INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations. and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning 300 North Zeeb Road, Ann Arbor, Mi 48106-1346 USA 800-521-0600

I M

--



Alignment and Packaging Techniques for Two-Dimensional Free-Space Optical Interconnects

Michael H. Ayliffe

Department of Computer and Electrical Engineering McGill University, Montréal, Canada April 2001

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Doctor of Philosophy

© Michael H. Ayliffe, 2001



National Library of Canada

Acquisitions and Bibliographic Services

395 Wellington Street Ottawa ON K1A 0N4 Canada Bibliothèque nationale du Canada

Acquisitions et services bibliographiques

395, rue Wellington Ottawa ON K1A 0N4 Canada

Your Sie Vore référence

Our lie Nore rélérence

The author has granted a nonexclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission. L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-75605-X



Abstract

Two-dimensional free-space optical interconnects (2D-FSOIs) promise to deliver tremendous gains in bandwidth and architectural freedom for applications such as telecommunication switches and massively parallel computing systems. One major obstacle preventing the commercial deployment of 2D-FSOI systems is the problem of optical alignment, which is further exacerbated by the requirements that these systems be fieldserviceable and able to sustain the harsh conditions of industrial environments.

This thesis proposes a broad range of solutions to alleviate this alignment problem. One important aspect of this work concerns the development of a generic packaging strategy, which consists of partitioning an optical system into separate modules in such a way that the loose tolerances are between the modules while the tight tolerances are between the components inside the modules. To accomplish this, novel alignment techniques are designed and demonstrated, including the use of integrated diffractive features, CMOS position detectors, ultrathick photoresist micro-structures, and semi-kinematic fixtures using dowel pins. In all cases, emphasis is placed on approaches that are amenable to lowcost manufacturing and high-volume production.

These techniques were developed in the context of a photonic backplane prototype experiment that demonstrated 1024 free-space interconnections between four optoelec-tronic-VLSI (OE-VLSI) chips. The design and implementation of a module integrating an OE-VLSI chip, a mini-lens array, a thermoelectric cooler and a heatsink is presented. Optomechanical, electrical and thermal characterization results are reported.

The other aspect of this work aims at identifying the types of optical designs that provide more generous misalignment tolerances. This is done by investigating various optical configurations for the design of the chip module. The central objective is to understand the underlying reasons that make one configuration more misalignment-tolerant than another. A significant outcome of this work is to show that the inherent misalignment tolerances of 2D-FSOI systems translates into an aspect-ratio limitation similar to the one found in electrical interconnects.

Résumé

Les interconnexions bidimensionelles optiques à l'air libre promettent de fournir une énorme augmentation de largeur de bande et une architecture plus flexible à des applications telles que les commutateurs de télécommunications et les systèmes massivement parallèles. Un obstacle majeur à la réalisation commerciale des interconnexions optiques à l'air libre demeure le problème de l'alignement optique, qui se complique en raison de la nécessité d'avoir des systèmes faciles d'entretien et résistants aux conditions difficiles que l'on retrouve dans les milieux industriels.

Cette thèse présente un éventail de solutions à ce problème d'alignement. Son premier volet propose une stratégie générique d'empaquetage optique, qui consiste à diviser un système optique en plusieurs modules tout en s'assurant que les tolérances d'alignement faciles à réaliser se retrouvent entre les modules tandis que les tolérances d'alignement difficiles sont situées entre les composantes d'un même module. Pour ce faire, de nouvelles techniques d'alignement, optiques et mécaniques, sont conçues et prouvées. Ces techniques utilisent des élements diffractifs intégrés, des détecteurs d'alignement de type CMOS, des microstructures faites de photorésine super épaisse ainsi qu'un dispositif de fixation intégrant des goupilles de positionnement. L'accent est mis sur les approches ayant un potentiel de fabrication à bas coût et de production en masse.

Ces techniques ont été conçues dans le cadre d'un système expérimental de fond de chassis photonique ayant réalisé 1024 interconnexions optiques à l'air libre entre quatre puces hybrides optoélectroniques-VLSI (OE-VLSI). La conception et la fabrication d'un module intégrant une puce OE-VLSI, une matrice de mini-lentilles, un refroidisseur thermoélectrique et un dissipateur thermique sont presentées, ainsi que les résultats de la caractérisation optomécanique, électrique et thermique de ce module.

Le second volet de cette thèse tente d'identifier les types de systèmes optiques offrant les meilleures tolérances d'alignement. Plusieurs configurations optiques sont étudiées. Le principal objectif est de comprendre les raisons sous-jacentes qui font en sorte qu'une configuration est plus facile à aligner qu'une autre. Une conclusion importante de cette étude révèle que les tolérances d'alignement inhérentes aux systèmes d'interconnexions optiques à l'air libre font en sorte que ceux-ci sont limités par un rapport largeur/longueur similaire à celui touchant les interconnexions électriques.

Acknowledgments

I wish to sincerely thank my supervisor and friend David Plant for his guidance and tireless encouragement over the course of my graduate studies. He is the reason why I joined the Photonics Systems Group and eventually decided to make a career in this field of research. I am grateful to have had the opportunity to work with him and am deeply indebted to him for the great trust he has always placed in me.

I owe my deepest thanks to Dr. Brian Robertson and Dr. Guillaume Boisset for invaluable technical discussions related to optics and optomechanics. Particular thanks go to Prof. Andrew Kirk and Prof. Frank Tooley for their advices, technical discussions and great sense of humour. I must also thank Don Pavlasek from the departmental machine shop for all the time spent teaching me the rudiments of mechanical design and dimensional tolerancing. Many thanks to Dr. Edwis Richard for the countless hours spent with me in the microlithography laboratory at l'Université de Montréal. I also thank Tsuyoshi Yamamoto for showing me how to apply glue properly.

I have had the good fortune to carry out this work in a friendly research group made of talented individuals. Many heartfelt thanks go to Guillaume Boisset, David Rolston, David Kabal, Michael Venditti, Frédéric Lacroix, Emmanuelle Laprise, Rajiv Iyer, Pritha Khurana, Daniel Filiatrault-Brosseau, Marc Châteauneuf, Julien Faucher, Feras Michael, Rhys Adams, Nam Kim, Frédéric Thomas-Dupuis, Eric Bernier, Yongsheng Liu, Tomasz Maj, Madeleine Mony, Eric Bisaillon, Danny Birdie, Julianna Lin, Greg Brady, Wayne Hsiao, Keivan Razavi, Robert Varano, Leo Lin, Mitch Salzberg, Xin Xue, Marcos Otazo, Lukas Chrostowski, and Alain Shang. Special thanks vont aussi à la toujours souriante Sylviane Duval and Kay Johnson from the CITR office.

Un gros merci à Stéphane, Taryn, Matthew, Richard, Stéphanie, Alexandra, Marc, and Stéphane pour votre présence et vos encouragements. The contributions, mostly athletic, of the Inflictors of Pain are also gratefully acknowledged.

Je suis infiniment reconnaissant envers ma meilleure amie et compagne, Brigitte, qui est là depuis le tout début du voyage. Sa patience, sa compréhension et son immense générosité sont pour moi une source d'inspiration.

Finalement, j'aimerais dédier cette thèse à mes parents et mon frère pour leur amour et leurs encouragements continuels durant toutes ces années. Merci.

Table of Content

Chapter 1: Introduction

1.1	Scope of the thesis	
1.2	The off-chip bandwidth bottleneck	2
1.3	Limitations of electrical interconnects	
	1.3.1 Consequences of electrical lines having a finite resistance	6
	1.3.2 Consequences of electrical lines having a low impedance	7
1.4	Physical reasons for optical interconnects	8
1.5	Applications for 2D-POIs	10
1.6	Technological Challenges for 2D-POIs.	10
1.7	Original Contributions	11
1.8	Thesis Organization	
1.9	References	14

Chapter 2: Two-dimensional parallel optical interconnect (2D-POI) technologies

2.1	Introd	luction	
2.2	Optoe	electronic-VLSI (OE-VLSI) technology	22
	2.2.1	Silicon versus III-V material systems	22
	2.2.2	Monolithic versus hybrid integration	22
2.3	Trans	mitter technologies	24
	2.3.1	Light-emitting diodes (LEDs)	24
	2.3.2	Vertical-cavity surface-emitting lasers (VCSELs)	25
	2.3.3	Electro-absorption (EA) modulators	
2.4	Detec	ctor technologies	
2.5	Optic	al receivers	34
2.6	Guide	ed-wave interconnection technologies	
	2.6.1	Two-dimensional fiber arrays	36
	2.6.2	Fiber image guides (FIGs)	
2.7	Free-	Space interconnection technologies	
	2.7.1	Conventional Lenses	39
	2.7.2	Microlens arrays	
	2.7.3	Hybrid lenses	41
	2.7.4	Mini-lens arrays	42

2.8	Planar optics	43
2.9	Optical interconnection medium hierarchy	44
2.10	Conclusion	46
2.11	References	47

Chapter 3: Design and testing of a free-space photonic backplane demonstrator

3.1	Introduction60			
3.2	Target application: a multiprocessor computing system60			
3.3	Interconnect topology			
3.4	System specifications and requirements	64		
3.5	Physical layout issues	65		
	3.5.1 OE-VLSI chip mounted directly on the motherboard	65		
	3.5.2 OE-VLSI chip packaged in a separate module	68		
3.6	Optoelectronic-VLSI chip	69		
	3.6.1 MQW modulator technology	69		
	3.6.2 CMOS technology	73		
	3.6.3 OE-VLSI chip layout	75		
	3.6.4 CMOS chip functionality	77		
3.7	Optical interconnect design			
	3.7.1 Selection of focal length	80		
	3.7.2 Power delivery system	81		
	3.7.3 Modulator-to-detector link efficiency	82		
3.8	Optical packaging considerations	82		
3.9	System implementation and performance	85		
	3.9.1 Optical modules	85		
	3.9.2 OE-VLSI chip	88		
	3.9.3 MQW modulator reflectivity	88		
	3.9.4 Chip module	89		
3.10	Author's contributions	90		
3.11	Conclusion	91		
3.12	Acknowledgments92			
3.13	References			

Chapter 4: Chip module design and testing

4.1	Introduction
-----	--------------

4.2	Desig	n choices and constraints	96
	4.2.1	Alignment issues	96
	4.2.2	Electrical issues	98
	4.2.3	Thermal issues	99
4.3	Chip	module design overview	
	4.3.1	Physical dimensions	
	4.3.2	Materials selection	100
	4.3.3	Chip module assembly	101
4.4	Electr	rical packaging and high-speed testing	102
	4.4.1	Electrical packaging design	102
	4.4.2	High-speed testing	105
	4.4.3	Crosstalk measurements	106
4.5	Therm	nal design and experimental evaluation	110
4.6	Conc	lusion	116
4.7	Ackn	owledgments	
4.8	Refer	ences	

Chapter 5: Intra-module alignment techniques

5.1	Introd	luction	119
5.2	Revie	w of previously published techniques	120
5.3	Proble	em definition: aligning a lens array to a OE-VLSI chip	123
5.4	Techr	nique #1: off-axis Fresnel lenses and on-chip quadrant detectors	124
	5.4.1	Position sensing using quadrant detectors	126
	5.4.2	Alignment procedure	128
	5.4.3	Design of CMOS-compatible quadrant detectors	129
	5.4.4	Responsivity measurements	131
	5.4.5	Discussion	132
5.5	Technique #2: on-chip off-axis linear Fresnel zone plates (FZPs)		133
	5.5.1	Sources of errors	136
	5.5.2	Design considerations	137
	5.5.3	Experimental setup requirements	137
	5.5.4	Retroreflected beam alignment technique	138
	5.5.5	In Situ beam alignment technique	140
	5.5.6	Worst-case misalignment errors	143
	5.5.7	Experimental results for the accuracy of technique #2	144
	5.5.8	Improved implementation of technique #2	145
	5.5.9	Automating alignment technique #2 using on-chip detectors	146

5.6	Conclusion	149
5.7	Acknowledgments	150
5.8	References	150

Chapter 6: Inter-module alignment techniques

6.1	Introd	uction	154
6.2	Revie	w of previously published techniques	155
	6.2.1	Mechanical methods	155
	6.2.2	Optical methods	158
	6.2.3	Array redundancy methods	161
6.3	Chip i	module interface requirements	
6.4	Desig	n #1: Kinematic fixture using alignment micro-structures	163
	6.4.1	The Kelvin clamp	164
	6.4.2	Design of the Kelvin clamp	165
	6.4.3	Fabrication of alignment micro-structures using SU-8 photoresist	166
	6.4.4	Discussion	170
6.5	Desig	m #2: Semi-kinematic fixture using dowel pins	173
	6.5.1	Experimental evaluation of insertion repeatability	175
6.6	Conci	lusion	177
6.7	Ackn	owledgments	179
6.8	Refer	ences	179

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

7.1	Introd	luction	
7.2	Motiv	/ation	
7.3	Defin	ing a misalignment metric	187
7.4	Defin	ing a figure of merit for alignability	188
7.5	Misal	lignment tolerance and scalability analysis	
	7.5.1	Design #1: no optics integrated with the chip	190
	7.5.2	Design #2: microchannel design	
	7.5.3	Design #3: clustering using mini-lens array	
	7.5.4	Design #4: microchannel telescope	
	7.5.5	Design #5: microchannel with field lens array	
	7.5.6	Summary of results	201
7.6	Discu	ission	201

7.7	Benefits of Gaussian relays	204
7.8	Invariance of the alignment product	205
7.9	Aspect ratio limitation of misaligned free-space optical interconnects	208
7.10	Conclusion	209
7.11	References	210

Chapter 8: Conclusion

8.1	Summary	
8.2	Design and assembly of a free-space photonic backplane	
8.3	Packaging strategy and alignment techniques	
8.4	Design of misalignment-tolerant 2D-FSOI modules	
8.5	Future research directions	
	8.5.1 Intra-module issues	
	8.5.2 Inter-module issues	
8.6	References	

Appendix A: Mechanical drawings of the chip module assembly

Mini-lens holder	A-1
Mounting spacer	A-7
Flex-PCB mount	A-10
Thermally isolating spacer	A-13
Copper heat spreader	A-15
Protective cover	A-17

Appendix B: Mathematical derivations for intra-module alignment technique #1

Coordinate system definitions	B-1
Solving for the spot positions	B-1
Coordinate system conversion	B-4
Solving for the chip misalignment	B-6
Summary	B-9



Associated Publications

The work reported in this thesis has been published or will be published in the form of the following journal articles and conference papers.

Journal articles

- [1] M. H. Ayliffe, D.V. Plant, "On the design of misalignment-tolerant free-space optical interconnects," to be published in Applied Optics.
- [2] M. H. Ayliffe, M. Châteauneuf, D. R. Rolston, A. G. Kirk, D.V. Plant, "Six-degreesof-freedom alignment of two-dimensional array components using in-situ off-axis diffractive structures," accepted for publication in Applied Optics, July 2001.
- [3] M. H. Ayliffe, D. R. Rolston, E. L. Chuah, E. Bernier, F. S. J. Michael, D. Kabal, A. G. Kirk, D. V. Plant, "Design and testing of a kinematic package supporting a 32 x 32 array of GaAs MQW modulators flip-chip bonded to a CMOS chip," to be published in IEEE J. of Lightwave Technology, October 2001 issue.
- [4] M. H. Ayliffe, D. Kabal, F. Lacroix, E. Bernier, P. Khurana, A. G. Kirk, F. A. P. Tooley, D. V. Plant, "Electrical, thermal and optomechanical packaging of large 2D optoelectronic device arrays for free-space optical interconnects," Journal of Optics A: Pure and Applied Optics, vol. 1, pp. 267-271 (1999).
- [5] D.F. Brosseau, F. Lacroix, M.H. Ayliffe, E. Bernier, B. Robertson, F.A.P. Tooley, D.V. Plant, A.G. Kirk, "Design, implementation, and characterization of a kinematically aligned, cascaded spot-array generator for a modulator-based free-space optical interconnect," Applied Optics, vol.39, no.5, pp. 733-745 (2000).
- [6] Y. Liu, B. Robertson, G. C. Boisset, M. H. Ayliffe, R. Iyer, D. V. Plant, "Design, implementation, and characterization of a hybrid optical interconnect for a four-stage free-space optical backplane demonstrator," Applied Optics, vol. 37, no. 14, pp. 2895-2914 (1998).
- [7] R. Iyer, Y. S. Liu, G. C. Boisset, D. J. Goodwill, M. H. Ayliffe, B. Robertson, W. M. Robertson, D. Kabal, F. Lacroix, D. V. Plant, "Design, implementation, and characterization of an optical power supply spot-array generator for a four-stage free-space optical backplane," Applied Optics, vol.36, no.35, pp. 9230-9242 (1997).

multistage free-space optical interconnection system," Optics in Computing 2000 Conference, June 18-23 Québec City, Qc., Canada.

- [15] A. G. Kirk, F. K. Lacroix, M. H. Ayliffe, E. Bernier, D. F-Brosseau, M. Chateauneuf,
 B. Robertson, F. A. P. Tooley, D. V. Plant, "A multistage free-stage optical interconnect for backplane applications: implementation issues," IEEE LEOS 1998, Orlando,
 FL, USA, 1-4 Dec. 1998, p.149-50.
- [16] M. B. Venditti, D. N. Kabal, M. H. Ayliffe, D. V. Plant, F. A. P. Tooley, E. Richard, J. Currie, A. J. Spring Thorpes, "Temperature dependence of QCSE device characteristics and performance," 1998 IEEE/LEOS Summer Topical Meeting. Monterey, CA, USA, 20-24 July 1998. p.17-20.
- [17] M. H. Ayliffe, D. Kabal, P. Khurana, F. Lacroix, A. G. Kirk, F. A. P. Tooley, D. V. Plant, "Optomechanical, electrical and thermal packaging of large 2D optoelectronic device arrays for free-space optical interconnects," Optics in Computing '98, Brugge, Belgium, 17-20 June 1998, p.502-505.
- [18] A. G. Kirk, D. F-Brosseau, F. K. Lacroix, E. Bernier, M. H. Ayliffe, B. Robertson, F. A. P. Tooley, D. V. Plant, "Design and implementation of a two-stage optical power supply spot array generator for a modulator-based free-space interconnect," Optics in Computing '98, Brugge, Belgium, 17-20 June 1998, p.48-51.
- [19] F. Lacroix, B. Robertson, M. H. Ayliffe, E. Bernier, F. A. P. Tooley, M. Chateauneuf, D. V. Plant, A. G. Kirk, "Design and implementation of a four-stage clustered free-space optical interconnect," Optics in Computing '98, Brugge, Belgium, 17-20 June 1998, pp.107-110.
- [20] D. Kabal, M.H. Ayliffe, G.C. Boisset, D.V. Plant, D.R. Rolston and M.B. Venditti, "Chip-on-board packaging of a Hybrid-SEED smart pixel array," Optical Computing 1997, Lake Tahoe, Nevada, March 18-21, 1997.
- [21] D. V. Plant, B. Robertson, H. S. Hinton, M. H. Ayliffe, G. C. Boisset, D. J. Goodwill, D. Kabal, R. Iyer, Y. S. Liu, D. R. Rolston, M. B. Venditti, T. H. Szymanski, W. M. Robertson, M. R. Taghizadeh, "Optical, optomechanical, and optoelectronic design testing of a multistage optical backplane demonstrator system," 1997 Photonics West, Hybrid and Monolithic OEICs, San Jose, CA, February 12-14, 1997, paper 3005-17.

- [22] D. V. Plant, B. Robertson, H. S. Hinton, M. H. Ayliffe, G. C. Boisset, D. J. Goodwill, D. Kabal, R. Iyer, Y. S. Liu, D. R. Rolston, M. B. Venditti, T. H. Szymanski, W. M. Robertson, M. R. Taghizadeh, "Optical, optomechanical, and optoelectronic design and operational testing of a multi-stage optical backplane demonstration system," Proceedings of the Third International Conference on Massively Parallel Processing Using Optical Interconnections (MPPOI '96). Maui, HI, USA, 27-29 Oct. 1996, p.306-312.
- [23] G. C. Boisset, M. H. Ayliffe, D. J. Goodwill, B. Robertson, R. Iyer, Y. S. Liu, D. Kabal, D. Pavlasek, W. M. Robertson, D. R. Rolston, H. S. Hinton and D. V. Plant, "Design, fabrication and characterization of optomechanics for a hybrid four-stage free-space optical backplane demonstrator," presented at the 1996 OSA Annual Meeting, Rochester, NY, paper MBBB4.
- [24] R. Iyer, D.J. Goodwill, B. Robertson, W.M. Robertson, Y.S. Liu, G.C. Boisset, M.H. Ayliffe, D. Kabal, D. Pavlasek, M.R. Taghizadeh, H.S. Hinton and D.V. Plant, "Characterization and measurement of an optical power delivery system for a freespace photonic backplane demonstrator," presented at the 1996 OSA Annual Meeting, Rochester, NY, paper MBBB5.
- [25] B. Robertson, Y.S. Liu, G.C. Boisset, D.J. Goodwill, M.H. Ayliffe, W.H. Hsiao, R. Iyer, D. Kabal, D. Pavlasek, D.R. Rolston, M.R. Taghizadeh, W.M. Robertson, H.S. Hinton and D.V. Plant, "Optical design and characterization of a compact free-space photonic backplane demonstrator," presented at the 1996 OSA Annual Meeting, Rochester, NY, paper MLL5.
- [26] D. R. Rolston, D. V. Plant, H. S. Hinton, W. S. Hsiao, M. H. Ayliffe, D. N. Kabal, T. H. Szymanski, A. V. Krishnamoorthy, K. W. Goossen, J. A. Walker, B. Tseng, S. P. Hui, J. C. Cunningham, W. Y. Jan, "Design and testing of a smart pixel array for a four-stage optical backplane demonstrator," IEEE/LEOS 1996 Summer Topical Meetings, Keystone, CO, USA, August 1996, p.30-31.
- [27] D. V. Plant, B. Robertson, H. S. Hinton, M. H. Ayliffe, G. C. Boisset, D. J. Goodwill,
 D. N. Kabal, R. Iyer, Y. S. Liu, D. R. Rolston, W. M. Robertson, M. R. Taghizadeh,
 "A multistage CMOS-SEED optical backplane demonstration system," in Proceedings of Optical Computing 1996, vol. 1, Sendai, Japan, April 1996. pp.14-15.

Chapter 1: Introduction

1.1 Scope of the thesis

This thesis considers the use of two-dimensional parallel optical interconnects (2D-POIs) applied to the problem of providing Tbit/s data communication between silicon VLSI chips. The idea of using 2D arrays of light beams to transfer data between silicon chips was proposed in the early 1980's, starting with the seminal paper of Goodman et al. [1]. Until recently, the physical advantages of 2D-POI systems had been offset by technological disadvantages, but 2D-POI technologies have now matured to the point where it becomes possible to contemplate commercial applications within a time frame of 5 years. This progress is primarily due to advances made in the areas of optoelectronic-VLSI (OE-VLSI) and micro-optical technologies. Today, large 2D arrays of vertical-cavity surfaceemitting lasers (VCSELs) or multiple-quantum-well (MQW) electro-absorption (EA) modulators can be routinely flip-chipped to CMOS chips [2]-[5]. Similarly, highly efficient multi-function diffractive and refractive micro-optical components are now available commercially [6][7]. Despite these achievements, the construction of highly parallel 2D-POI systems remains a difficult task, mostly due to the lack of practical and cost-effective packaging technologies which allow for 2D array components to be easily aligned to one another, similar to the ease with which connectorized fiber-optic components are assembled. Some even contend that research in this field has now reached a point where the complexity of implementing the packaging technologies far exceeds any shortcomings in the performance of the OE-VLSI and micro-optical technologies [8].

The origin of the alignment problem lies in the fact that 2D-POI systems require multiple array components to be individually aligned in three spatial and three angular coordinates with tolerances that often exceed the capabilities of standard packaging techniques. This thesis focuses on the development of a broad range of solutions to alleviate this alignment problem. One important aspect of this work concerns the development of novel alignment techniques for packaging components into modules (intra-module techniques) and for interfacing modules to one another (inter-module techniques), with a particular focus on approaches that are amenable to low-cost and high-volume manufacturing. In most cases, the techniques that are presented were developed in the context of a photonic backplane prototype experiment that demonstrated 1024 free-space optical interconnections between four OE-VLSI chips. This experimental prototype provided a realistic framework for developing and demonstrating new ideas, ensuring that these be subjected to system-level constraints and technological limitations.

The other aspect of this work aims at identifying the fundamental reasons behind the tolerances of 2D-POI systems in order to propose designs that are inherently more tolerant to misalignment. This approach will initially simplify alignment and may ultimately lead to systems that are passively assembled using "snap-together" modules, a pre-requisite for their commercial deployment. Research along this direction has lead to the formulation of general guidelines for the design of misalignment-tolerant 2D-POI modules.

The remainder of this chapter justifies in more detail the motivation behind this thesis. The next section demonstrates how the sustained exponential growth of on-chip processing capacity is imposing exceeding demands on the bandwidth capabilities of current electrical interconnect technology. The following sections bring forward the fundamental reasons behind the limitations of electrical interconnections and examine the physical properties of optics that motivate the use of 2D-POI technologies for data communication between silicon chips. Most of this discussion is derived from a large number of publications that address the potential benefits and limitations of optical interconnects versus electrical interconnects [1],[9]-[20]. The remaining sections address thesis organization and highlight the author's contributions to this area of research.

1.2 The off-chip bandwidth bottleneck

The prediction made by Gordon Moore in the early 1970s that transistor count on state-of-the-art integrated circuit technology would double every 18 months - a statement loosely referred to as "Moore's law" - has held remarkably well over the past 30 years. The continued increase in transistor count has been achieved through a combination of reduced transistor dimensions and an increase in chip size, both subjected to the constraint of reasonable yield. There is a large effort in the industry to maintain this trend; this is supported by the periodic publication of the Semiconductor Industry Association (SIA) road-

map that identifies the technological developments required over a 15-year horizon. Table 1 shows some of the technology requirements for cost-performance and high-performance microprocessor market segments, taken from the International Technology Roadmap for Semiconductors (ITRS) 1999 edition [21].

Year of Introduction <i>Technology node</i>	1999 180 nm	2002 130 nm	2005 1 <i>00 nm</i>	2008 70 nm	2011 <i>50 nm</i>	2014 35 nm		
Transistor density logic (Mtransistors/cm ²)								
Cost-performance	7	22	41	100	247	609		
High-performance	24	78	142	350	863	2,130		
Chip size (mm ²)								
Cost-performance	170	191	235	270	308	351		
High-performance	450	509	622	713	817	937		
Power consumption for a single-chip package (Watts)								
Cost-performance	48	75	96	1.04	1.02	115		
High-performance	90	130	160 -		174	183		
Power supply voltage (Volts)								
Cost-performance	1.8	1.5	12.7	0.9	06.	O.ScO.Co		
High-performance	1.8	1.5	1,2	-0.0	0,6	0,540,6		
Maximum number of package pins/balls								
Cost-performance	740	1012	1384	<u>. 1003</u>	2580	3541		
High-performance	1600	2248	31158	-4427	(<u>2</u> 22	8.75E		
On-chip clock frequency (MHz)								
Cost-performance	600	800	1100	<u> 1200</u>	0.031	2200		
High-performance	1200	1600	2000	2500	3(1)()	E600		
Chip-to-board (off-chip) speed for peripheral buses (MHz)								
Cost-performance	133	150	150	175	2.010	2.25		
High-performance	600	800	1,000	1,250	1.500	1.1:00		
Cost-performance = $< 3,000$ laptops, desktop PCs and telecommunications.								
High-performance $=$ > 3,000\$ workstations, servers, avionics, supercomputers.								
Sheared energy mean the and another solutions outs								

Table 1.1. Selected performance requirements of packaged microprocessor chips [22].



Although it seems likely that traditional transistor scaling will continue for the next 10 years (with a 30% reduction in feature size every third year), the data collected in table 1 indicates that silicon technology is beginning to reach fundamental limits. The shaded areas of the table, which correspond to technological requirements with no currently known solutions, are in many cases within a 5-year reach. The number and difficulty of the technical challenges continue to increase as technology moves forward. In their 1999 report, the roadmap coordinating group recognized that "it is becoming difficult for people in the semiconductor industry to imagine how we could continue to afford the historic trends (...) for another 15 years" [22].

Large digital systems are often implemented by distributing the amount of processing across multiple chips. In general, as the processing capacity of individual chips increases, so does the bandwidth requirement on the interconnections over which they communicate. This is especially true for global architectures, where data is communicated across the full scale of the system. With Moore's law continuing to hold, the sustained exponential growth of on-chip processing capacity is imposing exceeding demands on the bandwidth capabilities of current electrical interconnect technology. Already today, the performance of many digital systems is limited by the electrical interconnections rather than the chips at either end [23]. On-chip performance is often masked by the electrical interconnections. A simple example of this technology gap can be found inside computer systems, where data buses operate at rates much slower than the clock rate on the chips.

Using the data of table 1, it is easy to show how the existing technology gap will continue to grow with time. To do this, the following metrics are defined. First, the product of the number of transistors and the on-chip clock frequency is used to parametrize the ability of a chip to perform logic operations (although this does not imply that all transistors are operating at all time). Similarly, the product of the off-chip data rate with the package pin count is used to parametrize the ability of a chip to communicate data to another chip (although this is not to say that all package pins are used for data communications). Figure 1.1 shows both metrics being plotted on a logarithmic scale. Results from this first-order analysis clearly illustrate how conventional electrical interconnection technology cannot scale with the growing on-chip processing capacity. with 4000 optical I/O in a $7 \times 7 \text{ mm}^2$ area in 1996 [31] and a 375 Mb/s receiver-transmitter in 0.8 µm CMOS with less than 6 mW power dissipation in a $17 \times 18 \text{ µm}^2$ area in 1995 [32]. High interconnection density combined with low power dissipation make 2D-POI technology arguably the only interconnect solution able to keep pace with future generations of silicon integrated circuits [14].

1.3 Limitations of electrical interconnects.

Before discussing the benefits of optics to the interconnection problem, the physical reasons behind the bandwidth limitations of electrical interconnects are first examined. The purpose of this section is to show that most of the limitations of electrical interconnections (limited bit rates, high power dissipation, low connection density and high switching noise) are fundamentally due to the inherent low impedance and finite resistance of metal transmission lines.

1.3.1 Consequences of electrical lines having a finite resistance

All electrical lines have a finite resistance, which is dominated by bulk resistance in the case of RC lines and skin-effect resistance in the case of LC lines. This resistance limits the signal rise time which in turn puts an upper bound on the bit rate capacity of the line. The longer the electrical line, the larger the resistance, the longer the rise time and the lower the bit rate capacity. The larger the cross-sectional area of the line, the lower the resistance, the shorter the rise time and the higher the bit rate capacity. Relationships between length, area and bit rate were initially developed by Goodman et al. for RC lines [1] and have since been elegantly generalized to both RC and LC lines by Miller and Ozaktas [18] to show that the limit to the bit rate capacity of an electrical line depends only on the ratio of the cross-sectional dimension (\sqrt{A}) to the length (l) of the line - the "aspect ratio" of the interconnection. This limit is given approximately by:

$$B \cong B_o \times \frac{A}{l^2} \tag{1.1}$$

where $B_o \sim 10^{15}$ bit/s for LC lines (i.e. off-chip lines) and $B_o \sim 10^{16}$ bit/s for RC lines (i.e. on-chip lines). Note that this aspect ratio limitation is scale-invariant; it applies equally well to long coaxial cables, microstrip traces on a printed-circuit board (PCB) or short traces on a silicon chip. As an example, consider the bit rate limitation of a transmission line on a conventional electrical backplane with a horizontal dimension of 40 cm. A typical implementation would use microstrip traces that are 0.5 mm above a ground plane and pitched at 0.5 mm (limited by crosstalk requirements) [33]. The aspect ratio (A/l^2) of a line linking both ends of the backplane is then approximately equal to 1.5×10^{-6} , resulting in a bit rate limitation of ~ 1.5 Gbit/s.

As shown by the previous example, the "aspect ratio" limitation is already seen as a practical limit on sending high-speed signals over the length of current electrical backplanes and it is likely to become the limiting factor in the design of electronic systems approaching Tbit/s off-chip bandwidths. In fact, this "aspect ratio" limitation is largely responsible for the recent emergence of optical solutions to the frame-to-frame [34] and backplane interconnect problem [35].

1.3.2 Consequences of electrical lines having a low impedance

In addition to their finite resistivity, high-speed LC lines are usually designed with a low impedance, typically in the range 30 - 100 ohms. This design range is largely independent of the specific details of the design of the line; it follows from fundamental constants and the fact that line capacitance and inductance vary only logarithmically with the dimensions of the line [18]. The low impedance of electrical interconnects is problematic because it precludes the use of small-size driver circuits (with their naturally high output impedance) for sending high-speed signals off-chip. Redesigning the line to increase its characteristic impedance can be done but not without creating new problems. For example, increasing the separation between the conductor and ground plane does increase line impedance but at the expense of more crosstalk between lines. Alternatively, line impedance can be increased by reducing the size of the conductor, but this leads to a larger skineffect resistance with its associated reduction in bit rate performance.

The typical solution to this impedance mismatch problem is to use large-area output drivers (to reduce the driver output impedance) along with termination resistances to avoid wave reflection phenomena. This has a number of detrimental consequences: it results in large silicon area consumption, large current switching transients and high power dissipation with the associated undesirable consequences of ground bounce due to simultaneous switching noise (delta-I noise). Also, the fact that parallel electrical lines must follow planar geometries and are constrained by crosstalk issues limit the density of electrical interconnections. The above points are summarized in the influence diagram of figure 1.2



Figure 1.2. Consequences of electrical lines having a low impedance.

1.4 Physical reasons for optical interconnects

Optical interconnects are not limited by the aspect ratio of equation 1.1. Over the scale of an electronic system, optical interconnects do not suffer from distance-dependent optical loss or distortion (loss and dispersion in optical waveguides start appearing only over larger distances). In other words, if one can build parallel optical interconnections between two chips, then one is able to take those interconnections over essentially any "machine-scale" distance without any degradation. As a result, optics can provide highbandwidth interconnections between chips located as far as 10's of meters away with transceivers that are no different from the ones that would be used over shorter distances. This opens the possibility for revolutionary system architectures that are physically impossible to implement with electrical interconnections [36].

Additional advantages of optics originate from the electrical-to-optical (E-O) and optical-to-electrical (O-E) conversions occurring at both ends of the interconnect. The OE devices are used to effectively isolate the high impedance of small-size electronic circuits from the low impedance of the wave propagation. This is the so-called quantum impedance conversion phenomenon described in [11], and it has many important consequences:

- No impedance mismatch problem: small-size low-power transceivers can be used to send and receive high-speed optical data. Wave reflections are avoided by applying anti-reflection coatings on the optical surfaces.
- Low power dissipation and low silicon area consumption: dense 2D arrays of OE devices can be integrated directly on the silicon chip (1000's of channels/cm²), thereby exploiting the third spatial dimension of the interconnect. This large interconnection density allows for 2D-POI technology to keep up with the ever increasing computational bandwidth of silicon chips [14].
- Voltage isolation between interconnected chips: this results in the absence of ground bounce and other undesirable effects associated with pin inductance.
- No radio-frequency signal interference: detectors are OE quantum devices that behave as high-pass optical-frequency filters, generating electron-hole pairs only for photons with sufficiently high energies (~1 eV). All other low-energy "photons" - like the ones associated with RF interference signals - are simply filtered out.
- Absence of frequency-dependent loss and distortion: this is because OE sources automatically convert a baseband electronic signal into an optical-frequency (~300 THz) carrier-modulated signal. The frequency of light is very high compared to the data frequency and thus modulation makes no difference to the propagation of the signal. This means that an optical system designed for a 100 MHz clock signal will

continue to work equally well at 100 GHz, provided we can design transceivers that can run this fast. This same can certainly not be said of electrical interconnects.

Ability to implement wavelength-division multiplexing (WDM) systems: it is possible to exploit "a fourth dimension" (in addition to the three spatial dimensions) by designing OE devices that convert electrical signals into light beams of slightly different carrier frequencies (wavelengths) [37][38].

1.5 Applications for 2D-POIs

Target applications for 2D-POIs are ones that require the I/O bandwidth (off-chip capacity) to scale in proportion with the computational bandwidth (on-chip capacity). There exist numerous examples of such applications in switching, parallel processing, Fourier transform, sorting, matrix-vector processing, database searches and processor-to-memory applications [1][36][39]. For these applications, one significant advantage of 2D-POI systems is the ability to input entire 2D data array to the chip while using the third physical dimension for data propagation.

Successful implementations of 2D-POI systems have in common that they must take advantage of the high optical I/O bandwidth of OE-VLSI chips while not being limited by their order-of-magnitude-lower off-chip electrical capacity. One way to achieve this is by using 2D fiber arrays (section 2.6.1) to bring data optically on and off chips [40][41]. Another approach consists of using the aggregate electrical off-chip I/O capacity of many OE-VLSI chips to fill the optical I/O capacity of each. A good example of this is the concept of an optical backplane to interconnect multiple OE-VLSI chips residing on separate printed-circuit boards [42]. Numerous experimental demonstration of optical backplanes can be found in the literature, including [43]-[45].

1.6 Technological Challenges for 2D-POIs.

Although optics can provide many advantages, there still remains practical problems limiting the commercial deployment of 2D-POI systems. A list of the most significant technological challenges include:

- better integration of OE device arrays onto silicon VLSI chips. Desirable attributes are high yield, large and uniform arrays, low parasitic capacitance, low thermal resistance and compatibility with high-volume production at the wafer-level.
- high-speed, small-area, low-power transceivers. Crosstalk and noise immunity during large array operation are critical issues that have not been addressed sufficiently.
- dense and uniform arrays of temperature-tolerant emitters/modulators with operating voltages that can keep up with the scaling of silicon technology.
- the development of alignment techniques and cost-effective packaging methodologies compatible with high-volume manufacturing. This thesis focuses directly on this problem.

Although the above list is only partial, it is included here to provide a proper perspective on the research problems addressed in this thesis and their relation with the other research activities going on in this field.

1.7 Original Contributions

The approach taken in this thesis has been both theoretical and experimental and has led to the following original contributions to the field:

- First demonstration of a high-performance chip module integrating a 32 × 32 array of OE devices flip-chip bonded to a 9 × 9 mm² CMOS VLSI chip. The module integrates a microlens array, a high-speed flexible printed-circuit-board and a thermoelectric cooler. This work differentiates itself from previous implementations in that it simultaneously addresses optomechanical, electrical and thermal design issues (chapter 4).
- First demonstration of the use of on-chip off-axis linear Fresnel Zone Plates (FZPs) for achieving six-degrees-of-freedom (6-DOF) alignment of a microlens array to an OE-VLSI chip. This technique offers significant improvements over previous methods by providing high sensitivity in all DOFs in addition to minimizing silicon area usage (chapter 5).

- First attempt at automating the packaging of a microlens array to an OE-VLSI chip using off-axis alignment beams combined with three on-chip quadrant detectors. This technique uses the alignment information derived from the detector photocurrents to compensate for the chip misalignment in all six DOFs (chapter 5).
- First demonstration of a novel beam alignment technique that improves the accuracy with which a collimated beam is aligned orthogonal to a flat transparent substrate. Orthogonal beam alignment is required when aligning two out-of-plane components, such as a microlens array and an OE-VLSI chip. The technique uses *in situ* off-axis diffractive features fabricated on the microlens substrate. Angular beam misalignment can be precisely quantified simply by reading off the position of the focal line on a lithographically-defined ruler deposited on the same substrate (chapter 5).
- First demonstration of a truly robust, kinematic optical module supporting 1024 optical interconnections. The module can be manually inserted into and removed from a photonic backplane demonstrator without upsetting system alignment. During 50 insertion/ removal cycles of the module, the standard deviation of the spots on the OE devices was measured to be less than 3 µm (chapter 6).
- Development of a new formalism that quantifies the ability of a given optical module to be misalignment-tolerant. This is used to identify the class of 2D free-space optical interconnect (2D-FSOI) designs that are inherently tolerant to misalignment. Such designs require little alignment compensation mechanisms, and ultimately, may allow for modules to be passively aligned to one another. General design guidelines derived from this work are presented (chapter 7).

Evidence of the impact of the above contributions is manifested by the journal articles [47]-[50] and conference papers [51]-[58] that have resulted.

1.8 Thesis Organization

This thesis is organized as follows. Chapter 2 provides a wide overview of current 2D-POI technologies including flip-chip technology, OE devices (LED, VCSEL, MQW modlished inter-module alignment techniques, including both mechanical and optical methods and device array redundancy. Next, the problem of interfacing the chip module to the freespace backplane demonstrator is examined and two techniques are proposed. The first technique implements a kinematic fixture using lithographically-defined ultrathick photoresist micro-structures fabricated directly on the optical substrates. The second technique implements a semi-kinematic fixture using a pair of dowel pins and the front surface of the optical substrates to constrain the chip module in all six DOFs. Both approaches are implemented and tested.

The work presented in chapter 7 seeks to identify the types of optical designs that can lead to modules that are inherently misalignment-tolerant. The analysis focuses on the design of the chip module and five different optical configurations are examined and compared. The first objective is to determine which of these configurations is the most tolerant to misalignment. The second and more important objective is to understand the underlying reasons that makes one configuration more misalignment-tolerant than another. It is shown that the alignability of a module can be adequately specified by the product of its lateral and tilt misalignment tolerances, that this product is an invariant of the optical system, and that misalignment-tolerant modules require a proper balance between lateral and tilt misalignment tolerances. This provides general guidelines for the design of misalignment-tolerant 2D-FSOI modules. A significant outcome of this chapter is the demonstration that practical FSOI systems suffer from an aspect-ratio limitation similar to the one found in electrical interconnects.

Chapter 8 concludes this work. It summarizes and reiterates the thesis central ideas.

1.9 References

- J.W. Goodman, F.J. Leonberger, S.-Y. Kumg, and R.A. Athale, "Optical Interconnections for VLSI Systems," Proc. IEEE, vol. 72, pp. 850-866 (1984).
- [2] A. V. Krishnamoorthy, K. W. Goossen, "Optoelectronic-VLSI: photonics integrated with VLSI circuits," IEEE Journal of Selected Topics in Quantum Electronics, vol. 4, pp. 899-912 (1998).

۰<u>.</u>

- [3] J. A. Trezza, J. S. Powell, C. Garvin, K. Kang, R. Stack, "Large format smart pixel arrays and their applications," 1998 IEEE Aerospace Conference Proceedings (Snowmass at Aspen, CO, USA) vol. 5, pp. 299-310 (1998).
- [4] A. V. Krishnamoorthy, L. M. F. Chirovsky, W. S. Hobson, R. E. Leibengath, S. P. Hui, G. J. Zydzik, K. W. Goossen, J. D. Wynn, B. J. Tseng, J. Lopata, J. A. Walker, J. E. Cunningham, L. A. D'Asaro, "Vertical-cavity surface-emitting lasers flip-chip bonded to gigabit-per-second CMOS circuits," IEEE Phot. Tech. Letters, vol. 11, pp. 128-30, 1999.
- [5] D. V. Plant, J. A. Trezza, M. B. Venditti, E. Laprise, J. Faucher, K. Razavi, M. Châteauneuf, A. G. Kirk, and W. Luo, "A 256 Channel Bi-Directional Optical Interconnect Using VCSEL's and Photodiodes on CMOS," in Optics in Computing 2000, R. A. Lessard and T. Galstian eds., SPIE 4089,1046-1054 (2000).
- [6] MEMS Optical (Huntsville, AL, USA) is one commercial manufacturer of microoptics. (http://www.memsoptical.com/)
- [7] A. Erlich, "Micro-optical integration spurs mass production," Laser Focus World, vol. 34, pp.77-81 (1998).
- [8] F. A. P. Tooley, "Optical interconnects do not require improved optoelectronic devices," 1998 Optics in Computing Conference proceedings, Brugge, Belgium, pp. 14-17 (1998).
- [9] W. H. Hu, L. A. Bergman, A. R. Johnston, C. C. Guest, S. C. Esener, P. K. L. Yu, M.
 R. Feldman, S. H. Lee, "Implementation of Optical Interconnections for VLSI," IEEE Trans. Electron. Devices, vol. 34, pp. 706-714 (1987).
- [10] M. R. Feldman, S. C. Esener, C. C. Guest, and S. H. Lee, "Comparison between optical ans electrical interconnects based on power and speed considerations," Applied Optics, vol. 27, pp. 1742-1751 (1988).
- [11] D. A. B. Miller, "Optics for low-energy communication inside digital processors: quantum detectors, sources and modulators as efficient impedance converters," Optics Letters, vol. 14, pp. 146-148 (1989).

- [12] C. Fan, B. Mansoorian, D.A.Vanberklom, M.W. Hansen, V. H. Ozguz, S.C. Esener, G.C. Marsden, "Digital Free-Space Optical Interconnects: A Comparison of Transmitter Technologies," Applied Optics, vol. 34, pp. 3103-3115 (1995).
- [13] F. A. P. Tooley, "Challenges in optically interconnecting electronics," IEEE J. Selected Topics in Quantum Electronics, vol. 2, pp. 3-13 (1996).
- [14] A. V. Krishnamoorthy, D. A. B. Miller, "Scaling Optoelectronic-VLSI Circuits into the 21st Century: a Technology Roadmap," IEEE J. Selected Topics in Quantum Electronics, vol. 2, pp. 55-76 (1996).
- [15] D. A. B. Miller, "Physical reasons for optical interconnects", International J. of Optoelectronics, vol. 11, pp. 155-168 (1997).
- [16] H. M. Ozaktas, "Toward an optimal foundation architecture for optoelectronic computing. 1. Regularly interconnected device planes," Applied Optics, vol. 36, pp. 5682-5696 (1997).
- [17] H. M. Ozaktas, "Toward an optimal foundation architecture for optoelectronic computing. 2. Physical construction and application platforms," Applied Optics, vol. 36, pp. 5697-5705 (1997).
- [18] D. A. B. Miller and H. M. Ozaktas "Limit to the Bit-Rate Capacity of Electrical Interconnects from the Aspect Ratio of the System Architecture," Special Issue on Parallel Computing with Optical Interconnects, J. Parallel and Distributed Computing, vol. 41, pp. 42-52 (1997).
- [19] G. I. Yayla, P. J. Marchand, S. C. Esener, "Speed and energy analysis of digital interconnections: comparison of on-chip, off-chip and free-space technologies," Applied Optics, vol. 37, pp. 205-227 (1998).
- [20] A. V. Krishnamoorthy, "Application of optoelectronic-VLSI technologies," International J. of Optoelectronics, vol. 12, pp. 155-161 (1998).
- [21] Semiconductor Industry Association, The International Technology Roadmap for Semiconductors, 1999 edition, http://notes.sematech.org/ntrs/PublNTRS.nsf Data collected from pages 4, 6, 10, 12, 14, 217, 218, 219, 220.
- [22] Semiconductor Industry Association, *ibid.*, p.2.

- [23] J. W. Dally and W. J. Poulton, *Digital systems engineering*, 1998 Cambridge University Press, pp. 19-20.
- [24] D. P. Seraphim and D. E. Barr, "Interconnect and packaging technology in the 90's," in Proc. SPIE 1390, pp. 39-54 (1990).
- [25] M. Donlin, "Device packaging meets increased I/O and speeds demands," Computer Design, October 1995, pp. 32-37
- [26] R. O. C. Neugerbauer, R. A. Fillion, and T. R. Haller, "Multichip module designs for high performance applications," in *Multichip Modules, Compendium of 1989 papers*: International Electronic Packaging Society, pp. 149-163 (1989).
- [27] J. A. Gregus, M. Y. Lau, Y. Degani, K. L. Tai, "Chip-scale modules for high-level integration in the 21st century," Bell Labs Technical Journal, July-September 1998, pp. 116-123.
- [28] Irvine Sensors Corporation, "Design breakthrough: 3D chip packaging saves space, packs more memory," NASA Technical Briefs, vol. 17, no. 5, 1993.
- [29] B. J. Smith, "Interconnection networks for shared memory parallel computers," Proc. 2nd International Conf. on Massively Parallel Processing Using Optical Interconnections (MPPOI '95) (IEEE Computer Society Press, Los Alamitos, CA), pp. 255-256.
- [30] W. J. Dally, J. Poulton, IEEE Micro, January-February 1997, pp. 48-56.
- [31] 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, D. A. B. Miller, "Arrays of optoelectronic switching nodes comprised of flip-chip bonded MQW modulators and detectors on silicon CMOS circuitry," IEEE Photonics Technology Letters, vol. 8, pp. 221-223 (1996)
- [32] A. Krishnamoorthy, A. L. Lentine, K. W. Goossen, J. A. Walker, T. K. Woodward, J. E. Ford, G. F. Aplin, L. A. D'Asaro, S. P. Hui, B. Tseng, R. Leibenguth, D. D. Kossives, D. W. Dahringer, L. M. F. Chirovsky, D. A. B. Miller, "3D integration of MQW modulators over active submicron CMOS circuits: 375 Mb/s transimpedance receiver-transmitter circuit," IEEE Photonics Technology Letters, vol. 7, pp. 1288-1290 (1995).

- [43] A.G. Kirk, D. V. Plant, T. H. Szymanski, Z. G. Vranesic, J. A. Trezza, F. A. P. Tooley, D. R. Rolston, M. H. Ayliffe, F. Lacroix, D. Kabal, B. Robertson, E. Bernier, D. F.-Brosseau, F. S. J. Michael and E. L. Chuah, "A modulator-based multistage free-space optical interconnection system," in Optics in Computing 2000, R. A. Lessard and T. Galstian eds., SPIE 4089, 449-459 (2000).
- [44] S. Araki, M. Kajita, K. Kasahara, K. Kobota, K. Kurihara, I Redmond, E. Schenfeld, and T. Suzaki, "Experimental free-space optical network for massively parallel computers," Applied Optics, vol. 35, pp. 1269-1281, 1996.
- [45] K. Hamanaka, "Optical bus interconnection system using Selfoc lenses," Optics Letters, vol. 16, no. 16, pp. 1222-1224, 1991.
- [46] G. C. Boisset, B. Robertson, W. S. Hsiao, M. R. Taghizadeh, J. Simmons, K. Song, M. Matin, D. A. Thompson, D. V. Plant, "On-die diffractive alignment structures for packaging of microlens arrays with 2-D optoelectronics device arrays," IEEE Photonics Technology Letters 8, 918-920 (1996).
- [47] M. H. Ayliffe, D.V. Plant, "On the design of misalignment-tolerant free-space optical interconnects," to be published in Applied Optics.
- [48] M. H. Ayliffe, M. Châteauneuf, D. R. Rolston, A. G. Kirk, D.V. Plant, "Six-degreesof-freedom alignment of two-dimensional array components using in-situ off-axis diffractive structures," accepted for publication in Applied Optics, July 2001.
- [49] M. H. Ayliffe, D.R. Rolston, E.L. Chuah, E. Bernier, F.S.J. Michael, D. Kabal, A.G. Kirk, D.V. Plant, "Design and testing of a kinematic package supporting a 32 × 32 array of GaAs MQW modulators flip-chip bonded to a CMOS chip," to be published in IEEE J. of Lightwave Technology, October 2001 issue.
- [50] M. H. Ayliffe, D. Kabal, F. Lacroix, E. Bernier, P. Khurana, A. G. Kirk, F. A. P. Tooley, D. V. Plant, "Electrical, thermal and optomechanical packaging of large 2D optoelectronic device arrays for free-space optical interconnects," Journal of Optics A: Pure and Applied Optics, vol. 1, pp. 267-271 (1999).
- [51] M. H. Ayliffe, D. V. Plant, "On the design of misalignment-tolerant free-space optical interconnects," Optics in Computing 2000 Conference, June 18-23 Québec City, Qc., Canada.

Chapter 2: Two-dimensional parallel optical interconnect (2D-POI) technologies

2.1 Introduction

For long-distance optical communications, data is transmitted serially over a single optical fiber using a technology that is now very mature. The bandwidth capacity of long-distance links has recently exceeded Tbit/s data rates through the use of dense wavelength division multiplexing (DWDM), where data encoded on different wavelengths are multiplexed into a single fiber [1]. To implement Tbit/s links between silicon VLSI circuits over cm's distances, the same WDM technology does not seem viable because it requires several components to multiplex different wavelengths onto a single fiber, each wavelength being modulated at high bit rates (i.e. 10 Gbit/s) using sophisticated electronics.

A better approach to the problem of providing Tbit/s capacity between silicon VLSI chips is to use large 2D arrays of optical channels operating at data rates compatible with mainstream silicon VLSI technology (i.e. 32×32 channels operating at 1 Gbit/s). Using a large number of low-bandwidth channels avoids the need for data mux/demux, thereby reducing size, cost and power dissipation. In addition, the ability to transfer data in parallel 2D format is a significant advantage for the class of applications described in section 1.5.

To support 2D parallel optical interconnections (2D-POIs) between silicon VLSI circuits, two enabling technologies are being developed. First, there is a need for the close integration of large 2D arrays (i.e. 16×16 and larger) of surface-normal OE devices with silicon VLSI circuits. This is commonly referred to as optoelectronic-VLSI (OE-VLSI) technology [2]. Second, there is a requirement for guided-wave or free-space optical hardware, with its associated packaging, to relay 2D arrays of signal beams between chips.

The aim of this chapter is (i) to provide a good overview of the state-of-the-art 2D-POI technologies and (ii) to focus on the specific aspects of these technologies that may impact the alignment and packaging of 2D-POI systems.

2.2 Optoelectronic-VLSI (OE-VLSI) technology

The philosophy behind OE-VLSI technology is to take advantage of the density and functionality of microelectronics for data processing, while exploiting the quantum impedance conversion property of OE devices and the parallelism of optics for data communication. What follows is a brief discussion of OE devices and the techniques used to integrate arrays of them onto VLSI chips.

2.2.1 Silicon versus III-V material systems

While silicon has proven to be the material of choice for VLSI microelectronics, silicon-based OE devices have very limited capabilities. Photodetectors can be implemented directly in a standard CMOS process but the relatively long absorption length (~14 μ m at 850 nm [9]) compared to the shallow depletion region thickness leads to two significant problems: (i) low detector responsivity and (ii) slow response time due to diffusive carriers generated deep in the substrate, giving rise to long tails in the detector time response. Various techniques have been demonstrated to increase the speed of CMOS photodetectors but all are limited to moderate data rates (100's of Mbit/s) and very low responsivities (<0.1 A/W) [10]-[12]. In addition, using silicon to fabricate emitters (or modulators) is even more difficult due to the indirect-bandgap property of this material. Several schemes have been investigated (for a review, see [13]) but all result in very low emission efficiency (<10⁻⁴). The conclusion is that silicon-based OE devices (other than possibly photodetectors operating at moderate data rates) are not viable for 2D-POI systems.

For the foreseeable future, it seems likely that the required speed and efficiency of emitters, modulators and detectors will require the use of direct-bandgap III-V semiconductor material. This leads to a heterogeneous environment, where III-V OE devices are required to be integrated with silicon VLSI microelectronics.

2.2.2 Monolithic versus hybrid integration

Integration methods can be categorized as monolithic or hybrid. In the monolithic approach, III-V OE devices are fabricated directly on top of the VLSI circuits (i.e. the semiconductor materials are in direct contact). The main advantages of monolithic integration is a low parasitic capacitance and a good thermal path to the substrate, which ulti-

mately translates into higher device performance. A number of experiments have demonstrated the monolithic integration of GaAs OE devices on lattice-matched GaAs microelectronics [3][4]. The main drawback to this approach has to do with the poor maturity of GaAs microelectronics compared to silicon VLSI in terms of yield, transistor density, chip size, device modeling and availability. Keeping in mind that the primary motivation behind OE-VLSI technology is the ability to provide optical I/O where electrical I/O is insufficient, one may argue that if the electronics cannot be state-of-the-art then the exercise of integrating OE devices may serve little purpose. For this reason, a major effort has been invested in growing III-V materials directly on silicon. Heteroepitaxial growth is problematic because of the dislocations that are formed at the lattice-mismatched semiconductor interface; these dislocations tend to propagate and degrade the OE device performance over time. Although this does not seem to be a problem for multiplequantum-well (MQW) modulators on silicon substrates [5], there still remains significant practical obstacles to the growth of III-V devices on pre-fabricated CMOS VLSI chips [6][7]. At the time of this writing, a truly monolithic integration technology of OE devices on pre-fabricated CMOS chips has not yet emerged.

A more viable approach has been to focus on hybrid integration techniques. This is attractive because it requires no modification of the standard CMOS process. The growth and fabrication of OE devices are done separately from the CMOS chip fabrication, which allows greater flexibility in the choice and design of the OE devices. Various hybrid integration techniques have been developed (for a review, see [2][8]), but perhaps the simplest and most mature method is flip-chip bonding which is a derivative of the C4 (Controlled Collapse Chip Connection) process developed by IBM in the early 1960's [14].

The flip-chip process (also called bump-bonding or solder-bonding) works as follows. The starting point is two wafers (one with CMOS circuits, the other with III-V OE device arrays) covered with a passivation layer. Both wafers have metal pads at the locations of the flip-chip interconnections. Small openings are exposed in the passivation layer on top of the metal pads and a solder-wettable metal layer is deposited over the window. The solder-wettable metal usually contains a thin layer of a barrier metal (e.g. Cr, Ni, Pd) that act as a barrier that prevents solder from diffusing into the underlying devices. Solder is then

Chapter 2: Two-dimensional parallel optical interconnect (2D-POI) technologies

evaporated (or plated) on one or both wafers. The wafers are then diced into chips. One CMOS chip and one OE chip are placed face-to-face and brought into coarse alignment. The solder is then reflowed and both chips are pulled into precise alignment due to the surface tension forces of the melted solder. Alignment precision of the order of 1 μ m is achieved routinely. The size of the flip-chip metal pad can reliably be made as small as $15 \times 15 \mu$ m, resulting in high-density area interconnections with a parasitic capacitance as low as 50 fF [15].

2.3 Transmitter technologies

Various studies have been carried out to determine the optimum transmitter technology for 2D-POI systems [16][17]. There are three candidate technologies: light-emitting diodes (LEDs), vertical-cavity surface-emitting lasers (VCSELs) and electro-absorption (EA) modulators. What follows is a review of these technologies. LEDs are only briefly covered; the main focus is on VCSELs and EA modulators because only they have the potential to provide Tbit/s capacity between silicon chips.

2.3.1 Light-emitting diodes (LEDs)

The main benefits of LEDs are their low cost, high yield and good reliability due to a simple fabrication process and relaxed tolerances to process variations. One major obstacle to the use of LEDs is their slow response time which is determined by the electron-hole recombination lifetime. This leads to a maximum modulation rate limited to 100's of Mbit/s. The modulation rate can be improved by operating the device at high current densities such as to shorten the spontaneous lifetime but this reduces efficiency and increases power dissipation. A second impediment to the use of LEDs is the poor directionality of the light emission. Collecting the emitted light and focusing it onto a small-area detector is difficult to do with LEDs. Attempts at reducing the problems of poor efficiency and large angular spread have been sought through the use of microcavities (MC-LEDs) [18]. The conclusion is that MC-LEDs are a possible candidate technology for low-cost 2D-POI applications with low-density interconnections operating at moderate speeds (< 1 Gbit/s).

2.3.2 Vertical-cavity surface-emitting lasers (VCSELs)

VCSELs have optical cavities oriented orthogonal to those of conventional edge-emitting lasers [19][20]. This unique attribute leads to many advantages, including the capability of wafer-level testing before packaging and the ability of fabricating 2D device arrays. This results in a technology that is amenable to low-cost and high-volume manufacturing. In addition, VCSELs typically emit circularly-symmetric Gaussian beams which can be efficiently coupled into an optical system and focused down to a small spot size.



Figure 2.1. Cross-section of a selectively-oxidized top-emitting VCSEL.

A cross-sectional drawing of a VCSEL is shown in figure 2.1; it is composed of two high-reflectivity (~99%) distributed Bragg reflectors (DBRs) separated by a thin active gain region (~1- λ thick) to form a high-finesse Fabry-Perot cavity. A DBR consists of alternating pairs of quarter-wavelength-thick high and low-refractive-index monolithically-grown semiconductor layers. The active region usually contains 1 to 5 quantum wells (QWs) designed for light emission at a specific wavelength. A complete treatment of VCSELs can be found in [21].

The pumping current is injected into the metal contact deposited on the top DBR. Current flows through the DBR layers and reaches the active region where it creates a condition of population inversion. Two methods are used to achieve lateral current confinement within the VCSEL cavity [22]. First-generation VCSELs used ion implantation to create
crystalline damage thereby turning the region surrounding the laser cavity into an insulating material. Although this approach defines an electrical path for the current, it provides poor optical confinement which tends to limit the performance of the device in terms of threshold current and modulation characteristics. Ion-implanted VCSELs are readily available commercially from companies such as Honeywell and Mitel with threshold currents of 3-6 mA and 3-dB modulation bandwidths in the range of 2-6 GHz. Honeywell has reported a yield of 99.8% across 3-inch wafers with excellent long-term reliability [23].

Second-generation VCSELs employ selective oxidation of buried AlGaAs layers to form an oxide-confined aperture within the VCSEL cavity [24][25]. The VCSEL drawn in figure 2.1 is a selectively-oxidized VCSEL. The oxide layers have a low refractive index which provides both electrical and optical lateral confinement. The advent of oxide-confined VCSELs has produced remarkable performance advances: threshold current below 100 μ A [24], threshold voltage of 1.3 V [26], power conversion efficiency greater than 50% [27][28] and small-signal modulation bandwidth of 20 GHz [29]. Modulation rates of 10 Gbit/s have been reported [30] and will be commercially available as of 2001 [31].

The initial commercial application of VCSELs focused on their use as a low-cost and high-speed (> 1 Gbit/s) alternative to LEDs in local area networks (LANs). VCSELs have also been incorporated in parallel optical data links (ODLs) which use a linear array of devices coupled to a multimode fibre ribbon cable to support high-aggregate-bandwidth data communication over 100's of meters. Parallel ODLs with an aggregate bandwidth of 30 Gbit/s (12 channels at 2.5 Gbit/s) are currently available from companies such as Picolight, Infineon Technologies, Agilent and Mitel.

The main attractions for using VCSELs in 2D-POI systems is (i) their high contrast ratio (>10:1) which simplifies the design of the receiver and (ii) the fact that, unlike EA modulators, no external laser source is required which simplifies the design of the optics. Despite this, VCSELs have seen little use so far in 2D-POI systems. The main reason has to do with the fact that a VCSEL structure cannot readily be used as an efficient detector and so a different technology is typically required to implement the detector array on the OE-VLSI chip. One approach has been to fabricate metal-semiconductor-metal (MSM) detectors on the same substrate as the VCSELs [32]. Another approach has been to use separately-fabricated p-i-n photodiodes, but this requires an additional flip-chip step using a lower melting point solder. The hybrid integration of 256 VCSELs and 256 p-i-n photodiodes on CMOS has been successfully demonstrated using this technique [33]. Recent work has focused on methods of turning a VCSEL structure into an efficient resonant cavity-enhanced photodiode [34], thereby avoiding the problem of an additional flip-chip step. Despite these advances, there still remains a large number of practical considerations associated with the use of VCSELs in dense 2D-POI systems:

- Threshold current: dense 2D-POI systems (16×16 array and larger) will likely require threshold currents below 100 μ A. A low threshold removes the problem of pattern-dependent turn-on delay associated with bias-free modulation [35]. The ability to perform bias-free modulation is desirable because it removes the need for biasing circuitry with its associated power dissipation and chip area requirements. Bias-free modulation at 2.5 Gbit/s has been demonstrated [36].
- Thermal management: a problem with VCSELs is that the heat source is confined to a very small volume, resulting in a package thermal resistance that is dominated by the thermal spreading resistance in the region surrounding the device [37]. The thermal resistance of a flip-chip bonded VCSEL with a 15 µm aperture is >1000 °C/W [38]. Decreasing the aperture to further reduce the threshold current does not help; this increases the spreading thermal resistance and leads to a still higher operating temperature. High-temperature operation must be avoided because lifetime is reduced by 50% with every 10 °C increase in junction temperature [37].
- Wavelength control: lasing wavelength is a function of the cavity dimensions and operating temperature. Wavelength uniformity and stability is an important consideration if diffractive optics is to be used in the interconnect. An increase in temperature increases the refractive index which shifts the emission wavelength to longer wavelengths with a typical sensitivity of ~0.06 nm/°C [39].
- Transverse mode behaviour: at high current injection, VCSELs generally emit multiple transverse modes. This is undesirable because it leads to unstable output optical

strong change in absorption is what enables the use of flip-chip surface-normal QCSE devices for 2D-POI applications.

A review of QCSE modulators can be found in [46]. A surface-normal QCSE modulator is a reverse-biased p-i-n diode with the intrinsic region consisting of a MQW structure, which is why QCSE modulators are also referred to as MQW modulators. The MQW structure is a repeated sequence of layers of large bandgap and small bandgap semiconductor material whose individual layers are so thin (<100 Å) that electrons and holes are quantum mechanically confined to discrete energy states in the layers of the smaller bandgap materials. This quantization of the electron and hole energies leads to strong (excitonic) peaks in the absorption spectrum. The peak closest to the absorption edge is commonly referred to as *the* exciton peak; its presence provides an abrupt absorption edge which increases device performance. The basic principle behind the operation of QCSE modulators is the wavelength shift of the exciton absorption peak with the application of an electric field perpendicular to the MQW layers [44]. Associated with the wavelength shift is a reduction and broadening of the exciton absorption peak. These effects are shown in figure 2.2, where the measured absorption spectrum of a GaAs/AlGaAs QCSE modulator is plotted for different reverse-bias voltages.



Figure 2.2. Absorption spectrum of a GaAs/AlGaAs QCSE modulator at T = 40 °C.

A QCSE modulator provides intensity modulation by directing a narrow-linewidth continuous-wave (CW) laser beam (traditionally called a "read-off" beam) perpendicular to the MQW layers and modulating the voltage across the device. Figure 2.2 corresponds to a device optimized for operation at $\lambda = 852$ nm and T = 40 °C. In this case, a 5-V modulation voltage swing is applied from 1 V (high reflectivity state) to 6 V (low reflectivity state). MQW structures based on the GaAs/AlGaAs material system are most often used because they offer maximum modulation contrast and are easy to grow [47].

In 2D-POI applications, QCSE modulators are operated in reflection mode because this simplifies the integration with silicon VLSI. Figure 2.3 describes how a reflectionmode 850-nm GaAs/AlGaAs QCSE modulator is integrated onto a silicon chip using flipchip bonding techniques [48]. The first step consists of depositing solder on metal pads located on one or both of the chips to be bonded (figure 2.3(a)). Chips are then brought together and bonded under careful alignment and controlled temperature and pressure. Next, epoxy is flowed between the chips (figure 2.3(b)); this adds mechanical stability to the flip-chip structures and acts as an etch-protectant for the subsequent substrate removal step. Next, the entire GaAs substrate is removed using a selective chemical etch (figure 2.3(c)). This is required because GaAs is absorptive at 850 nm. The last step consists of depositing an anti-reflection (AR) coating to the top of the device (figure 2.3(d)).



Figure 2.3. Flip-chip bonding of GaAs/AlGaAs modulators onto silicon VLSI.

Chapter 2: Two-dimensional parallel optical interconnect (2D-POI) technologies

QCSE modulators have been used in several large-scale 2D-POI system demonstrations [49]-[52]. Three factors explain their extensive use:

- Modulators are efficient detectors: the same reverse-biased p-i-n diode structure can be used as a modulator or a detector. In fact, it is the underlying CMOS electronics that distinguishes between modulators and detectors. This property has been used to create a dual-function modulator-detector device [53]. Device homogeneity facilitates integration with silicon because devices are coplanar and a single flip-chip step is required.
- Large arrays and high yields: modulators have been developed for a long period and can be reliably produced in large arrays. Researchers at Bell Labs have produced zerodefect arrays containing more than 4000 devices with an average yield of 99.9% [54]. Lockheed Martin has demonstrated 256 × 256 arrays with an average device yield of 99.98% [55].
- High-speed operation: QCSE modulators can operate at very high data rates. The QCSE itself occurs over a sub-picosecond time scale [46]. In practice, the speed is only limited by the speed at which the electronics can charge and discharge the device capacitance. A 1-µm thick GaAs modulator has a capacitance per unit area equal to 0.12 fF/µm² [46]. Thus, a 20 × 20 µm² modulator has a capacitance of only 48 fF and is conveniently driven with a small-size CMOS inverter. QCSE modulators have been demonstrated with speeds in excess of 40 GHz [56], limited by the driving electronics.

An important consideration related to the choice between modulators and emitters is the relative complexity of the optical system. Modulators require the external generation of a 2D spot array; this can be done efficiently using a single high-power laser in combination with diffractive array generator elements [57][58]. The use of an external laser source is usually considered a disadvantage because it adds to the size of the system, increases the complexity of the optics and requires an additional alignment step. From a different perspective, the use of an external laser source can be considered an advantage because it is only necessary to control a single laser beam in terms of wavelength, polarization and modal quality. Modulators also avoids the problem of turn-on delay associated with VCSELs. Finally, the use of a mode-locked laser source allows for centralized clocking and resynchronization of signals at each node in the system [59]. The use of short optical pulses is also beneficial because it improves receiver sensitivity [60].

A major advantage of modulators is their low power dissipation which allows them to be used in dense arrays (1000's per chip) [61]. The amount of power dissipated is the sum of the power required to set up a modulation voltage V_m across the device $(C_{mod}V_m^2/2\Delta t)$ and the power consumed by the photocurrent $(V_{tot}I_{ph})$, where V_{tot} is the sum of the bias and modulation voltages. The latter term can usually be neglected considering the low optical power levels (100's of μ W) used in practical systems. For example, a 20 × 20 μ m² device $(C_{mod} = 48 \text{ fF})$ modulated with a 5-V swing at 1 Gbit/s (with a 10%-90% rise time equal to $\Delta t \sim 0.3$ ns) dissipates only 2 mW. In addition, the use of a remote laser source takes the thermal management problem away from the OE-VLSI chip.

Despite these attractive attributes, there still remains some practical considerations associated with the use of modulators in 2D-POI applications:

- Contrast ratio: absorption modulators exhibit poor contrast ratio; for flip-chip devices, it is typically 2:1 for a 5-V swing [62]. Low contrast combined with unpredictable and non-uniform losses along the optical path complicates the design of single-ended receivers. This has lead to the use of differential receivers where one data signal is encoded on two spatially separated optical beams at the expense of a 50% reduction in optical I/O [63]. Contrast can be improved by incorporating a Fabry-Perot cavity in the structure. A contrast of >100:1 has been demonstrated on individual devices [64]. Lockheed Martin has also demonstrated 10:1 contrast over a 256 × 256 array [65].
- Temperature sensitivity: modulators are usually designed for optimal performance at a given operating wavelength and temperature. Any variation in temperature results in a ~0.27 nm/°C shift of the excitonic peak [66] and a corresponding decrease in modulation contrast. Under constant biasing conditions and 5-V modulation, the useful temperature range is limited to about 10 °C [67]. This limited temperature range may represent a serious obstacle for adopting modulators in commercial systems.

Recent work has also focused on the use of resonant cavity-enhanced (RCE) structures to provide high-speed and high-efficiency photodetectors. The presence of an optical cavity increases responsivity over a certain wavelength range without degrading the electrical properties of the device. A RCE detector operating at 840 ± 5 nm with a 3-dB bandwidth of 50 GHz has been reported [78]. The ability to turn a VCSEL structure into an efficient RCE detector (or vice-versa) is an important step in the development of a low-cost emitter-based OE-VLSI technology [34].

2.5 Optical receivers

The types of optical receivers used in 2D-POI systems differ significantly from the ones used in long-distance optical telecommunications. This is not surprising considering that what is important for 2D-POI systems (high channel density, small area, low latency, low power, moderate speed, low cost) is generally not of primary importance to long-distance telecommunications. Desirable characteristics for 2D-POI receivers are [79]:

- Small area: in dense arrays, the size of the receiver circuit is limited to about 40×40 μ m², allowing it to fit underneath the detector and solder bump.
- Low power: power dissipation should be limited to <1 mW per receiver. Receiver power is usually the main factor limiting array scalability in 2D-POI systems [61].
- High sensitivity: a sensitivity better than 10 μ W (-20 dBm) is required. The higher the sensitivity, the lower the optical power requirement at the transmitter end.
- Moderate speed: the receiver must operate at on-chip clock data rates.

The most common receiver configuration is the transimpedance amplifier, which can be designed as a single-beam or two-beam receiver [80][81]. A review of optical receivers for 2D-POI applications can be found in [63].

The speed of a receiver is usually determined by its front-end stage where the signal rise time (τ) is limited by the *RC* time constant seen by the input photocurrent. The front-end input signal rise time is given by:

$$\tau = (ln9)R_{in}C_{in} \tag{2.1}$$

where R_{in} and C_{in} represent the input resistance and input capacitance seen by the photocurrent. The input capacitance C_{in} is the sum of the detector shunt capacitance (C_{det}) , the front-end transistor capacitance (C_{FET}) and the wiring capacitance (C_{wire}) , the latter being usually negligible compared to the previous two. The receiver bit rate, *B*, is limited by the signal rise time. Using equation 2.1 and the fact that $B < 1/\tau$, one can write:

$$R_{in}C_{in}B < constant \tag{2.2}$$

Equation 2.2 illustrates the design trade-off between speed ($\propto B$), sensitivity ($\propto R_{in}$) and detector area ($\propto C_{in}$). On the one hand, high bit rate operation is realized by minimizing R_{in} and C_{in} . On the other hand, lowering R_{in} reduces receiver sensitivity and lowering C_{in} (by reducing detector area) leads to a small spot size requirement and an optical system that is difficult to align.

The above constraints can be mitigated by using a current-mode front-end stage [82]. A current-mode receiver offers a very small input resistance. The front-end stage acts as a current buffer that decouples the detector capacitance from the high resistance of the gain stage [83][84]. This renders the receiver bit rate almost insensitive to detector capacitance, opening the way for large-area detectors and misalignment-tolerant systems.

2.6 Guided-wave interconnection technologies

One approach at providing 2D optical interconnections between a pair of OE-VLSI chips uses a flexible guided-wave interconnection, as shown in figure 2.4.





- Precision hole arrays: fiber arrays can be assembled by inserting individual fibers through a substrate having a 2D array of precision holes. The accuracy of this method depends on (i) the tolerance on the hole location, (ii) the clearance between the fiber cladding and the hole and (iii) the concentricity of the fiber. This approach has been demonstrated using various substrate materials: etched holes in silicon [91], etched holes in photosensitive glass [92], drilled hole in stainless steel [93], drilled holes in thin kevlar [94] and drilled holes in polyimide [91].
- Crossing grooves in glass plate: this method consists of inserting individual fibers at the crossings of horizontal and vertical grooves grounded in a glass plate [95]. Precise alignment is achieved by inserting a "fiber jack" into an "optical plug". The "fiber jack" is made by partially etching the core at the fiber tip and an "optical plug" is made by depositing polymide bumps using a self-align process.
- Microferrules in a square frame: individual fibers are first inserted into glass microferrules and fixed with adhesive. Glass microferrules are cheap and highly accurate. Next, the microferrules are placed inside a square frame made of zirconia plates on all sides. Sufficient pressure is applied on the outside frame in all four directions, forcing the fibers to make contact with each other and form a 2D array. Adhesive is then injected into the gaps between microferrules. An 8 × 8 array of singlemode polarization-maintaining fiber has been assembled using this technique [96].

The first three techniques are labour-intensive and typically provide fiber positioning accuracy of the order of 5 - 10 μ m, which is appropriate for multimode fiber arrays but not sufficient for singlemode applications. Stacking microferrules in a square frame simplifies assembly and provides increased positioning accuracy (2 μ m accuracy was achieved in [96]). 2D fiber arrays are commercially available from companies such as Furukawa Electric and Sumitomo Electric Industries.

2.6.2 Fiber image guides (FIGs)

The alternative guided-wave interconnection technology is the flexible fiber image guide (FIG), a technology that was originally developed for endoscopic medical applications. FIG-based parallel optical interconnections are being pursued by several researchers [97]-[100]. A FIG consists of an array of densely-packed fiber cores arranged in a hexagonal lattice, as shown in figure 2.6. Fiber core diameters are usually in the range 5 - 20 μ m; their relative position is maintained throughout the length of the bundle. Core densities in the range 2,000 - 15,000 per mm² are achievable.



Figure 2.6. A fiber image guide with 10-µm fiber core (after [100]).

As an optical interconnect medium, FIGs represent an oversampled approach where a single beam is transmitted over a cluster of fiber cores in the image guide. The use of FIGs relaxes alignment tolerances and allows for one or more fibers to be damaged without losing data transmission. The main drawback of FIGs is their high attenuation levels (0.4 dB/ meter has been measured in the 700 - 1100 nm range [101]) which limits the interconnection distance to about 50 m (20 dB link budget). Total pulse broadening (combining effects of skew and multimode dispersion) is of the order of 2 ps/m [102], allowing for Gbit/s bit rate operation. A brief description of the FIG fabrication process can be found in [99]. FIGs are commercially available from Schott Fiber Optics.

2.7 Free-Space interconnection technologies

For short-distance interconnects (<10 cm), free-space optics is often preferred over guided-wave technologies. The main reason for this is due to the cost of manufacturing and connectorizing fiber arrays and image guides. Another reason has to do with the difficulty in realizing parallel multi-point interconnections using 2D guided-wave interconnects. While fiber arrays and image guides are efficient at providing parallel point-to-point

links between two chips, they are not readily suitable for implementing multistage architectures across multiple chips (e.g. broadcast, multicast).

One potential solution to the architectural limitations of guided-wave technologies has been proposed recently; it consists of partitioning the optical I/O area of the OE-VLSI chip and using segments of flexible image guides bonded together to create a composite structure [103]. Each optical I/O partition can send and receive signals through a dedicated segment of the image guide. Image guide segments can be bonded to form various types of interconnect topologies (4-way fan out, ring, nearest-neighbour, crossbar). Unfortunately, this approach does not come without the problem of having a dense mesh of electrical interconnections running across the surface of the CMOS chip. This routing requirement rapidly becomes a connectivity bottleneck, limiting the scalability of this approach.

Free-space optics uses refractive and/or diffractive components to relay dense 2D array of beams between two or more OE-VLSI chips. Arbitrary interconnection patterns (space-variant interconnects) can be implemented between multiple chips. Input and output OE devices are usually placed close to one another on the OE-VLSI chip; this minimizes on-chip routing requirements while reducing latency and power dissipation.

The ideal free-space 2D-POI system possesses the following attributes: misalignmenttolerant (>25 μ m, enough for passive alignment), large field of view (>1 cm²), small spot size (10 μ m diameter), low loss (>10% link efficiency), light weight, compact, ease of integration into modules, and compatibility with high-volume production. Free-space optical relays have been implemented using conventional compound lenses, microlens arrays, hybrid lenses (combining conventional lenses with microlens arrays) and mini-lens arrays. These choices are described and compared below.

2.7.1 Conventional Lenses

Conventional lenses have been used in several free-space photonic switching systems [49][51][104][105]. A typical optical relay based on conventional lenses is shown in figure 2.7. The main advantage of conventional lenses is their ease of alignment due to their large clear aperture. Their circular symmetry also simplifies rotational alignment.

The major disadvantage of conventional lenses is their large size. This comes from the requirement of providing high-resolution imaging over a large field of view. High-resolu-

tion imaging requires the use of a low f-number lens (e.g. a 10 μ m spot size requires a f/5 lens at 850 nm) and a low f-number lens with a large field of view is necessarily multi-element. For example, it has been shown that a f/5 lens with a field of view of 8 mm requires four lens elements; the resulting lens is 25 mm in diameter and 80 mm long [105]. The size, weight and cost of this lens make it unsuitable for commercial applications.



Figure 2.7. Conventional lens relay.

2.7.2 Microlens arrays

The size and cost of the optics can be significantly reduced by realizing that high-resolution imaging is not required at every point across the image field but only at the locations of the optical I/O devices. This is to say that the large space-bandwidth product (SBWP) of conventional lenses is not being fully utilized [109]. The microlens-based relay in figure 2.8 is a means by which the required high-resolution imaging is supplied only at the optical I/O devices [106]. Microlens-based relays have been used in free-space system demonstrators [107][108]. The major advantages of microlenses are (i) their small size and low weight, (ii) their potential for low-cost high-volume production [110] and (iii) the fact that they are fabricated on flat substrates, which facilitates their packaging with chips, beamsplitters, etc.



Figure 2.8. Microlens relay.





Figure 2.9. Hybrid lens relay.

Drawbacks to hybrid systems are that (i) they require more components (higher complexity, higher losses) and (ii) they propagate beams at large off-axis angles through the beamsplitting element - this especially becomes a problem when polarizing beamsplitters and retarders are used. Hybrid imaging systems have been used in several free-space system demonstrations [111]-[113].

2.7.4 Mini-lens arrays

An alternative approach, which might prove to be superior, is shown in figure 2.10. It consists of grouping optical devices to form small clusters on the chip $(2 \times 2, 4 \times 4, \text{ etc.})$ and using a lens array with each lens imaging an entire cluster. The lenses are usually referred to as "mini-lenses" to underscore the fact that they relay multiple beams at once and thus usually have a larger aperture than the microlenses of figure 2.8. Packing optical devices into small clusters relaxes the requirement for a large image field, allowing for a compact and low-cost mini-lens array to be used. The use of mini-lens arrays in 2D-POI systems has been demonstrated recently [33][115].



Figure 2.10. Mini-lens relay.

Compared to hybrid systems, the mini-lens approach requires only two lens arrays, reduces off-axis beam propagation, and avoids the use of bulky conventional lenses which

results in compact packaging. Compared to microlens-based, the mini-lenses have a larger aperture which relaxes lateral misalignment tolerances and significantly improves the scalability in terms of array size and interconnect length [114].

As a comparison, the numerical example of section 2.7.2 (chip size = $10 \times 10 \text{ mm}^2$, interconnect length = 50 mm, spot size $2\omega_d = 10 \mu$ m, wavelength = 850 nm) can be used to show that a 32×32 device array can be easily accommodated using a 4×4 array of mini-lenses (2.50 mm aperture) with a 8×8 array of devices per cluster. This calculation assumes (i) a device diameter of $3\omega_d$, (ii) a device pitch equal to twice the device diameter ($6\omega_d$) and (iii) a mini-lens that operates over a field of view of 320 µm.

2.8 Planar optics

The free-space optical relays described previously have in common that the chips are located in different planes with beams propagating perpendicular to their surface. An alternative approach, commonly referred to as planar optics, consists of placing the chips in the same plane by mounting them on the surface of a flat transparent substrate. The optical system is thus folded into a 2D geometry. Optical elements (e.g. lenses, gratings, mirrors) are integrated on the surfaces of the substrate to route optical signals between chips. Optical beams propagate off-axis, following zig-zag paths inside the substrate. The key feature of planar optics is its 2D geometry which enables the use of standard manufacturing techniques such as microlithography, dry etching, flip-chip bonding, and thin film deposition. Examples of the planar optics concept can be found in [116]-[118].



Figure 2.11. Planar optical interconnections between multiple chips (after [118]).

onto a single fiber is prohibitively high. The cost of the WDM approach can be considerably reduced by using uncooled laser sources and a wide wavelength spacing (e.g. 20 nm) - this is the philosophy behind coarse WDM (CWDM) [119]. Over long distances (>500 m), parallel fiber ribbons are undesirable because of their high cost (the interconnection medium becomes more expensive than the transceivers at both ends) and their inability to be field-terminated (i.e. if a fiber breaks, a new fiber array must be reinstalled).



Figure 2.12. Optical technology hierarchy as a function of interconnection length.

The attributes of 2D fiber arrays and image guides (high parallelism, mechanically robust) make them a good fit for high-bandwidth point-to-point interconnects over distances of 10 cm - 10 m. Their high parallelism avoids the need of serializing/deserializing (SERDES) electronics with its associated cost, area, latency and power dissipation, making them attractive over 1D fiber ribbon links.

Over shorter distances (1 cm - 10 cm), free-space technologies are likely to be used in applications requiring highly parallel interconnections between multiple chips located in different physical planes. An important practical consideration for free-space applications will be the issue of chip separability, that is, the ability to remove and replace an OE-VLSI chip without having to realigning the optical system. A good example is the free-space backplane demonstrator described in chapter 3.

The attributes of planar optics (alignment-free, compact, low-cost) make it attractive for interconnecting multiple chips lying in the same physical plane. It is easy to envision the use of planar optics for inter-chip interconnections inside a multichip module. As an and the absence of a field termination technology. Over the scale of a multichip module (1 mm - 10 cm), the attributes of planar optics (alignment-free, compact, low-cost) make it the technology of choice for dense inter-chip interconnections.

Somewhere in the range 1 cm - 10 cm exist a variety of applications requiring highly parallel interconnections between multiple chips located in different physical planes (e.g. board-to-board interconnects). This niche of applications is likely to be fulfilled by free-space technologies. However, unlike guided-wave or planar optics, 2D free-space inter-connections (2D-FSOI) come with a significant alignment problem, which is further exacerbated by the requirement that OE-VLSI chips be replaceable. A broad range of solutions to this alignment problem are presented in chapters 5 and 6.

2.11 References

- A. R. Chraplyvy and R. W. Tkack, "Terabit/second transmission experiments," IEEE J. of Quantum Electronics, vol. 34, pp. 2103-2108 (1998).
- [2] A. V. Krishnamoorthy and K. W. Goossen, "Optoelectronic-VLSI: Photonics Integrated with VLSI Circuits," IEEE J. of Selected Topics in Quantum Electronics, vol. 4, pp. 899-912 (1998).
- [3] L. A. D'Asaro, L. M. F. Chirovsky, E. J. Laskowski, S. S. Pei, T. K. Woodward, A. L. Lentine, R. E. Leibenguth, M. W. Focht, J. M. Freund, G. G. Guth and L. E. Smith, "Batch fabrication and operation of GaAs-AlGaAs field-effect transistor-self-electro-optic effect device (FET-SEED) smart pixel arrays," IEEE J. Quantum Electronics, vol. 29, pp. 670-677 (1993).
- [4] S. Matsuo, T. Nakahara, Y. Kohama, Y. Ohiso, S. Fukushima, and T. Kurokawa, "Monolithically integrated photonic switch device using and MSM PD, MESFETs, and a VCSEL," IEEE Photonics Technology Letters, vol. 7, pp. 1165-1167 (1995).
- [5] K. W. Goossen, G. D. Boyd, J. E. Cunningham, W. Y. Jan, D. A. B. Miller, D. S. Chemla, and R. M. Lum, "GaAs/AlGaAs multiple-quantum-well reflection modulators grown on GaAs and silicon substrates," IEEE Photon. Technol. Lett., vol. 1, pp. 304-306, 1989.

- [6] K. W. Goossen, J. E. Cunningham, and W. Y. Yan, "GaAs 850 nm modulators solder-bonded to silicon," IEEE Photon. Technol. Lett., vol. 5, pp. 776-778, 1993.
- [7] D. A. B. Miller, "Rationale and challenges for optical interconnects to electronic chips," IEEE Proceedings, pp. 1-44 (2000).
- [8] D. L. Mathine, "The integration of III-V optoelectronics with silicon circuitry," IEEE
 J. of Sel. Quantum Electron., vol. 3, pp. 952-959 (1997).
- [9] S. M. Sze, *Physics of Semiconductor Devices*, New York: Wiley, 1981, Chapter 10.
- [10] K. Ayadi, M. Kjuik, P. Heremans, G. Bickel, G. Borghs, R. Vrounckx, "A monolithic optoelectronic receiver in standard 0.7 μm CMOS operating at 180 MHz and 176-fJ light input energy," IEEE Photonics technology Letters, vol. 9, pp. 88-90 (1997).
- [11] T. K. Woodward and A. V. Krishnamoorthy, "1Gb/s CMOS photoreceiver with integrated detectors operating at 850 nm," Electronics Letters, vol. 34, pp. 1252-1253 (1998).
- [12] M. Kjuik, D. Coppee, R. Vounckx, "Spatially modulated light detector in CMOS with sense-amplifier receiver operating at 180 Mb/s for optical data link applications and parallel optical interconnects between chips," IEEE J. Selected Topics in Quantum Electronics, vol. 4, pp. 1040-1045 (1998).
- [13] L. C. Kimerling, K. D. Kolenbrander, J. Michel, and J. Palm, "Light emission from silicon," Solid State Physics, vol. 50, pp. 333-381 (1997).
- [14] L. F. Miller, "Controlled collapse reflow chip joining," IBM J. Res. Develop., vol.13, pp. 239-250 (1969).
- [15] A. V. Krishnamoorthy, T. K. Woodward, R. Novotny, K. Goossen, J. Walker, A. Lentine, L. A. D'Asaro, S. Hui, B. Tseng, R. Leibenguth, D. Kossives, D. Dahringer, L. Chirovsky, G. Aplin, R. Rozier, F. Kiamilev, and D. A. B. Miller, "Ring oscillators with optical and electrical readout based on hybrid GaAs MQW modulators bonded to 0.8 micron silicon VLSI circuits," Electronics Letters, vol. 31, pp. 1917-1922 (1995).

- [16] C. Fan, B. Mansoorian, D. A. Van Blekom, M. W. Hansen, V. H. Ozguz, S. C. Esener, and G. C. Marsden, "Digital free-space optical interconnections," Applied Optics, vol. 34, pp. 3103-3115 (1995).
- [17] T. Nakaraha, S. Matsuo, S. Fukushima, and T. Kurokawa, "Performance comparison between MQW modulator-based and VCSEL-based smart pixels," Applied Optics, vol. 35, pp. 860-871 (1996).
- [18] R. Bockstaele, J. Derlyun, C. Sys, S. Verstuyft, I. Moerman, P. Vandaele, and R. Baets, "Realisation of highly-efficient 850nm top-emitting resonant-cavity light-emitting diodes," Electronics Letters, vol. 35, pp. 1564-1565 (1999).
- [19] K. Iga, F. Koyama, and S. Kinoshita, "Surface-emitting semiconductor lasers," IEEE J. of Quantum Electronics, vol. 24, pp. 1845-1855 (1988).
- [20] J. L. Jewell, J. P. Harbison, A. Scherer, Y. H. Lee, and L. T. Florez, "Vertical-cavity surface-emitting lasers: design, growth, fabrication, characterization," IEEE J. of Quantum Electronics, vol. 27, pp. 1332-1346 (1991).
- [21] T. E. Sale, Vertical-Cavity Surface-Emitting Lasers, Somerset, U.K.: Research studies (1995).
- [22] K. D. Choquette and H. Q. Hou, "Vertical-cavity surface-emitting lasers: moving from research to manufacturing," Proc. of IEEE, vol. 85, pp. 1730-1739 (1997).
- [23] M. K. Hibbs-Brenner, R. A. Morgan, R. A. Walterson, J. A. Lehman, E. L. Kalweit, S. Bounnak, T. Marta, R. Gieske, "Performance, uniformity and yield of 850 nm VCSELs deposited by MOVPE," IEEE Photonics Technology Letters, vol. 8, pp.7-9 (1996).
- [24] D. L. Huffaker, D. G. Deppe, K. Kumar and T. J. Rogers, "Native-oxide-defined ring contact for low-threshold vertical-cavity lasers," Applied Physics Letters, vol. 65, pp. 97-99 (1994).
- [25] W. W. Chow, K. D. Choquette, M. H. Crawford, K. L. Lear, and G. R. Hadley, "Design, fabrication, and performance of infrared and visible vertical-cavity surfaceemitting lasers" J. of Quantum Electronics, vol. 33, pp. 1810-1824 (1997).

- [26] K. D. Choquette, R. P. Schneider Jr., K. L. Lear, K. Geib, "Low threshold voltage vertical-cavity lasers fabricated by selective oxidation," Electronics Letters, vol. 30, pp. 2043-2044 (1994).
- [27] K. L. Lear, K. D. Choquette, R. P. Schneider Jr., S. P. Kilcoyne, K. M. Geib, "Selectively oxidized vertical cavity surface emitting lasers with 50% power conversion efficiency," Electronics Letters, vol. 31, pp. 208-209 (1995).
- [28] R. Jager, M. Grabherr, C. Jung, R. Michalzik, G. Reiner, B. Weigl, K. J. Ebeling, "57% wallplug efficiency oxide-confined 850 nm VCSELs," Electronics Letters, vol. 33, pp. 330-331 (1997).
- [29] K. L. Lear, A. Mar, K. D. Choquette, S. P. Kilcoyne, R. P. Schneider Jr., K. M. Geib, "High-frequency modulation of oxide-confined vertical cavity surface emitting lasers," Electronics Letters, vol. 32, pp. 457-458 (1996).
- [30] U. Fiedler, G. Reiner, P. Schinitzer, K. J. Ebeling, "Top-surface-emitting laser diodes for 10 Gbit/s data transmission," IEEE Photonics Technology Letters, vol. 8, pp. 746-748 (1996).
- [31] J. L. Jewell, Picolight Inc., Boulder, CO, USA; private communication.
- [32] S. Matsuo, T. Nakahara, K. Tateno, T. Kurokawa, "Novel technology for hybrid integration of photonic and electronic circuits," IEEE Photonics Technology Letters, vol. 8, pp. 214-216 (1996).
- [33] D. V. Plant, J. A. Trezza, M. B. Venditti, E. Laprise, J. Faucher, K. Razavi, M. Châteauneuf, A. G. Kirk, and W. Luo, "A 256 Channel Bi-Directional Optical Interconnect Using VCSEL's and Photodiodes on CMOS," in Optics in Computing 2000, R. A. Lessard and T. Galstian eds., SPIE 4089,1046-1054 (2000).
- [34] O. Sjolund, D. A. Louderback, E. R. Hegblom, J. Ko, and L. A. Coldren, "Monolithic integration of substrate input output resonant photodetectors and vertical-cavity lasers," IEEE J. of Quantum Electronics, vol. 35, pp. 1015-1023 (1999).
- [35] D. M. Cutrer and K. Y. Lau, "Ultralow power optical interconnect with zero-biased, ultralow threshold laser - how low a threshold is low enough?," IEEE Photonics Technology Letters, vol. 7, pp. 4-6 (1995).

- [36] P. Schnitzer, M. Grabherr, R. Jager, C. Jung, K. F. Ebeling, "Bias-free 2.5 Gbit/s data transmission using ployimide passivated GaAs VCSELs," Electronics Letters, vol. 19, pp. 573-575 (1998).
- [37] Y. C. Lee, S. E. Swirhun, W. S. Fu, T. A. Keyser, J. L. Jewell, W. E. Quinn, "Thermal management of VCSEL-based optoelectronic modules," IEEE Trans. on Components, Packaging, and Manufacturing Technology - Part B, vol. 19, pp. 540-547 (1996).
- [38] R. Pu, C. W. Wilmsen, K. M. Geic, K. D. Choquette, "Thermal resistance of VCSELs bonded to integrated circuits," IEEE Photonics Technology Letters, vol. 11, pp. 1554-1556 (1999).
- [39] L. A. Coldren and S. W. Corzine, Diode Lasers and Photonic Integrated Circuits, Chapter 5, Wiley, New York (1995).
- [40] M. S. Torre and H. F. Ranea-Sandoval, "Modulation response of multiple transverse modes in vetical-cavity surface-emitting lasers," IEEE J. of Quantum Electronics, vol. 36, pp. 112-117 (2000).
- [41] J. Martin-Regalado, F. Prati, M. San Miguel, and N. B. Abraham, "Polarization properties of vertical-cavity surface-emitting lasers," IEEE J. Quantum Electronics, vol. 33, pp. 765-783 (1997).
- [42] T. Yoshikawa, H. Kosaka, K. Kurihara, M. Kajita, Y. Sugimoto, and K. Kasahara, "Complete polarisation control of 8 × 8 VCSEL matrix arrays," Applied Physics Letters, vol. 66, pp. 908-910 (1996).
- [43] D. A. B. Miller, D. S. Chemla and S. Schmitt-Rink, "Relation between electroabsorption in bulk semiconductors and in quantum wells: the quantum-confined Franz-Keldysh effect," Physics Review, vol. B33, pp. 6976-6982 (1986).
- [44] D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood and C. A. Burrus, "Electric field dependence of optical absorption near the bandgap of quantum well structures," Phys. Rev., vol. B32, pp. 1043-1060 (1985).
- [45] A. V. Krishnamoorthy and K. W. Goossen, "Progress in optoelectronic-VLSI smart pixel technology based on GaAs/AlGaAs MQW modulators," International J. of Optoelectronics, vol. 11, pp. 181-198 (1997).

- [46] D. A. B. Miller, "Quantum well optoelectronic switching devices," International J. of High-Speed Electronics, vol. 1, pp. 19-46 (1990).
- [47] K. W. Goossen, M. B. Santos, J. E. Cunningham, and W. Y. Jan, "Independence of absorption coefficient-linewidth product to material system for multiple quantum wells with excitons from 850 nm to 1064 nm," IEEE Photonics Technology Letters, vol. 5, pp. 1392-1394.
- [48] K. W. Goossen, J. A. Walker, L. A. D'Asaro, S. P. Hui, B. Tseng, R. Leigenbuth, D. Kossives, D. D. Bacon, D. Dahringer, L. M. F. Chirovsky, A. L. Lentine, and D. A. B. Miller, "GaAs MQW modulators integrated with silicon CMOS," IEEE Photonics Technology Letters, vol. 7, pp. 360-362 (1995).
- [49] F. B. McCormick, T. J. Cloonan, F. A. P. Tooley, A. L. Lentine, J. M. Sasian, J. L. Brubaker, R. L. Morrison, S. L. Walker, R. J. Crisci, R. A. Novotny, S. J. Hinterlong, H. S. Hinton, and E. Kerbis, "Six-stage digital free-space optical switching network using symetrical self-electro-optic-effect devices," Applied Optics, vol. 32, pp. 5153-5171 (1993).
- [50] A. L. Lentine, K. W. Goossen, J. A. Walker, J. E. Cunningham, W. Y. Jan, T. K. Woodward, A. V. Krishnamoorthy, B. J. Tseng, S. P. Hui, R. E. Leibenguth, L. M. F. Chirovsky, R. A. Novotny, D. B. Buchholz, R. L. Morrison, "Optoelectronic VLSI switching chip with > 1Tbit/s potential optical I/O bandwidth," Electronics Letters, vol. 33, pp. 894-895 (1997).
- [51] A. C. Walker, T. Y. Yang, J. Gourlay, J. A. B. Dines, M. G. Forbes, S. M. Prince, D. A. Baillie, D. T. Neilson, R. Williams, L. C. Wilkinson, G. R. Smith, M. P. Y. Dezmulliez, G. S. Buller, M. R. Taghizadeh, A. Waddie, I. Underwood, C. R. Stanley, F. Pottier, B. Vogele, and W. Sibbett, "Optoelectronic systems based on InGaAs-complementary-metal-oxide-semiconductor smart-pixel arrays and free-space optical interconnects," Applied Optics, vol. 37, pp. 2822-2830 (1998).
- [52] A.G. Kirk, D. V. Plant, T. H. Szymanski, Z. G. Vranesic, J. A. Trezza, F. A. P. Tooley, D. R. Rolston, M. H. Ayliffe, F. Lacroix, D. Kabal, B. Robertson, E. Bernier, D. F.-Brosseau, F. S. J. Michael and E. L. Chuah, "A modulator-based multistage

free-space optical interconnection system," in Optics in Computing 2000, R. A. Lessard and T. Galstian eds., SPIE 4089, 449-459 (2000).

- [53] A. V. Krishnamoorthy, T. K. Woodward, K. W. Goossen, J. A. Walker, S. P. Hui, B. Tseng, J. E. Cunningham, W. Y. Jan, F. E. Kiamilev, and D. A. B. Miller, "Dualfunction detector-modulator smart-pixel module," Applied Optics, vol. 36, pp. 4866-4870 (1997).
- [54] K. W. Goossen, "Optoelectronic-VLSI," 1998 Electronic Components and Technology Conference, pp. 771-777 (1998).
- [55] J. A. Trezza, J. S. Powell, C. G. Garvin, K. Chang, R. D. Stack, "Creation and application of a very-large-format high-fill-factor GaAs-on-CMOS binary and gray-scale modulator and emiter arrays," in Optics in Computing 1998, P. Chavel, D. A. B. Miller and H. Thienpont, Editors, Proceedings SPIE vol. 3490, pp. 78-81 (1998).
- [56] K. Wakita, I. Kotaka, O. Mitomi, H. Asai, Y. Kawamura, and M Naganuma "Highspeed InGaAs/InAlAs multiple quantum well optical modulators with bandwidths in excess of 40 GHz at 1.55 μm," presented at CLEO '90, paper CTuC6, Anaheim, CA (1990).
- [57] D. F-Brosseau, F. Lacroix, M. H. Ayliffe, E. Bernier, B. Robertson, F. A. P. Tooley, D. V. Plant, A. G. Kirk, "Design, implementation, and characterization of a kinematically aligned, cascaded spot-array generator for a modulator-based free-space optical interconnect," Applied Optics, vol. 39, pp. 733-745 (2000).
- [58] M. Châteauneuf, F. Thomas-Dupuis, A.G. Kirk, "Design, implementation and characterization of a folded spot array generator for a modulator-based free-space optical interconnect," in Optics in Computing 2000, R. A. Lessard, T. Galstian, Proc. SPIE 4089, pp. 263-271 (2000).
- [59] G. A. Keeler, B. E. Nelson, D. Agarwal, and D. A. B. Miller, "Skew and jitter removal using short optical pulses for optical interconnection," IEEE Photonics Technology Letters, vol. 12, pp. 714-716 (2000).
- [60] L. Boivin, M. C. Nuss, J. Shah, D. A. B. Miller and H. A. Haus, "Receiver sensitivity improvement by impulsive coding," IEEE Photonics Technology Letters, vol. 9, pp. 684-686 (1997).

- [61] A. V. Krishnamoorthy and D. A. B. Miller, "Scaling optoelectronic-VLSI circuits inot the 21st century," IEEE J. of Selected Topics in Quantum Electronics, vol. 2, pp. 55-76 (1996).
- [62] A. V. Krishnamoorthy and K. W. Goossen, "Progress in optoelectronic-VLSI smart pixel technology based on GaAs/AlGaAs MQW modulators," Internation J. of Optoelectronics, vol. 11, pp. 181-198 (1997).
- [63] T. K. Woodward, A. V. Krishnamoorthy, A. L. Lentine, and L. M. F. Chirovsky, "Optical receivers for optoelectronic-VLSI," IEEE J. of Selected Topics in Quantum Electronics, vol. 2, pp. 106-116 (1996).
- [64] M. Whitehead, A. W. Rivers, G. Parry, J. S. Roberts, and C. Button, "Low voltage multiple quantum well modulator with on:off ratio >100:1," Electronics Letters, vol. 25, pp. 984-985 (1989).
- [65] T. L. Worchesky, K. J. Ritter, R. Martin, and B. Lane, "Large arrays of spatial light modulators hybridized to silicon IC's," Applied Optics, vol. 35, pp. 1180-1186 (1996).
- [66] Properties of Gallium Arsenide, EMIS Data Reviews Series, 2nd Edition London, UK: INSPEC, no. 2, 1990.
- [67] M. B. Venditti, "Temperature dependence of QCSE modulator and detector efficiency for free-space optical interconnect applications," M. Eng. Thesis, McGill University, Montréal, Canada, September 1999.
- [68] K. W. Goossen, J. E. Cunningham, and W. Y. Jan, "Stacked-diode electroabsorption modulator," IEEE Photonics Technology Letters, vol. 6, pp. 936-938 (1994).
- [69] T. K. Woodward, A. V. Krishnamoorthy, K. W. Goossen, J. A. Walker, B. Tseng, J. Lothian, S. Hui, and R. Leibenguth, "Modulator-driver circuits for optoelectronic VLSI," IEEE Photonics Technology Letters, vol. 9, pp. 839-841 (1997).
- [70] D. T. Neilson, "Optimization and tolerance analysis of QCSE modulators and detectors," IEEE J. of Quantum Electronics, vol. 33, pp. 1094-1103 (1997).
- [71] D. T. Neilson, D. J. Goodwill, L. C. Wilkinson, F. A. P. Tooley, A. C. Walker, C. R. Stanley, M. McElhinney, and F. Pottier, "InGaAs transceivers for smart pixels," presented at Optical Computing '95, Salt Lake City, UT, paper OTuC3.

- [72] M. K. Hibbs-Brenner, Y. Liu, R. Morgan, and J. Lehman, "VCSEL/MSM detector smart pixel arrays," in Broadband Optical Networks and Technologies, IEEE/LEOS Summer Topical Meetings, pp. 3-4 (1998).
- [73] C. Moglestue, J. Rosenzweig, J. Kuhl, M. Klingerstein, M. Lambsdorff, A. Axmann, J. Schneider, and A. Hulsmann, "Picosecond pulse response characteristics of GaAs metal-semiconductor-metal photodetectors," J. of Applied Physics, vol. 70, pp. 2435-2448 (1991).
- [74] R. B. Darling, H. J. Youn, and K. J. Kuhn, "Use of active loads with MSM photodetectors in digital GaAs MESFET photoreceivers," J. of Lightwave Technology, vol. 10, pp. 1597-1605 (1992).
- [75] J. Choi, B. J. Sheu, O. T.-C. Chen, "A monolithic GaAs receiver for optical interconnect systems," IEEE J. of Solid-State Circuits, vol. 29, pp. 328-331 (1994).
- [76] D. L. Rogers, "Integrated optical receivers using MSM detectors," J. of Lightwave Technology, vol. 9, pp. 1635-1638 (1991).
- [77] E. M. Hayes, R. D. Snyder, R. Jurrat, S. A. Feld, C. W. Wilmsen, K. D. Choquette, K. M. Geib, H. Q. Hou, "8 × 8 array of smart pixels fabricated through the Vitesse foundry integrating MESFET, MSM, and VCSEL elements," IEEE/LEOS 1996 Summer Topical Meetings (Cat. No.96TH8164) New York, NY, USA, pp.103-104 (1996).
- [78] M. Gökkavas, G. Ulu, and M. S. Ünlü, "Resonant cavity enhanced photodetectors with a flat spectral response," in Ultrafast Electronics and Optoelectronics Technical Digest, pp. 208-210 (1999).
- [79] F. A. P. Tooley, "Challenges in optically interconnecting electronics," IEEE J. Selected Topics in Quantum Electronics, vol. 2, pp. 3-13 (1996).
- [80] A. V. Krishnamoorthy, T. K. Wodward, K. W. Goossen, J. A. Walker, A. L. Lentine, L. M. F. Chirovsky, S. P. Hui, B. Tseng, R. Leigenbuth, J. E. Cunningham, and W. Y. Jan, "Operation of a single-ended 550 Mb/s, 41 fJ, hybrid CMOS/MQW receivertransmitter," Electronics Letters, vol. 32, pp. 764-765 (1996).
- [81] T. K. Woodward, A. V. Krishnamoorthy, A. L. Lentine, K. W. Goossen, J. A. Walker, J. E. Cunningham, W. Y. Jan, L. A. D'Asaro, L. M. F. Chirovsky, S. P. Hui,

B. Tseng, D. Kossives, D. Dahringer, and R. E. Leigenbuth, "1 Gb/s two-beam transimpedance smart-pixel optical receivers made from hybrid GaAs MQW modulators bonded to 0.8 micron silicon CMOS," IEEE Photonics Technology Letters, vol. 8, pp. 422-424 (1996).

- [82] C. Toumazou, F. J. Lidgey, and D. G. Haigh, eds., Analogue IC Design: the Current-Mode Approach, London: Peter Peregrinus (1990).
- [83] A. Z. Shang and F. A. P. Tooley, "Current-mode smart-pixel receivers," paper ThB4 Summer Topicals on Smart Pixels, August 8th, Keystone, CO, USA (1996).
- [84] T. Vanisri and C. Toumazou, "Integrated high-frequency low-noise current-mode optical transimpedance preamplifiers: theory and practice," IEEE J. of Solid-State Circuits, vol. 30, pp. 677-685 (1995).
- [85] J. M. Sasian, R. A. Novotny, M. G. Beckman, S. L. Walker, M. J. Wojcik, and S. J. Hinterlong, "Fabrication of fiber bundle arrays for free-space photonic switching systems," Optical Engineering, vol. 33, pp. 2979-2985 (1994).
- [86] H. S. Hinton, T. J. Cloonan, F. A. P. Tooley, F. B. McCormick and A. L. Lentine, "Free-space digital optical interconnections," *Proc. IEEE*, vol. 82, pp. 1632-1649, 1995.
- [87] A. C. Walker, M. R. Taghizadeh, F. A. P. Tooley, G. Smith, I. R. Redmond, D. J. Knight, B. Robertson, and C. P. Barett, "Optical and optomechanical design of a matrix-matrix crossbar interconnect," in European Optical Society, 8th Workshop on Optics in Computing, EOS Topical Meetings Digest, vol. 1, pp. 7-10 (1992).
- [88] T. Yamamoto, M. Yamaguchi, K. Hirabayashi, S. Matsuo, C. Amano, H. Iwamura, Y. Kohama, T. Kurokawa, and K. Koyabu, "High-density digital free-space photonic switches using microbeam optical interconnections," IEEE Photonics Technology Letters, vol. 8, pp. 358-360 (1996).
- [89] C. M. Miller, "Fiber optic array splicing with etched silicon chips," Bell Systems Technical Journal, vol. 57, pp. 75-90 (1978).
- [90] C. M. Schroeder, "Accurate silicon spacer chips for an optical fiber cable connector," Bell Systems Technical Journal, vol. 57, pp. 91-97 (1978).

- [101]K. Tatah, D. Filkins, B. Greiner, and M. Robinson, "Performance measurements of fiber image guides and fiber bundles on optical interconnect applications," in the Digest of the Topical Meeting on Optics in Computing (Optical Society of America, Washington DC, 1999), pp. 112-114.
- [102]S. Kawai, Y. Li, and T. Wang, "Skew-free optical interconnections using VCSELbased fiber image guides," in the Sixth Microoptics Conference and the Fourteenth Topical Meeting on Gradient-Index Optical Systems, MOC/GRIN Technical Digest (The Japan Society of Applied Physics, Tokyo, 1997), pp. 78-80.
- [103]D. M. Chiarulli, S. P. Levitan and M. Robinson, "Optoelectronic multi-chip modules based on imaging fiber bundle structures," in Optics in Computing 2000, R. A. Lessard and T. Galstian eds., SPIE 4089, 80-85 (2000).
- [104]D. J. Reiley and J. M. Sasian, "Optical design of a free-space photonic switching system," Applied Optics, vol. 36, pp. 4497-4504 (1997).
- [105]D. T. Neilson and C. P. Barrett, "Performance trade-offs for conventional lenses for free-space digital optics," Applied Optics, vol. 35, pp. 124-1248 (1996).
- [106]F. B. McCormick, F. A. P. Tooley, T. J. Cloonan, J. M. Sasian, and H. S. Hinton, "Optical interconnects using microlens arrays," Optical Quantum Electronics, vol. 24, pp. S465-S477 (1992).
- [107]N. C. Craft and A. Y. Feldblum, "Optical interconnects based on arrays of surfaceemitting lasers and lenslets," Applied Optics, vol. 31, pp. 1735-1739 (1992).
- [108]T. Yamamoto, M. Yamaguchi, K. Hirabayashi, S. Matsuo, C. Amano, H. Iwamura, T. Kurokawa, and K. Koyabu, "High-density digital free-space photonic switches using micro-beam optical interconnections," IEEE Photonics Technology Letters, vol. 8, pp. 358-360 (1996).
- [109]A. Lohmann, "Image formation of dilute arrays for optical information processing," Optical Communications, vol. 86, pp. 365-370 (1991).
- [110]W. Dascher and S. H. Lee, "Reproducing micro-optics in quantities by semiconductor fabrication techniques," in Technical Digest ICO Meeting, Optics in Computing '96, Sendai, Japan, paper OTuB3.

- [111]J. Jahns, F. Sauer, B. Tell, K. F. Brown-Goebeler, A. Y. Feldblum, C. R. Nijander, and W. P. Townsend, "Parallel optical interconnections using surface-emitting microlasers and a hybrid imaging system," Optics Communications, vol. 109, pp. 328-337 (1994).
- [112]G. C. Boisset, M. H. Ayliffe, B. Robertson, R. Iyer, Y. S. Liu, D. V. Plant, D. J. Goodwill, D. Kabal, D. Pavlasek, "Optomechanics for a four-stage hybrid self-electro-optic-device-based free-space optical backplane," Applied Optics, vol. 36, no. 29, pp.7341-7358 (1997).
- [113]S. Araki, M. Kajita, K. Kasahara, K. Kobota, K. Kurihara, I Redmond, E. Schenfeld, and T. Suzaki, "Experimental free-space optical network for massively parallel computers," Applied Optics, vol. 35, pp. 1269-1281 (1996).
- [114]D. R. Rolston, B. Robertson, H. S. Hinton, and D. V. Plant, "Analysis of a microchannel interconnect based on the clustering of smart pixel windows," Applied Optics, vol. 35, pp. 1220-1233 (1996).
- [115]B. R. Robertson, "Design of an optical interconnect for photonic backplane applications," Applied Optics, vol. 37, pp. 2974-2984 (1998).
- [116]J. Jahns and A. Huang, "Planar integration of free-space optical components," Applied Optics. vol. 28, pp. 1602-1605 (1989).
- [117]S. Sinzinger and J. Jahns, "Integrated micro-optical imaging system with a high interconnection capacity fabricated in planar optics," Applied Optics, vol. 36, pp. 4729-4735 (1997).
- [118]D. Frey, W. Erhard, M. Gruber, J. Jahns, H. Bartlet, G. Grimm, L. Hoppe, and S. Sinzinger, "Optical interconnects for neural and reconfigurable VLSI architectures," in Proceedings IEEE, vol. 88, pp. 838-848 (2000).
- [119]L. A. Buckman, B. E. Lemoff, D. W. Dolfi, "A low-cost compact multimode/singlemode transceiver modules for 10Gb/s applications," in Proceedings of the Optical Fiber Communications 2000 Conference (Baltimore, MA, USA), paper WE2-2.
- [120]S. J. Walker and J. Jahns, "Optical clock distribution using integrated free-space optics," Optics Communications, vol. 90, pp. 359-371 (1992).

Chapter 3: Design and testing of a free-space photonic backplane demonstrator system

3.1 Introduction

This chapter provides an overview of the design, implementation and performance of a free-space photonic backplane demonstrator system. The photonic backplane project was initiated in 1996; it is the result of the work of a large group of individuals from academia and industry, all of whom are gratefully acknowledged at the end of this chapter.

The chapter is organized as follows. First, the application that motivated the system demonstrator is presented (section 3.2). Next, the interconnect topology (section 3.3), the system specifications (section 3.4), and its physical layout (section 3.5) are described. An important objective is to underscore the large set of critical engineering choices and compromises required in turning a high-level topological description into a hardware implementation. The next sections present an overview of the OE-VLSI chip (section 3.6) and the design of the optical interconnect (section 3.7), and examines packaging and alignment issues (Section 3.8). The experimental performance of the demonstrator system is presented in section 3.9. The last section provides a list of the author's contributions.

3.2 Target application: a multiprocessor computing system

The photonic backplane was developed to replace a standard electrical backplane in a multiprocessor computing system based on a non-uniform memory access (NUMA) architecture [1]. In the NUMA architecture, processors have access to both local and non-local memories, so all processors can access all of the memory. In order to ensure that one processor is not operating on memory that has been modified by another processor, memory coherency is necessary. Latency of memory is thus of vital importance and a significant portion of that latency is due to the backplane interconnections. A high-bandwidth backplane is required to accommodate simultaneous transfer of data between multiple processing nodes and non-local memories. This results in a requirement for low-latency and high-

bandwidth interconnections between a large number of nodes, a perfect application for two-dimensional free-space optical interconnect (2D-FSOI) technologies.



Figure 3.1. Representation of an optically interconnected multiprocessor system.

Figure 3.1 is a high-level representation of the multiprocessor system, showing *N* processors being interconnected via a photonic backplane. Transfer of data across the photonic backplane is performed as follows. Electrical data originating from a processor/ memory is communicated to an OE-VLSI chip where it is converted into modulated optical signals. The optical signals propagate in free-space from one OE-VLSI chip to the next. At each stage, the optical signals are either (i) regenerated on-chip towards the next stage or (ii) converted back into electrical signals and directed off-chip towards the local processor/memory.

The transfer of data between a processor chip and an OE-VLSI chip is performed over electrical lines. As a result, the amount of data that can be transferred on and off the backplane is limited by the off-chip electrical I/O bandwidth (off-chip data rate \times number of I/O pads). On the other hand, the amount of data that can be transferred across the backplane is determined by the OE-VLSI chip optical I/O bandwidth (on-chip data rate \times number of transceivers in the array) which is typically 1 to 3 orders of magnitude higher. This apparent mismatch between the off-chip electrical I/O bandwidth and the backplane optical I/O bandwidth is successfully managed by properly "balancing" the system computational and communication bandwidths. This balance is achieved by using the aggregate electrical I/O capacity of *N* OE-VLSI chips to fill the high optical I/O bandwidth of each.

This system architecture is commonly referred to as an external data "firehose" [2]; the name "firehose" originates from the fact that the problem of interfacing low-bandwidth electronics to high-throughput parallel optical interconnections is reminiscent to the situation of trying to drink water from a high-pressure firehose.

3.3 Interconnect topology

For memory-coherent multiprocessor systems, the interconnect topology is generally not critical in terms of system performance; what matters most is the interconnect latency between nodes [1]. Also, the maximum rate of data transfer provided by a given topology might be less of a concern than the ability to broadcast data. This suggests that the choice of an interconnect topology should not only be based on its ability to optimize system performance in terms of latency and bandwidth but also on the ease with which it can be turned into a physical implementation using optics. What follows is a list of key requirements and constraints that impacted the choice of the interconnect topology.

- any node must be allowed to communicate to any other node.
- the system must be scalable to a larger number of nodes.
- considering the limited electrical I/O bandwidth between a processor chip and the photonic backplane, the interconnect topology must minimize the I/O bandwidth requirement at the processor/backplane interface.
- to facilitate the optical design, a topology offering a high degree of symmetry is desirable, such that it allows for identical optical modules to be used between nodes.
- the design of a 2D bi-directional optical interconnect between multiple nodes is nontrivial. Thus, a topology where optical data flows unidirectionally is preferred. This suggests that nodes be optically interconnected in a ring configuration.

The above has lead to the choice of the interconnect topology shown in figure 3.2, drawn for the case of a system interconnecting four processors (N=4).

Chapter 3: Design and testing of a free-space photonic backplane demonstrator system



Figure 3.2. Interconnect topology for the case N = 4.

In figure 3.2, direct point-to-point optical interconnections exist only between neighbouring nodes: optical signals propagate from one black dot to next, in a clockwise direction. Each circular line corresponds to an optical channel. The photonic backplane is made of N optical channels and each channel is M bits wide. At every black dot, optical signals are detected and either (i) directed off-chip towards the local processor or (ii) regenerated on-chip and transmitted optically towards the next node. Using this scheme, optical data flows across the backplane by "hopping" from one node to the next, in a clockwise direction. This allows for data originating from any node to reach any other node. This topology can be referred to as a unidirectional multi-hop optical ring.

An important feature of this topology is that each node has a different channel reserved for transmission, such that different nodes transmit on different channels. This is referred to as a "sender-reserve" scheme. In addition, each OE-VLSI chip can receive data on all N channels simultaneously. Received data is fed into an N:1 on-chip multiplexer allowing only one channel to be electrically routed off-chip towards the local processor/ memory while the other channels are blocked for immediate access (blocking of data can

be properly managed by providing on-chip memory buffers). This approach reduces the I/O requirements at the processor/backplane interface to 1 input and 1 output channel port, each M bits wide. This way, the electrical I/O bandwidth does not have to scale (it is independent of N) and is low enough that it can be supported by electrical connections on the periphery of the OE-VLSI chip. On the other hand, the optical I/O bandwidth is expected to scale with N; this can be accomodated using a dense 2D array of surface-normal OE devices on the OE-VLSI chip.

3.4 System specifications and requirements

The design of the photonic backplane was constrained by the following set of specifications and requirements:

- Number of nodes: in order to keep the system implementation to a reasonable scale, it
 was decided that the backplane be designed to interconnect 4 processors but be able to
 scale up to 8 nodes (N = 8).
- Electrical I/O: each node is designed to inject or extract 32 bits from the backplane. Consequently, the electrical I/O requirement at the processor/backplane interface is 32 inputs and 32 outputs (M = 32), resulting in 64 electrical I/O.
- Optical I/O: in order to support up to 8 nodes in a non-blocking manner with each node contributing 64 electrical I/O, the OE-VLSI chip is required to have 256 transmitters and 256 receivers for a total of 512 optical signal I/O.
- Cross-sectional area: in order to have a single OE-VLSI chip at each node, the crosssectional area of the optical interconnections must be kept below 1 cm².
- Interconnect length: the processor/memory chips are mounted on separate printed circuit boards (motherboards) which are inserted parallel to each other in a backplane chassis. The optical interconnect length is constrained by the motherboard separation, which is typically of the order of 50 mm.

• Field serviceability: a technician must be able to remove and insert a motherboard without the need for realignment. Replaceability of the OE-VLSI chip is also desirable. This is necessary because the OE-VLSI chip is the only active element in the photonic backplane; it is prone to failure and is likely to be upgraded with time.

3.5 Physical layout issues

This section examines the design choices and trade-offs involved in turning the topological description of figure 3.2 into a physical implementation. The problem consists of determining the best approach for implementing a scalable, low-latency, high-bandwidth free-space optical interconnect between the motherboard electronics.

A fundamental design consideration concerns the packaging of the OE-VLSI chip. The issue is to determine whether to (i) mount the OE-VLSI chip directly on the motherboard, next to the processing electronics, or (ii) package the OE-VLSI chip in a separate module that is incorporated with the interconnect optics. The following sections examine and compare both options.

3.5.1 OE-VLSI chip mounted directly on the motherboard

Mounting the OE-VLSI chip directly on the motherboard, a short distance away from the processing electronics offers a number of advantages: it facilitates routing of the data paths between the processor and the OE-VLSI chip, it minimizes signal transmission latency and mitigates signal integrity issues, paticularly those related to transmission line effects and electromagnetic interference. Another advantage is that it removes the need for backplane connectors with their limited pin-count and reliability problems.

This approach appears in figure 3.3, where three neighbouring nodes are shown. Because OE-VLSI chips are opaque, they must be mounted face-to-face, on both sides of the motherboards. The main advantage of this layout is the simplified optical design: light travels along a straight path from one motherboard to the next, avoiding the need to redirect the signal beams using mirrors and beamsplitters.

Unfortunately, this layout results in a backplane with very limited throughput. To see this, consider the transfer of data from node i-1 to node i+1. At node i, the incident optical signals are detected and converted into electrical signals, directed off-chip to the OE-VLSI chip located on the other side of the motherboard before being regenerated optically towards node i+1. As a result, the backplane throughput is limited to the off-chip electrical I/O bandwidth of the OE-VLSI chip. The physical layout of figure 3.3 does not take advantage of the high parallelism of optics. Simply put, the problem is that the firehose is pinched in the middle.



Figure 3.3. Proposed backplane physical layout showing data transfer from node i-1 to node i+1. This approach is not scalable.

This difficulty can only be resolved by ensuring that optical signals hopping from one node to the next along the backplane are never sent off-chip, unless of course they have reached their final destination node. This is to say that optical regeneration must be performed on-chip, requiring the integration of both receivers and transmitters on the same OE-VLSI chip. A physical layout that accomplishes this is shown in figure 3.4. In this case, mirrors, prisms and beamsplitters are required to direct light beams from node *i*-1 to node i+1. At stage *i*, incident optical signals are detected, converted into electrical signals; and immediately retransmitted optically towards the next node. There are several advantages associated with performing optical regeneration on-chip. First, signal latency is minimal, the major contribution coming from the propagation delay of the on-chip receiver and transmitter circuits. Second, power dissipation is minimized because no signals have to be driven off-chip through power-hungry CMOS output pad drivers. Most importantly, the backplane throughput is now determined by the optical I/O bandwidth of the OE-VLSI chip. Figure 3.4 represents a scalable implementation of an external data firehose.



Figure 3.4. Proposed backplane physical layout showing data transfer from node i-1 to node i+1. This approach is scalable but imposes severe alignment contraints.

The layout of figure 3.4 is difficult to implement in practice due to the alignment constraints it imposes on the motherboards. This problem is easily understood when we consider the alignment operations required for optical signals to be received and transmitted by the OE-VLSI chip at stage *i*. First, signals arriving from stage *i*-1 must be properly aligned to detectors at stage *i*. Second, signals transmitted from stage *i* must be aligned to detectors at stage *i*+1. Also, in the case of modulator-based systems, an additional alignment operation is required to position the continuous wave (CW) beams onto the modulator array at stage *i*. These operations are not trivial to perform since OE devices are small in size (10's of microns in diameter) and 2D device arrays need to be aligned in all six degrees of freedom (DOF): lateral (*x*, *y*), longitudinal (*z*), rotational (θ_z) and tilt (θ_x , θ_y). This problem is further exacerbated by the requirement that motherboards be replaceable: a person servicing the system must be able to remove and insert a motherboard without disturbing the alignment of the optics.

For all of the above reasons, attempts at mounting the OE-VLSI chip directly on the motherboard have been limited to:

 board-to-board demonstrators implementing a low number (<12) of optical channels with a generous misalignment budget [3]-[6]. Unfortunately, this approach fails at demonstrating the high parallelism of optics which is the principal objective here.

Chapter 3: Design and testing of a free-space photonic backplane demonstrator system

 systems employing some type of alignment compensation techniques such as liquidcrystal steering devices or array redundancy [7][8]. For large-scale 2D-FSOI systems interconnecting multiple nodes, a large number of individually-controlled beam-steering devices are needed and the corresponding increase in cost, volume and complexity makes it unclear whether adaptive optical techniques represent a viable solution. In this context, the idea of using redundant OE devices represents an interesting strategy (see chapter 6). This concept essentially amounts to trading off channel density for misalignment tolerances.

3.5.2 OE-VLSI chip packaged in a separate module

The alternative option consists of packaging the chip in a separate module and attaching it to the motherboard via a flexible interconnecting medium. This approach has been used in several system demontrations [9]-[11] and is shown in figure 3.5. The flexible interconnection provides the required mechanical isolation of the module from the motherboard: it avoids the alignment constraints placed on the motherboards and allows for PCBs to be replaced without disturbing the alignment of the optical system. This removes the need for alignment compensation techniques. Furthermore, OE-VLSI chips can now be oriented at 90° relative to the motherboards which reduces the complexity of the optical design. Figure 3.5 represents a scalable implementation of an external data firehose.



Figure 3.5. Physical layout used for the photonic backplane demonstrator. This approach is scalable and removes the alignment contraints on the motherboards.

Chapter 3: Design and testing of a free-space photonic backplane demonstrator system
The layout of figure 3.5 does not come without difficulties. The increased distance between the chip and the motherboard electronics can potentially aggravate signal integrity. The limited pin-count of the connector puts an upper bound on the amount of data that can come on and off the backplane, although this does not constitute a problem as long as the connector can support the required electrical I/O bandwidth between the processor and the backplane. Furthermore, the size of the chip module is now constrained by the motherboard pitch and this may complicate routing of data lines to the OE-VLSI chip.

The above discussion serves to demonstrate how the location of the OE-VLSI chip relative to the motherboard has a direct impact on the packaging and alignment of the OE-VLSI chip. Although packaging the OE-VLSI chip in a separate module is certainly not optimal, it does lead to a practical approach to the construction of a scalable photonic backplane and it is the scheme adopted in the system demonstrator described in the following sections.

3.6 Optoelectronic-VLSI chip

This section provides an overview of the OE-VLSI chip used in the photonic backplane demonstrator.

3.6.1 MQW modulator technology

At the time the backplane demonstrator project was initiated (1996), a VCSEL-based OE-VLSI technology (VCSELs and detectors on CMOS) was unavailable. MQW modulators were thus selected as the OE technology. This choice satisfied system requirements in terms of (i) large array size, (ii) high-speed operation, and (iii) the integration of both optical input and output devices on the same chip in a single flip-chip step.

The design of the MQW modulator devices was accomplished by Dr. Frank Tooley (now with TeraHertz Photonics, Edinburgh, UK) with the assistance of Dr. David Neilson (Bell Labs - Lucent Technologies, Holmdel, NJ, USA) and Dr. Anthony SpringThorpe (Nortel Networks, Kanata, Canada). A simplified version of the layer structure is shown in figure 3.6. The design was optimized to maximize the difference in reflectivity $\Delta R = R_{high}$ - R_{low} under a voltage swing of $V_{mod} = 5$ V at $\lambda = 852$ nm and T = 40 °C. The top 500Å thick GaAs layer (cap layer) is *p*-doped (5e18 Be); it provides an ohmic *p*-contact for the later deposition of a Ti(400Å)/Pt(400Å)/Au(1000Å) metal contact which is also used as the reflecting mirror. The *p* region is made of 1000Å AlGaAs (1e18 Be) while the *n* region is made of 1000Å AlGaAs (1e18 Si). Because the same structure is used for both modulators and detectors, the optimal thickness of the MQW region is a compromise, chosen to balance between responsivity, insertion losses and voltage swing [12]. The MQW region is made of 60 quantum wells made with alternating intrinsic layers of GaAs (90 Å) and AlGaAs (35 Å), for a total thickness of $L_{MQW} = 750$ nm. Finally, etch stop layers are grown between the *n* region and the GaAs substrate; these are required to perform the substrate removal step (see figure 2.3).



Figure 3.6. MQW modulator layer structure.

The layer structure was grown using molecular beam epitaxy (MBE) by Dr. SpringThorpe. Devices were subsequently fabricated using a wet-etch process developed by Dr. Edwis Richard (formerly at École Polytechnique, Montréal, Canada). A scanning electron microscope (SEM) photograph of a fabricated modulator device (prior to flip-chip) is shown in figure 3.7.

An important design objective was to allow for the measurement of R_{high} and R_{low} on fabricated devices prior to flip-chip; this allowed for GaAs chips exhibiting poor reflectivities to be discarded prior to hybridization. Unfortunately, performing a *direct* reflectivity measurement on the device of figure 3.7 is not possible because the GaAs substrate is absorptive at 852 nm. Instead, an *indirect* reflectivity measurement was performed on a dedicated test device, shown in figure 3.8. The test device was located on the same chip as the other modulators, and because both types of devices were fabricated at the same time, using the same process, it was reasonable to assume that their reflectivity spectrum would be almost identical. The test device differs from the modulator of figure 3.7 in that the *p*-contact covers only a small portion of the MQW region, allowing for a light beam to be focused on the MQW region (shown in figure 3.8). The physical dimensions of the test device were also made larger in order to facilitate the alignment of the input light beam.



Figure 3.7. SEM photograph of fabricated MQW modulator prior to flip-chip.



Figure 3.8. Photograph of a test device showing input beam incident on MQW region..

Using this test device, a direct measurement of the diode responsivity was performed. This measurement was carried out using an experimental setup constructed by Mr. Michael Venditti. The setup uses a monochromator to produce the light signal over the wavelength range of interest and a lock-in amplifier to measure the generated photocurrent as a function of incident power, applied voltage and temperature. A detailed description of the experimental setup and its operation is found in [13]. The responsivity data was used to calculate the absorption coefficient (α_{MQW}) of the MQW region. This was done over the range $\lambda = 840 - 860$ nm for voltages 0 - 8 V at T = 40 °C and the results were plotted previously in figure 2.2 (section 2.3.3). The final step consists of calculating the modulator reflectivity (R_{MOW}) from the absorption data using:

$$R_{MQW}(\lambda, V, T) = R_{mirror} \exp\left[-2\alpha_{MQW}(\lambda, V, T)L_{MQW}\right] \exp\left[-2\alpha_{GaAs}L_{cap}\right]$$
(3.1)

where the first exponential term accounts for the double-pass absorption in the MQW region while the second exponential term accounts for the double-pass absorption in the *p*-doped GaAs cap layer. The mirror reflectivity (i.e. at the interface between the GaAs cap layer and the Ti/Pt/Au *p*-contact) was estimated to be $R_{mirror} \sim 90\%$ (the same number was used in [12]). The results of equation (3.1) are plotted in figure 3.9; they predict that for $\lambda = 852$ nm and T = 40 °C, $R_{high} = 82\%$ (at 6V) and $R_{low} = 37\%$ (at 1V), corresponding to a change in reflectivity of $\Delta R = 45\%$ for a 5V voltage swing.



Figure 3.9. Reflectivity spectrum of a MQW modulator at $T = 40^{\circ}$ C. (Calculations based on responsivity measurement data; two data points per nm).

3.6.2 CMOS technology

Two generations of CMOS technology were used. In the first generation (1997-1998), the chip technology was a *p*-substrate, 5V, three-metal layer, 0.8 μ m BiCMOS process (although the bipolar capability was not used). The BiCMOS fabrication run was donated by the Canadian Microelectronics Corporation (CMC) located at Queen's University (Kingston, Ontario). The on-chip clock rate for this technology was 300 MHz.

The hybridization of modulator arrays with BiCMOS chips was performed by Dr. John Trezza's group (formely with Sanders Corp., a division of Lockheed Martin, Nashua, NH, USA; now with Xanoptix, Nashua, NH, USA). The flip-chip process was the same as the one described in figure 2.3 (section 2.3.3), except that indium was used for the bump metallurgy. Indium bumps were deposited on both GaAs and CMOS chips. The bumps were not reflowed; instead, the chips were brought in contact under sufficient temperature and pressure conditions for bonding to take place. Sanders also performed the steps of substrate removal and the application of an anti-reflection coating.

The experimental testing of the first-generation OE-VLSI chip revealed three major design flaws (a complete discussion can be found in [14]). First, an error was found in the layout of the BiCMOS chip. The consequence of this error was that data received optically could not be transmitted electrically off-chip. The second flaw concerned the poor sensitivity of the optical receiver (a clocked sense amplifier configuration was used [15]). Although simulations predicted high-speed operation with only 8 µW of optical power, experimental testing of the sense amplifier (performed by David Rolston) proved that the receivers were erratic and required over 130 μ W of optical power to switch, which exceeding by far the system's optical power budget (31 μ W, in the best case, see section 3.7.3). The origin of the problem could not be isolated, mostly because of the large overhead in digital BiCMOS logic surrounding the receiver. The third flaw concerned the poor reflectivity of the flip-chip modulator devices. Reflectivity measurements were performed on flip-chip devices using a tunable Ti:Sapphire laser over the range $\lambda = 840 - 860$ nm; the best devices exhibited $R_{high} = 15\%$ and $R_{low} = 7\%$ with a 5-V swing under optimal biasing and temperature conditions which is significantly lower than the predictions of figure 3.9. Two hypothesis were put forward to explain such low reflectivity: (i) the temperature and These disappointing results indicated that the annealing step was not the main reason for the mirror reflectivity problem. Further investigation was performed on different mirror samples; a description of these measurements along with the conclusion of our findings are presented in section 3.9.2.

The next sections present the layout of the OE devices and provide a basic-level description of the CMOS circuitry. In what follows and for the remaining of this thesis, all of the discussion pertains to the second-generation OE-VLSI chip.

3.6.3 OE-VLSI chip layout

A photograph of the CMOS chip is shown in figure 3.10.



Chapter 3: Design and testing of a free-space photonic backplane demonstrator system

The CMOS chip dimensions are $9 \times 9 \text{ mm}^2$. An important feature of the layout concerns the placement of the OE devices on the chip. Unlike conventional layout approaches, OE devices are not distributed uniformly across the surface of the chip; instead, devices are grouped into small clusters. As will be demonstrated in chapter 7, the use of a clustering configuration increases lateral misalignment tolerances and improves system scalability in terms of array size and interconnection length.

In our design, each cluster is composed of a 4×4 array of OE devices. Within a cluster, the pitch between OE devices is 90 µm in both lateral directions. Clusters are arranged to form an 8×8 array on the chip; the pitch between clusters is 800 µm in both lateral directions. This results in a total of 1024 OE devices on the chip, occupying an area of 6.4×6.4 mm² (a density of 2500 devices/cm²). Half of the OE devices are used as modulators, the other half as detectors. All devices within a cluster are of the same type (i.e. either all modulators or all detectors). Furthermore, clusters of the same type are organized in columns, with four columns of modulator clusters being interlaced with four columns of detector clusters (figure 3.10). This arrangement ensures that light beams reflected from modulators at one stage are imaged on detectors at the next stage. Figure 3.11(a) is a photograph of a region of the CMOS chip corresponding to a cluster of modulators. The location of the flip-chip points are determined by the 10×15 µm² openings in the top passivation layer of the CMOS process. A photograph of the same cluster, following the steps of flip-chip, substrate removal and AR coating is shown in figure 3.11(b).



Figure 3.11. Photograph of a modulator cluster (a) before and (b) after hybridization.

The active area of modulators is determined by the size of the *p*-contact mirror; it is circular in shape with a diameter of 52.5 μ m. Detectors are square with dimensions 66 × 66 μ m. The size of the modulator is chosen to be a compromise between device capacitance (determining switching speed) and misalignment tolerance. Detectors were made larger than modulators to further relax misalignment tolerances; their maximum size is limited by the fabrication design rules which specify the minimum separation between adjacent devices.

The design of the OE-VLSI chip is based on dual-rail encoding, where one data signal is encoded on two spatially separated optical beams [15]. Dual-rail encoding is required because OE-VLSI receivers are DC-coupled (2D arrays of AC-coupled receivers would be prohibitively large), modulation contrast is low (about 2:1) and the DC component of a single-ended detector photocurrent is difficult to control (it depends on the input optical power and losses along the optical path due to reflections, clipping, polarization and misalignment). The strategy behind dual-rail encoding is to use a pair of neighbouring optical beams with the data on one beam being the complement of the other; these two beams experience approximately the same losses as they propagate through the optical system and the DC component at the receiver front-end can be nulled by taking the difference between the two photocurrents. Dual-rail encoding ensures robust receiver operation under low modulation contrast but requires a pair of modulators and 512 detectors, 256 dual-rail transmitters and 256 dual-rail receivers, a capacity that matches the optical I/ O requirement listed in section 3.4.

3.6.4 CMOS chip functionality

The CMOS chip has three modes of operation: (i) incident optical data can be detected and extracted off-chip electrically (extract state), (ii) data originating from the processor/ memory elements can be injected electrically and transmitted optically towards the next stage (inject state) and (iii) incident optical data can be regenerated on-chip and transmitted towards the next stage (transparent state). A block diagram representation of the CMOS logic is shown in figure 3.12.



Figure 3.12. Modes of operation of a smart pixel: extract, inject and transparent states.

The building block formed by the combination of a receiver, CMOS logic and a transmitter is often referred to as a "smart pixel" [16]. Due to the clustering configuration used in our design, the dual-rail receiver is attached to a pair of detectors located in a detector cluster while the dual-rail transmitter is attached to a pair of modulators located in a neighbouring modulator cluster. The CMOS logic circuitry occupies the silicon area between clusters. Note that in this case, the term "smart pixel" (which suggests a small physical region) is misleading considering that the "pixel" is physically distributed across two clusters, a distance of 800 µm. A detailed description of the design and layout of the CMOS logic is found in [14].

3.7 Optical interconnect design

The design of the optical interconnect was done by Dr. Brian Robertson (now with TeraHertz Photonics, Edinburgh, UK) and is shown in figure 3.13. The design uses a large number of components to illuminate modulators at stage i and route the reflected signal

though the polarizing beamsplitter (PBS). A second quarter-wave plate (QWP2) converts the beams back to RHC polarization. A mini-lens array, located one focal length away from the PMG, focuses the CW beams onto modulators (label C). Following reflection from the modulators, signal beams are left-hand circularly (LHC) polarized; they are converted into s polarization by OWP2 and thus get reflected by the PBS (label D). Beams are relayed towards stage *i*+1 using a 4*f* telecentric mini-lens relay. A corner prism is inserted between the PBS and the mini-lens relay and is used to deflect the beams through an angle of 90°; this is required to form the optical ring topology. Risley prisms are inserted next to the corner prism and are required to perform the fine alignment of signal beams onto detectors at stage i+1. Beams emerging from the mini-lens relay (label E) are reflected by the PBS towards the mirror sections of the PMG element (label F). Note that figure 3.13 suggests that the PMG at stage i+1 must be inverted compared to the PMG at stage i for signal beams to land on the mirror sections. In reality, this is not the case (i.e. both PMG elements are oriented the exact same way) because a beam inversion is introduced by the mirror prism and cannot be shown in figure 3.13. Finally, upon reflection from the PMG element, beams are converted into p polarization by QWP1, transmitted through the PBS and focused on the detectors on the OE-VLSI chip (label G) at which point optical data can either be extracted off-chip or be regenerated towards the next stage.

3.7.1 Selection of focal length

The mini-lens focal length was chosen to be 8.50 mm; this choice represents a compromise between the need to generate a small spot size compared to the modulator active area ($\phi = 52.5 \,\mu$ m) and the requirement for an interconnect length that matches the motherboard pitch (~50 mm). Using $\omega_{OPS} = 176.5 \,\mu$ m and f = 8.50 mm, one calculates a beam radius in the plane of the chip to be $\omega_{CHIP} = \lambda f \pi \omega_{OPS} = 13.1 \,\mu$ m. Assuming a diffractionlimited spot size and a Gaussian intensity distribution, this corresponds to 99% of the optical power being located in diameter of $3\omega_{CHIP} = 39.3 \,\mu$ m, meaning that modulators are oversized by an amount of 13.2 μ m. The resulting optical path length between two stages is approximately equal to $4f = 34 \,\text{mm}$. As discussed in section 3.8, this distance can be increased by inserting high-refractive-index glass in the mini-lens relay.

3.7.2 Power delivery system

The photonic backplane was designed such that the entire four-stage system be supplied by a 1W 852 nm laser source. This requirement was fulfilled using two external-cavity tunable 500 mW diode lasers, one laser for two stages. The 500 mW lasers were purchased from Spectra Diode Labs (model SDL 8630). The free-space output of each 500 mW diode laser was directed through a 50:50 beamspliter and fiber coupled to the input of two optical power supply (OPS) modules using the optical delivery system shown in figure 3.14. The demagnifying telescope at the laser output was required to avoid clipping at the isolator aperture. A half-wave plate (HWP) was inserted at the isolator output to compensate for the polarization sensitivity of the 50:50 beamsplitter. Free-space beams were coupled into singlemode PM fibers. A HWP was required in front of each coupler to ensure that the input polarization be aligned with the polarization axis of the PM fiber. The optical power at different points in the delivery system was measured and the results are included in figure 3.14. The measured power at the OPS input is 128 mW with a polarization extinction ratio of >30 dB.



Figure 3.14. Power delivery system. Numbers in parentheses represent measured optical power (in mW) at different points in the delivery system.

The OPS module uses a diffraction fanout grating with 4 phase levels (theoretical maximum efficiency of 81%, down to 73% if uncoated) in combination with a multi-element Fourier lens and a mini-lens array to generate a 8×4 array of collimated beams. The fanout grating was left uncoated; its diffraction efficiency was measured to be 70% (thus only 3% below the theoretical limit). The efficiency of the Fourier lens was measured to

be 95.6%. The mini-lenses were diffractive Fresnel lens with 8 phase levels; they were AR coated for 852 nm and exhibited a diffraction efficiency of 87% (8% below the theoretical limit of 95%). Combining efficiency and fanout losses, this results in 2.33 mW of optical power in each beam at the OPS output. This result was confirmed experimentally.

Referring to figure 3.13, the OPS output beam must propagate through the 4×4 grating sections of the PMG element, through the QWP1/PBS/QWP2 assembly and through a mini-lens array before reaching the modulator array. The 4×4 PMG grating had 8 phase levels and was left uncoated; its diffraction efficiency was measured to be 73% (14% below theoretical limit). The losses due to the QWP1/PBS/QWP2 assembly are mainly due to imperfect *p* polarization and a small reflection leakage; its efficiency was measured to be 94%. The mini-lens was the same as the one used in the OPS (8 phase levels, AR coated, diffraction efficiency of 87%). Combining efficiency and fanout losses, this results in 87 μ W of optical power incident on each modulator device.

3.7.3 Modulator-to-detector link efficiency

Upon reflection from a modulator, a signal beam must pass through several optical components before reaching a detector at the next stage. The efficiency of individual components was measured. Using these results, the expected modulator-to-detector link efficiency was calculated to be 44%. However, this number must be interpreted as an upper bound because (i) it does not take into account the effects due to multiple-beam interference (cavity effects) and (ii) it neglects the fact that every component will suffer some misalignment which can lead to considerable clipping losses.

Using the above information and assuming a modulator reflectivities of $R_{high} = 82\%$ and $R_{low} = 37\%$ (section 3.6.1), an upper bound for the optical power at the detector is equal to $P_{high} = 31 \,\mu\text{W}$ and $P_{low} = 14 \,\mu\text{W}$.

3.8 Optical packaging considerations

The optical system of figure 3.13 uses a total of 11 components per stage, excluding the components of the OPS module. Each component is required to be aligned in multiple degrees of freedom (DOFs). In some cases, the alignment tolerances are not critical (e.g.

the lateral and longitudinal alignment of the QWPs); in other cases, components must be accurately aligned in all six DOFs (e.g. OE-VLSI chip and mini-lens array). Aligning each component individually is not practical because it results in a large number of DOFs to be precisely controlled and maintained over time. A better strategy consists of integrating components together to form rigid modular subassemblies. This reduces the total number of DOFs and facilitates system assembly and maintenance.

This approach raises the following question: what is the optimal way of partitioning the optical system into modules? To illustrate this, consider the system design of figure 3.13 and the following packaging alternatives: should the mini-lens array next to label G be integrated with the PBS or should it be packaged with the other mini-lens in the relay? Similarly, should the mini-lens array next to the OE-VLSI chip be integrated with the PBS/QWP assembly or integrated with the chip? Usually, the optimal packaging choice can be identified by considering the following aspects:

- Misalignment tolerances: the objective is to partition the optical system in such a way that it results in modules with large tolerances to misalignment. The work done by Neilson [19] on this subject is helpful here; his results indicate that the amount of tolerances to misalignment tolerances depends only on the propagation properties of the optical beams at the interface between modules. Specifically, two modules interconnected with wide collimated beams are tolerant to lateral misalignments but have tight tilt tolerances. Conversely, the opposite is true for a pair of modules separated by small focusing beams. This suggests a trade-off between lateral and tilt misalignment, a point which will be further examined in chapter 7.
- **Practicality**: in addition to the alignment issues, packaging choices are also influenced by their ease of implementation. Simply put, components with similar geometries are easier to integrate together. For example, integrating a bulk refractive lens with a beamsplitter cube is not practical; however, if the same lens can be fabricated as a diffractive Fresnel lens etched on a flat optical substrate, the integration is greatly simplified. This, in fact, is one of the main motivation behind the use of diffractive lenses in our system.

The above considerations has lead to the packaging choices shown in figure 3.15. Each stage is composed of four separate modules: optical power supply (OPS) module, beam combination module (BCM), chip module and relay module. Note that the optical system has been partitioned in such a way that the beams at the interface between any two modules are wide and collimated, resulting in modules that are more tolerant to lateral misalignment, but less to tilt. This fact was confirmed by a misalignment tolerance analysis [20]. Favoring lateral tolerances (at the expense of tilt tolerances) was desirable because it facilitated the integration of the modules on a common optomechanical base-plate fabricated using standard machining capabilities (see section 3.9.1).

Wherever possible, the air gaps within a module were removed to create a solid subassembly. To this end, a block of glass (material = SF56A, index = 1.7622 at 850 nm) was inserted between the pair of relay mini-lenses. The use of a high-index glass increases the stage-to-stage distance up to 47 mm. Similarly, the thickness of QWP1 was increased to facilitate the integration of the PMG in the BCM and the corner prism was also abutted to the PBS. Also, a flat transparent substrate (referred as the jointing plate) was fixed to QWP2. The purpose of the jointing plate is to provide a well-defined alignment plane at the interface between the BCM and the chip module (see chapter 6).



Figure 3.15. A modular implementation of the interconnect design of figure 3.13.

3.9 System implementation and performance

The OPS, BCM and relay modules were assembled and mounted in a stainless steel precision-machined optomechanical baseplate. A photograph of the demonstrator system appears in figure 3.16. Four motherboards are inserted in a standard chassis; the four-stage photonic backplane is located in the back of the chassis. The following sections present the performance obtained from various parts of the system.



Figure 3.16. Photograph of the final system implementation.

3.9.1 Optical modules

Four OPS modules were assembled and characterized. A photograph of an assembled OPS module is shown in figure 3.17. The components were precisely aligned to optimize the spot array at the output of the Fourier lens. At a laser wavelength of 852.0 ± 0.2 nm, the average array pitch was measured to be $1602.0 \pm 0.4 \ \mu\text{m} \times 800.8 \pm 0.2 \ \mu\text{m}$ (compared to a nominal pitch of $1600 \times 800 \ \mu\text{m}$). Using a beam profiler, the spot size was measured to be $13.2 \pm 0.1 \ \mu\text{m}$ (compared to the nominal value of $13.1 \ \mu\text{m}$, see section 3.7.1). The power uniformity of the array was quantified by measuring the power in each spot,

approximating a normal distribution and calculating a standard deviation; a standard deviation of $11 \pm 1\%$ was measured for all 4 OPS modules. The optomechanical interface between the OPS module and the baseplate was designed to be kinematic; this allowed for the OPS module to be removed and replaced repeatably. The OPS insertion repeatability was measured to be less than 0.018° angularly and 15 µm laterally. More details concerning the design and characterization of the OPS module can be found in [18][21].



Figure 3.17. Photograph of the OPS module (after [18]).

The BCM and relay modules were assembled by M. Frédéric Lacroix using an in-situ interferometric alignment technique originally developed by Dr. Brian Robertson [22]. Details of the experimental setup and technique can be found in [23]. Photographs of the assembled modules are shown in figure 3.18.



Figure 3.18. Photographs of (a) the BCM module and (b) the relay module (after [23]).

3.9.2 OE-VLSI chip

A dedicated optical rig was assembled to perform the experimental testing of the OE-VLSI chip. Most of the testing was performed by David Rolston; a detailed description of the testing procedure and results can be found in [14]. The first test consisted of verifying the *inject* mode of operation (electrical in, optical out). This was done by sending electrical data to a transmitter in the array and applying a CW beam to one of the modulator connected to that transmitter. The reflected modulated signal was captured on a high-speed avalanche photodetector. The maximum bit rate that was achieved was 56 Mb/s, consistent with the 50 MHz specification of the 1.5 μ m CMOS process. The measured rise and fall times were 1.6 ns and 1.8 ns respectively.

The second test verified the *extract* mode of operation (optical in, electrical out). Two modulated light beams were directed at a dual-rail receiver in the array. The optical power of the two beams were complementary to simulate dual-rail operation. The modulated light sources were commercial TO-can 850 nm VCSELs, each mounted on a dedicated printed circuit board capable of 1 Gb/s operation. The maximum receiver sensitivity was determined by operating the receiver with no feedback (this can be done by turning off the PMOS transistor that implements the feedback resistor) and operating the receiver at low bit rates. At 1 Mb/s, the receiver showed proper operation with incident beams $P_{high} = 4 \,\mu\text{W}$ and $P_{low} = 2 \,\mu\text{W}$. For higher bit operation, a low feedback resistance is required to reduced the input signal rise time. However, a lower feedback resistance means a lower gain and thus the incident optical power must be increased. Proper receiver operation was demonstrated at 60 Mb/s with $P_{high} = 44 \,\mu\text{W}$ and $P_{low} = 24 \,\mu\text{W}$.

3.9.3 MQW modulator reflectivity

The mirror reflectivity of flip-chip modulator devices was measured to be $R_{high} = 20\%$ and $R_{low} = 10\%$ for a 5V swing under optimal biasing and temperature conditions. Assuming an incident power of 87 μ W on the modulator (section 3.7.2) and a link efficiency of 23% (section 3.9.1), this results in $P_{high} = 4 \mu$ W and $P_{low} = 2 \mu$ W at the detector plane. This is already at the limit of receiver sensitivity, a situation which precluded the operation of the OE-VLSI chips in the optical backplane. Further investigation was carried out to determine the cause of the poor mirror reflectivity. Previous experiments had demonstrated that the reflectivity problem could not be attributed to the flip-chip process nor the annealing of the metal contact (see section 3.6.2). It was suggested that the mirror reflectivity problem was due to the presence of titanium in the *p*-contact mirror. The titanium layer had been included to improve the adhesion of the gold layer to the GaAs cap layer. To verify this hypothesis, two samples where prepared. The first sample was a GaAs substrate with a pure gold mirror; the other was a GaAs substrate with a titanium/gold mirror (the thickness of the titanium layer was not recorded). The mirror reflectivity was measured by directing a long-wavelength (>900 nm) beam from the substrate side. The reflectivity of the pure gold mirror was found to be $R_{du} = 96\%$ whereas the titanium layer degraded the reflectivity down to $R_{TU/du} = 26\%$.

Researchers at Bell Labs have found a similar problem with their devices [24]. They noticed that the performance of their flip-chip modulator devices was significantly lower than what was measured on laboratory devices ($R_{high} = 17\%$, $R_{low} = 9\%$ for flip-chip devices compared to $R_{high} = 76\%$, $R_{low} = 26\%$ for laboratory devices). In reference [24], they write that the decrease in mirror reflectivity for flip-chip devices "is due to the fact that, whereas our relatively large laboratory devices had a pure gold mirror, the flip-chip bonded devices require a sticking layer of titanium between the gold mirror and the device in order to achieve high yield. This layer of titanium lowers the mirror reflectivity and degrades performance."

A potential solution to this problem consists placing a distributed Bragg reflector (DBR) underneath the *p*-contact. The use of a DBR mirror effectively separates the tasks of providing a good mirror reflectivity from the one of providing a good ohmic *p*-contact.

3.9.4 Chip module

Despite the above difficulties, the chip module was assembled, integrating the chip, the mini-lens array, a copper heat spreader, a thermoelectric cooler (TEC) and an aluminum heatsink. A photograph of an assembled chip module is shown in figure 3.20.

The OE-VLSI chip was wirebonded to a flexible printed circuit board (flex-PCB) using a chip-on-board approach, providing 200 bond pad connections in a $44 \times 44 \text{ mm}^2$

area. Electrical lines on the flex-PCB are 50-ohm impedance-controlled and were operated at 1.0 Gb/s with a crosstalk of 4.0% between nearest-neighbour lines. The junction-to-TEC thermal resistance was measured to be 0.31 ± 0.02 °C/W, allowing for the use of a single-stage TEC to regulate the chip at an operating temperature of 40 °C under a maximum thermal load of 13.1 W. Electrical and thermal aspects of the design and testing of the chip module are presented in chapter 4.

The mini-lens array was aligned and packaged to the OE-VLSI chip using a novel six DOF technique described in chapter 5. The module was designed with a semi-kinematic interface; this allows for the manual insertion of the chip module into the system with no need for further adjustments. Details of the interface design are given in chapter 6.



Figure 3.20. Photograph of an assembled chip module.

3.10 Author's contributions

The work presented in this chapter is the result of the efforts of a large group of people. The purpose of this section is to explicitly specify the portion of this work that has been performed by the author of this thesis. These contributions are listed below.

- OE devices: defined geometry and layout of modulators and detectors; designed ebeam photomasks (on Autocad) used for device fabrication and flip-chip process; performed characterization of modulator devices (absorption and responsivity spectrum as a function of wavelength, applied voltage and temperature) before and after flip-chip.
- CMOS chip: designed, simulated and tested (with David Rolston) the dual-rail transimpedance amplifier receiver; performed the design, layout and testing of on-chip Fresnel zone plates and quadrant detectors (used for mini-lens array alignment).
- Optics: defined the layout and specifications of the fanout gratings, PMG, mini-lens array (with Frédéric Lacroix); determined and measured optical power budget; simulated and experimentally measured the polarization losses in the system (with Frédéric Lacroix); designed and assembled the diagnostic chip modules (DCM) used to align the optical ring; implemented and characterized the power delivery system; assisted in the implementation and testing of the OPS module.
- Chip module: designed and specified tolerances of all mechanical parts (on Autocad); performed alignment tolerancing analysis; developed techniques for aligning the minilens array to the chip; assisted in the design of the flex-PCB (the layout was done by Alan Chuah); developed a first-order thermal network for the package and validated the model experimentally; performed electrical high-speed characterization (time-domain reflectometry, eye diagrams, crosstalk); assembled and tested the chip module.

3.11 Conclusion

This chapter presented the design and implementation of a free-space photonic backplane demonstrator system, developed to provide low-latency high-bandwidth optical interconnections between four nodes in a multiprocessor system. The interconnect topology is a unidirectional multi-hop optical ring based on a sender-reserve scheme. The discussion of section 3.5 underscored the large set of considerations involved in turning the topological description of the system into a physical implementation, taking into account the relative strengths and weaknesses of the underlying technologies. aspects of the optical interconnect, including optical power budgetting, misalignment tolerancing, module alignment and assembly. Eric Bernier was responsible for the mechanical design of the baseplate, making sure that all modules would fit together properly. Daniel Brosseau designed and assembled the optical power supply modules. Alan Chuah and Feras Michael were responsible for the design and layout of the flex-PCB and motherboard electronics. Rhys Adams also assisted in assembling OPS modules and integrating optical modules in the baseplate. The following individuals are also acknowledged for stimulating technical discussions: Michael Venditti, Marc Châteauneuf, Emmanuelle Laprise, Julien Faucher, Pritha Khurana, Madeleine Mony, Tsuyoshi Yamamoto.

3.13 References

- [1] R. Grindley, T. Abdelrahman, S. Brown, S. Caranci, D. DeVries, B. Gamsa, A. Grbic, M. Gusat, R. Ho, O. Krieger, G. Lemieux, K. Loveless, N. Manjikian, P. McHardy, S. Srbljic, M. Stumm, Z. Vranesic, Z. Zilic, "The NUMAchine multiprocessor," in Proceedings of the International Conference on Parallel Processing (IEEE Computer Society), pp. 487-496 (2000).
- [2] A.V. Krishnamoorthy, D.A.B. Miller, "Firehose Architectures for Free-Space Optically Interconnected VLSI Circuits," J. of Parallel and Distributed Computing, vol. 41, pp. 109-114 (1997).
- [3] D.Z. Tsang, T.J. Goblick, "Free-space optical interconnection technology in parallel processing systems," Optical Engineering, vol. 33, pp. 1524-1531 (1994).
- [4] T. Sakano, T. Matsumoto, K. Noguchi, "Three-dimensional board-to-board freespace optical interconnects and their application to the prototype multiprocessor system: COSINE-III," Applied Optics, vol. 34, pp. 1815-1822 (1995).
- [5] D. J. Goodwill, D. Kabal, P. Palacharla, "Free-space optical interconnect at 1.25 Gb/ s/channel using adaptive alignment," in Proceedings of Optical Fiber Communication Conference, vol. 2, pp. 259-261 (1999).
- [6] X. Zheng, P. J. Marchand, D. Huang, O. Kibar, N. S. E. Ozkan, S. C. Esener, "Optomechanical design and characterization of a printed-circuit-board-based free-space optical interconnect package," Applied Optics, vol. 38, pp. 5631-5640 (1999).

- [15] T. K. Woodward, A. V. Krishnamoorthy, A. L. Lentine, and L. M. F. Chirovsky, "Optical receivers for optoelectronic-VLSI," IEEE J. of Selected Topics in Quantum Electronics, vol. 2, pp. 106-116 (1996).
- [16] H. S. Hinton, "Architectural considerations for photonic switching networks," IEEEJ. of Selected Areas of Communications, vol. 6, pp. 1209-1226 (1988).
- [17] B. Robertson, "Design of an optical interconnect for photonic backplane applications," Applied Optics, vol. 37, pp.2974-2984 (1998).
- [18] D. F-Brosseau, F. Lacroix, M. H. Ayliffe, E. Bernier, B. Robertson, F. A. P. Tooley, D. V. Plant, A. G. Kirk, "Design, implementation, and characterization of a kinematically aligned, cascaded spot-array generator for a modulator-based free-space optical interconnect," Applied Optics, vol. 39, pp. 733-745 (2000).
- [19] D. T. Neilson, "Tolerance of optical interconnections to misalignment," Applied Optics, vol. 38, pp. 2282-2290 (1999).
- [20] F. Lacroix, B. Robertson, M. H. Ayliffe, E. Bernier, F. A. P. Tooley, M. Chateauneuf, D. V. Plant, A. G. Kirk, "Design and implementation of a four-stage clustered free-space optical interconnect," Optics in Computing '98, Brugge, Belgium, 17-20 June 1998, pp.107-110.
- [21] E. Bernier, F. Lacroix, M. H. Ayliffe, B. Robertson, F. A. P. Tooley, D. V. Plant, A.
 G. Kirk, "Implementation of a compact, four-stage, scalable optical interconnect," Optics in Computing 2000 Conference, June 18-23 Québec City, Qc., Canada.
- [22] B. Robertson, Y. Liu, G. C. Boisset, M. R. Taghizadeh, D. V. Plant, "In situ interferometric alignment systems for the assembly of microchannel relay systems," Applied Optics, vol. 36, pp. 9253-9260 (1997).
- [23] F. Lacroix, "Design, analysis and implementation of free-space optical interconnects," Chapter 7, Ph.D. Thesis, McGill University, Montréal, Canada, 2000.
- [24] A. V. Krishnamoorthy and K. W. Goossen, "Progress in optoelectronic-VLSI smart pixel technology based on GaAs/AlGaAs MQW modulators," International J. of Optoelectronics, vol. 11, pp. 181-198 (1997).

Chapter 4: Chip module design and testing

4.1 Introduction

This chapter presents the design, implementation and testing of the chip module, with a focus on the electrical and thermal aspects. The work presented here builds on an earlier design that was developed for the first-generation BiCMOS chip [1]. The chapter is organized as follows. First, the design constraints in terms of mechanical alignment, electrical signal integrity and temperature stabilization are described (section 4.2). Next, the mechanical design of the chip module and its assembly are presented (section 4.3). Section 4.4 present the electrical design and include the result of signal integrity tests. The thermal design of the module including the experimental evaluation of its thermal resistance is found in section 4.5. The remaining section acknowledges other contributors to this work.

4.2 Design choices and constraints

4.2.1 Alignment issues

A critical design requirement was for the chip module to be manually replaceable. This is necessary because the OE-VLSI chip is the only active element in the photonic backplane; it is prone to failure and is likely to be upgraded with time. This requirement calls for a misalignment-tolerant design that uses a kinematic alignment scheme at the interface between the chip module and the photonic backplane.



Figure 4.1. Packaging alternatives: (a) mini-lens array can be integrated with the chip or (b) mini-lens array can be integrated with the BCM.

With the discussion of section 3.8 in mind, two different packaging choices were examined and are shown in figure 4.1. In the first case, the mini-lens array is aligned to the chip and integrated as part of the chip module (figure 4.1(a)). In the second case, the mini-lens array is integrated as part of the beam combination module (BCM) (figure 4.1(b)).

Packaging the mini-lens array with the BCM is certainly more practical (this is because the mini-lens array is directly adjacent to the BCM) but it does not lead to a misalignment-tolerant design. To see this, consider the results of the misalignment tolerance analysis shown in table 4.1. The tolerances correspond to the amount of misalignment required for the modulator-to-detector power throughput to drop by 1%. The analysis is performed by varying one degree of freedom (DOF) at a time. Only optical power losses caused by clipping effects are taken into account. Calculations are based on paraxial Gaussian beam propagation theory; geometrical optics is used for off-axis ray propagation. A detailed description of the tolerance analysis is found in [2][3].

Degree of freedom (DOF)	Mini-lens with chip (figure 4.1(a))	Mini-lens with BCM (figure 4.1(b))
lateral (x, y)	± 26 μm	± 8 µm
tilt (θ_x, θ_y)	± 0.03°	± 0.12°
rotational (θ_z)	± 0.36 °	± 0.10 °
longitudinal (z)	± 500 μm	± 125 μm

Table 4.1. Results of misalignment tolerance analysis.

The results of table 4.1 indicate that integrating the mini-lens array with the OE-VLSI chip relaxes lateral, longitudinal and rotational alignment tolerances by a factor of approximately 4 but tightens the tilt alignment tolerances by the same factor. These results are consistent with the discussion of section 3.8 to the effect that modules interconnected with collimated beams are tolerant to lateral misalignment but have tight tilt tolerances. This trade-off between lateral and tilt tolerances is inherent to the alignment problem and can be linked to the principle of optical invariance (see chapter 7). By choosing to integrate the mini-lens array with the chip, a kinematic interface with the BCM module becomes possible because the resulting lateral and rotational misalignment tolerances can be achieved with standard machining capabilities (the accuracy of modern computer numerically controlled (CNC) machines being limited to about $\pm 10 \mu m$ [4]). This choice also gives rise to the following design challenges. First, the mini-lens array must be accurately aligned to the OE-VLSI chip in all 6 DOFs. This is achieved by a novel alignment technique that uses off-axis diffractive elements placed on the CMOS chip along with metal alignment markers located on the mini-lens array substrate; this technique is presented in chapter 5. Secondly, the design of the interface mechanics must exhibit outstanding angular repeatability to satisfy the stringent tilt tolerance of the module (θ_x , $\theta_y = \pm 0.03^\circ$). This is accomplished by adding a flat transparent substrate to the BCM (referred to as the jointing plate) and taking advantage of the optical flatness of the mini-lens array and jointing plate substrates to use them as passive alignment planes that defines the tilt of the chip module (see chapter 6).

4.2.2 Electrical issues

The basic functions of the chip module are to provide stable mechanical support for the OE-VLSI chip, distribute power, control and data lines to the CMOS chip and offer an adequate means for removal of heat. The CMOS chip (figure 3.10) has a total of 232 bond pads, of which 207 are used during normal chip operation (25 unused pads are dedicated to device testing). In order to meet the spatial constraints imposed by the optical interconnect design, a packaging scheme capable of providing the required connectivity in an area smaller than 47×47 mm² was needed (this is because the pitch between stages is 47 mm, see section 3.8). In addition, the design had to support a mechanically flexible connection to the motherboard PCB.

These objectives were realized using a chip-on-board (COB) packaging scheme in combination with flexible PCB (flex-PCB) technology with impedance-controlled microstrip lines [5]. The OE-VLSI chip was mounted directly on a copper heat spreader that was subsequently inserted through an opening machined in the flex-PCB. The chip was wirebonded directly to the flex-PCB, thereby eliminating the need for an electronic chip carrier. This resulted in a small, simple, and low-cost method of chip packaging. This packaging scheme also improves high-frequency performance due to the absence of package lead inductance and the ability of placing surface mount components (i.e. decoupling capacitors and termination resistors) in close proximity to the chip bond pads.

4.2.3 Thermal issues

3

The basic principle behind the operation of MQW modulators is the wavelength shift of the excitonic absorption peak with applied voltage described in section 2.3.3. For MQW modulators based on a GaAs/AlGaAs quantum well structure, the position of the exciton peak is temperature dependent, with a sensitivity equal to ~0.27 nm/°C. This dependence on temperature is undesirable because it limits the temperature range over which MQW modulators can operate efficiently under constant biasing conditions.

The useful temperature range of MQW modulators has been investigated by Venditti [6]. His measurements were performed on GaAs/AlGaAs modulators with a quantum well structure identical to the one used here. Assuming constant biasing conditions and a 5 V swing, his results show that the operating temperature must be within ± 5 °C of the designed temperature in order for the difference in reflectivity (ΔR) not to fall below 90% of its optimal value. If the latter criterion is relaxed to 70%, the useful temperature range is effectively doubled to ± 10 °C. Thus, it can be concluded that without any means of controlling the operating temperature, it would be difficult, if not impossible, to operate MQW modulators efficiently. For this reason, a thermoelectric cooler (TEC) was incorporated into the chip module. This choice was motivated by the small size, the light weight, and the ability of TECs to achieve precise temperature stabilization when combined with a thermistor in a closed-loop configuration.

4.3 Chip module design overview

4.3.1 Physical dimensions

An exploded view of the chip module is shown in figure 4.2. The chip module integrates a mini-lens array, a OE-VLSI chip, a copper heat spreader, a TEC and a heatsink. The design results in a compact module with a footprint of 44×44 mm², satisfying the physical constraints imposed by the optical interconnect design. The length of an assembled module, measured from the front of the mini-lens array to the back of the heatsink, is 45 mm, half of which is occupied by the heatsink fins. The mechanical drawings specifying the dimensions and tolerances of each component can be found in appendix A.



Figure 4.2. Exploded view of the chip module design.

4.3.2 Materials selection

All components were custom-designed and fabricated on a CNC machine (except for the TEC and heatsink, which were available commercially). The mini-lens holder, mounting spacer, flex-PCB mount and protective cover were machined out of aluminum 6061, followed with a clear sulfuric anodization finish. The choice of this alloy was motivated by its ease of machining, lightness, low residual stress, low cost and availability. In addition, the high electrical conductivity and nonsparking properties of aluminum make it an excellent electrical shielding material. The purpose of the anodization step was to electrically isolate the mechanical components to eliminate any possibility of electrical shorts with the nearby chip-on-board electronics. Sulfuric anodization was chosen instead of hard anodization to maintain the tight machining tolerances specified on some of the components; this is because sulfuric anodic coatings are typically 5-15 μ m thick while hard anodizing coatings can be as much as 50 μ m thick [4].

The finned heatsink was made of 6063 aluminum, followed with a clear sulfuric anodization finish. The use of 6063 aluminum for heatsinking applications is common and is due to the fact that 6063 conducts 15% more heat than 6061 and can more easily be extruded into complex shapes. The heat spreader was made of a high-copper alloy (C18500), a choice motivated by its high thermal conductivity, its satisfactory machinability rating and its availability. High thermal conductivity was required to provide an efficient thermal path between the chip and the TEC. Also, the heat spreader was nickelplated to prevent copper oxidation.

4.3.3 Chip module assembly

A photograph of an assembled chip module was shown previously in figure 3.20. The design of the chip module offers a high degree of modularity. Components are joined to one another using dowel pins and screws. A module can be disassembled and reassembled in a few minutes. The design also allows for the OE-VLSI chip to be removed and replaced (of course, this requires the wirebonds to be reworked). The assembly process can be broken down in three main steps: (i) mounting and wirebonding the chip to the flex-PCB, (ii) mounting the mini-lens array onto its holder and (iii) alignment of the mini-lens array to the chip. What follows is a description of this assembly sequence.

Referring to figure 4.2, the chip was passively aligned and fixed to the heat spreader. Chips were precision-diced to within 50 μ m of the bond pad frame and the pedestal portion of the heat spreader was machined to within $\pm 10 \mu$ m of the chip's nominal dimensions. This allowed for the alignment of the chip to be performed manually using the sidewalls of the heat spreader pedestal as passive alignment references. The chip was attached using silver-filled conductive epoxy, providing an efficient electrical and thermal path to the heat spreader. Next, the chip was inserted through the opening of the flex-PCB, the latter having previously been glued onto the flex-PCB mount using high temperature thermal set lamination techniques. A 1.0 mm thick acetal spacer was placed between the flex-PCB mount and the heat spreader. The chip was then centered on the flex-PCB opening using a pair of acetal dowel pins and the heat spreader was fastened to the back of the flex-PCB mount using nylon screws. The use of plastic screws, pins and spacers is to avoid a "thermal short" between the heat spreader and the aluminum mount. The chip was then wirebonded to gold-on-nickel plated printed circuit using an aluminum wedge wirebonder. The TEC, protective cover and heatsink were then incorporated. The TEC was clamped in place by fastening the heatsink directly to the back of the flex-PCB mount using a pair of nylon screws. Thermally-conductive, self-adhesive, 250-µm thick elastomer interface pads (T-pli 210 from Thermagon Inc.) were placed on both sides of the TEC to ensure efficient heat flow across the interfaces.

Next, the mini-lens array was positioned and glued to its holder. The mini-lens substrate (fused silica, 1.0 mm thick) was precision-diced to within $\pm 25 \ \mu\text{m}$ of a lithographically-defined metal frame. The positioning of the mini-lens array was done manually using a semi-kinematic passive alignment scheme. As shown in figure 3.20, the lateral position (x, y) and rotation (θ_z) of the mini-lens substrate are constrained by three press-fit precision dowel pins inserted in the mini-lens holder (dowel pin positional tolerance = $\pm 10 \ \mu\text{m}$). The remaining DOFs, tilt (θ_x, θ_y) and longitudinal (z), are constrained by the mini-lens holder pedestal (flatness tolerance = $\pm 20 \ \mu\text{m}$). The mini-lens array was fixed in place using ultraviolet curing optical epoxy.

The last step involves the alignment of the mini-lens array to the chip. First, an aluminum spacer (see figure 4.2) was glued onto the flex-PCB. The mini-lens array was then positioned in front of the chip and the two were brought into alignment using a technique described in chapter 5. During the alignment process, the mini-lens holder was free to move in all six DOFs and did not come into physical contact with the aluminum spacer (there was a 400 μ m gap separating the two components). Once alignment was completed, the mini-lens holder was fixed to the mounting spacer using room temperature epoxy.

4.4 Electrical packaging and high-speed testing

4.4.1 Electrical packaging design

A photograph of the flex-PCB is shown in figure 4.3. The flex-PCB provides connectivity to 207 bond pads on the CMOS chip, 64 of which are high-speed lines. A total of 48 pads are dedicated to ground or power connections. To minimize routing area, power and ground wirebonding rings are placed around the periphery of the chip. The rings are directly connected to their respective copper plane through multiple vias. The use of ground and power rings reduces the number of wirebonding fingers down to 159, allowing for all signal traces to be routed using a four-layer flex-PCB in an area smaller than 44×44 mm². A photograph of the region surrounding the chip appears in figure 4.4.



Figure 4.3. Photograph of the flex-PCB assembly.





All four copper layers (signal, power, ground, signal) are 0.5 oz/ft² (17 μ m) in thickness and are separated by 3 mils (76 μ m) of kapton ($\varepsilon = 3.7$). The outside copper layers are coated with 2 mils (50.8 μ m) of kapton; this results in a total board thickness of 15.8 mils (400 μ m). Signal traces are 5.5 mils (140 μ m) in width and are placed at a minimum pitch of 10.5 mils (267 μ m). Given nominal material thicknesses and dielectric constant, the calculated impedance for this coated microstrip stackup is 47.3 Ω and the propagation delay is 5.2 ps/mm [7].

The electrical packaging has been designed for high-speed operation (> 100 Mbit/s), with fall and rise times shorter than 2 ns. The nominal physical length of a signal trace between the chip and the motherboard is about 180 mm. This corresponds to about half the effective length of a 2 ns rising edge; this means the flex-PCB must be treated as a distributed system requiring proper line terminations for both input and output signals. The choice of the line termination scheme is an important design issue. To minimize signal reflections and maximize switching speed, one would like to terminate high-speed signal lines with 50 Ω resistors. For our implementation, however, load termination was unsuitable for the following reasons. Input signals to the chip could not be load terminated because this required a large number of external 50 Ω resistors to be placed close to the chip, resulting in a prohibitively large footprint. Output signals from the chip could not be terminated with 50 Ω loads because this would have exceeded the current driving capability of the CMOS output pad drivers, not to mention the thermal problems associated with every output driver generating an average power of 0.25 W. For these reasons, series terminations were used for both inputs and outputs to the chip. Inputs to the chip were series terminated by placing a 33 Ω series resistance at the output of the motherboard driving electronics, this value corresponding to the difference between the line impedance (50 Ω) and the nominal output resistance (17 Ω) of the driving electronics. Outputs from the chip didn't require the use of external matching resistors because the CMOS output driver output resistance was already slightly larger than the line impedance.

An important signal integrity issue is related to the noise created by high supply current switching transients, often referred to as simultaneous switching noise (SSN) [8]. This noise is caused by a rapid change in current consumption of the circuit (due to many



Figure 4.5. Time-domain reflectometry (TDR) measurement showing typical response.



Figure 4.6. Eye diagram at 1.0 Gbit/s NRZ modulation rate (PRBS 2²³-1).

4.4.3 Crosstalk measurements

Considering the large number of closely packed high-speed lines on the flex-PCB, an important signal integrity issue is the amount of crosstalk between adjacent lines. Here,

we define crosstalk as the ratio of the induced voltage amplitude to the driving voltage amplitude. As a voltage pulse propagates down a line, it generates both forward and reverse propagating crosstalk on adjacent lines. In general, for traces above a ground plane, the inductive and capacitive components of reverse crosstalk are approximately equal, have the same polarity and therefore reinforce. The amount of reverse crosstalk can be calculated using [7]:

reverse crosstalk ~
$$\frac{1}{1 + (D/H)^2}$$
 (4.1)

where D is the center-to-center spacing between two lines and H is the trace height above the ground plane. For our design, the closest lines have D = 10.5 mils and H = 3mils, resulting in a worst-case theoretical reverse crosstalk of 7.5% between nearestneighbor lines. Unlike reverse crosstalk, the inductive and capacitive components of forward crosstalk are of opposite polarity and therefore tend to cancel. For microstrip lines, however, most of the electric field travels through air instead of through the dielectric; this reduces the capacitive component of forward crosstalk and usually results in small negative forward crosstalk [7].

The amount of forward and reverse crosstalk was measured using dedicated microstrip lines on the flex-PCB. As illustrated in figure 4.7, a 2 V pulse with 200 ps rising and falling edges was applied to the near-end of an aggressor line and the amplitude of the induced voltages, V_1 and V_2 , were measured at the far-end of victim lines 1 and 2 respectively. Crosstalk for other remote lines was negligible. The length of the interfering microstrip lines was 100 mm. Measurements were performed under different termination conditions.

First, the forward crosstalk component was determined by using 50 Ω terminations at both ends of the aggressor and victim lines ($R_{load} = 50 \Omega$ and $R_{source} = 50 \Omega$). This arrangement eliminates the near-end reflection of reverse crosstalk. This way, the crosstalk response measured at the far-end is mainly due to forward crosstalk. As shown in figure 4.8, forward crosstalk is negative; it appears as a narrow negative peak on a rising edge, the width of the peak being approximately equal to the rise time.



Figure 4.7. Experimental setup used to characterize forward and reverse crosstalk.



Figure 4.8. Crosstalk response on both victim lines with $R_{load} = 50 \Omega$ and $R_{source} = 50 \Omega$. In this case, the response is due to negative forward crosstalk.

Next, the near-end of the victim lines were shorted to ground ($R_{load} = 50 \ \Omega$ and $R_{source} = 0 \ \Omega$). This arrangement provides total reflection of the reverse crosstalk at the near-end and results in the superposition of both forward and reverse crosstalk at the far-end. As shown in figure 4.9, reverse crosstalk appears as a wide negative pulse at the far-end, the width of the pulse being approximately equal to twice the delay of the microstrip line. Reverse crosstalk is positive; it appears as a negative voltage at the far-end due to the negative reflection at the near-end. Reverse crosstalk amplitudes of -160 mV and -60 mV

These measurements indicate that 5 V NRZ data propagating down a line in our system will generate a 200 mV crosstalk pulse on its nearest lines and 75 mV on the next, the width of the pulse being equal to twice the line delay. This translates into 4.0% and 1.5% respectively, which is tolerable. In a worst-case situation, however, crosstalk components originating from nearby lines add together. In this case, rising edges occurring simultaneously on multiple lines on both sides of a victim line can generate as much as 550 mV of crosstalk, which is significant.

4.5 Thermal design and experimental evaluation

The MQW modulators were designed to provide a change in reflectivity ΔR of 45% when operated at $\lambda = 852$ nm and T = 40 °C (see section 3.6.1). For ΔR not to fall below 40%, the chip temperature must be stabilized to 40 ± 5 °C. Temperature stabilization is easily achieved with ± 0.01 °C precision by using a TEC (from Melcor Corp., P/N CP-1.0-71-05L) in combination with a precision mini-probe thermistor (from Betatherm Corp., P/N 10K3A1) in a closed-loop feedback configuration. The thermistor probe is 1.0 mm in diameter and 5.0 mm long; it is inserted and fixed using thermal epoxy in a small hole drilled on the side of the copper heat spreader pedestal, just underneath the chip. The thermal resistance between the thermistor probe and the chip surface must be minimized to ensure that the thermistor temperature reading is close to the actual chip temperature.

HSPICE simulations were performed to determine the average power dissipation of the CMOS chip. Results indicate a worst-case dissipation of about 7.5 W, of which 3.5 W is due to the pad drivers. Given this thermal load, minimizing the package thermal resistance is essential in order to avoid the need for a multi-stage TEC or a prohibitively large heatsink. The thermal resistance was minimized by mounting the chip directly onto a high-copper alloy (C18500) heat spreader using a thin layer of silver-filled epoxy (Epotek H20E). This arrangement provides an excellent thermal path to the TEC due to the high thermal conductivity of alloy C18500 (324 W/m/°C).

The thermal components of the chip module are shown in figure 4.11. A first-order thermal network model was developed to determine the thermal resistance between the surface of the silicon chip to the cold side of the TEC (this is referred to as the junction-to-

- epoxy layer: the thickness of the epoxy layer was estimated to be 10 µm. This was determined experimentally by applying equal layers of epoxy between a pair of microscope slides and using a high-precision micrometer to measure the total thickness of the slides with and without the epoxy. The thermal conductivity of the epoxy was available from the manufacturer: $k_{epoxy} = 2.0 \text{ W/(m} \cdot ^{\circ}\text{C})$. This results in $R_{epoxy} = 0.062 \text{ °C/W}$.
- copper heat spreader: the thermal conductivity of C18500 is $k_{copper} = 324$ W/(m·°C) [4]. The top portion of the heat spreader has an area 9.0×9.0 mm² and a thickness of 3.7 mm which yields $R_{top} = 0.14$ °C/W. The bottom portion of the heat spreader has an area 23×23 mm² (to match the TEC dimensions) and a thickness of 3.0 mm. Note that R_{bottom} cannot be calculated directly using equation (4.2) because the heat originating from the top section must spread over the area of the bottom section. One way of including the thermal spreading resistance in the calculation of R_{bottom} is to use Kennedy plots [5]. Using this approach, we find that $R_{bottom} = H_2/(k_{copper}\pi^{1/2}d)$, where d is the lateral dimension of the heat source (here, d = 9.0 mm) and H_2 is a parameter that depends on the aspect ratio of the component and that can be read directly from the Kennedy plots. In our case, $H_2 = 0.5$, which results in $R_{bottom} = 0.24$ °C/W.
- thermal interface pad: the manufacturer data sheet specifies a thermal resistance of $0.14 \,^{\circ}\text{C-in}^2/\text{W}$ for an elastomer pad of 20 mils (500 µm) thickness and assuming no contact pressure. For a 250 µm thick pad of area $23 \times 23 \,\text{mm}^2$, this translates into $R_{pad} = 0.085 \,^{\circ}\text{C/W}$. Note that the actual value of R_{pad} is expected to be lower because significant pressure results from the heatsink being clamped to the flex-PCB mount and the fact that thermal resistance decreases with increasing pressure [12].

Using the above, one calculates the junction-to-TEC thermal resistance as follows: $R_{junction-to-TEC} = R_{chip} + R_{epoxy} + R_{spreader} + R_{pad} = 0.47 \text{ °C/W}$. To validate this result, a calorimetric assembly was built to measure the thermal resistance along the thermal path. This was done by replacing the silicon chip with a square resistive heating element (0.5" × 0.5")
area, $R_{heater} = 7.5 \Omega$, from Minco Corporation, P/N HK5572R7.5L12B) and substituting the TEC with a brass spacer of identical size and known thermal resistance (for C36000 brass: $k_{brass} = 116$ W/(m·°C) [4]). Five thermistor mini-probes were inserted into small holes drilled at specific locations in the assembly, as shown in figure 4.12. The thermal resistance between any two thermistors is equal to the difference in temperature (ΔT) divided by the amount of heat flowing through the assembly.



Figure 4.12. Design of the calorimetric assembly.

The heating element was attached to a tall brass post made of C36000 brass which was attached to the heat spreader using a thin layer of silver-filled epoxy. The brass post served two purposes. First, it was used to convert the $0.5" \times 0.5"$ area of the heating element into the required $9.0 \times 9.0 \text{ mm}^2$ in order to properly simulate the heat generated from the chip. Secondly, by inserting a pair of thermistors at a known distance in the post (here, 20.0 mm), the amount of heat going into the module (Q_{in}) can be monitored according to the following equation:

$$Q_{in} = \frac{T_{post1} - T_{post2}}{R_{th,post}}$$
(4.3)

where T_{post1} and T_{post2} are the temperature readings of the thermistors and $R_{th,post}$ is the thermal resistance between the thermistors: $R_{th,post} = l/(k_{brass} \cdot area) = 2.13$ °C/W. A photograph of the assembled calorimetric assembly is shown in figure 4.13. The front of the assembly was thermally insulated to ensure that most of the heat generated by the heating element ($P_{heater} = V^2/R_{heater}$) flows through the heat spreader and is not lost to the surrounding environment through natural convection



Figure 4.13. Calorimetric assembly to measure thermal resistances.

Table 4.2. Thermistor temperature data for different power settings ($T_{room} = 23.6^{\circ}$ C).

P _{heater} (W)	T _{posti} (°C)	T _{post2} (°C)	Q _{in} (W)	T _{spreader} (°C)	T _{brass} (°C)	T _{heatsink} (°C)	R _{junction-} to-TEC (°C/W)
1.54	40.12	37.02	1.46	36.37	36.17	35.40	0.292
2.09	46.36	42.34	1.89	41.55	41.27	40.24	0.275
2.60	50.63	45.65	2.34	44.65	44.23	42.78	0.315
3.13	55.58	49.56	2.83	48.42	47.83	46.27	0.320
3.64	60.22	53.36	3.22	52.06	51.39	49.30	0.320
4.16	65.06	57.22	3.68	55.78	54.92	52.70	0.333

Chapter 4: Chip module design and testing

The heating element was set to a known power setting (P_{heater}) and the steady-state temperature data at each thermistor was recorded. This step was repeated for several power settings of the heating element and the results appear in table 4.2. Note that the amount of heat flowing through the brass post (Q_{in}) is systematically less (by about 5% to 10%) than the power generated by the heating element (P_{heater}) , indicating that some of the heat is being lost to the environment in spite of the thermal isolation. The data of table 4.2 can be used to derive the value of $R_{junction-to-TEC}$ according to:

$$R_{junction-to-TEC} = \left(\frac{T_{post2} - T_{brass}}{Q_{in}}\right) - R_{bottom, post} - R_{half, brass}$$
(4.4)

where $R_{bottom,post}$ corresponds to the thermal resistance between the thermistor at T_{post2} and the silver-filled epoxy and $R_{half,brass}$ corresponds to the thermal resistance between the thermistor at T_{brass} and the thermal interface pad. The thermistor at T_{post2} is located 2.5 mm from the heat spreader and so $R_{bottom,post} = 0.266$ °C/W. The thermistor at T_{brass} is located 1.6 mm into the 23 × 23 mm² brass spacer and so $R_{half,brass} = 0.026$ °C/W. Using the above, the experimental value of $R_{junction-to-TEC}$ can be calculated; the results appear in the last column of table 4.2. The average value of $R_{junction-to-TEC}$ is equal to 0.31 ± 0.02 °C/W, which is well below the theoretical calculation of 0.47 °C/W. A significant portion of this discrepancy is expected to originate from the interface pad having a much lower thermal resistance than expected due to the high contact pressure.

Another important quantity to determine is the maximum thermal load that the chip module can dissipate under forced air convection. To do this, the brass spacer was replaced by the TEC in the assembly of figure 4.13 and the temperature of the heat spreader thermistor ($T_{spreader}$) was stabilized at 40 °C. The power generated by the heating element was slowly increased and eventually reached a point where the TEC was unable to pump additional heat out of the module while keeping the thermistor at the set temperature. This point was found to correspond to $Q_{in} = 13.1$ W, which confirmed that the design can easily withstand the worst-case thermal load produced by the chip (7.5 W).

4.6 Conclusion

The design and testing of a high-performance chip module accommodating a 32×32 array of MQW modulators flip-chip bonded to a $9 \times 9 \text{ mm}^2$ CMOS chip was described. The module integrates a mini-lens array, a copper heat spreader, a thermoelectric cooler and an aluminum heatsink. The design simultaneously addresses the issues of mechanical alignment, electrical signal integrity and thermal heat dissipation. The key features of the design are listed below:

- The chip module assembly is simple and modular. Components are joined to one another using dowel pins and screws. A module can be assembled in a few minutes.
- The chip is mounted directly on a four-layer flex-PCB using a custom chip-on-board approach, providing 207 connections to the OE-VLSI chip in an area of 44 × 44 mm².
- The electrical design uses series terminations for both inputs and output data paths. Various techniques are used to minimize the amount of SSN.
- Impedance-controlled lines can be operated at 1.0 Gb/s with an open eye diagram. The measured crosstalk between nearest-neighbor lines is 4.0%.
- The junction-to-TEC thermal resistance was measured to be 0.31 ± 0.02 °C/W. This low thermal resistance allows for the use of a single-stage TEC and a compact heatsink to regulate the chip at an operating temperature of 40 °C. Under forced air convection, the module can dissipate a maximum thermal load of 13.1 W while keeping the chip temperature at T = 40 °C.
- The design implements a kinematic interface between the chip module and the interconnect optics, allowing for the module to be manually replaced in the free-space photonic backplane without the need for further adjustments. The design and experimental evaluation of the kinematic interface is presented in chapter 5.

4.7 Acknowledgments

The work presented in this chapter builds on an earlier design that was developed by David Kabal and the author of this thesis for the first-generation BiCMOS chip [13]; this work was performed in 1997 and is not included in this thesis. Many of the approaches pesented in this chapter were derived from the lessons learned on the first-generation chip module.

The following individuals contributed to the design of the second-generation chip module. David Rolston designed and performed the layout of the CMOS chip; his input into the electrical design and layout of the flex-PCB was invaluable. Alan Chuah was responsible for the design and layout of the flex-PCB; he managed to get everything right on the first try. Feras Michael also participated in the design of the flex-PCB and was particularly efficient at procuring parts such as the high-speed connector and surface-mount capacitors and resistors.

4.8 References

- M. H. Ayliffe, D. Kabal, F. Lacroix, E. Bernier, P. Khurana, A. G. Kirk, F.A.P. Tooley, and D.V. Plant, "Electrical, thermal and optomechanical packaging of large 2D optoelectronic device arrays for free-space optical interconnects," J. of Opt. A: Pure and Appl. Opt., vol. 1, pp.267-271, 1999.
- [2] F. Lacroix, B. Robertson, M. H. Ayliffe, E. Bernier, F. A. P. Tooley, M. Châteauneuf, D. V. Plant, and A. G. Kirk, "Design and implementation of a four-stage clustered free-space optical interconnect," in Proc. 1998 Optics in Computing Conference, Brugge, Belgium, June 1998, pp. 107-110.
- [3] F. Lacroix, "Design, analysis and implementation of free-space optical interconnects," Ph.D. Thesis, McGill University, Montréal, Canada, 2000.
- [4] R. E. Green, C. J. McCauley, E. Oberg, F. D. Jones, H. L. Horton, and H. H. Ryffel, Machinery's Handbook, 25th edition, Industrial Press, New York, 1996.
- [5] R. R. Tummala and E. J. Rymaszewski, Microelectronics Packaging Handbook, Van Nostrand Reinhold, New York, 1989.

Chapter 5: Intra-module alignment techniques

5.1 Introduction

The stringent alignment requirements of 2D-FSOI systems can be significantly alleviated if components are packaged together to form rigid modules. The philosophy behind this approach is to shift the burden of high-precision alignment towards the assembly of modules in order to relax misalignment tolerances between modules and thus facilitate system assembly and maintenance. This approach was used in the design of the photonic backplane system described in chapter 3. The alignment of this system was divided into two steps:

- Intra-module alignment: the first step consists of aligning components together to form rigid modules and is referred to as intra-module alignment. A good example is the assembly of the relay module, where two mini-lens arrays are accurately aligned and attached at each end of the block of glass. Another example is the assembly of the chip module, where a mini-lens array is aligned and packaged with the OE-VLSI chip. Intra-module alignment techniques usually require expensive high-precision staging equipment and dedicated assembly rigs.
- Inter-module alignment: the second step consists of aligning modules to one another and is referred to as inter-module alignment. A desirable attribute is for modules to have large misalignment tolerances, enough for passive alignment techniques to be used during system assembly. A good example of this is the OPS and chip modules which were designed with a kinematic interface, allowing them to be manually inserted and removed from the system repeatedly with no need for further adjustments.

This chapter focuses on intra-module alignment techniques. In the general case, intramodule techniques are used to aligned a pair of 2D array components in all six degrees of freedom (DOF). The separation between the components to be aligned may range from 10's of microns to 10's of millimeters. The chapter is organized as follows. Section 5.2 is a review of previously published intra-module alignment techniques, both active and passive. Section 5.3 examines the problem of aligning a lens array with an OE-VLSI chip and two techniques are proposed. The first technique uses off-axis Fresnel lenses on the lens array substrate in combination with silicon quadrant detectors on the chip (section 5.4). The second technique uses linear off-axis Fresnel zone plates (FZPs) implemented directly on the CMOS chip combined with metal alignment markers on the lens array substrate (section 5.5). The design and experimental evaluation of each technique is presented. Alternative implementations are also proposed (sections 5.5.8 and 5.5.9).

5.2 Review of previously published techniques

Several techniques for aligning lens arrays to 2D device arrays have been proposed in the past. These techniques can be categorized as either active or passive, the difference being that active alignment techniques require the activation of the OE devices, whereas this is not the case for passive alignment [1]. One common active technique consists of operating an array of emitters (typically VCSELs) above threshold and bringing the lens array into alignment while observing the far-field diffraction patterns of the laser beams with a CCD camera [2][3]. Another active method uses laser feedback effects due to reflections from the microlenses [4]. Drawbacks to active techniques are that they are time consuming, labor intensive, expensive, and not amenable to high-volume production.

For this reason, current research has focused on the development of passive alignment techniques which can be further divided into two categories: mechanical and optical methods. The most common mechanical method uses flip-chip technology to pull parts into position with an accuracy typically better than 2 μ m [5][6]. This technique takes advantage of the self-alignment capability arising from the high surface tension forces of reflowed solder. This is shown in figure 5.1 where a lens array substrate is aligned to an optoelectronic chip. The metal pads on the chip and their counterparts on the microlens substrate may be initially separated by as much as three times the average bump radius, but if the mating surfaces touch and are reasonably wettable then self-alignment will occur [7].



Figure 5.1. Reflowed solder self-alignment technique.

Other mechanical methods include the selective etching of silicon material to form grooves and hollows [8]-[10], the fabrication of micro-connecting plugs and sockets made with thick photoresist [11], and the use of deep proton lithography to fabricate mechanical alignment structures directly into PMMA (polymethylmethacrylate) components [12]. Mechanical methods are very promising because, beyond reducing cost, they provide an avenue for wafer-scale integration. However, they are generally limited to cases where the alignment planes are close together (i.e. less than a few 100's of microns) and thus they cannot be used in situations where components are separated by many millimeters.

Optical methods use lenses, gratings, fiducial markers, and other structure to transfer the alignment information of one component into the plane of the other. Optical methods are commonly used to achieve mask-to-wafer alignment in microlithography applications [13]. Other examples include techniques employing dual-side lithography for the monolithic integration of microlenses on the GaAs substrate [14] and the use of Moiré fringes [15]. These techniques were developed for applications where the gap between components is typically less than a few 100's of microns.

In situations where components are separated by many millimeters, other optical techniques must be employed. One simple approach consists of focusing an alignment microscope back and forth between the alignment planes and using fiducial markers on both components to perform the alignment [16]. This technique requires high-precision micropositioning equipment because the alignment accuracy relies on the microscope moving along a line exactly perpendicular to the alignment planes. The method is also very tedious and cannot be readily automated. The technique is significantly improved by introducing a beamsplitting element halfway between the alignment planes, allowing both planes to be imaged simultaneously with a static microscope. In one implementation of this approach, the components are moved away from each other, leaving enough space for a beamsplitter assembly to be placed halfway between the two [17][18]. Once both images are registered, the beamsplitter assembly is removed and the components are carefully brought in contact. Although this technique avoids the problem of having to focus a microscope back and forth between the alignment planes, it requires high-quality micropositioners and cannot detect misalignment errors that might occur during the last steps of the assembly process when both components are fixed together. This difficulty can be resolved by permanently inserting a partially reflecting surface halfway between the components [19]. Although this approach does not suffer from the drawbacks of the previous method, it requires an additional component may also introduce optical losses.

An alternative strategy relies on the use of dedicated Fresnel lenses fabricated directly on the array components. One technique uses identical Fresnel lenses on both components, with focal lengths equal to half the component separation [20]. The lenses are purposely designed such that a significant fraction of the incident light be diffracted into the zeroth order; this is easily achieved by fabricating the lenses with the wrong etch depth. The technique is based on the monitoring of interference fringes generated by the overlapping zeroth and first diffraction orders at the output. In the implementation described in [20], it achieved a lateral sensitivity of $\pm 3 \ \mu m$. This technique was used for the assembly of the BCM and relay modules of the photonic backplane demonstrator [21].

Another technique uses reflective on-axis circular FZPs on one component with alignment targets on the other [22]. In this case, the focal length of the FZPs is equal to the distance between components. Alignment is obtained by registering the spots generated by the reflective FZPs with the alignment targets. This method was demonstrated with a lateral sensitivity of $\pm 9 \mu m$ [22].

A significant advantage of the last two techniques is that they allow for the state of alignment to be monitored as the components are attached to one another. This is very convenient because it allows for misalignment errors (e.g. misalignment during glue curing due to glue shrinkage) to be detected and compensated for. While these techniques can provide accurate lateral and rotational alignment information, their sensitivity to longitudinal and tilt misalignment is limited by the *f*-number of the Fresnel lenses. For example, the sensitivity to longitudinal misalignment of the techniques demonstrated in [20] and [22] was of the order of \pm 50 µm, making them inappropriate for applications requiring accurate alignment in the longitudinal and tilt DOFs.

5.3 Problem definition: aligning a lens array to a OE-VLSI chip

The misalignment tolerance analysis of section 4.2.1 showed that integrating a minilens array with the OE-VLSI chip improved the lateral, rotational and longitudinal tolerances of the chip module at the expense of tighter tilt tolerances. The discussion of chapter 7 will formalize this classical trade-off between lateral and tilt tolerances; it will also be shown that a proper balance between lateral and tilt tolerances (a requirement for misalignment-tolerant systems) can only be achieved by integrating lens array components with the OE-VLSI chip. There is thus a need for cost-effective and accurate intra-module alignment techniques for aligning lens arrays with OE-VLSI chips.



Figure 5.2. Six-DOF alignment of a mini-lens array to an OE-VLSI chip.

Chapter 5: Intra-module alignment techniques

The following sections present two novel intra-module alignment techniques designed for the assembly of the chip module. The alignment problem they each try to solve is described in figure 5.2: an 8×8 array of diffractive mini-lenses fabricated on a $14 \times 14 \text{ mm}^2$ fused silica substrate is required to be aligned to a 32×32 array of GaAs MQW modulators flip-chip on a $9 \times 9 \text{ mm}^2$ OE-VLSI chip. The mini-lens array must be aligned in all six DOFs ($x, y, z, \theta_x, \theta_y, \theta_z$). The tolerances associated with each DOF are included in figure 5.2; they are the same as the tolerances found in the last column of table 4.1. This is because the tolerances of table 4.1 were calculated for the case of the chip being misaligned with respect to a perfectly aligned optical system, which is equivalent to a chip module having some misalignment between the chip and the mini-lens array.

Note that the techniques to be presented are not limited to the specific problem of aligning a lens array to a chip; they are general and can be used in a variety of other situations. Both alignment techniques use off-axis Fresnel lenses on one component to generate optical alignment features in the plane of the other. These techniques differentiate themselves from previous optical methods in that they are sensitive to all six DOFs and amenable to automated assembly.

5.4 Technique #1: off-axis Fresnel lenses and on-chip quadrant detectors

The first alignment technique is presented in figure 5.3 and works as follows. Four collimated beams are incident orthogonal to off-axis Fresnel lenses located on the periphery of the mini-lens substrate. As shown in figure 5.4, an off-axis Fresnel lens represents a portion of a larger on-axis Fresnel lens and focuses an incident beam at an off-axis angle ϕ . The focal length of the off-axis Fresnel lenses is chosen to be equal to that of the mini-lenses, $f_{mini} = 8.50$ mm.

The off-axis Fresnel lenses focus the incident beams on quadrant detectors located on the OE-VLSI chip. In our implementation, the quadrant detectors are silicon photodiodes fabricated using the 1.5 μ m CMOS process; details of the photodiode design are given in section 5.4.3. Because the off-axis lenses and quadrant detectors are fabricated using microlithographic techniques, their position is defined to micron precision. The design is such that under conditions of perfect alignment, each beam is focused exactly at the center of its respective quadrant detectors. An essential feature of this alignment technique is the off-axis operation of the Fresnel lenses, resulting in an increased sensitivity to tilt and lon-gitudinal misalignment.





tively. Each element has dimensions $d \times d$. Elements are separated by a small gap, g, whose minimum size is determined by the design rules of the CMOS process.



Figure 5.6. Representation of quadrant detector *i* and its relevant parameters.

It is assumed that the incident beams are focused into spots with a Gaussian intensity distribution of radius ω and that the detector elements have equal and uniform responsivities. It is useful to define the following functions:

$$M_{xi} = \frac{(I_{bi} + I_{ci}) - (I_{ai} + I_{di})}{I_{ai} + I_{bi} + I_{ci} + I_{di}}$$
(5.1)

$$M_{yi} = \frac{(I_{ai} + I_{bi}) - (I_{ci} + I_{di})}{I_{ai} + I_{bi} + I_{ci} + I_{di}}$$
(5.2)

Using the parameters defined in figure 5.6, it can be shown that (see appendix B):

$$M_{xi}(u_{xi}) = \frac{erf\left\{\frac{g/2 + d - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - u_{xi}}{\omega/\sqrt{2}}\right\} + erf\left\{\frac{g/2 + u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 + d + u_{xi}}{\omega/\sqrt{2}}\right\}}{erf\left\{\frac{g/2 + d - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 + u_{xi}}{\omega/\sqrt{2}}\right\} + erf\left\{\frac{g/2 + d + u_{xi}}{\omega/\sqrt{2}}\right\}}$$
(5.3)

$$M_{yi}(v_{yi}) = \frac{erf\left\{\frac{g/2 + d - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - v_{yi}}{\omega/\sqrt{2}}\right\} + erf\left\{\frac{g/2 + v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 + d + v_{yi}}{\omega/\sqrt{2}}\right\}}{erf\left\{\frac{g/2 + d - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 + v_{yi}}{\omega/\sqrt{2}}\right\} + erf\left\{\frac{g/2 + d + v_{yi}}{\omega/\sqrt{2}}\right\}}$$
(5.4)

Chapter 5: Intra-module alignment techniques

The final step consists of correcting for chip misalignment. This is achieved by mounting the chip on a six-DOF micropositioning stage. The best approach at implementing the alignment system is to use an analog-to-digital (A-to-D) converter card to sample the photocurrent signals and a computer to calculate the six-DOF misalignment of the chip. Furthermore, by controlling both the A-to-D converter and the micropositioning stage with the computer, it becomes possible to operate the alignment system as a closed feedback loop, opening the way for high-accuracy automated alignment in all six DOFs.

Considering that the accuracy of a quadrant detector improves as the beam gets closer to its center, the sequence of (i) measuring photocurrents, (ii) calculating chip misalignment and (iii) correcting for chip misalignment, is required to be iterated a few times for the alignment to be optimized.

5.4.3 Design of CMOS-compatible quadrant detectors

Four different types of vertical photodiodes can be implemented in a standard p-well CMOS process and they are shown in figure 5.7. The photodiodes of figure 5.7(a), (b) and (c) are created by forming a reverse-biased junction between a p-doped region and the n-type substrate. The structure of figure 5.7(d) is different; it comprises of two photodiodes operating in parallel, one formed between the n+ implant and the p-well, the other between the p-well and the n-type substrate.



Figure 5.7. Four different types of CMOS-compatible vertical photodiodes.

Note that lateral photodiodes are also possible in a CMOS process but were not considered here because of their limited photodetection area, making them poor candidates for their use as large-area alignment detectors.

Electrical and optical properties (dark current, capacitance, responsivity spectrum) of vertical CMOS-compatible photodiodes have been characterized by Aubert et al. [27]. The desirable attributes of a large-area alignment detector are a high responsivity and a low surface resistivity. Capacitance and dark current are not critical in this case because measurements are performed at low speeds and the signal photocurrent is many orders of magnitude larger than the dark current. Considering the long absorption length in silicon at 850 nm (~14 μ m [28]), a high responsivity requires a thick depletion region. The thickness of the depletion region is maximized by forming a junction between regions of low doping levels that extend as deep as possible into the substrate. In a CMOS process, the best way of achieving this is to form a junction between the *p*-well and the *n*-type substrate. For this reason, the photodiode structure of figure 5.7(c), formed between a shallow *p*+ implant and the *n*-type substrate, was promptly discarded due to its poor responsivity. The responsivities of the remaining three structures are approximately equal [27].



Figure 5.8. A quadrant detector implemented in a standard 1.5 µm CMOS process.

For large-area detectors, the surface resistivity of the anode contact must be minimized. This is required to facilitate the collection of charges generated far from the contact. Surface resistivity can be reduced by placing a p+ implant across the surface of the pwell. This is the reason why the photodiode structure of figure 5.7(b) was preferred over the other designs. A photograph and layout of the quadrant detector used in our alignment system is shown in figure 5.8.

To minimize the possibility of crosstalk between adjacent detector elements, distinct *p*-wells are used, separated by a *n*+ implant. Anode contacts are located at the far corner of each element, away from the detector center, avoiding shadowing effects as the spot converges towards perfect alignment. Parameters *d* and *g*, introduced in figure 5.6, are specified by the layout of the detector elements: *d* corresponds to the width of the *p*-well region and *g* corresponds to the width of the central metal contact. In our design, $d = 139.1 \,\mu\text{m}$ and $g = 10.9 \,\mu\text{m}$.

An important design detail concerns the presence of a silicide mask in the CMOS process [29]. As CMOS technology proceeds further into the deep sub-micron regime, polysilicon traces are increasingly being silicided to reduce their resistance. This causes a problem for CMOS detectors because silicided regions are opaque. The 1.5 μ m CMOS technology used in our design was non-silicided and so this was not issue. For silicided CMOS processes, this problem can be avoided by protecting the detector area by specifying appropriate blocking regions in the silicide mask.

5.4.4 Responsivity measurements

Detector responsivity was initially measured at 852 nm using a large Gaussian spot size $(3\omega_{spot} \sim 100 \ \mu\text{m})$, enough to cover most of the area of a detector element. A 5 V reverse-bias was applied to the diode. Measurements performed on multiple detector elements yielded an average responsivity of 0.15 ± 0.02 A/W.

Due to the large spot size used, the previous measurement was not sensitive to nonuniformities across the area of the detector. To characterize the detector uniformity, a measurement setup was constructed that allowed for a small spot size $(3\omega_{spot} \sim 21 \ \mu\text{m})$ to be scanned across the surface of the detector, following a high-resolution 2D grid. At each point on the grid, the incident optical power and the generated photocurrent on all four elements were measured. This process was automated using the GPIB capabilities of the optical power meter and micropositioning stage. Photocurrent signals were amplified and sampled using a 16-bit A-to-D card inserted in a computer. The measurement setup was automated using Labview.

A typical responsivity scan appears in figure 5.9. The responsivity is not uniform, with more than 50% variations across the surface of an element. Multiple detectors were characterized; in all cases, a responsivity profile similar to that of figure 5.9 was found.



Figure 5.9. Responsivity map of a four-element CMOS detector.

5.4.5 Discussion

The main motivation behind the use of on-chip quadrant detectors was the ease with which the alignment setup could be controlled by a computer. Furthermore, the alignment system can readily be operated as a closed feedback loop, with the computer also controlling the positioning stage holding the chip, resulting high-accuracy automated alignment in all six DOFs.

Unfortunately, the non-uniformities found in the detector responsivity were fatal to the proper operation of the alignment technique. The mathematical derivation of appendix B,

ments: (i) the use of linear FZPs (instead of circular FZPs), resulting in a more efficient use of silicon area and (ii) the use of off-axis FZPs (instead of on-axis FZPs), offering increased sensitivity in the longitudinal and tilt DOFs.



Figure 5.10. Description of alignment technique #2.

The alignment technique works as follows. A wide, collimated, and monochromatic beam is incident orthogonal to the OE-VLSI chip. The portion of the incident beam falling outside the region occupied by mini-lenses traverses the fused silica substrate and illuminates reflective off-axis linear FZPs located on the periphery of the OE-VLSI chip. The focal length of FZPs is equal to the focal length of the mini-lenses: $f_{mini} = 8.50$ mm.

The FZPs are fabricated using the top metal layer of the CMOS process, as shown in figure 5.11. The use of linear FZPs, as opposed to circular FZPs offer two significant advantages. First, linear FZPs are easier to design and their rectangular geometries are readily compatible with commercial integrated circuit layout CAD tools. Secondly, linear FZPs consume much less silicon area. For example, the FZP in our design could be made as small as 100 μ m × 1000 μ m. Assuming a pair of FZPs at each corner of the 9 × 9 mm²

chip, the area consumed by the FZPs represents less than 1% of the total silicon area. In comparison, four on-axis circular FZPs with the same *f*-number would consume 3.9% of the chip area, about four times more.



Figure 5.11. Photographs of the off-axis FZP fabricated on the CMOS chip.

A single linear FZP focuses the incident plane wave into a line; a pair of off-axis linear FZPs oriented at 90° to each other is used to generate a cross pattern in the focal plane. Due to the off-axis operation of the FZPs, the trajectory traversed by the center point of a cross pattern does not follow a line perpendicular to the chip. This means that the center point of the cross pattern shifts laterally as it is imaged at different planes along the optical axis, as shown in figure 5.12.

In the implementation of figure 5.10, one FZP pair is placed in every corner of the OE-VLSI chip. Chrome alignment targets, in the form of inverted crosses, are deposited on the mini-lens array substrate; these are lithographically-defined with respect to the minilenses. Six-DOF alignment is achieved by positioning the mini-lens array substrate such that all chrome targets are registered with the focused cross patterns. Note that only three FZP pairs are necessary; the fourth pair is redundant. This is because the center of three cross patterns in the focal plane uniquely specify the position and orientation (i.e. all six DOFs) of the mini-lens substrate with respect to the chip.



Figure 5.12. CCD images of the cross pattern at 3 different planes along the optical axis.

5.5.1 Sources of errors

Three sources of error limit the accuracy of this technique. The most important error is due to the incident beam not being perfectly orthogonal to the chip, resulting in the lateral shift of the optical crosses in the plane of the mini-lens array. We define the angle between the incident beam and the chip to be 90°- β . Techniques developed to minimize the value of β are presented in sections 5.5.4 and 5.5.5.

The second contribution is due to the error in judging when an optical cross is properly centered on its alignment target. This decentration error is defined by a quantity δ , expressed in microns. The value of δ depends on various features of the experimental setup, including the magnification of the imaging system, the width of the focal lines making up the cross patterns, the resolution of the CCD camera, the mechanical stability of the positioning equipment and the use of image processing software. The value of δ for the setup described in sections 5.5.4 and 5.5.5, which use a ×50 microscope objective, was estimated to be 2 µm. Finally, the third source of error is due to the finite accuracy of the microlithography processes that are used to fabricate the FZPs on the chip and the alignment targets on the mini-lens substrate. This last contribution is usually negligible compared to the previous two and so it is discarded in what follows.

5.5.2 Design considerations

The longitudinal and tilt alignment sensitivity improve as the off-axis angle, ϕ , increases. The off-axis angle is defined in figure 5.10; it corresponds to the angle between a vector normal to the chip and the vector connecting the center of the FZP to the center of its corresponding alignment target when perfectly aligned. The off-axis angle can be increased in two ways: (i) by increasing the diffraction angle on the outside edge of the FZP (θ_{edge}) by adding high-order zones and (ii) by removing low-order zones on the inside of the FZP. The maximum value of θ_{edge} is limited by the minimum feature size allowable by the fabrication technology. In our implementation, this limitation was dictated by the CMOS design rule for the top metal layer that specifies a minimum metal width of 2.0 µm and a minimum spacing of 2.0 µm. The relationship between θ_{edge} and the grating period, *p*, is specified by the grating equation [30]:

$$\sin(\theta_{edge}) = \frac{m\lambda}{p}$$
(5.6)

where *m* is the diffraction order and λ is wavelength. Assuming m = 1, $\lambda = 0.85 \,\mu\text{m}$ and $p = 4.0 \,\mu\text{m}$, this results in a maximum edge diffraction angle of $\theta_{edge} = 12.3^{\circ}$.

The removal of low-order zones increases the value of ϕ but reduces the FZP aperture; this increases the width of the focal lines forming the cross patterns, resulting in an increase of δ . This leads to the following design compromise. On the one hand, too small a value of ϕ yields a poor sensitivity to longitudinal (z) and tilt (θ_x , θ_y) misalignments. On the other hand, increasing the value of ϕ by removing low-order zones increases δ and eventually decreases the sensitivity in the lateral (x, y) and rotational (θ_z) DOFs. Determining the optimal value of ϕ using an analytic approach is difficult to do in general because it requires the prior knowledge of the relationship between ϕ and δ , which is specific to the experimental setup.

5.5.3 Experimental setup requirements

The experimental setup must fulfill the following set of objectives: generate a wide, collimated and monochromatic incident beam, minimize the value of β (defined in 5.5.1),

provide a high-magnification image of the optical crosses and the alignment markers, and provide accurate and stable 6-DOF positioning of the mini-lens array substrate.

Two techniques were developed to achieve the above; they differ in their approach at minimizing the value of β . The first approach (section 5.5.4) is derived from the technique used in [31]; it uses the retroreflection of an incident wave from the surface of the chip to determine and minimize the value of β . The second approach (section 5.5.5) is novel and uses *in situ* diffractive structures fabricated directly on the mini-lens substrate to control the orthogonality of the incident wave.

5.5.4 Retroreflected beam alignment technique

This technique is described in figure 5.13. The optical train generates a wide beam that is collimated using a lens of focal length f_c . First, the chip is inserted and secured in place (the mini-lens substrate is not present at this point). As shown by the dotted lines in figure 5.13, the beam reflected from the chip is refocused by the lens and can be observed (using CCD camera #1) at the front focal plane where a thin pellicle has been inserted. The pellicle is partially reflective; this allows for the retroreflected spot to be observed at the same time as the spot incident in the opposite direction. The pellicle is a few microns thick; this avoids undesirable ghost images of the spots. The use of a pellicle is preferred to the pinhole used in [31] because it allows for the retroreflected spot to remain visible even under conditions of large beam misalignments.



Figure 5.13. Experimental setup using the retroreflected beam alignment technique.

Next, Risley beam steerers are rotated until the retroreflected spot overlaps with the incident spot at the pellicle plane. At this point, the incident beam is orthogonal to the chip and the value of β is minimized. The worst-case value of β is given by:

$$\beta = \frac{1}{2} \operatorname{atan}\left(\frac{\varepsilon}{f_c}\right) \cong \frac{\varepsilon}{2f_c}$$
(5.7)

where ε corresponds to the decentration error involved in judging the alignment of the retroreflected spot relative to the incident spot. Equation (5.7) remains valid regardless of the distance separating the collimating lens and the chip (i.e. the chip is not required to be one focal length away from the collimating lens). Next, the mini-lens substrate is attached to a 6-DOF micropositioning stage and inserted in front of the chip. Final alignment is obtained by registering the alignment targets of the mini-lens substrate to the optical crosses generated by the on-chip FZPs using CCD camera #2.

An important limitation of the retroreflection technique is that it relies on the surface of the chip to be optically flat, which is not necessarily the case with standard CMOS chips. Furthermore, the addition of flip-chip OE devices on the CMOS chip further degrades the surface flatness. This translates into an abberated retroreflected spot, leading to a large value of ε . To demonstrate this, the setup of figure 5.13 was constructed using a collimating lens with $f_c = 100$ mm and the OE-VLSI chip of figure 3.10. A photograph of the retroreflected spot in the plane of the pellicle appears in figure 5.14(a), showing a large amount of aberrations and a spot diameter of ~120 µm.



Figure 5.14. Retroreflected spot from (a) the OE-VLSI chip and (b) a 100% mirror.

Taking the value of ε to be equal to half the spot diameter, equation (5.7) leads to a worst-case $\beta = 0.017^{\circ}$. Next, the OE-VLSI chip was replaced with a 100% dielectric mir-

ror with an optical flatness of $\lambda/10$. The resulting retroreflected spot, shown in figure 5.14(b), is much smaller with a diameter of ~20 µm, leading to a worst-case $\beta = 0.003^{\circ}$.

This serves to illustrate how poor chip flatness can significantly degrade the accuracy of the retroreflection technique. Although the accuracy can always be improved by further increasing the focal length f_{cr} this can only be done at the expense of a bulkier setup.

5.5.5 In Situ beam alignment technique

A novel alignment technique was developed to address the limitations of the previous method. A diagram of the optical train is shown in figure 5.15. The technique works as follows.



Figure 5.15. Experimental setup using the *in situ* beam alignment technique.

First, the mini-lens substrate and the chip are both inserted in the alignment setup. The chip is secured in place and the mini-lens is attached to a 6-DOF micropositioning stage. The next step consists of removing tilt misalignment between the mini-lens substrate and the chip. Tilt misalignment can be observed using the interference fringes produced by partial reflections from the mini-lens substrate and the chip surface, as shown in figure 5.16. This requires the use of a laser source with a coherence length that is at least $2 \times f_{mini-1}$

Interference fringes are difficult to observe if the reflectivity of the mini-lens substrate is low, which is especially true if the substrate has been anti-reflection coated at the operating wavelength of the OE devices. To avoid this problem and increase the fringe contrast, it is often necessary to design and operate all alignment features at a different wavelength (e.g. HeNe laser at 633 nm).



Figure 5.16. Interference fringes between the mini-lens substrate and the chip surface, due to tilt misalignment. The upper-left corner of the chip is shown.

The amount of tilt misalignment $(\Delta \theta_x, \Delta \theta_y)$ is determined by measuring the fringe separation; this can be conveniently done by counting the number of fringes, N, over a distance L along a line perpendicular to the fringe pattern. The tilt misalignment (about an axis parallel to the fringe pattern) is given by:

$$\Delta \theta = \operatorname{atan}\left(\frac{N\lambda}{2L}\right) \cong \frac{N\lambda}{2L}$$
(5.8)

For example, the fringe pattern in figure 5.16 shows 12 fringes (oriented at approximately 45° in the figure) over a distance of 2.2 mm, which translates in a tilt misalignment of 0.13° at $\lambda = 850$ nm. By carefully adjusting the tilt of the mini-lens substrate, the number of fringes was brought down to 2 over the same distance, leading to a worst-case tilt misalignment of 0.022°.

Next, the incident beam is aligned orthogonal to the chip. This is achieved using *in situ* linear FZPs fabricated directly on the mini-lens substrate, as shown in figure 5.17. The focal length of the FZPs is equal to $2 \times f_{mini}$. As a result, the incident beam is focused back into a line towards the center of the FZP. Using this scheme, any angular deviation between the incident beam and the chip results in the decentration of the focal line relative to the FZP center. Risley beam steerers are used to center the focal line on the FZP. To facilitate this adjustment, metal alignment targets are deposited in place of the central

Fresnel zone. As shown in figure 5.17, two linear FZPs, oriented at 90° to each other, are required to perform θ_x and θ_y beam angular alignment.



Figure 5.17. Orthogonal beam alignment using in situ off-axis linear FZPs.

The worst-case angular deviation between the incident beam and chip is given by:

$$\beta = \operatorname{atan}\left(\frac{\delta}{2f_{mini}}\right) \cong \frac{\delta}{2f_{mini}}$$
(5.9)

where δ corresponds to the error made in judging when the focal line is properly centered on the alignment targets. For the experimental setup of figure 5.15, it was found that $\delta = 2 \ \mu m$ and $f_{mini} = 8.50 \ mm$, which results in $\beta = 0.0067^{\circ}$. This is to be compared with $\beta = 0.017^{\circ}$ obtained using the retroreflection technique and with $\beta = 0.09^{\circ}$ obtained in [22]. Comparison of equations (5.7) and (5.9) shows that the *in situ* technique is more accurate than the retroreflection technique if $\delta/f_{mini} < \varepsilon/f_c$, which may be the case with a textured or non-planar chip surface.

A significant advantage of the *in situ* technique is its ease of use; it allows for all of the misalignment information (angular deviation of the incident beam and the 6-DOF mis-

alignment of the mini-lenses) to be simultaneously monitored, in the same observation plane, using a single CCD camera. Additionally, the *in situ* technique is more compact because it does not require the use of a long focal length collimating lens.

5.5.6 Worst-case misalignment errors

General closed-form expressions for the worst-case misalignment errors are derived as follows. We start by assuming a perfectly aligned mini-lens array, with an incident collimated beam making an angle 90° - β with the surface of the chip. For each DOF, the worst-case misalignment error is obtained by determining the maximum misalignment of the mini-lens substrate, in that DOF, which results in optical crosses being decentered by an amount equal to δ . Given the above, the worst-case lateral error (Δx , Δy) corresponds to the mini-lens array being laterally misaligned by an amount:

$$\Delta x = \Delta y = \delta + f_{mini} \tan(\beta)$$
(5.10)

The rotational error $(\Delta \theta_z)$ is derived by considering the rotational angle required for optical crosses located at opposite corners of the mini-lens array to be decentered by an amount δ :

$$\Delta \Theta_z = \operatorname{atan}\left(\frac{\sqrt{2}\delta}{d_t}\right) \cong \frac{\sqrt{2}\delta}{d_t}$$
(5.11)

where d_t corresponds to the center-to-center distance between a pair of alignment targets located on the same side of the mini-lens substrate. The longitudinal error (Δz) is given by the axial displacement required for optical crosses to be decentered by an amount less or equal to δ :

$$\Delta z = \frac{\delta}{\tan(\phi - \beta)} \cong \frac{\delta}{\tan\phi}$$
(5.12)

where ϕ is the off-axis angle of the linear FZPs, defined in section 5.5.1. In practice, a large value of ϕ is desirable because it reduces Δz and so we can assume $\phi >> \beta$.

The tilt misalignment error $(\Delta \theta_x, \Delta \theta_y)$ corresponds to alignment targets on opposite sides of the mini-lens substrate being defocused by an amount Δz in opposite directions:

$$\Delta \theta_x = \Delta \theta_y = \operatorname{atan}\left(\frac{\Delta z}{d_t/2}\right) \cong \frac{2\delta}{d_t \tan \phi}$$
(5.13)

As indicated by equations (5.12) and (5.13), the larger the off-axis angle, the more sensitive the technique to longitudinal and tilt misalignments.

5.5.7 Experimental results for the accuracy of technique #2

Given the value of β , equations (5.10)-(5.13) are used to calculate the worst-case misalignment error between the mini-lens array and the OE-VLSI chip. In our design, the separation between the mini-lenses and the chip is $f_{mini} = 8.50$ mm, and the center-to-center distance between a pair alignment targets on the same side of the mini-lens array substrate is $d_t = 7150 \,\mu$ m. The FZPs were designed with an off-axis angle $\phi = 8.90^\circ$. The value of δ was determined experimentally to be 2 μ m under × 50 magnification. The worst-case error is calculated for both orthogonal beam alignment techniques and the results are compared in table 5.1. The techniques differ mainly in their sensitivity to tilt misalignments, the use of interference fringes improving the tilt sensitivity by an order of magnitude in this case.

Worst-Case Misalignment Error	Retroreflected Beam Alignment Technique ($\delta = 2 \ \mu m, \epsilon = 50 \ \mu m$)	<i>In Situ</i> Beam Alignment Technique (δ = 2 μm)		
Beam Deviation (β)	0.017°	0.0067°		
Lateral (Δx , Δy)	4.5 μm	3.0 μm		
Rotational ($\Delta \theta_z$)	0.023°	0.023°		
Longitudinal (Δz)	13 µm	13 µm		
Tilt $(\Delta \Theta_x, \Delta \Theta_y)$	0.20°	0.022°		

m 11 / 1	337 .	• • •	C 1 .1	1 1*	
	WATCH AACA	micalianmant arr	へゃぐ すべゃ やべすれ	haam alla	iment techniquiec
	W UISL-LASE	IIIISAIIPIIIIICIR CH	ons ioi doun	DCAIL AILY	michi icimiucs.

The results of table 5.1 are to be compared with the tolerance budget listed in figure 5.2 and repeated here for convenience: Δx , $\Delta y = 8.0 \ \mu m$, $\Delta \theta_z = 0.10^\circ$, $\Delta z = 125 \ \mu m$ and $\Delta \theta_x$, $\Delta \theta_y = 0.12^\circ$. The results indicate that the *in situ* beam alignment easily satisfy all requirements of the tolerance budget. The *in situ* beam alignment technique is also preferred because of its ease of use and compactness.

5.5.8 Improved implementation of technique #2

The implementation of figure 5.10 was found to be awkward to use in practice because no alignment information could be inferred from measurements performed on a single alignment target. For example, the decentration of a cross pattern may be caused by a lateral misalignment, a longitudinal misalignment, or a combination of both. This ambiguity could only be resolved by measuring the decentration of another cross pattern, located at the other corner of the chip. The consequence of this is that the operator is forced to travel back and forth between two alignment targets to determine the misalignment in one DOF. This inconvenience has led to the proposal of an alternative implementation which is shown in figure 5.18.



Figure 5.18. An improved implementation of alignment technique #2.

This implementation uses pairs of off-axis linear FZPs on each side of the chip combined with metal alignment features fabricated on the mini-lens substrate. As before, a wide and collimated beam is incident orthogonal to the on-chip FZPs and the light is reflected from the FZPs and focused into lines in the plane of the mini-lens array. In this case, the alignment features are designed as metal rulers and are used to measure the position and separation between two parallel focal lines: the separation between focal lines is a measure of longitudinal misalignment whereas their position relative to the center of the metal ruler is a measure of lateral misalignment. The FZPs are designed such that perfect alignment of a metal ruler corresponds to a pair of focal lines being separated by a distance D and symmetrically positioned about the center of the metal ruler. The advantage of this type of layout is that both lateral and longitudinal misalignment information can be directly inferred from measurements performed on a single alignment feature. Defining x^+ and x^- to be the distance separating each focal lines from the center of the horizontal metal rulers (see inset of figure 5.18), the lateral (Δx) and longitudinal (Δz) misalignments of the metal ruler are calculated as follows:

$$\Delta x = x^{+} - x^{-} \tag{5.14}$$

$$\Delta z = \frac{x^+ + x^- - D}{2\tan\phi} \tag{5.15}$$

where ϕ is the off-axis angle defined earlier in section 5.5.2. Using the definition of equation (5.15) and the implementation of figure 5.18, a positive value of Δz corresponds to the mini-lens substrate being too close to the chip. Note that the lateral misalignment information from two alignment rulers located on opposite sides of the mini-lens array can be used to determine the rotational misalignment ($\Delta \theta_z$) of the mini-lens. Similarly, the longitudinal misalignment information from three alignment rulers is sufficient to determine the tilt misalignment ($\Delta \theta_x$, $\Delta \theta_y$) of the mini-lens substrate. Consequently, three alignment rulers, combined with three pairs of off-axis FZPs on the chip, are sufficient to determine the mini-lenses misalignment in all six DOFs.

5.5.9 Automating alignment technique #2 using on-chip detectors

Assuming that on-chip detectors with uniform responsivity are available, the implementation of figure 5.18 can be automated by placing the off-axis FZPs on the mini-lens substrate and position detectors on the chip, as shown in figure 5.19. Additional benefits to this approach are that (i) quadrant detectors can be made compact and thus consume less silicon area than FZPs and (ii) FZPs implemented as multiple phase level lenses on the mini-lens substrate offer a higher diffraction efficiency.



Figure 5.19. Novel implementation using on-chip quadrant detectors.

In the implementation of figure 5.19, the detectors are designed with elements that are triangular in shape. Each pair of off-axis FZPs on the mini-lens substrate produces two focal lines in the plane of chip, one line falling on elements a and b, the other on elements c and d. All four elements produces equal photocurrents when each focal line is centered on its respective pair of detector elements; this corresponds to focal lines being separated by a distance D and centered on the four-element detector. The four photocurrents (I_a , I_b , I_c , I_d) are used to calculate the lateral (Δx) and longitudinal (Δz) misalignments of that four-element detector relative to the mini-lenses:

$$\Delta x = \left(\frac{I_b + I_d - I_a - I_c}{I_a + I_b + I_c + I_d}\right) \times D$$
(5.16)

$$\Delta z = \left(\frac{I_b + I_d - I_a - I_c}{I_a + I_b + I_c + I_d}\right) \times \left(\frac{D}{2\tan\phi}\right)$$
(5.17)

Using previous arguments, it can be shown that three four-element detectors, a total of 12 photocurrents, are sufficient to determine the chip misalignment in all six DOFs.

Careful examination of figure 5.19 reveal that it is very similar to alignment technique #1 (figure 5.3) but with the following significant improvements. First, linear off-axis FZPs are easier to design than circular off-axis Fresnel lenses and consume much less area. Secondly, the use of a pair of focal lines on four triangular-shaped elements, as opposed to a small focal spot on four square-shaped elements, results in a more robust measurement technique. To see this, consider figure 5.20 which compares both types of detectors under conditions of perfect alignment. For the detector with square-shaped elements, most of the incident power is lost to the gap between elements; thus, as the spot converges towards the center of the quadrant detector, the signal-to-noise ratio deteriorates. This problem can be reduced by increasing spot size, but this is done at the expense of accuracy [26]. This signal-to-noise problem is not present if triangular-shaped elements are used because only a small amount of light is lost to the gap, irrespective of the focal line position. Furthermore, the use of focal lines, which extend over a larger area of the detector, make the measurement less sensitive to non-uniformities in responsivity.



Figure 5.20. Comparing two types of four-element detectors.

Finally, the use of detectors with triangular-shaped elements dramatically simplifies the calculation of the six-DOF misalignment of the chip, avoiding the intricacies of appendix B. There are two reasons for this simplification. First, the position of the focal lines can be solved explicitly from the photocurrent signals (see equations (5.16) and (5.17)), whereas a numerical solution was required in the case of quadrant detectors (see equations (5.3) and (5.4)), which prompted the use of look-up tables. More importantly, the use of focal lines whose length extends beyond the area of the triangular-shaped elements allows for DOFs to be decoupled from each other. For example, the photocurrents of the triangu-

= 13 µm, $(\Delta \theta_x, \Delta \theta_y) = 0.022^\circ$, $\Delta \theta_z = 0.023^\circ$. This satisfies the requirements of the misalignment tolerance budget listed in figure 5.2.

5.7 Acknowledgments

The following individuals contributed to various aspects of the development and implementation of the alignment techniques described in this chapter. Stéphane Ménard assisted for the implementation of alignment technique #1; he wrote a C++ program which interfaced with an A-to-D board and automatically calculated the six-DOF misalignment of the chip based on the mathematical derivation of appendix B. Julien Faucher and Robert Varano assisted in collecting the responsivity data of figure 5.9: Julien Faucher designed and implemented the photocurrent amplifying electronics, while Robert Varano was responsible for automating the responsivity measurement in Labview. Marc Château-neuf helped with the layout of the circular off-axis Fresnel lenses of figure 5.3. The design and layout of the quadrant detector of figure 5.8 was done with David Rolston.

5.8 References

- R. Boudreau, "Passive optical alignment methods," in Proceedings 3rd International Symposium on Advanced Packaging Materials Processes, Properties and Interfaces (Cat. no.97TH8263)(IEEE 1997, New York, 1997), pp. 180-181.
- [2] N. C. Craft, A. Y. Feldblum, "Optical interconnects based on arrays of surface-emitting lasers and lenslets," Applied Optics, vol. 31, pp. 1735-1739 (1992).
- [3] E. Bisaillon, T. Yamamoto, D. F.-Brosseau, E. Bernier, D. Godwill, A. G. Kirk and D. V. Plant, "Optical link for an adaptive redundant free-space interconnect," in Optics in Computing 2000, R. A. Lessard and T. Galstian eds., SPIE 4089, pp. 999-1009 (2000).
- [4] P. Sheer, T. Collette and P. Churoux, "Free-space optical interconnection within SIMD massively parallel computers," in Proceedings of the Fourth International Conference on Massively Parallel Processing using Optical Interconnects, J. Goodman, S. Hinton, T. Pinkston, E. Schenfeld, eds. (IEEE Computer Society, New York, 1997), pp. 167-177.

- [5] J. Jahns, R. A. Morgan, H. N. Nguyen, J. A. Walker, S. J. Walker and Y. M. Wong, "Hybrid integration of surface-emitting microlaser chip and planar optics substrate for interconnection applications," IEEE Photonics Technology Letters, vol. 4, pp. 1369-1372 (1992).
- [6] N. R. Basavanhally, M. F. Brady and D. B. Buchholz, "Optoelectronic packaging of two-dimensional surface active devices," IEEE Transactions on Components, Packaging, and Manufacturing Technology Part B, vol. 19, pp. 107-115 (1996).
- [7] L. S. Goldmann, "Self-alignment capability of controlled-collapse chip joining," Proceedings 22nd Electronic Comp. Conf., pp. 332 (1972).
- [8] N. R. Basavanhally, "Opto-mechanical alignment and assembly of 2D-array components," Sarnoff Symposium, 1993 IEEE Princeton Section, pp. 23-27 (1993).
- [9] R. A. Boudreau, H. Han, M. Kadar-Kallen, J. R. Rowlette, "Kinematic mounting of optical and optoelectronic elements on silicon waferboard," US Patent 5,574,561 (Nov. 12 1996).
- [10] C. Gonzales, R. J. Welty, R. L. Smith and S. D. Collins, "Microjoinery for optomechanical systems," in SPIE vol. 3008, pp. 171-178 (1997).
- [11] D. Miyazaki, S. Masuda, K. Matsushita, "Self-alignment with optical microconnectors for free-space optical interconnections," Applied Optics, vol. 37, pp. 228-232 (1998).
- [12] V. Baukens, G. Verschaffelt, P. Tuteleers, P. Vynck, H. Ottevaere, M. Kufner, S. Kufner, I. Veretennicoff, R. Bockstaele, A. Van Hove, B. Dhoedt, R. Baets, H. Thienpont, "Performances of optical multi-chip-module interconnects: comparing guided-wave and free-space pathways," J. Opt. A: Pure Appl. Opt., vol. 1, pp. 255-261 (1999).
- [13] J. T. M. Stevenson and J. R. Jordan, "Use of gratings and periodic structures as alignment targets on wafer steppers," Precision Engineering, vol. 11, pp. 63-69 (1989).
- [14] E. M. Strzelecka, D. A. Louderback, B. J. Thibeault, G. B. Thompson, K. Bertilsson and L. A. Coldren, "Parallel free-space optical interconect based on arrays of vertical-cavity lasers and detectors with monotlithic microlenses," Applied Optics, vol. 37, pp. 2811-2821 (1998).

- [15] S. K. Patra, J. Ma, V. H. Ozguz, S. H. Lee, "Alignment issues in packaging for freespace optical interconnects," Optical Engineering, vol. 33, pp. 1561-1570 (1994).
- [16] P. N. Everett and W. F. Delaney, "Aligning lithography on opposite surfaces of a substrate," Appl. Opt. 31, 7292-7294 (1992).
- [17] M. Yamaguchi, T. Yamamoto, K. Hirabayashi, S. Matsuo and K. Koyabu, "Highdensity digital free-space photonic-switching fabrics using exciton absorption reflection-switch (EARS) arrays and microbeam optical interconnections," IEEE J. of Selected Topics in Quantum Electronics, vol. 2, pp. 47-54 (1996).
- [18] Y. Y. Kipman, P. A. McDonald and R. D. Schuchatowitz, "System and method for aligning a first surface with respect to a second surface," US Patent 5,532,815 (July 2 1996).
- [19] J. M. Sasian, D. A. Baillie, "Simple technique for out-of-focus feature alignment," Opt. Eng. 34, 564-566 (1995).
- [20] B. Robertson, Y. Liu, G. C. Boisset, M. R. Taghizadeh, D. V. Plant, "In situ interferometric alignment systems for the assembly of microchannel relay systems," Applied Optics, vol. 36, pp. 9253-9260 (1997).
- [21] F. Lacroix, "Design, analysis and implementation of free-space optical interconnects," Chapter 7, Ph.D. Thesis, McGill University, Montréal, Canada, 2000.
- [22] G. C. Boisset, B. Robertson, W. S. Hsiao, M. R. Taghizadeh, J. Simmons, K. Song, M. Matin, D. A. Thompson, D. V. Plant, "On-die diffractive alignment structures for packaging of microlens arrays with 2-D optoelectronics device arrays," IEEE Photonics Technology Letters, vol. 8, pp. 918-920 (1996).
- [23] S. J. Bennett and J. W. C. Gates, "The design of detector arrays for laser alignment systems," J. of Physics E, vol. 3, pp. 65-68 (1970).
- [24] L. Bursanescu and V. Vasiliu, "Laser system for high accuracy alignment and positioning," Rev. Sci. Instrum., vol. 65, pp. 1686-1690 (1994).
- [25] A. J. Makynen, J. T. Kostamovaara, and T. E. Rahkonen, "CMOS photodetectors for industrial position sensing," IEEE Transactions on Intrumentation and Measurement, vol. 43, pp. 489-492 (1994).

- [26] N. K. S. Lee, Y. Cai, and A. Joneja, "High-resolution multidimensional displacement monitoring system," Optical Engineering, vol. 36, pp. 2287-2293 (1997).
- [27] P. Aubert, H. J. Oguey, and R. Vuilleumier, "Monolithic optical position encoder with on-chip photodiodes," IEEE J. of Solid-State Circuits, vol. 23, pp. 465-473 (1988).
- [28] S. M. Sze, *Physics of Semiconductor Devices*, New York: Wiley, 1981, Chapter 10.
- [29] T. K. Woodward and A. V. Krishnamoorthy, "1-Gb/s integrated optical detectors and receivers in commercial CMOS technologies", IEEE J. of Selected Topics in Quantum Electronics, vol. 5, p. 146-156 (1999).
- [30] E. Hecht, Optics, 3rd ed., (Addison-Wesley, Reading, Mass. USA, 1998).
- [31] J. Jahns and W. Däschner, "Precise alignment through thick wafers using an optical copying technique," Optics Letters, vol. 17, pp. 390-392 (1992).
Chapter 6: Inter-module alignment techniques

6.1 Introduction

The previous chapter presented various alignment techniques used to package components together to form rigid modules, a task referred to as intra-module alignment. The other aspect of the alignment problem, which is the subject of this chapter, concerns the alignment of modules to one another, a task referred to as inter-module alignment.

Intra-module and inter-module alignment techniques are inherently different in nature due to the physical scale of the problem they each try to solve. Intra-module alignment features (e.g. solder bumps, silicon V-grooves, fiducial makers) are typically of the microscopic scale. This introduces the demand for a complex micro-packaging infrastructure capable of performing the tasks of component handling, misalignment monitoring and micropositioning [1]. Once alignment is achieved, components are permanently fixed in place, usually through the use of adhesives or solder.

Conversely, inter-module alignment features are orders of magnitude larger in size (e.g. dowel pins, baseplate grooves). The requirement here is for modules to be brought into alignment without the need for complex optical setups and expensive staging equipment. Practical 2D-FSOI systems will need to be field-serviceable and therefore allow for the removal and replacement of a defective module without upsetting the alignment of the system. In that context, the ideal scenario is one where modules are manually aligned to each other, similar to the ease with which connectorized fiber components can be assembled.

To realize this, several inter-module alignment techniques have been demonstrated and they are described in section 6.2. The remainder of the chapter is organized as follows. Section 6.3 examines the problem of interfacing the chip module with the beam combination module (BCM) and two inter-module alignment techniques are proposed. Both techniques are passive and use mechanical alignment structures. The first technique (section 6.4) uses a kinematic fixture between the modules. The mechanical structures are implemented using miniature ball lenses and alignment grooves. The grooves are fabricated using thick photoresist and so their position is lithographically-defined. The second technique (section 6.5) uses a pair of dowel pins inserted in the baseplate and precision machined holes in the chip module. The design, limitations and experimental evaluation of each technique are presented.

6.2 Review of previously published techniques

Several inter-module alignment techniques have been demonstrated in the past. These techniques can be divided in three categories: mechanical methods, optical methods and array redundancy methods.

6.2.1 Mechanical methods

These methods use high-precision mechanical structures to align one module to another. The challenge is to fabricate mechanical structures with small positional and dimensional tolerances such that the worst-case misalignment between modules never exceeds the allowances of the misalignment budget.

One of the first mechanical methods used in the construction of 2D-FSOI systems is the slotted baseplate shown in figure 6.1(a). The baseplate approach consists of (i) mounting components together on cylindrical cells and (ii) laying out the cells on high-precision slots milled in a solid support structure which is typically about one inch thick. The cells are often held in place via the restoring force generated by magnets located at the bottom of the milled slots. This alignment scheme is semi-kinematic: the lateral (x, y) and tilt (θ_x, θ_y) degrees of freedom (DOFs) are constrained by the slot while the longitudinal (z) and rotational (θ_z) DOFs are adjustable. Materials used for machining baseplates include: aluminum [2], electroless nickel-plated steel [3], magnesium [4] and glass-filled nylon [5].

A significant drawback of the baseplate approach is the difficulty of assembling an optical module on a single cylindrical cell. Optical components are usually mounted on individual cells, resulting in a large number of DOFs to adjust and maintain. Besides lack-ing modularity, systems implemented on baseplates tend to be bulky and heavy.



Figure 6.1. Mechanical methods: (a) slotted baseplate and (b) L-shaped structure.

A different approach uses square-shaped frames mounted on a L-shaped mechanical structure [6][7], as shown in figure 6.1(b). Optical components are first aligned and fixed to their frame; this step corresponds to intra-module alignment and requires the use of a high-precision assembly setup and micropositioning equipment. Mounting frames can be made with precisely the same shape and size by machining them out of the same prefabricated block of material. The sides of the frame that come in contact with the L-shaped structure are polished using conventional techniques. An optical module is realized by stacking multiple frames into the L-shaped structure. The insertion of spacer frames of precise thicknesses results in having all frames passively aligned in all six DOFs. A pair of L-shaped modules may be aligned using the same approach, that is, by stacking the modules into another L-shaped structure. This approach was successfully demonstrated in [7].

Unlike the slotted baseplate, the L-shaped structure allows for modules to be rapidly assembled with no need for further adjustments. In addition, the geometry of the L-shaped structure is more compatible with square-shaped components such as optical substrates, beamsplitter cubes and electronic chips. This is illustrated in figure 6.1(b), where a beamsplitter cube is passively aligned by mounting it directly on the L-shaped structure. Although the L-shaped structure offers a high degree of modularity and easy assembly, the burden of precision is placed on the alignment of the optical components and the machining and polishing of the mounting frames. An approach similar to the L-shaped structure was used for aligning the relay module in the photonic backplane demonstrator system described in chapter 3. The approach is shown in figure 6.2; it uses three precision-ground rods mounted on the system baseplate. The use of three rods, as opposed to two perpendicular planes, is preferred because (i) it provides only three lines of contact which improves repeatability, (ii) it avoids the requirement for high-precision machining and polishing of mounting frames and (iii) precisionground rods are available commercially at low cost.

The alignment and attachment of the mini-lens arrays to the block of glass is described in detail in [8]. The mini-lens array substrates are precisely diced ($\pm 10 \mu$ m) along lithographically-defined chrome lines deposited on the substrates during the fabrication of the mini-lenses. The cross-sectional dimensions of the block of glass are specified to be slightly smaller than the dimensions of the diced substrates; this is to ensure that only the mini-lens substrates come in contact with the rods, thereby constraining the module in five DOFs. The remaining DOF (longitudinal position) can be constrained by inserting a precision dowel pin in the baseplate, as shown in figure 6.2. The repeatability of this assembly technique was demonstrated during the implementation of the photonic backplane demonstrator; it was shown that the relay module could be manually removed and reinserted with negligible misalignment of the spot array at the receiver plane.



Figure 6.2. Alignment of the relay module using an improved L-shaped structure.

The important aspect of this alignment method is that the edges of the mini-lens substrates (used as passive alignment features) are precisely defined (to within $\pm 10 \ \mu$ m) relative to the mini-lenses. This requirement is also true of other mechanical methods in general, that is, the mechanical alignment structures must be precisely defined relative to the optical elements. This suggests an approach where the mechanical structures can be integrated directly during the fabrication of the optical elements.

A technology that promises to fulfill this goal is high-precision injection molded plastic [9]. This technology can provide 1- μ m dimensional tolerances and is widely used today in the fabrication of fiber connectors [10]. High-precision injection molding also allows for the integration of refractive and diffractive optical elements [11], which means that it is possible to fabricate single-component plastic modules which combine multiple optical elements with built-in mechanical alignment structures. The main drawback to plastic molding - which is what has limited its use in demonstrator systems - is the large setup cost which can only amortized by producing parts in large volumes.

6.2.2 Optical methods

Optical methods achieve inter-module alignment by displacing light beams as they propagate from one module to the next. Conventional approaches include the use of wedge prisms [3][4] (also referred to as Risley prisms) and tilt plates [12]; both are shown in figure 6.3. A wedge prism with a wedge angle β will impart an angular deviation δ to a paraxial beam such that $\delta = \beta \times (n-1)$ where *n* is the index of refraction of the glass. A tilt plate positioned at an angle θ will laterally deviate a paraxial beam by an amount $\Delta = t \times \theta \times [1-(1/n)]$, where *t* is the thickness of tilt plate. These components are used in pairs and rotated manually about the optical axis, allowing for beam adjustments in both directions. There is an inherent trade-off between travel range and alignment resolution; a large wedge or tilt angle increases travel but decreases resolution, and vice-versa.





Stronger beam-steering action has been achieved using LC microprism arrays [19], as shown in figure 6.4. This device consists of a homogeneously aligned nematic LC cell sandwiched between two transparent glass plates. Microprisms are fabricated on the bottom plate and an indium tin oxide (ITO) transparent electrode is deposited on it. The top transparent glass plate is stripe-patterned with ITO electrodes. When assembled, a small gap (~10 μ m) exists between the plates and so the LC molecules are trapped by the boundaries defined by the microprisms; this is to ensure that the voltage applied to one electrode has no influence on neighbouring microprisms. Beam deflection is achieved by varying the voltage on an ITO electrode. This is because the presence of an electric field modifies the orientation of LC molecules which changes the index of refraction. This device was demonstrated with a maximum deflection angle of 10° at a RMS voltage of 2.8 V.



Figure 6.4. Liquid-crystal microprism array.

Recent advances in the field of micro-electro-mechanical systems (MEMS) open a new avenue for the development of miniature beam-steering devices. For example, the use of MEMS translation actuators has been considered for the active alignment of a microlens array on top of an OE-VLSI chip [20]. Another example uses a MEMS micro-mirror to actively compensate for misalignment between a 4×4 VCSEL array and an 4×4 MSM detector array, as shown in figure 6.5 [21]. In this case, the device arrays are interconnected using a hybrid optical relay (see section 2.7.3) where only half the aperture of the conventional lens is being used. The micro-mirror is positioned at the focal point between the conventional lenses.



Figure 6.5. Beam steering using a MEMS micro-mirror.

6.2.3 Array redundancy methods

A straightforward approach at relaxing misalignment tolerances and increasing system robustness would be to use very large detectors (e.g > 100 μ m). Unfortunately, large detectors are undesirable because their large size and high capacitance result in low-density interconnections and low-speed operation (see section 2.5). The advantages of small detectors may be retained while still increasing misalignment tolerances through the use of an oversized array of small detectors [22][23]. An example of this approach is illustrated in figure 6.6, where a 4 × 4 beam array is incident on a 11 × 11 detector array.



Figure 6.6. Misalignment-tolerant detector array: (a) perfect alignment conditions and (b) with misalignment in multiple degrees of freedom.

This approach may be used to remove the misalignment introduced during the assembly of modules. The method requires a beam tracking routine, which is performed as follows. First, the array of laser sources are activated, allowing for the detectors being hit by the signal beams to be identified. In the case where a signal beam overlaps with more than one detector, the detector receiving the most optical power is selected. A critical design consideration therefore concerns the problem of having a signal beam falling in the region between detector active areas. This problem is minimized by designing a spot size that is large compared to the gaps between detectors; these gaps are imposed by the design rules of the detectors [23]. Once the detectors are identified, their photocurrent signals are routed towards a receiver array using control electronics. Data transmission can then be initiated.

This technique may also be operated as a closed-loop system by continuously tracking the alignment of the incident beams during data transmission and compensating for any misalignment by re-routing the photocurrent signals. Adaptive tracking of the incident beams may be required to compensate for mechanical vibrations and thermal drift.

Although the previous discussion has focused on the use of redundant detectors to relax misalignment tolerances, redundant laser sources can also be used, or a combination of redundant laser and detector arrays. The latter scheme has been used to demonstrate misalignment-tolerant free-space optical links between two printed circuit boards [24]. Each link was implemented using a 3×3 VCSEL array interconnected to a 3×3 detector array, a factor of 9 redundancy. This system was designed to tolerate lateral misalignment of ± 1 mm and angular misalignment of $\pm 1^{\circ}$ between boards separated by a distance ranging from 5 to 22 cm.

The use of redundant lasers and detectors to compensate for misalignment is a good example of how the close integration and cooperation of optics and electronics may help improve the manufacturability of 2D-FSOI systems. As Tewksbury et al. write in [25]: "Just as optics may help relax some of the limitations confronting electrical interconnections within microelectronics systems, so also may microelectronics help relax the limitations which would confront purely optical solutions to the interconnect problem."

freedom (DOFs) it is intended to restrain, assuming that no two contact points duplicate the same restraining function. As a result, if two components are to be restrained in all six DOFs, then they must have exactly six points of contact.

6.4.1 The Kelvin clamp

One of the most common type of kinematic fixture is shown in figure 6.8 [26]. This type of fixture was invented by Lord Kevin and so it is often referred to as a Kelvin clamp. The fixture is based on having three ball points resting in a trihedral hollow, a V-groove and a planar surface, respectively. This arrangement provides a total a six contact points between the two components: three in the trihedral hollow, two on the faces of the groove and one on the contact plane. As a result, both components are coupled to each other with no DOFs and no unnecessary constraints.



Figure 6.8. Kelvin clamp implementation using a plane, a V-groove and a hollow.

The Kelvin clamp offers several advantages. The main advantage is an excellent insertion repeatability, less than 1 μ m [27]. In addition, any dimensional variations in the components due to thermal variations will simply cause the ball points on the planar surface and the V-groove to slide, avoiding any strain in the fixture. Finally, since only six contact points exist between the components, the probability of having foreign matter (e.g. dust) contaminating the interface and deteriorating the repeatability is minimal. Even in the case interferometric Fresnel lenses which are used to provide accurate alignment to the patterned mirror grating (PMG) element, as described in [28]. The alignment balls are implemented by mounting ruby balls on donut-shaped structures on the mini-lens substrate. Ruby balls are used because their red colour makes them easier to handle and they are commercially available at low cost in a variety of sizes with dimensional tolerances of ± 0.000025 " ($\pm 0.64 \mu$ m) [29]. Note that the fixture could have equally well been designed with the balls on the jointing plate and the grooves on the mini-lens substrate.

The alignment grooves and donuts structures are fabricated using SU-8 ultrathick photoresist. SU-8 is an epoxy-based negative-tone photoresist designed specifically for highaspect-ratio MEMS-type structures [30]. It is made from EPON SU-8 resin (from Shell Chemical) dissolved in an organic solvent GBL (gamma-butyloracton). It provides a lowcost alternative to the LIGA process (an acronym from the German *Li*thographie, *Gal*vanoforming, and *A*bformung) which requires an expensive X-ray source and a demanding mask technology to pattern very thick (e.g. 1 mm) PMMA (polymethylmethacrylate) layers [31]. The key property of SU-8 that makes it an enabling technology for highaspect-ratio micro-structures is its low absorption in the near-UV (400 nm), allowing for uniform exposure conditions over the total thickness of the photoresist layer, giving rise to vertical sidewalls and good dimensional control. Whereas standard resists used in microelectronics fabrication are a few microns thick, SU-8 layer thicknesses up to 700 µm have been achieved in a single-coat operation [32]. Following the development process, a baking operation (200°C for 30 minutes) turns the photoresist features into hard epoxy.

For the application at hand, the use of alignment structures fabricated using baked SU-8 photoresist is attractive because (i) it allows for the grooves and donuts to be lithographically defined relative to the mini-lenses, (ii) structures with heights of 100's of microns are possible in a single-coat operation, (iii) the process only requires standard lithography equipment such as a spinner and a mask aligner (iv) fabrication can be performed at the wafer level and (v) SU-8 photoresist is inexpensive and commercially available [33].

6.4.3 Fabrication of alignment micro-structures using SU-8 photoresist

As a chemically-enhanced negative-tone photoresist, SU-8 must be spun, pre-exposure baked, exposed, post-exposure baked and developed. Associated with these steps is a large

uniformity and repeatability problems were mostly caused by the very high viscosity of SU-8 photoresist (a viscosity of 990 cSt is similar to that of liquid honey). A high viscosity makes the photoresist hard to manipulate and difficult to apply in controlled amounts. It was found that layer uniformity could be improved by carefully depositing the photoresist across the surface of the sample prior to spinning.

Results of uniformity measurements taken on four different samples are shown in table 6.1. The layer thickness was measured over a $10 \times 10 \text{ mm}^2$ area located in the central region of the 25-mm diameter sample. Measurements were performed using a stylus profiler (Veeco Dektak 8). For each sample, the minimum and maximum layer thickness were recorded and a minimum of 12 thickness profiles were collected to estimate the average thickness.

Measurements	Sample 1	Sample 2	Sample 3	Sample 4
Lowest point	28 µm	49 µm	41 µm	37 μm
Highest point	70 µm	80 µm	90 µm	78 µm
Average thickness	47 µm	68 µm	67 µm	58 µm

Table 6.1. Layer uniformity data collected on four samples.

Various technical hurdles were encountered during the fabrication process. Foremost was the problem of having photoresist pouring over the edges of the sample and flowing between the sample and the hot plate during pre-exposure bake. In many cases, the surface of the resist following pre-exposure bake was too tacky and got stuck to the mask during contact printing. Finally, photoresist residues could be seen on most samples after development, suggesting that the exposure and developing steps were not optimal.

Next, the fabrication process was adjusted in order to improve uniformity and repeatability. The main objective was to develop a process with a layer thickness of 100 μ m or more. Thick layers are required in order for large micro-structures and ruby balls to be used; this is desirable if the insertion/removal of the chip module is to be performed by a human hand. Determining the optimal process parameters is a laborious procedure because there is a large set of parameters that can be varied. The approach taken was systematic: the process parameters were varied over a certain range and multiple combinations of parameters were tested:

- spinning speed = 500, 1000, 2000 and 3000 rpm.
- spinning time = 5, 10 and 20 seconds.
- pre-exposure bake time = 10, 20, 30, and 60 minutes.
- exposure = 60, 85, 120, and 200 seconds.
- post-exposure bake time = 5 and 10 minutes.
- development time = 10, 15, 20, and 25 minutes.

The best results were obtained with the following recipe: spin at 2000 rpm for 5 seconds, pre-exposure bake at 95°C for 60 minutes, exposure time of 60 seconds, post-exposure bake at 95°C for 10 minutes and development for 20 minutes. A longer pre-exposure bake removed more solvent; this reduced the tackiness of the resist and minimized the chances of the sample getting stuck in the mask aligner. The development time was found not to be critical. This recipe produced a layer thickness of $100 \pm 20 \mu m$. A SEM photograph of SU-8 micro-structures after development is shown in figure 6.10. Note that a random mask was used in this case and so the shape of the micro-structures is meaningless.



Figure 6.10. SEM photograph of 100-µm tall SU-8 micro-structures.

6.4.4 Discussion

Despite a large number of attempts, the layer uniformity was still found to be unsatisfactory. Assuming the Kelvin clamp implementation of figure 6.9 with pairs of alignment grooves separated by a distance of 10 mm, a thickness variation of \pm 20 µm translates in a worst-case tilt misalignment of $\Delta \theta_{x,y} = 0.22^{\circ}$, which far exceeds the provisions of the misalignment budget (0.03°). Note that a uniformity of \pm 20% is very poor compared to what has been achieved by others; for example, Electronics Vision Inc. quotes a typical uniformity of \pm 3% on 150 µm thick SU-8 layers [35].

With the uniformity problem still unresolved, a Kelvin clamp prototype was nevertheless implemented by fabricating SU-8 alignment micro-structures on two fused silica substrates. Following development, the micro-structures were baked at 200°C for 30 minutes; this hardened the micro-structures and increased their adhesion to the substrate. Next, the Kelvin clamp was tested by manually bringing the two substrates together. This experiment uncovered a fundamental design flaw: the fact that 100-µm alignment grooves are too small for the human hand to feel whether the balls are in the grooves or not. Still, the two substrates were brought together with moderate pressure a few times and the microstructures were subsequently examined under an optical microscope; the repeated mating operations showed significant damage to the grooves and in some cases complete removal from the substrate.

The lessons learned from this work are as follows. First, for inter-module alignment to be performed by a human hand, the alignment structures must be significantly larger, perhaps as much as 1 mm. This is to ensure that when modules are brought together, the balls fall somewhere on the slopes of the grooves. Fabricating micro-structures of this size is pushing the limits of SU-8 photoresist technology. A better approach would be to use a separate guiding mechanism to perform a coarse alignment step, as shown in figure 6.11. In this case, a pair of dowel pins press-fit in the baseplate are inserted in guiding holes machined in the chip module. The dowel pins should be a few millimeters in diameter. The guiding holes are oversized compared to the dowel pin diameter to facilitate manual insertion. The design is specified such that the misalignment of a dowel pin relative to the center of its guiding hole never exceeds the size of the V-groove slope (referring to figure 6.11, this is to say that $\Delta < \delta$ always). The design of figure 6.11 effectively uses the dowel pins for coarse alignment and the micro-structures for fine alignment.



Figure 6.11. Alternative kinematic design assisted by guiding dowel pins.

Another issue concerns the hardness of the material used to fabricate the alignment grooves. Although the use of six contact points constitutes sound kinematic design, it results in a large contact pressure being applied to small regions on the grooves. If the grooves are not hard enough (this was the case for SU-8 micro-structures), this leads to structural bending which deteriorates insertion repeatability and therefore defeats the purpose of the kinematic approach altogether.

A straightforward solution consists of fabricating the alignment grooves using a hard material. For example, the use of V-grooves etched in a silicon waferboard should be considered [36]. This, however, creates the additional problem of accurately aligning the silicon waferboard to the optical elements. This alignment step could be performed using flip-chip techniques followed with epoxy underflow to reinforce the solder joints. Alternatively, for systems operating at a wavelength above the bandgap of silicon (e.g. 1.3 or 1.55 μ m systems), this alignment step can be avoided by fabricating the optical elements directly on the silicon waferboard. A different approach would be to fabricate the grooves out of solid metal. This can be done to lithographic precision by evaporating metal pads on the fused silica substrate and using a thick layer of SU-8 photoresist as an electroplating mask [37].

A simpler strategy would be to replace the pure kinematic fixture by a semi-kinematic one, an example of which is shown in figure 6.12. The term semi-kinematic is used to reflect the fact that contact points are replaced by planes of contact. The use of a semikinematic fixture becomes possible by realizing that the precision of a pure kinematic design (<1 μ m) far exceeds the lateral requirement of the misalignment budget (26 μ m). In the design of figure 6.12, the fine alignment is performed by inserting a micro-pin in a donut. To facilitate assembly, the diameter of the micro-pin is made smaller than the opening of the donut by an amount equal to the lateral misalignment budget. Furthermore, by making the micro-pins slightly shorter than the thickness of the donuts, the contact pressure is distributed over the entire surface of the donut. The use of large donuts allow for a soft material (such as baked SU-8 photoresist) to be used.



Figure 6.12. Semi-kinematic design using pins and donuts micro-structures.

The final point concerns the issue of robustness. Micro-structures are small and fragile and therefore easily breakable. For example, any lateral force applied on the chip module of figure 6.12 is likely to cause enough stress on the micro-pins to break them. In general, the reliability of the inter-module fixture depends on the robustness of the micro-structures. Of course, the design of figure 6.12 can be modified by tightening the tolerances on the guiding holes so that any lateral stress is supported by the dowel pins (instead of the micro-pins). But if this can be achieved, then the alignment is determined by the dowel pins alone and the micro-pins are not required anymore. This is the approach considered in the next section. strate comes in full contact with the jointing plate. The dowel pins and alignment holes are specified with tight tolerances, enough to provide adequate lateral and rotational alignment of the chip module relative to the BCM. Hence, the dowel pins constrain the lateral and rotational DOFs of the chip module while the flat optical substrates constrain the tilt and longitudinal DOFs. The alignment of the chip module is achieved simply by applying a force directed towards the BCM. This force is provided by a pair of spring-loaded screws inserted from the back of the chip module through the clamping holes shown in figure 6.14.

A potential problem of semi-kinematic designs in general is the possibility for components to be overconstrained. In the design of figure 6.13, there will always be a slight tilt misalignment between the jointing plate and the dowel pins. If the fit between the pins and the holes is too close, then it may become impossible for the mini-lens substrate to come in full contact with the jointing plate due to obstruction of dowel pins by the holes. Increasing the size of the holes increases the angular play of the chip module, but this can only be done at the expense of lateral precision. There is thus a trade-off between lateral alignment and angular play.

Alignment Operations	Tolerances	Tolerances determined by
Dowel pin location	± 10 μm	CNC capability
BCM location	± 20 μm	BCM mounting technique
Precision hole location	± 10 μm	CNC capability
Mini-lens array location	$\pm 20 \ \mu m$	Mini-lens mounting technique
Pin-to-hole misalignment	±20 μm	Specifications of figure 6.13

Table 6.2. Tolerances contributing to the lateral misalignment of the chip module.

The worst-case lateral misalignment of the chip module relative to the BCM depends on (i) the locational accuracy of the dowel pins, (ii) the alignment accuracy of the BCM relative to the dowel pins, (iii) the locational accuracy of the precision holes, (iv) the alignment accuracy of the mini-lens array relative to the precision holes, and (v) the worstcase misalignment of the dowel pin inside a precision hole. Given the specifications of figure 6.13, this last item is calculated by considering the case when the size of the pin is minimum and the size of the hole is maximum, resulting in a worst-case pin-to-hole misalignment of $\pm 20 \ \mu\text{m}$. The tolerances associated with the above alignment operations are listed in table 6.2.

The sum of the tolerances listed in table 6.2 indicates a worst-case lateral misalignment of \pm 80 µm, which far exceeds the allowances of the misalignment budget (\pm 26 µm). To circumvent this problem, a pair of tilt plates are inserted in the OPS module (see figure 3.13) and used to adjust the lateral alignment of the beams incident on the chip module. In effect, the tilt plates allow for the first four entries in table 6.2 to be compensated for actively. Note that the tilt plates are required to be adjusted only once, during the first insertion of the chip module, in a manner similar to a calibration procedure. This results in subsequent insertions to have a worst-case lateral misalignment determined solely by the pin-to-hole misalignment (\pm 20 µm). Using the fact that the design of figure 6.14 uses dowel pins separated by a distance of 56.6 mm, the worst-case rotational misalignment is calculated to be \pm 0.04°. Thus, the worst-case lateral and rotational misalignments are well within the allowances of the misalignment budget.

The last point to be considered concerns the angular play of the chip module. The fixture was designed such that the penetration depth of the dowel pins in the precision holes is 3.0 mm. Thus, the minimum angular play is calculated to be $\pm 0.5^{\circ}$, which is plenty.

6.5.1 Experimental evaluation of insertion repeatability

To evaluate the repeatability of the fixture, a diagnostic chip module (DCM) was assembled using the alignment technique described in section 5.5. A DCM is the equivalent of a chip module for which the heatsink, TEC and heat spreader have been removed and the OE-VLSI chip replaced with a transparent fused silica substrate having lithographically-defined metal targets replicating the location and size of the MQW modulators and detectors on the chip. Thus, a DCM uses exactly the same optomechanics as a chip module and provides easy access to the back of the module, allowing for the spots position in the device plane to be directly observed with a CCD camera. The positional information of the spots is used to quantify the repeatability of the chip module fixture. Prior to the repeatability measurements, the alignment of the OPS beam array is performed. To do this, a DCM is inserted in the system and secured in place. The lateral alignment of the beams is adjusted using the OPS tilt plates; this is done by imaging the mini-lens array from the back of the DCM and centering the beams on their respective mini-lenses. The tilt alignment of the beams is adjusted using OPS Risley prisms; this is done by imaging the device plane from the back of the DCM and centering the spots on their respective metal targets. The result of this alignment step is shown in figure 6.15, showing all 512 spots aligned to the DCM metal targets.



Figure 6.15. CCD images showing alignment of spots on DCM metal targets.

Repeatability measurements were performed by removing the chip module completely, inserting it back in the system and securing it in place. This removal/insertion operation was repeated 50 times. For each removal and insertion cycle, the lateral misalignments of the two spots located at the opposite corners of the array were recorded. The upper-left and lower-right corner spots are separated by a distance of 7.17 mm across the chip. The misalignment data is plotted in figure 6.16. The radial misalignment was calculated for each data point according to $r = (x^2 + y^2)^{\frac{1}{2}}$ and the resulting standard deviation is $\sigma = 2.2 \ \mu m$. Assuming a random process and a normal distribution, this means that there is a probability of 99.7% (3 σ metric) for all spots to fall within a circle of diameter 6.6 μ m centered on the MQW modulators. Using the fact that the 3 ω spot diameter is equal to 39.3 μ m (see section 3.7.1), this corresponds to a minimum device diameter of 45.9 μ m. This requirement was satisfied by designing modulators with a diameter of 52.5 μ m and square detectors with dimensions 66 × 66 μ m (see section 3.6.3).



Figure 6.16. Repeatability data after 50 insertion/removal cycles.

The above results validate the semi-kinematic design of figure 6.13 by showing that the chip module can be manually inserted and removed from the optical system with minimal insertion loss and no need for further adjustments. This demonstrates that OE-VLSI chips with large arrays (32×32 devices) can be replaceable using a simple mechanical interface.

6.6 Conclusion

This chapter focused on the inter-module alignment problem. A review of previous techniques, both mechanical and optical, was presented. The main challenge behind the use inter-module mechanical techniques lies in the difficulty of accurately locating the mechanical structures relative to the optical elements. The use of high-precision molded

plastic technology is very promising because it allows for modules to combine both optical elements and built-in mechanical alignment structures in the same component.

The main challenge behind the use of inter-module optical techniques concerns the reliability, compactness and performance of the beam-steering devices. In multi-stage 2D-FSOI systems (ones interconnecting many OE-VLSI chips), a large number of beam-steering devices are required and the corresponding increase in cost, volume and complexity (e.g. consider the requirements in control electronics) makes it unclear whether beam-steering techniques represent a viable solution. However, the agility of beam-steering devices is likely to be exploited in specialized applications, where systems are subjected to extreme mechanical shock and temperature cycling (military and aerospace applications).

The use of redundant arrays of detectors and lasers represent an attractive alternative to the above. Redundancy techniques essentially trade-off interconnection density for more tolerance to misalignment. The use of a snap-together mechanical interface, complemented by a small degree of redundancy in the detector array, may represent the solution of choice in future 2D-FSOI systems.

The problem of interfacing the chip module to the BCM was considered and two novel inter-module techniques were presented. The first technique implemented a pure kinematic fixture (a Kelvin clamp) using lithographically-defined photoresist micro-structures (100 μ m thick) fabricated directly on the optical substrates. Problems related to the lack of uniformity, hardness and robustness of the photoresist micro-structures precluded the use of this technique in the system.

The second technique implemented a semi-kinematic fixture using a pair of dowel pins and the front surface of the optical substrates to constrain the chip module in all six DOFs. This technique was used to insert the chip module during final system assembly. Repeatability measurements (50 insertion/removal cycles) show a lateral misalignment of the spot array with a standard deviation of only $\sigma = 2.2 \,\mu\text{m}$. This is a significant result because it demonstrates that OE-VLSI chips with large arrays (32 × 32 devices) can be replaceable using simple inter-module mechanical techniques.

6.7 Acknowledgments

The author of this thesis gratefully acknowledges the assistance and patience of Dr. Edwis Richard in teaching him the rudiments of microlithography. The work performed with SU-8 photoresist would not have been possible without him. Dr. Richard is also acknowledged for taking the SEM photograph shown in figure 6.10. Rhys Adams is acknowledged for his assistance in collecting the data of figure 6.16.

6.8 References

- G. Tittelbach, R. Eberhardt, V. Guyenot, "Assembling of microoptical components," SPIE Proceedings vol. 3008, pp. 242-250 (1997).
- [2] F. B. McCormick, F. A. P. Tooley, T. J. Cloonan, J. L. Brubaker, A. L. Lentine, R. L. Morrison, S. J. Hinterlong, M. J. Herron, S. L. Walker, and J. M. Sasian, "Experimental investigation of a free-space optical switching network by using symmetric self-electro-optic-effect devices," Applied Optics, vol. 31, pp. 5431-5445 (1992).
- [3] F. B. McCormick, T. J. Cloonan, F. A. P. Tooley, A. L. Lentine, J. M. Sasian, J. L. Brubaker, S. L. Walker, R. J. Crisci, R. A. Novotny, S. J. Hinterlong, H. S. Hinton, and E. Kerbis, "Six-stage digital free-space optical switchig network using symmetric self-electro-optic-effect devices," Applied Optics, vol. 32, pp. 5153-5170 (1993).
- [4] G. C. Boisset, M. H. Ayliffe, B. Robertson, R. Iyer, Y. S. Liu, D. V. Plant, D. J. Goodwill, D. Kabal, D. Pavlasek, "Optomechanics for a four-stage hybrid-self-electro-optic-device-based free-space optical backplane," Applied Optics, vol. 36, pp. 7341-7358 (1997).
- [5] D. T. Neilson, S. M. Prince, D. A. Baillie, and F. A. P. Tooley, "Optical design of a 1024-channel free-space sorting demonstrator," Applied Optics, vol. 36, pp. 9243-9251 (1997).
- [6] M. Mizukami, K. Koyabu, M. Fukui, and K. Kitayama, "Free-space optical module configuration using a guide-frame assembly method", Applied Optics, vol. 34, pp. 1783-1787 (1995).
- [7] M. Yamaguchi, T. Yamamoto, K. Hirabayashi, S. Matsuo, K. Koyabu, "High-density digital free-space photonic-switching fabrics using exciton absorption reflection-

- [20] F. S. J. Michael, "A feasibility study: positioning a lenslet array above a target using MEMS to specify three of four degrees of freedom", M. Eng. Thesis, McGill University, Montréal, Canada, September 2000.
- [21] C. Berger, J. Ekman, X. Wang, P. Marchand, H. Spaanenburg, F. Kiamilev, ans S. Esener, "Parallel distributed free-space optoelectronic compute engine using flat plug-on-top optics package," in Optics in Computing 2000, R. A. Lessard, T. Galstian, Proc. SPIE 4089, pp. 1037-1045 (2000).
- [22] L. A. Hornak, S. K. Tewksbury, V. K. Konkimalla, "Active mini-MCM daughterboard for optical interconnect insertion into microelectronic systems", 1994 Electronic Components and Technology Conference, pp. 317-322 (1994).
- [23] F. A. P. Tooley, "Challenges in optically interconnecting electronics," IEEE Journal of Selected Topics in Quantum Electronics, vol. 2, pp. 3-13 (1996).
- [24] D. V. Plant, E. Bernier, E. Bisaillon, M. Mony, M. Salzberg, T. Yamamoto, D. J. Goodwill, and A. G. Kirk, "A 5 Gb/s, 2-channel bi-directional adaptive redundant FSOI demonstrator system," in Optics in Computing 2000, R. A. Lessard, T. Galstian, Proc. SPIE 4089, pp. 465-472 (2000).
- [25] S. K. Tewksbury, L. A. Hornak, H. E. Nariman, S. M. Langsjoen, N. J. Hall, J. J. Hall, and S. P. McGinnis, "Toward cointegration of optical interconnections within silicon microelectronic systems," J. of Parallel and Distributed Computing, vol. 17, pp. 188-199 (1993).
- [26] J. E. Furse, "Kinematic design of fine mechanisms in instruments," J. of Physics E -Scientific Instruments, vol. 14, pp. 264-272 (1981).
- [27] A. H. Slocum and A. Donmez, "Kinematic couplings for precision fixturing Part 2: experimental determination of repeatability and stiffness," Precision Engineering, vol. 10, pp. 115-122 (1988).
- [28] F. Lacroix, "Design, analysis and implementation of free-space optical interconnects," Chapter 7, Ph.D. Thesis, McGill University, Montréal, Canada, 2000.
- [29] For example, ruby ball lenses are available from Imetra, Elmsford, NY (USA).

- [30] J. M. Shaw, J. D. Gelorme, N. C. LaBianca, W. E. Conley, and S. J. Holmes, "Negative photoresists for optical lithography", IBM Journal of Research and Development, vol. 41, pp. 81-94 (1997).
- [31] E. W. Becker, W. Ehrfeld, P. Hagmann, A. Maner, and D. Munchmeyer, "Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming, and plastic moulding (LIGA process)", Microelectronic Engineering, vol. 4, pp. 35-56 (1986).
- [32] M. Despont, H. Lorenz, N. Fahrni, J. Brugger. P. Renaud, and P. Vettiger, "High aspect ratio ultrathick, negative-tone near-UV photoresist for MEMS applications," in Proceedings of MEMS'97 (IEEE: Nagoya), pp. 518-522 (1997).
- [33] SU-8 photoresist is available from Microlithography Chemical Corporation, Newton, MA, USA (http://www.microchem.com).
- [34] Units of viscosity are centistokes (cSt). Viscosity data obtained from Electronics Vision Inc., Phoenix, Arizona, USA (http://www.elvisions.com).
- [35] Refer to SU-8 bulletin board on Microlithography Chemical Corporation web page: (http://www.microchem.com/su8.cfm).
- [36] R. A. Boudreau, H. Han, M. Kadar-Kallen, and J. R. Rowlette Sr., "Kinematic mounting of optical and optoelectronic elements on silicon waferboard," U.S. patent #5,574,561, November 12 1996.
- [37] K. Lee, N. LaBianca, S. Rishton, and S. Zohlgharnain, "Micromachining applications for a high resolution ultra-thick photoresist", J. of Vacuum Science Technology Part B, vol. 13, pp. 3012-3016 (1995).

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

7.1 Introduction

The previous chapter presented a variety of techniques for aligning modules to one another, a task referred to as inter-module alignment. Ideally, 2D-FSOI modules should be assembled without requiring any adjustments, similar to the ease with which connectorized fiber-based components can be put together. Unfortunately, the accuracy of purely mechanical alignment has so far proven to be insufficient considering the stringent set of misalignment tolerances that is typical of large-scale 2D-FSOI systems. A direct consequence of this has been the large development effort invested in misalignment compensation techniques, such as active beam-steering devices and array redundancy techniques.

Surprisingly, while most of the research has been focusing on misalignment compensation techniques, much less attention has been devoted to the study of optical designs that are inherently tolerant to misalignment. Research along this direction will initial reduce the need for misalignment compensation and may ultimately lead to purely mechanical solutions to the alignment problem.

With this in mind, the work presented in this chapter seeks to identify the types of optical designs that provide a more generous misalignment tolerance budget. The analysis focuses on the design of the chip module and five different optical configurations are examined and compared. The first objective is to determine which of these configurations is the most tolerant to misalignment. The second and more relevant objective is to understand the underlying reasons that makes one configuration more misalignment-tolerant than another.

7.2 Motivation

The development of a 2D-FSOI system usually follows the basic design flow outlined in figure 7.1. The starting point consists of identifying the available technologies (CMOS, optoelectronics, micro-optics, packaging, etc.) and specifying the target performance specifications of the system. In general, there is a close inter-relationship between technology and system performance; one influences the other and vice-versa.



Figure 7.1. Typical design flow of 2D-POI systems.

Next, the optical design of the interconnect is carried out and the parameters (wavelength, focal length, aperture size, etc.) of each optical element is specified. These parameters are used as the input to an optical simulation software and the correct propagation of the beams in a perfectly aligned system is verified. In some cases, a simple ray tracing analysis may be sufficient; in other cases, a full diffractive analysis is required to accurately predict the spot sizes and determine the effects of clipping at the lens apertures. If the simulation results turn out to be unsatisfactory, the design is modified accordingly. This design cycle is iterated until the optical design meets the system requirements.

The next step consists of partitioning the optical system into separate modules. The chosen partitioning scheme must (i) be practical, (ii) minimize the number of DOFs, and (iii) relax inter-module misalignment tolerances. This step was illustrated earlier in section 3.8 for the case of the photonic backplane demonstrator of chapter 3.

The following step in the design flow consists of calculating the misalignment tolerance budget of each module; this specifies the amount of misalignment that is allowed in Few designers usually consider a third option, which consists of modifying the optical design in a way that relaxes the misalignment tolerance budget without sacrificing system performance. Reasons for this are as follows:

- Multiple design cycles are undesirable: the design flow of figure 7.1 is a long and elaborate process that may require weeks of design work. Contributing to this is the fact that a complete tolerance analysis, performed using a statistical method such as Monte-Carlo analysis, is laborious and computer-intensive. Thus, optimizing the misalignment budget by iterating through multiple design cycles is undesirable.
- Designers are not familiar with misalignment-tolerant design techniques: although the outcome of the tolerance analysis specifies the alignability of a design, it provides little or no information as to how the design needs to be modified to further relax tolerances. Contributing to this problem is the fact that the tolerance analysis uses complex models that must be solved numerically, which completely hides the mathematical relationships between the system parameters and the misalignment tolerances. The consequence of this is a lack of general knowledge on misalignment-tolerant design techniques. The only relevant work on misalignment-tolerant techniques was found in [23].

The above discussion motivates the need for a better understanding of the general properties of misalignment-tolerant designs. One approach would be to derive analytical expressions that fully describe the relationships between the system parameters and the misalignment tolerances. However, deriving accurate, general, analytical expressions would be extremely difficult, if not impossible, due to the complexity of the models and the number of DOFs to be considered simultaneously. To circumvent this, while retaining the benefits of analytical expressions, the models are greatly simplified if a straighforward sensitivity analysis is performed. Of course, the resulting analytical expressions will suffer a loss in accuracy, but this is acceptable because the matter of interest is the form of the expressions, not their numerical values.

The above definition of the misalignment metric is incomplete because it does not take into consideration the degradation in system performance resulting from optical crosstalk between neighbouring channels. Optical crosstalk originates from clipping losses at the microlens apertures. This is illustrated in figure 7.2(b), for the case of a chip module where a microlens array is integrated with the chip. The presence of optical crosstalk can be included in the misalignment metric by specifying a minimum clipping ratio, k_l , that must be observed at all lens apertures in the system. The clipping ratio, k_l , is defined as:

$$k_l = \frac{D_{\mu lens}}{2\omega_{\mu lens}} \tag{7.1}$$

where $D_{\mu lens}$ is the effective lens aperture and $\omega_{\mu lens}$ corresponds to the (1/e²) Gaussian beam radius [24]. To summarize, a system is considered to be misaligned if it satisfies one of the following conditions: (i) the link efficiency drops by 50%, or (ii) the clipping ratio at a lens aperture falls below k_l . Belland and Crenn [25] have shown that a minimum clipping ratio of $k_l = 2.12$ ensures that clipping losses are limited to less than 0.1% and that the diffraction effects do not significantly modify the beam propagation characteristics.

7.4 Defining a figure of merit for alignability

The first step towards quantifying the alignability of a module involves the formulation of an adequate alignment figure of merit (FOM). A general expression for the alignment FOM should take into account the misalignment tolerances in all six DOFs. However, the form of such an expression is likely to vary depending on the design and may require the use of weighing factors to account for the fact that some DOFs are more difficult to align than others. Fortunately, these difficulties are avoided by realizing that in the vast majority of cases, only lateral and tilt misalignments need to be considered. Note that this point is consistent with the misalignment budget of the chip module in table 4.1, and it can be understood by considering the following:

• In most interconnect designs, the aperture of lenses and the active area of OE devices have identical lateral dimensions, implying that $\Delta x = \Delta y$ and $\Delta \theta_x = \Delta \theta_y$. Hence, the

alignment information associated with Δy and $\Delta \theta_y$ is redundant and need not be included in the expression for the FOM.

- The rotational misalignment tolerance $(\Delta \theta_z)$ of a module corresponds to the angle associated with a beam located at the corner of the array being rotationally misaligned by an arc distance approximately equal to the lateral misalignment tolerance (Δx) . For example, the rotational misalignment tolerance of a two-dimensional array of beams incident onto an array of square detectors is approximately $\Delta \theta_z \sim 2\Delta x/W$, where W is the linear physical dimension of the array. Thus, the alignment information associated with $\Delta \theta_z$ is already contained in Δx and need not be included in the expression for the FOM.
- The longitudinal misalignment tolerance (Δz) is generally orders of magnitude larger than the lateral tolerance (Δx). To see this, consider a Gaussian beam being focused on a square detector, where the beam waist is ω_o and the detector linear dimension is d. Assume that d = 3ω_o, so that perfect alignment results in > 99% coupling efficiency. It can be shown that a lateral displacement of about half the beam waist yields a 90% coupling efficiency, so that Δx_{η=90%} = ω_o/2. Similarly, a longitudinal displacement equal to the Rayleigh range results in about the same reduction in coupling efficiency, so that Δz_{η=90%} = πω_o²/λ. Thus, even in the limit where the beam is focused down to a spot radius equal to the wavelength of light (ω_o ~ λ), then Δz would still be larger than Δx. Thus, in general, longitudinal tolerances are negligible compared to lateral tolerances and need not be included in the FOM.

The above discussion demonstrates that the alignability of a module can be adequately specified by its lateral (Δx) and tilt ($\Delta \theta_x$) misalignment tolerances, all other DOFs being either redundant, dependent or negligible. Accordingly, we define the alignment FOM to be equal to the product of lateral and tilt tolerances:

Alignment FOM =
$$\Delta x \cdot \Delta \theta$$
 (7.2)

where the subscript on the tilt tolerance has been dropped for simplicity. In what follows, the quantity of equation (7.2) will be referred to as the "alignment product".

7.5 Misalignment tolerance and scalability analysis

This section examines five different approaches to the design of a module integrating a $N \times N$ detector array. The objective is to identify scalable optical designs that maximize the alignment product. Designs are compared under a common set of system parameters, listed in table 7.1. These parameters are used to derive closed-form analytical expressions for lateral and tilt misalignment tolerances. In addition, expressions for the maximum array size (N_{max}) are derived; these are used to determine the scalability of each design.

Parameters	Definitions	
λ	operating wavelength	
Ν	linear size of detector array	
W	linear dimension of the chip	
L	interconnection optical length	
d	linear dimension of detector	
ω _d	beam waist at detector plane	
k_l	minimum clipping ratio at a lens	

Table 7.1. System parameters and definitions.

7.5.1 Design #1: no optics integrated with the chip

The first approach is shown in figure 7.3. It corresponds to a 4-f microchannel design where the microlens is integrated with the beamsplitter rather than the chip. It is assumed that the Gaussian mode is focused to a small spot size, such that $d = 3\omega_d$, ensuring a >99% power efficiency for a perfectly aligned system. A small spot size implies that the Rayleigh range at the detector plane is much shorter than the focal length of the microlens: $\pi\omega_d^2/\lambda \ll f_{\mu lens}$. This results in a simple expression relating the beam waist at the detector, ω_d , to the beam waist at the lens, $\omega_{\mu lens}$:

$$\omega_{\mu lens} = \omega_d \sqrt{1 + (\lambda f_{\mu lens} / \pi \omega_d^2)} \cong \lambda f_{\mu lens} / \pi \omega_d$$
(7.3)

To minimize clipping effects, the microlens diameter, $D_{\mu lens}$, is maximized by setting $D_{\mu lens} = W/N$. The microlens focal length is equal to one fourth of the interconnect optical path length: $f_{\mu lens} = L/4$ (in reality, L is slightly larger due to the refractive index of glass).



Figure 7.3. Design #1: module without any optics (shown with N = 4).

The lateral misalignment tolerance is limited by a 50% detector efficiency drop: $\Delta x = d/2$. Considering that small-size detectors are desirable (see section 2.5), the lateral tolerance is tight. Tilt tolerance is also determined by a 50% efficiency drop at a detector located on the periphery of the array. The tilt misalignment condition can be written as:

$$\frac{d}{2}\cos(\Delta\theta) = \frac{W}{2}[1 - \cos(\Delta\theta)]$$
(7.4)

Using trigonometry, it can be shown that equation (7.4) is well approximated by $\Delta \theta = \sqrt{2d/W}$. The square root function relaxes the tilt misalignment tolerance. The maximum size of the array, N_{max} , is limited by clipping losses at the microlenses:

$$N_{max} = \frac{W}{D_{\mu lens,min}} = \frac{W}{2k_l \omega_{lens}} = \frac{2\pi \omega_d W}{\lambda L k_l} = N^*$$
(7.5)

where equation (7.3) has been used. Equation (7.5) indicates that the scalability of this design is poor, because N_{max} decreases rapidly as the interconnection length (L) increases. In what follows, the quantity corresponding to equation (7.5) is referred to as N^* .

7.5.2 Design #2: microchannel design

A simple variation to the previous design consists of integrating the microlens array with the chip module. This is shown in figure 7.4 below. The microlens diameter and focal length are the same as in design #1.



Figure 7.4. Design #2: Module integrating a microlens array (shown with N = 4).

In this configuration, a lateral displacement of the module with respect to the incident beams results in spots still being focused on the center of the detectors. Thus, the module lateral tolerance is limited by the minimum clipping ratio (k_l) at the microlens aperture:

$$\Delta x = \frac{1}{2} (D_{\mu lens} - 2k_l \omega_{\mu lens}) = \frac{W}{2N} \left(1 - \frac{N}{N^*}\right)$$
(7.6)

where N^* is given in equation (7.5). Equation (7.6) illustrates a classical trade-off between Δx and N: a more dense array means smaller microlenses, resulting in tighter lateral tolerances. A tilt misalignment about the center of the microlens array results in spots being displaced on the detectors. Thus, the module tilt tolerance is limited by a 50% drop in detector efficiency:

$$\Delta \Theta = \frac{d/2}{f_{\mu lens}} = \frac{2d}{L} \tag{7.7}$$

Because $d \ll L$, the tilt tolerance is tight. The maximum array size, N_{max} , is limited by clipping losses at the microlens aperture and is given by equation (7.5). The scalability of design #2 is limited in the same way as in design #1.

7.5.3 Design #3: clustering using mini-lens array

A clustering configuration was used in the design of the chip module presented in chapter 4. This approach groups detectors into $m \times m$ clusters, as shown in figure 7.5 below. A mini-lens is used to relay an entire cluster of beams through the interconnect. Clipping effects are minimized by maximizing the mini-lens diameter: $D_{mini} = mW/N$. The mini-lens focal length is equal to one fourth the interconnection path length: $f_{mini} = L/4$.



Figure 7.5. Design #3: Clustering configuration with mini-lenses (with N = 4 and m = 2).

In this configuration, a lateral displacement of the module results in spots still being centered on the detectors. Assuming, a detector pitch equal to twice the detector size (detector pitch = 2d), the module lateral tolerance is given by:

$$\Delta x = \frac{1}{2} [D_{mini} - 2d(m-1) - 2k_l \omega_{mini}] = \frac{W}{2N} \left(m - \frac{N}{N^*}\right) - (m-1)d$$
(7.8)

where ω_{mini} is calculated according to equation (7.3). Considering that $W/N \gg d$, Δx is mostly dominated by the first term in equation (7.8). Note that by choosing a large cluster size ($m \gg 1$), the lateral tolerance is increased and can be made much larger than in the previous two designs. This can be understood by realizing that, for a given array size N, grouping detectors into larger clusters maximizes the mini-lens aperture, leaving more room on the periphery of the mini-lenses before the clipping condition is reached. In practice, the maximum value of m will be determined by the acceptance angle of the beamsplitter or the field of view of the mini-lens, whichever is smallest. For instance, if a commercial polarizing beamsplitter is used, polarization aberrations will limit the half

acceptance angle to about 5° [26]. In general, assuming a half acceptance angle of ϕ radians, the maximum cluster size, m_{max} , is given by:

$$m_{max} = \frac{\phi L}{4d} + 1 \tag{7.9}$$

The module tilt tolerance is identical to the situation of design #2 and is thus given by:

$$\Delta \Theta = \frac{d/2}{f_{mini}} = \frac{2d}{L} \tag{7.10}$$

Because $d \ll L$, the tilt tolerance is tight. The maximum array size, N_{max} , is limited by clipping losses at the microlenses. It can be calculated by equating Δx to zero in equation (7.8) and solving for N:

$$N_{max} = \frac{mN^*}{\frac{2d(m-1)N^*}{W} + 1}$$
(7.11)

where N^* was defined earlier in equation (7.5). Because $2d/W \ll 1$, equation (7.11) demonstrates that scalability improves proportionally with cluster size (*m*), a significant advantage of this design. To summarize, the use of mini-lenses with large clusters significantly improves lateral tolerance and scalability, without sacrificing on tilt tolerance. These results are consistent with the findings of Rolston et al. [22].

7.5.4 Design #4: microchannel telescope

While the clustering approach can offer loose lateral tolerances, it still suffers from tight tilt tolerances (same as design #2). This is explained by the fact that mini-lenses must have a long focal length, enough to meet the interconnection length requirement. The long focal length produces the equivalent of a lever arm, such that a small tilt of the module results in a large displacement of the spot on the detector. One way to solve this problem is to separate the tasks of relaying and focusing the beams by using two lens elements instead of one. This is shown in figure 7.6 below, where a microchannel telescope is formed by cascading a pair of microlens arrays of focal lengths $f_{\mu lensl}$ and $f_{\mu lens2}$.



Figure 7.6. Design #4: Microchannel telescope (shown with N = 4).

The first microlens array (µlens1) implements a Gaussian relay, so $f_{µlens1} = \pi \omega_o^{2/\lambda}$, while the second microlens array (µlens2) is used to generate a small spot size on the detector, $\omega_d = \lambda f_{µlens2}/\pi \omega_0$. This requires that $f_{µlens1} >> f_{µlens2}$, and so $f_{µlens1} \sim L/4$. Both microlenses have identical diameters, given by: $D_{µlens1} = D_{µlens2} = W/N$. To maximize lateral and tilt tolerances, the second microlens should be as fast as possible. In practice, the f-number is limited by the microlens technology and fabrication technique. It is assumed that f/2 microlenses are available which result in $f_{µlens2} = 2W/N$. The magnification factor, M, of the microchannel telescope is $M = f_{µlens1}/f_{µlens2} = NL/8W$.

As the chip module is displaced laterally, two different events occur simultaneously: (i) the spot is misaligned relative to the detector and (ii) the incident beam is clipped at the aperture of the first microlens. Which of these two events will satisfy the misalignment metric first depends on the value of the magnification factor, M. In the case where a 50% efficiency drop at the detector occurs first, we can write:

$$\Delta x_{efficiency} = \left(\frac{f_{\mu lens1}}{f_{\mu lens2}}\right) \frac{d}{2} = M \frac{d}{2}$$
(7.12)

Alternatively, in the case where the lateral tolerance is limited by clipping losses at the aperture of μ lens 1, we can write:

$$\Delta x_{clipping} = \frac{1}{2} (D_{\mu lens1} - 2k_l \omega_{\mu lens1}) = \frac{W}{2N} - k_l \sqrt{\frac{\lambda L}{2\pi}}$$
(7.13)

The module lateral tolerance is equal to the smallest of the two quantities given in equations (7.12) and (7.13):

$$\Delta x = \min(\max\{\Delta x_{efficiency}, \Delta x_{clipping}\}$$
(7.14)

A similar situation occurs for the tilt tolerance, which can be determined by either (i) a 50% efficiency drop at the detector or (ii) clipping losses at μ lens2. The tilt tolerance for these two cases are given by:

$$\Delta \theta_{efficiency} = \frac{d}{2f_{\mu lens2}} = M \frac{2d}{L}$$
(7.15)

$$\Delta \theta_{clipping} = \frac{D_{\mu lens2}/2 - k_l \omega_{\mu lens2}}{f_{\mu lens1} + f_{\mu lens2}} \cong \frac{\frac{2W}{NL} - k_l \sqrt{\frac{4\lambda}{\pi L}}}{1 + 1/M}$$
(7.16)

The module tilt tolerance is equal to the smallest of these two quantities:

$$\Delta \theta = \min(\max\{\Delta \theta_{efficiency}, \Delta \theta_{clipping}\}$$
(7.17)

The above results indicate that both lateral and tilt tolerance are improved by maximizing the magnification factor (M). If M is made large enough, then the tolerances are solely limited by clipping at the microlenses.

The maximum size of the array, N_{max} , is usually limited by clipping losses at µlens1. It can be calculated by setting Δx to zero in equation (7.13) and solving for N:

$$N_{max} = \frac{W}{k_l \sqrt{2\lambda L/\pi}}$$
(7.18)

Comparing this result with the 4-f microchannel of designs #1 and #2, one sees that a microchannel telescope is much more scalable: N_{max} is less sensitive to the interconnection length (L) due to the presence of the square root function in the denominator.

7.5.5 Design #5: microchannel with field lens array

A small variation of design #4 is obtained by locating μ lens2 exactly at the focal plane of μ lens1, as shown in figure 7.7. In this case, μ lens2 is referred to as a field lens [27]. The motivation for doing this is explained below.



Figure 7.7. Design #5: microchannel with field lens array (shown with N = 4).

As in design #4, both microlenses have identical diameters: $D_{\mu lens1} = D_{\mu lens2} = W/N$. The first microlens array implements a Gaussian relay, such that $f_{\mu lens1} = \pi \omega_0^2 / \lambda$. It is assumed that f/2 microlenses are available and the focal of $\mu lens2$ is set to $f_{\mu lens2} = 2W/N$. Since $f_{\mu lens1} >> f_{\mu lens2}$, then $f_{\mu lens1} \sim L/4$.

The distance between μ lens2 and the chip, s, is selected such that the image of the detector, imaged through μ lens2, falls exactly in the plane of μ lens1:

$$\frac{1}{s} + \frac{1}{f_{\mu lens1}} = \frac{1}{f_{\mu lens2}}$$
(7.19)

Since the image magnification is given by $M = f_{\mu lensl}/s$, equation (7.19) leads to:

$$s = \frac{f_{\mu lens1}}{(f_{\mu lens1} / f_{\mu lens2}) - 1} = \frac{f_{\mu lens1}}{M} \cong \frac{L}{4M}$$
(7.20)

which means that the magnification factor can be rewritten as:

$$M = \frac{f_{\mu lens1}}{f_{\mu lens2}} - 1 = \frac{LN}{8W} - 1$$
(7.21)

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

197
Unlike previous designs, the field lens approach of figure 7.7 is not a symmetric relay, and so the relationship of equation (7.3) does not apply and the detector spot size (ω_d) must be calculated using ABCD matrices [28]. The ABCD matrix corresponding to the relay of figure 7.7, from the mid-point of the beamsplitter (beam waist: ω_o) to the detector plane (beam waist: ω_d) is given by:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} -(s f_{\mu lens1}) & -s \\ -(1 f_{\mu lens1}) & -(f_{\mu lens1} f_{\mu lens2}) \end{bmatrix}$$
(7.22)

The detector spot size (ω_d) is calculated using the following set of equations:

$$q_d = \frac{Aq_o + B}{Cq_o + D} \tag{7.23}$$

$$\frac{1}{q_o} = -j\frac{\lambda}{\pi\omega_o^2} \tag{7.24}$$

$$Im\left\{\frac{1}{q_d}\right\} = -\frac{\lambda}{\pi\omega_d^2}$$
(7.25)

where "Im" means "the imaginary part of". This leads to the following expression for ω_d :

$$\omega_d = \sqrt{\frac{2\lambda s^2}{\pi f_{\mu lensl}}} = \frac{\sqrt{2}\omega_o}{M}$$
(7.26)

Considering that the beam waist at μ lens1 is $\sqrt{2}\omega_o$, equation (7.26) indicates that the magnification factor relating ω_d to ω_o is the same as what would be calculated using the laws of geometrical optics. Using the ABCD matrix method, it can also be shown that the fact that μ lens1 implements a Gaussian relay, combined with the choice of s in equation (7.20), leads to a minimum value for ω_d which is a fortunate consequence.

Depending on the magnification, the lateral tolerance may be limited by either a 50% drop in detector efficiency or clipping losses at μ lens1 (same as in design #4). As a result, equations (7.12), (7.13) and (7.14) apply, but with the value of M given in equation (7.21).

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

The motivation for using a field lens is that it can further relax tilt tolerances. This may be understood by examining the propagation of a Gaussian beam under conditions of lateral and tilt misalignments, as shown in figure 7.8. In the case of tilt misalignment about the center of μ lens1 (figure 7.8(b)), the field microlens (μ lens2) will always redirect the chief ray towards the center of the detector. This makes sense considering that *s* has been chosen such that the detector is imaged exactly in the plane of μ lens1; thus, if the chief ray traverses the center of μ lens1, then it must also terminate at the center of the detector.



Figure 7.8. Field microlens design under (a) lateral and (b) tilt misalignments.

As a result, the tilt tolerance is only limited by clipping losses at μ lens2:

$$\Delta \Theta = \frac{(D_{\mu lens2}/2) - k_l \omega_o}{f_{\mu lens1}} = \frac{2W}{NL} - k_l \sqrt{\frac{4\lambda}{\pi L}}$$
(7.27)

The important point is that the use of a field lens effectively desensitizes tilt tolerance from the size of the detector (i.e. $\Delta \theta$ is not dependent on *d* anymore). This allows for small-size detectors to be used without sacrificing any tilt tolerance (although lateral tolerance may be affected if $\Delta x_{efficiency} < \Delta x_{clipping}$ in equation (7.14)).

The maximum size of the array, N_{max} , is limited by clipping losses at the first microlens. It can be calculated by equating Δx to zero in equation (7.13) and leads to the same expression as equation (7.18), indicating a good scalability of design #5.

Table 7.2 C	losed-form expressions	s for lateral (Δx) and ti	lt misalignment ($\Delta \theta$) to	lerances and maximum	array size (N _{max}).
	Design #1	Design #2	Design #3	Design #4	Design #5 Field microlens
Lateral tolerance (Δx)	<u>d</u> 2	$\frac{W}{2N}\left(1-\frac{N}{N^*}\right)$	$\frac{W}{2N}\left(m-\frac{N}{N^*}\right)-(m-1)d$	$min:\left\{M\frac{d}{2}, \frac{W}{2N} - k_l \sqrt{\frac{\lambda L}{2\pi}}\right\}$	$min:\left\{M\frac{d}{2}, \frac{W}{2N}-k_l\sqrt{\frac{\lambda L}{2\pi}}\right\}$
Tilt tolerance $(\Delta \theta)$	$\sqrt{\frac{2d}{W}}$	<u>2d</u> L	$\frac{2d}{L}$	$min:\left\{M\frac{2d}{L},\frac{\frac{2W}{NL}-k_{I}\sqrt{\frac{4\lambda}{\pi L}}}{1+1/M}\right\}$	$\frac{2W}{NL} - k_l \sqrt{\frac{4\lambda}{\pi L}}$
Max. array size (N _{max})	$\frac{2\pi\omega_d W}{\lambda L k_l} = N^*$	$\frac{2\pi\omega_d W}{\lambda L k_l} = N^*$	$\frac{mN^*}{[2d(m-1)N^*]/W+1}$	$\frac{W}{k_{l}\sqrt{2\lambda L/\pi}}$	$\frac{W}{k_l\sqrt{2\lambda L/\pi}}$
Notes	Tight lateral; $\Delta x \propto d$ Loose tilt; square root Poorly scalable; $N \propto i/L$	Trade-off; N versus Δx Tight tilt: $\Delta \theta \propto d$ Poorly scalable; N $\propto 1/L$	m = cluster array size A large m is desirable Loose lateral: $\Delta x \propto m$ Tight tilt: $\Delta \theta \propto d$ Good scalability: $N \propto m$	M = NL/8W f-number of µlens2 = f/2 A large M is desirable Loose lateral and tilt Good scalability: $N \propto 1 \frac{1}{L}$	M = (NL/8W) - 1 f-number of µlens2 = f/2 A large M is desirable Loose lateral and tilt Good scalability: N∝ 1//L
Example #1 $W = 10$ mm, $\lambda = 850$ nm L = 25 mm, $d = 50$ µm $k_l = 2, 12, N = 16, m = 4$	$\Delta x = 25 \mu m$ $\Delta \theta = 344 arcmin$ $\Delta x \Delta \theta = 8594 \mu m, arcmin$	$\Delta x = 97 \mu m$ $\Delta \theta = 14 arcmin$ $\Delta x \Delta \theta = 1339 \mu m. arcmin$	$\Delta x = 882 \mu\text{m}$ $\Delta \theta = 14 \text{arcmin}$ $\Delta x \Delta \theta = 12129 \mu\text{m}, \text{arcmin}$	$\Delta x = 125 \mu m$ $\Delta \theta = 69 arcmin$ $\Delta x \Delta \theta = 8594 \mu m, arcmin$	$\Delta x = 100 \mu\text{m}$ $\Delta \theta = 124 \text{arcmin}$ $\Delta x \Delta \theta = 12394 \mu\text{m.arcmin}$
Example #2 $W = 10 \text{ mm}, \lambda = 850 \text{ nm}$ $L = 50 \text{ mm}, d = 20 \mu \text{ m}$ $k_1 = 2.12, N = 16, m = 4$	Not scalable to a 16 × 16 array	Not scalable to a 16 × 16 array	$\Delta x = 114 \mu m$ $\Delta \theta = 3 arcmin$ $\Delta x \Delta \theta = 315 \mu m. arcmin$	$\Delta x = 100 \mu m$ $\Delta \theta = 28 arcmin$ $\Delta x \Delta \theta = 2750 \mu m. arcmin$	$\Delta x = 90 \mu m$ $\Delta \theta = 52 arcmin$ $\Delta x \Delta \theta = 4683 \mu m, arcmin$

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

200

7.5.6 Summary of results

For convenience, the expressions for Δx , $\Delta \theta$, and N_{max} for all five designs are summarized in table 7.2. Relevant comments relative to the relative strengths and weaknesses of each design are also included.

In addition, two numerical examples are included in the table. In both examples, the chip measures 10 mm on a side and supports a 16×16 array of detectors. Following the conclusions of Belland and Crenn [25], the minimum clipping ratio is selected to be $k_l = 2.12$; this ensures that clipping losses are limited to less than 0.1% and that the diffraction effects do not significantly modify the Gaussian beam characteristics. The first example corresponds to an interconnection length of L = 25 mm and detectors with d = 50 µm. The second example assumes L = 50 mm and d = 20 µm.

Although the numerical values resulting from these examples do not represent the actual misalignment budget of a practical implementation (this is because the misalignment metric of section 7.3 is not conservative enough), they provide a common ground from which the different designs can be compared.

7.6 Discussion

One approach at comparing the alignability of the five designs is to consider the alignment product ($\Delta x \Delta \theta$) as the figure of merit. Using this metric, a close examination of the alignment product for the first example indicates that design #5 is the most misalignment-tolerant whereas design #2 is the least. However, using the value of $\Delta x \Delta \theta$ as the only measure of alignability can be misleading. For example, although the module of design #1 has the same value of $\Delta x \Delta \theta$ as design #4, it is far more difficult to align because of the tight lateral tolerance, especially if small detectors are used (remember that $\Delta x = 25 \,\mu m$ results from a 50% efficiency drop misalignment metric while in practice a 1% metric is more likely to be used). The module of design #4 is much easier to align because the tolerances are better distributed between the lateral and tilt DOFs. To summarize: a large value of $\Delta x \Delta \theta$ is a necessary but not sufficient condition for a misalignment-tolerant design, the additional condition being that the alignment product be well distributed between Δx and $\Delta \theta$.

Applying the latter conclusion to the results of the first example indicates that only designs #4 and #5 are misalignment-tolerant, the other designs suffering from either a tight Δx (design #1) or a tight $\Delta \theta$ (designs #2 and #3) because they are directly proportional to the size of the detector. This indicates that integrating a single lens array component with the chip does not solve the alignment problem; it simply means that the tight tolerance is being shifted from the lateral to the tilt DOFs.

Fundamentally, the poor alignability of designs #1, #2 and #3 originates from the fact that a single lens is used to perform both the tasks of beam relaying and spot focusing. These are conflicting requirements because (i) a relay lens requires a long focal length while (ii) a focusing lens requires a small *f*-number. Trying to satisfy both requirements simultaneously leads to large-diameter lenses which limits scalability. This is the reason why designs #1 and #2 cannot scale up to a 16×16 array in the second example.

The clustering configuration of design #3 mitigates this problem by allowing more than one beam per lens. The use of larger clusters (large m) maximizes the diameter of the mini-lens, which leads to: (i) an increase in lateral tolerance ($\Delta x \propto m$) and (ii) an improvement in scalability ($N_{max} \propto m$). This explains why design #3 can easily scale up to a 16 × 16 array in the second example. Nevertheless, design #3 still suffers from having a tight tolerance, due to the fact that $\Delta \theta$ is directly proportional to d.

The alignability can be improved by integrating a second lens array with the chip module (designs #4 and #5). This way, a slow lens is used for relaying while a fast lens is used for focusing. This has two significant advantages: First, it allows for a scalable design, as indicated by the expression for N_{max} which is now independent of ω_d and less sensitive to L (see equation (7.18)). Second, the lateral and tilt tolerances associated with the 50% efficiency drop metric are multiplied by the magnification factor, M, which results in loose tolerances and a better balance between Δx and $\Delta \theta$.

These observations are best visualized using figure 7.9, which shows the module of design #4 under lateral and tilt misalignment conditions. In this case, the detector is magnified (by a factor M) and imaged an optical distance $f_{\mu \text{lensI}}$ in front of the module. Because the microchannel telescope is a telecentric relay, the trajectory and magnification of the Gaussian beam is the same as that of the detector image. This means that the coupling effi-

ciency of the Gaussian spot size on the detector is equivalent to the overlap of the incident beam on the detector image. Using this and referring to figure 7.9, the lateral and tilt tolerances are readily seen to be equal to $\Delta x = Md/2$ and $\Delta \theta = (Md/2)/f_{\mu \text{lens1}} = 2Md/L$, which is the same as equations (7.12) and (7.15) respectively.



Figure 7.9. Design #4: (a) lateral and (b) tilt misalignment (50% efficiency drop metric).

Usually, $f_{\mu \text{lensi}}$ is fixed by the interconnection length ($f_{\mu \text{lensi}} \sim L/4$) and so the magnification factor can only be increased by reducing $f_{\mu \text{lens2}}$. Beyond a certain point, increasing *M* does not improve tolerances because Δx and $\Delta \theta$ are limited by clipping losses at μlens1 and μlens2 respectively. When this happens, Δx and $\Delta \theta$ become independent of *d* which means that a smaller detector size can be used.

Figure 7.9(b) helps visualizing the reason why the tilt tolerance of design #4 is limited by the size of the detector while this is not the case for design #5. This is because the detector image in design #4 is located one focal length in front of the plane where the chip module interfaces with the beamsplitter cube. Consequently, a tilt misalignment about the center of μ lens1 results in the detector image being displaced by an amount $\Delta\theta \times f_{\mu}$ with respect to the incident Gaussian beam. The displacement of the detector image under tilt misalignment is avoided in design #5 because the detector image is located exactly in the plane of μ lens1. The general conclusion is that the separation between μ lens2 and the chip should always be chosen such that the detector image falls at the interface between the modules because this desensitizes the tilt tolerance from the detector size.

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

misalignment tolerances. A proper balance is achieved by designing the interconnect with slow beams on both sides of the optical elements. This observation lead him to conclude that Gaussian relays represent the most misalignment-tolerant method of interconnecting free-space modules.

7.8 Invariance of the alignment product

Consider the simple case of the module in figure 7.11, which integrates of a single detector of size d. The detector is imaged (from right to left) through an optical relay which may consist of one or more lens elements. The optical design is such that the detector image falls in the plane of the inter-module interface. It is assumed that the magnification factor (M) results in the size of the detector image (d') being smaller than the clear aperture of the optics such that Δx is limited by d' and not by clipping losses.



Figure 7.11. Derivation of the relationship between $\Delta x \Delta \theta$ and the optical invariant.

The image of the detector is drawn using one oblique ray and one axial ray. The angle the oblique ray makes with the optical axis is equal to θ ' and θ in the image and object plane respectively. These angles correspond to the entrance and exit numerical aperture (NA) of the optical system. Using paraxial ray trace equations, it can be shown that the following relationship is always true, irrespective of the design of the optical relay [29]:

$$\frac{d}{2}\theta = \frac{d'}{2}\theta' \tag{7.28}$$

The quantity given in equation (7.28) does not vary with the optical system, and for this reason, it is referred to as the optical invariant.

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

The chief ray of a Gaussian beam propagating through an optical system always follows the same trajectory as that of a light ray [30]. Referring to figure 7.11, consider a Gaussian beam incident from the left and laterally misaligned by an amount Δx . This beam is coincident with the axial ray and so will follow the same trajectory and terminate at the edge of the detector, yielding a coupling efficiency of 50%. Note that this case is equivalent to a module being laterally misaligned by an amount equal to half of the detector image size: $\Delta x = d'/2$. Second, consider the Gaussian beam incident at an angle $\Delta \theta$; this beam is coincident with the oblique ray and thus terminates at the center of the detector. In this case, the Gaussian beam is clipped at the aperture of the optics and so the module tilt misalignment is directly proportional to the entrance NA: $\Delta \theta \propto \theta'$. Using this and equation (7.28), one can write:

$$\Delta x \Delta \theta \propto \frac{d}{2} \theta \tag{7.29}$$

Equation (7.29) indicates that the alignment product is directly related to the optical invariant, which is given by the product of the detector size and the NA of the optics on the side of the detector. This is an important result which signifies that: (i) the alignment product is also an invariant of the system and (ii) misalignment-tolerant modules have in common that they have a large optical invariant. The alignment product is maximized by:

- Increasing *d*: which means using large detectors, which is obvious. The detector size is limited by its associated capacitance which usually impacts the receiver bit rate (see section 2.5).
- Increasing θ: which means designing the optical relay with a large NA on the side of the detector. Fortunately, detectors have a very large acceptance angle, which far exceeds the NA of the optical relay. The use of high-NA optics effectively takes advantage of this attribute. This is to be compared to the situation where the detector is replaced with an optical fibre; in this case, the alignment product may be significantly reduced, due to the limited NA of the fibre.

The fact that the alignment product is directly linked to the optical invariant provides a quick and easy way of assessing the alignability of an optical design. One simply needs to determine the value of the optical invariant in the plane of the detector; if the invariant is small (due to a small detector and/or a limited NA of the optical relay) then the design is inherently difficult to align. For example, the poor alignability of designs #1 and #2 is easily justified by the fact that they both have a small optical invariant, limited by the small NA of the slow microlenses. Furthermore, the clustering configuration of design #3 increases the diameter of the lenses by a factor of m compared to design #2; this increases the optical invariant by m and so it should not be surprising that the alignment product is increased by about the same factor. Finally, it now becomes apparent that designs #4 and #5 are inherently misalignment-tolerant because they both use a low f-number (large NA) microlens close to the detector plane, leading to a large optical invariant. Note that the optical invariant of design #4 is slightly larger than the one of design #5 because $f_{\mu lens2} < s$ (see equation (7.19)). However, design #5 provides a superior alignment product because the field lens places the image of the detector exactly in the plane corresponding to the module interface.

Using the same assumptions used in table 7.2, expressions for the optical invariant have been derived and are given in table 7.3. Although the optical invariant of design #1 is large (due to the large detector NA), it is still very difficult to align due to the poor balance between lateral and tilt tolerances. Note that the integration of a second lens element (designs #4 and #5) results in the optical invariant being multiplied by the magnification factor M compared to the single-lens module of design #2.

Module Design	Optical Invariant	Notes
Design #1 (no optics)	(d/2)NA _{detector}	NA _{detector} is usually large
Design #2 (microchannel)	dW/NL	
Design #3 (clustering)	$\frac{dW}{NL}\left[m-\frac{2d(m-1)N}{W}\right]$	m = cluster size
Design #4 (micro-telescope)	M(dW/NL)	M = NL/8W
Design #5 (field microlens)	M(dW/NL)	M = (NL/8W) - 1

Table 🕻	7.3. Ex	pressions	for the o	ptical	invariant	toft	he mod	lule desig	ns.
---------	---------	-----------	-----------	--------	-----------	------	--------	------------	-----

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

This result is significant because it had previously been thought that optical interconnects did not suffer from an aspect-ratio limitation (see [32]). From a purely theoretical standpoint, if a FSOI system can be aligned with infinite accuracy, then $\Delta x = \Delta \theta = 0$, which makes the right side of equation (7.32) go to infinity. Thus, FSOI systems do not suffer from this aspect-ratio limitation only if they be can aligned with infinite accuracy. Since practical systems will invariably have some degree of misalignment, their bandwidth capacity will be limited by the aspect-ratio of the system in a way that is similar to electrical interconnects. The value of the constant relating both sides of equation (7.32) is likely to be orders of magnitude larger and also depends on the details of the optical design. For example, the corresponding expression for designs #4 and #5 is:

$$B \propto \frac{M^2}{(\Delta x \Delta \theta)^2} \times \frac{A}{L^2}$$
 (designs #4 and #5) (7.33)

indicating that designs #4 and #5 can relax the alignment product by a factor of M while still providing the same bandwidth performance as design #2. Conversely, if designs #4 and #5 can be aligned to the same accuracy as design #2, then their bandwidth capacity can be increased by a factor of M^2 (because designs #4 and #5 are more scalable).

7.10 Conclusion

This chapter addressed the issue of alignment in 2D-FSOI systems by investigating the alignability of different optical configurations used for the design of a chip module. It was shown that the alignability of a module can be adequately specified by the product of its lateral (Δx) and tilt ($\Delta \theta$) misalignment tolerance. This "alignment product" is an invariant of the optical system. Although the alignment product is the same at any plane in the optical system, what changes is the relative distribution between Δx and $\Delta \theta$. Thus, a necessary condition for misalignment-tolerant modules is that the alignment product be (i) large and (ii) properly balanced between Δx and $\Delta \theta$.

A large Δx is obtained at planes where the signal beams are large in size. Conversely, a large $\Delta \theta$ is obtained at planes where the signal beams are small in size. A good balance between these conflicting beam requirements is achieved by implementing a Gaussian relay at the inter-module interface. The consequence of this is that a second lens element is required to focused the beam to a small spot on the detector. This lens should be designed with the lowest *f*-number possible because this maximizes the alignment product. Gaussian relays have the additional advantage that they are scalable to dense arrays.

The use of a clustering configuration was shown to improve both misalignment tolerances and array scalability, the amount of improvement scaling proportionally with the size of the cluster. Improved designs can be realized by combining, for example, the advantages of field lenses with a clustering configuration. These novel designs may allow future FSOI systems to be assembled using purely mechanical alignment techniques.

A significant outcome of this work is the demonstration that practical FSOI systems suffer from an aspect-ratio limitation similar to the one found in electrical interconnects. The aspect-ratio limitation of FSOI systems can be relaxed by using misalignment-tolerant designs.

7.11 References

- F. Lacroix, M. Châteauneuf, X. Xue, and A. G. Kirk, "Experimental and numerical analyses of misalignment tolerances in free-space optical interconnects," Applied Optics, vol. 39, pp. 704-713 (2000).
- [2] F. Lacroix, B. Robertson, M. H. Ayliffe, E. Bernier, F. A. P. Tooley, M. Chateauneuf, D. V. Plant, A. G. Kirk, "Design and implementation of a four-stage clustered free-space optical interconnect," Optics in Computing '98, Brugge, Belgium, 17-20 June 1998, pp.107-110.
- [3] V. N. Morozov, Y.-C. Lee, J. A. Neff, D. O'Brien, T. S. McLaren, H. Zhou, "Tolerance analysis for three-dimensional optoelectronic systems packaging," Optical Engineering, vol. 35, pp. 2034-2043 (1996).
- [4] F. Lacroix and A. G. Kirk, "Tolerance stackup effects in optical interconnect systems," in Optics in Computing 2000, R. A. Lessard and T. Galstian eds., SPIE 4089, 896-904 (2000).

Chapter 7: Misalignment-tolerant modules for free-space optical interconnects

using symetrical self-electro-optic-effect devices," Applied Optics, vol. 32, pp. 5153-5171 (1993).

- [13] A. L. Lentine, K. W. Goossen, J. A. Walker, J. E. Cunningham, W. Y. Jan, T. K. Woodward, A. V. Krishnamoorthy, B. J. Tseng, S. P. Hui, R. E. Leibenguth, L. M. F. Chirovsky, R. A. Novotny, D. B. Buchholz, R. L. Morrison, "Optoelectronic VLSI switching chip with > 1 Tbit/s potential optical I/O bandwidth," Electronics Letters, vol. 33, pp. 894-895 (1997).
- [14] A. C. Walker, T. Y. Yang, J. Gourlay, J. A. B. Dines, M. G. Forbes, S. M. Prince, D. A. Baillie, D. T. Neilson, R. Williams, L. C. Wilkinson, G. R. Smith, M. P. Y. Dezmulliez, G. S. Buller, M. R. Taghizadeh, A. Waddie, I. Underwood, C. R. Stanley, F. Pottier, B. Vogele, and W. Sibbett, "Optoelectronic systems based on InGaAs-complementary-metal-oxide-semiconductor smart-pixel arrays and free-space optical interconnects," Applied Optics, vol. 37, pp. 2822-2830 (1998).
- [15] M. Yamaguchi, T. Yamamoto, K. Hirabayashi, S. Matsuo, K. Koyabu, "High-density digital free-space photonic-switching fabrics using exciton absorption reflectionswitch (EARS) arrays and microbeam optical interconnections," IEEE Journal of Selected Topics in Quantum Electronics, vol. 2, pp. 47-54 (1996)
- [16] D. Z. Tsang, T. J. Goblick, "Free-space optical interconnection technology in parallel processing systems," Optical Engineering, vol. 33, pp. 1524-1531 (1994).
- [17] T. Sakano, T. Matsumoto, K. Noguchi, "Three-dimensional board-to-board freespace optical interconnects and their application to the prototype multiprocessor system: COSINE-III," Applied Optics, vol. 34, pp. 1815-1822 (1995).
- [18] D. J. Goodwill, D. Kabal, P. Palacharla, "Free-space optical interconnect at 1.25 Gb/ s/channel using adaptive alignment," in Proceedings of Optical Fiber Communication Conference, vol. 2, pp. 259-261 (1999).
- [19] X. Zheng, P. J. Marchand, D. Huang, O. Kibar, N. S. E. Ozkan, S. C. Esener, "Optomechanical design and characterization of a printed-circuit-board-based free-space optical interconnect package," Applied Optics, vol. 38, pp. 5631-5640 (1999).

Parallel Computing with Optical Interconnects, J. Parallel and Distributed Computing, vol. 41, pp. 42-52 (1997).

[32] D. A. B. Miller, "Rationale and challenges for optical interconnects to electronic chips," IEEE Proceedings, pp. 1-44 (2000).



Chapter 8: Conclusion

8.1 Summary

In principle, two-dimensional parallel optical interconnects (2D-POIs) can solve most of the problems encountered in electrical interconnections, which includes: signal distortion, skew, crosstalk, attenuation, impedance matching, electromagnetic interference, limited interconnection density and high-power dissipation. 2D-POI technologies offer the promise of providing Tbit/s data communication between silicon VLSI chips using lowpower (~ 1 mW) dense interconnections (> 1000 channels/cm²) operating at on-chip data rates (> 1 Gbit/s). In particular, two-dimensional free-space optical interconnects (2D-FSOIs) are expected to fulfill the requirements of short-distance (1 - 10 cm) applications that require highly parallel multi-point interconnections between multiple VLSI chips located in different physical planes.

One major obstacle that prevents the commercial deployment of 2D-FSOI systems is the problem of alignment which originates from the fact that these systems are constructed using discrete array components that must be accurately aligned in six degrees of freedom (DOFs). This problem is further exacerbated by the requirements that the systems be fieldserviceable and able to sustain the harsh conditions of industrial environments.

The work presented in this thesis addresses the alignment problem of 2D-FSOI systems and has significance in three areas. First, it has made a major contribution to the design and implementation of a large-scale photonic backplane demonstrator system conceived as a vehicle to demonstrate Tbit/s free-space optical interconnections between silicon VLSI chips. Secondly, it proposes a generic packaging strategy, which consists of partitioning the optical system into separate modules in such a way that the loose tolerances are between the modules while the tight tolerances are between the components inside the modules. Novel mechanical and optical methods for aligning both components and modules have been designed and demonstrated. Thirdly, it identifies the fundamental reasons behind the tolerances of 2D-FSOI systems and proposes a set of guidelines for the design of modules that are inherently tolerant to misalignment. Conclusions drawn from this research are described below. supported by the simulation results obtained by Lacroix and Kirk [2]. Thus, for a more realistic misalignment budget to be calculated, a rigorous approach that includes the combined effects of all six DOFs is required (e.g. Monte-Carlo simulations).

- For purely mechanical alignment to be used in this system, the misalignment budget must be (i) further relaxed and (ii) better distributed between the lateral and tilt DOFs. The design of the system demonstrator was biased towards lateral misalignment tolerances at the expense of the tilt tolerances, making the overall alignment difficult.
- The semi-kinematic fixture of the chip module was shown to be highly repeatable. The design takes advantage of the optical-grade flatness of two optical substrates to fix the tilt and longitudinal DOFs, while a pair of dowel pins fix the lateral and rotational DOFs. The success of this approach relies on the fact that the accuracy of the alignment structures (precise substrate flatness, dowel pins having a loose fit) matches the requirements of the misalignment budget (tilt is tight, lateral is relaxed).
- The use of a kinematic fixture, fabricated using ultrathick photoresist micro-structures, was found to be inadequate for inter-module mechanical alignment, due to problems related to the lack of uniformity, hardness and robustness of the micro-structures. However, it is believed that ultrathick photoresist technology is a good candidate for intra-module mechanical alignment, where components do not have to be separable.
- The use of the chip-on-board (COB) approach for packaging OE-VLSI chips was proven to be successful. This packaging scheme allows for (i) the close integration of a mini-lens array with the chip, (ii) an excellent thermal connection to the back of the chip for temperature control, (iii) high-speed performance due to the absence of package lead inductance, and (iv) the ability of placing surface mount components (e.g. decoupling capacitors) in close proximity to the chip. The flex-PCB can support > 1 Gbit/s/line data rates while providing the required mechanical flexibility.
- Diffractive elements such as Fresnel Zone Plates (FZPs) or amplitude gratings can be laid out using the top-level metal layer of a standard CMOS process. These optical elements can be used to perform the alignment of a lens array to a chip. The use of off-axis diffractive elements can provide alignment sensitivity in all six DOFs.

8.3 Packaging strategy and alignment techniques

Building on the experience of the backplane demonstrator system, this thesis proposes a generic packaging strategy whose objective is to facilitate the alignment of 2D-FSOI systems. The strategy divides the alignment problem into three steps as follows:

- System partitioning: the optical system is first partitioned into separate modules in such a way that the misalignment tolerances between modules be relaxed as much as possible. This is done to facilitate the alignment of modules during system assembly and servicing.
- Intra-module alignment: the second step consists of aligning components into modules. Intra-module alignment requires a high level of accuracy and so there is a need for dedicated micro-packaging infrastructures using high-precision positioning equipment and specialized techniques for handling and fixing miniature components. In many cases, the use of optical techniques might be the only way to achieve the required level of accuracy. Techniques that are compatible with automated assembly are preferable. Wherever possible, the air gaps within a module should be filled with glass spacers to create a solid subassembly.
- Inter-module alignment: the last step concerns the alignment of modules to one another. To ensure field-serviceability of the system, it is important that the inter-module interface allows for the removal and replacement of a defective module without upsetting system alignment. To reduce cost and complexity, this alignment step should be accomplished using purely mechanical methods. To do this, misalignment-tolerant designs are required.

A broad range of intra-module and inter-module alignment techniques, both optical and mechanical, have been discussed in chapters 5 and 6. For convenience, a summary of these techniques is shown in table 8.1. The technique of device array redundancy falls in a category of its own.

Category	Intra-module alignment	Inter-module alignment		
Mechanical	flip-chip technique silicon micromachining micro-connectors	slotted baseplate L-shaped structure precision dowel pins		
Optical	fiducial markers in-situ interferometry diffractive structures Moiré techniques	wedge prisms, tilt plates variable-angle prisms liquid-crystal devices optical MEMS		
Electronic		device array redundancy		

Table 8.1. Summary of intra-module and inter-module alignment techniques.

8.4 Design of misalignment-tolerant 2D-FSOI modules

This work examined the alignment problem of 2D-FSOI modules by investigating various approaches to the design of misalignment-tolerant modules. The following conclusions are drawn, which are relevant to all designers of optical packaging:

- The alignability of a module can be adequately specified by the product of its lateral (Δx) and tilt (Δθ) misalignment tolerance.
- The alignment product of misalignment-tolerant modules must satisfy two conditions:
 (i) its value must be large and (ii) it must be properly distributed between Δx and Δθ.
- The alignment product is an invariant of the system. At different planes of the system, the product is conserved but the relative distribution between Δx and $\Delta \theta$ changes.
- To achieve a proper balance between Δx and Δθ, 2D-FSOI systems should be designed with Gaussian relays at the interface between modules. Gaussian relays have the additional advantage that they are scalable to dense arrays.
- Misalignment-tolerant chip modules must integrate two lens arrays. This allows to separate the tasks of beam relaying and spot focusing.
- The alignment product is maximized by (i) using large detectors and (ii) using the lowest possible *f*-number lens in front of the detector. The position of the focusing lens

be reliably produced in high volume [8]. The main disadvantage of injection-molded plastics is its large setup cost, making this technology prohibitively expensive when only a few pieces are required, as has been the case for 2D-FSOI demonstrator systems to date.

The general conclusion is that cost-effective intra-module alignment solutions require that all packaging aspects be taken into consideration during the design, development and processing of the optical components.

8.5.2 Inter-module issues

The work of chapter 7 should impact the optical designs of future 2D-FSOI systems. At the time of writing, the design of chip modules based on a microchannel telescope or using a field microlens have not yet been implemented. These designs and others should be prototyped and validated. VCSEL-based chip modules should be investigated as well. Progress made in this area is important because it will initial reduce the need for beamsteering devices and may ultimately lead to purely mechanical solutions to the inter-module alignment problem.

In addition, the use of array redundancy techniques must be investigated further. Previous laboratory prototypes [9][10] have been limited to a few optical channels. Largescale demonstrations, using dense arrays, are required. There still remains several open questions concerning the use of redundant arrays in large-scale systems, including the layout of the OE devices to avoid blind spots, the optimal method of monitoring beam alignment, the design of the control algorithm and the complexity of the circuitry that is required to route the selected photocurrents to the receiver array.

It is believed that the use of a snap-together mechanical interface, complemented by a small degree of array redundancy, may represent the solution of choice of future 2D-FSOI systems.

8.6 References

 E. Bernier, F. Lacroix, M. H. Ayliffe, B. Robertson, F. A. P. Tooley, D. V. Plant, A.
 G. Kirk, "Implementation of a compact, four-stage, scalable optical interconnect," Optics in Computing 2000 Conference, June 18-23 Québec City, Qc., Canada.

- [2] F. Lacroix and A. G. Kirk, "Tolerance stackup effects in optical interconnect systems," in Optics in Computing 2000, R. A. Lessard and T. Galstian eds., SPIE 4089, 896-904 (2000).
- [3] M. Châteauneuf, F. Thomas-Dupuis, A.G. Kirk, "Design, implementation and characterization of a folded spot array generator for a modulator-based free-space optical interconnect," in Optics in Computing 2000, R. A. Lessard, T. Galstian, Proc. SPIE 4089, pp. 263-271 (2000).
- [4] W. Hunziker, "Low-cost packaging of semiconductor laser arrays," IEEE Circuits and Devices Magazine, vol. 131, pp. 19-25 (1997).
- [5] K. Kurata, "Mass production techniques for optical modules," 48th IEEE Electronic Components & Technology Conference, pp. 572-580 (1998).
- [6] P. Tuteleers, P. Vynck, H. Ottevaere, V. Baukens, G. Verschaffelt, S. Kufner, M. kufner, A. Hermanne, I. Veretennicoff, H. Thienpont, "Technological aspects of deep proton lithography for the fabrication of micro-optical elements for photonics in computing applications," SPIE vol. 3490, pp.409-11 (1998).
- [7] D. T. Neilson, E. Schenfeld, "Plastic modules for free-space optical interconnects", Applied Optics, vol. 37, pp. 2944-2952 (1998).
- [8] G. Kritchevsky, J. Schaefer, "Plastic optics offer unique design freedom," Laser Focus World, Supplement to October 1997 issue, pp.S19-S24 (1997).
- [9] D. J. Goodwill, D. Kabal, and P. Palacharla, "Free-space optical interconnect at 1.25 Gb/s/channel using adaptive alignment," in Digest of Topical Meeting on Optical Fiber Communication (Optical Society of America, Washington, D. C., 1999), paper WM22.
- [10] D. V. Plant, E. Bernier, E. Bisaillon, M. Mony, M. Salzberg, T. Yamamoto, D. J. Goodwill, and A. G. Kirk, "A 5 Gb/s, 2-channel bi-directional adaptive redundant FSOI demonstrator system," in Optics in Computing 2000, R. A. Lessard, T. Galstian, Proc. SPIE 4089, pp. 465-472 (2000).

This appendix contains the mechanical drawings of the parts used in the chip module. The following drawings are included: the mini-lens holder, the mounting spacer, the flex-PCB mount, the thermally isolating spacer, the copper heat spreader and the protective cover. The heatsink was procured off-the-shelf.

The drawings were made using Autocad version 14. They conform to the ANSI Y14.5M 1982 standard.































Appendix A: Mechanical drawings of the chip module assembly



A-20








A-23

Appendix A: Mechanical drawings of the chip module assembly



A-24

Appendix B: Mathematical derivations for intra-module alignment technique #1

This appendix is a complement to section 5.4; it presents the mathematical equations that are used to calculate the six-DOF misalignment of the chip given the photocurrent measurements of three quadrant detectors.

B.1 Coordinate system definitions

In the derivation that follows, three different coordinate systems are used:

- (x, y, z) coordinate system: this is an *absolute* coordinate system, in the sense that it is independent of the position and orientation of the chip. The origin of this coordinate system is chosen to be centered with the mini-lens array, one focal length behind the mini-lenses. This origin corresponds to the center of a perfectly aligned chip. The x and y axis follow the row and column directions of the mini-lens array.
- (x', y', z') coordinate system: the origin of this coordinate system is always located at the center of the chip. The x' and y' axis follow the horizontal and vertical directions of the chip. The x'-y' plane is defined as the front surface of the chip. This is a *relative* coordinate system because it is defined with respect to the chip center.
- (u_i, v_i) coordinate system: the origin of this coordinate system is located at the center of quadrant detector *i* (with i = [1, 2, 3]). The u_i and v_i axis follow the horizontal and vertical directions of the detector. The u_i - v_i plane is defined as the front surface of the detector (which is the same as the front surface of the chip). This is a *relative* coordinate system because it is defined with respect to the detector center.

B.2 Solving for the spot positions

A quadrant detector is drawn in figure B.1. There is a minimum of three quadrant detectors on the chip and they are labelled i = 1, 2 and 3. The elements of detector i are labelled a_{i} , b_{i} , c_{i} , and d_{i} and they each generate a photocurrent I_{ai} , I_{bi} , I_{ci} , and I_{di} respec-

tively. The incident beam is focused to a spot located at coordinates (u_{xi}, v_{yi}) on detector *i*; the spot is assumed to have a Gaussian intensity profile and a radius ω :

$$I(u_i, v_i) = I_o \exp\left\{\frac{-2[(u_i - u_{xi})^2 + (v_i - v_{yi})^2]}{\omega^2}\right\}$$
(B.1)



Figure B.1. Representation of quadrant detector *i* and its relevant parameters.

The optical power incident on a given element is obtained by solving the overlap integral between the intensity distribution and the element area. For example, the optical power, P_{ai} , falling on element a_i is given by:

$$P_{ai} = \int_{\frac{g}{2} - \frac{g}{2} - d}^{\frac{g}{2} + d} \int_{0}^{-\frac{g}{2}} I_{o} \exp\left\{\frac{-2\left[\left(u_{i} - u_{xi}\right)^{2} + \left(v_{i} - v_{yi}\right)^{2}\right]}{\omega^{2}}\right\} du_{i} dv_{i}$$
(B.2)

which leads to:

$$P_{a\bar{i}} = \frac{I_o \omega^2 \pi}{8} \times \left[erf\left\{ \frac{-g/2 - u_{x\bar{i}}}{\omega/\sqrt{2}} \right\} - erf\left\{ \frac{-g/2 - d - u_{x\bar{i}}}{\omega/\sqrt{2}} \right\} \right] \times \left[erf\left\{ \frac{g/2 + d - v_{y\bar{i}}}{\omega/\sqrt{2}} \right\} - erf\left\{ \frac{g/2 - v_{y\bar{i}}}{\omega/\sqrt{2}} \right\} \right]$$
(B.3)

where the function erf(x) is defined as follows:

$$erf(x) = \int_{0}^{x} \exp[-t^{2}]dt$$
(B.4)

The photocurrents I_{ai} , I_{bi} , I_{ci} , and I_{di} are obtained by multiplying the overlapping optical power by the detector responsivity, R:

$$I_{ai} = \frac{RI_o\omega^2\pi}{8} \times \left[erf\left\{\frac{-g/2 - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{-g/2 - d - u_{xi}}{\omega/\sqrt{2}}\right\} \right] \times \left[erf\left\{\frac{g/2 + d - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - v_{yi}}{\omega/\sqrt{2}}\right\} \right]$$
(B.5)

$$I_{bi} = \frac{RI_o \omega^2 \pi}{8} \times \left[erf\left\{\frac{g/2 + d - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - u_{xi}}{\omega/\sqrt{2}}\right\} \right] \times \left[erf\left\{\frac{g/2 + d - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - v_{yi}}{\omega/\sqrt{2}}\right\} \right]$$
(B.6)

$$I_{ci} = \frac{RI_o\omega^2\pi}{8} \times \left[erf\left\{\frac{g/2 + d - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - u_{xi}}{\omega/\sqrt{2}}\right\} \right] \times \left[erf\left\{\frac{-g/2 - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{-g/2 - d - v_{yi}}{\omega/\sqrt{2}}\right\} \right]$$
(B.7)

$$I_{di} = \frac{RI_o\omega^2\pi}{8} \times \left[erf\left\{\frac{-g/2 - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{-g/2 - d - u_{xi}}{\omega/\sqrt{2}}\right\} \right] \times \left[erf\left\{\frac{-g/2 - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{-g/2 - d - v_{yi}}{\omega/\sqrt{2}}\right\} \right]$$
(B.8)

It is useful to define the following functions M_x and M_y :

$$M_{xi} = \frac{(I_{bi} + I_{ci}) - (I_{ai} + I_{di})}{I_{ai} + I_{bi} + I_{ci} + I_{di}}$$
(B.9)

$$M_{yi} = \frac{(I_{ai} + I_{bi}) - (I_{ci} + I_{di})}{I_{ai} + I_{bi} + I_{ci} + I_{di}}$$
(B.10)

which results in M_{xi} being a function of u_{xi} only, and M_{yi} a function of v_{yi} only:

$$M_{xi} = \frac{erf\left\{\frac{g/2 + d - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - u_{xi}}{\omega/\sqrt{2}}\right\} + erf\left\{\frac{g/2 + u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 + d + u_{xi}}{\omega/\sqrt{2}}\right\}}{erf\left\{\frac{g/2 + d - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - u_{xi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 + u_{xi}}{\omega/\sqrt{2}}\right\} + erf\left\{\frac{g/2 + d + u_{xi}}{\omega/\sqrt{2}}\right\}}$$
(B.11)

$$M_{yi} = \frac{erf\left\{\frac{g/2 + d - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - v_{yi}}{\omega/\sqrt{2}}\right\} + erf\left\{\frac{g/2 + v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 + d + v_{yi}}{\omega/\sqrt{2}}\right\}}{erf\left\{\frac{g/2 + d - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 - v_{yi}}{\omega/\sqrt{2}}\right\} - erf\left\{\frac{g/2 + v_{yi}}{\omega/\sqrt{2}}\right\} + erf\left\{\frac{g/2 + d + v_{yi}}{\omega/\sqrt{2}}\right\}}$$
(B.12)

In order to determine the six-DOF misalignment of the chip with respect to the minilens array, one must solve for the positional coordinates of points P_1 , P_2 , P_3 in the x-y-z coordinate system:

$$P_i = (x_i, y_i, z_i)$$
 i=1,2,3. (B.14)

This is done by equating the sides of triangle P_1 - P_2 - P_3 in both coordinate systems. The side $\overline{P_iP_j}$ of the triangle has a length:

$$\overline{P_i P_j} = \sqrt{\left[(x_{di}' + u_{xi}) - (x_{dj}' + u_{xj}) \right]^2 + \left[(y_{di}' + v_{yi}) - (y_{dj}' + v_{yj}) \right]^2} \quad i, j = 1, 2, 3 \text{ and } i \neq j$$
(B.15)

The parametric equations of line *i* in the *x*-*y*-*z* coordinate system are given by:

$$x_{i} = x_{oi} + m_{i} p_{xi}$$

$$y_{i} = y_{oi} + m_{i} p_{yi}$$

$$z_{i} = z_{oi} + m_{i} p_{zi}$$

(B.16)

where (x_{oi}, y_{oi}, z_{oi}) is a reference point on line *i* and $\hat{p} = \langle p_{xi}, p_{yi}, p_{zi} \rangle$ is a unit vector pointing towards the chip and parallel to line *i*. Both the reference point and the unit vector are known by design because they are determined by the specifications of the off-axis Fresnel lenses. Equations (B.14) and (B.16) are used to write the sides of triangle $P_1 - P_2 - P_3$ in the *x-y-z* coordinate system:

$$\overline{P_i P_j}^2 = [(x_{oi} + m_i p_{xi}) - (x_{oj} + m_j p_{xj})]^2 + [(y_{oi} + m_i p_{yi}) - (y_{oj} + m_j p_{yj})]^2 + [(z_{oi} + m_i p_{zi}) - (z_{oj} + m_j p_{zj})]^2 (B.17)$$

Equating (B.15) and (B.17) leads to a system of three non-linear equations with three unknowns: m_1 , m_2 , m_3 . This system of equation can be solved iteratively using the Newton-Raphson method. The solutions are $m_1 = m_1^*$, $m_2 = m_2^*$, $m_3 = m_3^*$. The positional coordinates of points P_1 , P_2 , P_3 in the x-y-z coordinate system are thus:

$$P_{i} = (x_{oi} + m_{i}^{*} p_{xi}, y_{oi} + m_{i}^{*} p_{yi}, z_{oi} + m_{i}^{*} p_{zi}) \ i = 1, 2, 3.$$
(B.18)

B.4 Solving for the chip misalignment

The last step consists of using the absolute coordinates of points P_1 , P_2 , P_3 (equation (B.18)) to solve for the six-DOF chip misalignment. To do this, one must solve for the center point of the chip in the absolute x-y-z coordinate system. Referring to figure B.3, the center point of the chip (the origin of the x'-y' coordinate system) is labelled C.



Figure B.3. Vector construction to solve for the chip center C and the chip normal \hat{n} . Vector $\overrightarrow{P_1C}$ can be written as:

$$\overrightarrow{P_1C} = k_{12}\hat{v}_{12} + k_{23}\hat{v}_{23}$$
(B.19)

where \hat{v}_{12} and \hat{v}_{23} are defined as:

$$\hat{v}_{12} = \frac{\overrightarrow{P_1P_2}}{\|P_1P_2\|} \text{ and } \hat{v}_{23} = \frac{\overrightarrow{P_2P_3}}{\|P_2P_3\|}$$
 (B.20)

Using equation (B.19), one can write:

$$\overrightarrow{P_1C} \cdot \hat{v}_{12} = k_{12} + k_{23}(\hat{v}_{23} \cdot \hat{v}_{12})$$
(B.21)

$$\overrightarrow{P_1C} \cdot \hat{v}_{23} = k_{12}(\hat{v}_{12} \cdot \hat{v}_{23}) + k_{23}$$
(B.22)

Appendix B: Mathematical derivations for intra-module alignment technique #1

B-6

Solving equations (B.21) and (B.22) for k_{12} and k_{23} leads to:

$$k_{12} = \frac{(\overrightarrow{P_1C} \cdot \hat{v}_{23})(\hat{v}_{12} \cdot \hat{v}_{23}) - (\overrightarrow{P_1C} \cdot \hat{v}_{12})}{(\hat{v}_{12} \cdot \hat{v}_{23})^2 - 1}$$
(B.23)

$$k_{23} = \frac{(\overrightarrow{P_1C} \cdot \hat{v}_{12})(\hat{v}_{12} \cdot \hat{v}_{23}) - (\overrightarrow{P_1C} \cdot \hat{v}_{23})}{(\hat{v}_{12} \cdot \hat{v}_{23})^2 - 1}$$
(B.24)

Equations (B.23) and (B.24) are first evaluated by expressing the terms on the righthand side in the relative x'-y' coordinate system: C = (0, 0) and with P_1 , P_2 , and P_3 defined in equation (B.13). The solutions are $k_{12} = k_{12}^*$, $k_{23} = k_{23}^*$. These scalar quantities are then substituted back into equation (B.19) along with P_1 , \hat{v}_{12} and \hat{v}_{23} expressed in the absolute x-y-z coordinate system (equation (B.18)). This allows for the absolute coordinate of the chip center $C = (x_c, y_c, z_c)$ to be solved for.

Note that the absolute coordinates of point C are unaffected by rotational and tilt misalignments; they correspond directly to the lateral (Δx and Δy) and longitudinal (Δz) misalignment of the chip with respect to the mini-lens array:

$$\Delta x = x_c \tag{B.25}$$

$$\Delta y = y_c \tag{B.26}$$

$$\Delta z = z_c \tag{B.27}$$

To solve for the remaining three angular DOFs ($\Delta \theta_x$, $\Delta \theta_y$, $\Delta \theta_z$), the unit vector normal to the chip, \hat{n} , is defined; it can be evaluated with the values of P_1 , P_2 , and P_3 defined in the absolute x-y-z coordinate system of equation (B.18):

$$\hat{n} = \frac{(\overrightarrow{P_1P_2}) \times (\overrightarrow{P_3P_2})}{\left\| (\overrightarrow{P_1P_2}) \times (\overrightarrow{P_3P_2}) \right\|} = \langle n_x, n_y, n_z \rangle$$
(B.28)

Under perfect alignment condition, the unit normal vector is $\hat{n}_o = \langle 0, 0, 1 \rangle$. The vector normal \hat{n} can be written as follows:

$$\hat{n} = \vec{\theta}_x \cdot \vec{\theta}_y \cdot \vec{\theta}_z \cdot \hat{n}_o \tag{B.29}$$

where $\vec{\theta_x}, \vec{\theta_y}, \vec{\theta_z}$ are the following three-dimensional rotation matrices:

$$\vec{\theta_x} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\Delta \theta_x) & -\sin(\Delta \theta_x) \\ 0 & \sin(\Delta \theta_x) & \cos(\Delta \theta_x) \end{bmatrix}$$
(B.30)

$$\vec{\theta_y} = \begin{bmatrix} \cos(\Delta \theta_y) & 0 & \sin(\Delta \theta_y) \\ 0 & 1 & 0 \\ -\sin(\Delta \theta_y) & 0 & \cos(\Delta \theta_y) \end{bmatrix}$$
(B.31)

$$\vec{\theta}_{z} = \begin{bmatrix} \cos(\Delta \theta_{z}) & -\sin(\Delta \theta_{z}) & 0\\ \sin(\Delta \theta_{z}) & \cos(\Delta \theta_{z}) & 0\\ 0 & 0 & 1 \end{bmatrix}$$
(B.32)

Since matrix multiplication is not commutative, the sequence of the matrix multiplication in equation (B.29) is critical. The chosen sequence is arbitrary; it implies that the chip tilt and rotational misalignment is calculated as if the chip had been misaligned by the following sequence of operations: (1) rotational misalignment of $\Delta \theta_z$, (2) y-tilt misalignment of $\Delta \theta_y$ and (3) x-tilt misalignment of $\Delta \theta_x$. These angular operations are defined with respect to the relative x-y-z coordinate system. Substituting matrices (B.30), (B.31) and (B.32) into equation (B.29) leads to:

$$\hat{n} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \begin{bmatrix} \sin(\Delta\theta_y) \\ -\sin(\Delta\theta_x)\cos(\Delta\theta_y) \\ \cos(\Delta\theta_y)\cos(\Delta\theta_y) \end{bmatrix}$$
(B.33)

which allows for the tilt misalignment of the chip ($\Delta \theta_x$ and $\Delta \theta_y$) to be calculated:

$$\Delta \theta_x = -\operatorname{atan}\left(\frac{n_y}{n_z}\right) \tag{B.34}$$

$$\Delta \theta_y = \operatorname{asin}(n_x) \tag{B.35}$$

The rotational misalignment of the chip is calculated as follows. Under conditions of perfect alignment, the vector joining the center of the chip to point P_1 , in absolute coordinates, is given by $\langle x'_{d1} + u_{xi}, y'_{d1} + v_{yi}, 0 \rangle$ (see equation (B.13)). Following a series of misalignment operations $(\Delta \theta_z \rightarrow \Delta \theta_y \rightarrow \Delta \theta_x \rightarrow \Delta z \rightarrow \Delta y \rightarrow \Delta x)$, this vector points to the location of P_1 calculated by equation (B.18). This can be expressed mathematical as:

$$\vec{\Theta}_{x} \cdot \vec{\Theta}_{y} \cdot \vec{\Theta}_{z} \cdot \begin{bmatrix} x'_{d1} + u_{xi} \\ y'_{d1} + v_{yi} \\ 0 \end{bmatrix} + \begin{bmatrix} x_{c} \\ y_{c} \\ z_{c} \end{bmatrix} = \begin{bmatrix} x_{o1} + m_{1}^{*} p_{x1} \\ y_{o1} + m_{1}^{*} p_{y1} \\ z_{o1} + m_{1}^{*} p_{z1} \end{bmatrix}$$
(B.36)

At this point, the only unknown in equation (B.36) is the angle $\Delta \theta_z$, allowing for rotational misalignment to be solved for directly.

B.5 Summary

Starting with the photocurrent measurements from three quadrant detectors (a total of 12 photocurrents), the six-DOF misalignment of the chip (Δx , Δy , Δz , $\Delta \theta_x$, $\Delta \theta_y$, $\Delta \theta_z$) has been derived and can be found in equations (B.25), (B.26), (B.27), (B.34), (B.35) and (B.36) respectively.