Photonics for technology: topological microwave circuits, chip-scale LIDAR, and optical neural networks

by

Scott Alexander Skirlo

B.S. MIT, Cambridge, MA, USA 2013

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Author .... Signature redacted

Department of Physics
May 15th, 2017

Certified by .... Signature redacted ......

Marin Soljačić
Professor of Physics
Thesis Supervisor

Accepted by .... Signature redacted

Prof. Nergis Mavalvala
Associate Department Head

Photonics for technology: topological microwave circuits, chip-scale LIDAR, and optical neural networks

by

Scott Alexander Skirlo

Submitted to the Department of Physics
on May 15th, 2017, in partial fulfillment of the
requirements for the degree of
DOCTOR OF PHILOSOPHY

Abstract

This thesis focuses on a wide range of contemporary topics in modern electromagnetics and technology including topologically protected one-way modes, integrated photonic LIDAR, and optical neural networks.

First, we numerically investigate large Chern numbers in photonic crystals and explore their origin from simultaneously gapping multiple band degeneracies. Following this, we perform microwave transmission measurements in the bulk and at the edge of ferrimagnetic photonic crystals. Bandgaps with large Chern numbers of 2, 3, and 4 are present in the experimental results which show excellent agreement with theory. We measure the mode profiles and Fourier transform them to produce dispersion relations of the edge modes, whose number and direction match our Chern number calculations. We use these waveguides to realize reflectionless power splitters and outline their application to general one-way circuits.

Next we create a new chip-scale LIDAR architecture in analogy to planar RF lenses. Instead of relying upon many continuously tuned thermal phase shifters to implement non-mechanical beam steering, we use aplanatic lenses excited in their focal plane feeding 1D gratings to generate discrete beams. We design devices which support up to 128 resolvable points in-plane and 80 resolvable points out-of-plane, which are currently being fabricated and tested. These devices have many advantages over conventional optical phased arrays including greatly increased optical output power and decreased electrical power for in-plane beamforming.

Finally we explore a new approach for implementing convolutional neural networks through an integrated photonics circuit consisting of Mach-Zehnder Interferometers, optical delay lines, and optical nonlinearity units. This new platform, should be able to perform the order of a thousand inferences per second, at μJ power levels per inference, with the nearest state of the art ASIC and GPU competitors operating 30 times slower and requiring three orders of magnitude more power.

Thesis Supervisor: Marin Soljačić
Title: Professor of Physics

Acknowledgments

I've been at MIT for a long time now (too long to feel comfortable mentioning out loud). It's been a crazy ride through a lot of subjects and places, and there are many people who I have to thank from along the way.

I am very grateful for Marin's guidance throughout my PhD. He allowed and encouraged me to explore a wide range of subjects. He was also very patient and understanding when I encountered difficulties.

In addition I'm very grateful for the guidance of my mentors from over the years including Falk Eilenberger and Ling Lu.

I greatly appreciate the assistance from my collaborators on many projects including Qinghui Yan, Yuichi Igarashi, Josue Lopez, Henry Everitt, Cheri Sorace, Ido Kaminer, Jeff Herd, Paul Juodawlkis, Simon Verghese, Yi Yang, Yichen Shen, Ognjen Ilic, Avi Massuda and Hangameh Bagherian, just to name a few.

One of my thesis committee members, Prof. John Joannopoulos provided critical advice and guidance on numerous projects. I also appreciate all of the considerate feedback from Prof. Lindley Winslow, another member of my thesis committee.

I was so lucky to have many good friends during my PhD including Nikhil, Nick, Ryan, Ben, Amanda, Eric, Jen, Belle, Aaron, Nancy, Reid and others who I learned so much from and who made my time here unforgettable.

I'm thankful for the guidance of my high school teachers, John Bell, John Dell, and Jonathan Osbourne from TJHSST who instilled in me a love for engineering, science, and math.

Most importantly I appreciate the love and support from my wife, Victoria Messuri, who has always helped me to keep my priorities straight and who I believe is my greatest discovery to date.

Contents

List of Figures

Chapter 1

Introduction

Although Maxwell's equations were complete over 150 years ago, modern advances in materials, computing and fabrication have enabled continuous scientific development that is to this day rapidly evolving in beautiful and completely unexpected ways. Often new discoveries have significant technological ramifications which transform society and lead to better science and even better technology. I find this interplay fascinating and while I have studied many different subjects over the course of my PhD, the common thread among them has been the scientific investigation of new systems and ideas which can serve as the basis of future technology, whether it be one-way microwave circuits, or integrated LIDAR.

I have been fortunate to work in a group which invented photonic crystals, periodic wavelength-scale structures which manipulate the flow of electromagnetic radiation. At the start of this thesis I examine a type of photonic crystal composed of ferrimagnetic materials, which are typically used to create RF circulators because of their nonreciprocal behavior under applied magnetic field. In chapter one, I describe the discovery of bandgaps in these crystals which support multiple one-way modes at their edge that result from the simultaneous opening of many band degeneracies. In chapter two, we experimentally realize these systems and map a range of bandgaps with Chern numbers ranging from -4 to 3. Finally, we create a topological power splitter circuit and present the first ever measurement of the dispersion-relations of multiple one-way edge modes.

Next, I delve into the rapidly growing field of integrated LIDAR. Advances in nanofabrication coupled to accelerating commercial and military demand for long range 3D depth sensors has driven a surge of interest in non-mechanical optical beamsteering. I provide my contribution to this rapidly growing field in the form of a new lens-based integrated photonics LIDAR in chapter three. This novel approach was inspired by passive RF beamforming and has many significant advantages over its more traditional phased array counterparts. Here I also describe analytic models, ray tracing calculations, and simulations used for the design and characterization of our system. These devices are currently being fabricated and tested by collaborators at MIT Lincoln Labs.

Interest in LIDAR has been driven in part by the development of advanced machine learning algorithms which require massive amounts of real world data. One such class of algorithms known as convolutional neural networks (CNNs) have in the past five years become extremely effective at classifying image data. In chapter four, we explore a concept for implementing CNNs optically through integrated photonics. The system we outline potentially lowers the power required per inference by several several orders of magnitude relative to state-of-the-art digital solutions. This technology may eventually enable the application of CNNs to a significant fraction of the image and video data generated by the Internet.

Chapter five focuses on the development of new alignment marks for e-beam lithography, which can potentially improve the fabrication tolerance of devices like those discussed in chapter three and four. We derive mathematical bounds on the performance of these types of markers, and identify the best performing ones through exhaustive search.

Finally chapter six reviews my research in Germany on cavity optical pulse extraction, which uses a trigger pulse at one frequency to extract energy from an optical cavity pumped at a different frequency through optical nonlinearity. I develop a new technique to numerically integrate the nonlinear coupled differential equations governing this phenomena and explore cavity designs to create frequency combs.

Chapter 2

Multimode one-way waveguides of large Chern numbers

2.1 Abstract

Current experimental realizations of the quantum anomalous Hall phase in both electronic and photonic systems have been limited to a Chern number of one. In photonics, this corresponds to a single-mode one-way edge waveguide. Here, we predict quantum anomalous Hall phases in photonic crystals with large Chern numbers of 2, 3 and 4. These new topological phases were found by simultaneously gapping multiple Dirac and quadratic points. We demonstrate a continuously tunable power splitter as a possible application of multimode one-way waveguides. All our findings are readily realizable at microwave frequencies. This work is reported in [7].

2.2 Main text

Like the integer quantum Hall effect, the quantum anomalous Hall effect (QAHE) has topologically-protected chiral edge states with transverse Hall conductance $\frac{Ce^2}{h}$, where C is the Chern number of the system. Unlike the integer quantum hall effect, the electronic QAHE requires no external magnetic field and has no Landau levels. The first theoretical model of the QAHE was proposed by Haldane in a honeycomb lattice [8]. Haldane subsequently extended this electronic topological phase to photonics [9], where the Chern number equals the number of one-way waveguide modes. This photonic analog to the QAHE was experimentally realized in a gyromagnetic photonic crystal at microwave frequencies [2]. Very recently the electronic QAHE has also been experimentally demonstrated in magnetic topological insulators [10]. Although the Chern number can in principle take arbitrary interger values, in the very large body of work on the QAHE all photonic and electronic realizations were limited to $|C| = 1$. Consequently, finding systems with $|C| \geq 1$ is a fundamental scientific goal in studying topological phases. The search for higher Chern numbers also has practical value. In electronics, having more chiral edge states would greatly reduce the contact resistance for circuit interconnects [11, 12]. For photonic applications, multimode one-way waveguides have an increased mode density and coupling efficiency. In addition they enable new devices such as reflectionless waveguide splitters, combiners or even "one-way photonic circuits".

In this letter, we report topological photonic bandgaps of large Chern numbers $|C| = 2$, 3 and 4 using ab-initio calculations. First, we describe two approaches to implement bulk bands which have multiple pairs of Dirac cones and multiple quadratic points. The simultaneous opening of these point degeneracies generates Chern numbers of large magnitudes. Next we present a topological gap map using the parameters corresponding to Yittrium Iron Garnet, a common microwave gyromagnetic material, to demonstrate that all examples in this letter are readily realizable. Finally we put forward an implementation of an adjustable power splitter as an illustration of a possible practical application and as a way to experimentally verify the existence of the $C = 2$ phase. The ideas and results presented can also be applied to other topological photonic systems [13–19].

It is now known [1, 20] that each of the two bands connected by one pair of Dirac point degeneracies can acquire nonzero quantized Berry flux, when gapped by T-breaking. Each Dirac point contributes π Berry flux for a total of 2π in each band. Because 2π Berry flux is exchanged across the bandgap, the Chern number associated with the gap, which we call the "gap Chern number", is ± 1. A general way to calculate the gap Chern number ($C_{gap} = \sum_i C_i$) is to sum the Chern numbers of all bands below the bandgap [21]. Since a quadratic point consists of two Dirac points [22], the Berry flux exchanged is 2π and $C_{gap} = \pm 1$ when

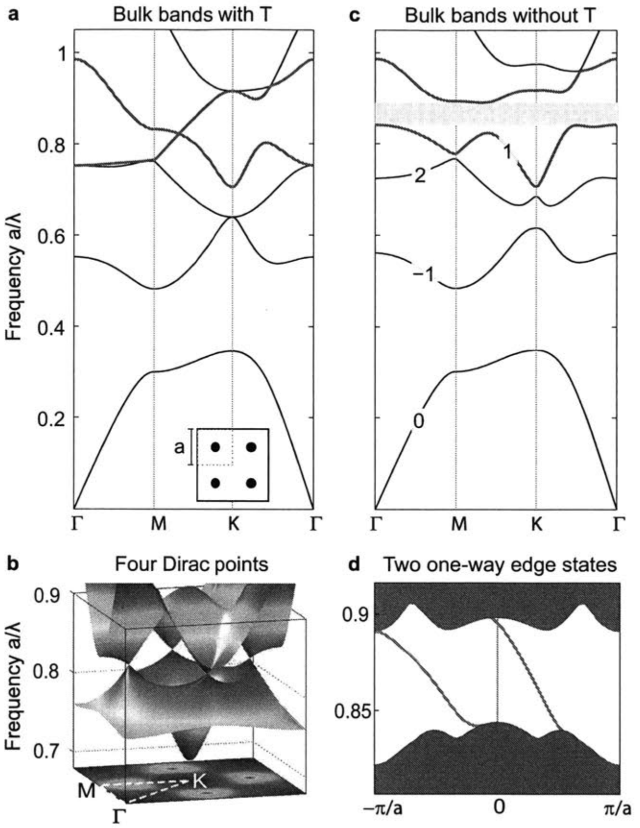


Figure 2-1: Bulk and edge TM bandstructures showing two one-way edge states ($C_{gap} = 2$) obtained from four Dirac points. The photonic crystal is a square lattice of rods with a radius of $0.13a$, $\epsilon = 13$ and $\mu = 1$. The T-breaking perturbation corresponds to adding $\mu_{12} = -\mu_{21} = 0.40i$ to the rods. a) Bulk bandstructure showing the Dirac point along $M - K$. The lower inset illustrates the lattice geometry. b) Four Dirac cones between the 4th and 5th bands plotted in the whole Brillouin zone. c) Bulk bandstructure under T-breaking perturbation opens a 5.5% complete gap highlighted in yellow. Each band is labeled with its Chern number. d) Two gapless one-way edge states (red) appear in the projected edge band diagram when the bulk is terminated by a metallic boundary.

the quadratic degeneracy is gaped by T-breaking.

When an edge is formed by joining two materials with bandgaps overlapping in frequency, the number of one-way edge states is equal to the difference of gap Chern numbers across the interface [21]. If one of the materials is topologically trivial ($C_{gap} = 0$) like metal or air, the other gap Chern number determines the number of topologically-protected one-way edge states. When a gapped Dirac or quadratic point system is interfaced with a trivial bandgap, the one-way waveguide formed is single-mode ($|C_{gap} = 1|$).

To create multimode one-way wavegudies, it is necessary to increase the magnitude of C_{gap} or, equivalently, increase the Berry flux exchanged between the bands. This requires the involvement of more pairs of Dirac points and more quadratic points. In the rest of this letter, we describe two strategies to achieve this goal. [1]

Every two-dimensional (2D) photonic crystal shown in the main text is a square lattice of gyromagnetic rods of lattice constant a. Results for the transverse magnetic (TM) modes are presented. The T-breaking perturbation is implemented by applying a static magnetic field perpendicular to the 2D plane. As a result, the gyromagnetic material develops off-diagonal imaginary terms ($\mu_{12} = -\mu_{21} \neq 0$) in the magnetic permeability. The full-wave simulations are done using the RF module of COMSOL Multiphysics. The Chern numbers of a single isolated band or multiple degenerate bands are calculated following the approach described in Ref. [23].

The first approach to obtain large gap Chern numbers is to increase the symmetry of the system so that point degeneracies can come in large multiples at the same frequency. Under T-breaking, the Berry flux contributed from each symmetry-related point-degeneracy is identical to the rest. Consequently, they add constructively, which increases C_{gap}. For example, in a system of four-fold rotational symmetry (C_4), Dirac points have to come in multiples of four. For two bands containing only Dirac degeneracies, the exchange of Berry flux comes in multiples of 4π and the gap Chern numbers are even.

An example of $C_{gap} = 2$ is shown in Fig. 3-11. Fig. 3-11a and Fig. 3-11b show four Dirac cones found in a system of four-fold rotational symmetry. These Dirac cones are between

---

[1] Note that a trivial way to create multimode one-way edge waveguides is to interface two identical topological gyromagnetic photonic crystals with their magnetic fields reversed. In this case, the number of one-way edge states is double the gap Chern number.

the 4th and 5th bands along the K-M line in the Brillouin zone. They are well-isolated from other bands. In Fig. 3-11c and 3-11d we open a complete gap by applying a T-breaking perturbation. In Fig. 3-11d, there are two edge states present when this bulk photonic crystal is interfaced with a metallic boundary (a perfect electric conductor).

By increasing the four-fold rotational symmetry to six-fold (C_6) in a triangular lattice, we found six Dirac cones and three one-way edge states ($C_{gap} = 3$). These results are presented in the supplemental material.

The second approach to obtain large gap Chern numbers is to tune multiple symmetry-unrelated point degeneracies to occur in the same frequency range, so that when a sufficient T-breaking perturbation is applied, a complete gap opens. Unlike the previous case of symmetry-related degeneracies, the Berry flux associated with gaping symmetry-unrelated degeneracies can either add constructively or destructively. When all Berry fluxes are of the same sign, large Chern numbers can occur. Using this approach, we obtained another example of $C_{gap} = 3$ shown in the supplemental material. There, we brought one quadratic point to a frequency in the vicinity of four Dirac points to obtain three one-way edge states by gapping all of them.

The example of $C_{gap} = 4$ is shown in Fig. 2. We shifted two quadratic points to frequencies near four Dirac points. Shown in Fig. 2a and 2b, there are four Dirac points along M-Γ, one quadratic point at the K point and another one at the Γ point. In Fig. 2c and 2d, the T-breaking opens a complete gap between the 6th and 7th bands, inside which four one-way edge states appear. The mode profiles of the four edge modes are plotted in Fig. 2e at a chosen frequency. They have different decay lengths into the bulk from the metallic boundary, determined largely by how close they are to the band edge in the band diagram.

All the findings above are readily realizable at microwave frequencies using gyromagnetic materials such as Yittrium Iron Garnet (YIG), which have been used in experimental studies of single-mode one-way edge waveguides. We note that the examples from the prior figures do not necessarily correspond to physical materials, since they were selected to show clear openings of the band degeneracies. To facilitate the eventual experimental realizations of multimode one-way edge waveguides, we construct, in Fig. 3, a topological gap map of

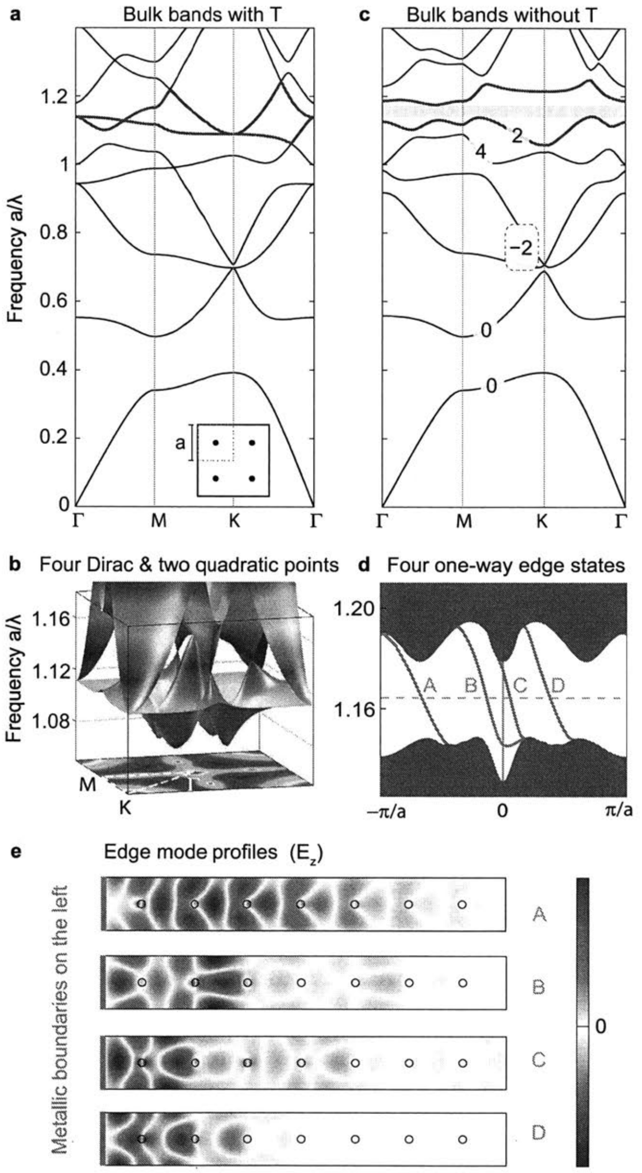


Figure 2-2: Bulk and edge TM bandstructures showing four one-way edge states ($C_{gap} = 4$) obtained from four Dirac points and two quadratic points. The photonic crystal is a square lattice of rods with a radius of $0.07a$, $\epsilon = 24$ and $\mu = 1$. The T-breaking perturbation corresponds to adding $\mu_{12} = -\mu_{21} = 0.40i$ to the rods. a) Bulk bandstructure showing the Dirac point along $\Gamma - M$ and two quadratic points at Γ and K respectively. The lower inset illustrates the lattice geometry. b) Four Dirac cones and two quadratic points between the 6th and 7th bands plotted in the whole Brillouin zone. c) Bulk bandstructure under T-breaking perturbation opens a 3.0% complete gap highlighted in yellow. Each band is labeled with its Chern number. The Chern number of degenerate bands is circled by a dotted line. d) Four gapless one-way edge states (red) appear in the projected edge band diagram when the bulk is terminated by a metallic boundary. e) Mode profiles of the four edge modes close to the metallic interfaces are plotted at the same frequency dotted in d).

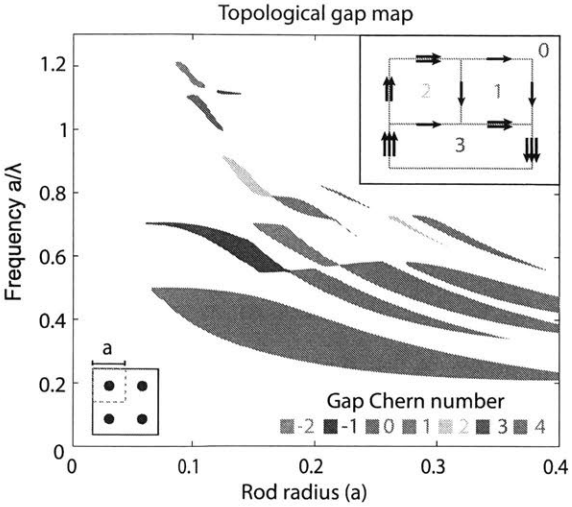


Figure 2-3: The TM topological gap map of a photonic crystal consisting of a square lattice of YIG rods under an applied static magnetic field. The gap Chern number labels each complete bandgap found. The previously reported single-mode one-way waveguide [1, 2] corresponds to the blue bandgap of $C_{gap} = -1$. The left lower inset illustrates the lattice geometry with a period of a. The upper right inset illustrates a 2D one-way circuit by interfacing various photonic crystals of different gap Chern numbers.

photonic crystals made of YIG rods. In the calculation, the material properties used for the YIG rods are $\mu = \begin{pmatrix} 0.84 & 0.41i & 0 \\ -0.41i & 0.84 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $\epsilon = 15$. This corresponds to the response of YIG at 14.5 GHz at 2020 Gauss of static magnetic field [24]. We scanned the radius of the YIG rods and mapped out the frequencies of all the complete TM bandgaps found up to the 7th bulk band. We did not find interesting features when going through higher bands. The area of each bandgap is colored according to its gap Chern number (C_{gap}). Shown in Fig. 3, a wide range of Chern numbers are found from -2, to +4. Negative Chern numbers represent the negative group velocities of the corresponding one-way edge states. We note that although the calculations here are done in 2D for TM modes, the designs can exactly be translated to 3D metallic waveguides in experiments [2, 25–28].

The material dispersion does not change the main features in the map, if the material operation frequency is placed at the center of the gap of interest. Taking into account dispersion from the gyromagnetic resonance [24], the widest gap for $C_{gap} = 2$ decreased from 5.4% to 3.5%, for $C_{gap} = 3$ from 3.6% to 2.7%, and finally for $C_{gap} = 4$ from 2.6% to 2.1%.

Since in general there are more band crossings in higher order bands, larger gap Chern numbers are expected to be found there. High-order bandgaps tend to be smaller due to the greater density of states there, so we expect to find larger gap Chern numbers in smaller gaps. These trends are clearly observed in Figs. 1,2, 3, and in the supplemental material.

The above multimode one-way waveguides significantly increase the transport channels that are topologically protected. We show as an example, in the appendix, a total number of eight one-way modes inside an edge waveguide constructed from domains of +4 and -4 gap Chern numbers. Compared to their single-mode counterparts, these one-way multimode waveguides have a much larger density of states which gives a much higher input coupling efficiency.

The wide range of gap Chern numbers found in the topological gap map offers the opportunity to make a "topological one-way circuit" [29, 30]. This idea is illustrated as an inset of Fig. 3, where bulk domains of different gap Chern numbers are joined together. One-way edge states flow around their interfaces. The number of the one-way waveguide modes equals the difference of the gap Chern numbers across the interface. At the junction between three bulk domains, one-way edge states merge together or branch off, enabling new device functionalities as signal combiners and splitters immune to backscattering from manufacturing imperfections.

In Fig. 4, we present a concrete design of a tunable power splitter showcasing a possible application of the multimode one-way waveguides. The splitter is at the junction beneath a metal wall between two domains with gap Chern numbers of +2 and +1. This junction couples the power from one multimode waveguide on the left into single-mode waveguides on the right (channel 1) and bottom (channel 2). A metal scatterer in the multimode waveguide tunes the power splitting between channel 1 and channel 2.

In the simulations, we placed a mono-frequency point source inside the left waveguide formed by the metal and $C_{gap} = 2$ domain. The source excites a linear combination of the two one-way modes. These modes propagate to the right and are scattered by the metal scatterer. Since one-way modes cannot backscatter, they only scatter into each other, changing their amplitudes and phases. The height of the metal scatterer controls the total mode profile at the junction, and consequently the power splitting between channel 1 and channel 2. In

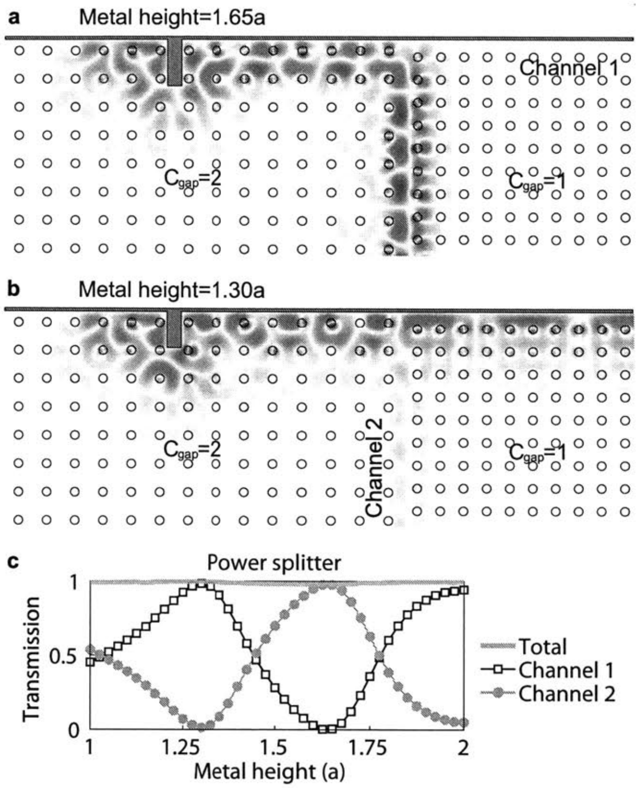


Figure 2-4: A power splitter implemented with $C_{gap} = 2$ and $C_{gap} = 1$ gyromagnetic photonic crystals bordered on the top by a metallic wall. A point source with frequency $\frac{a}{\lambda} = 0.82$ couples light into the multimode waveguide to the left of the metal scatterer. The unit cell size of the $C_{gap} = 2$ lattice is a, the unit cell size of the $C_{gap} = 1$ lattice is $0.805a$ and the rod radius for both lattices is $0.15a$. The operating parameters are the same as Fig. 3, so a in real units is 1.7 cm. a) For a metal scatterer with height $1.65a$, the majority of the light proceeds through channel 2 (bottom waveguide). b) For a metal scatterer with height $1.30a$, the majority of the light proceeds through channel 1 (right waveguide). c) The transmission to each waveguide as a function of metal scatterer height.

Fig. 4c, we plot the transmission into each channel as a function of the height of the metal scatterer. Since there is no reflection or absorption, the total power efficiency of the splitter is always 100%. Between a metal scatterer height of a and $2a$, the transmission into each channel oscillates between nearly 0 and 1. In Fig. 4a and 4b, we present two field profiles in which the transmission is maximized into either channel 1 or channel 2. We found that for a gyromagentic resonance linewidth $\Delta H = 0.5Oe$ and a dielectric loss tangent of 0.0002, that the attenuation length in the system varied between $100a$ and $400a$ depending on the modal profile. These attenuation distances are much longer than the dimensions of the device we propose here.

In the above tunable splitter, we placed the metal scatterer far away from the junction so the tuning from the metal scatter and the splitting at the junction are spatially seperated. This type of tuning is only possible when the scatterer is inside a multimode waveguide. If the left waveguide had only one mode, the metal scatterer could not have changed the field distribution at the junction, because the mode profile in a single-mode one-way waveguide cannot be changed by any scatterers far away. Therefore the experimental realization of the above tunable splitter can verify the existance of the large Chern number.

In conclusion, we predicted photonic analogs of the QAHE with large Chern number (up to 4) and constructed multimode one-way edge waveguides (up to 8 modes). The implementations are readily experimentally realizable using a square lattice of YIG rods at microwave frequencies. Using the discovered multimode one-way waveguides, we also predicted an adjustable power splitter with unity efficiency. Since the first demonstration of the quantum (anomalous) Hall phase in photonic crystals, there have been a wide range of theoretical [30–36] and experimental [25–28] efforts to investigate systems with single-mode one-way edge states. We expect our findings of large Chern numbers and multimode one-way waveguides to create even wider opportunities in topological photonics. Finally, our approach to create topological bandgaps of large Chern numbers can also be applied to electronic systems.

2.3 Appendix

2.3.1 Numerical Methods

We can rearrange Maxwell's equation into a matrix equation of the following form:

$$\begin{pmatrix} 0 & +i\vec{\nabla}\times \\ -i\vec{\nabla}\times & 0 \end{pmatrix} \begin{pmatrix} \vec{E} \\ \vec{H} \end{pmatrix} = \omega \begin{pmatrix} \epsilon & 0 \\ 0 & \mu \end{pmatrix} \begin{pmatrix} \vec{E} \\ \vec{H} \end{pmatrix} \tag{2.1}$$

Further we can identify $\hat{A} = \begin{pmatrix} 0 & +i\vec{\nabla}\times \\ -i\vec{\nabla}\times & 0 \end{pmatrix}$, $\hat{B} = \begin{pmatrix} \epsilon & 0 \\ 0 & \mu \end{pmatrix}$ and $|\phi\rangle = \begin{pmatrix} \vec{E} \\ \vec{H} \end{pmatrix}$. Since $\hat{A}^\dagger = \hat{A}$ and $\hat{B}^\dagger = \hat{B}$, we know that the following equation is a generalized Hermitian eigenvalue problem:

$$\hat{A}\,|\phi\rangle = \omega\hat{B}\,|\phi\rangle \tag{2.2}$$

As opposed to the regular eigenvalue problem where the inner product is defined as $\langle\phi_i|\phi_j\rangle$, here the inner product has to include the second Hermitian matrix: $\langle\phi_i|\,\hat{B}\,|\phi_j\rangle$ to have the expected ortho-normalization properties.

The chief task to numerically solve the system is to efficiently and correctly form the matrices $\hat{A}$ and $\hat{B}$, while also implementing correct periodic or metallic boundary conditions. The first step in this process is picking a convention for how to map the 2D real space cell into a 1D vector. Here we picked that each element of the N columns appears row by row for M rows. This looks like the following for a single component of the tensor ϵ:

$$[\epsilon_{zz}] = \begin{bmatrix} \epsilon_{zz}^{11} & 0 & 0 & \dots & 0 & \dots & 0 \\ 0 & \epsilon_{zz}^{21} & 0 & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \epsilon_{zz}^{N1} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & \dots & \epsilon_{zz}^{NM} \end{bmatrix}$$

Where ϵ_{zz}^{ij}, indicates the value of ϵ_{zz} in the i-th column and the j-th row.

Besides implementing the material tensors, the derivatives also need to be discretized appropriately. We implemented our discretization beginning with the standard difference matrix on the Yee-grid:

$$[D_x] = \frac{1}{\Delta x}\begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ 0 & 0 & 1 & -1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \exp(-ik_xL_x) & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Here we implemented periodic boundary conditions through the bottom left term which tracks the phase accumulated for a Bloch-mode of the form $u_k(x)\exp\left(-i\vec{k}\cdot\vec{x}\right)$. For numerical solutions of the bandstructure we instantiated a single $\hat{A}$ given the material geometry, and generated separate $\hat{B}$ for every k-point in the BZ we were interested in solving for.

To properly implement the discretized derivative using a consistent convention, we have to ensure that the proper elements between the columns and rows are matched. For the x-derivatives, we just need $[D_x]$ acting on each row, so we just copy this along the diagonals with a tensor product with an M by M identity matrix: $I_{MM}\otimes[D_x] = [M_x]$. The total y-difference matrix can be written similarly, except we swap the identity matrix to the second position, so that the difference between the *rows* of the Bloch mode is calculated: $[D_y]\otimes I_{NN} = [M_y]$.

We can use these matrices and specialize for the case of TM modes in our array of gyromagnetic rods, which have E_z, H_x and H_y components:

$$i\begin{pmatrix} 0 & -[M_y] & [M_x] \\ [M_y] & 0 & 0 \\ -[M_x] & 0 & 0 \end{pmatrix}\begin{pmatrix} E_z \\ H_x \\ H_y \end{pmatrix} = \omega\begin{pmatrix} [\epsilon_{zz}] & 0 & 0 \\ 0 & [\mu_{xx}] & +i[k] \\ 0 & -i[k] & [\mu_{xx}] \end{pmatrix}\begin{pmatrix} E_z \\ H_x \\ H_y \end{pmatrix} \tag{2.3}$$

Metallic boundary conditions can be enforced a number of ways, but we found that the easiest and most robust way was simply to set the rows of the matrix $\hat{A}$ to zero which were required by metallic boundaries on a particular edge.

We wrote a computer program which efficiently generated the $\hat{A}$ and $\hat{B}$ matrices based on specified locations for multiple dielectric shapes and cell geometry, for unit cell and supercell calculations, and then used the eigs method from MATLAB to find the smallest eigenvalues. Using this numerical finite-difference frequency-domain (FDFD) solution, we were able to calculate the photonic crystal bandstructure and corresponding eigenvectors for the gyromagnetic photonic crystal rapidly.

2.3.2 Chern number calculation

We can extract topological information about the bandstructure by computing the Chern number. Although well described in the literature, we briefly describe some key results and our method for numerically computing the Chern number. We begin by defining the Berry Connection:

$$\vec{A}_{n,n}(\vec{k}) = -i \left\langle u_{n,\vec{k}} \middle| \vec{\nabla}_k \middle| u_{n,\vec{k}} \right\rangle \tag{2.4}$$

We can integrate this quantity around the edge of the 1st BZ to get the Berry phase.

$$C_n = \frac{1}{2\pi} \oint_{BZ} d\vec{k} \cdot \vec{A}_{n,n}(\vec{k}) \tag{2.5}$$

Where c_n is an integer called the Chern number associated with the n-th band. To show that c_n has to be an integer, we use the observation that u_k on the BZ boundary can be related to each other through Bloch's theorem: $u_k = u_{k+G} \exp(i\phi_g)$. Using this we can write:

$$\frac{1}{2\pi} \oint_{BZ} d\vec{k} \cdot \vec{\nabla}_k \phi_k = c_n \tag{2.6}$$

Which must be an integer since ϕ_k is a single valued function mod 2π. This version of the Chern number calculation provides a simple intuition for what the Chern number does, it counts the number of phase vorticies contained within the BZ. Alternatively using Stokes theorem, we can rewrite the Chern number as an integral over the BZ.

$$\frac{1}{2\pi i} \int \int_{BZ} d^2 z (\vec{\nabla}_k \times \vec{A}_{n,n}(\vec{k})) \cdot \hat{z} \tag{2.7}$$

Where $\vec{\nabla}_k \times \vec{A}_{n,n}(\vec{k})$ is called the Berry Flux. This form of the Chern number lays the foundation for our numerical calculation. Fukui et al [23] showed that a gauge-invariant form of the above equation could be written in terms of inner products between u_k around small placates in the BZ. Specifically we find the Berry flux contribution from a small placate as:

$$F^i_{n,m} = \langle u^{i,1}_n | \hat{B} | u^{i,2}_m \rangle \langle u^{i,2}_n | \hat{B} | u^{i,3}_m \rangle \langle u^{i,3}_n | \hat{B} | u^{i,4}_m \rangle \langle u^{i,4}_n | \hat{B} | u^{i,1}_m \rangle \tag{2.8}$$

Here we have simplified the index notation such that i addresses a given corner of a grid and $1, 2, 3, 4$ indicate elements forming the placate. Note that we have included $\hat{B}$ in the inner product for our specific system whose eigenvectors come from a generalized Hermitian eigenvalue problem. $F^i_{n,n}$ essentially extracts the phase accumulated by making the small loop and can be thought of as $\exp(i(\theta_{1,2} + \theta_{2,3} + \theta_{3,4} + \theta_{4,1}))$. To calculate the total Chern number, we take the log and imaginary part of each $F^i_{n,n}$ to find the phase contributed by a single loop, and sum them to find the total Berry flux:

$$c_n = \frac{1}{2\pi} \sum_i Im\left[\log\left(\frac{F^i_{n,n}}{|F^i_{n,n}|} \right) \right] \tag{2.9}$$

The only other case we have to discuss applies to band degeneracies. When this happens, we cannot compute the Chern number for a single band anymore. We have to calculate the more general quantity:

$$F^i_{BD} = det(F^i_{l,m}) \tag{2.10}$$

Where $F^i_{l,m}$ is a matrix formed where we compute the above equation between pairs of entangled bands in our set. This method was useful for rapidly calculating the winding number (or the sum of the Chern numbers of all bands) below the bandgap, since this computation automatically and correctly sums the Chern numbers of the lower bands regardless of whether sets of them are entangled or not. This feature was very useful in practice for rapidly identifying non-trivial bandgaps in complicated band structures.

2.4 Acknowledgements

We acknowledge Liang Fu and Zheng Weng for discussions, and Xiangdong Liang and Wenjun Qiu for numerical assistance. S. S. was supported by MIT Tom Frank Fellowship. L. L. was

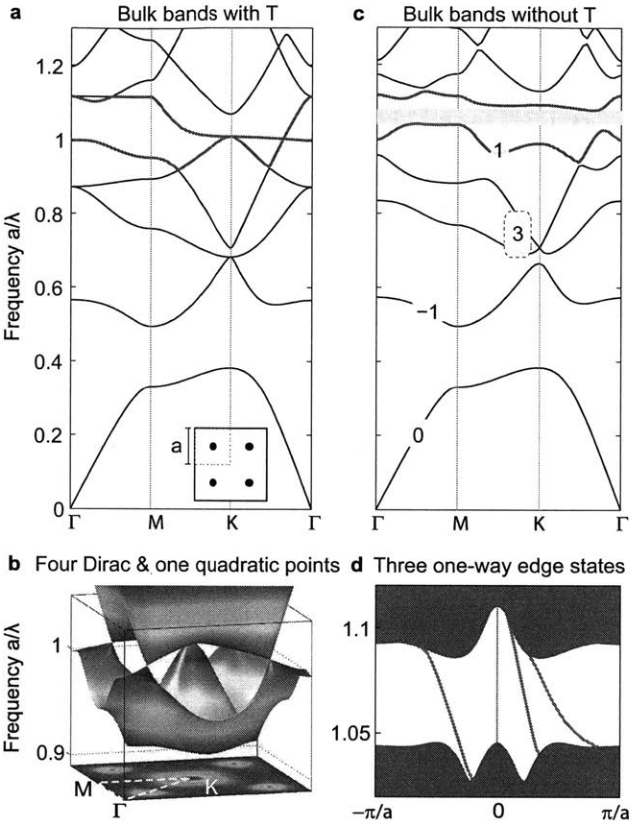


Figure 2-5: Bulk and edge TM bandstructures showing three one-way edge states ($C_{gap} = 3$) obtained from four Dirac points and one quadratic point. The photonic crystal is a square lattice of rods with a radius of $0.10a$, $\epsilon = 15$ and $\mu = 1$. The T-breaking perturbation corresponds to adding $\mu_{12} = -\mu_{21} = 0.45i$ to the rods. a) Bulk bandstructure showing the Dirac point along $\Gamma - K$ and quadratic point at K. The lower inset illustrates the lattice geometry. b) Four Dirac cones and one quadratic point between the 5th and 6th bands plotted in the whole Brillouin zone. c) Bulk bandstructure under T-breaking perturbation opens a 3.6% complete gap highlighted in yellow. Each band is labeled with its Chern number. The Chern number of degenerate bands is circled by a dotted line. d) Three gapless one-way edge states (red) appear in the projected edge band diagram when the bulk is terminated by a metallic boundary.

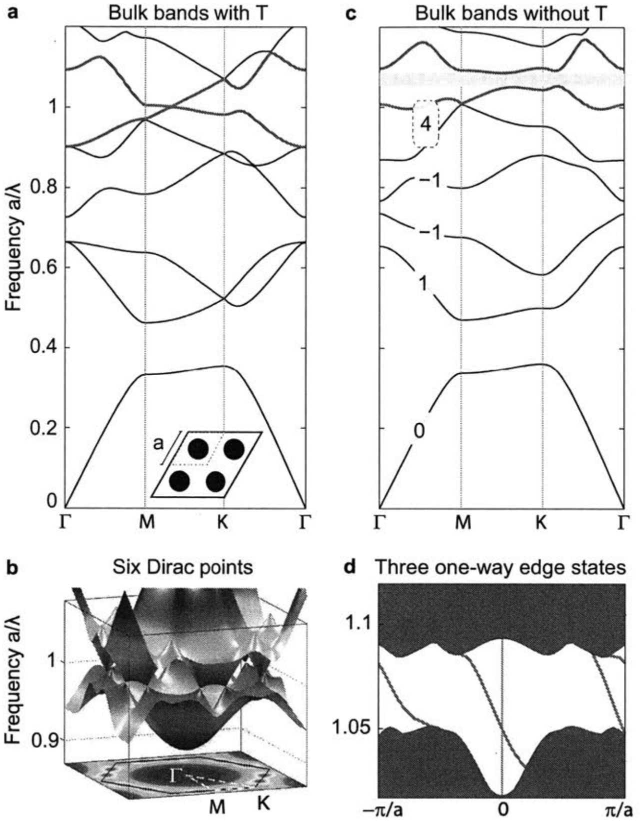


Figure 2-6: Bulk and edge TM bandstructures showing three one-way edge states ($C_{gap} = 3$) obtained from six Dirac points. The photonic crystal is a triangular lattice of rods with a radius of $0.27a$, $\epsilon = 5$ and $\mu = 1$. The T-breaking perturbation corresponds to adding $\mu_{12} = -\mu_{21} = 0.40i$ to the rods. a) Bulk bandstructure showing the Dirac point along $M - K$. The lower inset illustrates the lattice geometry. b) Six Dirac cones between the 6th and 7th bands plotted in the whole Brillouin zone. c) Bulk bandstructure under T-breaking perturbation opens a 3.0% complete gap highlighted in yellow. Each band is labeled with its Chern number. The Chern number of degenerate bands is circled by a dotted line. d) Three gapless one-way edge states (red) appear in the projected edge band diagram when the bulk is terminated by a metallic boundary.

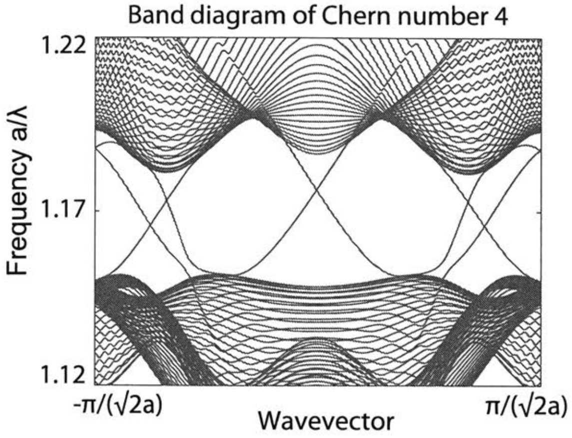


Figure 2-7: The complete bandstructure for a metal-terminated supercell of the $C_{gap} = 4$ photonic crystal from Fig. 2 of the main text. This edge waveguide is oriented along the $\Gamma - K$ direction. Unlike the other bandstructure plots in this letter, here we did not shade the bulk bands and we kept the edge modes from both ends of the supercell. This illustrates the connectivity between the bulk and edge bands.

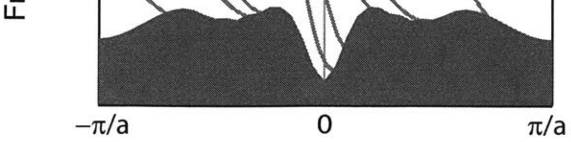


Figure 2-8: An example of eight one-way edge states constructed by photonic crystals of opposite gap Chern numbers ($C_{gap} = \pm 4$). We used the gyromagnetic photonic crystals of $C_{gap} = 4$ in Fig. 2 in the main text of this letter. Across the interface, the static magnetic field is reversed to get opposite Chern numbers.

supported in part by the MRSEC Program of the NSF under Award No. DMR-0819762. M. S. and L.L. were supported in part by the MIT S3TEC EFRC of DOE under Grant No. DE-SC0001299. Also supported in part by ARO through ISN, W911NF-13-D-0001.

Chapter 3

Experimental Observation of Large Chern numbers in Photonic Crystals

3.1 Abstract

Despite great interest in the quantum anomalous Hall phase and its analogs, all experimental studies in electronic and bosonic systems have been limited to a Chern number of one. Here, we perform microwave transmission measurements in the bulk and at the edge of ferrimagnetic photonic crystals. Bandgaps with large Chern numbers of 2, 3, and 4 are present in the experimental results which show excellent agreement with theory. We measure the mode profiles and Fourier transform them to produce dispersion relations of the edge modes, whose number and direction match our Chern number calculations. This work is reported in [37].

3.2 Main text

The Chern number [38] is an integer defining the topological phase in the quantum Hall effect (QHE) [39], which determines the number of topologically-protected chiral edge modes. The quantum anomalous Hall effect (QAHE) possesses these same properties as an intrinsic property of the bandstructure with time reversal symmetry breaking [8, 40]. Recent experiments have discovered the QAHE and its analogues in ferrimagnetic photonic crystals [2], magnetically-doped thin films [10] and in ultracold fermion lattices [41]. However, the Chern numbers observed in all of these systems were limited to ± 1. Finding larger Chern numbers would fundamentally expand the known topological phases [7,11,12,42,43].

Here, we provide the first experimental observation of Chern numbers of magnitude 2, 3 and 4, by measuring bulk transmission, edge transmission, and the edge mode dispersion relations in a ferrimagnetic photonic crystal. The excellent agreement between the experiment and modeling allows us to identify various topological bandgaps and map out the dispersion relations of one-way edge modes for the first time in *any* QHE or QAHE system in nature.

In a 2D system one can realize bands with nonzero Chern numbers, and generate the QAHE, by applying a T-breaking perturbation [1,9,20]. The Chern number is defined as the integral of the Berry flux over the entire Brillioun zone. When connected bands are gapped by a T-breaking perturbation, the bands will exchange equal and opposite Berry flux at each degenerate point, with the total Berry flux exchanged determining the Chern number. For instance, two isolated bands connected by one pair of Dirac points gapped by T-breaking will acquire $\pm 2\pi$ Berry flux (π from each Dirac point) and a Chern number associated with the bandgap ("gap Chern number") of ± 1. A general way to calculate the gap Chern number ($C_{gap} = \Sigma C_i$) is to sum the Chern numbers of all the bands below the bandgap [23]. A bandgap with $C_{gap} = 0$ is trivial, while a bandgap with $C_{gap} \neq 0$ is topologically nontrivial.

In our previous theoretical study we found that the magnitude of the gap Chern number can be increased above one by simultaneously gapping multiple sets of Dirac and quadratic degeneracies. If Berry flux from the gapped degeneracies adds constructively, C_{gap} can be large. In Fig. 1a we present a theoretical topological gap map for a 2D ferrimagnetic photonic crystal as a function of the externally applied magnetic field and the frequency, showing nontrivial bandgaps with C_{gap} from -4 to 3. We studied this same square lattice in an experiment, to verify these predictions.

The experimental configuration resembles a prior work which demonstrated $|C_{gap}| = 1$ [2]. A square lattice of ferrimagnetic garnet rods is placed between two conductive copper plates. This configuration forms a parallel-plate waveguide, with the electric field perpendicular to the plate. Since the electric field for the fundamental mode is constant as a function of height, this is equivalent to a 2D system. The modes in the photonic crystal are excited by

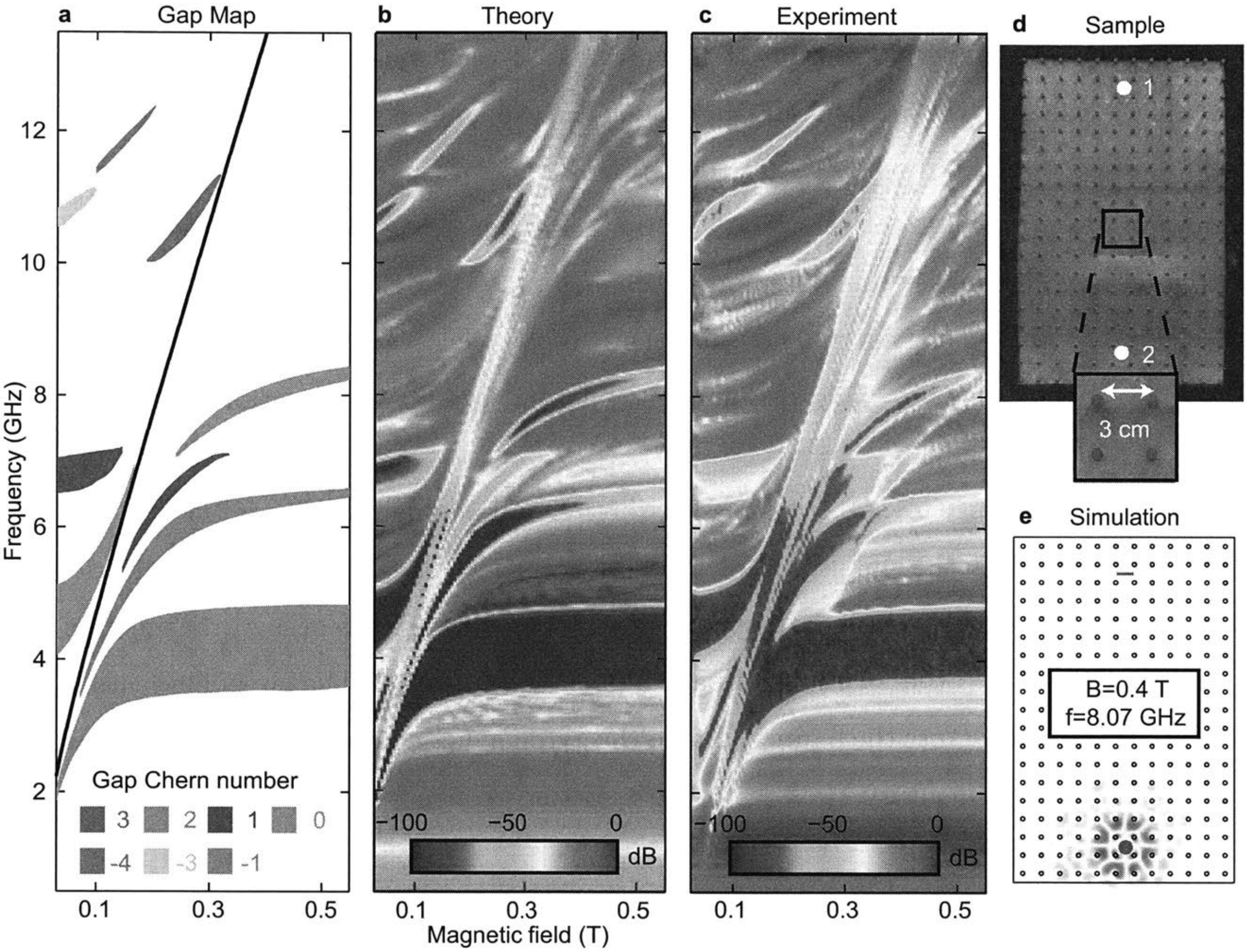


Figure 3-1: Comparison of theoretical gap map and bulk transmission to experimental transmission measurement in a 2D ferrimagnetic photonic crystal. a) Theoretical topological gap map as a function of the magnetic field and the frequency with each bandgap labeled by its gap Chern number. The diagonal black line indicates the resonance in the effective permeability (Supplementary Material) b) Theoretical bulk transmission c) Experimental bulk transmission d) Experimental configuration with the lattice geometry (top metal plate removed). The antenna locations are marked with "1" and "2". e) Simulation geometry with the green line representing the receiving antenna, and the green circle representing the transmitting antenna.

antennas attached to the top plate and fed to a network analyzer. Around the boundary of the system we placed an absorber to minimize reflections and outside interference. We include an overhead image of one of the crystals we constructed in Fig. 1d.

To observe the QAHE analog in the experiment, we break T-symmetry by applying a spatially uniform magnetic field to the ferrimagnetic garnet rods, which acquire off-diagonal imaginary parts in the permeability tensor [24]. Unlike electrons, the external magnetic field does not interact directly with photons. However, for this system, Maxwell's equations can be written in the exact same form as the Schrodinger equation with a periodic vector potential [1]. This makes the system an analogue of the QAHE. Our photonic crystals were placed in the MIT cyclotron magnet, and the magnetic field was swept between 0.03 T and 0.55 T to characterize the transmission of the photonic crystal as a function of the magnetic field and the frequency.

We show the experimental transmission through a bulk photonic crystal in Fig. 1c. Here the color illustrates the amplitude of the transmission between the antennas in decibels ($S12 = 20 \log \frac{E_{in}}{E_{out}}$). In the plots there are several deep blue regions of low transmission that clearly correspond to the locations of bandgaps in the gap map. The sweeping feature that extends diagonally across the figure is due to the gyromagnetic resonance of the ferrimagnetic garnet rods (Supplementary Material). The resonant frequency of the effective permeability is plotted with a black line in Fig. 1a.

In Fig. 1b we present the corresponding theoretical data for transmission through a lattice of the same size and dimension calculated with COMSOL. One of the transmission simulations is shown in Fig. 1e. For direct comparison, the transmission data in Fig. 1b is plotted with the same colorbar scale as the experiment in Fig. 1c. The slight offset of about 0.04 T in the magnetic field between the theoretical and experimental plots is caused by demagnetization (Supplementary Material). Clearly the theoretical and experimental transmission bear strong resemblance to each other and the topological gap map, showing that a square lattice of ferrimagnetic rods can contain a wide variety of different C_{gap} numbers.

Several nontrivial bandgaps in Fig. 1 and in the supplementary material occur even at low magnetic fields. This indicates that topological effects can be achieved at low applied magnetic fields(~ 0.03 T) enabling various studies and applications. Furthermore, these same

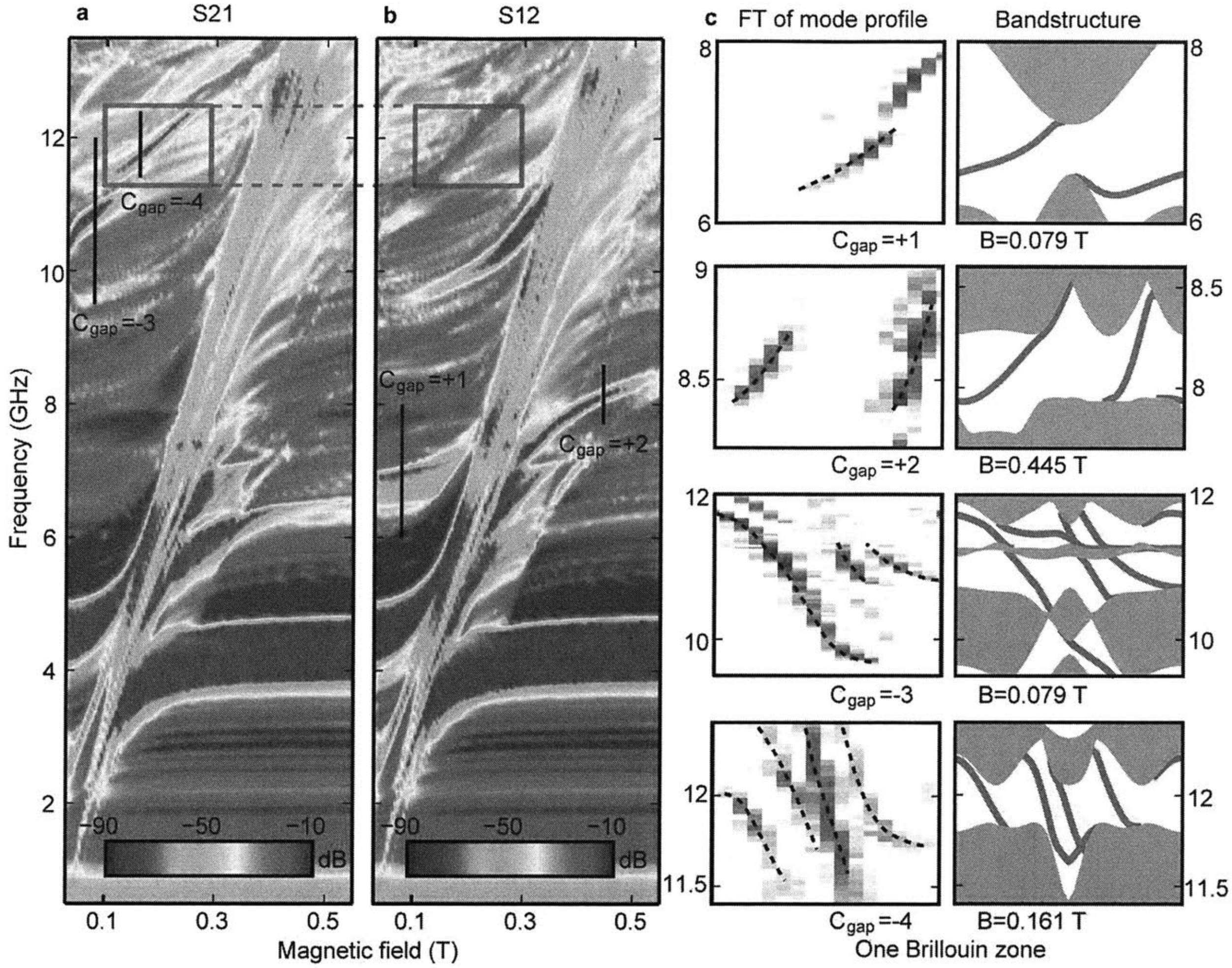


Figure 3-2: Experimental edge transmission measurement and Fourier transform (FT) of mode profiles along the copper boundary. a) S21. b) S12. The bandgaps that are nontrivial have direction-dependent edge transmission, because the interface of a nontrivial bandgap with a trivial bandgap (copper boundary) supports one-way modes. In a) and b) this causes the nontrivial bulk bandgaps from Fig. 1c to be present in one direction (e.g. S12) and be absent in the other (e.g. S21), which we highlight for the $C_{gap} = -4$ bandgap with black boxes. The trivial bandgaps around 4 GHz do not support one-way modes, and so do not exhibit direction-dependent transmission. c) Experimental FT of edge mode profiles and the theoretical edge bandstructures with the edge modes in red and the bulk bands in gray. The range of wavevectors included in both plots is the same and includes only one Brillouin zone. The number of one-way edge modes in both sets of plots agrees with $|C_{gap}|$ from Fig. 1a, while the sign of C_{gap} is consistent with the theoretical group velocity (from the edge mode dispersion) and the directional transmission in a) and b).

bandgaps would remain open at zero external magnetic field by using ferrimagnetic materials with remanent magnetization [44]; this way, a future experiment could be performed even without external magnetic fields.

One-way edge modes are present at the boundary between two crystals with a nontrivial and a trivial bandgap respectively, or at the boundary between crystals with nontrivial bandgaps with different C_{gap} [21, 40]. If the bandgaps of two neighboring crystals overlap in frequency, the number of edge states in the shared frequency gap is determined by the difference between the gap Chern numbers of each crystal. The sign of this difference determines the directions of the edge states. This means that with the nontrivial bandgaps we found, constructing one-way waveguides with up to seven modes is possible. If one of the materials is trivial ($C_{gap} = 0$), like metal or air, the number of edge states equals the gap Chern number of the crystal, with the sign of this number determining their directions.

To provide more evidence of the topological state of these bandgaps and the one-way modes we modified the setup to include a highly conductive copper boundary at the edge of the crystal. This boundary acts as a mirror with a trivial bandgap. We place two antennas near this edge on each side of the sample and measure the transmission between them. In Fig. 2a and Fig. 2b we present both the S12 and S21 parameters to describe the direction-dependent transmission of the edge modes along the metal boundary. S12 refers to exciting the second antenna and measuring with the first antenna, while S21 is the opposite.

The bandgaps that are nontrivial ($C_{gap} \neq 0$) can be identified in Fig. 2 because they will have direction-dependent edge transmission. Specifically the nontrivial bandgaps measured in Fig. 1c will appear in either Fig. 2a or b, but not both. We show this explicitly for the $C_{gap} = -4$ bandgap by highlighting the direction-dependent transmission with gray boxes. This arises from the directional edge states as follows. In one direction, the group velocity of the edge modes is opposite that required to travel to the receiving antenna, so the transmission measurement will record the bulk bandgap. However, in the other direction, the group velocity of the edge modes is in the same direction as is required to get to the receiving antenna, so the bandgap will appear to be nonexistant. Trivial bandgaps ($C_{gap} = 0$) around 4 GHz do not support one-way edge modes, and so do not exhibit direction-dependent transmission at the edge.

To further study the topological nature of these bandgaps we measured the mode profile at the edge of the photonic crystal. We accomplished this by mounting one antenna for excitation to the lower plate, and another small dipole antenna for measurement to the upper plate [45]. During the measurement, the upper plate was translated a total of 47 cm in 2.5mm steps. At each step both the phase and amplitude of the electric field was recorded (Supplementary Material). From this spatial data the mode profile in the waveguide can be reconstructed. The Fourier transform of the mode profile produces the dispersion relation of the waveguide which we present on the left-hand side of Fig. 2c.

In Fig. 2c on the right-hand side we include a comparison with the edge band calculations with the bulk bands in gray, and the edge modes in red. It is clear that the calculated edge-mode dispersion shows an excellent agreement with the dispersion relations extracted from experiments. The number of edge modes is equal to the gap Chern number for each inset. The sign of C_{gap} is consistent with the group velocity of the edge modes and agrees with the directional edge transmission data in Fig. 2a and 2b. In the supplementary material we present additional simulations validating these results for $C_{gap} = -3$ and -4. These results consistute the first direct measurement of one-way edge mode dispersion in any QHE system.

To further study the gap Chern numbers of the observed topological bandgaps, we construct a topological one-way circuit [7]. As illustrated in Fig. 3d, this consists of a $C_{gap} = 2$ (a=3.0 cm) crystal and $C_{gap} = 1$ (a=2.4 cm) crystal, with a copper boundary on the edge. We present the design and calculations characterizing the $C_{gap} = 1$ crystal in the Supplementary material, while the results from the $C_{gap} = 2$ crystal are shown in Fig. 1 and Fig. 2. From the rules described earlier, there will be two edge states flowing downwards between the metal boundary and the $C_{gap} = 2$ crystal as indicated with arrows in Fig. 3d. These edge states will "split" at the junction with one edge state flowing away along the boundary between the $C_{gap} = 1$ and the $C_{gap} = 2$ crystal, and the other continuing along the metal and $C_{gap} = 1$ crystal interface.

In Fig. 3a-c we present the transmission between ports 1-3 and a 4th port located at the junction as labeled in Fig. 3d. The highlighted yellow region indicates the shared bandgap between the $C_{gap} = 1$ crystal and the $C_{gap} = 2$ crystal. For each of the measurements, it is clear that in one direction we have a strong bandgap, with a signal level at the noise floor of

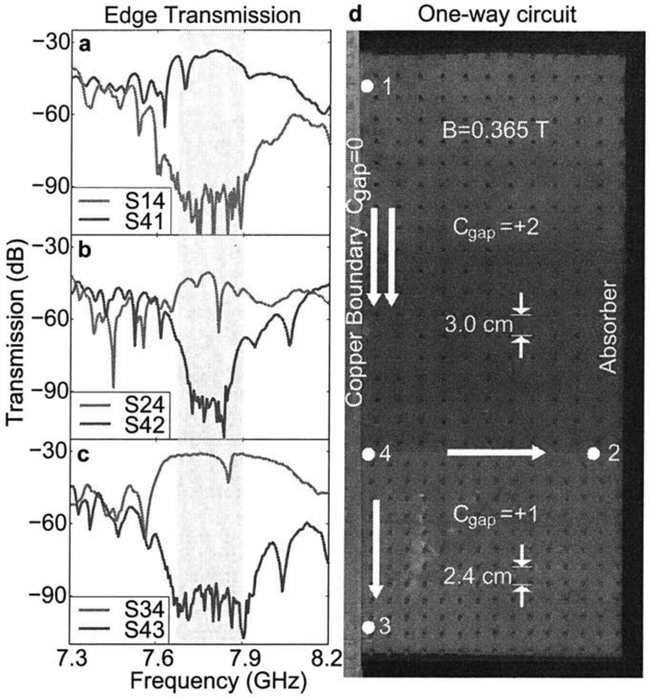


Figure 3-3: Topological one-way circuit implemented using $C_{gap} = 1$ (a=2.4 cm) and $C_{gap} = 2$ (a=3.0 cm) photonic crystal. a)-c) Transmission plots showing edge transmission between antennas at 1,2, and 3 and antenna 4 located at the center. Shared bulk bandgap for $C_{gap} = 1$ and $C_{gap} = 2$ crystals is highlighted in yellow. d) Experimental configuration illustrating crystals with copper boundary ($C_{gap} = 0$) on left and antenna locations 1-4. Arrows indicate the theoretical direction and the number of the edge states at each interface. The transmission data is consistent with predicted edge state directions, which confirms that the upper crystal has $C_{gap} > 1$.

about -100 dB, while in the opposite direction there is 50 to 60 dB more transmission. These edge state directions are consistent with the theoretical predictions and prove the existence of C_{gap}¿1 for the upper crystal. The results from Fig. 3 were obtained under an applied magnetic field of 0.365 T, although there was a window extending approximately from 0.32 T to 0.4 T where the $C_{gap} = 2$ and $C_{gap} = 1$ bandgaps from each crystal were well aligned (Supplementary Material).

In conclusion, we experimentally constructed a square lattice ferrimagnetic photonic crystal with a bandstructure comprised of high C_{gap} (-4 to 3) bandgaps and measured the dispersion relations of the multimode one-way edge waveguides. Fundamentally, having bandgaps with higher gap Chern numbers greatly expands the phases available for topological photonics. These results can potentially enable multi-mode one-way waveguides with high capacity and coupling efficiencies, as well as many other devices [26–28, 30]. A topological photonic circuit can also be made by interfacing photonic crystals of various C_{gap}, with one-way edge states combining together or splitting off at the junctions. Given the rapidly expanding literature on the QAHE and its analogs for $|C_{gap}| = 1$ [13–17, 19, 40], many more avenues of research are now possible because of the greater range of topological phases that can be investigated. Our approach can be readily extended to other systems of Bosonic particles such as magnons [46], excitons [47], and phonons [48, 49].

3.3 Appendix

3.3.1 Experimental Methods

The experimental setup, for the bulk and edge transmission measurements, consisted of two parallel copper sheets 1 m by 0.5 m attached to aluminum plates for mechanical stability and were separated by 0.635 cm (0.25 in). In the range of frequencies studied in this work, the sheets only supported modes which had a uniform electric field perpendicular to the plane and the magnetic field in-plane. Even though the experiment was in a 3D system, the modes are effectively uniform in the perpindicular direction and so can be represented by 2D TM simulations. Microwave absorbers placed around the boundary absorbed stray radiation and helped reduce the noise floor of the measurements.

The ferrimagnetic material used to make the crystals was obtained from TCI cermaics (NG-1850) and consisted of rods 0.795 and 0.823 cm in diameter and 0.635 cm tall. The smaller rods were used to construct the 2.4 cm square lattice used for the $C_{gap} = 1$ photonic crystal in Fig. 3, and in the supplementary material. The larger rods were used to construct the 3 cm lattice in Fig. 1-3. The rods had a saturation magnetization ($4\pi M_s$) of 1817 Gauss and a permittivity of 14.23. The loss tangent of the material was $2 \cdot 10^{-4}$ and the gyromagnetic resonance loss width (ΔH) was $17Oe$. The rods were produced from the same production batch so that the material parameters were uniform. They were machined down to different radii to optimize the overlap and size of the $C_{gap} = 1$ and $C_{gap} = 2$ bandgaps for the splitter experiment presented in Fig. 3. The material had very little hysteresis (and residual magnetization) overall since measurements conducted with increasing, versus decreasing, magnetic field were indistinguishable.

For the mode profile measurements we used a set of lighter and smaller aluminum plates to facilitate easier translation of the upper plate. The aluminum itself is sufficiently conductive for most microwave experiments. The transmitting antenna was mounted to the lower plate, while the receiving antenna was mounted to the upper plate. The upper plate slid along on top of two aluminum bars which were approximately 1 mm higher than the ferrimagnetic rods in the experiment. This gap between the upper plate and the rods was necessary for maintaining smooth translation, and does not invalidate the approximate 2D simulations we used for modeling [45]. The key for a successful mode profile measurement was smooth translation of the upper plate and a consistent metal to metal contact of the supporting bars and the upper plate.

We used a 8719C HP network analyzer for the transmission and mode profile measurements, and a LakeShore Model 410 Gaussmeter for measuring the magnetic field inside the MIT cyclotron magnet. Approximately one hundred transmission measurements were made while the magnetic field was swept between 0.03 T and 0.55 T. The frequency data was smoothed over the neighboring 4 points in post-processing. The experimental results presented in Fig. 1 and in the corresponding plot in the supplementary material were calibrated using the S12 parameters calculated for the antennas inside the parallel plate waveguide without the lattice of ferrimagnetic rods. This calibration helped remove the variation in transmission with frequency that results from the intrinsic impedance mismatch of the antenna. All transmission results were normalized by the maximum transmission. No such normalizations and smoothings were carried out for the mode profile results.

3.3.2 Material Model

Under an applied magnetic field, the permeability tensor of the ferrimagnetic garnet takes the following form [24]:

$$[\mu] = \begin{bmatrix} \mu & ik & 0 \\ -ik & \mu & 0 \\ 0 & 0 & \mu_{zz} \end{bmatrix}$$

Where $\mu = \mu_0(1 + \frac{\omega_0\omega_m}{\omega_0^2-\omega^2})$ and $k = \mu_0\frac{\omega\omega_m}{\omega_0^2-\omega^2}$. The gyromagnetic resonance frequency, $\omega_0(=\mu_0\gamma H_{int})$, and $\omega_m(=\mu_0\gamma M)$ are determined by the internal magnetic field H_{int} and the magnetization M where γ is the gyromagnetic ratio. The internal magnetic field, H_{int} $(=H-XM(H))$, is the difference between the external magnetic field H and the demagnetization field $XM(H)$, where X is the geometry-dependent demagnetization factor and $M(H)$ is the magnetization function. For our theoretical calculations we took $M(H) = M_{sat}$, and $H_{int} = H$, since we did not know the low field behavior of $M(H)$. This approximation causes the small offset of about 0.04 T visible between our transmission experiments and calculations in Fig. 1 and Fig. S3.

The diagonal sweeping features in Fig. 1,2, S3-4, and S7-8 that go from low frequency and low magnetic field, to high frequency and high magnetic field are caused by the resonant peak of the effective permeability ($\mu_{eff} = \frac{\mu^2-k^2}{\mu}$ [24]). The resonance frequency of the effective permeability is $\frac{\sqrt{\omega_0(\omega_0+\omega_m)}}{2\pi}$. It is important to note that this is different than the gyromagnetic resonance frequency$(=\frac{\omega_0}{2\pi})$, which is lower. Bandgaps to the left of this resonance, tend to move upwards in frequency as the magnetic field is increased because the index of the rods$(=\sqrt{\epsilon\mu_{eff}}\,)$ is becoming smaller as μ_{eff} decreases. On the other side of the resonance, the bandgap frequencies tend to shift downward as the magnetic field is decreased because the index is increasing as μ_{eff} increases.

Loss is included in all of the previous calculations by taking $\omega_0 \rightarrow \omega_0 + i\frac{\mu_0\gamma\Delta H}{2}$, where ΔH is the gyromagnetic resonance width.

3.3.3 Mode profile measurement and $C_{gap} = -3$ and -4 comparison to simulation

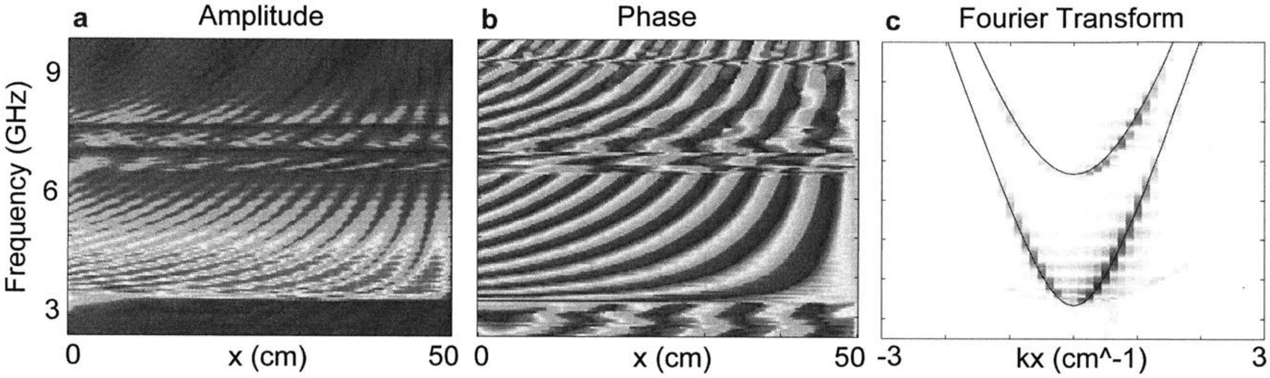


Figure 3-4: Phase and amplitude measurements for 4.5 cm by 0.735 cm rectangular waveguide with Fourier transform of calculated mode profile. a) Amplitude measured in waveguide from 0 to 50 cm b) Phase c) Fourier transform of electric field computed from the measured amplitude and phase data. The peaks of the Fourier transform appear in red, and their dispersion matches the theoretical predictions in black. Note that the peaks on the right side (the forward propagating waves) are higher in amplitude. The amplitude of the forward propagating wave is higher than the backward propagating one because of absorption at the end of the waveguide by the microwave absorber.

In Fig. S1 we present the procedure for extracting the dispersion relation of a rectangular waveguide from the mode profile. For this measurement we constructed a 50 cm long rectangular waveguide 0.735 cm tall and 4.5 cm wide. We placed microwave absorbers at the ends of the waveguide which minimized reflections.We first present the raw amplitude and phase data in Fig. S1a and b. The phase and amplitude for the entire range of frequencies was acquired during a single measurement at a specific position in the waveguide. We combined the phase and amplitude data to calculate the electric field along the edge. After we have the total mode profile, we take the Fourier transform to find the peaks in k-space corresponding to the propagating modes. In Fig. S1c, we plot the FT as a function of frequency and wavevector, allowing visualization of the mode dispersion in the waveguide.

The location of the FT peaks matches the theoretically-predicted dispersion relations which are in black. Note there is more energy on the forward propagating branch of the dispersion relation, than the backward propagating one. The phase and amplitude information from the measurement allows us to completely reconstruct the backward and forward propagating modes independently.

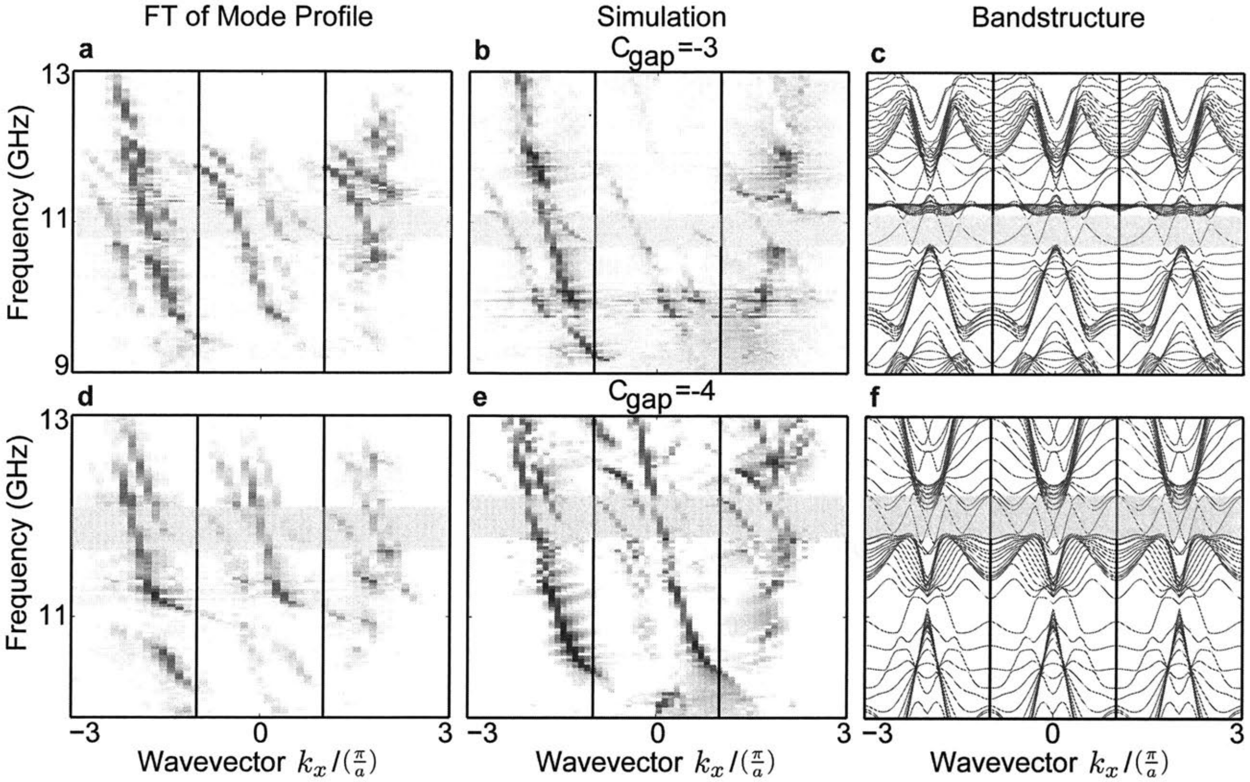


Figure 3-5: Fourier Transform of experimental and simulation data compared to computed bandstructure for $C_{gap} = -3$ and -4 from Fig. 2 in main text. Regions of interest are highlighted in yellow. a) FT of experimentally measured mode profile. b) FT of simulated mode profile in geometry closely matching experiment. c) Supercell calculation for bulk and edge bands. Note that in the bandgap we have forward and backward propagating modes, but that each set of these modes is localized to opposite edges. d)-f) The same data sets for $C_{gap} = -4$. The excellent agreement between these figures provides further evidence for the experimental observation of multiple one-way edge modes.

In Fig. S2 we present a comparison between Fourier transforms computed from simulation and experiment, and calculated supercell bandstructures for $C_{gap} = -3$ and -4 from Fig. 2 in the paper. We have highlighted the regions of interest in yellow. In interpreting the results it is important to understand some fundamental features of the Bloch modes propagating in the lattice. In general the electric field E_z in periodic media can be written as $u(x, y)e^{ikx}$

where $u(x,y) = u(x+a,y)$ is the envelope function and k is the wavevector, we can rewrite this by expanding $u(x,y)$ in its spatial Fourier components:

$$E_z(x,y) = u(x,y)e^{ikx} = \sum_G a_G(y)e^{i(k+G)x} \tag{3.1}$$

We can make several important observations about this result. The most important is that we will potentially observe multiple 'copies' of a single dispersion relation, shifted by reciprocal lattice vectors G. These copies are clearly visible in the experimental and theoretical results in Fig. S2a and S2b, where we have seperated each BZ by vertical black lines.

Another important fact is that the Fourier coeffecients a_G are functions of y. This means to reproduce the experimental results precisely that the exact positions of the excitation and measurement antennas have to be known, in addition to the distance between the metal boundary and the photonic crystal. In Fig. S2b we attempted to reproduce the experimental geometry as closely as possible. Despite many variables and nonidealities, the FT from the simulated mode profile bears close resemblance to that measured in experiment, providing further evidence for the observation of multiple one-way modes.

Finally in Fig. S2c we present supercell calculations for the bulk and edge bands. Careful study reveals that many bright-spots in the Fig. S2a and b can be traced back to certain bands which are being excited. Note that the supercell calculations in general include forward and backward propagating modes, but that inside the bandgap, one edge mode dispersion is localized to one end of the computational cell, while the other, oppositely traversing mode is on the opposite side, so in experiment and simulation only one set will be observed.

3.3.4 $C_{gap} = 1$ (a=2.4 cm) photonic crystal

In Fig. S3 and Fig. S4 we present the experimental results for the bottom photonic crystal used for the splitter in Fig. 3, which had a lattice period of 2.4 cm. Fig. S3 shows the agreement between the theoretical gap map, theoretical bulk transmission, and experimental transmission. This crystal was selected because the $C_{gap} = 1$ bandgap (the one to the right of the resonance) occurs in the same frequency and magnetic field ranges as the $C_{gap} = 2$

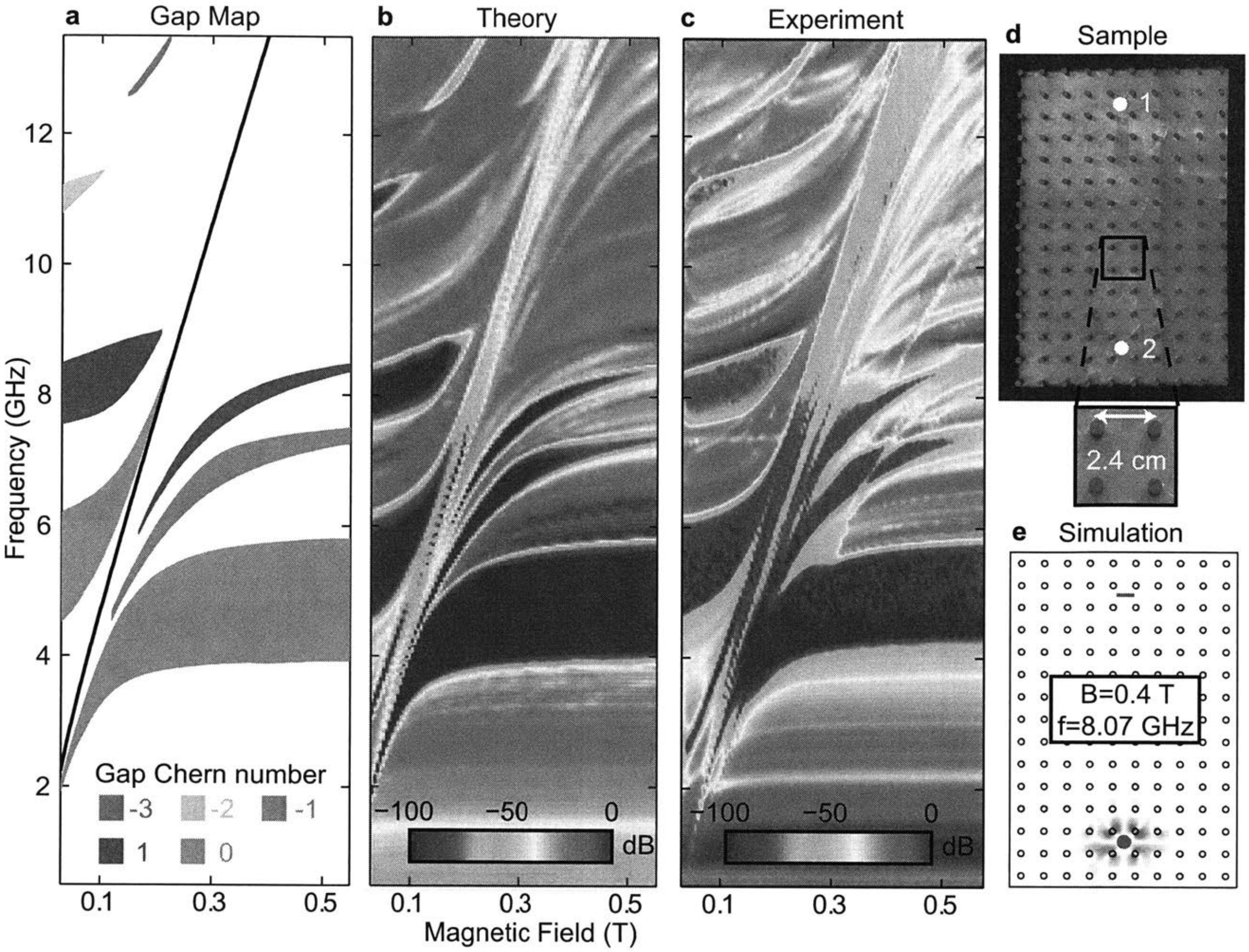


Figure 3-6: Comparison of theoretical gap maps and bulk transmission to experimental transmission measurement in a 2D ferrimagnetic photonic crystal for $C_{gap} = 1$ (a=2.4 cm) crystal from Fig. 3. a) Theoretical topological gap map as a function of the magnetic field and the frequency with each bandgap labeled by its gap Chern number. The black diagonal line indicates the resonance in the effective permeability (Appendix B) b) Theoretical bulk transmission c) Experimental bulk transmission d) Experimental configuration with the lattice geometry (top metal plate removed). The antenna locations are marked with "1" and "2". e) Simulation geometry with the green line representing the receiving antenna, and the green circle representing the transmitting antenna.

bandgap in the photonic crystal from Fig. 1 and 2, which we discuss in Appendix D. Fig. S4 shows the edge transmission measurements, and these are consistent with the Chern number and edge mode dispersion from the edge mode calculations. The large $C_{gap} = 1$ bandgap occurring at low magnetic fields is potentially useful for applications. This bandgap can alternatively potentially be opened without an external magnetic field by using a ferrimagnetic material with remanent magnetization.

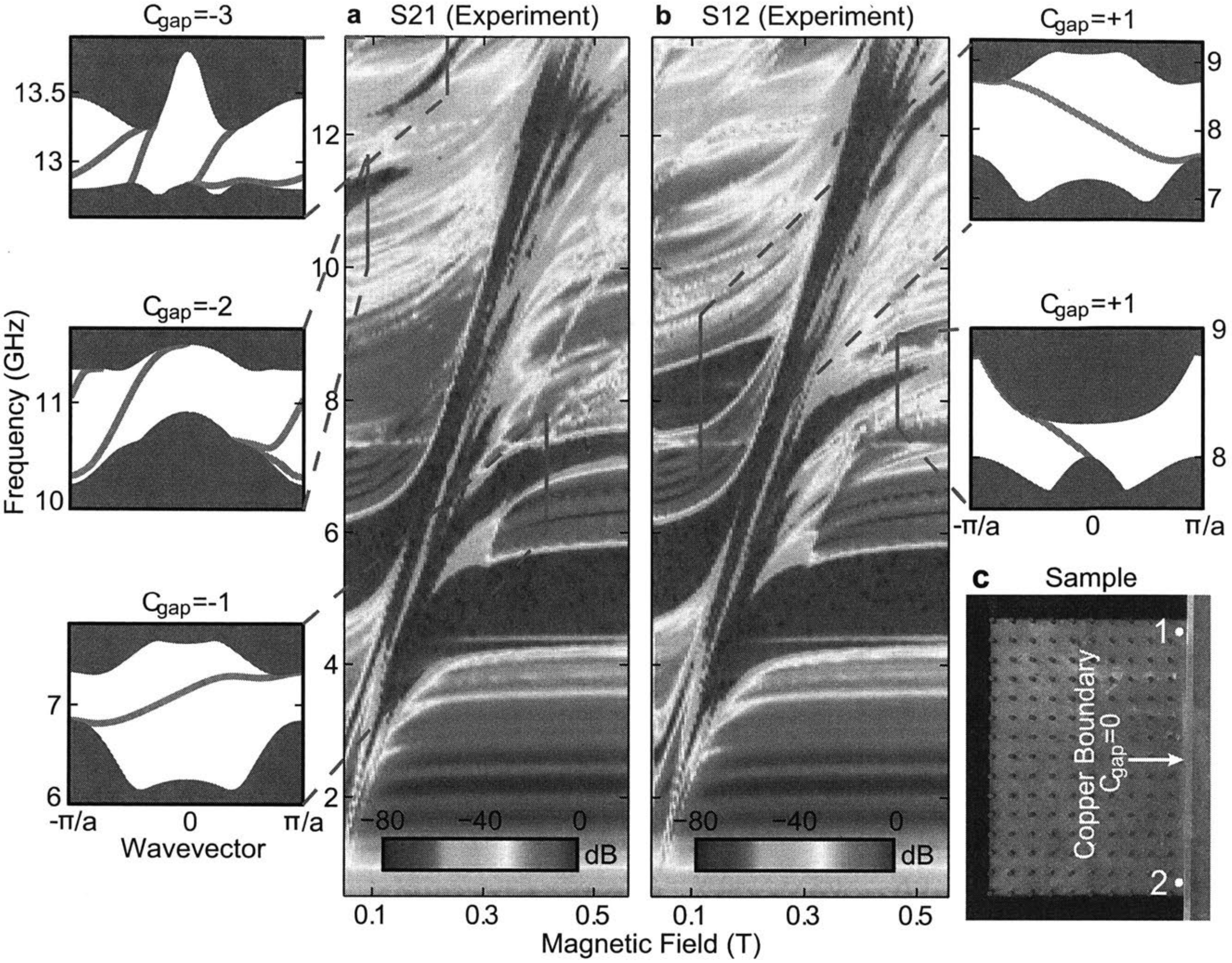


Figure 3-7: Experimental edge transmission measurement along copper boundary for $C_{gap} = 1$ (a=2.4 cm) crystal from Fig. 3. a) S21. b) S12. The bandgaps that are nontrivial have direction-dependent edge transmission, because the interface of a nontrivial bandgap with a trivial bandgap (copper boundary) supports one-way modes. In a) and b) this causes the nontrivial bulk bandgaps from Fig. S3 to be present in one direction (e.g. S12) and absent in the other (e.g. S21). The trivial bandgaps around 5 GHz do not support one-way modes, and so do not exhibit direction-dependent transmission. c) Experimental configuration with the antenna locations and the copper boundary. Insets on the left and right side of a) and b) respectively show edge band calculations with the edge modes in red and the bulk bands in blue. Each supercell calculation applies to a) and b) even though each is shown on only one side for clarity. The number of one-way edge modes corresponds to $|C_{gap}|$ from Fig. S3a, while the sign of C_{gap} is consistent with the theoretical group velocity (from the edge mode dispersion) and the directional transmission in a) and b).

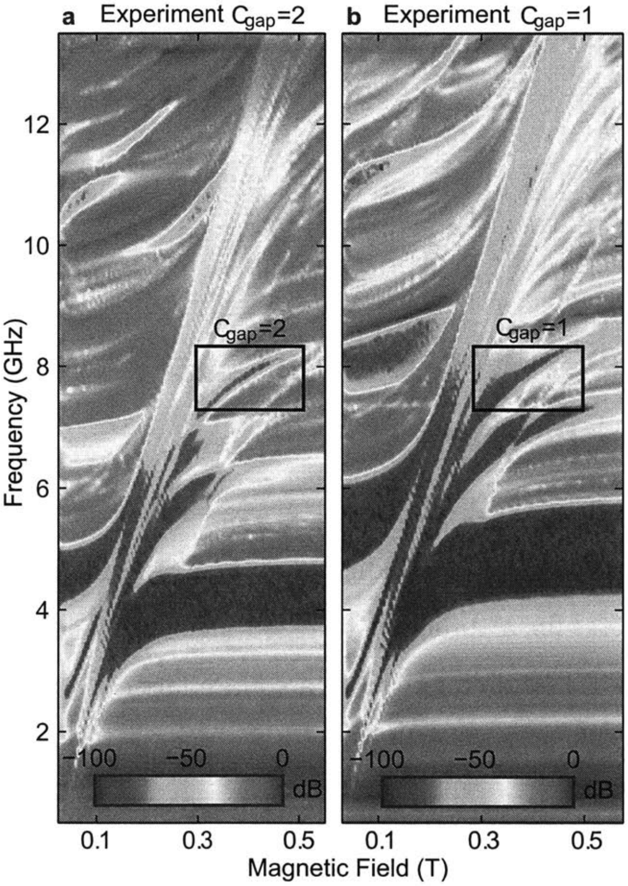


Figure 3-8: Alignment of bandgaps for $C_{gap} = 2$ (a=3.0 cm) crystal and $C_{gap} = 1$ (a=2.4 cm) crystal from Fig. 3 splitter. a) Transmission for a=3 cm crystal. The nontrivial $C_{gap} = 2$ bandgap used in the Fig. 3 splitter experiment is boxed in black. b) Transmission for a=2.4 cm crystal. The nontrivial $C_{gap} = 1$ bandgap used in the Fig. 3 splitter experiment is boxed in black. The $C_{gap} = 2$ bandgap and $C_{gap} = 1$ bandgap from each crystal show good frequency alignment over a range of magnetic fields from 0.32 T to 0.4 T

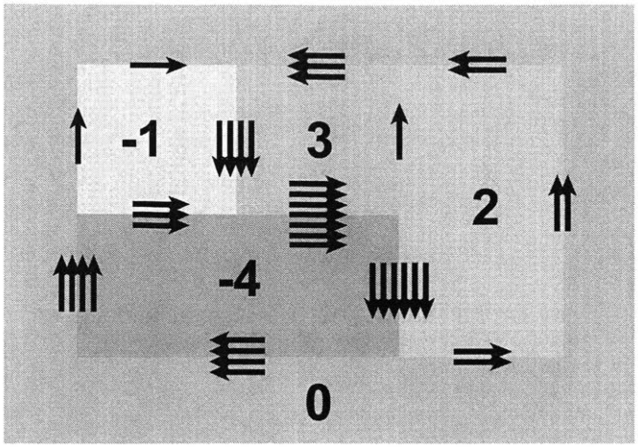


Figure 3-9: One-way Topological Photonic Circuit. Photonic one-way circuit which can be constructed using the topological bandgaps we found. Arrows represent one-way edge modes and each crystal is labeled by its gap Chern number.

3.3.5 Bandgap alignment between $C_{gap} = 2$ (a=3.0 cm) and $C_{gap} = 1$ (a=2.4 cm) photonic crystals

We designed each crystal in the Fig. 3 splitter experiment so that a nontrivial $C_{gap} = 2$ bandgap from one crystal would be well aligned at a certain magnetic field with the $C_{gap} = 1$ bandgap of another crystal for a sufficiently large range of frequencies. The $C_{gap} = 2$ crystal was studied in Fig. 1 and 2, and consisted of a square lattice of rods with lattice constant 3 cm. The $C_{gap} = 1$ crystal was studied in Fig. S3 and S4, and consisted of a square lattice of rods with lattice constant 2.4 cm. The transmission measurements showing the bandgaps of each crystal are presented in Fig. S5 with the location of each bandgap highlighted with a black box. For the $C_{gap} = 1$ bandgap and the $C_{gap} = 2$ bandgap, there is significant frequency overlap between the bandgaps for magnetic fields between approximately 0.32 T and 0.4 T. This overlap was acheived through theoretical calculations and experimental tuning of the lattice constant.

3.3.6 One-way Topological Photonic Circuit

We include an example of a topological photonic circuit using the topological bandgaps we found in Fig. S6. We assume in this figure that each topological bandgap for each topological "block" is perfectly aligned in frequency with all of the others. Each topological

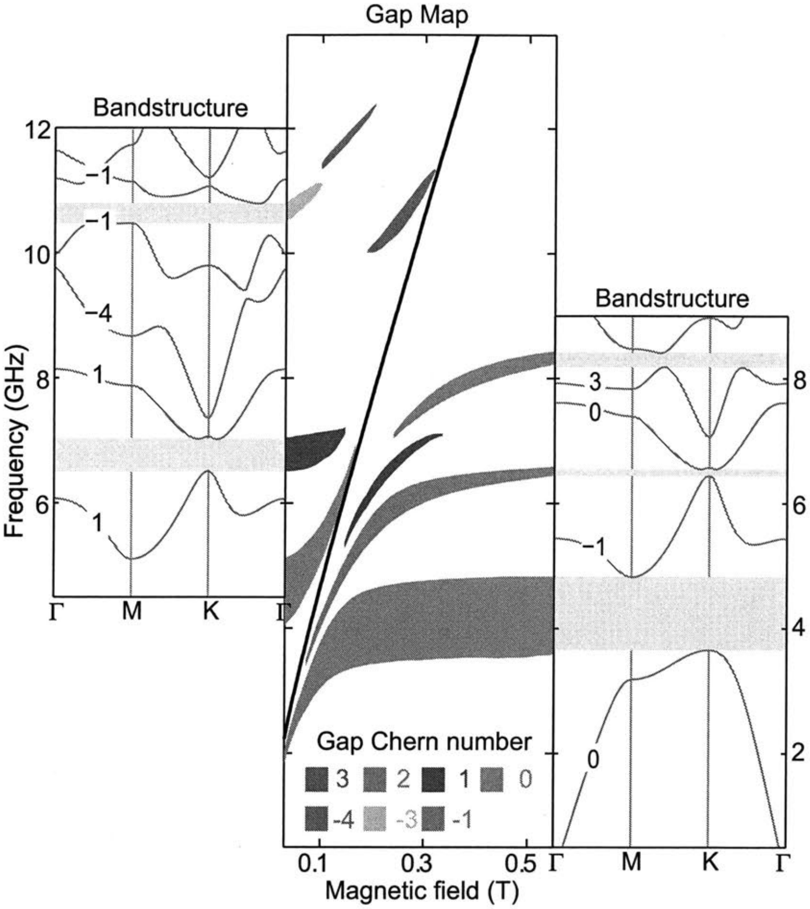


Figure 3-10: Calculation of C_{gap} by summing Chern numbers below bandgap for the $C_{gap} = 2$ (a=3.0 cm) crystal from Fig. 1,2, and 3. On each side of the gap map we have included bulk bands labeled with their Chern numbers. The sum of the Chern numbers of bands below a given complete bandgap is the gap Chern number.

"block" is labeled with its gap Chern number. At the junction between three bulk domains, one-way edge states merge together or branch off, enabling new device functionalities as signal combiners and photonic logic gates immune to backscattering from manufacturing imperfections.

3.3.7 Chern Number Sums

In Fig. S7 and S8 we present the topological gap maps for the photonic crystals we constructed along with insets illustrating the detailed bandstructure at the minimum and maximum magnetic fields. Each band in the insets is labeled with its Chern number. The sum of the Chern numbers of all bands below a given bandgap is the gap Chern number [21]. We

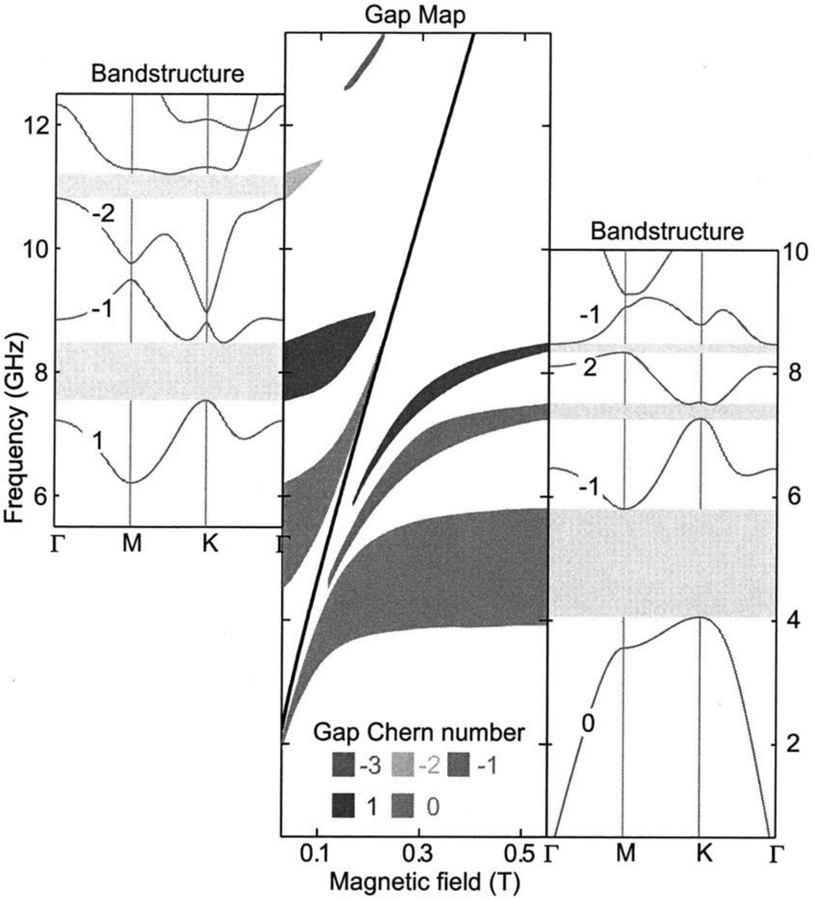


Figure 3-11: Calculation of C_{gap} by summing Chern numbers below bandgap for the $C_{gap} = 1$ (a=2.4 cm) crystal from Fig. 3, S3, and S4. On each side of the gap map we have included bulk bands labeled with their Chern numbers. The sum of the Chern numbers of bands below a given complete bandgap is the gap Chern number.

labeled each band in the gap map with its gap Chern number determined with this rule. The insets were computed using fixed point iteration, while the gap map was computed using transmission calculations as detailed in Appendix G.

3.3.8 Numerical Methods

The bandstructure and transmission plots in this work were calculated using COMSOL which utilizes the finite element method. The Chern numbers of each bulk band were calculated using a custom made finite-difference frequency-domain code.

For the bandstructure calculations, we solved a Hermitian eigenproblem of the following form in COMSOL.

$$\nabla \times (\mu^{-1} \nabla \times \vec{E}) = \epsilon \omega^2 \vec{E}$$

Since ferrimagnetic materials are dispersive (μ depends on ω), this is a nonlinear eigenproblem. Even though COMSOL can only solve linear eigenproblems, we can still find the solution to the nonlinear eigenproblem using fixed-point iteration. In this method the material parameters are evaluated at a trial frequency so that the normal linear eigenvalue problem can be solved. The new solution is used to evaluate the material parameters again, and if the initial guess is close enough, the frequency will converge with enough iterations. We applied fixed-point iteration to calculate the bandstructures in Appendix F.

For the supercell calculations in the paper, the linear eigenvalue problem was solved with material parameters evaluated at the midgap frequency. Material dispersion will cause the bandgaps to become smaller without effecting the topological properties. Consequently we did not include dispersion in the supercell calculations since they contained a large number of bands and would consequently take too many computing hours. The supercell calculations were computed using 15 unit cells of the photonic crystal joined at the ends either with periodic boundary conditions, in the case of the bulk calculations, or metallic boundary conditions, for edge-mode calculations.

The transmission was also calculated with COMSOL by solving for the time-harmonic solution at a given frequency and magnetic field. For these calculations the source "antenna" was represented by three point current sources of equal magnitude in the normal direction within a single unit cell of the crystal. The receiving antenna was represented by a short line on the other side of the crystal over which the power was integrated. The boundaries of the crystal were padded with perfectly matched layers. The effects of material loss and material dispersion were included in the calculation.

Additional transmission simulations were conducted to find the gap map in Fig. 1a and in the supplementary material. In order to sample the full k-space, the receiving boundary was expanded to receive all unique incoming waves from the current sources. The low transmission regions were identified and used to create the gap map. The locations of these gaps are consistent with the bandstructure calculations as shown in Fig. S7 and Fig. S8. Here the fixed-point iteration and transmission methods show good agreement in regimes of low dispersion and loss.

3.4 Acknowledgements

We acknowledge Carl E Patton, Liang Fu, Bo Zhen, Ido Kaminer, Yichen Shen, Zheng Weng, and Hongsheng Chen for discussions, Ulrich Becker and Peter Fisher for assistance with the MIT cyclotron magnet, and Xiangdong Liang for numerical assistance. S. S. was supported by MIT Tom Frank Fellowship and NSF Fellowship. L. L. was supported in part by the MRSEC Program of the NSF under Award No. DMR-1419807. Q. Y. was supported by the National Natural Science Foundation of China under Grants No. 61322501. Fabrication part of the effort was paid by the MIT S3TEC EFRC of DOE under Grant No. DE-SC0001299. Also supported in part by ARO through ISN, W911NF-13-D-0001.

Chapter 4

Lens-enabled Chip-scale LIDAR

4.1 Introduction

The meteoritic rise of autonomous navigation in real-world settings for self-driving cars and drones has propelled rapidly growing academic and commercial interest in LIDAR [50, 51]. One of the key application spaces that has yet to be filled, but is of great interest, is a non-mechanically steered LIDAR sensor which has substantial range (100-300 m), low power (1-10 W), low cost (100s of dollars), high resolution (10^4-10^6 pixels) and small size (10 cm^3). There are several candidate technologies including micro-mechanical mirrors [52, 53], liquid-crystal based devices [54, 55], and integrated photonics that are currently being explored academically and commercially to fill this niche.

Current state-of-the-art chip-scale integrated-photonic LIDARs are based on 1D or 2D phased array antennas [5, 56–63]. In this type of architecture, 1D or 2D arrays of dielectric grating antennas are connected to electrically-controlled thermo-optic (TO) or electro-optic phase shifters. These phase shifters are fed by waveguides splitting off from one main dielectric waveguide which brings power from an off-chip or on-chip source. By applying a gradient to the phases tuning each antenna, in-plane or out-of-plane beam-steering can be enabled. As shown in Fig. 1a and 1b, the direct predecessor of this architecture is RF-phased arrays developed for military and commercial RADARs [3]. Although the detailed implementation is different because RF primarily relies on metallic waveguides and structures whereas integrated photonics uses dielectrics, optical phased arrays are essentially based on directly replacing RF components with their optical equivalents. We note that this direct translation brings a significant disadvantage which is immediately obvious: whereas metallic waveguides can be spaced at sub-wavelength pitches, eliminating parasitic grating lobes, dielectric waveguides have to be separated by several wavelengths to prevent excessive coupling, resulting in significant grating lobes.

Here we study a new type of integrated photonics solution implied by Fig. 1c. Although the optical analogy to RF phased arrays has been well explored, there is an entire class of planar-lens based solutions developed in the RADAR literature that preforms the same function as RF phased arrays, whose integrated-photonics analogs have not yet been investigated [6, 64–68]. Here instead of relying upon many continuously tuned thermal phase shifters to steer beams, we use specially designed lenses excited in their focal plane to generate discrete far-field beams. We explore our new architecture and design in the following sections as well as the main advantages this new type of integrated LIDAR has over existing approaches.

4.2 Architecture and Operation

Inspired by Fig. 1c, we develop a new lens-enabled integrated photonic LIDAR architecture. We have illustrated the light propagation through each component of this architecture in Fig. 2. In the first stage we couple a fiber to an on-chip waveguide. The waveguide is formed by a 200 nm thick and 1 μm wide, Silicon Nitride (SiN) section encapsulated in SiO_2 (the upper SiO_2 layer is omitted from the diagram). This fiber connects to a tunable IR source centered about 1550 nm. The next stage consists of a switch matrix composed of Mach-Zehnder interferometers (MZI) which allow switching one input into 2^D output ports, where D is the depth of the tree [69]. The path length of one arm of a single MZI is controlled by an integrated TO phase shifter (external electronic control lines are omitted) which allows the optical signal to be electronically switched between two output ports. Each output from the switch tree feeds into a slab waveguide, patterned to form the focal surface of the aplanatic lens [70]. The light rapidly diffracts after entering the slab waveguide until being refocused by the aplanatic lens. The lens is formed by a patterned PolySi layer 20 nm or 40 nm thick.

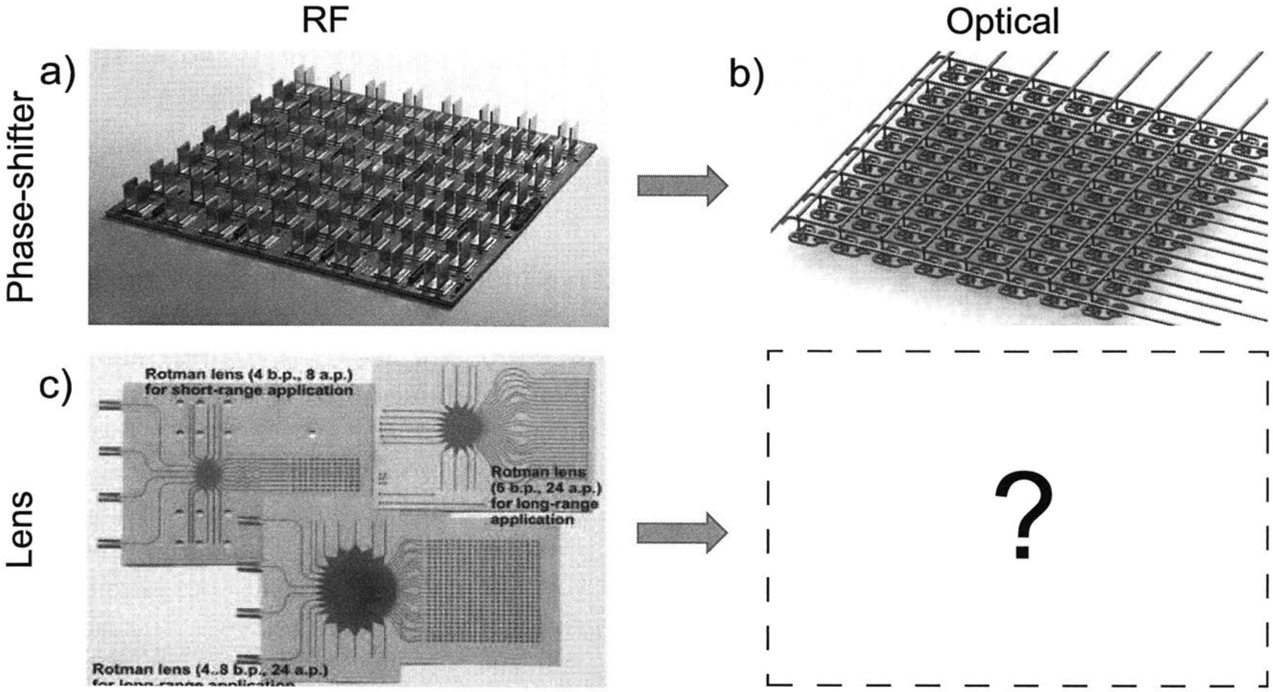


Figure 4-1: Missing analogy between RADAR literature and LIDAR literature. a) Phase-shifter based approaches from RADAR from [3,4]. There is an extensive literature developing phase-shifter based RADARs ranging from only a few active elements to thousands for large arrays. b) This phase-shifter based approach has been extensively adopted by integrated-LIDAR architects, who typically employ 1D or 2D arrays of phase shifters for beam forming. Imaged here is 2D integrated array from [5] with waveguide pitch 6 λ× 6 λ. c) Another class of RADAR solutions based on using lenses. Imaged here is Rotman lens design and patch antenna array from [6]. There is a very rich literature of 1D and 2D dielectric and metallic lens solutions developed in RADAR due to the expense and difficulty of implementing phase shifter solutions. Despite this extensive literature which contains many useful techniques, no analog planar integrated LIDAR solution exists in the literature. Our new solution fills in the missing piece by contributing a planar lens-based approach in direct analogy to these RADAR methods, which will enable new benefits and opportunities.

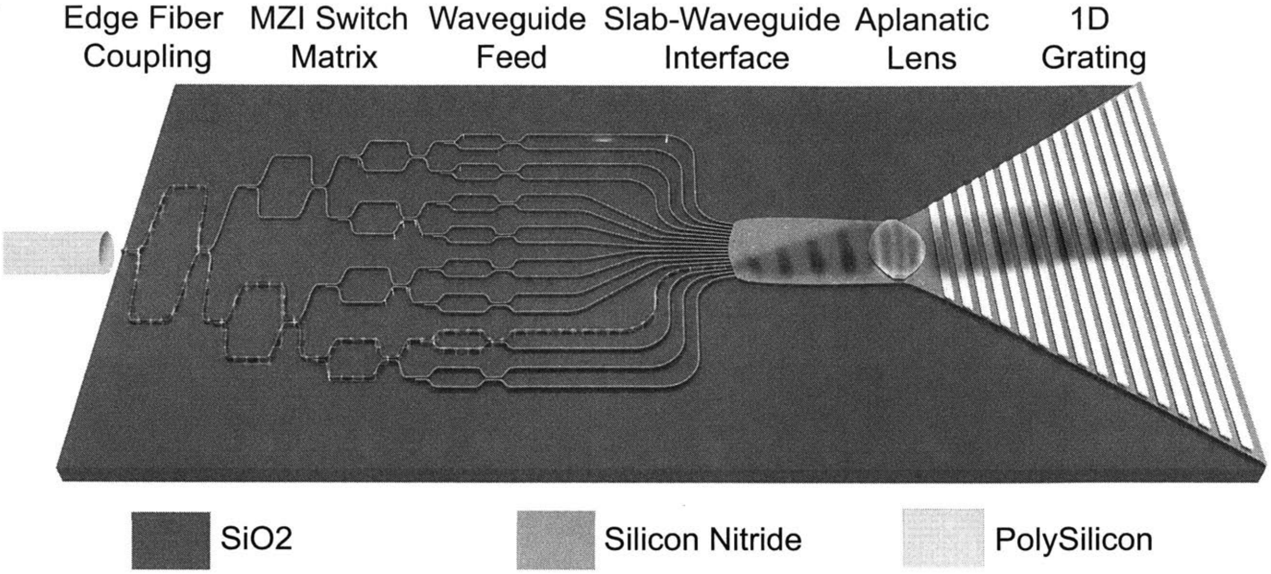


Figure 4-2: Lens-enabled chip scale LIDAR architecture. We illustrate the functionality of our optical chip through multiple stages. First a fiber containing an IR signal is edge coupled into the chip. Subsequently this signal is routed through an optical switch matrix formed by an array of electrically-controlled Mach-Zehnder interferometers (MZI). The electrical control lines required for the TO phase shifters on one arm of these interferometers is omitted. After routing, the beam is emitted at one point in the focal plane of the lens. The signal undergoes diffraction until being focused into a plane-wave by the planar aplanatic lens. Finally the infrared light is scattered out of plane by a 1D grating coupler. The relative dimensions of this figure have been distorted for visualization, but the SiN waveguides are 200 nm thick and 1μm wide, and the PolySi layer forming the aplanatic lens is 20-40 nm thick, with the gratings being the same thickness and material as the lens. An additional upper SiO_2 layer encapsulating all of the components is omitted.

After collimation, the beam propagates into the 1D grating where it is scattered out of the plane and into free-space. The grating is formed out of a PolySi layer the same height as the aplanatic lens. In Fig. 2b we depict the actual realization of the components of this device and test structures in a gdsii file, which we designed and are currently fabricating and testing at MIT Lincoln Labs.

As depicted in Fig. 3, non-mechanical beam steering for this chip is implemented by two mechanisms. The first is port switching, which changes the in-plane propagation direction of the beam as shown in Fig. 3a and 3b. As depicted in Fig. 3c, the wavelength roughly controls the out-of-plane angle, that is the angle between the beam center and the z-axis. These 2D beam steering mechanisms are related in spirit to those of RF Rotman lenses feeding arrays of patch antennas (see Fig. 1c) [71, 72]. The 3D directivity patterns of the generated beams

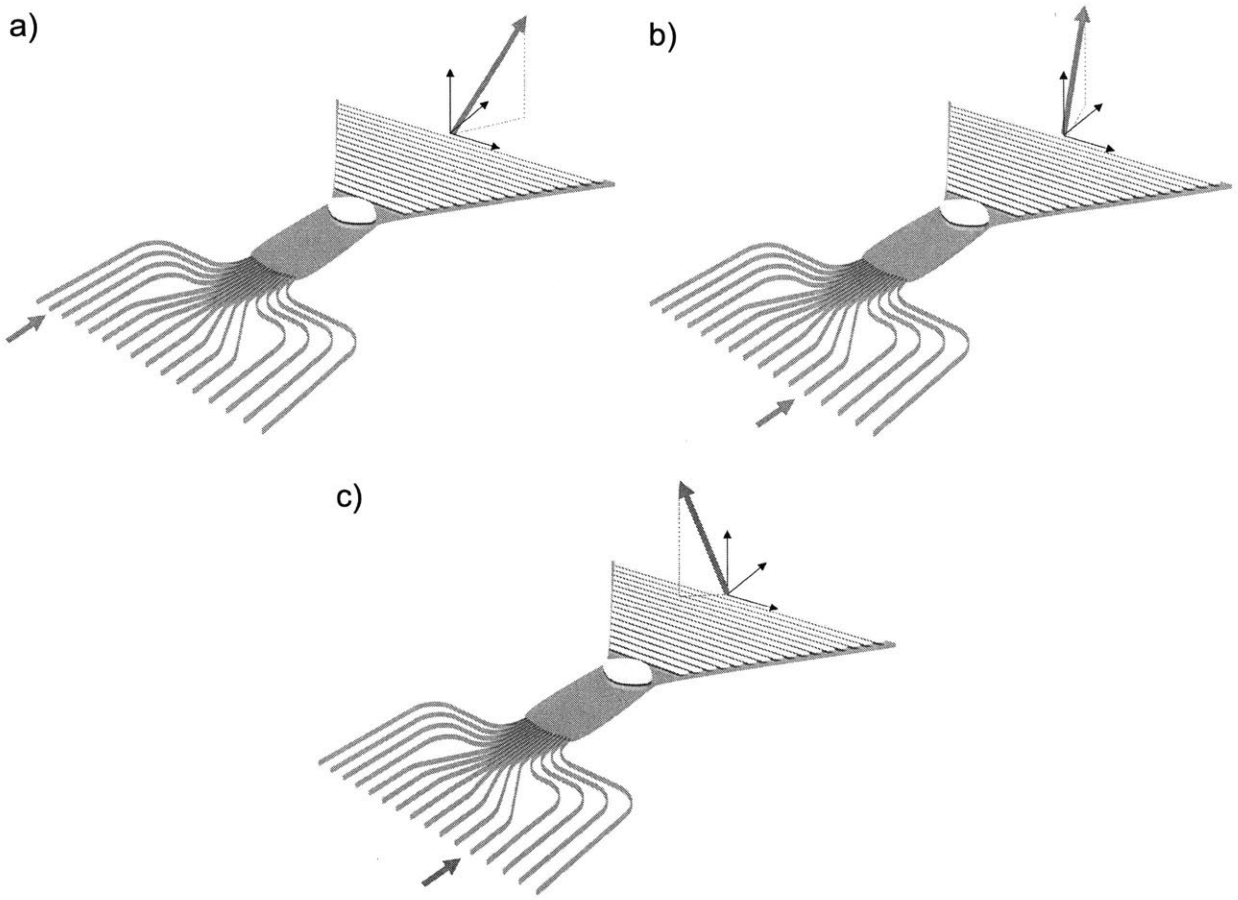


Figure 4-3: Non-mechanical beam steering principle for lens-enabled chip-scale LIDAR. Changing port excitation changes the in-plane beam steering direction, while tuning wavelength (red to blue, these are false colors since the device operates in the IR) roughly changes the out-of-plane angle. The 3D directivity pattern for the system in each excitation scenario is depicted above the output grating.

are depicted in each subfigure. The precise mathematical relationship between the emission angles and the analytic form of the directivity pattern are detailed in the next section.

4.3 Upper Bound on Performance

We begin our discussion of the system design by describing an ideally preforming aperture. The function of the lens is to generate a plane wave propagating at a finite angle. The scattered light from the plane wave propagation through the 1D grating forms the near-field of the radiation pattern. Assuming that the plane-wave emitted from the lens is uniform, that the lens introduces negligible aberrations, and that the lens and grating parameters are wavelength and angle independent, we can write down a simplified aperture pattern of the following form, which is depicted in Fig. 4a:

$$A(x,y) = \begin{cases} \exp(-qx)\exp(ik_0 u_{x0} x)\exp(ik_0 u_{y0} y) & 0 \le x \le L, -\frac{W}{2} + x\tan(\phi') \le y \le \frac{W}{2} + x\tan(\phi') \\ 0 & else \end{cases}$$

This pattern can be thought of as a parallelogram with uniform amplitude in the y-direction, and exponentially decaying amplitude in the x-direction, determined by the grating decay parameter q where L is the length of the grating and W is the width. The inclination of the parallelogram is determined by the grating propagation angle ϕ', which is derived in the Appendix and is close in magnitude to the propagation angle of the beam output from the lens ϕ_{in}.

$u_{x,0} = \sin(\phi_0)\cos(\theta_0)$ and $u_{y,0} = \sin(\phi_0)\sin(\theta_0)$ characterize the direction of the emitted mode, and can be calculated by tracking the phase accumulated by the collimated rays emitted from the lens and discretely sampling them at the grating teeth. As shown in the Appendix we can write these as:

$$\begin{aligned} u_{y,0} &= n_1 \sin(\phi_{in}) \\ u_{x,0} &= \frac{k_{x,avg}(\phi_{in})}{k_0} - \frac{\frac{2\pi m}{\Lambda}}{k_0} \end{aligned} \tag{4.1}$$

Where $k_{x,avg}$ is the average k component in the grating, n_1 is the effective index of the TE slab mode in the lens, m is the grating order, and Λ is the grating period. The function of the grating can be understood from this equation: it allows phase matching to radiating modes through the addition of the crystal momentum $\frac{2\pi m}{\Lambda}$.

Making the approximation $k_{x,avg} \approx n_{eff} k_0 cos(\phi_{in})$, where n_{eff} is the average effective index of the gratings, we can show that u_{x0} and u_{y0} satisfy an elliptical equation:

$$\left[\frac{u_{x,0} + \frac{m\lambda}{\Lambda}}{n_{eff}}\right]^2 + \left[\frac{u_{y,0}}{n_1}\right]^2 = 1 \tag{4.2}$$

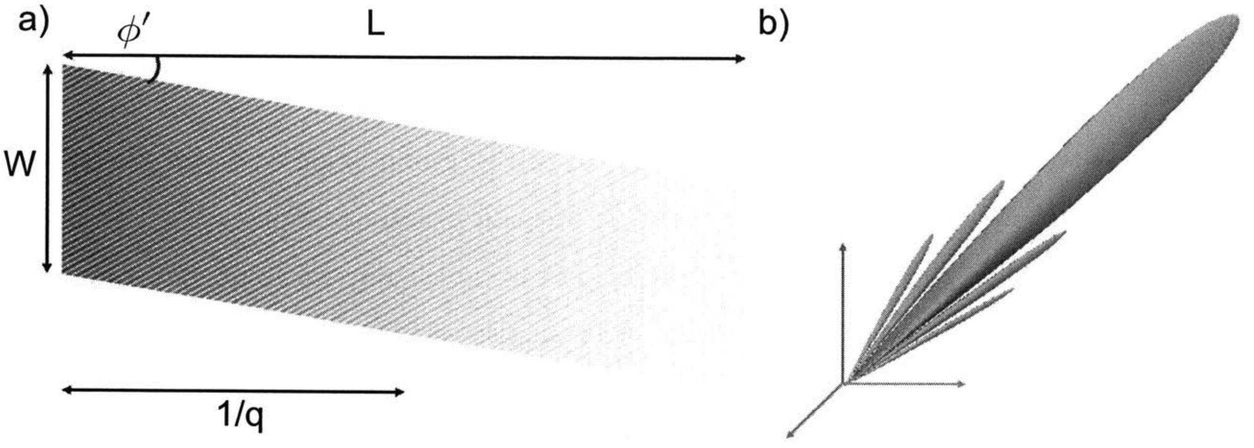


Figure 4-4: Ideal near-field aperture generated from system, the far-field beam angles, and the far-field directivity for this aperture. a) The ideal aperture pattern is a parallelogram with tilt $\phi_{\prime}$, width W, and length L, the exponential decay follows from radiation loss from grating emission in the x direction. The interference pattern results from the addition of a crystal momentum $\frac{2\pi m}{\Lambda}$ to the plane-wave momentum, where Λ is the grating period and m is the grating order. b) The computed far-field directivity for the ideal aperture given in a). The peaks to the left and right of the main beam are known as sidelobes and originate from the sinc factor in the directivity function.

This elliptical equation has a simple physical interpretation. Switching ports in-plane takes us along an elliptical arc in $u_{x,0}$ and $u_{y,0}$ space, while tuning the wavelength λ tunes this arc forward and backwards, which we depict in Fig. 6b.

We can analytically calculate the directivity, which characterizes the far-field distribution of radiation. We find:

$$\mathrm{D}(\Delta u_x, \Delta u_y) = \frac{W k_0^2 \cos(\theta_0)}{\pi q(1 - \exp\{(-2qL)\})} \frac{\mathrm{sinc}^2(\frac{W}{2} k_0 \Delta u_y)}{1 + \frac{k_0^2}{q^2}(\Delta u_x + \tan(\phi')\Delta u_y)^2} \times \\ \left(1 - 2\cos(k_0 L(\Delta u_x + \tan(\phi')\Delta u_y)) \exp\{(-qL)\} + \exp\{(-2qL)\}\right) \quad (4.3)$$

Where $\Delta u_x = u_x - u_{x,0}$ and $\Delta u_y = u_y - u_{y,0}$. We plot this in Fig. 4b. There are several recognizable features to this function such as the sinc from the rectangular aperture, and the $\frac{1}{k_0^2+q^2}$ from the exponential decay [73]. The $\tan\phi'$ components introduce a "shear" into the beam spots and come directly from the tilted aperture pattern.

Additionally we can use this result to calculate estimates for the number of resolvable points for port switching and wavelength steering. Specifically for wavelength tuning we find:

$$N_{wavelength} \approx Qn_{eff}\frac{\Delta\lambda}{\lambda_0}\frac{v_g}{c} + 1 \tag{4.4}$$

Where Q is the quality factor of the grating, v_g is the group velocity of propagation in the grating, and $\Delta\lambda$ is the bandwidth. This expression exactly resembles what would be extracted from other phase-shifter based architectures which rely on frequency tuning for beam-steering in one direction [56,58,59,63]. We can also estimate the number of resolvable points for steering in plane to be:

$$N_{in-plane} \approx \frac{2D_{peak}}{3} = \frac{2\pi W}{3\lambda} \tag{4.5}$$

4.4 Design

Although the ideal case describes the desired system operation, how to achieve this is less obvious. Thankfully the extensive literature on lens-based solutions in RADAR provides a rich source of design methods and work as was the promise from Fig. 1. There are many types of lens-based solutions invented for use in RADAR, including circularly symmetric Luneburg lenses [74,75], metallic lenses [65], bootlace lenses [76], etc. We focus primarily on the literature developed for shaped wide-angle dielectric lenses [66,67,77] since we can most easily implement these types of solutions in our planar photonic circuit platform.

We design our wide-angle dielectric lens by constraining the shape to satisfy the Abbe Sine condition, which eliminates the Coma aberration (see Appendix) [64,77]. Lenses designed this way tend to have good off angle performance to ± 20 or 30 °. In practice this quantity can translate to a field of view of 80° or more in ϕ_0 (see Appendix). The algorithm for generating the lens takes as input the focal length, lens thickness, and the lens index (the ratio of effective indices of a TE-mode in a SiN slab (n_2) to a SiN slab with a layer of PolySi (n_1)). After creating the lens, the focal plane is identified by conducting ray-tracing through the lens and optimizing the feed position and angle based on maximizing the 2D directivity (from the 1D aperture pattern computed from ray-tracing).

Following this, ray-tracing is done through the grating to compute the full 3D directivity for several optimized port locations and angles. We extract an aperture pattern from ray-tracing through the grating in the following way:

$$\mathrm{A}(x,y) = \left[\sum_{n=0}^{N_{rays}} \sum_{m=0}^{N_{grat}} \sqrt{P_{n,m}}\delta(x - m\Lambda)\delta(y - y_{n,m}) \exp(-qx) \exp\{\}(i\phi_{n,m}) \right] \tag{4.6}$$

Where we have discretely sampled the ray amplitudes $P_{n,m}$ and accumulated ray phases $\phi_{n,m}$ for all N_{ray} by N_{grat}, ray-grating intersections at $[x_{n,m}, y_{n,m}]$. The physical interpretation of this is that each ray-grating intersection acts as a point radiation source driven by the traveling wave (see Appendix). We have also artificially "added" an amplitude decay of $\exp(-qx)$ to account for the grating radiation as the rays propagate. The power associated with a given ray-grating intersection $P_{n,m}$ is calculated from the feed power based on conservation arguments: $P_{feed}(\phi)d\phi = P_{n,m}(y)dy_{n,m}$. We illustrate a full ray-tracing calculation and a 2D aperture pattern extracted by this method in Fig. 5a.

Ray-tracing through the grating is valid in the regime where the grating teeth individually cause low radiation loss, and small incoherent reflections (i.e. the excitation frequency is far from the Bragg bandgap). When correct, this method is useful because it can quickly compute the aperture pattern for a large many wavelength structure while including the effects of lens aberrations and a nonuniform power distribution, two features which would be difficult to model analytically, and very costly to simulate through 2D or 3D FDTD. Our use of ray-tracing to describe planar integrated optical devices builds on experimental and theoretical work conducted by Ulrich et al which directly verified its application to planar prism and lens devices [78], and Ura et al who used ray path length arguments to design the first focusing grating couplers [79].

Following standard RADAR design procedures, once a set of far-field 3D directivity patterns are calculated, new ports are placed to overlap the gain at each port by 3dB to provide suitable coverage in the field of view. To confirm the successful operation of this design, we calculate the aperture patterns for multiple wavelengths between 1500 and 1600 nm for all ports. We include the wavelength dependence of the effective indices and the grating decay factor q. Directivity patterns are plotted in u_x, u_y space for a range of wavelengths in 5b, where the 3dB overlapped ports lie along an elliptical arc, and where the arc is translated forward and backward by tuning the wavelength. These same beams are plotted in 3D in 5c. Note that beams towards the edge of the field of view tend to "fall-into" the device plane because of increasing in-plane momentum (see Appendix).

Ray-tracing was the main method used to design the optical beam-steering chip. The parameters needed for this method such as the port phase centers, feed beam width, grating decay length, and the effective indices were extracted from other calculations. In addition many other simulations were undertaken to validate the assumptions of our ray-tracing computations to account for second order effects. Finally the outcome of the ray-tracing calculations was compared to the analytically predicted directivity functions and beam directions to assess the performance and validity. Once a design was validated, cadence layouts of the necessary components were automatically generated and verified to ensure they satisfied design rule checks based on fabrication limitations and other physics-based constraints.

4.5 Advantages and Disadvantages

This architecture appears to have several advantages over the phase-shifter based solutions. One reason RF lenses were developed, was to minimize the use of phase shifters, which are expensive, lossy, complicated, power hungry and bulky. Some of the same considerations apply here: TO phase shifters are power hungry components, typically using 10 mW or more to achieve π phase shift [80]. For a thousand pixel device, this requires the order of 1000 phase shifters spread out in a 1mm aperture, dissipating 10 Watts. For our system we are restricted to using the same types of phase shifters to operate our MZI switch tree, but we only require $\log_2 N$ of these to be on at a given time, where N is the number of ports, as illustrated in Fig. 6. Consequently a lens-based device with 1000 resolvable points in plane will dissipate 100 times less power for in-plane steering compared to the phase-shifter based approach. In general for our lens-enabled LIDAR, the power budget will be dominated by the optical signal generation, whereas for the phase-shifter architectures, it scales primarily with the feed size.

Most practical phase-shifter approaches, require active feedback to maintain beam coherence, because thermal cross-talk causes changes in the path length of neighboring waveguides. This means either making a measurement of the relative phases on chip through

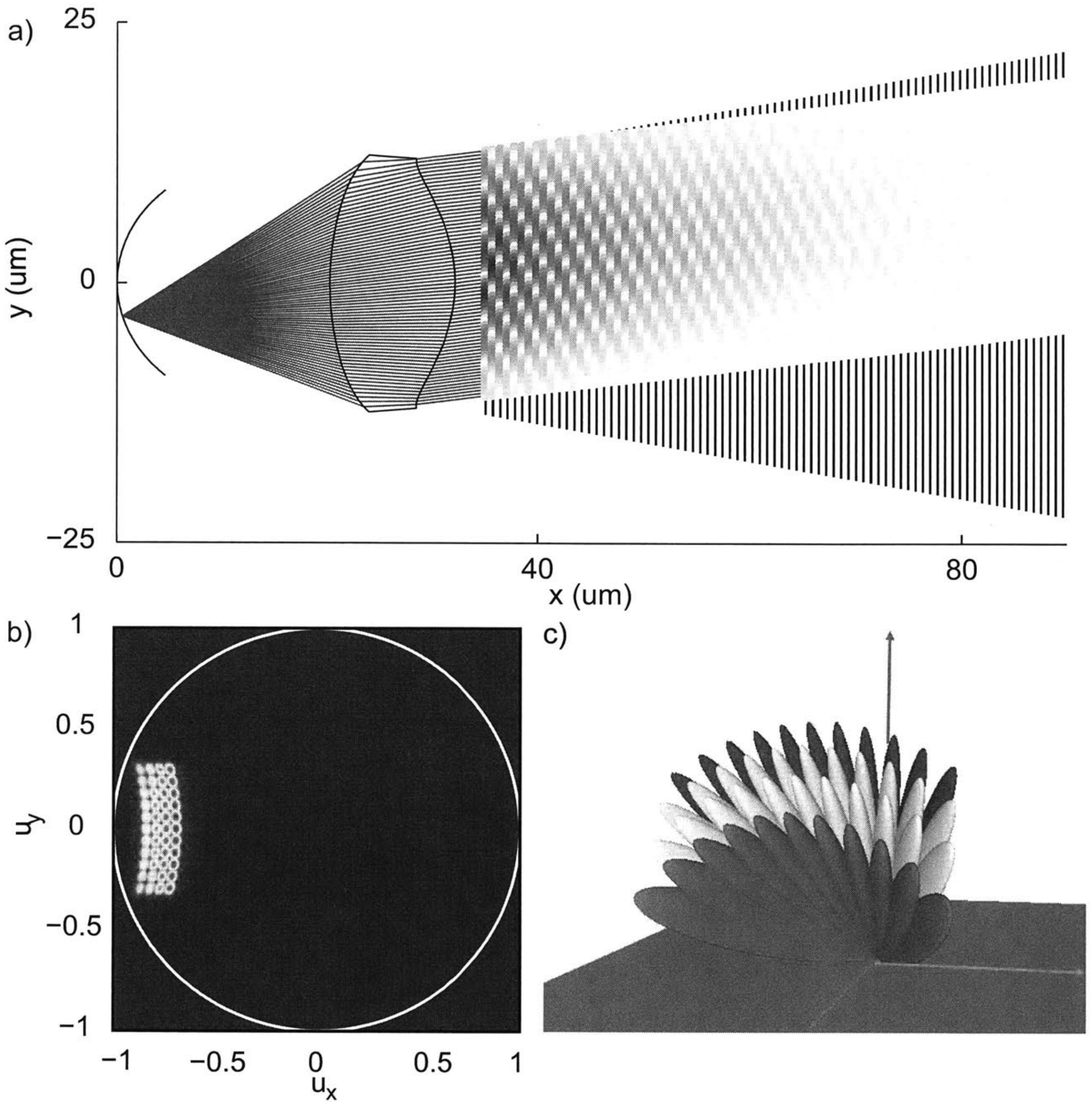


Figure 4-5: Illustration of theoretical ray-tracing and far-field for lens-enabled chip-scale LIDAR generated with $\lambda_0 = 1.55\ \mu$m, $q = 0.025\mu\text{m}^{-1}$, $\Lambda = 700$nm, dutycycle=0.1, feed beamwidth of 15°, and effective indicies $n_1 = 1.39$ and $n_2 = 1.96$ a) Ray-tracing simulations are used to determine the optimal port position and relative angle. These rays are traced through the grating and form an aperture pattern. The Fourier transform of the aperture gives the Far-field pattern. Numerical details of effective index calculations, port phase center, and feed patterns are detailed in the appendix. b) Heat plot showing far-field beam spots in u_x and u_y space. The location of these ports is governed by 4.2, where the beams along the elliptical curve are generated from port switching, while the points formed from translating the ellipse to the right and left correspond to frequency tuning over ±50 nm around $\lambda_0 = 1.55\mu$m. c) 3D beam patterns corresponding to those in b). Here the different colors indicate the different wavelengths used to generate the beam. The drooping effect of the beams as they turn off-axis is caused by increasing in-plane momentum.

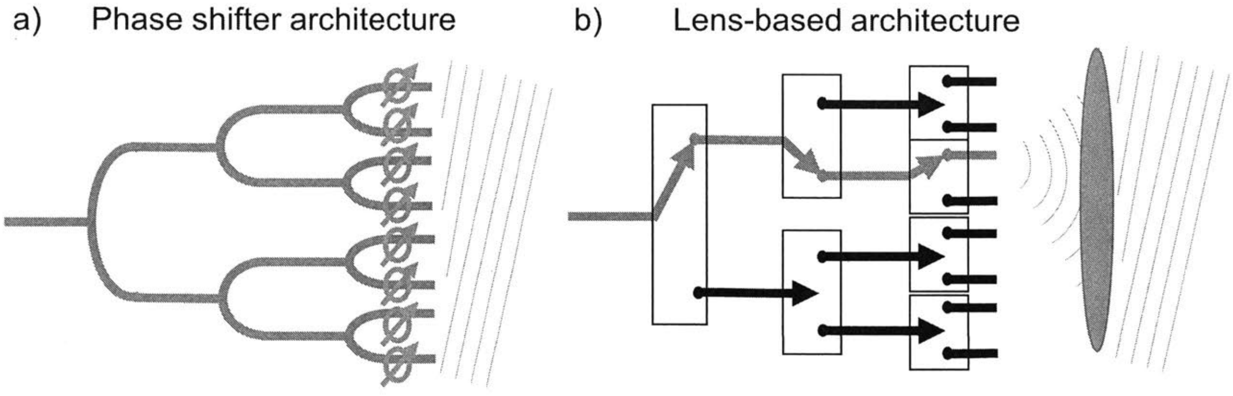


Figure 4-6: Schematic of power requirements for phase-shifter-based and lens-based optical beam-steering. a) In the phase-shifter approach power is uniformly fed to all output antenna elements through a 3dB splitting tree. Approximately N thermal phase shifters are actively cohered to implement in-plane beam steering over N resolvable points. b) In the lens-based approach, we achieve N resolvable points in-plane by switching with an MZI tree between N ports of a dielectric lens feed. The power requirements of this network scale like $\log_2 N$ as compared to N for the architecture in a) because only MZIs associated with the desired optical signal path need to be activated; rest are "off" and draw no power.

lenses and detectors or measuring the beam in free space through an IR camera, to provide feedback [56, 59]. Here we do not have to actively cohere thousands of elements: we can use "binary"-like switching to route the light to the appropriate port, which is a simpler control problem. Lower power consumption additionally makes thermal fluctuations less severe.

We note that the use of a solid $1D$ grating immediately eliminates the problem of grating lobes or high sidelobes that optical phased arrays suffer from. This is at the cost of not being able to "constrain" the ray path to be in the forward direction and results in having to use more material for the grating coupler, hence the triangle shape of the grating feed.

There is an alternate realization of this system, outlined in the Appendix, which can eliminate all TO phase shifters and parallelize the in-plane ports. This architecture parallelizes one scanning direction, as is commonly used in most commercial LIDARs to increase scanning speed. This modification is not possible with the phase-shifter based approach.

A final significant advantage arising from our architecture is the ability to use alternate material systems. One of the chief reasons for using Si for phased-array designs is its large TO coefficient which makes for lower power phase shifters. However, the maximum IR power a single Si waveguide can carry is 50 mW which significantly limits the LIDAR range. SiN has much better properties in the IR and can take the order of 10 W through a single waveguide, as recently demonstrated [81]. However, noting that the power required to operate the phase shifter goes like $\frac{dT}{dn}\sigma$, where $\frac{dn}{dT}$ is the thermo-optic coefficient and σ is the conductivity, we expect that phase shifters on the SiN platform to require at least three times more power than their equivalent Si versions [82,83]. This would exacerbate the power budget and control problems described above for any phase-shifter approach based on SiN. The lens-enabled design can still benefit from using SiN, and greatly improve the potential range, because the feed power is practically negligible.

No architecture is perfect, and there are several non-idealities which can alter the above story for our lens-based solution. The first is the nonuniform field of view of the device, which may cause problems for some applications. Another concern is scaling the number of resolvable points to thousands of pixels in each scanning direction. Although it is simple to ray-trace a lens which would support up to a thousand resolvable points for in-plane scanning, implementing such a lens in practice becomes more and more difficult because the required fabrication tolerances scale as $\frac{1}{N}$ (see Appendix). An additional concern is the impact of lens aberrations on the directivity degradation for the full aperture. Although it was captured by ray-tracing, it was not rigorously modeled to determine the required tolerances and behavior for high Q gratings.

4.6 Experiments and Future Work

In collaboration with Lincoln Labs the components from Fig. 2 are currently being fabricated and tested. We created 4 to 128 port devices based on 20 nm and 40 nm thick PolySi lenses, and gratings 10 nm, 20 nm, and 40 nm thick. The gratings also supported up to 300 resolvable points from wavelength tuning for the 10 nm variants over a 100 nm bandwidth. Because of fabrication constraints, the grating PolySi height had to be the same as the lens height. This resulted in tradeoffs, because thicker gratings had lower Qs, but thicker lenses had a better index contrast and could support more resolvable points. MZI switching trees of up to 128 ports are also being fabricated. Testing will verify the expected number of resolvable points from port switching and wavelength tuning and confirm successful operation of the MZI switch matrix. Finally InP amplifier arrays 128 amplifiers in size will be fabricated at Lincoln Labs and integrated in-between the SiN switch matrix and SiN passive beamforming chips [?]. Integrating sources with the SiN platform is an important step to generating a commercial integrated LIDAR chip.

The devices being fabricated this year are depicted in Fig. 7 in a solid black box, consisting of the switch matrix, amplifier array, and passive beam-forming chip. Future work on this optical beam-steering platform will include further exploring the literature of RADAR lens solutions, such as bifocal lenses, integrated Luneburg lenses, Rotman lenses, and many others [64, 66, 68, 71, 84, 85]. It also may be fruitful to investigate standard compound lens architectures such as achromatic doublets to deal with the high waveguide dispersion of the lens index, which becomes a problem for systems supporting 1000 resolvable points (see Appendix). Although the overall design approach is simple, there are a myriad of tasks ranging from enhancing the upward emission of the gratings to minimizing the spillover loss of the feed power on the lens, which can improve the efficiency, robustness, manufacturability, and performance of the system.

Future work at Lincoln Labs will consist of integrating tunable sources on chip, separate transmit and receive apertures or an on-chip circulator, integrated Ge detectors for heterodyne detection, and finally external electrical control and signal processing as shown in Fig. 6. Bringing all of these technologies together compactly, cheaply, and robustly will be challenging, but the reward will be a new sensor capable of supporting the next generation of autonomous machines.

4.7 Appendix

4.7.1 Alternate Realizations

In Fig. 8 we depict a realization of the system based on removing the MZI switch matrix and replacing this with parallel illumination of all ports through a 3dB splitting tree. Beam-steering of this “fan-beam” is accomplished through frequency tuning. A separate

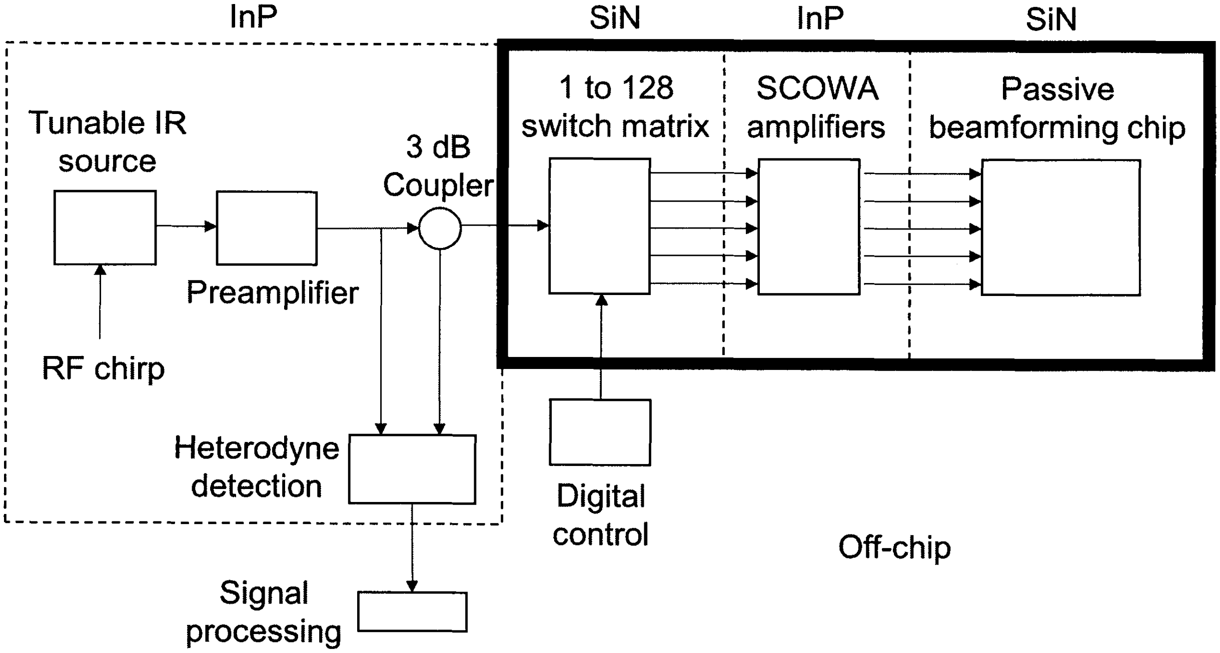


Figure 4-7: Block diagram of full integrated LIDAR chip. The products for this year's project are surrounded by a solid black line and consist of the SiN MZI switch matrix, the InP amplifier array, and the SiN feed, lens, and grating chip. These functional blocks implement a nonmechanical optical beam steering system. On the left hand side, we show several elements required to form a working LIDAR chip, including a tunable on-chip source, a source for an RF chirp to be modulated onto this source, heterodyne "IQ" detection (through III-V) diodes, and external electrical control to drive the optical-switch matrix and tunable source.

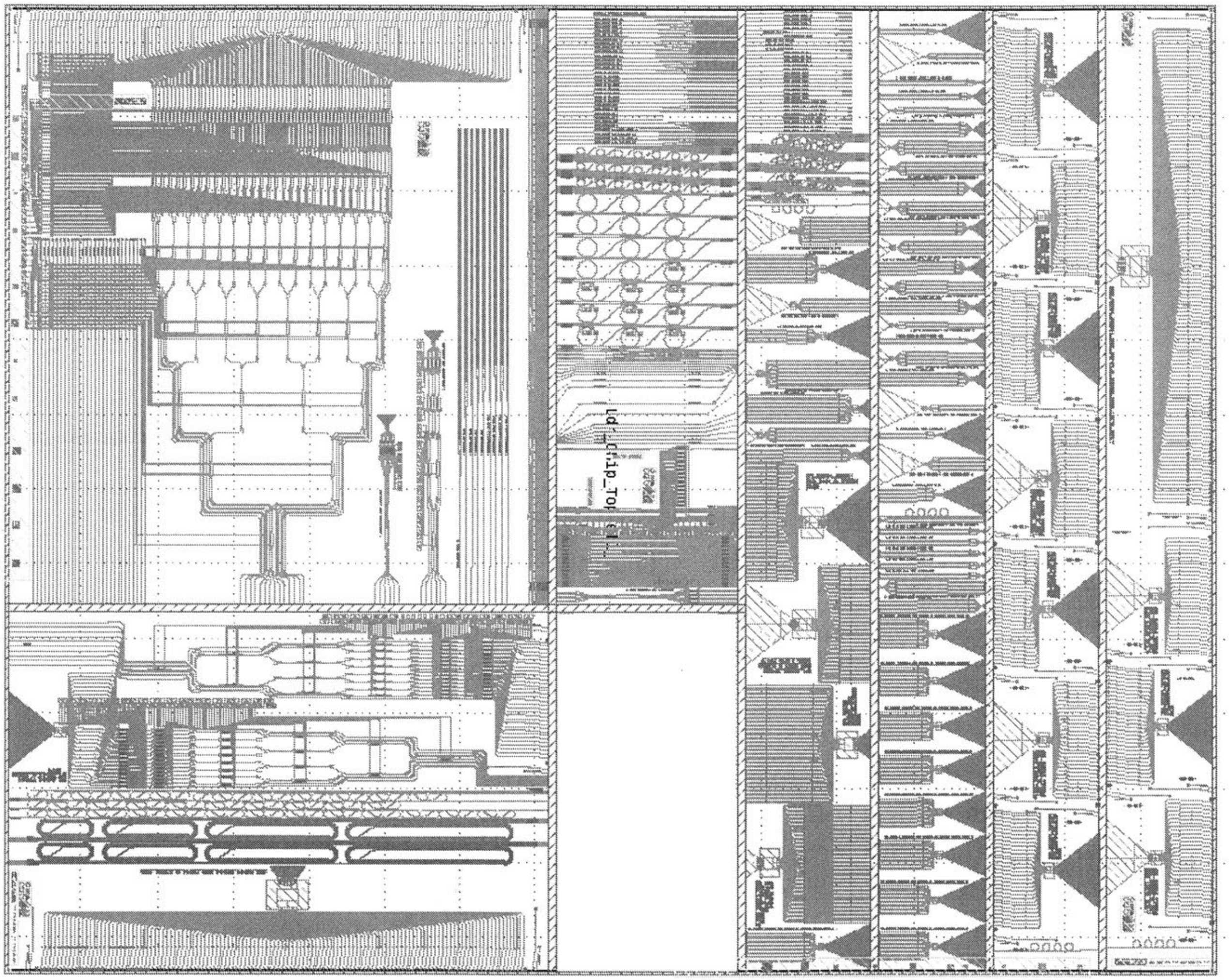


Figure 4-8: Silicon Nitride Lithography Mask currently being fabricated at MIT Lincoln Labs. Switch trees, passive beamforming chips, and other test structures are visible.

aperture with an array of detectors is employed for processing the LIDAR return. Most commercial LIDAR architectures employ some parallelization to increase the rate of point cloud acquisition. At 100 m, a raster scan for a megapixel scale sensor would take 1s, based on time-of-flight, far too slow for most applications, whereas parallelization can allow an acquisition on the scale of ms. The tradeoff is that the power requirements are increased by a factor of N, and more detector hardware is required.

4.7.2 Analytics

Far-Field Angles

The aperture phase is the most important quantity for our LIDAR antenna, and is completely determined by the initial ray directions, the grating parameters, and the wavelength. Since we are propagating through a straight grating, the plane wave k_y generated by the lens feed system will be conserved, so $k_{y,avg} = k_y$. k_x on the other hand, will be more complicated because it changes at each step of the grating. Assuming an initial in-plane angle of ϕ_{in}, an index of the starting medium n_1, an index of the steps n_2 and a step duty cycle d, we find that k_y and $k_{x,avg}$ are given by the following:

$$\begin{aligned} k_y &= n_1 k_0 \sin(\phi_{in}) \\ k_{x,avg} &= d n_1 k_0 \cos(\phi_{in}) + (1-d) n_2 k_0 \sqrt{\left(1 - \frac{n_1^2}{n_2^2}\sin^2(\phi_{in})\right)} \end{aligned} \tag{4.7}$$

It's important to note that the effective indices for the grating n_1 and n_2 are also functions of the wavelength. To compute the emission angle of this aperture, we perform phase matching between these wavevectors and those of a free-space plane wave with $\vec{k} = k_0[\sin(\theta_0)\cos(\phi_0), \sin(\theta_0)\sin(\phi_0), \cos(\theta_0)]$.

$$\begin{aligned} k_0 \sin(\theta_0)\sin(\phi_0) &= n_1 k_0 \sin(\phi_{in}) \\ k_0 \sin(\theta_0)\cos(\phi_0) &= k_{x,avg}(\phi_{in}) - \frac{2\pi m}{\Lambda} \end{aligned} \tag{4.8}$$

Here we have subtracted a crystal momentum $\frac{2\pi m}{\Lambda}$, which originates from the discrete and periodic sampling implemented by the scattering from each grating step. We can rearrange this to derive the following expressions for the far-field angles:

$$\begin{aligned}\theta_0 &= \text{asin}\left(\frac{\sqrt{(k_{avg,x}(\phi_{in}) - \frac{2\pi m}{\Lambda})^2 + (n_1 k_0 \sin(\phi_{in}))^2}}{k_0}\right)\\ \phi_0 &= \text{atan}\left(\frac{n_1 k_0 \sin(\phi_{in})}{k_{avg,x}(\phi_{in}) - \frac{2\pi m}{\Lambda}}\right)\end{aligned} \tag{4.9}$$

We want to understand how the far-field angles depend on the in-plane angle ϕ_{in} and the wavelength λ. We can immediately identify that ϕ_0 will be significantly greater than ϕ_{in}. This results from the grating momentum $\frac{2\pi m}{\Lambda}$ being subtracted from $k_{avg,x}$ in the denominator. This means that relatively small variations in the input angle will greatly change the output in-plane angle ϕ_0, sweeping it across the field-of-view. This feature ultimately allows us to use the lens in a small angle, aplanatic regime.

As we sweep ϕ_{in} we also expect variations in θ_0. By examining Eq. (4.9), we see that the argument of asin seems to increase with ϕ_{in}, and ultimately exceed 1, confining the beam in-plane. We can derive this cutoff condition (where $\theta_0 = 0$) more precisely by taking $k_{avg,x}(\phi_{in}) \approx n_{eff} k_0 cos(\phi_{in})$ where $n_{eff} = dn_1 + (1-d)n_2$. Even though n_{eff} is not truly constant, and it's variations significantly effect the far-field angles, qualitatively this description holds. We find that the cutoff angle ϕ_{cut} satisfies:

$$1 = \left[n_{eff}\cos(\phi_{cut}) - \frac{\lambda}{\Lambda}\right]^2 + n_1^2\left[1 - \cos(\phi_{cut})^2\right] \tag{4.10}$$

This can be easily rearranged into a quadratic equation and solved for ϕ_{cut}. To get more intuition into the behavior of this angle, we examine the case of normal (or broadside) emission at $\phi_{in} = 0$ and approximate $n_1 \approx n_{eff}$. Working this out we find:

$$\cos(\phi_{in}) \approx 1 - \frac{1}{2n_{eff}^2} \tag{4.11}$$

Where $\tan(\phi_0) \approx -\frac{n_{eff}}{2}$ gives the corresponding ϕ_0 at this point. It makes sense that the magnitude of the index will control the ϕ_{in} because it determines how rapidly k_y increases as we move off axis. Overall we can envision how $\vec{k}$ evolves as a function of ϕ_{in}: starting from emission normal to the surface, as we adjust ϕ_{in} away from 0, $\vec{k}$ turns rapidly to one side and falls into the plane.

We can further visualize this trajectory by rearranging Eq. (4.8). Taking $u_{x,0} = \sin(\theta_0)\cos(\phi_0)$ and $u_{y,0} = \sin(\theta_0)\sin(\phi_0)$, we can manipulate Eq. (4.8) to find:

$$\left[\frac{u_{x,0} + \frac{m\lambda}{\Lambda}}{n_{eff}}\right]^2 + \left[\frac{u_{y,0}}{n_1}\right]^2 = 1 \tag{4.12}$$

This is an ellipse centered at $[-\frac{m\lambda}{\Lambda}, 0]$. As ϕ_{in} is varied, the emission direction will traverse an arc of this ellipse in u_x, u_y space. Tuning the wavelength λ will translate this ellipse forward and backward in the u_x direction. The total field-of-view in u_x, u_y space will have the form of a curved band whose thickness will be controlled by the total wavelength tuning range. We discuss the number of 3dB overlapped beams we can fit inside this field-of-view in the Number of Resolvable Points section.

Far-Field directivity

We can begin our derivation of the far-field pattern by noting that we can completely specify the near-field amplitude to have the following form:

$$\mathrm{A}(x, y) = \exp(-qx)\exp(ik_0 u_{x0} x)\exp(ik_0 u_{y0} y) \tag{4.13}$$

Where $u_{x0} = \sin(\theta_0)\cos(\phi_0)$ and $u_{y0} = \sin(\theta_0)\sin(\phi_0)$

We have implicitly assumed a rectangular beam profile along the y-direction to simplify our calculations. In general we expect an additional function $f(y, x)$ to modulate the amplitude of the pattern according to the feed pattern, illumination position, and lens geometry. This derivation however captures the most critical features of the far-field pattern and establishes an upper bound on the gain. In addition the performance of the aperture is largely determined by its phase behavior, so smearing the amplitude distribution relative to the ideal tends to lead to small changes.

The physical aperture we are integrating over will be a parallelogram bounded by the following conditions:

$$0 \leq x \leq L \tag{4.14}$$

$$-\frac{W}{2} + x\tan(\phi') \leq y \leq \frac{W}{2} + x\tan(\phi') \tag{4.15}$$

Here ϕ' is equal to atan $\left(\frac{n_1 \sin(\phi_{in}) k_0}{k_{x,avg}}\right) \approx$ atan $\left(\frac{n_1 \sin(\phi_{in})}{n_{eff}\cos(\phi_{in})}\right)$ and is close in magnitude to ϕ_{in} from the previous section, but not identical because of the refraction at the grating steps. To find the far-field pattern we can compute the Fourier transform of this amplitude pattern over the domain:

$$\mathrm{F}(u_x - u_{x0}, u_y - u_{y0}) =$$

$$\int_0^L dx \int_{-\frac{W}{2}+x\tan\phi'}^{\frac{W}{2}+x\tan\phi'} dy \exp(-qx)\exp(-ik_0(u_x - u_{x0})x)\exp(-ik_0(u_y - u_{y0})y) \tag{4.16}$$

Where $u_x = \sin(\theta)\cos(\phi)$ and $u_y = \sin(\theta)\sin(\phi)$, which are the direction angles. For convenience, from here we denote $u_x - u_{x0}$ with Δu_x and $u_y - u_{y0}$ with Δu_y. We can evaluate these integrals easily to find:

$$\mathrm{F}(\Delta u_x, \Delta u_y) = \frac{\sin\left(k_0 \frac{W}{2}\Delta u_y\right)}{k_0 \Delta u_y} \frac{1 - \exp(-qL)\exp(ik_0 L(u_x + \tan(\phi')\Delta u_y))}{q + ik_0(u_x + \tan(\phi')\Delta u_y)} \tag{4.17}$$

The power of the far-field pattern is the magnitude of the field pattern squared, that is $\mathrm{P} = |\mathrm{F}|^2$. We use the power P below to compute the directivity of the far-field pattern with the following expression:

$$\mathrm{D}(\theta, \phi) = \frac{\mathrm{P}(\theta, \phi)}{\frac{1}{2\pi}\int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\phi \mathrm{P}(\theta, \phi)} \tag{4.18}$$

The directivity essentially gives the factor of the power emitted in a given direction relative to an isotropic radiator. A well-designed directional antenna will tend to maximize the peak gain, the directivity of the main lobe, and minimize the power into sidelobes, because these waste power and contribute to false detections. We will discuss the directivity more later concerning the range of the system and the number of resolvable points.

We can create a simpler expression by expanding the direction angles about the far-field peak at $\theta = \theta_0$ and $\phi = \phi_0$, we can also take the limits of the integral to infinity. This creates negligible error in the case of high-gain beams and ultimately allows many of these gain integrals to be evaluated analytically:

$$\mathrm{D}(\Delta\theta, \Delta\phi) \approx \frac{\mathrm{P}(\Delta\theta, \Delta\phi)}{\frac{\sin\theta_0}{2\pi}\int_{-\infty}^{\infty} d\Delta\theta \int_{-\infty}^{\infty} d\Delta\phi \mathrm{P}(\Delta\theta, \Delta\phi)} \tag{4.19}$$

It's illustrative to change coordinates of this expression from θ and ϕ to Δu_x and Δu_y. We can find:

$$\Delta u_x = \cos(\theta_0)\sin(\phi_0)\Delta\theta + \sin(\theta_0)\cos(\phi_0)\Delta\phi = u_x - u_{x,0} \tag{4.20}$$

$$\Delta u_y = \cos(\theta_0)\cos(\phi_0)\Delta\theta - \sin(\theta_0)\sin(\phi_0)\Delta\phi = u_y - u_{y,0} \tag{4.21}$$

Where we have taken $\phi = \phi_0 + \Delta\phi$ and $\theta = \theta_0 + \Delta\theta$. We can use these expressions to calculate the following Jacobian, where we have changed variables from ϕ and θ to $\Delta\phi$ and $\Delta\theta$:

$$\begin{aligned} d\Delta\theta d\Delta\phi \sin(\theta_0) &= d\Delta u_x d\Delta u_y \sin(\theta_0) \begin{vmatrix} \frac{\sin(\phi_0)}{\cos(\theta_0)} & \frac{\cos(\phi_0)}{\cos(\theta_0)} \\ \frac{\cos(\phi_0)}{\sin(\theta_0)} & -\frac{\sin(\phi_0)}{\sin(\theta_0)} \end{vmatrix} \\ &= \frac{d\Delta u_x d\Delta u_y}{\cos(\theta_0)} \end{aligned} \tag{4.22}$$

Taken together, we can use these results to rewrite our expression for the peak gain as a function of Δu_x and Δu_y:

$$\mathrm{D}(\Delta u_x, \Delta u_y) = \frac{2\pi\cos(\theta_0)\mathrm{P}(\Delta u_x, \Delta u_y)}{\int_{-\infty}^{\infty} d\Delta u_x \int_{-\infty}^{\infty} d\Delta u_y \mathrm{P}(\Delta u_x, \Delta u_y)} \tag{4.23}$$

This unsimplified expression can already tell us something very useful, namely that the peak gain of a given pattern is directly proportional to $\cos(\theta_0)$. This result emerges because the far-field gain in general is proportional to the projected area. To first order, neglecting additional aberrations and changes in the grating parameters, effective indices, reflections, and feed illumination, the peak gain fall-off as a function of angle is just determined by the angle between the emission vector and the z-axis. The other important feature of this equation is that the peak shape is essentially independent of the center of the main lobe: to leading order the pattern just changes by the $\cos(\theta_0)$ scale factor.

We can directly evaluate these integrals for our far-field pattern:

$$\begin{aligned}\int_{-\infty}^{\infty} d\Delta u_x \int_{-\infty}^{\infty} d\Delta u_y \mathrm{P}(\Delta u_x, \Delta u_y) = \\ \int_{-\infty}^{\infty} d\Delta u_x \int_{-\infty}^{\infty} d\Delta u_y \frac{\mathrm{sinc}^2(\frac{W}{2}k_0\Delta u_y)}{1+\frac{k_0^2}{q^2}(\Delta u_x + \tan(\phi)\Delta u_y)^2} \times \\ (1 - 2\cos(k_0 L(\Delta u_x + \tan(\phi)\Delta u_y))\exp\{(-qL)\} + \exp\{(-2qL)\}) \quad (4.24)\end{aligned}$$

We first start by performing a shear transformation on the integrating variables given by: $\Delta u_{x,s} = \Delta u_x + \tan(\phi)\Delta u_y$ and $\Delta u_{y,s} = \Delta u_y$. With this transformation the integral now becomes separable:

$$\begin{aligned}\int_{-\infty}^{\infty} d\Delta u_x \int_{-\infty}^{\infty} d\Delta u_x \mathrm{P}(\Delta u_x, \Delta u_y) = \\ \int_{-\infty}^{\infty} d\Delta u_{x,s} \frac{(1 - 2\cos(k_0 L \Delta u_{x,s})\exp\{(-qL)\} + \exp\{(-2qL)\})}{1+\frac{k_0^2}{q^2}\Delta u_{x,s}^2} \int_{-\infty}^{\infty} d\Delta u_{y,s} \mathrm{sinc}^2(\frac{W}{2}k_0 \Delta u_{y,s}) \\ (4.25)\end{aligned}$$

Note that the angle θ_p does not change the projected area of the aperture, since it just shears the emitting surface. Consequently we expect it to completely drop out of the integral, which is indeed the case. Next we remove the dimensions and break the integrals into parts and evaluate:

$$\begin{aligned}\frac{2q}{Wk_0^2}\int_{-\infty}^{\infty} dx \left(\frac{(1+\exp\{(-2qL)\})}{1+x^2} - \frac{2\cos(qLx)\exp\{(-qL)\}}{1+x^2}\right)\int_{-\infty}^{\infty} dy \left(\mathrm{sinc}^2(y)\right) \\ = \frac{2q\pi^2}{Wk_0^2}\left((1+\exp\{(-2qL)\}) - 2\exp\{(-2qL)\}\right) \quad (4.26)\end{aligned}$$

Using these results, finally we can write the directivity as:

$$\mathrm{D}(\Delta u_x, \Delta u_y) = \frac{W k_0^2 \cos(\theta_0)}{\pi q (1 - \exp\{(-2qL)\})} \frac{\mathrm{sinc}^2(\frac{W}{2} k_0 \Delta u_y)}{1 + \frac{k_0^2}{q^2}(\Delta u_x + \tan(\phi')\Delta u_y)^2} \times \left(1 - 2\cos(k_0 L(\Delta u_x + \tan(\phi')\Delta u_y)) \exp\{(-qL)\} + \exp\{(-2qL)\}\right) \quad (4.27)$$

We can also expression the peak directivity as:

$$D_{max} = \cos(\theta_0) \frac{W k_0^2}{\pi q} \frac{1}{1 + \frac{k_0^2}{q^2}} \frac{(1 - \exp\{(-qL)\})^2}{(1 - \exp\{(-2qL)\})} \quad (4.28)$$

To gain a little insight into how this function behaves, we can simplify it for large and small L. For $L << \frac{1}{q}$, we find:

$$\lim_{qL \to 0} \mathrm{D}(\Delta u_x, \Delta u_y) = \frac{W L k_0^2 \cos(\theta_0)}{2\pi} \mathrm{sinc}^2\left(\frac{W}{2} k_0 \Delta u_y\right) \mathrm{sinc}^2\left(\frac{L}{2} k_0 (\Delta u_x + \tan(\phi')\Delta u_y)\right) \quad (4.29)$$

This is just the directivity from a sheared rectangular aperture of length L and width W, note that the peak gain is $\frac{W L k_0^2 \cos(\theta_0)}{2\pi}$, which is directly proportional to the projected area $W L \cos(\theta_0)$. Taking the opposite limit, we can find another useful simplification:

$$\lim_{qL \to \infty} \mathrm{D}(\Delta u_x, \Delta u_y) = \frac{W k_0^2 \cos(\theta_0)}{\pi q} \frac{\mathrm{sinc}^2(\frac{W}{2} k_0 \Delta u_y)}{1 + \frac{k_0^2}{q^2}(\Delta u_x + \tan(\phi')\Delta u_y)^2} \quad (4.30)$$

Here the peak directivity scales as $\frac{W}{q}\cos(\theta_0)$, where $\frac{1}{q}$ becomes the effective length of the aperture. Even though these expressions are much simpler than the general one we derived, even if the aperture is several decay lengths long, the effect of the finite length of the directivity is significant and properly modeling it requires the full expression. An example of this is in computing the number of far-field resolvable points.

Number of Resolvable points with wavelength tuning

One of the most important properties of our system is the number of far-field resolvable points. There are some relatively simple expressions we can derive which will tightly bound the number of resolvable points we can achieve in a particular system as a function of the aperture parameters. We will first start with the number of resolvable points we can achieve through wavelength tuning. Assuming normal incidence from the feed, the far-field condition for the unit vector in the x-direction is just: $n_{eff}k_0 - \frac{2\pi}{\Lambda}m = k_0 u_x$, where u_x is the unit vector of the wavevector in the x-direction, n_{eff} is the effective index of the grating at normal incidence, Λ is the grating period, and m is the grating order.

We want to count the number of full-width half-maximums Δu_{FWHM}, we can fit inside a total tuning range of Δu_{range}. Δu_{range} in this case is just given by $\frac{m\Delta\lambda}{\Lambda}$, where $\Delta\lambda$ is the tuning wavelength, and is typically 50-100 nm for integrated tunable sources. With this we can write a simple expression for the number of resolvable points with wavelength tuning $N_{wavelength}$:

$$N_{wavelength} \approx \frac{m\Delta\lambda}{\Lambda_0 \Delta u_{FWHM}} + 1 \tag{4.31}$$

In the case of a finite length grating, Δu_{FWHM} has to be computed numerically from the full directivity formula to give a precise calculation of the number of resolvable points. However, in the case of a long grating, we can determine exactly that $\Delta u_{FWHM} = \frac{2q}{k_0}$. Plugging this in gives the following relationship:

$$N_{wavelength} \approx \pi \frac{\Delta\lambda}{\lambda_0} \frac{m}{q\Lambda_0} + 1 \tag{4.32}$$

Assuming that at λ_0, that the grating is emitting at normal incidence, and using our expression relating the decay length q to the grating quality factor Q, we find that:

$$N_{wavelength} \approx Q n_{eff} \frac{\Delta\lambda}{\lambda_0} \frac{v_g}{c} + 1 \tag{4.33}$$

Number of resolvable points from in-plane steering

We assume that we can determine the number of resolvable points from the field pattern at the lens aperture, as opposed to the pattern after being emitted from the grating.

In 1D the directivity of a far-field pattern $A(\theta)$ is defined by $\frac{A}{\frac{1}{\pi}\int A(\theta)d\theta}$. Assuming that the power is confined to a single lobe of angular width $\Delta\theta$, we can approximate D_{peak} as $\frac{\pi}{\Delta\theta}$. Neglecting lens aberrations, the directivity can be written:

$$\mathrm{D}(\theta) = \frac{\pi}{\Delta\theta}\cos(\theta) \tag{4.34}$$

The steering range in this situation is limited by the minimum acceptable gain usable by the system. Typically RADARs are designed to have a directivity fall-off of 3dB or 0.5 at the edge of their usable FOV. This gives an effective steering range of $\frac{2\pi}{3}$ radians. Conveniently approximating the beam-width to be constant, we find that:

$$N_{in-plane} \approx \frac{2D_{peak}}{3} = \frac{2\pi W}{3\lambda} \tag{4.35}$$

Where we have substituted in the peak directivity of a rectangular aperture of size W. This equation accurately reflects the scaling of the number of resolvable points when $\frac{a}{\lambda}$ is between 10 and 40 or so. Beyond this, the path error for off-axis scanning angles begins to become an appreciable fraction of the wavelength (since the error is directly proportional to the lens size). The 3dB scanning limit will be squeezed inwards as $\frac{a}{\lambda}$ increases.

Abbe Sine condition

As shown in [64, 77], we can shape a lens to satisfy the Abbe sine condition. Satisfying the Abbe sine condition eliminates Coma aberration on-axis and reduces it off axis in the regime where $\sin(\phi) = \phi$. We briefly outline the procedure for generating a shaped lens given input parameters thickness T, focal length F, effective focal length F_e, and index n. The inner surface of the lens is defined by r, θ, while the outer surface is defined by x, y. In this coordinate system, we satisfy the Abbe sine condition when $y = F_e \sin(\theta)$. We can further relate r and θ to x and y from the following expression calculated from ray-propagation:

$$r + n\sqrt{(y - r\sin(\theta))^2 + (x - r\cos(\theta))^2} - x = (n-1)T \tag{4.36}$$

This can be written as a quadratic equation for x and solved. Once x is solved, r can be advanced by computing:

$$\frac{dr}{d\theta} = \frac{nr\sin(\theta - \theta')}{n\cos(\theta - \theta') - 1} \tag{4.37}$$

Where:

$$\theta' = \tan^{-1}\left[\frac{(F_e - r)\sin(\theta)}{x - r\cos(\theta)}\right] \tag{4.38}$$

These equations can be solved iteratively to generate the entire lens surface, beginning with $\theta = 0$ and $r = F$. Other methods can be used to generate shaped lens surfaces such as designing the aperture power pattern based on the feed power pattern, or by forcing the lens to have two off axis focal points.

LIDAR Range

The minimum detectable received power $P_{r,min}$ from a LIDAR return determines the maximum range of the device. $P_{r,min}$ is determined by the integration time and sensor architecture, which can be based on frequency modulated continuous wave (FMCW) or pulsed direct detection type schemes. If a target has a cross section σ, the maximum range we can observe that target is given by the standard RADAR equation:

$$P_{r,min} = \frac{D(\theta,\phi)^2\eta^2}{(4\pi)^3}\frac{\lambda^2}{R_{max}^4}\sigma P_t \tag{4.39}$$

Where $D(\theta, \phi)$ is the directivity, η is the device efficiency, R_{max} is the maximum range, and P_t is the transmitter power. We see here that the primary determinant of the LIDAR preformance beyond the detection backend are the antenna characteristics given by $D(\theta, \phi)$ and η.

In the case that the beam spot from the LIDAR is contained completely within the target, which is a common application mode for LIDARs, we can derive an alternate constraint, which is more forgiving than the standard RADAR range equation in terms of distance

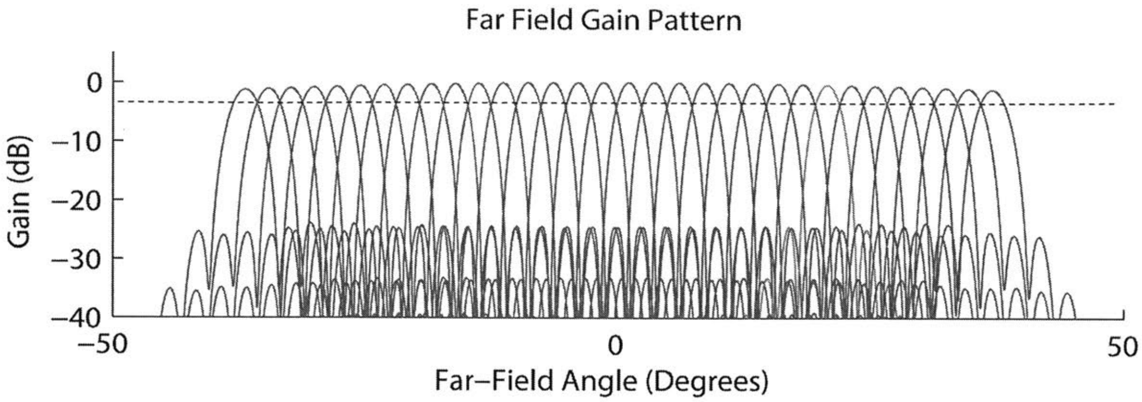


Figure 4-9: 32, 3-dB overlapped far-field beam patterns. Single port excitation is in red. Peaks to each side of main peak are sidelobes. These represent power radiated in unintended directions and may result in false detections. Dotted line indicates -3dB. Port positions are designed to overlap far-field resolvable spots by 3dB. For 2D aperture this is done along parameterized curve between each spot in u_x, u_y space.

falloff:

$$P_{r,min} = \frac{D(\theta, \phi)\eta}{(4\pi)^2} \frac{\lambda^2}{R_{max}^2} P_t \tag{4.40}$$

4.7.3 Overview of numerical methods and verification

Because of the large scale of the structures used and their lack of periodicity, full 3D FDTD simulations were not possible. However, we were able to do smaller 2D and 3D FDTD simulations of individual components to help verify the system performance.

First we conducted simulations of the waveguides generated from the routing algorithms to verify that they were defined with enough points, were not too close, and satisfied minimum bend radius requirements. Unfortunately having 3dB spaced far-field spots, results in wavelength spaced ports in the focal surface. Although the waveguides can be wavelength-spaced for short lengths without significant coupling, generally the feed geometry results in excessively high coupling between waveguides. We fixed this problem by decimating the ports by a factor of two.

Next we simulated the interface between the waveguide and the SiN slab. With these simulations, we visualized the creation of an effective point dipole source when the waveguide terminated at the slab, and determined the far-field radiation pattern and effective phase

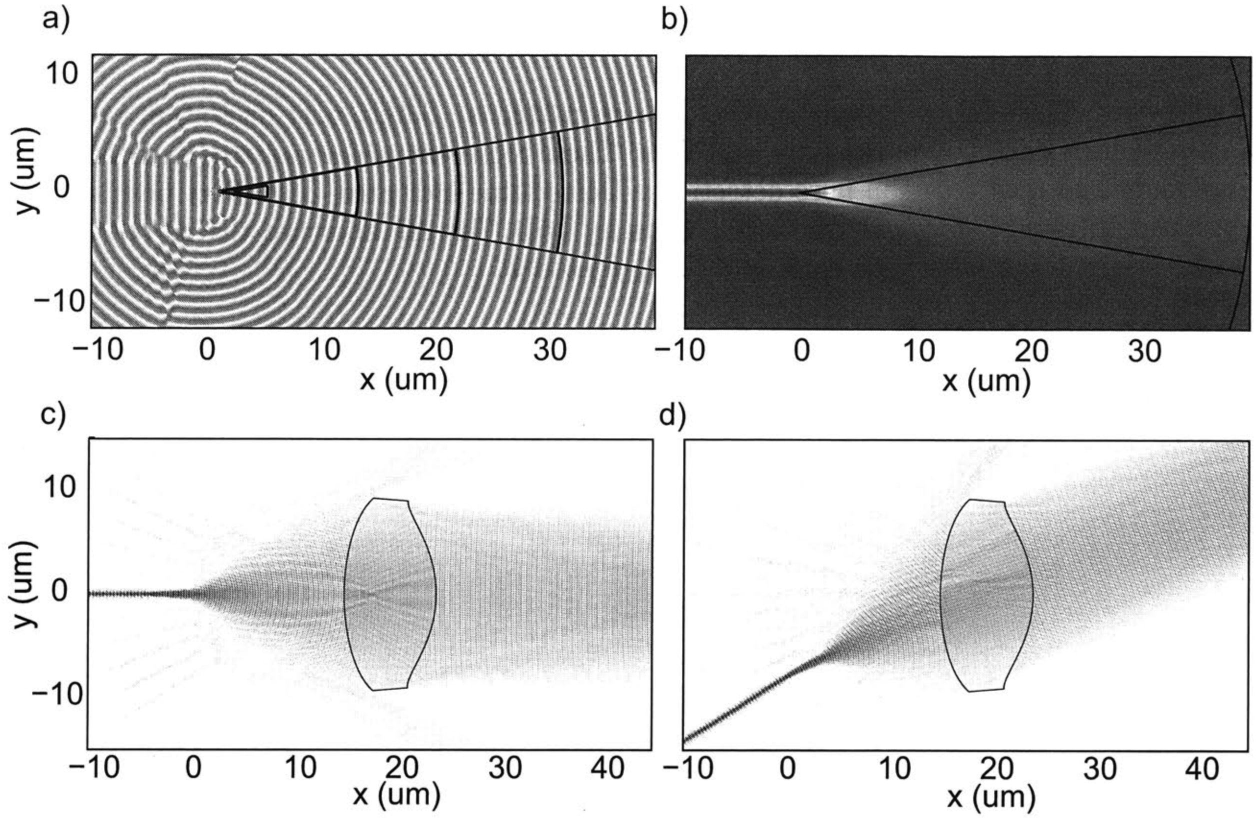


Figure 4-10: Simulations for design of chip-scale LIDAR. a) Simulation of far-field beam pattern to extract phase center. b) Simulation of far-field beam pattern to extract gaussian beamwidth. c) 2D simulation of on-axis port excitation of lens feed. d) 2D simulation of off-axis port excitation of lens feed.

center [86, 87]. By fitting circles to the phase front, we were able to extract a location for the phase center which we show in Fig. 10a. The phase front was approximately 1 μm behind the interface between the waveguide slab and the waveguide. The far-field power as a function of angle was well described by a Gaussian with a beam-width of 13.5 °, as shown in Fig. 10b. The phase center and beam-width did not change if the waveguide was incident at an angle on the interface: the phase center and beam-profile remained the same relative to the orientation of the waveguide. This feature was useful because it allowed us to angle the waveguides to minimize the spillover loss (the radiation that misses the lens), without having to be concerned about the beam-width or phase center changing.

Another set of simulations we preformed concerned the interface of the lens and the slab. One of the key assumptions of this project, which has also been a feature of other works on integrated planar lenses [78], is that we can describe the in-plane propagation in terms of the effective mode indices. We did several calculations of TE slab modes impacting 20 nm and 40 nm Si slab "steps" to verify this assertion, and to quantify the radiation loss at these interfaces. We found that for a wide range of angles, the radiation loss was less than 5 % in line with previous experimental results for incident angles less than 40 ° [88].

We also performed effective 2D FDTD simulations of the waveguides, the lens feed, and the lens itself to verify that beam-steering worked properly. We see this in Fig. 10c and d. We confirmed the directivity derived from these simulations closely matched those produced by ray-tracing. We also verified that the expected lens roughness for fabrication would not result in excessive gain degradation [89]. 2D grating simulations of a 1D grating were used to extract the grating Q as a function of wavelength. The emission angle was compared to that predicted from the average grating index and good agreement was obtained. Additionally we modeled the photonic bandstructure using meep to confirm that our excitation was far from the Bragg band edge.

Finally we extracted the grating Q as a function of angle and wavelength from meep calculations. We confirmed the on-axis performance matched that predicted from the FDTD simulations. Additionally we confirmed that the Q did not change too much for off-axis propagation. Generally the behavior within ±20 ° was well-behaved, but beyond that there were large fluctuations. For the ray-tracing simulations, in the regime we were interested in, the grating Q could be considered constant, but in general we found it was a complicated, rapidly varying function. Future studies should attempt to rigorously model Q as a function of angle because the dependence can be unexpectedly strong in certain regimes.

4.7.4 Index error

The effective index ratio n_2/n_1 of the experimental system will be different than that used in ray-tracing simulations, because of finite fabrication tolerance, wavelength dispersion, temperature variation, etc. In general an error in the index will cause the focal plane to shift by some amount. For a parabolic lens, we find that the change is:

$$\Delta f = \frac{f^2}{R}\Delta n \tag{4.41}$$

Since the depth of focus scales as λ, $R \sim f$, and $f \sim \lambda N$, where N is the number of resolvable points, we have that our effective index tolerance scales inversely with the number of resolvable points that the imaging system supports:

$$\Delta n \sim \frac{1}{N} \tag{4.42}$$

Without any kind of external tunablity, meeting this constraint for large N becomes increasingly difficult. For more than 100 ports, wavelength dispersion over a 100 nm bandwidth already exceeds this constraint for a 40 nm thick lens. In general for proper operation of a device with 100 ports at a single wavelength we require better than ±1 nm of precision in the layer heights, and better than 0.01 tolerance in the material index. Addressing these index tolerance issues is one of the most important engineering tasks required for scaling the system to 1000s of resolvable points.

Chapter 5

Optical Convolutional Neural Networks

5.1 Introduction

Exploration of neuromorphic computing architectures began in the late 1950s with the invention of the preceptron, which functioned as a binary classifier with a linear decision boundary [90]. The preceptron worked well for certain tasks, but further progress was hindered by a lack of understanding on how to handle multilayer versions. Progress slowed on neuromorphic computing until the 1990s, when Lecun et al introduced a new architecture based on convolving images with kernels, known as Convolutional Neural Networks (CNNs) [91–93]. This architecture consists of successive layers of convolution, nonlinearity, downsampling followed by fully connected layers (see Fig. 1a). The key to the success of CNNs was that convolution handled the translation invariance of image features efficiently, while the nonlinearity between layers allowed greater flexibility in training than the single-layer approaches.

Although the CNN architecture successfully managed to implement digit classification at human performance levels and compared favorably to other machine learning techniques, it was not until improvements in processing speeds and the creation of large human-labeled image databases from the Internet, that the full potential of CNNs became apparent [94]. Using GPU-accelerated backpropagtion, AlexNet achieved record breaking results on ImageNet for a thousand categories using a CNN architecture composed of five convolutional layers and three fully connected layers [94, 95]. Following AlexNet's lead, modern CNNs of dozens or hundreds of layers, and hundreds of millions to billions of parameters, can achieve better than human level performance in image classification tasks [96, 97]. Recent breakthroughs with DeepLearning in winning the game of Go, combining Q-learning and CNNs, have convinced many that these networks are the tools for a new machine learning golden age with applications ranging from pedestrian detection for self-driving cars to biomedical image analysis [98–104].

A big part of this success story was the advent of GPU-acceleration for large matrix-matrix multiplications, which are the essential and most time intensive step of back-propagation in CNN training. Despite significant gains, training large CNNs takes weeks utilizing large clusters of GPUs. More practically, GPU-accelerated CNN inference is still a computationally intensive task, making image analysis of the vast majority of the image and video data generated by the Internet prohibitively difficult. Youtube itself, in 2015 experienced uploads of 300 hours of video every minute [105]. This would require a cluster of 18000 Nvadia Titan X GPUs to process continuously with CNNs, drawing 4.5 Megawatts, with the hardware costing 44 million US dollars [106].

Given that this is just one company and that video traffic is predicted to grow to be 80% of the Internet by 2020 [107], this problem is going to get harder and will far outpace the current computing paradigm, requiring investment in specialized neuromorphic *hardware* architectures. There are many proposals and experimental demonstrations to accomplish this through analog circuits, digital ASIC designs, FPGAs, and other electronic technologies [108–114].

Here we explore a new approach for implementing CNNs through an integrated photonics circuit consisting of Mach-Zehnder Interferometers (MZIs), optical delay lines, and optical nonlinearity units [115]. The system we outline is closely related to a recently published work which implemented fully connected optical neural networks [116]. Here we bridge the gap between the fully connected optical neural network implementation and an optical CNN through the addition of precisely designed optical delay lines connecting together successive layers of optical matrix multiplication and nonlinearity. This new platform, should be able to perform the order of a thousand inferences per second (at detector limited rates), at 10μJ power levels per inference, with the nearest state of the art ASIC competitor operating 30 times slower and requiring 1000 times more power per inference [114]. 600 of these photonic circuits, if implemented with suitable electronic I/O, would be required to process all of the Youtube uploads in real time in the above example. Although the implementation of this photonic circuit would be an extreme engineering challenge, the benefits of successfully realizing it are difficult to understate.

Our work follows a long history of optical-accelerated computing which includes optical implementations of unitary matrix multiplication, optical memory, all optical switching, optical interconnects, photonic spike processing and reservoir computing [5,116–124]. We focus primarily on integrated photonics as a computation platform because it currently provides the highest raw bandwidth currently available of any technology that is mass manufacturable and has standardized components.

5.2 Architecture

As depicted in block diagram form in Fig. 1a, and pictorially in Fig. 1b, the CNN algorithm consists of several main steps, each of which we will need to execute optically. First the image is convolved with a set of kernels. The output is a new image with dimensions $[(W-K)/S+1] \times [(W-K)/S+1] \times [d]$, where W is the original image width, K is the kernel dimension, S is the convolution stride and d is the number of kernels. Next, the new image is subject to pooling, where the image produced by convolution is further downsampled by selecting the maximum value of a set of pixels (max pooling) or taking their average (average pooling). After pooling, a nonlinearity is applied to each pixel of the downsampled image. This nonlinearity can consist of the rectified linear unit (RELU), sigmoid, tanh, or other functions. Following nonlinearity, the entire process is repeated with new sets of kernels and the same nonlinearity. After a number of these convolution layers, a fully connected neural network is applied to the output to preform the final processing steps for classification.

We begin our discussion of an optical kernel convolution, with the description of a related

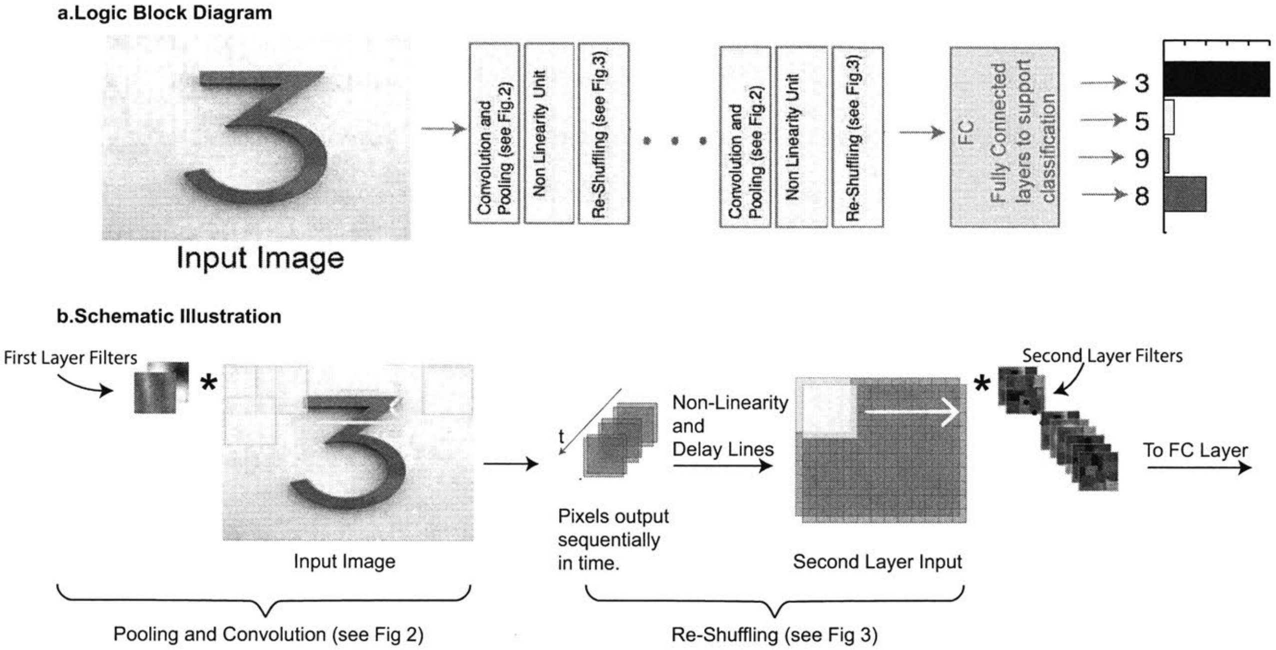


Figure 5-1: Convolutional Neural Net (CNN) Architecture. a. Logic Block Diagram: The input image, a three shown here, is passed through successive layers of convolution and pooling, nonlinearities (see Fig 2 for further description), and re-shuffling of the pixels (see Fig 3 for further description). A final fully connected layer maps the last stage of convolution output to a set of classification outputs. b Schematic Illustration: First part of CNN implements convolution of the image with a set of smaller filters. These produce a sequence of filter-kernel dot products which are passed through a nonlinearity and are re-shuffled into a new d-dimensional image, where d is the number of filters in the first layer. The process is then repeated on this new image for many subsequent layers.

GPU algorithm [125]. One of the chief advantages of utilizing GPUs for machine learning is their ability to execute large matrix-matrix multiplications rapidly. For fully connected neural networks, it is obvious how this capability can lead to large speedups. To see how this works for CNNs, we first depict the conversion of an image into a set of "patches", the same dimension as the kernels in Fig. 2a. In a GPU algorithm, these patches can be converted into a "patch matrix" by vectorizing and stacking each patch. The patch matrix can then be efficiently multiplied by a "kernel matrix", formed by vectorizing and stacking each kernel. The output is a matrix composed of kernel-patch dot products which can then be "re-patched" for multiplication by the next layer's kernel matrix.

Our recent work utilized networks of MZIs and variable loss waveguides for optical matrix multiplication for fully connected neural networks [115,116]. Here we propose using the same MZIs and variable loss waveguides to implement the kernel matrix, as depicted in Fig 2b. The patch matrix takes the form of a sequence of coherent optical pulses whose amplitude encodes the intensity of each patch pixel from an image. In turn, each output of the photonic circuit will correspond to a time series of Kernel-patch dot products with the amplitude and phase (0 or π) of each pulse encoding the output of the computation. Fig. 2b depicts this process schematically, where the photonic circuit outputs at different times have been drawn in their corresponding locations in the next layer's image.

In addition to implementing convolution with kernel matrices, the photonic circuit in Fig. 2 is also providing two other functions: pooling and nonlinearity. Pooling (or downsampling) is realized by taking the stride of the convolution to be greater than one. This step occurs when the patches are formed from the input image. This provides a necessary information filtering function required to reduce the image to only a few bits of information identifying its class. CNNs working this way have been shown to have the same performance as those using max or average pooling [126]. Finally, optical nonlinearity is applied to each output of the circuit, providing the remaining ingredient for a complete CNN layer.

The input for the next convolution layer has to be prepared by re-patching, that is reformatting the output data from the previous layer to form a patch matrix for multiplication by the next layer's kernel matrix. We propose to re-patch through a set of optical delay lines and splitters. The requirements for the splitting and delay procedure can be understood

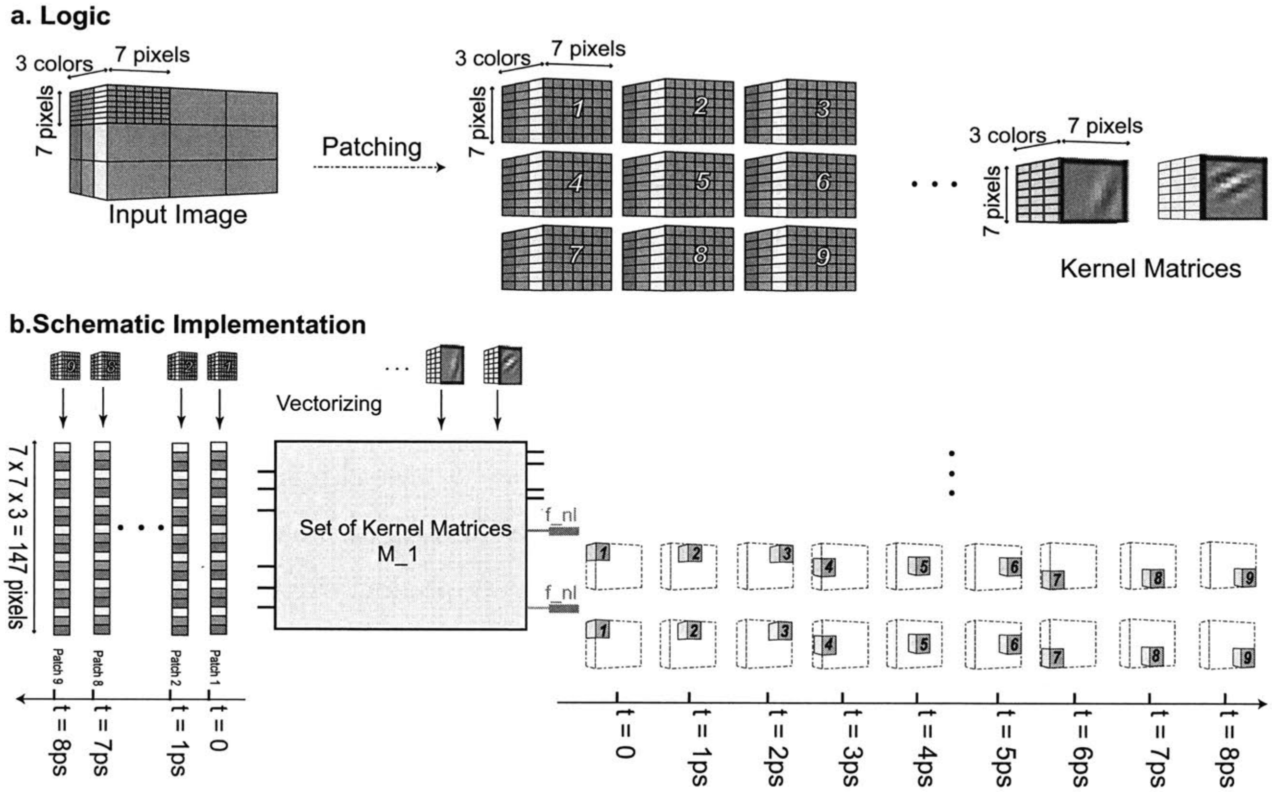


Figure 5-2: General Optical Matrix Multiplication a. Logic: The pixels of the input image on the left (21 x 21 x 3 colors) are grouped into smaller patches, which have the same dimension as the kernels of the first layer (depicted on the right-hand side). b. Schematic Implementation: Each of these patches is reshaped into a single column of data that is sequentially fed, patch by patch, into the optical interference unit. Signal propagation of the optical data column through the unit implements a dot product of the first layer kernels with the patch input vector. The result is a time series of optical signals whose amplitude is proportional to the dot products of the patches with the kernels. Each output port of the optical interference unit corresponds to a separate time series of dot products associated with a given kernel. Optical nonlinearity is applied to each output port of the optical interference unit.

from Fig. 3a. Here we depict the time sequence of kernel-patch dot products from Fig. 2b, as a single image, with each pixel of the image labeled with the timeslot associated with the computation. This image needs to be converted into four patches on the right of Fig. 3a, which will need to appear as a time sequence input for the next optical matrix multiplication.

How this is accomplished is illustrated in Fig. 3b. Here a given output from the previous layer is split into four separate waveguides and subject to different delays. Each delay line is selected such that at a given time, the outputs from the previous layer are synchronized in time and form a new patch for input into the next layer's kernel matrix. Since in this example we are forming two by two patches, a delay line of one time unit is required for the top right signal to arrive at the same time as the top left. Further delay lines of three and four time units are required for the bottom left and right signals to arrive with these. We illustrate the formation of new patches on the right hand side of Fig. 3b, where at specific times we have formed the four desired patches from Fig. 3a. Note that the grayed out section indicates a sampling time when an invalid patch is formed, that is the wrong set of pixels have arrived simultaneously. Since the original length of the patch matrix input is nine time units long, and since there are only four patches for input into the next system, there will be five such invalid sampling times in the period of the original input signal.

5.3 Physical Implementation

Figure 4 shows a full implementation of the envisioned architecture using optical matrix multiplication, optical nonlinearity and optical delay lines. To illustrate the fundamental features of the photonic integrated circuit, we have omitted the required optical amplifiers for each convolution layer, and some parts of the optical matrix multiplication.

The first part of the circuit consists of an optical interference unit. The matrix M_i encodes the kernels for a given convolution layer. From SVD decomposition we know that $M_i = U\Sigma V$, where V and U are unitary and Σ is some real diagonal matrix. U and V are implemented through Reck-encoding of the MZI matrix, and Σ through tunable waveguide loss [115, 116, 127, 128]. An incorrect realization of the unitary matrices (among many other possible errors) will degrade a network's inference capability, which we explore in the Appendix through a

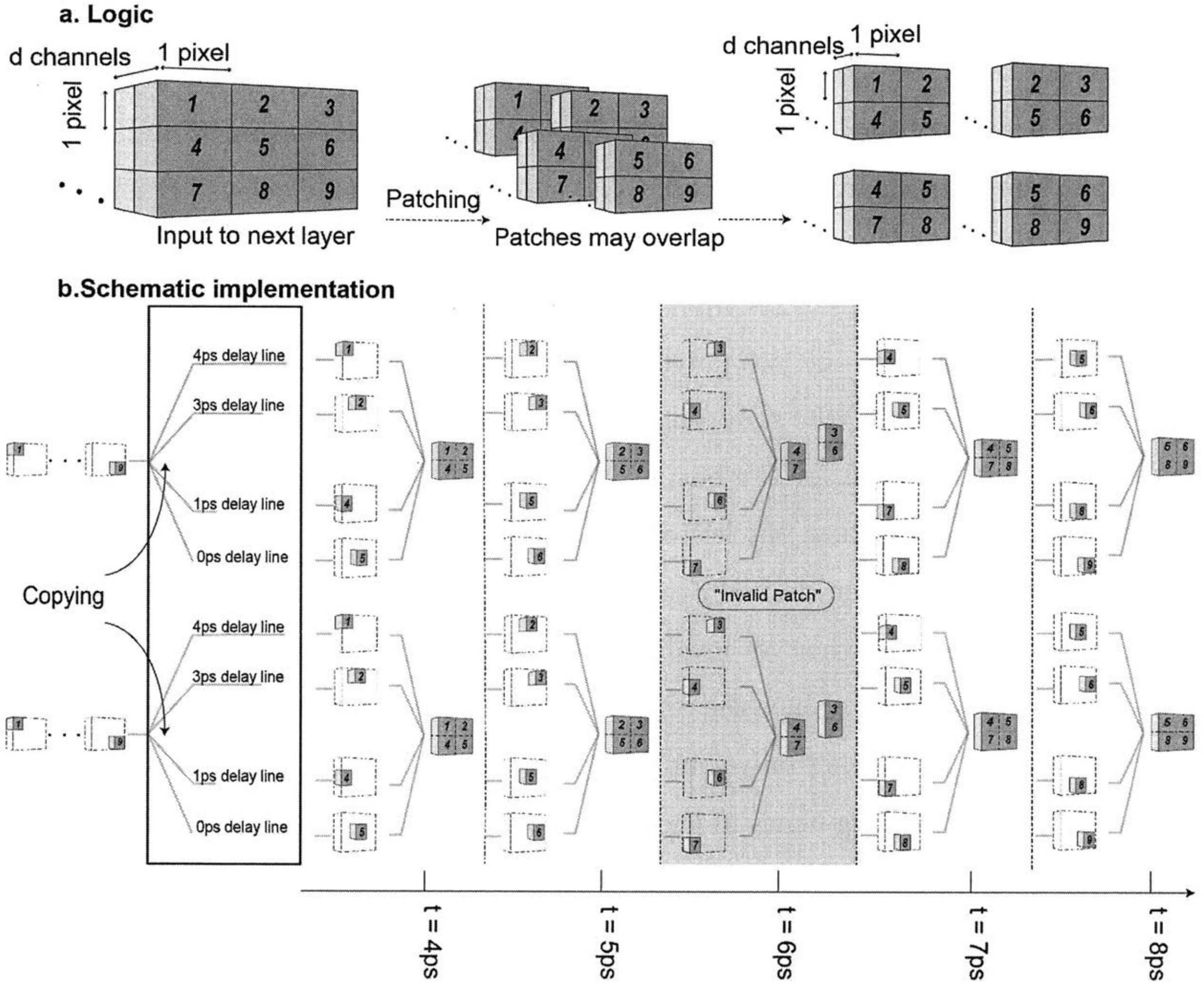


Figure 5-3: Optical Delay Lines for repatching a. Logic: The output kernel dot products from the first layer (right hand side Fig2.b) are depicted as cubes on the left in Fig3. Each cube is labeled with the timestep the corresponding kernel dot product was computed. The right hand side indicates the repatching procedure necessary to convert the set of kernel dot products on the left hand-side into input patches the same size as the next layer's kernels. b. Schematic Implementation: The optical delay lines are designed such that a sequence of kernel dot products can be reshuffled in time to form a new patch the same size as the next layer's kernels. Each delay line is connected through 3-dB splitters to the original signal line, allowing the data to be copied and then delayed for synchronization. The reshuffling procedure produces valid patches only at specific sampling times. The grayed out section at t=6ps indicates an invalid sampling interval, where the patches have partially wrapped around.

numerical example. In Fig. 4 we have omitted the optical implementations of Σ and V for simplicity.

The next stage consists of optical nonlinearity applied to each output waveguide of the MZI matrix. As suggested in our previous work, optical nonlinearity can be realized by using graphene, dye, or semiconductor saturable absorbers [129–133]. Additionally we showed that the nonlinear response of saturable absorbers has a suitable functional form for training neural networks. The power budget required for operating AlexNet, utilizing a saturable absorber nonlinearity operating at $\sim$ 1mW input powers is detailed in the Appendix. We find there that a single inference can be computed with P_{inf} power, which is given by:

$$4.17 \cdot 10^7 \Delta t P_0 = P_{inf} \tag{5.1}$$

Taking $P_0 = 1\text{mW}$, $\Delta t = \frac{1}{f} = 1$ ns, where $f \sim 1$ GHz is the throughput (from the maximum delay-line bandwidth), we find that we require 42 μJ per inference, which is over 1000 times more efficient than GPU-enabled inference with AlexNet and any other foreseeable digital implementation (see Appendix).

Finally a single convolution layer ends with repatching logic consisting of a tree of 3dB splitters feeding into variable length delay lines. The exact length of the delay lines for each layer can be determined (see Appendix), but the maximum delay line length can be estimated through the observation that the CNN algorithm will require image data at the beginning of the "patch matrix" data stream to interact with data at the end. This condition requires a delay line at least half the length of the patch matrix in time. Given that the initial layer of AlexNet consists of 55 × 55 input patches, this will require a delay line of $\sim 1000\Delta t$.

Delay lines have been engineered 1 ns long with a 3 GHz bandwidth using 200 ring resonators on 0.2mm^2 area [134–137]. Assuming we are using this technology, and $\Delta t \sim 1\text{ns}$, we will require about 1000 of these delay units for $1\mu\text{s}$ of delay. Since there are 256 outputs for the final AlexNet layer [95], we have to assume we would require at least this many maximum delay line chips with a total area of $\sim 500\text{cm}^2$. Obviously engineering and integrating delay lines of such length is an extreme engineering challenge.

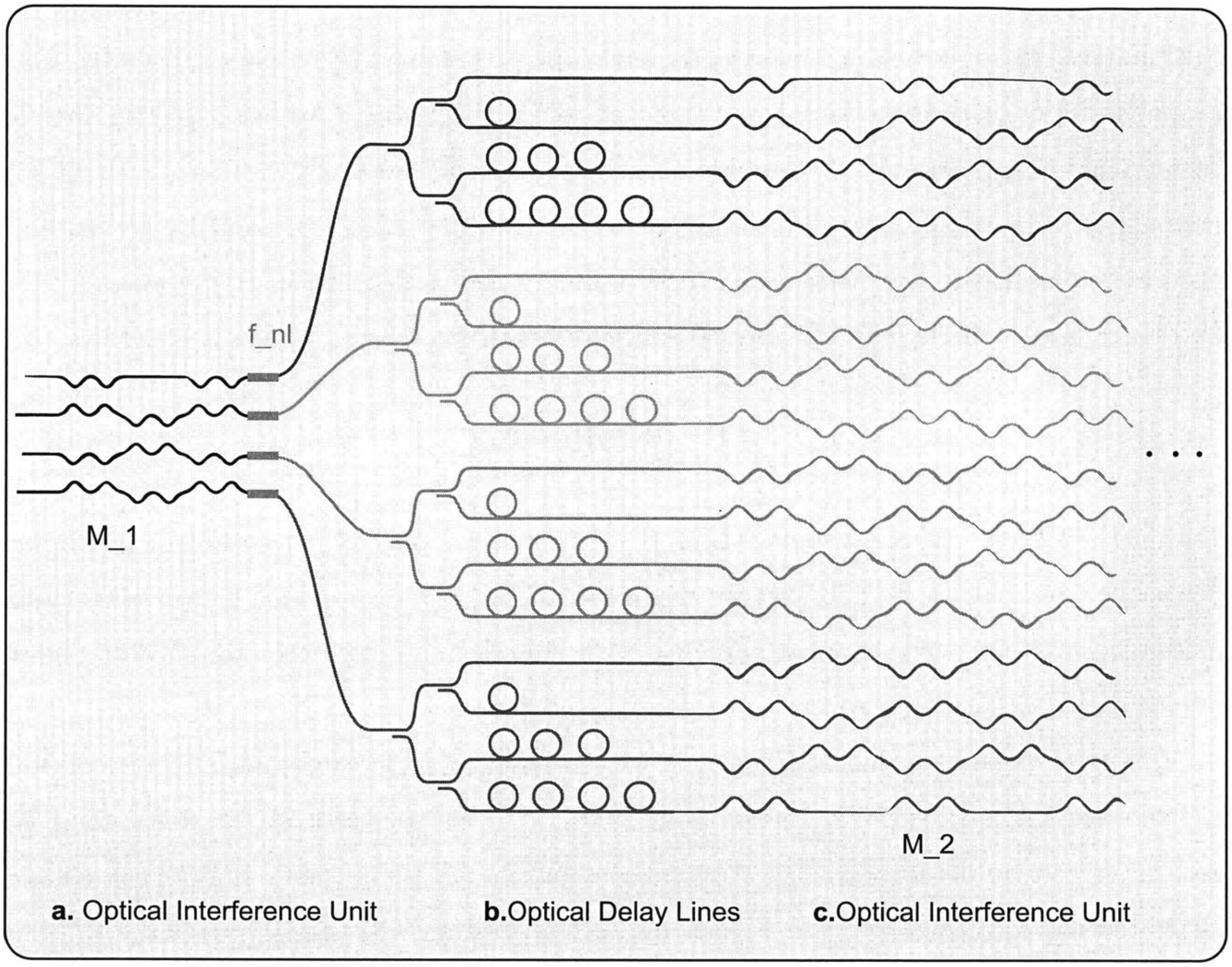


Figure 5-4: Illustration of optical interference unit with delay lines a. Optical interference unit: In the first stage an optical interference unit is used to implement a kernel matrix M_1 which processes the patches from the original image. The red segments on the output of M_1 are optical nonlinearity. b. Optical Delay Lines: In the second stage, optical delay lines properly reform the sequence of kernel dot products into new patches for input into the second kernel matrix M_2. c.Optical interference unit: In the third stage the next optical interference unit is used to implement M_2 (partially depicted here). For clarity the actual number of inputs and outputs have been reduced and the attenuator stage and subsequent additional optical interference units have been omitted from M_1 and M_2. Additionally we have omitted optical amplifiers required in each layer which boost the power sufficiently to trigger the optical nonlinearity.

5.4 Discussion

We have outlined the operation of a novel photonic integrated circuit that is capable of implementing high speed and low power inference for CNNs. This system could play a significant role in processing the thousands of terabytes of image and cat video data generated daily by the Internet.

Future work on this system and its components should focus on increasing the bandwidth. This task is critical for maximizing the inference rate of the system, shortening the delay line length, and decreasing the power consumption per inference. However every component, including the optical nonlinearity, MZIs, optical detectors, and signal generators, can potentially limit the bandwidth. Currently we are limited in bandwidth by the present state-of-the-art compact delay lines [136]. Beyond this, a significant limiting factor will be optical detectors which operate up to 120 GHz [138].

Assuming that we are able eventually to realize a 1 THz bandwidth through heroic engineering efforts, the power consumption could be reduced to nJ levels per inference, and the inference rate increased to over 10^6 frames per second. For comparison, the best digital implementation would still operate at GHz clock speeds, with a throughput of at best 10^3 frames per second, and require at minimum a mJ per inference (see Appendix). Consequently, when pushed to the physical limit, this architecture could provide a crucial capability for keeping up with the exponentially growing digital world.

5.5 Appendix

5.5.1 Analysis of MZI phase encoding and error

Although the analog nature of our optical CNN can allow for high precision computation, it also suffers significantly more than equivalent digital architectures from error propagation. There are many sources of error within our system that include variable waveguide loss, variable optical nonlinearity, variable amplification, and shot noise. We do not discuss these here, but for the purposes of providing a basic model of error propagation through the circuit, we calculate the classification error from incorrect phase settings of a Reck-encoded

MZI matrix [127,128,139].

To estimate how errors in MZI phase settings effect classification accuracy, a simulation was done on a toy model of a digit recognition CNN. Our toy CNN was comprised of two convolution layers and two fully connected layers and was trained on the MNIST digit recognition data set. After training the reference CNN (which had a classification accuracy of 97% on the test data set), the kernel matrix for each layer was exported to a phase extractor program that calculated the phase settings necessary to implement the unitary component of these matrices with MZIs (i.e. U and V from SVD decompostion of $M = U\Sigma V$).

We describe how we can calculate a phase encoding for any real valued unitary matrix. By applying a rotation matrix $T_{i,j}(\theta)$ to real unitary matrix U, with appropriate θ we can null elements in the ith row and jth column, where $T_{ij}(\theta)$, with $i = j - 1$, is an identity matrix with elements $T_{ii}, T_{ij}, T_{ji}, T_{jj}$ replaced by a two by two rotation matrix:

$$U' = \begin{pmatrix} 1 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & & & & \vdots \\ \vdots & & \cos(\theta_{j-1,j}) & \sin(\theta_{j-1,j}) & & \vdots \\ \vdots & & -\sin(\theta_{j-1,j}) & \cos(\theta_{j-1,j}) & & \vdots \\ \vdots & & & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & 1 \end{pmatrix} \begin{pmatrix} u_{1,1} & \cdots & \cdots & \cdots & \cdots & u_{1,j-1} & 0 & \cdots & 0 \\ & \ddots & & & & & \vdots & & \vdots \\ & & \ddots & & & & 0 & & \vdots \\ & & & \ddots & & & u_{i,j} & & \vdots \\ & & & & \ddots & & u_{i+1,j} & & \vdots \\ & & & & & \ddots & \vdots & & \vdots \\ & & & & & & d_k & & \vdots \\ & & & & & & & \ddots & 0 \\ & & & & & & & & d_N \end{pmatrix} \tag{5.2}$$

To determine θ in $T_{i,j}(\theta)$ such that the ijth element of the matrix is nulled, θ must satisfy the following equation:

$$\begin{aligned} u'_{i,j} = \cos(\theta_{j-1,j})u_{i,j} + \sin(\theta_{j-1,j})u_{i+1,j} &= 0 \\ \tan(\theta_{j-1,j}) &= \frac{-u_{i,j}}{u_{i+1,j}} \end{aligned} \tag{5.3}$$

If we apply these rotations starting with the first element of the far-right column and working downwards we find that:

$$T_{N,N-1}T_{N,N-2} \cdots T_{N,2}T_{N,1}U(N) = \left(\begin{array}{c|c} U(N-1) & 0 \\ & \vdots \\ \hline 0 \cdots & a \end{array}\right) \tag{5.4}$$

Now we note that the block matrix $U(N-1)$ can undergo the same process, and thus after $(N-1)+(N-2)\cdots=\frac{N(N-1)}{2}$ rotations the right hand side will turn into a diagonal real matrix D (this is only true if U is real, which it is for conventional CNNs). In total U can be written in the following way [127] [128]:

$$U=(T_{2,1}T_{3,2}T_{3,1}T_{4,3}\cdots T_{N,N-1}\cdots T_{N,1})^{-1}D \tag{5.5}$$

We can extract a matrix of phase encodings Θ with the following pseudocode:

```
function Phase_Extractor

for i from N to 2 :

    for j from 1 to i :
        $\Theta_{j,j+1} = \tan^{-1}(\frac{-U_{j,i}}{U_{j+1,i}})$
        Update $U = T_{j,j+1}(\Theta_{j,j+1})U$
    end for

end for

return $\Theta$
```

The algorithm nulls the elements of a given unitary matrix starting from $U_{1,N}$ and moving downward until reaching a diagonal element, upon which it moves to the next column to the left. The below schematic shows the order in which the algorithm nulls the elements and extracts the corresponding phases:

$$\begin{array}{c} \xleftarrow{\text{Higher Iterations}} \\ \begin{pmatrix} d_1 & u_{1,2}\big\downarrow & \cdots & u_{1,N-1}\big\downarrow & u_{1,N}\big\downarrow \\ & d_2 & \cdots & u_{2,N-1} & u_{2,N} \\ & & \ddots & & \\ & & & u_{N-2,N-1}\big\downarrow & \\ & & & d_{N-1} & u_{N-1,N}\big\downarrow \\ & & & & d_N \end{pmatrix} \end{array}$$

Each element of the resulting phase matrix Θ is randomly perturbed with a distribution given by $p(\Delta\Theta_{i,j}) = \frac{1}{\sqrt{2\pi\sigma}}\exp\Big\{\Big[-\frac{\Delta\Theta_{i,j}^2}{2\sigma^2}\Big]\Big\}$. This perturbed matrix is used to create a new U' (in $M = U'\Sigma V'$), which is in turn used to generate the kernel matrices of a perturbed CNN. The results of this perturbed network's digit classification on MNIST are then compared to those of the trained unperturbed CNN such that the network is assumed to be "error free" if it gives the same results as the reference network, not if it has 100% correct classification. These results are plotted in Fig. 5.

We find that for an error distribution with $\sigma > 0.01$, that the performance of the perturbed network is significantly degraded relative to the unperturbed version. This corresponds to about 8-bit accuracy in the phase settings, which has been achieved in our previous work on fully connected optical neural networks. These results are promising, but larger CNNs need to be examined with this method to assess their tolerance to the phase setting errors, and the others mentioned briefly above.

5.5.2 Power consumption for optical implementation of AlexNet

Since we are feeding each patch at a rate of $f \sim 1\text{GHz}$ each optical nonlinearity unit consumes roughly $E = 10\frac{P_0}{f} \approx 10^{-11}\text{J/waveguide/patch}$, where the factor of 10 comes from the amplifier efficiency, and $P_0 = 1\text{mW}$ is the power required to operate the optical nonlinearity. Consequently if we sum the number of kernel output waveguides multiplied by the number of patches in each layer, and scale them by this power factor, we will get an estimate for the power requirements for this optical CNN implementation:

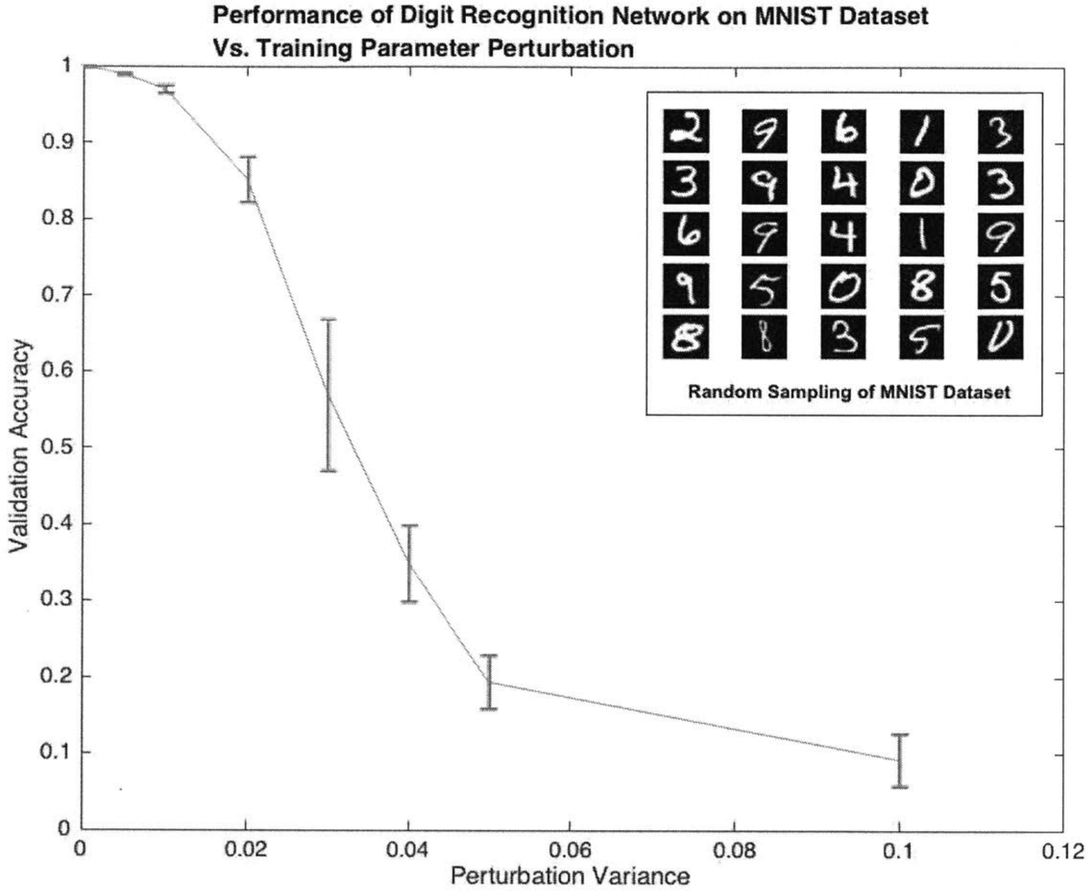


Figure 5-5: After the matrix of phases, Θ, has been returned by the phase extractor algorithm, the entries are perturbed and the resulting phase matrix is used in composing a new unitary matrix of weights that builds up a second CNN with desired perturbations. The perturbed CNN is tested and its inference performance on the MNIST dataset is then compared to those of the unperturbed CNN to analyze the error.

	# of RELU units	# of Input Patches	Layer Energy Consumption(J)
1st Conv	96	55×55	290400×10^{-11}
2nd Conv	256	55×55	774400×10^{-11}
3rd Cnv	384	55×55	1161600×10^{-11}
4th Conv	384	55×55	1161600×10^{-11}
5th Conv	256	55×55	774400×10^{-11}
1st FC	4096	1*	4096×10^{-11}
2nd FC	4096	1	4096×10^{-11}
3rd FC	1000	1	1000×10^{-11}
Total Energy Consumption			$4,171,592 \times 10^{-11}$

Which yields a total energy consumption per inference is 41 μJ for AlexNet [95]. We can further rewrite this as $4.17 \cdot 10^7 \Delta t P_0$.

The number of input patches decreases to one for the fully connected layers because we employ electro-optic converters to extract one valid patch from the convolutional layer output. In principle this technique could be employed to extract and re-emit only valid patches from previous layers, but we do not discuss this here, because we are primarily interested in an all-optical implementation.

5.5.3 Power consumption for electronic implementation of AlexNet

Electronic computers consume a fixed amount of energy per floating point operation. Since data movement (i.e. data transfer between hard drive and RAM, etc.) represents additional significant "overhead" for memory intensive algorithms like CNNs, and minimizing this is the chief objective of new neuromorphic digital architectures, calculating the total number of floating point operations required for AlexNet gives us a good estimation of the best case performance for a digital implementation of that algorithm:

	Kernel Size	Number of Input Patches	Number of Kernals	Layer FLOPs
1st Conv	$11 \times 11 \times 3$	55×55	96	105415200×2
2nd Conv	$5 \times 5 \times 96$	27×27	256	447897600×2
3rd Conv	$3 \times 3 \times 256$	13×13	384	149520384×2
4th Conv	$3 \times 3 \times 384$	13×13	384	224280576×2
5th Conv	$3 \times 3 \times 384$	13×13	256	149520384×2
1st FC	$13 \times 13 \times 256$	1×1	4096	177209344×2
2nd FC	$1 \times 1 \times 4096$	1×1	4096	16777216×2
3rd FC	$1 \times 1 \times 4096$	1×1	1000	4096000×2
Total FLOPs				$2,549,433,408$

Electronic computers have an average performance rate of $1\frac{\text{pJ}}{\text{FLOP}}$ [114], so the lower bound on the total energy consumed by a digital computer (ASIC, GPU, CPU) running AlexNet is $2,549,433,408$ FLOPs $\times\ 1\frac{\text{pJ}}{\text{FLOP}} \approx 2.55\text{mJ}$ [95].

Chapter 6

Binary matrices of optimal autocorrelations as alignment marks

6.1 Abstract

We define a new class of binary matrices by maximizing the peak-sidelobe distances in the aperiodic autocorrelations. These matrices can be used as robust position marks for in-plane translational spatial alignment. The optimal square matrices of dimensions up to 7 by 7 and optimal diagonally-symmetric matrices of 8 by 8 and 9 by 9 were found by exhaustive searches. This work is reported in [141].

6.2 Introduction

Binary sequences [142, 143] and matrices with good autocorrelation properties have key applications in digital communications (radar, sonar, CDMA and cryptography) [144] and in coded aperture imaging [145]. Several works have conducted exhaustive searches for the optimal matrices of these applications [146–149]. A less developed application of binary matrices with good aperiodic autocorrelations is two-dimensional (2D) translational spatial alignment. For example, it has been shown in electron-beam lithography [150–153] that position marks based on such binary matrices are immune to noise and manufacturing errors. However, the symbols that were used in these prior works were borrowed from different applications, noticablely the 1D Barker sequences of ±1 from communications. There have been no studies on the optimal patterns for translational alignment.

In this paper, we define and report the optimal binary matrices as alignment marks. Section 6.3 sets up the problem. Section 6.4 defines the criteria for the optimal matrices. Section 6.5 discusses previous work related to this problem. Section 6.6 works out the useful bounds. Section 6.7 explains the exhaustive computer searches and lists the results. Section 6.8 discusses several key observations of the optimal marks. Section 6.9 compares the performance of optimal and non-optimal marks through simulations. Section 6.10 discusses the potential applications of the matrices found. Section 6.11 concludes the paper.

6.3 Preliminaries

An alignment mark is made by creating a surface pattern different from the background so that the pattern information transforms into a two-level signal when a digital image is taken. This image can be represented as a binary matrix where 1 represents the (black) pattern pixels and 0 represents the (white) background pixels or vice versa.

The 2D aperiodic autocorrelation (A) of an M by N binary matrix with elements $R_{i,j}$ is defined as

$$A(\tau_1, \tau_2) = \sum_{i=1}^{M} \sum_{j=1}^{N} R_{i,j} R_{i+\tau_1, j+\tau_2} \tag{6.1}$$

where τ_1, τ_2 are integer shifts. The peak value is $A(0,0)$ while all other values are sidelobes. A is an inversion-symmetric $[A(\tau_1, \tau_2) = A(-\tau_1, -\tau_2)]$ $(2M-1)$ by $(2N-1)$ matrix. The crosscorrelation between R and the data image matrix $D_{i,j}$ is expressed as

$$C(\tau_1, \tau_2) = \sum_{i=1}^{M} \sum_{j=1}^{N} R_{i,j} D_{i+\tau_1, j+\tau_2}. \tag{6.2}$$

When the data D is a noisy version of the reference R, the peak value of the crosscorrelation determines the most probable position of the mark.

It is important to note that all the matrices are implicitly padded with 0s for all the matrix elements of indices exceeding their matrix dimensions.

A linear transformation of the data matrix results in a linear transformation of the correlation as long as the reference matrix is kept the same. This can be seen from

$$D'_{i,j} = cD_{i,j} + d \tag{6.3}$$

$$C'(\tau_1, \tau_2) = cC(\tau_1, \tau_2) + d\sum_{i=1}^{M}\sum_{j=1}^{N} R_{i,j} \tag{6.4}$$

where the second term of C' is a constant. The data matrix can thus be arbitrarily scaled ($c \neq 0$) while keeping the correlation equivalent and the alignment results identical.

6.4 Criteria for the optimal binary matrices

Depending on the quantities being optimized, the criteria for the optimal matrices are different. For alignment purposes, we list two criteria here. The first is to minimize the misalignment probability. The second is to minimize the misalignment deviation. The first criteria depends on the values of the autocorrelation sidelobes, while the second criteria also depends on their positions relative to the central peak.

In this paper, we chose to minimize the probability that misalignment happens. A misalignment occurs when one of the sidelobes exceeds the central peak [$p = A(0,0)$]; this probability is analytically expressed in Appendix 6.12.1. Under the same noise condition, the less the peak-sidelobe distance the higher the misalignment probability. Consequently the criteria for ranking the matrices is based on their peak-sidelobe distances.

The peak-sidelobe distances are illustrated in Fig. 6-1. We plotted an autocorrelation matrix $A(\tau)$ with peak value p and highest sidelobe value s. The shortest peak-sidelobe distance is denoted as d_1, where $d_1 = p - s$. The other distances are defined as $d_{i+1} = d_i + 1$ for $i \geq 1$, as shown for d_1 through d_4 in Fig. 6-1. n_i gives the number of times d_i occurs in the autocorrelation and $\sum\limits_i n_i = (2M-1)(2N-1) - 1$. The histogram of an autocorrelation matrix can be expressed as $\{d_1 | n_1, n_2, ..., n_{(s+1)}\}$.

The criteria for finding the optimal matrix is to maximize d_1 then minimize n_i sequentially in the dictionary order. This criteria is completely justified in the low noise limit in Appendix 6.12.1 , although a general criteria depends on the amount of noise in the data matrices.

Matrices of any size can be compared using this criteria. In general, the distances (d_i) of the autocorrelation increase with the size of the matrix. Without restricting the matrix dimension, the optimal matrix will diverge in size. Consequently, we study the optimal matrix for each *fixed dimension.* Interestingly, the optimal matrices found in this paper are unique as discussed in Sec. 6.8.

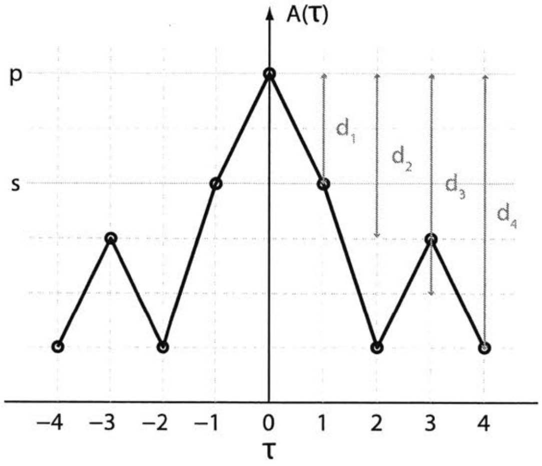


Figure 6-1: (Color Online) We illustrate an autocorrelation function $A(\tau)$, whose peak value is p, highest sidelobe value is s, and whose peak-sidelobe distances are d_i.

6.5 Related Work

Previous works on 1 and -1 matrices with 0 background [146, 149] in digital communications are different than our work on 1 and 0 matrices. The former representation has three levels (1,-1,0) while our binary matrices have only two levels. The aperiodic autocorrelations of these matrices are not equivalent.

Other works on binary matrices of 1s and 0s with aperiodic autocorrelations have used different criterias selected for applications in radar and sonar. In the Costas-array problem [147], only one black pixel is placed per column and row and the maximum sidelobe is fixed to one. In the Golomb-Rectangle problem [148], the number of black pixels is maximized with the restriction that the sidelobe still be fixed to one [154]. However, our criteria does bear some resemblance to those in some of the works on one dimensional -1 and 1 (three levels) sequences [143].

6.6 Two Upper bounds of $d_{1,\max}(p)$, $d_{1,\max}^{\text{upper,I}}(p)$ and $d_{1,\max}^{\text{upper,II}}(p)$

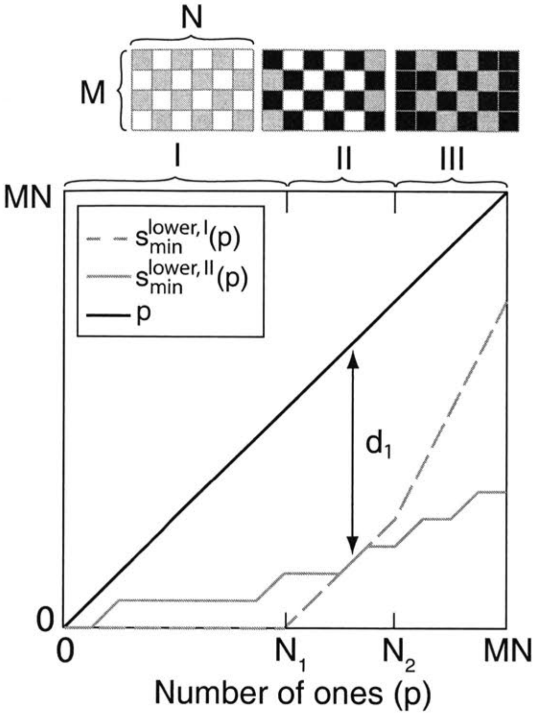


Figure 6-2: (Color Online) Lowerbounds of $s_{\min}(p)$, $s_{\min}^{\text{lower,I}}(p)$ and $s_{\min}^{\text{lower,II}}(p)$. p is the autocorrelation peak. The three matrices on top illustrate the methods of filling black pixels for regions I, II and III for the matrix construction of $s_{\min}^{\text{lower,I}}(p)$. The grey pixels show spots to be filled in that region, while the black pixels are spots that have been filled in previous regions.

For a binary matrix R, the peak value p of its autocorrelation A equals the number of ones in the matrix (R). The largest d_1 for all matrices with a given p, of a fixed dimension, is $d_{1,\max}(p)$. $d_{1,\max}(p) = p - s_{\min}(p)$, where $s_{\min}(p)$ is the minimum highest sidelobe value as a function of p.

In this section, we constructed an upperbound of $d_{1,\max}(p)$, $d_{1,\max}^{\text{upper,I}}(p)$, by maximizing $p - A(\pm 1, 0)$. The $A(\pm 1, 0)$ computed here forms a lower bound on $s_{\min}(p)$, $s_{\min}^{\text{lower,I}}(p)$. This construction is illustrated in Fig. 6-2, where we assume the matrix R used to construct our bound is of dimension $M \times N$ with $M \leq N$.

We find:

$$d_{1,\max}^{\text{upper,I}}(p) = \begin{cases} p, & p \in [0, N_1] \quad \text{I} \\ N_1, & p \in [N_1, N_2] \quad \text{II} \\ M(N+1) - p, & p \in [N_2, MN] \quad \text{III} \end{cases} \tag{6.5}$$

where $N_1 = \frac{MN}{2}, N_2 = \frac{MN}{2} + M$ when MN is even and $N_1 = \frac{MN+1}{2}, N_2 = \frac{MN+1}{2} + M - 1$ when MN is odd.

This upperbound can be derived by starting out with a matrix $R_{i,j} = 0$ for all (i, j) and 'filling in' with ones in a particular pattern. In region I, ones can be placed anywhere in $R_{i,j}$ where $i + j$ is odd. When $p = N_1$, we have formed a "checkerboard pattern". In region II, we place ones wherever $i+j$ is even for $i = 1$ or $i = N$. In region III, the remaining locations without ones are filled.

The autocorrelation function $A(\tau_1, \tau_2)$ equals the number of black squares that are connected by a displacement vector (τ_1, τ_2). We can use this property to construct a second lower bound $s_{\min}^{\text{lower,II}}(p)$. This approach is similiar to the method used in Ref. [154].

Since the autocorrelation is invariant under inversion, there are $((2M-1)(2N-1)-1)/2 = 2NM-N-M$ unique non-zero displacements; a matrix of p ones fills $p(p-1)/2$ of them. As p increases, there are repeated displacements because $p(p-1)/2$ quickly exceeds $2NM-N-M$.

We can find a lowerbound $s_{\min}^{\text{lower,II}}(p)$ by assuming that the displacements added to the autocorrelation function distribute uniformly, that is $|A(\tau_1, \tau_2) - A(\tau_2', \tau_2')| \leq 1$ for nonzero displacements. This gives $s_{\min}^{\text{lower,II}}(p) = ceil[\frac{p(p-1)}{4NM-2N-2M}]$, where $ceil[x]$ rounds to the nearest integer greater than x. Consequently, $d_{1,\max}^{\text{upper,II}} = p - ceil[\frac{p(p-1)}{4NM-2N-2M)}]$.

As illustrated in Fig. 6-2, $s_{\min}^{\text{lower,II}}(p)$ is a better bound for small p, while $s_{\min}^{\text{lower,I}}(p)$ is a better bound for large p. The first bound $s_{\min}^{\text{lower,I}}(p)$, which keeps track of the pixel positions, becomes exact when p approaches "MN" (filled) . While the second bound $s_{\min}^{\text{lower,II}}(p)$, which ignores the actual pixel locations, becomes exact when the matrix is sparse and p approaches "0" (empty).

6.7 Exhaustive computer searches for the optimal square matrices

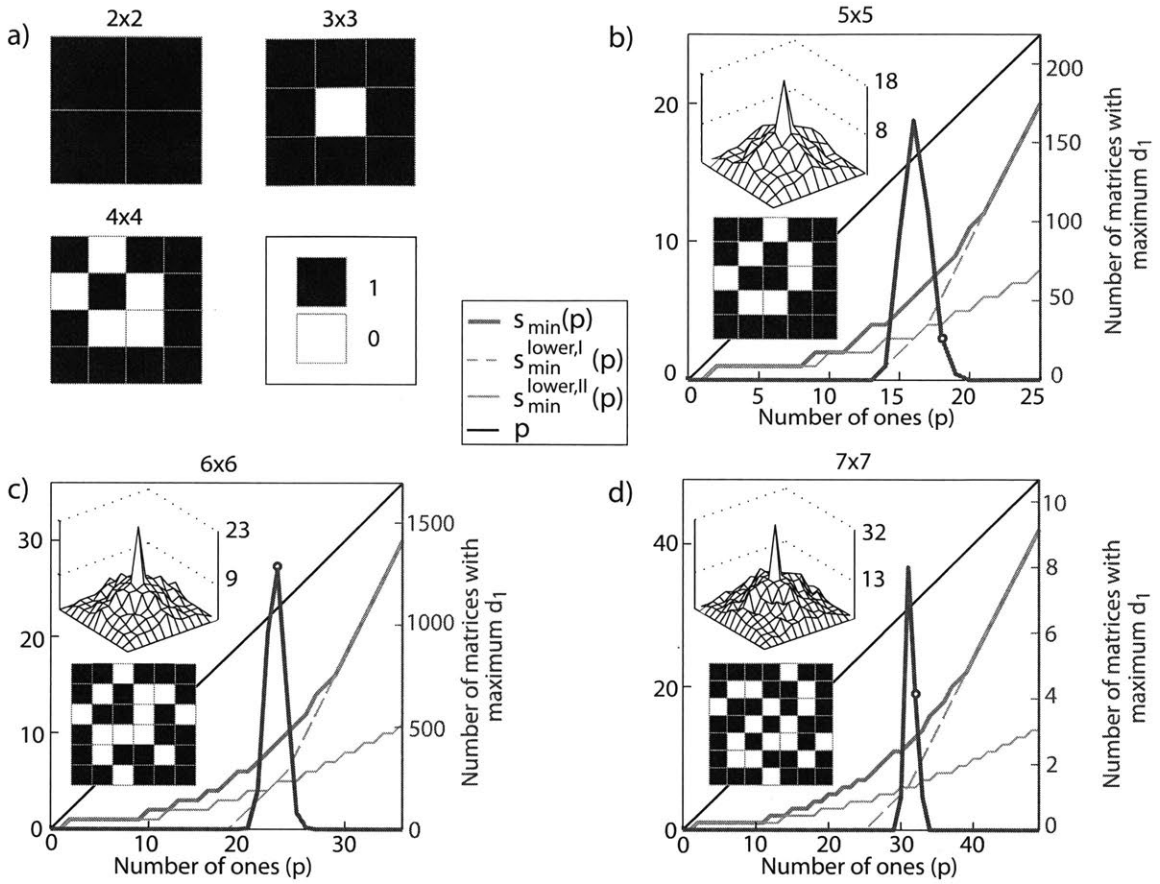


Figure 6-3: (Color Online) Results of the exhaustive searches for 2 by 2 to 7 by 7 matrices. a) The optimal matrices from 2 by 2 to 4 by 4 are shown. b), c) and d) $s_{\min}(p)$ is plotted in red. The solid grey line is $s_{\min}^{\text{lower,I}}(p)$ while the dotted grey line is $s_{\min}^{\text{lower,II}}(p)$. The number of the matrices having the maximum d_1 are plotted in blue. The circle specifies the location of the optimal matrix. The optimal matrices are presented as insets below their autocorrelations, which are labeled with their p and s values.

Physical in-plane alignment usually requires equal alignment accuracies in both directions; this calls for square matrices ($M = N$). We applied exhaustive searches to find the square matrices with the maximum $d_1[= max(d_{1,\max}(p))]$. The resulting matrices were ranked using the criteria in Sec. 6.4 to obtain the optimal matrices.

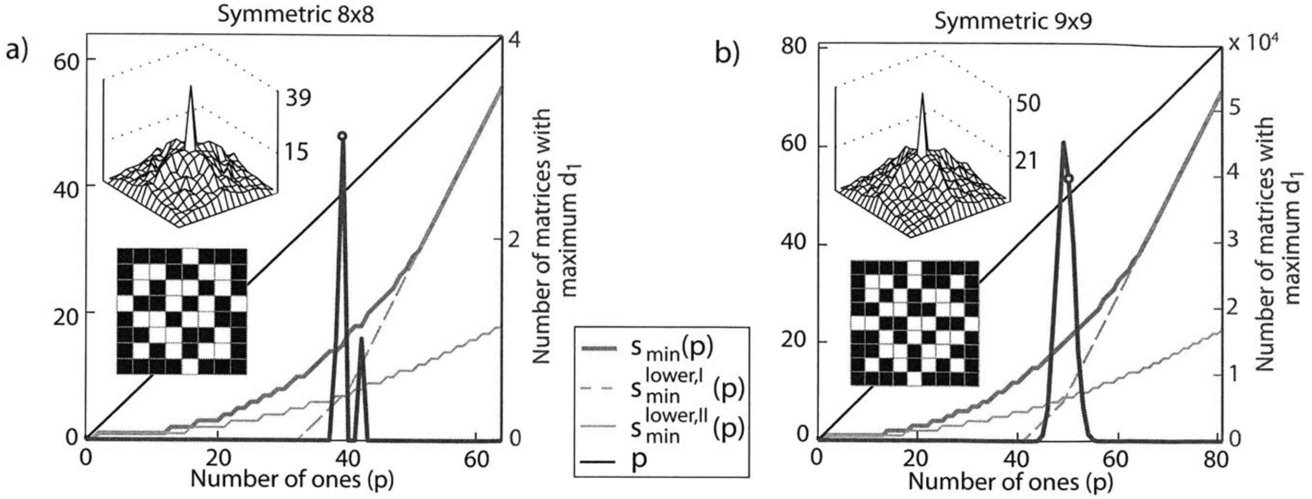


Figure 6-4: (Color Online) Results of the exhaustive searches for diagonally-symmetric 8 by 8 and 9 by 9 matrices.

Backtrack conditions based on symmetries and sidelobes have been found useful in exhaustive searches for binary matrices [146, 154, 155]. Matrices related by symmetry operations are considered the same matrix. The symmetry operations for square matrices are horizontal and vertical flips and rotations by multiples of 90 degrees. For this study, a backtrack condition based on eliminating redundant matrices related by horizontal flips was implemented. Backtrack conditions based on sidelobe levels are useful if the sidelobes are being minimized. However, we are maximizing the peak-sidelobe distance d_1, so the sidelobe backtrack condition was not used.

The search algorithm we implemented works by exhaustively generating matrices row by row. The algorithm continues generating rows until a backtrack condition occurs, or a matrix is completely specified. The matrix is stored for later ranking if it has the same or greater d_1 than the existing maximum d_1.

Several techniques were implemented to speed up the algorithm. Each matrix row was represented as a binary word so that fast bit-wise operations could be used. In addition lookup tables were created to calculate the horizontal flips and correlations of rows. For our binary matrices, the maximum sidelobes were typically located near the autocorrelation peak. Because of this, the sidelobe values were checked in a spiral pattern around the peak to quickly determine if a matrix had a d_1 less than the stored maximum.

The search results for square matrices of size up to 7 by 7 are presented in Fig. 6-3.

Fig. 6-3a) gives the optimal matrices for 2 by 2, 3 by 3 and 4 by 4. In Fig. 6-3b), c) and d) we plot, in red, $s_{\text{min}}(p)$ for matrices of sizes 5 by 5, 6 by 6 and 7 by 7. This red curve is indeed bounded from below by the grey $s_{\text{min}}^{\text{lower,I}}(p)$ and $s_{\text{min}}^{\text{lower,II}}(p)$ constructed in Sec. 6.6. The number of the matrices having the maximum d_1 is plotted in blue. This curve peaks around the intersection of the $d_{1,\text{max}}^{\text{upper,I}}$ and $d_{1,\text{max}}^{\text{upper,II}}$ upperbounds. The circle on the blue line specifies the location of the optimal matrix ranked first by the criteria in Sec. 6.4. The optimal matrices and their autocorrelations are shown as insets. The two numbers on the y-axes of the autocorrelation plots are the p and s values of the optimal matrices. The matrices ranked second and third and their distance spectra are listed in Appendix 6.12.2.

The runtime for 7 by 7 matrices was 3 hours on 1000 Intel EM64T Nodes with 2.6 GHz clock speed. Exhaustive searches of square matrices of size 8 by 8 are not accessible to us, since the size of the search space increases exponentially with the number of matrix elements as 2^{N^2}.

6.8 Observations on the optimal square matrices

The first interesting observation is that most top-ranked matrices in Fig. 6-3 and 6.12.2 are *diagonally symmetric.* Because of this if we restrict our searches to symmetric matrices of larger sizes, we still expect to find top-ranked matrices [155]. The search results for diagonally-symmetric matrices of 8 by 8 and 9 by 9 are presented in Fig. 6-4.

The second observation for our optimal matrices shown in Fig. 6-3, is that d_1 always occurred in the first four neighbors of the autocorrelation peak $[A(0, \pm 1), A(\pm 1, 0)]$. Since d_1 is the most likely point for misalignment, these matrices, although optimized for misalignment probability, also have low misalignment deviation discussed in Sec. 6.4. Another interesting property of the autocorrelation is that the ratio of $\frac{A(0,0)-A(\pm 1,0)}{N}$ or $\frac{A(0,0)-A(0,\pm 1)}{N}$ is invariant under symbol expansion (i.e. expanding the number of pixels making up the original marker pixel). This property allows us to define a new quantity for the optimal matrices in this work called *sharpness* $\Lambda = \frac{d_1}{N}$. Since Λ is scale-invariant, d_1 can be easily obtained for different scaling factors and used to evaluate the alignment performance. The sharpness (Λ) of the optimal matrices increases with the size of the matrices.

The third observation is that all of the optimal matrices shown in Figs. 6-3 and 6-4 are connected through their black pixel (1s) and all but 3 by 3 are connected through their white pixels. A pixel is connected if one or more of its eight neighboring pixels has the same value. *Connectedness* is a preferred topological property for alignment marks; it makes the marks self-supportive, suspendible and robust against mechanical disturbances.

The fourth observation is that the optimal matrices found in Figs. 6-3 and 6-4 are *unique*; there is only one matrix with the optimal histogram ranked by the criteria from Section 6.4 . In general, the mapping from histograms to correlations is not unique. For example the 2 by 2 matrices of $\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ have identical histograms. It is unclear whether this property holds for optimal matrices of all sizes.

6.9 Alignment accuracies of the optimal matrices

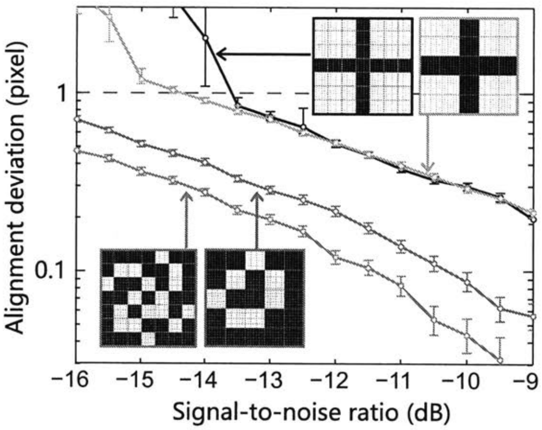


Figure 6-5: (Color Online) The "horizontal" alignment deviation is shown for the four alignment marks under various signal-to-noise ratios. The vertical deviation is almost identical. The color of each plot line borders the corresponding marker. All markers have been expanded to 35 by 35 pixels to illustrate the idea of pixel expansion. The top, black line, on the right edge, corresponds to the 7 by 7 cross, while the second to top, grey line corresponds to the 5 by 5 cross. The second to bottom, blue line corresponds to the optimal 5 by 5 marker, while the bottom, red line corresponds to the optimal 7 by 7 matrix.

We study the performance of the optimal matrices by comparing the optimal alignment marks to the cross patterns. The matrices were embedded in a white "0" background with a size 5 times that of the symbol. Uniform Gaussian noise was added to all pixels to simulate a noisey image. This was correlated with its noise-free version. The alignment accuracy was determined by the deviation of the correlation peak from the center for 10000 trials.

In Fig. 6-5, we plot the alignment deviation as a function of signal-to-noise ratio for two optimal marks from Fig. 6-3 and the crosses. The y-axis is the horizontal alignment deviation in pixels while the x-axis is the signal-to-noise ratio in decibels ($= 20log\frac{S}{N}$). At a signal-to-noise ratio of 0 dB, the markers are barely discernible by eye. All markers were expanded to the same area, of 35 by 35 total pixels and embedded in a background of 175 by 175 pixels, for direct comparison.

Applying the criteria from Section 6.4, using the expanded 35 by 35 symbols, the 7 by 7 mark is ranked first, followed by the 5 by 5 mark, and then the crosses. The quality of the optimal alignment marks should improve with increasing size, which provides a motivation to continue the search for larger optimal matrices.

6.10 Applications

Correlation detection from a digital image is a simple, efficient and reliable way to determine the position of an alignment mark. In practice, the crosscorrelations can be calculated by fast-Fourier-transforms. The peak of the correlation can further be interpolated to obtain an alignment accuracy better than the distance represented by a single pixel of the image [150]. The matrices reported in this paper are the desirable patterns to use in this context; they can replace the cross-type patterns widely in use today as position markers. Alignment using these matrices is very robust against noise in the imaging system and partial damage of the mark, providing the strongest peak signal for accurate sub-pixel interpolation. The potential applications of the matrices found in this paper include, but are not limited to, electron-beam lithography [151], planar alignment in manufacturing [156], synchronization [157] and digital watermarking [158].

6.11 Conclusions

We introduced a new class of binary matrices (two level signals) which have maximial peak-to-sidelobe distances in their aperiodic autocorrelation. Optimal square matrices of dimensions up to 7 by 7 and optimal diagonally-symmetric matrices of 8 by 8 and 9 by 9 were found using a backtrack algorithm. Useful bounds, notable properties and the performances of the optimal matrices were discussed.

6.12 Appendix

6.12.1 Probability of misalignment

The crosscorrelation between the data image and the reference matrix is denoted as $C(\tau_1, \tau_2)$. The autocorrelation of the binary reference matrix is denoted as $A(\tau_1, \tau_2)$. The data image is essentially a copy of the reference matrix with noise added to it. We assume the noise is Gaussian and the standard deviation for each pixel is σ. The "black" and "white" pixel values of the data image are denoted as b_i and w_i, whose expectation values are $\overline{b_i} = 1$, $\overline{w_i} = 0$ and $\overline{C} = A$.

Misalignment happens if $C(0,0) - C(\tau_1, \tau_2) = x_{\tau_1,\tau_2} \leq 0$, representing a sidelobe $[C(\tau_1, \tau_2)]$ exceeding the central peak $[(C(0,0))]$ in the crosscorrelation. Below we write this inequality in detail,

$$\begin{gathered} x_{\tau_1,\tau_2} = C(0,0) - C(\tau_1,\tau_2) = \\ \sum_{i=1}^{p} b_i - \left\{ \sum_{i=1}^{p-d_{\tau_1,\tau_2}} b_i^{(\tau_1,\tau_2)} + \sum_{i=1}^{d_{\tau_1,\tau_2}} w_i^{(\tau_1,\tau_2)} \right\} \leq 0 \\ p = A(0,0), d_{\tau_1,\tau_2} = A(0,0) - A(\tau_1,\tau_2) > 0 \end{gathered} \tag{6.6}$$

The first term in the inequality represents $C(0,0)$, where each element of the reference matrix with value 1 multiplies the corresponding b_i. The sum includes all p pixels of b_i. The two terms in the brackets represent $C(\tau_1, \tau_2)$, when the reference and data matrices are offset by (τ_1, τ_2). $b_i^{(\tau_1,\tau_2)}$ is a subset of b_i which multiply elements of value 1 in the reference matrix. $w_i^{(\tau_1,\tau_2)}$ is a subset of w_i which multiply the remaining elements of value 1 in the reference matrix.

x_{τ_1,τ_2} is a sum of Gaussian variables and so is also a Gaussian variable with an expectation value $\overline{x_{\tau_1,\tau_2}} = d_{\tau_1,\tau_2}$. By bookkeeping the terms in Eq. 6.6, one finds the standard deviation $\sigma^2_{x_{\tau_1,\tau_2}} = 2d_{\tau_1,\tau_2}\sigma^2$.

The probability of misalignment due to the sidelobe at (τ_1, τ_2) is $M(x_{\tau_1,\tau_2} \leq 0)$.

$$\begin{aligned} &M(x_{\tau_1,\tau_2} \leq 0 \mid \overline{x_{\tau_1,\tau_2}} = d_{\tau_1,\tau_2}) = \\ &\int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}\sigma_{x_{\tau_1,\tau_2}}} \exp\left[\frac{-(x_{\tau_1,\tau_2} - d_{\tau_1,\tau_2})^2}{2\sigma^2_{x_{\tau_1,\tau_2}}}\right] dx_{\tau_1,\tau_2} \\ &= \frac{1}{2}\,\mathrm{Erfc}(\frac{\sqrt{d_{\tau_1,\tau_2}}}{2\sigma}) = M(\frac{d_{\tau_1,\tau_2}}{\sigma^2}) \end{aligned}$$

Here, the complementary error function is $\mathrm{Erfc}(t) = \frac{2}{\sqrt{\pi}} \int_t^{\infty} dt' \exp\{(-t'^2)\}$.

The probability of misalignment (*PoM*) is the union of the probability in the spaces bounded by all the inequalities $(x_{\tau_1,\tau_2} \leq 0)$ at sidelobe positions $(\tau_1, \tau_2 \neq 0, 0)$. The individual spaces bounded by the inequalities overlap in general making the exact calculation of *PoM* difficult. However, it is easy to find an upper bound for the *PoM* by assuming no overlap between these spaces. Specifically, $PoM \leq \sum_{(\tau_1,\tau_2 \neq 0,0)} M(x_{\tau_1,\tau_2} \leq 0)$ [143] where the sum is over all sidelobes.

$M(\frac{d_{\tau_1,\tau_2}}{\sigma^2})$ decreases as the distance d_{τ_1,τ_2} increases. Consequently a good criteria should tend to maximize the overall d_i in order to minimize the probability of misalignment. Also, it is of higher priority to maximize the smaller distance, which contributes more to the *PoM*. This is the basis of our ranking criteria, which is completely justified in the low noise limit. Under the low noise limit, the terms of larger d_i make vanishingly small contributions compared to the term of smaller d_i. We show this in Eq. 6.7 by noticing that Erfc(t) can be approximated by $\frac{2}{\sqrt{\pi}} \frac{\exp\{(-t^2)\}}{t}$ for large t (or small σ).

$$\lim_{\sigma \to 0} \frac{M(d_{i+1}/\sigma^2)}{M(d_i/\sigma^2)} = \lim_{\sigma \to 0} \exp\left\{[-\frac{d_{i+1} - d_i}{2\sigma^2}]\right\} \frac{\sqrt{d_i}}{\sqrt{d_{i+1}}} = 0 \tag{6.7}$$

However, the ranking criteria, in general, depends on the noise level σ. We note, due to the central limit theorem, the above results still hold for non-Gaussian noise distributions, when the matrix size is large.

6.12.2 Distance spectra

In order to provide additional useful matrices and to illustrate our ranking criteria, we tabulated, in Table 6.1, part of the peak-sidelobe distance spectra for the top-three ranked square matrices from the exhaustive search results. The values of the first four distances (d_1, d_2, d_3, d_4) and the numbers (n_1, n_2, n_3, n_4) of the corresponding sidelobes are listed. Those top-three binary square matrices are shown in Fig. 6-3 and in Fig. 6-6.

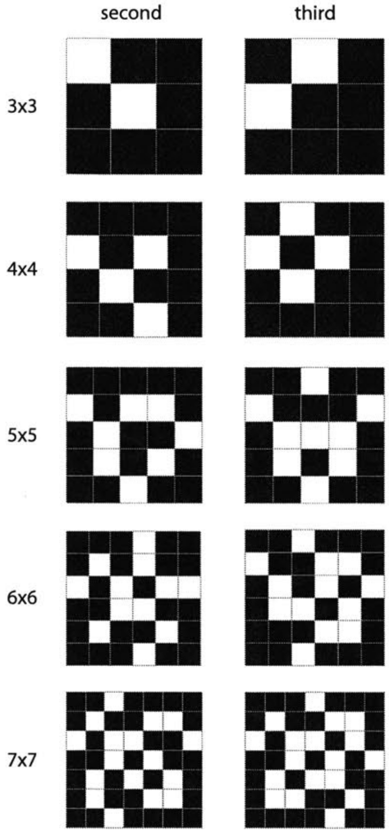


Figure 6-6: (Color Online) Matrices ranked second and third. The first-ranked optimal matrices are shown in Fig. 6-3.

6.13 Acknowledgments

We would like to thank John D. O'Brien, Robert A. Scholtz, Yuan Shen, Moe Win, Ramesh Raskar and Steven G. Johnson for useful discussions. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science

Table 6.1: Peak-sidelobe distance spectra of the top-three ranked square matrices from the exhaustive search results.

N × N	d_1	d_2	d_3	d_4
Ranking	n_1	n_2	n_3	n_4
3 × 3	4	5	6	7
First	4	4	12	4
Second	4	12	6	2
Third	6	6	12	0
4 × 4	7	8	9	10
First	8	8	22	10
Second	10	2	10	18
Third	12	0	8	20
5 × 5	10	11	12	13
First	4	6	10	6
Second	4	12	8	6
Third	4	12	16	12
6 × 6	14	15	16	17
First	4	16	4	2
Second	6	6	12	4
Third	6	8	12	6
7 × 7	19	20	21	22
First	14	8	6	0
Second	16	4	4	4
Third	16	4	8	4

Foundation grant number OCI-1053575. S.S. was supported by the MIT Undergraduate Research Opportunities program (UROP). This work was supported in part by the U.S.A.R.O. through the ISN, under Contract No. W911NF-07-D-0004. L.L. was supported in part by the MRSEC program of the NSF under Award No. DMR-0819762. L.L. and M.S. were partially supported by the MIT S3TEC Energy Research Frontier Center of the Department of Energy under Grant No. DE-SC0001299.

Chapter 7

Numerical Investigation of Cavity Optical Pulse Extraction

7.1 Abstract

Fiber Bragg gratings have been studied for many years in numerous contexts. Not only can they provide the simplest example of 1D bandgaps, they can also exhibit numerous interesting nonlinear effects, such as optical bistability. As shown recently by my collaborators at Friedrich-Schiller-University in Jena, a trigger pulse can be used to create a separate pulse from the light stored in a Bragg cavity defect. This effect, known as cavity-optical pulse extraction or COPE, is a new method for creating ultrashort pulses at arbitrary frequencies. Firstly, I discuss my work at Friedrich-Schiller-University concerning the creation of FDTD solvers to study COPE. After this I will review some particular insights I had in generating small frequency combs. Finally I will mention how this system resembles "White hole" physics and how the analogy can be improved further and explored.

7.2 Introduction

One of the major applications of the coupled mode equations in a Bragg grating is the description of the so-called optical pushbroom effect. In order to describe the optical pushbroom effect, it is instructive to go back to considerations first pointed out by Winful [159].

Winful showed that Bragg gratings can create large changes in the fiber dispersion and that they have regions of negative group velocity dispersion. SPM, and XPM can be used to compress pulses if a negative group velocity dispersion is present. In SPM the pulse will frequency chirp itself, leading to compression, while in XPM the chirp comes from another pulse.

The optical pushbroom effect was first predicted theoretically and simulated by de Sterke [160] [161]. He advanced the ideas of Winful by noting that XPM could be used to compress pulses in a fiber Bragg grating in the region of negative group velocity dispersion. The experimental arrangement the authors envisioned involved a slowly propagating probe pulse tuned below the bandgap. The "pushbroom" part of the effect comes in when a large pump pulse is applied far away from the bandgap. Through XPM, the front edge of the pump pulse lowers the slow probe's frequency, causing it to speed up through the negative group velocity dispersion. The pump pulse goes much faster than the probe pulse, so the net effect is that the low intensity continuous probe pulse, is swept up into one large pulse, that sits on the front of the probe pulse. Since de Sterke's prediction, there have been several experimental studies of the optical pushbroom effect [162] [163] [164] which have been confirmed by simulation.

Eilenberger et al [165] showed that the optical pushbroom effect could be enhanced by utilizing a cavity. The cavity essentially allows the output power to be multiplied by Q, the quality factor of the cavity, relative to the original pushbroom effect. As with the optical pushbroom effect, COPE can be used to generate ultrashort laser pulses at a greater range of frequencies. As long as a CW source is available at a given frequency, it can be converted into a pulsed source utilizing this effect. Eilenberger et al investigated some analytic models of the effect in certain regimes and provided simplified numerical examples to a set of two coupled 1st order differential equations for the forward propagating and backward propagating field envelopes. He also assumed that the only dominant nonlinear interactions were between the trigger field and the forward and backward field envelopes. General nonlinear interaction terms between each field were ignored.

We explored a more thorough numerical solution to the general 2nd order coupled differential equations describing COPE. In addition, we developed a solver to efficiently and

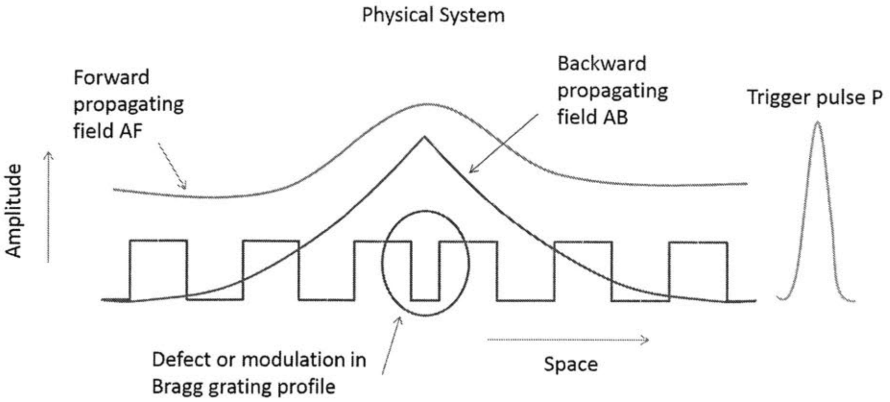


Figure 7-1: Schematic Illustration of Cavity Optical Pulse Extraction in a fiber Bragg grating. Here the Forward propagating mode is pumped from the right (red), and pumps a defect mode in the cavity with some quality factor. The defect mode will also support a local component of the backward propagating portion of the equations (blue) in steady state. We trigger extraction of power from the cavity by sending a trigger pulse green, which propagates from right to left, and which will sweep power from the cavity through cross-phase modulation.

correctly solve for the evolution of these three coupled field envelopes with strong nonlinearities. Further, we created new cavity engineering techniques to generate small frequency combs. In turn we used these techniques to investigate how trigger pulse parameters such power and pulse width influenced the output of the COPE process and the observation of a "double peaked" output pulse for certain parameter ranges.

7.3 Numerical solution to Nonlinear Coupled Mode Equations for Fiber Bragg Gratings with a defect

In this section we outline a thorough numerical investigation of the COPE scheme and details of how all nonlinear interactions, dispersion, Bragg gratings, and resonant cavity effects can be incorporated efficiently for three interacting field envelopes.

As shown in the Appendix, a forward propagating field envelope for the pump A_F, the backward propagation field envelope for the pump A_B, and the trigger pulse P, can be written as the following set of coupled second order differential equations:

$$[i\frac{d}{dz} + \delta(\omega_0) + i\beta_1\frac{d}{dt} - \frac{\beta_2}{2}\frac{d^2}{dt^2} + i\gamma_{AA}(2|A_B|^2 + |A_F|^2) + i2\gamma_{AP}|P|^2]A_F + \kappa(z)A_B = 0 \tag{7.1}$$

$$[-i\frac{d}{dz} + \delta(\omega_0) + i\beta_1\frac{d}{dt} - \frac{\beta_2}{2}\frac{d^2}{dt^2} + i\gamma_{AA}(2|A_F|^2 + |A_B|^2) + i2\gamma_{AP}|P|^2]A_B + \kappa(z)A_F = 0 \tag{7.2}$$

$$[-i\frac{d}{dz} + i\beta_{1P}\frac{d}{dt} - \frac{\beta_{2P}}{2}\frac{d^2}{dt^2} + i2\gamma_{AP}(|A_F|^2 + |A_B|^2) + i\gamma_{PP}|P|^2]P = 0 \tag{7.3}$$

Here β_1 and β_{1P} are the group velocities of the pump and trigger fields, β_2 and β_{2P} are the corresponding group velocity dispersions, $\delta(\omega_0)$ is the cavity detuning of the pump frequency from the band center of the Bragg grating, $\kappa(z)$ is the envelope of the coupling parameter for the Bragg grating, and the γs give nonlinear self-phase and cross-phase terms between the various field parameters. The $\kappa(z)$ term represents the change in the envelope of the Bragg grating strength as a function of position. For instance, a uniform grating would have $\kappa(z) = \frac{\pi\delta n}{\lambda_B}$, where λ_B is the Bragg grating period and δn is the Bragg grating strength. In general the grating strength δn, or the grating phase, can vary with position, and this will be encoded in $\kappa(z)$. The total system and associated fields is indicated schematically in Fig. 1.

Each of these equations support Soliton propagation, and could potentially be solved in the same way as typical Soliton propagation problems through the split-step method. However, a problem with employing this procedure is the presence of $\kappa(z)$, which is not present in most standard computations of Soliton propagation, and which makes the spatial Fourier transform step difficult. It turns out that the most efficient method for simulating this propagation is direct FDTD with Runge-Kutta integration.

Naively we may begin by writing down the set of coupled 1st order differential equations, and trying to solve them directly through a 1st order solver:

$$\frac{dP'}{dt} = \frac{2}{\beta_{2P}}[-i\frac{dP}{dz} + i2\gamma_{AP}(|A_F|^2 + |A_B|^2)P + i\gamma_{PP}|P|^2P + i\beta_{1P}P'] \tag{7.4}$$

$$\frac{dP}{dt} = P' \tag{7.5}$$

Unfortunately this method for splitting P up by its first and second derivatives is unstable since β_{2P} can in general be zero or very small. We can write $\frac{dP}{dt} \approx \frac{P_{i+1}-P_{i-1}}{2\Delta t}$, and $\frac{d^2P}{dt^2} \approx \frac{P_{i+1}+P_{i-1}-2P_i}{2\Delta t}$. Plugging these in, we can find the following stable solution for integrating in time [166]:

$$\frac{\Delta t[i2\gamma_{AP}(|A_F|^2 + |A_B|^2) + i\gamma_{PP}|P_i|^2 - i\frac{d}{dz} + \frac{\beta_{2P}}{\Delta t}]P_i - [\frac{\beta_{2P}}{2\Delta t} + \frac{i\beta_{1p}}{2}]P_{i-1}}{[\frac{\beta_{2P}}{2\Delta t} - \frac{i\beta_{1p}}{2}]} = P_{i+1} \tag{7.6}$$

We still need to employ spatial discretization on the $\frac{d}{dz}$ term. At first we might try to do a special discretization for the leftward and rightward propagating fields. That is lets say we defined the derivative such that the cells on the "receiving" end of the propagating wave are simply not included in the evolution. In turn the two cells at the opposite edge of the simulation are fed with the corresponding values necessary to introduce a constant pump signal, a Soliton, or a Gaussian mode.

$$\frac{dP_i}{dz} \approx \frac{-P_i^{n+2} + 4P_i^{n+1} - 3P_i^n}{2\Delta z} \tag{7.7}$$

This spatial discretization, although avoiding complications from introducing absorbing boundary conditions, is unstable. It turns out that the following symmetric spatial derivative is stable. We can introduce a constant field on the left or right boundaries by simply setting the value to a constant, a Gaussian mode, or a Soliton mode at the single pixel immediately outside the boundary.

$$\frac{dP_i}{dz} \approx \frac{P_i^{n+1} - P_i^{n-1}}{2\Delta z} \tag{7.8}$$

If the boundary is not fixed with a mode source, then it needs to be handled with a type of absorbing boundary condition. We found that 1D Mur absorbing boundary conditions of the following form were suitable for creating stable simulations which had minimum reflections at the boundary:

$$P_{i+1}^{m} = P_i^{m-1} + \frac{\frac{c\Delta t}{\Delta x} - 1}{\frac{c\Delta t}{\Delta x} + 1}[P_{i+1}^{m-1} - P_i^{m}] \tag{7.9}$$

In essence we are forcing the cells at the boundary to satisfy a one-way wave equation of the form $\frac{dP}{dx} = -\beta_{1P}\frac{dP}{dt}$. This works for a general non-dispersive wave equation, but with the inclusion of finite β_2, and nonlinearities, this means that the boundary conditions are not perfectly absorptive for our system. In practice the reflections were minimal.

We tested our solver with the Gaussian modes of the following form:

$$G(z,t) = \frac{\exp\left\{[-\frac{\sigma^2(t-\beta_{1P}z)^2)}{2}\right\}]}{\sqrt{1 - i\sigma^2\beta_{2P}z}} \tag{7.10}$$

We confirmed that the wavepacket broadening and phase evolution followed this expression exactly. We show a few spatial plots of the Gaussian pulse evolution in Fig. 2. We also verified Soliton propagation.

To prepare the system for a trigger pulse, the cavity needed to be stabilized over roughly Q cycles to fully charge the defect cavity. This could be accomplished by applying a constant power level to the forward propagating pump boundary condition. In practice, it was actually best to "precalculate" the expected defect mode shape, by solving the coupled equations for

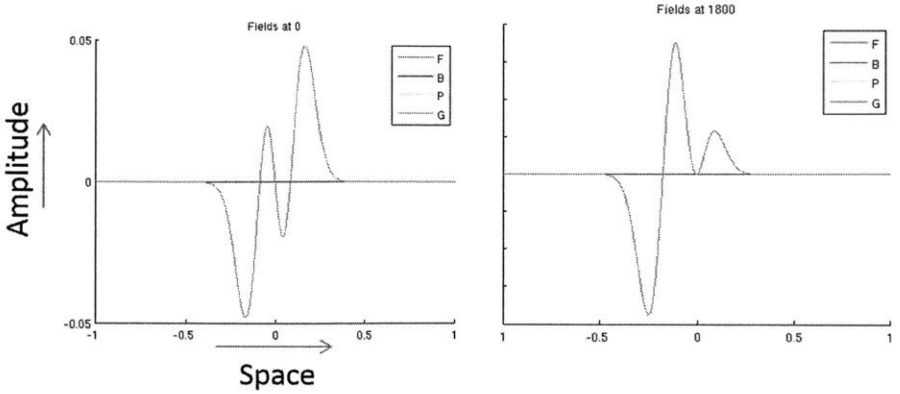


Figure 7-2: Gaussian mode evolution with 1D mode solver. The plot for the theoretical Gaussian mode (purple), completely overlaps the numerically integrated solution (Green).

the cavity at DC, neglecting nonlinearity:

$$+i\frac{dA_F}{dz} + \kappa(z)A_R = 0 \tag{7.11}$$

$$-i\frac{dA_R}{dz} + \kappa(z)A_F = 0 \tag{7.12}$$

Integrating these solutions, we find that the stable defect modes are given by: $A_F = A_0 cosh(\int_{-L}^{z} \kappa(z')dz')$ and $A_R = iA_0 sinh(\int_{-L}^{z} \kappa(z')dz')$. Even after initializing this mode shape, we found that we still had to wait for the system to equilibrate because of the additional optical nonlinearities, especially if the cavity Q was high.

7.4 Engineering Bragg Grating Defects for short Frequency Combs

In this next section we detail how the COPE process operates in certain regimes and how we can tailor the emission of the COPE process to create small frequency combs. We first start by providing a schematic illustration of how COPE transfers energy from the cavity to propagating Bragg grating modes, and then provide further numerical examples using the solver we developed above to give concrete examples.

Fig. 3 provides a schematic illustration of some phase matching conditions between propagating Bragg grating modes and cavity defects. The Bragg grating itself has a dispersion relation given by $\omega = \pm\sqrt{\kappa^2 + k^2}$, where κ is the grating strength, and determines the bandgap size. When a defect is introduced to the grating, defect modes are introduced into the bandgap. These modes do not propagate, but they do have a finite size, so they have some spread in k-space about the bandgap.

When the system is stationary, ω is a conserved quantity. However, when we introduce a traveling shock front in the system, implemented by our trigger pulse, ω is no longer a good conserved quantity because the "potential" induced by the trigger pulse is time-dependent. To understand what we expect to be conserved, we write simplified equations for the pump fields:

$$[i\frac{d}{dz} + \delta(\omega_0) + i\beta_1\frac{d}{dt} + i2\gamma_{AP}|P(x-vt)|^2]A_F + \kappa A_B = 0 \tag{7.13}$$

$$[-i\frac{d}{dz} + \delta(\omega_0) + i\beta_1\frac{d}{dt} + i2\gamma_{AP}|P(x-vt)|^2]A_B + \kappa A_F = 0 \tag{7.14}$$

If $P(x - vt) = 0$, we can immediately assume plane wave solutions for A_F and A_B as $\exp(i(\omega t + kz))$. If we took $v = 0$, this solution would still work, where ω would be conserved and $P(\Delta k) = FT(P(x))$, would serve to couple together different k solutions. We can still realize this useful property by transforming to the comoving reference frame with the pump field. Taking $z' = z + vt$, we can reexpress the plane-wave solutions in our old frame as $\exp(i(\omega - kv)t + kz'))$. This means that $\omega - kv$ should be conserved while, k will still be coupled to other k through $P(\Delta k)$.

In Fig. 3a and 3b we illustrate how this conservation property works in practice. In the stationary frame, lines of $\omega' = \omega - vk$, for a given ω are conserved. In Fig 3b, this result is presented more intuitively as conserved ω in the comoving reference frame. The key idea is that the pump pulse provides momentum δk to scatter modes from the defect state onto the propagating Bragg grating modes, in the same way any defect in a stationary system can scatter k-state modes. Since the stationary version is in a comoving frame, this means in practice that the scattering process will modify the frequency relative to the original defect mode frequency depending on its position in k-space.

In Fig. 4 we illustrate a more general example of phase matching for an arbitrary mismatch between the group velocities of the pump and trigger pulse. Taking the limiting group

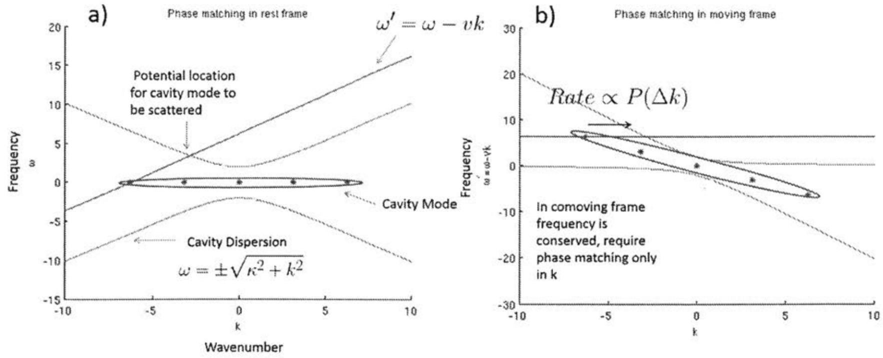


Figure 7-3: Phase matching between Cavity Modes and propagating Bragg grating modes. a) In the stationary frame b) In comoving reference frame.

velocity $\frac{1}{\beta_1} = 1$, we illustrate specific cases with phase matching conditions to the different branches of the Bragg grating dispersion for specific cases of $\beta_{1P} = 0.5$ and 2. If the trigger pulse travels faster than the pump modes, we have phase matching to two points above or below the band gap (Fig. 4a). In the other case, we have phase matching to one point above and one point below the bandgap (Fig. 4b).

We can use these general phase matching properties to investigate creating small frequency combs, by engineering $\kappa(z)$. Since the cavity mode k is translated into some ω around the grating frequency in a linear way, it means features in the cavity mode spectrum in k can be translated into frequency with the trigger pulse. Specifically we can engineer $\kappa(z)$ to be periodic. The periodicity w of the cavities creates a frequency comb separated by $\frac{2\pi v}{w}$ from a trigger pulse traveling at speed v. We illustrate one example of in Fig. 5 of how $\kappa(z)$ can be specified to yield periodic cavity modes.

We confirmed this intuition by directly simulating the COPE process for periodic $\kappa(z)$ in Fig. 6 to generate short frequency combs with four to six peaks. We swept over a range of group velocities for the pump pulse to reveal systematic changes in the generated frequency combs. This plot reveals several phase matching points which move with changing group velocity, and exactly follow our expected predictions for the phase matching conditions.

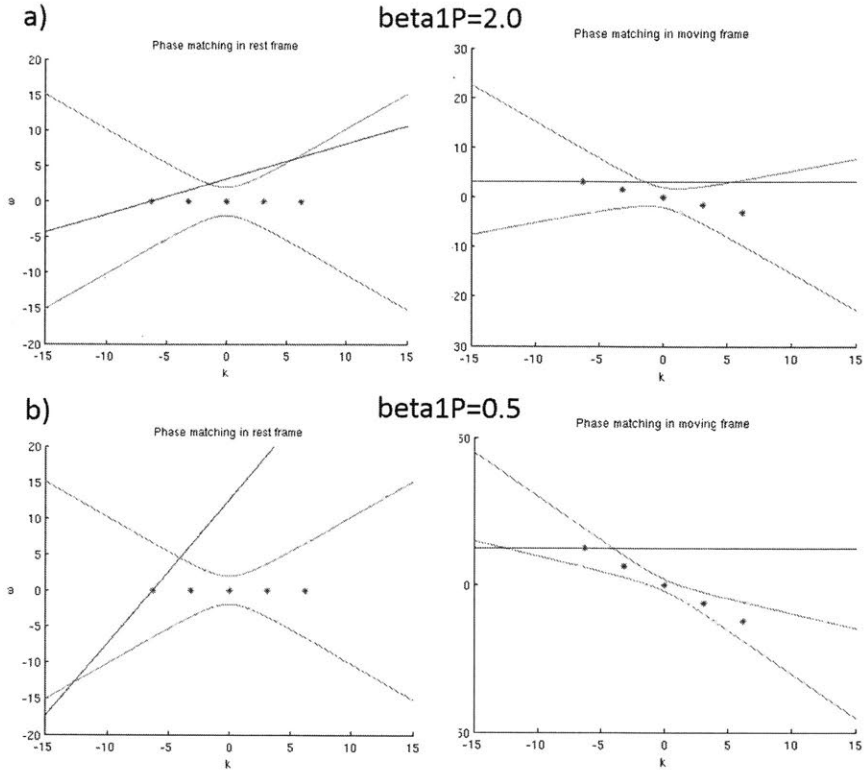


Figure 7-4: Phase matching between Cavity Modes and propagating Bragg grating modes for fast and slow trigger pulse cases. The maximum group velocity of the Bragg grating is fixed to 1. a) Slow case, phase matching occurs for two points above or below the band gap. b) Fast case, phase matching occurs at one point above and one point below the band gap.

7.5 Influence of trigger pulse parameters on COPE

We use some of the tools developed in previous sections to analyze how key trigger pulse parameters influence the generation of the frequency combs and the total COPE output power. To begin we plot in Fig. 7 a set of 6 simulations varying trigger power and pulse length in the exact same format as Fig. 6. Several features immediately become obvious. The first is an increasing asymmetry between the 1st and 3rd quadrants and the 2nd and 4th quadrants with increasing output power (between the 1st and 2nd lines of results). The second is a general broadening of the phased matched frequencies as the duration of the trigger pulse is increased (from left to right).

We can understand the behavior from left to right quite easily in terms of the trigger pulse serving as a "coupling" term between different k states. In the comoving reference frame, a pulse narrower in space, is broader in k-space, and so increases the rate at which

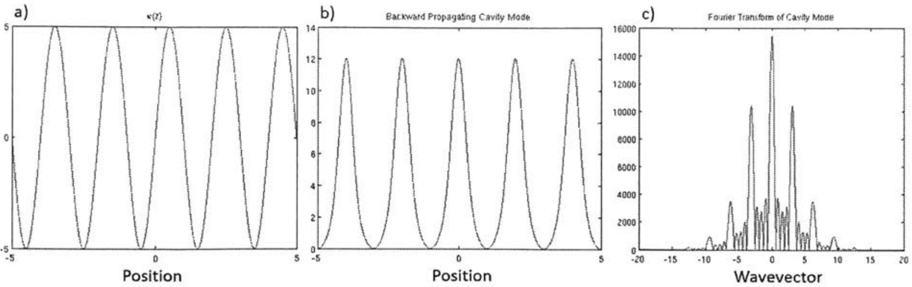


Figure 7-5: Engineering Bragg grating parameter to create periodic cavity modes and short frequency combs. a) Periodic variation of Bragg grating parameter along length of the fiber b) Periodic cavity modes supported by nonuniform, periodic Bragg grating parameter. c) Spatial Fourier transform of cavity modes, yielding a small comb. The COPE process translates this spatial periodicity into frequency.

power is able to be transferred between k states far apart in k-space. On the other hand, a broad pulse will be ineffective at providing the momentum necessary to connect widely separated states.

The asymmetry between the different quadrants is less clear, since the momentum conservation arguments above should apply equally well to positive and negative frequency. We can start to understand the origin of this asymmetry by studying the Bragg grating equations in the presence of a strong DC trigger field.

$$[i\frac{d}{dz} + \delta(\omega_0) + i\beta_1\frac{d}{dt} + i2\gamma_{AP}|P|^2]A_F + \kappa(z)A_B = 0 \tag{7.15}$$

$$[-i\frac{d}{dz} + \delta(\omega_0) + i\beta_1\frac{d}{dt} + i2\gamma_{AP}|P|^2]A_B + \kappa(z)A_F = 0 \tag{7.16}$$

When $|P|$ is a constant wrt time and space, it effectively contributes to the detuning parameter $\delta(\omega_0) = \omega - \omega_0$. We find that the Bragg grating dispersion is modified in the following way:

$$\omega - \omega_0 = \pm\frac{1}{\beta_{1P}}\sqrt{\kappa^2 + k^2} + \frac{2\gamma_{AP}}{\beta_{1P}}|P|^2 \tag{7.17}$$

This qualitatively illustrates how increasing probe power can cause asymmetry between the different scattering channels.

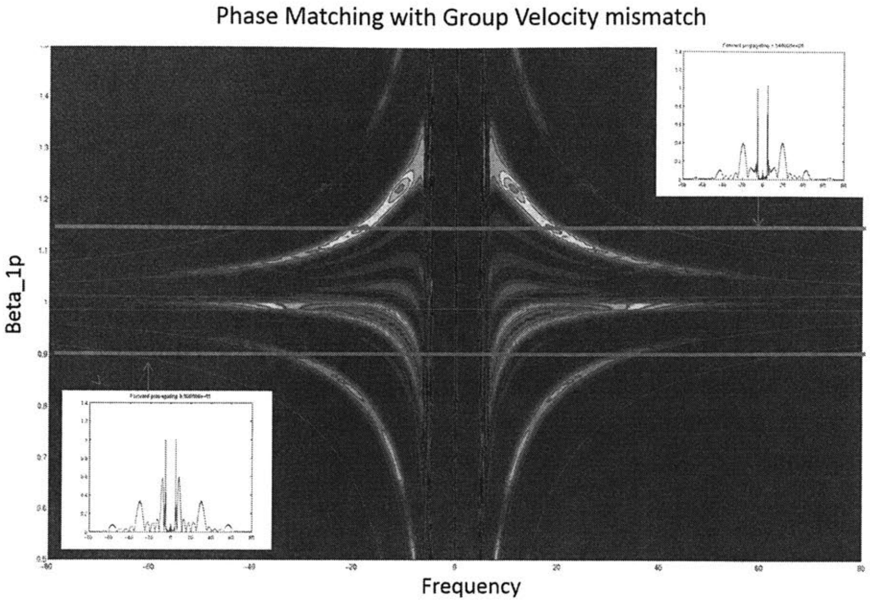


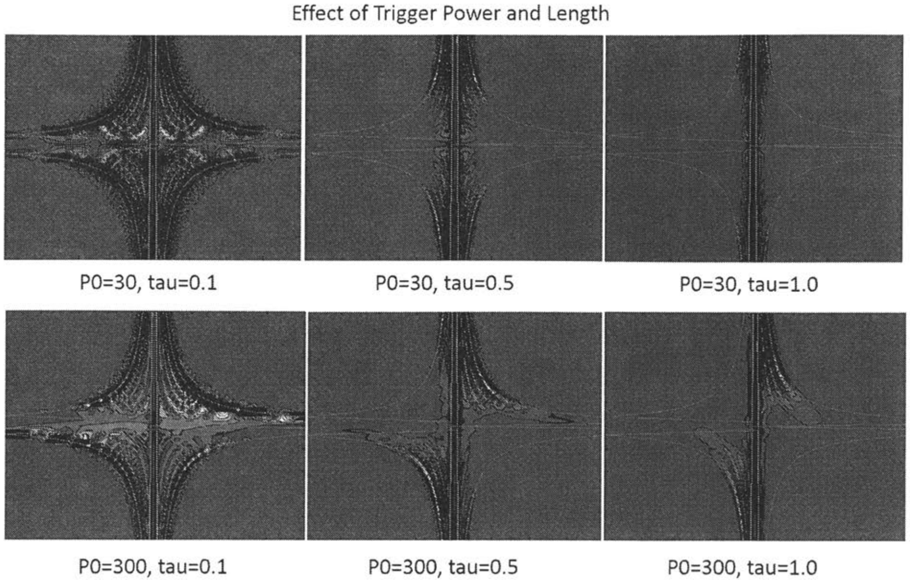
Figure 7-6: Generation of short frequency combs from COPE. Fourier transform of the output of the COPE process for different pump pulse group velocities. The red lines indicate cross sections which we plot as insets. The different colored streaks in the image indicate phase matching to separate orders of the defect mode spatial Fourier transform as a function of trigger group velocity and frequency. The location of these peaks corresponds exactly to the locations we predict from phase matching. The short frequency combs we generate here contain four to six peaks.

We next studied the peak trigger output power as a function of the system parameters and discovered an interesting relationship between the trigger power and defect cavity period at the peak COPE power. We have plotted this result in Fig 8.

To understand the origin of this relationship, we reanalyze the detuning discussion from earlier. An important feature of the result we found above in the simulations, was that it only converged to a fixed form for long trigger pulses. This implies that we should view the trigger pulse as a DC term in the coupled mode equations. Using the dispersion relation we calculated earlier for this situation, we want to analyze the case where the defect mode at $k = \frac{2\pi}{w}$ and $\omega = \omega_0$, is tuned to overlap the Bragg dispersion. This yields the following:

$$0 = \pm \frac{1}{\beta_{1P}} \sqrt{\kappa^2 + k^2} + \frac{2\gamma_{AP}}{\beta_{1P}} |P|^2 \tag{7.18}$$

We find that solving for P gives: $|P|^2 = \frac{1}{2\gamma_{AP}} \sqrt{\kappa^2 + (\frac{2\pi}{w})^2}$, which is exactly the form of the


Effect of Trigger Power and Length

P0=30, tau=0.1 P0=30, tau=0.5 P0=30, tau=1.0

P0=300, tau=0.1 P0=300, tau=0.5 P0=300, tau=1.0

Figure 7-7: Effect of trigger pulse power and length on COPE spectrum. The plots are in the same format as Fig. 6 with variable pulse trigger group velocity on the vertical axis, and frequency on the horizontal axis. From left to right we see that increasing trigger pulse length causes a narrowing in frequency of the COPE output spectrum. In addition from the first line to the second, with increasing trigger power, significantly more power appears in the 1st and 3rd quadrants vs the 2nd and 4th quadrants.

black curve we found in Fig. 8a. We can interpret this as the trigger pulse propagating and distorting the dispersion relation so as to directly touch the cavity mode in frequency and momentum, resulting in maximum power transfer to the COPE pulse.

We further investigated this concept by visualizing the propagation of the COPE pulse in the time domain in Fig. 9. Here we illustrate two examples with $w = 2.2$ and 4 with varying trigger pulse powers. Above each green trigger curve at a particular snapshot, we plot a horizontal line P_{crit}, indicating the power at which the cavity mode is exactly detuned to lie on the Bragg dispersion relation. Carrying this out for several power levels reveals a critical result, namely a splitting of the COPE pulse into two separate peaks, when the peak trigger pulse power exceeds P_{crit}. Intuitively the location of each separate peak corresponds to the value where the trigger pulse power equals P_{crit}. Each of these locations corresponds to the location where power transfer is maximized between the cavity modes and the Bragg dispersion relation. This is an alternate explanation for the same peak splitting phenomena

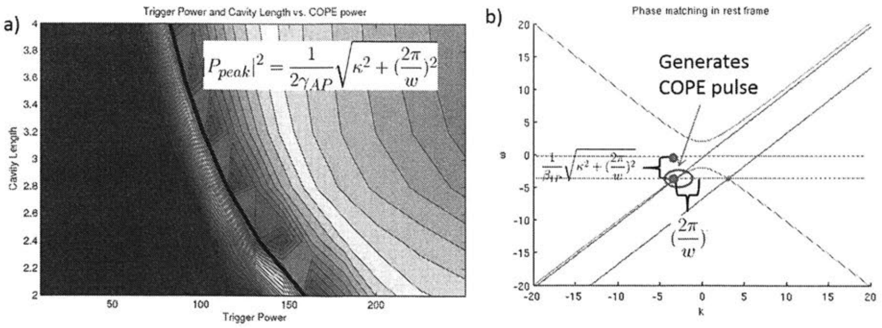


Figure 7-8: a) Peak COPE power vs cavity defect period and trigger power. The peak power follows a curve relating the cavity defect period w to the trigger power resulting in the peak COPE output power P_{peak} in the upper right hand corner. b) This relationship is exactly equivalent to the detuning required to couple a defect mode at $k = \frac{2\pi}{w}$ to the Bragg dispersion relation.

observed in the original work on COPE [165].

7.6 Analogy to "White Hole" Physics

Since event horizons are impossible to observe directly, there have been an entire class of analog physical systems explored which exhibit similar physics which, in certain regimes, might allow observations of exotic effects like Hawking radiation [167].

Most analogs of Hawking radiation exhibit a metric of the following form:

$$ds^2 = (c^2 - v(r,t)^2)dt^2 + 2v(r,t)drdt + dr^2 \tag{7.19}$$

Where $v(r,t)$ is the flow of some medium (say water), and c is the speed of excitations of that medium. If $v(r,t) > c$ at any point, we have an event horizon. The classical example is a fish falling down a waterfall. If the speed of the falling water at some point is greater than the speed of sound in the water, past a certain critical location, it will be impossible to hear the fish anymore because it's passed a "sonic event horizon".

This metric is equivalent to the following dispersion relation describing a frame at rest

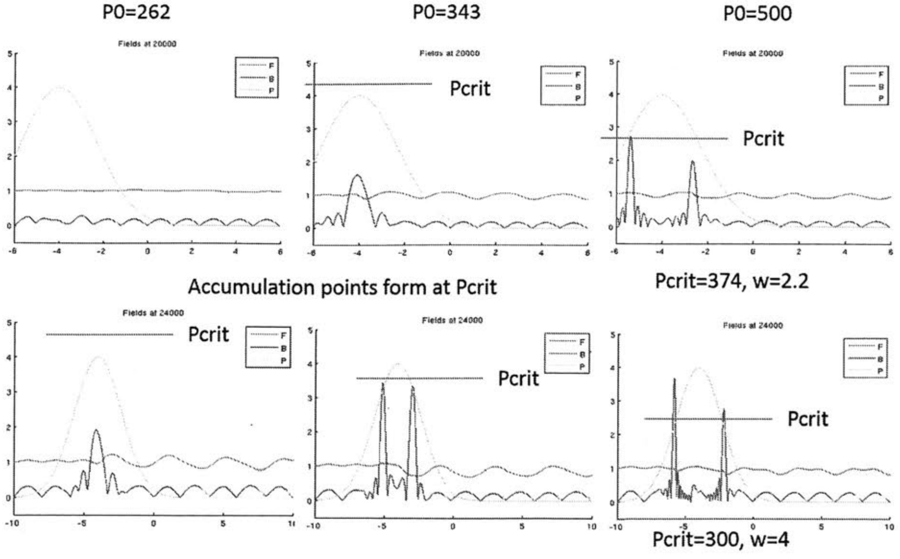


Figure 7-9: Snapshot of COPE process in fiber Bragg grating for $w = 2.2$ and $w = 4$ at several different trigger power levels. The trigger pulse is the green Gaussian curve in the left side of each image, while the forward and backward propagating pump modes are red and blue respectively. P_{crit}, the power calculated to exactly detune the dispersion relation to intercept the peak of the cavity defect mode at $\omega = \omega_0$ and $k = \frac{2\pi}{w}$ is indicated as a horizontal blue line. When the peak energy of the trigger pulse exceeds P_{crit}, we observe that the output pulse splits in two, with the location of each peak corresponding to where the trigger pulse power is equal to P_{crit}.

relative to the flowing fluid:

$$k = \frac{\omega}{\pm c + v(r,t)} \tag{7.20}$$

At the event horizon the group velocity goes to zero for "White holes". We want to see if we can find the analogous situation for our COPE system. Starting with the dispersion relation from earlier:

$$(\omega - \gamma|P|^2 + k)^2 = (k^2 + \kappa^2) \tag{7.21}$$

We want to put this in a form equivalent to the above description, where the event horizon is stationary. Since we believe our horizon will be formed by the trigger pulse, we should transform to the comoving reference frame with the trigger pulse, taking ω to $\omega - vk$. We can find a situation with this where k goes to infinity, if P goes to infinity. This corresponds to the case where an arbitrarily strong trigger pulse, causes light originally propagating on one part of the dispersion relation to travel with a faster and faster group velocity as the local band structure is detuned by the strong nonlinearity. This type of effect could be at work in stabilizing some of the peaks in Fig. 7. To explore this analogy further, it would be best to work with light that is already on the Bragg grating dispersion, instead of in a defect cavity mode. This makes analysis of the results clearer. Here the trigger pulse serves two functions: not only is it kicking the pump out of the defect cavity, it is also forming a horizon. All of the horizon and extraction effects are happening at once, making the horizon interpretation of COPE murky at best.

7.7 Appendix:Derivation of nonlinear coupled mode equations in fiber-Bragg grating

We begin our derivation of the Soliton Equation in fibers with an equation giving the relation of a polarization to the evolution of the $\vec{E}$ field in space [168]. In the derivation of this expression from Maxwell's equations, $\nabla \cdot \vec{E} = 0$ was assumed which is not true for arbitrary ϵ. The expression is written such that the polarization term can be interpreted as a source for the $\vec{E}$ field. A time derivative of polarization is a 'bound' current, since charges are being transfered in space to change the polarization. This is consistent with the fact that if we included $\vec{J}$, $\frac{d}{dt}\vec{J}$ would appear on the right side.

$$\nabla^2 \vec{E} - \frac{1}{c^2}\frac{d^2}{dt^2}\vec{E} = -\epsilon_0 \frac{d^2}{dt^2}\vec{P} \tag{7.22}$$

Here the nonlinear polarization is broken into a linear part and a nonlinear part.

$$\vec{P} = \vec{P_L} + \vec{P_{NL}} \tag{7.23}$$

Assuming that we have very broad pulses, which have a carrier frequency at some frequency ω_0, we can write the time differentiation in the frequency domain as multiplication by $i\omega_0$, since in the frequency domain we know the fourier components are tighly concentrated around ω_0:

$$\begin{aligned}
-\nabla^2\vec{E}(r,\omega) =& \frac{\omega_0^2}{c^2}(1+\chi^1(\omega))\vec{E}(r,\omega) + \vec{P}_{NL}(r,\omega) && (7.24)\\
=& \frac{\omega_0^2}{c^2}(1+\chi^1(\omega) + \chi^3|E(r,\omega_0)|^2)\vec{E}(r,\omega) && (7.25)\\
=& \frac{\omega_0^2}{c^2}\chi(r,\omega)\vec{E}(r,\omega) && (7.26)
\end{aligned}$$

To get the second term we assumed $\vec{P}_{NL}$ was of the form $\chi^3(\omega_0,-\omega_0,\omega)\vec{E}(r,\omega_0)\vec{E}(r,-\omega_0)\vec{E}(r,\omega)$ for simplicity, although in general, we can get phase modulation at ω from other fourier component combinations. It is important to note that we have not yet made the envelope approximation in time, otherwise we would have to have seperate time derivatives of the envelope, from the time derivatives of the carrier. In order to proceed, we break $\vec{E}(r,\omega)$ into two parts dependent on x,y and z seperately to represent progation of different modes along a fiber optic cable.

$$\begin{aligned}
\vec{E}(r,\omega) =& F(x,y)\vec{A}(z,\omega) && (7.27)\\
=& F(x,y)\vec{A}(z,\omega)\exp(i\beta(\omega_0)z) && (7.28)
\end{aligned}$$

For the second step, we made the slowing varying envelope approximation in the z direction, assuming the primary wavevector component is the $\beta(\omega_0)$ wavenumber. Using seperation of variables, we can find seperate eigenvalue equations for $A(z)$ and $F(x,y)$:

$$\begin{aligned}
-\nabla^2 F(x,y)\vec{A}(z,\omega)\exp(i\beta(\omega_0)z) =&\\
\frac{\omega_0^2}{c^2}\chi(\omega,r)F(x,y)\vec{A}(z,\omega)\exp(i\beta(\omega_0)z)&\\
(\frac{d^2}{dx^2}+\frac{d^2}{dy^2}+\frac{\omega_0^2}{c^2}\chi(\omega,r))F(x,y)\vec{A}(z,\omega)\exp(i\beta(\omega_0)z) =&\\
-\frac{d^2}{dz^2}F(x,y)\vec{A}(z,\omega)\exp(i\beta(\omega_0)z)& && (7.29)
\end{aligned}$$

To proceed, we expand out the right side, keeping the first derivative of the envelope:

$$= \exp(i\beta(\omega_0)z)(\beta(\omega_0)^2 - i\beta(\omega_0)\frac{d}{dz})\vec{A}(z,\omega)F(x,y) \tag{7.30}$$

$$\tag{7.31}$$

Now we divide each side by $F(x,y)\vec{A}(z,\omega)$:

$$\frac{(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{\omega_0^2}{c^2}\chi(\omega,r))F(x,y)}{F(x,y)} = \frac{(\beta(\omega_0)^2 - i\beta(\omega_0)\frac{d}{dz})\vec{A}(z,\omega)}{\vec{A}(z,\omega)}$$

$$= \beta_t(\omega)^2 \tag{7.32}$$

The last step follows from the fact that each side is a function of independent variables, and so must be equal to a constant. With this result we can finally write two eigenvalue equations:

$$(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{\omega_0^2}{c^2}\chi(\omega,r) - \beta_t(\omega)^2)F(x,y) = 0 \tag{7.33}$$

$$(\beta(\omega_0)^2 - \beta_t(\omega)^2 - 2i\beta(\omega_0)\frac{d}{dz})\vec{A}(z,\omega) = 0 \tag{7.34}$$

In general these equations are very difficult to solve, but we can first write down a solution of the problem without nonlinear effects, and then perturbatively add them back in. Without nonlinearity the equations reduce to the following:

$$(\frac{d^2}{dx^2} + \frac{d^2}{dy^2} + \frac{\omega_0^2}{c^2}(1 + \chi^1(\omega)) - \beta(\omega)^2)F(x,y) = 0 \tag{7.35}$$

$$(\beta(\omega_0)^2 - \beta(\omega)^2 - 2i\beta(\omega_0)\frac{d}{dz})\vec{A}(z,\omega) = 0 \tag{7.36}$$

The first equation can be solved in terms of Bessel and Neumann functions in cylindrical coordinates. The solution will give a dispersion relation $\beta(\omega)$.

It is important to clarify the definitions of the β symbols used here. $\beta_t(\omega)$ denotes the dispersion relation for the equation with nonlinearity, including corrections to all orders. $\beta_t(\omega)$ can be expanded as $\beta(\omega) + \beta_1(\omega) + \beta_2(\omega) + \ldots$. We will find the first order correction shortly.

With our zeroth order solution, $\beta(\omega)$, we now want to proceed to add back in nonlinear effects. It is first convienent to define $\chi(r,\omega)$ in terms of indices of refraction and losses:

$$\begin{aligned}\chi(r,\omega) &= (n + \Delta n + i\alpha\frac{c}{\omega})^2 && (7.37)\\ &\approx n^2 + 2n\Delta n + \frac{i2cn\alpha}{\omega} && (7.38)\end{aligned}$$

Where we have contrived the expression on the loss part to magically cancel a term later. We find that following correspondances, neglecting the imaginary part of χ^3.

$$\begin{aligned}n^2 =& \mathrm{Re}\{1 + \chi^1(\omega)\} && (7.39)\\ \alpha =& \frac{\omega}{2cn}\,\mathrm{Im}\{\chi^1(\omega)\} && (7.40)\\ \Delta n =& \frac{1}{2n}\,\mathrm{Re}\{\chi^3(\omega_0, -\omega_0, \omega)|E(r,\omega_0)|^2\} && (7.41)\end{aligned}$$

From here we can preform 1st order perturbation theory given some unperturbed eigenstate $F(x,y)$, and eigenvalue $\beta(\omega_0)$. In direct analogy to quantum mechanics we can write:

$$\begin{aligned}\hat{H}_0 &= \frac{d^2}{dx^2} + \frac{d}{dy^2} + \frac{\omega^2}{c^2}n^2 && (7.42)\\ \hat{H}_1 &= \frac{\omega^2}{c^2}(2n\Delta n) + i\frac{\omega}{c}n\alpha && (7.43)\end{aligned}$$

Where:

$$\begin{aligned}\frac{<\psi|\hat{H}_0|\psi>}{<\psi|\psi>} =& E_0 = \beta(\omega)^2 && (7.44)\\ \frac{<\psi|\hat{H}_1|\psi>}{<\psi|\psi>} =& E_1 = 2\beta_1\beta(\omega) && (7.45)\end{aligned}$$

Where β_1, is the first order perturbation to the eigenvalue. Using the definition of the inner product, we can write:

$$\beta_1 = \frac{\omega^2}{c^2}\frac{n(\omega)}{\beta(\omega)}\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\Delta n + i\frac{c}{\omega}\alpha)|F(x,y)|^2dxdy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|F(x,y)|^2dxdy} \qquad (7.46)$$

We can make a significant simplication to this expression if we assume $\beta(\omega)$ satisfies the free space dispersion relation: $\beta(\omega) = \frac{\omega n}{c}$. This follows if we assume $F(x, y)$ is a slowly varying function in x and y, because the x and y derivatives in the eigenvalue equation for $F(x, y)$ are negigible.

$$\begin{aligned}
\beta_1 =& \frac{\omega}{c}\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\Delta n + i\frac{c}{\omega}\alpha)|F(x,y)|^2 dxdy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|F(x,y)|^2 dxdy} && (7.47)\\
=& \frac{\omega}{c}\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{1}{2n}\chi^3|E(r,\omega_0)|^2|F(x,y)|^2 dxdy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|F(x,y)|^2 dxdy} + i\alpha && (7.48)\\
=& |A(z,\omega)|^2 \frac{\omega\chi^3}{2cn}\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|F(x,y)|^4 dxdy}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|F(x,y)|^2 dxdy} + i\alpha && (7.49)\\
=& \gamma|A(z,\omega)|^2 + i\alpha && (7.50)
\end{aligned}$$

Using these results we can rewrite the eigenvalue equation for $\vec{A}(z, \omega)$, using $\beta_t(\omega) \approx \beta(\omega) + \beta_1(\omega)$:

$$\begin{aligned}
0 \approx& ((\beta(\omega) + \beta_1(\omega))^2 - \beta(\omega_0)^2 - 2i\beta(\omega_0)\frac{d}{dz})A && (7.51)\\
=& ((\beta(\omega) + \beta_1(\omega) + \beta(\omega_0))(\beta(\omega) + \beta_1(\omega) - \beta(\omega_0)) && (7.52)\\
&- 2i\beta(\omega_0)\frac{d}{dz})A && (7.53)
\end{aligned}$$

To proceed, we assume that for the ω we are concerned with, we are close to ω_0 and that β_1 is small. This allows us to write:

$$\begin{aligned}
\approx& (2\beta(\omega_0)(\beta(\omega) + \beta_1(\omega) - \beta(\omega_0)) - 2i\beta(\omega_0)\frac{d}{dz})A && (7.54)\\
=& (\beta(\omega) + \beta_1(\omega) - \beta(\omega_0) - i\frac{d}{dz})A(z,\omega) && (7.55)
\end{aligned}$$

Now that we have this useful expression we would like to transform it back to the time domain. Since we are only interested in frequencies around ω_0, we can expand the dispersion relation for $\beta(\omega)$ to second order:

$$\beta(\omega) \approx \beta(\omega_0) + \frac{d}{d\omega}\beta(\omega)|_{\omega_0}(\omega-\omega_0) + \frac{1}{2}\frac{d^2}{d\omega^2}\beta(\omega)|_{\omega_0}(\omega-\omega_0)^2 \tag{7.56}$$

$$=\beta(\omega_0) + \frac{1}{v_g}(\omega-\omega_0) + \frac{1}{2g_d}(\omega-\omega_0)^2 \tag{7.57}$$

Where v_g is the wave group velocity and g_d is the group velocity dispersion at ω_0. We can insert this into our expression for $A(z,\omega)$ above:

$$0 =(\frac{1}{v_g}(\omega-\omega_0) + \frac{1}{2g_d}(\omega-\omega_0)^2 + \beta_1(\omega)) - i\frac{d}{dz})A(z,\omega) \tag{7.58}$$

To proceed we make the envelope approximation for $A(z,\omega)$:

$$A(z,\omega) = \int_{-\infty}^{\infty} A(z,t)\exp(-i\omega t) \tag{7.59}$$

$$= \int_{-\infty}^{\infty} A_e(z,t)\exp(i\omega_0 t)\exp(-i\omega t) \tag{7.60}$$

$$=A_e(z,\omega-\omega_0) \tag{7.61}$$

Where $A_e(z,\omega)$ is the time envelope. For simplicity we immediately drop the e subscript. We were able to neglect higher order terms in $\beta(\omega)$ because $A(z,\omega-\omega_0)$ is assumed to be a narrow function in frequency space.

$$0 =(\frac{1}{v_g}(\omega-\omega_0) + \frac{1}{2g_d}(\omega-\omega_0)^2 + \beta_1(\omega)) - i\frac{d}{dz})A(z,\omega-\omega_0) \tag{7.62}$$

We can know immediately preform a fourier transform on this function, and exhange multiplication by $\omega-\omega_0$ for $i\frac{d}{dt}$:

$$0 =(\frac{i}{v_g}\frac{d}{dt} - \frac{1}{2g_d}\frac{d^2}{dt^2} + \beta_1(t)) - i\frac{d}{dz})A(z,t) \tag{7.63}$$

$$=(\frac{1}{v_g}\frac{d}{dt} + i\frac{1}{2g_d}\frac{d^2}{dt^2} - i\gamma|A(z,t)|^2 + \alpha + \frac{d}{dz})A(z,t) \tag{7.64}$$

For the last step we inserted the expression we found for β_1 earlier using perturbation theory.

Through a similar procedure we can include XPM effects making the approximation that

$\gamma_{i,j} = \gamma$ if we assume the fiber modes have similiar shapes:

$$0 =\{\frac{1}{v_{g1}}\frac{d}{dt} + i\frac{1}{2g_{d1}}\frac{d^2}{dt^2} - i\gamma(|A_1(z,t)|^2 + 2|A_2(z,t)|^2) + \alpha_1 + \frac{d}{dz}\}A_1(z,t) \tag{7.65}$$

$$0 =\{\frac{1}{v_{g2}}\frac{d}{dt} + i\frac{1}{2g_{d2}}\frac{d^2}{dt^2} - i\gamma(|A_2(z,t)|^2 + 2|A_1(z,t)|^2) + \alpha_2 + \frac{d}{dz}\}A_2(z,t) \tag{7.66}$$

Now that we understand the effect of the Bragg grating on the dispersion relation, we can proceed to derive a set of coupled mode equations. For this derivation we follow the work of several authors [169] [160] [159] [170] [171]. Our starting point is the following equation for the total field $\vec{E}(r,t)$ fourier transformed in time:

$$-\nabla^2\vec{E}(r,\omega) =\frac{\omega^2}{c^2}\chi(r,\omega)\vec{E}(r,\omega) \tag{7.67}$$

Here $\chi(r,\omega)$ is now given by:

$$\chi(r,\omega) = (n + n_{NL} + \delta n\cos(2\beta_g z))^2 \tag{7.68}$$

$$\approx n^2 + 2nn_{NL} + 2n\delta n\cos(2\beta_g z) \tag{7.69}$$

Where we have now included the contribution to the polarization from the Bragg grating, and where:

$$n_{NL} =\frac{1}{2n}\operatorname{Re}\{\chi^3(\omega,-\omega,\omega)|E(r,\omega)|^2\} \tag{7.70}$$

Instead of considering two different frequency modes, as in our derivation of the general coupled mode equations, we are interested in the coupling of the forward propagating and backward propagating components at the same frequency around the zone edge:

$$\vec{E}(r,\omega) = F(x,y)\vec{A}_f(z)\exp(i\beta_g z) + F(x,y)\vec{A}_b(z)\exp(-i\beta_g z) \tag{7.71}$$

Here we have expanded in envelopes of the foward and backward propagating modes. The mode profiles $F(x,y)$ for each mode will be identical, and we take the carrier wavevector to be β_g, the Bragg grating wavevector, because the coupling between the forward and backward propagating modes will be strongest here.

The next step is to insert our expression for the total field into the main equation above and seperate out the components which have the same wavevector dependence:

$$\begin{aligned}
&F(x,y)^{-1}\exp(i\beta_g z)(\nabla + i\beta_g \hat{k})^2 \vec{A}_f(z,\omega)F(x,y) = \\
&\frac{\omega^2}{c^2}\exp(i\beta_g z)\{(n^2 + 2nn_{NL})\vec{A}_f(z,\omega) + \frac{n\delta n}{2}\vec{A}_b(z,\omega)\} \qquad (7.72) \\
&F(x,y)^{-1}\exp(-i\beta_g z)(\nabla - i\beta_g \hat{k})^2 \vec{A}_b(z,\omega)F(x,y) = \\
&\frac{\omega^2}{c^2}\exp(-i\beta_g z)\{(n^2 + 2nn_{NL})\vec{A}_b(z,\omega) + \frac{n\delta n}{2}\vec{A}_f(z,\omega)\} \qquad (7.73)
\end{aligned}$$

It is important to note here that the Bragg grating serves as a coupling between the forward and backward propagating modes because it connects a mode with $\exp(i\beta_g z)$ dependence to a mode with $\exp(-i\beta_g z)$ dependence and vice versa. Taking the first of these equations and applying the envelope approximation we find:

$$\begin{aligned}
&(\beta_g^2 + 2i\beta_g\frac{d}{dz} + \frac{d^2}{dx^2} + \frac{d^2}{dy^2})\vec{A}_f(z,\omega)F(x,y) \approx \\
&\frac{\omega^2}{c^2}\{(n^2 + 2nn_{NL})\vec{A}_f(z,\omega) + \frac{2n\delta n}{2}\vec{A}_b(z,\omega)\}F(x,y) \qquad (7.74)
\end{aligned}$$

We can split this into two seperate equations using the techniques from the first section. We skip some of these steps to write:

$$(-\beta_g^2 + \beta_t(\omega)^2 + 2i\beta_g\frac{d}{dz})\vec{A}_f(z,\omega) + \frac{\omega^2}{c^2}n\delta n\vec{A}_b(z,\omega) = 0 \qquad (7.75)$$

Here we remind ourselves that β_t is the unperturbed eigenvalue for the mode profile $F(x,y)$. We know make the approximation that $-\beta_g^2 + \beta_t(\omega)^2 \approx 2\beta_g(\beta_t(\omega) - \beta_g)$ so we can write:

$$\begin{aligned}
0 =& (\beta_t(\omega) - \beta_g + i\frac{d}{dz})\vec{A}_f(z,\omega) + \frac{n\delta n\omega^2}{2\beta_g c^2}\vec{A}_b(z,\omega) \qquad (7.76) \\
=& (\beta(\omega) + \beta_1(\omega) - \beta_g + i\frac{d}{dz})\vec{A}_f(z,\omega) + \frac{n\delta n\omega^2}{2\beta_g c^2}\vec{A}_b(z,\omega) \qquad (7.77)
\end{aligned}$$

In the second step we expanded β_t to the zeroth order dispersion relation for the fiber mode

$\beta(\omega)$ and the first nonlinear correction term $\beta_1(\omega)$. We now want to rewrite the above equation in the following way:

$$0 = (i\delta(\omega) + i\beta_1(\omega) - \frac{d}{dz})\vec{A}_f(z,\omega) + i\kappa\vec{A}_b(z,\omega) \tag{7.78}$$

$$0 = (i\delta(\omega) + i\beta_1(\omega) + \frac{d}{dz})\vec{A}_b(z,\omega) + i\kappa\vec{A}_f(z,\omega) \tag{7.79}$$

Where we have:

$$\delta(\omega) = \beta(\omega) - \beta_g \tag{7.80}$$

$$\kappa = \frac{n\delta n\omega^2}{2\beta_g c^2} \approx \frac{\delta n\beta_g}{2n} \tag{7.81}$$

Now we take the opportunity to do a sanity check of the above equations when the nonlinearity goes to zero. We find:

$$0 = (i\delta(\omega) - \frac{d}{dz})\vec{A}_f(z,\omega) + i\kappa\vec{A}_b(z,\omega) \tag{7.82}$$

$$0 = (i\delta(\omega) + \frac{d}{dz})\vec{A}_b(z,\omega) + i\kappa\vec{A}_f(z,\omega) \tag{7.83}$$

We can solve these equations using trial functions of the form $A_f(z,\omega) = A_1\exp(iqz) + A_2\exp(-iqz)$ and $A_b(z,\omega) = B_1\exp(iqz) + B_2\exp(-iqz)$. Upon substitution we identify the following constraints:

$$(q - \delta(\omega))A_1 = \kappa B_1 \tag{7.84}$$

$$-(q + \delta(\omega))B_1 = \kappa A_1 \tag{7.85}$$

Along with two other relations of the same form. This can easily be solved for nonzero field amplitudes to yield the constraint:

$$\delta(\omega) = \pm\sqrt{q^2 + \kappa^2} \tag{7.86}$$

$$\omega - \omega_g = \pm\frac{c}{n}\sqrt{(\beta - \beta_g)^2 + (\frac{\delta n\beta_g}{2n})^2} \tag{7.87}$$

Where in the second step, we assumed a free space dispersion, and expressed q in terms of the wavenumber β of the envelope. We can briefly analyze the limits of this dispersion relation. For $\omega \gg \omega_g$, we recover the free-space dispersion:

$$\omega \approx \frac{c}{n}\beta \tag{7.88}$$

We can also find an expression for the bandgap when $\beta = \beta_g$

$$\Delta = \frac{2c}{n}\frac{\delta n \beta_g}{2n} = \frac{c\beta_g \delta n}{n^2} \tag{7.89}$$

Which is exactly the same result we derived earlier using perturbation theory. Now that we are confident that these equations are correct, we can transform them to the time domain. To do this we first expand $\beta(\omega)$ about some center frequency ω_0 near the zone boundary: $\beta(\omega) \approx \beta(\omega_0) + \frac{1}{v_g}(\omega - \omega_0)$. We can substitute this into our fourier transformed expressions for the envelope evolution:

$$0 = (i\delta(\omega_0) + i\frac{1}{v_g}(\omega - \omega_0) + i\beta_1(\omega) - \frac{d}{dz})\vec{A}_f(z,\omega) + i\kappa\vec{A}_b(z,\omega) \tag{7.90}$$

$$0 = (i\delta(\omega_0) + i\frac{1}{v_g}(\omega - \omega_0) + i\beta_1(\omega) + \frac{d}{dz})\vec{A}_b(z,\omega) + i\kappa\vec{A}_f(z,\omega) \tag{7.91}$$

Using the envelope approximation, we can write this as:

$$0 = (i\delta(\omega_0) + i\frac{1}{v_g}(\omega - \omega_0) + i\beta_1(\omega) - \frac{d}{dz})\vec{A}_{ef}(z,\omega - \omega_0) + i\kappa\vec{A}_{eb}(z,\omega - \omega_0)$$

$$0 = (i\delta(\omega_0) + i\frac{1}{v_g}(\omega - \omega_0) + i\beta_1(\omega) + \frac{d}{dz})\vec{A}_{eb}(z,\omega - \omega_0) + i\kappa\vec{A}_{ef}(z,\omega - \omega_0)$$

Where the e subscript denotes an envelope in the time domain. We can easily fourier transform these equations to find, again dropping the e designation on the envelope:

$$0 = (i\delta(\omega_0) + i\gamma(|A_f(z,t)|^2 + 2|A_b(z,t)|^2) - \frac{d}{dz} - \frac{1}{v_g}\frac{d}{dt})\vec{A}_f(z,t) + i\kappa\vec{A}_b(z,t)$$

$$0 = (i\delta(\omega_0) + i\gamma(|A_b(z,t)|^2 + 2|A_f(z,t)|^2) + \frac{d}{dz} - \frac{1}{v_g}\frac{d}{dt})\vec{A}_b(z,t) + i\kappa\vec{A}_f(z,t)$$

Chapter 8

Future Work

Future work on 2D ferrimagnetic topological photonic crystals should focus on removing the external magnetic field. We investigated employing other varieties of gyromagnetic material to accomplish this, but unfortunately, their low coercivity prevented the rods from sustaining suitable magnetization. Thankfully, there is an ever growing variety of magnetic materials which potentially can be utilized to help realize the amazing properties of these crystals in more portable settings.

Additional studies on lens-enabled LIDAR should further examine phase error propagation through the grating. The high sensitivity to manufacturing errors and wavelength dispersion of the system needs to be addressed. The historical literature on advanced optics design may potentially help here. A part of the solution will also likely involve developing robust methods for tuning the focal plane position. Work by our collaborators at MIT Lincoln Labs in future years will focus on creating a full working system including integrated tunable sources and heterodyne detection. Packaging and integration of all of these different components will be highly nontrivial, but accomplishing this will be a critical step towards commercialization.

The most development remains for our proposal for optical CNNs. Although inference using networks of MZIs has been demonstrated experimentally, it was only for very small systems of a few inputs and outputs. Future work needs to address how this can be scaled to hundreds or thousands of inputs while keeping losses manageable and maintaining the required device tolerances. In addition robust methods for integrating optical amplifiers, optical nonlinearity, and extremely long delay lines need to be created.

Bibliography

[1] Zheng Wang, YD Chong, John D Joannopoulos, and Marin Soljačić. Reflection-free one-way edge modes in a gyromagnetic photonic crystal. *Phys. Rev. Lett.*, 100(1):013905, 2008.

[2] Zheng Wang, Yidong Chong, J. D. Joannopoulos, and Marin Soljačić. Observation of unidirectional backscattering-immune topological electromagnetic states. *Nature*, 461(7265):772–775, October 2009.

[3] Jeffrey S Herd and M David Conway. The evolution to modern phased array architectures. *Proceedings of the IEEE*, 104(3):519–529, 2016.

[4] Macom in active antennas. https://www.macom.com/activeantennas. Accessed: 2017-04-30.

[5] Jie Sun, Erman Timurdogan, Ami Yaacobi, Ehsan Shah Hosseini, and Michael R Watts. Large-scale nanophotonic phased array. *Nature*, 493(7431):195–199, 2013.

[6] Joerg Schoebel and Pablo Herrero. *Planar antenna technology for mm-wave automotive radar, sensing, and communications.* INTECH Open Access Publisher, 2010.

[7] Scott A. Skirlo, Ling Lu, and Marin Soljačić. Multimode one-way waveguides of large chern numbers. *Phys. Rev. Lett.*, 113:113904, Sep 2014.

[8] F. D. M. Haldane. Model for a quantum hall effect without landau levels: Condensed-matter realization of the "parity anomaly". *Phys. Rev. Lett.*, 61:2015–2018, Oct 1988.

[9] F. D. M. Haldane and S. Raghu. Possible realization of directional optical waveguides in photonic crystals with broken time-reversal symmetry. *Phys. Rev. Lett.*, 100:013904, Jan 2008.

[10] Cui-Zu Chang, Jinsong Zhang, Xiao Feng, Jie Shen, Zuocheng Zhang, Minghua Guo, Kang Li, Yunbo Ou, Pang Wei, Li-Li Wang, et al. Experimental observation of the quantum anomalous hall effect in a magnetic topological insulator. *Science*, 340(6129):167–170, 2013.

[11] Jing Wang, Biao Lian, Haijun Zhang, Yong Xu, and Shou-Cheng Zhang. Quantum anomalous hall effect with higher plateaus. *Phys. Rev. Lett.*, 111:136801, Sep 2013.

[12] Chen Fang, Matthew J. Gilbert, and B. Andrei Bernevig. Large-chern-number quantum anomalous hall effect in thin-film topological crystalline insulators. *Phys. Rev. Lett.*, 112(4):046801, 2014.

[13] Mohammad Hafezi, Eugene A Demler, Mikhail D Lukin, and Jacob M Taylor. Robust optical delay lines with topological protection. *Nat. Physics*, 7(11):907–912, 2011.

[14] Kejie Fang, Zongfu Yu, and Shanhui Fan. Realizing effective magnetic field for photons by controlling the phase of dynamic modulation. *Nat. Photonics*, 6(11):782–787, 2012.

[15] Alexander B Khanikaev, S Hossein Mousavi, Wang-Kong Tse, Mehdi Kargarian, Allan H MacDonald, and Gennady Shvets. Photonic topological insulators. *Nature Mater.*, 2012.

[16] Yaacov E Kraus, Yoav Lahini, Zohar Ringel, Mor Verbin, and Oded Zilberberg. Topological states and adiabatic pumping in quasicrystals. *Phys. Rev. Lett.*, 109(10):106402, 2012.

[17] Mikael C Rechtsman, Julia M Zeuner, Yonatan Plotnik, Yaakov Lumer, Daniel Podolsky, Felix Dreisow, Stefan Nolte, Mordechai Segev, and Alexander Szameit. Photonic floquet topological insulators. *Nature*, 496(7444):196–200, 2013.

[18] Ling Lu, Liang Fu, John D Joannopoulos, and Marin Soljačić. Weyl points and line nodes in gyroid photonic crystals. *Nat. Photonics*, 7(4):294–299, 2013.

[19] Wen-Jie Chen, Shao-Ji Jiang, Xiao-Dong Chen, Baocheng Zhu, Lei Zhou, Jian-Wen Dong, and CT Chan. Experimental realization of photonic topological insulator in a uniaxial metacrystal waveguide. *Nat. Commun.*, 5, 2014.

[20] S Raghu and FDM Haldane. Analogs of quantum-hall-effect edge states in photonic crystals. *Phys. Rev. A*, 78(3):033834, 2008.

[21] Yasuhiro Hatsugai. Chern number and edge states in the integer quantum hall effect. *Phys. Rev. Lett.*, 71(22):3697, 1993.

[22] Y. D. Chong, Xiao-Gang Wen, and Marin Soljacic. Effective theory of quadratic degeneracies. *Phys. Rev. B*, 77(23):235125, JUN 2008 2008.

[23] T. Fukui, Y. Hatsugai, and H. Suzuki. Chern numbers in discretized brillouin zone: Efficient method of computing (spin) hall conductances. *J. Phys. Soc. Jpn.*, 74(6):1674–1677, JUN 2005 2005.

[24] David M Pozar. Microwave engineering, 1998.

[25] Jin-Xin Fu, Rong-Juan Liu, and Zhi-Yuan Li. Robust one-way modes in gyromagnetic photonic crystal waveguides with different interfaces. *Appl. Phys. Lett.*, 97(4):041112, 2010.

[26] Yin Poo, Rui-xin Wu, Zhifang Lin, Yan Yang, and CT Chan. Experimental realization of self-guiding unidirectional electromagnetic edge states. *Phys. Rev. Lett.*, 106(9):93903, 2011.

[27] Jin-Xin Fu, Jin Lian, Rong-Juan Liu, Lin Gan, and Zhi-Yuan Li. Unidirectional channel-drop filter by one-way gyromagnetic photonic crystal waveguides. *Appl. Phys. Lett.*, 98(21):211104–211104, 2011.

[28] Yan Yang, Yin Poo, Rui-xin Wu, Yan Gu, and Ping Chen. Experimental demonstration of one-way slow wave in waveguide involving gyromagnetic photonic crystals. *Appl. Phys. Lett.*, 102:231113, 2013.

[29] Zhuoyuan Wang, L. Shen, Y. Zaihe, Xianmin Zhang, and Xiaodong Zheng. Highly efficient photonic-crystal splitters based on one-way waveguiding. *J. Opt. Soc. Am. B*, 30:173–176, Jan 2013.

[30] Cheng He, Xiao-Lin Chen, Ming-Hui Lu, Xue-Feng Li, Wei-Wei Wan, Xiao-Shi Qian, Ruo-Cheng Yin, and Yan-Feng Chen. Tunable one-way cross-waveguide splitter based on gyromagnetic photonic crystal. *Appl. Phys. Lett.*, 96(11):111111–111116, 2010.

[31] Xianyu Ao, Zhifang Lin, and CT Chan. One-way edge mode in a magneto-optical honeycomb photonic crystal. *Phys. Rev. B*, 80(3):033105, 2009.

[32] Kejie Fang, Zongfu Yu, and Shanhui Fan. Microscopic theory of photonic one-way edge mode. *Phys. Rev. B*, 84:075477, 2011.

[33] Kexin Liu, Linfang Shen, and Sailing He. One-way edge mode in a gyromagnetic photonic crystal slab. *Opt. Lett.*, 37(19):4110–4112, 2012.

[34] Ara A Asatryan, Lindsay C Botten, Kejie Fang, Shanhui Fan, and Ross C McPhedran. Local density of states of chiral hall edge states in gyrotropic photonic clusters. *Physical Review B*, 88(3):035127, 2013.

[35] Xiaofei Zang and Chun Jiang. Edge mode in nonreciprocal photonic crystal waveguide: manipulating the unidirectional electromagnetic pulse dynamically. *JOSA B*, 28(3):554–557, 2011.

[36] W Qiu, Z Wang, and M Soljačić. Broadband circulators based on directional coupling of one-way waveguides. *Opt. Express*, 19(22):22248, 2011.

[37] Scott A Skirlo, Ling Lu, Yuichi Igarashi, Qinghui Yan, John Joannopoulos, and Marin Soljačić. Experimental observation of large chern numbers in photonic crystals. *Physical review letters*, 115(25):253901, 2015.

[38] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. Quantized hall conductance in a two-dimensional periodic potential. *Phys. Rev. Lett.*, 49:405–408, Aug 1982.

[39] K v Klitzing, Gerhard Dorda, and Michael Pepper. New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance. *Phys. Rev. Lett.*, 45(6):494, 1980.

[40] Ling Lu, John D Joannopoulos, and Marin Soljačić. Topological photonics. *Nat. Photonics*, 2014.

[41] Gregor Jotzu, Michael Messer, Rémi Desbuquois, Martin Lebrat, Thomas Uehlinger, Daniel Greif, and Tilman Esslinger. Experimental realization of the topological haldane model with ultracold fermions. *Nature*, 515(7526):237–240, 2014.

[42] Hua Jiang, Zhenhua Qiao, Haiwen Liu, and Qian Niu. Quantum anomalous hall effect with tunable chern number in magnetic topological insulator film. *Physical Review B*, 85(4):045445, 2012.

[43] Pablo M Perez-Piskunow, LEF Foa Torres, and Gonzalo Usaj. Hierarchy of floquet gaps and edge states for driven honeycomb lattices. *Physical Review A*, 91(4):043625, 2015.

[44] Jerome J Green and Frank Sandy. Microwave characterization of partially magnetized ferrites. *IEEE Trans. Microw. Theory Techn.*, 22(6):641–645, 1974.

[45] Sucheng Li, Chendong Gu, Yadong Xu, Shahzad Anwar, Weixin Lu, Zhi Hong Hang, Bo Hou, and Huanyang Chen. Probing electric field in an enclosed field mapper for characterizing metamaterials. *International Journal of Antennas and Propagation*, 2014, 2014.

[46] Alexander Mook, Jürgen Henk, and Ingrid Mertig. Edge states in topological magnon insulators. *Phys. Rev. B*, 90(2):024412, 2014.

[47] Joel Yuen-Zhou, Semion K Saikin, Norman Y Yao, and Alán Aspuru-Guzik. Topologically protected excitons in porphyrin thin films. *Nature materials*, 13(11):1026–1032, 2014.

[48] Zhaoju Yang, Fei Gao, Xihang Shi, Xiao Lin, Zhen Gao, Yidong Chong, and Baile Zhang. Topological acoustics. *Phys. Rev. Lett.*, 114:114301, Mar 2015.

[49] Pai Wang, Ling Lu, and Katia Bertoldi. Topological phononic crystals with one-way elastic edge waves. *Physical review letters*, 115(10):104302, 2015.

[50] Ryan W Wolcott and Ryan M Eustice. Visual localization within lidar maps for automated urban driving. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 176–183. IEEE, 2014.

[51] Francesco Nex and Fabio Remondino. Uav for 3d mapping applications: a review. *Applied Geomatics*, 6(1):1–15, 2014.

[52] Adisorn Tuantranont, VM Bright, J Zhang, W Zhang, JA Neff, and YC Lee. Optical beam steering using mems-controllable microlens array. *Sensors and Actuators A: Physical*, 91(3):363–372, 2001.

[53] Alexander Wolter, Shu-Ting Hsu, Harald Schenk, and Hubert K Lakner. Applications and requirements for mems scanner mirrors. In *MOEMS-MEMS Micro & Nanofabrication*, pages 64–75. International Society for Optics and Photonics, 2005.

[54] Sean D Keller, Gerald P Uyeno, Ted Lynch, Scott R Davis, Scott D Rommel, and Juan Pino. Emerging liquid crystal waveguide technology for low swap active short-wave infrared imagers. In *SPIE OPTO*, pages 93840M–93840M. International Society for Optics and Photonics, 2015.

[55] Jihwan Kim, Chulwoo Oh, Michael J Escuti, Lance Hosting, and Steve Serati. Wide-angle, nonmechanical beam steering using thin liquid crystal polarization gratings. In *Proc. SPIE*, volume 7093, page 709302, 2008.

[56] David N Hutchison, Jie Sun, Jonathan K Doylend, Ranjeet Kumar, John Heck, Woosung Kim, Christopher T Phare, Avi Feshali, and Haisheng Rong. High-resolution aliasing-free optical beam steering. *Optica*, 3(8):887–890, 2016.

[57] Tin Komljenovic, Roger Helkey, Larry Coldren, and John E Bowers. Sparse aperiodic arrays for optical beam forming and lidar. *Optics Express*, 25(3):2511–2528, 2017.

[58] Jonathan K Doylend, MJR Heck, Jock T Bovington, Jonathan D Peters, LA Coldren, and JE Bowers. Two-dimensional free-space beam steering with an optical phased array on silicon-on-insulator. *Optics Express*, 19(22):21595–21604, 2011.

[59] JC Hulme, JK Doylend, MJR Heck, JD Peters, ML Davenport, JT Bovington, LA Coldren, and JE Bowers. Fully integrated hybrid silicon free-space beam steering source with 32-channel phased array. In *SPIE OPTO*, pages 898907–898907. International Society for Optics and Photonics, 2014.

[60] Karel Van Acoleyen, Hendrik Rogier, and Roel Baets. Two-dimensional optical phased array antenna on silicon-on-insulator. *Optics express*, 18(13):13655–13660, 2010.

[61] Weihua Guo, Pietro RA Binetti, Chad Althouse, Milan L Mašanović, Huub PMM Ambrosius, Leif A Johansson, and Larry A Coldren. Two-dimensional optical beam steering with inp-based photonic integrated circuits. *IEEE Journal of Selected Topics in Quantum Electronics*, 19(4):6100212–6100212, 2013.

[62] David Kwong, Amir Hosseini, John Covey, Yang Zhang, Xiaochuan Xu, Harish Subbaraman, and Ray T Chen. On-chip silicon optical phased array for two-dimensional beam steering. *Optics letters*, 39(4):941–944, 2014.

[63] Ami Yaacobi, Jie Sun, Michele Moresco, Gerald Leake, Douglas Coolbaugh, and Michael R Watts. Integrated phased array for wide-angle beam steering. *Optics letters*, 39(15):4575–4578, 2014.

[64] Yuen T Lo and SW Lee. *Antenna Handbook: theory, applications, and design.* Springer Science & Business Media, 2013.

[65] John Ruze. Wide-angle metal-plate optics. *Proceedings of the IRE*, 38(1):53–59, 1950.

[66] Ann Lee Peebles. A dielectric bifocal lens for multibeam antenna applications. *IEEE Transactions on Antennas and Propagation*, 36(5):599–606, 1988.

[67] Yousuke Tajima and Yoshihide Yamada. Design of shaped dielectric lens antenna for wide angle beam steering. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 89(2):1–12, 2006.

[68] Bernhard Schoenlinner, Xidong Wu, Jim P Ebling, George V Eleftheriades, and Gabriel M Rebeiz. Wide-scan spherical-lens antennas for automotive radars. *IEEE Transactions on microwave theory and techniques*, 50(9):2166–2175, 2002.

[69] P Dainesi, A Kung, M Chabloz, A Lagos, Ph Fluckiger, A Ionescu, P Fazan, M Declerq, Philippe Renaud, and Ph Robert. Cmos compatible fully integrated mach-zehnder interferometer in soi technology. *IEEE Photonics Technology Letters*, 12(6):660–662, 2000.

[70] Daniel Malacara-Hernández and Zacarías Malacara-Hernández. *Handbook of optical design.* CRC Press, 2016.

[71] Jaeheung Kim, Choon Sik Cho, and Frank S Barnes. Dielectric slab rotman lens for microwave/millimeter-wave applications. *IEEE transactions on microwave theory and techniques*, 53(8):2622–2627, 2005.

[72] Leonard T Hall, Hedley J Hansen, and Derek Abbott. 2d scanning rotman lens structure for smart collision avoidance sensors. In *Microelectronics, MEMS, and Nanotechnology*, pages 93–99. International Society for Optics and Photonics, 2004.

[73] Felix K Schwering and Song-Tsuen Peng. Design of dielectric grating antennas for millimeter-wave applications. *IEEE Transactions on Microwave Theory and Techniques*, 31(2):199–209, 1983.

[74] Young-Jin Park and Werner Wiesbeck. Offset cylindrical reflector antenna fed by a parallel-plate luneburg lens for automotive radar applications in millimeter-wave. *IEEE Transactions on Antennas and Propagation*, 51(9):2481–2483, 2003.

[75] John Sanford. A luneberg-lens update. *IEEE Antennas and Propagation Magazine*, 37(1):76–79, 1995.

[76] J Rao. Multifocal three-dimensional bootlace lenses. *IEEE Transactions on Antennas and Propagation*, 30(6):1050–1056, 1982.

[77] FG Friedlander. A dielectric-lens aerial for wide-angle beam scanning. *Journal of the Institution of Electrical Engineers-Part IIIA: Radiolocation*, 93(4):658–662, 1946.

[78] R Ulrich and RJ Martin. Geometrical optics in thin film light guides. *Applied Optics*, 10(9):2077–2085, 1971.

[79] Shogo Ura, Toshiaki Suhara, Hiroshi Nishihara, and Jiro Koyama. An integrated-optic disk pickup device. *Journal of lightwave technology*, 4(7):913–918, 1986.

[80] Michael R Watts, Jie Sun, Christopher DeRose, Douglas C Trotter, Ralph W Young, and Gregory N Nielson. Adiabatic thermo-optic mach–zehnder switch. *Optics letters*, 38(5):733–735, 2013.

[81] Christopher V Poulton, Matthew J Byrd, Manan Raval, Zhan Su, Nanxi Li, Erman Timurdogan, Douglas Coolbaugh, Diedrik Vermeulen, and Michael R Watts. Large-scale silicon nitride nanophotonic phased arrays at infrared and visible wavelengths. *Optics Letters*, 42(1):21–24, 2017.

[82] Amir Arbabi and Lynford L Goddard. Measurements of the refractive indices and thermo-optic coefficients of si 3 n 4 and sio x using microring resonances. *Optics letters*, 38(19):3878–3881, 2013.

[83] J Komma, C Schwarz, G Hofmann, D Heinert, and R Nawrodt. Thermo-optic coefficient of silicon at 1550 nm and cryogenic temperatures. *Applied Physics Letters*, 101(4):041905, 2012.

[84] Lucas H Gabrielli and Michal Lipson. Integrated luneburg lens via ultra-strong index gradient on silicon. *Optics express*, 19(21):20122–20127, 2011.

[85] Andrea Di Falco, Susanne C Kehr, and Ulf Leonhardt. Luneburg lens in silicon photonics. *Optics express*, 19(6):5156–5162, 2011.

[86] D Waidelich. The phase centers of aperture antennas. *IEEE Transactions on antennas and propagation*, 28(2):263–264, 1980.

[87] YY Hu. A method of determining phase centers and its application to electromagnetic horns. *Journal of the Franklin Institute*, 271(1):31–39, 1961.

[88] Lifu Zhou, Amit V Itagi, James A Bain, and Tuviah E Schlesinger. Characterization of refraction at a waveguide step for fabrication of mode index lenses. In *Optical Data Storage Topical Meeting 2004*, pages 697–703. International Society for Optics and Photonics, 2004.

[89] John Ruze. The effect of aperture errors on the antenna radiation pattern. *Il Nuovo Cimento (1943-1954)*, 9:364–380, 1952.

[90] Frank Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[91] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[92] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[93] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning.* MIT Press, 2016.

[94] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[95] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[96] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[97] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[98] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[99] Jürgen Schmidhuber. Deep learning in neural networks: An overview. *Neural networks*, 61:85–117, 2015.

[100] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

[101] Rodrigo Benenson, Mohamed Omran, Jan Hosang, and Bernt Schiele. Ten years of pedestrian detection, what have we learned? *arXiv preprint arXiv:1411.4304*, 2014.

[102] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118, 2017.

[103] Kun-Hsing Yu, Ce Zhang, Gerald J Berry, Russ B Altman, Christopher Ré, Daniel L Rubin, and Michael Snyder. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature Communications*, 7, 2016.

[104] Haoxiang Li, Zhe Lin, Xiaohui Shen, Jonathan Brandt, and Gang Hua. A convolutional neural network cascade for face detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5325–5334, 2015.

[105] Bree Brouwer. Youtube now sees 300 hours of video uploaded every minute. *Online unter: http://www. tubefilter. com/2014/12/01/YouTube-300-hours-video-per-minute/(22.10. 2015)*, 2014.

[106] Daniel Kang, John Emmons, Firas Abuzaid, Peter Bailis, and Matei Zaharia. Optimizing deep cnn-based queries over video streams at scale. *arXiv preprint arXiv:1703.02529*, 2017.

[107] I Cisco. Cisco visual networking index: Forecast and methodology, 2015–2020. *CISCO White paper*.

[108] Carver Mead. Neuromorphic electronic systems. *Proceedings of the IEEE*, 78(10):1629–1636, 1990.

[109] Chi-Sang Poon and Kuan Zhou. Neuromorphic silicon neurons and large-scale neural networks: challenges and opportunities. *Frontiers in neuroscience*, 5:108, 2011.

[110] Mark Horowitz. 1.1 computing's energy problem (and what we can do about it). In *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2014 IEEE International*, pages 10–14. IEEE, 2014.

[111] Ali Shafiee, Anirban Nag, Naveen Muralimanohar, Rajeev Balasubramonian, John Paul Strachan, Miao Hu, R Stanley Williams, and Vivek Srikumar. Isaac: A convolutional neural network accelerator with in-situ analog arithmetic in crossbars. In *Proceedings of the 43rd International Symposium on Computer Architecture*, pages 14–26. IEEE Press, 2016.

[112] Janardan Misra and Indranil Saha. Artificial neural networks in hardware: A survey of two decades of progress. *Neurocomputing*, 74(1):239–255, 2010.

[113] Paul Merolla, John Arthur, Filipp Akopyan, Nabil Imam, Rajit Manohar, and Dharmendra S Modha. A digital neurosynaptic core using embedded crossbar memory with 45pj per spike in 45nm. In *Custom Integrated Circuits Conference (CICC), 2011 IEEE*, pages 1–4. IEEE, 2011.

[114] Yu-Hsin Chen, Tushar Krishna, Joel S Emer, and Vivienne Sze. Eyeriss: An energy-efficient reconfigurable accelerator for deep convolutional neural networks. *IEEE Journal of Solid-State Circuits*, 2016.

[115] Nicholas C Harris, Gregory R Steinbrecher, Jacob Mower, Yoav Lahini, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Seth Lloyd, and Dirk Englund. Bosonic transport simulations in a large-scale programmable nanophotonic processor. *arXiv preprint arXiv:1507.03406*, 2015.

[116] Yichen Shen, Nicholas C Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, et al. Deep learning with coherent nanophotonic circuits. *arXiv preprint arXiv:1610.02365*, 2016.

[117] Kengo Nozaki, Takasumi Tanabe, Akihiko Shinya, Shinji Matsuo, Tomonari Sato, Hideaki Taniyama, and Masaya Notomi. Sub-femtojoule all-optical switching using a photonic-crystal nanocavity. *Nature Photonics*, 4(7):477–483, 2010.

[118] Carlos Ríos, Matthias Stegmaier, Peiman Hosseini, Di Wang, Torsten Scherer, C David Wright, Harish Bhaskaran, and Wolfram HP Pernice. Integrated all-photonic non-volatile multi-level memory. *Nature Photonics*, 9(11):725–732, 2015.

[119] Chen Sun, Mark T Wade, Yunsup Lee, Jason S Orcutt, Luca Alloatti, Michael S Georgas, Andrew S Waterman, Jeffrey M Shainline, Rimas R Avizienis, Sen Lin, et al. Single-chip microprocessor that communicates directly using light. *Nature*, 528(7583):534–538, 2015.

[120] Takasumi Tanabe, Masaya Notomi, Satoshi Mitsugi, Akihiko Shinya, and Eiichi Kuramochi. Fast bistable all-optical switch and memory on a silicon photonic crystal on-chip. *Optics letters*, 30(19):2575–2577, 2005.

[121] Alexander N Tait, Mitchell A Nahmias, Yue Tian, Bhavin J Shastri, and Paul R Prucnal. Photonic neuromorphic signal processing and computing. In *Nanophotonic Information Physics*, pages 183–222. Springer, 2014.

[122] Alexander N Tait, Mitchell A Nahmias, Bhavin J Shastri, and Paul R Prucnal. Broadcast and weight: an integrated network for scalable photonic spike processing. *Journal of Lightwave Technology*, 32(21):3427–3439, 2014.

[123] Paul R Prucnal, Bhavin J Shastri, Thomas Ferreira de Lima, Mitchell A Nahmias, and Alexander N Tait. Recent progress in semiconductor excitable lasers for photonic spike processing. *Advances in Optics and Photonics*, 8(2):228–299, 2016.

[124] Kristof Vandoorne, Pauline Mechet, Thomas Van Vaerenbergh, Martin Fiers, Geert Morthier, David Verstraeten, Benjamin Schrauwen, Joni Dambre, and Peter Bienstman. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nature communications*, 5, 2014.

[125] Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. cudnn: Efficient primitives for deep learning. *arXiv preprint arXiv:1410.0759*, 2014.

[126] Jost Tobias Springenberg, Alexey Dosovitskiy, Thomas Brox, and Martin Riedmiller. Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*, 2014.

[127] Michael Reck, Anton Zeilinger, Herbert J. Bernstein, and Philip Bertani. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.*, 73:58–61, Jul 1994.

[128] William R. Clements, Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley. An optimal design for universal multiport interferometers, 2016.

[129] Zhenzhou Cheng, Hon Ki Tsang, Xiaomu Wang, Ke Xu, and Jian-Bin Xu. In-plane optical absorption and free carrier absorption in graphene-on-silicon waveguides. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(1):43–48, 2014.

[130] Marin Soljačić, Mihai Ibanescu, Steven G Johnson, Yoel Fink, and JD Joannopoulos. Optimal bistable switching in nonlinear photonic crystals. *Physical Review E*, 66(5):055601, 2002.

[131] Robert W Schirmer and Alexander L Gaeta. Nonlinear mirror based on two-photon absorption. *JOSA B*, 14(11):2865–2868, 1997.

[132] Qiaoliang Bao, Han Zhang, Zhenhua Ni, Yu Wang, Lakshminarayana Polavarapu, Zexiang Shen, Qing-Hua Xu, Dingyuan Tang, and Kian Ping Loh. Monolayer graphene as a saturable absorber in a mode-locked laser. *Nano Research*, 4(3):297–307, 2011.

[133] AC Selden. Pulse transmission through a saturable absorber. *British Journal of Applied Physics*, 18(6):743, 1967.

[134] G Lenz, BJ Eggleton, C Kd Madsen, and RE Slusher. Optical delay lines based on optical filters. *IEEE Journal of Quantum Electronics*, 37(4):525–532, 2001.

[135] Takasumi Tanabe, Masaya Notomi, Eiichi Kuramochi, Akihiko Shinya, and Hideaki Taniyama. Trapping and delaying photons for one nanosecond in an ultrasmall high-q photonic-crystal nanocavity. *Nature Photonics*, 1(1):49–52, 2007.

[136] Fengnian Xia, Lidija Sekaric, and Yurii Vlasov. Ultracompact optical buffers on a silicon chip. *Nature photonics*, 1(1):65–71, 2007.

[137] Mahmoud S Rasras, Christi K Madsen, Mark A Cappuzzo, Evans Chen, Lou T Gomez, Edward J Laskowski, Andrew Griffin, Annjoe Wong-Foy, Arman Gasparyan, Albin Kasper, et al. Integrated resonance-enhanced variable optical delay lines. *IEEE photonics technology letters*, 17(4):834–836, 2005.

[138] Laurent Vivien, Andreas Polzer, Delphine Marris-Morini, Johann Osmond, Jean Michel Hartmann, Paul Crozat, Eric Cassan, Christophe Kopp, Horst Zimmermann, and Jean Marc Fédéli. Zero-bias 40gbit/s germanium waveguide photodetector on silicon. *Optics express*, 20(2):1096–1101, 2012.

[139] David AB Miller. Perfect optics with imperfect components. *Optica*, 2(8):747–750, 2015.

[140] Scott A Skirlo, Ling Lu, and Marin Soljačić. Binary matrices of optimal autocorrelations as alignment marks. *Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena*, 33(2):021601, 2015.

[141] RH Barker. Group synchronizing of binary digital systems. *Communication theory*, pages 273–287, 1953.

[142] RH Neuman and L Hofman. Group synchronizing of binary digital systems. *IEEE Aero. El. Sys. Mag.*, page 570, 1971.

[143] Solomon W Golomb and Guang Gong. *Signal design for good correlation: for wireless communication, cryptography, and radar.* Cambridge University Press, 2005.

[144] Stephen R Gottesman and EE Fenimore. New family of binary arrays for coded aperture imaging. *Applied optics*, 28(20):4344–4352, 1989.

[145] Sabah Alquaddoomi and Robert A Scholtz. On the nonexistence of barker arrays and related matters. *IEEE transactions on information theory*, 35(5):1048–1057, 1989.

[146] John P Costas. A study of a class of detection waveforms having nearly ideal range-doppler ambiguity properties. *Proceedings of the IEEE*, 72(8):996–1009, 1984.

[147] Solomon Golomb and Herbert Taylor. Two-dimensional synchronization patterns for minimum ambiguity. *IEEE Transactions on Information Theory*, 28(4):600–604, 1982.

[148] Gummadi S Ramakrishna and Wai Ho Mow. A new search for optimal binary arrays with minimum peak sidelobe levels. In *International Conference on Sequences and Their Applications*, pages 355–360. Springer, 2004.

[149] EH Anderson, D Ha, and JA Liddle. Sub-pixel alignment for direct-write electron beam lithography. *Microelectronic engineering*, 73:74–79, 2004.

[150] V Boegli and DP Kern. Automatic mark detection in electron beam nanolithography using digital image processing and correlation. *Journal of Vacuum Science & Technology B: Microelectronics Processing and Phenomena*, 8(6):1994–2001, 1990.

[151] Minghao Qi, Elefterios Lidorikis, Peter T Rakich, Steven G Johnson, JD Joannopoulos, Erich P Ippen, and Henry I Smith. A three-dimensional optical photonic crystal with designed point defects. *Nature*, 429(6991):538–542, 2004.

[152] Ling Lu. *Photonic crystal nanocavity lasers for integration.* 2010.

[153] John P. Robinson. Golomb rectangles as folded rulers. *IEEE Transactions on Information Theory*, 43(1):290–293, 1997.

[154] James B Shearer. Symmetric golomb squares. *IEEE transactions on information theory*, 50(8):1846–1847, 2004.

[155] Hiroshi Sakou, Takafumi Miyatake, Seiji Kashioka, and Masakazu Ejiri. A position recognition algorithm for semiconductor alignment based on structural pattern matching. *IEEE Transactions on Acoustics, Speech, and Signal processing*, 37(12):2148–2157, 1989.

[156] R Scholtz. Frame synchronization techniques. *IEEE Transactions on Communications*, 28(8):1204–1213, 1980.

[157] Andrew Z Tirkel, Charles F Osborne, and Thomas E Hall. Image and watermark registration. *Signal processing*, 66(3):373–383, 1998.

[158] Herbert G Winful. Pulse compression in optical fiber filters. *Applied Physics Letters*, 46(6):527–529, 1985.

[159] C Martijn de Sterke. Optical push broom. *Optics letters*, 17(13):914–916, 1992.

[160] MJ Steel and C Martijn de Sterke. Schrödinger equation description for cross-phase modulation in grating structures. *Physical Review A*, 49(6):5048, 1994.

[161] Neil GR Broderick, Domino Taverner, David J Richardson, and Morten Ibsen. Cross-phase modulation effects in nonlinear fiber bragg gratings. *JOSA B*, 17(3):345–353, 2000.

[162] NGR Broderick, D Taverner, DJ Richardson, M Ibsen, and RI Laming. Experimental observation of nonlinear pulse compression in nonuniform bragg gratings. *Optics letters*, 22(24):1837–1839, 1997.

[163] NGR Broderick, D Taverner, DJ Richardson, M Ibsen, and RI Laming. Optical pulse compression in fiber bragg gratings. *Physical review letters*, 79(23):4566, 1997.

[164] Falk Eilenberger, Irina V Kabakova, C Martijn de Sterke, Benjamin J Eggleton, and Thomas Pertsch. Cavity optical pulse extraction: ultra-short pulse generation as seeded hawking radiation. *Scientific reports*, 3, 2013.

[165] Rose M Joseph, Susan C Hagness, and Allen Taflove. Direct time integration of maxwell's equations in linear dispersive media with absorption for scattering and propagation of femtosecond electromagnetic pulses. *Optics Letters*, 16(18):1412–1414, 1991.

[166] Thomas G Philbin, Chris Kuklewicz, Scott Robertson, Stephen Hill, Friedrich König, and Ulf Leonhardt. Fiber-optical analog of the event horizon. *Science*, 319(5868):1367–1370, 2008.

[167] Govind Agrawal. Nonlinear fiber optics, (optics and photonics). 2006.

[168] X Tr Tran and NN Rosanov. Conservative and dissipative fiber bragg solitons (a review). *Optics and Spectroscopy*, 105(3):393–435, 2008.

[169] H Kogelnik and CV Shank. Coupled-wave theory of distributed feedback lasers. *Journal of applied physics*, 43(5):2327–2335, 1972.

[170] Herbert G Winful and Gene D Cooperman. Self-pulsing and chaos in distributed feedback bistable optical devices. *Applied Physics Letters*, 40(4):298–300, 1982.