UNIVERSITY OF CALIFORNIA, SAN DIEGO

Silicon Photonics with Applications to Data Center Networks

A dissertation
submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy

in

Electrical Engineering (Applied Physics)

by

Ryan Francis Aguinaldo

Committee in charge:

Shayan Mookherjea, Chair
Prabhakar Bandaru
Yeshaiahu Fainman
George Papen
George Porter

2014

The dissertation of Ryan Francis Aguinaldo is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2014

EPIGRAPH

There is no such thing as a photon.
Only a comedy of errors and historical accidents
led to its popularity among physicists and optical scientists...
There are very good substitute words
for "photon" (e.g., "radiation" or "light")
and for "photonics" (e.g., "optics" or "quantum optics").
—Willis Lamb, Jr. [1]

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation was defended on June 5, 2014. It was approximately three-and-a-quarter years before that when, looking for greener grass and more fertile fields, I first approached Prof. Shayan Mookherjea regarding admission into his group. To my benefit, he accepted and, under his advice, guidance, direction, and mentorship, I am able to present this work three-and-a-quarter years later. In addition to the research, I also assisted in teaching with Prof. Mookherjea. The great deal of discretion he gave me in those duties has allowed me to refine my skills as a teacher; for my betterment and that of my students. For all this, I am grateful.

The other members of my doctoral committee are Prof. Prab Bandaru, Prof. Shaya Fainman, Prof. George Papen, and Prof. George Porter. Prof. Bandaru has been very accommodating ever since I first ambushed him after one of his classes to schedule my candidacy examination. I have known Prof. Fainman since I took his class on Fourier optics. Ever since then, he has been an excellent professional connection; full of technical knowledge and friendly jokes! Prof. Papen and Prof. Porter have both been instrumental to this dissertation. Many of the experimental measurements made use of devices motivated by or interfaced with their MORDIA network. In that vein, Prof. Papen's student, Mr. Alex Forencich, has been helpful in making sure that sufficient optical power was available to perform the interfaced measurements.

Throughout the past three-and-a-quarter years, I have shared an office and a lab with some fine scientists. They include Dr. Jun Rong Ong, Mr. Ranjeet Kumar (Jaswal), Mr. Mark Schneider, Mr. Yiran Shen, Mr. Peter Weigel, Ms. Hannah Grant, and Dr. Mark Savanier. The opportunity to discuss ideas and their intricate details with one's colleagues cannot be praised enough! I should also make note of the scientific debates that I've shared with Mr. Kumar; truly the best way to learn and reinforce one's knowledge in a subject, these debates are rivaled only by those shared with Dr. Patrick Whiting in a previous time.

Many of the experimental measurements performed in this work made use of devices fabricated at Sandia National Laboratories, under the direction of Dr. Tony

Lentine and Dr. Chris DeRose. This collaboration marked the first time that outside parties (i.e. myself and colleagues at several other universities) were able to submit silicon-photonic mask layouts to be fabricated through Sandia's CMOS process. I even had the opportunity to work alongside Dr. Lentine and Dr. DeRose, onsite at Sandia, for a summer. In addition to the interesting work, I also got a chance to see Čerenkov radiation with my own eyes. What a treat!

The collaborations with Sandia and Profs. Fainman, Papen, and Porter, as well as partial support, were provided through the NSF Center for Integrated Access Networks (CIAN). Measurements that made use of the MORDIA network were performed in the CIAN Chip-Scale Photonics Testing Laboratory, University of California, San Diego.

I must thank my family for their love and support over the years. When they first sent a young layman off to get an engineering degree, I am sure that they never expected him to fall in love with science and end up pursuing a PhD. I know that I never expected it!

Finally, and very importantly, I must acknowledge my SO, Ms. Jennifer Driggers. She has accompanied me throughout these past three-and-a-quarter years (and somewhat longer than that!); putting up with late nights, early mornings, late nights that go into early mornings, time apart while I was in Albuquerque, etc. On her own accord, invoking her professional knowledge of wave physics, inter alia, she even proof-read this dissertation!

To all, many thanks — cheers!

Ryan Aguinaldo
San Diego, CA
June 5, 2014

Chapter 4, in part, contains material, published in the following, of which the dissertation author was the primary investigator:
R. Aguinaldo, Y. Shen, and S. Mookherjea, "Large Dispersion of Silicon Directional Couplers Obtained via Wideband Microring Parametric Characterization," *IEEE Photon. Technol. Lett.*, vol. 24, pp. 1242–1244, July 2012.

Chapter 5, in part, contains material, published in the following, of which the dissertation author was the primary investigator:
R. Aguinaldo, P. Weigel, H. R. Grant, C. T. DeRose, A. Lentine, A. Pomerene, A. Starbuck, and S. Mookherjea, "Characterization of a silicon-photonic multi-wavelength power monitor," in *IEEE Opt. Interconnects Conf.*, p. WD5, 2014.

Chapter 6, in part, contains material, published in the following, of which the dissertation author was the primary investigator:
R. Aguinaldo, A. Forencich, C. T. DeRose, A. Lentine, D. C. Trotter, Y. Fainman, G. Porter, G. Papen, and S. Mookherjea, "Wideband silicon-photonic thermo-optic switch in a wavelength-division multiplexed ring network," *Opt. Express*, vol. 22, no. 7, pp. 8205–8218, 2014.

Chapter 7, in part, contains material, to be submitted for publication as the following, of which the dissertation author was the primary investigator:
R. Aguinaldo, H. Grant, A. Forencich, A. Lentine, C. DeRose, D. Trotter, A. Pomerene, A. Starbuck, Y. Fainman, G. Papen, G. Porter, and S. Mookherjea, "A silicon photonic network node-on-a-chip for optically interconnected data-center networks," in-preparation (2014).

VITA

2004-2005	Student Technician Semiconductor & Microsystems Fabrication Laboratory Rochester Institute of Technology
2005	Undergraduate Student Researcher (Internship) College of Nanoscale Science & Engineering State University of New York
2005-2008	Undergraduate/Graduate Student Researcher Nanopower Research Laboratories Rochester Institute of Technology
2006	Process Engineer, 300-mm Ion Implant (Internship) IBM
2009	Bachelor of Science Microelectronic Engineering with minors in Physics and Electrical Engineering Rochester Institute of Technology
2009	Master of Science Materials Science Rochester Institute of Technology
2009-2014	Graduate Student Researcher Dept. of Electrical & Computer Engineering University of California, San Diego
2010	Master of Science Electrical Engineering (Nanoscale Devices & Systems) University of California, San Diego
2013	Graduate Student Researcher (Internship) Sandia National Laboratories
2014	Doctor of Philosophy Electrical Engineering (Applied Physics) University of California, San Diego

ABSTRACT OF THE DISSERTATION

Silicon Photonics with Applications to Data Center Networks

by

Ryan Francis Aguinaldo

Doctor of Philosophy in Electrical Engineering (Applied Physics)

University of California, San Diego, 2014

Shayan Mookherjea, Chair

In data center applications, fiber-based optical interconnects can be used to provide point-to-point links enabling high-bandwidth, inter-rack, data communications. In order to provide for future network scalability, which must be able to handle ultra-large data flows and bandwidth-intensive requests, optical technologies are increasingly introduced to different levels of the data center architecture to enable a variety of transparent network or all-optical networking schemes. However, the use of bulk optical components, which take up valuable rack-space real estate, can be extremely energy and cost prohibitive, especially when scaled up to the size of industrial warehouse-scale computing and considering that predictions of future data center networks are expected to contain millions of nodes. As such, we study chip-scale, silicon photonic, integrated circuits and their use as the optical hardware

in future data center implementations. This work describes aspects of the design and integration of silicon photonic devices, which can be used for high-bandwidth, multi-channel, wavelength division multiplexed, optical communications. Examples of silicon photonic subsystems are discussed, including the realization of an on-chip channelized spectrum monitor and a network-node-on-a-chip. These optical integrated circuits are meant to replace bulk optical components with their functional equivalents on monolithic silicon. This work demonstrates that silicon photonics may be advantageous in meeting the urgent hardware-scaling demands of high-bandwidth, multi-user, communication networks.

Chapter 1

Introduction

1.1 Silicon photonics and data centers

The present work is concerned with monolithically integrated optics on a silicon platform, i.e. silicon photonics. Fueled by modern society's consumption of digital electronics, silicon manufacturing technology represents an extremely mature process that has had half-a-century of continuous, law-like [4], exponential improvements and breakthroughs [5]. Indeed, as of this writing, the current state of the art, in the silicon manufacturing of end-user microelectronics, includes devices with critical dimensions of ∼14 nm, which are fabricated on 300-mm substrates, with 450-mm substrates on the horizon. Therefore, one of the benefits of silicon photonics is the ability to leverage an already mature and large-scale manufacturable process for the realization of optical integrated circuits. Additionally, since silicon waveguide widths tend to be between 200–2000 nm or larger, older fabrication technologies may be utilized; these older processes being more reliable, more accessible[1], and less costly. Of course, fabricating optics on a CMOS-compatible process allows for the monolithic integration of electronic CMOS circuitry with active optical devices such as switches, modulators, reconfigurable filters, etc.

Some of the original motivation for silicon photonics had to do with on-chip

[1] Indeed, it is likely that only the "select-few" key industry organizations would have access to a 14-nm process, whereas multi-project wafer runs on less advanced processes are commonly utilized by university research groups and small companies.

and off-chip optical interconnects [6, 7]. Upwards of 80% of the power consumption in a modern microprocessor goes into driving the long copper lines that route the electrical signals. This power dissipation occurs, in part, due to resistive losses and, although possibly minimized, still unavoidable impedance mismatches. Dielectric waveguiding, on the other hand, which is provided by silicon photonics, mitigates such blatant energy waste in the interconnect by completely removing from the equation the R's, L's, and C's of electronic wires.[2] These are the parameters that describe an electromagnetic wave guided by a conductor (i.e. an electronic signal propagating down a wire). In an optical dielectric waveguide, these parameters are replaced solely by the refractive index, which does not have associated with it I^2R losses or RC and L/R time delays. Furthermore, electronic interconnects exhibit maximum bit rates that scale as A/L^2 [6], where A is the cross-sectional area of the wire and L is its length. For denser integration, A can only get smaller while L might get larger. The bit rate of an optical waveguide, however, could better approach the Shannon limit.

While the foregoing application of silicon photonics is important, the present work focuses on a different application: data-center networks. In a sense, data centers are at the core of the internet, with virtually all internet traffic ending up in a data-center server. With the massive popularization, within the last decade, of cloud computing, cloud storage, streaming media, and mobile computing devices, data-center traffic has shifted from a linear growth model to a 31% compounded annual growth rate [8]. Such massive growth may be a challenge for future scalability and lead to network congestion. Additionally, the physical growth of the amount of hardware needed to support the usage trends has caused an exponential increase in the collective power consumption of data centers [9]. Indeed, internet traffic is projected to make up 10% of the world's energy usage by 2018 [10].

To help combat the rapid growth in the bandwidth demand of data-center networks, all-optical networking technology is being introduced, for rack-to-rack and intra-rack communications, to replace bandwidth-limiting electronic components and cabling [11, 12, 13]. However, future data-center networks are expected

[2] Resistance, inductance, and capacitance.

to scale to millions of nodes [14, 15], which will require their components to be compact, energy efficient, and easily manufacturable. This poses a problem for conventional off-the-shelf optical components, which are bulky, consume many watts of power, and are relatively expensive to manufacture. Therefore, silicon photonics, which can be manufactured relatively inexpensively due to wafer-scale processing[3], becomes an obvious candidate for use in next-generation data-center networks.

1.2 ITU-T 100-GHz telecommunications grid

The Telecommunication Standardization Sector (ITU-T) of the International Telecommunication Union (ITU) specifies several frequency grids (i.e. channel plans) for use in communications applications based on wavelength division multiplexing (WDM) [17, 18]. In dense WDM (DWDM) applications, the ITU-T channels are equally spaced in frequency; i.e. they are not equally spaced in wavelength. One of the more popular grids is the 100-GHz-spaced grid. Throughout this dissertation, we make use of the 100-GHz grid, and so we summarize the grid in Table 1.2 for quick reference. The MORDIA network, which we will reference on several occasions, also makes use of the 100-GHz grid. The channels in use by MORDIA are highlighted in Table 1.2.

Plus or minus a channel or so [19], channels 16–59 make up the telecommunications C-band; this is the so-called "erbium window," or the range of wavelengths most commonly amplifiable by erbium-doped fiber amplifiers. Lower frequency channels that are listed in Table 1.2 are part of the L-band; higher frequency channels in the table are part of the S-band. A convenient way to remember the table is to recognize that the channel numbers are precisely the last two digits of the corresponding frequency. For example, the last two digits of 195.1 THz are *5* and *1*; therefore, this corresponds to Ch. 51. To quickly remember the corresponding wavelength, one needs only to divide the velocity of light by the desired

[3] A colloquialism that has recently emerged, paraphrased ad lib. by the present author, is that, when under the purview of optics, "CMOS" should stand for "cheaply manufacturable optical subsystems" [16].

frequency on the fly.

Table 1.1: ITU-T 100-GHz telecommunications grid. Highlighted channels are used by the MORDIA network.

Channel No.	Frequency (THz)	Wavelength (nm)	Channel No.	Frequency (THz)	Wavelength (nm)
1	190.1	1577.03	37	193.7	1547.72
2	190.2	1576.20	38	193.8	1546.92
3	190.3	1575.37	39	193.9	1546.12
4	190.4	1574.54	40	194.0	1545.32
5	190.5	1573.71	41	194.1	1544.53
6	190.6	1572.89	42	194.2	1543.73
7	190.7	1572.06	43	194.3	1542.94
8	190.8	1571.24	44	194.4	1542.14
9	190.9	1570.42	45	194.5	1541.35
10	191.0	1569.59	46	194.6	1540.56
11	191.1	1568.77	47	194.7	1539.77
12	191.2	1567.95	48	194.8	1538.98
13	191.3	1567.13	49	194.9	1538.19
14	191.4	1566.31	50	195.0	1537.40
15	191.5	1565.50	51	195.1	1536.61
16	191.6	1564.68	52	195.2	1535.82
17	191.7	1563.86	53	195.3	1535.04
18	191.8	1563.05	54	195.4	1534.25
19	191.9	1562.23	55	195.5	1533.47
20	192.0	1561.42	56	195.6	1532.68
21	192.1	1560.61	57	195.7	1531.90
22	192.2	1559.79	58	195.8	1531.12
23	192.3	1558.98	59	195.9	1530.33
24	192.4	1558.17	60	196.0	1529.55
25	192.5	1557.36	61	196.1	1528.77
26	192.6	1556.55	62	196.2	1527.99
27	192.7	1555.75	63	196.3	1527.22
28	192.8	1554.94	64	196.4	1526.44
29	192.9	1554.13	65	196.5	1525.66
30	193.0	1553.33	66	196.6	1524.89
31	193.1	1552.52	67	196.7	1524.11
32	193.2	1551.72	68	196.8	1523.34
33	193.3	1550.92	69	196.9	1522.56
34	193.4	1550.12	70	197.0	1521.79
35	193.5	1549.32	71	197.1	1521.02
36	193.6	1548.51	72	197.2	1520.25

Chapter 2

Nonlinear Loss

2.1 Overview

Silicon photonics offers an interesting platform to study nonlinear optics [20, 21, 22, 23, 24, 25, 26, 27], due in part to its relatively strong $\chi^{(3)}$ and Raman nonlinearities [28] as well as the ability to observe a $\chi^{(2)}$ nonlinearity via lattice strain [29, 30, 31]. In addition to these qualities, silicon, being a semiconductor, suffers significantly from the "linear" impairment of free-carrier absorption (FCA) [32] and the nonlinear impairment of two-photon absorption (TPA) [22]. As we will show, the dependence of FCA on TPA causes the FCA mechanism to become highly nonlinear. The present chapter is concerned with the foregoing nonlinear impairments. Specifically, in the design of silicon photonic circuits, it is important to know, via design and simulation, the losses that can be expected as well as if those loses can be mitigated. In what follows, we discuss methods to predict and mitigate the effects of FCA and TPA at varying degrees of optical power. Additionally, we discuss the prediction of intensity-dependent loss, which occurs due to the nonlinearity of the extinction mechanisms.

2.2 Free carrier absorption

Intrinsic silicon, having a bandgap of 1.12 eV, is transparent to light that has a wavelength longer than $\sim$1.11 μm. This transparency occurs because of a

filled valence band and a nearly "empty" ($\sim 10^{10}$ electrons/cm^3) conduction band. Electrons in lower bands are unable to move into the fully occupied valence band; meanwhile, electrons in the valence band require more energy than the light is able to endow for them to move into the conduction band. If, however, the conduction band was not depleted, then the *free carriers* in that band could easily move to higher bands by absorbing the otherwise unabsorbable light. This absorption mechanism is termed free carrier absorption (FCA). A similar argument applies for the population of free holes in the valence band. These free carriers (i.e. conduction-band electrons or valence-band holes) can be populated by the usual means of doping, photogeneration, etc.

Applying the Drude model to a semiconductor yields the following expression for the intensity attenuation coefficient due to FCA [32, 20]:

$$\alpha_{\text{FCA}} = \frac{e^3}{\epsilon_0 c n \omega^2} \left(\frac{N}{\mu_N m_N^{*2}} + \frac{P}{\mu_P m_P^{*2}} \right), \tag{2.1}$$

where e is the elementary charge, ϵ_0 is the permittivity of free space, c is the velocity of light, n is the refractive index, ω is the angular frequency of light, N and P are electron and hole concentrations, μ is the carrier mobility, and m^* is the effective mass. In the literature, it is common to express the foregoing as

$$\alpha_{\text{FCA}} = \sigma_N N + \sigma_P P, \tag{2.2}$$

where σ is the FCA cross-section, which is an empirical parameter. The celebrated Soref-Bennett result is $\sigma_N = 8.5$ Mbarns and $\sigma_P = 6.0$ Mbarns at 1550 nm for silicon [32, 33]. At 1300 nm, $\sigma_N = 6.0$ Mbarns and $\sigma_P = 4.0$ Mbarns. Since the community is often interested in cases where $N = P$ (e.g. during photogeneration) and $\lambda = 1550$ nm, the literature sometimes represents (2.2) as

$$\alpha_{\text{FCA}} = \sigma N_{\text{EHP}}, \tag{2.3}$$

where $\sigma = 1.45 \times 10^{-17}$ cm^2 and N_{EHP} is the number of electron-hole pairs [20, 34]. In the present work, in order to keep as much generality as possible, we use (2.2) with

$$\begin{aligned} \sigma_N &= 1.24 \times 10^{-9} \times \frac{\lambda^2}{n}, \\ \sigma_P &= 8.50 \times 10^{-10} \times \frac{\lambda^2}{n}. \end{aligned} \tag{2.4}$$

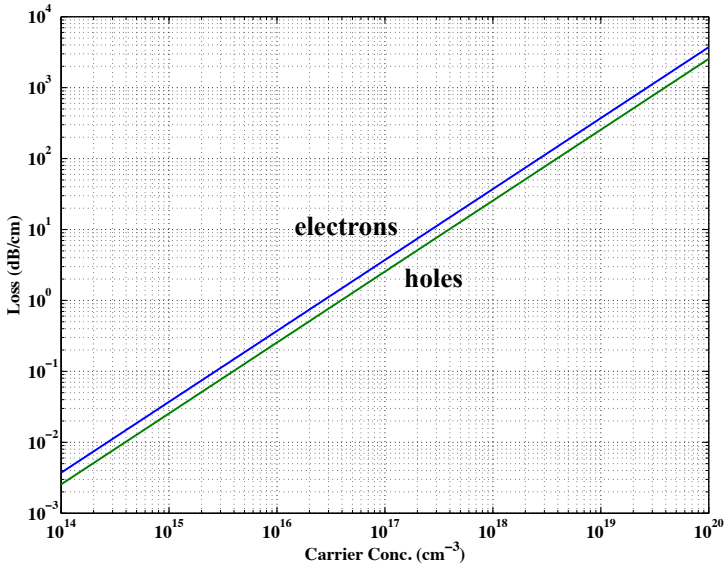

Figure 2.1: Optical attenuation in silicon, due to free carrier absorption, at a wavelength of 1550 nm. The effects of conduction-band electrons and valence-band holes are shown separately and can be summed as necessary.

These expressions for the cross-sections were obtained by curve fitting, using the four coefficients obtained by Soref and Bennett, while retaining the $\frac{\lambda^2}{n}N$ and $\frac{\lambda^2}{n}P$ dependencies seen in (2.1). A plot of FCA loss, using our expressions for the FCA cross-sections with $\lambda = 1550$ nm, is shown in Fig. 2.1. Strictly speaking, (2.4) is only applicable at near-infrared telecommunications frequencies. Further generalizations, appropriate in the mid-infrared, are provided in the literature [35].

2.3 Two-photon absorption

In semiconductor waveguiding, the frequency ν of the light that is used is chosen such that the photon energy $h\nu$ (h being Planck's constant) is less than the bandgap E_g of the semiconductor. The semiconductor is thus transparent to the light. There is a fundamental process, however, where normally transparent frequencies may be absorbed by the semiconductor.

Two-photon absorption (TPA) in semiconductors is a process in which sub-bandgap light (i.e. $h\nu < E_g$) is absorbed by the semiconductor and causes the

photogeneration of an electron-hole pair (EHP). This process occurs via the simultaneous absorption of two photons and requires that $2h\nu \geq E_g$; thus, the energy of two photons is transferred to the excitation of a single EHP. Physically, EHPs represent the existence of free carriers; i.e. conduction-band electrons and valence-band holes. These mobile electrons (holes) are easily excitable to higher (lower) bands through the standard single-photon absorption process; this is termed free carrier absorption (FCA). Therefore, in the presence of TPA, the light actually suffers from two distinct absorption mechanisms: TPA and TPA-induced FCA.

FCA, in an equilibrium sense, can usually be considered a linear absorption process that is governed by the usual Beer-Lambert law:

$$\frac{dI}{dz} = -\alpha_{\text{lin}} I, \tag{2.5}$$

where I is the optical intensity, z is the coordinate in which the light propagates with respect to, and

$$\alpha_{\text{lin}} = \alpha_{\text{FCA}} + \alpha_{\text{scat}} \tag{2.6}$$

is the usual intensity attenuation coefficient; α_{scat} contains the effects of other attenuation mechanisms — most commonly scattering. On the other hand, TPA is an inherently nonlinear absorption process [20, 22]:

$$\frac{dI}{dz} = -\alpha_{\text{TPA}} I^2, \tag{2.7}$$

where α_{TPA} is the TPA coefficient, which plays a similar role to α_{lin} in (2.5) (i.e. α_{TPA} determines the strength of the TPA process — larger values of α_{TPA} increase the number of TPA events that occur). In silicon, $\alpha_{\text{TPA}} \sim 0.5 - 1.5$ cm/GW at a wavelength of 1550 nm [28], signifying that TPA is a relatively weak process. For example, momentarily considering TPA as an isolated process and given a typical silicon waveguide cross-section of 0.1 μm^2, 250 mW of power must be coupled into the waveguide for the light to suffer a loss of 1 dB in 1 cm of propagation distance, which is on par with typical passive waveguide loss due to fabrication imperfections, sidewall scattering, etc. Actually, as we will elucidate below, a more appropriate estimate is 100 mW due to the additional effect of FCA. Throughout the present work, we will invoke a value of 1 cm/GW for α_{TPA}. Because

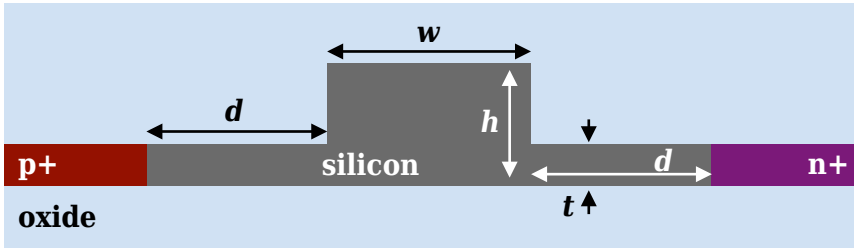

Figure 2.2: Ridge waveguide that doubles as a PIN diode. The effective lifetime of TPA-photogenerated free carriers in the core is reduced by the contact potential formed in the diode. A reverse bias on the diode aids in the free carrier extraction process.

we are talking about relatively large powers considering the size of the waveguide cross-section, we note parenthetically that the silicon breakdown intensity is ~ 6 GW/cm^2 [36].

Relatively high optical powers on the order of 10 – 100 mW are common in experiments involving Raman lasing [34, 37] and four-wave mixing [27] in silicon. In these experiments, TPA and FCA are detrimental effects that limit the efficiencies of the processes actually under investigation. The TPA process is fundamental to the presence of high optical intensities; however, the induced FCA process can be mitigated by simply sweeping the free carriers out of the vicinity of the optical mode before they are able to absorb any light. This carrier sweep out is accomplished by the structure in Fig. 2.2. In this ridge-waveguide structure, the waveguide core is intrinsic or background-doped ($\lesssim 10^{15}$ cm^{-3}) silicon with nearby p+ and n+ regions. These heavily doped regions are ideally far enough away to not perturb the optical mode but near enough to efficiently enable the extraction of free carriers from the core. As such, the silicon waveguide doubles as a PIN diode. Under zero bias, when free carriers are photogenerated via TPA, the built-in potential of the PIN junction sweeps these free carriers away from the core producing a reverse current; essentially a two-quantum version of the photovoltaic effect [38]. A reverse bias across the junction increases the electrostatic field in the core and thus the rate of removal of photogenerated carriers.

To model carrier sweep-out structures, we adopt a two-tier approach in the use of modeling tools. A standard vectorial mode solver yields the field distribution

over the cross-section of a waveguide. The corresponding time-averaged Poynting vector then yields the distribution of intensity $I(x, y)$. I can be scaled by an arbitrary multiplicative constant to satisfy

$$\int d^2 r I = P, \tag{2.8}$$

where P is the desired amount of power in the waveguide. We also recognize that the left-hand side of (2.7) can be interpreted as the change in optical energy, per unit volume, per unit time. Since we know that two quanta must participate in every absorption event, we may write

$$\frac{dI}{dz} = -2h\nu G_{\mathrm{TPA}}, \tag{2.9}$$

where G_{TPA} is the rate of absorption events per unit volume. G_{TPA} is therefore also the generation rate of EHPs per unit volume. Comparing the foregoing with (2.7), we find that the EHP generation rate due to TPA is

$$G_{\mathrm{TPA}} = \frac{\alpha_{\mathrm{TPA}}}{2h\nu} I^2; \tag{2.10}$$

i.e. the modal solution directly gives us the TPA generation rate, which is appropriate for use in a finite-element method (FEM) solution to the carrier dynamics.

A commercial semiconductor FEM tool, such Silvaco Atlas, can be used to model the structure of Fig. 2.2 once G_{TPA} is known. The FEM tool solves the carrier transport equations [38]:

$$\begin{aligned} \frac{\partial N}{\partial t} &= G - R + \mu_N N \nabla \cdot \mathbf{E} + \mu_N \mathbf{E} \cdot \nabla N + D_N \nabla^2 N, \\ \frac{\partial P}{\partial t} &= G - R - \mu_P P \nabla \cdot \mathbf{E} - \mu_P \mathbf{E} \cdot \nabla P + D_P \nabla^2 P, \end{aligned} \tag{2.11}$$

where $\mathbf{E}$ is the electrostatic field (which we consider to be distinct from the electric field of the optical mode), D is the carrier diffusivity, R is the EHP recombination rate, and $G = G_{\mathrm{TPA}} + G_{\mathrm{other}}$ is the EHP generation rate, which includes TPA as well as other effects such as thermal generation, impact ionization, etc. The solutions are the carrier concentrations N and P subject to the geometrical boundary conditions of the structure as well as the applied bias.

With N and P known throughout the structure, the effective lifetimes of the free carriers are [20, 39, 40]:

$$\begin{aligned} \tau_N^{\text{eff}} &= \frac{\langle N \rangle}{\langle G \rangle}, \\ \tau_P^{\text{eff}} &= \frac{\langle P \rangle}{\langle G \rangle}, \end{aligned} \tag{2.12}$$

where the notation

$$\langle \cdot \rangle = \int (\cdot) I d^2 r \tag{2.13}$$

represents a weighted average with respect to the optical mode. The physical interpretation of the effective lifetimes is that TPA-generated free carriers only remain in the vicinity of the optical mode for a limited amount of time, after which their presence can no longer affect the light. This temporal period intrinsically exists due to recombination, but can be shortened by the presence of the junction field that serves to sweep free carriers away from the optical mode. In summary, a mode solver implicitly yields the TPA generation rate, which is then fed into an FEM semiconductor solver. The semiconductor solutions, however, must then be averaged over the original modal solution to quantify the mitigation of TPA-induced FCA.

As an example, we model the structure in Fig. 2.2 with a background p-type concentration of 10^{15} cm^{-3}, 10^{18} cm^{-3} in the heavily doped regions, $w =$ 650 nm, $h = 220$ nm, $t = 70$ nm, $d = 900$ nm, and $\lambda = 1550$ nm. We also take $\alpha_{\text{TPA}} = 1$ cm/GW [28] and common values of 100 ns for the Shockley-Reed-Hall recombination lifetime [41, 42] and 100 cm/s for the surface recombination velocity [43, 44]. For this geometry, the fundamental TE mode exhibits a modal phase index of 2.648 and an effective area of 0.158 μm^2. Various results of coupling the modal solution with the semiconductor FEM solver are plotted in Fig. 2.3. Specifically, in Fig. 2.3.a we see that the spatially-averaged carrier concentration increases with increasing intensity as a result of TPA. In fact, intensity thresholds are apparent where the rate of increase occurs much faster; these are the intensities at which TPA is no longer a weak effect. Comparing Fig. 2.3.a to Fig. 2.3.b, we see the relation between increased carrier concentrations and the effective free-carrier lifetimes. When a reverse bias is applied, the carriers are swept away from the

optical mode and a decrease in lifetime is observed. For the sake of comparison, we also plot in Figs. 2.3.a–b the effects of forcing an open-circuit condition on the junction; this is a forward-biased condition such that no current is allowed to flow through the device. At low intensities, the effective lifetime is actually increased because the forward bias causes a low-level of injection of additional free carriers into the junction, which follows with it a decrease in the electric field throughout the junction. At all biases, the lifetimes are asymptotic at large intensities to the same lifetime. At these high optical intensities, the carrier concentrations are so large that they screen the electrostatic field of the junction. This lifetime asymptote is the lifetime that the structure would exhibit at all intensities if the p+/n+ regions were not present in the structure.

Going further, we analyze the foregoing example at zero bias and an optical intensity of 100 mW while varying either the background dopant concentration (Fig. 2.4.a) or the surface recombination velocity (Fig. 2.4.b). In the former case, we see that there is no benefit of switching to a more intrinsic substrate since the TPA lifetime does not vary much at background dopant concentrations lower than $\sim 10^{15}$ cm^{-3}. In the latter case, we see that the junction could be aided in decreasing TPA lifetime by increasing the surface recombination velocity. Such an effect, however, is very weak at 0 V and methods to increase the surface recombination, such as increased sidewall roughness, would add more in scattering loss to the structure than is saved in FCA loss.

As a final example, we change the core height to $h = 340$ nm and plot the total free-carrier concentration and average free-carrier lifetime, as functions of optical intensity, for different applied biases, in Fig. 2.5. For this new geometry, the modal phase index is 2.915 and the effective area is 0.172 μm^2. Similar trends, as in the previous example, are observed. The free-carrier lifetimes, however, are slightly longer due to the increased size of the waveguide core.

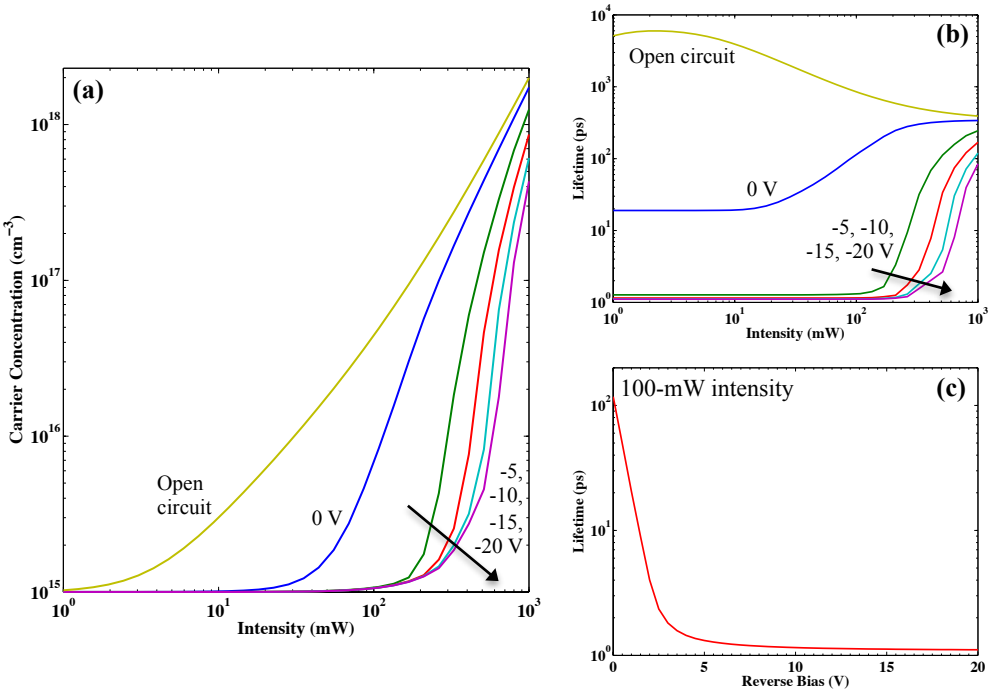

Figure 2.3: Simulation of the swept-carrier structure in Fig. 2.2 with 10^{15} cm^{-3} background p-type concentration, 10^{18} cm^{-3} p+/n+ concentrations, $w = 650$ nm, $h = 220$ nm, $t = 70$ nm, $d = 900$ nm, and $\lambda = 1550$ nm. This structure yields an effective modal area of 0.158 μm^2. **a)** The sum of $\langle N \rangle + \langle P \rangle$ is plotted. At low intensities, $\langle N \rangle + \langle P \rangle = \langle P \rangle$ is equal to the background doping. At high intensities, the sum is due almost completely to TPA. **b)** The mean of $\left\langle \tau_N^{\text{eff}} \right\rangle$ and $\left\langle \tau_P^{\text{eff}} \right\rangle$ is plotted. The high-intensity asymptote is the effective lifetime in the absence of the p+/n+ regions; it prevails at high intensities because the large number of TPA-generated free carriers serves to screen the electric field of the junction. At lower intensities, the junction serves to decrease the lifetime from the asymptotic value. Reverse biasing also helps to decrease the carrier lifetimes. **c)** The average lifetime of (b) plotted as a function of reverse bias at 100-mW intensity. A reverse bias of $\sim$5 V aids in decreasing carrier lifetime. Further increase of the reverse bias has only a minuscule effect in further decrease of the carrier lifetime.

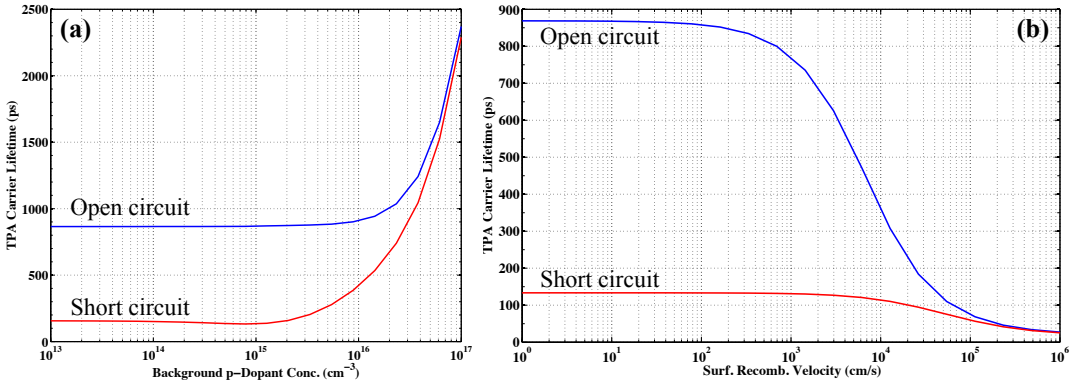

Figure 2.4: Simulation of the swept-carrier structure, as described in Fig. 2.3, with 100-mW optical intensity, showing the effects on TPA-generated carrier lifetimes from **a)** varied background dopant concentration and **b)** varied surface recombination velocity. Short circuit refers to zero applied bias across the junction; open circuit refers to zero current through the junction.

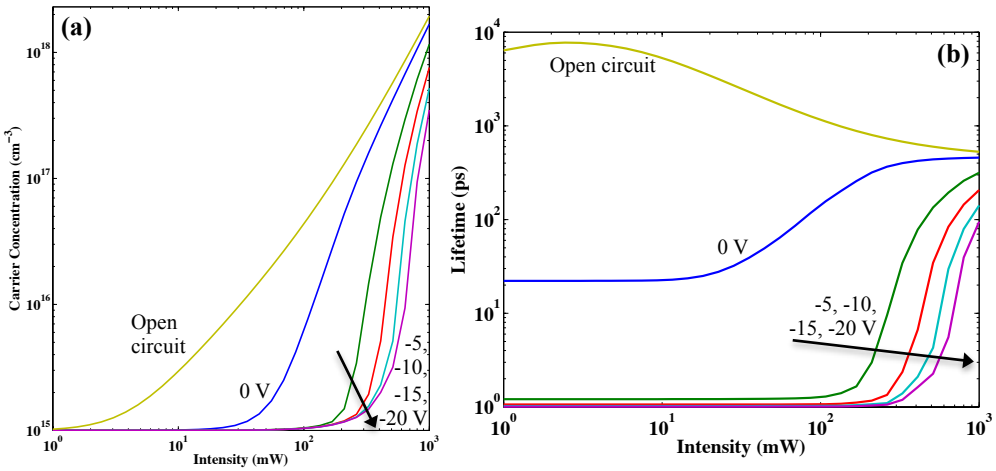

Figure 2.5: Simulation of the swept-carrier structure in Fig. 2.2, similar to Fig. 2.3 and with similar interpretations, but with $h = 340$ nm. For this example, the effective modal area is 0.172 μm^2.

2.4 Propagation loss

In this section, our goal is to predict the amount of loss that will be suffered in a given length of waveguide in the presence of loss due to linear scattering, TPA, and FCA. Combining (2.5) and (2.7) yields the governing equation that must be solved:

$$\frac{dI}{dz} = -\alpha_{\text{scat}} I - \alpha_{\text{TPA}} I^2 - \alpha_{\text{FCA}} I. \tag{2.14}$$

We have written the equation this way to signify that the FCA contribution, although momentarily appearing and sometimes being linear, will lead to higher order nonlinearities. Substituting (2.2), (2.10), and (2.12) into the foregoing, we obtain

$$\frac{dI}{dz} = -\alpha_{\text{scat}} I - \alpha_{\text{TPA}} I^2 - \left(\sigma_N \tau_N^{\text{eff}} + \sigma_P \tau_P^{\text{eff}}\right) \frac{\alpha_{\text{TPA}}}{2h\nu} I^3. \tag{2.15}$$

In writing the foregoing, we have recognized that

$$\alpha_{\text{FCA}} = \left(\sigma_N \tau_N^{\text{eff}} + \sigma_P \tau_P^{\text{eff}}\right) \frac{\alpha_{\text{TPA}}}{2h\nu} I^2 \tag{2.16}$$

is not constant with respect to intensity, whereas α_{scat} and α_{TPA} are true constants. Since the intensity is attenuated during propagation, α_{FCA} also becomes a spatially varying parameter. Furthermore, based on the results of Figs. 2.3.b and 2.5.b, τ^{eff} is also an intensity-dependent parameter thus making the nonlinearity of (2.15) possibly higher than order three. Although τ^{eff} can be modeled by a logistic function, a direct solution to (2.15) is still difficult. Instead, for the sake of analytical solutions, we describe a method of solution to (2.14) through a parameterization of α_{FCA}.

Without loss of generality, we consider the structure of Fig. 2.2 with a background p-type concentration of 10^{15} cm^{-3}, 10^{18} cm^{-3} in the heavily doped regions, $w = 600$ nm, $h = 250$ nm, $t = 90$ nm, $d = 900$ nm, and $\lambda = 1550$ nm. The average free-carrier lifetime is plotted in Fig. 2.6.a. Looking at this plot, we recognize that there are low and high intensity regimes where τ^{eff} is approximately constant; this makes $\alpha_{\text{FCA}} \propto I^2$ as per (2.16). There is also a medium intensity regime where τ^{eff} is approximately linear in I; this implies that $\alpha_{\text{FCA}} \propto I^3$. Indeed, we may use (2.2) to directly calculate α_{FCA}, as in Fig. 2.6.b. The calculation verifies

our assertion that

$$\alpha_{\text{FCA}} = \begin{cases} \gamma_1 I^2 & \text{Low intensity} \\ \delta I^3 & \text{Medium intensity} \\ \gamma_3 I^2 & \text{High intensity} \end{cases} \tag{2.17}$$

where γ_i and δ are fitting parameters as seen in Fig. 2.6.b. Therefore, we must solve

$$\frac{dI}{dz} = -\alpha_{\text{scat}} I - \alpha_{\text{TPA}} I^2 - \gamma_i I^3 \tag{2.18}$$

in the low ($i = 1$) and high ($i = 3$) intensity regimes and

$$\frac{dI}{dz} = -\alpha_{\text{scat}} I - \alpha_{\text{TPA}} I^2 - \delta I^4 \tag{2.19}$$

in the medium intensity regime.

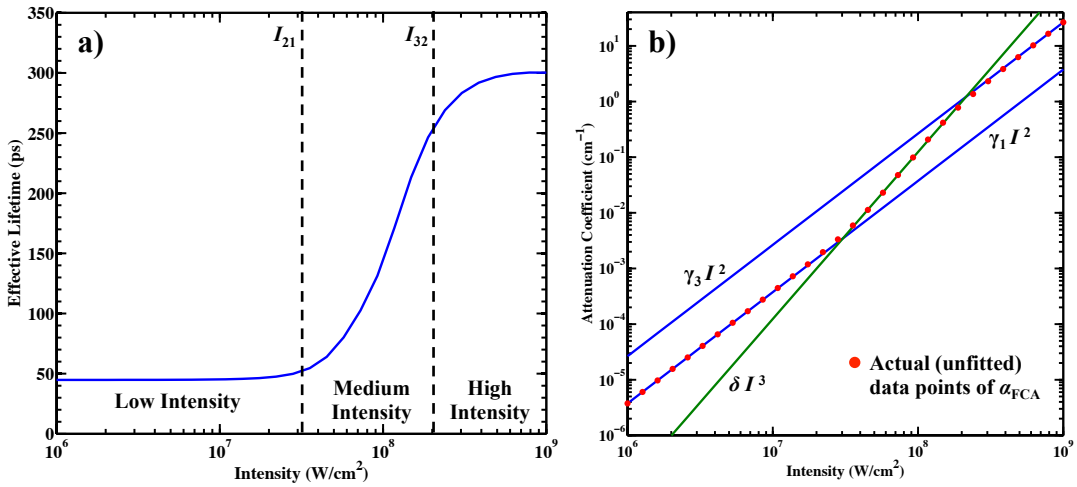

Figure 2.6: Simulation of the swept-carrier structure in Fig. 2.2 with a background p-type concentration of 10^{15} cm^{-3}, 10^{18} cm^{-3} in the heavily doped regions, $w = 600$ nm, $h = 250$ nm, $t = 90$ nm, $d = 900$ nm, and $\lambda = 1550$ nm. **a)** The average free-carrier lifetime is plotted with divisions signifying three district intensity regimes. I_{32} delineates the high and medium regimes; I_{21} delineates the medium and low regimes. **b)** The three intensity regimes carry over more apparently into the intensity-dependent FCA loss, which makes use of the individual carrier concentrations as per (2.2). The low and high intensity regimes correspond to the two regions where $\alpha_{\text{FCA}} \propto I^2$; the medium intensity regime corresponds to the region where $\alpha_{\text{FCA}} \propto I^3$.

Unfortunately, closed-form $I(z)$ solutions to (2.18) and (2.19) do not exist. We may, however, invert these equations and find analytical solutions to the inverse problem $z(I)$. Although unusual in the physical sense, solutions of the form $z(I)$ need not bother us in the modeling sense. As long as the inverse solution $z(I)$ exhibits a one-to-one relation between z and I, we may form a lookup table that effectively yields the desired information $I(z)$. For simplicity, we will first consider the case of $\alpha_{\text{scat}} = 0$; this greatly simplifies the solutions. We will completely relax this condition later on. Without loss of generality, we will assume that the initial input intensity is in the high regime. A piecewise solution is then obtained as the intensity is diminished to lower-intensity regimes. If the input intensity initially begins in a lower-intensity regime, we need only ignore higher-intensity regimes of the piecewise solution.

With an input intensity of $I_0 > I_{32}$ and $\alpha_{\text{scat}} = 0$, we must solve

$$\frac{dI}{dz} = -\alpha_{\text{TPA}} I^2 - \gamma_i I^3 \tag{2.20}$$

with $i = 3$. We may invert the foregoing and solve the integral equation

$$\int dz = -\int \frac{dI}{\alpha_{\text{TPA}} I^2 + \gamma_i I^3}. \tag{2.21}$$

Expanding the integrand on the right side,

$$z = k - \int dI \left[\frac{\gamma_i}{\alpha_{\text{TPA}}^2} \left(\frac{\gamma_i}{\alpha_{\text{TPA}} + \gamma_i I} - \frac{1}{I} \right) + \frac{1}{\alpha_{\text{TPA}} I^2} \right], \tag{2.22}$$

where k is a constant of integration. Completing the integration,

$$z = k + \frac{1}{\alpha_{\text{TPA}} I} - \frac{\gamma_i}{\alpha_{\text{TPA}}^2} \ln\left(1 + \frac{\alpha_{\text{TPA}}}{\gamma_i I} \right). \tag{2.23}$$

Invoking the boundary condition $I(z = 0) = I_0$ at the input, we can solve for k and arrive at the solution

$$z(I) = \frac{I^{-1} - I_0^{-1}}{\alpha_{\text{TPA}}} - \frac{\gamma_i}{\alpha_{\text{TPA}}^2} \ln\left(\frac{1 + \frac{\alpha_{\text{TPA}}}{\gamma_i I}}{1 + \frac{\alpha_{\text{TPA}}}{\gamma_i I_0}} \right). \tag{2.24}$$

This solution is valid from I_0 down to the point $I = I_{32}$; we define

$$L_{32} \equiv z(I_{32}) \tag{2.25}$$

as the corresponding spatial coordinate.

As we transition from the high to medium intensity regimes, we effectively begin with an initial intensity I_{32}; we must then solve

$$\frac{dI}{dz} = -\alpha_{\text{TPA}} I^2 - \delta I^4. \tag{2.26}$$

For simplicity in the notation, we will consider that $z(I_{32}) = 0$ and $z(I < I_{32}) > 0$; whatever $z(I)$ solution we obtain in this medium intensity regime, we must add to it L_{32}. As before, we invert the differential equation and expand pertinent terms into partial fractions to obtain

$$z = k - \int dI \left(\frac{1}{\alpha_{\text{TPA}} I^2} - \frac{\delta}{\alpha_{\text{TPA}}^2} \cdot \frac{1}{1 + \frac{\delta}{\alpha_{\text{TPA}}} I^2} \right). \tag{2.27}$$

Integrating,

$$z = k + \frac{1}{\alpha_{\text{TPA}} I} + \sqrt{\frac{\delta}{\alpha_{\text{TPA}}^3}} \tan^{-1} \left(\sqrt{\frac{\delta}{\alpha_{\text{TPA}}}} \cdot I \right). \tag{2.28}$$

Invoking the boundary condition $I(z = 0) = I_{32}$ at the input, we can solve for k and arrive at the solution

$$z(I) = \frac{I^{-1} - I_{32}^{-1}}{\alpha_{\text{TPA}}} - \sqrt{\frac{\delta}{\alpha_{\text{TPA}}^3}} \left(\tan^{-1} \left(\sqrt{\frac{\delta}{\alpha_{\text{TPA}}}} \cdot I_{32} \right) - \tan^{-1} \left(\sqrt{\frac{\delta}{\alpha_{\text{TPA}}}} \cdot I \right) \right). \tag{2.29}$$

This solution is valid from I_{32} down to the point $I = I_{21}$; we define

$$L_{21} \equiv z(I_{21}) \tag{2.30}$$

as the corresponding spatial coordinate.

As we transition from the medium to low intensity regimes, we effectively begin with an initial intensity I_{21} and we resort back to the differential equation (2.20), with solution (2.24), and $\gamma_i = \gamma_1$. In using this solution, we may make the substitution

$$I_0 \to I_{21} \tag{2.31}$$

and consider that $z(I_{21}) = 0$ and $z(I < I_{21}) > 0$; whatever $z(I)$ solution we obtain in this low intensity regime, we must add to it $L_{32} + L_{21}$.

In summary, the solution to the nonlinear loss problem of (2.14), with $\alpha_{\text{scat}} = 0$, is

$$z(I) = \begin{cases} \frac{I^{-1}-I_0^{-1}}{\alpha_{\text{TPA}}} - \frac{\gamma_3}{\alpha_{\text{TPA}}^2} \ln\left(\frac{1+\frac{\alpha_{\text{TPA}}}{\gamma_3 I}}{1+\frac{\alpha_{\text{TPA}}}{\gamma_3 I_0}}\right) & I_0 \geq I \geq I_{32} \\ L_{32} + \frac{I^{-1}-I_{32}^{-1}}{\alpha_{\text{TPA}}} - \sqrt{\frac{\delta}{\alpha_{\text{TPA}}^3}}(\tan^{-1}\left(\sqrt{\frac{\delta}{\alpha_{\text{TPA}}}} \cdot I_{32}\right) & \\ \qquad - \tan^{-1}\left(\sqrt{\frac{\delta}{\alpha_{\text{TPA}}}} \cdot I\right)) & I_{32} > I > I_{21} \\ L_{32} + L_{21} + \frac{I^{-1}-I_{21}^{-1}}{\alpha_{\text{TPA}}} - \frac{\gamma_1}{\alpha_{\text{TPA}}^2} \ln\left(\frac{1+\frac{\alpha_{\text{TPA}}}{\gamma_1 I}}{1+\frac{\alpha_{\text{TPA}}}{\gamma_1 I_{21}}}\right) & I \leq I_{21} \end{cases} \tag{2.32}$$

where I_0 is the initial input intensity at $z = 0$. If $I_0 < I_{32}$, then the first row is ignored and we make the substitutions $I_{32} \to I_0$ and $L_{32} \to 0$. If $I_0 < I_{21}$, then the first and second rows are ignored and we make the substitutions $I_{21} \to I_0$ and $L_{32} + L_{21} \to 0$.

Calculations of (2.32), using the data from Fig. 2.6, is plotted in Fig. 2.7 for several values of I_0, which correspond to the three different intensity regimes. This plot shows the clearly nonlinear behavior of the attenuation of highly intense optical radiation as it propagates down the waveguide. The larger intensities suffer much more attenuation as a result of contributing more free carriers via TPA. Normalizing these data by the propagation distance z yields an effective attenuation coefficient, as plotted in Fig. 2.8. This parameter, which lessens with propagation distance as the intensity lessens, is a useful metric for comparing the nonlinear loss with the usual linear loss parametrized by α_{scat}, which tends to be on the order of $\sim$1 dB/cm in a good fabrication process. From this normalization, we see that the high intensity suffers from a gigantic amount of loss for the first centimeter of propagation distance. The medium intensity suffers from a much smaller effective loss, but one that is still comparable to the linear loss. The low intensity adds only a small amount to the overall loss.

We now consider the solution of (2.14) with $\alpha_{\text{scat}} > 0$. In this case, we must solve the piecewise differential equations (2.18) and (2.19) depending on which intensity regime we are interested in. As before, the general method involves inverting (2.18) and (2.19) and integrating to find $z(I)$. For an initial input

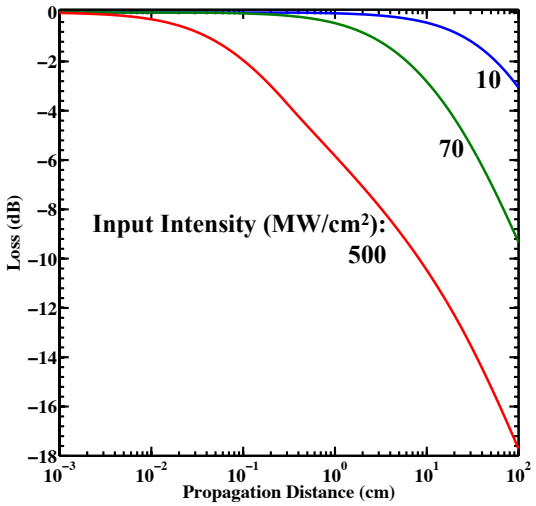

Figure 2.7: Calculations of (2.32) using the data from Fig. 2.6.

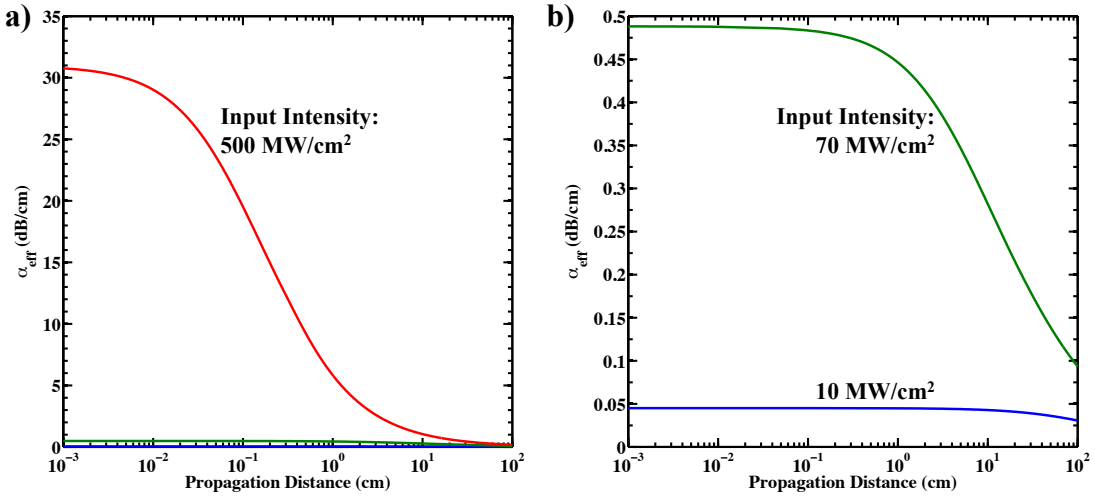

Figure 2.8: **a)** The data of Fig. 2.7 normalized by the propagation distance. This yields an effective attenuation coefficient that can be compared with the linear attenuation coefficient α_{scat}. **b)** A reduced-ordinate version of (a). Note that, in a good fabrication process, $\alpha_{\text{scat}} \sim 1$ dB/cm.

intensity I_0 at $z = 0$, the inverse solution is not pretty:

$$z(I) = \begin{cases} \frac{1}{\alpha_{\text{scat}}} \ln \frac{I_0}{I} - \frac{1}{2\alpha_{\text{scat}}} \ln \frac{\alpha_{\text{scat}}+\alpha_{\text{TPA}} I_0+\gamma_3 I_0^2}{\alpha_{\text{scat}}+\alpha_{\text{TPA}} I+\gamma_3 I^2} \\ \qquad -\frac{1}{\alpha_{\text{scat}}} \left(\frac{4\alpha_{\text{scat}}\gamma_3}{\alpha_{\text{TPA}}^2} - 1 \right)^{-1/2} (\tan^{-1} \frac{1+\frac{2\gamma_3 I_0}{\alpha_{\text{TPA}}}}{\sqrt{\frac{4\alpha_{\text{scat}}\gamma_3}{\alpha_{\text{TPA}}^2}-1}} \\ \qquad\qquad - \tan^{-1} \frac{1+\frac{2\gamma_3 I}{\alpha_{\text{TPA}}}}{\sqrt{\frac{4\alpha_{\text{scat}}\gamma_3}{\alpha_{\text{TPA}}^2}-1}}) & I_0 \geq I \geq I_{32} \\ L_{32} + \frac{1}{\alpha_{\text{scat}}} \ln \frac{I_{32}}{I} \\ \qquad -\frac{1}{\alpha_{\text{scat}}} \sum_{j:\delta j^3+\alpha_{\text{TPA}} j+\alpha_{\text{scat}}=0} \frac{\delta j^2+\alpha_{\text{TPA}}}{3\delta j^2+\alpha_{\text{TPA}}} \ln \frac{I_{32}-j}{I-j} & I_{32} > I > I_{21} \\ L_{32} + L_{21} + \frac{1}{\alpha_{\text{scat}}} \ln \frac{I_{21}}{I} - \frac{1}{2\alpha_{\text{scat}}} \ln \frac{\alpha_{\text{scat}}+\alpha_{\text{TPA}} I_{21}+\gamma_1 I_{21}^2}{\alpha_{\text{scat}}+\alpha_{\text{TPA}} I+\gamma_1 I^2} \\ \qquad -\frac{1}{\alpha_{\text{scat}}} \left(\frac{4\alpha_{\text{scat}}\gamma_1}{\alpha_{\text{TPA}}^2} - 1 \right)^{-1/2} (\tan^{-1} \frac{1+\frac{2\gamma_1 I_{21}}{\alpha_{\text{TPA}}}}{\sqrt{\frac{4\alpha_{\text{scat}}\gamma_1}{\alpha_{\text{TPA}}^2}-1}} \\ \qquad\qquad - \tan^{-1} \frac{1+\frac{2\gamma_1 I}{\alpha_{\text{TPA}}}}{\sqrt{\frac{4\alpha_{\text{scat}}\gamma_1}{\alpha_{\text{TPA}}^2}-1}}) & I \leq I_{21} \end{cases} \tag{2.33}$$

where the summation in the middle row is over the roots of the cubic equation appearing therein. If $I_0 < I_{32}$, then the first row is ignored and we make the substitutions $I_{32} \to I_0$ and $L_{32} \to 0$. If $I_0 < I_{21}$, then the first and second rows are ignored and we make the substitutions $I_{21} \to I_0$ and $L_{32} + L_{21} \to 0$.

As with the restricted solutions, we plot calculations of (2.33) in Fig. 2.9, for several values of I_0, using the data from Fig. 2.6. For these calculations, we assume $\alpha_{\text{scat}} = 1$ dB/cm. Compared to the restricted solutions, we see that including a non-zero value of α_{scat} causes a quicker convergence of the three plotted curves. Normalizing by the propagation distance z, as plotted in Fig. 2.10, shows that the three cases converge to α_{scat}. As the intensity is attenuated, TPA eventually becomes negligible and linear loss becomes the predominant extinction mechanism.

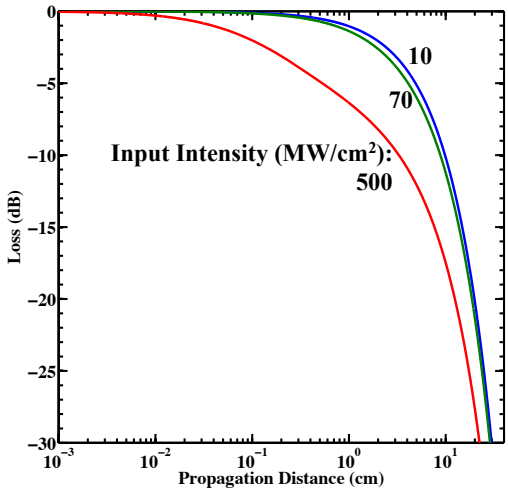

Figure 2.9: Calculations of (2.33) using the data from Fig. 2.6 and $\alpha_{\text{scat}} = 1$ dB/cm.

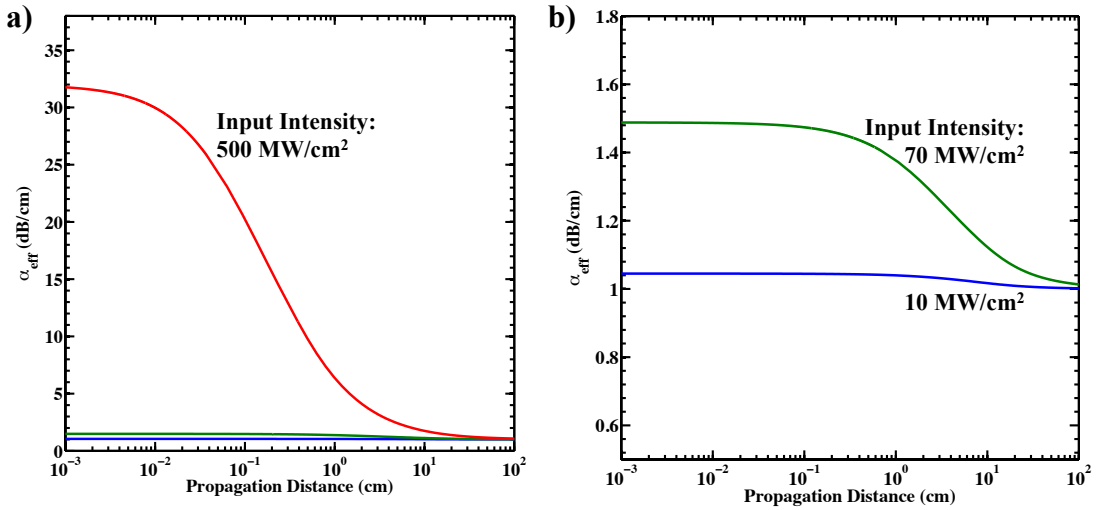

Figure 2.10: **a)** The data of Fig. 2.9 normalized by the propagation distance. This yields an effective attenuation coefficient that can be compared with the linear attenuation coefficient α_{scat}. **b)** A reduced-ordinate version of (a). Note that all three curves converge to α_{scat} as TPA becomes negligible due to the reduced intensity.

Chapter 3

Certain Aspects on the Design of Directional Couplers and Adiabatic Ring Resonators

3.1 Overview

The ring resonator is a fundamental component in chip-scale integrated optics [46, 47, 48]. On its own, the ring resonator is one of the simplest optical implementations of an add-drop filter. Silicon ring resonators, by virtue of resonant light exhibiting large intensities inside the resonator, and due in part to the tight mode confinement provided by the high index contrast of silicon photonics, allow for studies in nonlinear optics [25, 27, 49]. Lattices of coupled ring resonators give rise to high-order filters in which the filter transmission spectra can be tailored [50, 51, 52]. Various other types of configurations use ring resonators in more complicated coupling schemes to achieve more freedom in the filter design method [53, 54, 55, 56, 57, 58, 59]. At several points throughout the present work, we make use of a recently proposed architecture of a ring resonator, which is formed by a waveguide of non-constant width [60]. In the present chapter, we review key aspects of conventional ring resonators and the design of the newer implementation.

A related component, auxiliary to ring resonators, inter alia, is the direc-

tional coupler. In its simplest form, the coupler is merely two parallel waveguides in close proximity. The proximity of the waveguides allows for evanescent coupling of light in one waveguide to the other. As such, light can be routed from one waveguide to another. The topic of directional coupling in optics has a rich history of being well studied [61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72]. The standard textbook treatments of coupled mode theory [73, 74], as applied to directional couplers, treat one waveguide as a dielectric perturbation to the other, through which scattering between the modes may occur. Due to the mathematical complexities involved in coupled mode theory, finite-difference time-domain (FDTD) based methods are often done in which the simulated light is propagated down a three-dimensional structure. Both methods can be computationally expensive. Alternately, we champion an eigenvalue-based design method in which the existence of the normal (transverse) modes of the directional coupler are assumed a priori. This analysis yields a closed-form solution for the power-splitting ratio of the directional coupler, given that the modal refractive index eigenvalues are known. Such information can be easily obtained via a computationally inexpensive, cross-sectional mode solver. In what follows, we review the application of this method to the design of directional couplers.

3.2 Eigenvalue-based design of directional couplers

3.2.1 Basic theory and methodology

Suppose two single-mode waveguides are brought into close proximity of one another. The two waveguide modes exist as part of the two systems that are the individual waveguides. The newly formed directional coupler, however, may be considered as a system in its own right; this system exhibiting normal modes that are linear combinations of the two individual waveguide modes. The symmetric and anti-symmetric normal modes of the directional coupler may be expanded,

respectively, as

$$\begin{aligned} |s\rangle &= S_s|1\rangle + T_s|2\rangle, \\ |a\rangle &= S_a|1\rangle + T_a|2\rangle, \end{aligned} \tag{3.1}$$

where $|1\rangle$ and $|2\rangle$ are the modes of the two individual waveguides that constitute the coupler when they are each separately considered in isolation. Each eigenmode is normalized ($\langle i|i\rangle = 1;\ i = s, a, 1, 2$) and $|s\rangle$ and $|a\rangle$ form an orthonormal basis. The total field at an arbitrary point z throughout the length L of the coupler can be expanded in the basis of normal modes as

$$|E(z)\rangle = R_s e^{i\beta_s z}|s\rangle + R_a e^{i\beta_a z}|a\rangle, \tag{3.2}$$

where

$$\beta_i = \frac{n_i \omega}{c} \tag{3.3}$$

is the modal propagation constant and n_i is the modal refractive index for symmetric and anti-symmetric normal modes ($i = s, a$) and $0 \leq z \leq L$. The expansion coefficients $R, S,$ and T will be discussed below. For notational convenience, we also define the quantity

$$\mathcal{O} = \langle 2|1\rangle. \tag{3.4}$$

The amount of light in the mode $|2\rangle$ (i.e. in the second isolated-waveguide mode) is then a projection of $|2\rangle$ over the total coupler field:

$$\langle 2|E(z)\rangle = R_s(S_s\mathcal{O} + T_s)e^{i\beta_s z} + R_a(S_a\mathcal{O} + T_a)e^{i\beta_a z}. \tag{3.5}$$

We are most interested in the case in which light enters the coupler at $z = 0$ through Waveguide 1 (i.e. through mode $|1\rangle$) only; then, at $z = L$, (3.5) is a measure of how much light is transferred from $|1\rangle$ into $|2\rangle$. We thus have an expression for the dimensionless power coupling coefficient

$$\begin{aligned} |\kappa|^2 \equiv & |\langle 2|E(L)\rangle|^2 \\ = & \mathcal{O}^2[R_s^2 S_s^2 + R_a^2 S_a^2 + 2R_s R_a S_s S_a \cos((\beta_s - \beta_a)L)] \\ & + 2\mathcal{O}[R_s^2 S_s T_s + R_a^2 S_a T_a + R_s R_a (S_s T_a + S_a T_s)\cos((\beta_s - \beta_a)L)] \\ & + [R_s^2 T_s^2 + R_a^2 T_a^2 + 2R_s R_a T_s T_a \cos((\beta_s - \beta_a)L)]. \end{aligned} \tag{3.6}$$

Differentiating the foregoing with respect to wavelength gives the dispersion of the coupling coefficient

$$\begin{aligned}\frac{d|\kappa|^2}{d\omega} =& 2R_sR_a(S_sS_a\mathcal{O}^2 + (S_sT_a + S_aT_s)\mathcal{O} + T_sT_a) \\ &\times \frac{L}{c}(n_a^g - n_s^g)\sin((\beta_s - \beta_a)L),\end{aligned} \tag{3.7}$$

where $n_i^g = n_i + \omega\frac{dn_i}{d\omega}$ are the group indices of the symmetric ($i = s$) and anti-symmetric ($i = a$) normal modes. If weak coupling prevails, then $\mathcal{O}$ is vanishingly small [75] and (3.6) and (3.7) simplify to

$$\begin{aligned}|\kappa|^2 &= R_s^2T_s^2 + R_a^2T_a^2 + 2R_sR_aT_sT_a\cos((\beta_s - \beta_a)L), \\ \frac{d|\kappa|^2}{d\omega} &= 2R_sR_aT_sT_a\frac{L}{c}(n_a^g - n_s^g)\sin((\beta_s - \beta_a)L).\end{aligned} \tag{3.8}$$

Equation (3.8) is the main result. Regarding the expansion coefficients $R, S,$ and T – they are related via projections of the modal fields. For example,

$$R_s = \langle E(0)|s\rangle = S_s\langle E(0)|1\rangle + T_s\langle E(0)|2\rangle = S_s, \tag{3.9}$$

where, in the last step, we invoked the boundary condition that light input into the coupler is through Waveguide 1 only. Furthermore, if $\mathcal{O} \approx 0$, then

$$S_s \approx \langle s|1\rangle,\ T_s \approx \langle s|2\rangle. \tag{3.10}$$

Similarly,

$$R_a = S_a \approx \langle a|1\rangle,\ T_a \approx \langle a|2\rangle. \tag{3.11}$$

Therefore, knowledge of the isolated-waveguide and directional-coupler eigenmodes gives the coupling coefficient and its dispersion via substitution of (3.9)-(3.11) into (3.8).

As an example, consider a directional coupler made up of parallel, equal-geometry waveguides. Both of the normal modes should exhibit equal-power contributions from the two isolated-waveguide modes. Therefore, the (amplitude) expansion coefficients are $S_s = S_a = T_s = -T_a = 1/\sqrt{2}$. Substitution into (3.8) yields

$$|\kappa|^2 = \sin^2\left(\frac{(\beta_s - \beta_a)L}{2}\right) \tag{3.12}$$

and

$$\frac{d|\kappa|^2}{d\omega} = \frac{L}{2c}(n_a^g - n_s^g)\sin((\beta_s - \beta_a)L). \tag{3.13}$$

Equation (3.12) for the intensity coupling coefficient yields the immediate interpretation of the operation of the directional coupler as the beating two waves (the symmetric and anti-symmetric modes). As such, one may talk of the beat wavenumber

$$\beta_b \equiv \beta_s - \beta_a \tag{3.14}$$

(i.e. a *spatial* beat frequency) of the directional coupler. The beat period or, more appropriately named, beat length is then $L_b = 2\pi/\beta_b$; this is the length a directional coupler must be for no *net* transfer of power between the waveguides to occur (i.e. power transfers completely from one waveguide, to the other, and then completely back to the original waveguide). Of more practical interest is one-half of the beat length

$$L_c \equiv \frac{\pi}{\beta_s - \beta_a} = \frac{\lambda}{2(n_s - n_a)}; \tag{3.15}$$

this is the full-coupling length, or the length necessary for full power to transfer from one waveguide to the other. We note that L_c is proportional to $\lambda/2$.

Equation (3.12) also implies that the normal modes must propagate with different phase velocities for coupling to occur. Interestingly, there is no such constraint on the group velocities. In fact, (3.13) shows that if the normal modes propagated with the same group velocity, then the dispersion of the coupling coefficient would be completely mitigated.

As an example of using the ideas of this section in the design of a directional coupler, we consider the cross-sectional geometry of two, coupled, silicon ridge waveguides as diagrammed in Fig. 3.1.a. This geometry can be put into a numerical cross-sectional mode solver. The eigenvalue solutions that are obtained are the modal refractive indices of the symmetric and anti-symmetric normal modes[1],

[1] A cross-section mode solver (i.e. one that solves for *transverse* modes), for a given geometry, solves the Helmholtz equation, $\nabla_t^2 E = -\beta^2 E$, where the transverse (i.e. two-dimensional) Laplacian is utilized and E is the (three-dimensional) electric field vector. As such, the true eigenvalue solutions are the propagation constants β. The modal refractive indices, which are more common in practice, are simply a normalization of β as per (3.3).

as plotted in Fig. 3.1.b. Also plotted is the mode index for the single- (i.e. isolated) waveguide geometry. This is the refractive index that both waveguide modes would share if they were spaced reasonably far apart (instead of 300 nm, as in the diagram). The effect of evanescent coupling, between the two waveguides that compose the directional coupler, is that the degenerate single-waveguide mode index, which is shared by the two waveguides, is split into the two normal modes of the directional coupler.

With the knowledge of the refractive indices of the normal modes, one can proceed to substitute them into the relevant equations and design the length of the directional coupler. Invoking (3.15), we plot the full-coupling length in Fig. 3.1.c. Due to dispersion, as discussed above, we see that full coupling at one wavelength does not necessarily imply full coupling at another. To gain further insight into the dispersion of coupling, we invoke (3.12) and plot the intensity coupling coefficient, in Fig. 3.1.d, as a function of wavelength, for different lengths of the directional coupler. We see that longer lengths of the coupler lead to stronger dispersion of the coupling coefficient. Note that the longest coupling length that we have considered, 50 μm, is shorter than the shortest full coupling length plotted in Fig. 3.1.c. This implies that the argument of (3.12) is less than $\pi/2$. Longer coupling lengths would cause the curves in Fig. 3.1.c to "turn over," first exhibiting a local maximum, and then having completely negative slope, as the coupling coefficient is reduced back towards zero as the coupler length approaches the first beat length, and so forth. Further discussion of coupling dispersion[2] is provided in Chapter 4. Curves, such as the ones plotted in Fig. 3.1.c–d, are instrumental in properly designing a directional coupler.

[2] Dispersion, in the general-sense, has plagued optics since the time of Prof. Newton and when he decided that indigo was just as prominent as and easily distinguishable from blue and violet [76]. In the present work, the dispersion of directional coupling will be revisited on several occasions. We note in passing, however, that almost all optical phenomena will be, to some degree, dispersive.

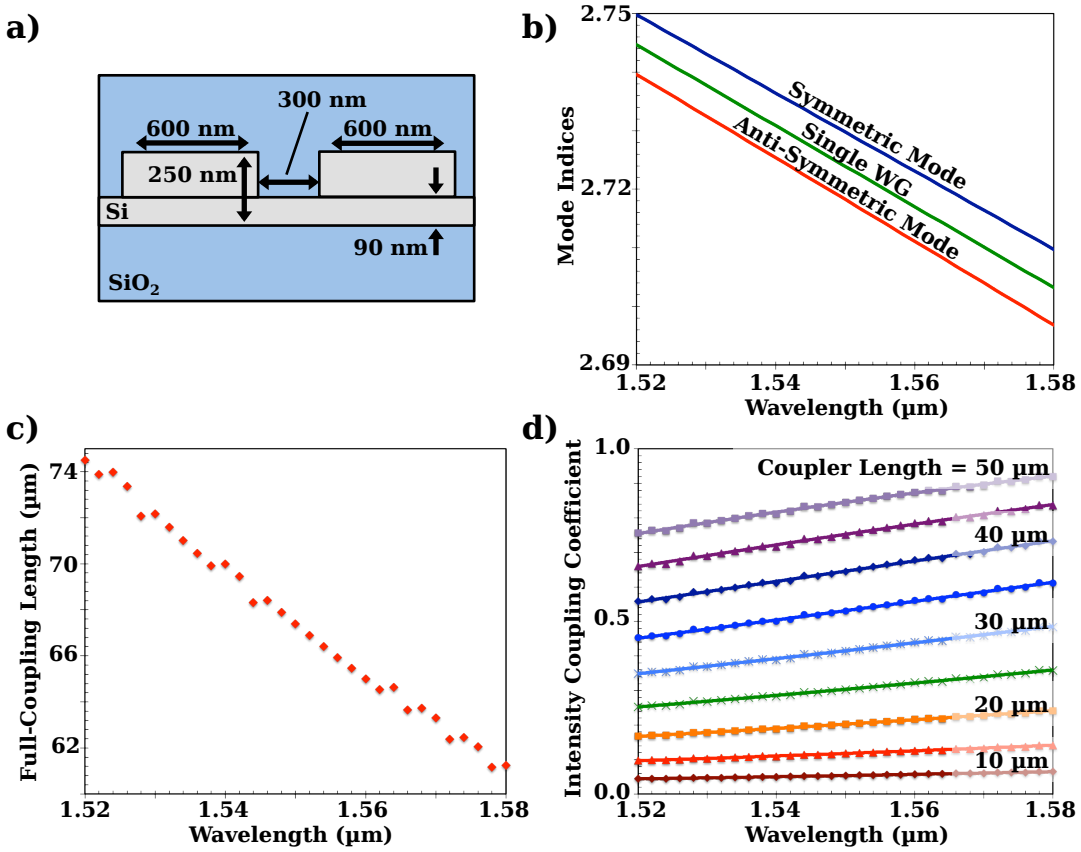

Figure 3.1: **a)** Cross-sectional diagram of a ridge-waveguide directional coupler. **b)** Refractive indices of the two directional-coupler normal modes (symmetric and anti-symmetric) as well as that of an isolated waveguide (WG). The evanescent coupling of the two waveguides lifts the degeneracy of the otherwise isolated waveguide modes into the the two non-degenerate normal modes. **c)** The full-coupling length (i.e one-half of the beat length), calculated by using the results of (b) in (3.15). **d)** Several dispersions of the intensity coupling coefficient, corresponding to different lengths of the directional coupler, calculated by substituting the previous results into (3.12).

3.2.2 Non-parallel sections

A realistic directional coupler, e.g. as in Fig. 3.2.b, is composed of both parallel and non-parallel sections. The analysis above is strictly applicable to the parallel section only. To generalize the analysis to the non-parallel waveguides, it is useful to consider that, in the parallel-waveguide case, the transfer of power between the two waveguides is due to the accumulation of phase

$$\phi_{\text{par}} = \int_0^{\phi_{par}} d\phi = \int_0^L \frac{1}{2}\beta_b dz = \frac{\beta_b L}{2}, \tag{3.16}$$

which is the argument of (3.12). With reference to the geometry of Fig. 3.2.b, it is therefore straightforward to generalize the foregoing to the non-parallel end sections of the directional coupler:

$$\phi_{\text{end}} = 2\int_0^{L_{end}/2} \frac{1}{2}\beta_{b,\text{end}}(z)dz, \tag{3.17}$$

where $\beta_{b,\text{end}}(0) = \beta_b$ and the factor of 2 occurs because there are two identical end sections. The intensity coupling coefficient of (3.12) then becomes

$$|\kappa|^2 = \sin^2\left(\frac{\beta_b L}{2} + \phi_{\text{end}}\right). \tag{3.18}$$

To determine ϕ_{end}, we assume a functional form for $\beta_{b,\text{end}}(z)$ [77],

$$\beta_{b,\text{end}}(z) = \beta_b e^{-g_{\text{end}}(z)/\gamma}, \tag{3.19}$$

where γ is a characteristic decay constant, which can be determined by modeling/measuring directional couplers of varying gaps, and $g_{\text{end}}(z)$ is the *additional* gap in the end section such that $g_{\text{end}}(0) = 0$ and $g(z) = g_{\text{straight}} + g_{\text{end}}(z)$. This assumption is justified because the field decays exponentially when moving away from the waveguide core, thus so too should the interaction decay as such when separating two waveguides. Furthermore, since we do not know L_{end} a priori, as a result of the rapid decease of (3.19), we may take the upper limit of integration in (3.17) to infinity, thus writing

$$\phi_{\text{end}} = \beta_b \int_0^{\infty} e^{-g_{\text{end}}(z)/\gamma} dz. \tag{3.20}$$

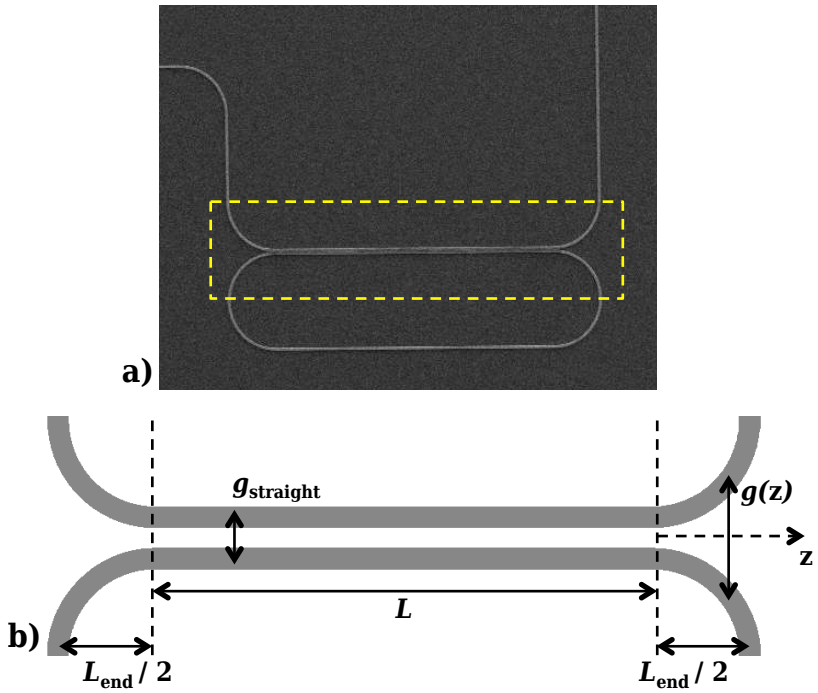

Figure 3.2: **a)** Ring resonator, evanescently coupled to a bus waveguide, as an example of a realistic application/analysis of a directional coupler. **b)** Diagram of the boxed section of (a); this is a realistic directional coupler in that the non-parallel sections are also considered. The total length of the directional coupler is $L + L_{\text{end}}$, where L is a physical (i.e. directly measurable) length, while L_{end} is an *effective* length that is proportional to the line integral of the waveguide gap $g(z)$. For the purpose of integration, $z = 0$ occurs at the origin of the as-drawn z-axis.

To find a closed-form solution to the foregoing, we may invoke a paraxial approximation. First, we recognize that, geometrically, each waveguide bend in Fig. 3.2 is one-quarter of a circle. Referring to the geometry of Fig. 3.3, we transform the integral over length to one over angle:

$$\phi_{\text{end}} = \beta_b R \int_0^{\pi/2} e^{-\frac{4R}{\gamma}\sin^2\frac{\theta}{2}} \cos\theta d\theta. \tag{3.21}$$

The physics of this equation is that the lessened exponential decay (relative to (3.20)) is aided by the $\cos\theta$ kernel, which acts as a type of obliquity factor (i.e. the coupling between the individual-waveguide modes tends to zero as their propagation vectors each bank through an angle that tends to 90°); thus the complete

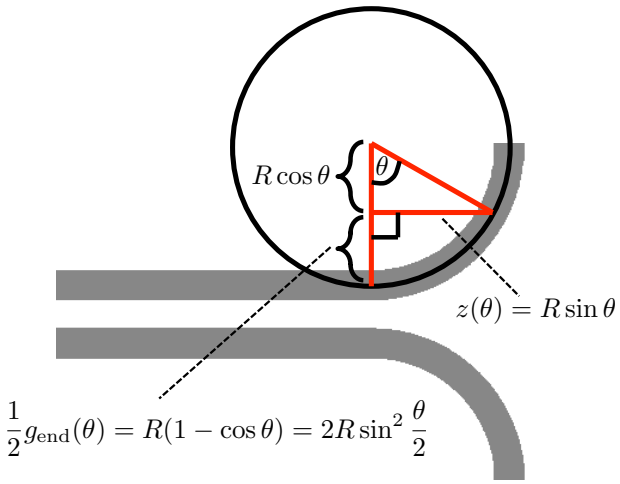

Figure 3.3: Right end of Fig. 3.2.b, showing the circular geometry of one of the waveguide bends. This geometry allows for the calculation of ϕ_{end} by transforming the integral over length to one over angle. The variables z and $g_{\text{end}}(z)$, as they appear in Fig. 3.2.b, are projected onto line segments inside the circle of radius R and parametrized with respect to the angle θ.

exponential decay seen in (3.20) is preserved. Under the paraxial approximation,

$$\theta \approx \frac{z}{R} \tag{3.22}$$

and

$$g(\theta) = 4R\sin^2\frac{\theta}{2} \tag{3.23}$$

$$\approx R\theta^2 \tag{3.24}$$

$$\approx \frac{z^2}{R}. \tag{3.25}$$

Therefore,

$$\phi_{\text{end}} \approx \frac{\beta_b}{2}\sqrt{\pi\gamma R}. \tag{3.26}$$

Substituting this result into (3.18), we see that the effect of the end sections is to give the directional coupler an effective length

$$L_{\text{eff}} = L + L_{\text{end}}, \tag{3.27}$$

where

$$L_{\text{end}} = \sqrt{\pi\gamma R}. \tag{3.28}$$

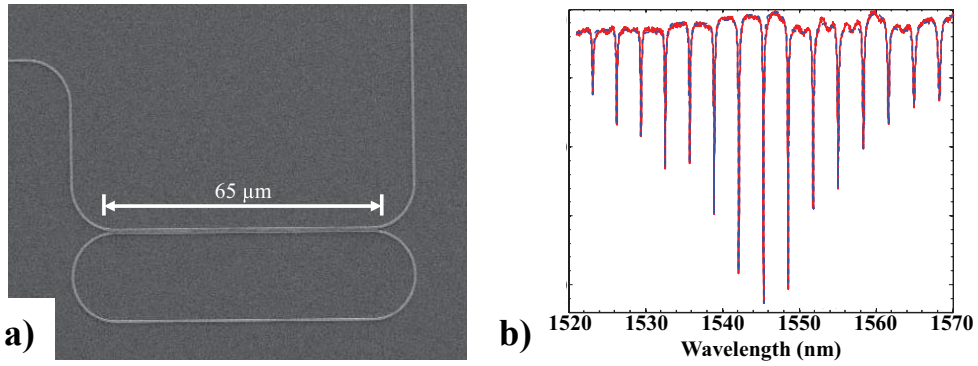

Figure 3.4: **a)** SEM micrograph of a standard implementation of a ring resonator. **b)** The corresponding transmission spectrum.

This is the same L_{end} that appears in (3.17), which was brushed underneath the rug in (3.20) because it is only known a posteriori. If, for example, the bottom waveguide in Fig. 3.2.b remained the same but the top waveguide were straight, even in the end sections, then a similar analysis to the foregoing would yield

$$L_{\text{end}} = \sqrt{2\pi\gamma R}. \tag{3.29}$$

Clearly, other geometries are possible.

3.3 Adiabatic ring resonators

3.3.1 A review of conventional ring resonators

A standard implementation of a ring resonator is shown in Fig. 3.4.a. This ring is in the "racetrack" configuration; i.e. two identical waveguide semi-circles sandwich a section of two parallel waveguides — the four waveguides actually being one and forming the ring. The (upper) bus waveguide is evanescently coupled to the ring. In the micrograph, the $\sim$65-μm region, where the bus waveguide and the ring are in close proximity, forms the directional coupler, which transfers light between the bus and the ring.

If optical loss in the ring is nonexistent, and because there is only one bus waveguide, the ring of Fig. 3.4.a would be an allpass filter; i.e. the transfer function of the ring would exhibit a flat unity magnitude response, albeit with a

non-trivial phase response. In the presence of loss, however, resonant light can be "dropped" via the loss mechanism thus yielding a resonant spectrum as seen in Fig. 3.4.b. Further details concerning the transmission spectra of ring resonators are covered in Chapter 4. For now, we concern ourselves with the calculation of the free spectral range (FSR) and, especially, adiabatic ring resonators, which we shall define shortly.

The usual kinematic phase condition for the m^{th} resonance (i.e. a null in Fig. 3.4.b) is

$$\phi = 2\pi m, \tag{3.30}$$

where m is an integer. The phase ϕ is related to the propagation constant β of the mode and the length L of the resonator via

$$\begin{aligned} \phi = \beta L &= n\frac{\omega}{c}L \\ &= n\frac{2\pi}{\lambda}L, \end{aligned} \tag{3.31}$$

where β has been expanded into terms involving the velocity of light c, angular frequency $\omega = 2\pi\nu$, free-space wavelength λ, and the modal refractive (phase) index n. At present, we are interested in the FSR, which is the spacing between adjacent resonances in the transmission spectrum of the resonator. Combining the foregoing two equations then subtracting the $m+1$ resonant frequency from the m^{th} resonant frequency yields the following expressions for the FSR in terms of either frequency or wavelength, respectively:

$$\begin{aligned} \Delta\nu &= \frac{c}{n_g L}, \\ \Delta\lambda &= \frac{\lambda^2}{n_g L}, \end{aligned} \tag{3.32}$$

where

$$n_g = n + \omega\frac{dn}{d\omega} \tag{3.33}$$

is the modal group index, which we have obtained by recognizing that the phase index n is a dispersive function of frequency.

Of paramount importance is (3.32). For example, from Fig. 3.4, we measure the ring length to be 186 μm and the FSR to be 3.24 nm; therefore, by (3.32) we

know the group index to be 3.99, without ever specifying details of the waveguide cross-section. For the purpose of filter design, however, we usually want to meet certain periodicity requirements in the transmission spectrum. In this case, the prescription is to first design the waveguide cross-section, which yields a specific n_g, then to chose the resonator length that yields the desired FSR as per (3.32).

3.3.2 Geometry and free spectral range of adiabatic rings

We now move on to the topic of adiabatic ring resonators [60], of which two examples are diagrammed in Fig. 3.5. Concentrating on the first example, this structure is a resonator in which the waveguide width transitions adiabatically between w_1 at its narrowest points to w_2 at its widest points. The purpose of such a transition is to allow electrical contact to be made directly to the ring via tethers that are located internal to the ring. For example, the entire structure of 3.5.a being silicon, the tethers, as well as narrow interior portions of the waveguide at both of its widest regions, could be electrically doped thus forming a resistive heater. Such a scheme allows for an efficient thermo-optic effect. Ideally, light suffers from little scatter loss in this structure because it hugs the outer wall of the waveguide as it propagates around a bend. To design such an adiabatic transition, the outer wall of the ring is geometrically a circle with radius r, while the inner wall is an ellipse with semi-major axis a and semi-minor axis b; the circle and ellipse sharing a common center. The circular and elliptical parameters are connected via

$$\begin{aligned} w_1 &= r - a \\ w_2 &= r - b. \end{aligned} \tag{3.34}$$

Furthermore, the nominal radius R (i.e. the radius that would track the center of the waveguide if $a = b$) is

$$R = r - \frac{w_1}{2}. \tag{3.35}$$

The second example, Fig. 3.5.b, shows an adiabatic ring resonator in the racetrack configuration. As with the conventional ring resonator, the present structure is just a non-racetrack resonator, split in half, with both ends sandwiching a parallel straight waveguide section of length L_s. The geometric parameters in

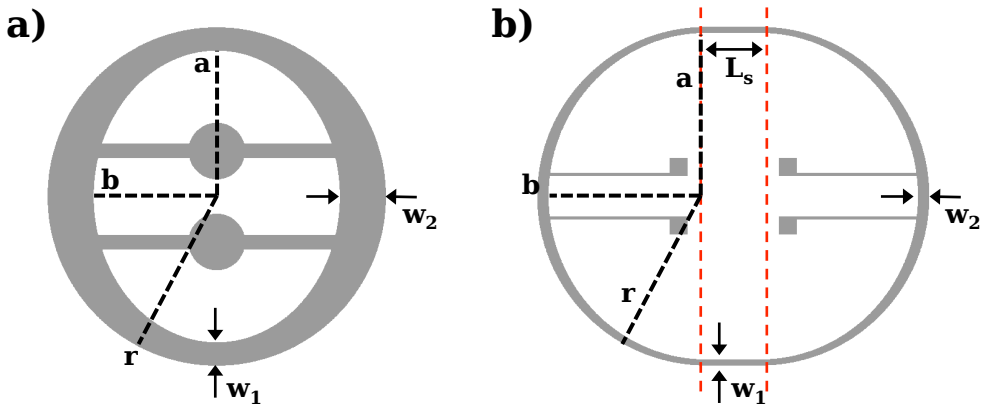

Figure 3.5: **a)** Design of an adiabatic ring resonator. The outer wall takes the shape of a circle with radius r. The inner wall is an ellipse with semi-major axis a and semi-minor axis b. The waveguide that composes the ring transitions from a width of w_1 at the narrowest points to w_2 at the widest points. The tethers, interior to the ring, provide for low-optical-loss electrical contact to a higher level metal layer. **b)** An adiabatic ring resonator in the "racetrack" configuration. As with convention ring resonators, a non-racetrack adiabatic resonator is split in half and a straight-waveguide section, of length L_s, is sandwiched between. The straight waveguides retain constant widths w_1. The other geometric parameters remain as before.

this structure carry over from before. Additionally, the straight section contains waveguides of constant width w_1; i.e. the adiabatic transition does not occur until one is outside of the straight section.

We now want to calculate the FSR of an adiabatic ring resonator; the FSR being dependent on n_g as per (3.32). As discussed above, n_g is related to β, which is the eigenvalue associated with the transverse modes; i.e. the value of n_g is dependent on the waveguide cross-section normal to the direction of propagation. This poses a dilemma because the waveguide cross-section changes as the width changes. To rectify this situation, we first rewrite the phase relation (3.31) as

$$\phi = \int_0^L \beta dl, \tag{3.36}$$

where the integration is performed over the length of the ring. Invoking the geometry of Fig. 3.5.b,

$$\phi = 2\beta\left(w_1\right) L_s + R\int_0^{2\pi} \beta d\theta, \tag{3.37}$$

where β is written as a function of waveguide width, which can be parameterized with respect to the angle θ subtended when propagating around the ring. In the foregoing, setting $L_s = 0$ yields the equation appropriate for Fig. 3.5.a. In analogy with (3.31), we may also write

$$\phi = \beta_{\text{eff}} L, \tag{3.38}$$

where β_{eff} is an effective, or path-integrated, propagation constant that yields the correct phase result for light propagating through the ring. Combining the foregoing with (3.37) yields

$$\beta_{\text{eff}} = 2\beta\left(w_1\right)\frac{L_s}{L} + \frac{R}{L}\int_0^{2\pi}\beta d\theta. \tag{3.39}$$

Finally, invoking the relation

$$n_g = \frac{c}{\frac{d\omega}{d\beta}} \tag{3.40}$$

yields an effective, or path-integrated, group index

$$n_g^{\text{eff}} = 2\frac{L_s}{L}n_g\left(w_1\right) + \frac{R}{L}\int_0^{2\pi} n_g d\theta, \tag{3.41}$$

where it is now n_g that is written as a function of waveguide width or parameterized with respect to angle subtended. This equation solves the problem we are interested in; n_g^{eff} can be substituted into (3.32) to determine the FSR of an adiabatic ring resonator. We note that, specifying f to be the fraction of the total length of the ring resonator that is a straight waveguide, we can rewrite the foregoing as

$$n_g^{\text{eff}} = fn_g\left(w_1\right) + \left(1-f\right)\frac{1}{2\pi}\int_0^{2\pi} n_g d\theta. \tag{3.42}$$

This has the useful interpretation that n_g^{eff} is simply the weighted sum of n_g around the length of the resonator.

Calculating n_g^{eff} via (3.42) is a straight-forward extension of performing waveguide-mode solving. For example; considering a silicon strip waveguide, 230-nm thick, surrounded by SiO_2, and a wavelength of 1550 nm; a mode solver yields solutions for the phase and group indices when varying the width of the waveguide, as plotted in Fig. 3.6. More accurately, the phase index is the normalized eigenvalue solution, while the group index is easily calculated from the former via (3.33).

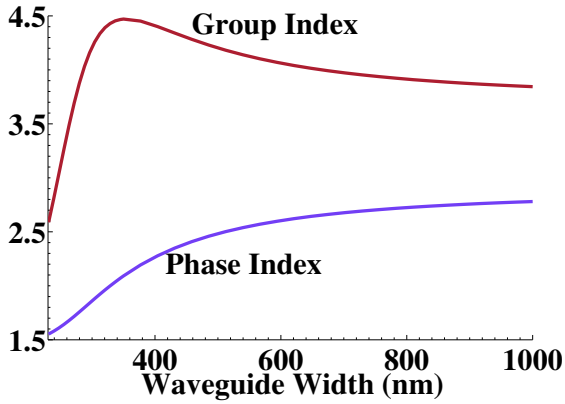

Figure 3.6: Modal refractive indices for an oxide-clad, 230-nm-thick, silicon strip waveguide, of varying width, at a wavelength of 1550 nm.

Having, what computationally amounts to, a lookup table for n_g^{eff} allows for (3.42) to be solved numerically.

We now show two experimentally confirmed examples of using the foregoing methods. For the first example, we take the ring of Fig. 3.5.a, iterated four times, and arranged in the configuration as shown in Fig. 3.7.a–b. For this example, $w_1 = 325$ nm, $w_2 = 650$ nm, $R = 2.25$ μm, and the waveguides are 230-nm thick so that we may use the previous result of group index vs. width. Invoking (3.42) yields $n_g^{\text{eff}} = 4.22$ and an FSR of 5.03 THz, which is consistent with the experimental measurement of Fig. 3.7.c.

As a second example, we use the ring of Fig. 3.5.b, arranged as a second-order lattice filter, as shown in Fig. 3.8.a. For this example, $w_1 = 400$ nm, $w_2 = 800$ nm, $R = 12.3$ μm, $L_s = 4.018$ μm, and the waveguides are again 230-nm thick. Invoking (3.42) yields $n_g^{\text{eff}} = 4.14$ and an FSR of 849 GHz, which is consistent with the experimental measurement of Fig. 3.8.b.

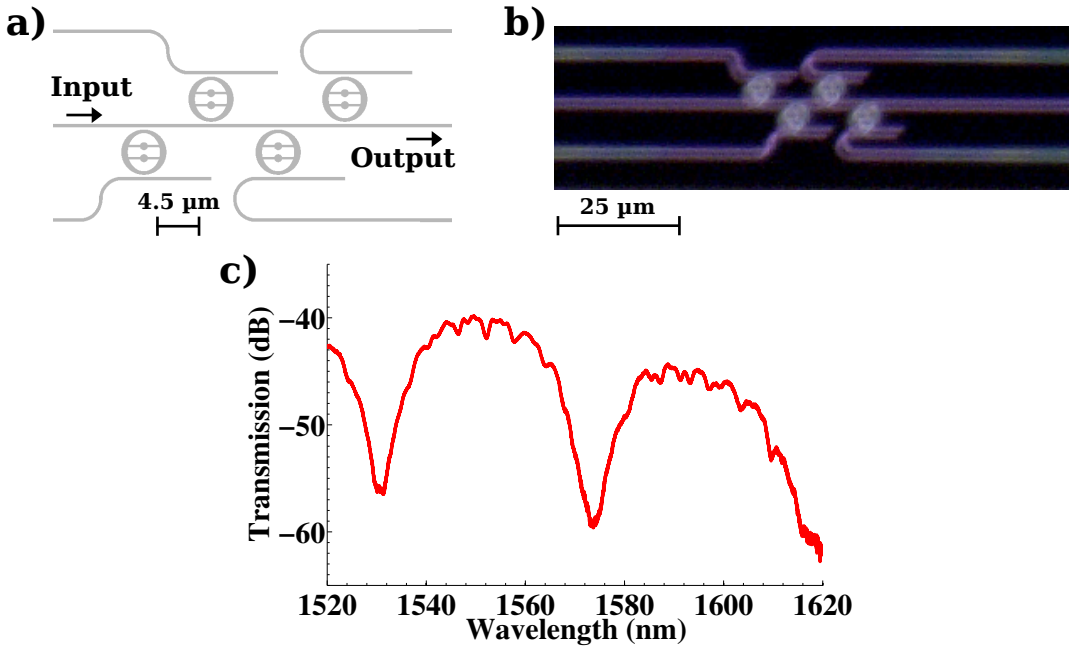

Figure 3.7: **a)** Diagram of a filtering scheme using the type of ring resonator diagrammed in Fig. 3.5.a. **b)** Dark-field micrograph of a fabricated device using the foregoing design. **c)** The corresponding measured transmission spectrum.

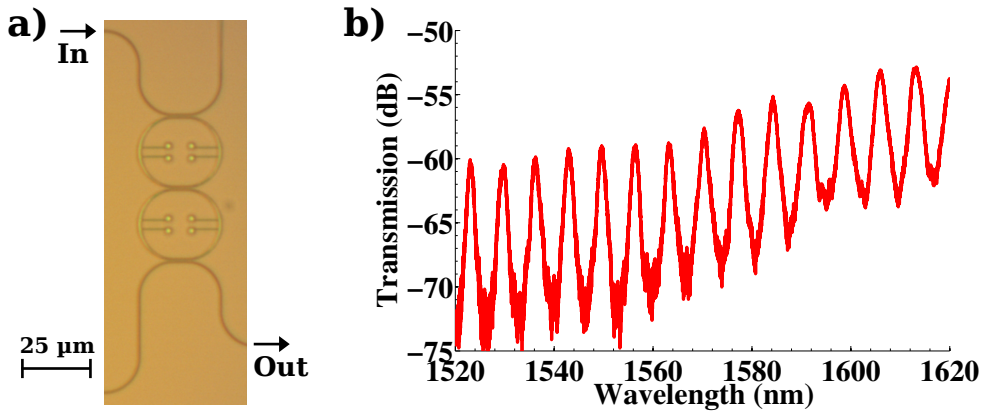

Figure 3.8: **a)** Bright-field micrograph of a fabricated device that uses the type of ring resonator diagrammed in Fig. 3.5.b. **b)** The corresponding measured transmission spectrum.

Chapter 4

An Investigation of Dispersion in Directional Couplers and a Wideband Parameter Extraction of Ring Resonators

4.1 Overview

Silicon ring resonators are key building blocks for modulators, filters, and resonantly-enhanced detectors. A wideband parametric characterization of these resonators is important for expanding their spectral ranges and total usable bandwidths for use in on-chip interconnects. Extending reported measurements that focus on fitting a transmission spectrum with Lorentzian functions throughout only one or two free spectral ranges (FSR) [46, 80, 81, 82, 83], we perform a wideband characterization that spans 15 FSRs and demonstrate that coupling-coefficient dispersions considerably exceed modal refractive-index dispersions in silicon directional couplers.

The device analyzed here is a single-waveguide-coupled silicon ring resonator, as shown in Fig. 4.1.a. Aside from being useful in its own right, such a device can be utilized as a common test structure in silicon-photonic integrated

circuits. Such test structures can exhibit relatively small footprints and the lack of multiple couplers eliminates the possibility of the confounding of multiple coupling coefficients [84]. Ridge waveguides were fabricated on a silicon-on-insulator wafer (250-nm silicon thickness) using electron-beam lithography and reactive ion etching (160-nm etch depth). Using a scanning electron microscope (SEM), the measured perimeter of the ring/racetrack was 186 μm with a 10-μm bending radius along the curves; waveguide widths were 530 nm (bus) and 550 nm (ring). An effective coupler length of 65 μm was inferred by matching finite-element-method (FEM) simulations of the measured geometry to the coupling-coefficient data obtained below. Light was coupled to the chip via polarization-maintaining tapered lensed fibers and polymer spot-size converters overlaid on tapered silicon waveguides. The transmission spectrum (Fig. 4.1.b) was measured using a swept-wavelength laser and photoreceiver.

4.2 Ring-resonator parameter extraction

Here, an algorithm is developed that is able to extract the coupling and through coefficients and loss of a ring resonator, that is coupled to a single waveguide, over arbitrarily many FSRs. Previous work on the single-bus structure followed a different track of focusing on the loss of the ring [85, 86]. Another study proposed a method for extracting the coupling coefficients of the single-bus resonator [80]; however, the model used did not focus on spanning and reconstructing several FSRs [46]. Other methods for extracting the ring parameters require that the resonator be coupled to two waveguides [80, 84, 87].

In developing the computational algorithm, we modified the previously published result [47], for the intensity transmission of a ring resonator coupled to a single waveguide, to include the effects of the intensity loss that is incurred in the coupling region. The intensity transmission is

$$|H(\lambda)|^2 = \left| \frac{t - \epsilon a e^{i\beta L}}{1 - a t^* e^{-i\beta L}} \right|^2, \tag{4.1}$$

where $\beta = \frac{2\pi n_{\text{eff}}}{\lambda}$, λ is the wavelength, n_{eff} is the effective mode index, L and a

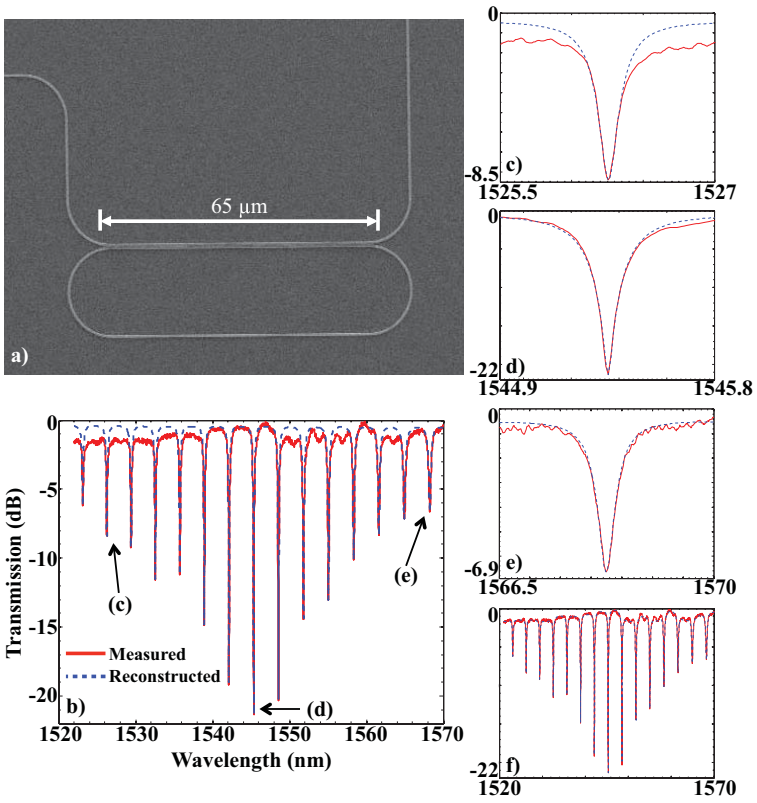

Figure 4.1: **a)** SEM micrograph of the ring resonator. **b)** Measured spectrum and its reconstruction via iterations on t and a. **c–e)** Lineshapes corresponding to the labels in (b). **f)** Measurement and its reconstruction with iterations on ϵ added.

are the perimeter and round-trip loss factor of the ring, κ and t are the length-integrated cross- and through-coupling coefficients of the field amplitude and $\epsilon = |t|^2 + |\kappa|^2$ characterizes the coupling intensity loss. Note that, except for L, the foregoing parameters are wavelength dependent. To determine the wavelength variations of a, t, and κ, we fit (4.1) to the experimentally measured data of Fig. 4.1.b, FSR-by-FSR, in a piecewise-continuous manner. Use of (4.1) is more accurate than Lorentzian fitting because the latter is valid only close to resonance [46].

Using (4.1), the values of a and t are iteratively computed at each resonant wavelength while inserting the measured values of $|H|^2$ into the left-hand side. At the resonances, the exponentials in (4.1) equal unity. Initial test values of a then determine the first-iteration values of t at the resonances. Interpolations of a and

t over the entire spectral domain then allow for the calculation of a transmission spectrum, i.e. a reconstructed $|H|^2$, using (4.1).

At this stage, the reconstructed $|H|^2$ exhibits identical resonant locations and depths as the measured $|H|^2$; however, the linewidths of the measured vs. calculated resonances will not match. The test values of a are then altered iteratively until the linewidths match. This completes one full sweep of t- followed by a-iterations. Following the sweep, the reconstructed depths are altered due to the iterations on a; thus, several more full sweeps are performed until convergence, yielding a reconstructed $|H|^2$ that matches the measured spectrum. Figures 4.1.b–e show the good agreement in matching the locations, depths, and widths of the resonances at distant wavelengths throughout the C-band. Three full sweeps over the entire spectrum were performed, with an average number of iterations at each resonance of 24, 19, and 14 for the first through third sweeps. The convergence condition was that both t and a converged to a relative error of less than 10^{-6}.

A linear expansion of the mode index is assumed:

$$n_{\text{eff}} = n_0 + (\lambda - \lambda_0)\frac{dn_{\text{eff}}}{d\lambda}, \tag{4.2}$$

where the derivative was approximately constant over the spectral domain of interest. The group index,

$$n_g = n_{\text{eff}} - \lambda\frac{dn_{\text{eff}}}{d\lambda} = 3.99, \tag{4.3}$$

was determined by calculating the FSR of the measured $|H|^2$; $\frac{dn_{\text{eff}}}{d\lambda} = -0.897\ \mu\text{m}^{-1}$ was computed via FEM simulations of the waveguide geometry.

A better fit in the off-resonant areas of the reconstructed $|H|^2$ to the measured $|H|^2$ can be obtained by including FSR-by-FSR iterations on ϵ, as in Fig. 4.1.f. This additional routine, however, leads to several unphysical values of $a > 1$. Such behavior occurs because the shapes/depths of the resonances are dependent on the product ϵa (e.g. critical coupling occurs when $t = \epsilon a$), whereas the baseline is primarily dependent on ϵ only. Once the baseline clamps the value of ϵ, the algorithm is forced to choose the unique, and possibly unphysical, value of a that conserves the loss product ϵa. Therefore, the baseline of the measured $|H|^2$ that fluctuates between 0 and -2 dB, which has been observed by others [81, 82],

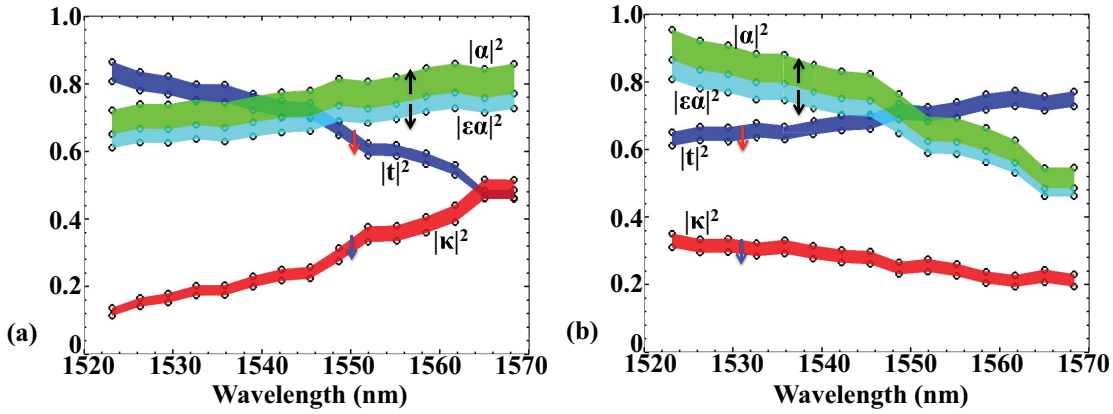

Figure 4.2: Extracted dispersions of the ring-resonator parameters. Mathematically, two sets of solutions are possible, both of which yield the same transmission spectrum. **a)** The solution corresponding to the fabricated devices. **b)** An alternate solution, which physically would correspond to a different device geometry.

does not occur solely due to the loss that is suffered in the directional coupler, but also due to loss mechanisms that are external to the coupler, e.g. fiber/multimode coupling, waveguide disorder, etc.

4.3 Dispersions of the ring-resonator parameters

Using the method described above, the extracted parameters are shown in Fig. 4.2. The parameters a and t were determined from the iteration algorithm and $|\kappa|^2 = \epsilon - |t|^2$. The shaded bands indicate parameter solutions corresponding to different assumed values of ϵ. The band edges correspond to $\epsilon = 1.0$ (arrow tails) and $\epsilon = 0.92$ (arrow heads), which represents the ideal lossless case and the typical loss incurred in a fused coupler [87], respectively. The parameters shift monotonically in the direction of the arrows as ϵ decreases from 1.0 to 0.92.

Owing to the fact that (4.1) exhibits two zeros, there are actually two convergent solutions to the algorithm [80, 85, 86]. The two solutions correspond to the case where the deepest resonance is either slightly under-coupled or slightly over-coupled; the solutions cannot be distinguished from a phase insensitive measurement of $|H|^2$ but may be inferred by the expected parameter trends based on the design. Note that a certain symmetry is exhibited by (4.1), and seen in

Fig. 4.2, in that the $|t|^2$ band in one solution set becomes the $|\epsilon a|^2$ band in the other solution set, and vice-versa. We determine that the results in Fig. 4.2.a are the correct parameter dispersions for the device in Fig. 4.1.a based on our design goal and FEM simulations.

From Fig. 4.2.a, we calculate a coupling-coefficient dispersion $d|\kappa|^2/d\lambda \approx 8\ \mu\text{m}^{-1}$; this is much larger in magnitude than the index dispersion $dn_{\text{eff}}/d\lambda = 0.897\ \mu\text{m}^{-1}$. To explain why the coupling dispersion is much larger than the index dispersion, we can invoke a coupled-mode formalism to show that

$$\frac{d|\kappa|^2}{d\lambda} \approx -4\pi I_{s1} I_{a1} I_{s2} I_{a2} \frac{L_c}{\lambda^2} \left(n_a^g - n_s^g\right) \sin\left(\frac{2\pi L_c}{\lambda}\left(n_s - n_a\right)\right), \tag{4.4}$$

where L_c is the length of the coupler, n_i and n_i^g are the phase and group refractive indices, and

$$I_{ij} = \int E_i^* \cdot E_j d^2 r \tag{4.5}$$

are overlap integrals over the waveguide modes E_i with the normalization that $I_{ii} = 1$; s, a, 1, and 2 indicate the symmetric and anti-symmetric supermodes and the isolated modes of the two waveguides that constitute the coupler, respectively. In a symmetric coupler, the product of the four I's equals $-\frac{1}{4}$; this product tends towards zero for asymmetries in the coupler geometry and, for our device, is -0.64, based on FEM simulations of the coupler and dimensional measurements using SEM. This analysis shows that large coupling dispersion is caused by the large prefactor $\pi L_c/\lambda^2 \sim 85\ \mu\text{m}^{-1}$, in addition to phase or group velocity mismatch between the supermodes.

The magnitude of (4.4) can be decreased by designing the two waveguides, which constitute the coupler, to have different cross-sectional geometries. Such a scheme decreases the magnitude of $\prod_{i,j} I_{ij}$; however, increased asymmetry of the waveguide geometries increases the $n_s - n_a$ difference.

Coupling dispersion may also be decreased by causing the values of n_s and n_a to approach each other. This is possible for a symmetric coupler by increasing the waveguide separation because the supermodes degenerate to isolated waveguide modes as the separation tends to infinity. Such a scheme, however, is impractical because the perturbation of one waveguide on the other is then exponentially

weaker; longer coupling lengths would be needed for a desired $|\kappa|^2$. Furthermore, as $n_s \rightarrow n_a$, (4.4) becomes proportional to a larger prefactor of $(\pi L_c)^2/\lambda^3$. An alternative approach to reducing the magnitude of (4.4) may be in causing n_s^g and n_a^g to approach each other; i.e. one can match either the phase velocities or group velocities of the supermodes.

4.4 Summary

A wideband parameterization algorithm for silicon microring resonators has been presented. The resultant parameter extraction has shown that the coupling dispersions of silicon directional couplers considerably exceed the modal refractive-index dispersions in magnitude. The analysis extends over 15 FSRs and covers the entire C-band. The results have also yielded valuable insight into coupling dispersion and may pave the way for future designs of wideband, dispersion-compensated, silicon photonic circuits.

Chapter 4, in part, contains material, published in the following, of which the dissertation author was the primary investigator:
R. Aguinaldo, Y. Shen, and S. Mookherjea, "Large Dispersion of Silicon Directional Couplers Obtained via Wideband Microring Parametric Characterization," *IEEE Photon. Technol. Lett.*, vol. 24, pp. 1242–1244, July 2012 [88].

Chapter 5

Channelized Spectrum Monitoring

5.1 Overview

In this chapter, we describe a device that can be described as a channelized spectrum monitor (CSM). The purpose of this device is to replace the optical spectrum analyzer (OSA) that may be found in an optical data center network or similar. While by no means an on-chip OSA, the CSM represents a functional equivalent to the OSA when used in a network that operates with known wavelength channels that originate from known sources. In this case, the only new information that the OSA can provide is how much power is carried in each channel. The CSM that we describe here provides this very same information, but with many orders-of-magnitude reduction in size, weight, and power. While rack-friendly optical performance monitors and channel monitors do exist commercially [89], these conventional off-the-shelf (COTS) components tend to operate on 100-μs time scales and still represent bulk optical components. The silicon-photonic solution that we describe here has the potential to be several orders-of-magnitude faster than the COTS solutions via the monolithic integration of high-speed Ge photodetectors. Additionally, a silicon-photonic solution allows for the monolithic integrability with other on-chip subsystems.

As motivation for the CSM, we turn to the MORDIA network at the University of California, San Diego. MORDIA (Microsecond Optical Reconfigurable Datacenter Interconnect Architecture) is an optical circuit-switched, dense wavelength division multiplexed (DWDM), ring network [90]. It supports up to 24 hosts, representing 24 wavelength channels located throughout the C-band on the 100-GHz-spaced ITU-T telecommunications grid, from channels 15 (1565.50 nm) to 58 (1531.12 nm). Each channel carries 10-Gbps data streams. The network is set up in a ring topology with six nodes and four hosts per node; however, 23 of the 24 hosts are utilized in the current implementation.

The MORDIA hardware is shown in Fig. 5.1.a. The eyesore in this picture is the OSA, which takes up more than its fair share of rack-space. Two separate outputs, taken for different network configurations, of the OSA are plotted in Fig. 5.1.b. Each line in the spectrum represents an ITU-T channel; adjacent channels are spaced by 100 GHz. Once the network is built and deemed "operational," the locations of the spectral lines do not change; however, their powers may drift due to network impairments, component failures, etc. Additionally, the noise floor, which is due to amplified spontaneous emission (ASE), is characterized as part of the initial setup and is not of concern to a network monitor who is only worried about the power contained in each channel. As such, instead of displaying the full spectra of Fig. 5.1.b, one could instead show only the pertinent information, as plotted in Fig. 5.1.c, which is simply a graph of powers contained in a discrete number of ITU-T channels.

In other words, the full power of the OSA is not necessary. Once the network is built, and through initial characterization, one knows the source lineshapes (i.e. lineshapes of the laser transceivers) as well as their spectral locations. One only needs to know how much power is contained in each channel at some arbitrary point in the network; i.e. Fig. 5.1.c instead of Fig. 5.1.b. The CSM is designed for this functionality.

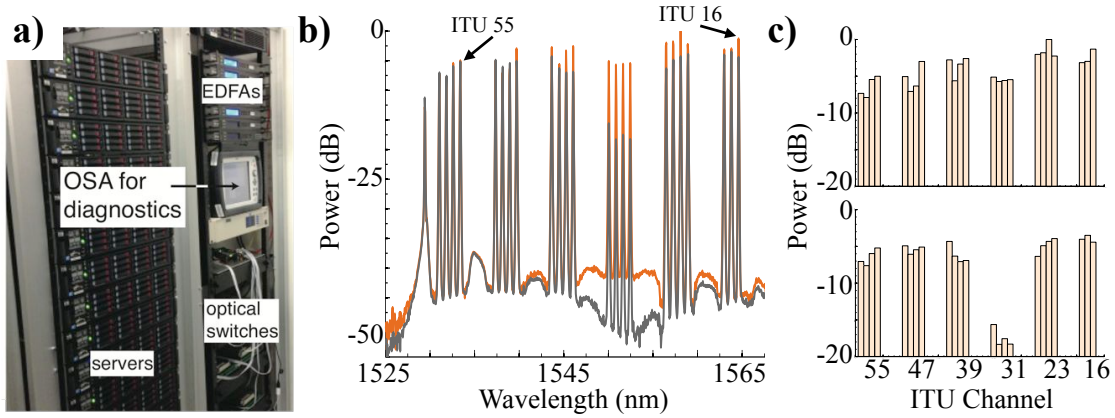

Figure 5.1: **a)** Hardware for the MORDIA network. In particular, we note that there is a large OSA, which takes up more than its fair share of rack space, to monitor the power in the various channels. **b)** Two spectra, of the MORDIA network, corresponding to two different network conditions. From a network-monitoring point-of-view, we are interested in the power variations of each line (i.e. channel), but not necessarily in the specifics of the lineshapes, ASE noise, etc. **c)** The same information in (b), but in the network-monitoring point-of-view (orange trace $\rightarrow$ top panel; grey trace $\rightarrow$ bottom panel). In going from (b) $\rightarrow$ (c), the ordinate remains the same but the abscissa has been discretized into corresponding ITU-T channel numbers.

5.2 The simplest, yet impractical, method to monitor a channelized spectrum

Here, we describe the conceptually easiest method for designing a CSM; however, as we will see, this method is impractical given the constraint of realistic fabrication processes [91, 7, 92, 93, 16]. Again, the goal of a CSM is to detect the powers in a priori known channels, i.e. over a discrete set of (center) wavelengths. In principle, the structure diagrammed in Fig. 5.2 should be able to perform the described function. In this structure, a bank of N ring resonators, which are isolated from each other, are coupled to a common bus waveguide. Each resonator is tuned to a different wavelength λ_i, with each wavelength representing a different DWDM channel. Each resonator also drops resonant light to a photodetector; one photodetector per resonator. Therefore, by dropping the N separate channels to N separate detectors, one is able to directly monitor the power levels in each of

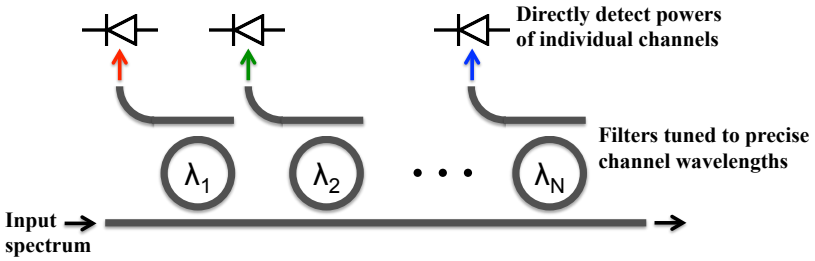

Figure 5.2: A conceptually simple implementation of a CSM. Each resonator drops a unique wavelength channel.

the N channels.

The foregoing prescription is impractical for two reasons. First, there is an implicit necessity for ultra-high-Q resonators. For the CSM to operate as described, there must be no cross talk between adjacent resonators or even between any two resonators. Given that our present application is geared towards a 100-GHz channel spacing, an extremely stringent requirement is therefore placed on the spectral properties (bandwidths, roll-offs, etc.) of the resonators. Second, it is generally impossible to predict the spectral locations of resonances pre-fabrication [91, 16]. Therefore, it is difficult to meet the requirement that each of the resonators correspond to a specific and unique λ_i. Even if the resonators are made to be tunable, the allowable distance in wavelength between the two furthest channels may then be limited by the tuning range.

We note, and show for completeness, that even though it is difficult to predict the resonant frequency of a to-be-fabricated ring of length L, one is usually able to predict the free spectral range (FSR)

$$\Delta\nu = \frac{c}{n_g L}, \tag{5.1}$$

where n_g is the group index of the waveguide mode and c is the velocity of light. The reason for unintended shifts, in either the m^{th} resonance ν_m or the FSR $\Delta\nu$, is due to deviations in the refractive (phase) index n of the waveguide mode, which can occur due to various process variations, sidewall roughness, waveguide disorder, or even non-uniformity in the buried oxide layer [91]. Differentiating (5.1) yields

$$\frac{d\Delta\nu}{dn} = -\frac{c}{n_g^2 L} \cdot \frac{dn_g}{dn}. \tag{5.2}$$

Meanwhile, the m^{th} resonance

$$\nu_m = \frac{mc}{nL} \tag{5.3}$$

has the derivative

$$\frac{d\nu_m}{dn} = -\frac{mc}{n^2 L}. \tag{5.4}$$

In a common silicon waveguide, the second factor in (5.2) is of magnitude-order unity and n_g is not much greater than $2n$, if even that large[1]. Then, comparing (5.2) to (5.4), we may write

$$\left|\frac{d\nu_m}{dn}\right| \sim m \left|\frac{d\Delta\nu}{dn}\right|. \tag{5.5}$$

From (5.3), we see that

$$m = \frac{nL}{\lambda}, \tag{5.6}$$

which could easily be on the order of 100, if not 1000. Therefore, (5.5) indicates that errors in nominal resonant frequencies are ∼100–1000 times stronger than errors in nominal FSRs. This confirms our assertion that, within an acceptable range of error, FSRs are well-predictable whereas specific resonant locations are not.

5.3 Device architecture

Before we describe the actual implementation of the CSM, we briefly describe the architecture of the experimentally measured device. A dark-field micrograph of the CSM is shown in Fig. 5.3.a. At first glance, it is not unlike the diagrammed example of the foregoing section. The CSM is a 24-element filter bank containing a total of 168 ring resonators. Each element of the filter bank is a seventh-order ring-lattice filter (i.e. seven rings per filter). A single filter-element is shown in the bright-field micrograph of Fig. 5.3.b. The rings exhibit a nominal bend radius of 1.69 μm and are of constant size within each filter, but vary in length from filter-to-filter. The ring lengths are linearly graded from 34.9 μm at the input down to 34.5 μm for the last filter. The gradation of the ring lengths are

[1] cf. Fig. 3.6.

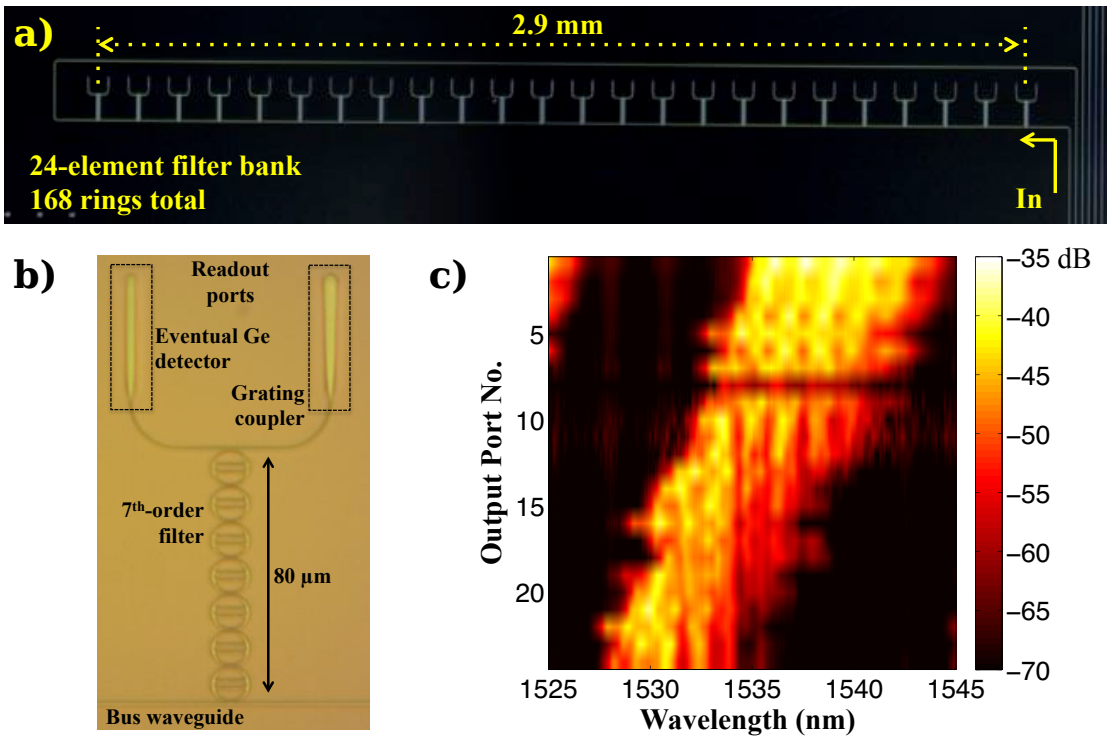

Figure 5.3: **a)** Dark-field micrograph of the CSM. **b)** Bright-field micrograph of a single filter element of the CSM. Light input, through the bus waveguide, from the right, is dropped to the grating coupler. Light input, from the left, is dropped to what will eventually be a Ge photodetector for high-speed readout. **c)** Transmission spectra, for the 24 filters, collected via each of the corresponding grating couplers.

achieved by grading the lengths of the coupling regions while keeping the bend radii constant. The fabrication was performed using a fully CMOS process, at Sandia National Laboratories, on a 6-in. silicon-on-insulator (SOI) wafer. The waveguides are fully etched strip waveguides, which were oxidation trimmed to 230-nm thickness and oxidation smoothened for sidewall-roughness reduction and mitigation of optical propagation loss. To mitigate dispersion in the directional couplers, the bus waveguide was given a width of 550 nm while the straight portions of the rings were given a width of 400 nm. The wall-to-wall gaps of all the directional couplers were 280 nm.

The CSM shown in Fig. 5.3 is a fully passive device; there are no dopant implants, Ge epitaxy, or metallization. The measurements performed below were

on this passive device. This design, however, was made with higher mask levels in mind. Specifically, the rings are based on an adiabatically transitioned design [60] in which the waveguide width evolves from 400 nm in the coupling region to 800 nm at 90° away. The inner portions of the wide-waveguide sections may be n+ doped thus forming a resistive heater that can thermo-optically affect the optical mode, which traverses primarily through the undoped outer portions. The tether-like structures interior to the rings may also be n+ doped to provide for electrical contact to a metal layer for electrical routing. This design allows for thermo-optic tuning of the filter passbands.

As seen in Fig. 5.3.b, opposite to the input bus waveguide, there are two options at the output port. If light is input, through the bus waveguide, from the right, then the dropped signal exits a grating coupler for fiber-based detection. If light is input from the left, then the dropped signal is sent to what will be, in a future iteration, a Ge photodetector. The use of on-chip detection will allow for robust and packagable high-speed operation. In the present work, we detect the output signal via grating-based fiber coupling to prove our concept.

The 24 filters each drop their signals to 24 corresponding detection ports, which are the grating couplers on a 125-μm pitch. For testing purposes, the signals from the gratings were individually coupled to a lensed fiber and detected with a high-sensitivity InGaAs photodiode. We also coupled the grating signals to an InGaAs sensor array, via a microscope, for the purpose of simultaneous readout. Both methods yielded similar information for the input signals to be discussed below.

5.4 Matrix implementation and experimental results

We now elucidate the actual implementation of the CSM, which is based on the solution of a matrix equation. For the sake of example, we also describe the experimental results concurrently.

The transmission spectra, of all 24 filters, as read out by coupling a fiber to

the grating couplers, are plotted in Fig. 5.3.c. Due to the filter-to-filter gradation of the ring lengths, the passband centers shift by ∼4 nm per filter. Additionally, as explained in Sec. 5.6, there is intentional passband ripple as a result of imperfect apodization in the filter design. In contrast with the necessities specified in Sec. 5.2, we see that the passbands in the experimental device are relatively broad (e.g. they pass several 100-GHz ITU-T channels) and that there is significant filter-to-filter crosstalk. As is common, Fig. 5.3.c was obtained by sweeping a narrowband tunable laser (ideally a delta function) over the range of wavelengths indicated. The output-port axis is clearly discrete, due to the finite number of filters present. The appearance, however, that the wavelength axis forms a continuum, is due to the experimental fact that many wavelength points are recorded. In reality though, any experimental recording must be inherently discrete. We can then interpret Fig. 5.3.c as a $24 \times N$ matrix, where the transmission axis forms the non-discretized entries of each element of the matrix and N is the number of wavelength data points recorded during each sweep of the tunable laser.

Without loss of generality, let us suppose that the input spectrum to the CSM is composed of five ITU-T channels, spaced 100-GHz apart, as plotted in Fig. 5.4.a. Because we seek a matrix interpretation to the problem of channel monitoring, we can discretize the input spectrum into the five channels of interest, as plotted in Fig. 5.4.b. The individual powers that are plotted are the individual spectral integrations of each respective source as would be obtained, for example, if one sends a single channel directly to a photodetector without first dispersing the spectrum. Regardless of the interpretation of the input signal, when the foregoing spectrum is sent through the CSM, the powers measured at the 24 output ports are as plotted in Fig. 5.4.c. In the continuous interpretation, the full spectrum (Fig. 5.4.a) is multiplied by each of the 24 rows in Fig. 5.3.c to yield the 24 output detector readings. In the discrete (i.e. matrix) interpretation, the discretized spectrum (Fig. 5.4.b) is multiplied by some appropriately channel-discretized version of Fig. 5.3.c, which will be interpreted as a 24×5 matrix below, to yield the 24 output detector readings.

Let us the call the desired 24×5 matrix $\mathbb{T}$. In analogy with the inter-

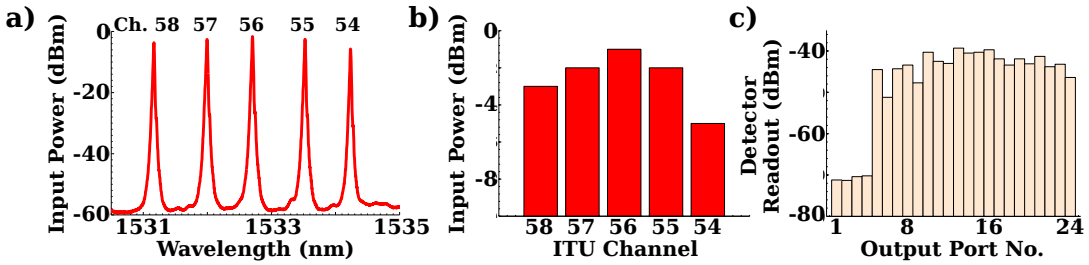

Figure 5.4: **a)** Five-channel input spectrum. **b)** The powers carried in each of the input channels. This plot corresponds to the data plotted in (a), but is spectrally resolved only over the channels of interest instead of over a wavelength continuum. Each bar represents the spectral integration of each respective source, performed individually. **c)** Powers read out at the 24 output ports when the input signal to the CSM corresponds to (a) and (b).

pretation that the wavelength axis of Fig. 5.3.c is formed via N wavelengths, we can form $\mathbb{T}$ by inputting our five channels of interest, normalized to 0 dBm each, into the CSM one-at-a-time. For each input channel, we measure the 24 outputs. This yields the 24×5 matrix $\mathbb{T}$, which mathematically connects Figs. 5.4.b–c. The foregoing measurements serve as the initial device calibration. We are now able to write the matrix equation

$$\mathbb{D} = \mathbb{T}\mathbb{P}, \tag{5.7}$$

where $\mathbb{P}$ is a five-element column vector containing the input powers (i.e. the vector representation of Fig. 5.4.b) and $\mathbb{D}$ is a five-element column vector containing the output-port detector readouts (i.e. the vector representation of Fig. 5.4.c). This equation is diagrammed in Fig. 5.5.

The foregoing equation is the mathematical statement that the CSM maps the input powers $\mathbb{P}$ to the output detectors $\mathbb{D}$. However, in the spirit of channel monitoring, we are interested in being able to reconstruct $\mathbb{P}$ from an arbitrary set of detector readouts $\mathbb{D}$. I.e. we are interested in the solution to the inverse problem

$$\mathbb{P} = \mathbb{T}^{-1}\mathbb{D}, \tag{5.8}$$

where, because we are working with a rectangular matrix, $\mathbb{T}^{-1}$ is the Moore-Penrose pseudo-inverse of $\mathbb{T}$ [94, 95]. It is evident that the problem of channel monitoring, as we have stated it, has been reduced to a problem in linear algebra.

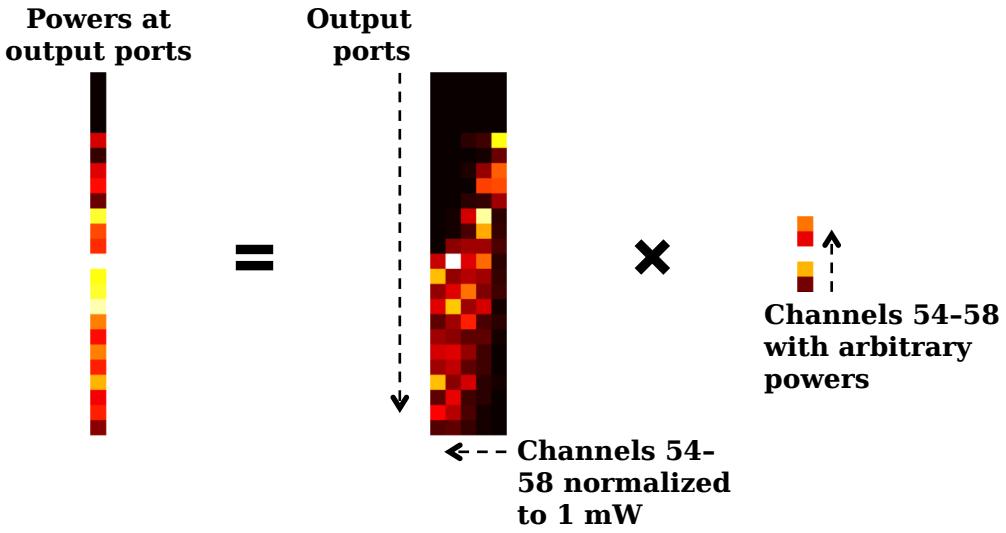

Figure 5.5: The matrix equation $\mathbb{D} = \mathbb{T} \times \mathbb{P}$, respectively. $\mathbb{T}$, which is obtained through the initial device calibration, allows for the unique prediction of $\mathbb{D}$, given $\mathbb{P}$. We are more interested, however, in the inverse problem of determining $\mathbb{P}$, given $\mathbb{D}$.

Instead of explicitly calculating the pseudo-inverse, (5.8) can be solved directly via the QR-factorization of $\mathbb{T}$ [96, 97]. Several experimental examples of solving (5.8) via QR-factorization are shown in Fig. 5.6. The left column of Fig. 5.6 shows detector readouts for three sets of inputs where the channel powers vary between sets. The right column shows the "unknown" input powers (bars) along with the powers that are inferred (stems) via the solution of (5.8). These data show that the matrix implementation of the CSM is able to determine the input powers to within an acceptable amount of error.

Looking at the bars in Fig. 5.6.b, we see that Ch. 55–58 are attenuated while Ch. 54 is held constant. Taking this example further, we continue to attenuate Ch. 55–58 while holding Ch. 54 constant. In Fig. 5.7.a, we show the CSM's ability to track the power changes in the relevant channels over a dynamic range of $\sim$17 dB, with a maximum error of $\sim$1 dB. The CSM also recognizes that Ch. 54 does not suffer any attenuation.

Finally, if one desires, the spectrum may be reconstructed from the data in Fig. 5.6.b. If one characterizes the normalized lineshapes of the individual channel sources (e.g. during initial system set-up), then the spectrum can be reconstructed

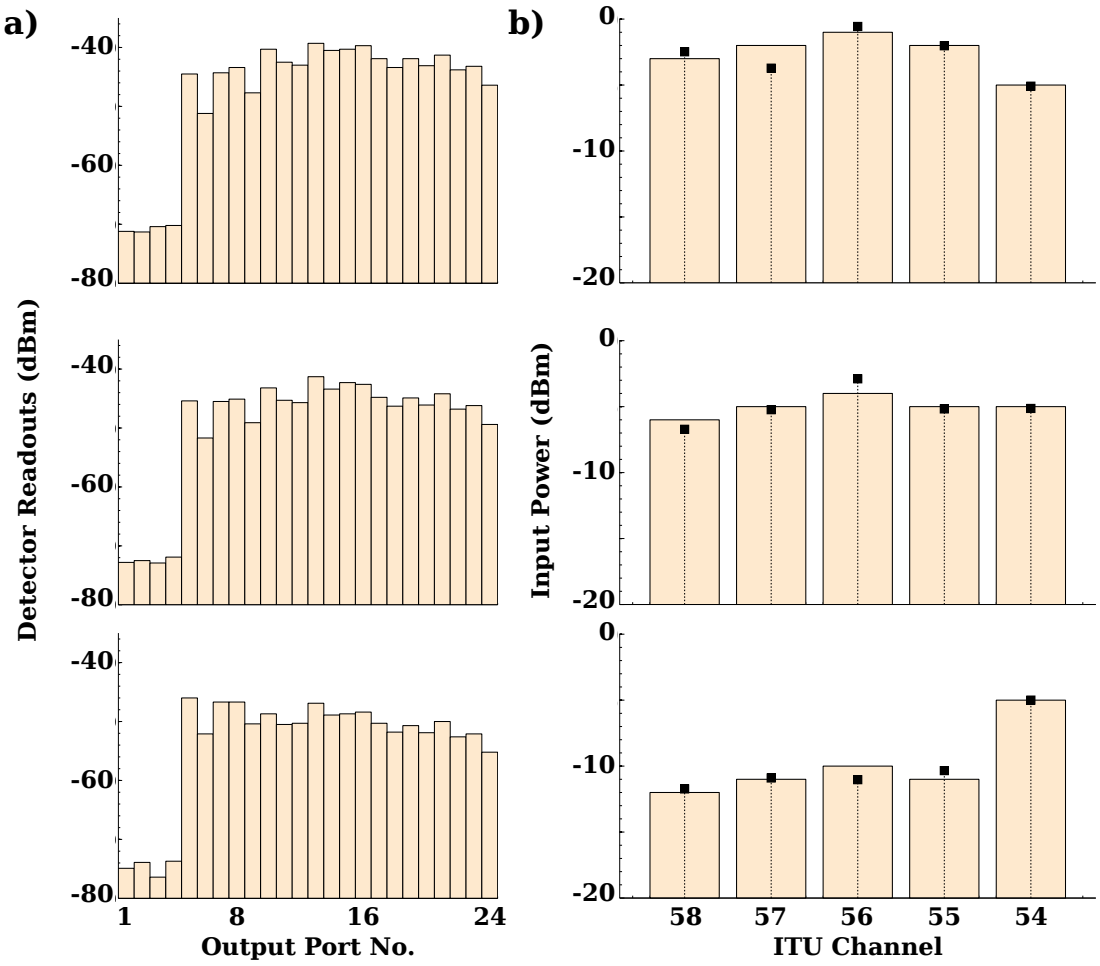

Figure 5.6: **a)** Experimental detector readouts. Each set corresponds to different amounts of powers carried by the five channels. **b)** Input powers in the five channels corresponding to (a). The bars show the actual "unknown" inputs. The stems show the results of computing the solutions to (5.8), thus indicating the realization of a channelized spectrum monitor.

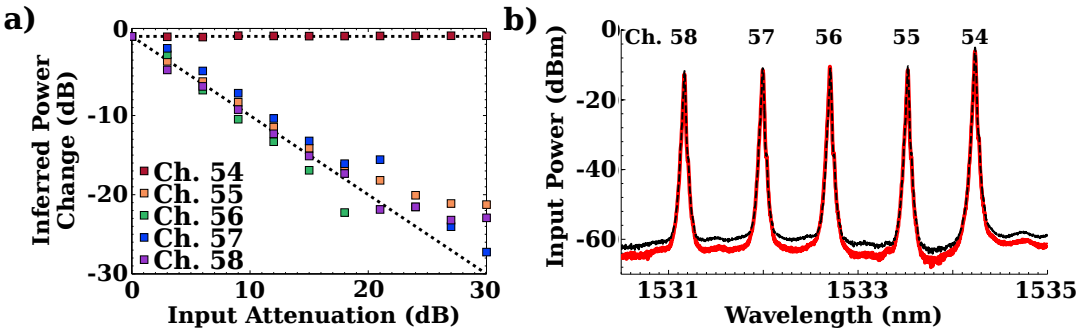

Figure 5.7: **a)** Tracking ability of the CSM. Channels 55–58 are attenuated while Ch. 54 is held constant. The dashed lines show the ideal result; the markers show the actual result of using the CSM. Reasonable tracking is observed with a dynamic range of ∼17 dB. **b)** Reconstruction (black trace) of the original input spectrum (red trace) by weighting pre-characterized source lineshapes.

by weighting the lineshapes by the corresponding inferred powers and then summing the lineshapes. An example of this, corresponding to the last row of Fig. 5.6, is plotted in Fig. 5.7.b. We see that this method accurately reconstructs the source lineshapes, but overestimates the ASE noise floor. The overestimation of ASE is of little concern because our original intent was to track the powers over a spectrum discretized by the number of channels and not over a continuous one.

5.5 Robustness to catastrophic failure of filter elements

The architecture of the CSM endows upon it a certain degree for fault tolerance. If several of the filters fail, as long as there are still more filters than there are input channels, it is possible that the CSM may still continue to operate as good as, if not better than, when the full filter bank was present[2]. Continuing with the example from the foregoing section, we assume that catastrophic failure occurs to 19 of the 24 filters. Specifically, we assume that we are able to measure light only from output ports 10, 14, 19, 22, and 23, as shown in Fig. 5.8.a. The

[2]We defer the mathematical conditions, under which the CSM operates nominally, to the next section.

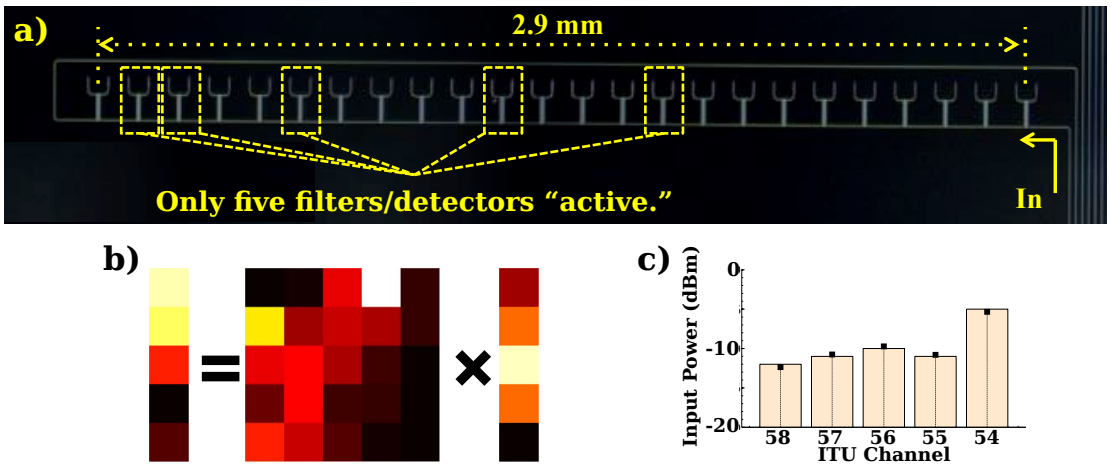

Figure 5.8: **a)** Dark-field micrograph of the CSM showing the filter elements used in the five-element test. The other 19 filters can be considered to be unusable due to catastrophic damage. **b)** The matrix equation $\mathbb{D} = \mathbb{T} \times \mathbb{P}$, respectively (cf. Fig. 5.5). Both $\mathbb{T}$ and $\mathbb{D}$ are reduced, from the previous section, from 24 to five rows but $\mathbb{P}$ remains the same size. **c)** Actual (bars) and inferred (stems) input powers, showing the CSM operating nominally under the condition of reduced output-port data.

effect of this is that the number of rows, in the matrices $\mathbb{D}$ and $\mathbb{T}$, are reduced from 24 to five. The number of rows in $\mathbb{P}$, however, remains five because the number of channels has not decreased. The new $\mathbb{T}$ is simply the 10^{th}, 14^{th}, 19^{th}, 22^{nd}, and 23^{rd} rows of the old $\mathbb{T}$. The new matrix multiplication is diagrammed in Fig. 5.8.b. The prescription at this point proceeds as before; i.e. solving the matrix equation

$$\mathbb{P} = \mathbb{T}^{-1}\mathbb{D}. \tag{5.9}$$

An example solution, showing nominal operation of the CSM, is plotted in Fig. 5.8.c.

We now perform the same attenuation experiment as before; i.e. Ch. 55–58 are attenuated while Ch. 54 is held constant. In Fig. 5.9, we show the CSM's ability to track the power changes in the relevant channels over a dynamic range of ∼21 dB. This range is slightly higher than when all 24 filters were used. Furthermore, the error within the tracking range is tighter (< 1 dB) than in the former case; although, there seems to be slightly more error in tracking Ch. 54, which does not suffer any physical attenuation. It is therefore evident that the CSM can operate with reduced output ports.

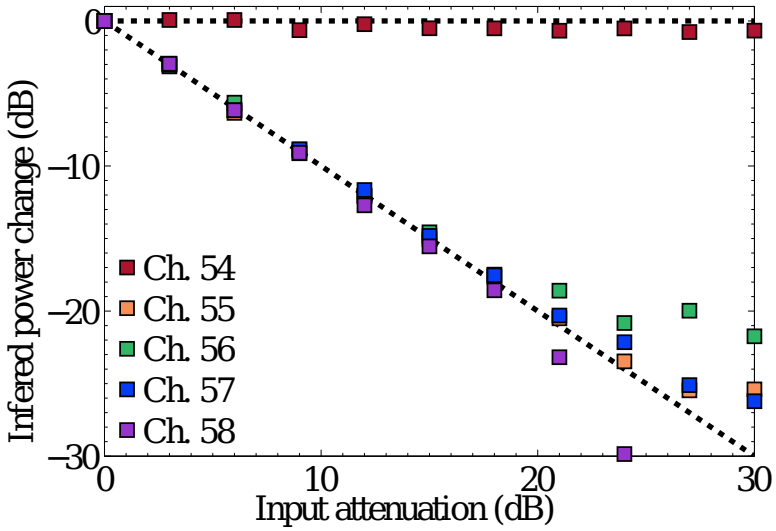

Figure 5.9: Tracking ability of the CSM while receiving data from only five of the 24 output ports. Channels 55–58 are attenuated while Ch. 54 is held constant. The dashed lines show the ideal result; the markers show the actual result of using the CSM. Reasonable tracking is observed with a dynamic range of ∼21 dB.

5.6 Mathematical elaborations

To summarize the matrix implementation of the CSM: An arbitrary, but unknown, power distribution is represented by the n-element column vector $\mathbb{P}$, where the n vector elements correspond to the n wavelength channels of the network. When this power distribution propagates through the CSM, power is routed to m detection ports via the transmission spectra of the filters. These detector readouts are represented by the m-element column vector

$$\mathbb{D} = \mathbb{T}\mathbb{P}, \tag{5.10}$$

where $\mathbb{T}$ is an $m \times n$ matrix that is measured as part of the initial device characterization. During real-time operation, $\mathbb{D}$ is measured; however, the desired information is $\mathbb{P}$. Therefore, (5.10) must be solved for the unknown vector $\mathbb{P}$.

In the experimental measurements, we took $n = 5$ and either $m = 24$ or $m = 5$; i.e. we imposed that $m \geq n$. Provided that $\mathbb{T}$ is of full rank, then the condition $m \geq n$ provides that (5.10) exhibits a unique solution for $\mathbb{P}$ [98]. If $m < n$, then the system of equations represented by (5.10) is under-determined

and the solutions for $\mathbb{P}$ may not be unique. Therefore, the CSM requires that $m \geq n$.

If $\mathbb{T}$ is rank deficient, then $\mathbb{P}$ may not have a unique solution even if $m \geq n$. Therefore, the filter design of the CSM must be performed with the intention that $\mathbb{T}$ must be of full rank. To ensure unique columns of $\mathbb{T}$, the rings were graded in length, from filter-to-filter, thus yielding the graded transmission spectra of Fig. 5.3.c or the $\mathbb{T}$ matrices of Figs. 5.5 and 5.8.b. Additionally, the coupling coefficients of the rings were chosen such that the transmission spectra of the filters were not apodized. This gave rise to significant, and somewhat random, passband ripple in the transmission spectra. This is essentially a dithering process to further aid in ensuring that $\mathbb{T}$ is of full rank. The ultimate resolution of the CSM is determined by the point at which two adjacent channels yield non-independent columns during the initial construction of $\mathbb{T}$.

If the above restrictions on $\mathbb{T}$ are satisfied, the solution of $\mathbb{P}$ is easily obtained. If the system is critically determined ($m = n$), then (5.10) can be directly solved by Gaussian elimination, LU-factorization, or similar [98]. If the system is over-determined ($m > n$), then QR-factorization is useful [96, 97]. In this case, $\mathbb{T}$ can be factorized such that

$$\mathbb{T} = \mathbb{Q}\mathbb{R}, \tag{5.11}$$

where $\mathbb{Q}$ is orthogonal and $\mathbb{R}$ is upper-triangular. This allows us to write

$$\mathbb{R}\mathbb{P} = \mathbb{Q}^{\mathrm{T}}\mathbb{D}, \tag{5.12}$$

which is easy to solve by backwards substitution[3]. Of course, although computationally more expensive, one could directly calculate the matrix inverse (for $m = n$) or the Moore-Penrose pseudo-inverse (for $m > n$) [94, 95] of $\mathbb{T}$ and determine the solution via

$$\mathbb{P} = \mathbb{T}^{-1}\mathbb{D}. \tag{5.13}$$

All of the foregoing methods provide unique, optimal solutions in the least-squares sense [98].

[3]The last row of $\mathbb{R}$ only has one element, thus yielding the solution to the last element of $\mathbb{P}$. This can be substituted into the second-to-last equation of the system, thus yielding the second-to-last element of $\mathbb{P}$, and so forth.

Chapter 5, in part, contains material, published in the following, of which the dissertation author was the primary investigator:
R. Aguinaldo, P. Weigel, H. R. Grant, C. T. DeRose, A. Lentine, A. Pomerene, A. Starbuck, and S. Mookherjea, "Characterization of a silicon-photonic multi-wavelength power monitor," in *IEEE Optical Interconnects Conference*, p. WD5, 2014 [99].

Chapter 6

Routing of a Data Center Network Through a Wideband Thermo-Optic Switch

6.1 Overview

Data-center networks with optical interconnects [11, 12] may lower energy consumption, and scale more efficiently if silicon photonic components can replace some of the conventional off-the-shelf components used today. MORDIA (Microsecond Optical Reconfigurable Datacenter Interconnect Architecture), shown in Fig. 6.1, is a multi-wavelength, multi-port optical circuit-switched network, designed to support a wide variety of all-to-all communication workloads, e.g. MapReduce, TritonSort, and data sorting and searching [100]. It may benefit scalability, re-configurability, and maintenance of networks if optical switching can be incorporated within the ring. Such switches have to be capable of supporting the full optical bandwidth in the ring (here, more than 30 nm) and be reconfigurable in a few microseconds (currently about 12 μs in the MEMS-based implementation [100]). It is also necessary that such a switch be energy efficient, and can be driven in a manner that is compliant with digital controllers.

While thermo-optic silica and polymer switches are mature technologies,

and have shown excellent output-port contrast [101], their large footprints, large power consumptions, and relatively slow switching times suggest areas for improvement. Silicon can be used to make a broadband, yet energy-efficient, microsecond-scale switch that is highly integrable and can operate on many wavelength channels. One method is to use a microring resonator, designed with a free-spectral range (FSR) equal to the inter-channel spacing. All-optical switching of 20 CW wavelengths (and one data channel) in this configuration has been demonstrated, albeit with a significant power penalty and change-of-slope in the bit-error-rate sensitivity curve [102]. In principle, the microring could be thermo-optically switched: energy-efficient thermo-optic tuning of a microring has been demonstrated with a power consumption of only 0.5 mW per nanometer of wavelength shift, and a 10%-90% switching time of about 1 μs [103]. However, this resonant structure may be subject to crosstalk between the optical channels, and the alignment of the resonances to the ITU-T wavelength grid may depend on temperature. Also, the coupling coefficient of a compact directional coupler, as typically used between a microring and a waveguide, tends to vary widely with wavelength [88], and longer (adiabatic) couplers, which overcome this limitation, may not allow the microring resonator to achieve the requisite free-spectral range.

A second approach is to use a Mach-Zehnder interferometer (MZI) [104]; implemented here with wideband 3-dB couplers, and an energy-efficient thermo-optic phase-shift mechanism in one arm. Compared to carrier injection in silicon MZIs, the thermo-optic MZI should have smaller insertion loss and greater scalability to larger switching fabrics, because the device is much smaller as a result of the larger magnitude of the thermo-optic effect, compared to the free carrier plasma dispersion effect. Here, the 3-dB couplers attempt to achieve wavelength-insensitive power splitting by lithographic design [105]. However, the thermo-optic phase shift should also be wavelength-insensitve, which can be fundamentally difficult to achieve. While care has been taken to investigate the nonlinearity of the thermo-optic phase shift with temperature [106], the wavelength variation of this effect remains relatively unexplored.

The differential phase shift $d\phi$ accumulated in an incremental distance dx

of optical path length is not only proportional to the change in the refractive index Δn induced by the temperature rise ΔT, but is also inversely proportional to the optical wavelength λ, i.e. $d\phi = (2\pi/\lambda)\Delta n\, dx$. Thus, to minimize the variation of the cumulative phase with wavelength, we need to increase the magnitude of Δn, so that integration over a long optical path length is not required to achieve π phase shift. This requires increasing the temperature range ΔT introduced by the heating source, and also minimizing the spread of the temperature away from the heating source along the optical path, so that the range of integration is minimized. Efficient and fast heating can be achieved by directly heating the silicon waveguide in close proximity to the optical mode [103, 105, 107, 108]. Reducing the heat spreading along the silicon waveguide can be achieved by varying the driving waveform at sub-microsecond time-scales, as discussed in Section 6.3.1.

Here, we demonstrate and characterize microsecond time-scale cross-bar digital switching of twenty 10 Gbit/s wavelength channels spanning the wavelength range from 1531.12 nm (ITU Channel 58) to 1563.05 nm (ITU Channel 18), using a thermo-optically driven wideband Mach-Zehnder interferometer, with an electrical power consumption of 15 mW and a 10%-90% switching time of 11 μs. We investigate the 2×2 switch as a building block of a larger switching fabric; therefore, optical coupling to and from the chip was not optimized and we incurred large losses (about 10 dB per coupler) when using lensed tapered fibers and multi-access nano-positioning stages. Electrical contacts to the chip were made using a multi-contact wedge. There was no need to stabilize the chip temperature in the demonstration (i.e. no power consumption for thermo-electric cooling).

6.2 A thermo-optic switch with cascaded phase shifters

The switch described here is based on a Mach-Zehnder interferometer (MZI), with adiabatic wideband 3-dB splitters [105]. The switch was operated by heating one arm of the MZI shown in the Fig. 6.2, which, in general, causes an amplitude and phase change in that arm. When no DC voltage was applied to the contact

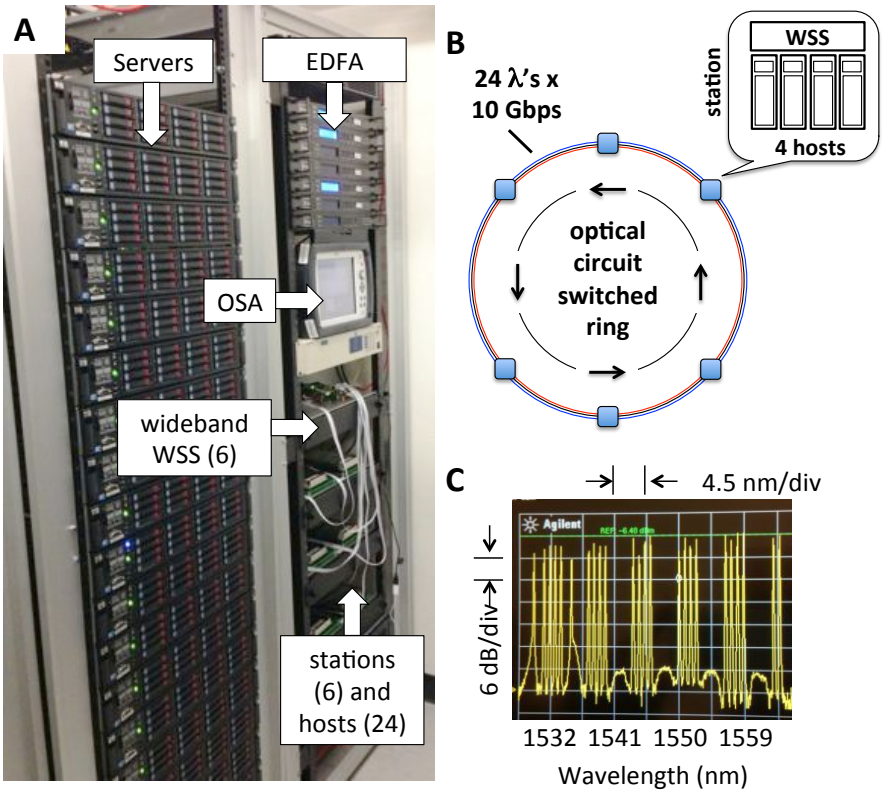

Figure 6.1: **A)** Hardware for the optical circuit-switched multi-wavelength MORDIA ring network at UC San Diego, including data servers, optical amplifiers (EDFAs), optical spectrum analyzer (OSA) for power monitoring, and hardware for wideband wavelength-selective switching (WSS). There are six nodes and four host stations per node. **B)** Schematic of the ring network topology, in which any of the nodes can access the full bandwidth of the ring (about 30-nm wavelength span). **C)** Optical spectrum of 20 data channels, each carrying 10-Gbit/s data, used in the switching demonstration (some extraneous channels, not carrying data, or at long wavelengths that lie outside the range of the tunable filters used to measure the individual eye patterns, also propagated through the chip but were not measured here).

pads labeled V_{mod} in Fig. 6.2, all the wavelengths exited the device in the "cross" port, i.e. the device is, to a good approximation, bias-voltage free. Upon applying a voltage to the contact pads shown in Fig. 6.2, all the wavelengths are switched over to emerge from the "bar" port. Unlike conventional thermo-optic MZIs, in which a metallic heater is fabricated at some distance from the silicon waveguide, and separated from it by a certain thickness of insulating oxide [109], current was driven through a dopant-implanted region of the waveguide itself, in close proximity to the optical mode. The waveguide was widened in certain regions from about 0.4 μm to about 1.0 μm, N-doped, and contacted with narrow N-doped silicon tethers connected to metal, through which an electrical current was directly injected in close proximity to the optical mode. The device was fabricated, using a fully CMOS-compatible process with 248-nm lithography at Sandia National Laboratories, on a 150-mm silicon-on-insulator wafer (250-nm active layer, 3-μm buried oxide). The waveguides were fully etched with nominal width $\times$ height dimensions of 400 $\times$ 230 nm^2. Additional details are described elsewhere [105].

An earlier version of the device with a single phase-shift element in each arm has been previously reported [105, 108]. The MZI used here comprises a sequence of five thermo-optic phase shifters in each arm. As shown schematically in Fig. 6.2, the five resistors were electrically driven in parallel ($R_{\text{total}} = 1.17$ kΩ), thus reducing the drive voltage needed to achieve π phase shift from $V_\pi = 20$ V for a single-element phase shifter to about $V_\pi = 4.25$ V. Characterization measurements showed that heaters could withstand no more than about 40 mW of electrical power before damage, and that the five-element parallel-heater structure was noticeably more robust than the single-element heater. However, there is a concern regarding the increased insertion loss, since each of the five phase-shifting elements imparts some loss to the optical transmission in the "hot" state, when an electrical current passes in close proximity to the optical mode. This issue is investigated in the following section.

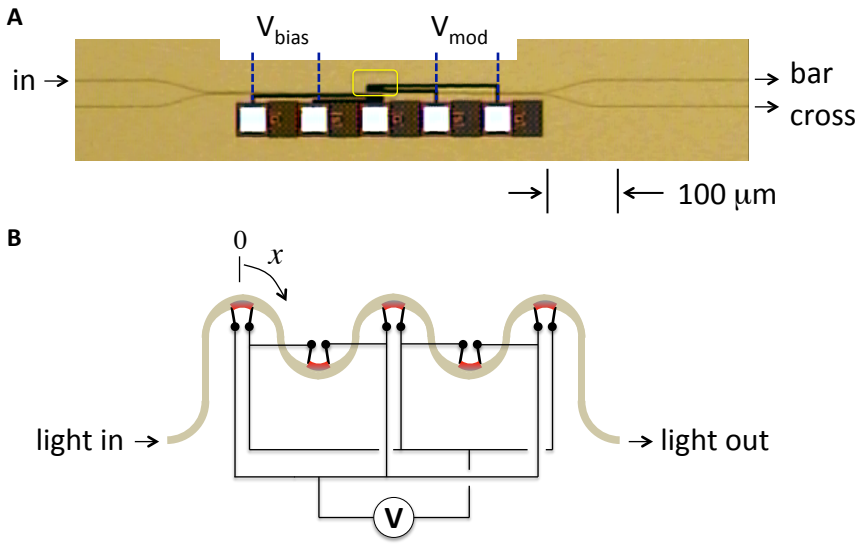

Figure 6.2: **A)** Mach-Zehnder interferometer thermo-optic silicon-photonic cross-bar switch with bias voltage (V_{bias}, unused) and switching voltage ($V_{mod} = 0$ V or $V_{mod} = 4.25$ V) indicated. The region highlighted in yellow contains a bank of five phase shifters, as shown schematically in (B). **B)** The optical field experiences a thermo-optic phase-shift in each of the widened arcs, the inside of which is doped to create a resistor. These resistive heaters are electrically wired in parallel, so as to reduce the switching voltage compared to a single heater. The axial co-ordinate x is referred to in Section 6.3.1.

6.2.1 Device modeling and parameter extraction

The switch structure shown in Fig. 6.2.A was modeled using transfer matrices. We label the optical field amplitude cross-coupling and through-coupling coefficients by κ and t, respectively [73]. The transfer function of the device can be calculated as a cascade of coupling ($\mathbb{C}$) and propagation ($\mathbb{P}$) matrices:

$$\begin{aligned}\begin{pmatrix}\text{bar}\\ \text{cross}\end{pmatrix} &= \overbrace{\begin{pmatrix} t & \kappa^* \\ \kappa & -t^* \end{pmatrix}}^{\mathbb{C}} \cdot \overbrace{\begin{pmatrix} a\,e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix}}^{\mathbb{P}} \cdot \overbrace{\begin{pmatrix} t & \kappa^* \\ \kappa & -t^* \end{pmatrix}}^{\mathbb{C}} \cdot \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} at^2 e^{i\phi} + |\kappa|^2 \\ a\kappa t e^{i\phi} - t^*\kappa \end{pmatrix} \end{aligned} \tag{6.1}$$

where ϕ is a function of the thermally-induced phase shift, and hence, of the applied voltage. The parameter a in matrix $\mathbb{P}$ represents the loss introduced by heating portions of one of the MZI arms, by driving an electrical current in close proximity

to the optical mode, as shown in Fig. 6.2.B. We take $a = 1$ when no voltage is applied, and also assume that the lower-right term of $\mathbb{P}$ (i.e. the unheated arm) has no thermally-induced amplitude attenuation. A common multiplicative factor (amplitude or phase) affecting all the terms of the $\mathbb{P}$ matrix has no bearing on the parameter extraction algorithm described below, which is based on ratios of measured bar and cross transmissions.

The measured quantities were $10 \cdot \log_{10} |\text{bar}|^2$ and $10 \cdot \log_{10} |\text{cross}|^2$, as functions of wavelength, and for different voltages as shown in Fig. 6.3. As with all MZI devices, at certain voltages, the bar transmission was minimized and the cross transmission was maximized; in the present device, this occurred, to a good approximation, with no voltage applied, i.e. $\exp(i\phi)$ was approximately -1, based on the algebraic form of (6.1). More precisely, we write $\exp(i\phi)_{V=0} = -\exp(i\delta_0)$ where the small parameter δ_0, a function of wavelength, represents a phase variation because it is generally impossible for lithography to achieve exact phase equality over a wide range of wavelengths (here, exceeding 30 nm). The bar transmission was maximized, and the cross transmission was minimized, when a voltage of $V_\pi \equiv 4.25$ V was applied. In this case, $\exp(i\phi)$ was approximately $+1$, and we write $\exp(i\phi)_{V=V_\pi} = +\exp(i\delta_V)$.

From (6.1), the algebraic expressions for these quantities are:

$$\begin{aligned} \text{bar}_{\text{on}} &= 20 \cdot \log_{10} \left| a\, t^2 e^{i\delta_V} + |\kappa|^2 \right|; \quad \text{cross}_{\text{off}} = 20 \cdot \log_{10} \left| -\kappa t e^{i\delta_0} - t^* \kappa \right|; \\ \text{bar}_{\text{off}} &= 20 \cdot \log_{10} \left| -t^2 e^{i\delta_0} + |\kappa|^2 \right|; \quad \text{cross}_{\text{on}} = 20 \cdot \log_{10} \left| a\, \kappa t e^{i\delta_V} - t^* \kappa \right|. \end{aligned} \tag{6.2}$$

To each of the expressions in (6.2) must be added the fiber-to-chip coupling losses at each interface. To factor out the latter, we defined four ratios: $\text{bar}_{\text{on}}/\text{bar}_{\text{off}}$, $\text{cross}_{\text{off}}/\text{cross}_{\text{on}}$, $\text{cross}_{\text{off}}/\text{bar}_{\text{off}}$, and $\text{bar}_{\text{on}}/\text{cross}_{\text{on}}$. In Fig. 6.3, we assumed $|\kappa|^2 + |t|^2 = 1$, i.e. the coupler sections were lossless, and relax this assumption in Fig. 6.4. We calculated the remaining four parameters ($|\kappa|^2$, a, δ_0, and δ_V) based on a Levenberg-Marquardt nonlinear fitting algorithm [110, 111]. We are interested in quantifying the wavelength variations of $|\kappa|^2$, a, and the "phase slip" parameters δ_0 and δ_V.

Figure 6.3 shows the extracted parameters for the measured device. There are two possible mathematical solutions at each wavelength for $|\kappa|^2$ and for a, which

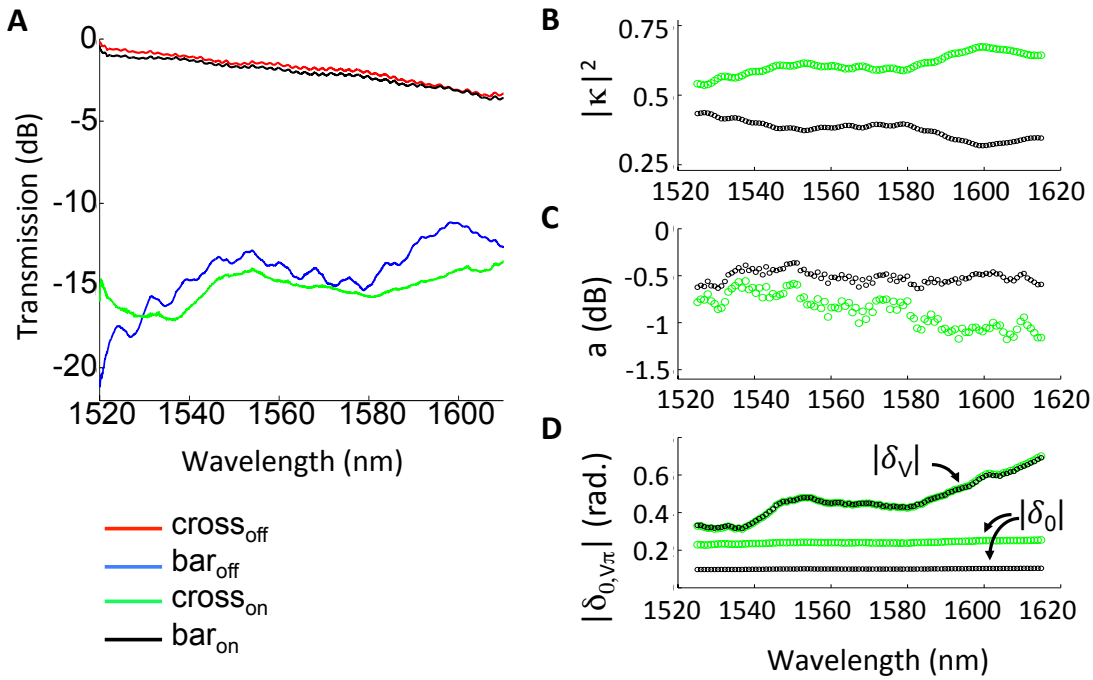

Figure 6.3: **A)** Transmission in the cross and bar output ports, at 0 V ($\text{cross}_{\text{off}}$ and bar_{off}), and $V_\pi = 4.25$ V (cross_{on} and bar_{on}) applied to the switching arm. Using the algorithm described in Section 6.2.1, the wavelength variation of the main device parameters were measured. There are two mathematical solutions, shown in black and green, and the physically meaningful ones are plotted in black. **B)** The coupling coefficient for the adiabatic 3-dB couplers (nominally 0.5). **C)** The loss induced in the "hot" state by the cascade of five phase shifters ($a = 0.5$ dB for five heaters implies 0.1 dB loss per heater section). **D)** The wavelength variations of the phase parameters which describe the phase slip from 0 or π phase. As shown by the flat lines for δ_0, there is no wavelength variation of the phase slip when no voltage is applied; however, there is significant variation with wavelength in δ_V. Note that both branches of the $|\kappa|^2$ solution result in the similar phase estimations for $|\delta_V|$.

are indicated by black and green colors[1]. The former was seen to be the physically-correct solution by performing an additional experiment: we measured the bar_{on} transmission (i.e. when voltage was applied to the heater) when the heater in the cross arm was heated, instead of in the bar arm[2]. In this new configuration, the bar arm was left unheated. (With reference to Fig. 6.2.A, we switched the role of the V_{bias} and the V_{mod} electrical contacts, while leaving the optical input, bar and cross pathways as indicated.) The bar_{on} transmission was seen to increase (by 0.5 dB) at all wavelengths. The two measurements are represented by

$$\begin{aligned} \text{bar}_{\text{on}}|_{(\text{bar heat})} &= 20 \cdot \log_{10} \left| a\, t^2 e^{i\delta_V} + |\kappa|^2 \right| \\ \text{bar}_{\text{on}}|_{(\text{cross heat})} &= 20 \cdot \log_{10} \left| t^2 + a\, |\kappa|^2 e^{i\delta_V} \right|. \end{aligned} \tag{6.3}$$

Assuming that the heaters are identical (and $0 < a < 1$), the observation that $\text{bar}_{\text{on}}|_{(\text{cross heat})}$ was greater than $\text{bar}_{\text{on}}|_{(\text{bar heat})}$ implies that $|\kappa|^2 < 0.5$. (On the other hand, $|\kappa|^2 > 0.5$ would result in the bar_{on} transmission decreasing when the cross arm was heated instead of the bar arm.)

As shown in Fig. 6.3.B, the nominally 3-dB couplers were, in fact, slightly imbalanced as a function of wavelength. However, even a 60-40 splitting imbalance has only a minor effect on $\text{cross}_{\text{off}}$, and mainly impacts bar_{off}, i.e. reduces the contrast between bar_{on} and bar_{off}. The ability to trim the splitting ratio may be useful in order to achieve higher contrast throughout a wide range of wavelengths.

Figure 6.3.C quantifies the attenuation in the "hot" arm of the MZI with $20 \cdot \log_{10}(a) \approx -0.5$ dB resulting from the cascade of five heaters in series. The loss of a single heater section is therefore about -0.1 dB, which is consistent with finite-element simulations [108]. This low loss suggests that a significant number of thermo-optic 2×2 switching elements can be cascaded, in order to build up a larger switching fabric.

A non-zero value of δ_0 limits the amount of interferometric cancellation between $-t^2$ and $|\kappa|^2$ in the expression for bar_{off} in (6.2), and also affects $\text{cross}_{\text{off}}$.

[1] For example, by examining the expression for 'bar_{off}' in (6.2) in the simple case when $\delta_0 = 0$, we see that $-t^2 + |\kappa|^2$ (for $|\kappa| > 0.5$ and $t < 0.5$) and $+t^2 - |\kappa|^2$ (for $|\kappa| < 0.5$ and $t > 0.5$) give the same numerical value.

[2] One could also heat the original arm while switching the input to the originally unused port, with similar arguments following.

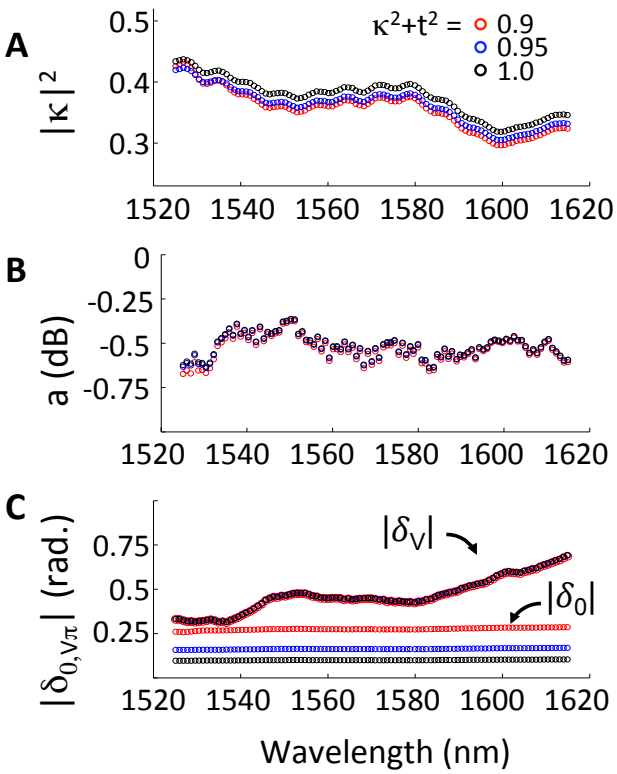

Figure 6.4: **A)** Similar spectral variations were extracted for the coupling coefficient under the three separate assumptions: no coupler loss ($|\kappa|^2 + |t|^2 = 1$, shown in black), or increasing amounts of loss, ($|\kappa|^2 + |t|^2 = 0.95$, shown in blue, and $|\kappa|^2 + |t|^2 = 0.90$, shown in red). **B)** For these three assumptions, the differences in the loss induced in the "hot" state were not significant. **C)** The three assumptions also gave essentially the same estimate regarding the variation of the phase slip with wavelength in the "hot" state, $|\delta_V|$. There is not much significance to the numerical value of the phase slip in the "cold state" $|\delta_0|$ since a spectrally-flat phase slip can be easily compensated for by heating the bias arm; however, the wavelength-dependent variations in $|\delta_V|$ cannot be compensated by a bias voltage simultaneously at all wavelengths and pose a fundamental limitation to the extinction ratio of the switch.

Going beyond this, we are interested in the difference between δ_V and δ_0, since ideally, the wavelength variation of both these quantities should be identical, if the heating-induced phase shift had no spectral dependency. However, Fig. 6.3.D shows that the relative phase slip, i.e. the difference between $|\delta_V|$ and $|\delta_0|$, is not spectrally flat. This is not unexpected, because the voltage-induced phase shift is both a function of wavelength and the effective index change due to temperature ($\Delta\phi \propto \Delta n_{\text{eff}}/\lambda$).

In Fig. 6.4, we investigate the assumption made earlier that $|\kappa|^2 + |t|^2 = 1$. Specifically, we made alternate assumptions that $|\kappa|^2 + |t|^2 = 0.95$ or $|\kappa|^2 + |t|^2 = 0.90$ (the latter implies nearly -0.5 dB loss per coupler, which is rather high for this fabrication technology). There was no observable difference in our estimate of a (heating induced loss), or of the wavelength variations of the phase-slip parameters (from 0 or π, respectively, when voltages of 0 and V_π volts were applied). Similar to previous investigations of coupler loss [88], under the assumptions of increased coupler loss, there were corresponding small reductions in the estimated $|\kappa|^2$ values, preserving the wavelength trends observed in Fig. 6.3. These results show that relaxing our earlier assumption that the couplers are lossless to practically-relevant coupler-loss numbers does not significantly change our conclusions regarding the wavelength variation of the switch parameters, or the approximate magnitude of the heating-induced loss.

Taken together, the results shown in Fig. 6.3.B–D suggest that the most significant reason for limited on-off contrast in Fig. 6.3.A was the heating-induced spectral slip. Although $|\delta_0|$ was flat over all wavelengths (since the couplers are adiabatic and the two arms of the MZI were fabricated with nearly exactly equal lengths), the wavelength variation of $|\delta_V|$ was clearly evident. For example, a phase imbalance of about 0.4 radians limits the on-off contrast to about 14 dB even if we were to take $a = 1$ and $|\kappa|^2 = |t|^2$. Referring back to (6.1), the bandwidth of this switch was limited by the properties of the matrix $\mathbb{P}$ (the phase shifter) rather than the matrix $\mathbb{C}$ (the couplers).

6.2.2 A digression on the choice of the coupling matrix

In (6.1) we used the transfer matrix [73]

$$\mathbb{C} = \begin{pmatrix} t & \kappa^* \\ \kappa & -t^* \end{pmatrix} \tag{6.4}$$

to model the directional couplers of the switch. This led to the curious result that, at zero bias, the upper arm experienced a phase difference of $\phi = \pi$ with respect to the lower arm, even though the two arms are equivalent. This zero-bias π phase shift comes about because of the negative sign in (6.4). Physically, $\mathbb{C}$ above says that light that passes "through" (i.e. not "coupled" across via κ or κ^*) the directional coupler will suffer no phase delay if traversing the top path via t and will suffer a π phase delay if traversing the bottom path via $-t^*$. It is because of this implicit π phase shift in $\mathbb{C}$ that the full MZI switch exhibits a zero-bias phase delay. This zero-bias phase delay can be eliminated by a redefinition of $\mathbb{C}$, which amounts to a choice of phase convention. Indeed, it is not uncommon to see other choices of $\mathbb{C}$ in the literature [47, 113].

In analogy with quantum mechanics, a proper transfer matrix should be both unitary ($\mathbb{C}^\dagger\mathbb{C} = \mathbf{1}$) and unimodular ($|\det \mathbb{C}| = 1$). From these conditions, the energy conservation condition, $|t|^2 + |\kappa|^2 = 1$, follows. E.g. enforcing unitarity and unimodularity on (6.4) immediately implies the energy conservation condition. With these constraints in mind, it is easy to write down other choices of $\mathbb{C}$. One such choice is

$$\mathbb{C} = \begin{pmatrix} t & \kappa \\ -\kappa^* & t^* \end{pmatrix}. \tag{6.5}$$

This matrix, when used in (6.1), yields the more common condition of $\phi = 0$ at zero bias. Additionally, when using the foregoing, $\det \mathbb{C} = 1$; however, when using (6.4), $\det \mathbb{C} = -1$. Apparently, $\det \mathbb{C}$ is directly related to the zero-bias phase shift of the MZI. Other valid choices for $\mathbb{C}$ can be obtained from (6.5) by moving the negative sign to any of the other matrix elements, as well as moving either or both of the complex conjugations to the corresponding opposite side of its diagonal, for a total of 16 possibilities. If the negative sign appears on an off-diagonal (diagonal) element, then $\det \mathbb{C} = 1$ ($\det \mathbb{C} = -1$).

Two other possible choices are

$$\mathbb{C} = \begin{pmatrix} t & \pm i\kappa \\ \pm i\kappa & t \end{pmatrix}, \tag{6.6}$$

where t and κ are both real. Two more choices are obtained by negating the t's in the foregoing. In all four of these cases, $\det \mathbb{C} = 1$.

In all of the above examples, the overall device operation remains unchanged. The only thing that changes is the initial zero-bias phase reference. Therefore, making a choice for the form of $\mathbb{C}$ is to make a choice for one's phase reference. In the foregoing, we have briefly discussed 20 of the most "common" choices[3]. Another 20 choices are possible by multiplying (6.5) by -1, and by moving the i's in (6.6) to the diagonal elements, and then stepping through the above discussions. Many more sadistic choices may be possible by appropriate use of unit-magnitude complex exponentials.

6.2.3 Comparison with silicon-photonic carrier-injection switches

The algorithm we have developed in Section 6.2.1 can be used to study the performance of any silicon photonic 2×2 switch. To demonstrate this utility, we have taken the data for two other broadband, energy-efficient, silicon, electro-optic switches from the literature [2, 3]. In both these cases, current was driven through the silicon waveguide cross-section in close proximity to the optical mode; however, both relied on the refractive index shift due to the plasma dispersion of injected free carriers [32]. The design goals are the same as those of our thermo-optic switch: to minimize the on-state insertion loss and to achieve wideband operation with low crosstalk.

Figure 6.5 shows the results of running the extraction algorithm on the present, as well as on the literature, devices. Subsequent reconstruction of the experimentally measured transmission spectra, which makes use of the extracted

[3] Realistically, (6.4)–(6.6) seem to be the most common choices. The other 18 are then already within the realm of strange and unusual. The many more after that, therefore, may be mathematically correct but practically unnecessary.

device parameters, are also shown; this confirms the validity of the extracted parameters. In the case of the literature devices, a similar procedure, as discussed in the foregoing section, was performed to ascertain the correct solution branch. In the respective reports, data is also included for the on-state bar transmission when the input port is switched to the originally "unused" port (e.g. switching the input port to the unlabeled port in Fig. 6.2.A). The correct solution branch is then identified by comparing the relative magnitudes of this second bar transmission to the on-state bar transmission in Fig. 6.5.

Several observations can be made by comparing the columns of Fig. 6.5. The slightly improved transmission contrast of the literature devices occurs as a result of values of $|\kappa|^2$ closer to 0.5 as well as more favorable phase slips. In the present device, the improved tracking of the cross_{on} response to the bar_{off} response occurs because of less on-state loss. In all three devices, the presence of loss limits the attainable transmission contrast as per (6.2).

The measured insertion loss a (in units of decibels) for a single 2×2 switch element can be used to estimate scalability. For example, the overall insertion loss L (in units of decibels, excluding waveguide crossing loss) in an $N \times N$ Cantor-type switch architecture scales with the number of ports N as

$$L = 2\left(\log_2(N-1) + \log_2\left(\log_2 N\right)\right) a. \tag{6.7}$$

For $N = 64$, the foregoing yields 8.5 dB of insertion loss for the current-driven thermo-optic switch reported here, and 17 dB [2] and 29 dB [3] for the carrier-injection switches.

6.3 Device response time and digital driving

Large scale switching fabrics constructed out of 2×2 building blocks will require many electrical control signals. It will be more convenient if these signals can be obtained from digital ports of an FPGA or microcontroller, e.g. via pulse width modulation, rather than analog output ports. However, a digital driving waveform potentially causes a concern with ringing, e.g. as is seen in the current MEMS implementation of switching in MORDIA [100]. In this section, we

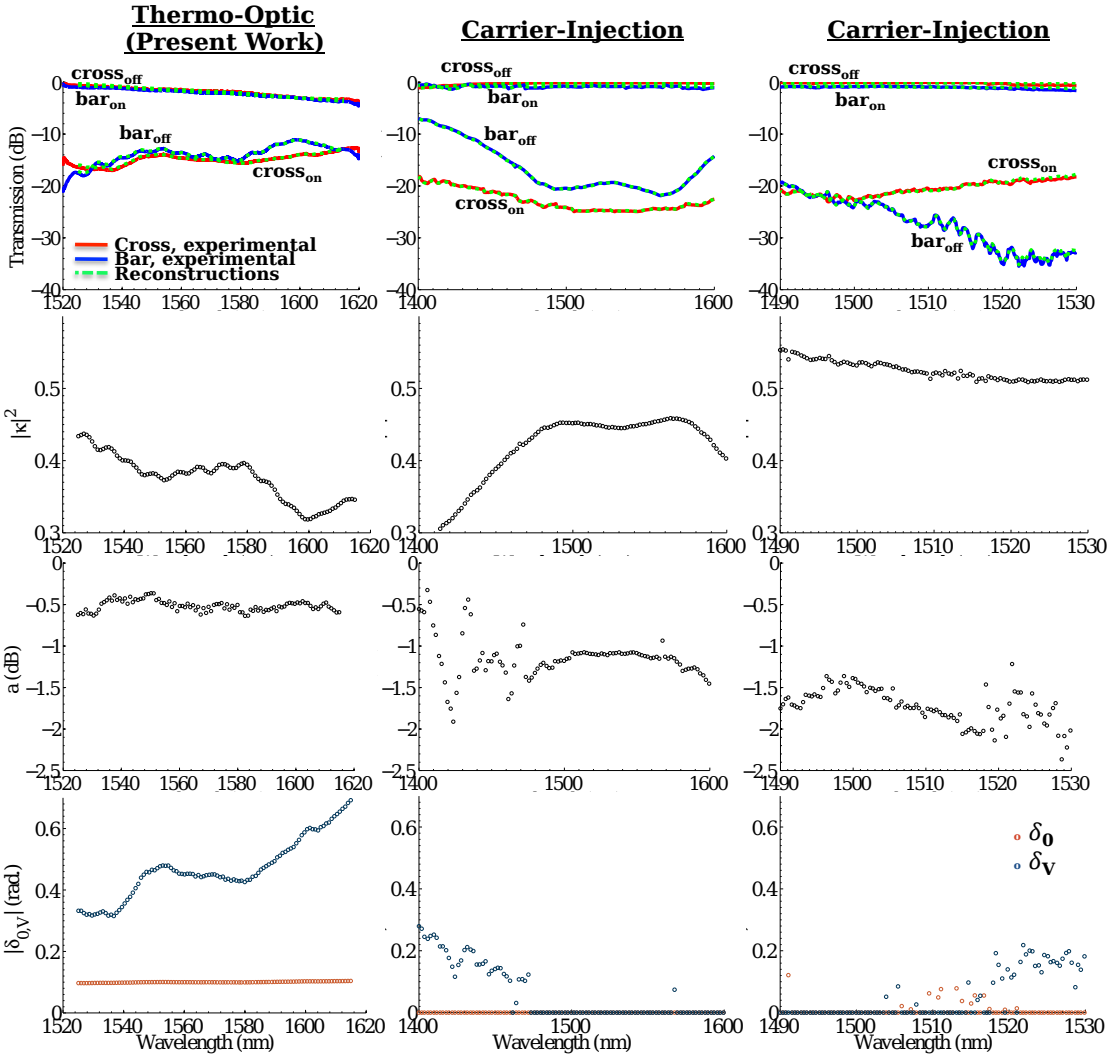

Figure 6.5: Comparison of transmission spectra, intensity coupling coefficients $|\kappa|^2$, on-state loss a, and phase slips δ_0 and δ_V for three different devices. **First column:** data for the present device. **Middle column:** transmission data from Ref. [2]. **Last column:** transmission data from Ref. [3]. The parameters in each column; $|\kappa|^2$, a, δ_0, and δ_V; were extracted from the respective transmission spectra in the first row. For the sake of device-to-device comparison, each set of transmission spectra is normalized to the maximum of its respective $\text{cross}_{\text{off}}$ response. Note that the abscissas are different column-to-column since the devices are optimized for different spectral regions.

discuss how to avoid ringing when driving the heater control using on-off voltage pulses. Experimental measurements include eye patterns and bit-error-rate sensitivity measurements of switching with both analog and digital drives.

6.3.1 Digital heating: analytic formulation

The goal of this analysis is to obtain insight into the physical mechanisms governing a compact heater driven by a time-varying waveform. For reasons discussed in Section 6.2.1, the length of the heated section should be kept short. The silicon waveguide itself is narrow, and the surrounding oxide is a relatively poor conductor of heat. For these reasons, it is reasonable to approximate each heated section of the waveguide by a one-dimensional model, with x being the spatial coordinate along the light path, and $x = 0$ defining the location of the heat source (see Fig. 6.2.B). Light propagating past the hot spot picks up a phase around $x = 0$ over a length scale that is defined by the temperature profile. We will show this length varies inversely with the square root of the driving frequency.

Because of the symmetry of the x coordinate, we can solve for the heat distribution $u(x,t)$ in the semi-infinite region $x > 0$ only, and extend the solution to negative x. We seek the solution to the homogeneous diffusion equation,

$$\frac{\partial u}{\partial t} = k\frac{\partial^2 u}{\partial x^2}, \qquad \text{with} \quad \begin{array}{l} u(0,t) = g(t), \\ u(x,0) = 0, \end{array} \tag{6.8}$$

where k is the thermal diffusivity of silicon (the thermal conductivity divided by the product of the mass density and the specific heat capacity) and $g(t)$ is the driving waveform, which is proportional to the product of the square of the electrical current and the electrical resistance, at $x = 0$. We expand $g(t)$ in a Fourier series,

$$g(t) = \sum_n A_n e^{i\Omega_n t}, \tag{6.9}$$

where, since $g(t)$ is a digital (e.g. pulse-width modulated) waveform with period T, $A_n \propto 1/n$ and $\Omega_n = n2\pi/T$.

We can solve (6.8) by expanding $u(x,t)$ in normal modes, paralleling (6.9),

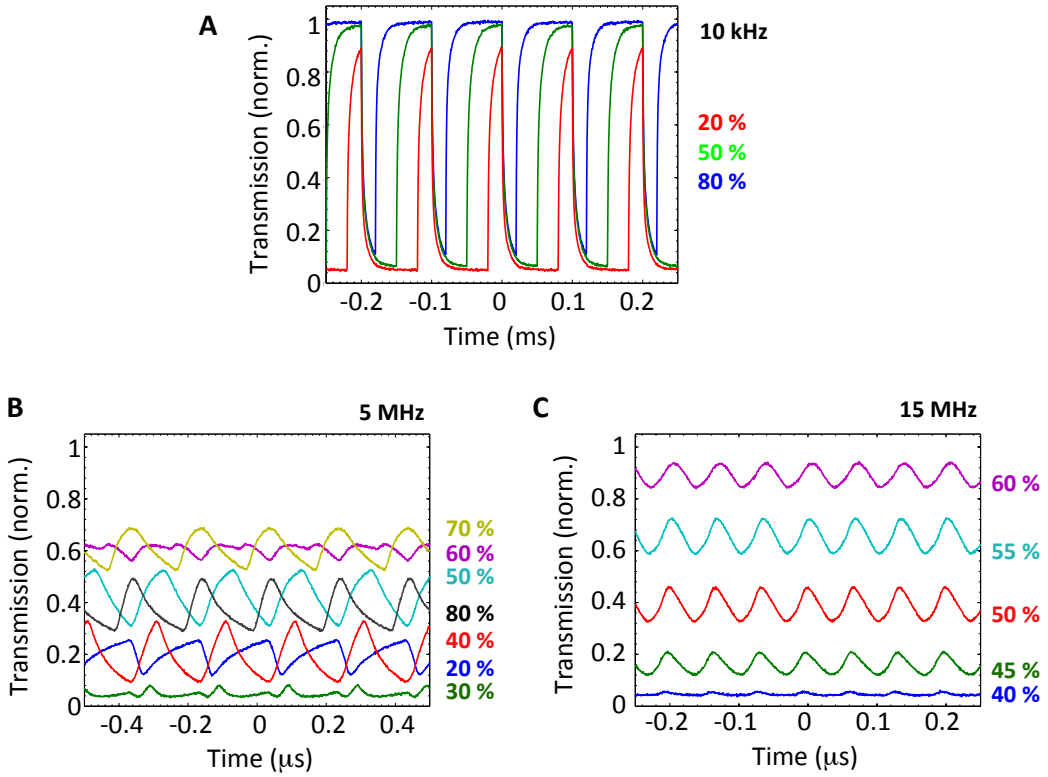

Figure 6.6: Pulse-width modulation of a digital heater drive (10 V amplitude), with different duty cycles as indicated by the percentages. **A)** Using a slow (10 kHz) drive, the rise and fall time constants were measured to be 11.1 μs and 11.3 μs, respectively, at 50% duty cycle. **B, C)** Here, both the drive frequencies were greater than the inverse of the time constants. The vertical axis shows the cross-state transmission when a heating voltage was applied, i.e. the desirable transmission was as close to 0 as possible with minimum ripple. The results show that at the lower frequency (**B**, 5 MHz), the residual ripple at the frequency of the drive signal was greater than at a higher frequency (**C**, 15 MHz), in accordance with the discussion in Section 6.3.1.

and taking the Fourier sine transform of each term, defined as

$$\tilde{u}_n(\omega, t) = \frac{2}{\pi} \int_0^\infty u_n(x, t) \sin \omega x \, dx, \tag{6.10}$$

resulting in

$$\frac{d\tilde{u}_n}{dt} + k\omega^2 \tilde{u}_n = \frac{2}{\pi} k\omega A_n e^{i\Omega_n t}. \tag{6.11}$$

Multiplying both sides by $\exp(k\omega^2 t)$ and integrating, we obtain,

$$\tilde{u}_n(\omega, t) = \left[\tilde{u}_n(\omega, 0) + i\frac{2}{\pi} \frac{k\omega \, A_n}{\Omega_n - ik\omega^2} \right] e^{-k\omega^2 t} + \frac{2}{\pi} \frac{\omega \, A_n}{\omega^2 + i\Omega_n/k} e^{i\Omega_n t}. \tag{6.12}$$

Since the first term on the right-hand side quickly decays to zero, we inverse-transform the second term, resulting in

$$u_n(x, t) \approx A_n \exp\left(-\sqrt{\frac{\Omega_n}{2k}} x \right) \cos\left(\Omega_n t - \sqrt{\frac{\Omega_n}{2k}} x \right). \tag{6.13}$$

Thus, for each harmonic mode of the driving waveform, the heat spreads out to a distance of about $L_n \equiv \sqrt{2k/\Omega_n}$ on either side of the origin. Based on k (silicon) = 0.8 cm^2/s, $2L_1 = 2.6$ μm for the $n = 1$ Fourier component for an alternating current at 15 MHz, as used in the experiment described in Section 6.3.2. In general, the effective heater length (and hence the thermally-induced phase shift picked up by the light as it propagates past the hot spot) can be controlled not only by amplitude but also by frequency, and by shaping the waveform, i.e. the amplitude coefficients in the harmonic mode expansion.

To achieve error-free switching of 10 Gbps data, we are only interested in the DC term and inhibition of ringing. For higher Fourier components, $L_n \propto n^{-1/2}$, and further, $A_n \propto n^{-1}$, and only odd integers contribute in the case of a pulse-width modulated square wave, i.e. the effective length and amplitude of the heater for the higher harmonic components are too small to affect light propagation. As such, we may expect that for AC frequencies that are *high enough*, there is no Gibbs phenomenon associated with a digital current drive, whereas some ripples may be seen if the frequency is *reduced*. In the latter case, the heat waves can spread out over a longer distance, and thus, interact with the optical field over a longer length, imparting a harmonic oscillation to the optical field. This is indeed seen in the experimental traces shown in Fig. 6.6, where some ringing was observed at 5 MHz, but not at 15 MHz.

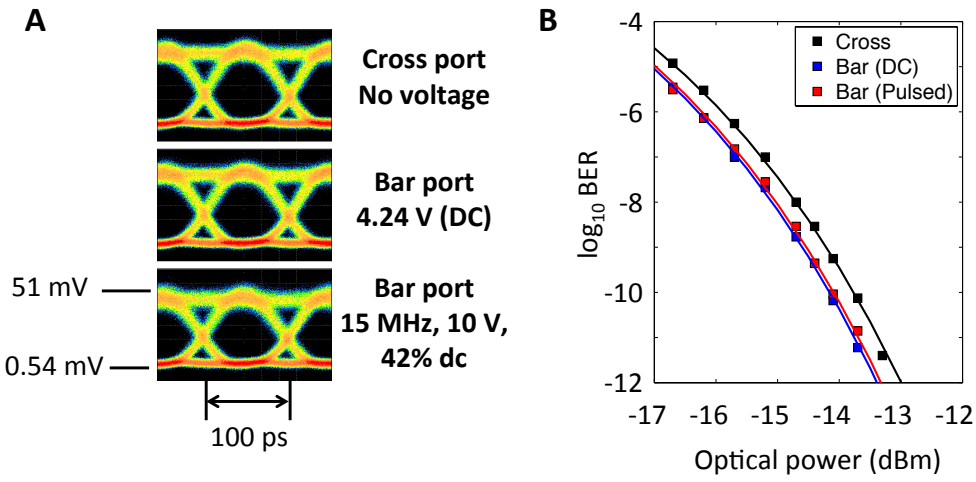

Figure 6.7: **A)** 10 Gbit/s eye patterns of cross and bar states (analog and digital drives) for a selected channel at 1558 nm. **B)** Bit-error-rate (BER) power sensitivity curves, showing no penalty between analog and digital voltages for switching. The optical power labeled on the horizontal axis was measured at the detector.

6.3.2 Experimental verification

Figure 6.7.A shows a typical eye pattern of a single 10 Gbit/s channel, demonstrating that the same bar switching behavior was achieved by driving the heaters with a 10 V, 42% duty cycle, rectangular waveform at 15 MHz, thus mimicking $V_\pi = 4.2$ V. In Fig. 6.7.B, bit-error-rate sensitivity curves versus power are shown in the three cases, at a representative wavelength of 1550 nm, using a $2^{31} - 1$ pseudo-random bit sequence. There was no observable penalty difference between the DC drive and pulsed-drive bar states. The bar states show slightly better sensitivity curves than the cross state. The cause for this difference is still under investigation; one possible factor is that the switching voltage V_π to enable bar transmission was fine-tuned to its optimum value, whereas no voltage was applied to the MZI in the cross state, resulting in a small phase-slip in the latter case at the measured wavelength.

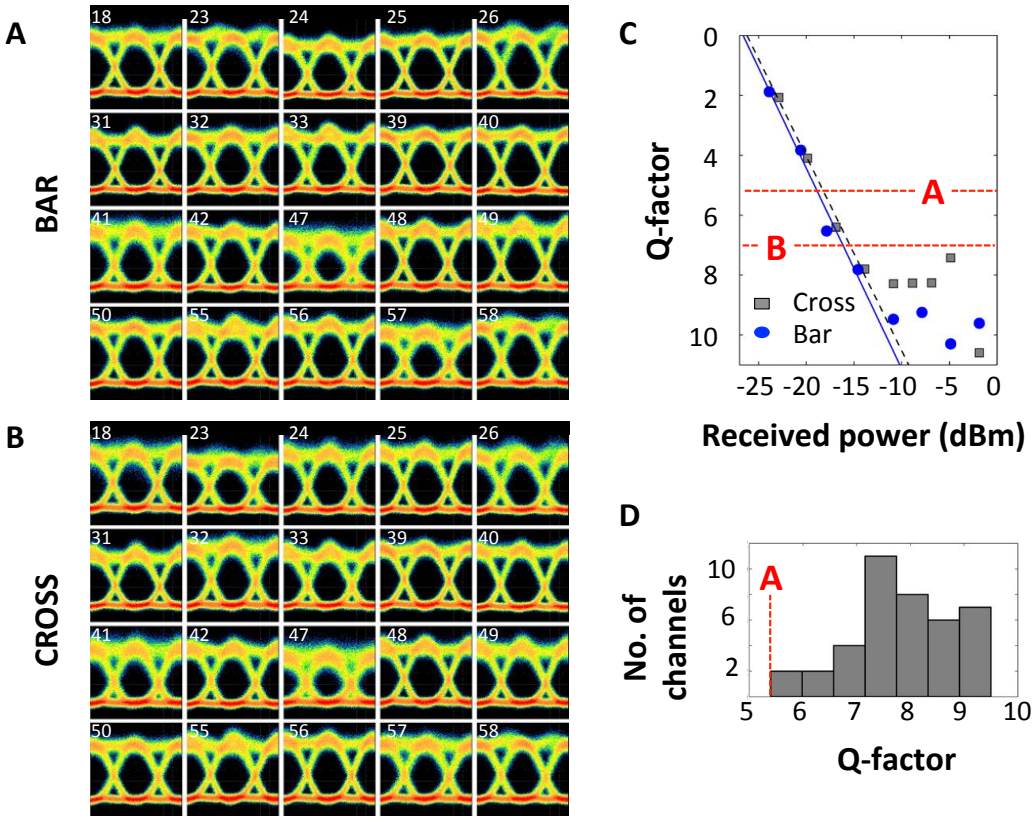

Figure 6.8: 10 Gbit eye patterns (labeled by ITU-T DWDM channel number) in the bar (**A**) and cross (**B**) states for server-driven data. Channel-to-channel differences correspond to normal variations in the ring (cf. Fig. 6.1.C). **C)** For a single channel at 1558 nm, Q-factor versus received power curves for the cross and bar states are nearly identical. Horizontal red dashed lines 'A' and 'B' refer to an estimated packet loss rate of 10^{-4} and estimated BER of 10^{-12}. **D)** The histogram of Q-factors, with all channels above the 'A' threshold.

6.4 Performance in the network

The device was tested during live operation of the 20-wavelength network, i.e. all wavelengths "loading" the device simultaneously. The network's wavelength channels all lie within the highest transmission-contrast region of Fig. 6.3.A. The transmitters were commercial DWDM SFP+ form-factor modules transmitting 10 Gbit/s data, fed from computer servers, in 9000-bit-length TCP packets over single-mode fiber at a power level between 1-3 dBm. Since no line-by-line equalization of power levels in the ring is performed under normal operating conditions, this results in non-uniform channel powers, as shown in Fig. 6.1.C. Figure 6.8 shows open eyes for all indicated channels (20 x 10 Gbps) in both the cross and bar states. As there was no difference between the bar (Fig. 6.8.A) and cross (Fig. 6.8.B) eyes for all channels, the channel-to-channel variations can be attributed to the normal differences in circulating power in the network itself.

Further evidence of the satisfactory performance of the silicon switch was obtained from the Q versus received power measurements, where Q is the signal-to-noise ratio measured from each eye pattern. In these measurements, a linear PIN detector was used with 15 GHz bandwidth, 300 V/W responsivity, and noise equivalent power NEP = 30 pW/Hz$^{1/2}$, with optical pre-amplification using an erbium-doped fiber amplifier and ASE filtering. As shown in Fig. 6.8.C, both sensitivity lines have very similar slopes. The dashed red line labeled 'B' in Fig. 6.8.C represents a bit-error-rate (BER) of 10^{-12}, assuming Gaussian statistics, and the line labeled 'A' represents a packet loss probability $p = 1 - (1 - \mathrm{BER})^L = 10^{-4}$, where $L = 9000$ is the number of bits in a packet. All the eyes measured in Fig. 6.8.A and Fig. 6.8.B were above this threshold, an order-of-magnitude lower (better) than the typical packet error rate under normal software operating conditions or due to congestion, buffer overflows, TCP incast, etc.

6.5 Summary

This chapter summarized the testing and component-level modeling of an energy-efficient wideband silicon-photonic switching element used in a wavelength-

division multiplexed ring network, whose good performance imposes no penalty on the normal operating error threshold of the network. The parameter-extraction method shows how to relate transmission measurements of switches (which are commonly reported) to the intrinsic device parameters (which are less commonly reported). The extracted parameters of this thermo-optic switch were compared to those of two recently-published electro-optic switches. In all these cases, an electrical current is driven in close proximity to the optical mode, in order to increase the efficiency and reduce the switching speed of the underlying physical phenomena. Minimizing wavelength variations of the passive couplers (e.g. by improving the adiabatic coupler design or by using MZI splitters [2], multi-mode interference couplers [3], or three-waveguide couplers [115, 116]), as well as of the phase-shifting mechanism itself, will be important to achieve truly wideband crossbar switching and fully exploit the spectral transparency advantage that optical switching has over electronic switching.

In this demonstration, twenty server-driven 10 Gbit/s wavelengths between 1531 nm and 1563 nm with TCP packets were loaded onto the device and simultaneously switched between cross and bar ports in a footprint of 0.6 mm $\times$ 0.05 mm. Although five heating elements were cascaded and driven in parallel to reduce the required voltage, the cascaded losses in the thermo-optic "hot" arm were only about 0.5 dB. The power consumption of this "fat pipe" switch is about 15 mW, and no thermo-optic cooling or temperature stabilization was required. The measured on-off switching time constant of the chip-based switch was 11 μs, which is about the same as the loss-of-light time in the current bulk-optics implementation of the MORDIA network architecture [100]. Possible improvements in the device performance may be made by reducing the contact resistance [117], or by implementing a dual-drive operation [118, 119]. In summary, this compact, broadband, microsecond-scale switch can be a low insertion-loss building block for scalable switching fabrics using silicon photonics.

Chapter 6, in part, contains material, published in the following, of which the dissertation author was the primary investigator:
R. Aguinaldo, A. Forencich, C. T. DeRose, A. Lentine, D. C. Trotter, Y. Fainman,

G. Porter, G. Papen, and S. Mookherjea, "Wideband silicon-photonic thermo-optic switch in a wavelength-division multiplexed ring network," *Opt. Express*, vol. 22, no. 7, pp. 8205–8218, 2014 [120].

Chapter 7

A Network-Node-on-a-Chip

7.1 Overview

In data center applications, optical interconnects have secured their place providing point-to-point links between the top-of-rack (ToR) switches that route rack-to-rack data communications. More recently, optical circuit switching has been proposed and demonstrated to relieve data-center network congestion by taking over the routing of larger flows of data traffic [12, 121, 90]. The use of bulk optical components, however, can be energy and cost prohibitive when scaled up to the size of industrial warehouse-scale computing. As such, chip-scale solutions for data centers have been discussed with the envisioning of an optical rack-on-a-chip [122] or network-on-a-chip [123]. As a first step towards these goals, we present an optical network-node-on-a-chip (NNoC) that has been designed for the MORDIA (Microsecond Optical Reconfigurable Datacenter Interconnect Architecture) network at the University of California, San Diego. This optical application-specific integrated circuit replaces the conventional off-the-shelf (COTS) optical hardware in a network node by functional equivalents in silicon photonics. When manufactured in volume, silicon photonic chips are expected to be greatly inexpensive, compared to their COTS counterparts [124, 93].

MORDIA is an optical circuit-switched, dense wavelength division multiplexed (DWDM), ring network [90]. It supports up to 24 hosts, representing 24 wavelength channels located throughout the C-band on the 100-GHz-spaced ITU-T

telecommunications grid, from channels 15 (1565.50 nm) to 58 (1531.12 nm). Each channel carries 10-Gbps data streams. The network is set up in a ring topology with six nodes and four hosts per node; however, 23 of the 24 hosts are utilized in the current implementation. In the present work, we scale the COTS hardware of a single node to a chip-scale solution (Fig. 7.1.A-B). Like the COTS architecture, our NNoC has the same architectural blue print for all six nodes and can be dynamically reconfigured to address the pertinent channels in a specific node. As such, the maintenance complexity and replacement-upon-failure costs are minimized.

The on-chip network node is functionally divided into three sections in Fig. 7.1.C: *VOA, Drop*, and *Add*. The five input ports (labeled Drop/Express, A, B, C, D) lead directly to variable optical attenuators (VOA), which allow for electrically controlled equalization of the input signal powers. The Drop/Express input accepts, from the network, the full bandwidth of 23 channels and leads to the DROP section where four adjacent channels ($\sim$400 GHz bandwidth) are removed from the data stream and 19 are "expressed" to the output. Four new channels, corresponding to the dropped wavelengths, are added to the express channels via inputs A, B, C, D and the ADD section. The foregoing functionalities are summarized in Fig. 7.1.D. The various sections are elaborated on below. The output of the ADD section (containing the four multiplexed ADD channels) is combined with the output of the DROP section (containing the 19 express channels) via a wavelength-flattened, adiabatic, 3-dB coupler [105].

The fabrication of the NNoC was performed using a fully CMOS process, at Sandia National Laboratories, on a 6-in. silicon-on-insulator (SOI) wafer. Fully etched, 400-nm wide, silicon strip waveguides were oxidation trimmed to 230-nm thickness and oxidation smoothened for sidewall-roughness reduction and mitigation of optical propagation loss. Dopant implants were used for the formation of electronic diodes and resistive heaters. A single metal layer provided for low-resistance Ohmic contacts and electrical connections. Light was coupled on/off the chip via inverse-tapered waveguides, which were all located on one side of the chip on a standard 127-μm pitch to facilitate coupling with a multi-fiber v-groove array. An increased insertion loss of $\sim$10 dB/facet was suffered because of a waveguide-

to-fiber mode-field-diameter mismatch of 2.5 to 10.4 μm. Although this loss did not impact the reported measurements, improvements are generally desirable and can be attained with tapered-fiber arrays [126].

7.2 Drop functionality

A scale diagram of the ring resonators that constitute the DROP section is shown in Fig. 7.2.A. The 4.5-μm diameter rings are based on an adiabatically transitioned design [60] in which the waveguide width evolves from 325 nm in the coupling region to 650 nm at 90° away. The inner portions of the wide-waveguide sections are n+ doped thus forming a resistive heater that can thermo-optically affect the optical mode, which traverses primarily through the undoped outer portions. The tether-like structures interior to the rings are also n+ doped and provide for electrical contact to the metal layer through which the spectra of the rings can be electrically tuned. Wavelength channels that are resonant with the rings are dropped to grating couplers (visible as yellow terminations in Fig. 7.1.C), which can be used for diagnostics.

The full transmission spectrum of the DROP section (i.e. the sum effect of the four rings) is plotted in Fig. 7.2.B for several applied voltages. The unmodified spectrum of the MORDIA channels is plotted in grey in Fig. 7.2.C. The DROP section is designed to extinguish a bank of four adjacent channels (∼400-GHz bandwidth) while allowing the remaining 19 to propagate through. The three cases in Fig. 7.2.C, which correspond to three different applied voltages, show this functionality. A wide free spectral range (FSR) of 5.4 THz (43 nm) allows for any group of adjacent MORDIA channels to be dropped without inadvertently dropping another group. Nominally, the rings are electrically connected in parallel ($R_{\mathrm{eq}} \approx 160\ \Omega$); they exhibit an electrical tuning efficiency of 17 mW per band of four channels (21 mW/THz) when biased in the Ohmic regime.

Eye patterns for Ch. 34, corresponding to the top two panels of Fig. 7.2.C, are plotted in Fig. 7.2.D. When Ch. 34 is allowed to propagate (Fig. 7.2.C, middle), an open eye is observed at the output with a signal-to-noise ratio (SNR) of 7.6.

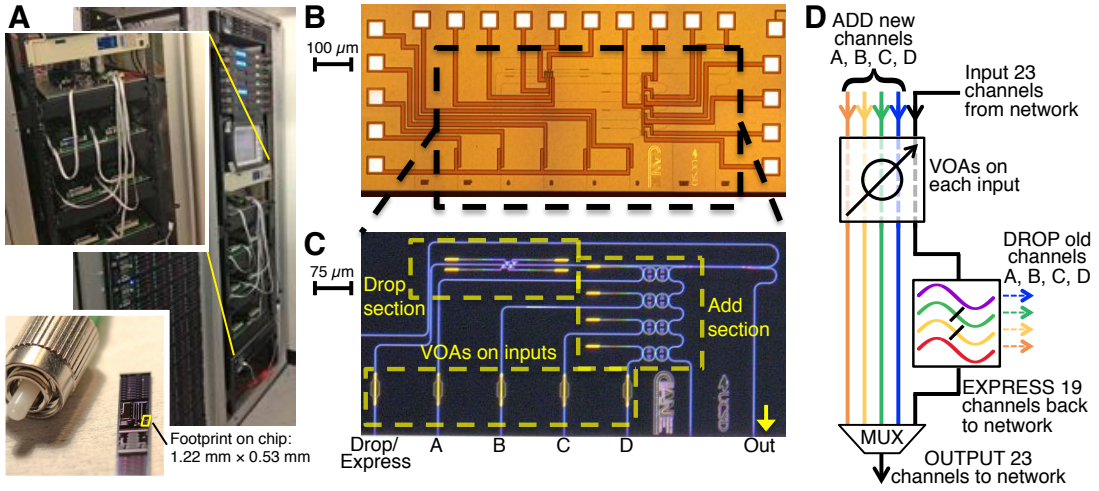

Figure 7.1: **A)** Rack mounted hardware that makes up the MORDIA network. The top half of the main image contains, from top to bottom, erbium-doped fiber amplifiers and an optical spectrum analyzer. The top inset shows the remaining optical hardware used in each host. For size comparisons, the bottom inset shows the chip that contains the NNoC next to the FC-connectorized end of a fiber patch cord. **B)** Bright-field micrograph of the optical integrated circuit. The waveguides are mostly visually blocked by the metal lines that provide electrical connections to the VOAs and waveguide heaters. Note that the bottom edge of the micrograph is the actual edge of the chip where light may be coupled to optical fiber. **C)** Dark-field micrograph, of the boxed section in (B), for another chip that did not undergo metallization. The optical circuitry is more readily visible. The major sections, as well as the input/output ports, are marked. **D)** Diagram of the NNoC's functionality. Note that the MUX symbol is a diagrammatic simplification; actual multiplexing on the chip occurs in several stages. The add channels are multiplexed by virtue of the four add filters outputting to a common bus waveguide. After the ADD section, this bus waveguide leads to the first input of a 3-dB coupler. The other input of the coupler is connected to the output of the DROP section. Thus, the add and express channels are multiplexed at the 3-dB coupler.

When Ch. 34 is "extinguished," the eye closes to zero SNR; i.e. although "full" extinction of dropped channels does not occur as per Fig. 7.2.C, Fig. 7.2.D indicates that the dropped channel is indeed extinguished in the sense that the amplitude of the dropped signal is completely within the decision threshold of a bit 0.

The eye patterns in Fig. 7.2.E correspond to the top panel of Fig. 7.2.C. Although expressed bands adjacent to the dropped band suffer from slight attenuation, this is seen to not be troublesome at the system level because clear open eyes are measured for the adjacent expressed bands at the output port. Similarly, the eyes in Fig. 7.2.F, which correspond to the middle panel of Fig. 7.2.C., are also open regardless of slight attenuation leaking over from the dropped spectral region.

7.3 Add functionality

The ADD section in Fig. 7.1.C consists of four parallel input paths that each lead to a two-stage (i.e. two-ring) ring-lattice filter. Similar to the DROP rings, the 85.3-μm-long ADD rings also contain integrated heater-waveguides by virtue of adiabatically transitioning the widths from 400 to 800 nm over each 90° arc. Within each individual filter, the two constituent rings are electrically driven in parallel ($R_{\text{eq}} \approx 2.2\text{k}\Omega$); however, the individual filters are electrically biased by different amounts such that the filter passbands correspond to the input wavelengths of the corresponding host transceivers. The 850-GHz (6.8-nm) FSR of the rings allows for an efficient tuning scheme in which the ADD section can be reconfigured to pass any desired four-element subset of the MORDIA channels. The four filters all output to a common bus waveguide, thus multiplexing the four ADD channels. The output signal from one filter is not dropped when it passes by the output of another filter because the filters are tuned to different resonant wavelengths. Another method of multiplexing the four ADD channels would be to use a 4-to-1 waveguide combiner; however, the present use of rings mitigates the minimum 6-dB/channel insertion loss of such a combiner.

The transmission spectra of the ADD filters are plotted, in Fig. 7.3.A,

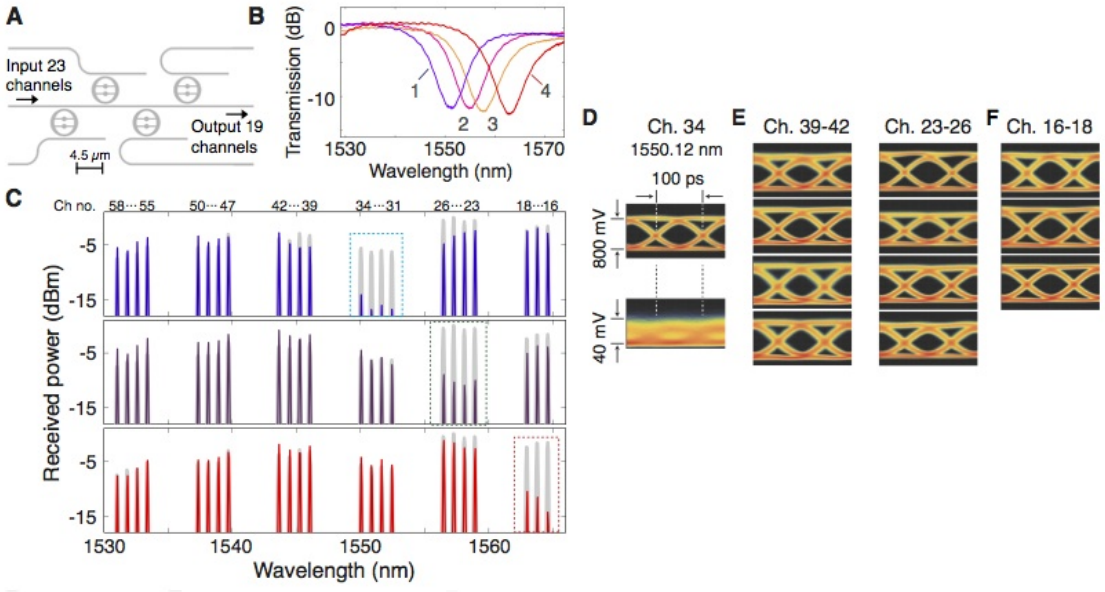

Figure 7.2: **A)** Scale diagram of the silicon layer of the DROP section. Each ring resonator drops resonant light to grating couplers, which are not shown. **B)** Transmission spectra of the DROP section corresponding to injected currents (i.e. under varying degrees of heating) of 0, 8, 10, and 13 mA for labels 1–4, respectively. **C)** MORDIA channel spectra measured before (grey) and after (colors) the drop section. From top to bottom, the panels correspond to the transmission spectra in (B) that are labeled 1, 3, and 4, respectively. As the DROP filters are tuned, a different group of four adjacent channels are dropped. **D)** Eye patterns for Ch. 34 corresponding to the first two panels of (C). As the filters are tuned from the middle to the top panel in (C), the eye closes by ∼26 dB indicating that a new Ch. 34 may be added to the network with impunity. **E)** Eye patterns corresponding to the top panel of (C), which indicate that adjacent (i.e. expressed or non-dropped) bands traverse through the DROP section without signal degradation. **F)** Eye patterns corresponding to the middle panel of (C), also indicating non-degradation of express channels.

under the specific configuration scheme to add channels 23–26 to the network (cf. Fig. 7.2.C, middle panel). At an ambient temperature of 25 °C, the pass-band of filter A coincided with Ch. 26. Filters B–D were electrically tuned, with an efficiency of 0.11 mW/GHz, to accept Ch. 25, 24, and 23, respectively. To test the ADD functionality, we modulated 10-Gbps non-return-to-zero (NRZ) pseudo-random bit sequences (PRBS), of length $2^{31} - 1$, onto a laser tuned to the aforementioned ITU-T channels and input through the corresponding filters. Eye patterns, plotted in Fig. 7.3.B, were obtained from the NNoC output and show large SNRs of 8–9.2. A bit-error-rate (BER) sensitivity test, plotted in Fig. 7.3.C, was performed by attenuating the input signals. The magnitude of the swing of the BER degrades only slightly, from 1.4 to 0.85 dB/decade, in moving from input A to D, indicating that the four ADD filters provide equal performance when using reasonable input powers. The degradation of the BER swing can be further mitigated by using filters with steeper transition bands.

7.4 Channel equalization

The VOA section of the NNoC is present because it is not uncommon for channel-to-channel power variations to occur in the network (cf. original power levels in Fig. 7.2.C). The full channel spectrum can then be equalized, line-by-line, via appropriate attenuation at the input/VOA stage. A micrograph of one of the VOAs is shown in Fig. 7.4.A. The VOA is formed by adiabatically transitioning from a fully etched strip waveguide to a ridge waveguide. A 150-nm-thick slab tapers in over a 25-μm length. Either side of the slab is doped n+ or p+ thus forming an electronic *p-i-n* diode. The optical mode propagates though the $\sim$2-μm-wide *i*-layer / depletion region of the diode with nominally zero additional loss. Under forward bias, electrons (holes) are injected into the depletion region from the cathode (anode) thus increasing the optical loss through free-carrier absorption [32]. The VOA cross-section is diagramed in Fig. 7.5.

A representative electrical *I-V* characteristic of the VOAs, when light is completely absent from the device, is plotted in Fig. 7.4.B. Using a theoretical

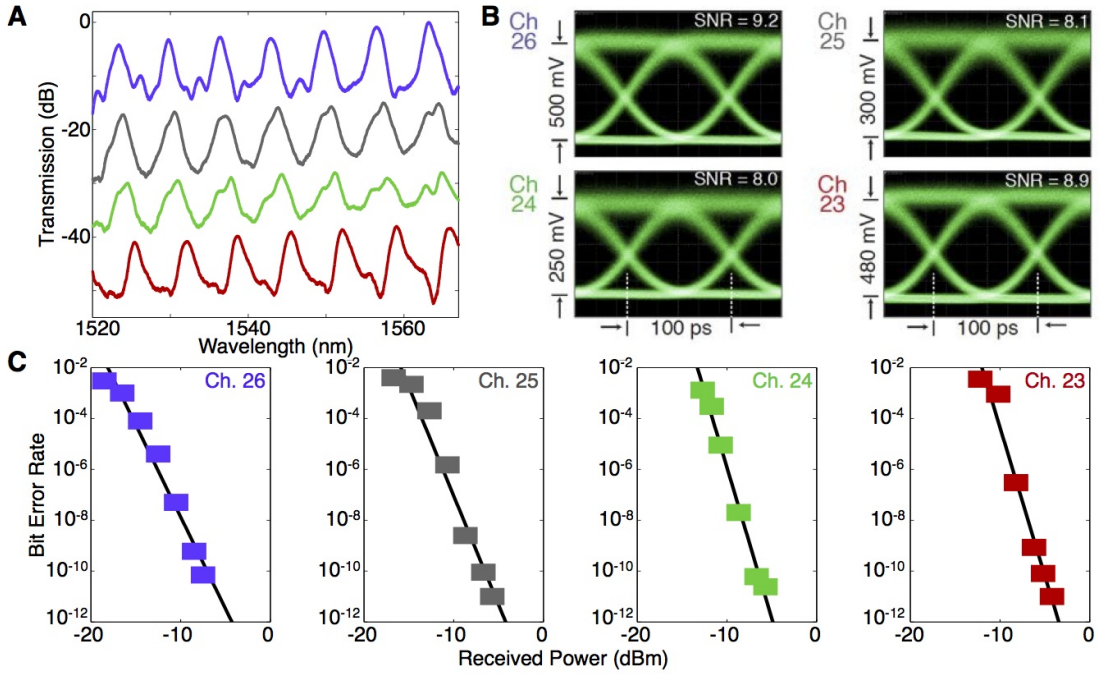

Figure 7.3: **A)** Transmission spectra of the four add channels, corresponding, from top to bottom, to inputs A–D, respectively. The spectra are vertically offset in the plot for clarity. **B)** Simultaneously multiplexed 10-Gbps NRZ eye patterns, for input Ch. 23–26, through filters D, C, B, and A, respectively. **C)** Bit error rates corresponding to the signals in (B).

least-squares fit (elucidated in the following section), the ideality factor n, reverse-bias saturation current I_0, and parasitic series resistance R_s of the diode were extracted from the measured data, as listed in Fig. 7.4.B. The value of $n = 2.06$, which is very close to the theoretical value of 2.0, indicates that the current is dominated by electron-hole recombination in the depletion region [127]. The contact resistance to individual contact pads was also determined to be 3.8 Ω; therefore, R_s occurs primarily because of the pad contact resistance indicating a very low contact resistance between the silicon and metal layers.

When light is passed through the VOA, the optical signal experiences an exponential decrease as a forward current is injected, as seen in Fig. 7.4.C, with an attenuation efficiency of 0.13 dB/mA. To exemplify the use of multiple VOAs for channel equalization, we concurrently inserted ITU-T channels 55–58, into the ADD inputs, with varying line-to-line powers. The grey trace in Fig. 7.4.D shows

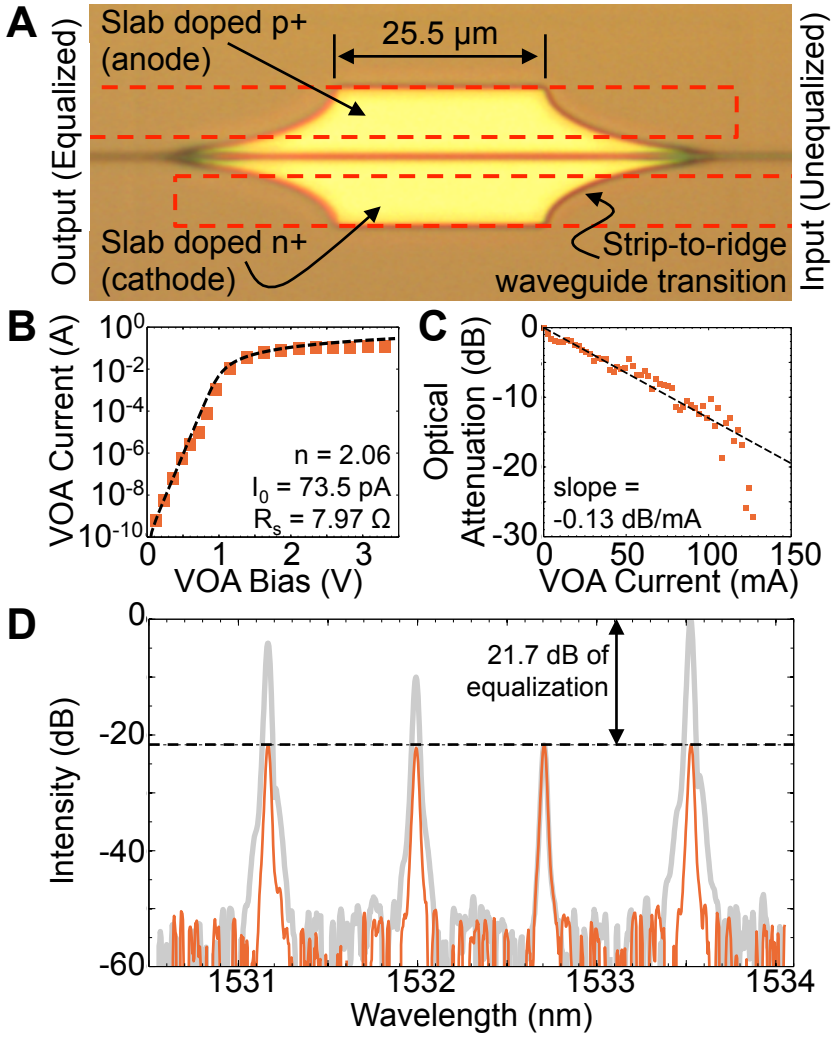

Figure 7.4: **A)** Micrograph of a VOA without the metal layer. The dashed lines indicate the positions of the metal lines. **B)** Dark current under forward bias; measurement (markers) and parameterized fit (dashed curve). The extracted diode ideality factor n and saturation current I_0 indicate an electrically well-functioning device. The extracted series resistance R_s is close to the measured contact resistance indicating low intrinsic parasitics. **C)** Optical attenuation as a function of injected forward current; measurement (markers) and a fitted line. Optical power is exponentially decreased by current injection up to sizable currents exceeding ~100 mA. **D)** An example of spectral equalization using the VOAs. The grey spectrum is the original multiplexed output of the ADD section. The orange spectrum is obtained by forward biasing the VOAs that correspond to the first, second, and fourth channels.

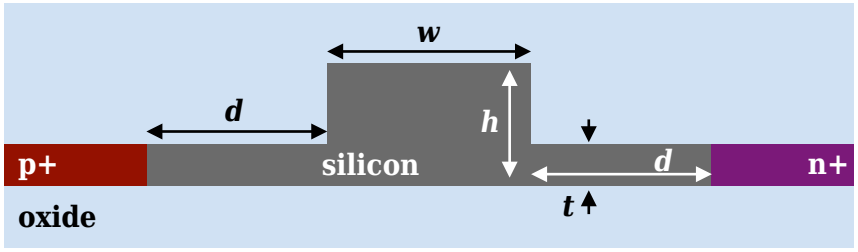

Figure 7.5: Cross-section of the VOA; $h = 230$ nm, $w = 400$ nm, $t = 150$ nm, $d = 1$ μm.

the spectrum at the output port. This spectrum is undesirable for network operation if consistent powers are needed among the channels. However, by individually biasing each of the VOAs at the inputs, the output spectrum can be equalized, as shown in the orange trace. Since replicates of the chip are to be used at each network node, the entire network can be equalized by the foregoing method.

7.5 Extraction of VOA's diode parameters

In the preceding section, the diode ideality factor n, reverse-bias saturation current I_0, and parasitic series resistance R_s were extracted from the measured dark I-V curve of the VOA in Fig. 7.4.B. Following a similar method [128], the p-i-n diode that constitutes the VOA can be modeled as an ideal diode in series with a parasitic resistor, as diagramed in Fig. 7.6. In the diagram, V is the applied bias and I is the measured current; these are the parameters that are plotted in Fig. 7.4.B. We may directly write down the Shockley equation:

$$\begin{aligned} I &= I_0 \left(\exp\left(\frac{eV_D}{nkT} \right) - 1 \right) \\ &\approx I_0 \exp \frac{eV_D}{nkT} \\ &\approx I_0 \exp \frac{e\left(V - IR_s\right)}{nkT}, \end{aligned} \tag{7.1}$$

where e is the elementary charge, k is Boltzmann's constant, T is the temperature, and the potential V_D across the diode (i.e. across the depletion region in the physical device) is not equal to the applied bias because of the parasitic resistance (e.g. voltage drops may occur over the quasi-neutral regions of the physical diode

or over contact regions). Differentiating the foregoing yields

$$I\frac{dV}{dI} = IR + \frac{kT}{e}n. \tag{7.2}$$

Therefore, making a plot $I\frac{dV}{dI}$ vs. I and fitting a line to the linear portion yields two of the desired parameters: the slope of the fitted line is precisely R_s and the intercept to the ordinate is proportional to n. For the data corresponding to Fig. 7.4.B, we use the region between 2–40 mA yielding $R_s = 7.97\ \Omega$ and $n = 1.74$. Note that this value of n is only an estimate, which will be improved upon in the following, because the electrical region in which series resistance is easily extractable (i.e. the linear region of $I\frac{dV}{dI}$ vs. I) corresponds to relatively large cases of carrier/current injection where the diode is no longer operating in an "ideal" sense.

The saturation current I_0, as well as a more accurate n, may be determined by first rewriting the last line of (7.1):

$$\ln I = \frac{e}{nkT}V_D + \ln I_0. \tag{7.3}$$

Since R_s is now known, and I and V are experimental parameters, a plot can be made of $\ln I$ vs. V_D. Similar to before, a line is to be fitted to the linear portion; n is calculated from the slope and I_0 from the ordinate intercept. Care should be taken to make sure that the fitted region does not include values of V_D that are too small, otherwise the approximation that led to the foregoing equation may be invalid. For the data corresponding to Fig. 7.4.B, we use the region between 300–500 mV yielding $I_0 = 73.5$ pA and $n = 2.06$. The theoretical I-V curve, making use of the first line of (7.1) along with the extracted parameters, is plotted atop the experimental data in Fig. 7.4.B showing a good fit between the model and experiment.

7.6 Summary

The architecture of the reported NNoC achieves the needed functionality of a MORDIA network node on a small footprint while providing for easy tuning ad-

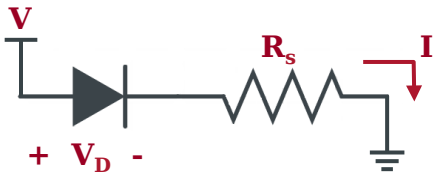

Figure 7.6: Equivalent circuit used to model a diode having a parasitic series resistance. The applied bias V is equal to V_D only if the parasitics are negligible.

justments to compensate for fabrication imperfections and to allow for network reconfigurations. The electronic tunability of the primary ADD and DROP functions allows replicates of the NNoC to be distributed to varying nodes, which contain different host channel wavelengths. This tuning functionality may also stabilize the chip's operation under ambient temperature fluctuations. The electronically controlled VOAs allow for full network equalization. After a simple calibration procedure, all the required tuning voltages and currents can be programmed via a look-up table for quick access. In summary, we have demonstrated the feasibility and usefulness of introducing an on-chip silicon-photonic network node into data center architectures to replace most of the conventional, off-the-shelf, rack-mounted network hardware.

Chapter 7, in part, contains material, to be submitted for publication as the following, of which the dissertation author was the primary investigator:
R. Aguinaldo, H. Grant, A. Forencich, A. Lentine, C. DeRose, D. Trotter, A. Pomerene, A. Starbuck, Y. Fainman, G. Papen, G. Porter, and S. Mookherjea, "A silicon photonic network node-on-a-chip for optically interconnected data-center networks," in-preparation (2014).

Bibliography

[1] W. E. Lamb, "Anti-photon," *Appl. Phys. B*, vol. 60, no. 2-3, pp. 77–84, 1995.

[2] J. Van Campenhout, W. M. Green, S. Assefa, and Y. A. Vlasov, "Low-power, 2x2 silicon electro-optic switch with 110-nm bandwidth for broadband reconfigurable optical networks," *Opt. Express*, vol. 17, no. 26, pp. 24020–24029, 2009.

[3] P. Dong, S. Liao, H. Liang, R. Shafiiha, D. Feng, G. Li, X. Zheng, A. V. Krishnamoorthy, and M. Asghari, "Submilliwatt, ultrafast and broadband electro-optic silicon switches," *Opt. Express*, vol. 18, no. 24, pp. 25225–25231, 2010.

[4] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, pp. 114–117, Apr. 1965.

[5] C. A. Mack, "Fifty Years of Moore's Law," *IEEE Trans. Semicond. Manufact.*, vol. 24, pp. 202–207, May 2011.

[6] D. Miller, "Device Requirements for Optical Interconnects to Silicon Chips," *Proc. IEEE*, vol. 97, no. 7, pp. 1166–1185, 2009.

[7] A. V. Krishnamoorthy, R. Ho, X. Zheng, H. Schwetman, J. Lexau, P. Koka, G. Li, I. Shubin, and J. E. Cunningham, "Computer Systems Based on Silicon Photonic Interconnects," *Proc. IEEE*, vol. 97, pp. 1337–1361, July 2009.

[8] "Corning-Intel White Paper: Corning ClearCurve LX Fiber and the Corning MXC Connector," 2013.

[9] B. Jalali, "Can silicon change photonics?," *Phys. Stat. Sol. A*, vol. 205, pp. 213–224, Feb. 2008.

[10] M. Asghari and A. V. Krishnamoorthy, "Silicon photonics: Energy-efficient communication," *Nature Photon.*, vol. 5, pp. 268–270, May 2011.

[11] A. Vahdat, H. Liu, X. Zhao, and C. Johnson, "The emerging optical data center," in *Opt. Fiber Comm. Conf.*, p. OTuH2, 2011.

[12] G. Wang, D. G. Andersen, M. Kaminsky, K. Papagiannaki, T. S. Ng, M. Kozuch, and M. Ryan, "c-Through: Part-time optics in data centers," *ACM SIGCOMM Computer Comm. Rev.*, vol. 40, no. 4, pp. 327–338, 2010.

[13] N. Farrington, *Optics in data center network architecture*. PhD thesis, University of California, San Diego, 2012.

[14] L. Barroso, J. Clidaras, and U. Holzle, *The Datacenter as a Computer*. Morgan & Claypool, 2 ed., 2013.

[15] N. Farrington and A. Andreyev, "Facebook's data center network architecture," in *IEEE Opt. Interconnects Conf.*, p. TuB5, 2013.

[16] A. V. Krishnamoorthy, X. Zheng, G. Li, J. Yao, T. Pinguet, A. Mekis, H. Thacker, I. Shubin, Y. Luo, K. Raj, and J. E. Cunningham, "Exploiting CMOS Manufacturing to Reduce Tuning Requirements for Resonant Optical Devices," *IEEE Photon. J.*, vol. 3, pp. 567–579, June 2011.

[17] ITU-T Recommendation G.694.1, *Spectral grids for WDM applications: DWDM frequency grid*. International Telecommunication Union, 2 ed., Feb. 2012.

[18] ITU-T Recommendation G.694.2, *Spectral grids for WDM applications: CWDM wavelength grid*. International Telecommunication Union, Dec. 2003.

[19] ITU-T Supplement G-39, *Optical system design and engineering considerations*. International Telecommunication Union, Sept. 2012.

[20] Q. Lin, O. J. Painter, and G. P. Agrawal, "Nonlinear optical phenomena in silicon waveguides: Modeling and applications," *Opt. Express*, vol. 15, no. 25, pp. 16604–16644, 2007.

[21] I. D. Rukhlenko, M. Premaratne, and G. P. Agrawal, "Nonlinear Silicon Photonics: Analytical Tools," *IEEE J. Sel. Topics Quantum Electron.*, vol. 16, pp. 200–215, Jan. 2010.

[22] R. M. Osgood Jr, N. C. Panoiu, J. I. Dadap, X. Liu, X. Chen, I.-W. Hsieh, E. Dulkeith, W. M. Green, and Y. A. Vlasov, "Engineering nonlinearities in nanoscale optical systems: physics and applications in dispersion-engineered silicon nanophotonic wires," *Adv. Opt. Photon.*, vol. 1, no. 1, p. 162, 2009.

[23] J. Park, *Nonlinear silicon photonics from the near to mid infrared*. PhD thesis, University of California, San Diego, 2010.

[24] S. Zlatanovic, J. S. Park, S. Moro, J. M. C. Boggio, I. B. Divliansky, N. Alic, S. Mookherjea, and S. Radic, "Mid-infrared wavelength conversion in silicon waveguides using ultracompact telecom-band-derived pump source," *Nature Photon.*, vol. 4, pp. 561–564, Aug. 2010.

[25] J. R. Ong, M. L. Cooper, G. Gupta, W. M. J. Green, S. Assefa, F. Xia, and S. Mookherjea, "Low-power continuous-wave four-wave mixing in silicon coupled-resonator optical waveguides," *Opt. Lett.*, vol. 36, no. 15, pp. 2964–2966, 2011.

[26] J. Leuthold, C. Koos, and W. Freude, "Nonlinear silicon photonics," *Nature Photon.*, vol. 4, no. 8, pp. 535–544, 2010.

[27] J. R. Ong, R. Kumar, R. Aguinaldo, and S. Mookherjea, "Efficient CW Four-Wave Mixing in Silicon-on-Insulator Micro-Rings With Active Carrier Removal," *IEEE Photon. Technol. Lett.*, vol. 25, pp. 1699–1702, Sept. 2013.

[28] N. K. Hon, R. Soref, and B. Jalali, "The third-order nonlinear optical coefficients of Si, Ge, and $Si_{1-x}Ge_x$ in the midwave and longwave infrared," *J. Appl. Phys.*, vol. 110, no. 1, p. 011301, 2011.

[29] R. S. Jacobsen, K. N. Andersen, P. I. Borel, J. Fage-Pedersen, L. H. Frandsen, O. Hansen, M. Kristensen, A. V. Lavrinenko, G. Moulin, H. Ou, C. Peucheret, B. Zsigri, and A. Bjarklev, "Strained silicon as a new electro-optic material," *Nature*, vol. 441, pp. 199–202, May 2006.

[30] B. Chmielak, M. Waldow, C. Matheisen, C. Ripperda, J. Bolten, T. Wahlbrink, M. Nagel, F. Merget, and H. Kurz, "Pockels effect based fully integrated, strained silicon electro-optic modulator," *Opt. Express*, vol. 19, no. 18, pp. 17212–17219, 2011.

[31] M. W. Puckett, J. S. T. Smalley, M. Abashin, A. Grieco, and Y. Fainman, "Tensor of the second-order nonlinear susceptibility in asymmetrically strained silicon waveguides: analysis and experimental validation," *Opt. Lett.*, vol. 39, no. 6, pp. 1693–1696, 2014.

[32] R. Soref and B. Bennett, "Electrooptical effects in silicon," *IEEE J. Quantum Electron.*, vol. 23, no. 1, pp. 123–129, 1987.

[33] R. Soref and B. Bennett, "Kramers-Kronig Analysis Of Electro-Optical Switching In Silicon," *Proc. SPIE*, vol. 704, pp. 32–37, Mar. 1987.

[34] H. S. Rong, A. S. Liu, R. Jones, O. Cohen, D. Hak, R. Nicolaescu, A. Fang, and M. Paniccia, "An all-silicon Raman laser," *Nature*, vol. 433, pp. 292–294, Jan. 2005.

[35] M. Nedeljkovic, R. Soref, and G. Z. Mashanovich, "Free-Carrier Electrorefraction and Electroabsorption Modulation Predictions for Silicon Over the 1-14-m Infrared Wavelength Range," *IEEE Photon. J.*, vol. 3, pp. 1171–1180, Dec. 2011.

[36] S. Sze and K. Ng, *Physics of Semiconductor Devices*. Hoboken, NJ: John Wiley & Sons, 3 ed., 2007.

[37] H. Rong, S. Xu, Y.-H. Kuo, V. Sih, O. Cohen, O. Raday, and M. Paniccia, "Low-threshold continuous-wave Raman silicon laser," *Nature Photon.*, vol. 1, pp. 232–237, Apr. 2007.

[38] R. Aguinaldo, "Modeling solutions and simulations for advanced III-V photovoltaics based on nanostructures," Master's thesis, Rochester Institute of Technology, 2008.

[39] D. Dimitropoulos, R. Jhaveri, R. Claps, J. C. S. Woo, and B. Jalali, "Lifetime of photogenerated carriers in silicon-on-insulator rib waveguides," *Appl. Phys. Lett.*, vol. 86, no. 7, p. 071115, 2005.

[40] D. Dimitropoulos, S. Fathpour, and B. Jalali, "Limitations of active carrier removal in silicon Raman amplifiers and lasers," *Appl. Phys. Lett.*, vol. 87, no. 26, p. 261108, 2005.

[41] W. Shockley and W. T. Read, "Statistics of the Recombination of Holes and Electrons," *Phys. Rev.*, vol. 87, no. 5, pp. 835–842, 1952.

[42] W. Shockley, "Electrons, holes, and traps," *Proc. IRE*, vol. 46, no. 6, pp. 973–990, 1958.

[43] O. Palais and A. Arcari, "Contactless measurement of bulk lifetime and surface recombination velocity in silicon wafers," *J. Appl. Phys.*, vol. 93, no. 8, p. 4686, 2003.

[44] T. Kuwayama, M. Ichimura, and E. Arai, "Interface recombination velocity of silicon-on-insulator wafers measured by microwave reflectance photoconductivity decay method with electric field," *Appl. Phys. Lett.*, vol. 83, no. 5, p. 928, 2003.

[45] R. Aguinaldo, Y. Shen, and S. Mookherjea, "Large Dispersion of Silicon Waveguide Directional Couplers," *Conf. Lasers and Electro-Optics*, p. CM4A.2, 2012.

[46] B. Little, S. Chu, H. Haus, J. Foresi, and J. Laine, "Microring resonator channel dropping filters," *J. Lightw. Technol.*, vol. 15, no. 6, pp. 998–1005, 1997.

[47] A. Yariv, "Universal relations for coupling of optical power between microresonators and dielectric waveguides," *Electron. Lett.*, vol. 36, no. 4, pp. 321–322, 2000.

[48] W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, "Silicon microring resonators," *Laser Photon. Rev.*, vol. 6, pp. 47–73, Sept. 2011.

[49] S. Mookherjea and M. A. Schneider, "The nonlinear microring add-drop filter," *Opt. Express*, vol. 16, no. 19, pp. 15130–15136, 2008.

[50] C. Madsen and J. Zhao, "A general planar waveguide autoregressive optical filter," *J. Lightw. Technol.*, vol. 14, no. 3, pp. 437–447, 1996.

[51] B. E. Little, S. T. Chu, P. P. Absil, J. V. Hryniewicz, F. G. Johnson, F. Seiferth, D. Gill, V. Van, O. King, and M. Trakalo, "Very High-Order Microring Resonator Filters for WDM Applications," *IEEE Photon. Technol. Lett.*, vol. 16, pp. 2263–2265, Oct. 2004.

[52] H.-C. Liu and A. Yariv, "Synthesis of high-order bandpass filters based on coupled-resonator optical waveguides (CROWs)," *Opt. Express*, vol. 19, pp. 17653–17668, Aug. 2011.

[53] B. Little, "A VLSI photonics platform," in *Optical Fiber Comm. Conf.*, pp. 444–445, 2003.

[54] C. Madsen, "Efficient architectures for exactly realizing optical filters with optimum bandpass designs," *IEEE Photon. Technol. Lett.*, vol. 10, no. 8, pp. 1136–1138, 1998.

[55] Y. Wang, R. Aguinaldo, and T. Nguyen, "Realistic Photonic Filter Design Based on Allpass Substructures With Waveguide Loss Compensation," *J. Lightw. Technol.*, vol. 32, pp. 1024–1031, Mar. 2014.

[56] G. Paloczi, Y. Huang, A. Yariv, and S. Mookherjea, "Polymeric Mach-Zehnder interferometer using serially coupled microring resonators," *Opt. Express*, vol. 11, no. 21, pp. 2666–2671, 2003.

[57] A. Vörckel, M. Mönster, W. Henschel, P. H. Bolivar, and H. Kurz, "Asymmetrically coupled silicon-on-insulator microring resonators for compact add-drop multiplexers," *IEEE Photon. Technol. Lett.*, vol. 15, no. 7, pp. 921–923, 2003.

[58] S.-Y. Cho and R. Soref, "Apodized SCISSORs for filtering and switching," *Opt. Express*, vol. 16, no. 23, pp. 19078–19090, 2008.

[59] M. Mancinelli, R. Guider, M. Masi, P. Bettotti, M. R. Vanacharla, J.-M. Fedeli, and L. Pavesi, "Optical characterization of a SCISSOR device," *Opt. Express*, vol. 19, no. 14, pp. 13664–13674, 2011.

[60] M. R. Watts, "Adiabatic microring resonators," *Opt. Lett.*, vol. 35, pp. 3231–3233, Oct. 2010.

[61] E. A. Marcatili, "Dielectric rectangular waveguide and directional coupler for integrated optics," *Bell Syst. Tech. J.*, vol. 48, no. 7, pp. 2071–2102, 1969.

[62] A. Yariv, "Coupled-Mode Theory for Guided-Wave Optics," *IEEE J. Quantum Electron.*, vol. QE 9, no. 9, pp. 919–933, 1973.

[63] A. Hardy and W. Streifer, "Coupled mode theory of parallel waveguides," *J. Lightw. Technol.*, vol. 3, no. 5, pp. 1135–1146, 1985.

[64] E. Marcatili, "Improved Coupled-Mode Equations for Dielectric Guides," *IEEE J. Quantum Electron.*, vol. 22, no. 6, pp. 988–993, 1986.

[65] H. Haus, W. Huang, S. Kawakami, and N. Whitaker, "Coupled-mode theory of optical waveguides," *J. Lightw. Technol.*, vol. 5, no. 1, pp. 16–23, 1987.

[66] H. Haus, "Coupled-Mode Theory Revisited," *Proc. SPIE*, vol. 704, pp. 68–78, Mar. 1987.

[67] Haus and W. Huang, "Coupled-mode theory," *Proc. IEEE*, vol. 79, no. 10, pp. 1505–1518, 1991.

[68] W. P. Yuen, "On the Different Formulations of the Coupled-Mode Theory for Parallel Dielectric Wave-Guides," *J. Lightw. Technol.*, vol. 12, no. 1, pp. 82–85, 1994.

[69] W. Huang, "Coupled-mode theory for optical waveguides: an overview," *J. Opt. Soc. Am. A*, vol. 11, no. 3, pp. 963–983, 1994.

[70] B. Little and W. Huang, "Coupled-mode theory for optical waveguides," *Prog. Electromag. Res.*, vol. 10, pp. 217–270, 1995.

[71] W.-P. Huang and J. Mu, "Complex coupled-mode theory for optical waveguides," *Opt. Express*, vol. 17, no. 21, pp. 19134–19152, 2009.

[72] M. L. Cooper and S. Mookherjea, "Numerically-assisted coupled-mode theory for silicon waveguide couplers and arrayed waveguides," *Opt. Express*, vol. 17, no. 3, pp. 1583–1599, 2009.

[73] A. Yariv and P. Yeh, *Photonics: Optical Electronics in Modern Communications*. New York: Oxford University Press, 6 ed., 2007.

[74] L. Coldren, S. Corzine, and M. Masanovic, *Diode Lasers and Photonic Integrated Circuits*. Hoboken, NJ: John Wiley & Sons, 2 ed., 2012.

[75] M. Cooper, *Modeling and Characterization of Strongly Coupled Silicon-on-insulator Nanophotonic Devices*. PhD thesis, University of California, San Diego, 2010.

[76] I. Newton, *Opticks: or, a Treatise of the Reflections, Refractions, Inflections and Colours of Light*. London: William Innys, West-End, St. Paul's, 4 ed., 1730.

[77] Y. P. Li and C. H. Henry, "Silica-based optical integrated circuits," *IEE Proc. Optoelectron.*, vol. 143, pp. 263–280, Oct. 1996.

[78] S. Mookherjea, J. R. Ong, and R. Aguinaldo, "High performance silicon multi-ring filters," in *IEEE Avionics, Fiber-Optics, and Photonics Technology Conference (AVFOP)*, (Cocoa Beach, FL), p. WB5, Sept. 2012.

[79] R. Aguinaldo, P. Weigel, H. R. Grant, C. T. DeRose, A. Lentine, A. Pomerene, A. Starbuck, A. Tkacenko, and S. Mookherjea, "A silicon photonic channelized spectrum monitor for UCSD's multi-wavelength ring network," in *Conf. Lasers and Electro-Optics*, p. STu1 G.4, 2014.

[80] S. Xiao, M. H. Khan, H. Shen, and M. Qi, "Modeling and measurement of losses in silicon-on-insulator resonators and bends," *Opt. Express*, vol. 15, no. 17, pp. 10553–10561, 2007.

[81] R. Sun, P. Dong, N.-n. Feng, C.-y. Hong, J. Michel, M. Lipson, and L. Kimerling, "Horizontal single and multiple slot waveguides: optical transmission at λ=1550 nm," *Opt. Express*, vol. 15, no. 26, pp. 17967–17972, 2007.

[82] J. Palaci, G. Villanueva, J. Galán, J. Marti, and B. Vidal, "Single Bandpass Photonic Microwave Filter Based on a Notch Ring Resonator," *IEEE Photon. Technol. Lett.*, vol. 22, no. 17, pp. 1276–1278, 2010.

[83] J. Hu, X. Sun, A. Agarwal, and L. C. Kimerling, "Design guidelines for optical resonator biochemical sensors," *J. Opt. Soc. Am. B*, vol. 26, no. 5, pp. 1032–1041, 2009.

[84] A. M. Prabhu, H. L. Liew, and V. Van, "Experimental determination of coupled-microring filter parameters via pole-zero extraction," *Opt. Express*, vol. 16, no. 19, pp. 14588–14596, 2008.

[85] R. Adar, Y. Shani, C. Henry, R. Kistler, G. Blonder, and N. Olsson, "Measurement of very low-loss silica on silicon waveguides with a ring resonator," *Appl. Phys. Lett.*, vol. 58, no. 5, pp. 444–445, 1991.

[86] R. Adar, M. Serbin, and V. Mizrahi, "Less than 1 dB per meter propagation loss of silica waveguides measured using a ring resonator," *J. Lightw. Technol.*, vol. 12, no. 8, pp. 1369–1372, 1994.

[87] F. Xia, L. Sekaric, and Y. Vlasov, "Mode conversion losses in silicon-on-insulator photonic wire based racetrack resonators," *Opt. Express*, vol. 14, no. 9, pp. 3872–3886, 2006.

[88] R. Aguinaldo, Y. Shen, and S. Mookherjea, "Large Dispersion of Silicon Directional Couplers Obtained via Wideband Microring Parametric Characterization," *IEEE Photon. Technol. Lett.*, vol. 24, pp. 1242–1244, July 2012.

[89] D. C. Kilper, R. Bach, D. J. Blumenthal, D. Einstein, T. Landolsi, L. Ostar, M. Preiss, and A. E. Willner, "Optical Performance Monitoring," *J. Lightw. Technol.*, vol. 22, pp. 294–304, Jan. 2004.

[90] N. Farrington, A. Forencich, G. Porter, P. C. Sun, J. E. Ford, Y. Fainman, G. C. Papen, and A. Vahdat, "A Multiport Microsecond Optical Circuit Switch for Data Center Networking," *IEEE Photon. Technol. Lett.*, vol. 25, pp. 1589–1592, Aug. 2013.

[91] W. A. Zortman, D. C. Trotter, and M. R. Watts, "Silicon photonics manufacturing," *Opt. Express*, vol. 18, no. 23, pp. 23598–23607, 2010.

[92] M. Hochberg and T. Baehr-Jones, "Towards fabless silicon photonics," *Nature Photon.*, vol. 4, pp. 492–494, Aug. 2010.

[93] T. Baehr-Jones, T. Pinguet, and P. Guo-Qiang, "Myths and rumours of silicon photonics," *Nature Photon.*, 2012.

[94] E. H. Moore, "On the reciprocal of the general algebraic matrix," *Bull. Amer. Math. Soc.*, vol. 20, pp. 394–395, 1920.

[95] R. Penrose, "A generalized inverse for matrices," *Proc. Camb. Philos. Soc.*, vol. 51, no. 3, pp. 406–413, 1955.

[96] J. G. Francis, "The QR transformation: a unitary analogue to the LR transformation—Part 1," *The Comp. J.*, vol. 4, no. 3, pp. 265–271, 1961.

[97] J. G. Francis, "The QR transformation—part 2," *The Comp. J.*, vol. 4, no. 4, pp. 332–345, 1962.

[98] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge University Press, 1985.

[99] R. Aguinaldo, P. Weigel, H. R. Grant, C. T. DeRose, A. Lentine, A. Pomerene, A. Starbuck, and S. Mookherjea, "Characterization of a silicon-photonic multi-wavelength power monitor," in *IEEE Opt. Interconnects Conf.*, p. WD5, 2014.

[100] N. Farrington, G. Porter, P.-C. Sun, A. Forencich, J. Ford, Y. Fainman, G. Papen, and A. Vahdat, "A demonstration of ultra-low-latency data center optical circuit switching," *ACM SIGCOMM Computer Comm. Rev.*, vol. 42, no. 4, pp. 95–96, 2012.

[101] Y.-O. Noh, H.-J. Lee, Y.-H. Won, and M.-C. Oh, "Polymer waveguide thermo-optic switches with -70dB optical crosstalk," *Optics Comm.*, vol. 258, pp. 18–22, Feb. 2006.

[102] B. G. Lee, A. Biberman, P. Dong, M. Lipson, and K. Bergman, "All-Optical Comb Switch for Multiwavelength Message Routing in Silicon Photonic Networks," *IEEE Photon. Technol. Lett.*, vol. 20, pp. 767–769, May 2008.

[103] M. R. Watts, W. A. Zortman, D. C. Trotter, G. N. Nielson, D. L. Luck, and R. W. Young, "Adiabatic resonant microrings (ARMs) with directly integrated thermal microphotonics," in *Conf. Lasers Electro-Optics*, p. CPDB10, Optical Society of America, 2009.

[104] Y. Shoji, K. Kintaka, S. Suda, H. Kawashima, T. Hasama, and H. Ishikawa, "Low-crosstalk 2 x 2 thermo-optic switch with silicon wire waveguides," *Opt. Express*, vol. 18, no. 9, pp. 9071–9075, 2010.

[105] C. T. DeRose, M. Watts, R. W. Young, D. C. Trotter, G. N. Nielson, W. A. Zortman, and R. D. Kekatpure, "Low power and broadband 2 x 2 silicon thermo-optic switch," in *Opt. Fiber Comm. Conf.*, p. OThM3, Optical Society of America, 2011.

[106] G. Coppola, "Advance in thermo-optical switches: principles, materials, design, and device structure," *Opt. Eng.*, vol. 50, p. 071112, July 2011.

[107] M. W. Geis, S. J. Spector, R. C. Williamson, and T. M. Lyszczarz, "Submicrosecond Submilliwatt Silicon-on-Insulator Thermooptic Switch," *IEEE Photon. Technol. Lett.*, vol. 16, pp. 2514–2516, Nov. 2004.

[108] M. R. Watts, J. Sun, C. T. DeRose, D. C. Trotter, R. W. Young, and G. N. Nielson, "Adiabatic thermo-optic Mach-Zehnder switch," *Opt. Lett.*, vol. 38, pp. 733–735, Mar. 2013.

[109] R. L. Espinola, M. C. Tsai, J. T. Yardley, and R. M. Osgood, "Fast and low-power thermooptic switch on thin silicon-on-insulator," *IEEE Photon. Technol. Lett.*, vol. 15, pp. 1366–1368, Oct. 2003.

[110] K. Levenberg, "A method for the solution of certain problems in least squares," *Quart. Appl. Math.*, vol. 2, pp. 164–168, 1944.

[111] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Indus. Appl. Math.*, vol. 11, no. 2, pp. 431–441, 1963.

[112] R. Aguinaldo, A. Forencich, C. T. DeRose, A. Lentine, A. Starbuck, Y. Fainman, G. Porter, G. Papen, and S. Mookherjea, "Characterization of a silicon-photonic wideband switch in UCSD's MORDIA ring network," in *IEEE Opt. Interconnects Conf.*, p. TuD4, 2014.

[113] J. Van Campenhout, W. M. J. Green, and Y. A. Vlasov, "Design of a digital, ultra-broadband electro-optic switch for reconfigurable optical networks-on-chip," *Opt. Express*, vol. 17, no. 26, pp. 23793–23808, 2009.

[114] R. Aguinaldo, G. Porter, G. Papen, Y. Fainman, and S. Mookherjea, "Characterization of electrically-driven silicon photonic Mach-Zehnder switches," in *Conf. Lasers and Electro-Optics*, p. STh4M.5, Feb. 2014.

[115] P. Ganguly, J. C. Biswas, S. Das, and S. K. Lahiri, "A three-waveguide polarization independent power splitter on lithium niobate substrate," *Optics Comm.*, vol. 168, no. 5-6, pp. 349–354, 1999.

[116] G. Calo, A. D'Orazio, and V. Petruzzelli, "Broadband Mach-Zehnder Switch for Photonic Networks on Chip," *J. Lightw. Technol.*, vol. 30, pp. 944–952, Apr. 2012.

[117] W. A. Zortman, A. Lentine, D. Trotter, and M. Watts, "Integrated CMOS Comaptible Low Power 10Gbps Silicon Photonic Heater Modulator," in *Opt. Fiber Comm. Conf.*, p. OW4I, Optical Society of America, 2012.

[118] M. Harjanne, M. Kapulainen, T. Aalto, and P. Heimala, "Sub-μs Switching Time in Silicon-on-Insulator Mach-Zehnder Thermooptic Switch," *IEEE Photon. Technol. Lett.*, vol. 16, pp. 2039–2041, Sept. 2004.

[119] Y. Li, J. Yu, S. Chen, Y. Li, and Y. Chen, "Submicrosecond rearrangeable nonblocking silicon-on-insulator thermo-optic 4×4 switch matrix," *Opt. Lett.*, vol. 32, no. 6, pp. 603–604, 2007.

[120] R. Aguinaldo, A. Forencich, C. T. DeRose, A. Lentine, D. C. Trotter, Y. Fainman, G. Porter, G. Papen, and S. Mookherjea, "Wideband silicon-photonic thermo-optic switch in a wavelength-division multiplexed ring network," *Opt. Express*, vol. 22, no. 7, pp. 8205–8218, 2014.

[121] G. Porter, R. Strong, N. Farrington, A. Forencich, P. Chen-Sun, T. Rosing, Y. Fainman, G. Papen, and A. Vahdat, "Integrating Microsecond Circuit Switching into the Data Center," *ACM SIGCOMM Computer Comm. Rev.*, vol. 43, pp. 447–458, Oct. 2013.

[122] Y. Fainman and G. Porter, "Directing Data Center Traffic," *Science*, vol. 342, pp. 202–203, 2013.

[123] M. Glick, "Optical Interconnects in Next Generation Data Centers: An End to End View," in *Optical Interconnects for Future Data Center Networks* (C. Kachris, K. Bergman, and I. Tomkos, eds.), New York: Springer, 2013.

[124] R. Soref, "The Past, Present, and Future of Silicon Photonics," *IEEE J. Sel. Topics Quantum Electron.*, vol. 12, pp. 1678–1687, Nov. 2006.

[125] R. Aguinaldo, A. Forencich, C. T. DeRose, A. Lentine, D. C. Trotter, A. Starbuck, Y. Fainman, G. Porter, G. Papen, and S. Mookherjea, "Energy-efficient, digitally-driven "fat pipe" silicon photonic circuit switch in the UCSD MORDIA data-center network," in *Conf. Lasers and Electro-Optics*, p. STu1 J.2, 2014.

[126] F. E. Doany, B. G. Lee, S. Assefa, W. M. J. Green, M. Yang, C. L. Schow, C. V. Jahnes, S. Zhang, J. Singer, V. I. Kopp, J. A. Kash, and Y. A. Vlasov, "Multichannel High-Bandwidth Coupling of Ultradense Silicon Photonic Waveguide Array to Standard-Pitch Fiber Array," *J. Lightw. Technol.*, vol. 29, pp. 475–482, Feb. 2011.

[127] C. Sah, R. Noyce, and W. Shockley, "Carrier generation and recombination in pn junctions and pn junction characteristics," *Proc. IRE*, vol. 45, no. 9, pp. 1228–1243, 1957.

[128] S. Mookherjea, J. R. Ong, X. Luo, and G. Q. Lo, "Electronic control of optical Anderson localization modes," *Nature Nanotech.*, pp. 1–7, Mar. 2014.