Spring 4-1-2015

Devices and Systems-on-Chip for Photonic Communication Links in a Microprocessor

Mark T. Wade
University of Colorado Boulder, mtw1088@gmail.com

Follow this and additional works at: https://scholar.colorado.edu/ecn_gradetds

Part of the Semiconductor and Optical Materials Commons

Recommended Citation
https://scholar.colorado.edu/ecn_gradetds/139

This Dissertation is brought to you for free and open access by Electrical, Computer & Energy Engineering at CU Scholar. It has been accepted for inclusion in Electrical, Computer & Energy Engineering Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Devices and systems-on-chip for photonic communication links in a microprocessor

by

Mark T. Wade

M.S., University of Colorado, 2013
B.S. (Physics), Louisiana Tech University, 2010
B.S. (Electrical Engineering), Louisiana Tech University, 2010

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Electrical, Computer, and Energy Engineering

2015
This thesis entitled:
Devices and systems-on-chip for photonic communication links in a microprocessor
written by Mark T. Wade
has been approved for the Department of Electrical, Computer, and Energy Engineering

Prof. Miloš Popović
Assistant Professor of Electrical Engineering
University of Colorado
Thesis Advisor

Prof. Vladimir Stojanović
Associate Professor of Electrical Engineering and Computer Science
University of California, Berkeley

Prof. Anatol Khilo
Assistant Professor of Microsystems Engineering
Masdar Institute, Abu Dhabi

Prof. Kelvin Wagner
Professor of Electrical Engineering
University of Colorado
Thesis Committee Chairman

Date __________________

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.
For the first time, high-performance photonic devices and electronic-photonic systems-on-chip are monolithically integrated in an advanced CMOS microelectronics fabrication process. This includes a silicon optical resonator termed the “spoked-ring” cavity that meets the constraints of thin-SOI microelectronics CMOS processes and enables energy efficient modulators and thermally tunable filters. For low-loss fiber-to-chip optical coupling, a phased-array antenna concept is demonstrated, and the 45 nm CMOS microelectronics process is shown to support a near ideal implementation of the device using the crystalline silicon and polysilicon material layers that comprise the active region and gate, respectively, of the native MOSFET transistors.

The active devices and vertical grating couplers are implemented in large-scale electronic-photonic systems-on-chip to demonstrate a wavelength stabilized, microring-based chip-to-chip communications link and an 11-channel wavelength division multiplexed (WDM) transmitter. The link is shown to be robust against thermal environmental variations which is critical for operation in realistic systems. The chip-to-chip link is then used to demonstrate a CPU-to-memory communication link, the first demonstration of its kind. The first microprocessor with photonic I/O is demonstrated as part of this work, with substantial implications for computer architecture.

Advanced photonic device technology demonstrations, including photonic crystals, a quantum-correlated photon-pair source, an active photonic device platform in a 32 nm SOI node, and a 180 nm bulk silicon process, are presented to show the wide range of applications that monolithic integration could support in the future of photonics.

These results taken together show that monolithic integration directly into CMOS microelectronics processes does allow high performance photonics, and is a viable approach to build large-scale electronic-photonic systems with a realistic path to commercial technologies.
Dedication

To my wife, Katie

to my parents, brother, and sister
Acknowledgements

Many people have supported and encouraged me in my pursuit of science and engineering, and I fear that this acknowledgement is sure to leave out some. Much of the work presented in this thesis is the result of collaboration and team efforts by many talented people. In particular, Jeff Shainline mentored me in my early years and has been a close friend ever since. Jason Orcutt provided priceless guidance and patience when I first started working on the POEM project, and his ability to manage large amounts of work and stress left a lasting impression on me. Mike Georgas always provided a voice of reason in trying times, and I valued his mentorship and friendship. Working with Chen Sun has been a privilege, and we spent countless nights designing chips that led to many of the results in this thesis. Dr. Davis Harbour once told me that I would go on to do big things. Although the veracity of that statement is yet to be seen, it has served as continued motivation. I thank Alex Bohl, Chris Hertlein, and Michael Willis for their many years of friendship and encouragement. I am in gratitude to my cousin, Aaron Morris, for giving me a place to live in Boston while I worked at MIT and the countless hours of fun. It was great to have family so far from home.

My fellow group members at CU Boulder made my Ph.D. enjoyable with many lifelong memories – Cale Gentry, Fabio Pavanello, Rajesh Kumar, Xiaoge “Savvy” Zeng, Yangyang Liu, Nathan Dostart, Imbert Wang, Gil Triginer, and Josep Fargas – I could not ask for a better set of colleagues and friends. My advisor, Miloš Popović, taught me how to think about problems with his tireless search for understanding and his “first principles” approach to design.

Several people aided in the completion of this thesis with their thoughtful feedback and helpful discussions. In particular, I thank committee members Prof. Anatoly Khilo and Prof. Vladimir Stojanović. I would like to thank Prof. Rajeev Ram for hosting me at MIT as a visiting student. My time spent at MIT was very productive and led to many exciting opportunities.

I owe the deepest thanks to my family. My parents, Philip and Carol Wade, have always provided a loving, stable home that has served as the backbone of my various life pursuits. I see their influence in myself more and more as I continue to grow. My wife, Katie, has made my life infinitely better, and I look forward to our continued adventures together.

Mark Wade
Boulder, Colorado
November 1, 2015
## Contents

### Chapter

1. **Introduction**
   - 1.1 Evolution of silicon photonics ................................................. 2
   - 1.2 Chip-to-chip I/O bandwidth limitations ....................................... 3
   - 1.3 “Zero-change” CMOS integration ................................................... 4
   - 1.4 Outline of thesis ........................................................................ 7

2. **Active microring-based devices in thin-SOI CMOS** ................................. 10
   - 2.1 Introduction ............................................................................... 10
   - 2.2 Basic properties of microring-based active devices ......................... 11
     - 2.2.1 Resonance shift given an index perturbation .............................. 11
     - 2.2.2 Relation between material absorption and resonator quality factor ................................................................. 13
     - 2.2.3 Perturbation of silicon refractive index due to free carriers .......... 15
   - 2.3 Basic operation of depletion-mode microring modulators .................. 18
   - 2.4 Designing waveguides and ring resonators in thin-SOI CMOS ............ 21
   - 2.5 Modulator design: fitting the constraints of the 45 nm CMOS process .... 27
     - 2.5.1 Resonance shift of interleaved junction ..................................... 30
     - 2.5.2 Design flow of spoked-ring cavity .......................................... 32
     - 2.5.3 CAD flow integration ............................................................... 36
     - 2.5.4 Photonics CAD in CMOS design flows ..................................... 38
2.6 Experimental demonstrations .................................................. 40
  2.6.1 First depletion-mode modulators in 45 nm CMOS ....................... 40
  2.6.2 Efficient thermally tunable filter ........................................ 45
2.7 Conclusions ............................................................................. 51

3 Vertical grating couplers .............................................................. 53
  3.1 Introduction and background ..................................................... 53
  3.2 Performance and design trade-offs in vertical grating couplers ........ 57
    3.2.1 General expressions for grating-to-fiber coupling efficiency .......... 59
    3.2.2 Calculating the bandwidth of the grating-to-fiber coupling coefficient . 61
    3.2.3 Optimum overlap for uniform gratings .................................. 63
    3.2.4 Performance and design trade-offs for Gaussian gratings .......... 65
    3.2.5 Summary of equations for key grating coupler properties .......... 72
  3.3 Vertical grating couplers in zero-change CMOS ......................... 75
    3.3.1 Using transistor layers to build vertical grating couplers .......... 75
    3.3.2 Single layer grating (SLG) demonstration ............................. 77
  3.4 Breaking the vertical symmetry: antenna-array inspired designs .......... 79
    3.4.1 Dual layer grating (DLG) first generation design ................. 81
    3.4.2 Second generation DLG design .......................................... 83
    3.4.3 Second generation DLG experimental characterization .......... 84
  3.5 Manufacturability and fabrication sensitivities ........................... 86
  3.6 Conclusions ............................................................................. 88

4 System demonstrations ............................................................... 91
  4.1 Introduction ........................................................................... 91
    4.1.1 Chip-to-chip photonic wavelength-division multiplexed communication links . 91
    4.1.2 Design considerations for monolithically integrated systems-on-chip .... 96
    4.1.3 WDM transceivers ......................................................... 104
4.1.4 Photonic layout integration with on-chip electronics ........................................ 106
4.2 Experimental demonstrations ............................................................................. 107
  4.2.1 A monolithic chip-to-chip link ...................................................................... 108
  4.2.2 The first optically connected CPU chip ....................................................... 121
4.3 Conclusions ......................................................................................................... 128

5 Advanced microring devices for on-chip communication systems ......................... 130
  5.1 Pole-zero filters for WDM multiplexers/demultiplexers ....................................... 131
    5.1.1 Introduction ............................................................................................... 131
    5.1.2 Coupling of modes in time model ............................................................. 133
    5.1.3 Design of a serial demultiplexer based on asymmetric response stages .... 143
    5.1.4 Alternative topologies for a filter response with two poles and one zero .... 145
    5.1.5 Experimental results ................................................................................... 148
    5.1.6 Conclusion .................................................................................................. 149
  5.2 Wavelength conversion in coupled microring modulators and their design via an equivalent linear filter representation ..................................................... 150

6 Outlook on monolithic integration ........................................................................... 161
  6.1 Extended photonic device platform demonstrations in 45 nm CMOS ................. 162
    6.1.1 Photonic crystals in 45 nm CMOS ............................................................. 162
  6.2 Nonlinear four-wave mixing and correlated single photon pair source .................. 165
  6.3 Scaling zero-change photonics: an active photonics platform in a 32 nm microelectronics SOI CMOS process ................................................................. 168
  6.4 Photonics in bulk silicon CMOS processes ....................................................... 170
    6.4.1 A depletion-mode microring modulator in a bulk CMOS process ............... 172
    6.4.2 The first monolithic chip-to-chip link in a bulk silicon process .................... 174
  6.5 Conclusions ......................................................................................................... 177
7 Conclusions

7.1 Summary of major accomplishments ........................................... 178

7.2 Remaining challenges and future directions for research ................. 180

Bibliography ................................................................................. 183
Tables

Table

2.1 Equations for quality factor, $Q$, and 3-dB linewidth, $\delta f_{3dB}$ of a resonator given common loss variables. $\omega_0$ is the angular frequency of light, $\gamma_l$ is the loss decay rate, $\alpha$ is the field spatial decay, $\kappa$ is the imaginary part of the refractive index, $\omega$ is the complex resonance frequency, and $\gamma$ is the complex angular propagation constant. 

2.2 Values for constants in Eq. 2.16 for $\lambda = 1300$ nm and $\lambda = 1550$ nm.

2.3 Values for polynomial fit coefficients corresponding to ring-bus simulations in Fig. 2.8.

The gap is in units of $\mu$m.

2.4 Values for polynomial fit coefficients corresponding to ring-ring simulations in Fig. 2.8.

The gap is in units of $\mu$m.

3.1 Summary of derived equations.

3.2 Summary of variables.

4.1 Summary of sub-system performance in chip-to-chip link. The table shows results from standalone characterization of each sub-system. The actual chip-to-chip link was run at 5 Gb/s as shown in Fig. 4.19. The receiver sensitivity is reported as measured sensitivity with the on-chip photodiode as well as “normalized sensitivity” which estimates the sensitivity of the receiver circuitry if the photodiode produces 1 A/W. This is commonly done for comparison between receiver circuits independent of the photodiode.
4.2 Summary of transceiver performance in CPU-to-memory communication link and standalone transceivers. The devices in standalone transceivers will be integrated in the next iteration of the system transceivers. Optical Modulation Amplitude (OMA) sensitivity is \(10 \log_{10}(P_1 - P_0)\) where \(P_1\) and \(P_0\) are the optical powers (mW) corresponding to a “1” bit and a “0” bit, respectively.

5.1 Comparison of filter design specifications with measured results
Figures

Figure

1.1 Waveguide integration in a (a) bulk silicon process and (b) silicon-on-insulator process.

1.2 Illustration of the material layer stack cross-section for the 45 nm SOI fabrication process showing the MOSFET transistors (NMOS and PMOS) as well as the photonics implementation.

2.1 Perturbation to refractive index (left) and absorption (right) for $\lambda = 1300$ nm (top) and $\lambda = 1550$ nm (bottom). These plots are generated from fits to data from [123].

2.2 Intrinsic quality factor versus carrier concentration for (a) $\lambda = 1300$ nm and (b) $\lambda = 1550$ nm with assumed $n_g = 3.0$.

2.3 (a) Resonance frequency shift versus electron and hole concentration for $\lambda = 1300$ nm and $\lambda = 1550$ nm. (b) Intrinsic 3 dB linewidth of resonance. (c) Ratio of resonance shift and 3 dB linewidth which is a useful figure of merit.

2.4 (a) Schematic of microring model used for CMT equations. (b) p-n junction configurations for depletion-mode modulators. (c) Cross-sectional view of reverse-bias state showing index and absorption perturbation. The gray regions are depleted of free carriers. (d) Illustration of resonance shift with applied voltage. Carriers blue-shift the resonance. A cw laser frequency is drawn on the graph to show how on-off keying is achieved. (e) Resulting amplitude modulation of optical power at the output of the microring modulator.
2.5 Effective index of waveguide transverse modes at $\lambda = 1180, 1310, 1550\,\text{nm}$. Single-mode widths are indicated by the blue dots, and $E_x$, the x-component of the electric field, contours are shown below, overlapped with the waveguide cross-section. The widths are 380 nm, 460 nm, and 700 nm for $\lambda=1180$, 1310, and 1550 nm, respectively.

2.6 Radiation limited quality factor of fundamental TE-like transverse mode as a function of ring geometry (width and radius) for (a) $\lambda = 1180\,\text{nm}$, (b) $\lambda = 1310\,\text{nm}$, (c) $\lambda = 1550\,\text{nm}$. Our choices for specific geometries are indicated by the red dot, and the contour plots of the $E_x$ component of the corresponding bent modes are plotted in (d)-(f). The selected widths ($w$) and outer radii ($R$) are $w = 460, 560, 700\,\text{nm}$ and $R = 4, 5, 8\,\mu\text{m}$ for $\lambda = 1180, 1310, 1550$, respectively. These dimensions are chosen to achieve $Q > 10^6$, but the exact dimensions of a microring depend on specifications such as the total $Q$ and free spectral range required for the application.

2.7 (a) Illustration of simulation configuration for calculating power coupling coefficient between a ring and a bus waveguide. (b) Snapshot of ring-bus 3D FDTD simulation. (c) Illustration of simulation configuration for calculating power coupling coefficient between two rings. (d) Snapshot of ring-ring 3D FDTD simulation.

2.8 Ring-bus (left column) and ring-ring (right column) power coupling coefficients for $\lambda = 1180$ (top), 1310 (middle), and 1550 (bottom) nm.

2.9 Schematic of the spoked-ring microcavity: (a) passive cavity; (b) formation of interleaved junctions for a modulator, with inner radius vias; (c) extended interaction length input coupler to excite only the fundamental mode. (Illustration: M. Popović)

2.10 Resonance shifts for $\lambda = 1180\,\text{nm}$. (a) Resonance shift with 1 V applied electrical reverse bias, relative to the built-in potential, for interleaved junction. A unit cell periodicity of 500 nm is used as is achievable in the 45 nm process. (b) Ratio of resonance shift to intrinsic linewidth. An intrinsic (passive, undoped) quality factor of 100,000 is assumed, and the group index is $n_g \approx 3.0$. 

2.11 Supermodes of disk/waveguide system. Spatial electric field mode profiles of the (a) symmetric (top) and (b) anti-symmetric (bottom) supermodes. (c) Angular propagation constant anti-crossing between symmetric and anti-symmetric supermodes. This splitting occurs when there is efficient coupling between the disk and the waveguide indicating phase-matching. (Original simulation by Jeff Shainline. Modified for use here.)

2.12 Spoked-ring width versus Q. (a) 3D FDTD simulation showing the fundamental mode avoiding the contacts. (b) Radiation quality factor versus the width of the ring (outer radius fixed at 5\(\mu\)m). As the width decreases, the fundamental mode scatters more light and the Q degrades. Narrowing the ring leads to reduced RC time constants for driving the reverse-biased junctions from the electrical contacts at the inner radius, at the cost of broader linewidth resonances. (Original simulation by Jeff Shainline. Modified for use here.)

2.13 FDTD simulations showing various coupling schemes to the spoked-ring cavity for \(\lambda = 1280\, nm\). (a) Point coupling scheme where the input coupling waveguide is straight and is brought in close proximity to the spoked-ring microcavity. This approach excites higher-order modes and is lossy. (b) Phase-matched approach. The input coupling waveguide has a radius of curvature larger than that of the spoked-ring cavity. This approach efficiently excites the fundamental mode and does not introduce significant loss. (c) Sinusoidal-coupler approach with 1.2\(\mu\)m ring width. This method still achieves phase matching between the input waveguide and the spoked-ring cavity, and it minimizes straight to bend loss by slowly changing the radius of curvature along the input waveguide. It is also simpler to implement in a layout. (d) Power coupling from bus to ring vs. gap from the FDTD simulation shown in part (c). Coupling to the 1st order mode is about two orders of magnitude stronger than coupling to the 2nd order mode over the range of relevant coupling coefficients. (Original simulation by Jeff Shainline. Modified for use here.)
2.14 Layout of spoked-ring cavity and modulator. (a) Passive body silicon structure showing the silicon spokes that will be used for electrical contact. The inner circular silicon structure is an integrated heater for thermal tuning. (b) Layout view of the doping shapes to create a carrier-plasma optical modulator. The p/n regions are discretized to ≈200 nm grid to meet design rules imposed by the process. The silicon heater is blanket doped with a high concentration n-type dopant to create a low resistance heater that can be driven by on-chip circuits. (c) Zoom in of the modulator showing a close up of the p/n regions and the electrical wiring. The silicon “spokes” are covered with high concentration dopants and silicide to create highly conductive contacts to the p-n junctions. Note that the width and pitch of the metal layers are of critical importance to allow the tightly packed wiring.

2.15 Typical design flow for integrating photonics into a CMOS electronics design flow.

2.16 (a) 3D render of GDS layout of actual device submitted for fabrication. The electrical wires and vias form a coliseum-like structure around the inner radius of the microring. The anode (p-type connection) and cathode (n-type connection) wires are shown. There is a resistive thermal tuner integrated inside the ring which can be seen through the vias. (b) Zoom-in on the microring showing the radially extending p-n junctions. Level 1 (contacting the silicon) vias can be seen as well as the resistive silicon heater. (c) Optical micrograph of the fabricated device. The dark ring on the inner sidewall shows the silicide regions where the vias contact the silicon. The embedded heater and its contacts can also be seen. The green structures are density fill shapes to avoid dishing during fabrication. (d) Optical micrograph of the top-side of the fabricated chip. (e) Optical micrograph of the back-side of the fabricated chip after substrate removal. The silicon devices can be seen. (f) Scanning electron microscope (SEM) image of the fabricated device. The inner radius contacts to the spoked-ring cavity can be seen. (SEM image by Xiaoge Zeng.)
2.17 Overview of spoked-ring modulator experimental characterization. (a) Passive spectrum of first generation device (G1) showing an FSR of 3.2 THz and \(\approx 8\) dB of extinction at the resonance near \(\lambda = 1264\) nm. The inset shows the resonance shift with applied DC voltage. (b) Passive spectrum of second generation device (G2) showing an FSR of 3.3 THz and \(\approx 12\) dB of extinction. The inset shows the resonance shift with applied DC voltage. (c) 5 Gbps eye diagram of first generation device with PRBS7 modulation data stream. The extinction ratio (E.R.) is \(\approx 5.2\) dB with \(\approx 4.5\) dB insertion loss (I.L.) (d) 5 Gbps eye diagram of second generation device with PRBS7 modulation data stream. The E.R. is \(\approx 7.3\) dB with \(\approx 3\) dB I.L.

2.18 Summary of static modulator measurements. (a) Through port optical spectrum showing five resonances. (b) Each resonance is fit with a Lorentzian through port model Eq. 2.21 which is used to calculate quality factors. (c) DC voltage is applied to the p-n junctions to measure the resonance shift. Lorentzian through ports are fit to each resonance, and the resonance frequency is extracted to calculate the resonance shift per volt. A shift of 4.1 GHz/V is measured with -0.5V to 0.5V applied voltage bias. (d) Approximate extinction ratio (E.R.) and insertion loss (I.L.) calculated from the DC shifts from resonance 2.

2.19 Approximate extinction ratio (E.R.) and insertion loss (I.L.) when modulating resonances 1-5 from Fig. 2.18(a), calculated from experimentally measured spectra under 0-state and 1-state drive. This static analysis clearly shows that resonance 2 will yield the best modulation. Resonance 2 is the same as shown in Fig. 2.18(d).
2.20 Illustration of efficient tunable cavities. (a) The microring resonator blanket doped with contacts on the inner sidewall. (b) Inner portion of the ring is doped (either p-type or n-type) to avoid excess optical absorption. (c) Inner portion of ring is doped, and many electrical contacts are placed around the inner sidewall of the ring. This allows a lower resistance when the contacts are wired in parallel. (d) Input and output waveguides shown. The input waveguide has an extended coupling region to selectively excite the fundamental transverse mode of the ring, and the output waveguide is a point-like coupler. (e) Schematic of model used to derive filter response.

2.21 Drop port insertion loss versus filter 3 dB bandwidth and (a) electron implant concentration (b) hole implant concentration.

2.22 (a) 3D render of GDS layout submitted for fabrication. The cavity is blanket doped with a p-type implant and contacted on either side of the inner sidewall. The drop port waveguide is not an extended coupler since only the fundamental mode of the cavity is excited, so the single mode coupling is preserved at the drop port. The heater wires are shown which connect the cavity to upper-level metal pads. (b) Optical transmission spectrum showing $\approx 75 \text{ GHz}$ 3 dB bandwidth with $<0.5 \text{ dB I.L.}$ (c) Drop port spectra versus tuning power. (d) Change in resonant wavelength versus tuning power showing $2.73 \text{ nm/mW}$ tuning efficiency.

3.1 Cross-sectional schematic of chip-to-fiber coupling through a vertical grating coupler. Light from the waveguide is incident on a period array of scattering elements (conceptually represented by point-like scatterers for illustration of concept). In the simple case shown, the structure is vertically symmetric which produces scattered light in both the up and down directions which can be collected by an optical fiber. The point-like scatterers are sources of spherically expanding phase-fronts which determine the output angle of the radiation.
3.2 Configurations for on-chip grating couplers in silicon photonics showing optical power radiated up \( (P_{up}) \) and down \( (P_{down}) \). (a) Single-etch step resulting in vertically symmetric radiation \( (P_{up} = P_{down}) \). (b) Metal or dielectric reflectors introduced in the vertical cross-section to achieve directional radiation \( (P_{up} \neq P_{down}) \). (c) Two-etch steps resulting in offset scattering sites which radiate asymmetrically in the vertical direction \( (P_{up} \neq P_{down}) \).

3.3 Highly directional grating couplers using two independently patternable layers to achieve (a) up and (b) down directionality.

3.4 (a) Orientation of coordinate system for equation derivations. (b) \( k \)-space picture for deriving equations in Sec. 3.2. \( \beta(\omega) \) is the propagation constant of the waveguide mode incident on the grating, \( K_G = 2\pi/\Lambda \) is the grating momentum, \( \theta_o \) is the optimal angle aligned with the output fiber, \( k_x \) and \( k_z \) are the x and z-components of the \( k \)-vector of the light radiated from the grating, \( \vec{k}_{cl} \) is the \( k \)-vector of the output radiation, and \( |\vec{k}_{cl}| = (2\pi/\lambda)n_{cl} \) is the radius of \( k \)-sphere in the cladding material with index of refraction \( n_{cl} \).

3.5 Two-sided 1-dB bandwidth versus Gaussian beam waist size for different group indices, \( n_g \), of the grating Bloch mode.

3.6 Optimal exponential field for maximizing the overlap with a Gaussian profile. The exponential and Gaussian curves are normalized such that \( \int_{-\infty}^{\infty} G_f(x) \, dx = 1 \) and \( \int_{-\infty}^{\infty} E_g(x - x_m) \, dx = 1 \).

3.7 Peak coupling efficiency (blue) and corresponding 1-dB bandwidth for fiber waist sizes of 2.5 \( \mu m \) (red) and 5 \( \mu m \) (yellow).

3.8 Grating-to-fiber peak power coupling efficiency \( (\eta_{peak}) \) as a function of normalized bandwidth (blue). Coupling efficiency as a function of 1-dB bandwidth for \( w_o = 5 \mu m \) (red) and \( w_o = 2.5 \mu m \) (yellow). \( n_g \) assumed to be 3.0.
3.9 Positional tolerance $\delta$ as a function of waist size ratio $r$ for several values of coupling efficiency. For each required efficiency, there is an optimal waist size ratio that gives the maximum positional tolerance.

3.10 Positional tolerance $\delta$ versus waist-size ratio $r$ and minimum coupling efficiency $\eta_{\text{min}}$. For a constant $\eta_{\text{min}}$, there is an optimal $r$ to achieve the largest positional tolerance $\delta$.

3.11 Coupling efficiencies $\eta_{\text{peak}}$, $\eta_{\text{min}}$, and $\eta_{\text{match}}$ versus $\delta$ directly showing the trade-off between $\delta$ and $\eta_{\text{peak}}$. $\eta_{\text{match}}$ is the coupling efficiency corresponding to $\delta$ when the waist-size ratio is $r = 1$ (i.e. matched Gaussian waist sizes) and is included for comparison.

3.12 1-dB angular tolerance versus normalized wavelength. Air cladding (red) and oxide cladding (blue) curves are shown.

3.13 Power coupling efficiency in decibels, $10 \log_{10}(\eta)$, shown for $\lambda = 1180, 1310, 1550$ nm (from top to bottom) and $w_o = 2.5, 5.0$ $\mu$m (from left to right), shown as a function of positional and angular misalignment $\delta x$ and $\delta \theta$.

3.14 Front-end device layers used to build transistors and optical devices. (a) Illustration of a MOSFET in a thin-SOI process. (b) Illustration of patterned crystalline body crystalline silicon (active region of transistor) to form a vertical grating coupler. (c) Illustration of patterned polycrystalline silicon (gate of transistor) and crystalline silicon to break the vertical symmetry for high directionality radiation. (d) TEM image showing the front end layers, the silicon substrate, and a few upper level dielectrics. (e) Zoom-in showing a unit cell of the single layer grating. (f) Zoom-in showing a unit cell of the dual layer grating.

3.15 Single layer grating demonstration. (a) Optical micrograph of fabricated device. (b) TEM image of several periods of fabricated device. (c) Transmission spectrum comparing design and experiment.
3.16 Illustration of arrayed antennas concept and building grating couplers from transistor layers. (a) Achieving upward directivity with two scatterers. The scatterers are positioned such that the scattered light is out of phase in the downward direction and in phase in the upward direction. (b) Downward directivity is achieved by changing the position of the scatterers. (c) Illustration of CMOS MOSFET and vertical grating couplers built in the same physical layers. (d) Simulated directionality to illustrate the concept used in the actual designed device which confirms that the conceptual model shown in (a) and (b) can be used to guide designs.

3.17 Summary of first generation dual layer grating. (a) Layout view. (b) Optical micrograph of fabricated device. (c) TEM image showing several periods of fabricated device. 3D FDTD simulation showing magnitude of the electric field at the output plane (bottom of chip) (d) without p-Si gaps (e) and with p-Si gaps showing significant distortion in the output field. (f) Comparison between measured and simulated transmission spectra. For $\lambda > 1200$ nm, the measured spectrum shows increasing ripple which indicates large reflections at the interface between the adiabatic waveguide taper and the vertical grating coupler.

3.18 Summary of simulation results of chosen grating design. (a) FDTD simulation showing majority of radiation is directed downward. (b) Zoom-in of unit cell from simulation showing dimensions of the designed grating. The period is 650 nm. (c) Reflection into waveguide and downward directivity ($P_{\text{down}}/P_{\text{up}}$). (d) Angular spectrum of output radiation for $\lambda = 1310$ nm. (e) Chip-to-fiber coupling efficiency versus angle and wavelength. (f) Peak efficiency of $\approx -1.06$ dB near $\lambda = 1320$ nm.
3.19 (a) Optical micrograph of the grating coupler. The single-mode waveguide is connected to the grating through a parabolic adiabatic taper. An AR tooth is placed at the front of the grating to minimize reflections. (b) Comparison between experiment and simulation at 11° coupling angle. 4th-order polynomial fit is used to calculate bandwidths. (c) Simulated and (d) measured transmission response as a function of λ and angle. (e) Comparison between simulated and measured peak coupling efficiencies versus fiber coupling angle. (f) Comparison between simulated and measured peak wavelengths versus fiber coupling angle.

3.20 (a) Coupling efficiency for λ = 1320 nm, relative to global maximum, versus dimensional errors in c-Si and p-Si widths. (b) Optimal coupling angle for each point in (a) versus dimensional errors in c-Si and p-Si widths. The −1 dB contour from (a) is shown in red.

3.21 (a) Relative coupling efficiency as a function of fiber translational misalignment relative to the optimal position. A 3D FDTD simulation was used to simulate the output radiation of the grating coupler which was then overlapped with a Gaussian fiber mode varying in position. The coupling efficiency is normalized to the global maximum. (b) x and y-misalignment from (a) compared to predicted tolerance using Eq. 3.39 (c) Angular tolerance simulated with 2D FDTD and compared with predicted tolerance from Eq. 3.39.

3.22 Comparison to other recently demonstrated devices found in literature. The results are color-coded for CMOS compatibility. Gratings that would require severe modifications to a CMOS process (e.g. custom wafers, dimensional tolerances unachievable with photolithography, precise placement of dielectric/metal reflectors) are shown in red. Moderate modifications are shown in yellow (e.g. silicon nitride patterning and custom etch steps). Green shows results that were demonstrated on the same chip with working transistors.
4.1 Schematic view of chip-to-chip WDM communication link.

4.2 Schematic for calculating optical link budget. Light is supplied by an off-chip laser and is coupled into the chip through a grating coupler with $T_g$ (dB) losses. The light propagates through a waveguide and is modulated by a microring. The modulator achieves an extinction ratio, $ER$ (dB), and attenuates the light by some insertion loss, $IL$ (dB). The light exits the chip through a second grating coupler, propagates through a fiber, and couples into the receiving chip through a third grating coupler. The detector is assumed to be resonant with responsivity, $R_{det}$ (A/W), defined as the ratio of photocurrent produced, $i_{pd}$ (A) to optical power in the waveguide, $P_{wg}$ (W). The receiver circuitry has a minimum peak-to-peak current sensitivity, $S_{rec}$ (A), required for error-free operation. The propagation losses are given by $(L_1 + L_2) T_p$ where $L_1$ ($L_2$) is the length traversed through chip 1 (chip 2), and $T_p$ is the propagation loss in dB/cm.

4.3 (a) Schematic of an 8-λ WDM row. This row can be either be a transmitter with resonant modulators or a receiver with resonant detectors. (b) Through port spectrum with the free spectral range (FSR) and channel spacing labeled. The usable data channels must fit within the FSR of the resonators.

4.4 Channel spacing as a function of spoked-ring radius increment.

4.5 (a) Passive (untuned) through port spectrum of 11-λ WDM transmitter row. The extinction on all resonances is between 8-12dB which is near critical coupling. A Lorentzian model is used to fit each resonance, as shown in (b). The 3dB linewidth is calculated from the Lorentzian fit and is shown in (c). The gray region shows the confidence intervals from the fit.
4.6  (a) Cross-section of mode solver simulation used to calculate the propagation loss as a function of copper distance. (b) Results of mode solver simulation for \( \lambda = 1180, 1310, 1550 \text{nm} \). The blue traces show propagation loss in dB/cm; the red traces show the approximate intrinsic quality factor corresponding to the propagation losses. The dotted line indicates an intrinsic quality factor of 1 million.

4.7  Cross-section of 45 nm process showing the first eight metal layers and a crystalline silicon waveguide (dimensions to scale). A minimum region of 1 \( \mu \text{m} \) is cleared around the waveguide to avoid excess propagation losses from metals.

4.8  Schematic of WDM compatible transceiver modulators/detectors based on microrings. The microrings have an embedded resistive thermal tuner to allow for resonance stabilization using on-chip feedback loops. (a) Detector configuration with passive microring filter drop port connected to broadband data detector and a weakly coupled port connected to a second broadband detector for thermal tuning feedback. (b) Resonant detector where photocurrent is generated directly in the resonator, and the feedback detector is implemented as a portion of the circumference of the ring. (c) Modulator with weakly coupled drop port connected to a broadband detector for thermal tuning feedback. (d) Modulator with portion of ring used as a feedback detector.

4.9  Example hierarchical design flow for enabling full electronic-photonic systems. (a) Basic photonic components. (b) Photonic devices. (c) Photonic sub-systems. (d) Electronic-photonic systems.

4.10 (a) Optical micrograph showing a 1 \( \lambda \) optical test site for transmitter characterization. (b) 3D render of the modulator layout. The metal wires that connect the optical device to the on-chip circuits can be seen (heater +/- and modulator anode/cathode). (c) Zoom in showing p-n junctions in the modulator. (d) Passive optical spectrum of modulator. The FSR is 3.07 THz (14.6 nm). (e) Resonance shift versus applied bias voltage resulting in 27 pm/V (5.7 GHz/V).
4.11 Modulator layout. (a) Crystalline silicon mask shapes with labeled dimensions. The drop port is weakly coupled and connected to a broadband photodetector used in a wavelength locking feedback loop. (b) Zoom in showing c-Si contacts to the p-type and n-type regions. (c) Layout view showing n-type, p-type, metal 1, and metal 2 mask shapes and the anode (p-type) and cathode (n-type) contacts. (d) Zoom in showing the implant regions and metal connectivity.

4.12 Schematic of transmitter on-chip experiment. An off-chip FPGA is used to communicate with on-chip circuits to configure experiments. The modulated optical output is viewed on a sampling oscilloscope.

4.13 (a) Through port spectra with several applied heater DAC codes. Maximum tuning range corresponds to \( \approx 50 \text{K} \) of temperature change. (b) Resonance wavelength as a function of heater DAC codes (blue); dissipated heater power (red). (c) Through port transient response under applied step voltage (DAC code 0 to 511).

4.14 (a) Optical micrograph of on-chip transmitter characterization site showing the optical modulator and the on-chip circuitry (data drivers, tuners, controllers, data generation). (b) PRBS31 5Gbps optical eye diagram taken with an off-chip optical sampling oscilloscope. The E.R. and I.L. is 6.5 dB and 3 dB, respectively, with a voltage swing of -0.7 V to 0.5 V. The drive inverters and the 8-to-1 serializer consume 30 fJ/bit and 140 fJ/bit, respectively.

4.15 5Gbps eye diagrams with 64-bit patterns of varying ratio of transmitted zeros. (a) Average photocurrent tracking. (b) Eye-height tracking. (c) Eye-height and self-heating cancellation.
4.16 11λ DWDM transmit row. (a) GDS layout view with macro slices labeled showing each transmit unit cell. (b) Optical micrograph of chip showing DWDM row, input/output gratings, cascaded microrings, and backend circuits. (c) Tuned (red) and untuned (gray) through port optical spectra of DWDM row. (d) 8 Gbps eye diagrams of each transmit slice giving $11 \times 8 = 88$ Gbps aggregate bandwidth coupled into a single fiber.

4.17 (a) Optical micrograph of SiGe photodetector. The absorbing region in the photodetector is 50 $\mu$m long. (b) Cross-section of photodetector showing doping regions. The carriers are generated in the SiGe and swept out by the reverse biased p-n junction and collected through the metal contacts. The polysilicon region is used to concentrate the optical mode away from the metal contacts to reduce propagation loss. (c) Optical micrograph showing on-chip receiver test site.

4.18 Receiver characterization. (a) Schematic of receiver test setup. An off-chip FPGA generates a 10 Gbps PRBS31 signal which is amplified and used to drive an MZI modulator (optical eye diagram shown). The modulated light is coupled into the chip through a vertical grating coupler terminated by the photodiode. The photocurrent is received by an on-chip receiving frontend which performs in situ bit error rate (BER) checking and exports the results to an off-chip test FPGA for viewing. (b) Bit errors as a function of sampling time offset and receiver current decision threshold. (c) Bathtub curves for several datarates. (d) BER as a function of average photocurrent for several datarates.

4.19 5 Gb/s chip-to-chip link experiment. (a) Schematic view of the chip-to-chip link testing. The optical power at several critical stages is shown in (b), and the receiver BER eye is shown in (c).
4.20 Thermal stress tests of chip-to-chip link. (a) Bit-errors with (blue) and without (red) thermal tuning locking. (b) Output of thermal tuner controller showing dynamic adjustment to on-chip power fluctuations shown in (c). The optical eye diagrams with tuning disabled (d) and enabled (e) show that the eye diagram is completely closed and unstable without the thermal tuner.

4.21 Microprocessor with photonic I/O. (a) Optical micrograph image of the fabricated 6 × 3 mm chip. The image is taken from the backside of the chip after full substrate removal. The bottom portion of the chip contains a dual-core RISC-V processor which utilizes the processor + memory transceiver banks to communicate with the 1 MB memory bank. Independent transceiver test sites are included at the top of the chip for prototyping new designs. (b) Zoom-in of the critical components in the processor + memory transceiver banks showing the WDM transmitter bank (top) and the WDM receiver bank (bottom). (c) Photonic components that comprise the communication link.

4.22 (a) Post-processing XeF₂ etch step to enable simultaneous optical and electrical functionality. (b) Optical micrograph of the packaged chip after XeF₂ etch.

4.23 Schematic of chip-to-chip photonic communication link. The off-chip control FPGA communicates with chip 1 to configure experiments and allow user input. Chip 1 is in “processor mode” which utilizes the RISC-V processor. The processor makes read/write requests to chip 2, which is in “memory mode”, using the photonic link which is accessed through a memory controller. A 1λ continuous wave laser is split 50/50 to power both transceiver paths. An optical amplifier is used between chips to overcome the link losses.
4.24 (a) Steps for program execution using the photonic chip-to-chip link. In step 1, the control FPGA writes the program’s instructions into memory on chip 2 through direct memory access (DMA) through the memory controller on chip 1. After loading the program, the FPGA issues a reset signal to the processor, and the processor fetches the first program instruction from memory address 0x00002000 (step 2). In step 3, the processor executes the program, writing and reading to and from memory as well as reading instructions from memory. (b) Terminal output of hello world! and STREAM programs utilizing the optical link. (c) Graphical output of the teapot renderer program which pixel shades a 3D teapot using the Blinn-Phong shading model. The rendered image is output to the control FPGA which is displayed to the user.

4.25 Laboratory setup for chip-to-chip communication testing. (a) Optical table with important components labeled. (b) Optical data signals for processor-to-memory and memory-to-processor communication links.

4.26 1000 seconds processor stress test. (a) Processor power and resulting microring heater power for transmitter stabilization. (b) Processor power and resulting transmission error rates with thermal tuning on and off. Without thermal tuning, the communication link is practically unusable.

5.1 Resonant systems capable of a 2-pole, 1-zero response: (a) abstract representation of a 2-pole, 1-zero photonic circuit; (b) a physical implementation that uses a weak tap coupler to give rise to interference that produces the transmission zero.

5.2 Higher-order filters with a drop-port transmission zero: (a) a device architecture that can produce one drop-port zero in arbitrarily high order filters; (b) example response of a 2nd-order filter using Eqs. 5.6, 5.7 and a zero placed at $\delta \omega_{zd} = 10r_i$. 
5.3 Abstract photonic circuit used to derive the T-matrix of the tapped-filter: (a) schematic of a 2-ring filter with 3 input and 3 output ports; (b) graphical representation of the drop-port zero location in the complex-$\delta\omega$ plane.

5.4 Comparison of a pole-zero filter with an asymmetric response, and an all-pole Butterworth filter. In the pole-zero filter, the zero is placed at $\delta\omega/r_i = 2.3$.

5.5 Constructing a serial demultiplexer with symmetrized, densely packed passbands by using asymmetric response filters: (a) illustration of a two-channel demultiplexer; (b) Channel 1 drop port response, $|T_{31}|^2$; (c) Channel 1 through port response, $|T_{21}|^2$; (d) Channel 2 drop port response, $|T_{61}|^2$, showing a highly selective response due to the transmission zero on the right, and through-port extinction of the previous stage on the left.

5.6 Design example demonstrating higher bandwidth density and denser channel packing in a serial demultiplexer based on asymmetric second-order filter stages: (a) example demultiplexer design using pole-zero filters shows 20GHz passbands with 44GHz channel spacing; (b) example design using conventional, all-pole Butterworth filters (same passband shape and bandwidth) is limited to 106GHz channel spacing.

5.7 Proposed device topologies that support a 2-pole, 1-zero drop-port response: (a) general, abstract design with all possible (non-trivial) degrees of freedom; (b) tap-coupler implementation (analyzed in detail in this paper), (c) phased parallel-coupled-ring implementation, (d) 2-poles, 1-zero with minimal degrees of freedom. (b-d) are limiting cases of the general geometry in (a).

5.8 Optical micrographs of the fabricated devices. (a) Full chip made in 45nm SOI CMOS; (b) magnified image showing single device.

5.9 Measured transmission spectra of pole-zero devices. (a) Response with zero detuning at 26GHz; (b) response with zero detuning at 45.6GHz; (c) response with zero detuning at 74.4GHz.
5.10 (a) Physical implementation showing two coupled rings that are modulated anti-
symmetrically. (b) Abstract representation of wavelength conversion where \( b_1 \) and 
\( b_2 \) are the supermode amplitudes of the unmodulated \( a_1, a_2 \) system. (c) Conceptual 
figure showing conversion from \( \omega_1 \) to \( \omega_2 \). Since the supermode system only supports 
two resonances, the lower generated sideband is unused. (d) \( |S_{21}|^2 \). Maximally flat 
(left), Chebyshev (middle), and optimal design with loss (right). Center frequencies 
are offset for clarity.

5.11 Wavelength conversion efficiency as a function of modulation strength normalized to 
intrinsic bandwidth.

5.12 (a) Three coupled rings with the outer rings modulated antisymmetrically. (b) Ab-
stract representation of wavelength conversion where \( b_1, b_2, \) and \( b_3 \) are the supermode 
amplitudes of the unmodulated \( a_1, a_2, a_3 \) system. (c) Conceptual figure showing con-
version between the three supermode resonances. To match the conversion efficiency 
of the two ring case, light is sent in to the \( \omega_2 \) resonance (green) and \( \omega_1 \) (red) and \( \omega_3 \) 
(blue) are generated.

5.13 (a) Simplified model for carrier-plasma modulator calculations. In the approxima-
tions used, the cavity goes from a negligibly small depletion width (top) to fully 
depleted (middle) which introduces a complex perturbation to the refractive index 
(bottom). Calculated resonance shift normalized to the total 3 dB linewidth for 
intrinsic quality factors \( b) Q_o = 10^4 \) (c) \( Q_o = 10^6 \) (d) \( Q_o = 10^6 \).

6.1 (a) Layout view of photonic crystal nanobeam and transverse mode of input and 
output excitation waveguides. (b) Nominal design dimensions labeled showing the 
square hole tapering from the middle (320 nm) to the ends of the nanobeam (258 nm). 
(c) Simulated fundamental and third order longitudinal modes of nanobeam cavity. 
(d) Measured through (green) and drop (yellow) port spectra of fabricated device. 
(e) Zoom-in of fundamental mode resonance compared to a calculated lossless response.
6.2 (a) Configuration of two standing-wave cavities to result in the equivalent of a (b) traveling-wave microring cavity.  (c) Optical micrograph of fabricated device and (d) 3D renders of the tuning arms.  (e) Passive (untuned) spectra.  (f) Phase arms tuned to approximate correct relative phase, $|\phi_1 - \phi_2| = \pi$.  The phase is also tuned to result in (g) the lowest drop port insertion loss and (h) the highest drop port suppression.

6.3 Optical micrograph of fabricated 3×3 mm chip.  (a) Top side view.  (b) Bottom side view.  (c) Zoom-in of silicon microring device used for stimulated and spontaneous FWM experiments.  (d) Simulated optical mode of the FWM microring.  The outer radius is 22 $\mu$m, and the waveguide width is 1.08 $\mu$m.

6.4 (a) Passive through port transmission spectrum.  (b) Stimulated FWM efficiency.

6.5 Coincidence-to-accidental ratio measurements with time windows of (a) 240 ps and (b) 611 ps showing CARs greater than 50 are achievable.

6.6 (a) Top and (b) bottom view of fabricated die.  (c) Fundamental electric field profile of typical single mode waveguide.  (d) 4-channel tunable filter bank.  (e) Vertical grating coupler.  (f) Microring tunable filter.  (g) Depletion-mode optical modulator.

6.7 (a) Spectral response of single vertical grating coupler showing $\approx 40\%$ peak coupling efficiency.  (b) Filter bank through port response.  (c) Through and drop of one filter in the filter bank.  (d) Passive spectrum of modulator.  (e) Resonance shift with applied voltage bias.  (f) I-V curve of modulator.  (f) 1 Gbps (g) 5 Gbps (h) 8 Gbps eye diagrams.
6.8 (a) 3D render of fabricated device. The upper left shows a zoom-in of the p-n junctions integrated in the multi-mode straight sections. The bottom right shows a top view of the cavity. (b) Simulated electric field pattern showing the “wiggling” mode in the multi-mode region avoiding the electrical contacts. (c) Passive optical spectra of 5-period and 15-period cavities. (d) Optical micrograph of fabricated 5-period device. (e) Optical micrograph of fabricated 15-period device. (f) I-V curves showing successful formation of rectifying junctions.

6.9 (a) Resonance shift with applied DC voltage bias. (b) 5 Gbps eye diagram using 15-period device. (c) 5 Gbps eye diagram using 5-period device.

6.10 Schematic of chip-to-chip test setup and resulting eye diagrams from the split receivers showing better than $10^{-10}$ BER.

6.11 Schematic of WDM transmitter testing and optical micrograph of on-chip test macros. Individual eye diagrams of each transmitter are shown indicating that the system can achieve 45 Gbps (9×5 Gbps) aggregate transmission bandwidth on a single fiber.
Chapter 1

Introduction

Integrated photonics has a rich history of impacts in technology and basic science, beginning with a flurry of activity in the late 1960’s [71,72,80]. In the decades since Miller’s original paper titled “Integrated Optics: An Introduction,” all of the basic device concepts (e.g. waveguides, resonators, distributed Bragg reflectors, phase shifters, directional couplers, semiconductor lasers, etc.) have been realized and form the core of integrated photonics technology. To date, most integrated photonics found in commercial technology is based on III-V compound semiconductors, materials with $\chi^2$ electro-optic non-linearity, or silica-based waveguides which found success in long-haul telecommunications applications and light sources. These types of devices, however, are struggling to meet modern communications challenges due to high costs, power consumption, and small areal bandwidth densities. Silicon photonics has emerged as a technology that can meet the staggering demand for bandwidth coming from modern digital chips from microprocessors that span consumer and supercomputer needs, to network chips for data centers. This thesis presents photonic devices and electronic/photonic systems built directly in CMOS microelectronics chips and demonstrates the feasibility of this approach to integration to build complex, high-performance electronic/photonic systems-on-chip (SoC’s). The two key results are: 1) demonstration of high performance components in CMOS; and 2) demonstration of the largest scale electronic/photonic integration to date. Taken together, these resulted in the first microprocessor with photonic I/O and represent an important milestone to enabling large-scale silicon photonic systems on chip.
1.1 Evolution of silicon photonics

Although integrated photonics has been an active field of research since the 1960’s, silicon photonics was not a popular field of research until fabrication technologies were developed that could support high-index contrast devices. In particular, silicon-on-insulator (SOI) wafers were beginning to be widely commercialized in the early 1990’s as a technology for enhanced transistor performance [11]. Optical researchers started to explore integrated photonic devices in silicon around the same time. SOI technology offered lithographically patternable crystalline silicon which paved the way for feasible high-index contrast photonics in silicon. There were considerable difficulties, however, in enabling high-index contrast devices due to their sensitivities to fabrication tolerances. Namely, lithography induced sidewall roughness, and its impact on propagation loss, polarization issues, and challenges with resonators (control of absolute resonance frequency, critical coupling, etc.) were some of the early obstacles researchers had to overcome. For example, the first silicon microring-based filters demonstrated loaded Q’s of ≈ 250 in 1998 and were achieved with modest fabrication fidelity [31,65], while today, Q’s of > 100,000 are commonplace.

After the feasibility of passive silicon photonic devices were demonstrated, the next obstacle was enabling electrically active devices (phase shifters, tunable filters, detectors, modulators). The large thermo-optic coefficient of silicon provides a mechanism for introducing phase shifts to an optical wave through temperature change. On a silicon chip, in turn, localized temperature change can be achieved by designing resistive microheaters, i.e. through simple resistive Joule-heating. Thermal control of photonic phase shifts for resonance tuning, Mach-Zehnder interferometers, etc., but it is limited to $\mu$s response times. Thus, a mechanism for faster (GHz speeds) dynamics is also needed, and for data modulation, one was found in the carrier-plasma effect [123]. Soref’s watershed publication that estimated the carrier-plasma dispersion effect in crystalline silicon offered a vision for building electrically controlled photonic devices. This was a critical development for active photonics in silicon, since crystalline silicon, due to crystal inversion symmetry, lacks a $\chi_2$ nonlinearity that is the dominant mechanism for modulation in many active photonic devices. The
carrier-plasma effect is now widely used in silicon photonic modulators, and since carrier implantation is extremely well understood and engineered in CMOS-like fabrication technologies, there is a clear path for mass-production of silicon photonic modulators.

Today, silicon photonics seems poised to experience wide commercial success. Large companies (e.g. Intel, IBM, Cisco, Huawei, Oracle) have made investments in silicon photonics, and some are forming product divisions based on the need for high bandwidth communications. However, the economics of silicon photonics as a high-volume communication interconnect technology has not been fully resolved with key aspects such as packaging, reliability, and light sources remaining a challenge. Regardless, the core technology has been demonstrated, and there is confidence among those in the field that the remaining practical issues will be solved.

1.2 Chip-to-chip I/O bandwidth limitations

The majority of the work in this thesis was motivated by a problem that emerged in the mid 2000’s. As computational power followed Moore’s Law throughout the latter half of the 20th century, a serious problem occurred in the early 2000’s. The clock rate of digital circuits hit what is known as the “power wall” \[93\]. In the previous decades, computational power was mostly increased through increasing the switching speed of the transistors that comprised the digital circuits. This was made possible through the steady improvement of photolithography which achieved smaller feature sizes year after year, thereby reducing the size of the transistor which increased the RC-limited switching speed. Eventually, the dynamic power consumption of switching the transistor became the bottleneck. The dynamic power consumptions scales with \(CV^2f\) with \(C\) capacitance, \(V\) voltage, and a \(f\) switching frequency of a switch/transistor. To circumvent this bottleneck, multi-core processor chips were introduced and ushered in a paradigm shift in computer architecture and programming \[2\]. While the multi-core paradigm has brought many challenges, it has allowed computational power to continue to scale through lithography improvements since more cores could fit on a chip from generation to generation.

Although the compute power of a single chip was able to continue scaling through multi-
core technology, the next bottleneck was immediately recognized: chip-to-chip I/O bandwidth. In a simplified computer system, there is a processor chip with a small amount of on-chip memory (cache), and a set of memory chips that provide the processor with a large pool of dynamic random access memory (DRAM). In commercially available systems today, the DRAM chips are connected to the processor chip through a printed circuit board (PCB), the motherboard. The PCB contains copper wires of cm-scale that provide the interconnection. The copper wires, however, suffer from prohibitively high losses at >GHz switching speeds. This has led to processor chips evolving to have as much on-chip memory as possible since any read/write from/to off-chip memory comes with large bandwidth and latency penalties. It is this bandwidth limitation that silicon photonics can potentially alleviate.

1.3 “Zero-change” CMOS integration

One key aspect of silicon photonics technology is the strategy used to integrate the photonic devices with electronics. The photonics community has fiercely debated this issue, and there are two main approaches: monolithic integration and heterogeneous integration.

Monolithic integration offers an elegant solution for building electro-optic systems, but it has been difficult to achieve due to pragmatic engineering constraints. Joint optimization of fabrication processes for high-performance transistors and high-performance photonic devices has proven challenging, and there are many trade-offs that must be considered. Moreover, it is exceedingly costly due to small market sizes, and as a result, custom developed monolithic photonics electronics processes have been limited to older process nodes with limited transistor performance. However, the reduced parasitic resistances and capacitances resulting from short-reach metal interconnects offer compelling advantages in terms of energy consumption and fabrication yield.

Heterogeneous integration attempts to factorize the fabrication optimization; the electronics fabrication process is optimized independently from the photonics fabrication process such that

\[ \text{[\text{In this thesis, “monolithic integration” will refer to all components being integrated on the same chip except for the laser.}]} \]
optimal performance is achieved for both the electronics and the photonics. In practice, this approach has had limited success to date due to the complexity of connecting the two chips together in a way that preserves the low parasitics and large connection density needed to leverage the full benefits of photonics.

In this thesis, the feasibility of a “zero-change” approach to photonic integration in building high-performance electro-optic systems is investigated and successfully demonstrated. This approach was pioneered by Orcutt et al. [85, 86, 88, 89], and the early results showed promise but limited performance. The main idea of zero-change CMOS photonics is to use existing advanced microelectronics fabrication processes to directly integrate photonics alongside electronics by reusing the same silicon device layers (and mask levels) that form the transistors. This allows photonics designers to leverage the mature infrastructure developed in microelectronics and leads to high-yield, reproducible devices integrated alongside arbitrarily complex digital and analog systems on chip. This approach was explored for both bulk silicon processes [86] and SOI processes [88]. Due to unfavorable optical properties of polysilicon, integration of photonics into bulk silicon processes proved difficult [86]; after much effort, however, the first monolithic chip-to-chip link in bulk silicon was demonstrated, with contributions from work that forms this thesis, using a modified version of Micron Technology’s NOR Flash memory periphery process [126]. Since SOI wafers have a crystalline silicon device layer for the active region of the MOSFET transistor, a more direct path to photonic functionality was possible [88] in IBM’s 45 nm SOI platform [61], by using this device layer as the optical guiding layer. Early demonstrations showed compelling results that hinted at the potential of the platform; the first monolithically integrated transmitter showed 2.5 Gbps modulation with 1.23 pJ/bit energy consumption [81]. The optical modulator was a ridge-type microring resonator using the polysilicon gate layer as the ridge. The polysilicon, however, is lossy which gives a wide linewidth resonance. The device had to be operated in carrier-injection mode (forward bias of the p-n junction) to shift the resonance enough to get appreciable modulation of the optical guided wave. This injection design is what leads to the modest energy consumption of 1.23 pJ/bit. However, the initial feasibility of the platform was demonstrated.
Figure 1.1 shows a simplified illustration of optical waveguide integration into bulk silicon [Fig. 1.1(a)] and an SOI platform [Fig. 1.1(b)] CMOS fabrication processes. Other than a brief description of highlighted results on bulk CMOS in Ch. 6, the work presented in this thesis focuses on integration in SOI CMOS processes, and a more detailed integration cross-section is shown in Fig. 1.2 which contains the main features of the SOI process. The photonic devices are patterned and built using the same silicon layers as the transistors; three example waveguides and a vertical grating coupler are shown. Special implant blocking masks must be accessed to prevent the optical devices receiving unintended doping and the resulting high propagation loss.

Although the polysilicon (transistor gate device layer) can be patterned to form full-etch rectangular waveguides as well as ridge waveguides (in conjunction with the crystalline silicon), the propagation loss is prohibitively high (> 50 dB/cm) for most applications and hence polysilicon is typically avoided in design. The exception is the vertical grating coupler; since the polysilicon shapes are used mostly as scattering elements rather than waveguides with significant propagation lengths, the polysilicon’s high propagation loss has a negligible effect on the grating coupler, but it is critical to providing the phased array function and directionality of radiation. The polysilicon was also used as a guiding layer in early photodetector designs [3], but its optical loss contributes to the low responsivity of the devices.

The crystalline silicon device layer is used as the primary material layer for constructing
The photonics implemented in the 45 nm SOI fabrication process was previously measured to support single-mode waveguides with propagation losses of $2 - 3 \text{ dB/cm}$ and $3 - 4 \text{ dB/cm}$ near $\lambda = 1310\text{ nm}$ and $\lambda = 1550\text{ nm}$, respectively, which is near the propagation losses obtained in the common customized silicon photonics processes. The transistor doping implants can be utilized to build active photonic devices (modulators, tunable filters, detectors). Finally, the metal interconnects (of which there are 10 levels to support the complex interconnectivity of microprocessors for which the 12SOI process is designed) are used to electrically connect the transistor circuits to the photonic devices.

### 1.4 Outline of thesis

In this thesis, several devices critical to building electronic/photonic communication systems on chip are designed and experimentally demonstrated, and their use in chip-to-chip communications as part of complex electronic-photonic systems on chip is demonstrated.

In Chapter 2, the basics of active photonic devices in the 45 nm CMOS process are presented.
This includes waveguide and microring resonator simulations and designs, a thermally tunable microring filter with record tuning efficiency, and an energy-efficient depletion-mode microring modulator with an embedded thermal tuner. These device demonstrations form the basis of building larger integrated systems.

In Chapter 3, simulations, designs, and experimental demonstrations are shown for a two-level vertical grating couplers that are implemented using the polysilicon gate and crystalline silicon active region of the MOSFET transistor. High coupling efficiencies ($-1.1\,\text{dB}$) and wide $1\,\text{dB}$ bandwidths (14 THz) are achieved. The two-level grating design is inspired by antenna phased arrays and results in a structure that achieves $> 99\%$ directionality of the launched radiation. The structure can be applied to any dielectric cross-section that supports two independently patternable layers.

In Chapter 4, electronic/photonic systems utilizing the devices presented in Chapters 2 and 3 are experimentally demonstrated. This includes a monolithically integrated chip-to-chip communication link and the world’s first microprocessor chip to communicate with the outside world using light. These demonstrations show that, despite challenges in current state-of-the-art, microring-based wavelength division multiplexed (WDM) communication systems are viable and can be successfully stabilized using on-chip feedback control techniques.

In Chapter 5, two novel microring devices are proposed and theoretically explored. The first is an add/drop filter that is particularly well-suited for WDM filter banks. By cascading these filters, adjacent channel spacings are decreased by $\approx 2\times$ which allows for greater bandwidth utilization. A single filter, as opposed to the full bank, is experimentally demonstrated and shown to be realizable. The second proposed device is a wavelength conversion scheme using two coupled microring modulators, and its design and predicted performance is presented. This device may have impact in RF signal processing applications – a potential application area for electronic-photonic SoC’s.

In Chapter 6, an outlook on monolithic integration is presented. The relevance of the 45 nm process to devices and applications beyond data communications is experimentally demonstrated
through: 1) a nanobeam photonic crystal cavity and 2) a nonlinear microring-based source of quantum-correlated photon pairs. Next, the “zero-change” approach to monolithic integration is extended to IBM’s 32 nm SOI technology node, and the first photonic devices in this platform are experimentally characterized. Finally, highlights from work on monolithic integration in a bulk CMOS process are presented which serve as an important milestone to enabling commercially viable photonic systems for lower cost, higher volume applications than is accessible to SOI.

Chapter 7 concludes the thesis, and the main results from Chapters 2-6 are summarized and placed into context. Several remaining challenges and directions for future research are identified.
Chapter 2

Active microring-based devices in thin-SOI CMOS

2.1 Introduction

One of the most exciting devices in silicon photonics is the microring resonator. In research, it has been used in filtering \[65,78,95,99\], modulation \[132,141,147\], photodetection \[75\], nonlinear four-wave mixing \[7,17,26\], signal processing \[66\], and even as a generator of non-classical states of light for quantum systems \[43,59\]. Although the silicon microring resonator has gone through a proliferation of applications in research, it has yet to enter into real-world technologies. This is partly due to the overwhelming majority of research results being obtained using commercially unrealistic fabrication processes; in addition, the microring resonator is extremely sensitive to fabrication tolerances and thermal perturbations, making simple demonstrations with the help of plenty of laboratory equipment doable, but stable, repeatable results without use of expensive table-top equipment a challenge.

In this chapter, we present two active devices based on silicon microrings that are critical building blocks in realistic communication systems. They are realized using 45 nm CMOS technology based on deep-UV photolithography, which is immediately manufacturable and scalable to high volumes. These devices form the building blocks for larger system demonstrations on chip (covered in Ch. 4) culminating with the first CPU with photonic I/O. Thus, this demonstration of a CMOS active photonic device platform represents an important milestone for silicon photonics finally breaking into real-world technologies.

First, an introduction to the basic principles of active silicon microring devices is presented
in Sec. 2.2 and 2.3. Next, Sec. 2.4 presents designs for waveguides and microrings in thin-SOI CMOS which form the basic photonic circuit building blocks. Section 2.5 reviews the constraints of a thin-SOI CMOS microelectronics process and introduces the “spoked-ring” microcavity as a suitable variant of the family of microring resonators to build active devices in CMOS. In Sec. 2.5.1, simple equations to approximate the resonance shift of the “spoked-ring” modulator are derived. Finally, Sec. 2.6 presents experimental results of the first depletion-mode modulators in a 45 nm CMOS process which achieve near state-of-the-art energy efficiencies even when compared to state-of-the-art silicon photonics results in custom photonics processes. Notably, combined with on-chip electronic drivers, they eclipse custom photonics showing record energy/bit [37,125]. Additionally, a tunable microring filter is also presented which achieves a record tuning efficiency of 2 µW/GHz.

2.2 Basic properties of microring-based active devices

This section introduces fundamental principles of microring-based active devices, in particular devices built with a silicon core. We derive several useful equations related to dielectric semiconductor resonators that will be used throughout this chapter.

2.2.1 Resonance shift given an index perturbation

Optical phase modulators typically work by having some mechanism change the refractive index in an optical path to achieve phase modulation. An interferometer (e.g. Mach-Zehnder) converts this phase to amplitude modulation for data encoding. In resonant modulators, such as microring modulators, a given perturbation to the refractive index leads to a change in the resonance frequency. If the ring is in a notch filter configuration, this can be used to modulate the attenuation that laser light experiences as it passes the resonator, thus resulting in amplitude modulation. In this section, we derive the equation for the change in resonance frequency given a perturbation to the refractive index. This equation will be used to approximate the characteristics of microring modulators using the carrier-plasma effect, so it is worthwhile to present its derivation in detail.

Assuming a static system $\epsilon(r)$ (i.e. modulator at rest), we start with the time dependent
form of Maxwell’s equations.
\[ \nabla \times \mathbf{E} = -\mu \frac{\partial \mathbf{H}}{\partial t} \quad (2.1) \]
\[ \nabla \times \mathbf{H} = \epsilon(r) \frac{\partial \mathbf{E}}{\partial t} \quad (2.2) \]

Taking the curl of Eq. 2.1 and plugging Eq. 2.2 into the result gives
\[ \nabla \times \nabla \times \mathbf{E} = -\mu \epsilon \frac{\partial^2 \mathbf{E}}{\partial t^2} \quad (2.3) \]

Assuming time harmonic solutions, \( \mathbf{E} = \mathbf{E}(r) \exp(j\omega t) \), the time derivative goes to \( \frac{\partial^2}{\partial t^2} \to -\omega^2 \) giving
\[ \nabla \times \nabla \times \mathbf{E} = \mu \epsilon(r) \omega^2 \mathbf{E} \quad (2.4) \]

This is the eigenvalue equation for the resonant modes of the material configuration \( \epsilon(r) \). Dot-producting Eq. 2.4 by \( \mathbf{E}^* \) and taking the volume integral over all space yields an equation for \( \omega^2 \).

This is a stationary integral for \( \omega^2 \) \[53\].

\[ \omega^2 = \frac{1}{\mu} \frac{\int_V \mathbf{E}^* \cdot \nabla \times \nabla \times \mathbf{E} \, dV}{\int_V \mathbf{E}^* \cdot \epsilon(r) \mathbf{E} \, dV} \quad (2.5) \]

We assumed static \( \epsilon(r) \) because the modulation is weak, \( \epsilon(r, t) = \epsilon(r) + \delta\epsilon(r, t) \) and changes the resonance frequency but not the mode field distribution \( \mathbf{E}(r) \). This is because the high index contrast of silicon to (typically) silicon dioxide (about 3.5:1.45) defines the modes whereas \( \delta\epsilon \) corresponds to \( \delta n \) on the order of < 0.01 with the type of modulation we are using. Next, we are interested in approximating the change in resonance frequency, \( \delta\omega \), for a perturbation to the dielectric constant, \( \delta\epsilon \). We let \( \omega \to \omega_o + \delta\omega \) and \( \omega^2 \approx \omega_o^2 + 2 \delta\omega \omega_o \) where we have neglected the \( \delta\omega^2 \) term, and \( \epsilon(r) \to \epsilon_o(r) + \delta\epsilon(r) \). Substituting these terms into Eq. 2.5 gives

\[ \omega_o^2 + 2 \delta\omega \omega_o = \frac{1}{\mu} \frac{\int_V \mathbf{E}^* \cdot \nabla \times \nabla \times \mathbf{E} \, dV}{\int_V \mathbf{E}^* \cdot [\epsilon_o(r) + \delta\epsilon(r)] \mathbf{E} \, dV} \quad (2.6) \]

Rearranging Eq. 2.6 gives

\[ \omega_o^2 + \omega_o^2 \frac{\int_V \mathbf{E}^* \cdot \delta\epsilon(r) \mathbf{E} \, dV}{\int_V \mathbf{E}^* \cdot \epsilon_o \mathbf{E} \, dV} + 2 \delta\omega \omega_o = \frac{\int_V \mathbf{E}^* \cdot \nabla \times \nabla \times \mathbf{E} \, dV}{1/\mu \int_V \mathbf{E}^* \cdot \epsilon(r) \mathbf{E} \, dV} \quad (2.7) \]
The term on the right-hand side of Eq. 2.7 is recognized to be equal to \( \omega_o^2 \) from Eq. 2.5 and cancels the \( \omega_o^2 \) term on the left-hand side of Eq. 2.7. Simplifying the resulting equation yields the desired equation for \( \delta \omega \) as a function of \( \delta \epsilon \):

\[
\frac{\delta \omega}{\omega_o} = -\frac{1}{2} \int_V E^* \cdot \delta \epsilon E \, dV \quad \frac{\omega_o}{2} \int_V E^* \cdot \epsilon_o E \, dV \tag{2.8}
\]

Note that \( \delta \epsilon \) can be a complex number which means the material is lossy. This leads to an imaginary component in \( \delta \omega \) which can be used to calculate the decay rate and thus the quality factor of a resonant mode of a cavity.

### 2.2.2 Relation between material absorption and resonator quality factor

In this section, several equations are derived that relate the propagation loss in a material to the quality factor of a resonator built using that material.

The quality factor of a resonator is defined as

\[
Q \equiv \frac{2\pi}{2} \frac{\text{energy stored in resonator}}{\text{energy lost per cycle}} \tag{2.9}
\]

If the electric field in the resonator decays at rate \( r_l \) (s\(^{-1}\)), then the energy decays at rate \( 2 r_l \) where \( r_l \) has units of Hz. If the total energy in the cavity is given by \( |a|^2 \), then the energy lost in one cycle is given by \( |a|^2 \cdot 2 r_l T \) where \( T = 1/f_o \), \( T \) is the period in seconds and \( f_o \) is the resonance frequency. Plugging this into Eq. 2.9 gives

\[
Q = \frac{\pi}{r_l T} = \frac{\pi f_o}{r_l} = \frac{\pi c}{r_l \lambda} = \frac{\omega_o}{2 r_l} \tag{2.10}
\]

where \( c \) is the speed of light in vacuum and \( \lambda \) is the wavelength of light in vacuum. Equation 2.10 can be put in terms of more commonly used values for losses such as the field absorption coefficient \( \alpha \) (cm\(^{-1}\)) and \( \kappa \), the complex part of the refractive index. We first need to find the relation between the field decay rate, \( r_l \), and the spatial decay coefficient, \( \alpha \) (m\(^{-1}\)). This is done by recognizing that in a given cycle, the energy in the resonator will have decayed by \( 2 r_l T |a|^2 \) joules. We can use the group velocity, \( v_g = c/n_g \) where \( n_g \) is the group index, to relate the spatial decay coefficient to the energy decay rate. In time \( t = T \), the wave travels a distance \( z(t = T) = v_g T \). The absorption
coefficient is used to calculate the fraction of power dissipated: $2\alpha z$ where $2\alpha$ is the fraction of power dissipated per unit length. We can now relate the energy decay rate to the absorption coefficient: $2r_l T = 2\alpha v_g T$ which gives

$$r_l = \frac{c}{n_g} \alpha = v_g \alpha$$  \hspace{1cm} (2.11)

Plugging Eq. 2.11 into Eq. 2.10 gives

$$Q = \frac{\pi n_g}{\lambda}$$  \hspace{1cm} (2.12)

Using the alternate definition of the quality factor, $Q = f_o/\delta f_{3dB}$, the 3-dB linewidth of the resonance can be calculated as

$$\delta f_{3dB} = \frac{c \alpha}{\pi n_g}$$  \hspace{1cm} (2.13)

When simulating resonators, a finite difference frequency domain (FDFD) mode solver can be used to solve Eq. 2.4 numerically. When solving for bent, leaky modes, the solver returns either a complex resonance frequency (if solving Eq. 2.4) or a complex angular propagation constant (if solving the waveguide equivalent) as the eigenvalue. Thus, it is useful to calculate the intrinsic quality factor as a function of these eigenvalues. The quality factor is given by

$$Q = \frac{\omega_R}{2\omega_I}$$  \hspace{1cm} (2.14)

$$Q = \frac{1}{2\gamma_I} \frac{\partial}{\partial k_0} (k_0 \gamma_R) \approx \frac{\gamma_R}{2\gamma_I}$$  \hspace{1cm} (2.15)

Where $\omega_R$ ($\gamma_R$) is the real part of the complex resonance frequency (angular propagation constant), and $\omega_I$ ($\gamma_I$) is the imaginary part of the complex resonance frequency (angular propagation constant). The approximation in Eq. 2.15 is correct if $n_g \approx n_{eff}$.

Table 2.1 summarizes the equations that can be used to calculate the quality factor and the 3 dB linewidth of a resonator given commonly used loss variables. The third column includes $\kappa$, the imaginary part of the refractive index (also called the “extinction coefficient”), which is related to $\alpha$ through $\kappa = \frac{\lambda \alpha}{4\pi}$. 
Table 2.1: Equations for quality factor, $Q$, and 3-dB linewidth, $\delta f_{3dB}$ of a resonator given common loss variables. $\omega_o$ is the angular frequency of light, $r_l$ is the loss decay rate, $\alpha$ is the field spatial decay, $\kappa$ is the imaginary part of the refractive index, $\omega$ is the complex resonance frequency, and $\gamma$ is the complex angular propagation constant.

<table>
<thead>
<tr>
<th>$Q$</th>
<th>$\frac{\omega_o}{2 r_l}$</th>
<th>$\frac{\pi n_g}{\lambda \alpha}$</th>
<th>$\frac{n_g}{4 \kappa}$</th>
<th>$\frac{\text{Re}(\omega_o)}{2 \text{Im}(\omega_o)}$</th>
<th>$Q = \frac{1}{2 \text{Im}(\gamma)} \frac{\partial}{\partial k_o} (k_o \text{Re}(\gamma)) \approx \frac{\text{Re}(\gamma)}{2 \text{Im}(\gamma)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta f_{3dB}$</td>
<td>$\frac{2 r_l}{2 \pi}$</td>
<td>$\frac{e \alpha}{\pi n_g}$</td>
<td>$\frac{4 f_o}{n_g} \kappa$</td>
<td>$\frac{\text{Im}(\omega_o)}{\pi}$</td>
<td>$\approx \frac{\text{Im}(\gamma)}{\pi}$</td>
</tr>
</tbody>
</table>

2.2.3 Perturbation of silicon refractive index due to free carriers

The perturbation on the refractive index of silicon caused by the density of free carriers forms the basis of many active devices in silicon. The carrier concentration due to implanted dopants can be precisely controlled in modern CMOS fabrication processes since there has been decades of research and development focused on optimizing carrier concentrations in transistors. This allows photonic devices to be built using fabrication technologies that are extremely well understood and matured by the evolution of microelectronics technology. For this reason, much research has gone into building active silicon photonic devices based on implanted dopants rather than other physical phenomena, such as the active devices based on the electro-optic effect (silicon has no electro-optic effect due to its centrosymmetric crystal structure).

A landmark study was done by Soref and Bennet \[123\] decades ago that studied the change perturbation to the refractive index of silicon as a function of electron and hole concentrations. The results of this study are widely used in designing carrier-plasma based active devices in silicon. The following equation summarizes the key result of \[123\], and Table 2.2 gives the value of the coefficients as presented in \[108\].

\[ -\delta n = -(\delta n_e + \delta n_h) = A N_e^B + C N_h^D \]
\[ \delta \alpha = \delta \alpha_e + \delta \alpha_h = E N_e^F + G N_h^H \]  

(2.16)  

where $\delta n_e$ and $\delta n_h$ are the perturbation to the refractive index due to electrons and holes, respectively, $\delta \alpha_e$ and $\delta \alpha_h$ are the power absorption coefficients in cm$^{-1}$ due to electrons and holes, and $N_e$
Equation 2.16 is plotted in Fig. 2.1 Since the carriers introduce absorption, any resonant cavity built using the p/n-type implants will have a limited quality factor. Equation 2.16 can be plugged into Eq. 2.12 to calculate the intrinsic quality factor for a given carrier concentration, and the resulting quality factors are plotted in Fig. 2.2.

The shift in resonance frequency as a function of carrier concentration can be calculated using Eq. 2.8 As an approximation, a constant value is assumed for the electric field inside the core of the waveguide and is assumed to be zero outside the core, and Eq. 2.8 is reduced to

$$\frac{\delta \omega}{\omega_0} = -\frac{1}{2} \frac{\delta \epsilon}{\epsilon_{core}} \approx -\frac{\delta n}{n_{core}}$$  \hspace{0.5cm} (2.17)

where $\delta n$ is the perturbation to the index of refraction, $n_{core} = \sqrt{\epsilon_{core}}$, and $\delta \epsilon \approx 2n\delta n$ was used. Now, several useful quantities can be calculated related to the impact of carriers on silicon resonators including the intrinsic quality factor, the intrinsic linewidth, and the resonance shift. These plots are shown in Fig. 2.3  Figure 2.3(a) shows the resonance frequency shift as a function carrier concentration calculated using Eqs. 2.17 and 2.16 Figure 2.3(b) shows the intrinsic linewidth of a cavity that is doped. Equation 2.13 was used to calculate the intrinsic linewidths. Finally, a useful figure of merit \[142\] is the ratio of the resonance shift to the intrinsic linewidth, which can be calculated from Figs. 2.3(a),(b), and it is shown in Fig. 2.3(c). This is a useful figure of merit because the resonance needs to shift on the order of a linewidth to achieve data modulation. As seen, in Fig. 2.3(c), holes have a larger resonance shift for a given loss compared to electrons. It is for this reason that designers typically choose to favor larger p-type regions in modulators.

<table>
<thead>
<tr>
<th>$\lambda$ (nm)</th>
<th>$A$ (cm$^3$)</th>
<th>$B$</th>
<th>$C$ (cm$^3$)</th>
<th>$D$</th>
<th>$E$ (cm$^2$)</th>
<th>$F$</th>
<th>$G$ (cm$^2$)</th>
<th>$H$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1300</td>
<td>$6.2 \times 10^{-22}$</td>
<td>1</td>
<td>$6.0 \times 10^{-18}$</td>
<td>0.8</td>
<td>$6.0 \times 10^{-18}$</td>
<td>1</td>
<td>$4.0 \times 10^{-18}$</td>
<td>1</td>
</tr>
<tr>
<td>1550</td>
<td>$8.8 \times 10^{-22}$</td>
<td>1</td>
<td>$8.5 \times 10^{-18}$</td>
<td>1</td>
<td>$8.5 \times 10^{-18}$</td>
<td>1</td>
<td>$6.0 \times 10^{-18}$</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2.2: Values for constants in Eq. 2.16 for $\lambda = 1300$ nm and $\lambda = 1550$ nm.

and $N_h$ are the free electron and hole concentrations in cm$^{-3}$. The values for $A, B, C, D, E, F, G, H$ are given in Table 2.2 $A, C$ have units of cm$^{-3}$, $E, G$ have units of cm$^{-2}$, and $B, D, F, H$ are unitless exponents.
as opposed to n-type. The p-type carriers have a reduced electrical mobility compared to the n-type carriers which affects the RC limited electrical switching speed, but the achievable resistances support GHz switching speeds in most designs. The next section presents the basic operation of microring depletion-mode modulators.
2.3 Basic operation of depletion-mode microring modulators

This section presents an overview of depletion-mode modulators and their basic operation. Figure 2.4 gives a simplified overview of depletion-mode modulators and their use in on-off keying (OOK) modulation schemes.

We will use a coupling of modes in time (CMT) model [65] to analyze the behavior of microring modulators. Figure 2.4(a) shows the schematic view of a microring modulator. The dynamics are described by

\[ \frac{d}{dt} a = j \left[ \omega_0 + j \left( r_i + r_l \right) \right] a - j \sqrt{2 r_i} s_+ \]

\[ s_- = s_+ - j \sqrt{2 r_i} a \]

where \( a \) is the energy amplitude in the resonator normalized such that \( |a|^2 \) gives the energy, \( s_+ \), \( s_- \) are the amplitudes of the input and output waves, respectively, with \( |s|^2 \) giving the power, \( r_i \) is the input amplitude coupling rate, \( r_l \) is the amplitude decay rate due to losses, and \( \omega_0 \) is the resonance frequency of the cavity, which depends on the carrier concentration. To find the frequency response
of the ring, we let $d/dt \to j\omega$ and solve for $a$.

$$a = \frac{-\sqrt{2}r_i}{\delta\omega - j(r_i + r_l)} s_+$$

(2.20)

where $\delta\omega = \omega - \omega_0$. Plugging this equation into Eq. 2.19 to solve for the through port response of the modulator gives

$$\frac{|s_-|^2}{s_+^2} = 1 - \frac{4r_i r_l}{\delta\omega^2 + (r_i + r_l)^2} = \frac{\delta\omega^2 + (r_i - r_l)^2}{\delta\omega^2 + (r_i + r_l)^2}$$

(2.21)

Equation 2.21 is recognized as an inverted Lorentzian lineshape, and the critical coupling condition [46] is seen to occur when the input coupling rate matches the loss decay rate, $r_i = r_l$. At critical coupling when on resonance, all of the input light is absorbed or radiated in the cavity, and there

Figure 2.3: (a) Resonance frequency shift versus electron and hole concentration for $\lambda = 1300$ nm and $\lambda = 1550$ nm. (b) Intrinsic 3 dB linewidth of resonance. (c) Ratio of resonance shift and 3 dB linewidth which is a useful figure of merit.
To shift the resonance of the cavity, p-n junctions are integrated into the ring and can be electrically contacted. Figure 2.4(b) shows the most used p-n junction configurations in microring modulators including lateral [147, 149], vertical [131, 146], and interleaved junction configurations [119]. When a bias voltage is applied to the device (contacts not shown here), the carrier concentration changes as shown in Fig. 2.4(c) (polarity of applied voltage indicated with +/−). A change in the carrier concentration causes a refractive index change, which causes a resonance frequency shift as shown in Fig. 2.4(d). If a continuous-wave (cw) laser is tuned to be resonant with the device, electro-optic signal conversion is achieved by changing the resonance frequency of

Figure 2.4: (a) Schematic of microring model used for CMT equations. (b) p-n junction configurations for depletion-mode modulators. (c) Cross-sectional view of reverse-bias state showing index and absorption perturbation. The gray regions are depleted of free carriers. (d) Illustration of resonance shift with applied voltage. Carriers blue-shift the resonance. A cw laser frequency is drawn on the graph to show how on-off keying is achieved. (e) Resulting amplitude modulation of optical power at the output of the microring modulator.
the cavity which varies the attenuation that the input laser light experiences and achieves on-off keying (OOK) modulation. The output light is amplitude modulated as illustrated in Fig. 2.4(e). For such modulation, several performance characteristics can now be defined. The extinction ratio (E.R.) is typically reported in dB, and is defined as

$$E.R. = 10 \log_{10} \left( \frac{P_1}{P_0} \right)$$

where $P_1$ is the optical power corresponding to a “1” bit and $P_0$ is the optical power corresponding to a “0” bit. Since the input laser frequency is typically still within the Lorentzian resonance even for an optical “1”, there is a power penalty commonly referred to as the insertion loss. The insertion loss (I.L.) is defined as

$$I.L. = 10 \log_{10} \left( \frac{P_{\text{off}}}{P_1} \right)$$

where $P_{\text{off}}$ is defined as the optical power when detuned far from the resonance as seen in Fig. 2.4(e). This is a good approximation to the total I.L. of the modulator because off-resonance, the transmission loss of a high-Q resonator is small.

### 2.4 Designing waveguides and ring resonators in thin-SOI CMOS

The majority of silicon photonic devices and systems today are designed in 220 nm thick silicon-on-insulator (SOI) processes [24]. Device geometries such as single-mode widths and radii for bends and resonators greatly depend on the thickness of the silicon. In advanced microelectronics process, thin-SOI with silicon thicknesses < 100 nm emerged to enhance transistor performance [39]. The thin silicon requires waveguide and resonator geometries that are substantially different than “traditional” silicon photonics. Furthermore, our main electronic-photonic systems require a non-standard wavelength of 1180 nm as opposed to 1310 and 1550 nm (because of availability of detectors in the unmodified CMOS platform). The thin silicon and the non-standard wavelength cause our device geometries to deviate from commonly found geometries in literature. In this section, guidelines for waveguide geometries and resonator geometries are presented for $\lambda = 1180$, 1310, and 1550 nm. Although devices and systems are mostly demonstrated at 1180 nm and near 1310 nm,
1550 nm designs are included since it is the wavelength with which the silicon photonics community is most familiar.

Figure 2.5 gives an overview of the bound modes for straight waveguides simulated with a FDFD mode solver\(^1\) [94]. When designing single-mode waveguides, the designer usually chooses the largest waveguide width just before the second-order transverse mode is guided. This is to have the largest confinement of the mode in the waveguide which reduces sidewall scattering loss and generally lower ring-waveguide coupling losses [96]. A slightly narrower width ensures single-mode operation when considering fabrication errors and over a wide wavelength range. Figure 2.5(a)-(c) shows the effective index of several transverse modes. The choice of the width depends on the wavelength, and the selected widths are indicated by blue dots. The electric fields of bound TE\(_1\) modes of the chosen widths are shown in Fig. 2.5(d)-(f) with 0.1\(E_x^{max}\) to 0.9\(E_x^{max}\) contours shown in steps of 0.1\(E_x^{max}\). Over the widths simulated, there are no TM-like modes guided, and the TE\(_3\) mode is also unbound. Although single-mode waveguides comprise the vast majority of on-chip photonic devices, there are many devices that rely on multi-mode interference [67, 68, 120]. For such devices, larger widths can be simulated to find the effective indexes of the desired high-order modes.

When designing resonant structures, the intrinsic quality factor (related to the intrinsic linewidth and finesse) is one of the most important properties of the device. The quality factor affects the insertion loss in filters, the energy efficiency of modulators, and many other linear and nonlinear applications. Figure 2.6 shows the radiation limited quality factor (i.e. the quality factor due to bending loss) of microrings with varying widths and radii and at three different wavelengths. The FDFD mode solver in cylindrical coordinates calculates a complex angular propagation constant, \(\gamma\), as the eigenvalue; the real part of \(\gamma\) is an integer if a resonance condition is enforced. The quality factor is calculated using Eq. 2.15. In the 45 nm process, we have measured intrinsic quality factors of 227,000 and 112,000 at \(\lambda = 1280\) nm and \(\lambda = 1550\) nm, respectively [88].

---

\(^1\) The mode solver was written by Miloš Popović and is described in [94]. It is currently maintained by the Nanophotonic Systems Laboratory research group (P.I. Miloš Popović). The mode solver is used many times throughout this thesis.
Figure 2.5: Effective index of waveguide transverse modes at $\lambda = 1180, 1310, 1550$ nm. Single-mode widths are indicated by the blue dots, and $E_x$, the x-component of the electric field, contours are shown below, overlapped with the waveguide cross-section. The widths are 380 nm, 460 nm, and 700 nm for $\lambda=1180$, 1310, and 1550 nm, respectively.

Typically, the radiation limited quality factor is chosen to have negligible impact on the resulting overall quality factor. Since doping or target bandwidths will limit the total Q to $\approx 10k$ to 30k, we choose microring geometries with $Q_{rad} > 10^6$. The choice of $Q_{rad} = 10^6$ allows the devices to be robust against fabrication variance (i.e. thinner than expected silicon layers will reduce $Q_{rad}$); $Q_{rad}$ could be chosen more aggressively (e.g. $10^5$) if the statistical variation of the silicon layer thickness is known\textsuperscript{2}. These geometries are indicated by red dots in Fig. 2.6.

When designing devices with coupled waveguides, a designer usually has a photonic circuit model of their device/system (e.g. coupled mode theory, transfer matrix models, etc.) that abstracts away the geometry and represents it as a coupling coefficient. Once the designer knows what coupling coefficient values are needed for their device/system, the coupling coefficient can be mapped onto a physical device. This mapping is achieved by simulating the power coupling between waveguides as a function of the gap between the waveguides. In low-index contrast devices,\textsuperscript{2} Currently, this information is known by the foundry, but it is proprietary information and is not disclosed to customer designers.
Figure 2.6: Radiation limited quality factor of fundamental TE-like transverse mode as a function of ring geometry (width and radius) for (a) $\lambda = 1180$ nm, (b) $\lambda = 1310$ nm, (c) $\lambda = 1550$ nm. Our choices for specific geometries are indicated by the red dot, and the contour plots of the $E_x$ component of the corresponding bent modes are plotted in (d)-(f). The selected widths ($w$) and outer radii ($R$) are $w = 460, 560, 700$ nm $R = 4, 5, 8 \mu$m for $\lambda = 1180, 1310, 1550$, respectively. These dimensions are chosen to achieve $Q > 10^6$, but the exact dimensions of a microring depend on specifications such as the total $Q$ and free spectral range required for the application.

coupling coefficients can be accurately calculated using perturbation techniques (e.g. coupling of modes in space), but in high-index contrast devices, such as silicon surrounded by silicon dioxide, perturbation calculations are inaccurate. Therefore, 3D FDTD simulations are used to accurately simulate the coupling coefficients \[96\]. Figure 2.7 gives an overview of the most common simulation configurations to simulate coupling between waveguides. Figure 2.7(a) shows the simulation domain for simulating the coupling coefficient between a ring and a straight waveguide (referred to as a bus). The fundamental mode of the ring waveguide is excited with power $P_{in}$, and the wave propagates around the ring. When the ring waveguide is in close proximity to the straight waveguide, a fraction of the optical power, $P_{drop}$, is coupled into the fundamental mode of the straight waveguide. The remaining power, $P_{thru}$, continues in the ring waveguide. There is a small
amount of coupling loss given by $P_{\text{loss}}$. Figure 2.7(c) shows the configuration for simulating the coupling coefficient between two rings, and Figs. 2.7(b),(d) show field snapshots from the FDTD simulations with a pulse excitation. The following equations define the power coupling coefficient, $\kappa^2$, and $P_{\text{loss}}$.

\[ P_{\text{drop}} = \kappa^2 P_{\text{in}} \] (2.24)

\[ P_{\text{loss}} = P_{\text{in}} - P_{\text{thru}} - P_{\text{drop}} \] (2.25)

In FDTD simulations, there is some numerical loss due to artificial roughness introduced in curves by the discretized Cartesian grid. This can be taken into account by simulating the transmission of a portion of a single ring (with no coupling to other waveguides) and normalizing Eq. 2.25 to correct for the numerical loss [96]. If $T_{\text{prop}}$ is the power transmission after propagating around 1/4 of the ring, then Eq. 2.25 becomes

\[ P_{\text{loss}} = P_{\text{in}} - \frac{P_{\text{thru}}}{T_{\text{prop}}^2} \frac{P_{\text{drop}}}{T_{\text{prop}}} \] (2.26)

Figure 2.8 summarizes the power coupling simulation results. A polynomial fit is used to map arbitrary gaps to their power coupling. The power coupling has a well-behaved exponential
behavior versus gap with a slight curvature (on a logarithmic scale) for ring-ring coupling. The polynomial fit is applied to the natural logarithm of $\kappa^2$ such that

$$\kappa^2(g) = \exp (P_2 g^2 + P_1 g + P_0)$$

where $\kappa^2$ is the power coupling coefficient, $g$ is the gap between waveguides, and $P_2, P_1, P_0$ are constant coefficients from the polynomial fit.

Table 2.3 gives the values for the polynomial fits corresponding to ring-bus simulations in Fig. 2.8 and Table 2.4 gives the values for the polynomial fits corresponding to ring-ring simulations.
Table 2.3: Values for polynomial fit coefficients corresponding to ring-bus simulations in Fig. 2.8. The gap is in units of µm.

<table>
<thead>
<tr>
<th>λ (nm)</th>
<th>$P_2$</th>
<th>$P_1$</th>
<th>$P_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1180</td>
<td>$\kappa^2$</td>
<td>0.24</td>
<td>−13.64</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{loss}}$</td>
<td>n/a</td>
<td>−6.18</td>
</tr>
<tr>
<td>1310</td>
<td>$\kappa^2$</td>
<td>0.14</td>
<td>−11.67</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{loss}}$</td>
<td>n/a</td>
<td>−6.90</td>
</tr>
<tr>
<td>1550</td>
<td>$\kappa^2$</td>
<td>0.20</td>
<td>−9.82</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{loss}}$</td>
<td>n/a</td>
<td>−5.11</td>
</tr>
</tbody>
</table>

Table 2.4: Values for polynomial fit coefficients corresponding to ring-ring simulations in Fig. 2.8. The gap is in units of µm.

<table>
<thead>
<tr>
<th>λ (nm)</th>
<th>$P_2$</th>
<th>$P_1$</th>
<th>$P_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1180</td>
<td>$\kappa^2$</td>
<td>2.23</td>
<td>−14.09</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{loss}}$</td>
<td>n/a</td>
<td>−9.23</td>
</tr>
<tr>
<td>1310</td>
<td>$\kappa^2$</td>
<td>1.95</td>
<td>−11.96</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{loss}}$</td>
<td>n/a</td>
<td>−8.01</td>
</tr>
<tr>
<td>1550</td>
<td>$\kappa^2$</td>
<td>1.10</td>
<td>−8.85</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{loss}}$</td>
<td>n/a</td>
<td>−6.97</td>
</tr>
</tbody>
</table>

in Fig. 2.8. The geometries and coupling coefficients presented in this chapter can be used to design arbitrary coupled-ring systems and give representative geometries for each wavelength of interest.

2.5 Modulator design: fitting the constraints of the 45 nm CMOS process

The 12SOI process has a transistor body silicon layer thickness that is <100 nm whereas typical custom SOI wafers used for silicon photonics have a device layer thickness of >200 nm. Since the process does not have a partial etch step for the silicon device layer, ridge waveguide structures that are often used to create optical modulators [149] cannot be made out of the crystalline body silicon. The polysilicon transistor gate, also sub-100 nm in thickness, can be patterned atop a crystalline body silicon slab to construct a ridge-like waveguide and was previously shown to
enable injection-mode microring modulators [88]. However, the polysilicon in 12SOI is highly lossy optically (> 50 dB/cm propagation loss in a polysilicon only waveguide) which translates to broad linewidth resonances. This in turn requires injection-mode operation to get sufficient shift and thus lowers the energy efficiency and speed of modulation limited by the minority carrier lifetime. Since the polysilicon is unusable for applications requiring low optical loss, the optical devices must be made entirely in the thin crystalline body silicon layer. In an SOI process, the crystalline silicon forms the active region (source/drain and channel) of the electrical transistors. In our approach, electrical circuits and photonic circuits are made in the same physical layer, and they can be tightly integrated on the same silicon die. Working in thin SOI introduces another potentially significant design constraint when designing active photonic components: the set of doping profiles native to the process were designed for transistor performance. This limits the available options for doped optical waveguides and cavities to enable optoelectronic devices. We show that native doping profiles enable efficient optoelectronic devices, eliminating the need for process customization. Finally, energy efficient depletion-mode modulators require a significant fraction of the cavity volume to be depleted of carriers. One efficient modulator geometry available in custom photonics processes is a vertical junction modulator [146]. In thin SOI CMOS, the active layer is too thin to support traditional vertical junctions. As a result, lateral junctions with one or more of the dopant concentrations optimized for transistors are what must be used. Recently, a depletion modulator based on a wiggler-mode racetrack microcavity, which is compatible with a single-lithographic layer and low-resolution junctions, was demonstrated in bulk CMOS [120]. Some disadvantages of this resonator are its limited Q demonstrated so far, dead regions that ensure imperfect optical field overlap with the junction, and large cavity size which increases junction capacitance and energy consumption. In the following, we introduce an optical cavity that is compatible with SOI CMOS design constraints described above, and that maximally employs the high mask resolution for dopants made available in a 45 nm CMOS process. By enabling implementation in unmodified SOI CMOS, many active optoelectronic applications are achievable.

Figure 2.9 shows a schematic representation of what we have termed the “spoked-ring” mi-
Figure 2.9: Schematic of the spoked-ring microcavity: (a) passive cavity; (b) formation of interleaved junctions for a modulator, with inner radius vias; (c) extended interaction length input coupler to excite only the fundamental mode. (Illustration: M. Popović)

crocavity. The width of the ring is wider than the single-mode width, so multiple higher order spatial modes are guided. However, the higher order modes have significant electric field on the inner portion of the ring which has “spoked” electrical contacts that result in high optical loss for those modes, suppressing them by “spoiling” their Q. When designed properly, the fundamental mode is guided as a whispering-gallery mode, similar to what happens in microdisks, and the field does not extend to the inner portion of the ring. Since the fundamental mode does not significantly scatter off the contacts, high-Q resonances are maintained. However, the next few modes may also have appreciable Q, but they are left unexcited by the coupling geometry.

One use for the cavity is to create depletion-mode carrier plasma modulators for efficient communication links. Figure 2.9(b) shows how a depletion modulator can be implemented with this cavity. Lateral junctions are created using interleaved radially extending p and n doped sections. As previously stated, only dopant concentrations that already exist in the process can be utilized. Fortunately, the transistor n-well and p-well dopants have concentrations of $\approx 5 \times 10^{17} \text{ cm}^{-3}$ which is sufficient to make modulators and other active devices without accruing too much optical loss. Ideally, the interleaved junctions would be narrow enough to fully deplete the active regions under reverse bias, thereby introducing the maximum perturbation to the refractive index and achieving the maximum resonance shift for a given voltage. In practice, the design rules for the dopant masks
limit the minimum width of the p and n doped sections of the waveguide. For 12SOI, the minimum width resolution is \( \approx 250 \text{ nm} \). This resolution allows the creation of narrow p and n regions, and a typical microring can include >80 junctions. The widest depletion width at reverse bias in typical operation is >150 nm, so an efficient design can be obtained, where over half the cavity sees active carrier density modulation. Specialty circuits can be designed to provide larger reverse bias voltages (\( \approx 2 \text{ V} \)) to deplete a larger fraction of the cavity at the expense of larger power consumption. The fraction of the cavity that is depleted can also be increased by reducing the minimum feature size of the dopant masks, but this requires modification to the fabrication process.

Since the cavity is multimode, the optical input coupling of light from the waveguide to the cavity must be carefully designed to avoid excitation of higher order modes [48,119]. This is achieved by having a long interaction length input coupler which phase-matches the input coupler with the angular propagation constant of the spoked-rings fundamental mode. The extended interaction length input coupler allows an input excitation in one waveguide to create a long polarization current in the adjacent perturbing waveguide. The spatially extended polarization current then has a small spread in \( k \)-space, and thus it selectively excites only the fundamental mode, to which it is well matched in \( k \)-space (i.e. in propagation constant). If an output coupler is needed, the extended coupler does not need to be used since only the fundamental mode is excited in the cavity, and the coupling is approximately single-mode to single-mode. Leaving the resonator, the higher-order-mode excitation is a second order coupling effect, and typically negligible.

The spoked-ring cavity geometry meets all of the constraints of a thin-SOI microelectronics process, and can be seamlessly integrated into existing SOI CMOS processes. In the next section, a design flow for the spoked ring cavity is presented.

2.5.1 Resonance shift of interleaved junction

In this section, simple equations are derived to predict the expected performance of the microring modulator based on the interleaved junctions. We begin with Eq. 2.8 and make a plane-wave approximation. This assumes a uniform electric field throughout the dielectric core of the
waveguide (neglecting the mode shape), and the volume integral is computed over the core of the microring cavity. Equation 2.8 is reduced to
\[
\frac{\delta \omega}{\omega_o} = -\frac{1}{2} \frac{\delta \epsilon}{\epsilon_o}
\] (2.29)
where \(\delta \epsilon\) is the volume permittivity perturbation to the silicon introduced by carriers, and \(\epsilon\) is the permittivity of silicon. Using \(\delta \epsilon/\epsilon \approx 2 \delta n/n\), Eq. 2.29 is converted to an equation in terms of the index of refraction
\[
\frac{\delta \omega}{\omega_o} = -\frac{\delta n}{n_o}
\] (2.30)
where \(\delta n\) can now be calculated using Eq. 2.16 and \(n\) is the refractive index of silicon.

Next, equations for calculating \(\delta n\) are derived. A single unit cell of the interleaved junction cavity is considered to calculate \(\delta n\). In a unit cell, the index perturbation only occurs in a fraction of the volume that has the carriers depleted under applied voltage. Thus, the index perturbation is given by
\[
\delta n = \frac{(\delta n_e x_n + \delta n_h x_p) \ w \ t}{L \ w \ t}
\] (2.31)
where \(x_n, x_p\) are the n-type and p-type depletion widths, respectively, \(w\) is the width of the microring cavity, \(t\) is the thickness of the microring cavity, and \(L\) is the period of the unit cell; \(w\) and \(t\) cancel from Eq. 2.31 resulting in
\[
\delta n = \frac{(\delta n_e x_n + \delta n_h x_p)}{L}
\] (2.32)
showing that \(\delta n\) is proportional to the ratio of the depletion widths to the period of the unit cell, or equivalently, the fraction of the unit cell that is depleted under applied voltage.

To calculate the depletion widths, the standard electrostatic approximation is used [134], and the p-type and n-type depletion widths are
\[
x_p = \sqrt{\frac{2 \epsilon \ N_e \ \Delta V}{q \ N_h \ N_e + N_h}}
\] (2.33)
\[
x_n = \sqrt{\frac{2 \epsilon \ N_h \ \Delta V}{q \ N_e \ N_e + N_h}}
\] (2.34)
where $\epsilon_0$ is the permittivity of silicon, $q$ is the electron charge, $N_e$ and $N_h$ are the electron and hole concentrations (cm$^{-3}$), respectively, and $\Delta V \equiv V_{bi} - V_{applied}$ is the applied reverse bias voltage relative to the built-in potential. The full depletion width is $w_d = x_p + x_n$. The resonance shift can now be written as

$$\delta \omega = -\frac{\omega_0}{n_o} \frac{\delta n_e x_n + \delta n_h x_p}{L}$$

(2.35)

Figure 2.10(a) shows the calculated resonance shift using Eqs. 2.16, 2.33–2.35. The unit cell periodicity, $L$, is set to 500 nm which corresponds to the the unit cell used in the experimental devices presented later, and the voltage swing is $\Delta V = V_{bi} - V_{applied} = 1$ V which is approximately the achievable swing using on-chip circuits. The resonance shift is calculated for electron and hole concentrations ranging from $10^{17}$ to $10^{18}$ cm$^{-3}$ since that is the range of transistor well implant concentrations found in the 45 nm process and used to build the modulator. The ratio of the resonance shift to the intrinsic linewidth, a useful figure-of-merit that can be used to calculate various performance metrics [142], is shown in Fig. 2.10(b) and calculated using Eqs. 2.16, 2.33–2.35 and Table 2.1. An intrinsic quality factor of 100,000 is assumed which is consistent with measured intrinsic quality factors [88]. Figure 2.10 shows that transistor well implant concentrations are well-suited for building modulators, and the resonance shift with 1 V applied bias voltage is 5-7 GHz.

2.5.2 Design flow of spoked-ring cavity

The design of the spoked-ring cavity involves a combination of mode solver and FDTD simulations. First, a 2D cylindrical mode solver [94] is used to simulate the geometry of the ring. At this point, a disk supporting whispering-gallery modes is used since the fundamental mode of the disk will be nearly identical to the fundamental mode of the final structure. The bending loss versus radius is calculated to ensure the radiation limited intrinsic Q is high enough to not degrade the performance of the device. Next, a second waveguide, the input waveguide, is included in the cylindrical geometry mode solver simulations. As with any coupled waveguides, the supermodes of the two-waveguide system are analogous to symmetric and antisymmetric modes, with some defor-
Figure 2.10: Resonance shifts for $\lambda = 1180$ nm. (a) Resonance shift with 1 V applied electrical reverse bias, relative to the built-in potential, for interleaved junction. A unit cell periodicity of 500 nm is used as is achievable in the 45 nm process. (b) Ratio of resonance shift to intrinsic linewidth. An intrinsic (passive, undoped) quality factor of 100,000 is assumed, and the group index is $n_g \approx 3.0$. 

mation because of the curved geometry. With coupled waveguides, the coupling between the two waveguides is efficient only if the propagation constants are matched. The propagation constants of the “symmetric” and “antisymmetric” supermodes differ by an amount equal to the coupling between the waveguides when the waveguides have degenerate uncoupled propagation constants. Here, the complicating factor is that each waveguide sees a shift in its own uncoupled propagation constant due to presence of the other waveguide [100]. In a pair of straight, symmetric waveguides, the two waveguides see the same self-shift, so a pair of identical waveguides stays matched. For bent waveguides, the shifts are different. With bent waveguides, the angular propagation constants should match to achieve efficient coupling, and the matching condition is more subtle in the cylindrical geometry. First, one must look for dimensions of each of the disk and the wrapping waveguide that give the same propagation constant in isolation. However, the coupling shifts the propagation constants of the disk and the waveguide which changes the matching of the angular propagation constants. Thus, one needs to correctly simulate the coupled system to ensure that
phase matching has been achieved in the coupled geometry. This could be done using coupled mode theory approximately, but is more directly addressed by solving the coupled system in the mode solver. The angular propagation constant of the supermodes is plotted versus the gap between the microdisk and the input waveguide, while keeping the radii of the disk and the input waveguide constant, in Fig. 2.11(b), and the anti-crossing shows a region of gaps (≈ 250 to 350 nm) at which the angular propagation constants are well-matched and efficient single-mode excitation occurs. To determine the range of geometries that result in phase-matching and efficient coupling, similar plots are generated that vary the radii and keep the gap constant. These plots are used to guide the geometries to simulate in (time-consuming) 3D FDTD simulations.

Figure 2.11: Supermodes of disk/waveguide system. Spatial electric field mode profiles of the (a) symmetric (top) and (b) anti-symmetric (bottom) supermodes. (c) Angular propagation constant anti-crossing between symmetric and anti-symmetric supermodes. This splitting occurs when there is efficient coupling between the disk and the waveguide indicating phase-matching. (Original simulation by Jeff Shainline. Modified for use here.)

Next, 3D finite-difference time-domain (FDTD) simulations are run to verify the results from the mode solver simulations and to design the position of the inner sidewall contacts. To ensure extra loss is not being introduced by the electrical contacts, the fundamental mode is launched into the simulation domain and propagates around the spoked-ring. The width of the ring is varied to increase the distance between the optical mode and the inner sidewall contacts, and the transmission from input to output is calculated. This simulation is illustrated in Fig. 2.12. Figure 2.12(a) shows
that the spokes do not significantly scatter the fundamental mode, and Fig. 2.12(b) shows the quality factor of the resonator as a function of ring width. As expected, the Q decreases as the contacts get closer to the optical mode. For some applications, the designer might choose to be more aggressive with the location of the spokes. For instance, in a modulator configuration, the width of the spoked-ring cavity, \( w \), directly corresponds to the series resistance and the junction capacitance of the diodes. If the spoked-ring cavity width is narrowed, both the resistance and capacitance are decreased which decreases the RC time constant of the diode. At some point, however, the contacts will introduce loss to the fundamental mode and broaden the resonance. Thus, there is a tradeoff between energy efficiency (related to the linewidth of the resonance) and the electrical speed of the device.

![Spoked-ring width versus Q](image)

**Figure 2.12:** Spoked-ring width versus Q. (a) 3D FDTD simulation showing the fundamental mode avoiding the contacts. (b) Radiation quality factor versus the width of the ring (outer radius fixed at 5\( \mu \)m). As the width decreases, the fundamental mode scatters more light and the Q degrades. Narrowing the ring leads to reduced RC time constants for driving the reverse-biased junctions from the electrical contacts at the inner radius, at the cost of broader linewidth resonances. (Original simulation by Jeff Shainline. Modified for use here.)

Figures 2.13(a),(b) show snapshots of FDTD simulations comparing a point coupler that excites higher order modes and an extended coupler that only excites the fundamental mode. Figure 2.13(c) shows a more compact layout implementation to achieve the phased-matched coupling.
As confirmed by the figures, the extended coupler is critical to achieving a low-loss efficient coupler. To adjust the amount of power coupled from the input waveguide to the spoked-ring cavity, the interaction length can be extended or the gap between the spoked-ring and the input waveguide can be adjusted. Although changing the gap affects the phase matching, FDTD simulations showed that the effect is small for perturbations around a particular design. Thus to arrive at design variants, we adjust the coupling strength with the gap alone which simplifies the layout. Figure 2.13(d) shows the power coupling coefficient to the fundamental mode and the second order mode as a function of the gap between the spoked-ring cavity and the input waveguide, for the design in Fig. 2.13(c). For typical coupling strengths per unit length, i.e. typical gaps, the total coupling to the fundamental mode is almost two orders of magnitude higher than the unwanted coupling to the second order mode.

Once the cavity is designed, the next step is to implement it in a standard microelectronics CAD flow for this process, as described in the next section.

2.5.3 CAD flow integration

Advanced CMOS processes were developed entirely for electronics. This includes the complex tools that are used to design, synthesize, lay out, and verify electronic circuits. For a given process, the designer receives a process design kit (PDK) that contains details of many aspects of the process including transistor models, front end (silicon layers) and back end (metal layers and dielectrics) information, design rules, fabrication mask sets, and the process flow. The entire PDK has been written and developed to be as convenient as possible for circuit designers. This infrastructure presents challenges and needs to be adapted for making customized structures like photonics.

To design an electronic-photonic chip in an advanced process, one must use the existing toolset that has been developed for the electronics industry. Since the IBM 12SOI process is an electronics only process, its PDK provides no guidance on how to make optical structures, and making simple structures is obfuscated by the steps involved in turning a layout (which uses design layers) into a mask set. For example, if one simply draws a rectangle of body silicon in the layout,
Figure 2.13: FDTD simulations showing various coupling schemes to the spoked-ring cavity for $\lambda = 1280\ nm$. (a) Point coupling scheme where the input coupling waveguide is straight and is brought in close proximity to the spoked-ring microcavity. This approach excites higher-order modes and is lossy. (b) Phase-matched approach. The input coupling waveguide has a radius of curvature larger than that of the spoked-ring cavity. This approach efficiently excites the fundamental mode and does not introduce significant loss. (c) Sinusoidal-coupler approach with 1.2$\mu$m ring width. This method still achieves phase matching between the input waveguide and the spoked-ring cavity, and it minimizes straight to bend loss by slowly changing the radius of curvature along the input waveguide. It is also simpler to implement in a layout. (d) Power coupling from bus to ring vs. gap from the FDTD simulation shown in part (c). Coupling to the 1st order mode is about two orders of magnitude stronger than coupling to the 2nd order mode over the range of relevant coupling coefficients. (Original simulation by Jeff Shainline. Modified for use here.)

the resulting structure that is fabricated in the 12SOI process will be heavily doped by default, because it is presumed to be part of a transistor. For photonics, additional dopant blocking shapes are added on other design levels. Many photonic structures involve smooth bends which introduce numerous design rule violations, since the design rule set was developed for electronics which are typically comprised of rectangular shapes. In our early attempts at implementing photonics in advanced CMOS, in this work and e.g. [86–88], close communication with foundry engineers was required to ensure that apparent design rule violations that arise in photonics are artifacts of the electronics oriented PDK and do not introduce manufacturability problems. As silicon photonics becomes more prevalent, it is likely that the advanced processes will officially support photonics in
their PDK, and many of these difficulties will be alleviated.

Figure 2.14: Layout of spoked-ring cavity and modulator. (a) Passive body silicon structure showing the silicon spokes that will be used for electrical contact. The inner circular silicon structure is an integrated heater for thermal tuning. (b) Layout view of the doping shapes to create a carrier-plasma optical modulator. The p/n regions are discretized to ≈200 nm grid to meet design rules imposed by the process. The silicon heater is blanket doped with a high concentration n-type dopant to create a low resistance heater that can be driven by on-chip circuits. (c) Zoom in of the modulator showing a close up of the p/n regions and the electrical wiring. The silicon “spokes” are covered with high concentration dopants and silicide to create highly conductive contacts to the p-n junctions. Note that the width and pitch of the metal layers are of critical importance to allow the tightly packed wiring.

Figure 2.14 shows a spoked-ring modulator layout. Figure 2.14(a) shows the body silicon structure. This includes the spoked electrical contacts as well as an electrically contacted circular strip of silicon that is doped and used for thermal tuning. The mask shapes that form the p/n regions are shown in Figs. 2.14(b),(c). Figure 2.14(c) shows a close-up of the p/n junctions and the wiring between the many p/n regions to connect all p shapes together and all n shapes together.

2.5.4 Photonics CAD in CMOS design flows

At the time this work was conducted, there were no suitable commercial CAD tools specifically designed for integrating photonic devices and systems with VLSI electronics. Therefore, we had to create our own custom tools and become compatible with existing electronics CAD software. This mainly involved creating a code base of custom photonics layouts\(^3\) which formed the basis

\(^3\) The first generation of this code base was almost solely developed by Jason Orcutt and is covered in \(^{90}\), and the second generation of the code base is covered in \(^{4}\) which greatly simplifies the end-user experience.
of a component library. The layouts can be parameterized (so-called “p-cells”) and dynamically change depending on parameters passed to them (e.g. the radius of a ring would be a parameter of a microring resonator).

A typical design flow for implementing photonics in a CMOS process is shown in Fig. 2.15. The designer starts with photonics design which includes electromagnetic simulations, photonic circuit models, etc. Next, the designer uses the code base previously mentioned to implement their device as a physical layout. This requires the designer to physically draw shapes that correspond to the desired device. Once the physical layout is complete, commercial software is used to make sure the layout is compliant with the design rules imposed by the foundry (minimum width of shapes, minimum gap between shapes, etc.). If the layout fails DRC, the designer might have to return to the very beginning since the constraints imposed by DRC can potentially affect the photonic device dimensions, thus, requiring the device to be re-simulated. This loop can be extremely time-consuming (days/weeks) to get layouts “DRC clean.”

Once the layout is compliant with DRC, the next step is verification. In this step, especially when connected to on-chip electronics, the connectivity is verified. Verification algorithms compare a schematic with intended electrical connections to the actual connections made in the physical layout. This comparison catches simple mistakes that frequently occur (wires shorted, connections backwards, etc.). In future iterations, once physical layouts have been characterized well enough to develop compact models, verification will also include simulating the connected system (both electronic and photonic) to ensure the desired functionality is achieved (data rate, channel cross-talk, energy consumption, etc.). After passing verification, the layout is complete and ready to be submitted or integrated into a larger layout.

Figure 2.15: Typical design flow for integrating photonics into a CMOS electronics design flow.
2.6 Experimental demonstrations

In this section, two demonstrations of high-performance microring-based active devices are presented, a depletion-mode modulator and an efficient thermally tunable filter\(^4\). These devices and results are the first demonstrations that showed high-performance (i.e. comparable with state-of-the-art photonics in fully customized processes) photonic devices in an advanced CMOS micro-electronics process. The implication is that existing high-volume CMOS fabrication foundries can be used to build integrated electronic-photonic systems without changing their processes, and the resulting photonic devices can rival those built in custom photonics fabrication processes.

2.6.1 First depletion-mode modulators in 45 nm CMOS

We present two generations of the modulator. The modulator occupies 80 \(\mu\text{m}^2\) in area and runs at 5 Gbps. The first generation (G1) achieved a 5.2 dB extinction ratio and estimated energy consumption of 40 fJ/bit \(^{119}\). The second generation (G2) achieves a 7.3 dB extinction ratio and estimated energy consumption of <5 fJ/bit. The large improvement in energy efficiency is due to optimization of the optical cavity (linewidth) and coupling gaps based on results from the G1 devices. G1 devices were overcoupled which resulted in wider linewidth resonances and decreased extinction ratios. Characterization of G1 devices allowed the gaps to be tuned for critical coupling to achieve higher Q and extinction in G2 devices.

The modulator design is covered in Sec. 2.5.2. We utilize disk-like whispering-gallery modes of a microring of multi-modal width formed in the sub-90 nm-thick transistor body device layer of the SOI CMOS process. Figure 2.16 summarizes the fabricated device. A 3D render of the device is shown in Fig. 2.16(a). Eighty-four lateral p-n junctions extend radially out and are azimuthally distributed around the ring as shown in Fig. 2.16(b). The number of p-n junctions is limited by the minimum width and spacing rules on the p-n dopant masks and the silicon contacts on the inner sidewall. The ring cavity is wider than the single-mode width to allow electrical contacts

\(^4\) This work was presented in \[^{119,141}\].
placed at the inner-radius edge to impart minimal optical loss to the fundamental mode as seen in Fig. 2.12(a). Although the higher-order spatial modes are suppressed in Q by scattering from these contacts and bending loss, they remain high enough Q to have an undesirable spectral signature, if excited. Hence, excitation of only the fundamental mode, and suppression of the higher order modes, is further accomplished by a suitably designed coupler. A propagation-constant-matched, curved bus-to-resonator coupler with a long interaction length has a small $k$-space spread of the perturbation and does not excite the higher order, low-Q resonances [14]. The 5 µm outer ring
radius is larger than the minimum permitted by bending loss to accommodate an efficient coupler design. Optical micrographs of the fabricated device and full chip are shown in Figs. 2.16(c)-(e), and a scanning electron microscope image is shown in Fig. 2.16(f).

The optical transmission spectra in Figs. 2.17(a,b) show that only the fundamental family of TE modes are excited in the cavity, as designed. For G1, Fig. 2.17(a), the measured free spectral range (FSR) is 17 nm (3.2 THz) near 1260 nm, and the 3 dB linewidth is 26 GHz near critical coupling, indicating an intrinsic Q of 18,000. The lower through-port extinction for the two resonances at longer wavelengths indicates that the cavity is overcoupled, and the coupling gap was adjusted for G2. For G2, Fig. 2.17(b), the FSR is 18 nm (3.3 THz) near 1280 nm, and the 3 dB linewidth is 15.5 GHz near critical coupling, indicating an intrinsic Q of 29,500. For both generations, through port optical extinction exceeds 8 dB near the design wavelength.

Figure 2.17: Overview of spoked-ring modulator experimental characterization. (a) Passive spectrum of first generation device (G1) showing an FSR of 3.2 THz and ≈8 dB of extinction at the resonance near \(\lambda = 1264\) nm. The inset shows the resonance shift with applied DC voltage. (b) Passive spectrum of second generation device (G2) showing an FSR of 3.3 THz and ≈12 dB of extinction. The inset shows the resonance shift with applied DC voltage. (c) 5 Gbps eye diagram of first generation device with PRBS7 modulation data stream. The extinction ratio (E.R.) is ≈5.2 dB with ≈4.5 dB insertion loss (I.L.) (d) 5 Gbps eye diagram of second generation device with PRBS7 modulation data stream. The E.R. is ≈7.3 dB with ≈3 dB I.L.
Figure 2.17(c) shows an eye diagram from the G1 modulator when driven with a 5 Gbps, $2^7 - 1$ bit PRBS, using a 40-GHz GSG probe. The voltage swing seen by the modulator was -3V to +0.6V. Under these operating conditions, 5.2 dB modulation depth was measured with 4.5 dB insertion loss. The average switching energy ($1/4CV_{pp}^2$ for NRZ data) is estimated from the voltage swing and a computed device capacitance of 12 fF to be 40 fJ/bit, assuming full voltage doubling (worst case). Figure 2.17(d) shows a 5 Gbps eye diagram from the G2 modulator. The voltage swing was -0.6V to +0.6V. The G2 modulator achieved 7.3 dB modulation depth with 3 dB insertion loss and <5 fJ/bit energy consumption. This is comparable to state-of-the-art custom-process devices [131], and shows that high-performance devices can be designed with no process changes by leveraging advanced lithography and innovations in device topology. Furthermore, through monolithic integration and low-voltage operation, these devices promise very low driver circuit complexity and link energy-cost [81].

In a WDM transmitter, modulators cascaded such that the through port of each microring is connected to the input of the successive microring which multiplexes the modulated light, and each resonance is an independent communication channel. To align the resonances to their nominal channel spacing as well as stabilizing the resonance, the microring resonance must be tunable. For active resonance tuning, a resistive microheater was integrated in the modulator cavity, as mentioned earlier. The heater is formed in the body silicon layer and uses the process source/drain implants to create a 10 kΩ resistor to use for Joule heating. While tuning on the order of a channel spacing ($\approx$1 nm) is needed, these heaters are capable of tuning over a full FSR with tuning efficiencies of 1.6 nm/mW (3.4 $\mu$W/GHz) [119].

The static (DC) characterization of the G2 device is shown in Fig. 2.18. Although static testing gives only an approximate prediction of the dynamic behavior, it is useful when narrowing down promising device variants to extensively test since static testing is fast (several minutes) while extensively testing the dynamic behavior is time consuming (several days). Static testing includes

(1) measure optical transmission spectrum without voltage applied to find resonances with suitable extinction, Fig. 2.18(a),(b)
Figure 2.18: Summary of static modulator measurements. (a) Through port optical spectrum showing five resonances. (b) Each resonance is fit with a Lorentzian through port model Eq. 2.21 which is used to calculate quality factors. (c) DC voltage is applied to the p-n junctions to measure the resonance shift. Lorentzian through ports are fit to each resonance, and the resonance frequency is extracted to calculate the resonance shift per volt. A shift of 4.1 GHz/V is measured with -0.5V to 0.5V applied voltage bias. (d) Approximate extinction ratio (E.R.) and insertion loss (I.L.) calculated from the DC shifts from resonance 2.

(2) measure optical spectra as a function of applied bias voltage, Fig. 2.18(c)

(3) analyze data to calculate quality factors, Fig. 2.18(b),(c)

(4) calculate approximate modulation performance, Fig. 2.18(d)

A least-squares fit is applied to the through port model of a single ring resonator [Eq. 2.21], and the resulting fit parameters are used to calculate the quality factor and the resonance frequency shifts. For the G2 device, these are shown in Fig. 2.18(b),(c). The total quality factor is ≈15,900 while the extracted intrinsic quality factor is ≈43,000. The intrinsic quality factor matches well with what is expected from mid-level transistor implants (2e17 to 1e18), as shown in Fig. 2.2(a).

When testing the modulators, it is useful to calculate the approximate modulation perfor-
mance from the static characterization for all of the resonances. Figure 2.19 shows the approximate modulation performance for all of the resonances shown in Fig. 2.18(a). From Fig. 2.18(a), resonances 1-4 appear to be promising for modulation, but after calculating the approximate performance, it is quickly seen that resonance 2, Fig. 2.19(b), has the best performance for modulation. From Fig. 2.19(b), the best achievable extinction ratio is $\approx 7$ dB with an insertion loss of $\approx 4$ dB. This matches well with Fig. 2.17(d), noting that the drive voltage was slightly higher than used in the DC shifts.

![Figure 2.19: Approximate extinction ratio (E.R.) and insertion loss (I.L.) when modulating resonances 1-5 from Fig. 2.18(a), calculated from experimentally measured spectra under 0-state and 1-state drive. This static analysis clearly shows that resonance 2 will yield the best modulation. Resonance 2 is the same as shown in Fig. 2.18(d).](image)

### 2.6.2 Efficient thermally tunable filter

Also critical to WDM communication links are efficient tunable demultiplexing filters. The filters presented here achieve large efficiency due to the heat generation overlapping directly with the optical mode. Thus, for a given tuning power, the cavity is efficiently heated. This is achieved by heating the ring resonator itself by driving current through it. The cavity is made conductive and electrical contacts are made to the inner sidewall, and electrical current is driven through the
same silicon that is confining the optical mode, as opposed to a proximity heater used in the modulator, see Fig. 2.16(b).

Figure 2.20: Illustration of efficient tunable cavities. (a) The microring resonator blanket doped with contacts on the inner sidewall. (b) Inner portion of the ring is doped (either p-type or n-type) to avoid excess optical absorption. (c) Inner portion of ring is doped, and many electrical contacts are placed around the inner sidewall of the ring. This allows a lower resistance when the contacts are wired in parallel. (d) Input and output waveguides shown. The input waveguide has an extended coupling region to selectively excite the fundamental transverse mode of the ring, and the output waveguide is a point-like coupler. (e) Schematic of model used to derive filter response.

The basic concept for the efficient tunable filter is illustrated in Fig. 2.20. Using a similar cavity as the modulators, a wide ring that supports a disk-like whispering gallery mode allows electrical contacts to be placed on the inner sidewall far enough from the optical mode to avoid scattering. As was the case with the modulator, the cavity supports multiple higher-order transverse modes, so an extended interaction length input coupler is used to selectively excite the fundamental mode. Since the cavity is being used as a tunable filter, there must also be an output (“drop”)
port. The drop port waveguide does not need to be an extended coupler as is used for the input waveguide. The reason is that if the extended input waveguide is designed properly, only the fundamental mode of the cavity is excited. Thus, at the output port coupling, the coupling is single-mode. This allows a simple point-like coupler, as shown in Fig. 2.20 (d), and the short (point-like) coupler quickly moves away from the microring which prevents a second order coupling effect where the fundamental mode of the output waveguide couples back into higher order modes of the microring.

The simplest implementation for the tunable filter is shown in Fig. 2.20 (a). The cavity is blanket doped and electrically contacted on the inner sidewall of the ring. The blanket doping allows a simple shape drawn on the layout file which avoids DRC complications. Also, it ensures that current is flowing through the regions where the optical mode exists, thus generating heat directly in the optical mode. The downside, however, is that the optical mode experiences higher propagation loss due to the implants used to make the cavity conducting. For this reason, it is important to understand the expected transmission of the filter when the cavity is lossy.

For microring filters, the insertion loss that the drop port experiences depends on the intrinsic linewidth of the cavity and the linewidth of the filter response. A narrower intrinsic linewidth allows for narrower filter responses to be designed with lower insertion loss. When the intrinsic linewidth is comparable to the linewidth of the desired filter response, substantial insertion loss in the drop port occurs.

Figure 2.20 (e) shows the schematic of a first order filter that we will use to calculate the effect of implant concentration on the filter insertion loss. The CMT equations for Fig. 2.20 (e) are

\[ \frac{d}{dt} a = j \left[ \omega_o + j (r_t + r_i + r_d) \right] a - j \sqrt{2} r_i s_i \quad (2.36) \]

\[ s_t = s_i - j \sqrt{2} r_i a \quad (2.37) \]

\[ s_d = - j \sqrt{2} r_d a \quad (2.38) \]
where \( s_i \) is the input wave, \( s_t \) is the through port wave, \( s_d \) is the drop port wave, \( a \) is the energy amplitude in the resonator, \( r_i \) is the input coupling rate, \( r_d \) is the drop port decay rate (output coupling), and \( r_l \) is the decay rate due to losses (scattering and absorption). The input, through, and drop port waves are normalized such that \(|s_{i,t,d}|^2\) is the power, and the energy amplitude in the cavity is normalized such that \(|a|^2\) is the energy in the cavity.

The above equations can be used to design arbitrary first order filter responses \[65\], but we are specifically interested in the insertion loss at the drop port. When Eqs. 2.36, 2.37, 2.38 are solved, the drop port response on resonance is given by

\[
\left| \frac{s_d}{s_i} \right|^2 = \frac{4r_i^2}{(2r_i + r_l)^2} \quad (2.39)
\]

where it was assumed that \( r_i = r_d \). Recognizing that \( 2r_i + r_l = 2\pi\delta f_{\text{tot}} \) and \( 2r_i = 2\pi\delta f_e \), where \( \delta f_{\text{tot}} \) is the total 3 dB bandwidth of the filter and \( \delta f_e \) is the external linewidth of the cavity (equal to the total bandwidth of the filter in the lossless case), Eq. 2.39 becomes

\[
\left| \frac{s_d}{s_i} \right|^2 = \left( \frac{\delta f_e}{\delta f_{\text{tot}}} \right)^2 = \left( \frac{\delta f_e}{\delta f_e + \delta f_o} \right)^2 \quad (2.40)
\]

**Figure 2.21:** Drop port insertion loss versus filter 3 dB bandwidth and (a) electron implant concentration (b) hole implant concentration.

5 This is valid if the input and output couplers are properly designed to achieve the same coupling coefficient. Special care must be taken, however, since the input coupler is an extended interaction length sine coupler, and the output coupler is a point-like coupler.
Equation 2.40 allows us to directly calculate the drop port insertion loss as a function of filter bandwidth and doping level. The total bandwidth, $\delta f_{\text{tot}}$, is a function of $\delta f_o$, related to the carrier concentration used to make the cavity conductive [plotted in Fig. 2.3(b)], and the external linewidth, $\delta f_e$. Using Eq. 2.40 the equations in Table 2.1 and Eq. 2.16 the insertion loss is calculated as a function of filter bandwidth and free carrier concentration and is shown in Fig. 2.21. Two cases are considered: donor implants, Fig. 2.21(a), and acceptor implants, Fig. 2.21(b). In both cases, the approximation is made that the entire optical mode experiences the absorption caused by the implants. For a given carrier concentration, holes have a lower absorption coefficient as seen in Fig. 2.1. Figure 2.21 considers implant concentrations of $1 \times 10^{17}$ to $1 \times 10^{19}$ cm$^{-3}$ since these are commonly found in transistors. Typical transistor well implants are in the range of $1 \times 10^{17}$ to $2 \times 10^{18}$ cm$^{-3}$. To achieve better than 1 dB insertion loss, the required filter bandwidth is $> 20$ GHz when using donors and $> 10$ GHz when using acceptors with carrier concentrations $< 1 \times 10^{18}$ cm$^{-3}$, which shows that well implants are suitable.

A slightly refined form of this design is shown in Fig. 2.20(b). The basic idea is to only dope the inner portion of the silicon ring such that the optical mode has negligible overlap with the implants. This will result in a narrower intrinsic linewidth which will relax the required filter bandwidths shown in Fig. 2.21 while maintaining the efficient thermal tuning.

The total resistance of the conductive path must also be considered if the device is meant to be driven by on-chip circuits. In the 45 nm process, the on-chip circuits require the heater resistance to be $\approx 1 \, \text{k}\Omega$ [125] for maximum power delivery and tuning range. To control the total resistance, multiple contacts can be placed on the inner sidewall of the cavity as shown in Fig. 2.20(c). The contacts can then be wired in parallel in the backend metals to achieve an overall reduced resistance. If the resistance of the circumference of the conductive path is $R_{\text{tot}}$, the range of achievable resistances is then $R_{\text{tot}}/N$ for $N$ the number of contacts used. $N$ will have an upper bound limited by the design rules of the fabrication process. The following equation can be used to approximate the resistance.

$$R_{\text{tot}} = \frac{2\pi \left( R_{\text{out}} - \frac{w}{2} \right) \rho}{t \, w \, N}$$  (2.41)
where $R_{\text{out}}$ is the outer radius of the doped region, $w$ is the width of the doped region, $\rho$ is the resistivity of the doped silicon, $t$ is the thickness of the silicon, and $N$ is the number of contacts. For example, assuming a mid-level p-type concentration of $1 \times 10^{18} \text{ cm}^{-3}$, an outer ring radius of $5 \mu\text{m}$, a ring width of $1.2 \mu\text{m}$, a maximum $N$ of 80, and blanket doping, the range of achievable resistances is $R_{\text{tot}} \approx 1.5$ to $60 \text{k}\Omega$.

In thru drop

<table>
<thead>
<tr>
<th>Power (mW)</th>
<th>Δλ0 (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 mW</td>
<td>0</td>
</tr>
<tr>
<td>2.4 mW</td>
<td>3.5</td>
</tr>
<tr>
<td>5.2 mW</td>
<td>6.3</td>
</tr>
</tbody>
</table>

Figure 2.22: (a) 3D render of GDS layout submitted for fabrication. The cavity is blanket doped with a p-type implant and contacted on either side of the inner sidewall. The drop port waveguide is not an extended coupler since only the fundamental mode of the cavity is excited, so the single mode coupling is preserved at the drop port. The heater wires are shown which connect the cavity to upper-level metal pads. (b) Optical transmission spectrum showing $\approx 75$ GHz 3 dB bandwidth with $<0.5$ dB I.L. (c) Drop port spectra versus tuning power. (d) Change in resonant wavelength versus tuning power showing 2.73 nm/mW tuning efficiency.

The implementation shown in Fig. 2.20(a) was designed and fabricated in the 45 nm process. As a first demonstration, the simplest case was chosen to avoid DRC complications and explore the feasibility of the device. Figure 2.22(a) shows a 3D rendering of the actual layout file submitted for fabrication. As previously discussed, the input coupler has an extended interaction length, and the drop port coupler is point-like. Although several contacts are made on either side of the inner
sidewall, as seen in Fig. 2.22(a), the adjacent contacts are wired together in the backend metals such that the effective contact number is $N = 2$.

A mid-level p-type implant with concentration $<1e18$ was used to blanket dope the entire cavity. The input and output coupling gaps were chosen to achieve a 3 dB bandwidth of 80 GHz such that the resulting insertion loss was nominally better than 0.5 dB, as seen in Fig. 2.21(b). The contacts are routed to top-level metal pads for testing with microprobes. The fabricated resistance was $\approx 65 \text{k}\Omega$ indicating a p-type concentration of $\approx 8e17 \text{cm}^{-3}$.

Figure 2.22(c) shows the through and drop port spectra. The 3 dB bandwidth is 75 GHz which is close to the designed bandwidth. The insertion loss is $\approx 0.5 \text{dB}$ as expected. DC voltage was applied to the metal pads to measure the resonance shift as a function of applied electrical power where the electrical power was calculated by monitoring the applied voltage and the current drawn such that $P = IV$. Figure 2.22(d) shows the drop port responses for several applied heater powers, and Fig. 2.22(e) shows the change in resonance wavelength versus applied power. The slope of the fitted line in Fig. 2.22(e) gives a tuning efficiency of 2.73 nm/mW (2 $\mu$W/GHz).

The results summarized in Fig. 2.22 is a promising first demonstration, but improving on the design is required before the device is suitable to be integrated with circuits due to the large resistance. To achieve a lower resistance for integration with on-chip driver circuits, more contacts wired in parallel can be added as previously discussed and shown in Fig. 2.20(c). For densely packed WDM communication mux/demux filters, a 3 dB bandwidth of 80 GHz is too large to avoid adjacent channel crosstalk. A relevant 3 dB bandwidth is $<30 \text{GHz}$. However, to achieve suitable insertion loss, the cavity cannot be blanket doped, as made clear in Fig. 2.21. It is expected to be achievable through the implementation shown in Fig. 2.20(b).

2.7 Conclusions

The basics of active microring devices in a thin-SOI process have been presented. The waveguide and microring geometries presented in Figs. 2.5, 2.6 can be used in conjunction with the power coupling simulations (Tables 2.3, 2.4) to design a wide variety of silicon photonic devices.
A simple model was derived to predict the resonance shift of the “spoked-ring” modulator in the 45 nm CMOS process, and it was shown that transistor well implants result in near optimal carrier concentrations for a depletion-mode microring modulator (Fig. 2.10). The constraints of a thin-SOI microelectronics process were reviewed, and the spoked-ring cavity was introduced to satisfy those constraints. Using the spoked-ring cavity, energy efficient modulators and tunable filters were presented. The modulator achieved 5 Gb/s modulation rates with < 5 fJ/bit energy consumption which is near state-of-the-art energy efficiency compared to devices made in fully customized photonics processes [132]. Additionally, the embedded resistive heater for thermal tuning achieves a tuning efficiency of 3.3 μW/GHz which is the highest published in a working microring modulator, ≈ 20× better than those in [63, 92]. With an FSR of 3.2 THz, the cavity is well-suited for implementation in on-chip WDM communication transmitters, as presented in Ch. 4. The tunable microring filter achieved a tuning efficiency of 2 μW/GHz which is the second best efficiency published to date in a silicon microring tunable filter (best is [23]) while noting that the device in [23] achieves ≈ 1.7 μW/GHz but requires full suspension in air, and there are no other devices working on the same chip due to the extensive undercut.

We have demonstrated that the 45 nm process supports high-performance active photonic devices with zero in-foundry process changes. Thus, these devices open the door to large-scale, monolithically integrated electronic-photonic systems with a realistic path to commercial manufacturing.
Chapter 3

Vertical grating couplers

3.1 Introduction and background

Structures for redirecting light to and from planar waveguides have been studied for decades \[5, 152\], yet today remain one of the most challenging devices for enabling realistic large-scale integrated optical systems. Low-loss, broadband optical coupling to on-chip optics is a gating issue that severely limits the viability of integrated photonics \[125, 127\]. Although much research has been devoted to this problem in recent years, the majority of results have been achieved with tools and fabrication techniques that are difficult to transfer out of niche research environments. Thus, there is a need for high-performance designs to be explored within realistic manufacturing constraints. This chapter focuses on the design and experimental demonstrations of the first high-performance vertical grating couplers designed in an advanced CMOS microelectronics process. Although edge coupling is an alternative approach, it requires breaking the guard ring in electronic chips and does not support wafer scale test, both issues resolved by vertical grating couplers. The demonstrations presented in this chapter show that state-of-the-art loss is possible in zero change CMOS and as such are a milestone in enabling fully integrated electronic-photonic systems, and validate unmodified CMOS as a viable photonics platform.

The basic function of a grating coupler is to redirect light from planar, on-chip waveguides to out-of-plane radiating beams which can then be collected by an optical fiber. This is achieved with periodic perturbations to the waveguide that produce scattered light which, as seen in Fig. 3.1, constructively interferes at some angle relative to the grating. The angle is determined by the
well-known grating law \[46\],

\[
\frac{n_{cl} \sin(\theta)}{\lambda} = \frac{n_{eff}}{\lambda} \pm \frac{m}{\Lambda} \tag{3.1}
\]

where \(\theta\) is the output angle of the radiation relative to the direction perpendicular to the waveguide propagation, \(n_{cl}\) is the cladding refractive index, \(n_{eff}\) is the effective index of the waveguide mode, \(\lambda\) is the wavelength of light in vacuum, \(\Lambda\) is the spatial periodicity of the grating, and \(m\) is the diffraction order. In silicon photonics, gratings are formed by etching through the waveguide, either

![Cross-sectional schematic of chip-to-fiber coupling through a vertical grating coupler.](image)

Figure 3.1: Cross-sectional schematic of chip-to-fiber coupling through a vertical grating coupler. Light from the waveguide is incident on a period array of scattering elements (conceptually represented by point-like scatterers for illustration of concept). In the simple case shown, the structure is vertically symmetric which produces scattered light in both the up and down directions which can be collected by an optical fiber. The point-like scatterers are sources of spherically expanding phase-fronts which determine the output angle of the radiation.

fully or partially, as summarized in Fig. 3.2. The simplest grating uses a single full-etch step and can be defined by the same masks that define the waveguides, Fig. 3.2(a). These devices, however, are limited to around 50% in the maximum chip-to-fiber coupling efficiency that can be achieved due to their near vertically symmetric structure, and thus radiation pattern \[129\]. One widely used technique to achieve highly directional radiation is to introduce a reflective layer below or above the grating coupler resulting in upwards and downwards directionality, respectively. Alternatively, two etch steps can be used to allow scattering from vertically offset positions in the silicon layer,
and the relative phase between the scattering sites determines the directionality, Fig. 3.2(c).

![Figure 3.2: Configurations for on-chip grating couplers in silicon photonics showing optical power radiated up ($P_{up}$) and down ($P_{down}$). (a) Single-etch step resulting in vertically symmetric radiation ($P_{up} = P_{down}$). (b) Metal or dielectric reflectors introduced in the vertical cross-section to achieve directional radiation ($P_{up} \neq P_{down}$). (c) Two-etch steps resulting in offset scattering sites which radiate asymmetrically in the vertical direction ($P_{up} \neq P_{down}$).](image)

In low complexity fabrication processes, such as those that are typically practical in an academic setting, a grating coupler with broken vertical and horizontal symmetry, as shown in Fig. 3.2(c), is defined by two lithographic etch steps into a single, thick material layer [27, 28, 74]. Such processes impose geometric constraints on the design of the grating coupler that limit their ability to form unidirectional, Gaussian beam radiation. In [27], independently patterned layers were suggested to be attractive for gratings, but no explicit designs were produced. Here, we propose to use of two independently patternable CMOS device layers, one above the other as shown in Fig. 3.3 which removes the design constraints and permits higher efficiency designs. This strategy requires a chemical mechanical polishing (CMP) step between the patterning of the first device layer and its overcladding, and the deposition and patterning of the second device layer. CMP is an advanced processing step that requires careful tuning, wafer leveling and therefore complicates fabrication. However, it is a standard feature in all advanced electronic CMOS processes. Therefore,
our proposed strategy of using two independently patternable layers for efficient grating couplers [27], although complex for a university demonstration, comes at no extra cost if one uses an advanced electronic CMOS process. Specifically, the crystalline body silicon and gate polysilicon layers in SOI CMOS are well suited.

Figure 3.3: Highly directional grating couplers using two independently patternable layers to achieve (a) up and (b) down directionality.

In this chapter, we show that not only can such grating couplers be realized, but they provide extremely high efficiency (1.1 dB loss) and record bandwidth demonstrated to date in grating couplers of 14 THz (80 nm) between 1 dB roll-off points. This is accomplished in a 45 nm CMOS microelectronics process [61] by utilizing silicon layers that comprise the electrical transistors native to the process. We use the independently patternable crystalline silicon (c-Si) and polyysilicon (p-Si) layers, which form the active region and gate of the transistor, respectively, to create a highly directional, uniform vertical grating coupler whose design is inspired by phased array antennas. Our scheme thus avoids the need for custom SOI wafers [77] that increase cost, or any other schemes like bottom or top mirrors [13,21,54,76,106,115,135,154,155,157]. It is also realized using the CMOS deep UV photolithography, in contrast to most other high performance grating couplers which make use of e-beam lithography [21,155] and are thus limited to research demonstrations. As a result, we demonstrate an approach that can allow high-efficiency fiber-to-CMOS coupling in current commercial microelectronics foundries, providing an important milestone for the zero-change CMOS photonics approach [88]. The method, however, is also compatible with most other approaches to photonics-electronics integration that permit the integration of multi-
ple independently patternable device layers, whether in the transistor device layers [74, 77] or in separate photonics layers [22, 111, 158].

In Sec. 3.2, we begin by deriving a set of simple equations that encapsulate many grating-to-fiber coupling including bandwidth and coupling efficiencies versus mode-size mismatch, fiber positional tolerances, and fiber angular tolerances. To a first approximation, we show that the performance of grating-to-fiber coupling depends only on simple variables related to the grating and the fiber (geometry of the grating beam and fiber mode, group index of the mode in the grating, etc.). In Sec. 3.3, we cover the details of building grating couplers directly in the transistor layers of a CMOS process and show the first grating results obtained in the 45 nm process. In Sec. 3.4, we describe in detail the antenna-array concept used to design the high-performance gratings covered in Secs. 3.4.2 and 3.4.3. The results shown in Sec. 3.4.3 represent the first demonstration of high-performance coupling to an advanced CMOS microelectronics chip. Section 3.5 contains analysis of the fabrication sensitivities of the grating coupler presented in Sec. 3.4. These sensitivities impact the expected variation in performance across a full silicon wafer as well as packaging tolerances. The rigorous simulations are compared to the approximate equations derived in Sec. 3.2 and the approximate equations are shown to agree well with simulation.

3.2 Performance and design trade-offs in vertical grating couplers

The goal of this section is to arrive at a set of simple approximate equations to calculate bandwidth of coupling, positional tolerances (translational and angular), and parameters of Gaussian beam to optimize coupling parameters. The derived equations can be used to better understand the design trade-offs and how the basic parameters involved with grating couplers (size of grating, group index, wavelength of light) affect the fiber coupling behavior. A summary of the main equations that result from this section is given in Table 3.2.5. These equations are later used in Sec. 3.5 to show that they give similar results as full FDTD simulations.

We seek the simplest set of equations which describe grating performance that capture the dominant effects in grating-to-fiber coupling. This requires making intelligent approximations for
optical fields in fibers and gratings. Figure 3.4(a) shows the coordinate system orientation used to derive the equations in the next section. We use the following major assumptions and approximations:

1. We assume the optical field radiated by the grating coupler and the optical mode in the fiber can be represented by a function that gives the field profile (e.g. Gaussian or exponential) multiplied by a linear phase factor, \( \exp(j k_x x) \), that contains the angular dependence, i.e. tilted but flat phase front \( (k_x = 0 \text{ being normal incidence}) \). This approximation allows simple analytic overlap integrals to be used to calculate the grating-to-fiber coupling.

2. Constant waist size of the Gaussian beam versus frequency of light and angular detuning. The radiated output beam from the grating is assumed to have a constant spatial variation with a varying output angle. The small increase in waist size as a function of angular mismatch between grating and fiber optical fields is neglected, which is valid at small angles.

Figure 3.4(b) shows the simplified \( k \)-space diagram of the scattering process and defines the notation used below [46]. The waveguide incident on the grating coupler has a propagation constant,
\( \beta(\omega) \), that varies with the frequency of the light. When the light in the waveguide is incident on the grating, it is scattered into modes (either propagating waveguide modes of the uncorrugated waveguide, or radiation modes in the space above/below the grating) with 
\( k_x = \beta(\omega) + m K_G \),
where \( m \) is the diffraction order and is set to be \( m = -1 \) for the remaining analysis, for momentum conservation. There is (typically) no supported mode at \( \beta(\omega) + K_G \), and if \( \beta(\omega) - K_G \) falls within the cladding \( k \)-sphere, it produces radiation escaping into the cladding material. The second-order diffraction, \( k_x = \beta(\omega) - 2K_G \), ideally lies outside of the cladding \( k \)-sphere and does not produce propagating waves in the cladding. It can, however, excite a fraction of the backwards propagating waveguide mode which can be suppressed with an anti-reflection scatterer at the front of the grating. This simplified \( k \)-space model is used to derive expressions that describe the performance of vertical grating couplers. The next section introduces all of the equations that will be used as the basis for the following derivations.

### 3.2.1 General expressions for grating-to-fiber coupling efficiency

The electric field mode in the optical fiber is approximated as a normalized Gaussian beam:

\[
G_f(x, w_o, \theta) = \sqrt{\frac{1}{w_o}} \left( \frac{2}{\pi} \right)^{1/4} \exp \left[ -\left( x \frac{1}{w_o} \right)^2 \right] \exp \left[ -jk_{cl} \sin(\theta) x \right] \tag{3.2}
\]

where \( w_o \) is the mode-field radius of the Gaussian beam as defined in Eq. 5.3 in [46] (MFD = 2\( w_o \)), \( k_{cl} = (2\pi/\lambda)n_{cl} \) is the \( k \)-vector in the on-chip cladding, and \( \theta \) is the angle of the Gaussian beam relative to normal incidence. The normalization is chosen so that \( \int E_f^* E_f \, dx = 1 \). We assume small \( \theta \) so that effective waist size changes, \( w_o \), versus \( \theta \) and the Gouy phase can be neglected.

The output beam of the grating coupler is approximated as a normalized Gaussian mode given by:

\[
G_g(x, w_1, k_x) = \sqrt{\frac{1}{w_1}} \left( \frac{2}{\pi} \right)^{1/4} \exp \left[ -\left( x \frac{1}{w_1} \right)^2 \right] \exp (-jk_{x} x) \tag{3.3}
\]

where \( w_1 \) is the mode-field radius of the Gaussian beam (MFD = 2\( w_1 \)), and \( k_x \) is the \( x \)-component of the \( k \)-vector. Synthesizing Gaussian output beams by varying the dimensions of the scattering elements along the direction of propagation has been shown to very nearly produce ideal Gaussian
output beams \cite{83,130}; thus for the analysis presented here, it assumed that the output beam from the grating can be tailored to arbitrary Gaussian beam shapes. The width of the silicon waveguide is designed to match the width of the Gaussian fiber mode. Since the shape of the approximate Gaussian output beam can be entirely controlled, we assume radial symmetry (where the radial direction is orthogonal to the direction of the output radiation) which allows one dimensional equations [Eq. 3.3].

Also important is the beam produced by a uniform grating which has a periodic unit cell in the direction of propagation along the grating coupler. Uniform gratings produce close to exponentially decaying output beams, and the decay rate is a function of the scattering strength of the unit cell. The exponential output beam produced by the grating is approximated as:

\[
E_g(x, x_m, w_e, k_x) = \sqrt{\frac{2}{w_e}} \exp \left( -\frac{x - x_m}{w_e} \right) \exp(-jk_x x) H(x - x_m) \tag{3.4}
\]

where \(x_m\) is an offset in the x direction, \(w_e\) is the exponential decay length (1/e width), \(k_x\) is the x-component of the \(k\)-vector (derived in the next section), and \(H(x)\) is the unit (Heaviside) step function. For maximal overlap with Eq. 3.2, \(x_m\) and \(w_e\) must be optimally chosen.

For scalar fields, as we use here for TE polarized 2D structures, the fraction of launched field \(F_2(x)\) that excites mode \(F_1(x)\) in a new basis is given by the following scalar overlap integral:

\[
K = \left| \int_{-\infty}^{\infty} F_1^*(x) F_2(x) dx \right|^2 \tag{3.5}
\]

where \(F_1^*(x)\) is the complex conjugate of \(F_1\). The coupling between the exponential field generated by the grating and the Gaussian field of the fiber is given by the scalar overlap integral

\[
K_{GE}(x_m, w_e, w_o, \theta, k_x) = \left| \int_{-\infty}^{\infty} G_f^*(x, w_o, \theta) E_g(x, x_m, w_e, k_x) dx \right|^2 \tag{3.6}
\]

The analogous Gaussian-Gaussian overlap is

\[
K_{GG}(w_o, w_1, \delta x, \theta, k_x) = \left| \int_{-\infty}^{\infty} G_f^*(x - \delta x, w_o, \theta) G_g(x, w_1, k_x) dx \right|^2 \tag{3.7}
\]

where \(w_o\) is the waist size of the fiber Gaussian mode, and \(w_1\) is the waist size of the Gaussian beam produced by the grating. The optimal peak coupling is always achieved when \(\delta x = 0\) (i.e. the
Gaussian profiles are aligned). Later, in Sec. 3.2.4.2 positional misalignments ($\delta x \neq 0$) between Gaussian beams will be studied to analyze the effects of realistic alignment tolerances.

These equations will be used in the next sections to derive simple equations to calculate trade-offs between coupling efficiency and bandwidth as well as positional tolerances.

### 3.2.2 Calculating the bandwidth of the grating-to-fiber coupling coefficient

The goal of this section is to arrive at simple expressions describing the bandwidth of the grating-to-fiber coupling coefficient, we need an expression for $k_x(\omega)$ in Eq. 3.3. From the grating $k$-space picture

\[ k_x(\omega) = k_{cl} \sin \theta = \beta(\omega) - K_G \tag{3.8} \]

where $k_{cl}$ is the magnitude of the output $k$-vector $\vec{k}_{cl}$, $k_x$ is its $x$-component, $\theta$ is the angle between the normal to the grating and the $k$-vector of the output beam, $\beta(\omega)$ is the propagation constant at frequency $\omega$ in the grating region, and $K_G$ is the grating momentum defined as $2\pi/\Lambda$ where $\Lambda$ is the period of the grating. According to Eq. 3.8 the angle of the output beam depends on frequency $\omega$. We can assume that the fiber is optimally aligned at the angle $\theta = \theta_o$ corresponding to the center frequency of the beam $\omega = \omega_o$; so that

\[ k_x(\omega_o) = k_{cl} \sin(\theta_o) = \beta(\omega_o) - K_G \tag{3.9} \]

and $K_G$ is designed to be as

\[ K_G = \beta(\omega_o) - k_{cl} \sin \theta_o \tag{3.10} \]

We now inspect the frequency dependence of $k_x(\omega)$. First, we write the first-order Taylor series approximation for $\beta(\omega)$.

\[ \beta(\omega) \approx \beta(\omega_o) + \frac{\partial \beta}{\partial \omega} \bigg|_{\omega=\omega_o} \delta \omega, \tag{3.11} \]

\[ \beta(\omega) \approx \beta(\omega_o) + \frac{n_g}{c} \delta \omega, \]

where $n_g$ is the group index of the waveguide (in the weak perturbation approximation\(^1\)) at $\omega = \omega_o$, $c$ is the speed of light in vacuum, and $\delta \omega$ is a small deviation from $\omega_o$. Substituting Eqs. 3.10 3.11

\(^1\) The weak perturbation approximation is not valid for high index contrast as we have in silicon devices. It is likely that this group index is rigorously the group index of the Bloch mode of an infinite periodic grating.
into Eq. 3.8 gives

\[ k_x(\omega) = \beta(\omega_o) + \frac{ng}{c} \delta \omega - K_G \]

\[ = \beta(\omega_o) + \frac{ng}{c} \delta \omega - \beta(\omega_o) + k_{cl} \sin \theta_o \]

\[ = \frac{ng}{c} \delta \omega + k_{cl} \sin \theta_o \]

which provides change in the grating’s output beam angle, \( \theta \), versus the frequency of light, \( \omega \).

We can now plug Eq. 3.12 into Eq. 3.3 and use Eq. 3.7 to find the coupling coefficient for the Gaussian-Gaussian case, if \( w_1 = w_o \).

\[ K_{GG}(w_o, k_x) = \frac{ng}{c} \delta \omega + k_{cl} \sin \theta_o; w_1 = w_o, \delta_x = 0, \theta = \theta_o \] = \[ \cdot \left| \frac{1}{w_o} \right| \sqrt{\frac{2}{\pi}} \exp \left[ -\frac{1}{8} \left( \frac{ng \delta \omega w_o}{c} \right)^2 \right] \]

(3.13)

Note that the dependence on the input angle, \( \theta_o \), vanishes due to the conjugation in the overlap integral. Since we are ultimately interested in the roll-off of Eq. 3.13 as a function of \( \delta \omega \), we will drop the constants in front of the integral. We recognize 3.13 as the Fourier transform of the Gaussian function \( \exp \left[ -\frac{a}{2} x^2 \right] \) with the transform domain variable equal to \( \frac{ng}{c} \delta \omega \). The Fourier transform of a Gaussian is a Gaussian and is given by

\[ \mathcal{F} \{ \exp[-ax^2] \} = \sqrt{\frac{1}{2a}} \exp \left( -\frac{k_x^2}{4a} \right) \]

(3.14)

Substituting \( a = 2/w_o^2 \) and the transform domain variable \( k_x = \frac{ng}{c} \delta \omega \), we arrive at

\[ K_{GG}(w_o, \delta \omega) \propto \left| \frac{w_o}{2} \exp \left[ -\frac{1}{8} \left( \frac{ng \delta \omega w_o}{c} \right)^2 \right] \right|^2 \]

(3.15)

Taking only the exponential portion of Eq. 3.15 and setting it equal to a value in decibels gives

\[ \exp \left[ -\frac{\pi^2}{2} \left( \frac{ng \delta f w_o}{c} \right)^2 \right] = 10^{-\frac{L_{dB}}{10}} \]

(3.16)

Which verifies that there is no loss (0 dB) at \( \delta f = 0 \) (i.e. \( \omega = \omega_o \)). We substituted \( \delta \omega = 2\pi \delta f \), and \( L_{dB} \) is the attenuation in decibels when detuned in frequency by \( \delta f \). Equation 3.16 can be solved for \( \delta f \)

\[ \delta f = \sqrt{\frac{2L_{dB} \ln(10)}{5} \frac{c}{2\pi ng w_o}} \]

(3.17)
The full two-sided bandwidth is given by $\Delta f = 2\delta f$. Equation 3.17 is similar to Eq. 3 in [22], but is simpler. It also shows that the group index and the waist size are all that is needed to approximate the bandwidth of the coupling whereas the equation in [22] requires the numerical aperture of the fiber, the cladding index, the effective index of the grating, the wavelength of light, and the output angle.

Figure 3.5 plots the two-sided 1-dB bandwidth versus the waist size of the Gaussian beam. Widely used fiber mode-field diameters (MFD = 2\(w_o\)) are 5\(\mu\)m and 10\(\mu\)m corresponding to 1-dB bandwidths of \(\approx 11 - 15\) THz and \(\approx 6 - 8\) THz, respectively, depending on the group index. As expected, the smaller the waist size, the larger the bandwidth. However, smaller waist sizes also result in more stringent positional tolerances, as discussed later.

![Figure 3.5](image)

Figure 3.5: Two-sided 1-dB bandwidth versus Gaussian beam waist size for different group indices, \(n_g\), of the grating Bloch mode.

### 3.2.3 Optimum overlap for uniform gratings

Uniform (non-apodized) grating couplers are widely used due to their relative design simplicity as well as potentially reduced fabrication tolerances and lesser demands on high resolution lithography. The radiated field from a uniform grating exponentially decays since the scattering
strength is constant along the grating. Thus, the output beam has an exponential shape that couples into the approximate Gaussian mode of the optical fiber. In our idealized model, we assume an abrupt start to the exponential output field of the grating [Fig. 3.6], but in practice, the field is more rounded.

We first derive expressions for the exponential profile of the grating output beam that optimizes coupling to a Gaussian profile. This is done by maximizing Eq. 3.6. Equation 3.6 can be reduced to the following form

\[ K_{GE} = \sqrt{\frac{\pi}{2 w_e^2}} \exp \left( \frac{w_o^2 + 4 x_m w_e}{2 w_e^2} \right) \left[ \text{erfc} \left( \frac{x_m}{w_o} + \frac{w_o}{2 w_e} \right) \right] \]  (3.18)

where \( \text{erfc}(x) \) is the complementary error function, \( \text{erfc}(x) = 1 - \text{erf}(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) \, dt \).

\( x_m \) and \( w_e \) can be scaled by \( w_o \) to give

\[ K_{GE}' = \sqrt{\frac{\pi}{2 w_e'}} \exp \left( \frac{1 + 4 x_m' w_e'}{2 w_e'^2} \right) \left[ \text{erfc} \left( x_m' + \frac{1}{2 w_e'} \right) \right] \]  (3.19)

where \( x_m' = x_m/w_o \) and \( w_e' = w_e/w_o \).

Although the optimal \( x_m \) and \( w_e \) in Eq. 3.19 cannot be obtained analytically, it is readily calculated numerically. This results in the following numerically obtained optimal values for \( w_e \) and \( x_m \) as a function of \( w_o \).

\[ x_m(w_o) = -0.7312 w_o \]

\[ w_e(w_o) = 1.4625 w_o \]  (3.20)

Figure 3.6 plots the optimal exponential field for coupling to a Gaussian using values from Eq. 3.20 for \( x_m \) and \( w_e \). The optimal exponential field overlap with a Gaussian field yields a coupling efficiency of \( \approx 80\% \). This sets the upper limit for coupling efficiencies achievable with uniform gratings\(^2\).

\(^2\) Actual grating couplers will not achieve a sharp step-like start to the radiated field profile – it will be more gently curved; thus, the highest achievable coupling with a fabricated uniform grating coupler is likely slightly higher than 80%. 

Figure 3.6: Optimal exponential field for maximizing the overlap with a Gaussian profile. The exponential and Gaussian curves are normalized such that \( \int_{-\infty}^{\infty} G_f(x) \, dx = 1 \) and \( \int_{-\infty}^{\infty} E_g(x - x_m) \, dx = 1 \)

3.2.4 Performance and design trade-offs for Gaussian gratings

It is important to understand the basic trade-offs in peak coupling efficiency, bandwidth, and positional tolerances, both translational and angular. This section focuses on deriving simple equations that explore these trade-offs when the output radiation of the grating coupler is assumed to have a Gaussian profile \([83, 130]\), and is coupled to a fiber mode which is also assumed to have a Gaussian profile.

3.2.4.1 Coupling efficiency versus bandwidth

To examine the trade-off between coupling efficiency and bandwidth, we begin by expressing the coupling coefficient as the overlap integral

\[
K_{GG}(w_o, w_1, k_x) = \left[ \int_{-\infty}^{\infty} G_f^{*}(x, w_o) G_g(x, w_1, k_x) \, dx \right]^2 \tag{3.21}
\]

where the Gaussian fiber mode has waist size \( w_o \), and the Gaussian beam produced by the grating has waist size \( w_1 \). Following a similar procedure outlined in Sec. 3.2.2, we arrive at the equation
for the coupling efficiency

\[ K_{GG}(r, w_1, k_x) = 2 \frac{r}{1 + r^2} \exp \left( -\frac{k_x^2 w_1^2}{1 + r^2} \right) \]  \hspace{1cm} (3.22)

where \( r \equiv w_1/w_o \) and \( k_x = \delta \omega n_g/c \) as previously discussed. The peak coupling efficiency is then

\[ K_{GG}(r, k_x = 0) = \eta_{peak} = 2 \frac{r}{1 + r^2} \]  \hspace{1cm} (3.23)

and the two-sided bandwidth in Hz is

\[ \Delta f_{\text{dB}} = \frac{c}{\pi n_g w_o r} \sqrt{\frac{L_{\text{dB}}}{5} \ln(10)(1 + r^2)} \]  \hspace{1cm} (3.24)

where \( L_{\text{dB}} \) is the attenuation corresponding to the two-side bandwidth in Hz.

Figure 3.7 plots Eq. 3.23 (blue) and Eq. 3.24 (red and yellow) for commonly used fiber waist sizes. When \( w < 1 \) (i.e. the waist size of the radiation from the grating coupler is smaller than the waist size of the fiber mode), the bandwidth of the coupling increases due to a smaller exponential roll-off, as can be seen in Eq. 3.22. However, the increase in bandwidth comes with a trade-off in peak coupling efficiency as seen in the blue trace in Fig. 3.7. Equation 3.24 can be solved in terms...
of the waist size ratio $r$ and plugged into Eq. 3.23 to give the peak coupling efficiency as a function of bandwidth.

$$\eta_{\text{peak}}(\Delta f_{\text{LdB}}) = \frac{2c^2 L_{\text{dB}} \ln(10)}{5(\Delta f n_g \pi w_o)^2} \sqrt{\frac{5(\Delta f n_g \pi w_o)^2}{c^2 L_{\text{dB}} \ln(10)}} - 1$$

(3.25) 

To simplify the above equation, we can define a normalized square bandwidth

$$\Delta f_{\text{LdB}}^2 = \frac{5(\Delta f n_g \pi w_o)^2}{c^2 L_{\text{dB}} \ln(10)}$$

(3.26) 

and plug this into Eq. 3.25. This results in

$$\eta_{\text{peak}}(\Delta f_{\text{LdB}}^') = \frac{2\sqrt{\Delta f_{\text{LdB}}^2 - 1}}{\Delta f_{\text{LdB}}^2}$$

(3.27) 

The coupling efficiency as a function of 1-dB bandwidth is shown in Fig. 3.8. The blue trace shows $\eta_{\text{peak}}$ as a function of the normalized bandwidth as defined in Eq. 3.26 and the red and yellow traces show $\eta_{\text{peak}}$ as a function of the two-sided 1-dB bandwidth for waist sizes 5 $\mu$m and 2.5 $\mu$m, respectively. Figure 3.8 clearly shows the trade-off between bandwidth and peak coupling efficiency. For example, for $w_o = 2.5$ $\mu$m, 100% efficiency is possible with a bandwidth of $\approx 12$ THz, but if the peak coupling efficiency is reduced to $\approx 78\%$, the 1-dB bandwidth increases to 20 THz. Note that for larger bandwidths, the fiber waist size is always smaller than the waist size of the beam produced by the grating ($r < 1$). This physically make sense because waist size and fiber acceptance angle are inversely related through a Fourier transform, and the wide acceptance angle is responsible for the larger bandwidth since detuning $\omega$ from $\omega_o$ results in an angular detuning from $\theta_o$.

### 3.2.4.2 Positional tolerance

In this section, we will calculate the positional tolerance as a function of the minimum required coupling efficiency, $\eta_{\text{min}}$, and waist size ratio, $r$. This leads to an optimal solution for $w$ for a required coupling efficiency (i.e. the minimum required coupling efficiency is achieved over a range of positional tolerances). The overlap integral including positional detuning is given by

$$K_{GG}(w_o,w_1,\delta x) = \left| \int_{-\infty}^{\infty} G_f^*(x-\delta x,w_o)G_g(x,w_1)dx \right|^2$$

(3.28)
where we allow the fiber mode, $G_f$, to deviate from the optimal position by $\delta x$. Calculating this integral gives

$$K_{GG}(w_o, w_1, \delta x) = 2 \frac{w_o w_1}{w_o^2 + w_1^2} \exp \left( -2 \frac{\delta x^2}{w_o^2 + w_1^2} \right)$$

Defining $\delta \equiv \delta x/w_o$ and $r \equiv w_1/w_o$ gives

$$K_{GG}(r, \delta) = \eta = 2 \frac{r}{1 + r^2} \exp \left( -2 \frac{\delta^2}{1 + r^2} \right)$$

where $\eta$ is the power coupling coefficient. We can solve Eq. 3.30 for $\delta$ in terms of $\eta$ and $r$.

$$\delta = \sqrt{\frac{(1 + r^2)}{2} \ln \left( \frac{2r}{\eta (1 + r^2)} \right)}$$

Assuming monotonic decrease in $\eta$ with positional offset, then the minimum efficiency, $\eta_{\text{min}}$ occurs at the maximum offset, $\delta$. For a given minimum efficiency $\eta_{\text{min}}$, positional tolerance in Eq. 3.31 can be maximized by choosing the correct ratio of mode sizes, $r$. This allows the designer to trade-off coupling efficiency with positional tolerance. Figure 3.9 plots the positional tolerance as defined in Eq. 3.31 versus $r$ for several efficiencies. This plot shows the range of positional offsets ($\pm \delta x$) for

Figure 3.8: Grating-to-fiber peak power coupling efficiency ($\eta_{\text{peak}}$) as a function of normalized bandwidth (blue). Coupling efficiency as a function of 1-dB bandwidth for $w_o = 5\mu m$ (red) and $w_o = 2.5\mu m$ (yellow). $n_g$ assumed to be 3.0.
which the coupling efficiency is at least $\eta_{\text{min}}$. For example, if the minimum required efficiency is 60% (shown in red), the positional tolerance is maximized near $r = 1.75$ (point A in Fig. 3.9). This gives a $\approx 20\%$ improvement in positional tolerance compared to the matched case ($r = 1$, point B in Fig. 3.9). Figure 3.10 plots the positional tolerance versus $r$ and $\eta_{\text{min}}$. The contours shown in Fig. 3.9 are lines of constant $\eta_{\text{min}}$ from Fig. 3.10.

![Figure 3.9: Positional tolerance $\delta$ as a function of waist size ratio $r$ for several values of coupling efficiency. For each required efficiency, there is an optimal waist size ratio that gives the maximum positional tolerance.](image)

Figures 3.9 and 3.10 show the positional tolerances, $\delta$, while maintaining $> \eta_{\text{min}}$ coupling efficiencies. This leads to an optimal choice for the waist size ratio, $r$, which results in a decrease of the peak coupling efficiency $\eta_{\text{peak}}$. Thus, the designer can choose to trade-off $\eta_{\text{peak}}$ to achieve $> \eta_{\text{min}}$ over a larger range of $\delta$ by properly choosing $r$. For a given $\eta_{\text{min}}$ in Figs. 3.9 and 3.10, the optimal $r$ is plugged into Eq. 3.30 with $\delta x = 0$ to calculate the peak coupling coefficient $\eta_{\text{peak}}$ to directly see the trade-off between $\eta_{\text{peak}}$ and $\delta$. This is plotted in Fig. 3.11.

### 3.2.4.3 Angular tolerance

In this section, we will first analyze the angular tolerance when the mode sizes are matched, $w_o = w_1$, for clarity. We will then analyze the scenario where the mode sizes are mismatched.
Figure 3.10: Positional tolerance $\delta$ versus waist-size ratio $r$ and minimum coupling efficiency $\eta_{\text{min}}$. For a constant $\eta_{\text{min}}$, there is an optimal $r$ to achieve the largest positional tolerance $\delta$.

Figure 3.11: Coupling efficiencies $\eta_{\text{peak}}$, $\eta_{\text{min}}$, and $\eta_{\text{match}}$ versus $\delta$ directly showing the trade-off between $\delta$ and $\eta_{\text{peak}}$. $\eta_{\text{match}}$ is the coupling efficiency corresponding to $\delta$ when the waist-size ratio is $r = 1$ (i.e. matched Gaussian waist sizes) and is included for comparison.

**Matched mode sizes, $w_o = w_1$**

To calculate the angular tolerance, we allow the Gaussian fiber mode to deviate from the optimal angle by $\delta \theta$. The power coupling equation is then

$$K_{GG}(w_o, w_1 = w_o, \delta \theta; \delta x = 0, k_x = 0) = \left| \int_{-\infty}^{\infty} G_f^*(x, w_o, \delta \theta) G_g(x, w_o) \, dx \right|^2$$

(3.32)
where $\delta \theta$ is included in the linear phase term so that the Gaussian fiber mode is given by

$$G_f(x, w_o, \delta \theta) = \sqrt{\frac{1}{w_o}} \left( \frac{2}{\pi} \right)^{1/4} \exp \left[ -\left( \frac{x}{w_o} \right)^2 \right] \exp (-jk_{cl} \sin (\delta \theta) x) \quad (3.33)$$

where $k_{cl} = (2\pi/\lambda)n_{cl}$, $\lambda$ is the free space wavelength of the light, and $n_{cl}$ is the cladding index. The reference plane is rotated to align with $\theta_o$, and $\delta \theta$ contains the angular mismatch between the grating and fiber modes. Substituting Eq. 3.33 into Eq. 3.32 and simplifying gives

$$\eta = \exp \left[ -\left( \frac{\pi n_{cl} \sin(\delta \theta) w_o}{\lambda} \right)^2 \right] \quad (3.34)$$

A fiber positional offset, $\delta x$, can also be included, as discussed in Sec. 3.2.4.2, which yields the following equation showing the effect of waist size on the angular and positional tolerance

$$\left[ \frac{\pi n_{cl} \sin(\delta \theta) w_o}{\lambda} \right]^2 + \left( \frac{\delta x}{w_o} \right)^2 = -\ln (\eta) \quad (3.35)$$

This equation shows that a larger waist size ($w_o$) increases the effect of angular misalignment ($\delta \theta$) while decreasing the effect of positional misalignment ($\delta x$).

Equation 3.34 can be solved for $\delta \theta$ to arrive at an expression for the maximum angular tolerance for a given $\eta$.

$$\delta \theta = \sin^{-1} \left( \frac{\lambda}{2\pi w_o n_{cl}} \sqrt{\frac{2\ln(10) L_{\text{dB}}}{5}} \right) \quad (3.36)$$

where $L_{\text{dB}}$ is the maximum tolerable attenuation in decibels, $L_{\text{dB}} = -10 \log_{10}(\eta)$. Equation 3.36 is plotted in Fig. 3.12 and shows that as expected, $\delta \theta \sim 1/w_o$. Also, $\delta \theta \sim 1/n_{cl}$ which is expected since the divergence angle of Gaussian beams is inversely proportional to the cladding index.

**Mismatched mode sizes, $w_o \neq w_1$**

The same analysis as in the previous section can be performed while allowing the mode size of the grating mode to be different than the fiber mode.

$$K_{GG}(w_o, w_1, \delta \theta) = \left| \int_{-\infty}^{\infty} G_f^*(x, w_o, \delta \theta) G_g(x, w_1) dx \right|^2 \quad (3.37)$$

Following a similar procedure as before, $\delta \theta$ can be calculated as a function of the tolerable insertion loss in decibels, $L_{\text{dB}}$.

$$\delta \theta = \sin^{-1} \left\{ \frac{\lambda}{2\pi n_{cl} w_o} \sqrt{\frac{(r'f^2 + 1) L_{\text{dB}}}{5} \ln \left[ \frac{5 (r'f^2 + 1)}{r'} \right]} \right\} \quad (3.38)$$
where \( r' \equiv w_o / w_1 \).

Lastly, we calculate the coupling efficiency as a function of the angular and positional tolerance while allowing \( w_o \neq w_1 \).

\[
\eta = \frac{2r}{1 + r^2} \exp \left[ - \frac{(k_{cl} \sin \theta r w_o)^2 + 4 \left( \frac{\delta x}{w_o} \right)^2}{2 (1 + r^2)} \right] \tag{3.39}
\]

where \( r \equiv w_1 / w_o \) and \( k_{cl} = 2\pi n_{cl} / \lambda \). The coupling efficiency given by Eq. 3.39 is plotted in Fig. 3.13 for \( r = 1 \) and several choices of \( \lambda \) and \( w_o \).

### 3.2.5 Summary of equations for key grating coupler properties
two-sided bandwidth

\[ \Delta f = \sqrt{\frac{L_{dB} \ln(10)}{5}} (1 + r^2) \frac{c}{\pi n_g w_o r} \]

peak efficiency as a function of bandwidth

\[ \eta_{peak} = \frac{2 \sqrt{\Delta f'^2 - 1}}{\Delta f'^2} \]

\[ \Delta f'^2 = \frac{5 (\Delta f n_g w_o)^2}{c^2 L_{dB} \ln(10)} \]

positional tolerance

\[ \delta = \sqrt{\frac{(1+r^2)}{2} \ln \left( \frac{2r}{\eta(1+r^2)} \right)} \]

\[ \delta \equiv \delta x/w_o, \ r \equiv w_1/w_o \]

angular tolerance

\[ \delta \theta = \sin^{-1} \left\{ \frac{\lambda}{2 \pi n_{cl} w_o} \sqrt{\frac{(r''^2+1)}{5} L_{dB} \ln \left[ \frac{5(r''^2+1)}{r''^2} \right]} \right\} \]

\[ r'' \equiv w_o/w_1 \]

efficiency as function of positional and angular misalignment

\[ \eta = \frac{2r}{1+r^2} \exp \left[ -\frac{(k_{cl} \sin \delta \theta r w_o)^2 + 4d^2}{2(1+r^2)} \right] \]

\[ r \equiv w_1/w_o, \ \delta \equiv \delta x/w_o \]

constrained efficiency equation

\[ \left[ \frac{\pi n_{cl} \sin(\delta \theta) w_o}{\lambda} \right]^2 + \left( \frac{\delta x}{w_o} \right)^2 = -\ln (\eta) \]

\[ w_o = w_1 \]

Table 3.1: Summary of derived equations.

<table>
<thead>
<tr>
<th>symbol</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( w_o )</td>
<td>1/e waist radius of fiber Gaussian mode</td>
</tr>
<tr>
<td>( w_1 )</td>
<td>1/e waist radius of grating Gaussian mode</td>
</tr>
<tr>
<td>( r )</td>
<td>( w_1/w_o )</td>
</tr>
<tr>
<td>( r'' )</td>
<td>( w_o/w_1 )</td>
</tr>
<tr>
<td>( \eta )</td>
<td>power coupling efficiency</td>
</tr>
<tr>
<td>( \eta_{peak} )</td>
<td>peak power coupling efficiency</td>
</tr>
<tr>
<td>( \eta_{min} )</td>
<td>minimum power coupling efficiency</td>
</tr>
<tr>
<td>( L_{dB} )</td>
<td>power coupling coefficient in decibels</td>
</tr>
<tr>
<td>( \delta x )</td>
<td>positional misalignment</td>
</tr>
<tr>
<td>( \delta )</td>
<td>normalized positional misalignment, ( \delta x/w_o )</td>
</tr>
<tr>
<td>( \delta \theta )</td>
<td>angular misalignment</td>
</tr>
<tr>
<td>( \Delta f_{L_{dB}} )</td>
<td>two-sided bandwidth</td>
</tr>
<tr>
<td>( n_g )</td>
<td>guided group index</td>
</tr>
<tr>
<td>( n_{cl} )</td>
<td>cladding refractive index</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of variables.
Figure 3.13: Power coupling efficiency in decibels, $10 \log_{10}(\eta)$, shown for $\lambda = 1180, 1310, 1550$ nm (from top to bottom) and $w_0 = 2.5, 5.0$ $\mu$m (from left to right), shown as a function of positional and angular misalignment $\delta x$ and $\delta \theta$. 

(a) $\lambda = 1180$ nm, MFD = $2w_0 = 5.0$ $\mu$m  
(b) $\lambda = 1180$ nm, MFD = $2w_0 = 10.0$ $\mu$m  
(c) $\lambda = 1310$ nm  
(d) $\lambda = 1310$ nm  
(e) $\lambda = 1550$ nm  
(f) $\lambda = 1550$ nm
### 3.3 Vertical grating couplers in zero-change CMOS

The following sections present experimental demonstrations of vertical grating couplers in the 45 nm CMOS process and show progress towards ultra-high efficiency grating couplers that meet the requirements for energy efficient chip-to-chip photonic interconnects. Recently, a chip-to-chip link between monolithically integrated photonic transceivers in bulk CMOS was demonstrated in [125]. However, high input/output losses in the grating couplers required an optical amplifier to be used between the chips. The optical amplifier is unrealistic in a commercial system due to its huge impact on energy efficiency of the links as well as its bulkiness and cost. Furthermore, optical losses must be overcome by increasing the input laser power. Lasers typically have $< 20\%$ wall-plug efficiencies (ratio of output optical power to total dissipated electrical power), and decreasing the input laser power results in significant energy savings. In this section, we describe how the front-end silicon transistor layers that exist natively in sub-100 nm silicon-on-insulator (SOI) CMOS nodes can be used to build vertical grating couplers. We describe multiple gratings designed in the 45 nm process and show the evolution of their performance.  

---

3.3.1 Using transistor layers to build vertical grating couplers

Sub-100 nm SOI microelectronics processes offer a near ideal dielectric environment for building high-performance vertical grating couplers because they support two independently patternable silicon layers which we show can be designed to produce radiation with $> 99\%$ power radiated up or down. Figure 3.14(a-c) illustrates a standard MOSFET in the 45 nm SOI process and how its silicon layers are used to build gratings. The MOSFET, shown in Fig. 3.14(a), consists of a crystalline silicon active region, where the source and drain connections are made, as well as a polycrystalline silicon gate. These layers comprise the “front-end” device layers. The crystalline silicon was previously shown to have low optical propagation loss [88] that is suitable for building optical waveguides, resonators, and transceivers. The polycrystalline silicon exhibits higher loss, but it can

---

\[ ^4 \text{All of the gratings in this section are designed to be used with 5 \mu m mode-field diameter fibers. This spot size is chosen to allow for more densely packed devices while prototyping designs. More standard 10 \mu m spot sizes can be designed and used in the future for better compatibility with standard fiber optic spot sizes.} \]
be selectively used for short propagation lengths such as those in a vertical grating coupler.

Figure 3.14: Front-end device layers used to build transistors and optical devices. (a) Illustration of a MOSFET in a thin-SOI process. (b) Illustration of patterned crystalline body crystalline silicon (active region of transistor) to form a vertical grating coupler. (c) Illustration of patterned polycrystalline silicon (gate of transistor) and crystalline silicon to break the vertical symmetry for high directionality radiation. (d) TEM image showing the front end layers, the silicon substrate, and a few upper level dielectrics. (e) Zoom-in showing a unit cell of the single layer grating. (f) Zoom-in showing a unit cell of the dual layer grating.

A large challenge to implementing photonics in the “zero-change” approach is that the CAD tools and Process Design Kit (PDK) that is received from the foundry do not (at this time) officially support any photonic design. Thus, the photonic devices must pass thousands of design-rule-check (DRC) rules before the foundry will accept the design. To minimize DRC complexity, in the simplest case, the crystalline silicon can be patterned to form a periodic array, and this can be used as vertical grating coupler. This grating is illustrated in Fig. 3.14(b) and will be referred to as a “Single Layer Grating (SLG)”. The main advantage of the SLG is that the design is contained to a single silicon layer; therefore, its design and implementation are greatly simplified. The main disadvantage of the SLG is that the structure is an almost vertically symmetric device and therefore is limited to, at best, around 50% radiation directivity in each of the up and down directions. Although this is useful during research phases when devices are tested from both the top-side and the bottom-side of the chips [see Fig. 3.1], ultimately, high directivity of the output optical beam is needed for
low-loss coupling to fibers.

To achieve highly directional vertical grating couplers, discussed in detail in Sec. 3.4.2, both the crystalline silicon and the polysilicon transistor layers can be independently patterned as illustrated in Fig. 3.14(c) and referred to as a “Dual Layer Grating (DLG)”. This allows the directionality of the radiation to be controlled entirely by the design of the widths and relative positions of the front-end silicon layers.

Figures 3.14(e-f) show transmission electron microscope (TEM) images of the dielectric environment in the cross-section of the gratings. A SiN dual-stress liner can be seen which is conformal to the polysilicon and the buried oxide. This liner is in the CMOS process to produce compressive or tensile strain and increase the mobility of the electrical carriers (holes) for better transistor performance [61], and it is necessary to include it in design simulations. The buried oxide is much thinner than the buried oxide of traditional custom SOI wafers used for photonics. To eliminate optical tunneling from optical waveguides into the Si substrate through the thin oxide, which would result in extremely high waveguide propagation loss, the region of silicon substrate under the optical devices must be removed as discussed in Sec. 1.3. This is accomplished either locally or globally with a XeF₂ etch [47], and it also enables back-side optical access to the chips. In the full systems using the optical devices [37, 125], the optical signals exit the chip through the bottom-side of the chip, and the electrical signals exit through the top-side of the chip through the normal backend metal pads. Since the optical and electrical signals exit through the bottom and the top of the chip, respectively, complex interaction between optical paths and electrical paths is avoided.

3.3.2 Single layer grating (SLG) demonstration

As previously discussed, the SLG is attractive due to its design simplicity and its ability for the same designed to be used from the top-side or the back-side of the chip. To design such a grating, two-dimensional finite-difference time-domain (2D FDTD) simulations were used. Simulation parameter sweeps exploring the width and period of the crystalline silicon shapes were utilized to
Figure 3.15: Single layer grating demonstration. (a) Optical micrograph of fabricated device. (b) TEM image of several periods of fabricated device. (c) Transmission spectrum comparing design and experiment.

Figure 3.15 shows the fabricated SLG and its characterization. The grating was designed to have a tooth width of 470 nm and a tooth gap of 250 nm giving a period of 720 nm. An optical micrograph taken from the back-side of the chip is shown in Fig. 3.15(a), and the SLG is shown in the inset. The grating is connected to a single-mode waveguide through an adiabatic taper [Fig. 3.15(a)]; the taper converts the mode from the fundamental mode of a wide, multimode waveguide to that of a single-mode waveguide.

The results of the experimental characterization of the SLG, performed from the back-side of the chip, are shown in Fig. 3.15(c). The peak coupling efficiency closely matches the simulated performance and is $\approx 35\%$. The ripple in the measured spectrum is from reflection off of the grating which forms a Fabry-Perot cavity between the two on-chip gratings. This ripple can be suppressed in a next design iteration of the device by including an anti-reflective scatterer at the front of the grating which cancels the back reflection over the wavelength range of interest. The bandwidth is smaller than expected which is likely due to weak reflections from the upper dielectric layers. The process supports 10 metal levels, and if a metal shape is not used, there is a layer of silicon nitride as

---

4 Early generation designs were made by S. Urosević and K. Nammari. The present design represents the most advanced generation and was carried out with M. Wade (author) in collaboration with Rajesh Kumar.
shown in Fig. 3.15(b). TEM images were used to measure the fabricated dimensions of the grating and the upper level dielectrics, and the dimensions were used to simulate the device. When the upper level dielectrics are included in the simulation, the bandwidth narrows as seen in the red trace in Fig. 3.15(c). Overall, this device was sufficient to test prototype optical devices (modulators, detectors, filters, etc.), but a more efficient grating coupler was needed to build energy-efficient systems.

3.4 Breaking the vertical symmetry: antenna-array inspired designs

Peak coupling efficiency from chip-to-fiber depends on two key aspects: 1) directionality of the radiation from the chip and 2) the radiation overlap with the optical fiber mode. To achieve high upwards or downwards directionality, the device must be vertically asymmetric. Although the output beam overlap with the optical fiber mode can be optimized using apodized grating designs, we chose a uniform grating for this work for design simplicity and relaxed fabrication tolerance; as is shown, high efficiency is still achieved in the second generation of the device.

The transistor layers offer a convenient set of layers to design highly directional gratings. The c-Si and the p-Si can be independently patterned (within DRC constraints), thus offering a near ideal design space for building highly efficient grating couplers. The relative positions of the c-Si and p-Si scatterers are designed to control the directionality of the radiation. A simple model to intuitively guide the design and behavior of grating couplers is that of phased array antennas familiar to radio frequency (RF) engineers. In a phased array, the radiating elements are volume current sources of radiation with certain amplitudes and phase relationships. In our case, the radiating volume currents are created by perturbations to the refractive index, where the wave incident on the grating in the on-chip waveguide sees the grating as a perturbation and produces a volume polarization current distribution \[30\]. It is this current distribution that is viewed as the phased array whose radiation pattern we wish to design. In the context of high index contrast, this picture is only approximate and actual design needs to be done with rigorous numerical tools.

Figure 3.16 presents the basic concept used \[27\,28\] to guide the design of the DLG for
directionality. Figure 3.16(a),(b) illustrate how two scatterers can be placed relative to each other to control whether the resulting radiation pattern exhibits upwards or downwards directivity, where directivity is defined as $P_{\text{up}}/P_{\text{down}}$. To illustrate the concept, the scatterers are offset by $\lambda/4$ in both the vertical and horizontal directions. If they are placed as shown in Fig. 3.16(a), the light will destructively interfere in the downward direction which results in upwards directivity. The alternate configuration is shown in Fig 3.16(b) which results in downwards directivity.

This simple concept can be mapped directly onto the c-Si (source/drain region of the transistor) and the polysilicon (gate of the transistor) in an advanced SOI CMOS process. The point scatterers in Figs. 3.16(a),(b) are approximated by rectangular shapes created using the transistor layers as shown in Fig. 3.16(c). By controlling the relative in-plane positions of the gate and the body regions, the directionality of the radiation can be optimized for either the up or the down direction. Using a band structure solver \cite{84}, Fig. 3.16(d) shows the simulated directivity of a representative unit-cell in our fabrication process. As expected from the phased array antennas model, negative offsets correspond to upwards directionality, and positive offsets correspond to downwards directionality. Thus, the phased-array antennas model can be used as a guide for design. In an actual design, the widths of the two layers must be optimized in conjunction with their positional offsets.

This section presents two generations of a uniform (non-apodized) vertical grating design that achieves large directionality and high coupling efficiency. The first generation of the design, presented in Sec. 3.4.1, demonstrated the basic concept that the crystalline silicon and polysilicon could be patterned independently to build vertical grating couplers. The first generation device, however, suffered from reduced coupling efficiency due to gaps in the polysilicon shapes required by foundry design rule checks.

\footnote{Several previous “generation” designs (2010-present) exist but did not show substantial light transmission and were deemed non-functional.}
3.4.1 Dual layer grating (DLG) first generation design

As previously mentioned, the system demonstrations couple light in and out of the back-side of the chip which requires a downward grating design. The first attempt at a downward design is summarized in Fig. 3.17. This device illustrates one of the challenges in meeting DRC rules that were designed for electronic devices – for polysilicon strips longer than several microns, the strip must be segmented into multiple smaller strips. To meet this rule, the polysilicon was segmented as shown in Fig. 3.17(a),(b). The gaps between the polysilicon shapes deteriorated the output optical beam as shown by the 3D FDTD simulations in Fig. 3.17(d),(e). Figure 3.17(d) shows the electric field profile at the output plane at the back-side of the chip for the ideal design with no gaps between polysilicon shapes. This design achieves $\approx -1.8$ dB coupling efficiency in simulation, as shown in Fig. 3.17(f). The gaps between polysilicon shapes decreases the peak coupling efficiency to $\approx -3.1$ dB which is close to what was measured in experiment. After showing that the polysilicon

---

6 Designs led by Kareem Nammari, and experimental characterization done by Mark Wade (author) in collaboration with Rajesh Kumar.
gaps created a critical performance issue for the vertical grating couplers, we secured a DRC rule waiver that allowed longer continuous strips of polysilicon. This means that the infringing design rule violation was not one that would detrimentally affect the transistor process performance or yield. If the design rule could not be waived, the polysilicon gaps could be randomized in a next design iteration and their effect on the output radiation could be designed to be minimal\footnote{Thanks to Prof. Kelvin Wagner for the idea to randomize the polysilicon gaps.}.

Figure 3.17: Summary of first generation dual layer grating. (a) Layout view. (b) Optical micrograph of fabricated device. (c) TEM image showing several periods of fabricated device. 3D FDTD simulation showing magnitude of the electric field at the output plane (bottom of chip) (d) without p-Si gaps (e) and with p-Si gaps showing significant distortion in the output field. (f) Comparison between measured and simulated transmission spectra. For $\lambda > 1200$ nm, the measured spectrum shows increasing ripple which indicates large reflections at the interface between the adiabatic waveguide taper and the vertical grating coupler.

In summary, the SLG [Sec. 3.3.2] and DLG experimental demonstrations shown in this section confirmed that the front-end transistor layers can be utilized to build vertical grating couplers. The
SLG is convenient for early prototype testing when devices are tested from both the top-side of the chip as well as the back-side of the chip. The first demonstration of the DLG suffered from decreased coupling efficiency due to DRC constraints, but the second iteration of the device, presented in the next section, removed the constraint and achieves excellent performance that rivals state-of-the-art custom devices.

3.4.2 Second generation DLG design

Two-dimensional FDTD simulations were used to further optimize, as compared to the first generation design, the widths of the c-Si and p-Si as well as the relative offset between the layers. The period of the device was controlled to achieve the desired peak wavelength and output angle, and lastly, an anti-reflection tooth was added at the front of the grating to reduce the reflection to prevent unwanted Fabry-Perot ripple in the spectrum.

Figure 3.18 summarizes the simulated performance of the device. The directionality of the radiation can be seen in Fig. 3.18(a) and is simulated to be > 96% in the downward direction. The unit cell with labeled dimensions is shown in Fig. 3.18(b). The design dimensions are well within the lithographic resolution of the 45 nm process. Figure 3.18(c) shows high directivity and low reflection near $\lambda = 1310$ nm, and Fig. 3.18(d) shows the angular power spectrum of the radiation. Figure 3.18(d) shows the simulated chip-to-fiber coupling efficiency versus angle and wavelength. To calculate the coupling efficiency, the output radiation is overlapped with a 5$\mu$m MFD fiber at varying angles. The fiber angle can be used to tune the center wavelength, and > −2 dB coupling efficiency is achieved between 1240 – 1360 nm which covers the entire O-band telecommunications window. The peak efficiency is −1.06 dB, and the −1 dB and −3 dB bandwidths are 11.2 THz (65 nm) and 18.8 THz (109 nm), respectively. This efficiency and bandwidth is suitable for dense WDM as well as coarse WDM communication applications.

---

8 Work led by Mark Wade (author) in collaboration with Rajesh Kumar (design, experimental characterization) and Fabio Pavanello (experimental characterization).
Figure 3.18: Summary of simulation results of chosen grating design. (a) FDTD simulation showing majority of radiation is directed downward. (b) Zoom-in of unit cell from simulation showing dimensions of the designed grating. The period is 650 nm. (c) Reflection into waveguide and downward directivity \( \frac{P_{\text{down}}}{P_{\text{up}}} \). (d) Angular spectrum of output radiation for \( \lambda = 1310 \text{ nm} \). (e) Chip-to-fiber coupling efficiency versus angle and wavelength. (f) Peak efficiency of \( \approx -1.06 \text{ dB} \) near \( \lambda = 1320 \text{ nm} \).

### 3.4.3 Second generation DLG experimental characterization

To characterize the transmission response of the grating couplers, light is coupled into and out of the chip through two gratings. The power coupled out of the chip is normalized to the input power, and the single chip-to-fiber linear coupling efficiency is given by \( \sqrt{\frac{P_{\text{out}}}{P_{\text{in}}}} \). The gratings are connected by an on-chip waveguide that is short enough to allow neglecting propagation losses. Lensed fibers with a \( \approx 5 \mu\text{m} \) spot size are used, and the gratings were designed to radiate a 5 \( \mu\text{m} \) beam into air. A similar design can target emission into silica for packaging strategies where the space between the coupler and the fiber is filled with silica or an index matching fluid or epoxy.

Figure 3.19 summarizes the experimental results. An optical micrograph of the fabricated device is shown in Figure 3.19(a). A 25 \( \mu\text{m} \) long adiabatic taper connects the grating to a single-mode waveguide. The anti-reflection (AR) tooth can be seen at the front of the grating. The dimensions (width of scatterer and placement relative to adiabatic taper and grating) of the AR
Figure 3.19: (a) Optical micrograph of the grating coupler. The single-mode waveguide is connected to the grating through a parabolic adiabatic taper. An AR tooth is placed at the front of the grating to minimize reflections. (b) Comparison between experiment and simulation at 11° coupling angle. 4th-order polynomial fit is used to calculate bandwidths. (c) Simulated and (d) measured transmission response as a function of $\lambda$ and angle. (e) Comparison between simulated and measured peak coupling efficiencies versus fiber coupling angle. (f) Comparison between simulated and measured peak wavelengths versus fiber coupling angle.

Tooth are varied and simulated to minimize reflection over the band of operation as shown in Fig. 3.18(c). Figure 3.19(b)-(f) compares the experimental results to simulation. The peak coupling efficiency, seen in Fig. 3.19(b), is $-1.1$ dB which is close to simulation, and there is virtually no ripple, indicating that the AR tooth is successfully suppressing reflections from the grating. The $-1$ dB and $-3$ dB bandwidths are 14 THz (80 nm) and 23 THz (133 nm), respectively. The measured bandwidths are 25% larger than predicted in simulation. This is possibly due to thinner than expected silicon and silicon nitride layers which result in a lower group index of the waveguide. A lower group index results in larger bandwidths as discussed earlier in this chapter [see Eq. 3.17 and illustrated in Fig. 3.5]

Transmission spectra were taken versus fiber angle and compared to simulation as shown in Fig. 3.19(c),(d) where the peak efficiency is normalized at each angle to more easily compare
the angular dependence. The angular dependence shows good agreement between experiment and simulation. The general trend is as expected – larger angles shift the center wavelength to lower values. Figure 3.19(e) shows the measured and simulated peak coupling efficiency versus angle. We estimate ±0.25 dB uncertainty on the peak efficiency measurement. The measured peak wavelength also agrees well with experiment, as seen in Fig. 3.19(f).

3.5 Manufacturability and fabrication sensitivities

Vertical grating couplers are sensitive to small dimensional variations in fabrication. The location of the peak coupling efficiency (wavelength and angle) is particularly sensitive to dimensional errors. Thus, it is important for a grating coupler design to be able to tolerate expected dimensional variations while maintaining acceptable coupling efficiencies. Depending on how the grating coupler is being used, these tolerances can be critically important. For use in laboratory prototyping where a user (or machine) is actively controlling the angle of the input fiber to optimize coupling efficiency, variations in the output angle are less critical. For viable high-volume manufacturing, fibers must be fixed to the chips through a packaging scheme. In this situation, actively optimizing the fiber coupling angle is either impossible or extremely difficult or costly; therefore, angular tolerance is critical. The same is true for positional tolerance. A user (or machine) in a lab environment can actively optimize the position of an optical fiber for coupling into a vertical grating coupler. For high-volume manufacturing, however, the position of the fiber is set through an automated packaging process, preferably passive. The packaging process has some achievable positional tolerance which ultimately affects the variability of the coupling efficiency of a given grating coupler. In this section, the fabrication tolerance of the device presented in Sec. 3.4 is explored. Due to the limited number of die we receive through multi-project wafer (MPW) fabrication runs, wafer-scale experiments are not feasible to build statistics with large sample sizes. Thus, the fabrication sensitivity is explored through simulation, and tolerances are established to guide future technology development.

In silicon fabrication processes (e.g. photolithography, e-beam lithography), the dimensional
control widely varies depending on the tools being used. The emerging standard for high-volume photonic fabrication is 193 nm immersion lithography \[113\] to achieve low dimensional variability required for device uniformity across a wafer. This is the same lithography technology used in advanced CMOS, such as the 45 nm node, and has been shown to achieve acceptable dimensional tolerances (silicon widths and gaps) across 200 and 300 mm wafers \[113\].

First, the dimensional variations of the c-Si and p-Si widths and their effect on the fiber coupling efficiency is studied. 2D FDTD simulations sweeps are used to simulate the effect of the dimensional variations. The width of the c-Si and p-Si are swept between \(\pm 150\) nm of their optimal values. The resulting coupling efficiency is plotted in Fig. 3.20(a). The coupling efficiency is more sensitive to errors in the p-Si shapes than the c-Si shapes; the c-Si and p-Si dimensions need to be between \(\approx -90\) to \(>150\) nm and \(\approx -25\) to \(90\) nm, respectively, of their nominal values to be within the \(-1\) dB contour. These tolerances are well within the dimensional control of 193 nm immersion lithography \[113\] which is used in the 45 nm SOI fabrication process \[61\]. Figure 3.20(b) shows how the optimal coupling angle changes [for each point in Fig. 3.20(a)] versus errors in the silicon widths, and the \(-1\) dB contour is repeated from Fig. 3.20(a).

Since the device dimensional tolerances presented in Fig. 3.20 are comfortably within the lithographic resolution, repeatability, and layer to layer alignment tolerances of the 45 nm process, fiber positional and angular misalignment tolerances related to packaging become more important to investigate. Figure 3.21 shows the simulated effects of fiber positional and angular misalignment. Figure 3.21(a) plots the coupling efficiency when the fiber is translated from its optimum position. The plot is normalized to the peak coupling efficiency. The \(-1\) dB contour is within \(\pm 1.4\) \(\mu\)m in the \(x\)-direction (direction of propagation in the waveguide) and \(\pm 1.2\) \(\mu\)m in the \(y\)-direction (transverse to direction of propagation). Since the vertical grating coupler studied here is a uniform grating with an approximately exponential decay, the \(x\) and \(y\) tolerances are different. Despite the abrupt exponential field decay, the equations derived for Gaussian-Gaussian coupling in Sec. 3.2 with \footnote{9} The global maximum does not occur at \((0,0)\) in 3.20(a) because a coarser design parameter sweep used in the original design simulations, which is slightly off the optimum design point.
Figure 3.20: (a) Coupling efficiency for \( \lambda = 1320 \) nm, relative to global maximum, versus dimensional errors in c-Si and p-Si widths. (b) Optimal coupling angle for each point in (a) versus dimensional errors in c-Si and p-Si widths. The \(-1\) dB contour from (a) is shown in red.

\( r = 1 \) agree well with the rigorous simulation results which allows the approximate equations to be confidently used in place of time-consuming FDTD simulations. Figure 3.21(b) shows the x (at y-misalignment=0) and y (at x-misalignment=0) coupling efficiencies from Fig. 3.21(a) as well as the predicted efficiency versus positional offset calculated with Eq. 3.39. The fiber angular tolerance is shown in Fig. 3.21(c) where the transmission is normalized to the nominal fiber angle, 13°. The \(-1\) dB roll-off occurs at \( \approx \pm 4.45^\circ \) angular offset from nominal. Equation 3.39 is again used to compare the approximate prediction for the angular tolerance versus the rigorous FDTD simulation. As shown, Eq. 3.39 agrees well with the simulation and can be used in place of full FDTD simulations with Fig. 3.13(c) showing the approximate dependence of the coupling efficiency on positional and angular offsets (assuming x and y-offsets have approximately the same effect as seen to be true in Fig. 3.21(b)).

3.6 Conclusions

The first high-performance vertical grating coupler that is suitable for building electro-optic systems was presented. We showed that independently patternable transistor layers native to
existing advanced CMOS SOI processes offer near ideal degrees of freedom for designing vertical grating couplers, and we used the antenna-array design concept as a guiding principle for vertical grating design using two layers for the first time. Although the devices presented in this chapter were built in a CMOS microelectronics process, the antenna-array design concept is broadly applicable to any fabrication process that allows two patternable layers to achieve vertical asymmetry. In comparison with other recent vertical grating coupler results, the device achieves very high coupling efficiencies while demonstrating a record 1 dB bandwidth, as shown in Fig. 3.22.

The devices presented here were uniform grating couplers that produce an exponential field pattern. The next iteration of the design is an apodized grating for better mode-matching to a fiber mode, and early apodized designs have already been explored [83]. Furthermore, designs for a more standard fiber mode size (e.g. $MFD = 2w_o = 10\mu m$) need to be explored for better compatibility with existing parts, packaging vendors, etc. The main benefit OF using a 10$\mu$m MFD mode size is the relaxed positional tolerances, as seen in Fig. 3.13. Since emerging fiber-to-chip packaging solutions are significantly more sensitive to positional offsets, the contours shown in Fig. 3.13(b,d,f), corresponding to a 10$\mu$m MFD mode size, are strongly preferred over the contours
in Fig. 3.13(a,c,e), 5 µm MFD.

Figure 3.22: Comparison to other recently demonstrated devices found in literature. The results are color-coded for CMOS compatibility. Gratings that would require severe modifications to a CMOS process (e.g. custom wafers, dimensional tolerances unachievable with photolithography, precise placement of dielectric/metal reflectors) are shown in red. Moderate modifications are shown in yellow (e.g. silicon nitride patterning and custom etch steps). Green shows results that were demonstrated on the same chip with working transistors.

The equations presented in Sec. 3.2 can be used to guide designers decisions regarding how to optimally trade-off aspects of the grating performance. As on-chip gratings mature and start to move to products in industry, the devices should be expected to be extremely optimized. For example, a designer might want to trade-off an acceptable amount of coupling efficiency to increase the positional tolerance. An increased positional tolerance will translate to faster throughput in packaging, leading to higher volumes and lower cost. Other likely trade-offs include coupling efficiency versus bandwidth for applications such as sensing and imaging.
4.1 Introduction

The devices presented in the previous chapters were developed for implementation in large-scale electro-optic systems. The main application is energy efficient, high bandwidth transceivers for chip-to-chip communications. Energy efficiency is enabled by the monolithic integration of electronics and photonics on the same physical die which allows the photonic devices to have very short wires connecting to the electronics, thus minimizing the parasitic resistances and capacitances. The parasitics impact the achievable electrical switching speed as well as the energy consumed for data transmission [79]. In this chapter, we present energy efficient electronic/photonic systems-on-chip that are capable of scaling to huge areal bandwidth densities, on the order of 1 Terabit per second leaving 1 mm$^2$ of chip area, and are suitable to meet the demands of chip-to-chip interconnects found in CPU-to-memory systems, high bandwidth switches and routers, and emerging computer architectures based on “disaggregation” which requires large bandwidth communication links spanning longer distances than are possible with electrical high-speed I/O.

4.1.1 Chip-to-chip photonic wavelength-division multiplexed communication links

High areal bandwidth density is enabled through use of compact ($\approx 10 \mu$m) silicon microring resonators for modulators (in Tx) and detectors (in Rx). Active devices, including modulators, tunable filters, and resonant detectors, can be designed with microrings, as covered in Ch. 2, to form the building blocks of communication links, and they are particularly well-suited for wavelength-
division multiplexed (WDM) communication links. Figure 4.1 shows a simplified schematic view of a chip-to-chip photonic WDM link. For clarity, only the chip 1 (Tx) to chip 2 (Rx) path is shown; in a full system, there is also a Tx-Rx path from chip 2 to chip 1. An off-chip laser supplies continuous wave (cw) light at multiple wavelengths which are coupled into the chip through a vertical grating coupler. Each wavelength is resonant with a specific microring modulator which provides individual communication channels. All of the wavelengths share the same optical waveguide and exit the chip through a vertical grating coupler into single-mode fiber. Since single-mode fiber is used, chip 2 can be large distances (several kilometers) separated from chip 1 assuming the latency can be tolerated in the application. The modulated light is incident on a grating coupler on chip 2 which directs the light into a waveguide for detection. Microrings are again used to select (demultiplex) the different wavelengths for detection. The detector can be built directly in the resonator, or the resonator can be a passive filter with the “drop” port connected to a broadband detector \[3\]. The two cases are conceptually equivalent but differ in their physical implementation and layout.

There are several important features shown in Fig. 4.1 that are required to ensure operation. First, the microring resonators must have properly spaced resonance frequencies that coincide with the cw laser comb input wavelengths. The relative resonance frequency spacing can be designed, as discussed later, but the absolute resonance frequencies are impractical to expect to know pre-fabrication. This is due to the extreme sensitivity of absolute resonance frequencies to dimensional variations in fabrication. Ignoring resolution and repeatability of lithography, the wafer-scale silicon thickness variation is too large to guarantee absolute resonance frequencies \([58, 96, 114]\). Furthermore, the thermal environment can experience large swings in temperature due to heat generated by surrounding electronic circuits under varying workloads as well as nonlinear optical effects in the optical resonators such as self-heating induced by two photon absorption and associated free absorption. The changing temperature causes dynamic instability in the resonance frequencies.

To overcome this, a resistive thermal tuner is integrated inside each microring modulator, as discussed in Ch. 2. This allows a feedback loop to be implemented with on-chip control circuitry (“ring tuning control” in Fig. 4.1). The feedback loop controls and stabilizes each individual
resonance frequency and maintains alignment with the incoming wavelengths. A system such as this allows the input laser wavelengths to drift which may allow uncooled temperature operation of the laser, which may have further overall system energy advantages.

Handling the data clock is also important. Electronic modulator drivers control the optical microring modulators to convert signals in the electrical domain into signals in the optical domain. In Fig. 4.1 five wavelengths are shown incident on the chip. In this example, four wavelengths are used for data, and one wavelength, $\lambda_3$, is used to forward the clock [38]. This scheme obviates the need for clock-data recovery (CDR) at the receiver which saves significant power and circuit
complexity. On chip 2, the clock is detected in the same manner as the other data signals, but its output is fed into the clock buffer tree that supplies the clock to the data receivers. This allows simple phase-adjust retiming across the data channels [38] which allows the detected data signals at the receiver to be realigned in time to correct for any small time distortions in transmission. In an actual WDM system based on microring modulators, there can be up to $\approx 32$ wavelengths per fiber, so using one wavelength for the clock is minimal overhead in usable data bandwidth. The maximum usable wavelengths is limited by the free spectral range of the resonant devices in the WDM path and the crosstalk between channels [104].

One critical aspect of the link is the optical power needed to transmit data with close to no bit-errors. The characteristics of each optical component impacts the required optical power from the off-chip laser. The wall-plug efficiency of lasers, the ratio of output optical power to supplied electrical power, is very low, with 10-20% efficiency considered state-of-the-art. Thus, reducing the required optical power needed to complete the link has a large impact on reducing the electrical power drawn by the off-chip laser. Figure 4.2 shows a schematic view of one communication channel that is used to derive a simple equation for the required optical power. The optical path is traced in red beginning at the laser and ending at the receiver.

The receiver has some sensitivity, $S_{rec} (\mu A)$, which is the minimum peak-to-peak photocurrent, corresponding to the current produced by optical 1 and 0 power levels, needed by the circuitry to achieve error-free operation. Thus, we will start with $S_{rec}$ and trace back through the optical path to calculate the approximate output power needed at the laser. The photodetector has a responsivity, $R_{det} (A/W)$, which, for our purposes, is defined as $i_{pd}/P_{wg}$ with units of $A/W$; $i_{pd}$ is the generated photocurrent when there is optical power $P_{wg}$ in the input waveguide. The optical wave experiences propagation loss through the straight waveguides, where $\alpha$ is the loss in $\text{dB/cm}$ and $l_{1,2}$ are the waveguide lengths. The loss coupling into and out of a grating coupler is given by $L_g$ in $\text{dB}$. Lastly, the modulator has some extinction ratio, $ER$, and insertion loss, $L_m$, as defined
Figure 4.2: Schematic for calculating optical link budget. Light is supplied by an off-chip laser and is coupled into the chip through a grating coupler with $T_g$ (dB) losses. The light propagates through a waveguide and is modulated by a microring. The modulator achieves an extinction ratio, $ER$ (dB), and attenuates the light by some insertion loss, $IL$ (dB). The light exits the chip through a second grating coupler, propagates through a fiber, and couples into the receiving chip through a third grating coupler. The detector is assumed to be resonant with responsivity, $R_{det}$ (A/W), defined as the ratio of photocurrent produced, $i_{pd}$ (A) to optical power in the waveguide, $P_{wg}$ (W). The receiver circuitry has a minimum peak-to-peak current sensitivity, $S_{rec}$ (A), required for error-free operation. The propagation losses are given by $(L_1 + L_2)\, T_p$ where $L_1$ ($L_2$) is the length traversed through chip 1 (chip 2), and $T_p$ is the propagation loss in dB/cm.

An equation for the required laser power is now written as

$$P_{laser} = 10^{\frac{1}{10}(3\, L_g + L_m + L_p)} \, 10^{\frac{ER}{10}} \, \frac{S_{rec}}{R_{det}}$$

(4.1)

For sufficiently small lengths and/or propagation loss, $L_p = (l_1 + l_2)\alpha$ can be neglected. Typical lengths are on the order of a few millimeters, and crystalline silicon waveguide propagation loss in 12SOI is $<3$ dB/cm [88]. This gives $<1$ dB for the propagation losses. In other cases, for example polysilicon waveguides, the propagation loss can be on the order of 10 to 20 dB/cm which makes the propagation loss a much more substantial portion of the total loss budget. $S_{rec}$ is a function of required bit error rate and data rate.

Equation 4.1 also shows the importance of high performance grating couplers. The simplified schematic shown in Fig. 4.2 includes three grating couplers. A more realistic system might include up to five grating couplers, depending on the configuration[1]. Thus, improving losses through grating couplers substantially decreases the amount of power needed from the off-chip laser. For example, our early grating designs in the 45 nm platform resulted in $\approx 4$ dB of loss per grating [137].

---

1 In one possible system implementation, there is only one input laser that powers all of the transceiver links. The laser light thus encounters a grating at the input and output of chip 1, the input and output of chip 2, and the again input of chip 1 to complete the link. This results in five total grating coupler losses.
whereas later generations of the device improved to \( \approx 1.1 \text{dB} \) of loss per grating \( 138 \), as presented in Ch. 3. In a system with three grating couplers, this results in an \( 8.7 \text{dB} \) \((\approx 7.4 \times)\) reduction in laser power.

### 4.1.2 Design considerations for monolithically integrated systems-on-chip

There are several details that must be considered when designing WDM links such as the one shown in Fig. 4.1. Since microrings have a periodic spectrum, all of the communication channels (wavelengths) must fit within one free spectral range (FSR) of the microring. An example 8-\( \lambda \) WDM row is shown in Fig. 4.3. The rings in Fig. 4.3(a) are incremented in radius to separate the resonance frequencies for use as isolated communication channels. The distance between two resonances is called the channel spacing. The number of channels that can fit within an FSR is limited by the minimum tolerable adjacent channel rejection; the adjacent channel rejection is determined by the linewidth of the resonance and the spacing between resonances \( 104 \). For \( N \) channels, the nominal channel spacing is \( \text{FSR}/N \).

To design the channel spacing, finite-difference frequency-domain \( 94 \) mode solver simulations are used to calculate the shift in resonance frequency as a function of the radius of the microring. Figure 4.4 shows the resulting channel spacing as a function of microring radius increment, \( dR \), such that the radius of the rings is

\[
R(i) = R_o + dR
\]

where \( i \) is the \( i^{th} \) ring in the WDM row (Fig. 4.3) and \( R_o \) is 5 \( \mu \text{m} \) corresponding to the modulator presented in Ch. 2. Figure 4.4 is used to calculate the desired channel spacing. For a nominal channel spacing of 200 GHz, the radius is incremented by 5.4 nm. The layout file, however, is rounded to a 1 nm grid, so a radius increment of 5 nm was chosen which gives a channel spacing of \( \approx 185 \text{GHz} \). The radius increment can also be calculated analytically with the following equation \( 3 \)

\[
\text{More sophisticated techniques can be used to effectively double the FSR of each microring} \ 145, \text{ but the extra phase required introduces additional system complexity.} \]
Figure 4.3: (a) Schematic of an 8-λ WDM row. This row can be either be a transmitter with resonant modulators or a receiver with resonant detectors. (b) Through port spectrum with the free spectral range (FSR) and channel spacing labeled. The usable data channels must fit within the FSR of the resonators.

\[
\frac{\delta \omega}{\omega} = -\frac{n_{ef}}{n_g} \left( \frac{\delta R}{R + \delta R} \right)
\]  

(4.3)

where \( \omega \) is the resonance frequency corresponding to a ring with radius \( R \), and \( \delta \omega \) is the change in resonance frequency corresponding to a small change in radius \( \delta R \), \( n_{ef} \) is the effective index of the waveguide, and \( n_g \) is the group index of the waveguide. Typically, mode solver simulations are used to calculate the effective index and group index, so while the analytic form in Eq. 4.3 is simple, it is more straightforward to directly calculate the change in resonance frequency as a function of the radius using a mode solver [Fig. 4.4].

Figure 4.5(a) shows the measured transmit (Tx) WDM row. All 11 channels (microrings) are

---

Thanks to Yossi Ehrlichman for the approximate analytic equation.
shown, untuned. A through port Lorentzian model is fit to each individual Lorentzian, as shown in Fig. 4.5(b). The Lorentzian fits allow the resonance frequencies and the 3 dB linewidths of each resonance to be accurately extracted. The difference in neighboring resonance frequencies, i.e. the channel spacing, is computed and compared to the design. The average channel spacing is 189 GHz with a standard deviation of 41.6 GHz. The average channel spacing is very close to the nominal designed spacing, 185 GHz. This provides confidence that, on average, the 5 nm microring radius increment was resolved in the photolithography. We can predict an upper limit for the tuning power required to align the channels to have equal channel spacing of 189 GHz. If the resonances fully span an FSR, the worst case tuning is that each channel is tuned a full channel spacing (assuming that the channels always are in order after fabrication and are detuned less than a channel spacing from their nominal resonances, as seen in Fig. 4.5). Thus, assuming a 1.6 nm/mW (≈3.5 µW/GHz) tuning efficiency, as presented in Ch. 2, the worst-case power required per channel is 661 µW. Therefore, for an 11-channel bank, it requires at most ≈7.3 mW of tuning power. This is the required power dropped across the resistive heater, so the efficiency at which the circuit drivers can deliver this power will affect the total wall-plug power required for tuning. The static tuning power is amortized across the aggregate data rate of the WDM row; assuming 10 Gb/s data rates, this gives ≈66 µW/Gb/s (66 fJ/bit) of worst-case static tuning power. There is also dynamic tuning
power required to keep the microring resonance locked to the laser during thermal fluctuations (environmental and self-heating), and this further increases the required tuning power as discussed and presented later in Fig. 4.15.

Figure 4.5(c) shows the 3 dB linewidths resulting from the Lorentzian fits. The resonances have 24-32 GHz 3 dB linewidths, and the variation is likely due to small fabrication tolerances in the coupling gaps at the input and the tap drop port (tap drop port discussed in Sec. 4.1.3). Since the coupling coefficient has an exponential dependence on the gap [see Fig. 2.8], nm-scale variations in the gap will noticeably affect the linewidth.

![Figure 4.5](image)

Figure 4.5: (a) Passive (untuned) through port spectrum of 11-λ WDM transmitter row. The extinction on all resonances is between 8-12 dB which is near critical coupling. A Lorentzian model is used to fit each resonance, as shown in (b). The 3 dB linewidth is calculated from the Lorentzian fit and is shown in (c). The gray region shows the confidence intervals from the fit.

Another important consideration for monolithically integrated systems-on-chip is the treat-
ment of metal wire routing. For circuits, the routing and connectivity is synthesized by automated place and route tools. Other than minimizing parasitic resistances and capacitances, the tools are mostly free to route wires without constraint. For optical devices, however, the metal routing must be done carefully. Metal is very lossy at optical frequencies, and it is critical to ensure that the optical modes of the waveguides and resonators avoid the metal shapes.

Figure 4.6 shows design guidelines for how close optical devices can come to metal. A mode solver simulation is used to vary the distance from a straight silicon waveguide to copper, Fig. 4.6(a). The mode solver simulation uses the geometries discussed and shown in Fig. 2.5 for \( \lambda = 1180, 1310, 1550 \) nm. Since copper has a significant imaginary part of its refractive index, the propagation constant that is calculated by the mode solver is complex. The complex part of the propagation constant is converted to an equivalent propagation loss in dB/cm through \( \frac{20 \text{imag}(\beta)}{\ln(10)} \) where \( \beta \) has units of rad/cm. Figure 4.6(b) shows the propagation loss in blue as a function of the distance between the top of the waveguide and the bottom of the copper [Fig. 4.6(a)].

![Figure 4.6](image_url)

Figure 4.6: (a) Cross-section of mode solver simulation used to calculate the propagation loss as a function of copper distance. (b) Results of mode solver simulation for \( \lambda = 1180, 1310, 1550 \) nm. The blue traces show propagation loss in dB/cm; the red traces show the approximate intrinsic quality factor corresponding to the propagation losses. The dotted line indicates an intrinsic quality factor of 1 million.

---

4 Place and route tools are commonly used in VLSI design to automatically position and connect electrical circuits. The algorithms are mature and can optimize placement and wiring to e.g. minimize parasitics.
The propagation loss is converted to an approximate quality factor of a resonator through the following equation (Appendix B in [96])

\[ Q = \left( \frac{\lambda_o \ln 10}{\pi n_g 0.2 L_{dB/cm}} \right)^{-1} \]  \hspace{1cm} (4.4)

where \( \lambda_o \) is the wavelength in meters, \( n_g \) is the group index, and \( L_{dB/cm} \) is the propagation loss in dB/cm. The group index was assumed to be \( n_g = 3.0 \) for Fig. 4.6(b).
Figure 4.7: Cross-section of 45 nm process showing the first eight metal layers and a crystalline silicon waveguide (dimensions to scale). A minimum region of 1 μm is cleared around the waveguide to avoid excess propagation losses from metals.
A good rule-of-thumb choice for intrinsic quality factor is $10^6$. Since the intrinsic quality factors of microrings in the process are $< 250k$, $Q = 10^6$ ensures that the copper does not significantly impact the total quality factor. Also, Fig. 4.6 presents a worst case scenario. The mode solver simulation assumes invariance in the direction of propagation meaning the copper is always present. In typical device layouts, a metal wire comes in close proximity to a waveguide for a very short distance (e.g. crosses over a waveguide). Thus, $Q = 10^6$ is a conservative value to establish minimum distances between waveguides and metal wires.

Figure 4.7 shows the immediate impact of the minimum copper distance. Advanced microelectronics processes have $> 10$ layers of metal interconnect, and they are organized into levels, as shown in Fig. 4.7. The lower levels have finer lithographic resolution such that they can support very small pitches between transistor source, drain, and gate connections, and they have the smallest thicknesses ($< 150$ nm). As the metal levels increase, their distance from the front-end silicon also increases.

Figure 4.7 shows that a wire should not cross over a waveguide until at least metal 4. This can be controlled by applying constraints to the automated electronic place and route software tools; regions of the layout can be specified to not allow metal wire routing on certain levels in that region. The region surrounding the photonics is constrained to route metals on metal 4 or above. The second more subtle implication is that foundry applied density shapes must also be controlled. When the layout file is submitted to the foundry, it undergoes processing which fills in empty spaces on each layer with “dummy fill” shapes. These fill shapes are present to prevent wafer dishing during chemical-mechanical polishing, and there are minimum density manufacturing design rules that must be satisfied (e.g. 20% metal 1 in a $25 \times 25 \mu m$ region). Thus, when integrating optics directly in the front-end transistor layers, the automatic fill must be disabled. There are special design layers that the designer can access while constructing the layout that prevent the automatic fill from being placed on a given layer. These layers must be used to block metal fill on at least metals 1-3, as well as automatic density fill crystalline silicon and polysilicon layers.
4.1.3 WDM transceivers

Figure 4.3 can conceptually be a transmitter (Tx) or a receiver (Rx) WDM row. The Tx row consists of microring modulators that shift their resonance back and forth to modulate data onto each laser wavelength. The Rx row consists of two units: a microring filter and a detector. There are two possible implementations: 1) the microring filter is a tunable filter with its drop port connected to a broadband detector; 2) the detector is integrated directly into the microring filter such that there is no drop port.

![Diagram of WDM transceivers](image)

Figure 4.8: Schematic of WDM compatible transceiver modulators/detectors based on microrings. The microrings have an embedded resistive thermal tuner to allow for resonance stabilization using on-chip feedback loops. (a) Detector configuration with passive microring filter drop port connected to broadband data detector and a weakly coupled port connected to a second broadband detector for thermal tuning feedback. (b) Resonant detector where photocurrent is generated directly in the resonator, and the feedback detector is implemented as a portion of the circumference of the ring. (c) Modulator with weakly coupled drop port connected to a broadband detector for thermal tuning feedback. (d) Modulator with portion of ring used as a feedback detector.

Figure 4.8 shows schematic illustrations of the different $1\lambda$ modulator and detector transceiver implementations. Fig. 4.8(a) uses the passive filter with a broadband detector attached to its
drop port, and Fig. 4.8(b) uses a resonant detector. In the resonant detector implementation, the photocurrent is generated inside the microring cavity, and similar to the microring active devices presented in Ch. 2, the cavity must be electrically contacted to extract the photocurrent. Additionally, as demonstrated later, the microrings in both implementations require a method to complete a feedback loop for on-chip circuit controlled resonance locking. This is accomplished with the integrated resistive thermal tuner and the secondary detector, labeled as the feedback detector. In Fig. 4.8(a), the feedback detector is implemented through a weakly coupled drop port. In Fig. 4.8(b), the feedback detector is integrated along a portion of the circumference of the microring detector. Alternatively, the resonant detector can use the same weakly coupled drop port scheme as shown in Fig. 4.8(a). The resonant implementation is a compact, elegant solution, but early designs used the broadband implementation for design and layout simplicity.

The microring modulator has similar configuration options, Fig. 4.8(c),(d). For wavelength stabilization, the weakly coupled drop port connects to a feedback detector which allows a control loop to maintain a resonance lock with the incoming laser wavelength. Alternatively, the feedback detector can be integrated directly in the ring at the sacrifice of a fraction of the ring circumference. In the Fig. 4.8(d) scenario, the microring cavity must have both a modulation and a detection mechanism. In Fig. 4.8(d), most of the ring is configured for modulation in the same way as the spoked-ring modulator presented in Ch. 2, but a few p-n junctions are devoted to photodetection and are wired to be reverse biased. This configuration allows the modulator site to be much smaller in size and is more robust to fabrication tolerances since there is no drop port (with some achievable gap tolerance) connected to the device. However, devices presented in system experiments later in this chapter utilize the configuration shown in Fig. 4.8(c), and the fully integrated detector is left as future research. In the resonant implementation, care must be taken to provide electrical isolation between the feedback detector and the data detector. The feedback detector is occupying a fraction of the ring that could be used for modulation, but there is no weakly coupled drop port that increases the linewidth of the resonance. Thus, there is likely an optimal configuration to implement the feedback detector (either in the ring or connected to the weakly coupled drop port).
which minimizes the effect on the modulation efficiency.

### 4.1.4 Photonic layout integration with on-chip electronics

Since we are monolithically integrating photonics and electronics on the same die, the photonics must be implemented alongside electronics in the same CAD environment. A typical design flow for the photonics devices was discussed surrounding Fig. 2.15 and at the end of this process, the photonics sub-system is ready for implementation into a larger sub-system that contains the electronics. This hierarchical approach to design and layout is common in electronics, and it is illustrated in Fig. 4.9 with a focus on the photonics.

We developed custom photonics layout codes in Cadence Virtuoso, one of the most popular electronics design automation (EDA) tools. This allows seamless integration of photonics into larger electronic-photonic systems-on-chip. The photonics designer begins with a basic set of photonic components (waveguides, tapers, bends, rings, etc.) in a device library, as shown in Fig. 4.9(a). These basic components can be combined to construct photonic devices, such as gratings, ring filters, and modulators [Fig. 4.9(b)]. Next, the photonic devices can be used to build larger photonic sub-systems, as shown in Fig. 4.9(c). As an example, WDM modulator and detector rows are shown which are built by cascading the modulator and detector devices from the lower hierarchical level. Finally, the photonic sub-systems are integrated with the electronic sub-systems and connected through metal wires [Fig. 4.9(d)]. At each of level of the hierarchy, design rule check (DRC) algorithms are run as well as verification, as discussed previously in Fig. 2.15.

The ability to integrate the photonic devices in the same CAD and EDA tools used for designing VLSI circuits is critical to enabling efficient photonic WDM communication systems. There are many nuanced challenges that must be overcome to arrive at a full chip layout (millions of transistors and thousands of optical devices) that the foundry will accept for fabrication. These challenges are discussed in more detail in [4][88][90].

---

5 The first generation of these codes was developed by Jason Orcutt and is presented in [90]. The second generation of the codes was led by Luca Alloatti with large development contributions from this author and can be found in [4].
4.2 Experimental demonstrations

Several system/sub-system level demonstrations have used the photonic devices presented in Ch. 23 and the previous sections. The following sections present results from a chip-to-chip optical link [125] and a CPU-to-memory link from the world’s first microprocessor with photonic I/O.
4.2.1 A monolithic chip-to-chip link

For a proof-of-concept demonstration of the WDM system shown in Fig. 4.1, a 1λ chip-to-chip link was demonstrated with a microring modulator, broadband detector, and integrated driver/receiver circuits.

![Diagram of chip-to-chip link](image)

Figure 4.10: (a) Optical micrograph showing a 1λ optical test site for transmitter characterization. (b) 3D render of the modulator layout. The metal wires that connect the optical device to the on-chip circuits can be seen (heater +/- and modulator anode/cathode). (c) Zoom in showing p-n junctions in the modulator. (d) Passive optical spectrum of modulator. The FSR is 3.07 THz (14.6 nm). (e) Resonance shift versus applied bias voltage resulting in 27 pm/V (5.7 GHz/V).

Figure 4.10 gives an overview of the optical device used in the on-chip transmitter. An optical micrograph is shown in Fig. 4.10(a) which shows the main features. The microring modulator is the generation 2 (G2) device as presented in Ch. 2 and a drop port is integrated to allow a feedback photodetector for thermal tuning. The device achieves ≈ 5.7 GHz/V shift which is consistent with the model presented in Sec. 2.5.1.

Figure 4.11 shows a layout view of the main features of the microring modulator used in the

---

6 This work was presented in [125] which contains detail on the on-chip circuitry used in these experiments as well as the thermal tuning feedback loop. Chen Sun led the top-level circuit builds, and this author (MW) led the photonic system designs as well as device designs for the vertical grating coupler and microring modulator. Many figures in this section are modified and reproduced from [125].
system experiments with detailed dimensions. The device is nearly identical to the device presented in Ch. 2 with the notable difference being the addition of a drop port as discussed in Fig. 4.8. The weakly coupled drop port is designed to drop $\approx 10\%$ of the optical power from the cavity to feed into a feedback receiver for wavelength locking.

![Diagram](image)

Figure 4.11: Modulator layout. (a) Crystalline silicon mask shapes with labeled dimensions. The drop port is weakly coupled and connected to a broadband photodetector used in a wavelength locking feedback loop. (b) Zoom in showing c-Si contacts to the p-type and n-type regions. (c) Layout view showing n-type, p-type, metal 1, and metal 2 mask shapes and the anode (p-type) and cathode (n-type) contacts. (d) Zoom in showing the implant regions and metal connectivity.

Figure 4.11(a) shows the mask shapes on the crystalline silicon layer that define the modulator, waveguides, and integrated thermal tuner. All of the dimensions are designed with rigorous FDTD simulations as covered in Ch. 2. Figure 4.11(b) shows a zoom in of a region of the microring modulator. The contacts on the inner sidewall can be seen with labeled width and pitch. These small features are enabled by the resolution of the 45 nm process, allowing individual contact to the p-n junctions. Figure 4.11(c) shows the cavity with p-type, n-type, metal 1, and metal 2 shapes.
The resistive heater is doped to a large n-type carrier concentration ($> 10^{19}$ cm$^{-3}$) and covered in silicide (not shown). This results in a target resistance of $\approx 400 \, \Omega$. Figure 4.11(d) shows a zoom in of the p-n junctions with metal connectivity. Metal 1 (see Fig. 4.7) is used to connect the n-type regions, and metal 2 is used to connect the p-type regions. The minimum feature on the implant masks is 230 nm, so the p-n junctions are snapped to a coarse grid to prevent design rule violations.

The Tx portion of the link was characterized with the experimental setup shown in Fig. 4.12. An off-chip, single wavelength laser is coupled into a “transmit site” which consists of a single modulator. An off-chip field-programmable gate array (FPGA) is used to communicate with and configure the on-chip digital circuits. An on-chip pseudo-random binary sequence (PRBS) generator is seeded with a value from the FPGA and generates a $2^{31} - 1$ bit binary sequence. The data signal goes to an 8-to-1 serializer which goes to a modulator driver circuit that drives the microring modulator. The modulated optical wave exits the chip through a grating coupler and is coupled into single-mode fiber, and an optical sampling oscilloscope is used to view the modulated optical signal.

![Figure 4.12: Schematic of transmitter on-chip experiment. An off-chip FPGA is used to communicate with on-chip circuits to configure experiments. The modulated optical output is viewed on a sampling oscilloscope.](image)

There is a weakly coupled drop port on the microring modulator, as discussed in Fig. 4.8, which is connected to a broadband photodetector. The generated photocurrent is sent to thermal
tuning circuitry which implements resonance locking to keep the microring modulator’s resonance frequency locked to the incoming laser light. This dynamic wavelength locking is critical for robust performance.

The thermal tuning transient characteristics of the modulator and heater digital-to-analog converter (DAC) are shown in Fig. 4.13. The DAC can deliver a maximum 2 mW of heater power which results in 2.5 nm (524 GHz) of tuning range, as shown in Fig. 4.13(b), corresponding to ≈ 50 K of temperature change. The tuning efficiency is 3.8 µW/ GHz (1.25 nm/mW) which includes the DAC power consumption. The 9-bit heater DAC has a resonance frequency control resolution of 524 GHz/2^9 bits = 1.02 GHz/LSB. The spoked-ring modulator has a thermal time constant (1−1/e) of 14.7 µs calculated using the through-port frequency response model and the transient response shown in Fig. 4.13(c) while assuming an exponential settling time for the thermal response.

Figure 4.14(a) shows an optical micrograph of the on-chip transmitter test site with several critical components labeled. Figure 4.14(b) shows a 5 Gbps optical eye diagram corresponding to a 2^{31}−1 PRBS (“PRBS31”) transmitter experiment (see Fig. 4.12). The experiment can be configured through the config I/O to transmit PRBS31 or a programmed 64-bit pattern. The status of the thermal tuner can also be monitored through the config I/O.

The thermal tuner uses bit-statistical thermal tuning which can handle non-DC balanced data. The digital controller goes through an initialization routine that finds the optimal heater power to maximize the difference between the “1” and “0” optical levels which give the transmit eye-height. The controller then uses the heater power to maintain a lock such that the eye-height stays at its optimal value. This control mechanism achieves robust microring stabilization under thermal variations in the surrounding environment, which are substantial as on-chip circuitry experiences dynamic workloads.

In addition to the optimal eye-height controller just described, self-heating cancellation is included as part of the thermal tuner feedback control. When the ratio of zeros, r_o, (averaged over N bits where N sets the bandwidth of the cancellation feedback loop) in the transmitted data changes, there is a fast (photon lifetime of the cavity) change in the time-averaged optical power
dissipated in the microring modulator. This causes fast self-heating which changes the temperature of the ring much faster than environmental temperature changes. To counteract this effect, the transmitted bit-statistics are monitored to calculate $r_o$, and sudden changes to $r_o$ result in a fast change to the target heater power already specified by the controller.

To experimentally demonstrate the necessity of the thermal tuning circuitry, the ratio of zeros, $r_o$, in the 64-bit data pattern is changed every 25 ms. Figure 4.15 shows the resulting 5 Gbps eye-diagrams from different operating modes used by the controller. When the controller is set to track the average photocurrent, Fig. 4.15(a), the eye-diagram is almost completely closed. This is due to the controller being unable to distinguish changes in the photocurrent due to $r_o$ from changes due to resonance drift. When the eye-height tracking is enabled, the eye-diagram greatly improves,
Figure 4.14: (a) Optical micrograph of on-chip transmitter characterization site showing the optical modulator and the on-chip circuitry (data drivers, tuners, controllers, data generation). (b) PRBS31 5 Gbps optical eye diagram taken with an off-chip optical sampling oscilloscope. The E.R. and I.L. is 6.5 dB and 3 dB, respectively, with a voltage swing of -0.7 V to 0.5 V. The drive inverters and the 8-to-1 serializer consume 30 fJ/bit and 140 fJ/bit, respectively.

but there is still some inter-symbol interference as shown in Fig. 4.15(b). When eye-height tracking and self-heating cancellation is enabled, the eye-diagram is completely open and free of inter-symbol interference, as seen in Fig. 4.15 and error-free. Thus, the self-heating cancellation is critical to the transmitter handling arbitrary data patterns; if the data is encoded to balance the number of 1’s and 0’s transmitted ($r_o = 0.5$), self-heating cancellation is not necessarily needed.

Figure 4.15: 5 Gbps eye diagrams with 64-bit patterns of varying ratio of transmitted zeros. (a) Average photocurrent tracking. (b) Eye-height tracking. (c) Eye-height and self-heating cancellation.

The wavelength-locked transmitter components are used to build an 11λ dense WDM (DWDM) transmitter, shown in Fig. 4.16. The GDS layout view of the DWDM row is shown in Fig. 4.16(a) with crystalline silicon and several interconnect metal layers shown. An optical micrograph taken
Figure 4.16: 11λ DWDM transmit row. (a) GDS layout view with macro slices labeled showing each transmit unit cell. (b) Optical micrograph of chip showing DWDM row, input/output gratings, cascaded microrings, and backend circuits. (c) Tuned (red) and untuned (gray) through port optical spectra of DWDM row. (d) 8 Gbps eye diagrams of each transmit slice giving $11 \times 8 = 88$ Gbps aggregate bandwidth coupled into a single fiber.

from the backside of the chip is shown in Fig. 4.16(b). The DWDM row consists of 11 transmitter circuit macros (as individually characterized in Fig. 4.14) cascaded to utilize 11 microring modulators all sharing the same through port, as schematically shown in Fig. 4.3. Although each successive microring is incremented in radius to design the nominal channel spacing [see Fig. 4.4], the resonance frequencies are slightly perturbed from their nominal design values due to fabrication tolerances, as discussed regarding Fig. 4.5. The resonance frequencies are tuned to the correct spacing using the embedded heater and the thermal tuning circuitry, and the tuned and untuned optical spectra are shown in Fig. 4.16(c). The tuned microring modulators are each individually demonstrated in transmit mode with full eye-tracking and self-heating cancellation. Figure 4.16(d) shows 11 eye diagrams each at 8 Gbps resulting in an aggregate 88 Gbps exiting the output grating if all concurrently operate. Currently, concurrent operation is not possible due to the unavailability of a multi-λ laser source in the 1180 nm wavelength range. Given the microring FSR of 3.07 THz,
the 11 × 8 Gb/s DWDM row achieves a spectral efficiency of 0.018 bits/s/Hz. The full DWDM row, which includes all devices required for operation (digital backend, thermal tuning, transceiver frontends, and optical devices) is 1500 µm × 150 µm resulting in a 391 Gb/s/mm² areal bandwidth density.

Figure 4.17: (a) Optical micrograph of SiGe photodetector. The absorbing region in the photodetector is 50 µm long. (b) Cross-section of photodetector showing doping regions. The carriers are generated in the SiGe and swept out by the reverse biased p-n junction and collected through the metal contacts. The polysilicon region is used to concentrate the optical mode away from the metal contacts to reduce propagation loss. (c) Optical micrograph showing on-chip receiver test site.

The receiver portion of the link is based on a broadband photodetector, shown in Fig. 4.17, that uses SiGe that is native in the process for silicon channel strain engineering, which enhances PMOS carrier mobility. An input grating coupler is connected directly to the broadband photodetector through a linear taper which adiabatically transitions from a crystalline silicon single-mode waveguide to a wide ridge waveguide where the polysilicon forms the ridge, as seen in Fig. 4.17(b). The optical mode overlaps with the SiGe region in which electron-hole pairs are produced and collected as photocurrent. Details on the design of this device can be found in [3,37,85].

The photodetector used in these experiments is an early generation design and provides a

\[ \text{BER checker} \]

receive site

input grating
coupler

receiver

frontend

50 µm

2:8

Figure 4.17: (a) Optical micrograph of SiGe photodetector. The absorbing region in the photodetector is 50 µm long. (b) Cross-section of photodetector showing doping regions. The carriers are generated in the SiGe and swept out by the reverse biased p-n junction and collected through the metal contacts. The polysilicon region is used to concentrate the optical mode away from the metal contacts to reduce propagation loss. (c) Optical micrograph showing on-chip receiver test site.

The receiver portion of the link is based on a broadband photodetector, shown in Fig. 4.17, that uses SiGe that is native in the process for silicon channel strain engineering, which enhances PMOS carrier mobility. An input grating coupler is connected directly to the broadband photodetector through a linear taper which adiabatically transitions from a crystalline silicon single-mode waveguide to a wide ridge waveguide where the polysilicon forms the ridge, as seen in Fig. 4.17(b). The optical mode overlaps with the SiGe region in which electron-hole pairs are produced and collected as photocurrent. Details on the design of this device can be found in [3,37,85].

The photodetector used in these experiments is an early generation design and provides a

\[ \text{BER checker} \]

receive site

input grating
coupler

receiver

frontend

50 µm

2:8

7 The SiGe photodetector designs were led by Rajeev Ram’s group at MIT including Jason Orcutt and Luca Alloatti.
responsivity of 0.023 A/W. An off-chip lithium niobate Mach-Zehnder Interferometer (MZI) traveling wave modulator is used to characterize the receiver. The experimental characterization setup is shown in Fig. 4.18(a). An FPGA is used to drive the MZI modulator up to 10 Gbps, and the data is coupled into the chip through a grating coupler that then terminates in the photodetector. The photocurrent is received by frontend circuitry which consists of a transimpedance amplifier (TIA) followed by two data samplers that sample the data signal on even and odd clock phases. The receiver bandwidth is $\approx 7$ GHz (including PD and parasitic wiring capacitance). The bit-error-rate checker provides in situ BER measurements which are read out through the config I/O and the test FPGA. Figure 4.18(b) shows the 10 Gbps optical BER eye as a function of sampling timing and receiver decision threshold. Figures 4.18(c),(d) shows the bathtub curves\(^8\) and receiver photocurrent as a function of average photocurrent for several datarates.

\(^8\) Bathtub curves are commonly used to plot the sensitivity of BER to timing variation of the signal used to trigger
tocurrent sensitivity, respectively, for three data rates. A BER of $< 10^{-12}$ is achieved at 10 Gbps with an average photocurrent sensitivity of 9 $\mu$A. The photocurrent is measured by monitoring the current flowing through the photodiode bias port.

![Diagram of chip-to-chip link](image)

Figure 4.19: 5 Gb/s chip-to-chip link experiment. (a) Schematic view of the chip-to-chip link testing. The optical power at several critical stages is shown in (b), and the receiver BER eye is shown in (c).

The transmitter and receiver are combined to demonstrate a chip-to-chip link between two identical chips that each contain Tx and Rx sites. An off-chip laser is coupled into chip 1 and the decision to determine if the bit is a “1” or a “0”. It is called a “bathtub” curve since the shape resembles the cross-section of a bathtub.
modulated at 5 Gbps by a spoked-ring modulator. The output is coupled into a single-mode optical fiber through a grating coupler. The gratings used in this system are not the best that we demonstrated in this process and have ≈4 dB of loss\textsuperscript{9}. The light is coupled into chip 2 through another grating coupler and is incident on the 0.023 A/W SiGe photodetector. Due to the limited responsivity of the photodetector and the relatively high loss of the grating couplers, an optical amplifier is necessary between the chips to complete the link. The amplifier provides ≈8 dB of gain while degrading the extinction ratio by 0.6 dB. In the next iteration of this system, the grating couplers that were presented in Ch. 3 will be used to lower the loss from 4 dB to 1.1 dB per coupler, and the amplifier between chips will not be needed. Also, using resonant photodetectors with > 0.1 A/W responsivity (tested at standalone sites on the same chip, but not in the system path) will obviate the need for an amplifier and reduce the required laser power, as discussed in Eq. 4.1. Table 4.1 summarizes the sub-system level performance of the chip-to-chip link critical components (Tx, Rx, and tuning). The chip-to-chip link was run at 5 Gb/s which was limited by the transmitter speed. The receiver, however, was tested standalone at 10 Gb/s. Thus, improvements to the transmitter speed will enable the link to run at up to 10 Gb/s with the same receiver presented here.

\textsuperscript{9} This is because of the evolution of best-performing standalone devices on each chip generation being implemented in the system on the next chip generation. Thus, although we demonstrated better performing grating couplers [see Ch. 3 on the same chip as the systems, they were not yet implemented in the system. This mitigates the risk of new device having unintended consequences in the full system.
### sub-system specification

<table>
<thead>
<tr>
<th>sub-system</th>
<th>specification</th>
<th>value</th>
</tr>
</thead>
<tbody>
<tr>
<td>transmitter</td>
<td>data rate</td>
<td>5 Gb/s</td>
</tr>
<tr>
<td></td>
<td>ring radius</td>
<td>5 µm</td>
</tr>
<tr>
<td></td>
<td>modulation mode</td>
<td>depletion</td>
</tr>
<tr>
<td></td>
<td>ring loaded quality factor</td>
<td>11,600</td>
</tr>
<tr>
<td></td>
<td>WDM channels</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>extinction</td>
<td>&gt; 6 dB</td>
</tr>
<tr>
<td></td>
<td>insertion loss</td>
<td>3-4 dB</td>
</tr>
<tr>
<td></td>
<td>driver power</td>
<td>0.15 mW, 0.03 pJ/bit</td>
</tr>
<tr>
<td></td>
<td>serializer power</td>
<td>0.70 mW, 0.14 pJ/bit</td>
</tr>
<tr>
<td></td>
<td>data rate</td>
<td>10 Gb/s</td>
</tr>
<tr>
<td></td>
<td>PD responsivity</td>
<td>0.023 A/W</td>
</tr>
<tr>
<td></td>
<td>PD + input capacitance</td>
<td>15-20 fF</td>
</tr>
<tr>
<td></td>
<td>sensitivity</td>
<td>-6.5 dBm@8Gb/s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-4.1 dBm@10Gb/s</td>
</tr>
<tr>
<td></td>
<td>normalized sensitivity</td>
<td>-22.8 dBm@8Gb/s</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-20.5 dBm@10Gb/s</td>
</tr>
<tr>
<td></td>
<td>power</td>
<td>3.48 mW, 0.35 pJ/bit</td>
</tr>
<tr>
<td></td>
<td>deserializer power</td>
<td>1.50 mW, 0.15 pJ/bit</td>
</tr>
<tr>
<td></td>
<td>tuner type</td>
<td>thermal</td>
</tr>
<tr>
<td></td>
<td>tunable range</td>
<td>524 GHz, 2.5 nm</td>
</tr>
<tr>
<td></td>
<td>max heater power</td>
<td>2 mW</td>
</tr>
<tr>
<td></td>
<td>tuning efficiency</td>
<td>3.8 µW/GHz, 1.25 nm/mW</td>
</tr>
<tr>
<td></td>
<td>control power</td>
<td>0.72 mW</td>
</tr>
</tbody>
</table>

Table 4.1: Summary of sub-system performance in chip-to-chip link. The table shows results from standalone characterization of each sub-system. The actual chip-to-chip link was run at 5 Gb/s as shown in Fig. 4.19. The receiver sensitivity is reported as measured sensitivity with the on-chip photodiode as well as “normalized sensitivity” which estimates the sensitivity of the receiver circuitry if the photodiode produces 1 A/W. This is commonly done for comparison between receiver circuits independent of the photodiode.
Figure 4.20: Thermal stress tests of chip-to-chip link. (a) Bit-errors with (blue) and without (red) thermal tuning locking. (b) Output of thermal tuner controller showing dynamic adjustment to on-chip power fluctuations shown in (c). The optical eye diagrams with tuning disabled (d) and enabled (e) show that the eye diagram is completely closed and unstable without the thermal tuner.

The thermal stability of the chip is stress-tested by ramping an on-chip supply voltage that is irrelevant to the functionality of the link. The supply voltage creates two coupled thermal “hot spots” on chip 1 that perturb the microring modulator at different time constants, $\tau_1$ and $\tau_2$. Figure 4.20 shows the behavior of the link during the thermal stress-test. The instantaneous bit-error-rate is shown in Fig. 4.20(a) and can be seen to have no errors while the locking is enabled (blue) and very high error-rates without locking (red). The heater DAC code is shown in Fig. 4.20(b) and is seen to respond to the power fluctuations plotted in Fig. 4.20(c). The transmit optical eye
digram during no thermal tuning, Fig. 4.20(d), is shown to be completely closed while the thermal tuning maintains an open eye-diagram, Fig. 4.20(e). The total change in temperature during this stress-test is estimated to be 6 K which is well within the $\approx 50$ K tuning range of the thermal tuner, as discussed in Fig. 4.13.

4.2.2 The first optically connected CPU chip

In this section, we present an operational CPU-to-memory communication link. This demonstration is the world’s first CPU chip to communicate to the outside world with light - a major milestone in demonstrating the viability of microring-based silicon photonics systems, and to date, the most complex functionality demonstrated in an silicon-based electronic-photonic system. Section 4.2.1 described a chip-to-chip transceiver monolithically integrated in the 45 nm platform, which laid the groundwork for demonstrating a functional communication link utilized by on-chip digital electronics. Thus, this section will focus on full system details and demonstrations, and Sec. 4.2.1 can be referred to for details on the sub-system functionality (transmitters, receivers, etc.).

Figure 4.21 summarizes the fabricated chip. The full chip is shown in Fig. 4.21(a). The chip contains $>70$M transistors and $\approx 850$ photonic components. The CPU is a custom dual-core RISC-V instruction set architecture [144], and the independent memory bank consists of 1 MB of static random access memory (SRAM). The optical transmitter and receiver banks, shown in Fig. 4.21(b), contain microring modulator transmitters and SiGe detectors, as discussed in Sec. 4.2.1. The individual critical photonic components are shown in Fig. 4.21(c) and are vertical grating couplers [see Ch. 3], SiGe photodetectors (Sec. 4.2.1 and [3,37,85]), and microring modulators [see Ch. 2].

To enable both optical and electrical functionality, a robust packaging strategy is required. Since the buried oxide is not thick enough to adequately confine an optical mode to the c-Si waveguides, the silicon substrate must be removed beneath all optical devices. The silicon substrate

---

10 At the time of this writing, the work presented in this section has been accepted for publication in Nature. Chen Sun led the system-level experimental demonstrations and top-level builds of the chip, and Yunsup Lee led the processor development. Mark Wade led the photonic system integration and designs for the modulator and grating couplers.
Figure 4.21: Microprocessor with photonic I/O. (a) Optical micrograph image of the fabricated 6 × 3 mm chip. The image is taken from the backside of the chip after full substrate removal. The bottom portion of the chip contains a dual-core RISC-V processor which utilizes the processor + memory transceiver banks to communicate with the 1 MB memory bank. Independent transceiver test sites are included at the top of the chip for prototyping new designs. (b) Zoom-in of the critical components in the processor + memory transceiver banks showing the WDM transmitter bank (top) and the WDM receiver bank (bottom). (c) Photonic components that comprise the communication link.

removal, however, has several beneficial effects. Since all optical input/output occurs through the backside of the chip, the electrical signals can exit through the top of the chip unconstrained, avoiding the complexity of having both optical and electrical signals exiting the chip through the same surface. Additionally, since the thermally conductive path to the substrate is removed, the thermal tuning control of microring modulators is significantly more efficient. The silicon substrate cannot be removed from the entire chip since there are portions of the chip that need a thermally conductive path for heat dissipation; namely, the dual-core processor and the SRAM.

To meet these requirements, a simple backside XeF$_2$ etch post-processing step is used.

---

11 Sen Lin performed the backside etch on these chips.
Figure 4.22 summarizes this process. When the die is received from the foundry, it is sent to an electrical packaging house to flip-chip mount the die on a printed circuit board (PCB). Since the die is flip-chip packaged, the silicon substrate is now oriented upwards, as shown in Fig. 4.22(a). Kapton tape is applied to the back of the die, covering the processor and SRAM regions, to act as an etch mask for the XeF$_2$ backside etch. The packaged chip is placed in an etch chamber where XeF$_2$ isotropically etches the exposed silicon substrate. The final packaged chip exposes the areas with photonic devices and leaves the silicon substrate covering the processor and SRAM bank, as seen in Fig. 4.22(b), allowing a heat sink to contact the substrate over the CPU and memory. This enables simultaneous operation of photonic devices and the processor and SRAM.

The experimental test setup is schematically shown in Fig. 4.23. Two chips are used: chip 1 operates in processor mode, and chip 2 operates in memory mode. In processor mode, the 1 MB memory bank is inactive, and the CPU utilizes the photonic transmitter and receiver through a memory controller. In memory mode, the CPU is inactive, and the 1 MB memory bank is active.
The memory bank is accessible through a communication interface that communicates with the photonic transmitter and receiver.

A single wavelength laser is used as the light source. It is split 50/50 and sent into input grating couplers on each chip’s transmitter site. The output of the transmitter is sent through an off-chip optical amplifier to overcome the insertion loss of the grating couplers and the low responsivity of the photodetector. As discussed in previous sections, improved grating couplers and photodetectors that have already been demonstrated as standalone devices will eliminate the need for optical amplification.

Figure 4.23: Schematic of chip-to-chip photonic communication link. The off-chip control FPGA communicates with chip 1 to configure experiments and allow user input. Chip 1 is in “processor mode” which utilizes the RISC-V processor. The processor makes read/write requests to chip 2, which is in “memory mode”, using the photonic link which is accessed through a memory controller. A 1 λ continuous wave laser is split 50/50 to power both transceiver paths. An optical amplifier is used between chips to overcome the link losses.

All communication between the chips is done optically. The RISC-V CPU on chip 1 can read and write to the memory bank on chip 2 through the optical link. To demonstrate this, programs compiled to the RISC-V instruction set are executed. Figure 4.24(a) shows the steps of executing a program over the optical link. First the off-chip FPGA uses chip 1 to write program instructions to the memory bank on chip 2. Second, the FPGA resets the CPU, and the CPU begins
executing instructions at the beginning of the memory bank. Chip 1 requests the instruction stored at the start of the memory bank, and chip 2 transmits it over the optical link to be received by chip 1. After this, the CPU executes the program that is in memory which involves writing and reading instructions and data to and from the memory bank in chip 2. The processor-to-memory and memory-to-processor links both run at 2.5 Gbps resulting in an aggregate 5 Gbps of memory bandwidth. The demonstration uses only one wavelength of light; if all 11 channels are active simultaneously (requiring an 11 λ laser), the memory bandwidth would be 55 Gbps per fiber.

Figure 4.24(b) shows several example programs executing. First, a simple “hello world with photonics!” program is executed which confirms correct operation. Second, a more stringent memory bandwidth benchmark program (STREAM) is executed which continuously reads and writes varying bit patterns to the memory bank in chip 2, and it confirms that there are no errors in operation. The final program that is demonstrated is a real-time video rendering of a teapot. The user can interact with the teapot through keyboard inputs sent through the FPGA to chip 1, and the color and light source position can be changed. Figure 4.24(c) shows one particular color and light source position for the teapot render.

The test setup at Berkeley Wireless Research Center, where the full chip-to-chip experimental characterization took place, is shown in Fig. 4.25. A real-time optical microscope image can be seen on the right side of Fig. 4.25(a) that shows the transmit microring lighting up since it is on resonance. The “eye diagrams” monitoring link communication, both processor to memory and memory to processor, are shown in Fig. 4.25(b).

The thermal environment on a CPU chip can fluctuate drastically depending on the workload of the CPU. As discussed in Sec. 4.2.1, stabilizing the microring resonance frequency is critical to error-free operation. Figure 4.26 demonstrates the microring stabilization under varying processor power consumption. The heater power can be seen responding to the fluctuating processor dissipated power in Fig. 4.26(a). Figure 4.26(b) shows the necessity of the thermal tuning. With tuning on, there are no errors over a test time of 1000 seconds. With no tuning, the instantaneous bit-error-rate fluctuates wildly as the processor varies its workload. The link was clocked at 2.5 Gbps
Figure 4.24: (a) Steps for program execution using the photonic chip-to-chip link. In step 1, the control FPGA writes the program's instructions into memory on chip 2 through direct memory access (DMA) through the memory controller on chip 1. After loading the program, the FPGA issues a reset signal to the processor, and the processor fetches the first program instruction from memory address 0x00002000 (step 2). In step 3, the processor executes the program, writing and reading to and from memory as well as reading instructions from memory. (b) Terminal output of hello world! and STREAM programs utilizing the optical link. (c) Graphical output of the teapot renderer program which pixel shades a 3D teapot using the Blinn-Phong shading model. The rendered image is output to the control FPGA which is displayed to the user.

for this test and the processor was clocked at 31.2 MHz due to a locked 1-to-80 clock frequency ratio between the processor and communication link which was a decision made to simplify engineering efforts during chip design. The processor can run at a maximum speed of 1.65 GHz when operated without the photonic link, and should be able to operate with photonics if this ratio is removed.
Figure 4.25: Laboratory setup for chip-to-chip communication testing. (a) Optical table with important components labeled. (b) Optical data signals for processor-to-memory and memory-to-processor communication links.

Figure 4.26: 1000 seconds processor stress test. (a) Processor power and resulting microring heater power for transmitter stabilization. (b) Processor power and resulting transmission error rates with thermal tuning on and off. Without thermal tuning, the communication link is practically unusable.
4.3 Conclusions

The system results in the previous sections show the first experimental demonstrations of a CPU chip to communicate with the outside world using light. The system was demonstrated using early generation optical device designs and electrical circuit designs, and there is much room for improvement. Table 4.2 gives an overview of the devices in the demonstrated CPU-to-memory system. Also shown is the measured performance of standalone transceivers on the same chip that represent the best to-date device performance. These devices will be implemented in the next iteration of these communication systems, and for example, should alleviate the need for an optical amplifier between chips.

<table>
<thead>
<tr>
<th></th>
<th>System transceivers</th>
<th>Standalone transceivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>propagation loss</td>
<td>4.3 dB/cm</td>
<td>4.3 dB/cm</td>
</tr>
<tr>
<td>grating coupler loss</td>
<td>4-6 dB</td>
<td>1.1 dB</td>
</tr>
<tr>
<td>Tx data rate</td>
<td>2.5 Gb/s</td>
<td>5 Gb/s</td>
</tr>
<tr>
<td>Extinction ratio</td>
<td>6 dB</td>
<td>&gt; 6 dB</td>
</tr>
<tr>
<td>Tx insertion loss</td>
<td>3 dB</td>
<td>3 dB</td>
</tr>
<tr>
<td>Tx power</td>
<td>0.02 pJ/bit</td>
<td>0.03 pJ/bit</td>
</tr>
<tr>
<td>Rx data rate</td>
<td>2.5 Gb/s</td>
<td>10 Gb/s</td>
</tr>
<tr>
<td>PD responsivity</td>
<td>0.023 A/W</td>
<td>0.10 A/W</td>
</tr>
<tr>
<td>Rx OMA sensitivity</td>
<td>-5 dBm@2.5 Gb/s</td>
<td>-7.2 dBm@10 Gb/s</td>
</tr>
<tr>
<td>Rx power</td>
<td>0.50 pJ/bit</td>
<td>0.30 pJ/bit</td>
</tr>
<tr>
<td>Ring tuning range</td>
<td>≈ 3.0 nm</td>
<td>≈ 3.0 nm</td>
</tr>
<tr>
<td>Heater tuning efficiency</td>
<td>1.25 nm/mW</td>
<td>1.25 nm/mW</td>
</tr>
<tr>
<td>Ring tuning control power</td>
<td>0.19 pJ/bit</td>
<td>0.14 pJ/bit</td>
</tr>
</tbody>
</table>

Table 4.2: Summary of transceiver performance in CPU-to-memory communication link and standalone transceivers. The devices in standalone transceivers will be integrated in the next iteration of the system transceivers. Optical Modulation Amplitude (OMA) sensitivity is $10 \log_{10}(P_1 - P_0)$ where $P_1$ and $P_0$ are the optical powers (mW) corresponding to a “1” bit and a “0” bit, respectively.

Next steps in system-level demonstrations include simultaneous use of WDM transceivers to communicate using several wavelengths at once. A critical challenge to doing this is the availability of multi-wavelength sources that meet the criteria for use in the system (channel spacing, power per wavelength, etc.). Although there are currently no commercial solutions available, there are
emerging products that show promise [12]. The CPU-to-memory system presented here was a “proof-of-concept” in that it used a 1 MB SRAM bank to emulate off-chip main memory. Ideally, the off-chip memory is DRAM with several GB of accessible memory to demonstrate the full power of this type of interconnect. However, no such DRAM chip is currently available, although we have made progress towards implementing photonics directly in a DRAM process as presented in Ch. 6.4 [120, 125].
Chapter  5

Advanced microring devices for on-chip communication systems

The previous chapters presented designs of microring-based active devices [Ch. 2] and grating
couplers [Ch. 3], and their experimental demonstration in large-scale electro-optic systems [Ch. 4].
The system demonstrations usher in a new paradigm in optical design; we now have the opportunity
to build optical devices directly integrated alongside complex analog/digital electrical circuits. This
presents the opportunity to explore systems and sub-systems whose functionality is an intimate mix
of electronics and photonics.

In this chapter, we explore directions in enabling a more diverse photonics platform, by
investigating new, advanced device concepts that allow efficient signal processing but would require
complex control, as the CMOS platform allows. Here, two advanced microring-based device designs
are presented. The first, presented in Sec. 5.1, presents a unique microring filter design that is
particularly well-suited for dense wavelength division multiplexed add/drop filters. In this device,
however, several phases must be controlled for the intended behavior. For this reason, it is likely to
need some kind of on-chip circuit feedback and control to be robustly deployed. The second device
is presented in Sec. 5.2. This device proposes a conceptually simple and efficient photonic circuit to
achieve wavelength conversion through two coupled microring modulators, and an intuitive design
method is presented that maps to equivalent linear time-invariant (LTI) filters making design
straightforward – unusual for a time-variant (LTV) system. Combined with on-chip photodetectors
and electrical feedback, this device could form the basis for RF oscillators and optical frequency
shifters. For both devices, the potential applications benefit from and are expanded by the intimate
integration with advanced electronics on the same chip.

5.1 Pole-zero filters for WDM multiplexers/demultiplexers

We propose integrated photonic wavelength multiplexers based on serially cascaded channel add-drop filters with an asymmetric frequency response. By utilizing the through-port rejection of the previous channel to advantage, the asymmetric response provides optimal rejection of the adjacent channels at each wavelength channel. We show theoretically the basic requirements to realize an asymmetric filter response, and propose and evaluate the possible implementations using coupled resonators. For one implementation, we provide detailed design formulas based on a coupled-mode theory model, and more generally we provide broad guidelines that enumerate all structures that can provide asymmetric passbands in the context of a pole-zero design approach to engineering the device response. Using second-order microring resonator filter stages as an example, we show that the asymmetric multiplexer can provide 2.4 times higher channel packing (bandwidth) density than a multiplexer using the same order stages (number of resonators) using conventional all-pole maximally-flat designs. We also address the sensitivities and constraints of various implementations of our proposed approach, as it affects their applicability to CMOS photonic interconnects.

5.1.1 Introduction

The integration of photonics and CMOS electronics has been an active area of research in recent years [63, 81, 86, 88, 110, 122, 159]. With cloud based computing driving the production of larger and more bandwidth intensive data centers as well as the increasing number of processors in multi-core CPUs, the need for more energy efficient and higher bandwidth density communication links between CPUs and RAM has motivated research into photonic CPU-memory communication links [10, 79] and the first generation of monolithic electronics-photonics integration [81, 86, 88]. Photonic communication links, using wavelength division multiplexing (WDM), have the potential to greatly increase the bandwidth density and energy efficiency compared to electrical links [9].

1 This section is nearly verbatim reproduced from reference [139, 140].
At the heart of many photonic communication links and network implementations are wavelength (de)multiplexers typically comprising serially cascaded microring filter stages. Microring filters have a free spectral range (FSR) that is determined by the ring circumference and the guided mode group index (i.e. dispersion). In a WDM communication link, the FSR, adjacent channel rejection, and required filter bandwidth with a certain maximum insertion loss determine how many WDM channels can fit in one FSR of the microring-based filters. The total bandwidth (and bandwidth utilization, Gbps data/GHz optical bandwidth) increases with an increasing number of WDM channels in a given optical wavelength range; for this reason, a designer would like to use higher order filter responses to permit denser channel spacing. Higher order filters, however, require a larger number of microring resonators which requires additional thermal tuning to compensate for fabrication variations and align to a WDM grid. Thermal tuning has a substantial energy cost and significantly impacts the energy efficiency of a proposed photonic link design. Therefore, it is of interest to investigate methods to achieve maximal bandwidth efficiency with a given number of resonator elements.

In this work, we propose a type of multiplexer/demultiplexer that we will call a “pole-zero” (de)multiplexer. It relies on the cascade of a number of stages of a novel filter design that enables asymmetric response shapes, which we will refer to as a “pole-zero” filter. A pole-zero (de)multiplexer enables very dense wavelength channel packing using low-order filters, denser by a factor of 2.4 than conventional Butterworth designs of the same order when using second-order stages. We choose a Butterworth response for comparison since it is a commonly used passband shape.

To design a pole-zero filter, we directly control the placement of the resonant frequencies (poles) and transmission zeros of the filter response (S-matrix element of interest) in the complex-frequency plane. Utilizing poles and zeros in the complex frequency plane is familiar in electrical and RF/microwave circuits, and has been previously explored in photonic devices \cite{69,95,96,103}. In this paper, we focus on the simplest way to achieve an asymmetric response by the introduction of a single transmission zero on one side of the passband. We present a coupling of modes in
time (CMT) model and use it to design an asymmetric filter response that lends itself to design of WDM demultiplexers with very densely packed channels. The advantage of this approach is that it provides a physically intuitive design technique that naturally leads to efficient implementations based on various criteria and constraints, as described later. The resulting asymmetric response is equivalent to that of previously studied asymmetric-response RF filters based on standing wave cavities [60]. Based on some general guidelines for designing pole-zero filters [96], we also consider all possible topologies of a second-order design. These alternative designs have advantages and disadvantages, e.g. in terms of sensitivity to various fabrication parameters, number of degrees of freedom that need to be controlled, etc.

5.1.2 Coupling of modes in time model

A CMT model [44] is used to design the filter response, namely, the shape of the passband and the location of the transmission zero. First, we show what is required in a general photonic system to achieve one finite-detuning transmission zero in the drop port, and we present a physical implementation that can achieve this. We then consider an approximate solution to the CMT equations for an \( N^{th} \)-order system. The specific device we are interested in is a 2\(^{nd} \)-order implementation for which we rigorously derive the CMT model and solve the full design equations.

Note that in the CMT model we consider a single resonance per resonant cavity; accordingly, when referring to a single pole or some number of zeros, this refers in a real cavity to a certain number per mode, i.e. per FSR of the system. We assume a narrowband approximation, i.e. that the passband is much smaller than the FSR, so that the adjacent azimuthal modes do not contribute to the same passband. However, these constraints are artificial and the same approach can be applied if a single cavity is used to supply multiple resonances that contribute to the passband, for example.
5.1.2.1 Designing a response with one transmission zero at finite detuning from the passband center

Consider an abstract photonic circuit representing a filter with one input port and two output ports as shown in Fig. 5.1(a). The two outputs are the familiar through and drop ports. Since the system has two resonances, all of the ports share the same two poles in the complex-frequency plane. The ports can have 0 to $N$ finite-detuning transmission zeros. Assuming a lossless system, we choose to constrain the system to have two real zeros in the through port to ensure 100% transmission at those frequencies in the drop-port passband by power conservation, with a second-order rolloff. We put one real zero in the drop port transfer function. The zero is placed on one side of the passband to make the response function asymmetric, for reasons that will become clear in Section 5.1.3.

Next, one must determine a physical implementation of a photonic circuit that can achieve the desired response. A simple rule can be used to determine the number of finite-detuning transmission zeros in the response function from the input to a given output port, i.e. in each $s$-parameter, $S_{j,\text{input}}$, $j \in \{\text{thru,drop}\}$. In general, the number of finite transmission zeros in each $s$-parameter is equal to $N$, the resonant order of the system, minus the minimum number of resonators that must be traversed from input to output [96]. Using this rule as a guide, the circuit shown in Fig. 5.1(b) can achieve the desired transmission response. Specifically, the resonant order of the system is
$N = 2$, and the minimum number of resonators that the light must pass through is one to the drop port, and zero to the through port.

In the drop port response, the light can take the blue path shown in Fig. 5.1(b) and bypass the second resonator. Using our general rule, this results in: $(2 \text{ poles}) - (1 \text{ minimum resonator traversed to drop}) = 1$ transmission zero in the drop port response. Similarly, we find two zeros in the through port, which create the familiar rejection band in the same way as for regular serially-coupled ring filters [18,65].

**5.1.2.2 Approximate design equations for an $N^{th}$-order system**

Because the transmission zero is placed off resonance, to enhance the drop-port response rejection band, it is possible to find a simple model for the position of the transmission zero, by assuming off-resonant excitation of the resonances in the system.

The time evolution of the mode energy amplitudes in a lossless $N^{th}$ order resonant system of serially coupled resonators can be written as $N$ first order differential equations,

\[
\begin{align*}
\frac{d}{dt} a_1 &= (j\omega_1 - r_i) a_1 - j\mu_{12} a_2 - j\sqrt{2} r_i s_i \\
\frac{d}{dt} a_2 &= j\omega_2 a_2 - j\mu_{21} a_1 - j\mu_{23} a_3 \\
\vdots \\
\frac{d}{dt} a_N &= (j\omega_N - r_d) a_N - j\mu_{(N)(N-1)} a_{N-1} - j\sqrt{2} r_d s_d
\end{align*}
\]

(5.1)

where $a_k$ is the energy amplitude in the $k^{th}$ ring, $\omega_k$ is the resonant frequency of the $k^{th}$ ring, $r_{i,d}$ are the decay rates to the input bus and drop bus, respectively, $\mu_{kl}$ is the energy coupling rate to the $k^{th}$ ring from the $l^{th}$ ring, $s_i$ is the amplitude of the input wave, and $s_d$ is the amplitude of the drop-port output wave. If the resonant frequencies of all rings are set equal, as is the case for typical square passband responses, the desired filter shape is synthesized through choice of the ring-ring couplings, $\mu_{kl}$, and the input and drop port decay rates, $r_i$ and $r_d$.

When a monochromatic input wave is sufficiently far detuned in wavelength from the passband center wavelength, the coupling in each equation is dominated by the forward coupling from
one ring to another (i.e. $|\mu(N)(N+1)a_{N+1}|/|\mu(N)(N-1)a_{N-1}| << 1$). This is because off-resonance the rings do not like to exchange energy (i.e. when the detuning is much larger than the coupling rate [46]), so coupling from ring 1 to ring 2 is weak, and back from ring 2 back to ring 1 is weaker still because it is a second-order effect in the detuning-induced suppression of coupling. Hence, in the coupling equations we can assume dominant coupling from the ring energy amplitude that is closer to the input bus. This simplifies Eq. 5.1 by completely decoupling the equations, and should perfectly recover the response in the off-resonant wings of the passband (only). Our goal is to design a circuit to achieve one finite transmission zero in the drop port of the device. Figure 5.2(a) shows an extension of Fig. 5.1(b) to achieve this for increasing order microring filters. In all of these filters, bypassing the $N^{th}$ ring with a tap at the $(N-1)^{th}$ ring coupled directly to the drop port enables the asymmetric response by ensuring a single drop-port response function zero. The weaker the tap coupling, the further detuned the transmission zero is from the passband. In the limit of zero tap coupling to the $(N-1)^{th}$ ring, the standard symmetric response is recovered.

Figure 5.2: Higher-order filters with a drop-port transmission zero: (a) a device architecture that can produce one drop-port zero in arbitrarily high order filters; (b) example response of a 2$^{nd}$-order filter using Eqs. 5.6, 5.7 and a zero placed at $\delta \omega_{zd} = 10r_1$.

Equations 5.1, with one modification, can be solved for the asymmetric response of the circuits shown in Fig. 5.2(a). The modification is an additional term that describes the direct coupling of the drop port to the $(N-1)^{th}$ ring. After further making the foregoing off-resonant approximation, i.e. that the energy amplitude $a_k$ is excited primarily by the previous energy amplitude $a_{k-1}$,
Eqs. 5.1 can be simplified to

\[
\begin{align*}
\frac{d}{dt}a_1 &= j(\omega_1 + j r_1) a_1 - j \sqrt{2 r_t} s_i \\
\frac{d}{dt}a_2 &= j \omega_2 a_2 - j \mu_2 a_1 \\
&\quad \vdots \\
\frac{d}{dt}a_{N-1} &= j \omega_{N-1} a_{N-1} - j \mu_{(N-1)(N-2)} a_{N-2} - r_t a_{N-1} \\
\frac{d}{dt}a_N &= j (\omega_{N-1} + j r_d) a_N - j \mu_{(N-1)(N-1)} a_{N-1} - j \sqrt{2 r_d s_d' e^{-j \phi}} \\
\end{align*}
\]

(5.2)

where \( r_t \) is the decay rate to the tap port, \( \phi \) is the propagation phase accumulated in the interference arm, and \( s'_d \) [see Fig. 5.3(a)] is given by

\[ s'_d = -j \sqrt{2 r_t} a_{N-1}. \] (5.3)

The output wave, \( s_d \), can be then be found from

\[ s_d = s'_d e^{-j \phi} - j \sqrt{2 r_d} a_N. \] (5.4)

Letting \( d/dt \rightarrow j \omega \) to solve for the steady state frequency response of the system, Eqs. 5.2–5.4 can be solved for the transfer function, \( S_{d,i}(\omega) \equiv s_d/s_i \) (valid off resonance)

\[
\frac{s_d}{s_i} = \frac{\mu_{N-2}}{j \delta \omega_{N-1} + r_t} \left( \prod_{k=1}^{N-3} \frac{\mu_k}{j \delta \omega_{k+1}} \right) \left[ -j \sqrt{2 r_t} \left( \frac{j \mu_{N-1} + 2 \sqrt{r_d} r_te^{-j \phi}}{j \delta \omega_N + r_d} \right) - \sqrt{2 r_t} e^{-j \phi} \right].
\]

(5.5)

The root of the numerator in Eq. 5.5 gives the frequency position of the transmission zero which, since it is off resonant, can be found from this approximate model. Setting the imaginary part of the root to zero to place the transmission zero on the real frequency axis and introducing \( \delta \omega_{zd} \) as the desired detuning from the passband (resonant) frequency to the transmission zero, two simple design equations can be derived that give the phase delay needed in the interference arm [see Fig. 5.1(b)] as well as the decay rate to the tap port:

\[
\cos \phi = \frac{\delta \omega_{zd}}{\sqrt{\delta \omega_{zd}^2 + r_d^2}} \quad (5.6)
\]

\[ r_t = \frac{r_d \mu_{N-1}}{r_d^2 + \delta \omega_{zd}^2}. \] (5.7)
The remaining decay rates and ring-ring couplings can be taken from the standard all-pole design synthesis techniques [65,78,91]. Figure 5.2(b) shows the transmission for a 2\textsuperscript{nd}-order pole-zero filter using Eqs. 5.6 and 5.7 with a design zero location of $\delta \omega / r_i = 10$.

The specific reason for our choice of this implementation is because it converges, in the limit of zero coupling at the tap, to a standard all-pole design. This means that any fabrication uncertainties introduced in the additional tap interferometer will affect only the position of the zero to first order, and will require weak coupling for substantially detuned zeros, making the design fairly insensitive to variations (or at least not considerably more so than a standard all-pole design). We will consider alternative geometries in Section 5.1.4.

5.1.2.3 Rigorous solution of the 2\textsuperscript{nd}-order filter synthesis problem

In the previous section, approximate design equations were derived, and it was shown that a finite transmission zero in the drop port can be achieved with the proper choice of the tap decay rate and the interference phase. This approximate model, however, is not applicable when it is desirable to place the zero close to the resonant frequency (since the approximate model is valid far from resonance). To derive the full design equations, we begin with a 3 × 3 system whose tap port is not connected to the drop port as shown in Fig. 5.3(a).

![Figure 5.3](image_url)

Figure 5.3: Abstract photonic circuit used to derive the T-matrix of the tapped-filter: (a) schematic of a 2-ring filter with 3 input and 3 output ports; (b) graphical representation of the drop-port zero location in the complex-$\delta \omega$ plane.
The CMT equations for the $3 \times 3$ system can be written in state-variable form $[96, 124]$: 

$$\frac{d}{dt} \vec{a} = j \vec{H} \cdot \vec{a} - j \vec{M}_1 \cdot \vec{s}_+$$

(5.8)

$$\vec{s}_- = -j \vec{M}_o \cdot \vec{a} + \vec{I} \cdot \vec{s}_+$$

(5.9)

where $\vec{I}$ is the identity matrix, and $\vec{H}, \vec{M}_1, \vec{M}_o, \vec{a}, \vec{s}_+, \text{and} \vec{s}_-$ are defined as follows:

$$\vec{H} = \begin{bmatrix}
\omega_1 + j (r_i + r_t) & -\mu \\
-\mu & \omega_2 + j r_d
\end{bmatrix}$$

$$\vec{M}_1 = \begin{bmatrix}
\sqrt{2} r_i e^{j \phi_1} & 0 & \sqrt{2} r_t e^{j \phi_3} \\
0 & \sqrt{2} r_d e^{j \phi_2} & 0
\end{bmatrix}$$

$$\vec{M}_o = \begin{bmatrix}
\sqrt{2} r_i e^{-j \phi_1} & 0 \\
0 & \sqrt{2} r_d e^{-j \phi_2} \\
\sqrt{2} r_t e^{-j \phi_3} & 0
\end{bmatrix}$$

$$\vec{a} = \begin{bmatrix}
a_1 \\
a_2
\end{bmatrix}$$

$$\vec{s}_+ = \begin{bmatrix}
s_i \\
s'_a \\
s_a \\
\end{bmatrix}$$

$$\vec{s}_- = \begin{bmatrix}
s_t \\
s_d \\
s'_d
\end{bmatrix}$$

(5.10)

The state variables $a_{1,2}$ are the energy amplitudes in rings 1 and 2; $s_i, s_a, \text{and} s'_a$ are input port wave amplitudes; and $s_t, s'_d, \text{and} s_d$ are output port wave amplitudes. $s'_a \text{ and } s'_d$ are temporary ports which will be later connected together. The ring resonant frequencies are $\omega_1$ and $\omega_2$, the ring decay rates due to coupling to waveguides are $r_i, r_t, \text{and} r_d$, and the related couplings are given by $\sqrt{2} r_i, \sqrt{2} r_t, \text{and} \sqrt{2} r_d [46, 65, 124]$. The coupling phases, $\phi_{1,2,3}$, can be set to zero without loss of generality.
With \( d/dt \rightarrow j\omega \), a transmission matrix \( \mathbf{T}_{3\times 3} \), where \( \mathbf{s} = \mathbf{T}_{3\times 3} \cdot \mathbf{s} \), can be derived,

\[
\mathbf{T}_{3\times 3} = \mathbf{I} + j\mathbf{M}_0 \cdot \left( \delta\omega \mathbf{I} - \mathbf{H} \right)^{-1} \cdot \mathbf{M}_1.
\] (5.11)

The coupling arm can now be connected. That is, port \( s'_d \) is connected to \( s'_a \) after a finite propagation distance. For our 3-port model, this eliminates one input and output port, and reduces the model to a \(2 \times 2\) T-matrix, described by

\[
\mathbf{T} = \mathbf{P} \cdot \left( \mathbf{I} - \mathbf{T}_{3\times 3} \cdot \mathbf{B} \right)^{-1} \cdot \mathbf{T}_{3\times 3} \cdot \mathbf{A}
\] (5.12)

where

\[
\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & e^{-j\phi} \\ 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}
\] (5.13)

where \( \phi (\equiv \beta L) \) is the phase accumulated in the coupling arm. The reduced transmission matrix is

\[
\begin{bmatrix} s_t \\ s_d \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \cdot \begin{bmatrix} s_i \\ s_a \end{bmatrix}
\] (5.14)

We are interested in the through port response, \( T_{11} \), and the drop port response, \( T_{21} \), given by

\[
T_{11}(\delta\omega) = \frac{-j2\sqrt{r_d r_t \mu} + e^{j\phi} \left[(jr_t - jr_i \delta\omega - j\delta\omega') \cdot (r_d + \delta\omega + j\delta\omega') + \mu^2 \right]}{-j2\sqrt{r_d r_t \mu} + e^{j\phi} \left[(r_t + jr_i + j\delta\omega - j\delta\omega') \cdot (r_d + j\delta\omega + j\delta\omega') + \mu^2 \right]},
\] (5.15)

\[
T_{21}(\delta\omega) = \frac{2\sqrt{r_i} \left[\sqrt{r_i} (r_d - j\delta\omega - j\delta\omega') + j e^{j\phi} \sqrt{r_d r_t \mu} \right]}{-j2\sqrt{r_d r_t \mu} + e^{j\phi} \left[(r_t + jr_i + j\delta\omega - j\delta\omega') \cdot (r_d + j\delta\omega + j\delta\omega') + \mu^2 \right]},
\] (5.16)

where \( \delta\omega' \) is the relative detuning of the two rings such that their resonances are at \( \omega_1 = \omega_0 + \delta\omega' \) and \( \omega_2 = \omega_0 - \delta\omega' \).

We can now require that the through-port zeros be placed on the real frequency axis to ensure 100% power transfer to the drop port (in the lossless approximation). To first find the zeros, the numerator of \( T_{11} \) is set to 0 and solved for \( \delta\omega \), giving

\[
\delta\omega_{z,\text{thru}} = \frac{1}{2} \left[ j \left( r_d - r_i + r_t \right) + \sqrt{-j8 e^{-j\phi} \sqrt{r_d r_t \mu} - \left( r_d + r_i - r_t + j2\delta\omega' \right)^2 - 4 \mu^2} \right]
\] (5.17)
Figure 5.4: Comparison of a pole-zero filter with an asymmetric response, and an all-pole Butterworth filter. In the pole-zero filter, the zero is placed at $\delta \omega / r_i = 2.3$. 
This equation produces two constraints that must be satisfied in order to have real zeros in the through port:

\[ r_d - r_i + r_t = 0 \]  \hspace{1cm} (5.18)

\[
\text{Im} \left\{ \sqrt{-j8e^{-j\phi}\sqrt{r_d r_i \mu}} - (r_d + r_i - r_t + j2\delta\omega')^2 - 4\mu^2 \right\} = 0
\]  \hspace{1cm} (5.19)

The constraint in Eq. 5.19 simplifies to

\[(r_d + r_i - r_t) \delta\omega' + 2\sqrt{r_d r_i \mu} \cos \phi = 0 \]  \hspace{1cm} (5.20)

which leads to a design equation that determines \( \delta\omega' \):

\[ \delta\omega' = -\frac{2\sqrt{r_d r_i \mu} \cos \phi}{r_d + r_i - r_t} \]  \hspace{1cm} (5.21)

We next consider the drop port response \( T_{21}(\delta\omega) \), given by Eq. 5.16. Its transmission zero is given by

\[ \delta\omega = -(\delta\omega' + jr_d) + e^{j\phi} \mu \sqrt{\frac{r_d}{r_t}} \equiv \delta\omega_{zd}. \]  \hspace{1cm} (5.22)

Since \( \phi \) is not yet determined, this equation for the zero location can be interpreted graphically as a root locus in the complex-\( \delta\omega \) plane. For various \( \phi \), the zero is located on a circle with radius \( \mu \sqrt{r_d/r_t} \) centered at \(-(\delta\omega', r_d)\), as shown in Fig. 5.3(b). This interpretation leads to a simple design equation for \( \phi \), the phase accumulated in the interference arm, to place the zero on the real frequency axis:

\[ \cos \phi = \frac{\delta\omega' + \delta\omega_{zd}}{\sqrt{\left(\delta\omega' + \delta\omega_{zd}\right)^2 + r_d^2}}. \]  \hspace{1cm} (5.23)

as well as for the tap coupling, \( r_t \),

\[ r_t = \frac{\mu^2 r_d}{\left(\delta\omega' + \delta\omega_{zd}\right)^2 + r_d^2}. \]  \hspace{1cm} (5.24)

At this point, we have fixed all degrees of freedom of the model. The total list of parameters of the device relevant in our synthesis includes \( r_i, \mu, \delta\omega_{zd}, r_t, r_d, \delta\omega' \) and \( \phi \). The first three \( (r_i, \mu, \delta\omega_{zd}) \) are chosen to be inputs to the model. The choice of \( r_i \) and \( \mu \) largely determines the passband
shape (maximally flat, equiripple, bandwidth, etc.), and these exist in all-pole (serially-coupled) ring filters. Detuning \( \delta \omega_{zd} \) is the desired location of the drop port zero. Without loss of generality, here, we choose \( r_i \) and \( \mu \) to be those of an all-pole 2\textsuperscript{nd}-order Butterworth filter. This is given by \( r_i = \mu \) \( [65, 78] \). After the three input parameters are chosen, Eqs. 5.18, 5.21, 5.23, and 5.24 are solved for \( r_t, r_d, \delta \omega' \) and \( \phi \) using the derived expressions. Figure 5.4 shows the transmission of a representative pole-zero filter, and a 2\textsuperscript{nd}-order Butterworth filter of equal 3 dB bandwidth for comparison.

In general, when the transmission zero is very close the passband, the \( r_i \) and \( \mu \) no longer give exactly the bandwidth and passband ripple that a Butterworth or Chebyshev design sets for them, but they are close enough for all practical designs that they can either be used as is or adjusted slightly to get the exact desired parameters. The fixing of the zeros ensures that the passband is not distorted and has (in principle) complete dropping of wavelengths in the passband.

5.1.3 Design of a serial demultiplexer based on asymmetric response stages

![Diagram of a serial demultiplexer](image)

Figure 5.5: Constructing a serial demultiplexer with symmetrized, densely packed passbands by using asymmetric response filters: (a) illustration of a two-channel demultiplexer; (b) Channel 1 drop port response, \( |T_{31}(\delta \omega)|^2 \); (c) Channel 1 through port response, \( |T_{21}(\delta \omega)|^2 \); (d) Channel 2 drop port response, \( |T_{61}(\delta \omega)|^2 \), showing a highly selective response due to the transmission zero on the right, and through-port extinction of the previous stage on the left.
In this section, we show the advantages that can be obtained from using filters with the asymmetric response shown in Fig. 5.4 to design an efficient serial wavelength demultiplexer. In the drop port response, the transmission rolls off more slowly than a standard 2nd-order Butterworth response on the left side of the passband, and it rolls off much faster on the right side between the center frequency and the zero location. To the right of the zero, there is again an increase in transmission. If this increase in transmission is detrimental in a particular application, e.g. as a crosstalk level, the designer must be able to set bounds on the maximum tolerable out-of-band transmission. In general, there is a tradeoff between how close a zero is to the passband (allowing a sharper rolloff) and the worst-case off-resonant rejection out of band. For the purposes of a serial demultiplexer, this will affect the adjacent channel rejection. The zero location in Fig. 5.4 was chosen to ensure a minimum 20 dB adjacent channel rejection.

Using the asymmetric-response filter as a building block, a serial demultiplexer can be designed to achieve a symmetrized response in the drop port that has fast rolloff on both sides of the center frequency. Figure 5.5 shows a 2-channel serial demultiplexer and the transfer function for Channel 1 through and drop port, $|T_{21}|^2$ and $|T_{31}|^2$, and Channel 2 drop port, $|T_{61}|^2 = |T_{64}T_{21}|^2$, where $|T_{64}|^2$ (and $|T_{31}|^2$) has the response shown in Fig. 5.5b. The through port response of the Channel 1 filter shapes the left side of the drop port response at the Channel 2 filter. This outcome is achieved when the channel spacing is set equal to the detuning of the zero from the passband center.

Figure 5.6a) shows the drop port responses from a 4 channel serial demultiplexer where each successive channel has a resonant frequency that is detuned from the previous channel’s center wavelength by the zero detuning. To make clear the advantage gained from using the pole-zero filters in comparison to conventional all-pole Butterworth filters, it is helpful to convert the normalized plots shown thus into an actual example implementation. For a filter bank that has a passband of 20 GHz defined at a 0.05dB ripple and at least 20 dB adjacent channel rejection, the multiplexer based on pole-zero filters can achieve a channel spacing of 44 GHz, or 45% bandwidth utilization. The all-pole 2nd-order Butterworth filter achieves a channel spacing of 106 GHz, i.e.
bandwidth utilization under 19%. The pole-zero filter bank gives a 2.4 times denser channel packing, i.e. higher bandwidth density, with no increase in filter order. Figures 5.6(a) and (b) show the responses of the example pole-zero and Butterworth demultiplexers based on second-order filter stages for comparison. It should be noted that although the pole-zero filters were derived from the Butterworth design, the transmission zero causes there to be a slight ripple in the passband. Comparing the pole-zero filter to a Chebyshev filter with approximately the same ripple, the channel packing is about 1.8 times denser using a pole-zero filter bank compared to a Chebyshev filter bank.

If the interference arm path length is chosen to give the correct accumulated phase and the two rings only need to be slightly tuned to line up their resonance frequencies, then the filter can be tuned with only 2 heaters on the ring resonators and still achieve reasonable extinction at the zero location. Since a benefit of this design is in part in the reduction of the number of thermal tuning elements (thermal power) compared to using a higher order filter, it is of interest to investigate both the sensitivity of the zero placing mechanism (the interferometer arm and tap coupling) to fabrication variations, and to investigate the possible geometries that can realize the proposed asymmetric response in search of the simplest, and most symmetric, possible implementation.

5.1.4 Alternative topologies for a filter response with two poles and one zero

The topology shown in Fig. 5.1 is not the only physical implementation that supports one transmission zero in the drop-port response with a finite detuning from the passband. Using the rule discussed in Section 5.1.2.1 for the number of finite-detuning zeros in a given S-matrix transfer function (i.e. in transmission to a given port), in a simple system such as this, it is straightforward to consider all device topologies that can result in one transmission zero. Figure 5.7(a) shows all of the degrees of freedom for a photonic circuit based on 2 traveling-wave resonators, that can produce a 2-pole, 1-zero response. This calls for two waveguides (2 pairs of input and output ports), with each waveguide coupled to each resonator. The designer has access to 5 couplings and 2 phases. Not all of the degrees of freedom are needed to design a response with one finite-detuning transmission
Figure 5.6: Design example demonstrating higher bandwidth density and denser channel packing in a serial demultiplexer based on asymmetric second-order filter stages: (a) example demultiplexer design using pole-zero filters shows 20 GHz passbands with 44 GHz channel spacing; (b) example design using conventional, all-pole Butterworth filters (same passband shape and bandwidth) is limited to 106 GHz channel spacing.

Abstract topology with all degrees of freedom

Figure 5.7: Proposed device topologies that support a 2-pole, 1-zero drop-port response: (a) general, abstract design with all possible (non-trivial) degrees of freedom; (b) tap-coupler implementation (analyzed in detail in this paper), (c) phased parallel-coupled-ring implementation, (d) 2-poles, 1-zero with minimal degrees of freedom. (b-d) are limiting cases of the general geometry in (a).
Figures 5.7(b)-(d) show alternate configurations with Fig. 5.7(b) being the configuration that was analyzed in detail in this paper. The configurations are grouped by which couplings they use. All of them share the characteristic that a minimum of one ring must be traversed from the input port to the drop port. Each configuration has strengths and weaknesses. Figure 5.7(b) can be viewed as a perturbation to the standard all-pole filters, so intuition used for the all-pole filter mostly applies. The drop port adjacent channel rejection is robust because a serially cascaded ring array naturally rejects signal off resonance (there is no requirement for a carefully tuned phase relationship between the energy amplitudes in ring 1 and ring 2, as is the case in some geometries like parallel-coupled rings [64]). Figure 5.7(c) has a robust through port rejection since the light is forced to pass through both of the rings, but the 2nd-order portion of the drop port rolloff is sensitive due to the necessity of the proper phase relationship between rings 1 and 2. This filter geometry has been used previously for filtering and for its high-Q states, but the responses realized either had no transmission zero [16, 64], or one in the center of the passband not usable for bandpass filtering [16, 148]. In this work, we propose that this geometry can be used to design asymmetric bandpass filters, subject to appropriate choice of the phases between the two resonators. The second configuration in Fig. 5.7(c) uses the same coupling as the first configuration in part (c). Design of the phases here requires unequal waveguide lengths and hence some adjustments to the waveguide geometry. This crossed configuration lends itself to a convenient network topology. Filter designs like this geometry, with waveguides crossed orthogonally (only), have been demonstrated [121], but they have again been used exclusively for all-pole filter design. Our work shows that in principle, pole-zero filters are realizable with different phase shifts. Figure 5.7(d) is the simplest implementation requiring only 3 coupling points which may make it more tolerant to fabrication uncertainties. The degenerate case of this implementation has also been previously studied [25]. Although Fig. 5.7(d) can achieve a transmission zero, it is forced to be between the split resonances of the supermodes of the cavity. Hence, it is not useful for the type of spectral response we are pursuing in this paper. All of these configurations have a first-order asymptotic rolloff far from the zero.
center frequency, and second order rolloff when detuned closer to the passband than the transmission zero.

It is outside the scope of this paper to consider in detail specific implementations and fabrication processes and conditions, but future work will investigate the best designs in terms of robustness to various common types of error in realistic fabrication processes such as those seen in monolithic integration efforts [86,88].

5.1.5 Experimental results

The proposed device was realized in the IBM 45 nm 12SOI CMOS process [61] with no process changes in foundry. Fig. 5.8(a) shows an optical micrograph of the fabricated chip (top metal side). Fig. 5.8(b) shows an optical micrograph of one of the fabricated devices. Each ring has a circular doped-silicon heater in the center to control the resonant frequencies, and the coupling arm has a heater to control the phase shift. Three devices were designed with different target offset frequencies from the passband center for the transmission zero. Fig. 5.9 shows measured through- and drop-port responses for the three devices. Measured values are summarized in Table 5.1, showing close agreement of design to measured data. The insertion losses are higher than expected and are likely due to the limited tunability of the individual microring resonances. The fabricated integrated heater resistance was much higher than intended (≈ 500 MΩ) which required high-voltage to be applied, ultimately limited by dielectric breakdown on-chip. Resistive heaters have been corrected.
to result in lower resistances [see Ch. 2] and would allow a much wider tuning range of this device, leading to drop port passbands closer to the nominal shape with lower insertion losses.

Figure 5.9: Measured transmission spectra of pole-zero devices. (a) Response with zero detuning at 26GHz; (b) response with zero detuning at 45.6GHz; (c) response with zero detuning at 74.4GHz.

Table 5.1: Comparison of filter design specifications with measured results

<table>
<thead>
<tr>
<th>Device</th>
<th>BW, design (GHz)</th>
<th>BW, meas. (GHz)</th>
<th>Zero location, design (GHz)</th>
<th>Zero location, meas. (GHz)</th>
<th>IL (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>40</td>
<td>41.7</td>
<td>24</td>
<td>26</td>
<td>2.38</td>
</tr>
<tr>
<td>2</td>
<td>40</td>
<td>49.5</td>
<td>40</td>
<td>45.6</td>
<td>1.88</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>44.3</td>
<td>83.7</td>
<td>74.4</td>
<td>2.87</td>
</tr>
</tbody>
</table>

5.1.6 Conclusion

We have proposed pole-zero filter designs with asymmetric filter responses, based on coupled resonators on chip. We have developed a physical model and design approach that gives insight into passband design and sensitivity. We have also proposed, and through simulations demonstrated, the benefit of a pole-zero based design in increasing the bandwidth density of serial wavelength demultiplexers, without an increase in filter order. A presented example design shows a factor of 2.4 reduction in the channel spacing, i.e. increase in bandwidth density, needed when using a pole-zero filter in comparison to an all-pole Butterworth filter of the same order. We have also shown that there is a finite number of possible implementations, at an abstract level (independent of resonator type), that can produce responses with two poles and one finite detuning transmission zero in the
drop port. These designs can be applied to many physical geometries and implementations and can be extended to apply to standing-wave cavities in a straightforward way. Some of the investigated geometries enforce restrictions on where the zeros can be placed, to the designer’s advantage or disadvantage.

The demand for very dense WDM communication links will push designers to using higher order filters; asymmetric response filters and (de)multiplexers can provide more densely packed channels using fewer resonators compared to all-pole designs. Because of the importance of energy considerations, these designs could make an impact in on-chip systems and interconnects, if simple and robust implementations can be demonstrated that are competitive in sensitivity to standard all-pole (and higher order) designs.

5.2 Wavelength conversion in coupled microring modulators and their design via an equivalent linear filter representation

We propose wavelength converters based on modulated coupled resonators that achieve conversion by matching the modulation frequency to the frequency splitting of the supermodes of the unmodulated system. Using temporal coupled-mode theory, we show that these time-variant systems have an equivalent linear, time-invariant filter representation that simplifies the optimal engineering of design parameters for realistic systems. Applying our model to carrier plasma-dispersion modulators as an example implementation, we calculate conversion efficiencies between $-5.4 \, \text{dB}$ to $-1.7 \, \text{dB}$ for intrinsic quality factors of $10^4$ to $10^6$. We show that the ratio of the resonance shift to the total linewidth is the most important parameter when determining conversion efficiency. Finally, we discuss how this model can be used to design devices such as frequency shifters, widely tunable radio frequency oscillators, and frequency combs.

Coupled microring resonators were first applied to high-order bandpass filter design \cite{65} and slow-light propagation \cite{153}, but have recently given rise to an increasingly wide range of novel de-

\footnote{This work was supported in part by DARPA POEM program award W911NF-10-1-0412; a University of Colorado Boulder/NIST Measurement Science and Engineering Fellowship; and University of Colorado College of Engineering startup funding. We thank J.M. Shainline and C.M. Gentry for helpful discussions.}

\footnote{This section is reproduced nearly verbatim from \cite{142}.}
vice concepts including advanced filter configurations [139], novel laser concepts [33], robust optical delay lines using analogies to edge states in topological insulators [42], and nonlinear classical and quantum light generation [19, 156]. Active microrings are dominant in optical modulator research and electronics-photonics integration, but the majority of work to date has focused on utilizing single resonators for efficient modulation [132, 141]. Active microrings that resonantly enhance modulation induced sideband generation have been shown to slightly increase the maximum modulation rate compared to the photon lifetime limited modulation rate [8, 133]. In [133], adjacent order resonances are used to enhance the sideband generation efficiency, but this results in a very large cavity because the free spectral range must be small enough to match the radio-frequency (RF) drive frequency. The large cavity size limits both the integration and switching energy efficiency. Coupled-cavity systems in [8] use doublet and triplet supermode resonances to suggest a similar enhancement in a smaller geometry. The previous work, however, does not provide a simple model that derives the key parameters and their scaling to enable optimized designs, and resonant modulator-based wavelength converters are not investigated. Previous research in wavelength conversion has been focused on traveling-wave modulated waveguides [109] and conversion through nonlinear processes, both resonant and non-resonant, [32, 151]. However, resonant, modulated systems have not been investigated as wavelength converters.

In this paper, we investigate the simplest coupled-cavity systems enabling single and dual sideband generation and analyze their performance as wavelength converters. We show that a coupling of modes in time (CMT) model [45] for coupled, modulated resonators (a time-variant linear system) can be mapped to an equivalent linear, time-invariant (LTI) filter, which enables straightforward design and calculation of wavelength conversion bandwidth and efficiency. The CMT equations for two coupled, lossless resonators with resonance frequencies modulated anti-symmetrically are given by

$$\frac{d}{dt} \mathbf{a} = j\mathbf{\bar{\omega}} \cdot \mathbf{a} - j\mathbf{\bar{\mu}} \cdot \mathbf{a} \quad (5.25)$$
Figure 5.10: (a) Physical implementation showing two coupled rings that are modulated anti-symmetrically. (b) Abstract representation of wavelength conversion where $b_1$ and $b_2$ are the supermode amplitudes of the unmodulated $a_1, a_2$ system. (c) Conceptual figure showing conversion from $\omega_1$ to $\omega_2$. Since the supermode system only supports two resonances, the lower generated sideband is unused. (d) $|S_{21}|^2$. Maximally flat (left), Chebyshev (middle), and optimal design with loss (right). Center frequencies are offset for clarity.
where

\[
\vec{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \quad \vec{\mu} = \begin{pmatrix} 0 & \mu \\ \mu & 0 \end{pmatrix}, \quad \vec{\omega} = \begin{pmatrix} \omega_o + \frac{\delta \omega_m}{2} \cos(\omega_m t) & 0 \\ 0 & \omega_o - \frac{\delta \omega_m}{2} \cos(\omega_m t) \end{pmatrix}
\]

where the two resonators have energy amplitudes \(a_1\) and \(a_2\) and are nominally resonant at \(\omega_o\), \(\delta \omega_m\) is the range of resonance frequency shift in each ring due to phase modulation, \(\omega_m\) is the RF modulation speed of the phase modulators, and \(\mu\) is the coupling between resonators.

The unmodulated (\(\delta \omega_m = 0\)) normalized supermodes of the coupled ring system in Eq. (5.25) are

\[
\begin{align*}
\bar{b}_1 &= \frac{1}{\sqrt{2}} (a_1 + a_2), \quad \omega_1 = \omega_o - \mu \\
\bar{b}_2 &= \frac{-1}{\sqrt{2}} (a_1 - a_2), \quad \omega_2 = \omega_o + \mu
\end{align*}
\]

where \(\bar{b}_1\) (\(\bar{b}_2\)) is the energy amplitude of the symmetric (antisymmetric) supermode with resonance frequency \(\omega_1\) (\(\omega_2\)). To simplify the model, we first solve for \(a_1, a_2\) in terms of \(\bar{b}_1, \bar{b}_2\) and substitute into Eq. (5.25) to find the dynamics for the supermode amplitudes:

\[
\frac{d}{dt} \vec{b} = j \vec{\omega} \cdot \vec{b} + j \vec{\mu} \cdot \vec{b}
\]

where

\[
\begin{align*}
\vec{b} &= \begin{pmatrix} \bar{b}_1 \\ \bar{b}_2 \end{pmatrix}, \quad \vec{\omega} = \begin{pmatrix} \omega_o - \mu & 0 \\ 0 & \omega_o + \mu \end{pmatrix}, \\
\vec{\mu} &= \begin{pmatrix} 0 & \frac{\delta \omega_m}{2} \cos(\omega_m t) \\ \frac{\delta \omega_m}{2} \cos(\omega_m t) & 0 \end{pmatrix}
\end{align*}
\]

This shows that modulating the rings in anti-phase results in coupling of the supermode resonances. A clearer picture is obtained by recasting the system in terms of supermode envelopes. By substituting \(\bar{b}_1 = B_1 e^{j(\omega_o - \mu)t}\) and \(\bar{b}_2 = B_2 e^{j(\omega_o + \mu)t}\) (i.e. factoring out the center frequency of \(\bar{b}_1\) and \(\bar{b}_2\)), we arrive at the evolution equations for the complex envelope amplitudes, \(B_1\) and \(B_2\).
To further simplify the envelope amplitude equations, each \( \cos(\omega_m t) \) coupling term is expanded into two exponential terms with arguments \( \propto (\omega_m \pm 2\mu) t \). Since the envelope amplitudes are centered around zero frequency, we keep only terms that can achieve phase matching, \( e^{\pm j(\omega_m - 2\mu) t} \), a temporal version of a similar spatial coupling approximation used for gratings [46]. The envelope amplitude equations are now

\[
\frac{d}{dt} \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} = j \frac{\delta \omega_m}{4} \begin{pmatrix} 0 & e^{j(\omega_m - 2\mu)t} \\ e^{-j(\omega_m - 2\mu)t} & 0 \end{pmatrix} \cdot \begin{pmatrix} B_1 \\ B_2 \end{pmatrix} \tag{5.27}
\]

By setting \( \omega_m = 2\mu \) for optimal conversion efficiency (i.e. the modulation rate equal to the frequency splitting between the symmetric and anti-symmetric supermode), the equations reduce to the same equations as derived for the envelope amplitudes of a static, 2-ring coupled resonator add-drop filter [65], where the modulation strength \( \delta \omega_m \) serves the role of “ring-to-ring” coupling strength.

Thus far, we have investigated conversion of energy from one resonant mode to another in a closed system. We now add ports (and loss) and consider the conversion of a wavelength incident in one port to another wavelength exiting a different port to find the transfer functions which give the efficiency of wavelength conversion.

\[
\frac{d}{dt} \bar{B} = j \bar{H} \cdot \bar{B} - j \bar{M}_i \bar{S}_+ \\
\bar{S}_- = -j \bar{M}_o \cdot \bar{B} + \bar{S}_+ \tag{5.29}
\]

where

\[
\bar{B} = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \quad \bar{H} = \begin{pmatrix} j(r_1 + r_o) & \frac{\delta \omega_m}{4} \\ \frac{\delta \omega_m}{4} & j(r_2 + r_o) \end{pmatrix},
\]

\[
\bar{M}_i = \begin{pmatrix} \sqrt{2r_1} \\ \sqrt{2r_2} \end{pmatrix}, \quad \bar{M}_o = \bar{M}_i^T.
\]

\[
\bar{S}_+ = \begin{pmatrix} S_{+1} \\ S_{+2} \end{pmatrix}, \quad \bar{S}_- = \begin{pmatrix} S_{-1} \\ S_{-2} \end{pmatrix}
\]

The decay rates, \( r_1 \) and \( r_2 \), are decay rates from the symmetric and anti-symmetric supermodes to ports 1 and 2. The losses are represented by the decay rate \( r_o \). If \( r_1 \) and \( r_2 \) need to
be different, the input coupling must be properly designed to achieve different coupling to each supermode. In the simplest case, as shown in Fig. 5.10(a), $r_1 = r_2$. The input and output ports have power normalized envelope amplitudes $\vec{S}_+\text{ and } \vec{S}_-$, respectively. Note that, since amplitudes $B_1$ and $B_2$ are at two different wavelengths, port 1 envelope amplitudes are defined around $\omega_1$ center frequency while port 2 is around $\omega_2$.

To calculate the wavelength conversion efficiency of interest, we set $S_{+2} = 0$ and send light only into $S_{+1}$. To find the steady-state solutions, we let $d/dt \rightarrow j\Delta\omega$ where $\Delta\omega$ is both the detuning of the input wave, $S_{+1}$, from the center frequency of $b_1$ and the detuning of the output wave, $S_{-2}$, from the center frequency of $b_2$. Solving for $|S_{21}|^2$ and setting $r_1 = r_2 \equiv r$ for simplicity, the “drop” port, which gives the conversion efficiency from port 1 at frequency $\omega_o - \mu$ to port 2 at frequency $\omega_o + \mu$, gives

$$|S_{21}|^2 = \frac{1}{4} \frac{r^2 \delta \omega_m^2}{(r + r_o)^2 + \left(\frac{\delta \omega_m}{4}\right)^2 - \Delta \omega^2} + 4\Delta \omega^2(r + r_o)^2$$ \hspace{1cm} (5.30)

The lossless case is given when $r_o = 0$. Standard filter synthesis for coupled resonators can now be used which gives the relationship between the ring-bus couplings (decay rates), $r_{1,2}$, and the modulation strength, $\delta \omega_m$. For a maximally-flat Butterworth filter, $r_1 = r_2 = \delta \omega_m/4$ and is shown in the left plot of Fig. 5.10(d). For a given ripple, a Chebyshev design can be achieved by setting $r_1 = r_2 < \delta \omega_m/4$ which is illustrated in the middle plot in Fig. 5.10(d). If $\delta \omega_m/4$ is close to $2r_o$ (i.e. the modulation frequency shift is on the same order as the intrinsic linewidth), the optimal solution for conversion efficiency at $\Delta \omega = 0$ is found by maximizing Eq. (5.30), after normalizing to $2r_o$, which results in $r_1 = r_2 = \sqrt{\delta \omega_m^2 + 4/4}$. For $\delta \omega_m \gg 2r_o$, this term converges to the case for the Butterworth filter. The optimal solution results in a design where $r_1 = r_2 > \delta \omega_m/4$. This case is shown in the right plot of Fig. 5.10(d). Equation (5.30) can be normalized to the intrinsic linewidth, $2r_o$, such that all variables are scaled by the losses that the cavity experiences. This reveals that $\delta \omega_m/2r_o$, the ratio of the achievable resonance shift and the intrinsic linewidth, is one key figure of merit of this design problem. We can calculate the maximum conversion efficiency as a function of $\delta \omega_m/2r_o$ by setting $r_{1,2}$ equal to their optimal values as previously discussed. The
blue trace in Fig. 5.11 shows the conversion efficiency for the two resonator case. Three points are marked on the plot which are relevant for silicon-based, carrier plasma-dispersion effect modulators discussed later. The key conclusion is that a large resonance shift with respect to the intrinsic linewidth is needed for good conversion efficiencies.

![Figure 5.11: Wavelength conversion efficiency as a function of modulation strength normalized to intrinsic bandwidth.](image)

The same analysis can be performed for three coupled resonators to explore possible conversion efficiency improvements due to dual sideband conversion. A physical model of three coupled rings is shown in Fig. 5.12(a). For simplicity, the input coupling is not shown as it requires more complex excitation of the supermodes as compared to the two resonator case. The supermode with resonance frequency $\omega_0$ has energy in the outer rings, $a_1$ and $a_3$. Thus, the outer rings are modulated antisymmetrically, and energy can be transferred to the lower and upper frequency supermodes. Following the same analysis presented for the two resonator case, we arrive at the equivalent conceptual linear filter model as shown in Fig. 5.12(b). In this case, we have a choice of how to achieve conversion between the supermodes. If we couple into $B_1$ and convert to $B_3$, this is a multi-step conversion process through $B_2$, as shown in Fig. 5.12(c), which results in lower conversion efficiency as shown in the green trace in Fig. 5.11. Alternatively, we can couple into $B_2$ which results in a single step conversion process to both $B_1$ and $B_3$. This conversion results in the
same efficiency as the two resonator case as shown by the yellow trace in Fig. 5.11. $|S_{12}|^2$ and $|S_{32}|^2$ for the 3-ring structure (yellow trace) are 3 dB below $|S_{21}|^2$ for the 2-ring structure (blue trace) due to the output power being split between two waves. Since there is no increase in conversion efficiency when using the three resonator implementation and there is significant complexity added, the two resonator system is the more appealing implementation. The only benefit to using the three resonator implementation is it achieves a larger frequency difference between the lower and upper supermodes for a given coupling strength between the resonators. However, the resonator coupler strength will not be a limiting factor in actual designs; usable splitting will be limited first by modulation speed.

To estimate the conversion efficiency in realistic systems, we consider carrier plasma-dispersion effect modulators that are widely used in silicon photonics (e.g. [147]). With this type of modulator, there is a tradeoff between the achievable resonance shift and the broadening of the intrinsic linewidth due to the losses introduced by the implants. The ratio of the resonance shift and the intrinsic linewidth is the most important parameter in determining conversion efficiency as previously discussed and shown in Fig. 5.11. We can derive a simplified model of carrier plasma-dispersion modulators to predict the maximum conversion efficiencies achievable with a given passive intrinsic quality factor and varying the p and n type implant concentrations. Using perturbation theory and assuming a modulator with azimuthal symmetry, the resonance shift for a given mode due to a perturbation to the permittivity is given by

$$\frac{\delta \omega}{\omega_o} = -\frac{1}{2} \frac{\int \bar{e}^* \cdot \delta \epsilon \cdot \bar{e} \, dA}{\int \bar{e}^* \cdot \epsilon \cdot \bar{e} \, dA}$$

(5.31)

where $\delta \omega$ is the resonance frequency shift, $\omega_o$ is the unperturbed resonance frequency, $\bar{e}$ is the cross-sectional electric field mode profile, $\delta \epsilon$ is the permittivity perturbation, and $\epsilon$ is the permittivity of the unperturbed resonator. To simplify the equations, we assume a constant electric field mode profile, and we replace $\delta \epsilon$ with $1/2(\delta \epsilon_p + \delta \epsilon_n)$, with the 1/2 factor coming from an assumed symmetric p-n junction. With these simplifications, Eq. (5.31) becomes

$$\frac{\delta \omega}{\omega_o} = -\frac{1}{4} \frac{\delta \epsilon_p + \delta \epsilon_n}{\epsilon}$$

(5.32)
By introducing the quality factor, $\omega_o/\delta \omega_{3dB}$, and converting to a perturbation in refractive index, Eq. (5.32) can be converted to

\[
\frac{\delta f}{\delta f_{3dB}} = -\frac{1}{2n} [\delta n_p(N_a) + \delta n_n(N_d)] Q(N_a, N_d, Q_o)
\]  

(5.33)

where $\delta f$ and $\delta f_{3dB}$ are the resonance frequency shift and the 3dB linewidth, respectively, $n$ is the unperturbed real refractive index, $\delta n_p(N_a)$ ($\delta n_n(N_d)$) is the real perturbation of the refractive index due to acceptor (donor) concentration $N_a$ ($N_d$), and the total quality factor is given by $Q(N_a, N_d, Q_o) = [(Q_p(N_a)^{-1} + Q_n(N_d)^{-1})/2 + Q_o^{-1}]^{-1}$. $Q_p(N_a)$ and $Q_n(N_d)$ are quality factors due to absorption from the implants, and $Q_o$ is the intrinsic (passive, undoped) quality factor of the resonator.

Using experimental fits for $\delta n(N_{a,d})$ and $\delta \alpha(N_{a,d})$ from [123] and converting the absorption, $\alpha$, to quality factors $Q_p(N_a)$ and $Q_n(N_d)$, curves for the achievable resonance shift normalized to the intrinsic linewidth are generated using Eq. (5.33) and shown in Fig. 5.13(b)-(d). The combination of Figs. 5.11 and 5.13(b)-(d) gives the estimation of conversion efficiency for plasma-dispersion

Figure 5.12: (a) Three coupled rings with the outer rings modulated antisymmetrically. (b) Abstract representation of wavelength conversion where $b_1$, $b_2$, and $b_3$ are the supermode amplitudes of the unmodulated $a_1$, $a_2$, $a_3$ system. (c) Conceptual figure showing conversion between the three supermode resonances. To match the conversion efficiency of the two ring case, light is sent in to the $\omega_2$ resonance (green) and $\omega_1$ (red) and $\omega_3$ (blue) are generated.
Figure 5.13: (a) Simplified model for carrier-plasma modulator calculations. In the approximations used, the cavity goes from a negligibly small depletion width (top) to fully depleted (middle) which introduces a complex perturbation to the refractive index (bottom). Calculated resonance shift normalized to the total 3 dB linewidth for intrinsic quality factors (b) $Q_o = 10^4$ (c) $Q_o = 10^5$ (d) $Q_o = 10^6$.

Effect resonant modulators in silicon. In generating Fig. 5.13(b)-(d), we assumed the cavity is fully depleted as shown in Fig. 5.13(a). However, for large implant concentrations ($> 10^{19}$ cm$^{-3}$) and large asymmetry in $N_a$ and $N_d$, the applied voltage required to fully deplete the cavity becomes unrealistic for some applications. Thus, the particular constraints in a given design must be considered jointly with Fig. 5.13(b)-(d). The maximum normalized frequency shift ranges from $\delta f/\delta f_{3dB} = 3$ to 10 in Fig. 5.13. This maps directly to the x-axis in Fig. 5.11 and corresponds to conversion efficiencies between $-5.4$ dB to $-1.7$ dB which is an upper-limit for resonators with intrinsic quality factors ranging from $10^4$ to $10^6$. For example, using Eq. 5.30 with optimal values for $r$ and setting $r_o = 0$, incoming light carrying amplitude modulated data can be frequency shifted by 100 GHz with a 132 GHz 3 dB conversion bandwidth (i.e. the center frequency can shift by 100 GHz while
supporting datarates up to $\approx 190\text{ Gbps}$) assuming a 1 V swing and the modulator has similar performance as shown in [132]. The conversion bandwidth improves with a larger voltage swing; thus, the electrical circuit design is a critical aspect of the overall performance. Assuming $Q_0 = 10^5$, Fig. 5.11 shows the conversion efficiency is close to $-3\text{ dB}$. Some practical considerations must be addressed in the design of a real device including two photon absorption (TPA) and thermal instability. TPA ultimately limits the input optical power. Closed-loop feedback circuitry [126] which thermally locks the resonators to the incoming light is a promising way to mitigate thermal instability.

We have presented the concept of wavelength conversion in modulated coupled resonators and showed how the conversion process maps to an equivalent time-invariant filter. Using the filter model, coupling strengths can be optimally chosen to design the spectral shape of the conversion, and the conversion efficiency can be easily calculated while including loss. Combining this analysis with approximate calculations for plasma-dispersion effect modulators, conversion efficiencies ranging from $-5.4\text{ dB}$ to $-1.7\text{ dB}$ were predicted for typical silicon photonic modulators. These efficiencies are suitable for many applications including novel types of optoelectronic RF oscillators built around this device, and frequency shifters. Higher efficiencies would be necessary for applications like comb generation. For such applications, electro-optic (Pockel’s effect) modulators may be suitable which do not intrinsically couple the real and imaginary parts of the perturbed permittivity (e.g. based on strain-induced $\chi^{(2)}$ or EO polymers). Designs based on the presented model can be readily implemented in fully integrated platforms [37,141] where the tight integration with electronics makes the antisymmetric modulation trivial, and taking advantage of integrated photodiodes and fast transistors allows for applications requiring closed-loop feedback. More generally, this work shows that some time-variant systems can be reformulated to take advantage of the decades of advances in synthesis techniques for LTI systems.
Chapter 6

Outlook on monolithic integration

The previous chapters presented devices and systems monolithically integrated in IBM’s 45 nm SOI CMOS process. This fabrication process has been used to build several commercially successful microprocessor products (PlayStation 3 Cell processor, Nintendo WiiU Espresso processor, Power7 processors, all by IBM), and it is also the process that, at the time of this writing, is used to build the processors in five out of ten of the world’s most powerful supercomputers. The work in this thesis contributed to demonstrations of (1) high performance active photonics in this platform, (2) electronic-photonic subsystems (Tx,Rx), and (3) the first microprocessor with optical I/O, demonstrating the largest scale of electronic-photonic integration to date by far, thus validating the zero-change CMOS approach. The CPU is also likely to be a major milestone in computer architecture. Finally, the zero-change approach allowed design of photonics in the most advanced fabrication process to date.

In this chapter, an outlook on monolithic electronic-photonic integration is given. A linear photonic crystal cavity and a correlated single-photon pair source are presented in Sec. 6.1 to show the viability of the 45 nm photonics platform to extend beyond data communication applications. In Sec. 6.3, we present the first photonic device demonstrations in IBM’s 32 nm node (waveguides, grating couplers, tunable filters, and modulators). Finally, results are highlighted from our efforts to monolithically integrate photonics into bulk silicon processes, which are potentially important as silicon photonics approaches commercial viability and high-volume production.

1 World rankings of supercomputers can be found at www.top500.org and is updated quarterly.
6.1 Extended photonic device platform demonstrations in 45 nm CMOS

The suitability of the 45 nm node to support basic photonic devices (waveguides, grating couplers, microring filters, active devices, etc.) needed to build on-chip data communications systems have been demonstrated in the previous chapters, but the process is also suitable for more intricate photonic devices as well as nonlinear and non-classical photonics applications. In this section, we present the first photonic crystals demonstrated in the 45 nm platform and their use in a configuration that allows them to be treated as traveling wave cavities suitable for WDM applications. Next, a microring-based correlated photon-pair source, based on nonlinear four-wave mixing, is presented. It is shown to produce correlated single-photon pairs (with record low pump power) – a potentially important building block to enabling non-classical photonic systems-on-chip.

6.1.1 Photonic crystals in 45 nm CMOS

Photonic crystals have been an active field of research since their conception in the late 1980’s and early 1990’s. Their realization in silicon photonics has always been plagued with fabrication tolerances and suspicion that they have limited practical use. In research, most devices are fabricated using electron-beam lithography, and little work has been done to demonstrate photonic crystals in a volume process based on photolithography due to extreme dimensional sensitivities and resolution requirements. The 45 nm process, however, offers extremely fine resolution photolithography, optical proximity correction (OPC), and carefully optimized process variations. It is of interest to demonstrate a photonic crystal platform in unmodified CMOS. That is done here for the first time.

Figure 6.1 summarizes fabricated photonic crystal nanobeam and the measured experimental results. The device is realized in the crystalline silicon region, and the design uses patterned holes in the silicon nanobeam, as shown in Fig. 6.1(a),(b), to approximate a parabolic optical potential and support Hermite-Gaussian resonant modes, Fig. 6.1(c). The optical through and drop port

---

2 Christopher Poulton led the work on photonic crystals, and Mark Wade aided in their implementation in the 45 nm platform and the experimental characterization. Details on the photonic crystals can be found in [101,102].

3 In this section, figures are modified and reproduced from [101,102] with permission from Chris Poulton.
spectra are shown in Fig. 6.1(d) showing five resonances corresponding to the first five supported longitudinal modes of the cavity. Their quality factors are extracted showing intrinsic Q’s of \( \approx 100,000 \) are obtainable for the fundamental longitudinal mode, Fig. 6.1(e). This demonstrates photonic crystal cavities as a viable building block for photonic circuits in 45 nm SOI CMOS.

![Figure 6.1](image)

Figure 6.1: (a) Layout view of photonic crystal nanobeam and transverse mode of input and output excitation waveguides. (b) Nominal design dimensions labeled showing the square hole tapering from the middle (320 nm) to the ends of the nanobeam (258 nm). (c) Simulated fundamental and third order longitudinal modes of nanobeam cavity. (d) Measured through (green) and drop (yellow) port spectra of fabricated device. (e) Zoom-in of fundamental mode resonance compared to a calculated lossless response.

One shortcoming of photonic crystal nanobeam cavities is their standing-wave characteristics which radiate bi-directionally to all coupled waveguides. In most practical communications applications, single direction traveling-wave cavities (e.g. microrings) are preferred since they result in simplified configurations (i.e. no need for isolators) and avoid unnecessary insertion loss. Thus, it is of interest to show ring-like port-separating operation in photonic crystals. Based on the photonic...
crystal nanobeam cavity shown in Fig. 6.1, a traveling-wave emulating configuration is demonstrated which behaves analogous to a traveling-wave microring. The traveling-wave configuration is an implementation realizing a concept first proposed by Manolatou et al. in 1999 \cite{70}. The fabricated device and its experimental characterization is summarized in Fig. 6.2. The device includes integrated resistors on the phase arms to achieve optimal phase differences between the cavities that are necessary for its microring equivalent operation. This demonstration shows that photonic crystal nanobeam cavities can be configured for use in traveling-wave WDM systems integrated in the 45 nm CMOS platform, as drop-in replacements for functions suitable to microring resonators.

---

**Figure 6.2:** (a) Configuration of two standing-wave cavities to result in the equivalent of a (b) traveling-wave microring cavity. (c) Optical micrograph of fabricated device and (d) 3D renders of the tuning arms. (e) Passive (untuned) spectra. (f) Phase arms tuned to approximate correct relative phase, $|\phi_1 - \phi_2| = \pi$. The phase is also tuned to result in (g) the lowest drop port insertion loss and (h) the highest drop port suppression.
6.2 Nonlinear four-wave mixing and correlated single photon pair source

In this section, we present experimental results that form the basis for beyond classical photonic devices and experiments and demonstrate that the 45 nm platform is a promising platform for nonlinear and non-classical applications. First, stimulated four-wave mixing (FWM) in a silicon microring cavity is demonstrated and shown to achieve wavelength conversion efficiencies as high as -42 dB at -2 dBm pump power in the waveguide. This result is an important step towards nonlinear optical signal processing in CMOS (amplification, wavelength conversion, etc.). Second, spontaneous four-wave mixing is shown to produce correlated single photon pairs which is an important building block for quantum photonic systems.

Figure 6.3 shows a summary of the fabricated chip and device. The silicon microring is realized in the sub-100 nm thick crystalline silicon layer, and the geometry is designed to achieve minimal dispersion between adjacent pump resonances used for the signal and idler photons, which is critical for efficient conversion. The radius of the ring is 22 \( \mu \text{m} \), as shown in Fig. 6.3(c), and the waveguide width is 1.08 \( \mu \text{m} \) shown in Fig. 6.3(d).

The passive optical spectrum of the device is shown in Fig. 6.4(a). A Lorentzian through port model is fit to the measured data, and a total quality factor of 31,000 is extracted along with an intrinsic quality factor of 114,000. The adjacent resonances are measured to be separated from the pump resonance by 756.6 GHz and 758.4 GHz giving a difference of 1.8 GHz, which is within the linewidths of the resonances and supports FWM. Figure 6.4(b) reports the measured stimulated FWM efficiency versus pump power compared to theory, where deviation at high powers are due to parasitic nonlinearities (e.g. two-photon absorption and free-carrier absorption).

The microring was also experimentally characterized as a quantum-correlated photon pair generator based on spontaneous FWM. Spontaneous FWM was tested by performing time-resolved coincidence measurements between generated photon pairs. The pump wavelength is coupled into

---

4 This work was led by Cale Gentry as part of a collaboration with NIST Boulder. The stimulated four-wave mixing results were presented in [35], and the correlated single photon results have been accepted for publication in Optica at the time of this writing, and a pre-publication version can be found in [34]. The author’s contribution to this work was the implementation in the 45 nm process (layout, waveguide design, coupling gaps) of the microring designed by Cale Gentry.
Figure 6.3: Optical micrograph of fabricated 3×3 mm chip. (a) Top side view. (b) Bottom side view. (c) Zoom-in of silicon microring device used for stimulated and spontaneous FWM experiments. (d) Simulated optical mode of the FWM microring. The outer radius is 22 \( \mu \)m, and the waveguide width is 1.08 \( \mu \)m.

Figure 6.4: (a) Passive through port transmission spectrum. (b) Stimulated FWM efficiency.

the device after a series of two C-band filters used to suppress the laser spontaneous emission noise at the signal and idler wavelengths. The generated photon pairs are coupled out of the chip through a vertical grating coupler and filtered through cascaded telecom filters to further suppress noise from the pump wavelength. The generated signal and idler photon pairs are then sent to
81% and 87% efficient WSi superconducting single photon detectors (SNSPDs) \cite{73}, respectively. One useful measure of correlated photon pairs is the coincidence-to-accidentals ratio (CAR) which measures the detected ratio of correlated photon pairs to non-correlated photon pairs in a given timing window. Figure 6.5 shows the measured CARs versus pump power for two measurement time windows, 240 ps and 611 ps, respectively.

![Coincidences-to-Accidentals Ratio (CAR)](image.png)

Figure 6.5: Coincidence-to-accidental ratio measurements with time windows of (a) 240 ps and (b) 611 ps showing CARs greater than 50 are achievable.

This demonstration is the first source of quantum-correlated photons in a commercially available advanced CMOS process, and it opens the door to integration of non-classical states of light with high-performance electronics on a single chip. The 45 nm SOI CMOS process may provide a platform for cryogenic experiments because transistors in it have been demonstrated to operate without carrier freezeout down to 2.8 K \cite{136} as well as reconfigurable feed-forward systems \cite{105,117} where detection of a heralding signal photon can be processed by on-chip digital circuits and route the corresponding idler photon for a specific operation.
6.3    Scaling zero-change photonics: an active photonics platform in a 32 nm microelectronics SOI CMOS process

One of the reasons for success of the CMOS microelectronics technology has been its “scaling” – the ability to port the transistor technology (nearly) seamlessly to the next smaller process node. Photonic devices are limited in size by the optical wavelength. Nevertheless, it is interesting to ask whether our CMOS photonics platform and approach can be ported to a next process node, to take advantage of next-node transistors and process improvements. Although the photonics would not get smaller as transistors do, much of the CMOS photonics design infrastructure (parameterized cells used in layouts, design rule management, design levels used for photonics) could be reused, allowing minimal development. In custom processes, this is an issue. For example, Luxtera spent years developing a custom monolithic process with Freescale Semiconductor [41], but when the process became too slow (as other CMOS technology kept advancing), Luxtera decided not to develop another process and chose to use hybrid integration (connect an electronics chip to a photonics chip) instead. The zero-change approach makes the transition between technology nodes feasible.

In this section, we report the first migration of zero-change CMOS photonics integration to a more advanced technology node. As CMOS microelectronics processes continue to evolve, the “zero-change” approach to photonic integration will have its limits. In particular, fully depleted SOI (FDSOI) technologies have approached < 10 nm thick crystalline silicon active regions [29] which is inhospitable to integrated optics. Fundamentally, this is due to the waveguide dimensions being set largely bey the wavelength of operation. Thus, optical devices do not scale in geometry the same way that electronic devices benefit from dimensional scaling. A side-effect of the smaller technology nodes, however, is increased lithographic resolution in the silicon and dopant masks, and many photonic devices benefit from this (e.g. the “spoked-ring” modulator presented in Ch. 2), as well as better control of process variability which is critical to photonics.

Despite being unlikely to scale the zero-change approach into FDSOI nodes, IBM’s 32 nm
The 32 nm node is a partially-depleted SOI technology with cross-sectional dimensions very similar to the 45 nm process. Performance benefits of the 32 nm compared to the 45 nm node include \( \approx 20\% \) higher transistor density \([40]\) and the introduction of embedded DRAM (eDRAM), a dense on-chip memory \([143]\). One critical difference is the introduction of a metal gate in place of the polysilicon gate. This prevents the type of grating design that was shown in Ch. \([3]\) since the metal gate cannot act as an individually patternable scattering element. There is, however, polysilicon used in the gate formation \([40]\) such that minor modifications to the process would allow efficient grating designs.

The following experimental results represent a proof-of-concept demonstration that the 32 nm node is suitable for photonic integration.

Fig. \([6.6]\) gives an overview of the fabricated chip and devices. In particular, Figs. \([6.6](a)\) and \([b]\) show an optical micrograph of the top side and bottom side of the chips. To enable optical waveguides, a XeF2 etch is used to remove the silicon substrate which gives access to the front-end device layers as seen in Fig. \([6.6](b)\) \([47]\). In partially depleted SOI processes, the thickness of the crystalline silicon layer that forms the transistor body is \(< 100 \) nm. This layer is exploited for photonic devices, and it is significantly thinner than in typical photonics SOI wafers where the active Si layer thickness is usually 220 nm. This difference translates in a less confined mode as it can be seen in Fig. \([6.6](c)\) where the fundamental TE-like mode of a typical waveguide in this process is shown at \(\lambda=1280 \) nm. Although devices were demonstrated at both \(\lambda=1280 \) nm and 1550 nm, here we focus on the 1280 nm devices for clarity. Figures \([5.10](d)-(f)\) show optical micrograph images of the devices corresponding to experimental data in Fig. \([6.7]\).

In Fig. \([6.7]\) experimental data for several photonic devices are reported including a vertical grating coupler achieving \(\approx 40\%\) peak coupling efficiency per grating near \(\lambda=1260 \) nm, a 4-channel widely tunable filter bank with \(\approx 1.7 \) nm/mW tuning efficiencies and \(< 3 \) dB insertion loss, and a depletion-mode modulator \([120]\). The modulator achieves modest performance as shown in Fig. \([6.7](d)-(h)\), but it is promising as a first device demonstration in this platform, and it is the first demonstration of a “wiggler” modulator in SOI CMOS (the device was originally demonstrated in bulk CMOS as covered in Sec. \([6.4.2]\) and \([120]\)). The drive voltage for these measurements was
≈ −5 V to +0.5 V, and the device shows up to 8 Gb/s with 5-6 dB I.L. and 1.5-3 dB E.R. Higher than expected propagation loss was measured (≈ 35 – 40 dB/cm) which limits the performance of the filters and the modulator and is possibly caused by an unintentional mid-level implant in all devices. Future efforts will focus on achieving low propagation loss (similar to those in [88] is expected) which will allow optimized designs and higher performance photonic devices.

### 6.4 Photonics in bulk silicon CMOS processes

Up to this point, all of the photonics work presented has been in SOI CMOS fabrication processes. However, the majority of electronic integrated circuits that are fabricated today are based on bulk silicon wafer processes due to wafers being significantly cheaper than SOI wafers. Bulk silicon processes differ from SOI processes in several key aspects. First, the only patternable silicon is polysilicon rather than crystalline silicon. This poses an immediate challenge to building
photonic devices since the optical loss of polysilicon is complex and strongly depends on processing steps that affect the polycrystalline grain structure (size and orientation) and associated electronic defect states in the band gap. The activation of dopants as a function of implant concentrations is $< 100\%$ and follows a well-known, highly nonlinear function [116]. This requires fine-tuning the fabrication process steps to achieve carrier concentrations required for efficient active photonic devices, which is extremely difficult in the context of a full CMOS fabrication flow.

Despite these difficulties, bulk silicon is so ubiquitous that the long term prospects of silicon
photonics might depend on its compatibility with polysilicon. One immediate motivation for integrating photonics into bulk silicon is that all of today’s DRAM is built in bulk silicon processes. This is due to the extreme cost sensitivity of the DRAM commercial market, and for viable economical reasons, the cost of a fully finished DRAM wafer is cheaper than the starting cost of an SOI wafer. Since CPU-to-DRAM interconnects are a bottleneck in today’s computing systems, integrating photonics directly onto DRAM is of interest. In the following sections, several key results are highlighted from our work to enable photonic devices in bulk silicon platforms.

Micron Technology’s flash periphery process (180 nm technology node) is used as the platform for prototyping silicon photonic integration. The process was previously used to build flash memory products, and minimal process changes were introduced in order to not diverge the process far from a commercially viable process (e.g. low-cost mask additions, no epitaxial growth, silicon-based detectors – no germanium, etc.). The first depletion-mode optical modulator in a bulk silicon platform is presented followed by the first optical chip-to-chip link in bulk silicon. The developed fabrication process is described in [74]. The bulk silicon photonics work is being continued in collaboration with the College of Nanoscale Science and Engineering in Albany, New York, based on a IBM 65 nm node, and early microring modulator results have been obtained.

6.4.1 A depletion-mode microring modulator in a bulk CMOS process

Depletion-mode microring modulators are at the core of efficient WDM communication links. The devices presented in this section were demonstrated in a Micron Technology short-flow polysilicon process that emulates a full-flow commercial DRAM process. The DRAM emulation process was subject to many of the constraints and processing steps that are found in a full-flow process (i.e. all the way through backend metals), including no crystalline silicon, no partial silicon etch, and no vertical p-n junction implants, most of which are utilized in other published microring modulators [120][131][141].

5 This work was led by Fabio Pavenello, and at the time of this writing, the results have been submitted to a peer-reviewed optics conference.
6 The results presented in this section were led by Jeffrey Shainline, and the figures are reproduced, with permission, from [120].
The device is shown in Fig. 6.8(a). To meet the constraints of the process, a cavity based on regions that support multiple transverse optical modes is used. The cavity is comprised of these multi-mode straight regions, which are electrically contacted, and single-mode waveguides used to support tight bends. A zoom-in of the multi-mode, contacted region is seen in Fig. 6.8(b) which shows a propagating optical wave simulated with 3D FDTD. In the simulation, the optical optical mode is inserted on the left side, and propagates through a non-adiabatic taper when connecting to the multi-mode region. The taper is carefully designed to excite the optimal ratio of the fundamental TE-like and the 2\textsuperscript{nd}-order TE-like transverse waveguide modes. The superposition of these two modes results in the “wiggling” pattern as shown in Fig. 6.8(b). The field pattern results in nulls on the sidewalls where the magnitude of the electric field is very close to zero. This allows electrical contacts to be placed at these positions, as seen in the upper left of Fig. 6.8(a), which minimizes optical scattering loss. The p-n junction is placed in the middle of the multi-mode region, and the modulation is achieved by applying a voltage bias that modulates the depletion width of the p-n junction, thus modulating the resonance frequency of the device (see Fig. 2.4).

The static characterization for two different “wiggler” devices is shown in Figs. 6.8(c)-(f). The passive optical spectra shown in Fig. 6.8(c) shows that the 5-period and 15-period devices achieve FSR’s of 587 THz and 1.09 THz, respectively. These FSR’s support approximately 4-6 WDM communication channels, and although they cannot support as many channels as a larger FSR device (e.g. the spoked-ring cavities presented in previous chapters), the smaller FSR results in a smaller power enhancement which result in decreased instabilities due to non-linearities. Optical micrographs of the fabricated devices are shown for the 5-period and 15-period devices in Figs. 6.8(d),(e), respectively. The rectifying I-V curves for these devices are shown in Fig. 6.8(f) showing successful formation of the p-n junctions. Hall measurements indicate that activated dopant concentrations were achieved in the ranges of $7-9 \times 10^{17}$ cm$^{-3}$ n-type and $6-8 \times 10^{17}$ cm$^{-3}$ p-type. Loaded quality factors of the 5 and 15-period devices are measured to be 5900 and 4300, respectively.

Figure 6.9 summarizes the modulation results. The resonance shift versus applied DC bias voltage is shown in Fig. 6.9(a). To get enough shift for substantial optical modulation, the voltage
swing is from -3.5 to 0.5 V where a negative voltage is reverse bias. The -3.5 to 0.5 V swing achieves 5 Gbps eye diagrams with 4.2 dB and 3.1 dB extinction ratios for the 15 and 5-period devices, respectively, as shown in Fig. 6.9(b),(c). The insertion loss is 4.0 dB and 1.5 dB for the 15 and 5-period devices, respectively. To approximate the energy consumption, $\frac{1}{4}CV^2$ is used for non-return-to-zero (NRZ) data where $C$ is the 0 V junction capacitance at the high-end of the measured dopant concentration range, and $V$ is the peak-to-peak voltage swing. With these values, the estimated energy is 60 and 160 fJ/bit for the 5 and 15-period modulators, respectively.

This demonstration was important to show that a bulk silicon process can be optimized to support photonic devices. The next important demonstration is to show that the process can simultaneously support photonic devices and transistors.

6.4.2 The first monolithic chip-to-chip link in a bulk silicon process

In this section, we present the world’s first monolithic chip-to-chip link in a bulk silicon CMOS process. A tremendous amount of work was accomplished to reach this demonstration spanning...
several research groups (Rajeev Ram at MIT, Vladimir Stojanović at Univ. of California, and Miloš Popović at Univ. of Colorado Boulder) and Micron Technology, and only the highlights are presented here. In particular, the process work can be found in [74], the system level results are given in full detail in [126], and the photodetector used in the receiver is presented in [75].

Figure 6.9 summarizes the chip-to-chip link. An off-chip laser is coupled into chip 1 through vertical grating couplers, the cw light is modulated by a ridge-type depletion-mode microring modulator. The modulator achieves a 5 Gbps modulation rate with 7.6 dB of extinction and 1.6 dB of insertion loss. The total transmitter energy (including electrical circuits) is 1.22 pJ/bit. The modulated light is coupled out of the chip through a vertical grating coupler and into a single-mode fiber. Due to high grating coupler losses (5 dB loss per coupler), an optical amplifier with 8 dB of gain is required to complete the link (a link without an amplifier was completed at 2 Gbps). The light is then split through a 90/10 power splitter where the 10% tap is used to monitor the transmitted eye diagrams on an external sampling oscilloscope. The light enters a chip 2 receive

\[ \text{Figure 6.9: (a) Resonance shift with applied DC voltage bias. (b) 5 Gbps eye diagram using 15-period device. (c) 5 Gbps eye diagram using 5-period device.} \]
waveguide through a vertical grating coupler, and the waveguide is terminated by a receiver \[75\]. The receiver uses a split photodiode for double data rate (DDR) sampling to sample the 5 Gbps data with two 2.5 Gbps receiving frontends which sample out of phase with each other. This is done to overcome the electrical speed limitations in the 180 nm equivalent process. The link was demonstrated to have better than $10^{-10}$ BER at 5 Gbps. Improvements to vertical grating couplers, using similar ideas as presented in Ch. 3, would obviate the need for an optical amplifier and greatly improve the total system energy efficiency. The total energy efficiency was measured to be 2.8 pJ/bit (Tx+Rx), neglecting the optical amplifier.

Figure 6.10: Schematic of chip-to-chip test setup and resulting eye diagrams from the split receivers showing better than $10^{-10}$ BER.

The potential WDM capabilities were demonstrated for both transmitters and receivers by individually testing each transmitter and receiver channel\[8\]. The WDM transmitter is presented in Fig. 6.11 and shows a 45 Gbps (9×5) aggregate bandwidth transmitter with individual eye diagrams of each channel shown. The WDM receiver can be seen in \[125\]. It is worth noting that 45 Gbps

\[8\] The channels could not be tested simultaneously due to the lack of a suitable laser source with multiple output wavelengths at the correct channel spacing.
is a larger aggregate bandwidth than modern X8 DDR4 DRAM chips and is distance insensitive, while DDR4 chips can only support cm-scale communication distances. This demonstration was achieved with early generation device designs, and there is much room for improvement.

Figure 6.11: Schematic of WDM transmitter testing and optical micrograph of on-chip test macros. Individual eye diagrams of each transmitter are shown indicating that the system can achieve 45 Gbps (9×5 Gbps) aggregate transmission bandwidth on a single fiber.

6.5 Conclusions

In this chapter, the viability of monolithic integration to extend to a wide portfolio of photonic devices and applications was demonstrated. Also, the first demonstration of photonic devices in a 32nm SOI CMOS technology node was shown which was the first time zero change CMOS photonics integration has been ported to a more advanced node. Scaling past partially-depleted SOI will likely prove difficult, and is an interesting next phase of research for this work. One promising direction might be backend polysilicon integration which would allow the photonics to be integrated in different processing steps than the transistors [62]. Devices and systems were also demonstrated in a bulk silicon platform which represents a potentially important milestone in the long-term economic viability of silicon photonics.
Chapter 7

Conclusions

7.1 Summary of major accomplishments

Silicon photonics has been a rapidly growing field of research since the late 1990’s, and it has promised to impact several areas of technology including low energy, high bandwidth chip-to-chip interconnects. Microring resonators have long been recognized as a compact device that can enable ultra-high bandwidth densities and energy efficiencies. Practical uses of microrings, however, have been plagued by fabrication tolerances, thermal sensitivities, and a lack of a good integration strategy with electronics.

In this thesis, the major accomplishment was to demonstrate the feasibility of microring-based communication systems built through a “zero-change” approach to monolithic integration. A disk-like cavity, termed the “spoked-ring”, was introduced as a suitable cavity that meets all of the constraints in a thin-SOI microelectronics process. The spoked-ring was used to demonstrate energy efficient, tunable depletion-mode modulators and tunable microring filters with record thermal efficiency. The modulators achieve 5 Gbps modulation with < 5 fJ/bit modulation efficiency as estimated from the junction capacitance and the voltage swing. This energy consumption is low enough to essentially be neglected in the entire system that uses these devices. The modulators have a free spectral range (FSR) of > 3 THz which can comfortably support 16 WDM channels, and the embedded heater achieves a tuning efficiency of 3.33 µW/GHz which allows for energy efficient on-chip feedback for wavelength stabilization. The tunable filter was shown to achieve an efficiency of 2 µW/GHz which is a record for microring based tunable filters without suspending the device.
in air.

The spoked-ring modulator was used as an enabling device to demonstrate robust communication links in realistic thermal environments using an on-chip feedback loop for maintaining alignment with the input laser wavelength, and it was shown that error-free operation critically relies on this stabilization. Using these devices and sub-systems, along with a SiGe photodetector and on-chip receiver, a chip-to-chip communication link was demonstrated which achieved a transmitter energy efficiency of 30 fJ/bit, a receiver energy efficiency of 350 fJ/bit, and a thermal tuning feedback loop with 524 GHz tuning range and 3.8 \( \mu \)W/GHz tuning efficiency (including electrical circuit efficiency). These results enabled the world’s first CPU chip to communicate with light. Along with a monolithically integrated RISC-V processor, a CPU-to-memory photonic link was demonstrated. The link was shown to achieve error-free operation during thermal stress tests, and arbitrary programs were executed that used the photonic link for writing and reading to and from main memory. This opens the door to a huge range of applications in communications, sensing, imaging, and possibly more fundamental physics experiments (photon pair generation with feed-forward detection and dynamic control, cryogenic optical readout of on-chip circuits, and more).

Next, two novel devices based on microrings were explored which could potentially aid from the on-chip integration with electronics. First, a microring filter configuration was presented that allows more densely packed WDM multiplexers/demultiplexers. This device was termed a “pole-zero” filter due to the presence of transmission zeros in the drop port, whereas microring filters today are mostly based on all-pole designs. When designed properly and cascaded in a WDM mux/demux row, \( \approx 2 \times \) denser channel packing is achieved with \(-20\) dB adjacent channel rejection, compared to the standard all-pole microring filters. Second, a two-microring system was theoretically explored for efficient wavelength conversion. By coupling two microring modulators together (e.g. the spoked-ring modulators), wavelength conversion efficiencies between -5.4 to -2 dB were predicted for depletion-mode silicon modulators (although modulators based on the electrooptic effect could potentially be even more efficient). It was shown that this particular device can be reduced to an equivalent time independent, linear filter model and well-known microring filter synthesis techniques
can be applied. This leads to straight-forward designs and performance predictions. Combined with on-chip circuitry, this device can form the basis of frequency shifters and optoelectronic RF oscillators.

Finally, we presented several key results that were achieved in our monolithic integration efforts. This includes demonstrations of photonic crystals, a “microring” equivalent photonic crystal configuration, non-linear four-wave mixing and generation of correlated photon pairs for quantum applications, extension of the “zero-change” approach to integration to the 32 nm SOI microelectronics process, and several key demonstrations in a bulk silicon platform including the first depletion-mode modulator demonstrated in bulk silicon and a WDM enabled chip-to-chip link. All of these demonstrations validate monolithic integration and serve as proof that high-performance systems can be built using these approaches. Many challenges and interesting directions for research still exist and are covered in the next section.

7.2 Remaining challenges and future directions for research

There are several key challenges that remain to fully realize the potential of the devices and systems presented in the previous chapters. These include robust fiber-to-chip packaging schemes that maintain high coupling efficiencies, further optimization of devices to achieve maximum performance, and a suitable laser source to provide multiple wavelengths of light so that the large bandwidth density enabled by WDM devices can be wholly utilized. For fiber-to-chip packaging, many research groups have shown promising results with techniques providing fixed fiber pig-tailing to vertical grating couplers [55, 57, 82], and there are large, on-going collaborative efforts to achieve scalable solutions [160].

The photonic microring active devices and the vertical grating couplers presented should be considered early generation devices, and there are several improvements that are possible through further engineering efforts. For the spoked-ring modulators, optimizing the electrical RC parasitics will result in faster modulation rates. This can be done through reducing the parasitic wiring capacitance of the device as well as better understanding of the dynamics of the radially extending
p-n junctions. Also, there are a wide variety of well-implants available in the process that are worth exploring to fine-tune the junction characteristics. Several masks can be used in conjunction to tailor the implant concentrations and their spatial profiles. Ultimately, the modulation speed is limited by the optical photon lifetime of the cavity (how long it takes to build up and deplete optical power in the cavity) which is set by the linewidth of the cavity. Microring structures that break the photon lifetime limitation are beginning to be researched \[98, 112\]. Also, concepts presented in the pole-zero filter design (Sec. 5.1) can be applied to modulator design resulting in full 1-to-0 modulation which achieves lower insertion loss over a smaller frequency detuning \[97\].

The efficient thermally tunable filter can be optimized to enable narrower linewidth filters, as proposed in Figs. 2.20,2.21 and post-processing can be explored to undercut the surrounding silicon dioxide to further improve the tuning efficiency. The device has previously been used to form the cavity in an external cavity laser (ECL) \[15\], and it could also be used to build Fourier Domain Mode-locked Lasers \[49\].

The vertical grating couplers presented in Ch. 3 achieve the limit of what is expected from a uniform (non-apodized) grating. Further engineering of this device is likely needed to make it suitable for emerging fiber-pigtailing packaging technologies such as redesigning the device for a 10 \(\mu m\) mode-field diameter. The next major design iteration for this device is a fully apodized design, and the 45 nm platform has been recently shown to support extremely high coupling efficiencies using apodized designs \[83\]. Furthermore, advanced grating coupler devices such as producing focusing output beams and positional misalignment insensitive designs are worth exploring.

Chapter 6 presented several early experimental demonstrations that contain a plethora of opportunities for additional research. A quantum correlated photon-pair source was demonstrated that was directly integrated in the 45 nm process. This photon-pair source can be combined with on-chip microring filters to spectrally separate (and reject the pump wavelength) the generated photon-pairs for further use in on-chip experiments. Since the transistors in the 45 nm process have been shown to work at cryogenic temperatures \[136\], on-chip circuitry can potentially be used in dynamically controlled quantum experiments \[105, 117\].
Although the first demonstrations were shown in the 32 nm SOI fabrication process, achieving high-performance in that process remains as further engineering work. The next iteration of a chip designed in the 32 nm process should focus on achieving low-loss waveguides and understanding PDK design levels and mask generation since the higher than expected propagation loss on the first prototype chip was likely due to a mistake in dopant masks.

The results presented in this thesis showed unprecedented levels of electronic/photonic integration which opens the door to a wide range of future research directions in on-chip optical signal processing and communications such as photonic analog-to-digital converters and large-scale optical phased arrays. Additionally, new computer architectures that have already been theoretically investigated to take advantage of high-bandwidth chip-to-chip optical interconnects will need to be implemented and their practical challenges and limitations explored. Since the bandwidth-distance trade-off between a processor and memory can largely be broken with optical interconnects, new processor architectures that can take advantage of huge memory bandwidths but with $\propto \mu s$ latencies will likely be investigated.

In summary, the devices and systems demonstrated in this thesis usher in a new era of highly integrated electronic-photonic systems-on-chip, where the electronics and photonics work in concert to enable the next generation of efficient communications and signal processing systems.
Bibliography


193


