

Advanced Silicon Photonic Modulators

by

Cheryl M. Sorace

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

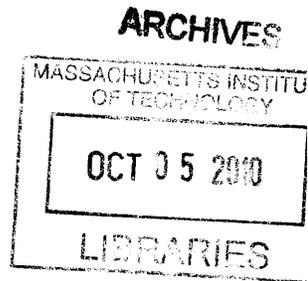
Master of Science in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2010

© Massachusetts Institute of Technology 2010. All rights reserved.



Author.....

Department of Electrical Engineering and Computer Science
September 3, 2010

Certified by.....

Franz X. Kärtner
Professor of Electrical Engineering
Thesis Supervisor

Accepted by

Terry P. Orlando
Chairman, Department Committee on Graduate Students

ADVANCED SILICON PHOTONIC MODULATORS

by

CHERYL SORACE

Submitted to the Department of Electrical Engineering and Computer Science
on September 3, 2010 in partial fulfillment of the
requirements for the degree of Master of Science in
Electrical Engineering and Computer Science

Abstract

Various electrical and optical schemes used in Mach-Zehnder (MZ) silicon plasma dispersion effect modulators are explored. A rib waveguide reverse biased silicon diode modulator is designed, tested and found to operate at speeds up to 13 GHz with a $V_{\pi}L$ of 1.2 Vcm. MOS capacitor modulator designs are investigated as an alternative, but are not found to offer significant advantages. Modulators are also designed for fabrication in an actual CMOS process—a crucial step in the quest for low-cost integration with modern electronic devices. Photonic crystal structures, which promise smaller footprint sizes and lower power requirements, are also investigated, but it proves difficult to obtain a physically feasible design. Finally, a linearization scheme for Mach-Zehnder modulators is proposed to significantly improve signal fidelity in analog applications. Simulations are used to demonstrate the effectiveness of this scheme for reverse biased silicon diode modulators.

Thesis Supervisor: Franz X. Kärtner
Title: Professor of Electrical Engineering

ACKNOWLEDGEMENTS

First, I'd like to thank my advisor, Franz X. Kärtner, for all his fantastic guidance and for giving me the opportunity to work on this project. He has been a patient mentor, a receptive sounding board, and an invaluable source of direction in my research.

So far at MIT, I've had the opportunity to work with many wonderful people and I'd like to express my gratitude to them. Anatoly Khilo, especially, has been my go-to person for patiently explaining everything from basic physical concepts to MIT logistics. He should really be a coauthor on the Linearity Section. I am also grateful to Amir Nejadmalayeri for his valuable explanations and experimental help. Jason Orcutt from Rajeev Ram's group has also provided me with helpful insights and many fruitful conversations. Additionally, I'd like to thank him for letting me borrow his measurement setup. I'd like to thank Dorothy and Donna for all the behind the scenes work that keeps the group running smoothly. Special thanks, or maybe recriminations, go to Dorothy for fueling my mid-afternoon candy habit. Finally, I'd like to thank all the members of the OQE group for making my time so far enjoyable and productive. I look forward to continuing to work with them.

I'd also like to thank our collaborators at Lincoln Laboratory; Dr. Steven Spector, Dr. Michael Geis, Dr. Matthew Grein, and Dr. Jade Wang, for the countless hours they have spent working on this project with us, and particularly for fabricating our devices and for characterizing many of them.

Additionally, I'd like to thank all the wonderful teachers I have had at Cornell and MIT that have prepared me so well for research. In particular, I'd like to thank Professor David Muller for all his advice, and Professor Michel Louge for giving me a great undergraduate introduction to research.

I'd like to thank my friends and family for all the incredible support they have provided me over the years. I'd especially like to thank Kendra for being my voice of reason, Tom for always seeing the bright side, and LEM for providing me with my first MIT home. Finally, I'd like to thank my fiancé, Ameya, for celebrating the good times with me and pulling me through the bad times.

SUPPORT

This work was funded by DARPA as part of the EPIC program and by a National Science Foundation Graduate Research Fellowship. Additional support was provided by the MIT Ida M. Green Fellowship.

CONTENTS

INTRODUCTION	11
APPLICATIONS OF MODULATORS	11
<i>Photonic ADC Converters</i>	11
SILICON PHOTONICS.....	13
SCOPE OF THESIS	15
MODULATOR BASICS.....	16
PHYSICAL EFFECTS USEFUL FOR MODULATION.....	16
<i>Pockels and Kerr Effects</i>	16
<i>Quantum Confined Stark Effect and Franz-Keldysh Effect</i>	16
<i>Thermo-Optic Effect</i>	17
<i>Plasma-Dispersion Effect</i>	17
MACH-ZEHNDER MODULATORS – RATIONAL AND TRANSFER FUNCTION	19
FIGURES OF MERIT.....	22
FUNDAMENTAL LIMITATIONS OF PLASMA-DISPERSION MZ MODULATORS.....	22
MODULATOR MODELING.....	29
ELECTRICAL SIMULATION DETAILS	30
OPTICAL SIMULATION DETAILS	33
RIB WAVEGUIDE MZ MODULATOR DESIGNS	36
MODAL OVERLAP CONSIDERATIONS	36
ACTIVE DEVICES FOR MOVING CARRIERS	40
<i>Diode Modulators</i>	41
<i>MOS Capacitor Modulators</i>	54
<i>Other Modulator Electrical Structures in the Literature</i>	63
LINEARIZATION OF MZ MODULATORS.....	64
SCHEME PROPOSED IN THIS WORK.....	64
APPLICATION OF SCHEME TO REVERSE BIASED SILICON MZ MODULATOR.....	70
MZ MODULATORS IN THE NATIVE CMOS PROCESS	75
PHOTONIC CRYSTAL MODULATORS.....	83
BACKGROUND.....	83
USE AS SLOW WAVE STRUCTURES AND SLOW WAVE MODULATORS.....	87
SLOW-WAVE MODULATOR DESIGN.....	89
<i>Achievable Slowdown</i>	89
<i>Contacting Issues</i>	93
CONCLUSIONS AND FUTURE WORK.....	97
APPENDICES.....	98
A: SELECT PARAMETERS USED IN ELECTRICAL SIMULATIONS	98
B: EXAMPLE ELECTRICAL SIMULATION CODE	103
C: OPTICAL PROCESSING CODE	109
D: EXAMPLE “MIT PHOTONICS BANDS” CODES	118
REFERENCES	120

INTRODUCTION

Applications of Modulators

As electronics become more complex and performance demands become more exacting, it becomes increasingly important that each component in a system be as accurate and efficient as possible. Sometimes the performance of traditional electrical components is acceptable for the desired specifications. Increasingly, however, the use of non-traditional components is necessary or will be necessary to meet performance targets [1]. Optical components, in particular, provide a path forward towards achieving much-desired performance gains. Including optical components in a system, however, requires successfully passing the requisite information back and forth between the optical and the electrical components in the system. Such communication is accomplished by the use of photodetectors, which allow optical signals to be passed to electrical systems, and modulators, which pass electrical signals to optical systems. While any photonic link would need to consist of both a modulator and a photodetector, this work will be focused on the design and implementation of electrical to optical modulators.

There are many applications in which opto-electronic integration is desirable. One such application is communication between multiple processing cores and memory modules in integrated circuits, particularly when these components are located on different chips, or far apart on the same chip. Currently data transmission between these components is accomplished via electrical interconnects. However, as transistor density and the bandwidth required increase, power density and interconnect latency become problematic. Photonic links are likely to be faster, have higher bandwidth, and use less power than traditional electronic solutions [1].

Photonic ADC Converters

ADCs are the means by which analog signals (from, say, an imaging system or an antenna) are digitized for further processing. Many different effects, including thermal effects, ambiguity, quantum uncertainty and jitter, place limits on the performance of ADC systems [38]. These limits for available technologies are summarized in Figure 1, which also summarizes the performance of current electronic ADCs. As can be seen, aperture jitter – random deviation from the nominal sampling period – is the effect that currently limits electrical ADC performance. At present, aperture jitter in electronic ADCs is on the order of 100 fs and improves only slowly [7]. Optical sources can be much more precisely timed, with jitters of less

than 100 attoseconds, and, therefore, offer a way around this problem. Optical schemes to implement or improve ADC performance have been around for several decades, but only recently have mode-locked lasers been stable enough for these schemes to be practical [7]. Optical ADC schemes generally fall into one of four categories: (1) where photonics simply "assist" an existing electrical component, (2) where optics are used to sample and electronics to quantize the signal, (3) the reverse, where optics quantize and electronics sample the signal, and (4) where optics are used to both quantize and sample the signal [39].

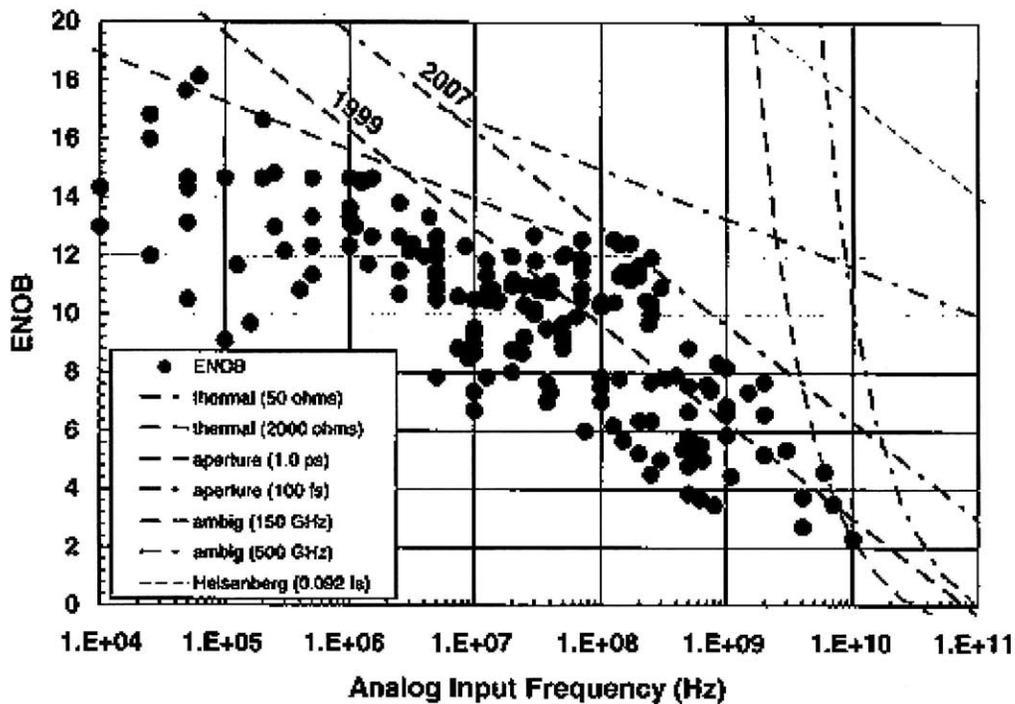


Figure 1: Performance of electrical ADC and performance limits. From [38].

Our group is currently working on designing an ADC of the second type; in which optical sampling is combined with electrical quantization. The optical sampling allows for a significant increase in the ADC bandwidth at a given accuracy. Specifically, the jitter problem is overcome by the use of a stable mode-locked laser to determine the sampling times, thus increasing the timing precision, and by the use wavelength division multiplexing (WDM) to split the signal between several electronic ADCs and thereby multiply the effective sampling rate achievable at a given ENOB level. To be more precise, a mode locked laser and a dispersive fiber are used to generate a sequence of chirped pulses. An electro-optic modulator, driven by the RF signal to be

sampled, imprints the signal on the chirped pulse stream. The modulated signal is then passed through an optical filter bank where different channels pick out the different frequencies of the chirp. Each channel passes its frequency to a photodetector, which converts it back into an electrical signal. Finally, this electrical signal is quantized using an electronic ADC. Because the temporal relationship between the frequencies is known, the digital output from all of the different channels can be stitched back together to create a finely-sampled version of the signal. The sampling rate of the ADC is then that of the electronic ADCs multiplied by the number of different channels of the filter bank [7]. For example, if the system had 20 channels and each channel had an electrical ADC sampling at 2 GHz, the overall sampling rate would be 40 GHz. The achievable ENOB could then be as high as the 8-10 effective bits 2 GHz electrical ADCs can achieve, instead of the 4 or so effective bits higher speed electronic ADCs can obtain [38].

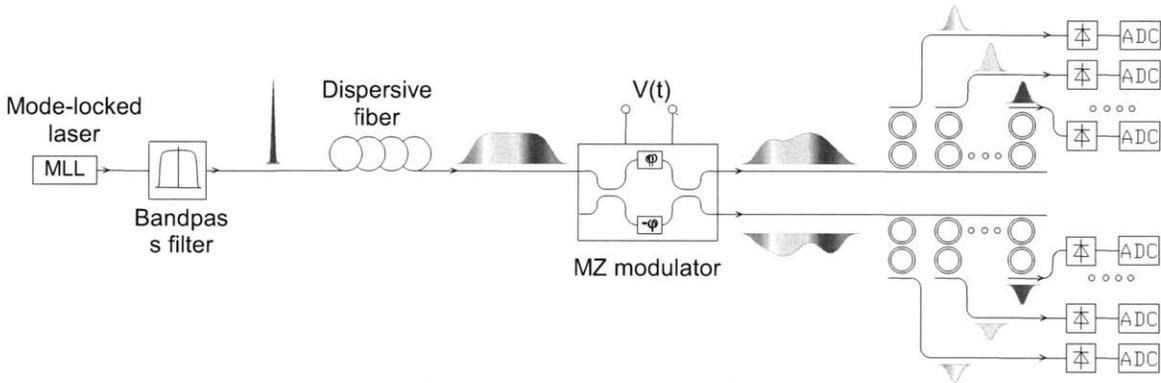


Figure 2 Schematic of optically sampled, electrically quantized ADC. Knowledge of the precise temporal location of each different optical frequency allows the outputs from all of the electronic ADCs to be recombined to yield a finely sampled version of the original signal. From [7].

Silicon Photonics

Much attention has been paid to silicon photonics because it promises to integrate optical and electronic devices together on the same chip, or even optical and electronic functions into the same device. Such integration would allow for the cheap, efficient and effective use of optics to improve the performance of what are presently all-electronic systems where the electronics are hitting their fundamental performance limits. Additionally, silicon photonic structures can take advantage of silicon fabrication techniques that have been developed into maturity for the electronics industry [5]. This allows silicon photonic devices to be made easily and cheaply.

There are other reasons that silicon is a good candidate for optical device design. The high index contrast between silicon and air, or, more importantly, between silicon and its native oxide

(SiO₂) allows for high index contrast (HIC) waveguides to be easily fabricated. HIC structures have the advantage of tight confinement (which decreases necessary power), small size and a tight bending radius (which decreases device size—key for any sort of integration with electronics or with other optical devices on the same chip) [7]. The tight bending radius also allows for the creation of ring filters with large free-spectral-range and of high-Q ring resonators (which, among other things, are useful for ring modulators) [1][7]. Finally, silicon is transparent at most of the traditional telecommunications wavelengths, allowing for on chip processing of signals sent over fibers without any sort of wavelength conversion, and for the use of laser sources already developed to maturity for the telecommunications industry [6].

However, silicon also has several drawbacks. Perhaps the most significant one is the lack of a direct bandgap, which makes the use of silicon to *create* optical sources all but impossible (requiring either an outside source, or that a source be bonded onto the silicon wafer) [6]. For better or for worse, the nonlinear coefficients of silicon are quite low, precluding most useful nonlinear effects (e.g., electro-optic modulators using the Pockels effect or the Kerr effect are not practical, see below) [1]. Moreover, HIC structures like those in silicon can be highly sensitive to fabrication variations and the tight tolerance specifications can be difficult or impossible to meet [7]. Often this requires some post-fabrication way to correct errors. Thermal control is a common solution, but leads to decreased power efficiency, more complicated designs, and degraded performance as it must be maintained without drift [6]. Furthermore, coupling into and out of HIC structures from fiber, often necessary due to the lack of on-chip sources, can be difficult and is in many cases associated with high losses (on the order of a few decibels.) Moreover, because HIC structures are often highly polarization dependent, some sort of polarization control must be maintained in the fiber parts of the system [7]. Finally, propagation in the waveguides is also associated with loss on the order of 0.1-0.5 dB/cm at minimum. This loss is mainly caused by sidewall roughness (but loss from impurities can also be high) [6]. In some applications these losses are problematic.

In all, silicon is a very promising material for optics due to its high index contrast, mature fabrication processes, and easy integration with electronic components. Use of silicon photonics can lead to important functional improvements to current electronic systems. However, there are still many challenges that need to be overcome before silicon photonics can be practically implemented at large scale.

Scope of Thesis

This work concentrates on Mach-Zehnder modulator structures that use the plasma dispersion effect in silicon, for applications in integrated photonic systems such as the ADC system outlined above. The next chapter, “Modulator Basics” outlines basic modulator designs and the limits of what is theoretically possible using silicon phase shifters using the plasma dispersion effect. The following chapter, “Modulator Modeling” explains the physical and numerical models used for the simulations carried out in this thesis. “Rib Waveguide MZ Modulator Designs” explains the basic structures that can be used to inject or extract carriers. It focuses on simulated and experimental results for diode modulators and simulation results for MOS capacitor modulators. “Linearization of MZ Modulators” shows how linear operation can be achieved in a simple MZ structure, and applies this to the main diode design of the previous chapter. “MZ Modulators in the Native CMOS Process” explores the integration of silicon plasma dispersion effect MZ modulators into a current silicon electronics fabrication process. Finally, slow wave structure options are discussed in “Photonic Crystal Modulators”.

MODULATOR BASICS

Physical Effects Useful for Modulation

There are many different physical effects that can be exploited to make modulators, and a variety of ways that each effect can be implemented. The most commonly used physical effects in modulation include the Pockels effect, the Kerr effect, the Franz-Keldysh effect, the quantum confined Stark effect (QCSE), the thermo-optic effect and the plasma dispersion effect. All of these effects use physical phenomena to change either the index of refraction or the absorption of the material or both. In general, modulation schemes that rely on changes in the index of refraction, while trying to minimize absorption, are considered to be more useful for analog modulation.

Pockels and Kerr Effects

In the Pockels and Kerr effects the anharmonic nature of the electron potential allows for an applied electric field to change the polarization of a material and, thus, change its refractive index. Modulation speeds can be very fast since the effects are based on bound electronic states in the material, so no carrier transport is involved. For the same reason, modulators based on these effects consume very little power. However, the Pockels effect is only present in crystals that lack inversion symmetry. In centrosymmetric (unstrained) silicon – the material addressed in this work – the Pockels effect vanishes. The Kerr effect is present in all materials, including silicon. However, in silicon it is very weak–yielding index changes on the order of 10^{-6} to 10^{-4} at the most [1].

Quantum Confined Stark Effect and Franz-Keldysh Effect

The quantum confined stark effect (QCSE) and the Franz-Keldysh effect both use an electric field to modify the energy bands of a semiconductor in order to shift the absorption spectra of the material to longer wavelengths. In the Franz-Keldysh effect, the applied electric field tilts the bands throughout the crystal material, leading the electron and hole wavefunctions, which do not tilt, to extend slightly further into the bandgap. Photon-assisted tunneling may then occur at slightly longer wavelengths than before, and light at these wavelengths will be absorbed instead of transmitted [1]. Similarly, in the QCSE, the applied electric field changes the bands in the quantum well in such a way as to lower the electron energy and raise the hole energy. Thus, the

energy needed to excite an electron decreases and longer wavelengths are absorbed during photon-assisted tunneling. In silicon the absorption due to the Franz-Keldysh effect is relatively small – demanding large device sizes to work – and only works at certain wavelengths. The QCSE is routinely used in III-V materials and has been demonstrated to be relatively strong in Ge-SiGe quantum wells [40]. However, Ge-SiGe quantum wells are difficult to fabricate, and there are still no good schemes for coupling into the quantum well structures from standard, thin silicon waveguides [1].

Because a change in absorption also yields a change in refractive index through the Kramers-Kronig relations, the QSCE and Franz-Keldysh effect can also be exploited to change the index of refraction of the material. For wavelengths that are much longer than the bandgap the index change has a larger effect than the absorption change, and phase modulator designs work better. Such is the case for modulation in silicon at the telecommunications wavelengths. However, the Franz-Keldysh effect is still quite small – on the order of 10^{-5} for an electric field of 100,000 V/cm [1]. The QCSE is larger, but the same coupling and fabrication issues remain.

Thermo-Optic Effect

The thermo-optic effect refers to the impact of a change in temperature on the optical properties of a material. Thermal expansion changes the optical path length through the material. More importantly, a change in temperature causes a change in electron distribution because it alters the band structure and the electron-phonon interaction coefficients. This change in electron distribution translates into a change in index of refraction, and is the dominant of the two effects in both amorphous and crystalline silicon [33]. The thermo-optic coefficient is defined as the effective change in refractive index for 1°C change in temperature of the material. At telecommunication wavelengths in room-temperature silicon, the thermo-optic coefficient is about $2 \times 10^{-4} \text{ K}^{-1}$ [6]. Despite the comparatively large effect, the thermo-optic effect is not a good modulator candidate due to the time it takes to effect large, uniform temperature changes in a material. In the end, device operation is limited to about 1 MHz [9].

Plasma-Dispersion Effect

The plasma-dispersion effect, caused by free carrier injection or extraction, can also change the index of refraction and absorption of a material. A theoretical interpretation of this effect can be derived starting from the Drude model of electrical conduction [1] [6]. The Drude model

states that electron movement in a crystal is governed by: $e\mathbf{E} = m \frac{d\mathbf{v}}{dt} + m \frac{\mathbf{v}}{\tau}$. When this equation is substituted into Maxwell's equations, a carrier-dependent index of refraction and absorption result. After a little algebra one eventually finds that:

$$\Delta n = -\frac{e^2 \lambda^2}{8\pi^2 c^2 \epsilon_0 n_0 m} \left(\frac{\Delta N_e}{m_{ce}^*} + \frac{\Delta N_h}{m_{ch}^*} \right) \text{ and}$$

$$\Delta \alpha = \frac{e^3 \lambda^2}{4\pi^2 c^3 \epsilon_0 n_0 m^2 \mu} \left(\frac{\Delta N_e}{m_{ce}^{2*} \mu_e} + \frac{\Delta N_h}{m_{ch}^{2*} \mu_h} \right) \quad [1],[6].$$

However, the Drude model makes many approximations, most of which work best for metals, and also does not account for many physical phenomena. Because of this, the standard empirical results of Soref et. al. [10] for silicon in the 1.55 μm telecommunications band are more useful for that material.

To obtain these results, Soref et al. compiled, from the literature, measured values of absorption, α , as a function of wavelength and doping. They then used the Kramers-Kronig relations to calculate the expected change in refractive index, and plotted these results along with the absorption data on a series of graphs. Interpolating, they came up with a set of data that relates free carrier concentration to $\Delta\alpha$ and to Δn in the two telecommunications bands (1.3 and 1.55 μm). Their results for 1550nm are shown in Figure 3.

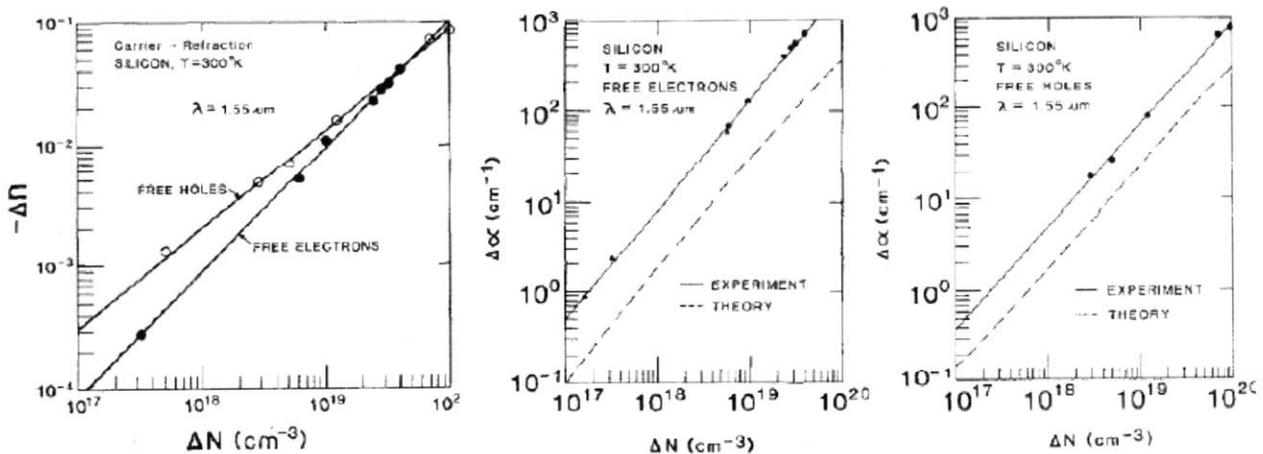


Figure 3 Left to right: fits to the change in refractive index as the concentration of electrons and holes varies (Soref et al. Fig.11), the change in absorption coefficient with a change in free electron concentration(Soref et al. Fig.14), and the change in absorption coefficient with a change in free hole concentration (Soref et al. Fig.15). From [10].

The data can be fit to get simple equations relating these quantities. These equations are usually given as [1], [6]:

$$\Delta n = -8.8e-22 \cdot \Delta N_e - 8.5e-18 \cdot (\Delta N_h)^{0.8}, \text{ and}$$

$$\Delta \alpha = 8.5e-18 \cdot (\Delta N_e) + 6.0e-18 \cdot (\Delta N_h)$$

However, refitting Soref et.al.'s absorption data shows that a better fit is actually given by [5]:

$$\Delta \alpha = 9.1e-22 \cdot (\Delta N_e)^{1.22} + 2.5e-20 \cdot (\Delta N_h)^{1.13}$$

It is useful to note that neither of these sets of equations is presented in Soref et. al.'s original paper, though the first set of equations is used in later papers by Soref (for example, [11]). Soref et. al. did do a similar fit to the data in their original paper, but ended up with slightly different numbers, which are not, generally, used in the literature.

Index changes caused by carrier changes can be as high as 2×10^{-3} for 1×10^{18} p-type doping in silicon. Modulation speeds, which are determined by the time necessary to move the free carriers, can be in the GHz range [5]. Thus, the plasma-dispersion effect is the most widely used modulation mechanism in the current silicon modulator literature. It is this mechanism that is used for all the designs in this thesis.

Mach-Zehnder Modulators – Rational and Transfer Function

There are many different device designs for exploiting a change in index to create a working modulator. Possibly the biggest split in these structures is between resonant and non-resonant modulators. Resonant modulators are usually made in rings or photonic crystal cavities [5] [20] [18]. Non-resonant modulators are usually made in a Mach-Zehnder (MZ) configuration. Resonant modulators have a smaller footprint and require less power, but have smaller bandwidth and are more sensitive to temperature and fabrication variations. Mach-Zehnder modulators require more power and space, but have larger bandwidth and are more robust to variations [2]. Because the relevant application requires a large bandwidth, this work addresses Mach-Zehnder modulators

A Mach-Zehnder (MZ) modulator is created by splitting the power from an incoming light pulse between two paths of different effective length, then recombining it at the output. When the pulse recombines, the difference in effective length has caused a phase difference, which yields either constructive or destructive interference, thereby changing the output amplitude. In

order to dynamically change the output signal, it is necessary to dynamically change the effective length of one or both arms. Such a change is accomplished by including a device which employs one of the phase-affecting modulation effects described above: in our case the plasma-dispersion effect. For the remainder of this work "MZ modulator" will refer to the entire modulator structure, while "modulator" or "phase-shifter" will refer to the device present in one or both arms to modulate the effective path length. In order to ensure that full modulation depth can be achieved—that is that full cancelation is possible when the arms are 180° out of phase—the light must be split evenly between the two arms. This can be accomplished with either a 3dB coupler or a Y-junction. In our system (depicted in Figure 4) 3dB couplers are used as the 50-50 splitter at both the input and the output.

Modulators are included in both MZ arms of our system [7]. While, technically, only one arm needs to contain a phase shift (and thus a modulator), operating in both arms allows for push-pull operation. In push-pull, opposite biases are applied to the arms around some DC bias point, adding more phase in one arm and less in the other. A π -phase change can then be achieved in half the modulator length. To take advantage of the most linear part of the MZ transfer function, the MZ modulator is operated in quadrature. To place the MZ modulator in quadrature, a titanium heater is fabricated over one of the modulator arms. The heater is then used to adjust the accumulated phase change using the thermo-optic effect (see above) until the outputs indicate that the device is in quadrature [7]. After the second 3 dB coupler the desired output (the optical version of the input electrical RF signal) is transmitted in the top waveguide and its complement in the bottom one. Often both outputs are used in future processing to minimize errors [7].

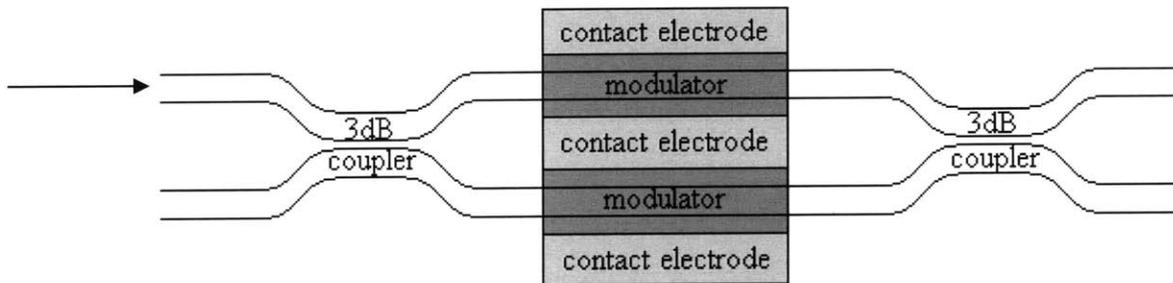


Figure 4: Schematic of Mach-Zehnder modulator waveguide structure as implemented on chip.

The MZ transfer function itself can be calculated as follows. It is assumed that all light is input on the upper channel, that the MZ modulator is operated at quadrature in push-pull, and that the 3dB input and output couplers are ideal. The phase shift on the upper arm is denoted by φ_1 and that on the lower arm by φ_2 and the input power is denoted by P_0 . The requirement that the MZ modulator be operated in quadrature places an extra $\pi/2$ phase-shift on one of the arms. In this calculation it is placed on the lower arm. For simplicity it is assumed that the modulators are lossless. The effects of loss will be discussed in the “Linearization of MZ Modulators” section below. After the first 3dB coupler the amplitude in the upper arm is given by

$$A_{t1} = \sqrt{P_0} / \sqrt{2}$$

and that on the lower arm by:

$$A_{b1} = \sqrt{P_0} e^{i\pi/2} / \sqrt{2}.$$

The heater and the modulators add a phase change to these signals:

$$A_{t2} = \sqrt{P_0} e^{i\varphi_1} / \sqrt{2}$$

$$A_{b2} = \sqrt{P_0} e^{i\pi+i\varphi_2} / \sqrt{2}.$$

At the second 3dB coupler part of the light in each arm stays in that arm and part is coupled across to the opposite arm. The final output amplitudes are :

$$A_{t3} = \sqrt{P_0} [e^{i\varphi_1} - e^{i\varphi_2+i\pi/2}] / \sqrt{2} = \sqrt{P_0} e^{i\pi/2} e^{i\left(\frac{\varphi_1+\varphi_2+\pi}{2}\right)} \sin\left(\frac{\varphi_1 - \varphi_2}{2} - \frac{\pi}{4}\right)$$

$$A_{b3} = \sqrt{P_0} [e^{i\varphi_1} + e^{i\varphi_2+i\pi/2}] / \sqrt{2} = \sqrt{P_0} e^{i\pi/2} e^{i\left(\frac{\varphi_1+\varphi_2+\pi}{2}\right)} \cos\left(\frac{\varphi_1 - \varphi_2}{2} - \frac{\pi}{4}\right)$$

Multiplying by the complex conjugates yields the final output powers:

$$P_{upper} = P_0 \sin^2\left(\frac{\varphi_1 - \varphi_2}{2} - \frac{\pi}{4}\right)$$

$$P_{lower} = P_0 \cos^2\left(\frac{\varphi_1 - \varphi_2}{2} - \frac{\pi}{4}\right)$$

It can be seen that the two outputs are complementary, and that the MZ transfer function is sinusoidal, and not linear, in shape. This has important implications for modulator accuracy which will be discussed in "Linearization of MZ Modulators" section below.

Figures of Merit

There are several common figures of merit (FOM) that are used when evaluating modulator performance. The most important ones are $V_\pi L$, the 3dB optical bandwidth, the 3dB RF bandwidth, and the insertion losses incurred by the device. These last three are more or less self-explanatory. The 3dB optical bandwidth is the wavelength range over which the device works. The 3dB RF bandwidth is the fastest electrical signal that can be faithfully modulated. The insertion loss is the total optical loss through the device, including all waveguide loss mechanisms and input and output coupling. It is usually given in dB/cm, and, especially in smaller devices, can be quite high [26]. $V_\pi L$ is meant to be a measure of device sensitivity and corresponds to the length of the device times the voltage which, when placed across it, yields a π -phase-shift (thus allowing the MZ output in one channel to be changed across the full range from zero to the input power using a phase-shift section in only one arm). In linear, electro-optic modulators (like LiNbO_3), $V_\pi L$ is a very good metric; increases in device length translate to decreases in voltage in such a way that the $V_\pi L$ of a given modulation scheme stays constant even as device length varies. However, in silicon plasma dispersion modulators, this FOM is problematic because such a nice relationship does not hold. More specifically, because the carrier concentrations are not linear functions of bias voltage, some voltage steps create larger changes in phase than others. The DC operating point and the amplitude of the RF voltage swing, thus, both matter when determining the phase change achieved. There are then locations on this phase-change versus voltage curve where increasing the voltage has more or less effect than increasing the length. For example, in a forward biased diode modulator, the relationship between voltage and phase is exponential and a small change in voltage can swamp even a large change in length [5]. Alternatively, in a reverse biased diode modulator, once the waveguide is largely depleted, increasing the voltage across the device does little in comparison to increasing the device length. Because, in the end, the quantity of interest is the power required to operate the device, a much better FOM would simply be the total RF power required to achieve a π -phase-shift. The footprint of the device, which is more relevant than the length in determining packing density, could then be reported separately

Fundamental Limitations of Plasma-Dispersion MZ Modulators

The plasma dispersion effect relies on injection or extraction of free carriers in order to create a change in refractive index. In order for a modulator based on the plasma dispersion effect to

work, there must exist both a way of injecting and extracting carriers at high speed, and a way to exploit the generated change in refractive index to create modulation.

The generation of free carriers is usually accomplished through the inclusion in the design of one or more electronic devices which inject or deplete them. Diodes are by far the most common structures used in the literature [5] [4] [1] [19] [20]. MOS capacitors are also common [9] [14] [15] [16] and a few researchers have used specialty structures (e.g. [12] [13] [2] [37].) In non-integratable modulator designs carrier creation is sometimes accomplished by use of an external light source (such as a pumping laser) [17]. The work in this thesis focuses on diode structures, with some work on MOS capacitors.

The details of diode and MOS capacitor modulator structures are discussed below. However, it is first useful to consider the limitations imposed by the use of the plasma dispersion effect itself.

The first restriction to note is that changes in refractive index are coupled with changes in the loss. This is always the case – refractive index and loss are related through the Kramers-Kronig relations. In the case of the plasma dispersion effect, it is clear what the physical source of much of this loss is: free carriers can be excited optically within the bands and then relax back thermally through the emission of one or more phonons [6]. The more free carriers there are, the more undergo this process, and the higher the loss is. Because a certain amount of index change is associated with the presence (at some point) of a given number of carriers, it will also be associated with a minimum amount of loss. In general, holes yield both a larger change in index and a lower loss. It is therefore desirable that phase changes be accomplished entirely using holes. While this is not possible in practice (though it is possible to *mainly* use holes), to discuss fundamental limitations it will be assumed for the remainder of this section that all free carriers are holes.

Given the maximum tolerable loss in a silicon phase-shifter that uses the plasma dispersion effect, the maximum achievable phase change can be calculated by first assuming that all losses in the waveguide are the result of free carriers (with no contributions from side-wall roughness, insertion loss, etc.) It is further assumed that the dependence of the index of refraction change on free carriers has the form:

$$\Delta n = A_n (\Delta p)^{p_n}$$

and that of loss on carriers has the form:

$$\Delta\alpha = A_a (\Delta p)^{p_a}$$

Where Δp denotes the change in the hole concentration and A_n , A_a , p_n , and p_a are coefficients. The maximum phase shift achievable for a given amount of loss, M_{loss} , is then:

$$\Delta\varphi = k_0 LC_f A_n \left(\frac{\left(\frac{M_{loss}}{10} \right) \ln(10)}{LC_f A_p} \right)^{\frac{p_n}{p_a}}$$

where L is the length of the phase shifter, C_f represents how much of the mode is confined in the region of carrier change, and M_{loss} is the total loss, in dB, tolerated through the device. From this equation we note that if $p_a \neq p_n$ the achievable phase shift will be a function of device length and of confinement factor. Because these coefficients are, in fact, not the same for holes in silicon at 1550 nm, we can see that a confinement factor of one (the highest allowed) will yield the most phase shift, and that longer modulators with lower carrier densities will achieve larger phase shifts per loss than shorter ones with higher carrier densities. Specifically, at 1550 nm the above equation becomes:

$$\Delta\varphi = k_0 LC_f \cdot -8.5e-18 \left(\frac{\left(\frac{M_{loss}}{10} \right) \ln(10)}{6.0e-18 \cdot LC_f} \right)^{0.8}$$

The achievable phase-shift per loss is shown in Figure 5.

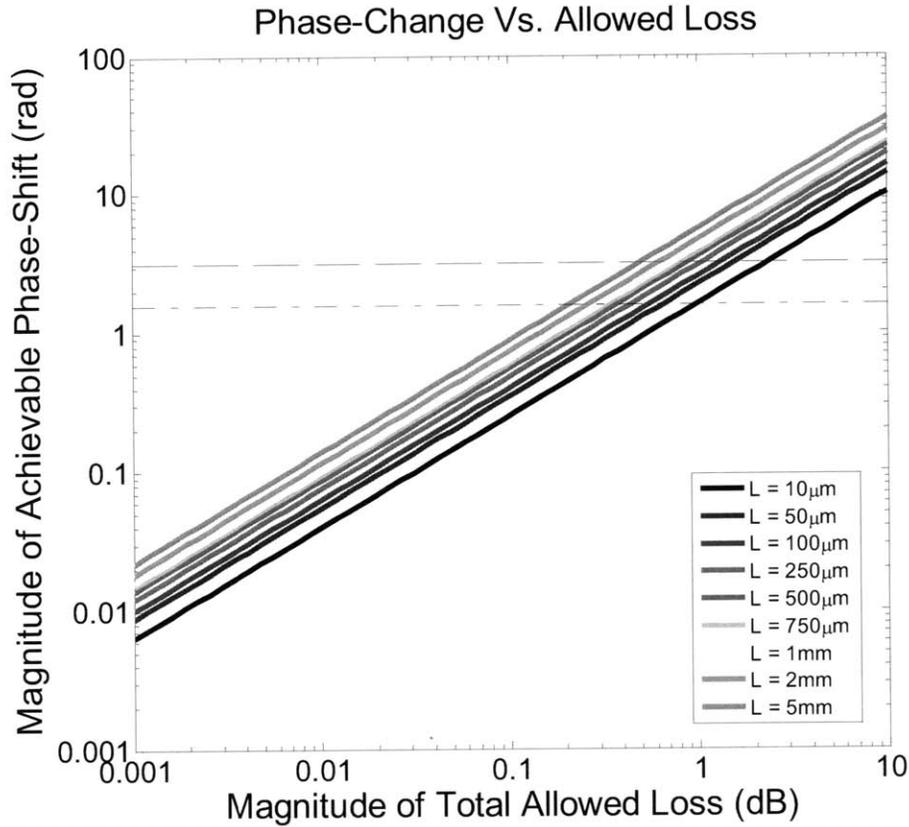


Figure 5: Dashed line corresponds to a π shift, dashed-dotted line to a $\pi/2$ shift. In a push-pull MZ modulator a $\pi/2$ shift in each arm is enough to achieve full modulation depth.

The $\pi/2$ radian shift needed to achieve maximum modulation depth in a MZ modulator driven in push-pull, is theoretically achievable for all considered lengths at a total loss of less than 1dB. However, almost three times as much phase shift is achievable in the longest device. To see why such a difference is present, it is useful to consider carrier concentration and the effective index change as a function of the total losses. These are shown in Figure 6.

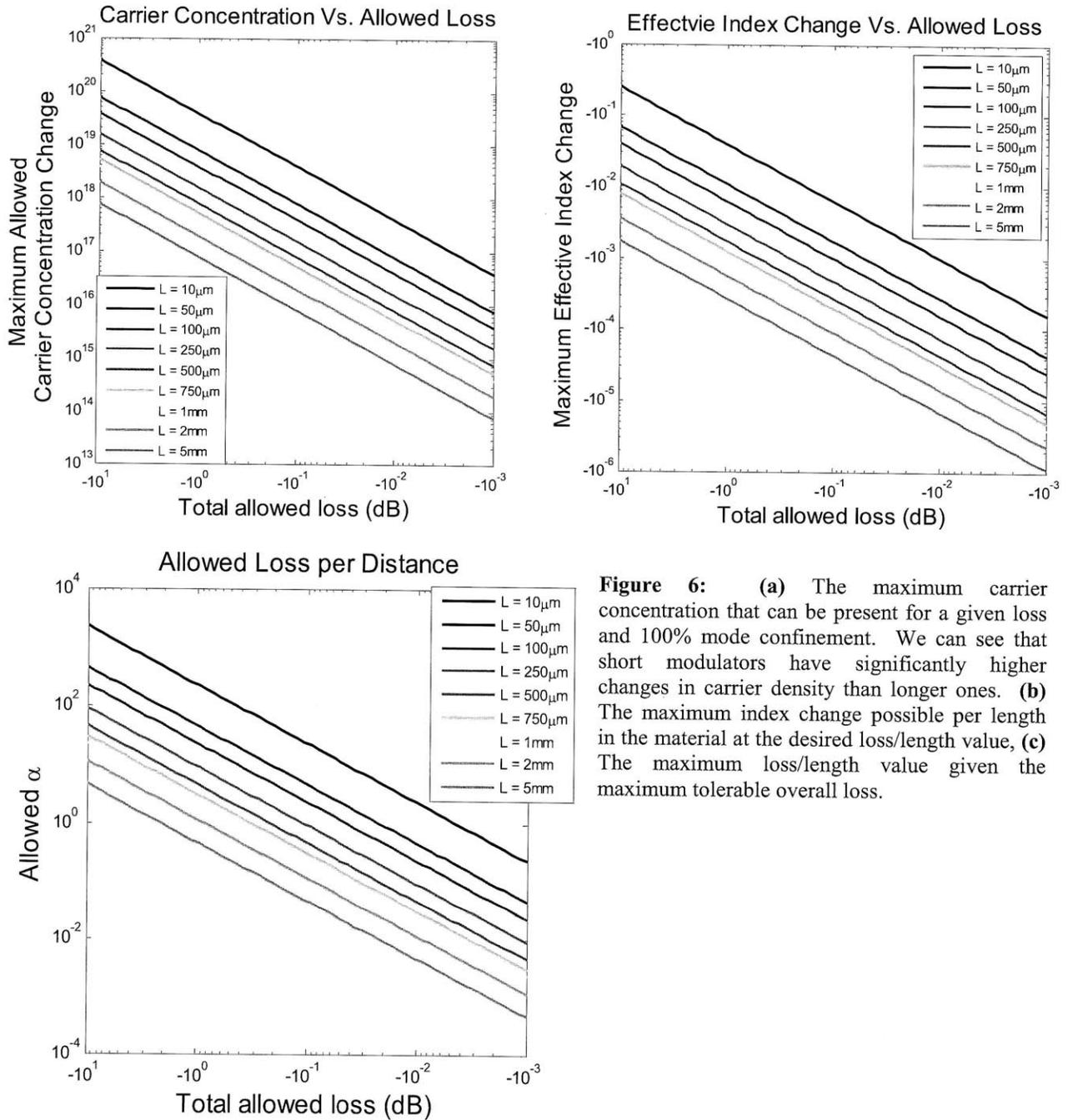


Figure 6: (a) The maximum carrier concentration that can be present for a given loss and 100% mode confinement. We can see that short modulators have significantly higher changes in carrier density than longer ones. (b) The maximum index change possible per length in the material at the desired loss/length value, (c) The maximum loss/length value given the maximum tolerable overall loss.

It can be seen that in the shorter structures, higher carrier concentrations are used to make up for the shorter length. However, because loss increases at a faster rate with carrier concentration than index change does, these higher concentrations are not able to compensate completely for the length difference. In the case where the loss and index change relations are of the same order, as is the case for electrons, increasing the carrier concentration while shortening the length leaves the overall phase shift unchanged.

The above calculations represent a theoretical limit and are, of course, unrealistically optimistic. The modal confinement to a modulated carrier region is usually much less than 50%, not the 100% assumed here. Furthermore, the modulated region, and thus the confinement to it, changes in volume as device doping concentrations are changed. Often, though not always, higher doping concentrations yield poorer overlaps, preventing the full benefit of having more carriers present from being realized. Making longer devices is also problematic, as not only is small size a desired modulator property, but there are often other losses (such as those from sidewall roughness or impurities) that limit practical device length.

Fundamental limitations to device speed also result from the use of the plasma dispersion effect. Because the index change is caused by a change in carrier concentration, it is necessary to move carriers into and out of some region of space. The speed that these carriers can be moved will determine how fast the device can be operated. There are several physical mechanisms that can limit speed. In the case of minority carrier presence, these carriers must either travel out of the region or recombine. Since both diffusion and recombination are slow processes, clearing a volume of minority carriers can take on the order of 100ps-1 μ s and can place a severe limit on device speed – limiting performance to a few GHz at best. However, as modulation can be accomplished with majority instead of minority carriers, lifetimes do not provide a fundamental limit on device speed.

When only majority carriers are present, speed is usually limited by the dielectric relaxation rate of silicon, the RC time constant of the device, or by the saturation velocity of the carriers. The dielectric relaxation time is a measure of how fast a perturbation in the distribution of majority carriers can recover to the equilibrium distribution. It is a function of doping and is generally well known (it is given by ϵ/σ). In p-doped silicon, with $N_A = 1 \times 10^{18} \text{ cm}^{-3}$, the dielectric relaxation time is around $4 \times 10^{-14} \text{ s}$, which, if dominant, would allow the device to operate at about 30 THz. The RC time constant places a speed-limit based on the capacitance of the device. While, in practice, this is often the limiting factor, there is no theoretical reason why it needs to be. Presumably, a low-capacitance, low-resistance design could be realized by either good device engineering (particularly if holes and electrons could be made to move in the same direction making the carrier plasma charge neutral and, thus, capacitance free), by the use of traveling wave electrodes to cut capacitance, or even, for some device designs, simply if the contacts have low enough capacitance and resistance that the intrinsic RC time constant of the

device (which can be quite low) dominates. It is, therefore, assumed that the RC time constant does not limit device performance. However, carriers cannot move arbitrarily fast – moving carriers scatter off the vibrations of the silicon lattice, not matter how pure the material, how high the field, or how few other carriers are present; the resulting limit on their speed is known as the saturation velocity. In situations where carriers need to move across large distances to restore equilibrium, saturation velocity ultimately provides the theoretical limit to device performance. In silicon, the saturation velocity for electrons is 1×10^7 cm/s (or 1/1000 the speed of light). If the modulated region they need to move in and out of is 400 nm wide, device operation is then limited to 95 GHz. For holes it is 6×10^6 cm/s. For the same modulated region width the device operation would then be limited to 57 GHz [1]. In general, if the distance the carriers must move is known, the maximum device speed can be determined from:

$$f_{3dB} = 2.4 / 2\pi\tau \quad \text{where } \tau = W / v_{sat}$$

where W is the distance the carriers must move and v_{sat} is the saturation velocity [1]. Note that the distance that needs to be moved might not be the length of the guide. Particularly for movement of majority carriers in an accumulation layer or movement of the depletion region edge in a highly doped material this distance might be quite short. However, for other situations, such as large depletion width changes, a longer distance must be used. Again, saturation speed represents best case performance. Usually, device speed is limited to much lower values by the RC time constant.

One final consideration is the limit on optical bandwidth from the plasma dispersion effect. Although there is a limit placed on the optical bandwidth usable at a given speed from the wavelength dependence of Soref's equations, this usually is dominated in the MZ case by the wavelength dependence of the waveguide propagation constants and, more importantly, of the couplers into and out of the phase shifting sections. In general, the optical bandwidth usable in a given phase-shifter design will depend on waveguide geometry and system tolerances.

MODULATOR MODELING

Accurate modulator modeling involves modeling both the electronic carrier responses of the device to an applied bias and the optical responses of the modes to those carrier changes. The modeling work in this thesis was split into two different parts. First, an electronic structure was designed and the carrier distributions were calculated either analytically or using Synopsys's Sentaurus™ TCAD software suite. Then, those carrier distributions were imported into MATLAB™ and used to calculate the optical response of the device. The work in this thesis focused on the design of the phase shifting modulator sections. Design of the overall MZ modulator structure, including couplers, was not undertaken. However, some analysis of the impact of the phase shifting section design on overall MZ performance was undertaken in the section entitled "Linear Modulators" below.

Electrical simulations were carried out using both analytic models and full scale simulations. Full scale electrical simulations were implemented in Synopsys's Sentaurus™ software suite. Optical simulations for straight and rib waveguides were performed in MATLAB™, where a mode solver previously developed by Milos Popovic was used to calculate effective indices and losses. Because Sentaurus does not output directly to a MATLAB-compatible format it was necessary to parse the output files into a readable form. It was also necessary to re-grid the carrier distributions from the finite element grid used in Sentaurus to an evenly spaced finite difference grid for input into Dr. Popovic's mode solver (MMS). The overall simulation flow for these files is outlined in Figure 7.

Optical simulations for photonic crystal waveguides were carried out using the MIT Photonic Bands software suite developed by Steven Johnson's group, the output of which was analyzed in MATLAB and in MayaVi™.

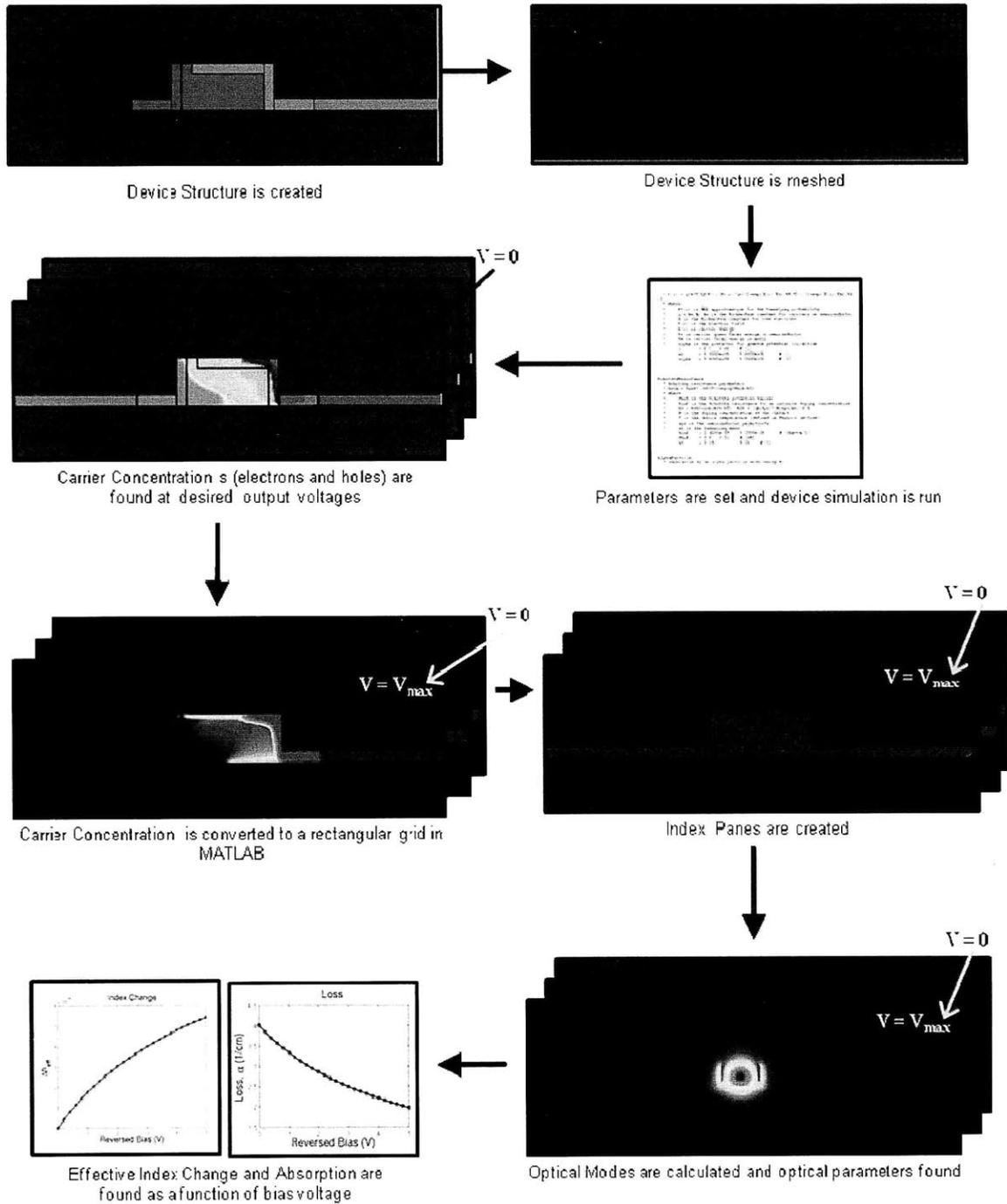


Figure 7: Flow diagram of simulation process

Electrical Simulation Details

Electrical simulations were carried out using both Sentaurus Structure Editor and Sentaurus Device. The accompanying Synopsys plotting software, Tecplot, was used in the processing of the output files. Sentaurus Workbench, a wrapper program, was used to organize the larger runs.

Device structures were created and meshed using Sentaurus Structure Editor. To create a device structure, a Scheme script placing material blocks, doping profiles, and contacts was written (see appendix B). For large parameter sweeps, Sentaurus Workbench was used to vary the desired input variables without the need to hand code each desired structure. To simplify device creation and analysis all doping profiles were assumed to be constant and abrupt. The boundary geometry specified for the device was then meshed using a Delaunay meshing algorithm suitable for finite element simulations [46]. A fine mesh was used in areas of particular interest, such as diode junctions. The mesh was allowed to expand away from these areas. An example boundary file and mesh (expanded for illustration) are shown in Figure 8. To keep simulation times down, the mesh was made as large as possible while still maintaining the desired accuracy. This accuracy varied from simulation to simulation: with the least accuracy demanded for original exploratory simulations and the most for rigorous optimization simulations.

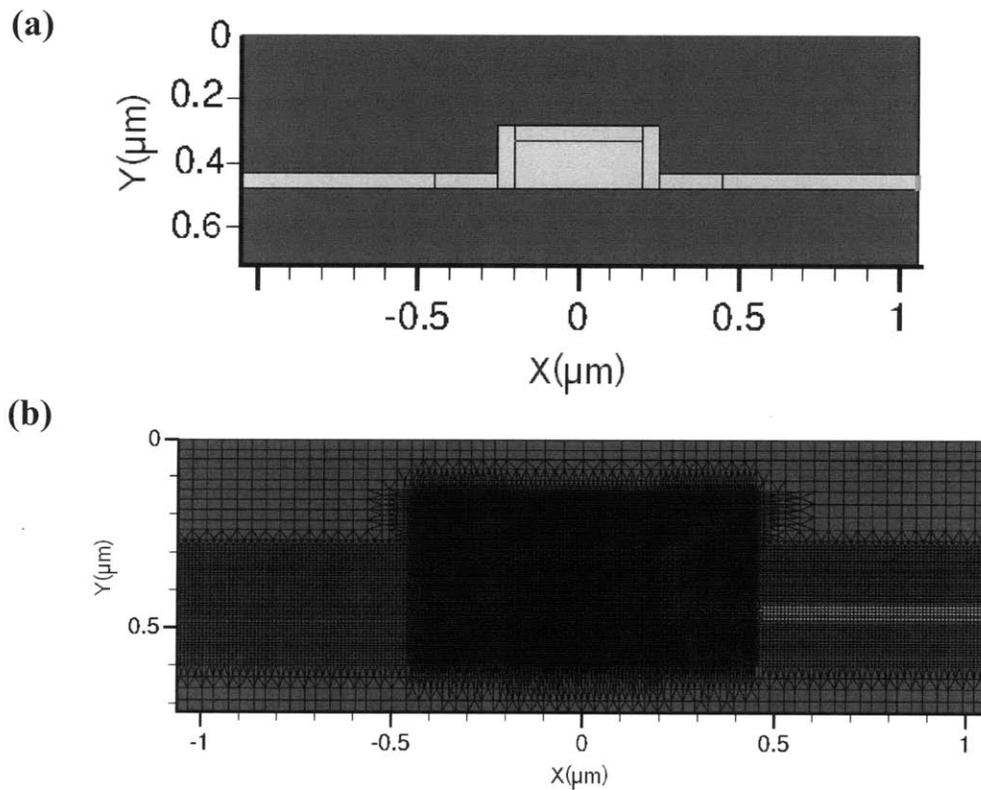


Figure 8: (a) Boundary geometry of example device structure, (b) Example Delaunay simulation mesh. Mesh is finest in the junction area and largest in the oxide area away from the active device.

After the device structure was created and meshed, its electrical performance was simulated using Sentaurus Device (Sdevice). In Sdevice, the Poisson equation and the electron and hole continuity equations were simultaneously solved. In many simulations, the heating equation was also solved. However, for most devices of interest (reverse biased diodes and MOS capacitors) its inclusion did not have a significant impact on simulation results.

A number of physical and numerical parameters can be set from the command and parameter files of an Sdevice simulation. A few of the more important ones will be highlighted here. (For an example command file, see Appendix B). Because of the high doping levels used in parts of the modulator, it was necessary to force the use of the more accurate Fermi statistics over the computationally more efficient Boltzmann statistics when calculating carrier densities. This was done by specifying the keyword "Fermi" in the physics section of the file. By default the mobility is a function of temperature only; this is not physically correct. Mobility degradation due to doping, high electric fields, and proximity to surfaces was added into this model. Similarly, all recombination mechanisms need to be explicitly specified: Shockley-Read-Hall (SRH), surface and Auger recombination were included. Avalanche and band to band generation were also incorporated into many simulations. Doping dependence was included in the SRH calculations. In all silicon simulations, the carrier lifetimes were left at the accepted values of 10 μ s for electrons and 3 μ s for holes, which are also the defaults. To find the surface recombination velocity at the silicon-SiO₂ interfaces, simulation results were fit to previous simulations that were fit to measurements of forward biased diode performance for diodes fabricated at Lincoln Laboratory. The resulting surface recombination velocity was 2×10^5 cm/s. This value is important in minority carrier dominated situations, as surface recombination is the dominant recombination process for minority carriers in silicon devices of this size. Using this surface recombination velocity the overall minority carrier lifetime was found to be around 1 ns. Devices whose responses are characterized mainly by majority carrier dynamics are not sensitive to this value. In forward bias diode simulations that involved polysilicon, the surface recombination was turned off, and the carrier lifetime in the SRH model varied by hand in the 100 ps-10 ns range. This was to account for the fact that the large number of flaw and trap states present in polysilicon dominate the lifetime. Because the number and type of these flaw and trap states depends heavily on the specifics of the device processing, predicting the lifetime values beforehand is impractical. Instead, this lifetime range was considered and devices designed that

could work throughout it. A list of important physical parameters input to simulations can be found in Appendix A.

Optical Simulation Details

As stated above, Sentauros does not output its results in a MATLAB readable format. It was, thus, necessary to first convert the output files from ".tdr" format to Tecplot's ".dat" format and then to parse them into a usable array. Because the optical mode solver needs a rectangular grid, and Sentauros simulations use a finite element grid, it is necessary to re-grid the electrical output before it is input into the mode solver. It is also necessary to convert the calculated carrier distributions into index and absorption distributions.

Each output file corresponds to a specific structure at a specific voltage. To re-grid these files, the location of and the variable values at each grid point are placed into a list and exported in ASCII form. A rectangular grid of blocks of the desired size (usually 2 nm × 2 nm) is created in MATLAB and the finite element grid locations in the list are placed onto this new grid. If more than one location falls within the same grid square, the values are averaged. To fill in squares in which no grid points fall, a linear interpolation scheme is used.

Once the rectangular grid is created, it is necessary to prepare the electrical output for input into the MMS. The complex refractive index at each point is determined according to

$$\tilde{n}(x, y, \lambda) = n_0(x, y, \lambda) + \Delta n(x, y, \lambda) + i \Delta \alpha(x, y, \lambda) / 2k_0$$

where n_0 is the material index of refraction, and Δn and $\Delta \alpha$ are, respectively, the change in index of refraction and absorption due to carriers, and are calculated using Soref's equations. At 1550nm, the standard equations from Soref presented above (in the section entitled "Plasma-Dispersion Effect") are used. However, there are no such standard equations at other wavelengths. Soref's original paper does contain graphs (but not equations) that relate change in refractive index and absorption to carrier density at a variety of wavelengths. An interpolation program based off data read from these graphs was implemented by Eugen Zraggen, a visitor to Rajeev Ram's group, and is used here when results at wavelengths other than 1550 nm are desired. Because this program yields slightly different results at 1550 nm than the standard form of Soref's equations (see Figure 9), the results at 1550 nm are recalculated using this interpolation program when comparisons across wavelengths are desired (for, say, optical bandwidth calculations) to ensure consistency. However, in order to facilitate comparisons to

the literature and to previous results in the group, when only 1550 nm wavelength results are wanted, the standard form of Soref's equations is used.

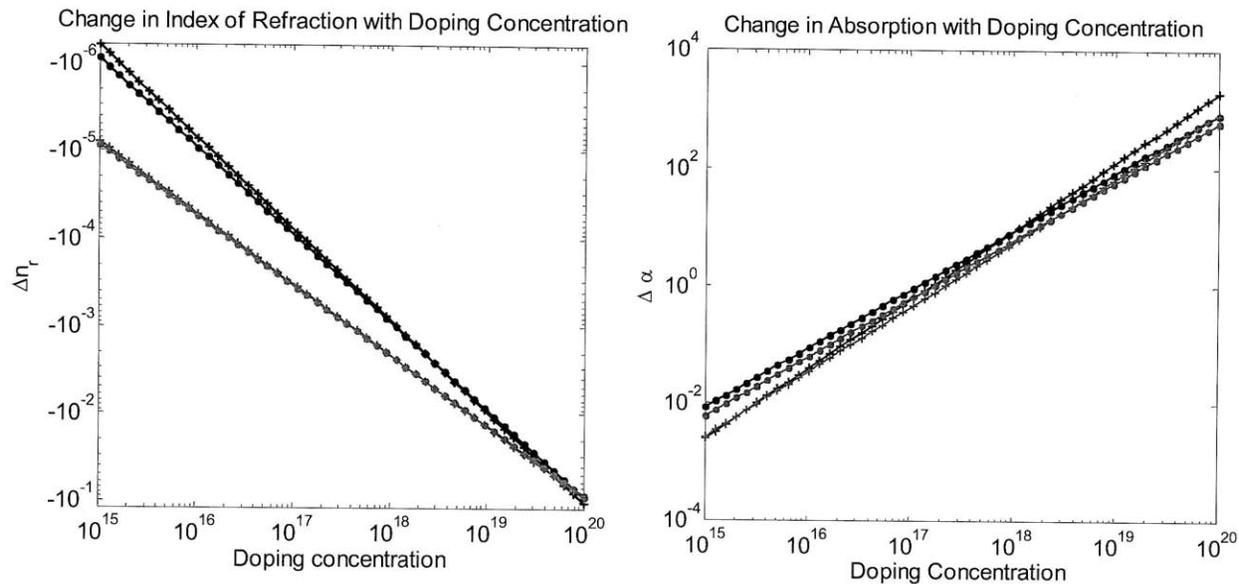


Figure 9: Comparison of change in index and absorption with carrier concentration at $\lambda=1550$ nm as given by the accepted forms of Soref's equations (circles) and the program from E.Zgraggen, which interpolates from Soref's graphs (crosses). Green lines indicate changes with hole concentrations and black lines those with electron concentration. As can be seen, there is good agreement in the area of 1×10^{18} carriers per cubic cm, but progressively less agreement further away from this range. At the edges of the graph (concentrations of $1 \times 10^{15} \text{cm}^{-3}$ or $1 \times 10^{20} \text{cm}^{-3}$) values may be off by a factor of two or more. Agreement is much better for index variations than for absorption variations.

After the carrier distributions are converted into a complex index distribution, that index distribution must be averaged for input into the mode solver. This is necessary because the mode solver uses an iterative Arnoldi algorithm and abrupt changes in matrix values hurt convergence and accuracy. Due to time and memory constraints, the size of the grid input into MMS is often coarser than the newly created grid of the Sentaurus outputs. If the two grids use different x- or y- direction steps, the complex index distribution is first re-gridded onto the MMS grid. Changes in MMS grid size below $10 \text{ nm} \times 10 \text{ nm}$ grid squares seem to have little impact on results (see Figure 10). Furthermore, because the computational domain for Sentaurus electrical simulations usually does not extend much beyond the silicon regions (while the modes can), extra oxide padding is added around the Sentaurus computational domain region to ensure that no problems are caused by computational domain boundaries. Once this is done, standard area-arithmetic index averaging is applied. The averaged index distribution is then passes to MMS where the

eigenvalues and eigenvectors of the index distribution are found using an Arnoldi algorithm in shift-invert mode, and the complex electric and magnetic fields along with the complex propagation constant are returned.

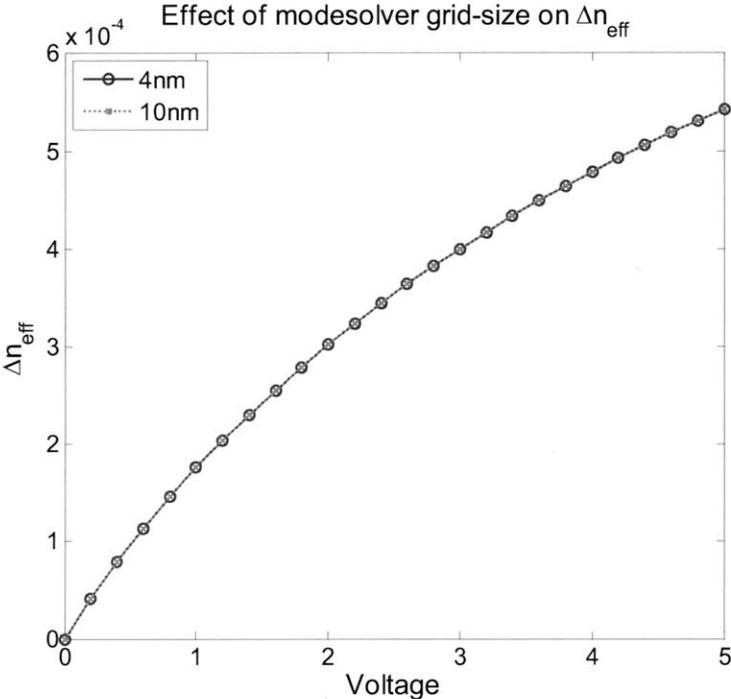


Figure 10: Below 10 nm × 10 nm grid squares, x- and y-direction step size of the grid input into MMS has little impact on results. Here the index change calculated for an example modulator design is shown for both a 10 nm × 10 nm and a 4 nm × 4 nm mode solver grid spacing.

RIB WAVEGUIDE MZ MODULATOR DESIGNS

A common waveguide design for modulator structures is the rib or ridge waveguide, pictured in Figure 11. This structure is popular because light is guided in and confined by the rib region, while the slab allows electrical contact to be made to an intrinsic device in such a way that the highly-doped and metal contact regions are kept separate (or largely separate) from the optical mode, thus, limiting losses to acceptable values.

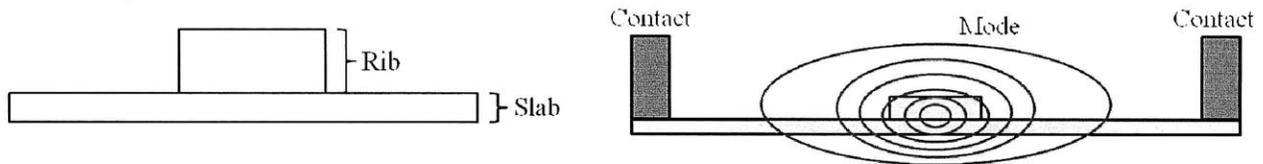


Figure 11: (a) A standard Rib-waveguide consists of a slab section with a rib placed on top, (b) the rib serves to confine the mode away from the ends of the slab, allowing lossy metal contacts to be placed there and successful electrical contact to be made to the intrinsic modulator device.

When designing a rib waveguide, there are two main design spaces: that of the electrical device and that of the optical waveguide. The key is to achieve regions of high index modulation from the electrical device and to integrate the two such that the optical mode has maximal overlap with the area of the device that sees index modulation.

Modal Overlap Considerations

Overlap between the mode and the modulated region is, of course, important in all modulator structures. It will be explored here in the context of the plasma dispersion effect in rib waveguides. Because only part of a mode is carried in the modulated, or active, region of the waveguide, it is necessary to calculate an effective index change from the refractive index change present in the active region. The effective index change reflects the actual change in phase accumulated by a mode traveling through the waveguide. When refractive index changes are small enough to be considered perturbations (that is the mode shape does not change noticeably), the effective index change can be calculated as the change in refractive index of the region times the confinement factor of the mode to the region. What percentage of the mode experiences what index change depends significantly on both the size and location of the active region and on the geometry of the waveguide itself.

The two graphs in Figure 12 show the effect of waveguide geometry on effective index change. The height and width of the rib in a rib waveguide are varied and the effective index of a plain silicon guide with the given geometry is compared to that of a perturbed guide. Two perturbations are considered. In the first, the perturbed region covers the bottom 100 nm of the guide. In the second, it covers the whole guide. Because the mode is more confined in a larger waveguide, when the perturbation occurs throughout the waveguide, the effective index change simply increases as the size of the guide increases and there is no optimal value for either dimension. However, when the active region only covers a portion of the waveguide, there can exist optimal values for waveguide dimensions. Here, the 100 nm profile is constant over the width of the guide, but varies in height. Therefore, there is no optimal width – wider is always better as it pulls more and more of the mode into the guide and into the perturbed region – but there is an optimal height. The maximum in effective index change with height reflects the trade-off between obtaining the best overlap between the peak of the mode and the active region and maintaining high overall confinement of the mode in the guide thereby keeping the modal peak high. Too short, and there isn't enough light in the waveguide to get optimal modulation even when the modal maximum is in the active region. Too high, and the modal peak moves higher in the guide leaving only the mode's edges to see the index change present in the bottom. Were the region of refractive index change to be a vertical strip instead of a horizontal one, there would be no optimal height, but there would be an optimal width.

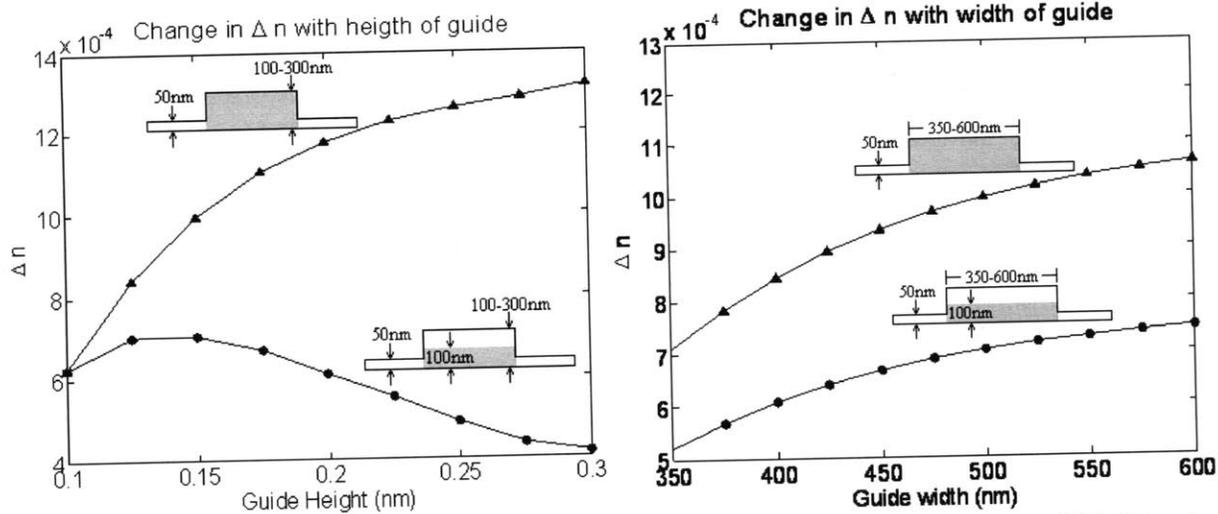


Figure 12 (a) Change in effective index (from the plain, unperturbed silicon waveguide case) with height when an index perturbation equivalent to the injection of 5×10^{17} holes is applied across the bottom 100 nm of the waveguide (circles) and across the whole waveguide (triangles). The rib width is fixed at 500 nm with a slab height of 50 nm surrounded by SiO_2 . Guide height reflects the entire height of the center of the guide (from the SiO_2 substrate to the top SiO_2 cladding). Note that, for a 100 nm perturbation there is a maximum Δn , while for perturbation of the whole waveguide, Δn increases continuously with height. **(b)** Change in effective index under the same index perturbations as a function of guide width. Here the guide height was held constant at 150 nm. In both cases effective index change simply increases with width.

Within a particular waveguide design, the details of the perturbation layer location and strength obviously also have a large impact on the mode. The three graphs below demonstrate this impact on effective index given a specific waveguide geometry. In the first graph, a 50 nm layer of refractive index change is moved from the bottom to the top of the waveguide. An analogous situation in an actual device would be moving the vertical location of a junction. In the second graph, the size of the index perturbation in a 100 nm layer at the bottom of a waveguide is varied. Changing doping levels to cause more or less carrier injection or depletion would cause the same type of difference in a real structure. Finally, the third graph shows how effective index change changes as the active region is increased to cover the whole guide. Increasing bias voltage to affect a larger device area would accomplish the same thing in an actual device.

Looking at the first graph, it should be noted that the index perturbation has a clear peak. Because the mode is concentrated in the center of the guide, as the layer of refractive index change approaches this region, the effective index change increases, and as it moves away it decreases. The difference in effective index change between the location of maximum overlap and that of minimum overlap is greater than 30%. Clearly the location of the junction relative to the modal maximum is important. The slope of the third graph also illustrates this point. Here,

the effective index change increases quickly as the perturbation layer moves across the modal maximum at the center of the guide, then rolls off as new index change is placed in the top of the guide where there is not as much of the optical mode. Finally, the second graph shows the effective index increase with increase in the strength of the index perturbation. The important thing to notice here is that this increase is linear with refractive index change. This is to be expected as the effective index change is approximately the refractive index change times the confinement factor, and the confinement factor is not changing.

In general, then, the placement of regions of carrier change and the geometry of the waveguide are important parameter spaces that need to be considered carefully during modulator design.

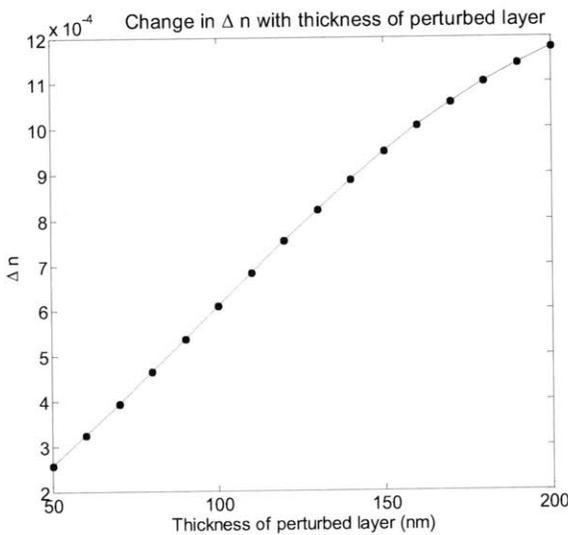
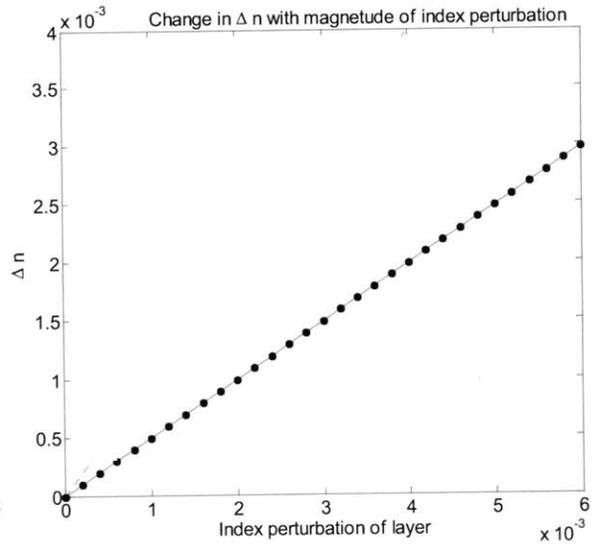
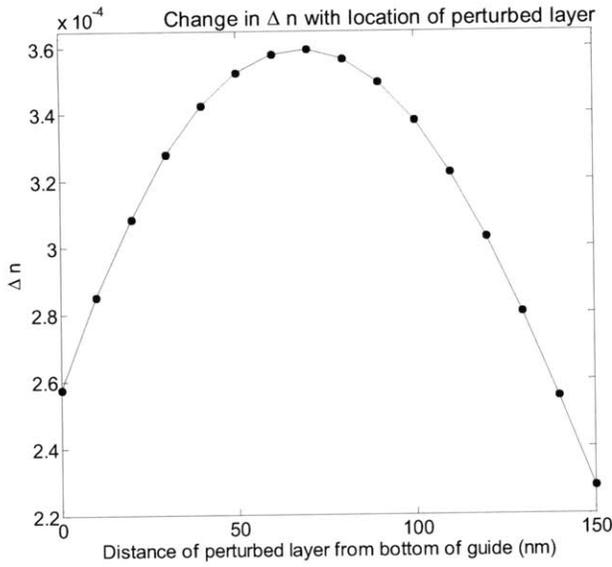


Figure 13: (a) Change in Δn_{eff} as a 50 nm perturbation equivalent to the injection of 5×10^{17} holes is moved from the bottom to the top of a silicon rib waveguide 500 nm wide \times 200 nm tall with a slab height of 50 nm. The guide is surrounded by SiO_2 . (b) Change in Δn_{eff} for the same guide as the strength of the index perturbation is increased from 0 to 6×10^{-3} (equivalent to injecting 3.7×10^{18} holes). (c) Change in Δn_{eff} for the same guide with an index perturbation equivalent to the injection of 5×10^{17} holes, as the affected portion of the guide is expanded from the bottom 50 nm to cover the whole 200 nm guide height.

Active Devices for Moving Carriers

Intrinsic electronic devices can be used to dynamically control carrier concentrations. There are several different devices that can be used and they can be biased in one of several operating regions. Diodes can be operated in forward or reverse bias; MOS capacitors in depletion, inversion, or accumulation. Other devices come with their own operating region choices. Each device and operating region has its own benefits and drawbacks. In particular, it is important to consider achievable speed, sensitivity, and loss. The work performed in this thesis is

predominately concerned with reverse biased diode modulators and MOS capacitor modulators biased in accumulation.

Diode Modulators

In general, diode modulators are nice because they are easy to fabricate and analyze, and because they can achieve acceptable speed and sensitivity performance. Diodes can be operated in forward or reverse bias. In forward bias, carriers injected across the junction contribute most of the index change seen. However, in some designs, modulation of the depletion region size can still have a notable effect. In the first set of diodes fabricated in our group (pictured in Figure 14), depletion width modulation accounted for nearly half of the index change observed at low forward biases. In reverse bias, it is only the modulation of the depletion region that is responsible for the index change.

Device speed is determined by the rate at which the carriers can be moved into or out of the interaction area. Forward biased devices are slower as this rate is dominated by the minority carrier lifetime (which is determined from the minority carrier recombination time – the time it takes injected carriers to recombine – and the minority carrier transit time – the time it takes injected carriers to move across the diode and out of the region of interest) [5]. In previous devices fabricated in the group, the surface recombination rate was found to be the dominant component of the minority carrier lifetime. The total lifetime was on the order of on 1 ns, which allowed for 160 MHz modulation [4].

In reverse bias, speed is determined either by the dielectric relaxation rate of silicon, the saturation velocity, or by the RC time constant of the device. While the saturation velocity places the theoretical limit on device speed (as discussed above), in practice, it is actually the RC time constant that tends to dominate in these devices. When calculating the capacitance of the device, the capacitance of the junction as well as the contacts (and, for experimental results, the measurement set-up) needs to be included. The resistance needs to include the resistance of the silicon regions between the junction and the contacts, and the contacts themselves, (and again, in practice, the measurement set-up, which typically includes a 50 Ω probe) [5]. Previous reverse biased Mach-Zehnder diode modulators designed in our group have had operation frequencies of up to 26GHz. These operation frequencies are measurement limited due to the 50 Ω probe in the measurement setup, and would, presumably, increase if the probe resistance could be lowered [5].

While reverse biased diodes are significantly faster than forward biased ones, they are also much less sensitive, requiring significantly larger operating voltages and device footprints to work. A typical forward biased diode is operated between 0 and 1 V; a typical reverse biased diode between 0 and -5 to -10 V [5]. Furthermore, a forward biased diode with p-type doping of $N_A=1\times 10^{18} \text{ cm}^{-3}$, and n-type doping of $N_D=2\times 10^{17} \text{ cm}^{-3}$ will inject on the order of $1\times 10^{19} \text{ cm}^{-3}$ holes, while a reverse biased diode at -10 V will still only deplete the doping value of $N_D=2\times 10^{17} \text{ cm}^{-3}$. The reverse biased case is aided slightly by the fact that, at higher voltages, depletion can occur across the entirety of the guide, while the number of injected carriers in the forward biased case will always fall off sharply away from the junction. However, it turns out that forward biased modulators are still much more sensitive. Previous results in our group obtained a $V_\pi L$ value in forward bias of 0.005 Vcm when the bias point was chosen to be 0.95V and a value in reverse bias of 4 Vcm at bias point of -7V.

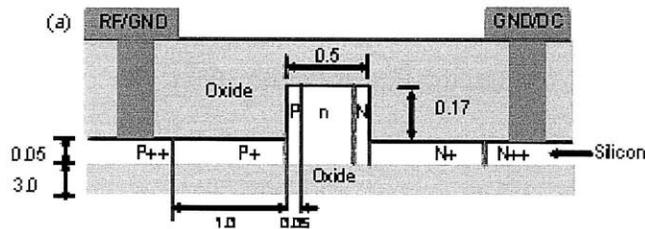


Figure 14: Previous modulator design. From [5]

Work in this thesis concentrated on improving the reverse biased diode sensitivity. The previous reverse biased modulator design used a vertical p-n junction at the side of a single mode rib-waveguide (see Figure 14). Carriers were injected or depleted from this vertical junction area. This junction structure does not optimally overlap the optical mode traveling through the device: much of the index change, in both forward and reverse bias, is “wasted” in areas with relatively little light. To increase optical mode overlap, several different junction schemes were considered (see Figure 15). Ease of fabrication as well as expected sensitivity improvement were considered for all cases. Because sensitivity improvement is associated with the injection or depletion of a larger number of carriers, increased sensitivity is usually associated with increased loss (as described above) and with decreased speed. This last tradeoff comes from the fact that increasing the number of carriers moved into or out of a reverse biased diode by a given voltage change is usually associated with increased device capacitance. Physically, the increase in

capacitance results either from increased junction area or from increased doping levels causing a decrease in depletion region width (see below). The higher capacitance then decreases overall speed. These tradeoffs were also taken into account when picking the final modulator design.

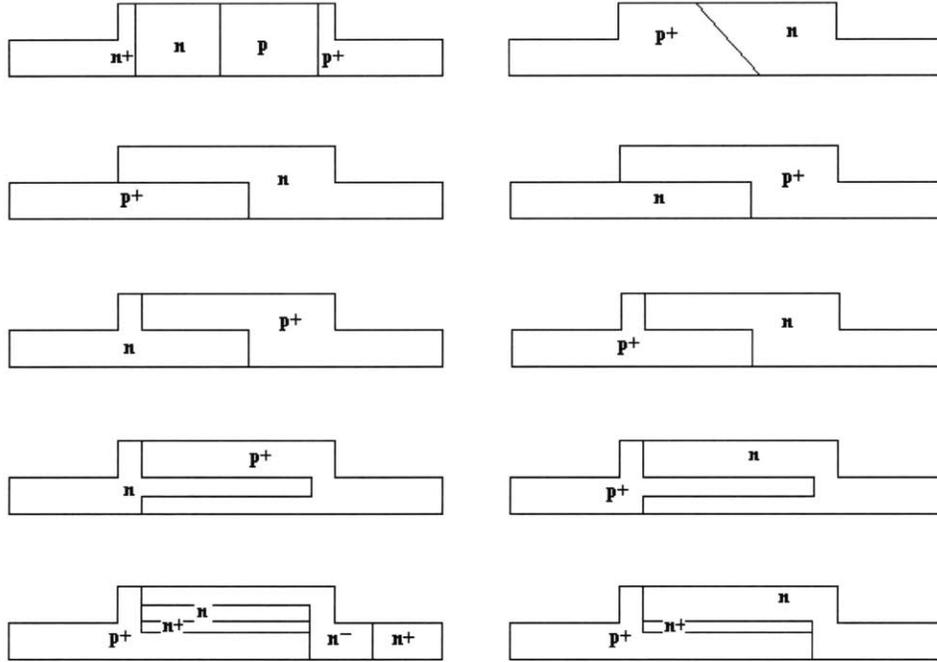


Figure 15: Diagram of proposed diode modulator schemes

To obtain a rough estimate of modulator sensitivity and speed the Schottky diode model can be used. In this model, the doping profiles are assumed to be abrupt with a line denoting the junction between the n-type and the p-type materials. The depletion region is modeled as a region of some width, x_n , on the n-side of the junction and, x_p , on the p-side of the junction that contains zero carriers. The width of this region as a function of voltage is given by:

$$x_n = \sqrt{\frac{2\epsilon_{Si}N_A}{qN_D(N_A + N_D)}(\phi_{bi} - V)} \quad \text{and} \quad x_p = \sqrt{\frac{2\epsilon_{Si}N_D}{qN_A(N_A + N_D)}(\phi_{bi} - V)}.$$

where N_D is the doping concentration of the n-type material, N_A is that of the p-type material, ϵ_{Si} is the dielectric constant of silicon, q is the electron charge, ϕ_{bi} is the built-in bias of the junction, and V is the applied voltage. From these widths, the area of the modulated region can be found and, together with the doping concentration and the modal confinement factor to that area, the effective index change can be determined. The losses accumulated by the mode can be calculated from the confinement in regions that still contain carriers.

In reverse bias, the capacitance can be considered to be equal to the parallel plate capacitance for two sheets of charge separated by the width of the depletion region.

$$C = \frac{A_{\text{junct}} \epsilon_{\text{Si}}}{(x_n + x_p)}$$

where A_{junct} is the area of the junction. Assuming a 50Ω load, the predicted measurement-limited RF speed of the modulator can be determined. Rough simulations using the full electrical and optical model described in the "Modulator Modeling" section of this work, were run for the structures that seemed promising.

A horizontal junction placed across the top of the waveguide (for ease of fabrication) was chosen as the desired structure (see Figure 16) [5]. This structure was then optimized using more precise simulations.

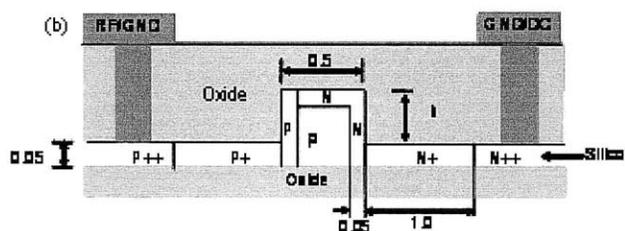


Figure 16: Proposed Modulator Design. From [5].

The device width was fixed by other system parameters. However, the height and the doping levels were optimized. To start, the n and p regions were switched as compared to the previous design; the hole region was placed in the center of the mode to take advantage of the larger effective index change caused by holes. To help keep losses low, the horizontal high-doping layer was kept as thin as possible (50 nm), and the height changed to maximize the modal overlap. The doping level of the high doping layer was set to $1 \times 10^{18} \text{cm}^{-3}$, again to minimize losses, and the doping level of the center of the guide was varied to determine maximum sensitivity. A two dimensional parameter sweep through doping and waveguide height values was run. The effective index changes as a function of voltage for various doping levels at one height and of waveguide heights at one doping level are shown, for example, in Figure 17.

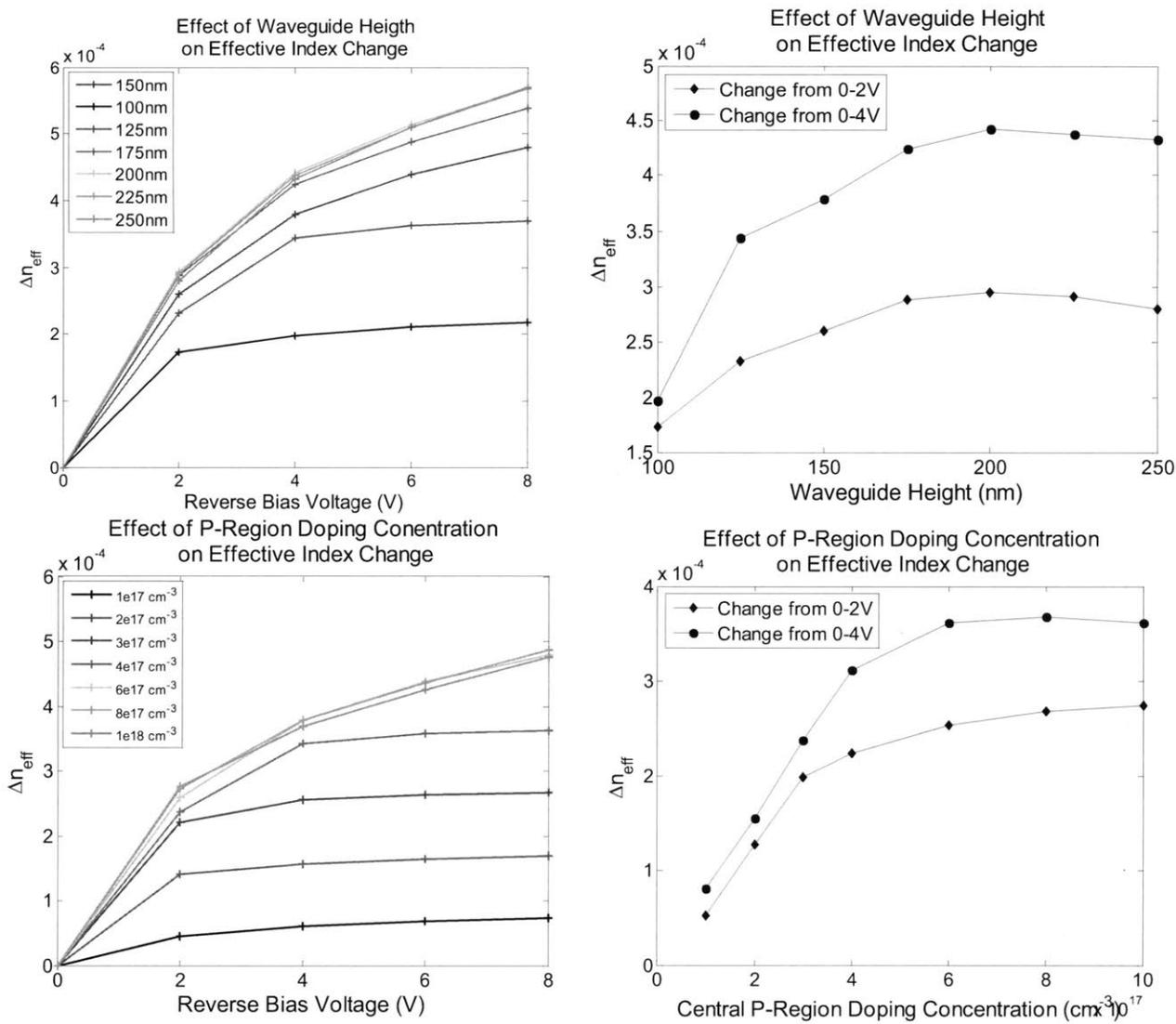


Figure 17: (a) Change in effective index vs. bias voltage at various doping levels for a 500 nm wide 150 nm high silicon rib waveguide, (b) Change in effective index vs. doping level for at various bias differences for same waveguide, (c) Change in effective index vs. bias voltage at various heights for a 500 nm wide waveguide with a center doping of $6 \times 10^{17} \text{ cm}^{-3}$, (d) Change in effective index vs. height at various bias differences for same waveguide

From Figure 17, it can be seen that there is an optimal height at this doping level, as expected. Increasing the waveguide height increases the modulation depth quickly. However, after the maximal height, the decrease is much slower. In part, this imbalance comes from the fact that the depletion region edge at zero volts is some distance away from the junction. A junction placed at the center of a narrow guide then yields index change not at the guide center but, rather, at the bottom of the guide (in this case 35 nm or more below the junction). The short waveguides not only have poor confinement, but also poor modal overlap. Additionally, in short

waveguides, the guide may become fully depleted after a certain voltage: that is, the depletion region extends across all the silicon in the guide. Increasing the applied voltage then yields little change in the carrier concentration (except at the extreme edges of the mode) and the index change levels off. This effect can be observed in the bottom two curves of graph (a) in Figure 17. After a certain height, the index modulation in this design seems to be fairly insensitive to changes in height – an advantage when it comes to fabrication tolerances. This is because the better modal confinement and the poorer modal overlap balance each other somewhat over this region. If the guide height were to increase further, performance would eventually decrease significantly.

When the doping concentration is varied at a particular height, the story is a bit more complex. Increasing the doping concentration both increases the size of the index change in the modulated region and decreases the size of the modulated region. In general, the depletion region on one side of the junction is proportional to $1/\sqrt{N_{\text{dop}}}$. If the mode were spread evenly throughout the guide, increasing the doping would always help as the effective index (which would be proportional to the affected area times the index change) would be roughly proportional to $\sqrt{N_{\text{dop}}}$. However, because the mode is not evenly spread, the decrease in doping region size may also hurt overlap between the index and the mode. Therefore, increasing the doping may not always help, and may sometimes hurt. Indeed, in graphs (c) and (d) of Figure 17, the response to the doping increase eventually levels off. The change in modulated region width also explains why a two-dimensional parameter sweep was necessary: the modal overlap changes with both doping and height. One other consideration of increased doping is increased loss. The mode in the undepleted region of the waveguide (which increases in size at higher dopings) sees loss from the carriers contributed by the doping in this area (also higher at higher dopings). Because of this, the lowest acceptable doping level should usually be used.

From the optimization simulations, a diode height of 200 nm and a doping concentration of $6 \times 10^{17} \text{ cm}^{-3}$ p-type was chosen for the center and $1 \times 10^{18} \text{ cm}^{-3}$ n-type for the top layer. The device's simulated DC performance was compared to the simulated performance of the previous design used in the group. As can be seen in Figure 18, the performance is significantly improved. We expect about a seven fold increase in device sensitivity ($V_{\text{pi}}L$) over the vertical design: half of which is due to the change in doping and half to the horizontal junction [5].

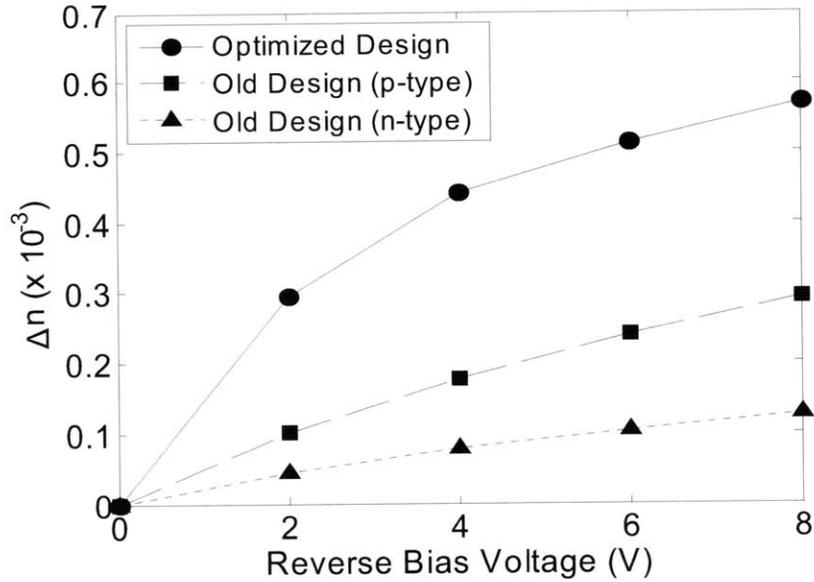


Figure 18: Simulation results comparing the old and new modulator design sensitivity in reverse biased operation. The old design is plotted both as originally fabricated (triangles) and with the doping polarizations switched (squares) to take advantage of the greater index modulation from holes. In the new design the center of the waveguide is doped p-type. From [5].

It is worthwhile to explore the predicted RF characteristics of this device. As mentioned above, in a reverse biased diode the operating frequency is limited by the RC time constant of the diode. Because the internal resistance of the modulators is quite low (below about $0.3 \Omega\text{-cm}$), it can be neglected in comparison to the 50Ω load from the measurement set up [5]. It is then the capacitance of the device that dominates the frequency cutoff of operations. Provided good, low capacitance contacts are made to the device, the capacitance will be dominated by the diode junction capacitance. It will, thus, increase as the area of the junction increases and as the equilibrium width of the depletion region decreases (that is, as the doping level rises.) Turning the junction from vertical to horizontal thus increases the capacitance, as does increasing the doping level to achieve better sensitivity. The capacitance of the new device was calculated from AC simulations carried out in Sentaurus to be 4.9 pF/cm (of device length.) For comparison, the previous design had a capacitance of 1.9 pF/cm . Therefore, it is predicted that a 500 micron long version of this new device will only work up to 13 GHz, and not out to the 26 GHz achieved before [5]. If the resistance present in our measurement probes could be reduced, the device could, presumably, be operated at significantly higher frequencies. The frequency cut off of the

intrinsic device (not counting contacts or measurement probes), was simulated using Sentaurus and found to be 270 GHz.

The proposed device was fabricated at Lincoln Laboratory. The final device height was 220 nm, the p-type central doping was $5 \times 10^{17} \text{ cm}^{-3}$, the n-type doping across the top was $1.5 \times 10^{18} \text{ cm}^{-3}$ and did not extend all the way across, but was instead 50 nm short of the opposite side. The expected index change with voltage for this device variation was simulated and is shown in Figure 19. In this design the losses are expected to be between 16.7 and 5.8 dB/cm depending on applied voltage. The expected $V_{\pi}L$ for this structure is 0.7 Vcm when biased from 0 V to -4 V, 1.1 Vcm when biased from 0 to -10V, and 0.5 Vcm when biased from 0 to -2 V.

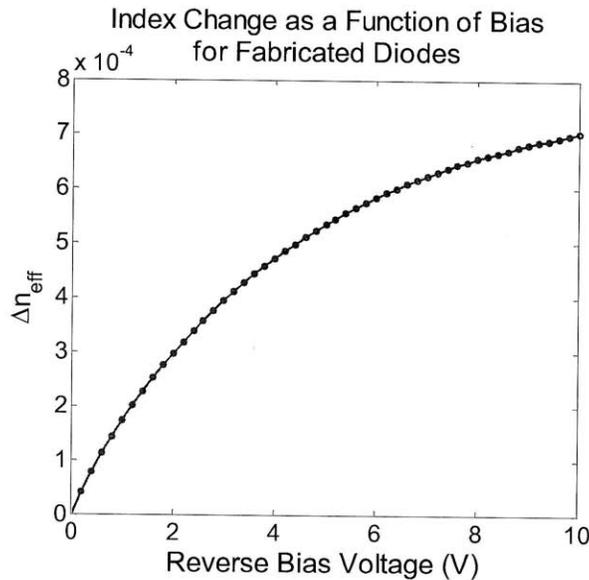


Figure 19: Simulated effective Index change for fabricated diodes

The devices were measured using a set-up designed by Jason Orcutt in Rajeev Rams's group and diagrammed in Figure 20. In this setup, a tunable laser was used to take data at several wavelengths from 1549.9 to 1550.1 nm to account for possible effects from Fabry-Perot fringes. Because such fringes were not found to be a problem, the data at 1550 nm were used during analysis. Light was coupled into and out of the waveguides through the use of lensed fiber. The polarization and the coupling were adjusted until the output power reached a maximum. An IR camera was used to visually inspect the light traveling through the waveguides. Several bright scattering spots were observed, corresponding to possible defects in the waveguides. The wavelength and the optical power out of the laser were monitored during measurements, and relatively little drift was observed. Finally, only the top modulator output was measured.

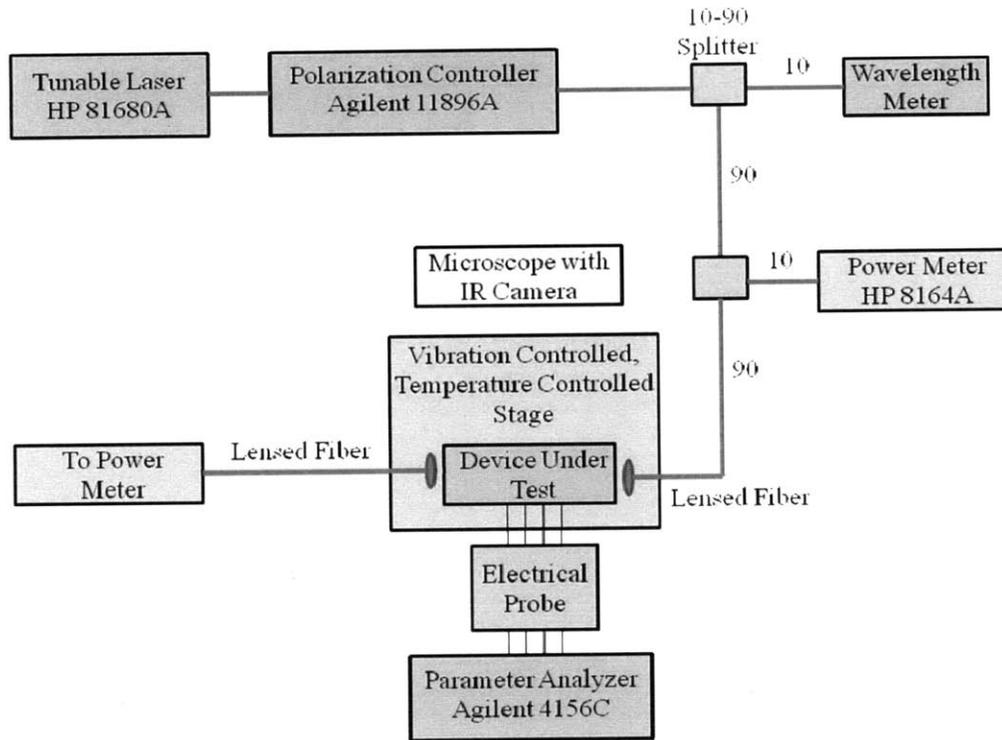


Figure 20: Diagram of modulator test set-up.

To characterize the DC response of the modulator, the arms were driven in push-pull: a DC bias point, V_{DC} , was chosen, the top contact of the top diode was placed at $2V_{DC}$, the bottom contact of the bottom diode was set to ground, and the middle, joint contact was stepped through from 0 to $2V_{DC}$ (see Figure 21). This is equivalent to biasing at V_{DC} with a swing of $\pm V_{DC}$. This sweep was repeated for several different heater bias currents in order to determine the quadrature bias-point of the heater and to get a complete picture of the DC response of the device. The experimental data for a V_{DC} bias-point of 5 V and 2 V are shown in Figure 24 and Figure 25. To check the stability of the measurement, two complete runs at a V_{DC} of 5 V were taken on the same modulator. These are plotted against each other in Figure 22. As can be seen, the measurements are consistent.

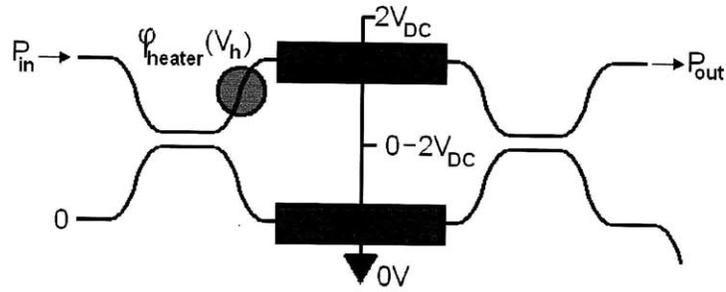


Figure 21: Diagram of biasing in experimental set-up

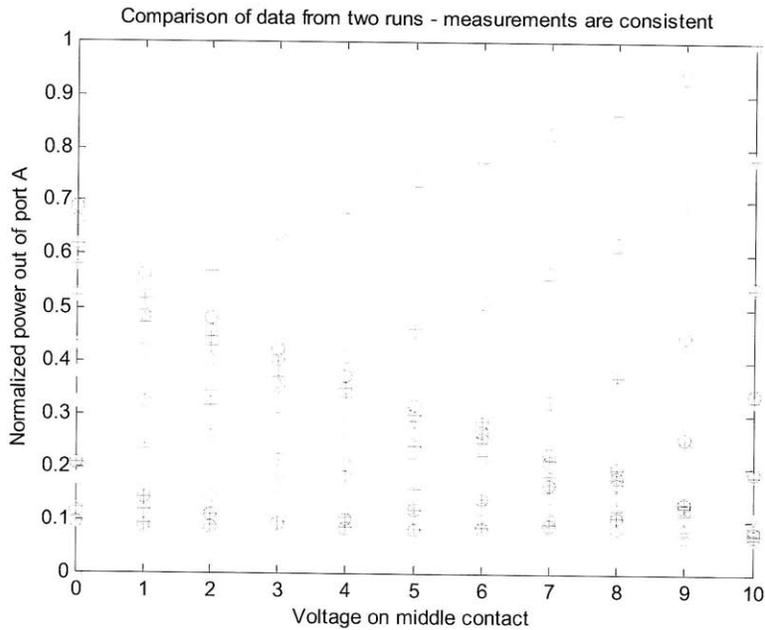


Figure 22: Comparison of two different sets of data taken on the same modulator in the same set up. Measurement drift is minimal.

To fit the experimental data, it was first necessary to determine the effect of heater bias on the phase of the output, and to normalize the output. The heater phase is expected to be proportional to the square of the current put into the heater. Thus, the output power as a function of the square of the heater current should be sinusoidal when the bias across both arms is the same, as it is at the DC bias point. A sine-function was fit to this curve (see Figure 23). From this fit, the phase offset of the two arms at zero heater bias, $2F_x F_m$, the proportionality factor between the heater bias and the heater phase, F_x , and the amplitude of the output sine wave at the DC bias point, F_A , can be determined. The theoretically predicted output power for a given heater current and voltage input on the center contact, is then given by:

$$P_{out} = \frac{F_A \exp\left(-\frac{L}{2}(\alpha_{V_{top}} + \alpha_{V_{bot}})\right)}{\exp(-L\alpha_{V_{DC}})} \sin^2\left(F_x \left[I_{heater}^2 + F_m\right] + \frac{\varphi_{V_{top}} - \varphi_{V_{bot}}}{2}\right) + \frac{F_A}{4 \exp(-L\alpha_{V_{DC}})} \left[\exp\left(-L\frac{\alpha_{V_{top}}}{2}\right) - \exp\left(-L\frac{\alpha_{V_{bot}}}{2}\right) \right]^2$$

where L is the modulator length, I_{heater} is the heater current, $\alpha_{V_{top}}$ is the predicted loss in the modulator in the top arm at the applied bias, $\alpha_{V_{bot}}$ is the same for the bottom arm, $\varphi_{V_{top}}$ is the predicted phase modulation in the top arm, $\varphi_{V_{bot}}$ is the same in the bottom arm, and $\alpha_{V_{DC}}$ is the predicted loss at the DC bias point.

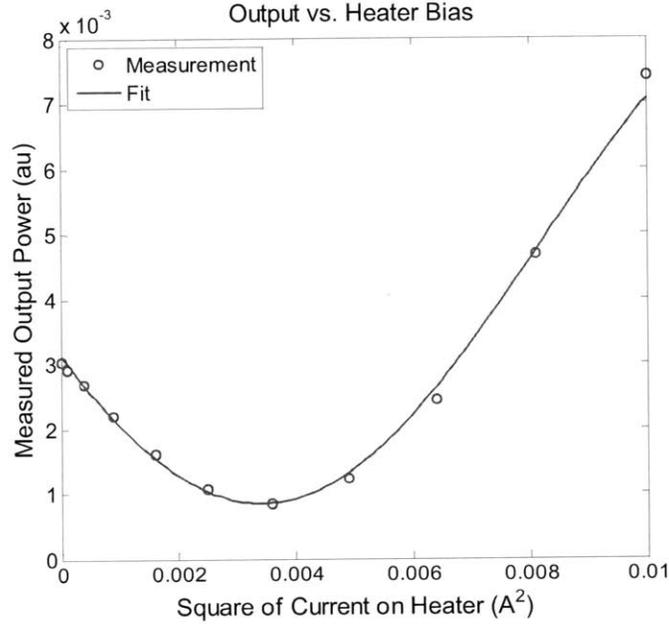


Figure 23: Fit to output as a function of the square of the heater current

The theoretical fit to the experimental data taken at the V_{DC} bias-points of 5 V and 2 V is also shown in Figure 24 and Figure 25. The quadrature point was determined to correspond to 90 mA placed on the heater. The measured value for $V_{\pi}L$ was determined to be 1.2 Vcm (at the 2 V bias), which is almost a factor of 4 improvement over our previous designs, but almost 2 times higher than the predicted value of 0.7.

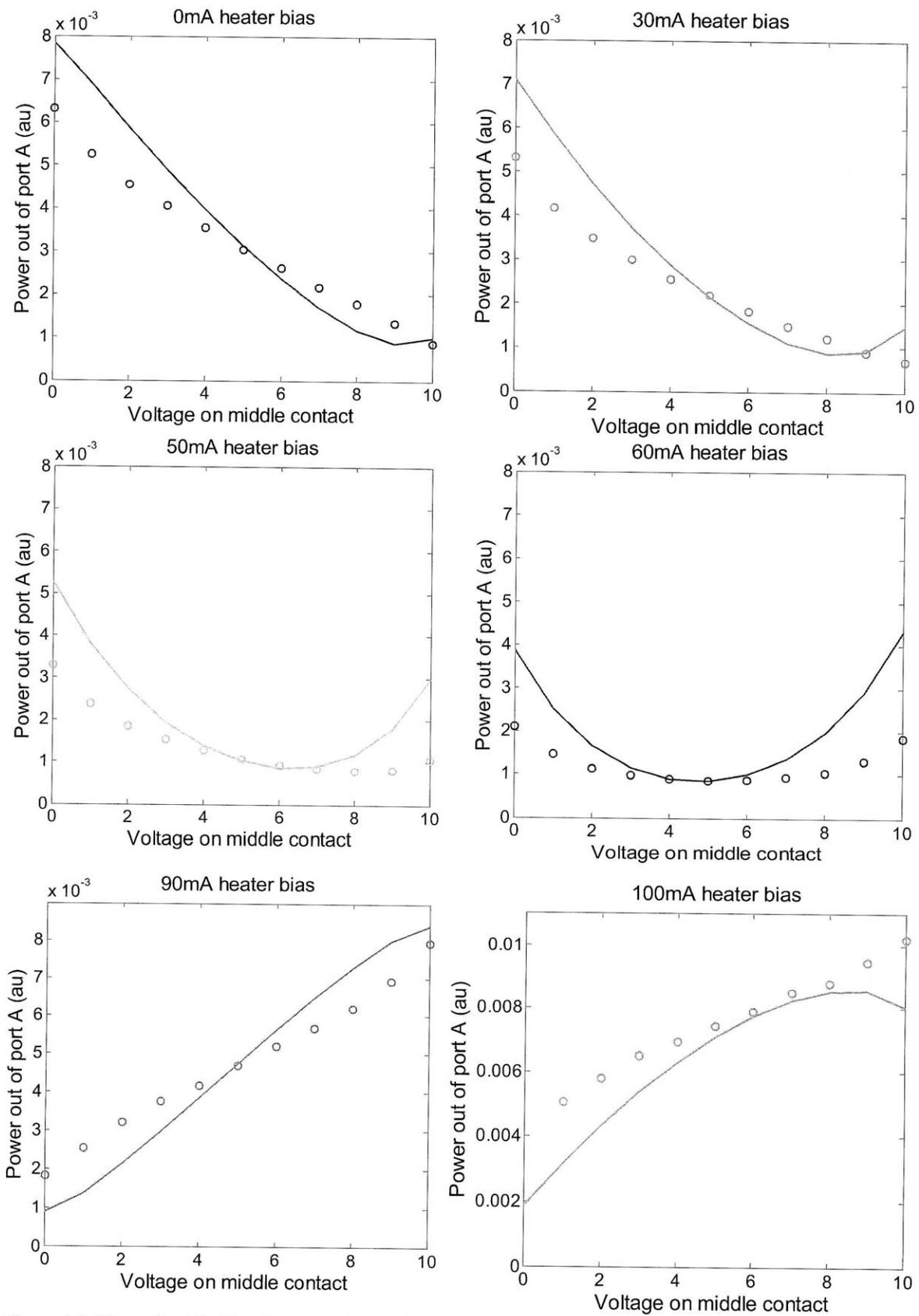


Figure 24: Theoretical fit (lines) to experimental data (circles) for $V_{DC} = 5V$ and a variety of heater bias currents.

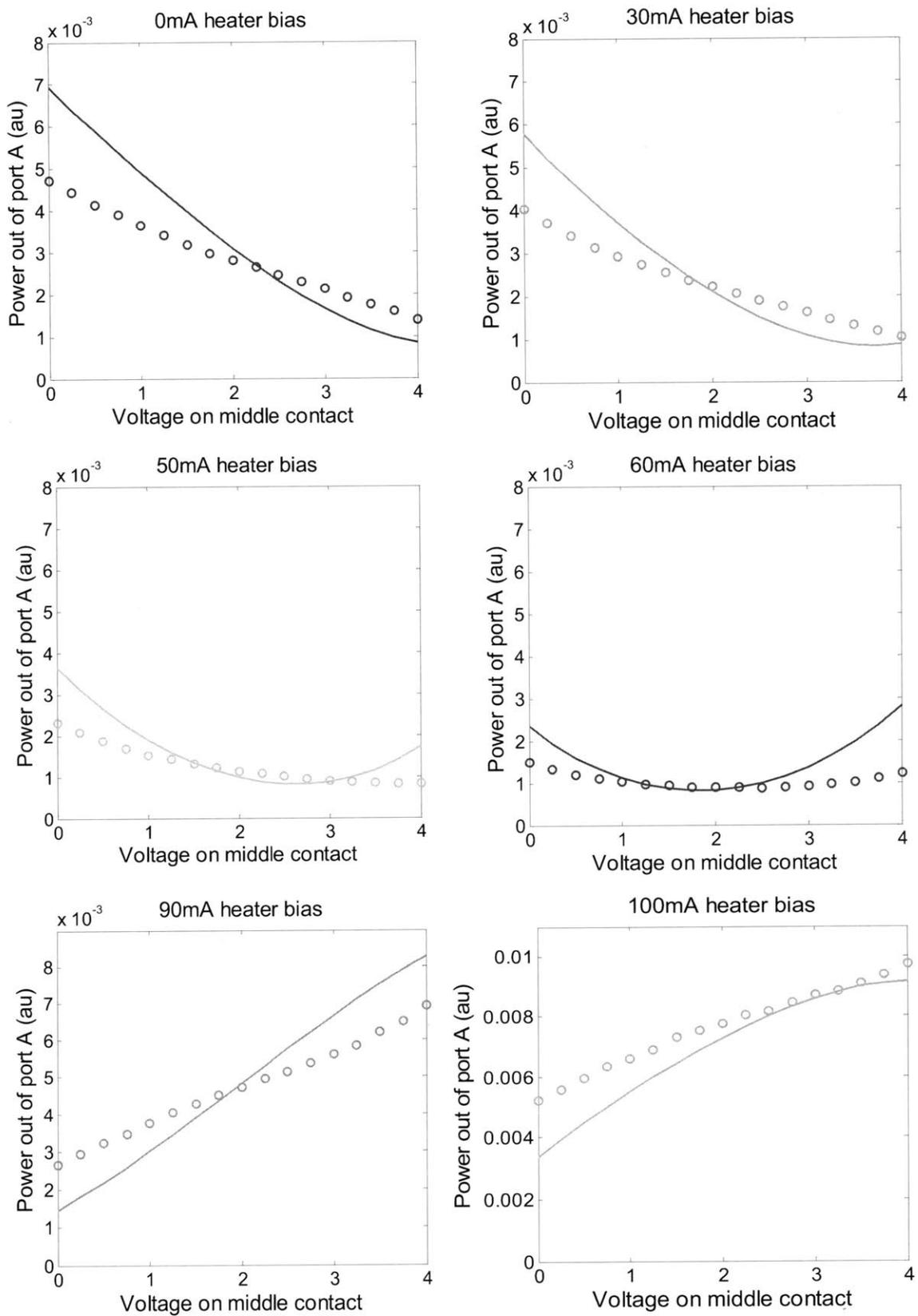


Figure 25: Theoretical fit (lines) to experimental data (circles) for $V_{DC} = 2$ V and a variety of heater bias currents.

As can be seen, the theoretical fit follows the general trend of the measurements well. However, it tends to overestimate the phase-change observed. Why this is the case is currently being investigated. The frequency response of these diodes was measured at LL and the 3 dB roll-off point was determined to be 12 GHz, which is in good agreement with the predicted value.

MOS Capacitor Modulators

MOS capacitor modulators can be run in the depletion, inversion, or accumulation regime. Operation in the depletion regime is similar to operation in the reverse bias regime for a diode. However, while the depletion region in a reverse biased diode continues to expand with additional applied voltage, there is a maximum depletion region width in a MOS capacitor that corresponds to the onset of strong inversion. After the depletion region hits its maximum width, it expands minimally with additional voltage, the device is in inversion, and the situation changes. Additionally, the depletion region width in a MOS capacitor biased some amount above its flatband voltage is similar to the change in depletion region width of an equivalently doped diode biased the same amount above zero [8]. This can be seen in Figure 26. This means that the device is not significantly more sensitive than its diode counterpart in the depletion regime. Furthermore, because the depletion width may not expand to the desired extent (inversion may be hit first), depletion of enough of the waveguide to create a π or $\pi/2$ phase-change may not occur at any voltage. Finally, because oxide thicknesses are often kept small for fabrication reasons, the total capacitance of a MOS capacitor device in depletion is often higher than that of an equivalently sensitive reverse biased diode. Thus, the MOS capacitor will have a higher RC time constant, and be slower. Using a MOS capacitor device biased in depletion is therefore not expected to be beneficial. However, if fabrication permits large oxide thicknesses, higher speed operation can be achieved and significantly faster speeds at the same sensitivity can be possible.

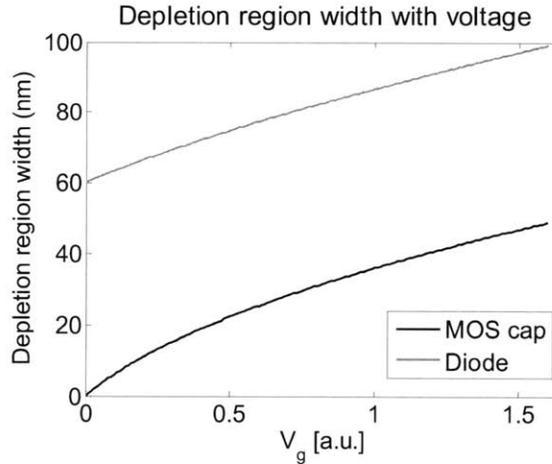


Figure 26: Calculated width of the depletion region of a diode with $N_A=5\times 10^{17}$, $N_D=1\times 10^{18}$ biased from 0 to 1.6 volts (red) and the same for an n+-polysilicon-gated p-type MOS capacitor with $N_A=5\times 10^{17}$ and oxide thickness of 6nm biased from flatband at -1 V to the onset of strong inversion at +0.6 V (black).

In both the accumulation and inversion regimes, high levels of charge are confined in a thin layer near the junction. One might think that this is similar to the forward bias regime, but this is not quite the case. In a MOS capacitor, this layer of charge is confined much more closely to the junction itself – within a few dozen nanometers at most, as illustrated in Figure 27 [8].

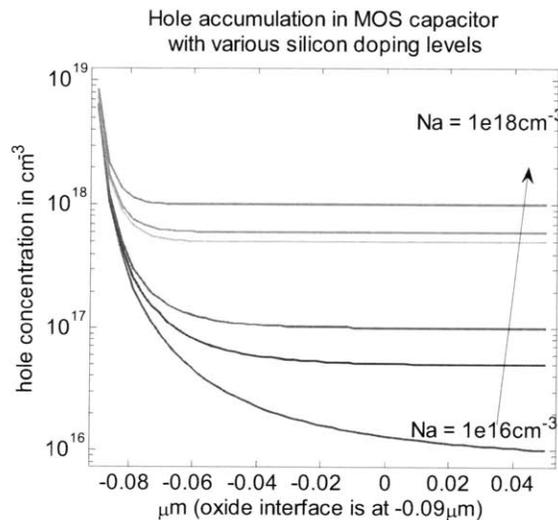


Figure 27: Cross section of hole concentration in a MOS capacitor device in accumulation biased 2V from the flatband voltage. Various doping levels are shown. The thinness of the accumulation layer, especially at high doping concentrations can be seen.

In the accumulation regime the accumulated thin charge layer is made up of *majority* carriers – this means that the layer can be created and destroyed at high speed, and the limiting factor, in practice, is again that of the RC time constant. In accumulation, to first order, the capacitance

per area is simply ϵ/t_{ox} , where t_{ox} is the oxide thickness. In reality, there will be a correction to account for the finite thickness of the accumulation layer in the semiconductor (and, in a polysilicon gate, for the finite thickness of the gate charge layer). This correction will lower the overall capacitance, so the above value can be treated as a maximum [42].

In the inversion regime, the charge layer is made up of minority carriers. However, unlike the diode case, where carriers are pulled across the junction, the inversion layer is isolated by the depletion region and does not have easy access to a source of minority carriers. While the inversion charges can recombine quickly, it can take on the order of 1 μ s - 10 ms to generate these carriers [8]. Speed can be improved by shining a light on the junction (which generates electron-hole pairs to replenish the inversion layer), or by attaching an oppositely doped piece of silicon to the edge of the inversion layer to act as a source of minority carriers (see Figure 28). Once a source of minority carriers is provided, the capacitance of the inversion layer can also be approximated simply as the parallel plate capacitance between the inversion layer at the semiconductor-oxide interface and the equivalent charge layer at the metal-oxide interface [8], and high-speed operation can be realized.

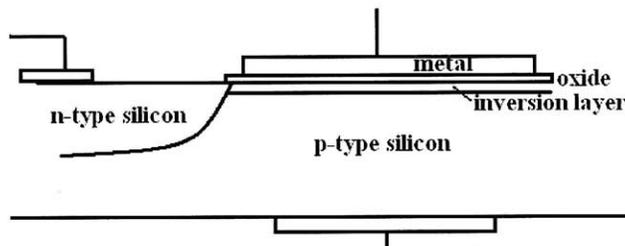


Figure 28: A MOS capacitor in inversion with a source of minority carriers provided to the junction.

Because inversion regime operation is more complicated and offers no obvious advantage over accumulation regime operation, only accumulation regime operation will be considered here for modulator implementation. To start, the speed of a given MOS capacitor modulator section is limited by the RC time constant of the device. As stated above, the capacitance per area can be well approximated by ϵ/t_{ox} , this value as a function of oxide thickness is plotted in Figure 29:

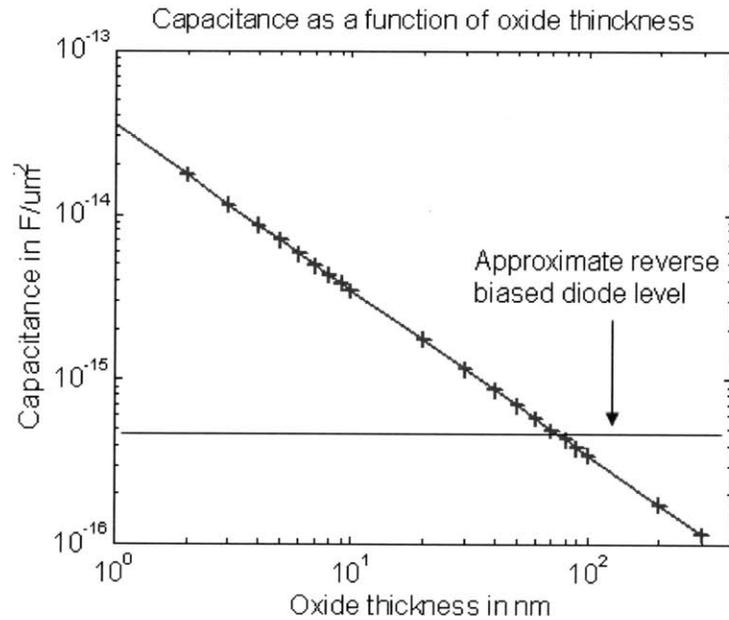


Figure 29: Capacitance per junction area as a function of oxide thickness. Note that for thin oxides, this value is an over estimate as the finite accumulation layer thickness will increase the effective oxide thickness slightly

If we assume again that the 50Ω resistance of the measurement probe dominates, then we expect to see a measurement-limited speed of 246 MHz to 13 GHz for oxide thicknesses of 2 to 100 nm and a junction area of $250 \mu\text{m}^2$. Of more interest is the intrinsic speed of the device, which was found to be around 140 GHz for a 10nm oxide via Sentaurus simulations. This makes this structure faster than a forward biased diode modulator, but slower than a reverse biased one.

To obtain a good estimate of possible MOS capacitor modulator performance, it was assumed that the contacts could be placed where desired, without regard to geometric constraints. Therefore, a horizontal junction (allowing for more accumulation layer overlap with the mode) was implemented here as well. The basic geometry of the simulated structure is shown in Figure 30.

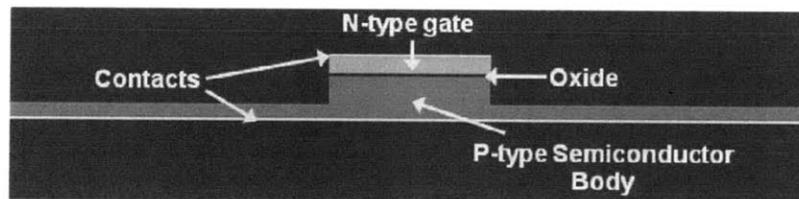


Figure 30: Geometry of simulated MOS capacitor phase-shift structure.

Because this is not a physically realizable structure, rigorous optimization of the structure parameters was not undertaken. Instead the relative impacts of some of the parameters were explored and a general idea of the performance achievable in a MOS capacitor modulator was formed. Since modal overlap is important for modulator performance, and because the accumulation layer is so thin, a reasonable junction location and rib height need to be chosen. The gate height is again picked to be 50 nm in order to minimize losses. To pick the height of the semiconductor rib, simulations were carried out in MATLAB using an approximation to the accumulation layer width and charge density as described below.

Though the thickness of the accumulation layer can be approximated to first order as [43]

$$t_{acc} = \sqrt{2}L_D$$

where L_D is the Debye length of the material, the charge is not actually spread evenly throughout the accumulation layer. The resulting accumulation layer thickness obtained from this approximation (shown in Figure 31), seem to imply that the thickness of the accumulation charge at low doping is significantly larger than that at higher ones (the width is inversely proportional to the square root of the doping concentration). However, if Sentaurus is used to plot the precise carrier distribution of the accumulation layer (see Figure 27), it can be seen that, though there is still accumulation at these distances, the vast majority of charge at all dopings is concentrated within about 10 nm of the junction. This, therefore, was the accumulation layer thickness chosen for use at all dopings in the MATLAB simulations.

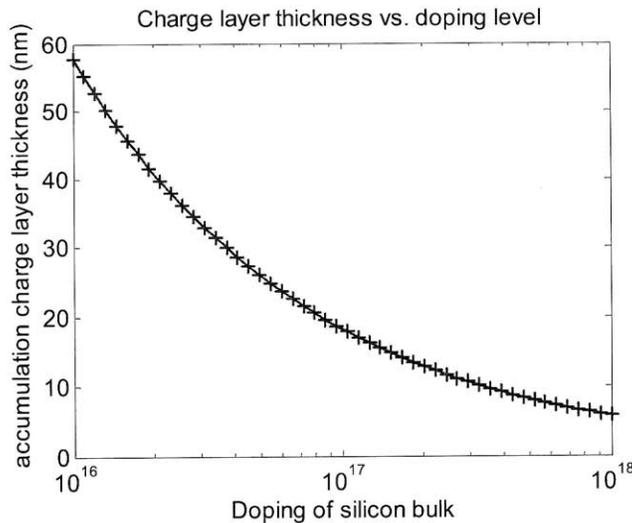


Figure 31: Accumulation Layer thickness as a function of semiconductor doping level.

Because a MOS capacitor is a capacitor, the amount of charge it stores, per area, is approximated well by

$$Q_{acc} = C_{ox}(V_{FB} - V)$$

where V_{FB} is the flat-band voltage of the MOS capacitor – that is the voltage at which the energy bands of the device are flat, and there is neither depletion nor accumulation at the oxide interface (there is no charge on the capacitor). $C_{ox} = \epsilon/t_{ox}$ as explained above. The flatband voltage is a function of the doping concentration (see Figure 32). Lower doping in the p-type region means that less voltage is necessary to pull its band into line with the n^+ -doped region being used as the "gate." Thus, the overall voltage sensitivity will depend on the doping. However, after the flatband voltage, the effect of a change in voltage on the accumulation layer charge is negligible to first order. This means that, were the mode evenly spread in the guide, for equivalent changes in the voltage above flatband, doping values would not matter. Because practical accumulation layer width is not a strong function of doping, it is expected that MOS capacitor modulator sensitivity to voltage above flatband, in any design, will not be a strong function of doping. However, a lower doping means lower baseline losses contributed from carriers outside of the accumulation layer. On the whole, then, it is expected that lower doping levels in the semiconductor region will yield more desirable performance.

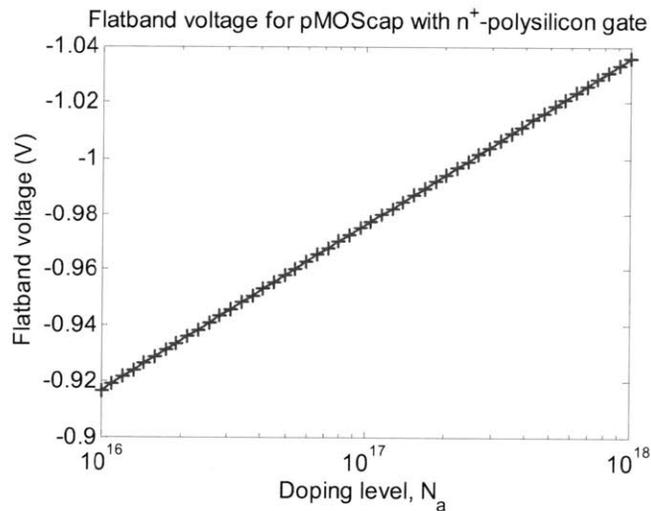


Figure 32: Flatband voltage as a function of doping level. Note that the flatband voltage is independent of oxide thickness.

It was stated above that the accumulation layer charge, to first order, is independent of everything except the oxide thickness. However, a more accurate approximation is given in [43].

The resulting values from these two equations for a doping of $N_A=1\times 10^{18} \text{ cm}^{-3}$ and an oxide thickness of 6 nm are plotted in Figure 33. As can be seen, the agreement is quite good. It is therefore determined that the original approximation can be used.

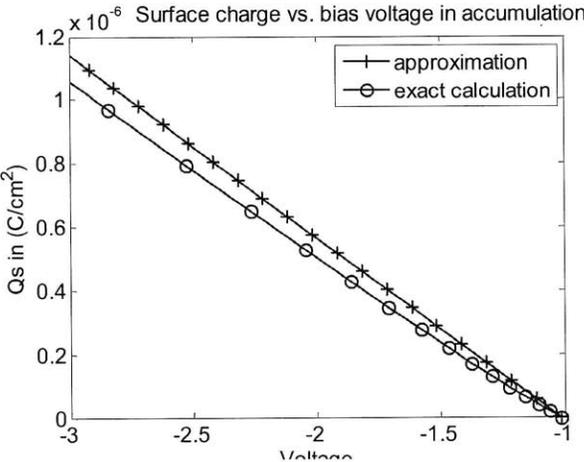


Figure 33: Check of rough approximation against more accurate approximation. About the same results are found.

Using the above calculations for accumulation layer width and charge (at 1 V above flatband), the rough dependence of modulator response as a function of height can be plotted (see Figure 34). A 150 nm height is chosen for use in later electrical simulations.

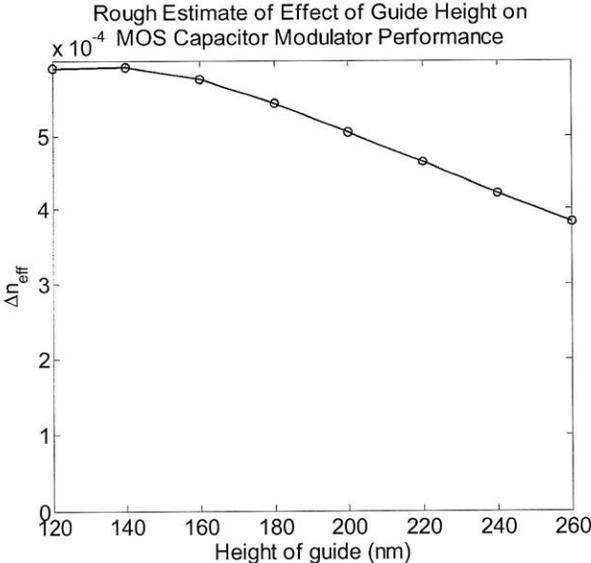


Figure 34: Estimated effect of guide height on MOS capacitor modulator performance. Note that while the effect is noticeable, it is not particularly strong.

To determine the actual performance of the MOS capacitor modulator, simulations were run in Sentaurus for various semiconductor doping levels. The results are plotted in Figure 35.

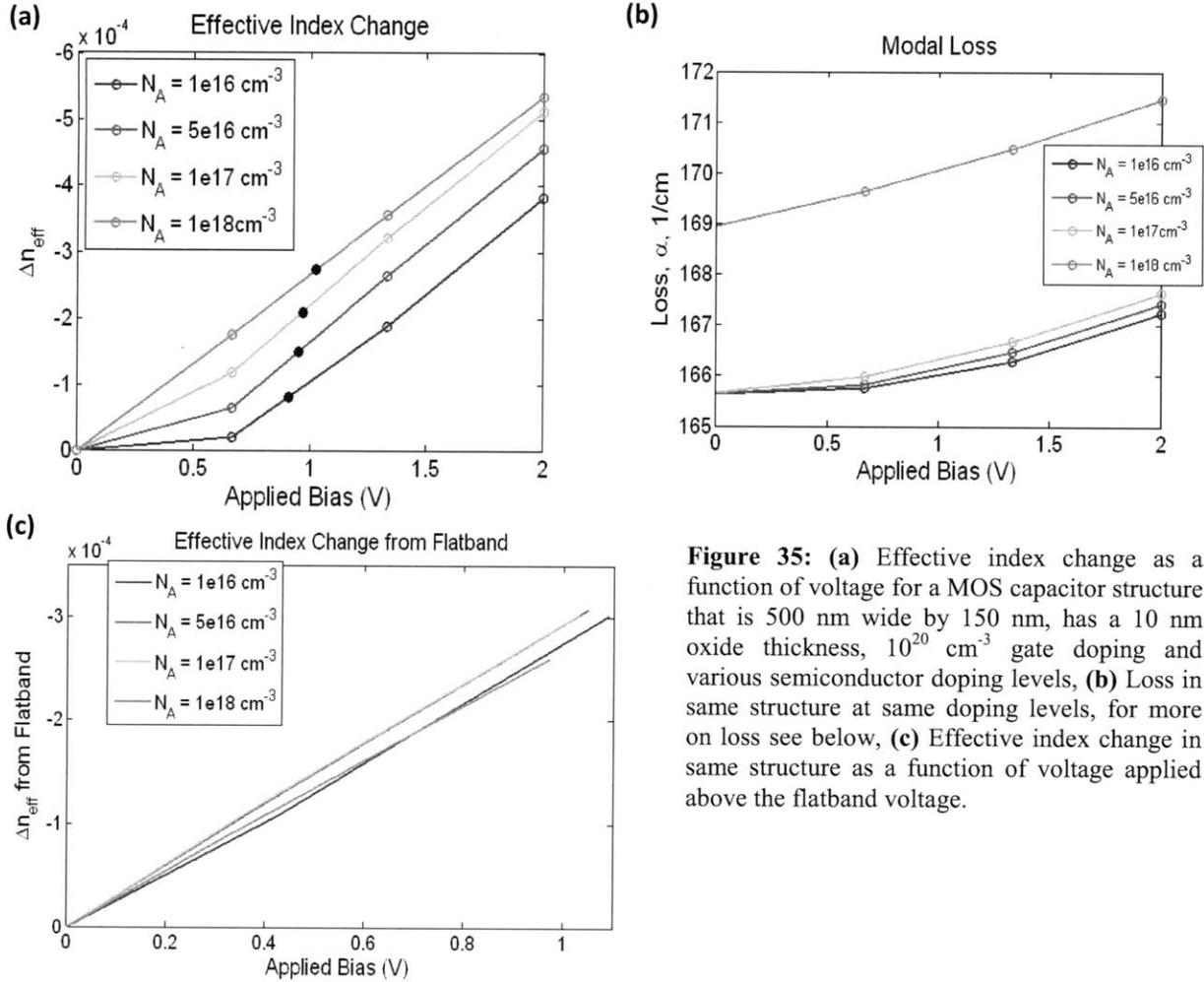


Figure 35: (a) Effective index change as a function of voltage for a MOS capacitor structure that is 500 nm wide by 150 nm, has a 10 nm oxide thickness, 10^{20} cm^{-3} gate doping and various semiconductor doping levels, (b) Loss in same structure at same doping levels, for more on loss see below, (c) Effective index change in same structure as a function of voltage applied above the flatband voltage.

As can be seen, below the flatband voltage (black circles), the effective index change is dominated by the change in the depletion layer width (as it decreases from its equilibrium value to zero). It turns out that this difference more than makes up for variation caused by the difference in flatband voltage location or modal overlap. Thus, it ends up being the higher doping levels that show a larger overall index change from 0 V. If just the voltage above the flatband voltage is considered, we see that the response is relatively insensitive to doping concentration, as expected. The sensitivity of this set of modulators falls somewhere in between that of forward biased diodes (which are more sensitive) and that of reverse bias diodes (which are less so). As stated above, the operating speed of these modulators also falls in between the

two cases. It is worth noting that the figures of merit achieved seem to be more similar to those of a reverse biased modulator than those of a forward biased one. While MOS capacitor modulators may be capable of some improvement over diode modulators and may be favored for some applications, movement to a MOS capacitor design does not appear to present a path towards next-generation modulator performance.

Finally, it is necessary to consider the losses. In the simulations above, a gate doping of $1 \times 10^{20} \text{ cm}^{-3}$ was used. This led to very high predicted losses for the modulators—on the order of -725 dB/cm or -36 dB through a 500 micron device—with lower semiconductor doping levels yielding lower losses as expected. Increasing guide height (to decrease modal overlap with the gate) is one way to lower the losses. If the guide height is raised to 200 nm, the loss is reduced to -356 dB/cm while the index modulation only decreases by around 10%. However, lowering the gate doping provides a much better way to decrease the loss. As long as the gate contains enough charge, electrical performance and modulation depth will be similar [42]. Because of this, the same simulations were re-run with a variety of different gate dopings and are plotted in Figure 36 for a semiconductor doping of $1 \times 10^{16} \text{ cm}^{-3}$.

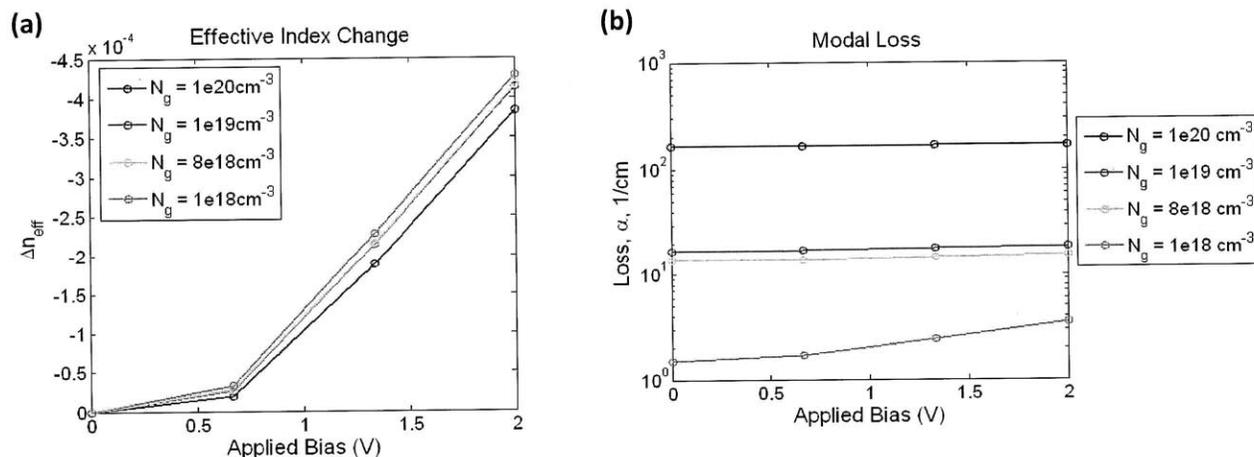


Figure 36: Effective index change, (a), and modal loss, (b), as a function of applied bias for a 500 nm wide by 150 nm tall MOS capacitor structure with a 10 nm oxide, a semiconductor doping of 10^{16} cm^{-3} , and various gate doping levels.

As can be seen, changing the doping of the gate did not impact the effective index results significantly, but did have a dramatic, beneficial impact on the loss levels. Because of this, if a MOS capacitor modulator was to be designed and fabricated, low doping levels should be used for the gate region.

Other Modulator Electrical Structures in the Literature

In addition to diodes and MOS capacitors, there are several other published junction designs that are aimed at maximizing modulator performance. One design recently put forward in [12] uses a P-I-N-I-P electrical junction in a double ring modulator (only one ring of which is optically active). The design exploits the high fields created by the device to sweep carriers in and out of the intrinsic region of the active ring at nearly the saturation velocity of silicon. 100 GHz operation is predicted. In [13], modulation in modified MOS structures using multiple SOI layers is theoretically explored, and modulation speeds of 10 Gbit/sec with $V_{\pi}L < 1$ V-cm are predicted. In [24], modulation is accomplished in ring resonators by modulating the coupling coefficient instead of the ring region. This has been proposed several times in the past. Finally, in [37], the performance of a SiGe heterojunction bipolar transistor modulator is modeled and a $V_{\pi}L$ of 0.01 Vcm is achieved for speeds up to a few gigahertz. In all, these structures are very promising and movement away from diode and MOS capacitors toward a more complicated electrical structure is, most likely, necessary to meet desired modulator performance targets.

LINEARIZATION OF MZ MODULATORS

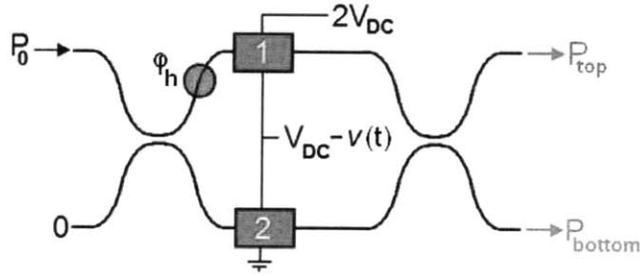
Linear modulators are important in all analog optical applications including CATV and analog optical links. Linear modulator performance is also important to the accuracy of our ADC system; if the modulator is not linear, it will not imprint the applied RF-signal onto the optical carrier accurately. Specifically, the nonlinearities will distort the signal slightly as it is transferred, creating errors in optical pulse height, and, thus, pulse energy. When the pulse is then detected and quantized, these slight errors will translate into errors in the ADC output, thus decreasing the effective number of bits achieved by the system. If 10 ENOB operation is desired, all harmonics need to be suppressed to *at most* -60dBc, and in reality below than this [7]. One solution is to decrease modulation depth, staying on the more linear parts of the curve. However, to maintain a high signal-to-noise ratio (SNR), modulation depth must be kept large. This is important, because if the SNR becomes too low, it will take over from total harmonic distortion or spur free dynamic range as the limiting effect for ENOB and may also create problems in reaching desired accuracy levels.

Currently proposed schemes to linearize modulator output can be split into one of two categories: electrical schemes and optical schemes [35]. Electrical schemes involve the use of electrical circuits to somehow compensate for the Mach-Zehnder transfer function. Examples include pre-emphasis of the signal to cancel out MZ linearities, feedback schemes, and inverting the MZ transfer function in post-processing [35]. Optical schemes involve designing an optical system that has a more linear output than a simple MZ modulator with linear phase-shifting sections does on its own. Such schemes usually involve cascading modulators, splitting signals, and including other elements, such as non-3 dB couplers, in an attempt to achieve linear output. The resulting systems are usually more complicated and use more power than a single MZ modulator [35].

Scheme Proposed in this Work

Examination of the nonlinearities present in a MZ modulator with nonlinear phase shifters shows that phase shifter nonlinearity can be used to cancel out the nonlinearity inherent in the sinusoidal nature of the Mach-Zehnder transfer function. To see why this is so, the MZ transfer function is recalculated to include the effects of modulator sections with arbitrary voltage dependent phase and loss.

The MZ modulator is assumed to be biased in push-pull, with a heater section that adds some additional phase given by φ_h , and with a phase-shifting section in each arm. Let $\varphi_1 = \varphi(v_{DC} + v)$ represent the phase added by the top arm as a function of voltage, where v_{DC} is the dc bias voltage and v is the rf signal. Let $\varphi_2 = \varphi(v_{DC} - v)$ be the equivalent for the bottom arm, and $\alpha_1 = \alpha(v_{DC} + v)$ and $\alpha_2 = \alpha(v_{DC} - v)$ be the voltage dependent losses through the two modulators. It is further assumed that the input and the output couplers are ideal 50-50 splitters, that light is input only on the top arm, and that the input light has power P_0 .



Performing the same mathematical steps as in the "Mach-Zehnder Modulators-Rational and Transfer Function" section above, the MZ transfer function including losses is found to be [36]:

$$P_{top} = P_0 \exp\left[-\left(\frac{\alpha_1}{2} + \frac{\alpha_2}{2}\right)\right] \sin^2\left(\frac{\varphi_h}{2} + \frac{\varphi_1 - \varphi_2}{2}\right) + \frac{P_0}{4} \left[\exp\left(-\frac{\alpha_1}{2}\right) - \exp\left(-\frac{\alpha_2}{2}\right)\right]^2$$

$$P_{bottom} = P_0 \exp\left[-\left(\frac{\alpha_1}{2} + \frac{\alpha_2}{2}\right)\right] \cos^2\left(\frac{\varphi_h}{2} + \frac{\varphi_1 - \varphi_2}{2}\right) + \frac{P_0}{4} \left[\exp\left(-\frac{\alpha_1}{2}\right) - \exp\left(-\frac{\alpha_2}{2}\right)\right]^2$$

It can be seen that the transfer-function is still generally sinusoidal in nature and that the two outputs are still complimentary. However, it can also be seen that the sinusoidal term (which is main term responsible for imprinting the RF signal onto the optical mode) is multiplied by the combined and voltage dependent loss in the two arms. In an ideal, lossless modulator, this multiplier would be equal to one. In the case were the losses were constant in voltage, this multiplier would be a simple constant and could be wrapped into P_0 , where, other than its effect on SNR (due to decreasing signal power), it would have no effect on accuracy. Furthermore, there is a second term due purely to the voltage-dependent losses. This term captures the fact that, if the losses in the two arms are different, the amplitude of the output from the arm with more loss will not be high enough to cancel the output from arm with less loss completely or to

the extent that it ought. This adds a non-constant voltage-dependent offset to the output, and prevents full modulation depth from being achieved [36].

The question of the modulator's linearity can be addressed by Taylor expanding these expressions. The math is equivalent for the upper and lower outputs of the MZ modulator. It will, therefore, be performed for the top output only, with the results for the bottom output simply stated without proof where necessary.

First, the phase and voltage are Taylor expanded, and the terms up through the third order nonlinearity are kept. The phase and absorption effects of the modulator sections are then given by:

$$\varphi_m(v) \approx \left(\varphi_{mDC} + a_m v + \frac{b_m v^2}{2} + \frac{c_m v^3}{6} \right) L$$

$$\alpha_m(v) \approx \left(\alpha_{mDC} + x_m v + \frac{y_m v^2}{2} + \frac{z_m v^3}{6} \right) L$$

where m is an index into the different phase-shift sections; a , b , c , x , y , and z are the coefficients of the Taylor expansion of the phase-change per length and absorption per length respectively around the desired DC bias point; and L denotes the length of the modulator itself. These expansions next need to be substituted into the transfer-functions obtained above, and the resulting equations Taylor expanded to obtain the linear and nonlinear output terms.

In Quadrature with Balanced Arms

For a first analysis, it is assumed that the modulator is in quadrature, and that the phase-shifting sections are identical—that is that the phase change and absorption created when a certain voltage is applied is the same on both arms: $\varphi_1(v) = \varphi_2(v)$ and $\alpha_1(v) = \alpha_2(v)$. Substituting the above expansions into the transfer function then yields:

$$\begin{aligned}
P_{top} &= P_0 \exp \left[- \left(\frac{\alpha_{DC} + \cancel{xv} + \frac{yv^2}{2} + \frac{\cancel{zv^3}}{6} + \alpha_{DC} - \cancel{xv} + \frac{yv^2}{2} - \frac{\cancel{zv^3}}{6}}{2} \right) \right] \\
&\times \sin^2 \left(\frac{\pi}{4} + L \frac{\cancel{\phi_{DC}} + av + \frac{bv^2}{2} + \frac{cv^3}{6} - \cancel{\phi_{DC}} + av - \frac{bv^2}{2} + \frac{cv^3}{6}}{2} \right) \\
&+ \frac{P_0}{4} \left[\exp \left(- \frac{\alpha_{DC} + xv + \frac{yv^2}{2} + \frac{zv^3}{6}}{2} \right) - \exp \left(- \frac{\alpha_{DC} - xv + \frac{yv^2}{2} - \frac{zv^3}{6}}{2} \right) \right]^2
\end{aligned}$$

A key point to notice at this stage is that there are no harmonic contributions from the second derivative of the phase; for identical diodes the push-pull configuration causes all even derivatives of the phase as well as the DC contribution to cancel. In push-pull the voltages applied to the two diodes are exactly equal and opposite (usually the center contact is shared). If the responses of the diodes are also exactly equivalent, subtraction will cause all the odd derivatives to double and all the even ones to cancel. This is important because, for many modulator designs, such as the reverse biased diode modulator of the next section, the second order nonlinearity of the phase is the largest nonlinearity and would otherwise dominate the response [36].

Taylor expanding the above equation, and keeping the first three terms, then yields:

$$\begin{aligned}
P_{top} &\approx P_0 e^{-\alpha_{DC}L} \left[\frac{1}{2} + (aL)v + \frac{1}{4}(x^2L^2 - yL)v^2 + \left(-\frac{2}{3}a^3L^3 + \frac{c}{6}L - \frac{1}{2}ayL^2 \right)v^3 \right] \\
P_{bottom} &\approx P_0 e^{-\alpha_{DC}L} \left[\frac{1}{2} - (aL)v + \frac{1}{4}(x^2L^2 - yL)v^2 - \left(-\frac{2}{3}a^3L^3 + \frac{c}{6}L - \frac{1}{2}ayL^2 \right)v^3 \right]
\end{aligned}$$

The first term corresponds to the output at the DC bias point (orange), where quadrature simply splits the input evenly between the two channels. The second term (yellow) is the desired, undistorted, linear response of the modulator: it's simply the input RF voltage scaled by a constant. The third term (blue) is the second order nonlinearity and, assuming all higher harmonics are negligible, gives the second harmonic response of the device. As can be seen, the

second order nonlinearity comes only from the voltage-dependent loss, and would not be present were the phase-shift sections lossless. Importantly, this term has the same sign in the two arms, while the linear terms have opposite signs. Therefore, measuring both outputs and subtracting can cancel this harmonic. Finally, the last set of terms represents the third order nonlinearity, and, if higher order nonlinearities can be ignored, the third harmonic. The first of these terms (red) is the third-order nonlinearity from the MZ transfer function itself – it comes from the interaction of the linear phase change with the sinusoidal transfer function. In a MZ modulator with ideal, lossless, linear phase-shifters, this is the only contribution to the third harmonic that would be present, and it is the contribution (and the harmonic) that would ultimately limit modulator linearity and, if modulator nonlinearity dominated, system performance. The second contribution to the third order nonlinearity (green) comes from the third-order nonlinearity of the RF-signal-to-phase-shifter transfer function and is a direct result of the phase shifters not being linear themselves. The final contribution (purple) is a cross term between the linear phase response and the loss and comes from the fact that the sinusoidal transfer function is multiplied by a loss that is not constant in voltage and, thus, adds some additional variation to the outputs [36].

Looking at this equation, one can see that the presence of either a cubic nonlinearity in the phase response, or a quadratic nonlinearity in the absorption response of the modulator, or both will yield an equation for a length at which the 3rd harmonic term will go to zero. This length is given by:

$$L = \frac{1}{4a^2} \left(\frac{-3y}{2} \pm \sqrt{\frac{9y^2}{4} + 4ca} \right)$$

Where a , c , and y are the Taylor coefficients for the first and third derivatives of the phase and the second derivative of the loss respectively *at that given bias point*. This is important because it means that a modulator of a given length will only have a linear response at a particular DC bias point. For any given phase-shifter section the question remains as to whether the length

given by the equation above is physical (that is positive and real) and realizable (that is practical to use in an actual device) [36]. In at least one case, that of a reverse biased silicon diode modulator, this length is indeed both (see the section entitled "Application of Scheme to Reverse Biased Silicon MZ Modulator" below for more details) [36].

Out of Quadrature with Balanced Arms

It was mentioned in the previous section that one way to cancel the remaining second harmonic from the loss was to use both complementary outputs and subtract. Another option commonly used is to adjust the phase added by the bias element (that is the heater) such that the second harmonic is canceled in one of the outputs. This method is particularly advantageous when there are differences in the losses between the output of the MZ structure and the location where that output is used (such as different coupling losses into the next element) as it does not require that the two outputs be balanced. However, it has the disadvantage of moving away from the more linear part of the sine curve, which can make other harmonics worse and can also decrease the achievable modulation depth (the linear part of a sine corresponds to the area of largest slope). If we denote the bias added by the heater as φ_h and consider the Taylor expansion of the top output at an arbitrary heater bias, we find:

$$P_{top} = P_0 e^{-\alpha_{DC}L} \left\{ \begin{array}{l} \sin^2\left(\frac{\varphi_h}{2}\right) + 2\sin\left(\frac{\varphi_h}{2}\right)\cos\left(\frac{\varphi_h}{2}\right)aLv \\ \left[\left(\frac{x^2 L^2}{4} - \frac{1}{2}\sin^2\left(\frac{\varphi_h}{2}\right)Ly \right) + \left(\cos^2\left(\frac{\varphi_h}{2}\right) - \sin^2\left(\frac{\varphi_h}{2}\right) \right) aL \right] v^2 \\ + \sin\left(\frac{\varphi_h}{2}\right)\cos\left(\frac{\varphi_h}{2}\right) \left[-\frac{4}{3}a^3 L^3 + \frac{1}{3}cL - ayL^2 \right] v^3 \end{array} \right\}$$

We note that, *regardless of the bias phase* we achieve cancellation of the 3rd order nonlinearity at the same length and DC bias.

The second harmonic can be canceled using this method as long as:

$$\varphi_h = 2 \sin^{-1} \left(\sqrt{\frac{x^2 L + 4a}{2y + 8a}} \right)$$

yields an achievable value for ϕ_h – that is the quantity in the arcsine is real and in the range $[-1,1]$, and the heater or other bias element is capable of adding the necessary phase-shifts (not all bias elements can tune across the full phase range).

Unbalanced Phase Responses

In practice, of course, one cannot fabricate two diodes that have absolutely identical responses. Because of this it is useful to consider the impact of slight differences between the phase-shifters on this linearity scheme. If the differences are small enough, the impact on linearity will be minimal; if they are large, achieving linear operation may not be possible. Furthermore, if the phase responses are different, use of the push-pull scheme will no longer allow for cancellation of the large second-order nonlinearity in the response of the phase. This will then affect the cancellation that can be achieved. Performing the same analysis as above yields as a set of coupled, transcendental equations for the second and third order nonlinearity. Whether an acceptable length and bias phase can be found to solve these depends on the coefficients. Actual simulation or measurement of a real modulator is most likely a faster way to determine linear performance than solving these equations. It should be noted that most of the extra terms come not from the phase, but from the voltage-dependent loss.

If the loss is considered to be constant, the third harmonic again vanishes at

$$L = \left(\frac{1}{4a^2} \pm \frac{\sqrt{ca}}{2a^2} \right)$$

where a and c are the averages of the coefficients in each arm ($a=(a_1+a_2)/2, c=(c_1+c_2)/2$) and it is assumed that the heater is tuned to make the total DC phase difference between the two arms $\pi/2$. This is an equivalent length to the lengths found in both cases above. The second harmonic would then be equal to $P_0 \exp(\alpha_{DCtotal})L(b_1 - b_2)/2$, and the difference between the quadratic coefficients of the phase on the two arms would limit performance. Alternatively, if the device is biased out of quadrature it might be possible to cancel both harmonics at once.

Application of Scheme to Reverse Biased Silicon MZ Modulator

The analysis of the previous sections is general and can be applied to any MZ modulator with non-ideal phase shifters. A reverse-biased silicon diode with the cross-section shown in Figure

37 is now used to illustrate how the principles described above can be applied to achieve linear modulation in practice. The diode cross-section was not designed with linearization in mind [36].

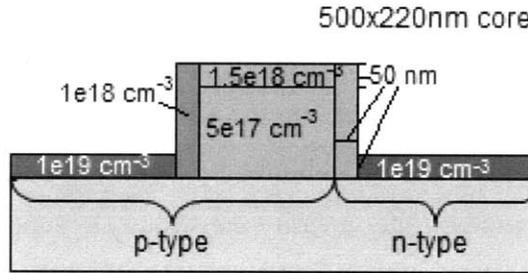


Figure 37: Cross-section of reverse-biased silicon phase-shifter section. After [36].

The DC phase change $\varphi(v)$ and absorption $\alpha(v)$ caused by the diode as a function of reverse bias voltage are shown in Figure 38. Both $\varphi(v)$ and $\alpha(v)$ are highly non-linear functions of voltage. The expansion coefficients a and c have the same sign; thus, cancellation of the cubic nonlinearity is possible at some L [36].

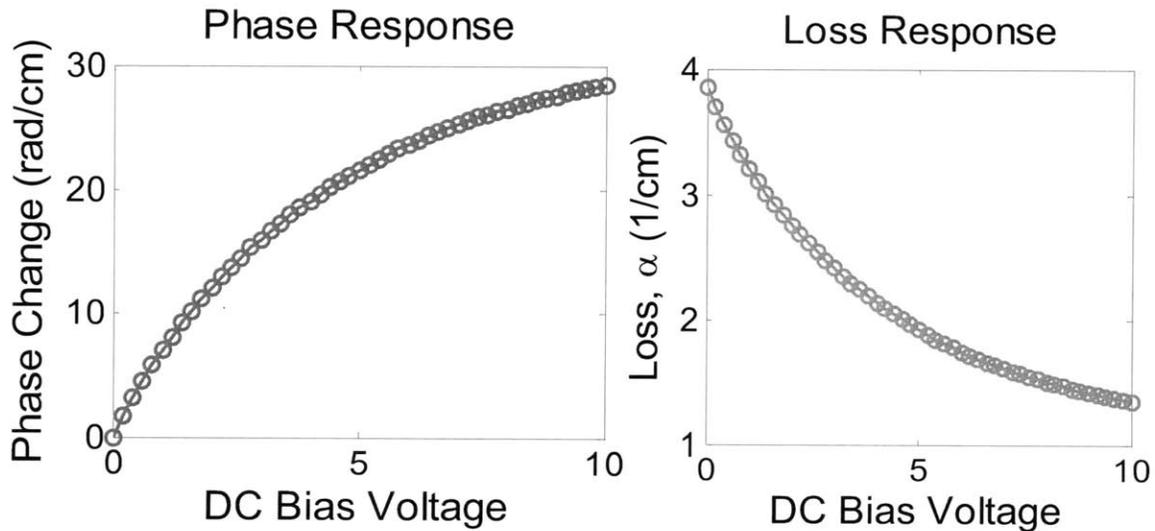


Figure 38: Phase and loss response of the modulator structure in figure X, as a function of DC bias voltage. After [36].

To find this length a high-order polynomial fit was made to the simulated response. This fit was then used to give the response of the phase-shifter to a numerically applied sinusoidal RF tone of varying amplitude. The responses were then used to calculate the output of the overall MZ modulator structure assuming that the 3 dB input and output couplers contributed no error, and that the structure was exactly in quadrature. The harmonics present in this output were then measured [36].

Figure 39 shows the calculated optimal length, L , for the above modulator as a function of bias voltage. This length corresponds to the length that keeps the calculated 3rd harmonic below -60dBc while maximizing achievable modulation depth. Modulation depth is maximized because large modulation depth is important for maintaining high SNR and because the third harmonic can always be decreased further by decreasing the RF voltage range. The achievable modulation depths are also shown in Figure 39. It should be noted that out to a V_{DC} level of 4.5 V, modulation depth is not limited by the linearity constraint (to suppress the 3rd harmonic to less than -60 dBc). Instead, in this region, modulation depth is limited by the need to keep the modulator in reverse bias – that is the RF amplitude cannot be larger in magnitude than the DC operating point or the applied voltage will be positive at some point in the swing [36].

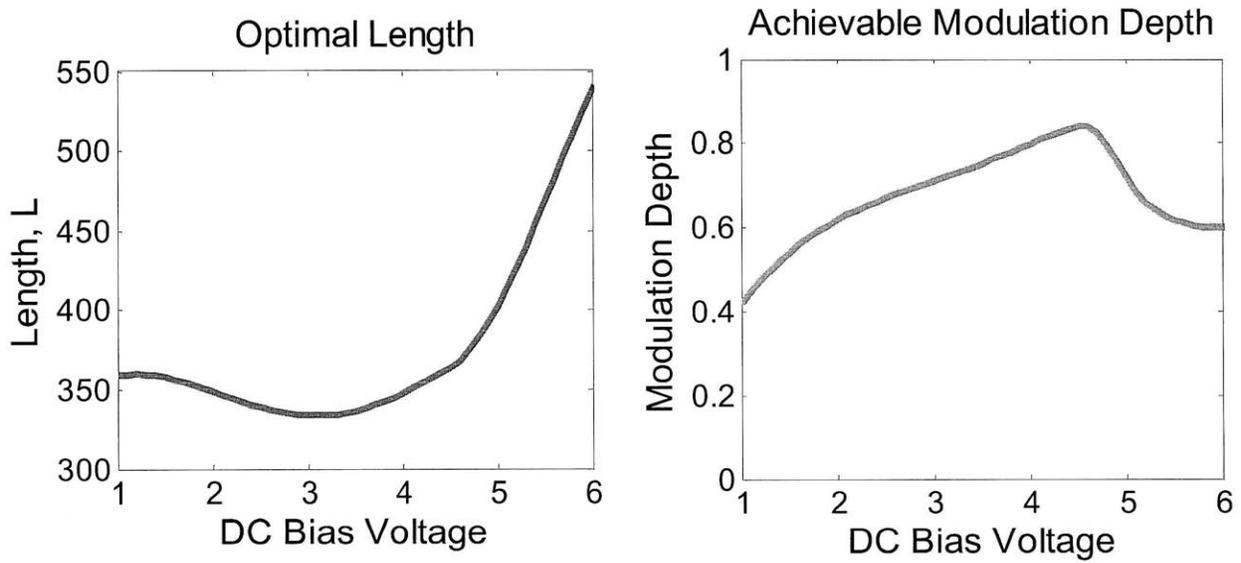


Figure 39: The maximum modulation depth achievable for third harmonic suppression to below -60 dBc (red line) and the length at which that modulation depth occurs (green line). From [36].

The lengths calculated are quite reasonable for a silicon MZ modulator. Specifically, the lengths are short enough that the losses are tolerable and the footprint is manageable, yet long enough to allow for a reasonable modulation depth. Indeed, modulators of these lengths have been successfully fabricated and demonstrated in the past (see, for example, [5]). Because the nonlinearity varies as a function of bias voltage, the optimal length also varies. This means that the bias point can be tuned to correct for errors in fabricated length, and possibly for other minor fabrication errors [36].

Figure 40 illustrates the suppression of the third harmonic in a silicon MZ modulator with a 348 μm length and a V_{DC} of 4 V. For comparison, it also shows the 3rd harmonic for an “ideal”

MZ modulator – that is a modulator with perfectly linear phase shifters – biased to achieve the same modulation depth. The 3rd harmonic of the optimized Si MZ modulator is 10-30 dB below that of the “ideal” MZ modulator at all RF amplitudes, with the most improvement seen at large amplitudes. Given that it was predicted that there would be a length at which the 3rd order nonlinearity went to zero, it might seem surprising that any 3rd harmonic is present at all. However, higher order nonlinear terms in the Taylor expansion, such as the fifth order term, contribute to the 3rd harmonic, so it does not completely vanish. For completeness, the 2nd harmonic, caused by the voltage-dependent absorption, is also plotted. As explained above, the 2nd harmonic can be completely removed either by using a differential detection scheme, or by tuning the MZ operating point [36].

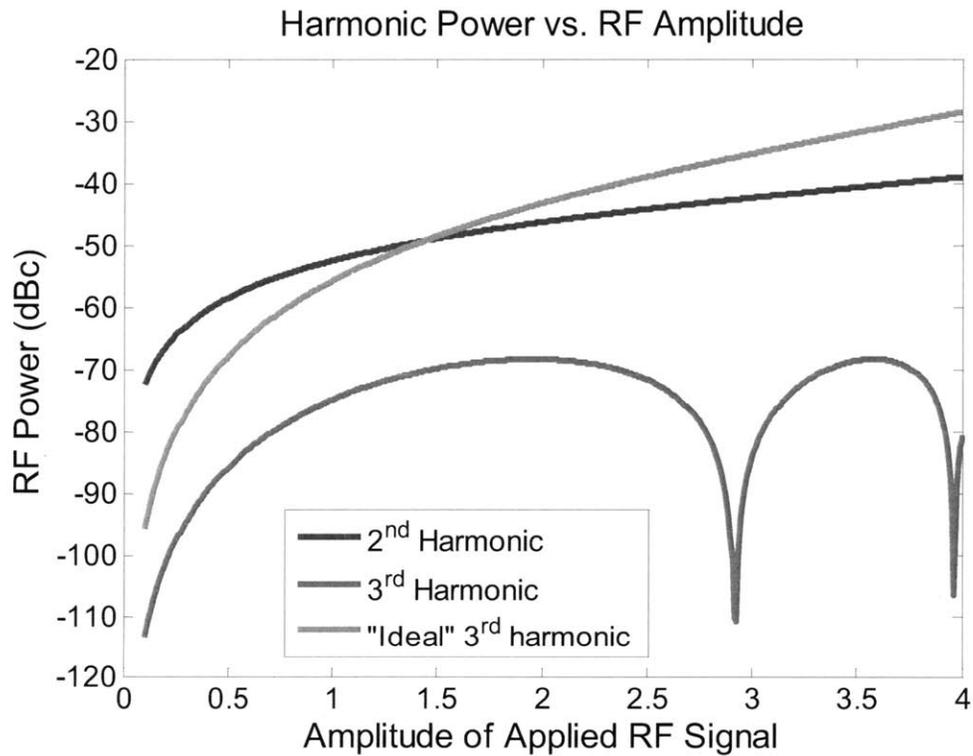


Figure 40: Suppression of 3rd harmonic in a 348 μm long MZ modulator biased at 4 V for various RF signal amplitudes. As can be seen, the 3rd harmonic of the linearized structure is suppressed well below that of a MZ modulator with ideal, linear phase shifters biased to achieve the same modulation depth. The second harmonic can be canceled by subtracting the two complimentary outputs. From [36].

The optical bandwidth over which this scheme will work is also important, particularly for the ADC application mentioned above, where a number of wavelengths need to be modulated. To address this, equivalent simulations were run for wavelengths of 1500 nm and 1600 nm. These simulations assumed that the 3 dB couplers and the heater element were not wavelength

dependent. The resulting 3rd harmonics are plotted in Figure 41. As can be seen, up to an RF amplitude of 2V (modulation depth of 39%), the 3rd harmonic is suppressed to below -60dBc for all 100 nm of bandwidth. Above this RF amplitude, the third harmonic creeps up slightly [36].

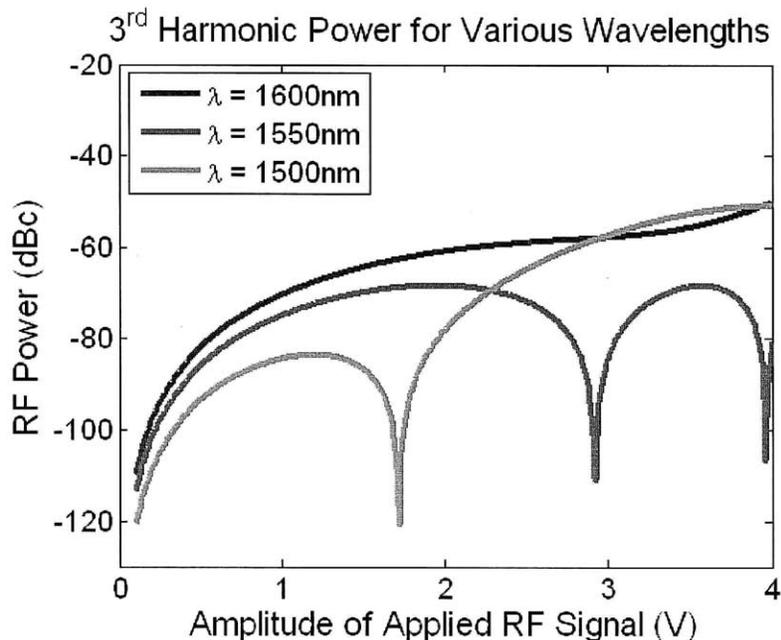


Figure 41: 3rd harmonic power in a 348 μm long MZ modulator biased at 4 V as a function of RF amplitude for several different wavelengths. From [36].

In all, it has been demonstrated that a highly linear response (both second and third harmonics suppressed to below -60dBc) and large modulation depth can be simultaneously obtained over a broad optical bandwidth in a MZ modulator using reverse biased silicon diode phase-shifters [36].

MZ MODULATORS IN THE NATIVE CMOS PROCESS

In previous chapters of this thesis, design work was undertaken for silicon modulators with the stated goal of allowing for integration between silicon photonics and electronics. These designs were made in silicon, and the fabrication steps required to make them would not theoretically hurt the performance of silicon electronics; however, they cannot be fabricated in the industry standard native CMOS process. Due to the huge cost of changing fabrication processes as well as to the importance of electronic device performance (for which the CMOS process is heavily optimized) it is unlikely that the semiconductor industry will be willing to make large changes to this fabrication process. It is therefore necessary that silicon modulators can be fabricated in the actual CMOS process as opposed to simply in compatible processes. Here modulators were designed for fabrication in the IBM 45 nanometer 12SOI CMOS process. They are intended to work at 1200 nm as opposed to 1550 nm.

The full layer stack for this process is shown in Figure 42; a more detailed view of just the device layers is shown in Figure 43. The height of the silicon and poly-silicon layers is fixed in this process: silicon layers are 80 nm tall and poly-silicon layers are 65 nm tall. Additionally, there is only one layer of each. When both the silicon and the poly-silicon layers are present in the same location, there is a 1.16-2.5 nm oxide layer between them. A conformal nitride layer is deposited over the silicon and poly-silicon layers. The oxide layer separating the silicon guide from the silicon substrate is only 145 nm [45]. Because this is thin enough that light will couple from the silicon waveguides to the substrate, it is necessary to remove the substrate from under the photonic devices [44]. However, due to heating effects, the electronics need a substrate capable of conducting heat away from them, which air is not. Because of this, the substrate needs to be removed from under the photonic structures and left in place under the electronic ones. This idea and process, called "localized substrate removal," was developed at MIT and will be used on the fabricated chips to allow for both chip electronics and chip photonics to work [44].

Like the dimensions, the doping concentrations available are constrained. Here, there is the added problem that not all of the doping concentrations are precisely known. Of relevance to modulator design, contact dopings of $1 \times 10^{20} \text{ cm}^{-3}$ of both n-type and p-type are available for use. "Intrinsic" poly-silicon regions doped at (probably much) less than $1 \times 10^{16} \text{ cm}^{-3}$ can be used.

"Intrinsic" silicon regions are also present, but this doping level is less well-known. When designs were completed, it was thought to be p-type and in the range of $1 \times 10^{17} \text{ cm}^{-3}$ to $5 \times 10^{18} \text{ cm}^{-3}$. Since then, it has been learned that this doping is n-type and much lower, most likely between 1×10^{16} and $1 \times 10^{17} \text{ cm}^{-3}$, though the exact value is still unknown [45].

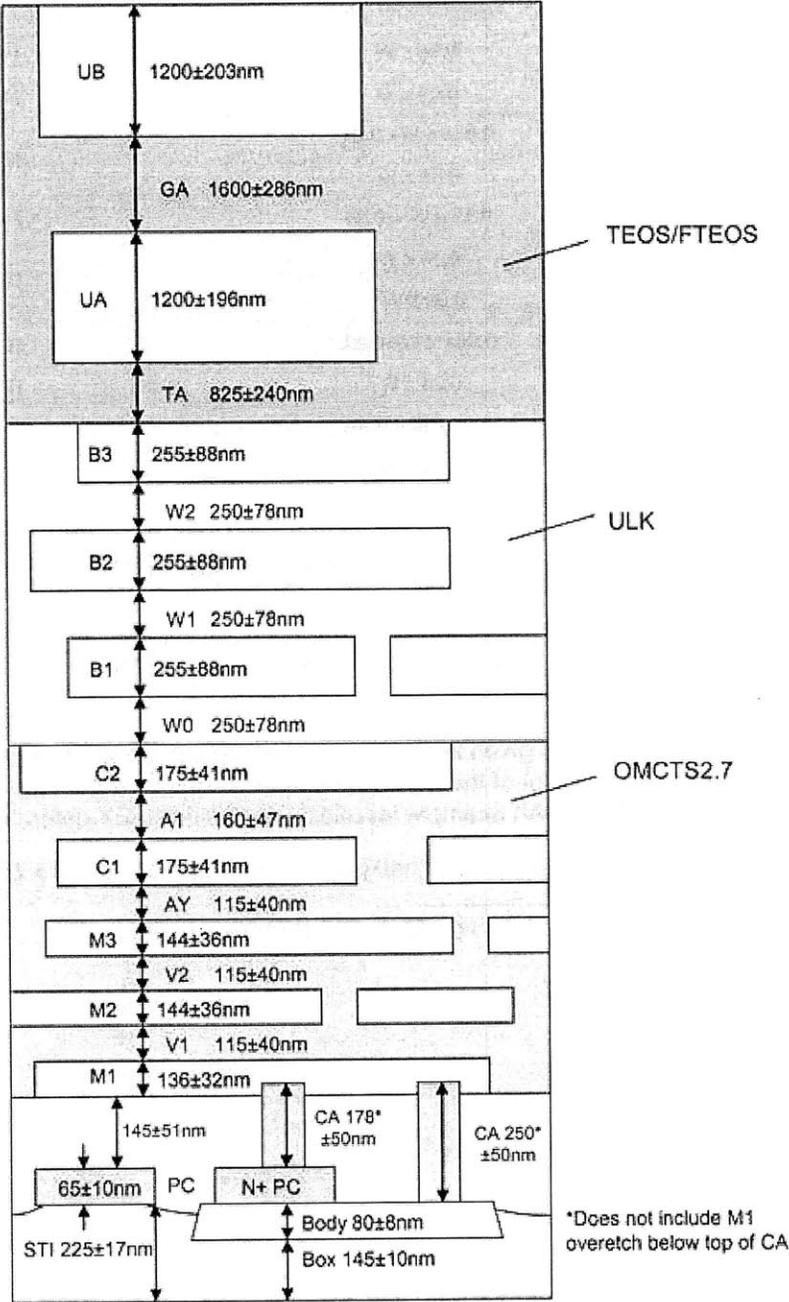


Figure 42: Top level overview of the fabrication stack-up in the IBM 45 nanometer 12SOI CMOS process. Device layers are in the bottom-most box. Metal layers occur above them. [45]

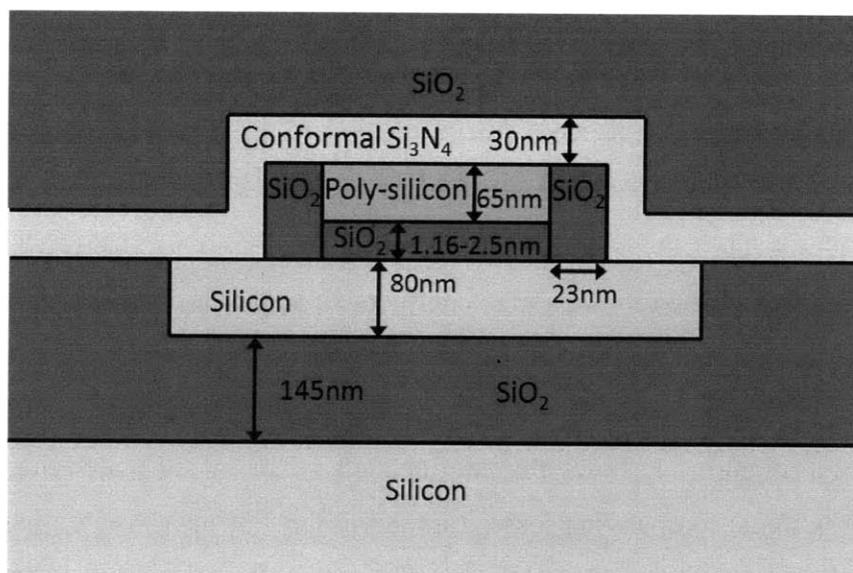


Figure 43: Detail of available silicon and poly-silicon device layers. Note that it is not necessary to use both silicon and poly-silicon; the two layers can be included separately. The bottom silicon layer is locally removed in post-processing.

The poly-silicon layer exists in this process to serve as the gate for MOSFET structures, however it can be used instead as the rib part of a rib waveguide. Importantly, when this is done electrical contact is not made to the poly-silicon and the poly-silicon is not in electrical contact with the silicon. This means that carriers present in the poly-silicon will not be available for use in modulation. Furthermore, measurements of poly-silicon waveguides in other CMOS processes have shown that there is a high modal loss of 50-100 dB/cm associated with the poly-silicon mainly due to surface roughness [45]. Finally, the presence of the oxide spacers on either side of the poly-silicon layer can lead to field enhancement as in the slot waveguide case. This effect does not seem to be large enough hurt modal confinement in the silicon, but it is not a beneficial effect.

Electrically, contact doping levels and intrinsic doping levels are known to exist. Because no intermediate doping levels are present, the electrical structure is constrained to have the $p^+ - i - n^+$ form common in the literature (for the diode case). Because it is not known if the intrinsic doping level is high enough to support reverse bias diode operation, device performance was analyzed for both the forward and reverse bias cases.

There are relatively few free parameters in this system. The width of the rib is one important one. The width of the "intrinsic" silicon region present between the contacts is another. In general, a wider rib provided better confinement to the center of the waveguide. This was

beneficial in that it kept more of the mode out of the lossy contact regions, but harmful in that it pulled more of the mode from the modulated, intrinsic silicon region into the unmodulated polysilicon region. Widening the "intrinsic" silicon region had the benefit of significantly decreasing the loss (by moving the heavily-doped contact layers further away from the mode), but severely decreased the device speed in the case where forward bias operation became necessary.

The rib width was chosen to maximize confinement in the intrinsic region of the underlying silicon layer. As expected, if the rib was too narrow, modulation depth suffered and losses increased. If the rib was too wide, however, then the modulation depth also suffered (though the loss stayed low). Figure 44 shows the confinement factor in the underlying intrinsic silicon region as well as the accompanying losses (assuming the "intrinsic" region has no carriers) for various poly-silicon rib widths. This loss is a lower bound as the intrinsic region will contain some number of carriers. The carriers in the contacts (with a doping level of $1 \times 10^{20} \text{ cm}^{-3}$) and the scattering losses due to the poly-silicon and silicon regions were included in the loss calculation. The poly-silicon was assumed to contribute a 75 dB/cm modal loss: as stated above, this is about what has been measured previously in similar structures [45]. The effect of the thin (1-2 nm) oxide layer on the mode and the modal loss was ignored.

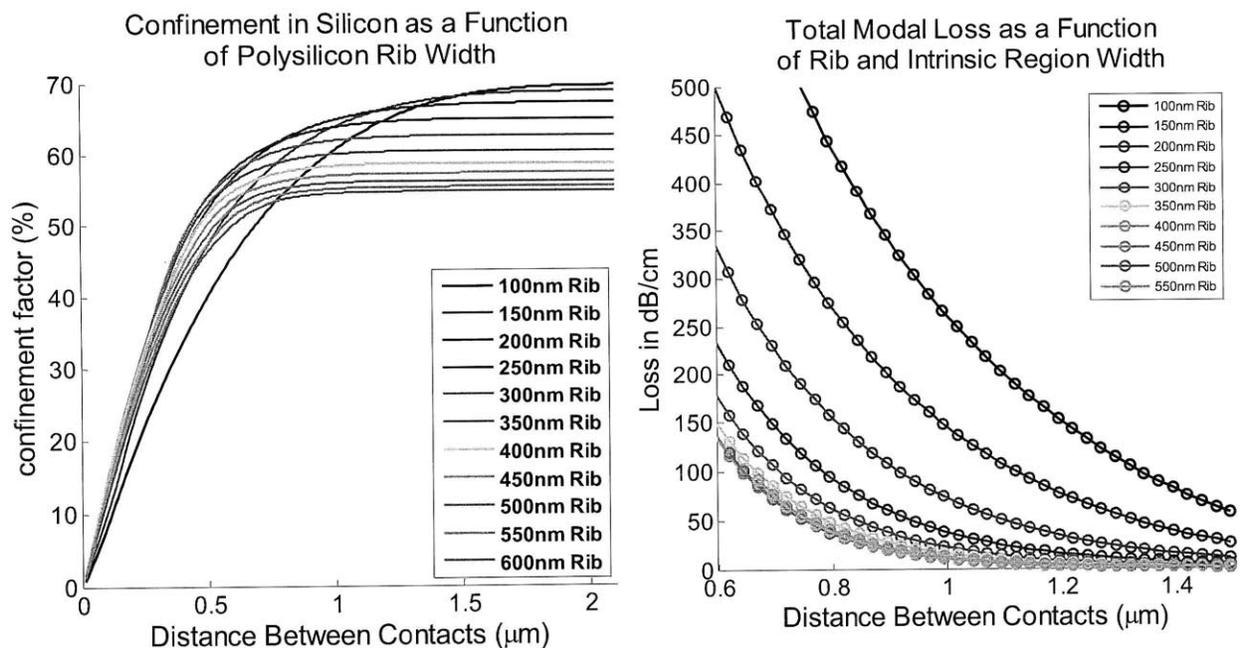


Figure 44: (a) Modal confinement to intrinsic silicon region as a function of intrinsic silicon region width for various rib sizes. (b) Total loss in structure (not including effects from intrinsic region carriers) as a function of same

It was determined from system requirements that no more than 10 dB, and preferably no more than 6 dB, of loss could be accumulated in the modulator. The rib and the intrinsic silicon region need to be wide enough to keep the losses below this level. However, because there is a good chance that these modulators will need to work in the forward bias mode, and because high speed operation involves moving minority carriers across the intrinsic region, the intrinsic region needs to be kept short. From Figure 44, it was estimated that rib widths of around 300-500 nm and intrinsic regions of 600-1000 nm will work best.

To estimate device performance, electrical simulations were run in Sentaurus to calculate modulator sensitivity and speed. Because the "intrinsic" doping level was not known these simulations were run for both $1 \times 10^{17} \text{ cm}^{-3}$ and $1 \times 10^{18} \text{ cm}^{-3}$ p-type doping to find an estimated performance range. From the simulation results, a 400 nm rib width and an 800 nm intrinsic region width were chosen. The effective index change and absorption as a function of voltage for the $1 \times 10^{17} \text{ cm}^{-3}$ and $1 \times 10^{18} \text{ cm}^{-3}$ intrinsic region doping cases are summarized in Figure 45. Note that in reverse bias about half the loss comes from the modal loss in the poly-silicon and half from overlap with the contacts. The intrinsic region doping level ends up having little impact on the loss. In forward bias, the injected carriers eventually dominate the loss through the device.

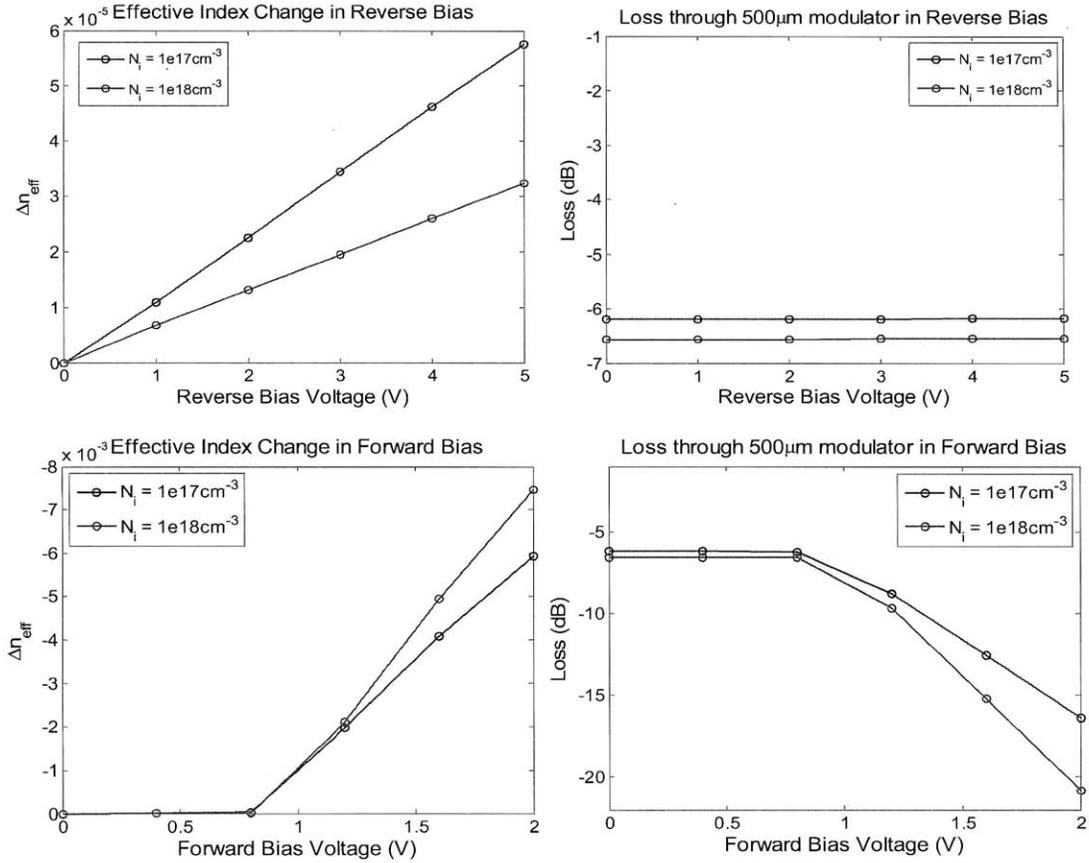


Figure 45: Effective index change and total loss in both forward and reverse bias for proposed diode structure. Two different intrinsic region doping levels are considered, the intrinsic region width was chosen to be 800 nm and the poly-silicon rib width to be 400 nm.

It should be noted that the low modulation achieved in reverse bias in the $1 \times 10^{18} \text{ cm}^{-3}$ doping case comes, in large part, from the poor overlap of the mode with the narrow depletion region near the junction. Offsetting the rib from the center of the guide by some amount to place it closer to the junction side provides a way of increasing achievable modulation depth while minimizing the associated increase in loss. Such a shift also benefits the forward biased results. The phase-change and loss curves for an 800 nm intrinsic region width and a 400 nm rib width at the $1 \times 10^{18} \text{ cm}^{-3}$ doping level are shown in Figure 46 for two different rib offsets. As can be seen, increasing the rib offset helps increase index change considerably.

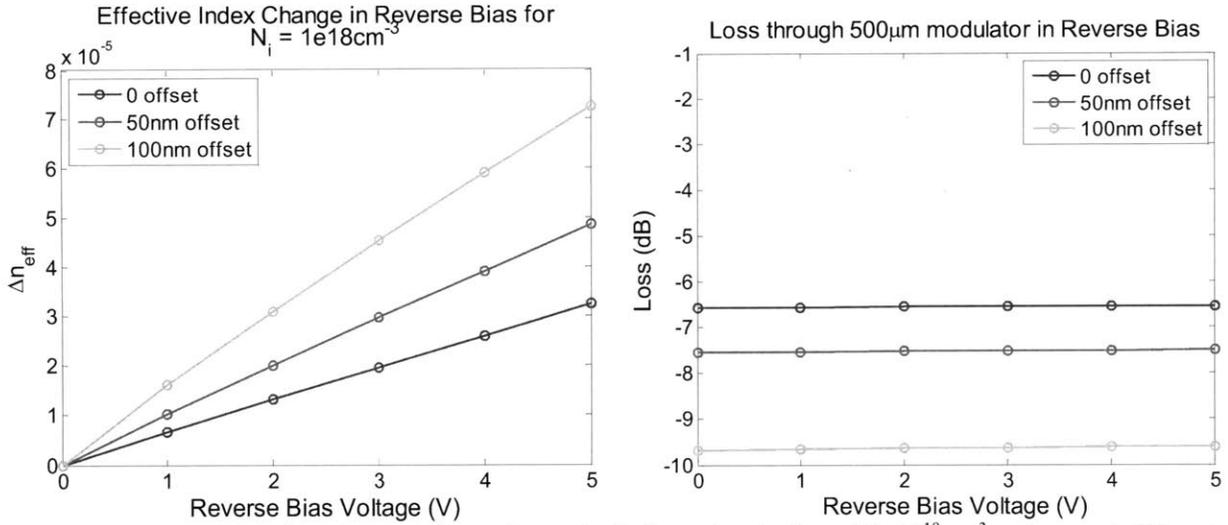


Figure 46: Effective index change and loss for an intrinsic region doping of $1 \times 10^{18} \text{ cm}^{-3}$ and several different rib offsets.

For the $1 \times 10^{17} \text{ cm}^{-3}$ intrinsic region doping, offsetting the rib raised the losses while failing to notably improve the modulation depth. Because of this, a 50 nm rib offset (half way between the desired offset of 100 nm for the $1 \times 10^{18} \text{ cm}^{-3}$ case and that of 0 nm for the $1 \times 10^{17} \text{ cm}^{-3}$ case) was chosen. The cross section is shown in Figure 47. Table 1 outlines the device performance. It should be noted that the multimode nature of the waveguide means that there may be losses beyond those predicted for the desired fundamental mode case. On the other hand, the maximum losses calculated for a forward bias voltage correspond to more than 100% modulation depth. Because the extra modulation is unwanted and unnecessary, there is no need to inject so many carriers and this loss number can be lowered. Finally, it is worth noting that the forward bias current is likely to be limited by some sort of contact or series resistance before it reaches the listed value.

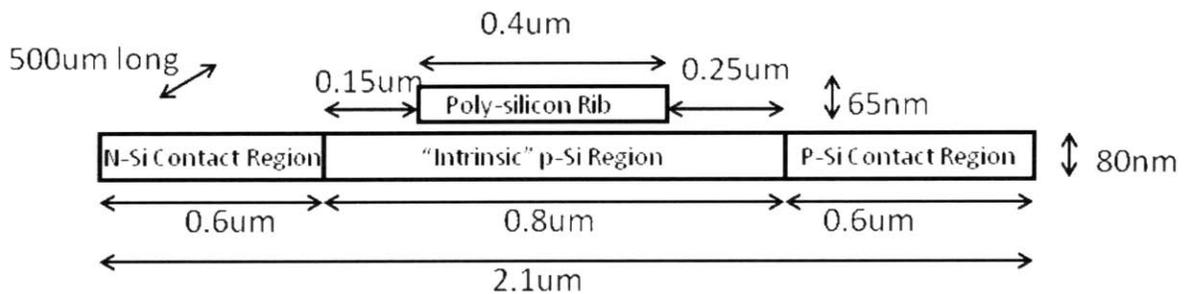


Figure 47: Cross section of designed CMOS silicon diode modulator

"Intrinsic" Doping level	Loss at 0V bias/ (max loss)	Reverse bias phase change – push-pull (rad.)	Reverse bias speed	RB "SNR/ (photon level)"	Forward Bias mod. Depth – single arm	Forward Bias speed	Forward Bias "SNR/ (photon level)"	Is/ I(1V)
1×10^{17}	-7.2 dB/ (-9.7 dB)	.17	>20GHz	.0022	1	1.5GHz	.1	83pA / 40 mA
1×10^{18}	-7.5dB/ (-10.7dB)	.13	>20GHz	.0012	1	1GHz	.085	14nA / 31 mA

Table 1: Predicted performance for modulator design shown in Figure 47.

At the last minute, it was learned that the doping level was n-type and much lower. The design was changed to that for the $1 \times 10^{17} \text{ cm}^{-3}$ p-type silicon doping, shown in Figure 48. The device performance parameters for a $1 \times 10^{17} \text{ cm}^{-3}$ p-type intrinsic doping are shown in Table 2. Because the actual doping is still not known, performance values were not calculated for any of the possible n-type cases.

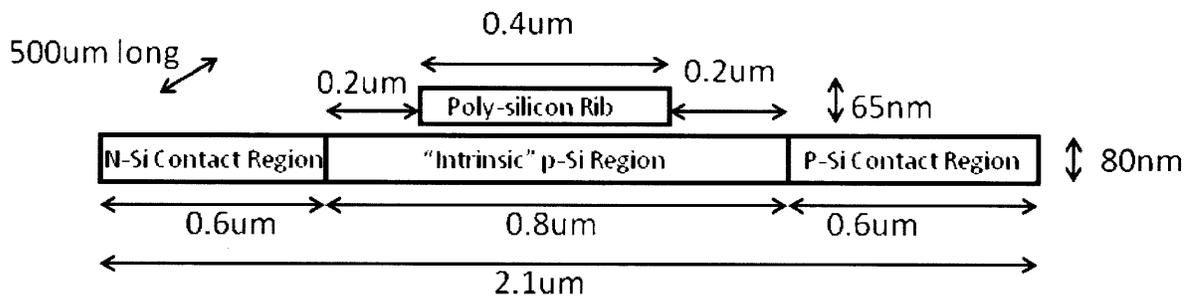


Figure 48: Cross-section of CMOS silicon diode modulator sent to be fabricated

"Intrinsic" Doping level	Loss at 0V bias/(max loss)	Reverse bias phase change – push-pull (rad.)	Reverse bias speed	RB "SNR/ (photon levels)"	Forward Bias mod. Depth – single arm	Forward Bias speed	Forward Bias "SNR/ (photon levels)"	Is/ I(1V)
1×10^{17}	-6.2dB/(-8.8 dB)	.15	>20GHz	.0022	1	1.5GHz	.1	83pA/ 40mA

Table 2: Predicted performance for modulator design shown in Figure 48.

PHOTONIC CRYSTAL MODULATORS

Background

A photonic crystal (PC) is a periodic repetition of a dielectric pattern, and is basically a way to dispersion-engineer a medium. As light travels through a PC it diffracts in such a way that at a given frequency, ω , it can only propagate in certain directions. Careful design of a PC thus allows one to set the dispersion relationship; the relationship between the wavevectors, k , and frequencies ω , that are allowed to propagate thorough the PC [21]. This relationship gives the *band diagram* of that particular PC (dielectric structure), and is calculated by solving the

eigenvalue problem for the magnetic field [21]: $\nabla \times \left(\frac{1}{\varepsilon(\mathbf{r})} \nabla \times \mathbf{H}(\mathbf{r}) \right) = \left(\frac{\omega}{c} \right)^2 \mathbf{H}(\mathbf{r})$ subject to the

constraints $\nabla \cdot \mathbf{H}(\mathbf{r}) = 0$ and $\nabla \cdot [\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r})] = 0$. It should be noted that this eigenvalue problem and constraints are simply an alternative way of writing Maxwell's equations for a location dependent ε and constant μ . To ensure that the solutions contain the translational symmetry of the crystal they are required to be of the form given by Bloch's equation: $\mathbf{H}_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k} \cdot \mathbf{r}} \mathbf{u}_{\mathbf{k}}(\mathbf{r})$, where $\mathbf{u}_{\mathbf{k}}(\mathbf{r})$ is a function with the periodicity of the lattice [21]. Because the crystal is periodic, k vectors that differ by an integer number of periods will propagate at the same frequencies. Therefore, one need only to look at one period in k -space to see what is happening in the crystal; the values must repeat further on. In general, we look at the period centered around $k = 0$, which is called the *Brillouin Zone* [21]. We also assume that non-linear optical effects can be ignored, that the dielectric materials are isotropic, that the structure is macroscopic, that the material dielectric constants do not vary with optical wavelength, and that $\varepsilon(\mathbf{r})$ is real and positive to within small perturbations [21].

Photonic crystals are analogous to semiconductor crystals in many ways. Most importantly, just as there are energy levels forbidden to electrons in semiconductor crystals, there can be energy levels, corresponding to frequencies, forbidden to photons in photonic crystals [21]. When such forbidden frequencies exist, there is said to be a *bandgap*, and the frequencies at which light may not propagate are said to be *in the bandgap*. In general, the analogy to semiconductor crystals works very well with one exception. In semiconductors there is a uniquely determined length scale set by the radius of the atoms. Under Maxwell's relations there is no fundamental length scale, and a structure can simply be rescaled (in absolute dimensions) to work with different parts of the electromagnetic spectrum [21]. Because of this, for a given

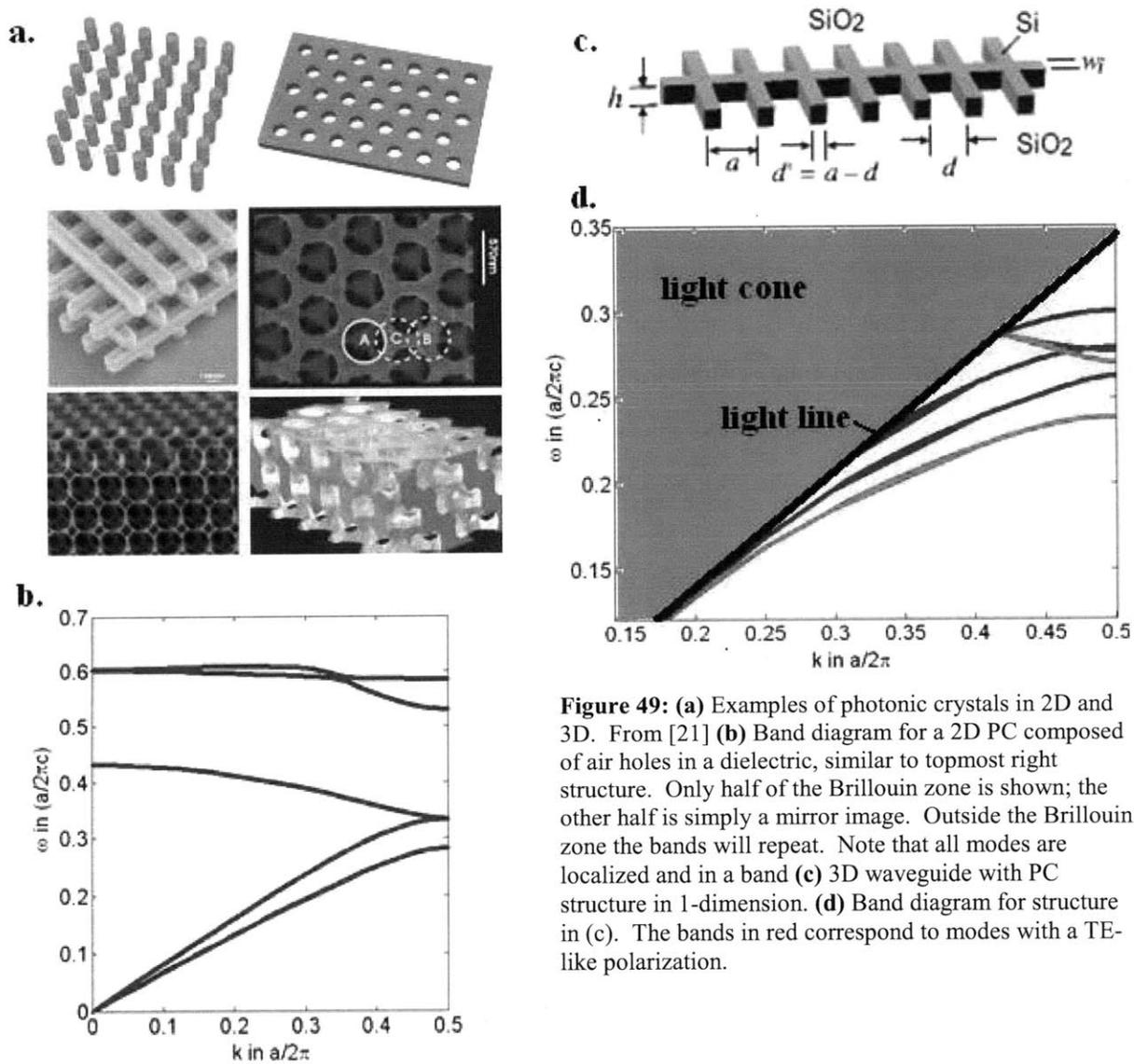


Figure 49: (a) Examples of photonic crystals in 2D and 3D. From [21] (b) Band diagram for a 2D PC composed of air holes in a dielectric, similar to topmost right structure. Only half of the Brillouin zone is shown; the other half is simply a mirror image. Outside the Brillouin zone the bands will repeat. Note that all modes are localized and in a band (c) 3D waveguide with PC structure in 1-dimension. (d) Band diagram for structure in (c). The bands in red correspond to modes with a TE-like polarization.

geometry, the absolute size of the bandgap may vary based on the size of the device, but the relative size of the bandgap (that is the absolute size divided by the frequency located in the middle of the gap) will not. This ratio, given in percent and called the *gap-midgap ratio*, is often used to characterize bandgap size [21].

Much, but not all, of the relevant information about a photonic structure is present in the band diagram (see Figure 49b-d). In 1D, 2D, or 3D photonic crystals the band diagram is simply made up of a set of lines showing allowed ω and k (as in figure 9b). However, for a structure like that shown in Figure 49c, where the photonic crystal only extends in one or two directions, but the actual structure exists in 3D, the situation changes. Specifically, these structures rely on index guiding to confine the mode in one or more directions. There then exists a *light cone* in which index guiding fails and extended states propagate through both the core and cladding. The bottom of this *light cone*, is called the *light line*. This line will intersect the edge of the Brillouin zone. Because the dispersion relation must repeat, the light line defines the maximum frequency that can be guided at a given k . This means that there are a finite number of guided modes in such a photonic crystal waveguide regardless of operation frequency (in contrast, a regular rectangular dielectric waveguide allows for the propagation of an infinite number of modes across the frequency spectrum) [21].

Looking at the band diagram, one can see that the group velocity of the propagating waves is also affected by the photonic crystal structure. In general $v_g = \partial\omega/\partial k$. Thus, a photonic crystal has the ability to slow the light propagating through it; that is, to create slow wave structures. Experimentally, structures have been demonstrated that slow light by more than a factor of 1000 [29]. We note that, because the bands must bend at the Brillouin zone edge in order to repeat, the group velocity in this area of the band diagram will always be very low, approaching zero as one picks operation frequencies closer and closer to the band edge [21]. Usually, a structure needs to operate over a range of frequencies; thus, a given photonic crystal device is designed to operate a distance from the Brillouin zone edge that corresponds to at least the desired bandwidth. In practice, one also needs to choose the operating point such that there are no other k -states in which light at the desired frequencies can propagate [21]. To create a slow light structure one then seeks flat areas near the Brillouin zone edge that correspond to the edge of a band. In practice, for 3D structures that contain a photonic crystal along one dimension (like

gratings in a photonic wire configuration,) these are more or less impossible to find. Fortunately, bandgaps can exist for given symmetries or polarizations. If light is polarized before being inserted into the waveguide, and the guide contains relatively few defects, coupling into the other polarization state will be relatively low, and bands corresponding to that polarization can be ignored [21].

Many device designs, particularly modulator designs, also need to account for dispersion.

Dispersion is generally given by: $D = \frac{d}{d\lambda} \left(\frac{q}{v_g} \right) = -\frac{\lambda}{c} \frac{\partial^2 n}{\partial \lambda^2}$. If the dispersion is too great, the frequencies placed into the device will walk off of each other and the pulse shape will degrade. One way to avoid problems caused by dispersion is to employ a dispersion compensating structure; that is a second waveguide with opposite dispersion that essentially moves the frequencies back into place [30]. In a PC this can be accomplished by changing the crystal dimensions or period such that the bands change shape. In particular, in the structure pictured in figure 9c, it is not uncommon for the first and second TE bands to have opposite curvature (figure 9d). One can then pass the light through a PC where its frequency travels in the second band after passing it through a PC where it travels in the first, thus compensating the original dispersion [30]. It is important to note that since light cannot propagate in the band gap, it is impossible to transition directly from the first structure to the second. One must instead transition back to a regular waveguide before transitioning to the second PC guide [30].

It is also important to note that such a structure will not solve all the problems caused by dispersion in our application. In particular, this structure would only fix the dispersion of the optical pulse. The walk off of the various frequencies in the first PC waveguide would still damage the precise mapping between frequency and time that we rely on to properly encode the electrical signal, causing the wrong frequency to pick up part of the electrical information at each point in the modulator. This mistake would not be fixed by afterward moving the incorrectly modulated pieces back into their correct temporal locations. Because of this, it is important that the dispersion be minimal. In general, however, we do not expect the time-frequency errors to be problematic.

Finally, it is important to note that photonic crystals can be difficult to fabricate, and fabrication errors may have a large effect on device performance. Therefore, it is important to examine the robustness of a PC design before sending it to be fabricated. In particular, if the

theoretical bandgap has a gap-midgap ratio below about 5%, there is a chance that there will not be any bandgap in the fabricated device [21].

Use as Slow Wave Structures and Slow Wave Modulators

Slow wave structures have been used in time delay links, in modulators, and in other devices to increase sensitivity and decrease device length. Because light travels through a slow wave structure very slowly, the interaction time in a modulator between the light and the index perturbation is larger, resulting in a larger overall phase shift for a given index perturbation at a given length. There are several important results in the literature on both slow wave structures and slow wave modulators. The slow wave modulators found so far in the literature all use a MZ configuration.

Most slow wave structures that have been reported, and all slow wave modulators that the author found, use a rectangular or hexagonal pattern of air holes in a dielectric (usually silicon) as the basis for the photonic crystal structure. In these situations, a waveguide is created by removing one (or more) of the columns of holes to create a defect state through which light of the desired wavelength can propagate. In [3], a slow light structure was used to enhance non-linear interactions in silicon, and achieved a group velocity of $c/40$ (which would mean an increase in sensitivity of a factor of 10 to 20, depending on geometry, over a plain, rectangular silicon waveguide of equal length). In this structure, in addition to removing rows of holes to create the waveguiding region, the rows of holes immediately adjacent to the waveguide were displaced in order to enhance the slow light effect. In [29] the group velocity and dispersion of a 2D SOI photonic crystal waveguide are characterized and a group velocity of less than $c/1000$ is found.

In PC modulators, a 2D photonic crystal waveguide of this type is included in the phase-shifter sections of the Mach-Zehnder arms to create a slow light interaction. In [17], waveguides with both a hexagonal and a square grid of holes were used to create a low-power, compact optical switch in an AlGaAs crystal. The switching was accomplished by injection with free carriers excited by an external pumping source. Best results were achieved with the square pattern of holes. More comprehensive work on PC modulators has been done in silicon in [22]. Here a hexagonal PC waveguide section is added to the arms of a MZ modulator, and a p-i-n diode configuration is used to create a phase-shift through the plasma dispersion effect. The

slow light enhancement allows for a π phase-shift to be achieved in 80 μm at 2 V, yielding a $V_{\pi}L$ of .016 Vcm. We note that this sensitivity is not better than results reported for regular, forward biased diode modulators [5]. This lack of improvement probably has to do with the large width of the intrinsic region used – it was 4 μm as opposed to the more common 0.5 μm . Like most other forward biased modulators, this modulator operates with a speed around 1-2 GHz.

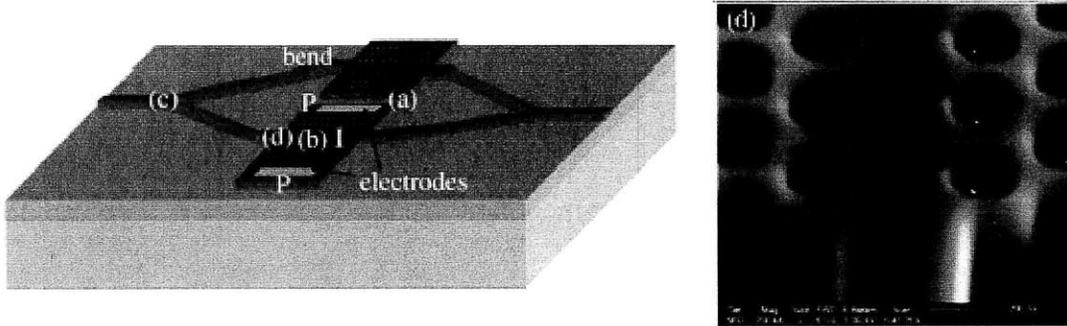


Figure 50: A basic PC waveguide MZ modulator structure similar to that implemented in the papers described above. From [27].

Neither of these papers mentions the insertion loss or dispersion of the device. However, these are important parameters to consider. The importance of mode matching when entering a PC should be obvious: large mismatch leads to high optical reflection—in some cases there is as much as a 30 dB drop in transmitted power, though 6 dB is a more reasonable figure. Tapering into the structure can significantly minimize loss, to approximately 1 dB [26] [18] [30]. However, tapering into a PC waveguide made of holes can be difficult as holes cannot be fabricated below a certain diameter. Thus, structures which decrease more gracefully are of use. Furthermore, photonic crystal waveguides of this nature usually need to be patterned with electron-beam lithography, a slow, expensive process that is generally not compatible with mass production.

In [30], a different type of photonic crystal was used to create a slow light region for use in tunable time delays. In this case, the PC was made up of a series of flanges attached to a normal, rectangular waveguide (see Figure 51). In this structure reductions in group velocity were considerably smaller – light was in the range of 2-4 times slower. The device was operated 4% from the band edge; thus these numbers could be improved somewhat, though not dramatically. However, this structure does retain some advantages over the structures above. First, this design can easily be tapered into and out of; there is no problem of minimum fabrication feature size. Because of this, reflections upon entering the device can be less than 0.1% for a 10 period taper

and less than 0.001% for a 50 period taper. A dispersion compensation section is also successfully included in this paper, and seems to work over a broad bandwidth. Furthermore, such a structure could be fabricated using conventional high-throughput lithography. It is therefore desired that a modulator be integrated into this photonic crystal design.

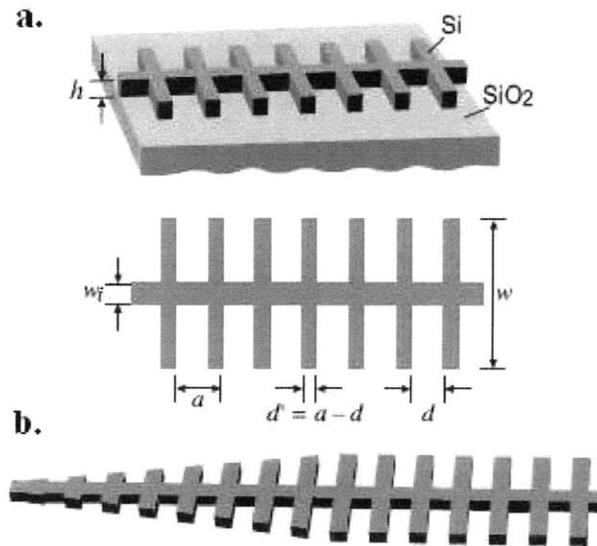


Figure 51: (a) Flange structure implemented in [Povinelli] (b) Example of flange structure tapering. From [Povinelli]

Slow-Wave Modulator Design

Achievable Slowdown

The PC structure in [30] is optimized to achieve the maximum slowdown possible for an oxide substrate and an air cladding. It is also chosen so that a dispersion compensating scheme is possible. Moreover, it is not electrically contacted. To start modifying this design for incorporation into a modulator, it is desired that the maximum achievable slowdown be found. In our system, both the substrate and the cladding are oxide. Horizontal features are constrained to be at least 100nm in width, and vertical ones can be no more than a few hundred nanometers tall. Simulations were run using the MIT Photonic Bands [34] software produced by Steven Johnson's group to maximize the achievable slowdown. There were two types of modes available that corresponded roughly to two different modal symmetries: the first kind of modes had an electric field with odd symmetry with respect to a plane down the center of the device, while the second kind had even symmetry. Two of the modes with odd symmetry were well localized to the center of the structure, while only one of the modes with even symmetry possessed this

property, which is necessary to prevent modes from interacting significantly with the contacts. For this reason, the odd-symmetry modes were chosen. Structures with bandgaps smaller than 5% were ignored (see above.)

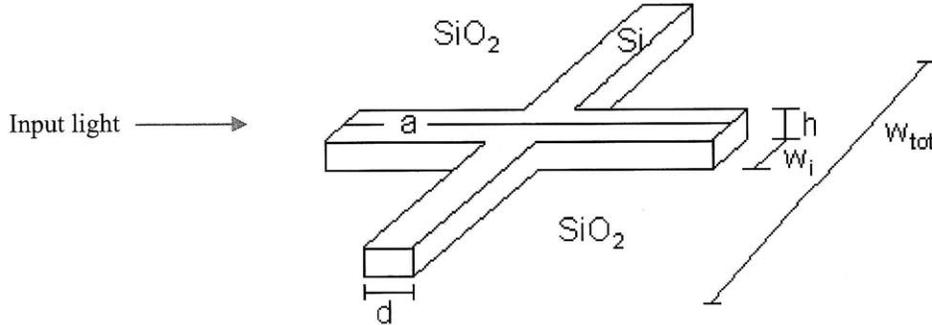


Figure 52: Diagram of PC unit cell consisting of a silicon structure surrounded by oxide cladding. The period is given by a ; other relevant dimensions are marked.

A diagram of the PC unit cell is shown in Figure 52. The minimum speed found in this structure was $0.075c$, or about 4 times slower than the speed of light in silicon. This speed was obtained for a structure with $a = 340$ nm, $h = 272$ nm, $w_i = 170$ nm, $w_{tot} = 1.02$ μm , and $d = 100$ nm. The structure had a 10% bandgap. Structures with similar dimensions yielded similar results. Perhaps not unexpectedly, the optimized dimensions for this case are only slightly different than those for the situation in [30].

It is generally assumed that larger bandgaps correspond to larger slowdowns, but this is not always the case. Figure 53 illustrates the reasoning behind the standard assumption and its breakdown in certain cases. In Figure 53a, there are only two bands in the guided region, and the bandgap is large. As a result, the slopes of both bands are small, so the propagation velocity $d\omega/dk$ is small. In Figure 53b, the same two bands are pictured in the guided region, but with a smaller bandgap. As a result, the slopes are now larger, leading to faster propagation. Figure 53c demonstrates how the appearance of a third band in the guided region can lead to a breakdown in the connection between small bandgaps and slow propagation. The first two bands are the same as in Figure 53a, so their propagation velocities are still small. However, the presence of the third band in the guided region has caused a decrease in the bandgap with no associated increase in velocity. Figure 53d shows this phenomenon carried to its extreme. In this case, mode 3 appears in such a way as to completely erase the bandgap, rendering the device useless.

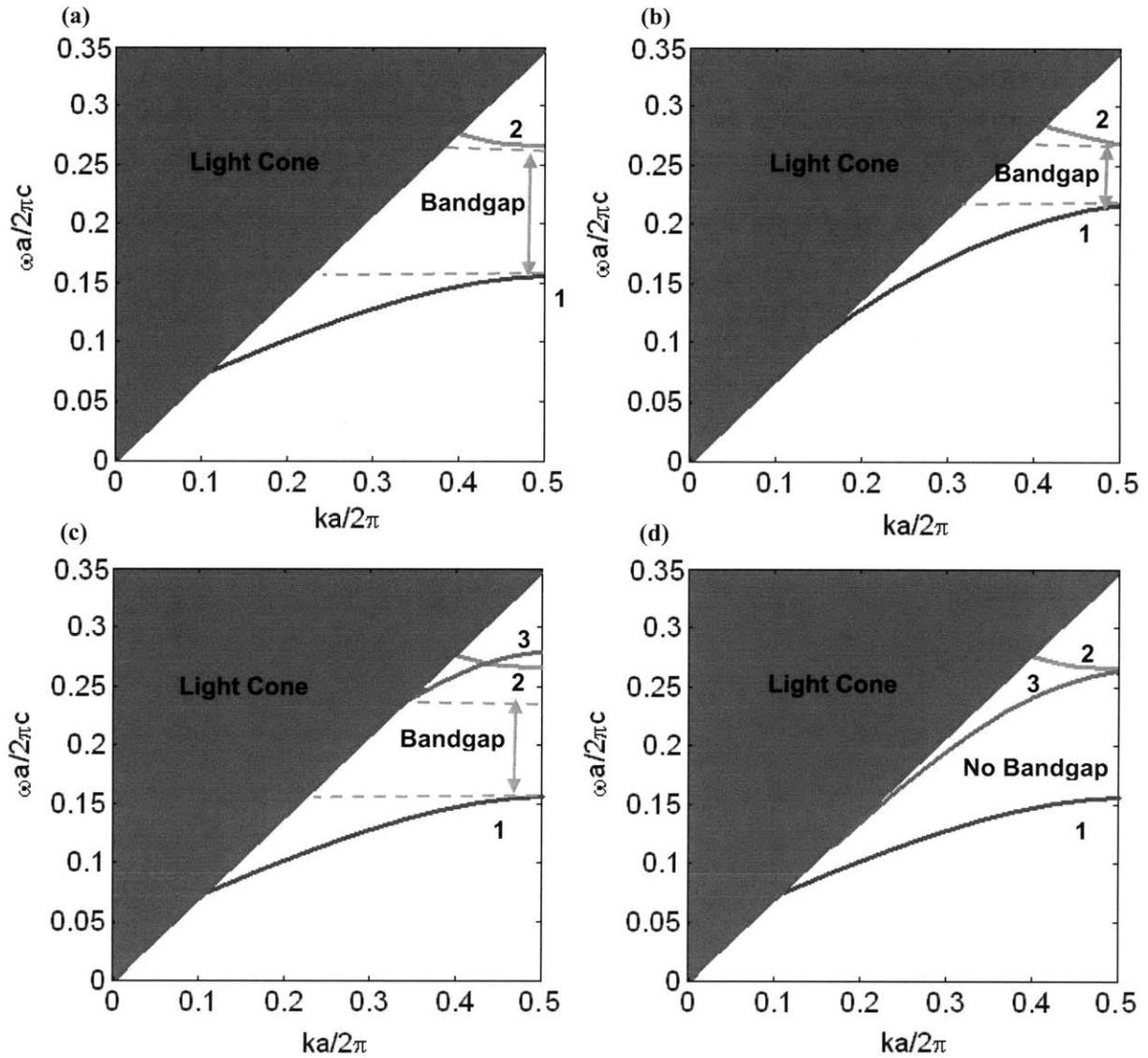


Figure 53: Illustration of the relationship between group velocity and bandgap size.

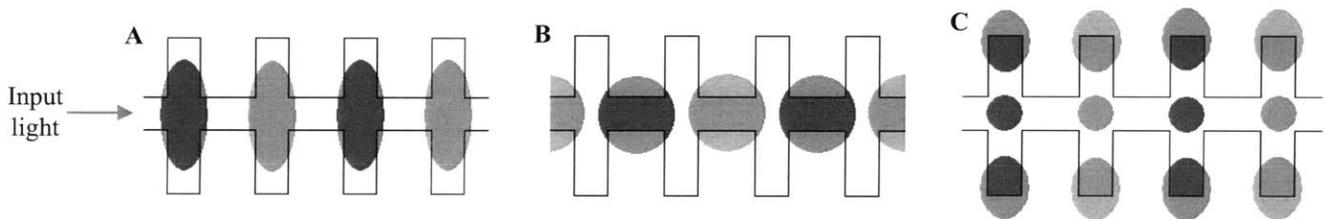


Figure 54: Diagram of the time-averaged electric-field energy density of the modes corresponding to the Brillouin zone edge of the lowest three bands (in most variations of this structure). The leftmost mode corresponds to the lowest band in nearly all cases. The middle and right most modes correspond to the second and third bands, but which is which varies by structure.

Several trends could be ascertained from the output of the simulations. In general, a mode more tightly confined to the dielectric will have a lower frequency than a mode carried substantially in the cladding. Usually, the greater the difference in frequency between the modes, the more the bands must bend and the larger the slowdown achieved. The trends observed in the simulation sweeps show this. In most simulations mode A corresponded to the first band and mode B corresponded to the second band (see Figure 54). It was noted that the speed of light through the structure tended to decrease as the ratio of the waveguide height, h , to width, w_i , increased. The increase in height caused an increase in the fraction of the modes carried in the dielectric and therefore a decrease in frequency. However, this effect was more pronounced in mode A than in mode B, resulting in a wider bandgap and therefore a lower speed. This trend is outlined in Figure 55. The achieved slowdown is also increased as d is decreased in relation to a , up to a point. Here, the decrease in rib width initially increases the frequency of both modes, but the effect is more pronounced in mode B (which is kicked substantially into the cladding) than in mode A (which remains localized to the guide.) However, as d continues to shrink, more of mode A is moved into the cladding, and its frequency starts to increase at a faster rate than that of mode B (which is already dominated by the cladding nearly as much as it will be.) At this point, the speed starts to increase again. Increasing w_{tot} in relation to w_i or a lowers the frequency of mode A and has little effect on the frequency of mode B. Therefore, increasing this width also decreases the light's speed. However, a large w_{tot} pulls mode C (far left in Figure 54) down from the air band, destroying the bandgap and ruining device performance. Because of this, these fingers need to remain short. Finally, as w_i increases more of mode B is carried in the dielectric, lowering its frequency toward that of mode A, while little impact is had on the frequency of mode A. Because of this, light travels faster at larger values of w_i . All of these trends are outlined in Figure 55.

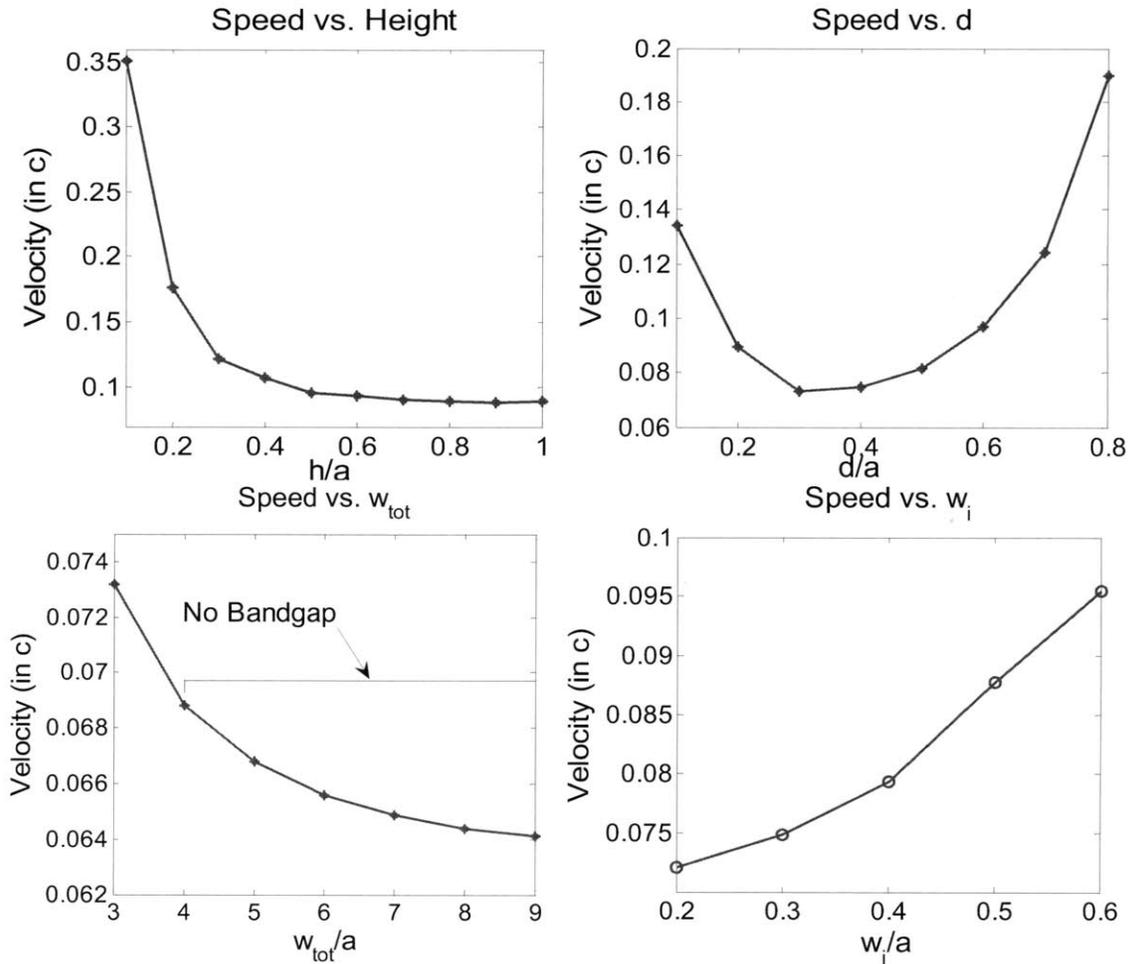


Figure 55: Trends in speed of light in the structure as a function of various dimensions. Note that not all data points correspond to structures that can be fabricated or that have a useable bandgap. In particular, large relative values of w_{tot} do not yield bandgaps.

Contacting Issues

The structure found in the previous section still needs to be electrically contacted; being able to drive the modulator structure electrically is vital for integration. The structure, thus, needs to be modified in such a way that some part of it that does *not* contain a large amount of the fields extends to a contact region. Contacts are usually located a micron or more away from the center of the device. In most structures, the main motivation for keeping contacts so far out is to minimize loss. However, in PC structures it is also desired that contacts be far enough from the mode that they do not affect the periodicity of the structure.

Finding an appropriate contacting scheme for this structure has proven to be problematic. In particular, extra silicon must be added to the edges of the structure to make contact. This extra silicon lowers the frequency of mode C (or its equivalent) in Figure 54, pulling it from the air band to lie across the bandgap. The bandgap is then destroyed, and the structure no longer functions as a slow wave structure. To be more precise, Figure 56 shows the actual distribution of the time-averaged energy-density in the electric fields of the first three bands present in this

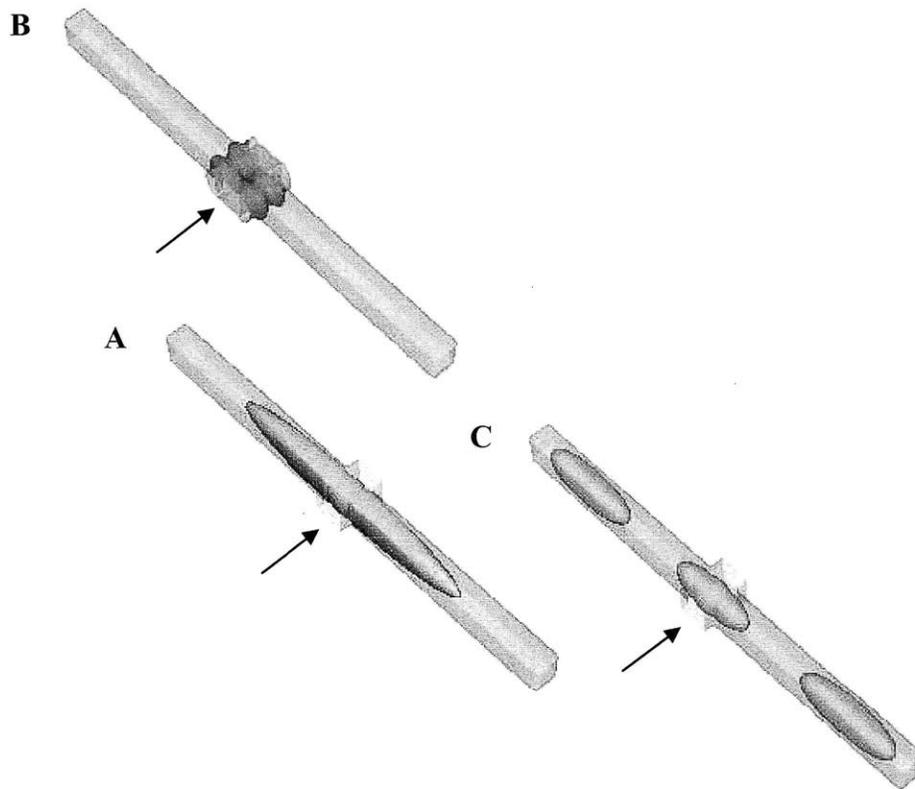


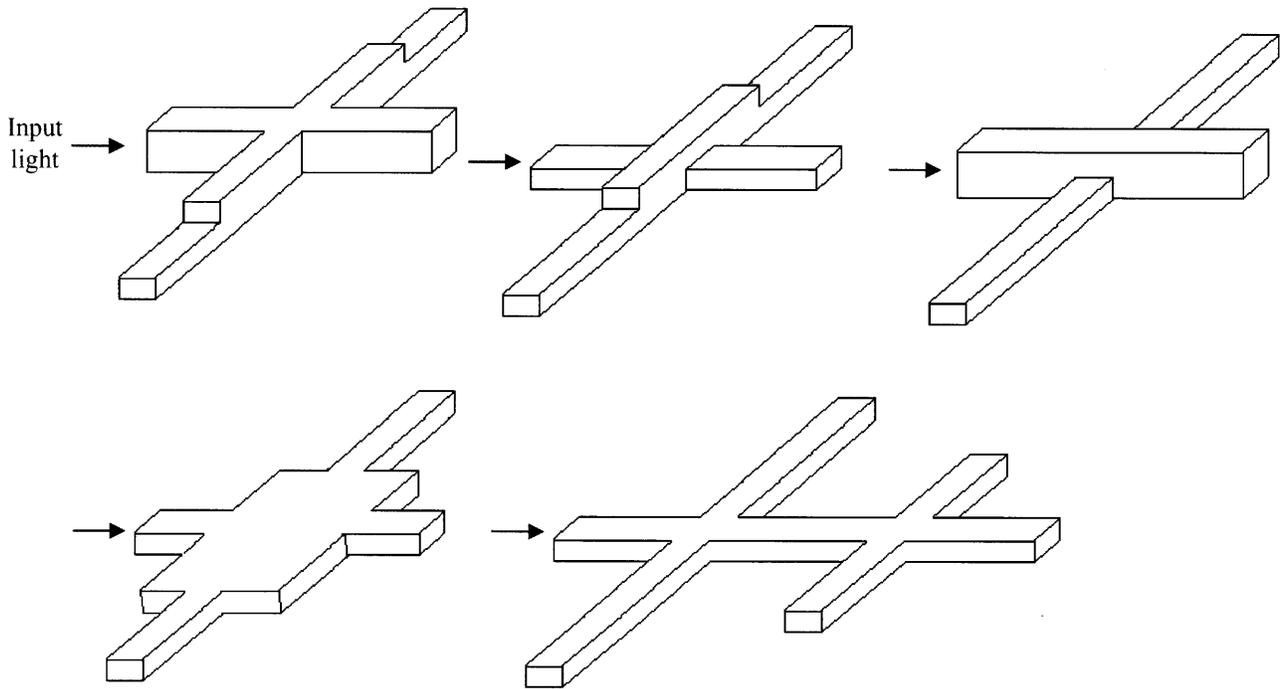
Figure 56: Power in electric field in the three lowest frequency modes at the Brillion Zone edge ($k = \pi/a$).

fingered structure. When there is little enough dielectric material away from the center of the guide, the first band is A and the second is B, both of which are nicely localized to the center of the guide. Band C is located in the light cone for most of the Brillion zone, and is confined to the guide only near the edges. However, as more material is added, mode C becomes the second band. This is problematic for two reasons. First, band C tends to intersect with band B, destroying the bandgap. Secondly, even when a bandgap is present, band C has the same general shape as band A but at higher frequencies. This prevents mode A from being used, and mode C

cannot be used for propagation instead because it has a lot of light out near the contacting regions. Moreover, shrinking d or the height of the fingered regions until mode C is no longer supported at frequencies near those of A and B yields fingers that are too small to fabricate or to use for the necessary amounts of carrier injection or extraction.

The structure was varied in several ways to attempt to circumvent this problem. Rigorous simulations were not run for each of these structures. Instead, several different parameter combinations were run to see if the new form was likely to help, and, if promising results were not found, the structure was abandoned. The various structures tried are diagrammed in Figure 57. The first set of structures were aimed at keeping band B below band C. Unfortunately, while most of these structures succeeded in lowering band B, band C stayed in the same place (because the contacting arms needed to be maintained at a minimum thickness and width), crossing whatever bandgap would otherwise have been present. Moreover, these structures generally also dropped the frequency of band B in relation to that of band A, often destroying the bandgap anyway (and raising the speed of light through the device significantly, even when a bandgap remained). Another approach was to drop both bands A and B below band C, and use the gap between band C and the next highest mode (band D) to carry light; this was not tenable as both of those modes carried a significant amount of light in the contact region. The next idea was to try to drop band C below bands A and B leaving bands A and B as the band edges. However, doing so usually pulled other, additional modes similar to C down into the gap, even when the order of A and C was successfully changed. In all, a bandgap of usable size (greater than 5%) was not found in any of these structures.

Structures Attempting to Keep Band (b) below band (c)



Structure Attempting to Lower Band (c) Below Band (a)

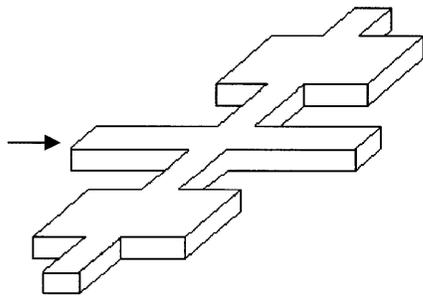


Figure 57: Structure variations considered to allow for electrical contact to be made to the device.

If work is to continue on this project, more rigorous simulations should be run to make sure that none of the structures considered work for parameter combinations not tested. Some more rigorous proof that contacts cannot be successfully added is desired. Once this is obtained, it will probably be necessary to move to a different PC structure entirely.

CONCLUSIONS AND FUTURE WORK

This thesis investigated the design of various electrical and optical schemes used in modulators. In rib waveguides, a silicon diode modulator was designed and tested. It operated at speeds up to 13 GHz and had a $V_{\pi}L$ of 1.2 Vcm. MOS capacitor modulators were investigated as an alternative but were found not to offer substantial improvements, yielding slightly higher sensitivities at slightly lower speeds. In addition to the standard rib waveguide structure, modulators were designed for fabrication in an actual CMOS process. If these modulators work as desired, they will be a big step forward towards on-chip integration of electronics and optics. Photonic crystal structures were also investigated. While the structure explored here ultimately proved unfeasible, slow light structures in general promise to lower footprint size and power requirements. Therefore, other photonic crystal structures should be explored in the future. Meanwhile, since linear modulator performance is important in analog applications, a linearization scheme for MZ modulators was proposed, and simulations demonstrated that it should be effective for reverse biased silicon diode modulators.

To further improve modulator performance (for use in a given general waveguide geometry,) more complicated electronic structures, such as transistors or diode variants such as $p-n^+-n-n^+$ structures, need to be explored. Outside of the plasma dispersion effect, recent work from other groups on using the QCSE in SiGe structures looks promising.

Future work in integrating modulator designs into a real CMOS process must start with obtaining a better knowledge of the foundry process. Tests need to be run to determine the actual doping levels of the various layers (and the layers themselves need to be further explored so there is better knowledge of which can and cannot be successfully used) as well as the losses inherent in the process. A successful way to make electrical contact to the modulators, other than the fake rib-structure, would probably also be helpful as avoiding the use of the high loss polysilicon layer will be beneficial to performance.

Finally, experimental verification of the linearization scheme proposed in this thesis would be valuable. Future theoretical work needs to include a more detailed and practical analysis of what happens if the two phase-shifter sections are unbalanced and, in particular, how much imbalance can be tolerated. The ability to linearize different modulator structures should also be explored.

APPENDICES

A: Select Parameters Used In Electrical Simulations

```
Material = "Silicon" {
Bandgap
{ *Eg = Eg0 + dEg0 + alpha Tpar^2 / (beta + Tpar) - alpha T^2 / (beta + T)
  * dEg0(<bgn_model_name>) is a band gap correction term. It is used together with
  * an appropriate BGN model, if this BGN model is chosen in Physics section
  * Parameter 'Tpar' specifies the value of lattice
  * temperature, at which parameters below are defined
  * Chi0 is electron affinity.
    Chi0 = 4.05 # [eV]
    Bgn2Chi = 0.5 # [1]
    Eg0 = 1.16964 # [eV]
    dEg0(Bennett) = 0.0000e+00 # [eV]
    dEg0(Slotboom) = -4.7950e-03 # [eV]
    dEg0(OldSlotboom) = -1.5950e-02 # [eV]
    dEg0(delAlamo) = -1.4070e-02 # [eV]
    alpha = 4.7300e-04 # [eV K^-1]
    beta = 6.3600e+02 # [K]
    Tpar = 0.0000e+00 # [K]
}

OldSlotboom
{ * deltaEg = dEg0 + Ebgn ( ln(N/Nref) + [ (ln(N/Nref))^2 + 0.5]^1/2 )
  * dEg0 is defined in BandGap section
    Ebgn = 9.0000e-03 # [eV]
    Nref = 1.0000e+17 # [cm^(-3)]
}

eDOSMass
{ *Documentation says to use formula 1 in simulations of silicon. Default parameters
are supposed to be correct. ?? Check this.
  * For effective mass specificatition Formula1 (me approximation):
  * or Formula2 (Nc300) can be used :
    Formula = 1 # [1]
  * Formula1:
  * me/m0 = [ (6 * mt)^2 * ml ]^(1/3) + mm
  * mt = a[Eg(0)/Eg(T)]
  * Nc(T) = 2(2pi*kB/h_Planck^2*me*T)^3/2 = 2.540e19 ((me/m0)*(T/300))^3/2
    a = 0.1905 # [1]
    ml = 0.9163 # [1]
    mm = 0.0000e+00 # [1]
}

hDOSMass
{ * Documentation says to use formula 1 in simulations of silicon. Default parameters
are supposed to be correct. ??Check this.
  * For effective mass specificatition Formula1 (mh approximation):
  * or Formula2 (Nv300) can be used :
    Formula = 1 # [1]
  * Formula1:
  * mh = m0*{[(a+bT+cT^2+dT^3+eT^4)/(1+fT+gT^2+hT^3+iT^4)]^(2/3) + mm}
  * Nv(T) = 2(2pi*kB/h_Planck^2*mh*T)^3/2 = 2.540e19 ((mh/m0)*(T/300))^3/2
    a = 0.443587 # [1]
    b = 0.003609528 # [K^-1]
    c = 0.0001173515 # [K^-2]
    d = 1.263218e-06 # [K^-3]
    e = 3.025581e-09 # [K^-4]
```

```

f      = 0.004683382 # [K^-1]
g      = 0.0002286895 # [K^-2]
h      = 7.469271e-07 # [K^-3]
i      = 1.727481e-09 # [K^-4]
mm     = 0 # [1]
}

ConstantMobility:
{ * Constant mobility, where limit is from phonon scattering only.
  * mu_const = mumax (T/T0)^(-Exponent)
    mumax = 1.4170e+03 , 4.7050e+02 # [cm^2/(Vs)]
    Exponent = 2.5 , 2.2 # [1]
    mutunnel = 0.05 , 0.05 # [cm^2/(Vs)]
}

DopingDependence:
{ * Dopant dependent mobility scattering, three models total, though one of them, the
  University of Bologna model takes a different key word in the physics section
  * (UniBoDopingDependence instead of Doping Dependence), can from either the Massetti
  model (??better for silicon) or the Arora model (??Better for GaAs) here. ??Check
  values, check that using the right option. To see how much difference mobility model
  makes, run both. If similar just use Massetti. If not, still use Masetti, but look at
  this in a lot more detail and be aware that mobility has a big impact on results.
  * For doping dependent mobility model three formulas
  * can be used. Formula1 is based on Masetti et al. approximation.
  * Formula2 uses approximation, suggested by Arora.
    formula = 1 , 1 # [1]
  * If formula=1, model suggested by Masetti et al. is used:
  * mu_dop = mumin1 exp(-Pc/N) + (mu_const - mumin2)/(1+(N/Cr)^alpha)
  * - mul/(1+(Cs/N)^beta)
  * with mu_const from ConstantMobility
    mumin1 = 52.2 , 44.9 # [cm^2/Vs]
    mumin2 = 52.2 , 0.0000e+00 # [cm^2/Vs]
    mul = 43.4 , 29 # [cm^2/Vs]
    Pc = 0.0000e+00 , 9.2300e+16 # [cm^3]
    Cr = 9.6800e+16 , 2.2300e+17 # [cm^3]
    Cs = 3.4300e+20 , 6.1000e+20 # [cm^3]
    alpha = 0.68 , 0.719 # [1]
    beta = 2 , 2 # [1]
}

EnormalDependence:
{ * Mobility degradation due to interfaces (semiconductor-oxide. ??This should be
  turned on in our simulations. (also we probably want it flaged to be perpendicular to
  the boundrys and not the current.
  * mu_Enorm^(-1) = mu_ac^(-1) + mu_sr^(-1) with:
  * mu_ac = B / Enorm + C (T/T0)^(-k) (N/N0)^lambda / Enorm^(1/3)
  * mu_sr^(-1) = Enorm^(A+alpha*n/(N+N1)^nu) / delta + Enorm^3 / eta
  * EnormalDependence is added with factor exp(-l/l_crit), where l is
  * the distance to the nearest point of semiconductor/insulator interface.
  * Factor is equal to 1 if l_crit > 100.
    B = 4.7500e+07 , 9.9250e+06 # [cm/s]
    C = 5.8000e+02 , 2.9470e+03 # [cm^(5/3)/(V^(2/3)s)]
    N0 = 1 , 1 # [cm^(-3)]
    lambda = 0.125 , 0.0317 # [1]
    k = 1 , 1 # [1]
    delta = 5.8200e+14 , 2.0546e+14 # [V/s]
    A = 2 , 2 # [1]
    alpha = 0.0000e+00 , 0.0000e+00 # [1]
    N1 = 1 , 1 # [cm^(-3)]
    nu = 1 , 1 # [1]
    eta = 5.8200e+30 , 2.0546e+30 # [V^2/cm*s]
    l_crit = 1.0000e-06 , 1.0000e-06 # [cm]
}

```

```

}

HighFieldDependence:
{ *
  * Deals with velocity saturation due to high electric fields. ??We need this turned
on. Which method should we be using?
  * Caughey-Thomas model:
  * mu_highfield = ( (alpha+1)*mu_lowfield ) /
  *   ( alpha + ( 1 + ( (alpha+1)*mu_lowfield*E/vsat)^beta )^(1/beta) )
  * beta = beta0 (T/T0)^betaexp.
    beta0 = 1.109 ,    1.213 # [1]
    betaexp = 0.66 ,    0.17 # [1]
    alpha = 0.0000e+00 ,    0.0000e+00 # [1]

  * Smoothing parameter for HydroHighField Caughey-Thomas model:
  * if Tl < Tc < (1+K_dT)*Tl, then smoothing between low field mobility
  * and HydroHighField mobility is used.
    K_dT = 0.2 ,    0.2 # [1]
  * Transferred-Electron Effect:
  * mu_highfield = (mu_lowfield+(vsat/E)*(E/E0_TrEf)^4)/(1+(E/E0_TrEf)^4)
    E0_TrEf = 4.0000e+03 ,    4.0000e+03 # [1]
    Ksmooth_TrEf = 1 , 1 # [1]

  * For vsat either Formula1 or Formula2 can be used.
    Vsat_Formula = 1 , 1 # [1]
  * Formula1 for saturation velocity:
  *   vsat = vsat0 (T/T0)^(-Vsatexp)
  * (Parameter Vsat_Formula has to be not equal to 2)
    vsat0 = 1.0700e+07 ,    8.3700e+06 # [1]
    vsatexp = 0.87 ,    0.52 # [1]
}

Scharfetter * relation and trap level for SRH recombination:
{ * tau = taumin + ( taumax - taumin ) / ( 1 + ( N/Nref )^gamma)
  * tau(T) = tau * ( (T/300)^Talpha ) (TempDep)
  * tau(T) = tau * exp( Tcoeff * ((T/300)-1) ) (ExpTempDep)
    taumin = 0.0000e+00 ,    0.0000e+00 # [s]
    taumax = 1.0000e-05, 3.0000e-06 # [s]
    *taumax = 1.0000e-09, 1.0000e-09 # [s]
    Nref = 1.0000e+16 ,    1.0000e+16 # [cm^(-3)]
    gamma = 1 , 1 # [1]
    Talpha = -1.5000e+00 ,    -1.5000e+00 # [1]
    Tcoeff = 2.55 ,    2.55 # [1]
    Etrap = 0.0000e+00 # [eV]
}

Auger * coefficients:
{ * R_Auger = ( C_n n + C_p p ) ( n p - ni_eff^2)
  * with C_n,p = (A + B (T/T0) + C (T/T0)^2) (1 + H exp(-{n,p}/N0))
    A = 6.7000e-32 ,    7.2000e-32 # [cm^6/s]
    B = 2.4500e-31 ,    4.5000e-33 # [cm^6/s]
    C = -2.2000e-32 ,    2.6300e-32 # [cm^6/s]
    H = 3.46667 ,    8.25688 # [1]
    N0 = 1.0000e+18 ,    1.0000e+18 # [cm^(-3)]
}

Band2BandTunneling
{ * See Sentaurus Device manual `Band-To-Band Tunneling'
  A = 8.9770e+20 # [cm / (s V^2)]
  B = 2.1466e+07 # [eV^(-3/2) V/cm]
  hbarOmega = 0.0186 # [eV]

  * Traditional models for the following keywords in input file:

```

```

* Band2Band(E1) : A1*E*exp(-B1/E)
* Band2Band(E1_5): A1_5*E^1.5*exp(-B1_5/E)
* Band2Band(E2) : A2*E^2*exp(-B2/E)
  A1      = 1.1000e+27 # [1/cm/sec/V]
  B1      = 2.1300e+07 # [V/cm]
  A1_5    = 1.9000e+24 # [1/cm/sec/V^1.5]
  B1_5    = 2.1900e+07 # [V/cm]
  A2      = 3.5000e+21 # [1/cm/sec/V^2]
  B2      = 2.2500e+07 # [V/cm]

* Hurkx model for the following keywords in input file:
* Band2Band(Hurkx) : -Agen*D*(E/E0)^Pgen*exp(-Bgen*(Eg/Eg300)^1.5/E) if D < 0
*                  -Arec*D*(E/E0)^Prec*exp(-Brec*(Eg/Eg300)^1.5/E) if D > 0
*                  D = (n*p-ni^2)/(n+ni)/(p+ni)*(1-|alpha|)+alpha, E0 = 1 V/cm
*                  So, if alpha = 0, it's original Hurkx model,
*                  if alpha = -1, it's only generation,
*                  if alpha = +1, it's only recombination.
  Agen    = 3.5000e+21 # [1/cm^3/sec]
  Bgen    = 2.2500e+07 # [V/cm]
  Pgen    = 2          # [1]
  Arec    = 3.5000e+21 # [1/cm^3/sec]
  Brec    = 2.2500e+07 # [V/cm]
  Prec    = 2          # [1]
  alpha   = 0.0000e+00 # [1]

* min length to interfaces (for traditional & Hurkx models):
  dDist   = 0.0000e+00 # [cm]
* min potential difference on length dPot/E (for traditional & Hurkx models):
  dPot    = 0.0000e+00 # [V]
}
}

MaterialInterface = "Oxide/Silicon" {
SurfaceRecombination * surface SRH recombination:
{ * s = S0 ( 1 + Sref ( N/Nref )^gamma ) recombination velocity
  S0      = 2.0e+4 , 2.0e+4 # [cm/s]
  Sref    = 1.0000e-03 # [1]
  Nref    = 1.0000e+16 # [cm^(-3)]
  gamma   = 1          # [1]
  Etrap   = 0.0000e+00 # [eV]
}
}

Material = "PolySi" {
Bandgap
{ * Eg = Eg0 + dEg0 + alpha Tpar^2 / (beta + Tpar) - alpha T^2 / (beta + T)
* dEg0(<bgn_model_name>) is a band gap correction term. It is used together with
* an appropriate BGN model, if this BGN model is chosen in Physics section
* Parameter 'Tpar' specifies the value of lattice
* temperature, at which parameters below are defined
* Chi0 is electron affinity.
  Chi0    = 4.05 # [eV]
  Bgn2Chi = 0.5 # [1]
  Eg0     = 1.16964 # [eV]
  dEg0(Bennett) = 0.0000e+00 # [eV]
  dEg0(Slotboom) = -4.7950e-03 # [eV]
  dEg0(OldSlotboom) = -1.5950e-02 # [eV]
  dEg0(delAlamo) = -1.4070e-02 # [eV]
  alpha   = 4.7300e-04 # [eV K^-1]
  beta    = 6.3600e+02 # [K]
  Tpar    = 0.0000e+00 # [K]
}
}

```

```

OldSlotboom
{ * deltaEg = dEg0 + Ebgn ( ln(N/Nref) + [ (ln(N/Nref))^2 + 0.5]^1/2 )
  * dEg0 is defined in BandGap section
    Ebgn   = 9.0000e-03 # [eV]
    Nref   = 1.0000e+17 # [cm^(-3)]
}

eDOSMass
{
  * For effective mass specificatition Formula1 (me approximation):
  * or Formula2 (Nc300) can be used :
    Formula   = 1 # [1]
  * Formula1:
  * me/m0 = [ (6 * mt)^2 * ml ]^(1/3) + mm
  * mt = a[Eg(0)/Eg(T)]
  * Nc(T) = 2(2pi*kB/h_Planck^2*me*T)^3/2 = 2.540e19 ((me/m0)*(T/300))^3/2
    a        = 0.1905 # [1]
    ml       = 0.9163 # [1]
    mm       = 0.0000e+00 # [1]
}

hDOSMass
{
  * For effective mass specificatition Formula1 (mh approximation):
  * or Formula2 (Nv300) can be used :
    Formula   = 1 # [1]
  * Formula1:
  * mh = m0*[(a+bT+cT^2+dT^3+eT^4)/(1+fT+gT^2+hT^3+iT^4)]^(2/3) + mm
  * Nv(T) = 2(2pi*kB/h_Planck^2*mh*T)^3/2 = 2.540e19 ((mh/m0)*(T/300))^3/2
    a        = 0.443587 # [1]
    b        = 0.003609528 # [K^-1]
    c        = 0.0001173515 # [K^-2]
    d        = 1.263218e-06 # [K^-3]
    e        = 3.025581e-09 # [K^-4]
    f        = 0.004683382 # [K^-1]
    g        = 0.0002286895 # [K^-2]
    h        = 7.469271e-07 # [K^-3]
    i        = 1.727481e-09 # [K^-4]
    mm       = 0 # [1]
}

ConstantMobility:
{ * mu_const = mumax (T/T0)^(-Exponent)
  mumax = 1.4170e+03 , 4.7050e+02 # [cm^2/(Vs)]
  Exponent = 2.5 , 2.2 # [1]
  mutunnel = 0.05 , 0.05 # [cm^2/(Vs)]
}

DopingDependence:
{
  * For doping dependent mobility model three formulas
  * can be used. Formula1 is based on Masetti et al. approximation.
  * Formula2 uses approximation, suggested by Arora.
    formula = 1 , 1 # [1]
  * If formula=1, model suggested by Masetti et al. is used:
  * mu_dop = mumin1 exp(-Pc/N) + (mu_const - mumin2)/(1+(N/Cr)^alpha)
  * - mul/(1+(Cs/N)^beta)
  * with mu_const from ConstantMobility
    mumin1 = 52.2 , 44.9 # [cm^2/Vs]
    mumin2 = 52.2 , 0.0000e+00 # [cm^2/Vs]
    mul = 43.4 , 29 # [cm^2/Vs]
    Pc = 0.0000e+00 , 9.2300e+16 # [cm^3]
}

```

```

        Cr      = 9.6800e+16 ,      2.2300e+17   # [cm^3]
        Cs      = 3.4300e+20 ,      6.1000e+20   # [cm^3]
        alpha   = 0.68 ,      0.719   # [1]
        beta    = 2 , 2          # [1]
    }

HighFieldDependence:
{ * Caughey-Thomas model:
  * mu_highfield = ( (alpha+1)*mu_lowfield ) /
  *   ( alpha + ( 1 + ( (alpha+1)*mu_lowfield*E/vsat)^beta )^(1/beta) )
  * beta = beta0 (T/T0)^betaexp.
    beta0 = 1.109 ,      1.213   # [1]
    betaexp = 0.66 ,      0.17   # [1]
    alpha = 0.0000e+00 ,      0.0000e+00   # [1]

  * Smoothing parameter for HydroHighField Caughey-Thomas model:
  * if Tl < Tc < (1+K_dT)*Tl, then smoothing between low field mobility
  * and HydroHighField mobility is used.
    K_dT = 0.2 ,      0.2   # [1]
  * Transferred-Electron Effect:
  * mu_highfield = (mu_lowfield+(vsat/E)*(E/E0_TrEf)^4)/(1+(E/E0_TrEf)^4)
    E0_TrEf = 4.0000e+03 ,      4.0000e+03   # [1]
    Ksmooth_TrEf = 1 , 1   # [1]

  * For vsat either Formulal or Formula2 can be used.
    Vsat_Formula = 1 , 1   # [1]
  * Formulal for saturation velocity:
  *   vsat = vsat0 (T/T0)^(-Vsatexp)
  * (Parameter Vsat_Formula has to be not equal to 2)
    vsat0 = 1.0700e+07 ,      8.3700e+06   # [1]
    vsatexp = 0.87 ,      0.52   # [1]
}

Scharfetter * relation and trap level for SRH recombination:
{ * tau = taumin + ( taumax - taumin ) / ( 1 + ( N/Nref )^gamma)
  * tau(T) = tau * ( (T/300)^Talpha ) (TempDep)
  * tau(T) = tau * exp( Tcoeff * ((T/300)-1) ) (ExpTempDep)
    taumin = 0.0000e+00 ,      0.0000e+00   # [s]
    taumax = 200.00e-12 ,      200.00e-12   # [s] * varies from 1e-9 to 10e-12
    Nref = 1.0000e+16 ,      1.0000e+16   # [cm^(-3)]
    gamma = 1 , 1   # [1]
    Talpha = -1.5000e+00 ,      -1.5000e+00   # [1]
    Tcoeff = 2.55 ,      2.55   # [1]
    Etrap = 0.0000e+00 # [eV]
}
}

```

B: Example Electrical Simulation Code

Example Sentaurus Structure Editor Input File:

```

;; Reinitializing SDE
(sde:clear)

;; Setting parameters
; - lateral
(define Ltot 2.12) ; [um] Lateral extend total, check that
(define Lguide 0.500) ; [um] Guide width
(define Ltoj 0.05) ; [um] distance from left side of guide to end of encroachment
(define Lsp 0.20) ; [um] distance from guide to highly doped contact (spacing
width)
(define Lleftside 0.05) ; [um] thickness of the side wall on the left

```

```

(define Lrightside 0.05); [um] thickness of the side wall on the right
;(define Lgap 0.0) ; [um] gap across the top of the layer

;; - layers
(define Htot 0.72) ; [um] Vertical thickness total
(define Hsub 0.24) ; [um] Substrate thickness
(define Hsilicon 0.05) ; [um] EPI thickness
(define Hguide 0.220) ; [um] Gate thickness
(define Hjunct 0.05) ; [um] thickness of horizontal junction (total).
;(define Hj bott 0.080) ; [um] thickness of the bottom layer of the horizontal
junction

;minimum junction meshing resolution
(define minmesh 0.001) ; [um] consider dropping to 1nm.
(define medmesh 0.005)
(define maxmesh 0.01)

;; Dopings
;doping level at contacts (will be highest level in device).
(define rightContDop 1e19) ; [1/cm^3]
(define leftContDop 1e19) ; [1/cm^3]
;doping level of side layers (middle doping)
(define rightSideLayerDop 1.5e18) ; [1/cm^3]
(define leftSideLayerDop 1e18) ; [1/cm^3]
;doping level of side walls (middle or lower doping)
(define rightSideWallDop 1.5e18) ; [1/cm^3]
(define leftSideWallDop 1e18) ; [1/cm^3]
;doping level of the guide (lowest level)
(define GuideDop 5e17) ; [1/cm^3]

;define doping types
(define doctype_p "BoronActiveConcentration")
(define doctype_n "ArsenicActiveConcentration")

;-----
;; Derived quantities
(define Xmax (/ Ltot 2.0))
(define Xmin (* Xmax -1.0))
(define Xguide (/ Lguide 2.0))
(define Xjunct (+(* Xguide -1.0) Ltoj))
(define Xleftside (+ (* Xguide -1.0) Lleftside))
(define Xrightside (- Xguide Lrightside))
(define Xsp (+ Xguide Lsp))

(define Ymax Htot)
(define Ymin 0.0)
(define Ysub (- Ymax Hsub)) ;distance from top to bottom of Si
(define Yguide (- Ysub Hguide )) ;distance from top to top of guide
(define Ysilicon (- Ysub Hsilicon )) ;distance from top to top of side layers
(define Yjunct (+ Yguide Hjunct)) ;distance from top to bottom of junction layer

;----Next build the actual device----.
;-----
(display "Generating layer structure... \n")
; Overlap resolution: New replaces Old
(sdegeo:set-default-boolean "ABA")

;-----
; Creating oxide region
(sdegeo:create-rectangle
(position Xmin Ymin 0.0 )
(position Xmax Ymax 0.0 ) "SiO2" "oxide_region_1" )

```

```

; Creating Si gate
(sdegeo:create-rectangle
  (position (* Xguide -1.0) Yguide 0.0)
  (position Xguide Ysub 0.0)
  "Silicon" "guide_active_layer")

; Creating guide Silicon encroach layer
(sdegeo:create-rectangle
  (position Xrightside Yjunct 0.0 )
  (position Xjunct Yguide 0.0 )
  "Silicon" "guideEncroachTopLayer"
)

; Creating Left side-wall
(sdegeo:create-rectangle
  (position (* Xguide -1.0) Yguide 0.0 )
  (position Xleftside Ysub 0.0 )
  "Silicon" "guideLeftSideWall"
)

; Creating Right side-wall
(sdegeo:create-rectangle
  (position Xguide Ysub 0.0 )
  (position Xrightside Yguide 0.0 )
  "Silicon" "guideRightSideWall"
)

; Creating Side Silicon layer right
(sdegeo:create-rectangle
  (position Xguide Ysilicon 0.0 )
  (position Xsp Ysub 0.0 )
  "Silicon" "side_spacer_right"
)

; Creating Spacing Silicon layer left
(sdegeo:create-rectangle
  (position (* Xguide -1.0) Ysilicon 0.0 )
  (position (* Xsp -1.0) Ysub 0.0 )
  "Silicon" "side_spacer_left"
)

; Creating Contact Silicon layer right
(sdegeo:create-rectangle
  (position Xsp Ysilicon 0.0 )
  (position Xmax Ysub 0.0 )
  "Silicon" "contact_layer_right"
)

; Creating Contact Silicon layer left
(sdegeo:create-rectangle
  (position (* Xsp -1.0) Ysilicon 0.0)
  (position Xmin Ysub 0.0)
  "Silicon" "contact_layer_left"
)

;-----
(display "Defining contacts\n")
; Contact declarations
(sdegeo:define-contact-set "Anode"
  4.0 (color:rgb 1.0 0.0 0.0 ) "##")

```

```

(sdegeo:define-contact-set "Cathode"
  4.0 (color:rgb 0.0 1.0 0.0) "###")

;(sdegeo:define-contact-set "P_thermal"
; 4.0 (color:rgb 0.0 0.0 1.0) "###")

;(sdegeo:define-contact-set "N_thermal"
; 4.0 (color:rgb 1.0 1.0 0.0) "###")

;-----
; Contact settings
; Electrical Contacts
(sdegeo:set-current-contact-set "Anode")
(sdegeo:define-2d-contact
 (find-edge-id (position Xmin (* 0.5 (+ Ysub Ysilicon)) 0.0))
 "Anode" )

(sdegeo:set-current-contact-set "Cathode")
(sdegeo:define-2d-contact
 (find-edge-id (position Xmax (* 0.5 (+ Ysub Ysilicon)) 0.0))
 "Cathode")
(display "Contact define completed.\n")

;-----
; Separating lumps
(sde:assign-material-and-region-names "all")

;-----
; Setting region names
(sde:addmaterial
 (find-body-id (position 0.0 (/ (+ Ysub Ymax) 2) 0.0))
 "SiO2" "oxide_substrate_region")

(sde:addmaterial
 (find-body-id (position 0.0 (/ (+ Ymin Yguide) 2) 0.0))
 "SiO2" "oxide_cap_region")

(display "Completed region re-name.\n")
;-----
; Profiles:isUndernTyp

;Guide layer
(sdedr:define-constant-profile "Const.active_guide"
 doctype_p GuideDop)
(sdedr:define-constant-profile-region "Place.active_guide"
 "Const.active_guide" "guide_active_layer" )

;Guide Encroach layer
(sdedr:define-constant-profile "Const.guideEncroachTopLayer"
 doctype_n rightSideWallDop)
(sdedr:define-constant-profile-region "Place.guideEncroachTopLayer"
 "Const.guideEncroachTopLayer" "guideEncroachTopLayer" )

; Sidewall area doping
(sdedr:define-constant-profile "Const.guideLeftSideWall"
 doctype_p leftSideWallDop)
(sdedr:define-constant-profile-region "Place.guideLeftSideWall"
 "Const.guideLeftSideWall" "guideLeftSideWall" )

;other side wall
(sdedr:define-constant-profile "Const.guideRightSideWall"
 doctype_n rightSideWallDop)
(sdedr:define-constant-profile-region "Place.guideRightSideWall"

```

```

"Const.guideRightSideWall" "guideRightSideWall" )

; Spacer area doping
;p-type spacer
(sdedr:define-constant-profile "Const.left_Spacer"
 doctype_p leftSideLayerDop)
(sdedr:define-constant-profile-region "Place.left_Spacer"
 "Const.left_Spacer" "side_spacer_left" )
;n-type spacer
(sdedr:define-constant-profile "Const.right_Spacer"
 doctype_n rightSideLayerDop)
(sdedr:define-constant-profile-region "Place.right_Spacer"
 "Const.right_Spacer" "side_spacer_right" )

;insert doping for other complicated regions here if needed

;Contact area doping
(sdedr:define-constant-profile "Const.left_Contact"
 doctype_p leftContDop )
(sdedr:define-constant-profile-region "Place.left_Contact"
 "Const.left_Contact" "contact_layer_left" )

(sdedr:define-constant-profile "Const.right_Contact"
 doctype_n rightContDop )
(sdedr:define-constant-profile-region "Place.right_Contact"
 "Const.right_Contact" "contact_layer_right" )

;; Z2007.03 version
;;Saving BND file
;(sdeio:save-tdr-bnd (get-body-list) "")

(display "Completed doping definitions.\n")
;-----
; Meshing Strategy:
; note that refinement numbers go as xmax ymax xmin ymin.

;Over-all mesh:
(sdedr:define-refeaval-window "Total_mesh_win" "Rectangle"
 (position Xmin Ymin 0.0 )
 (position Xmax Ymax 0.0 ) )
(sdedr:define-refinement-size "Total_guide_win_Def" maxmesh maxmesh medmesh medmesh)
(sdedr:define-refinement-placement "Total_win_place" "Total_guide_win_Def"
"Total_mesh_win")

;Mesh around guide -- start from side layers
(sdedr:define-refeaval-window "Guide_mesh_win" "Rectangle"
 (position (* Xsp -1.0) (- Yguide (/ Yguide 2)) 0.0 )
 (position Xsp (/ (+ Ysub Ymax) 2.0) 0.0 ) )
(sdedr:define-refinement-size "Guide_guide_win_Def" medmesh medmesh minmesh minmesh)
(sdedr:define-refinement-placement "Guide_win_place" "Guide_guide_win_Def"
"Guide_mesh_win")

;Mesh side parts
(sdedr:define-refeaval-window "Left_side_mesh_win" "Rectangle"
 (position Xmin (- Ysilicon (/ Yguide 2)) 0.0 )
 (position (* Xsp -1.0) (/ (+ Ysub Ymax) 2.0 ) 0.0 ) )
(sdedr:define-refinement-size "Left_side_win_Def" medmesh medmesh minmesh minmesh)
(sdedr:define-refinement-placement "Left_side_win_place" "Left_side_win_Def"
"Left_side_mesh_win")
;;and the other side
(sdedr:define-refeaval-window "Right_side_mesh_win" "Rectangle"
 (position Xsp (- Ysilicon (/ Yguide 2)) 0.0 )
 (position Xmax (/ (+ Ysub Ymax) 2.0 ) 0.0 ) )

```

```

(sdedr:define-refinement-size "Right_side_win_Def" medmesh medmesh minmesh minmesh)
(sdedr:define-refinement-placement "Right_side_win_place" "Right_side_win_Def"
"Right_side_mesh_win")

(display "Finished defining meshing strategy.\n")
;;-----
;save files and generate mesh:

(sdeio:save-tdr-bnd (get-body-list) "n@node@_bnd") ;save BND file
(sdedr:write-cmd-file "n@node@_cmd") ; Save CMD file
(sdedr:append-cmd-file "") ;if don't have this line, everything dies
;(display ".bnd and .cmd files saved.\n")

(sde:build-mesh "mesh" "-F tdr -s " "n@node@_msh")
(display "Mesh generation completed.\n")
;end of file

```

Example Sentaurus Device Input File:

```

Electrode {
    { Name="Anode" Voltage=0.0}
    { Name="Cathode" Voltage= 0.0}
}

Thermode {
    {Name="Anode" Temperature = 300 SurfaceResistance = 1.4e-4}
    {Name="Cathode" Temperature = 300 SurfaceResistance = 1.4e-4}
}

File {
    Grid = @tdr@
    Parameters = "sdevice.par"
    *Need parameter's file here
    Current = @plot@
    Plot = @tdrdat@
    Output = @log@
}

Physics (MaterialInterface = "Silicon/Oxide"){
    Recombination( surfaceSRH )
}

Physics {
    Thermodynamic
    Mobility ( DopingDep
        Enormal
        eHighFieldsat(GradQuasiFermi)
        hHighFieldsat(GradQuasiFermi) )

    Recombination(
        SRH(DopingDep)
        Auger
        surfaceSRH
        Avalanche( Okuto )
        Band2Band(E1)
    )
    EffectiveIntrinsicDensity ( OldSlotBoom )
    Fermi
}

Math {
    Digits = 9
}

```

```

    Extrapolate
    Derivatives
    RecomputeQFP
    RelErrControl
        Iterations = 30
    BreakCriteria{ Current(Contact = "Anode" AbsVal = 1e-5)}
    NewDiscretization
    Wallclock
    Method = ParDiSo
}

Solve {
    Poisson
    Coupled { Poisson Electron Hole }
    Coupled {Poisson Electron Hole Temperature}
    Quasistationary (
        InitialStep = 0.002
        MaxStep = 0.01
        MinStep = 1e-4
        Goal {Name="Anode" Voltage =-5.0}
        Plot { Range = (0 1) Intervals = 25}
    )
    {Coupled {Poisson Electron Hole Temperature}}
}

Plot {
    eDensity hDensity eCurrent hCurrent
    Potential SpaceCharge ElectricField
    eMobility hMobility eVelocity/Vector hVelocity/Vector eLifetime hLifetime
    Doping DonorConcentration AcceptorConcentration
    *Refractiveindex eTemperature eRelativeEffectiveMass
    *hRelativeEffectiveMass eAugerRecombination SurfaceRecombination
}

```

C: Optical Processing Code

```

function tecplot_output_process_1(myfilenames, numvar, numel, myvars, myxsize,
myysize)
%File takes in tecplot ".dat" files and converts the data in them to 2D graphs.
%.dat file must be in ascii point form (when saving tecplot file this is
%one of the options)
%It then, making some assumptions converts the carrier panes into index
%and absorption panes and outputs these...
%Inputs:
% myfilenames - list of tecplot data files that want to process (cell array of
strings). Files
% do not need to be in any specific order, but they all must have the
% same format (same number of tecplot variables, ones desired in the
% same location on the list, same desired output gridding)
% numvar - number of variables in the tecplot output file (can see this
% by opening the file in wordpad a: the variable numbers and names are
% listed in the header information.
% numel - number of vertices a FE element has (4 if quadrilateral, 3 if
% a triangle).
% myvars - location of the variables that want to extract and grid in
% tecplot's list of variables. The first number MUST correspond to the
% location of the electron density and the second one MUST be the
% location of the hole density (usually these are 4 and 5 respectively).
% After that can have whatever variables want.
% myxsize - size, in um, of the x-dimension of the output blocks of the
% grid. Please note that it is important to pick this parameter
% correctly.
% myysize - size in um, of the y-dimension of the output blocks of the

```

```

% grid. Please note that it is important to pick this parameter
% correctly.
%Outputs:
%not output variables are generated
%Output Files:
%"input_name2test.mat": straight up parsed output file -- simply contains
%lists of the variables and coordinates , in order. Not yet gridded
%"input_name_processed.mat": contains 2D grids of variable , in order, in
%a 3D array. Also conatins index and absorption panes created for 1550nm
%light in silicon on a SiO2 background.

%parse files into MATLAB variable list
for ij = 1:length(myfilenames)
    mycurfile = myfilenames{ij};
    Create_reformatted_files(mycurfile, numvar, numel);
end

%Copy variables into array
for ij = 1:length(myfilenames)
    filename = myfilenames{ij};
    filename = [filename(1:(end-4)), '2test.mat'];
    [C, my_ymap, my_xmap] = Create_structure_arrays2(filename, myxsize, myysize,
myvars); %create arrays with known data points
    C = grid_smoother(C); %interpolate between those points
    filename = [filename(1:(end-9)), '_processed.mat'];
    save(filename, 'C', 'my_xmap', 'my_ymap', 'myvars', 'myxsize', 'myysize');
end

makesoref_panes = 1; %boolean that allows you to turn this part of the code off, if
would rather make index panes on own separately
for ij = 1:length(myfilenames)
    filename = myfilenames{ij};
    filename = [filename(1:(end-4)), '2test.mat'];
    filename = [filename(1:(end-9)), '_processed.mat'];
    load(filename);
    if(makesoref_panes == 1)
        %use soref model and get modal index - figure out how many decimals of
        %accuracy have here...
        mytemp_screne = (C(:,:,1) ~= 0);
        mytemp_opscrene = (C(:,:,1) == 0);
        mytemp_screne = mytemp_screne.*3.48; %correct to three sig-figs? Change from
flat silicon
        mytemp_opscrene = mytemp_opscrene.*1.445;
        mytemp_screne = mytemp_screne + mytemp_opscrene;
        My_index_pane = (-8.8e-22.*C(:,:,1) -8.5e-18.*C(:,:,2).^0.8) +mytemp_screne;
        My_delta_absorption_alpha_pane = (8.5e-18.*C(:,:,1)+6.0e-18.*C(:,:,2));

    end
    save(filename, 'C', 'my_xmap', 'my_ymap', 'myvars', 'myxsize', 'myysize',
'My_index_pane', 'My_delta_absorption_alpha_pane');
end
end

function tecplot_output_process_1_lambda_var(myfilenames, mylambda)
%Function takes output of and tecplot_output_process_1.m and creates a new
%output file where the index and absorption panes correspond to a specified
%wavelength instead of to 1550 nm.
%Inputs:
% -myfilenames - list of ".mat" output files from tecplot_output_process_1.mat
% in general these will have the form "input_name_processed.mat"
% -mylambda - desired wavelength for index and absorption in microns
% (fits are only good between 1.2 and 1.6 microns)
%Outputs:

```

```

%no output variables are generated
%Output Files:
%"input_name_processed_lambda_'mylambda'.mat": contains 2D grids of
%variables, in order, in a 3D array. Also contains index and absorption
%panes created at specified wavelength, assuming a SiO2 background.

for ij = 1:length(myfilenames)
    disp(['on file ', num2str(ij)]);
    filename = myfilenames{ij};
    load(filename);
    save_filename = [filename(1:end-4), '_lambda_', num2str(mylambda), '.mat'];
    if(1)
        %use soref model and get modal index
        mytemp_screne = (C(:,:,1) ~= 0);
        mytemp_opscrene = (C(:,:,1) == 0);
        mytemp_screne = mytemp_screne.*3.48; %correct to three sig-figs? Change from
flat silicon
        mytemp_opscrene = mytemp_opscrene.*1.445;
        mytemp_screne = mytemp_screne + mytemp_opscrene;
        %My_index_pane = (-8.8e-22.*C(:,:,1) -8.5e-18.*C(:,:,2).^0.8) +mytemp_screne;
        %My_delta_absorption_alpha_pane = (8.5e-18.*C(:,:,1)+6.0e-18.*C(:,:,2));
        [dne, dae] = E_soref_n_si_for_matrix(C(:,:,1), mylambda);
        [dnp, dap] = E_soref_p_si_for_matrix(C(:,:,2), mylambda);
        My_index_pane = dne+dnp+mytemp_screne;
        My_delta_absorption_alpha_pane = dae+dap; %insert desired background loss here
    end
    save(save_filename, 'C', 'my_xmap', 'my_ymap', 'myvars', 'myxsize', 'mysize',
'My_index_pane', 'My_delta_absorption_alpha_pane');
    %note that in the end, all this really does is open the results at
    %1550nm, overwrite the index and absorption panes, and save the results
    %in a different file labeled with the new lambda value...
end
end

```

```

function my_neff = tecplot_output_process_2(my_filenames, my_volts, output_filename)
%File takes in output ".mat" file from tecplot_output_process_1.m
%It finds the refractive index and absorption for the structure contained
%therein and saves then in an array. (While only these two values are
%saved it does get the full mode solution from Milos's mode solver, so if wanted to,
%could easily save other variables).
%
%Inputs:
% my_filenames - list of MATLAB files that want to process (cell array of strings).
Note that
% files need to be related and in order. (I usually input all the
% outputs of one device in order from lowest to highest bias voltage.
% While the x-axis variable is called "volts" if could be used to store
% whatever (frequency, doping level, the like). The simple point I'm
% trying to make here is that there *is* an x-axis variable.
% my_volts - corresponding voltage for each file. (could also be used to
% store another x-axis variable (see above)).
% output_filename - string containing the name of the file to which to
% save the outputs
%Outputs:
% my_neff - the (non-complex) index change along my_volts (I find having
% this useful, though not strictly necessary)
%Output Files:
%"output_filename" - contains copies of "my_filenames" and "my_volts" as
%well as the corresponding effect index (my_neff) and the effective absorption
%(my_alpha) of the structures. *my_alpha is in 1/cm.*
%Note that to get the CHANGE in index or absorption, one needs to subtract
%the first value from the array

```

%IMPORTANT NOTE: PARTICULARLY FOR HIGH INDEX STRUCTURES, MILOS'S MODE
 %SOLVER GIVES THE CORRECT *RELATIVE* BUT NOT NECESSARILY THE CORRECT
 %*ABSOLUTE* BETA (INDEX AND ABSORPTION). BECAUSE OF THIS IT IS BETTER TO
 %COMPARE *CHANGES* WHEN ACCURACY IS DESIRED

```

%initialize variables:
k0 = 2*pi/1.55; %BE CAREFUL HARD WIRED!!!
num_files = length(my_filenames);
my_neff = zeros(1, num_files);
my_alpha = zeros(1, num_files);
%then loop through files:
for ij = 1:1:num_files
    disp(['completed file ', num2str(ij), ' of ' num2str(num_files)]); %so don't go
    crazy
    clear my_temp_grid my_add_around2 my_add_around Nn Nx Ny N F mtemp mtemp2 myipane
    mytempipane
    filename = my_filenames{ij};
    load(filename);
    %change grid
    sf = 5; %BE CAREFUL HARD WIRED!!!
    yc = mod(size(My_index_pane), sf); yc1 = [floor(yc(1)/2), ceil(yc(1)/2)]; yc2 =
    [floor(yc(2)/2), ceil(yc(2)/2)];
    My_index_pane = My_index_pane((1+yc1(1)):end-yc1(2)), (1+yc2(1)):end-yc2(2));
    My_delta_absorption_alpha_pane = My_delta_absorption_alpha_pane((1+yc1(1)):end-
    yc1(2)), (1+yc2(1)):end-yc2(2));
    myipane = My_index_pane';
    myapane = 1e-4*My_delta_absorption_alpha_pane';
    % myipane = [myipane; myipane(end,:)];
    mytempipane = 0;
    mytempapane = 0;
    for ij2 = 1:1:sf
        for jk = 1:1:sf
            mytempipane = mytempipane+myipane(ij2:sf:end, jk:sf:end).^2;
            mytempapane = mytempapane+myapane(ij2:sf:end, jk:sf:end);
        end
    end
    N3 = mytempipane./sf^2;
    N3a = mytempapane./sf^2;
    %convert alpha absorption to a complex index
    N3a = i*N3a/(2*k0);
    %apply "simple" index averaging algorithm used in Milos's wrapper function
    my_temp_grid = zeros(2*size(N3));
    my_temp_grid_a = zeros(2*size(N3a));
    my_temp_grid(2:2:end, 2:2:end) = sqrt(N3);
    my_temp_grid = [my_temp_grid(:, 11:end), my_temp_grid(:, 1:10)];
    my_temp_grid(3:2:end-2, :) = sqrt( (my_temp_grid(2:2:end-
    3, :).^2+my_temp_grid(4:2:end-1, :).^2)/2);
    my_temp_grid(:, 3:2:end-2) = sqrt( (my_temp_grid(:, 2:2:end-
    3).^2+my_temp_grid(:, 4:2:end-1).^2)/2);
    my_temp_grid(1, :) = my_temp_grid(2, :); my_temp_grid(end, :) = my_temp_grid(end-
    1, :); my_temp_grid(:, 1) = my_temp_grid(:, 2); my_temp_grid(:, end) = my_temp_grid(:, end-
    1);
    my_temp_grid_a(2:2:end, 2:2:end) = N3a;
    my_temp_grid_a = [my_temp_grid_a(:, 11:end), my_temp_grid_a(:, 1:10)];
    my_temp_grid_a(3:2:end-2, :) = (my_temp_grid_a(2:2:end-3, :)+my_temp_grid_a(4:2:end-
    1, :))/2;
    my_temp_grid_a(:, 3:2:end-2) = (my_temp_grid_a(:, 2:2:end-
    3)+my_temp_grid_a(:, 4:2:end-1))/2;
    my_temp_grid_a(1, :) = my_temp_grid_a(2, :); my_temp_grid_a(end, :) =
    my_temp_grid_a(end-1, :); my_temp_grid_a(:, 1) = my_temp_grid_a(:, 2);
    my_temp_grid_a(:, end) = my_temp_grid_a(:, end-1);
    my_temp_grid = my_temp_grid+ my_temp_grid_a;

```

```

    %the next few lines expand boundaries of the grid because Milos's mode solver
needs
    %some padding and often sentaurus outputs don't have it
    s1 = size(my_temp_grid,1);
    s2 = size(my_temp_grid, 2);
    mtemp = my_temp_grid(1,1)*ones(4*size(my_temp_grid));
    mtemp((s1+s1/2+1):(2*s1+s1/2-2), (s2+s2/2+1):(2*s2+s2/2-2)) = my_temp_grid(1:end-
2,1:end-2);
    my_temp_grid = mtemp;
    %set up Mode solver variables
    dxy = [myxsize*sf, myysize*sf];
    options.NMODES_CALC = 1; % how many modes to calculate
    options.PMLwidth = [0 0.0 0.0 0.0]; % left right bottom top; we don't need PMLs
for straight guide
    options.PMLsigma = [0.3 0.3];
    options.mu_guess = k0*2.5; % wave number guess (i.e. k0*neff)
    %options.operver = 'm2dpmloperR2'; options.enginever = 'm2wcylR2'--%NEVER USE THIS
IT IS BROKEN!!!!
    %fix dimensions so line up:
    Nn = my_temp_grid;
    Nn = [Nn, Nn(:,end)];
    Nn = [Nn;Nn(end,:)];
    nny = size(Nn,1);
    nnx = size(Nn,2);
    Ny = (myysize*sf)/2:(myysize*sf)/2:(nny*(myysize*sf)/2);
    Nx = (myxsize*sf)/2:(myxsize*sf)/2:(nnx*(myxsize*sf)/2);
    %get the mode and record the variables:
    [N, F] = sisolver3d_mod(Nn, Ny, Nx, dxy, k0, options);
    my_neff(ij) = real(F.beta/F.k0);
    my_alpha(ij) = 1e4*2*F.k0*imag(F.beta/F.k0);
end

save(output_filename, 'my_filenames', 'my_volts', 'my_neff', 'my_alpha',
'my_neff_cut', 'my_alpha_cut');
end
end

function B = Create_reformatted_files(filename, mynumvar, mynumelsides)
A = textread(filename, '%s', 'delimiter', '\n');
filename2 = [filename(1:(end-4)), '2test.mat'];
B(1:300000,1:mynumvar) = 0;
B2(1:300000, 1:mynumelsides) = 0;
mycount = 1;
mycount2 = 1;
debugline = 5;
myadder = 0;
mycorrectadder = 0;
myprev = 0;
%myblock = 0;
for ij = 1:1:length(A)
    if( mod(ij, 10000)==0)
        disp(['First ', num2str(ij), ' rows completed']);
    end
    if((~isnan(str2double(A{ij}(1)))) || (~isnan(str2double(A{ij}(2)))) &&
(A{ij}(1)=='-')) )
        b = strread(A{ij}, '%f');
        if(length(b) == mynumvar)
            if(myprev == 0)
                mycorrectadder = myadder;
            end
            B(mycount,:) = b;
            mycount = mycount+1;
            myadder = myadder+1;
        end
    end
end

```

```

        myprev = 1;
    elseif(length(b) == mynumelsides)
        if(myprev ==1)
            myprev = 0;
        end
        B2(mycount2,:) = b + mycorrectadder;
        mycount2 = mycount2 +1;
    end
end
end
save(filename2, 'B', 'B2');
B = B(1:(mycount-1), :);
B2 = B2(1:(mycount2-1), :);
save(filename2, 'B', 'B2');
end
end

```

```

function [C, my_ymap, my_xmap] = Create_structure_arrays2(filename, x_element_size,
y_element_size, myvarmap)
%Takes MATLAB variable arrays from Tecplot files and parses them into 2D
%arrays
%Inputs:
% filename- name of MATLAB file containing B and B2 (the array outputs of
% Create_reformatted_file.m")
% x_element_size and y_element_size - are the x and y element sizes (in um)
% of the desired mesh Please note that depending on the tecplot mesh and
% the desired accuracy and speed, these need to be chosen rather
% carefully.
% myvarmap - numeric (column) positions of the variables that want to
% grid (that is *which* tecplot variables to grid)
%Outputs:
% C - an array of 2D output panes. The first of which corresponds to
% the first variable specified in myvarmap, the second to the second, and
% etc...
% my_ymap and my_xmap - the x and y axis of the 2D output panes in um.
%Output files:
%contains all the outputs as well as myvarmap. saved as
%"input_name_processed.mat"
%
%IMPORTANT NOTE: THIS FILE ONLY GRIDS THE OUTPUT POINTS FROM TECPLOT ONTO
%THE GRID. IT DOES NOT FILL IN THE BLANK GRID SPACES BETWEEN THESE POINTS.
%THIS NEEDS TO BE DONE IN A DIFFERENT FILE (SEE FOR EXAMPLE:
%GRID_SMOOTHER.M)
%THIS WAS DONE TO ALLOW DIFFERENT INTERPOLATION FUNCTIONS TO BE USED IF
%DESIRED OR NECESSARY. FINALLY, WE NOTE THAT IN THE CASE OF A LARGE GRID
%IN WHCIH MANY TECPLOT VETRICES FALL INTO THE SAME SQUARE, THE VALUE OUTPUT
%FOR THAT SQUARE WILL BE A SIMPLE ARAITHMATIC AVERAGE OF THOSE POINTS, AND
%MAY NOT ACCURACTLY REPRESENT THE GEOMETRIC AREAS EACH POINT COVERS (THAT
%IS IF THERE ARE MANY POINTS IN ONE SQUARE THE WEIGHTING WILL BE OFF).
%BECAUSE OF THIS IT IS RECOMMENDED THAT THE GRID BE FINE ENOUGH THAT NO
%MORE THAN A FEW POINTS FALL INTO A SQUARE. IF THAT GRIDDING IS TOO FINE,
%I RECOMMEND MANUALLY GRIDDING THE THING DOWN **AFTER** FILLING IN
%THE FULL ARRAY.

load(filename); %get data
x_min = min(B(:,1)); %find overall size of the array
x_max = max(B(:,1));
y_min = min(B(:,2));
y_max = max(B(:,2));
x_length = x_max-x_min;
x_size = floor(x_length/x_element_size);
y_length = y_max-y_min;
y_size = floor(y_length/y_element_size);

```

```

if(x_size <=0 || y_size <=0) %get upset if anything tried to give a negative length
    error('Incorrect input signs or order');
end

my_device = zeros([y_size, x_size]); %create actual grid
                                %recall that matlab is funky and that
                                %the horizontal variable goes second
my_xmap = x_min:x_element_size:x_max; %create mapping onto actual grid
my_ymap = y_min:y_element_size:y_max; %note that point (y_min, x_min) is
                                %the upper left hand corner

%initialize output array
C = zeros([y_size, x_size, length(myvarmap)]); %actual grid times number of variables.

%then loop through the different variables:
for ijk = 1:1:length(myvarmap)
    my_repeats_map = my_device;
    my_variable_map = my_device;
    for ij = 1:1:size(B2,1) %loop through all the different squares
        %find their location on the map
        qpos(1) = B2(ij, 1);
        qpos(2) = B2(ij, 2);
        qpos(3) = B2(ij, 3);
        qpos(4) = B2(ij, 4);
        for ij2 = 1:1:4
            my_x_position(ij2) = sum(B(B2(ij, ij2),1) >= my_xmap); %number of bars
            pass tell location:
            my_y_position(ij2) = sum(B(B2(ij, ij2), 2) >= my_ymap);
            %check that location was valid
            if(B(B2(ij, ij2),1) == x_max)
                my_x_position(ij2) = my_x_position(ij2) -1;
            end
            if(B(B2(ij, ij2),2) == y_max)
                my_y_position = my_y_position -1;
            end
        end
    end
    try
        for ij3 = 1:1:4
            Q1x = my_x_position(ij3);
            Q1y = my_y_position(ij3);
            if (ij3 == 4)
                Q2x = my_x_position(1);
                Q2y = my_y_position(1);
            else
                Q2x = my_x_position(ij3+1);
                Q2y = my_y_position(ij3+1);
            end
            if(Q1x> Q2x)
                mywaystep = -1;
            else
                mywaystep = 1;
            end
            for mybloc = Q1x:mywaystep:Q2x
                if(Q1x == Q2x) %xvalue doesn't change = vertical line -- place
                    first point only.
                        my_y_here = Q1y;
                        my_variable_map(my_y_here, mybloc) =
my_variable_map(my_y_here, mybloc)+ B(qpos(ij3), myvarmap(ijk) );
                        my_repeats_map(my_y_here, mybloc) = my_repeats_map(my_y_here,
mybloc) +1;
                    else
                        my_y_here = round( ((Q2y-Q1y)/(Q2x-Q1x))*(mybloc-Q1x) + Q1y);

```

```

        my_variable_map(my_y_here, mybloc) =
my_variable_map(my_y_here, mybloc)+ B(qpos(ij3), myvarmap(ijk) );
        my_repeats_map(my_y_here, mybloc) = my_repeats_map(my_y_here,
mybloc) +1;
        end
    end
    %now catch the vertical lines
    if(Q1y> Q2y)
        mywaystep = -1;
    else
        mywaystep = 1;
    end
    for mybloc2 = Q1y:mywaystep:Q2y
        if(Q1y == Q2y) %horizontal line, assume taken care of above.
            my_x_here = Q1x;
            my_variable_map(mybloc2, my_x_here) = my_variable_map(mybloc2,
my_x_here)+ B(qpos(ij3), myvarmap(ijk) );
            my_repeats_map(mybloc2, my_x_here) = my_repeats_map(mybloc2,
my_x_here) +1;
        else
            my_x_here = round( ((Q2x-Q1x)/(Q2y-Q1y))*(mybloc2-Q1y) + Q1x);
            my_variable_map(mybloc2, my_x_here) = my_variable_map(mybloc2,
my_x_here)+ B(qpos(ij3), myvarmap(ijk) );
            my_repeats_map(mybloc2, my_x_here) = my_repeats_map(mybloc2,
my_x_here) +1;
        end
    end
    end
    %Above based off of equation for line is: (y-y1) = (y2-y1)/(x2-x1)*(x-
x1);
    end
    catch
        warning('Error placing location on map.');
```

```

    end
    warning off MATLAB:divideByZero
    C(:, :, ijk) = my_variable_map./my_repeats_map; %note that this line sets all
squares that had no input to NaN (which is good).
    warning on MATLAB:divideByZero
    %next take care of averaging pieces that need to be averaged over
    %do this in another file so that can change method easily
    %C(:, :, ijk) = grid_smoother(C(:, :, ijk));
end
filename2 = [filename(1:(end-9)), '_processed.mat'];
save(filename2, 'C', 'my_xmap', 'my_ymap', 'myvarmap');
end
end
```

```

function Cout = grid_smoother(C)
%fills in grid via weighted linear interpolation. There
%must be at least 1 non-NaN entry on each side of the grid. Interpolation
%is done linearly around the edges in a loop.
my_num_var = size(C, 3);
x_size = size(C,2);
y_size = size(C,1);
for kn = 1:1:my_num_var
    %handel edges first:
    mt = C(1, :, kn);
    mb = C(end, :, kn);
    ml = C(:, 1:kn);
    mr = C(:, end, kn);
    %start with corners: note that if corners have values, the code below
    %won't change them. %do need to weight them, though
    %top corner
```

```

l1 = min(find(~isnan(mt))); mval1 = mt(l1);
l2 = min(find(~isnan(ml))); mval2 = ml(l2);
if(mval1 == mval2) %line is flat...
    mt(1) = mval1; ml(1) = mt(1);
else
    m = (mval2 - mval1)/(l1+l2-2); b = mval2-m*(-l2+1);
    mt(1) = m*0+b; ml(1) = mt(1);
end
%bottom courner
l1 = max(find(~isnan(mb))); mval1 = mb(l1); l1 = x_size+1-l1;
l2 = max(find(~isnan(mr))); mval2 = mr(l2); l2 = y_size+1-l2;
if(mval2 == mval1) %line is flat...
    mb(end) = mval2; mr(end) = mb(end); %C(end,end, kn) = m*0+b;
else
    m = (mval1 - mval2)/(l1+l2-2); b = mval2-m*(-l2+1);
    mb(end) = m*0+b; mr(end) = mb(end); %C(end,end, kn) = m*0+b;
end
%other two corners:
l1 = min(find(~isnan(mb))); mval1 = mb(l1);
l2 = max(find(~isnan(ml))); mval2 = ml(l2); l2 = y_size+1-l2;
if(mval2 == mval1) %line is flat...
    mb(1) = mval1; ml(end) = mb(1); %C(end,1, kn) = m*0+b;
else
    m = (mval1 - mval2)/(l1+l2-2); b = mval2-m*(-l2+1);
    mb(1) = m*0+b; ml(end) = mb(1); %C(end,1, kn) = m*0+b;
end
l1 = max(find(~isnan(mt))); mval1 = mt(l1); l1 = x_size+1-l1;
l2 = min(find(~isnan(mr))); mval2 = mr(l2);
if(mval2 == mval1) %line is flat...
    mt(end) = mval1; mr(1) = mt(end); %C(1,end, kn) = m*0+b;
else
    m = (mval1 - mval2)/(l1+l2-2); b = mval2-m*(-l2+1);
    mt(end) = m*0+b; mr(1) = mt(end); %C(1,end, kn) = m*0+b;
end
for ij = 2:1:(x_size-1)
    if (isnan(mt(ij)))
        lt = min(find(~isnan(mt(ij+1:end))));
        m = (mt(ij+lt)-mt(ij-1))/(ij+lt - (ij-1));
        b = mt(ij+lt)-m*(ij+lt);
        mt(ij) = m*ij+b;
    end
    if(isnan(mb(ij)))
        lt = min(find(~isnan(mb(ij+1:end))));
        m = (mb(ij+lt)-mb(ij-1))/(ij+lt - (ij-1));
        b = mb(ij+lt)-m*(ij+lt);
        mb(ij) = m*ij+b;
    end
end
for ij = 2:1:(y_size-1)
    if (isnan(ml(ij)))
        lt = min(find(~isnan(ml(ij+1:end))));
        m = (ml(ij+lt)-ml(ij-1))/(ij+lt - (ij-1));
        b = ml(ij+lt)-m*(ij+lt);
        ml(ij) = m*ij+b;
    end
    if(isnan(mr(ij)))
        lt = min(find(~isnan(mr(ij+1:end))));
        m = (mr(ij+lt)-mr(ij-1))/(ij+lt - (ij-1));
        b = mr(ij+lt)-m*(ij+lt);
        mr(ij) = m*ij+b;
    end
end
end
C(1, :, kn) = mt;

```

```

C(end,:,kn) = mb;
C(:,1:kn) = ml;
C(:,end,kn) = mr;
for nm = 2:1:(x_size-1) %for each element
    for ml = 2:1:(y_size-1)
        if(isnan(C(ml,nm, kn))) %see if it has a value
            %squares in the middle
            myneighbors = 0;
            ij = 0;
            while ((ij>-1)&&isnan(C(ml+ij, nm, kn)))
                ij = ij+1; %can't run off the edge because edges are set to
numerical values above
            end
            m = (C(ml+ij, nm, kn)-C(ml-1, nm, kn))/(ml+ij-(ml-1));
            b = C(ml+ij, nm, kn)-m*(ml+ij);
            myneighbors = myneighbors+ (m*ml+b);
            ij = 0;
            while ((ij>-1)&&isnan(C(ml, nm+ij, kn)))
                ij = ij+1;
            end
            m = (C(ml, nm+ij, kn)-C(ml, nm-1, kn))/(nm+ij-(nm-1));
            b = C(ml, nm+ij, kn)-m*(nm+ij);

            myneighbors = myneighbors + (m*nm+b);
            C(ml, nm, kn) = myneighbors/2;
        end
    end
end
end
Cout = C;
end

```

D: Example “MIT Photonics Bands” Codes

```

;set general simulation parameters
(set-param! num-bands 3)
(set-param! resolution 16)
(set-param! mesh-size 7)
(set-param! default-material (make dielectric (index 1.445)))
(define Sil (make dielectric (index 3.4845)))

;set main period parameters:
(define-param um 1.0)
(define-param a (* 1.0 um)) ;let a = 1.0 and scale things off of a

;define k-points:
(set! k-points (list (vector3 0 0 0) (vector3 0.5 0 0)))
(define-param k-points-interp-num 100)
(set! k-points (interpolate k-points-interp-num k-points))

;set PC unit-cell sizes (height, width, length, etc...)
(define-param a_bar (* 0.3 a))
(define-param h_tot (* a .8))
(define-param w_guide (* a .5))
(define-param w_tot (* a 3))

(set! geometry (list
    (make block (center 0 0 0) (material Sil) (size a_bar w_tot h_tot))
    (make block (center 0 0 0) (material Sil) (size 1.0 w_guide h_tot))
))
;note: took advantage of new-replaces-old definitions to have fewer blocks.

;supercell parameters

```

```
(define-param scheight (* (/ h_tot a) 14) )  
(define-param scwidth (* (/ w_tot a) 3) )  
  
;define lattice  
(set! geometry-lattice (make lattice (size 1.0 scwidth scheight) (basis-size a a a) ))  
  
;define simulation and let it run  
(run-yodd display-group-velocities)
```

REFERENCES

- [1] F. Gan, High-Speed Silicon Electro-Optic Modulator for Electronic Photonic Intergrated Circuits, Ph.D. thesis, Massachusetts Institute of Technology, June 2007.
- [2] D. Marris-Morini, L. Vivien, J. M. Fédéli, E. Cassan, P. Lyan, S. Laval, "Low loss and high speed silicon optical modulator based on a lateral carrier depletion structure," *Optics Express*, Vol, 16, pp. 334-339, Jan. 2008
- [3] D. Pudo, B. Corcoran, C. Monat, M. Pelusi, D. J. Moss, B. J. Eggleton, T. P. White, L. O'Faolain, T. F. Krauss, " Nonlinear transfer function in slow light silicon photonic crystal waveguides at 10 Gbit/s," CThBB2, CLEO/IQEC 2009.
- [4] G.-R. Zhou, M. W. Geis, S. J. Spector, F. Gan, M. E. Grein, R. T. Schulein, J. S. Orcutt, J. U. Yoon, D. M. Lennon, T. M. Lyszczarz, E. P. Ippen, and F. X. Kärtner, "Effect of carrier lifetime on forward-biased silicon Mach-Zehnder modulators," *Optics Express*, Vol, 16, pp. 5218-5226, Apr. 2008.
- [5] S.J. Spector, C.M. Sorace, M.W. Geis, M.E. Grein, J.U. Yoon, T.M. Lyszczarz, E.P. Ippen, F.X. Kärtner, "Operation and Optimization of Silicon-Diode-Based Optical Modulators" Accepted to JSTQE.
- [6] L.Pavesi, D.J. Lockwood (Eds.), *Silicon Photonics*, Ed. 1, Springer, Heidelberg, Germany, c2004.
- [7] A. Khilo, Integrated Optical Analog-to-Digital Converter, SM Thesis, Massachusetts Institute of Technology, February 2008.
- [8] Y. Taur, T.H. Ning, *Fundamentals of Modern VLSI Devices*, Cambridge University Press, New York, New York, 2006 (c1998).
- [9] C.A.Barrios and M.Lipson, "Modeling and analysis of high-speed electro-optic modulation in high confinement silicon waveguides using metal-oxide-semiconductor configuration," *J. Appl. Phys.*, Vol. 96, pp.6008-6015, Dec. 2004.
- [10] R.A. Soref and B.R. Bennett, "Electrooptical Effects in Silicon," *IEEE Journal of Quantum Electronics*, Vol. QE-23, pp. 123-129, Jan. 1987.
- [11] S.R. Giguere, L. Friedman, R.A. Soref, J.P. Lorenzo, "Simulation studies of silicon electro-optic waveguide devices," *J. Appl. Phys.*, Vol. 68, pp. 4964-4970, Nov. 1990.
- [12] S. Manipatruni, Q. Xu, M. Lipson, "PINIP based high-speed high-extinction ratio micron-size silicon electro-optic modulator," *Optics Express*, Vol. 15, pp. 13035-13042, Oct. 2007.

- [13] C.A. Barrios, "Electrooptic Modulation of Multisilicon-on-Insulator Photonic Wires," *Journal of Lightwave Technology*, Vol. 24, pp. 2146-2155, May 2006.
- [14] A. Liu, R. Jones, L. Liao, D. Samara-Rubio, D. Rubin, O. Cohen, R. Nicolaescu, M. Paniccia, "A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor," *Nature*, Vol. 427, pp.615-618, Feb. 2004.
- [15] A. Liu, D. Samara-Rubio, L. Liao, and M. Paniccia, "Scaling the Modulation Bandwidth and Phase Efficiency of a Silicon Optical Modulator," *IEEE Journal of Selected Topics in Quantum Electronics*, Vol. 11, pp. 367-372, Mar-Apr. 2005.
- [16] X. Tu, S. Chen, L. Zhao, F. Sun, J. Yu, and Q. Wang, "A High-Performance Si-Based MOS Electrooptic Phase Modulator With a Shunt-Capacitor Configuration," *Journal of Lightwave Technology*, Vol. 24, pp. 1000-1007, Feb. 2006.
- [17] D.M. Szymanski, B.D. Jones, D. O'Brien, M.S. Skolnick, J.S. Roberts, A.M. Fox, T.F. Krauss, "Ultrafast All-Optical Switching in AlGaAs Photonic Crystal Waveguide Interferometers," IThA2, CLEO/IQEC 2009.
- [18] A.R.M. Zain, N.P. Johnson, M.Sorel and R.M. De La Rue, "Ultra high quality factor one dimensional photonic crystal/photonic wire micro-cavities in silicon-on-insulator (SOI)," *Optics Express*, Vol. 16, pp. 12084-12089, Aug. 2008.
- [19] S.P. Anderson and P.M. Fauchet, "Conformal P-N Junctions for Low Energy Electro-optic Switching," JTuD51, CLEO/IQEC 2009.
- [20] Q. Xu, S. Manipatruni, B. Schmidt, J. Shakya, M. Lipson, "12.5 Gbit/s carrier injection-based silicon microring silicon modulators," *Optics Express*, Vol. 15, pp.430-436, Jan. 2007.
- [21] J.D. Joannopoulos, S.G. Johnson, J.N. Winn, R.D. Meade, *Photonic Crystals: Molding the Flow of Light*, Ed.2. Princeton University Press, Princeton, NJ. c2008.
- [22] L. Gu, W. Jiang, X. Chen, L. Wang, R.T. Chen, "High speed silicon photonic crystal waveguide modulator for low voltage operation," *Appl. Phys. Lett.*, Vol. 90, Feb. 2007.
- [23] A. Liu¹, L. Liao, D. Rubin, H. Nguyen, B. Ciftcioglu, Y. Chetrit, N. Izhaky, M. Paniccia¹, "High-speed optical modulation based on carrier depletion in a silicon waveguide," *Optics Express*, Vol.15, pp. 660-668, Jan. 2007.
- [24] Wesley D. Sacher and Joyce K. S. Poon, " Designing High-Speed, Low-Chirp, Low-Distortion Microring Modulators," CTuBB4, CLEO/IQEC 2009.
- [25] M.R. Watts, W.A. Zortman, D.C. Trotter, R.W. Young, and A.L. Lentine, "Low-Voltage, Vertical-Junction, Depletion-Mode, Silicon Mach-Zehnder Modulator with Complementary Outputs," PDPC5, IPNRA/NLO/SL 2009

- [26] W. Jiang, L. Gu, X. Chen, R.T. Chen, "Photonic crystal waveguide modulators for silicon photonics: Device physics and some recent progress," *Solid-State Electronics*, Vol. 51 pp.1278–1286, Sept. 2007.
- [27] Y. Jiang, W. Jiang, L. Gu, Xiaonan Chen, R.T. Chen, "80-micron interaction length silicon photonic crystal waveguide modulator," *Appl. Phys. Lett.*, Vol. 87, Nov. 2005.
- [28] L. Liao, A. Liu, D. Rubin, J. Basak, Y. Chetrit, H. Nguyen, R. Cohen, N. Izhaky, M. Paniccia, "40 Gbit/s silicon optical modulator for highspeed applications," *Electronic Letters*, Vol. 43, No. 22, Oct. 2007.
- [29] H. Gersen, T. J. Karle, R. J. P. Engelen, W. Bogaerts, J. P. Korterik, N. F. van Hulst, T. F. Krauss, and L. Kuipers, "Real-Space Observation of Ultraslow Light in Photonic Crystal Waveguides," *PRL*, Vol. 94, Fed. 2005.
- [30] M. L. Povinelli, S.G. Johnson, J. D. Joannopoulos, "Slow-light, band-edge waveguides for tunable time delays," *Optics Express*, Vol. 13, pp. 7145-7159, Sept. 2005.
- [31] E. Kuramochi, M. Notomi, S. Mitsugi, A. Shinya, T. Tanabe, T. Watanabe, "Ultra-high- Q photonic crystal nanocavities realized by the local width modulation of a line defect," *Appl. Phys. Lett.*, Vol. 88, Jan. 2006.
- [32] P.B. Deotare, M.W. McCutcheon, I.W. Frank, M. Khan, M. Lončar, "High quality factor photonic crystal nanobeam cavities," *Appl. Phys. Lett.*, Vol. 94, Mar. 2009.
- [33] R. Amatya, Unpublished Thesis Chapter. Massachusetts Institute of Technology, Oct. 2009.
- [34] S.G. Johnson and J. D. Joannopoulos, "Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis," *Optics Express* **8**, no. 3, 173-190 (2001).
- [35] C.H. Cox III, *Analog Optical Links: Theory and Practice*. Cambridge University Press. New York, NY. 2006.
- [36] C.Sorace, A.Khilo, F.X.Kaertner, "Broadband Linear Silicon Mach-Zehnder Modulators", IWA4, Integrated Photonics Research, July 2010, Monterey, CA.
- [37] S.Deng, T.G.Neogi, J.Novak, J.McDonald, and R.Huang, "Design of a short electro-optic modulator based on SiGe HBT structure," *Opt. Express* **18**, 1994-2001 (2010)
- [38] R.H. Walden, "Analog-to-digital conversion in the early twenty-first century," *Wiley Encyclopedia of Computer Science and Engineering*, John Wiley & Sons, Inc., 2008
- [39] G.C. Valley, "Photonic analog-to-digital converters," *Opt. Express* **15**, 1955-1982 (2007)
- [40] G.T.Reed, G. Mashanovich, F.Y. Gardes and D.J. Thomson, "Silicon Optical Modulators," *Nature Photonics*, **4**, 518-526 (2010); published online: 30 July 2010.

[42] S.M. Sze, K.K. Ng, *Physics of Semiconductor Devices*, Ed. 3, John Wiley & Sons, Inc., New Jersey, USA. c2007.

[43] J. del Alamo. Lecture Notes. 6.720: Integrated Microelectronic Devices, MIT, Fall 2008.

[44] C. W. Holzwarth, J. S. Orcutt, H. Li, M. A. Popovic, V. Stojanovic, J. L. Hoyt, R. J. Ram, and H. I. Smith, "Localized Substrate Removal Technique Enabling Strong-Confinement Microphotonics in Bulk Si CMOS Processes," CThKK5, CLEO/QELSC and PAST, OSA Technical Digest (CD), 2008.

[45] Jason Orcutt. Private Communication.

[46] Synopsys, *Mesh Generation Tools User Guide*, Z-2007.03 ed., March 2007.