

ADJOINT-BASED OPTIMIZATION AND INVERSE DESIGN OF PHOTONIC  
DEVICES

A DISSERTATION  
SUBMITTED TO THE DEPARTMENT OF APPLIED PHYSICS  
AND THE COMMITTEE ON GRADUATE STUDIES  
OF STANFORD UNIVERSITY  
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS  
FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY

Tyler William Hughes  
June 2019

© Copyright by Tyler William Hughes 2019  
All Rights Reserved

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

(Shanhui Fan) Principal Adviser

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

(Robert L. Byer)

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

---

(Olav Solgaard)

Approved for the Stanford University Committee on Graduate Studies

# Preface

Photonic devices are important to a wide range of applications, from communications to fundamental science. However, the design of such devices is traditionally done by hand, using physical intuition and trial and error. The development of ‘inverse design’, where computational optimization techniques are used to design devices based on certain specifications, has led to the discovery of many compact, non-intuitive structures with superior performance. Among various methods, large-scale, gradient-based optimization techniques have been one of the most important ways to design a structure containing a vast number of degrees of freedom. These techniques are made possible by the *adjoint method*, in which the gradient of an objective function with respect to all design degrees of freedom can be computed using only two full-field simulations.

In this thesis, we will discuss the application of inverse design to two emerging photonic technologies and discuss the generalization of the adjoint method to new scenarios. First, we will present the inverse design of laser-driven particle accelerators on a chip as well as our efforts to scale this technology to higher energy gains using photonic integrated circuits. Next, we will discuss how the adjoint method may be used to perform on-chip training of optical neural networks. We will show that this technique corresponds to a physical implementation of the backpropagation algorithm, commonly used in traditional neural networks. A procedure for measuring the gradient determined by the adjoint method will be introduced. Then, we will discuss the application of this technique to design general wave systems capable of performing machine learning computation on sequence data in the form of time series signals. Finally, we will discuss the generalization of the adjoint method to nonlinear optical phenomena and show that this may be used to devise compact photonic switches in a Kerr nonlinear material.

# Acknowledgments

Thanks to

# Contents

<b>Preface</b>	<b>iv</b>
<b>Acknowledgments</b>	<b>v</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Photonic Device Design . . . . .	2
1.1.1 Traditional Design Approach . . . . .	2
1.1.2 Inverse Design Approach . . . . .	2
1.2 Adjoint Method . . . . .	3
1.3 Thesis Overview . . . . .	4
<b>2 Adjoint-Based Optimization of Laser-Driven Particle Accelerators</b>	<b>5</b>
2.1 Dielectric Laser Acceleration . . . . .	5
2.2 Adjoint Method for Accelerator . . . . .	7
2.2.1 Mathematical Definition . . . . .	7
2.2.2 Adjoint Formalism . . . . .	10
2.2.3 Interpretation of the Adjoint Field as Radiation . . . . .	12
2.2.4 Application to Design Parameters . . . . .	13
2.3 Inverse Design of Accelerator . . . . .	13
2.3.1 Maximization of Acceleration Gradient . . . . .	13
2.3.2 Maximization of Acceleration Factor . . . . .	16
2.4 Conclusions . . . . .	19
<b>3 Integrated Photonic Circuit for Accelerators on a Chip</b>	<b>22</b>
3.1 On-Chip Laser Coupling Device . . . . .	22
3.1.1 Device Components . . . . .	24
3.1.2 Practical Considerations . . . . .	35
3.1.3 Parameter Study . . . . .	39
3.2 Reconfigurable Control System for Accelerator . . . . .	48

3.2.1	DLA System with Power Equalization . . . . .	49
3.2.2	Power Distribution Protocol . . . . .	53
3.2.3	Numerical Demonstration . . . . .	56
3.2.4	Performance Improvement . . . . .	57
3.3	Outlook and Conclusions . . . . .	63
<b>4</b>	<b>Training of Optical Neural Networks</b>	<b>64</b>
4.1	Optical Neural Networks . . . . .	65
4.2	Training of Optical Hardware . . . . .	69
4.3	Experimental Measurement of Adjoint Gradient . . . . .	71
4.3.1	Numerical Demonstrations . . . . .	76
4.3.2	Discussion . . . . .	81
4.4	Conclusion . . . . .	82
<b>5</b>	<b>Wave-Based Analog Recurrent Neural Networks</b>	<b>83</b>
5.1	Wave Physics as a Recurrent Neural Network . . . . .	83
5.2	Vowel Recognition Demonstration . . . . .	87
5.3	Comparison of Analog and Digital Implementations . . . . .	93
<b>6</b>	<b>Extension of Adjoint Method to Nonlinear Systems</b>	<b>95</b>
6.1	Mathematical Formulation . . . . .	96
6.2	Inverse Design of Nonlinear Switches . . . . .	100
6.3	Summary and Outlook . . . . .	106
<b>7</b>	<b>Conclusion and Final Remarks</b>	<b>108</b>
<b>A</b>	<b>Backpropagation for Nonholomorphic Activations</b>	<b>110</b>
<b>B</b>	<b>Simulation of Photonic Neural Networks</b>	<b>113</b>

# List of Tables

2.1	Acceleration factor ( $f_A$ ) before and after maximization. . . . .	19
3.1	Parameters assumed in the study. . . . .	40
3.2	Optimal results from the parameter study, for waveguides fabricated from SOI and SiN. .	44
3.3	Set of parameters used in the analysis of Fig. 3.15 . . . . .	61
5.1	Comparison of scalar wave model and conventional RNN on vowel recognition task. .	93
6.1	Parameters used in the optimization study. Column ‘2-port’ refers to the device from Fig. 2. Column ‘3-port’ refers to the device from Fig. 3 . . . . .	107

# List of Figures

2.1	Diagram outlining the system setup for side-coupled DLA with an arbitrary dielectric structure $\epsilon(x, y)$ (green). A charged particle moves through the vacuum gap with speed $\beta c_0$ . The periodicity is set at $\beta\lambda$ where $\lambda$ is the central wavelength of the laser pulse. . . . .	6
2.2	Definition of the vector field, $\vec{\eta}$ , which defines the position of the electron in the frequency domain. The green regions represent domains where we will optimize the material properties using the adjoint method. The central gap is constrained to vacuum to allow passage of the electron beam. The red arrow signifies the driving laser. . . . .	9
2.3	Demonstration of adjoint method in calculating accelerator sensitivities. (a) The acceleration gradient ( $G$ ) of a square accelerator structure (inset) as a function of the square's relative permittivity. We express the acceleration gradient in its dimensionless form, normalized by the electric field amplitude of the incident plane wave ( $E_0$ ). The particle traverses along the dotted line with a velocity of $c_0$ ( $\beta = 1$ ) and a plane wave is incident from the bottom of the structure. (b) The sensitivity $\frac{dG}{d\epsilon}$ of the gradient with respect to changing the square relative permittivity for direct central difference (solid line) $\frac{dG}{d\epsilon} = \frac{G(\epsilon+\Delta\epsilon)-G(\epsilon-\Delta\epsilon)}{2\Delta\epsilon}$ and using the adjoint method (circles). The two calculations agree with excellent precision. The dotted line at $\frac{dG}{d\epsilon} = 0$ , corresponds to local minima and maxima of $G(\epsilon)$ above. . . . .	14
2.4	Optimization routine for DLA. . . . .	15

2.5 Demonstration of the structure optimization for $\beta = 0.5$ , laser wavelength $\lambda = 2 \mu\text{m}$ , and a gap size of 400 nm. A plane wave is incident from the bottom in all cases. (a) Acceleration gradient as a function of iteration number for different maximum relative permittivity values, corresponding to those of Si, $\text{Si}_3\text{N}_4$ , and $\text{SiO}_2$ at the laser wavelength. The acceleration gradient is normalized by the electric field amplitude of the incident plane wave ( $E_0$ ). The optimizations converge after about five-hundred iterations. (b-d) Final structure permittivity distributions (white = vacuum, black = $\epsilon_m$ ) corresponding to the three curves in (a). Eight periods are shown, corresponding to four laser wavelengths. For each (b-d), design region widths on each side of the particle gap were given by 1, 2, and 4 $\mu\text{m}$ for Si, $\text{Si}_3\text{N}_4$ , and $\text{SiO}_2$ , respectively. . . . .	17
2.6 Similarity between DLA structures designed using adjoint method (left) and those independently proposed using human intuition. . . . .	18
2.7 Demonstration of the final structures after optimization for (a) maximizing gradient only, (b) maximizing the acceleration factor. $\beta = 0.5$ , laser wavelength $\lambda = 2 \mu\text{m}$ , gap size of 400 nm. $\epsilon_m = 2.1$ , corresponding to $\text{SiO}_2$ . In (a), the high gradients are achieved using reflective dielectric mirrors to confine and enhance the fields in the center region. In (b), these dielectric mirrors are removed and the pillar structures are augmented. The structure in (b) shows a 23% increase in the acceleration factor in the material region when compared to (a). . . . .	20
3.1 Two stages of the DLA laser coupling ‘tree-network’ structure. The electron beam travels along the z-axis through the center of this structure. The laser pulses are side coupled with optical power shown in red. Black regions define the on-chip waveguide network. Blue circles represent the optical phase shifters used to tune the phase of the laser pulse. This geometry serves to reproduce the pulse-front-tilt laser delivery system outlined in [33] in an integrated optics platform. . . . .	23
3.2 Waveguide geometries and corresponding horizontal electric field components [56]. (a-b) Strongly confined modes. (c-d) Weakly confined modes. (a) and (c) are SOI material platforms whereas (b) and (d) are $\text{Si}_3\text{N}_4/\text{SiO}_2$ materials. Waveguide core heights in (a-d) are given by 220, 400, 60, and 100 nm, respectively. Waveguide core widths are given by 0.78, 1.6, 2, and 4 $\mu\text{m}$ , respectively. . . . .	26
3.3 Diagram of a single bend in the tree-network structure with an optical pulse incident from the left. The bend has radius $R$ , accomplishes a vertical climb of $h$ over a horizontal distance $d$ . The total length of the bent section is $L$ . The electron travels from bottom to top in this configuration. We wish to find an $R$ such that an optical pulse traveling through the bent section is delayed by the same amount of time for the electron to travel the vertical distance $h$ . . . . .	29

3.4	(a) Electric field amplitude for a strongly guiding SOI waveguide. (b) Electric field amplitude for a weakly guiding SOI waveguide. (c) Comparison of bending loss as a function of bend radius for the 4 waveguides from Fig. 3.2. . . . .	30
3.5	Idealized schematic of a feedback system for automatic phase control. A dedicated light extraction section is added to the accelerator. Light is radiated from the electron beam transversing the DLA structures and the frequency content and/or timing of the light is sent to a controller. The phase shifts of each waveguide are optimized with respect to either the frequency or the delay of the signal. . . . .	32
3.6	(a) A schematic of the waveguide to DLA connection. Silicon dual pillars of optimized radius of 981 nm and gap size of 400 nm are used. (b) The accelerating electric field during one time step. (c) Absolute value of the transverse magnetic field. (d) Absolute value of the acceleration gradient as a function of frequency, normalized by the peak electric field in the waveguide. A Lorentzian line shape is fit to the square of this plot. The square root of this fit is shown in red. Based on the Lorentzian fit, a Q-factor of $152 \pm 29$ was determined. As computed following the derivation in [152], but with the waveguide mode impedance and effective area in place of the plane wave values, this structure has a shunt impedance, $Z_S$ , of 449.1 $\Omega$ over 3 periods and a $Z_S/Q$ value of 2.95 $\Omega$ . . . . .	33
3.7	Results from the parameter study. A single stage of the tree-network structure is simulated, with stage length of 192 $\mu\text{m}$ , corresponding to 5 power splits and $2^5 = 32$ output ports. In (a-b), Silicon-on-Insulator (SOI) waveguides are assumed. In (c-d), $\text{Si}_3\text{N}_4/\text{SiO}_2$ waveguides are assumed. For each Q-factor and pulse duration, we compute the maximum input field achievable before damage or nonlinearity occurs. The different colored regimes in (a) and (c) correspond to different limiting constraints as labeled in the plots. The dotted line corresponds to the minimum pulse duration before the pulse bandwidth exceeds the DLA resonator bandwidth. The energy gain from one section is plotted in (b) and (d). . . . .	43
3.8	Scaling of optimal parameters as a function of the stage length. (a) The optimal energy gains and acceleration gradients as a function of stage length for both SOI and SiN structures. (b) The optimal set of pulse duration and Q-factor corresponding to the highest energy gain and acceleration gradient at each stage length. The curves for SOI and SiN are overlaid. (c) The number of stages required to reach 1 MeV of total energy gain as a function of individual stage length. . . . .	45

3.9 Schematic of a hybrid structure for DLA laser coupling. Center: an SOI tree-network / DLA geometry optimized for tight bends and compact waveguides. This is fed by a  $\text{Si}_3\text{N}_4/\text{SiO}_2$  waveguide section with relatively higher damage threshold, and lower nonlinearities. This section is then fed by an all  $\text{SiO}_2$  power delivery section as described in the discussion section. Coarse and fine phase shifters are used in different splitting sections. . . . .

46

3.10 **Schematic of the proposed DLA laser delivery control system.** Comparison of setup from Ref [84] (**a**) and this work (**b**). **a**, In the splitting structure of Ref. [84], an input pulse (large red arrow) is focused to a single grating coupler on the chip surface. The pulse is split a series of times and bends in the waveguide to create group velocity delay. The phase of the pulses are corrected before injection into the accelerator. **b**, In the schematic proposed in this work, the pulse is directly coupled into several waveguides through a grating coupler array. To mitigate the large variance in coupled powers, the pulses next enter a mesh of MZIs, which is sequentially adjusted, using the protocol from this paper, to provide uniform powers in each waveguide. As before, phase control and group delay are performed using integrated phase shifters and lithographically-defined bends before coupling to the accelerator channel. The second scheme eliminates the damage and nonlinearity bottleneck present in **a** directly after the input coupler. . . . .

50

3.11 **Diagram of MZI mesh for power distribution.** **a**, Diagram of a single MZI consisting of two input ports and two output ports. Where the two arms come together, a 50-50 beam-splitter operation is performed. The blue regions indicate tunable optical phase shifters with added phases marked as  $\theta$  and  $\phi$ . **b**, The MZI represents a tunable unitary transformation on its inputs  $[u_1, u_2]^T$  to give outputs  $[v_1, v_2]^T$ . **c**, Simplified diagram representing an MZI in the following figures. One may think of this element as a tunable power switch. **d**, Individual MZIs are combined into meshes, which may implement tunable unitary operations on several inputs. Shown is a 'Clements' Mesh geometry, which was used in this work. **e**, Schematic of the Clements mesh in **d** using the simple MZI diagram in **c**. . . . .

51

3.12 Sequential algorithm for tuning single MZI.	The red square outlines the MZI being tuned. Black arrows on the left indicate input power to the current layer, black arrows on the right indicate desired output powers (uniform in this case). The dotted line separates the mesh into powers above and below the MZI. <b>a</b> When there is more power needed in the ports above the MZI than are supplied in the ports leading into the MZI and above, the MZI should output all of its power to its top output port. <b>b</b> When there is more power needed in the ports below the MZI than are supplied in the ports leading into the MZI and below, the MZI should send all of its power to the bottom output port. <b>c</b> In the intermediate case, the MZI should output just enough power to its top and bottom output ports to match the target power requirements. . . . .	54
3.13 Demonstration of mesh optimization for equal power distribution.	<b>a-f</b> The powers in each port after optimizing the mesh for uniform output power given several random input powers. The colorbar represents the fractional power in each port, where the total power is normalized to 1. The vertical axis represents the input port index and the horizontal axis represents the layer index. Power flows from left to right. Equalization is achieved, on average, after only around 5 layers given this network with 30 ports. . . . .	56
3.14 Analysis of mesh performance vs. number of input ports.	We plot the mean-squared-error between the power in each layer and the target power after the mesh has been optimized. These are averaged over 5 random input values. We see that the number of layers ( $M$ ) needed to equalize the power in the mesh increases slowly with respect to the number of ports ( $N$ ). . . . .	58
3.15 Figure of merit scaling for different laser coupling architectures.	Solid blue lines refer to the results using the system of Ref [84], in which all optical power is initially coupled at a single input facet. Red dotted lines refer to the structure from this work. <b>a</b> The acceleration gradient as a function of electron energy gain. For the splitting structure, the acceleration gradient diminishes rapidly as the length of acceleration is increased to match the desired energy gain. However, the structure from this work achieves uniform gradient in principle. <b>b</b> The number of output waveguides ( $N$ ) required to achieve an given energy gain of $\Delta_E$ . <b>c</b> The acceleration length ( $L$ ) required to achieve an given energy gain of $\Delta_E$ . . . . .	62

- 4.1 (a) A schematic of the ANN architecture demonstrated in Ref [170]. The boxed regions correspond to OIUs that perform a linear operation represented by the matrix  $\hat{W}_l$ . Integrated phase shifters (blue) are used to control the OIU and train the network. The red regions correspond to nonlinear activations  $f_l(\cdot)$ . (b) Illustration of operation and gradient computation in an ANN. The top and bottom rows correspond to the forward and backward propagation steps, respectively. Propagation through a square cell corresponds to matrix multiplication. Propagation through a rounded region corresponds to activation.  $\odot$  is element-wise vector multiplication. . . . . 67
- 4.2 (a) Block diagram of a feedforward neural network of  $L$  layers. Each layer consists of a  $\hat{W}_i$  block representing a linear matrix which multiplies vector inputs  $x_{i-1}$ . The  $f_i$  block in each layer represents an element-wise nonlinear activation function operating on vectors  $z_i$  to produce outputs  $x_i$ . (b) Schematic of the optical interferometer mesh implementation of a single layer of the feedforward neural network. (c) Schematic of the proposed optical-to-optical activation function which achieves a nonlinear response by converting a small portion of the optical input,  $z$  into an electrical signal, and then intensity modulating the remaining portion of the original optical signal as it passes through an interferometer. . . . . 70
- 4.3 Schematic illustration of our proposed method for experimental measurement of gradient information. The box region represents the OIU. The colored ovals represent tunable phase shifters, and we illustrate computing the gradient with respect to the red and the yellow phase shifters, labeled 1 and 2, respectively. (a): We send the original set of amplitudes  $\mathbf{X}_{l-1}$  and measure the constant intensity terms at each phase shifter. (b): We send the adjoint mode amplitudes, given by  $\delta_l$ , through the output side of our device, recording  $\mathbf{X}_{TR}^*$  from the opposite side, as well as  $|\mathbf{e}_{aj}|^2$  in each phase-shifter. (c): We send in  $\mathbf{X}_{l-1} + \mathbf{X}_{TR}$ , interfering  $\mathbf{e}_{og}$  and  $\mathbf{e}_{aj}^*$  inside the device and recovering the gradient information for all phase shifters simultaneously. . . . . 74

4.4	Numerical demonstration of the time-reversal procedure. (a): Relative permittivity distribution for three MZIs arranged to perform a 3x3 linear operation. Blue boxes represent where phase shifters would be placed in this system. As an example, we compute the gradient information for a layer with $\mathbf{X}_{l-1} = [0 \ 0 \ 1]^T$ and $\boldsymbol{\delta}_l = [0 \ 1 \ 0]^T$ , corresponding to the bottom left and middle right port, respectively. (b): Real part of the simulated electric field $E_z$ corresponding to injection from the bottom left port. (c): Real part of the adjoint $E_z$ , corresponding to injection from the middle right port. (d): Time-reversed adjoint field as constructed by our method, fed in through all three ports on the left. (e): The gradient information $\frac{d\mathcal{L}}{d\epsilon_l}(x, y)$ as obtained directly by the adjoint method, normalized by its maximum absolute value. (f): The gradient information as obtained by the method introduced in this work, normalized by its maximum absolute value. Namely, the field pattern from (b) is interfered with the time-reversed adjoint field of (d) and the constant intensity terms are subtracted from the resulting intensity pattern. Panels (e) and (f) match with high precision. . . . .	75
4.5	Numerical demonstration of a photonic ANN implementing an XOR gate using the backpropagation algorithm and adjoint method described in this work. (a) The architecture of the ANN. Two layers of $3 \times 3$ OIUs with $z^2$ activations. (b) The mean-squared error (MSE) between the predictions and targets as a function of training iterations. (c) The absolute value of the network predictions (blue circles) and targets (black crosses) before training. (d) The absolute value of the network predictions after training, showing that the network has successfully learned the XOR function.	78
4.6	Numerical demonstration of a photonic ANN learning to classify an oblong ring. (a) The architecture of the ANN. Six layers of $4 \times 4$ OIUs with $ z $ activations. A final softmax activation is applied at the very end. (b) The loss function of Eq. (4.28) over training iterations. (c) The training examples, blue and red dots correspond to $y = 0$ and $y = 1$ labels on a given $x_1$ and $x_2$ input. The background shows the prediction of the network on a continuum of $x_1$ and $x_2$ pairs, with colors representing the corresponding predictions. One can see that the ring was learned successfully without overfitting. . . . .	80

5.1	<b>Conceptual comparison of a standard recurrent neural network and a wave-based physical system.</b> (A) Diagram of a recurrent neural network (RNN) cell operating on a discrete input sequence and producing a discrete output sequence. (B) Internal components of the RNN cell, consisting of trainable dense matrices $\mathbf{W}^{(h)}$ , $\mathbf{W}^{(x)}$ , and $\mathbf{W}^{(y)}$ . Activation functions for the hidden state and output are represented by $\sigma^{(h)}$ and $\sigma^{(y)}$ , respectively. (C) Diagram of the directed graph of the RNN cell. (D) Diagram of a recurrent representation of a continuous physical system operating on a continuous input sequence and producing a continuous output sequence. (E) Internal components of the recurrence relation for the wave equation when discretized using finite differences. (F) Diagram of the directed graph of discrete time steps of the continuous physical system. . . . .	85
5.2	<b>Schematic of the vowel recognition system and the training procedure.</b> (A) Raw audio waveforms of spoken vowel samples from three classes. (B) Layout of the vowel recognition system. Vowel samples are independently injected at the source, located at the left of the domain, and propagate through the center region, indicated in green, where a material distribution is optimized during training. The dark gray region represents an absorbing boundary layer. (C) For classification, the time-integrated power at each probe is measured and normalized to be interpreted as a probability distribution over the vowel classes. (D) Using automatic differentiation, the gradient of the loss function with respect to the density of material in the green region is computed. The material density is updated iteratively, using gradient-based stochastic optimization techniques, until convergence. . . . .	88
5.3	Mean energy spectrum for the <i>ae</i> , <i>ei</i> , and <i>iy</i> vowel classes. . . . .	91
5.4	<b>Vowel recognition system training results.</b> Confusion matrix over the training and testing datasets for the initial structure (A),(B) and final structure (C),(D), indicating the percentage of correct (diagonal) and incorrect (off-diagonal). Cross validated training results showing the mean (solid line) and standard deviation (shaded region) of the (E) cross entropy loss and (F) prediction accuracy over 20 training epochs and 5 folds of the dataset, which consists of a total of 272 total vowel samples of male and female speakers. (G)-(I) The time-integrated intensity distribution for a randomly selected input (G) <i>ae</i> vowel, (H) <i>ei</i> vowel, and (I) <i>iy</i> vowel. . . . .	92

6.1	Illustration of the adjoint field computation for a linear and a nonlinear system. (a) The linear system driven by a point source $\mathbf{b}$ with an objective function $\mathcal{L}$ given by the field intensity at a measuring point. (b) The adjoint problem for the linear system: the same system driven by a point source given by $-\partial\mathcal{L}/\partial\mathbf{e}$ located at the measuring point. (c) The nonlinear system containing a medium with Kerr nonlinearity (red). The electric fields are the solution to a nonlinear equation. (d) The adjoint problem for the nonlinear system, which is a <i>linear</i> system of equations for the adjoint field and its complex conjugate. The Kerr medium is replaced by a linear region whose permittivity depends on the nonlinear fields. . . . .	99
6.2	<b>Inverse design demonstration of a <math>1 \rightarrow 1</math> port switch.</b> (a) Optical power is input into the left port (purple arrow). The goal of optimizing the design region (blue square) is to maximize power transmission in the linear regime (blue arrow) and minimize transmission in the nonlinear regime (red arrow). The final permittivity distribution after optimizing is also shown. The black regions are chalcogenide with a relative permittivity of 5.95 and a $\chi^{(3)}$ of $4.1 \times 10^{-19} \text{ m}^2/\text{V}^2$ . The waveguide regions outside the design region have a width of $0.3\mu\text{m}$ . The operating wavelength is $2\mu\text{m}$ . (b) The transmission as a function of input power, demonstrating the switching behavior at around $10^{-3} \text{ W}/\mu\text{m}$ . The dashed black line indicates the input power used for the high-power regime in the optimization. (c-d) The amplitude of the simulated electric field of the final structure, in the linear (c) and nonlinear (d) regimes, respectively – with input power of $10^{-9} \text{ W}/\mu\text{m}$ and $0.157 \text{ W}/\mu\text{m}$ , respectively. $E_z$ corresponds to the out-of-plane electric field in the 2D simulation. . . . .	101
6.3	Filtering and projection of an example design density, $\rho$ . (top left) the original density $\rho$ before processing. (top right) the density after applying a low pass filter, $\tilde{\rho}$ . (bottom left) the density $\bar{\rho}$ after applying projection. (bottom right) the final relative permittivity distribution $\epsilon_r$ . The parameters used are $R = 200\text{nm}$ , $\beta = 100$ , $\eta = 0.5$ . . . . .	103
6.4	<b>Inverse design demonstration of a <math>1 \rightarrow 2</math> port switch.</b> (a) Optical power is input into the left port (purple arrow). The goal of optimizing the design region (blue square) is to maximize the power transmission to the right port (blue arrow) in the linear regime and maximize transmission to the bottom port (red arrow) in the nonlinear regime. The final permittivity distribution after optimizing is also shown. (b) The transmission in the right (blue) and bottom (red) ports as a function of input power, demonstrating the switching behavior at around $2 \times 10^{-2} \text{ W}/\mu\text{m}$ . The dashed black line indicates the input power used for the high-power regime in the optimization. (c-d) The amplitude of the simulated electric field of the final structure, in the linear (c) and nonlinear (d) regimes, respectively – with input power of $10^{-9} \text{ W}/\mu\text{m}$ and $0.057 \text{ W}/\mu\text{m}$ , respectively. . . . .	104

6.5	Objective function vs. iteration of the optimization for (a) 2-port device of Fig. 2 and (b) 3-port device of Fig. 3 of the main text. . . . .	105
6.6	(a) Transmission spectrum through the 2-port device for the low power (linear) regime. (b) Transmission spectrum through the right (blue) and bottom (red) ports of the 3-port device in the low power (linear) regime. The x-axis represents the difference in frequency with respect to the design frequency. . . . .	106
B.1	(a): An OIU implementing a universal $3 \times 3$ unitary operation, parameterized as in Ref. [39]. (b): Illustration of Eq. (B.4) for the computation of the gradient with respect to a given phase shifter. . . . .	114

# Chapter 1

## Introduction

The field of photonics is concerned with the study and manipulation of light. This endeavor has given rise to countless technologies of great scientific and practical interest. Perhaps most prominently, the use of light as an information carrier has enabled high speed and low loss communications [5]. Light is also used extensively for precise detection and measurement in scientific studies. For example, X-ray radiation is now used to observe femtosecond dynamics in chemical reactions [94], and laser interferometry was recently used to measure gravitational waves emitted from black hole mergers [112]. Apart from these examples, there are many applications of photonics with significant practical importance ranging from renewable energy [30, 223] to passive refrigeration [160, 75].

One of the most important achievements of photonics in the past few decades has been the development of *integrated* photonic devices [153]. In this paradigm, rather than constructing devices using macroscopic components, such as lenses and mirrors, they are created on the surface of a chip using techniques common to the semiconductor industry. Such an approach is appealing as it allows for compact, low-cost, and highly functional devices that are also easier to integrate with existing electronic platforms based on composite metal on semiconductor (CMOS) technology [209]. The field of *silicon photonics* has especially generated much interest in recent years, in which photonic devices integrated on Silicon are employed in applications ranging from optical interconnects for fast data transfer between microchips to large scale integrated photonic circuits [87].

Here, we will primarily explore two emerging technologies based on integrated photonics, (1) Laser-driven particle accelerators on a chip, and (2) optical hardware for machine learning applications. The approach to laser-driven particle acceleration examined here is referred to as *dielectric laser acceleration*, in which charged particles are accelerated by the near field of a patterned dielectric structure driven by an external laser. As we will show, this technology may benefit greatly from the use of integrated photonic platforms for its eventual practical applications. Integrated photonics is also a promising candidate for building hardware platforms specialized for performing machine learning computation. As the transmission of an image through an optical lens applies a Fourier

transform with minimal energy consumption, one may design reconfigurable integrated photonic devices, which perform arbitrary linear operations through the passive propagation of optical signals through their domain. As machine learning models are often dominated by linear operations, this technology may provide a platform with higher processing speed, lower energy usage when compared to conventional digital electronics.

## 1.1 Photonic Device Design

### 1.1.1 Traditional Design Approach

In any of these applications, it is of critical importance to design the photonic device to perform the desired functionality. The typical approach to such a process is to use physical intuition to propose a device design. This device may be parameterized by several *design variables*, such as geometric or material parameters. Then, these parameters may optimized, using numerical simulation or experiment, until they converge on a final design that performs well on the given task, while satisfying other constraints, such as minimum achievable feature size, for example. As an example, if one is interested in designing a device routes input light to different ports for different input wavelengths, one such approach would be to combine several wavelength filters into one device and tune their parameters until the functionality is achieved. Such an approach, while intuitive, has a number of potential drawbacks. First, it is dependent on the designer having significant physical intuition about the problem, which is not always available, especially in novel applications. Second, the method of tuning parameters by hand is tedious and the time needed to complete such a task generally has poor scaling with the number of design variables. This fact means that the designer is practically limited to examining a small number of design variables or only a few select combinations. The use of few design variables further limits the designer to consider devices within a fixed parameterization. For example, if one were to designing a device for tailored diffraction or transmission characteristics, he or she may decide to explore grating structures parameterized by tooth height, width, and duty cycle, while ignoring other possible designs.

### 1.1.2 Inverse Design Approach

*Inverse design* is a radically different approach that has become popular in photonics within the past decade [133]. In this scheme, the overall performance of the device is defined mathematically through an *objective function*, which is then either maximized or minimized using computational and mathematical optimization techniques. This approach allows for automated design of photonic devices that are often more compact and more performant than their traditionally-designed alternatives. Furthermore, this approach allows one to search through a much larger parameter space, typically on the order of thousands to millions of design variables, which enables the design algorithms to

often find complex structures that are beyond the reach of human intuition.

The use of inverse design has a long history in other fields, such as mechanics [198], aerodynamics [88], and heat transfer [144]. However, its introduction to photonics has only occurred relatively recently. A few early examples include the use of inverse design to engineer wavelength splitters [148], perfect 90 degree bends in dielectric waveguides [89], or the design of photonic crystals [20]. More recently, it was applied to engineer more exotic phenomena, nonlinear optical responses [114], and metamaterials [177]. For a thorough overview of the progress of inverse design in photonics at the time of publishing, we refer the reader to Ref. 133.

## 1.2 Adjoint Method

As we will explore in detail, the ability to perform inverse design is largely enabled by the ability to efficiently search such a large parameter space. Typically, this is performed using *gradient-based optimization* techniques, which use local gradient information to iteratively progress through the design space. In design problems with several degrees of freedom, gradient-based methods typically converge on local minima much faster than more general optimization techniques such as particle swarm optimization [1] or genetic algorithms [211], which don't typically use local gradient information.

In problems constrained by physics described by linear systems or differential equations, the *adjoint method* is typically used to compute these gradients. The adjoint method allows one to compute gradients of the objective function with respect to each of the design parameters in a complexity that is, in practice, independent of the size of the design space. As such, it is the cornerstone of the inverse design works in photonics and other fields.

Here we give a brief introduction to the mathematics behind the adjoint method before discussing it in greater detail in this thesis. Many engineering systems can be described by a linear system of equations  $A(\phi)\mathbf{x} = \mathbf{b}$ , where  $A$  is a sparse matrix that depends on a set of parameters describing the system,  $\phi$ . Solving this equation with a driving source  $\mathbf{b}$  results in the solution  $\mathbf{x}$ , from which an objective function  $J = J(\mathbf{x})$  can be computed.

The optimization of this system corresponds to maximizing or minimizing  $J$  with respect to the set of parameters  $\phi$ . For this purpose, the adjoint method allows one to calculate the gradient of the objective function  $\nabla_\phi J$  for an arbitrary number of parameters. Crucially, this gradient may be obtained with the computational cost of solving only one additional linear system  $\hat{A}^T \bar{\mathbf{x}} = -\frac{\partial J}{\partial \mathbf{x}}$ , which is often called the ‘adjoint’ problem. With this adjoint solution in hand, the final gradient is given by  $\nabla_\phi J = \bar{\mathbf{x}}^T \frac{\partial A}{\partial \phi} \mathbf{x}$  where  $\frac{\partial A}{\partial \phi}$  is typically trivial to compute.

In the case of performing inverse design of electromagnetic devices,  $A$  represents Maxwell's equations in the presence of the device,  $\mathbf{x}$  are the electromagnetic fields, and  $\mathbf{b}$  is the electric current source driving the system.

### 1.3 Thesis Overview

In this thesis, we will discuss the application of the adjoint method to new applications in photonics. We will also introduce extensions to the adjoint method, which allow it to be applied to new systems and implemented experimentally. The thesis is organized as follows: In Chapter 2, we will introduce the mathematical details behind adjoint-based optimization. To give a concrete example, we will focus on its application to laser-driven particle accelerators on a chip. To continue this discussion, in Chapter 3, we will explore the use of photonic integrated circuits to scale laser-driven particle accelerators to longer length scales and practical applications. This discussion will motivate the need to use inverse design for new components, and we will discuss efforts to use such techniques to build these systems experimentally. In Chapter 4, we will discuss optical hardware platforms for machine learning applications. The adjoint method will be explored in the context of training an optical neural network, and we will show that its implementation corresponds to the backpropagation algorithm of conventional neural networks. A novel method for experimentally measuring the gradients obtained through the adjoint method will be introduced in the context of machine learning hardware and we will also discuss our exploration of nonlinear optical activation functions. In Chapter 5, the use of time-domain adjoint methods will then be introduced in the context of using wave physics to perform analog machine learning on raw time-series signals. Then, in Chapter 6, we will explore the extension of the adjoint method to general systems, in particular those described by a nonlinear system of equations. We will conclude and provide an outlook of these methods in Chapter 7.

## Chapter 2

# Adjoint-Based Optimization of Laser-Driven Particle Accelerators

### 2.1 Dielectric Laser Acceleration

In the public sphere, particle accelerators most commonly conjure images of giant facilities for performing particle physics experiments, such as the Large Hadron Collider. However, in fact, most particle accelerators are used in other applications, such as radiotherapy, X-ray generation, and ion implantation for semiconductor device fabrication [53]. Conventional radio-frequency (RF) accelerators use a metal or superconducting cavity, driven with microwave radiation, to provide sustained acceleration to charged particles traversing the structure. However, the amount of achievable acceleration per unit length is fundamentally limited by the material breakdown and damage limit of the cavity. Therefore, for an accelerator already driven at its damage threshold, the only option to achieve high total energy gains from an accelerator is to make the device longer. The largest current particle accelerators reach several of kilometers in length, requiring substantial resources to operate and maintain.

Dielectric laser acceleration (DLA) is an emerging method that seeks to revolutionize particle accelerator technology by exploiting the decades of progress in nanofabrication, materials science, and laser technology. In DLA, instead of using metal structures driven by microwaves, dielectric structures are illuminated with infrared laser light, which creates an electromagnetic field pattern in their vicinity that may accelerate electrically charged particles, such as electrons. When compared to metal surfaces at microwave frequencies, dielectric materials have very high damage thresholds at short pulse durations and infrared wavelengths [118, 182]. This fact allows DLAs to achieve energy gains per length that are between 10 to 100 times higher than those found in conventional radio frequency (RF) accelerators.

Experimental demonstrations of these acceleration gradients have been made practical in recent years by the availability of robust nanofabrication techniques combined with modern solid state laser systems [44]. As a result, the development of DLA may lead to compact particle accelerators that enable new applications. By providing the potential for generating relativistic electron beams in relatively short length scales, DLA technology is projected to have numerous applications where tabletop accelerators may be useful, including medical imaging, radiation therapy, and X-ray generation [151, 53].

Several recently demonstrated candidate DLA structures consist of a planar dielectric structure that is periodic along the particle axis with either a semi-open geometry or a narrow (micron to sub-micron) vacuum gap in which the particles travel [152, 147, 118, 110, 35, 24, 25, 100]. These structures are then side-illuminated by laser pulses. Fig. 2.1 shows a schematic of the setup, with a laser pulse incident from the bottom.

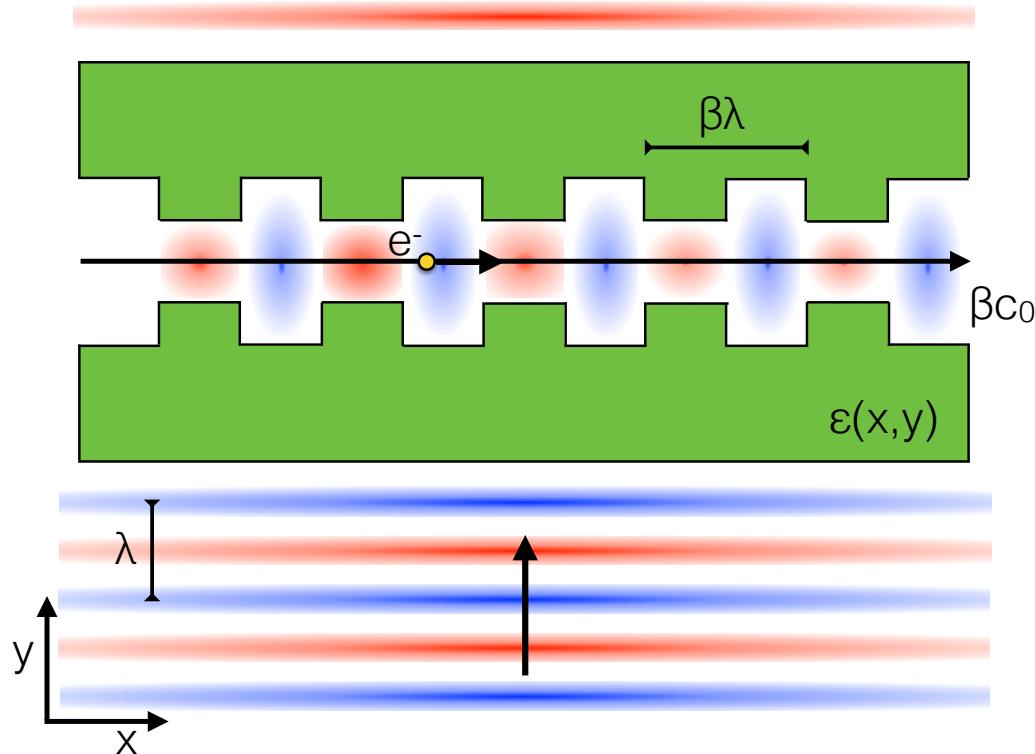


Figure 2.1: Diagram outlining the system setup for side-coupled DLA with an arbitrary dielectric structure  $\epsilon(x, y)$  (green). A charged particle moves through the vacuum gap with speed  $\beta c_0$ . The periodicity is set at  $\beta\lambda$  where  $\lambda$  is the central wavelength of the laser pulse.

The laser field may also be treated with a pulse front tilt [70, 6] to enable group velocity matching over a distance greater than the laser's pulse length.

## 2.2 Adjoint Method for Accelerator

### 2.2.1 Mathematical Definition

To achieve high energy gain in a compact size, it is of principle interest to design structures that may produce the largest acceleration gradients possible without exceeding their respective damage thresholds. Here we will discuss the use of the adjoint method and inverse design to design such a structure, as explained originally in Ref. 77. To begin, we must first define the optimization figure of merit and design parameters.

We first seek to maximize the *acceleration gradient* of the device, which is defined as the amount of energy gain per unit length achieved by a particle that is phased correctly with the driving field. For acceleration to occur, the dielectric structure must be designed such that the particle feels an electric field that is largely parallel to its trajectory over many optical periods. In a general DLA device, the acceleration gradient 'G' over a time period 'T' is defined mathematically as

$$G = \frac{1}{T} \int_0^T E_{||}(\vec{r}(t), t) dt, \quad (2.1)$$

where  $\vec{r}(t)$  is the position of the electron over time and  $E_{||}$  signifies the (real) electric field component parallel to the electron trajectory.

Since we assume the structure is invariant in the  $\hat{z}$  direction, we may work in two dimensions, examining only the  $H_z$ ,  $E_x$  and  $E_y$  field components. While this approximation neglects fringing fields that will be present in any fabricated device, it is a good approximation for the fields experienced by particles traversing the center of the acceleration channel. For an approximately monochromatic input laser source with angular frequency  $\omega$ , the electric fields are, in general, of the form

$$\vec{E}(\vec{r}, t) = \Re \left\{ \tilde{E}(\vec{r}) e^{i\omega t} \right\}, \quad (2.2)$$

where  $\tilde{E}$  is complex-valued.

Let us assume the particle we wish to accelerate is moving on the line  $y = 0$  with velocity  $\vec{v} = \beta c_0 \hat{x}$ , where  $c_0$  is the speed of light in vacuum and  $\beta \leq 1$ . The  $x$  position of the particle as a function of time is given by  $x(t) = x_0 + \beta c_0 t$ , where  $x_0$  represents an arbitrary choice of initial starting position. For normal incidence of the laser (laser propagating in the  $+\hat{y}$  direction), phase velocity matching between the particle and the electromagnetic fields is established by introducing a spatial periodicity in our structure of period  $\beta\lambda$  along  $\hat{x}$ , where  $\lambda$  is the laser wavelength. In the limit of an infinitely long structure (or equivalently,  $T \rightarrow \infty$ ) we may rewrite our expression for the

gradient in Eq. (2.1) as an integral over one spatial period, given by

$$G = \frac{1}{\beta\lambda} \Re \left\{ e^{-i\phi_0} \int_0^{\beta\lambda} dx E_x(x, 0) e^{i\frac{2\pi}{\beta\lambda}x} \right\}. \quad (2.3)$$

Here the quantity  $\phi_0 = \frac{2\pi x_0}{\beta\lambda}$  is representative of the phase of the particle as it enters the spatial period. In further calculations, we set  $\phi_0 = 0$ , only examining the acceleration gradients experienced by particles entering the accelerator with this specific phase. Since we have arbitrarily control over our input laser phase, this does not impose any constraint on the acceleration gradient attainable.

To simplify the following derivations, we define the following inner product operation involving the integral over two vector quantities  $\vec{a}$  and  $\vec{b}$  over a single period volume  $V'$

$$\langle \vec{a}, \vec{b} \rangle = \int_{V'} dv \cdot (\vec{a} \cdot \vec{b}) = \int_0^{\beta\lambda} dx \int_{-\infty}^{\infty} dy (\vec{a} \cdot \vec{b}). \quad (2.4)$$

With this definition, we then have the gradient

$$G = \Re \left\{ \langle \vec{E}, \vec{\eta} \rangle \right\}, \quad (2.5)$$

where we define the vector field ' $\vec{\eta}$ ' to signify the position and phase of the moving electron as

$$\vec{\eta}(x, y) = \frac{1}{\beta\lambda} e^{i\frac{2\pi}{\beta\lambda}x} \delta(y) \hat{x}. \quad (2.6)$$

The physical interpretation of  $\vec{\eta}$  is digrammed in Fig. 2.2.

Our goal in designing the accelerator for maximum acceleration gradient is to create a permittivity distribution that maximizes  $G$  subject to a few constraints. We assume geometry of the DLA structure is represented by a spatially varying dielectric constant  $\epsilon(x, y)$ . As mentioned, we assume invariance in one coordinate ( $\hat{z}$ ) in keeping with the planar symmetry of most current designs. However the methodology we present can be extended to include a third dimension. We consider performing this optimization in a small design region surrounding a small gap defined for the electron to travel through the structure. Secondly, we assume that the structure has a finite extent along the direction of the incoming laser beam. We also consider realizing this device through the patterning of a material with permittivity  $\epsilon_{\max}$ . Therefore, the final device should have permittivity of either 1 or  $\epsilon_{\max}$  at all points.

To perform this optimization task, we discretize our entire spatial domain into a rectangular grid, which will be necessary for numerical simulation. We define our design parameters,  $\phi$ , as the relative permittivity of each grid cell within the design region. Our problem then becomes finding the permittivity of each cell that will maximize the acceleration gradient, subject to each grid cell having a permittivity value of either 1 or  $\epsilon_{\max}$ .

To accomplish this, the most naive approach would involve performing a direct search over the

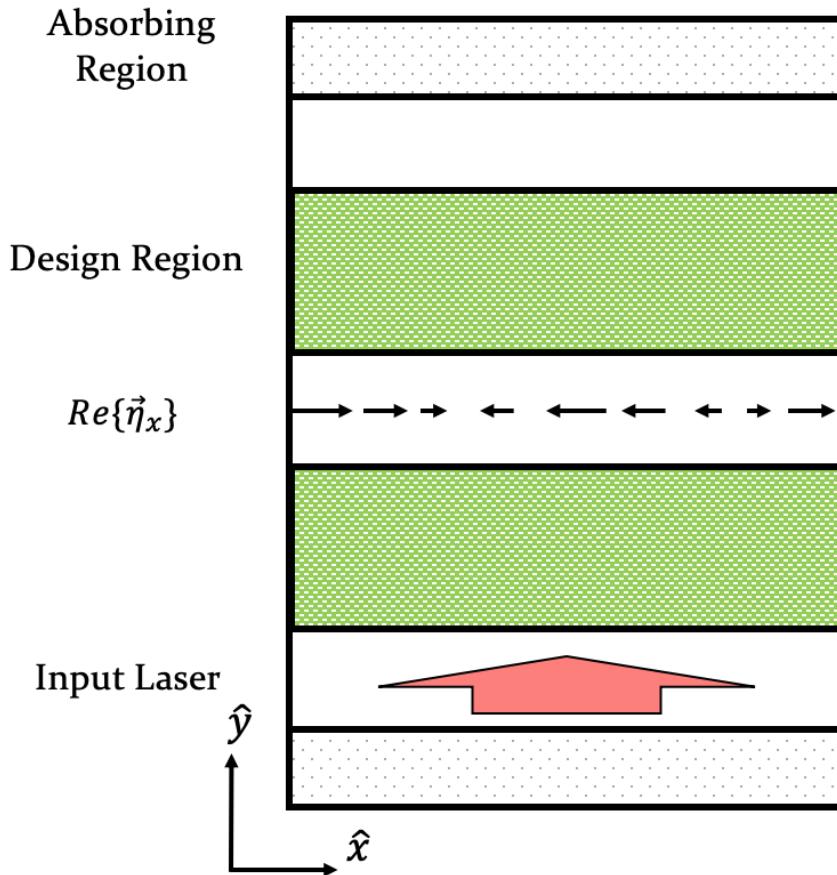


Figure 2.2: Definition of the vector field,  $\vec{\eta}$ , which defines the position of the electron in the frequency domain. The green regions represent domains where we will optimize the material properties using the adjoint method. The central gap is constrained to vacuum to allow passage of the electron beam. The red arrow signifies the driving laser.

full design space. For example, one could label each cell within the design region with an identifier ‘0’ or ‘1’ corresponding to ‘vacuum’ and ‘material’, respectively. Then, one would generate all possible structures and check their respective acceleration gradients. However, as one can imagine, this method would be far too computationally expensive to perform in practice. For example, even considering a very small design region consisting of  $10 \times 10 = 100$  grid cells would result in  $2^{100} \approx 10^{30}$  device simulations, which is far too many to realistically perform. While one may consider smart ways of searching through this device space without checking each structure, for example using global optimization approaches such as genetic algorithms [211] or particle swarm optimization [1], this problem is still quite computationally expensive and the size of design space becomes exponentially larger as the number of design parameters are increased.

As mentioned, a smarter approach involves performing *gradient-based optimization*, in which we search the design space according to the local gradient of the figure of merit with respect to each of the parameters. For example, we may start with an initially random device, compute how the performance will change with respect to a change in the permittivity of each cell in the design region, and make a small update. This process may be repeated until convergence on a locally optimal solution. If the design space contains several local optima, then this whole process may be repeated several times with different initial conditions.

For gradient-based optimization to be useful, one would like an efficient means to compute the gradient of the figure of merit with respect to the design parameters. For photonic devices, one may achieve this using the *adjoint method*, which allows one to analytically compute the gradient directly from Maxwell's Equations and evaluate the result with only one additional electromagnetic simulation. This remarkable efficiency is largely responsible for the success of inverse design in photonics.

### 2.2.2 Adjoint Formalism

The adjoint method is typically introduced for linear optical systems, although, as we will show in a later chapter, it may be extended to nonlinear systems without much additional complication. In the frequency domain, Maxwell's equations may be written as

$$\nabla \times \nabla \times \vec{E}(\vec{r}) - k_0^2 \epsilon_r(\vec{r}) \vec{E}(\vec{r}) \equiv A\vec{E}(\vec{r}) = -i\mu_0\omega\vec{J}(\vec{r}), \quad (2.7)$$

Here,  $\vec{E}(\vec{r})$  and  $\vec{J}(\vec{r})$  are the electric field and electric current distributions, respectively.  $k_0 = \omega/c_0$ ,  $\epsilon_r$  is the relative permittivity and a non-magnetic material is assumed ( $\mu = \mu_0$ ). This formalism is referred to as *finite-difference frequency-domain* (FDFD) [174, 192].

More abstractly, we may write Eq. (2.7) as

$$A\mathbf{x} = \mathbf{b}, \quad (2.8)$$

where  $A$  is a sparse, complex symmetric matrix that encodes Maxwell's equations in the presence of the device.  $\mathbf{x}$  is a vector containing the electromagnetic fields at each position in the domain, which are the solution to Eq. (2.8) given the vector  $\mathbf{b}$  describing the electric current source distribution in the domain. When the number of grid points is small, Eq. (2.8) is typically solved using lower-upper (LU) decomposition techniques. However, in several cases, it must be solved using iterative methods instead [174].

Our device is described by a set of design variables  $\phi$ , which influence the system matrix,  $A = A(\phi)$ . Differentiating Eq. (2.7) with respect to  $\phi$ , and assuming that the current source,  $\mathbf{b}$ , does

not depend on  $\phi$ , we may recover the change in the solution with respect to the parameters as

$$\frac{d\mathbf{x}}{d\phi} = -A^{-1} \frac{\partial A}{\partial \phi} A^{-1} \mathbf{b} = -A^{-1} \frac{\partial A}{\partial \phi} \mathbf{x} \quad (2.9)$$

Now, we consider differentiating an objective function  $J = J(\mathbf{x})$  that depends explicitly on the field solution. By the chain rule, this gives

$$\frac{dJ}{d\phi} = -\Re \left\{ \frac{\partial J}{\partial \mathbf{x}} \frac{d\mathbf{x}}{d\phi} \right\} = 2\Re \left\{ -\frac{\partial J}{\partial \mathbf{x}} A^{-1} \frac{\partial A}{\partial \phi} \mathbf{x} \right\}. \quad (2.10)$$

We note that the factor of  $2\Re \{\cdot\}$  comes from the fact that  $\mathbf{x}$  is complex-valued, and therefore in taking the derivative of (real-valued)  $J$ , one must consider the dependence on both  $\mathbf{x}$  and  $\mathbf{x}^*$ .

To evaluate Eq. (2.10), we define a second simulation with source term  $-\frac{\partial J}{\partial \mathbf{x}}^T$ ,

$$A^T \mathbf{x}_{aj} = A \mathbf{x}_{aj} = -\frac{\partial J}{\partial \mathbf{x}}^T, \quad (2.11)$$

where we have used the fact that  $A$  is symmetric. Then, the field solution,  $\mathbf{x}_{aj} = -A^{-1} \frac{\partial J}{\partial \mathbf{x}}^T$ , can be easily identified in Eq. (2.10), which gives the expression

$$\frac{dG}{d\phi} = 2\Re \left\{ \mathbf{x}_{aj}^T \frac{\partial A}{\partial \phi} \mathbf{x} \right\}. \quad (2.12)$$

The only quantity in this expression that depends on the parameter  $\phi$  is  $\frac{\partial A}{\partial \phi}$ . As we will soon discuss, this quantity will generally be trivial to compute. On the other hand, the full field calculations of  $\mathbf{x}$  and  $\mathbf{x}_{aj}$  are computationally expensive, but may be computed once and used for an arbitrarily large set of parameters  $\phi_i$ . This gives the adjoint method significant scaling advantage with respect to traditional direct sensitivity methods, such as finite difference, which require a separate full-field calculation for each parameter being investigated.

Previously, we expressed the acceleration gradient as the inner product of the electric fields  $\vec{E}(\vec{r})$  and our vector field  $\vec{\eta}(\vec{r})$  as

$$G = \Re \left\{ \langle \vec{E}, \vec{\eta} \rangle \right\}, \quad (2.13)$$

which we may express in matrix notation as

$$G = \Re \{ \boldsymbol{\eta}^T \mathbf{x} \}, \quad (2.14)$$

where  $\boldsymbol{\eta}$  is a vector representing  $\vec{\eta}$  on the finite difference grid and  $\mathbf{x}$  is a vector containing the electric fields, as before.

In this notation, the adjoint problem is therefore

$$A^T \mathbf{x}_{\text{aj}} = -\frac{\partial G^T}{\partial \mathbf{x}} = -\boldsymbol{\eta}. \quad (2.15)$$

### 2.2.3 Interpretation of the Adjoint Field as Radiation

Intuitively, the adjoint source,  $-\boldsymbol{\eta}$ , represents a line of electric current located at the central gap where the accelerator is traversing the structure. To make this connection to radiation more explicit, let us now consider the fields radiated by a point particle of charge  $q$  flowing through our domain at  $y = 0$  with velocity  $\vec{v} = \beta c_0 \hat{x}$ . In the time domain, we can represent the current density of this particle as

$$\vec{J}_{\text{rad}}(x, y; t) = q\beta c_0 \delta(x - x_0 - c_0 \beta t) \delta(y) \hat{x}. \quad (2.16)$$

To express this in the frequency domain, we take the Fourier transform of  $\vec{J}_{\text{rad}}$  with respect to time, giving

$$\vec{J}_{\text{rad}}(x, y; \omega) = q\beta c_0 \delta(y) \hat{x} \int_{-\infty}^{\infty} dt \exp(i\omega t) \delta(x - x_0 - c_0 \beta t) \quad (2.17)$$

$$= q \exp\left(i\frac{\omega(x - x_0)}{c_0 \beta}\right) \delta(y) \hat{x} \quad (2.18)$$

$$= q \exp\left(i\frac{2\pi}{\beta\lambda}x\right) \exp(-i\phi_0) \delta(y) \hat{x}. \quad (2.19)$$

Comparing with the source of our adjoint problem,  $\vec{J}_{\text{aj}} = \frac{-i}{\omega\mu_0} \vec{\eta}$ , we see that

$$\vec{J}_{\text{aj}} = \frac{-i \exp(i\phi_0)}{2\pi q\beta c_0 \mu_0} \vec{J}_{\text{rad}}. \quad (2.20)$$

This finding shows that the adjoint field solution ( $\vec{E}_{\text{aj}}$ ) corresponds (up to a complex constant) to the field radiating from a test particle flowing through the accelerator structure. To put this another way, in order to calculate the acceleration gradient sensitivity with the adjoint method, we must simulate the same structure operating both as an accelerator ( $A\vec{E} = -i\omega\mu_0 \vec{J}_{\text{acc}}$ ) and as a radiator ( $A\vec{E}_{\text{aj}} = -i\omega\mu_0 \vec{J}_{\text{aj}}$ ).

It is understood that one way to create acceleration is to run a radiative process in reverse. Indeed, this is the working principle behind accelerator schemes such as inverse free electron lasers [135, 42], inverse Cherenkov accelerators [97, 58], and inverse Smith-Purcell accelerators [14, 132]. Here, we see that this relationship can be expressed in an elegant fashion using the adjoint method.

### 2.2.4 Application to Design Parameters

To test our adjoint formalism, we examine a simple accelerator geometry composed of two opposing dielectric squares each of relative permittivity  $\epsilon$ . We take a single  $\phi$  parameter to be the relative permittivity of the entire square region. Because we only change the region inside the dielectric square, we may identify the  $\frac{\partial A}{\partial \phi}$  operator from Eq. (2.7) as

$$\frac{dA}{d\epsilon}(\vec{r}) = \begin{cases} -2k_0^2 & \text{if } \vec{r} \text{ in square} \\ 0 & \text{otherwise} \end{cases}. \quad (2.21)$$

Thus, given the form of the sensitivity of the acceleration gradient from Eq. (2.12), combined with Eq. (2.21), the change in acceleration gradient with respect to changing the entire square permittivity is simply given by the integral of the two field solutions over the square region, labeled ‘sq’

$$\frac{dG}{d\epsilon_{sq}} = -2k_0^2 \Re \left\{ \int_{sq} d^2\vec{r} \cdot \vec{E}(\vec{r}) \cdot \vec{E}_{aj}(\vec{r}) \right\}. \quad (2.22)$$

In Fig. Fig. 2.3 we compare this result with the direct sensitivity calculation where the system is manually changed and simulated again, using a finite difference derivative. The two methods agree with excellent precision, which confirms that the adjoint formalism is giving the correct results.

Extending this example to the general case of perturbing the permittivity at an arbitrary position  $\vec{r}$ , we see that

$$\frac{dG}{d\epsilon}(\vec{r}) = -k_0^2 \Re \left\{ \int d^2\vec{r}' \cdot \vec{E}(\vec{r}') \cdot \vec{E}_{aj}(\vec{r}') \delta(\vec{r} - \vec{r}') \right\} \quad (2.23)$$

$$= -2k_0^2 \Re \left\{ \vec{E}(\vec{r}) \cdot \vec{E}_{aj}(\vec{r}) \right\}. \quad (2.24)$$

## 2.3 Inverse Design of Accelerator

### 2.3.1 Maximization of Acceleration Gradient

With the mathematical form of the adjoint problem discussed, now we focus on the optimization and inverse design of the DLA device. In our FDFD simulation, we use a grid spacing that corresponds to 200 grid points per free space wavelength in each dimension. Perfectly matched layers are implemented as absorbing regions on the edges parallel to the electron trajectory, with periodic boundary conditions employed on boundaries perpendicular to the electron trajectory. A total-field scattered-field [192] formalism is used to create a perfect plane wave input for the acceleration mode.

Since the adjoint method gives us a highly efficient method to calculate  $\frac{dG}{d\epsilon_i}$ , we use this information in an iterative optimization procedure. During each iteration, we first calculate  $\frac{dG}{d\epsilon_i}$  for all

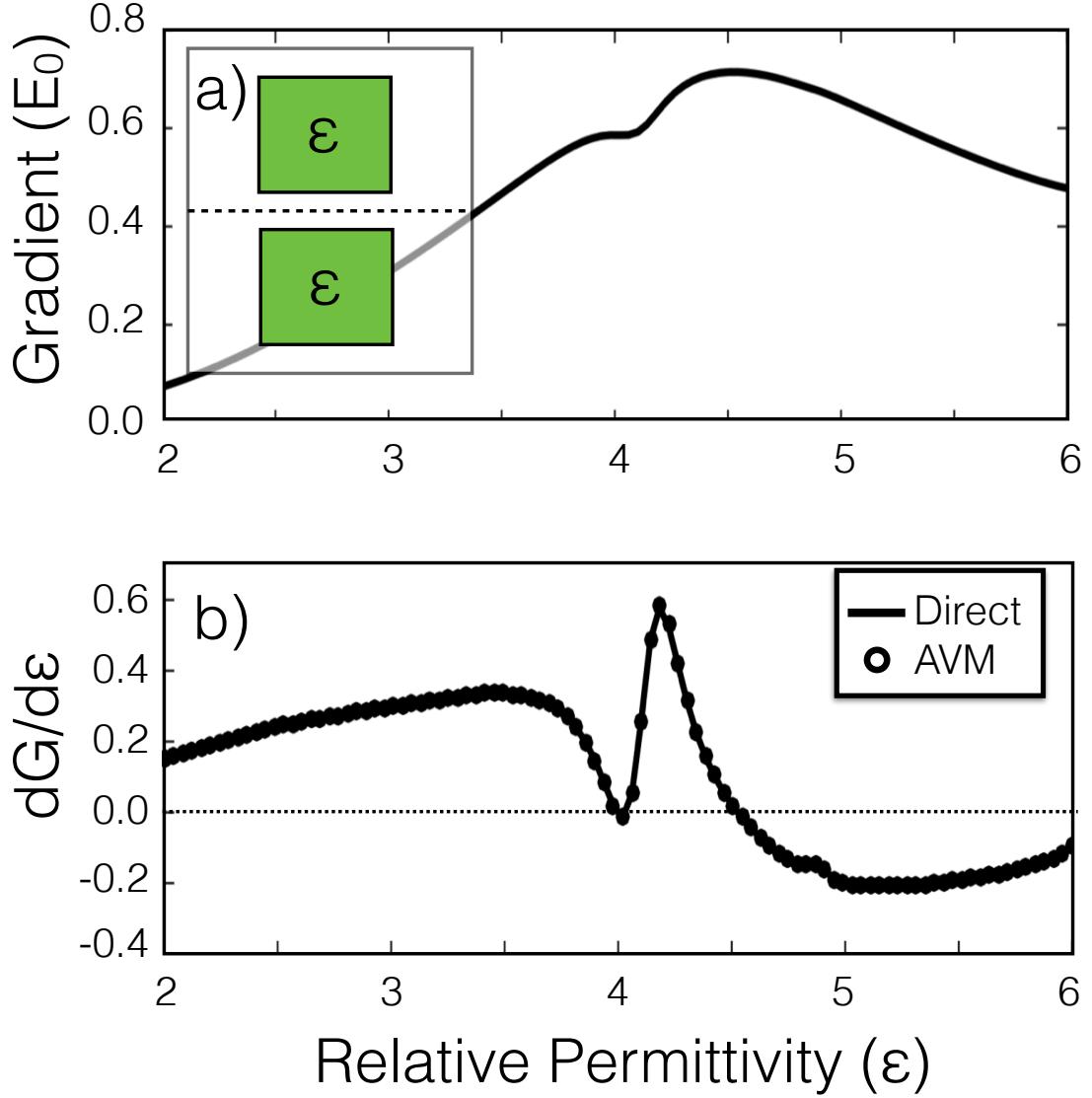


Figure 2.3: Demonstration of adjoint method in calculating accelerator sensitivities. (a) The acceleration gradient ( $G$ ) of a square accelerator structure (inset) as a function of the square's relative permittivity. We express the acceleration gradient in its dimensionless form, normalized by the electric field amplitude of the incident plane wave ( $E_0$ ). The particle traverses along the dotted line with a velocity of  $c_0$  ( $\beta = 1$ ) and a plane wave is incident from the bottom of the structure. (b) The sensitivity  $\frac{dG}{d\epsilon}$  of the gradient with respect to changing the square relative permittivity for direct central difference (solid line)  $\frac{dG}{d\epsilon} = \frac{G(\epsilon + \Delta\epsilon) - G(\epsilon - \Delta\epsilon)}{2\Delta\epsilon}$  and using the adjoint method (circles). The two calculations agree with excellent precision. The dotted line at  $\frac{dG}{d\epsilon} = 0$ , corresponds to local minima and maxima of  $G(\epsilon)$  above.

pixels ‘ $i$ ’ within some specified design region. Then, we update the  $\epsilon_i$  of each grid cell as follows

$$\epsilon_i := \epsilon_i + \alpha \frac{dG}{d\epsilon_i}. \quad (2.25)$$

Here,  $\alpha$  is a step parameter that we can tune. We need  $\alpha$  to be small enough to find local maxima, but large enough to have the optimization run in reasonable amount of time. This process is repeated until convergence on  $G$ , as diagrammed in Fig. 2.4.

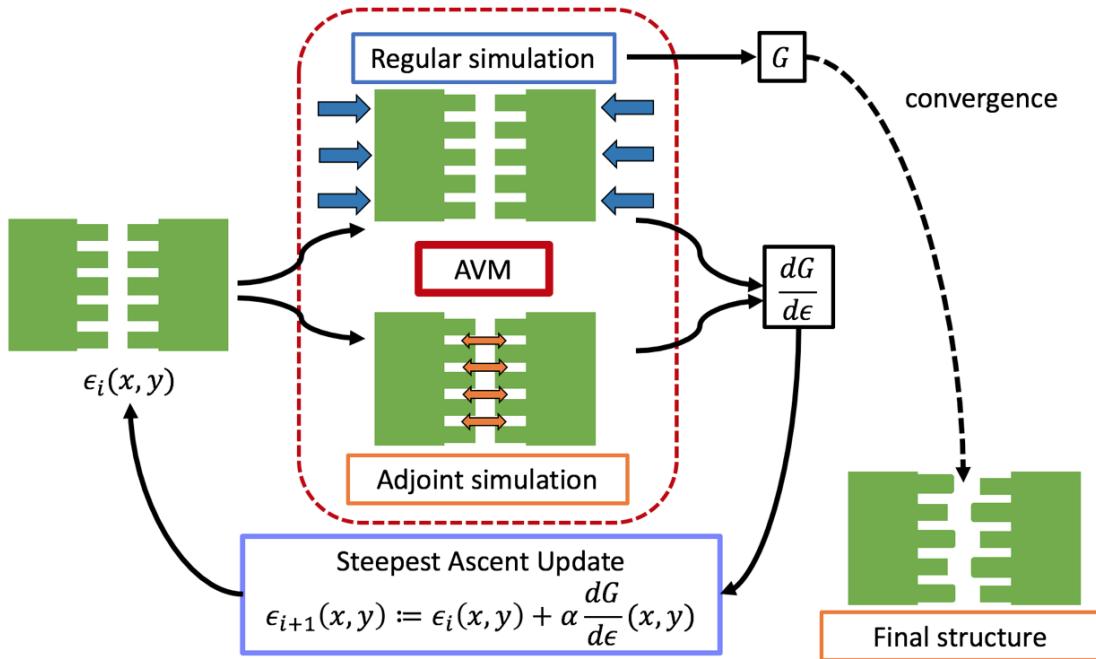


Figure 2.4: Optimization routine for DLA.

During the course of optimization, the permittivity distribution is considered as a continuous variable, which is not realistic in physical devices. To address this issue, we employ a permittivity capping scheme during optimization. We define a maximum permittivity ‘ $\epsilon_m$ ’ corresponding to a material of interest. During the iterative process, if the relative permittivity of any cell becomes either less than 1 (vacuum) or greater than  $\epsilon_m$ , that cell is pushed back into the acceptable range. It was found that with this capping scheme, the structures converged to binary (each pixel being either vacuum or material with a permittivity of  $\epsilon_m$ ) after a number of iterations without specifying this choice of binary materials as a requirement of the optimization. Therefore, only minimal post-processing of the structures was required. Since high index contrast is favorable for maximizing the acceleration gradient, the optimization often favors increasing index contrast within the structure, which leads to each pixel being pushed towards either 1 or  $\epsilon_m$ . This is a possible explanation for

why the structures converge to binary distributions under our optimization procedure.

The results of this optimization scheme are shown in Figs. 2.5(b)-2.5(d) for three different  $\epsilon_m$  values corresponding to commonly explored DLA materials. The design region was taken to be a rectangle fully surrounding but not including the particle gap. The design region was made smaller for higher index materials, since making it too large led to divergence during the iteration. We found that a totally vacuum initial structure worked well for these optimizations. However, initially random values between 1 and  $\epsilon_m$  for each pixel within the design region also gave reasonable results. For materials of Si,  $\text{Si}_3\text{N}_4$ , and  $\text{SiO}_2$ , the achieved gradients (normalized by the incident field,  $E_0$ ) were 0.90, 0.56, and 0.31 ( $E_0$ ), respectively. Assuming incident field values consistent with the laser damage thresholds given in Ref. [182], these correspond to acceleration gradients of 0.25, 0.83, and  $1.33 \text{ GV m}^{-1}$ , respectively. Without the dielectric mirrors, these structures give normalized acceleration gradients of 0.28, 0.15, and 0.07 ( $E_0$ ), respectively. For comparison, the Si dual pillar structures presented in Ref. [110] give gradients of around 0.3 ( $E_0$ ). Therefore, the adjoint optimized structures show about a three-fold improvement in acceleration gradient over established structures.

This optimization scheme seems to favor geometries consisting of a staggered array of field-reversing pillars surrounding the vacuum gap, which is already a popular geometry for DLA. However, these optimal designs also include reflective mirrors on either side of the pillar array, which suggests that for strictly higher acceleration gradients, it is useful to use dielectric mirrors to resonantly enhance the fields in the gap.

It was observed that for random initial starting permittivity distributions, the same structures as shown in Fig. 2.5 are generated every time. Furthermore, as shown in Fig. 2.6, the geometries found using this method are remarkably similar to those recently proposed through human design using physical intuition [101]. These findings together suggest that the proposed structures may be close to the globally optimal structure for maximizing  $G$ .

It was further found that convergence could be achieved faster by a factor of about ten by including a ‘momentum’ term in the update equation. This term corresponds to the sensitivity calculated at the last iteration multiplied by a constant,  $\alpha' < 1$ . Explicitly, for iteration number ‘ $j$ ’ and pixel ‘ $i$ ’

$$\epsilon_i^{(j+1)} := \epsilon_i^{(j)} + \alpha \left[ \frac{dG^{(j)}}{d\epsilon_i} + \alpha' \frac{dG^{(j-1)}}{d\epsilon_i} \right]. \quad (2.26)$$

### 2.3.2 Maximization of Acceleration Factor

DLas are often driven with the highest input field possible before damage occurs. Therefore, another highly relevant quantity to maximize is the *acceleration factor*, given by the acceleration gradient divided by the maximum electric field amplitude in the system. This quantity will ultimately limit the amount of acceleration gradient we can achieve when running at damage threshold. Explicitly, the acceleration factor is given by

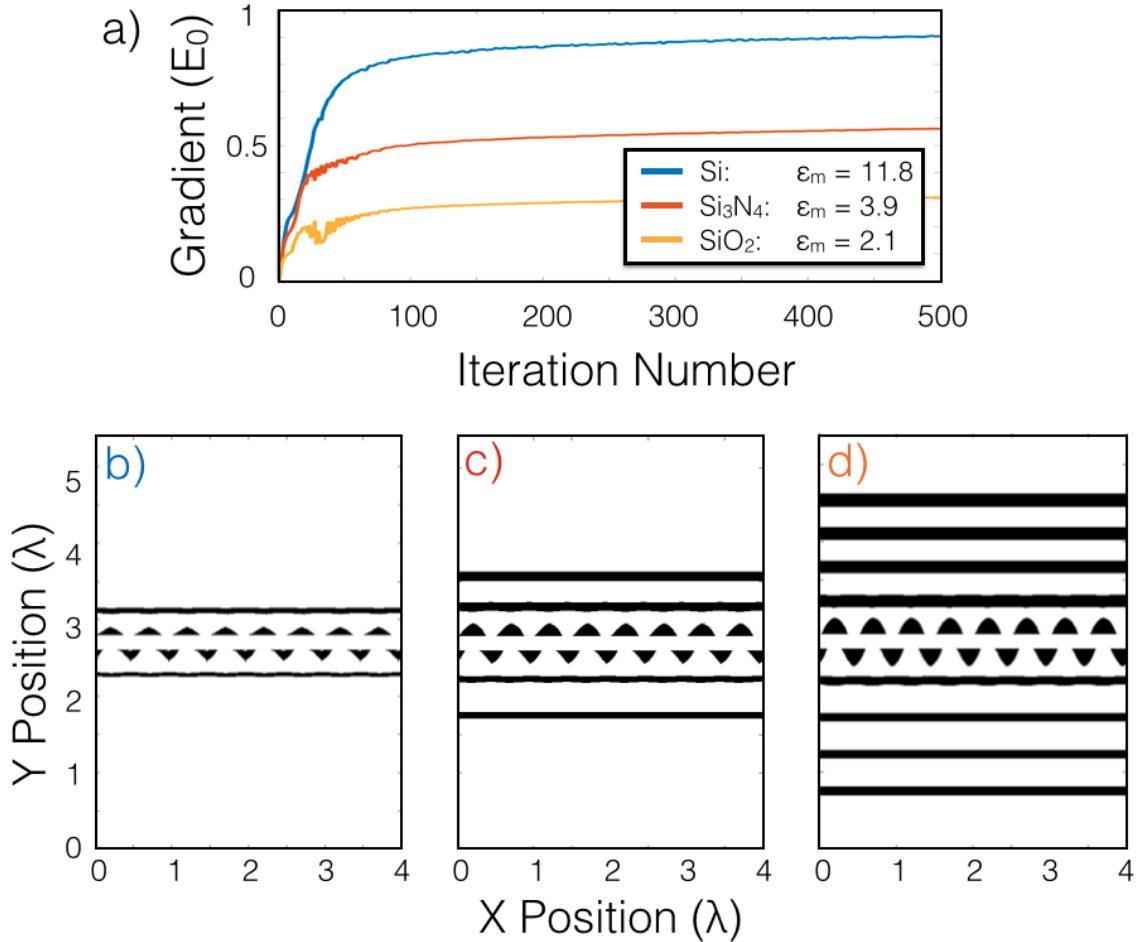


Figure 2.5: Demonstration of the structure optimization for  $\beta = 0.5$ , laser wavelength  $\lambda = 2 \mu\text{m}$ , and a gap size of 400 nm. A plane wave is incident from the bottom in all cases. (a) Acceleration gradient as a function of iteration number for different maximum relative permittivity values, corresponding to those of Si,  $\text{Si}_3\text{N}_4$ , and  $\text{SiO}_2$  at the laser wavelength. The acceleration gradient is normalized by the electric field amplitude of the incident plane wave ( $E_0$ ). The optimizations converge after about five-hundred iterations. (b-d) Final structure permittivity distributions (white = vacuum, black =  $\epsilon_m$ ) corresponding to the three curves in (a). Eight periods are shown, corresponding to four laser wavelengths. For each (b-d), design region widths on each side of the particle gap were given by 1, 2, and 4  $\mu\text{m}$  for Si,  $\text{Si}_3\text{N}_4$ , and  $\text{SiO}_2$ , respectively.

$$f_A = \frac{G}{\max\{|\vec{E}|\}}. \quad (2.27)$$

Here,  $|\vec{E}|$  is a vector of electric field amplitudes in our system, which are normalized, as in the case of  $G$ , by the electric field amplitude of the incident plane wave ( $E_0$ ). The  $\max\{\cdot\}$  function is designed to pick out the highest value of this vector in either our design or material region, depending on

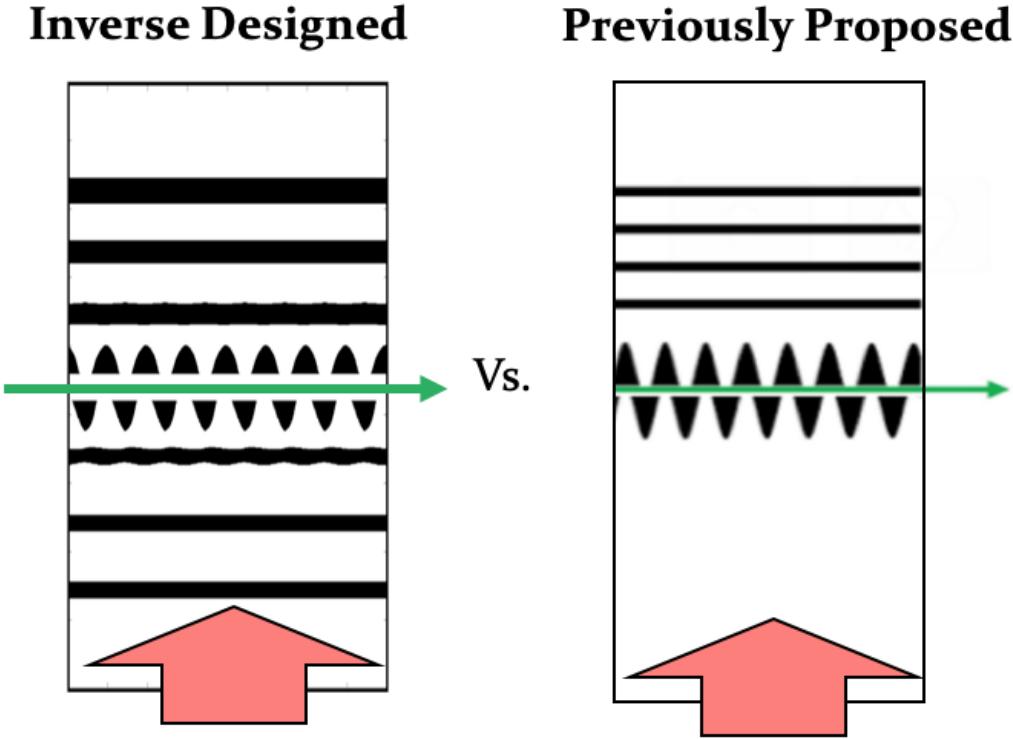


Figure 2.6: Similarity between DLA structures designed using adjoint method (left) and those independently proposed using human intuition.

the context. The *design region* is defined as the total region outside of the particle gap where the permittivity is updated. The *material region* is defined as any region where the permittivity is equal to  $\epsilon_m$ . We would like to use the same basic formalism to maximize  $f_A$ . However, since the  $\max\{\}$  function is not differentiable, this is not possible directly. Instead we may use a ‘smooth-max’ function to approximate  $\max\{\}$  as a weighted sum of vector components

$$\max\{|\vec{E}|\} \approx \frac{\sum_i |\vec{E}_i| \exp(a|\vec{E}_i|)}{\sum_i \exp(a|\vec{E}_i|)}. \quad (2.28)$$

Here, the parameter  $a \geq 0$  controls the relative strength of the exponential sum terms, for  $a = 0$ , this function simply gives the average value of the field amplitudes. By sweeping  $a$  and examining the acceleration factors of the resulting optimized structures, we determined that  $a = 3$  gave the best improvement in  $f_A$ . If  $a$  is too large, the calculation may induce floating point overflow or rounding error issues.

Using this smooth-max function, one may calculate  $\frac{df_A}{d\epsilon_i}$  analytically and perform structure optimizations in the same way that was discussed previously. Two structures with identical parameters but optimized, respectively, for maximum  $G$  and  $f_A$  are shown in Fig. 2.7. On the left, we see that the  $G$  maximized structure shows the characteristic dielectric mirrors, giving resonant field enhancement. On the right is the structure optimized for  $f_A$ , which has eliminated most of its dielectric mirrors and also introduces interesting pillar shapes. In Table 2.1 the main DLA performance quantities of interest are compared between these two structures. Whereas the acceleration gradient is greatly reduced when maximizing for  $f_A$ , the  $f_A$  value itself is improved by about 25% or 23% depending on whether one measures the maximum field in the design region or the material region, respectively. As a comparison with the geometry from Ref. [110], a SiO<sub>2</sub> dual pillar structure with a 400 nm gap was optimized for acceleration gradient, giving an acceleration gradient of 0.025 ( $E_0$ ) with a pillar radius of 670 nm. This structure was found to have  $\max\{|\vec{E}|\}$  in the design region of 1.612 ( $E_0$ ) and  $\max\{|\vec{E}|\}$  in the material region of 1.209 ( $E_0$ ), leading to acceleration factors of 0.016 and 0.021, respectively. Comparing these numbers to those in Table 2.1, it is apparent that the optimized structures perform better than the dual pillar structures in terms of both gradient and acceleration factor.

These findings suggest that the inverse design strategy is effective in designing not only resonant, high acceleration gradient structures, but also non-resonant structures that are more damage resistant. In the future, when more components of DLA are moved on-chip (such as the optical power delivery), it will be important to have control over the resonance characteristics of the DLA structures to prevent damage breakdown at the input facet. Our technique may be invaluable in designing structures with tailor-made quality factors for this application.

Table 2.1: Acceleration factor ( $f_A$ ) before and after maximization.

Quantity	Value (max $G$ )	Value (max $f_A$ )	Chang.
Gradient ( $E_0$ )	0.1774	0.0970	-45.32%
$\max\{ \vec{E} \}$ in design region ( $E_0$ )	4.1263	1.7940	-56.52%
$\max\{ \vec{E} \}$ in material region ( $E_0$ )	2.7923	1.2385	-55.84%
$f_A$ in design region	0.0430	0.0541	+25.81%
$f_A$ in material region	0.0635	0.0783	+23.31%

## 2.4 Conclusions

We found that the adjoint method is a reliable method for optimizing DLA structures for both maximum acceleration gradient and also acceleration factor. The optimization algorithm discussed shows good convergence and rarely requires further post-processing of structures to create binary permittivity distributions. Therefore, it is a simple and effective method for designing DLAs. Whereas most structure optimization in this field uses parameter sweeps to search the design space, the efficiency of

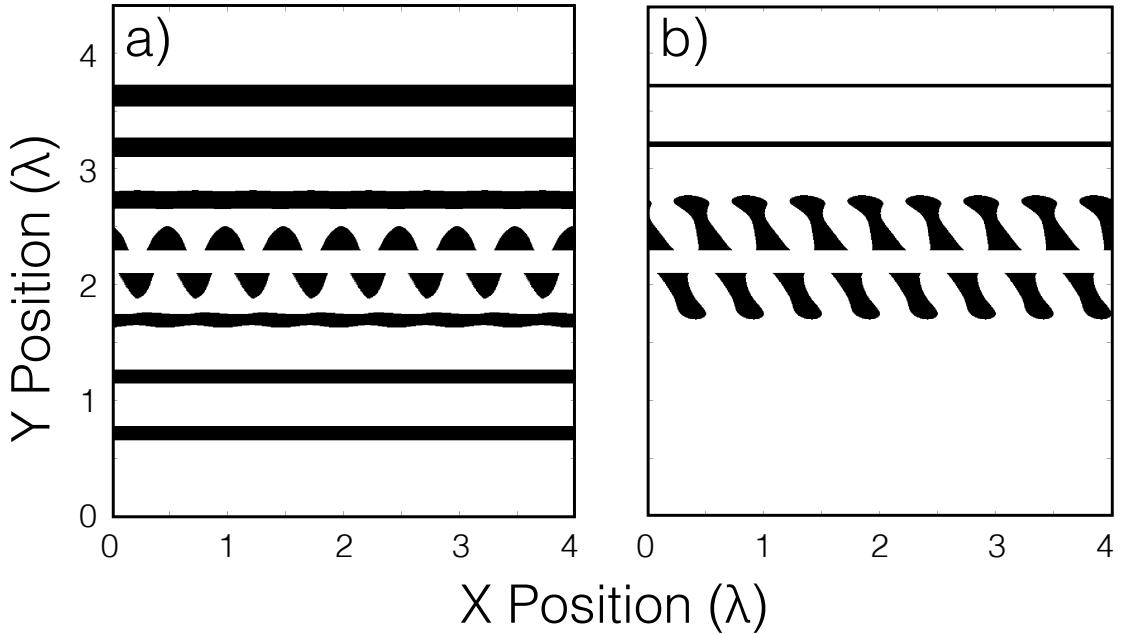


Figure 2.7: Demonstration of the final structures after optimization for (a) maximizing gradient only, (b) maximizing the acceleration factor.  $\beta = 0.5$ , laser wavelength  $\lambda = 2 \mu\text{m}$ , gap size of 400 nm.  $\epsilon_m = 2.1$ , corresponding to  $\text{SiO}_2$ . In (a), the high gradients are achieved using reflective dielectric mirrors to confine and enhance the fields in the center region. In (b), these dielectric mirrors are removed and the pillar structures are augmented. The structure in (b) shows a 23% increase in the acceleration factor in the material region when compared to (a).

our method allows us to more intelligently find optimal geometries without shape parameterization. Furthermore, the structures that we design are fabricable.

Although no DLA structures have been tested at the proposed wavelength of  $2 \mu\text{m}$ , both simulations [152] and experimental results from other wavelengths [110] show gradients far below those presented here. We had limited success designing DLA structures in the relativistic ( $\beta \approx 1$ ) regime, especially for higher index materials, such as Si. We believe this is largely due to the stronger coupling between electron beam and incident plane wave at this energy. The characteristics of the adjoint source change dramatically at the  $\beta = 1$  point. Whereas in the sub-relativistic regime, the adjoint source generates an evanescent near-field extending from the gap particle position, at  $\beta \geq 1$ , the adjoint fields become propagating by process of Cherenkov radiation. Upon using the above described algorithm, the gradients diverge before returning to low values, no matter the step size  $\alpha$ . The only way to mitigate this problem is to choose prohibitively small design regions or low index materials, such as  $\text{SiO}_2$ .

In addition to the side-incident geometry explored, this technique is applicable to designing other

dielectric-based accelerator structures. This includes particle-laser co-propagating schemes [43] and dielectric wakefield acceleration [224], among others. Therefore, we expect that our results may find use in the larger advanced accelerator community.

In this chapter, we have introduced the adjoint variable method as a powerful tool for designing dielectric laser accelerators for high gradient acceleration and high acceleration factor. We have further shown that the adjoint simulation is sourced by a point charge flowing through the accelerator, which quantifies the reciprocal relationship between an accelerator and a radiator.

Optimization algorithms built on this approach allow us to search a substantially larger design space and generate structures that give acceleration gradients far above those normally used for each material. Furthermore, the structures designed by adjoint are fundamentally not constrained by shape parameterization, allowing never-before-seen geometries to be generated and tested. This theme will continue to be explored in other chapters.

## Chapter 3

# Integrated Photonic Circuit for Accelerators on a Chip

In the previous chapter, we discussed the basic working principles of DLA and described its optimization using the adjoint method. Here, we will describe the use of integrated optics to power and control such a device over an extended length.

### 3.1 On-Chip Laser Coupling Device

As mentioned, since DLA structures are already driven at their damage thresholds, apart from finding methods to increase material damage thresholds, achieving high total energy gain from DLA will fundamentally require extending the interaction length between the incoming laser pulse and the particle beam. This interaction length is limited not only by the longitudinal and transverse stability of the electron beam [139, 137], but also by the laser delivery system, which we will focus on in this Chapter. Several proof of principle DLA experiments [213, 147] have demonstrated high acceleration gradients using free-space manipulation of the laser pulse, including lensing, pulse-front-tilting [70, 6, 33], or multiple driving lasers [109, 118]. However, these techniques require extensive experimental effort to perform and the system is exceedingly sensitive to angular alignment, thermal fluctuations, and mechanical noise. By replacing free-space manipulation with precise nanofabrication techniques, an on-chip laser power delivery system would allow for orders of magnitude increases in the achievable interaction lengths and energy gains from DLA.

In designing any laser power delivery system for DLA, there are a few major requirements to consider. (1) The optical power spatial profile must have good overlap with the electron beam side profile. (2) The laser pulses must be appropriately delayed along the length of the accelerator to arrive at the same time as the moving electron bunches. (3) The optical fields along each section

of the accelerator must, ideally, be of the correct phase to avoid dephasing between the electrons and incoming laser fields. To accomplish all three of these requirements, here we discuss a method for on-chip power delivery, which is based on a fractal *tree-network* geometry. Furthermore, we provide a systematic study of the structure's operating principles, the optimal range of operating parameters, and the fundamental trade-offs that must be considered for any on-chip laser coupling strategy of the same class [84]. Through detailed numerical modeling of this design, we estimate that the proposed structure may achieve 1 MeV of energy gain over a distance less than 1 cm by sequentially illuminating 49 identical structures.

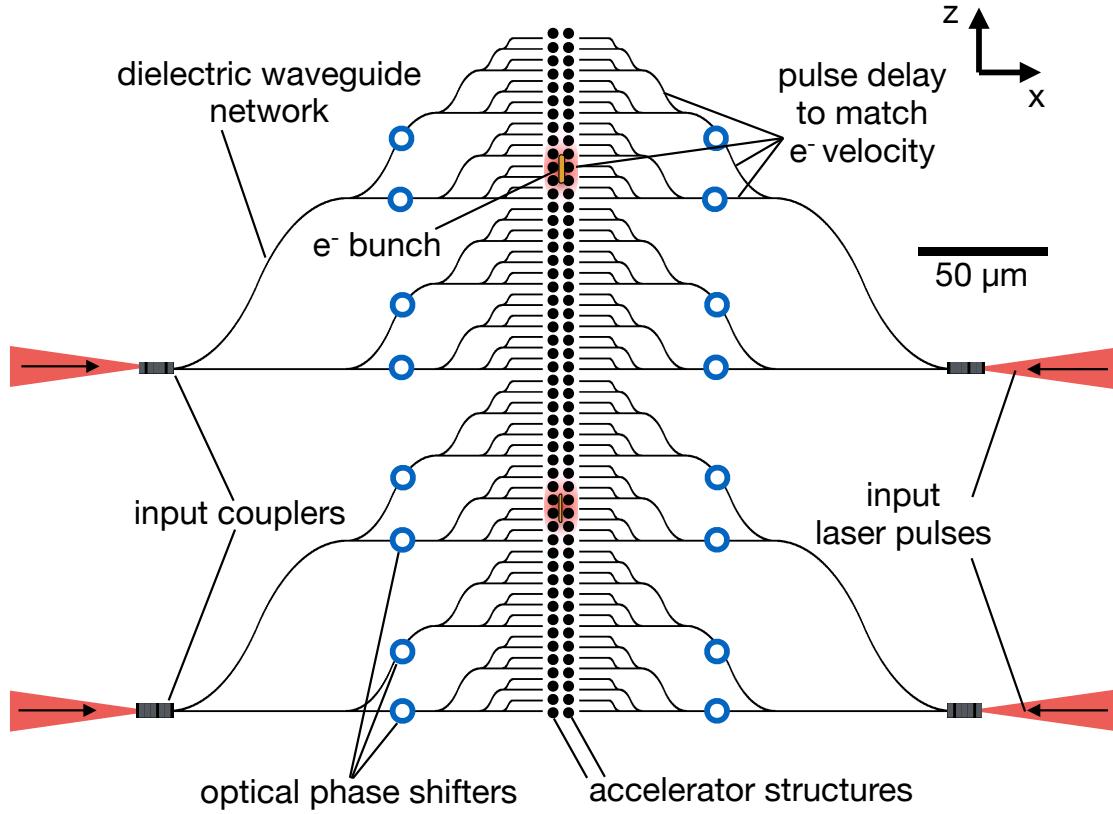


Figure 3.1: Two stages of the DLA laser coupling ‘tree-network’ structure. The electron beam travels along the z-axis through the center of this structure. The laser pulses are side coupled with optical power shown in red. Black regions define the on-chip waveguide network. Blue circles represent the optical phase shifters used to tune the phase of the laser pulse. This geometry serves to reproduce the pulse-front-tilt laser delivery system outlined in [33] in an integrated optics platform.

We first introduce the proposed *tree-network* waveguide geometry, which is diagrammed in Fig. 3.1. The electron beam to be accelerated is propagating along the z-axis in the central accelerator gap. We first couple the laser pulses to the on-chip dielectric waveguides by use of input

couplers. The optical power is then split a series of times and directed by waveguide bends to illuminate the entire length of the accelerator gap. A fractal waveguide geometry is chosen as it evenly illuminates the accelerator gap with minimal use of 50-50 splitters. Furthermore, the waveguide bends are designed such that the laser pulse arrival at the accelerator gap is delayed to coincide with the arrival of the electron bunch as it propagates through the structure.

Integrated phase shifters are used to tune the phase of each pulse upon exiting the waveguides and may be optimized for maximum acceleration. The accelerating structures are placed adjacent to the waveguide outputs. In this study, we choose to investigate silicon dual-pillar accelerator structures, similar to those used in [110]. The entire device is mirrored over the center plane and is driven by laser inputs on each side. Two stages of the structure are shown in Fig. 3.1, although several more may be implemented in series, assuming availability of several phase-locked laser sources. Electron beam focusing elements may be implemented between stages as needed.

### 3.1.1 Device Components

We will now discuss the individual components involved in the on-chip laser coupling system.

**Input Coupling:** The proposed structure first requires a strategy to couple light from the pump laser to the on-chip optical waveguides. We focus on free-space coupling to the input facet via a surface grating, eliminating the need for single mode fiber delivery. Our laser and macroscopic optical components are capable of handling pulse energies far beyond enough to cause damage to the structure. Bare single mode fibers also have damage thresholds high enough to withstand these laser pulses, but the large amount of dispersion introduced (associated with the relatively long length of  $> 1$  mm) will make them unsuitable for delivery to the chip.

In general, couplers must have (1) high coupling efficiency, (2) a bandwidth large enough to couple entire pulse spectrum, and (3) high power handling and minimized hot spots. Input coupling may be accomplished by use of end coupling, focusing the laser beam directly onto the waveguide cross section, or vertical coupling schemes, such as grating couplers. In SOI systems, end coupling can achieve insertion losses as low as 0.66 dB (85.9%) over a bandwidth of roughly 10 THz [158], but is cumbersome to perform experimentally for a large number of inputs and constrains the input and output coupling ports to be located on the edges of the chip. Vertical couplers provide the benefit of relative flexibility in alignment and positioning on chip. The coupling efficiency of these devices varies drastically depending on the complexity of the grating coupler design, from an efficiency of  $> 30\%$  to  $> 90\%$  [193]. However, highly efficient broadband couplers capable of sustaining large bandwidths still provide design challenges, with the state-of-the-art fully-etched structures able to provide 67% coupling efficiency with a 3 dB bandwidth of 60 nm at 1550 nm [46].

In the following parameter study, we assume a coupling power efficiency of 60% with a substantially wide bandwidth to accommodate that of our pulse (up to about 117 nm for a 50 fs pulse),

which is reasonably achievable with end coupling. Additional investigation into the design of ultra-broadband vertical couplers must be considered to guarantee coupling of the femtosecond pulsed lasers.

**Waveguides:** Waveguides are a critical component of laser coupling. Schematics of the waveguide cross-sections and their field distributions are shown in Fig. 3.2. We have explored two general classes of wave-guiding systems: (1) tightly confined systems and (2) weakly confined systems. Weakly confined waveguide modes have a small difference between mode effective index and cladding index, which results in the optical power being spread over a larger area and into the cladding material, which generally has preferable damage and nonlinearity properties. However, as we will discuss in the next section, our simulations show that weakly confined modes, with  $n_{\text{eff}} - n_{\text{core}}$  of about 0.1, have almost 0% power transmission for bend radii less than 10  $\mu\text{m}$ . In our tree-network structure, we require bend radii on this order to achieve the required pulse delay to matching to the electron bunch, therefore weakly guided waveguides were not considered for the particular tree-network structure in this parameter study.

We explored material systems of SOI and  $\text{Si}_3\text{N}_4/\text{SiO}_2$  due to their common use as waveguide core materials. SOI-based waveguides would be simpler to integrate with the silicon DLA structure and electron gun and there exists a much larger body of previous work on fabrication of silicon material systems for applications such as phase control, especially in the LIDAR community [215, 104]. However,  $\text{Si}_3\text{N}_4/\text{SiO}_2$  waveguides have favorable nonlinear and damage properties when compared to SOI. As mentioned, there are several other material systems that could also be explored for low loss, low nonlinearity, and high damage thresholds.  $\text{Ta}_2\text{O}_5$  [18] and  $\text{Ga}_2\text{O}_3$  are promising candidates that will be investigated in future studies.

To study waveguide nonlinearity, we solve a version of the nonlinear Schrödinger equation (NLSE), which is typically used for describing nonlinear propagation of a pulse of duration between 10 fs and 10 ns. In this particular treatment, the solution for the electric field is assumed to be of form in Eq. 3.1, where the slowly varying envelope approximation and separation of variables of the modal distribution  $F(x, y)$  and envelope  $A(z, t)$  are used [4].

$$\mathbf{E}(\mathbf{r}, t) = \frac{\hat{x}}{2} \{ F(x, y) A(z, t) \exp[i(\beta_0 z - \omega_0 t)] + \text{c.c.} \}, \quad (3.1)$$

where  $x, y$  are the transverse directions,  $z$  is the propagation direction,  $\beta_0$  is the propagation constant and  $\omega_0$  is the optical frequency.

The slowly varying envelop  $A(z, t)$  obeys the form of the NLSE given in Eq. 3.2, which can be

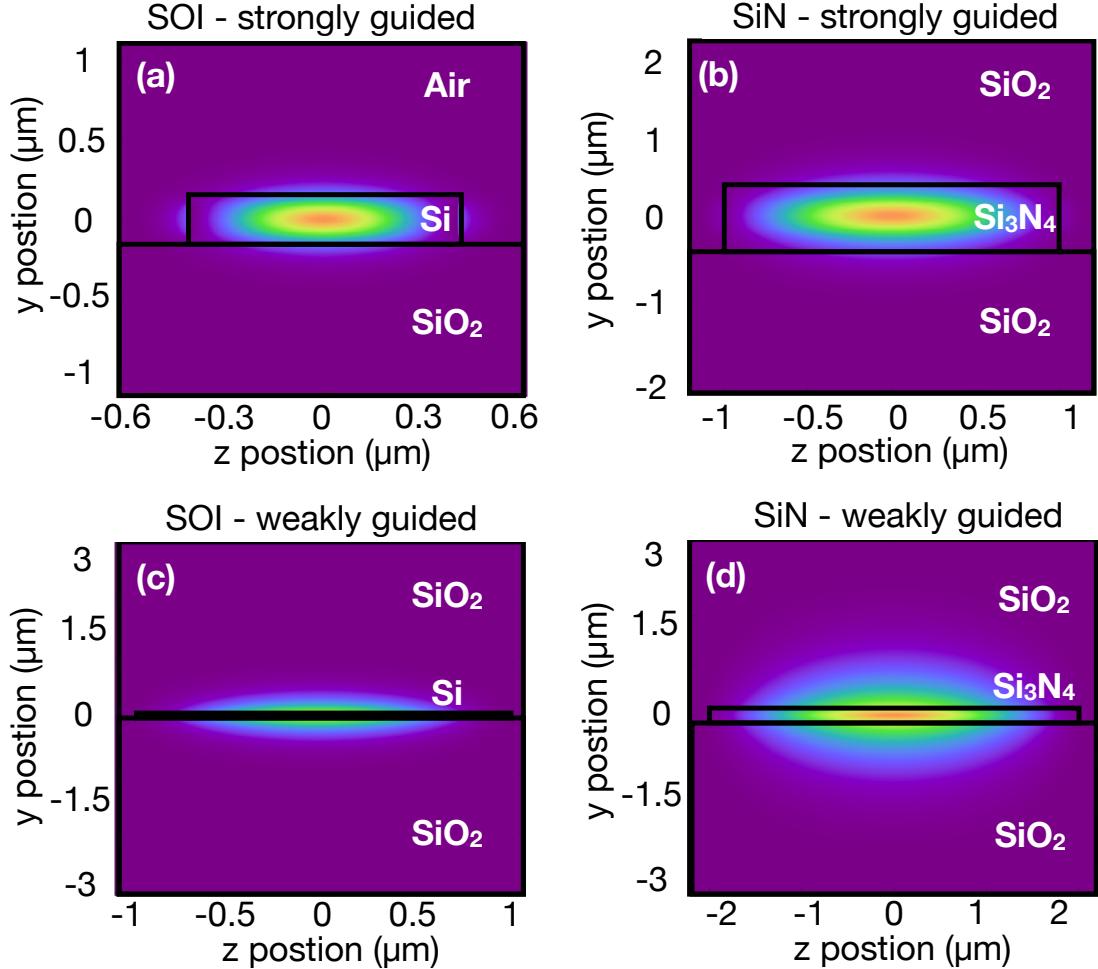


Figure 3.2: Waveguide geometries and corresponding horizontal electric field components [56]. (a-b) Strongly confined modes. (c-d) Weakly confined modes. (a) and (c) are SOI material platforms whereas (b) and (d) are Si<sub>3</sub>N<sub>4</sub>/SiO<sub>2</sub> materials. Waveguide core heights in (a-d) are given by 220, 400, 60, and 100 nm, respectively. Waveguide core widths are given by 0.78, 1.6, 2, and 4 μm, respectively.

solved by the split-step method [208].

$$\begin{aligned} \frac{\partial A}{\partial z} + \frac{\alpha}{2} A + \frac{i\beta_2}{2} \frac{\partial^2 A}{\partial T^2} - \frac{\beta_3}{6} \frac{\partial^3 A}{\partial T^3} \\ = i\gamma \left( |A|^2 A + \frac{i}{\omega_0} \frac{\partial}{\partial T} (|A|^2 A) - T_R A \frac{\partial |A|^2}{\partial T} \right), \end{aligned} \quad (3.2)$$

where  $T = t - z/v_g$  is the time in retarded frame with  $v_g$  being the group velocity,  $\gamma = 2\pi n_2/(\lambda A_{\text{eff}})$  is the nonlinear parameter per unit length and power, and  $A_{\text{eff}}$  is the effective modal area.  $T_R$  is

the Raman time constant and has an approximated value of 3 fs [11]. On the left hand side of this equation, the loss is incorporated into the second term with  $\alpha$  being the loss of the waveguide in units of  $m^{-1}$ . The 3rd and 4th terms indicate second and third order dispersion, with  $\beta_2$  and  $\beta_3$  being the respective dispersion coefficients. On the right hand side of the equation, the 1st term is SPM, the 2nd term is self-steepening, and the 3rd term is Raman scattering.

For our proposed structure, the overall length of the waveguide is short ( $\ll 1$  m), hence material loss  $\alpha$  can be neglected. The dispersion terms come from both the material dispersion and waveguide dispersion. These terms,  $\beta_{2,wg}$  and  $\beta_{3,wg}$ , can be obtained from numerically solving for effective refractive index as a function of wavelength  $n_{\text{eff}}(\lambda)$ , and are explicitly given as

$$\beta_{2,wg} = \frac{\lambda^3}{2\pi c^2} \frac{d^2 n_{\text{eff}}}{d\lambda^2}, \quad (3.3)$$

$$\beta_{3,wg} = -\frac{3\lambda^4}{4\pi^2 c^3} \frac{d^2 n_{\text{eff}}}{d\lambda^2} - \frac{\lambda^5}{4\pi^2 c} \frac{d^3 n_{\text{eff}}}{d\lambda^3}. \quad (3.4)$$

We note that the contribution of dispersion and SPM is generally compared through the  $N^2$  parameter [4]:

$$N^2 = \frac{L_D}{L_{\text{NL}}} = \frac{\gamma P_0 \tau^2}{|\beta_2|}, \quad (3.5)$$

where  $\tau$  is the pulse duration. When the dispersion length,  $L_D$ , is larger than nonlinear length,  $L_{\text{NL}}$ , SPM is dominant over dispersion and  $N^2 > 1$ . SPM is typically large in strongly-guiding and high  $n_2$  materials, such as the strongly-guiding SOI waveguide. For the weakly-guiding, lower  $n_2$  SiN waveguides, SPM will be less prominent, yet still larger than dispersive effects for the range of peak powers we consider. Using typical experimental parameters and examining the material considered in this study with the lowest nonlinearity (SiO<sub>2</sub>), we have:  $n_2(\text{SiO}_2) = 2.6 \text{ m}^2/\text{W}$ ,  $A_{\text{eff}} \sim 7 \mu\text{m}^2$ ,  $|\beta_2| = 76 \text{ fs}^2/\text{mm}$  and a peak power of  $P_0 = 80 \text{ kW}$ , the corresponding  $N^2 = 758$ , indicating that SPM is highly dominant over dispersion.

Alternatively, by turning on and off each term in Eq. 3.2 to investigate its contribution, we find that for both SOI and SiN cases, SPM is the dominant contribution to the nonlinearity, other terms does not yield a significant difference to the results for a propagation distance on the order of hundreds of  $\mu\text{m}$ .

**Splitters:** After the initial input coupling step, splitters are used to distribute the laser power along the DLA structure. Splitters further contribute to insertion loss, with experimental characterization of Y-splitters indicating losses on the order of 1 dB [225]. However, recent advances in inverse design and optimization techniques have allowed for new designs with much higher efficiencies. Using particle swarm optimization [48], devices have been produced with theoretical insertion losses of 0.13 dB and an experimentally determined value of  $0.28 \pm 0.02$  dB [225]. As even more sophisticated techniques of optimization have been developed, the insertion loss of simulated designs

has reached 0.07 dB [105]. Adjoint-based optimization methods have been further expanded to enforce fabrication constraints on the permitted designs, thus allowing one to expect greater agreement between simulated and fabricated structures [149]. As a consequence of the rapid progress made in this field and the efforts to ensure robustness of device to fabrication tolerance, we have used an insertion loss per splitter of 0.22 dB, or 95% efficiency, for the following parameter study.

**Bends:** Waveguide bends are necessary to provide group delay to the laser pulses to ensure their arrival at the accelerator gap at the same time as a moving electron beam. Here, using the circular bending scheme drawn in Fig. 3.3, we calculate the required geometry needed to satisfy this condition. Specifically, for an electron speed  $\beta c_0$ , an electron propagation distance  $h$ , and waveguide group index  $n_g$ , we seek to set a condition on the radius  $R$  to accomplish this. First, we may establish the value of the bend angle ‘ $\theta$ ’ as

$$\theta = \begin{cases} \cos^{-1}(1 - h/2R) & \text{if } h < 2R \\ \pi/2 & \text{if } h \geq 2R \end{cases}. \quad (3.6)$$

When  $h \geq 2R$ , we use two 90-degree bends and extend the intermittent length with a vertical waveguide section. From this, we can express the horizontal distance  $d$  as

$$d = 2R \sin(\theta), \quad (3.7)$$

and the total length of the bent waveguide as

$$L = \begin{cases} 2R\theta & \text{if } h < 2R \\ h + (\pi - 2)R & \text{if } h \geq 2R \end{cases}. \quad (3.8)$$

To now set a condition on  $R$ , we insist that the pulse timing delay between the curved waveguide and the straight waveguide is equal to the time needed for the electron to travel a distance  $h$ . The difference in length between the curved waveguide and straight waveguide is simply  $L - d$ , thus the timing delay of the pulse is given by

$$\begin{aligned} \Delta t_{\text{pulse}} &= \frac{n_g}{c_0}(L - d) \\ &= \frac{n_g}{c_0} \begin{cases} 2R(\theta - \sin(\theta)) & \text{if } h < 2R \\ h + R(\pi - 4) & \text{if } h \geq 2R \end{cases}. \end{aligned} \quad (3.9)$$

The electron has a velocity of  $\beta c_0$ , so it’s timing delay is given by

$$\Delta t_{e^-} = \frac{h}{\beta c_0} \quad (3.10)$$

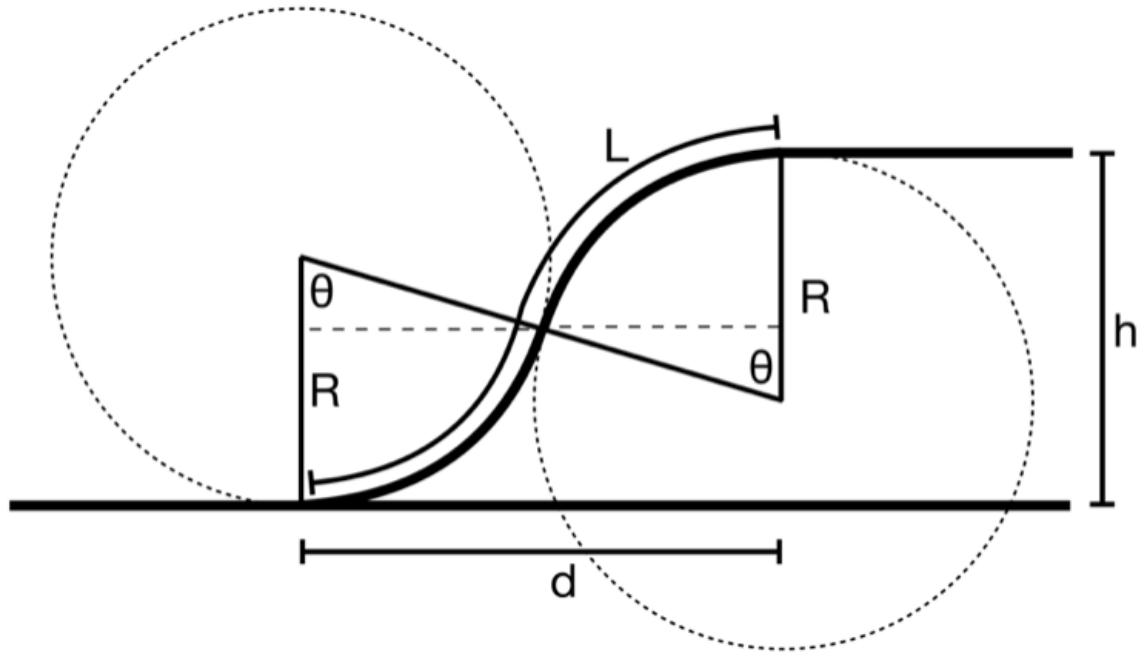


Figure 3.3: Diagram of a single bend in the tree-network structure with an optical pulse incident from the left. The bend has radius  $R$ , accomplishes a vertical climb of  $h$  over a horizontal distance  $d$ . The total length of the bent section is  $L$ . The electron travels from bottom to top in this configuration. We wish to find an  $R$  such that an optical pulse traveling through the bent section is delayed by the same amount of time for the electron to travel the vertical distance  $h$ .

Setting these two equal and solving for ‘ $R$ ’, we find that

$$R = \frac{h}{\beta n_g} \begin{cases} 2(\theta - \sin(\theta))^{-1} & \text{if } h < 2R \\ \frac{\beta n_g - 1}{4 - \pi} & \text{if } h \geq 2R \end{cases}. \quad (3.11)$$

Thus, for extended interaction lengths where  $h \gg 2R$ , we require that  $\beta n_g > 1$  for a positive (and physical) solution for  $R$ . Equivalently, for low  $\beta$ , we require large  $n_g$  in order to sufficiently delay the pulse in order to match the low electron velocity.

With this geometry, there is thus a condition on the group index of the waveguide system that may achieve the required delay given an electron speed

$$n_g \beta \geq 1. \quad (3.12)$$

Thus, for sub-relativistic electrons ( $\beta < 1$ ), higher index materials are required for the waveguides. For example, for a  $\beta$  of  $1/3$ , a group index of  $n_g > 3$  is required, which may not be satisfied by a standard SiN waveguide geometry. Thus, in sub-relativistic regimes, SOI waveguides are the optimal

choice.

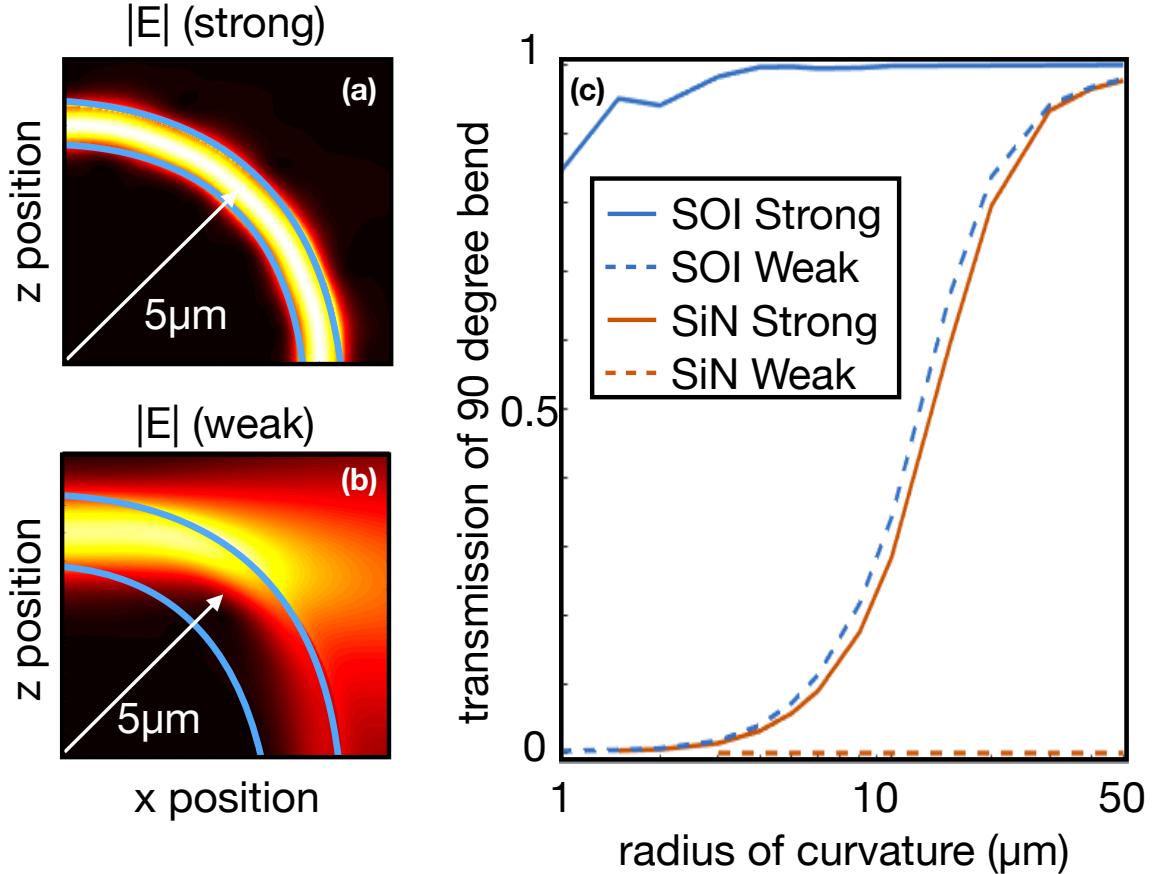


Figure 3.4: (a) Electric field amplitude for a strongly guiding SOI waveguide. (b) Electric field amplitude for a weakly guiding SOI waveguide. (c) Comparison of bending loss as a function of bend radius for the 4 waveguides from Fig. 3.2.

Fig. 3.4 shows the optical power transmission through a series of bends and waveguide geometries using the finite-difference frequency-domain method (FDFD) [174] and an established two-dimensional approximation to the three-dimensional structure [180]. For tightly confined SOI waveguide modes, the bending radius can reach as low as 2  $\mu\text{m}$  before there is significant loss. However, for weakly confined SOI modes and strongly confined SiN modes, the power transmission is less than 50% until the radius exceeds 20  $\mu\text{m}$ . For our purposes, this kind of bending loss is unacceptable as radii on the order of 10  $\mu\text{m}$  are required close to the DLA structure to perfectly match the electron velocity. However, if we relax the delay requirement in favor of larger bend radii, we may still use strongly confined SiN modes. Based on a simple calculation following the bending radius discussion, if we wish to keep all SiN waveguides above 40  $\mu\text{m}$  radius of curvature, we will experience a 25 fs

mismatch in peak pulse arrival to electron arrival. For a pulse duration of 250 fs, this will have negligible effect on the acceleration gradient. Therefore, in our parameter study, we assume strongly confined waveguide modes and bends that are large enough to achieve transmission of 95%. Many of these issues may be reconciled by choosing a hybrid waveguide system, as shown in Fig. 3.9, in which different materials and waveguide modes are used at different distances from the central DLA structure. We did not consider these options directly in our following parameter study.

**Phase Shifters:** Phase shifters are an essential component in the DLA system for ensuring proper phase matching between the electrons and photons. While it is simple to do phase tuning in free-space for a single stage DLA with macroscopic delay stages, waveguide-integrated phase shifters for long interaction or multi-stage DLAs will be experimentally complicated. To achieve a sizable energy gain and gradient over a given interaction length, a high level of precision and stability in the phase of each section is required.

To illuminate the importance of precision phase shifters, a Monte Carlo simulation was performed in which the output phase of each waveguide was perturbed from its optimal value by a random amount. This study found that, for a stage length of 1 mm, phase stability and precision of greater than 1/100 of a radian (0.16% of a cycle) was required to achieve sustained energy gain within 90% of the maximum achievable amount.

There are a few strategies to implement integrated phase shifters, including the use of (1) thermal/thermal-optic effect [104, 103], (2) electro-optic effect, and (3) mechanical techniques, such as piezo controlled elements [154]. For this application, we will require a full  $2\pi$  range of phase control of each output port with a resolution of 1/100 of a radian, and a modulation bandwidth of  $\sim 1$  kHz to correct for environmental perturbations.

Rather than supplying each waveguide output port with a phase shifter with these properties, it may be possible to have dedicated ‘fine’ and ‘coarse’ phase shifters as we move through the splitting structure. Furthermore, some degree of relative fixed phase between output ports may be accomplished by precision fabrication.

To further mitigate the challenges associated with operating these multiple phase shifters during acceleration, we may implement a feedback control loop, which is described in Fig. 3.5. In this setup, the quantity of interest, such as electron energy gain, can be measured at the end of a section and optimized with respect to the individual phase shifters in the power delivery system without explicit knowledge of the electron beam dynamics.

**DLA Structures:** We assume silicon dual-pillar DLA structures in the following parameter study, but the choice is arbitrary and can be changed to other materials or designs depending on the fabrication constraints. In Fig. 3.6, we show an example of the setup considered in the parameter study, simulated with FDFD. The pillar radius is 981 nm and the gap width is 400 nm. 3 periods of DLA are powered by a single waveguide and periodic boundary conditions are used in the z

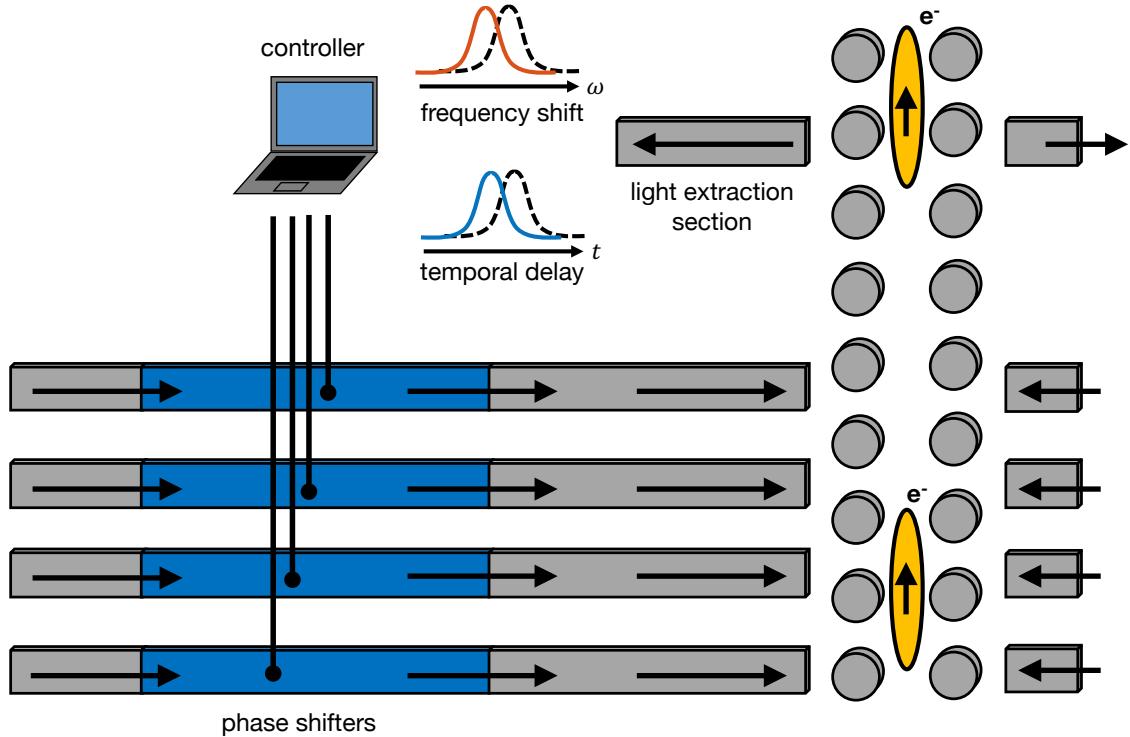


Figure 3.5: Idealized schematic of a feedback system for automatic phase control. A dedicated light extraction section is added to the accelerator. Light is radiated from the electron beam transversing the DLA structures and the frequency content and/or timing of the light is sent to a controller. The phase shifts of each waveguide are optimized with respect to either the frequency or the delay of the signal.

direction. Wakefields and transverse deflections are ignored for simplicity as these simulations are intended to provide an estimate of the resonant enhancement, acceleration gradient, and accelerator damage threshold. The waveguide refractive index was approximated using [180].

Resonant enhancement in the dual pillars is clearly visible and can be accomplished by optimizing the spacing and radius parameters. To investigate the effect of resonance in the DLA structures, here we derive the analytical form of the resonant field enhancement in the accelerator gap and verify that it is approximately proportional to  $\sqrt{Q}$ . The resonant nature of the acceleration structure can be described by coupled mode theory [69, 190]. We denote the amplitude of the resonant mode as  $s$ , where  $|s|^2$  represents the energy stored in the resonant mode, and the amplitudes of incoming and outgoing waves as  $a$  and  $b$  respectively, where  $a^\dagger a$  and  $b^\dagger b$  represent the power of incoming and

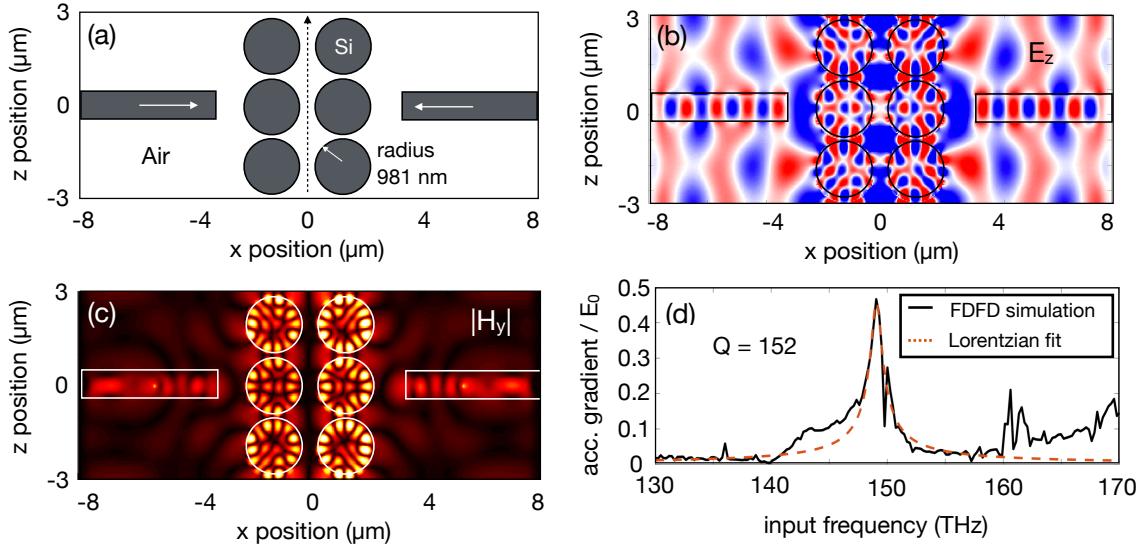


Figure 3.6: (a) A schematic of the waveguide to DLA connection. Silicon dual pillars of optimized radius of 981 nm and gap size of 400 nm are used. (b) The accelerating electric field during one time step. (c) Absolute value of the transverse magnetic field. (d) Absolute value of the acceleration gradient as a function of frequency, normalized by the peak electric field in the waveguide. A Lorentzian line shape is fit to the square of this plot. The square root of this fit is shown in red. Based on the Lorentzian fit, a Q-factor of  $152 \pm 29$  was determined. As computed following the derivation in [152], but with the waveguide mode impedance and effective area in place of the plane wave values, this structure has a shunt impedance,  $Z_S$ , of  $449.1 \Omega$  over 3 periods and a  $Z_S/Q$  value of  $2.95 \Omega$ .

outgoing waves. The dynamics of the resonant mode can be described as

$$\frac{ds}{dt} = (-i\omega_0 - \gamma_s)s + \boldsymbol{\kappa}^T \mathbf{a}, \quad (3.13a)$$

$$\mathbf{b} = B\mathbf{a} + s\mathbf{d}, \quad (3.13b)$$

where  $\omega_0$  is the resonant frequency of the acceleration mode,  $\gamma_s$  is the leakage rate resulting from the coupling to outgoing waves, and  $B$  is the background scattering matrix including direct pathways.  $\boldsymbol{\kappa}$  and  $\mathbf{d}$  are coupling coefficients for incoming and outgoing waves. In a reciprocal system with lossless materials [190], which is the case for DLA,

$$\boldsymbol{\kappa} = \mathbf{d}, \quad (3.14a)$$

$$\mathbf{d}^\dagger \mathbf{d} = 2\gamma_s. \quad (3.14b)$$

The periodic acceleration structure has two channels for incoming and outgoing propagation

waves to couple to the resonant mode. As the acceleration mode is an even mode which has non-vanishing longitudinal electric field at the mirror plane, the incident wave from the left and right should have equal amplitude and phase to efficiently excite the acceleration mode. Thus,  $\mathbf{a} = [1, 1]^T a_1$ , where  $|a_1|^2$  represents the power of incoming waves from the left channel. The even mode also couples equally to the left and right channel. As a result,  $\mathbf{d} = [1, 1]^T d_1$ , where  $d_1$  is the coupling coefficient for the outgoing waves in the left channel and  $|d_1|^2 = \gamma_s$  according to Eq. 3.14.

From Eq. 3.13 we obtain  $s(\omega) = \frac{\kappa^T \mathbf{a}^+(\omega)}{-i(\omega - \omega_0) + \gamma_s}$  for each frequency component. Based on the preceding analysis, the spectrum of energy stored in the resonant mode is

$$|s(\omega)|^2 = \frac{4\gamma|a_1^+(\omega)|^2}{(\omega - \omega_0)^2 + \gamma_s^2}. \quad (3.15)$$

To give an explicit expression about the field enhancement, we denote the maximum electric field amplitude at the output port of the power delivery waveguide as  $E_{\text{out}}$  and the maximum amplitude of the electric field inside acceleration structure as  $E_{\text{mat}}$ . We introduce the effective incident spot area ( $S$ ) such that the incident power from, say, the left channel is  $\frac{1}{2\eta_0}|E_{\text{out}}|^2 S$ , and define the mode volume ( $V$ ) of the resonant mode so that the energy stored in the resonant mode is  $\frac{1}{2}\epsilon_r\epsilon_0|E_{\text{mat}}|^2 V$  [86], where  $\epsilon_r$  is the relative permittivity of the dielectric accelerator. Thus,

$$|E_{\text{mat}}(\omega)| = \left( \frac{8cS}{\epsilon_r\omega_0 V} \right)^{1/2} \left[ \frac{\gamma_s^2}{(\omega - \omega_0)^2 + \gamma_s^2} \right]^{1/2} \sqrt{Q}|E_{\text{out}}(\omega)|, \quad (3.16)$$

where the quality factor  $Q$  is inversely proportional to the resonant mode leakage rate, i.e.  $Q = \frac{\omega_0}{2\gamma_s}$ . Eq. 3.16 shows that the field enhancement in the resonant accelerator structure is proportional to  $\sqrt{Q}$  and has a bandwidth that decreases with increasing  $Q$ , where the frequency dependence is the square root of a Lorentzian line shape. From our Lorentzian fit, a  $Q$  value of  $152 \pm 29$  was determined.

From the field plots of Fig. 3.6, it is also clear that the two surrounding DLA cells are slightly out of phase with the center cell. This effect is caused by the lack of translational symmetry in the input optical beam in the  $z$  direction and will lower the acceleration gradient.

Coupling efficiently from waveguides to DLA structures may be done by optimizing the structure parameters. For an optimized structure, back reflection may be minimized. Furthermore, one waveguide is generally able to power multiple DLA periods. However, simulations suggest that additional periods of DLA per waveguide do not significantly increase the total energy gain achievable from a single waveguide. Thus, the spacing between waveguides must be large enough to eliminate cross-talk, but small enough to ensure high acceleration gradients.

It will be of great importance in future experiments to integrate the waveguide system and the DLA structure on the same chip. Thus, the height of the pillar structure may be constrained to be equal to that of the waveguide core and 500 nm thick SOI platforms may be a good starting point

for testing these integrated systems.

### 3.1.2 Practical Considerations

Here we discuss practical considerations of this system and formalize the calculation of the acceleration gradient used in the following parameter study. First, we show how to reason about the acceleration gradient with both arbitrary, finite-duration input pulses and finite stage length structures. In both derivations, we assume an input pulse  $E_0(t)$ , which leads to the creation of an accelerating field in the gap of a unit cell  $E_z(z, t)$  through the convolution with the corresponding impulse response function  $f(z, t)$ . In the frequency domain, this is done via multiplication of the pulse spectrum  $E_0(\omega)$  with the transfer function  $F(z, \omega)$

$$E_z(z, t) = E_0(t) * f(z, t) \quad (3.17)$$

$$E_z(z, \omega) = E_0(\omega)F(z, \omega). \quad (3.18)$$

**Finite pulse duration:** Here we wish to derive the correspondence between the time-domain description of the acceleration gradient, given an arbitrary input pulse, and the frequency domain approach that is used in the previous Chapter and other works [77, 152].

In the time domain, the acceleration gradient is expressed as an integral over the accelerating electric field over the particle's trajectory.

$$G = \frac{1}{L} \int_{-L/2}^{L/2} dz E_z(z, t(z)). \quad (3.19)$$

If the electron moves uniformly in  $\hat{z}$  with speed  $\beta c_0$ , then  $z(t) = z_0 + \beta c_0 t$  and we may express the acceleration gradient as a function of the starting time,  $t_0$ , as

$$\begin{aligned} G(t_0) &= \frac{1}{L} \int_{-L/2}^{L/2} dz E_z(z, t_0 + z/\beta c_0) \\ &= \frac{1}{L} \int_{-L/2}^{L/2} dz \int_{-\infty}^{\infty} dt E_z(z, t) \delta(t - t_0 - z/\beta c_0). \end{aligned} \quad (3.20)$$

In previous works, such as Ref. 152, the acceleration gradient is computed by first performing a *finite-difference time-domain* (FDTD) simulation to record  $E_z(z, t)$  along the gap for a series of time, and then maximizing the integral in Eq. 3.20 with respect to  $t_0$ . However, we may equivalently do the computation in the frequency domain by Fourier transforming this equation with respect to

$t_0$ , which yields

$$\begin{aligned} G(\omega) &= \frac{1}{L} \int_{-L/2}^{L/2} dz \int_{-\infty}^{\infty} dt E_z(z, t) e^{i\omega(t-z/\beta c_0)} \\ &= \frac{1}{L} \int_{-L/2}^{L/2} dx e^{-i\omega z/\beta c_0} \int_{-\infty}^{\infty} dt E_z(z, t) e^{i\omega t} \\ &= \frac{1}{L} \int_{-L/2}^{L/2} dz e^{-i\omega z/\beta c_0} E_0(\omega) F(z, \omega) \\ &\equiv g(\omega) E_0(\omega) \end{aligned} \quad (3.21)$$

Here  $g(\omega)$  is the gradient normalized by the incident electric field at that frequency,  $E_0(\omega)$ , which is also described in the following subsection. Now, by performing a series of FDFD simulations at discrete frequencies, we may estimate  $F(z, \omega)$ . Then, using the known pulse amplitude spectrum and phase information in  $E_0(\omega)$ , we can compute  $G(\omega)$  as described. Finally,  $G(t_0)$  can be determined by applying a inverse discrete Fourier transform on  $G(\omega)$ , and the acceleration gradient can then be found by taking the maximum of the absolute value of this quantity. Explicitly,

$$G = \max_{t_0} |\mathcal{F}^{-1}\{g(\omega)E_0(\omega)\}|. \quad (3.22)$$

**Finite stage length:** Now, let us assume that we have a DLA with a stage length of  $L$  along  $\hat{z}$  with an incident laser pulse of the form  $E_0(t)$  with spectrum  $E_0(\omega)$ . The laser is assumed to be spatially uniform along the entire interaction length. We use the same transfer function formalism as introduced at the beginning of this section.

The DLA structure is further assumed to be periodic in  $\hat{z}$  with a periodicity of  $\Lambda_z = \beta\lambda = 2\pi c_0/\omega_0$ . Thus, the fields can be expressed as a Fourier series.

$$E_z(z, \omega) = E_0(\omega) \sum_{m=-\infty}^{\infty} T_m(\omega) e^{imz\omega_0/\beta c_0} \quad (3.23)$$

where the  $T_m(\omega)$  terms are the spatial Fourier amplitudes of the transfer function  $F(z, \omega)$ . See Ref. [150] for a similar discussion.

The acceleration gradient at frequency  $\omega$ ,  $G(\omega)$ , can be written as the average  $E_z$  felt by the particle as it moves with velocity  $\beta c_0 \hat{z}$  through the entire interaction length of the structure from  $z = -L/2$  to  $z = L/2$ .

$$\begin{aligned} G(\omega) &= \frac{1}{L} \int_{-L/2}^{L/2} dz E_z(z, \omega) e^{iz\omega/\beta c_0} \\ &= \frac{1}{L} \int_{-L/2}^{L/2} dz E_0(\omega) \sum_{m=-\infty}^{\infty} T_m(\omega) e^{i(m\omega_0 + \omega)z/\beta c_0}. \end{aligned} \quad (3.24)$$

Rearranging the integral and defining the normalized gradient  $g(\omega) \equiv G(\omega)/E_0(\omega)$ ,

$$\begin{aligned} g(\omega) &= \frac{1}{L} \sum_{m=-\infty}^{\infty} T_m(\omega) \int_{-L/2}^{L/2} dz e^{i(m\omega_0 + \omega)z/\beta c_0} \\ &= \sum_{m=-\infty}^{\infty} T_m(\omega) \frac{2\beta c_0 \sin\left(\frac{L}{2\beta c_0}(m\omega_0 + \omega)\right)}{L(m\omega_0 + \omega)} \\ &= \sum_{m=-\infty}^{\infty} T_m(\omega) \operatorname{sinc}\left(\frac{L}{2\beta c_0}(m\omega_0 + \omega)\right). \end{aligned} \quad (3.25)$$

We reasonably assume that the input pulse power is centered around  $\omega_0$ . In this case, then only the  $m = -1$  will contribute to the accelerating mode. We could have also chosen a higher order  $m = -2, -3, \dots$  for the accelerating mode, as was demonstrated previously [26, 117], but  $m = -1$  is chosen for simplicity. Thus, as the interaction length increases, the  $\operatorname{sinc}()$  function becomes more tightly centered around  $\omega = \omega_0$ . This has the effect of limiting the available bandwidth of the input pulse.

Under this assumption, the final form of the normalized gradient becomes

$$g(\omega) = T_{-1}(\omega) \operatorname{sinc}\left(\frac{L}{2\beta c_0}(\omega - \omega_0)\right). \quad (3.26)$$

Assuming  $T_{-1}(\omega)$  is relatively constant over a bandwidth larger than our input pulse, then we see that the gradient falls to zero at  $\omega = \omega_0 \pm \frac{2\pi\beta c_0}{L}$ . For a Gaussian pulse of duration  $\tau$  with a time-bandwidth product of 0.44, the gradient would fall to zero at

$$L = \tau \frac{4\pi\beta c_0}{0.44}. \quad (3.27)$$

For a  $\tau$  of 250 fs and  $\beta$  of 1, this corresponds to a stage length of 2.14 mm. Thus, to satisfy the bandwidth requirement,  $L$  must be much less than 2.14 mm if no pulse delay techniques are used.

This result can be compared to the following back-of-the-envelope calculation: An electron traveling over a length  $L$  with speed  $\beta c_0$  will spend  $\Delta t_{e^-} = \frac{L}{\beta c_0}$  of time in the channel. The input pulse will spend approximately  $\tau$  seconds in the gap. Thus, for the fields to be present during the whole duration

$$L < \tau\beta c_0. \quad (3.28)$$

This scales with  $\tau$ ,  $\beta$ , and  $c_0$  in the same fashion as Eq. 3.27, which serves as a sanity check. However, the full expression in Eq. 3.26 can be used to rigorously compute the effect that a finite stage length structure will have on the acceleration gradient.

**Beam Loading and Longitudinal Wakes:** Here we consider the effect of feedback from electron radiation in the device. The fundamental unit cell of the proposed accelerator design, depicted in

Fig. 3.6, consists of a structure segment of three periods  $\Delta z = 3\lambda$  fed by a single laser pulse of the multi-branch network with duration  $\tau = 250$  fs. It is shown in Ref. 152 that the coupling efficiency of the laser field to a point charge  $q$  for the side-coupled geometry used here is analogous to Eq. (7) of Ref. 176, which considers a traveling wave mode in a cylindrical structure with group velocity  $\beta_g c$ , under the substitution  $\beta_g/(1 - \beta_g) \rightarrow \Delta z/\tau c$ , which gives a coupling efficiency  $\eta_q = qG\Delta z/P\tau$ . Here  $P$  is the laser mode power and  $G = G_0 - G_H$  is the loaded gradient where  $G_0$  is the unloaded value and  $G_H$  is a retarding field that accounts for the longitudinal wake induced in the structure by the beam. These may be written

$$G_0 = \sqrt{\frac{Z_C P}{\lambda^2}} \quad , \quad G_H = \frac{qcZ_H}{\lambda^2}. \quad (3.29)$$

where  $Z_C$  is the characteristic impedance and  $Z_H$  is the Cherenkov wake impedance. A conservative approximation for the latter  $Z_H \approx \pi Z_0 \lambda^2 / (16a^2)$  is provided by Ref. 15 for the case of a flat (2D) geometry with a beam charge  $q$  in a narrow channel, where  $Z_0$  is the impedance of free space and we take  $a = 200$  nm to be the half-width of the accelerating channel. The resulting efficiency  $\eta_q$  is then quadratic in the charge  $q$ . Solving for the maximal value gives optimal bunch charge and efficiency

$$q_{\text{opt}} = \frac{G_0 \lambda^2}{2cZ_H} \quad , \quad \eta_{q_{\text{opt}}} = \frac{1}{4} \frac{\Delta z}{c\tau} \frac{Z_C}{Z_H}. \quad (3.30)$$

For the present case, with  $Z_C = 149 \Omega$ ,  $Z_H = 7402 \Omega$ ,  $G_0 = 108 \text{ MV/m}$ , we obtain  $q_{\text{opt}} \approx 0.1 \text{ fC}$  and  $\eta_{q_{\text{opt}}} \approx 0.04\%$ , corresponding to a retarding gradient  $G_H = 54 \text{ MV/m}$  and thus a beam loaded gradient  $G = G_0/2$ . The optimal charge corresponds to 608 electrons, which is consistent with achieved laser-triggered emission from nanotip electron sources. As shown in [136] under multi-bunch operation with structures designed for higher gradients, efficiencies can theoretically be in the tens of percents. The structure design considered here was intended to illustrate the basic principles of constructing a multi-guided wave system and was not optimized for efficient beam coupling. Even so, efficiencies this order are still acceptable for possible near-term applications, such as a 1 to 10 MeV medical linac, where requisite beam powers are less than 1 Watt.

**Heat Dissipation:** For the pulse parameters used in the following parameter study, the laser input pulse energy at each stage of length  $L = 192 \mu\text{m}$  is  $E_p = 11 \text{ nJ}$ . We assume a repetition rate  $f_{\text{rep}} = 10 \text{ MHz}$ , which is consistent with commercially available solid state fiber lasers at micro-Joule pulse energies. Given that there are two input laser couplings per stage of length  $L$  in the configuration of Fig. 3.1, the average laser power per unit length of accelerator is  $dP/dz \approx 11 \text{ W/cm}$ .

Making a conservative assumption that all of this power passes through solid silicon, which has an absorption coefficient of  $\alpha_{\text{Si}} = 0.027 \text{ cm}^{-1}$  at  $\lambda = 2 \mu\text{m}$ , the corresponding absorbed power is of order  $6 \text{ mW/cm}^2$ . This is more than 5 orders of magnitude lower than the technological limit for heat dissipation from planar surfaces where  $1 \text{ kW/cm}^2$  is typical [49, 165]. Prior work has

shown that near-critical coupling to silicon dielectric accelerator structures using SOI waveguides is possible with appropriate phase adjustment to produce a traveling wave match between input and output couplers [214]. The latter work was for a structure design based on a 3D photonic crystal, but illustrates the principle that more sophisticated power handling techniques can potentially be employed in future designs to remove laser power from the wafer and safely dump it away from the accelerator.

### 3.1.3 Parameter Study

Here we estimate the performance of the device described in the previous section using working parameters consistent with that discussion. We will also use this analysis to investigate the interplay between various constraints that exist in the system. This will allow us to gain important intuition about the general working principles of an integrated accelerator on a chip, which may guide future studies.

#### Constraints

In the analysis of our system, we consider four main factors that will ultimately limit the acceleration gradients and energy gains attainable.

**1. Laser-induced damage of the DLA and waveguide materials:** To avoid damage of the structure, the electric fields in the system may never exceed the damage thresholds of the dielectrics used. The laser damage threshold for dielectric materials is highly favorable at short pulse durations, with sustainable peak powers that scale roughly as  $\tau^{-1/2}$  for  $\tau > 1$  ps and approach  $\tau^{-1}$  scaling for fs pulses [189, 188]. Amongst the materials considered in this study,  $\text{SiO}_2$  has the highest damage fluence threshold of  $2.5 \text{ J/cm}^2$  at 800 nm wavelength, followed by  $\text{Si}_3\text{N}_4$  at  $0.65 \text{ J/cm}^2$  and Si at  $0.18 \text{ J/cm}^2$  [182]. For a 100 fs pulse propagating in vacuum, these correspond to peak fields of 13.7, 7.0, and 3.7 GV/m, respectively.

**2. Optical nonlinearities in the materials:** Optical nonlinear effects may be encountered by the optical pulse as it propagates through the waveguides, which may cause significant pulse distortion, resulting in either damage or dramatic reduction of the acceleration gradient. Through the full treatment given earlier, we found that the most prominent nonlinear effect in our structure is self-phase modulation (SPM). For a pulse with a given peak power, the effects of SPM scale in proportion to the lengths of the waveguide sections.

**3. Power loss:** The tree-network structure introduces several sources of power loss: (1) input coupling loss, (2) splitting loss, (3) bending loss, and (4) waveguide scattering loss. Waveguide power loss due to scattering must be considered for structures with stage lengths greater than the

cm scale [216]. However, we neglect scattering loss in this study because the waveguide sections in question are shorter than one millimeter.

**4. DLA structure bandwidth:** As discussed, the DLA structures are designed to resonantly enhance the optical fields. The field enhancement is proportional to the square root of the quality factor of the DLA structures (similar to an optical cavity) which can be approximated by a Lorentzian spectrum. This resonance is used to increase the acceleration gradient while avoiding damage at the input facet. However, if the pulse bandwidth is large with respect to the bandwidth of the accelerator, the pulse will not efficiently couple into the DLA structure.

### Findings

With the system components and constraints introduced, we now present a parameter study to understand the fundamental trade-offs and optimal working parameters of an on-chip optical power delivery system for DLA of this class. A software package [76] was written to separately simulate each component and combine the results to generate an estimate for the acceleration gradient and energy gain assuming a set of parameters, which are outlined in Table 3.1 and chosen based on the considerations outlined at the beginning of this chapter.

Table 3.1: Parameters assumed in the study.

Parameter	Symbol	Value	Units
Wavelength	$\lambda$	2	$\mu\text{m}$
Electron speed / speed of light	$\beta$	1	-
DLA periods per waveguide	M	3	-
Input coupler efficiency	$\eta_c$	0.6	-
Splitting efficiency	$\eta_s$	0.95	-
Bending efficiency	$\eta_b$	0.95	-
Accelerating gradient at $Q = 1$	$G_{Q=1}$	0.0357	$E_0$
Input coupler - first split length	$L_0$	10	$\mu\text{m}$
DLA pillar radius	$R_{\text{pillar}}$	981	nm
DLA acceleration gap material / gap field enhancement factor	$d$	400	nm
	$f_m$	2	-

For a given pulse duration ( $\tau$ ) and DLA quality factor ( $Q$ ), the minimum peak electric field of the input pulse ( $E_0$ ) required to encounter each damage or nonlinearity constraint are modeled using the assumptions described below:

**(1) Input damage:** Fields at the input will be damaged if they exceed the damage threshold of the coupling material ( $E_d$ ). Thus, we enforce that

$$E_0 < E_d(\tau). \quad (3.31)$$

**(2) Accelerator damage:** With a given tree-network structure, we introduce a total of  $N_s$

separate 1→2 power splits for an input pulse. Furthermore, we introduce some optical power loss characterized by the power efficiencies of the input coupler ( $\eta_c$ ), splitters ( $\eta_s$ ), and bends ( $\eta_b$ ). Thus, the field at the output port of the laser coupling structure ( $E_{out}$ ) is given by

$$E_{out} = E_0 \left( 2^{-N_s} \eta_c \eta_s^{N_s} \eta_b^{N_s} \right)^{1/2}. \quad (3.32)$$

As we showed earlier, resonance in the DLA structures with quality factor  $Q$  will lead to a field enhancement in the accelerator gap that scales as  $\sqrt{Q}$ . Since our damage will be caused by the maximum field in the DLA materials, we assume there is another constant factor,  $f_m$ , relating the maximum field in the DLA material to the average field in the accelerator gap. From simulations, we estimate the value of  $f_m$  to be 2. Thus, the maximum field in the DLA material is

$$\begin{aligned} E_{mat} &= E_{out} f_m \sqrt{Q} \\ &= E_0 f_m \sqrt{Q} \left( 2^{-N_s} \eta_c \eta_s^{N_s} \eta_b^{N_s} \right)^{1/2}. \end{aligned} \quad (3.33)$$

We require the maximum field in the DLA material to be lower than the damage threshold, giving the constraint that

$$E_0 < E_d(\tau) \frac{2^{N_s/2}}{f_m \sqrt{Q}} \left( \eta_c \eta_s^{N_s} \eta_b^{N_s} \right)^{-1/2}. \quad (3.34)$$

**(3) Self-phase modulation:** For a wave of power  $P_0$  and wavelength  $\lambda$  traveling a distance  $L$  in a material with cross sectional area  $A$ , and nonlinear refractive index  $n_2$ , the accumulated SPM phase is given by [199]

$$\Delta\phi_{SPM} = 2\pi \frac{n_2 P L}{A \lambda}. \quad (3.35)$$

Since the optical power in our waveguides have optical power traveling in several materials, each with a different nonlinear refractive index, we define an effective  $n_2$  for modeling that is given by

$$n_2^{(eff)} = \frac{1}{P^{(tot)}} \sum_{j=1}^{\text{num. mat.}} n_2^{(j)} P^{(j)}, \quad (3.36)$$

where  $P^{(tot)}$  is the total optical power carried by the waveguide and  $P^{(j)}$  is the amount of power traveling in material ‘ $j$ ’.

Furthermore, the optical power is being split in half at each bend, so we must take this into account in our SPM calculation. Taking into account the losses in our system, the final expression for the amount of SPM phase is

$$\Delta\phi_{SPM} = 2\pi \frac{n_2^{(eff)} P_0 \eta_c}{A_{\text{eff}} \lambda} \sum_{i=0}^{N_s} \frac{\eta_s^i \eta_b^i L_i}{2^i}. \quad (3.37)$$

Once the SPM phase reaches a value of  $2\pi$ , we notice pulse deformation leading to degradation of the acceleration gradient. This is confirmed by full simulations with our NLSE solver. Thus, the constraint on our input field to avoid SPM effects is given by

$$E_0 < \left( \frac{2\lambda}{n_2^{(\text{eff})} n c_0 \epsilon_0 \eta_c} \sum_{i=0}^{N_s} \frac{2^i}{\eta_s^i \eta_b^i L_i} \right)^{1/2}. \quad (3.38)$$

To model the DLA structures and estimate the acceleration gradient achievable in this geometry, we use a two-dimensional finite-difference frequency-domain method (FDFD) [174] to simulate a waveguide feeding Si dual pillar structures, as in Fig. 3.6. The pillars are assumed to have infinite extent out of the plane, neglecting fringing effects. The phase at each output waveguide is assumed to be at its optimal value for maximum acceleration through the entire section. To compute the acceleration gradient, we must: (1) Use FDFD to compute the acceleration gradient over a discrete range of frequencies. (2) Fit a Lorentzian to the frequency response of the DLA structure, following the discussion of DLA resonances in the previous section. (3) Using the parameters extracted from this fit, scale the response to the Q-factor of interest. (4) Use the input pulse spectrum and fit parameters to compute the acceleration gradient.

We first examine a single stage with a length of 192  $\mu\text{m}$ . In this work we define a ‘stage’ as an accelerator section with a single input laser. This number is chosen as it gives a reasonable balance between acceleration gradient and energy gain. Over a range of pulse durations ( $\tau$ ) and quality factors ( $Q$ ), we first compute the minimum peak electric field at input that will cause either damage or nonlinear pulse distortion using Eqs. 1-4. Then, for relativistic electrons, we use the assumed parameters to compute the achievable acceleration gradient and energy gain. In Fig. 3.7, we show the limiting constraints for each  $\tau$  and  $Q$ , as well as the energy gain from a single stage. This is presented separately for waveguide core materials of Si and  $\text{Si}_3\text{N}_4$ .

From Fig. 3.7, we see that, for a given geometry, there is an optimal combination of  $\tau$  and  $Q$  where the energy gains and acceleration gradients are maximized. For a structure with a stage length 192  $\mu\text{m}$ , this point is at  $\tau = 341 / 322$  fs and  $Q = 157 / 154$  for waveguide cores made of Si /  $\text{Si}_3\text{N}_4$ . A full list of the results are displayed in Table 3.2. Using a SiN waveguide system, we may expect to achieve 1 MeV of energy gain at 108 MV/m gradients by running 49 stages in series. However, these are conservative values based upon a few well-established waveguide approaches and materials, and therefore represent a lower bound on the achievable gradient.

There are several competing effects that lead to the existence of this optimal point. First, for a given pulse peak power, shorter pulse durations will generally lead to higher acceleration gradients because the materials will exhibit higher electric field damage thresholds. However, this effect is limited by the occurrence of SPM at a certain input field. Furthermore, if the pulse is too short with respect to the Q-factor of the DLA structures, the pulse will not couple efficiently to the accelerator gap due to the pulse bandwidth being larger than the structural bandwidth. Secondly, higher

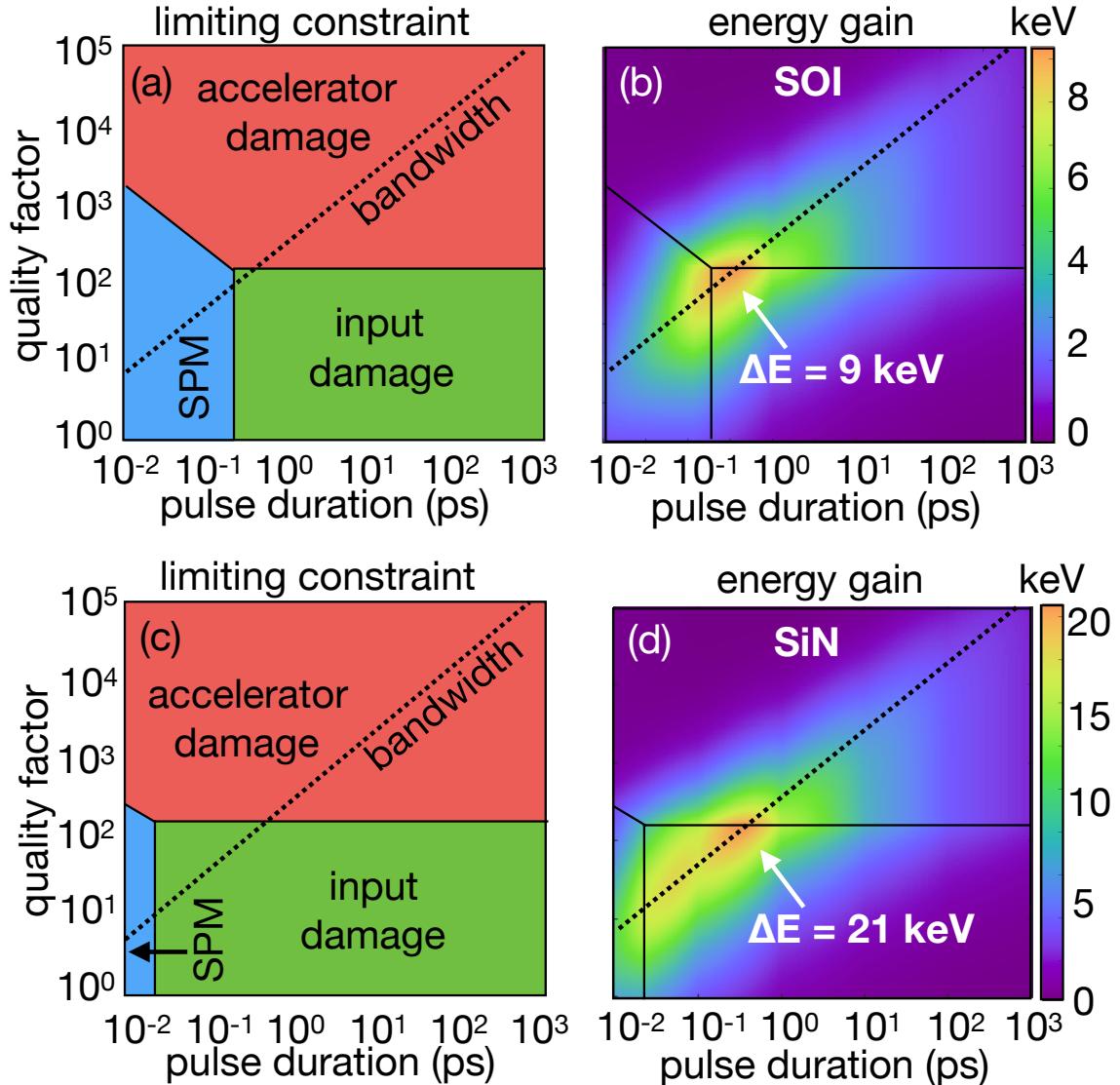


Figure 3.7: Results from the parameter study. A single stage of the tree-network structure is simulated, with stage length of  $192 \mu\text{m}$ , corresponding to 5 power splits and  $2^5 = 32$  output ports. In (a-b), Silicon-on-Insulator (SOI) waveguides are assumed. In (c-d),  $\text{Si}_3\text{N}_4/\text{SiO}_2$  waveguides are assumed. For each Q-factor and pulse duration, we compute the maximum input field achievable before damage or nonlinearity occurs. The different colored regimes in (a) and (c) correspond to different limiting constraints as labeled in the plots. The dotted line corresponds to the minimum pulse duration before the pulse bandwidth exceeds the DLA resonator bandwidth. The energy gain from one section is plotted in (b) and (d).

Q-factors lead to resonantly enhanced fields inside of the DLA structure and higher acceleration gradients as a result [45]. However, if the Q-factor is too high, these enhanced fields will cause the

Table 3.2: Optimal results from the parameter study, for waveguides fabricated from SOI and SiN.

Metric	Value (SOI)	Value (SiN)	Units
Acceleration gradient	45.3	107.5	MV/m
Energy gain per stage	8.7	20.6	keV
Input peak electric field	1.0	2.4	GV/m
Pulse duration	341	322	fs
DLA Q-factor	156.7	154.0	-
Pulse energy at input coupler	0.36	11.3	nJ
Number of stages for 1 MeV	116	49	-
Stage length	192	192	$\mu\text{m}$
Waveguide core width	0.78	2	$\mu\text{m}$
Waveguide core height	220	400	nm

accelerator structures to damage.

To investigate how these results depend on the stage length, we run several of these simulations over a range of structures with different numbers of splits, keeping track of the optimal  $\tau$ ,  $Q$ , acceleration gradient, and energy gain of each structure. The results are presented in Fig. 3.8 where the red dotted line corresponds to the stage length used in this study (192  $\mu\text{m}$ ).

From Fig. 3.8a, we note that as the stage lengths become longer, the achievable acceleration gradients decrease due to the increased losses introduced by the greater number of splits, combined with the increased nonlinearities and concentration of optical power at the input facet. On the other hand, the energy gain increases with greater stage length. Thus, there is an intrinsic trade-off between having a high acceleration gradient and a large energy gain per laser input, suggesting that the choice of stage length should be determined by the acceleration gradients and energy gains required by the application. For instances where high acceleration gradient is preferred, a smaller stage length per laser is optimal, meaning less splits. However, for applications where high total energy gain is a more important figure of merit, it may be beneficial to use a coupling structure with many splits and long stage length, but lower acceleration gradient. These metrics will also depend on the availability of several phase-locked laser sources and the experimental difficulties associated with coupling them to several input couplers. Because of the challenges introduced by concentrating the optical power at a single input facet, there would be significant improvement on these results by considering input schemes that may couple a single beam directly into several waveguides. While this is outside of the scope of this study, it is a promising avenue to explore for these systems.

From inspecting Fig. 3.8b, we see that the optimal  $\tau$  and  $Q$  increases as the structure becomes larger. Thus, the longer the stage length we wish to supply with this tree-network geometry, the more resonance we require in the DLA structures. For a longer stage length, more splits must be performed, which puts additional burden on the input facet relative to the DLA structure. This, in turn, requires greater resonant enhancement at the accelerator gap to offset, and a subsequently larger  $\tau$  to match the structural bandwidth.

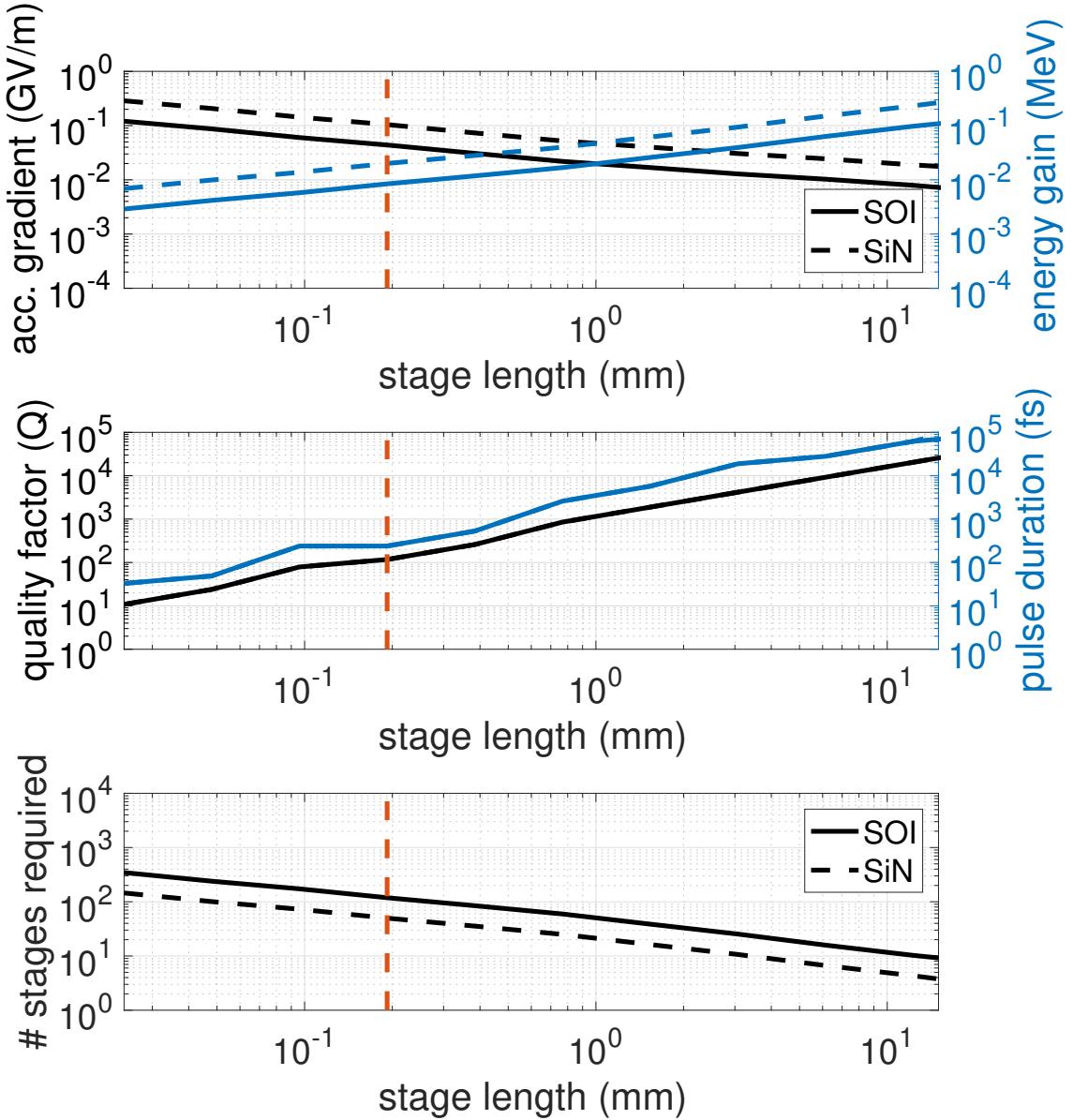


Figure 3.8: Scaling of optimal parameters as a function of the stage length. (a) The optimal energy gains and acceleration gradients as a function of stage length for both SOI and SiN structures. (b) The optimal set of pulse duration and Q-factor corresponding to the highest energy gain and acceleration gradient at each stage length. The curves for SOI and SiN are overlaid. (c) The number of stages required to reach 1 MeV of total energy gain as a function of individual stage length.

We now discuss the outlook of these results and present some methods for improving on the findings. First, we notice that SiN waveguide systems may supply much higher acceleration gradients than SOI systems. This is due to the favorable damage and nonlinear properties of  $\text{Si}_3\text{N}_4$  compared to Si. However, SiN waveguides have high bending loss at bend radii below 50  $\mu\text{m}$  due to the low refractive index of  $\text{Si}_3\text{N}_4$  compared to Si. Therefore, to mitigate the effects of damage and nonlinearities in our waveguide system while maintaining the bending radii required for pulse delay, one solution is to implement a hybrid system comprising of a laser power delivery system optimized for high power handling to feed a series of smaller tree-network structures optimized for tight bends. A diagram of this setup is given in Fig. 3.9.

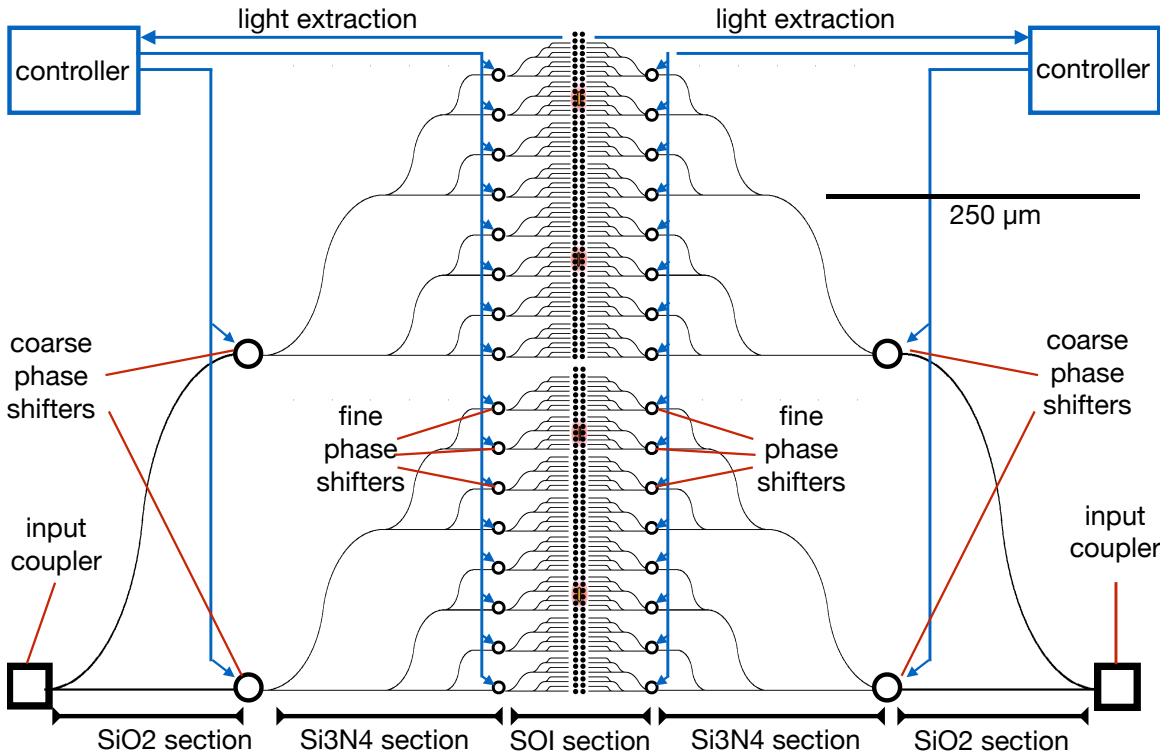


Figure 3.9: Schematic of a hybrid structure for DLA laser coupling. Center: an SOI tree-network / DLA geometry optimized for tight bends and compact waveguides. This is fed by a  $\text{Si}_3\text{N}_4/\text{SiO}_2$  waveguide section with relatively higher damage threshold, and lower nonlinearities. This section is then fed by an all  $\text{SiO}_2$  power delivery section as described in the discussion section. Coarse and fine phase shifters are used in different splitting sections.

Waveguiding systems for this high power handling region may be based on hollow-core photonic crystals, high damage threshold materials, such as silica or silicon nitride, or weakly-guided waveguide modes. The section closer to the DLA could then be implemented in SOI allowing for tight bending radii, compact waveguide networks, and fine phase control. The DLA structures may also

be integrated directly on the same chip as the inner power delivery system. Multiple of these hybrid systems may be driven in series, each with an individual driving laser. The relative merits of large stage length power delivery systems vs. multiple driving lasers will depend on their respective engineering challenges, such as chip-to-chip coupling [175, 166], alignment and stability of input coupling multiple lasers, and availability of these sources.

Furthermore, based on the presented geometry, there is a clear need for resonant DLA structures to enhance the fields at the accelerator gap. For the parameters discussed, the optimal Q-factors were shown to be around 150. Previous work on optimizing DLA structures for high acceleration gradient has shown that periodic dielectric mirrors may be useful in raising quality factors and field enhancement in DLA structures [77, 131, 138, 207]. However, achieving DLA structures with these Q-factors may be difficult with current fabrication tolerances. Furthermore, even slight deformation due to both electron collision with the DLA structure and the presence of high power optical pulses would degrade the Q-factors of fabricated structures. Therefore, experimental verification is required to determine whether such resonant structures can survive operation in a DLA.

One final set of attractive options for further improving the acceleration gradients and energy gains achievable with an on-chip waveguide power delivery system involve engineering the group velocity dispersion (GVD) of the waveguides. One strategy involves pre-chirping the input pulse to compensate for the GVD. Then, the optical power may be initially spread in the temporal domain, mitigating damage bottlenecks near the input facet. Later, with the presence of GVD, the structure may be designed such that the pulse re-compresses at the accelerator structure. Additionally, we may use GVD to balance out SPM effects in our waveguides. With the proper amount of GVD, a temporal soliton may be formed for a given power, which will propagate without distortion, potentially allowing for higher operating powers and acceleration gradients. A similar technique was recently demonstrated to compensate for the SPM effects in short DLA structures [32].

These are promising avenues for exploration, but were not considered in this work with the intention of establishing a conservative baseline for the merits of on-chip laser coupling. Since publishing, experimental verification was performed for many of the parameters assumed in this work [197]. In particular, the waveguide damage thresholds, and nonlinear effects were shown to be consistent (and slightly more optimistic) than those used here. Additional exploration of other material systems, such as  $Ta_2O_5$  [18] and  $Ga_2O_3$ , may offer waveguides and components with loss, nonlinearity, and damage threshold characteristics superior to the material systems assumed in this work. With these issues investigated, a proof of principle optical test will be performed on a simple system before acceleration experiments with electron beams are performed.

We have presented a method for accomplishing chip-based, optical laser power delivery for DLA applications along with a systematic study investigating the damage and nonlinearity constraints and the trade-off between pulse characteristics and DLA resonance. For a stage length of  $192\ \mu m$ , our method predicts acceleration gradients greater than  $100\ MV/m$ , and  $1\ MeV$  of energy gain in

less than 1 cm with 49 structures integrated in series.

We conclude that an on-chip laser coupling system is a promising avenue of exploration for DLA technology. Using the known parameters of existing waveguide technology, we may couple laser sources to an accelerator on a chip with a reasonable acceleration gradient. Additionally, our proposal has a major advantage over free-space laser coupling techniques in that it provides an on chip solution for scalable stage length, which enables access to longer interaction lengths, better integration with DLA structures and greater total energy gains. These findings are a crucial and necessary step towards bringing DLA from proof-of-principle to application stage. However, they may be improved by considering novel components, which we will explore in the next section.

## 3.2 Reconfigurable Control System for Accelerator

The *tree-network* design proposed in the previous section has two serious limitations. First, the design relies on a single input facet to couple the laser beam to a single waveguide before it is split. The optical power concentration in the input facet thus creates a bottleneck leading to optical damage and nonlinearities, which limits the maximum input power and, in turn, energy gain of the accelerator. Secondly, while the phase of each output pulse is tuned using integrated optical phase shifters, the power distribution of the output waveguides is fixed by the fabrication of the splits and bends in the structure and may not be easily tuned later if there are errors.

Therefore, an ideal laser coupling scheme for DLA would involve a coupling scheme that directly couples a single laser pulse into several waveguides. This would greatly ease the input power limitations present in the power-splitting proposal of Ref. [84] by eliminating the optical damage and nonlinearity bottleneck of the input coupler. While such coupling techniques exist in the form of grating coupler arrays [201] or combined grating coupler and splitter devices [185], a tunable *power* distribution system would also be necessary to mitigate the variations in input coupling efficiency, fabrication errors, and drift experienced during the experiment. In conjunction with optical phase control and long-range beam-focusing [140], a reconfigurable power distribution mechanism would be an essential component for scaling DLA from proof-of principle length scales of 100's of  $\mu\text{m}$  to the scale of modern accelerators.

To accomplish this, we propose and present a numerical analysis of an integrated power distribution system and corresponding protocol for accomplishing automated, on-chip phase and power control for DLA and other applications [79]. For this, we utilize a mesh of integrated Mach-Zehnder interferometers (MZIs), a device that allows for reconfigurable unitary operations on-chip [125, 126]. MZI meshes are becoming a fundamental component in integrated, reconfigurable optics for mode sorting [121, 8, 129, 130], quantum information processing [67, 120, 10, 141], and optical machine learning [171, 80]. Here, we explore the novel application of these reconfigurable optics systems to high power pulse delivery and control for accelerators on a chip, which brings its own interesting set

of constraints and challenges. In conjunction with the previous section, this section gives a roadmap for transitioning the control of DLA systems away from hand-tuned, free-space optical setups to precise, automatically configured integrated optical components. As such, it points to the opportunity for the use of on-chip reconfigurable optics to dramatically scale up the acceleration length and energy gains of DLA, which is of crucial importance in realizing the exciting applications of these miniature accelerators.

### 3.2.1 DLA System with Power Equalization

Here we will give an overview of the proposed reconfigurable laser coupling scheme for DLA. A schematic of our system is shown in Fig. 3.10. The system consists of sequential stages for input coupling, power distribution, phase control, group delay control, and electron acceleration.

In our design, the driving laser pulse is first focused onto an input element that directly splits the optical power into several waveguides. Compared with schemes where the laser is first coupled to a single waveguide, as in the previous section, this coupling strategy can greatly improve the power that can be safely supplied by the driving laser. This element could take the form of a grating coupler array or combined grating coupler and power splitter geometry [184]. Adjoint-based optimization techniques [167] may be employed to design novel input coupling components with improved coupling efficiency, less variation between powers, or significantly more output waveguides.

After coupling, due to fabrication and alignment errors, there will be variation in the power distribution of each waveguide. To ensure that an equal amount of power is supplied to the accelerator, we introduce a power distribution component that is comprised of a mesh of MZIs. As diagrammed in Fig. 3.11 **a-c**, each individual MZI is comprised of two beam-splitters and two optical phase shifters,  $\theta$  and  $\phi$ , which can be electrically adjusted to perform the following unitary operation on its two inputs [161, 40, 171, 145]

$$\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = U(\theta, \phi) \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = e^{i\frac{\theta}{2}} e^{i\frac{\phi}{2}} \begin{bmatrix} e^{i\frac{\phi}{2}} \cos \frac{\theta}{2} & e^{i\frac{\phi}{2}} \sin \frac{\theta}{2} \\ -e^{i\frac{\phi}{2}} \sin \frac{\theta}{2} & e^{-i\frac{\phi}{2}} \cos \frac{\theta}{2} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}. \quad (3.39)$$

As shown in Figs. 3.11 **d,e**, several MZIs may be combined in a mesh, which is capable of performing unitary operations over an arbitrary large number of inputs. There are several possible configurations of the MZI mesh, each with their own benefits and drawbacks. In this application, the mesh must be compact enough to fit on the chip and also have a bandwidth large enough to handle sub-picosecond pulses. Whereas many meshes are capable of performing arbitrary unitary operations, for this power delivery problem it is only necessary to sort a single random input into a uniform output. With these considerations in mind, the ‘Clements’ mesh geometry [40], in which

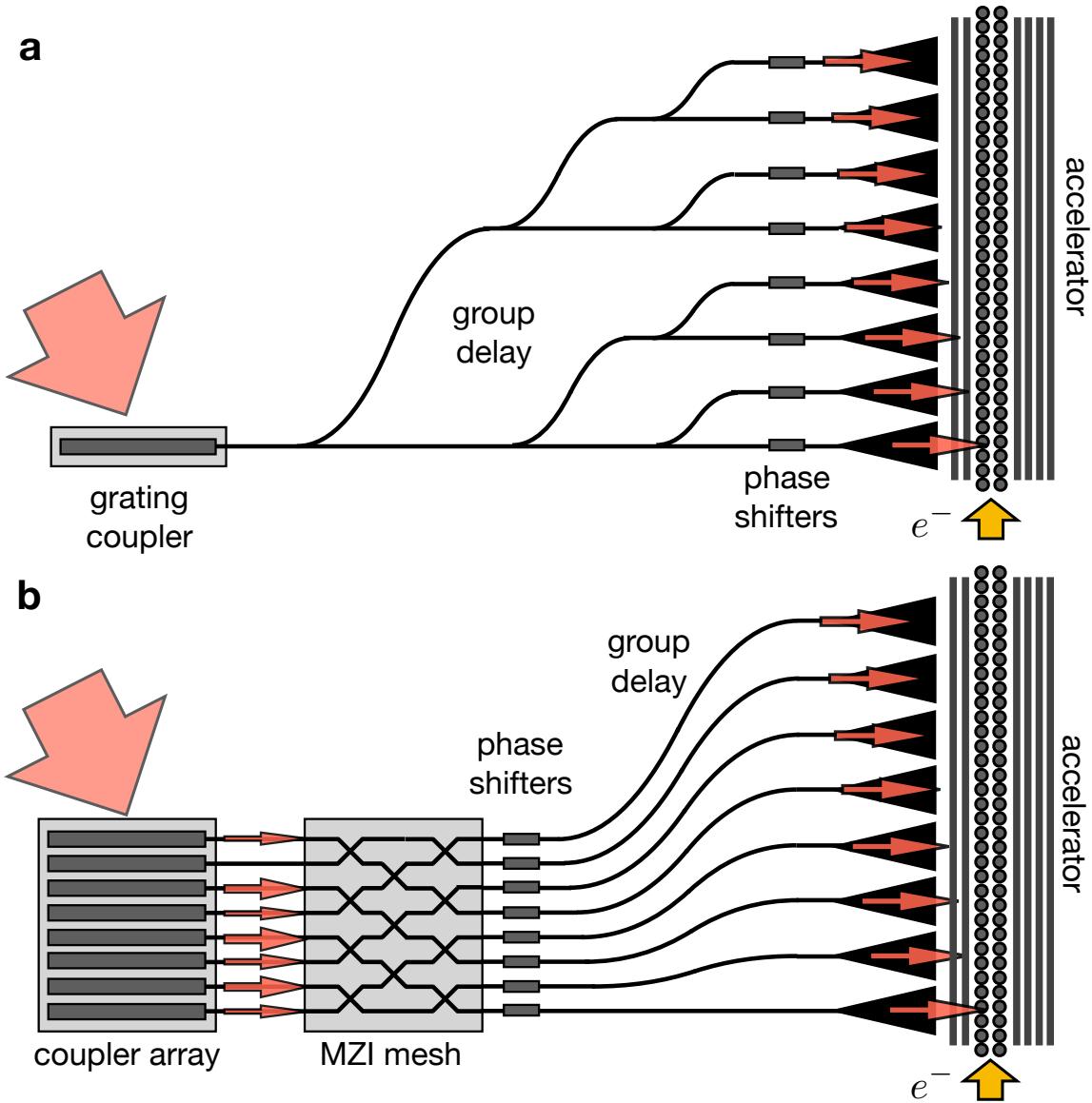


Figure 3.10: **Schematic of the proposed DLA laser delivery control system.** Comparison of setup from Ref [84] (a) and this work (b). **a**, In the splitting structure of Ref. [84], an input pulse (large red arrow) is focused to a single grating coupler on the chip surface. The pulse is split a series of times and bends in the waveguide to create group velocity delay. The phase of the pulses are corrected before injection into the accelerator. **b**, In the schematic proposed in this work, the pulse is directly coupled into several waveguides through a grating coupler array. To mitigate the large variance in coupled powers, the pulses next enter a mesh of MZIs, which is sequentially adjusted, using the protocol from this paper, to provide uniform powers in each waveguide. As before, phase control and group delay are performed using integrated phase shifters and lithographically-defined bends before coupling to the accelerator channel. The second scheme eliminates the damage and nonlinearity bottleneck present in **a** directly after the input coupler.

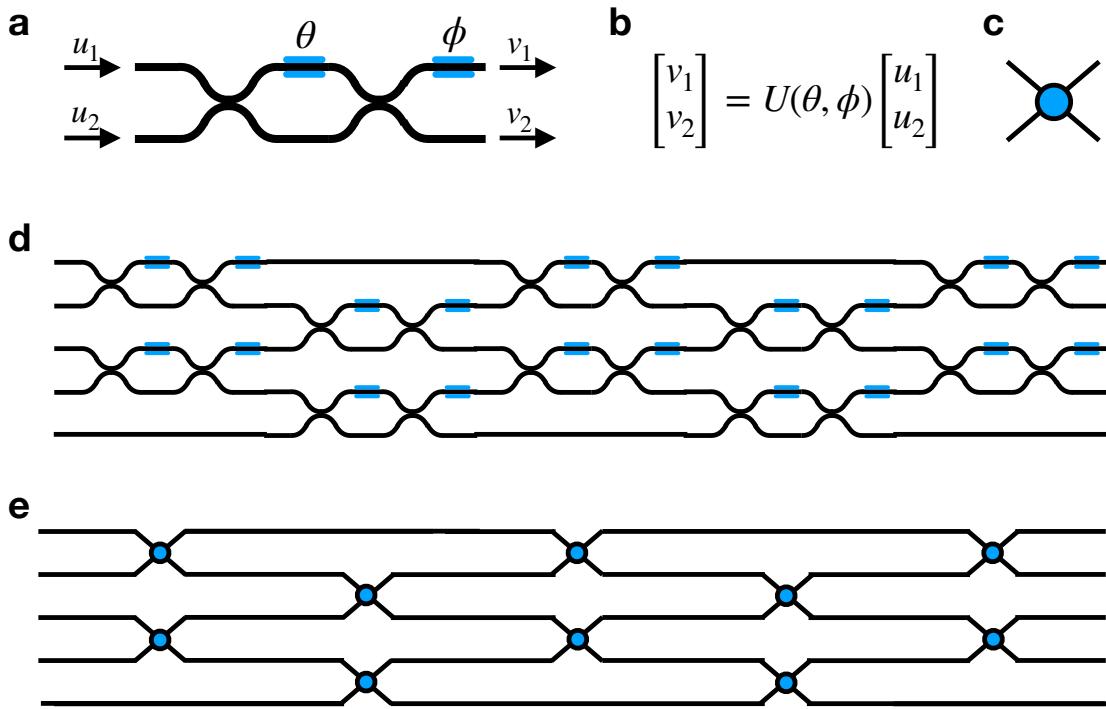


Figure 3.11: **Diagram of MZI mesh for power distribution.** **a**, Diagram of a single MZI consisting of two input ports and two output ports. Where the two arms come together, a 50-50 beam-splitter operation is performed. The blue regions indicate tunable optical phase shifters with added phases marked as  $\theta$  and  $\phi$ . **b**, The MZI represents a tunable unitary transformation on its inputs  $[u_1, u_2]^T$  to give outputs  $[v_1, v_2]^T$ . **c**, Simplified diagram representing an MZI in the following figures. One may think of this element as a tunable power switch. **d**, Individual MZIs are combined into meshes, which may implement tunable unitary operations on several inputs. Shown is a 'Clements' Mesh geometry, which was used in this work. **e**, Schematic of the Clements mesh in **d** using the simple MZI diagram in **c**.

MZIs are configured in a rectangular mesh, is best suited for this application. As diagrammed in Fig. 3.11 **d,e**, the Clements mesh requires fewer layers than other designs [161] and also is more robust to optical losses because of its symmetric layout. Furthermore, it may be implemented in a ‘shallow’ mesh with fewer layers than input ports. This is especially useful for DLA power distribution problem as we will show that only a few layers are required for adequate power equalization.

The MZI meshes used in previous demonstrations [8, 171] were fabricated using Silicon waveguide platforms. In these works, electrically controlled thermal heating was used to implement the tunable phase shifters. In the system presented in this work, the presence of ultra-fast, high power pulses will introduce constraints on the fabrication. For example,  $\text{Si}_3\text{N}_4$  or  $\text{SiO}_2$  materials would be preferred to Si for their high damage and nonlinearity thresholds. Also, thermal phase shifters might suffer from drift or shot-to-shot variations when high power pulses are used. While phase shifters using the electro-optic effect are another option, there may be unwanted interaction between the pulses and the phase shifter. While these issues will be explored in a future study, one promising option may be mechanical phase shifters [65] based on MEMS actuation, which are likely to be less sensitive to pulse propagation, more compact, and have a large enough bandwidth for DLA applications.

We will soon describe a sequential protocol for optimizing the Clements mesh for power distribution. As our protocol uses local information about the power within the network, we require the inclusion of integrated optical photodetectors within the MZI mesh, such as those used in [8]. Alternatively, an imaging system, in conjunction with scattering elements, may be used to gather information about the power distribution from above the chip.

Once the power in each waveguide is equalized, we may use integrated optical phase shifters to correct the phase of each laser pulse such that it is synchronous with the electron beam. While the phase shifters may be adjusted to maximize energy gain, they may also be used to incorporate other functionality, such as beam focusing, total beam transmission, deflection, or diagnostics [183, 220]. Beam energy measurements, performed periodically along the accelerator, may be used as a signal to automatically configure the optical phase shifters.

Once the power and phase are sufficiently controlled, we must delay the laser pulse in each waveguide so that it arrives at the accelerator gap the same time as the moving electron beam. To do this, we introduce lithographically-defined bends in the waveguides to provide a delay that is matched to the electron arrival, producing the integrated optics analogue of a pulse-front tilt [34]. The mathematical details of the bend design are described in the previous section.

The final stage involves coupling the waveguide mode into the acceleration channel. Ideally, an inverse taper may be placed at the end of each waveguide to spread the mode area to match the spot size of the laser pulse to the dimensions of the DLA. Then, an accelerator structure may be placed adjacent to the end of the waveguides. Alternatively, the accelerator structures and tapers may be part of the same system and may be designed following ideas from proposed buried grating structures [36], or using inverse design techniques, such as what was shown Ref. [77] for free-space

coupling. Dielectric mirrors may be used to design the resonance of the acceleration cavity and provide back reflection if a single-sided drive is used [222] (as pictured in Fig. 3.10). Ref. [84] argued that a moderately resonant acceleration cavity with quality factor of several hundred would be necessary to lower the input power to mitigate the input facet bottleneck introduced by that design. However, in this proposal, because that bottleneck is eliminated, one might not require such a resonant structure.

### 3.2.2 Power Distribution Protocol

With the control system defined, here we describe a power distribution protocol that may be implemented on a Clements mesh to optimally equalize the power coming from a random input source. The goal of this power equalization stage is to find the settings of each of the integrated phase shifters such that, given an input to the mesh, there is an equal power in each output port. A naive implementation of this may involve performing a global optimization over each of the phase shifters. However, when the number of degrees of freedom increases (for example, for a very large accelerator), this becomes unfeasible as the dimension of the search space scales linearly with the number of MZIs. Fortunately, the protocol presented in this work allows each MZI to be tuned individually and in sequence, layer by layer, from input to output.

Our protocol is diagrammed in Fig. 3.12, in which we show how to tune a single MZI (red) in a given layer of the network based purely on information about the powers coming into that layer. For generality, we assume that we wish to tune this mesh to achieve an arbitrary target output power distribution  $T_{\text{out}}^{(i)}$  for each of the  $N$  ports  $i \in \{1 .. N\}$ . Here in Fig. 3 the top port is at index 1 and the bottom port is at index  $N$ . We also assume that we have knowledge of the power at each of the ports coming into this layer, labelled  $P_{\text{in}}^{(i)}$ .

The essential idea of the protocol is to tune each MZI to locally direct power to either its top output port or its bottom output port depending on where power is deficient in the input to this layer and where it is needed in the final output layer. To visually represent this idea, in Fig. 3.12, we show a horizontal line bisecting the mesh through the MZI in question. Assuming that the MZI is located vertically within the mesh with its top input port at index  $j$ , we may define the sum of power input to this specific MZI as

$$P_{\text{MZI}} \equiv P_{\text{in}}^{(j)} + P_{\text{in}}^{(j+1)}. \quad (3.40)$$

The sum of the powers input to this layer both above and below this MZI, respectively, are defined

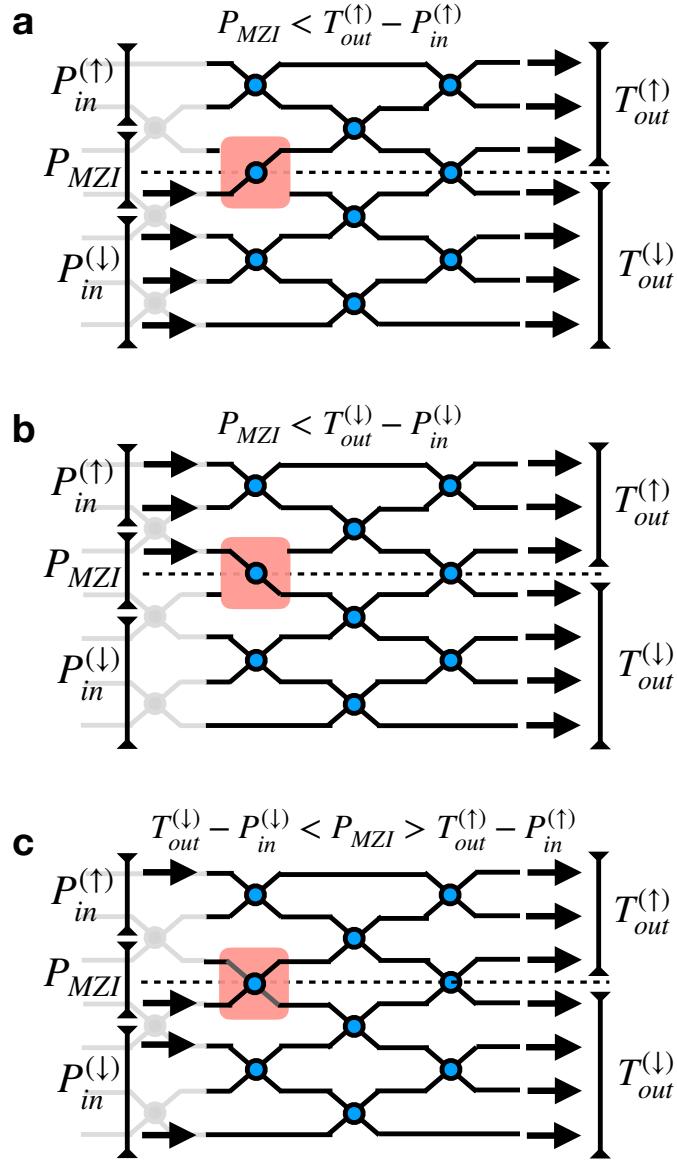


Figure 3.12: **Sequential algorithm for tuning single MZI.** The red square outlines the MZI being tuned. Black arrows on the left indicate input power to the current layer, black arrows on the right indicate desired output powers (uniform in this case). The dotted line separates the mesh into powers above and below the MZI. **a** When there is more power needed in the ports above the MZI than are supplied in the ports leading into the MZI and above, the MZI should output all of its power to its top output port. **b** When there is more power needed in the ports below the MZI than are supplied in the ports leading into the MZI and below, the MZI should send all of its power to the bottom output port. **c** In the intermediate case, the MZI should output just enough power to its top and bottom output ports to match the target power requirements.

as

$$P_{\text{in}}^{(\uparrow)} \equiv \sum_{i=1}^{j-1} P_{\text{in}}^{(i)} \quad (3.41)$$

$$P_{\text{in}}^{(\downarrow)} \equiv \sum_{i=j+2}^N P_{\text{in}}^{(i)}. \quad (3.42)$$

Finally, the target powers above and below this MZI, respectively, are defined as

$$T_{\text{out}}^{(\uparrow)} \equiv \sum_{i=1}^j T_{\text{out}}^{(i)} \quad (3.43)$$

$$T_{\text{out}}^{(\downarrow)} \equiv \sum_{i=j+1}^N T_{\text{out}}^{(i)}. \quad (3.44)$$

Now, using these values, we give a prescription for directing the power out of the MZI to optimally match the target. We notice that there are three distinct cases to consider. These are each diagrammed separately in the subplots of Fig. 3.12.

Case 1 (Fig. 3.12a): the sum of power supplied to the MZI and the ports *above* it is less than the sum of power needed at the final output *above* the MZI. This is also written as

$$P_{\text{MZI}} < T_{\text{out}}^{(\uparrow)} - P_{\text{in}}^{(\uparrow)}. \quad (3.45)$$

In this case, we require that the MZI direct all of its power to the *top* output port, as shown in the red box of Fig. 3.12a.

Case 2 (Fig. 3.12b): the sum of power supplied to the MZI and the ports *below* it is less than the sum of power needed at the final output *below* the MZI. This is also written as

$$P_{\text{MZI}} < T_{\text{out}}^{(\downarrow)} - P_{\text{in}}^{(\downarrow)}. \quad (3.46)$$

In this case, we require that the MZI directs all of its power to the *bottom* output port, as shown in the red box of Fig. 3.12 b.

Case 3 (Fig. 3.12c): when neither of the two cases above are satisfied, i.e.:

$$\begin{aligned} P_{\text{MZI}} &\geq T_{\text{out}}^{(\uparrow)} - P_{\text{in}}^{(\uparrow)} \quad \text{and} \\ P_{\text{MZI}} &\geq T_{\text{out}}^{(\downarrow)} - P_{\text{in}}^{(\downarrow)}, \end{aligned} \quad (3.47)$$

we require the MZI only supply  $T_{\text{out}}^{(\uparrow)} - P_{\text{in}}^{(\uparrow)}$  of its power to the top port. The leftover power may be transmitted to the down port, which, by power conservation, will be equal to  $T_{\text{out}}^{(\downarrow)} - P_{\text{in}}^{(\downarrow)}$  since  $\sum_{i=1}^N P_{\text{in}}^{(i)} = \sum_{i=1}^N T_{\text{out}}^{(i)}$ . This is demonstrated in Fig 3.12c where the MZI performs a partial splitting

of power.

With this protocol, one may thus optimize each MZI sequentially through the mesh. To do this optimally, the MZIs must be tuned layer-by-layer from input to output. Within each layer, a set of integrated photodetectors must be used to measure  $P_{\text{in}}^{(i)}$  for all ports  $i$ . Then, the individual MZIs in this layer may be tuned in parallel or in any order desired.

While the algorithm presented above assumes lossless components, in practice, variations in loss between MZIs in each layer will introduce discrepancies between the expected and measured output power. Therefore, to compensate for this, in practice, a calibration of each MZI's loss, and corresponding update to the algorithm, may be necessary to achieve good performance.

### 3.2.3 Numerical Demonstration

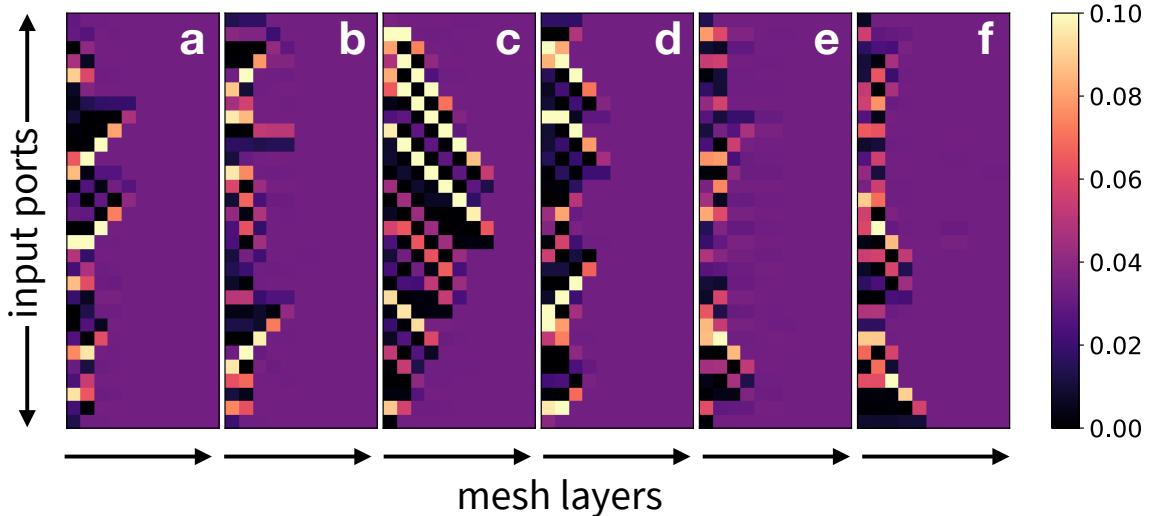


Figure 3.13: **Demonstration of mesh optimization for equal power distribution.** a-f The powers in each port after optimizing the mesh for uniform output power given several random input powers. The colorbar represents the fractional power in each port, where the total power is normalized to 1. The vertical axis represents the input port index and the horizontal axis represents the layer index. Power flows from left to right. Equalization is achieved, on average, after only around 5 layers given this network with 30 ports.

To demonstrate our protocol, we perform numerical simulations of a Clements mesh and optimize it for power equalization from random inputs to uniform outputs. A software package [78] was written to simulate the mesh and perform the optimizations. This package was written such that it may eventually be augmented to interface with a physical MZI mesh to act as a control mechanism. The result of 6 independent runs is shown in Fig 3.13, in which we plot the power in each port within the network after it is optimized using this procedure.

For each run, we initialize a Clements mesh with  $N = 30$  input ports and  $M = 10$  layers. Then, we generate a vector of random input powers to couple to the mesh,  $P_{\text{in}}$ . When constructing  $P_{\text{in}}$ , each element is chosen uniformly at random between 0 and 1 and then the whole vector is normalized such that it sums to 1. The phase of each input mode is set to 0. For demonstration purposes, we optimize the mesh to output to a uniform output target  $T_{\text{out}}$ , where each element of  $T_{\text{out}}$  is equal to  $1/N$ , such that both the input and target powers are normalized to each sum to 1. Although a uniform output target was chosen as it is most applicable to DLA applications, the same protocol may be equally applied to other targets with similar results.

We then step through the mesh from input to output, tuning all MZIs in a given layer according to the protocol introduced in the previous section before moving to the next layer. To perform the tuning, we use a simple downhill simplex algorithm [12] to tune the phase shifters ( $\theta$  and  $\phi$  in Eq. 3.39) in the MZI until they output the correct power as described by the protocol. From Fig. 3.13, we notice that equalization is achieved after only around 5 layers on average.

To understand how the device operation scales with the number of layers, we studied the performance of the equalization routine as the number of ports are increased. The results are shown in Fig. 3.14 where we show the mean-squared-error between the power in each layer, defined as

$$MSE = \frac{1}{N} \sum_{i=1}^N \left( P_{\text{in}}^{(i)} - T_{\text{out}}^{(i)} \right)^2. \quad (3.48)$$

We then sweep through different mesh sizes and average the data for each mesh over 5 runs with different random inputs. The number of layers needed for equalization grows slowly as the network size is increased. This suggests that for very large networks (with  $> 100\text{-}1000$  ports), only tens of layers of MZIs may be needed in the Clements mesh to perform power equalization.

As evidenced by Fig. 3.13, extreme input conditions, such as the full power concentrated at a single port on the edge of the mesh, will need a full quadratic mesh to fully equalize power. However, these edge cases are not only far less likely than the more uniform input conditions, but also represent failure modes caused by severely distorted input laser modes, which are likely to damage the structure. In practice, we therefore do not expect the MZI mesh to compensate for them.

### 3.2.4 Performance Improvement

To quantify the benefit of this new power distribution component, we now compare the performance of a DLA with the power splitting approach of Ref. [84] to one with direct coupling and power control from this work and as diagrammed in Fig. 3.10. For each of these approaches, we give an estimate of the acceleration gradient, number of output waveguides, and acceleration length needed to achieve a given energy gain with a single driving laser.

Following the analysis of Ref. [84], we first consider a splitting design in which a single input

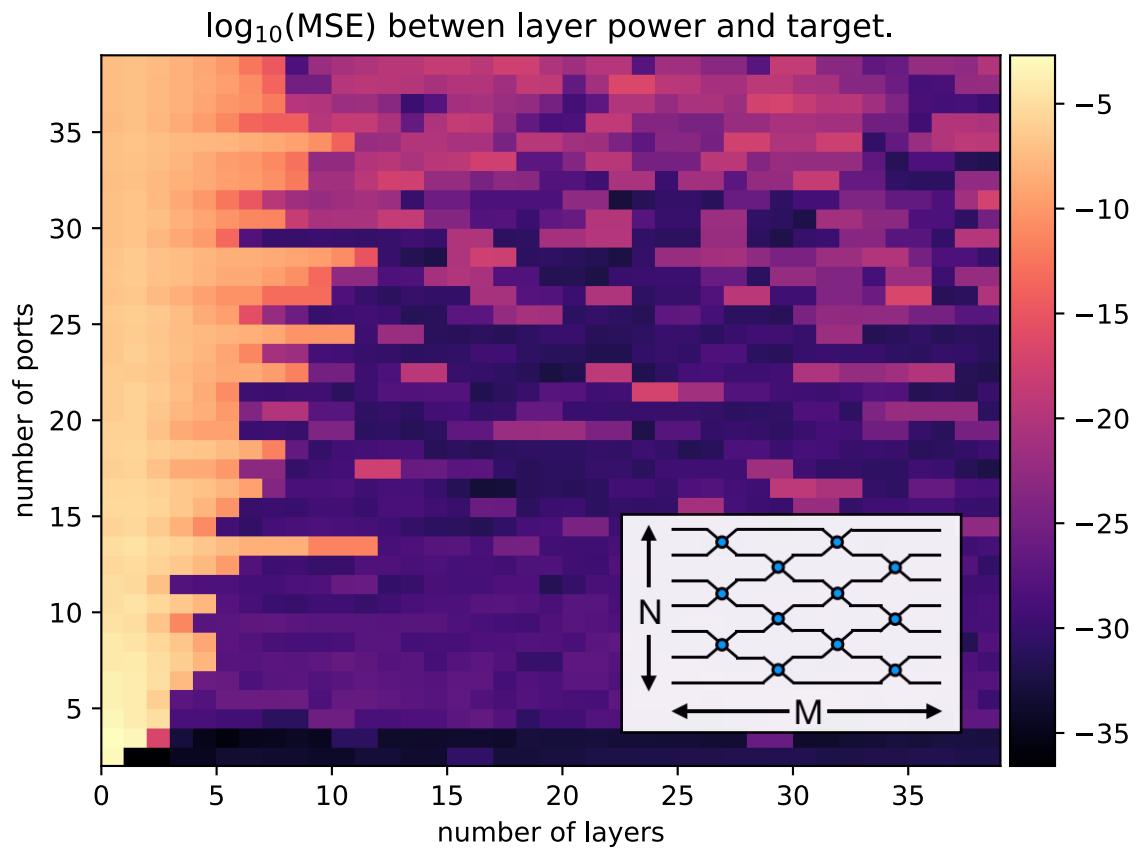


Figure 3.14: **Analysis of mesh performance vs. number of input ports.** We plot the mean-squared-error between the power in each layer and the target power after the mesh has been optimized. These are averaged over 5 random input values. We see that the number of layers ( $M$ ) needed to equalize the power in the mesh increases slowly with respect to the number of ports ( $N$ ).

waveguide is split  $N_s$  times to cover the full length of acceleration. We assume that an optical pulse with energy  $U_d$  and temporal duration  $\tau$  is coupled into the input waveguide. We must ensure that  $U_d$  is below the damage and nonlinearity limit of the waveguides, which was recently measured to be 20 nJ in SiN waveguides using pulses of duration of 250 fs [197]. This corresponds to a fluence of  $0.12 \text{ J/cm}^2$  for a mode area of  $17.2 \mu\text{m}^2$ , as used in this reference. We then assume that this initial waveguide is split  $N_s$  times, in a binary fashion, to couple to  $N = 2^{N_s}$  final waveguides, which are directly coupled to the accelerator. The power coupling efficiency factors corresponding to waveguide splitting and bending loss are denoted by  $\eta_s$  and  $\eta_b$  respectively. The values used for these efficiencies are given in Table 3.1 and are consistent with those of Ref. [84].

Next, we assume each waveguide powers  $N_p$  periods of the accelerator, each of height  $h$  and width  $\beta\lambda_0$  to satisfy the synchronicity condition, where  $\beta$  is the ratio of the electron speed to the speed of light and  $\lambda_0$  is the free space wavelength of the driving laser. This gives a total acceleration length of

$$L = N N_p \beta \lambda_0 \quad (3.49)$$

In terms of these parameters and the peak electric field of the laser pulse incident on the accelerator channel ( $E_0$ ), the corresponding power intercepting the DLA is  $P_0 \simeq \zeta E_0^2 h N_p \lambda \beta n / (2Z_0)$ , where  $Z_0 = 377\Omega$  is the impedance of free space,  $n$  is the refractive index, and  $\zeta$  is a dimensionless factor of order unity (for a Gaussian  $\zeta = \pi/2$ ). Considering losses between the input coupling of a pulse with energy  $U_d$  and duration  $\tau$ , we may compute an expression for  $P_0$  and solve for  $E_0$  to obtain

$$E_0 = \sqrt{\frac{2U_d Z_0 (\eta_s \eta_b)^{N_s}}{2^{N_s} \zeta N_p \tau \lambda_0 \beta h n}}, \quad (3.50)$$

The acceleration gradient, defined as the energy gain ( $\Delta_E$ ) per unit length, may be written in terms of the acceleration length  $L$ , input electric field  $E_0$ , and elementary charge  $e$  as

$$G = \frac{\Delta_E}{eL} = \kappa E_0, \quad (3.51)$$

The proportionality constant or *structure factor*  $\kappa$  is a dimensionless quantity that denotes the coupling of incident field to the structure. Structure factors of  $\kappa = 0.2$  have been experimentally demonstrated [34] and  $\kappa = 1.34$  theoretically predicted [17] for various structure designs, and by use of a resonant structure with a quality factor  $Q$  it can be further enhanced as  $\sqrt{Q}$  [84]. We here assume a moderate value of  $\kappa = 1$ . With the expression above for the gradient, along with Eqs. (3.49) and (3.50), we may solve for the number of waveguide splits required to accomplish the energy gain of  $\Delta_E$  as

$$N_s = \log_2 \left[ \frac{\Delta_E^2 \tau h n \zeta}{2 \kappa^2 U_d N_p \beta \lambda_0 Z_0} \right] / \log_2(2\eta_s \eta_b) \quad (3.52)$$

The total number of waveguides feeding the DLA is then  $N = 2^{N_s}$ . We can see that in the limit

where the efficiency terms are equal to unity,  $N$  reduces to the quantity contained inside the first logarithm in Eq. (3.52).

In contrast, using a direct coupling method with power distribution components introduced in this work, there is no need for on-chip splitting of the optical power. Because of this, we may couple  $\eta_{\text{MZI}} \times U_d$  of optical energy separately into each waveguide, where  $\eta_{\text{MZI}}$  is the throughput efficiency of the MZI. In the experimental demonstrations of Ref. [8, 171], less than 50% of power loss was demonstrated for a 4-layer MZI mesh. As a conservative estimate, we assume here  $\eta_{\text{MZI}} = 0.25$  for 10 layers. Additionally, negligible amounts of reflection and internal backscattering were observed in these implementations, and therefore are not considered in this analysis. Following a similar analysis, the electric field incident on the accelerator, given by Eq. (3.50), loses its  $N$  dependence and thus the number of required output waveguides needed scales as

$$N^{(\text{direct})} = \frac{\Delta_E}{\kappa} \sqrt{\frac{\zeta \tau h n}{2U_d \eta_{\text{MZI}} N_p \beta \lambda_0 Z_0}}. \quad (3.53)$$

In Fig. 3.15 we show the scaling of the acceleration gradient, number of output waveguides, and acceleration length as a function of required energy gain for a DLA given realistic parameters, which are supplied in Table 3.1. Based on previous demonstrations of these components [8, 171] the reflection and backscattering is negligible, and therefore not considered in this analysis.

To connect these results with prior work, we note that for the split waveguide approach, Eqs. (3.49), (3.50), (3.51), (3.52) give nearly identical results as those of the more involved analysis of [84] for the case of silicon nitride waveguides for a single-stage energy gain of 20 keV, when the same input parameters are used. In Ref. [197], the damage and nonlinearity constraints using similar pulse parameters was studied experimentally. It was found that a silicon nitride waveguide could sustain a pulse of 20 nJ for several millimeters of propagation distance. Therefore, while this proposed power equalization system includes more components (beam splitters and phase shifters), it is likely to be well within the damage threshold.

However, the results in Fig. 3.15 for the split waveguide approach (solid blue curves) are somewhat more optimistic than those of the prior work, as recent experiments [197, 17] have provided evidence for larger values of  $U_d$  and  $\kappa$  than were previously assumed. These more optimistic parameters are reflected in Table 3.3. However, due to the  $N_s$  scaling of the power loss for the split waveguide approach, higher energy gains per stage rapidly become prohibitive as the resulting gradient drops exponentially and the number of required waveguides increases. In Ref. [84] this difficulty was addressed by cascading multiple laser-coupled stages in series to achieve a reasonable energy gain of 1 MeV. In contrast, as seen in Fig. 3.15 the direct coupling approach can in principle achieve this with one acceleration stage and one input laser, requiring 144 waveguides over a length of 2.8 mm with a gradient of 350 MeV/m. These are promising numbers for a DLA structure and may be improved primarily by optimizing for higher values of  $\kappa$  and  $N_p$ , which may be accomplished using

inverse design techniques [77].

Metric	Description	Value	Units
$\beta$	$e^-$ speed / $c_0$	1	-
$\lambda_0$	free space wavelength	2	$\mu\text{m}$
$N_p$	# DLA periods / waveguide	10	-
$\kappa$	structure coupling factor	1	-
$U_d$	input pulse energy	20	nJ
$\tau$	input pulse duration	250	fs
$h$	DLA structure height	2	$\mu\text{m}$
$n$	waveguide refractive index	2	-
$\eta_s$	waveguide splitting efficiency	0.95	-
$\eta_b$	bend loss efficiency	0.95	-
$\eta_{\text{MZI}}$	MZI transmission efficiency	0.25	-
$\zeta$	geometrical mode factor	$\pi/2$	-

Table 3.3: Set of parameters used in the analysis of Fig. 3.15

Our proposal provides a method for automated control of DLA systems and eliminates the major issues with previous laser coupling schemes. We show that MZI meshes are a promising candidate for a power distribution system for DLA and our findings may be applied to other applications in integrated optics requiring power routing. The proposal given here only uses existing optical components and, therefore, should be feasible for an experimental demonstration.

The procedure we introduce for optimizing the MZI mesh to achieve arbitrary power distribution is highly efficient in terms of number of measurements and phase shifter tunings. For a mesh with  $N$  ports and  $M$  layers, our protocol replaces one large optimization problem with  $2NM$  degrees of freedom to  $MN$  independent optimization problems with 2 degrees of freedom each. This greatly improves the feasibility of implementing this protocol on large meshes with thousands of input ports, for example, which may be needed for large-scale DLA. In comparison with other mode sorting schemes [8, 122] in which a Reck (triangular) mesh architecture is used, our algorithm provides the opportunity to perform similar functionality on a rectangular mesh, which benefits from more uniform loss and bandwidth across pathways through the device.

We show that a shallow Clements mesh is sufficient for equalizing power from random inputs. This is of crucial importance for DLA applications where there are limitations on the chip space available for phase shifters and electrical contacts. It also means that a shorter propagation length within the waveguides can be achieved, which reduces the possible nonlinear effects. Having a shallow mesh also allows the device to retain a large bandwidth, which decreases with the number of MZIs in an optical path. This is crucially important for handling sub-picosecond driving pulses used in DLA. However, the bandwidth of MZI meshes and its scaling with respect to network size is not yet fully understood. This will need to be tested experimentally or with a separate numerical study.

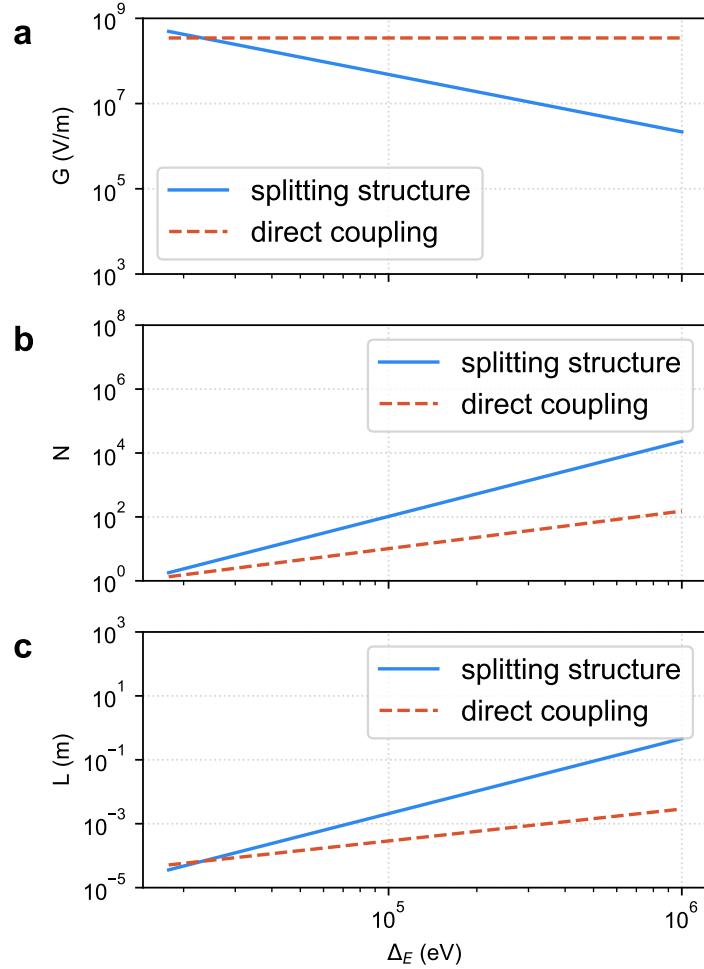


Figure 3.15: **Figure of merit scaling for different laser coupling architectures.** Solid blue lines refer to the results using the system of Ref [84], in which all optical power is initially coupled at a single input facet. Red dotted lines refer to the structure from this work. **a** The acceleration gradient as a function of electron energy gain. For the splitting structure, the acceleration gradient diminishes rapidly as the length of acceleration is increased to match the desired energy gain. However, the structure from this work achieves uniform gradient in principle. **b** The number of output waveguides ( $N$ ) required to achieve an given energy gain of  $\Delta_E$ . **c** The acceleration length ( $L$ ) required to achieve an given energy gain of  $\Delta_E$ .

The optimization protocol presented here decomposes the global optimization problem of tuning

the entire mesh into several subproblems involving tuning the individual MZIs. However, alternatively a gradient-based approach may potentially be used to train the full mesh. It was shown previously [80] that the gradient of the output of an MZI mesh with respect to the dielectric function of each of the phase shifters may be measured experimentally using adjoint fields [204]. Interestingly, for the case of maximizing the acceleration of a DLA, it was independently shown that the corresponding adjoint fields are given exactly by the fields radiated by the electron beam [77]. This suggests an interesting approach to optimizing the MZI mesh towards maximum acceleration by first measuring the radiation from test electron beam, and then using the protocol from Ref. [80] to measure the derivative of the acceleration with respect to each of the phase shifters. With this, one may do parallel, gradient-based updates of the phase shifters and optimize arbitrarily large grids with high efficiency. This idea may be explored in a future study.

Overall, these studies indicates that integrated optical power delivery systems are worth continuing to pursue for DLA. We presented a path towards automatic power distribution, which is an essential component towards scaling DLA to longer length scales and exciting applications. We also provide a novel application of the MZI mesh, which is already finding many applications in other exciting reconfigurable optics applications. Our efficient protocol for optimizing an MZI for arbitrary power distribution may also find many applications beyond DLA.

### 3.3 Outlook and Conclusions

Integrated optics, and reconfigurable optics in general, allows unique opportunities for accelerators on a chip to take advantage of high precision control and automatic compensation for errors from fabrication, alignment, or drift. These studies presented two promising avenues for accomplishing extended acceleration lengths for these accelerators and eventually may enable future applications of DLA technology.

Since publishing, recent works have further validated the feasibility and usefulness of using integrated optics platforms as a power delivery system for DLA. A recent study [226] proposed the use of dielectric waveguides as a co-propagating accelerating scheme. Here, the optical pulses are coupled into a slot waveguide configuration and the evanescent fields are used to provide energy gain to the electron beams. The waveguide widths are tapered along the propagation axis, which allows for phase matching to an accelerating, subrelativistic electron beam. Furthermore, experimental damage and nonlinearity testing of SiN waveguides has shown that the waveguides may sustain pulses similar to those used in these parameter studies [197] without damage or significant nonlinear effects. Finally, recent demonstrations of the first waveguide-coupled DLA were performed using a grating coupler and accelerator structure designed using inverse design [167, 168]. Further experimental studies are underway, with the goal of realizing a multi-stage integrated accelerator within the next two years.

## Chapter 4

# Training of Optical Neural Networks

In this chapter, we will shift gears and discuss optical hardware platforms for machine learning applications. The field of machine learning has seen an impressive revival in the past decade, with several applications from machine translation [27], computer vision [57], health and medicine [55], and game playing [179]. While the basic ideas and models of machine learning, and *deep learning* [61, 108] have existed for several decades, this recent success may be largely attributed to (1) the widespread availability of hardware platforms capable of running these models efficiently, and (2) the accessibility of large datasets needed to train these models.

Indeed, with the new applications and models introduced in recent years, the hardware demand continues to rise exponentially over time. One common approach to satisfying this demand involves running the computation on graphical processing units (GPUs), which creates benefits from massive parallelization. Along this line, companies are beginning to invest in hardware platforms that are specially designed to run machine learning software platforms. One such example is the *tensor processing unit* (TPU), which runs Google's Tensorflow package.

Alternatively, technologies are being considered that would replace the typical digital electronics hardware architectures with those more suited for machine learning computation. This is evidenced by the renewed interest in analog computing and quantum machine learning [19], in which a physical system is used to emulate the machine learning model. Other groups have attempted to create *neuromorphic* hardware, which is designed to vaguely mimick the circuitry of the human brain [195].

Of great interest among these new platforms are so-called *linear nanophotonic processors* [66]. The idea here is to replace the linear elements common to most machine learning models with linear

optical devices that may perform linear operations via transmission of light through their components. While this idea has existed for quite some time [121, 127, 123, 126, 122], its implementation for deep learning was recently demonstrated in a photonic integrated circuit [172] and has been used in recent years for quantum information processing [141]. The main advantage of this technique is that it can, in principle, perform these crucial operations faster, with lower latency, and with a low energy cost when compared to digital electronics. A further benefit is that the operations, themselves, may be reconfigured using standard optical components, such as phase shifters. This ability will be necessary for the training of these platforms, as we will discuss.

## 4.1 Optical Neural Networks

There are significant efforts in constructing artificial neural network architectures using various electronic solid-state platforms [119, 156], but ever since the conception of ANNs, a hardware implementation using optical signals has also been considered [3, 93]. In this domain, some of the recent work has been devoted to photonic spike processing [163, 196] and photonic reservoir computing [28, 202], as well as to devising universal, chip-integrated photonic platforms that can implement any arbitrary ANN [169, 170]. Photonic implementations benefit from the fact that, due to the non-interacting nature of photons, linear operations – like the repeated matrix multiplications found in every neural network algorithm – can be performed in parallel, and at a lower energy cost, when using light as opposed to electrons.

A key requirement for the utility of any ANN platform is the ability to train the network using algorithms such as error backpropagation [164]. Such training typically demands significant computational time and resources and it is generally desirable for error backpropagation to be implemented on the same platform. This is indeed possible for the technologies of Refs. [119, 62, 71] and has also been demonstrated e.g. in memristive devices [7, 156]. In optics, as early as three decades ago, an adaptive platform that could approximately implement the backpropagation algorithm experimentally was proposed [206, 157]. However, this algorithm requires a number of complex optical operations that are difficult to implement, particularly in integrated optics platforms. Thus, the current implementation of a photonic neural network using integrated optics has been trained using a model of the system simulated on a regular computer [170]. This is inefficient for two reasons. First, this strategy depends entirely on the accuracy of the model representation of the physical system. Second, unless one is interested in deploying a large number of identical, fixed copies of the ANN, any advantage in speed or energy associated with using the photonic circuit is lost if the training must be done on a regular computer. Alternatively, training using a brute force, *in situ* computation of the gradient of the objective function has been proposed [170]. However, this strategy involves sequentially perturbing each individual parameter of the circuit, which is highly inefficient for large systems.

To overcome this, we proposed [80] a procedure to compute the gradient of the cost function of a photonic ANN by use of only *in situ* intensity measurements. Our procedure works by *physically* implementing the adjoint method in the device [59, 205, 77]. As discussed, the adjoint method scales in constant time with respect to the number of parameters, which allows for backpropagation to be efficiently implemented in a hybrid opto-electronic network. Although we focus our discussion on a particular hardware implementation of a photonic ANN, our conclusions are derived starting from Maxwell's equations, and may therefore be extended to other photonic platforms.

First, we introduce the operation and gradient computation of a feed-forward photonic ANN. In its most general case, a feed-forward ANN maps an input vector to an output vector via an alternating sequence of linear operations and element-wise nonlinear functions of the vectors, also called ‘activations’. A cost function,  $\mathcal{L}$ , is defined over the outputs of the ANN and the matrix elements involved in the linear operations are tuned to minimize  $\mathcal{L}$  over a number of training examples via gradient-based optimization. The *backpropagation algorithm* is typically used to compute these gradients analytically by sequentially utilizing the chain rule from the output layer backwards to the input layer.

Here, we will outline these steps mathematically for a single training example, with the procedure diagrammed in Fig. 4.1a. We focus our discussion on the photonic hardware platform presented in [170], which performs the linear operations using optical interference units (OIUs). The OIU is a mesh of controllable Mach-Zehnder interferometers (MZIs) integrated in a silicon photonic circuit. It is worth noting that this is the same component proposed in the previous Chapter to control power delivery to the accelerator on a chip. By tuning the phase shifters integrated in the MZIs, any unitary  $N \times N$  operation on the input can be implemented [162, 39], which finds applications both in classical and quantum photonics [31, 68]. In the photonic ANN implementation from Ref. [170], an OIU is used for each linear matrix-vector multiplication, whereas the nonlinear activations are performed using an electronic circuit, which involves measuring the optical state before activation, performing the nonlinear activation function on an electronic circuit such as a digital computer, and preparing the resulting optical state to be injected to the next stage of the ANN.

We first introduce the notation used to describe the OIU, which consists of a number,  $N$ , of single-mode waveguide input ports coupled to the same number of single-mode output ports through a linear and lossless device. In principle, the device may also be extended to operate on a different number of inputs and outputs. We further assume directional propagation such that all power flows exclusively from the input ports to the output ports, which is a typical assumption for the devices of Refs. [123, 170, 68, 31, 162, 124, 39]. In its most general form, the device implements the linear operation

$$\hat{W}\mathbf{X}_{\text{in}} = \mathbf{Z}_{\text{out}}, \quad (4.1)$$

where  $\mathbf{X}_{\text{in}}$  and  $\mathbf{Z}_{\text{out}}$  are the modal amplitudes at the input and output ports, respectively, and  $\hat{W}$ , which we will refer to as the transfer matrix, is the off-diagonal block of the system’s full scattering

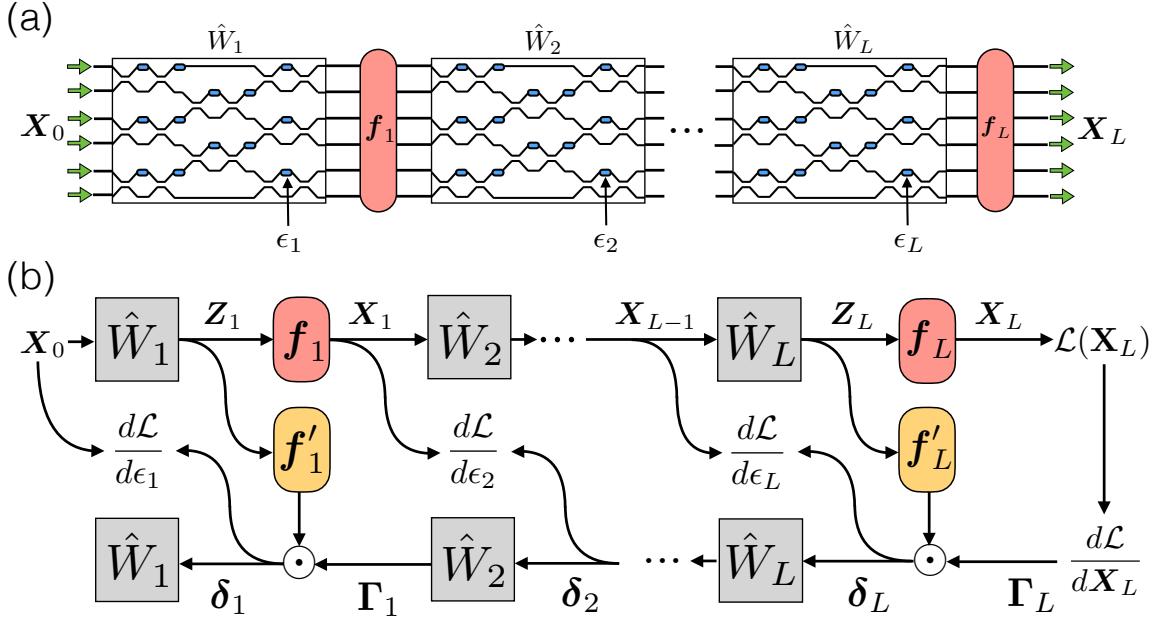


Figure 4.1: (a) A schematic of the ANN architecture demonstrated in Ref [170]. The boxed regions correspond to OIUs that perform a linear operation represented by the matrix  $\hat{W}_l$ . Integrated phase shifters (blue) are used to control the OIU and train the network. The red regions correspond to nonlinear activations  $f_l(\cdot)$ . (b) Illustration of operation and gradient computation in an ANN. The top and bottom rows correspond to the forward and backward propagation steps, respectively. Propagation through a square cell corresponds to matrix multiplication. Propagation through a rounded region corresponds to activation.  $\odot$  is element-wise vector multiplication.

matrix,

$$\begin{pmatrix} \mathbf{X}_{\text{out}} \\ \mathbf{Z}_{\text{out}} \end{pmatrix} = \begin{pmatrix} 0 & \hat{W}^T \\ \hat{W} & 0 \end{pmatrix} \begin{pmatrix} \mathbf{X}_{\text{in}} \\ \mathbf{Z}_{\text{in}} \end{pmatrix}. \quad (4.2)$$

Here, the diagonal blocks are zero because we assume forward-only propagation, while the off-diagonal blocks are the transpose of each other because we assume a reciprocal system.  $\mathbf{Z}_{\text{in}}$  and  $\mathbf{X}_{\text{out}}$  correspond to the input and output modal amplitudes, respectively, if we were to run this device in reverse, i.e. sending a signal in from the output ports.

Now we may use this notation to describe the forward and backward propagation steps in a photonic ANN. In the forward propagation step, we start with an initial input to the system,  $\mathbf{X}_0$ , and perform a linear operation on this input using an OIU represented by the matrix  $\hat{W}_1$ . This is followed by the application of a element-wise nonlinear activation,  $f_1(\cdot)$ , on the outputs, giving the input to the next layer. This process repeats for the each layer  $l$  until the output layer,  $L$ . Written compactly, for  $l = 1 \dots L$

$$\mathbf{X}_l = f_l(\hat{W}_l \mathbf{X}_{l-1}) \equiv f_l(\mathbf{Z}_l). \quad (4.3)$$

Finally, our cost function  $\mathcal{L}$  is an explicit function of the outputs from the last layer,  $\mathcal{L} = \mathcal{L}(\mathbf{X}_L)$ . This process is shown in Fig. 4.1(a).

The nonlinear activation functions,  $f_l()$ , play a key role in ANNs by enabling them to learn complex mappings between their inputs and outputs. Whereas digital processors have the expressiveness to trivially apply nonlinearities such as the widely-used `sigmoid`, `ReLU`, and `tanh` functions, the realization of nonlinearities in optical hardware platforms is more challenging. One reason for this is that optical nonlinearities are relatively weak, necessitating a combination of large interaction lengths and high signal powers, which impose lower bounds on the physical footprint and the energy consumption, respectively. Although it is possible to resonantly enhance optical nonlinearities, this comes with an unavoidable trade-off in reducing the operating bandwidth, thereby limiting the information processing capacity of an ONN. Additionally, maintaining uniform resonant responses across many elements of an optical circuit necessitates additional control circuitry for calibrating each element [159].

A more fundamental limitation of optical nonlinearities is that their responses tend to be fixed during device fabrication. This limited tunability of the nonlinear optical response prevents an ONN from being reprogrammed to realize different forms of nonlinear activation functions, which may be important for tailoring ONNs for different machine learning tasks. Similarly, a fixed nonlinear response may also limit the performance of very deep ONNs with many layers of activation functions since the optical signal power drops below the activation threshold, where nonlinearity is strongest, in later layers due to loss in previous layers. For example, with optical saturable absorption from 2D materials in waveguides, the activation threshold is on the order of 1-10 mW [16, 146, 90], meaning that the strength of the nonlinearity in each subsequent layer will be successively weaker as the transmitted power falls below the threshold.

In light of these challenges, the ONN demonstrated in Ref. 171 implemented its activation functions by detecting each optical signal, feeding them through a conventional digital computer to apply the nonlinearity, and then modulating new optical signals for the subsequent layer. Although this approach benefits from the flexibility of digital signal processing, conventional processors have a limited number of input and output channels, which make it challenging to scale this approach to very large matrix dimensions, which corresponds to a large number of optical inputs. Moreover, digitally applied nonlinearities add latency from the analog-to-digital conversion process and constrain the computational speed of the neural network to the same GHz-scale clock rates which ONNs seek to overcome. Thus, a hardware nonlinear optical activation, which doesn't require repeated bidirectional optical-electronic signal conversion, is of fundamental interest for making integrated ONNs a viable machine learning platform.

For this purpose, as diagrammed in Fig. 4.2, we proposed an electro-optic architecture for synthesizing optical-to-optical nonlinearities which alleviates the issues discussed above. Our architecture features complete *on-off* contrast in signal transmission, a variety of nonlinear response curves,

and a low activation threshold. Rather than using traditional optical nonlinearities, our scheme operates by measuring a small portion of the incoming optical signal power and using electro-optic modulators to modulate the original optical signal, without any reduction in operating bandwidth or computational speed. Additionally, our scheme allows for the possibility of performing additional nonlinear transformations on the signal using analog electrical components. Related electro-optical architectures for generating optical nonlinearities have been previously considered [111, 116, 194]. While we focused on the application of our architecture as an element-wise activation in a feed-forward ONN, but the synthesis of low-threshold optical nonlinearities could be of broader interest to optical computing and information processing. For more information on this work, we refer the reader to Ref. 212.

## 4.2 Training of Optical Hardware

To train the network, we must minimize the cost function with respect to the linear operators,  $\hat{W}_l$ , which may be adjusted by tuning the integrated phase shifters within the OIUs. While a number of recent papers have clarified how an individual OIU can be tuned by sequential, *in situ* methods to perform an arbitrary, pre-defined operation [124, 123, 127, 9], these strategies do not straightforwardly apply to the training of ANNs, where nonlinearities and several layers of computation are present. In particular, the training of ANN requires gradient information which is not provided directly in the methods of Ref. [124, 123, 127, 9].

In Ref. [170], the training of the ANN was done *ex situ* on a computer model of the system, which was used to find the optimal weight matrices  $\hat{W}_l$  for a given cost function. Then, the final weights were recreated in the physical device, using an idealized model that relates the matrix elements to the phase shifters. Ref. [170] also discusses a possible *in situ* method for computing the gradient of the ANN cost function through a serial perturbation of every individual phase shifter ('brute force' gradient computation). However, this gradient computation has an unavoidable linear scaling with the number of parameters of the system. The training method that we propose here operates without resorting to an external model of the system, while allowing for the tuning of each parameter to be done in parallel, therefore scaling significantly better with respect to the number of parameters when compared to the brute force gradient computation.

To introduce our training method we first use the backpropagation algorithm to derive an expression for the gradient of the cost function with respect to the permittivities of the phase shifters in the OIUs. In the following, we denote  $\epsilon_l$  as the permittivity of a single, arbitrarily chosen phase shifter in layer  $l$ , as the same derivation holds for each of the phase shifters present in that layer. Note that  $\hat{W}_l$  has an explicit dependence on  $\epsilon_l$ , but all field components in the subsequent layers also depend implicitly on  $\epsilon_l$ .

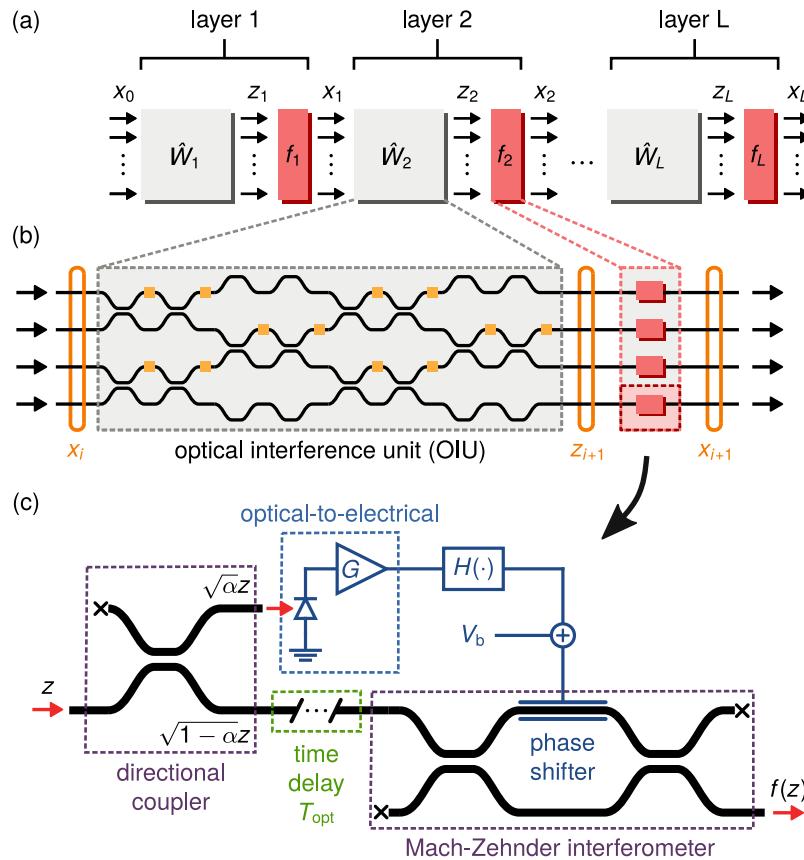


Figure 4.2: (a) Block diagram of a feedforward neural network of  $L$  layers. Each layer consists of a  $\hat{W}_i$  block representing a linear matrix which multiplies vector inputs  $x_{i-1}$ . The  $f_i$  block in each layer represents an element-wise nonlinear activation function operating on vectors  $z_i$  to produce outputs  $x_i$ . (b) Schematic of the optical interferometer mesh implementation of a single layer of the feedforward neural network. (c) Schematic of the proposed optical-to-optical activation function which achieves a nonlinear response by converting a small portion of the optical input,  $z$  into an electrical signal, and then intensity modulating the remaining portion of the original optical signal as it passes through an interferometer.

As a demonstration, we take a mean squared cost function

$$\mathcal{L} = \frac{1}{2} (\mathbf{X}_L - \mathbf{T})^\dagger (\mathbf{X}_L - \mathbf{T}), \quad (4.4)$$

where  $\mathbf{T}$  is a complex-valued target vector corresponding to the desired output of our system given input  $\mathbf{X}_0$ .

Starting from the last layer in the circuit, the derivative of the cost function with respect to the permittivity  $\epsilon_L$  of one of the phase shifters in the last layer is given by

$$\frac{d\mathcal{L}}{d\epsilon_L} = \mathcal{R} \left\{ (\mathbf{X}_L - \mathbf{T})^\dagger \frac{d\mathbf{X}_L}{d\epsilon_L} \right\} \quad (4.5)$$

$$= \mathcal{R} \left\{ \left( \boldsymbol{\Gamma}_L \odot \mathbf{f}_L'(\mathbf{Z}_L) \right)^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right\} \quad (4.6)$$

$$\equiv \mathcal{R} \left\{ \boldsymbol{\delta}_L^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right\}, \quad (4.7)$$

where  $\odot$  is element-wise vector multiplication, defined such that, for vectors  $\mathbf{a}$  and  $\mathbf{b}$ , the  $i$ -th element of the vector  $\mathbf{a} \odot \mathbf{b}$  is given by  $a_i b_i$ .  $\mathcal{R}\{\cdot\}$  gives the real part,  $\mathbf{f}_l'(\cdot)$  is the derivative of the  $l$ th layer activation function with respect to its (complex) argument. We define the vector  $\boldsymbol{\delta}_L \equiv \boldsymbol{\Gamma}_L \odot \mathbf{f}_L'$  in terms of the error vector  $\boldsymbol{\Gamma}_L \equiv (\mathbf{X}_L - \mathbf{T})^*$ .

For any layer  $l < L$ , we may use the chain rule to perform a recursive calculation of the gradients

$$\boldsymbol{\Gamma}_l = \hat{W}_{l+1}^T \boldsymbol{\delta}_{l+1} \quad (4.8)$$

$$\boldsymbol{\delta}_l = \boldsymbol{\Gamma}_l \odot \mathbf{f}_l'(\mathbf{Z}_l) \quad (4.9)$$

$$\frac{d\mathcal{L}}{d\epsilon_l} = \mathcal{R} \left\{ \boldsymbol{\delta}_l^T \frac{d\hat{W}_l}{d\epsilon_l} \mathbf{X}_{l-1} \right\}. \quad (4.10)$$

Figure 4.1(b) diagrams this process, which computes the  $\boldsymbol{\delta}_l$  vectors sequentially from the output layer to the input layer. A treatment for non-holomorphic activations is derived in the Appendix.

We note that the computation of  $\boldsymbol{\delta}_l$  requires performing the operation  $\boldsymbol{\Gamma}_l = \hat{W}_{l+1}^T \boldsymbol{\delta}_{l+1}$ , which corresponds physically to sending  $\boldsymbol{\delta}_{l+1}$  into the output end of the OIU in layer  $l + 1$ . In this way, our procedure ‘backpropagates’ the vectors  $\boldsymbol{\delta}_l$  and  $\boldsymbol{\Gamma}_l$  physically through the entire circuit.

### 4.3 Experimental Measurement of Adjoint Gradient

In the previous Section, we showed that the crucial step in training the ANN is computing gradient terms of the form  $\mathcal{R} \left\{ \boldsymbol{\delta}_l^T \frac{d\hat{W}_l}{d\epsilon_l} \mathbf{X}_{l-1} \right\}$ , which contain derivatives with respect to the permittivity of the phase shifters in the OIUs. In this Section, we show how this gradient may be expressed as the

solution to an electromagnetic adjoint problem.

The OIU used to implement the matrix  $\hat{W}_l$ , relating the complex mode amplitudes of input and output ports, can be described using first-principles electrodynamics. This will allow us to compute its gradient with respect to each  $\epsilon_l$ , as these are the physically adjustable parameters in the system. Assuming a source at frequency  $\omega$ , at steady state Maxwell's equations take the form

$$\left[ \hat{\nabla} \times \hat{\nabla} \times - k_0^2 \hat{\epsilon}_r \right] \mathbf{e} = -i\omega\mu_0 \mathbf{j}, \quad (4.11)$$

which can be written more succinctly as

$$\hat{A}(\epsilon_r) \mathbf{e} = \mathbf{b}. \quad (4.12)$$

Here,  $\hat{\epsilon}_r$  describes the spatial distribution of the relative permittivity ( $\epsilon_r$ ),  $k_0 = \omega^2/c^2$  is the free-space wavenumber,  $\mathbf{e}$  is the electric field distribution,  $\mathbf{j}$  is the electric current density, and  $\hat{A} = \hat{A}^T$  due to Lorentz reciprocity. Eq. (4.12) is the starting point of the finite-difference frequency-domain (FDFD) simulation technique [174], where it is discretized on a spatial grid, and the electric field  $\mathbf{e}$  is solved given a particular permittivity distribution,  $\epsilon_r$ , and source,  $\mathbf{b}$ .

To relate this formulation to the transfer matrix  $\hat{W}$ , we now define source terms  $\mathbf{b}_i$ ,  $i \in 1 \dots 2N$ , that correspond to a source placed in one of the input or output ports. Here we assume a total of  $N$  input and  $N$  output waveguides. The spatial distribution of the source term,  $\mathbf{b}_i$ , matches the mode of the  $i$ -th single-mode waveguide. Thus, the electric field amplitude in port  $i$  is given by  $\mathbf{b}_i^T \mathbf{e}$ , and we may establish a relationship between  $\mathbf{e}$  and  $\mathbf{X}_{\text{in}}$ , as

$$X_{\text{in},i} = \mathbf{b}_i^T \mathbf{e} \quad (4.13)$$

for  $i = 1 \dots N$  over the input port indices, where  $X_{\text{in},i}$  is the  $i$ -th component of  $\mathbf{X}_{\text{in}}$ . Or more compactly,

$$\mathbf{X}_{\text{in}} \equiv \hat{P}_{\text{in}} \mathbf{e}, \quad (4.14)$$

Similarly, we can define

$$Z_{\text{out},i} = \mathbf{b}_{i+N}^T \mathbf{e} \quad (4.15)$$

for  $i + N = (N + 1) \dots 2N$  over the output port indices, or,

$$\mathbf{Z}_{\text{out}} \equiv \hat{P}_{\text{out}} \mathbf{e}, \quad (4.16)$$

and, with this notation, Eq. (4.1) becomes

$$\hat{W} \hat{P}_{\text{in}} \mathbf{e} = \hat{P}_{\text{out}} \mathbf{e} \quad (4.17)$$

We now use the above definitions to evaluate the cost function gradient in Eq. (4.10). In particular, with Eqs. (4.10) and (4.17), we arrive at

$$\frac{d\mathcal{L}}{d\epsilon_l} = -\mathcal{R} \left\{ \boldsymbol{\delta}_l^T \hat{P}_{\text{out}} \hat{A}^{-1} \frac{d\hat{A}}{d\epsilon_l} \hat{A}^{-1} \mathbf{b}_{x,l-1} \right\}. \quad (4.18)$$

Here  $\mathbf{b}_{x,l-1}$  is the modal source profile that creates the input field amplitudes  $\mathbf{X}_{l-1}$  at the input ports.

The key insight of the adjoint variable method is that we may interpret this expression as an operation involving the field solutions of two electromagnetic simulations, which we refer to as the ‘original’ (og) and the ‘adjoint’ (aj)

$$\hat{A}\mathbf{e}_{\text{og}} = \mathbf{b}_{x,l-1} \quad (4.19)$$

$$\hat{A}\mathbf{e}_{\text{aj}} = \hat{P}_{\text{out}}^T \boldsymbol{\delta}, \quad (4.20)$$

where we have made use of the symmetric property of  $\hat{A}$ .

Eq. (4.18) can now be expressed in a compact form as

$$\frac{d\mathcal{L}}{d\epsilon_l} = -\mathcal{R} \left\{ \mathbf{e}_{\text{aj}}^T \frac{d\hat{A}}{d\epsilon_l} \mathbf{e}_{\text{og}} \right\}. \quad (4.21)$$

If we assume that this phase shifter spans a set of points,  $\mathbf{r}_\phi$  in our system, then, from Eq. (4.11), we obtain

$$\frac{d\hat{A}}{d\epsilon_l} = -k_0^2 \sum_{\mathbf{r}' \in \mathbf{r}_\phi} \hat{\delta}_{\mathbf{r}, \mathbf{r}'}, \quad (4.22)$$

where  $\hat{\delta}_{\mathbf{r}, \mathbf{r}'}$  is the Kronecker delta.

Inserting this into Eq. (4.21), we thus find that the gradient is given by the overlap of the two fields over the phase-shifter positions

$$\frac{d\mathcal{L}}{d\epsilon_l} = k_0^2 \mathcal{R} \left\{ \sum_{\mathbf{r} \in \mathbf{r}_\phi} \mathbf{e}_{\text{aj}}(\mathbf{r}) \mathbf{e}_{\text{og}}(\mathbf{r}) \right\}. \quad (4.23)$$

This result now allows for the computation in parallel of the gradient of the loss function with respect to *all* phase shifters in the system, given knowledge of the original and adjoint fields.

We now propose a method to compute the gradient from the previous section through *in situ* intensity measurements. This represents the most significant result of this paper. Specifically, we wish to generate an intensity pattern with the form  $\mathcal{R}\{\mathbf{e}_{\text{og}} \mathbf{e}_{\text{aj}}\}$ , matching that of Eq. (4.23). We

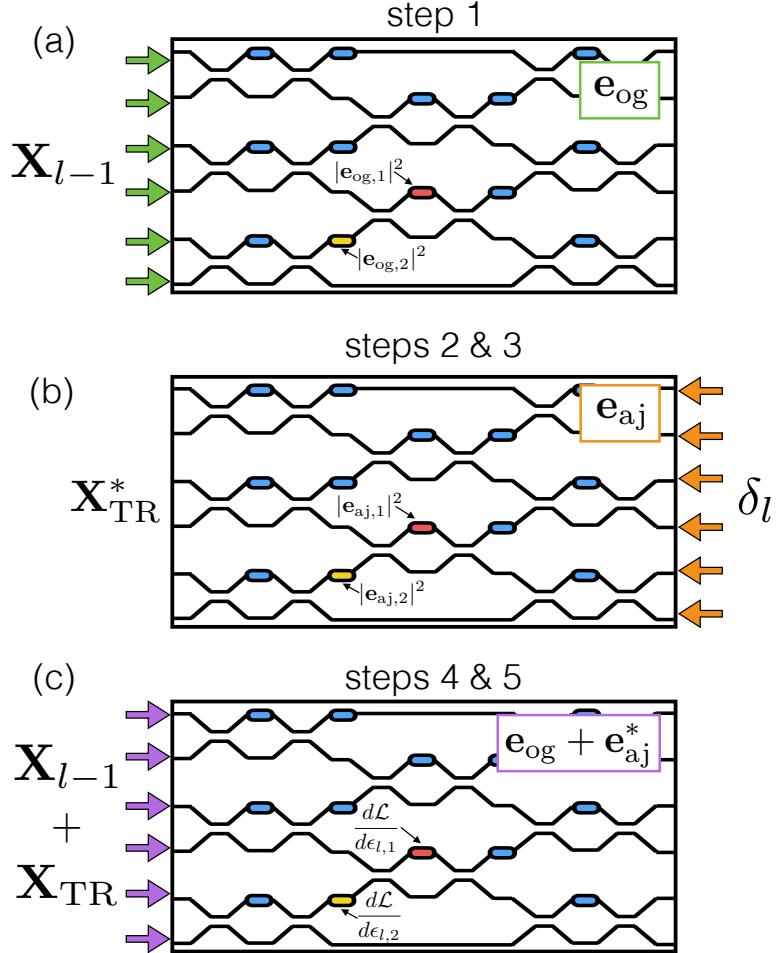


Figure 4.3: Schematic illustration of our proposed method for experimental measurement of gradient information. The box region represents the OIU. The colored ovals represent tunable phase shifters, and we illustrate computing the gradient with respect to the red and the yellow phase shifters, labeled 1 and 2, respectively. (a): We send the original set of amplitudes  $\mathbf{X}_{l-1}$  and measure the constant intensity terms at each phase shifter. (b): We send the adjoint mode amplitudes, given by  $\delta_l$ , through the output side of our device, recording  $\mathbf{X}_{TR}^*$  from the opposite side, as well as  $|e_{aj}|^2$  in each phase-shifter. (c): We send in  $\mathbf{X}_{l-1} + \mathbf{X}_{TR}$ , interfering  $e_{og}$  and  $e_{aj}^*$  inside the device and recovering the gradient information for all phase shifters simultaneously.

note that interfering  $e_{og}$  and  $e_{aj}^*$  directly in the system results in the intensity pattern:

$$I = |e_{og}|^2 + |e_{aj}|^2 + 2\Re\{e_{og}e_{aj}\}, \quad (4.24)$$

the last term of which matches Eq. (4.23). Thus, the gradient can be computed purely through intensity measurements if the field  $e_{aj}^*$  can be generated in the OIU.

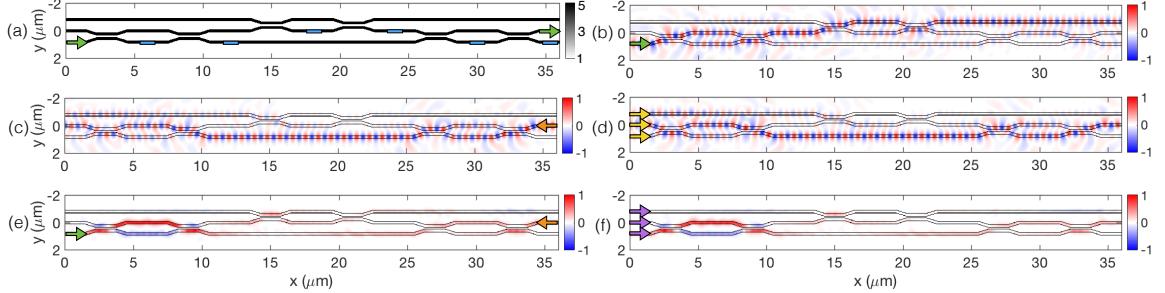


Figure 4.4: Numerical demonstration of the time-reversal procedure. (a): Relative permittivity distribution for three MZIs arranged to perform a 3x3 linear operation. Blue boxes represent where phase shifters would be placed in this system. As an example, we compute the gradient information for a layer with  $\mathbf{X}_{l-1} = [0 \ 0 \ 1]^T$  and  $\boldsymbol{\delta}_l = [0 \ 1 \ 0]^T$ , corresponding to the bottom left and middle right port, respectively. (b): Real part of the simulated electric field  $E_z$  corresponding to injection from the bottom left port. (c): Real part of the adjoint  $E_z$ , corresponding to injection from the middle right port. (d): Time-reversed adjoint field as constructed by our method, fed in through all three ports on the left. (e): The gradient information  $\frac{d\mathcal{L}}{d\epsilon_l}(x, y)$  as obtained directly by the adjoint method, normalized by its maximum absolute value. (f): The gradient information as obtained by the method introduced in this work, normalized by its maximum absolute value. Namely, the field pattern from (b) is interfered with the time-reversed adjoint field of (d) and the constant intensity terms are subtracted from the resulting intensity pattern. Panels (e) and (f) match with high precision.

The adjoint field for our problem,  $\mathbf{e}_{aj}$ , as defined in Eq. (4.20), is sourced by  $\hat{P}_{out}^T \boldsymbol{\delta}_l$ , meaning that it physically corresponds to a mode sent into the system from the output ports. As complex conjugation in the frequency domain corresponds to time-reversal of the fields, we expect  $\mathbf{e}_{aj}^*$  to be sent in from the input ports. Formally, to generate  $\mathbf{e}_{aj}^*$ , we wish to find a set of input source amplitudes,  $\mathbf{X}_{TR}$ , such that the output port source amplitudes,  $\mathbf{Z}_{TR} = \hat{W} \mathbf{X}_{TR}$ , are equal to the complex conjugate of the adjoint amplitudes, or  $\boldsymbol{\delta}_l^*$ . Using the unitarity property of transfer matrix  $\hat{W}_l$  for a lossless system, along with the fact that  $\hat{P}_{out} \hat{P}_{out}^T = \hat{I}$  for output modes, the input mode amplitudes for the time-reversed adjoint can be computed as

$$\mathbf{X}_{TR}^* = \hat{W}_l^T \boldsymbol{\delta}_l. \quad (4.25)$$

As discussed earlier,  $\hat{W}_l^T$  is the transfer matrix from output ports to input ports. Thus, we can experimentally determine  $\mathbf{X}_{TR}$  by sending  $\boldsymbol{\delta}_l$  into the device output ports, measuring the output at the input ports, and taking the complex conjugate of the result.

We now summarize the procedure for experimentally measuring the gradient of an OIU layer in the ANN with respect to the permittivities of this layer's integrated phase shifters:

1. Send in the original field amplitudes  $\mathbf{X}_{l-1}$  and measure and store the intensities at each phase shifter.
2. Send  $\boldsymbol{\delta}_l$  into the output ports and measure and store the intensities at each phase shifter.

3. Compute the time-reversed adjoint input field amplitudes as in Eq. (4.25).
4. Interfere the original and the time-reversed adjoint fields in the device, measuring again the resulting intensities at each phase shifter.
5. Subtract the constant intensity terms from steps 1 and 2 and multiply by  $k_0^2$  to recover the gradient as in Eq. (4.23).

This procedure is also illustrated in Fig. 4.3.

### 4.3.1 Numerical Demonstrations

We numerically demonstrate this procedure in Fig. 4.4 with a series of FDFD simulations of an OIU implementing a  $3 \times 3$  unitary matrix [162]. These simulations are intended to represent the gradient computation corresponding to one OIU in a single layer,  $l$ , of a neural network with input  $\mathbf{X}_{l-1}$  and delta vector  $\boldsymbol{\delta}_l$ . In these simulations, we use absorbing boundary conditions on the outer edges of the system to eliminate back-reflections. The relative permittivity distribution is shown in Fig. 4.4(a) with the positions of the variable phase shifters in blue. For demonstration, we simulate a specific case where  $\mathbf{X}_{l-1} = [0 \ 0 \ 1]^T$ , with unit amplitude in the bottom port and we choose  $\boldsymbol{\delta}_l = [0 \ 1 \ 0]^T$ . In Fig. 4.4(b), we display the real part of  $\mathbf{e}_{\text{og}}$ , corresponding to the original, forward field.

The real part of the adjoint field,  $\mathbf{e}_{\text{aj}}$ , corresponding to the cost function  $\mathcal{L} = \mathcal{R} \left\{ \boldsymbol{\delta}_l^T \hat{W}_l \mathbf{X}_{l-1} \right\}$  is shown in Fig. 4.4(c). In Fig. 4.4(d) we show the real part of the time-reversed copy of  $\mathbf{e}_{\text{aj}}$  as computed by the method described in the previous section, in which  $\mathbf{X}_{TR}^*$  is sent in through the input ports. There is excellent agreement, up to a constant, between the complex conjugate of the field pattern of (c) and the field pattern of (d), as expected.

In Fig. 4.4(e), we display the gradient of the objective function with respect to the permittivity of each point of space in the system, as computed with the adjoint method, described in Eq. (4.23). In Fig. 4.4(f), we show the same gradient information, but instead computed with the method described in the previous section. Namely, we interfere the field pattern from panel (b) with the field pattern from panel (d), subtract constant intensity terms, and multiply by the appropriate constants. Again, (b) and (d) agree with good precision.

We note that in a realistic system, the gradient must be constant for any stretch of waveguide between waveguide couplers because the interfering fields are at the same frequency and are traveling in the same direction. Thus, there should be no distance dependence in the corresponding intensity distribution. This is largely observed in our simulation, although small fluctuations are visible because of the proximity of the waveguides and the sharp bends, which were needed to make the structure compact enough for simulation within a reasonable time. In practice, the importance of this constant intensity is that it can be detected *after* each phase shifter, instead of inside of it.

Finally, we note that this numerically generated system experiences a total power loss of 41% due to scattering caused by very sharp bends and stair-casing of the structure in the simulation. We

also observe approximately 5-10% mode-dependent loss, as determined by measuring the difference in total transmitted power corresponding to injection at different input ports. Minimal amounts of reflection are also visible in the field plots. Nevertheless, the time-reversal interference procedure still reconstructs the adjoint sensitivity with very good fidelity.

Finally, we use the techniques from the previous Sections to numerically demonstrate the training of a photonic ANN to implement a logical XOR gate, defined by the following input to target ( $\mathbf{X}_0 \rightarrow \mathbf{T}$ ) pairs

$$[0\ 0]^T \rightarrow 0, \quad [0\ 1]^T \rightarrow 1, \quad [1\ 0]^T \rightarrow 1, \quad [1\ 1]^T \rightarrow 0. \quad (4.26)$$

This problem was chosen as demonstration of learning a nonlinear mapping from input to output [203] and is simple enough to be solved with a small network with only four training examples.

As diagrammed in Fig. 4.6a, we choose a network architecture consisting of two  $3 \times 3$  unitary OIUs. On the forward propagation step, the binary representation of the inputs,  $\mathbf{X}_0$ , is sent into the first two input elements of the ANN and a constant value of 1 is sent into the third input element, which serves to introduce artificial bias terms into the network. These inputs are sent through a  $3 \times 3$  unitary OIU and then the element-wise activation  $f(z) = z^2$  is applied. The output of this step is sent to another  $3 \times 3$  OIU and sent through another activation of the same form. Finally, the first output element is taken to be our prediction,  $\mathbf{X}_L$ , ignoring the last two output elements. Our network is repeatedly trained on the four training examples defined in Eq. (4.26) and using the mean-squared cost function presented in Eq. (4.4).

For this demonstration, we utilized a matrix model of the system, as described in [162, 39]. This model allows us to compute an output of the system given an input mode and the settings of each phase shifter. Although this is not a first-principle electromagnetic simulation of the system, it provides information about the complex fields at specific reference points within the circuit, which enables us to implement training using the backpropagation method, combined with the adjoint gradient calculation. Using these methods, at each iteration of training we compute the gradient of our cost function with respect to the phases of each of the integrated phase shifters, and sum them over the four training examples. Then, we perform a simple steepest-descent update to the phase shifters, in accordance with the gradient information. This is consistent with the standard training protocol for an ANN implemented on a conventional computer. Our network successfully learned the XOR gate in around 400 iterations. The results of the training are shown in Fig. 4.6b-d.

Next, we demonstrate training on a more complex problem. Specifically, we generate a set of one thousand training examples represented by input and target ( $\mathbf{X}_0 \rightarrow \mathbf{T}$ ) pairs. Here,  $\mathbf{X}_0 = [x_1, x_2, P, 0]^T$  where  $x_1$  and  $x_2$  are the independent inputs, which we constrain to be real for simplicity, and  $P(x_1, x_2) = \sqrt{P_0 - x_1^2 - x_2^2}$  represents a mode added to the third port to make the norm of  $\mathbf{X}_0$  the same for each training example. In this case, we choose  $P_0 = 10$ . Each training example has a corresponding label,  $y \in \{0, 1\}$  which is encoded in the desired output,  $\mathbf{T}$ , as  $[1, 0, 0, 0]^T$  and  $[0, 1, 0, 0]^T$  for  $y = 0$  and  $y = 1$  respectively.

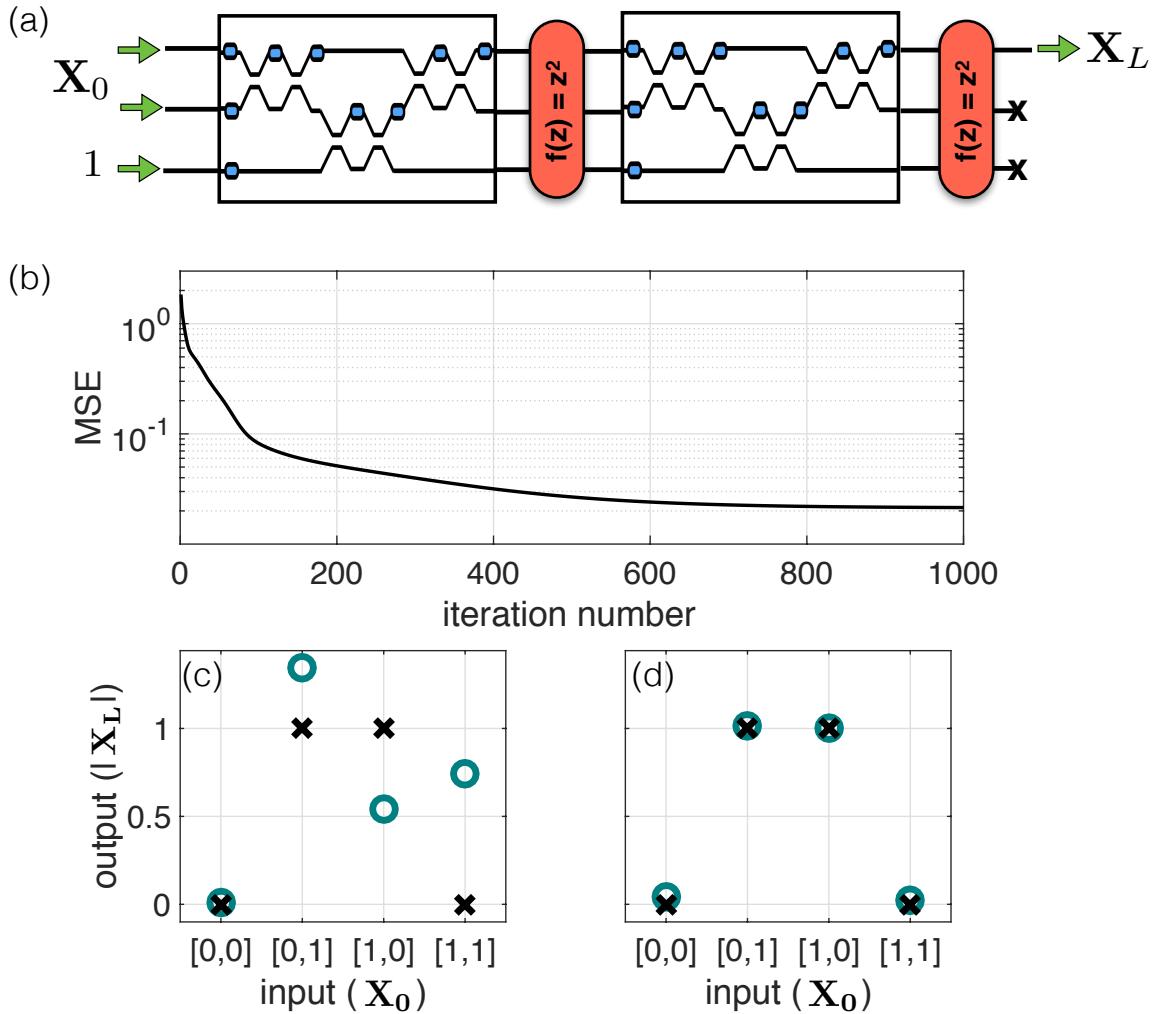


Figure 4.5: Numerical demonstration of a photonic ANN implementing an XOR gate using the backpropagation algorithm and adjoint method described in this work. (a) The architecture of the ANN. Two layers of  $3 \times 3$  OIUs with  $z^2$  activations. (b) The mean-squared error (MSE) between the predictions and targets as a function of training iterations. (c) The absolute value of the network predictions (blue circles) and targets (black crosses) before training. (d) The absolute value of the network predictions after training, showing that the network has successfully learned the XOR function.

For a given  $x_1$  and  $x_2$ , we define  $r$  and  $\phi$  as the magnitude and phase of the vector  $(x_1, x_2)$  in the 2D-plane, respectively. To generate the corresponding class label, we first generate a uniform random variable between 0 and 1, labeled  $\mathcal{U}$ , and then set  $y = 1$  if

$$\exp\left(-\frac{(r - r_0 - \Delta \sin(2\phi))^2}{2\sigma^2}\right) + 0.1 \mathcal{U} > 0.5. \quad (4.27)$$

Otherwise, we set  $y = 0$ . For the demonstration,  $r_0 = 0.6$ ,  $\Delta = 0.15$ , and  $\sigma = 0.2$ . The underlying distribution thus resembles an oblong ring centered around  $x_1 = x_2 = 0$ , with added noise.

As diagrammed in Fig. 4.6(a), we use a network architecture consisting of six  $4 \times 4$  layers of unitary OIUs, with an element-wise activation  $f(z) = |z|$  after each unitary transformation except for the last in the series, which has an activation of  $f(z) = |z|^2$ . After the final activation, we apply an additional ‘softmax’ activation, which gives a normalized probability distribution corresponding to the predicted class of  $\mathbf{X}_0$ . Specifically, these are given by  $s(z_i) = \exp(z_i) / (\sum_j \exp(z_j))$ , where  $z_{i=1,2}$  is the first/second element of the output vector of the last activation (the other two elements are ignored). The ANN prediction for the input  $\mathbf{X}_0$  is set as the larger one of these two outputs, while the total cost function is defined in the cross-entropy form

$$\mathcal{L} = \frac{1}{M} \sum_{m=1}^M \mathcal{L}^{(m)} = \frac{1}{M} \sum_{m=1}^M -\log(s(z_{m,t})), \quad (4.28)$$

where  $\mathcal{L}^{(m)}$  is the cost function of the  $m$ -th example, the summation is over all training examples, and  $z_{m,t}$  is the output from the target port,  $t$ , as defined by the target output  $\mathbf{T}^{(m)}$  of the  $m$ -th example. We randomly split our generated examples into a training set containing 75% of the originally generated training examples, while the remaining 25% are used as a test set to evaluate the performance of our network on unseen examples.

At each iteration of training we compute the gradient of the cost function with respect to the phases of each of the integrated phase shifters, and sum this over each of the training examples. For the backpropagation through the activation functions, since  $|z|$  and  $|z|^2$  are non-holomorphic, we use the notes from Appendix A to obtain

$$\boldsymbol{\delta}_L = 2\mathbf{Z}_L^* \odot \mathcal{R}\{\boldsymbol{\Gamma}_L\} \quad (4.29)$$

$$\boldsymbol{\delta}_l = \exp(-i\boldsymbol{\phi}_l) \odot \mathcal{R}\{\boldsymbol{\Gamma}_l\}, \quad (4.30)$$

where  $\boldsymbol{\phi}_l$  is a vector containing the phases of  $\mathbf{Z}_l$  and  $\boldsymbol{\Gamma}_L$  is given by the derivative of the cross-entropy loss function for a single training example

$$\boldsymbol{\Gamma}_L = \frac{\partial \mathcal{L}^{(m)}}{\partial z_{m,i}} = s(z_{m,i}) - \delta_{i,t}, \quad (4.31)$$

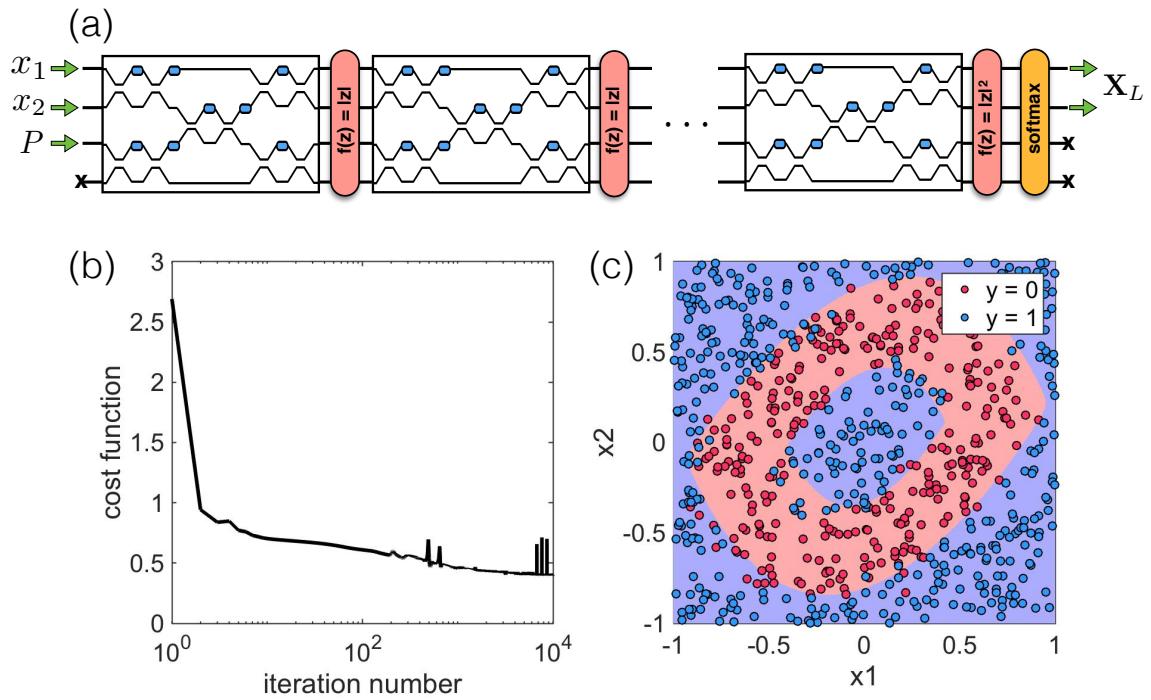


Figure 4.6: Numerical demonstration of a photonic ANN learning to classify an oblong ring. (a) The architecture of the ANN. Six layers of  $4 \times 4$  OIUs with  $|z|$  activations. A final softmax activation is applied at the very end. (b) The loss function of Eq. (4.28) over training iterations. (c) The training examples, blue and red dots correspond to  $y = 0$  and  $y = 1$  labels on a given  $x_1$  and  $x_2$  input. The background shows the prediction of the network on a continuum of  $x_1$  and  $x_2$  pairs, with colors representing the corresponding predictions. One can see that the ring was learned successfully without overfitting.

where  $\delta_{i,t}$  is the Kronecker delta.

With this, we can now compute the gradient of the loss function of eq. 4.28 with respect to all trainable parameters, and perform a parallel, steepest-descent update to the phase shifters, in accordance with the gradient information. Our network successfully learned the this task in around 4000 iterations. The results of the training are shown in Fig. 4.6(b). We achieved a training and test accuracy of 91% on both the training and test sets, indicating that the network was not overfitting to the dataset. This can also be confirmed visually from Fig. 4.6(c). The lack of perfect predictions is likely due to the inclusion of noise.

### 4.3.2 Discussion

Here, we justify some of the assumptions made in this work. Our strategy for training a photonic ANN relies on the ability to create arbitrary complex inputs. We note that a device for accomplishing this has been proposed and discussed in [128]. Our recovery technique further requires an integrated intensity detection scheme to occur in parallel and with virtually no loss. This may be implemented by integrated, transparent photo-detectors, which have already been demonstrated in similar systems [9]. Furthermore, as discussed, this measurement may occur in the waveguide regions directly after the phase shifters, which eliminates the need for phase shifter and photodetector components at the same location. Finally, in our procedure for experimentally measuring the gradient information, we suggested running isolated forward and adjoint steps, storing the intensities at each phase shifter for each step, and then subtracting this information from the final interference intensity. Alternatively, one may bypass the need to store these constant intensities by introducing a low-frequency modulation on top of one of the two interfering fields in Fig. 4.3(c), such that the product term of Eq. (4.24) can be directly measured from the low-frequency signal. A similar technique was used in [9].

We now discuss some of the limitations of our method. In the derivation, we had assumed the  $\hat{W}$  operator to be unitary, which corresponds to a lossless OIU. In fact, we note that our procedure is exact in the limit of a lossless, feed-forward, and reciprocal system. However, with the addition of any amount of uniform loss,  $\hat{W}$  is still unitary up to a constant, and our procedure may still be performed with the added step of scaling the measured gradients depending on this loss (see a related discussion in Ref. [128]). Uniform loss conditions are satisfied in the OIUs experimentally demonstrated in Refs. [170, 124]. Mode-dependent loss, such as asymmetry in the MZI mesh layout or fabrication errors, should be avoided as its presence limits the ability to accurately reconstruct the time-reversed adjoint field. Nevertheless, our simulation in Fig. 4.4 indicates that an accurate gradient can be obtained even in the presence of significant mode-dependent loss. In the experimental structures of Refs. [170, 124], the mode-dependent loss is made much lower due to the choice of the MZI mesh. Thus we expect our protocol to work in practical systems. Our method, in principle, computes gradients in parallel and scales in constant time. In practice, to get this scaling would

require careful design of the circuits controlling the OIUs.

Conveniently, since our method does not directly assume any specific model for the linear operations, it may gracefully handle imperfections in the OIUs, such as deviations from perfect 50-50 splits in the MZIs. Lastly, while we chose to make an explicit distinction between the input ports and the output ports, i.e. we assume no backscattering in the system, this requirement is not strictly necessary. Our formalism can be extended to the full scattering matrix. However, this would require special treatment for subtracting the backscattering.

The problem of overfitting is one that must be addressed by ‘regularization’ in any practical realization of a neural network. Photonic ANNs of this class provide a convenient approach to regularization based on ‘dropout’ [186]. In the dropout procedure, certain nodes are probabilistically and temporarily ‘deleted’ from the network during train time, which has the effect of forcing the network to find alternative paths to solve the problem at hand. This has a strong regularization effect and has become popular in conventional ANNs. Dropout may be implemented simply in the photonic ANN by ‘shutting off’ channels in the activation functions during training. Specifically, at each time step and for each layer  $l$  and element  $i$ , one may set  $f_l(Z_i) = 0$  with some fixed probability.

## 4.4 Conclusion

In this Chapter, we have demonstrated a method for performing backpropagation in an ANN based on a photonic circuit. This method works by physically propagating the adjoint field and interfering its time-reversed copy with the original field. The gradient information can then be directly measured out as an *in-situ* intensity measurement. While we chose to demonstrate this procedure in the context of ANNs, it is broadly applicable to any reconfigurable photonic system. One could imagine this setup being used to tune phased arrays [191], optical delivery systems for dielectric laser accelerators [83], or other systems that rely on large meshes of integrated optical phase shifters. Furthermore, it may be applied to sensitivity analysis of photonic devices, enabling spatial sensitivity information to be measured as an intensity in the device.

Our work should enhance the appeal of photonic circuits in deep learning applications, allowing for training to happen directly inside the device in an efficient and scalable manner. Furthermore, this method is broadly applicable to integrated and adaptive optical systems, enabling the possibility for automatic self-configuration and optimization without resorting to brute force gradient computation or model-based methods, which often do not perfectly represent the physical system.

## Chapter 5

# Wave-Based Analog Recurrent Neural Networks

In the previous Chapter, we demonstrated the use of the adjoint method to compute the gradient of a device that is parameterized by optical phase shifters. We showed that this formalism allows one to train a common class of optical neural network based on MZI meshes. In this Chapter, we will continue to explore the use of the adjoint method in training of optical machine learning hardware, but will take a more general approach. Specifically, we will discuss the use of an optical system (or a wave system, more generally), as an analog computing platform for processing time series signals. As we will show, this capability is made possible by the use of the adjoint method, which allows one to come up with a structure to solve a particular machine learning task through inverse design and optimization.

### 5.1 Wave Physics as a Recurrent Neural Network

Analog computing platforms, which use the natural evolution of a continuous physical system to perform calculations, are emerging as important direction for the implementation of machine learning models [171, 19, 107, 113, 96]. Such a scheme has a number of potential benefits. Namely, they are able to *passively* process signals and information in their native domain, without analog-to-digital conversion, which should result in a significant gain in speed and a reduction in power consumption.

Here, we identify a mapping between the dynamics of wave-based physical phenomena, such as acoustics and optics, and the computation in a recurrent neural network (RNN). RNNs are one of the most important machine learning models and have been widely used to perform tasks such as natural language processing [219] and time-series prediction [85, 47, 41]. Here we show that wave-based physical systems can be trained to operate as an RNN, and as a result can *passively* process

signals and information in their native domain, without analog-to-digital conversion, which should result in a significant gain in speed and a reduction in power consumption.

An RNN converts a sequence of inputs into a sequence of outputs by applying the same basic operation to each member of the input sequence in a step-by-step fashion (Fig 5.1A). Memory of previous time steps is encoded into the RNN's *hidden state*, which is updated at each step. The information in the hidden state allows for the RNN to learn temporal structure and long-range dependencies in data [51, 92]. At a given time step,  $t$ , the RNN operates on the current input vector in the sequence,  $\mathbf{x}_t$ , and the hidden state vector from the previous step,  $\mathbf{h}_{t-1}$ , to produce an output vector,  $\mathbf{y}_t$ , as well as an updated hidden state,  $\mathbf{h}_t$ . While many variations of RNNs exist, a common implementation [60] is described by the following update equations

$$\mathbf{h}_t = \sigma^{(h)} \left( \mathbf{W}^{(h)} \cdot \mathbf{h}_{t-1} + \mathbf{W}^{(x)} \cdot \mathbf{x}_t \right) \quad (5.1)$$

$$\mathbf{y}_t = \sigma^{(y)} \left( \mathbf{W}^{(y)} \cdot \mathbf{h}_t \right), \quad (5.2)$$

which are diagrammed in Fig. 5.1B. As in the standard feedforward neural network case, the dense matrices defined by  $\mathbf{W}^{(h)}$ ,  $\mathbf{W}^{(x)}$ , and  $\mathbf{W}^{(y)}$  are optimized during training while  $\sigma^{(h)}(\cdot)$  and  $\sigma^{(y)}(\cdot)$  are fixed nonlinear activation functions. The operation defined by Eq. 5.1 and Eq. 5.2, when applied to each element of an input sequence, can be described by the directed graph shown in Fig. 5.1C.

We now discuss the connection between the dynamics in the RNN as described by Eqs. 5.1 and 5.2, and the dynamics of a wave-based physical system. We explore the dynamics of a scalar field distribution,  $u = u(x, y, z, t)$ , as governed by the damped wave equation [52]

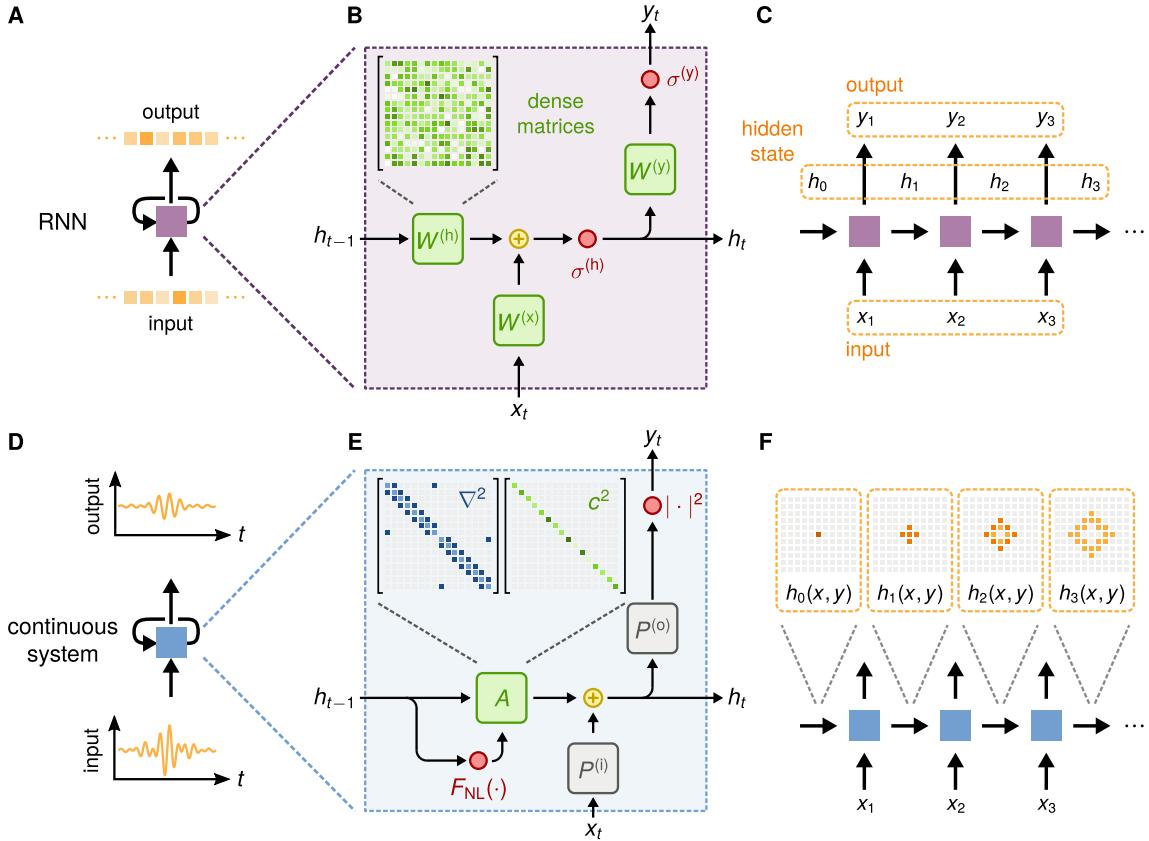
$$\frac{\partial^2 u}{\partial t^2} + 2b \cdot \frac{\partial u}{\partial t} = c^2 \cdot \nabla^2 u + f, \quad (5.3)$$

where  $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$  is the Laplacian operator.  $c = c(x, y, z)$  is the spatial distribution of the wave speed and  $f = f(x, y, z, t)$  is a source term.  $b = b(x, y, z)$  is the dampening coefficient which is spatially varying but frequency-independent. For a time step indexed by  $t$ , Eq. 5.3 is discretized using *centered* finite differences in time to give

$$\frac{u_{t+1} - 2u_t + u_{t-1}}{\Delta t^2} + 2b \frac{u_{t+1} - u_{t-1}}{2\Delta t} = c^2 \nabla^2 u_t + f_t, \quad (5.4)$$

where the subscript in  $(\cdot)_t$  is used to indicate the value of a scalar field at time step  $t$ .

From Eq. 5.4, we may form a recurrence relation in terms of  $u_{t+1}$ , which leads to the following



**Figure 5.1: Conceptual comparison of a standard recurrent neural network and a wave-based physical system.** (A) Diagram of a recurrent neural network (RNN) cell operating on a discrete input sequence and producing a discrete output sequence. (B) Internal components of the RNN cell, consisting of trainable dense matrices  $\mathbf{W}^{(h)}$ ,  $\mathbf{W}^{(x)}$ , and  $\mathbf{W}^{(y)}$ . Activation functions for the hidden state and output are represented by  $\sigma^{(h)}$  and  $\sigma^{(y)}$ , respectively. (C) Diagram of the directed graph of the RNN cell. (D) Diagram of a recurrent representation of a continuous physical system operating on a continuous input sequence and producing a continuous output sequence. (E) Internal components of the recurrence relation for the wave equation when discretized using finite differences. (F) Diagram of the directed graph of discrete time steps of the continuous physical system.

update equation

$$\begin{aligned} \left( \frac{1}{\Delta t^2} + \frac{b}{\Delta t} \right) u_{t+1} - \frac{2}{\Delta t^2} u_t + \left( \frac{1}{\Delta t^2} - \frac{b}{\Delta t} \right) u_{t-1} &= c^2 \cdot \nabla^2 u_t + f_t \\ \left( \frac{1}{\Delta t^2} + \frac{b}{\Delta t} \right) u_{t+1} &= \frac{2}{\Delta t^2} u_t - \left( \frac{1}{\Delta t^2} - \frac{b}{\Delta t} \right) u_{t-1} + c^2 \cdot \nabla^2 u_t + f_t \\ u_{t+1} &= \left( \frac{1}{\Delta t^2} + \frac{b}{\Delta t} \right)^{-1} \left[ \frac{2}{\Delta t^2} u_t - \left( \frac{1}{\Delta t^2} - \frac{b}{\Delta t} \right) u_{t-1} + c^2 \cdot \nabla^2 u_t + f_t \right]. \end{aligned} \quad (5.5)$$

Equation 5.5 therefore represents the discretized update equation for the scalar wave equation with damping. In matrix form, we may express Eq. (5.5) as

$$\begin{bmatrix} u_{t+1} \\ u_t \end{bmatrix} = \begin{bmatrix} \frac{2+\Delta t^2 \cdot c^2 \cdot \nabla^2}{1+\Delta t \cdot b} & \frac{-1-\Delta t \cdot b}{1+\Delta t \cdot b} \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} u_t \\ u_{t-1} \end{bmatrix} + \Delta t^2 \cdot \begin{bmatrix} f_t \\ 0 \end{bmatrix}. \quad (5.6)$$

Nonlinearity may be introduced into the system by assuming an intensity-dependent wave speed of the form  $c(x, y) = c_{\text{linear}} + c_{\text{nonlinear}} \cdot |u_t(x, y)|^2$ , where  $c_{\text{nonlinear}}$  is exhibited in regions containing nonlinear materials. In practice, this class of nonlinearity is encountered in a variety of wave physics, such as shallow water waves [200] or nonlinear optics via the Kerr effect [21].

To make the connection to the standard RNN more apparent, we define the wave system's *hidden state* as the concatenation of the field distributions at the current and immediately preceding time steps,  $\mathbf{h}_t \equiv [\mathbf{u}_t, \mathbf{u}_{t-1}]^T$ , where  $\mathbf{u}_t$  and  $\mathbf{u}_{t-1}$  are vectors representing the *flattened* fields,  $u_t$  and  $u_{t-1}$ , discretized in the spatial domain. With this, we may express the update of the wave equation from Eq. (5.6) as

$$\mathbf{h}_t = \mathbf{A}(\mathbf{h}_{t-1}) \cdot \mathbf{h}_{t-1} + \mathbf{P}^{(i)} \cdot \mathbf{x}_t \quad (5.7)$$

$$\mathbf{y}_t = |\mathbf{P}^{(o)} \cdot \mathbf{h}_t|^2, \quad (5.8)$$

where we have defined  $\mathbf{A}$  as the matrix appearing in Eq. (5.6) and its dependence on  $\mathbf{h}_{t-1}$  is caused by the inclusion of the nonlinear wave speed as described above.

Like the standard RNN, the connections between the hidden state and the input and output of the wave equation are also defined by linear operators, given by  $\mathbf{P}^{(i)}$  and  $\mathbf{P}^{(o)}$ . These matrices define injection and measuring points within the spatial domain. Unlike the standard RNN, where the input and output matrices are dense, the input and output matrices of the wave equation are usually sparse and, moreover, unchanged by the training process. The nonlinear relationship between the hidden state,  $\mathbf{h}_t$ , and the output,  $\mathbf{y}_t$ , of the wave equation is typical in wave physics since the output usually corresponds to a measurement of the wave intensity.

To more explicitly define the input and output of this RNN, we introduce the linear operators,  $\mathbf{M}^{(i)}$  and  $\mathbf{M}^{(o)}$ , each column of which define the respective spatial distributions of the injection and

measurement points in this flattened basis. With this, we can write the injection of the input vector ( $\mathbf{x}_t$ ) as a matrix-vector multiplication

$$\Delta t^2 \mathbf{f}_t \equiv \mathbf{M}^{(i)} \cdot \mathbf{x}_t. \quad (5.9)$$

Similarly, as the output of the RNN at each time step is given by an intensity measurement of the scalar fields, we may express this in terms of the flattened scalar field as

$$\mathbf{y}_t = |\mathbf{M}^{(o)T} \cdot \mathbf{u}_t|^2. \quad (5.10)$$

As the wave equation *hidden state*,  $\mathbf{h}_t$  is defined as the concatenation of  $\mathbf{u}_t$  and  $\mathbf{u}_{t-1}$ , we define the following matrices for convenience, as they only act on the  $\mathbf{u}_t$  portion of  $\mathbf{h}_t$

$$\mathbf{P}^{(i)} \equiv \begin{bmatrix} \mathbf{M}^{(i)} \\ \mathbf{0} \end{bmatrix} \quad (5.11)$$

$$\mathbf{P}^{(o)} \equiv [\mathbf{M}^{(o)T}, \mathbf{0}], \quad (5.12)$$

where  $\mathbf{0}$  is a matrix of all zeros. These matrices are used in the injection and measurement stages of the scalar wave update equations of the main text and thus serve a similar role to the  $\mathbf{W}^{(x)}$  and  $\mathbf{W}^{(y)}$  matrices of the traditional RNN in Eqs. (5.1) and (5.2). However, unlike  $\mathbf{W}^{(x)}$  and  $\mathbf{W}^{(y)}$ , these matrices are fixed and not trainable parameters.

When training the system, we take the wave speed distribution,  $c(x, y, z)$  as design parameters, which are optimized for a given machine learning task, physically corresponding to a patterning of materials within the domain. Thus, when modeled numerically in discrete time (Fig. 5.1E), the wave equation defines an operation which maps into that of an RNN (Fig. 5.1B). Similarly to the RNN, the full time dynamics of the wave equation may be represented as a directed graph (Fig. 5.1F).

## 5.2 Vowel Recognition Demonstration

We now demonstrate how the dynamics of the wave equation can be trained to classify vowels through the construction of an inhomogeneous material distribution. For this task, we utilize a dataset consisting of 930 raw audio recordings of 10 vowel classes from 45 different male speakers and 48 different female speakers [73]. For our learning task, we select a subset of 279 recordings corresponding to three vowel classes, represented by the vowel sounds *ae*, *ei*, and *iy*, as contained in the words *had*, *hayed*, and *heed*, respectively (Fig. 5.2A). Before injection into the system, each vowel waveform is downsampled from its original recording with a 16 kHz sampling rate to a sampling rate of 10 kHz.

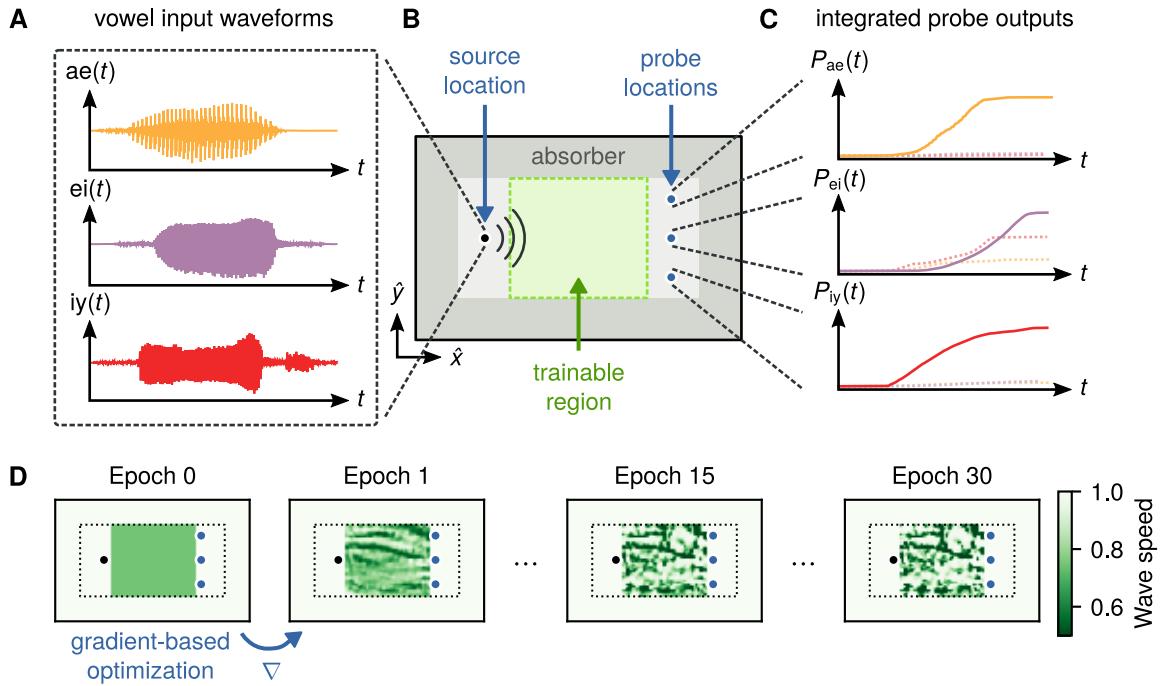


Figure 5.2: **Schematic of the vowel recognition system and the training procedure.** (A) Raw audio waveforms of spoken vowel samples from three classes. (B) Layout of the vowel recognition system. Vowel samples are independently injected at the source, located at the left of the domain, and propagate through the center region, indicated in green, where a material distribution is optimized during training. The dark gray region represents an absorbing boundary layer. (C) For classification, the time-integrated power at each probe is measured and normalized to be interpreted as a probability distribution over the vowel classes. (D) Using automatic differentiation, the gradient of the loss function with respect to the density of material in the green region is computed. The material density is updated iteratively, using gradient-based stochastic optimization techniques, until convergence.

Our goal is to design a system such that when we inject the raw audio corresponding to each vowel (Fig. 5.2A), the power is successfully directed to one of three measuring points. The physical layout of the vowel recognition system consists of a two-dimensional domain in the  $x$ - $y$  plane, infinitely extended along the  $z$ -direction (Fig. 5.2B). The audio waveform of each vowel, represented by  $\mathbf{x}^{(i)}$ , is injected by a source at a single grid cell on the left side of the domain, emitting waveforms which propagate through a central region with a trainable distribution of the wave speed, indicated by the light green region in Fig. 5.2B. An absorbing region is introduced to approximate an open boundary condition [143], corresponding to the grey region in Fig. 5.2B. This region is defined by a dampening coefficient,  $b(x, y)$ , which has a cubic dependence on the distance from the interior boundary of the layer. The presence of absorption prevents energy from building up inside the computational domain. Three probe points are defined on the right hand side of this region, each assigned to one of the three vowel classes. To determine the system's output,  $\mathbf{y}^{(i)}$ , the time-integrated power at each probe is measured (Fig. 5.2C). After the simulation evolves for the full duration of the vowel recording, this integral gives a non-negative vector of length 3, which is then normalized by its sum and interpreted as the system's predicted probability distribution over the vowel classes.

For the purposes of our numerical demonstration, we consider binarized systems consisting of two materials: a background material with a normalized wave speed  $c_0 = 1.0$ , and a second material with  $c_1 = 0.5$ . We assume that the second material has a nonlinear parameter,  $c_{\text{nonlinear}} = -30$ , while the background material has a linear response. In practice, the wave speeds would be modified to correspond to different materials being used. For example, in an acoustic setting the material distribution could consist of air, where the sound speed is 331 m/s, and porous silicone rubber, where the sound speed is 150 m/s [13]. The initial distribution of the wave speed consists of a uniform region of material with a speed which is midway between those of the two materials (Fig. 5.2D).

To create realistic devices with sufficiently large minimum feature sizes and binarized materials, we utilized spatial filtering and thresholding schemes. A material density parameter  $\rho(x, y)$  is defined for the structure, which may vary from 0 to 1. Then, we apply a low pass spatial filter to  $\rho(x, y)$ , which gives a filtered density, labelled  $\tilde{\rho}(x, y)$

$$\tilde{\rho}(x, y) = \begin{bmatrix} 0 & 1/8 & 0 \\ 1/8 & 1/2 & 1/8 \\ 0 & 1/8 & 0 \end{bmatrix} * \rho(x, y). \quad (5.13)$$

For binarization of the resulting structure, a projection scheme is used to recreate the final wave speed from the filtered density. We define  $\bar{\rho}(x, y)$  as the projected density, which is created from  $\tilde{\rho}(x, y)$  as

$$\bar{\rho}_i = \frac{\tanh(\beta\eta) + \tanh(\beta[\tilde{\rho}_i - \eta])}{\tanh(\beta\eta) + \tanh(\beta[1 - \eta])}. \quad (5.14)$$

Here,  $\eta$  is a parameter between 0 and 1 that controls the mid-point of the projection, typically

0.5, and  $\beta$  controls the strength of the projection, typically around 100. The wave speed can be determined from  $\bar{\rho}$  as

$$c(x, y) = (c_1(x, y) - c_0(x, y))\bar{\rho} + c_0(x, y), \quad (5.15)$$

where  $c_0$  and  $c_1$  are the background and optimized material wave speed, respectively.

To train the system, we perform backpropagation through the model of the wave equation to compute the gradient of the cross entropy loss function of the measured outputs with respect to the density of material in each pixel of the trainable region. As we discussed to in the previous Chapter, this process is mathematically equivalent performing the adjoint method. Then, we use this gradient information update the material density using the the Adam optimization algorithm [98], repeating until convergence on a final structure (Fig. 5.2D).

Numerical modeling and simulation of the wave equation physics was performed using a custom software package written in Python [2]. The software package was developed on top of the popular machine learning library, `pytorch`, to compute the gradients of the loss function with respect to the material distribution via reverse-mode automatic differentiation. Using of a machine learning platform to perform the numerical simulation greatly reduces opportunities for errors in the analytic derivation or numerical implementation of the gradient.

All windowed samples from the training set are run through the simulation in batches of 9 and the categorical cross entropy loss between the output probe probability distribution and the correct one-hot vector for each vowel sample is computed. To encourage the optimizer to produce a binarized distribution of the wave speed with relatively large feature sizes, the optimizer minimizes this loss function with respect to a material density distribution,  $\rho(x, y)$  within a central region of the simulation domain, indicated by the green region in Fig. 5.2B. We use the Adam algorithm [99] with learning rate 0.0004 to perform the optimization in batches of 9. Fig. 5.2D illustrates the optimization process over several epochs, during which, the wave velocity distribution converges to a final structure. At the end of each epoch, the classification accuracy is computed over both the testing and training set. Unlike the training set, the full length of each vowel sample from the testing set is used.

The mean energy spectrum of the three vowel classes after downsampling to 10 kHz is shown in Fig. 5.3. We observe that the majority of the energy for all vowel classes is below 1 kHz and that there is strong overlap between the mean peak energy of the *ei* and *iy* vowel classes. Moreover, the mean peak energy of the *ae* vowel class is very close to the peak energy of the other two vowels. Therefore, the vowel recognition task learned by the system in the main text is non-trivial.

To validate the performance of our system, the full dataset of (3 classes)  $\times$  (45 males + 48 females) = 272 vowel samples is divided into 5 groups of approximately equal size. Cross validated training is performed with 4 out of the 5 sample groups forming a training set and 1 out of the 5 sample groups forming a testing set. Independent training runs are performed with each of the 5 groups serving as the testing set, with results being averaged over all training runs. During each

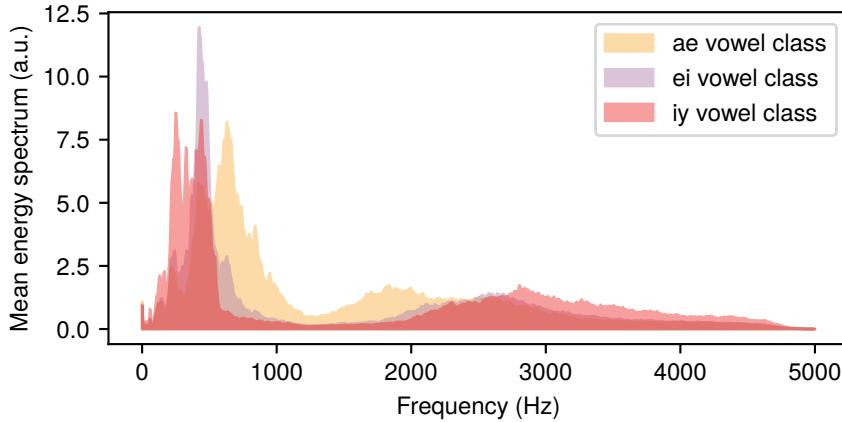
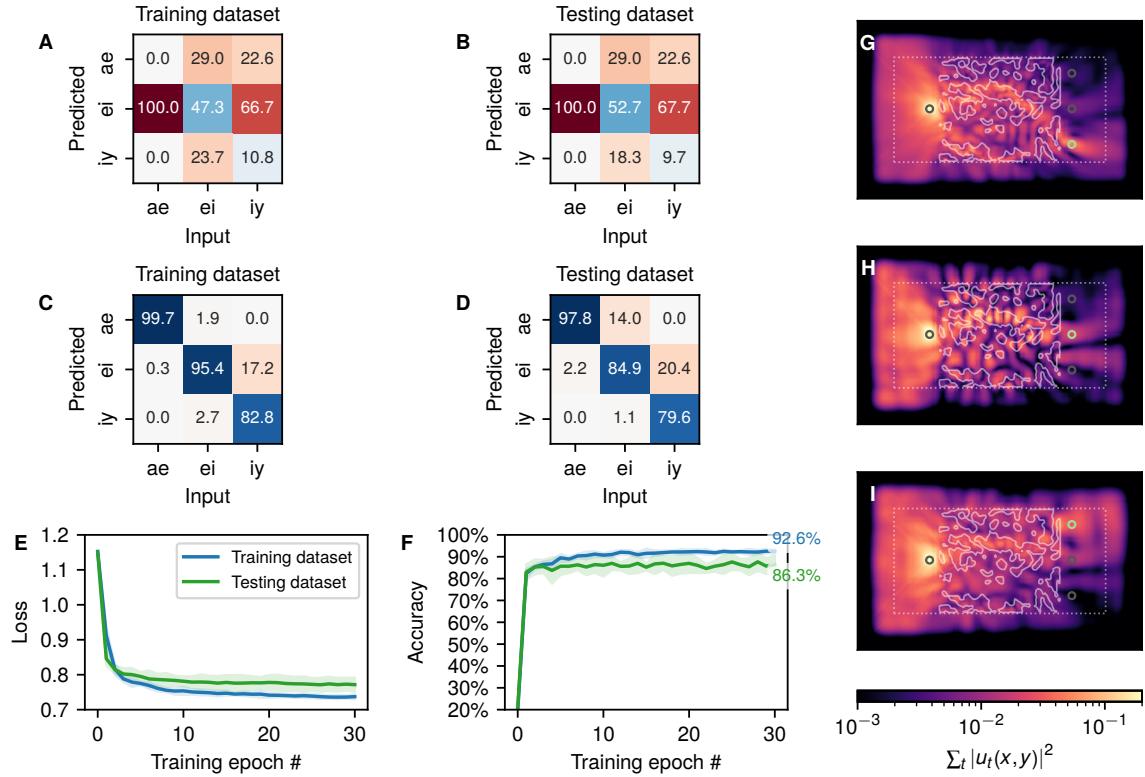


Figure 5.3: Mean energy spectrum for the *ae*, *ei*, and *iy* vowel classes.

epoch, every sample vowel sequence from the training set is windowed to a length of 1000, taken from the center of the sequence, but the testing is performed on the full sequence.

The confusion matrices over the training and testing sets for the starting structure are shown in Fig. 5.4A and Fig. 5.4B, averaged over five cross-validated training runs. Here, the confusion matrix defines the percentage of correctly predicted vowels along its diagonal entries and the percentage of incorrectly predicted vowels for each class in its off-diagonal entries. Clearly the starting structure can not perform the recognition task. Fig. 5.4C and Fig. 5.4D show the final confusion matrices after optimization for the testing and training sets, averaged over five cross validated training runs. The trained confusion matrices are diagonally dominant, indicating that the structure can indeed perform vowel recognition.

Fig. 5.4E and Fig. 5.4F show the cross entropy loss value and the prediction accuracy, respectively, as a function of the training epoch over the testing and training datasets, where the solid line indicates the mean and the shaded region corresponds to the standard deviation over the cross-validated training runs. Interestingly, we observe that the first epoch results in the largest reduction of the loss function and the largest gain in prediction accuracy. From Fig. 5.4F we see that the system obtains a mean accuracy of  $92.6\% \pm 1.1\%$  over the training dataset and a mean accuracy of  $86.3\% \pm 4.3\%$  over the testing dataset. From Fig. 5.4C and Fig. 5.4D we observe that the system attains near perfect prediction performance on the *ae* vowel and is able to differentiate the *iy* vowel from the *ei* vowel, but with less accuracy, especially in unseen samples from the testing dataset. Fig. 5.4G, Fig. 5.4H, and Fig. 5.4I show the distribution of the integrated field intensity,  $\sum_t |u_t(x, y)|^2$ , when a representative sample from each vowel class is injected into the trained structure. We thus provide visual confirmation that the optimization procedure produces a structure which routes the majority of the signal energy to the correct probe. As a performance benchmark, a conventional RNN was trained on the same task, achieving comparable classification accuracy to that of the wave



**Figure 5.4: Vowel recognition system training results.** Confusion matrix over the training and testing datasets for the initial structure (A),(B) and final structure (C),(D), indicating the percentage of correct (diagonal) and incorrect (off-diagonal). Cross validated training results showing the mean (solid line) and standard deviation (shaded region) of the (E) cross entropy loss and (F) prediction accuracy over 20 training epochs and 5 folds of the dataset, which consists of a total of 272 total vowel samples of male and female speakers. (G)-(I) The time-integrated intensity distribution for a randomly selected input (G) *ae* vowel, (H) *ei* vowel, and (I) *iy* vowel.

equation. However, a larger number of free parameters was required. Additionally, we observed that a comparable classification accuracy was obtained when training a linear wave equation.

We compare the wave equation results to a conventional RNN model as defined in Eq. (5.1) and Eq. (5.2). The number of trainable parameters in the model is determined by the hidden state size  $N_h$ , as the model is given by three matrices  $\mathbf{W}^{(x)}$ ,  $\mathbf{W}^{(h)}$ , and  $\mathbf{W}^{(y)}$  of size  $[N_h, 1]$ ,  $[N_h, N_h]$  and  $[3, N_h]$ , respectively. We tried  $N_h = 70$ , for which the total number of RNN free parameters is 5250, and  $N_h = 100$ , with 10500 free parameters. The RNN was implemented and trained using `pytorch`. In Table 5.1 we show the results of a standard RNN on the vowel recognition task and compare them to the scalar wave. We find that the conventional RNN achieves a performance comparable to the wave equation. However, this performance is highly dependent on the number of trainable parameters. For a similar number of trainable parameters, the conventional RNN achieves about 6%

Model	Nonlinearity	# parameters	Accuracy	
			Training	Testing
<b>Wave Equation</b>	linear wave speed	4200	93.1%	86.6%
	nonlinear wave speed	4200	92.6%	86.3%
<b>Conventional RNN</b>	linear	5250	78.8%	79.4%
	leaky ReLU	5250	82.6%	80.2%
	linear	10500	88.9%	88.2%
	leaky ReLU	10500	89.4%	89.4%

Table 5.1: Comparison of scalar wave model and conventional RNN on vowel recognition task.

lower classification accuracy. However, when the number of free parameters is increased to about 2 times that of the scalar wave, the accuracy is higher by about 3 %. We note that it is possible that more advanced recurrent models like long short-term memory (LSTM) [74] or gated recurrent unit (GRU) [38] could have a better performance with a smaller number of parameters, but exploring this is outside the scope of this study.

### 5.3 Comparison of Analog and Digital Implementations

The conventional RNN and the one implemented by a scalar wave equation have many qualitative differences. We discuss some of those below. First, in the RNN case, the trainable parameters are given by the elements of the weight matrices. In the wave equation case, we choose to use the wave velocity,  $c(x, y, z)$ , as trainable parameters, because a specific distribution of  $c$  can be physically implemented after the training process. In acoustic or optical systems, this can be practically realized using technologies such as 3D printing or nanolithography. Furthermore, whereas the RNN free parameters define a matrix which is multiplied by the input, output, and hidden state vectors, in the wave equation case, the free parameters are multiplied element-wise with the hidden state, which limits the influence of each individual parameter over the full dynamics.

For a given amount of expressive power, the size of the hidden state in the wave equation must arguably be much larger than that in the RNN case. This is because the amount of information that can be encoded in the spatial distribution of  $u_t$  is constrained by the diffraction limit for wave systems. It follows that a single RNN hidden state element may be analogous to several grid cells in the scalar wave equation. Furthermore, the wave update matrix  $A$  is sparse and only contains non-zeros on diagonal elements (self coupling) and those corresponding to neighbor-to-neighbor coupling between spatial grid cells. Because of this, information in a given cell of  $u_t$  will take several time steps to reach other cells, as determined by the wave velocity and the distance between them. The presence of this form of causality practically means that one must wait longer for a full ‘mixing’ of information between cells in the domain, suggesting that in the our numerical simulations, a larger

number of time steps may be needed as compared to the typical RNN.

The form of nonlinearity used in the wave equation is different from that in the typical RNN, which involves the application of the nonlinear function,  $\sigma^{(h)}(\cdot)$ , as in Eq. (5.1). In the wave equation, nonlinearity is provided by making the wave velocity,  $c$ , or damping dependent,  $b$ , to be dependent on the instantaneous wave intensity  $|u_t|^2$ . For example  $c = c(|u_t|^2)$ , or  $b = b(|u_t|^2)$ . In optics, these nonlinearities may be implemented using  $\chi^{(3)}$  materials or saturable absorption, respectively. With this addition, the update matrix of Eq. (5.7),  $\mathbf{A} = \mathbf{A}(\mathbf{h}_{t-1})$ , becomes a function of the solution at that time step, making the dynamics nonlinear. Nonlinearity is introduced into the output of the wave system ( $\mathbf{y}_t$ ) by measuring the intensity of the wave field, which involves a squaring operation. One may also consider directly discretizing a nonlinear wave equation using the same technique.

The wave-based RNN presented here has a number of favorable qualities that make it a good candidate for processing sequential data. Unlike the standard RNN, the update of the wave equation from one time step to the next enforces a nearest-neighbor coupling between elements of the hidden state through the Laplacian operator, which is represented by a sparse matrix in Fig. 5.1E. This is a direct consequence of the fact that the wave equation is a hyperbolic partial differential equation in which information propagates with a finite velocity. Thus, the size of the analog RNN’s hidden state, and therefore its memory capacity, is directly determined by the size of the propagation medium. Additionally, unlike the conventional RNN, the wave equation enforces an energy conservation constraint, preventing unbounded growth of the norm of the hidden state and the output signal. In contrast, the unconstrained dense matrices defining the update relationship of the standard RNN lead to vanishing and exploding gradients, which pose a major challenge for training traditional RNNs [91].

In this chapter, we have shown that the dynamics of the wave equation are conceptually equivalent to those of a recurrent neural network. This conceptual connection opens up the opportunity for a new class of analog hardware platform, in which evolving time dynamics play a significant role in both the physics and the dataset. While we have focused on a specific example of the scalar wave equation, our results apply broadly to other wave-like physics. Such an approach of using physics to perform computation [178, 72, 63, 113, 102, 54] may inspire a new platform for analog machine learning devices, with the potential to perform computation far more naturally and efficiently than their digital counterparts. The generality of the approach further suggests that many physical systems may be attractive candidates for performing RNN-like computation on naturally occurring sequences, such as optical, acoustic, or seismic signals.

## Chapter 6

# Extension of Adjoint Method to Nonlinear Systems

Up to now [82], in photonics the adjoint method has been mostly applied to gradient-based optimization of linear optical devices. The generalization of the adjoint method to nonlinear optical devices would create new possibilities in several exciting fields such as on-chip lasers [217], frequency combs [142], spectroscopy [134], neural computing [95], and quantum information processing [64]. To this end, several recent works [114, 115, 23] have applied adjoint methods to engineer *linear* devices to display favorable properties for nonlinear optical applications, such as high quality factors, small mode volume, or large field overlap between the modes of interest. However, these works do not directly optimize the nonlinear systems.

To solve for the adjoint sensitivity of a nonlinear system, the standard option is to work within a time-domain adjoint formalism, which entails simulating an additional linear system with a time-varying permittivity [50]. However, as this formalism requires the storage of the fields at each time step, it has substantial memory requirements. Furthermore, because in many cases the steady-state behavior of the system is of interest, a frequency-domain approach is preferred as the steady state response can be obtained directly, without the need for going through a large number of time steps as in a time-domain simulation. The general mathematical formalism for the adjoint method in nonlinear systems is known in the applied mathematics literature [187]. But, with the exception of a very recent preprint that seeks to design a nonlinear element in an optical neural network [95], such a formalism has not been previously applied to nonlinear photonic device optimizations.

In this Chapter we outline, in detail, how the adjoint method may be used to optimize the steady-state response of a nonlinear optical device in the frequency domain. We first outline a formalism for generalizing adjoint problems to arbitrary nonlinear problems. Then, as a demonstration, we use our method to inverse-design photonic switches with Kerr nonlinearity. Our results may be applied

more generally to other objective functions and sources of nonlinearity and provides new possibilities for designing novel nonlinear optical devices.

## 6.1 Mathematical Formulation

We first outline the formulation of the adjoint method for the inverse design of nonlinear optical devices. The goal of inverse design is to find a set of real-valued design variables  $\varphi$  that maximize a real-valued objective function  $\mathcal{L} = \mathcal{L}(\mathbf{e}, \mathbf{e}^*, \varphi)$ , where the complex-valued vector  $\mathbf{e}$  is given by the solution to the equation

$$\mathbf{f}(\mathbf{e}, \mathbf{e}^*, \varphi) = 0. \quad (6.1)$$

For example, Eq. (6.1) may represent the steady-state Maxwell's equations with an intensity-dependent permittivity distribution where  $\mathbf{e}$  is the electric field distribution. The solution to Eq. (6.1) may be found with any nonlinear equation solver, such as with the Newton-Raphson method [155]. We further note that the treatment of  $\mathbf{e}$  and its complex conjugate as independent variables is necessary for differentiation as will be shown later.

As before, the aim of the optimization is to maximize the objective function with respect to the design variables  $\varphi$ . For this purpose, it is essential to compute the sensitivity of  $\mathcal{L}$  with respect to each element of  $\varphi$ . For simplicity, we derive the derivative of the objective function with respect to a single parameter  $\varphi$ , which is written

$$\frac{d\mathcal{L}}{d\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} + \frac{\partial\mathcal{L}}{\partial\mathbf{e}} \frac{d\mathbf{e}}{d\varphi} + \frac{\partial\mathcal{L}}{\partial\mathbf{e}^*} \frac{d\mathbf{e}^*}{d\varphi}. \quad (6.2)$$

Or, in matrix form as

$$\frac{d\mathcal{L}}{d\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} + \begin{bmatrix} \partial\mathcal{L}/\partial\mathbf{e} & \partial\mathcal{L}/\partial\mathbf{e}^* \end{bmatrix} \begin{bmatrix} d\mathbf{e}/d\varphi \\ d\mathbf{e}^*/d\varphi \end{bmatrix}. \quad (6.3)$$

To compute  $d\mathbf{e}/d\varphi$  and  $d\mathbf{e}^*/d\varphi$ , we differentiate Eq. (6.1):

$$\frac{d\mathbf{f}}{d\varphi} = 0 = \frac{\partial\mathbf{f}}{\partial\varphi} + \frac{\partial\mathbf{f}}{\partial\mathbf{e}} \frac{d\mathbf{e}}{d\varphi} + \frac{\partial\mathbf{f}}{\partial\mathbf{e}^*} \frac{d\mathbf{e}^*}{d\varphi}. \quad (6.4)$$

Eq. (6.4) together with its complex conjugate then yields

$$\begin{bmatrix} \partial\mathbf{f}/\partial\mathbf{e} & \partial\mathbf{f}/\partial\mathbf{e}^* \\ \partial\mathbf{f}^*/\partial\mathbf{e} & \partial\mathbf{f}^*/\partial\mathbf{e}^* \end{bmatrix} \begin{bmatrix} d\mathbf{e}/d\varphi \\ d\mathbf{e}^*/d\varphi \end{bmatrix} = - \begin{bmatrix} \partial\mathbf{f}/\partial\varphi \\ \partial\mathbf{f}^*/\partial\varphi \end{bmatrix}. \quad (6.5)$$

Thus, formally we can rewrite Eq. (6.3) as

$$\frac{d\mathcal{L}}{d\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} - \begin{bmatrix} \partial\mathbf{f}/\partial\mathbf{e} & \partial\mathbf{f}/\partial\mathbf{e}^* \\ \partial\mathbf{f}^*/\partial\mathbf{e} & \partial\mathbf{f}^*/\partial\mathbf{e}^* \end{bmatrix}^{-1} \begin{bmatrix} \partial\mathbf{f}/\partial\varphi \\ \partial\mathbf{f}^*/\partial\varphi \end{bmatrix}. \quad (6.6)$$

In analogy with the linear adjoint method, we can now compute the gradient by solving an additional linear system. We define a complex-valued adjoint field  $\mathbf{e}_{aj}$  as the solution to

$$\begin{bmatrix} \partial\mathbf{f}/\partial\mathbf{e} & \partial\mathbf{f}/\partial\mathbf{e}^* \\ \partial\mathbf{f}^*/\partial\mathbf{e} & \partial\mathbf{f}^*/\partial\mathbf{e}^* \end{bmatrix}^T \begin{bmatrix} \mathbf{e}_{aj} \\ \mathbf{e}_{aj}^* \end{bmatrix} = - \begin{bmatrix} \partial\mathcal{L}/\partial\mathbf{e}^T \\ \partial\mathcal{L}/\partial\mathbf{e}^{*T} \end{bmatrix}, \quad (6.7)$$

and the gradient of the objective function is then

$$\frac{d\mathcal{L}}{d\varphi} = \frac{\partial\mathcal{L}}{\partial\varphi} + 2\mathcal{R}\left(\mathbf{e}_{aj}^T \frac{\partial\mathbf{f}}{\partial\varphi}\right), \quad (6.8)$$

where  $\mathcal{R}$  denotes taking the real part. In deriving Eq. (6.8), we have used the fact that both  $\mathcal{L}$  and  $\varphi$  are real. In the case of multiple parameters  $\boldsymbol{\varphi}$ , we can simply replace  $\partial\mathbf{f}/\partial\varphi$  with the matrix  $\partial\mathbf{f}/\partial\boldsymbol{\varphi}$ . Since  $\mathbf{e}_{aj}$  only needs to be solved once regardless of the number of parameters, gradients may be computed with very little marginal cost for an arbitrary number of free parameters.

We now apply the general formalism as discussed above to the optimization of nonlinear optical systems. Since the formalism is applicable to linear optical systems as well, for illustration purposes here we use it to treat both the linear and the nonlinear cases, in order to highlight aspects that are unique to nonlinear systems. A schematic outlining the two cases is presented in Fig. 6.1. For a linear system, Maxwell's equations for the steady-state at a frequency  $\omega_0$  may be written as

$$\mu_0^{-1} \nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \omega_0^2 \epsilon_0 \epsilon_r(\mathbf{r}) \mathbf{E}(\mathbf{r}) = -i\omega_0 \mathbf{J}(\mathbf{r}), \quad (6.9)$$

where  $\mathbf{E}(\mathbf{r})$  is the electric field,  $\mathbf{J}(\mathbf{r})$  is the electric current source,  $\epsilon_r(\mathbf{r})$  is the relative dielectric permittivity, and we have assumed relative permeability  $\mu_r = 1$  everywhere. Compactly, and to make connection to the general formalism in the previous section, this can be written in matrix form as

$$\mathbf{f}(\mathbf{e}, \mathbf{e}^*, \boldsymbol{\varphi}) = A(\epsilon_r)\mathbf{e} - \mathbf{b} = 0, \quad (6.10)$$

where  $A$  is a linear operator, vectors  $\mathbf{e}$  and  $\epsilon_r$  now contain the electric fields and the relative permittivity, respectively, and  $\mathbf{b}$  is a vector proportional to the current source. The design parameters  $\boldsymbol{\varphi}$  in this case is the permittivity distribution  $\epsilon_r$ . Eq. (6.10) can be solved to obtain the electric fields  $\mathbf{e}$ , as diagrammed by Fig. 6.1(a).

We assume an objective function  $\mathcal{L}$  that depends on the field solution to Eq. (6.10) and we take the linear relative permittivity distribution as the set of design variables. Because  $\partial \mathbf{f} / \partial \mathbf{e} = A$  and  $\partial \mathbf{f} / \partial \mathbf{e}^* = 0$  for the linear system, from Eq. (6.7), the adjoint field may be written simply as the solution to the equation

$$A^T(\epsilon_r) \mathbf{e}_{aj} = -(\partial \mathcal{L} / \partial \mathbf{e})^T, \quad (6.11)$$

as shown in Fig. 6.1(b). For a reciprocal system,  $A^T = A$ , thus the original and the adjoint fields are solutions to the same linear problem but with different source terms. Note that the source for the adjoint field,  $-(\partial \mathcal{L} / \partial \mathbf{e})^T$  depends on both the objective function and the original solution.

Once the adjoint field is computed, the gradient of the objective function with respect to the permittivity distribution is given, through Eq. (6.8), by

$$\frac{d\mathcal{L}}{d\epsilon_r} = \frac{\partial \mathcal{L}}{\partial \epsilon_r} + 2\mathcal{R}\left(\mathbf{e}_{aj}^T \frac{\partial A}{\partial \epsilon_r} \mathbf{e}\right) \quad (6.12)$$

$$= \frac{\partial \mathcal{L}}{\partial \epsilon_r} - 2\omega_0^2 \epsilon_0 \mathcal{R}(\mathbf{e}_{aj}^T \mathbf{e}). \quad (6.13)$$

Having reviewed the adjoint formalism for linear optical systems we now consider nonlinear optical systems. As an example, we introduce Kerr nonlinearity into the system [22], which corresponds to an intensity-dependent permittivity

$$\tilde{\epsilon}_r(\mathbf{r}) = \epsilon_r(\mathbf{r}) + 3\omega_0^2 \epsilon_0 \chi^{(3)}(\mathbf{r}) |\mathbf{E}(\mathbf{r})|^2, \quad (6.14)$$

where  $\chi^{(3)}(\mathbf{r})$  is the nonlinear susceptibility distribution. Other types of nonlinear terms can also be treated with the formalism outlined above. Replacing  $\epsilon_r(\mathbf{r})$  in Eq. (6.9) with  $\tilde{\epsilon}_r(\mathbf{r})$  in Eq. (6.14), our system is then described by the equation:

$$\mathbf{f}(\mathbf{e}, \mathbf{e}^*, \boldsymbol{\varphi}) = A_{nl}(\epsilon_r, \chi, \mathbf{e}) \mathbf{e} - \mathbf{b} = 0, \quad (6.15)$$

where  $A_{nl} \equiv [A(\epsilon_r) - \text{diag}(\chi \odot |\mathbf{e}|^2)]$ . Here,  $\odot$  is element-wise vector multiplication and  $\text{diag}(\mathbf{v})$  represents a diagonal matrix with vector  $\mathbf{v}$  on the main diagonal. The vector  $\chi$  corresponds to the term  $3\omega_0^2 \epsilon_0 \chi^{(3)}(\mathbf{r})$  and  $|\mathbf{e}|^2 \equiv \mathbf{e} \odot \mathbf{e}^*$ . Again, for concreteness, the design parameters  $\boldsymbol{\varphi}$  correspond to the permittivity  $\epsilon_r$ . The solution to this problem is diagrammed in Fig. 6.1(c).

From Eq. (6.7) we may now compute the partial derivatives of  $\mathbf{f}$  with respect to the electric fields  $\mathbf{e}$ , which is needed to construct the adjoint problem.

$$\partial \mathbf{f} / \partial \mathbf{e} = A - 2\text{diag}(\chi \odot |\mathbf{e}|^2) \quad (6.16)$$

$$\partial \mathbf{f}^* / \partial \mathbf{e} = -\text{diag}(\chi \odot \mathbf{e}^* \odot \mathbf{e}^*). \quad (6.17)$$

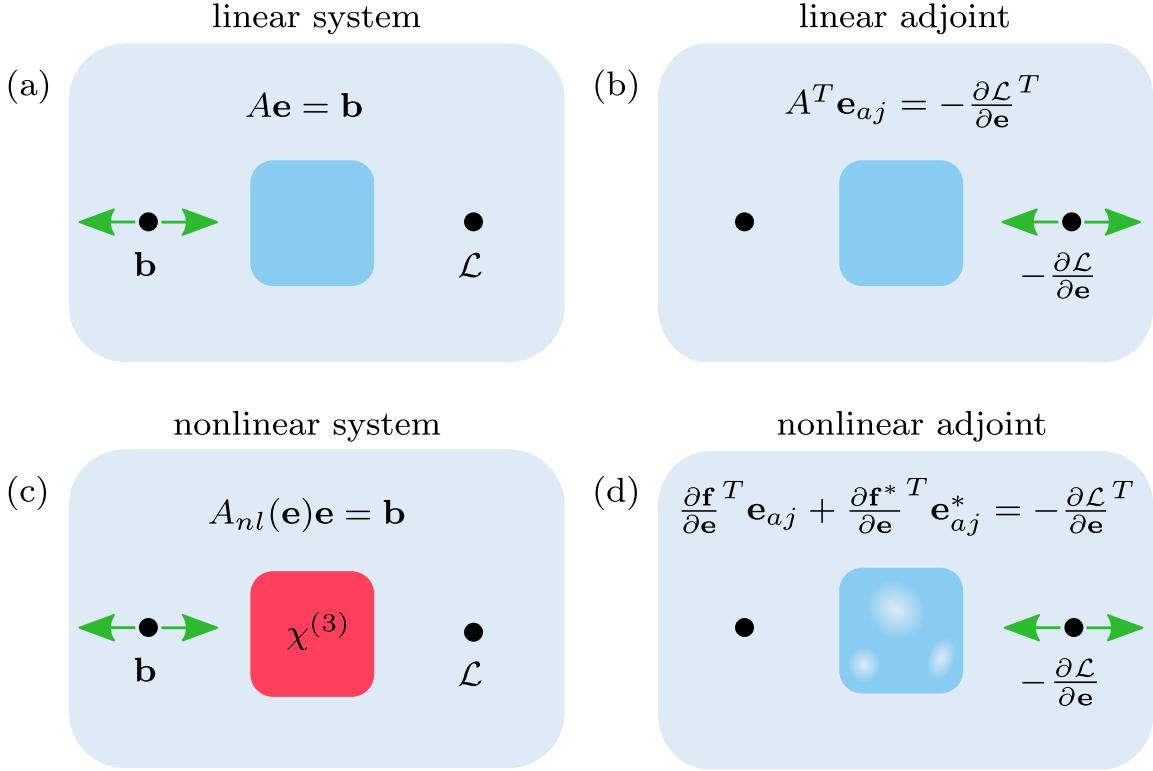


Figure 6.1: Illustration of the adjoint field computation for a linear and a nonlinear system. (a) The linear system driven by a point source  $\mathbf{b}$  with an objective function  $\mathcal{L}$  given by the field intensity at a measuring point. (b) The adjoint problem for the linear system: the same system driven by a point source given by  $-\partial\mathcal{L}/\partial\mathbf{e}$  located at the measuring point. (c) The nonlinear system containing a medium with Kerr nonlinearity (red). The electric fields are the solution to a nonlinear equation. (d) The adjoint problem for the nonlinear system, which is a *linear* system of equations for the adjoint field and its complex conjugate. The Kerr medium is replaced by a linear region whose permittivity depends on the nonlinear fields.

With this, we then express the adjoint field as a solution to the linear system

$$(\partial\mathbf{f}/\partial\mathbf{e})^T \mathbf{e}_{aj} + (\partial\mathbf{f}^*/\partial\mathbf{e})^T \mathbf{e}_{aj}^* = -(\partial\mathcal{L}/\partial\mathbf{e})^T, \quad (6.18)$$

which is diagrammed in Fig. 6.1(d).

For a nonlinear system, to obtain the field  $\mathbf{e}$ , one will need to solve a nonlinear equation (e.g. Eq. (6.15)). However, we emphasize that the adjoint problem, as required to determine the derivative of the objective function, is a *linear* problem. The size of the adjoint problem is twice as large as the corresponding linear problem of Eq. (6.10), but it is of a similar form, with the source dependent upon the solution  $\mathbf{e}$ .

Once the adjoint field is computed, the gradient  $d\mathcal{L}/d\epsilon_r$  is evaluated from Eq. (6.13) as in the

linear case. Here for simplicity we do not assume any explicit dependence of the nonlinearity on the design variable. However, the formalism is straightforward to extend to that case.

For example, assume the nonlinear susceptibility vector  $\chi$  may be written in terms of the scalar magnitude of the nonlinear susceptibility  $\chi^{(3)}$  and the relative permittivity vector  $\epsilon_r$  explicitly as

$$\chi = 3\omega_0^2 \epsilon_0 \chi^{(3)} \frac{\epsilon_r - \mathbf{1}}{\epsilon_m - 1} \quad (6.19)$$

where  $\mathbf{1}$  is a vector of all ones and  $\epsilon_m$  is the maximum relative permittivity allowed in the optimization, corresponding to the material relative permittivity.

As the nonlinear adjoint problem requires the calculation of the partial derivatives  $\partial \mathbf{f} / \partial \mathbf{e}$ ,  $\partial \mathbf{f}^* / \partial \mathbf{e}$ , and  $\partial \mathbf{f} / \partial \epsilon_r$ . When choosing the form of  $\chi$  from Eq. (6.19), the partial derivatives  $\partial \mathbf{f} / \partial \mathbf{e}$  and  $\partial \mathbf{f}^* / \partial \mathbf{e}$  are the same as derived in the main text. However, the term  $\partial \mathbf{f} / \partial \epsilon_r$  takes on a more complicated form given by

$$\frac{\partial \mathbf{f}}{\partial \epsilon_r} = \frac{\partial}{\partial \epsilon_r} \left[ A(\epsilon_r) \mathbf{e} - 3\omega_0^2 \epsilon_0 \chi^{(3)} \frac{\epsilon_r - \mathbf{1}}{\epsilon_m - 1} \odot \mathbf{e} \odot \mathbf{e} \odot \mathbf{e}^* - b \right] \quad (6.20)$$

$$= \frac{\partial A}{\partial \epsilon_r} \mathbf{e} - 3\omega_0^2 \epsilon_0 \chi^{(3)} \frac{1}{\epsilon_m - 1} \text{diag}(\mathbf{e} \odot |\mathbf{e}|^2) \quad (6.21)$$

$$= \text{diag}(\mathbf{e}) - 3\omega_0^2 \epsilon_0 \chi^{(3)} \frac{1}{\epsilon_m - 1} \text{diag}(\mathbf{e} \odot |\mathbf{e}|^2), \quad (6.22)$$

where we make use of the fact that  $\left(\frac{\partial A}{\partial \epsilon_r}\right)_{ijk} = \delta_{ij} \delta_{jk}$ , where  $\delta$  is the Kronecker delta.

Thus, for this case, while the adjoint field  $\mathbf{e}_{aj}$  will have the same form as described earlier, when computing the sensitivity as in Eq. (8), one must insert the form of  $\partial \mathbf{f} / \partial \epsilon_r$  from Eq. (6.22). This is in contrast to the usual case where the nonlinear susceptibility is fixed,  $\partial \mathbf{f} / \partial \epsilon_r = \text{diag}(\mathbf{e})$ .

## 6.2 Inverse Design of Nonlinear Switches

We now demonstrate the use of this nonlinear adjoint formalism to inverse design optical switches with desired power-dependent performance characteristics. In Figs. 6.2 and 6.4, we show the optimization procedures and performance characteristics of a  $1 \rightarrow 1$  and  $1 \rightarrow 2$  port device, respectively. The operating frequency for both devices correspond to a free-space wavelength of  $2\mu\text{m}$ .

For each device, we seek to maximize the corresponding objective function with respect to the permittivity distribution within a fixed design region. To perform the numerical optimization of the structure, we use the finite-difference frequency-domain method (FDFD) [174], where the fields and operators of Eq. (6.9) are spatially discretized using a Yee lattice [221]. For simplicity, we restrict our study to two-dimensional structures (i.e. structures with infinite extent in the third dimension), and transverse-magnetic polarization, which has only non-zero out-of-plane electric field components. In the optimization process, we start with an initial relative permittivity in the design region. We solve

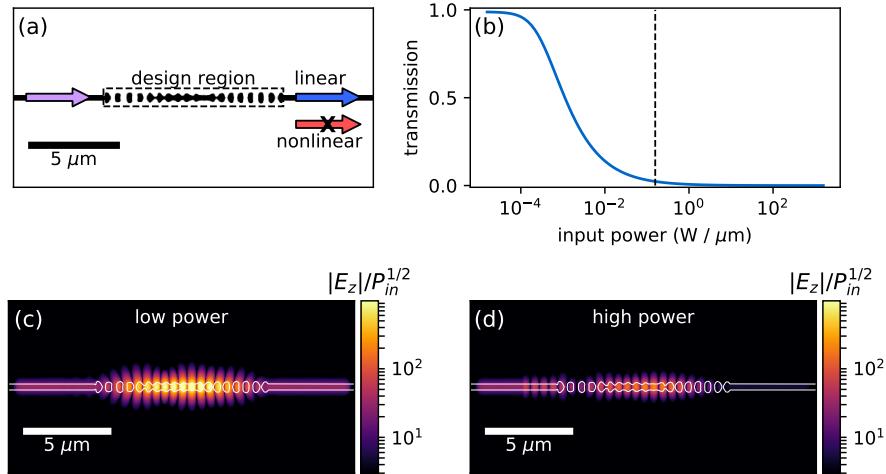


Figure 6.2: **Inverse design demonstration of a  $1 \rightarrow 1$  port switch.** (a) Optical power is input into the left port (purple arrow). The goal of optimizing the design region (blue square) is to maximize power transmission in the linear regime (blue arrow) and minimize transmission in the nonlinear regime (red arrow). The final permittivity distribution after optimizing is also shown. The black regions are chalcogenide with a relative permittivity of 5.95 and a  $\chi^{(3)}$  of  $4.1 \times 10^{-19} \text{ m}^2/\text{V}^2$ . The waveguide regions outside the design region have a width of  $0.3\mu\text{m}$ . The operating wavelength is  $2\mu\text{m}$ . (b) The transmission as a function of input power, demonstrating the switching behavior at around  $10^{-3} \text{ W}/\mu\text{m}$ . The dashed black line indicates the input power used for the high-power regime in the optimization. (c-d) The amplitude of the simulated electric field of the final structure, in the linear (c) and nonlinear (d) regimes, respectively – with input power of  $10^{-9} \text{ W}/\mu\text{m}$  and  $0.157 \text{ W}/\mu\text{m}$ , respectively.  $E_z$  corresponds to the out-of-plane electric field in the 2D simulation.

the electric field distribution in the structure by solving the nonlinear equation (Eq. (6.15)). Then, we compute the gradient of  $\mathcal{L}$  with respect to the relative permittivity distribution in this region using Eq. (6.13). With the gradient information, we perform updates of the design variables using the limited-memory BFGS [29] algorithm, although a simple gradient ascent algorithm would also suffice. This procedure is repeated until convergence on a final structure.

We choose optimization parameters corresponding to a device made from chalcogenide glass ( $\text{As}_2\text{S}_3$ ), which exhibits a strong  $\chi^{(3)}$  response and high damage threshold [210, 106, 22]. During the optimization, the relative permittivity was constrained to lie between 1 (air) and 5.95 ( $\text{As}_2\text{S}_3$ ). We further assume that the materials exhibit nonlinearity only within the design regions outlined in Fig. 6.2(a) and 6.4(a).

To create a more realistic final structure, the strength of the nonlinear susceptibility was assumed to be proportional to the "density" of material,  $\rho$ , defined as

$$\rho(\mathbf{r}) = \frac{\epsilon_r(\mathbf{r}) - 1}{\epsilon_m - 1}, \quad (6.23)$$

where  $\epsilon_m$  is the permittivity of the material. This assumption ensures that air regions do not exhibit a nonlinear refractive index. Eq. (6.23) adds an  $\epsilon_r$  dependence in the nonlinear susceptibility, which is straightforwardly treated in the adjoint method. Low-pass spatial filtering and projection techniques [227] were applied during optimization to create binarized (air and chalcogenide) final structures with large, smoothed features.

To create realistic devices with sufficiently large minimum feature sizes and binarized permittivity distributions, we employed filtering and projection schemes during our optimization. These schemes are discussed in great detail in [227] and related works.

Rather than updating the permittivity distribution directly, one may instead choose to update a design density  $\rho$ , which varies between 0 and 1 within the design region. To create a structure with larger feature sizes, a low pass filter can be applied to  $\rho$  to create a filtered density, labelled  $\tilde{\rho}$ :

$$\tilde{\rho}_i = \frac{\sum_{j \in \mathcal{D}} W_{ij} \rho_j}{\sum_{j \in \mathcal{D}} W_{ij}}, \quad (6.24)$$

where  $\mathcal{D}$  denotes the design region, and  $W$  is the spatial filter, defined for a feature size of  $R$  as

$$W_{ij} = \begin{cases} R - |r_i - r_j| & \text{if } |r_i - r_j| \leq R \\ 0, & \text{otherwise} \end{cases} \quad (6.25)$$

with  $|r_i - r_j|$  being the distance between points  $i$  and  $j$ . This defines a low-pass spatial filter on  $\rho$  with the effect of smoothing out features with length scale below  $R$ .

Now, for binarization of the structure, a projection scheme is used to recreate the final permittivity from the filtered density. For this we define  $\bar{\rho}$  as the projected density, which is created from  $\tilde{\rho}$  as

$$\bar{\rho}_i = \frac{\tanh(\beta\eta) + \tanh(\beta[\tilde{\rho}_i - \eta])}{\tanh(\beta\eta) + \tanh(\beta[1 - \eta])}. \quad (6.26)$$

Here,  $\eta$  is a parameter between 0 and 1 that controls the mid-point of the projection, typically 0.5, and  $\beta$  controls the strength of the projection, typically around 100.

The relative permittivity can then be determined from  $\bar{\rho}$  as

$$\epsilon_r = (\epsilon_m - 1)\bar{\rho} + 1, \quad (6.27)$$

where  $\epsilon_m$  is the maximum permittivity.

The effect of these filtering and the binarization techniques on a sample permittivity set is illustrated in Fig. 6.3. In the optimizations of the main text, these techniques were performed only within the design region and required minimal modifications to the adjoint sensitivity. The determination of  $\partial\epsilon_r/\partial\bar{\rho}$ ,  $\partial\bar{\rho}/\partial\tilde{\rho}$ , and  $\partial\tilde{\rho}/\partial\rho$  were required to compute the derivatives of the objective function with respect to the underlying  $\rho$ . For more details, see the software package accompanying

this work [81].

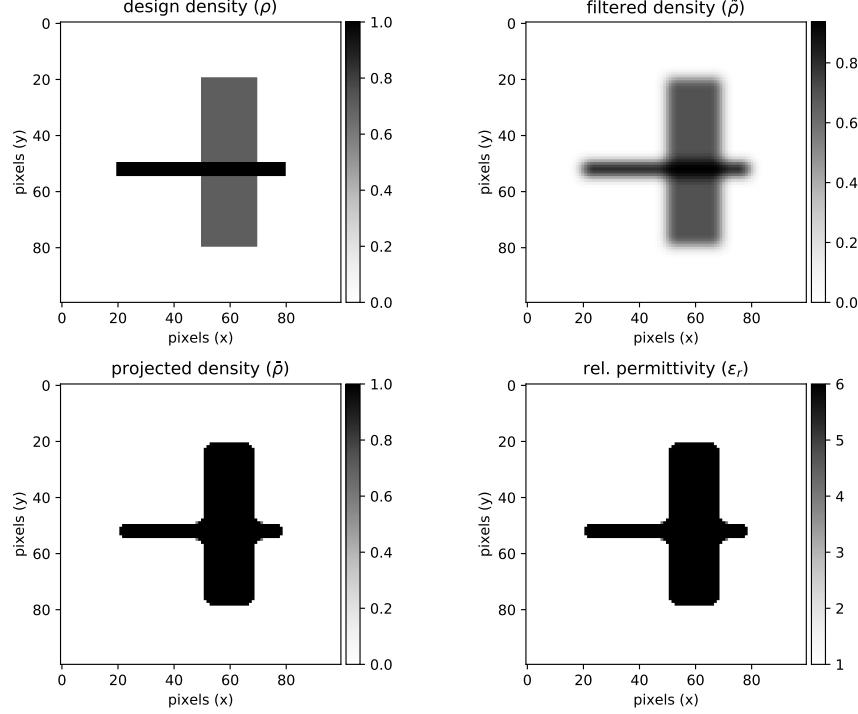


Figure 6.3: Filtering and projection of an example design density,  $\rho$ . (top left) the original density  $\rho$  before processing. (top right) the density after applying a low pass filter,  $\tilde{\rho}$ . (bottom left) the density  $\bar{\rho}$  after applying projection. (bottom right) the final relative permittivity distribution  $\epsilon_r$ . The parameters used are  $R = 200\text{nm}$ ,  $\beta = 100$ ,  $\eta = 0.5$ .

Our first device, as shown in Fig. 6.2, consists of a waveguide-fed  $1 \rightarrow 1$  port geometry with a central design region. We optimize this design region to maximize power transmission in the linear regime when the incident power is sufficiently low such that the nonlinear terms do not affect the transmission, and minimize transmission in the nonlinear regime when the incident power is at a specific high value such that the nonlinearity plays a significant role. This corresponds to an objective function of the form

$$\mathcal{L}(\mathbf{e}_{\text{low}}, \mathbf{e}_{\text{high}}) = |\mathbf{m}^T \mathbf{e}_{\text{low}}|^2 - |\mathbf{m}^T \mathbf{e}_{\text{high}}|^2, \quad (6.28)$$

where  $\mathbf{e}_{\text{low}}$  and  $\mathbf{e}_{\text{high}}$  are the simulated fields with a low and a high input power, respectively,  $\mathbf{m}$  is the modal profile of the electric field for the waveguide in the output port, and the objective function is normalized such that its maximum value is 1. The optimization setup and the optimized structure are diagrammed in Fig. 6.2(a). The final structure resembles a resonator between two Bragg mirrors, effectively acting like a bistable switch [181, 218]. Fig. 6.2(b) shows the transmission as a function

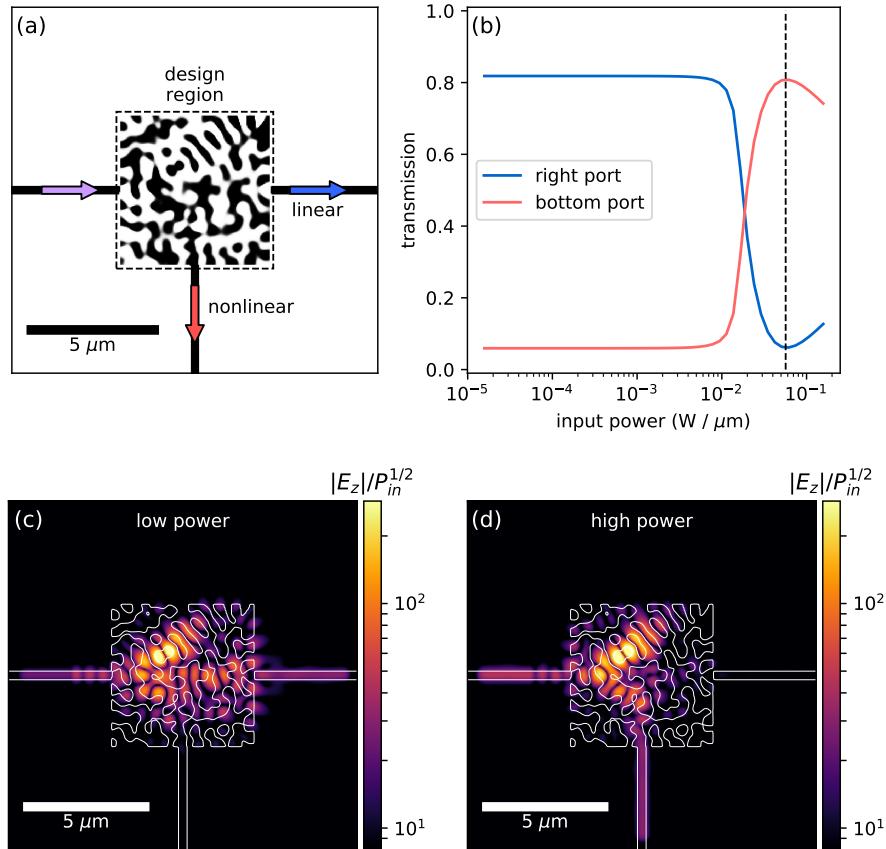


Figure 6.4: **Inverse design demonstration of a  $1 \rightarrow 2$  port switch.** (a) Optical power is input into the left port (purple arrow). The goal of optimizing the design region (blue square) is to maximize the power transmission to the right port (blue arrow) in the linear regime and maximize transmission to the bottom port (red arrow) in the nonlinear regime. The final permittivity distribution after optimizing is also shown. (b) The transmission in the right (blue) and bottom (red) ports as a function of input power, demonstrating the switching behavior at around  $2 \times 10^{-2}$  W/μm. The dashed black line indicates the input power used for the high-power regime in the optimization. (c-d) The amplitude of the simulated electric field of the final structure, in the linear (c) and nonlinear (d) regimes, respectively – with input power of  $10^{-9}$  W/μm and  $0.057$  W/μm, respectively.

of the input power, and it clearly switches from high to low as the input power increases. This is also illustrated in panels (c)-(d), where we plot the field amplitude distributions in the low-power (high-transmission) regime and in the high-power (low-transmission) regime, respectively. The computed power transmission coefficients for these two panels are 98.2% and 3.1%, respectively. The value of the input power used in the optimization and in panel (d) is shown by a dashed line in panel (b). At this input power, the device exhibits a maximum nonlinear refractive index shift of  $4.0 \times 10^{-3}$ ,

which is below the damage threshold for  $\text{As}_2\text{S}_3$  using sub-nanosecond pulses [37].

Here we estimate the maximum nonlinear index shift of chalcogenide ( $\text{As}_2\text{S}_3$ ) materials. Based on [22, 106, 210],  $\text{As}_2\text{S}_3$  has a nonlinear index ( $n_2$ ) between  $3 \times 10^{-14}$  and  $2 \times 10^{-13} \text{ cm}^2/\text{W}$ . From [37], the damage threshold is  $2.5 \text{ J/cm}^2$ . At a pulse duration of 100 ps, this damage threshold corresponds to  $2.5 \times 10^{10} \text{ W/cm}^2$ . Together with the nonlinear refractive index, the maximum refractive index shift sustainable by the material is approximately

$$\Delta n = n_2 I_{\text{damage}} \approx 7.5 \times 10^{-4} - 5 \times 10^{-3} \quad (6.29)$$

with a corresponding pulse bandwidth (for a bandwidth-limited Gaussian pulse) of  $\approx 4.5 \text{ GHz}$ .

Our final structures exhibit maximum refractive index shifts below  $5 \times 10^{-3}$  and their objective functions have FWHM bandwidths above 10 GHz. This suggests that they should exhibit their desired switching effects without damage using pulse durations on the order of 100 ps and input powers on the order of 100 mW/ $\mu\text{m}$ .

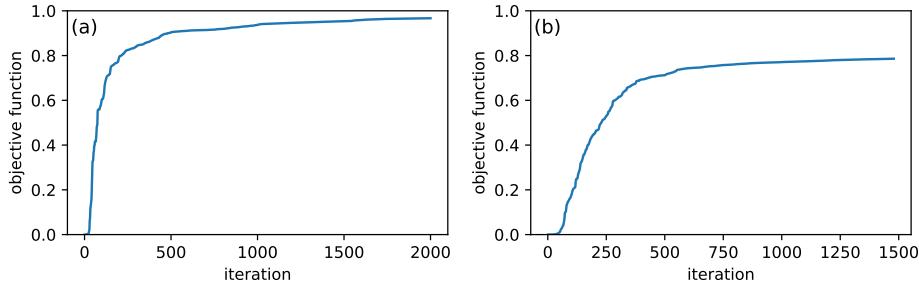


Figure 6.5: Objective function vs. iteration of the optimization for (a) 2-port device of Fig. 2 and (b) 3-port device of Fig. 3 of the main text.

Reaching the final optimized structure shown in Fig. 6.2 required the evaluation of 2000 structures, but a reasonably high-performing structure is already reached after only a few hundred iterations.

We also use the same technique for the inverse design of a  $1 \rightarrow 2$  port switch where light is guided to the right port in the linear regime and to the bottom port in the nonlinear regime. To achieve this design, we define the objective function as

$$\begin{aligned} \mathcal{L}(\mathbf{e}_{\text{low}}, \mathbf{e}_{\text{high}}) = & |\mathbf{m}_r^T \mathbf{e}_{\text{low}}|^2 - |\mathbf{m}_r^T \mathbf{e}_{\text{high}}|^2 \\ & - |\mathbf{m}_b^T \mathbf{e}_{\text{low}}|^2 + |\mathbf{m}_b^T \mathbf{e}_{\text{high}}|^2, \end{aligned} \quad (6.30)$$

where  $\mathbf{m}_r$  and  $\mathbf{m}_b$  denote the mode profiles of the waveguides in the right and bottom ports, respectively. These are normalized such that the objective function has a maximum value of 1 for a perfect switching operation. The setup of the optimization problem and the final design are

diagrammed in Fig. 6.4(a). We note that the device displays a non-intuitive geometry while retaining large features and good binarization.

In Fig. 6.4(b), we plot the transmission through the right and through the bottom ports as a function of input power. This clearly shows the switching of power from the right port to the bottom port in the linear and nonlinear regimes, respectively. Specifically, in the linear regime, the device has a power transmission of 81.8% and 5.9% to the right and bottom ports, respectively, while in the nonlinear regime, at the input power marked by the dashed line in Fig. 6.4(b), these values are 6.1% and 80.8%, respectively. The electric field amplitudes for linear and nonlinear regimes are displayed in 6.4(c)-(d). The operational bandwidth for this device is approximately 90 GHz. The device exhibits a maximum nonlinear refractive index shift of  $3.9 \times 10^{-3}$ , which is also below the acceptable damage threshold for  $\text{As}_2\text{S}_3$  using sub-nanosecond pulses [37].

The transmission vs. frequency for the two devices is shown in Fig. 6.6 in the linear regime.

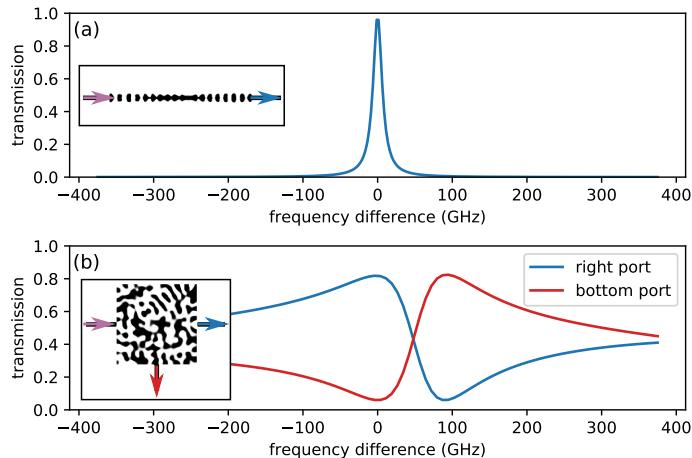


Figure 6.6: (a) Transmission spectrum through the 2-port device for the low power (linear) regime. (b) Transmission spectrum through the right (blue) and bottom (red) ports of the 3-port device in the low power (linear) regime. The x-axis represents the difference in frequency with respect to the design frequency.

Finally, Table 6.1 contains the parameters used in the inverse design demonstrations of Fig. 2 and Fig. 3. The values of the objective function vs. iteration are shown in Fig. 6.5 for both the 2-port and 3-port devices.

### 6.3 Summary and Outlook

We have presented an extension to the adjoint variable method applied to the optimization of an electromagnetic system with Kerr nonlinearity. Our approach can be straightforwardly applied to

parameter	symbol	value (2-port)	value (3-port)	units
max relative permittivity	$\epsilon_m$	5.95	5.95	-
nonlinear susceptibility	$\chi^{(3)}$	$4.1 \times 10^{-19}$	$4.1 \times 10^{-19}$	$\text{m}^2 \text{ V}^{-2}$
input power	$P_0$	157	57	$\text{mW}/\mu\text{m}$
free space wavelength	$\lambda_0$	2	2	$\mu\text{m}$
design region length	$L$	10	6	$\mu\text{m}$
design region height	$H$	1.6	6	$\mu\text{m}$
waveguide width	$w$	300	300	nm
grid size	$g$	40	40	nm
low-pass filter feature size	$R$	160	200	nm
projection strength	$\beta$	100	500	-
projection mid-point	$\eta$	0.5	0.5	-

Table 6.1: Parameters used in the optimization study. Column ‘2-port’ refers to the device from Fig. 2. Column ‘3-port’ refers to the device from Fig. 3

other types of nonlinearities which do not mix frequencies, such as saturable gain or absorption. Moreover, the methods here should be straightforwardly generalizable to treat nonlinear problems involving frequency mixing. For example, one can imagine implementing a similar adjoint method in combination with the multi-frequency finite-difference frequency-domain implementations for nonlinear wave interactions [173].

In addition to the design of optical switches, our formalism may prove useful for many other interesting problems in nonlinear photonics. For example, one could apply our approach to design nonlinear elements in optical neural networks [172] with specific forms of activation functions. Another interesting application is power regulation in the photonic networks described in Chapter 3 for laser-driven particle accelerators [84]. As those systems must be able to handle large input powers, it may be of interest to use our approach to design compact optical limiters in these networks. For the purposes of exploring these and many other potential applications, we have made publicly available a software package that implements the algorithms discussed here [81].

## Chapter 7

# Conclusion and Final Remarks

The adjoint method is an incredibly powerful tool for creating novel integrated photonics devices through inverse design. As we have demonstrated throughout this thesis, the application of this method in photonics may give rise to compact, performant devices in a wide range of applications.

We first showed that the adjoint method could be applied to optimize the performance of a laser-driven particle accelerator. In this study, we developed optical structures that have a factor of 3 improvement over existing structures. Furthermore, these structures had a final design that was incredibly close to those designed independently through human intuition. Another interesting observation was that the same final structures were found no matter the starting condition, which suggests that this problem may be convex. The convexity of different inverse design problems in photonics is an interesting area that may be explored in future works. Finally, we showed that the adjoint current source for this problem corresponded to the current source of an electron traversing the structure. This makes intuitive sense as radiation of charged particles is the reciprocal problem to acceleration.

We then discussed, in great detail, the use of integrated photonic circuits to scale DLA technology to long interaction lengths. This is a promising avenue for performing reconfigurable manipulation and delivery of the laser pulses using precision components made through nanofabrication. This conceptual advance may allow DLA to move from its current stage as a proof-of-principle technology to many exciting practical applications. We discussed the use of a tree-network structure of dielectric waveguides and considered its many opposing constraints and practical considerations. Through a parameter study, we showed that there are optimal operating conditions for such a device, which require a moderate amount of resonance in the accelerator structure to counterbalance the effect of damage at the input couplers. Feedback mechanisms may be necessary to optimize the phase of each stage of this accelerator, which may be implemented directly on chip. We also considered the use of reconfigurable photonics platforms for optimized power delivery to such an accelerator, which are projected to provide substantial improvements to the performance of such a device.

We then discussed optical hardware for machine learning applications. We showed that, for optical neural networks implemented using meshes of Mach Zehnder Interferometers, one may perform training of such a system using what we call *in situ backpropagation*. In this formalism, we may express the gradient calculation of the neural network as an adjoint problem. We further introduced a general method for experimentally measuring the adjoint gradients of any linear optical system through *in situ* intensity measurements. This innovation therefore allows efficient, model-free training of optical neural network hardware.

We showed, more generally, that adjoint method may be used to design optical and acoustic devices for performing machine learning on sequence data. By patterning a wave system using a design obtained through adjoint-based optimization, we showed that one may perform vowel classification through the propagation of raw audio waveforms through a device. This finding opens up the possibility of implementing analog computers for machine learning in wave systems, such as optical, acoustic, or even fluidic systems. Such devices may benefit from higher processing speed, lower latency, and higher energy efficiency when compared to conventional digital electronic platforms.

Finally, we discussed the extension of the adjoint method to nonlinear optical systems. This formalism serves as a more general adjoint framework that may be incredibly useful in designing a new class of nonlinear photonic structures.

In conclusion, the adjoint method is an important conceptual and practical tool for the field of photonics. There are many novel areas that it may benefit, from machine learning hardware to dielectric laser acceleration. The simulation, design, and optimization of optical structures using the adjoint method has proven to be both a fruitful area of research and provides a rich theoretical framework that will continue to be worth exploring for years to come.

## Appendix A

# Backpropagation for Nonholomorphic Activations

In the previous derivation, we have assumed that the functions  $\mathbf{f}_l(\cdot)$  are holomorphic. For each element of input  $\mathbf{Z}_l$ , labeled  $z$ , this means that the derivative of  $\mathbf{f}_l(z)$  with respect to its complex argument is well defined, or the derivative

$$\frac{d\mathbf{f}_l}{dz} = \lim_{\Delta z \rightarrow 0} \frac{\mathbf{f}_l(z + \Delta z) - \mathbf{f}_l(z - \Delta z)}{2\Delta z} \quad (\text{A.1})$$

does not depend on the direction that  $\Delta z$  approaches 0 in the complex plane.

Here we show how to extend the backpropagation derivation to non-holomorphic activation functions. We first examine the starting point of the backpropagation algorithm, considering the change in the mean-squared loss function with respect to the permittivity of a phase shifter in the last layer OIU, as written in Eq. (7) of the main text as

$$\frac{d\mathcal{L}}{d\epsilon_L} = \mathcal{R} \left\{ \mathbf{\Gamma}_L^T \frac{d\mathbf{X}_L}{d\epsilon_L} \right\} \quad (\text{A.2})$$

Where we had defined the error vector  $\mathbf{\Gamma}_L \equiv (\mathbf{X}_L - \mathbf{T})^*$  for simplicity and  $\mathbf{X}_L = \mathbf{f}_L(\hat{W}_L \mathbf{X}_{L-1})$  is the output of the final layer.

To evaluate this expression for non-holomorphic activation functions, we split  $\mathbf{f}_L(\mathbf{Z})$  and its argument into their real and imaginary parts

$$\mathbf{f}_L(\mathbf{Z}) = \mathbf{u}(\boldsymbol{\alpha}, \boldsymbol{\beta}) + i\mathbf{v}(\boldsymbol{\alpha}, \boldsymbol{\beta}), \quad (\text{A.3})$$

where  $i$  is the imaginary unit and  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are the real and imaginary parts of  $\mathbf{Z}_L$ , respectively.

We now wish to evaluate  $\frac{d\mathbf{X}_L}{d\epsilon_L} = \frac{d\mathbf{f}_L(\mathbf{Z})}{d\epsilon_L}$ , which gives the following via the chain rule

$$\frac{d\mathbf{f}}{d\epsilon} = \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} \odot \frac{d\boldsymbol{\alpha}}{d\epsilon} + \frac{d\mathbf{u}}{d\boldsymbol{\beta}} \odot \frac{d\boldsymbol{\beta}}{d\epsilon} + i \frac{d\mathbf{v}}{d\boldsymbol{\alpha}} \odot \frac{d\boldsymbol{\alpha}}{d\epsilon} + i \frac{d\mathbf{v}}{d\boldsymbol{\beta}} \odot \frac{d\boldsymbol{\beta}}{d\epsilon}, \quad (\text{A.4})$$

where we have dropped the layer index for simplicity. Here, terms of the form  $\frac{d\mathbf{x}}{d\mathbf{y}}$  correspond to element-wise differentiation of the vector  $\mathbf{x}$  with respect to the vector  $\mathbf{y}$ . For example, the  $i$ -th element of the vector  $\frac{d\mathbf{x}}{d\mathbf{y}}$  is given by  $\frac{dx_i}{dy_i}$ .

Now, inserting into Eq. (A.2), we have

$$\frac{d\mathcal{L}}{d\epsilon_L} = \mathcal{R} \left\{ \mathbf{\Gamma}_L^T \odot \left( \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} + i \frac{d\mathbf{v}}{d\boldsymbol{\alpha}} \right)^T \frac{d\boldsymbol{\alpha}}{d\epsilon_L} \right. \quad (\text{A.5})$$

$$\left. + \mathbf{\Gamma}_L^T \odot \left( \frac{d\mathbf{u}}{d\boldsymbol{\beta}} + i \frac{d\mathbf{v}}{d\boldsymbol{\beta}} \right)^T \frac{d\boldsymbol{\beta}}{d\epsilon_L} \right\}. \quad (\text{A.6})$$

We now define real and imaginary parts of  $\mathbf{\Gamma}_L$  as  $\mathbf{\Gamma}_R$  and  $\mathbf{\Gamma}_I$ , respectively. Inserting the definitions of  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  in terms of  $\hat{W}_L$  and  $\mathbf{X}_{L-1}$  and doing some algebra, we recover

$$\frac{d\mathcal{L}}{d\epsilon_L} = \mathcal{R} \left\{ \left( \mathbf{\Gamma}_R \odot \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} \right)^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right. \quad (\text{A.7})$$

$$\left. - \left( \mathbf{\Gamma}_I \odot \frac{d\mathbf{v}}{d\boldsymbol{\alpha}} \right)^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right\} \quad (\text{A.8})$$

$$-i \left( \mathbf{\Gamma}_R \odot \frac{d\mathbf{u}}{d\boldsymbol{\beta}} \right)^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \quad (\text{A.9})$$

$$+i \left( \mathbf{\Gamma}_I \odot \frac{d\mathbf{v}}{d\boldsymbol{\beta}} \right)^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \Big\}. \quad (\text{A.10})$$

Finally, the expression simplifies to

$$\frac{d\mathcal{L}}{d\epsilon_L} = \mathcal{R} \left\{ \left[ \mathbf{\Gamma}_R \odot \left( \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} - i \frac{d\mathbf{u}}{d\boldsymbol{\beta}} \right) \right. \right. \quad (\text{A.11})$$

$$\left. \left. + \mathbf{\Gamma}_I \odot \left( -\frac{d\mathbf{v}}{d\boldsymbol{\alpha}} + i \frac{d\mathbf{v}}{d\boldsymbol{\beta}} \right) \right]^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right\}. \quad (\text{A.12})$$

As a check, if we insert the conditions for  $\mathbf{f}_L(\mathbf{Z})$  to be holomorphic, namely

$$\frac{d\mathbf{u}}{d\boldsymbol{\alpha}} = \frac{d\mathbf{v}}{d\boldsymbol{\beta}}, \quad \text{and} \quad \frac{d\mathbf{u}}{d\boldsymbol{\beta}} = -\frac{d\mathbf{v}}{d\boldsymbol{\alpha}}, \quad (\text{A.13})$$

Eq. (A.10) simplifies to

$$\frac{d\mathcal{L}}{d\epsilon_L} = \mathcal{R} \left\{ \left[ \boldsymbol{\Gamma}_R \odot \left( \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} + i \frac{d\mathbf{v}}{d\boldsymbol{\alpha}} \right) + \right. \right. \quad (A.14)$$

$$\left. \left. \boldsymbol{\Gamma}_I \odot \left( -\frac{d\mathbf{v}}{d\boldsymbol{\alpha}} + i \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} \right) \right]^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right\} \quad (A.15)$$

$$= \mathcal{R} \left\{ \left[ \boldsymbol{\Gamma}_L \odot \left( \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} + i \frac{d\mathbf{v}}{d\boldsymbol{\alpha}} \right) \right]^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right\} \quad (A.16)$$

$$= \mathcal{R} \left\{ \left[ \boldsymbol{\Gamma}_L \odot \mathbf{f}_l'(\mathbf{Z}_L) \right]^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right\} \quad (A.17)$$

$$= \mathcal{R} \left\{ \boldsymbol{\delta}_L^T \frac{d\hat{W}_L}{d\epsilon_L} \mathbf{X}_{L-1} \right\} \quad (A.18)$$

as before.

This derivation may be similarly extended to any layer  $l$  in the network. For holomorphic activation functions, whereas we originally defined the  $\boldsymbol{\delta}$  vectors as

$$\boldsymbol{\delta}_l = \boldsymbol{\Gamma}_l \odot \mathbf{f}_l'(\mathbf{Z}_l), \quad (A.19)$$

for non-holomorphic activation functions, the respective definition is

$$\boldsymbol{\delta}_l = \boldsymbol{\Gamma}_R \odot \left( \frac{d\mathbf{u}}{d\boldsymbol{\alpha}} - i \frac{d\mathbf{u}}{d\boldsymbol{\beta}} \right) + \boldsymbol{\Gamma}_I \odot \left( -\frac{d\mathbf{v}}{d\boldsymbol{\alpha}} + i \frac{d\mathbf{v}}{d\boldsymbol{\beta}} \right), \quad (A.20)$$

where  $\boldsymbol{\Gamma}_R$  and  $\boldsymbol{\Gamma}_I$  are the respective real and imaginary parts of  $\boldsymbol{\Gamma}_l$ ,  $\mathbf{u}$  and  $\mathbf{v}$  are the real and imaginary parts of  $\mathbf{f}_l(\cdot)$ , and  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  are the real and imaginary parts of  $\mathbf{Z}_l$ , respectively.

We can write this more simply as

$$\boldsymbol{\delta}_l = \mathcal{R} \left\{ \boldsymbol{\Gamma}_l \odot \frac{d\mathbf{f}}{d\boldsymbol{\alpha}} \right\} - i \mathcal{R} \left\{ \boldsymbol{\Gamma}_l \odot \frac{d\mathbf{f}}{d\boldsymbol{\beta}} \right\}. \quad (A.21)$$

In polar coordinates where  $\mathbf{Z} = \mathbf{r} \exp(i\phi)$  and  $\mathbf{f} = \mathbf{f}(\mathbf{r}, \phi)$ , this equation becomes

$$\boldsymbol{\delta}_l = \exp(-i\phi) \left( \mathcal{R} \left\{ \boldsymbol{\Gamma}_l \odot \frac{d\mathbf{f}}{d\mathbf{r}} \right\} - i \mathcal{R} \left\{ \boldsymbol{\Gamma}_l \odot \frac{1}{\mathbf{r}} \frac{d\mathbf{f}}{d\phi} \right\} \right) \quad (A.22)$$

where all operations are element-wise.

## Appendix B

# Simulation of Photonic Neural Networks

In Sections 4 and 5 of the main text, we have shown, starting from Maxwell's equations, how the gradient information defined for an arbitrary problem can be obtained through electric field intensity measurements. However, since the full electromagnetic problem is too large to solve repeatedly, for the purposes of demonstration of a functioning neural network, in Section 6 we use the analytic, matrix representation of a mesh of MZIs as described in Ref. [39]. Namely, for an even  $N$ , the matrix  $\hat{W}$  of the OIU is parametrized as the product of  $N + 1$  unitary matrices:

$$\hat{W} = \hat{R}_N \hat{R}_{N-1} \dots \hat{R}_2 \hat{R}_1 \hat{D}, \quad (\text{B.1})$$

where each  $\hat{R}_i$  implements a number of two-by-two unitary operations corresponding to a given MZI, and  $\hat{D}$  is a diagonal matrix corresponding to an arbitrary phase delay added to each port. This is schematically illustrated in Fig. B.1(a) for  $N = 3$ . For the ANN training, we need to compute terms of the form

$$\frac{d\mathcal{L}}{d\phi} = \mathcal{R} \left\{ \mathbf{Y}^T \frac{d\hat{W}}{d\phi} \mathbf{X} \right\}, \quad (\text{B.2})$$

for an arbitrary phase  $\phi$  and vectors  $\mathbf{X}$  and  $\mathbf{Y}$  defined following the steps in the main text. Because of the feed-forward nature of the OIU-s, the matrix  $\hat{W}$  can also be split as

$$\hat{W} = \hat{W}_2 \hat{F}_\phi \hat{W}_1, \quad (\text{B.3})$$

where  $\hat{F}_\phi$  is a diagonal matrix which applies a phase shift  $e^{i\phi}$  in port  $i$  (the other elements are independent of  $\phi$ ), while  $\hat{W}_1$  and  $\hat{W}_2$  are the parts that precede and follow the phase shifter, respectively

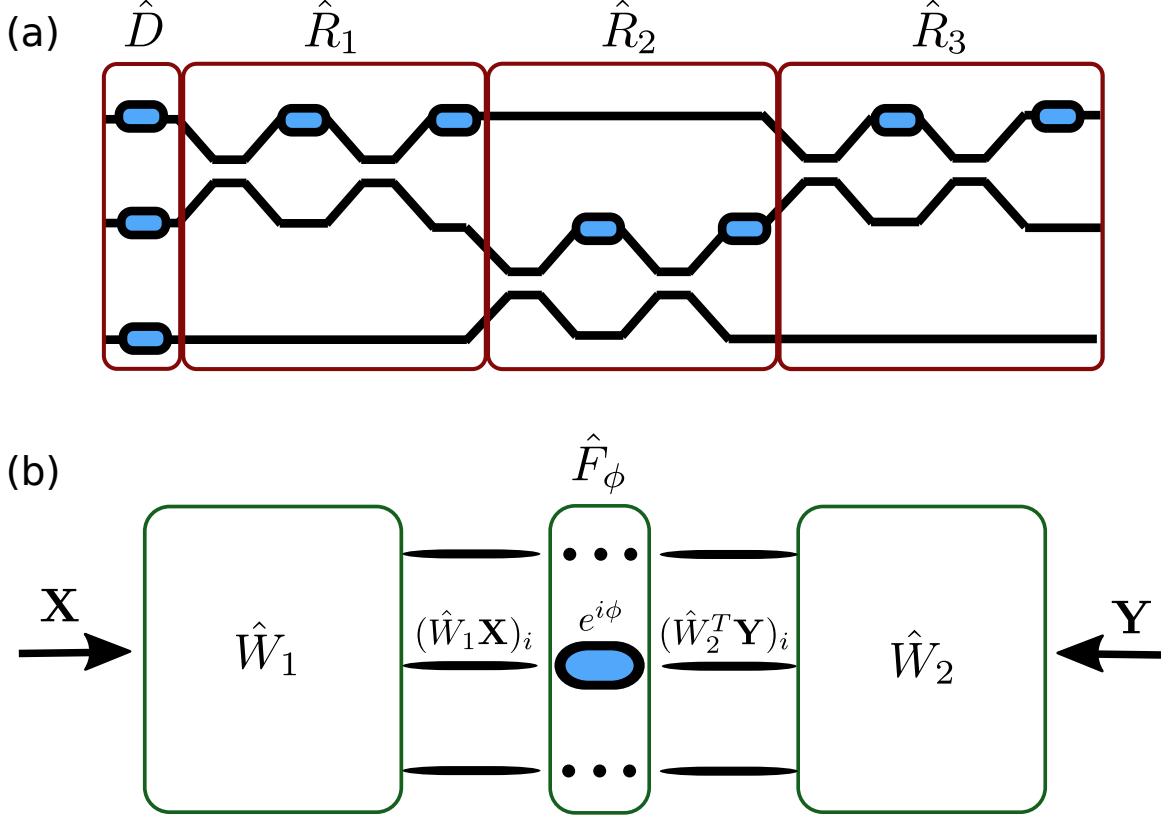


Figure B.1: (a): An OIU implementing a universal  $3 \times 3$  unitary operation, parameterized as in Ref. [39]. (b): Illustration of Eq. (B.4) for the computation of the gradient with respect to a given phase shifter.

(Fig. B.1(b)). Thus, Eq. (B.2) becomes

$$\begin{aligned} \mathcal{R} \left\{ \mathbf{Y}^T \frac{d\hat{W}}{d\phi} \mathbf{X} \right\} &= \mathcal{R} \left\{ \mathbf{Y}^T \hat{W}_2 \frac{d\hat{F}_\phi}{d\phi} \hat{W}_1 \mathbf{X} \right\} \\ &= -\mathcal{I} \left\{ (\hat{W}_2^T \mathbf{Y})_i e^{i\phi} (\hat{W}_1 \mathbf{X})_i \right\}, \end{aligned} \quad (\text{B.4})$$

where  $(\mathbf{V})_i$  is the  $i$ -th element of the vector  $\mathbf{V}$ , and  $\mathcal{I}$  denotes the imaginary part. This result can be written more intuitively in a notation similar to the main text. Namely, if  $A_\phi$  is the field amplitude generated by input  $\mathbf{X}$  from the right, measured right after the phase shifter corresponding to  $\phi$ , while  $A_\phi^{\text{adj}}$  is the field amplitude generated by input  $\mathbf{Y}$  from the right, measured at the same point, then

$$\frac{d\mathcal{L}}{d\phi} = -\mathcal{I} \left\{ A_\phi A_\phi^{\text{adj}} \right\} \quad (\text{B.5})$$

By recording the amplitudes in all ports during the forward and the backward propagation, we can thus compute in parallel the gradient with respect to every phase shifter. Notice that, within this computational model, we do not need to go through the full procedure outlined in Section 4 of the main text. However, this procedure is crucial for the *in situ* measurement of the gradients, and works even in cases which cannot be correctly captured by the simplified matrix model used here.

# Bibliography

- [1] Particle swarm optimization: developments, applications and resources - IEEE Conference Publication.
- [2] `wavetorch`: A Python framework for simulating and back propagating through the wave equation. <https://github.com/fancompute/wavetorch>.
- [3] Y. S. Abu-Mostafa and D. Psaltis. Optical neural computers. *Scientific American*, 256:88–95, 1987.
- [4] Govind P Agrawal. *Nonlinear fiber optics*. Academic press, 2007.
- [5] Govind P. Agrawal. *Fiber-Optic Communication Systems*. John Wiley & Sons, February 2012.
- [6] Selcuk Akturk et al. Pulse-front tilt caused by spatial and temporal chirp. *Opt. Express*, 12(19):4399–4410, 2004.
- [7] Fabien Alibart, Elham Zamanidoost, and Dmitri B Strukov. Pattern classification by memristive crossbar circuits using ex situ and in situ training. *Nature Communications*, 4:2072, 2013.
- [8] Andrea Annoni, Emanuele Guglielmi, Marco Carminati, Giorgio Ferrari, Marco Sampietro, David AB Miller, Andrea Melloni, and Francesco Morichetti. Unscrambling light - automatically undoing strong mixing between modes. *Light: Science & Applications*, 6(12):e17110, December 2017.
- [9] Andrea Annoni, Emanuele Guglielmi, Marco Carminati, Giorgio Ferrari, Marco Sampietro, David AB Miller, Andrea Melloni, and Francesco Morichetti. Unscrambling light—automatically undoing strong mixing between modes. *Light: Science & Applications*, 6(12):e17110, 2017.
- [10] Alán Aspuru-Guzik and Philip Walther. Photonic quantum simulators. *Nature Physics*, 8(4):285–291, April 2012.

- [11] AK Atieh, P Myslinski, J Chrostowski, and P Galko. Measuring the raman time constant ( $t_{\text{sub } r}$ ) for soliton pulses in standard single-mode fiber. *Journal of lightwave technology*, 17(2):216–221, 1999.
- [12] Mordecai Avriel. *Nonlinear programming: analysis and methods*. Courier Corporation, 2003.
- [13] Abdoulaye Ba, Artem Kovalenko, Christophe Aristégui, Olivier Mondain-Monval, and Thomas Brunet. Soft porous silicone rubbers with ultra-low sound speeds in acoustic metamaterials. *Scientific Reports*, 7:40106, January 2017.
- [14] J Bae et al. Experimental verification of the theory on the inverse Smith–Purcell effect at a submillimeter wavelength. *Appl. Phys. Lett.*, 61(7):870–872, 1992.
- [15] K. Bane and G. Stupakov. Using surface impedance for calculating wakefields in flat geometry. *Phys. Rev. ST-AB*, 18:034401, 2015.
- [16] 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, March 2011.
- [17] Doron Bar-Lev, R Joel England, Kent P Wootton, Weihao Liu, Avraham Gover, Robert Byer, Ken J Leedle, D Black, and Jacob Scheuer. Design of a plasmonic metasurface laser accelerator with a tapered phase velocity for subrelativistic particles. *Physical Review Accelerators and Beams*, 22(2):021303, 2019.
- [18] Michael Belt, Michael L Davenport, John E Bowers, and Daniel J Blumenthal. Ultra-low-loss ta 2 o 5-core/sio 2-clad planar waveguides on si substrates. *Optica*, 4(5):532–536, 2017.
- [19] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, September 2017.
- [20] P. I. Borel, A. Harpøth, L. H. Frandsen, M. Kristensen, P. Shi, J. S. Jensen, and O. Sig-mund. Topology optimization and fabrication of photonic crystal structures. *Optics Express*, 12(9):1996–2001, May 2004.
- [21] Robert W. Boyd. *Nonlinear Optics*. Academic Press, May 2008.
- [22] Robert W. Boyd. *Nonlinear Optics*. Academic Press, May 2008.
- [23] Jorge Bravo-Abad, Alejandro Rodriguez, Peter Bermel, Steven G. Johnson, John D. Joannopoulos, and Marin Soljačić. Enhanced nonlinear optics in photonic-crystal microcavities. *Optics Express*, 15(24):16161–16176, November 2007.

- [24] John Breuer et al. Dielectric laser acceleration of electrons in the vicinity of single and double grating structures-theory and simulations. *J. Phys. B: At. Mol. Opt. Phys.*, 47(23):234004, 2014.
- [25] John Breuer et al. Dielectric laser acceleration of nonrelativistic electrons at a single fused silica grating structure: Experimental part. *Phys. Rev. ST Accel. Beams*, 17(2):021301, 2014.
- [26] John Breuer and Peter Hommelhoff. Laser-based acceleration of nonrelativistic electrons at a dielectric structure. *Phys. Rev. Lett.*, 111(13):134803, 2013.
- [27] Peter F Brown, Vincent J Della Pietra, Stephen A Della Pietra, and Robert L Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational linguistics*, 19(2):263–311, 1993.
- [28] Daniel Brunner, Miguel C. Soriano, Claudio R. Mirasso, and Ingo Fischer. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nature Communications*, 4:1364–1367, 2013.
- [29] Richard H Byrd, Peihuang Lu, Jorge Nocedal, and Ciyou Zhu. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16(5):1190–1208, 1995.
- [30] D. E. Carlson and C. R. Wronski. Amorphous silicon solar cell. *Applied Physics Letters*, 28(11):671–673, June 1976.
- [31] Jacques Carolan, Christopher Harrold, Chris Sparrow, Enrique Martín-López, Nicholas J. Russell, Joshua W. Silverstone, Peter J. Shadbolt, Nobuyuki Matsuda, Manabu Oguma, Miki-taka Itoh, Graham D. Marshall, Mark G. Thompson, Jonathan C.F. Matthews, Toshikazu Hashimoto, Jeremy L. O’Brien, and Anthony Laing. Universal linear optics. *Science*, 349(6249):711–716, 2015.
- [32] D. Cesar, S. Custodio, J. Maxson, P. Musumeci, X. Shen, E. Threlkeld, R. J. England, A. Hanuka, I. V. Makasyuk, E. A. Peralta, K. P. Wootton, and Z. Wu. Nonlinear response in high-field dielectric laser accelerators. 2017.
- [33] D Cesar, J Maxson, P Musumeci, X Shen, RJ England, and KP Wootton. Optical design for increased interaction length in a high gradient dielectric laser accelerator. *arXiv preprint arXiv:1801.01115*, 2018.
- [34] D Cesar, J Maxson, P Musumeci, X Shen, RJ England, and KP Wootton. Optical design for increased interaction length in a high gradient dielectric laser accelerator. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 909:252–256, 2018.

- [35] Chia-Ming Chang and Olav Solgaard. Silicon buried gratings for dielectric laser electron accelerators. *Appl. Phys. Lett.*, 104(18):184102, 2014.
- [36] Chia Ming Chang and Olav Solgaard. Silicon buried gratings for dielectric laser electron accelerators. *Applied Physics Letters*, 104(18):184102, May 2014.
- [37] Marine Choré, Thomas Lantertier, Eric Lavastre, Nicolas Bonod, Bruno Bousquet, and Jérôme Néauport. Robust optimization of the laser induced damage threshold of dielectric mirrors for high power lasers. *Optics express*, 26(9):11764–11774, 2018.
- [38] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv:1412.3555*, 2014.
- [39] William R. Clements, Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley. Optimal design for universal multiport interferometers. *Optica*, 3(12):1460, dec 2016.
- [40] William R. Clements, Peter C. Humphreys, Benjamin J. Metcalf, W. Steven Kolthammer, and Ian A. Walmsley. Optimal design for universal multiport interferometers. *Optica*, 3(12):1460, December 2016.
- [41] J. T. Connor, R. D. Martin, and L. E. Atlas. Recurrent neural networks and robust time series prediction. *IEEE Transactions on Neural Networks*, 5(2):240–254, March 1994.
- [42] E. D. Courant et al. High-energy inverse free-electron laser accelerator. *AIP Conf. Proc.*, 127(1):849–874, 1985.
- [43] Benjamin M Cowan. Three-dimensional dielectric photonic crystal structures for laser-driven acceleration. *Phys. Rev. ST Accel. and Beams*, 11(1):011301, 2008.
- [44] Jay W Dawson et al. Analysis of the scalability of diffraction-limited fiber lasers and amplifiers to high average power. *Opt. Express*, 16(17):13240–13266, 2008.
- [45] Huiyang Deng, Jiaqi Jiang, Yu Miao, Kenneth J Leedle, Hongquan Li, Olav Solgaard, Robert L Byer, and James S Harris. Design of racetrack ring resonator based dielectric laser accelerators. 2017.
- [46] Yunhong Ding, Haiyan Ou, and Christophe Peucheret. Ultrahigh-efficiency apodized grating coupler using fully etched photonic crystals. *Opt. Lett.*, 38(15):2732–2734, 2013.
- [47] Georg Dorffner. Neural Networks for Time Series Processing. *Neural Network World*, 6:447–468, 1996.

- [48] Russell Eberhart and James Kennedy. A new optimizer using particle swarm theory. In *Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on*, pages 39–43. IEEE, 1995.
- [49] J. M. Eggleston, T. J. Kan, K. Kuhn, J. Unternahrer, and R. L. Byer. The slab geometry laser - part i: Theory. *IEEE Journal of Quantum Electronics*, QE-20:289–301, 1984.
- [50] Y. Elesin, B. S. Lazarov, J. S. Jensen, and O. Sigmund. Design of robust and efficient photonic switches using topology optimization. *Photonics and Nanostructures - Fundamentals and Applications*, 10(1):153–165, January 2012.
- [51] Jeffrey L Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.
- [52] William C. Elmore and Mark A. Heald. *Physics of Waves*. Courier Corporation, April 2012.
- [53] R Joel England et al. Dielectric laser accelerators. *Rev. Mod. Phys.*, 86(4):1337, 2014.
- [54] Nasim Mohammadi Estakhri, Brian Edwards, and Nader Engheta. Inverse-designed metasstructures that solve equations. *Science*, 363(6433):1333–1338, March 2019.
- [55] 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, 2017.
- [56] Arman B Fallahkhair, Kai S Li, and Thomas E Murphy. Vector finite difference modesolver for anisotropic dielectric waveguides. *Journal of Lightwave Technology*, 26(11):1423–1431, 2008.
- [57] Olivier Faugeras and OLIVIER AUTOR FAUGERAS. *Three-dimensional computer vision: a geometric viewpoint*. MIT press, 1993.
- [58] J R Fontana and R H Pantell. A high-energy, laser accelerator for electrons using the inverse Cherenkov effect. *J. Appl. Phys.*, 54(8):4285–4288, 1983.
- [59] N.K. Georgieva, S. Glavic, M.H. Bakr, and J.W. Bandler. Feasible adjoint sensitivity technique for EM design optimization. *IEEE Transactions on Microwave Theory and Techniques*, 50(12):2751–2758, dec 2002.
- [60] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [61] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [62] Alex Graves, Greg Wayne, Malcolm Reynolds, Tim Harley, Ivo Danihelka, Agnieszka Grabska-Barwińska, Sergio Gómez Colmenarejo, Edward Grefenstette, Tiago Ramalho, John Agapiou, Adrià Puigdomènech Badia, Karl Moritz Hermann, Yori Zwols, Georg Ostrovski, Adam Cain,

- Helen King, Christopher Summerfield, Phil Blunsom, Koray Kavukcuoglu, and Demis Hassabis. Hybrid computing using a neural network with dynamic external memory. *Nature*, 538(7626):471–476, oct 2016.
- [63] Cheng Guo, Meng Xiao, Momchil Minkov, Yu Shi, and Shanhui Fan. Photonic crystal slab Laplace operator for image differentiation. *Optica*, 5(3):251–256, March 2018.
- [64] Xiang Guo, Chang-Ling Zou, Hojoong Jung, and Hong X. Tang. On-Chip Strong Coupling and Efficient Frequency Conversion between Telecom and Visible Optical Modes. *Physical Review Letters*, 117(12), September 2016.
- [65] Sangyoon Han, Tae Joon Seok, Niels Quack, Byung-Wook Yoo, and Ming C Wu. Large-scale silicon photonic switches with movable directional couplers. *Optica*, 2(4):370–375, 2015.
- [66] Nicholas C Harris, Jacques Carolan, Darius Bunandar, Mihika Prabhu, Michael Hochberg, Tom Baehr-Jones, Michael L Fanto, A Matthew Smith, Christopher C Tison, Paul M Alsing, et al. Linear programmable nanophotonic processors. *Optica*, 5(12):1623–1631, 2018.
- [67] Nicholas C. Harris, Gregory R. Steinbrecher, Mihika Prabhu, Yoav Lahini, Jacob Mower, Darius Bunandar, Changchen Chen, Franco N. C. Wong, Tom Baehr-Jones, Michael Hochberg, Seth Lloyd, and Dirk Englund. Quantum transport simulations in a programmable nanophotonic processor. *Nature Photonics*, 11(7):447–452, June 2017.
- [68] Nicholas C. Harris, Gregory R. Steinbrecher, Mihika Prabhu, Yoav Lahini, Jacob Mower, Darius Bunandar, Changchen Chen, Franco N. C. Wong, Tom Baehr-Jones, Michael Hochberg, Seth Lloyd, and Dirk Englund. Quantum transport simulations in a programmable nanophotonic processor. *Nature Photonics*, 11:447, Jun 2017.
- [69] Hermann A Haus. *Waves and fields in optoelectronics*. Prentice-Hall,, 1984.
- [70] J Hebling. Derivation of the pulse front tilt caused by angular dispersion. *Opt. Quant. Electron.*, 28(12):1759–1763, 1996.
- [71] Michiel Hermans, Michaël Burm, Thomas Van Vaerenbergh, Joni Dambre, and Peter Bienstman. Trainable hardware for dynamical computing using error backpropagation through physical media. *Nature Communications*, 6:6729, 2015.
- [72] Michiel Hermans, Michaël Burm, Thomas Van Vaerenbergh, Joni Dambre, and Peter Bienstman. Trainable hardware for dynamical computing using error backpropagation through physical media. *Nature Communications*, 6:6729, March 2015.
- [73] James Hillenbrand, Laura A. Getty, Michael J. Clark, and Kimberlee Wheeler. Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5):3099–3111, May 1995.

- [74] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [75] Po-Chun Hsu, Alex Y. Song, Peter B. Catrysse, Chong Liu, Yucan Peng, Jin Xie, Shanhui Fan, and Yi Cui. Radiative human body cooling by nanoporous polyethylene textile. *Science*, 353(6303):1019–1023, September 2016.
- [76] T. W. Hughes. DLA Laser Coupling Simulation Software. <https://github.com/twhughes/DLA-Laser-Coupling-Simulation-Software>, 2017.
- [77] Tyler Hughes, Georgios Veronis, Kent P. Wootton, R. Joel England, and Shanhui Fan. Method for computationally efficient design of dielectric laser accelerator structures. *Optics Express*, 25(13):15414–15427, June 2017.
- [78] Tyler W Hughes. Accelerator-control. <https://github.com/twhughes/Accelerator-Control>, 2018.
- [79] Tyler W. Hughes, R. Joel England, and Shanhui Fan. Reconfigurable Photonic Circuit for Controlled Power Delivery to Laser-Driven Accelerators on a Chip. *Physical Review Applied*, 11(6):064014, June 2019.
- [80] Tyler W Hughes, Momchil Minkov, Yu Shi, and Shanhui Fan. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica*, 5(7):864–871, 2018.
- [81] Tyler W Hughes, Momchil Minkov, and Ian A D Williamson. Angler – an adjoint nonlinear gradient open-source package. <https://github.com/fancompute/angler>, 2018.
- [82] Tyler W Hughes, Momchil Minkov, Ian AD Williamson, and Shanhui Fan. Adjoint method and inverse design for nonlinear nanophotonic devices. *ACS Photonics*, 5(12):4781–4787, 2018.
- [83] Tyler W Hughes, Si Tan, Zhixin Zhao, Neil V Sapra, Yun Jo Lee, Kenneth J Leedle, Huiyang Deng, Yu Miao, Dylan S Black, Minghao Qi, Olav Solgaard, James S Harris, Jelena Vuckovic, Robert L Byer, and Shanhui Fan. On-chip laser power delivery system for dielectric laser accelerators. *Physical Review Applied*, Accepted 2018.
- [84] Tyler W. Hughes, Si Tan, Zhixin Zhao, Neil V. Sapra, Kenneth J. Leedle, Huiyang Deng, Yu Miao, Dylan S. Black, Olav Solgaard, James S. Harris, Jelena Vuckovic, Robert L. Byer, Shanhui Fan, R. Joel England, Yun Jo Lee, and Minghao Qi. On-Chip Laser-Power Delivery System for Dielectric Laser Accelerators. *Physical Review Applied*, 9(5):054017, May 2018.
- [85] Michael Hüskens and Peter Stagge. Recurrent neural networks for time series classification. *Neurocomputing*, 50:223–235, January 2003.

- [86] John David Jackson. *Classical electrodynamics*. John Wiley & Sons, 2007.
- [87] B. Jalali and S. Fathpour. Silicon Photonics. *Journal of Lightwave Technology*, 24(12):4600–4615, December 2006.
- [88] Antony Jameson. Aerodynamic Shape Optimization Using the Adjoint Method. page 30.
- [89] J. S. Jensen, O. Sigmund, L. H. Frandsen, P. I. Borel, A. Harporth, and M. Kristensen. Topology design and fabrication of an efficient double 90/spl deg/ photonic Crystal waveguide bend. *IEEE Photonics Technology Letters*, 17(6):1202–1204, June 2005.
- [90] Xiantao Jiang, Simon Gross, Michael J. Withford, Han Zhang, Dong-Il Yeom, Fabian Rotermund, and Alexander Fuerbach. Low-dimensional nanomaterial saturable absorbers for ultrashort-pulsed waveguide lasers. *Optical Materials Express*, 8(10):3055, October 2018.
- [91] Li Jing, Yichen Shen, Tena Dubcek, John Peurifoy, Scott Skirlo, Yann LeCun, Max Tegmark, and Marin Soljačić. Tunable efficient unitary neural networks (eunn) and their application to rnns. In *Proceedings of the 34th International Conference on Machine Learning- Volume 70*, pages 1733–1741. JMLR. org, 2017.
- [92] Michael I Jordan. Serial order: A parallel distributed processing approach. In *Advances in psychology*, volume 121, pages 471–495. Elsevier, 1997.
- [93] S Jutamulia. Overview of hybrid optical neural networks. *Science*, 28(2), 1996.
- [94] Jan Kern, Ruchira Chatterjee, Iris D. Young, Franklin D. Fuller, Louise Lassalle, Mohamed Ibrahim, Sheraz Gul, Thomas Fransson, Aaron S. Brewster, Roberto Alonso-Mori, Rana Hussein, Miao Zhang, Lacey Douthit, Casper de Lichtenberg, Mun Hon Cheah, Dmitry Shevela, Julia Wersig, Ina Seuffert, Dimosthenis Sokaras, Ernest Pastor, Clemens Weninger, Thomas Kroll, Raymond G. Sierra, Pierre Aller, Agata Butryn, Allen M. Orville, Mengning Liang, Alexander Batyuk, Jason E. Koglin, Sergio Carbajo, Sébastien Boutet, Nigel W. Moriarty, James M. Holton, Holger Dobbek, Paul D. Adams, Uwe Bergmann, Nicholas K. Sauter, Athina Zouni, Johannes Messinger, Junko Yano, and Vittal K. Yachandra. Structures of the intermediates of Kok’s photosynthetic water oxidation clock. *Nature*, 563(7731):421, November 2018.
- [95] Erfan Khoram, Ang Chen, Dianjing Liu, Qiqi Wang, and Zongfu Yu. Stochastic optimization of nonlinear nanophotonic media for artificial neural inference. *arXiv preprint arXiv:1810.07815*, 2018.
- [96] Erfan Khoram, Ang Chen, Dianjing Liu, Qiqi Wang, Ming yuan, and Zongfu Yu. Stochastic Optimization of Nonlinear Nanophotonic Media for Artificial Neural Inference. *arXiv:1810.07815 [physics]*, October 2018.

- [97] W D Kimura et al. Laser acceleration of relativistic electrons using the inverse Cherenkov effect. *Phys. Rev. Lett.*, 74(4):546, 1995.
- [98] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [99] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, December 2014.
- [100] M Kozák et al. Dielectric laser acceleration of sub-relativistic electrons by few-cycle laser pulses. *Nucl. Instrum. Methods Phys. Res., Sect. A*, 2016. DOI: 10.1016/j.nima.2016.12.051.
- [101] Martin Kozak. Status report erlangen (hommelhoff group). ACHIP 3rd Collaboration Meeting, 2016.
- [102] Hoyeong Kwon, Dimitrios Sounas, Andrea Cordaro, Albert Polman, and Andrea Alù. Nonlocal Metasurfaces for Optical Signal Processing. *Physical Review Letters*, 121(17):173004, October 2018.
- [103] 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. *Opt. Lett.*, 39(4):941–944, 2014.
- [104] David Kwong, Amir Hosseini, Yang Zhang, and Ray T Chen. 1 × 12 unequally spaced waveguide array for actively tuned optical phased array on a silicon nanomembrane. *Appl. Phys. Lett.*, 99(5):051104, 2011.
- [105] Christopher M Lalau-Keraly et al. Adjoint shape optimization applied to electromagnetic design. *Opt. Express*, 21(18):21693–21701, 2013.
- [106] Michael R. Lamont, Barry Luther-Davies, Duk-Yong Choi, Steve Madden, and Benjamin J. Eggleton. Supercontinuum generation in dispersion engineered highly nonlinear ( $\gamma = 10 \text{ w/m}$ )  $\text{As}_2\text{S}_3$  chalcogenide planar waveguide. *Optics Express*, 16(19):14938, September 2008.
- [107] Floris Laporte, Andrew Katumba, Joni Dambre, and Peter Bienstman. Numerical demonstration of neuromorphic computing with photonic crystal cavities. *Optics express*, 26(7):7955–7964, 2018.
- [108] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
- [109] Ken Leedle. Private communication.
- [110] Kenneth J Leedle et al. Dielectric laser acceleration of sub-100 keV electrons with silicon dual-pillar grating structures. *Opt. Lett.*, 40(18):4344–4347, 2015.

- [111] A. L. Lentine and D. A. B. Miller. Evolution of the SEED technology: Bistable logic gates to optoelectronic smart pixels. *IEEE Journal of Quantum Electronics*, 29(2):655–669, February 1993.
- [112] LIGO Scientific Collaboration and Virgo Collaboration, B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, V. B. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, L. Aiello, A. Ain, P. Ajith, B. Allen, A. Allocca, P. A. Altin, S. B. Anderson, W. G. Anderson, K. Arai, M. A. Arain, Araya, et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116(6):061102, February 2016.
- [113] Xing Lin, Yair Rivenson, Nezih T Yardimci, Muhammed Veli, Yi Luo, Mona Jarrahi, and Aydogan Ozcan. All-optical machine learning using diffractive deep neural networks. *Science*, 361(6406):1004–1008, 2018.
- [114] Zin Lin, Xiangdong Liang, Marko Lončar, Steven G. Johnson, and Alejandro W. Rodriguez. Cavity-enhanced second-harmonic generation via nonlinear-overlap optimization. *Optica*, 3(3):233–238, March 2016.
- [115] Zin Lin, Marko Lončar, and Alejandro W. Rodriguez. Topology optimization of multi-track ring resonators and 2d microcavities for nonlinear frequency conversion. *Optics Letters*, 42(14):2818, July 2017.
- [116] Arka Majumdar and Armand Rundquist. Cavity-enabled self-electro-optic bistability in silicon photonics. *Optics Letters*, 39(13):3864, July 2014.
- [117] Josh McNeur, Martin Kozak, Norbert Schönenberger, Ang Li, Alexander Tafel, and Peter Hommelhoff. Laser-driven acceleration of subrelativistic electrons near a nanostructured dielectric grating: From acceleration via higher spatial harmonics to necessary elements of a dielectric accelerator. *Nucl. Instrum. Methods Phys. Res., Sect. A*, 829:50 – 51, 2016.
- [118] Joshua McNeur et al. Elements of a dielectric laser accelerator. *arXiv preprint arXiv:1604.07684*, 2016.
- [119] Paul A. Merolla, John V. Arthur, Rodrigo Alvarez-Icaza, Andrew S. Cassidy, Jun Sawada, Filipp Akopyan, Bryan L. Jackson, Nabil Imam, Chen Guo, Yutaka Nakamura, Bernard Brezzo, Ivan Vo, Steven K. Esser, Rathinakumar Appuswamy, Brian Taba, Arnon Amir, Myron D. Flickner, William P. Risk, Rajit Manohar, and Dharmendra S. Modha. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, 345(6197):668–673, 2014.

- [120] Benjamin J. Metcalf, Nicholas Thomas-Peter, Justin B. Spring, Dmytro Kundys, Matthew A. Broome, Peter C. Humphreys, Xian-Min Jin, Marco Barbieri, W. Steven Kolthammer, James C. Gates, Brian J. Smith, Nathan K. Langford, Peter G.R. Smith, and Ian A. Walmsley. Multiphoton quantum interference in a multiport integrated photonic device. *Nature Communications*, 4:1356, January 2013.
- [121] D. A. B. Miller. Sorting out light. *Science*, 347(6229):1423–1424, March 2015.
- [122] David A. B. Miller. Self-aligning universal beam coupler. *Optics Express*, 21(5):6360, March 2013.
- [123] David A B Miller. Self-aligning universal beam coupler. *Optics Express*, 21(5):6360–70, 2013.
- [124] David A. B. Miller. Self-configuring universal linear optical component. *Photonics Research*, 1(1):1, 2013.
- [125] David A. B. Miller. Self-configuring universal linear optical component [Invited]. *Photonics Research*, 1(1):1, June 2013.
- [126] David A. B. Miller. Perfect optics with imperfect components. *Optica*, 2(8):747, August 2015.
- [127] David A. B. Miller. Perfect optics with imperfect components. *Optica*, 2(8):747, 2015.
- [128] David A. B. Miller. Setting up meshes of interferometers; reversed local light interference method. *Opt. Express*, 25(23):29233–29248, Nov 2017.
- [129] David A. B. Miller. Setting up meshes of interferometers – reversed local light interference method. *Optics Express*, 25(23):29233, November 2017.
- [130] David A. B. Miller. Self-configuring complex optics (Conference Presentation). In *Quantum Sensing and Nano Electronics and Photonics XV*, volume 10540, page 1054017. International Society for Optics and Photonics, March 2018.
- [131] Amit Mizrahi and Levi Schächter. Optical bragg accelerators. *Physical Review E*, 70(1):016505, 2004.
- [132] K Mizuno et al. Experimental evidence of the inverse Smith–Purcell effect. *Nature*, 328(6125):45–47, 1987.
- [133] Sean Molesky, Zin Lin, Alexander Y. Piggott, Weiliang Jin, Jelena Vucković, and Alejandro W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659, November 2018.
- [134] Joong Ho Moon, Jin Ho Kim, Ki-jeong Kim, Tai-Hee Kang, Bongsoo Kim, Chan-Ho Kim, Jong Hoon Hahn, and Joon Won Park. Absolute Surface Density of the Amine Group of the Aminosilylated Thin Layers: Ultraviolet-Visible Spectroscopy, Second Harmonic Generation,

- and Synchrotron-Radiation Photoelectron Spectroscopy Study. *Langmuir*, 13(16):4305–4310, August 1997.
- [135] P Musumeci et al. High energy gain of trapped electrons in a tapered, diffraction-dominated inverse-free-electron laser. *Phys. Rev. Lett.*, 94(15):154801, 2005.
  - [136] Y. C. N. Na, R. H. Siemann, and R. L. Byer. Energy efficiency of an intracavity coupled, laser-driven linear accelerator pumped by an external laser. *Phys. Rev. ST-AB*, 8:031301, 2005.
  - [137] B Naranjo, A Valloni, S Putterman, and JB Rosenzweig. Stable charged-particle acceleration and focusing in a laser accelerator using spatial harmonics. *Physical review letters*, 109(16):164803, 2012.
  - [138] Uwe Niedermayer, Oliver Boine-Frankenheim, and Thilo Egenolf. Designing a dielectric laser accelerator on a chip. In *8th Int. Particle Accelerator Conf.(IPAC'17), Copenhagen, Denmark*, 2017.
  - [139] Uwe Niedermayer, Thilo Egenolf, and Oliver Boine-Frankenheim. Beam dynamics analysis of dielectric laser acceleration using a fast 6d tracking scheme. *Physical Review Accelerators and Beams*, 20(11):111302, 2017.
  - [140] Uwe Niedermayer, Thilo Egenolf, Oliver Boine-Frankenheim, and Peter Hommelhoff. Alternating-phase focusing for dielectric-laser acceleration. *Physical Review Letters*, 121(21):214801, 2018.
  - [141] Jeremy L. O'Brien, Akira Furusawa, and Jelena Vučkovic. Photonic quantum technologies. *Nature Photonics*, 3(12):687–695, December 2009.
  - [142] Yoshitomo Okawachi, Kasturi Saha, Jacob S. Levy, Y. Henry Wen, Michal Lipson, and Alexander L. Gaeta. Octave-spanning frequency comb generation in a silicon nitride chip. *Optics Letters*, 36(17):3398, September 2011.
  - [143] Ardavan F. Oskooi, Lei Zhang, Yehuda Avniel, and Steven G. Johnson. The failure of perfectly matched layers, and towards their redemption by adiabatic absorbers. *Optics Express*, 16(15):11376–11392, 2008.
  - [144] M. Necat Ozisik. *Inverse Heat Transfer : Fundamentals and Applications*. Routledge, May 2018.
  - [145] Sunil Pai, Ben Bartlett, Olav Solgaard, and David AB Miller. Matrix optimization on universal unitary photonic devices. *arXiv preprint arXiv:1808.00458*, 2018.

- [146] Nam Hun Park, Hwanseong Jeong, Sun Young Choi, Mi Hye Kim, Fabian Rotermund, and Dong-Il Yeom. Monolayer graphene saturable absorbers with strongly enhanced evanescent-field interaction for ultrafast fiber laser mode-locking. *Optics Express*, 23(15):19806, July 2015.
- [147] E. A. Peralta et al. Demonstration of electron acceleration in a laser-driven dielectric microstructure. *Nature*, 503(7474):91–94, 2013.
- [148] Alexander Y. Piggott, Jesse Lu, Konstantinos G. Lagoudakis, Jan Petykiewicz, Thomas M. Babinec, and Jelena Vučković. Inverse design and demonstration of a compact and broadband on-chip wavelength demultiplexer. *Nature Photonics*, 9(6):374–377, June 2015.
- [149] Alexander Y Piggott, Jan Petykiewicz, Logan Su, and Jelena Vučković. Fabrication-constrained nanophotonic inverse design. *Sci. Rep.*, 7:1786, 2017.
- [150] T. Plettner, R. L. Byer, C. McGuinness, and P. Hommelhoff. Photonic-based laser driven electron beam deflection and focusing structures. *Phys. Rev. ST Accel. Beams*, 12(10):101302, 2009.
- [151] T Plettner and RL Byer. Microstructure-based laser-driven free-electron laser. *Nucl. Instrum. Methods Phys. Res., Sect. A*, 593(1):63–66, 2008.
- [152] T Plettner et al. Proposed few-optical cycle laser-driven particle accelerator structure. *Phys. Rev. ST Accel. Beams*, 9(11):111301, 2006.
- [153] C. R. Pollock and Michal Lipson. *Integrated photonics*. Kluwer Academic, Norwell, Mass., 2010. OCLC: 854553854.
- [154] Menno Poot and Hong X Tang. Broadband nanoelectromechanical phase shifting of light on a chip. *Appl. Phys. Lett.*, 104(6):061101, 2014.
- [155] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.
- [156] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov. Training and operation of an integrated neuromorphic network based on metal-oxide memristors. *Nature*, 521(7550):61–64, 2015.
- [157] Demetri Psaltis, David Brady, and Kelvin Wagner. Adaptive optical networks using photorefractive crystals. *Applied Optics*, 27(9):1752–1759, 1988.
- [158] Minhao Pu, Liu Liu, Haiyan Ou, Kresten Yvind, and Jørn M Hvam. Ultra-low-loss inverted taper coupler for silicon-on-insulator ridge waveguide. *Opt. Commun.*, 283(19):3678–3682, 2010.

- [159] Marina Radulaski, Ranojoy Bose, Tho Tran, Thomas Van Vaerenbergh, David Kielinski, and Raymond G. Beausoleil. Thermally Tunable Hybrid Photonic Architecture for Nonlinear Optical Circuits. *ACS Photonics*, 5(11):4323–4329, November 2018.
- [160] Aaswath P. Raman, Marc Abou Anoma, Linxiao Zhu, Eden Rephaeli, and Shanhui Fan. Passive radiative cooling below ambient air temperature under direct sunlight. *Nature*, 515(7528):540–544, November 2014.
- [161] Michael Reck, Anton Zeilinger, Herbert J. Bernstein, and Philip Bertani. Experimental realization of any discrete unitary operator. *Physical Review Letters*, 73(1):58–61, July 1994.
- [162] Michael Reck, Anton Zeilinger, Herbert J. Bernstein, and Philip Bertani. Experimental realization of any discrete unitary operator. *Physical Review Letters*, 73(1):58–61, jul 1994.
- [163] David Rosenbluth, Konstantin Kravtsov, Mable P Fok, and Paul R Prucnal. A high performance photonic pulse processing device. *Optics Express*, 17(25):22767–22772, 2009.
- [164] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. *Parallel Distributed Processing*, volume 1. MIT Press, 1986.
- [165] T. S. Rutherford, W. M. Tulloch, E. K. Gustafson, and R. L. Byer. Edge-pumped quasi-three-level slab lasers: Design and power scaling. *IEEE Journal of Quantum Electronics*, 36:205–219, 2000.
- [166] Wesley D Sacher, Jared C Mikkelsen, Patrick Dumais, Jia Jiang, Dominic Goodwill, Xianshu Luo, Ying Huang, Yisu Yang, Antoine Bois, Patrick Guo-Qiang Lo, et al. Tri-layer silicon nitride-on-silicon photonic platform for ultra-low-loss crossings and interlayer transitions. *Optics express*, 25(25):30862–30875, 2017.
- [167] Neil V Sapra, Dries Vercruyse, Logan Su, Ki Youl Yang, Jinhe Skarda, Alexander Piggott, and Jelena Vučković. Inverse design and demonstration of broadband grating couplers. *IEEE Journal of Selected Topics in Quantum Electronics*, 2019.
- [168] Neil V. Sapra, Ki Youl Yang, Dries Vercruyse, Kenneth J. Leedle, Dylan S. Black, R. Joel England, Logan Su, Yu Miao, Olav Solgaard, Robert L. Byer, and Jelena Vučković. On-chip integrated laser-driven particle accelerator. *arXiv:1905.12822 [physics]*, May 2019. arXiv: 1905.12822.
- [169] Jeffrey M. Shainline, Sonia M. Buckley, Richard P. Mirin, and Sae Woo Nam. Superconducting Optoelectronic Circuits for Neuromorphic Computing. *Physical Review Applied*, 7(3):1–27, 2017.

- [170] Yichen Shen, Nicholas C. Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, and Marin Soljacic. Deep Learning with Coherent Nanophotonic Circuits. *Nature Photonics*, 11:441, jun 2017.
- [171] Yichen Shen, Nicholas C. Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, and Marin Soljacic. Deep learning with coherent nanophotonic circuits. *Nature Photonics*, 11(7):441–446, June 2017.
- [172] Yichen Shen, Nicholas C Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund, and Marin Soljačić. Deep learning with coherent nanophotonic circuits. *Nature Photonics*, 11(7):441, 2017.
- [173] Yu Shi, Wonseok Shin, and Shanhui Fan. Multi-frequency finite-difference frequency-domain algorithm for active nanophotonic device simulations. *Optica*, 3(11):1256–1259, Nov 2016.
- [174] Wonseok Shin and Shanhui Fan. Choice of the perfectly matched layer boundary condition for frequency-domain Maxwell’s equations solvers. *J. Comput. Phys.*, 231(8):3406–3431, 2012.
- [175] T Shoji, T Tsuchizawa, T Watanabe, K Yamada, and H Morita. Low loss mode size converter from  $0.3\ \mu\text{m}$  square si wire waveguides to singlemode fibres. *Electronics Letters*, 38(25):1669–1670, 2002.
- [176] R. H. Siemann. Energy efficiency of laser driven, structure based accelerators. *Phys. Rev. ST-AB*, 7:061303, 2004.
- [177] Ole Sigmund. Systematic Design of Metamaterials by Topology Optimization. In R. Pyrz and J. C. Rauhe, editors, *IUTAM Symposium on Modelling Nanomaterials and Nanosystems*, IUTAM Bookseries, pages 151–159. Springer Netherlands, 2009.
- [178] Alexandre Silva, Francesco Monticone, Giuseppe Castaldi, Vincenzo Galdi, Andrea Alù, and Nader Engheta. Performing Mathematical Operations with Metamaterials. *Science*, 343(6167):160–163, January 2014.
- [179] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *Nature*, 550(7676):354, 2017.
- [180] Elena I Smotrova, Alexander I Nosich, Trevor M Benson, and Phillip Sewell. Cold-cavity thresholds of microdisks with uniform and nonuniform gain: quasi-3-d modeling with accurate 2-d analysis. *IEEE J. Sel. Top. Quantum Electron.*, 11(5):1135–1142, 2005.
- [181] Marin Soljačić, Mihai Ibanescu, Steven G. Johnson, Yoel Fink, and J. D. Joannopoulos. Optimal bistable switching in nonlinear photonic crystals. *Phys. Rev. E*, 66:055601, Nov 2002.

- [182] Ken Soong et al. Laser damage threshold measurements of optical materials for direct laser accelerators. *AIP Conf. Proc.*, 1507(1):511–515, 2012.
- [183] Ken Soong, Edgar A Peralta, R Joel England, Ziran Wu, Eric R Colby, Igor Makasyuk, James P MacArthur, Andrew Ceballos, and Robert L Byer. Electron beam position monitor for a dielectric microaccelerator. *Optics letters*, 39(16):4747–4750, 2014.
- [184] T. Spuesens, S. Pathak, M. Vanslembrouck, P. Dumon, and W. Bogaerts. Grating Couplers With an Integrated Power Splitter for High-Intensity Optical Power Distribution. *IEEE Photonics Technology Letters*, 28(11):1173–1176, June 2016.
- [185] Thijs Spuesens, Shubnath Pathak, Michael Vanslembrouck, Pieter Dumon, and Wim Bogaerts. Grating couplers with an integrated power splitter for high-intensity optical power distribution. *IEEE Photon. Technol. Lett.*, 28(11):1173–1176, 2016.
- [186] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(1):1929–1958, 2014.
- [187] Gilbert Strang. *Computational Science and Engineering*. Wellesley-Cambridge Press, 2007. Chapter 8.
- [188] B. C. Stuart, M. D. Feit, S. Herman, A. M. Rubenchik, B. W. Shore, and M. D. Perry. Nanosecond-to-femtosecond laser-induced breakdown in dielectrics. *Phys. Rev. B*, 53(4):1749, 1996.
- [189] B. C. Stuart, M. D. Feit, A. M. Rubenchik, B. W. Shore, and M. D. Perry. Laser-induced damage in dielectrics with nanosecond to subpicosecond pulses. *Phys. Rev. Lett.*, 74:2248–2251, Mar 1995.
- [190] Wonjoo Suh, Zheng Wang, and Shanhui Fan. Temporal coupled-mode theory and the presence of non-orthogonal modes in lossless multimode cavities. *IEEE Journal of Quantum Electronics*, 40(10):1511–1518, 2004.
- [191] Jie Sun, Erman Timurdogan, Ami Yaacobi, Ehsan Shah Hosseini, and Michael R Watts. Large-scale nanophotonic phased array. *Nature*, 493(7431):195, 2013.
- [192] Allen Taflove and Susan C Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Artech House Publishers, Norwood, MA, 2nd edition, 2000.
- [193] Dirk Taillaert, Frederik Van Laere, Melanie Ayre, Wim Bogaerts, Dries Van Thourhout, Peter Bienstman, and Roel Baets. Grating couplers for coupling between optical fibers and nanophotonic waveguides. *Japanese Journal of Applied Physics*, 45(8R):6071, 2006.

- [194] Alexander N. Tait, Thomas Ferreira de Lima, Mitchell A. Nahmias, Heidi B. Miller, Hsuan-Tung Peng, Bhavin J. Shastri, and Paul R. Prucnal. A silicon photonic modulator neuron. *arXiv:1812.11898 [physics]*, December 2018.
- [195] Alexander N. Tait, Thomas Ferreira de Lima, Ellen Zhou, Allie X. Wu, Mitchell A. Nahmias, Bhavin J. Shastri, and Paul R. Prucnal. Neuromorphic photonic networks using silicon photonic weight banks. *Scientific Reports*, 7(1):7430, August 2017.
- [196] 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.
- [197] Si Tan, Zhixin Zhao, Karel Urbanek, Tyler Hughes, Yun Jo Lee, Shanhui Fan, James S Harris, and Robert L Byer. Silicon nitride waveguide as a power delivery component for on-chip dielectric laser accelerators. *Optics letters*, 44(2):335–338, 2019.
- [198] Masataka Tanaka and G. S. Dulikravich. *Inverse Problems in Engineering Mechanics*. Elsevier, November 1998. Google-Books-ID: tyOpk49g0vsC.
- [199] Malvin Carl Teich and B Saleh. Fundamentals of photonics. *Canada, Wiley Interscience*, 3, 1991.
- [200] F. Ursell. The long-wave paradox in the theory of gravity waves. *Mathematical Proceedings of the Cambridge Philosophical Society*, 49(04):685, October 1953.
- [201] Karel Van Acleyen, DC O'Brien, F Payne, Wim Bogaerts, and Roel Baets. Optical retroreflective marker fabricated on silicon-on-insulator. *IEEE Photonics Journal*, 3(5):789–798, 2011.
- [202] 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*, 5:1–6, 2014.
- [203] Kristof Vandoorne, Pauline Mehet, 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*, 5:3541, 2014.
- [204] Georgios Veronis, Robert W. Dutton, and Shanhui Fan. Method for sensitivity analysis of photonic crystal devices. *Optics Letters*, 29(19):2288, October 2004.
- [205] Georgios Veronis, Robert W. Dutton, and Shanhui Fan. Method for sensitivity analysis of photonic crystal devices. *Optics Letters*, 29(19):2288, oct 2004.

- [206] Kelvin Wagner and Demetri Psaltis. Multilayer optical learning networks. *Applied Optics*, 26(23):5061–5076, 1987.
- [207] Y Wei, G Xia, JDA Smith, and CP Welsch. Dual-gratings with a bragg reflector for dielectric laser-driven accelerators. *Physics of Plasmas*, 24(7):073115, 2017.
- [208] J. A. C. Weideman and B. M. Herbst. Split-step methods for the solution of the nonlinear schrödinger equation. *SIAM J. Numer. Anal.*, 23(3):485–507, 1986.
- [209] N. H. E. Weste and K. Eshraghian. Principles of CMOS VLSI design: A systems perspective. *NASA STI/Recon Technical Report A*, 85, 1985.
- [210] Richard T. White and Tanya M. Monro. Cascaded Raman shifting of high-peak-power nanosecond pulses in  $\text{As}_2\text{S}_3$  and  $\text{As}_2\text{Se}_3$  optical fibers. *Optics Letters*, 36(12):2351, June 2011.
- [211] Darrell Whitley. A genetic algorithm tutorial. *Statistics and Computing*, 4(2):65–85, June 1994.
- [212] Ian A. D. Williamson, Tyler W. Hughes, Momchil Minkov, Ben Bartlett, Sunil Pai, and Shanhui Fan. Reprogrammable Electro-Optic Nonlinear Activation Functions for Optical Neural Networks. *arXiv:1903.04579 [physics]*, March 2019. arXiv: 1903.04579.
- [213] K. P. Wootton, D. B. Cesar, B. M. Cowan, A. Hanuka, I. V. Makasyuk, J. Maxson, E. A. Peralta, K. Soong, Z. Wu, R. L. Byer, P. Musumeci, and R. J. England. Recent demonstration of record high gradients in dielectric laser accelerating structures. 1812(1):060006, 2017.
- [214] Z. Wu, R. J. England, C-K. Ng, B. Cowan, C. McGuinness, C. Lee, M. Qi, and S. Tantawi. Coupling power into accelerating mode of a three-dimensional silicon woodpile photonics bandgap waveguide. *Phys. Rev. ST-AB*, 17:081301, 2014.
- [215] Ami Yaacobi, Jie Sun, Michele Moresco, Gerald Leake, Douglas Coolbaugh, and Michael R Watts. Integrated phased array for wide-angle beam steering. *Opt. Lett.*, 39(15):4575–4578, 2014.
- [216] Koji Yamada. Silicon photonic wire waveguides: fundamentals and applications. In *Silicon Photonics II*, pages 1–29. Springer, 2011.
- [217] Daiki Yamashita, Yasushi Takahashi, Takashi Asano, and Susumu Noda. Raman shift and strain effect in high-Q photonic crystal silicon nanocavity. *Optics Express*, 23(4):3951, February 2015.
- [218] Mehmet Fatih Yanik, Shanhui Fan, Marin Soljačić, and J. D. Joannopoulos. All-optical transistor action with bistable switching in a photonic crystal cross-waveguide geometry. *Optics Letters*, 28(24):2506, December 2003.

- [219] Kaisheng Yao, Geoffrey Zweig, Mei-Yuh Hwang, Yangyang Shi, and Dong Yu. Recurrent neural networks for language understanding. In *Interspeech*, pages 2524–2528, 2013.
- [220] Yu Ye, Fang Liu, Mengxuan Wang, and Yidong Huang. Deep ultraviolet smith-purcell radiation. In *CLEO: Science and Innovations*, pages SM4I–4. Optical Society of America, 2018.
- [221] Kane Yee. Numerical solution of initial boundary value problems involving maxwell's equations in isotropic media. *IEEE Transactions on antennas and propagation*, 14(3):302–307, 1966.
- [222] Peyman Yousefi, Norbert Schönenberger, Joshua Mcneur, Martin Kozák, Uwe Niedermayer, and Peter Hommelhoff. Dielectric laser electron acceleration in a dual pillar grating with a distributed bragg reflector. *Optics letters*, 44(6):1520–1523, 2019.
- [223] Zongfu Yu, Aaswath Raman, and Shanhui Fan. Fundamental limit of nanophotonic light trapping in solar cells. *Proceedings of the National Academy of Sciences*, 107(41):17491–17496, October 2010.
- [224] T B Zhang et al. Stimulated dielectric wake-field accelerator. *Phys. Rev. E*, 56(4):4647, 1997.
- [225] Yi Zhang, Shuyu Yang, Andy Eu-Jin Lim, Guo-Qiang Lo, Christophe Galland, Tom Baehr-Jones, and Michael Hochberg. A compact and low loss y-junction for submicron silicon waveguide. *Opt. Express*, 21(1):1310–1316, 2013.
- [226] Zhixin Zhao, Tyler W. Hughes, Si Tan, Huiyang Deng, Neil Sapra, R. Joel England, Jelena Vuckovic, James S. Harris, Robert L. Byer, and Shanhui Fan. Design of a tapered slot waveguide dielectric laser accelerator for sub-relativistic electrons. *Optics Express*, 26(18):22801–22815, September 2018.
- [227] Mingdong Zhou, Boyan S Lazarov, Fengwen Wang, and Ole Sigmund. Minimum length scale in topology optimization by geometric constraints. *Computer Methods in Applied Mechanics and Engineering*, 293:266–282, 2015.