# Demonstrating Effective All-Optical Processing in Ultrafast Data Networks using Semiconductor Optical Amplifiers

by

Jade P. Wang

Submitted to the Department of Electrical Engineering and Computer Science

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2008

Author . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Department of Electrical Engineering and Computer Science
May 23, 2008

Certified by. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Erich P. Ippen
Elihu Thomson Professor of Electrical Engineering
Thesis Supervisor

Certified by. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Scott A. Hamilton
MIT Lincoln Laboratory Assistant Group Leader
Thesis Supervisor

Accepted by . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
Terry P. Orlando
Chairman, Department Committee on Graduate Students

# Demonstrating Effective All-Optical Processing in Ultrafast Data Networks using Semiconductor Optical Amplifiers

by

Jade P. Wang

## Abstract

The demand for bandwidth in worldwide data networks continues to increase due to growing Internet use and high-bandwidth applications such as video. All-optical signal processing is one promising technique for providing the necessary capacity and offers payload transparency, power consumption which scales efficiently with increasing bit rates, reduced processing latency, and ultrafast performance. In this thesis, we focus on using semiconductor optical amplifier-based logic gates to address both routing and regeneration needs in ultrafast data networks.

To address routing needs, we demonstrate a scalable, multi-packet all-optical header processing unit operating at a line rate of 40 Gb/s. For this experiment, we used the ultrafast nonlinear interferometer (UNI) gate, a discrete optical logic gate which has been demonstrated at speeds of 100 Gb/s for bit-wise switching. However, for all-optical switching to become a reality, integration is necessary to significantly reduce the cost of manufacturing, installation, and operation. One promising integrated all-optical logic gate is the semiconductor optical amplifier Mach-Zehnder interferometer (SOA-MZI). This gate has previously been demonstrated capable of up to 80 Gb/s bit-wise switching operation. To enable simple installation and operation of this gate, we developed a performance optimization method which can quickly and accurately pinpoint the optimal operating point of the switch. This eliminates the need for a time-intensive search over a large parameter space and significantly simplifies the operation of the switch. With this method, we demonstrate the ability of a single SOA-MZI logic gate to regenerate ultrafast pulses over 100 passes and 10,000 km in a regenerative loop.

Ultimately, all-optical logic gates must be integrated on a single low-cost platform and demonstrated in cascaded, multi-gate operation for increased functionality. This requires low-loss monolithic integration. Our approach to this involves an asymmetric twin waveguide (ATG) design. This design also has the potential for high-yields as a result of a high tolerance for fabrication errors. We present our characterization results of ATG waveguides and proposals for future improvements.

Thesis Supervisor: Erich P. Ippen
Title: Elihu Thomson Professor of Electrical Engineering

Thesis Supervisor: Scott A. Hamilton
Title: MIT Lincoln Laboratory Assistant Group Leader

# Acknowledgments

There are many people without whom this thesis would never have been completed. First and foremost, I would like to thank Professor Erich Ippen and Dr. Scott Hamilton, my thesis advisors. Professor Ippen's keen intelligence and insightful comments have been critical to my development as an engineer and scientist. Without his patient advice and encouragement, I would never have managed to finish this degree. Dr. Scott Hamilton's contribution to my intellectual and professional development cannot be overstated. Eight years ago, I applied for an internship in his group on the basis of his presentation on all-optical logic. Since then, his tireless efforts on our research group's behalf have given me the freedom to pursue my research interests. His guidance in all areas, from the details of my experiments to general career advice, has been indispensable. His unfailing sense of optimism, kindness, and cheerfulness never fails to inspire me. I would also like to thank Professor Rajeev Ram for serving on my thesis committee and advising me through my qualifying exams. His sharp insights have always left me with a new understanding of the field and his breadth of knowledge is truly impressive.

I would also like to thank Professor Jin Au Kong for his enthusiastic help over the years as my graduate counselor and for my first introduction to electromagnetism and Maxwell's equations. I will miss him.

A short paragraph is insufficient to describe the depth of appreciation I have for my colleagues at MIT Lincoln Laboratory. They have taught me much of what I know of optical communications as well as science, research, and even philosophy. Dr. Shelby Savage was my fellow graduate student and labmate for most of my time here. His methodical approach and willingness to answer the simplest of questions with thoroughness and rigor has taught me as much any course I have taken. His sense of humor has always enlivened the darkest hours in lab. Without his company and support, I would have never succeeded in this endeavor. Dr. Bryan Robinson has been invaluable in his knowledge of semiconductor physics and his vast skill with all types of laboratory equipment. He has the rare ability to take any problem and

distill it down to its critical components. My discussions with him have always been fruitful and have often redefined the problem I am trying to solve. I would also like to thank Claudia Fennelly for her expertise and advice in both splicing, soldering, and wiring, and Chad Demers, for his help and guidance in my first attempts at splicing fiber. Dr. Todd Ulmer, Dr. Matthew Grein, Dr. Dave Caplan, Dr. Mark Stevens, and Dr. Neal Spellmeyer have all provided invaluable advice, loaned me instruments, and offered their expertise and help for a wide variety of questions and problems in the laboratory. I would also like to thank Dr. Juliet Gopinath, for her willingness to offer advice and support from the perspective of someone who has gone through all this before. Her presence has been an inspiration and her advice is always solid. My group leaders, Dr. William Keicher and Dr. Don Boroson, both deserve thanks for their continued support throughout my PhD thesis. Finally, much thanks is due to Cheryl Abruzzese, for her patience and support in dealing with the administrative aspects of my research here.

Some of the experimental work presented in this thesis has been the product of collaborations with other groups and companies. I would first like to thank the folks at Alphion Corporation for their generosity in providing us with their SOA-MZI devices. Dr. Boris Stefanov, Dr. Leo Spiekman, Dr. Ruomei Mu, and Dr. Hongsheng Wang also provided their expertise and advice in the operation and design of SOA-based devices in order to make our collaboration a success. I would also like to thank the people in the Integrated Photonic Materials and Devices group at MIT; in particular, Professor Leslie Kolodziejski, Dr. Gale Petrich, Dr. Aleksandra Markina, Dr. Ryan Williams, and PhD candidate Ta-Ming Shih. They are the giants upon whom our work with integrated SOA-MZI chips stand. Finally, I would also like to thank Jason Plant, Dr. Paul Juodawlkis, and Reuel Swint from Group 83 at MIT Lincoln Laboratory. Jason has been particularly patient at enlightening me about chip fabrication techniques. His careful work and good advice have been instrumental in the progress made on this research project.

Several students from MIT deserve special thanks. Ali Motamedi spent several weeks at Lincoln Laboratory helping me with my experiments. He was a pleasure to

work with and his drive was enviable. I also consider it a privilege to have worked with Tauhid Zaman. His forthrightness and intelligence are both valuable traits which I can only seek to emulate. I would also like to thank Reja Amatya for her help with thermal measurements and simulations. Last, but definitely not least, I would like to thank the all students in Professor Ippen's group on campus, who have often gone out of their way to make me feel extraordinarily welcome in their offices and on the ski slopes. Many of my most enjoyable hours were spent in your company.

My housemates past and present, deserve much thanks for helping me make my house into a home. They have provided me with much-needed company, commiseration, celebration, and support throughout. I also wish to particularly thank Julia Liberman, for helping me correct my hopelessly tangled tenses in this thesis. I am deeply grateful to have fallen in with a caring community of friends who continue to support me and enrich my life. Most of all, however, I wish to thank Jay Muchnij, my partner in all things. He has stood by me and supported me in all my endeavors, offering me advice, perspective, proofreading, and love. I only hope that I can be as giving and faithful a partner for him as he is for me.

Finally, I am deeply grateful to my family. My father, mother, and brother have provided me with a nurturing and loving home where I grew up learning to take responsibility for my actions, get along with others, go after my dreams, and share my toys. My father has listened and given advice on my research when it became frustrating. My mother has provided good advice on time management and dealing with burnout as well as the best food you can find anywhere. My brother has become a successful and savvy businessman with well-reasoned opinions and advice for someone who has yet to step foot out of her ivory tower. Without them, I would not have gotten to where I am today.

Thank you.

# Contents

# List of Figures

17

19

# List of Tables

# Chapter 1

# Introduction

Today, data networks transmit thousands of petabytes of data each day in the form of email, video, pictures voice, web pages, data files, and gaming content. This is accomplished through a combination of networks of various sizes, from global networks such as those owned by Global Crossing or Tyco, down to tiny networks such as those run by a single university. In most networks, high-rate data is transmitted over optical fiber. Figure 1-1 shows a diagram of a simple network. Electronic data packets are encoded in optical pulses through an electronic-to-optical conversion and aggregated in tens of separate wavelength channels in a single fiber. To direct the data packets from source to the correct destination, routers are used to process the address of each packet and forward the packets to the correct output port. At each router, the many wavelength channels are demultiplexed and converted back into the electronic domain. Each data packet is stored in memory until it can be processed. Once a data packet is processed, it is switched to the correct output port and re-converted to the optical domain. This occurs multiple times before the data arrives at its destination. As the data packets propagate down long lengths of fiber, loss and distortions occur. Erbium-doped fiber amplifiers (EDFAs) capable of providing gain over a wide wavelength range are used to counteract loss in the fibers. To counteract linear and nonlinear degradations, electronic regenerators are used. The optical pulses are demultiplexed, converted to electronic signals and processed electronically to regenerate the data. The regenerated data is then converted back to the optical domain.
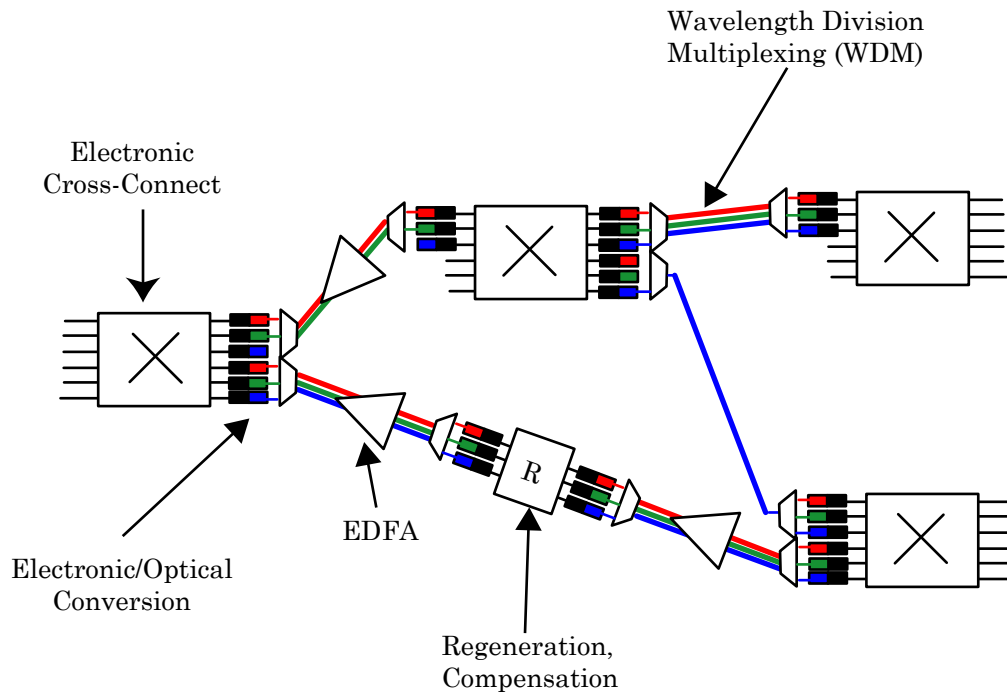
Figure 1-1: Diagram of networks today. EDFA stands for erbium-doped fiber amplifier.

There is a steadily increasing demand for bandwidth driving the need for higher network capacities at lower costs [1, 2, 3, 4]. This demand arises largely from increasing video traffic as a result of video-on-demand and user-generated content on sites such as YouTube. Future video needs are likely to result in even higher bandwidth demands from the introduction of the high-definition video standards. Increasing use of storage networks, grid computing, and scientific data exchange will also contribute to increased demand. Furthermore, this increased bandwidth demand does not correspond to a willingness to pay higher costs for service. Thus, future networks must be capable of delivering higher bandwidths at lower or equivalent costs.

To accomplish this, both router capacity and fiber transmission capacity must increase while costs must be kept low. Capacity can be increased by increasing channel bit rates, number of channels per fiber, and by laying down more fiber. The high cost of duplicating regenerators, amplifiers, and other equipment encourages research into increasing channel bit rates as well as increasing the number of channels in each fiber.

As a result, many networks in the United States have moved from 2.5 Gb/s to 10 Gb/s and 40 Gb/s bit rates while also moving from to dense WDM channel spacings of 50 GHz or 25 GHz, allowing over 100 channels on a single fiber. This continuing trend of increasing channel bit rate and channels per fiber is beginning to significantly impact the power consumption, size, and weight of electronic routers and regenerators [5, 6]. A conversion between optical and electronic domains (O/E/O) is required at every input and output of each router and regenerator, adding to the power consumption and complexity of each device. Furthermore, electronic switching speeds are currently only capable of 10-40 Gb/s operation. As channel bit rates continue to increase, this forces a parallel architecture and further contributes to the complexity and cost of electronic routers and regenerators.

We believe that optical signal processing techniques may be one approach for alleviating these issues. First of all, all-optical switches are extremely fast when compared with electronic switches. Bit-wise switching has been demonstrated in optical switches at 100 Gb/s and higher [7], while wavelength conversion has been demonstrated at 640 Gb/s [8]. This has the potential to reduce power consumption and complexity by eliminating the need for parallel architectures. Secondly, the bit-wise optical switching energy for these optical switches is extremely low, on the order of tens of femtojoules. The majority of the power consumed in semiconductor-based all-optical switches is in the DC bias current, which is approximately independent of bit rate. Thus, as bit rates increase, the energy per bit required for switching actually decreases. This implies that optical switches will scale well with continually increasing bit rates. Third, by supplementing electronic processing techniques with optical techniques, we can potentially eliminate many costly O/E/O conversions. Finally, optical techniques offer increased transparency. For instance, without an O/E/O converter, we can transparently pass the payload of the packet through the router, handling any bit rate or modulation format. In other situations, optical add-drop multiplexers (OADMs) can add or drop entire wavelengths without further processing. Again, the exact encoding of data on the particular wavelengths is irrelevant. This allows us to design a flexible network which can adapt well to improvements in modulation format

and channel bit rates.

However, all-optical techniques currently remain a fairly immature technology compared with electronic processing techniques. Before all-optical techniques can be applicable to real-world networks, two major challenges must be met. First of all, to make a compelling case of adopting optical signal processing techniques into current data networks, all-optical processing must offer additional or equivalent functionality to the electronic devices it will be replacing. Though an impressive amount of all-optical functionality has been demonstrated thus far, more work still needs to be done in this area if all-optical signal processing techniques are to match the functionality offered by more mature electronic processing techniques. Secondly, the practical implementation of all-optical switching must be addressed. Integration and mass-production of all-optical logic gates will be required to reduce the cost and footprint of all-optical technology. Optical logic gates must also be designed to be simple to operate and optimize, in order to reduce labor costs associated with installation and maintenance. Integration and simple operation will also allow practical multi-gate logic demonstrations, thus further improving the functionality of all-optical logic. In this thesis, we will demonstrate the effectiveness of using all-optical techniques for supplementing electronic processing in data networks and address this need for practical implementation and integration of the switches in real-world operation.

In Chapter 2, we investigate how all-optical signal processing can be used to supplement routing and switching. We demonstrate a simple all-optical header processing unit which requires only two all-optical logic gates to simultaneously forward two 40-Gb/s packet streams through a 2x2 spatial electro-optic switch. This demonstrates a scalable, ultrafast all-optical solution which provides payload transparency and shorter packet-processing latencies.

Next, in Chapter 3, we examine the need for integration of all-optical switches. We focus on SOA-based (semiconductor optical amplifier) switches and describe the SOA-MZI (semiconductor optical amplifier Mach-Zehnder interferometer) logic gate. We demonstrate a practical and straightforward optimization method for quickly arriving at optical switch performance.

In Chapter 4, we turn to the function of regeneration in data networks. By using the optimization method described in Chapter 3, we are able to find an optimal operating point and demonstrate error-free regeneration using a single SOA-MZI gate over 10,000 km and 100 passes in a recirculating loop. This also successfully demonstrates the feasibility of cascaded operation of the SOA-MZI logic gate.

Finally, in Chapter 5, we describe our progress in developing a monolithic integration platform for designing multi-gate optical logic. In particular, we target low-power, high yield devices using asymmetric twin waveguide techniques to integrate our SOA active devices with passive waveguides and couplers.

We conclude with Chapter 6, where we summarize our contributions and discuss what is required for future developments in all-optical logic.

# Bibliography

[1] R. J. Shapiro, "The internet's capacity to handle fast-rising demand for bandwidth." `http://www.usiia.org/pubs/Demand.pdf`, 2007.

[2] Cisco, "The exabyte era." white paper, Jan. 2008.

[3] Cisco, "Global ip traffic forecast and methodology, 2006-2011." white paper, Jan. 2008.

[4] A. M. Odlyzko, "Internet growth trends and moore's law." `http://www.dtc.umn.edu/mints/igrowth.html`.

[5] G. Epps, "System power challenges," in *Cisco Routing Research Symposium*, Aug. 2006.

[6] C. Minkenberg, R. P. Luijten, F. Abel, W. Denzel, and M. Gusat, "Current issues in packet switch design," in *ACM SIGCOMM Computer Communication Review*, vol. 33, pp. 119–124, 2003.

[7] K. L. Hall and K. A. Rauschenbach, "100-Gbit/s bitwise logic," *Optics Letters*, vol. 23, 1998.

[8] M. Galili, H. C. H. Mulvad, L. K. Oxenlowe, H. Ji, A. T. Clausen, and P. Jeppensen, "640 Gbit/s wavelength conversion," in *Optical Fiber Communication Conference (OFC '08)*, 2008.

# Chapter 2

# All-Optical Techniques for Switching and Routing

In the introduction, we suggested that all-optical techniques may be able to supplement the electronic methods currently used in data networks and enable transparent, ultrafast (40 Gb/s and higher) networks with lower size, weight and power requirements. In this chapter, we focus on the potential of optical processing techniques for supplementing routing and switching in data networks. We first provide a brief overview of the basics of networks, switching, and contention resolution in data networks today and illustrate where optical signal processing techniques can supplement or even replace current technologies. As an example, we investigate the capabilities of optical switching for header processing. We successfully demonstrate a scalable all-optical header processing technique which correctly forwards two simultaneous 40-Gb/s packet streams through a single 2x2 spatial switch.

## 2.1   Network Design

To transmit information from source to destination in a network, some agreed-upon protocol must exist as to how to locate the destination and how to determine the path by with the information is transmitted. In a telephone network, the address is the telephone number: a single 10-digit number is assigned to every possible location.

When the number is dialed, this tells the switches in the network to connect the calling phone with the dialed phone. In the same way, the postal network observes the address on each letter and thus knows the destination of the letter. In computer networks, this address is the IP address, which is a 32-bit address (in the IPv4 standard). This IP address does not uniquely identify a single device, just as a phone number does not identify a single phone.

Networks can be roughly divided into two types: circuit-switched and packet-switched. Figure 2-1 illustrates the difference between the two. The type of traffic
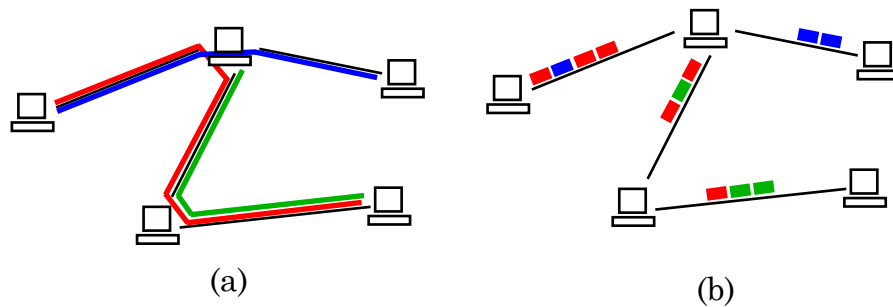


Figure 2-1: (a) A circuit-switched network. (b) A packet-switched network.

a network serves is used to determine the correct design choices for how routing and switching is implemented in the network. The telephone network is an example of a circuit-switched network. When a connection is made, a reserved circuit is created between the source and destination until the call is ended. These circuits change very slowly, on the order of milliseconds to seconds. This circuit-switched connection-based network is well suited to telephone traffic for several reasons. First of all, telephone callers wish to communicate with a minimum of delay between responses. Thus, telephone traffic needs to be low-latency, which means the time it takes for information to travel between point A to point B needs to be short. Furthermore, this latency needs to remain relatively consistent. Secondly, a dropped or interrupted call is much worse than a call which cannot be completed. Thus, circuits are reserved, and in cases of high traffic, new circuits are prevented from accessing the network (busy signal). Finally, information is almost continuously transmitted during a conversation

— either one person is speaking or the other person is speaking. Thus, each circuit uses the bandwidth reserved for it efficiently. A circuit guarantees a certain minimum bandwidth and delay once it is set up.

The postal network and current computer networks are examples of packet-switched networks. Information is encapsulated into packets (or envelopes) and the packets are individually delivered. This is also known as a connection-less network, as no continuous connection exists. Each packet consists of a header, which contains the address and control bits, and the payload, which consists of the data to be transmitted. Each node makes the decision of how to forward each packet to the next node, based only on the packet address and a local routing table. Thus, decisions are decentralized and made on-the-fly. This is well-suited to the bursty nature of Internet traffic such as emails, web pages traffic, and file transfers [1]. In these situations, reserving an entire circuit for a single traffic source/destination pair would be a highly inefficient use of resources. Furthermore, it is not critical to have a low latency. It does not matter exactly when an email arrives at its destination. An inconsistent or variable latency is also not an issue. Finally, a dropped packet is not as costly as a dropped phone call, since the packet can be sent again, causing only extra delay. Packets can even arrive out of order, taking different paths as the network traffic patterns change, without necessarily impinging upon the perceived performance. However, the recent popularity of voice and video traffic (such as YouTube) on data networks [2, 3, 4] means that a more connection-based architecture may make more sense in the future.

Packet-switched networks can further be divided into networks which use variable-sized packets (IP networks) or fixed-size packets. Networks can also be asynchronous (packets can arrive at any time) or synchronized (packets are timed to arrive simultaneously at each node.) An asynchronous network with variable-sized packets allows each node to operate in a highly-decentralized manner. However, each switching node is less efficient, since a long packet can block many shorter packets from being forwarded until it is transmitted. At the other end of the spectrum is a synchronized network with fixed-length packets. Synchronization requires a network clock signal and some method for clock recovery at each node. This results in a more complex

architecture. However, switching can be much more efficient, since at each time slot, packets will enter and exit the switch. Today's data networks rely on variable-sized, asynchronous packets.

We distinguish between "routing" and "switching". Routing is the determination of the path by which a packet is sent from the source to its final destination. Switching occurs in each individual node, as inputs are connected to the proper outputs for the packets arriving and departing the node. Routes for different packets are currently determined by a routing table which is updated whenever nodes appear or disappear from the network. Routing requires a large memory for the routing table and relies on longest-match algorithms, which make search algorithms more time-intensive and costly. These routing tables could theoretically change quickly, but in actuality these routes can remain stable for a day or more at a time, changing only in response to new nodes, node failures, and line cuts [5, 6]. Frequent updates of the routing table can result in an unacceptable amount of overhead in the network or network instability. Switching, on the other hand, requires fast updates (often changing switch configuration per packet arrival) and only requires memory sufficient for the port count of the individual node. Design of switch architecture requires balancing the need for reducing cost and complexity with the need for speed and effectiveness. In this thesis, we focus on the techniques of switching and forwarding, as these are areas where optical techniques can make a significant difference.

Data networks today consist of entirely electronic routers connected by wavelength-division-multiplexed optical fiber links (Figure 2-2). At each router, optical channels are demultiplexed and converted into electronic signals by a bank of photodetectors. Often, each channel is further demultiplexed to match the processing rates of router electronics. The electronic packets are stored in memory while the header is processed and the electronic cross-connect is configured to correctly forward the packets to the output ports. Packets which wish to exit at the same output port simultaneously are delayed in buffers. At the output, these packets are re-converted to the optical domain through modulators and laser diodes. Each channel is then re-assembled and multiplexed onto the output fibers. As routers increase in capacity in response
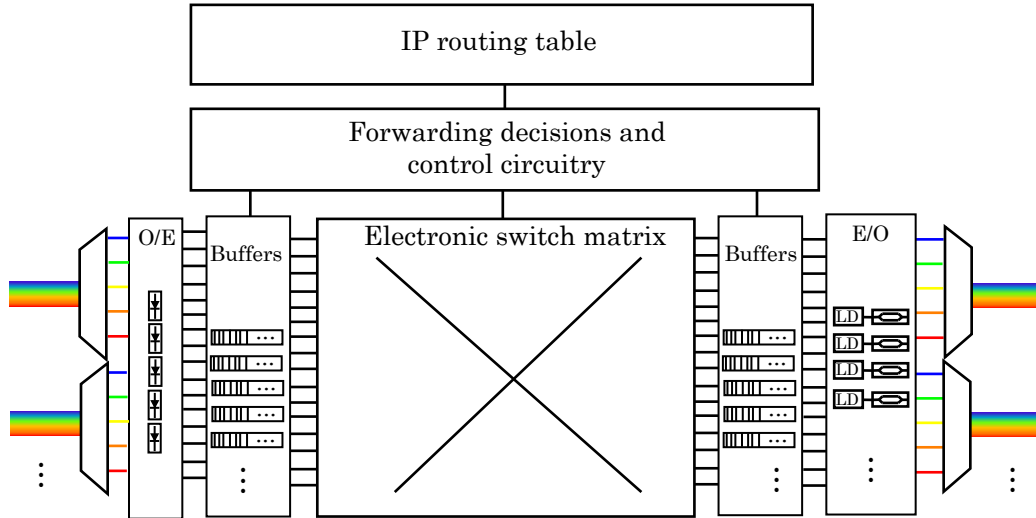
Figure 2-2: Schematic of an electronic packet router.

to bandwidth demand, methods to reduce power consumption and complexity will be needed. Optical techniques can remove the need for optical-electronic-optical (O/E/O) conversions and are capable of much higher speeds than electronic switching. Where electronic switches may require a parallel architecture to handle the switching capacity, a single optical switch may be sufficient. Furthermore, improved modulation formats and higher bit rates are constantly being researched. The ability to quickly upgrade and incorporate new research into the existing network will be extremely useful for future, more demanding networks. Optical techniques can provide this ability by creating payload-transparent networks. Currently, the process of O/E/O conversion requires a set modulation format for the payload as well as the header of packets. By moving to optical techniques, we can eliminate the O/E/O conversion and pass the payload transparently through the router. This means that as modulation formats and bit rates change, the optical network can easily keep up by simply changing the transmitter and receiver nodes, rather than all possible nodes in the network.

## 2.2  Switching

The function of a switch is to forward incoming packets to the correct output port. Switches can range in size from 2-port and 4-port switches in metro networks to hundreds of ports in core routers. To correctly forward packets, a switch needs to (1) process the packet header or address, (2) configure a switch matrix to produce the correct path or paths for the incoming packets, and (3) resolve any contentions between packets. The performance of a switch can be measured by several metrics: scalability, throughput, and latency. The scalability of a switch determines how many extra components are required to build larger and larger switches of the same type. A very complex switch is not very scalable, as past a certain port count the cost of the components and the complexity of switch operation becomes prohibitive. The throughput of a switch describes how much information the switch can transmit per unit time from input to output. The ideal throughput of a switch can be claimed as the capacity of each input port multiplied by the number of input ports. However, the actual throughput is generally much lower, limited by issues such as contention, blocking, and even traffic statistics. For instance, as shown in Figure 2-3, if all input packets wish to exit at the same output port (a), the throughput of the switch will be dramatically lower than if they wish to exit on different output ports (b). Finally, the latency of the switch is the amount of time it takes a packet to move from the input of the switch to the output. This includes the header processing time, the time it takes for the switch matrix to be configured, the travel time of the packet through the matrix, and any additional delays the packet experiences due to contention. For some applications the latency variation also matters. For instance, with video, if all packets experience a 5-second latency, the application can add an additional 5-second buffer and the user will not observe any degradation. However, if each packet alternates between experiencing a 1-second and a 10-second latency, it is very inefficient use a 10-second buffer but a shorter buffer will result in an erratic video stream. Furthermore, variations in packet latency can result in packets arriving significantly out of order, which the current network congestion
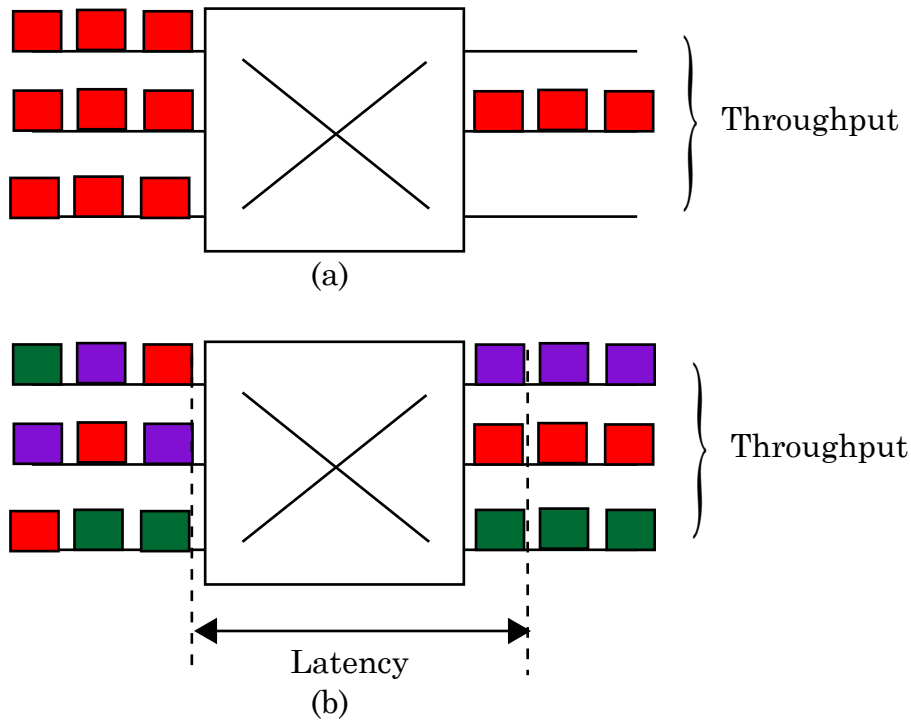
Figure 2-3: Illustration of the concepts of switch throughput and latency. In (a), all input packets wish to exit on the same output port. Thus, 2/3 of the packets must be buffered or dropped, significantly decreasing switch throughput. In (b), the input packets are all addressed to different output ports, resulting in a much higher throughput.

control protocol (TCP) cannot handle effectively.

## 2.2.1   Header Processing

The header of a packet usually consists of the IP address, packet length field, and various other checksums and control bits which are used to verify packet integrity, control router behavior and prevent indefinite looping. Typically, upon entering the switch, the packet is buffered while the IP routing table is consulted to determine the destination port for that packet. This decision is then used to configure the switch matrix.

The effectiveness of the packet header processing unit affects both the latency of the switch and the throughput. Typically, the IP address lookup involves a time-

intensive search through a large routing table. This search time will only increase as the network expands; for example, the IPv6 protocol, which expands IP addresses to 128-bit addresses, is likely to become standard within the next few years. As this header processing delay increases, the packet latency increases and switch throughput tends to decrease.

One potential method for avoiding the time-intensive search is to move such processing to the edges of the network. In the core of the network, traffic tends to be aggregated and large bandwidths are essential. At the edge of a network, routers are under fewer constraints and thus can provide complex packet processing without significantly slowing the entire network. There have been several suggestions for how to accomplish this, often proposed in the context of all-optical switching. Multi-protocol label switching (MPLS) (also often GMPLS, generalized multi-protocol label switching, and MP$\lambda$S when referring to optical networks) is a method by which short labels are applied to each packet at label edge routers. These labels are then used to determine the path a packet takes through the network. By switching from an IP table lookup to a label lookup, we are able to switch from a longest-path search to an exact match search, such that a much simpler search algorithm can be implemented. Furthermore, this technique also allows a greater variety of routers to be incorporated in a single network. It is also very attractive for all-optical networks, as one major challenge for optical techniques is in the lack of a simple optical memory, making IP routing difficult to accomplish with purely optical methods. With MPLS, however, an optical switch only needs to match a short label [7, 8]. Currently, MPLS is a fairly common service provided in many network routers.

Burst switching is another option, in which packets would be aggregated into long bursts formed of packets with similar destinations at the network edges. It is a hybrid of packet-switched and circuit-switched systems. The main advantage of this method lies in reducing buffering requirements in the core of the network and potentially reducing latencies. Various protocols have been suggested in the context of optical burst switching, the most popular of which is known as JET, or Just-Enough-Time switching. In this method, a control packet is sent out before the burst, reserving

a channel for when the burst arrives at each node and only for the length of the burst. Thus, bandwidth is reserved in the most efficient manner. After a certain delay equal to the sum of the delays required at each node, the burst is sent. Thus, the burst is buffered at the edge nodes and does not require buffering in the core network nodes [9, 10, 11].

Whether or not an IP address lookup is involved, routers need to process some sort of label or address in order to correctly configure the switch matrix. All-optical techniques can reduce the header processing latency significantly when compared to electrical techniques due to faster switching capabilities and no need for memory access. Previous optical header processing demonstrations have focused on address keyword recognition for broadcast-and-select networks [12, 13, 14, 15, 16]. In Section 2.3, we demonstrate all-optical header processing at 40 Gb/s for more general (mesh) network architectures.

### 2.2.2 Switch Matrix

A switch matrix is a rearrangeable combination of connections and simple switches that are used to connect input ports with output ports. The switch architecture determines how much blocking exists, the latency of the packets through the switch, and the scalability of the switch. The implementation of the switch affects these qualities as well as the speed of the switch and the extinction ratio for packets exiting the switch.

We use the example of a Banyan switch architecture (Figure 2-4) to illustrate these concepts. The Banyan switch is organized such that the $j$th column uses only the $j$th address bit to forward packets through the switch. This means that the control logic for a Banyan switch is very straightforward in design. The Banyan switch matrix is a rearrangably non-blocking switch. This means that any input can be connected with any output only with a rearrangement of the switch state. When one input is connected with one output, other connections are blocked. In this case, if input port 2 is connected with output port 5 (the red path), a packet coming in on port 1 is blocked from all of the bottom 4 output ports. In contrast, a wide-sense non-blocking

45
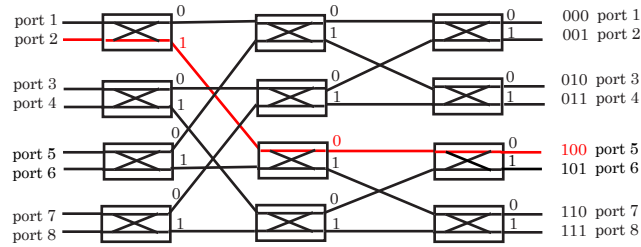
Figure 2-4: An 8x8 Banyan switch. As indicated by the red line, each 2x2 switch in the $j$th column only uses the $j$th address bit to forward the packets correctly to the destination address.

switch is a switch where all inputs can be connected with the correct outputs as long as the connections are all made in an intelligent manner. Finally, a strict-sense or strictly non-blocking switch is a switch where any input can be connected to any output without ever blocking any other possible connections. One example of this is the crossbar interconnect switch, shown in Figure 2-5. Each black dot represents a potential connection. A strict-sense non-blocking switch will reduce latencies and increase throughput in the switch because all packets will be able to immediately pass through the switch. However, strictly non-blocking switches require many more components (in this case, $N^2$), reducing scalability and increasing cost.

The scalability of a switch is determined by how many 2x2 (or 1x2 and 2x1) components an N-port switch requires. For the Banyan switch architecture, one can see that an $N$x$N$ switch will require $\frac{N}{2}\log_2 N$ 2x2 switches. Each column adds another bit to the address space, thus doubling the addressable outputs. A 3-column Banyan switch is an 8x8 switch, as shown in Figure 2-4. Each column will have N ports. Since each 2x2 switch gives us 2 ports, each column will have $\frac{N}{2}$ components. The total number of components is thus $\frac{N}{2}\log N$, giving us a very scalable switch architecture.

Latency describes the time it takes for a packet to exit the switch matrix. Header processing, buffering, and switch blocking can all add to this time. The shortest latency in the Banyan architecture, then, is $T\times\log N$, where $T$ is the time it takes a packet to traverse a single column. Since all packets must travel all the Banyan switch columns once, there is no latency variation inherent in the Banyan switch
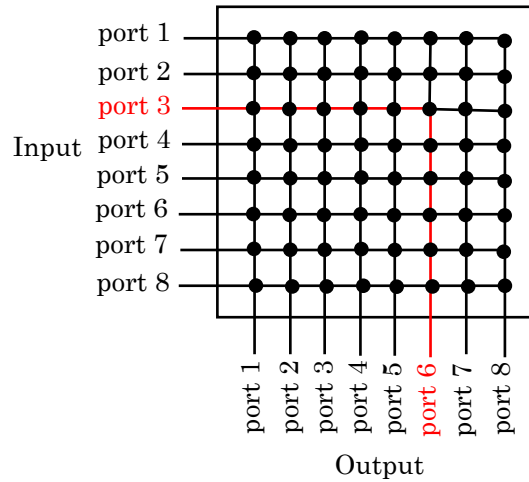
Figure 2-5: An 8x8 crossbar interconnect switch. Each black circle represents a potential connection. Any input can be switched to any output without blocking any other path.

architecture. The effects of blocking, however, will introduce variations in latency between packets.

The speed of a switch determines the guard band requirements of the router. As switches cannot switch instantaneously, packets must be separated by the amount of time needed to reconfigure the switch in order to avoid partial switching and errors. This concept is illustrated in Figure 2-6(a). This guard band should be as short as possible in order to maximize the packet rate through the switch. However,
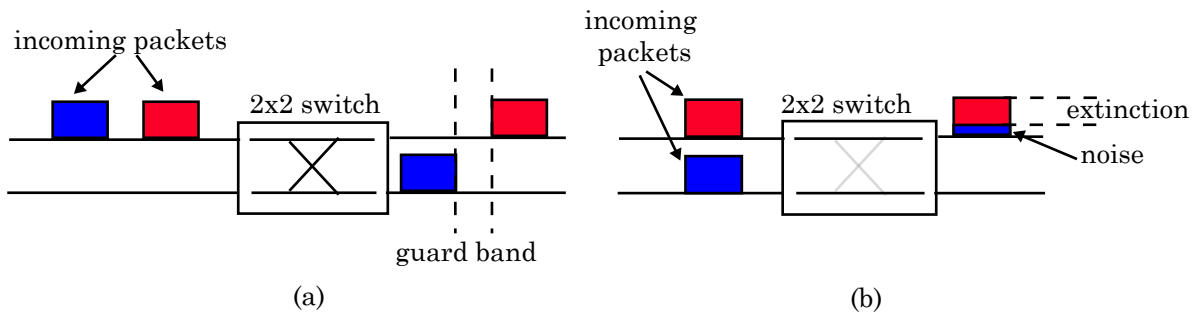


Figure 2-6: (a) Illustration of guard band requirements associated with switch matrices. (b) Illustration of switch extinction ratio.

switch matrix speeds do not need to approach those of header processing speeds or regeneration speeds, as they only need to change state with each new incoming packet, rather than each new incoming bit. The smallest IP packet is 20 bytes (minimum size header with no data) while the largest IP packet is 65,536 bits ($2^{16}$), because the length field in IP headers occupies 16 bits. The average packet length is 300-400 bytes, but the distribution of packets dominated by very short packets (40-44 bytes). Almost no packets are longer than 1500 bytes, due to the popularity of the Ethernet protocol for which 1500 bytes is the maximum length [17, 18]. At 40 Gb/s, a 20-byte packet occupies 4 ns. At 100 Gb/s, a 20-byte packet shortens to 1.6 ns. Thus, a switch matrix should have speeds on the nanosecond time scale.

Another key metric for switch matrices is the extinction ratio. This determines how well the switch matrix preserves the quality of the packets which pass through it. A switch matrix with poor extinction will incur higher costs because more hardware for regeneration and error correction will need to be implemented. Figure 2-6(b) illustrates the extinction ratio of a switch.

Currently, switch matrices are implemented entirely in electronics, resulting in high losses and the need for expensive O/E/O conversions. An alternative is to use electro-optic switches, typically in lithium niobate ($LiNbO_3$), which can provide optical transparency, high extinction, and reasonably fast switching. Furthermore, electronic control means that they can be used with existing electronic processing. Electro-optic switches rely on the electro-optic effect, in which an electric field is used to change the index of refraction of the optical waveguide. This can be used to create phase delays as well as interferometric switches (Figure 2-7). In the Mach-Zehnder interferometer design, DC voltage sets the bias of the interferometer such that the input light either recombines constructively or destructively. Thus, the input A will exit either on output A or output B. The phase relationship between the upper and lower arms of the interferometer is then changed by exactly $\pi$ through the application of the correct voltage ($V_\pi$). This changes the interferometer from constructive interference to destructive interference or vice versa. Thus, the input A is sent to the other output port. These switches usually have speeds of about 1 GHz, though
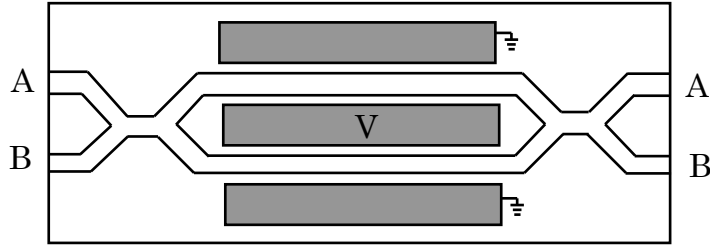
Figure 2-7: Schematic of a 2x2 electro-optic crossbar Mach-Zehnder switch. By applying a voltage across the electro-optic material, we can change index of refraction in the upper and lower waveguides. If we induce an additional $\pi$ phase difference between the upper and lower waveguides, we change the interferometer from constructive to destructive interference.

40-GHz switches are possible by using traveling-wave electrode designs [19, 20]. For fast switching, however, electro-optic switches generally require significant voltages, resulting in high power consumption. Some research into electro-optic switches has focused on using polymers to achieve low power, ultra-high bandwidth switches [21]. Switches having a <1 V $V_\pi$ [22] and switching at over 100 GHz [23] have both been demonstrated.

Semiconductor optical amplifier (SOA) technology can also be used to create switch matrices. Figure 2-8 shows a design for a 2x2 switch using SOAs [24]. Electrical
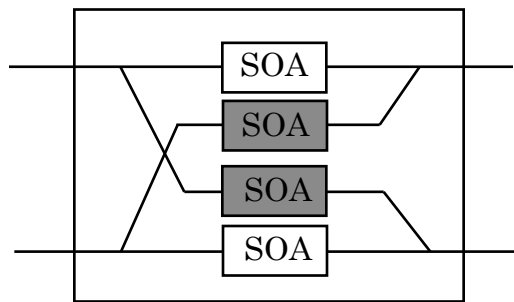


Figure 2-8: One design for an SOA-based 2x2 switching matrix with high extinction. The white and grey SOAs turn on and off simultaneously to switch between "Cross" and "Bar" operation. When the white SOAs are on, the switch is in the "Bar" state. When the grey SOAs are on, the switch is in the "Cross" state.

current to the white and gray SOAs is turned on and off to switch between "Cross"

and "Bar" switch states. When an SOA is pumped with current, it acts as an amplifier. When the current is turned off, the SOA material is very lossy. When the white SOAs are turned on and the grey SOAs are off, the switch is in the "Bar" state. When the grey SOAs are on and the white SOAs are off, the switch is in a "Cross" state. One major advantage of an SOA-based switch matrix is the ability to fully integrate a future all-optical router. Furthermore, an SOA-based switch matrix provides gain to the packets as they progress through the matrix. SOA switch matrices offer similar speeds and extinctions to those of electro-optic switches, and are also designed to be controlled electrically [25]. However, control can be more complex and the SOAs add additional noise to the packets.

Thus far, we have mainly focused on spatial switches, as that is the predominant switch matrix methodology. However, there are other options, such as wavelength switching, where wavelengths are used to differentiate the output port paths, as shown in Figure 2-9. Wavelength switching has the advantage of being well suited to optical
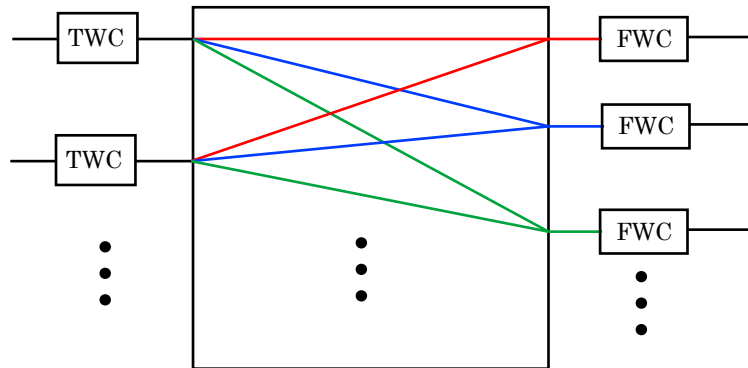


Figure 2-9: A wavelength switch design. TWC = tunable wavelength converter. FWC = fixed wavelength converter. As wavelength channels enter the switch, the tunable wavelength converter converts each packet to the appropriate wavelength, which is then forwarded to the correct output port through an arrayed-waveguide grating. Fixed wavelength converters at the output convert the packet back to the correct channel wavelength for transmission.

techniques, where it is relatively simple to differentiate between different wavelengths. However, it requires fast tunable laser diodes or filters and wavelength converters, which can still be a challenge. Packet switching using this technology has been demonstrated in the LASOR and IRIS programs [26, 27].

### 2.2.3  Contention and Buffering

Contention occurs in a switch when two or more packets need be switched to the same output port or when a packet is blocked. This can happen at even very light loads, but obviously occurs much more frequently with heavy traffic loads. There is a variety of methods for dealing with contention in a switch, and the choice of methods can significantly affect the latency, throughput, and complexity of a switch. The naive response to congestion is to drop one of the two packets which are contending for the same output port. We examine a simple example of this by assuming packets arrive independently and with a uniform distribution, i.e., that a packet arriving at port A has a 50% probability of being addressed to port C and 50% probability of being addressed to port D. Furthermore, we assume maximal load, such that there is always a packet at the input of the switch matrix ($P(0) = 1$). This is known as saturation analysis. We assume the switch has $m$ stages, and $P(m)$ is the probability that there is a packet at the input to the $m$th stage. We look at a single $m$th-stage 2×2 switch (Figure 2-10(a)) in this switch matrix, where A and B are input ports and C is the output port under examination.

We wish to calculate the throughput of the switch. Under this maximally loaded situation, the throughput is simply the probability that there is a packet at the output port. The probability that there is no packet at port C is $1 - P(m + 1)$. As
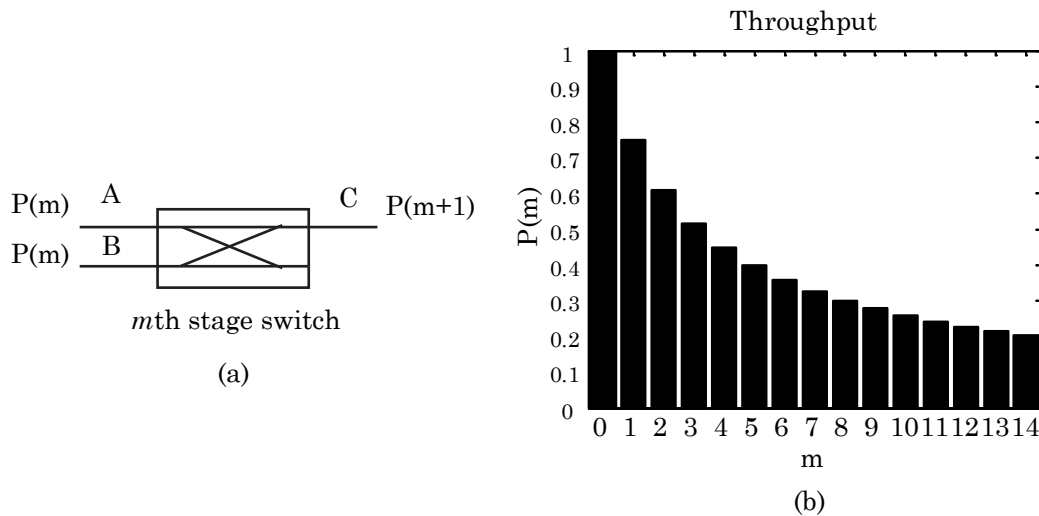
Figure 2-10: (a) A single crossbar 2x2 switch in the $m$th stage of a switch fabric. (b) Throughput decreases dramatically without contention resolution.

the incoming packets are independent and identical,

$$1 - P(m+1) = \text{Probability neither A nor B sent a packet to C} \tag{2.1}$$

$$= (1 - \text{Probability A sent a packet to C}) \times \tag{2.2}$$

$$(1 - \text{Probability B sent a packet to C}) \tag{2.3}$$

$$= (1 - \frac{1}{2} \times \text{Probability there was a packet at A})^2 \tag{2.4}$$

$$= (1 - \frac{1}{2}P(m))^2 \tag{2.5}$$

Since $P(0) = 1$, we can use this equation to recursively solve for $P(m+1)$, or the throughput of the switch. We plot this result in Figure 2-10(b), where we can see that the throughput declines dramatically as the number of switch stages increases. This result can be worse given a non-uniform distribution, since a higher demand for one output port will limit the throughput even more strongly. This severely limits the size of our switch fabric and makes a strong argument for alternative solutions.

Another method for dealing with contention is to use deflection routing. With deflection routing, when two packets contend, the losing packet simply gets forwarded to the wrong port in the hope that the next node will have the capacity to forward it

on towards the correct destination. This method requires no buffering; in a sense, it is using the network itself as the buffer. However, one might expect that large delays can be incurred by the packet, especially for heavily-congested networks. Calculating bounds for this delay can be complex, depending on the routing algorithm, network topology, and deflection strategy [28, 29]. However, experimental simulations have also shown that it can be possible to achieve up to 80% of the throughput of a buffered system [30]. Deflection routing is naturally very appealing for all-optical systems, as it requires little to no memory and takes advantage of the extremely fast switch speeds. Furthermore, deflection routing naturally adapts to traffic patterns and avoids hot-spots in the network by deflecting packets away from regions of high congestion. Depending on how deflection strategies are implemented, it can also allow networks to assign different priorities to packets and guarantee services. However, one major disadvantage of deflection routing is that packets can arrive significantly out-of-order at the end nodes. Packet-switched networks can deal with occasional rearrangement of packets, but current protocols (especially TCP) handle out of order packets very poorly.

Buffering is the third and most common approach to contention resolution. Buffers can be placed at the input to a switch, the output of the switch, and also internally distributed through the switch. Buffers can also be arranged in various ways, such as first-in-first-out (FIFO), last-in-first-out (LIFO), or can be accessed using a more complex buffer management system. Input-buffered switches suffer from several disadvantages, including head-of-line (HOL) blocking. This occurs when packets at the front of the buffer contend, blocking later packets in the buffer from being switched (Figure 2-11). Head-of-line blocking limits the maximum throughput to 56% of the total available switch throughput (which is the number of links × link capacity) [31]. Sophisticated buffer management schemes can be designed to avoid HOL blocking at the cost of increasing complexity, latency, and reducing scalability in the switching node [32].

Internal buffering means that each buffer can be short, but it usually requires more individual buffers. This makes memory access quick, but buffer control can be
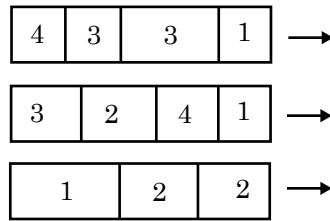
Figure 2-11: Packets at the front of the queue contend, blocking later packets from being switched. In this example, two packets with an address of 1 contend, but the packets behind them (with addresses of 3 and 4) are also affected even though they can be switched without impeding the first two packets.

more complex. With electronic memory, this is also more costly than fewer, longer buffers.

Output buffering means that buffering occurs after the switch correctly forwards the packets. This is more efficient due to the fact that each output port has its own queue, and thus never blocks another packet heading elsewhere. The disadvantage is in order to provide a high throughput, the switch must operate much faster than the incoming packet rate in order to handle every possible packet arriving at every input link. One example is the case where all N incoming packets wish to exit on the same output port. If our non-blocking NxN switch can operate N times faster than the incoming packet rate, we can switch all incoming packets to the correct output buffer before the next set of packets arrive. Otherwise, we will need to drop or deflect packets. Output buffering has been shown to result in smaller mean waiting times than input buffering [31] and is thus used in the majority of buffers today.

Optical buffers are still fairly rudimentary when compared to electronic buffers. Two options for all-optical buffering are a recirculating loop or a series of delay lines (Figure 2-12). A recirculating loop buffer consists of a loop which is the length of multiple packets. Packets enter and are kept circulating (with a regenerator to compensate for the fiber distortions) until switched out. A simple optical regenerator can be used in this buffer. Another possible arrangement consists of a series of switched delay lines in a base-2 sequence. The first delay line is one packet length, the second is two packet lengths, the third is four packet lengths, etc. By controlling

54
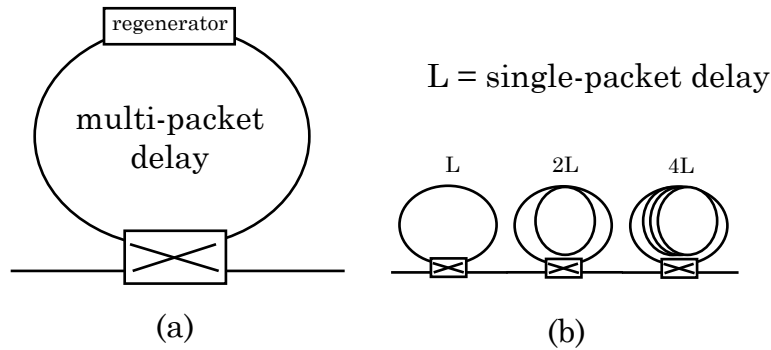
Figure 2-12: Two types of optical buffers which can be implemented today. (a) A multi-packet loop, which is essentially a FIFO delay. A regenerator is necessary to maintain the packet power and signal quality since the packet may spend multiple passes in the loop. (b) A multi-loop system which enables a finer granularity for buffer delays. Each packet can be delayed for any multiple of L, the packet-length. Furthermore, each packet can be delayed a different amount.

the sequence of delay lines a packet passes through, we can control exactly how much delay each packet sees. For instance, if we switch the packet into the first delay line and then the third one, we get a total delay of five packet-lengths. However, this latter approach requires fixed-length packets.

Another option that has been demonstrated to be capable of long, variable delays is to use wavelength converters and dispersion to determine the delay experienced by a packet. With this technique, a fast wavelength converter converts the input packet to an intermediate wavelength, which is then transmitted down a length of fiber. The wavelength of the packet determines the group delay it experiences due to dispersion in the length of fiber. At the output, a second wavelength converter converts the packet back to its original wavelength. Careful dispersion and chirp compensation is necessary to ameliorate higher-order dispersion and nonlinearities induced by fiber lengths. Discrete delays can be more easily compensated for and have been demonstrated at up to 7 $\mu$s of total delay [33, 34]. Continuously variable delays are more difficult to accurately compensate over the entire delay range and have been demonstrated at over 40 ns of buffering [35].

Recently, some excitement has been generated by the research into "slow light" for optical buffering. In materials, a sharp change in either gain or loss is usually

accompanied by a sharp change in the index of refraction. By exploiting this resonance, large delays or advancement of light pulses can be achieved. Tunable delays can be achieved by using either optical or electronic techniques to tune the index of refraction. Currently, the majority of slow light demonstrations have achieved delays which are less than a pulse width, though some results have demonstrated delays of around one pulse width [36, 37, 38, 39]. While such short delays are not useful for optical buffer implementations, they can still be important for synchronization. Furthermore, arbitrarily long delays have been shown to be theoretically possible [40].

Since buffering using optical techniques is still rudimentary, all-optical networks should involve as few buffers as possible, potentially using an entirely bufferless architecture. Reducing buffer requirements in networks also helps electronic processing as well. One important question, then, is exactly how much buffering any network needs. Recent research suggests that a significant reduction in buffer size can be achieved without impacting network performance and throughput. The rule of thumb currently used for buffers is $RTT \times C$, where $RTT$ is the round-trip-time of the link and $C$ is the capacity of the link. This is a result of the windowing method of TCP, the congestion control protocol in wide use on the Internet. However, Mckeown et al. [41] shows that in situations where there are $N$ TCP flows (as there are especially in core routes), this rule of thumb can be significantly reduced to $\frac{RTT \times C}{\sqrt{N}}$ with negligible impact on throughput due to the desynchronized nature of multiple TCP flows. This can reduce the buffer requirement from roughly a half-million packets (a 10 Gb/s channel with a 250 ms $RTT$ and assuming an average packet size of 400 bytes) to 5000 packets (assuming 10,000 TCP flows). In certain conditions, networks may need only buffers of 20 packets [42].

## 2.3    All-Optical Header Processing for Packet Routing

In the previous sections, we described the fundamentals of switching and routing in data networks, going into some detail as to where optical techniques can supplement or replace electronic processing in order to meet growing demand. Here, we put these ideas together in a demonstration of an all-optical switch which can simultaneously forward two 40-Gb/s optical packet streams to the correct output ports. We demonstrate highly scalable all-optical header processing using only two SOA-based optical logic gates operating at 40 Gb/s [43, 44]. In this experiment, we focus on packet forwarding and all-optical header processing, and do not implement techniques for buffering or contention resolution. Thus, the packet switch demonstrated here can be used in a network with deflection routing.

We briefly summarize the many advantages of optical processing described previously. First of all, header processing using optical techniques can provide ultra-low processing latencies [45], which reduces buffering requirements in the switch. Optical processing also offers other advantages: power consumption which scales well with increasing channel data rates, fewer O/E/O conversions, transparency to payload bit rate and modulation format [46, 47], shorter buffer lengths, and a more straightforward switch structure requiring fewer input/output ports for the same bandwidth due to the ability to process higher data rates per channel.

Figure 2-13 shows a generic design for an all-optical switch. Packets arriving on the left first enter an input interface, in which synchronization, amplification or regeneration, and buffering occurs. The header of the packet is tapped off and enters the header processing unit, which configures the transparent switch matrix for the payload. The packets exit on the right, having been correctly forwarded. Depending on the loss and degradations in the switch matrix, an output interface may need to provide additional amplification or regeneration. Output buffering may also be implemented in the output interface.
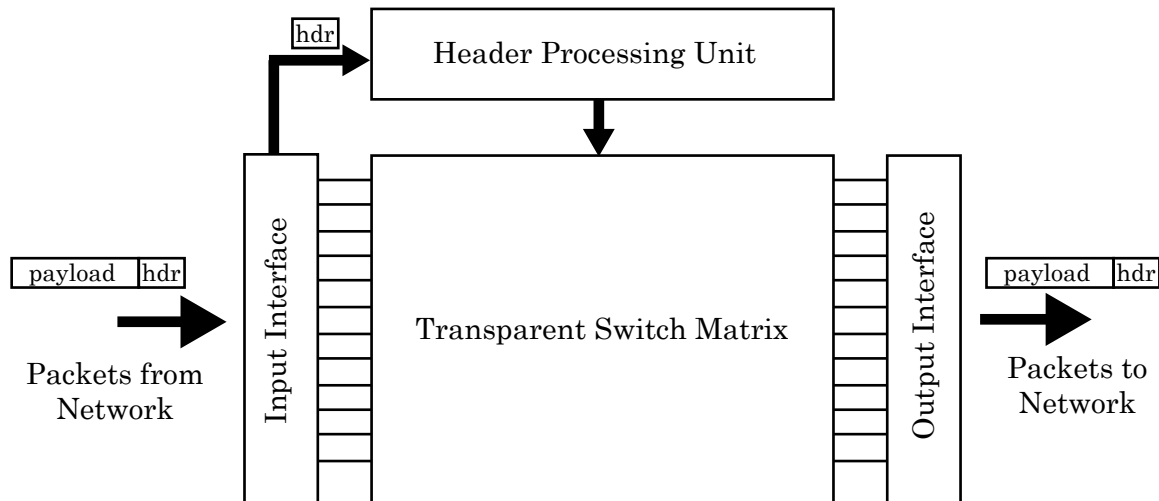
Figure 2-13: Generic all-optical switch design.

## 2.3.1 Implementation and Design

For our implementation, we choose to use the Banyan switch architecture for the transparent switch matrix. As described previously in Section 2.2.2, header processing is modular: only the $j$th address bit is needed to process the correct orientation of the $j$th column in the Banyan. We implement this switch matrix using lithium niobate 2x2 electro-optic switches, providing a transparent switch matrix with high extinction and low loss. Figure 2-14 illustrates our implementation of the optical switch.

To perform header processing with as few all-optical switches as possible, we define an "Empty/Full" bit, which indicates the presence or absence of data in a packet. By defining an "Empty/Full" bit, we are able to eliminate several possibilities in the Karnaugh map for the header processor (Figure 2-15) and perform header processing using only two optical logic gates. For example, when both packets are "Empty", any forwarding decision is acceptable. When both packets are "Full" and contending, we define a default forwarding decision and can resolve the contention either through input buffering, output buffering, or deflection routing. This resolution contention method can be defined around our choice of logic and implemented later. By using the "Empty/Full" bit, we can perform the necessary header processing by using only two
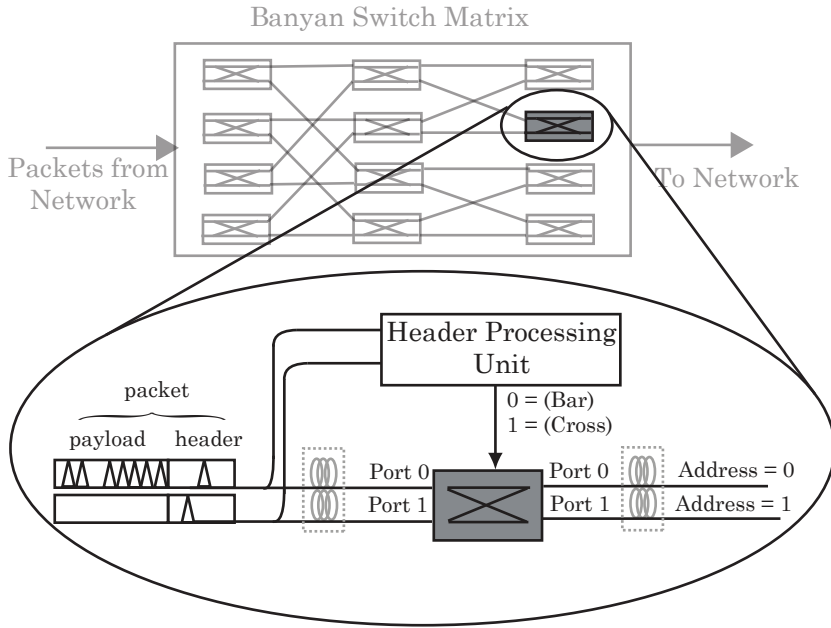
Figure 2-14: Diagram of the network architecture for a unit cell of the router. Buffering and contention resolution were not implemented for this experiment.

all-optical switches. We use one switch to generate the logic ($A_0 \cdot E_0$) and another to generate the logic ($\bar{A}_1 \cdot E_1$). A simple passive coupler can perform the OR operation.

Thus, we obtain the proper logical processing for the two incoming packet streams: if an arriving packet is "Empty" ($E_k = 0$), the corresponding address ($A_k$) is ignored. If both arriving packets are "Empty", the logic defaults to a "Bar" switch state (Output $= 0$). If a "Full" packet ($E_k=1$) arrives, the state of the 2x2 spatial switch is set by the packet address if it arrives on port 0 ($A_0$) or by the inverse of the packet address if it arrives on port 1 ($A_1$). This is because packets arriving on different input ports of the 2x2 spatial switch require opposite switch states ("Bar" vs. "Cross") to be forwarded to the same output address. A packet arriving on port 0 and desiring address 0 ($A_0=0$) requires the "Bar" state (output $= 0$), where a packet arriving on port 1 and desiring the same address ($A_1=0$) requires the "Cross" state (output $= 1$). If both incoming packets are "Full" ($E_0 = E_1 = 1$) and addressed to the same output ($A_0=A_1$), we prioritize the "Cross" switch state. In this experiment, we implement the "Empty/Full" control bit as an extra header bit. An alternative method is to use
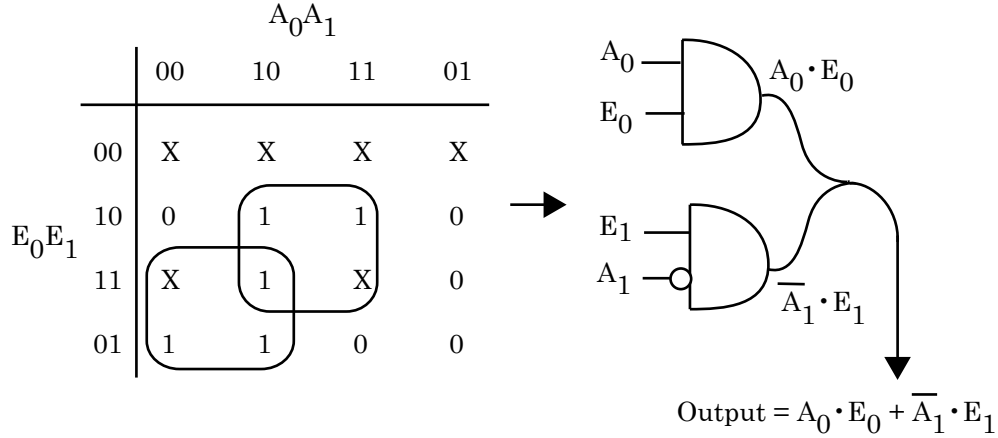
Figure 2-15: Illustration of the advantages of an "Empty/Full" bit. This allows us to implement 2x2 switching with only 2 all-optical logic gates. In this figure, $A_k$ and $E_k$ indicate the address and "Empty/Full" bit for packets arriving at port k, respectively.

an average energy detection method to generate the control bit. The latter reduces header overhead, but increases the complexity of the router.

For the two optical logic gates needed, we use two ultrafast nonlinear interferometers (UNI) (Figure 2-16). The ultrafast nonlinear interferometer is a SOA-based gate capable of ultrafast operation at 100 Gb/s and higher [48, 49, 50]. The signal pulses (shown in red) enter the UNI and are split by the birefringent fiber into two orthogonally-polarized and temporally separated pulses, which then enter the SOA. In the absence of a control pulse, the two signal pulses recombine either constructively or destructively depending on the bias of the SOA. When the control pulse is coupled into the SOA as well, it is timed to arrive in between the two signal pulses. The control pulse power affects the gain and phase of the SOA, but only the second of the signal pulses experiences the effect. The control pulse power is set to induce a $\pi$ phase shift between the two signal pulses. Thus, when the two signal pulses recombine, in the presence of the control pulse, the output is changed from either constructive to destructive interference (non-inverting mode) or destructive to constructive interference (inverting mode). The control pulse is filtered out at the output. Thus, the UNI can either produce the logic $S{\cdot}C$ (in the non-inverting mode) or $S{\cdot}\bar{C}$ (in the inverting mode). This allows us to achieve the correct optical logic for header processing.
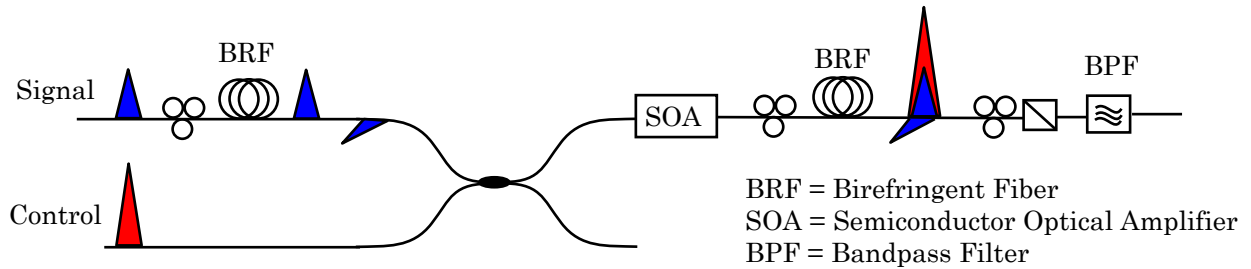
Figure 2-16: Schematic diagram of an ultrafast nonlinear interferometer (UNI). The signal pulse, shown in red, enters the UNI and is split into two orthogonally-polarized, temporally separated copies. The control pulse, shown in blue, is timed to arrive at the SOA in between the two signal pulse copies. This induces a gain and phase shift in the SOA which affects the second pulse but not the first. When the two signal pulse copies are recombined, they interfere constructively or destructively depending on their induced relative phases.

In this configuration, our "Empty/Full" and address bits need to be on different wavelengths, in order to distinguish the signal from the control pulses at the output. (It is possible to use the UNI gates in a counter-propagating configuration with same wavelength signal and control pulses, but this limits the speed of the switch due to a trade-off between achieving a full $\pi$ phase shift and the propagation distance in which the signal pulse is affected by the control pulse.) In the co-propagating situation, we can use a single optical logic gate to demultiplex the "Empty/Full" and address bits from the header onto two separate wavelengths [51]. For simplicity, we do not implement this demultiplexer and instead create our "Empty/Full" and address bits by using two, synchronized mode-locked fiber lasers tuned to separate wavelengths. In an average power detection scheme, the "Empty/Full" bit can be generated on an arbitrary wavelength.

## 2.3.2   Experimental Setup

Figure 2-17 shows the schematic of this experimental implementation of 40-Gb/s all-optical header processing, including the packet and address generator, the "Empty/Full" bit generator, header processor, and 2x2 spatial electro-optic switch. For packet and
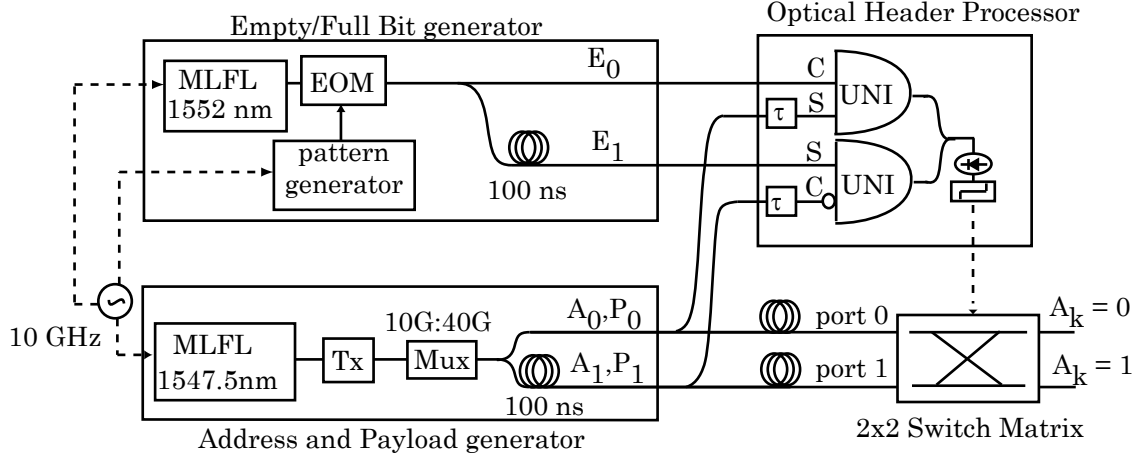
Figure 2-17: Experimental schematic of our implementation of ultrafast all-optical header processing. MLFL is a mode-locked fiber laser, Tx is the transmitter, EOM is the electro-optic modulator, MUX is a 1:4 free-space multiplexer and UNI is the ultrafast nonlinear interferometer. At the UNI inputs, C indicates the control while S indicates the signal. $\tau$ shows a fixed delay determining the relevant address bit for the 2x2 electro-optic switch. Finally, $A_k$, $E_k$, and $P_k$ are the address, "Empty/Full" bit, and payload of the packets arriving at port $k$ of the 2x2 spatial switch.

address generation, we use a mode-locked fiber laser which produces 2-ps pulses at 1547.5 nm modulated at 10 Gb/s with a $2^7$-1 pseudo-random bit sequence. We multiplex this sequence in a free-space optical multiplexer to 40 Gb/s. By using the proper delays in the multiplexer, we avoid placing adjacent bits next to each other and thus maintain the pseudo-random statistics of the bit-pattern [52]. We define our packets as fixed-length, 100-ns packets with one address bit and "Empty/Full" bit per packet. This results in each packet consisting of 3999 bits of payload, 1 address bit, and 1 "Empty/Full" bit (at a different wavelength). Since $2^7$-1 (127) is prime with respect to 4000, or 1000, each address bit will cycle through the entire pseudo-random bit sequence. This tests the robustness of the header-processing unit against address-pattern-dependent effects for forwarding packets with 7-bit addresses through a full-scale router with 128 input ports. To create two distinct packet streams, we split the output of the packet and address generator and introduce a single-packet delay of 100 ns between the port 0 and port 1 packet streams.

We create the "Empty/Full" bits using a second mode-locked fiber laser at 1550 nm, synchronized to the first laser. We modulate the pulses with a "1001" pattern at 10 MHz, which corresponds to one bit every 100 ns. Again, we introduce a single-packet delay between the port 0 and port 1 packet streams. Since this pattern is prime with the $2^7$-1 pattern, we therefore ensure that we explore all possible combinations of address and "Empty/Full" bits in this demonstration.

The optical header processor unit, as described earlier, consists of two UNI logic gates, a non-inverting UNI performing ($A_0 \cdot E_0$) and an inverting UNI performing the logic ($\bar{A}_1 \cdot E_1$). For both UNI logic gates, we use Alcatel SOAs [53] with recovery times on the order of 100 ps at a bias current of 200 mA. The switching energies for the logic gates are 27.5 fJ (non-inverting) and 33 fJ (inverting). The birefringent fiber in the UNI set the delay between the signal pulse copies. For the non-inverting gate, we use a 10-ps delay while for the inverting gate, we use a 5-ps delay. This difference was due to the availability of devices. A longer delay results in higher timing tolerance for the control pulse input but limits the speed of the UNI since we cannot allow adjacent signal pulse copies to overlap. Furthermore, a longer delay means the two signal pulses may experience different operating conditions in the SOA. At our operating speed of 40 Gb/s, however, we can obtain effective switching with either 5-ps or 10-ps delays.

Fixed delay lines on the address-bit inputs to the two UNIs determine which address the header processor unit operates on. The assignment of signal and control for the inverting UNI is fully defined: since the address bit ($A_1$) is to be inverted, it has to be the control input of the UNI. The output then is at the wavelength of the signal input ($E_1$), 1552 nm. The assignment of the signal and control for the non-inverting UNI, however, is not defined. We choose to reverse the assignment, using the address bit ($A_0$) as the signal, and the "Empty/Full" bit ($E_0$) as the control. Thus, the output is at the address wavelength, 1547.5 nm, so that we minimize the interferometric effects of the coupler. The coupler output provides the desired logic, $A_0 \cdot E_0 + \bar{A}_1 \cdot E_1$. Note that this output is a control bit which occurs once per packet, or every 100 ns.

We next need to use this output to drive the operation of our 2x2 spatial switch. We choose to use a commercially-available electro-optic lithium niobate 2x2 spatial switch with a fast switching speed (1 ns) and high extinction (> 25 dB). A 1-ns guard band is only 1% of the packet length. To measure the switching speed, we send in CW light and drive the 2x2 switch with a -4 dBm RF sinusoid over a range of frequencies. We measure the output amplitude of the CW light and locate the 3-dB point. The extinction of the switch determines the crosstalk between the ports. To measure the extinction, we send CW light into each input port and measure the output power across all DC bias voltages. Figure 2-18 shows the measurement of extinction and switching speed of our switch.
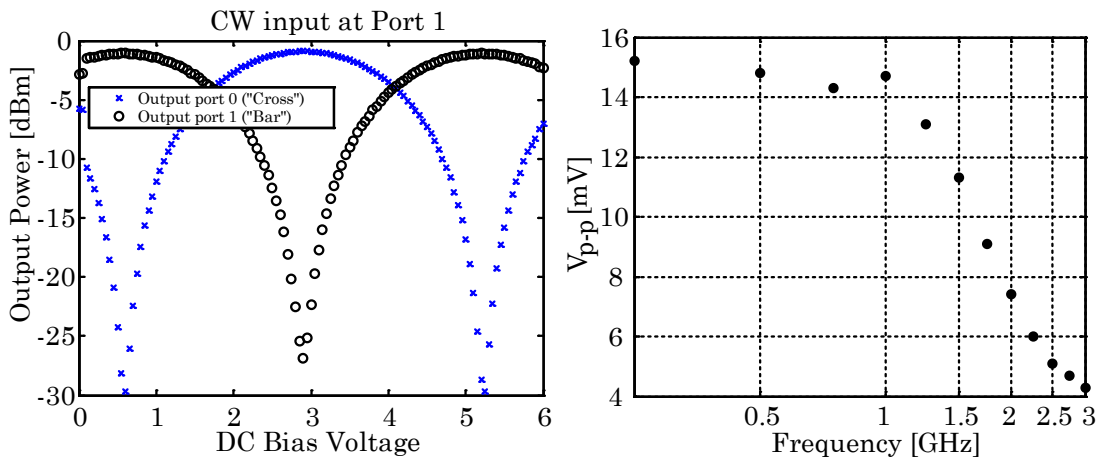


Figure 2-18: (a) Extinction of the electro-optic switch in "Cross" and "Bar" configurations with CW light at 1550 nm into port 1. We see an extinction of >25 dB. (b) Switching speed of the 2x2 switch is measured to be < 1 ns.

To drive an electro-optic switch, we need to convert the optical output (a 2-ps optical pulse) into an electronic pulse which is temporally matched to the length of the packet. This electronic pulse needs to hold the switch in the correct state for the duration of the packet being switched. We use a high-speed 18-GHz photodetector followed by a 10-GHz electronic D-flip-flop. The photodetector converts the optical pulse into a longer electronic pulse, which is then sampled by the D-flip-flop and held for the packet duration. The threshold on this D-flip-flop is set such that the output of the coupler results in an OR operation: the optical power of either UNI output as

well as both UNI outputs combined all result in a "1" detection. Only when there is no output from either UNI is there a "0" detection at the D-flip-flop. A high-speed detector and flip-flop is necessary here in order to maintain the small guard band and high extinction of the 2x2 electro-optic switch.

### 2.3.3   Results

The output of each of the two UNI logic gates is a 2-ps pulse every 100 ns, representing the decision logic made per packet. We measure the eye diagram of each output (Figure 2-19(a) and (b)) using a sampling scope. We also measure the packet error rate of the logic gates to be $1.7x10^{-9}$ for the non-inverting UNI and $3x10^{-9}$ for the inverting UNI. To measure this packet error rate we use our 10-Gb/s pre-amplified receiver programmed with the correct pattern. In Figure 2-19(c), we show the electronic output of the all-optical header processor, first with each UNI and then with the total combined output. This demonstrates the accuracy of the OR operation. We also see the operation of our high-speed photodetector and D-flip-flop do indeed convert the short optical pulses to long, packet-length electronic pulses with short rise and fall times. Using a 10-MHz error detector programmed with the correct pattern, we detect the total packet error rate of the header processing unit to be $1x10^{-6}$, comparable with electronic routers today. The degradation in packet error rate is likely due to timing and thresholding errors in the D-flip-flop as well as additional noise from the photodetector and interference in the coupler.

To show routing through the system, we use a 40-Gsample real-time oscilloscope and two 50-GHz photodetectors to measure the packets at the output of the 2x2 spatial electro-optical switch driven by the all-optical ultrafast header processor. Since the packet payloads are at 40 Gb/s, we are undersampling the output significantly. This is unavoidable due to our lack of higher-speed electronic oscilloscopes. To observe routing, we need to visually distinguish between packets addressed to the upper output port ($A_k = 0$) and those addressed to the lower output port ($A_k = 1$). We use two distinct patterns which can be visually identified even with undersampling: "11110000" (packet M) and "10" (packet N). We then program our packet stream
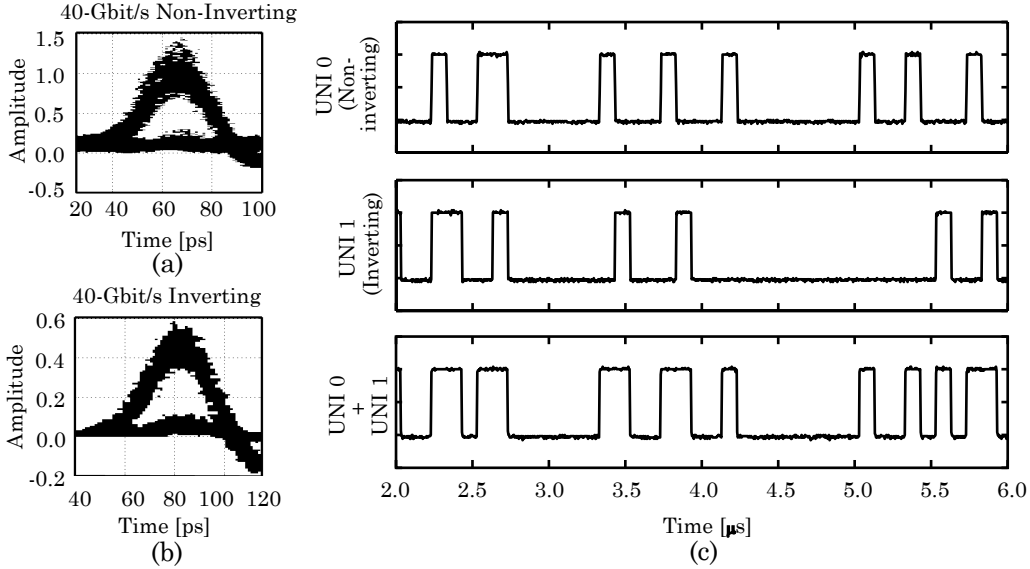
Figure 2-19: (a) Eye diagram of non-inverting ($A_0 \cdot E_0$) UNI. Packet error rate measured to be $1.7 \times 10^{-9}$. (b) Eye diagram of inverting ($\bar{A}_1 \cdot E_1$) UNI. Packet error rate measured to be $3 \times 10^{-9}$. (c) Electronic output of photodetector and D-flip-flop for each UNI alone and the combined coupled output of both UNIs. The packet error rate is measured to be $1 \times 10^{-6}$, comparable with current electronic routers.

in an "NMMN" pattern, with a 1.5-ns guard band. Thus, when our two distinct packet streams are created with a single-packet delay, we obtain as the input to port 0 "NMMN" and as the input to port 1 "MMNN". We continue to use "1001" as our "Empty/Full" bit pattern. We use variable delays on the address inputs to the UNI logic gates to select one bit in the packet as the address bit and to ensure a non-degenerate result from the header-processing logic. Previously, we tested all possible combinations of "Empty/Full" bits and address bits. Here, we only intend to visually demonstrate correct routing and thus present only 4 interesting cases: (1) both packets are full and contending (2) the port 0 packet is full and desires the "Cross" state while the port 1 packet is empty (3) both packets are empty and (4) the port 0 packet is empty while the port 1 packet is full and desires the "Bar" state. We show the results of these measurements in Figure 2-20.
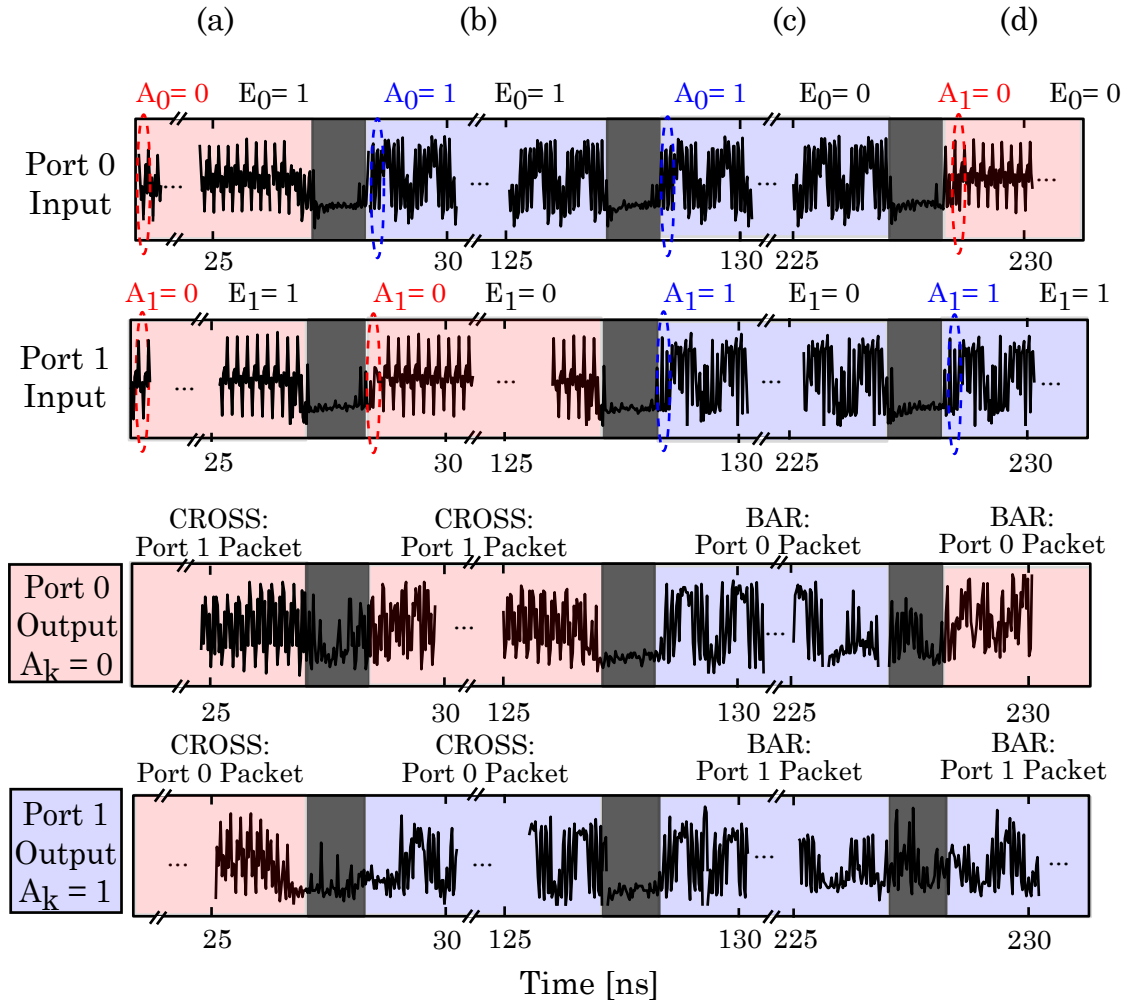
Figure 2-20: System demonstration of how packets are routed by the ultrafast all-optical header processing unit. Red/pink indicates the packet address is "0" and payload is "1010", while blue/pale-blue indicates a packet address of "1" and payload of "11110000". Dark grey indicates the guard band. We demonstrate 4 situations: (a) both packets are full and contending (b) the port 0 packet is full and desires the "Cross" state while the port 1 packet is empty (c) both packets are empty and (d) the port 0 packet is empty while the port 1 packet is full and desires the "Bar" state.

## 2.4 Conclusion

In this chapter, we have described how optical processing can help supplement routing and switching in networks. To demonstrate this point, we have proven the capability of all-optical header processing to accurately handle 2 simultaneous 40-Gb/s packet streams with a packet error rate of $1\mathrm{x}10^{-6}$, on par with electronic routers today [54]. This experiment also expands optical processing functionality to multi-packet routing. As core network channel data rates increase beyond 40 Gb/s, such all-optical processing techniques will become critical.

In this demonstration, our optical signal processing techniques outstrip electronic capabilities in terms of speed and switching energies, but not complexity, footprint, or cost. Furthermore, one of the main motivations for all-optical header processing is to reduce the packet-processing latency of the header processing unit. In this experiment, our packet processing delay is dominated by the 62.5-ns fiber propagation time in these discrete UNI logic gates. Integrated all-optical logic gates can reduce this delay to below 1-ns, since propagation distances involved are much shorter. By enabling mass-production of these logic gates, costs can also be brought down. Thus, Chapter 3, we turn our attention to integrated logic gates. We focus mainly on the semiconductor optical amplifier, which provides high nonlinearities, gain, and ease of integration.

# Bibliography

[1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. v. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, pp. 1–15, 1994.

[2] Cisco, "The exabyte era." white paper, Jan. 2008.

[3] Cisco, "Global ip traffic forecast and methodology, 2006-2011." white paper, Jan. 2008.

[4] J. V. der Merwe, S. Sen, and C. Kalmanek, "Streaming video traffic: Characterization and network impact," in *Proceedings of the 7th International Workshop on Web Content Caching and Distribution, Boulder, CO, USA*, Aug. 2002.

[5] R. Govindan and A. Reddy, "An analysis of internet inter-domain topology and route stability," in *Proceedings of the Sixteenth Annual joint Conference of the IEEE Computer and Communications Societies (INFOCOM '97)*, pp. 850–857, Apr. 1997.

[6] C. Labovitz, A. Ahuja, and F. Jahanian, "Experimental study of internet stability and backbone failures," in *Twenty-Ninth Annual International Symposium on Fault- Tolerant Computing*, p. 278, 1999.

[7] L. L. Peterson and B. S. Davie, *Computer Networks: A Systems Approach*, section 4.5, pp. 340–346. Morgan Kaufmann Publishers, third ed., 2003.

[8] A. Banerjee, J. Drake, J. P. Lang, B. Turner, K. Kompella, and Y. Rekhter, "Generalized multiprotocol label switching: An overview of routing and management enhancements," *IEEE Communications Magazine*, vol. 39, no. 1, pp. 144–150, 2001.

[9] C. Qiao and M. Yoo, "Choices, features and issues in optical burst switching," *Optical Networks Magazine*, vol. 1, pp. 36–44, 2000.

[10] C. Qiao and M. Yoo, "Optical burst switching (OBS) - a new paradigm for an optical internet," *Journal of High Speed Networks*, vol. 8, no. 1, pp. 69–84, 1999.

[11] A. Ge, F. Callegati, and L. S. Tamil, "On optical burst switching and self-similar traffic," *Electronics Letters*, vol. 4, no. 3, pp. 98–100, 2000.

[12] M. C. Cardakli, S. Lee, A. E. Wilner, V. Grubsky, D. Starodubov, and J. Feinberg, "Reconfigurable optical packet header recognition and routing using time-to-wavelength mapping and tunable fiber Bragg gratings for correlation decoding," *IEEE Photonics Technology Letters*, vol. 12, no. 5, pp. 552–554, 2000.

[13] I. Glesk, J. P. Solokoff, and P. R. Prucnal, "All-optical address recognition and self-routing in a 250 Gbit/s packet-switched network," *Electronics Letters*, vol. 30, no. 16, pp. 1322–1323, 1994.

[14] D. Cotter, J. K. Lucek, M. Shabeer, K. Smith, D. C. Rogers, D. Nesset, and P. Gunning, "Self-routing of 100 Gbit/s packets using 6 bit 'keyword' address recognition," *Electronics Letters*, vol. 31, no. 25, pp. 2201–2202, 1995.

[15] S. A. Hamilton and B. S. Robinson, "40-Gb/s all-optical packet synchronization and address comparison for otdm networks," *IEEE Photonics Technology Letters*, vol. 14, no. 2, pp. 209–211, 2002.

[16] F. Ramos, E. Kehayas, J. M. Martinez, R. Clavero, J. Marti, L. Stampoulidis, D. Tsiokos, H. Avramopoulos, J. Zhang, P. V. Holm-Nielsen, N. Chi, P. Jeppesen, N. Yan, I. T. Monroy, A. M. J. Koonen, M. T. Hill, Y. Liu, H. J. S. Dorren, R. V. Caenegem, D. Colle, M. Pickavet, and B. Riposati, "IST-LASAGNE: Towards all-optical label swapping employing optical logic gates and optical flip-flops," *Journal of Lightwave Technology*, vol. 23, no. 10, pp. 2993–3011, 2005.

[17] A. Kos, M. Pustisek, and J. Bester, "Characteristics of real packet traffic captured at different network locations," in *Eurocomn 2003*, vol. 1, p. 164, 2003.

[18] K. Claffy, G. Miller, and K. Thompson, "The nature of the beast: Recent traffic measurements from an internet backbone." `http://www.caida.org/publications/papers/1998/Inet98/`.

[19] R. Alferness, S. Korotky, and E. Marcatili, "Velocity-matching techniques for integrated optic traveling wave switch/modulators," *IEEE Journal of Quantum Electronics*, no. 3, pp. 301–309, 1984.

[20] E. L. Wooten, K. M. Kissa, A. Yi-Yan, E. J. Murphy, D. A. Lafaw, P. F. Hallemeier, D. Maack, D. V. Attanasio, D. J. Fritz, G. J. McBrien, and D. E. Bossi, "A review of lithium niobate modulators for fiber-optic communications systems," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 6, no. 1, pp. 69–82, 2000.

[21] Y. Shi, C. Zhang, H. Zhang, J. H. Bechtel, L. R. Dalton, B. H. Robinson, and W. H. Steier, "Low (Sub-1-volt) halfwave voltage polymeric electro-optic modulators achieved by controlling chromophore shape," *Science*, vol. 288, pp. 119–122, 2008.

[22] R. A. Norwood, "Electro-optic polymer modulators for telecommunications applications," in *Optical Fiber Communications Conference (OFC '08)*, 2008.

[23] D. Chen, H. R. Fetterman, A. Chen, W. H. Steier, L. R. Dalton, W. Wang, and Y. Shi, "Demonstration of 110 GHz electro-optic polymer modulators," *Applied Physics Letters*, vol. 70, p. 3335, 1997.

[24] W. van Berlo, M. Janson, L. Lundgren, A.-C. Morner, J. Terlecki, M. Gustavsson, P. Granestrand, and P. Svensson, "Polarization-insensitive, monolithic 4×4 InGaAsP-InP laser amplifier gate switch matrix," *IEEE Photonics Technology Letters*, vol. 7, no. 11, pp. 1291–1293, 1995.

[25] K. A. Williams, G. F. Roberts, T. Lin, R. V. Penty, I. H. White, M. Glick, and D. McAuley, "Integrated optical 2×2 switch for wavelength multiplexed interconnects," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 11, no. 1, pp. 78–85, 2005.

[26] Z. Pan, H. Yang, J. Yang, J. Hu, Z. Zhu, J. Cao, K. Okamoto, S. Yamano, V. Akella, and S. J. B. Yoo, "Advanced optical-label routing system supporting multicast, optical TTL and multimedia applications," *Journal of Lightwave Technology*, vol. 23, no. 10, pp. 3270–3281, 2005.

[27] D. T. Neilson, D. Stiliadis, and P. Bernasconi, "Ultra-high capacity optical IP routers for the networks of tomorrow: IRIS project," in *31st European Conference on Optical Communications*, vol. 5, p. 45, 2005.

[28] B. Hayek, "Bounds on evacuation time for deflection routing," *Distributed Computing*, vol. 5, pp. 1–6, 1991.

[29] J. T. Brassil and R. L. Cruz, "Bounds on maximum delay in networks with deflection routing," *IEEE Transactions on Parellel and Distributed Systems*, vol. 6, no. 7, pp. 724–732, 1995.

[30] N. F. Maxemchuk, "Comparison of deflection and store-and-forward techniques in the Manhattan Street and Shuffle-Exchange networks," in *Proceedings of the Eigth Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '89)*, vol. 3, p. 800, 1989.

[31] M. J. Karol, M. G. Hluchyj, and S. P. Morgan, "Input versus output queueing on a space-division packet switch," *IEEE Transactions on Communications*, no. 12, pp. 1347–1356, 1987.

[32] J. Duato, J. Flich, and T. Nachiondo, "A cost-effective technique to reduce HOL blocking in single-stage and multistage switch fabrics," in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium*, p. 48, Feb. 2004.

[33] C.-H. Chen, L. A. Johansson, V. Lal, M. L. Masanovic, D. J. Blumenthal, and L. A. Coldren, "Programmable optical buffering using fiber bragg gratings combined with a widely-tunable wavelength converter," in *Optical Fiber Communication Conference (OFC '05)*, Mar. 2005.

[34] C.-H. Chen, D. Wolfson, L. A. Johansson, and D. J. Blumenthal, "Demonstration of 40 Gbit/s optical packet synchronisation using fibr bragg gratings and fast-tunable wavelength converters," *Electronics Letters*, vol. 42, no. 6, pp. 367–369, 2006.

[35] Y. Wang, C. Yu, L. Yan, A. E. Willner, R. Roussev, C. Langrock, M. M. Fejer, J. E. Sharping, and A. L. Gaeta, "44-ns continuously tunable dispersionless optical delay element using a PPLN waveguide with two-pump configuration, DCF, and a dispersion compensator," *IEEE Photonics Technology Letters*, vol. 19, no. 11, pp. 861–863, 2007.

[36] Y. O. M. S. Bigelow, J. E. Sharping, Z. Zhu, A. Schweinsberg, D. J. Gauthier, R. W. Boyd, and A. L. Gaeta, "Tunable all-optical delays via brillouin slow light in an optical fiber," *Physical Review Letters*, vol. 94, p. 153902, 2005.

[37] M. F. Yanik and S. Fan, "Stopping light all optically," *Physical Review Letters*, vol. 92, no. 8, p. 083901, 2004.

[38] Y. Okawachi, M. A. Foster, J. E. Sharping, and A. L. Gaeta, "All-optical slow-light on a photonic chip," *Optics Express*, vol. 14, no. 6, pp. 2317–2322, 2006.

[39] D. Dahan and G. Eisenstein, "Tunable all optical delay via slow and fast light propagation in a Raman assisted fiber optical parametric amplifier: a route to all optical buffering," *Optics Express*, vol. 13, no. 16, pp. 6234–6249, 2005.

[40] R. W. Boyd, D. J. Gauthier, A. L. Gaeta, and A. E. Willner, "Maximum time delay achievable on propagation through a slow-light medium," *Physical Review A*, vol. 71, p. 023801, 2005.

[41] G. Appenzeller, I. Keslassy, and N. McKeown, "Sizing router buffers," in *ACM SIGCOMM Computer Communication Review*, vol. 34, pp. 281–292, 2004.

[42] N. Beheshti, Y. Ganjali, R. Rajaduray, D. Blumenthal, and N. McKeown, "Buffer sizing in all-optical packet switches," in *Optical Fiber Communication Conference*, 2006.

[43] J. P. Wang, B. S. Robinson, S. A. Hamilton, and E. P. Ippen, "40-Gbit/s all-optical header processing for packet routing," in *Optical Fiber Communication Conference (OFC)*, 2006.

[44] J. P. Wang, B. S. Robinson, S. A. Hamilton, and E. P. Ippen, "Demonstration of 40-Gbit/s packet routing using all-optical header processing," *IEEE Photonics Technology Letters*, vol. 18, no. 21, pp. 2275–2277, 2006.

[45] B. A. Small, A. Shacham, and K. Bergman, "Ultra-low latency optical packet switching node," *IEEE Photonics Technology Letters*, vol. 17, no. 7, pp. 1564–1566, 2005.

[46] D. J. Blumenthal, P. R. Prucnal, and J. R. Sauer, "Photonic packet switches: architectures and experimental implementations," in *Proceedings of the IEEE*, 1994.

[47] E. Kehayas, G. T. Kanellos, L. Stampoulidis, D. Tsiokos, N. Pleros, G. Guekos, and H. Avramopoulos, "Packet-format and network-traffic transparent optical signal processing," *Journal of Lightwave Technology*, vol. 22, no. 11, pp. 2548–2556, 2004.

[48] S. A. Hamilton, B. S. Robinson, T. E. Murphy, S. J. Savage, and E. P. Ippen, "100-Gb/s optical time-division multiplexed networks," *Journal of Lightwave Technology*, vol. 20, no. 12, pp. 2086–2100, 2002.

[49] K. L. Hall and K. A. Rauschenbach, "100-Gbit/s bitwise logic," *Optics Letters*, vol. 23, 1998.

[50] C. Schubert, S. Diez, J. Berger, R. Ludwig, U. Fieste, H. G. Weber, G. Toptchiyski, K. Peterman, and V. Krajinovic, "160-Gb/s all-optical demultiplexing using a gain-transparent ultrafast-nonlinear interferometer (GT-UNI)," *IEEE Photonic Technology Letters*, vol. 13, no. 5, pp. 475–477, 2001.

[51] B. S. Robinson, S. A. Hamilton, and E. P. Ippen, "Multiple wavelength demultiplexing using an ultrafast nonlinear interferometer," in *Conference on Lasers and Electro-Optics (CLEO '01)*, p. 528, 2001.

[52] K. Suzuki, K. Iwatsuki, S. nishi, M. Saruwatari, and T. Kitoh, "160 Gb/s sub-picosecond transform-limited pulse signal generation utilizing adiabatic soliton compression and optical time-division multiplexing," *IEEE Photonics Technology Letters*, vol. 6, no. 3, pp. 352–354, 1994.

[53] J. Emery, T. Ducellier, M. Bachmann, P. Doussiere, F. Pommereau, R. Ngo, F. Gaborit, L. Goldstein, G. Laube, and J. Barrau, "High performance $1.55\mu$m polarisation insensitive semiconductor optical amplifier based on low-tensile-strained bulk GaInAsP," *Electronic Letters*, vol. 33, pp. 1093–1084, 1997.

[54] Cisco, "Advanced PHY layer technologies for high speed over cable." white paper, 2005.

# Chapter 3

# Integrated All-Optical Switches: Semiconductor Optical Amplifiers

In Chapter 2, we discussed methods by which all-optical processing can improve data network routing and switching by replacing or supplementing electronic router functions. We demonstrated a very scalable optical header processing technique which uses only two optical logic gates for routing two incoming packet streams. In that experiment, all-optical switching was performed using a discrete optical logic gate. However, to fully take advantage of optical signal processing techniques, we must move from discrete optical switches to fully-integrated logic gates. For practical implementation and deployment in the field, integration is critical. First, it allows easy mass-production and standardization of logic gates. This means that installation can be a much simpler process than with discrete optical switches. Secondly, integration of multiple devices on a single chip can also save costly packaging and alignment steps. Finally, mass production is necessary for optical devices to supplement and compete with the mass-produced silicon electronic chips currently used today.

Integration is also needed to fully take advantage of optical signal processing techniques. As explained in the previous chapter, discrete optical logic gates can experience undesired latency as a result of requiring long lengths of fiber. Integrated gates are scaled much smaller, and as a result the latencies are negligible. Furthermore, optical logic gates often rely on interferometric effects. As a result, environmental

considerations such as temperature changes can significantly degrade performance. Integrated gates can provide higher stability because in the smaller device, a larger proportional change in length is required to significantly alter the performance. For instance, a change of 750 nm will cause a $\pi$ phase shift at 1550 nm in an interferometric device. This is a much larger percentage of a 1-mm propagation length than it is of a 20-ft propagation length. Thus, it is easier to maintain the proper interferometric relationship in a 1-mm device. It is also much easier to maintain a constant temperature across a smaller area through the use of temperature controllers.

Semiconductor optical amplifiers are highly promising devices for creating integrated optical logic gates. Semiconductor processing is a fairly mature process and thus yields can be high. Furthermore, SOAs are versatile, capable of operating as linear amplifiers, nonlinear gates, and lasers (with mirrored facets as opposed to anti-reflective coating on the facets) with only minor changes in the design. For these reasons, we have decided to investigate integration of SOA-based logic. In this chapter, we first present some basic background for semiconductor optical amplifier operation and describe the range of useful potential switch designs for integrated SOA-based switching. We then investigate the SOA Mach-Zehnder interferometer design and develop a technique to characterize and optimize switch performance and simplify switch development and operation.

## 3.1   SOA Device Physics

Semiconductor optical amplifier operation requires population inversion. The first semiconductor optical amplifiers achieved this through the use of a p-n junction. This was later refined as the heterojunction p-i-n structure, which we will discuss later on. An n-type semiconductor is doped with donor atoms which provide more electrons to the conduction band. A p-type semiconductor is doped with acceptor atoms which provide more holes (i.e., fewer electrons) in the valence band. These impurities change the chemical potential (also known as the Fermi level) of the semiconductor (Figure 3-1(a)). When the two materials are brought into contact, the electrons and holes
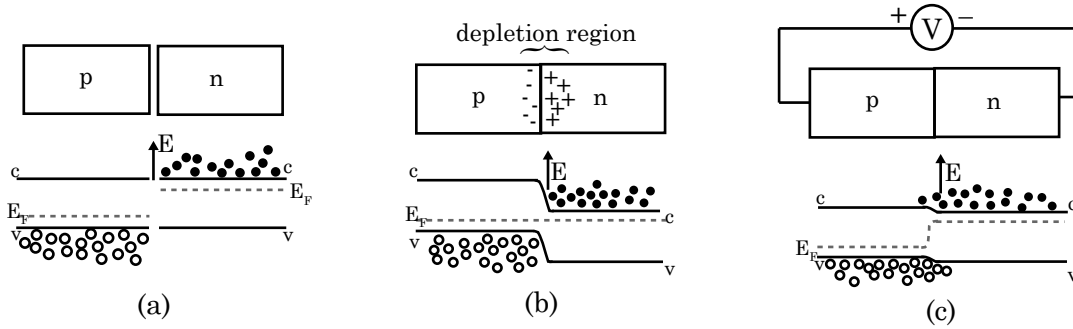
Figure 3-1: (a) Diagram of a p-type and n-type semiconductor before they are brought into contact. Holes are represented by empty circles and electrons are represented by filled circles. c and v denote the conduction and valence bands. $E_F$ represents the chemical potential of the two materials, also known as the Fermi level. E represents energy. (b) When the p and n semiconductors are brought into contact, charges diffuse across the junction forming an electric field that prevents further electron/hole movement across the junction. (c) Forward biasing the junction reduces this electric field and allows free diffusion of electrons and holes across the junction. This sets up the potential for optical amplification under the proper conditions.

diffuse across the barrier due to the difference in carrier concentrations between the semiconductors. The small region in which this occurs is known as the depletion region. The presence of additional holes in the p-type semiconductor and additional electrons in the n-type semiconductor results in an electric field across the barrier. When the force of the electric field equals the field due to carrier diffusion, the system is at equilibrium and further diffusion of electrons and holes is prevented by the potential barrier, as shown in Figure 3-1(b). When we forward bias the p-n junction, we apply a voltage to counteract this natural electric field created by the diffusion of carriers. This allows more carriers to travel across the junction, resulting in a small region where large quantities of holes and electrons can interact, allowing us to achieve optical amplification under certain conditions (Figure 3-1(c)).

The first semiconductor lasers were created by using a simple p-n junction, referred to as a homostructure. To improve the amplification and increase the interaction of light with carriers, researchers developed heterostructure devices (Figure 3-2). In these devices, a layer of intrinsic semiconductor material is inserted in between the n-type and p-type semiconductors and is referred to as the active region. The material
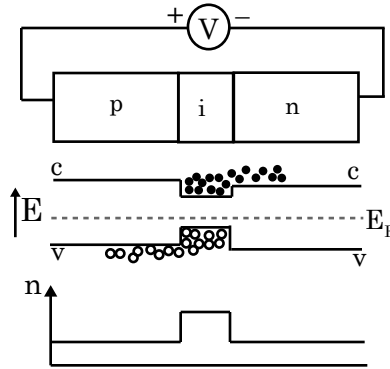
Figure 3-2: Diagram of a heterostructure semiconductor amplifier. A layer of intrinsic semiconductor with a smaller bandgap and higher index of refraction (n) is sandwiched between the p-type and n-type semiconductors to confine light and carriers.

is chosen to have a smaller bandgap and higher index than the other materials. Thus, when carriers are pushed into this region, they tend to remain there, trapped by the higher potential on either side. Furthermore, the higher index of refraction of the material means that optical modes are confined to the same region, thus increasing the interaction between photons and carriers.

Optical gain is achieved by inducing population inversion in the active region (or the depletion region in a homostructure device). To gain a basic understanding of the operation of semiconductor amplifiers, we can consider the conduction and valence bands in this region as a simple two-level system. Under normal conditions, light of the correct wavelength entering the system is absorbed, pumping electrons from the lower energy level ($E_1$) into the higher energy level ($E_2$) as shown in Figure 3-3(a). When the higher energy level is occupied, electrons can spontaneously drop to the lower energy level, emitting energy in the form of a photon. This is known as spontaneous emission (Figure 3-3(b)). An entering photon can stimulate this process, in which case the emitted photon will be of the same energy and phase as the entering photon (Figure 3-3(c)). This is known as stimulated emission and is the fundamental process behind lasing and optical amplification. By applying a strong forward bias and pumping the p-i-n junction with current, we inject carriers into the device. When the number of free carriers in the conduction band exceeds that in the lower energy
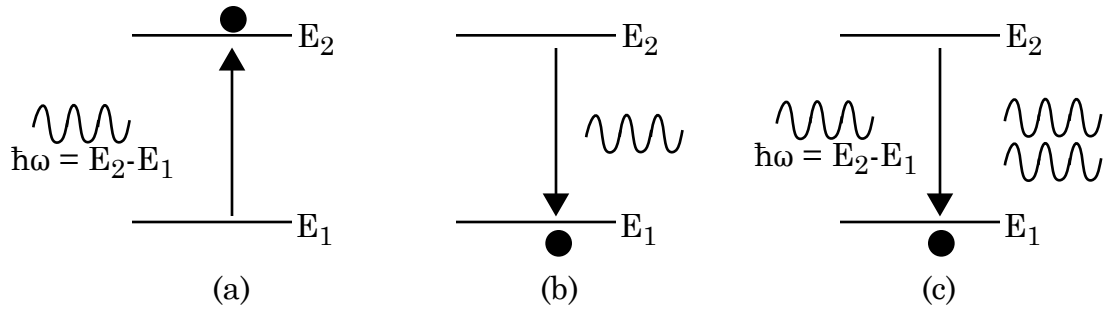
Figure 3-3: Diagram of (a) absorption, (b) spontaneous emission, and (c) stimulated emission in a two-level system.

states, we achieve population inversion. In this situation, it is more likely for an entering photon to stimulate emission than to be absorbed. This results in optical gain.

In a semiconductor optical amplifier, the conduction and valence bands are not single states but rather large bands of available energy states. This is because the atoms are arranged so closely that their allowable energy states combine into large bands of available states. As a result, SOAs can interact with more than one wavelength of light. A band diagram of a direct bandgap semiconductor is shown in Figure 3-4. This diagram shows the relationship between the energy of the conduction and valence band states and the wave vector k. Note that there is a continuum of energies. A direct-bandgap semiconductor (as shown here) indicates that the peak of the valence band is directly below the lowest point of the conduction band at $k = 0$. This means that only photons need to be present for stimulated emission. An indirect bandgap, where the peak and nulls are offset by some $k \neq 0$, means that for stimulated emission a change in momentum must be involved.

The occupation of these allowable energy states determines the effect of injected current and optical pulses on the SOA. We can calculate the occupation of the conduction and valence bands in a unit volume of the material. First, we derive the density of states per unit volume at each energy level, $\rho(E)$. We multiply the density of states by the probability of occupation at each energy $f(E)$ and obtain the
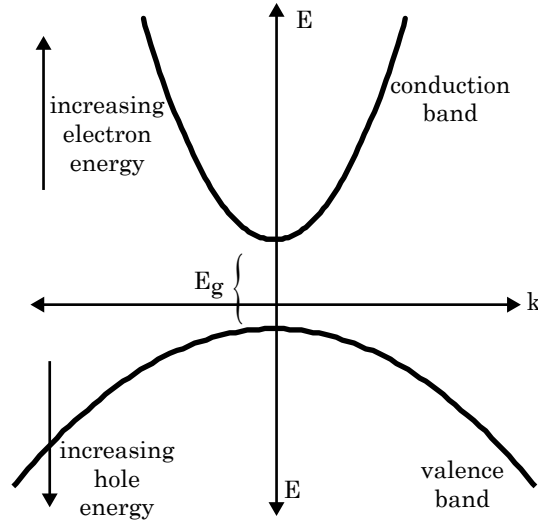
Figure 3-4: Band diagram of a direct bandgap semiconductor. E is the energy, $E_g$ is the bandgap, and k is the wave vector.

occupation of states at each energy level in a unit volume of semiconductor material.

If we assume a parabolic dispersion relationship between energy $E$ and the wave-vector $k$ (as shown in Figure 3-4), we can write the density of states as

$$\rho(E) = \frac{1}{2\pi^2}(\frac{2m}{\hbar^2})^{\frac{3}{2}} E^{\frac{1}{2}}, \tag{3.1}$$

where $m$ is the effective mass of the electron or hole and $\hbar$ is Planck's constant [1]. From this, we see that more available states exist at higher energies.

To calculate the probability of occupation of a particular energy state, we begin by noting that electrons obey the Pauli Exclusion principle and thus their distribution can be described by Fermi-Dirac statistics. For the following development, we focus on the calculation for n-type semiconductors. The calculation for p-type semiconductors is similar. For the n-type semiconductor, the probability that an electron occupies a state with an energy $E$ is thus

$$f(E) = \frac{1}{1 + \exp(\frac{E-E_F}{k_B T})}, \tag{3.2}$$

where $f(E)$ is the probability of occupation of the energy state $E$, $E_F$ is the Fermi

energy (the energy at which the probability of occupation is exactly $\frac{1}{2}$), $k_B$ is the Boltzmann constant and $T$ is the temperature of the electrons [2]. We see here that the probability of occupation decreases with increasing energy.

To calculate the distribution of carriers in the available states, we multiply the probability of occupation of a state with energy $E$ by the density of available states at the same energy:

$$N(E) = \rho(E)f(E) = \frac{1}{2\pi^2}(\frac{2m}{\hbar^2})^{\frac{3}{2}}E^{\frac{1}{2}} \times \frac{1}{1 + \exp(\frac{E-E_F}{k_B T})}. \tag{3.3}$$

Figure 3-5 shows this result graphically. As we can see from this Figure, at low energies above the bandgap, there is a high probability of occupation but fewer states exist. At higher energies, there are many more available states but the probability of occupation is much lower. Thus, the peak of the distribution lies above the bottom of the band, as shown in Figure 3-5(c).

This development shows the carrier distributions at equilibrium. Our intended application, ultrafast switching, is fast enough that our devices may not be in equilibrium during the interaction with our incoming signal pulses. To fully understand this interaction requires that we continue with this framework to calculate the carrier distributions under different carrier injection rates with ultrashort optical pulse excitation. From the calculation of carrier distributions, we can then calculate quantities such as gain, phase shift, and noise and thus understand their effects on the optical input pulses. This is a complex calculation and requires numerical simulation to achieve an accurate result. More information on this calculation can be found in [3]. We began with this description in order to provide an intuitive basis for understanding basic SOA operation. However, we can obtain a reasonable approximation of SOA behavior for short optical pulse propagation using a simpler phenomenological model valid under most conditions in photonic switching. The following development was first published in [4].
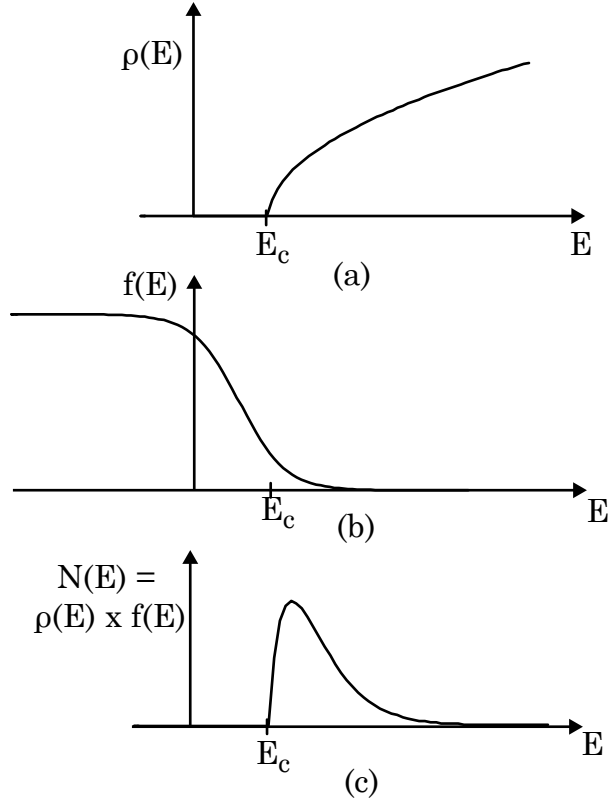
Figure 3-5: Calculation of the distribution of carriers in semiconductors. (a) shows the density of states with respect to energy. (b) shows the probability of occupation for a state of energy $E$. (c) shows the carrier occupation distribution over the available energies.

### 3.1.1    A Phenomenological Model of the SOA

To describe optical pulse propagation in SOAs, we need to first describe the effect of an optical field within the SOA upon the carrier distributions.  Also, we need to describe how the carrier distributions affect the propagation of that optical field. This will result in a set of coupled-wave equations for the SOA gain, phase, and the optical power within the device. In this treatment, we have assumed that the optical pulse duration is much longer than the intraband relaxation time that governs the semiconductor dynamics. This relaxation time is typically $< 0.1$ ps. Thus, our pulse propagation model is valid for pulses that are 1 ps or longer. This is well justified for the pulse durations commonly used for ultrafast switching.  A picosecond pulse can theoretically be used at data rates approaching 1 Tb/s. To use shorter pulses may

require us to use a more exact theory of semiconductor dynamics.

Before we proceed further, we first define the transverse and longitudinal directions of this device and place it in the reference of a coordinate system as shown in Figure 3-6. Optical pulses travel longitudinally along the device in the $z$ direction. Current is injected into the device along the $y$ axis, uniformly across the $z$ dimension.
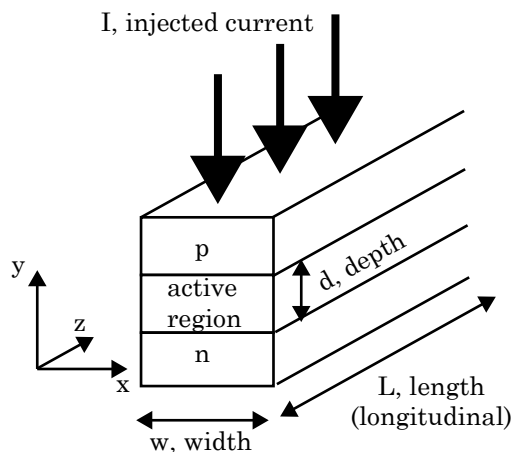


Figure 3-6: Coordinate system and description of the SOA being modeled. The longitudinal direction is in the direction of propagation of the optical pulses, $z$. The transverse plane $(x,y)$ is perpendicular to $z$. Current is uniformly injected from above.

The rate equation which describes the evolution of carriers in an SOA in response to current injection and optical power is given by

$$\frac{\partial N}{\partial t} = D\nabla^2 N + \frac{I}{qV} - R(N), \tag{3.4}$$

where $N$ is the carrier density, $D\nabla^2 N$ describes the carrier diffusion rate, $I$ is the injected current, $q$ is the electron charge, $V$ is the volume of the active region, and $R(N)$ describes the recombination mechanisms of the semiconductor which create or destroy free carriers. Note that $N$ is a function of $(x, y, z, t)$. The recombination mechanisms can be expressed as

$$R(N) = A_{nr}N + BN^2 + CN^3 + R_{st}N_{ph}, \tag{3.5}$$

where $A_{nr}N$ describes the non-radiative recombination processes such as recombina-

83

tion at defects and surface recombination, $BN^2$ describes spontaneous emission processes, $CN^3$ describes the non-radiative Auger recombination processes, and $R_{st}N_{ph}$ describes stimulated emission in the presence of a photon ($N_{ph}$). Non-radiative recombination occurs when electrons and holes recombine and release energy in the form of heat, or vibrations, in the semiconductor crystal lattice. Spontaneous emission occurs when electrons in the conduction band recombine with holes in the valence band, emitting photons. Auger recombination occurs when the energy released from the recombination of an electron and hole stimulates a second electron or hole to a higher energy state. This electron or hole then relaxes, releasing energy into the semiconductor lattice. Yamada [5] derives this rate equation from first principles.

We can write this equation in terms of the incident electric field $E$:

$$\frac{\partial N}{\partial t} = D\nabla^2 N + \frac{I}{qV} - \frac{N}{\tau_c} - \frac{a(N - N_o)}{\hbar\omega}|E|^2, \tag{3.6}$$

where we have collapsed all recombination terms except the stimulated recombination term into a single term $N/\tau_c$, where $\tau_c$ describes a combined carrier recombination rate. We have modeled stimulated emission (or absorption) as a reduction in the carrier density linearly proportional to the number of entering photons with the constant of proportionality $a$. $N_o$ is the carrier density at transparency, where the stimulated emission is exactly equal to absorption. If the carrier density is below $N_o$, the entering photons add carriers to the conduction band; if the carrier density is above $N_o$, the entering photons remove carriers from the conduction band. For simplicity, we assume the input electric field is linearly polarized and only consider a single polarization within the amplifier. We also assume that the electric field is quasi-monochromatic, consisting of a slowly-varying envelope combined with a monochromatic plane wave. The electric field can thus be written as,

$$E(x, y, z, t) = \hat{x}\frac{1}{2}F(x, y)A(z, t)e^{i(k_o z - \omega_o t)} + \text{c.c.}, \tag{3.7}$$

where $F(x, y)$ is the transverse mode profile of the optical pulse and $A(z, t)$ is the slowly-varying envelope.

We can simplify the carrier rate equation by assuming that $L \gg w, d$ and much longer than the carrier diffusion length. Thus, the carrier density is roughly uniform in the transverse dimension. We can neglect carrier diffusion and average our equation over the transverse active region dimensions to obtain

$$\frac{\partial N}{\partial t} = \frac{I}{qV} - \frac{N}{\tau_c} - \frac{a(N - N_o)}{\hbar\omega}\Gamma|A(z, t)|^2, \tag{3.8}$$

where $N$ is now only a function of $(z, t)$. We define $\Gamma$, the confinement factor, as the fraction of the power in the active region of the SOA, averaged over the transverse dimensions.

$$\Gamma = \frac{\displaystyle\int_0^w \int_0^d |F(x, y)|^2 dxdy}{\displaystyle\int_{-\infty}^\infty \int_{-\infty}^\infty |F(x, y)|^2 dxdy}. \tag{3.9}$$

We are interested in calculating the effect of light on the semiconductor material in terms of the gain of the material, rather than the carrier density. We define the gain as linearly proportional to the increased carrier density

$$g(N) = \Gamma a(N - N_o) \tag{3.10}$$

and write the carrier rate equation in terms of the gain.

$$\frac{\partial g}{\partial t} = \Gamma a \frac{I}{qV} - \frac{g + \Gamma a N_o}{\tau_c} - \frac{\Gamma a g}{\hbar\omega}|A(z, t)|^2. \tag{3.11}$$

We can simplify this by defining some constants. We define $g_o$, the small signal gain, $E_{sat}$, the saturation energy of the amplifier, $\sigma$, the mode cross section, and $I_o$, the

current required for transparency.

$$g_o = \Gamma a N_o(\frac{I}{I_o} - 1) \tag{3.12}$$

$$E_{sat} = \frac{\hbar \omega \sigma}{a} \tag{3.13}$$

$$\sigma = \frac{wd}{\Gamma} \tag{3.14}$$

$$I_o = \frac{qV N_o}{\tau_c}. \tag{3.15}$$

We also normalize the amplitude $A(z,t)$ such that $|\overline{A(z,t)}|^2$ is the power, by dividing by the mode cross-sectional area $\sigma$. Thus, we obtain:

$$\frac{\partial g(N)}{\partial t} = \frac{g_o - g(N)}{\tau_c} - \frac{g(N)|\overline{A(z,t)}|^2}{E_{sat}}. \tag{3.16}$$

Now that we have arrived at a description of how optical power affects the semiconductor carrier densities, we turn our attention to describing how these changes in the semiconductor material affect the incident optical pulse. We first begin with a general framework of how material nonlinearities affect optical power and then specify the effect of semiconductor nonlinearities on an input optical pulse. We begin by considering the general wave equation,

$$\nabla^2 E(\mathbf{r},t) - \frac{1}{c^2}\frac{\partial^2 E(\mathbf{r},t)}{\partial t^2} = -\mu_o\frac{\partial^2 P(\mathbf{r},t)}{\partial t^2}, \tag{3.17}$$

where we have defined an electric polarization $P$, which describes the effect of dipoles in a material upon an incident electric field. The relationship between $P$ and $E$ can be written as:

$$P(\mathbf{r},t) = P^{(1)}(\mathbf{r},t) + P^{(2)}(\mathbf{r},t) + P^{(3)}(\mathbf{r},t) + \ldots \tag{3.18}$$

$$P(\mathbf{r}, t) = \epsilon_o \int R^{(1)}(\mathbf{r}, \tau) E(\mathbf{r}, t - \tau) d\tau \qquad (3.19)$$

$$+ \epsilon_o \int \int R^{(2)}(\mathbf{r}, \tau_1, \tau_2) E(\mathbf{r}, t - \tau_1) E(\mathbf{r}, t - \tau_2) d\tau_1 d\tau_2$$

$$+ \epsilon_o \int \int \int R^{(3)}(\mathbf{r}, \tau_1, \tau_2, \tau_3) E(\mathbf{r}, t - \tau_1) E(\mathbf{r}, t - \tau_2) E(\mathbf{r}, t - \tau_3) d\tau_1 d\tau_2 d\tau_3$$

$$+ \dots.$$

Here, we have separated out the multiple orders of the electric polarization, where the $n$th order refers to the number of distinct electric fields which are involved in the process. It is simpler to write this in the frequency domain, which also clarifies the nonlinear frequency mixing that occurs. We can write the electric polarizations as a sum of frequencies, where for each frequency $\omega$ we have

$$P(\mathbf{r}, \omega) = \epsilon_o \chi^{(1)}(\omega : \omega_1) \cdot E(\mathbf{r}, \omega_1) \qquad (3.20)$$

$$+ \epsilon_o \chi^{(2)}(\omega : \omega_1, \omega_2) : E(\mathbf{r}, \omega_1) E(\mathbf{r}, \omega_2)$$

$$+ \epsilon_o \chi^{(3)}(\omega : \omega_1, \omega_2, \omega_3) \vdots E(\mathbf{r}, \omega_1) E(\mathbf{r}, \omega_2) E(\mathbf{r}, \omega_3)$$

$$+ \dots, \qquad (3.21)$$

Here, we can see that the $n$th order of the polarization refers to the number of frequencies involved in the interaction. The first order polarization describes a linear interaction, meaning that no new frequencies are created or destroyed. This is often included in the dielectric constant and forms the basis for the different index of refraction in different materials. The higher order polarizations involve nonlinear interactions between the different frequencies of the electric field. For example, the third-order nonlinearity can create a nonlinear polarization at the frequency $\omega$ through all possible sums and differences of three arbitrary frequencies $\omega_1, \omega_2, \omega_3$, as long as they sum to $\omega$. Each distinct sum of the possible frequencies results in a different value of the nonlinear coefficient $\chi^{(3)}$. Higher order nonlinearities tend to be weaker.

In the semiconductor optical amplifier, the second-order nonlinearity does not exist in the co-polarized case [6]. Thus, we shall only consider the first and third non-linearities. Furthermore, for now, we shall only consider the degenerate $\chi^{(3)}(\omega)$, where

we only consider a single wavelength involved in the introduction of semiconductor nonlinearities. In this development, we only consider a single linear polarization. A more complete treatment can be found in [7, 8]. The degeneracies we have included result in a coefficient of $\frac{3}{4}$ for our electric polarization $P(\mathbf{r}, \omega)$:

$$P(\mathbf{r}, \omega) = \epsilon_o \frac{3}{4} \chi^{(3)}(\omega = \omega - \omega + \omega) \vdots |E(\mathbf{r}, \omega)|^2 E(\mathbf{r}, \omega). \qquad (3.22)$$

If we combine these equations with the wave equation in the frequency domain, we obtain

$$\nabla^2 E(\mathbf{r}, \omega) - \omega^2 \frac{1}{c^2} \left\{ 1 + \chi^{(1)}(\omega) + \frac{3}{4} \chi^{(3)}(\omega) |E(\mathbf{r}, \omega)|^2 \right\} E(\mathbf{r}, \omega) = 0. \qquad (3.23)$$

Typically, the index of refraction of a material in the absence of nonlinear polarization is defined as

$$n_o = \sqrt{1 + \text{Re}\{\chi^{(1)}(\omega)\}}. \qquad (3.24)$$

The imaginary component of the linear electric polarization gives us the background loss in the material. We can simplify the wave equation to

$$\nabla^2 E(\mathbf{r}, \omega) - \omega^2 \frac{1}{c^2} \left\{ n_o^2 + i\text{Im}\{\chi^{(1)}\} + \frac{3}{4} \chi^{(3)}(\omega) |E(\mathbf{r}, \omega)|^2 \right\} E(\mathbf{r}, \omega) = 0. \qquad (3.25)$$

We can see from this that the real part of the nonlinear polarization leads to an intensity dependence in the index of refraction. This is also known as self-phase modulation and describes the effect of nonlinearities induced in the semiconductor amplifier by an incoming pulse. These nonlinearities then affect the pulse itself, as we shall see in the following development. The imaginary part of the nonlinear polarization leads similarly to an intensity-dependent loss (or gain) in the material.

We are interested in modeling this intensity dependence in terms of the carrier density, so we can relate the effect of the semiconductor dynamics to the electric field. We can fold these effects into a complex dielectric constant, $\epsilon$, where

$$\epsilon = n_o^2 + \chi(N), \qquad (3.26)$$

88

where we assume that $\chi(N)$ changes slowly with respect to frequency. This essentially indicates that we are neglecting the frequency effects in self-phase modulation in the SOA and assuming all gain or phase change effects on an incoming optical pulse are constant across the frequencies in the pulse. We can then write our wave equation as

$$\nabla^2 E - \frac{\epsilon}{c^2}\frac{\partial^2 E}{\partial t^2} = 0. \tag{3.27}$$

We model $\chi(N)$ as linearly dependent upon the carrier density. We also separate out the imaginary and real components of $\chi(N)$. The real part of $\chi(N)$ describes the change in the index of refraction of the SOA due to carrier density changes, while the imaginary part describes the loss or gain in the SOA due to the change in the carrier densities.

$$\chi(N) = -\frac{\bar{n}c}{\omega_o}[\alpha a(N - N_o) + ia(N - N_o)], \tag{3.28}$$

where $\bar{n}$ is the effective mode index, and $\alpha$ is the linewidth enhancement factor. From this, one can see that our model approximates the gain and phase as linear with respect to the carrier density, and thus proportional to each other by the linewidth enhancement factor, $\alpha$ [4].

We need to substitute our description of the electric field from Equation 3.7 into the wave propagation equation (Equation 3.27) to describe the propagation of our electric field in the SOA. Neglecting the second derivatives of $A(z,t)$ with respect to $z$ and $t$ due to the slowly-varying envelope assumption, and integrating over the transverse dimensions, we obtain the following coupled equations:

$$\frac{\partial^2 F(x,y)}{\partial x^2} + \frac{\partial^2 F(x,y)}{\partial y^2} + (n_b^2 - \bar{n}^2)\frac{\omega_o^2}{c^2}F(x,y) = 0 \tag{3.29}$$

$$\frac{\partial A(z,t)}{\partial z} + \frac{1}{v_g}\frac{\partial A(z,t)}{\partial t} = \frac{i\omega_o\Gamma}{2\bar{n}c}\chi(N)A(z,t) - \frac{1}{2}\alpha_{int}A(z,t), \tag{3.30}$$

where $v_g$ is the group velocity $\frac{c}{n_g}$, $n_g$ is the effective group index $\bar{n} + \omega_o\frac{\partial \bar{n}}{\partial \omega}$, and $\alpha_{int}$ is the internal loss of the electric field when there are no carrier effects ($\chi(N) = 0$).

We further simplify this equation by transforming into a moving reference frame:

$$\tau = t - \frac{1}{v_g}z \tag{3.31}$$

and separating out the amplitude and phase of the optical pulse:

$$\overline{A(z,\tau)} = \sqrt{P(z,\tau)}e^{i\phi(z,\tau)}, \tag{3.32}$$

where $P(z,\tau)$ is the power of the optical pulse. By separating out the imaginary and real components of Equation 3.30, we obtain:

$$\frac{\partial P(z,\tau)}{\partial z} = (g(z,\tau) - \alpha_{int})P(z,\tau) \tag{3.33}$$

$$\frac{\partial \phi(z,\tau)}{\partial z} = -\frac{1}{2}\alpha g(z,\tau). \tag{3.34}$$

If we apply Equations 3.31 and 3.32 to our previously-derived carrier rate equation (Equation 3.16), we obtain the third equation:

$$\frac{\partial g(z,\tau)}{\partial \tau} = \frac{g_o - g(z,\tau)}{\tau_c} - \frac{g(z,\tau)P(z,\tau)}{E_{sat}}. \tag{3.35}$$

Together, Equations 3.33 - 3.35 describe pulse propagation in semiconductor optical amplifiers. Equations 3.33 and 3.34 describe the effect of gain in the SOA upon the propagating optical pulse amplitude and phase while Equation 3.35 describes the effect of the propagating pulse on the SOA gain. The presence of the product $g(z,\tau)P(z,\tau)$ makes solving this set of differential equations difficult.

We can make some additional simplifications to solve this equation easily without losing significant accuracy. First, we assume $\alpha_{int} \ll g(z,\tau)$ and thus we set $\alpha_{int} = 0$ in Equation 3.33. We also define an integrated gain $h(\tau)$ and thus ignore gain variations in the longitudinal direction:

$$h(\tau) = \int_0^L g(z,\tau)dz. \tag{3.36}$$

We can integrate Equations 3.33 and 3.34 to get the output power and phase in terms of the integrated gain:

$$P_{out}(\tau) = P_{in}(\tau)e^{h(\tau)} \tag{3.37}$$

$$\phi_{out}(\tau) = \phi_{in} - \frac{1}{2}\alpha h(\tau). \tag{3.38}$$

We derive the evolution of the integrated gain across $\tau$ by integrating Equation 3.35 over $z$ and substitute Equation 3.33 for the product $g(z,\tau)P(z,\tau)$:

$$\frac{dh(\tau)}{dt} = \frac{g_o L - h(\tau)}{\tau_c} - \frac{P_{in}(\tau)}{E_{sat}}[e^{h(\tau)} - 1]. \tag{3.39}$$

This allows us to solve these equations numerically using the 4th-order Runge-Kutta method. For our simulations, we choose typical SOA values: $E_{sat}$=1 pJ, $\tau_c = 80$ ps, $G_o = 10^3$ (30 dB unsaturated integrated gain), $L = 1$ mm, and $\alpha = 5$. We use 2-ps input pulses with a repetition period of 25 ps, forming a 40-GHz pulse train. We picked an input pulse energy of 5 fJ to illustrate the cumulative effects of gain recovery in the SOA. We plot the evolution of the integrated gain over time normalized to the incident pulse width and the effect of the change in gain on the propagating pulses in Figure 3-7.

The first thing to note is that as the pulse propagates through the amplifier, the amplifier gain saturates and then recovers with the carrier recovery time $\tau_c$. As can be seen by the plots of the integrated gain $h(\tau)$ and the output pulses, the effect of gain saturation causes pulse patterning, in which the amplitude of the output pulse is greatly affected by the exact sequence of bits in the sequence. This significantly affects the performance of ultrafast optical switching. Methods for overcoming this pulse patterning will be discussed in Section 3.2.

The next thing to note is that this gain saturation effect also distorts the pulses as they exit the switch. Figure 3-8 illustrates these effects. Recall that $\tau$ is defined in terms of our moving reference frame, such that smaller values of $\tau$ correspond to higher values of $z$. We thus refer to the negative values of $\tau$ in our figures as the
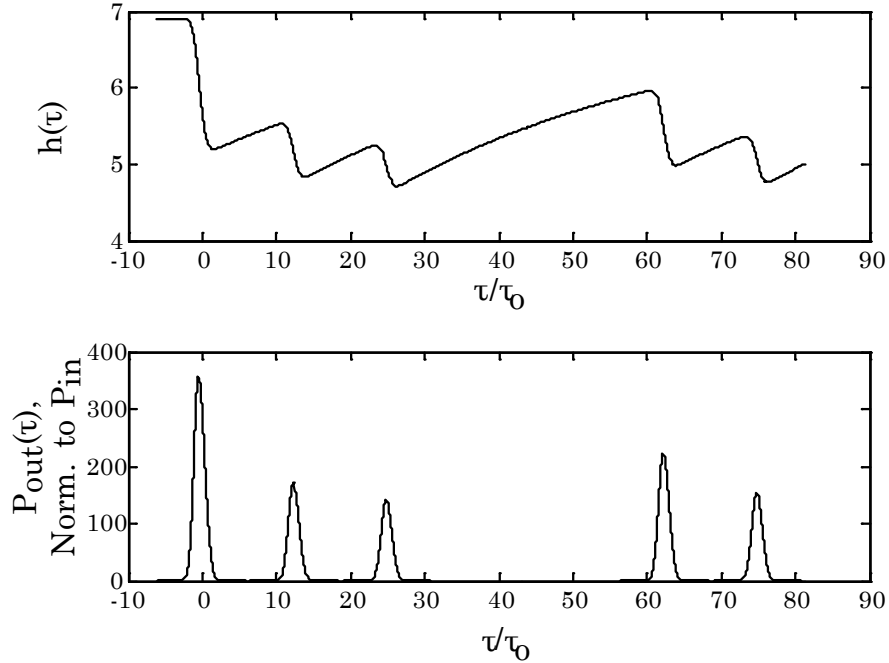
Figure 3-7: Simulation of gain saturation in an SOA using the Runge Kutta method. The upper plot shows the integrated gain in the SOA as a function of $\tau$ over $\tau$ normalized to the pulse width $\tau_o$. The lower plot shows the output optical pulses normalized to the input pulse power.

"leading" edge of the pulse and positive values of $\tau$ as the "lagging" part of the pulse. First of all, since the pulse power changes the gain of the SOA, the leading edge of the pulse will see higher gain than the trailing edge of the pulse. This has the effect of steepening the leading edge of the pulse while broadening the trailing edge. Secondly, since the added phase of the outgoing pulse is proportional to the negative of the integrated gain, the saturation of the integrated gain provides a change in the phase across the pulse. This adds a chirp to the output pulse. Since the aggregate phase shift is negative, this also causes a frequency shift that we can observe in the spectral domain.

Now that we have a basic understanding of how a semiconductor optical amplifier operates, we can turn to investigation of the operation of ultrafast optical switches using SOAs.
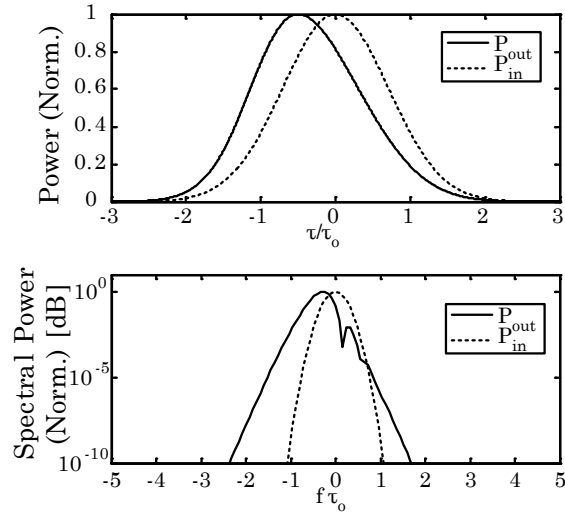
Figure 3-8: Effect of gain saturation on optical pulse shape. Pulse amplitude is shown in the upper plot and the pulse spectrum is shown in the lower plot, normalized to pulse width and power.

## 3.2 Optical Switching using SOAs

In the previous section, we described how an incoming optical pulse both changes the SOA characteristics and is changed by them through self-phase modulation and gain saturation. These effects can be used to change the characteristics of a second incoming optical pulse. This allows us to use the SOA as a key element in optical switching.

Previously, we described self-phase modulation, which arises from the degenerate $\chi^{(3)}$ nonlinearity. Here, we expand the description to include cross-phase modulation (XPM) and four-wave mixing (FWM). We also describe the use of cross-gain modulation (XGM), which uses the gain saturation effect for switching. Our main interest is in implementing integrated all-optical logic, meaning that we wish to implement Boolean logic functions such as AND, NAND, NOR, etc. Most of the switch configurations which perform optical logic can also involve a wavelength conversion. However, as we showed in Chapter 2, there may be situations where only a wavelength conversion is required. In those cases, a simpler configuration can be used, thus potentially reducing cost. In the following sections, we describe a few of the

most promising switch configurations which use XGM, XPM, or FWM in an SOA.

## 3.2.1   Cross-gain Modulation (XGM)

Cross-gain modulation is the simplest to understand: the saturation of the SOA gain as a result of an incident pulse (referred to as the control pulse) can be used to control the amplification seen by a second pulse (referred to as the signal pulse). A diagram of cross-gain modulation is shown in Figure 3-9. When the control pulse (shown in
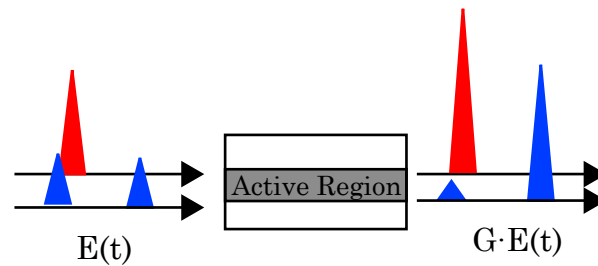


Figure 3-9: Diagram of gain-switching in an SOA with cross-gain modulation (XGM). The control pulse is shown in red and the signal pulse is shown in blue. The logical output is $S \cdot \bar{C}$.

red) is not present, the SOA amplifies the weak signal pulse with very little change in the gain response. When the strong control pulse enters the SOA, the gain saturates. Thus, when a second, weaker signal pulse arrives just afterwards, it sees a much lower gain. Note that a bandpass filter is required at the output of this switch to strip off the control pulse. This results in the logic of $S \cdot \bar{C}$, where $S$ refers to the signal and $C$ refers to the control. This is the simplest method of all-optical switching in an SOA. However, one major disadvantage results from the carrier recovery time of the SOA. When a strong control pulse saturates the SOA, the carrier recovery time determines the length of time needed before the SOA recovers to its previous state. We can observe this by the measurement of the switching window in an SOA.

The switching window measures the change in the transmission of the switch caused by the control input pulse. For optimal switching, the signal input should be timed to arrive at the peak of the switching window, thus experiencing the maximum transmission change due to the control pulse. In order to measure the switching

window, we look at the transmission of the switch before, during, and after the incident control pulse. The schematic for this measurement is shown in Figure 3-10. We use clock pulses (without a bit pattern) as both the signal and control inputs. We
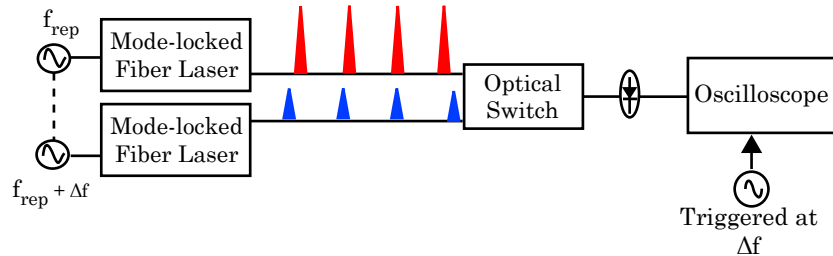


Figure 3-10: Schematic for the switching-window measurement on an optical switch. The signal and control pulse trains are created by two synchronized mode-locked fiber lasers. The two lasers are driven at slightly offset repetition rates, causing the two pulse trains to walk through each other. The average power at the output of the SOA is detected and plotted as a function of the pulse train offsets. This results in a plot of the transmission of the SOA switch vs. control-signal pulse delay.

offset the repetition rate of the signal pulse train from the clock pulse train. Thus, the two pulse trains will slowly scan through each other at a rate equal to the offset. The signal input is set to be much lower power than the control input, so that it has a negligible effect on the SOA. Thus, the switch is affected mainly by the control pulse. At the output, we measure the average power of the signal output. This average power varies depending on the signal pulse's position in relation to the control pulse. If the control pulse arrives behind the signal pulse, the signal pulse sees very little effect from the control pulse. If the control pulse arrives slightly before the signal pulse, the signal pulse sees the maximum change in transmission of the switch due to the control pulse. We measure the average power of the signal pulse as a function of the control-signal pulse delay, which gives us the switching window. The width of this switching window limits the speed of the optical switch. If a second signal pulse arrives before the end of the first switching window, the effects of that first control pulse will interfere and degrade switch performance. Figure 3-11 shows the switching window for a single SOA at several different pump currents. This figure shows that the speed of a gain-switched SOA is entirely determined by the carrier recovery time.

95
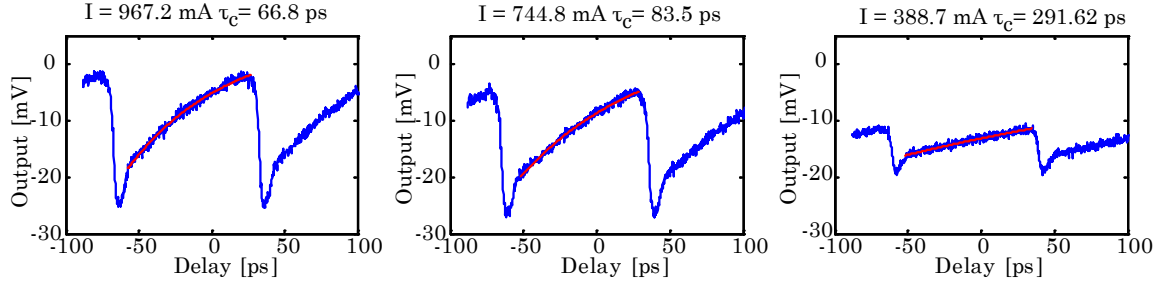
Figure 3-11: Switching windows at different pump currents for a gain-switched SOA.

To obtain a fit of the carrier recovery time, we use an exponential recovery model [9]. Thus, even for very high currents, the speed of gain-switching is limited to about 10 Gb/s.

To improve the speed of SOA switches, we can place our SOAs in an interferometric configuration. These type of switches rely on cross-phase modulation (XPM).

### 3.2.2   Cross-Phase Modulation (XPM)

Cross-phase modulation is a result of using a control pulse to change the index of refraction of the SOA and affect the phase of the signal pulse. This can be described in terms of the nonlinear polarization as

$$P(\mathbf{r}, \omega_2) = \epsilon_o \chi^{(3)}(\omega_2 = \omega_1 - \omega_1 + \omega_2) \vdots |E_{control}(\mathbf{r}, \omega_1)|^2 E_{signal}(\mathbf{r}, \omega_2). \qquad (3.40)$$

To utilize cross-phase modulation (XPM) for optical switching, the phase change must be converted into an intensity change to indicate a change in the bit values (for on-off-keyed data). Thus, an interferometric technique must be used. Furthermore, an interferometric configuration can overcome this carrier recovery speed limitation. A Mach-Zehnder interferometer is shown in Figure 3-12(a). This SOA-MZI gate has been demonstrated to be capable of ultrafast switching: demultiplexing at speeds up to 336 Gb/s [10, 11] as well as bit-wise switching for 80 Gb/s AND [12] and 40 Gb/s XOR [13] operation. Though we introduce this switch in the context of cross-phase modulation, we should keep in mind that all the nonlinearities are present in the SOA and thus effect the result. So, self-phase modulation and cross-gain modulation can
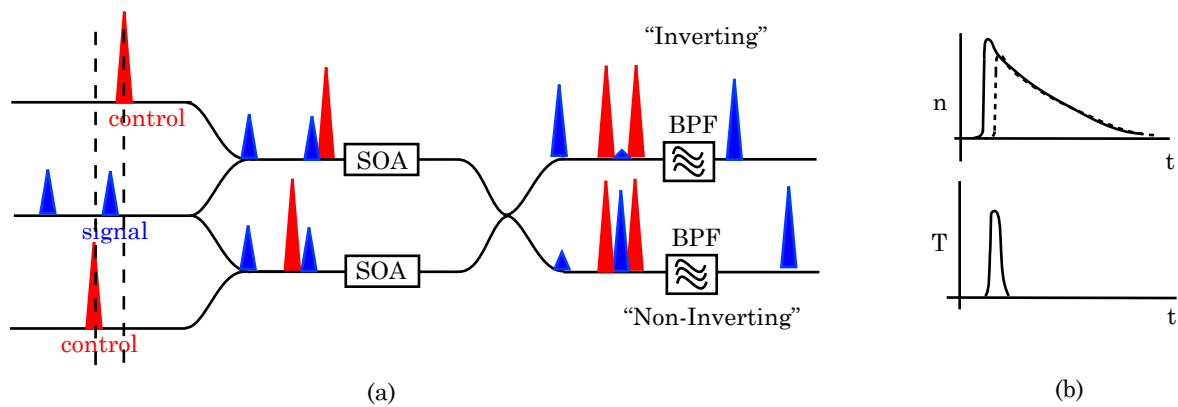
Figure 3-12: (a) An SOA-based Mach-Zehnder interferometer switch. Both inverting and non-inverting operation are shown. The control pulses are shown in red and the signal pulses are shown in blue. The logical output for the inverting operation is $S \cdot \bar{C}$ and the logical output for the non-inverting operation is $S \cdot C$. Only one control pulse is necessary for switching; the second control pulse is used to speed up switch operation, as shown in (b). The upper plot shows the change in the index of refraction over time in the upper SOA (solid line) and the lower SOA (dashed line). The lower plot shows the transmission over time in the non-inverting case as a result.

both add to or detract from the switch.

The switch operation can be understood as follows: the presence of the control pulse changes the phase of the signal pulse by exactly $\pi$. This interferometer can be biased in two ways, which we refer to as "Inverting" and "Non-inverting". When biased in inverting mode, in the absence of the control pulse, the signal pulse is split into two halves and recombines constructively at the output. When the control pulse appears in one arm, it causes a $\pi$ phase shift in one of the signal pulses. Thus, when the signal pulse recombines, it does so destructively. This creates the logic $S \cdot \bar{C}$. When biased in non-inverting mode, the opposite occurs: the presence of the control pulse causes the signal pulse to recombine constructively. This gives us an AND gate: $S \cdot C$. To ameliorate the carrier recovery time limitation, we inject a second control pulse into the bottom arm of the interferometer, delayed from the first control pulse to arrive after the signal pulse. This shifts the phase induced in the bottom arm of the interferometer to again match that of the first arm, thus overcoming the recovery time limitation as illustrated in Figure 3-12(b). Though the speed of this gate can be

97

significantly increased by this mechanism, ultimately, the carrier recovery time still limits switch operation. Though the two arms of the interferometer are now matched again, the two SOAs are still in saturation until the SOA recovers, thus reducing the effect of the control optical pulse. If a sequence of "one" bits at a high repetition rate is sent into the switch, each pulse will drive the switch deeper and deeper into saturation. By the last pulse, the SOAs will be deep in saturation and the same control input power will change the phase by a much smaller amount than for the first pulse. This introduces pulse patterning and degrades the output.

As before, we can plot the switching window of this switch for both the inverting and non-inverting configurations (Figure 3-13). The adjustment of the bias of the
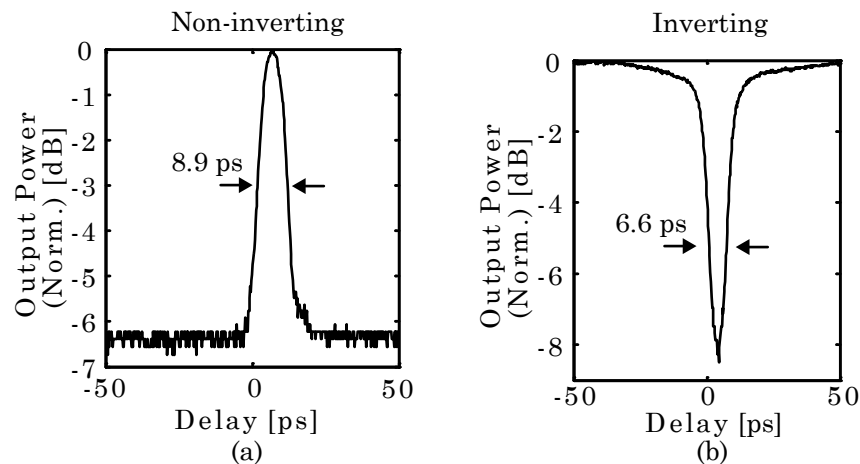


Figure 3-13: Inverting and non-inverting switching windows for the SOA-MZI gate.

interferometer was accomplished by changing the current bias of the two SOAs. Since this affects the carrier populations of the SOAs, it determines the gain and phase induced by each SOA in the absence of any optical pulse. Both switching windows are below 10 ps, indicating that this switch configuration can be used up to 100 Gb/s without significant degradation. We also note that the inverting switching window is both shorter in duration and larger in magnitude. This is because cross-gain modulation aids the main cross-phase modulation effect in inverting mode. Thus, inverting operation of the SOA-MZI is more effective than non-inverting operation.

The SOA-MZI can also be used in counter-propagating operation to eliminate

the need for a filter, as shown in Figure 3-14. Integration of filters adds complexity
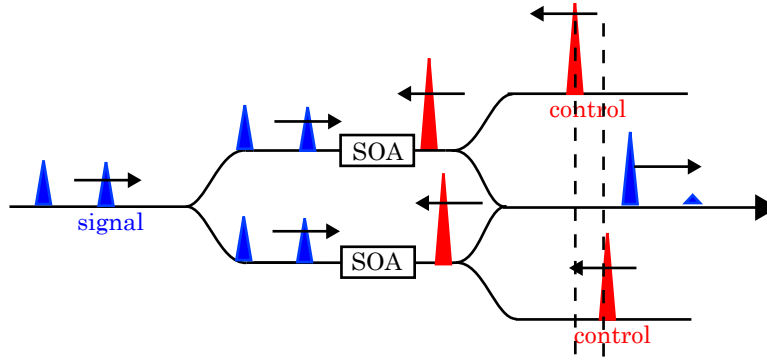


Figure 3-14: Counter-propagating operation of the SOA MZI switch. The signal pulses (in blue) are propagating to the right and the control pulses (in red) propagate to the left. This avoids the need for a filter but adds limitations to the speed of the switch.

and counter-propagating operation of the SOA-MZI can avoid this need. However, in this mode, longitudinal effects of the SOA limit the speed of the SOA and add additional constraints on SOA operation, since the length of the SOA is also related to the amount of phase shift or gain saturation resulting from the incident pulse. We illustrate this in Figure 3-15(a) and (b). In the Figure 3-15(a), we see that the effect of a control pulse on the signal pulse exists for twice the transit time ($T_o$) of the SOA. If the signal pulse arrives just as the control pulse is exiting, it still sees the effect of the control pulse at the left edge of the SOA. If the signal pulse arrives as the control pulse is $T_o$ away from the SOA, the control pulse will be entering the right edge of the SOA just as the signal pulse is leaving and thus exerting an effect on the signal pulse. In contrast, for a co-propagating configuration, the control and signal pulses propagate through the SOA together and thus the effect of the control pulse is very close to just the length of the control pulse. In Figure 3-15(b), we see that the effect of the control pulse on the signal pulse is necessarily less than that of a co-propagating pulse for the same pulse energy. This is because in the co-propagating configuration, the signal and control pulses cross and the signal pulse may propagate through some length of the SOA without encountering the effect of the control pulse. Thus, the effective length of the SOA is shortened. In the co-propagating configuration, the

99
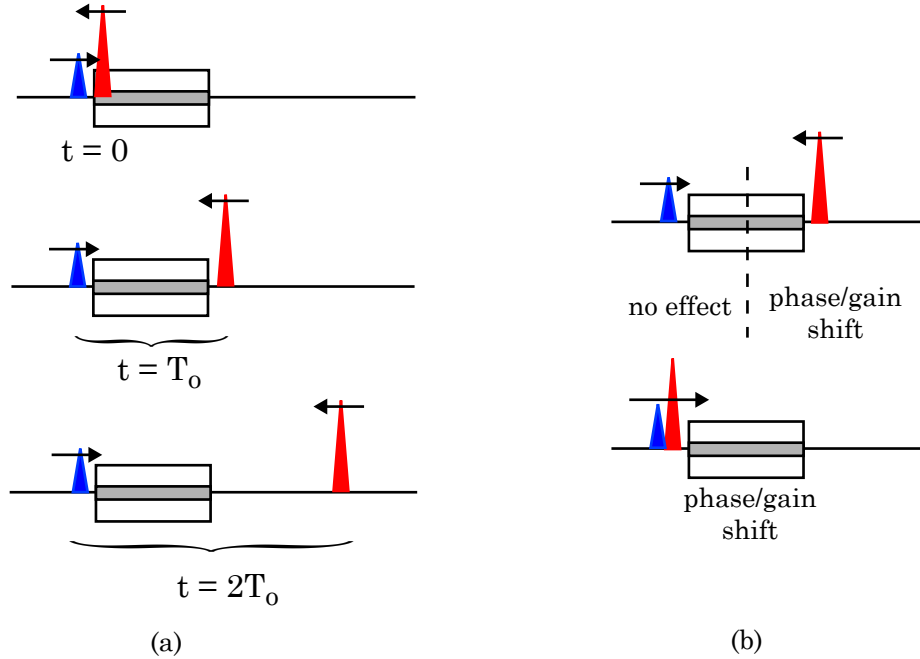
Figure 3-15: Illustration of the limits of the counter-propagating configuration. (a) shows that the control pulse will affect the signal pulse over a time $2T_o$, where $T_o$ is the transit time of the SOA. (b) shows that the magnitude of the control pulse effect on the SOA will be smaller when compared with the co-propagating configuration because less of the SOA length will be in use.

control and signal pulses are constantly affecting each other through the entire length of the SOA.

We also wish to know how accurately each these switches operate. This is determined by using a bit-error rate (BER) measurement. A general measurement setup is shown in Figure 3-16. A pseudo-random bit sequence (PRBS), generated by the pattern generator, is modulated onto a sequence of pulses, and passed into the logic gate. A second sequence is likewise generated and sent into the second port of the logic gate. The switch performs a Boolean operation on the inputs. At the output, the resulting pulse sequence is amplified and then detected by a photodetector. The output of the photodetector is filtered and then sent into an error detector, which compares the bit sequence with the expected bit sequence and measures the errors per bit. This arrangement is known as a pre-amplified receiver. Typically, a $10^{-9}$ BER is considered "error-free" for uncoded optical signaling. This threshold is set
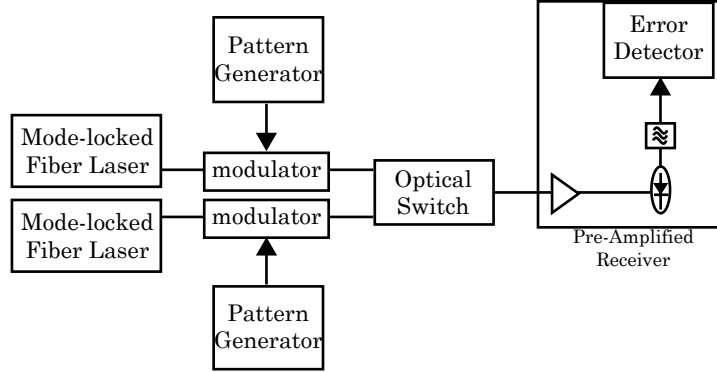
Figure 3-16: Schematic for bit-error rate measurement of an optical switch. The input pulse trains are created by pattern generators modulating a pseudo-random bit sequence on a sequence of optical pulses. The output of the optical switch is detected by our pre-amplified receiver, which measures the errors per bit in the resulting pattern.

by the network protocols which rely on this physical layer for accurate transmission. We measure the bit-error rate over a range of receiver input average optical powers and compare the resulting curve to the measurement without the device under test (known as the "back-to-back" measurement). This allows us to see how much additional average power would be necessary to compensate for the added noise from the optical switch. The additional power necessary is known as the power penalty of the switch. In networks, this penalty informs us how much additional optical amplification will be necessary in the link as a result of this switch. Figure 3-17 shows the bit-error rate plot for the co-propagating and counter-propagating configurations as compared with the back-to-back at 10 Gb/s. In our experiments, we generally simplify this BER measurement by using a clock as one of our inputs (usually the signal input) instead of needing a second pattern generator and modulator. At these low rates, the counter-propagating and co-propagating performance is similar. We see a power penalty of about 3 dB. The optimization of the various parameters of this switch can make a large difference in the observed power penalties. Later in this chapter, we develop a method which allows us to pinpoint the optimal operating point of this type of switch and improve this result.

One thing to note about the SOA-MZI configuration is that it is difficult to implement using discrete optics. The SOA-MZI is primarily an integrated device. This due

101
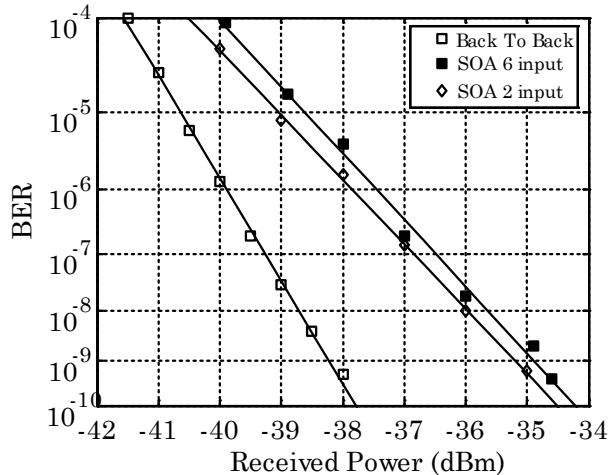
Figure 3-17: Co-propagating and counter-propagating bit-error rate measurements for the SOA-MZI at 10 Gb/s. Results for co-propagating and counter-propagating configurations are similar at low data rates.

to the fact that for this switch to perform well, both arms of the interferometer must be exactly matched. Implementing this interferometer in free space or fiber results in significant drift due to environmental factors such as temperature or vibration. Integration ameliorates these effects by reducing the entire configuration in size and allowing us to package the device for better stability.

An alternative, but equivalent configuration is the ultrafast nonlinear interferometer (UNI) previously described as a discrete optical switch in Chapter 2. The UNI gate has been demonstrated as a discrete logic gate at speeds up to 100 Gb/s for AND and INVERT operation [14] and also as NOR, OR [15], and XOR [16, 17] logic gates. The schematic is repeated here for reference (Figure 3-18). The conceptual operation of the UNI is identical to that of the SOA-MZI, with the difference in implementation being that the two arms of the interferometer are separated by time rather than space. Thus, instead of using two separate control pulses, we split the signal pulse into two orthogonally-polarized and delayed copies. The control pulse is coupled to arrive after the first signal pulse and before the second signal pulse. Thus, the first signal pulse sees no effect from the control pulse while the second signal pulse is strongly affected by the control pulse. Furthermore, since the two signal pulse copies are usually separated by only 5-10 ps, any environmental changes seen by

102
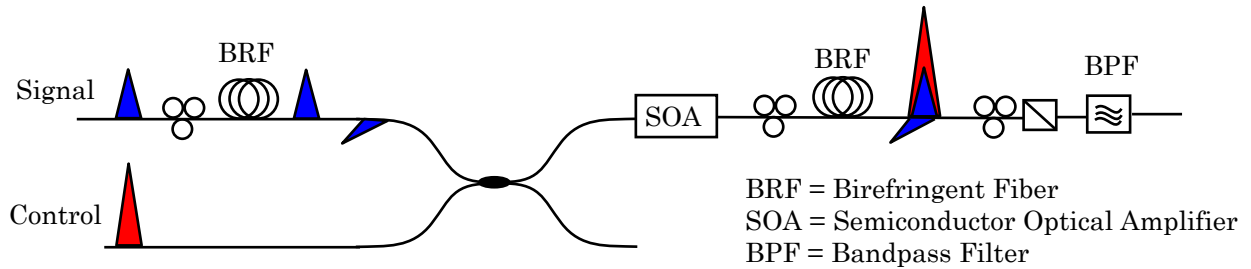
Figure 3-18: Schematic diagram of an ultrafast nonlinear interferometer (UNI). The signal pulse, shown in blue, enters the UNI and is split into two orthogonally-polarized, temporally separated copies. The control pulse, shown in red, is timed to arrive at the SOA in between the two signal pulse copies. This induces a gain and phase shift in the SOA which affects the second pulse but not the first. When the two signal pulse copies are recombined, they interfere constructively or destructively depending on their induced relative phases.

one signal pulse are usually also seen by the second signal pulse. This means that the interferometer remains perfectly balanced. At the output, the two signal pulses are recombined and the control pulse is filtered out. This switch can also be configured to operate in counter-propagating mode with similar advantages and disadvantages compared with the SOA-MZI.

Figure 3-19 shows the BER measurements for the UNI switches in both co- and counter-propagating configurations and inverting and non-inverting configurations at 10 Gb/s. In this case, we have used clock pulses for the signal pulse and modulated the control pulse with a $2^7 - 1$ PRBS. At 10 Gb/s, the counter-propagating configuration does not have a significant disadvantage. In contrast to the SOA-MZI, the UNI interferometer bias cannot be adjusted by changing the SOA current bias, since that would affect both signal pulses equally. Instead, we can adjust the interferometer bias by changing the output polarizer, such that we choose either the constructive or destructive port. In these figures, we can see that the power penalties range from 1.2 dB to about 4 dB, due to the optimization of the switch in that configuration.

Integration of the single arm interferometer UNI is also possible, but faces a significant challenge compared to the SOA-MZI. The UNI relies on separating the
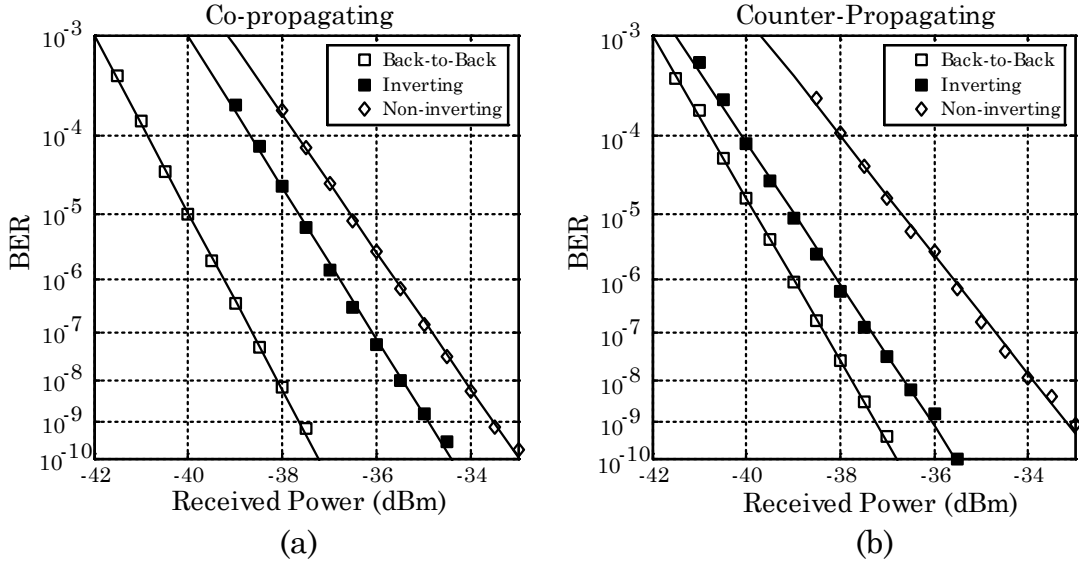
Figure 3-19: Representive (a) co-propagating and (b) counter-propagating BER plots for the ultrafast nonlinear interferometer. Inverting and non-inverting operation is shown. Power penalties range from 1.2 dB to 4 dB, due to optimization of the switch.

two signal pulses in time, which is accomplished in the discrete gate by a birefringent fiber. An integrated birefringence is still not easily available. The UNI has two advantages over the SOA-MZI, however: first, it requires only one SOA, which means that the static power required for switch operation is half that needed for the SOA-MZI. This is a significant fraction of the power consumed by each gate, since an SOA is typically biased to 200 mA or more. Secondly, no waveguide crossing is required. In the SOA-MZI, generating the second control pulse requires a waveguide crossing to send the second control pulse into the bottom arm of the interferometer. Thus, the control pulse will cross over the signal input arm, which can significantly increase crosstalk and degrade the performance of the switch.

As described in Chapter 2, there may be situations where we only require a wavelength conversion rather than a full optical logic gate. While the switches described previously can also be used to convert the control pulses to the signal wavelength if we use a continuous-wave (CW) beam as the signal input, simpler switches capable of ultrafast wavelength conversion can be advantageous because of reduced cost and operating complexity. Two of the most promising cross-phase modulation wavelength-

conversion schemes are the delayed-interference signal-wavelength converter (DISC) and the filtered cross-phase modulation method.

The delayed-interference signal-wavelength converter [18] is shown in Figure 3-20. It is a very simple device, relying only on a nonlinear SOA followed by a delay line.
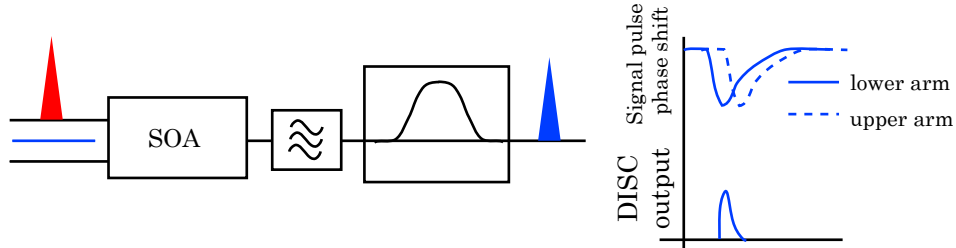


Figure 3-20: Schematic and explanation of the delayed-interference signal-wavelength converter (DISC). An input control pulse (in red) is combined with the CW signal (blue) at the destination wavelength. The control pulse changes the phase of part of the CW signal through XPM, and is filtered out after the SOA. The CW signal is then split and delayed, creating a phase difference which forms the output pulse.

The method works through a differential cross-phase modulation similarly to the UNI and the SOA-MZI. Two inputs, a CW signal and a control pulse train are incident on the SOA. The control pulse train contains the data and the CW signal contains the destination wavelength. The presence of the control pulse induces a phase shift in the CW signal at time $t_o$. After the SOA, an interferometer splits the CW signal and delays one copy with respect to the other copy. The phase shift induced by the control pulse is thus offset between the two CW signal copies. We adjust the control pulse power such that this differential phase shift is exactly $\pi$. When the CW light interferes at the output, the phase difference between the two arms of the interferometer changes the output from destructive interference to constructive interference, forming a pulse. Thus, the delay in the interferometer determines the pulse width of the output pulses. Provided the interferometer at the output is kept stable, the speed of this method is only limited by the control pulse width and the delay in the interferometer, which limits the output pulse width. Leuthold et al. [19] have demonstrated this switch operation in integrated form at 100 Gb/s.

Another method which uses cross phase modulation in a simple configuration

involves a filtered cross-phase modulation, shown in Figure 3-21 [20, 21, 22, 23]. As
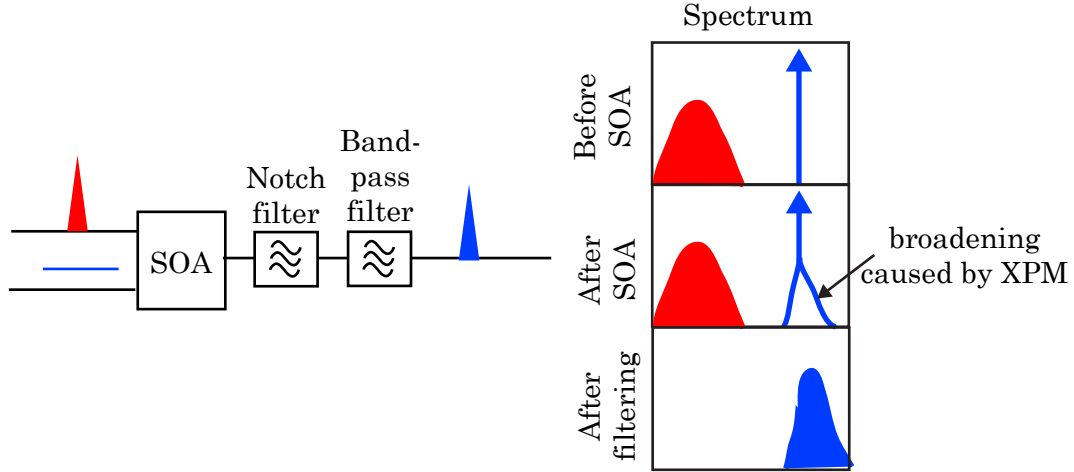


Figure 3-21: Schematic and explanation of XPM-based switch. A control pulse (red) is coupled into the SOA along with a CW signal (blue). Cross-phase modulation between the control and signal broadens the CW line of the signal. By filtering out one side of the broadened CW line with a bandpass filter and knocking out the CW line with a notch filter, we are able to recover the data pattern, now at a new wavelength.

with the DISC method, a CW line is coupled into the SOA along with a control pulse train. In the SOA, cross-phase modulation causes the control pulses to mix with the CW line, broadening the spectrum. By using a notch filter to remove the CW line and a bandpass filter to select one side of the broadened spectrum, we recover the pulse train at the new wavelength. This method has been demonstrated at a remarkable 640 Gb/s [24]. The one major disadvantage of this method is the difficulty for integration, due to the difficulty of integrating filters.

### 3.2.3 Four-Wave Mixing (FWM)

Finally, four-wave mixing involves an interaction between two input electric fields and results in two new frequencies:

$$P(\mathbf{r},\omega_3) = \epsilon_o \chi^{(3)}(\omega_3 = 2\omega_1 - \omega_2) E_2(\mathbf{r},\omega_2)^2 E_1^*(\mathbf{r},\omega_1) \tag{3.41}$$

$$P(\mathbf{r},\omega_4) = \epsilon_o \chi^{(3)}(\omega_4 = 2\omega_2 - \omega_1) E_1(\mathbf{r},\omega_2)^2 E_2^*(\mathbf{r},\omega_1). \tag{3.42}$$

If the optical inputs $E_1$ and $E_2$ are the signal and control pulse trains, this process can be used to obtain the logic $S \cdot C$ if we take the output at a separate wavelength. Four-wave mixing is an attractive switch due to its broad bandwidth in the SOA and its transparency to modulation formats and bit rates. Since FWM is a parametric process, it can preserve the phase of incoming signals. This makes it suitable for modulation formats which encode information in the phase, such as differential phase shift keying (DPSK). It has been demonstrated for DPSK XOR at 20 Gb/s [25] and for XNOR, AND, NOR, and NOT at 10 Gb/s [26]. It can also be used as a simple wavelength converter. If we use a CW line as the signal and send in our data on the control port, the data will be perfectly replicated (with a conjugation) at one of the two generated wavelengths. This is a huge advantage over the cross-gain and cross-phase modulation which are only well-suited to on-off keyed modulation formats. Figure 3-22 shows the spectrum of a wavelength conversion using four-wave mixing. Here, we can also observe some broadening in the CW line due to cross-phase modulation.
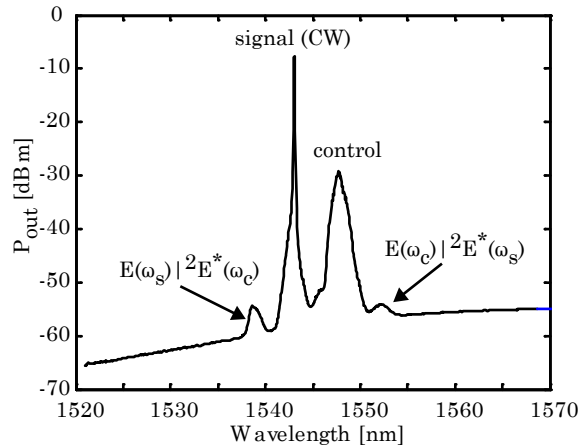


Figure 3-22: Spectrum of four-wave mixing wavelength conversion. The signal CW line mixes with the control pulses to create replicas of the control pulses at $2\omega_s - \omega_c$.

However, as shown in Figure 3-22, conversion efficiency remains a major challenge for four-wave mixing. Typical conversion efficiencies are on the order of -20 dB, which represents a large loss in a switch. Furthermore, any logic automatically involves a wavelength conversion. This makes the switch less versatile than cross-phase modu-

lation methods which can be operated in counter-propagating mode to maintain the wavelength.

## 3.3 Performance Optimization of the SOA-MZI

In the previous section, we have covered the basics of how to use SOA nonlinearities to form optical switches for logic as well as for wavelength conversion. Currently, for integrated all-optical logic, the SOA-MZI is the most promising switch due to its simple design, ultrafast performance, and versatility. In order to prepare for more practical operation of the SOA-MZI, we need to fully characterize the device and investigate the development of the SOA-MZI from a research project in the laboratory to a device that can be easily deployed and operated in the field. Though this device operation is straightforward in concept, in practice, determining the optimal point for a particular device can be very time-consuming. The optimization is performed over a large parameter-space which includes the signal and control pulse energies, their relative delay, and the bias currents for the switching SOAs. In particular, changing the bias currents for the switching SOAs can completely change the interferometer bias from inverting to non-inverting. As seen in Figure 3-11, the bias currents can also significantly change the recovery time of an SOA.

In light of this, we have developed a technique for optimizing operation of this switch and thoroughly characterizing its behavior [27]. We can measure a bias map of the constructive and destructive interference fringes of the interferometer by measuring the transmission through the SOA-MZI at all values of the SOA current biases. By repeating this for each value of the signal-control delay, we can observe how this bias map changes over all possible delays. This gives us insight into the device dynamics at all the possible bias points for the SOA-MZI. Furthermore, this measurement allows us to quickly compare different bias points for switching efficiency and carrier recovery times. In order to measure the effect of the carrier recovery time, we must use the switch in single-ended mode, with only one control pulse.

### 3.3.1   Static Bias Scan: Method and Analysis

We used a semiconductor optical amplifier Mach-Zehnder interferometer (SOA-MZI) provided by Alphion Corporation (Fig. 3-23). This device included amplifying SOAs (shown in yellow) as well as switching SOAs (shown in orange) to compensate for the coupling losses and waveguide losses in the MZI. In single-ended operation, SOA 3 is unused.
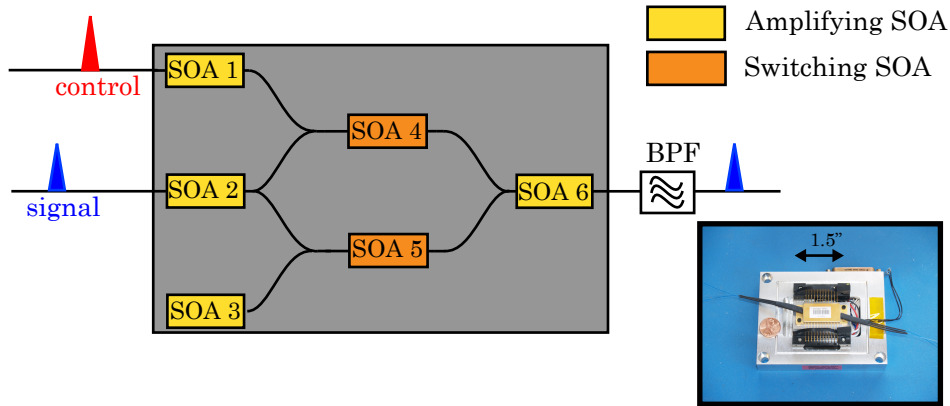


Figure 3-23: Schematic of the semiconductor optical amplifier Mach-Zehnder interferometer (SOA-MZI). Inset shows a photograph of the packaged device. SOAs 4,5 perform switching while SOAs 1-3, and 6 amplify. SOA 3 is not used in single-ended operation. BPF is a band-pass filter.

To determine the optimal operating point of this device, we first measure the bias of the interferometer without a control input. We refer to this as the static bias scan. Our original scan involved stepping the current biases of the two switching SOAs over all possible values (from 0 to 1000 mA) while recording the average output power at each combination of values. We controlled the instruments using GPIB and a laptop. Unfortunately, the speed of the GPIB control combined with the speed of the current sources proved to be prohibitively slow, resulting in a full scan time of about 9 hours. In order to improve the speed of the scan, we applied a sawtooth modulation to one of the current sources in order to quickly sweep the current on one SOA while stepping the current on the other SOA. Figure 3-24(a) shows the schematic of this measurement. We use a Hall-effect probe to measure the current sweep caused by

the modulation with an oscilloscope. We calibrate this measurement to the current driving SOA 4, $I_4$. At the output of the SOA-MZI, we use a photodetector and an oscilloscope in order to measure the average output power. We calibrated the voltage measurement on the oscilloscope to a power measurement performed by an optical spectrum analyzer. We triggered the oscilloscope at the same frequency as the sawtooth modulation. Thus, for each value of the current on SOA 5 (the "slow axis"), we sweep the current on SOA 4 (the "fast axis") and measure the transmission through the interferometer across all values of $I_4$. The blue trace in Figure 3-24(a) shows an example of a single sweep across $I_4$. By doing this for each value of $I_5$, we build up the static bias scan, a two-dimensional plot of the interferometer bias as shown in Figure 3-24(b).
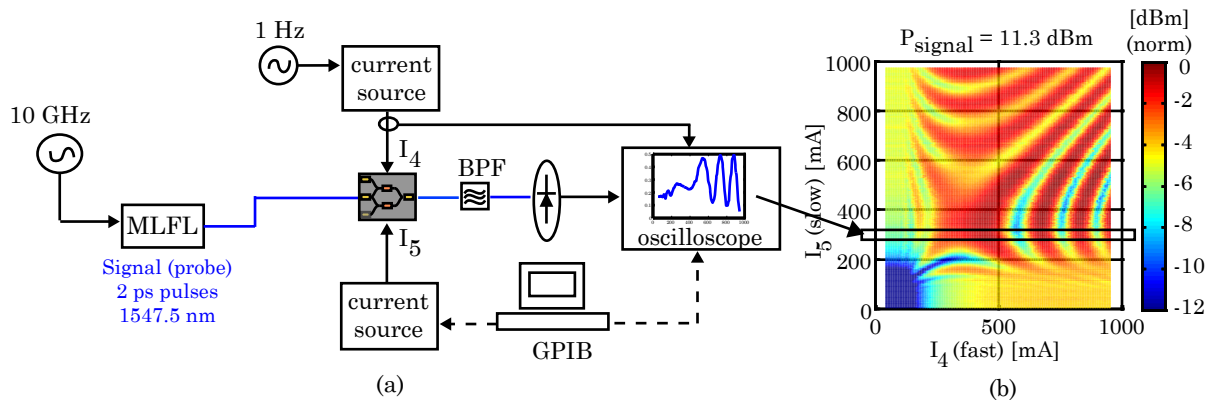


Figure 3-24: (a) Experimental schematic for the static bias scan. BPF stands for bandpass filter. (b) Static bias map of the SOA-MZI interferometer. Areas of constructive interference show up as deep red while areas of destructive interference show up as deep blue.

We used a sweep frequency of 1 Hz for this scan, thus reducing each static bias scan to about 7 minutes, or 185 points per second. The speed of this sweep frequency is limited by two factors: first, by the modulation bandwidth of the current source and secondly, by thermal effects in the SOA-MZI. In sweeping the current from 0 to 1 Amp in the device, we are creating a significant amount of local heating. Since the carrier densities are highly dependent on temperature, local large changes in temper-

ature can significantly change the gain and phase effects and thus the interferometer bias. In typical operation, the SOAs would be maintained at a single current, i.e., DC operation. Therefore, we have two options for our current sweep: operate as close to DC as possible while still maintaining a reasonable sweep speed, or operate fast enough to avoid thermal heating or cooling. The latter is impractical with our hardware implementation due to the limit on the current modulation bandwidth (<50 kHz). To determine the maximum current sweep frequency we can use, we set $I_5$ to a single value and compared one slice of the bias scan using the current sweep with another slice of the bias scan using the slow scan method. Figure 3-25(a) shows the sweep comparison for the slow scan versus a scan at 100 Hz. As can be seen from this plot, the thermal effects from the current sweep can make the scan useless for pinpointing the correct bias. Furthermore, as seen in Figure 3-25(b), even at the slow 1 Hz scan, the slow scan differed by a small amount. Since a scan rate of 7 minutes per scan is already very long for a measurement of multiple scans, we decided to correct the remaining thermal effects in post-processing, by calibrating the 1 Hz scan to the slow scan.
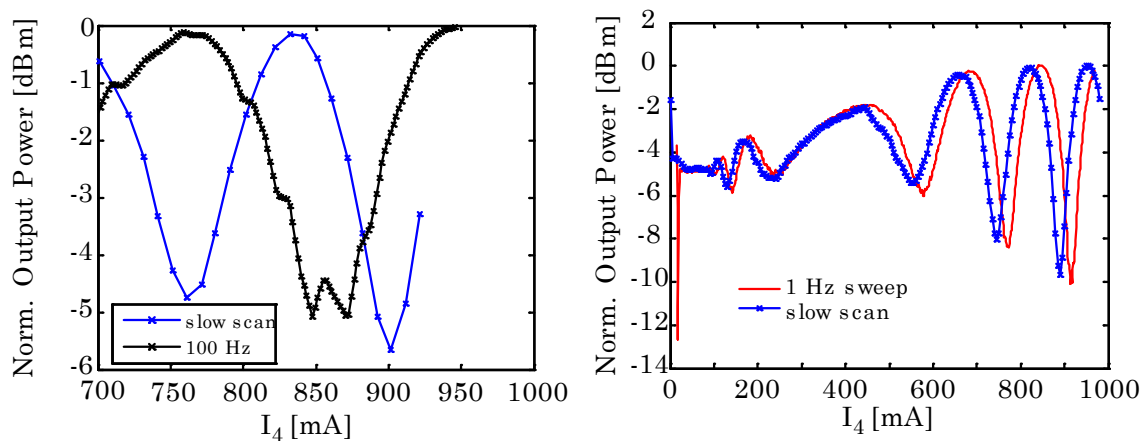


Figure 3-25: (a) shows the thermal effects of the current sweep at 100 Hz. Note that the nulls become peaks due to this effect. (b) shows the thermal effects of the current sweep at 1 Hz. The offset is much smaller, but can be compensated in post-processing.

Even this static bias scan can provide significant information about the switch operation. In Figure 3-26, we compare the differences in the static bias scan between

sending our signal input into SOA 2 in Figure 3-23 and reversing the SOA-MZI and sending our signal input into SOA 6. In Figure 3-26(a), we note that the SOA 2 input



(a) SOA 2 input
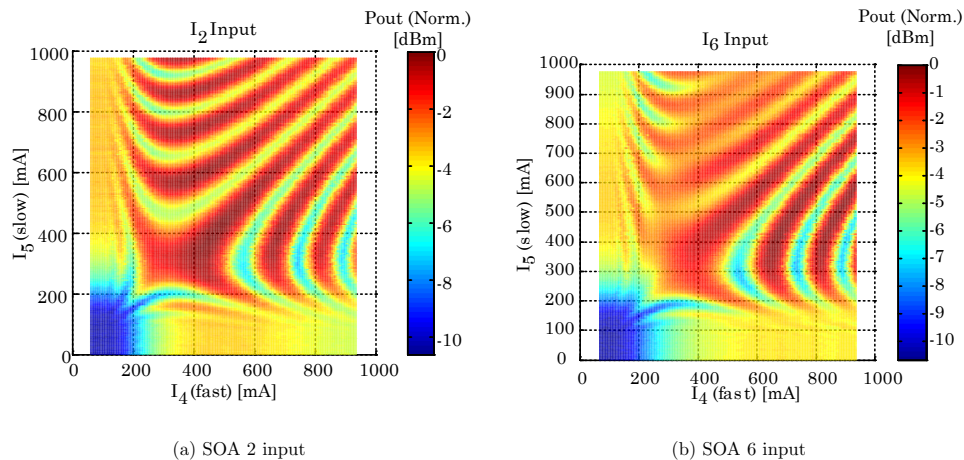
(b) SOA 6 input

Figure 3-26: (a)Static bias scan where the signal pulse is input into SOA 2, as shown in Figure 3-24. (b) Static bias scan where the SOA-MZI is reversed and the signal pulse is input into SOA 6.

scan is much more symmetric than the SOA 6 input scan. This asymmetry indicates that SOA 4 cannot entirely compensate for the gain introduced by SOA 5. As the current in SOA 5 increases, the peaks and nulls of the interferometer are washed out as seen in the upper part of the figure. This indicates that the output from the two arms of the interferometer are not well matched in amplitude and thus do not completely cancel or add when interfered. In contrast, even at high SOA 4 currents, we see high contrast fringes for the interferometer. This means that the two combining pulses are evenly matched, making it possible to observe high constructive or destructive interference. On the other hand, when we observe Figure 3-26(b), where we have sent our signal into SOA 2 instead of SOA 6, we see a much more symmetric bias scan. This indicates that the two interferometer arms are better matched. Since the SOAs in use are identical, the reason for this must lie in the interferometer couplers. If the input coupler is unbalanced, this has a smaller effect on the output pulse power due to the nonlinearity of gain saturation. Thus, by reversing directions of the static bias scan, we can determine which coupler is better matched, and choose to use the device in that direction.

### 3.3.2 Dynamic Bias Scan: Method and Analysis

Our main interest is in how the optical control pulses induce switching in the SOA-MZI. To observe this, we measure a dynamic bias map by inserting control pulses and measuring a static bias map at each value of the signal-control delay $\tau$. This essentially combines the switching window measurement with the static bias scan and allows us to simultaneously measure the switching window at all possible biases of the SOA-MZI. Figure 3-27 shows the addition of the control pulse train and the variable delay in the setup. By arranging our series of static bias scans in order
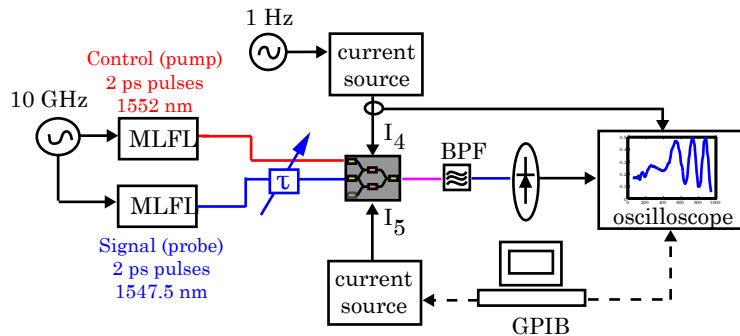


Figure 3-27: Schematic for the dynamic bias scan. We introduce the control pulse train and vary its delay with respect to the signal pulse train through the use of a variable delay $\tau$. At each value of the signal-control delay, we measure the static bias scan.

of increasing signal-control delay, we can create a dynamic movie of the all-optical switching dynamics, as shown in Fig. 3-28(a). For those reading this in print, I have included all the movie frames in Appendix A. For this measurement, the control pulses are generated by a mode-locked fiber laser at 1552 nm with a repetition rate of 10 Gbit/s and are inserted into one arm of the interferometer. Signal pulses are generated with a second, synchronized mode-locked fiber laser at 1547.5 nm with a repetition rate of 10 Gbit/s. The average power of the control pulses is 0.4 dBm and the average power of the signal pulses is -8.1 dBm. Figure 3-28(b) shows the relative control-signal delay for each frame of the dynamic bias map and Fig. 3-28(c) shows the switching window at the operating point "X" on Fig. 3-28(a). The switching window is the plot of the output signal power at each control-signal delay. Here,
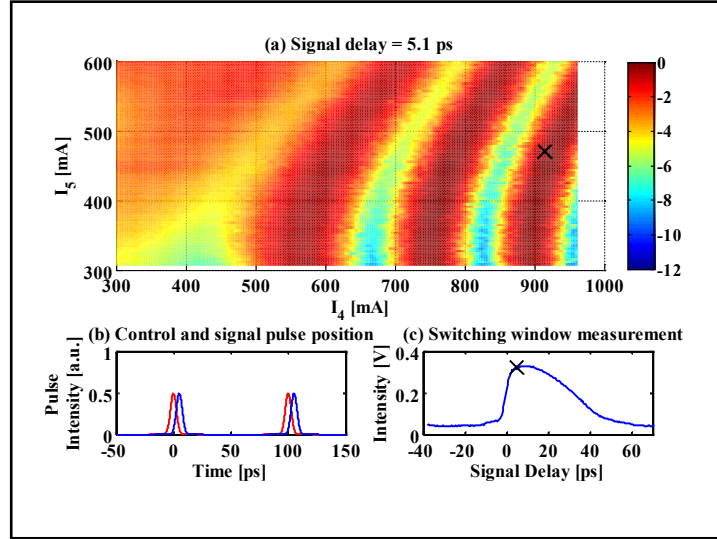
Figure 3-28: (a) Dynamic bias map of the SOA-MZI interferometer. (b) Relative signal and control pulse positions corresponding to each frame of the bias map. (c) Switching window plot at position "X". (*Size: 848 KB.*)

we have zoomed into a region of interest in order to increase the resolution of the measurement and to increase the speed of the entire measurement. We picked this region of interest since the contrast between the interference fringes is higher than anywhere else in the static bias scan. We see that the presence of the control pulse shifts the bias of the interferometer and moves the signal pulse at point "X" from destructive to constructive interference. This results in non-inverting operation. We can also observe the slow carrier recovery time of the switch, limiting single-ended switching performance in the SOA-MZI to about 20 Gb/s at our operating point.

We also note that in the static bias map, the distance between the fringes in the $I_5$ direction is larger than in the $I_4$ direction, indicating that a larger control pulse energy would be needed if we used the lower interferometer arm as opposed to the upper interferometer arm. We sent our control pulses into the upper arm of the interferometer and thus shifted the bias of SOA 4. This is why the interferometer fringes move along the $I_4$ axis as opposed to the $I_5$ axis.

The varying widths of the fringes also indicate that at different bias currents, a different amount of control pulse power will be needed to achieve full switching. Furthermore, when the control pulse energy first hits the device, we see a bleaching

114

of the fringe pattern before the fringes shift. The peaks and nulls of the interferometer are washed out to some extent. This indicates that we are seeing a gain effect slightly before we observe the phase shift due to the incoming control pulse. If the two output pulses are slightly offset in power, complete constructive or destructive interference will no longer occur, and thus the interference fringes will not be as deep.

### 3.3.3 SOA-MZI optimization

In addition to information about switching dynamics, we can use these data to accurately pinpoint the optimal operating point for high-extinction ultrafast switching. For the extinction map, we calculate the extinction at each operating point and subtracting the minimum power output from the maximum power output [28, 29]. For non-inverting switching, the maximum output occurs where the control pulse is overlapped with the signal pulse, while the minimum output occurs when the control pulse arrives behind the signal pulse and thus affects the signal pulse the least. Aside from cumulative gain saturation effects, this approximates the difference between when the control pulse exists and when there is no control pulse. The power difference between the two states indicates the power difference between a "1" and a "0" bit at the output. The larger this extinction, the lower the bit error rate.

Figure 3-29(a) shows the extinction map of this measurement. Regions of high extinction occur for both non-inverting and inverting operation of the switch, and are separated by regions of low extinction. We see that higher extinctions (deeper reds) exist for inverting operation than for non-inverting operation. As explained previously, in inverting operation, cross-gain modulation aids the switch and thus better performance is achieved than in non-inverting operation. However, for applications, an inversion is unwanted and would require an additional inverting switch to restore the polarity of the signal. Thus, we are mainly interested in non-inverting switch operation. The optimal extinction for non-inverting operation was found to be 8.8 dB at $I_4$=893.5 mA and $I_5$=470.0 mA, marked by the black square. To verify the accuracy of this measurement, we perform a 10-Gbit/s wavelength conversion at this operating point and compare it to a wavelength conversion performed at a nearby, non-optimal

operating point. We found this nearby point through the typical process without this performance method by adjusting the currents, delay, and powers in sequence to minimize the bit-error rate. This nearby operating point was found to be $I_4$=905.5mA and $I_5$=476.6mA, marked by the black triangle. Figure 3-29(b) compares the BER at both operating points. The non-optimal operating point results in an additional 1-dB penalty as compared to the optimal point found using this optimization method.
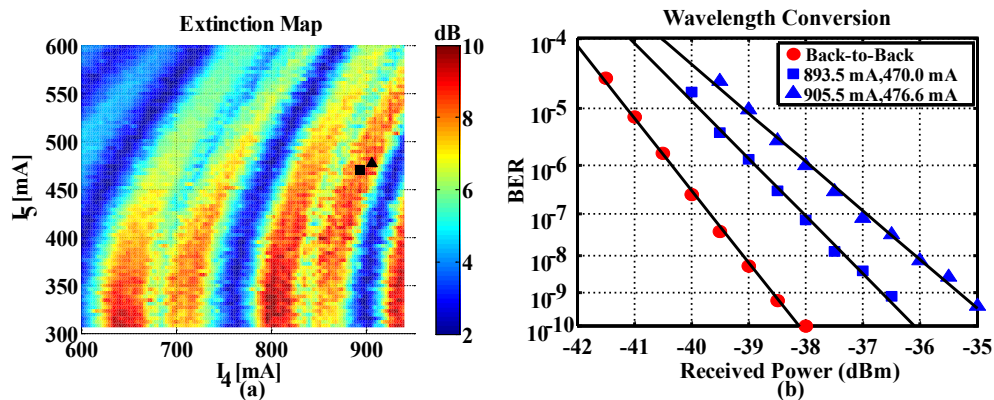


Figure 3-29: (a) Extinction map. Regions of high extinction (red) are separated by regions of low extinction (blue). (b) Bit-error rate plot of wavelength conversion performed at the optimal operating point for non-inverting switch operation (square) as compared with wavelength conversion performed at a nearby non-optimal point (triangle).

## 3.4 Conclusion

For all-optical ultrafast switching to become a reality, all-optical logic gates must be integrated for ease of manufacturing, installation, and operation in the field. One of the most promising logic gates is the semiconductor optical amplifier Mach-Zehnder interferometer, which uses cross-phase and cross-gain modulation to achieve ultrafast switching speeds. It has been demonstrated to have ultrafast switching capabilities and can implement a variety of logic operations. Furthermore, we have found the switch to be highly stable in its optimal operating point. In the field, these switches cannot be constantly adjusted to optimize switch performance. We found, however, that the optimal operating currents for the SOA-MZI remained stable over several

days, requiring little to no adjusting for optimal operation. This further indicates the potential of the SOA-MZI for future all-optical processing techniques.

In this chapter, we have presented a brief description of the underlying physics of SOA-based logic gates and described a few of the most common switching configurations for optical logic as well as wavelength conversion. Finally, we have presented a method for optimization and characterization of the SOA-MZI gates which allows us to quickly determine the optimal bias points for high-extinction ultrafast switching. This method, while useful for single-gate logic, becomes critical for signal processing requiring multiple logic gates. We have illustrated the effectiveness of the method by comparing the bit-error rate found using a traditional blind optimization method with the pin-point accuracy offered by the dynamic bias map.

In the next chapter, we shall use this technique to address the issue of regeneration for optical networks. This allows us to investigate the cascadability of the SOA-MZI switch while researching its potential for alleviating the cost of optical transmission.

# Bibliography

[1] S. M. Sze, *Semiconductor devices, physics and technology.* Wiley, 2002.

[2] N. W. Ashcroft and N. D. Mermin, *Solid State Physics.* Saunders College, 1976.

[3] B. S. Robinson, *Semiconductor-Based All-Optical Switching for Optical Time-Division Multiplexed Networks.* PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, June-Aug. 2003.

[4] G. P. Agrawal and N. A. Olsson, "Self-phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers," *IEEE Journal of Quantum Electronics*, vol. 25, no. 11, pp. 2297–2306, 1989.

[5] M. Yamada, "Transverse and longitudinal mode control in semiconductor injection lasers," *IEEE Journal of Quantum Electronics*, no. 9, pp. 1365–1380, 1983.

[6] Y. R. Shen, *The Principles of Nonlinear Optics.* John Wiley and Sons.

[7] J. P. Wang, "Complete characterization of optical pulses in the picosecond regime for ultrafast communication systems," Master's thesis, Massachusetts Institute of Technology, 2002.

[8] R. W. Boyd, *Nonlinear Optics.* Academic Press, second ed., 2003.

[9] K. L. Hall, G. Lenz, A. M. Darwish, and E. P. Ippen, "Subpicosecond gain and index nonlinearities in InGaAsP diode lasers," *Optics Communications*, vol. 111, no. 5, 1994.

[10] S. Nakamura, Y. Ueno, and K. Tajima, "Error-free all-optical demultiplexing at 336 Gb/s with a hybrid-integrated symmetric-mach-zehnder switch," in *Optical Fiber Communications Conference (OFC '02)*, 2000.

[11] M. Heid, S. L. Jansen, S. Spälter, E. Meissner, W. Vogt, and H. Melchior, "160-Gbit/s demultiplexing to base rates of 10 and 40 Gbit/s with a monolithically integrated SOA-mach-zehnder interferometer," in *28th European Conference on Optical Communication (ECOC '02)*, vol. 3, p. 1, 2002.

[12] H. Dong, H. Sun, Q. Wang, N. K. Dutta, and J. Jacques, "80 Gb/s all-optical logic AND operation using mach-zehnder interferometer with differential scheme," *Optics Communications*, vol. 265, pp. 79–83, 2006.

[13] R. P. Webb, R. J. Manning, G. D. Maxwell, and A. J. Poustie, "40 Gbit/s all-optical XOR gate based on hybrid-integrated mach-zehnder interferometer," *Electronics Letters*, vol. 29, no. 1, pp. 79–81, 2003.

[14] K. L. Hall and K. A. Rauschenbach, "100-Gbit/s bitwise logic," *Optics Letters*, vol. 23, 1998.

[15] N. S. Patel, K. L. Hall, and K. A. Rauschenbach, "40-Gbit/s cascadable all-optical logic with an ultrafast nonlinear interferometer," *Optics Letters*, vol. 21, no. 18, pp. 1466–1468, 1996.

[16] C. Bintjas, M. Kalyvas, G. Theophilopoulos, T. Stathopoulos, H. Avramopoulos, L. Occhi, L. Schares, G. Guekos, S. Hansmann, and R. Dall'Ara, "20 Gb/s all-optical XOR with UNI gate," *IEEE Photonics Technology Letters*, vol. 12, no. 7, pp. 834–836, 2000.

[17] G. Theophilopoulos, K. Yiannopoulos, M. Kalyvas, C. Bintjas, G. Kalogerakis, H. Avramopoulos, L. Occhi, L. Schares, G. Guekos, S. Hansmann, and R. Dall'Ara, "40 GHz all-optical XOR with UNI gate," in *Optical Fiber Communications Conference (OFC '01)*, vol. 1, 2001.

[18] Y. Ueno, S. Nakamura, K. Tajima, and S. Kitamura, "3.8-THz wavelength conversion of picosecond pulses using a semiconductor delayed-interferometer signal-wavelength converter DISC," *IEEE Photonics Technology Letters*, vol. 10, no. 3, pp. 346–348, 1998.

[19] J. Leuthold, C. H. Joyner, B. Mikkelsen, G. Raybon, J. L. Pleumeekers, B. I. Miller, K. Dreyer, and C. A. Burrus, "100 Gb/s all-optical wavelength conversion with integrated soa delayed-interference configuration," *Electronics Letters*, vol. 36, no. 13, pp. 1129–1130, 2000.

[20] J. Leuthold, D. Marom, S. Cabot, J. Jaques, R. Ryf, and C. Giles, "All-optical wavelength conversion using a pulse reformatting optical filter," *Journal of Lightwave Technology*, vol. 22, 2004.

[21] M. Nielsen, B. Lavigne, and B. Dagens, "Polarity-preserving soa-based wavelength conversion at 40 Gbit/s using bandpass filtering," *Electronics Letters*, vol. 39, no. 18, pp. 1334–1335, 2003.

[22] Y. Liu, E. Tangdiongga, Z. Li, S. Zhang, H. deWaardt, G. Khoe, and H. Dorren, "Error-free all-optical wavelength conversion at 160 Gb/s using a semiconductor optical amplifier and an optical bandpass filter," *Journal of Lightwave Technology*, vol. 24, 2006.

[23] Y. Liu, E. Tangdiongga, Z. Li, H. deWaardt, A. Koonen, G. Khoe, H. Dorren, X. Shu, and I. Bennion, "Error-free 320 Gb/s SOA-based wavelength conversion using optical filtering," in *Conference on Optical Fiber Communications (OFC)*, 2006.

[24] M. Galili, H. C. H. Mulvad, L. K. Oxenlowe, H. Ji, A. T. Clausen, and P. Jeppensen, "640 Gbit/s wavelength conversion," in *Optical Fiber Communication Conference (OFC '08)*, 2008.

[25] K. Chan, C.-K. Chan, L. K. Chen, and F. Tong, "Demonstration of 20-Gb/s all-optical XOR gate by four-wave mixing in semiconductor optical amplifier with RZ-DPSK modulated inputs," *IEEE Photonics Technology Letters*, vol. 16, no. 3, pp. 897–899, 2004.

[26] G. Berrettini, A. Simi, A. Malacarne, A. Bogoni, and L. Poti, "Ultrafast integrable and reconfigurable XNOR, AND, NOR, and NOT photonic logic gate," *IEEE Photonics Technology Letters*, vol. 18, no. 8, pp. 917–919, 2006.

[27] J. P. Wang, B. S. Robinson, S. J. Savage, S. A. Hamilton, E. P. Ippen, R. Mu, H. Wang, J. Sarathy, and B. B. Stefanov, "A performance optimization method for SOA-MZI devices," in *Optical Fiber Communication Conference (OFC)*, 2007.

[28] J. P. Wang, B. S. Robinson, S. J. Savage, S. A. Hamilton, E. P. Ippen, R. Mu, H. Wang, J. Sarathy, and B. B. Stefanov, "Efficient performance optimization for ultrafast all-optical switching in SOA-MZI devices," in *IEEE LEOS Photonics in Switching Conference*, 2007.

[29] J. P. Wang, B. S. Robinson, S. J. Savage, S. A. Hamilton, E. P. Ippen, R. Mu, H. Wang, J. Sarathy, and B. B. Stefanov, "Efficient performance optimization of SOA-MZI devices," *Optics Express*, vol. 16, no. 5, pp. 3288–3292, 2008.

# Chapter 4

# Data Regeneration with Integrated Optical Logic Gates

In networks today, data is transmitted from source to destination through electronic routers and wavelength-division-multiplexed (WDM) fiber connections. In Chapter 2, we focused on optical switching techniques for routing and switching. Here, we focus on the transmission between routers in a network. Over long distances, fiber has entirely replaced copper links due to ultra-low loss ($<$ 0.25 dB/km) and high capacity ( $>$ 20 THz bandwidth). Prior to the development of the wide-bandwidth erbium-doped fiber amplifier (EDFA) in the late 1980s and early 1990s, electronic regenerators were necessary every 60-80 km (15-20 dB) to compensate for loss, dispersion, and nonlinearities in fiber and to maintain low bit error rates in the network. A single EDFA could simultaneously compensate for 20-30 dB of loss across multiple wavelength channels. Combined with simple dispersion compensation modules consisting of a length of specially-designed optical fiber to cancel second-order dispersion, this allowed network designers to replace many electronic regenerators and significantly reduce costs. However, regeneration is still necessary over long distances to compensate for higher-order dispersion effects as well as nonlinearities. Currently, EDFAs are spaced by 80-100 km and electronic regenerators are placed after 2-3 amplifiers (Figure 4-1). The distance between regenerators is known as the regenerator reach.

80 km

320 km
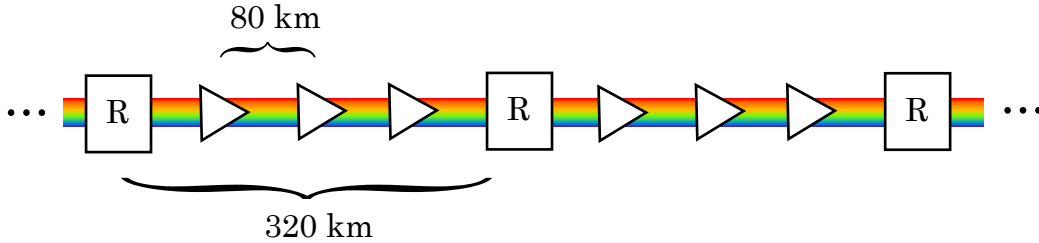
Figure 4-1: Schematic of regeneration and amplification in a WDM fiber link. R stands for regenerator. The triangles represent erbium-doped fiber amplifiers.

As we will show in Section 4.1, the demand for increasing network capacity tends to lead to decreased regenerator reach due to increased sensitivity to interference, dispersion, and nonlinear effects. Tighter channel spacing increases the inter-channel interference from nonlinear effects. Higher channel bit rates result in higher sensitivity to dispersion and larger nonlinear effects due to higher peak-to-average power ratios. In order to maintain low infrastructure costs, we can either increase regenerator reach or reduce the cost of each regenerator.

Research towards increasing the reach of regenerators has focused on investigating pre-/post-compensation techniques and on alternative modulation formats which are less sensitive to fiber effects. Compensation techniques involve careful modeling of expected fiber distortions in the data pattern and reversing the distortions either at the receiver or transmitter. For instance, if we know the phase effects the pulse will accumulate, we can generate a pulse at the transmitter with the opposite phase effects. Thus, as the pulse propagates down the known length of fiber, the distortions are reversed. This requires accurate knowledge of the path the pulses take. In point-to-point links, such as undersea links, this can eliminate the need for regeneration entirely [1, 2]. However, these techniques are not suited for data networks, where different packets may travel along very different paths before reaching the same destination.

Alternative modulation formats are more promising for increasing regenerator reach in data networks. Fiber transmission originally used non-return-to-zero (NRZ) modulation, in which each bit was encoded in a pulse that extended through the entire

124

bit period (Figure 4-2(a)). As bit rates increase, the close proximity of bits means that NRZ is more sensitive to inter-symbol interference from dispersion and nonlinearities. Return-to-zero (RZ) modulation (Figure 4-2(b)) can support higher bit rates per channel. However, the shorter pulses used in RZ modulation result in a broader spectrum, which limits the density of wavelength channels for WDM transmission. Recently, differential phase-shift keying (DPSK) has emerged as a highly promising modulation format for increasing regenerator reach in ultrafast networks [3]. With DPSK, data is encoded as a differential phase, usually over RZ pulses (Figure 4-2(c). A $\pi$ phase shift between two pulses encodes a one bit, whereas a 0 phase shift encodes a zero bit. To demodulate and detect the bit pattern, a 1-bit interferometer causes adjacent pulses to interfere. This differential modulation results in a natural 3-dB reduction in the power penalty, corresponding to a potential doubling in the transmission span.
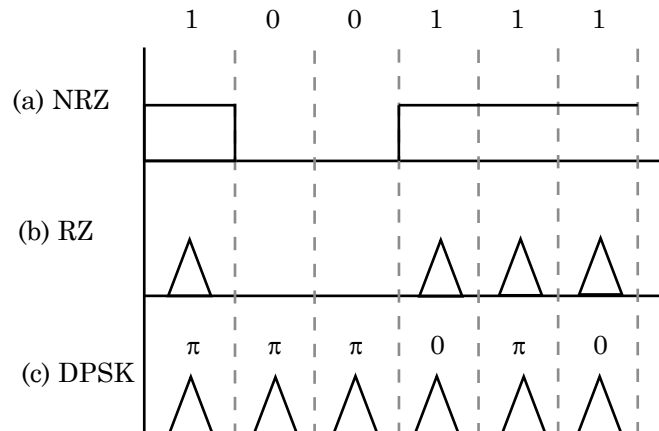


Figure 4-2: Diagram of three different modulation formats: NRZ (non-return-to-zero), RZ (return-to-zero) and DPSK (differential phase-shift keying).

In this chapter, we focus on the potential of optical processing techniques to reduce regenerator costs. Current regenerators are implemented entirely in electronics and, similarly to routers, involve expensive O/E/O conversions and demultiplexing to bring channel bit rates down to electronic switch rates. Figure 4-3 shows a schematic of an electronic regenerator. Wavelength-multiplexed data enters from the left on one fiber and must be demultiplexed into individual wavelength channels. This is a
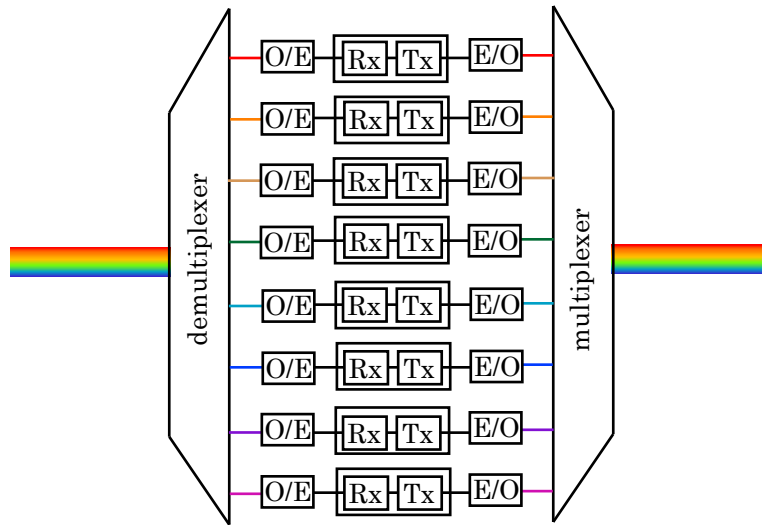
Figure 4-3: Schematic of electronic regenerator. O/E is the optical-to-electronic converter, generally a photodetector and filter. E/O is the electronic-to-optical converter, generally a laser diode and modulator. Rx is the receiver which detects the incoming data. Tx is the transmitter. Occasionally traffic monitoring and grooming is done between the receiver and transmitter.

relatively simple operation and can be accomplished by a simple arrayed waveguide grating. Each channel needs a separate O/E/O conversion and regeneration. The O/E conversion usually involves a photodetector as well as electronic filters and amplifiers. The receiver detects the electronic pulses and determines if each pulse is a "1" or a "0". The transmitter receives the bit information and encodes it in a series of electronic pulses. This is then followed by a E/O conversion, usually involving a laser diode and an external electro-optic modulator. Between the receiver and transmitter, traffic monitoring, traffic grooming, and other types of packet processing can be added.

As channel bit rates increase, a parallel architecture is required to allow electronic processing to keep up with capacity. Thus, a single wavelength channel may be demultiplexed further to four parallel streams, each of which requires its own O/E/O conversion and receiver/transmitter pair. We believe that all-optical techniques can help reduce the size, complexity, and power required for these functions. Integrated all-optical switches can replace the complex hardware required for electronic switching in regenerators. Furthermore, as discussed in the previous chapter, integrated all-

optical switches have been shown capable of ultrafast operation at speeds > 100 Gb/s, thus eliminating the need for parallel architectures. The ultimate goal here would be a multi-wavelength all-optical regenerator which could regenerate all wavelength channels simultaneously, similar to the EDFA. However, few options currently exist for multi-wavelength optical switching in either electronic or optical domains [4].

In this chapter, we shall demonstrate how the integrated optical logic gate presented in Chapter 3 performs as an all-optical regenerator. We first begin with a brief background on the linear and nonlinear effects in optical fiber and show how they limit regenerator spans at high bit rates. Then, we give a brief overview of optical regeneration requirements. Finally, we show how the SOA-MZI can meet these requirements by demonstrating its performance as an all-optical regenerator in a re-circulating loop. We demonstrated error-free regeneration at 10 Gb/s over 10,000 km in 100 passes.

## 4.1 Pulse Propagation in Optical Fiber

Optical fiber is formed of a core and a cladding, using doped silica glass to create a different index of refraction in each region. A cross-section of step-index fiber is shown in Figure 4-4. The higher index of refraction confines the optical mode to the
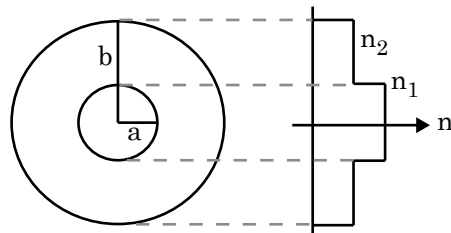


Figure 4-4: Cross-section of optical fiber. The inner core (radius $a$) is doped to have a higher index of refraction ($n_1$) than the cladding (outer radius $b$) with an index of refraction ($n_2$).

core of the fiber. The radius of the core determines the number of modes which can fit in the fiber at a particular wavelength. Most telecommunications fibers are designed to be single-mode. The two main effects which govern pulse propagation in optical

fiber are dispersion and nonlinearity. Dispersion arises from the fact that the index of refraction $n$ is wavelength-dependent. Figure 4-5 shows the measured dependence of $n$ on wavelength for fused silica [5]. Like semiconductor optical amplifiers, the main
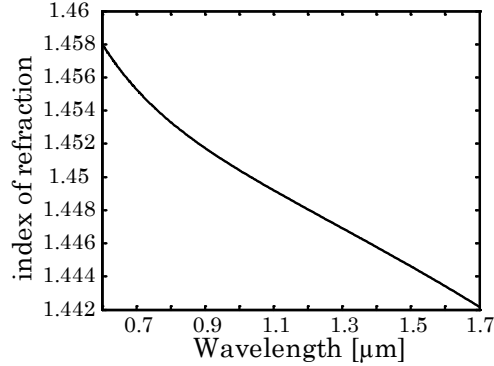


Figure 4-5: Wavelength-dependence of the index of refraction for undoped fused silica.

nonlinearity in optical fiber is the third order nonlinearity, $\chi^{(3)}$. Both dispersion and nonlinearity are weak effects, but due to the large distances involved in fiber transmission, they become influential and must be taken into consideration.

To calculate the propagation of pulses in fiber, we again return to the wave equation:

$$\nabla^2 E(\mathbf{r}, t) - \frac{1}{c^2} \frac{\partial E(\mathbf{r}, t)^2}{\partial^2 t} = \mu_o \frac{\partial P(\mathbf{r}, t)^2}{\partial^2 t}, \tag{4.1}$$

We first examine the effect of the $\chi^{(3)}$ nonlinearity on pulse propagation. We follow the development in Chapter 3 to arrive at

$$\nabla^2 E(\mathbf{r}, \omega) - \omega^2 \frac{1}{c^2} \left\{ 1 + \chi^{(1)}(\omega) + \frac{3}{4} \chi^{(3)}(\omega) |E(\mathbf{r}, \omega)|^2 \right\} E(\mathbf{r}, \omega) = 0, \tag{4.2}$$

where we have transformed the wave equation to the frequency domain. To do this, we assume the nonlinear effect is small and approximate it as a constant. We define our dielectric constant $\epsilon(\omega)$, where

$$\epsilon(\omega) = 1 + \chi^{(1)}(\omega) + \frac{3}{4} \chi^{(3)}(\omega) |E(\mathbf{r}, \omega)|^2. \tag{4.3}$$

This simplifies our wave equation to

$$\nabla^2 E(\mathbf{r}, \omega) - \epsilon(\omega)\frac{\omega^2}{c^2}E(\mathbf{r}, \omega) = 0. \tag{4.4}$$

We wish to calculate the effect of the nonlinear polarization in terms of the physical quantities of loss and the change in the index of refraction. In fiber, the intensity-dependence of loss (also known as two-photon absorption) is small, and can be neglected. We can write the dielectric constant as

$$\epsilon(\omega) = \left(n_o + n_2|E(\mathbf{r}, \omega)|^2 + i\frac{\alpha c}{2\omega}\right)^2. \tag{4.5}$$

Since the loss and intensity-dependent index of refraction is small compared with $n_o$, we can define a small perturbation $\Delta n$ as

$$\Delta n = n_2|E(\mathbf{r}, \omega)|^2 + i\frac{\alpha c}{2\omega} \tag{4.6}$$

and approximate $\epsilon(\omega)$ as

$$\epsilon(\omega) = (n_o + \Delta n)^2 \simeq n_o^2 + 2n_o\Delta n. \tag{4.7}$$

Remember from Chapter 3 that the background index of refraction is defined as:

$$n_o = \sqrt{1 + \mathrm{Re}\{\chi^{(1)}(\omega)\}}. \tag{4.8}$$

Equating Equations 4.3 and 4.7, we arrive at values for loss $\alpha$ and the intensity-dependent index of refraction $n_2$ in terms of the nonlinear polarization $\chi$.

$$\alpha = \frac{\omega}{n_o c}\mathrm{Im}\{\chi^{(1)}(\omega)\} \tag{4.9}$$

$$n_2 = \frac{3}{8n_o}\mathrm{Re}\{\chi^{(3)}(\omega)\} \tag{4.10}$$

To investigate the evolution of a pulse in this medium, we define our input as

129

a quasi-monochromatic electric field centered around $\omega_o$, where the transverse and longitudinal envelopes are separable. We multiply the slowly-varying envelopes by a monochromatic plane wave with wave vector $\beta_o$ and wavelength $\omega_o$:

$$E(x, y, z, t) = F(x, y)A(z, t)e^{i(\beta_o z - \omega_o t)}. \tag{4.11}$$

The Fourier transform gives us

$$E(x, y, z, \omega) = F(x, y)A(z, \omega - \omega_o)e^{i\beta_o z}. \tag{4.12}$$

If we use this form of the electric field and substitute it into the wave equation found in Equation 4.4, we obtain:

$$\frac{1}{F(x,y)}\left[\frac{\partial^2 F(x,y)}{\partial x^2} + \frac{\partial^2 F(x,y)}{\partial y^2} + \epsilon(\omega)\frac{\omega^2}{c^2}F(x,y)\right] = \tag{4.13}$$

$$-\frac{1}{A(z,\omega - \omega_o)}\left[\frac{\partial^2 A(z,\omega - \omega_o)}{\partial z^2} + 2i\beta_o\frac{\partial A(z,\omega - \omega_o)}{\partial z} - \beta_o^2 A(z,\omega - \omega_o)\right]. \tag{4.14}$$

We can solve this using the method of separation of variables, where we define $\beta^2$ as the separation constant. We shall see shortly that $\beta$ turns out to be the wave vector of the electric field within the fiber. We obtain equations for the transverse mode $F(x, y)$ and the propagating envelope $A(z, \omega - \omega_o)$.

$$\frac{\partial^2 F(x,y)}{\partial x^2} + \frac{\partial^2 F(x,y)}{\partial y^2} + \left(\epsilon(\omega)\frac{\omega^2}{c^2} - \beta^2\right)F(x,y) = 0 \tag{4.15}$$

$$2i\beta_o\frac{\partial A(z,\omega - \omega_o)}{\partial z} + (\beta^2 - \beta_o^2)A(z,\omega - \omega_o) = 0, \tag{4.16}$$

where we have assumed that the second derivative of the slowly-varying envelope $A(z, \omega - \omega_o)$ is much smaller than the first derivative and thus can be neglected. Equation 4.15 can be solved using perturbation theory to reveal the transverse modes of the fiber and the wave vector $\beta$, which we can then use to solve Equation 4.16 to describe the longitudinal propagation of the electric field with time. We first solve

130

Equation 4.15 in the absence of $\Delta n$ to obtain the unperturbed transverse mode in fiber and a corresponding unperturbed wave vector $\bar{\beta}$. These unperturbed transverse modes are most easily solved using a cylindrical coordinate system and are described by Bessel functions. A thorough derivation can be found in [6]. For single-mode fiber, only the lowest order mode is supported, and can be approximated as a Gaussian in the $x$ and $y$ directions:

$$F(x, y) \approx \exp\left(-\frac{(x^2 + y^2)}{s}\right). \tag{4.17}$$

In this Equation, $s$ is a free parameter for fitting the Gaussian approximation to the mode. Figure 4-6 shows a drawing of this solution.
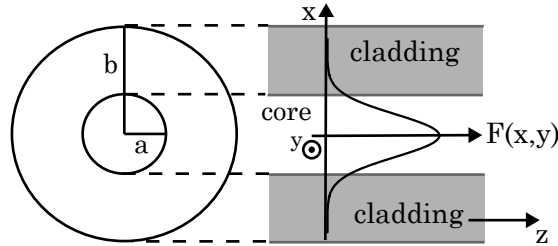


Figure 4-6: Drawing of the Gaussian approximation of the transverse fiber mode $F(x, y)$.

The next step is to include the perturbational effect of $\Delta n$ due to the nonlinearity and loss in the fiber. To the first order, this does not affect the transverse mode, but does affect the eigenvalue, or the wave vector, $\beta$. The solution becomes:

$$\beta = \bar{\beta} + \Delta\beta, \tag{4.18}$$

where

$$\Delta\beta = \frac{\frac{\omega}{c}\int\int \Delta n |F(x, y)|^2 dxdy}{\int\int |F(x, y)|^2 dxdy}. \tag{4.19}$$

131

We can evaluate this using our definition of $\Delta n$ from Equation 4.6:

$$\Delta\beta = \frac{\frac{\omega}{c}\int\int\left(i\frac{\alpha c}{2\omega} + n_2|E(\mathbf{r},\omega)|^2\right)|F(x,y)|^2 dxdy}{\int\int |F(x,y)|^2 dxdy}. \qquad (4.20)$$

This reduces to

$$\Delta\beta = i\frac{\alpha}{2} + \frac{\omega n_2}{cA_{eff}}|A(z,\omega - \omega_o)|^2, \qquad (4.21)$$

where we have defined

$$A_{eff} = \frac{\left(\int\int |F(x,y)|^2 dxdy\right)^2}{\int\int |F(x,y)|^4 dxdy}, \qquad (4.22)$$

such that $|A(z,\omega - \omega_o)|^2$ is normalized over the modal area to be equal to the optical power of the electric field.

We can now use our solution for $\beta$ to solve Equation 4.16 for the propagation of the electric field in fiber. We substitute Equation 4.15 into Equation 4.16 and assume that $\beta$ is close to $\beta_o$, meaning that the fiber propagation effects on the wave vector are small. This allows us to make the approximation:

$$\beta^2 - \beta_o^2 \simeq 2\beta_o\left(\beta - \beta_o\right). \qquad (4.23)$$

Using this, Equation 4.16 becomes:

$$2i\beta_o\frac{\partial A(z,\omega - \omega_o)}{\partial z} + 2\beta_o\left(\bar{\beta} + \Delta\beta - \beta_o\right)A(z,\omega - \omega_o) = 0. \qquad (4.24)$$

Substituting in our solution of $\Delta\beta$, we obtain

$$\frac{\partial A(z,\omega - \omega_o)}{\partial z} - i\left(\bar{\beta} - \beta_o\right)A(z,\omega - \omega_o) + \frac{\alpha}{2}A(z,\omega - \omega_o)$$
$$= i\gamma|A(z,\omega - \omega_o)|^2 A(z,\omega - \omega_o), \quad (4.25)$$

132

where we have defined a nonlinear coefficient $\gamma$, which describes the third order non-linearity in fiber, as

$$\gamma = \frac{\omega n_2}{c A_{eff}}. \tag{4.26}$$

Thus far, we have derived the propagation in fiber in the presence of the third-order nonlinearity. As mentioned earlier, the long transmission distances involved in fiber propagation results in linear dispersion playing a significant role. This dispersive effect shows up in the wave vector $\beta$, which is related to the index of refraction and thus also dependent on wavelength. We can include this effect by expanding the unperturbed, linear wave vector $\bar{\beta}$ in a Taylor series around $\omega_o$, the central frequency of the electric field:

$$\bar{\beta}(\omega) = \beta_o + \left.\frac{\partial \beta(\omega)}{\partial \omega}\right|_{\omega_o} (\omega - \omega_o) + \frac{1}{2} \left.\frac{\partial^2 \beta(\omega)}{\partial \omega^2}\right|_{\omega_o} (\omega - \omega_o)^2$$
$$+ \frac{1}{6} \left.\frac{\partial^3 \beta(\omega)}{\partial \omega^3}\right|_{\omega_o} (\omega - \omega_o)^3 + \ldots, \quad (4.27)$$

where the unperturbed wave vector $\bar{\beta}(\omega)$ evaluated at $\omega_o$ is simply the wave vector $\beta_o$ of the monochromatic plane wave. We define:

$$\beta_1 = \frac{1}{v_g} = \left.\frac{\partial \beta(\omega)}{\partial \omega}\right|_{\omega_o} \tag{4.28}$$

$$\beta_2 = \left.\frac{\partial^2 \beta(\omega)}{\partial \omega^2}\right|_{\omega_o}, \tag{4.29}$$

$$\beta_3 = \left.\frac{\partial^3 \beta(\omega)}{\partial \omega^3}\right|_{\omega_o}, \tag{4.30}$$

where $\beta_1$ describes the group velocity of the wavelengths, $\beta_2$ describes the change in the group velocity across different wavelengths, and $\beta_3$ is the third-order dispersion. For the distances and pulse-widths under investigation, the higher terms are negligible. Substituting our expansion of $\bar{\beta}$ into the propagation equation 4.25, we

133

obtain

$$\frac{\partial A(z, \omega - \omega_o)}{\partial z} - i \left( \beta_1(\omega - \omega_o) + \frac{1}{2}\beta_2(\omega - \omega_o)^2 + \frac{1}{6}\beta_3(\omega - \omega_o)^3 \right) A(z, \omega - \omega_o)$$
$$+ \frac{\alpha}{2} A(z, \omega - \omega_o) = i\gamma |A(z, \omega - \omega_o)|^2 A(z, \omega - \omega_o). \quad (4.31)$$

We ultimately wish to determine the propagation of the electric field in time. Taking the inverse Fourier transform, we move back into the time domain and obtain:

$$\frac{\partial A(z,t)}{\partial z} + \beta_1 \frac{\partial A(z,t)}{\partial t} + i\frac{1}{2}\beta_2 \frac{\partial^2 A(z,t)}{\partial t^2} - \frac{1}{6}\beta_3 \frac{\partial^3 A(z,t)}{\partial t^3} + \frac{\alpha}{2} A(z,t)$$
$$= i\gamma |A(z,t)|^2 A(z,t). \quad (4.32)$$

To simplify the solution, we transform this to a moving reference frame, where

$$\tau = t - \beta_1 z. \quad (4.33)$$

Thus, at any location along the fiber $z$, the center of the pulse remains at $\tau = 0$. We use the chain rule to calculate the derivatives:

$$\frac{\partial A(z,t)}{\partial z} = \frac{\partial A(z,\tau)}{\partial z}\frac{\partial z}{\partial z} + \frac{\partial A(z,\tau)}{\partial \tau}\frac{\partial \tau}{\partial z} \quad (4.34)$$
$$= \frac{\partial A(z,\tau)}{\partial z} - \beta_1 \frac{\partial A(z,\tau)}{\partial \tau}, \quad (4.35)$$

and

$$\frac{\partial A(z,t)}{\partial t} = \frac{\partial A(z,\tau)}{\partial \tau}\frac{\partial \tau}{\partial t} \quad (4.36)$$
$$= \frac{\partial A(z,\tau)}{\partial \tau}. \quad (4.37)$$

We see that Equation 4.36 is of a different form than the previous equation. This is due to the fact that in Equation 4.34, we are taking the partial derivative with respect to $z$ while treating $t$ as a constant. In writing the partial derivative in terms of $A(z, \tau)$, we introduce the variable $\tau$ which is not a constant and thus must be treated

134

as a variable. In contrast, in Equation 4.36, we are taking the partial derivative with respect to $t$, where $z$ is a constant. We write the partial derivative in terms of $A(z, \tau)$, where $z$ remains a constant and can thus be ignored. The derivation of higher order derivatives with respect to $t$ are similarly derived to obtain:

$$\frac{\partial A(z, \tau)}{\partial z} + i\frac{1}{2}\beta_2\frac{\partial^2 A(z, \tau)}{\partial \tau^2} - \frac{1}{6}\beta_3\frac{\partial^3 A(z, \tau)}{\partial \tau^3} + \frac{\alpha}{2}A(z, \tau) = i\gamma|A(z, \tau)|^2 A(z, \tau). \quad (4.38)$$

Equation 4.38 is known as the nonlinear Schroedinger's Equation. We have now derived the propagation of an optical pulse in the presence of dispersion and nonlinearity in fiber.

## 4.1.1 Split-Step Fourier Transform

We now wish to solve this nonlinear partial differential equation. To do this, we employ the numerical split-step Fourier method [6] and assume that the nonlinear and linear effects can be considered independently given a small enough length of fiber $\Delta z$. This allows us to consider the linear and nonlinear effects separately. The linear effects include dispersion and loss, while the nonlinear effect includes self-phase modulation:

$$\frac{\partial A(z, \tau)}{\partial z} = \left(\hat{L} + \hat{N}\right)A(z, \tau), \quad (4.39)$$

where

$$\hat{L} = -i\frac{1}{2}\beta_2\frac{\partial^2}{\partial \tau^2} + \frac{1}{6}\beta_3\frac{\partial^3}{\partial \tau^3} - \frac{\alpha}{2} \quad (4.40)$$

$$\hat{N} = i\gamma|A(z, \tau)|^2. \quad (4.41)$$

$\hat{L}$ and $\hat{N}$ are the linear and nonlinear operators, respectively. We divide the length of fiber into a sequence of small segments of length $\Delta z$. In each segment, we solve the linear portion of the NLSE for the first half. We include the nonlinear contribution throughout the segment as a lump contribution at the midpoint. Finally, we solve the linear portion for the second half of the segment. We repeat this procedure for each consecutive segment until we reach the distance of interest. Figure 4-7 visually illustrates this method. This is known as the symmetrized split-step Fourier transform

135
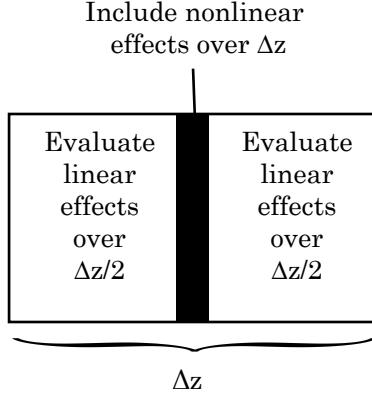
Figure 4-7: Diagram of an incremental segment of fiber $\Delta z$. In the first half of the segment, we calculate the linear effects of the length $\Delta z/2$. Then, at the midpoint, we include the nonlinear effects for the entire length $\Delta z$. Finally, we calculate the linear effects for the second half of the segment, $\Delta z/2$.

method, and results in a leading error term which is third-order in the step size $\Delta z$.

To solve the linear equation, we can take the Fourier transform and obtain

$$\frac{\partial A(z,\omega)}{\partial z} = i\frac{1}{2}\beta_2\omega^2 A(z,\omega) - i\frac{1}{6}\beta_3\omega^3 A(z,\omega) - \frac{\alpha}{2}A(z,\omega) \tag{4.42}$$

The solution to this is of the form $Ce^{gz}$, where

$$g = \frac{i}{2}\beta_2\omega^2 - \frac{i}{6}\beta_3\omega^3 - \frac{\alpha}{2}, \tag{4.43}$$

and $C$ is the initial pulse $A(z_o,\omega)$ in the frequency domain. Thus, to calculate the linear effects of fiber on a pulse over a distance $\Delta z/2$, we simply multiply the Fourier transform of the initial pulse by

$$\exp\left(\frac{i}{2}\beta_2\omega^2 - \frac{i}{6}\beta_3\omega^3 - \frac{\alpha}{2}\right)\frac{\Delta z}{2}. \tag{4.44}$$

Then, we take the inverse Fourier transform of the result to return it to the time domain:

$$A(z_o + \frac{\Delta z}{2}, \tau) = F^{-1}\left\{A(z_o,\omega)e^{(\frac{i}{2}\beta_2\omega^2 - \frac{i}{6}\beta_3\omega^3 - \frac{\alpha}{2})\frac{\Delta z}{2}}\right\}. \tag{4.45}$$

To solve the nonlinear portion, we drop the linear terms and are left with

$$\frac{\partial A(z,\tau)}{\partial z} = i\gamma |A(z,\tau)|^2 A(z,\tau).$$

(4.46)

Again, the solution to this is of the form $Ce^{hz}$, where

$$h = i\gamma |A|^2$$

(4.47)

and $C$ is the initial pulse $A(z_o, \tau)$ in the time domain. Thus, to calculate the nonlinear effects in fiber over a distance $\Delta z$, we simply multiply the initial pulse $A(z_o, \tau)$ (in the time domain) by $e^{(i\gamma |A|^2)L}$:

$$A(z + \Delta z) = A(z_o, \tau)e^{(i\gamma |A|^2)L}$$

(4.48)

We outline the entire process in Figure 4-8.
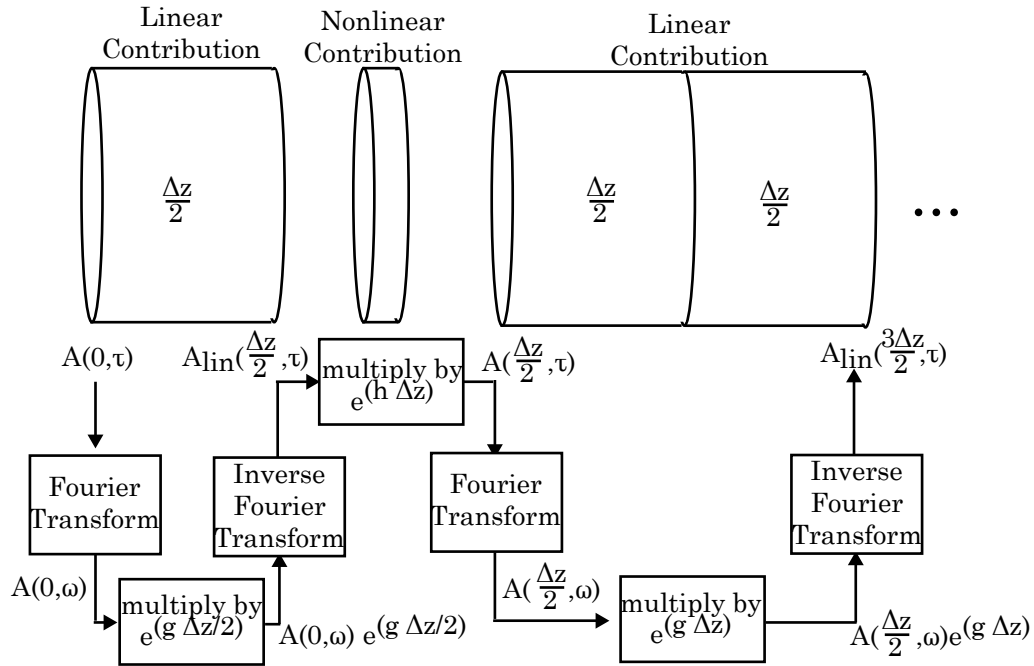


Figure 4-8: Diagram describing the split-step Fourier Transform method for solving the nonlinear Schroedinger equation.

Using this method, we can observe the effects of dispersion and self-phase modulation on fiber propagation. In Figure 4-9 we show the effects of chromatic dispersion, third-order dispersion, and self-phase modulation on a single optical pulse. Typical values for these effects at 1550 nm in single-mode fiber are: $\beta_2$ = -20 ps$^2$/km, $\beta_3$ = 0.1 ps$^3$/km, and $\gamma$ = 3 W$^{-1}$km$^{-1}$. Loss is typically 0.25 dB/km or less.

From Figure 4-9, we can see that as the transmission distance increases and pulse widths decrease, we observe more distortions in the pulses. First of all, $\beta_2$, the chromatic dispersion, is the strongest effect, and causes the pulse to spread in time. This is due to the fact that the different frequencies of the pulse are traveling at different velocities. As the pulse spreads, it can interfere with its adjacent pulses. As bit rates increase, we encounter two difficulties: first, the pulses arrive closer to each other, increasing the chance of interference between adjacent pulses. Second, shorter pulses are more susceptible to dispersion and spread faster. We can see this in the first column of Figure 4-9, where we have compared the dispersion between a 2-ps pulse with the dispersion of a 10-ps pulse across the same distance.

We next observe the effect of $\beta_3$, the third-order dispersion. This is a much weaker effect than chromatic dispersion, and thus we propagate for a significantly longer distance before observing its effects. We see that $\beta_3$ causes a ringing effect on the pulse, which can also cause inter-symbol interference. Again, the shorter 2-ps pulse is more strongly affected than the longer 10-ps pulse.

Finally, we observe the effect of the third-order nonlinearity, which causes self-phase modulation. This is a phase affect, and thus is only observed in the frequency domain. We see that this self-phase modulation causes a ringing and spreading in the spectral domain. Our use of wavelength-division multiplexing means that this broadening of the spectrum can cause interchannel interference. Since fiber has a third-order nonlinearity, four wave mixing can also be observed. Though we do not simulate it here, it can also cause significant interference in our wavelength channels. The strength of these nonlinear effects depends on the peak power of our pulses. A higher peak power results in a stronger nonlinear affect, as shown in Figure 4-9. As pulse widths decrease, we see a higher peak power for the same average power.
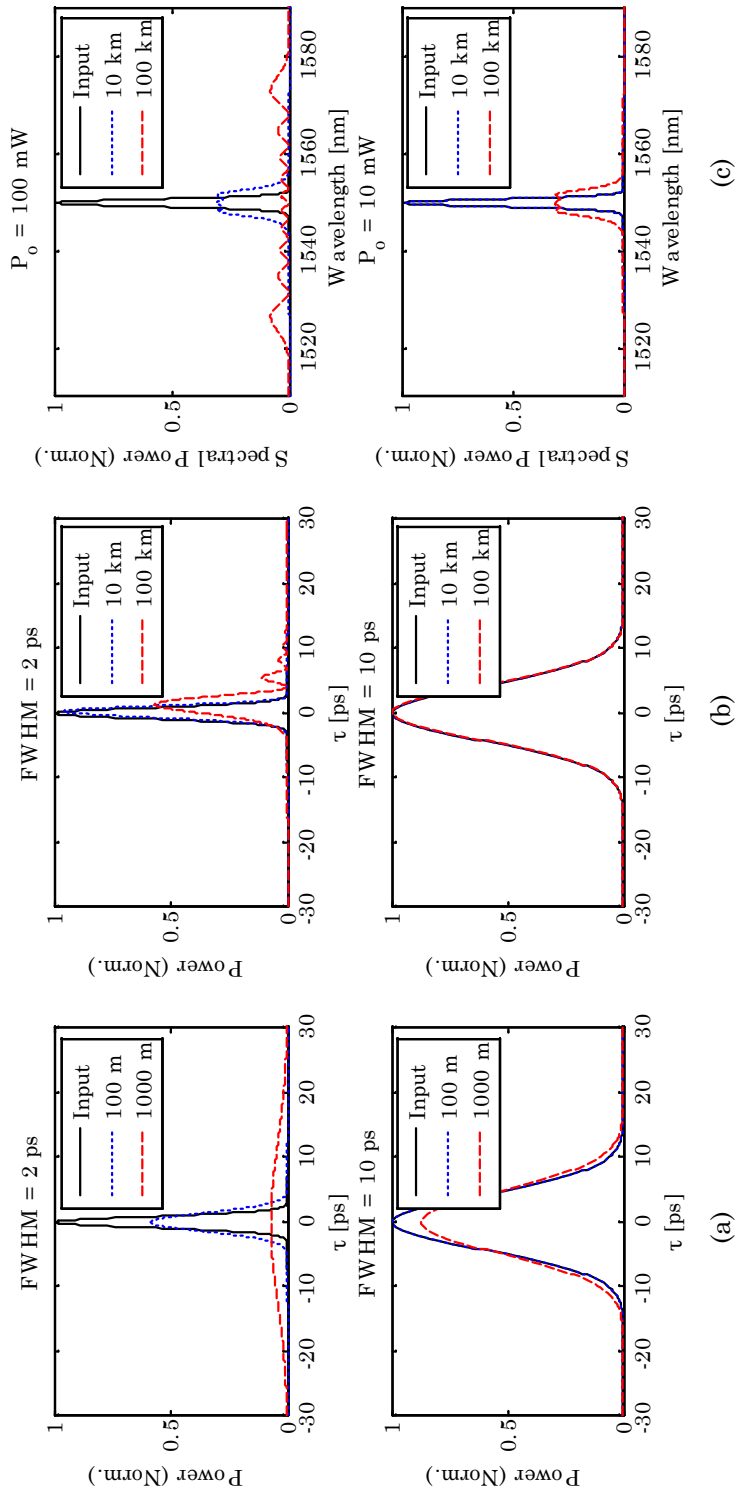
Figure 4-9: A set of plots showing the effects of (a) chromatic dispersion ($\beta_2$) in the first column, (b) third order dispersion ($\beta_3$) in the second column, and (c) self-phase modulation ($\gamma$) in the third column. We compare the magnitude of these effects between a pulse of 2 ps full-width half max (FWHM) in the upper row and a pulse of 10-ps FWHM in the lower row for dispersive effects. For the nonlinearity, we compare the effects of higher peak power (100 mW) and lower peak power (10 mW). It is evident that dispersive degradations increase with shorter pulses and longer transmission lengths. Nonlinear degradations increase with higher peak powers and longer transmission lengths.

The interaction of these separate effects with a data pattern leads to amplitude patterning and the appearance of power in between pulses. To illustrate the combined effects, we simulate a 100-Gb/s data stream with 2-ps pulses through two spans, including a chromatic dispersion module. The simulated span is shown in Figure 4-10. We choose fiber characteristics which roughly correspond to TrueWave XL fiber from
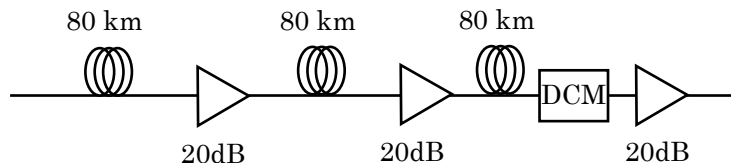


Figure 4-10: Schematic for simulation of fiber propagation effects on a data pattern in a span. DCM is the dispersion compensation module which exactly matches second-order dispersion without contributing to other effects.

OFS [7]. We choose the following values: $\beta_2 = 5$ ps$^2$/km, $\beta_3 = 0.014$ ps$^3$/km, $\gamma = 3$ W/km, and loss is 0.25 dB/km. Each span involves 160 km of fiber, with EDFAs spaced 80 km apart. After 80 km, we see a 20 dB loss, which the EDFAs will compensate for. We do not simulate the effects of EDFA noise on pulse propagation. Our dispersion-compensation module is an ideal compensator which exactly matches the second-order dispersion in the preceding lengths without contributing to other effects. The results of this simulation are shown in Figure 4-11. We see that the interaction of nonlinear and linear effects in fiber result in pulse-to-pulse amplitude variations as well as pulse width variations. Furthermore, optical power begins to show up between pulses, reducing extinction. As seen earlier, these effects are more pronounced at higher data rates with shorter pulses. With the 100-Gb/s data rate, without regeneration, significant distortions are introduced after only 360 km of propagation. This provides a pressure to create more effective and cheaper regenerators in order to compensate for increasing need for regeneration. As described earlier, all-optical regenerators offer the potential for reducing costly O/E/O conversions and simplifying regenerator architecture. In the next section, we briefly provide an overview of the necessary components of all-optical regeneration.
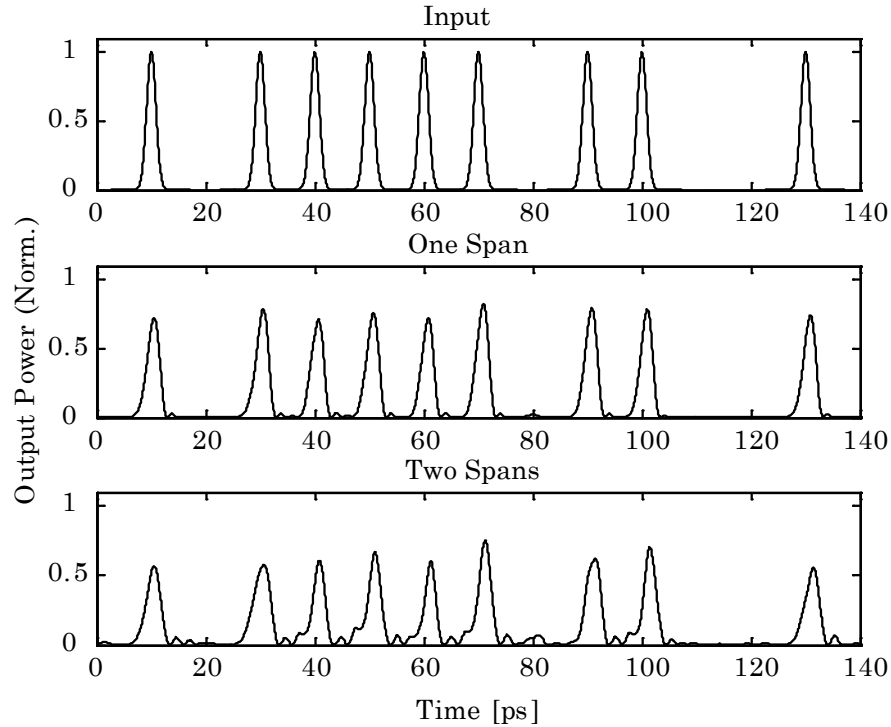
Figure 4-11: Simulation of data pattern propagation after one and two spans of fiber with amplification and second-order dispersion compensation.

## 4.2 All-Optical Regeneration

Traditionally, there are three aspects to regeneration of data signals: re-amplification, re-timing, and re-shaping [8]. Re-amplification is required to compensate for variations in loss or gain between packets and pulses within a packet. Different packets take different routes through the network and experience different amounts of loss, gain, and noise. Within a packet, different pulses can experience differing effects due to pattern-dependent interference from the interaction of nonlinearities and dispersion in fiber, as seen in the previous section. In order to reduce these amplitude variations, we require a thresholding power transfer function such as the one shown in Figure 4-12. As can be seen in the figure, this transfer function will reduce the variation in the power of both the zero bits and the one bits ($\Delta P_0$ and $\Delta P_1$, respectively) of the incoming data stream. One example of this type of transfer function is
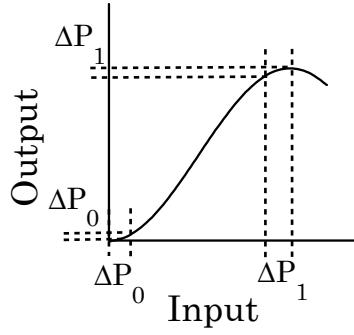
Figure 4-12: Power transfer function for re-amplification in a regenerator. $\Delta P_0$ refers to the variation in the power of the zero bits of the data stream and $\Delta P_1$ refers to the variation in the power of the one bits. This sort of thresholding transfer function will reduce the amplitude variation.

the raised sinusoid, as generated by a simple delayed interferometer:

$$E_{out} = E_{in} + E_{in}e^{i \cdot n \cdot L}, \tag{4.49}$$

where $n$ is the index of refraction of the material and L is the additional length in one arm of the interferometer. This is equivalent to

$$E_{out} = E_{in}e^{i \cdot n \cdot \frac{L}{2}} 2\cos\left(\frac{nL}{2}\right). \tag{4.50}$$

This power transfer function is simply a squared cosine.

The power transfer function of an SOA can also reduce amplitude patterning in the ones through the saturation effects in the SOA. We can see this by looking at the steady-state operation of an SOA. We return to Equation 3.35, derived in Chapter 3, which describes the gain evolution:

$$\frac{\partial g(z,\tau)}{\partial \tau} = \frac{g_o - g(z,\tau)}{\tau_c} - \frac{g(z,\tau)P(z,\tau)}{E_{sat}}. \tag{4.51}$$

In the steady state, $\dfrac{\partial g(z,\tau)}{\partial \tau} = 0$, and the steady state gain $g_s$ becomes

$$g_s(z) = \frac{g_o}{1 + P_{in}/P_{sat}}, \tag{4.52}$$

where $P_{sat} = E_{sat}/\tau_c$. If $P_{in} \ll P_{sat}$, the gain approaches the unsaturated gain $g_o$. When the input power is the same as $P_{sat}$, the gain is cut in half. As the input power increases, the gain continues to decrease. We can calculate the power transfer function by multiplying the gain by the input power:

$$P_{out} = g_s P_{in} = \frac{g_o P_{in}}{1 + P_{in}/P_{sat}}. \tag{4.53}$$

This transfer function is plotted in Figure 4-13, where we see that it will result in a reduction of variation in the ones, but not the zeros. In fact, we see that SOAs can increase the variation in the amplitude of the zeros. Thus, a single SOA will be insufficient for re-amplification on its own, but may be useful when paired with other devices.
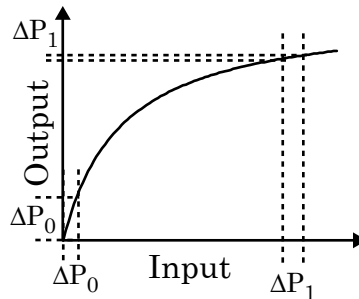


Figure 4-13: Power transfer function for a single SOA. Note that the SOA can reduce variation in the one bits, but increases amplitude variations in the zeros.

The second aspect of regeneration is re-timing, which is necessary to compensate for variations in pulse arrival times, known as timing jitter. As the pulses propagate, amplified spontaneous emission from the EDFAs can nonlinearly interact with the pulses to cause varying frequency shifts and thus introduce timing jitter. In order to re-time pulses, we require a local, undistorted clock. Since we wish to operate our network asynchronously, this means that clock recovery techniques are required. The simplest all-optical clock recovery technique relies on using a Fabry-Perot cavity followed by an SOA [9] as shown in Figure 4-14. The length of the Fabry-Perot is designed to resonate at the repetition rate of the incoming data. Given one incoming
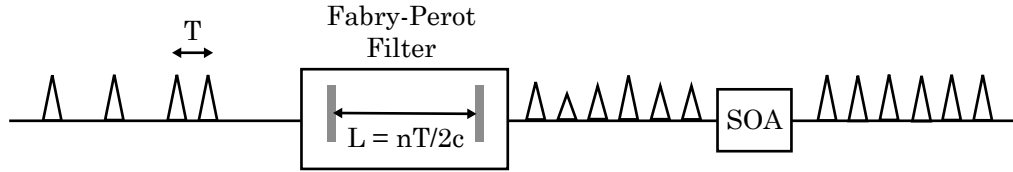
Figure 4-14: Clock recovery using a Fabry-Perot filter. The length of the resonating cavity is designed to be $L = \frac{n}{2c}T$, where $n$ is the index of refraction of the cavity, $c$ is the speed of light, and $T$ is the period of the incoming pulse train. The SOA reduces amplitude variation through saturation.

pulse $E_{in}$, the output will be a pulse at time $t_o$, and a copy of that pulse at time $3t_o$, followed by another copy at time $5t_o$. Each successive copy of the pulse contains less power due to loss and less than perfect reflectivity in the mirrors. Thus, the ones in the incoming data pattern provide the pulse energy to fill in the zeros. At the output of the Fabry-Perot, there can be significant amplitude variation in the pulses. An SOA or other thresholding switch can be used to smooth out these variations. This method has been demonstrated at clock speeds up to 40 Gb/s [10]

Ideally, the clock should be recovered with low amplitude variation and with a minimum of delay. Typically, it takes time before the amplitude of the clock grows to its steady-state value. This is known as the rise time or locking time of the clock recovery. Ideally, this time should be short, as it determines the delay a packet will experience in the regenerator. The amplitude variation and locking time are related through the finesse of the Fabry-Perot [11]. A high finesse means that more energy is contained in the Fabry-Perot cavity. This leads to a slow recovery time, but less amplitude variation in the steady state.

Another key design parameter of the Fabry-Perot filter is the cavity length. The cavity length must be well-controlled to match the repetition rate of the incoming pulse train and generate an accurate clock. This implies that the cavity must be designed to a particular bit rate, thus reducing the transparency of the network. Tuning of the cavity length through temperature or mechanical methods can allow us to accommodate different bit rates with the same Fabry-Perot, but this adds significant complexity to the system.

144

Once we have generated a local clock, we can re-time our incoming data pattern. Our method of accomplishing this is to use a single all-optical logic gate in which our data inputs are the control and our clock inputs are the signal (Figure 4-15). As long as the switch is insensitive to timing jitter on the data input port, the output timing jitter will be reduced. Figure 4-15 shows a wavelength conversion as part
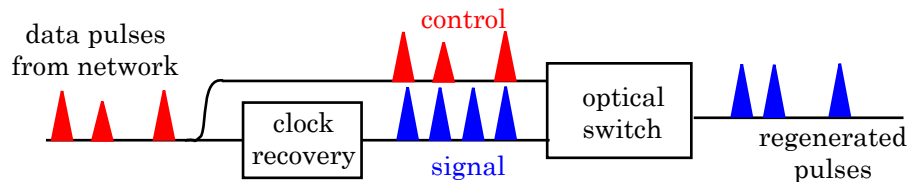


Figure 4-15: All-optical 3R regeneration using an all-optical switch. The data pulses act as the control input, modulating the appropriate data on the locally-generated clock pulses.

of the regeneration. This is generally undesired, but can be easily avoided through either counter-propagating operation of the all-optical switch or through a simple wavelength conversion.

With this method, we also simultaneously re-amplify and re-shape the incoming pulses, since we have completely replaced our input pulses with new, undistorted clock pulses. Re-shaping is necessary to compensate for dispersion and chirping effects as described in Section 4.1. Though we focus on using all-optical switching techniques for regeneration in this section, there are several other techniques for accomplishing this 3R regeneration, most notably synchronous modulation [12].

A regenerator capable of accomplishing all three tasks (re-amplification, re-timing, and re-shaping) is known as a 3R regenerator. S. J. Savage et al. [13] introduces the idea of a 4R regenerator, suggesting the need for re-polarization as well as re-amplification, re-timing, and re-shaping. Many all-optical switches are polarization sensitive, and varying polarizations at the input can be translated into amplitude patterning at the output. In order to provide re-polarization, S. J. Savage includes a polarization-insensitive front-end which replaces the pulses with pulses of a single polarization. In the next section, we describe our collaboration with S. J. Savage, where we build upon his previous work with 4R regeneration to demonstrate the

utility of our integrated logic gate and the effectiveness of our optimization technique for developing all-optical regenerators.

## 4.3   All-Optical Regeneration Using the SOA-MZI

Ideally, the test of regenerator performance would occur in a real system, involving many regeneration spans and packets from multiple locations with differing distortions. As this involves a large number of fiber and amplifiers, it is not feasible in a small laboratory setting. Instead, we use a recirculating loop to simulate conditions over many regenerator spans. We implement one regenerator span and then recirculate the pulses through the span to simulate transmission over multiple spans.

In this experiment, our goal is to demonstrate the effectiveness of the SOA-MZI integrated logic gate for improving transmission in data networks by providing a simple regenerator that can operate at line rates. This reduces the need for O/E/O conversions and can reduce the need for parallel architectures caused by limited electronic data rates. Furthermore, we wish to demonstrate the effectiveness of the optimization method developed in Chapter 3 and investigate the cascaded operation of the logic gate. Regeneration using SOA-MZI logic gates has been demonstrated previously in recirculating loops using a pair of inverting gates at speeds up to 40 Gb/s [14] and distances up to 400,000 km [15, 16].

Here, we demonstrate successful 3R regeneration at 10 Gb/s in a single non-inverting SOA-MZI over a 100-km recirculating loop [17] with a total distance of 10,000 km. By combining the XPM SOA wavelength converter described in Chapter 3 with the SOA-MZI, we achieve re-polarization as well. This also allows us to use the SOA-MZI in the faster co-propagating configuration. To illustrate the accuracy of this regeneration, we present bit error rate measurements directly after the regenerator (0 passes through the loop) and after 10 and 100 passes in the loop, showing less than 0.5-dB penalty [18]. This indicates the effectiveness of the SOA-MZI for regeneration and for cascaded operation.

## 4.3.1 Experimental Setup

The experimental setup is shown in Figure 4-16. We generate a data pulse pattern
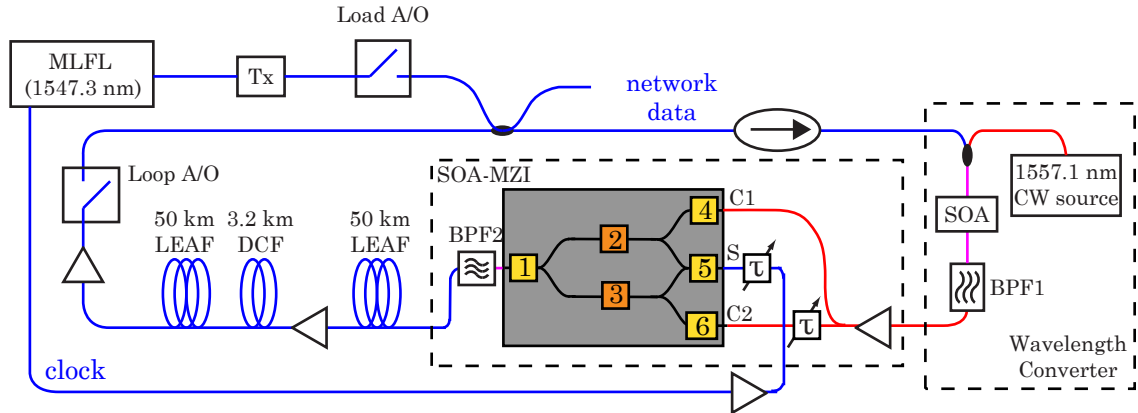


Figure 4-16: Experimental schematic of the recirculating fiber loop with an SOA-MZI optical regenerator. MLFL is mode-locked fiber laser, BPF stands for bandpass filter, A/O is acousto-optic modulator, C1 and C2 are the first and second control inputs to the SOA-MZI, and S is the signal input. The numbered boxes represent SOAs within the SOA-MZI device.

using a mode-locked fiber laser followed by our transmitter. The fiber laser generates 3-ps pulses at 1547.3 nm with a repetition rate of 10 Gb/s. The transmitter consists of a pattern generator and an electro-optic modulator which modulates a $2^{31}-1$ pseudorandom bit sequence onto the clock pulses generated by the mode-locked laser. These pulses are loaded into the loop through an acousto-optic modulator. An isolator prevents reflections in the loop. The pulses are then regenerated by the combination of the SOA-based cross-phase modulation wavelength converter followed by the SOA-MZI optical logic gate. The wavelength converter uses a CW source at 1557.1 nm. Cross-phase modulation in the SOA causes a broadening in the CW line. The wavelength-converted pulses are recovered by a combination of a notch filter, which filters out the CW line, and a band-pass filter at the long-wavelength side. The output spectrum of the wavelength converter is shown in Figure 4-17(a). We also present the bit error rate plot of the XPM wavelength converter as compared with the transmitter output (back-to-back test) in Figure 4-17(b). We see only about a 1.5 dB penalty at $10^{-9}$. The output of the SOA consists of slightly broadened data pulses centered at 1557.5 nm. This broadening is advantageous, as it reduces the sensitivity of this

regenerator to timing jitter.
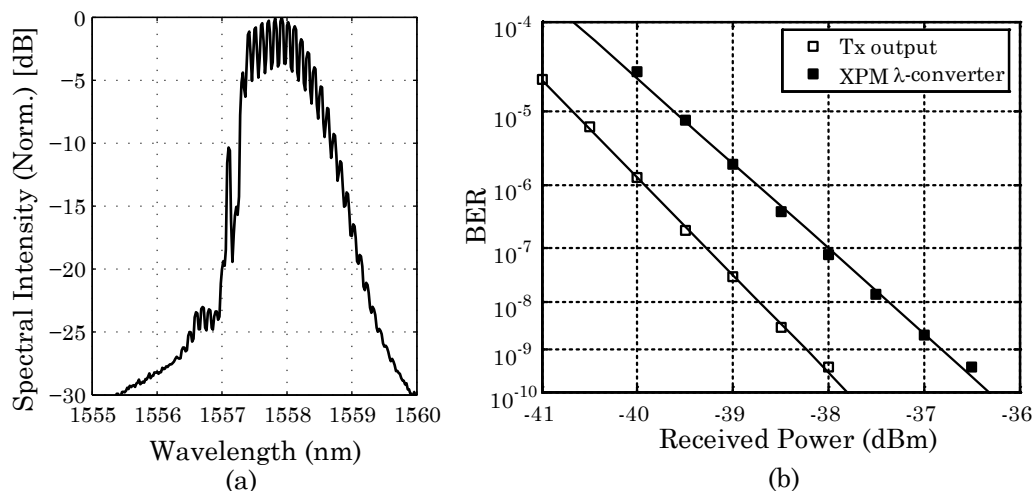


Figure 4-17: (a) Output spectrum of the XPM SOA wavelength converter. (b) BER plot showing only a 1.5 dB penalty at $10^{-9}$.

These wavelength-converted pulses are then used as the control inputs to the SOA-MZI, which provides the 3R regeneration. In this experiment, we focus on evaluating the performance of the SOA-MZI as a regenerator, and thus do not implement clock recovery. Instead, we use the same mode-locked fiber laser output before modulation as our clock input, which provides a perfect clock to our regenerator. A clock generated through clock recovery methods will likely degrade performance slightly. We send this clock into the signal port of the switch. We operate the SOA-MZI in differential mode and adjust the timing between the two control inputs with the use of a manual delay. As described in Chapter 3, a pulse entering the upper control input (C1) of the SOA-MZI pushes the interferometer out of extinction, allowing the signal pulse to pass. A second copy of the control pulse enters the lower control input (C2) after the signal pulse, pulling it back into extinction and eliminating the effect of the long SOA carrier-recovery time. The timing between the signal clock pulses and the data pulses is tuned using a second variable delay line. We use the dynamic bias scan technique described in Chapter 3 to discover the optimal switching point for the SOA-MZI at the powers used in the recirculating loop. We find the optimal switching

currents to be $I_2 = 692.3$ mA and $I_3 = 708.1$ mA. The pre- and post-amplifying SOA currents are set to $I_5 = 81.4$ mA (input), $I_4 = 107.8$ mA (C1), $I_6 = 107.8$ mA (C2), and $I_1 = 150.3$ mA (output). The input signal power is 3.5 dBm (448 fJ/bit) and the control input powers are 3.2 dBm (418 fJ/bit) for C1 and -2.9dBm (103 fJ/bit) for C2. In differential operation, a lower C2 input power is necessary to compensate for the recovery of the SOA during the delay between C1 and C2 pulses. After passing through the wavelength converter and the SOA-MZI, we have now regenerated our data pulses by replacing the distorted pulses with undistorted clock pulses at the same wavelength modulated with the correct pattern.

These regenerated pulses then pass through the fiber span, which consists of two 50-km sections of LEAF (Large Effective Area Fiber) [19] separated by EDFAs. The characteristics of LEAF fiber are similar to the TrueWave fiber simulated earlier: at 1547.3 nm, $\beta_2 = 4.7$ ps$^2$/km and $A_{eff} = 72$ $\mu$m$^2$. Dispersion compensation is accomplished to second-order with 3.2 km of dispersion-compensating fiber (DCF), which has a $\beta_2$ of -161.3 ps$^2$/km. A second acousto-optic modulator is used to empty the loop after a certain number of passes. This loop modulator opens when the load modulator closes, which loads the data pattern into the loop and simultaneously flushes any previous data in the loop. Then, after the pattern is loaded, the loop modulator closes and the load modulator opens. We allow the pulses in the loop to circulate for a set number of passes, and then repeat the cycle. At the output of the regenerator, we use a gated bit error rate detector and a cross correlator to test the data pattern after a certain number of passes.

## 4.3.2 Results

We use a cross-correlator to look at the evolution of the pulses over multiple passes in the loop. We sample the data pattern with clock pulses from the mode-locked fiber laser which have been compressed to about 150 fs. This allows us to observe asymmetries and distortions to the pulse shape. Figure 4-18 shows the cross-correlation with and without the regenerator. From this comparison, we can see how critical regeneration is to fiber transmission for short pulses. By the 36th pass (roughly 3,600
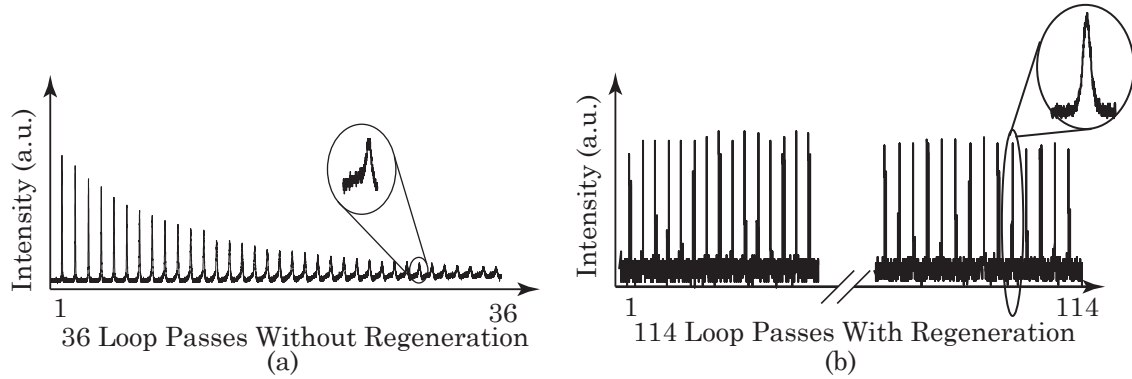
Figure 4-18: Cross-correlations of the loop output (a) without regeneration and (b) with regeneration.

km), despite compensating for loss, the pulses have experienced significant dispersion and can no longer be detected. We show a detail of the pulses in the 29th pass and see significant asymmetry due to third-order dispersion. In contrast, with the regenerator in the loop, we see that the pulses remain short and undistorted even by the 114th pulse.

To verify accurate regeneration, we perform a bit error rate test. We gate our error detector such that it will record the errors for a single pass. This significantly increases the amount of time it takes for a full measurement. In un-gated operation, at 10 Gb/s, the error detector receives 10 Gb of data each second. If we assume that the data has an error rate of $10^{-9}$ errors/second, we will detect ten errors every second. We thus obtain a statistically significant number of errors quickly. In a gated measurement, to measure the errors on a single pass through the loop, we have significantly less data to measure. For instance, in order to test the tenth pass, we must allow the pulses to circulate for 10 passes and measure the errors on every tenth pass. Thus, in the same amount of time, we only obtain 1/10 of the data as an un-gated measurement. As the number of passes increase, it becomes more and more difficult to obtain the same accuracy. Thus, by the 100th pass, we are only able to obtain data down to an error rate of $10^{-8}$. Figure 4-19 shows the bit error rate test comparing the output of the regenerator without any fiber degradations with the bit error rate plots from the 10th and 100th pass. From this, we see that we achieved only
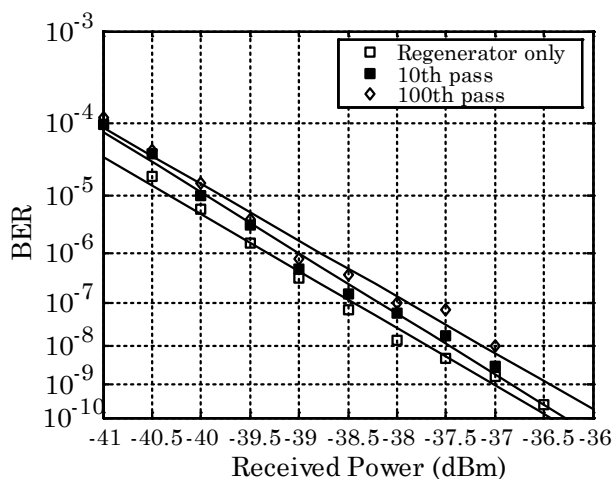
Figure 4-19: BER plots for the regenerator output at the 10th and 100th pass compared with the output of the regenerator alone.

a 0.5-dB penalty after 100 passes at a BER of $10^{-9}$, showing successful regeneration over 10,000 km.

## 4.4    Conclusion

In this chapter, we have demonstrated the capability of the SOA-MZI to successfully provide stable, simple regeneration for fiber transmission of ultrashort pulses over a distance of 10,000 km. The demonstration of the SOA-MZI in a recirculating loop indicates its potential as a key building block for all-optical processing in future ultrafast data networks. Using a single all-optical switch for ultrafast regeneration can lead to reduced costs for regeneration in data networks.

However, for future multi-gate processing needs, the SOA-MZI as currently designed suffers from the need for additional amplification to compensate for high losses in the waveguide. This significantly increases the power consumption of the gates. Furthermore, additional noise from the amplifiers necessitates filtering and degrades performance. In an on-going collaboration with the Integrated Photonic Materials and Devices Group on MIT campus, we investigated improvements to the design and

fabrication of the SOA-MZI switch. In the next chapter, we present some preliminary results from the testing of the asymmetric twin waveguide SOA-MZI switches and discuss future directions for fabrication and integration of multi-gate SOA-MZI switches.

# Bibliography

[1] B. Bakhshi, M. Manna, G. Mohs, D. I. Kovsh, R. Lynch, M. Vaa, E. Golovchenko, W. W. Patterson, W. T. Anderson, P. Corbett, S. Jiang, M. M. Sanders, H. Li, G. T. Harvey, A. Lucero, and S. M. Abbott, "First dispersion-flattened transpacific undersea system: From design to terabit/s field trial," *Journal of Lightwave Technology*, vol. 22, no. 1, pp. 233–241, 2004.

[2] J.-X. Cai, D. G. Foursa, L. Liu, C. R. Davidson, Y. Cai, W. W. Patterson, A. J. Lucero, B. Bakhshi, G. Mohs, P. C. Corbett, V. Gupta, W. Anderson, M. Vaa, G. Domagala, M. Mazurczyk, H. Li, S. Jiang, M. Nissov, A. N. Pilipetskii, and N. Bergano, "RZ-DPSK field trial over 13 100 km of installed non-slope-matched submarine fibers," *Journal of Lightwave Technology*, vol. 23, no. 1, pp. 95–103, 2005.

[3] C. Rasmussen, T. Fjelde, J. Bennike, F. Liu, S. Dey, B. Mikkelsen, P. Mamyshev, P. Serbe, P. van der Wagt, Y. Akasaka, D. Harris, D. Gapontsev, V. Ivshin, and P. Reeves-Hall, "DWDM 40G transmission over trans-pacific distance (10 000 km) using CSRZ-DPSK, enhanced FEC, and all-raman-amplified 100-km UltraWave fiber spans," *Journal of Lightwave Technology*, vol. 22, no. 1, pp. 203–207, 2004.

[4] M. Vasilyev, T. I. Lakoba, and P. G. Patki, "Multi-wavelength all-optical regeneration," in *Optical Fiber Communications Conference (OFC '08)*, 2008.

[5] I. H. Malitson, "Interspecimen comparison of the refractive index of fused silica," *Journal of the Optical Society of America*, vol. 55, no. 10, pp. 1205–1209, 1965.

[6] G. P. Agrawal, *Nonlinear Fiber Optics*. Academic Press, second ed., 1995.

[7] "Truewave XL ocean fiber product description." `www.ofsoptics.com/resources/TrueWaveXLOceanFiber-122.pdf`.

[8] O. Leclerc, B. Lavigne, E. Balmefrezol, P. Brindel, L. Pierre, D. Rouvillain, and F. Seguineou, "Optical regeneration at 40 Gb/s and beyond," *Journal of Lightwave Technology*, vol. 21, no. 11, pp. 2779–2790, 2003.

[9] G. Contestabile, A. D'Errico, M. Presi, and E. Ciaramella, "40-GHz all-optical clock extraction using a semiconductor- assisted fabry-perot filter," *IEEE Photonics Technology Letters*, vol. 16, no. 11, pp. 2523–2525, 2004.

[10] E. Kehayas, L. Stampoulidis, H. Avramopoulos, Y. Liu, E. Tangdiongga, and H. J. S. Dorren, "40 Gb/s all-optical packet clock recovery with ultrafast lock-in time and low inter-packet guardbands," *Optics Express*, vol. 13, no. 2, pp. 475–480, 2005.

[11] X. Zhou, C. Lu, P. Shum, H. H. M. Shalaby, T. H. Cheng, and P. Ye, "A performance analysis of an all-optical clock extraction circuit based on fabry-perot filter," *Journal of Lightwave Technology*, vol. 19, no. 5, pp. 603–613, 2001.

[12] G. Raybon, Y. Su, J. Leuthold, R.-J. Essiambre, T. Her, C. Joergensen, P. Stein-verzel, K. Dreyer, and K. Feder, "40 Gbit/s pseudo-linear transmission over one million kilometers," in *Optical Fiber Communications Conference (OFC '02)*, 2002.

[13] S. J. Savage, B. S. Robinson, S. A. Hamilton, and E. P. Ippen, "Wavelength-maintaining polarization-insensitive all-optical 3R regenerator," *Optics Express*, vol. 14, no. 5, pp. 1748–1754, 2006.

[14] B. Lavigne, P. Guerber, P. Brindel, E. Balmefrezol, and B. Dagens, "Cascade of 100 optical 3R regenerators at 40 Gbit/s based on all-active Mach-Zehnder interferometers," in *Proceedings of the 27th European Conference on Optical Communications (ECOC '01)*, p. 290, 2001.

[15] M. Funabashi, Z. Zhu, Z. Pan, S. J. B. Yoo, and L. Paraschis, "All-optical 3R regeneration in monolithic SOA-MZI to achieve 0.4 million km fiber transmission," in *The 18th Annual Meeting of the IEEE Lasers and Electro-Optics Society (LEOS '04)*, 2005.

[16] Z. Zhu, M. Funabashi, Z. Pan, L. Paraschis, and S. J. B. Yoo, "1000 cascaded stages of optical 3R regeneration with soa-mzi- based clock-enhancement to achieve 10-Gb/s 125,000-km dispersion uncompensated transmission," *IEEE Photonic Technology Letters*, vol. 18, no. 20, pp. 2159–2161, 2006.

[17] S. J. Savage, B. S. Robinson, S. A. Hamilton, and E. P. Ippen, "200-Pass picosecond-pulse transmission through a regenerative recirculating fiber loop," in *Conference on Lasers and Electro-Optics (CLEO '06)*, 2006.

[18] J. P. Wang, B. S. Robinson, S. J. Savage, S. A. Hamilton, E. P. Ippen, R. Mu, H. Wang, L. Spiekman, and B. B. Stefanov, "Regeneration using an SOA-MZI in a 100-pass 10,000-km recirculating fiber loop," in *Conference on Lasers and Electro-Optics (CLEO)*, 2007.

[19] "Corning LEAF optical fiber product information." `www.corning.com/docs/opticalfiber/pi1107.pdf`.

# Chapter 5

# Designing Integrated Logic Gates for Multi-Gate Processing

In previous chapters, we discussed and demonstrated some of the advantages offered by all-optical signal processing. However, there are also challenges which must be overcome before all-optical processing can be adopted into existing data networks. First of all, cost must come down. Integration and mass-production will be required to reduce the cost of a single all-optical switch. Secondly, optical logic gates must be designed to be simple to operate and optimize. This reduces labor costs associated with installing and maintaining these switches in the field. Third, with power consumption a key constraint for future electronic routers [1, 2], minimizing the power consumption of all-optical switches is required for a truly compelling argument for the use of all-optical processing in future data networks. Finally, cascaded multi-gate processing must be demonstrated. Data networks will require more complex processing than can be accomplished with a single logic gate. To that end, integrating multiple gates on a single chip will be critical.

In designing integrated logic gates, all these considerations must be taken into account. As described in previous chapters, the SOA-MZI logic gate is currently the most promising solution for practical all-optical signal processing. It is an integrated logic gate with conceptually straightforward operation. Furthermore, our optimization method developed in Chapter 3 allows efficient optimization of individual logic

gates. However, our SOA-MZI gates currently require additional amplifiers to compensate for waveguide losses. This increases power consumption per gate and can degrade performance for cascaded gates. To address this issue, we investigated alternative integration techniques for lower losses. The capability to design and fabricate our own gates will allow us to demonstrate integrated, multi-gate logic and to optimize SOA design for low-power, ultrafast nonlinear switching. We collaborated with the Integrated Photonic Materials and Devices Group at MIT to design and fabricate low-loss SOA-MZI logic gates. The ultimate goal of our collaboration is the capability to design and fabricate all-optical chips for more complex, multi-gate all-optical processing.

In this chapter, we describe some potential integration methods for reducing loss in SOA-MZI gates and give background on our choice of the asymmetric twin waveguide design. We finish by describing our efforts characterizing the fabricated chips and provide recommendations for future chip design.

## 5.1  Integration Techniques for the SOA-MZI

In considering the fabrication of the SOA-MZI, we note that this logic gate consists of two types of components: active and passive. The active components, the SOAs, are required to produce the nonlinear effect required for switching. The passive components, which consist of the couplers and waveguides, guide light and form the interferometer. Fundamentally, the requirements for passive and active materials are very different. Active materials should possess a high nonlinearity, which means we must operate at photon energies above the bandgap of the material to maximize the interaction between photons and electrons. Passive materials, however, should be highly linear and possess low loss, which implies operating at energies far below the material bandgap.

Integration techniques for the SOA-MZI can be split into two broad categories: hybrid and monolithic integration. Hybrid integration refers to integration of two incompatible materials in a single device. This incompatibility usually results from a

mismatch in the lattice constant, temperature coefficient, or both. Thus, these materials cannot be grown on the same substrate or fabricated using the same processes. The advantage of hybrid integration is that we can use the best passive materials for the passive components in the SOA-MZI while choosing the best active materials for the active components. The challenge then arises from the problem of aligning the passive and active devices. Furthermore, because of additional fabrication needs, hybrid devices are generally more complex and costly to build.

In contrast, monolithic integration uses compatible materials for both active and passive components. Different doping profiles are used to tailor the material characteristics. Figure 5-1 shows the compatibilities of different semiconductor materials used in optical devices and the wavelengths they operate at. Monolithic integration results in a much simpler fabrication process and fewer alignment steps, but limits our choice for active and passive materials. Currently, monolithic integration has focused on using InP-based (indium phosphide) materials for better active-material performance.

### 5.1.1 Hybrid Integration

The two most promising materials for passive components are silicon and silica-on-silicon. The advantage of silicon lies in its compatibility with electronic processing techniques. Silicon also possesses a high confinement factor due to its high index of refraction ($\sim3.5$) compared with silica ($\sim1.5$). This high confinement factor allows tight bends and thus reduces the footprint of photonic devices. Waveguides in silicon also offer low propagation losses (0.6 dB/cm or less [4, 5]) due to the large bandgap of silicon at optical wavelengths. However, coupling between silicon waveguides and fiber as well as active material waveguides is difficult due to the differences in the mode size, which results from the high confinement in silicon waveguides. To compensate for this, mode converters are required, which adds complexity to the design.

Silica-on-silicon techniques address this issue by using silica glass fabricated on a silicon substrate. This naturally results in waveguide modes which are well-matched to fiber modes. Furthermore, losses are much lower than in silicon ($< 0.01$ dB/cm) [6].
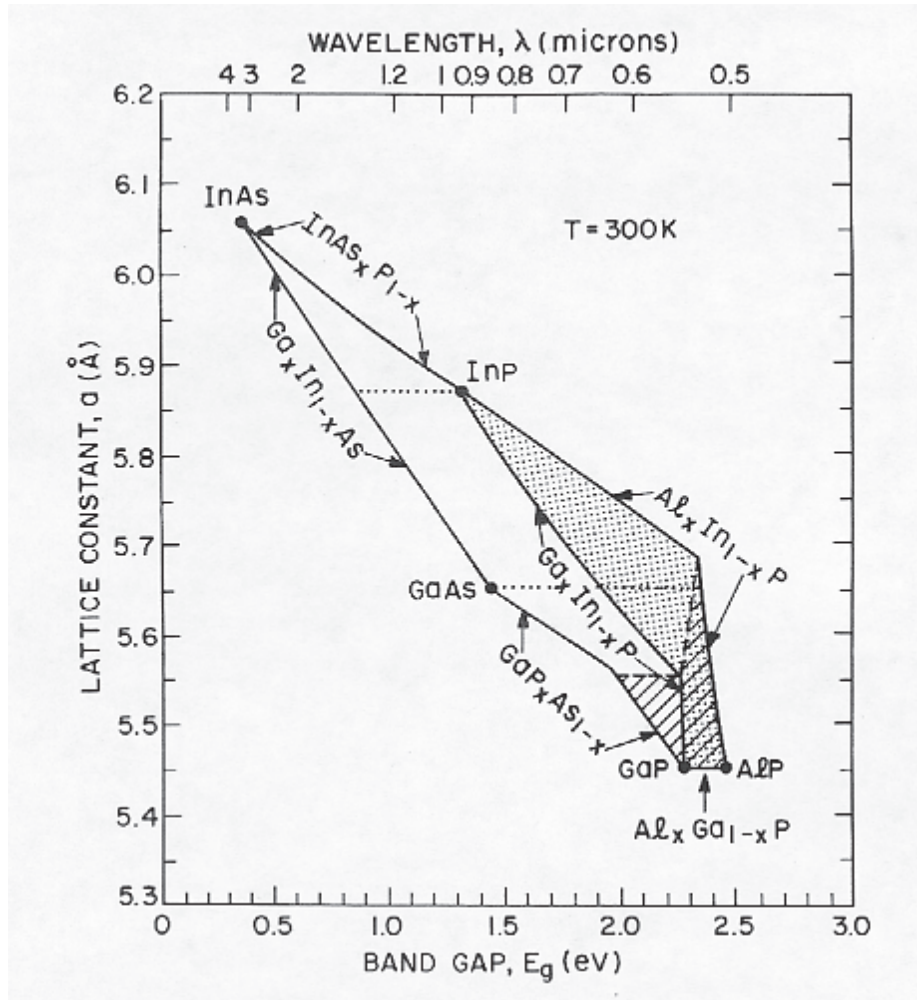
Figure 5-1: Different semiconductor materials and their properties. This information is used to design active optical devices. Figure from [3].

The disadvantages are that it requires more complex processing than silicon and has a lower confinement factor resulting in larger devices.

For the active material components, we use InGaAsP, which provides a good compromise between structural integrity, high nonlinear coefficients, and a bandgap near optical communication wavelengths. Fabrication with very low defects is required for reliability and higher yields. The exact composition of the material is designed such that the optical communication wavelengths are of slightly higher energy than the material bandgap. This results in much interaction between the material and light, allowing us to achieve high nonlinearities and large optical gain.

As the different materials are fabricated using different processes, the main chal-

lenge with hybrid integration lies with the alignment of passive and active devices. The cost of this alignment can be great enough that it eliminates any cost savings from integration. In some cases, the active material is inserted into a hole etched in the passive structure. The alignment between the active device and passive waveguide can be accomplished by flip-chip bonding, magnetic or geometric self-assembling chips, or by pick-and-place techniques. With flip-chip techniques, the active material is flipped and pressed against the passive material. Alignment is achieved with index alignment, mechanical contact alignment, or solder bump self-alignment techniques [7]. With magnetic or geometric self-assembling methods, the active devices are placed on the passive structure in a fluid and capillary or magnetic action is used to draw the devices into alignment [8, 9, 10]. Finally, pick-and-place techniques involve picking up each device and manually placing it in the correct location. This last is the least scalable of the options.

Using a silica-on-silicon platform, the Center for Integrated Photonics has demonstrated the flip-chip hybrid alignment method for fabricating SOA-MZIs. They use a motherboard/daughterboard structure, where the motherboard is silica-on-silicon and consists of the passive components. The daughterboard is silicon, upon which InGaAsP SOAs have been bonded. The silicon daughterboard is flipped onto the motherboard (Figure 5-2). Precision machining ensures good alignment between the passive and active waveguides. This has resulted in < 1.5-dB coupling losses between the waveguide and SOA [11]. Using this technique, they have successfully integrated 4 SOA-MZIs on a single chip in parallel [12].

An alternative to these techniques has been introduced by the Bowers group [14] and uses wafer bonding to enable simultaneous production of a large number of hybrid devices without requiring careful alignment. First, passive waveguides are defined in the silicon wafer. Next, a 2-inch wafer of active material is bonded to the silicon wafer through a low-temperature wafer-bonding process. Finally, the active material is etched and contacts are created. A cross-section schematic of the device structure is shown in Figure 5-3. In this design, light travels in the silicon waveguide and evanescently couples to the active material. The degree of coupling is determined by
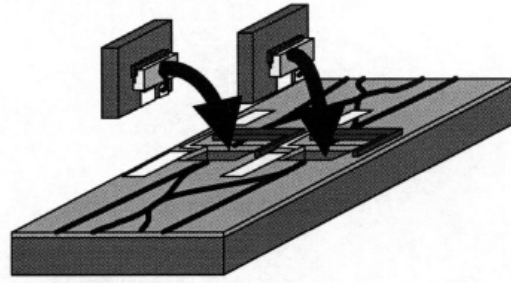
Figure 5-2: Diagram of flip-chip technique with active devices being flipped onto the silica-on-silicon motherboard. Figure from [13].
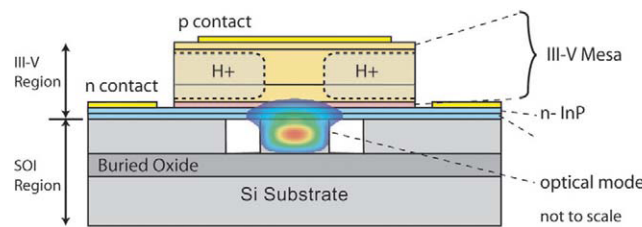


Figure 5-3: Cross-section of the wafer-bonding hybrid integration method. Figure from [15].

the silicon waveguide dimensions, as shown in Figure 5-4. There are two advantages to this structure: first, careful alignment is not necessary, since the optical mode is largely defined by the silicon waveguide. Second, a large number of devices can be formed simultaneously due to wafer-scale processing. However, evanescent coupling is naturally a low-efficiency process, and it is unclear whether nonlinearities high enough for optical switching can be obtained in this manner. Laser diodes, amplifiers, and photodetectors have all been demonstrated with this technique [14, 15, 17].

## 5.1.2 Monolithic Integration

Monolithic integration uses compatible materials for both active and passive devices and thus all fabrication can be done using the same processes or all growth done on the same substrate. Though we generally have to compromise either active or passive device performance slightly, monolithic integration allows us to avoid a costly

160
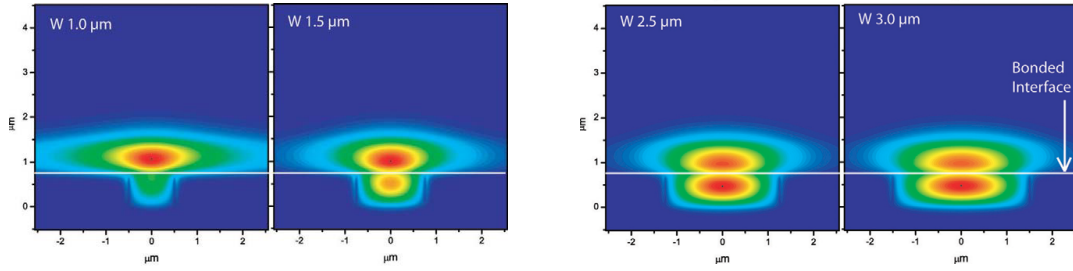
Figure 5-4: Simulation showing the effect of silicon waveguide width on coupling into the active material using the wafer-bonding hybrid integration method. From left to right, waveguide widths of $1.0\mu$m, $1.5\mu$m, $2.5\mu$m, and $3.0\mu$m are shown. Figure from [16].

alignment step and use simpler fabrication processes.

The ideal monolithically-integrated device involves using silicon for both active and passive devices due to its compatibility with electronic processes. However, the indirect bandgap of silicon has thus far prevented the development of efficient active devices. Silica-on-silicon devices could also be promising for monolithic integration due to the possibility of doping silica to create highly nonlinear media, but again, active devices have yet to be monolithically integrated. Currently, monolithic integrated devices are almost entirely created with materials compatible with InP, which provides good active devices. Passive waveguides and couplers can be created by using a different doping profile and a combination of materials which results in a bandgap around 1100 nm. Thus, the two main challenges in monolithic techniques using InP-compatible materials are to maintain low-loss and to increase yield through streamlining fabrication processes.

In the previous chapters, we described an SOA-MZI logic gate where loss in the device is compensated for via amplifying SOAs. These additional amplifiers add noise and complexity to the device. In order to improve upon the design, we examine methods by which device loss can be reduced. One major factor in device loss is the coupling between the active and passive waveguides. The simplest method of fabricating both active and passive waveguides on the same wafer is in using a butt-joint regrowth method in which we first grow one type of material and then etch out regions where the other material needs to be placed. The second type of material is

then grown in the etched regions. Coupling between the active and passive materials is in this case accomplished by the butt-joint interface (Figure 5-5. This method requires careful etching and selective-area regrowth which can reduce yields. Previous demonstrations of butt-joint regrowth have achieved over 90% coupling efficiencies [18] and good extinction in the SOA-MZI [19].
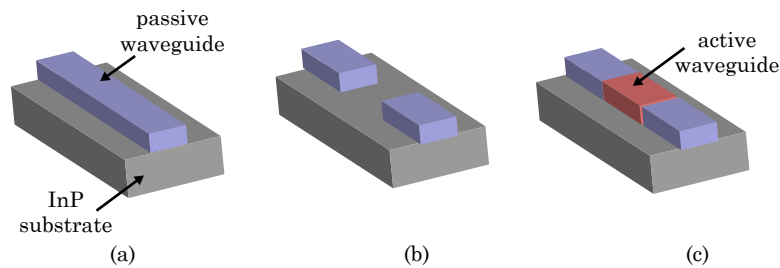


Figure 5-5: Illustration of regrowth fabrication process. In (a), the passive waveguide is created. In (b), we etch a gap in the passive waveguide for the active material. In (c), we grow the active material in the gap.

An alternative technique known as the asymmetric twin-waveguide technique involves fabricating the active waveguide above the passive waveguide. Adiabatic coupling is used to couple light from the passive waveguide to the active waveguide and back by changing the index, width or height of the waveguides. The passive and active materials are grown together and etched to create the required design. Typically, it is easiest to change the width of the waveguide. Figure 5-6 illustrates the structure of the resulting asymmetric twin waveguide. This technique has the advantage of requiring fewer fabrication steps and has the potential to achieve very high coupling efficiencies with large tolerances to misalignment and other fabrication errors [20]. For these reasons, we choose to focus on the asymmetric twin waveguide approach. In the next section, we describe the design of our SOA-MZI gates.
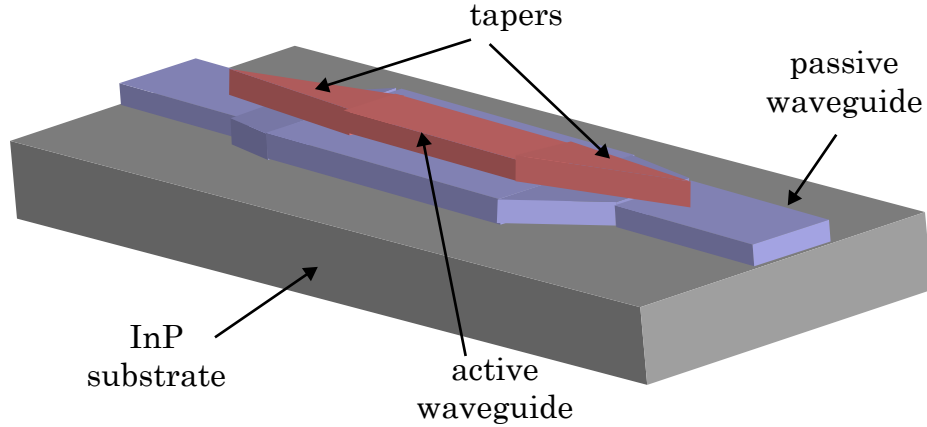
Figure 5-6: Illustration of asymmetric twin waveguide structure. Light couples from the blue passive waveguide up into the red active waveguide due to the change in width of the active waveguide taper.

## 5.2 Asymmetric Twin Waveguide Approach

### 5.2.1 Design

The asymmetric twin waveguide for SOA-MZI was developed by Studenkov et al. [21, 22]. Our device was designed and simulated by A. Markina et al. [20] and the first and second generations were fabricated by R. Williams et al. [23] with the help of Jason Plant and the Electro-Optical Materials and Devices group at Lincoln Laboratory. The design of the active and passive waveguide structure is shown in Figure 5-7. In this design, we choose to use a dilute ridge waveguide structure for both the passive and active waveguides. Using a dilute waveguide allows us to fabricate single-mode waveguides with a large cross-section for easy coupling to fiber. It further allows precise control over the effective index of the waveguide through control of layer thicknesses. Finally, we also obtain a higher-quality semiconductor by inserting InP layers between InGaAsP layers to relieve the strain caused by small, unintended lattice mismatching between the two materials [20].

For the SOA-MZI logic gate, we need to design several different components as shown in Figure 5-8. For passive components, we need to design the waveguide and fabricate bends and couplers (1×2, 2×1, and 2×2). For active components, we need

163
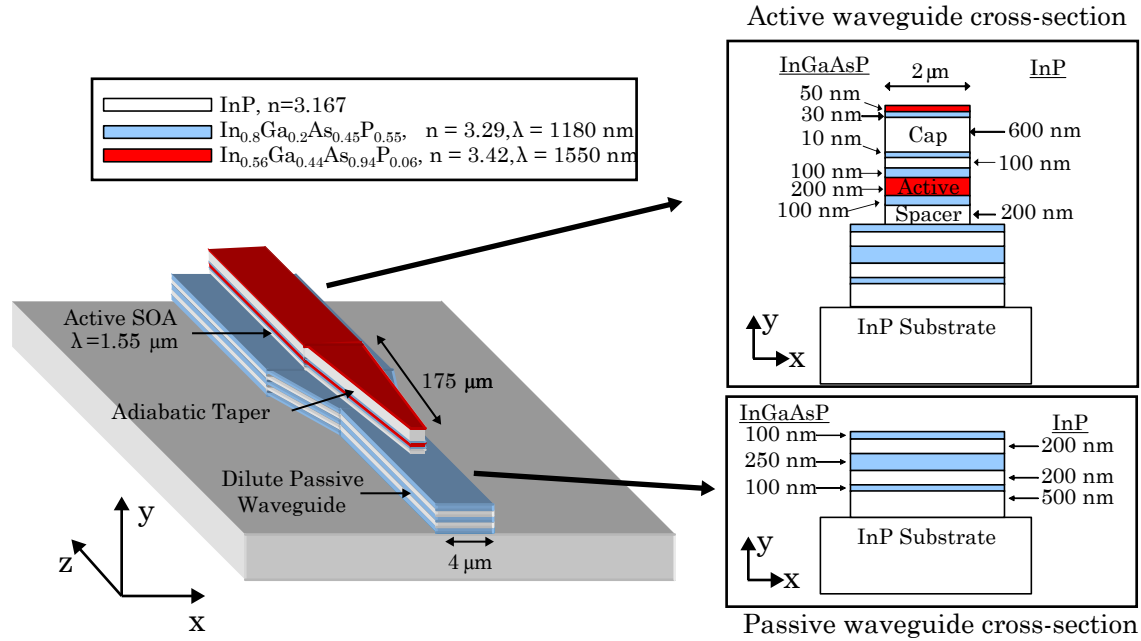
Figure 5-7: Design and cross-section of the active and passive waveguides for the asymmetric twin waveguide approach. A dilute waveguide structure was used. Figure is adapted from figures in [23].

to design the SOA characteristics. Finally, we also need to design the adiabatic tapers which couple the SOA to the passive waveguides. The details of this design are described in Markina's thesis [20]. Here, we present only a brief overview to provide context for the results of our characterization efforts. The passive waveguide was designed to be 4 $\mu$m wide and about 1 $\mu$m high, with a bandgap at 1180 nm to minimize loss at optical communication wavelengths. The bends were modeled using a finite-difference method with cylindrical perfectly-matched layers to allow for the leaky nature of the bends. The design of the bends needs to compromise between lower loss, which gives more efficient devices, and a tight radius, which allows us to create more compact devices. As the radius of the bend decreases, loss increases due to leakage. Figure 5-9 shows this tradeoff for our waveguide structures as simulated by Markina [20].

Multi-mode interference (MMI) couplers were chosen for the passive coupling in the SOA-MZI. MMI couplers have more relaxed fabrication requirements and superior performance when compared with directional and adiabatic couplers [24]. The MMI
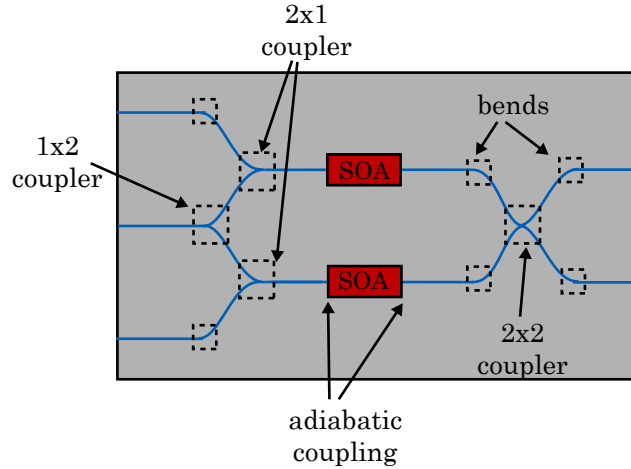
Figure 5-8: Components involved in fabrication of the SOA-MZI. Dashed boxes highlight passive components such as couplers and bends.

coupler is designed such that light entering will excite multiple modes, which beat within the coupler. By carefully controlling the length of the coupler, we can design the coupling ratio between the output ports. The width of the MMI coupler only has to be large enough to accommodate multiple modes and provide sufficient tolerance to inaccuracies in fabrication. Figure 5-10 shows example simulations for the 1x2 coupler and the 2x2 coupler. For the 1x2 coupler, simulations show 48.8% power in each of the output waveguides with an MMI length of 157.4 $\mu$m and width of 12 $\mu$m. For the 2x2 coupler, simulations show 48.8% power in waveguide 1 and 47.2% power in waveguide 2 when operated as a 1x2 coupler. The optimal length is found to be 480 $\mu$m with a width of 18 $\mu$m. For the 2x1 coupler, the phase of the input light can cause large back-reflections by changing the interference of the modes in the MMI. To avoid this possibility, a 2x2 coupler is used instead.

The coupling between the active and passive waveguides is accomplished through the use of asymmetric twin waveguides. In this design, an active waveguide is fabricated above the passive waveguide. In a symmetric design, where the two waveguides are identical, a periodic power transfer between the upper and lower waveguides occurs through vertical modal interference between the odd and even modes. Similar to the operation of the MMI couplers, the amount of power transferred is heavily dependent on the length of the waveguides. By using asymmetric waveguides we can
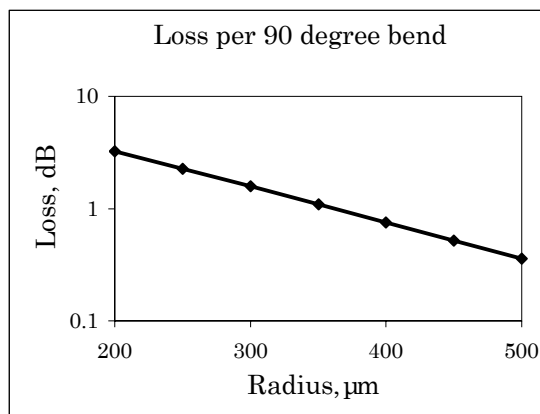
Figure 5-9: Simulation of bend losses in the dilute passive waveguide. As bend radius decreases, loss increases [20].
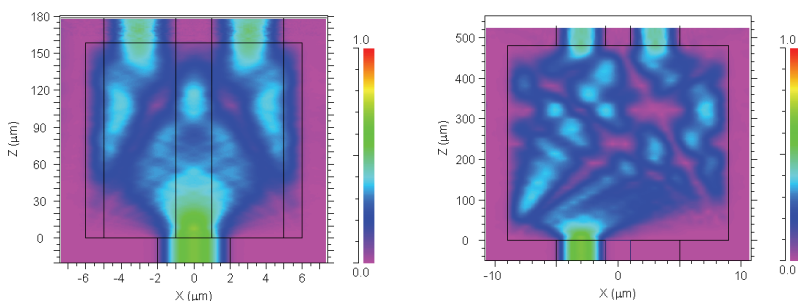


Figure 5-10: Simulation of MMI couplers using BeamProp, a finite-difference beam propagation simulation tool designed by RSoft [20].

reduce this dependence on length. If the effective index of the active waveguide is higher than that of the passive waveguide, we can cause the even mode to be confined in the upper waveguide while the odd mode is largely confined in the passive waveguide. The gain in the active waveguide then causes the even mode to dominate and reduces interference from the even mode effects. Finally, tapering the active waveguide allows for more efficient coupling between the passive waveguide region and the twin-waveguide region. The taper creates a variation in the effective index of the upper waveguide, which allows the fundamental (even) mode of the passive waveguide to move up into the active waveguide. We change the taper widths slowly to avoid the introduction of higher-order modes. The design of the length of the taper,

type of taper, and width of waveguides is optimized to obtain close to 100% coupling efficiency. Figure 5-11 shows an example of the simulation of coupling efficiency.
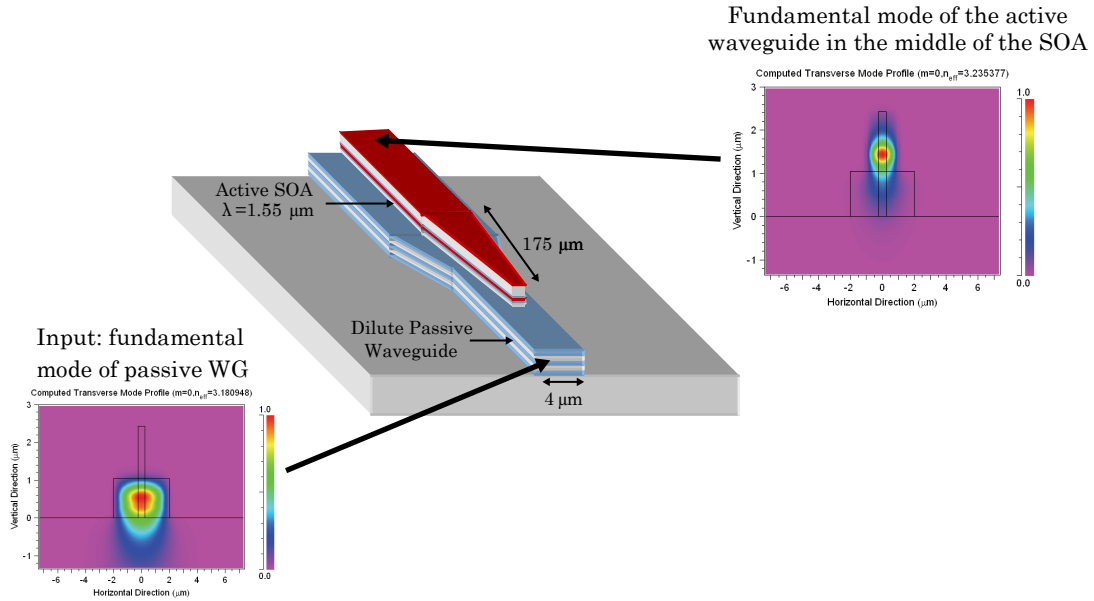


Figure 5-11: Simulation of active-passive adiabatic coupling using BeamProp [20]. WG stands for waveguide.

The design of a good nonlinear SOA for switching depends on several variables which can be difficult to isolate. In these devices, we have initially focused on verifying basic operation of the SOA-MZI logic gate before optimizing SOA operation. The quantities which are important for SOA operation in ultrafast switching include the carrier recovery time $\tau$, the linewidth enhancement factor $\alpha$, the gain $g$, and the saturation energy $E_{sat}$. As shown in Chapter 3, the carrier recovery time directly affects the speed of the SOA switch, especially in single-ended operation of the SOA-MZI. With differential operation, this effect is reduced but not completely eliminated. In differential operation, a long sequence of closely-timed pulses will still drive the SOAs deeper into saturation. Both SOAs are saturated to the same extent, which means that the pulses in the interferometer are matched in phase. However, saturation means that the next arriving control pulse will have a smaller effect on the phase and amplitude of the signal pulse. Since the control pulse powers are not adaptively

adjusted to account for the saturation, this will affect the extinction of the switch. A shorter carrier recovery time means faster recovery between pulses, which means that the same sequence of pulses will not drive the SOA as deeply into saturation. The carrier recovery time is dependent on the material used and can also be affected by the confinement factor of the SOA [25].

A higher saturation energy means that a longer sequence of pulses is required before saturation is reached and switch effectiveness is degraded. This saturation energy can also be similarly affected by the confinement factor and active area of the SOA. A larger active area or larger mode means the saturation energy is higher. The length of the SOA also affects the ease of SOA saturation due to the generation of ASE within the device. A very long SOA can even be self-saturating due to the buildup of ASE along the length. Finally, the linewidth enhancement factor controls the switching energy required to provide a $\pi$ phase shift. This switching energy is dependent on the device material as well as the SOA length. A longer SOA means that there is more distance over which the pulses will experience the nonlinearity. Thus, a smaller switching energy can be used to achieve a $\pi$ phase shift.

## 5.2.2  Fabrication

In the first and second generation devices, Williams et al. designed several test structures for aiding in the characterization of the separate passive and active components [23]. The mask layout of a single second-generation chip is shown in Figure 5-12. To reduce reflections, a 7° angle was introduced between the passive waveguide and the edge of the chip. In addition to the different test structures, different lengths of SOAs were fabricated to examine the effect of SOA length on performance. Furthermore, different widths were used (2 $\mu$m and 4 $\mu$m) to allow comparisons to be made for coupling efficiencies and device performance.

Epitaxial growth was done on InP wafers by IQE, a semiconductor wafer growth company, and consist of the active material grown above the passive material on an InP substrate. The layer structure of the wafer is shown in Appendix B. The fabrication of the wafer involved the following steps. This summary follows the description
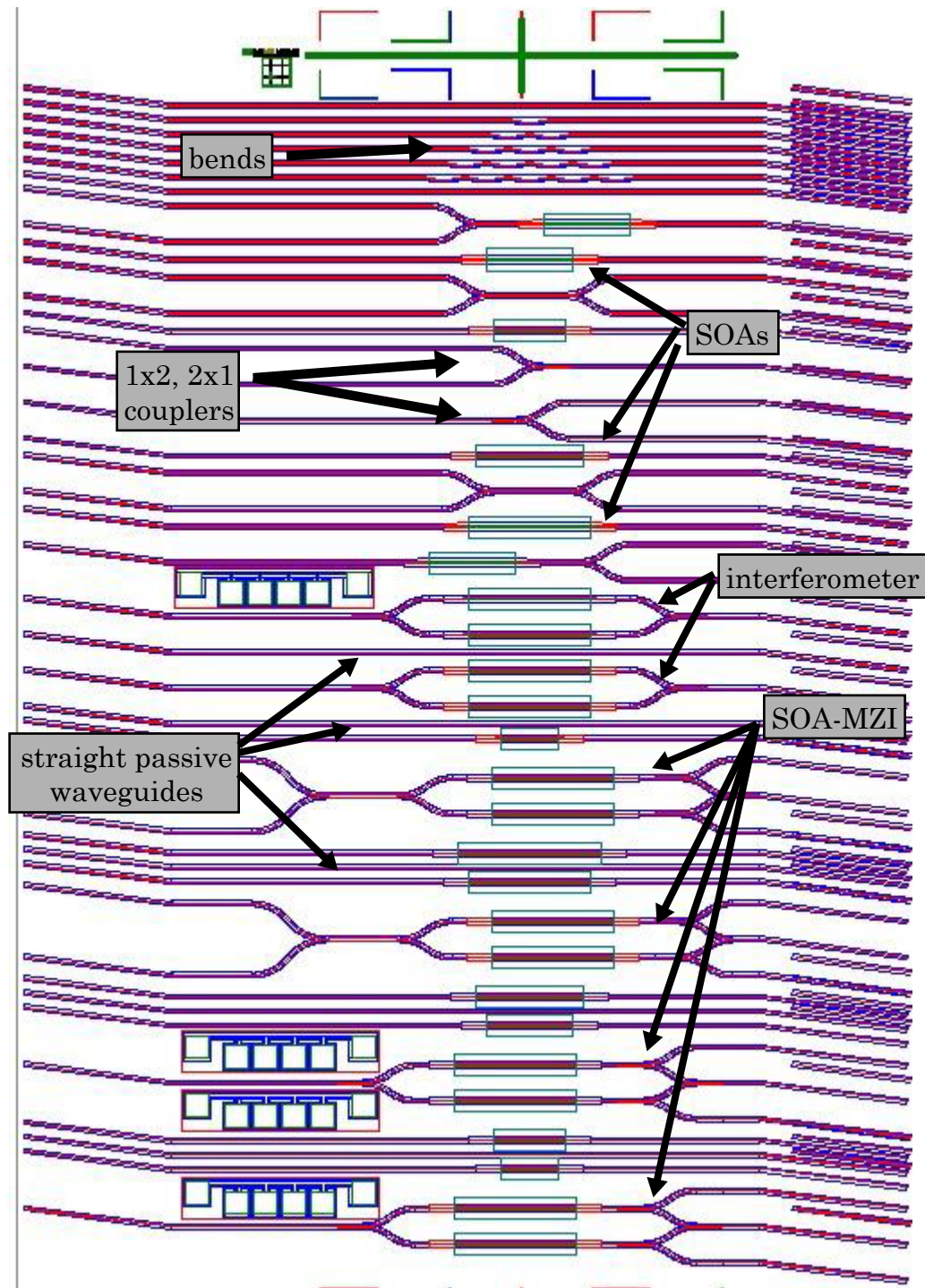
Figure 5-12: Layout of a single chip designed by Williams. Several test structures are included to characterize performance of various components for the SOA-MZI [23].

first set forth in Williams' thesis [23]. First, a Ti/Pt base metal is deposited on the SOAs by applying a resist to the other areas of the wafer and then depositing 20 nm of Ti and 20 nm of Pt on the wafer. An organic solvent is used to remove the resist and the metal on top of the resist, leaving only the Ti/Pt metal on top of the SOAs. Next, we deposit a thick layer of oxide ($SiO_2$) using a process known as PECVD (plasma-enhanced chemical vapor deposition), which deposits material on the top and sides of the structures on the wafer. We also deposit a patterned photoresist to create the active region tapers and trenches for the waveguides. These trenches are a critical part of the second-generation design, allowing us to achieve higher accuracies with regard to taper sidewalls. A flourine-based reactive ion etch (RIE) is used to etch away the oxide not covered by the photoresist. This defines the location of the active waveguide and tapers, as shown in Figure 5-13(a). Next, we use a chlorine-based inductively-coupled plasma reactive ion etch (ICP-RIE) process to etch the active material not covered by the oxide, as shown in Figure 5-13(b).

The next step is to etch the passive waveguides. We deposit another layer of oxide using PECVD to protect the active waveguide sidewalls. We then repeat the process above to create the passive waveguide and couplers. This is shown in Figure 5-13(c). Next, the oxide masks are removed using a hydrofluoric (HF) etch (Figure 5-13(d)). The Ti/Pt metal contact above the active waveguide can be accidentally removed at this step. Finally, the electrical contacts for pumping the SOA devices must be fabricated. In order to create a large contact pad and avoid shorting the p-type and n-type materials in the active region, we planarize the wafer using benzocyclobutane (BCB). BCB is an insulator which is spun on in liquid form and then cured at high temperature. After a thick layer of BCB is spun on and cured, a plasma etch is used to etch back the BCB until only the top of the active waveguide is revealed (Figure 5-13(e)). Finally, a Ti/Pt/Au composite p-type metal contact is evaporated onto the top of the active region. The back-side n-type contact (Ni/Au/Ge/Au) is sputtered on after lapping the substrate to reduce the thickness. The wafer is then rapid-thermal annealed to alloy the contacts (Figure 5-13(f)). A scanning electron micrograph (SEM) of the taper before planarization is shown in Figure 5-14(a).
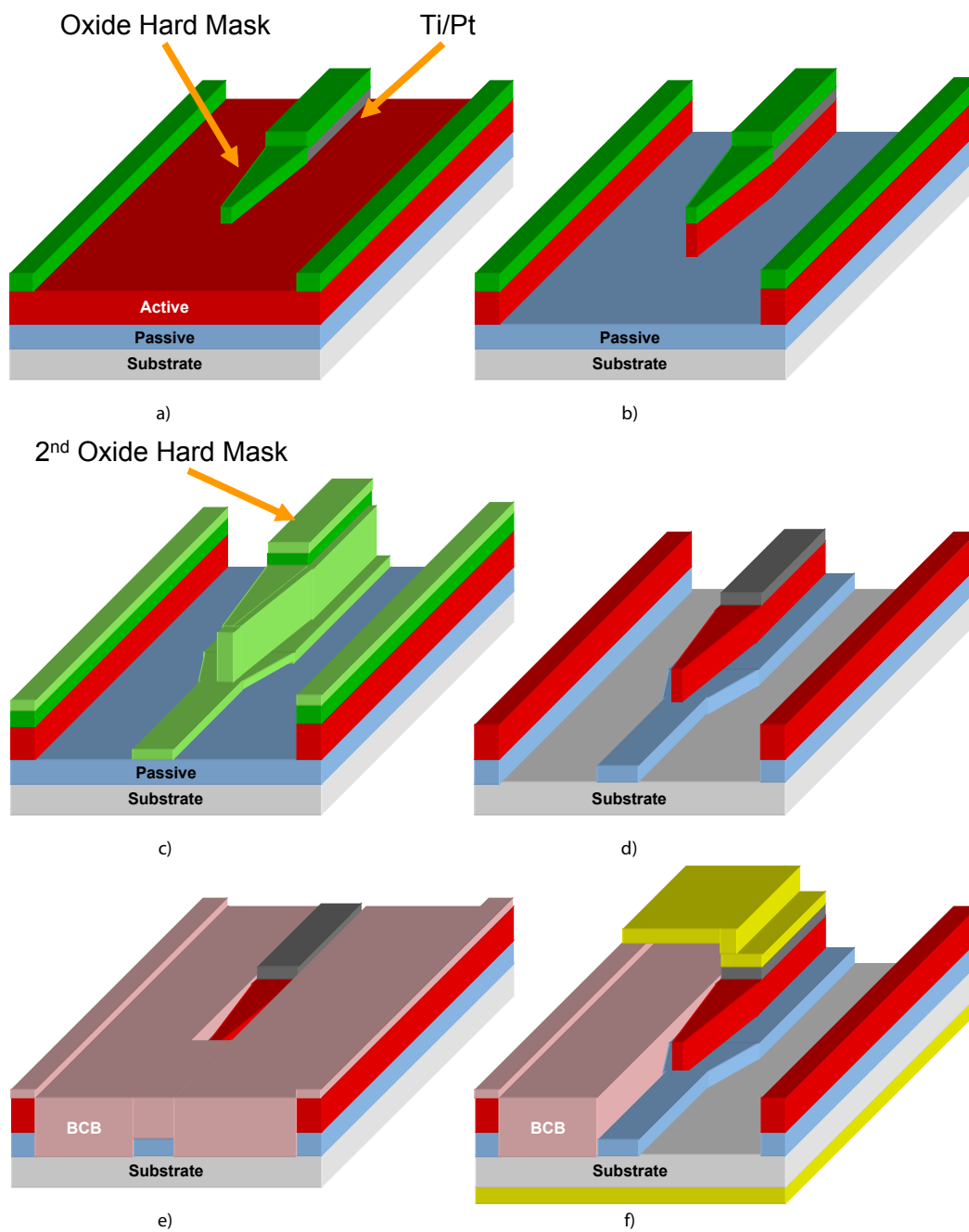
Figure 5-13: Figure from [23]. Illustration of fabrication steps for the asymmetric twin waveguide SOA-MZI design.
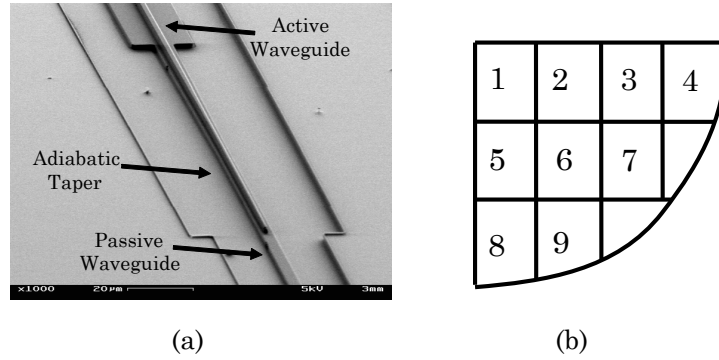
Figure 5-14: (a) SEM of fabricated taper before planarization [23]. (b) Illustration of quarter-wafer with nine potential chips.

Processing was done on a quarter-wafer at a time to minimize waste and allow experimentation with different processing steps. Each quarter wafer consists of nine potential chips, though the chips at the edge are less reliable (Figure 5-14(b)). With the help of J. Plant, the quarter-wafer is cleaved and the chips are mounted. For purposes of testing, some chips were straight-cleaved, such that the waveguides met the chip facet at a 90° angle, while others were angle-cleaved to minimize reflections. In the next section, we describe the progress thus far in characterizing our chip design, verifying operation, and noting potential areas for improvement.

## 5.3  Characterization and Testing

We divide our characterization efforts into the testing of passive components and testing of active components.

### 5.3.1  Passive Components

We first determine the loss of the passive waveguide using the Hakki-Paoli [26] method of measuring waveguide loss. This method models the waveguide as a Fabry-Perot resonator with a loss element. As we vary the wavelength of the input light, the transmission of the resonator reaches peaks and valleys depending on constructive or destructive interference between the reflected waves (Figure 5-15). The fringe spacing
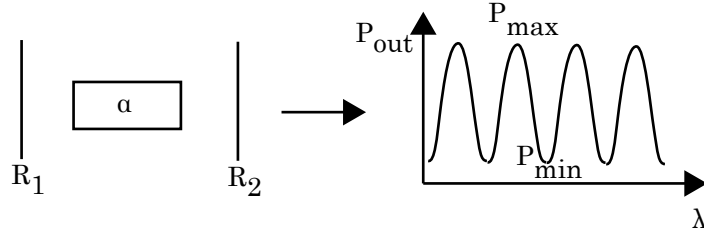
Figure 5-15: Hakki-Paoli method for measuring loss in a waveguide.

depends on the length of the device. The value of the maximum and minimum powers depends only on the loss, reflectivities, and length. Thus, if we know the output powers, reflectivities, and length, we can extract the total loss of the waveguide:

$$\alpha(\lambda) = \frac{1}{L}\ln\left(\frac{\sqrt{P_{max}} + \sqrt{P_{min}}}{\sqrt{P_{max}} - \sqrt{P_{min}}}\right) + \frac{1}{2L}\ln\left(R_1 R_2\right), \tag{5.1}$$

where $\alpha$ is defined by

$$P_{out} = e^{-\alpha L} P_{in}. \tag{5.2}$$

The major advantage of this measurement method is that it is completely independent of coupling losses into and out of the waveguide. An additional variation is known as the cut-back method and involves making the same measurement over different lengths. This would allow us to eliminate the reflectivities as variables. However, it is a destructive test and due to the chip layout, we choose to avoid this method for now.

We use a straight-cleaved chip to maximize the reflections and ensure a good measurement. A high-resolution tunable laser is aligned to the waveguide through a "laser optical fiber interface" (LOFI) which consists of a fiber and lens to focus the light on the facet. The output light is collected by a 5x microscope objective and a power meter. The schematic is shown in Figure 5-16. An IR camera is used to image the output of the facet as an aid in alignment. We need to amplify the tunable laser in order to overcome the large coupling losses in the setup. Since the output of the tunable laser is large enough to saturate the amplifier, the additional

173
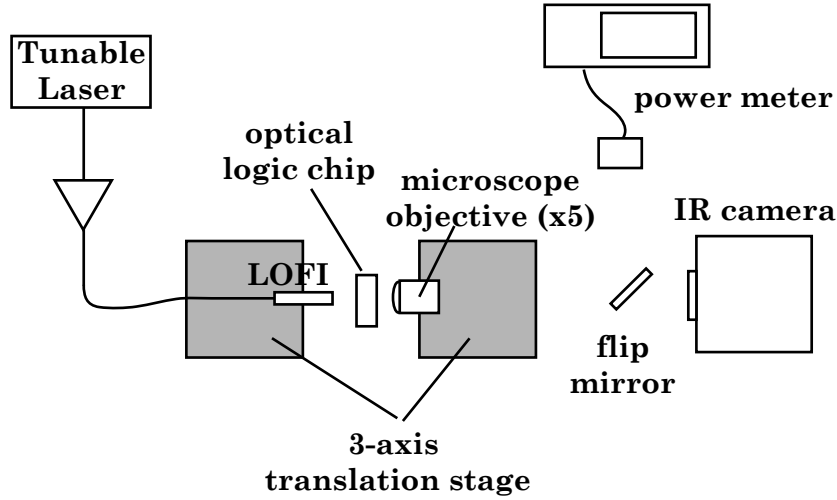
Figure 5-16: Schematic for waveguide loss measurement.

wavelengths introduced by the ASE of the amplifier are at much lower power and do not significantly interfere with the measurement. The output of the measurement is shown in Figure 5-17. The additional slower modulation on the fringes is probably due to additional reflections in the experimental setup. We take an average of the loss calculated from three of the fringes near 1550 nm and obtain a loss of 0.95 cm$^{-1}$ or 4.1 dB/cm. We calculate the reflections to be $R_1 = R_2 = 0.272$ assuming a 90° angle to the facet and an effective index of 3.18 in the waveguide. The length of the device is 3.5 mm.

Next, we measure the bend losses in the device. We make use of the bend test structures in the upper part of the chip, where different numbers of bends are fabricated on each waveguide. Several different bend radii and degrees are used in fabricating the SOA-MZI. The larger bends have radii of roughly 330 $\mu$m and the smaller bends have radii of roughly 296 $\mu$m. The bends for the test structure have a radii of roughly 285 $\mu$m. The angle of the bends in the SOA-MZI device and the test structures differ significantly. Thus, the measurement of loss in the bends of the test structure may not provide useful information pertaining to the losses in the SOA-MZI structure. Figure 5-18 shows microscope images of the bends in the test structures and the bends in the SOA-MZI, to scale. Nevertheless, the bend losses in the test-
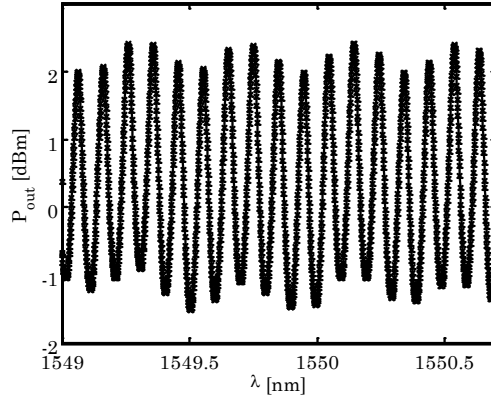
Figure 5-17: Hakki-Paoli measurement of the passive waveguide loss. The additional modulation on the fringe is probably due to additional reflections in the measurement set-up.

structure may give us an order-of-magnitude understanding of the bend losses in the SOA-MZI device. We measure the bend loss by using the Hakki-Paoli method to measure the additional loss in the waveguide introduced by increasing numbers of bends. We define a single bend as an S-bend shown in Figure 5-18. The sequence of test structures provides us with waveguides of 2, 4, 6, and 8 bends. With 8 bends, the losses are too high for our detector to make an accurate measurement of the fringes. The Hakki-Paoli method allows us to ignore differences in coupling efficiencies between waveguide measurements, but if the coupling angle varies between waveguide alignments, the reflectivities may be different. Figure 5-19 shows the loss measured for the waveguides with different numbers of bends. We use a least-squares fit to arrive at a loss per bend of 1.17 dB, assuming an accurate measurement of straight waveguide loss.

We have not yet quantified the MMI coupler efficiency of the passive device. We have qualitatively observed coupler outputs for the 2x2 and 1x2 on the infrared camera. Figure 5-20 shows the output of a 2x2 coupler. Note that there is significant additional light not coupled into the waveguide. This is due to a mismatch in the spot size of the focusing lens and the waveguide dimensions. For the measurement of waveguide loss, this is not significant because we care about the difference between
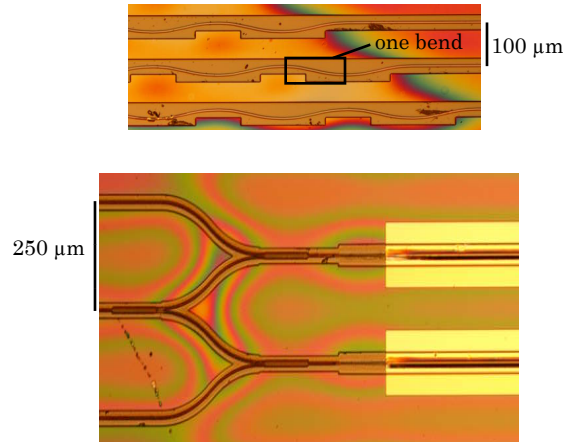
Figure 5-18: Microscope images of the bends in the test structures and the SOA-MZI bends on a chip. Images are roughly to scale.
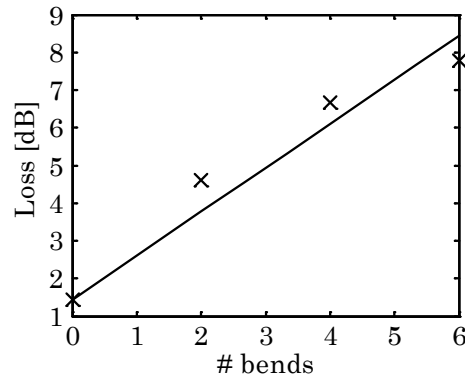


Figure 5-19: Measurement and fit of loss per bend in the test structures. Loss measurements are marked with "X" and the line illustrates the least-squares fit assuming an accurate straight waveguide loss measurement (0 bends).

maximum and minimum powers due to the reflections of the facet, rather than the total power collected by the detector. In a measurement of coupler performance, however, the additional light significantly degrades the accuracy of the measurement. A higher magnification lens and iris will be needed to reduce the excess light on the detector. Alternatively, a digital IR camera can be used to spatially isolate the power in the waveguide.

176

<div align="center">(a)            (b)</div>
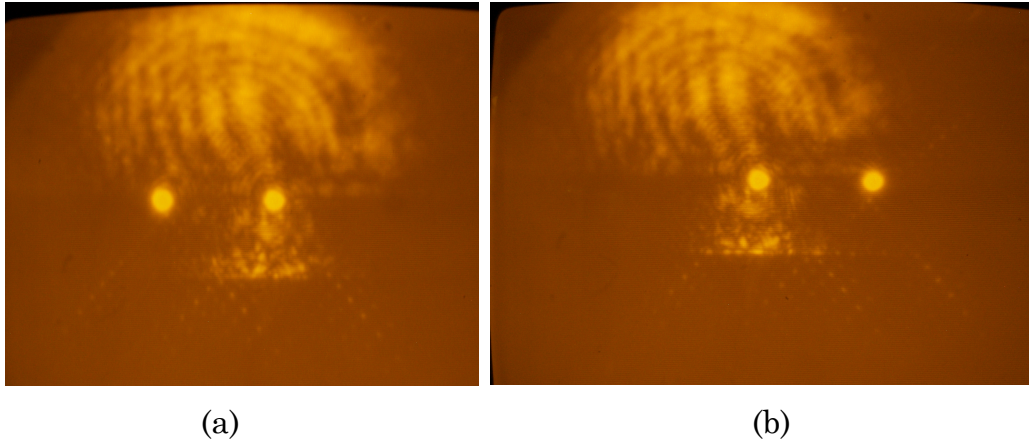
Figure 5-20: Image of the output facet of a 2x2 coupler using an IR camera and a 5x microscope objective. (a) Output facet with light coupled into the "left" input. (b) Output facet with light coupled into the "right" input.

## 5.3.2 Active-Passive Coupling

It is important to measure the active-passive coupling efficiency of the asymmetric twin waveguide design. This is difficult without knowing the coupling efficiency into the waveguide, the active material loss, or the percentage of light which remains in the passive waveguide. In order to decouple some of these variables, we perform the following measurements. First, we cleave the chip in half so we can isolate the passive-to-active coupling process from the active-to-passive coupling process and so we can access the active waveguide. The asymmetric design of the twin waveguide should ensure that interference between the odd mode in the passive waveguide and the even, fundamental mode in the active waveguide is minimal. This means that the coupling between the active and passive waveguides is insensitive to length and we can assume that the power in the active waveguide observed at any length is roughly the same. This is borne out in simulations performed by A. Markina [20]. Figure 5-21 illustrates a cut-away view of the a cleaved device. We also use a single SOA test structure to avoid complications from couplers or interferometers.

We make three measurements: the loss of the entire waveguide, the total fiber-to-fiber loss below the bandgap, and the total fiber-to-fiber loss above bandgap. For these measurements, we use a LOFI to couple light from a wide-bandwidth source
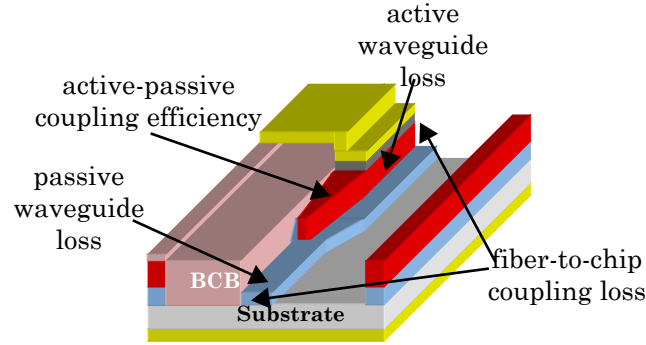
Figure 5-21: Cut-away view of the variables involved in measuring the active-passive coupling efficiency.

(ASE from an EDFA) into the waveguide. To measure the total loss of the entire waveguide, we use a variation of the previously described Hakki-Paoli technique. We use a second LOFI to couple light from the output of the waveguide into a optical spectrum analyzer. The advantage of this method over the tunable laser method is that we can obtain a set of fringes nearly instantaneously. The limiting factor in this measurement is the resolution of the optical spectrum analyzer. Without a high enough resolution, the peaks and valleys are not clearly defined and the measurement of loss will be inaccurate. We used a spectrometer with a resolution of 0.08 nm. The peaks and valleys of the fringes were comparable to that measured with a resolution of 0.01 nm. This gives us confidence that our resolution is high enough for an accurate measurement. We measure the Hakki-Paoli fringes near 1561 nm to obtain a total waveguide loss of 16.6 dB. Figure 5-22(a) shows the fringe measurements. The curve on the fringes is due to the curve in the input spectrum at those wavelengths. We again calculated our reflectivities to be 0.272 and measured the length of the total device from the fringe spacing to be 0.213 cm. This loss measurement includes the passive waveguide loss, active-passive coupling loss, and active waveguide loss. We estimate the length of the active region to be roughly 500 $\mu$m. Since we have previously measured the passive waveguide loss to be 4.1 dB/cm, this means that the contribution of the passive waveguide to the total loss is about 0.67 dB. Thus, the combination of the active waveguide loss and the active-passive coupling loss is 15.9

178
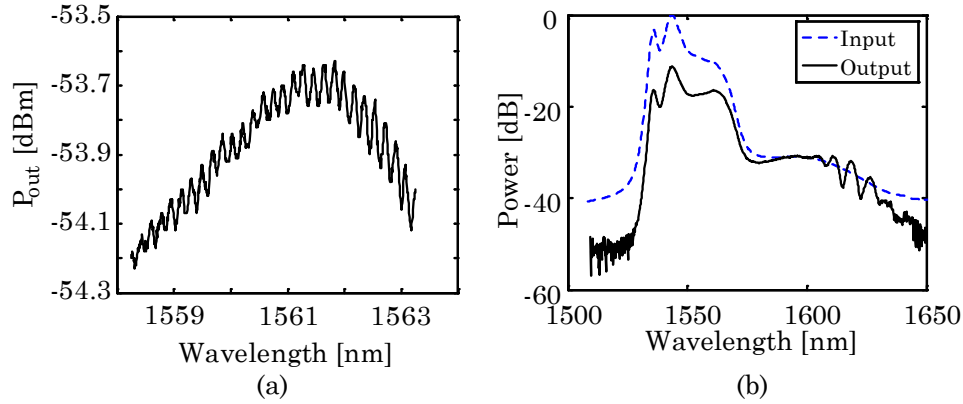
Figure 5-22: (a) Fringe measurements of the cleaved waveguide. (b) Input and output spectrum through the cleaved waveguide. Input spectrum is normalized. Output spectrum is normalized to 0 loss at 1598 nm.

dB.

Next, we measure the input spectrum and compare it against the output spectrum (Figure 5-22(b)). We assume that the fiber-waveguide coupling losses and active-passive coupling losses are independent of wavelength. This may not be accurate over a wide wavelength range, as shorter wavelengths may result in multi-mode behavior and longer wavelengths may experience more loss. However, we perform our measurement across 100 nm, which is small compared with the $4mu$m of the waveguide. Furthermore, we assume that the active material behaves passively at energies below the bandgap. Thus, if we subtract out all passive losses at 1598 nm, we will be left with the additional loss of the active material at energies above the bandgap. We zero the loss at 1598 nm and subtract the output spectrum from the input spectrum. The remaining loss should be the additional loss of the active waveguide from absorption, shown in Figure 5-23. This rough measurement resulted in an active material loss of 6.5 dB and an active-passive coupling loss of 9.4 dB. The largest source of error in this measurement is our assumption that the input signal was either absorbed by the active waveguide or lost in the coupling from passive to active waveguide. The output LOFI has a spot size which is large enough to collect light from both the passive and active waveguides. An accurate determination of the passive-to-active
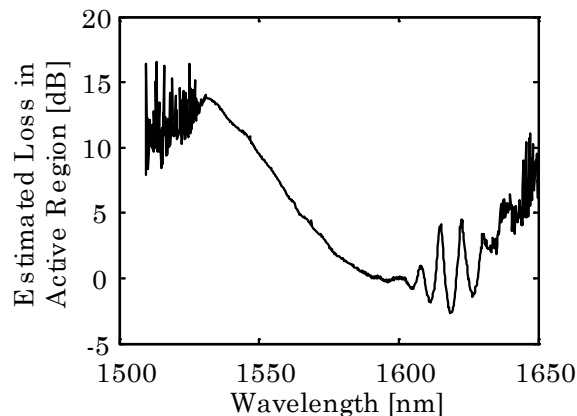
Figure 5-23: Estimated loss in the active waveguide.

coupling efficiency cannot be made if we are measuring light which did not couple up into the active waveguide. We note that our measurement of the active material loss is 6.5 dB over 500 $\mu$m, which corresponds to a loss of 130 dB/cm. Previously-measured losses for similar materials are an order of magnitude higher, about 4000 dB/cm [27]. This could suggest that significant light propagates in the passive waveguide and thus experiences much less loss.

In order to address these issues, we must be able to quantify the loss in the active material. We must also be able to accurately distinguish between the active and passive waveguide outputs. The first can be accomplished by fabricating a test structure which is simply a length of active waveguide. A short length may be necessary such that losses are low enough for us to measure the output accurately. Then, Hakki-Paoli measurements can be performed. To accurately distinguish between the optical outputs of the active and passive waveguides, we require a microscope objective of at least 100x magnification. This gives us a resolution of roughly 1 $\mu$m. Since our waveguides are each 1 $\mu$m in height, this is the minimum resolution at which they can be distinguished. Ideally, we would like a higher magnification, but such objectives are extremely difficult to manufacture.

Our preliminary results, however, indicate that there may be a significant coupling loss between the active and passive waveguides. Investigation revealed that loss in

an un-pumped active taper could be significant, thus reducing the coupling efficiency. These simulations were performed by T. Shih and are reproduced in Figure 5-24. A 4 $\mu$m wide active waveguide similar to the SOA tested was simulated using an active material loss of 1000 cm$^{-1}$. From Figure 5-24, we see that this high loss reduces
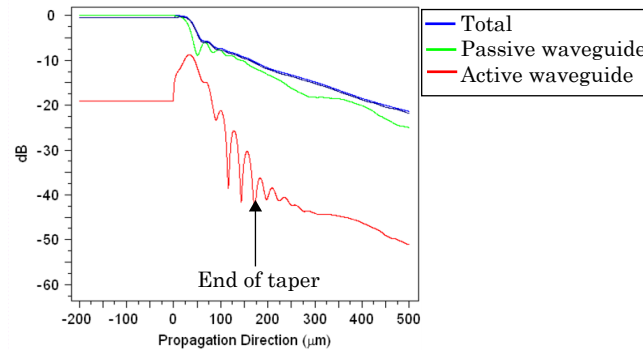


Figure 5-24: Simulation of coupling between the active and passive waveguides performed using BeamProp. Figure from T. Shih.

coupling into the active region and ensures that the majority of the measured output power remains in the passive waveguide. Furthermore, we see that significant loss is experienced within the taper itself. As a result, we conclude that actively pumping the taper will be necessary for proper operation of the SOA-MZI device.

### 5.3.3 Active Device Testing

To test the performance of the active devices, we first test the electrical current response. To do this, we must form an electric contact with the chip. The back-side contact is made through the mount of the chip. The top-side p-type contact, however, must be made through either wire-bonding or using a probe. Our attempts at wire-bonding reveal a poor mechanical bond between the SOA and the p-type contact. The force of the wire-bonding mechanism lifts the p-type contact from the surface of the diode, as shown in Figure 5-25. A better mechanical bond between the alloyed top contact and the material underneath can be achieved through an alternative contact material and an additional fabrication step. In the meantime, probing of the device is sufficient to make an electric connection.
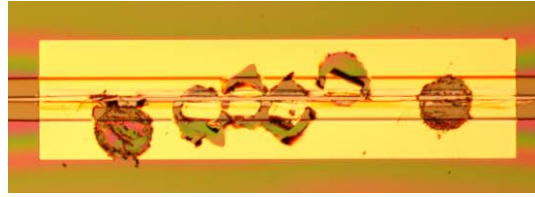
Figure 5-25: Microscope image of wire bonding attempt on a SOA.

We measure the diode response of the SOAs by measuring the output voltage over different values of the pumping current. From this diode response, we can measure the resistance of the diode and the threshold current. In performing this measurement, we discover that in several cases, the contact pad did not make electrical contact with the diode. Further investigation revealed that in some cases, the contact pads separated from the diode contacts, as shown in the SEM photograph taken by T. Shih in Figure 5-26. We also observe that the BCB used in planarization had pulled
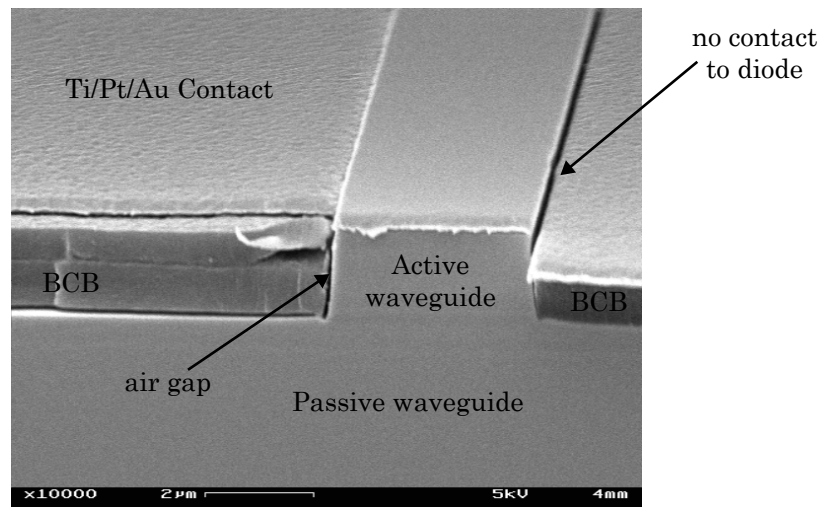


Figure 5-26: SEM photograph of a SOA showing a disconnect from the contact pad and the shrinkage of the BCB. Photograph from T. Shih.

away from the waveguide, revealing an air gap. The rapid thermal anneal for the contacts in the last step of the fabrication caused the BCB to shrink. This caused strain in the contact above the diode and the BCB, which in some cases resulted

182

in a complete break. In other cases, thin connections remained. There are several potential solutions to this issue First, a different contact alloy which does not require as high temperatures could be used to minimize BCB shrinkage. A second option is to investigate a different planarization material. Third, we can electroplate or evaporate a much thicker gold contact such that any strain would not cause a complete disconnection between the diode contact and the contact pad. These options are currently under investigation by T. Shih.

We focus on diodes which have an electrical contact with the contact pad. (Probing the active waveguide contact directly is likely to result in damage to the waveguide.) With these diodes, we discover a maximum current of about 80-100 mA before the diode failure. Failure in this case meant the destruction of the waveguide such that no optical light could propagate as well as an electrical failure such that contact with the diode could no longer be made. We took a microscope photograph of a failed diode (Figure 5-27). Two reasons for this failure exist: first, electromigration may
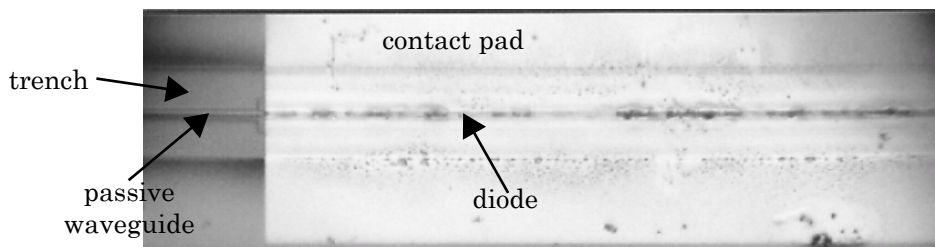


Figure 5-27: Microscope image of burned-out SOA at 16x magnification.

have caused an uneven distribution of the current to the diode, resulting in an arc which destroyed the waveguide. Alternatively, high temperatures due to high resistance in the contact may have caused melting and alloying of the active waveguide. Electromigration can be alleviated with a thicker contact layer. Thermal issues may be more difficult to solve, given that BCB and InP conduct heat poorly. We choose to delay a more careful investigation of this effect, as that would have likely led to many more device failures and we wish to ensure that some devices remain for other characterization purposes. Instead, we maintain our pumping currents below 50 mA to provide a margin of safety.

We next test the V-I characteristics of several SOAs by injecting current and measuring the voltage across the diode. This allows us to measure the resistance of the diode and verify proper device operation. We use a four-point measurement to reduce the effect of probe resistances. Figure 5-28 shows a sample plot of the V-I characteristic of an SOA. By performing a linear fit to the upper portion of the curve,
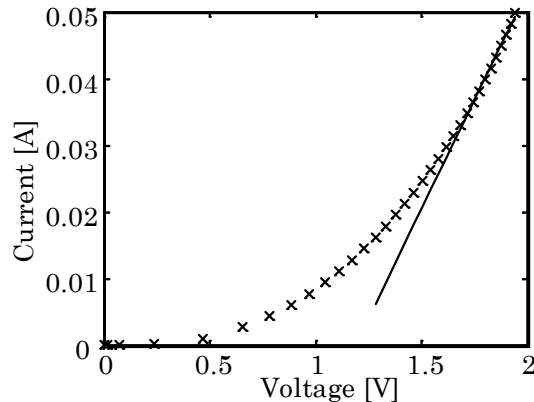


Figure 5-28: V-I plot of a SOA, with a linear fit for the resistance.

we find that the resistance of the SOA is 15.3 ohms. This total resistance can be separated into the junction resistance internal to the SOA and the contact resistance which describes the resistance between the contact and the diode. The majority of our measured 15.3 ohm resistance probably arises from the contact resistance due to the relatively low doping of the top of the diode (about $2 \times 10^{18}$ cm$^{-3}$). This level of doping, when used with a Ti/Pt/Au contact, results in a specific contact resistance on the order of $10^{-4} \Omega$cm$^2$ [28, 29]. The area of the device measured in Figure 5-28 was $850 \mu$m by 1 $\mu$m, which would correspond to $1.4 \times 10^{-4} \Omega$cm$^2$. A high contact resistance degrades performance of the SOA by increasing localized heating. This decreases the gain of the device by moving carriers to higher energy levels, thus decreasing carriers available for stimulated emission. In order to reduce this contact resistance, we can use a different metalization (Ge/Pd/Zn/Pd) which promises to achieve contact resistances on the order of 0.1 ohms [30]. An alternative solution is to dope the InGaAsP film to over $10^{19}$ cm$^{-3}$, which should have a similar effect [31, 32].

184

We next investigate the optical response to current injection. We begin with a measurement of the amplified spontaneous emission using an optical spectrum analyzer (Figure 5-29). The peak of the ASE spectrum should be near the band edge, where the number of electrons in the conduction band is the highest. We measure our ASE peak at 1574 nm.
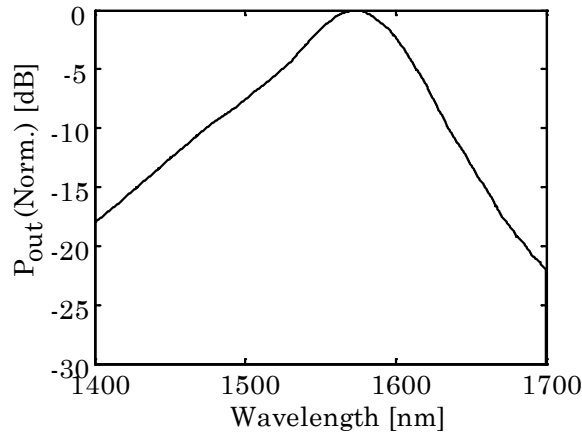


Figure 5-29: Spectrum of the amplified spontaneous emission with input current of 42.2 mA.

We also wish to measure the stimulated emission and determine if we can achieve gain in the device. To do this, we perform a transparency point measurement as shown in Figure 5-30. We focus CW light from a tunable laser into the SOA using
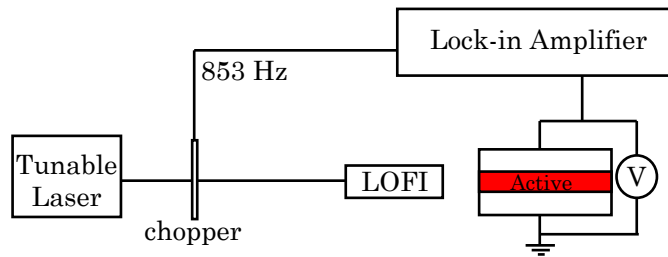


Figure 5-30: Schematic for the transparency point measurement. LOFI stands for "laser optical fiber interface" and consists of a fiber and a set of lenses to focus light from the fiber onto a waveguide.

a LOFI. We use a chopper to chop this input signal at a slow frequency of 360 Hz and measure the change in the voltage across the SOA at that frequency. In the

absorption regime, photons from the tunable laser are absorbed, causing carriers to be pumped up into the conduction band. This increases the number of carriers and the Fermi level of the conduction band, causing an increase in the voltage across the diode. In the gain regime, photons stimulate emission, removing carriers from the conduction band. This creates a decrease in the voltage across the diode. We chop the input light to observe the change in the voltage between the presence and absence of incident photons. If the presence of photons causes a positive voltage change, the SOA is in the absorption regime. If the presence of photons causes a negative voltage change, the SOA is in the gain regime.

The two factors which affect which regime the SOA is in are the energy of the photons and the pumping current, as shown Figure 5-31 from [33]. Below the bandgap,
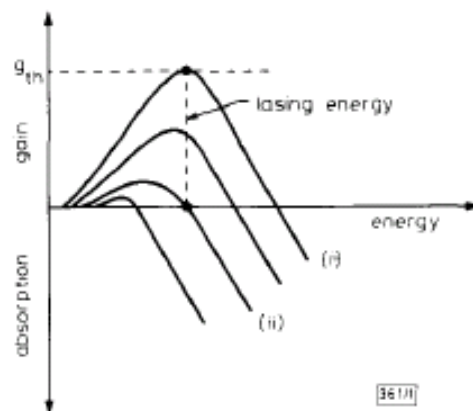


Figure 5-31: Diagram of gain and absorption in the SOA over different values of pumping current and wavelength. Figure from [33].

the photons do not interact with the SOA and thus neither gain nor absorption is observed. Above the bandgap, photons can experience either gain or absorption depending on the pumping current. Increasing current pumps more electrons into the conduction band, increasing the energies over which photons can experience gain. SOA gain increases with increasing current until saturation is reached.

This experiment allows us to accurately determine if we can successfully achieve internal gain. We measure the change in voltage in our device over several wavelengths and pumping currents. The wavelength range in our tunable laser limits the range of

186

our measurement. Figure 5-32(a) shows the result of our measurement. We see from this measurement that we do see internal gain. However, we note that the measured internal gain is small. We compare our measurements to those of an Alcatel SOA [34] in Figure 5-32(b), which we have previously used to perform ultrafast switching. We note that both the measured loss and gain are small compared to that in the Alcatel SOA, suggesting that we have high coupling losses into the active waveguide. This would result in a smaller voltage response in both directions.
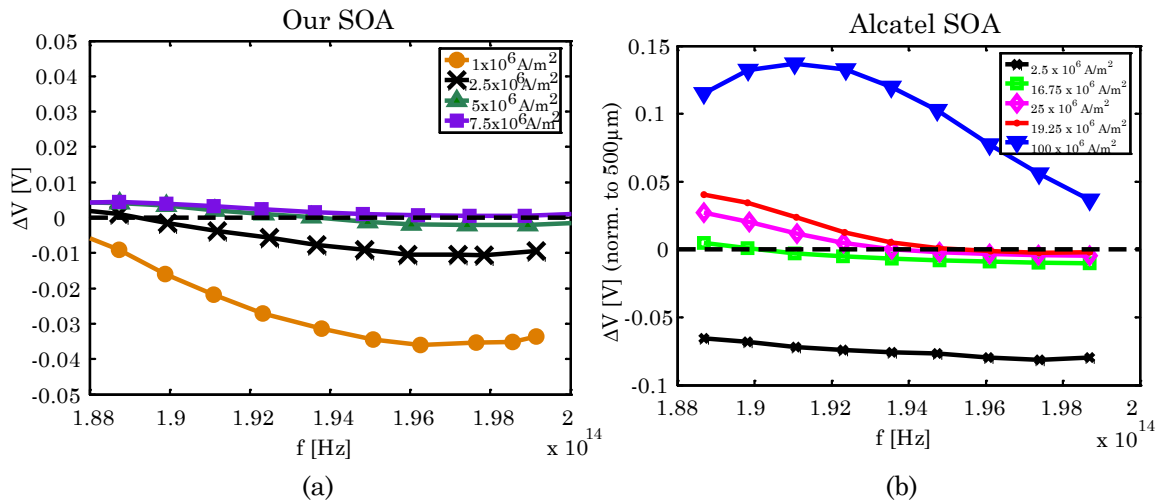


Figure 5-32: (a) Plot of our SOA loss and gain over various currents and wavelengths. (b) Plot of loss and gain in an Alcatel SOA over various currents and wavelengths. We normalize the length of the Alcatel SOA to the length of our SOA.

### 5.3.4   Conclusions and Future Directions

Our progress thus far in characterizing our SOA-MZI design has shown us that there are several areas of improvement needed before ultrafast switching can be demonstrated. First, we need to reduce the coupling losses between the active and passive waveguides. This can be accomplished by fabricating contacts on top of the tapers as well as the SOA devices. Second, we need to increase the thickness of the contact to reduce electromigration issues and allow us to inject more current into the device and achieve more population inversion. This should give us higher gains and

nonlinearities, both of which are beneficial to ultrafast switches. Finally, we need to reduce the contact resistance between the contact and the diode in order to reduce local heating and power consumption. Better mechanical bonding between the contact and the chip as well as larger contact pads will also be useful for simplifying the testing apparatus. Furthermore, the addition of a simple active waveguide will allow us to separately characterize the loss in the active material. Currently, T. Shih has implemented many of these suggestions and is in the process of fabricating the next set of chips for testing.

## 5.4   Future of Integrated Photonics

In this chapter, we have described progress in our on-going collaboration with the Integrated Photonic Materials and Devices Group at MIT to design, fabricate, and test our own SOA-MZI devices. This will allow us to experiment with optimization of SOA devices for ultrafast switching and to eventually develop integrated multi-gate devices for optical logic demonstrations.

Integration of optical devices is critical for reducing costs of optical networks as bandwidth demands increase. Ideally, electronic and optical devices will be integrated on the same chips, which can decrease the cost of optical-electronic-optical conversions and allow us to use the strengths of each technology easily. The difficulty lies in the fundamentally passive nature of silicon at optical wavelengths, which were optimized for low-loss fiber transmission. Hybrid techniques tend to result in higher costs due to alignment and complex fabrication processes. As a result, much research is currently focused on developing active silicon optical devices as well as in methods for lowering hybrid integration costs.

Ultimately, we need to develop a low-cost, standardized monolithic platform which can easily be used for any combination of integrated optical devices such as filters, isolators, amplifiers, lasers, and photodetectors. One major challenge to this is the difficulty of integrating devices which have fundamentally different requirements. Furthermore, the complexity of optical devices means that device yields remain low com-

pared with electronic fabrication. Finally, another major challenge is in maintaining a low power density. Power densities in electronic devices are already reaching the threshold of thermal cooling abilities and are currently a limiting factor in further miniaturization [1, 2]. For good performance in optical logic gates, high bias currents are often required. Such powers will need to be brought down before optical logic gates can be used to supplement the electronic processors currently used in networks today.

# Bibliography

[1] G. Epps, "System power challenges," in *Cisco Routing Research Symposium*, Aug. 2006.

[2] C. Minkenberg, R. P. Luijten, F. Abel, W. Denzel, and M. Gusat, "Current issues in packet switch design," in *ACM SIGCOMM Computer Communication Review*, vol. 33, pp. 119–124, 2003.

[3] G. P. Agrawal and N. K. Dutta, *Semiconductor Lasers*. Van Nostrand Reinhold, second ed., 1993.

[4] M. Lipson, "Guiding, modulating, and emitting light on silicon — challenges and opportunities," *Journal of Lightwave Technology*, vol. 24, 2005.

[5] U. Fischer, T. Zinke, J.-R. Kropp, F. Arndt, and K. Petermann, "0.1 dB/cm waveguide losses in single-mode SOI rib waveguides," *IEEE Photonics Technology Letters*, vol. 8, no. 5, pp. 647–648, 1996.

[6] W. Lin, C. J. Sun, and K. M. Schmidt, "Hybrid integration platform based on silica-on-silicon planar lightwave circuit," in *Photonics West 2007*, p. 6476, 2007.

[7] T. Hashimoto, Y. Nakasuga, Y. Yamada, H. Terui, M. Yanagisawa, Y. Akahori, Y. Tohmori, K. Kato, and Y. Suzuki, "Multichip optical hybrid integration technique with planar lightwave circuit platform," *Journal of Lightwave Technology*, vol. 16, pp. 1249–1258, 1998.

[8] H.-J. Yeh and J. S. Smith, "Fluidic self-assembly for the integration of GaAs light-emitting diodes on Si substrates," *IEEE Photonic Technology Letters*, vol. 6, no. 6, pp. 706–708, 1994.

[9] A. Greiner, J. Lienemann, J. G. Korvink, X. Xiong, Y. Hanein, and K. F. Bohringer, "Capillary forces in micro-fluidic self-assembly," in *Technical Proceedings of the 2002 International Conference on Modeling and Simulation of Microsystems*, p. 764, 2002.

[10] J. Rumpler, J. M. Perkins, and C. G. F. Jr., "Optoelectronic integration using statistical assembly and magnetic retention of heterostructure pills," in *Conference on Lasers and Electro-Optics (CLEO '04)*, p. CThT2, 2004.

[11] G. Maxwell, "Hybrid integration technology for high functionality devices in optical communications," in *Optical Fiber Communications Conference (OFC '08)*, 2008.

[12] P. Zakynthinos, D. Petrantonakis, D. Apostolopoulos, A. Poustie, G. Maxwell, and H. Avramopoulos, "Four-wavelength 3R burst mode regenerator using three integrated quad MZI arrays," in *Optical Fiber Communication Conference (OFC)*, 2008.

[13] G. Maxwell, B. Manning, M. Nield, M. Harlow, C. Ford, M. Clements, S. Lucas, P. Townley, R. McDougall, S. Oliver, R. Cecil, L. Johnston, A. Poustie, R. Webb, I. Lealman, L. Rivers, J. King, S. Perrin, R. Moore, I. Reid, and D. Scrase, "Very low coupling loss, hybrid-integrated all-optical regenerator with passive assembly," in *28th European Conference on Optical Communication (ECOC '02)*, 2002.

[14] H. Park, A. W. Fang, O. Cohen, R. Jones, M. J. Paniccia, and J. E. Bowers, "Design and fabrication of optically pumped hybrid silicon-AlGaInAs evanescent lasers," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 12, no. 7, pp. 1657–1663, 2006.

[15] H. Park, A. W. Fang, O. Cohen, R. Jones, M. J. Paniccia, and J. E. Bowers, "A hybrid AlGaInAs-silicon evanescent amplifier," *IEEE Photonic Technology Letters*, vol. 19, no. 4, pp. 230–232, 2007.

[16] A. W. Fang, H. Park, Y. Kuo, R. Jones, O. Cohen, D. Liang, O. Raday, M. N. Sysak, M. J. Paniccia, and J. E. Bowers, "Hybrid silicon evanescent devices," *Materials Today*, vol. 10, no. 7, 2007.

[17] H. Park, A. W. Fang, R. Jones, O. Cohen, O. Raday, M. N. Sysak, M. J. Paniccia, and J. E. Bowers, "A hybrid AlGaInAs-silicon evanescent waveguide photodetector," *Optics Express*, vol. 15, no. 10, pp. 6044–6052, 2007.

[18] T. Brenner, E. Gini, and H. Melchior, "Low coupling losses between InP/InGaAsP optical amplifiers and monolithically integrated waveguides," *IEEE Photonics Technology Letters*, vol. 5, no. 1, pp. 212–214, 1993.

[19] J. Leuthold, J. Eckner, C. Holtmann, R. Hess, and H. Melchior, "All-optical 2x2 switches with 20-dB extinction ratios," *Electronics Letters*, vol. 32, no. 24, pp. 2235–2236, 1996.

[20] A. Markina, *Design and Simulation for the Fabrication of Integrated Semiconductor Optical Logic Gates*. PhD dissertation, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2005.

[21] P. V. Studenkov, M. R. Gokhale, J. C. Dries, and S. R. Forrest, "Monolithic integration of a quantum-well laser and an optical amplifier using an asymmetric twin-waveguide structure," *IEEE Photonics Technology Letters*, vol. 10, no. 8, pp. 1088–1090, 1998.

[22] F. Xia, V. M. Menon, and S. R. Forrest, "Photonic integration using asymmetric twin waveguide (ATG) technology: Part 1–concepts and theory," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 11, no. 1, pp. 17–29, 2005.

[23] R. D. Williams, *Photonic Integrated Circuits for Optical Logic Applications*. PhD dissertation, Massachusetts Institute of Technology, Department of Materials Science and Engineering, 2007.

[24] J. Leuthold and C. H. Joyner, "Multimode interference couplers with tunable power splitting ratios," *Journal of Lightwave Technology*, vol. 19, no. 5, pp. 700–707, 2001.

[25] R. Giller, R. J. Manning, G. Talli, R. P. Webb, and M. J. Adams, "Analysis of the dimensional dependence of semiconductor optical amplifier recovery speeds," *Optics Express*, vol. 15, no. 4, pp. 1773–1782, 2007.

[26] B. W. Hakki and T. L. Paoli, "CW degredation at 300°K of GaAs double-heterostructure junction lasers II: Electronic gain," *Journal of Applied Physics*, vol. 44, no. 9, pp. 4113–4119, 1972.

[27] S. Adachi, *Physical properties of III-V semiconductor compounds: InP, InAs, GaAs, InGaAs, and InGaAsP*. Wiley, 1992.

[28] P. W. Leech, G. K. Reeves, W. Zhou, and P. Ressel, "Thermal stability of Pd/Zn and Pt based contacts to p-In$_0$.53Ga$_0$.47As/InP with various barrier layers," *Journal of Vacuum Science Technology B*, vol. 16, no. 1, pp. 227–231, 1998.

[29] T. C. Shen, G. B. Gao, and c. H. Morko "Recent developments in ohmic contacts for III-V compound semiconductors," *Journal of Vacuum Science Technology B*, vol. 10, no. 5, pp. 2113–2132, 1992.

[30] P. Ressel, P. H. Hao, M. H. Park, Z. C. Yang, L. C. Wang, W. Osterle, P. Kurpas, E. Richter, E. Kuphal, and H. L. Hartnagel, "Pd/Sb(Zn) and Pd/Ge(Zn) ohmic contacts on p-Type indium gallium arsenide: The employment of the solid phase regrowth principle to achieve optimum electrical and metallurgical properties," *Journal of Electronic Materials*, vol. 29, no. 7, pp. 964–972, 2000.

[31] A. G. Baca, F. Ren, J. C. Zolper, R. D. Briggs, and S. J. Pearton, "A survey of ohmic contacts to III-V compound semiconductors," *Thin Solid Films*, 1997.

[32] A. Katz, P. M. Thomas, S. N. G. Chu, W. C. Dautremont-Smith, R. G. Sobers, and S. G. Napholtz, "Pt/Ti ohmic contact to p$^{++}$-InGaAsP (1.3$\mu$m) formed by rapid thermal processing," *Journal of Applied Physics*, vol. 67, 1990.

[33] P. Andrekson, N. A. Olsson, T. Tanbun-Ek, R. A. Logan, D. Coblentz, and H. Temkin, "Novel technique for determining internal loss of individual semiconductor lasers," *Electronic Letters*, vol. 28, no. 2, pp. 171–172, 1992.

[34] J. Emery, T. Ducellier, M. Bachmann, P. Doussiere, F. Pommereau, R. Ngo, F. Gaborit, L. Goldstein, G. Laube, and J. Barrau, "High performance $1.55\mu$m polarisation insensitive semiconductor optical amplifier based on low-tensile-strained bulk GaInAsP," *Electronic Letters*, vol. 33, pp. 1093–1084, 1997.

# Chapter 6

# Conclusion

We have described the use of optical signal processing for routing and regeneration in data networks. As bandwidth demands continue to increase, networks must provide higher capacities at lower costs. With the electronic techniques currently used in data networks, these requirements are becoming increasingly difficult to meet due to cost, power consumption, and size/weight limitations. The relatively slow speed of electronic switching (10 Gb/s) forces a parallel architecture which significantly increases cost, power consumption, and size. Optical/electronic converters are also required at every step, further increasing costs and power consumption.

Electronic processing is very effective at parallel processing, random access memory, and complex computing requiring millions of logic gates. However, many functions in data networks do not require the unique advantages of electronic processing and may benefit from the unique advantages of optical processing techniques. These advantages include: ultrafast switching (100 Gb/s and higher), power consumption which scales well with increasing bit rates, elimination of O/E/O converters, wavelength-flexibility, and transparency to bit-rate and modulation format.

In this thesis, we investigated how optical signal processing techniques can supplement electronic methods in data networks. We described some of the challenges we need to meet with all-optical signal processing techniques. For practical implementation in real-world networks, these optical gates must have a small footprint, ultra-low power consumption, and competitive costs with electronic methods. Furthermore,

these optical gates must be simple to operate and optimize. Finally, multi-gate, cascaded operation must be demonstrated to expand the functionality of all-optical techniques.

We began by investigating the use of optical processing in routers, which need to perform header processing, routing, forwarding, and contention resolution. Optical signal processing can be used to eliminate the majority of the optical-electronic conversions and provide a transparent payload path by replacing the electronic switch matrices with optical or electro-optical switch matrices. In this case, header processing can be accomplished either electronically, to take advantage of electronic memory for storing routing tables, or optically, to reduce the processing latency. In the former case, the main challenge lies in providing long buffers for buffering optical packets while the headers are processed. In the latter case, slow routing table lookups are pushed to label edge routers and each core router only needs to process a short label. However, for either solution, contention resolution remains a challenge for optical techniques. In this thesis, we demonstrated the capability of all-optical techniques for performing header processing for two incoming 40-Gb/s packet streams. By demonstrating accurate forwarding using only two optical logic gates, we show that all-optical processing is capable of scalable routing functionality at ultra-fast rates.

To take full advantage of the potential for low-latency header processing, and to provide a practical implementation for real-world networks, integrated gates must be designed, optimized, and demonstrated. We chose to focus on SOA-based logic gates for their strong nonlinearity, versatility, and ease of integration. The SOA-MZI design has been demonstrated to operate in a variety of configurations and provide ultrafast AND, INVERT, and XOR functionality. Conceptually, it is identical to the UNI, which has been further demonstrated as a NOR, and XOR gate. We developed a method to quickly optimize operation of the SOA-MZI gate and demonstrated its effectiveness in using the SOA-MZI as a wavelength converter.

With this tool, we investigated the use of optical signal processing in regeneration. As channel bit rates increase to alleviate bandwidth demands, regenerator

reach decreases due to added sensitivity to interference, dispersion, and nonlinearities. Alternative modulation formats can extend regenerator reach, while optical switching can potentially reduce regenerator costs by eliminating O/E/O conversions and reducing the need for parallel processing by providing ultrafast switching capable of operating at channel bit rates. For all-optical regeneration to compete with electronic techniques, optical logic gate installation and operating costs must be reduced. In this thesis, we addressed the need for scalable, ultrafast regeneration by demonstrating the use of a simple all-optical regenerator consisting only of a single SOA-MZI and an additional SOA as wavelength converter with operation at 10 Gb/s line rates.

Ultimately, an effective monolithic platform for integrating optical processing components will be required to fully realize a low-cost and scalable optical processing solutions. We focused on InP-based fabrication and described our progress in designing and fabricating a low-loss, multi-gate design in collaboration with the Integrated Photonic Materials and Devices group at MIT. We characterized the waveguide loss, active-passive coupling efficiencies, active waveguide performance, and provided recommendations for the next generation of fabricated devices.

## 6.1  Future Research Directions

Much work remains to be done before optical signal processing can be effective for supplementing electronic techniques in future data networks. First, there remain several critical functions which need to be developed before optical packet switching can become a reality. Furthermore, routers continue to come under pressure to provide more services, which means that optical techniques must supplement electronic capabilities or enable similar functionality. In addition, improvements in the performance of optical gates themselves are required, including improving the speed and power consumption of individual gates. Finally, as mentioned in Chapter 5, a cohesive monolithically-integrated platform needs to be developed to enable lower costs and practical operation.

### 6.1.1 Increasing Functionality

Several critical functions must still be demonstrated in a practical manner before optical packet switching can become a reality. First of all, as mentioned earlier, a compelling approach for contention resolution has yet to be fully realized. Optical buffering would require simple electronic or optical control logic, sufficient buffer depth, and effective performance for increasing throughput on a network level. Since buffering needs to occur at each input or output port on the router, a single buffer would ideally be a low-cost and integrated device. The largest challenge to optical buffering is in the long lengths required for storage due to fast propagation speeds. This results in distortions and loss which must both be compensated for. One example which uses a recirculating buffer is presented by J. Yang et al. [1]. An alternative approach focuses on avoiding contention by deflecting packets into an unused wavelength [2]. This has implications for packet routing which have yet to be addressed. An alternative approach is to investigate hybrid strategies in which simple optical logic is used to detect contention but contending packets are stored in the electrical domain. For low levels of contention, this can still provide speed improvements to routers. One such approach was taken by Takahashi et al. [3].

Another key technology required is tunable delay lines, which are needed to provide synchronization on both packet and bit time scales. These delay lines need to be tunable over many nanoseconds, to provide packet-level synchronization, or hundreds of picoseconds, to provide bit-level synchronization. Furthermore, we require low losses, fast tuning times, and either electrical or optical control. Currently, controllable delay lines are implemented through mechanically moving mirrors to increase propagation distances through free space or by stretching fiber. Thermal effects can also be used to change the length or index of refraction of a material, but generally do not provide a significantly large delay. Discrete-tunable delays can also be achieved by using a switch over different lengths of fiber, or via a tunable wavelength conversion followed by a wavelength-dependent propagation time. Integration of a variable delay will also be necessary to lower costs and decrease footprint.

Finally, we also need to demonstrate clock recovery for regeneration. Integration of a Fabry-Perot clock-recovery filter along with the SOA-MZI used for regeneration will make for a compelling and compact 3R regenerator.

In addition to pressure to significantly increase capacity, networks also come under pressure to provide more services to satisfy the increasingly diverse customer base. Network management also becomes increasingly complex as networks grow and become more interconnected. Optical techniques must either supplement these new demands or provide similar functionality. This requires more research into both increasing the functionality of optical processing techniques as well as research into better methods of integrating optical techniques with existing electronic capabilities. The availability of stable, integrated optical logic gates may aid in demonstrating more complex processing than can be achieved by 1-3 logic gates. One example of this is the relatively recent rise of Quality of Service (QoS) functionality. With increasing video traffic, this will become more of a priority, since it introduces significantly different requirements for different types of video traffic. For instance, delays on the order of seconds are very costly for live video traffic but largely inconsequential for email traffic. Thus, routers need to distinguish between types of traffic in order to optimize network performance. Optical signal processing techniques have yet to address this question. Simple optical header processing logic could potentially be used to determine packet priorities and affect routing behavior.

## 6.1.2  Improving Switch Operation

Research is still required for improving the operation of a single optical switch. The key requirements for optical switches are ultrafast and low-power performance. As described in Chapter 3, the fundamental limitation for SOA switching is simply the carrier recovery time of the switch. Quantum dot SOAs have been measured with carrier recovery times on the order of 15 ps [4, 5], which suggests their potential for ultrafast switching. Fabrication of quantum dots, however, is complex and provides low yields. Another possibility involves highly nonlinear fiber such as bismuth oxide glass [6] or chalcogenide glass [7, 8]. These doped glasses are passive, and rely only

on a nearly-instantaneous $\chi^{(3)}$ effect. Thus, switching times can be extremely fast. The challenge here is to provide a strong enough nonlinearity such that very short lengths and low powers can be used. These high nonlinearities often coexist with large dispersion, which makes it difficult to maintain undistorted pulses in the fiber. Furthermore, alignment and splicing of sections of bismuth oxide or chalcogenide glass is difficult. Potentially, these doped fibers can be integrated into a silica-on-silicon integration platform, though that has yet to be investigated in depth. Finally, photonic crystals offer another option for creating highly nonlinear materials with ultrafast recovery times through achieving very slow group velocities, which allows for increased phase shifts [9]. Fabrication and integration of photonic crystal devices remains a challenge, however.

As power dissipation and consumption are becoming major limitations for electronic routing and regeneration, research on ultra-low power switches will also be necessary. For SOA-based logic, this means reducing waveguide losses and increasing coupling efficiencies. Furthermore, more efficient carrier pumping methods can be researched to create lower-power SOA biases. Passive switching devices which do not require electronic biasing may provide a better solution. Both nonlinear doped glasses and photonic crystals are possibilities. Relatively high conversion efficiencies in four wave mixing have been achieved in passive InP-based waveguides [10] and may offer a solution. To compensate for loss in these passive switches, we will require highly efficient linear amplification as well. Ideally, such amplifiers will be integrated and capable of amplification over multiple wavelengths.

## 6.1.3  Integration Platform

As discussed briefly in Chapter 5, we ultimately wish to develop a low-cost, standardized monolithic platform which can be used for any combination of optical devices including fixed and tunable filters, modulators, isolators, fixed and tunable delay lines, amplifiers, lasers (both tunable and fixed), and photodetectors. Filters are required to manage noise in the optical chip as well as for copropagating SOA-MZI operation. Tunable filters can be used to select different wavelengths for a more flexible network

architecture. Modulators are needed to modulate a data pattern onto the optical pulse train. Isolators can reduce the impact of reflections and noise from backward-propagating waves in a multi-gate logic chip. Delay lines will be necessary to time the signal and control pulses into the optical logic switch. Tunable delay lines are necessary for synchronization of data packets and bits. Amplification is required to compensate for the inevitable losses in the chip. Lasers are required to generate optical packets, and photodetectors are required to eventually convert the optical data into the electronic domain.

Much work in this direction has already been accomplished. However, there is not yet a single cohesive platform where all these different functionalities have been demonstrated. Some researchers have focused on developing the field of silicon optics, aiming for a monolithic platform in silicon which can be easily integrated with CMOS technologies for electronic circuitry. This approach can also allow us to significantly reduce the cost of O/E/O conversions, and thus more easily use optical and electronic techniques for their respective strengths. Effective static and tunable filters, passive couplers and waveguides, and modulators have all been designed and demonstrated on a silicon photonics platform [11, 12]. Fixed delay lines using a resonant ring design have also been demonstrated [13]. However, lasers and amplifiers remain a challenge due to the indirect bandgap of silicon. Thus far, active amplification and lasing in silicon largely relies on optical pumping using Raman amplification, an inefficient process [14, 15, 16]. If this challenge can be overcome, silicon would clearly be the most effective platform for photonic integration.

An alternative focus has been on using an InP-based platform, which is very effective for active devices but more lossy than silicon. InP fabrication also tends to produce lower yields and is a less mature process. This also ensures a different process from electronic circuitry. However, due to the lack of a efficient silicon laser, InP is currently the most feasible platform for monolithic integration of photonic devices. Integration of diode lasers and photodetectors with amplifiers and modulators has been achieved commercially by Infinera [17], thus significantly reducing the cost of O/E/O conversion. Passive devices such as filters have been fabricated by either

etching air gaps to create Fabry-Perot filters [18] or fabricating ring resonators with SOAs to compensate for losses [19]. The major challenge for InP-based integration is in reducing losses and increasing yields.

There are still a few critical devices which have not yet been successfully integrated. Isolators, for instance, remain discrete devices due to the difficulty of creating a Faraday rotator to create non-reciprocal polarization rotation in either InP or silicon. Current research in integrated isolators has largely focused on incompatible materials such as YIG. One potential solution has been proposed by Guo et al. [20] and involves using magnetically doped semiconductors to achieve this polarization rotation. Fixed and tunable delay lines also remain difficult to integrate due to the need for long lengths and the means to significantly and quickly change the lengths or the index of refraction in the material. For packet-level synchronization, delays must be able to tune over nanosecond scales. Bit-level synchronization has less strict requirements, requiring only picosecond-scale tuning. Much attention has recently been paid towards slow light experiments, which uses the resonance of a material to allow us to change the index of refraction significantly and induce a large delay. Thus far, however, the delays induced require significant experimental apparatus and do not extend much further than a pulsewidth.

## 6.2   Conclusion

Ultimately, the need for capacity at low costs will means that we must adopt alternative technologies. Optical signal processing remains a promising path for development. With this thesis, we have demonstrated significant progress in the implementation of all-optical signal processing for data networks. In this concluding chapter, we provided a brief overview of the interesting research issues still remaining in this field for the benefit of future researchers.

# Bibliography

[1] J. Yang, B. Xiang, T. He, A. Karalar, X. Ye, and S. J. B. Yoo, "All-optical contention resolution with TTL-Aware selective 3R regeneration in optical-label switching router networks," in *Optical Fiber Communication Conference (OFC)*, 2008.

[2] L. Stampoulidis, E. Kehayas, D. Apostolopoulos, P. Bakopoulos, K. Vyrsokinos, and H. Avramopoulos, "On-the-fly all-optical contention resolution for NRZ and RZ data formats using packet envelope detection and integrated optical switches," *IEEE Photonics Technology Letters*, vol. 19, no. 8, pp. 538–540, 2007.

[3] R. Takahashi, T. Nakahara, K. Takahata, H. Takenouchi, T. Yasui, N. Kondo, and H. Suzuki, "Photonic random access memory for 40-Gb/s 16-b burst optical packets," *IEEE Photonics Technology Letters*, vol. 16, no. 4, pp. 1185–1187, 2004.

[4] M. Sugawara, N. Hatori, M. Ishida, H. Ebe, Y. Arakawa, T. Akiyama, K. Otsubo, T. Yamamoto, and Y. Nakata, "Recent progress in self-assembled quantum-dot optical devices for optical telecommunication: temperature insensitive 10 gb s$^{-1}$ directly modulated lasers and 40 gb s$^{-1}$ signal-regenerative amplifiers," *Journal of Physics D: Applied Physics*, vol. 38, pp. 2126–2134, 2005.

[5] A. J. Zilkie, J. Meier, P. W. E. Smith, M. Mojahedi, P. J. Poole, C. N. Allen, P. Barrios, and D. Poitras, "Femtosecond gain and index dynamics in an InAs/InGaAsP quantum dot amplifier operating at 1.55 $\mu$m," *Optics Express*, vol. 23, no. 14, pp. 11453–11459, 2006.

[6] J. H. Lee, T. Nagashima, T. Hasegawa, S. Ohara, N. Sugimoto, and K. Kikuchi, "Four-wave-mixing-based wavelength conversion of 40-Gb/s nonreturn-to-zero signal using 40-cm bismuth oxide nonlinear optical fiber," *IEEE Photonic Technology Letters*, vol. 17, no. 7, pp. 1474–1476, 2005.

[7] R. E. Slusher, G. Lenz, J. Hodelin, J. Sanghera, L. B. Shaw, and I. D. Aggarwal, "Large raman gain and nonlinear phase shifts in high-purity as$_2$se$_3$ chalcogenide fibers," *Journal of the Optical Society of America B*, vol. 21, no. 6, pp. 1146–1155, 2004.

[8] K. Ogusu, J. Yamasaki, and S. Maeda, "Linear and nonlinear optical properties of Ag-As-Se chalcogenide glasses for all-optical switching," *Optics Letters*, vol. 29, no. 3, pp. 265–267, 2004.

[9] M. Soljačić, S. G. Johnson, S. Fan, M. Ibanescu, E. Ippen, and J. D. Joannopoulos, "Photonic-crystal slow-light enhancement of nonlinear phase sensitivity," *Journal of Optical Society of America B*, vol. 19, no. 9, pp. 2052–2059, 2002.

[10] A. M. Darwish, E. P. Ippen, H. Q. Le, J. P. Donnelly, and S. H. Groves, "Optimization of four-wave mixing conversion efficiency in the presence of nonlinear loss," *Applied Physics Letters*, vol. 69, no. 6, pp. 737–739, 1996.

[11] M. Lipson, "Guiding, modulating, and emitting light on silicon — challenges and opportunities," *Journal of Lightwave Technology*, vol. 24, 2005.

[12] T. Barwicz, H. Byun, F. Gan, C. W. Holzwarth, M. A. Popović, P. T. Rakich, M. R. Watts, E. P. Ippen, F. X. Kärtner, H. I. Smith, J. S. Orcutt, R. J. Ram, V. Stojanovic, O. O. Olubuyide, J. L. Hoyt, S. Spector, M. Geis, M. Grein, T. Lyszczarcz, and J. U. Yoon, "Silicon photonics for compact, energy-efficient interconnects," *Journal of Optical Networking*, vol. 6, no. 1, pp. 63–73, 2007.

[13] F. Xia, L. Sekaric, and Y. Vlasov, "Ultracompact optical buffers on a silicon chip," *Nature Photonics*, vol. 1, pp. 65–71, 2007.

[14] R. L. Espinola, J. I. Dadap, R. M. O. Jr., S. J. McNab, and Y. A. Vlasov, "Raman amplification in ultrasmall silicon-on-insulator wire waveguides," *Optics Express*, vol. 12, no. 6, pp. 3713–3718, 2004.

[15] O. Boyraz and B. Jalali, "Demonstration of directly modulated silicon Raman laser," *Optics Express*, vol. 13, no. 3, pp. 796–800, 2005.

[16] H. Rong, R. Jones, A. Liu, O. Cohen, D. Hak, A. Fang, and M. Paniccia, "A continuous-wave raman silicon laser," *Nature*, vol. 433, pp. 725–728, 2005.

[17] M. Kato, R. Nagarajan, J. Pleumeekers, P. Evans, A. Chen, A. Mathur, A. Dentai, S. Hurtt, D. Lambert, P. Chavarkar, M. Missey, J. Bäck, R. Muthiah, S. Murthy, R. Salvatore, C. Joyner, J. Rossi, R. Schneider, M. Ziari, F. Kish, and D. Welch, "40-channel transmitter and receiver photonic integrated circuits operating at a per channel data rate 12.5 Gbit/s," in *Optical Fiber Communications Conference (OFC '07)*, p. 1, 2007.

[18] M. Strassner, J. Daleiden, N. Chitica, D. Keiper, B. S. lnacke, S. Greek, and K. Hjort, "III-V semiconductor material for tunable fabry-perot filters for coarse and dense WDM systems," *Sensors and Actuators A: Physical*, vol. 85, no. 1, 2000.

[19] D. G. Rabus, M. Hamacher, U. Troppenz, and H. heidrich, "Optical filters based on ring resonators with integrated semiconductor optical amplifiers in GaInAsP-InP," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 8, no. 6, pp. 1405–1411, 2002.

[20] X. Guo, T. Zaman, and R. J. Ram, "Magneto-optical semiconductor waveguides for integrated isolators," in *Proceedings of the SPIE*, pp. 152–158, 2005.

# Appendix A

# Dynamic Bias Scan Data

For those unable to access the multimedia content of this thesis, we present here all the frames of Figure 3-28 from Chapter 3. This set of figures illustrates the effect of the control pulse power on the interferometer bias of the SOA-MZI gate at every signal-control delay. In each frame, (a) shows the interferometer bias scan at the signal delay as labeled. (b) shows the relative position of the signal and control pulses, where the signal pulse is shown in blue and the control pulse is shown in red. (c) shows the corresponding switching window measurement at point "X" on the interferometer bias scan.
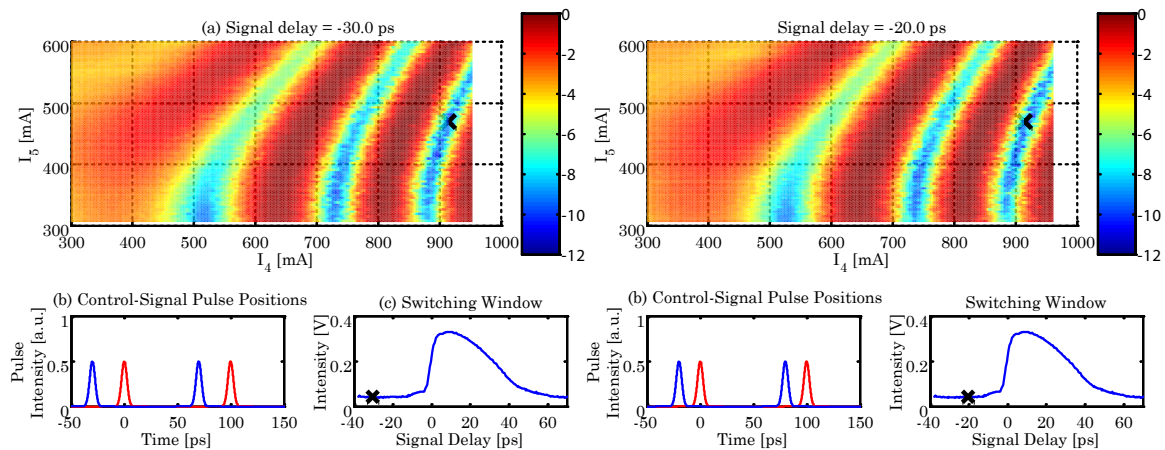
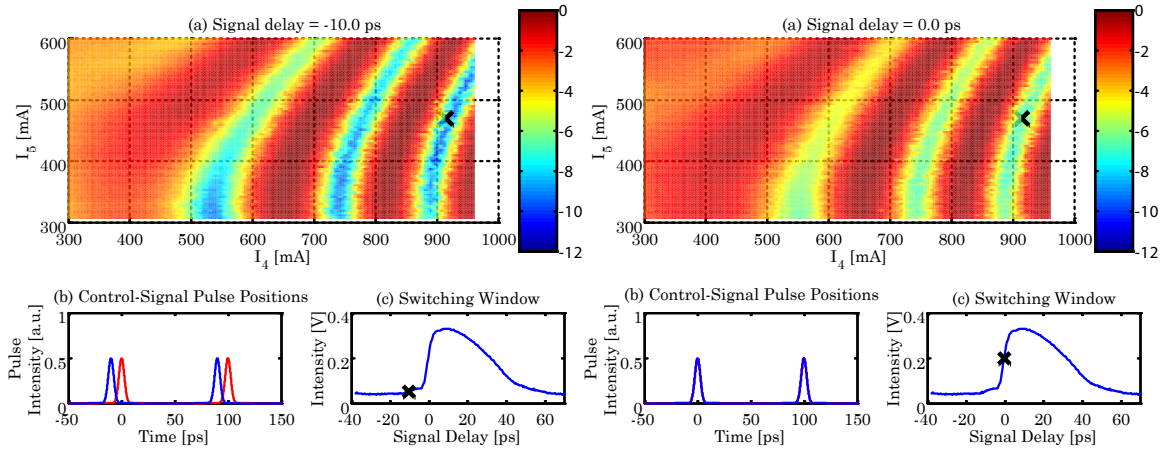Figure A-1: Frames corresponding to a signal delay of -30 ps and -20 ps.

Figure A-2: Frames corresponding to a signal delay of -10 ps and 0 ps.
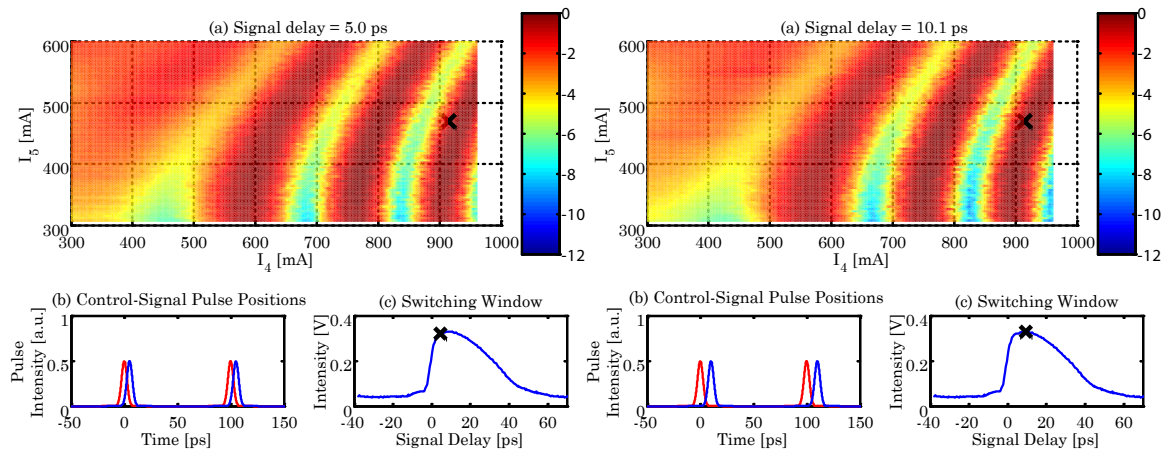


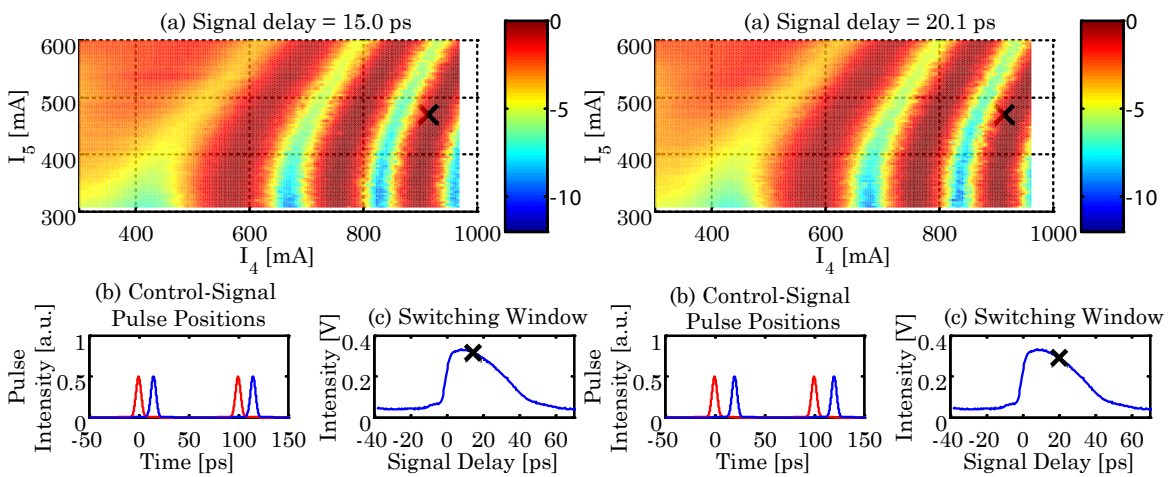Figure A-3: Frames corresponding to a signal delay of 5 ps and 10.1 ps



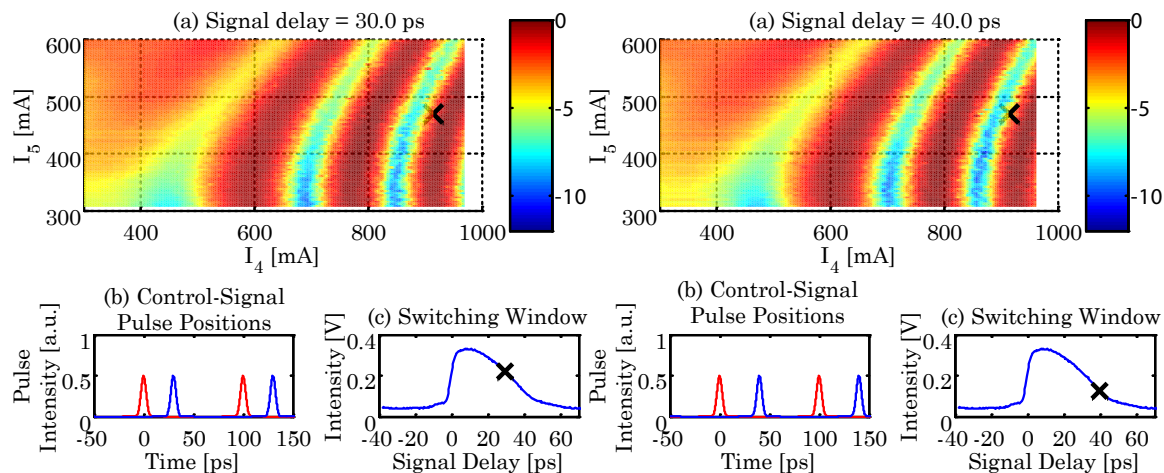Figure A-4: Frames corresponding to a signal delay of 15 ps and 20.1 ps

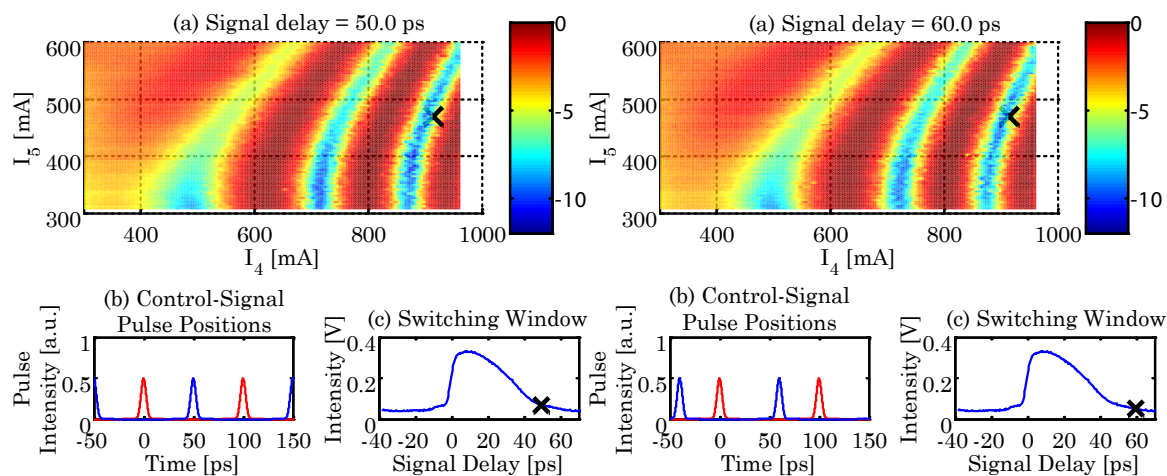Figure A-5: Frames corresponding to a signal delay of 30 ps and 40 ps



Figure A-6: Frames corresponding to a signal delay of 50 ps and 60 ps
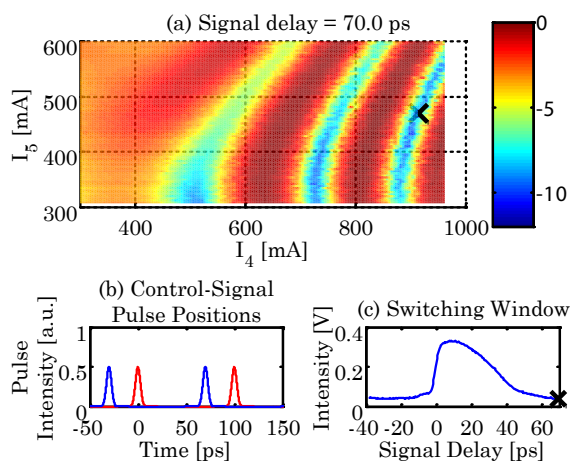


Figure A-7: Frame corresponding to a signal delay of 70 ps.

# Appendix B

# SOA-MZI Wafer Composition

We include here the layer structure of the asymmetric twin waveguide design for the SOA-MZI logic gate in Chapter 5. Figure B shows the cross-section of the passive and active waveguides. The composition of each layer is described in Table B.
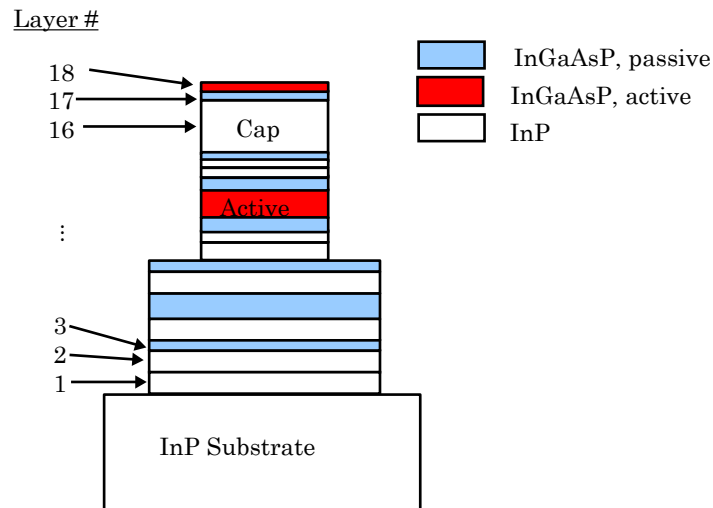
Figure B-1: Cross-section of the SOA-MZI layers.

| Layer | Material | x | y | PL (nm) | Thickness ($\mu$m) | Doping (cm$^{-3}$) | Dopant |
|---|---|---|---|---|---|---|---|
| 18 | GaIn(x)As(y)P | 0.56 | 0.94 | 1550 $\pm$ 50 | 0.05 | >2x10$^{18}$ | Zn |
| 17 | GaIn(x)As(y)P | 0.80 | 0.45 | 1180 $\pm$ 50 | 0.03 | >1x10$^{18}$ | Zn |
| 16 | InP | | | | 0.60 | 2.5-10 x 10$^{17}$ | Zn |
| 15 | GaIn(x)As(y)P | 0.80 | 0.45 | 1180 $\pm$ 50 | 0.01 | 2.5x10$^{17}$ | Zn |
| 14 | InP | | | | 0.05 | 2.5x10$^{17}$ | Zn |
| 13 | InP | | | | 0.05 | none | N/A |
| 12 | GaIn(x)As(y)P | 0.80 | 0.45 | 1180 $\pm$ 50 | 0.10 | none | N/A |
| 11 | GaIn(x)As(y)P | 0.56 | 0.94 | 1550 $\pm$ 50 | 0.20 | none | N/A |
| 10 | GaIn(x)As(y)P | 0.80 | 0.45 | 1180 $\pm$ 50 | 0.10 | none | N/A |
| 9 | InP | | | | 0.05 | none | N/A |
| 8 | InP | | | | 0.15 | 5x10$^{17}$ | S |
| 7 | GaIn(x)As(y)P | 0.80 | 0.45 | 1180 $\pm$ 50 | 0.10 | 5x10$^{17}$ | S |
| 6 | InP | | | | 0.20 | 5x10$^{17}$ | S |
| 5 | GaIn(x)As(y)P | 0.80 | 0.45 | 1180 $\pm$ 50 | 0.25 | 5x10$^{17}$ | S |
| 4 | InP | | | | 0.20 | 5x10$^{17}$ | S |
| 3 | GaIn(x)As(y)P | 0.80 | 0.45 | 1180 $\pm$ 50 | 0.10 | 5x10$^{17}$ | S |
| 2 | InP | | | | 0.20 | 5x10$^{17}$ | S |
| 1 | InP | | | | 0.30 | 5x10$^{17}$ | S |

Table B.1: Composition of layers in asymmetric twin waveguide design for SOA-MZI devices