UNIVERSITY OF CALIFORNIA

Santa Barbara

Edge Interoperability for High-Performance Optical Core Network Routers

A Dissertation submitted in partial satisfaction of

the requirements for the degree

Doctor of Philosophy in Electrical and Computer Engineering

by

John M. Garcia

Committee in charge:

Professor Daniel J. Blumenthal, chair

Professor John E. Bowers

Professor Nadir Dagli

Doctor Brian R. Koch

June 2013

UMI Number: 3596137

UMI

Dissertation Publishing

UMI 3596137

ProQuest®

The dissertation of John M. Garcia is approved.

_____

Professor Daniel J. Blumenthal

_____

Professor John E. Bowers

_____

Professor Nadir Dagli

_____

Doctor Brian R. Koch

June 2013

Edge Interoperability for High-Performance Optical Core Network Routers

John M. Garcia

Acknowledgments

There are several individuals that contributed, in the form of mentorship and friendship, to the work presented in this dissertation. I first and foremost would like to thank and acknowledge my mother and stepfather. Although they may not exactly understand what I have been working on these past years, I hope they know that their hard work and sacrifices have shaped me into the person I am today.

There was a sequence of fortuitous events (still unknown to me) that landed my parents in Tulsa, Oklahoma, where I was eventually born. Shortly after my birth, my parents decided to leave Oklahoma and head back to Mexico to be closer to family. It was then that the relationship between my mother and my biological father would become contentious and would eventually end in divorce after the birth of my younger brother, Omar. The three of us lost contact with my father and spent the following years traveling back and forth between the my grandmother's and Aunt Martha's (on my mother's side) home as my mother looked for work. We would eventually settle in my aunt's home where they ran a business dedicated to fabricating religious sculptures. My mother would perform odd jobs ranging from sweeping floors to hand painting sculptures to make sure show appreciation for their hospitality. Several years afterward, my uncles (Guillermo and Jesus) were making plans to return to the US to find work. At that instant, my mother decided it was time to return to the US so that my brother and I could obtain a decent education.

When I was about 7, my mother, my brother (4 years) and myself returned to the US and settled in Santa Maria, California. Nowadays, Santa Maria is known for their tri-tip and the signature Santa Maria Grill, but back then it was better know then for its agricultural industry. My mother was able to

find work as a farm worker in a broccoli farm while I attended elementary school. Anyone that has ever worked in the agricultural industry will attest to the unforgiving work conditions and the minimal wages. Six days a week, my mother would go to work at six in the morning and be back home by five in the evening to make sure we had a decent meal for dinner. She was a farm worker for over ten years and all she ever asked of us is to make sure we excelled in school so that we would not have to be farm workers ourselves. There were low points in my graduate career where I contemplated taking the easy way out and walking away. It was during those moments where her unrelenting work ethic reminded me that no matter how "difficult" the situation became, it cannot compare to what she has endured. I am greatful for her hard work, and this dissertation is dedicated to her.

I would not be the man I am today if it were not for my stepfather, Salvador. Our lives changed for the better when my mother met him while working in a lettuce and strawberry farm. Even as a child, I remember him being quite the stand up guy. He was respectful, a hard worker, didn't drink or smoke, and was always reading encyclopedias (I know, who reads encyclopedias?!). Since then, I have always had the utmost respect for the person he is and the way he carries himself. It takes a remarkable man to not only marry a single mother with 4- and 7-year old sons, but to also raise them as his own. In fact, one of the fondest memories I have of him involves him teaching me long multiplication and division while on a "date" with my mother. He and my mother would eventually have a daughter (Andrea) who he loves as much as he did my brother and myself. I am very fortunate to be his son, and I hope that this dissertation serves as some form of appreciation of him being an amazing father.

At no fault of their own, my parents were unable to become involved in my

academic studies. Their long work shifts made it difficult to attend parent-teacher conferences, and at a certain point they were unable to directly help with things like math homework and college applications. Luckily, I was fortunate enough to have extraordinary high school teachers that excelled at their craft. Mrs. Gaylen Clarke taught the most enjoyable math class I have ever taken. I would not be as adept (inept?) in French if it wasn't for Madame Marianne Angel who was patient enough to endure my sense of humor. Mr. Ben Wieman, a fellow Gaucho, unknowingly put me on the path towards an engineering career with his physics class. My favorite problem he ever assigned involved calculating the velocity of a freshman hitting the ground after being thrown off a helicopter hovering 300 feet in the air. I would also like to acknowledge the counselors in the College Center (Eric Blanco and Vicky Trejo) who helped this clueless individual with the college application process.

I would like to also acknowledge my two best friends from high school Daniel Macias and Juan Lopez, who share my twisted sense of humor. Even though we sometimes go years without hanging out, we still manage to pick up right were we left off.

During my time as an undergrad at UCSB I met and worked with several people who contributed to this dissertation. The year and half at I spent as a dishwasher at two of UCSB dining commons taught me how to balance work, school, and social life. It also taught me that very few things, including grad school, can be worse than working a dish room during the lunch and dinner rush hours. I would like to acknowledge Computer Science professors such as Eliot Jacobson and Fred Carlin for assigning the epiphanic programming projects that taught me how to program. I realized if one can implement a graphing calculator in MIPS assembly, then one can implement it in any

programming language. The guys from the Del Norte House were the best and possibly the worst mentors one could have: Alex B., Adam, Ferni, Chris, Cris, Luis, Josh, Captain Chonie, Alex S., and Ronnie. All the intramural basketball teams I was a part of played a great role in helping me keep my sanity. We had some rough, winless seasons, but we managed to improve to the point where we reached the championship round. I am also thankful for the time I invested in campus organizations such as Los Ingenieros (LI) and the National Society of Black Engineers (NSBE).

I would like to thank my advisor Dan Blumenthal for giving me the opportunity to perform cutting edge research within his group. His guidance and support over the years has made me a better student and researcher. I would also like to thank the rest of my committee members. It has been a pleasure to work with John Bowers and his research group under the DARPA-funded LASOR project. He and Dan have fostered a special collaborative relationship not only between themselves, but also between their groups. Working with the Bowers Group has been one of my favorite experiences in graduate school. I would also like to thank Professor Dagli for his guidance and support. It is no coincidence that the courses where I was most challenged and engaged happened to be taught by him. I would also like to acknowledge Doctor Brian Koch for taking the time to serve on my committee. I was under his tutelage during my time at Intel, and learned a great deal under him. My experience at Intel was a great learning experience and he had a direct hand in that.

There are several members (past and present) of the Blumenthal Group that have been outstanding colleagues and friends. First, I would like to acknowledge my mentor Henrik Poulsen. I had the pleasure of working alongside Henrik as far back as Junior year in undergrad. After working with him,

This dissertation is dedicated to my parents,
Guadalupe & Salvador

Vita of John M. Garcia

Email: johngarcia@ece.ucsb.edu

EDUCATION

Doctor of Philosophy in Electrical and Computer Engineering

Institution: University of California, Santa Barbara, CA 93106, U.S.A.

Date Awarded: June 2013

Master of Science in Electrical and Computer Engineering

Institution: University of California, Santa Barbara, CA 93106, U.S.A.

Date Awarded: June 2009

Bachelor of Science in Computer Science

Institution: University of California, Santa Barbara, CA 93106, U.S.A.

Date Awarded: June 2007

High School Diploma

Institution: Santa Maria High School, Santa Maria, CA 93454, U.S.A.

Date Awarded: June 2003

TECHNICAL EXPERIENCE

Graduate Researcher (Sept. 2007 – June 1013)

ECE Department, UC Santa Barbara, CA 93106, U.S.A.

Hybrid Silicon Laser Testing Intern (Aug. 2011 – Feb. 2012)

Intel Corporation, Santa Clara, CA 95054, U.S.A.

Software Engineer Intern (July 2007 – Sept. 2007)

Raytheon Electronic Warfare Systems, Goleta, CA 93117, U.S.A

Undergraduate Researcher (April 2006 – June 2007)

ECE Department, UC Santa Barbara, CA 93106, U.S.A.

Web Developer / Programmer (April 2005 – April 2007)

ECE Department, UC Santa Barbara, CA 93106, U.S.A.

JOURNAL PUBLICATIONS

1. J.M. Garcia, K.N Nguyen, T. Huffman, J.P. Mack, E.F. Burmeister, G. Kurczveil, J.S. Barton, H.N. Poulsen, D.J. Blumenthal, "A High-Performance Edge Router Demonstrating Buffering, Forwarding, and 3R Regeneration of Labeled 40Gb/s Optical Packets," to be submitted to *Optics Express*, June 2013.

2. D.J. Blumenthal, J. Barton, N. Behesti, J.E. Bowers, E. Burmeister, L.A. Coldren, M. Dummer, G. Epps, A. Fang, Y. Ganjali, J.M. Garcia, B. Koch, V. Lal, E. Lively, J. Mack, M. Masanovic, N. McKeown, K. Nguyen, S.C. Nicholes, H. Park, B. Stamenic, A. Tauke-Pedretti, H. Poulsen and M. Sysak, "Integrated Photonics for Low-Power Packet Networking," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 17(2), pp. 458-471, March 2011.

3. T. Kise, K.N. Nguyen, J.M. Garcia, H.N. Poulsen, and D.J. Blumenthal, "Cascadability properties of MZI-SOA-based all-optical 3R regenerators for RZ-DPSK signals," *Optics Express*, vol. 19, pp. 9330-9335 (2011).

4. G. Kurczveil, M.J.R. Heck, J.D. Peters, J.M. Garcia, D. Spencer and J.E. Bowers, "An Integrated Hybrid Silicon Multiwavelength AWG Laser," *Selected Topics in Quantum Electronics, IEEE Journal of* , vol. 17, Issue 6, pp.1-7, November 2011.

5. J.P. Mack, E.F. Burmeister, J.M. Garcia, H.N. Poulsen, B. Stamenic, G. Kurczveil, K.N. Nguyen, K. Hollar, J.E. Bowers and D.J. Blumenthal, "Synchronous Optical Packet Buffers," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 16(5): pp. 1413-1421, September 2010.

CONFERENCE PAPERS

1. J.M. Garcia, K.N. Nguyen, T. Huffman, M. Belt, J. S. Barton, H.N. Poulsen, and D.J. Blumenthal, "Demonstration of Edge Interoperability, Re-Shaping and Re-Timing using Hybrid Mode-Locking within a 40Gb/s Optical Packet Router," *Proc. Optical Fiber Communications Conference, OFC '13*, paper OTh4D, March 2013.

2. J.M. Garcia, H.N. Poulsen, D.J. Blumenthal, "Demonstration of End-to-End Interoperability between Legacy 100MbE and a 40Gb/s Optical Label Switched Network Layer," *Optical Communication (ECOC), 2011 36th European Conference and Exhibition on*, September 2011.

3. K.N. Nguyen, J.M. Garcia, E. Lively, H.N. Poulsen, D. M. Baney, and D.J. Blumenthal, "Monolithically Integrated Dual-Quadrature Coherent Receiver on InP with 30 nm Tunable SG-DBR Local Oscillator," *Optical Communication (ECOC), 2011 36th European Conference and Exhibition on*, September 2011.

4. J.M. Garcia, H.N. Poulsen, D.J. Blumenthal, "An Adaptation Layer for Real-Time Interoperability Between Legacy 100MbE and 40Gb/s (and Beyond) Optical Label Switched Networks," *Photonics Society Summer Topical Meeting Series, 2011 IEEE*, July 2011.

5. K.N. Nguyen, T. Kise, J.M. Garcia, H. Poulsen, and D.J. Blumenthal, "All-Optical 2R Regeneration of BPSK and QPSK Data using a 90° Optical Hybrid and Integrated SOA-MZI Wavelength Converter Pairs," *Proc. Optical Fiber Communications Conference, OFC '11*, paper OMT3, 2011.

6. T. Kise, K.N. Nguyen, J.M. Garcia, H. Poulsen, and D.J. Blumenthal, "Demonstration of Cascadability and Phase Regeneration of SOA-Based All-Optical DPSK Wavelength Converters," *Proc. Optical Fiber Communications Conference, OFC '11*, paper OThY3, 2011.

7. G. Kurczveil, M. J. R. Heck, J.M. Garcia, H.N. Poulsen, H. Park, D.J. Blumenthal, and J.E. Bowers, "Integrated recirculating optical hybrid silicon buffers," *Proc. SPIE*, paper 79430U, 2011.

8. J.M. Garcia, J.P. Mack, K.N. Nguyen, K. Hollar, E. F. Burmeister, B. Stamenic, G. Kurczveil, H.N. Poulsen, J.E. Bowers and D.J. Blumenthal, "A Real-Time Asynchronous Dynamically Re-Sizable Optical Buffer for Variable Length 40Gbps Optical Packets," *Proc. Optical Fiber Communication Conference, OFC '10*, Paper OThN4, March 2010.

9. J.P. Mack, K.N. Nguyen, J.M. Garcia, E. F. Burmeister, M. M. Dummer, H.N. Poulsen, B. Stamenic, G. Kurczveil, K. Hollar, L. A. Col-

dren, J.E. Bowers and D.J. Blumenthal, "Asynchronous 2x2 Optical Packet Synchronization, Buffering, and Forwarding," *Proc. Optical Fiber Communications Conference, OFC '10*, Paper OThN1, March 2010.

10. G. Kurczveil, M. J.R. Heck, J. D. Peters, J.M. Garcia, J.E. Bowers, "A Fully Integrated Hybrid Silicon AWG Based Multiwavelength Laser," *IEEE International Semiconductor Laser Conference*, Paper WA 4, September 2010.

11. J.P. Mack, J.M. Garcia, H.N. Poulsen, E. F. Burmeister, B. Stamenic, G. Kurczveil, J.E. Bowers and D.J. Blumenthal, "End-to-end asynchronous optical packet transmission, scheduling, and buffering," *Proc. Optical Fiber Communication Conference, OFC '09*, Paper OWA2, March 2009.

Abstract

Edge Interoperability for High-Performance Optical Core Network Routers

by

John M. Garcia

Traffic in today's backbone networks is dominated by Internet protocol (IP) applications and is continuously increasing. The increasing demand is driven by video streaming, cellular phone technologies, and the vast amounts of information stored in data centers. Traffic volumes exceeding 1PB ($> 1,000$ Terabytes) per business day have been reported and are expected to increase by more than tenfold over the next decade. Current electronic core network routers rely on a store-and-forward technique that requires optical-electronic-optical (OEO) conversions of high-speed data in order to route packets. This technique requires each bit of incoming data to be processed by high-speed electronics which tend to be rather power intensive and do not scale efficiently with the increasing trends in traffic demand. Optical data routers (ODRs) and optical packet switching (OPS) are currently being investigated as potentially scalable, energy efficient alternatives to current electronic routers. The OPS scheme utilizes a packet format consisting of a low-speed header followed by high-speed payload. The routing of such a packet is performed by processing the header via low-speed electronics that are relatively inexpensive in terms of cost and power dissipation while allowing the payload to be transparently processed entirely in

the optical domain. Adaptation layers are required at the edge of core optical networks to facilitate the interoperability between current electronic and future optical networks. Successful router demonstrations have relied heavily on fixed-length packet formats, which require edge adaptation layers to perform computationally complex fragmentation algorithms to break larger variable-length packets into smaller fixed-length packets. The fragmentation process is viewed as detrimental to performance because of its inherent latency penalties and by the fact that re-transmission of entire datasets is required if individual fragments are lost.

This dissertation presents enabling technology required for low-latency, low packet loss, interoperability between today's electronic networks and future optical networks. Adaptation between legacy 100 megabit Ethernet (MbE) and 40Gb/s OPS formats is successfully achieved by utilizing custom, FPGA-based Edge Adaptation Layers that demonstrate latencies below 300ns and packet loss rates less than 10^{-5}. Latency penalties are minimized by forgoing the fragmentation process at the edge and enforcing variable-length packet compatibility within optical core routers. A re-sizable optical buffer is designed and implemented to enable all-optical routers to accommodate packet lengths ranging from 40 to 800 bytes with less than 5% packet loss. This dissertation culminates in an end-to-end ODR link demonstration utilizing an energy efficient FPGA-based electronic control and UCSB-fabricated photonic integrated circuits (PICs) housed in custom package sub-mounts and FPGA-based driving circuitry. The end-to-end link demonstration successfully achieves adaptation between 100MbE and 40Gb/s optical packet formats, optical packet buffering, forwarding, and 3R regeneration (re-amplification, re-shaping, and re-timing) of 40Gb/s optical packets at packet loss rates below 1%.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Information in today's networks consists primarily of user-generated Internet Protocol (IP) traffic that is continually increasing. This increasing demand is driven by on-demand video streaming, data center, and mobile services. In fact, the second largest wireless telecommunications provider in the United states recently reported it supports over 23PB (1PB = 1,024 terabytes) of data per business day [1]. Moreover, the regression analysis presented in [2] predicts a 12X increase of optical backbone traffic within the 2012-2022 time period. These gargantuan amounts of IP traffic are routed through current backbone networks using sophisticated electronic routers that utilize a store-and-forward scheme to route information. An opto-electronic (OE) conversion is performed on incoming IP traffic before each bit is stored in memory. Once routing is completed, packets are read from memory and an electro-optic (EO) conversion is performed to transmit IP packets over wavelength division multiplexed (WDM) optical links. Scaling current electronic routing technologies to meet the increasing network demand requires additional engineering to upgrade electronics towards faster switching speeds, which may be costly in terms of power and monetary consumption.

The optical communications community widely believes that energy consumption, rather than component cost, is the major detractor of continued growth [3]. As a result, all-optical photonic processing is being investigated as a means of potentially scaling core router bandwidth and capacity independently of power consumption. The growth of traffic demand was initially met in the late 1990s by leveraging WDM technology capable of transmitting over 10Tb/s of data across multiple densely spaced optical channels [4]. The next-generation core network technology will benefit from dynamic reconfiguration of all-optical router nodes where high-speed optical packets are forwarded while avoiding costly OEO data conversions. In this scenario, an optical routing fabric operates at switching speeds much lower than the data rate and is controlled by a low-speed electronic control plane that is inexpensive in terms of cost and power consumption. The ideal all-optical routing system should support arbitration of asynchronous, variable-length IP packets [5].

As it becomes feasible to deploy a core backbone consisting of all-optical routers, adaptation layers will be required at the core edges. The edge adaptation layers serve as a means of converting between the packet formats used in current electronic networks and future all-optical networks. This enables edge-to-edge communication between non-adjacent electronic networks where packet routing is solely performed via all-optical networks. Current optical router demonstrations showing end-to-end adaptation utilize fixed-length packet formats. As a result, the edge adaptation layers need to perform computationally complex fragmentation algorithms to break up varying-length packets into smaller fixed-length datagrams (and vice-versa). The (de-)fragmentation process is viewed as detrimental to throughput performance as it introduces computational latency penalties. Additionally, re-

2

transmission of entire data sets is often required if individual fragments are lost. Furthermore, previous work has yet to show forwarding of IP traffic through an all-optical core router capable of contention resolution and signal regeneration.

1.1   Summary and Thesis Organization

The work presented in this dissertation demonstrates enabling technology necessary for low-latency, low-packet-loss interoperability between current electronic networks and future optical networks. Edge-to-edge communication between non-adjacent electronic networks is carried out via an all-optical data router node capable of buffering, forwarding, and regenerating high-speed, variable-length optical packets.

A discussion of several optical transport paradigms is presented in Chapter 2. Here, the advantages and shortcomings of optical circuit switching (OCS), optical burst switching (OBS), and optical packet switching (OPS) are discussed to provide some rationale behind the fundamental design directions taken in this work. Chapter 3 discusses the background of previous interoperability demonstrations and how the work presented in this dissertation impacts optical communications. Chapter 4 presents functional descriptions for each building block used to implement the optical packet router. Here, a comparison of past demonstrations of each subsystem is carried out to provide some insight into the optical router architecture design chosen in this work. In Chapter 5, the basic operating principle behind high-performance edge interoperability is illustratively discussed along with its requirements and implementation challenges. Chapter 6 presents a detailed description of the adaptation layer architectural design and discusses how implementation

challenges are addressed. In Chapter 7, the characterization of the end-to-end adaptation process is presented. Here, the back-to-back packet loss and latency performance of each adaptation building block required for interoperability is evaluated. Chapter 8 presents the progress made towards demonstrating optical storage capable of accommodating variable-length packets. Here, a dynamically re-sizable, re-circulating optical packet buffer is successfully demonstrated with asynchronously arriving packets ranging from 40 to 800 bytes in length. Chapter 9 demonstrates the end-to-end adaptation performance of a 2x2 label swapped optical router. Here, the design and implementation aspects of each subsystem is shown along with their respective end-to-end performance results. Chapter 10 concludes the dissertation by summarizing the edge interoperability results obtained in this work and discussing their limitations. Additional work necessary to address performance limitations is also discussed to ensure that the technology presented here continues to advance.

References

[1] K.Rinne, "Building next generation mobility networks: Lessons from the Bonneville speedway," presented at *Optical Fiber Communications Conference, (OFC) 2011*, Plenary, 2011.

[2] S. K. Korotky, "Traffic Trends: Drivers and Measures of Cost-Effective and Energy-Efficient Technologies and Architectures for Backbone Optical Networks," *Optical Fiber Communication Conference (OFC) 2012*, paper OM2G.1, March 2012.

[3] G. Shen, R. S. Tucker, "Energy-Minimized Design for IP Over WDM Networks," *Optical Communications and Networking, IEEE/OSA Journal of*, vol.1, no.1, pp.176-186, June 2009.

[4] S. J. B. Yoo, "Optical Packet and Burst Switching Technologies for the Future Photonic Internet," *IEEE Journal of Lightwave Technology*, vol. 24, no. 12, December 2006.

[5] S. J. B. Yoo, F. Xue, Y. Bansal, J. Taylor, Z. Pan, J. Cao, M. Jeon, T. Nady, G. Goncher, K. Boyer, K. Okamoto, S. Kamei, V. Akella, "High-Performance Optical-Label Switching Packet Routers and Smart Edge Routers for the Next-Generation Internet," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 7, September 2003.

Chapter 2

Towards More Efficient Optical Communication Systems

This chapter presents the electronic routing technology used in current core networks and introduces the optical label swapping (OLS) technique as a potentially energy efficient alternative. An architecture overview of a conventional electronic router is shown along with its technical shortcomings. The reader is then introduced to three different transport paradigms that utilize OLS to route high-speed optical packets by avoiding costly optical-electro-optic (OEO) conversions. Significant reductions in power consumption may be obtained by employing optical switching fabrics that are configured via a low-speed electronic control plane. A description of circuit switching (OCS), optical burst switching (OBS), and optical packet switching (OPS) is then carried out in detail. The OCS transport paradigm is promptly ruled out as a potential high-utilization, next-generation all-optical routing technology since rather long setup and tear-down latencies are required to reserve edge-to-edge light paths. Within OBS, IP packets with common destinations are aggregated into optical payload bursts that are preceded by a low-speed

control header. OBS improves upon OCS throughput at the node-level by transmitting data in long, high-speed payload bursts. However, throughput is limited by latency penalties observed at the network edges resulting from payload aggregation and disassembly. In OPS, each IP packet is transported as a high-speed optical payload trailing a low-speed optical header. The OPS transport mechanism is ultimately selected in this work since it can potentially achieve higher throughput than OBS and OCS. However, OPS requires fast switching speeds with fine temporal resolution that may result in a complex electronic control at the routing node. A detailed discussion of several methods of OPS packet encoding is then presented with representative examples implementing time domain signaling, frequency division modulation (FDM), wavelength division modulation (WDM), and orthogonal modulation. The trade-offs between the differing methods are evaluated based on ease of implementation and signal quality. The packet encoding format used in this work makes use of the time domain method since it provides good spectral efficiency and high header-payload isolation. This method, however, calls for strict timing requirements for header extraction and processing within the OPS node fabric.

2.1   Current Electronic Packet Routers

The data in today's core network is forwarded using electronic packet routers that rely on a store-and-forward method that utilizes optical-electro-optical (OEO) conversions to successfully route high-speed data packets [1]. Each router consists of large chassis shelf housing several line cards with multiple network ports each capable of transporting 10s of gigabits of information per second (Gb/s). The front-end of each routing line card consists of high-speed

7

latching electronics that operate at a speed equal to the payload data rate (10s of Gb/s). Incoming data packets of varying sizes are fragmented into smaller, fixed-length packets and parallelized into several data lines operating at lower speeds (100s of MHz). Packets are then written to random access memory (RAM) before routing and forwarding algorithms are applied. Large banks of RAM are supplied to enable the router to perform arbitration and resolution of potential packet collisions. While in RAM, packet information is accessible to the router to enable it to carry out complex queuing and service oriented scheduling in addition to its core routing algorithms. The current state of the art electronic packet router is the Cisco CRS-3 which touts 2.24Tb/s of routing capacity via sixteen 40Gb/s line cards. The chassis occupies $213{\times}60{\times}91\text{cm}^3$, weighs 777kg and consumes 16.8kW of power when fully stocked and cooled [2].

2.1.1 Electronic Core Router Evolution

The plot in Figure 2.1, provided courtesy of Garry Epps of Cisco Systems, shows the evolution of core router performance over several years. One can observe that the system bandwidth of a router in 1993 was about 1Gb/s while it only consumed about 1kW of power. In 2010, system bandwidth exceeded several terabits per second (Tb/s) while exceeding 10kW of system power. Operating frequency-to-power ratio of silicon CMOS transistors has evolved from 10MHz/W in 1998 to 100MHz/W in 2009 by deceasing the gate size from 250nm to 45nm respectively. It is becoming increasingly difficult to obtain simultaneous improvement in transistor switching speed while maintaining constant power dissipation since the rate of transistor scaling has begun to slow down. As CMOS devices are miniaturized, it is difficult to avoid the power dissipation resulting from carrier tunneling and leakage
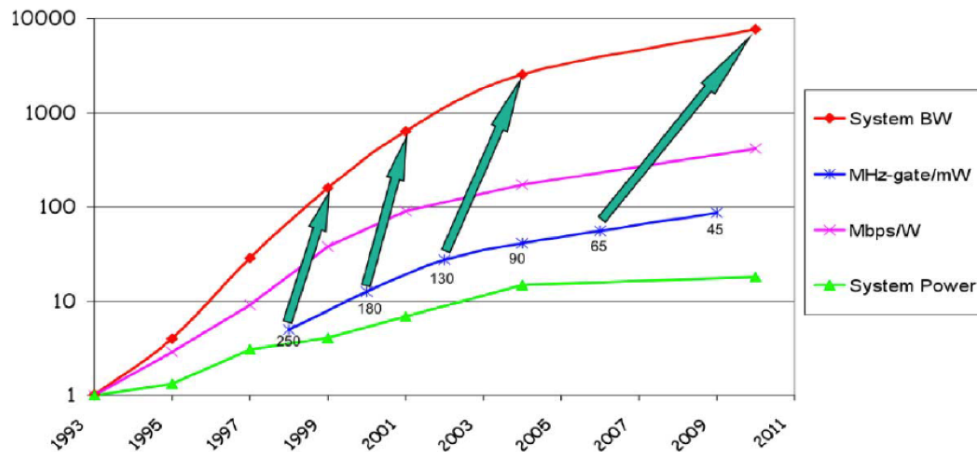
Figure 2.1: *Evolution of electronic core routers (courtesy of Garry Epps of Cisco Systems).*

currents. Development in electronic router technology has reached a point where it is not feasible to achieve a consistent increase in bandwidth while maintaining relatively constant system power consumption. In fact, field engineers are deploying clustered router configurations to keep pace with the growing demand in network bandwidth. Therefore, an alternative technology is required to mitigate the observed scaling limitations of bandwidth and power dissipation of conventional electronic core routers.

2.2 All-Optical Transport Paradigms

Several optical transport paradigms are under investigation as a means of meeting the increasing demand of high-utilization core network routers. Optical circuit switching (OCS), optical burst switching (OBS), and optical packet switching (OPS) in conjunction with optical label swapping (OLS) are viewed as potential next-generation optical routing methodologies employing Internet Protocol-over-wavelength division multiplexed (IP-over-WDM)
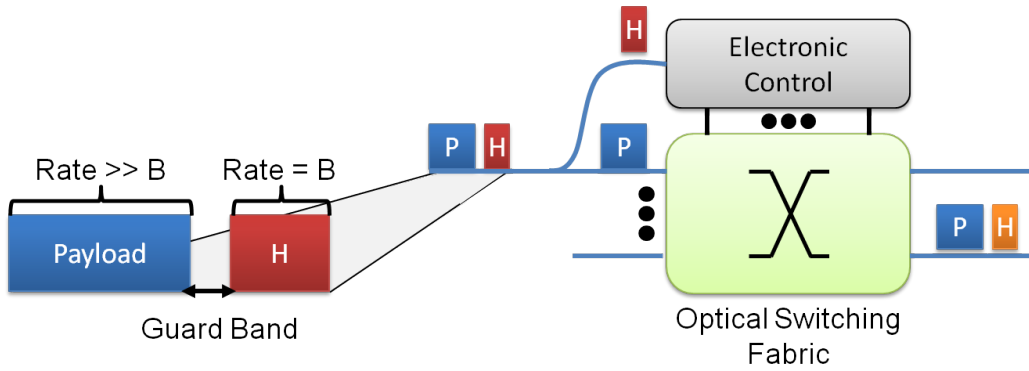
Figure 2.2: *Diagram illustrating operating principle behind optical label swapping (OLS).*

transportation services. The next-generation switching technology should provide high bandwidth and throughput utilization, low configuration latencies, and scalability among other requirements.

2.2.1 Optical Label Swapping (OLS)

Optical label swapping (OLS) is a technique used to route high-speed packets entirely in the optical domain through reconfigurable switching fabrics that are controlled via slow-switching electronics. Figure 2.2 shows the basic operation of a label-swapped switching fabric. In OLS, optical packets are comprised of a labeled header (H) and a high-speed payload (P). The header contains an optical label, implemented as a short sequence of bits, that represents a packet's desired output port and output wavelength. The header is extracted from incoming packets and is processed by low-speed electronics that are used to configure the optical switching fabric. The payload is then transparently switched towards its destination and a new labeled header is written to the packet. The manner in which the header extraction and rewrite is carried out will depend on how the packet is encoded and transported. The

10
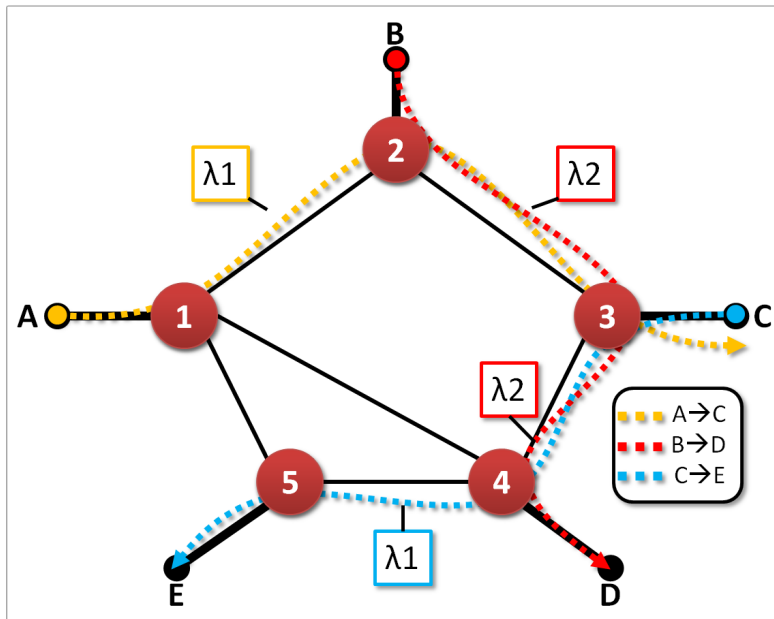
Figure 2.3: *An optical circuit switching (OCS) WDM network consisting of routing nodes interconnected by point-to-point fiber optic links [3].*

following sections will discuss the performance advantages and shortfalls of OLS implemented via the OCS, OBS, and OPS transport paradigms.

2.2.2 Optical Circuit Switching (OCS)

Optical circuit switching (OCS) is a technique that was introduced to transmit the growing Internet traffic using high-capacity IP over WDM architectures [3,4]. The OCS scheme functions in a fashion very similar to telephonic circuit switching. A source-to-destination data path is established on a per-request-basis to facilitate the transfer of information. The connection remains open for the duration of the transfer, and is immediately terminated when no longer in use. Figure 2.3 shows the point-to-point connections utilized in an OCS wavelength-routed WDM network. Network nodes are interconnected via WDM links that serve as the transport medium of optical payloads. Each

connection is assigned a specific path and wavelength such that shared WDM links are configured to disparate wavelengths. Demonstrations using wavelength conversion [5] and deflection [6] have been utilized in an attempt to enhance contention performance.

2.2.3 Optical Burst Switching (OBS)

Optical burst switching (OBS) allows one to transport large amounts of data via reconfigurable, all-optical switching fabrics. Unlike OCS, bursts of data are routed without requiring long-standing, end-to-end optical circuit connections. At the ingress of an OBS network, electronic edge routers store and hold incoming data packets until enough packets with identical destinations can be collected. The aggregated packets are used to assemble high-speed optical payload bursts that are preceded by a low-speed labeled header. Optical router nodes then use the information stored in the header to allocate a forwarding datapath. At the egress of the OBS network, edge routers perform the function of disassembling the payload bursts before transmitting them through an electronic channel.

Figure 2.4 shows a qualitative timing diagram of the OBS packet format. The OBS format was originally envisioned within a WDM implementation, where an out-of-band wavelength channel is reserved for means of router node control and configuration (top) while the remaining channels are comprised of payload bursts (bottom) consisting of multiple IP packets [7]. The header is extracted and processed using a low-speed control plane, while the payload bursts are switched entirely in the optical domain. The Control Header (CH) specifies the destination, wavelength channel, and duration of an incoming payload burst. A payload burst is transmitted immediately after the CH separated by a pre-calculated guard band (GB) delay. The header-to-payload

12
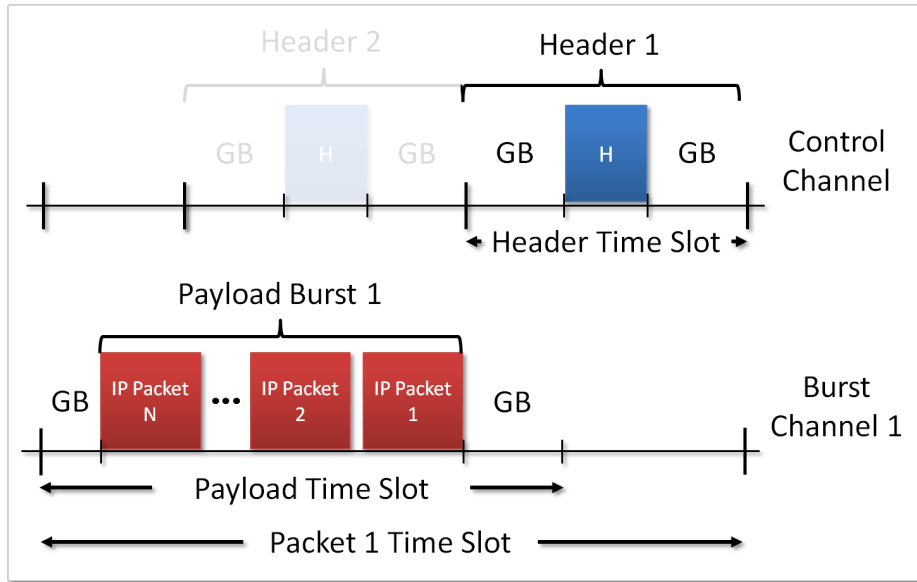
Figure 2.4: *Diagram of packet format used in optical burst switching (OBS) showing labeled low-speed header (H), high-speed payload bursts, and guard bands (GB).*

GB accounts for the extraction time required by the electronic control. An additional GB is inserted after the Nth burst to account for the longest transient observed within the optical switching fabric.

Before transmitting the payload burst, the CH is sent via an idle access link. The CH is received by an OBS node and is used to configure a light path in order to route a payload burst towards its desired destination. The burst arrives at the node shortly after, while the CH is forwarded to a separate node down the line. This process is repeated until the burst reaches its desired destination. Several OBS transmission protocols have been previously discussed to facilitate transmittance of OBS packets. In [8], Duser *et al.* used a tell-and-wait (TAW) method that required an acknowledgment (ACK) of resource allocation before transmitting the payload burst. In [7], Turner *et al.* used a tell-and-go (TAG) method that transmits the burst after a
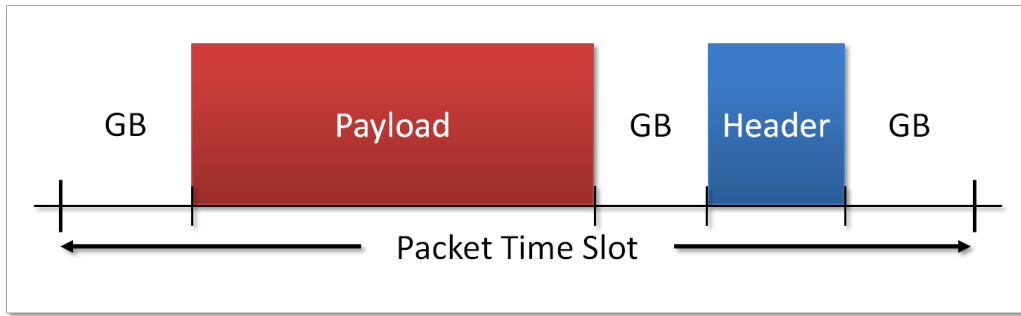
Figure 2.5: *Diagram of packet format used in optical packet switching (OPS) showing low-speed labeled header, high-speed payload, and guard bands (GB).*

predetermined time offset dictated by the processing time of the electronic control plane. The processing delay can be embedded into the header-payload guard band or it can be accounted for within the routing node. If the switch is in use and it is not possible to allocate the necessary path and bandwidth, the contention can be resolved by means of optical buffers or deflection towards a separate wavelength or output port.

2.2.4   Optical Packet Switching (OPS)

The packet format used in OPS is demonstrated in Figure 2.5. The format consists of a low-speed header followed by a high-speed payload. The routing and forwarding information is located within the header, which is extracted and processed using low-speed electronics that are relatively inexpensive in terms of monetary cost and power dissipation. This allows the high-speed payload to be processed transparently and completely in the optical domain using a switching fabric that is operated at a rate much lower than the payload data rate. A finite guard band (GB) of idle transmission is placed between the header and payload to allow the electronic control of an

14

OPS routing node ample time to extract and process the forwarding information stored in the header. Moreover, leading and trailing guard bands are inserted to account for the setup and hold times of the optical fabric that are dominated by inherent switching transients in the optical routing fabric.

2.2.5 Transport Paradigm Comparisons

When using OCS, each path request consists of a circuit setup and tear-down that requires the static allocation of bandwidth resources for the entire duration of the connection. As a result, OCS is ideal for applications such as voice where traffic is expected to be relatively continuous. However, the circuit setup and tear-down latencies result in router operation that is not optimized for data-driven, burst mode traffic.

The utilization of OBS allows one to obtain the benefits of OCS while avoiding some of its shortcomings. The bandwidth utilization is considerably higher since data is transmitted in multiple payload bursts as soon as the resources of a single node become available. Here, one does not need to suspend transmission while end-to-end communication is in progress. Though contention resolution within an OBS node is possible, it is primarily performed within the edge routers as IP packets are stored in electronic memory before being aggregated into payload bursts. On the other hand, the assembly and disassembly of payload bursts burdens the edge adaptation layers with additional latencies that curb throughput as optical channels remain idle while IP packets are aggregated into a payload burst.

The truest implementation of IP-over-WDM is achieved by employing OPS. Additionally, this technique allows one to achieve the highest possible throughput since node configuration occurs on a packet-by-packet basis and edge routers do not incur packet aggregation latencies. The main shortfall

of this approach is that it requires strict timing with minimal guard bands to reduce overhead and maximize throughput. Nanosecond-scale switching speeds with sub-nanosecond resolution accuracies are crucial for operation, which are likely to increase the complexity and cost of the electronic control plane.

The work presented in this dissertation makes use of the OPS transport paradigm. The OPS and OBS techniques are relatively comparable except OPS requires heightened timing constraints while OBS enhances the edge router complexity. The work presented in later chapters will demonstrate an electronic control plane capable of achieving nanosecond-scale switching accuracy and speeds while incurring minimal penalties in power consumption. This dissertation shows the implementation of low-latency edge adaptation layers in conjunction with an optical core node that satisfies the strict OPS timing requirements.

2.3 Optical Packet Encoding Formats

Having opted for the OPS routing scheme, one needs to select a suitable optical packet encoding format that allows a single header-payload pair to be transmitted per packet time slot. The more prominent packet encoding methods can be categorized as either serial or parallel. The main serial header-payload encoding method takes advantage of time domain multiplexing (TDM), while frequency division modulation (FDM), wavelength division modulation (WDM), and orthogonal modulation are commonly used encoding schemes that are classified as parallel formats.
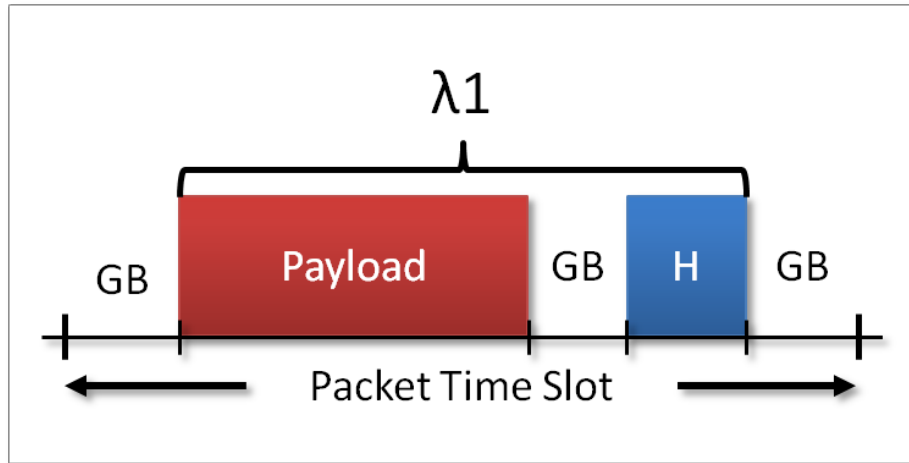
Figure 2.6: *Time domain method of encoding OPS packets showing header and payload on same optical channel with guard bands (GB).*

2.3.1 Time Domain

Initial OPS demonstrations have utilized encoding schemes where the header is transmitted immediately before the payload over a single optical wavelength, as shown in Figure 2.6. In [9], Ha *et al.* demonstrated a 2×2 all-optical packet switch using a 100Mb/s header followed by a 700Mb/s payload in bit-serial fashion with 240ns of inter-packet guard bands. More recently, Mack *et al.* demonstrated a 2×2 buffered optical switch that utilized 10Gb/s headers followed by 40Gb/s payloads with 43.2ns packet guard bands [10].

2.3.2 Frequency Division Modulation

The sub-carrier multiplexing (SCM) technique, shown in Figure 2.7, is one of the more prominent FDM methods used in header-payload encoding. The SCM technique consists of transmitting the header and payload as parallel radio frequencies on a single optical channel. The payload is modulated at the based band while the header is mixed with a sub-carrier whose fre-
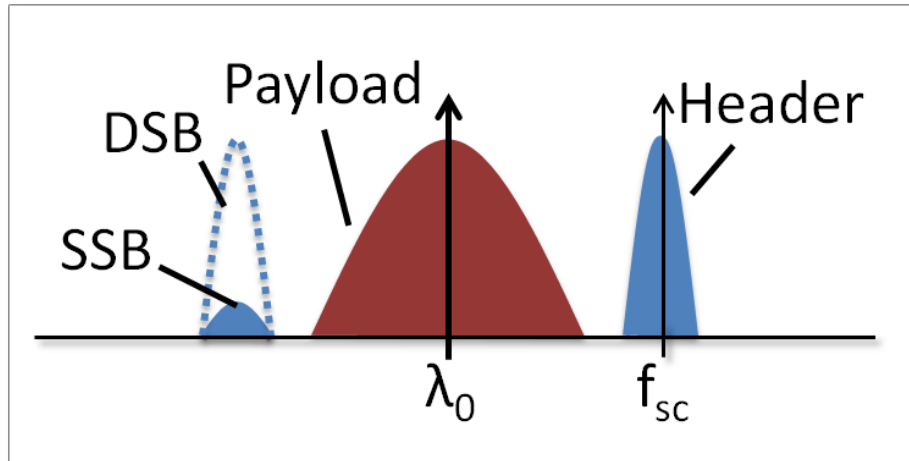
Figure 2.7: *Packet encoding method utilizing frequency division modulation (FDM) showing sub-carrier multiplexing (SCM) on single channel (λ_0), double sideband (DSB) and single sideband (SSB) regimes, and sub-carrier frequency (f_{SC}).*

quency (f_{SC}) spacing is determined by the data rate of the payload. The baseband and sub-carrier are then combined and used to modulate an optical signal. Initial all-optical switch demonstrations utilized optical double-sideband (ODSB) techniques using FDM header encoding. In [11], Budman *et al.* demonstrated error-free operation of a 1×2 optical packet switch where a 40Mb/s NRZ header was combined with a 3GHz carrier and the payload was modulated at a 2.56Gb/s NRZ baseband over a 1310nm optical channel. Shortly after, Zhu *et al.* demonstrated header extraction and rewrite of optical packets consisting of 155Mb/s headers and 10Gb/s payloads. The header was mixed with a 14GHz sub-carrier and combined with the payload baseband frequency before being transmitted on a 1549.2nm optical signal [12]. More recently, a routing node utilizing the double sideband SCM technique was demonstrated with 155Mb/s headers mixed with a 17.8GHz sub-carrier

while the payload was present on the 10Gb/s baseband [13].

Carrier-suppressed SCM has also been utilized in the SCM encoding scheme in order to improve spectral efficiency and CD performance. In [14], Xiao *et al.* used hyperfine optical blocking filter to convert obtain an optical single sideband (OSSB) modulated signal that resulted in carrier side band suppression greater than 20dB. In [15, 16], Smith *et al.* introduced a novel technique using an external Mach-Zehnder modulator (MZM) to obtain an OSSB modulated signal that is able to reduce CD degradations without the use of dispersion compensation. The technique involves driving a dual-electrode MZM at quadrature with an RF signal that is split and applied to each electrode with a $\pm\dfrac{\pi}{2}$ phase shift between each arm.

2.3.3 Wavelength Division Modulation

Figure 2.8 demonstrates an out-of-band signaling method used to encode OPS packets where the header and payload are transmitted on separate wavelength channels. Initial packet switching demonstrations of WDM packet encoding involved transporting 155Mb/s headers at 1300nm and 933Mb/s payloads at 1550nm [17]. Shortly after, Okada *et al.* described all-optical packet routing in an AWG-based network using out-of-band packet signaling where 1Gb/s headers and payloads are encoded onto datacomm (1300nm) and telecomm (1550nm) optical wave bands. With this format, header extraction is easily performed via wavelength selective elements such as optical notch filters and fiber Bragg gratings.
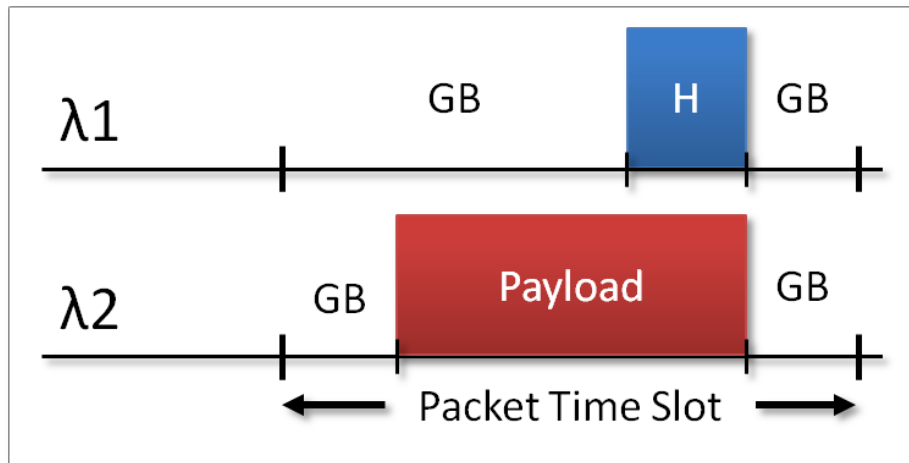
Figure 2.8: *Diagram of OPS encoding scheme using wavelength division modulation (WDM) showing header and payload on separate optical wavelengths.*

2.3.4 Orthogonal Modulation
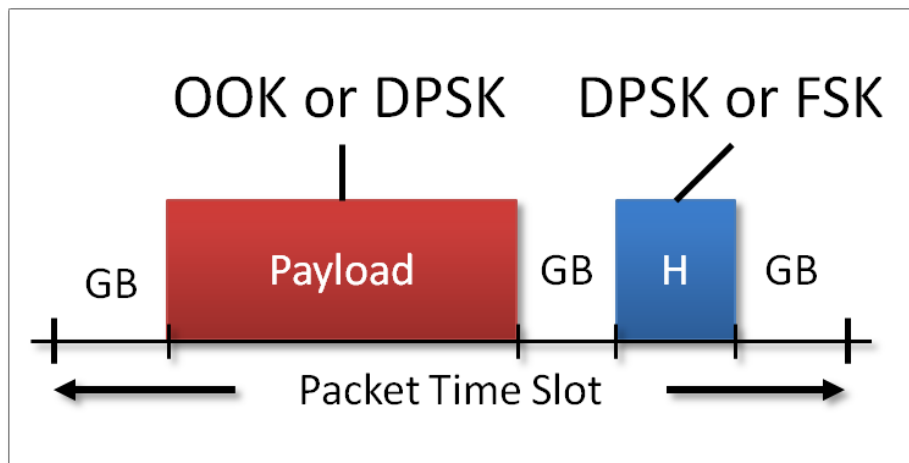
Figure 2.9: *Diagram of OPS encoding scheme using orthogonal modulation showing header and payload encoded in on-off-keying (OOK), differential phase shift keying (DPSK), and frequency shift keying (FSK) formats.*

Anther approach to parallel encoding of OPS header and payloads can be implemented by writing information on orthogonal modulation formats as shown in Figure 2.9. The idea was introduced in [19] where 155Mb/s headers are encoded in differential phase-shift keying (DPSK) or frequency-shift keying (FSK) formats while 10Gb/s payloads are on-off keying (OOK). Shortly after, Chie *et al.* demonstrated header extraction and rewriting for OPS packets consisting of 2.5Gb/s DPSK headers and 10Gb/s amplitude-shift keying (ASK) payloads. Header erasure was performed using cross-gain modulation (XGM) in an SOA that is not phase preserving. Header rewrite was performed by phase-encoding the header information onto an optical signal while the payload was transferred over by utilizing cross-absorption modulation (XAM) within an electro-absorption modulator (EAM). Shortly after, Leet *et al.* demonstrated simultaneous label extraction and rewriting of OPS packets consisting of 1.25Gb/s non-return-to-zero (NRZ) ASK labeled headers and 10Gb/s NRZ-DPSK payloads using a single reflective SOA (RSOA). The high-pass filter (HPF) properties of the RSOA are used to erase the low-speed headers while the payload remains unaffected. A new header is written by directly modulating the carrier density of the RSOA [20]. Recently, Zhang *et al.* introduced a packet encoding format using an 8-PSK modulation format to further increase spectral efficiency [21].

2.3.5 Encoding Format Comparisons

Packet encoding in the time domain is an attractive approach because it is relatively straightforward to implement and exhibits favorable spectral efficiency and low header-to-payload crosstalk. The main drawback of utilizing this approach is that the process of extracting and re-writing headers requires precise, nanosecond-scale timing. To reduce the overhead penalty in

forwarding capacity, header durations are typically on the order of 100s of nanoseconds, while header-payload guard bands are usually 10s of nanoseconds. Satisfying these stringent timing requirements may result in complex OPS node electronic control circuitry.

The ODSB sub-carrier technique is an attractive header-payload encoding approach because it offers ease of implementation and acceptable levels of spectral efficiency and low header-to-payload crosstalk. Unfortunately, this approach suffers from susceptibilities to signal walk-off resulting from chromatic dispersion (CD). Though this technique does not need to adhere to strict timing requirement, it calls for several optical filtering components in order to separate the header and payload. Additionally, this scheme may not scale well with payload data rate since it is necessary to utilize microwave components that are capable of operating at a frequency beyond the payload baseband.

The WDM approach is attractive because it potentially provides the simplest method of header extraction and rewriting while offering significantly low label-to-payload crosstalk. Though the timing requirements may not be as strict as the TDM approach, they are dominated by the compensation of header-payload walk-off caused by dispersion. This approach also exhibits an inferior spectral efficiency performance compared to the other methods since it requires additional optical bandwidth to transport the low-speed header.

The orthogonal packet encoding approach is attractive since it offers relatively good spectral efficiency. However, the most significant experimental demonstrations were carried out using payload extinction ratios (ER) ranging from 2-4dB to minimize inter-modulation crosstalk with the header. Moreover, minimizing the ER of the payload limits the distance one is able to transmit the OPS packet. Additionally, it becomes difficult to perform

phase-preserving packet routing via integrated wavelength conversion as the modulation formats become more complex. Most straight-forward demonstrations of wavelength conversion for advanced modulation formats have exploited nonlinear optical properties of bulky, fiber-based systems.

The OPS packet encoding format used in this work is employed in the time domain. The single-sideband SCM technique exhibits some promise, but requires complex header extraction techniques that do not necessarily scale well with payload data rate. The OPS core router is implemented with fast-switching integrated components such as SOAs and sampled-grating distributed Bragg reflector (SG-DBR) lasers. Furthermore, the integrated components are housed within custom FPGA-based driver boards to meet the stringent requirements of the time domain packet encoding method. Nanosecond-scale header extraction (6.4ns) and rewrite with sub-nanosecond accuracy (100ps) is demonstrated in subsequent chapters.

2.4 Chapter Summary

The technology representative of electronic routers used in current core networks has been presented along with its advantages and shortcomings with respect to bandwidth scaling. Additionally, all-optical transport methodologies were introduced as next-generation technology whose power consumption may be potentially independent of system bandwidth. A detailed comparison between optical circuit switching (OCS), optical burst switching (OBS), and optical packet switching (OPS) has been carried out with respect to throughput efficiency and implementation complexity. The OPS transport scheme is ultimately selected since it may achieve higher bandwidth utilization than OCS and OBS. A hit in system throughput is observed in OCS because of

the long setup and tear-down times required when establishing end-to-end optical circuit paths. Moreover, aggregation and disassembly of optical payload bursts at the edges of OBS network cores limits the end-to-end system throughput performance. High bandwidth utilization of OPS router nodes is achieved by minimizing packet guard bands, which requires strict optimization of optical and electronic components to achieve nanosecond-scale switching speeds and with sub-nanosecond temporal resolution. Several methods of OPS packet encoding have been presented, but the scheme utilizing bit-serial transmission in the time domain is favored. This method exhibits good spectral efficiency and low crosstalk between low-speed headers and high-speed payloads. However, extraction, processing, and rewriting of optical headers requires stringent timing that may complicate electronic control implementations.

References

[1] S. J. B. Yoo, "Energy Efficiency in the Future Internet: The Role of Optical Packet Switching and Optical-Label Switching," *IEEE Journal of Selected Topics In Quantum Electronics*, vol. 17, no. 2, April 2011.

[2] Cisco Systems. (2010). *Cisco CRS-3 carrier routing system.* Available: http://www.cisco.com/en/US/products/ps5763/products_data_sheets_list.html.

[3] R. Ramaswami, K. N. Sivarajan, "Routing and wavelength assignment in all-optical networks," *IEEE/ACM Trans. Networking*, vol.3, pp.858-867, October 1996.

[4] B. Mukherjee, *Optical Communication Networks*, New York: McGraw-Hill Publisher, 1997.

[5] M. Kovacevic, A. Acampora, "Benefits of wavelength translation in all-optical clear-channel networks," *IEEE Journal on Selected Areas in Communications*, vol.14, no. 5, pp. 868-880, 1996.

[6] R. Ramamurthy, B. Mukherjee, "Fixed-alternate routing and wavelength conversion in wavelength-routed optical networks," *IEEE Global Telecommunications Conference, (GLOBECOM) 1998*, vol.4, pp. 2295-2302, 1998.

[7] J. S. Turner, "Terabit Burst Switching," *Journal of High Speed Networks*, vol. 8, no. 1, pp. 316, March. 1999.

[8] M. Duser, P. Bayvel, "Performance of a dynamically wavelength routed optical burst switched network," *IEEE Photonics Technology Letters*, vol. 14, no. 2, pp. 239241, Feb. 2002.

[9] W. L. Ha, R. M. Fortenberry, R. S. Tucker, "Demonstration of photonic fast packet switching at 700 Mbit/s data rate," *Electronics Letters*, vol. 27, no. 10, pp. 789-790, May 1991.

[10] J. P. Mack, K. N. Nguyen, M. M. Dummer, E. F. Burmeister, H. N. Poulsen, B. Stamenic, G. Kurczveil, J. E. Bowers, L. A. Coldren, D. J. Blumenthal, "40 Gb/s Buffered 2x2 Optical Packet Switching Using Photonic Integrated Circuits," *Conference on Lasers and Electro-Optics/International Quantum Electronics Conference*, Paper CMJJ6, May 2009.

[11] A. Budman, E. Eichen, J. Schalafer, R. Olshansky, F. McAleavey, "Multigigabit optical packet switch for self-routing network with subcarrier addressing," in *Optical Fiber Communications Conference, (OFC) 1992*, vol. 5, Paper TuO4, 1992.

[12] Z. Zhu, Z. Pan, S. J. B. Yoo, "A Compact All-Optical Subcarrier Label-Swapping System Using an Integrated EML for 10-Gb/s Optical Label-Switching Networks," *IEEE Photonics Technology Letters*, vol. 17, no. 2, February 2005.

[13] G. Kovács, G. Puerto, T. Bánky, A. Martinez, M. Csörnyei, M. D. Manzanedo, D. Pastor, B. Ortega, T. Berceli, J. Capmany, "Subcarrier

multiplexed optical label swapping networks," *IET Optoelectronics*, vol. 4, no. 6, pp. 235-246, December 2010.

[14] S. Xiao, A. M. Weiner, "Four-User ∼3-GHz-Spaced Subcarrier Multiplexing (SCM) Using Optical Direct-Detection via Hyperfine WDM," *IEEE Photonics Technology Letters*, vol. 17, no. 10, pp. 2218-2220, October 2005.

[15] G. H. Smith, D. Novak, Z. Ahmed, "Novel technique for generation of optical SSB with carrier using a single MZM to overcome fiber chromatic dispersion," *Microwave Photonics, 1996, International Topical Meeting on*, pp. 5-8, December 1996.

[16] G. H. Smith, D. Novak, "Broad-Band Millimeter-Wave (38 GHz) Fiber-Wireless Transmission System Using Electrical and Optical SSB Modulation to Overcome Dispersion Effects," *IEEE Photonics Technology Letters*, vol. 10, no. 1, pp. 141-143, January 1998.

[17] C. J. Moss, L. J. S. Ville, K. S. Man, I. M. Burnet, "Experimental results for fast, high-capacity optical switching architectures," *Topical Meeting Photonics in Switching*, Palm Springs, CA, 1993.

[18] A. Okada, "All-optical packet routing in AWG-based wavelength routing networks using an out-of-band optical label," *Optical Fiber Communication Conference, (OFC) 2002*, pp. 213-215, Paper WG1, March 2002.

[19] T. Koonen, G. Morthier, J. Jennen, H. de Waardt, P. Demeester, "Optical packet routing in IP-over-WDM networks deploying two-level optical labeling," *Proceedings of 27th European Conference on Optical Communications, (ECOC) 2001*, vol. 4, pp. 608-609 Th.L.2.1, 2001.

[20] K. L. Lee, C. Lim, E. Wong,A. Nirmalathas, "Simultaneous Label Erasure and Rewriting using a Single Reflective Semiconductor Optical Amplifier for DPSK/ASK Optical Label Switching," *Lasers and Electro-Optics Society Meeting, (LEOS) 2006*, pp. 851-852, October 2006.

[21] L. Zhang, C. Yu, X. Xin, L. Bo, "High-speed optical label switching based on the 8PSK/ASK orthogonal modulation format," *The 14th OptoElectronics and Communications Conference, (OECC) 2009*, pp. 1-2, July 2009.

Chapter 3

Background

Thus far, most of the work presented in the field of optical packet switching (OPS) has been focused on backbone core router technology. There has been a limited amount of published demonstrations showing edge-to-edge transmission of Internet Protocol (IP) traffic trough an OPS core network, but the obtained results are presented in this chapter as the background for the work included in this dissertation. End-to-end transmission requires adaptation layers at the core edges to enable interoperability between legacy and future network formats. A successful implementation must support large traffic capacity with minimal performance degradation, must be scalable, modular, reconfigurable, and should potentially provide some form of traffic shaping or engineering. Currently, there is no commercially available technology capable of allowing an OPS network to inter-operate with electronic legacy networks, which requires the development of custom high-speed electronics. Several research groups have presented intricate OPS experiments showing crucial optical core router functionalities such as optical packet forwarding, synchronization, contention resolution, and all-optical signal regeneration. However, end-to-end communication demonstrations so far have been real-

ized at relatively low data rates and/or without at least one of the afore-mentioned OPS router functionalities needed for successful communication between optical core edges. Edge interoperability was initially demonstrated through OPS nodes consisting of 1×2 optical forwarding fabrics without contention resolution, or signal regeneration. IP traffic was successfully routed through such switches after being adapted into labeled optical payloads operating at data rates of 250Mb/s [1] and 2.5Gb/s [2,3]. All-optical forwarding and successful interoperability between an IP format and a labeled optical packet format, consisting of 3.125Gb/s labeled headers and 12.5Gb/s payloads, has been shown through edge adaptation layers with traffic shaping and real-time configuration capabilities [4–6]. Contention resolution and forwarding of 2.5Gb/s [7–9] and 80Gb/s [10] labeled optical packets have been shown in conjunction with end-to-end adaptation resulting in low packet loss rates. The work presented in this dissertation demonstrates the next steps required to achieve edge-to-edge forwarding through a regenerative, buffered OPS core node.

3.1 Packet Adaptation and Forwarding

Transmission of IP-based traffic over an optical label switching format was initially demonstrated by Cheng *et al.* in [1] by using four computer-based hosts and two $LiNbO_3$ optical cross-connects that served as backbone core routers. The IP payloads were modulated at 250Mb/s, while the headers were multiplexed at a 3GHz sub-carrier frequency. End-to-end transmission between all hosts and real-time reconfiguration of routing tables was successfully achieved with observed latencies of $0.9\mu s$ dominated by the optical switches.
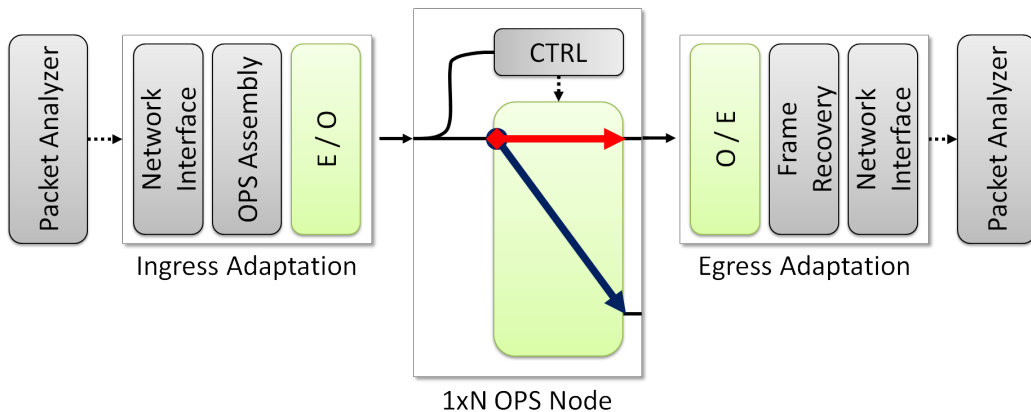
Figure 3.1: *Experimental setup for end-to-end optical packet adaptation, and forwarding showing two edge routers, an OPS node, an OPS electronic control plane (CTRL), electro-optic (E/O) conversion, and opto-electric (O/E) conversion.*

The initial work by UC Santa Barbara made significant contributions to the aforementioned proof-of-concept adaptation demonstration by focusing on the functionality, performance, and capacity of the adaptation layers and the OPS switching fabric. End-to-end transmission of 2.5Gb/s asynchronous, variable-length traffic was demonstrated trough a dynamically controlled 1×2 OPS switch in [2, 3]. The experimental setup used in that demonstration is shown in Figure 3.1. Frames using the OC-48 (2.5Gb/s) Packet over SONET (POS) format were converted to 2.5Gb/s optical payloads that were preceded by a labeled header. The OPS core router node consists of a two-stage tunable wavelength conversion (TWC) unit connected to an arrayed waveguide grating (AWG). Optical idlers were inserted to fill inter-packet gaps to avoid transients at the optical amplifiers and the burst mode receiver. Excellent throughput performance was observed with $0.79\mu s$ of latency through the core node.

Though this work improved existing adaptation technology, by increasing data rate, throughput, and latency performance. It, however, was a minimalistic demonstration of OPS switch functionality. A high-capacity OPS core router needs to operate at higher data rates and must be able to demonstrate contention resolution beyond deflection routing. Furthermore, this implementation did not include all-optical signal regeneration, which is crucial to extend the reach of packets traversing an OPS core network.

3.2 Packet Adaptation, Forwarding, and Traffic Shaping

A demonstration of end-to-end communication, traffic shaping, and optical packet forwarding has been shown for variable-length Gigabit Ethernet (1GbE) traffic [4–6], where the experimental setup is shown in Figure 3.2. Incoming 1GbE frames enter the Ingress Edge router where labeled OPS packets are dynamically generated using electronic table lookups that map a destination IP address to a desired 3.125Gb/s labeled header and output wavelength. Packets are stored in memory and are aggregated into high-speed payloads at 12.5Gb/s depending on destination and class of service (CoS). Packets are transparently forwarded using a 1×N OPS switching fabric that is configured via a low-speed electronic control plane (CTRL). Routed OPS packets then enter the Egress Edge router where 1GbE frames are extracted from the payloads before being transmitted to a packet analyzer for performance measurements.

The OPS switching fabric consists of a fast-switching, Mach-Zehnder interferometer tunable wavelength converter (MZI-TWC) and an arrayed waveguide grating (AWG). The electronic control plane requires 42ns to per-
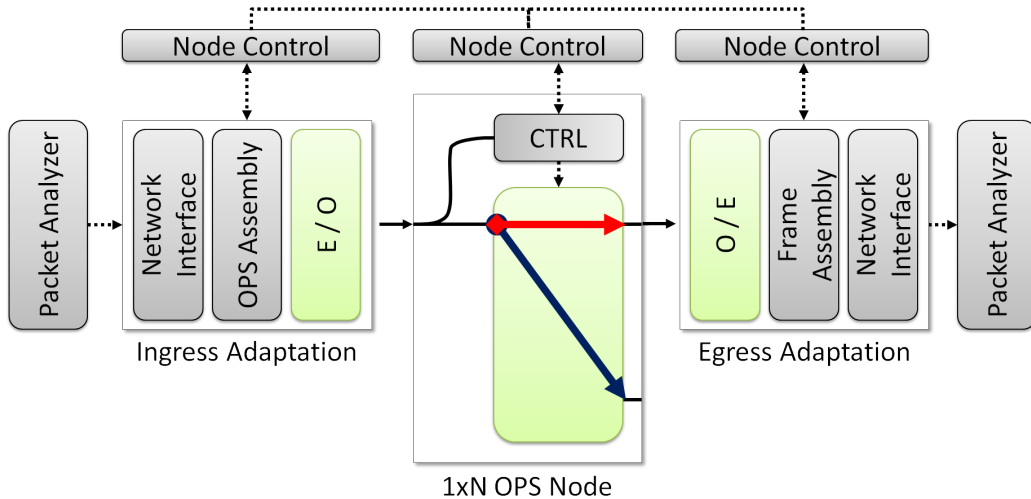
Figure 3.2: *Experimental setup for end-to-end optical packet adaptation, traffic shaping, and forwarding showing two edge routers, an OPS node, an OPS electronic control plane (CTRL), electro-optic (E/O) and opto-electronic (O/E) converters.*

form asynchronous recovery and extraction of labeled headers, while the MZI-TWC is capable of average λ-switching times on the order of 8ns. Therefore, a 60-nanosecond guard band is inserted between the header and payloads to allow the core node enough time to extract and process labeled headers. An extinction ratio greater than 16dB was maintained after switching and label insertion. Traffic shaping was performed by collecting information regarding observed traffic, which was forwarded to Node Control stages responsible for updating electronic lookup tables within the ingress edge and the CTRL. End-to-end performance was evaluated via bit error rates and showed error-free operation ($< 10^{-9}$) at a receiver power sensitivity of -12dBm.

This previous work is the most significant demonstration of edge adaptation technology, but falls short of presenting end-to-end transmission through practical OPS core router technology. This work utilized payload data rates

of 12.5Gb/s, while speeds greater than 40Gb/s are required to take advantage of the power scalability that OPS potentially provides. Furthermore, there is no sign of packet contention resolution (other than deflection routing) since only one input port of the core router is utilized. Also, there is no demonstration of signal regeneration, which will be required for multi-hop operation. Though error-free operation was shown via BER measurements, end-to-end performance was not evaluated using Layer-III metrics that extend beyond the OPS core.

3.3   Packet Adaptation, Forwarding, and Buffering

The previous work by UC Davis, which is presented in [7–9], successfully showed edge-to-edge transmission through a buffered optical switching fabric. The experimental setup used is shown schematically in Figure 3.3. A commercially available packet analyzer was utilized to generate OC-48 frames, which were forwarded to the Ingress Adaptation Layer to be converted into optical packets comprised of 1.55Mb/s labeled headers and 2.5Gb/s payloads. The adapted packets were then sent to a 2×2 optical router node, where they were switched towards the Egress Adaptation Layer or one of several drop ports. The Egress Layer extracted OC-48 frames from incoming payloads, which were then sent to a packet analyzer for performance measurements.

The Ingress Layer was capable of performing real-time generation of labeled headers by utilizing electronic table lookups. The Egress Layer recovered the encapsulated frames from incoming payloads, but it is unclear whether it was performed asynchronously with a CDR stage. The optical switching fabric consisted of an electronic control plane (CTRL) that was
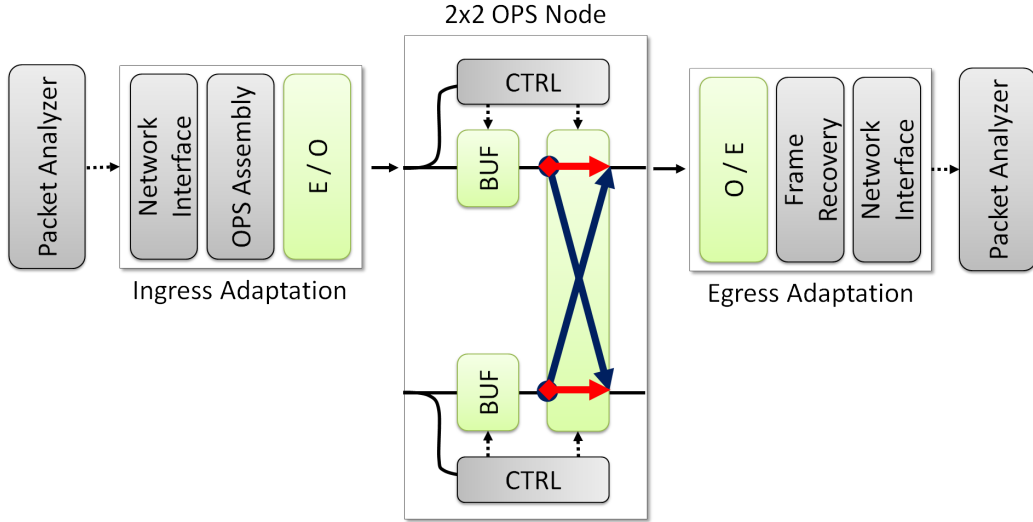
Figure 3.3: *Experimental setup for end-to-end optical packet adaptation, forwarding, and buffering showing four edge routers, a 2×2 OPS node, optical buffers (BUF), OPS electronic control planes (CTRL), electro-optic (E/O) and opto-electronic (O/E) converters.*

used to arbitrate operation of components in the optical data path. Packets were routed and forwarded by utilizing an AWG and TWCs based on cross-gain modulation (XGM) in semiconductor optical amplifiers (SOAs) and cross-phase modulation (XGM) in MZI-TWCs. Time-domain contention resolution was implemented by coupling a fiber delay line (FDL) to an AWG drop-port and feeding it back towards an add-port via a TWC. Though contention resolution was not demonstrated, the buffer connection was used to emulate short-reaching, end-to-end communication through two node hops. In [7], error-free operation (BER $< 10^{-9}$) was successfully achieved with a receiver sensitivity of -21 and -20dBm for 1 and 2 hops, respectively. The error-free, Layer-I BER results can be extrapolated to a Layer-III packet error rate performance of 10^{-4}. In [8], more than 99.77% of packets were

successfully recovered ($< 2.34 \times 10^{-03}\%$ lost) after two node hops. Finally, the observed packet loss rate percentage varied from 10^{-2} and 9×10^{-2} for the best and worst case scenarios respectively in [9].

The contributions by UC Davis allowed for end-to-end communication through multiple all-optical router nodes while only incurring frame loss percentages below 9×10^{-2}. However, the data rate was limited to 2.5Gb/s and some sort of signal regeneration was required to minimize the observed power penalty associated with multi-hop operation.

The results achieved by Furukawa *et al.* in [10], sought to improve upon the high-speed performance of buffered end-to-end OPS demonstration by UC Davis. Here, an experimental setup similar to Figure 3.3 was used. Except, an Ingress Adaptation Layer was placed at each input of the 2×2 OPS router, while two Egress Adaptation Layers were placed at two of its outputs. Also, the switching fabric was implemented spatially with interferometric 1×2 LiNbO$_3$ switches, while contention resolution was carried out with output buffers employing three stages of feed-forward fiber delays. The Ingress Layer converted 10GbE frames into a packet format consisting of 10Mb/s labeled headers and 80Gb/s high-speed payloads. The payload encoding scheme utilized wavelength division multiplexing (WDM) where incoming frames were segmented into an $8\lambda \times 10$Gb/s optical signal. The eight optical channels utilized the optical bandwidth between 1547.72 and 1553.33nm with 100GHz of channel spacing, while the low-speed header was transmitted on a 1556nm optical channel. End-to-end adaptation and transmission through an buffered OPS node was successfully demonstrated with a packet loss percentage below 5.41×10^{-7}.

Though this demonstration showed improvements in data rate, it suffers from several fundamental system-level shortfalls, which are addressed in

other chapters of this dissertation. Long haul system performance may be degraded since spectral efficiency is diminished and signal walk-off must be accounted for when using out-of-band WDM signaling, as discussed in Section 2.3.5. Also, the LiNbO$_3$ switches utilize interferometric schemes that may exhibit high crosstalk (10-20dB), details in Section 4.2.3. The optical buffer design could benefit from employing a feed-backward configuration utilizing fast, high-isolating switches, addressed in Section 4.2.2. Moreover, signal regeneration stages will be required to enable multiple hops through an OPS core network.

3.4 Impact of this Work on Edge Interoperability and Optical Packet Switching

The work in this dissertation presents the next critical steps required to achieve the first demonstration of edge-to-edge forwarding through a buffered, regenerative OPS core router. The experimental setup used in this demonstration is shown in Figure 3.4. A commercially available packet analyzer is utilized to transmit IP-based traffic through an Ingress adaptation Layer, which adapts incoming frames into an OPS packet format consisting of 10Gb/s low-speed labeled headers and 40Gb/s high-speed payloads. The labeled packets are then forwarded through a 2×2 OPS core router. Furthermore, contention resolution is performed in the time domain via optical packet buffers (BUF), while forwarding is carried out in the optical domain by integrated TWCs and an AWG. Data signals are then regenerated by employing re-amplification, re-shaping, and re-timing (3R) at each node output. Each of the optical subsystems are comprised of photonic integrated circuits (PICs) packaged in custom FPGA-based drivers. The electro-optic control
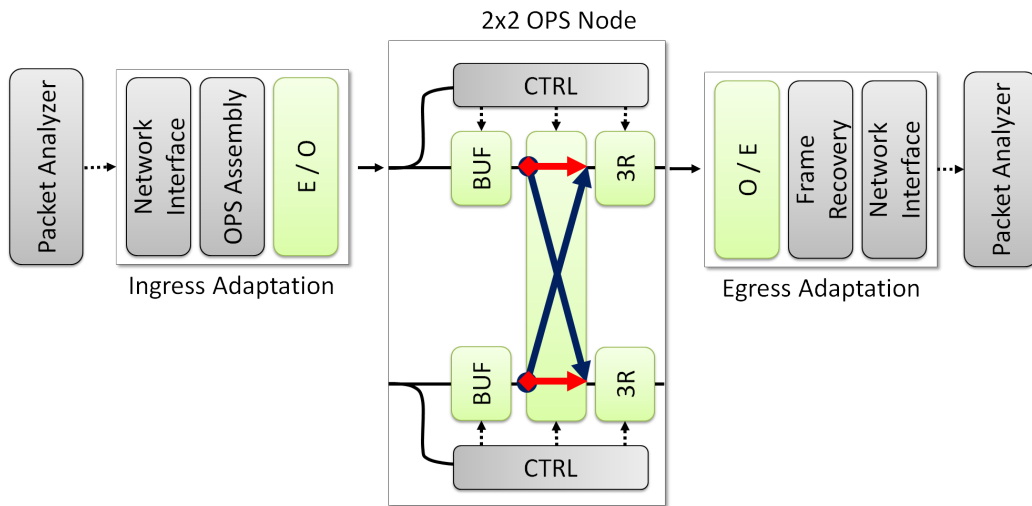
Figure 3.4: *Experimental setup for end-to-end optical packet adaptation, forwarding, and buffering showing four edge routers, a 2×2 OPS node, optical buffers (BUF), OPS electronic control planes (CTRL), electro-optic (E/O) and opto-electronic (O/E) converters.*

and interfacing is handled by a low-speed electronic control plane (CTRL), which extracts and processes packet headers to perform the routing functionality. Finally, packets exiting the router node are then sent to an Egress Adaptation Layer where IP frames are recovered and extracted from the optical payloads before being forwarded to the commercial packet analyzer for Layer-III performance measurements.

	Measurement	Comments
Layer-I	BER	• Repeating PRBS pattern • Metric: (# errors) / (# bits transmitted) • Pro: quick & good for Physical Layer • Con: system-level more difficult
Layer-II (Prev Work)	Header Recovery (Identifier)	• Burst mode, packet-based data stream • Detect sequence located in header • Metric: (# detected / # transmitted) • Pro: less stringent sync requirements • Con: does not detect payload errors
Layer-III (This Work)	Frame Recovery (CRC)	• Burst mode, packet-based data stream • Check-sequence imbedded in payload • Detect: seq in payload vs. derived seq • Pro: single erroneous bit → frame error

Figure 3.5: *Measurements used in this dissertation to evaluate router performance.*

3.5 Performance Measurements

The work presented in this dissertation uses several types of measurements to evaluate overall performance at the physical layer and at the system level. The following sections discuss the details of three measurement techniques by addressing their advantages and drawbacks, which are summarized in Figure 3.5.

3.5.1 Layer-I

Performance of physical link interconnects are evaluated using bit-level Layer-I measurements. Bit-error-rate (BER) tests are widely accepted as a means of ascertaining Layer-I performance. Such a measurement is performed by transmitting a continuously repeating pseudo-random bit sequence (PRBS) through a device that is under test (DUT) and noting the rate of observed

errors compared to total bits transmitted ($\frac{\text{\# errors}}{\text{\# total bits}}$). Long-haul optical transmission applications require $BER \leq 10^{-9}$ to regard operation as error-free. Since the detection of errors is a statistical process, a BER value is measured with 100% of certainty when the number of bits transmitted approaches infinity. However, a confidence level of approximately 95% is the industry standard that allows for more practical measurement durations. The number of transmitted bits (N) required to obtain a BER with a confidence level (CL) for a known error count (E) is shown in (3.1) [11]. By assuming zero errors (E = 0) and knowing $N = B \times T$, the duration (T) of a BER measurement for a certain confidence level at a known bit rate (B) is shown in Equation 3.2.

$$\text{N} = \frac{1}{BER} \left\{ -ln(1 - CL) + ln \left[\sum_{k=0}^{E} \frac{(N \times BER)^k}{k!} \right] \right\} \tag{3.1}$$

$$\text{Time (s)} = \frac{-ln(1 - CL)}{B \times BER} \tag{3.2}$$

The BER measurement has been standardized and several commercial implementations are readily available and capable of providing quick, real-time physical layer (Layer-I) performance evaluations for at data rates up to 60Gb/s. The only caveat is that one needs to provided a continuous data-clock signal pair that is synchronous. Burst mode BER measurements are typically performed by utilizing data bursts with durations on the order of 100s of microseconds to allow the error analyzer (EA) ample amount of time for data-to-clock synchronization. Therefore, it is impractical to utilize a BER tester to obtain reliable system-level performance characterizations with traffic consisting of packet bursts lasting 10-100s of nanoseconds at data rates beyond 40Gb/s.

3.5.2   Layer-II

The previous work by Mack *et al.* utilized Layer-II packet recovery measurements to obtain performance evaluations of a transmission system using a burst mode, packet-based data stream. The transmitter inserted a check-sequence within the header of an OPS packet, while the receiver interpreted the detection of the sequence as a successfully recovered packet. Furthermore, system-level performance was evaluated using packet recovery percentage $(\frac{\# \text{ detected}}{\# \text{ transmitted}})$ as a metric. This method allowed one to detect a sequence as short as 100 bits in length, which provided more flexibility with respect to data-clock synchronization. As a result, it was possible to obtain header recovery (Layer-II) measurements for traffic consisting of data bursts with durations comparable to IP data packet lengths. While this method serves as an attractive means of evaluating packet recovery performance of label swapping within an optical network, it is not capable of detecting errors within high-speed payloads since they are not processed electronically. Therefore, a method of evaluating performance beyond optical network edges is needed.

3.5.3   Layer-III

The work presented in this dissertation utilizes Layer-III packet recovery measurements to evaluate the end-to-end performance of an ODR network beyond the edges. These measurements utilize a continuous stream of IP payloads encapsulated by Ethernet frames. Each frame includes a checksum derived from the payload and is appended at the end of the frame. A data stream consisting of such frames is generated at one network edge and is then transmitted through the system under consideration before exiting through

a separate network edge. Layer-III measurements are then performed by deriving a check-sequence from the received frame and comparing it to the embedded sequence. A frame is then successfully recovered when the two sequences are identical. This approach is attractive because it allows one to abstract end-to-end performance into a single metric without the need of extrapolation. Additionally, this measurement is more applicable when evaluating the performance of an OPS core network from the point of view of an end user. Layer-II and -III measurements are presented through out this work where applicable.

3.6   Chapter Summary

This chapter has discussed the previously published work demonstrating interoperability between OPS and legacy networks, which is required for edge-to-edge optical core communication. The research community widely accepts the fact that all-optical forwarding, contention resolution, and signal regeneration of high-speed optical payloads is crucial to implementing a successful OPS core router. However, previous demonstrations of edge-to-edge transmissions have been performed at payload data rates well below 40Gb/s and have not shown forwarding in conjunction with contention resolution and signal regeneration. Some form of time-domain contention resolution is required for multi-port core routers and signal regeneration is crucial to achieve transmission through multiple nodes of an OPS core network. This dissertation presents the steps needed to achieve the first demonstration of edge-to-edge communication through an OPS core router capable of all-optical forwarding, buffering, and signal regeneration of 40Gb/s optical packets. Contention resolution, forwarding, and regenerators are implemented with

integrated devices in order to take advantage of potential benefits in reduced power consumption. Also, optical devices are packaged in custom, FPGA-based controller boards which is crucial to obtaining stability, flexibility, and switching efficiency.

References

[1] G. K. Chang, G. Ellinas, B. Meagher, W. Xin, S. J. B. Yoo, M. Z. Iqbal, J. Young, H. Dai, Y. J. Chen, C. C. Lee, X. Yang, A. Chowdhury, T. F. Chen, "A proof-of-concept, ultra-low latency optical label switching testbed demonstration for next generation Internet networks," *Optical Fiber Communications Conference, (OFC) 2000*, vol. 2, pp. 56-58 vol.2, Paper WD5-1, 2000.

[2] S. Rangarajan, H. N. Poulsen, P. G. Donner, R. Gyurek, V. Lal, M. L. Masanovic, D. J. Blumenthal, "End-to-end layer-3 (IP) packet throughput and latency performance measurements in an all-optical label switched network with dynamic forwarding," *Optical Fiber Communication Conference, (OFC) 2005*, vol. 3, Paper OWC5, March 2005.

[3] H. N. Poulsen, S. Rangarajan, M. L. Masanovic, V. Lal, D. J. Blumenthal, "Performance of a label erase and wavelength switching sub-system for layer-3 all-optical label switching using a two stage InP wavelength converter," *Optical Fiber Communications Conference, OFC 2005*, vol. 2, Paper OTuC2, March 2005.

[4] R. Nejabati, D. Klonidis, D. Simeonidou, M. O'Mahony, "Hybrid edge IP/optical packet generator in wavelength routed networks," *Optical*

Fiber Communications Conference, (OFC) 2003, vol. 2, Paper ThV1, February 2004.

[5] R. Nejabati, D. Klonidis, D. Simeonidou, M. OMahony, "Demonstration of an Agile Hybrid IP-Optical Packet Construction Mechanism in Wavelength Routed Optical Packet Switched Networks," *IEEE Communications Letters*, vol. 9, no. 6, pp. 552-554, June 2005.

[6] R. Nejabati, G. Zervas, D. Simeonidou, M. J. O'Mahony, D. Klonidis, "The 'OPORON' Project: Demonstration of a Fully Functional End-to-End Asynchronous Optical Packet-Switched Network," *Journal of Lightwave Technology*, vol. 25, no. 11, November 2007.

[7] J. Taylor, Y. Bansal, M. Y. Jeon, Z. Pan, J. Cao, V. J. Hernandez, Z. Zhu, Z. Wang, S. J. B. Yoo, V. Akella, T. Nady, G. Goncher, K. Ervin, K. Boyer, B. Davies, "Demonstration of IP client-to-IP client packet transport over an optical label-switching network with edge routers," *European Conference on Optical Communications, ECOC 2003*, Paper Mo3.4.2, 2003.

[8] S. J. B. Yoo, F. Xue, Y. Bansal, J. Taylor, Z. Pan, J. Cao, M. Jeon, T. Nady, G. Goncher, K. Boyer, K. Okamoto,S. Kamei, V. Akella, "High-Performance Optical-Label Switching Packet Routers and Smart Edge Routers for the Next-Generation Internet," *IEEE Journal of Selected Areas in Communications*, vol. 21, no. 7, pp. 1041-1051, September 2003.

[9] J. Hu, Z. Pan, Z. Zhu, H. Yang, V. Akella, S. J. B. Yoo, "First experimental demonstration of combined multicast and unicast video streaming over an optical-label switching network," *Optical Fiber Communications Conference, (OFC) 2006*, March 2006.

[10] H. Furukawa, N. Wada, H. Harai, M. Naruse, H. Otsuki, K. Ikezawa, A. Toyama, N. Itou, H. Shimizu, H. Fujinuma,H. Iizuka, T. Miyazaki, "Demonstration of 10 Gbit Ethernet/Optical-Packet Converter for IP Over Optical Packet Switching Network," *Journal of Lightwave Technology*, vol. 27, no. 13, pp. 2379-2390, July 2009.

[11] J. Redd, "Calculating statistical confidence levels for error-probability estimates," *Lightwave Magazine*, vol. 17, Issue 5, pp. 110-114, April 2000.

Chapter 4

The Label Swapped Optical Router (LASOR)

In this chapter, the architecture of a label swapped optical router (LASOR)
and its subsystems is presented in detail. The work previously demonstrated
for each subsystem is shown and compared to what is used in this work. The
LASOR router utilizes the optical packet switching (OPS) scheme as a means
of efficiently forwarding and routing optical packets. The OPS scheme makes
use of a labeled packet format comprised of a relatively low-speed header fol-
lowed by a high-speed payload. The forwarding information pertaining to
each packet is encoded as a label within the header, which is processed by
low-speed electronics that are relatively inexpensive in terms of cost and
power dissipation. This allows the payload to be processed transparently
in the optical domain. Routing and forwarding is successfully achieved by
extracting the labeled header and via electronic processing to determine the
desired output port of a particular packet. Incoming packets are synchro-
nized, via temporal alignment, to the local time-slot of the optical router.
Packet contention arises when multiple packets from disparate input ports
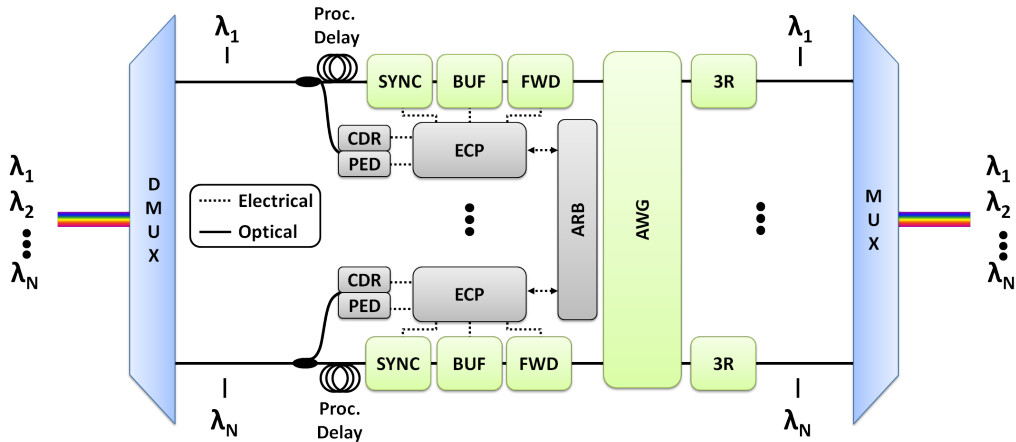
Figure 4.1: *Schematic representation of the label swapped optical router (LASOR) architecture.*

issue identical output port requests. Collisions are typically mitigated via contention resolution employing a combination of time delays, wavelength conversion, and deflection routing. Buffered payloads are routed towards the requested output port and new labeled headers are written via an optical packet forwarding plane. Signal regeneration via re-amplification, re-shaping, and re-timing (3R) is performed at each of the router output ports.

4.1 Architecture Overview

A schematic representation of an NxN LASOR router is illustrated in Figure 4.1. The router input consists of packets that are transmitted in a wavelength division multiplexed (WDM) fashion and are spatially separated using a wavelength de-multiplexer (DMUX). The DMUX output is then evenly distributed to each arm of the optical router, which consists of an optical data path and an electronic control path. The optical path is comprised of a fixed processing delay, an optical packet synchronizer (SYNC), buffer (BUF), for-

warding plane (FWD), and a 3R signal regenerator. The synchronizer is utilized to align the incoming packets to the local time-slot of the router, while the optical packet buffer is used to resolve any contention resulting from conflicting port requests. The packet forwarding plane performs tunable wavelength conversion and header rewrite before sending packets through an arrayed waveguide grating (AWG) that routes packets by utilizing its wavelength dependent properties.

The data path is tapped using a 3dB coupler in order to direct incoming packets towards the electronic control plane that generates controlling signals for the optical path on a per-packet-basis. An opto-electronic(OE) conversion is performed on incoming packets before they enter the clock and data recovery (CDR) and payload envelope detection (PED) stages whose output is forwarded to an electronic channel processor (ECP). A label is extracted from the header and is used to perform an electronic table lookup to resolve the port request. The PED signal is used by the ECP to determine the temporal location of the high-speed payload as it traverses the optical data path. Each ECP forwards the port request to an overseeing arbitration (ARB) stage that dictates router operation based on a prioritizing algorithm. The ARB stage settles any packet contention by instructing the ECPs to perform forwarding, buffer-writes, or buffer-reads depending on a predetermined precedence scheme. Once proper course of action is determined, each ECP produces appropriate control signaling for the optical components in the data path. Routed packets are regenerated via re-amplification, re-shaping, and re-timing (3R) to maintain high signal quality throughout multiple routing nodes. Each 3R stage performs a fixed wavelength conversion where the output is tuned to match the original packet wavelength ($\lambda_{IN} = \lambda_{OUT}$). Finally, regenerated packet streams are converted to a WDM format by utilizing an

optical multiplexer (MUX) to combine all output wavelengths.

4.2 Optical Data Path

The optical data path of the LASOR router consists of a packet synchronizer that is utilized to time-align incoming packets the local router time slot. Packet contention resulting from conflicting port requests is resolved by means of optical packet buffers, while routing is performed via a packet forwarding plane. Routed packets are then subjected to re-amplification, reshaping, and re-timing (3R) to maintain or improve a packet's signal quality.

4.2.1 Packet Synchronizer

It is undesirable to fragment variable-length optical packets into fixed length packets at the edge of core optical networks. Complex, resource intensive defragmentation algorithms are required at the egress edge of the core, and the loss of a single fragment may require the re-transmission of large datasets. Therefore, optical routers must be capable accommodating asynchronously arriving variable-length optical packets [1]. Figure 4.2 illustrates the basic operating principle behind optical packet synchronization. Asynchronously arriving packets (left) are shown as being misaligned with the local router time slots. The synchronizer forwards packets through a series of variable delays where configuration is set based on the amount of misalignment. Synchronized packets (right) are then aligned to the router time slot to enable precise, efficient switching through the remaining components within the optical data path.

Optical synchronizers typically consists of designs where a distribution of several delays is either laid out in series or parallel configuration. The
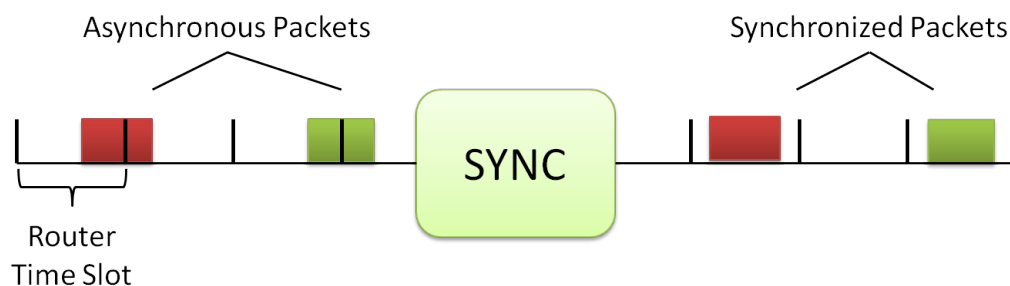
Figure 4.2: *Illustration showing basic operating principle behind packet synchronization. Asynchronously arriving packets (left) are aligned to the local router time slot (right).*

switching mechanism for both types is generally implemented using tunable wavelength converters connected to wavelength dependent elements or optical cross-point switches. Delays can also be configured in either feed-forward or feed-backward fashion. Feed-backward configurations tend to have smaller footprints where utilization of delays is done in an efficient manner. However, extra care must be taken to ensure that the leading bits of optical packets do not collide with the bits in the trailing portion of the packet. As a result, feed-forward designs tend to be more compliant with variable-length packets.

A schematic representation of a packet synchronizer in feed-forward configuration utilizing wavelength conversion can be seen in Figure 4.3. The wavelength of incoming packets is switched depending the on the delay required via tunable wavelength conversion (TWC). A 1×N de-multiplexer (DMUX) is then used to spatially route the incoming packets to an appropriate delay, while an N×1 multiplexer (MUX) is used to combine the output of each delay. An optional fixed wavelength conversion (FWC) is performed to preserve the wavelength of incoming packets if it is necessary. This design utilizes N delays to generate a total configurable time delay $T_D = (N-1)\Delta$
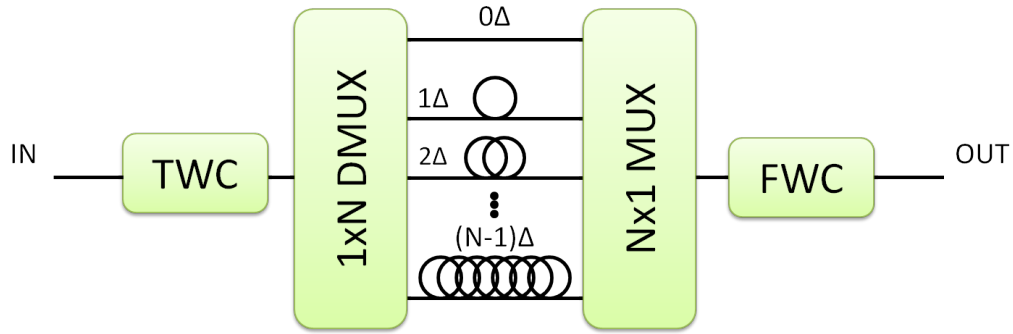
Figure 4.3: *Schematic of a feed-forward, wavelength conversion-based optical packet synchronizer (TWC: tunable wavelength converter, FWC: fixed wavelength converter).*
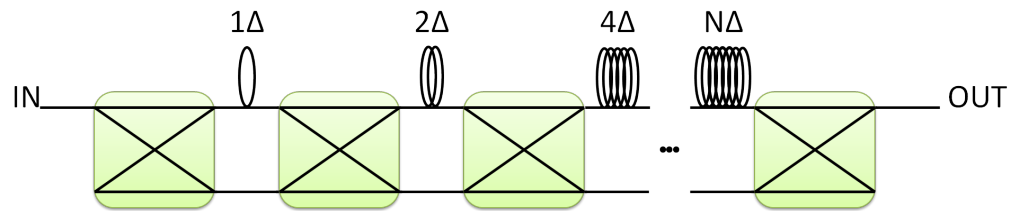
Figure 4.4: *Schematic of a feed-forward, optical packet synchronizer based on cross-bar switches.*

with a temporal resolution of Δ.

Figure 4.4 shows a diagram of a packet synchronizer in a feed-forward configuration with cross-bar switches. The design includes N delay stages that consist of a cross-bar switch that routes packets towards either a delayed or non-delayed path. The amount of delay increments as powers of two throughout each stage resulting in a binary delay configuration with a reduced foot print. Hence, a total of $\text{LOG}_2(\text{N})$ stages are required to achieve a total configurable time delay of $\text{T}_D = \Delta(2^N - 1)$ with a temporal resolution of Δ.

Optical packet synchronization of 10Gb/s optical packets across a 1008ns span with a time resolution of 16ns has been previously demonstrated using wavelength and space switching [2]. A tunable wavelength converter leads into a coarsely configurable delay (128ns) consisting of a 1×8 optical coupler that allows packets to be switch to one of eight delays using SOAs before being combined using an 8×1 coupler. Fine tuning is achieved by routing packets into a secondary delay stage consisting of eight 16ns delays using an AWG. Synchronization of variable-length 40Gb/s optical packets was previously achieved with 853ps of resolution over a dynamic tuning range of 12.8ns [3]. The design is based on a four-stage, feed forward configuration utilizing SOA-based cross-point switches and fiber-based delays that increment as powers of two throughout each stage. More recently, a configurable delay has been implemented on an ultra-low-loss SiN platform that may potentially synchronize variable length packets over a 12ns dynamic range with an 800ps temporal resolution [4]. The configurable delay is also implemented as a four-stage, feed-forward logarithmic design, but switching is carried out via interferometric, thermo-optic switches.

The work presented in this thesis does not utilize optical packet synchronizers. However, it is instructive to discuss the varying implementations since a re-sizable optical packet buffer is later demonstrated by utilizing a re-circulating variable delay that is based on optical packet synchronizer technologies. Although the wavelength conversion-based synchronizer design reduces the number of active components, its fabrication process is far more complex, expensive, and bulky relative to the logarithmic delay line approach. The configurable delay technology used in this work is based on the four-stage feed-forward logarithmic design utilizing SOA-based switching. The logarithmic delay configuration allows one to maximize the number

of allowable delays while the stage count scales on the order of $LOG_2(N)$. Utilizing SOA-based switching not only provides signal gain to compensate for splitting losses, but also reduces the crosstalk between parallel paths (<-40dB). Alternatively, the thermo-optic switch-based design suffers from low isolation (<20dB) and switching speeds that are on the order of hundreds of microseconds to tens of milliseconds. The variable delay performance in this work is limited by the accumulation of ASE and patterning caused by the SOA switches.

4.2.2 Packet Buffer

Contention resolution within an OPS optical router occurs when multiple incoming packets request identical output ports. Contention may be potentially mitigated by using three resolution techniques: optical buffering, wavelength conversion, and deflection routing [5]. The latter method consists of mitigating contention by routing a contending packet towards an undesired output port. This method essentially "passes the buck" with the hopes that another optical router can successfully route the packet. The wavelength conversion method converts all but one of the contending packets to a separate wavelength in order to resolve contention. This method is desirable because it does not introduce excess delays in the data path and does not require packet sequencing. It does, however, demand extra wavelength conversion optics and requires that the router be compatible with this functionality. The most straightforward method of contention resolution is performed by utilizing the time dimension with optical packet buffering.

Figure 4.5 illustrates the basic operating principle behind contention resolution via optical packet buffers. Two optical packets arrive at different input ports of the router (left) requesting the same output port (OUT1). One of
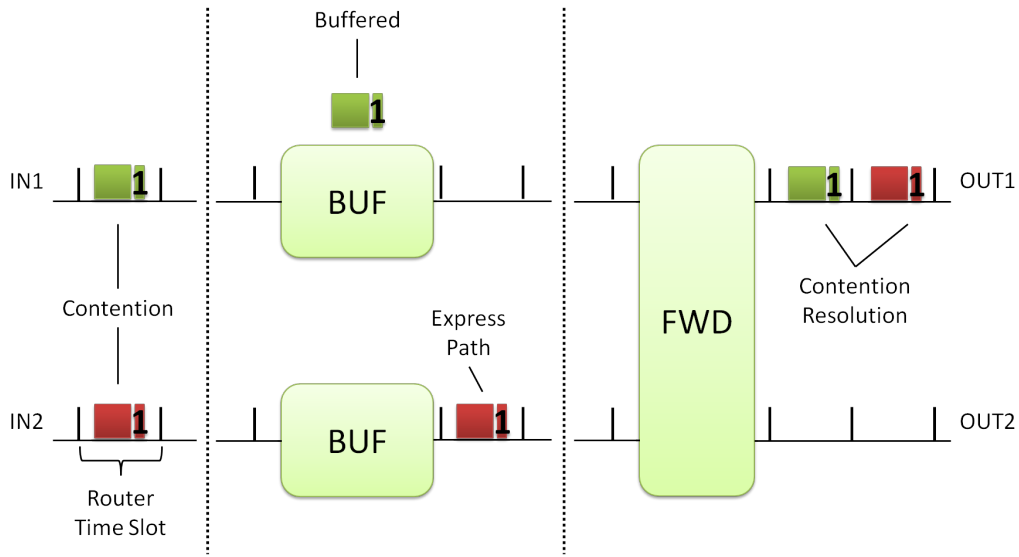
Figure 4.5: *Illustration showing basic operating principle behind packet contention resolution. Packet contention (left), buffering (middle), and contention resolution (right) are described.*

the contending packets is buffered while the other is granted routing precedence (middle) in order to successfully achieve contention resolution (right). The primary all-optical packet buffering approaches consist of slow light and delay line buffers that delay packets by decreasing the group velocity or increasing the physical length respectively [6]. Though there have been several advances in slow light optical packet buffers [7–9], it is a technology that is far from realization due to its fundamental limitations [10].

Delay line buffers can be classified as either feed-forward or feed-backward designs shown in Figure 4.6. When using a feed-forward design, a packet may be buffered only once where the maximum storage time is determined by the number of delay stages. The feed-backward design allows packets to be buffered multiple times by utilizing a re-circulating delay configuration where the maximum allowable storage time is determined by the number
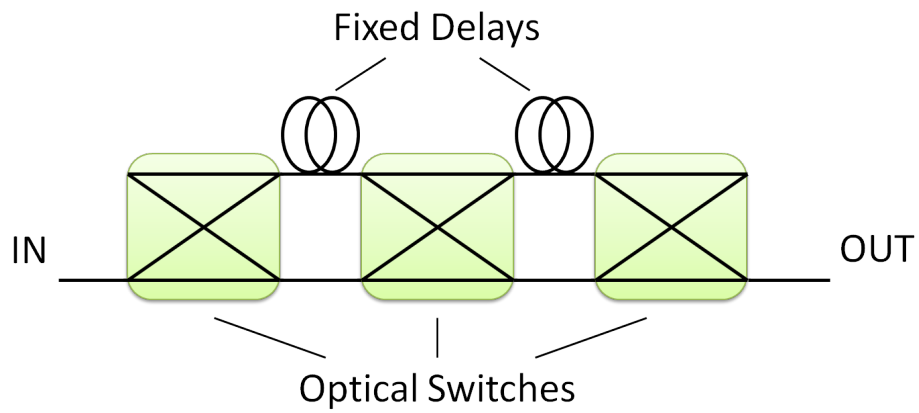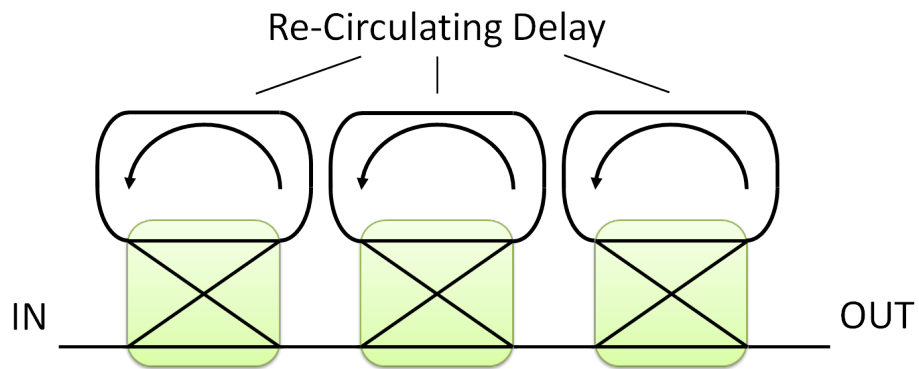
Figure 4.6: *(a) Cascaded two-stage feed-forward and (b) three-stage feed-backward optical packet buffer designs shown with fixed delays and optical switches.*

of circulations. The work presented in this thesis makes use of the feed-backward optical buffer design since it provides advantages of scalability, flexibility, smaller footprint, and lower component count.
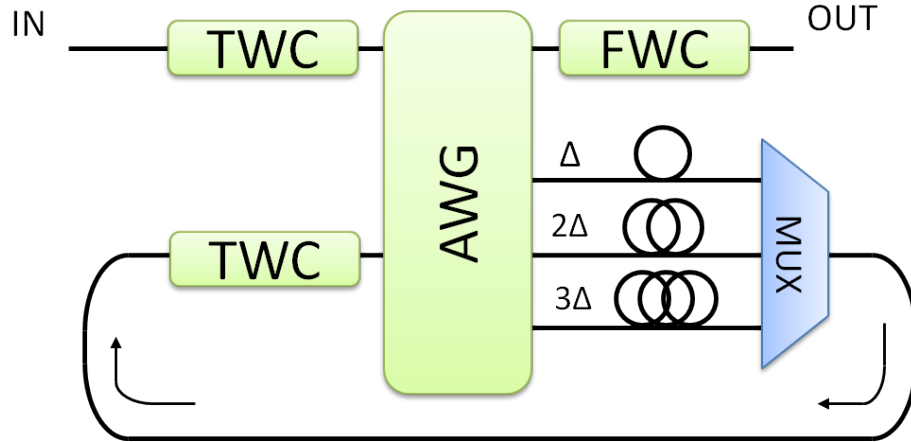
Figure 4.7: *Feed-backward implementation of a variable-length optical packet buffer based on wavelength conversion (TWC: tunable wavelength converter, FWC: fixed wavelength conversion, AWG: arrayed waveguide grating, MUX: multiplexer, DMUX: de-multiplexer).*

Feed-backward implementations can be further classified into single or multiple-wavelength designs utilizing cross-point switches or wavelength conversion respectively. Figure 4.7 shows a schematic representation of a feed-backward, variable-length optical packet buffer based on wavelength conversion. Input packets are spatially switched towards one of three re-circulating delays or the buffer output using a tunable wavelength converter (TWC) connected to an arrayed waveguide grating (AWG). The output from the delay stage is multiplexed (MUX) and sent to a separate TWC. The wavelength of buffered packets is switched to allow them either exit the buffer or re-enter the re-circulating delay. Packets exiting the buffer pass through an optional fixed wavelength conversion (FWC) if the router requires the packet

wavelength to be preserved.

Contention resolution of 100-byte 2.5Gb/s payloads has been previously demonstrated using wavelength conversion-based packet buffers achieving 14 circulations within a switching node [11]. The node employed a design consisting of two switching planes. The first plane routed packets towards the appropriate switch output port or towards a contention resolution path, while the second resembled Figure 4.7. Routing was carried out by connecting SOA-based, counter-propagating XGM wavelength converters (XGM-WC) with an AWG. A similar testbed for 40Gb/s optical packets was later demonstrated in [12] by utilizing integrated devices instead of XGM-WCs. The tunable wavelength converter was implemented using a multi-frequency laser that combines a passive AWG with an array of SOA gain sections capable of fast switching (<1ns). The tunable laser was then integrated with a Mach-Zehnder interferometer wavelength converter (MZI-WC) with the re-circulating buffer delay measuring $3.15\mu s$ in length.

A re-circulating optical packet buffer utilizing active vertical couplers (AVC) as a 4×4 cross-point switch achieved contention resolution for 10Gb/s optical packets [13]. The switch advantages included switching times below 1.5ns, low crosstalk (<-50dB), and a compact size ($500 \times 500\mu m$ / switch). Shortly after, packet contention for 40-byte 40Gb/s packets was demonstrated with a packet recovery greater than 98% across eight buffer circulations [14]. The implementation consisted of a 2×2 InP SOA cross-bar switch coupled to a re-circulating fiber delay. Performance was shown to improve by two circulations when inserting an optical bandpass filter within the buffer delay. More recently, a 2×2 cross-bar switch has been integrated with a meter delay on a hybrid silicon platform [15, 16]. Though contention resolution was not demonstrated, error-free operation was achieved through

every possible switch configuration.

An approach based on feed-backward packet buffer utilizing wavelength conversion has the advantages of nano-scale switching times and reduced number of active elements. However, the fabrication process of an MZI-WC is complex and susceptible to yield issues. A switch design based on either SOAs or AVCs would far simpler to fabricate and both would exhibit desirable qualities such as high isolation and fast switching speeds. However, the SOA-based implementation provides the added benefit of on-chip optical gain that can be utilized to compensate for splitting losses within the switch. As a result, the optical packet buffer technology used in this work is largely based on the previous work done by Mack *et al.* A functionally packaged prototype has been demonstrated showing packet loss percentages below 1% and is utilized in this work [14]. The implementation performance is limited by the accumulation of ASE and SOA patterning, which will be discussed in forthcoming chapters.

4.2.3 Packet Routing and Forwarding

Figure 4.8 illustrates the basic operating principle behind optical packet forwarding. The desired output port of an incoming packet (left) is extracted from the labeled header. The packet is then forwarded to a Header Erasure stage where the low-speed header is removed before forwarding the payload. A new labeled header is written to the packet before it is routed using the forwarding fabric. The forwarding fabric is used to route packet bursts ranging from 10s to 100s of nanoseconds at nanosecond-scale switching speeds with sub-nanosecond accuracy. The routing fabric must be transparent to payload data rate and must support multiple wavelengths and output ports. The most promising packet forwarding technologies can be represented as
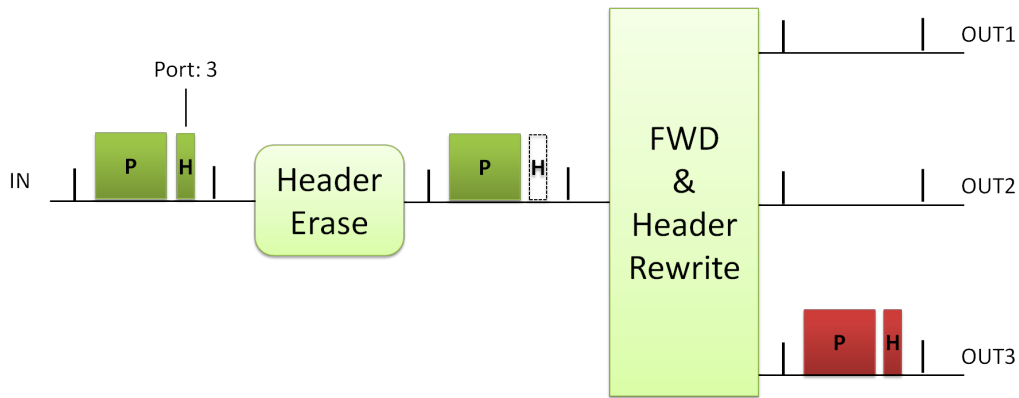
Figure 4.8: *Illustration showing principle behind optical packet forwarding. Incoming packets (left) consist of labeled headers (H) followed by payloads (P). Packets undergo header erasure before being forwarded with newly written headers.*
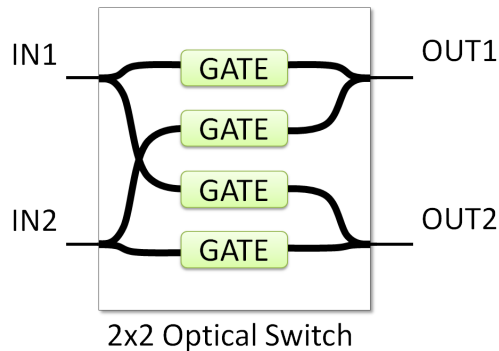
Figure 4.9: *A 2×2 optical switch implemented with 3dB couplers and optical gates.*

space switching or wavelength routing fabrics.

Space switching fabrics consist of feed-forward stages that broadcast packets to each possible path where routing is performed by enabling the path(s) leading to the desired output port via fast-switching optical gates. Figure 4.9 shows a schematic design of a 2×2 optical cross-point switch where

couplers are used as the broadcasting mechanism. The fast-switching optical gates have been demonstrated using SOAs or interferometers (directional couplers) where switches with an input port count of N typically require N^2 gating elements.

The use of switches based on directional coupler structures possess the advantage of achieving low insertion losses because the entire signal power is switched to the desired output without the need of splitting. This characteristic allows the insertion loss to remain relatively constant as the port count is scaled. The main disadvantage is that the extinction ratio between an ON and an OFF state is between -20 and -30dB. This is limited by fabrication uncertainties that prevent the phase and amplitude of interfering signals from being exact when being combined.

Fast-switching optical ON-OFF gates have also been implemented using the gain and absorption properties in SOAs. The gain is used to compensate for the splitting losses while the absorption properties allows one to obtain signal crosstalk levels below -40dB. The main drawback of utilizing SOA-based ON-OFF switches lies in the degradation of OSNR caused by the addition of amplified spontaneous emission (ASE). One must scale the SOA switch count as the number of signal splits increases as a result of scaling the switching fabric port count.

The most promising demonstrations consist of switch fabric designs utilizing hybrid implementations that simultaneously leverage the advantages of directional coupler and SOA ON-OFF switches. Error-free operation at 10Gb/s of a 4×4 integrated optical cross-point switch has been demonstrated using switching cells implemented with active vertical couplers [17]. The signal losses caused by splitting were eliminated by utilizing a grid of vertical couplers that demonstrated an extinction ration of roughly 30dB. Each
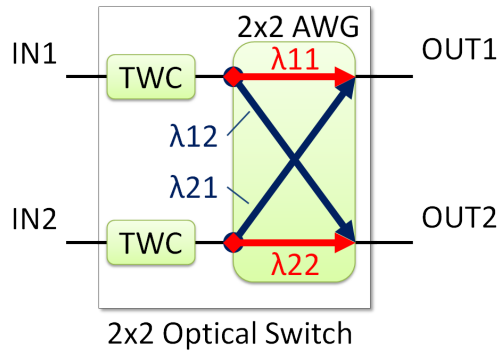
Figure 4.10: *A 2×2 optical switch implemented with tunable wavelength converters (TWCs) and an arrayed waveguide grating (AWG).*

cross-point contained active SOA elements that provided slight gain and an additional 40dB of isolation that potentially results in signal crosstalk levels below -70dB.

The wavelength routing fabrics are implemented using fast tunable wavelength converters (TWC) in conjunction with an arrayed waveguide grating (AWG) where a packet is switched to different output ports based its resultant wavelength. This approach is demonstrated in Figure 4.10 where λMN corresponds to a wavelength that is routed from input port M towards output port N. This approach is chosen in this work since it scales linearly with the switch port count. However, each TWC component is more complicated than its spatial switching counterpart resulting in a trade-off between scalability and device complexity.

Previous wavelength switching fabrics have been demonstrated utilizing cross-gain modulation (XGM) in SOAs and cross-phase modulation (XPM) in Mach-Zehnder interferometers (MZI). All-optical forwarding of 40Gb/s payloads and erasure of 10Gb/s headers was successfully achieved on a monolithically integrated InP platform by Lal *et al.* in [18]. The chip consisted of

a differential Mach-Zehnder interferometer wavelength converter (MZI-WC) integrated with a widely tunable sampled-grating distributed Bragg reflector (SG-DBR) laser and an electrical MZI modulator to perform the header rewrite. Recently, a monolithically integrated 8×8 all-optical switching fabric for 40Gb/s optical packets was shown by Nicholes *et al.* in [19]. Eight TWCs consisting of SG-DBR lasers and MZI-WCs were integrated with an 8×8 AWG on the same InP platform. Error-free operation was achieved through several channels which exhibited power penalties as low as 4.5dB while consuming less than 2W of drive power per channel.

The work presented in this dissertation utilizes an all-optical switching fabric similar to the work previously shown by Lal and Nicholes at UCSB. The forwarding plane in this work is implemented in a discrete fashion with an integrated widely tunable SG-DBR, an integrated MZI-WC, and an external LiNbO$_3$ Mach-Zehnder modulator (MZM). The tunable lasers are packaged in custom FPGA-based controller boards to achieve thermal stability and fast λ-switching speeds below 10ns. This implementation has been shown to be amenable to integration, but low-yielding fabrication led us to employ a discrete configuration.

4.2.4   Optical 3R Regeneration

All-optical signal regeneration is required to extend the reach of optical data routers (ODRs) and optical packet switching (OPS). Table 4.1 lists the different levels of signal regeneration along with a short descriptive passage. Signal re-amplification is classified as 1R regeneration. Simultaneous re-shaping and amplification is considered to be 2R regeneration. Re-shaping entails the improvement of signal quality by increasing signal-to-noise ratio (SNR). Finally, re-timing in addition to 2R results in 3R signal regeneration.
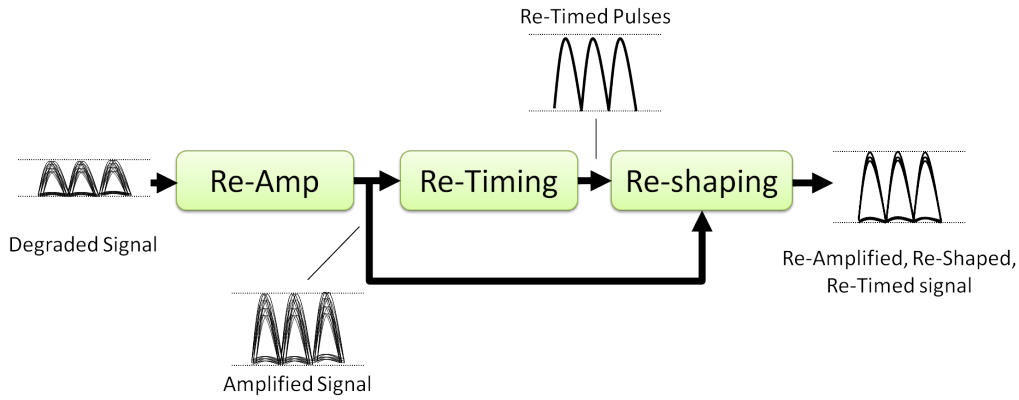
Figure 4.11: *Flow diagram showing optical re-amplification, re-shaping, and re-timing (3R).*

Re-timing is achieved by successfully demonstrating jitter reduction.

Figure 4.11 illustratively shows the configuration of a typical 3R signal regeneration circuit. A incoming data signal is shown with excessive amplitude noise and timing jitter. The degraded signal is transmitted through a re-amplification (1R) stage may be carried out by an integrated SOA or an erbium-doped fiber amplifier (EDFA). A re-timing stage is then used to extract a train of re-shaped pulses with reduced timing jitter. The re-timed output is then forwarded to an optical gate that is utilized to encode data onto the recovered pulses. The amplified output is tapped and used to drive

Table 4.1: *Requirements for several levels of signal regeneration.*

Regeneration Type	Description
1R	Re-Amplification
2R	Re-Shaping (noise reduction) + 1R
3R	Re-Timing (jitter reduction) + 2R

the optical gate. Ideally, the gate possesses re-shaping properties to reduce the amount of jitter and amplitude noise transferred onto the pulse train. Finally, gated pulses exit as regenerated signals that have been re-amplified, re-shaped, and re-timed.

4.2.4.1 Re-Amplification

Optical amplifiers are utilized in telecommunication systems to compensate for the losses observed in transmission medium and optical components. Amplification of optical signals is typically employed via EDFAs and SOAs. An EDFA typically provides polarization insensitive amplification with larger output powers, lower noise figure, and more optical gain bandwidth. Alternatively, an SOA is significantly more compact, exhibits lower carrier recovery lifetimes (good for optical processing), provides stronger nonlinear distortions (good for optical processing), and amplification is provided via electrical carrier pumping and not via a high-power optical pump. The work in this dissertation utilizes EDFAs to provide re-amplification since the 3R stage is implemented with discrete integrated devices. However, future monolithically implementations should benefit from integrated SOA technology.

4.2.4.2 Re-Shaping

A key functionality in signal re-shaping consists of increasing SNR by means of suppression of zero-level noise and one-level amplitude variations (noise re-distribution), extinction ratio (ER) enhancement, and pulse compression.

The basic operating principle behind signal reshaping is shown in Figure 4.12. A re-shaping component possesses an ideal step transfer function (top-left) where the output zero- and one-levels are clamped at constant values. For example, an incoming signal (bottom-left) is shown with zero- and
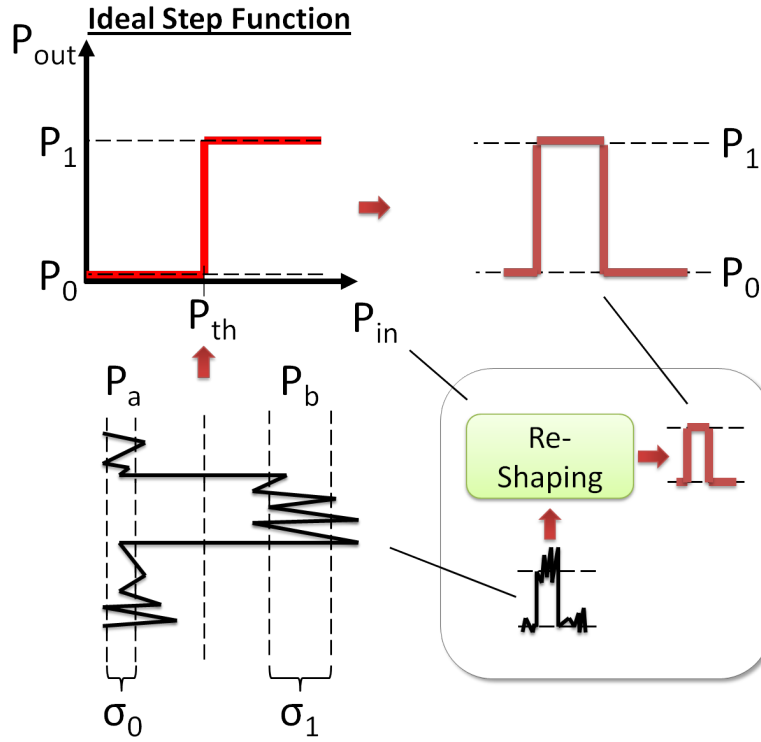
Figure 4.12: *Transfer functions used in signal re-shaping showing (left) ideal step, (middle) sinusoidal, and (right) gain saturation.*

one-levels, centered about P_a and P_b with root-mean-square (RMS) noise bands of σ_0 and σ_1 respectively. When a signal falls below the P_{th} input threshold, it results in instantaneous output clamped at P_0, while signals above the threshold produce a constant output of P_1. This leads to an ER enhancement provided $\dfrac{P_1}{P_0} > \dfrac{P_b}{P_a}$. Additionally, noise redistribution is successfully achieved since the RMS noise in the zero- and one-levels is suppressed at the output.

Figure 4.13 shows more realistic step transfer functions that are utilized in demonstrations of optical signal re-shaping. Different techniques utilizing the effects of optical nonlinearities, interferometry, or optical semiconductor carrier dynamics have been employed to produce the non-ideal step resem-
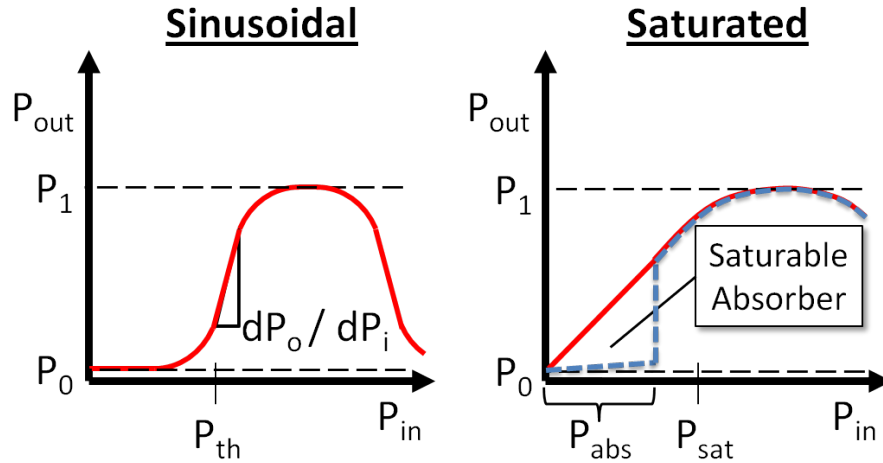
Figure 4.13: *Nonlinear power transfer functions used in signal re-shaping showing sinusoidal (left) and saturated (right) step functions.*

bling sinusoidal (left) and saturated (right) transfer functions to obtain signal re-shaping. The non-linear regimes of the sinusoid can be utilized to perform noise redistribution within the zero- and one-levels while ER can is increased if the slope of the linear regime approaches infinity ($\frac{dP_o}{dP_i} \to \infty$). Noise suppression in the one-level can be obtained by utilizing the saturation regime of the saturated transfer function, while zero-level noise reduction can be achieved by obtaining a transfer regime that behaves like a saturable absorber (SA).

Initial demonstrations of optical signal re-shaping, using the saturated step function, took advantage of fiber nonlinear phenomena and are shown in Figure 4.14. Optical re-shaping via wavelength conversion of 10 and 40Gb/s signals using cross-phase modulation (XPM) in highly nonlinear dispersion shifted fiber (HNL-DSF) was demonstrated in [20, 21]. A probe (λ1) and a pump (λ2) are co-propagated through a HNL-DSF. The incident probe signal induces a phase shift upon the pump via XPM and generates optical
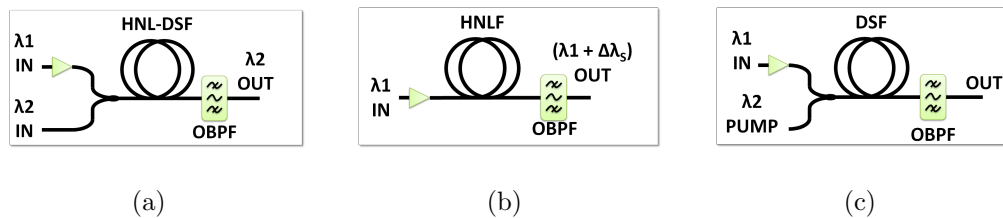
Figure 4.14: *(a) Schematics of fiber-based optical re-shaping utilizing cross-phase modulation (XPM), (b) self-phase modulation (SPM), and (c) four-wave mixing (FWM).*

side bands about $\lambda 2$. The bandpass filter (OBPF) is aligned to the $\lambda 2$ side lobes effectively converts phase modulation into amplitude modulation. A setup similar to (b) was used in [24, 25] to successfully redistribute amplitude noise via self-phase modulation (SPM) of 10 and 40Gb/s return-to-zero (RZ) signals. Pulses ($\lambda 1$) are transmitted through highly nonlinear fiber (HNLF) where spectral broadening ($\Delta\lambda_{SPM}$) is brought about by SPM. An OBPF is placed after the nonlinear medium, which is tuned to an offset wavelength ($\lambda 1 + \Delta\lambda_s$). Four-wave mixing (FWM) in dispersion shifted fibers (DSFs) was introduced in [27] as a scheme for signal all-optical re-shaping. These fiber-based reshaping implementations are well established and are capable of format preserving, high-speed operation beyond 40Gb/s. Yet, these bulky setups require high-power signals making it impractical to deploy in applications where energy efficiency and device footprint are crucial to system performance.

Integrated all-optical devices are needed to realize low-cost, scalable re-shaping implementations with compact physical footprints. The bulk of integrated devices demonstrating 2R re-shaping have utilized optical nonlinearities, akin to their fiber counterparts, in addition to inherent carrier dynamics within optical semiconductor materials. Demonstrations of signal re-shaping

have been previously carried out via integrated all-optical devices exploiting nonlinear optical Kerr effects that are typically realized using fiber lengths on the order of hundreds of meters.

Ta'eed *et al.* successfully demonstrated integrated all-optical 2R regenerators based on optical nonlinearities by utilizing chalcogenide glass (As_2S_3) waveguides integrated with Bragg grating filters. This waveguide material is transparent to infrared (IR) signals while it exhibits a high refractive index (n = 2-3), large third-order nonlinearities (n_2), and low two-photon absorption. Fabricated devices 6cm in length were utilized to obtain a re-shaping nonlinear power transfer function by inducing XPM and SPM within 50W optical 9MHz pulses [22,23]. More recently, Contestabile *et al.* demonstrated regenerative operation within a quantum dot SOA (QD-SOA) for 10, 20, and 40Gb/s RZ optical signals. The QD-SOA is biased into gain saturation with 2A of current to produce strong waveform distortions caused by SPM [26].

Initial high-speed demonstrations using wavelength conversion via FWM in monolithic devices were shown by Stephens *et al.* in [28]. The device was approximately 1.3mm in length and consisted of an SOA integrated with a distributed feedback (DFB) laser pump source. Shortly after, improvement in extinction ratio greater than 3dB of 2.5Gb/s optical signals was demonstrated by via FWM in an SOA [29]. All-optical signal re-shaping via ER enhancement of 10Gb/s RZ data was previously demonstrated by Salem *et al.* in [30]. A 300nm×500nm×1.6cm silicon waveguide is utilized to achieve a high nonlinear refractive index (n_2) along with a 2,000X enhancement of the nonlinear parameter (γ). A 100-milliwatt data signal and a 15-milliwatt pump are combined within the silicon waveguide to result in FWM wavelength conversion. An improvement in quality factor (Q) of about 1dB is observed, which is then extrapolated to an improvement in BER from

1.5×10^{-13} to 5.4×10^{-17}.

Regeneration of high-speed signals, which had been previously limited to bulky fiber-based setups, has been successfully demonstrated by utilizing optical nonlinearities within integrated devices. However, they require a prohibitive amount of optical power or electrical current that may compromise the lifetime and packaging reliability of devices under such operating conditions. As a result, numerous re-shaping experiments have taken advantage of carrier dynamics within integrated optical semiconductor devices to avoid using large amounts of optical power and biasing currents.

Optical signal re-shaping has been shown via suppression of zero-level noise by utilizing the transmission properties of integrated SAs. Early demonstrations achieved an improvement in link power penalty greater than 2dB for an RZ 10Gb/s optical link using a $200\mu m$ long SA waveguide exhibiting carrier recovery times below 9ps [31]. In the following years, regeneration of an RZ dispersion managed 40Gb/s long-haul wavelength division multiplexed (WDM) optical link was demonstrated using an integrated SA with carrier response times shorter than 5ps [32]. More recently, a vertically-coupled SA was utilized to achieve a transmission distance improvement factor greater than 3 for an RZ 42.6Gb/s transmission link spanning more than 10nm of optical bandwidth [33]. In [34], noise redistribution within the zero- and one-level was simultaneously achieved by integrating slightly reverse-biased SA sections with SOAs operated in saturation. Two SOA-SA pairs were utilized to obtain an ER enhancement greater than 5dB resulting in more than 4dB of power penalty improvement within an NRZ 10GB/s optical transmission link. A Q-factor of 13dB was achieved for four WDM channels after 1,300 and 7,600km of transmission distance with and without optical 2R regeneration, respectively.

All-optical signal re-shaping in the zero- and one-levels can be also simultaneously obtained by utilizing interferometric methods resulting in a sinusoidal transfer function. Integrated wavelength converters based on Mach-Zehnder interferometers (MZI) offer a power transfer function that is closest to being step-like. Wavelength conversion utilizing XPM within integrated SOAs have achieved regeneration of optical signals ranging from 10 to 20Gb/s where the operating speed was limited by the SOA carrier recovery [35]. In [36], a differential MZI configuration was introduced to increase operation recovery-limited bandwidth beyond 40Gb/s. Shortly after, all-optical 2R regeneration of RZ 40Gb/s signals was achieved using an MZI wavelength converter (MZI-WC) integrated on an InP platform [37].

Performing re-shaping via nonlinear optical phenomena in integrated devices has the potential of achieving higher data bit rates since the dynamic recovery times are much shorter compared to carrier lifetimes in optical semiconductors. However, this approach requires high-power optical pump and probe signals along with large amounts of bias currents. Utilizing integrated SOA-SA pairs for noise redistribution of the zero and one-levels is an attractive alternative that is limited primarily by carrier recovery dynamics that can be engineered for high-speed operation. Similar engineering schemes can be applied to the MZI-WC approach to achieve regeneration for signals beyond 20Gb/s. The switching bandwidth can be enhanced past 40Gb/s via differential push-pull signaling. Additionally, the MZI approach allows one to achieve a power transfer function that most closely resembles the ideal step since the slope of the linear transfer regime $(\frac{dP_o}{dP_i})$ can be engineered to be relatively steep. Achieving power balance between each MZI arm is a major challenge that must be mitigated to achieve near optimal constructive and destructive interference. Consequently, uncertainties in the fabrication
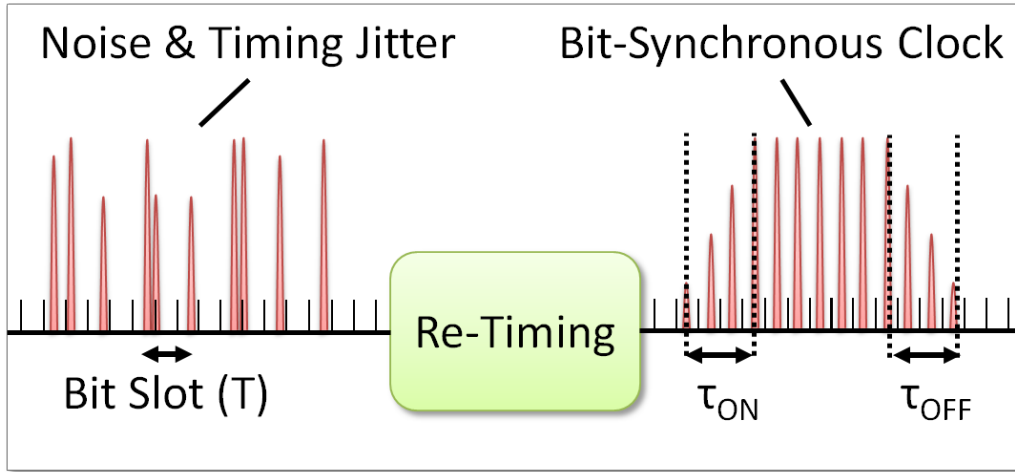
Figure 4.15: *Diagram of basic principal of operation behind signal re-timing showing incoming data (left) exhibiting timing jitter and re-timed clock tone aligned to bit slots (T).*

process may lead to large performance variations from wafer to wafer.

The 2R regenerators used in this work consist of integrated MZI-WCs. Optical 1R signal regeneration is successfully demonstrated by achieving re-amplification where the modulation depth of incoming signals is increased from 1 to 2mW. Additionally, 2R regeneration is obtained via signal re-shaping where zero-level noise is suppressed. The performance is limited by imbalance within the interferometer arms resulting in non-optimal constructive/destructive interference which is addressed in later chapters.

4.2.4.3 Re-Timing

Optical 2R regeneration facilitates the compensation of signal noise that accumulates from nodes to node, but it is unable to inhibit the accumulation of timing jitter. As a result, every node needs to apply signal re-timing onto incoming data signals to reduce timing jitter to further increase transmis-

sion quality and distance. Figure 4.15 shows a basic diagram describing the operating principle behind signal re-timing. An incoming optical data signal (left) enters the re-timing stage exhibiting timing jitter and amplitude noise varying from bit to bit. The re-timing stage recovers an optical clock signal consisting of a pulse train whose pulses are phase aligned to the data bit slot (T). The clock tone is recovered within finite turn-on and turn-off times τ_{ON} and τ_{OFF} respectively.
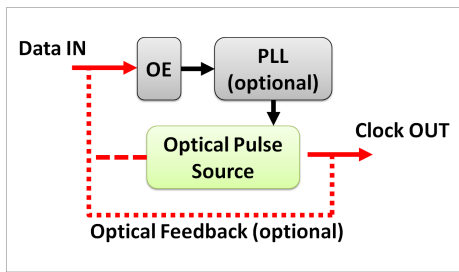
Figure 4.16: *Optical clock recovery employing an opto-electronic (OE) conversion.*
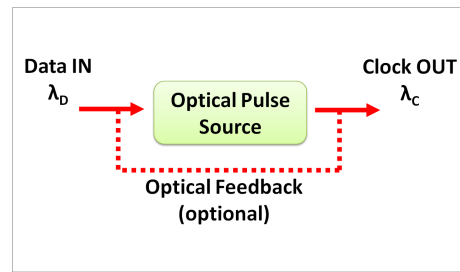
Figure 4.17: *All-optical clock recovery performed via injection locking.*

The most promising work in the field of optical clock recovery can be categorized into implementations using opto-electronic (OE) conversions and all-optical injection locking in order to successfully achieve optical clock recovery (OCR). Figure 4.16 shows the diagram of an optical clock recovery circuit where incoming data is converted to an electronic signal via an OE conversion. The electrical signal may then pass through an optional phase-locked loop (PLL) to lock a voltage controlled oscillator (VCO) reference signal before using it to drive the optical pulse source. If the incoming input signals consists of burst mode data, then the OCR output will consist of clock bursts with finite turn-on and turn-off times. The duration of the clock bursts may be extended by applying an optional optical feedback loop that is

fed into either the pulse source or the OE conversion stage. Alternatively, the circuit diagram in Figure 4.17 describes an implementation of OCR where the incoming data signal (λ_D) is used to injection-lock the optical pulse source. The injected data induces some form of modulation within the pulse source resulting in a recovered optical clock (λ_C). Finally, the duration of optical clock bursts resulting from burst mode input can be elongated by providing an optional optical feedback path towards the input of the pulse source.

Previously, a demonstration of optical 3R regeneration utilizing OE-based re-timing was shown for 40Gb/s RZ optical signals [38]. The re-timing stage was implemented by performing an OE conversion of incoming data to lock a 10GHz VCO electronic feedback loop, similar to [39]. The recovered 10GHz clock was then sent through two frequency doublers to obtain a 40GHz clock tone that was used to drive two LiNbO$_3$ modulators, connected in tandem, serving as optical pulse carvers. The recovered pulse train was then counter-propagated against the data signal within a re-shaping MZI-WC in push-pull (differential) configuration. A couple years afterward, 40Gb/s 3R regeneration was successfully achieved using a MLL-based OCR while an electro-absorption modulator (EAM) provided a nonlinear gating function [40]. The OCR stage was implemented with an optical phase locked loop (OPLL) where the OCR output (λ_C) and the data (λ_D) are used as inputs to a balanced detector. The detector provides an error signal to a VCO, which in turn drives an integrated MLL. The recovered clock signal is then counter-propagated against the data within the EAM where cross-absorption modulation (XAM) is used to transfer the data onto the pulse train. Several years later, Hu *et al.* demonstrated 40Gb/s OCR results using methods similar to the aforementioned work where a traveling wave EAM (TW-EAM), a coplanar waveguide (CPW) Q-filter, and a 40GHz narrow band amplifier (NB-AMP) were in-

tegrated on a 3.5×10mm AlN carrier [41, 42]. A circulating resonator was created by connecting one end of the TW-EAM transmission line to the NB-AMP, which led to the Q-filter. The filter was then connected to the opposite end of the TW-EAM. Lastly, Koch *et al.* demonstrated 10 [43] and 40Gb/s [44] OCR where significant reduction in timing was observed via hybrid locking of MLLs. The clock recovery implementation consisted of an OE conversion stage leading to a NB-AMP whose output was used to modulate a SA within in the laser cavity.

Past all-optical clock extraction experiments have also been accomplished via pulse injection locking of integrated mode-locked lasers. Clock recovery and jitter reduction have been demonstrated by Arahira *et al.* at 48.5 [45] and 160GHz [46–49] by utilizing sub-harmonic optical signal injection of 10 and 40Gb/s optical data signals respectively. Both experiments utilized integrated MLLs, fabricated on an InP substrate. The laser cavity was defined by etched facets and it included two gain sections and a saturable absorber section that was modulated by incoming data pulses. A few years later, Koch *et al.* successfully achieved all-optical clock recovery with reduced timing jitter by utilizing monolithic MLLs possessing repetition rates ranging from 30.4 [50] to 35Gb/s [51], which were fabricated on hybrid silicon and InP integration platforms, respectively. The hybrid silicon laser was designed as a ring laser with two gain sections and a saturable absorber that was used to phase-lock the longitudinal modes. An ER enhancement greater than 6dB was observed with a reduction in RMS jitter from 14 to 1.7ps. The InP-based MLL design consisted of a 1mm long cavity lithographically defined by DBR mirrors rather than by etched facets. The cavity incorporated two gain sections, a phase section, and a saturable absorber that served as a variable loss element. The jitter of the recovered clock was comparable to the near-ideal

input data signal, but a chip gain greater than 10dB was observed as an SOA was integrated with the MLL.

The most practical approaches to all-optical signal re-timing employ clock recovery stages utilizing integrated mode-locked lasers (MLLs). While it is difficult to electronically generate narrow pulse trains with high-repetition rates, it is relatively straightforward to implement in the optical domain. For example, the MLL repetition rate (f_R) can be increased to higher frequencies by scaling down the length of the laser cavity. Moreover, the pulse widths can be shortened by increasing the optical bandwidth.

Generally, less than 1mW of injected optical power is required to lock the laser, but this locking sensitivity will increase as additional optical components are integrated with the MLL. For example, the integration of SOAs will result in unwanted optical feedback in the form of amplified spontaneous emission (ASE) noise and backward-propagating pulses reflected at downstream interfaces. As a result, one would need to increase the power of the injected pulses, which may become problematic if they travel beyond the laser cavity. An attractive compromise would be to utilize an OE conversion to enable hybrid locking of the MLL while the optical data signal is isolated from the laser cavity. This scheme is potentially disadvantageous as it is difficult to scale the operating frequency (f_L) of the hybrid locking circuit beyond 40GHz. Alternatively, one could utilize sub-harmonic hybrid locking where an ultra-fast MLL ($f_R >$100GHz) is locked by modulating the SA section at a frequency that is a multiple of the repetition rate ($f_L < $40GHz).

The work presented in this dissertation utilizes InP-based MLLs, similar to the ones demonstrated by Koch *et al.*, to perform optical clock recovery and re-timing of 40Gb/s optical data signals. A hybrid locking circuit consisting of an OE conversion and two RF amplifier-filter pairs is used to

modulate a saturable absorber region. The recovered pulse train is encoded with data by utilizing a re-shaping MZI-WC optical gate. A measured RMS jitter reduction greater than 0.2ps is successfully achieved limited by the jitter transfer bandwidth and the on-chip optical gain of the integrated MLL. A detailed description of these limitations is discussed in the upcoming chapters.

4.3 Electronic Control Plane

The electronic control plane of the LASOR router consists of parallel clock and data recovery (CDR) and payload envelope detection (PED) stages that are used to recover the labeled header and the temporal location of incoming optical packets. The header and PED signals are then sent to the electronic channel processor to process the port request of each packet. Port requests are forwarded to the ARB, which dictates packet precedence and appropriate ECP signaling.

4.3.1 Clock and Data Recovery

A burst mode CDR circuit is required in order to successfully recover the labeled header containing the forwarding information of each packet. A clock tone with frequency and phase comparable to the bit rate of the header is used to latch recovered bits into the ECP. Figure 4.18 shows an illustration describing the CDR principle of operation. An optical signal (left) enters the CDR stage, which outputs its electrical equivalent (top-right) along with a corresponding clock burst. Every bit-level transition of the recovered data is aligned to the phase of the clock burst that exhibits a finite turn-on (τ_{ON}) and turn-off (τ_{OFF}) delay corresponding to the beginning an end of the in-

77
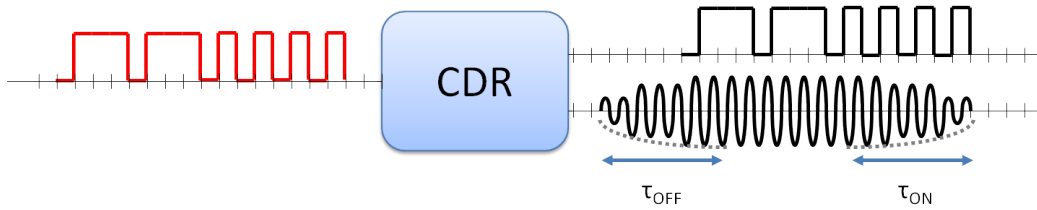
Figure 4.18: *Diagram showing optical data signal (left) and recovered electronic data (top-right) and clock (bottom-right) showing finite turn-on (τ_{ON}) and turn-off (τ_{OFF}) times.*

coming burst of data. Typical receiver designs rely on an approximate clock reference that is phase-aligned to the received data by means of phase-locked loop circuits, which is sufficient for data streams that are either continuous or arrive in relatively large bursts. This approach becomes impractical when attempting to receive burst mode data whose durations are 100s of nanoseconds. The burst mode clock and data recovery implemented in this work consists of a 10GHz CDR circuit similar to what is demonstrated in [52]. The design consists of a photo-detector followed by a limiting amplifier that evenly splits between a clock and a data path. The clock recovery path is comprised of a passive 10GHz narrow band filter ($\Delta f = 400$MHz) followed by a 10Gb/s trans-impedance amplifier that is again filtered. The recovered clock tone is then amplified via a 10GHz limiting amplifier in an attempt to reduce τ_{ON} and τ_{OFF}. The packet format used in this work is utilizes 32-bit sequences of alternating zeros and ones (0xAAAAAAAA) at the beginning of the header and payload to serve as initialization bits for the clock recovery process.
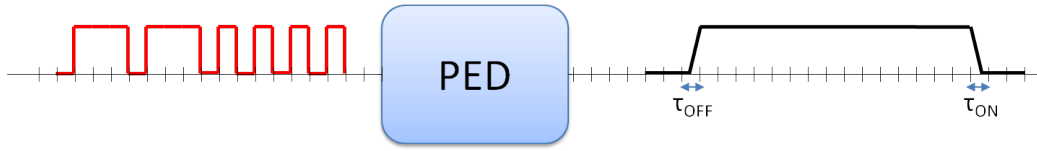
Figure 4.19: *Illustration showing optical data signal (left) and recovered electronic payload envelope (right) displaying finite turn-on (τ_{ON}) and turn-off (τ_{OFF}) times.*

4.3.2 Payload Envelope Detection

Payload envelope detection is an effective method of determining the approximate temporal location of an optical packet without having to process each bit of the high-speed payload. Figure 4.19 describes the basic operating principle behind payload envelope detection. A high-speed optical signal (left) exhibiting several data transitions enters the PED circuit, which outputs an electrical envelope function where the data transitions are suppressed. Finite turn-on (τ_{ON}) and turn-off (τ_{OFF}) delays are observed at the beginning and end of the PED signal, which are dominated by the inherent dynamics of the method used to implement the circuit. Minimal rise/fall times are desirable traits that can potentially lead to improved temporal detection accuracy and reduced guard bands.

Fast-switching payload envelope detection schemes have been previously implemented via photonic integrated circuits (PICs). The method utilized in [53,54] consists of a 40GHz resonant circuit that is used to generate a PED signal from 40Gb/s payloads while out-of-resonance 10Gb/s labeled headers are detected at the drop port. Integrated resonant laser cavities utilized in conjunction with cross-gain modulation (XGM) in a semiconductor optical amplifier (SOA) have also been included in PED implementations [55]. The
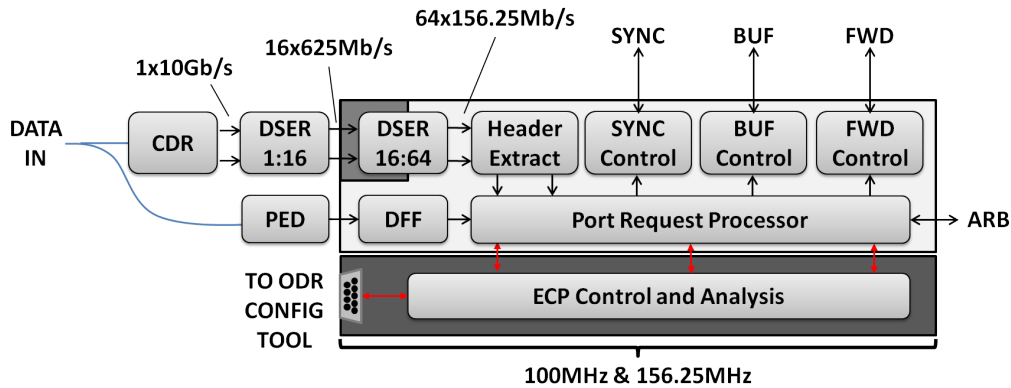
Figure 4.20: *Schematic representation of FPGA-based electronic channel processor (ECP) implementation.*

PED switching times demonstrated in these designs were dominated by recovery times inherent to InP semiconductor material, which resulted in sub-nanosecond switching times. The PED circuit in [56] utilizes a bandwidth limited design consisting of a 10GHz photo-detector followed by a 2.5GHz limiting amplifier to obtain PED signals with nanosecond-scale turn on and off times. Here, PED switching speed was sacrificed to reduce implementation costs by utilizing commercial components with bandwidths much lower than payload bit rates.

The work presented in this thesis makes use of a PED circuit similar to the bandwidth limited design previously mentioned. Additionally, an all-optical PED implementation consisting of two XGM feed-forward SOA stages is also carried out in order to achieve switching speeds on the order of 350ps.

4.3.3 Electronic Lookup and Arbitration

Figure 4.20 shows the process of header and payload recovery along with a schematic representation of the logical building blocks within the FPGA-

80

based ECP implementation. Incoming labeled optical packets are evenly distributed between the CDR and the PED paths using a 3dB optical coupler. The CDR stage is used to generate a 10Gb/s clock-data pair that is phase-aligned at the bit-level transitions. The CDR output is forwarded to a 1:16 de-serializer (DSER) that is used to parallelize the 10Gb/s data stream into sixteen 625Mb/s streams. The clock-data pair from the DSER is then sent towards the input of an FPGA-based 16:64 DSER that converts the 16 data streams into sixty-four 156.25Mb/s parallel data lines. The 156.25Mb/s clock-data pair is transmitted to the Header Extraction stage where burst mode recovery of the low-speed header is performed. When successful detection is achieved, a 6.4ns HEADER-DETECTED trigger pulse is forwarded to the Port Request Processor (PRQ).

The PED signal derived from incoming packets is latched into the PRQ using a 156.25MHz D flip-flop (DFF). The leading edge of the PED signal is used to ascertain the amount of synchronization needed at the SYNC stage to temporally align payloads to the local router time slot. The PRQ then uses the label from the recovered header to perform and electronic table look up to determine the desired port of the incoming packet. The port request and the latched PED signal are then forwarded to the ARB where packet contention is detected and mitigated. The PRQ then generates the appropriate BUF and FWD control signals that reflect the commands issued by the ARB.

4.4   Chapter Summary

The architecture of a 2×2 label swapped optical router has been presented where each core subsystem is discussed in detail. A brief overview describing the research previously done for each subsystem has been shown along with

reasoning behind the design paths chosen for this work. Each arm of the router consists of an optical data path and an electronic control path. The optical data path is comprised of a synchronizer, buffer, forwarding plane, and a 3R signal generator. The synchronizer is utilized to align incoming labeled optical packets to the local router time slot to minimize the packet guard bands required to efficiently switch packets through the routing fabric. Synchronization is performed by switching packets through a variable delay depending on the amount of packet-to-time slot misalignment. Optical packet buffers are needed to resolve contention between packets requesting identical output ports. A buffer design consisting of a re-circulating fiber delay coupled to an integrated switch matrix has been demonstrated as a means of mitigating contention between 40Gb/s optical packets. Though packet synchronization is not performed in this work, the variable delay technology shown is inserted in place of the re-circulating fiber delay to implement a re-sizable optical buffer. As a result, variable-length 40Gb/s packets up to 800 bytes in length can be buffered by utilizing a variable delay designed in a four-stage, feed-forward configuration using commercially available components. A switching fabric comprised of integrated, fast-switching, tunable wavelength converters is used to route and forward outgoing packets at nanosecond-scale speeds with sub-nanosecond accuracy. The electronic control path consists of a clock and data recovery stage (CDR), a payload envelope detection (PED) stage, an electronic channel processor (ECP), and an arbitration unit (ARB). A 10Gb/s CDR circuit is presented that extracts a 10GHz clock burst that is phase-aligned to the 10Gb/s data to allow for the successful detection of labeled headers. A PED circuit comprised on low-cost, commercially available components is presented as a means of detecting the temporal location of 10 and 40Gb/s labeled headers and high-speed payloads

respectively. The ECP is provides FPGA-based electronic lookup functionality where labels extracted from detected headers are used to process port requests at a local clock rate of 156.25MHz. Port requests from each ECP are forwarded to an arbitration FPGA, that detects contention and delegates packet priority. System-level performance is evaluated via frame recovery (Layer-III) measurements which builds upon the header recover (Layer-II) measurement scheme previously used to quantify performance at the optical data link layer.

References

[1] H. Yang, V. Akella, C. N. Chuah, S. J. B. Yoo, "Scheduling optical packets in wavelength, time, and space domains for all-optical packet switching routers," *IEEE International Conference on Communications, (ICC) 2005*, vol. 3, pp. 1836-1842, 2005.

[2] T. Sakamoto, A. Okada, M. Hirayami, Y. Sakai, O. Moriwaki, I. Ogawa, R. Sato, K. Noguchi, M. Matsuoka, "Optical Packet Synchronizer Using Wavelength and Space Switching," *IEEE Photonics Technology Letters*, vol. 14, no. 9, pp. 1360-1362, September 2002.

[3] J. P. Mack, H. N. Poulsen, D. J. Blumenthal, "Variable Length Optical Packet Synchronizer," *IEEE Photonics Technology Letters*, vol. 20, no. 14, pp. 1252-1254, July 2008.

[4] R. L. Moreira, J. M. Garcia, W. Li, J. S. Barton, D. J. Blumenthal, "Longest Fully Integrated Ultra-Low-Loss 4-Bit Tunable Delay for Broadband Phased Array Antenna Applications," *IEEE Photonics Technology Letters*, vol. 25, no. 12, June 2013.

[5] A. Pattavina, "Architectures and Performance of Optical Packet Switching Nodes for IP Networks," *Journal of Lightwave Technology*, vol. 23, no. 3, pp. 1023-1032, March 2005.

[6] E. F. Burmeister, D. J. Blumenthal, J. E. Bowers, "A comparison of optical buffering technologies," *Optical Switching and Networking*, vol. 6, pp. 10-18, January 1, (2007).

[7] J. Mork, R. Kjaer, M. van der Poel, K. Yvind, "Slow light in a semiconductor waveguide at gigahertz frequencies," *Optics Express*, vol. 13, no. 20, pp. 8136-8145, October 2005.

[8] Y. Okawachi, J. E. Sharping, A. L. Gaeta, M. S. Bigelow, A. Schweinsberg, R. W. Boyd, Z. Zhaoming, D. J. Gauthier, "All-optical tunable slow-light delays via stimulated scattering," *Optics Photonics News*, vol. 16, no. 12, p. 42, 2005.

[9] Y. A. Vlasov, S. J. McNab, "Coupling into the slow light mode in slab-type photonic crystal waveguides," *Optics Letters*, vol. 31, no. 1, pp. 50-52, January 2006.

[10] R. S. Tucker, "Slow-Light Optical Buffers: Capabilities and Fundamental Limitations," *Journal of Lightwave Technology*, vol. 23, no. 12, pp. 4046-4066, December 2005.

[11] M. O'Mahony, K. M. Guild, D. Hunter, I. Andanovic, I. White, R. Penty, "An optical packet switched network (WASPNET)-concept and realisation." *Optical Networks Magazine* 2, no. 6 pp. 46-53, 2001.

[12] J. Gripp, D. Stiliadis, J. E. Simsarian, P. Bernasconi, J. D. Le Grange, L. Zhang, L. Buhl, D. T. Neilson, "IRIS optical packet router," *Journal of Optical Networking*, vol. 5, Issue 8, pp. 589-597, 2006.

[13] N. Chi, Z. Wang, S. Yu, "A Large Variable Delay, Fast Reconfigurable Optical Buffer Based on Multi-Loop Configuration and an Optical Cross-

point Switch Matrix," *Optical Fiber Communications Conference, (OFC) 2006*, Paper OFO7, March 2006.

[14] E. F. Burmeister, J. P. Mack, H. N. Poulsen, J. Klamkin, L. A. Coldren, D. J. Blumenthal, J. E. Bowers, "SOA gate array recirculating buffer with fiber delay loop," *Optical Fiber Communications Conference, (OFC) 2008*, Paper OWe4, February 2008.

[15] Hyundai Park, John P. Mack, Daniel J. Blumenthal, J. E. Bowers, "An integrated recirculating optical buffer," *Optics Express*, vol. 16, Issue 15, pp. 11124-11131, July 2008.

[16] G. Kurczveil, "Hybrid Silicon AWG Lasers and Buffers," Ph.D. Dissertation, University of California, Santa Barbara, 2008.

[17] R. Varrazza, I. B. Djordjevic, S. Yu, "Active Vertical-Coupler-Based Optical Crosspoint Switch Matrix for Optical Packet-Switching Applications," *Journal of Lightwave Technology*, vol. 22, no. 9, pp. 2034-2042, September 2004.

[18] V. Lal, M. Masanovic, D. Wolfson, G. Fish, C. Coldren, D. J. Blumenthal, "Monolithic Widely Tunable Optical Packet Forwarding Chip in InP for All-Optical Label Switching with 40 Gbps Payloads and 10 Gbps Labels," *European Conference on Optical Communications, (ECOC) 2005*, vol. 6, Paper Th4.2.1, pp. 25-26, September 2005.

[19] S. C. Nicholes, M. L. Mašanović, B. Jevremović, E. Lively, L. A. Coldren, D. J. Blumenthal, "An 8×8 InP Monolithic Tunable Optical Router (MOTOR) Packet Forwarding Chip," *Journal of Lightwave Technology*, vol. 28, no. 4, pp. 641-650, February 2010.

[20] B. Olsson, P. Ölén, L. Rau, D. J. Blumenthal, "Simple and Robust 40-Gb/s Wavelength Converter Using Fiber Cross-Phase Modulation and Optical Filtering," *IEEE Photonics Technology Letters*, vol. 12, no. 7, pp. 846-848, July 2000.

[21] J. Suzuki, T. Tanemura, K. Taira, Y. Ozeki, K. Kikuchi, "All-Optical Regenerator Using Wavelength Shift Induced by Cross-Phase Modulation in Highly Nonlinear Dispersion-Shifted Fiber," *IEEE Photonics Technology Letters*, vol. 17, no. 2, pp. 423-425, February 2005.

[22] V. G. Taeed, M. R. E. Lamont, D. J. Moss, B. J. Eggleton, D.-Y. Choi, S. Madden, B. Luther-Davies, "All optical wavelength conversion via cross phase modulation in chalcogenide glass rib waveguides," *Optics Express*, vol. 14, no. 23, pp. 11242-11247, 2006.

[23] V. G. Taeed, M. Shokooh-Saremi, L. Fu, I. C. M. Littler, D. J. Moss, M. Rochette, B. J. Eggleton, Y. Ruan, B. Luther-Davies, "All optical wavelength conversion via cross phase modulation in chalcogenide glass rib waveguides," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 12, no. 3, pp. 11242-11247, May 2006.

[24] P. V. Mamyshev, "All-optical data regeneration based on self-phase modulation effect," *European Conference on Optical Communications, (ECOC) 1998*, pp. 20-24 September 1998.

[25] T. Her, L. Lens, G. Raybon, J.-C. Bouteiller, C. Jorgensen, K. Feder, K. Brar, P. Steinvurzel, D. Patel, N. M. Litchinister, P. S. Westbrook, L. E. Nelson, C. Headley, B. J. Eggleton, "Enhanced 40-Gbit/s Receiver Sensitivity with All-fiber Optical 2R Regenerator," *Conference on Lasers*

and Electro-Optics, (CLEO) 2002, vol. 1, pp. 534-535, Paper CThO42, 2002.

[26] G. Contestabile, A. Maruta, S. Sekiguchi, K. Morito, M. Sugawara, K. Kitayama, "Regenerative Amplification by Using Self-Phase Modulation in a Quantum-Dot SOA," *IEEE Photonics Technology Letters*, vol. 22, no. 7, pp. 492-494, April 2010.

[27] E. Ciaramella, "A new scheme for all-optical signal reshaping based on wavelength conversion in optical fibers," *Optical Fiber Communications Conference, (OFC) 2000*, vol. 2, pp. 320-322, 2000.

[28] M. F. C. Stephens, D. Nesset, R. V. Penty, I.H. White, M. J. Fice, "Wavelength conversion at 40Gbit/s via four wave mixing in semiconductor optical amplifier with integrated pump laser," *IEEE Electronics Letters*, vol. 35, no. 5, pp. 420-421, March 1999.

[29] C. Gosset, G.-H. Duan, "Extinction Ratio Improvement and Wavelength Conversion Based on Four-Wave Mixing in a Semiconductor Optical Amplifier," *IEEE Photonics Technology Letters*, vol. 13, no. 2, pp. 139-141, February 2001.

[30] R. Salem, M. A. Foster, A. C. Turner, D. F. Geraghty, M. Lipson, A. L. Gaeta, "Signal regeneration using low-power four-wave mixing on silicon chip," *Nature Photonics*, vol. 2, no. 1, pp. 35-38, January 2008.

[31] H. Yokoyama, Y. Hashimoto, H. Kurita, "Noise reduction in optical pulses and biterror-rate improvement with a semiconductor-waveguide saturable absorber," *Lasers and Electro-Optics, (CLEO) 1998*, pp. 502-503, Paper CFB5, May 1998.

[32] D. Rouvillain, F. Seguineau, L. Pierre, P. Brinde, H. Choumane, G. Aubin, J.-L. Oudar, O. Leclerc, "40Gbit/s Optical 2R Regenerator based on passive Saturable Absorber for WDM long-haul transmissions," *Optical Fiber Communications Conference, (OFC) 2002*, pp. FD11-1, FD11-3, March 2002.

[33] L. Bramerie, Q. T. Le, M. Gay, A. OHare, S. Lobo, M. Joindot, J.-C. Simon, H.-T. Nguyen, J.-L. Oudar, "All-optical 2R regeneration with a vertical microcavity based saturable absorber," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 18, no. 2, pp. 870-883, July 2011.

[34] T. Vivero, N. Calabretta, I. T. Monroy, G. Kassar, F. ÖhÖman, A. P. González-Marcos, J. Mørk, "2R-Regeneration in a monolithically integrated four-section SOAEA chip," *Optics Communications*, vol. 282, no. 1, pp. 117-121, January 2009.

[35] J. De Merlier, G. Morthier, S. Verstuyft, T. Van Caenegem, I. Moennan, P. Van Daele, R. Baets, "A novel 2R regenerator based on an asymmetric Mach-Zehnder interferometer incorporating an MMI-SOA," *IEEE Lasers and Electro-Optics Society, (LEOS) 2001*, pp. 366-367, 2001.

[36] B. Mikkelsen, K. S. Jepsen, M. Vaa, H. N. Poulsen, K. E. Stubkjaer, R. Hess, M. Duelk, W. Vogt, E. Gamper, E. Gini, P. A. Besse, H. Melchior, S. Bouchoule, F. Devau, "All-optical wavelength converter scheme for high speed RZ signal formats," *Electronics Letters*, vol. 33, no. 25, pp. 2137-2139, December 1997.

[37] D. Wolfson, P. B. Hansen, T. Fjelde, A. Kloch, C. Janz, A. Coquelin, I. Guillemot, F. Gaborit, F. Poingt, M. Renaud, "40 Gbit/s all-optical 2R regeneration in an SOA-based all-active Mach-Zehnder interferometer,"

Communications, 1999. APCC/OECC 'Asia-Pacific Conference on ... and Fourth Optoelectronics and Communications Conference*, vol. 1, pp. 456-457, October 1999.

[38] S. Fischer, M. Diilk, E. Gamper, W. Vogt, E. Gini, H. Melchior, W. Hunziker, D. Nesset, A.D. Ellis, "Optical 3R regenerator for 40Gbit/s networks," *IEEE Electronics Letters*, vol. 35, no. 23, pp. 2047-2049, November 1999.

[39] A. D. Ellis, T. Widdowson, X. Shan, G. E. Wickens, D. M. Spirit, "Transmission of a true single polarisation 40 Gbit/s soliton data signal over 205 km using a stabilised erbium fibre ring laser and 40 GHz electronic timing recovery," *IEEE Electronics Letters*, vol. 29, no. 11, pp. 990-992, May 1993.

[40] E. S. Awad, P. S. Cho, C. Richardson, N. Moulton, J. Goldhar, "Optical 3R Regeneration Using a Single EAM for All-Optical Timing Extraction With Simultaneous Reshaping and Wavelength Conversion," *IEEE Photonics Technology Letters*, vol. 14, no. 9, September 2002.

[41] Z. Hu, H.-F. Chou, K. Nishimura, M. Usami, J. E. Bowers, D. J. Blumenthal, "Optical Clock Recovery Circuits Using Traveling-Wave Electroabsorption Modulator-Based Ring Oscillators for 3R Regeneration," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 11, no. 2, pp. 329-337, April 2005.

[42] Z. Hu, B. R. Koch, J. E. Bowers, D. J. Blumenthal, "Integrated Photonic/RF 40-Gb/s Burst-mode Optical Clock Recovery for Asynchronous Optical Packet Switching," *Optical Fiber Communications Conference, (OFC) 2005*, Paper OTuO, March 2005.

[43] B. R. Koch, A. W. Fang, O. Cohen, J. E. Bowers, "Mode-locked silicon evanescent lasers," *Optics Express*, vol. 15, no. 18, pp. 11225-11233, September 2007.

[44] B. R. Koch, J. S. Barton, M. Mašanović, Z. Hu, J. E. Bowers, D. J. Blumenthal, "Monolithic Mode-Locked Laser and Optical Amplifier for Regenerative Pulsed Optical Clock Recovery," *IEEE Photonics Technology Letters*, vol. 19, no. 9, pp. 641-643, May 2007.

[45] S. Arahira, S. Kutsuzawa, Y. Ogawa, "Extreme Timing Jitter Reduction of a Passively Mode-Locked Laser Diode by Optical Pulse Injection," *IEEE Journal of Quantum Electronics*, vol. 35, no. 12, pp. 1805-1811, December 1999.

[46] S. Arahira, Y. Katoh, D. Kunimatsu, S. Kutsuzawa, Y. Ogawa, "160 GHz mode-locked laser diode stabilized by subharmonic-frequency optical pulse injection," *Lasers and Electro-Optics Society, (LEOS) 1999*, vol. 2, pp. 836-837, 1999.

[47] S. Arahira, Y. Katoh, D. Kunimatsu, Y. Ogawa, "Stabilization and timing jitter reduction of 160 GHz colliding-pulse mode-locked laser diode by subharmonic-frequency optical pulse injection," *IEICE Trans. Electron.*, vol. E83-C, pp. 966973, 2000.

[48] S. Arahira, S. Sasaki, K. Tachibana, Y. Ogawa, "All-Optical 160-Gb/s Clock Extraction With a Mode-Locked Laser Diode Module," *IEEE Photonics Technology Letters*, vol. 16, no. 6, pp. 1558-1560, June 2004.

[49] S. Arahira, Y. Ogawa, "Retiming and Reshaping Function of All-Optical Clock Extraction at 160 Gb/s in Monolithic Mode-Locked Laser Diode,"

IEEE Journal of Quantum Electronics, vol. 41, no. 7, pp. 937-944, July 2005.

[50] B. R. Koch, A. W. Fang, H. N. Poulsen, H. Park, D. J. Blumenthal, J. E. Bowers, R. Jones, M. J. Paniccia, O. Cohen, "All-Optical Clock Recovery with Retiming and Reshaping Using a Silicon Evanescent Mode Locked Ring Laser," *Optical Fiber Communications Conference, (OFC) 2008*, Paper OMN1 (Invited), February 2008.

[51] B. R. Koch, J. S. Barton, M. Masanovic, Z. Hu, J. E. Bowers, D. J. Blumenthal, "35 Gb/s Monolithic All-Optical Clock Recovery Pulse Source," *Optical Fiber Communications Conference, (OFC) 2007*, Paper OWP2, March 2007.

[52] H.N. Poulsen, D. Wolfson, S. Rangarajan, D. J. Bumenthal, "Burst Mode 10Gbps Optical Header Recovery and Lookup Processing for Asynchronous Variable-Length 40 Gbps Optical Packet Switching," *Optical Fiber Communications Conference, (OFC) 2006*, Paper, OThS7, March 2006.

[53] Z. Hu, R. Doshi, H-F Chou, H. N. Poulsen, D. Wolfson, J. E. Bowers, D. J. Blumenthal, "Optical label swapping using payload envelope detection circuits," *IEEE Photonics Technology Letters*, vol. 17, no. 7, pp. 1537-1539, 2005.

[54] B. R. Koch, Z. Hu, J. E. Bowers, D. J. Blumenthal, "Integrated optical payload envelope detection and label recovery device for optical packet switching networks," *Optics Express*, vol. 14, Issue 12, pp. 5073-5078, 2006.

[55] B. R. Koch, Z. Hu, J. E. Bowers, D. J. Blumenthal, "All-Optical Payload Envelope Detection for Variable Length 40-Gb/s Optically Labeled Packets," *IEEE Photonics Technology Letters*, vol. 18, no. 17, pp. 1846-1848, September 2006.

[56] J. P. Mack, E. F. Burmeister, J. M. Garcia, H. N. Poulsen, B. Stamenic, G. Kurczveil, K. N. Nguyen, K. Hollar, J. E. Bowers, D. J. Blumenthal, "Synchronous Optical Packet Buffers," *IEEE Journal of Selected Topics Quantum Electronics*, vol. 16, no. pp. 1413-1421, October 2010.

Chapter 5

Edge Adaptation Background

This chapter presents a background overview of required adaptation layer functionality along with associated implementation challenges. Optical packet switching (OPS) in conjunction with all-optical data routers (ODRs) is currently being explored as a data routing technology capable of meeting the increasing power-bandwidth demand projected for future backbone networks. Adaptation layers will be required at the perimeter of such optical networks to allow intercommunication between electronic legacy networks located at the edges. An adaptation layer at the ingress of the OPS network converts packets in a legacy format into one that utilizes labeled headers and high-speed payloads. Alternatively, the egress adaptation performs the opposite operation of converting labeled optical packets back into an electrical legacy format. A set of implementation requirements are outlined to be utilized as design guidelines for successfully demonstrating end-to-end packet adaptation. Some areas that are addressed include, but are not limited to format transparency, transmission capacity, dynamic operation, traffic engineering, latency performance, and relevant performance metrics. The work in this dissertation uses a previously demonstrated 10Gb/s burst mode FPGA-based

transceiver as foundational technology. The FPGA logic associated with data generation and detection is expanded to perform dynamic packet adaptation, while a hierarchical (de-)serialization configuration is utilized to scale the operational data rate beyond 40Gb/s. The challenges associated with employing (de-)serialization hierarchies are then separately discussed with respect to the transmitter and receiver.

5.1   Principle of Operation
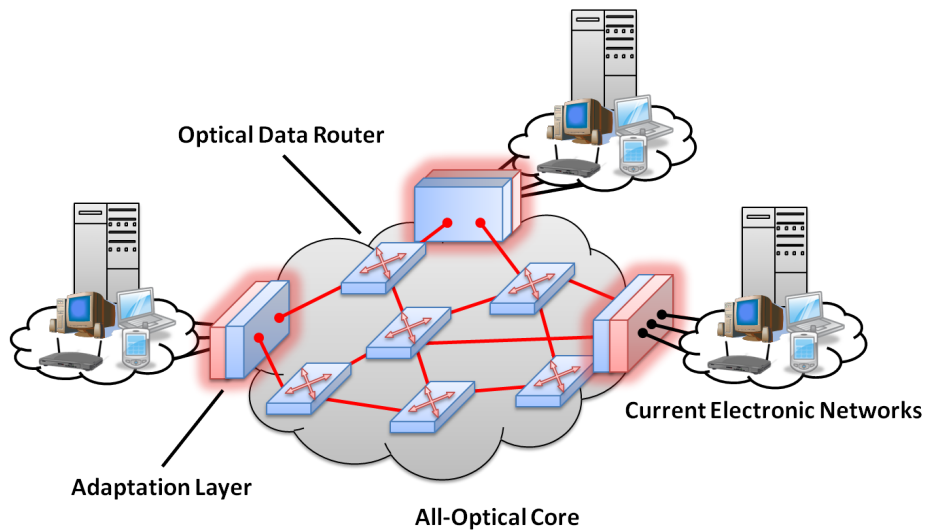
Figure 5.1: *Adaptation Layers (highlighted) are placed at the edge of optical core networks to enable interoperability between current electronic networks and future all-optical networks*

Figure 5.1 shows a network topology where legacy electronic networks are placed at the edge of an all-optical core network comprised of interconnected optical data routers (ODRs). Communication between non-adjacent electronic networks is established by forwarding packet data through the optical core routers. Adaptation layers are inserted at the interfacing edges to

enable interoperability between legacy and future all-optical core networks. The adaptation layer at the ingress of the optical core converts incoming frames into optical packets employing the label-swapping format used by the optical routers. Alternatively, the adaptation layer at the egress of the optical core performs the inverse operation of translating labeled packet formats into a conforming electronic format.
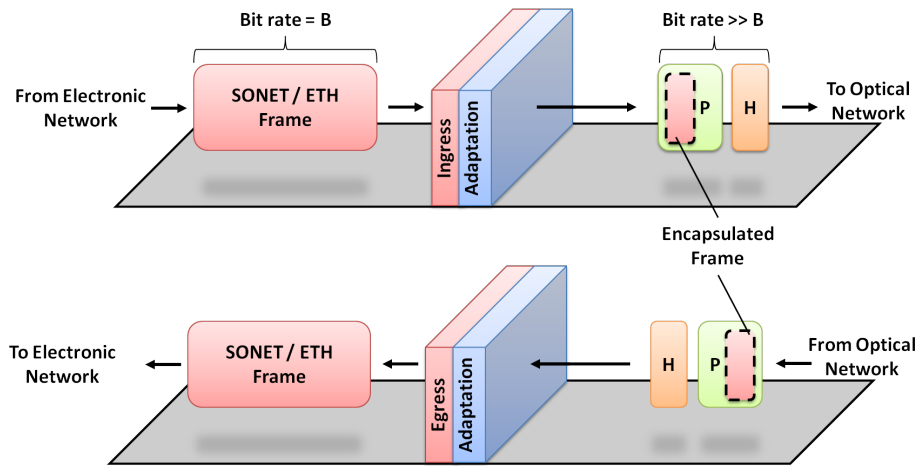
Figure 5.2: *Ingress Adaptation Layer (top) converts frames in current network formats to a labeled optical format where the frame is encapsulated into a high-bit rate payload (P) preceded by a labeled header (H). The Egress Adaptation Layer (bottom) adapts labeled optical packets into frames that conform to current network standards.*

The figure in Figure 5.2 shows the operating principle for the Ingress (top) and Egress (bottom) Adaptation Layers. Frames from legacy networks (either SONET or Ethernet) enter the Ingress Adaptation Layer where they are converted to a label-swapping optical packet format. The adaptation layer is designed to perform a temporal compression by up-converting the frame data rate to a significantly larger bit rate. This action allows traffic

from multiple legacy network interfaces to be aggregated onto a single, high-speed optical channel. Once compressed, an input frame is inserted into the high-speed payload (P) that is preceded by a relatively low-speed labeled header (H). The labeled header is generated via an electronic table lookup that maps a destination IP address to a corresponding optical label. The Egress Layer receives the labeled packets at its input and extracts the frames from the incoming high-speed payloads. A data rate down-conversion is then performed to allow the frames to be transmitted through the legacy network interface.

5.1.1 Implementation Requirements

5.1.1.1 Transparency

An adaptation layer must enable end-to-end inter-network communication between different topologies, data bandwidths and other network parameters [1]. The optical label swapping (OLS) scheme was conceived with the belief that substantial benefits may be gained by forwarding high-speed payloads transparently through all-optical switching fabrics regardless of data rate or protocol. Hence, edge adaptation layers must serve as universal access points capable of supplying end-to-end network-specific connectivity such as Ethernet, SONET, and ATM. Traffic from varying network formats can then be encapsulated and aggregated as common OPS payloads that are transparently forwarded throughout the OPS core network.

5.1.1.2 Capacity

Though access network data rates have evolved over time to provide much higher bandwidths via passive optical networks (PON), legacy data links such

as 100 megabit Ethernet (MbE) continue to be ubiquitous [2]. Therefore, a successful adaptation layer implementation will need to successfully enable interoperability between a high-speed (>40Gb/s) labeled optical packet format and a legacy network format such as 100MbE.

5.1.1.3   Dynamic Operation

It is crucial for the optical core routers work seamlessly with the edge adaptation layers in order to maximize end-to-end network performance. The Ingress Layer requires dynamic electronic lookup to generate labeled headers on a per-packet-basis. Moreover, on-the-fly configuration of electronic lookup tables is conducive to performance flexibility. Finally, real-time recovery and extraction of frames from high-speed payloads is needed to ensure near-optimal end-to-end recovery performance beyond the OPS core.

5.1.1.4   Traffic Engineering

End-to-end OPS performance may be improved by leveraging available electronic memory, within the Ingress Adaptation Layer, to employ traffic shaping. Typical Ingress Adaptation Layer implementations are electronic and are equipped with random access memory (RAM) banks that allow for successful format conversions before performing electro-optic conversions. For example, management and allocation of bandwidth resources along with enforcement of quality of service (QoS) and connection reliability can be carried out at the network edge. Additionally, packet queuing and scheduling may also be performed at the edge to reduce the amount of buffering required at optical router nodes. With that said, adaptation layer designs need to be wary of incurring excess latencies associated with the aforementioned traffic shaping functionalities.
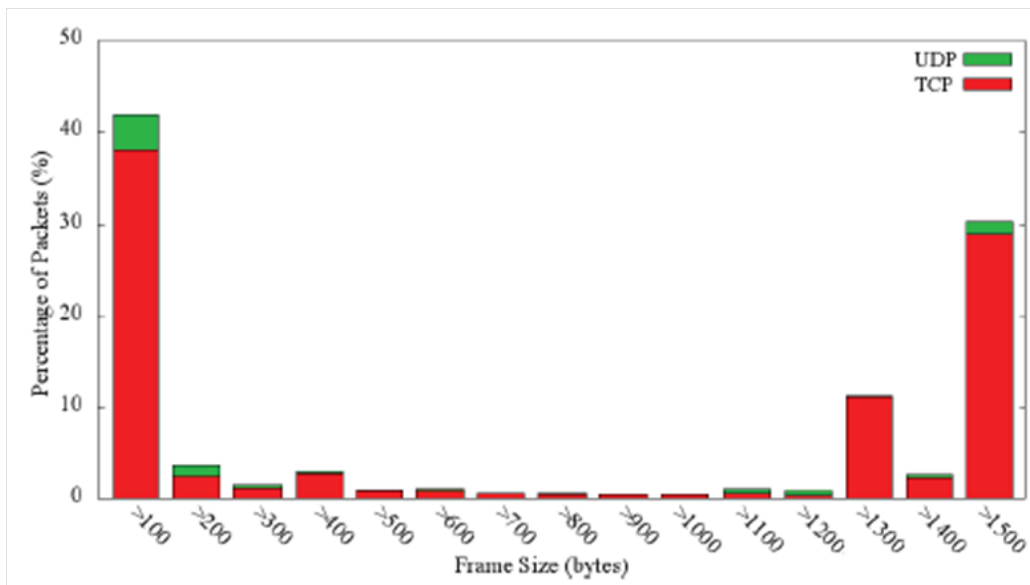
Figure 5.3: *Packet length variability observed in current networks [5–7].*

5.1.1.5 Fragmentation Latency

Packet adaptation must occur with high transparency and while minimizing the penalties associated with increased overhead and added latency. This requirement becomes difficult to accommodate when operating on packets that may vary in length. Figure 5.3 shows a distribution of packet lengths representative of current network traffic. Previous passive and active network traffic measurements have found the distribution of packet sizes to be trimodal with the dominant packet sizes being 40-100 (40%) bytes, 572 bytes (6%), and 1500 bytes (10%) [3, 4]. More recent measurements have found that network traffic may be shifting to a distribution that is more bimodal about 40-100 bytes (44%) and 1400-1500 bytes (37%) [5–7]. It is critical that the adaptation layers accommodate packet size distributions that are representative of current networks without performing additional fragmentation. Further fragmentation can debilitate system performance by inefficiently al-

locating resources to the fragmentation effort. Furthermore, the loss packet fragments may lead to performance degradation if re-assembly is not carried out in an efficient manner [8]. This requires OPS core nodes to be compatible with burst mode traffic consisting of varying packet lengths, which may be difficult to implement all-optically.

5.1.1.6 Performance Metrics

With OPS being well suited for IP-over-WDM deployment, one needs to consider metrics that evaluate performance beyond the OPS network. A successful realization of an optical core network must not only aim to minimize packet loss encountered at each router node, but should also be mindful of the performance observed from the vantage point of the end user. Consequently, any demonstration presented should evaluate end-to-end adaptation functionality via recovery rate at the frame level (Layer-III).

5.2 Adaptation Framework

The work presented in this dissertation makes use of the burst mode transmitter (TX) and receiver (RX) technology previously demonstrated in [9]. Figure 5.4 shows schematic representations of the custom, 10Gb/s FPGA-based TX/RX developed by Mack *et al*. A field-programmable gate array (FPGA) is utilized to generate a repetitious, reconfigurable packet stream consisting of a 32-bit idler, 64-bit identifier, and a repeating 2^7-1 pseudo-random bit sequence (PRBS) payload. Data out of the FPGA is forwarded to a 16:1 serializer (SER) as sixteen 625Mb/s streams. The SER is used to time-division multiplex the parallel data streams into a single non-return to zero (NRZ) 10Gb/s signal that modulates the output of a laser via a Mach-Zehnder
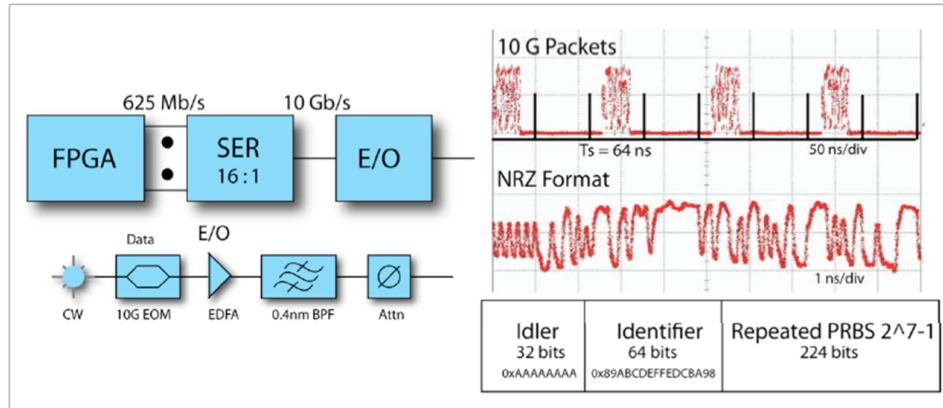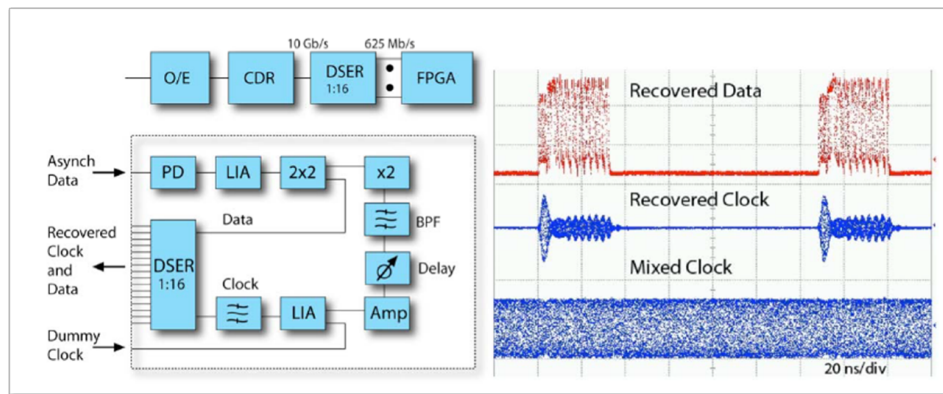
Figure 5.4: *(a) Schematic representation (left) and oscilloscope traces (right) of an asynchronous, 10Gb/s burst mode transmitter and (b) receiver [9].*

101

electro-optic modulator (MZM). The asynchronous, burst mode receiver in (b) utilizes a 10GHz photo-detector (PD) to perform an optical-to-electrical (OE) conversion on incoming optical packets. The electrical 10Gb/s signal is then sent to a clock and data recovery (CDR) stage that is used to extract a bit-aligned 10GHz clock bursts from the packet stream. The clock-data pair is then forwarded to a 1:16 de-serializer (DSER) that is used to parallelize the 10Gb/s stream into sixteen 625Mb/s data lines that are eventually forwarded to an FPGA-based packet analyzer. The analysis interprets the detection of the 64-bit identifier as successful recovery of a packet regardless of payload contents.

The previously discussed transceiver technology is used as a foundation for engineering high capacity adaptation layers to facilitate edge-to-edge interoperability between legacy and future core networks. A hierarchical serialization approach is needed to achieve high-capacity operation utilizing a 10Gb/s transceiver understructure. For example, one is able to achieve single data line operating at 40Gb/s by employing a twofold hierarchy consisting of four 16:1 SERs in parallel whose 10Gb/s outputs are multiplexed by a 4:1 SER. Furthermore, a single 40Gb/s data line can be parallelized into sixty-four 625Mb/s data lines utilizing a separate 2-stage DSER hierarchy comprised of a 1:4 DSER whose outputs are further de-multiplexed by four 1:16 DSERs configured in parallel. The 10Gb/s transceiver FPGA data processing capabilities can be easily scaled to 40Gb/s as long as enough high-speed physical ports exist to accommodate sixty-four 625Mb/s I/O ports. The FPGA logic used to implement the packet generation functionality within the transmitter can be expanded to achieve autonomous creation of labeled packets required of the Ingress Adaptation Layer. In a like manner, dynamic frame extraction performed in the Egress Adaptation Layer can be carried

102

out by extending the digital logic used to perform packet recovery within the previously discussed burst mode receiver.

5.3 Challenges of Edge Adaptation

Successful realization of low-latency, high-capacity interoperability requires one to address issues associated with the implementation of burst mode receivers utilizing hierarchical (de-)serialization schemes.

5.3.1 Hierarchical (De-)Serialization

Figure 5.5 contains diagrams illustrating the basic operating principle of (de-)serialization. The SER multiplexes four parallel data streams operating at a rate of $\frac{1}{T}$ into a single $\frac{1}{4T}$ data line, while the DSER performs the opposite function. Internally, the (de-)serializer essentially functions as a 4:1 (1:4) switch that iteratively establishes electrical connections between each of the input and output port(s). The timing diagrams show resulting (de-)serializer output grouped into time slots of duration T. The (D)SER time slot begins and terminates when an output circuit connection is established with ports 1 and 4 respectively. Each (de-)serializer is gated with a clocking signal whose phase is bit-aligned to the incoming data rate. The clock is either supplied directly from the synthesizer used to generate the data or through a CDR stage. Although it is relatively straightforward to provide a bit-aligned clock-data pair to (de-)serializers, it is challenging to ensure that the incoming data time slot is aligned to the time slot of the (de-)serializer.
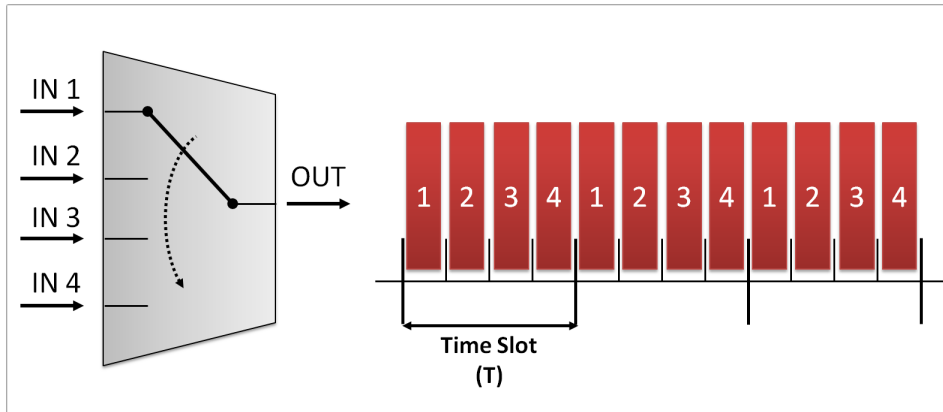
(a)
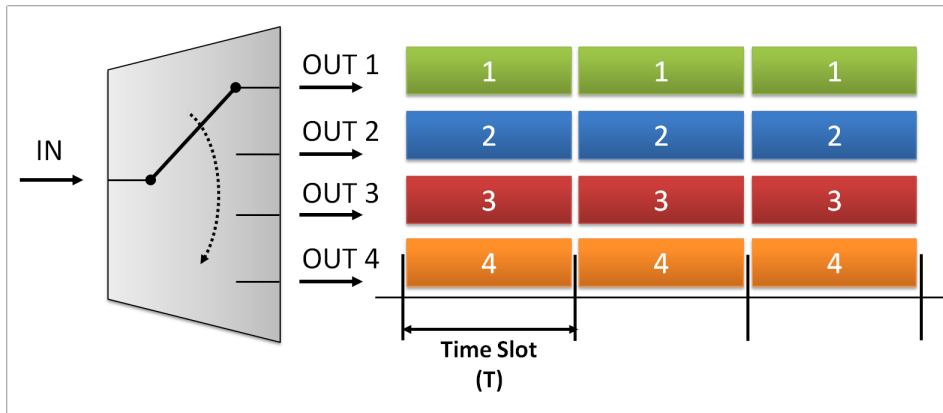
(b)

Figure 5.5: *(a) Diagrams describing internal functionality (left) and timing (right) of 4:1 serializer (b) and 1:4 de-serializer.*

Figure 5.6: *Timing diagram describing the un-synchronized operation of a 2-stage serializer hierarchy. The diagram shows SER input (left) where data time slots may not necessarily be aligned to N:1 SER time slots, resulting bit-level temporal skewing (middle) relative to time slot of 4:1 SER, and the serialized output of 4:1 SER (right) compared to output without skewing.*

5.3.1.1 Serialization

Figure 5.6 shows a timing diagram of an asynchronous 2-stage serializer hierarchy consisting of a 4:1 SER forming connections to four N:1 SERs configured in parallel. Four sets of N data streams (left) arrive, aligned to a data time slot (T_D), at the input of four N:1 SERs. Each data burst is overlaid

with a horizontal axis qualitatively depicting the time slot of each DSER. It is critical for the reader to note that each DSER is aligned at the bit level, while the time slots are not since each DSER is operated autonomously. In this example, the time slot of SER1 is perfectly aligned to the data time slot while SER2, SER3, and SER4 are arbitrarily designated as leading, lagging, and leading, respectively. The middle of the figure shows the output of the N:1 SERs and their temporal skew relative to the 4:1 SER time slot, which is overlaid horizontally. The right-most part of the figure qualitatively demonstrates the serialized out put resulting in a scrambled data stream. If a receiver expects an ideally-serialized data stream (inset), it may regard the scrambled data stream as illegible if no previous knowledge of skewing exists.

Some form of pre-skewing at the transmitter or skew compensation at the receiver is required to successfully recover data that has been scrambled by the serializer hierarchy. The amount of skew observed is deterministic, and remains unchanged as long the SERs are kept in continuous operation. Therefore, an approach utilizing pre-skewing is more practical than implementing skew compensating algorithms at the receiver. The latter approach incurs additional computational latency and power penalties since skew from different SER hierarchy configurations need to be accounted for. The work in this dissertation utilizes reconfigurable bit-delay registers as a means of pre-skewing, which is discussed in Section 6.4.2.

5.3.1.2   De-Serialization

The previous section addresses data skew within serialization hierarchies that results from the asynchronous operation of SERs configured in parallel. This section, however, shows that DSERs in both the lower and upper echelons of the DSER hierarchy affect the organization of outgoing data streams.
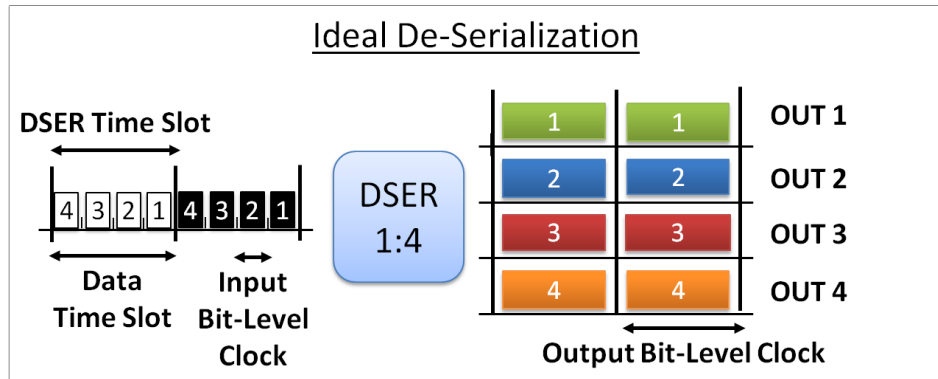
(a)
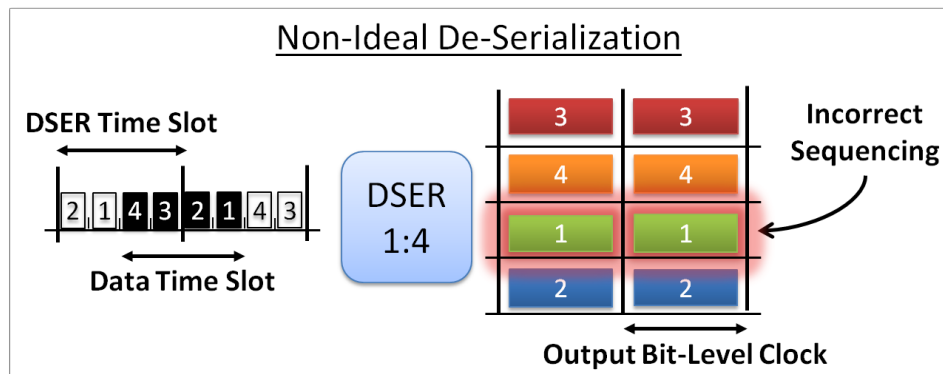
(b)

Figure 5.7: *(a) Timing diagrams showing de-serializer output when data and DSER time slots are synchronous (ideal) and (b) asynchronous (non-ideal).*

The illustration in Figure 5.7 shows the de-serialization process of a DSER located in the upper-echelon of the hierarchy where data and DSER time slots are synchronous (ideal) and asynchronous (non-ideal). The scenario illustrated in (a) consists of an incoming data stream of 4-bit nibbles whose time slot begins and terminates with bits 1 and 4, respectively. The DSER time slots are represented by overlaying a horizontal time axis with the data nibbles. When both time slots are synchronous, the data is evenly distributed between the DSER output ports where bits 1 and 4 exit through ports 1 and 4, respectively. Alternatively, (b) illustrates a scenario where the data and DSER time slots are asynchronous with respect to each other. The DSER output will be shifted and out of sequence depending on the amount of time difference between the two time slots. In this example, there is a 2-bit skew between time slots resulting in a DSER output where data bits 1 and 4 exit via ports 3 and 2, respectively. In practice, incoming packets will be transmitted from different sources with varying path delays, which may result in a diverse distribution of time slot skews. Therefore, one needs to account for the sequence order of streams in order to successfully recover data packets. Section 6.5.1 describes how this work performs sequence correction on asynchronously arriving payloads.

The timing diagram in Figure 5.8 shows an asynchronous 2-stage de-serializer hierarchy consisting of a 1:4 DSER followed by four 1:N DSERs configured in parallel. Four data bursts arrive, aligned to the data time slot (T_D), at the input of four 1:N DSERs. Each data burst is overlaid over a horizontal time axis corresponding to the time slot of each DSER. Each DSER is operated autonomously with a common gating clock source, resulting in bit-synchronous operation with misaligned time slots. In this example, the time slot of DSER1 is perfectly aligned to the data time slot
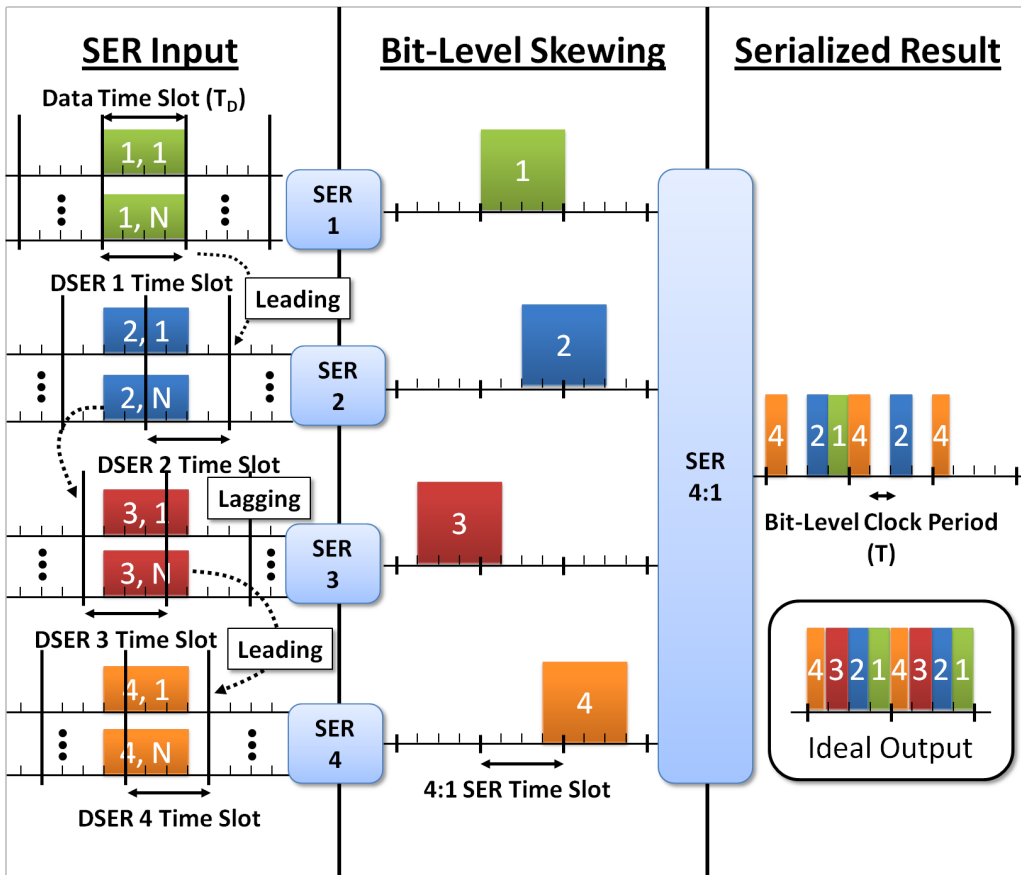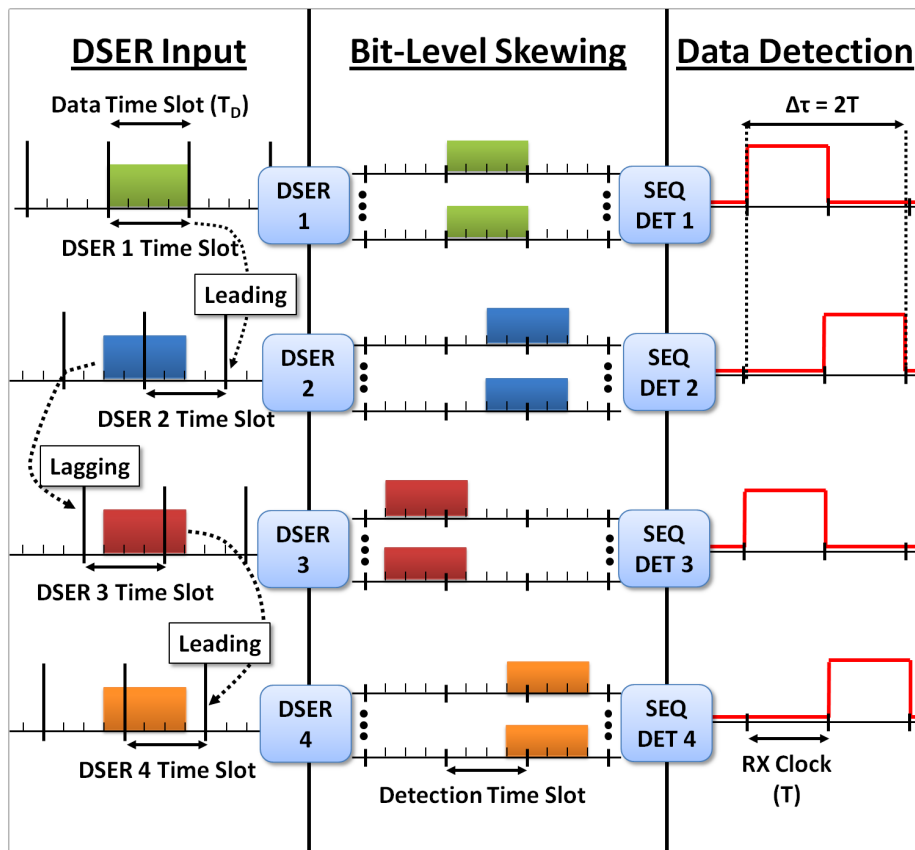
Figure 5.8: *Timing diagram describing the un-synchronized operation of a two-stage de-serializer hierarchy. The diagram shows DSER input (left) where data time slots are not aligned to the 1:N DSER time slots, bit-level temporal skewing (middle) relative to the 1:4 time slot, and digital signals that are enabled when a data burst is recovered.*

while DSER2, DSER3, and DSER4 are arbitrarily designated as leading, lagging, and leading, respectively. The middle of the figure shows the output of the 1:N DSERs and their temporal skew relative to the detection time slots of a receiver, which are horizontally overlaid. The right-most part of the figure shows the timing diagram of a digital signal that is enabled when the data burst is detected at by the receiver. The receiver detection circuit

109

is clocked with a signal whose timing is more coarse than the bit-level skew caused by the asynchronous DSERs. Hence, the data detection logic will need to also account for skew that may span two detection time slots ($\Delta\tau = 2\text{T}$). Section 6.5.1 describes how compensation of DSER skew is performed.

5.4 Chapter Summary

A discussion of the basic operating principles behind packet adaptation have been presented in detail. The framework technology used in this work is shown and the challenges associated with scaling its performance beyond 40Gb/s are addressed. This chapter has briefly discussed the requirements that need to be satisfied to realize a practical implementation of edge adaptation technology. Successful packet adaptation must accommodate a variety of electronic legacy formats along with a wide range of topologies, data bandwidths, and transport protocols. Payloads operating at data rates beyond 40Gb/s are essential to take advantage of the power consumption properties of an OPS data router that may potentially scale well with payload data rate. Dynamic de-framing of IP traffic and frame extraction from optical payloads within the Ingress and Egress Layers respectively is necessary for seamless, edge-to-edge communication exhibiting low packet loss. The amount of packet loss observed in OPS core nodes (due to contention) can be reduced by utilizing the volatile memory housed in the ingress edge to perform traffic engineering via bandwidth-scheduling algorithms, and enforcement of QoS requirements. The end-to-end adaptation process must also accommodate packets of varying lengths by avoiding packet fragmentation. This is widely viewed as a performance detriment since it inefficiently consumes system resources for the fragmentation and de-fragmentation processes. Additional

performance penalties may be incurred since the loss of a single data fragment could result in re-transmission of entire data sets. End-to-end data recovery measurements need to be performed at the frame level (Layer-III) to evaluate system-level performance from the perspective of the end user beyond the OPS core.

Previously shown 10Gb/s results are presented as the framework for the work in this dissertation and the issues associated with scaling such technology have been presented. The digital logic and (de-)serialization architecture is expanded to perform packet adaptation at data rates beyond 40Gb/s. An illustrative example of bit-level skewing that results from the asynchronous operation of a two-stage serialization hierarchy has been presented as an Ingress Layer implementation concern. Similar bit-skewing is observed at the Egress Layer, but the asynchronous configuration of the de-serialization hierarchy also results in parallel streams that are out of sequence or order. Potential solutions that mitigate these issues are introduced and discussed in detail within the following chapters.

References

[1] S. Yao, F. Xue, B. Mukherjee, S. J. B. Yoo, "Electrical Ingress Buffering and Traffic Aggregation for Optical Packet Switching and Their Effect on TCP-Level Performance in Optical Mesh Networks," *IEEE Communications Magazine*, vol. 40, no. 9, pp. 66-72, September 2002.

[2] A. Banerjee, Y. Park, F. Clarke, H. Song, S. Yang, G. Kramer, K. Kim, B. Mukherjee, "Wavelength-division-multiplexed passive optical network (WDM-PON) technologies for broadband access: a review [Invited]," *Journal of Optical Networking*, vol. 4, Issue 11, pp. 737-758, 2005.

[3] K. Thompson, G. Miller, R. Wilder, "Wide Area Internet Traffic Patterns and Characteristics", *IEEE Network*, vol. 11, pp. 10-23, November 1997.

[4] C. Fraleigh, S. Moon, B. Lyles, C. Cotton, M. Khan, D. Moll, R. Rockell, T. Seely, S. C. Diot, "Packet-level traffic measurements from the Sprint IP backbone," *IEEE Network*, vol. 17, no. 6, pp. 6-16, November 2003.

[5] K. Pentikousis, H. Badr, "Quantifying the Deployment of TCP Options - A Comparative Study," *IEEE Communications Letters*, vol. 8, no. 10, pp. 647-649, October 2004.

[6] W. John, S. Tafvelin, "Analysis of internet backbone traffic and header anomalies observed," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pp. 111-116, ACM, 2007.

[7] D. Murray, T. Koziniec, "The state of enterprise network traffic in 2012," *Asia-Pacific Communications Conference, (APCC), 2012*, pp. 179-184, October 2012.

[8] C. A. Kent, J. C. Mogul, "Fragmentation considered harmful," in *ACM SIGCOMM*, pp. 390401, 1987.

[9] J. P. Mack, Asynchronous Packet Routers, Ph.D. Dissertation, University of California, Santa Barbara, 2009.

Chapter 6

Adaptation Layer Implementation

In this chapter, the technology developed to enable end-to-end interoperability between legacy and future optical packet switching networks (OPS) is presented. An Ingress Adaptation Layer was designed to adapt present-day packet formats to the optical labeled format used by an optical label-swapping (OLS) core router, while the Egress Adaptation Layer performs the inverse function. The interoperability layers are implemented using custom FPGA-based designs with discrete, external 40Gb/s (de-)serialization stages. The Ingress Adaptation Layer extracts the destination Internet Protocol (IP) address from incoming 100 megabit Ethernet (MbE) frames and performs an electronic table lookup to generate a 10Gb/s labeled header. The data skewing caused by the external serialization stages is corrected via synchronization registers that perform several bit-wise shift operations to achieve a skew correction span of 192ns with a resolution of 100ps. Alternatively, the skewing caused by the de-serialization stages is corrected in the Egress Layer by utilizing Flip-Flop based configurable delays that demonstrate skew correction up

to 19.2ns with 6.4ns of resolution. Real-time sequence correction of 40Gb/s payloads is performed by means of burst mode electronic identifier recovery. Recovered payloads are then reconstructed and converted to conforming 100MbE frames. The main focus of this technology is on the interface between the optics and electronics that are relevant to a label-swapping edge router. As such, no traffic shaping or aggregation is performed.

6.1 Introduction

The technology developed here is based on the previous FPGA-based asynchronous receiver results obtained in [1]. An incoming stream of packets, consisting of a labeled payload format, was used as input to a 10Gb/s burst mode receiver. The stream was then forwarded to an external 16:1 de-serializer (DSER) and then to an FPGA-based 4:1 DSER to obtain a 64-bit data stream operating at a rate of 156.25Mb/s (6.4ns period). As mentioned in Section 5.3.1.2, the time slot of the incoming data may not be aligned with the DSER time slot, which results in a data recovery process that spans two receiver clock cycles (12.8ns). Hence, a sample window larger than the 64-bit stream was utilized to successfully detect the incoming packets.

The previously used payload detection scheme is shown in Figure 6.1, where it is assumed that data and DSER time slots are out of alignment. Time T = 0ns, displays an empty 128-bit detection register whose lower 64-bit word is filled from the output of the DSER stage. At time T = 6.4ns, a new 64-bit word is inserted into the upper 64 bits of the shift register. At this point in time, the burst mode receiver does not detect the payload since only a portion of incoming data is valid (dotted rectangle). At time T = 12.8ns, the previously inserted 64-bit word is bit-wise shifted towards
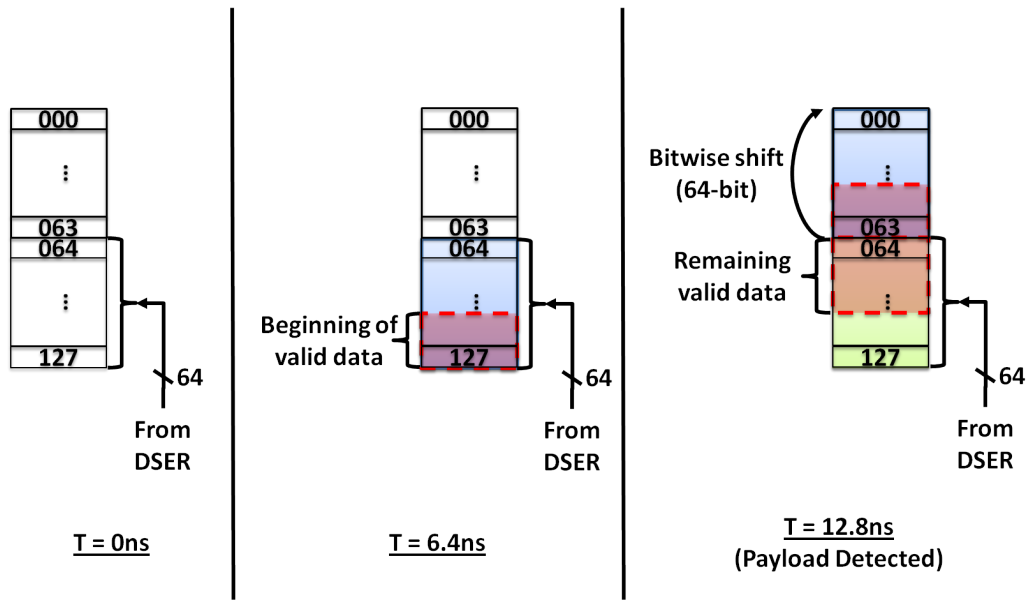
Figure 6.1: *Electronic detection scheme used to detect asynchronously arriving 10Gb/s payloads.*

the unused portion of the detection register and a new 64-bit word is written from the DSER output. The newly inserted word contains the remainder of the valid payload data (dotted rectangle), allowing the receiver FPGA to successfully detect the incoming payload by continuously scanning the 128-bit register. Therefore, the receiver FPGA was able to successfully detect incoming payloads with a timing uncertainty of two clock cycles (12.8ns).

The work described in this chapter, uses the previously discussed 10Gb/s electronic payload recovery scheme as a foundation and scales it to 40Gb/s by replicating the hardware and logic fourfold. A detailed description of the parallelized detection scheme is presented.
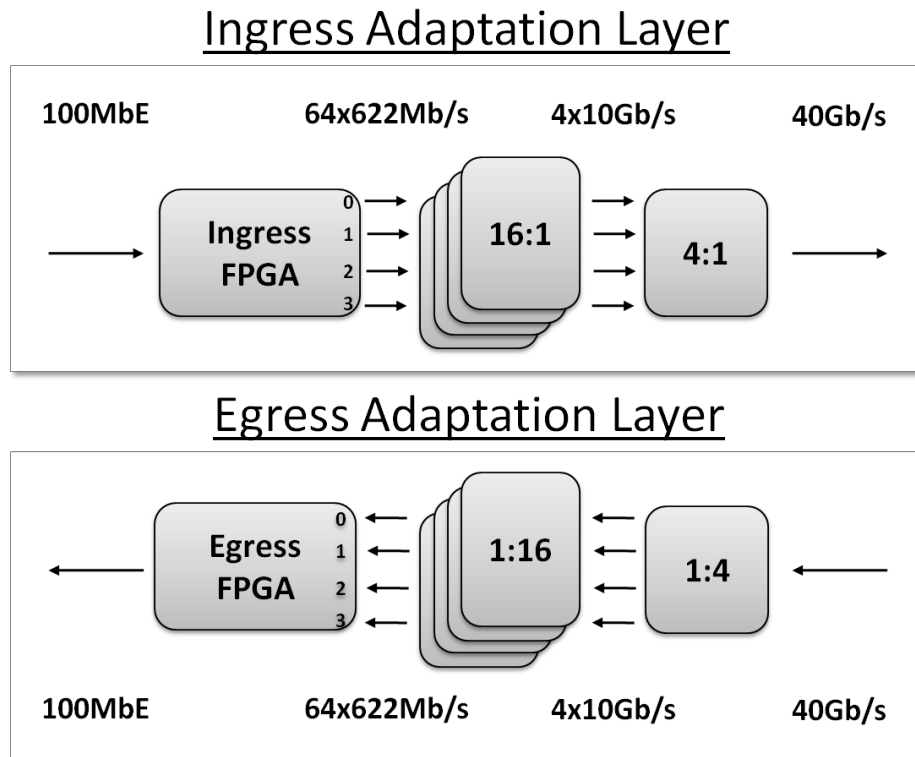
Figure 6.2: *Schematic representation of the Interoperability Adaptation Layers and external 40Gb/s (de-)serialization stages*

6.2 Adaptation Layers

Figure 6.2 shows a schematic representation of the Interoperability Adaptation Layers. Ethernet frames enter the Ingress FPGA through a legacy network interface and are then converted to a labeled packet format utilized by optical packet switching (OPS) core routers. The labeled packets are then forwarded to external serialization stages in order to up-convert the legacy frames to high-speed 40Gb/s labeled packets. Conversely, the Egress Adaptation Layer accepts the high-speed labeled packets and down-converts them to a data rate that is more easily managed by low-speed electronics. Output from the de-serialization stages is then forwarded to the Egress FPGA where

117

Optical header: 128 bits (12.8ns) 10Gb/s NRZ

Idler 32 bits 0xAAAAAAAA	Header Identifier 64 bits 0x89ABCDEFFEDCBA98	TTL 6 bit	Label 10 bits	Checksum 16 bits

Optical payload: 76-1512 bytes 40Gb/s RZ

Idler 32 bit 0xAAAAAAAA	Payload Identifier 64 bits 0x9BDFECA88ACEFDB9	Ethernet Frame 64 – 1500 bytes

(a)

Ethernet Frame Structure

Frame Header 14 bytes	IP Payload	Frame Check Sequence 32-bit CRC

IP Packet Structure

IP Header 96-bits	Source Address 32-bits	Destination Address 32-bits	Data

(b)

Figure 6.3: *(a) Labeled optical packet format: 10Gb/s optical header (top) and 40Gb/s optical payload (bottom) (b) Labeled optical packet format: 10Gb/s optical header (top) and 40Gb/s optical payload (bottom)*

Ethernet payloads are recovered and reconstructed.

6.3 Labeled Packet Format

The optical labeled packet format used by the OPS core router in this work is shown in Figure 6.3. It consists of a 128-bit low-speed 10Gb/s non-return-to-zero (NRZ) header and a high-speed 40Gb/s return-to-zero (RZ) payload. The Optical Header includes a 32-bit idler (0xAAAAAAAA), a unique 64-

bit sequence (0x89ABCDEFFEDCBA98), a 6-bit time-to-live (TTL) field, a 10-bit Label, and a 16-bit Header Checksum. The 32-bit idler is used for the purposes of initializing burst mode clock recovery and capacitive charging. When detected, the Optical Header Identifier allows the router's Electronic Channel Processor (ECP) to determine when a packet is present within 100ps of temporal accuracy. The TTL field is incremented at every router node and is used to collect statistics with respect to node hop count. The 10-bit label contains the forwarding information as it is used to determine a packet's desired output port and wavelength. Lastly, the 16-bit checksum is included in the header to detect any corruptions that might have occurred during transmission or header extraction. The checksum is calculated by maintaining a running sum of the header in 16-bit words similar to the checksum used in IP packets [1]. This allows the router to confidently perform arbitration. The Optical Payload consists of a 32-bit idler (0xAAAAAAAA), a unique 64-bit sequence (0x9BDFECA88ACEFDB9) and payload data comprised entirely of a 64- to 1500-byte Ethernet frame. The 32-bit idlers is again idler utilized for initialization of mode clock recovery and capacitive charging. When detected, the Payload Identifier allows the Egress Layer detect the presence of the Optical Payload within 25ps of accuracy. Furthermore, the identifier is used to successfully reconstruct Ethernet frames from data streams that may be out of sequence.

6.4 Ingress Adaptation Layer

The logical building blocks of the Ingress FPGA can be seen in Figure 6.4. Ethernet frames enter through the 100MbE network interface or are artificially generated via an internal frame generator. Frames from either the net-

119
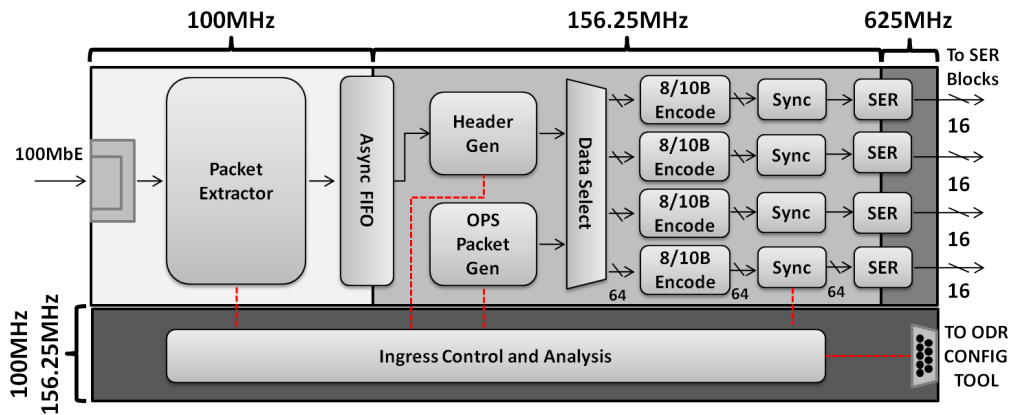
Figure 6.4: *Functional schematic of Ingress FPGA*

work interface are then written to an asynchronous first-in-first-out (aFIFO) memory block where they are transferred from the 100Mb/s clock domain to the internal clock frequency of the FPGA. Data out of the aFIFO is then forwarded to the Header Generation stage where 10Gb/s Optical Headers are generated based on the frame destination in order to perform the conversion to the OPS format. Labeled packets are then sent to the 8/10B Encoding and a Data Synchronization stags before they are forwarded to the serialization hierarchy.

6.4.1 Packet Extraction and Header Generation

Input Ethernet frames originate either externally from a 100MbE interface or from an internal OPS Packet Generator that used for debug and development purposes. The data from the network interface is recovered using a BCM5221 10/100 Ethernet PHY that is configured for 100Mb/s operation in duplex mode. The Ethernet frames arrive every 6.72μs and are available to the FPGA as four parallel 25Mb/s streams that are written to a 256-bit wide intermediary register. The 4-bit words are written to the register until each

bit of the IP Header is collected (208 bits). A checksum is derived from the IP header and is compared to the 2-byte Checksum field embedded in the header. The data received so far is discarded (along with the remaining data) if any corruptions are detected. Otherwise, the Datagram Length (L_D) field is used to confidently determine the size of IP payload, which in turn allows the size of the entire Ethernet frame (LOF) to be computed by adding the number of bits in the Frame Header and the Frame Check Sequence: ($LOF = L_D \times 8 + 14 \times 8 + 32$). Meanwhile, 4-bit words continue to be collected from the PHY until the intermediary register is filled. When filled, its contents are committed to the aFIFO consisting of 32kB of storage distributed across 64×4×128-byte memory blocks. Once an entire frame has been stored, a FRAME-READY pulse is enabled and forwarded to the Header Generation stage to notify it of the frame's availability.

Once the Header Generation stage detects the FRAME-READY pulse, it begins to read the contents of the aFIFO at a rate of 156.25MHz into a separate 256-bit intermediate register. It is assumed that the IP Header stored in memory contains valid data, so the frame length is once again derived from the L_D field to determine a stopping point for the memory reads. The destination IP address (DestIP) field is extracted from the IP Header and is then used to perform an electronic table lookup of label-to-IP mappings, which are listed in Table 6.1. The 32-bit Idler, 64-bit Header Identifier, and 6-bit Time to Live (TTL) fields are then generated along with a 16-bit Header Checksum computed from the aforementioned fields. Once the Optical Header is created, a 32-bit Idler and the unique 64-bit Payload Identifier are attached to the Ethernet frame to serve as the Optical Payload. The frames are converted to a labeled packet format with a configurable header-to-payload guard band of 12.8ns between the labeled header and payload.
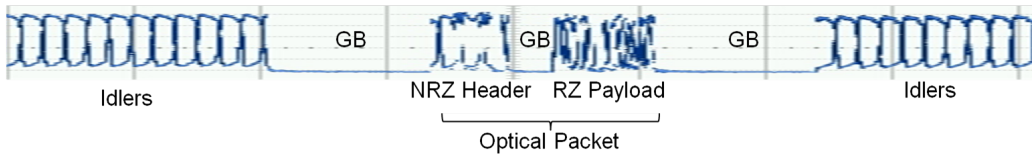
Figure 6.5: *Oscilloscope trace showing optical packet with NRZ header, RZ payload, guard bands (GB), and idlers.*

The data stream consisting of external 100MbE frames has a duty cycle of about 90% (64-byte frames), and a duty cycle of 0.63% when converted to the 40Gb/s labeled OPS format. Therefore, a 10Gb/s repeating idler pattern of alternating ones and zeros is inserted between packets to achieve DC balance and avoid transients within the optical and electronic components in the link. An oscilloscope trace of the resulting labeled optical packet is shown in Figure 6.5.

The internal OPS Packet Generator stage is used to emulate the functionality provided by the PHY-to-Header Gen data path. However, the packet generator has the capability of creating packets at a variable rate that may be different from what is enforced by the network interface. The Packet Generator is capable of generating a burst or a continuous stream of OPS packets consisting of pre-engineered header and payloads with configurable

Table 6.1: *Lookup table containing the mapping between destination IP addresses and Optical Labels.*

Destination IP Address	Optical Label	Destination IP Address	Optical Label
192.168.168.100	1100101011	192.168.168.104	1000101111
192.168.168.101	1011011010	192.168.168.105	0011100111
192.168.168.102	1010111011	192.168.168.106	0011001111
192.168.168.103	1110100011	192.168.168.107	1000001111

inter-packet gaps (IPG). The purpose of this functionality is to have the freedom to test the adaptation process at varying amounts of link utilization without being constrained by the requirements of the network interface.

A logical Data-Select stage is used to choose between the output of the Header Generator or the OPS Packet Generator to provide data for the remainder of the adaptation process. Data from the selector consists of four 64-bit wide parallel streams that are sent to an 8/10B Encoder stage to ensure that the reserved 64-bit identifiers (0x89ABCDEFFEDCBA98 and 0x9BDFECA88ACEFDB9) do not appear in the Ethernet frame. Once encoded, each data streams is sent to a Synchronization stage and then to a $4 \times 16:1$ serialization (SER) stage where an up-conversion to 64×625Mb/s is performed.

6.4.2 Data Synchronization

As previously mentioned in Section 5.3.1.1, bit-level skewing is an issue that arises when utilizing a SER hierarchy with multiple, independent serializers. Once the payloads are 8/10B encoded, they are passed through a synchronization stage that applies pre-skewing mitigate this issue. Figure 6.6 shows a functional diagram of the synchronization stage at different moments of operation. At T = 0ns, an empty 256-bit register is shown where the 64 least significant bits (boldly highlighted) are used as inputs to the SER hierarchy. At T = 6.4ns, the 64-bit output from 8/10B stage is inserted into the synchronization register at an offset j $(0 \leq j \leq 192)$, while dummy data is forwarded to the SER stage. At T = 12.8ns, the previously inserted 64-bit word is bit-wise shifted by 64 bits and a new word is inserted from the 8/10B stage at the j offset. One can see at this juncture, the upper portion of the valid 64-bit word has been shifted into the output section of
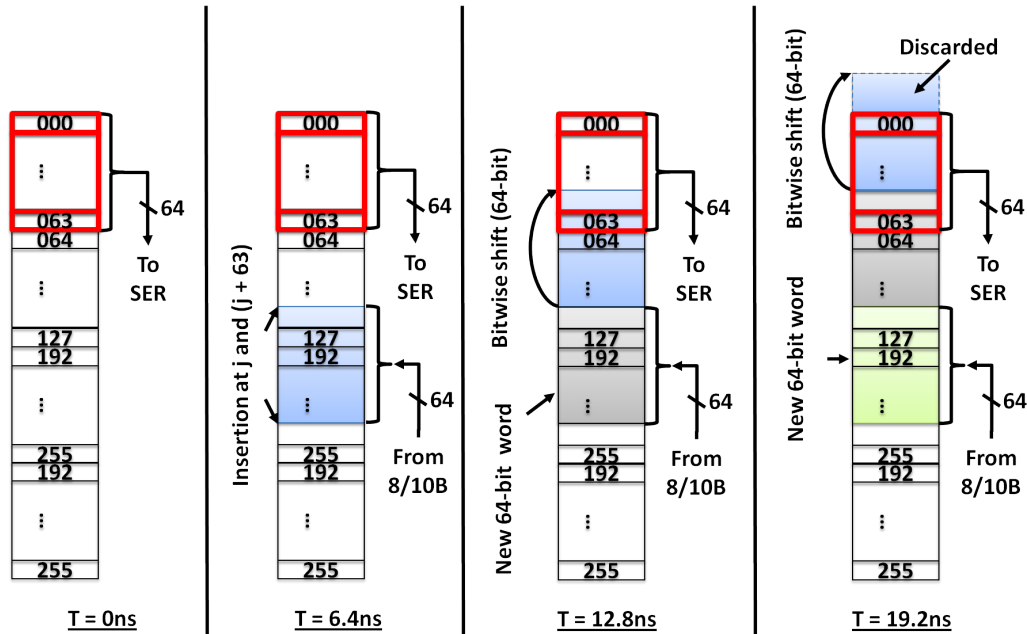
Figure 6.6: *Synchronization stage used to correct skew in serialization output. Labeled packets are inserted into a 128-bit wide shift register at an offset j. Register is shifted by 64 bits each clock cycle to allow for bit-level tuning via configurable parameter j.*

the synchronization register. At T = 19.2ns, the 64-bit words in the register continue to be shifted by 64-bits while a new word is again inserted at the j offset. Data shifted beyond the physical limits of the register are shown to be discarded. Utilizing a synchronization stage such as this allows for bit-wise skew correction over a 19.2ns span with 100ps of resolution by varying the insertion offset j. Furthermore, this parameter is held constant once pre-skewing is complete since skew in the SER hierarchy is constant. Once the serialization skew has been pre-compensated, the output from the four Ingress Synchronization stages is sent to the external SER hierarchy to obtain 40Gb/s labeled packets.

Special care needs to be taken to insure that the 10Gb/s header and
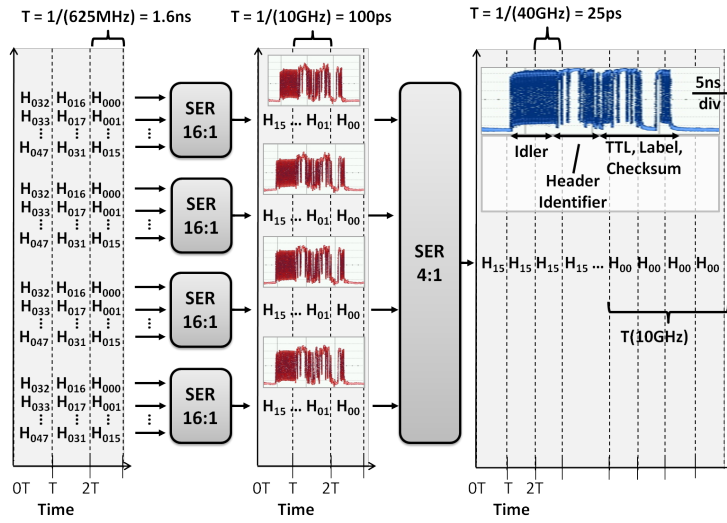
(a)
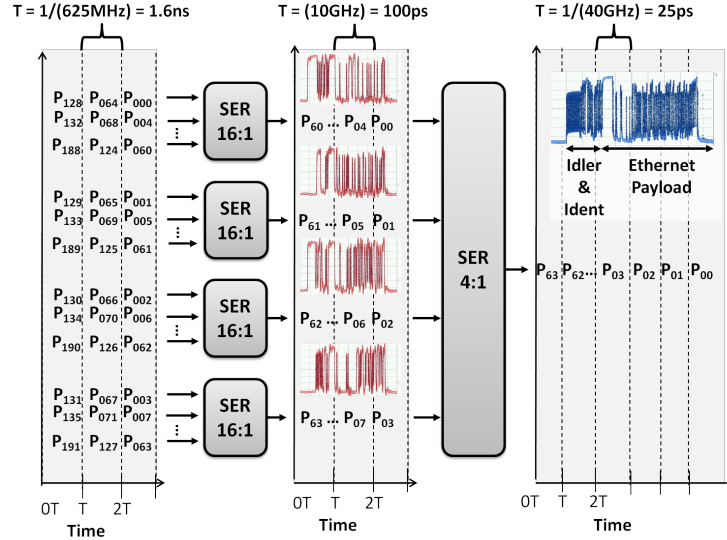
(b)

Figure 6.7: *(a) Serialization scheme used to generate 10Gb/s labeled Optical Headers with resulting oscilloscope traces (5ns/div). (b) Serialization scheme used to generate 40Gb/s Optical Payloads with resulting oscilloscope traces (2ns/div)*

40Gb/s payloads were correctly generated when using the external serialization hierarchy. Figure 6.7 shows timing diagrams of the serialization schemes used to generate the 10Gb/s Optical Headers and the 40Gb/s Optical Payloads. The timing diagram in (a) shows how data was forwarded to the SER hierarchy to generate the 10Gb/s Optical Headers. Each bit of the Optical Header (with the least significant bit denoted as H_{000}) is sent to the external 16:1 serializers 16 bits at a time. The input to all four SER stages is made identical to ensure a 10Gb/s signal is obtained from the multiplexed output of the hierarchy. To obtain a 40Gb/s Optical Payload, each bit of the payload (least significant bit denoted as P_{000}) was also forwarded to the external 16:1 serializers sixteen bits at a time. Here, each payload bit was evenly distributed between the serializer inputs to ensure that the payload remain identical to the original Ethernet with every bit transition occurring at 40Gb/s rather than 100Mb/s. An illustrative description of this serialization scheme can be seen in (b) along with resulting oscilloscope traces.

6.5   Egress Adaptation Layer

A functional block diagram of the Egress FPGA can be seen in Figure 6.8. Labeled 40Gb/s optical packets enter the FPGA, from the external deserialization (DSER) hierarchy, in four 16-bit 625Mb/s parallel streams. Then an additional down-conversion, via an internal DSER stage, is performed to obtain 256×156.25MHz parallel data lines operating at the FPGA clock rate. Data is then forwarded to the Payload Detection stage where electronic identifier recovery is performed to determine the sequence of the incoming stream. The Synchronization stage ensures that the streams are aligned in time before they are forwarded to the Payload Sequencing stage
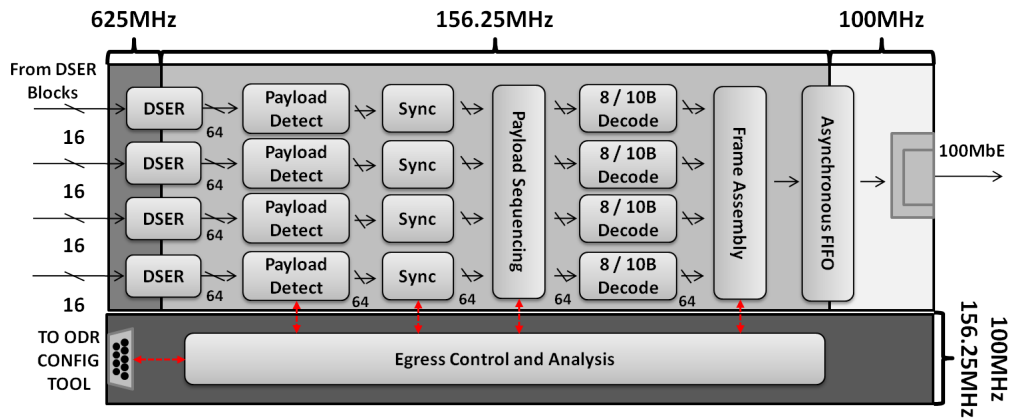
Figure 6.8: *Functional block diagram of Egress FPGA*

where they are re-sequenced. An 8/10B decoding is then performed before
the sequenced data is sent to the Frame Assembly stage where Ethernet
frames are recovered from incoming Optical Payloads. The reconstructed
data is then written to an aFIFO that is used to transfer the frames from
the FPGA local clock rate to the 100Mb/s data rate utilized by the Ethernet
network interface.

6.5.1 Payload Recovery, Synchronization, and Sequencing

The labeled packets first enter the Egress Adaptation layer through the ex-
ternal DSER stages where they are down-converted from a single 40Gb/s to
four sets of 16×625Mb/s parallel streams. An additional down-conversion is
performed, via FPGA-based 16:64 DSER stages, to transfer incoming data
to the local FPGA clock rate resulting in four sets of 64×156.25Mb/s paral-
lel data lines. At this point, the payloads are evenly distributed across four
parallel 64-bit streams that may be out of sequence due the fact that the
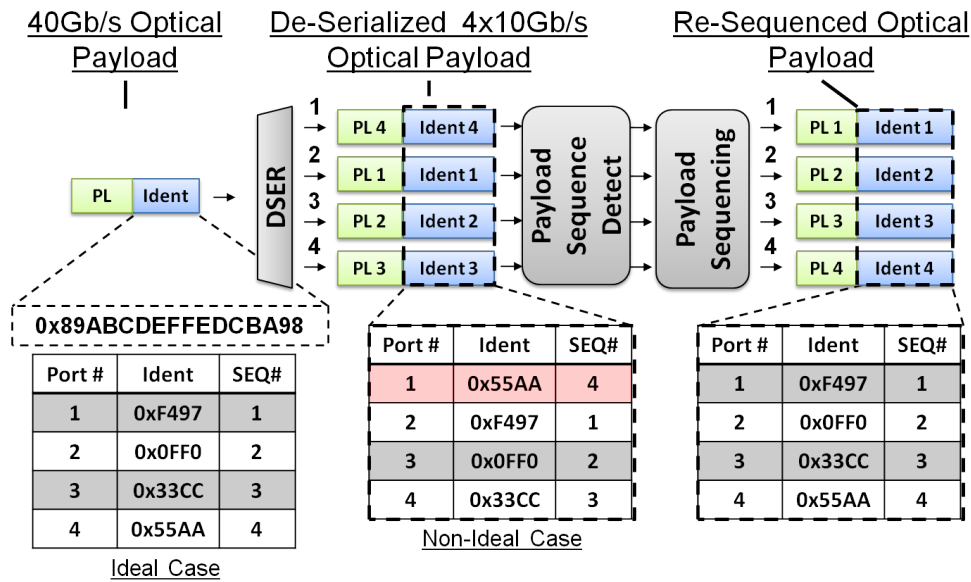
Figure 6.9: *Functional diagram of payloads sequencing method used to correct out-of-sequence payloads.*

external DSERs are operating asynchronously independent from each other, as discussed in Section 5.3.1.2. Each 64-bit stream is then forwarded to the Payload Sequencing stage where this issue is mitigated.

The sequencing method is performed by utilizing the unique 64-bit Payload Identifier that was attached to the Optical Payload at the Ingress FPGA. Figure 6.9 shows a flow chart of the method used to determine and correct payload sequencing. An incoming Optical Payload is shown, on the far left, consisting of an Ethernet payload (PL) and the 64-bit Payload Identifier (0x89ABCDEFFEDCBA98). The 40Gb/s payload is sent through the external 1:4 DSER stage and exits as a 4×10Gb/s signal. At this point, ideal de-serialization would have resulted in the payload being evenly distributed between each of the DSER ports with the first and fourth payload bits exiting from DSER ports 1 and 4, respectively. Had this occurred, the 64-bit

128

Payload Identifier would be de-multiplexed into the four 16-bit identifiers listed in the bottom-left table. The 16-bit identifier 0xF497 (exiting through PORT1) corresponds to the least significant bits (0, 4, 8, ..., 60), while 0x55AA (exiting from PORT4) corresponds to the most significant bits (3, 7, 10, ..., 63) of the 64-bit Payload Identifier. Since the payload was 8/10B encoded to ensure the Payload Identifier remained unique, one can be assured that the four 16-bit identifiers are also unique. Hence, they can be uniquely referred to as SEQ1 (0xF497) through SEQ4 (0x55AA) for simplicity. The reader will note that SEQ1 should exit from PORT1 in the ideal case, but actually exits from PORT2 in the non-ideal case (middle table) resulting in out-of-sequence streams. To obtain correct sequencing, the out-of-order streams are sent to a stage that continually scans the 4×10Gb/s data lines searching for the unique 16-bit identifiers (SEQ1-SEQ4). When an identifier is detected, a trigger pulse is enabled and forwarded, along with the data stream and its detected sequence number, to a separate stage responsible for sequence correction.

Output streams from the external DSER stages may not only be out of sequence, but they may also be susceptible to a bit-level skew resulting from the asynchronous operation of the 1:16 DSER stages. As discussed in Section 5.3.1.2, the maximum amount of skew that one would observe is equal to two clock cycles (12.8ns). Therefore, the Payload Synchronization stage in Figure 6.10 is implemented to correct the alignment skew caused by the asynchronous DSER hierarchy. The relative misalignment of each stream is determined by analyzing the amount of clock cycles between the trigger signals generated by the Sequence Detection blocks. The streams are then sent through a D Flip-Flop (DFF) configurable delay where input streams can be set to one of three time delays: 0.0ns, 6.4ns, or 12.8ns. The
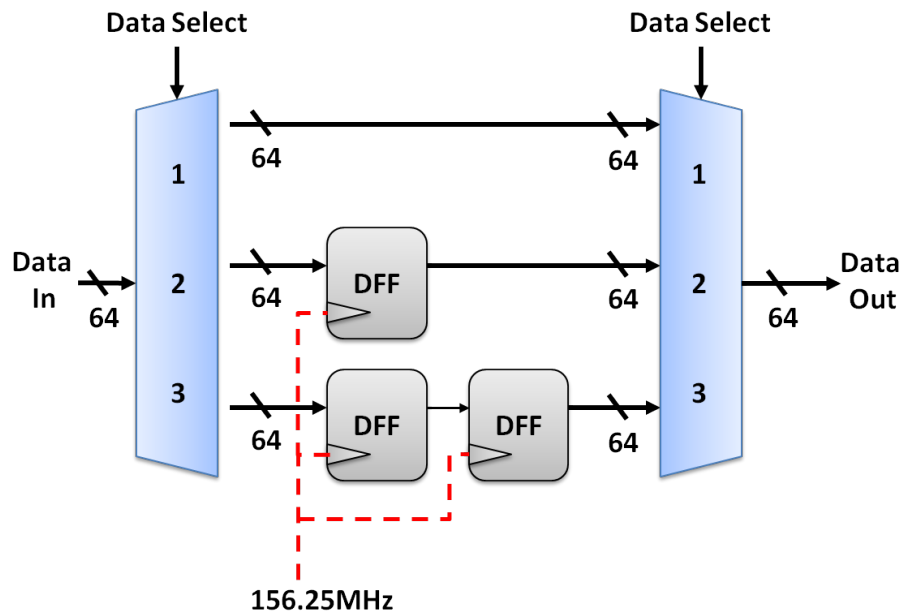
Figure 6.10: *Egress payload synchronization stage used to correct the skew caused by operating the external de-serialization stages independently.*

synchronized streams and trigger signals are then forwarded to the Payload Sequencer stage to perform the sequence correction operation.

For a payload to be considered detected, the Payload Sequencer requires that all four trigger signals originating from the Payload Detection stages be enabled simultaneously. Otherwise, the payload is considered corrupt and is discarded. When a payload is successfully detected, the sequences are analyzed and corrected if necessary. The synchronized, correctly-sequenced payload streams are then forwarded to an 8/10B Decoder stage and then to the Frame Assembly logical block where frames are extracted and reconstructed.

6.5.2 Frame Assembly and Extraction

The Frame Assembler strips the 10Gb/s Optical Header and begins to restore the Ethernet frame back to its original state. Data is gated out of the Sequencer stage and into an intermediate 256-bit register where the frames begin to be reconstructed. As reconstruction occurs, several fields are extracted from the IP header before it is written to the aFIFO memory bank. As in the Ingress Layer, the aFIFO accounts for 32kB of volatile storage distributed across $64 \times 4 \times 128$-byte memory blocks. First, a checksum is derived from the IP header and is then compared to the checksum embedded within the header. If corruption is detected, the payload is dropped and frame assembly halts. Otherwise, the length of the Ethernet frame is then extracted and used to determine how many words need to be committed into the aFIFO. Frames are then written, in 256-bit word increments, to memory at the FPGA clock frequency of 156.25MHz. Once an entire frame has been written to memory, a FRAME-READY pulse is generated and forwarded to the Frame Extraction stage responsible for transmitting the successfully reconstructed frames through the network interface. Once the Frame Extraction stage detects the pulse, it begins to read from the aFIFO a rate of 25MHz to accommodate the network interface. The duration of the memory-read is determined by computing the frame length (LOF) via the IP header contents. At this point, no further validation of the IP header is performed since it is assumed that it was not corrupted by the Frame Assembly stage. Data is read into a separate 256-bit intermediary register and then forwarded to the Ethernet PHY 4-bits at a time. When the contents of the intermediate register have been expunged, a new 256-bit word is read from memory and data continues to be forwarded to through the network interface until the entire frame has been transmitted.

6.6 Chapter Summary

In this chapter, a detailed implementation description of scalable, FPGA-based Adaptation Layers has been presented. This technology was designed to enable interoperability between legacy and future 40Gb/s labeled optical packet switching networks. The Ingress Adaptation Layer accepts frames through a 100MbE network interface and proceeds to convert them to a labeled packet format before utilizing an external serialization hierarchy to achieve a single 40Gb/s data stream. The packet format consists of a low-speed 10Gb/s labeled optical header followed by a high-speed 40Gb/s optical payload. The frame's Destination IP address is extracted and used to dynamically create the low-speed headers via a label-to-IP electronic table lookup, while the frame was inserted into the high-speed payload. A data serialization scheme that enabled correct generation of low-speed headers followed by high-speed payloads using a single external serialization hierarchy has also been presented. Additionally, a series of synchronization registers are utilized to correct the timing skew caused by operating the external hierarchy in an independently asynchronous fashion. A bit-wise skew correction over a span of 19.2ns has been successfully achieved with 100ps of resolution.

The Egress Adaptation Layer accepts 40Gb/s labeled packets through an external de-serialization hierarchy and converts them to a 100MbE packet format. The asynchronous operation of the de-serializers resulted in parallel data streams that were out of sequence and not temporally aligned. An burst mode electronic identifier detection scheme is used to not only successfully recover payloads, but is also utilized to correct out-of-sequence data streams. Temporal alignment is achieved via a DFF-based configurable delay capable of real-time data synchronization over a span of 12.8ns with 6.4ns of resolu-

tion. Ethernet frames were then reconstructed from the correctly-sequenced, synchronous 40Gb/s payloads and forwarded through a 100MbE network interface.

References

[1] J. P. Mack, J. M. Garcia, H. N. Poulsen, E. F. Burmeister, B. Stamenic, G. Kurczveil, J. E. Bowers, D. J. Blumenthal, "End-to-End Asynchronous Optical Packet Transmission, Scheduling, and Buffering," *Optical Fiber Communications Conference, (OFC) 2009*, Paper OWA2, March 2009.

[2] "Internet Header Format," *Internet Protocol DARPA Internet program protocol specification* IETF, pp. 14. STD 5, RFC 791, September 1981.

Chapter 7

End-to-End Adaptation Demonstration

In this chapter, the characterization of the end-to-end adaptation is presented. Layer-II and III recovery along with latency and link utilization characterizations are performed by inserting an optical link between the adaptation layers. A commercial packet tester is used to transmit 64-byte frames to the Ingress Adaptation Layer, where they are received via a 100MbE network interface. The Ingress Layer then converts the frames into optical packets consisting of a 10Gb/s NRZ labeled header and 40Gb/s high-speed payload. Packets at the output of the optical link are transferred to the Egress Adaptation Layer which converts them into 100MbE frames. Layer-II recovery measurements are then carried out at every functional stage of the Egress FPGA. The Payload Recovery stages, responsible for detecting the sequence of incoming 40Gb/s payloads, recovered over 99.9999% of incoming 16-bit Payload Identifiers. Once payloads are time-aligned by the Payload Synchronizer stage, they are re-ordered within the Payload Sequencing stage. The Sequencer exhibits negligible loss of payloads at receiver power levels

135

larger than -30dBm while 5% additional loss is observed at power levels around -35dB. The Frame Assembly stage, where the IP header of sequenced frames are validated before being written to memory, achieves a frame loss clamped at 0.3% at received power levels above -30dBm. The Frame Extraction stage, which performs a separate validation of frames read from memory before transmitting them via network interface, shows a constant loss floor located near 0.6%. Layer-III frame recovery measurements, performed at the packet tester via 32-bit CRC validations, demonstrate an end-to-end adaptation frame loss less than 0.6% for receiver power levels larger than -30dBm. Power penalty measurements are carried out relative to the Layer-III performance evaluated at the packet tester. An overall power penalty of 3dB is exhibited by the Egress Adaptation Layer. An incremental penalty of 0.5dB is observed in the Sequencer, Assembler, and Extractor stages. A total end-to-end latency of 272.2ns is observed with memory accesses (102.8ns) being the major contributors. Link utilization simulations, empirically based on current implementation, reveal 40% additional packet loss when increasing link utilization from 0.63% to 1%. Simulations performed assuming 100% link utilization suggest that a memory increase from 32kB to tens of megabytes is required to achieve zero packet loss for bursts longer than 1,000,000 packets.

7.1   End-to-End Adaptation Results

Figure 7.1 shows a schematic representation of the experimental setup used to characterize the end-to-end performance of the Interoperability Adaptation Layers. A stream of 50 million 64-byte Ethernet frames at a 100Mb/s data rate with $0.96\mu s$ of inter-packet gap (IPG) is generated by a commercial packet tester (SmartBits 6000). The 100MbE frames are sent to the Ingress
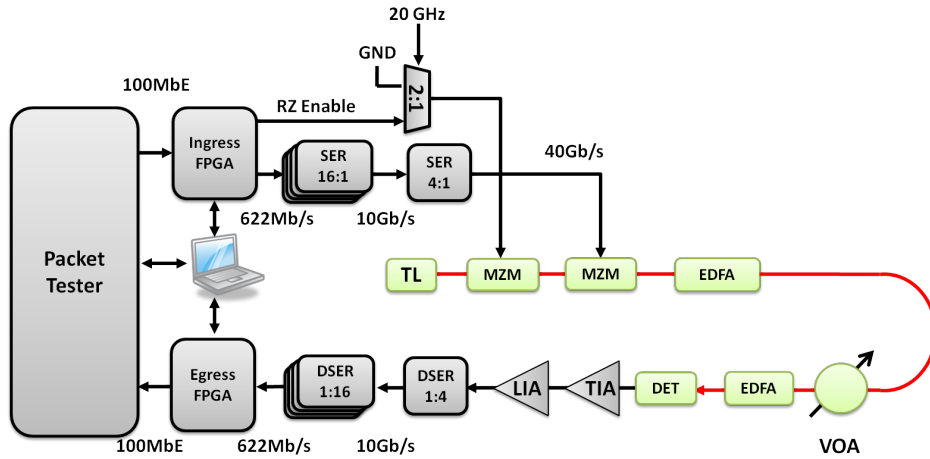
Figure 7.1: *Experimental setup of Interoperability demonstration.*

Adaptation Layer to be converted to a format that consists of a 40Gb/s RZ payload and a 10Gb/s NRZ labeled header. The guard band between the Optical Header and the Optical Payload is set to 12.8ns. The electrical signals from the external serialization (SER) hierarchy are used to drive an optical transmitter comprised of a tunable laser (TL), followed by two Mach-Zehnder modulators (MZM) connected in tandem, and an Erbium-doped fiber amplifier (EDFA). The alternating NRZ/RZ signals are generated by a 20Gb/s 2:1 multiplexer with an RZ-enable signal and a 50Ω-to-ground termination as its inputs. This creates a burst of 20GHz clock tones ($V_{PP} = 2V_{\pi}$) that are time-aligned with the 40Gb/s payloads in order to carve RZ the pulses. Oscilloscope traces of the Ingress Adaptation Layer output can be seen in Figure 7.2. The traces in (a) show a magnified view of the 40Gb/s Optical payload. The top traces correspond to the 10Gb/s output from each 16:1 SER stage, while the bottom trace shows the serialized 40Gb/s output from the 4:1 SER stage. The traces in (b) were taken by decreasing the magnification in order to obtain an overview of the 10Gb/s header and 40Gb/s
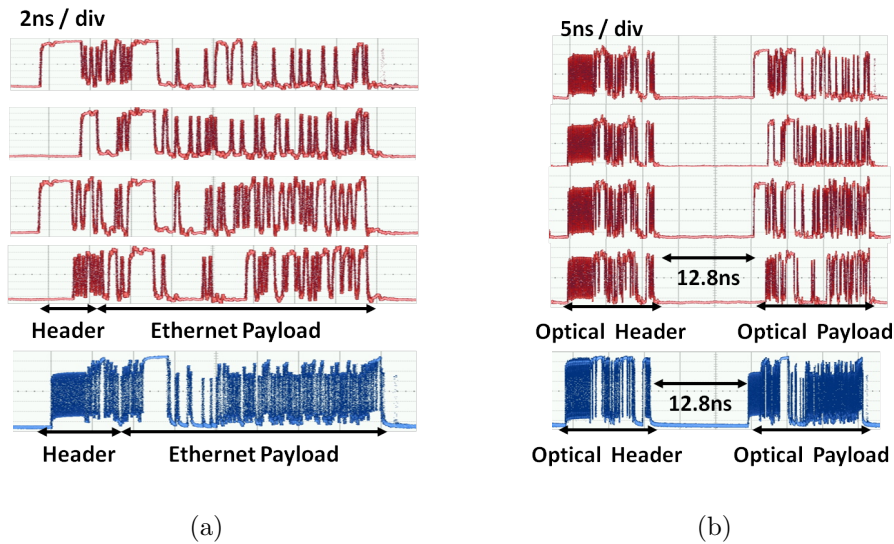
Figure 7.2: *(a) Parallel 4x10Gb/s payloads (top) and serialized 40Gb/s payloads (bottom). (b) Parallel optical header and payloads (top) and serialized 10Gb/s header and 40Gb/s payload (bottom).*

payload.

The optical packets are transmitted at an average power of -3dBm to a pre-amplified optical receiver that consists of a EDFA followed by a 50GHz photo-detector (DET), a trans-impedance amplifier (TIA), and a limiting amplifier (LIA). The receiver output is then directly connected to the Egress Adaptation Layer. A variable optical attenuator (VOA) is placed before the pre-amplified receiver to vary the optical signal to noise ratio (OSNR) of the signals entering the receiver. The Egress Layer then converts the 40Gb/s labeled packets back into Ethernet frames and sends them directly to the commercial packet tester where Layer-III recovery measurements are performed via 32-bit CRC validations. In addition to the Layer-III measurements performed by the packet tester, the Egress Layer performance is monitored at several validation points where one would expect packets to be dropped if corrupted.

138

7.1.1 Payload Recovery Stages Layer-II Results

Output from the external 40Gb/s DSER hierarchy enters the FPGA-based Egress Adaptation Layer as four 16×625Mb/s parallel streams. An additional DSER stage, internal to the FPGA, is then used to further parallelize the data into four parallel 64×156.25Mb/s streams. The data contained in the parallel streams may not be synchronized and may also be out of sequence, as discussed in Section 5.3.1.2. Each data stream is then forwarded to a Payload Recovery stage that is used to detect the sequence of incoming payloads using a unique de-serialized 16-bit payload identifier, as shown in Section 6.5.1. When a payload sequence is detected, a 6.4ns pulse trigger is then enabled and forwarded to the Payload Synchronization and Sequencing Stages. The trigger is also forwarded to the Ingress Control Logic for Layer-II statistical analyses.

The Layer-II results for the Payload Recovery stage are shown in Figure 7.3. The plot shows a successful payload detection percentage above 99% over a dynamic receiver power range larger than 27dB. The individual identifier detection performance varied within each stage with the third Payload Detection stage (ID3) showing the worst performance. The other stages exhibited payload loss percentages below 10^{-4} at certain points. There were also instances of zero loss, where one would have benefited from having longer datasets to obtain a more accurate performance assessment for the stages ID1, ID2, and ID4. Therefore, it is clear to see that the end-to-end performance will be governed by the amount of payloads dropped at the third Payload Recovery stage (ID3). This loss is mainly caused by timing uncertainties in the external DSER hierarchy, and non-deterministic propagation delays internal to the FPGA that varied each time the firmware was synthesized. Variability in propagation delays was minimized, but not eliminated,
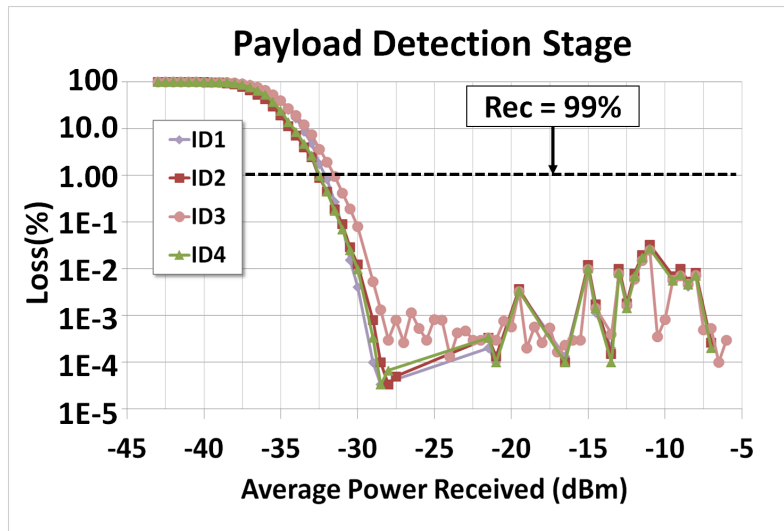
Figure 7.3: *Percentage of 16-bit Payload Identifiers lost at the Payload Recovery stages.*

by manually restricting the placement of FPGA logic.

7.1.2 Payload Sequencing Stage Layer-II Results

Payloads detected in the Payload Recovery stage are forwarded the Payload Synchronization stage that time-aligns the parallelized data streams. Once aligned, they are sent to the Payload Sequencer to correct the data streams that may be out of sequence. Sequencing occurs when four trigger pulses from Payload Recovery are simultaneously detected. A separate 6.4ns trigger is enabled to symbolize a successfully detected 40Gb/s payload. The trigger signal is then forwarded to the Assembler to begin frame assembly. Additionally, the trigger is sent to the Ingress Control Logic to perform Layer-II statistical analyses relevant to the Sequencing stage.

Figure 7.4 displays the results of the Layer-II measurements taken at the Payload Sequencing stage. The curve shows the percentage of 40Gb/s pay-
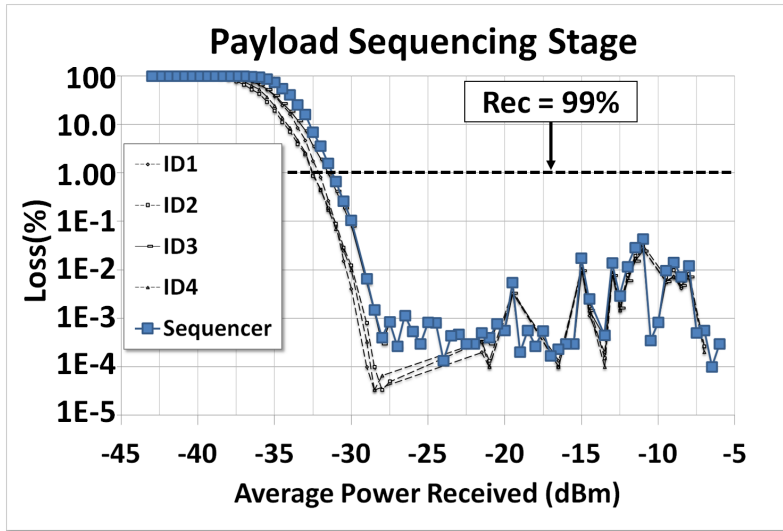
Figure 7.4: *Percentage of payloads dropped at the Payload Sequencing stage (bold trace).*

loads correctly recovered by the Sequencer (bold) overlaid with the identifier recovery from the Payload Detection stages (dotted). The results show that more than 99% of payloads were correctly detected and sequenced over a receiver power range greater than 27dB. A payload is interpreted as lost when at least one of the Payload Detection stages fails to successfully detect an incoming 16-bit identifier. Therefore, it is not unexpected that the recovery performance of the Sequencer is limited by the ID3 drop rate of the previous stage.

7.1.3 Frame Assembly Stage Layer-II Results

Once 40Gb/s payloads have been sequenced a trigger signal is enabled to notify the Frame Assembly stage of a successfully detected and sequenced payload. Once in the Assembler, payloads are converted into frames and undergo an additional stage of validation before being committed to memory.

141
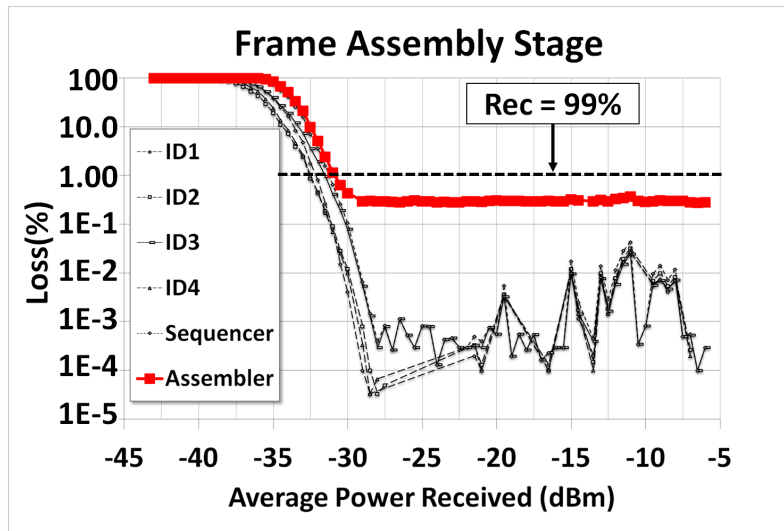
Figure 7.5: *Percentage of frames dropped at the Frame Assembly stage (bold trace).*

The number of bytes to write to memory is determined by extracting the IP Datagram Length field from the IP header. The header, however, is first validated by comparing the Header Checksum field against version derived from the incoming data. If it has not been corrupted, it is written to memory and a FRAME-READY trigger signal is forwarded to the Frame Extraction stage. The trigger is also sent to the Egress Control Logic where real-time Layer-II statistical analyses are performed.

Figure 7.5 shows the results obtained from the Layer-II measurements performed at the Frame Assembly (bold) stage, which are overlaid with the results from previous stages for reference (dotted). More than 99% of incoming frames are successfully validated, via IP Header validation, and written to memory. This is achieved across a 27dB span of received power. The percentage of dropped frames appears to be less than 1% for receiver power levels above -30dBm. However, there is an observed loss floor approximately

located at 0.3%. The loss floor is an evident result of the heightened valida-
tion requirements of the Assembler compared to previous stages.

7.1.4 Frame Extraction Stage Layer-II Results

The Frame Extraction stage begins to read frames from memory when it
detects the FRAME-READY trigger generated by the Assembler. An ad-
ditional IP header validation is performed to compute the duration of the
frame length used to determine when to stop reading from memory. If vali-
dation is successful, the frame is read from memory and transmitted via the
100MbE network interface. Otherwise, the frame is discarded and expunged
from memory. A trigger signal is enabled when transmission over the network
interface occurs to notify the Egress Control Logic of successful adaptation.
The trigger signal is then used to collect Layer-II statistics for the Frame
Extraction stage.

The Frame Extraction (bold) Layer-II results are overlaid with the mea-
surements from the previous stages (dotted), and are logarithmically ren-
dered in Figure 7.6. According to the plot, a frame drop rate less than 1%
(>99% recovery) is observed over a dynamic receiver power range greater
than 26dB. The Extractor performance also exhibits a loss floor around 0.6%.
The observed 0.3% increase in drop rate, relative to the previous stage, is
attributed to a timing aberration stemming from the asynchronous transfer
of frames between two asymmetric clock domains. A well documented mem-
ory irregularity in the FPGA-based aFIFO occurs when write operations are
performed at rates larger than read rates [2]. This irregularity prevents the
memory pointer from advancing, which in turn causes data overwrite. This
results in further data corruption when several memory blocks are utilized
in parallel, which is the case in this work.

143
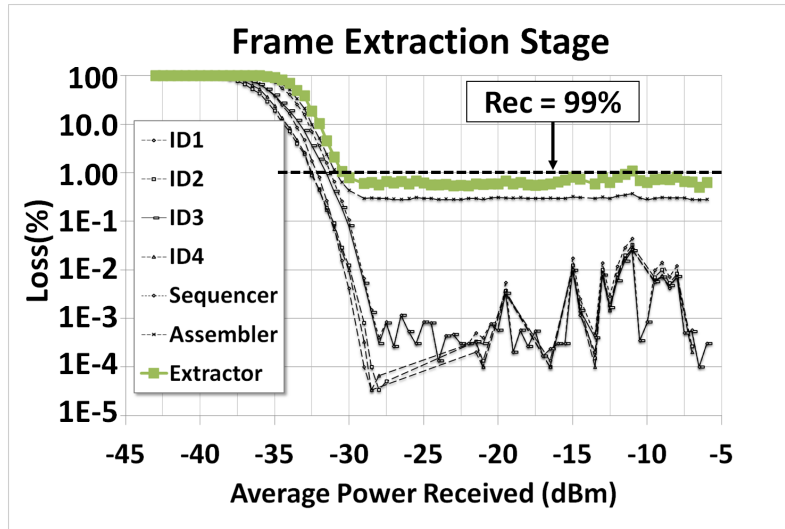
Figure 7.6: *Percentage of frames dropped at the Frame Extraction stage (bold trace).*

7.1.5   End-to-End Adaptation Layer-III Results

Ethernet frames transmitted through the 100MbE network interface are delivered to the SmartBits 6000 (SMB6000) commercial packet tester. The SMB6000 computes a running sum of incoming 32-bit words that will equate to an 8-byte CRC, which is then compared to the CRC sequence transported within the Ethernet frame. A successful recovery is contingent upon obtaining identical sequences. Alternatively, a CRC error is triggered when the sequences do not match.

Figure 7.7 shows the Layer-III frame recovery results measured by the SMB6000 (bold) overlaid with the Layer-II results measured by the Egress FPGA (dotted). The results demonstrate successful end-to-end adaptation for more than 99% of transmitted frames. These results are demonstrated over a dynamic receiver power larger than 25dB, where the observed sensitivity (loss $\leq$ 1%) is approximately -30dBm. A loss floor of about 0.3% is
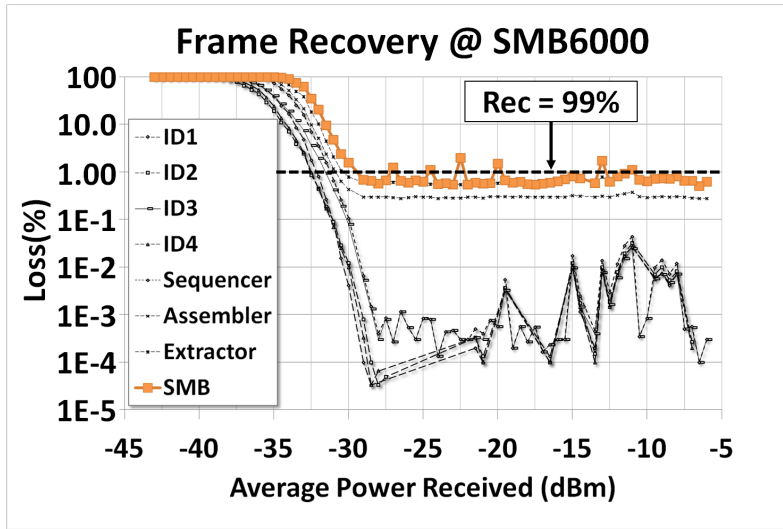
Figure 7.7: *Percentage of Ethernet frames lost during end-to-end transmission (bold trace). Results obtained using the SMB6000 commercial packet tester.*

observed during this operating regime, which is comparable to the Layer-II performance of the Frame Extraction stage. The loss floor is stems from the SMB6000's requirements for successful frame recovery, which are more stringent than the FPGA-based Layer-II measurements. The amount of validation performed at the Egress FPGA is less stringent in an effort to reduce adaptation latency.

Figure 7.8 provides a summary of the Egress Adaptation Layer performance used to estimate penalty incurred at each functional stage. Power penalty was measured at 70, 80, 90, and 99% recovery points, and evaluated relative to the SMB6000 (SMB) performance. The overall adaptation performance is observed to be within 3.5dB of the SMB performance. As previously stated, the adaptation performance at higher received powers (recovery > 99%) is governed by the third Payload Recovery stage (ID3).

145
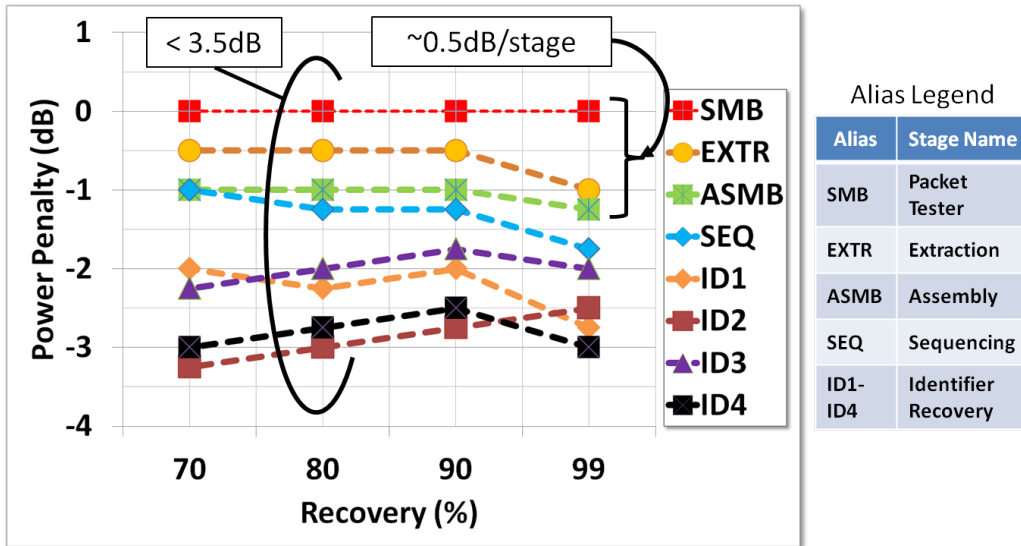
Figure 7.8: *Power penalty performance of each Egress Layer stage (SMB: SmartBits Recovery, EXTR: Frame Extraction, ASMB: Frame Assembly, SEQ: Payload Sequencing, ID1-ID4: Payload Recovery stages).*

This is demonstrated by the fact that ID3 and the Sequencer performance (SEQ) are within 0.5dB at high levels of receiver power. As received power decreases, the Payload Recovery stages (ID1-ID4) behave more erratic, and a divergence of SEQ from ID3 is observed. There is also an observed incremental power penalty of 0.5dB through the each stage after the Assembler (ASMB → EXTR → SMB), which can be attributed to the increasingly stringent validation requirements of each stage.

7.2 Latency & Link Utilization Performance

7.2.1 Latency Performance

The setup used to measure the latency performance and the packet loss rate of the end-to-end adaptation process is shown in Figure 7.9. The SMB6000
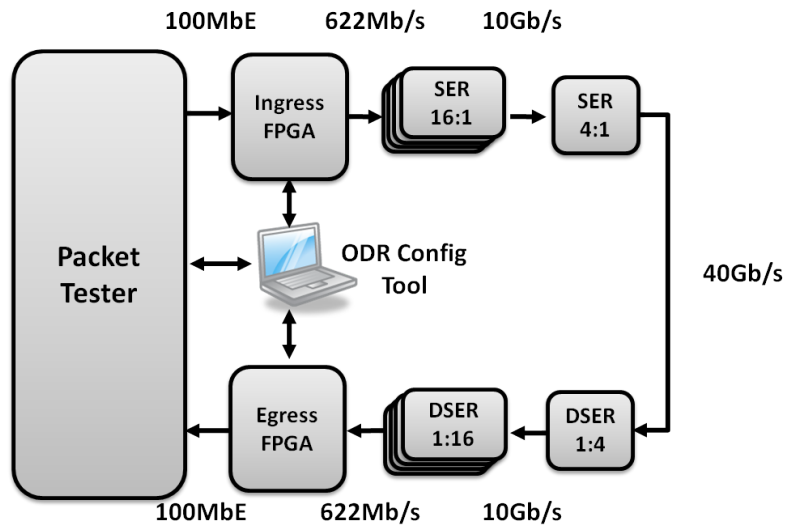
Figure 7.9: *Experimental setup used to measure the end-to-end packet loss rate and latency.*

commercial packet tester is used to generate a continuous stream of 64-byte 100Mb/s Ethernet frames that contain a sequence within the IP payload, which is utilized as a time stamp. The frames then enter the Ingress Adaptation Layer via the 100MbE network interface of the Ingress FPGA and are then converted to a labeled packets format. The labeled packet signals are then multiplexed to 40Gb/s via an external SER stage, which is directly connected to a 40GHz DSER hierarchy. Packets from the DSER hierarchy go through a series of payload detection, synchronization, and sequencing before frames can be extracted from incoming payloads. Extracted frames are then transmitted to the packet tester via a 100MbE interface where SMB then utilizes the time stamp sequence to derive the end-to-end adaptation latency. High-resolution latency measurements were also performed by evaluating the time delay between several trigger signals generated within the Ingress and Egress FPGAs via a sampling oscilloscope.

147

The packet tester reported a latency value of 6.72μs where 6.45μs are spent transmitting the 64-byte frame via the 100MbE network interface. Therefore, the resulting end-to-end adaptation process exhibited an excess latency of 272.2ns. The latency tributaries within the Adaptation Layers are broken down in Figure 7.10. The primary contributors of excess latency in the Ingress Layer are the Packet Extraction (EXTR) and the Synchronization stages. The EXTR stage receives reads an Ethernet frame from the network PHY without incurring any latency, but requires 80ns to commit two 256-bit words into memory at a rate of 25MHz. Alternatively, the Header Generation (HGEN) stage performs the same amount of memory reads, but at an expedited rate of 156.25MHz. The Synchronization stage requires, at most, 19.2ns in order to perform three register shifts to achieve data synchronization, as discussed in Section 6.4.2. The serialization (SER) hierarchy used to achieve 40Gb/s signaling incurs the least amount of latency with 8.10ns.

The major latency contributor in the Egress Layer (b) is the Frame Extraction stage (EXTR), which requires 80ns to read two 256-bit words from memory before transmitting them through the Ethernet interface. Conversely, the Frame Assembly (ASMB) stage only requires 12.8ns to write the two words into memory at a rate of 156.25MHz. The second largest tributary is the Payload Synchronization (SYNC) stage which latches (6.4ns) packets into a configurable delay that consists of up to 12.8ns of DFF-based delay. The Payload Recovery stages (ID) require two 156.25MHz clock cycles (12.8ns) to successfully detect a payload identifier from a de-serialized stream, as discussed in Section 5.3.1.2. Once payload sequences have been determined, the Sequencer (SEQ) stage only requires 6.4ns to correct any out-of-sequence data streams. Finally, the de-serializer hierarchy requires 8.10ns to parallelize a single 40Gb/s into four 64×156.25Mb/s data streams.
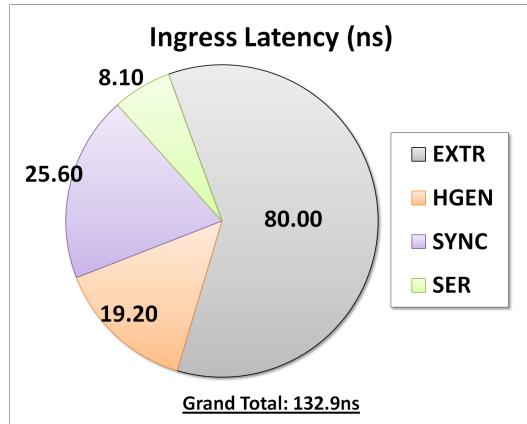
(a)
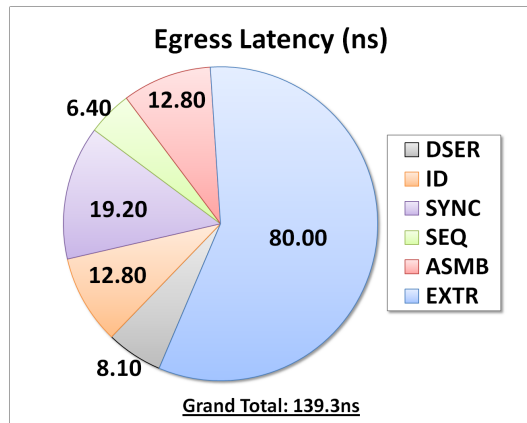
(b)

Figure 7.10: *(a) Measured latencies of Ingress (132.9ns) and (b) Egress (139.3ns) Adaptation Layers.*

7.2.2 Link Utilization Performance

The end-to-end adaptation performance measurements carried out thus far have consisted of point-to-point experiments where there exists one transmitter/receiver pair. Additionally, the rate at which packets are received by the Egress Adaptation Layer has been less than or equal to the rate of packets being transmitted by the Ingress Layer. In this implementation, 100MbE frames are received by the Ingress Adaptation Layer at a link utilization of 89.3% and are then up-converted to 40Gb/s labeled packets resulting in a diminished utilization of 0.63%. In practice, input to the Egress Layer will be aggregated from multiple core router nodes leading to a link utilization relatively larger than what is observed in present characterizations. The current allowable throughput of the Egress Adaptation Layer is limited to 148,000.81 packets per second (P/s), which is enforced by the throughput of the 100MbE network interface. One can foresee a performance degradation as link utilization is increased beyond the maximum allowable throughput, which will lead to memory storage overflow. Memory collisions will in turn lead an increase of dropped packets.

This section evaluates the Egress Layer performance under varying levels of link utilization to determine the amount of local packet storage required. A first-order performance simulation is carried out by utilizing parameters that are based on the current Egress Layer implementation. Additionally, the received power is assumed to be 1mW to ensure performance degradation is solely a result of memory overflow. The utilization performance can be estimated using the following equation: $\text{PR}(\%) = \frac{100}{B} \times \left[M + \frac{T}{R} \times (B - M) \right]$. Here, the packet recovery (PR) is proportional to the incoming packet rate (R), outgoing packet transmission rate (T), size of incoming packet burst (B), and total packet storage (M). Assuming the incoming packet rate is

greater than the outgoing rate ($R > T$), performance can be classified under two operation regimes: 1)memory is capable of storing the entire burst of incoming packets or 2)input burst is larger than available storage. In the first regime, the entire burst is written to memory resulting in zero packet loss. In the second regime, a limited amount of packets will be stored at the beginning of the burst resulting in dropped packets once memory is filled. Therefore, recovery performance is proportional to the sum of initially stored packets (M) and packets successfully recovered once memory is full $\left[\frac{T}{R} \times (B - M)\right]$. The PR value is held constant at 100% during instances when PR value is exceeds 100%. A volatile memory capacity of 32kB is assumed for 40.8-nanosecond packets consisting of 10Gb/s headers (H), 40Gb/s payloads(P), and 12.8-nanosecond header-to-payload guard bands (GB). Table 7.1 lists all the simulation parameters used.

The results from the link utilization performance simulations are shown in Figure 7.11. A packet recovery of 100% is achieved for link utilization values below the 0.63% critical point where R = T. When considering the

Table 7.1: *Simulation parameters used to estimate link utilization performance of Egress Adaptation Layer.*

Parameter	Value	Description
PR	$0\% \leq PR \leq 100\%$	Packet recovery
M	32kB	Available memory storage
B	$1 \leq B \leq 10^6$	Input packet burst length
T	148.81kP/s	Packets transmission rate at network interface
R	$\dfrac{U}{100 \times LOP}$	Rate packets arrive at Egress Layer input
U	$0\% \leq U \leq 100\%$	Link utilization or stream duty cycle
LOP	40.8ns	Packet duration (H:128ns, GB: 12.8ns, P: 15.2ns)
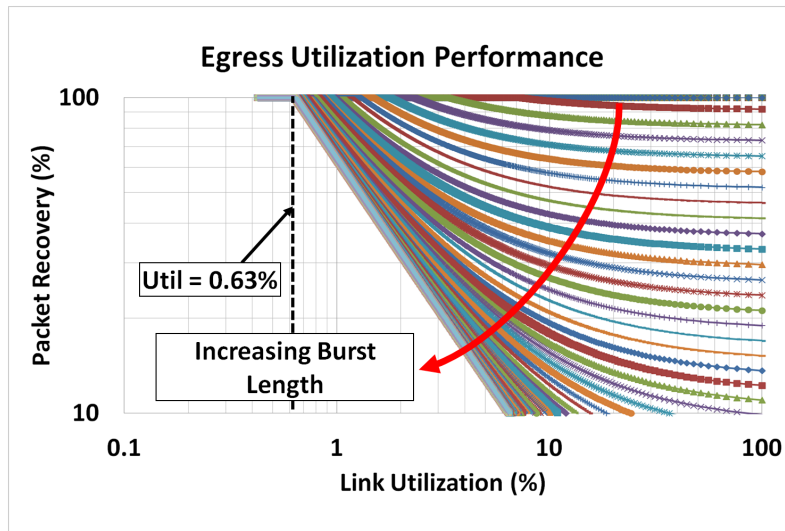
Figure 7.11: *Simulations showing end-to-end performance of current Egress Layer implementation under varying levels of link utilization.*

largest burst length, the packet recovery is inversely proportional to R. One will also note that perfect packet recovery can be achieved at higher levels of utilization as long as there is sufficient memory to accommodate shorter bursts of packets.

Figure 7.12 shows the expected performance at full (100%) link utilization evaluated at varying amounts of memory capacities. The packet burst size is varied from 1 to 10 million packets while the memory capacity ranges from 16kb to 8MB. Results demonstrate that the current Egress Layer implementation (32kB) is capable of accommodating a burst of 512 packets before incurring packet loss caused by memory overflow. They also suggest that increasing memory capacity beyond 8MB is crucial to maintaining loss-free performance under heavy link utilization. Packet storage is currently implemented via FPGA-based logic blocks in order to minimize the latency incurred during memory access. However, external memory banks capable
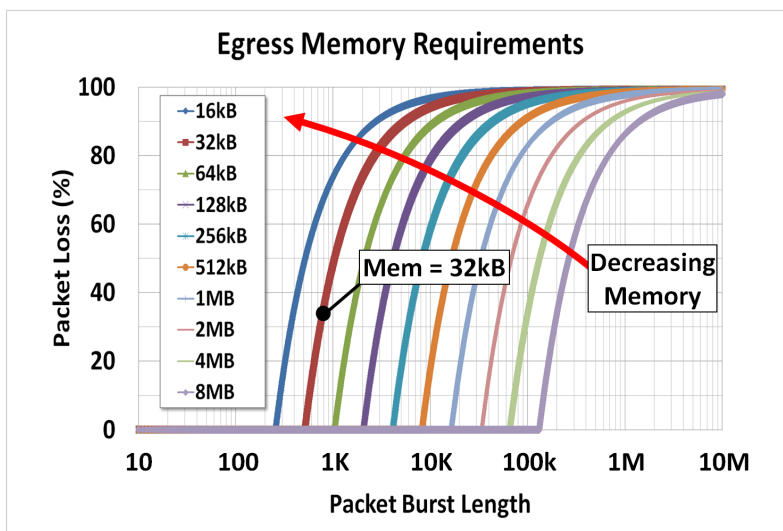
Figure 7.12: *Simulations showing end-to-end adaptation performance of current Egress Layer implementation as a function of packet burst lengths and available memory assuming full link utilization.*

of simultaneous, asynchronous read/write access will need to be utilized to obtain a memory capacity on the order of tens of megabytes. Doing so will result in increased memory access latencies that already serve as the major contributors of latency within the end-to-end adaptation process.

7.3 Chapter Summary

The performance characterizations of the end-to-end adaptation process has been presented in this chapter. The characterizations consist of Layer-II and III recovery in addition to latency and link utilization measurements. A commercial packet tester is used to transmit 64-byte 100MbE frames to the Ingress Adaptation Layer, where they are converted into optical packets consisting of a 10Gb/s NRZ labeled header and 40Gb/s high-speed payload. The packets were then transmitted though an optical link forwarded to the Egress

Adaptation Layer. Layer-II recovery measurements have been carried out at every functional stage of the Egress FPGA. The Payload Recovery stages achieved a payload loss percentage less than 10^{-4}. The Sequencer stage exhibited small loss of payloads at receiver power levels larger than -30dBm and 5% additional loss otherwise. The Frame Assembly stage, achieved frame a loss clamped at 0.3% at received power levels above -30dBm. The Frame Extraction stage showed a constant loss floor located near 0.6%. Layer-III frame recovery measurements, performed at the packet tester via 32-bit CRC validations, demonstrated an overall adaptation frame loss less than 0.6% for receiver power levels larger than -30dBm. Power penalty measurements carried out relative to the Layer-III performance, evaluated at the packet tester, exhibited an overall power penalty distribution within 3.5dB. The Sequencer, Assembler, and Extractor stages each incurred 0.5dB of incremental power penalty. Furthermore, an end-to-end latency of 272.2ns dominated by memory accesses (102.8ns) was observed. Link utilization simulations revealed a 40% packet loss penalty when increasing link utilization from 0.63% to 1%. Simulations performed assuming 100% link utilization suggested that a memory increase from 32kB to tens of megabytes is required to overcome performance degradation. A trade-off between link utilization performance and excess latency needs to be considered when increasing memory capacity in future Egress Layer implementations.

References

[1] "Internet Header Format," *Internet Protocol DARPA Internet program protocol specification* IETF, pp. 14. STD 5, RFC 791, September 1981.

[2] Xilinx, "Virtex-4 FPGA User Guide," http://www.xilinx.com/support/documentation/user_guides/ug070.pdf, December 2008.

Chapter 8

A Dynamically Re-Sizable Optical Buffer

Traffic in today's networks consist of packets sizes spanning lengths between 40 to 1500 bytes. Optical packet buffers with variable storage times are crucial to avoid the overhead incurred when fragmenting packets into fixed lengths. This chapter presents the progress made towards demonstrating optical storage capable of accommodating variable-length packets. The re-sizable buffer technology shown is based on a design consisting of an integrated 2x2 SOA switch with an external fiber-based delay configured in a re-circulating fashion. A discrete, SOA-based, variable delay is inserted into the recirculating loop to facilitate buffer re-sizing. The variable delay is shown to achieve a tuning range between 232.47ns and 323.82ns with 6.09ns of temporal resolution. This corresponds to a maximum storage space of 807 bytes with 32 bytes of resolution (evaluated at 40Gb/s). Dynamic buffer operation is demonstrated utilizing packet sizes ranging from 40 to 800 bytes. An all-optical payload envelope detection circuit (AO-PED) is used to determine the location and duration of incoming packets. A packet recovery

greater than 95% is achieved over one buffer circulation across a 4dB range of receiver power for 40-byte packets. On the other hand, two circulations are demonstrated over a 5dB range of receiver power for 800-byte packets. The optical signal OSNR performance is limited by ASE accumulation where the SOA switching dynamics demonstrated improved performance with high-duty cycle data streams. Additionally an OSNR degradation of 1.67dB per buffer circulation is observed.

8.1 Introduction

Optical packet buffering is one important approach to contention resolution and blocking in an optical router [1]. To date, packet buffer solutions have depended heavily on re-circulating loop configurations that must be designed to accommodate the largest packet size while trading off latency penalties smaller packets. Current IP networks consist of average packet sizes that vary from 40 bytes (40%) to 1500 bytes (10%) [2]. The majority of work in this field has provided several novel synchronous designs with theoretic analysis at data rates of 10Gb/s [3–5]. The work presented here, demonstrates a truly asynchronous, real-time, dynamically re-sizable optical buffer that adjusts its storage size based on the length of incoming packets. The re-sizable optical buffer is based on a digitally programmable delay matrix that is coupled in a re-circulating configuration to an integrated 2x2 cross-bar switch. A payload envelope detection (PED) technique is not only used to temporally locate a packet, but it is also used to determine its length. The ascertained location and length of a packet are then used to perform an electronic table lookup to carry out buffer re-sizing. The demonstration uses integrated optic technology and is amenable to chip-level integration.

8.2 Principle of Operation

The challenge of a re-sizable optical buffer lies in the trade off between a compact design and its flexibility. Larger packet lengths will require longer optical delays that introduce excess loss. Additionally, several lengths of delay are required to provide an acceptable level of storage time resolution, which can potentially result in an enhanced footprint. One could use a single length of fiber to accommodate the longest possible packet duration that will effectively allow for subjacent packet lengths to be stored. However, such a configuration may result in a diminished buffer throughput when storing packets with durations that are much shorter than the buffer storage time.

The illustration in Figure 8.1 shows the basic operation of the re-sizable buffer presented here. The left-most part of the figure shows two separate packets, of varying lengths (Δ and 3Δ), about to enter the buffer. The packets are evenly split between the buffer input and the control path consisting of a payload envelope detector (PED) and the electronic buffer control stage. The PED circuit provides an electrical representation of the incoming packets in the form of an envelope function whose on-time is comparable to the duration of its corresponding packet. The buffer controller is then able to sample the envelope signal to determine the location of the packet in addition to its length. Once the location and duration of a packet is determined, the buffer controller proceeds to generate the necessary switching signals to allow packets to be gated through suitable lengths of delays.
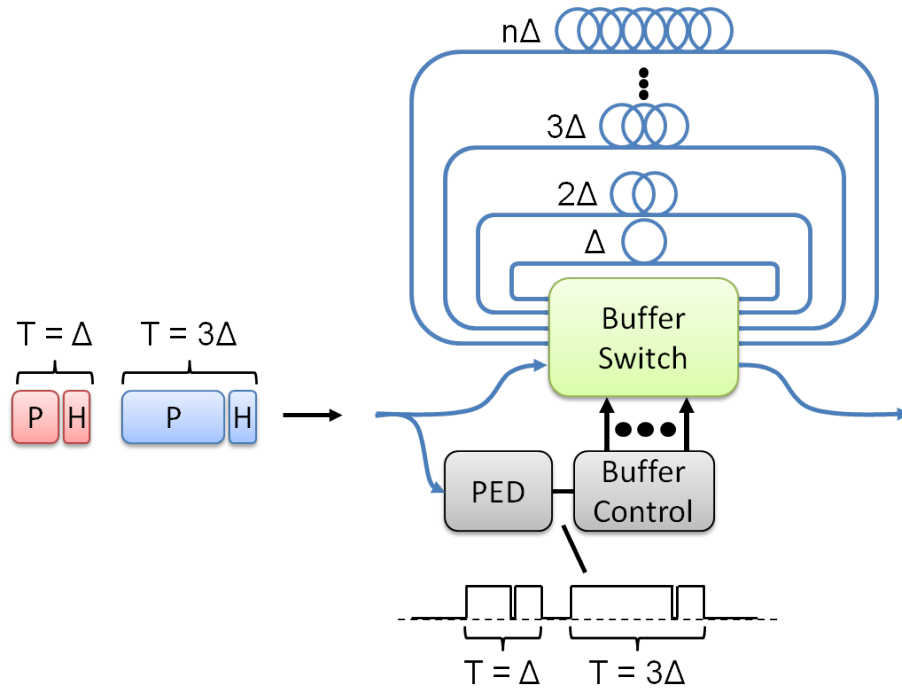
Figure 8.1: *Re-sizable optical packet buffer principle of operation (PED: payload envelope detection).*

8.3   Re-Sizable Buffer Implementation

A schematic representation of the proposed re-sizable optical buffer is shown in Figure 8.2. A variable delay is coupled to a 2x2 semiconductor optical amplifier (SOA) crossbar switch in a re-circulating fashion. A polarization controller is placed within the delay in order to maintain a TE polarization heading into the SOA gate array. Figure 8.3 shows the physical layout of the InP SOA-based 2x2 crossbar switch. The SOAs located near the device facets serve as booster amplifiers biased at constant currents of 80mA while the SOAs near the center of the device are designed to be high-speed optical switches. The switch is fabricated on the offset quantum well (OQW) platform where the SOAs have been previously demonstrated at 40Gb/s and meet

the requirements of low crosstalk (<-40dB), high extinction ratios (>40dB), and fast switching times (<2ns) [6]. During operation, input signals are gated either towards the output port or towards the fiber loop. Re-circulations are achieved by directing packets back towards fiber delay.
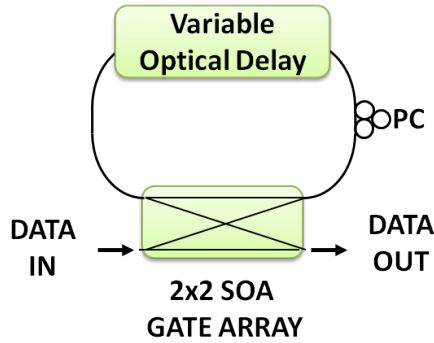
Figure 8.2: *Schematic represen-
tation of a re-sizable optical packet
buffer.*
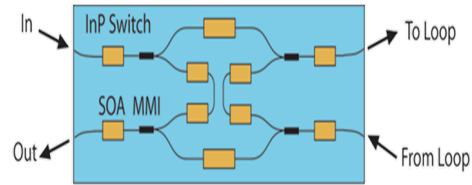
Figure 8.3: *Physical layout of InP
SOA-based 2x2 cross-bar switch [6].*

The re-sizing function is carried out by inserting the four-stage, feed-forward switched delay depicted in Figure 8.4 [7,8]. Each stage of the variable delay consists of a pass arm in addition to a delay arm with a Δ-delay that increases by powers of two at each stage. The SOAs serve as on/off switches for the packets as well as a means of loss compensation. Isolators (ISO)
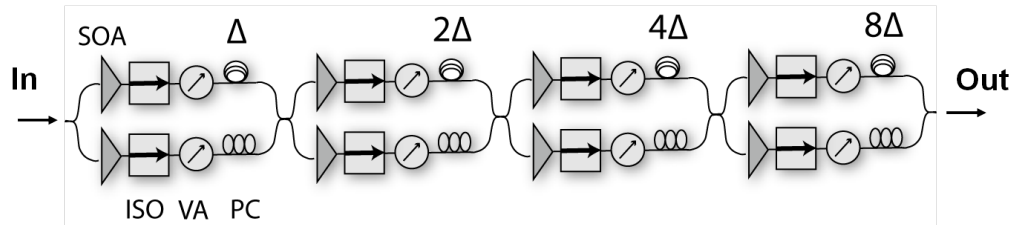
Figure 8.4: *Implementation of feed forward SOA-based variable delay (ISO:
isolator, VA: variable attenuator, PC: polarization controller) [7, 8].*

160

are included to eliminate back-reflections, while VAs are used to match the power through any path combination within 1dB. Such a configuration results in delays of $T_{N\Delta} = T_{0\Delta} + N \times \Delta$, where $T_{0\Delta}$ and Δ have been measured as 232.47ns and 6.09ns respectively. Thus, the shortest path through the synchronizer is $T_{0\Delta} = 232.47$ns while the longest path is $T_{15\Delta} = 323.82$ns with a resolution of 6.09ns.
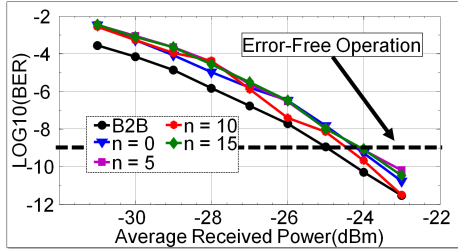
Figure 8.5: *BER characterizations of the variable delay configurable path lengths.*
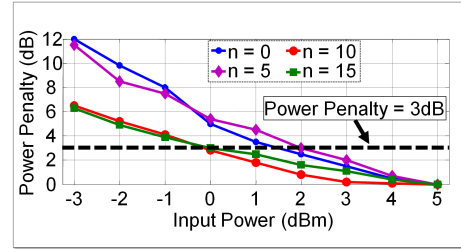
Figure 8.6: *Power penalty characterizations of variable delay evaluated at BER = 10^{-9}.*

The Layer-I performance of the variable delay was determined via bit error rate (BER) measurements, for a subset of possible delays, using a 40Gb/s non-return-to-zero (NRZ) $2^{15} - 1$ pseudo random binary sequence (PRBS) data stream. Figure 8.5 shows the results of the BER measurement performed at an input power of 5dBm for values of $n = 0$, 5, 10, and 15. The BER results show that each measured path achieved error-free operation (BER$\leq 10^{-9}$) with an observed power penalty within 1dB of the back-to-back measurement taken by removing the variable delay from the link. Figure 8.6 depicts the results of the power penalty measurements taken as a function of average power delivered to the variable delay. The power penalty was evaluated relative to an expected BER value of 10^{-9}. The degradation of the optical signal to noise ratio (OSNR) limits the switched delay to a dynamic power range of 3dB of decreased power to operate below a power penalty of 3dB.

161
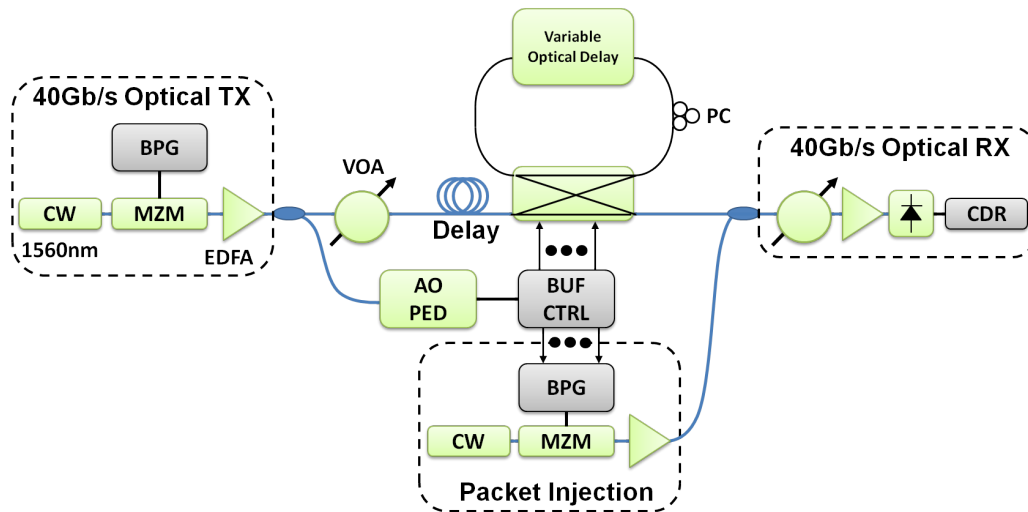
Figure 8.7: *Experimental setup used to measure the re-sizable optical packet buffer.*

A diagram of the experimental setup used to demonstrate and characterize the re-sizable optical packet buffer is shown in Figure 8.7. A bit pattern generator (BPG) is used to modulate a 40Gb/s packet stream onto a CW optical signal. The CW output is set to 1560nm to coincide with the gain peak of the optical switch. The packet stream is then evenly distributed between the optical path and the electronic control path of the buffer using a 3dB coupler. The optical path consists of a processing delay ($T_{PROC} = 190$ns) that is immediately followed by a polarization controller and the re-sizable packet buffer. The electronic control path is comprised of an all-optical payload envelope detection (AO-PED) circuit and the FPGA-based electronic buffer controller (BUF-CTRL). The recovered PED signals are sampled by the BUF-CTRL to establish an estimate of the temporal location and the duration of each packet. The sampling occurs at the local FPGA clock rate of 156.25MHz, which allows for a temporal resolution of 6.4ns (32 bytes at 40Gb/s). The amount of clock cycles that coincide with the on-time of a

162

PED signal are monitored and used to determine the length of incoming packets. An electronic lookup table is used to determine the appropriate buffer storage capacity based on the PED duration, which is shown in Table 8.1. Output from the re-sizable buffer is then forwarded to an optical receiver where packet recovery measurements are performed. The receiver consists of a variable optical attenuator (VOA) followed by an pre-amplified photo-detector and a clock and data recovery (CDR) stage. A separate 10Gb/s optical transmitter is used to inject dummy 40-byte packets in between valid packets to in crease the stream duty cycle. Thus, DC-balance is achieved and burst mode transients are avoided.

Table 8.1: *Electronic lookup table used by the buffer controller to configure storage size of optical packet buffer.*

PED Duration	Δ	Delay(ns)	Storage (bytes)
≤17	0	230	≤576
18	2	242	≤605
19	4	254	≤635
20	6	266	≤665
21	8	278	≤695
22	10	290	≤725
23	13	302	≤755
24	15	323	≤807

8.4 Packet Recovery Performance

The maximum allowable buffer size is limited by the longest path through the variable delay ($T_{15\Delta} = 323.82$ns). The buffer fiber delay is chosen to be twice the temporal length of a packet to allow for less stringent contention resolution requirements. Packet sizes of 40, 400, 600, 650, 700, 750, and 800 bytes are chosen to test buffer reconfigurability. Each packet consists of the following structure: a 32-bit idler (0xAAAAAAAA), 64-bit unique packet identifier (0x89ABCDEFFEDCBA98) and a repeating $2^7 - 1$ PRBS pattern. The amount of space between packets is determined by the BC processing time (T_{PROC}) and the time required for the maximum amount of circulations (3 x $T_{15\Delta}$). This, therefore, results in stream duty cycles of 9% and 0.5% for 800- and 40-byte packet streams respectively. Additionally, the re-sizable buffer is programmed to store all incoming packets a fixed number of times during each measurement in order to study the performance of the buffer as a function of delay circulations.

Previous PED implementations were realized by utilizing a bandwidth limited 10GHz photo-detector followed by a 2.5GHz limiting amplifier [9]. The data stream used in this has a relatively low duty cycle, which causes the signal-to-noise ratio (SNR) of the limiting amplifier to drastically deteriorate. An all-optical design is, therefore, utilized in this demonstration. Figure 8.8 shows a block diagram of the AO-PED circuit which consists of two stages of cross-gain modulated (XGM) SOAs operating in a saturated regime. The input 1560nm signal is combined with an intermediate 1550nm using a 3dB coupler. The signal pairs are then forwarded to an SOA where an inverted version of the input signal is transferred to the 1550nm pump via XGM. A filter is placed directly after the SOA to suppress the 1560nm input

signal before being amplified by an EDFA. The signal power is enhanced in order to saturate the gain of the second SOA, which allows one to obtain an envelope function. An additional inversion operation is performed by utilizing a separate 1560nm optical signal as the pump input to the secondary XGM stage. An optical filter is placed at the output of the second XGM stage to suppress the first 1550nm pump allowing the 1560nm optical PED signal to be forwarded to a 10Gb/s photo-detector.
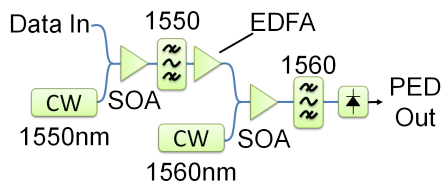
Figure 8.8: *Schematic diagram of all-optical payload envelope detection circuit (AO PED).*
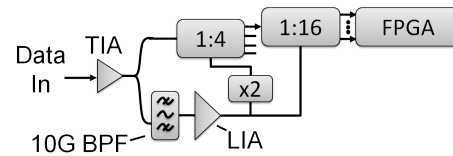
Figure 8.9: *Schematic diagram of 40Gb/s asynchronous clock and data recovery (CDR) circuit.*

The CDR stage used in the optical receiver is shown in Figure 8.9. Packets entering the CDR stage are amplified and forwarded to a 1:4 DSER to convert the single 40Gb/s data stream into four 10Gb/s parallel data streams. A lack of resources only allowed for the monitoring of one 10Gb/s stream, while the other three were discarded via AC-coupled 50Ω-to-ground terminations. The single 10Gb/s stream is then sent through a 1:16 de-serializer whose output was connected to an FPGA where real-time, Layer-II packet recovery measurements are performed. Furthermore, identifier recovery measurements were performed on only 16-bits of the packet identifier (0x55AA) since only $\frac{1}{4}$ of the data stream was utilized at the 1:4 DSER.

Verification of dynamic delay configuration is verified visually using a 50GHz sampling oscilloscope. Figure 8.10 shows the input optical stream consisting of variable-length packets (top) in addition to its corresponding

165

electronic PED signals (bottom). The resulting PED signals exhibits rise/fall times <350ps regardless of the data stream duty cycle. Furthermore, Figure 8.11 shows traces of the optical signals at various points of the experimental setup. The top-most trace shows the injected dummy packets with a brief period of zero-signal used as a placeholder for packets exiting the re-sizable buffer. The three traces below show the buffered 40-, 400-, and 800-byte packets combined with the stream of dummy packets.
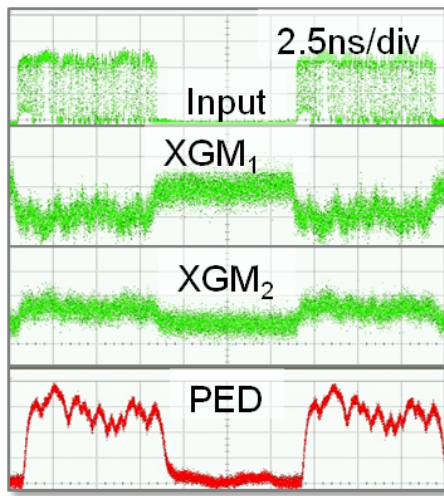
Figure 8.10: *Oscilloscope traces of AO-PED signals (bottom).*
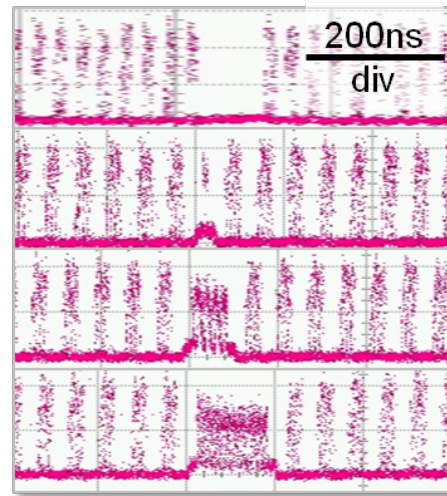
Figure 8.11: *Oscilloscope traces of packet injection output (top), and buffered packets (bottom three).*

Identifier recovery results performed for packet lengths of 40 and 800 bytes are shown in Figure 8.12 and Figure 8.13, respectively. Results show that 40-byte packets can be circulated up to one time with a packet recovery greater than 95% over a dynamic power range of 4dB, while 800-byte packets are capable of achieving up to two circulations over a 5dB dynamic range. Saturation of the receiver TIA, at higher levels of received power results in degradation of packet recovery performance. Previous fixed-delay
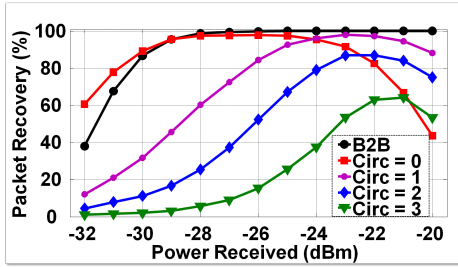
Figure 8.12: *Packet recovery results for 40-byte packets.*
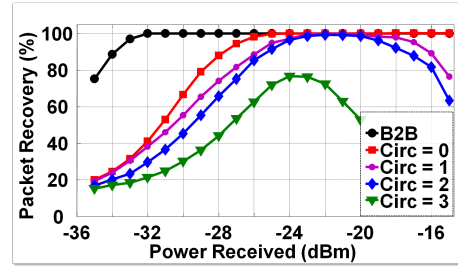
Figure 8.13: *Packet recovery results for 800-byte packets.*
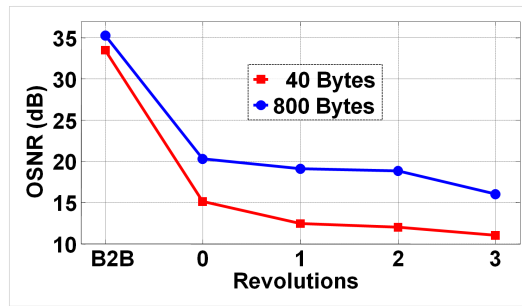
Figure 8.14: *OSNR as a function of buffer revolutions for an input power of -4dBm.*

measurements have demonstrated up to eight circulations limited by the accumulation of ASE [6]. The current buffer configuration further exacerbates the ASE buildup since packets must pass through four additional SOAs per revolution. The discrepancy in packet recovery between the two streams can also be explained by difference in optical signal to noise OSNR of the two streams caused by the differing stream duty cycles.

The OSNR of the buffer output was monitored by taking the ratio of power between the signal ($\lambda = 1560nm$) and the noise ($\lambda \pm 1nm$) via an optical spectrum analyzer (OSA). Figure 8.14 shows the OSNR variance as a function of buffer revolutions for packet sizes of 40 and 800 bytes. The results show an OSNR degradation of 15 and 20dB when entering the SOA switch

(zero revolutions) for 800- and 40-byte packets respectively. Once inside the buffer, the OSNR degrades at an approximate rate of 1.67dB per revolution.

8.5   Chapter Summary

An asynchronously loaded re-sizable packet buffer for 40Gb/s optical packets has been presented. The buffer design consists of an InP SOA cross-bar switch coupled to a re-circulating fiber-based delay. The re-sizing functionality was realized by inserting a variable delay based on a four-stage, feed-forward SOA switch design. A tuning range between 230 and 323ns was demonstrated with 6.09ns of resolution. This corresponded to a maximum storage capacity of 807 bytes with 32 bytes of resolution (evaluated at 40Gb/s). An all-optical payload envelope detection scheme has also been used to determine the location and the duration of incoming packets within 6.4ns (32 bytes) of accuracy. Furthermore, utilization of an all-optical circuit resulted in PED signals with rise/fall times below 350ns. Real-time, burst mode payload identifier recovery was performed to determine buffer performance for 40- and 800-byte packets. One buffer revolution was achieved with greater than 95% recovery rate over a 4dB dynamic range of receiver power for 40-byte packets, while two circulations were demonstrated over 5dB of received power for 800-byte packets. The performance was limited due to the accumulation of ASE resulting from the utilization of multiple SOA-based feed-forward switching stages. Streams with larger duty cycles exhibited higher OSNR values after passing through the SOA switches. An OSNR degradation of about 1.67dB per buffer circulation was then observed regardless of stream duty cycle. The design presented, successfully demonstrated potential for contention resolution within an ODR for variable-length

optical packets ranging from 40 to 800 bytes. Furthermore, accommodation of packets lengths reaching 1500 bytes could be attained by inserting an fifth feed-forward stage with a 16Δ delay.

References

[1] D. K. Hunter, M.C. Chia, I. Andonovic, "Buffering in optical packet switches," *IEEE Journal of Lightwave Technology*, vol.16, pp. 2081-2094, December 1998.

[2] K. Thompson *et al*, "Wide-Area Internet Traffic Patterns and Characteristics," *IEEE Network*, vol. 11, no. 6, pp. 10-23, August 2006.

[3] X. Li, L. Peng, J. Chen, S. Wang, G. Wu, J. Lu, Y. Kim, "A Novel Fast Programmable Optical Buffer with Variable Delays," in *National Fiber Optic Engineers Conference, (NFOEC) 2008*, Paper JThA41, 2008.

[4] C. J. Chang-Hasnain et al, "Variable optical buffer using slow light in semiconductor nanostructures," *Proceedings of the IEEE*, vol. 91, no. 11, pp. 1884-1897, Nov. 2004.

[5] J. Yang *et al*, "Continuously Tunable, Wavelength-Selective Buffering in Optical Packet Switching Networks," *IEEE Photonics Technology Letters*, vol. 20, no. 12, pp. 1030-1032, May 2008.

[6] E. F. Burmeister, J. P. Mack, H. N. Poulsen, J. Klamkin, L. A. Coldren, D. J. Blumenthal, J. E. Bowers, "SOA gate array recirculating buffer with fiber delay loop," *Optics Express*, vol. 16, pp. 8451-8456, 2008.

[7] J. P. Mack, H. N. Poulsen, D.J. Blumenthal, "40Gbps autonomous optical packet synchronizer," in *Optical Fiber Communications Conference, (OFC) 2008*, Paper OTuD3, 2008.

[8] J. P. Mack, H. N. Poulsen, D. J. Blumenthal, "Variable Length Optical Packet Synchronizer," *IEEE Photonics Technology Letters*, vol. 20, no. 14, pp. 1252-1254, July 15, 2008.

[9] D. Wolfson, V. Lal, M. Masonovic, H.N. Poulsen, C. Coldren, G. Epps, D. Civello, P. Donner, D.J. Blumenthal, "All-Optical Asynchronous Variable-Length Optically Labeled 40 Gb/s Switch," *European Conference on Optical Communications, ECOC 2005*, Paper Th.4.5.1, September 2005.

Chapter 9

End-to-End Adaptation, Forwarding, Buffering, and 3R Regeneration

In this chapter, the technology required to achieve the first edge-to-edge demonstration of Internet Protocol (IP) traffic transmission through a regenerative, buffered optical packet switched (OPS) router is presented. The end-to-end performance of a 2x2 40Gb/s label swapped optical router is presented along with detailed discussions of design and implementation of each subsystem. The optical router consists of packet buffers that are utilized to resolve any contention that may occur when multiple packets make an identical port request. The routing and header rewrite functionalities are executed by a set of packet forwarding planes that consist of high-speed wavelength converters connected to an arrayed waveguide grating. Signal regeneration is performed via 3R stages that amplify, reshape, and re-time incoming packets. The end-to-end performance of each major optical router component is characterized using the Adaptation Layers presented in Chapter 6. Both optical

packet buffers are packaged in a custom driver board and demonstrate compatibility with the end-to-end adaptation process by exhibiting greater than 99% Ethernet frame recovery for at least 128ns of storage time (two circulations). A buffer time of 192ns (three revolutions) can be achieved if one is willing to incur a 10% frame recovery penalty. The packet forwarding planes consist of 40Gb/s wavelength conversion and 10Gb/s header rewrite stages. Both sections use a widely tunable lasers as pump input that is packaged in a custom, FPGA-based driver board. The lasers demonstrates tuning ranges of about 20nm (2.5THz) with nanosecond-scale (<10ns) wavelength switching times. Header erasure and rewrite is successfully demonstrated with 100ps of temporal accuracy. Both packet forwarding planes show compatibility with end-to-end adaptation by achieving greater than 99% Ethernet frame recovery for at least five output wavelengths. The 3R regeneration stage consists of a high-speed wavelength converter with a 40GHz optical clock recovery block as its pump input. The optical clock recovery circuit is comprised of an integrated mode-locked laser operated in a hybrid locking configuration. The 3R stage shows amplification (1R), limited noise reduction(2R), and about 0.20ps of jitter reduction (3R). End-to-end adaptation compatibility of the 3R regeneration stage is demonstrated by achieving greater than 99%. A 5dB improvement in receiver sensitivity is observed when comparing the 3R circuit against a 2R implementation that employs a CW laser source as its pump input. All subsystems configured in with a custom, multi-threaded application allowing real-time configuration and statistical analysis of the 2x2 optical router.
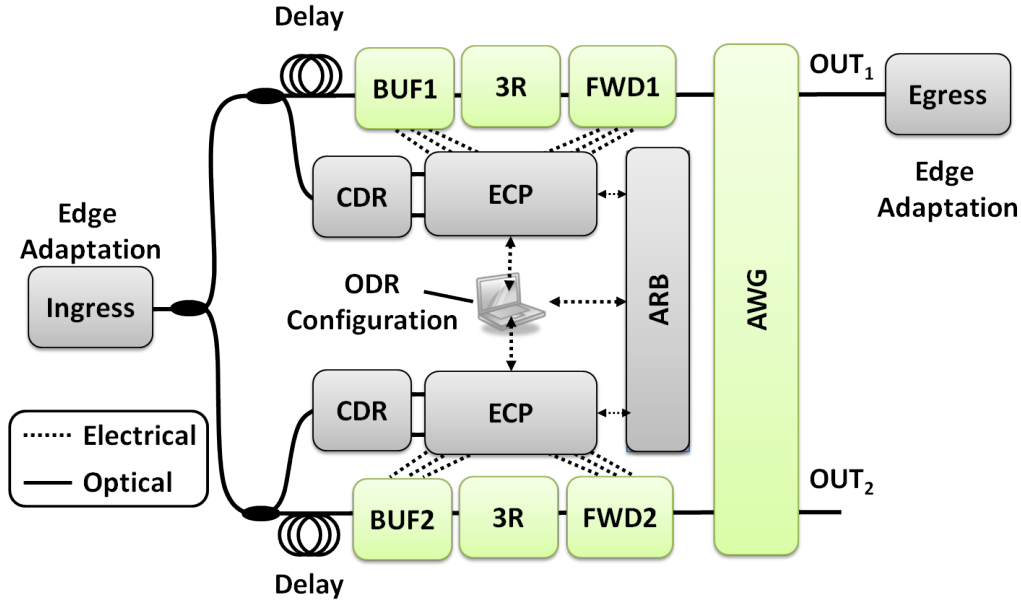
Figure 9.1: *A 40Gb/s 2x2 optical data router (ODR).*

9.1 Architecture Overview

A schematic representation of a 40Gb/s 2x2 label switched optical router (LASOR) is illustrated in Figure 9.1. Labeled 40Gb/s optical packets, in the format described in Section 6.3, enter the ODR through a 3dB optical splitter. Packets were evenly distributed between each ODR arm in order to ensure that packets are aligned to the local time slot of the ODR. Otherwise, an optical packet synchronization would be required in each optical path [1]. Each arm of the ODR consists of an optical data path and an electronic control path where high-speed payloads and low-speed headers are respectively processed. The optical path contains an optical packet buffer (BUF), a 3R regeneration stage, and a packet forwarding plane (FWD). The BUF stage consists of a re-circulating fiber-based delay packet buffer used to resolve contention resolution, while the 3R stage is utilized for signal regeneration. The packet forwarding plane consists of wavelength conversion and header

174

rewrite stages that are connected to an arrayed waveguide grating router (AWGR) to perform the key routing functionality. In this demonstration, the 3R regeneration is placed before the switching fabric instead of the output ports. This is done to preserve the formatting of the non-return-to-zero (NRZ) 10Gb/s headers, which would be altered by the re-timing stage.

The electronic control plane consists of a clock and data recovery (CDR) stage, an electronic channel processor (ECP), and an arbitration stage (ARB). The CDR stage performs burst mode clock and data recovery on incoming low-speed headers and forwards the data request to the ECP. The ECP validates the integrity of the header and proceeds to extract the forwarding information located in the labeled header. Before the ECP can generate the necessary signals for the optical path, it must first forward its port request to the ARB. The ARB stage then examines all of the port requests to determine if there is any possibility of contention. If a collision is detected, the ARB determines which packets to forward, store, or drop. Once arbitration is complete, the ECPs are informed of the appropriate course of action to take when generating control signals for the optical path.

In this chapter, the aliases BUF1 and FWD1 are used to represent the packet buffer and forwarding plane in the top arm of the LASOR router, respectively. Similarly, BUF2 and FWD2 correspond to the buffer and forwarding plane in the bottom arm.

9.2 Optical Packet Buffer

Figure 9.2 shows a schematic of the optical packet buffers used in the LASOR router. Each buffer is configured with a re-circulating fiber-based delay coupled to a 2x2 optical cross bar switch. The delay is chosen to be 64ns long
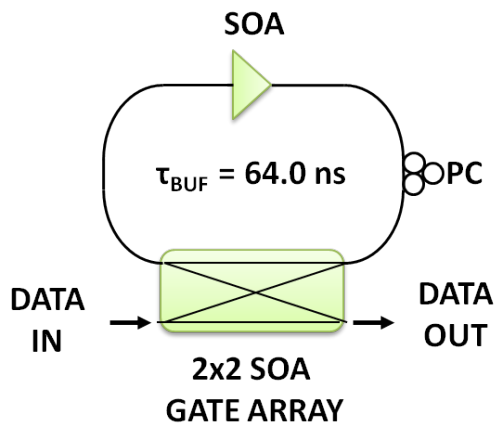
Figure 9.2: *Schematic representation of a re-circulating optical packet buffer*

to allow ample time to switch packets in and out of the buffer while meeting the guard band requirements. The 2x2 optical switch consists of an InP semiconductor optical amplifier (SOA) switch matrix. The fiber delay contains an off-the-shelf bulk SOA (G = 20dB) that is used to account for any losses within the external delay. Additionally, a polarization controller (PC) is placed in the fiber loop to maintain a transverse electric (TE) polarization when re-entering the cross-bar switch. The integrated SOA switches in this work, previously demonstrated nanosecond-scale switching times with less than -40dB crosstalk [2]. The cross-bar configuration of the switches allows input packets to either bypass the external fiber delay or allows them to be switched into the fiber delay. Packets can then be either re-circulated towards the delay or they can be forwarded towards the switch output.

9.2.1 End-to-End Characterization

The experimental setup used to characterize the optical packet buffers is displayed in Figure 9.3. Adapted packets from the Ingress Layer are evenly

176
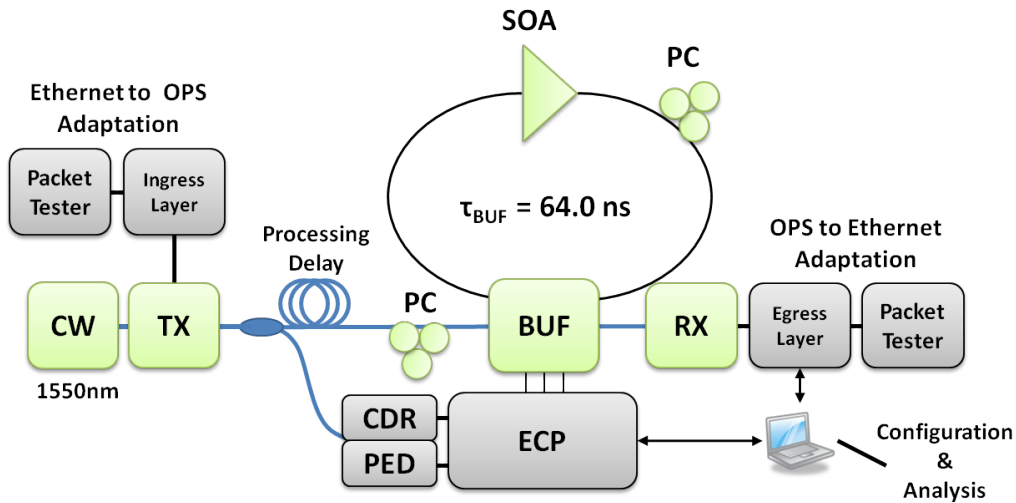
Figure 9.3: *Schematic diagram of buffer characterization setup.*

split between the optical data path and the electronic control path using a 3dB splitter. The optical path consists of about 300ns of processing delay followed by a polarization controller leading to the packet buffer (BUF). The electronic control path leads to the clock and data recovery (CDR) stage that is used to detect and recover the 10Gb/s labeled header. The temporal location of the 40Gb/s payload is monitored within 6.4ns of accuracy using the payload envelope detection (PED) stage. The recovered header and PED signal is then forwarded to the Electronic Channel Processor (ECP) that is used to generate the control signals for the optical buffer. The purpose of this measurement was to determine the performance limits of the buffers within the end-to-end adaptation process, rather than demonstrate contention resolution. As a result, the behavior of the buffer is engineered to store every incoming packet a fixed number of times n where $n \geq 0$. The buffer configurations and the signal fine-tuning is set using a custom multi-threaded ODR Configuration Tool that interfaces with the ECP via a USB-to-RS232 bridge controller.

177
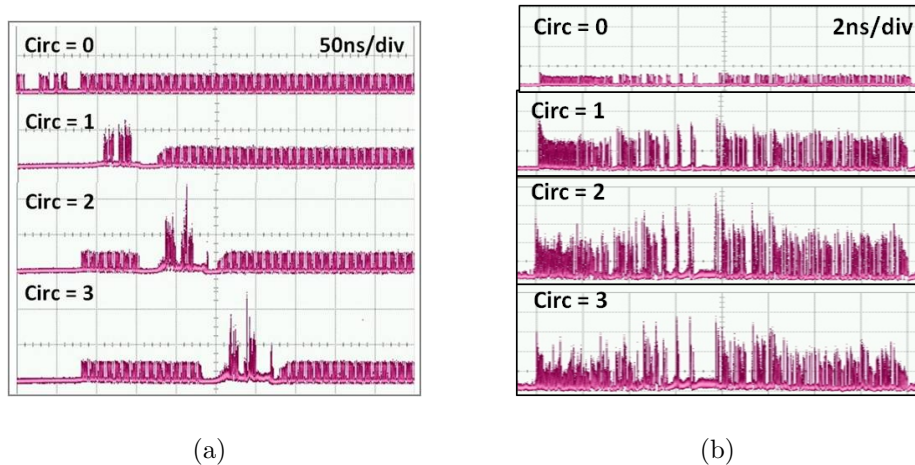
Figure 9.4: *(a) Oscilloscope traces verifying buffer functionality and (b) showing signal degradation over multiple buffer circulations.*

Visual verification of buffer functionality is demonstrated via the oscilloscope (OSC) traces in Figure 9.4. The traces in (*a*) show the output of an optical buffer where a packet is able to achieve up to 192ns of storage time (3 circulations) with 64ns of temporal resolution. The traces in (*b*) show a magnified view of the optical packet as it iterates through several buffer circulations. One can see that there is a net gain when between the packets that bypass the delay (Circ = 0) and those that are optically buffered (Circ > 0). Additionally, a patterning effect stemming from the carrier recovery dynamics of the integrated SOA is observed. A spike in optical power is present within leading pulses that follow a strand of zero-level bits, while trailing pulses deplete carriers faster than they are replenished. This leads to gain saturation and lower optical powers for trailing pulses. Furthermore, this effect is compounded through each buffer circulation, which further degrades signal quality. Hence, optical packet buffer performance seems to be initially limited by SOA patterning instead of amplified spontaneous emission (ASE) noise.
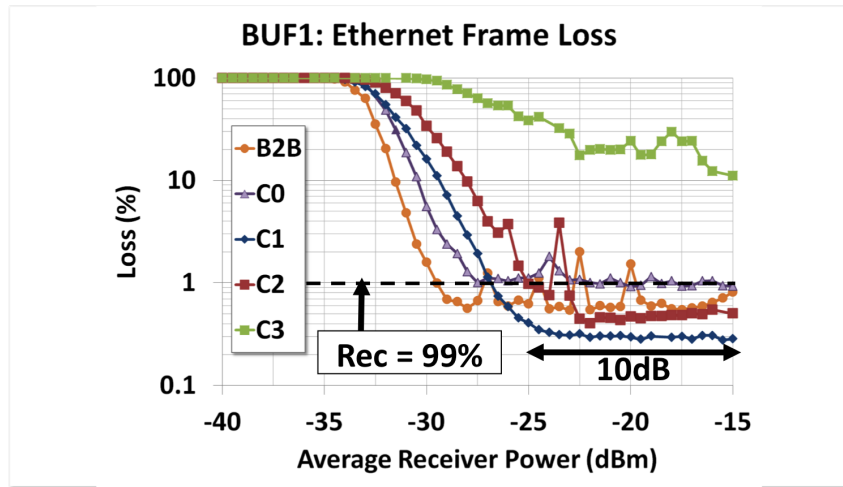
(a)
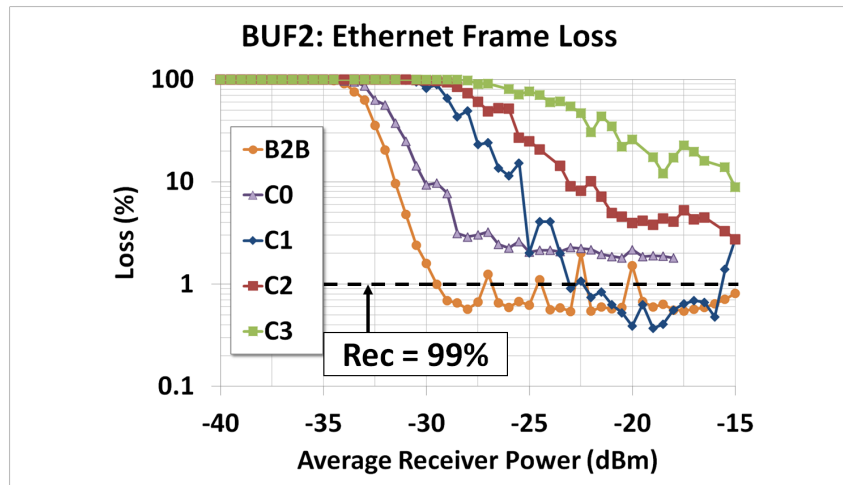
(b)

Figure 9.5: *(a) Averaged Ethernet frame recovery results of end-to-end adaptation as a function of buffer circulations for BUF1 and (b) BUF2.*

9.2.1.1 Results and Discussion

The end-to-end characterization results for buffers BUF1 and BUF2 are shown in Figure 9.5. The plots show the percentage of Ethernet frames lost as a function of receiver power and buffer circulations (C0-C3) where

179

the back-to-back (B2B) measurement was obtained by removing the buffer from the optical link. The results in (a) demonstrate BUF1 storage times up to 128ns (two circulations) at a frame loss percentage below 1 and 0.4% over a dynamic receiver range greater than 10 and 5dB respectively. Moreover, a storage time of 192ns (three circulations) is possible if one is willing to incur a 10% penalty in frame loss. An improvement in performance can be viewed in buffered optical packets, in contrast to those that bypass the fiber delay, when evaluated at higher levels of receiver power. At -20dBm, for example, the frame loss percentage corresponding to zero circulations (C0) appears to be clamped at 1%. For identical received powers, frame loss floors centered about 0.3 and 0.5% are obtained for one (C1) and two (C2) circulations, respectively. This is a direct result of net gain present within the re-circulating delay. The results in (b) show the BUF2 end-to-end performance, where a storage of 64ns (one circulation) is achieved at a frame loss rate below 1%. Furthermore, storage times of 128 and 192ns are achieved at loss percentages below 4 and 8%, respectively.

The power penalty performance, shown in Figure 9.6, was evaluated at the points of 70, 80, 90, and 99% of frame recovery for buffers BUF1 (a) and BUF2 (b). If the traces corresponding to three circulations (C3) are excluded, it is clear to see that the variability in performance is significantly less for BUF1 when compared to BUF2. The overall distribution of power penalty for BUF1 appears to be within 3-5dB, while the variance is on the order of 6-9dB for BUF2. This discrepancy can be attributed to uncertainties in packaging procedures. The coupling losses in BUF2 were measured to be significantly higher than BUF1. Additionally, BUF2 suffered from mechanical anomalies on the driver PCB board that affected the quality of the RF switching signals.
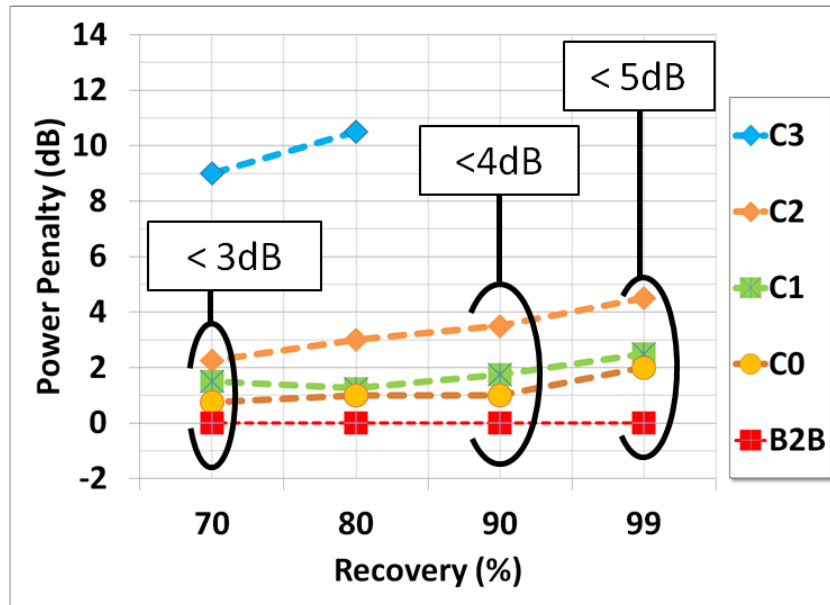
(a)
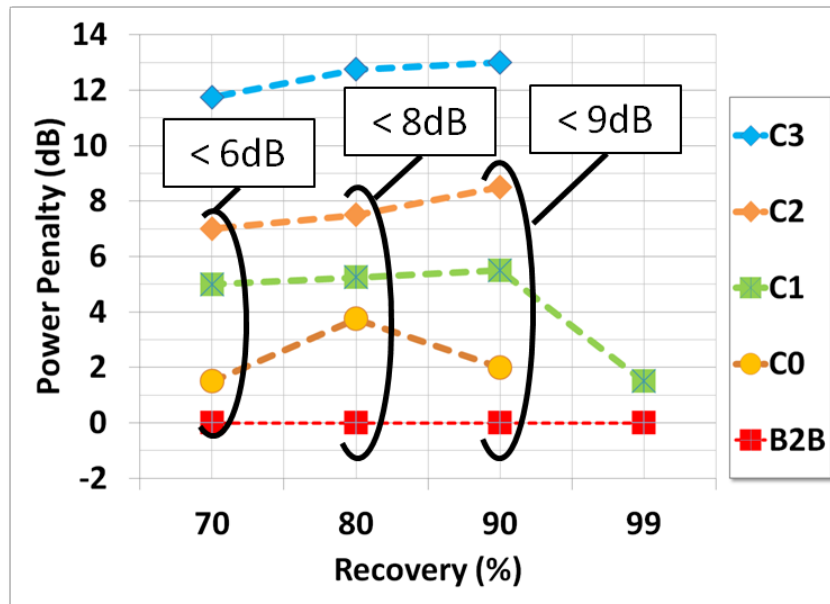
(b)

Figure 9.6: *(a) Power penalty measurements evaluated at different recovery percentages as a function of buffer circulations for BUF1 and (b) BUF2.*

181
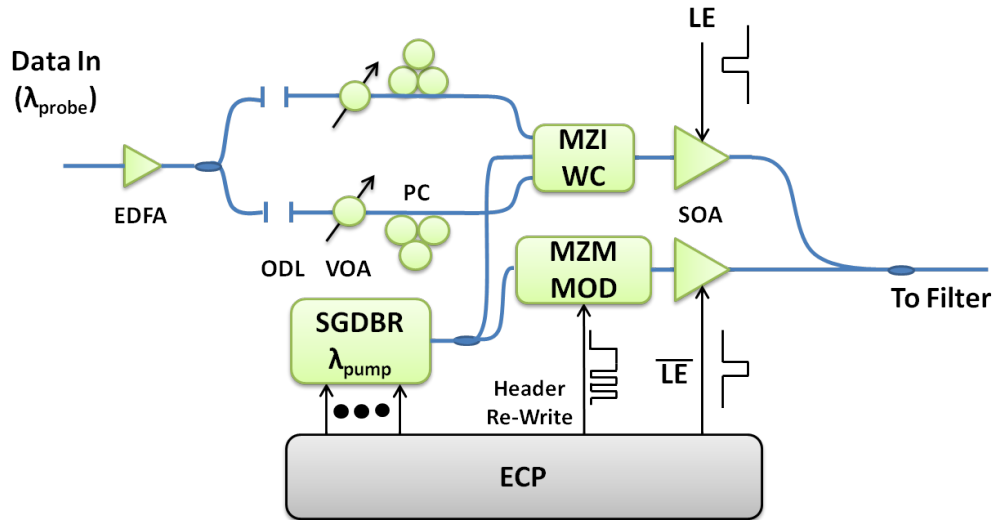
Figure 9.7: *Schematic diagram of optical packet forwarding plane with header erasure and rewrite.*

9.3 Packet Forwarding Plane

Figure 9.7 shows a schematic diagram describing the discrete implementation of the Optical Packet Forwarding Planes FWD1 and FWD2. The physical layout can be separated into two functional sections: wavelength conversion and header rewrite. Wavelength conversion is performed via a commercially available Mach-Zehnder interferometer wavelength converter (MZI-WC) connected in differential configuration to achieve 40Gb/s RZ operation. The probe inputs (λ_{PROBE}) consist of incoming 40Gb/s labeled packets that are amplified and evenly distributed between the MZI-WC differential inputs. The input arms are separated by about 10-12ps, corresponding to the width of input pulses, using free-space optical delay lines (ODL). The amplitude of each input is adjusted through a variable optical attenuator (VOA), while the input polarizations are rotated towards a TE polarization using polarization controllers (PC). The pump input (λ_{PUMP}) is sourced from a fast-switching,

182

widely-tunable, sampled-grating distributed Bragg reflector (SG-DBR) laser. The SG-DBR laser was designed to achieve nanosecond-scale switching over a 20nm span with more than 40dB of side-mode suppression ratio (SMSR) [3].

The header rewrite section is performed through an externally driven LiNbO$_3$ Mach-Zehnder modulator (MZM). The pump output from the SG-DBR laser is evenly split between the MZI-WC and the header rewrite inputs using a 3dB coupler. The Electronic Channel Processor (ECP) drives the MZM with a 10Gb/s NRZ signal in order to encode the optical header onto the pump output (λ_{PUMP}). A separate 3dB coupler is then used to combine the newly written 10Gb/s labeled header with the newly λ-converted 40Gb/s payload exiting the MZI-WC. For this process to be successful, it is crucial to obtain a high extinction ratio (ER) between the low-speed header and the high-speed payload. However, the measured ER of the λ-converted payloads is about 9-12dB, while the ER of the newly written 10Gb/s header is about 14-17dB.

The path leading from the SG-DBR pump to the point where newly written headers are combined with λ-converted payloads is essentially a fiber-based MZI that is phase unstable. Combining the header-payload pairs, which exhibit ERs below 20dB, would result in amplitude fluctuations that are detrimental to system performance. To mitigate this issue, a pair of highly absorbent bulk SOAs are inserted before the signal coupling point. The SOA at the output of the MZI-WC is driven with a label-erase (LE) signal that is constantly enabled, except when one needs to erase the previous (λ_{PROBE}) header. The SOA at the output of the header rewrite section is driven with a complimentary label-erase ($\overline{LE}$) that is enabled only in the presence of the newly written header (λ_{PUMP}). Since the SOAs are never enabled simultaneously, the SOA on-off extinction (30-40dB) properties

183
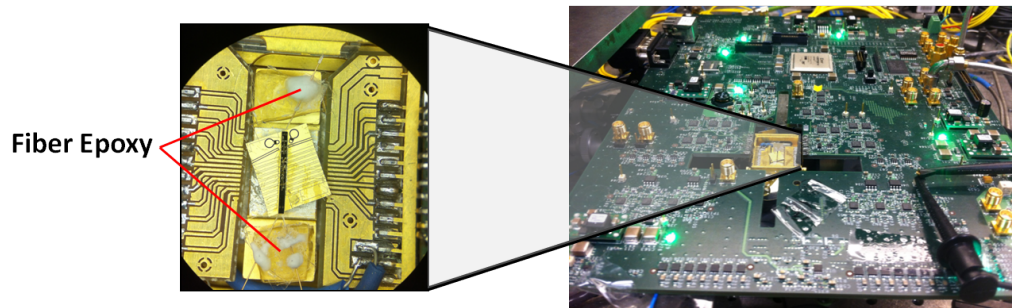
Figure 9.8: *Packaged SG-DBR laser mounted on custom, FPGA-based driver board for speed and flexibility.*

ensure low header-to-payload crosstalk is achieved when combined. Once the header and payload are combined, they are forwarded to the arrayed waveguide grating (AWG) to complete the routing function.

9.3.1 Implementation and Initial Characterizations

9.3.1.1 Device Packaging

The widely-tunable SG-DBR laser was placed in a butterfly package and driven with a custom FPGA-based controller board to simultaneously enable thermal stability and fast wavelength switching. Figure 9.8 shows an image of the FPGA-based laser controller board with an inset magnifying the packaged tunable laser. Each laser sub-component is driven with the digital-to-analog converter (DAC) hierarchy shown in Figure 9.9(a). The back mirror (BM), front mirror (FM), and phase (P) sections of the laser are driven using 10-bit switching DACs. These DACs are used as tunable current sources capable of fast (<10ns) switching speeds at maximum current levels of 40mA. Alternatively, the laser gain (G) and a booster SOA (bSOA) are driven with high-current 12-bit DACs. These DACs are used as DC current sources with maximum levels of 150mA. The FPGA-based DAC
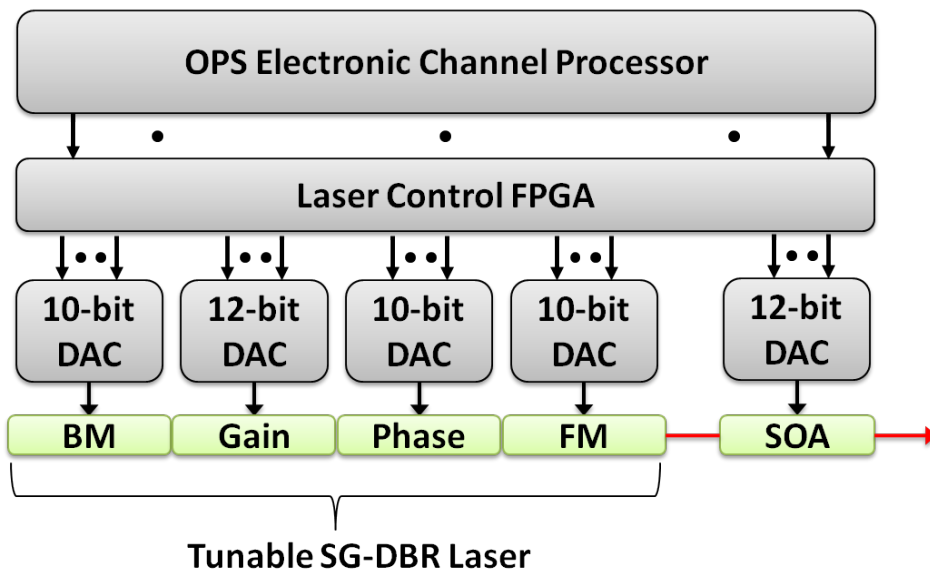
(a)
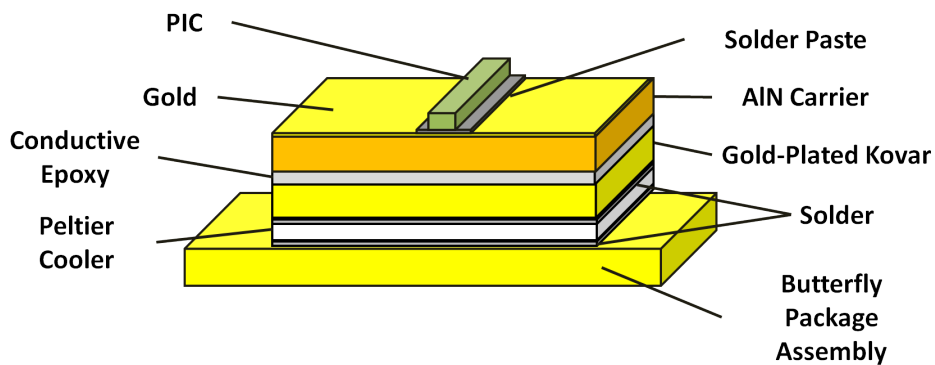
(b)

Figure 9.9: *(a) Hierarchy of FPGA-based tunable laser control (BM: back mirror, FM: front mirror, DAC: digital-to-analog converter). (b) Packaging stack used to obtain thermal stability during operation of tunable laser.*

hierarchy is configured and fine-tuned using a custom, multi-threaded ODR Configuration Tool that communicates to the ECP via an USB-to-RS232 bridge.

The diagram in (b) shows the packaging stack used to obtain thermal

185

stability during laser operation. The InP photonic integrated chip (PIC) was soldered onto an AlN carrier at 150°C. The carrier was then attached to a gold-coated Kovar block via a conductive epoxy layer that was cured at 110°C. The block of Kovar was then mounted onto an 18-pin butterfly package by inserting a solder-coated Peltier heating element between them at 80°C. A fanout printed circuit board (PCB) with gold traces was then attached to the pins of the butterfly package. Gold wire bonds were then used to ensure an electrical connection between the PIC, the carrier pads, the traces on the fanout PCB, and the butterfly package pins. Once thermal stability was verified, a lensed fiber was actively aligned to the output facet of the PIC and then held in place using UV-curing epoxy. The inset in Figure 9.8 shows a top-down view of the resulting package.

9.3.1.2 Tuning Range

Figure 9.10 shows optical spectrum analyzer (OSA) traces displaying the tuning ranges of the Optical Forwarding Planes FWD1 and FWD2. A tuning range of about 20nm (2.5THz) with a channel spacing of about 3.25nm (400GHz) is achieved with an SMSR greater than 40dB for all possible optical channels. A 5dB difference in coupled power was observed between FWD1 and FWD2, which is attributed to inconsistencies in packaging and coupling efficiencies.

The AWG (20°C) optical passband was measured by separately connecting an Erbium (Er^{3+}) ASE source at input ports 1 and 2 and observing each AWG output through an OSA. The AWG insertion loss was measured by taking the difference between the AWG-filtered spectral output and a back-to-back spectral measurement performed by removing the AWG from the optical link. Figure 9.10 shows the AWG pass band overlaid with the

186
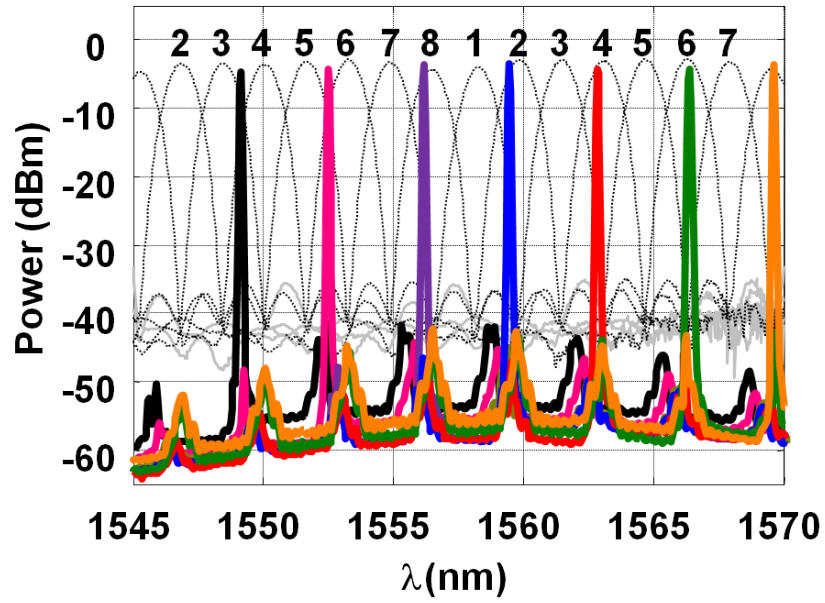
(a)
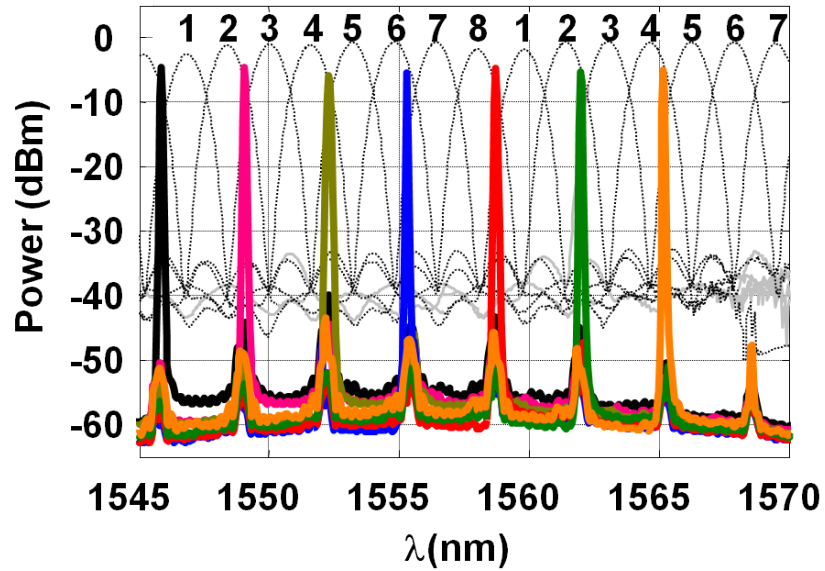
(b)

Figure 9.10: *(a) SG-DBR tuning range (solid) and AWG passband (dashed) alignment of Optical Packet Forwarding Planes FWD1 and (b) FWD2.*

187

aforementioned SG-DBR laser tuning ranges, where each output port is enumerated for simplicity. The figure in (a) shows the alignment of the AWG passband with the laser tuning range of FWD1, while (b) shows the passband alignment relative to FWD2. The excess loss in port-1 of the AWG is shown to be about 5dB, while port-2 only exhibits about 2dB. The laser output wavelengths are aligned relatively well with the AWG output ports that have even enumerations. FWD1 exhibits better alignment with the AWG passband when compared to FWD2 despite the fact that the phase section of each laser was used to fine-tune the alignment. However, the SG-DBR lasers did not possess anti-reflective (AR) coatings on the facets. Hence, a compromise was made between passband alignment and minimization of laser relative intensity noise (RIN). Table 9.1 shows the wavelengths that best matched the AWG passband in terms of alignment.

Table 9.1: *Wavelength values corresponding to the AWG output ports of forwarding planes FWD1 and FWD2.*

FWD1		FWD2	
AWG Port #	λ(nm)	AWG Port #	λ(nm)
2	1559.95	2	1562.15
4	1563.10	4	1565.55
6	1566.75	6	1555.35
8	1556.60	8	1559.00

9.3.1.3 Switching Dynamics
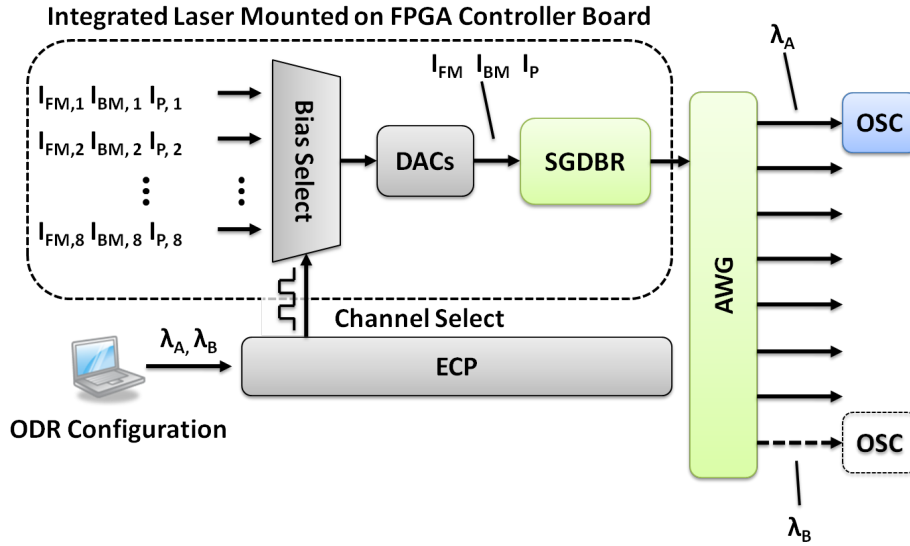
Figure 9.11: *SG-DBR laser switching measurement setup.*

The SG-DBR wavelength switching speeds were measured using the setup displayed in Figure 9.11. Each of the eight output wavelengths is mapped to one of the AWG outputs and a channel enumeration is determined based on the AWG output port. Laser bias points for the FM, BM, and phase sections are predetermined for each channel ($[I_{FM1}, I_{BM1}, I_{P1}], \ldots, [I_{FM8}, I_{BM8}, I_{P8}]$) and an electronic lookup table is generated based on those configurations. Each set of channel biases is connected to the input of a bias selector that maps one set of bias points the set of biases (I_{FM}, I_{BM}, I_P) used to drive the SG-DBR laser. The custom, multi-threaded ODR Configuration Tool is then used to select between two possible wavelength states: λA and λB ($\lambda A, \lambda B \in \lambda 1, \lambda 2, \ldots, \lambda 8$ and $\lambda A \neq \lambda B$), where λN represents the wavelength channel exiting AWG port N. The two state settings are forwarded to the ECP to generate a Channel Select signal that corresponds to one of the two λ-states and is used as the data-select input for the bias selector. In other

189
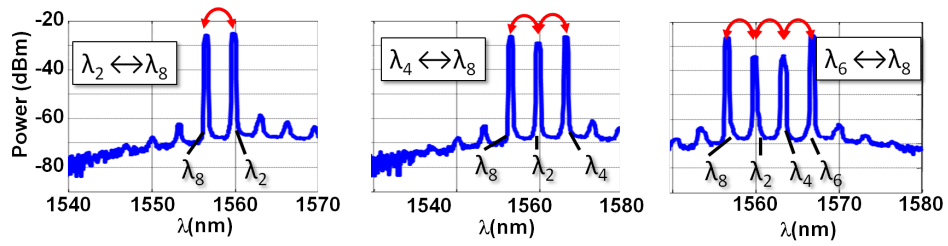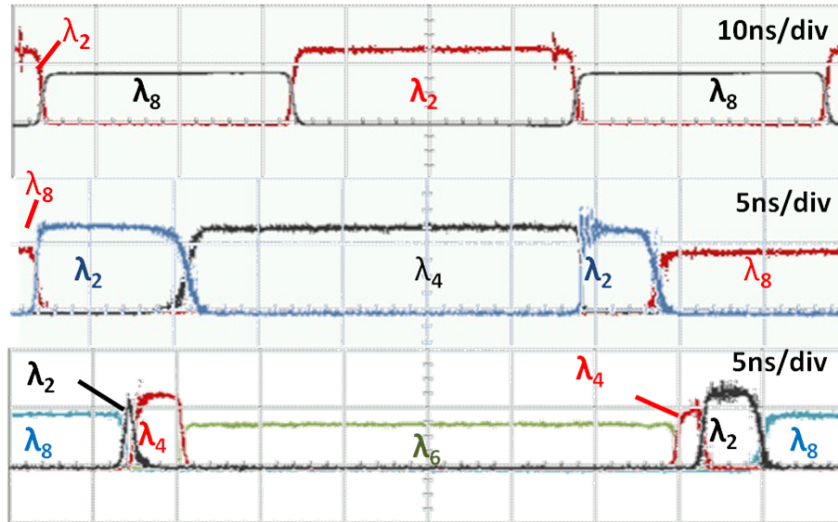
Figure 9.12: *(a) OSA traces showing FWD1 switching dynamics between several wavelength pairs and their intermediate states. (b) Oscilloscope traces used to determine FWD1 switching speeds between wavelength pairs $\lambda_2 \leftrightarrow \lambda_8$ (top), $\lambda_4 \leftrightarrow \lambda_8$ (middle), and $\lambda_6 \leftrightarrow \lambda_8$ (bottom).*

words, the Channel Select signal oscillates between λA and λB at a frequency equal to 15.625MHz with a 50% duty cycle. The selected bias points are then sent to the DAC current sources that are used to bias the SG-DBR laser. The output of the laser is sent through an AWG where each laser wavelength is spatially separated and monitored via an oscilloscope one at a time.

The optical spectrum analyzer (OSA) traces in Figure 9.12 (a) are used to

190

evaluate the switching dynamics of the SG-DBR tunable lasers. The left-most trace shows the switching dynamics between the neighboring channels $\lambda 8$ (1556.60nm) and $\lambda 2$ (1559.95nm). The middle trace not only demonstrates switching between the two non-adjacent channels $\lambda 8$ and $\lambda 4$ (1563.10nm), but also shows an intermediate transition to $\lambda 2$. The right-most trace shows the laser output iterating through intermediate wavelengths $\lambda 2$ and $\lambda 4$ when switching between non-adjacent channels $\lambda 8$ and $\lambda 6$ (1566.75nm). The switching times were measured, using the oscilloscope traces in (b), from the 50% point of the λA on-to-off transition to the 50% point of the λB off-to-on transition (and vice-versa).

The results from the laser wavelength switching measurements are shown in Figure 9.13. The switching time is shown as a function of spectral distance of four output channels. The spectral distance is defined as the difference between the two λ-states chosen to measure wavelength switching speeds ($|\lambda A - \lambda B|$). Only four wavelengths were considered as the SG-DBR lasing channels only coincided with the even-numbered output ports of the AWG. The switching results for FWD1 and FWD2, shown in (a) and (b) respectively, demonstrate nanosecond-scale switching times below 10ns. Switching wavelengths between adjacent channels generally requires relatively smaller index perturbations within the Bragg mirrors resulting in increased switching times for non-adjacent channels.
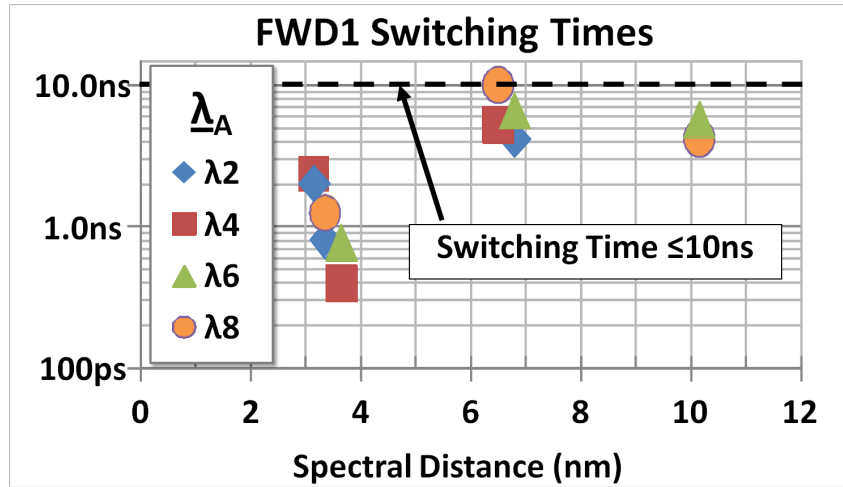
(a)
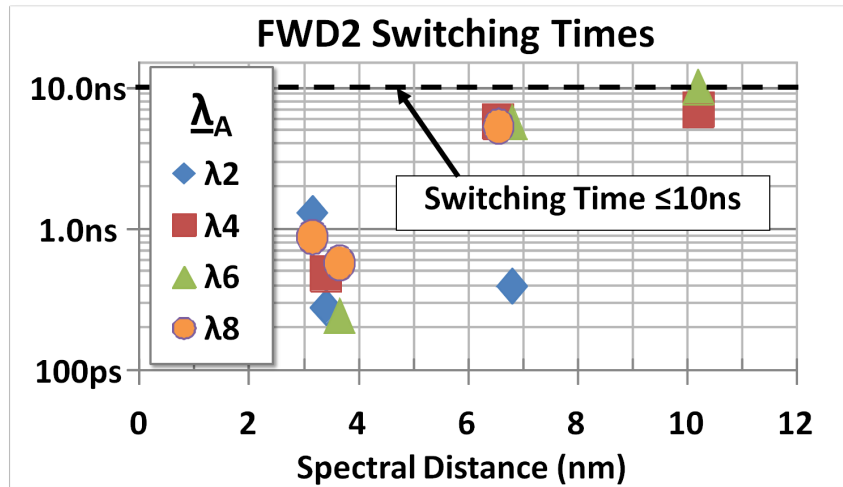
(b)

Figure 9.13: *(a) Measured SG-DBR laser switching speeds of FWD1 and (b) FWD2.*

192
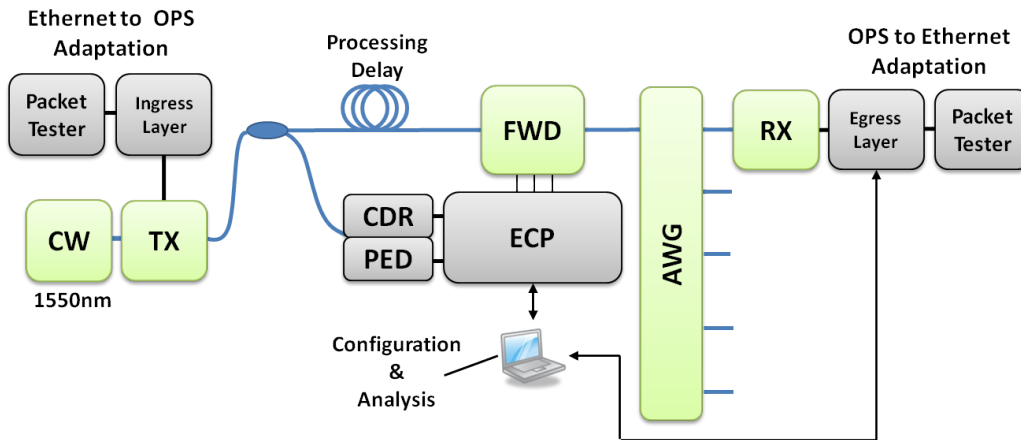
Figure 9.14: *Experimental setup used to characterize the optical packet forwarding plane.*

9.3.2 End-to-End Characterization

The setup used to experimentally characterize Packet Forwarding Planes FWD1 and FWD2 is shown schematically in Figure 9.14. Each forwarding plane is placed between the end-to-end adaptation setup characterized in Section 7.1. Adapted packets from the Ingress Layer are evenly split between the Optical Path and the Electronic Control Path using a 3dB splitter. The optical path consists of about 300ns processing delay followed that leads to the packet forwarding plane (FWD). The FWD is responsible for performing payload wavelength conversion along with header erasure and rewrite. Once labeled packets are converted to the desired output wavelength, they are forwarded to an arrayed waveguide grating (AWG) where routing function is executed. Routed packets are then sent to the Egress Layer where they are converted from labeled packets into Ethernet frames. Adapted frames are then transmitted to a commercial packet tester for frame recovery measurements. Recovery measurements are performed on each of the AWG output

193

channels one at a time. The electronic control path leads to the clock and data recovery (CDR) stage that is used to detect and recover the 10Gb/s Optical header. Header validation is performed and the 10-bit label is used to designate an output wavelength. The temporal location of the 40Gb/s payload is monitored within 6.4ns of accuracy using the payload envelope detection (PED) stage. The recovered header and PED signal is then forwarded to the Electronic Channel Processor (ECP) that is used to generate the control signals for the forwarding plane. The forwarding plane configurations and the signal fine-tuning is set using a custom, multi-threaded ODR Configuration Tool that communicates to the ECP via a USB-to-RS232 serial connection.

9.3.2.1 Results and Discussion

End-to-end packet recovery results for optical forwarding planes FWD1 and FWD2 are shown in Figure 9.15. The average Ethernet frame recovery percentage is shown as a function of average receiver power for each output wavelength of the SG-DBR laser. A frame recovery percentage greater than 99.9% was successfully achieved by both forwarding planes across multiple output wavelengths.

Figure 9.16 shows the power penalty of forwarding planes FWD1 (a) and FWD2 (b) evaluated at the 70, 80, 90, and 99% recovery points relative to the back-to-back (B2B) measurement. The B2B curve was taken by removing the FWD stage while leaving the AWG in place. The curves show that the performance of both forwarding planes is relatively comparable. At higher received powers (recovery $\geq 99\%$), the variance in power penalty is within 10 and 11dB for FWD1 and FWD2 respectively. At lower receiver powers (recovery $< 99\%$), the variabilities are within 6-7dB for both forwarding planes

194
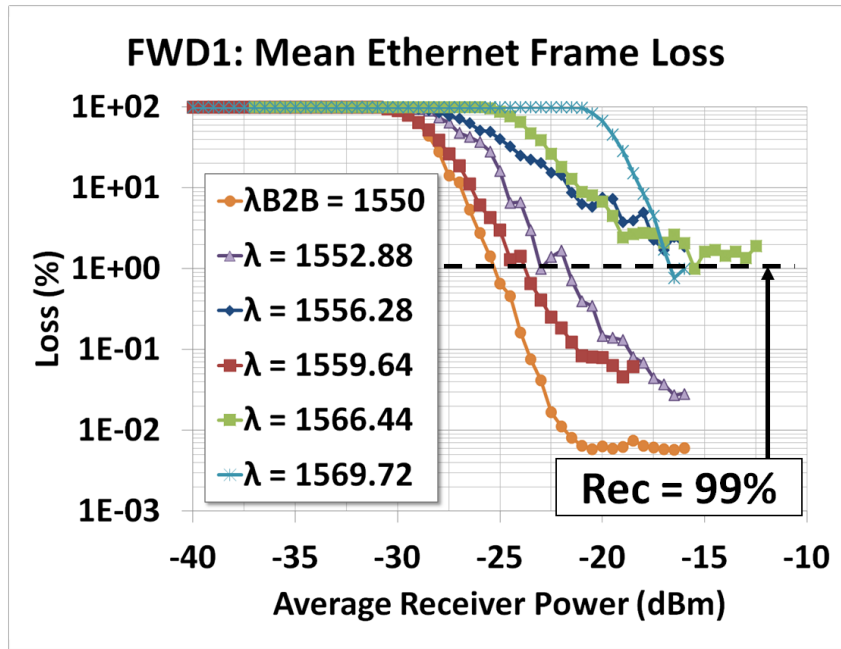
(a)
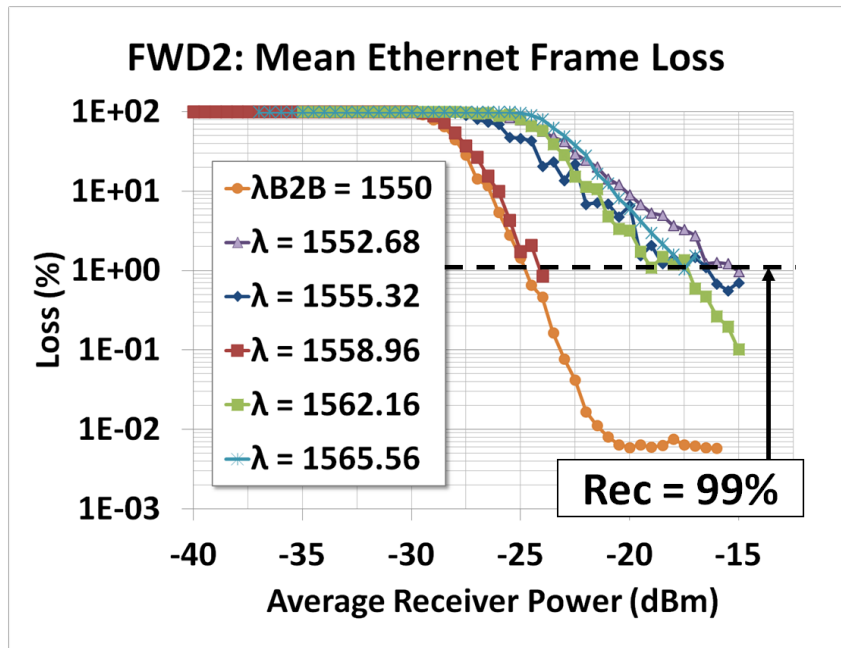
(b)

Figure 9.15: *(a) Averaged Ethernet frame recovery performance of optical packet forwarding planes FWD1 and (b) FWD2.*

195
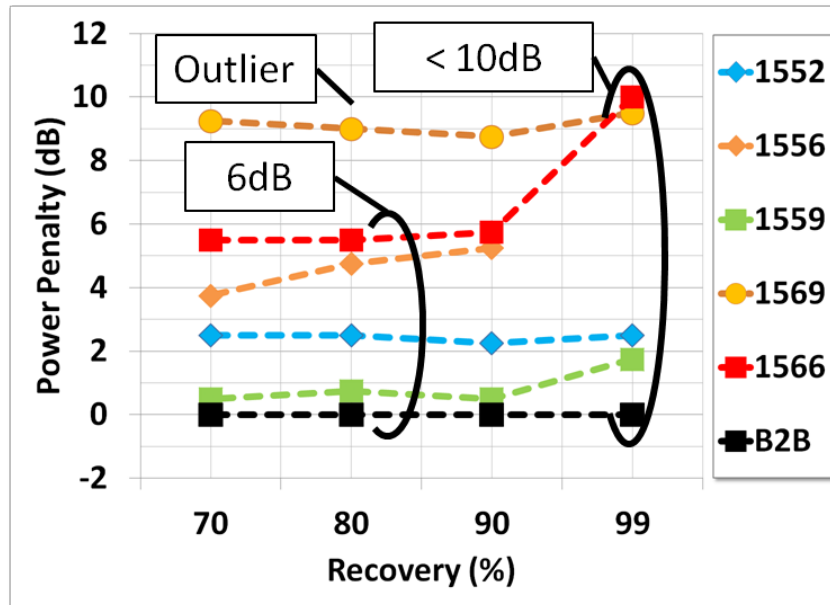
(a)
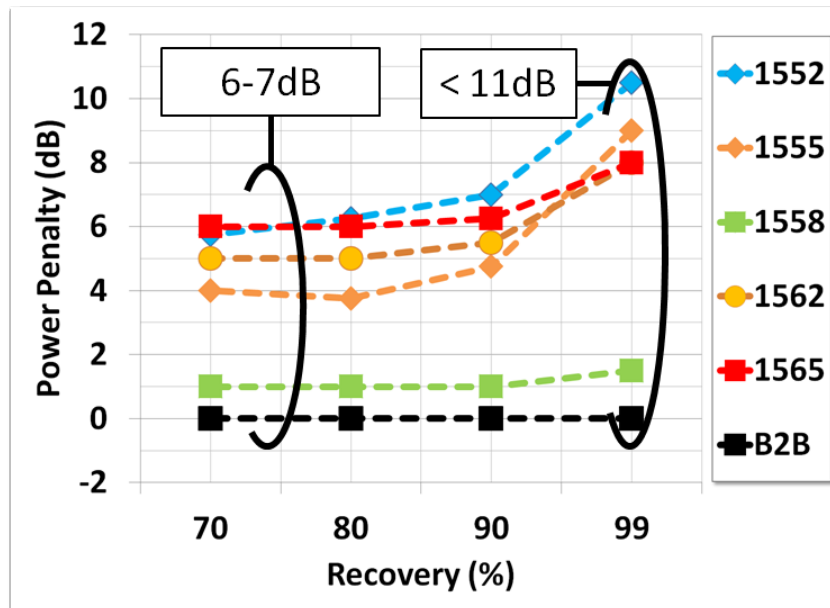
(b)

Figure 9.16: *(a) Averaged Ethernet frame recovery performance of optical packet forwarding planes FWD1 and (b) FWD2.*

as long as the outlier in FWD1 ($\lambda = 1569$nm) is excluded. With the exception of $\lambda = 1552$nm, mutual wavelength (±1nm) pairs between FWD1 and FWD2 appear to exhibit comparable power penalty performances. Additionally, peak performance seems to correspond to wavelength channels that are near the lasing peak of its respective SG-DBR laser (~1560nm).

The observed performance variability is believed to have resulted from physical layer limitations that are artifacts of the discrete implementation. The interconnection between the SG-DBR pump and the MZI-WC consisted of couplers, fiber patch cords, and a polarization controller (PC). The wavelength dependence of the PC contributed some inter-channel variability as it was difficult to manually optimize polarization during operation. Furthermore, the MZI-WC was operated in DC bias when wavelength-aware adaptive biasing of the MZI phase section is typically needed. Yet another contribution of variability is believed to have stemmed from nonuniform, inter-channel RIN. The RIN measured along the tuning range of a tunable SG-DBR laser has been shown to be rather uniform with the exception of peaks near mode hopping boundaries [4]. However, the lasers used in this demonstration were without AR facet coatings, which made it challenging to locate lasing modes that simultaneously met the requirements of high SMSR ($>$40dB), low RIN ($\leq$-130dB/Hz), and maximized alignment with AWG passband.

9.4 Optical 3R Regeneration

Figure 9.17 (a) schematically displays how optical regeneration, reshaping and re-timing is performed within the OPS router in this work. The 3R signal regenerator consists of an EDFA for re-amplification, an MZI-WC for re-shaping, and an optical clock recovery (OCR) stage for re-timing. The
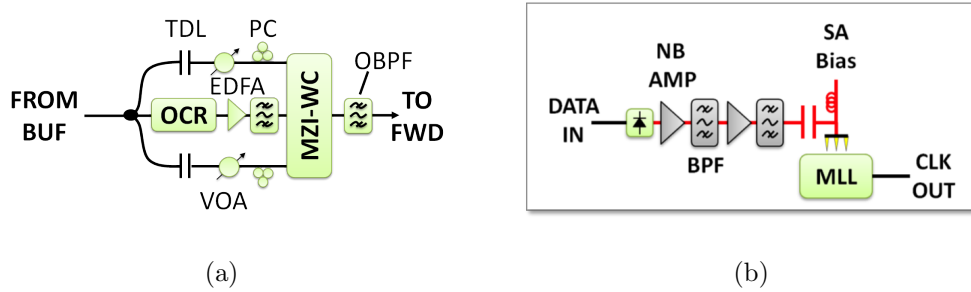
Figure 9.17: *(a) Optical regeneration, reshaping, and re-timing setup and (b) optical clock recovery via hybrid locking circuit.*

degraded signal from the optical packet buffer (BUF) is evenly split between the differential inputs of the MZI-WC and the input of the OCR stage by utilizing a 3dB coupler. The OCR output consists of a re-timed pulse train that is recovered from incoming packets, which is then amplified and filtered via an EDFA and an optical bandpass filter (OBPF), respectively. The pulse train is then used as the pump input to the MZI-WC, which is configured with differential input signaling to enable 40Gb/s return-to-zero (RZ) fixed wavelength conversion (FWC). The differential inputs are configured to have a relative 10-12ps time delay between them via tunable delay lines (TDL). Furthermore, the amplitude in each arm is manually adjusted using a variable optical attenuator (VOA) while the polarization is TE-rotated using a pair of polarization controllers (PC). The MZI-WC is effectively utilized as an optical gate for the purpose of transferring the amplitude information from the input signal onto the recovered pulse train. Its sinusoidal transfer characteristics also provide a re-shaping functionality to minimize the amount of noise and jitter transferred onto the re-timed pulse train. Finally, an OBPF is placed at the output of the MZI-WC to suppress the differential inputs while transferring the regenerated, λ-converted packets.

Optical clock recovery is performed by utilizing an integrated 40GHz
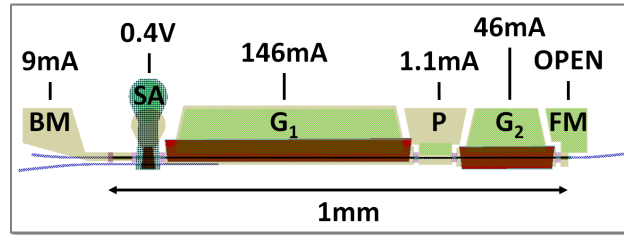
(a)
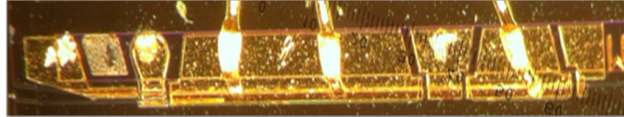
(b)

Figure 9.18: *(a) Mask layout and bias levels of 40GHz integrated mode-locked laser used for optical clock recovery (BM: back mirror, FM: front mirror, SA: saturable absorber, G: gain, P: phase). (b) Probed device in clock recovery setup.*

mode-locked laser (MLL) in conjunction with the hybrid locking circuit shown in (b). An optoelectronic (OE) conversion of the 40Gb/s labeled optical packets is first performed via a 50GHz photo-detector. Its output is then forwarded through a set of two filter-amplifier stages consisting of a 40GHz narrow-band amplifier (NB-AMP) and a 40GHz bandpass filter (BPF). Their output is electronically coupled to the saturable absorber (SA) of the MLL via a 50GHz ground-signal-ground (GSG) RF probe.

9.4.1 Implementation and Initial Characterizations

The mask layout and the bias settings of the integrated MLL, used in the OCR circuit, are shown in Figure 9.18. The device is of a self-colliding pulse design based on previous implementations that consisted of a 1mm long cavity designed for 38-40GHz operation [5]. The back mirror (BM)

199

and front mirror (FM) are distributed Bragg reflectors (DBR) of lengths $41.5\mu m$ ($R_{BM} = 90\%$) and $15.5\mu m$ ($R_{FM} = 38\%$) respectively. The gain sections (G_1 and G_2) are $500\mu m$ and $220\mu m$, the P phase section is $73\mu m$, and the saturable absorber (SA) is $50\mu m$ long. Anti-reflection (AR) coatings and angled waveguides are utilized at the facets of the device to minimize reflections from the semiconductor-to-air interfaces. The power coupled via a lensed fiber was measured to be -4dBm.

Signal re-shaping requires that the recovered pulse train exhibit characteristics of high extinction ratio and large modulation depth. A series of DC characterizations were carried on the MLL to empirically determine an operating range that would be conducive to re-shaping. A light versus current (LI) measurement was performed at various SA bias points, and the results are displayed in Figure 9.19. The plots show a contour map corresponding to the coupled optical power as a function of laser gain and absorber bias. Even at slight reverse biases, the MLL requires large amounts of injected current ($\frac{G_1}{2}$, $G_2 \sim$ 80-100mA) to obtain an output power on the order of 1mW. It is not until the SA is biased to at least 0.3V, that one begins to observe a regime where the laser can be operated under safe bias conditions ($\frac{G_1}{2}$, $G_2 \sim$50-70mA) in order to obtain 0.5-1mW out of the front facet. One can continue to forward bias the absorber up to 1V, but this would result in a pulse train exhibiting larger widths, lower extinction ratio, and a diminished modulation depth. If the SA were to be modulated with a sinusoidal function, one would need to set the DC offset such that the null point would bias the SA slightly below transparency. The LI results suggest that driving the SA with a $2V_{PP}$ signal centered about 0.3V, while biasing the laser gain sections at about 50 and 140mA, could potentially result in a pulse train with the desired signal qualities.
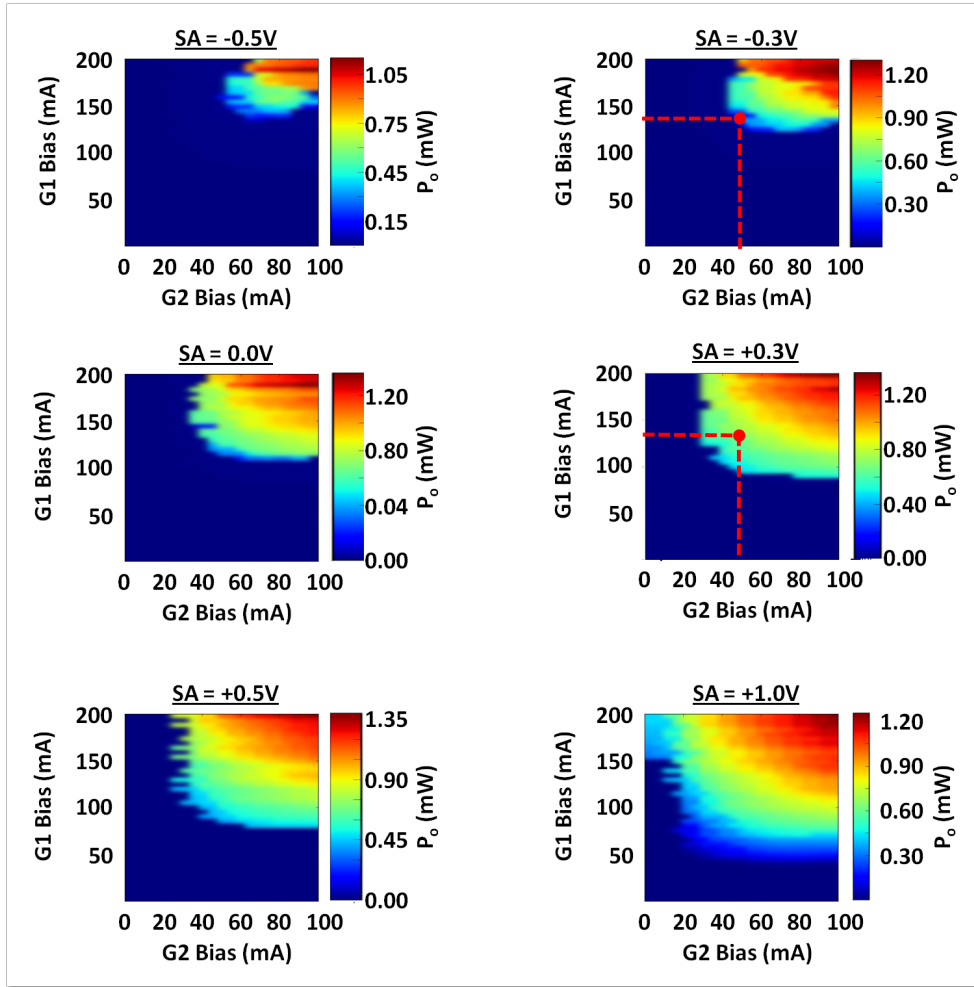
Figure 9.19: *LI measurements of integrated 40GHz mode-locked laser taken at various saturable absorber (SA) bias points.*

9.4.2 Verification of 3R Functionality

Figure 9.20 shows the experimental setup used to verify amplification, re-shaping, and re-timing characteristics of the 3R regeneration stage. A $2^{15}-1$ pseudo-random bit sequence (PRBS) is encoded onto the 1560nm output of a CW laser by driving an optical transmitter using a 40Gb/s bit pattern generator (BPG). The optical transmitter consists of two Mach-Zehnder modulators (MZM) connected in tandem. The first MZM is driven with the non-return-

201
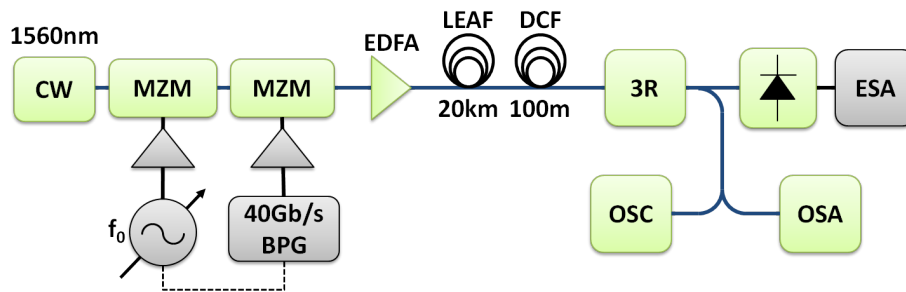
Figure 9.20: *Experimental setup used for verification of 3R regeneration (BPG: bit patter generator, OSC: oscilloscope, OSA: optical spectrum analyzer, ESA: electrical spectrum analyzer)*

to-zero (NRZ) output from the BPG while the second MZM is driven by a clock synthesizer (CLK) operating at a frequency (f_0) that is configurable between 38GHz and 40GHz. The same clock output is used as the clock input for the BPG to allow the data rate to also vary between 38Gb/s and 40Gb/s. The signal quality of the optical PRBS data stream is then degraded using 20km of large effective area fiber (LEAF) followed by 100m of dispersion compensating fiber (DCF). The degraded PRBS stream is then sent through the 3R regeneration stage. The regenerated signals are then transmitted to an oscilloscope (OSC) to verify amplification and reshaping, while a 50GHz detector followed by an electrical spectrum analyzer (ESA) is used to verify re-timing via single side-band (SSB) phase noise measurements.

The MLL was found to be passively mode-locked at a free-running frequency ($f`_R$) of 39.8182GHz by utilizing an ESA with 30kHz of resolution bandwidth (RBW). The setup's clock synthesizer frequency (f_0) was then set to the free-running frequency of the MLL to ensure that the data rate of the PRBS stream was within its locking range. Figure 9.21 shows ESA traces of the MLL taken during free-running (red) and hybrid locking (blue)
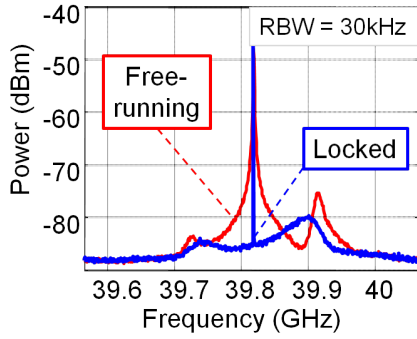
Figure 9.21: *ESA traces of mode-locked laser during free-running and hybrid locking operation.*
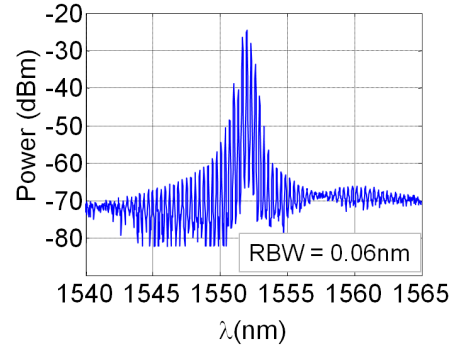
Figure 9.22: *OSA trace of mode-locked laser during hybrid locking operation.*

modes of operation. The wide frequency tone about f_R suggests that the passively locked laser exhibits a significant amount of frequency and phase noise. When locked, the f_R tone increases in amplitude and is narrowed resulting in a frequency-stable pulse train with significantly lower phase noise. The OSA trace in Figure 9.22 shows the MLL during locked operation and exhibits an optical bandwidth, evaluated at the 10dB point, of 0.6nm ($\sim$75GHz) centered about 1552nm.

9.4.2.1   Re-Amplification and Re-Shaping

Oscilloscope traces of the PRBS data were taken at various stages of 3R regeneration and are on display in Figure 9.23. The trace in (a) shows the input signal after having traveled through 20km of delay. One can see that the modulation amplitude has degraded to a value close to 1mW while the extinction ratio (ER) is about 11.68dB. The middle trace (b) presents the clock that has been recovered from the input signal using the hybrid mode locking circuit. The clock pulses have increased to a modulation depth of about 2mW while the ER has decreased to 6dB. Finally, the right-most
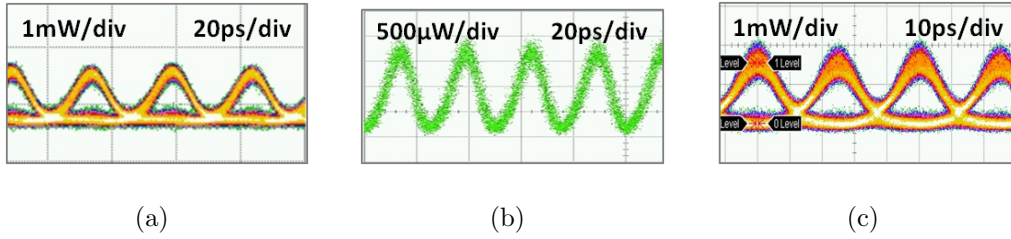
Figure 9.23: *(a) Oscilloscope traces of input signal, (b) recovered optical clock, and (c) reshaped & re-timed pulses*

trace (c) shows the output of the MZI-WC. It is clear that 1R amplification is achieved as the increase in amplitude modulation of 1mW is maintained through the λ-conversion. Though ER is still 6dB, the zero-level noise present in the input signal appears to be slightly suppressed at the output of the MZI-WC. Additionally, the amplitude noise is larger at the output compared to the input signal. Therefore, it is difficult to verify successful 2R regeneration without further characterizations.

There are two classes of signal regenerators: those that improve signal quality with a single pass, and those that degrade an input signal during the first pass, but minimize noise accumulation during subsequent passes [6]. The regenerator in this work may be filed under the latter classification, and requires λ-preserving, re-circulating loop characterizations to verify reshaping. This requires the insertion of a second λ-converter stage to facilitate λ-preservation [7].

The 2R performance in this work is limited by the degradation of extinction ratio within the MZI-WC and the OCR circuit. Here, the MZI-WC performance is shown to be limited to 6-9dB caused by imbalance in the MZI arms. This imbalance leads to non-optimal constructive and destructive interference, which limits the ER and re-shaping characteristics of the

MZI optical gate. Pulse trains obtained from the OCR circuit exhibited ER of about 6-7dB limited by the hybrid locking efficiency. Typically, the SA section of an MLL is operated at a slight reverse bias to to obtain narrow pulse shapes with high extinction. In this work, the SA required a positive bias of about 0.5V to maintain stable operation while reducing SA absorption. Furthermore, RF drive power is not efficiently delivered to the SA as the integrated MLL is implemented without impedance-matched contacts or top-side n-contacts. Therefore, the combination of inefficient delivery of RF power and limited absorption modulation resulted in wide sinusoidal pulses (8-10ps) with low extinction ratios (6-7dB).

9.4.2.2 Re-Timing

To verify re-timing, SSB phase noise measurements were performed on the input signal, the recovered optical clock, and the output of the 3R stage. The measurements were carried out over the offset frequency span of 1kHz to 100MHz where the RF signal fell below the ESA noise floor. The results in Figure 9.24 show that the phase noise for all three signals is relatively similar for lower offset frequencies resulting, while the signals are divergent at offsets beyond 1MHz. Hence, any phase noise present in the input signal within the 1MHz bandwidth will be transferred onto the recovered pulses. Integrating over the phase noise, from an offset frequency of 1kHz to 100MHz, produces a root-mean-squared (RMS) jitter value of 1.077ps and 0.8273ps for the degraded signal and the regenerated signals, respectively.

The re-timing performance shown in this work is limited by the jitter transferred observed within the MZI-WC and the hybrid-locked MLL. A jitter transfer bandwidth of about 1MHz is observed in the integrated MLL, which allows input phase noise within that offset frequency bandwidth to be
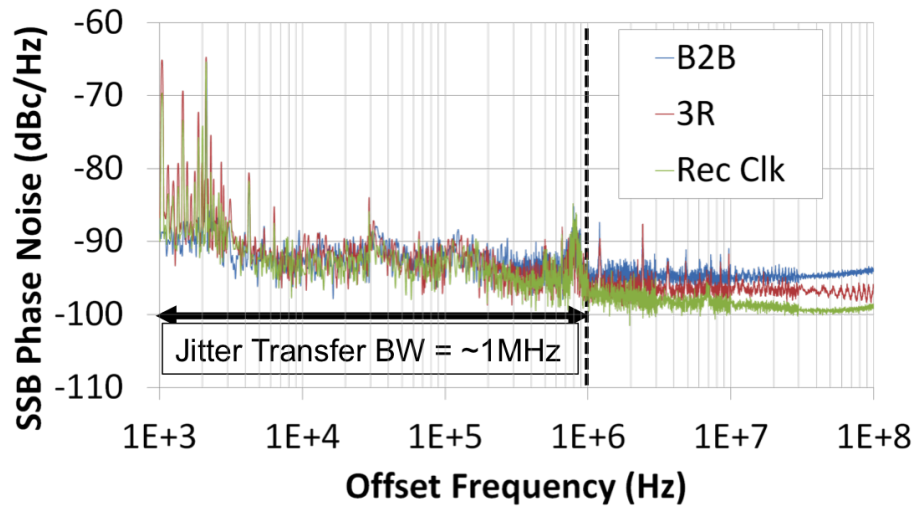
Figure 9.24: *Single side band phase noise measurement results of the back-to-back and regenerated signals along with the recovered optical clock*

carried over to the recovered pulses. The timing MLL characteristics were observed to be extremely susceptible to optical feedback. The phase and amplitude of the recovered pulse train was drastically affected by reflections from the lensed fiber. In fact, operation was most stable when the lensed fiber was laterally offset from optimal coupling, which would drift over time. It is believed that a significant amount of jitter is transferred from the MZI-WC as its gating window size is comparable to the pulse widths observed in the OCR stage. Ideally, the re-timed pulse width should be narrow enough to fit comfortably within the jittery gating window. As previously discussed, the OCR pulse widths in this work are approximately 8-10ns wide, which are comparable to the MZI-WC gating window size of 10-12ps.
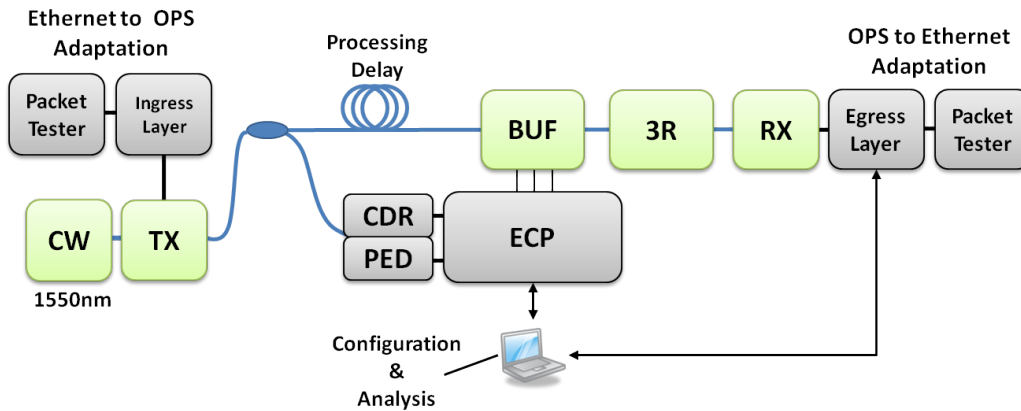
Figure 9.25: *Experimental setup used to characterize regenerative properties of the ODR*

9.4.3 End-to-End Characterization

Having successfully characterized the regenerative properties and limitations, the 3R stage is inserted into the ODR setup to evaluate end-to-end adaptation performance. Figure 9.25 shows the experimental setup used to characterize the regenerative properties of the 3R Regeneration stage when inserted into the ODR. The optical packet buffer BUF1 from Section 9.2 and the 3R Regeneration stage are placed between the interoperability setup shown in Section 7.1. Adapted packets from the Ingress Layer are distorted by sending them through 20km of LEAF fiber and then through 100m of DCF. The distorted packets are then evenly split between the optical data path and the electronic control path using a 3dB splitter. The optical path consists of a 300ns processing delay followed by a polarization controller leading to the optical packet buffer (BUF). The buffer is configured to a zero-circulation state by allowing packets to bypass the re-circulating fiber delay. Packets are then sent to the 3R regeneration stage where amplification, reshaping, and re-timing is performed. Regenerated packets are then sent to the Egress

Layer where they are converted to Ethernet frames and forwarded to a commercial packet tester for CRC-based recovery measurements. The electronic control path leads to the clock and data recovery (CDR) stage that is used to detect and recover the 10Gb/s Optical header while the temporal location of the 40Gb/s payload is monitored within 6.4ns of accuracy using the payload envelope detection (PED) stage. The recovered header and PED signal are then forwarded to the ECP that is used to generate the control signals for the BUF stage.

9.4.3.1 Results and Discussion

End-to-end performance is evaluated for three regenerator configurations. First, a back-to-back (B2B) measurements is carried out by removing the buffer and 3R regeneration circuit, while keeping the signal degradation stage intact. Second, the buffer and 3R regenerator are inserted into the 40Gb/s end-to-end adaptation setup. Finally, the OCR stage is removed from 3R regenerator and replaced with a tunable laser (TL) and essentially limiting its regenerative properties to re-amplification and re-shaping. The third configuration is effectively identical to the forwarding plane (FWD) configuration consisting of a tunable laser connected to a differential λ-converter (WC).

Figure 9.26 shows the end-to-end Layer-III frame recovery results for the back-to-back (B2B), 3R regeneration, 2R regeneration configurations. A frame recovery percentage greater than 99% is successfully achieved over a 1 and 5dB span of dynamic receiver range for the 2R and the 3R configurations, respectively. Additionally, an enhancement in receiver sensitivity evaluated, at the 80% recovery mark, greater than 8dB is observed when the 2R configuration (TL→WC) is upgraded to a 3R configuration (OCR→WC). Furthermore, a 5dB enhancement in sensitivity, evaluated at a recovery of
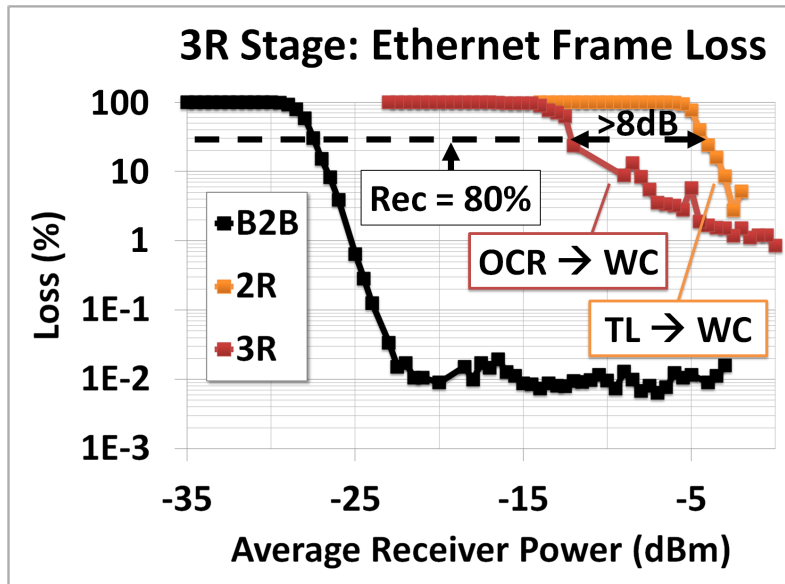
Figure 9.26: *Ethernet frame recovery results of 40Gb/s 3R regeneration stage measured for a degraded input signal (B2B) and the regenerator stage with (3R) and without (2R) optical clock recovery. The optical packet buffer is configured to bypass the re-circulating delay in these measurements.*

90%, is obtained when employing 3R regeneration.

9.5 Chapter Summary

The technology crucial to realizing the first demonstration of end-to-end transmission of IP traffic through a 40Gb/s 2x2 regenerative, buffered optical router has been presented in this chapter. The design and implementation of each OPS router subsystem has also been presented in great detail. The optical router consists of packet buffers that are utilized to resolve any contention that may occur when multiple packets make identical port requests. The routing and header rewrite functionalities were executed by a set of packet forwarding planes utilizing high-speed wavelength conversion

and AWG-based routing. Signal regeneration was performed via 3R stages that amplify, reshape, and re-time incoming signals. The end-to-end adaptation performance of each major optical router subsystem was separately characterized using the Adaptation Layers presented in previous chapters.

Both optical packet buffers were packaged in custom driver boards and demonstrated compatibility with the end-to-end adaptation process by exhibiting greater than 99% Ethernet frame recovery for at least 128ns of storage time. A buffer time of 192ns could be achieved by incurring a an additional 10% of frame loss penalty. Buffer performance was governed by SOA patterning rather than accumulation of ASE.

The packet forwarding planes are implemented with discretely, packaged SG-DBR lasers, MZI-WCs, MZMs, and an AWG to perform packet routing through label swapping. The widely tunable SG-DBR lasers used in the packet forwarding planes were packaged in a custom FPGA-based driver board and demonstrated a tuning range of about 20nm (2.5THz) with nanosecond-scale (<10ns) wavelength switching speeds. Header erasure and rewrite was successfully demonstrated with 100ps of temporal accuracy and more than 30dB of header-payload isolation. Both packet forwarding planes showed compatibility with end-to-end adaptation by obtaining greater than 99% Ethernet frame recovery for at least five output wavelengths. Power penalty variability within 10dB is caused by the physical layer limitations of the discrete implementation. The MZI-WC performance was degraded by λ-dependent PDL and non-adaptive biasing, while the SG-DBR performance was limited by optical feedback.

A 3R regeneration stage consisting of a high-speed wavelength conversion stage with a 40GHz optical clock recovery block as its pump input has also been presented. The optical clock recovery circuit included an integrated

mode-locked laser operated in a hybrid locking configuration. The 3R stage achieved 1mW of modulation depth enhancement (re-amplification), limited suppression of zero-level noise (re-shaping), and about 0.20ps of slight jitter reduction (re-timing). The regenerator re-shaping is limited by extinction ratio degradation caused by power imbalance in the MZI-WC and inefficient RF modulation of the MLL saturable absorber. Re-timing is limited by jitter transfer bandwidth of MLL and its susceptibility to optical feedback. Additional jitter transfer occurs as the recovered optical clock exhibits pulse widths comparable to the jittery gating window of the MZI-WC.

End-to-end adaptation compatibility has been demonstrated by achieving greater than 99% packet recovery through the 3R regeneration block. A 5dB improvement in receiver sensitivity, evaluated at a 90% frame recovery, was observed when comparing the 3R regeneration stage performance to a 2R configuration consisting of a tunable laser and an MZI-WC.

References

[1] J. P. Mack, H. N. Poulsen, D. J. Blumenthal, "Variable Length Optical Packet Synchronizer," *IEEE Photonics Technology Letters*, vol. 20, no. 14, pp. 1252-1254, 2008.

[2] E. F. Burmeister, J. P. Mack, H. N. Poulsen, J. Klamkin, L. A. Coldren, D. J. Blumenthal, and J. E. Bowers, "SOA gate array recirculating buffer with fiber delay loop," *Optics Express*, vol. 16, pp. 8451-8456. 2008.

[3] K. N. Nguyen, P. J. Skahan, J. M. Garcia, E. Lively, H. N. Poulsen, D. M. Baney, D. J. Blumenthal, "Monolithically integrated dual-quadrature receiver on InP with 30 nm tunable local oscillator," *Optics Express*, Vol. 19, Issue 26, pp. B716-B721, (2011).

[4] H. Shi, D. Cohen, J. Barton, M. Majewski, L. A. Coldren, M. C. Larson, and G. A. Fish, "Relative Intensity Noise Measurements of a Widely Tunable Sampled-Grating DBR Laser," *IEEE Photonics Technology Letters*, vol. 14, no. 6, pp. 759-761 (2002).

[5] B. R. Koch, J. S. Barton, M. Masanovic, Z. Hu, J. E. Bowers, and D. J. Blumenthal, "Monolithic Mode-Locked Laser and Optical Amplifier for Regenerative Pulsed Optical Clock Recovery," *IEEE Photonics Technology Letters*, vol. 19, no. 9, May 1 2007.

[6] M. Rochette, L. Fu, V. Taeed, D. J. Moss, B. J. Eggleton, "2R Optical Regeneration: An All-Optical Solution for BER Improvement," *IEEE Journal of Selected Topics In Quantum Electronics*, vol. 12, no. 4, pp. 736-744, July 2006.

[7] N. S. Bergano, J. Aspell, C. R. Davidson, P. R. Trischitta, B. M. Nyman, F. W. Kerfoot, "A 9000 km 5 Gb/s and 21,000 km 2.4 Gb/s Feasibility Demonstration of Transoceanic EDFA Systems Using a Circulating Loop," *Optical Fiber Communications Conference, (OFC) 1991*, Paper PD13-1, February 1991.

Chapter 10

Conclusions and Future Work

This dissertation has presented the technology needed to take significant steps towards realizing end-to-end communication through next-generation optical packet switching (OPS) core networks. This dissertation is organized into two main sections that discuss the details behind the design and implementation of two edge adaptation layers and an all-optical OPS router. Transmission of Internet Protocol (IP) traffic across an all-optical network is facilitated by edge adaptation layers that are capable of dynamic conversion between legacy 100MbE and future 40Gb/s OPS formats while achieving low packet loss with minimal latency penalties. The high-speed optical payloads are transparently switched by utilizing an OPS core router consisting of optical buffers, forwarding planes, and 3R regenerators. The first proof-of-concept demonstration of end-to-end transmission of IP datagrams encapsulated in Ethernet frames was shown through a 2×2 buffered, regenerative OPS core router. This demonstration was achieved with less than 1% of frame loss through each subsystem and end-to-end adaptation latencies below 273ns. A dynamically re-sizable optical packet buffer was also designed to allow the optical data outer to accommodate variable-length

214

packets to reduce the latencies incurred by packet fragmentation at the core edge. Contention resolution of 40Gb/s packets up to 800 bytes in length was achieved at packet loss rates below 5%. However, the end-to-end demonstration showed in this work is limited by several factors that currently prevent this technology from becoming practical for deployment. Though this work was implemented by utilizing photonic integrated chips (PICs), it was carried out via hybrid rather than monolithic integration. As a result, the performance of the optical switching elements was dominated by limitations at the physical layer such as loss, polarization dependence, gain saturation, noise accumulation, data patterning, and optical feedback. Finally, the sensitivity of the adaptation circuit and logic is limited to a performance regime that is currently not practical for long haul applications that require multiple node hops.

10.1 Performance Limitations

10.1.1 High-Speed Packet Generation and Recovery

The edge adaptation layers were designed and implemented with discrete electronic components that negatively affected the packet recovery sensitivity. The Layer-II packet loss sensitivity was observed to be 10^{-5} corresponding to a recovery percentage of 99.99999%. Additionally, the Layer-III frame recovery performance was clamped at 99.7% which corresponds to a frame loss percentage below 3×10^{-1}. The first order relationship between packet error loss (PEL) and bit-error-rate (BER) is estimated by the basic formula shown in (10.1), where N is the number of bits in a packet [1]. Figure 10.1 shows a graph of this relationship plotted for 64-byte frames packet where the observed measurement limitations of this work appear as dashed lines.
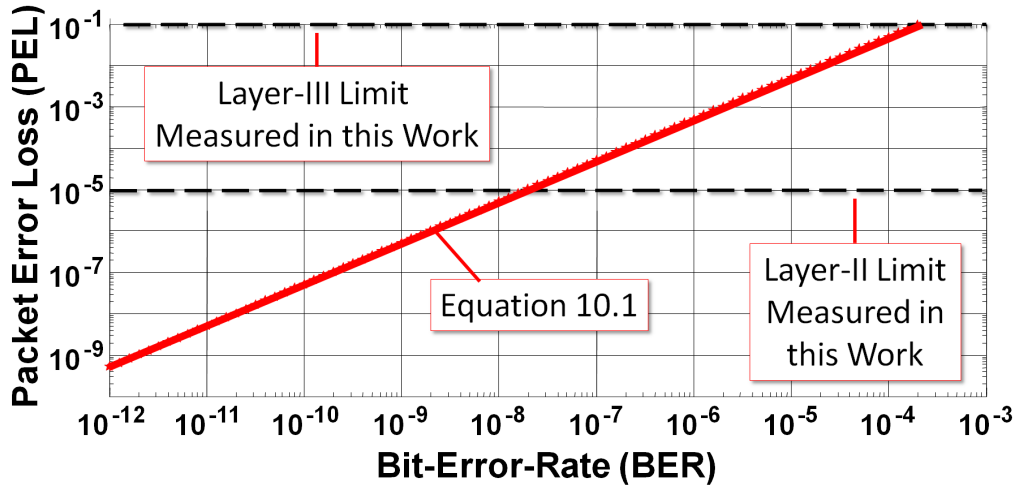
Figure 10.1: *Relationship between bit-error-rate (BER) and packet error loss (PEL) for 64-byte packets, using* (10.1), *is shown with the Layer-II and -III experimental sensitivities measured in this work (dashed).*

The Layer-II and -III sensitivities in this work are limited to BERs above 10^{-8} and 10^{-4} respectively. However, a BER of 10^{-9} is widely accepted as the threshold for error-free operation for applications involving long haul optical communications.

$$PEL = 1 - (1 - BER)^N \tag{10.1}$$

The adaptation layer limitations stem from temporal skewing between clock and data lines during the (de-)serialization process. The FPGA logic in this work performs time-critical operations on hundreds of parallel data signals that need to be bit-aligned with high precision. Therefore, efficient distribution of high-quality clock signals and uniform propagation delays are critical for high-performance applications. The electronic hardware in this work utilizes an on-board clock synthesizer to generate a 625MHz clock signal (of modest quality) that is distributed throughout the FPGA. Several

FPGA clocking primitives (buffers, dividers, multipliers, etc.) were utilized at 625Mb/s despite only being validated for clock rates up to 400MHz. This degraded not only the resulting clock signal, but also the components being clock by such signals. Adding to that, a daughter clock fanout board was designed and fabricated to distribute the 625MHz clocking signal to each of the four 16:1 serializers (SER). It was difficult to maintain equal path lengths within the daughter bard, which resulted in varying phase differences between each of the four clock-data pairs utilized by the SER hierarchy. In a like manner, varying phase differences between clock-data pairs were observed within the daughter board at the Egress Layer. Additionally, the utilization of automated logic placement resulted in non-deterministic signal propagation delays that varied each time firmware compilation and synthesis was carried out. Therefore, manual placement and FPGA logic was employed to mitigate this issue and obtain repeatable results.

10.1.2   Physical Layer

In a real system, input polarization needs to be optimized to reduce PDL within integrated devices. However, polarization was manually rotated to a transverse electric polarization via polarization controllers. The dynamic performance of the integrated devices within OPS core router is highly susceptible to input polarization state, which is randomized after propagating through single mode fiber. The active elements comprising the integrated devices consist of a quantum well design optimized for the transverse-electric (TE) polarization state. Polarization controllers (PCs) were incorporated within any experimental setup using fiber-based elements in conjunction with polarization sensitive PICs.

The optical packet buffers where implemented by including a PC within

the re-circulating loop to allow packets to reenter the buffer switch as TE-polarized signals. Likewise, the re-sizable packet buffer required a polarization controller in each path within the variable delay to minimize the polarization-dependent loss (PDL) of the integrated switches. Additionally, propagation losses of each delay path were balanced within 1dB by utilizing variable optical attenuators (VOAs). Though it was possible to obtain uniform PDL and propagation loss through each delay, the wavelength dependence of the PC restricted operation to a small range of optical bandwidth. Once loop losses (propagation and polarization) were compensated, buffer performance was limited by SOA carrier dynamics. The fixed delay buffer achieved a very limited amount of circulations as a result of data patterning caused the integrated SOA carrier recovery. The accumulation of amplified spontaneous (ASE) noise contributed to performance degradation as the number of achievable circulations decreased when additional SOA gates were inserted in the fiber delay to make the buffer re-sizable.

The performance of the optical packet forwarding plane was primarily limited by wavelength dependent PDL. The TE-polarized signal out of the widely tunable sampled grating-distributed Bragg reflector (SG-DBR) laser was randomized after traversing through meters of fiber patch cords. A PC was, therefore, connected in series with the laser output to maintain a TE polarization heading into the integrated Mach-Zehnder interferometer wavelength converter (MZI-WC). Consequently, it became difficult to obtain a cross-phase modulation (XPM) efficiency independent of SG-DBR output wavelength. As a result, a variation in receiver sensitivity greater within 10dB was observed when evaluating the end-to-end performance of the packet forwarding plane. Assuming wavelength dependent PDL could be eliminated by monolithically integrating the SG-DBR with the differential

MZI [2], performance is dictated by the output extinction ratio (ER). Currently, the dynamic ER of the MZI-WCs is limited to 6-9dB since it is difficult to obtain power balance between the MZI arms to achieve near-optimal destructive interference.

The 3R regeneration stage performance is not only limited by the reshaping properties of the MZI-WC, but mainly by the quality of the pulse train recovered in the re-timing stage. The 3R stage was implemented using a discretely integrated mode-locked laser (MLL) connected to an MZI-WC. Hence, the MZI-WC performance was affected by polarization dependence and non-optimal destructive interference. The MLL bias settings required for minimal timing jitter were sensitive to changes in bias made to components outside the laser cavity. Additionally, the device behavior was observed as having hysteresis as it became difficult to restore previous operation regimes following instances of configuration drift and power cycling. Typical MLL operation calls for a slightly reverse-biased saturable absorber (SA) to obtain narrow, high-ER pulses. The MLL used in this work was fabricated with an SA region that was longer than expected, which required a 0.5V forward bias resulting in 8-10 picosecond pulses with ER on the order of 6-9dB. Thus, a significant amount of jitter was transferred onto the recovered pulses since the pulse durations were comparable to the noisy, jittery gating window of the MZI-WC input signal. The most critical performance limitation observed within the MLL consisted of optical feedback. The free-running frequency, phase, and timing jitter of the recovered pulse train consistently varied with the amount of optical power reflected back into the laser cavity. Even after applying an anti-reflection (AR) coating, stable device operation was obtained by coupling power out of the device with a lensed fiber that was laterally offset from optimal coupling.

219

This makes packaging significantly difficult and expensive since active alignment with high-speed instruments may be required to monitor frequency, phase, and jitter along with coupled power. Moreover, this makes it prohibitively difficult to obtain dense integration with other active components while simultaneously achieving stable MLL operation.

10.2 Future Work

Integration is the first key step that needs to be taken to improve upon the physical layer limitations of the work presented in this dissertation. Constraining the physical placement of the FPGA logical building blocks led to the performance improvement within the edge adaptation layers. While the work presented here is sufficient for a proof-of-concept prototype, an application specific integrated circuit (ASIC) implementation would be necessary for practical deployment purposes. Integration of (de-)serialization hierarchies with control logic would greatly minimize non-deterministic path length differences that result in varying propagation delays. Doing so would ultimately lead to more reliable payload recovery at the Egress Layer, which processes a large number of parallel data streams that need to be bit-aligned with high precision.

Monolithic integration of optical devices used in this work would lead to reduction in footprint, power consumption, and physical layer limitations that are detriments overall system performance. Integration of the re-circulating delay and the SOA cross-bar switches into an optical packet buffer would eliminate the coupling losses and the PDL introduced by the insertion of the bulky fiber-based delay. Data patterning caused by SOA carrier dynamics can be reduced by utilizing monolithically integrated gain clamped

SOAs (GC-SOAs) as optical switches [3, 4] resulting in a buffer implementation limited by accumulation of waveguide loss and ASE noise. Such a device may be further improved via fabrication on an ultra low-loss platform similar to UCSB's iPhOD integration platform featuring planar waveguides with less than 0.1dB/m of propagation loss [5]. The switch matrix may be potentially implemented on the same platform that was recently heterogeneously integrated with active hybrid silicon optical components [6]. Finally, the issue associated with the accumulation of ASE may be mitigated by monolithically integrating a 2R regenerative structure such as a SA-SOA pair. Eventually, one would need to integrate an all-optical 3R regeneration circuit to compensate for the accumulation of timing jitter, but such integrated technology is far from maturity.

Monolithic integration of a 40Gb/s packet forwarding plane has been previously demonstrated by UCSB, but the PIC complexity is approaching the limit of what can be confidently fabricated in a research facility. As PIC functionality grows in complexity, it is becoming increasingly difficult to reliably fabricate sub-components that simultaneously meet specified requirements necessary for successful operation. For example, an MZI-WC has specific requirements including but not limited to pump power, differential input delay, waveguide loss, MMI splitting ratios, and SOA gain saturation regimes. Complex devices such as these need low-cost, reliable, generic integration technology to increase reliability and de-risk the fabrication process [7]. Performance variability may be minimized by utilizing an ASIC-like design space consisting of optical components whose performance tolerances are vetted and well known.

Integrated mode-locked lasers are the most promising technology for generating narrow, highly repetitive re-timed optical pulses, but are extremely

221

susceptible to optical feedback. The results presented in this work utilized a discrete implementation where an isolated EDFA was used to minimize optical feedback. However, a monolithically integrated solution is needed to achieve a smaller footprint and lower power consumption. Except, dense integration with MLLs becomes prohibitively difficult to attain without some form of on-chip isolation. Recently, potentially enabling integration technology consisting of magneto-optic isolators achieving more than 9 and 18dB of on-chip isolation in silicon-based waveguides has been demonstrated in [8–10] and [11], respectively.

10.3   Closing Remarks

This work has demonstrated successful forwarding, contention resolution, and regeneration of IP-based traffic with low end-to-end packet loss and latency. However, further research developments are needed to mature OPS router technology towards a point where it becomes viable for deployment in the core. Hence, the OPS router technology presented in this work is best suited for short-term applications that call for fast, efficient forwarding rather than routing. Certain data center networks are implemented as flat, low-latency topologies that provide any-to-any single-hop connectivity. Such network configurations stand to benefit from the forwarding technology presented in this work. The fast-switching, wavelength agile TWC paired with a passive AWG may provide improvements in reduced footprint and power consumption when integrated monolithically. Current electronic routers utilize vast memory storage banks to execute complex functionalities such as scheduling, routing, and signal regeneration. The implementation of OPS routers within backbone networks will be far from practical until integrated optical buffer

technology performance is at least comparable to its electronic counterpart.

References

[1] J. G. Proakis, *Digital Communications*, McGraw-Hill, New York, 1983.

[2] V. Lal, "Monolithic Wavelength Converters for High-Speed Packet Switched Optical Networks," Ph.D. Dissertation, University of California, Santa Barbara, June 2006.

[3] D. Wolfson, "Detailed Theoretical Investigation and Comparison of the Cascadability of Conventional and Gain-Clamped SOA Gates in Multi-wavelength Optical Networks," *IEEE Photonics Technology Letters*, vol. 11, no. 11, pp. 1494-1496, November 1999.

[4] F. Dorgeuille, L. Noirie, J.-P. Faure, A. Ambrosy, S. Rabaron, F. Bouval, M. Schilling, and C. Artigue, "1.28 Tbit/s throughput 8×8 optical switch based on arrays of gain-clamped semiconductor optical amplifier gates," *Optical Fiber Communications Conference, (OFC) 2000*, Paper PD18-1, pp. 221-223, 2000.

[5] J. Bauters *et al.*, "Planar waveguides with less than 0.1 dB/m propagation loss fabricated with wafer bonding," *Optics Express*, vol. 19, no.24, pp. 24090-24101, November 2011.

[6] M. L. Davenport, J. Bauters, M. Piels, A. Chen, A. Fang, J. E. Bowers, "A 400 Gb/s WDM Receiver Using a Low Loss Silicon Nitride AWG

Integrated with Hybrid Silicon Photodetectors," *Optical Fiber Communications Conference, (OFC) 2013*, Paper PDP5C.5, March 2013.

[7] M. Smit, X. Leijtens, E. Bente, J. Van der Tol, H. Ambrosius, D. Robbins, M. Wale, N. Grote, M. Schell, "Generic foundry model for InP-based photonics," *European Integrated Optoelectronics, Conference on*, vol. 5, no. 5, pp. 187-194, 2011.

[8] M.-C. Tien, T. Mizumoto, P. Pintus, H. Kromer, J. E. Bowers, "Silicon ring isolators with bonded nonreciprocal magneto-optic garnets," *Optics Express*, vol. 19, no. 12, pp. 11740-11745, June 2011.

[9] P. Pintus, M.-C. Tien, and J. Bowers, Design of Magneto-Optical Ring Isolator on SOI Based on the Finite-Element Method, *IEEE Photonics Technology Letters*, vol. 23, no. 22, pp. 1670-1672, November 2011 .

[10] H. Kroemer, J. E. Bowers, and M.-C. Tien, Ring Resonator Based Optical Isolator and Circulator, patent number 8,396,337, March 12, 2013.

[11] Y. Shoji, T. Mizumoto, H. Yokoi, I.-W. Hsieh, and R. M. Osgood, "Magneto-optical isolator with silicon waveguides fabricated by direct bonding," *Applied Physics Letters*, vol. 92, no. 7, 2008.