Integrated filters for the on-chip silicon photonics platform

A DISSERTATION PRESENTED
BY
Ian Ward Frank
TO
The School of Engineering and Applied Sciences
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy
IN THE SUBJECT OF
Applied Physics

Harvard University
Cambridge, Massachusetts
September 2013
Integrated filters for the on-chip silicon photonics platform

Abstract

We investigate the properties of integrated dielectric filters for the purposes of on-chip routing of photons. We started with the use of high quality factor tunable photonic crystal nanobeam cavities and moving on to examine a new class of reflection based reverse designed filters that maintain the footprint of a waveguide while allowing for arbitrary amplitude and phase response.

Photonic crystal nanobeam cavities are shown to exhibit high quality factors while maintaining a strong coupling to the feeder waveguide. This leads to two-fold advantages. The localized and strongly interacting light within the highly confined cavity modes allows for significant tuning to take place while only affecting a small physical area. The nanobeam cavities are also shown to have high quality factor TE and TM modes, a property that allows for interesting applications in nonlinear optics. As a case study, we examine using graphene to modulate the signal transmitted through nanobeam cavities. We show that by flooding graphene with free carriers we can electrically effect the optical properties of the nanobeam cavity.

We examine several methods for mechanical tuning of coupled photonic crystal nanobeam cavities, starting from capacitive force actuation, moving onto gradient force manipulation, and finally ending in optomechanical motion. The gradient force will facilitate high speeds and will work with insulators such as nitrides or diamond. Optomechanical actuation will display the advantage of
light-light control without electrical signals being involved, although it will suffer from disadvantages such as constant power consumption and in silicon will still be affected non-linear heating mechanisms.

Finally we will examine the theory behind, and many applications of arbitrary amplitude filters that are achieved by periodic width modulation in waveguides. We examine ways of improving the reverse design method to maintain better fidelity with target design spectra as well as examining some of the challenges in actually achieving arbitrary design such as moving toward sharper resonances. Additionally, we look at applications in ultra-fast pulse shaping and look for methods that will allow for dynamic or quasi-static applications.
Contents

1 Introduction 1

2 Photonic Crystal Nanobeam Cavities 4
   2.1 Prelude to the modern photonic crystal nanobeam cavity . . . . 6
   2.2 Increasing the quality factor(Q) . . . . . . . . . . . . . . . . . . 8
   2.3 Strong coupling between a waveguide and a cavity . . . . . . 10
   2.4 Design . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11
   2.5 Experimental characterization . . . . . . . . . . . . . . . . . . 13
   2.6 TE/TM cavities . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
   2.7 Electro-optic modulation, an application of PCNCs . . . . . . 22
   2.8 Optical properties of graphene . . . . . . . . . . . . . . . . . 22
   2.9 Electronic structure of graphene . . . . . . . . . . . . . . . . . 23
   2.10 Design considerations for a PCNC based graphene EOM . . . 25
   2.11 Static band filling . . . . . . . . . . . . . . . . . . . . . . . . . . . . 27
   2.12 An integrated graphene EOM . . . . . . . . . . . . . . . . . . 30
   2.13 Mid-Infrared . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33

3 Resonance Tuning and coupled PCNCs 34
   3.1 Mechanical deformation . . . . . . . . . . . . . . . . . . . . . . . 35
   3.2 Coupled photonic crystal nanobeam cavities . . . . . . . . . . 37
   3.3 The mechanical vibrations of coupled PCNCs . . . . . . . . . 42
4 Electro-Optomechanics of PCNCs  
4.1 Capacitive force ................................. 47  
4.2 The gradient force ............................... 56  

5 Arbitrary response filters  
5.1 Waveguide filters ............................... 77  
5.2 Improving the design for large r ............... 81  
5.3 Comparing the master equation to FDTD ........ 84  
5.4 Filter fabrication and testing ................... 86  
5.5 Application to ultra-fast pulse shaping ......... 88  
5.6 Controlling the phase of the filters ............ 90  
5.7 Overcoming the scale of width modulations .... 96  
5.8 Designing a reconfigurable filter ............... 99  
5.9 Outlook and utility .............................. 103  

6 Conclusion  

References  

vi
Author List

The following authors contributed to Chapter 2: Murray W. McCutcheon, Parag B. Deotare, Qimin Quan, Yinan Zhang, Yi Song, and Jing Kong.
The following authors contributed to Chapter 3: Parag B. Deotare, Murray W. McCutcheon, and Qimin Quan.
The following authors contributed to Chapter 4: Parag B. Deotare, Murray W. McCutcheon, Irfan Bulu, Rob Ilic, and Qimin Quan.
The following authors contributed to Chapter 5: Yinan Zhang.
## Listing of figures

2.0.1 Photonic Crystal Nanobeam Cavities ........................................... 5
2.2.1 Effect of tapering on scattering from a photonic crystal ............... 9
2.4.1 Example PCNC modes of different order .................................. 12
2.4.2 Fundamental dielectric- and air-band of a 1D photonic crystal ..... 14
2.5.1 Resonant scattering configuration. ............................................. 15
2.5.2 Resonant scattering cross section and Fano line shapes .............. 17
2.5.3 Butt coupling configuration ......................................................... 18
2.6.1 SEM Micrograph of a TE/TM PCNC along with measured spectra of the PCNC ................................................................. 20
2.6.2 TE/TM PCNC high Q modes ......................................................... 21
2.9.1 Illustration of variable absorption of graphene through the bandfilling effect. ................................................................. 24
2.10.1 Upper and lower limits for transmission contrast in a graphene EOM ....................................................................... 27
2.11.1 Transmission spectra from a device with graphene deposited. .... 28
2.11.2 Optical micrograph of a PCNC with graphene deposited. .......... 29
2.12.1 3D mock-up of a graphene based PCNC electro-optic modulator. 31
2.12.2 SEM micrographs of fabricated graphene EOMs ...................... 31
2.12.3 Mid-infrared graphene electro-optic modulator ....................... 32
3.1.1 Graphical definition of PCNC dimensions. ................................ 36
3.1.2 FEM simulation of the linear response of a doubly clamped nanobeam to an applied force. .......................................................... 37
3.1.3 Normalized strain in deformed PCNCs ................................. 38
3.2.1 Field layout and wavelength dependence of fundamental super-modes of coupled PCNCs .................................................. 39
3.2.2 Vibrational modes of photonic crystal nanobeam cavities .......... 41
3.3.1 Coupled vibrational modes of PCNCs ................................. 43
3.3.2 Bright and dark modes of coupled PCNCs ............................ 44
4.1.1 PCNC movement and optical mode shift with respect to applied voltage ................................................................. 49
4.1.2 Bending beams and supermode field profiles ........................ 50
4.1.3 SEM micrographs of fabricated, capacitively-actuated PCNCs .......... 51
4.1.4 Observed and simulated tuning of an example device .............. 52
4.1.5 SEM micrographs of capacitively actuated PCNCs showing the displacement ................................................................. 54
4.2.1 Cross-section of electric field intensity in gradient force configura-tion ................................................................. 57
4.2.2 Simulated displacements of coupled PCNCs due to the RF gra-dient force ................................................................. 58
4.2.3 Demonstration of potential switching at low power using high Q cavity modes ............................................................. 59
4.2.4 Close-up image of a gradient force device ............................. 61
4.2.5 Complete gradient force actuation device .............................. 64
4.2.6 Maxwell’s Stress Tensor .................................................... 66
4.2.7 Pump-probe configuration .................................................. 69
4.2.8 Temperature tuning of the probe-mode with high-frequency mod-ulation ................................................................. 71
4.2.9 Circuit model for temperature modulation ........................... 73
5.1.1 Comparison of TE and TM mode profiles in a 2D slab ............. 78
5.1.2 An example patterned waveguide and design process ............. 82
5.2.1 The iterative design process for waveguide filter design. . . . . . . 83
5.3.1 Comparing the 1D approximation to a full solution . . . . . . . 85
5.3.2 Comparing the 3D TE and TM filter designs . . . . . . . . . . . 86
5.4.1 Experimental configuration and measured spectra . . . . . . . . . 87
5.4.2 Measured filters based on the TM mode . . . . . . . . . . . . . 88
5.5.1 Femtosecond pulse shaping simulations . . . . . . . . . . . . . . 89
5.6.1 Simulated phase and amplitude of a filter . . . . . . . . . . . . . . 91
5.6.2 Design of an on-chip interference experiment . . . . . . . . . . 92
5.6.3 Simulations of on-chip interference . . . . . . . . . . . . . . . . 94
5.6.4 Experimental verification of on-chip interference . . . . . . . . . 95
5.7.1 Relationship between the bandwidth and width profile of our re-
verse designed filters . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 96
5.7.2 Width modulation entirely on one side of a waveguide . . . . . . . 97
5.7.3 Modulation of the evanescent field . . . . . . . . . . . . . . . . . 98
5.8.1 PIN junction based reconfigurable filter proposal . . . . . . . . . 99
5.8.2 Realistic reconfigurable filter design . . . . . . . . . . . . . . . 101
For Ken and Betty Frank, my loving and supportive parents
For Irene Frank, my favorite sister and childhood companion
And
For Rigel Patterson, my darling love.
Acknowledgments

First and foremost: thank you Marko! Marko Loncar has been a fantastic advisor throughout my time at Harvard. He was a constant source of encouragement, wisdom, and a fount of ideas to try (some of them were even good!). I owe Marko a debt of gratitude for bringing me to Harvard and giving me the resources I needed to pursue my intellectual dreams. I also would like to thank Prof. Federico Capasso, Prof. Evelyn Hu, and Prof. Donhee Ham, who have provided resources, help, and guidance along the way.

I also owe a great deal to Parag Deotare who taught me a large portion of what I know about nanofabrication and working in the cleanroom environment. We collaborated on many experiments and enjoyed many good times while commiserating through the hard times.

Thomas Babinec was a great friend to me. We never collaborated scientifically but he made my social life my first few years at Harvard, and taught me how to navigate the intricacies of being a new PhD student.

I want to thank Irfan Bulu for his knowledge and friendship. Any problem I had, he had a solution for. Sagol dostum.

Murray McCutcheon and Mughees Khan were postdocs in the group when I arrived and taught me about many things. Most of all they were role models in terms of kindness and professionalism.

Raji Shankar, Jennifer Choy, Birgit Hausmann, and Qimin Quan were my companions on a journey through a PhD. They offered support and advice, and I
will always be grateful for their presence in my life.

Yinan Zhang is indescribable, so I will not attempt to do so. I thank him for putting me on the path of an exciting new project and giving me hours of entertainment and laughter.

Everyone else in the Loncar group has been a source of inspiration and kindness at one time or another; thank you all for putting up with me! If I went through everyone in detail this section would kill too many trees. So thank you: Ian Burgess, Wallace Hui, Eiji Iwase, Motoo Aiba, Dan Floyd, Shota Kita, Cheng Wang, David Woolf, Daniel Ramos, Haig Atikian, Michael Burek, Arolyn Conwill, Vivek Venkataraman, Anna Shneidman, Ray Ng, I-Chun Huang, Srujan Meesala, and Young-Ik Sohn. My apologies to anyone I forgot to mention.

Kathleen Masse and Xiomara Forbez were so good at their jobs that I never suffered from any administrative difficulties. Thank you for making my life so easy.

Finally I must thank the CNS staff that kept the RIEs and the e-beam writers going through all these years. Without them I would have nothing to write about today.
In the last century communication was revolutionized a number of times. First, the wireless allowed long distance communications with no material connection between two parties. Second, copper was laid across the planet, allowing fast and clean communication independent of weather conditions. Third, the copper was replaced with glass. The switch from electrons to photons allowed for orders of magnitude higher bandwidths and all of the bits and bytes of the information age were able to fly around the globe in the blink of an eye. In addition to the bandwidth advantages, optical communication also resulted in a reduction of loss due to Ohmic interactions.

While long distance communications are carried over a fiber optic cable for a large proportion of the distance, the final few yards are still restricted to copper. Additionally, inside of your computer the signals are still carried on copper, and at the level of the microchips copper still dominates. This leaves a great deal of
space for improvement. At every step there are new challenges in converting the photons to electrons and back, and controlling the flow of photons.

Many decades and billions upon billions of dollars have been invested in optimizing copper and semiconductors for computation and communication. With such a head start for electrons, it is not feasible that photons will quickly displace them. However, there are a few important functions where photons could be integrated, creating a beneficial hybridization. A solid example of such a situation is optical interconnects. It is now very common to have multiple CPU cores on one chip. The communication between these cores is expensive and consumes vast amounts of energy. Photonics could be used to transfer information between the cores in an efficient, expedient manner. This does not require performing computation with the photons. All that is required is generating them, routing them to the correct core, and detecting them.

To accomplish this routing on-chip filters are required. Throughout this work we will examine possible candidates for filters that can efficiently and adjustably control the flow of light on-chip. We will start by examining photonic crystal nanobeam cavities (PCNCs). Their sharp resonances will allow selective interaction with light. We will show that not only can these PCNCs have extremely high Qs, they can be integrated into on-chip photonic circuitry seamlessly, with no little to no loss of light off-chip.

We will then examine the effect of placing graphene on a PCNC and see how an electrical signal can control the transmission properties of a PCNC coupled to a waveguide. This configuration allows for a large interaction in a small footprint and with high-quality graphene will allow modulation at fantastic speeds.

We will examine effective ways of controlling these PCNCs, focusing on mechanical motion as the control of the optical resonance. We will see both electrical and all optical methods of effecting this control and discuss their limitations and great strengths including high efficiency, excellent dynamic range, and speeds that cover the entire mechanical response range.

Finally, we will examine a new class of filters that allows for versatile customization of response in phase and amplitude. We will show that the design
is extremely computationally efficient. In addition to statically defined filters we show that it possible to achieve dynamic reconfiguration of these filters. We will also examine exciting applications such as ultra-fast pulse shaping with an eye toward quantum control experiments.
In which the author learns the value of defects.

Photonic Crystal Nanobeam Cavities

Since the Photonic Crystal Nanobeam Cavity (PCNC) is such an integral component of the work described below, it is appropriate to introduce it here by developing the concepts behind and the design, operation, and efficacy of such devices. Particular insight can be gained by examining the mechanisms that lead to strong coupling to a feeder waveguide. In addition we will examine the effect of multiple polarizations and propose a novel application by depositing graphene on a PCNC.
Figure 2.0.1: Photonic Crystal Nanobeam Cavities. a) The first PCNC (reproduced from Ref. [1]). The PCNC is coupled to a waveguide and rests on a glass substrate. b) A more advanced design for coupling to the waveguide; the holes taper out from the cavity center. Losses to the waveguide and to free space are shown. These losses are essential to interacting with the cavity; we merely need to control the dominant mechanism. c) A high Q PCNC. The substrate has been etched away so that this PCNC is suspended in vacuum. It is not coupled to a waveguide and is intended to be characterized using free-space optics.
2.1 Prelude to the Modern Photonic Crystal Nanobeam Cavity

In condensed matter physics, it has long been known that the periodic arrangement of atoms in crystals or quasi-crystals leads to the development of band structure for electrons[2]. Researchers soon discovered that a periodic arrangement of materials at the wavelength scale of light also led to the development of Bloch modes, band-gaps, and pass-bands for photons[3]. In three dimensional structures a high level of index contrast is required to open up a “complete band-gap” or a continuum of energies where photons are not allowed to exist inside the photonic material[4]. After the three dimensional stage, there was a search for more integrable alternatives that were easier to fabricate. The result was 2D slab photonic crystals[5, 6] and 1D PCNCs[7–9]. These structures maintain a periodic structure in a smaller number of dimensions but use the natural confinement of index contrast in the remaining dimensions[4].

Total internal reflection is a well studied phenomenon. Light incident from a high index medium will be refracted at the interface with a lower index material. Snell’s law determines that refraction will pull light away from the normal to this interface. Based on the ratio of the indices of refraction we can determine an angle at which no light is transmitted into the low index medium. At angles of incidence greater than this critical angle all of the light will be reflected back into the high index material as if the interface is a perfect mirror[10]. While in most nanoscale structures the geometry is smaller than the wavelength so that this simple ray optics explanation is no longer valid, the fundamental property of the high index material confining the light is still observable[11]. Of course there is a limit to this confinement: when given a dimension becomes less than $\lambda/2n$ - the so called “diffraction limit” - the light light “leaks out”, with the evanescent tails the dominant location for the energy. However, as long as the diffraction limit is respected, 1D PCNCs and 2D slab photonic crystals can be used to develop extremely strong partial band-gaps. The PCNC distinguishes itself in this respect. Because we only care about the $\Gamma - X$ portion of any Brillouin zone, a photonic
band-gap can occur regardless of the index contrast in the periodic perturbation\(^1\).

The next step is to construct the “cavity” part of the PCNC. A perfectly periodic structure along the waveguide will introduce a band-gap or stop band that will not allow light of certain frequencies to propagate along the waveguide. By introducing a defect, or a change in the periodicity of the structure, we can create a place where light that would not otherwise be allowed to exist in the periodic region can be trapped, and in some cases quite effectively.

The most primitive form of such a defect is shown in Fig. 2.0.1a. In this case the photonic crystal is formed by circular holes etched in the waveguide. The six holes visible are divided into two groups by an increased gap at the center. The periodic arrangement of holes provide the band-gap while the defect at the center provides the cavity. Examining this device we can see why the PCNC is so useful: the footprint is precisely that of a blank waveguide. Given a certain density of waveguides available for an on-chip integrated light circuit, PCNCs can be added while incurring no cost in terms of required real-estate.

The problem with this simple cavity design lay in the extremely low quality factor (Q) of the photonic cavity\(^1\). As with all resonant systems, we can characterize the efficiency of our PCNCs in terms of the ratio of the amount of energy lost by the cavity to the amount of energy stored in the cavity over one optical cycle (period). This Q tells us how many cycles are required for the energy stored in the cavity to reach a factor of \(e^{-1}\). This first PCNC had an extremely low Q of only 265. Because of the Fourier relationship between time and frequency, this meant that it also interacted with a large band of frequencies adjacent to the central resonant frequency. The essential rule in this case is that the more frequencies that a cavity can trap, the lower the Q. While it is not always advantageous to have an enormous Q, as we will see later, a Q of at least 10,000 is generally desired.

\(^1\)This band-gap may be extremely weak.
2.2 Increasing the Quality Factor (Q)

In order to understand why the cavity in 2.0.1a does not exhibit a high quality factor we need not look further than the defects in the fabrication. Clearly, there is significant edge roughness, some stigmation in the circularity of the holes, and a change in the width of the waveguide where the holes are written\(^2\). However, even a flawless fabrication of the same design would not have a high Q. The reason is a mode mismatch between the pure waveguide mode and the Bloch modes of the photonic crystal.

Let us consider how light that is launched into a waveguide reacts when it encounters a band-gap. Fig. 2.2.1a shows that a considerable amount of scattering occurs. But if the light is only gradually introduced to the photonic crystal, which we can effect by “tapering” it in with smaller holes, then we see that this scattering loss can be eliminated to a large extent. An adiabatic transition is almost always favorable when passing from one medium to another; photonics is no different. A study of the cross sections of the mode in the waveguide and the photonic crystal (the Bloch mode) shows that when the overlap is the smallest there is the highest scattering loss\[^7\]. Therefore, the taper can be viewed as gradually altering the cross-section of the mode so that the overlap mismatch never becomes significant.

Introducing this tapering has created a nearly ideal mirror. In order to create a trap for light we can place a mirror on either side of the region where we want to trap the light. This can induce the light to bounce back and forth between the two mirrors with little chance of escaping. The quality of the mirrors will determine just how long this process continues. Because our tapering allows us to create such high quality mirrors, by placing them on a waveguide we can create a highly effective cavity.

When such a taper is added to the defect, it allows us to create extremely high Qs. For silicon in air we expect Qs of greater than \(10^7\)\[^12\], while in lower index

\(^2\)Indicative of proximity effects in a positive electron-beam lithography resist
Figure 2.2.1: Effect of tapering on scattering from photonic crystal. In (a), (b), and (c), a mode is launched from the left toward a photonic crystal. Because of the band-gap, this light must either scatter or reflect. A cross section of the mode in the waveguide and the Bloch mode of the first period of the photonic crystal are compared for each configuration. a) No taper introduced, large levels of scattered light visible, the comparison of the modes shows significant mismatch. c) A three-hole taper has been introduced, resulting in a significant decrease in the level of scattered light. d) A five-hole taper is used, and virtually no scattered light is visible. The cross-section shows that the two modes are nearly spatially identical.
materials like Si$_3$N$_4$ Qs greater than $10^6$ have been designed. These numbers are generally not replicated experimentally because of fabrication defects. As the quality of fabrication increases higher and higher Qs are being demonstrated. We have demonstrated a Q of 750,000 in silicon.

2.3 Strong coupling between a waveguide and a cavity

Even though tapered designs allow for nearly arbitrarily high Qs in PCNCs, our work is not complete. We can trap light, but we do not have an effective way of getting the light in and out of the trap. Picture, once again, what the light “sees” as it propagates down a unperturbed waveguide and suddenly runs into a region where it cannot propagate. As we previously explained, there is a large mode mismatch, and a great proportion of the light scatters. If we include a well designed taper of holes, the light will be cleanly reflected without scattering. This reflection will span the band-gap of our photonic crystal.

Because our resonance lies inside the photonic band-gap we would assume that no light is able to pass from the waveguide to the cavity. Interestingly, if there is a cavity where the light can exist, there will be a resonant tunnelling effect in which the light will hop to the cavity, stay there for a certain number of cycles, and then hop to the other side of the cavity (such as in Fig. 2.0.1b, represented by the light yellow arrows). Therefore we add a taper similar to what we saw when defining the cavity but on the outsides. This matches the impedance and mode profile of the waveguide to the cavity and allows us to reach the “strong coupling” regime where the amount of light that escapes the cavity through the waveguide channel is far greater than the light escaping into free space.

Strong coupling requires not just the coupling of light from the waveguide to the PCNC, but also that this channel of coupling be the dominant form of coupling. There are many mechanisms for light to enter or exit a PCNC. Fig. 2.0.1b shows two of them. The scattering loss is out of the waveguide into free-space or the substrate. Another possibility is absorption loss where the material of the PCNC absorbs photons to generate excitons or phonons. In most
cases it is our goal to limit these non-waveguide loss channels. If we can accomplish this to a significant degree, any photon entering the cavity from the waveguide will also exit to the waveguide. Monitoring the transmission through such a cavity will show that at resonance the PCNC transmits close to 100% of the light incident upon it.

In most cases 100% transmission is not achieved. This becomes especially true at high Qs. As the waveguide losses are limited in order to achieve a higher Q, the scattering and material losses need to be lowered commensurately to maintain strong coupling. The result is that at a Q of 1,000 strong coupling is trivial, but at a Q of 10,000,000 it becomes virtually impossible.

2.4 Design

Many parameters go into the design of a PCNC. Depending on the required application we need to determine the wavelength, the number of resonances, the polarization of the resonances, the size, and the Q. Because running parameter sweeps with 3D optical simulations is computationally expensive, many semi-analytical methods have been developed[14, 15]. Of particular interest is the approach in Quan et al. [15] as it guarantees strong coupling to the waveguide and offers a choice of locating the energy of the cavity in the “air band” or the “dielectric band” (shown in Fig. 2.4.2). As their name’s suggest the dielectric band locates the field inside of the high index dielectric whereas the air band locates a significant amount of the mode in the low index material. Depending on a specific application we may want to interact more or less strongly with our surroundings (for example, in sensing experiments) or the material that forms the PCNC (for example, when probing the material properties).

In order to determine a cavity’s resonance, according to Quan et al., we start with a material and thickness, usually a predetermined factor. For instance, we often use 220 nm thick Silicon. We then choose a width — often determined by the “single-mode limit”[11]. Then a hole-radius and periodicity are used to find
Figure 2.4.1: When the dielectric band is pulled up into the band-gap, a harmonic potential is created, and as such, different orders of mode exist and can be characterized by their symmetry about the dotted mirror line above.

a) The first three TE modes of a PCNC strongly coupled to a waveguide.

b) An example of the quantum mechanical solutions to the harmonic potential; the plotted wavefunctions match the envelopes of the modes from (a) Image is adapted from http://en.wikipedia.org/wiki/File:HarmOsziFunktionen.png.
the band-edge modes. We want to cause the frequency of either the air- or
dielectric- bands at the edge of the Brillouin zone to be as close as possible to the
desired resonance. This process can be very fast as it only requires calculating a
the band-edge of a single unit cell of an otherwise infinite photonic crystal.

Once we have found a suitable hole size, then it is a simple matter of either
pulling the dielectric band “up” into the band-gap, or an air band “down” into the
band-gap through an appropriate taper. Fig. 2.4.2 illustrates this concept while
Fig. 2.4.1 shows the resulting resonances. In the case of dielectric band this is
achieved by reducing the square-root of the hole radius linearly as you move away
from the center of the PCNC (tapering down). If an extremely high Q is required
the holes can tapered down to nothing over a large distance. In practice the actual
length of the cavity need not exceed 30 μm.

2.5 Experimental characterization

In our experiments we used two separate methods for characterizing the
performance of our PCNCs. The first involved coupling into the cavities from
free-space using a high numerical aperture (NA) objective lens. This method is
difficult, because high Q cavities are difficult to design so that their far-field
radiation patterns overlap with near-normal incidence light\([16]\) (up to a 30 angle
with the normal for 0.55 NA). However, by making use of a cross-polarization
setup where only the light that is coupled into the cavity passes through an
analyzing polarizer, the signal to noise ratio can be raised enough to characterize
PCNCs, even at Qs approaching \(10^6\)\([12]\). The configuration is shown in Fig.
2.5.1.

The process is as follows. A tunable semiconductor laser outputs light, and an
input polarizer is used to set-up a reference polarization. The light is passed
through a half-wave-plate (HWP) to rotate the polarization by 45°. This places
the orientation of the electric field at a 45°angle with the major field component
of the PCNC (\(E_y\) in Fig. 2.4.1). At this polarization there is a 50% chance that a
Figure 2.4.2: The band diagram for the fundamental TE air- and dielectric-bands of a 1D photonic crystal. A nanobeam cavity is formed by pulling the dielectric-band up or the air-band down into the band-gap. The air-band has most of the field localized in the holes, while the dielectric-band localizes the field in the high index dielectric of the nanobeam. The major field component, $E_y$, is plotted. The direction of propagation is $x$. 
Figure 2.5.1: Resonant scattering configuration a) A schematic of the optical setup. (OBJ: microscope objective, HWP: half-wave plate, BS: beamsplitter, P1 and P2: polarizers, D: detector) b) 3D representation of PCNCs suspended above a substrate with light coming down from normal incidence at a 45°polarization with respect to the axis of the PCNCs.
photon can enter the cavity (this is an additional factor in the coupling efficiency so the total probability of coupling is much less). When the photon decays from the cavity, it passes back up into the objective lens and through the HWP again. This rotates the polarization by -45°, and then the light is then sent through a beam-splitter to an analyzing polarizer, which is cross-polarized with respect to the reference. Only the light which was coupled into and subsequently re-emitted from the cavity has a polarization that matches the analyzer and is transmitted\(^{[17, 18]}\). This greatly reduces the amount of purely scattered light that does not interact with the cavity mode that reaches the photodetector, thereby making it much easier to observe the resonance, despite the small amount of light able to couple into the cavity through the free-space channel.

While linear optical resonances generally exhibit Lorentzian line shapes, in resonant scattering more care needs to be taken when characterizing the resonances. We do not always observe the Lorentzian shape of the resonances that we would expect from a PCNC. The reason is that the substrate under the PCNC affects the measurement. Despite under-cutting the SiO\(_2\), there is still bulk Si substrate \(2\ \mu m\) below the PCNCs. The incoming signal reflects off this back-plane. Additionally there is light scattering off the PCNCs that did not couple to the cavity resonance under analysis. These reflections form a continuum of states. This is in contrast to the cavity resonance that only exists for a very small range of frequencies. The coupling between these two mechanisms forms a Fano resonance\(^{[19–22]}\), which can be fit to the form of:

\[
F(\omega) = A_0 + F_0 \left( \frac{q + \frac{1}{2} (\omega - \omega_0)/\Gamma}{1 + \left( \frac{1}{2} (\omega - \omega_0)/\Gamma \right)^2} \right)
\]

where \(A_0\) and \(F_0\) are constant parameters, \(\omega_0\) is the resonance frequency, and \(\Gamma\) is the resonance line-width. \(q\) is a dimensionless parameter that describes the relative strength of the resonant cavity interaction and the non-resonant standing wave reflecting off the substrate\(^{[22]}\). If \(|q|\) is very small then the non-resonant scattering is dominant and the line shape will be a negative Lorentzian. If \(|q|\) is near unity the strengths of the two processes are similar and a highly asymmetric
Figure 2.5.2: The processes behind Fano line shapes. a) Fano line shapes in different regimes of resonant and non-resonant scattering strengths. At \( q = 0 \) the non-resonant process dominates; at \( q = 1 \) the strengths are equal; and at \( q = 100 \) the resonant process dominates. The resonances have been shifted for clarity. b) Resonant scattering cross-section of a coupled PCNC excited at normal incidence. The field reflecting from the back-plane is significant and contributes to Fano lineshapes in the reflected signal.

lineshape emerges, and finally, if \( |q| \) is very large then the resonant process dominates and the usual Lorentzian emerges. The three different regimes are plotted in Fig. 2.5.2a.

The second method again uses a tunable semiconductor laser; however, in this case the PCNCs are strongly coupled to a waveguide. Light is brought onto the waveguide using a tapered, lensed fiber. These lensed fibers generate a spot size around 1.7 \( \mu \)m. In order to match this spot size, large polymer waveguides are patterned with a 3 x 3 \( \mu \)m cross-section. The lensed fibers are butt-coupled to the facets of these on-chip polymer waveguides. With the matching of the physical size this can be an extremely efficient coupling method[15, 23]. The method is visualized in Fig. 2.5.3. On-chip, the light is transferred from the large polymer waveguide to smaller, high-index-contrast dielectric waveguides through an inverse tapering scheme as detailed in Fig. 2.5.3. This scheme has the advantage of much higher signal-to-noise ratios, as well as easy integration of fiber components such as polarization rotators, circulators, and so forth, that are
Figure 2.5.3: Butt coupled PCNCs. a) A schematic of the setup including fibers, polymer spot-size-converters and PCNCs coupled to waveguides. b) A close up of the transition from the large polymer waveguide to the smaller silicon waveguide. c) A close-up SEM of the PCNC with the tapering outside the cavity to allow for strong coupling to the waveguide. d) An optical micrograph showing the polymer and silicon waveguides. The PCNCs can just be seen as a discoloration. e) An example transmission spectrum from a PCNC. A high Q cavity mode as well as band edge modes can be seen.

Inexpensive and plentiful in the near infrared (NIR) regime. It has the additional benefit of easily probing cavity modes of different polarizations, whereas resonant scattering is generally limited to TE modes.

2.6 TE/TM Cavities

It is instructive to consider the polarization of our light on-chip. Up to this point we have focused exclusively on so called Transverse Electric (TE) light. Given the nature of SOI waveguides it is worth considering the definition of Transverse Electric and Transverse Magnetic. In the case of hollow metal waveguides the
definition is given by the two linearly independent modes where the electric field is completely transverse to the direction of propagation of light in the case of TE, and the magnetic field is completely transverse to the direction of propagation in the case of TM. However, in highly confined dielectric modes these definitions become more blurred. Any mode calculated for such a waveguide will have both electric and magnetic field components in the propagation direction.

We therefore apply the definition of TE-like and TM-like modes. The definition is based upon the slab plane in which our PCNCs are defined. Therefore if the dominant electric field vector is in the plane of the substrate we will call it TE (the dominant magnetic field will be perpendicular to the substrate), and if the dominant electric field is perpendicular to the plane of the substrate will call this TM (the dominant magnetic field will be parallel to the substrate).

When designing our PCNCs we generally choose to interact either with the TE or TM mode. For a given choice of thickness, width, periodicity, and hole radius the TE and TM band structures are generally very different and the band gaps do not overlap. However, we can find a subspace of parameters where highly confined and high Q modes can exist for both TE and TM cavities\([24]\). This is in sharp contrast with 2D slab photonic crystals where the TM band-gap tends to close near the M-point, preventing a complete band-gap\([4]\). This is not true for periodic arrangements of high index materials such as posts. But for the case of a high index material with periodic holes the 1D photonic crystal is superior in this respect.

Using the design methodology from Ref. [15] we can come up with a cavity exhibiting high Q modes with both TE and TM polarization. Some care needs to be taken with choosing a thickness of the silicon in order to have the band-gaps to overlap. Fig. 2.6.1b shows simulated TM modes from such a structure. Figs. 2.6.1a,c show an SEM of a fabricated PCNC and its measured transmission spectra, with the TE and TM spectra represented in two different scans.
Figure 2.6.1: a) SEM micrograph of a PCNC that supports simultaneous TE and TM modes. It is somewhat thicker than other PCNCs mentioned in this work (500 nm vs. 220 nm), because the TM mode is more sensitive to thickness than the width. b) $E_z$ - the out of plane component of the electric field for the first three TM modes of the cavity. c) Measured spectra show multiple modes in both the TE and TM bands, and the high Q modes exhibit very poor coupling ($T \ll 1$) to the waveguide. The TE band-gap is blue-shifted with respect to the TM band-gap. By altering the parameters of the PCNC this relationship can be altered.
Figure 2.6.2: Transmission scans of fundamental modes of a TE/TM PCNC. In this case the TM mode exhibits higher Q but worse transmission.

TE/TM cavities tend to have clean transmission spectra. One of the main sources of noise in a cavity that is resting on a substrate is TE/TM coupling due to the asymmetric index of the cladding (air above oxide below)\(^{25}\). Because of the simultaneous band-gap for the TE and TM modes this noise is almost completely disappears. However, resting on the oxide substrate does lower the Q. Simulations suggest that undercutting these PCNCs could more than double the already excellent Q values. Fig. 2.6.2 shows detailed scans of the fundamental mode of each polarization. These Q values are already higher than our previous record\(^{26}\), once again, this is without undercutting.

The ability to probe these two different polarizations is significant. Materials such as graphene are theoretically predicted to have different responses to electric fields that are applied parallel to the plane of the graphene and perpendicular to it\(^{27-29}\). These cavities also enable a wide range of nonlinear optics. There are proposals to use TE/TM cavities for nonlinear frequency conversion\(^{30, 31}\). TE/TM PCNCs also allow for interaction with phonons such as in Raman scattering\(^{32}\), an important property for building Raman lasers\(^{33}\). Finally, these devices could have applications in quantum cascade lasers, which emit TM polarized light\(^{34, 35}\).
2.7 Electro-optic modulation, an application of PCNCs

A PCNC measured in transmission can have sharp peaks of transmitted wavelengths in the middle of a band-gap where no light is transmitted. Depending on the design of the PCNC the width of this peak and the level of transmission is controllable. Because of the high degree of localization of the light in the PCNC we are able to achieve strong interaction between any material in that vicinity — including the material of the PCNC — and this light. This means that even small disturbances can lead to large changes in transmission for a specific wavelength.

If PCNCs can perturbed electrically they can become an extremely low power, narrow-band electro-optic modulator (EOM). As a specific example we will examine the use of graphene as an electrically controlled absorber that will be used to turn the transmission of a sharp resonance peak on and off. It is also possible to shift the value of the resonance peak by altering the real part of the refractive index of the PCNC as will be discussed in the next chapter. Here, we will examine the effects of adding a variable absorptive layer and thereby controlling the imaginary refractive index.

2.8 Optical properties of graphene

Graphene, a two-dimensional hexagonal array of carbon atoms is the target of a great deal of experimental study. Its unique electrical[36], mechanical[37, 38], and optical properties[39, 40] make it attractive across many different fields of study. We are most attracted to its fascinating elecro-optical interactions. Most of these arise due to graphene’s linear band structure around the \(K\)-point[41]. The optical loss in graphene is constant across a broad spectrum[42]. Every wavelength is absorbed due the lack of a band-gap. However, at normal incidence this absorption is limited to approximately 2.3%, because the one-atom thick layer is has a vanishingly small absorption cross-section. When the light is no longer at normal incidence to the graphene, but is instead shifted into the plane of
the graphene we can exponentially increase this interaction length[43].

Recently, by utilizing a stretch of waveguide coated with graphene, scientists have been able to make extremely high speed EOMs[44]. In this case graphene is a worthy choice because high carrier mobilities can lead to truly fantastic speeds of operation. However, we could make the footprint even smaller by creating the same interaction length, with a shorter “real” length, through the use of a PCNC. In this case there is a “gain-bandwidth” trade off, where if the cavity has too high of a Q, the EOM will not be switched on or off quickly enough as the load/unload time will lag behind any applied modulation. The Q of the cavity will also limit the frequencies which can be controlled, perhaps the more significant bandwidth trade-off.

2.9 Electronic structure of graphene

The crystal structure of graphene is simple enough that the band-structure can be fully calculated. The band-structure gives us some interesting qualitative intuition. As a semi-metal, there is no band-gap and the valence band is completely full while the conduction band is entirely empty. This means that any photon incident on the graphene can be absorbed by promoting an electron to the conduction band. This is true regardless of the energy of the photon. However, this excitonic absorption can be turned on and off electronically[45]. By emptying the valence band or filling the conduction band, the absorption of photons of certain frequencies can be turned off. Fig. 2.9.1 shows how the shift in the Fermi level can lead to a total loss of excitonic absorption up to a level of \( \Delta \varepsilon_f / 2^3 \).

Because of the energy shifts required it will be easier to induce such a change at

\[ \Delta \varepsilon_f / 2^3. \]

\[^3\text{A curious reader might ask why the electron has to be promoted from } -\hbar \omega / 2 \text{ to } \hbar \omega / 2, \text{ and not, for instance, from } -4\hbar \omega / 5 \text{ to } 4\hbar \omega / 5 \text{ thereby making the cut-off for } \varepsilon_{	ext{fill}} \text{ the full photon energy. The reason is that the electrons are confined to the white lines in Fig. 2.9.1, and based on where they are on that line they have a specific } k \text{ or momentum. Since light carries virtually no momentum when compared to electrons, the light can only mediate an electron straight up and down; given the shape of the band structure this leads to the factor of two.}\]
Figure 2.9.1: Illustration of variable absorption of graphene through the band-filling effect. a) The natural state of graphene: the valence band is completely full (red), and every state in the conduction band is ready to receive an electron. Any photon with energy $\hbar \omega$ can be absorbed by creating an exciton. b) Electrons have been injected into the graphene filling the conduction band up to the energy level $\epsilon_{\text{fill}}$; photons with energy less than $2\epsilon_{\text{fill}}$ can no longer be absorbed by creating an exciton as there are no empty conduction band states available. c) Electrons have been removed from the graphene leaving the valence band empty down to a level of $-\epsilon_{\text{fill}}$; photons with energy less than $2\epsilon_{\text{fill}}$ can no longer be absorbed by creating an exciton as there are no electrons available to be promoted to the conduction band states within reach.
longer wavelengths. Current technology makes it possible in the NIR but visible energies are too high. Mid-infrared frequencies are even better and have the added benefit of not interacting as strongly with the optical phonons in graphene. In particular at wavelengths longer than $6 \, \mu m$ the optical phonon loss is completely turned off, this means that turning excitonic loss on and off leads to much higher contrast.

2.10 Design considerations for a PCNC based graphene EOM

When operating at $1.5 \, \mu m$ we can take advantage of a growing number of publications that have characterized the optical losses due graphene deposited on-top of photonic crystal cavities[46, 47] and SOI waveguides[43]. This data is extremely useful because graphene — especially graphene grown by chemical vapor deposition (CVD) — has a wide range of quality. This makes it difficult to quantify exactly how much optical loss to expect, as well as the contrast expected from an electronic on/off switch.

By taking minimum loss and maximum loss from literature we can develop a window to target with the design of our PCNCs. Because we want to operate at NIR frequencies optical phonon loss is always significant, even if we can completely eliminate excitonic absorption[48]. In reported results that do not use a cavity (i.e. just a straight waveguide) some contrast was observed, but the loss even in the “on” state was significant[43, 44].

As we will now show, this low-contrast loss means the high Q PCNCs must be close to strongly coupled to the waveguide in order to show high contrast switching behavior. It is possible to have a high Q cavity with bad transmission properties and the converse is true. However, a normalized transmission measurement gives us the relative amount of loss into the waveguide:

$$T = \left( \frac{Q_{total}}{Q_{wave}} \right)^2 \quad (2.2)$$

The total $Q$ is given by the FWHM measurement of the transmitted
Lorentzian. As the Q of a cavity is related to the inverse of the loss rate, we add the loss rates together to get a total Q as follows:

\[
\frac{1}{Q_{\text{total}}} = \frac{1}{Q_{\text{wave}}} + \frac{1}{Q_{\text{scatter}}} + \frac{1}{Q_{\text{material}}} + \ldots
\]  \hspace{1cm} (2.3)

\(Q_{\text{scatter}}\) is determined by the loss rate into free-space due to scattering and \(Q_{\text{material}}\) is related to the absorption loss in the material. \(Q_{\text{material}}\) is considered infinite in silicon at NIR wavelengths (i.e. no loss), but once graphene is added to the system it will become finite and contribute to the total Q. If we are attempting to observe a change in transmission from a cavity, such as the fundamental mode from Fig. 2.6.1, we can see that a considerable change in the \(Q_{\text{material}}\) will be needed:

\[
T = \left( \frac{\frac{1}{Q_{\text{wave}}} + \frac{1}{Q_{\text{scatter}}} + \frac{1}{Q_{\text{material}}}}{Q_{\text{wave}}} \right)^{-1}
\]  \hspace{1cm} (2.4)

This is because \(Q_{\text{scatter}}\) is already much smaller than \(Q_{\text{wave}}\). If we estimate \(Q_{\text{material}}\) based on waveguide loss figures from [43] we can see with our specific PCNCs why there will be no visible change in the “on” and “off” states of graphene in our system. In order to observe a difference two possible tacks are possible. First, a cavity with much better transmission properties can be designed. Second, the interaction of graphene and light can be altered such that the phononic loss mechanisms become less significant. This will changed by either getting a better quality of CVD graphene, or attempting to work at longer wavelengths. Fig. 2.10.1 shows a numerical analysis of the normalized transmission contrast as a function of both the intrinsic Q and intrinsic transmission properties of a PCNC.

Based on the calculations show in Fig. 2.10.1 we know that the high Q cavities, such as from the one shown in Fig. 2.6.1, are not suitable because their on-off contrast would be vanishingly small. However, because of the high number of lower Q, higher transmission peak as the wavelength approaches the band edge.
Figure 2.10.1: Upper and lower limits for transmission contrast in a graphene EOM. There is a sweet spot of maximum transmission contrast where \( Q_{\text{material}} \) due to the graphene is comparable to the intrinsic \( Q \). The reason for this is easily inferable from Eq. 2.4. a) Estimates based on low quality CVD graphene. The material \( Q \) is taken as 136 in the off-state, and 182 in the on-state[43]. b) Higher quality CVD graphene projects to have much better characteristics. \( Q \) is taken as 650 in the off-state and 1800 in the on-state. These numbers were extracted from[46, 47] with some adjustment for different modal overlaps in the 2D and 1D photonic crystals.

they are still worth exploring.

2.1.1 Static Band Filling

The Fermi level of graphene is rarely exactly at the Dirac point (the state from 2.9.1a). When it is in contact with any material that does not have precisely the same work function as itself, there will be an equilibration at the interface between the two materials depending on the relationship of their respective work functions. The graphene will either be doped p or n. In the case of p doping this equivalent to removing electrons (the state from 2.9.1c), whereas n doping has the opposite effect. This means that depending on the level of the band filling the excitonic absorption will be off up to a certain frequency of light[49].

We can see this when we transfer graphene onto our TE/TM cavities from Fig. 2.6.1. The graphene is grown by chemical vapor deposition (CVD) on top of
Figure 2.11.1: Transmission spectra from a device with graphene deposited. 
a) Before the supporting PMMA layer is removed. The PMMA acts a natural 
dopant and fills the bands of the graphene. This allows some light to avoid 
absorption and optical cavity modes are visible, though greatly degraded when 
compared to the clear cavity in Fig. 2.6.1. b) The PMMA has been removed 
and virtually all traces of the cavity spectrum have disappeared. By injecting 
a lot of light we are able to see the pass-band of the PCNC as well as some 
band-edge modes, but no localized modes can be distinguished.
Figure 2.11.2: Depositing graphene on a PCNC results in the optical modes disappearing in the transmission spectrum. a) An optical micrograph of a TE/TM PCNC with a patterned strip of graphene overlaying it. The graphene is just visible due the interference effects with the substrate. b) A transmission spectrum of the TE mode of the PCNC, with a “before and after” comparison. While graphene is expected to blue-shift the mode, this scan range would cover any such change.

copper foil. PMMA is then spun onto the graphene side of the foil, and the foil is chemically etched away. This leaves the graphene attached to a layer of PMMA and allows mechanical manipulation without damaging or losing the graphene. The graphene, supported by the PMMA, is then deposited on-top of our bare TE/TM cavities, if the PMMA is left on, then it will significantly dope the graphene. Fig. 2.11.1 shows spectra with the PMMA still on, and then after the PMMA is removed.

The PMMA fills the bands of the graphene in the areas where it is in contact with the graphene. This effect is reversible and is due to the discrepancy in the work function of the PMMA and the graphene. In order to equalize the Fermi level across the graphene-PMMA interface, charge flows into the graphene, stopping the excitonic absorption of IR light. With the PMMA doping the graphene we can still make out cavity modes, although their Q is diminished. In this way we see the effect of the optical phonon loss despite having limited or turned off the excitonic loss mechanism.

After the PMMA is removed we can pattern the graphene using
photolithography to mask the areas we want to keep, and burning the rest off with an oxygen plasma. Fig. 2.11.2a shows an optical micrograph of a PCNC with a patterned strip of graphene running over it. The graphene is visible due to interference effects although its contrast is low. A before and after scan of the device, Fig. 2.11.2b, reveals that the high Q mode is completely damped out by the graphene.

2.12 AN INTEGRATED GRAPHENE EOM

If we want to move beyond the binary choice of either having PMMA on the graphene or not, we need to find a way to dope the graphene dynamically. One of the best integrated technologies for this doping is based upon the idea of the metal-oxide-semiconductor field-effect-transistor (MOSFET), where a thin layer of oxide acts to create a capacitor. Here the semiconductor layer is simply the silicon in the TE/TM cavity. The graphene plays the role of the metal layer, and oxide is deposited between the silicon PCNC and the graphene. In the area of overlap between the silicon and the graphene, the graphene will be doped when a voltage is applied between the graphene and the silicon. Fig. 2.12.1 shows a cartoon of a proposed configuration.

The fabrication itself is a long and involved process. The TE/TM cavities are patterned in 500 nm thick SOI in a similar manner to other PCNCs. However, in this instance the etch is not performed fully, leaving a 30 - 50 nm layer of silicon behind. This thin layer is then selectively removed, leaving behind a “shelf” connected to the PCNC that allows for electrical contact to be made. After the residual etch mask is removed electrical contacts are put down by lift-off.

Atomic layer deposition (ALD) is used to deposit a dielectric oxide such as AlₓOᵧ at low temperature, so that photoresist can be used as a lift-off mechanism. Graphene is transferred to the substrate and the PMMA removed in a gentle acetone vapor etch; this protects the graphene. Then the graphene is patterned using photolithography, removing the excess graphene not covering the PCNC.
Figure 2.12.1: 3D mock-up of a graphene based PCNC electro-optic modulator. A PCNC is strongly coupled to a waveguide, and in electrical contact with a “gate.” The PCNC is isolated from the graphene by a thin layer of oxide. By applying a potential between the PCNC and the graphene the band can be filled or emptied.

Figure 2.12.2: SEM micrographs of fabricated graphene EOMs. Graphene is visible as a darker layer, while the gate-oxide has slightly less contrast.
**Figure 2.12.3:** Below: a mockup of the electrical contact to the graphene and photonic crystal. Above: 100 kHz trace of modulation in the reflectivity of the cavity.

While the final device seems to be complete in the SEM images (see Fig. 2.12.2), we have been unable to observe contrast between the “on” and “off” states where the voltage is turned on and off. This is likely due to the unsuitability of the specific cavity’s transmission properties as discussed in Sec. 2.10. In the future further improvements in design, such as lower Q and higher T, as well as creating a more symmetric cladding, may lead to vastly improved performance. However, there is one more consideration. Because the doping in the graphene is due to a capacitor forming between the graphene and the silicon, the graphene will only be doped in the areas immediately above the silicon. This doping effect is highly local and above the holes that form the PCNC, there will still be lossy areas\[^{[50]}\]. The use of the dielectric mode should limit the loss from this
inconsistency, but it should be kept in mind as a possible reason for device failure.

Note that people have observed EOM effects in silicon cavities with CVD graphene, however, they used ionic gels to achieve much higher levels of shifts in the Fermi level\cite{46,47}. This was achieved in reflection, not transmission, so the limits in detection are much more relaxed. Additionally, instead of an integrated, high-speed carrier injection, ionic gels were used to gate the graphene. This leads to very high levels of doping but is slow. Depending on the type of gel used it could be limited to DC operation or a few $kHz$. While an excellent first step, these devices are not an integrated solution.

2.13 Mid-Infrared

With a similar configuration of graphene and silicon we have examined the effects of graphene at 4.4 $\mu m$. Instead of a PCNC, a 2D L3 cavity was used. The characterization was performed in a resonant scattering setup, and instead of transmission a change in the reflection was monitored. Fig 2.12.3 shows a schematic of the device as well as a 100 $kHz$ modulation measured as a voltage was applied. While this experiment was performed in resonant scattering the technology to move this to a completely integrated configuration is well within reach. Working at MIR frequencies with graphene is attractive due the diminished phononic losses and this may well be a future avenue.
We will now explore some methods for tuning the resonances of PCNCs. We ended the last chapter by proposing a PCNC and graphene based EOM. The operating mechanism for such a device is based on changing the absorptivity of the cavity. This is equivalent to controlling the imaginary part of the refractive index. We will now consider the effects of changing the real part of the refractive index. This will lead to changes in the resonance frequency of the PCNCs without inducing absorptive losses. In particular we will focus on mechanically deforming PCNCs. We will see that it is advantageous to study a system of coupled PCNCs. We will then discuss some of the mechanical properties and the overlap between the optical and mechanical degrees of freedom.

There are a number of reasons why we may want to tune the frequencies at which our PCNCs resonate. We may have been targeting a certain wavelength
and due to fabrication tolerances have missed that window. We may want to actively react to different wavelengths over time. There are a number of ways we can accomplish this tuning. We could, for instance, run liquids with different indexes of refraction over the cavity\cite{51} and in that way control the wavelength. However, there are more local and elegant methods of accomplishing this effect.

For instance, a PN junction could be used to inject carriers into the cavity region\cite{52}. Free-carrier dispersion would then shift the refractive index of the nanobeam material, leading to a shift in the resonance. Another method would be to use either the absorption of light or an Ohmic heater to locally heat the material, which through the thermo-optic effect would also lead to a shift in the resonance\cite{53-55}. Both of these methods have the advantage of operating in an integrated manner, where electrical signals control the optical resonances. However, they have the disadvantage of constantly consuming power. One of the reasons for making the transition from electronics to photonics is to lose these ohmic losses.

One method that would not consume power would be to selectively remove some of the material of the nanobeam. With silicon this could be achieved by oxidizing the surface and using a selective etch to remove this oxidation. This process can be used to shift the resonance of a PCNC finely in what is called “digital tuning.” While this maintains a zero power consumption after tuning, it suffers from being irreversible\cite{56, 57}.

If we are seeking a low power method of tuning that is dynamic and fully reversible then mechanical deformation is worthy of consideration.

3.1 Mechanical deformation

Mechanical deformation is a natural avenue when the geometry of the PCNC is considered. As in our resonant scattering setup, when the beam is undercut it is already the quintessential MEMS device: a doubly clamped beam. Because of the small size and mass, a minuscule amount of force is required to displace the center of the beam significantly.
A doubly clamped beam can be modeled as a simple spring obeying Hook's law. In this case the 3D deformation tensor must be modeled by a single parameter. For this purpose we can consider the one dimensional motion of the beam’s center point. By applying a force along the direction of deformation (perpendicular to the axis of the beam) we can see a linear relationship between the displacement of the beam and the applied force (Fig. 3.1.1), as confirmed in Finite Element Analysis (FEM, COMSOL Multiphysics) simulations. The slope of this line gives us an effective spring constant from Hooke’s Law:

$$ F = -k_{eff} \chi $$ \hspace{1cm} (3.1)

The $k_{eff}$ of the spring constant will be highly dependent upon the profile of the applied force. A “point force” applied directly at the center will lead to a much lower $k_{eff}$ than the same magnitude of force applied along the whole length of the beam. Therefore, depending on the way in which a force is applied we need to be careful in the parametrization of the motion.

Unfortunately, we can see in Fig. 3.1.3 that in a doubly clamped beam all of the strain due to deformation occurs adjacent to the clamping points. The means that while displacement the center of the beam, where the cavity is located, does elongate the beam, none of this elongation takes place where it would be useful. One alternative would be to suspend the beam above the substrate, but still close enough that the PCNC’s modes are dependent on the separation. Then through

---

**Figure 3.1.1**: Graphical definition of PCNC dimensions. $L$: length. $w$: width. $t$: height or thickness.
Figure 3.1.2: FEM simulation of the linear response of a doubly clamped nanobeam to an applied force. Dimensions are given in Fig. 3.1.1

various contrivances, we could move the beam closer to or further from the substrate. This change in the refractive index profile of the mode will result in tuning. However, yet another alternative exists.

3.2 Coupled photonic crystal nanobeam cavities

If two PCNCs are brought together in parallel, and the gap between them is small enough that their evanescent fields can interact, then the system will not be described by two independent optical resonances; the resonances will instead couple, leading to supermodes that are even and odd about the mirror-symmetry plane as seen in Fig. 3.2.1. The even mode will have a lower energy (red-shifted) while the odd-mode will be pushed to higher energy (blue-shifted)\[8\]. The magnitude of the separation should increase as the beams get closer together, as this increases the coupling strength of the modes. To test this theory we fabricated pairs of PCNCs that were separated by varying distances. When we tracked the modes we were surprised to find that while the even mode did red-shift considerably, the odd-mode never moved much lower than the value at

\[k_{\text{eff}} = 4.5289 \text{ N/m}\]

\[w = 450 \text{ nm}\]

\[t = 220 \text{ nm}\]

\[L = 15 \mu m\]
Figure 3.1.3: A colormap of normalized strain in PCNCs. a) A PCNC that is deformed “in-plane”. One edge experiences strain, but the cavity length at the center is unaltered. b) A side view of a PCNC deformed “out-of-plane.” c) Top view of the out-of-plane deformed PCNC. There is virtually no deformation at the center of the beam.
Figure 3.2.1: Images of coupled PCNCs and the field layout of their fundamental supermodes. a) SEM micrograph of fabricated coupled beams. b) Resonant wavelength of fundamental supermodes as a function of separation (“d”). c) Close up of center showing the tapered cavity region. d) Even supermode with field strongly localized in the gap. e) Odd supermodes with node in the gap.

infinite separation. This trend was confirmed in both FDTD and FEM simulations as can be seen in Fig. 3.2.1.

From an examination of the mode profile this behaviour makes a certain amount of sense. There is a large field component in the even mode that lies between the two beams. This means that the mode is highly sensitive to perturbations in the gap, whereas the odd mode has almost all of its field components inside the beams, which makes it less susceptible to any change in the relative position of the beams. An expression for the change in resonance
frequency for a device as a function of dielectric index perturbation is given by

\[
\Delta \omega = \frac{\omega}{2} \int d^3r \Delta \varepsilon(r) |E(r)|^2 + O(\Delta \varepsilon^2) \tag{3.2}
\]

where \( \omega \) is the angular frequency, \( \varepsilon(r) \) is the permittivity of the region where the cavity is located and \( \Delta \varepsilon(r) \) is the small perturbation to this permittivity, caused by the motion of the beams, for example. By examining the overlap between \( \Delta \varepsilon(r) \) and \( |E(r)|^2 \), we can quickly see that the even mode should be affected by movement while the odd mode should remain unperturbed. Eq. 3.2 also shows why a single beam is difficult to tune. At small deformations each \( \Delta \varepsilon(r) \) on one side of the PCNC is exactly cancelled out by a corresponding \( \Delta \varepsilon(r) \) on the other side of the beam.

Another way to think about the process for the double beams is through Coupling Induced Frequency shift\[59]. The idea is that as the beams move closer together there are two effects. The increasing coupling strength is blue-shifting the odd mode while red-shifting the even mode, yet at the same time when we examine the mode of each individual beam (the components that form the supermodes), there is a large \( \Delta \varepsilon(r) \) due to the presence of the second beam. This means that each individual beam is being red-shifted regardless of the coupling between the cavities. We happen to be operating in a regime where the magnitude of the two effects is essentially equal. This means that the odd-mode is essentially not tuned whereas the even mode is tuned twice as much compared to tuning from coupling effects alone.

While we have shown that pairs of PCNCs that are lithographically defined to be at a given separation do follow the trend from Fig. 3.2.1, this is not reversible dynamic tuning. In the next few chapters we will discuss several methods for achieving this dynamically controlled separation.
Figure 3.2.2: Some of the vibrational modes of a suspended, doubly-clamped PCNC. The colormap is for normalized displacement as a function of position. a) The first two “in-plane” modes, as defined by the plane of the substrate. The second-order mode is marked by the node at the center. b) The out-of-plane modes. In this case their frequency is lower than the in-plane modes because this PCNC is wider than it is thick. c) The first two torsional modes of this PCNC. The beam is twisting along its axis. d) The fundamental “phononic mode”, at these frequencies there is negligible displacement and interestingly the periodic structure that leads to the photonic band-gap can also contain these mechanical vibrations [60].
3.3 **The mechanical vibrations of coupled PCNCs**

In addition to the static measurement of coupled PCNCs, we can also consider how the beams vibrate. As doubly-clamped resonators they have a number of different transverse, torsional, and “phononic” modes[60]. These can be seen in Fig. 3.2.2. These modes increase in frequency with the fundamental transverse modes having frequencies in the $MHz$, the torsional modes in the $100s$ of $MHz$ and the breathing modes reaching into the $GHz$. This range of frequencies means that the transverse vibrations are the largest in amplitude and therefore interact much more strongly than the higher frequency modes. In order to see the effect of the higher frequency modes more and more sensitive equipment is required. As a first order approximation we can consider only the transverse vibrations. Because of the geometry of our PCNCs we can consider the in-plane and out-of-plane vibrations separately.

Similar to coupled optical modes, if there is a connection between the mechanical vibrations they will also form supermodes. This connection could be material, such as in Figs. 3.2.1 and 3.3.1, or it could be mediated by the optical modes.¹ We therefore see mechanical modes as displayed in Fig. 3.3.1. If the motion of the mode leads to a change in the separation of the beams this is called a “bright mode” because it will be discernible in the optical signal. If there is no differential movement, then it will be difficult to discern the motion in the optical signal and these vibrational modes are called “dark modes.” We will discuss one method of observing these modes in the following section.

3.3.1 **Observing transverse vibrations**

When PCNCs are free-standing they have a number of different transverse vibrational modes. At a finite temperature these modes will all be active due to random thermal fluctuations. Therefore, if they have any “brightness,” they will

¹As we will discuss in Chapter 4 the dispersion of an optical mode with respect to the separation of the PCNCs leads to a mechanical force.
Figure 3.3.1: A few of the coupled transverse vibration modes of PCNCs. In this case the coupling is due a physical connection between the two beams. a) The even (+) fundamental mode for both the in- and out-of-plane modes. With virtually no differential motion between the beams these modes are dark. b) The odd (-) fundamental mode for the in- and out-of-plane modes. The mode in the top row is generally the brightest mode of a coupled PCNC because it directly changes the gap at the center of the beams length. c) Some examples of in- and out-of-plane modes of higher order. Depending on how they overlap with the optical mode, these can be bright (second row) or dark (first row).
Figure 3.3.2: Bright and dark modes of coupled PCNCs. A tunable laser is set at point “A” (inset), the largest slope on the transmitted signal. Due to the high Q and sensitivity to motion the tiny vibrations with non-zero $g_{om}$ can be seen when the photodetector signal is fed into a spectrum analyzer. NLT refers to non-linear transduction and is a harmonic of IP1 because the change induced by IP1 is so large that it is no longer a linear perturbation. a) The odd mechanical mode has a large effect on the even optical mode and can be seen very clearly on the transduced signal. b) The even mechanical mode has no differential motion and therefore is not visible on the transduced signal.

be leaving a small but measurable imprint on the optical signal transmitting through the PCNCs. If we place our tunable laser at the maximum slope of the Lorentzian (Point A in the inset of Fig. 3.3.2), and plug the output of a high-bandwidth photodetector into a spectrum analyzer, we see the modulation at the frequencies corresponding to the bright, transverse modes.

Fig. 3.3.2 shows an example spectrum as well as the fundamental, simulated, mechanical mode shapes that were matched to the spectrum. Fig. 3.3.1 shows a larger selection of modes labeled with the same nomenclature. The process of simulating the vibrational modes lead to the conclusion that our SOI device layer is compressively stressed. Matching the correct peaks required simulating with a built-in stress. Confirmation of the compressive stress in our SOI devices comes from observed mechanical motion in the following chapter. The amplitude of the peaks in the spectrum are based on two things: first the amount of movement
random thermal fluctuations impart, and second, the $g_{om}$ associated with each mode.

For small displacements we know that our PCNCs act as an ideal spring. Therefore they have the following relationship between the frequency and spring constant of a given mode:

\[
f = \frac{1}{2\pi} \sqrt{\frac{k_{\text{eff}}}{m_{\text{eff}}}}
\]

\[
k_{\text{eff}} \propto f^2
\]

$m_{\text{eff}}$ is the effective mass of the vibrational mode. This means that the spring constant increases quadratically with frequency. We also know that random thermal fluctuations impart an energy determined by the equipartition theorem \[61\]:

\[
E_{\text{dof}} = \frac{1}{2} k_B T
\]

$E_{\text{dof}}$ is the energy per quadratic degree of freedom (in our case a single vibrational mode modeled as a spring with an attached mass which has one quadratic degree of freedom), $k_B$ is Boltzmann’s constant and $T$ is the absolute temperature. This energy is the same for every mode, so based on the spring constant we can expect enough motion ($\Delta x$) to satisfy the following equation:

\[
\frac{1}{2} k_B T = \frac{1}{2} k_{\text{eff}} \Delta x^2
\]

\[
\Delta x = \sqrt{\frac{k_B T}{k_{\text{eff}}}}
\]

\[
\Delta x \propto \frac{1}{f}
\]

Where in the last step we substituted the relationship Eq. 3.4. So at a fixed
temperature we expect the motion in vibrational modes to fall off as $1/\nu$. On the other hand, the $g_{om}$ is entirely dependent on the overlap between the optical mode shape of the PCNC and the deformation profile of the vibrational mode. So based on our knowledge of $g_{om}$s we can match up the peaks we see in Fig. 3.3.2 to simulations of the transverse vibrations.
There are two main sections to this chapter. First we will examine a capacitive method for inducing motion in PCNCs. We will see that this is an extremely efficient method that is particularly applicable to silicon based devices. Second we will examine a similar geometry that makes use of the gradient force which loses a measure of its efficiency but allows for a generalization to insulating materials. In this second section we will examine applying reconfiguring forces applied as voltages as well as through the optical modes of the PCNCs.

4.1 Capacitive Force

Capacitive force is a common method for actuation at micro- and nano-scales. The origin of the force can be explained by examining the energy dependence of
a parallel plate capacitor\[^{[62]}\]:

\[
U = \frac{1}{2} CV^2 = \frac{\varepsilon_r \varepsilon_0 AV^2}{2d} = \frac{\varepsilon_r \varepsilon_0 AV^2}{2} (x_0 - x)^{-1} \tag{4.1}
\]

\(x_0\) is the initial separation and \(x\) is the differential movement of the capacitor plates (toward each other is increasing \(x\)). The resulting force can be calculated using the gradient of the potential\[^{1}\]. By limiting ourselves to one dimension we have:

\[
F = -\nabla \frac{1}{2} CV^2 = -\frac{\varepsilon_r \varepsilon_0 AV^2}{2} (x_0 - x)^{-2} \tag{4.2}
\]

From the negative sign we infer that the force is attractive. We can also see that as the plates move closer together the force increases. This leads to highly favorable operating conditions because devices with small gaps can be actuated with low power. Another element of note is that the force is quadratically dependent on the voltage. This means that a small time varying signal on-top of an applied DC voltage will be magnified proportional to the DC voltage\[^{[37, 63]}\]:

\[
F \approx -\frac{1}{2} C' (V_{DC})^2 + C' V_{DC} V_{AC} \tag{4.3}
\]

for the sake of clarity we have collected the capacitance terms into \(C'\), denoting the positional derivative of the capacitance\[^{2}\].

As discussed in Chapter 1 it is advantageous to mechanically tune coupled PCNCs. The tuning dependent on the gap between two parallel beams is significant not only due to supermode splitting, but also due to the coupling induced frequency shift (CIFS). When casting the problem in terms of mechanical actuation with capacitative forces, we see that the extremely small gaps where the even mode is highly dispersive is also where the most force can be obtained with the least voltage. The challenges of smaller gaps will be discussed in Sec. 4.2.3.

\[^{1}\text{In the long time limit where the capacitor charge is constant, the total energy is equivalent to the potential energy}\]

\[^{2}\text{A simple application of a Taylor expansion yields this expression.}\]
Armed with our knowledge of $k_{ef}$ and the capacitive force expected for different applied voltages, we can quickly extract the voltages required to move the beams a useful amount. It turns out that small voltages are required. In an easy to fabricate geometry we simulate that just $\Delta V$ should lead to over $1\mu m$ of tuning. One of the major strengths of this method is that once the beams have reconfigured, no power is required to keep them in position because the force is electrostatic. For the high Q cavities that we are capable of fabricating, $1\mu m$ of tuning is over 100 line-widths; this means that with a $\Delta V$ dynamic range potentially 100 channels could addressed.

Fig. 4.1.1 shows the simulated change in gap as well as the tuning of the even mode as a function of applied voltage. This significant tuning at sub-$100 \, nm$ gaps is one of the reasons that this method is so attractive. The reasons for the highly non-linear behavior has been discussed in the previous paragraphs. Simulations are carried out in COMSOL multi-physics, which allows for complete coupling

---

3This is why Qs in the range of 10,000 are useful: they allow a large number of channels.
Figure 4.1.2: Simulated deformation of beams under capacitive forces. Field profiles are the $E_y$ component. The actual displacement is exaggerated for clarity of direction and shape of the deformation.

of the electrostatic, mechanical, and even the optical eigenfrequency calculations. The calculations are self-consistent, meaning that changes in the geometry are fed back into the calculation to update the forces. This is necessary for the capacitive force because this force increases as the beams move closer together. Fig. 4.1.1 shows how the resonance of the even mode evolves, taking into account the deformed geometry due to the applied voltage.\(^4\)

\(^4\)There is very little difference between deforming two PCNCs so that their center's move closer together, and moving them closer without deforming them. This is because they they are much longer than the distance they are displaced, so curvature is minimal.
4.1.1 Device fabrication and characterization

These devices were fabricated and tested. Figure 4.1.3 shows an SEM of an example device. As can be seen, the two PCNCs are electrically isolated, but are extremely close to each other. Gaps as low as 50 nm were measured. The fabrication process involved several steps. Starting with 220 nm SOI (SOITEC), electron-beam (e-beam) lithography was used to pattern a mask in Flow-able Oxide (Corning FOX-16). After a high-contrast develop in 25% tetra-methyl ammonium hydroxide (TMAH), the mask was transferred into the Silicon device layer through inductively coupled plasma reactive ion etching (ICP-RIE). The remaining etch mask was removed with buffered oxide etch, and metal contacts were thermally evaporated onto the chip. After dicing the chip a HF vapor etching process was used to undercut the devices. Finally, the chip was attached to a chip carrier and wire bonded at low-power to prevent damage to the PCNCs.

Figure 4.1.4 shows the resulting tuning observed when a voltage was applied across the beams. We see that the observed tuning tracks the theoretical predictions remarkably well. In addition to observing the tuning by monitoring the resonance frequencies of the even and odd modes, we were also able to
Figure 4.1.4: Observed and simulated tuning of the fundamental even and odd optical modes of our devices. The simulations were carried out based on dimensions measured, post-fabrication, in the SEM. There is a small difference in the absolute value of the experimental and simulated resonances, but the differential behaviour is well reproduced.

confirm the motion by placing the devices inside of a SEM. By using an electrical feed-through we applied a driving voltage to the beams while imaging them. This allowed us to visually confirm the nanometer scale motion of the beams as a function of the applied voltage. Fig. 4.1.5b shows some stills of the beams moving closer together.

4.1.2 Analyzing the mechanics

In generating the simulations shown in Fig. 4.1.4, we noticed a considerable discrepancy in that we experimentally observed significantly higher tuning at a given voltage than coupled FEM simulations predicted. All of the physical parameters such the PCNC separation, actual beam widths, thicknesses, suspended length, hole size and position were confirmed to within a few nm by
high resolution electron microscopy. We were able to conclude that there was significant compressive stress in the device layer of our SOI. Through parameter sweeps of uniaxial compressive stress we able to determine a value of 30 MPa compressive stress that fit the experimentally determined behavior. Analysis of the same batch of SOI later yielded similar results[64].

The compressive stress led to larger than expected tuning because stress affects the $k_{\text{eff}}$ of a doubly clamped beam as follows[38]:

$$k_{\text{eff}} = \frac{AEw^3}{L^3} + \frac{BT}{L}$$

(4.4)

A and B are geometry dependent constants, $w$, $t$, and $L$ are the dimensions of the beam as defined in Fig. 3.1.1, $E$ is the Young’s Modulus of the material, and $T$ is the applied stress. Compressive stress is taken to be negative; thus a compressively stressed beam will have a lower $k_{\text{eff}}$ than an unstressed beam and therefore move more than expected for a given applied force.

4.1.3 LIMITS

While this actuation is extremely low power and has the advantage of working well with silicon, there are limits to its usefulness. First of all, it is hard to generalize to materials other than semiconductors. Any insulator such as Diamond or Si$_3$N$_4$ cannot be used as a capacitor, and therefore this kind of force would be inefficient at best. Efforts have been made to overcome such problems by depositing a low loss conductor on top of the insulator[65], but such actions have added complexity in fabrication and optical purity.

The second problem is the speed of charging the capacitor. The electrons need to be pushed through the silicon, which with considerable resistance leads to ohmic losses. In our experiments the silicon had a resistivity of 10 $\Omega \cdot cm$. This leads a resistance of $10^7$ $\Omega$. Even with the tiny capacitance inherent in the coupled beams this limits the reaction times to above a ns. This can be mitigated

---

5The presence of compressive stress was confirmed in Sec. 3.3.1 where we observed the resonance frequencies of suspended PCNCs
Figure 4.1.5: SEM micrographs of fabricated capacitively actuated PCNCs showing the displacement. The top figure shows a “pinched-off” device where the electrical attractive force has overcome the mechanical restorative force completely, and the PCNCs have become stuck together through Van der Waals interactions. The close-ups show a reduction in 67 nm gap as a function of applied voltage. The image contrast for the upper and lower beams is altered with applied voltage. The potential difference leads to a different secondary electron signature.
by doping the silicon, but there is a limit to the level of doping before the optical losses induced become a limiting factor in the device. Doping the silicon to a level of $10^{18} \text{ cm}^{-3}$ reduces the resistivity by two orders of magnitude while still allowing optical Qs as high as $10^5$ \cite{66}, a value high enough for most applications. If the resistance obtained from doping the semiconductor is still not low enough a switch to the actuation methods of the next section may be required.

The final limit is speed. The RC time constant of the device can be reduced to the point where the limiting factor becomes mechanical response, through the use of doping. However, the silicon beams are just not stiff enough to be able to operate at high frequencies. A good measure of their operating speed is $10 \text{ MHz}$. Beyond this frequency the response will "roll-off" at $20 \text{ dB}$ per decade or $6 \text{ dB}$ per octave (for instance a tenfold increase in frequency will produce a 100 times smaller response for the same input power or a doubling of the frequency will produce a 4 times smaller response\textsuperscript{6}). This means that as frequency is pushed higher and higher there are diminishing returns on the mechanical motion. While measures such as shortening the suspended length of the beams and making them wider can be taken to increase the stiffness, this also results in reduced mechanical motion and therefore less tuning of the optical mode.

4.1.4 Outlook of Capacitive Actuation

It is instructive to compare the potential performance of capacitive actuation to the state of the art in other optical modulators. Using graphene, charge injection in silicon waveguides, and silicon MOSFET devices modulation speeds of tens to hundreds of GHz at extremely low power is available \cite{44, 67, 68}. Opto-mechanical effects will never be able to compete in terms of speed with these devices. The mechanical roll-off means that significant displacement cannot occur at these frequencies while keeping the power consumption low.

Another effect to keep in mind is that the mechanical tuning is altering the real refractive index, whereas other examples of high speed modulators are changing

\textsuperscript{6}A an inverse quadratic relationship
the imaginary refractive index and inducing absorption. This means that if the mechanical modulator is not transmitting the signal is being reflected back. Many optical systems will be sensitive to back reflected light, so this would not be an ideal setup for our mechanical filters either\(^7\).

When the sum of the strengths and weaknesses of the capacitive actuation of mechanical tuning is taken we can conclude that these are potentially incredibly useful in a variety of applications, including adjusting the operational wavelength to compensate for fabrication tolerances, drifts due to temperature changes, or configuring optical circuits that do not need to be continuously altered. The low power, high dynamic range nature of the operation is simply not matched by any other dynamic method of resonance tuning. Additionally, it should be stressed again that after reconfiguration there is zero power consumption, so a given optical circuit could be held without constantly dissipating power.

### 4.2 The Gradient Force

Although the mechanical limitations cannot be overcome in a meaningful way within the paradigm of mechanical tuning, we can address the first two issues of material choice and ohmic losses in loading the capacitor.

If, instead of charging the PCNCs directly to induce a capacitive force, we could create a field distribution around the beams that would induce them to pull closer together or push further apart (as seen in Fig. 4.2.1), we could decouple the driving mechanism from the beams themselves. Because the beams are made from dielectrics, they become polarized when located in an externally applied electric field. This polarization is anti-aligned to the applied field, so the system’s energy is minimized when the field overlap with the dielectric is maximized\(^[10]\).

\(^7\)Future quantum computation gates that require no loss of information to operate may not be compatible with absorptive modulators and mechanical reconfiguration may need to be considered for these applications
Figure 4.2.1: Cross-section of electric field intensity in gradient force applied by adjacent electrodes. The high-field intensity between the beams leads to an attractive force in FEM simulations.

This leads to the following force (called the gradient force):

\[ F \propto -\nabla |E(r)|^2 \] (4.5)

We can further generalize this force by considering the effect of the optical modes without an applied external field. We have already extensively discussed the idea of causing a PCNC cavity to redshift by mechanically moving the beams together. This is a clear indication that the closer configuration of the beams constitutes a lower energy state of the optical system. Therefore, a sufficient amount of light in the optical modes will shift the beams toward each other. Obviously, this will be counteracted by restorative force that builds up in the elastic beams, but given enough power we should see mechanical motion.

In the context of Eq. 4.5 we can think of the field distribution of the even supermode and see that the gradient of the field pulls the beams toward the center of the gap where the field is maximized. On the other hand, the odd mode does not have a field maximum in the gap and therefore there is no attractive force pushing to close the gap.
Figure 4.2.2: Simulated displacements of coupled PCNCs due to the DC gradient force. The simulations are carried out self-consistently, with the parameters updating as the PCNCs move. a) Change in gap as a function of applied voltage as well as detuning of the even supermode. $k_{\text{eff}}$ can be used to determine the applied force. b) A perspective view; parts as labeled. Wire-frame outline shows unperturbed position. c) A top view of the displaced PCNCs. The white areas show the high field from to the applied voltage. Of interest is the sliver of white in the gap between the beams. $G$ is the distance between the metal contacts and the PCNCs. $g$ is the gap between then PCNCs. Both affect the total gradient force.

4.2.1 Static and RF gradient forces

Inducing the gradient force through applied voltages has the advantage of extremely low power operation and allows us to efficiently combine electrical and optical signals. As discussed in Sec. 4.1.3 the PCNCs are able to effectively respond to signals on the order of $10 \text{ MHz}$. After this limit it becomes very difficult to effect mechanical changes. So signals up to this level are able to dynamically control the optical resonant frequencies of the coupled PCNCs. If
Figure 4.2.3: Demonstration of potential switching at low voltages using high Q cavity modes. The red line represents light at a single frequency (such as a CW laser). The optical Q is 10,000. The shift in the optical resonance is based on FEM simulation for a coupled PCNC gap of 100 nm and a metal contact gap of 750 nm.

we than to go to higher frequencies than this, then there will be a trade-off between stiffness and mechanical response.

When we calculate typical forces in FEM simulations we find that they are significant yet small compared to capacitive forces. Fig. 4.2.2 shows FEM simulations of the gradient force. In the capacitive configuration we were seeing 10 nm of tuning at less than 5 V, but now we are down to 1 nm at 80 V. Yet, given the sensitivity of the coupled PCNCs to their separation, there is a meaningful amount of movement at reasonable voltages. Fig. 4.2.3 shows how these small changes can be used with a high Q cavity to control whether a signal is transmitted, creating an electrically controlled switch for an optical signal.

4.2.2 Fabrication of Gradient-Force Tunable Filters

The challenge for implementing this scheme lies in the difficult fabrication. In contrast to the capacitive actuation scheme there are two extra properties that considerably complicate the fabrication. First, by moving toward complete integrated operation, the design was changed from resonant scattering to strong waveguide coupling. This also simplifies measurement. On the other hand it also

8 100-1000 times smaller for similar geometries at the same voltages
adds fabrication steps with the definition of polymer waveguides, dicing clean facets, and forcing a switch from HF vapor etching to wet etching followed by critical point drying. Second, the metal contacts need to be much closer to the PCNCs. The requirement for the resulting small $G$ (as defined in Fig. 4.2.2) requires much more precise alignment and lift-off.

A step by step fabrication procedure follows. E-beam lithography is used to pattern an etch mask in Flowable-Oxide (FOx). After a high contrast development, the pattern for the PCNCs, feeder waveguides, and silicon support pads for the metal contacts are transferred to the device layer of 220 nm thick SOI using an ICP RIE etch. Following the RIE a short etch in 100:1 H$_2$O : 49% HF was used to remove the remaining etch mask without undercutting the thermal-oxide (BOX) layer. This selective removal of the remaining FOx is essential for the later undercutting steps. The 100:1 dilution is necessary because the inverse-tapers of the silicon waveguides are susceptible to be bent and broken if they are undercut at all.

Metal contacts were defined using e-beam lithography. In contrast to the the capacitive actuation where photolithography was used, e-beam lithography allows for the precise alignment required for sub-micron $G$ lengths. 200-400 nm of gold with a chrome or titanium adhesion layer was thermally evaporated and lifted-off. Then SU8 polymer waveguides were deposited, and the chip diced. The dicing process yields clean facets at the chip edge allowing high-efficiency butt-coupling. To this point we have generated an almost complete device, except that the PCNCs are not suspended. Ideally the undercutting would be performed by an HF vapor etch because of the small gaps and large elasticity of silicon. After a regular wet-etch the sample needs to be dried. This drying process can be highly destructive at the nanoscale because of the surface tension of the evaporating liquids. In the resonant scattering configuration we were able to overcome this through a vapor-phase etch, where no liquid surface exists to cause the close PCNCs to stick to each other. Now that we want to have feeder waveguides on-chip the vapor phase etch becomes unreasonable. The vapor of
Figure 4.2.4: Close-up image of a gradient force device. Metal contacts are slightly misaligned but within tolerance.
HF is so aggressive that it is very difficult to deposit a mask that would both protect the on-chip optical circuitry and not interfere with the propagating light once the undercut is complete.

In order to achieve a clean undercut we are forced to use a wet-etch. 7:1 buffered oxide etch (BOE) allows for a gentle undercut of the BOX without attacking the silicon. Additionally, photoresist can be used to mask off the areas of the chip that we do not want undercut. We can open up a small window over the PCNCs and the rest of the chip will be very well protected. At this point it is essential that we removed the remaining FOx, otherwise the BOE will etch along the waveguides where the FOx gives it a path, thereby making the undercut much longer than expected. If it is too long the beams will not be stiff enough to withstand the next step.

After the BOE undercut the sample is transferred — without ever allowing it to dry — to a series of solvent baths. This removes the photoresist and prepares the sample for critical point drying (CPD). The final solvent soak is for several hours in Ethanol. Ethanol is used in the critical point dryer as an initial medium. It is important to soak the sample in ethanol because small traces of other liquids can be trapped in the PCNCs and this will degrade the effectiveness of the CPD process.

The CPD process starts by slowly flushing out the ethanol (keeping the sample wet) by replacing it with liquid CO₂ at high pressures. The two liquids are imiscible and flooding the chamber of the CPD completely removes all traces of ethanol. When the sample is completely flushed in liquid CO₂ a process of heat and pressure cycles is used to maneuver the CO₂ around its critical point so that there is a state change from liquid to gas without ever crossing a phase change line. This eliminates any kind of surface tension and allows the coupled PCNCs to remain separate and free standing.

Fig. 4.2.4 and Fig. 4.2.5 show SEMs of various finished devices. The samples were characterized by a tunable semi-conductor laser in the butt-coupling configuration. A probe station was constructed to make contact on-chip, and up to 300 V was applied across the contacts. Over 10 chips with measurable optical
resonances were fabricated. On this chips there were 5-10 devices that appeared to be free standing and have gaps (g) of 100 nm or less. Based on the G values of 750 nm to 2 μm simulations suggest that this should have led to appreciable tuning in the observable optical modes, but none was ever observed.

It is likely that the complicated fabrication led to small discrepancies that are not easily visible in SEM images. For instance, it is possible that in the small gaps the RIE did not completely release the two beams and they are held together by an extremely thin layer of silicon that is penetrated by the SEM beam and therefore does not show up on images. This would stop any differential actuation.

The largest difficulty in perfecting these devices lay in iterating the fabrication process. Over the course of two years we successfully fabricated — where success is defined by the devices passing visual inspection in the SEM — 10 chips. There were at least that many unsuccessful attempts. This slog meant that it was and is very difficult to get to the bottom of why exactly we never observed motion.

One of the best improvements that can be made is to decrease the gap size between the PCNCs. Through the CPD process we were able to reliably make gaps of around 100 nm. Lengths under this were prone to stiction. At gaps of 120 nm or more the force can change sign and become repulsive. Given fabrication tolerances a 100 nm gaps is on the limit of acceptable gaps. Decreasing the gap has the dual advantages of increasing the force at a given voltage and increasing the optical tuning for a given force. A new fabrication technique that pre-undercuts the SOI is allowing our lab to reliably create gaps in the 60-80 nm range. This is allowing us to resurrect this project and may produce some results moving forward.

4.2.3 The challenges of fabricating small gaps

The limiting factor in fabricating our electro and optomechanical devices is the smallest gaps that can be obtained. Using the HF vapor etching technique we are able to obtain 50 nm gaps, while critical point drying limits our fabrication to 70 nm gaps. Smaller gaps are lithographically possible with e-beam techniques but
**Figure 4.2.5:** A complete view of a gradient force actuation device. SU8 polymer waveguides act as spot-size converters for lensed fibers. Light is transferred to silicon waveguides through an inverse taper. The symmetric curved sections act as single mode couplers between the single waveguide and the double waveguide section where the coupled PCNCs are. Two isolated contacts provide the field profile necessary for creating an attractive force in the PCNC.
stiction effects lead sticking when random fluctuations of charge in the air create enough force to stick the beams together\cite{69, 70}. There are passive and active techniques for combating the stiction effect\cite{71, 72}. For our purposes we are generally able to fabricate small enough gaps through critical point drying and HF vapor etching, but future applications that require smaller gaps may need to take in the added complexity of actively reducing stiction.

4.2.4 Optical frequency gradient force

It is worth noting that we can explore interesting applications while simplifying the fabrication process from the previous section. By removing the metal contacts that were designed to apply a gradient force we can remove several critical steps of fabrication and thus increase our yield. We will now be inducing the force purely through optical means.

The optical force arising from pumping the PCNCs at around 1.6 µm (186 THz) is oscillating so quickly that the PCNCs merely see a time-averaged force based on the RMS field amplitude, or the intensity of the light.

4.2.5 Calculating the force

The general calculation of the force exerted by an electromagnetic field is calculated using Maxwell’s Stress Tensor\cite{10}:

\[
T_{\alpha\beta} = \varepsilon_0 \left[ E_\alpha E_\beta - \frac{1}{2} \left( \vec{E} \cdot \vec{E} + c^2 \vec{B} \cdot \vec{B} \right) \delta_{\alpha\beta} \right]
\]

the indices \( \alpha \) and \( \beta \) indicate the three spatial components of the vectors, and \( \delta \) is the Kronecker delta. This tensor gives the momentum flow (or force per unit area) in the \( \alpha \)th direction (e.g. \( F_x/m^2 \), \( F_y/m^2 \), and \( F_z/m^2 \)) through a surface normal to the \( \beta \) direction. In order to determine the electromagnetic force exerted on an object it is enclosed in a box and \( T \cdot \vec{n} \) is integrated over the surface of the box. Fig. 4.2.6 illustrates the concept of enclosing the object of interest.
inside the box:

\[ F_a = \oint \mathbf{d}S \sum_{\beta} T_{a\beta} n_\beta \quad (4.7) \]

This stress tensor method of calculating forces has two interesting properties. The first is that as long as the surface completely encloses the object of interest, the actual shape is immaterial. For ease of calculation the surface is often chosen to be of a similar shape to the object of interest. Second, it is possible to have forces that are not perpendicular to a surface acting upon a surface. These “shear” forces are one of the more unique aspects of electrodynamics.

In the case of our coupled PCNC cavities we can choose two rectangular prisms that engulf the individual PCNCs and with the use of Eq. 4.7 calculate the forces exerted by the various optical modes. We indeed find that any kind of non-dispersive mode with a negligible \( g_{om} \) will not have any appreciable force arising from the electromagnetic mode. The limitation of this calculation method is that it is static. If the force from the mode is significant it will cause the PCNC beams to rearrange themselves, leading to a slightly different optical mode, as well as loading up the mechanical spring. In the case of capacitive forces we were able to incorporate the changing geometry by the use of self-consistent calculations. If, instead, we must re-calculate the optical modes continuously, the problem
becomes intractable. This is because the optical mode calculations are significantly more computationally intensive than the mechanical and electrical calculations.

### 4.2.6 Optomechanical Coupling Strength

To deal with these issues the optical and mechanical modes need to be treated equally, and in a way that removes the complete recalculation of the optical mode after every displacement. Fig. 3.2.1 first introduced the idea that the resonance of the supermodes of coupled PCNCs is controlled by their separation. We saw that the odd-mode exhibited little tuning, whereas the even-mode was highly dependent on the position. We are now going to introduce the concept of the optomechanical coupling strength, or $g_{om}$. This is a measure of the amount of shift of the wavelength of the optical resonance for a unit change in the separation. The even mode has a very high $g_{om}$, that increases as the gap decreases. The odd-mode on the other hand has a small and relatively constant $g_{om}$.

If we generalize our scope to any interaction where some form of mechanical motion leads to a change in the resonant wavelength of an optical mode, then this will lead to a force that will actuate the mechanical motion in such a way as to lower the energy of the optical mode[73]. In the case of the coupled PCNCs and the even-mode in particular, this leads to an attractive force between the two PCNCs. If there are enough photons in the optical mode, each having an energy of $h\nu$ (where $h$ is Planck’s constant and $\nu$ is frequency), then any decrease in the frequency of the mode would be available to do work through mechanical motion; loading the “spring.” At a certain point the force required to further load the spring would increase to the level where the total energy of the system would increase if the beams moved any closer together, and the motion would cease.

$g_{om}$ is the figure of merit when calculating the force exerted by the optical mode on the mechanical, and it will give a convenient measure of how much energy can be transferred between the modes. A positive $g_{om}$ will lead to an attractive force, whereas a negative $g_{om}$ will lead to a repulsive force. If the mode
is not dispersive, that is, if $g_{om}$ is zero, there will be no force. If the $g_{om}$ is in units of $Hz/m$ instead of $m/m$, that is, we move from the wavelength picture to the frequency picture, then the sign convention will be switched.

The complete development of the optomechanical theory is far beyond the scope of this work, and the interested reader can find further details in Deotare [74] and the supplementary material of Deotare et al. [23]. However, it is possible to develop a full theory of the interaction between a mechanical and optical mode. One of the nicer features of this full theory is that calculating the force exerted on the mechanical mode by the optical mode becomes considerably simpler:

$$F = 2g_{om} U_{optical} \frac{\omega_{optical}}{\omega_{optical}}$$

(4.8)

$U_{optical}$ and $\omega_{optical}$ are the respective energy and angular frequency of the optical mode. This means that if the $g_{om}$ of a cavity is well characterized, then calculating the force due to pumping energy into this cavity becomes trivial [23, 75].

4.2.7 Multiple modes

At this point it is worth considering exactly what will happen if we dump a large amount of optical power into a high Q cavity. Because of the Q there will be a significant amount of energy stored in the optical mode. This will lead to a force that, as we intend, will cause some level of mechanical motion. This will cause the resonance frequency to shift. If we have previously used a tunable laser to locate the resonance so that we can start to pump it with energy, we will no longer be targeting the correct wavelength. So some kind of feed-back based system that actively updates the wavelength of the laser is required to keep it in resonance with the shifting cavity [75–78]. One alternative would be to find another mode, that by virtue of its much lower Q, will not be as sensitive to this shifting.

The PCNC designs we use happen to have a large number of modes. So far we have concentrated on the “fundamental” odd- and even-modes. However, for
Figure 4.2.7: The pump-probe configuration and results. a) The experimental apparatus. A tunable laser source is used as both the pump (red) and probe (blue). The pump is amplified with an erbium doped fiber amplifier (EDFA) and then polarization adjusted. It is then multiplexed (MUX) with the polarization controlled probe light. On-chip the light interacts with the device under test (DUT) and then is de-MUXed. The separate signals are probed with photodiodes and the result read out on a computer. b) The delocalized pump mode. The symmetry is similar to the even cavity mode, but it is much more spread out and has a lower Q. c) A transmission spectrum from an example device. The probe mode is at the far left, and the pump mode corresponds to one of the broad peaks around 1,580 nm. The shading indicates the EDFA range. d) Observed tuning of the even (inset and red line) and odd (blue line) supermodes. The odd supermode is used as a temperature calibrator, and the optomechanical tuning (black line) is extracted.
every symmetry plane in the system it is possible to find higher order modes[4]. One of these can be viewed in Fig. 4.2.7b. As the order of the mode increases, the modes become less localized and have lower Qs. Eventually the modes reach a continuum state as part of the pass-band Bloch-modes of the photonic crystal. It worth remembering that these modes are still able to exert an optical force as long as they have a non-zero $g_{om}$.

In this pump-probe methodology where a weak laser is used to monitor a high Q cavity mode, while a strong laser is used actuate mechanical motion, we have traded the complicated process of a feed-back system for the complication of bringing the output of two separate lasers to our devices. For a single device this trade-off is neutral; however, if you consider having large arrays of coupled PCNCs, perhaps each with a different wavelength of high Q mode, then it would quickly become prohibitive to monitor each individually, but a single, powerful beam could be used to actuate them through their low Q modes.

### 4.2.8 Bringing silicon into the equation

In addition to dealing with the extra optical mode introduced in the pump-probe system, our optomechanical model needs to take into account real material properties. To this point we have been treating our PCNCs as if they were composed of perfectly linear material. However, with high Q cavities and with the large optical power being sent into the pump mode we need to consider silicon as a more complicated material. Silicon possesses a strong $\chi^3$, or third order nonlinearity[80]. Normally, silicon's indirect band-gap means that it is transparent throughout the near-infrared. However, with sufficiently high intensities two-photon absorption (TPA) becomes an issue. TPA is a particular problem because as free-carriers are created the absorption cross-section increases and the silicon starts heats up. The thermo-optic effect will shift all of the resonance frequencies of the PCNCs. This means that simply monitoring the

---

$^9$Pairs of waveguides on their own (for instance, PCNCs without the holes) can also create an attractive and repulsive force; however it will be much weaker than any effect based on a cavity resonance, since the stored energy is much lower.[79]
**Figure 4.2.8:** Temperature tuning of the probe-mode with high-frequency modulation. a) The modulation is scanned over a wide-frequency range. Several regions emerge and are explained in the text. b) Line scans from the indicated frequencies in (a). In region 2 the thermal response has essentially ended, so the broadening of the red scan over the blue one is due to mechanical vibration that is averaged out. The transition from region 2 to 3 shows mechanical resonance. Then the mechanical response will start to roll-off, leaving behind a narrow blue scan from region 3.

shift of the resonance frequencies is not sufficient to determine the optomechanical contribution to the tuning of the probe mode.

All of the various non-linearities in silicon can be integrated into the optomechanical model, and predictions can be made to determine the relative contributions of the TPA versus optomechanical effects. The reader is once again referred to [23, 74] for the minutiae. However, it is worth noting that the worst of the non-linearities — TPA — can be circumvented by modulating the pump beam at a frequency beyond the thermal response time of the PCNCs.

Fig. 4.2.8 shows the measured spectrum of the probe mode as a function of frequency of pump mode modulation. This picture is quite interesting and all of the features are not immediately obvious. There are several dynamics at play. First, the probe beam is scanning across a small wavelength range at a speed that is much slower than the most of the modulation frequencies.

At low frequencies of modulation the scanning probe laser can keep up with the modulating pump beam. This leads to a general “smear” in the spectrum. In the second stage two distinct bright bands emerge. This is because the
modulation is now fast enough that the scanning beam merely sees an averaging of the detuning of the resonance. As with any sine-wave modulation the “velocity” of modulation is fastest around the neutral point, whereas the velocity falls to zero at the end-points. This means that a time average view of the modulation will see more time spent near the end-points. This leads to these bright bands. The bands narrow together as the frequency continues to increase. This is due to the thermal-frequency limiting (see the RC time constant analogy in the next section), and the beams are unable to heat up and cool down fast enough to reach the full detuning limit. Eventually there is a complete cut-off and there will be no detuning. The frequency becomes high enough that a constant temperature is reached, based on the root-mean-square intensity of the pump beam. In the case of the even mode we see a large distortion at the mechanical resonance frequency. After mechanical resonance, the mechanical response also rolls-off, leaving a probe mode that is essentially unaffected by temperature and mechanical changes.

4.2.9 Thermal response

There is a direct analogy between electrical circuit theory and thermal circuitry. Electrical charge is analogous to heat, potential is mapped to temperature, electrical conductivity to thermal conductivity, and capacitance to heat capacity. Therefore our silicon PCNCs will have a heat source in the optical mode where the pump-beam is being absorbed, which is equivalent to a current source. The heating of the beams will be the equivalent of a charging capacitor, and in order to cool down the heat will have to flow across the silicon, which will act as a resistor. There is, therefore, a time constant (similar to an RC time constant) with a cut-off frequency and response similar to the “roll-off” discussed in Sec. 4.1.3. This cut-off point can be tested experimentally as in the previous section, but we can also model it in an FEM solver such as COMSOL Multiphysics. Fig. 4.2.9 compares the results of the two approaches.

Based on widely available parameters, Fig. 4.2.9 shows the results of simulating
Figure 4.2.9: Circuit model for temperature modulation. a) Simulated temperature response to modulated heat source as well as fit based on the electrical circuit theory. A thermal conductivity of $\kappa$ was used. The $3\, \text{dB}$ cut-off is found at $97\, \text{kHz}$. b) Experimentally measured response of optical mode to modulated heating. Agreement with theory is excellent with the measured $3\, \text{dB}$ cut-off at $101\, \text{KHz}$. c) The electrical circuit representation after which the thermal response is modeled. d) The elements of the electrical-thermal analogy are highlighted.
the thermal response, including the temperature modulation as function of input frequency. The one point of interest was the adjustment of the thermal conductivity of silicon. While the bulk-value is well known, surface effects become significant and the value decreases rapidly for silicon less that 200 nm thick\[81–83\]. The bulk value is approximately 150$W/m \cdot K$. Li et al. \[84\], have measured single-crystal silicon nanowires’ thermal conductivity and found at 115 nm diameter the thermal conductivity is 40$W/m \cdot K$, at 56 nm it is 25$W/m \cdot K$, and at a diameter of 37 nm the value is under 20$W/m \cdot K$. It is important to remember that these values are strongly dependent on surface treatments, so they should only be compared on an order of magnitude basis. However, they do give credence to our usage of 20$W/m \cdot K$ in arriving at the excellent agreement between theory and experiment in Fig. 4.2.9a,b.

4.2.10 Utility and outlook

Our system of coupled PCNCs is a rich playground for studying the various effects of optomechanics and electro-optomechanics. If placed in a vacuum the mechanical Q of our various dark and bright modes can reach tens of thousands. Even in a damped atmospheric setting we can observe optical spring constants, and stiffening and softening the mechanical modes etc.\[23, 74\]. However, the purpose of this work is to study the utility of these systems as on-chip filters. In this sense these filters are far from ideal. While they do provide and excellent all optical platform, their power consumption and total tunability fall short when compared to either capacitive or gradient force tuning. Once the PCNCs have been reconfigured to a desired wavelength, the pump beam must still provide considerable power to hold the position. This also results in significant thermal loss due to TPA.

Therefore, in applications where reconfiguration at the 10 MHz level or lower is specified: the capacitive reconfiguration offers a low power, high dynamic range option. If we want to avoid having to degrade the optical properties of the PCNC and instead wish to have a pure semiconductor or an insulating optical
material the DC and RF gradient force configuration offers an attractive alternative. In both cases we can further improve the performance by operating with a DC bias and taking advantage of the amplification shown in Eq. 4.3 to achieve dynamic results with even smaller voltage swings.
In which the author learns the value of doing the unexpected.

We are now departing from the friendly confines of the photonic crystal. Despite the many virtues of the PCNC, including its easy design, the varied ways of tuning the optical resonances, and the advantages that can be gained by localizing light to a high degree. It is a fundamentally limited system. A stop-band is created and then a defect is used to create a subset of frequencies within the stop-band that are able to tunnel across the device. This means that the number of spectral shapes we can generate are severely limited. Additional fields such as plasmonics[85, 86] and metamaterials[87, 88] provide methods for designing structures that can manipulate light, but even there we see limitations on the diversity of spectral response from a single device.

A significant advance is the possibility of designing a structure that can reflect spectral features of arbitrary bandwidth, amplitude, and phase. We present a
simple and fast method for designing such a structure. While this method can be
generally applied to any material system where the refractive index profile can be
controlled along one axis, we have concentrated on implementing arbitrary
reflective filters in compact, on-chip silicon-on-insulator (SOI) waveguides.
These integrated filters have myriad applications from on-chip signal routing to
compact and ultra-fast pulse shaping[89], the latter being of significant interest in
quantum control experiments[90].

The search for a new paradigm leads us back to the basics. In Ref. [91] the
process of deriving the Fourier transform relationship between a reflection
spectrum and the impedance profile of the interface on which the light incident
was laid out in detail. In this derivation two assumptions are made: first, that the
wave impedance is that of a traveling plane wave and second, that the amplitude
of the reflection is small. These are rigorous assumptions when dealing with
free-space and anti-reflection coatings, and they led to excellent results in terms
of deriving the best impedance gradient between two materials with different
wave impedances.

5.1 Waveguide filters

At this point we can make a few leaps of faith. First, we will attempt to try this
design method in instances where the reflection coefficient is not small; indeed,
we will see what happens when we push the limits to unity. Second, we will map
the impedance — a well defined property of the plane wave — onto the effective
index of a waveguide. Because of the discontinuities in the electric field in a
dielectric waveguide, the wave impedance is not a well defined property (in
contrast to free space or the modes of waveguides framed by conductors).

By considering two dimensional “slab” modes we can visualize exactly how
and why the wave impedance is not defined. In Fig. 5.1.1 we can see the TE and
TM modes of a slab dielectric waveguide. At the interfaces between materials we
know the $\vec{D}_\perp$ and $\vec{B}_\perp$ fields are continuous. Using the linear constitutive
relationships, we can recast this as continuity in $\varepsilon_\sigma \varepsilon J_{\perp}$ and $\mu_\sigma \mu \vec{H}_\perp$. In a
Figure 5.1.1: Comparison of TE and TM mode profiles in a 2D slab. a) The TM mode has no discontinuities so the wave-impedance is well defined. b) The TE mode has a discontinuity at the boundary of the high and low index material. This makes it difficult to define the wave-impedance.

dielectric waveguide $\varepsilon$, has a discontinuity at the boundary, whereas the $\mu$, is continuous for dielectric materials like silicon or glass. Therefore, $\vec{H}$ is always continuous, whereas depending on the polarization of the mode there may or may not be a discontinuity in the $\vec{E}$ field ($\vec{E}$, is always continuous).

When we move to a three dimensional case where the waveguide is no longer an infinite slab but has a dielectric boundary in both the $y$ and $z$ directions, there is always a discontinuity in the electric field. This means that modeling the wave impedance as inversely proportional to $n_{\text{eff}}$ becomes a significant approximation. Despite this difficulty we will move forward with the derivation of a reverse design method based on reducing the waveguide system to a 1-dimensional system described by the $n_{\text{eff}}$.

We start from the master differential equation for the ratio of the reverse propagating wave to the forward propagating wave as derived in \( \text{[92]} \). Some important variables in this equation are

\[
L = \int_0^1 n_{\text{eff}}(x) \, dx
\]

\[
u = \frac{1}{L} \int_0^x n_{\text{eff}}(t) \, dt
\]
$L$ is the optical path length and $u$ is a normalized position coordinate along the waveguide. When $\omega$ is the angular frequency the master equation is

$$
\frac{d}{du} r(u, \omega) + \frac{2L\omega}{c} r(u, \omega) = -\frac{1}{2} (1 - r(u, \omega)^2) \frac{d}{du} \ln \left( \frac{1}{n_{\text{eff}}(u, \omega)} \right)
$$

(5.3)

From now on "′" will refer to differentiation with respect to $u$. This is a Ricatti equation with no known general solution[93]. However, if we neglect the $r^2$ term, we can then choose an appropriate integrating factor:

$$
r(u, \omega) e^{i \frac{2L\omega}{c} u} = \frac{1}{2} \ln \left( n_{\text{eff}}(u, \omega) \right) e^{i \frac{2L\omega}{c} \ln \left( n_{\text{eff}}(u, \omega) \right)}
$$

(5.4)

We have also simplified the logarithmic expression. The left side of the equation is easily integrable over the region of interest giving us:

$$
r(u, \omega) e^{i \frac{2L\omega}{c} u} \bigg|_{u=1}^{u=0} = \int_0^1 \frac{1}{2} \ln \left( n_{\text{eff}}(u, \omega) \right) e^{i \frac{2L\omega}{c} \ln \left( n_{\text{eff}}(u, \omega) \right)} du
$$

(5.5)

where we have integrated over the length of the modulated region. We proceed by plugging in boundary conditions: since there is no light incident from the right hand side: $r(u = 1) = 0$. So our evaluation becomes

$$
r(0, \omega) = \int_0^1 \frac{1}{2} \ln \left( n_{\text{eff}}(u, \omega) \right) e^{i \frac{2L\omega}{c} \ln \left( n_{\text{eff}}(u, \omega) \right)} du
$$

(5.6)

Now, in order to get useful results out of this equation we need to transfer back from a normalized position coordinate to real position. The $u$ in the exponential cannot be transformed exactly. However, by assuming that we operate at a central $n_{\text{eff}} = n_o$ that is much larger than the modulation that we will apply to it, we can make the assumption that $u = \frac{n_o}{L} x$. This is only effective because we are moving from a uniform waveguide into an area with small perturbations and finally into a uniform waveguide with the same effective index.
With the appropriate substitutions we arrive at:

\[
  r(f) = -\frac{1}{2} \int_{x=0}^{x=1} \frac{1}{n_{\text{eff}}(x)} \frac{dn_{\text{eff}}(x,f)}{dx} \exp \left( i \frac{2\eta_0}{c} f x \right) \, dx \quad (5.7)
\]

Here we have recognized that \( r(\omega, \omega) \) is identically the reflection coefficient of the entire modulated region, and we have dropped the positional dependence of \( r \). The final step is to recognize that the derivative of \( n_{\text{eff}} \) is zero outside of the modulated region so that we can extend the integration limits to infinity, explicitly showing the Fourier (or in this case inverse-Fourier) transform relationship between the gradient of \( n_{\text{eff}}(x,f) \) and the \( r(f) \) (we also switch from \( \omega \) to \( f \), in order to bring the factor of \( 2\pi \) out):

\[
  r(f) = \frac{1}{2} \int_{-\infty}^{\infty} -\frac{1}{n_{\text{eff}}(x,f)} \frac{dn_{\text{eff}}(x)}{dx} \exp \left( i \frac{2\eta_0}{c} f x \right) \, dx \quad (5.8)
\]

We can further refine this, because in reality we will not be directly controlling the \( n_{\text{eff}} \) but will instead be controlling the width of a waveguide. Therefore, using an eigenfrequency analysis of the cross-section of the waveguide, we can map the width of the waveguide to a specific \( n_{\text{eff}} \). However, this value will vary as a function of the wavelength; therefore we need to take into account this wavelength dispersion. We then use the chain rule to expand: \( \frac{dn_{\text{eff}}}{dx} = \frac{dn_{\text{eff}}}{dW} \frac{dW}{dx} \), where \( W \) is the width of the waveguide. Once again we must utilize our assumption that \( n_{\text{eff}} \) is essentially constant in position, allowing us to assume that the \( W \) derivative can be pulled out of the integral as a constant in position. This gives us

\[
  r(f) = \frac{1}{2} J(f) \int_{-\infty}^{\infty} \frac{dW}{dx} \exp \left( i \frac{2\eta_0}{c} f x \right) \, dx \quad (5.9)
\]

where \( J(f) = -\frac{1}{n_{\text{eff}}(f,W)} \frac{dn_{\text{eff}}(f,W)}{dW} \). Any frequency, geometric, and material dispersion in the waveguide is now contained within the \( J \) function. If we want to obtain a width profile \( (W(x)) \) with a reflection spectrum of \( r \), it is now a simple matter of inverting the Fourier transform and numerically integrating the result.
with appropriate boundary conditions:

\[
\frac{dW}{dx} = \int_{-\infty}^{\infty} \frac{2r(f)}{f(f)} \exp \left( -i2\pi \frac{2n_0}{c} f x \right) dx \tag{5.10}
\]

### 5.2 Improving the Design for Large \( r \)

At this point it is worth exploring the effect of dropping the \( r^2 \) term of Eq. 5.3. For example, in our experimental designs we aimed for a peak reflectivity of unity. However, when this design is tested by numerically solving Eq. 5.3, we see that it only results in a peak of \( r^2 = 0.6 \). We therefore use a design higher than \( r^2 = 1 \). For instance, in order to target a reflection of unity, we resorted to a design goal of \( r^2 = 4 \), a number higher than is physically possible. This had the effect of driving the reflection values up, at the cost of some level of distortion. This is possible because the mathematics of the design process are not constrained to physically realistic parameters.

We can also use an iterative design process where we can scale the amplitude independently at each wavelength and feed that back into Eq. 5.10. This process allows a more faithful reproduction of the target \( r(\lambda) \) at the cost of running each iteration through an FDTD simulation to check the real reflection spectrum. Fig. 5.1.2d shows how the simulated reflection spectrum evolves over 4 iterations. Fig. 5.2.1 is a graphical representation of the iteration process.

If we want to avoid the cost of running a full FDTD (Lumerical Mode Solutions is an excellent tool for waveguide-type simulations) simulation at every iteration of the design process, then there are some modifications that we can make to the design process to improve the result on the first iteration. We divide both sides of Eq. 5.3 by \( 1 - r^2 \):

\[
\frac{r(u, \omega)'}{(1 - r(u, \omega)^2)} + \frac{2L\omega}{c} \frac{r(u, \omega)}{(1 - r(u, \omega)^2)} = \frac{1}{2} \ln \left( n_{\text{eff}}(u, \omega) \right)'
\tag{5.11}
\]
Figure 5.1.2: Width modulated waveguide filter example and design. a) SEM of a width modulated region of a SOI waveguide. b) An example target spectrum of Reflectance. c) The width profile obtained by plugging the spectrum from (a) into Eq. 5.10. d) Successive iterations of correcting the spectrum that emerges from Eq. 5.10 as detailed in Fig. 5.2.1. The initial iteration yields both a lower than expected amplitude and distortions in the shape. The second iteration multiplied the target amplitude by four, and the third by seven. This largely corrected the amplitude issues but distorted the shape. In the last iteration the correction was wavelength dependent and went a long way toward fixing the distortion in the shape. This can continue until the target is almost exactly matched.
Figure 5.2.1: Width modulated waveguide filter design methodology to overcome our incorrect assumptions about impedance and small reflectances.

The first term on the left is identically \( \tanh^{-1}(r(u, \omega))' \). We now can make the approximation that \( r/(1 - r^2) = \tanh^{-1}(r) \). This is especially valid for small values of \( r \) but maintains its validity at larger values when compared to \( r^2 = 0 \).

This gives us a new master equation:

\[
\tanh^{-1}(r(u, \omega))' + i{\frac{2L\omega}{c}}\tanh^{-1}(r(u, \omega)) = \frac{1}{2}\ln(n_{eff})'
\]  

(5.12)

The similarity in form to Eq. 5.3 permits us logically follow all the steps of the previous section to arrive at

\[
\tanh^{-1}(r(f)) = \frac{1}{2}f(f) \int_{-\infty}^{\infty} dW \frac{dx}{dx} \exp \left( i\frac{2n_{o}}{c}f x \right) dx
\]  

(5.13)

or

\[
\frac{dW}{dx} = \int_{-\infty}^{\infty} \frac{2\tanh^{-1}(r(f))}{f(f)} \exp \left( -i\frac{2n_{o}}{c}f x \right) dx
\]  

(5.14)

where we have inverted the Fourier transform. Therefore, given a target reflectance spectrum and a target width, the width profile can be quickly calculated using any number of discrete Fourier transform methods available in
mathematical software packages.

It should be noted that this design method places limitations on the target \( r^2 \). Because \( \tanh^{-1} \) is defined from -1 to 1, we must limit our target spectrum of \( r \) to this range. This means that we cannot incorporate the tricks of the iterative design process in quite the same way. Fortunately, this design process is so much superior that only one iteration is needed.

The relative value of this solution can be easily tested when we compare the results of a design by equation 5.14 with that of equation 5.10 by solving equation 5.3 numerically or by running a design through FDTD. The difference between solving 5.3 numerically and running a full FDTD solution is discussed in the following section.

5.3 Comparing the Master Equation to FDTD

While numerically solving equation 5.3 gives us information as to the efficacy of our assumption that \( r \ll 1 \) or \( \tanh^{-1}(r) = r/(1 - r^2) \), this equation contains no information about the effect of assuming that the wave impedance is given by the \( n_{\text{eff}} \) of the waveguide mode. This information requires that Maxwell’s equation be solved in at least two dimensions. To this end we will look at 2D FDTD simulations of the design obtained from Eq. 5.14 for TE and TM modes. Fig. 5.3.1 shows how the first iteration of Eq. 5.14 compares to a target \( R(\lambda) \) spectrum.

Additionally, the solution is run through a 1D numerical solution of Eq. 5.3. The design method from Eq. 5.14 evidently does a much better job in a single iteration than the design method from Eq. 5.10\(^1\). Additionally we see that the ODE solution is virtually identical to the TM design, whereas the TE design deviates considerably. This shows that for the TM mode where the wave impedance is properly defined the \( n_{\text{eff}} \) calculation allows us to treat the 2D problem as 1D. However, due to the discontinuities in the TE mode this is a deficient, yet clearly still useful, approximation.

\(^1\)Compare to the initial iteration in Fig. 5.1.2d
Figure 5.3.1: Comparing the 1D approximation to a full solution. Designs based on Eq. 5.14 show that the TM mode is approximated extremely well by the 1D approach. The TE mode is not nearly as effective in reproducing the shape of the intended reflection spectrum.
Figure 5.3.2: Comparing the 3D TE and TM filter designs based on Eq. 5.14. The agreement for the TM mode is far superior to that of the TE mode. There is a factor of 3-5 in the resulting amplitude to be gained by using the TM mode.

Once the problem is moved from 2D to 3D, there will always be a discontinuity; therefore there is no exact solution as in the 2D TM case. However, we can consider the nature of the discontinuity. In the TE case the modulation is at the exact location of the discontinuity. The TM mode, on the other hand, has a constant discontinuity and the modulation is across a continuous field profile. Fig. 5.3.2 shows the advantage of using the TM mode.

5.4 Filter fabrication and testing

Filters based on designs from Eq. 5.10 with a TE mode profile and Eq. 5.14 with a TM mode profile were fabricated and tested. The TE modes were designed with $R = 4$ to raise the reflectance values. Because the filters operate in the reflection regime we employed a 3 dB directional coupler to tap into the reflected signal at an output port. One of the advantages of this configuration is that the effective index modulation for the TE and TM modes are different enough that their reflectance spectra do not overlap and the signal in the reflected arm had a
Figure 5.4.1: Experimental configuration and measured spectra. (a) A set of five target spectra. The intensity is in a linear saw-tooth pattern. (b) Normalized, measured reflections from fabricated devices. The dashed lines indicate the uncertainty in the measurement. (c) SEM micrograph of example device; the inset shows a magnification of the width modulated region.

much lower noise floor than in the transmitted arm. Fig. 5.4.1c shows an SEM image of fabricated devices.

The fabrication process is very similar to the waveguide coupled PCNCs. A FOx mask is defined with e-beam lithography. Because the reflection profile is highly dependent on the width of the modulated region, great care needs to be taken with scaling the e-beam pattern. After developing the FOx in high contrast TMAH the pattern is etched into the SOI device layer using ICP-RIE. SU8 waveguides are written using e-beam lithography for precision alignment to the silicon inverse tapers. Finally, SiO$_2$ is deposited to a thickness of 2 - 4 $\mu$m using PECVD.

The PECVD oxide allows the SU8 to be polished after the wafer is diced. The polishing process ensures a uniform facet for butt coupling onto the chip. It also has the advantage of increasing the yield of the chip because SU8 that peeks off away from the edge of chip can brought back to the facet through extensive polishing.

Filter characterization was performed with a scanned tunable laser (Agilent 81682). Light was coupled onto the chip through a tapered, lensed fiber (Oz optics). Figs. 5.4.1a,b show example target spectra alongside experimentally
Figure 5.4.2: Measured filters based on the TM mode. Target amplitudes are shown with dotted lines. The designs are based on Eq. 5.14 with a single iteration.

measured spectra for the TE mode with a designed amplitude of $R = 4$. The agreement between the two is excellent. The signal is distorted by Fabry-Perot resonances in the system as well as the general deficiency of the TE mode due to discontinuities at the boundary of the waveguide. The absolute values of measured reflectance are based on average values for transmission through unmodulated waveguides with the uncertainty in the measurement arising from variations in the polymer waveguide facets and from insertion and extraction losses that arise from this coupling method.

The TM mode filters were fabricated in an identical fashion. The measurements were made with a Santec (TLS- próp) tunable laser. The resulting amplitudes of the filters fit with the target spectra much better than the TE mode as can be seen in Figs. 5.4.2a,b. Even smaller features are reproduced, though in a slightly degraded manner.

5.5 Application to ultra-fast pulse shaping

An important application of our approach, and arbitrary filters in general, is in shaping ultra-fast pulses. Bulky apparatus is currently used for ultra-fast shaping, and it requires precision alignment [94, 95]. By using the SOI waveguide platform our filters allow us to generate arbitrary pulse shapes in an integrated,
Figure 5.5.1: Femtosecond pulse shaping simulations. (a) Cartoon of devices filtering out blue light by reflecting it back; the red light is transmitted. (b) Time domain, reflected, arbitrary pulse shapes obtained from 3D FDTD simulations. (c) Gaussian input pulse used to obtain the pulses in (b). (d) Reflectance spectrum of the filters used to convert the input Gaussian pulse into the output pulse shapes.

on-chip fashion. The small footprint of the filters additionally permits a single external pulse to excite many different pulse shapes in parallel. Ultra-fast pulse shaping is of particular interest in quantum coherence control and other quantum optics experiments where a wider range of pulse shapes beyond “transform limited” is required [96, 97].

The key to pulse shaping is the control of amplitude and phase over a wide wavelength range[98]. Since Eq. 5.10 and 5.14 solve for r (the reflection coefficient) rather than R (the reflectivity, or |r|^2), the necessary conditions for
ultra-fast pulse shaping are met.

To illustrate the concept theoretically we consider two cases: i) we realize triple pulses with a Hamming\(^2\) envelope, and ii) a single pulse with saw-tooth envelope, starting both times with a simple Gaussian pulse. In both cases, the filter is designed using our inverse-design principle based on Eq. 5.10 as discussed above. To validate our data, we present finite difference time domain (FDTD, Lumerical Inc.) simulations showing a single Gaussian pulse launched into two different width-modulated SOI waveguides, and converted into three distinct Hamming pulses and a linear (saw-tooth) pulse, respectively. Fig. 5.5.1 shows the time domain representation of the input and the two simulated reflected pulses. All simulations are three dimensional with the mesh grid size of 1 nm, which is similar to the e-beam lithography resolution available to us. This is an excellent first demonstration for the viability of integrated femtosecond pulse shaping. Because of the inverse relationship between the width of the pulse and the broadness of the frequency spectrum, our corrugated waveguides lend themselves to shaping faster pulses. Shaping slower (for instance picosecond) pulses is possible but requires longer filters with smaller modulation of the width.

### 5.6 Controlling the Phase of the Filters

At this point I would like to once again stress that we are able to control the design of not just the reflectivity (\(R\)) but also the reflection coefficient (\(r\)), which contains information about the phase of the reflected wave. In particular, while designing our filters we always convolve the amplitude we are aiming for with a phase that increases linearly with frequency. Without this phase the resulting designs do not have a continuous filter shape. This strange design behavior is rooted in the way in which discrete Fourier transforms operate. However, it is possible to add non-linear phase terms while still maintaining good design practices. The non-linear phase shift can be used to counteract material dispersion and keep pulse shapes from distorting as they propagate through a

\[ w(n) = A \left(0.54 - 0.46 \cos \left(\frac{2\pi n}{N}\right)\right), \quad 0 \leq n \leq N \]
Figure 5.6.1: Simulated phase and amplitude of a filter. The amplitude has three “Dolph-Chebyshev” peaks. The design is a single iteration of Eq. 5.14. This is based on FDTD simulations (Lumerical Mode Solutions). The phase is remarkably linear across the designed region.

waveguide. This is a well known process called “chirping.”

FDTD simulations of our filters enable us to extract the phase of monitored signals. Fig. 5.6.1 clearly demonstrates that the phase proceeds linearly with wavelength. This linear phase response is desirable because it means that shapes of pulses coming in and out of the filter will not be distorted. This is because the slope of the phase gives the phase-group velocity, so a linear filter means that all of the frequency components travel at the same velocity.

In addition to checking our assumptions in FDTD simulations, we can also set-up phase sensitive experiments that will indirectly give us information about the phase response of the filters. One possibility is to split a single coherent input, reflect it off two distinct filters, such as the ones in Figure 5.6.2a, and then combine these reflected signals. If we compare the signals that were combined on-chip, to the sums of the individual outputs, we should see a significant difference as illustrated in Fig. 5.6.2c. One measure of coherence between the
Figure 5.6.2: Design of an on-chip interference experiment. (a) Proposed reflection coefficient spectra. (b) Resulting width modulation obtained from Eq. 5.14. (c) There are two possible outcomes from interfering the signal from the two filters. If they are in phase, then they will coherently add. If if phase is completely random, then they will incoherently add.
two signals will be the truth of the following inequality:

$$\langle I_1 + I_2 \rangle \neq \langle I_1 \rangle + \langle I_2 \rangle$$  \hspace{1cm} (5.15)

where the angled brackets refer to averaging over a long time, the effect of measuring the light on a photodetector with a bandwidth much, much smaller than the frequency of the incident light. In the event that the inequality is violated we know that the signals from the two filters are not coherent. There are two possibilities for such an outcome. The first is that the path difference of the input laser signal after the original split is longer than the coherence length of the laser. This would yield to the equality

$$\langle I_1 + I_2 \rangle = \langle I_1 \rangle + \langle I_2 \rangle$$  \hspace{1cm} (5.16)

However, this should not occur as we are able to control the length of the waveguides precisely. The second option is that the phase coming off of our filters is not controlled by us, and we are only really controlling the amplitude. The result of the interference in this case still maintains the inequality from Eq. 5.15, however, large fluctuations in the spectrum and not a clean outcome such as seen in Fig. 5.6.2c would occur.

This experiment was set-up to run in 3D FDTD. Fig. 5.6.3a-c show the resulting signals in FDTD as well the differential phase between the two filters. The middle peak is close to destroyed due to interference effects. Fig. 5.6.3d also shows an optical micrograph of an example fabricated device. Fig. 5.6.4 show the experimentally measured spectra of the device in Fig. 5.6.3d. The constructive and destructive interference are clearly visible.

While this is not a direct confirmation that the phase is linear, we do show that we at least have control over the phase, and coupled with FDTD simulations this strongly suggests that we do in fact have a linear phase response.
Figure 5.6.3: Simulations of on-chip interference. a) The simulated reflected signal for a triple-peak spectrum with all the peaks positive. b) The combined signal from the filter in a) and c). The differential phase shows that where the signal interferes the two filters are out of phase. c) The reflectance spectrum of the second filter where the middle peak is negative, making it out of phase with the first filter. d) An optical micrograph of an example fabricated device with labels showing the individual and combined outputs. e) A high magnification SEM of the width modulated filter region.
Figure 5.6.4: Experimental verification of on-chip interference. The individual outputs from Fig. 5.6.3c or on the left and right while the combined signal is shown in the center plot. The destructive interference in the combined signal is significant.
Figure 5.7.1: Relationship between the bandwidth and width profile of our reverse designed filters. a) An 80 \( \text{nm} \) band-width filter. The maximum modulation is \( \pm 9 \text{nm} \) and the filter length is 150 \( \mu\text{m} \). b) A 40 \( \text{nm} \) band-width filter of the same shape as (a). The maximum modulation is halved while the filter length is doubled.

5.7 Overcoming the Scale of Width Modulations

The magnitude of the width modulations required of these filters is a limiting factor in the fabrication. So far we have shown filters with features sizes over 20 \( \text{nm} \). This is determined by the width of the waveguides as well as the band-width of the filters. The width is often determined by the single-mode limit. The theory of the filters does not require a single-mode waveguide, but in practice a multi-mode waveguide will add noise and make the filters harder to measure accurately.

The Fourier transform nature of the design process allows significant intuition when it comes to designing filters with large width modulations. Fig. 5.7.1 shows how doubling the band-width doubles the size of the modulation while halving the length of the filter. This means that wider-band filters are much easier to fabricate.
Figure 5.7.2: Width modulation entirely on one side of a waveguide. This simple alteration doubles the feature size of the filters and makes fabrication significantly easier.

As can be seen in Fig. 5.7.1, when the filter bandwidth is down to 40 nm we are already pushing the limits of fabrication with \( \pm 5 \) nm of width modulation. A trivial way of improving the fabrication is to simply put all the width modulation on one side of the waveguide. This process increased the feature size by a factor of two. Fig. 5.7.2 shows a close up example of such a fabricated waveguide. There is no part of the theory we have developed in this chapter that would require a symmetric width modulation. All that is required is that the effective index be known at every point.

While a factor of two can be critical in achieving successful fabrication, it does not scale well as the bandwidth decreases. If, for instance, a filter was to be designed to shape a relatively long pulse with a duration of 100 ps, we might find that a sub-nm modulation would be necessary and a factor of two would not significantly make the fabrication easier. Motivated by a search for well controlled, small index modulations we can look to modulate a high index structure that only interacts with the evanescent field of the mode in the waveguide, rather than the large fields confined within the waveguide. Fig 5.7.3 shows an example of such a structure. Because the modulated “combs” can be placed arbitrarily far from the waveguides we can make the feature sizes arbitrarily large and still achieve extremely small modulations in the \( n_{eff} \) of the waveguide. Other possibilities would be to work in lower index materials such as polymers which will results in larger single-mode waveguides and larger
Figure 5.7.3: When very small modulations in the $n_{\text{eff}}$ are required it can become impossible to fabricate this directly by width modulating the waveguide. By designing a structure such as the one in this SEM we can make the feature size arbitrarily large because we are only interacting with the weak evanescent field of the guided mode.
modulations to achieve similar modulations in the $n_{eff}$.

5.8 Designing a Reconfigurable Filter

While we have demonstrated considerable promise in reverse-design filters based on width modulated waveguides, it would a significant step to be able to update the filters in real-time so that they could respond to different incoming signals or the need for a different pulse shape. The difficulty in implementing this methodology is in the extremely high density and number of individual modulations needed to construct a single filter. Essentially several hundred independently controlled segments of a waveguide, with a periodicity of $\lambda/(4n_{eff})$, are required. While accomplishing any one of the segments is trivial with current technology, when they are all put together the complexity becomes an issue.

Several different methods for locally controlling the index of a waveguide exist. These include liquid crystal cladding manipulation[99] and altering the index of ITO with an applied electric field[100]. But a design using $P^+IN^+$ junctions that
is feasible was developed. PIN junctions were chosen because it is possible to operate them with extremely high carrier density contrast, leading to a significant dynamic range in the possible refractive indices of the sections of the waveguide. Fig. 5.8.1 is a representation of the proposed scheme. The problem with this raw setup is also one of silicon’s strengths. Because of the indirect band-gap of silicon the free-carrier lifetime is extremely high, and at finite temperature this leads to a drift along the axis of the waveguide well outside of the intended localization of carrier infusion. On the other hand this effect is also extremely helpful in terms of the ability to inject such a high carrier density. It is important to keep in mind that passivating the silicon is required, we will later show that it is also beneficial to deposit an oxide to accomplish this for other reasons[101].

Fortunately there is a workaround for the issue of crosstalk. It is a simple matter of physically etching away the silicon in between the individual segments. One might think that this would lead to significant scattering events, the reality is that Bloch modes will develop, and as long as sufficient care is taken in impedance matching between the periodic waveguide and the continuous waveguide, little or no scattering will occur. It now becomes a problem of choosing the correct parameters so that the frequencies at which we want the filter to operate do not lie in the stop-band of the photonic crystal. Impedance matching can be accomplished by bringing in the periodic etching, in much the way we tapered our photonic crystals, slowly. For instance, if in the bulk the region will be 50 nm wide. We can start with a 25 nm trench and over the course of 5 periods increase the trench width to 50 nm. This will all but eliminate any reflection due to the periodic etching.

This also leads to an interesting change in the design. Instead of looking for an $n_{\text{eff}}$ at every point along the waveguide, we have now changed the problem to a number of discrete unit cells, each slightly perturbed from the underlying Bloch mode. Instead of a 2-D eigenfrequency calculation of the cross section we must perform a 3-D eigenfrequency calculation with periodic boundary conditions, calculate the associated k-vector for each frequency and from this relationship extract the phase velocity. This is still a feasible mapping and need be done only
once for a given waveguide geometry.

An interesting problem arises. Because we want the periodicity to be $\frac{\lambda}{(4n_{\text{eff}})}$, with $n_{\text{eff}} = \frac{k\lambda}{\pi}$, the periodicity required will change depending on the k-vector of the frequency in question. Therefore, we need to choose a periodicity based on our center-wavelength. We pick our waveguide dimensions and periodicity such that our center-wavelength (frequency) has a k vector exactly in the center of the unit-cell ($\frac{\pi}{2a}$, where $a$ is the period).

In addition to etching the trenches between the parallel $P^+IN^+$ junctions, it is also advantageous to use a process such as atomic layer deposition (ALD) to fill in these small spaces. Fig. 5.8.2 shows what the structure might look like with etched trenches filled with an ALD oxide. The benefits of this are two-fold: first the index of the material that is deposited can shift the band-structure up and down in frequency, allowing for an extra degree of freedom in the search for a “sweet spot”, where the center-wavelength has the appropriate k-vector; second, passivating the surface of the silicon can lower the surface recombination
velocity[101], making it easier to inject high carrier densities. Simulations suggest that the index of Al_{x}O_{y} matches silicon well and creates an easily accessible dielectric band around 1550 nm. Given that this is also an excellent material for ALD it would appear to be a great choice. However, there is a wide parameter space in terms of period, waveguide width, SOI thickness, cladding to name a few. This gives us great freedom in choosing an oxide to fill the trenches with and for passivating the silicon.

Now that all of the elements are in place, let us look at the actual implementation of a working reconfigurable filter. We assume that we can inject enough carriers to induce a shift of \( \Delta n = \frac{4x10^{-2}}{[102]} \). Then based on our desired reflectivity we are after we have a choice. If we want a high reflectivity then we would opt for a longer, narrow-band device that will suppress the amplitude of modulation required to create the requisite reflectivity. If we are willing to have a lower reflectivity then we can keep our device short and try to wring as many free-carriers out of the PIN junctions as possible.

In the large reflectivity range we are faced with many hundreds of unit cells. This means that on-chip we need to either locate sufficient bond pads for external electronic interface, or we need to fabricate on-chip drivers and logic to run so many devices independently. This is beyond our modest fabrication facilities at Harvard.

The astute digital signal processor will have no doubt identified that as long as a linear phase relationship is maintained, the filters are mirror-symmetric about the central point. This is a loss by a factor of two in the number of required independent drivers; however, this factor of two does not significantly alter the complexity.

One thing to keep in mind is that these devices would not be extremely fast. The high density of the PIN junctions will lead to large capacitive coupling so the operation would be in a more quasi-static regime where a desired configuration could be obtained but then be used for a long period of time. PIN junctions also have the disadvantage of slow recovery. While a large reverse bias can be used to remove carriers, the recovery is still smaller that of depletion based PN junctions.
The reader is asked to keep in mind that this is merely a single proposal to show that reconfiguration of these filters is feasible. Future work could easily improve on this design or utilize different methods of index control to obtain a more efficient or effective method of achieving reconfigurability.

5.9 Outlook and Utility

We have shown that a reverse design method allows for extremely fast design of arbitrary response filters in high index contrast waveguides. These filters also control the phase and can be used to shape ultra-fast pulses. This ultra-fast pulse shaping shows great promise in simplifying the process of pulse shaping by drastically reducing the size and complexity of such apparatus. We showed the TM mode allows us to get excellent results in terms of the fabricated filters matching their design goals.
Conclusion

Over the course of this work we have developed numerous methods for controlling light on-chip. We examined PCNCs that allowed us to localize light to a high degree. These PCNCs were excellent candidates for sharp frequency filters when strongly coupled to waveguides. We showed that their design was well developed and was at the point where heavy computation was no longer required.

We also showed that due to the one dimensional aspect of the photonic crystal it is possible to fabricate high quality factor optical cavities in both TE and TM polarizations. This opens up the door for many nonlinear optics experiments as well as allowing us to probe a wider range of dynamics on-chip.

We presented a proposal for a PCNC based EOM that made use of the unique electrical and optical properties of graphene. We showed a viable fabrication technique that coupled with high quality graphene and well designed PCNCs could lead to truly amazing performance levels in terms of energy and speed.
Our next step took us toward examining the mechanics of PCNCs. We saw how they responded to applied forces and showed that it was advantageous to use coupled PCNCs. The optical supermodes of these coupled PCNCs showed amazing potential for tunability because of their strong dependence on the separation between the PCNCs.

Next we examined various methods of achieving mechanical reconfiguration of these filters resulting in control of the optical frequency of some of the modes. We analyzed a capacitive method that allowed coupled PCNCs to be actuated with significant dynamic range at low powers, while allowing for speeds that pushed the mechanical limits.

After examining the capacitive force we considered a few aspects of the gradient force. We showed how electrical and optical signals could interact at high speeds and extremely low powers. Then we considered moving away from integrated electronics and inducing motion purely through optical means.

We wrapped up by deriving a new class of waveguide filters that could take on arbitrary shapes. The spectral freedom of these filters is matched only by the efficiency of the computation in designing them. We showed these filters to be excellent candidates for on-chip ultra-fast pulse shaping. We then proposed a device that would allow real-time reconfiguration in response to applied electrical fields.

Of the cornucopia photonic circuitry being developed today we examined but a small sliver. However, we did see promising results in terms of developing low power mechanical switches and opening new vistas in arbitrary filters.
References


