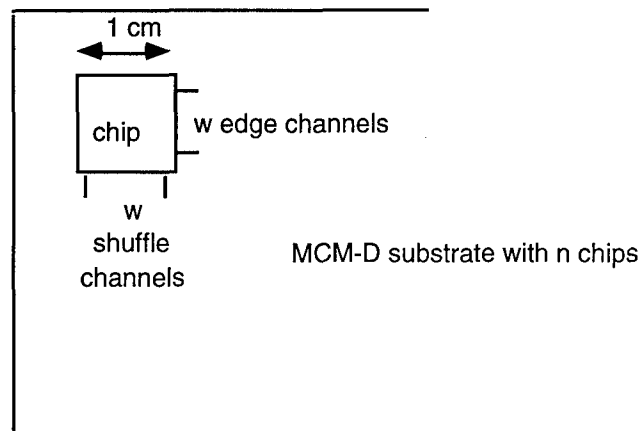


Figure 5-26. The architecture of the MCM-D layout.

The layout of the shuffle-exchange network on an MCM requires $(n/\log n)^2$ area, where n is the number of chips or nodes. There is an additional w^2 expansion of the area for the w -wide parallel channels. There are $w/4$ power/ground pins per chip along with the shuffle and exchange communication channels, and there is no routing underneath the chips. The solder bonds to the chips have a low failure rate.

This process flow shown in Fig. 5-27 illustrates solder bonding known good die to an MCM-D substrate. The silicon costs and defect densities are from the SIA roadmap.⁽⁵⁻⁴⁸⁾ The silicon yield model is from [5-44, 5-45]. The solder cost is the ARPA sponsored goal of \$35 per wafer,⁽⁵⁻⁴⁹⁾ and the die attach cost and MCM cost⁽⁵⁻⁵⁰⁾ are from the MCM literature. The MCM yield is assumed to follow the same model as the silicon, with the effective rate reduced due to the fewer lithography steps.

The optical implementation had a glass MCM substrate for the exchange channels and a simple free-space optical setup for the shuffle channels as shown in Figs. 5-28 and 5-29. To keep the cost of the optical shuffle exchange implementation as low as possible we consider a system with VCSELs and simple optomechanics.

The optical interconnect system begins with a micro-lenslet array fabricated by etching a computer generated hologram (CGH) which cost $\$4/\text{cm}^2$.⁽⁵⁻⁵¹⁾ The VCSEL output propagates through the beamsplitter and until it meets the second CGH, which acts as four macro-lenses to perform the folded shuffle. The fold mirrors reflect the light back to the beamsplitter. The beamsplitter reflects the light back through the micro-lenslet array and onto the detectors. To reduce the optical crosstalk, polarizers and waveplates can be used with a polarizing beamsplitter.

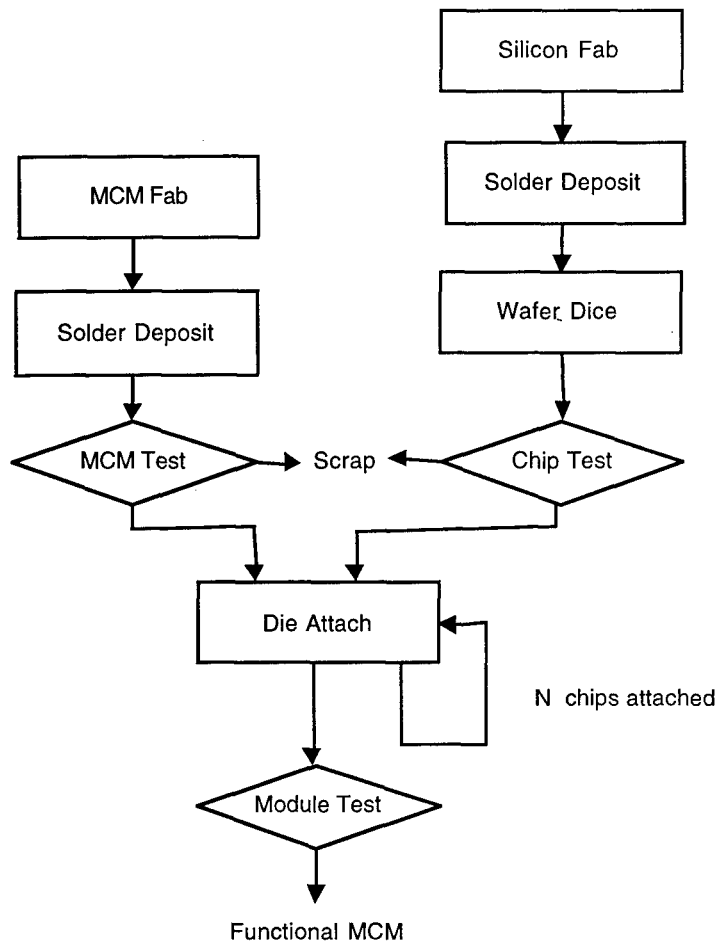


Figure 5-27. Process flow for a module for CMOS chips flip-chip solder-bumped to MCM-D.

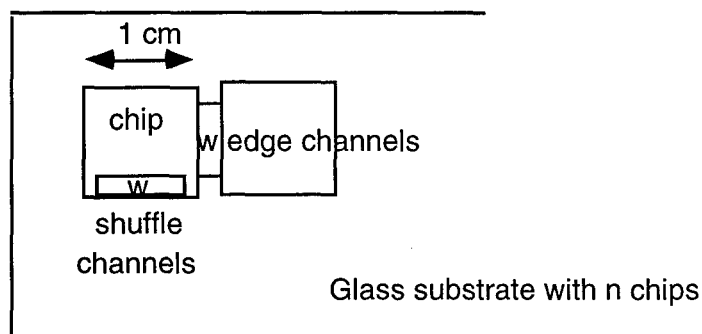


Figure 5-28. Architecture of optoelectronic MCM.

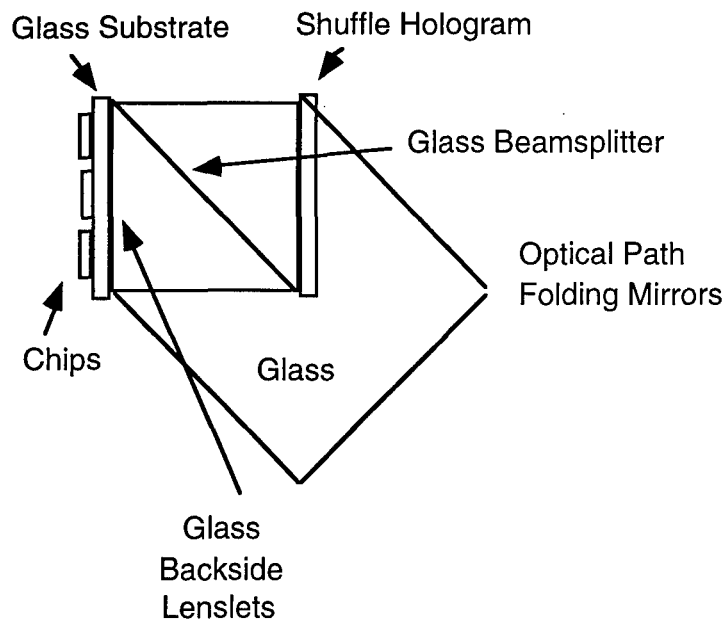


Figure 5-29. Architecture of free-space optical shuffle network.

The beamsplitter and fold mirror cost the same per area as the CGH, and the four optomechanical mounts and alignment costs \$10. It is not a coincidence that the silicon, MCM and the optics cost nearly the same per unit area, since they are made by lithography on similar quality material. The difference with the all-electronic MCM case is that in the optical system, the MCM area is smaller and the yield is not a function of area for the optics.

The cost model addresses solder bonding VCSELs onto commercial silicon using the process sequence shown in Fig. 5-30. The VCSELs are on epitaxially grown GaAs/AlGaAs wafers. After hybridization and testing, the known good die are solder bumped to a glass substrate that provides the exchange power and ground connections. The shuffle signal I/O is through the VCSELs.

The main result of the final system analysis is that for more than 20 nodes, the models predict that the cost of the optical implementation of the shuffle-exchange network will be less than an all-electronic MCM (see Fig. 5-31).

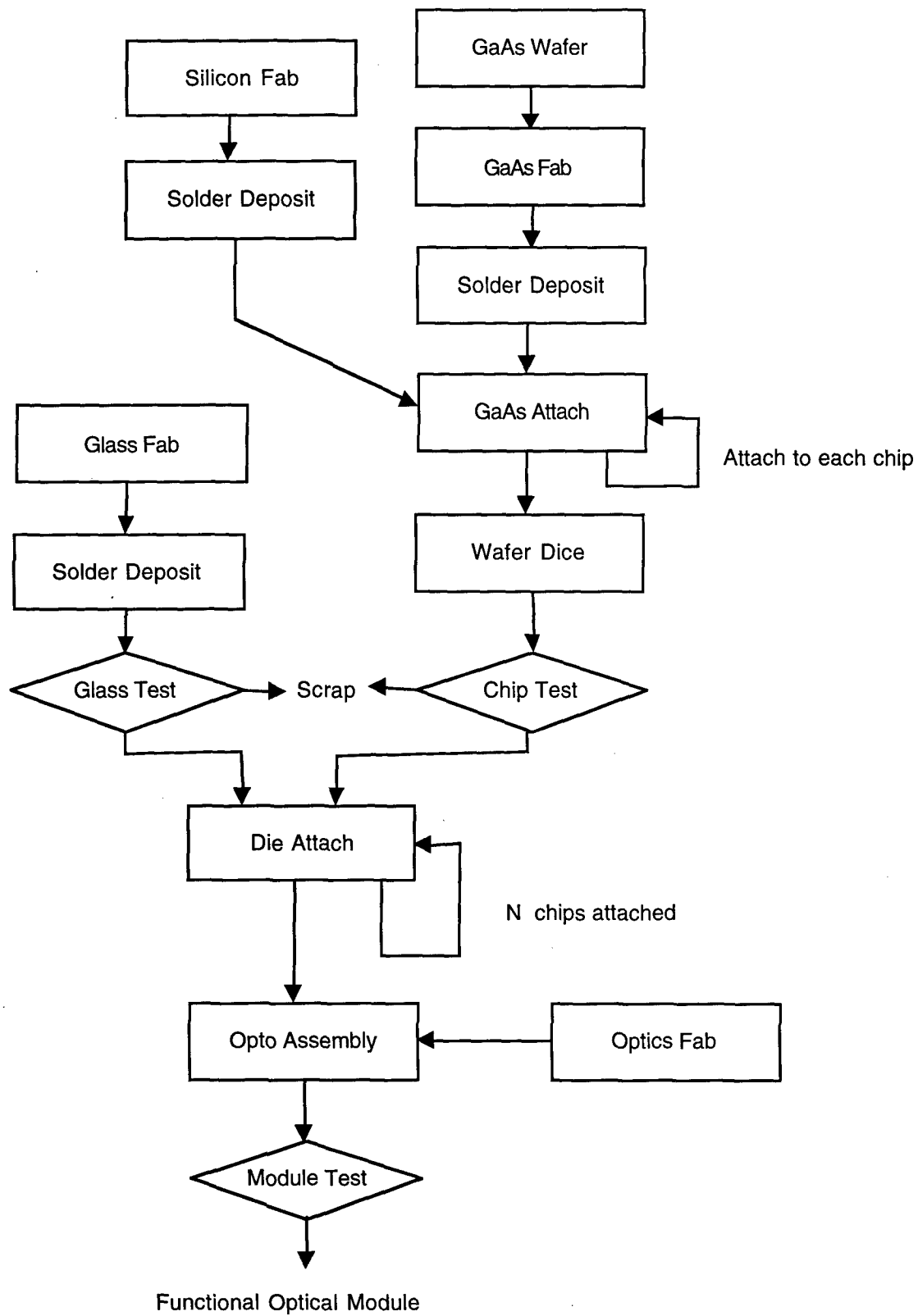


Figure 5-30. Process flow for optoelectronic MCM manufacture.

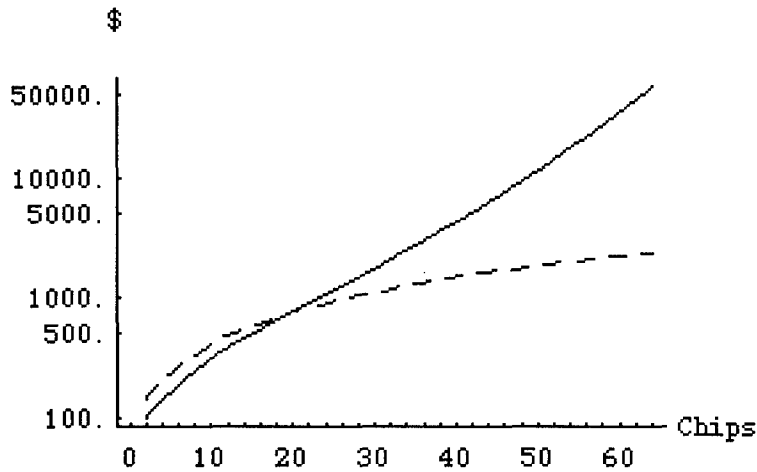


Figure 5-31. Comparison of manufacturing cost of VCSEL/CMOS (dashed line) and CMOS chips (solid line) interconnected by the shuffle-exchange network with optics/MCM or an MCM, respectively.

5.3.3 Conclusions

This program demonstrated that cost models for computer architectures can be constructed that contain fabrication costs and yields as well as system architecture parameters. The models depend on industry standard costs and yield models, as well as physical insight into fabrication processes and device operation. First we built cost models for the fundamental technologies such as CMOS and VCSELs. For the VCSELs, we determined the optimum chip size to fabricate a multichip laser array with 256 lasers. For MQW modulators, the cost models predicted that solder bump hybrid CMOS-SEED was lower cost than the monolithic FET-SEED integration. Then we showed how the individual models can be combined with an architectural model to produce a cost representation of complex systems. Using this representation, we performed a technology tradeoff analysis that determined that an optoelectronic VCSEL/CMOS/MCM/Optics implementation of the shuffle-exchange network is lower cost than the all-electronic CMOS/MCM alternative for greater than 20 computational nodes.

SECTION 6

REFERENCES

SECTION 2

- 2-1 K.W. Goossen, J.A. Walker, L.A. D'Asaro, S.P. Hui, B. Tseng, R. Leibenguth, D. Kossives, D.D. Bacon, D. Dahringer, L.M.F. Chirovski, A.L. Lentine, and D.A.B. Miller, "GaAs MQW modulators integrated with silicon CMOS," *IEEE Photon. Technol. Lett.*, Vol. 7, No. 4, pp.360-362, April 1995.
- 2-2 A.V. Krishnamoorthy, P.J. Marchand, F.E. Kiamilev, and S.C. Esener, "Grain-size considerations for optoelectronic multistage interconnection networks," *Applied Optics*, Vol 31, No. 26, pp. 5480-5507, 1992.
- 2-3 G.C. Marsden, P.J. Marchand, P. Harvey, and S.C. Esener, "Optical Transpose Interconnection System Architectures," *Optics Letters*, Vol. 18, No. 13, pp. 1083-1085, 1993.
- 2-4 O. Kibar, P.J. Marchand, and S.C. Esener, "High-Speed 2-D CMOS Designs of Bypass-and-Exchange Switch Arrays for Free-Space Optoelectronic MINs," presented at the *OSA Topical Meeting on Photonics in Switching*, Salt Lake City, March 1995.
- 2-5 H.B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley, Reading, Mass., 1990, Chap. 5.
- 2-6 W. Lee Hendrick, O. Kibar, P. Marchand, C. Fan, D. Van Blerkom, F. McCormick, I. Cokgor, M. Hansen, and S. Esener, "Modeling and Optimization of the Optical Transpose Interconnection System," *Optoelectronic Technology Center*, Cornell University, September 1995.
- 2-7 A.L.Lentine, K.W. Goossen, J.A. Walker, L.M.F. Chirovsky, L.A.D'Asaro, S.P. Hui, B.T. Tseng, R.E. Leibenguth, D.P. Kossives, D.W. Dahringer, D.D. Bacon, T.K. Woodward, and D.A.B. Miller, "700 Mb/s operation of optoelectronic switching nodes comprised of flip-chip-bonded GaAs/AlGaAs MQW modulators and detectors on silicon CMOS circuitry," Conference on Lasers and ElectroOptics, postdeadline paper CPD11, 1995.

SECTION 3

- 3-1. Code V® is a registered trademark of Optical Research Associates.
- 3-2. G.C. Marsden, P.J. Marchand, P. Harvey, and S. C. Esener, "Optical transpose interconnection system architectures," *Optics Letters*, **18**, 1083-1085, 1993.
- 3-3 P.J. Marchand, A.V. Krishnamoorthy, S.C. Esener, and U. Efron, "Optically augmented 3-D computer: technology and architecture," *First International Workshop on Massively Parallel Processing Using Optical Interconnections*, 133-139, Cancún, Mexico, 1994.
- 3-4. A.V. Krishnamoorthy, P.J. Marchand, F.E. Kiamilev, and S.C. Esener, "Grain-size considerations for optoelectronic multistage interconnection networks," *Applied Optics*, **31**, 5480-5507, 1992.
- 3-5 F. Xu, J.E. Ford, and Y. Fainman, "Polarization-selective computer-generated holograms: design fabrication, and applications," *Applied Optics*, **34**, 256-266, 1995.

- 3-6 F.B. McCormick, T.J. Cloonan, A.L. Lentine, J. M. Sasian, R.L. Morrison, M.G. Beckman, S L. Walker, M.J. Wojcik, S.J. Hinterlong, R.J. Crisci, R.A. Novotny, and H. S. Hinton, "5-Stage free-space optical switching network with field effect transistor self-electro-optic-effect device smart-pixel arrays," *Applied Optics*, **33**, 1601-1618, 1994.
- 3-7 Results were obtained by use of 8×8 segments of two 20×36 lenslet arrays (part no. AOA 400-3.2S) Adaptive Optics Associates, 54 Cambridge Park Drive, Cambridge, Mass. 02140.
- 3-8 Gary C. Marsden, Philippe J. Marchand, Phil Harvey, and Sadik C. Esener, "Optical transpose interconnection system architectures," *Optics Letters* **18**, 1083-1085 (1993).
- 3-9 J. Larry Pezzaniti and Russell A. Chipman, "Angular Dependence Of Polarizing Beam-Splitter Cubes", *Applied Optics* **33**, (no. 10), 1916-1929 (1994).
- 3-10 W. Lee Hendrick, Philippe J. Marchand, Frederick B. McCormick, Ilkan Çokgör, and Sadik C. Esener, "Optical Transpose Interconnection System: System Design and Component Development," OSA Topical Meeting on Optical Computing - Salt Lake City, 3/95.
- 3-11 Pochi Yeh, *Introduction To Photorefractive Nonlinear Optics*, Ch. 3, Wiley Interscience, New York (1993).
- 3-12 P. Yeh, "Generalized model for wire grid polarizers," *SPIE Vol. 307*, 13 (1981).

SECTION 4

- 4-1 M.S. Welkowsky et al, "Status of the Hughes CCD-addressed Liquid Crystal Light Valve," *Opt. Eng.* **26**, 414 (1987).
- 4-2 U. Efron et al, "The CCD-addressed Liquid Crystal Light Valve, - an update," *SPIE Vol. 1455*, 237-247 (1991).
- 4-3 K. Sayyah, M.S. Welkowsky, P.G. reif, and N.W. Goodwin, "High performance single crystal silicon liquid crystal light valve with god image uniformity," *Applied Optics* **28**, (no. 22), 4748-4756 (1989).
- 4-4 Z-Epoxy (ZXUV.102.3012Au) from Zymet, 7 Great Meadow Ln., E. Hanover, NJ 07936

SECTION 5

- 5-1 G.Marsden, P. Marchand, P. Harvey, and S. Esener, "Optical Transpose Interconnection System architectures," *Optics Letters*, **18**, 13, 1 July 1993, 1083-1085.
- 5-2 A. Krishnamoorthy, P. Marchand, F. Kiamilev, and S. Esener, "Grain-size considerations for optoelectronic multistage interconnection networks," *Applied Optics*, **31**, 26, September 1992, 5480-5507.
- 5-3 M. Feldman, S. Esener, C. Guest, and S. Lee, "Comparison between electrical and free-space optical interconnects based on power and speed considerations," *Applied Optics*, **27**, 9, May 1988, 1742-1751.
- 5-4 F. Kiamilev, P. Marchand, A. Krishnamoorthy, S. Esener, and S. Lee, "Performance comparison between optoelectronic and VLSI multistage interconnection networks," *Journal of Lightwave Technology*, **9**, 12 December 1991, 1674-1692.
- 5-5 H. Hinton, *An Introduction to photonic switching fabrics*, Plenum Press Ed., 1993.
- 5-6 S. Hinterlong, "High Performance SEED-based optical computing systems," *1995 ARPA MTO Program Review*, Big Sky, Montana, July 1995.

- 5-7 W. Hendrick, O. Kibar, P. Marchand, C. Fan, D. Van Blerkom, F. McCormick, I. Cokgor, M. Hansen, and S. Esener, "Modeling and optimization of the Optical Transpose Interconnection System," Optoelectronic Technology Center, Program Review, Cornell University, September 1995.
- 5-8 O. Kibar, P. Marchand, and S. Esener, "High-speed CMOS switch designs for free-space optoelectronic MINs," submitted to *IEEE transactions on VLSI*.
- 5-9 F.T. Leighton, *Introduction to Parallel Algorithms and Architectures*, Morgan Kaufmann Publishers, 1992.
- 5-10 K. Ghose, and K.R. Desai, "Hierarchical Cubic Networks," *IEEE Transactions on Parallel and Distributed Systems*, 6, 4, April 1995, 427--435.
- 5-11 K. Batcher, "Sorting Networks and their applications," Proceedings of the AFIPS Spring Joint Conference, 32, 1968. 307-314.
- 5-12 M. Ajtai, J. Komlos, and E. Szemerédi, "Sorting in clogn parallel steps," *Combinatorica*, 3, (1983), 1-19.
- 5-13 E. Upfal, "An $O(\log N)$ deterministic packet routing scheme," *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, Seattle, WA, May 1989, 241-250.
- 5-14 F.T. Leighton, and B. Maggs, "Expanders might be practical: Fast algorithms for routing around faults on multibutterflies and randomly-wired splitter networks," *IEEE Transactions on Computers*, 41, 5, (May 1992), 1-10.
- 5-15 E. Brewer, F.T., Chong and F.T. Leighton, "Scalable expanders: exploiting hierarchical random wiring," *Proceedings of the 26th Annual ACM Symposium on the Theory of Computing*, Montreal, Quebec, May 1994, 144-152.
- 5-16 J.W. Goodman, F.I. Leonberger, S.Y. Kung, and R.A. Athale, "Optical Interconnections for VLSI Systems," *Proc. IEEE* 72, 850 (1984).
- 5-17 L.A Bergman *et al.*, "Holographic Optical Interconnects in VLSI," *Opt. Eng.* 25, 1109 (1986).
- 5-18 W.H. Wu *et al.*, "Implementation of optical Interconnections for VLSI," *IEEE Trans. Electron Devices* ED-34, 706 (1987).
- 5-19 R K. Kostuk, J.W. Goodman, and L. Hesselink, "Optical Imaging Applied to Microelectric Chip-to-Chip Interconnections," *App. Opt.* 24, 2851 (1985).
- 5-20 F.B. McCormick, "Free-space interconnection techniques," in *Photonics in Switching*, Vol. II, John E. Midwinter, ed., (Academic Press, New York, 1993), pp. 169-250.
- 5-21 F. Kiamilev, P. Marchand, A. Krishnamoorthy, S. Esener, and S.H. Lee, "Performance comparison between optoelectronic and VLSI multistage interconnection networks," *IEEE Journal of Lightwave Technology* 9, 1674-1692, December 1991.
- 5-22 A. Krishnamoorthy, P. Marchand, F. Kiamilev, K. S. Urquhart, S. Esener, and S. H. Lee, "Grain-size study for a 2-D shuffle-exchange optoelectronic multistage interconnection network," *Applied Optics* 31, 5480-5507, September 1992.
- 5-23 K. Urquhart, P. Marchand, Y. Fainman, and S. H. Lee, "Diffractive Optics applied to free-space optical interconnects," *Applied Optics* 33, 3670-3682, June 1994.
- 5-24 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, No.9, May 1988.
- 5-25 G. Yayla, P. Marchand, and S. Esener, "Energy Requirements and Speed Analysis of Digital Electrical and Free-Space Optical Interconnections," to be published in Optical

Interconnections and Parallel Processing: The Interface, P. Berthome and A. Ferreira Eds., Kluwer Academic Publishers, December 1997.

- 5-26 H.J. Veendrick, "Short-circuit dissipation of static CMOS circuitry and its impact on the design of buffer circuits," *IEEE J. Solid-State Circuits*, **SC-19**, no. 4, pp.468-474, Aug. 1984.
- 5-27 N.C. LI, G.L. Haviland, A.A. Tuszynski, "CMOS tapered buffer," *IEEE J. Solid-State Circuits*, vol. 25, no. 4, pp.1005-1008, Aug. 1990.
- 5-28 R. Geiger, P. Allen, and N. Stroder, *VLSI design techniques for analog and digital circuits* (McGraw Hill, New York, 1990), pp. 590-593.
- 5-29 H.B. Bakoglu, *Circuits, Interconnections and Packaging for VLSI*, (Addison-Wesley Publishing Company), 1990.
- 5-30 C. Fan, D.W. Shih, M.W. Hansen, S.C. Esener, and H.H. Wieder, "Quantum-confined stark effect modulators at 1.06 μm on GaAs," *IEEE Photon. Technol. Lett.* **5**, 1383-1385 (1993).
- 5-31 D.S. Chemla, D.A.B. Miller, P.W. Smith, A.C. Gossard, and W. Wiegmann, "Room temperature excitonic nonlinear absorption and refraction in GaAs/AlGaAs multiple quantum well structures," *IEEE J. Quantum Electron.* **QE-20**, 265-275 (1984).
- 5-32 A.V. Krishnamoorthy et al., "Operation of a single-ended 550 Mbits/sec, 41 fJ, hybrid CMOS/MQW receiver-transmitter," *Electronics Letters*, **32**, 8, 764-765, 11 April 1996.
- 5-33 B. Pezeshki, D. Thomas, and J.S. Harris, Jr., "Optimization of modulation ratio and insertion loss in reflective electroabsorption modulators," *App. Phys. Lett.*, **57**, no. 15, pp. 1491-2, Oct. 1990.
- 5-34 P.J. Stevens and G. Parry, "Limits to normal incidence electroabsorption modulation in GaAs/(GaAl) as multiple quantum well diodes," *J. Lightwave Technol.* **7**, 1101-1108 (1989).
- 5-35 T.H. Wood, J.Z. Pastalan, C.A. Burrus, Jr., B.C. Johnson, B.I. Miller, J. L. deMiguel, U. Koren, and M. G. Young, "Electric field screening by photogenerated holes in multiple quantum wells: A new mechanism for absorption saturation," *App. Phys. Lett.* **57**, 1081-1083 (1990).
- 5-36 L. Coldren, S. Corzine, R. Feels, A.C. Fonard, K.K. Law, J. Merz, J. Scott, R. Simes, and R.H. Yan, "High efficiency vertical cavity lasers and modulators," *Proc. Soc. Photo-Opt. Instrum. Eng.* **1362**, 79-92 (1990).
- 5-37 J. Jewell, G. Olbright, "Vertical cavity surface emitting lasers," *IEEE J. Quantum Electron.* **27**, 1332-1346 (1991).
- 5-38 D.B. Young, J.W. Scott, F.H. Peters, M.G. Peters, M.L. Majewski, B. J. Thibeault, S.W. Corzine, and L.A. Coldren, "Enhanced Performance of Offset-Gain High-Barrier Vertical-Cavity Surface-Emitting Lasers," *IEEE J. Quantum Electronics*, **29**, no. 6, pp. 2013-2022, June 1993.
- 5-39 L. Colren, S. Corzine, R. Feels, A.C. Fonard, K.K. Law, J. Merz, J. Scott, R. Simes, and R. H. Yan, "High efficiency vertical cavity lasers and modulators," *Proc. Soc. Photo-Opt. Instrum. Eng.* **1362**, 79-92 (1990).
- 5-40 R. Geels and L. Coldren, "Submilliwatt threshold vertical cavity laser diodes," *App. Phys. Lett.* **57**, 1605-1607 (1990).
- 5-41 C. Fan, B. Mansoorian, D.A. Van Blerkom, M. W. Hansen, V.H. Ozguz, S.C. Esener, and G.C. Marsden, "A Comparison of Transmitter Technologies for Digital Free-Space Optical Interconnections," *Applied Optics*, **34**, No. 17, pp. 3103-3115, June 1995.

- 5-42 Daniel A. Van Blerkom, Osman Kibar, Chi Fan, Philippe J. Marchand and Sadik C. Esener, "Power Optimization of Digital Free-Space Optoelectronic Interconnections," *OSA Topical Meeting on Spatial Light Modulators*, Lake Tahoe, March 1997, and to be published in *Journal of Lightwave Technology*.
- 5-43 Daniel Van Blerkom, Chi Fan, Matthias Blume, and Sadik C. Esener, "Optimization of Smart Pixel Receivers," *IEEE LEOS Summer Topical Meeting on Smart Pixels*, Keystone, August 1996, and to be published in *Journal of Lightwave Technology*.
- 5-44 C.H. Stapper, Defect density distribution for LSI yield calculations, *IEEE Trans. Electron Devices*, vol. ED-20, p. 665-657, July 1973.
- 5-45 J.A. Cunningham, The use and evaluation of yield models in integrated circuit manufacturing, *IEEE Trans. Semicond. Manuf.* Vol. 3, No. 2, pp 60-71, May 1990.
- 5-46 Semiconductor Technology Workshop Conclusions, Semiconductor Industry Association, 1993.
- 5-47 C.W. Stirk, CAD tool for low-cost optoelectronic and optomechanical manufacturing processes, *Proc. Electronic Components Techn. Conf.*, May 30, 1996.
- 5-48 The National Technology Roadmap for Semiconductors, Semiconductor Industry Association, 1994.
- 5-49 C.L. Lassen, Global and commercial developments with flip chip technology, *IEEE ECTC*, 1056-1058 (1996).
- 5-50 D A. Doane and R. D. Franzon, eds., *Multichip Module Technologies and Alternatives--The Basics*, Van Nostrand Reinhold, New York, 1993.
- 5-51 R. TeKolste, *Proceedings of the MANTECH Conference on Optical Manufacturing*, Redstone Arsenal, 1994.

Appendix A

1. PROBABILITY, P_i , OF INPUT ARRIVING AT STAGE I:

stage i	probability = p_i
0	1
1	$4/5$
2	$2/3$
3	$4/7$
4	$1/2$
5	$4/9$
6	$2/5$
7	$4/11$
8	$1/3$
9	$4/13$
10	$2/7$
11	$4/15$
12	$1/4$

2. THE LENGTH OF WIRE BETWEEN PARTNER HALF-SWITCHES AT STAGE I IN TERMS OF Δ (I.E. LENGTH OF WIRE AT STAGE I IS = $F_i * \Delta$):

i	$f_i (S=8)$	$f_i (S=12)$
0	1	1
1	2	4
2	2	4
3	1	2
4	1	2
5	2	1
6	2	1
7	1	4
8	--	4
9	--	2
10	--	2
11	--	1

Appendix B

SUPERBUFFER DESIGN

A superbuffer, that is, a chain of inverters with increasing sizes is widely used to reduce propagation delays when driving large capacitive loads. In this appendix, we use some of the results of a superbuffer design reported in [5-27]. A superbuffer circuit with a tapering factor β is illustrated in Fig. (B-1).

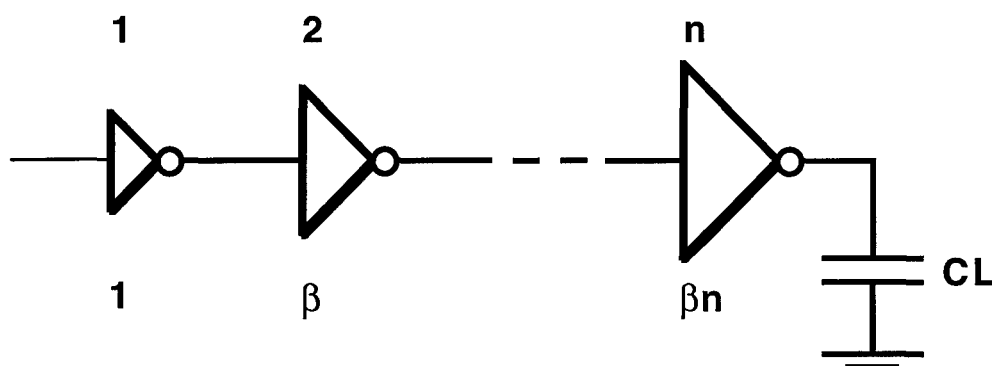


Figure B-1. An n -stage superbuffer used to drive large capacitive loads.

The minimized propagation delay through the superbuffer is given as:⁽⁵⁻²⁷⁾

$$t_{sb,p} = n \cdot RC_{\min} \cdot \alpha \quad (B-1)$$

where n : number of inverter stages, RC_{\min} : minimum logic RC time constant, and α is defined as:

$$\alpha \equiv 1 + p(\beta - 1), \quad (B-2)$$

where β : tapering factor, $p = \frac{C_{\min,o}}{C_{\min,i} + C_{\min,o}}$, $C_{\min,i}$ = minimum inverter input capacitance,

$C_{\min,o}$ = minimum inverter output capacitance.

The number of stages n (excluding the minimum first stage) can be calculated as:⁽⁵⁻²⁷⁾

$$n = \frac{1}{\ln \beta} \ln \left\{ \frac{C_L}{C_{\min,i}} \right\} - 1 \quad (B-3)$$

where C_L is the output load capacitance of the last superbuffer stage.

Because the signals within the superbuffer can be treated using the lumped capacitor approximation, the rise time t_r of the signals in the superbuffer is approximately twice the propagation delay of a single stage:

$$t_r = 2 \cdot RC_{\min} \cdot \alpha \quad (\text{B-4})$$

The output capacitance of the last superbuffers stage, which is β^{n-1} times larger than minimum, is calculated as:

$$C_{sb,o} = \beta^{n-1} \cdot C_{\min,o} \quad (\text{B-5})$$

The total capacitance of the superbuffers on the signal path is the sum of the input and output capacitances of all the stages:

$$C_{sb} = (C_{\min,i} + C_{\min,o}) \sum_{i=1}^{n-1} \beta^i = (C_{\min,i} + C_{\min,o}) \beta \frac{\beta^{n-1} - 1}{\beta - 1} \quad (\text{B-6})$$

The effective transconductance of the superbuffers, which we define as the sum of the transconductances of all the stages, is calculated as:

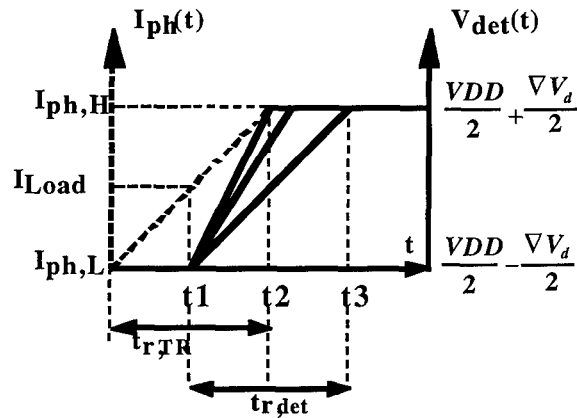
$$k_{eff} = k_{\min} \sum_{i=1}^{n-1} \beta^i = k_{\min} \frac{(\beta^n - \beta)}{(\beta - 1)} \quad (\text{B-7})$$

where k_{\min} is the transconductance parameter of minimum geometry inverter.

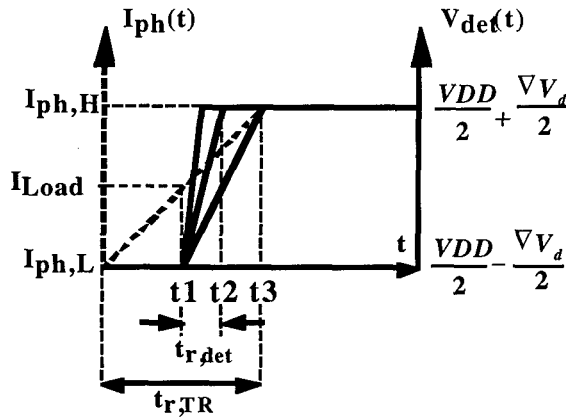
Appendix C

DETECTOR REQUIREMENTS

In this appendix, we express the detector photocurrent dynamic range requirement as a function of detector parasitics and speed of operation. Figure 5-15 illustrates the detector circuit considered. Figure C-1 shows the presumed detector transfer characteristic, where the optical input transition between a low and a high intensity level is linear, thus resulting in a linear photocurrent signal transition through the photodiode.



a: $t_{r,det} \geq \frac{t_{r,TR}}{2}$



b: $t_{r,det} < \frac{t_{r,TR}}{2}$

Figure C-1. Photodetector input/output waveforms used in the calculations. (a) The transmitter rise time is less than twice the detector rise time, (b) otherwise.

If the optical interconnection system does not alter the timing characteristics and the photodiode device limits are not pushed, the photocurrent rise time is equal to the rise time of the optical signal at the transmitter output. This is illustrated in Figure C-1, where $t_{r,TR}$ denotes the transmitter (or photocurrent) 0% to 100% rise time. Figure C-1 illustrates the detector operation where the photodiode output voltage rise time $t_{r,det}$ is larger than or equal to the half transmitter rise time. Applying the capacitance charging equation to the photodiode output node yields:

$$\frac{\frac{V_{DD} + \nabla V_d}{2} - \frac{V_{DD} - \nabla V_d}{2}}{C_{det}} dV = \frac{1}{C_{det}} \left[\int_{t_1}^{t_2} [i_1(t) - I_L] dt + \int_{t_2}^{t_3} [i_2(t) - I_L] dt \right] \quad (C-1)$$

where

$$i_1(t) = I_L + \frac{t - t_1}{t_2 - t_1} (I_{ph,H} - I_L)$$

$$i_2(t) = I_{ph,H} - I_L$$

and ∇V_d is the photodiode output voltage swing, C_{det} is the total photodiode output capacitance and I_{Load} is the photodiode load current. Integrating Eq. (C-1) and using the definitions of rise times (in Fig. C-1a) yields:

$$I_{ph,H} - I_{ph,L} = \frac{2\nabla V_d \cdot C_{det}}{t_{r,det} - \frac{t_{r,TR}}{4}} ;$$

$$t_{r,det} \geq \frac{t_{r,TR}}{2} \quad (C-2)$$

Similarly, for $t_{r,det} < t_{r,TR}/2$:

$$\frac{\frac{V_{DD} + \nabla V_d}{2} - \frac{V_{DD} - \nabla V_d}{2}}{C_{det}} dV = \frac{1}{C_{det}} \left[\int_{t_1}^{t_2} [i_3(t) - I_L] dt \right] \quad (C-3)$$

where ;

$$i_3(t) = I_L + \frac{t - t_1}{t_3 - t_1} (I_{ph,H} - I_L)$$

Integrating with the help of Fig. C-1b results in:

$$I_{ph,H} - I_{ph,L} = 2\nabla V_d \cdot C_{det} \frac{t_{r,TR}}{t_{r,det}}$$

$$t_{r,det} < \frac{t_{r,TR}}{2} \quad (C-4)$$

In Fig. C-2, we plotted Eq. (C-2) and Eq. (C-4). As observed from Fig. C-2, operating the detector at a shorter rise time than that of the transmitter requires increasingly large optical power dynamic range, whereas operating the detector slower than the transmitter results only in a small power saving. Therefore, operating at equal transmitter and detector rise times is nearly optimal in terms of speed and energy. This is the region of operation defined by Eq. (C-2).

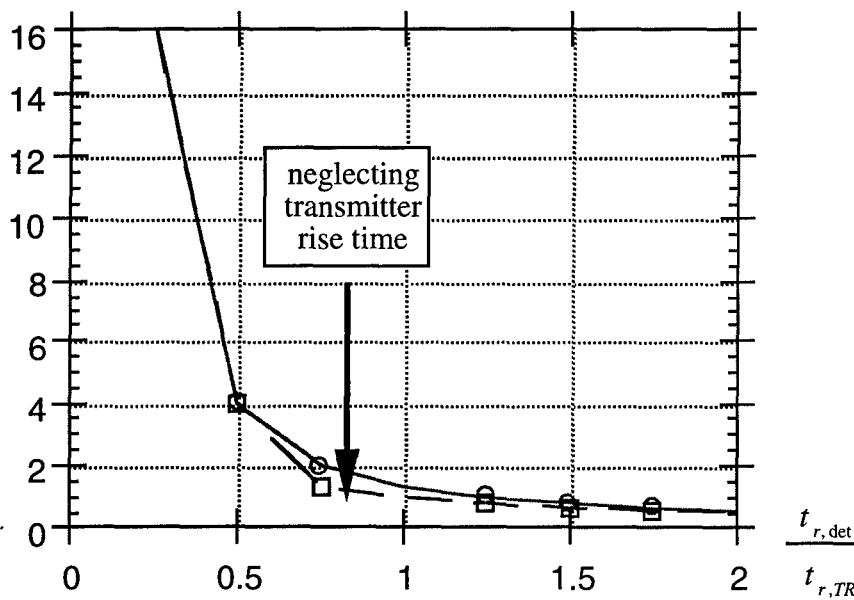


Figure C-2. Plot of Eq. (C-2) and Eq. (C-4).

Also plotted in Fig. C-2 is Eq. (C-2) when neglecting $t_{r,TR}$. We observe that in the interested region of operation, neglecting the rise time of the input signal to the detector causes only a small error while simplifying the calculations. Therefore, for practical purposes, we can assume that the photocurrent dynamic range DR_I can be represented by:

$$DR_I = I_{ph,H} - I_{ph,L} = \frac{2\nabla V_d \cdot C_{det}}{t_{r,det}} \quad (C-5)$$

and the photodiode propagation delay can be approximated as

$$t_{p,det} \approx \frac{t_{r,det}}{2} \quad (C-6)$$

Appendix D

TRANSMITTER AND RECEIVER STEADY-STATE CURRENTS DUE TO AMPLIFICATION

Consider the interconnection model illustrated in Fig. 5-15, the first inverter of the superbuffer at the transmitter site has to amplify its input signal (driven by a minimum logic gate) from the VLSI supply level to the transmitter supply level. This is generally around 10 V to achieve an acceptable optical modulation depth. Since this value is only a few times larger than the VLSI supply levels (3-5V), a single inverter is sufficient to perform the amplification. Similarly, at the receiver site, the thresholding inverter consumes steady-state currents due to the amplification of the photodiode output voltage whose swing is limited to around 330 mV by the clamping diodes.

In any case, a steady-state current results from the fact that the input voltage swing of an inverter is limited around half supply level. If the input voltage swings by an amount around mid-supply level, then it is possible to show that the average steady state current through the receiver site inverter is:

$$I_{RC} = \frac{k_{\min}}{2} (V_{DD}/2 - \Delta V - V_T)^2 \quad (D-1)$$

For the receiver site, where we assumed 330 mV voltage swing, $\Delta V = \Delta V_d/2 = 165$ mV. For the transmitter site, we replace V_{DD} in Eq. (D-1) with V_{TR} :

$$I_{TR} = \frac{k_{\min}}{2} \left(\frac{V_{TR}}{2} - \Delta V - V_T \right)^2 \quad (D-2)$$

where $\Delta V = V_{DD}/2$.

APPENDIX E

ESTIMATION OF OPTICAL TIME-OF-FLIGHT DELAY

We assume that two points on a plane separated by L_{int} can be interconnected optically via an optical element located at a distance L_{int} from the plane in the third dimension (see Fig.E-1). Thus, the length of the optical path between the two points is calculated as;

$$L_{opt} = 2\sqrt{L_{int}^2 + \left(\frac{L_{int}}{2}\right)^2} = 2.2L_{int} \quad (E-1)$$

which results in an optical time-of-flight delay of:

$$t_{fopt} = \frac{L_{opt}}{c} \quad (E-2)$$

where c is the speed of light propagation in the medium.

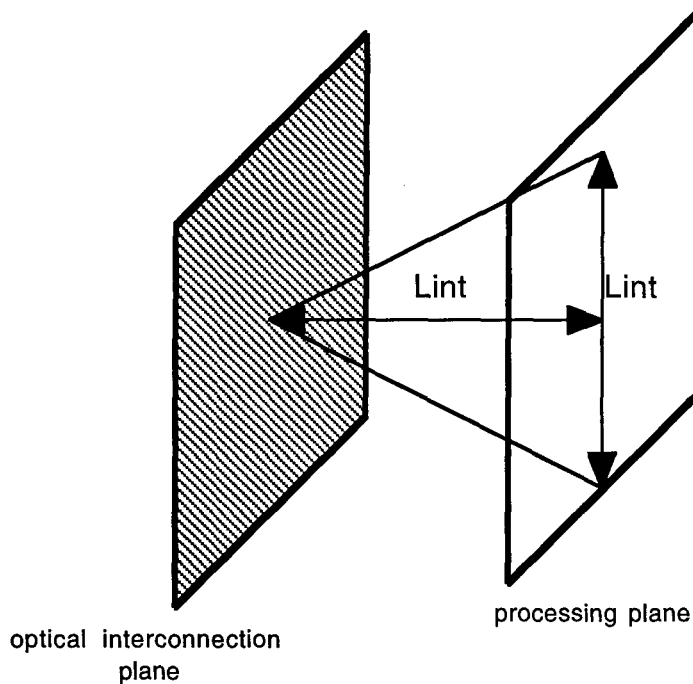


Figure E-1. Geometrical assumptions of the optical interconnect scheme.

Appendix F

MQW MODULATOR DRIVER DESIGN

In this appendix, we design an inverter stage to drive the MQW light modulator, as illustrated in Fig. F-1. The size of the inverter depends on the MQW current because of optical absorption, device, integration parasitics, and speed of operation. The MQW modulator input-to-output optical power efficiency (for high and low voltage states) is given as:⁽⁵⁻⁴¹⁾

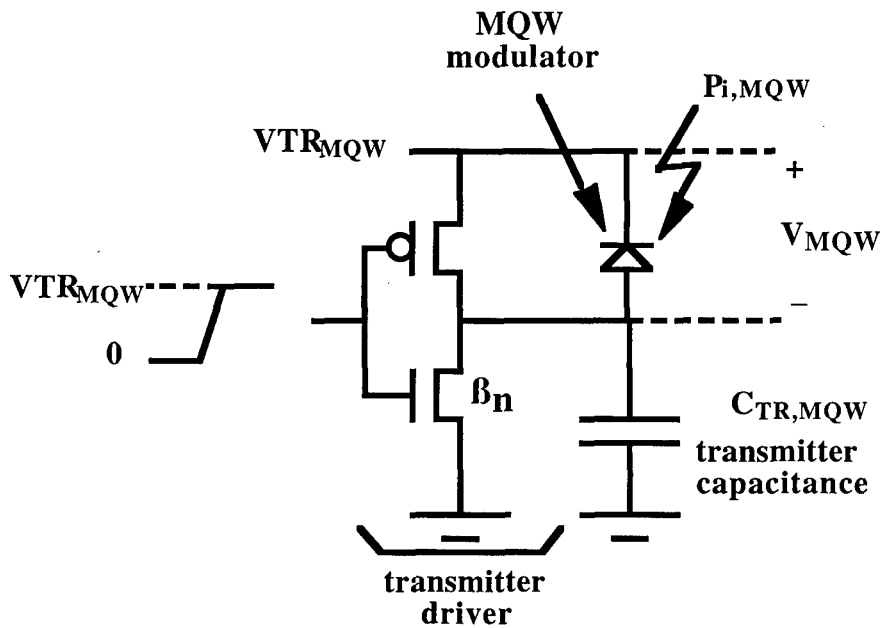


Figure F-1. CMOS MQW modulator driver circuit.

$$\eta_{MQW,H} = \frac{k(0)K_m}{1 + \frac{P_{i,MQW}}{A_{MQW}I_S(V_H)}} ; V_{MQW}=V_H \quad (F-1)$$

$$\eta_{MQW,L} = \frac{k(0)}{1 + \frac{P_{i,MQW}}{A_{MQW}I_S(0)}} ; V_{MQW}=V_L=0 \quad (F-2)$$

where V_H and V_L are the high and low level voltages across the modulator, $k(0)$ is the absorption slope at zero voltage, K_m is the absorption slope ratio, $P_{i,MQW}$ is the input optical power to the

modulator, A_{MQW} is the modulator area, and $I_S(V)$ is the saturation intensity as a function of the modulator voltage. To make the efficiency independent of the input optical power or modulator area, we will make the following practical assumption:

$$\frac{P_{i,MQW}}{A_{MQW} \times I_S(V_H)} = 0.2 \quad (F-3)$$

Note that reducing the above ratio further results in a larger modulator area for a given power with only a small gain in efficiency. Therefore, when the output power requirement from the modulator increases, the area of the modulator is also increased according to Eq. (F-3) in order to keep the modulator efficiency constant. Using Eq. (F-3) in Eq. (F-1) and Eq. (F-2) yields:

$$\eta_{MQW,H} = 0.83 \cdot k(0) \cdot K_m \quad (F-4)$$

$$\eta_{MQW,L} = 0.9 \cdot k(0) \quad (F-5)$$

On the other hand, the modulator high level absorption current can be estimated as:

$$I_{MQW,H} = r_{MQW} \cdot P_{i,MQW} \cdot \eta_{MQW,H} \quad (F-6)$$

where r_{MQW} is the modulator responsivity. The NMOS transistor of the driver inverter should be large enough to sink this current in order to satisfy the DC condition $V_{MQW} \geq V_H$. Equating Eq. (F-6) to the NMOS current in linear region, and solving for NMOS transconductance yields:

$$k_n^{DC} = \frac{I_{MQW,H}}{(V_{TR} - V_T)V_{dslow} - 0.5 \cdot V_{dslow}^2} \quad (F-7)$$

where V_{TR} is the transmitter power supply voltage, V_T the transistor threshold voltage, and $V_{dslow} = V_{TR} - V_H$.

In addition to the DC requirement, the driver should also satisfy the switching speed, or AC requirement. If the NMOS transistor is in the linear region of operation during switching, its resistance can be estimated as:

$$R_{NMOS} = \frac{2}{k_n^{AC} \cdot (V_{TR} - V_T)} \quad (F-8)$$

where we included a factor of two to better approximate the transistor resistance, since during switching, the average value of the input voltage is less than V_{TR} , and the PMOS current confronts the NMOS current. The rise time of the signal at the driver output can then be estimated as:

$$t_{r,dr} = 2.3R_{NMOS}C_{TR} \quad (F-9)$$

where C_{TR} is the total transmitter capacitance composed of:

$$C_{TR} = C_{bond} + A_{MQW}C_{MQW} + \frac{k_n^{AC}}{k_{min}} C_{min,o} \quad (F-10)$$

where C_{bond} is the flip-chip bond capacitance, C_{MQW} is the modulator capacitance per unit area, $C_{min,o}$ is the minimum inverter output capacitance and k_{min} is the minimum transistor transconductance parameter. It is not meaningful to operate the driver faster than a minimum inverter stage. Thus, equating Eq. (F-9) to the rise time of the signals at the output of a minimum inverter (driving another minimum inverter), using Eq. (F-8) and Eq. (F-10) in Eq. (F-9), and solving for the transistor transconductance gives:

$$k_n^{AC} = \frac{C_{bond} + A_{MQW}C_{MQW}}{0.5RC_{min} \cdot (V_{TR} - V_T) - \frac{C_{min,o}}{k_{min}}} \quad (F-11)$$

The NMOS transistor is therefore sized based on the stricter of the DC and AC conditions given by Eq. (F-7) and Eq. (F-11). To complete the driver design, the PMOS transistor is sized twice as big as the NMOS transistor to compensate for its lower mobility. The design of the driver gives the transmitter driver input capacitance as:

$$C_{TR,i} = \frac{k_{dr}}{k_{min}} C_{min,i} \quad (F-12)$$

where k_{dr} is the larger of the DC and AC transconductances, and $C_{min,i}$ is the minimum-size inverter input capacitance

Appendix G

VCSEL DRIVER DESIGN

The laser driver circuitry considered is shown in Fig. G-1, where a single NMOS transistor is used to switch the laser diode. The driver should satisfy the DC and the minimum switching speed requirements: the NMOS transistor should be large enough to conduct the high-level laser current, and switch the laser in no more than one minimum inverter propagation delay. In this analysis, we will assume that the laser is operated at the maximum output power level for a certain laser diameter. Therefore, the laser diameter is increased to achieve higher optical output power from the laser, as needed to accommodate larger fanout or faster operation. We assume that the output power vs. laser diameter and the threshold current vs. laser diameter characteristics are linear:

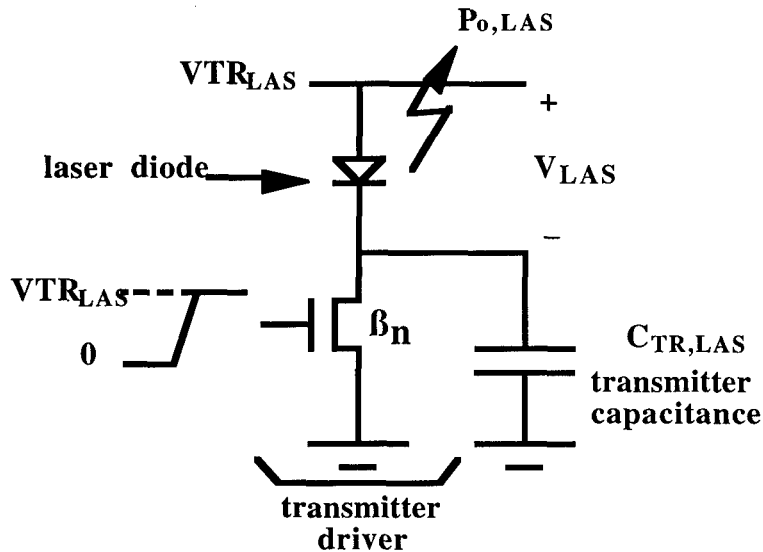


Figure G-1. CMOS VCSEL driver circuit.

$$P_{OH} = \gamma D \tag{G-1}$$

$$I_{th} = \phi D \tag{G-2}$$

where P_{OH} is the optical output high power level, I_{th} is the laser threshold current, D is the laser diameter, and γ and ϕ are the slopes of the characteristic curves. The laser Current to Power characteristic is assumed as:

$$P_{OH} = (I_{TR} - I_{th})\eta_{LI}, \quad I_{TR} \geq I_{th} \quad (G-3)$$

$$P_{OH} = 0, \quad I_{TR} < I_{th} \quad (G-4)$$

where η_{LI} is the slope of the laser L-I curve.

For a given laser output power requirement, laser diameter, threshold current, and operating DC current are found using (G-1) to (G-4) as follows:

$$D = \frac{P_{OH}}{\gamma} \quad (G-5)$$

$$I_{th} = \frac{\phi}{\gamma} P_{OH} \quad (G-6)$$

$$I_{TR} = P_{OH} \left(\frac{1}{\eta_{LI}} + \frac{\phi}{\gamma} \right) \quad (G-7)$$

The NMOS transistor should satisfy the DC condition $I_{DS} = I_{TR}$. Equating (G-7) to the linear region transistor current, neglecting the small quadratic term, and solving for the transistor transconductance yields:

$$k_n = \frac{I_{TR}}{(V_{TRLAS} - V_T)V_L} \quad (G-8)$$

where V_L is the NMOS drain voltage when it's "ON". Due to the low efficiency of the optical link, the resulting laser current is high. For this reason, we will assume that the design of the switch transistor based on (G-8) also satisfies the minimum switching speed requirement. If the transient capacitive current is approximately one fourth of the high-level laser current I_{TR} , then the rise time of the signal at the driver output can be estimated as:

$$t_r = 4C_{TRLAS} \frac{V_H - V_L}{I_{TR}} \quad (G-9)$$

where V_H is the high-level driver output voltage; $V_H = V_{TR} - V_{th}$, and C_{TR} is the laser driver output capacitance composed of flip-chip bond capacitance, NMOS transistor junction capacitance and the laser device capacitance:

$$C_{TR} = C_{bond} + \frac{k_n}{k_{min}} C_{min N,i} + A_{las} \cdot C_{las} \quad (G-10)$$

where $C_{\min N,i}$ and k_{\min} are the minimum NMOS transistor junction capacitance and transconductance, C_{las} is the laser device capacitance per area and A_{las} is the laser area;

$$A_{las} = \pi \cdot D \quad (G-11)$$

Finally, the laser driver input capacitance can be calculated as:

$$C_{TR,i} = \frac{k_n}{k_{\min}} C_{\min N,i} \quad (G-12)$$

where $C_{\min N,i}$ is the minimum NMOS transistor input capacitance.

Appendix H MATHEMATICA CODE

This appendix contains the Mathematica code that was used to represent the cost models and produce the figures.

Figure 5-20. Comparison of negative binomial and Poisson yield models for a 1994 CMOS process.

```
poisson := E^(- rate area fillfactor)
negativebinomial := 1 / (1 +
  rate area fillfactor/cluster)^cluster
cluster := 4
rate := 0.3
fillfactor := 0.8
Plot[{negativebinomial,poisson}, {area, 0.1, 4},
  PlotStyle ->{GrayLevel[0], Dashing[ {.03}]},
  PlotLegend -> {"Binomial", "Poisson"}, LegendPosition -> {0,0},
  AxesLabel->{"Area[cm^2]", "Yield"}]
```

Figure 5-21. The cost of 1994 high volume CMOS as a function of chip area including yield.

```
runcost := 38000
wafers := 20
waferdiameter := 8 2.54
negativebinomial := 1 / (1 + rate area fillfactor / cluster)^cluster
cluster := 4
rate := 0.3
fillfactor := 0.8
chips := wafers (Ceiling[Pi waferdiameter^2 / (4 area)] -
  Ceiling[Pi waferdiameter/Sqrt[area]])
chipcost := (runcost / chips)/(negativebinomial)
Plot[{chipcost},{area, 0.1, 4},
  AxesLabel -> {"Area [cm^2]", "Cost [$]"}]
```

Figure 5-22. Cost of a 250 μm pitch array of 50 μm^2 VCSELs as a function of the array size.

```

wafercost := 5000
runcost := 1250
waferdiameter := 2 2.54
density := 1000
spacing := 250 10^-4
vcSEL := 50 10^-8
area := lasers spacing^2
yield := E^(- density vcSEL lasers)
chips := (Ceiling[Pi waferdiameter^2 / (4 area)] -
  Ceiling[Pi waferdiameter/Sqrt[area]])
chipcost := (runcost + wafercost) / chips /yield
Plot[{chipcost},{lasers, 16, 256},
  AxesLabel -> {"VCSELS/Chip", "Chip Cost [$"]}

```

Figure 5-23. Cost of a 256-element VCSEL array as function of the number of chips solder bumped in the array.

```

wafercost := 5000
runcost := 1250
waferdiameter := 2 2.54
packagecost := 5
density := 1000
spacing := 250 10^-4
vcSEL := 50 10^-8
area := lasers spacing^2
yield := E^(- density vcSEL lasers)
chips := (Ceiling[Pi waferdiameter^2 / (4 area)] -
  Ceiling[Pi waferdiameter/Sqrt[area]])
chipcost := ((runcost + wafercost + soldercost) /
  chips + testcost)/yield
substratecost := chipspermodule area 4.8
bonds := 2 + lasers
lasers := 256 / chipspermodule
solderfailure := 1 - .99998
soldercost := 35
testcost := 1
dieattach := 0.25

```

```

bondyield := E^(- solderfailure bonds)
modulecost := ((chipcost + packagecost + dieattach)      chipspermodule +
substratecost + testcost +
    soldercost chipspermodule/chips )/ bondyield^chipspermodule
Plot[modulecost, {chipspermodule, 1, 4},
    AxesLabel -> {"Chips/Array", "Module Cost [$]"},
    PlotPoints -> 200, PlotRange -> All]

```

Figure 5-24. Yield of solder bumps, silicon and MQW modulators.

```

solderyield := E^(- solderfailure bumps)
bumps := 2 area fillfactor / (0.0125 0.00625)
solderfailure := 1 - 0.99998
negativebinomial := 1 / (1 + rate area fillfactor/cluster)^cluster
cluster := 4
rate := 0.43
fillfactor := 0.8
modyield := E^(-subdefect bumps /2 modarea)
subdefect := 100
modarea := 0.002 0.005
Plot[{negativebinomial, solderyield, modyield}, {area, 0.1, 4},
    PlotStyle ->{GrayLevel[0], Dashing[ {.015}], Dashing[ {0.05}]},
    PlotLegend -> {"Silicon", "Solder", "MQW"},
    LegendPosition -> {0.6,-0.25}, AxesLabel -> {"Area [cm^2]", "Yield"}]

```

Figure 5-25. Cost comparison of monolithic HFET-SEED GaAs vs hybrid CMOS-SEED as the chip size increases.

```

solderyield := E^(- solderfailure bumps)
bumps := 2 area fillfactor / (0.0125 0.00625)
solderfailure := 1 - 0.99998
soldercost := 5
negativebinomial := 1 / (1 + rate area fillfactor/cluster)^cluster
cluster := 4
rate := 0.43
fillfactor := 0.8
d := 2 2.54
aw := Pi (d/2)^2
gaaswafercost := 5000
monolithicprocess := 1250
w := 6
t := 35
dr := 2
ad := nn (t 10^-6 + w 4 10^-6 + dr 2 10^-6) / m
nn := Ceiling[area m / 0.000441]
usablechips := Ceiling[aw/area] - Ceiling[(Pi d)/Sqrt[area]]
ct := 5
b := 7
cp := monolithicprocess + gaaswafercost
lambda := 100
yield := E^(- lambda ad)
m := 1
chipcost := (cp / usablechips + ct) / yield
modityield := E^(-subdefect bumps /2 modarea)
subdefect := 100
modarea := 0.002 0.005
fillfactor := 0.8
siruncost := 25000
simetalcost := 200
runcost := (siruncost + simetalcost wafers)
wafers := 20
sidiameter := 8 2.54
gaasdiameter := 2 2.54
gaaswafercost := 5000
gaasprocess := 1000
gaascost := (gaaswafercost + gaasprocess) (sidiameter
/ gaasdiameter)^2 wafers

```

```

chips := wafers (Ceiling[Pi sidiameter^2
  /(4 area)] - Ceiling[Pi sidiameter/Sqrt[area]])
goossencost := ((runcost + gaascost) / chips + soldercost)/
  (negativebinomial solderyield modyield)
Plot[{chipcost,goossencost}, {area,0.04, 0.3},
  AxesLabel -> {"Area [cm^2]","Cost [$]"}, PlotRange -> All,
  PlotStyle -> {GrayLevel[0], Dashing[{0.03}]},
  PlotLegend -> {"Monolithic","Hybrid"}, LegendPosition -> {-0.6,0}

```

Figure 5-31. Comparison of manufacturing cost of VCSEL/CMOS (solid line) and CMOS chips (dashed line) interconnected by the shuffle-exchange network with optics/MCM or an MCM, respectively.

```

w := 64
sicost := 1250
sidiameter := 8 2.54
area := 1
negativebinomial := 1 / (1 + rate area fillfactor/cluster)^cluster
cluster := 4
rate := 0.3
fillfactor := 0.8
chips := (Ceiling[Pi sidiameter^2 / (4 area)] -
  Ceiling[Pi sidiameter/Sqrt[area]])
testcost := 1
soldercost := 35
chipcost := ((sicost + soldercost) / chips +
  testcost)/(negativebinomial)
mcmsub := 30 / 2.54^2
mcmyield := E^(- rate mcmlayers mcmcrit)
d := 0.0125
mcmlayers := 4/20
mcmcrit := (n w d/Log[2,n])^2
mcmarea := (n w d/Log[2,n])^2 + n area
mcmcost := (mcmarea mcmsub + soldercost + testcost) / mcmyield
solderyield := E^(- solderfailure bumps)
bumps := 2 w + w/4
solderfailure := 1 - 0.99998
dieattach := 0.25

```

```

modulecost := (n (chipcost + dieattach) +
  mcmcost + testcost)/solderyield^n
vcselyield := E^(-subdefect w vcselarea)
subdefect := 1000
vcselarea := 0.002 0.005
gaasdiameter := 2 2.54
gaaswafercost := 5000
gaasprocess := 1250
gaascost := (gaaswafercost + gaasprocess + soldercost)
gaasarea := 2 w vcselarea
gaaschips := (Ceiling[Pi
  gaasdiameter^2 / (4 gaasarea)] -
  Ceiling[Pi gaasdiameter/Sqrt[gaasarea]])
sicost := 1250
sidiameter := 8 2.54
area := 1
negativebinomial := 1 / (1 + rate area fillfactor/cluster)^cluster
cluster := 4
rate := 0.3
fillfactor := 0.8
chips := (Ceiling[Pi sidiameter^2 / (4 area)] -
  Ceiling[Pi sidiameter/Sqrt[area]])
testcost := 1
soldercost := 35
hybridyield := E^(- solderfailure w 2)
chipcost := ((sicost + soldercost) / chips +
  testcost + dieattach + gaascost / gaaschips)/
  (negativebinomial hybridyield vcselyield)
glasssub := 10 / 2.54^2
glassyield := E^(- rate glassfactor critarea)
d := 0.0125
glassfactor := 2/20
chipspace := 0.1
critarea := 2 d n Sqrt[area] + w d chipspace
glassarea := n area
glasscost := (glasssub glassarea +

```

```

    soldercost + testcost) / glassyield
solderyield := E^(- solderfailure bumps)
bumps := w/4 + w
solderfailure := 1 - 0.99998
dieattach := 0.25
cghcost := 4
optomount := 10
opticscost := (n (chipcost + dieattach) +
    glasscost + testcost)/solderyield^n + 4 (n cghcost + optomount)
LinearLogPlot[{modulecost,opticscost}, {n,2,64},
    PlotStyle ->{GrayLevel[0], Dashing[ {.03}]},
    PlotLegend -> {"mcm", "optics"}, LegendPosition -> {-0.5,0.25},
    AxesLabel->{"Chips", "$"}]

```

DISTRIBUTION LIST

addresses	number of copies
PAUL L. REPAK ROME LABORATORY/OCPC 25 ELECTRONIC PKY ROME NY 13441-4515	5
U. EFRON HUGHES RESEARCH LABORATORIES 3014 MALIBU CANYON ROAD MALIBU CA 90265	5
ROME LABORATORY/SUL TECHNICAL LIBRARY 26 ELECTRONIC PKY ROME NY 13441-4514	1
ATTENTION: DTIC-OCC DEFENSE TECHNICAL INFO CENTER 8725 JOHN J. KINGMAN ROAD, STE 0944 FT. BELVOIR, VA 22060-6218	2
ADVANCED RESEARCH PROJECTS AGENCY 3701 NORTH FAIRFAX DRIVE ARLINGTON VA 22203-1714	1
RELIABILITY ANALYSIS CENTER 201 MILL ST. ROME NY 13440-8200	1
ROME LABORATORY/C3A8 525 BROOKS RD ROME NY 13441-4505	1
ATTN: GWEN NGUYEN GIDEP P.O. BOX 8000 CORONA CA 91718-8000	1

AFIT ACADEMIC LIBRARY/LDEE 1
2950 P STREET
AREA B, BLDG 642
WRIGHT-PATTERSON AFB OH 45433-7765

ATTN: R.L. DENISON 1
WRIGHT LABORATORY/MLPD, BLDG. 651
3005 P STREET, STE 6
WRIGHT-PATTERSON AFB OH 45433-7707

WRIGHT LABORATORY/MTM, BLDG 653 1
2977 P STREET, STE 6
WRIGHT-PATTERSON AFB OH 45433-7739

ATTN: GILBERT G. KUPERMAN 1
AL/CFHI, BLDG. 248
2255 H STREET
WRIGHT-PATTERSON AFB OH 45433-7022

ATTN: TECHNICAL DOCUMENTS CENTER 1
OL AL HSC/HRG
2698 G STREET
WRIGHT-PATTERSON AFB OH 45433-7604

US ARMY SSSC 1
P.O. BOX 1500
ATTN: CSSD-IM-PA
HUNTSVILLE AL 35807-3801

COMMANDING OFFICER 1
NCCDSC RDT&E DIVISION
ATTN: TECHNICAL LIBRARY, CODE D0274
53560 HULL STREET
SAN DIEGO CA 92152-5001

NAVAL AIR WARFARE CENTER 1
WEAPONS DIVISION
CODE 48L000D
1 ADMINISTRATION CIRCLE
CHINA LAKE CA 93555-6100

SPACE & NAVAL WARFARE SYSTEMS CMD 2
ATTN: PMW163-1 (R. SKIANO)RM 1044A
53560 HULL ST.
SAN DIEGO, CA 92152-5002

SPACE & NAVAL WARFARE SYSTEMS 1
COMMAND, EXECUTIVE DIRECTOR (PD13A)
ATTN: MR. CARL ANDRIANI
2451 CRYSTAL DRIVE
ARLINGTON VA 22245-5200

COMMANDER, SPACE & NAVAL WARFARE 1
SYSTEMS COMMAND (CODE 32)
2451 CRYSTAL DRIVE
ARLINGTON VA 22245-5200

CDR, US ARMY MISSILE COMMAND 2
REDSTONE SCIENTIFIC INFORMATION CTR
ATTN: AMSMI-RD-CS-R, DOCS
REDSTONE ARSENAL AL 35898-5241

ADVISORY GROUP ON ELECTRON DEVICES 1
SUITE 500
1745 JEFFERSON DAVIS HIGHWAY
ARLINGTON VA 22202

REPORT COLLECTION, CIC-14 1
MS P364
LOS ALAMOS NATIONAL LABORATORY
LOS ALAMOS NM 87545

AEDC LIBRARY 1
TECHNICAL REPORTS FILE
100 KINDEL DRIVE, SUITE C211
ARNOLD AFB TN 37389-3211

COMMANDER 1
USAISC
ASHC-IMD-L, BLDG 61801
FT HUACHUCA AZ 85613-5000

US DEPT OF TRANSPORTATION LIBRARY 1
FB10A, M-457, RM 930
800 INDEPENDENCE AVE, SW
WASH DC 22591

AWS TECHNICAL LIBRARY 1
859 BUCHANAN STREET, RM. 427
SCOTT AFB IL 62225-5118

AFIWC/MSY 1
102 HALL BLVD, STE 315
SAN ANTONIO TX 78243-7016

SOFTWARE ENGINEERING INSTITUTE 1
CARNEGIE MELLON UNIVERSITY
4500 FIFTH AVENUE
PITTSBURGH PA 15213

NSA/CSS 1
KI
FT MEADE MD 20755-6000

ATTN: DM CHAUHAN 1
DCMC WICHITA
271 WEST THIRD STREET NORTH
SUITE 5000
WICHITA KS 67202-1212

PHILLIPS LABORATORY 1
PL/TL (LIBRARY)
5 WRIGHT STREET
HANSCOM AFB MA 01731-3004

ATTN: EILEEN LADUKE/0460 1
MITRE CORPORATION
202 BURLINGTON RD
BEDFORD MA 01730

DUSO(P)/D TSA/DUTD 2
ATTN: PATRICK G. SULLIVAN, JR.
400 ARMY NAVY DRIVE
SUITE 300
ARLINGTON VA 22202

ROME LABORATORY/ERD 1
ATTN: RICHARD PAYNE
HANSCOM AFB, MA 01731-5000

ROME LABORATORY/EROC 1
ATTN: JOSEPH P. LORENZO, JR.
HANSCOM AFB, MA 01731-5000

ROME LABORATORY/EROP 1
ATTN: JOSEPH L. HORNER
HANSCOM AFB, MA 01731-5000

ROME LABORATORY/EROC 1
ATTN: RICHARD A. SOREF
HANSCOM AFB, MA 01731-5000

ROME LABORATORY/ERXE 1
ATTN: JOHN J. LARKIN
HANSCOM AFB, MA 01731-5000

ROME LABORATORY/ERDR 1
ATTN: DANIEL J. BURNS
525 BROOKS RD
ROME NY 13441-4505

ROME LABORATORY/IRAP 1
ATTN: ALBERT A. JAMBERDINO
32 HANGAR RD
ROME NY 13441-4114

ROME LABORATORY/C38C 1
ATTN: ROBERT L. KAMINSKI
525 BROOKS RD
ROME NY 13441-4505

ROME LABORATORY/DCP 1
ATTN: MAJOR GARY D. BARMORE
25 ELECTRONIC PKY
ROME NY 13441-4515

ROME LABORATORY/DCP 1
ATTN: JOANNE L. ROSSI
25 ELECTRONIC PKY
ROME NY 13441-4515

NY PHOTONIC DEVELOPMENT CORP 1
MVCC ROME CAMPUS
UPPER FLOYD AVE
ROME, NY 13440