

Rowan University

Rowan Digital Works

Theses and Dissertations

6-28-2024

MATRIX PROCESSING WITH PHOTONIC ANALOG COMPUTING

James Michael Garofolo

Rowan University

Follow this and additional works at: <https://rdw.rowan.edu/etd>



Part of the [Computer Engineering Commons](#), and the [Electrical and Computer Engineering Commons](#)

Recommended Citation

Garofolo, James Michael, "MATRIX PROCESSING WITH PHOTONIC ANALOG COMPUTING" (2024).

Theses and Dissertations. 3267.

<https://rdw.rowan.edu/etd/3267>

This Thesis is brought to you for free and open access by Rowan Digital Works. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Rowan Digital Works. For more information, please contact graduateresearch@rowan.edu.

MATRIX PROCESSING WITH PHOTONIC ANALOG COMPUTING

by

James M. Garofolo

A Thesis

Submitted to the
Department of Electrical and Computer Engineering
College of Engineering
In partial fulfillment of the requirement
For the degree of
Master of Science in Electrical and Computer Engineering
at
Rowan University
May 30, 2024

Thesis Chair: Ben Wu, Ph.D., Associate Professor, Department of Electrical and
Computer Engineering

Committee Members:

Huaxia Wang, Ph.D., Assistant Professor, Department of Electrical and Computer
Engineering

Guimu Guo, Ph.D., Assistant Professor, Department of Computer Science

© 2024 James M. Garofolo

Dedication

I would like to dedicate this thesis to Gianna Emily Figueroa. Every one of my achievements has been underpinned by the unwavering love and support she provided me during my time as a researcher, and I will work tirelessly to repay her kindness as we move on in life together.

Acknowledgements

I would like to give my deepest thanks to my advisor, Dr. Ben Wu, for both sponsoring and guiding me along this journey. He enlightened me to pursuits in science that I didn't know existed, and enabled me to grow into a driving force for those pursuits with my colleagues. Understanding the risks of taking me under his wing, I can only hope I've sufficiently rewarded his efforts with my own.

Secondly, I would like to thank the many researchers who have mentored me in their craft throughout my education. Though all of my professors have affected me in this way, Dr. Dwaipayan Chakraborty and Dr. Huaxia Wang stick in my mind as the most impactful mentors of my time at Rowan. Their passion for their craft inspired my own, and their encouragement of my passions drove me to succeed. From devoting precious time to hearing out ideas to pushing me to publish the results of course projects, these researchers' actions have shaped me into the courageous and devoted researcher I am today.

Third, I would like to thank my friends and colleagues in the Rowan Photonic Computing and Communication Laboratory, Taichu Shi and Dr. Yang Qi. Their assistance with and contributions to my work are eclipsed only by the enriching friendships we've grown through our time together. I consider myself lucky to have grown alongside them, and can only hope I contributed as meaningfully to their success as they have to my own.

Lastly, I would like to thank all of the researchers who contributed to my pursuits with their own time, data, ideas, and techniques. Dr. Paul Prucnal and Weipeng Zhang, along with the other members of the Princeton Lightwave communication laboratory, have all contributed invaluable efforts, ideas, and support to my works in the field of photonic computing. Dr. Dimah Dera, Dr. Jacob Epifano, Dr. Giuseppina Carannante and Christopher Angleini also contributed excellent works of their own in the field of machine learning, standing as the foundations upon which my related works could stand. Without these contributions to science and myself, none of the presented works would be possible.

Abstract

James M. Garofolo
MATRIX PROCESSING WITH PHOTONIC ANALOG COMPUTING
2023-2024
Ben Wu, Ph.D.
Master of Science in Electrical and Computer Engineering

In the digital age, a wide variety of engineering problems have been solved, to a great deal of success, by digital computing techniques. The flexibility of software and relatively low cost of digital computing hardware make it an ideal starting point for solving a majority of tasks, and the numerical stability of software solutions make it highly appealing as the major workhorse for computational tasks. Despite this, many problems are actually sub-optimally solved by digital methods, leading to systems with high latency, low throughput, power hungry parallel processing units and an excess of memory for discretizing sensor inputs.

Computational photonic circuits are an emerging field of study which holds a number of advantages over modern digital computations. Their high bandwidth allows for the implementation of operations from basic arithmetic to frequency domain manipulation at speeds and efficiencies that their electrical counterparts are unable to approach, while still consuming less power per operation. This high bandwidth also enables parallelized computations in the same waveguide through the use of Wavelength-Division Multiplexing. Their analog nature allows for signal processing in continuous time, and eliminates the cost, memory requirements and precision loss resulting from the need to digitize massive amounts of data. This thesis will examine applications in which these circuits can provide a more optimal solution to tasks traditionally handled by software, and propose circuit architectures and control techniques that facilitate these more optimal solutions.

Table of Contents

Abstract	v
List of Figures	viii
Chapter 1: Introduction and Background	1
1.1 Motivation for Analog Computing	1
1.2 Mechanisms for Analog Matrix Processing	3
1.2.1 Vector Multiplication	3
1.2.2 Kirchoff Current Summation	8
1.2.3 Common Architectures.....	10
1.3 Analog Technologies in Industry	12
1.4 Advantages of Photonics	14
Chapter 2: Interference Cancellation.....	17
2.1 Introduction	17
2.2 Principle and System Setup.....	21
2.2.1 Principle: Photonic Blind Source Separation	21
2.2.2 Experimental Setup.....	22
2.3 Results and Discussion	25
2.4 Conclusion	30
Chapter 3: Variational Neural Networks	31
3.1 Introduction	31

Table of Contents (Continued)

3.2	Principle and System Setup.....	34
3.2.1	Principle: Variational Density Propagation	34
3.2.2	System Setup	36
3.3	Results and Discussion	39
3.4	Conclusion	45
Chapter 4:	Conclusions	46
References	48

List of Figures

Figure	Page
Figure 1. Geometry, Nomenclature and Behavior of a Microring Resonator	6
Figure 2. Arrangement of Phase-Change Material Based Ring Resonator Weight Bank	8
Figure 3. Illustration of a Crossbar Matrix Vector Multiplier	11
Figure 4. Schematic of a Digital Electronic and Analog Photonic Tensor Processor, Performing an n-by-m Vector-Vector Multiplication.....	12
Figure 5. Application Scenario of the Proposed System.....	20
Figure 6. System Setup for Testing the Interference Cancelling LiDAR Receiver	23
Figure 7. Position of Optical Lenses and Mirrors in the Experimental Setup	24
Figure 8. Transmission Spectrum for Each Channel	26
Figure 9. Results of De-Mixing With a PRBS Interference Signal	28
Figure 10. Results of De-Mixing With a Gaussian Noise Interference Signal....	29
Figure 11. Variational Density Propagation System Schematic	36
Figure 12. Architecture of the Simulated Two-Quadrant Photonic Multiplier Circuit.....	37
Figure 13. Average Predictive Variance Versus Signal-to-Noise Ratio for Gaussian Noise and Adversarial Attacks	40
Figure 14. Average Predictive Variance for 8 Different Corruptions from MNIST-C	42
Figure 15. Evaluation Under Semantic Shift	43

Chapter 1

Introduction and Background

1.1 Motivation for Analog Computing

The rapid reconfigurability, high precision and deterministic operation of digital computing has caused a revolution of applied mathematics and engineering feats that would not have been possible otherwise. Custom hardware algorithms[1], parallelized processing[2], and in-memory or near-memory computing architectures [3] have pushed the bounds of possibility further, making things possible like real-time signal processing, computer vision, and autonomous driving. Though the advantages of digital computing have led to much advancement in the field of engineering, there are many applications for which it is digital computers sub-optimally handle the necessary computations. This sub-optimality has caused many to push for exploration of implementing certain computations more efficiently and with higher performance using analog circuits.

One advantage of analog computation is the data rates and latencies it enables. Because signals travel through analog circuits at the rate of electron flow through wires, analog circuits naturally have low latency in comparison to digital technologies which require each logic stage to reach a quiescent state before signal begins to flow to the next stage[4]. In addition, the removal of the Nyquist limitation allows for much higher bandwidths of operation, leading to much higher data rates[5]. These advantages enable well explored concepts in digital computation to be deployed in analog to applications such as LiDAR based object detection [6, 7, 8, 9], communication signal processing and interference management [10, 11, 12, 13], and autonomous robotic control [14, 15], which all require high resolution data to be processed with low latency and increasingly high throughput in order to ensure robust time-domain control.

Another large advantage of analog solutions to complex computational problems

is the drastic reduction in hardware requirements, including both the number of devices as well as the density of the wires themselves. Digital devices must use binary to encode numbers for processing, requiring busses of 64 wires and hundreds of transistors per mathematical operation. Extending this to a process like signal processing, the task would require an analog-to-digital converter that could satisfy the Nyquist requirements of the signal being processed[1], along with enough memory to hold a digitized copy of the signal and kernel being used for filtering, a processor fast enough to implement each of the multiply-accumulate operations in the vector product before the next time sample needs to enter the system, and a digital-to-analog converter to reproduce the modified signal if necessary for the application. By contrast, analog circuits only require a single wire to represent a signal amplitude with respect to ground, and can implement a wide array of frequency domain functions with a handful of carefully selected components, or occasionally nothing but PCB traces [16]. With recent pushes toward configurability in analog, these elegant frequency domain solutions have even been demonstrated to be programmable[17], significantly narrowing the gap in capability between digital and analog signal processing solutions. This advantage scales particularly well with large data tasks like matrix/tensor operations, as the smaller footprint per operation allows a greater number of operations to be implemented with parallel hardware.

One advantage of analog computing that is easily overlooked is the numerical precision of the computing paradigm in comparison to digital. Because digital computations are relatively immune to the thermal and shot noises and electromagnetic interference which commonly affect analog circuits, they are commonly thought to be more robust in terms of computational precision. That said, in order to discretize a signal, digital computers have to reduce the infinitely resolute continuous amplitude space to a set of discrete points, moving intermediate values to the closest representable point[5]. This effectively adds noise to the measurement, decaying the precision with which mathematical operations can be performed. In addition, when performing iterative numerical analysis, for exam-

ple optimizing an objective function with gradient descent or approximating a differential equation with Euler's method, steps can only be taken in discrete intervals, incurring some approximation error due to the violated assumption of derivative information generalizing beyond the point of inference, with the error scaling strongly with the size of the discrete steps taken [18]. By performing computations in both continuous time and amplitude, incurred errors due to discretization are eliminated. If the analog system performing a given computation can be thoroughly electromagnetically isolated and sufficiently noise free, the decision to use it over digital will lead to a more numerically stable solution than digital computation can manage.

1.2 Mechanisms for Analog Matrix Processing

1.2.1 Vector Multiplication

An essential operation for matrix and tensor processing is the parallel multiplication of vectors and other scalars, vectors or matrices. This operation contributes a majority of the time complexity of many common tasks like signal filtration and neural network inference. Digital multiplication increases in worst-case delay path length, and thus latency, linearly with respect to the number of bits in the operands, making it a very expensive operation[4]. In addition, without multi-threading or specialized hardware, digital multiplications would need to be done sequentially, which wastes unnecessary time with the highly parallel operation that is vector multiplication. By contrast, single-quadrant analog multiplication can be done using a single resistive element, and many can be performed at once without any loss of generality by simply adding the elements in parallel. In this paradigm, the conductance of the resistive element acts as the weight to be multiplied, the voltage input is taken to be the other operand, and the resulting product is represented as an electrical current flowing toward the negative power rail. [19, 20, 21, 22]

The key to analog multiplication is designing a circuit element whose conductance can be tuned reliably and repeatedly, such that they can be used to load operands

and perform the proposed operation continually as a co-processor, rather than requiring a new application specific integrated circuit for every new operation that must be implemented. Fortunately, it is common in modern applications of linear algebra for one set of operands to change much more slowly than others, so it is not necessary to be able to modulate these tunable resistive elements at high frequencies. Observing this pattern, many designers choose to prioritize non-volatility in their resistive elements of choice, adding the benefits of in-memory computing to their architectures. Common resistive elements include phase-change crystalline cells which change resistance in response to power dissipation[21], memristors which increase or decrease in resistance in response to the bias direction of high voltage programming pulses[23, 19], charge trap transistors which remove charges from the conduction channel of the transistor structure in response to high gate voltage[22], and floating gate MOS transistors which store charges tunneled through the gate oxide in a floating gate to slightly bias the transistor with or without any gate voltage applied[20]. These programmable elements, each with their own advantages and drawbacks, enable the massively parallel vector weighting of voltage inputs in many analog computing paradigms.

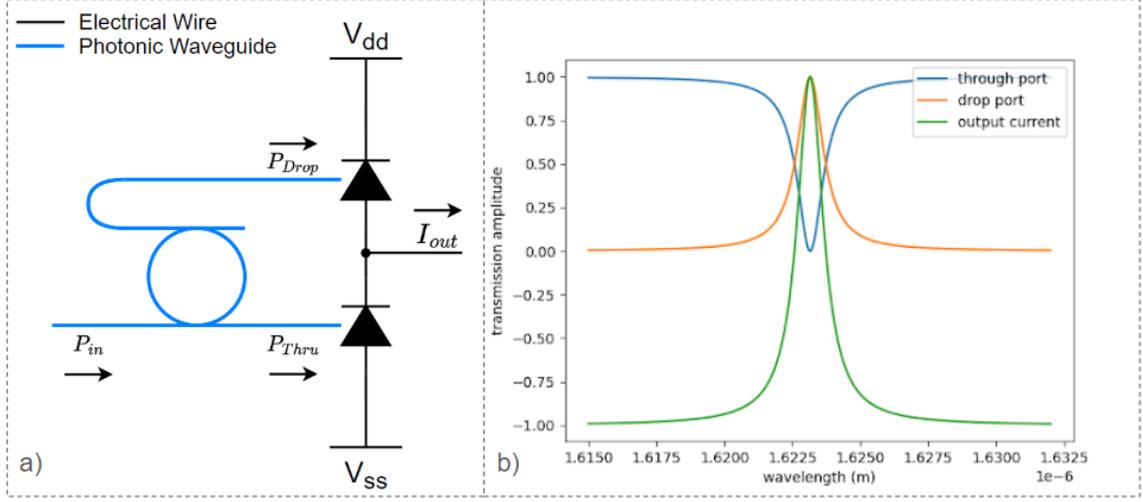
Designing analog vector multipliers with photonic circuits is a bit more in depth, due to the difference in signal medium. Electrical voltage makes an excellent candidate for a medium with which to supply a scalar-vector multiplier with its scalar input because it is commonly safe to assume that voltage remains constant within the same electrical node at low frequencies with small electrical nodes. By contrast, optical power is subject to loss throughout the transmission medium, including both absorption by the medium itself as well as leakage from the surrounding cladding[24]. In addition, as optical power is distributed over the surface plane orthogonal to the direction of travel, any splits in the signal path will cause only a portion of the optical power to follow each signal path. Due to this, it is necessary to fan optical signals out into the desired number of inputs before weighting using optical power splitters that are designed to achieve the desired ratio of

power passed on to each branch. This method of distributing optical power works well with Mach-Zehnder interferometer based weights, as splitters and combiners can easily be converted to Mach-Zehnder elements in integrated photonics by adding an electrically modulated phase shift element. Using triangle decomposition, a mesh of interferometers can be created to perform arbitrary matrix multiplications using optical powers of the same wavelength as inputs [25].

One downside of the Mach-Zehnder interferometer mesh method of weighing optical signals is that the interferometry operation is highly wavelength dependent. Tunable optical delay elements implement a time delay invariant of the input wavelength, meaning that a half-wave shift for one wavelength would equate to a quarter-wave shift with double that wavelength. The attenuation operation of Mach-Zehnder interferometers is based on the phase shift introduced in the test path, so using wavelength division multiplexing with interferometer weights would produce a variety of attenuation constants that could vary anywhere between 1 and 0 based on the wavelength being used to carry the signal[24]. While the wavelength dependence of interferometry is a limiting factor of Mach-Zehnder based matrix processors, there is another technique for photonic signal weighing which actually uses wavelength dependent interferometry to its advantage. Microring resonators are an interferometer structure which guides incoming light into a circular pattern, forcing it to interfere with itself. The geometry of these devices is shown in Figure 1a. In this figure, the resonator has both ports connected to output waveguides, feeding the excitatory and inhibitory inputs of a balanced photodetector. This creates a two-quadrant multiplier which is capable of implementing any weight on the interval $(-1, 1)$, though single quadrant multipliers can be implemented just as easily by terminating either port with an ideal absorber and taking the other as the output.

Figure 1

Geometry, Nomenclature and Behavior of a Microring Resonator



In this operation, the majority of wavelengths experience destructive interference as they traverse the ring, leading to negligible total optical power being transmitted by it. On-resonance wavelengths, however, will travel around the circle in an integer number of wave cycles, fully overlapping themselves and transferring power through the ring to another port while critically destructively interfering with light that has passed by the resonator structure. The equations that describe the port interactions of this structure are as follows

$$T_{Thru} = \frac{(ar)^2 - 2r^2 a \cos(\phi) + r^2}{(ar)^2 - 2r^2 a \cos(\phi) + (r^2 a)^2} \quad (1)$$

$$T_{Drop} = \frac{(1 - r^2)^2 a}{(ar)^2 - 2r^2 a \cos(\phi) + (r^2 a)^2} \quad (2)$$

where T_{Thru} and T_{Drop} are the attenuation coefficients applied to the input signal upon transfer to the Thru and Drop ports respectively, r is the self-coupling coefficient for the

resonator, a is the insertion loss of the structure, and ϕ encodes the ratio of the input wavelength to the on-resonance wavelength of the resonator. This ratio can be expressed as:

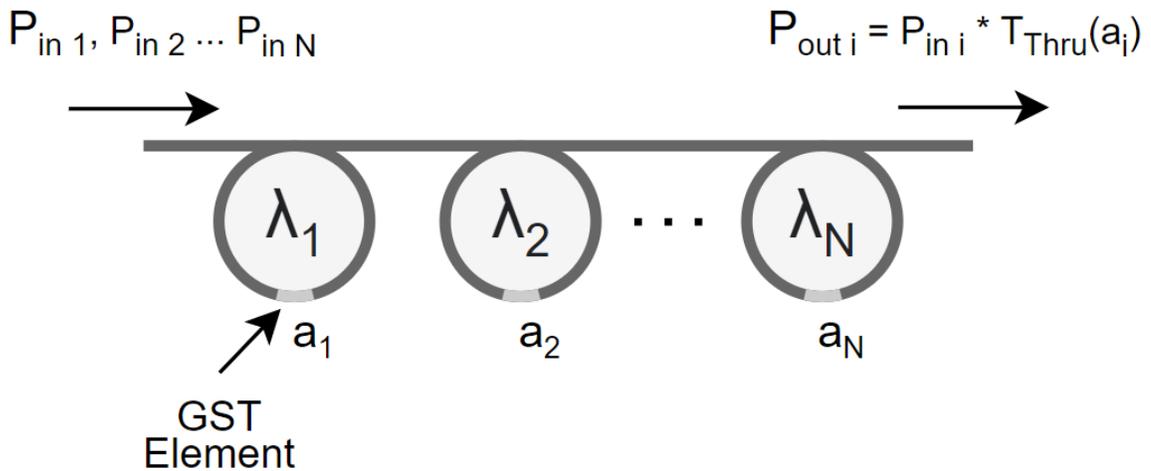
$$\phi = \frac{4\pi^2 d n_{eff}}{\lambda} \quad (3)$$

where d is the diameter of the resonator, n_{eff} is the effective index of refraction between the waveguide and the surrounding cladding, and λ is the wavelength of the input light [26]. The ideal transfer curves with respect to input wavelength for a lossless resonator multiplier in the Thru-Drop configuration are shown in Figure 1b. This peak-valley pattern repeats as wavelength continues in either end, forming a plurality of resonance points: for example, a ring with a resonance point at 1000nm would also have a resonance point at 500nm. Critical to this design is the fact that the thru port transfer constant is close to 1 for wavelengths sufficiently far from the on-resonance wavelength. This allows the resonator to weigh signals of a certain wavelength while leaving others relatively unchanged, enabling a wavelength-division-multiplexed array of inputs to be individually weighed by a bank of microring resonators stacked next to one another with their thru and drop ports cascaded together. This scalability is constrained by the spectral distance between resonance points of a given resonator, as the ring's ability to represent all weight values will be hindered by an overlap between its optical spectral response and that of a neighboring ring[26, 27]. These devices can be tuned to apply a desired weight to a given wavelength by changing the resonance point of the ring. This can be done by either heating the waveguide to change the refractive index, or by modulating it with a P-type-Insulator-N-type diode junction in reverse bias[28, 29]. Alternatively, the resonance point can be left constant, and the resonator can be used only for its wavelength-division-demultiplexing behaviors by adding a tunable attenuator element like a germanium antimony tellurium crystal in the ring structure to variably decrease the on-resonance wavelength interference[27]. This allows for a higher density of rings in the weight bank, though this technique tends to suffer

from a loss of precision as phase change crystals tend to be imprecise to program. This type of architecture is illustrated in Figure 2.

Figure 2

Arrangement of Phase-Change Material Based Ring Resonator Weight Bank[27, 30]



1.2.2 Kirchoff Current Summation

Along with multiplication, matrix processors must be able to perform many-variable additions. This is the "accumulate" operation referred to by the phrase "multiply-and-accumulate" used to describe a vector-vector multiplication operation. This is an operation that can become painfully sequential on a CPU, as each addition result must be stored back in memory before being recalled as one of the operands in the next. Digital hardware accelerators can make this operation somewhat more efficient by funneling the additions down two by two in parallel, but the process remains somewhat sequential due to the fact that addition is not easily implemented with more than two operands at once in digital. This limitation makes analog an especially appealing alternative for many-variable accumulation, as

Kirchoff's current law provides a mechanism for fully parallelized analog addition[31].

In both the electronic and photonic cases discussed in the previous section, output signals from analog multiplications were represented as current signals, either produced from a voltage being placed across a tunable resistive element, or from a photodiode experiencing stimulated absorption. With the desired products encoded in this medium, accumulating them is as simple as forcing them to use the same path to ground. Once the currents are accumulated in the same path, they can be duplicated and scaled using current mirror structures[32], or converted back to voltage signals using transimpedance amplifiers where they can be buffered or sampled for further use elsewhere[33, 34]. For neural network operations which require a nonlinearity, output currents can also be modified using diodes to imitate rectified linear units, winner-take-all transistor circuits to perform a comparison operation similar to softmax[35], or microring modulators to imitate a sigmoid nonlinearity[36].

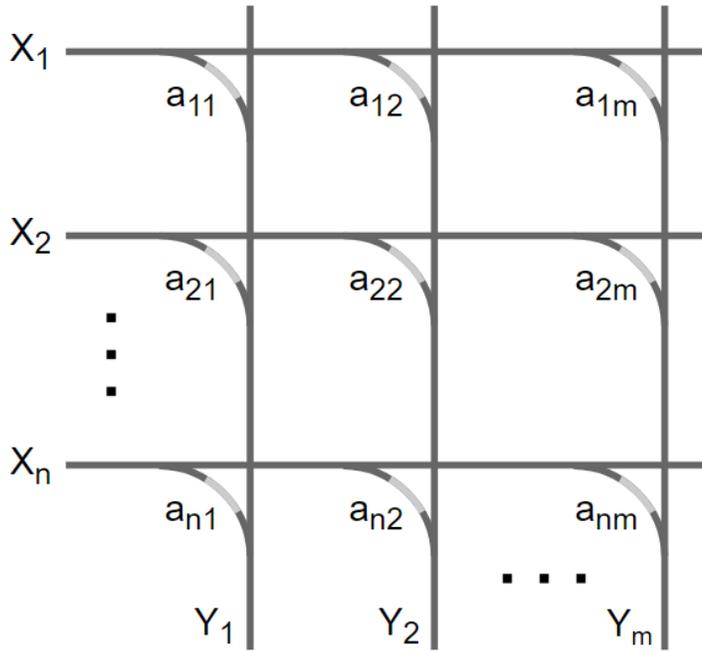
Using balanced photodetectors to generate currents from wavelength-division multiplexed optical powers automatically produces the accumulated output for all of the inputs supplied. Unlike electrical matrix processor output currents, which tend to be electrically short-wave, these currents tend to be in higher frequency bands, meaning they are subject to reflected power if their load impedance is mismatched. For this reason, it is common to terminate current nodes in transimpedance amplifiers with characteristic input impedances regardless of any future operations that need to be performed on the current. If the number of inputs represented by a single dot product engine is sufficient to perform the desired operation, no further modification is necessary, but if a larger operation must be performed, parallel engines can be used to create intermediate output voltages which can be Kirchoff-summed using resistors of equal value.[26]

1.2.3 Common Architectures

Given the convenience of parallel weighting and current summation for voltage signals, a naturally emerging architecture that leverages these techniques is the crossbar vector-matrix-multiplier circuit[37, 23]. This architecture implements parallel vector multiplications by connecting the input voltages to the same output node using tunable resistive elements, as illustrated in Figure 3. In cases of two-terminal devices like memristors, it is common to attempt integrating them vertically, allowing them to act as tunably resistive vias to connect neighboring layers of metal and reducing the footprint limit of the architecture to the minimum spacing allowed by the design rules of the process. This architecture also closely resembles a routing junction, inspiring modern analog designers to implement matrix-vector multiplier circuits in the routing of a larger programmable analog array[38]. Crossbar architectures have also been implemented in photonics with tunable optical attenuators[39], though they are less common due to the lower precision of fiber-embedded tunable attenuators and the difficulty of integrating precision power splitters with a ratio other than 1:2.

Figure 3

Illustration of a Crossbar Matrix Vector Multiplier[37, 23, 39]

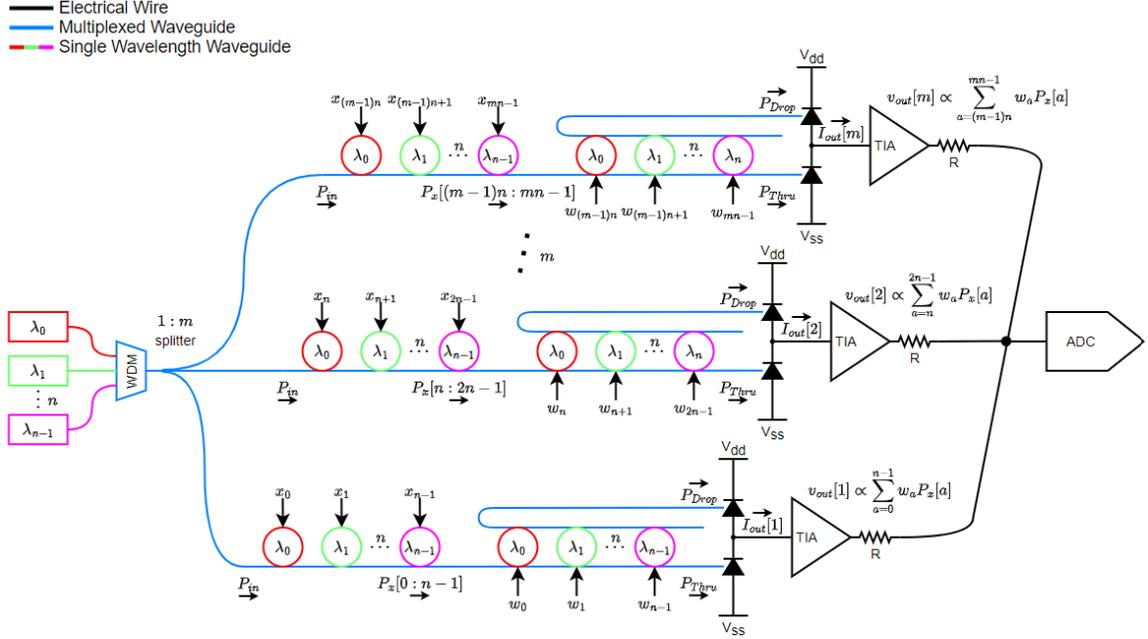


With the benefits of wavelength division multiplexing, photonic computers can encode all of the desired inputs in the same waveguide, rather than requiring separate wires for each input. Multiplexed inputs are demultiplexed and weighed by ring resonators tuned to each wavelength, summed by Kirchoff's law in a balanced photodetector circuit, and converted back to voltage for sampling and/or further processing in analog[40]. In this paradigm, physically separated transmission media are used to parallelize the processing of a given input or set of inputs, rather than to add more inputs to be processed in parallel[29]. This method of parallel processing can be used to implement full matrix vector multipliers, though it is especially adept at performing 2d convolution operations as the input data can change much more quickly than the weights[26, 28]. A common architecture that performs this operation with the support of digital electronic control circuitry is depicted in Figure 4. This architecture is referred to by its inventors as a Digital Electronic and Analog Photonic,

or DEAP, tensor processor.

Figure 4

Schematic of a Digital Electronic and Analog Photonic Tensor Processor[26], Performing an n -by- m Vector-Vector Multiplication



1.3 Analog Technologies in Industry

Due to the emergence of neural networks and artificial intelligence in industry, analog matrix processors have found a great deal of success in accelerating machine learning algorithms at the edge. One of the first companies to find success in industry with this technology is Mythic AI[41]. This company uses charge trap transistors to both store neural network weights and apply them to voltage inputs during runtime, sampling the resulting signals with analog-to-digital converters to store the results back into memory to avoid noise cascading. Their tensor processors act as co-processors for computer systems,

similar to modern GPUs. The improvements their products provide to the speed and power efficiency of neural network inference allows for a variety of methods to be available at the edge, and their recent ventures have been focused on improving the viability of generative AI algorithms.

Okika technologies takes the concept of analog function accelerators further by implementing a full analog instruction set which can be used to make a datapath[42]. Inspired by the success of field-programmable gate arrays, these circuits are commonly referred to as field-programmable analog arrays. The available analog function block set includes blocks as abstract as arbitrary frequency filters, as well as those as low level as single transistors. This allows designers to effectively create custom analog integrated circuits using only software, and reconfigure them as necessary without having to fabricate a new design. The current state of the art field-programmable analog arrays still require the use of sampling and reproduction to allow functions to generalize to non-sequential data paths, but their voltage storage cells are implemented in analog, allowing for some of the numerical stability of continuous processing to be preserved. Their currently available products include a stand-alone development board and a Raspberry Pi hat, and ongoing efforts are being made to bring full mixed-signal systems-on-chips to the consumer market.

Another company that has found industry success in analog computer design is Aspinity. Unlike the previous two companies, Aspinity's AML100 analog processor implements a fully analog datapath, including feature extractors, multi-layer perceptrons, and winner-take-all classifiers, only resorting to digital circuitry to provide control signals and programming interfaces. This grants the processor the full range of benefits associated with analog computing, including low latency, high speed, numerical stability, and, thanks to the sub-threshold operation of their MOSFET devices, a power demand on the order of microamps. This drastically lower power consumption makes the AML100 ideal for sensor systems that are always on, such as voice detection and recognition[43] as well as object detection with sonar sensors in vehicles[44].

Due to the relative immaturity of integrated photonic computing technologies compared to electronic computing, most advanced photonic tensor processors are relegated to research ventures and military applications; however, some companies have implemented photonic computing technologies during research and development with the intent to leverage them commercially in the future. The most promising company in the business of analog photonic computing is IBM, as they have made several strides in the field of in-memory photonic matrix processing with phase-change material cells[45]. These emerging circuits show promise for a variety of applications, especially as general purpose photonic integrated circuits rise into prevalence in industry. While companies like IBM research emerging applications of photonic integrated circuits, manufacturers like GlobalFoundries are expanding their stake in integrated photonics by unveiling new processes for chip designers to use[46]. Unveiled in 2022, GF's "Fotonix" silicon photonics platform targets a wide variety of already prevalent applications, such as telecommunications and optical fiber interconnect, while enabling external designers to pursue the emerging methods that make photonics such a promising computing paradigm.

1.4 Advantages of Photonics

Despite its relative immaturity, integrated photonics presents a number of unique advantages that differentiate the technology from electronic analog computing. Primary among these advantages is the high modulation speed possible with integrated photonics. With modulator bandwidths commonly spanning into the gigahertz range[29], high frequency signals like LiDAR scanners[9] and fiber communication signals[11] can be processed in real time, and convolutional neural network models with large input sizes and relatively small weight tensors can be processed much more quickly than electrical analog acceleration can manage[26]. These signals also propagate through integrated optical fibers much faster than electricity moves through wires[24], meaning the latency of these circuits is extremely low. This advantage allows signal processing circuits to be placed

directly in line with optical fiber without disrupting the speed of signal flow through the communication line at all[11].

Another benefit of photonic circuitry over electronic is the simplicity and size of designs for high speed applications. Once electrical traces become long relative to the signal wavelength, assumptions of constant node voltage are violated, power is reflected back toward potentially sensitive signal sources if impedances are not matched, and neighboring traces begin to interfere with one another due to capacitive coupling[4, 16]. By contrast, neighboring optical waveguides do not interfere with one another so long as the criteria of total internal reflection are satisfied, and reflected power can be avoided by terminating unused waveguides with optical absorbers to dissipate any received optical power[24]. In addition, wavelength division multiplexing allows integrated photonic circuits to route a large number of signals through the same physical medium, drastically reducing the footprint of interconnect circuits between functions on a chip or chips on a circuit board. While electronics are generally more scalable in terms of compute density with respect to area due to the need of photonics to satisfy total internal reflection[24, 27, 26, 19], photonics provides unique advantages that makes smaller, more high performing designs possible.

Power consumption is a unique factor to analyze with photonic computing, because it achieves computing energy efficiency through different means than electronic computing. Electronic analog computers tend to increase energy efficiency by reducing power consumed during operation, thus performing the same operations in a comparable amount of time to a digital computer while consuming less energy per second[19, 41, 43]. Photonic circuits do tend to consume less power overall compared to digital computers[26], but they are still out-performed by ultra-low power electrical analog designs. Despite this, photonic circuits can claim a very low amount of energy required to perform a given operation because the operation occurs so much more quickly[40]. While this means that photonics is not very suitable for edge computing the way analog electronics is, server applications may find photonics to be more efficient in terms of energy per task, especially for tasks that

require large amounts of data to be processed at once.

Chapter 2

Interference Cancellation

The material in this chapter is used with permission from *Frontiers in Optics / Laser Science* and *IEEE Photonics Technology Letters*, previously published in *Conference [7]*, and *Journal [8]*. ©2023 IEEE

2.1 Introduction

Light detection and Ranging (LiDAR) is a powerful technology for computer vision and the modeling of 3D space, associated with many emerging technologies like autonomous driving [14]. It uses the constant and measurable travel speed of light to generate collections of points in space, referred to as "point clouds," which can be used to infer the presence of different objects in said sample space. The most elegant and cost-effective way to take this measurement is to use short optical pulses, and measure the time between rising edges, which can be detected using a thresholding mechanism. This method of measuring distance is primarily referred to as time-of-flight (ToF) LiDAR. While this solution works well in an optically isolated environment, problems arise when the quality of the signal is corrupted by external optical interference. This interference can originate from the ambient electro-optical noise in the environment [33], the emerging presence of free-space optical (FSO) communication links [47, 13], other LiDAR sensors in the same environment [48], or even from adversarial actors intending to cause harm with the interference they introduce [49, 50]. The presence of this interference forces LiDAR manufacturers to increase their transmission power to provide an acceptable Signal-To-Interference Ratio (SIR) in exchange for more power draw and cost of manufacture. This trade-off limits the viability of LiDAR sensors in technologies for which they are otherwise very well-suited, and thus warrants exploration into methods of removing interference from LiDAR signals.

The current methods for mitigating the effects of interference on LiDAR signals tend to fall into one of three categories: pattern recognition, digital signal processing, and physical layer signal processing. Pattern recognition practices in LiDAR signal processing can create noise immunity by way of making a uniquely identifiable signal to use in place of a pulse, and then recognizing that signal amongst the interference that it is measured with [51]. While this does create a system that is more robust to noise, it does add latency to the system, and is still not immune to noise-related errors in measurement. It also does nothing to improve the sensor's vulnerability to adversarial attacks, which can be designed to mimic any LiDAR signal provided to them without any circuit reconstruction necessary [49]. With adversarial attacks left intact, the only action that can be taken to prevent their influence is to detect the data as adversarial and discard it [50]. Digital signal processing techniques focus more on removing noise from the signal of interest (SOI), by way of techniques like wavelet domain spatial filtering [52]. These techniques, while effective, tend to be computationally demanding, adding an amount of latency to the system that may be incompatible with real-time applications. Additionally, because digital signal processing techniques require the exact time-domain behavior of the processed signals to be captured, systems that leverage them are required to have analog-to-digital converters (ADC's) that adhere to the LiDAR signal's Nyquist sampling limit. This enforces a trade-off between system price and spatial resolution, as the precision of ToF LiDAR and RaDAR signals is proportional to their pulse width [53]. Physical layer signal processing aims to mitigate the shortcomings of its digital counterpart by implementing the denoising processes in analog rather than digital. The removal of the digitization delay and the increased speed of analog computation results in a negligible addition of latency over the noisy system, making the ideal choice for time sensitive applications like autonomous driving. Photonic analog circuitry is particularly effective for these applications, as it can support throughputs similar to that of optical fiber communication cable, and is versatile enough to handle a wide variety of tasks, as demonstrated by [54]. Despite this advantage, research on physical

layer improvements to LiDAR systems has mostly focused on removing noise originating from the LiDAR system itself [33]. This is likely due to the difficulty of implementing more complex or time-dependent denoising techniques in analog.

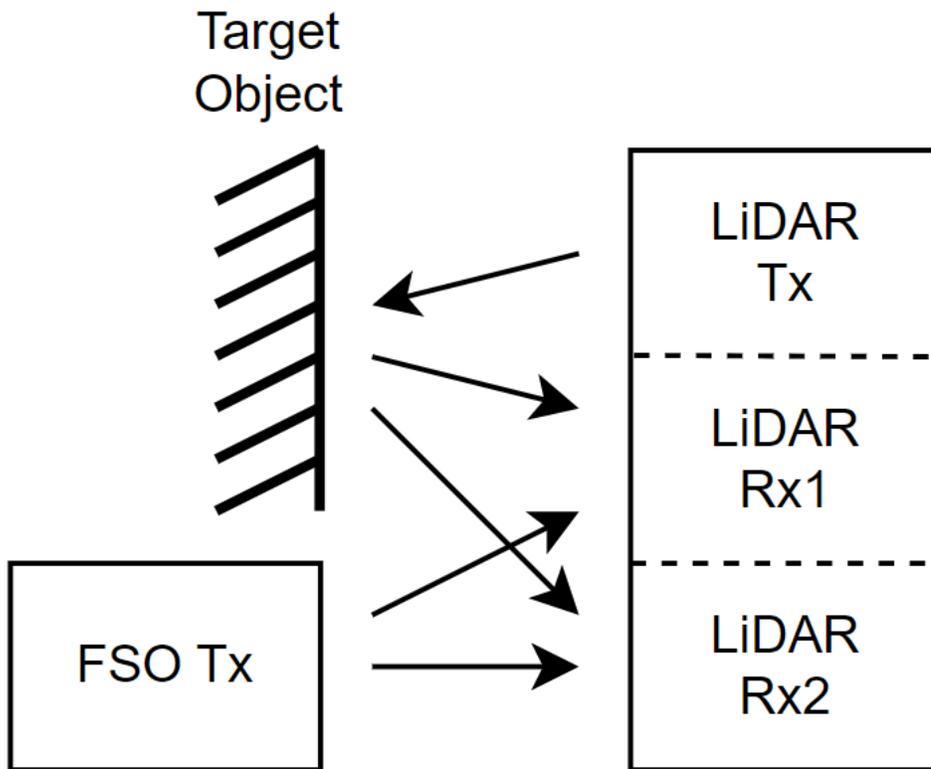
Blind source separation is a method by which a multi-input, multi-output (MIMO) system can differentiate signals from one another without the need for time or frequency division multiplexing, and without any prior knowledge of the incoming signals' characteristics. This is commonly done by way of using the second and fourth order moments of the signals with respect to one another. This technique generates a de-mixing matrix, which can be multiplied by the received signal vector to recover the original signals as they were prior to combination. This method of de-mixing signals is especially elegant, as the linear transform that separates the signals can be easily implemented in the physical layer or built into an application-specific integrated circuit, as demonstrated by [12] and [55] respectively. This technique has been utilized heavily in the field of optical communications, as it adds next to no latency, and requires sampling of only a small portion of the mixed signals for the derivation of the de-mixing matrix in the case of static systems [47]. Additionally, because only statistical information about the signal amplitudes is used to calculate the de-mixing matrices, this technique can be done using sub-Nyquist sampling circuits. This enables cost-effective noise cancellation in frequency bands for which digital signal processing techniques would not be feasible [56].

In this paper, we propose a multi-input LiDAR receiver that is capable of removing interference signals from one another through the use of photonic BSS. The receiver uses a pair of spatially separated collimator lenses to generate a received signal vector out of the combination of the LiDAR pulses and all other optical signals. By spatially separating the lenses, the receiver increases the likelihood of separability of the incoming signals. The separation of the signals is implemented using a balanced photodetector and optical tunable attenuators and delays, adding virtually no latency to the measurement of distances while improving the SIR of the measurements drastically. This system is also compatible

with sub-Nyquist sampling circuitry, removing the trade-off between spatial resolution and the cost of the LiDAR receiver. The organization and use case of the proposed system is illustrated in Figure 5.

Figure 5

Application Scenario of the Proposed System



Note. ASE: Amplified Spontaneous Emission source, BERT: Bit Error Rate Tester, IM: Intensity Modulator, EDFA; Erbium-Doped Fiber Amplifier, DFB: Distributed Feedback Laser, TA: Tunable Attenuator, TD: Tunable Delay, BPD: Bipolar Photodiode.

2.2 Principle and System Setup

2.2.1 Principle: Photonic Blind Source Separation

The mixing of two signals in a MIMO system can be described as a linear transform that adds a scaled version of each of the sent signals together to create each of the received signals. More formally,

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \quad (4)$$

where x_1 and x_2 are received signals, s_1 and s_2 are transmitted signals, and the a parameters are functions with respect to frequency that describe the transmission coefficients of each signal in the system as it propagates toward one of the receiver ports. These parameters are often referred to in vector/matrix notation as \mathbf{X} , \mathbf{S} and \mathbf{A} . The "de-mixing" of these signals in this model can then be simplified down to performing the inverse of this linear transform, or more formally,

$$\mathbf{S} = \mathbf{A}^{-1}\mathbf{X} \quad (5)$$

where the inverse matrix \mathbf{A}^{-1} is known as the "de-mixing matrix" for the system. These matrices can be found using the second and fourth order moment of the received signals with respect to one another, as described in [12]. This yields the best results if the receivers are sufficiently spatially separated, such that $a_{11} \neq a_{21}$ and $a_{12} \neq a_{22}$. This strategy of statistical measurement is compatible with any two arbitrary signals, but ToF LiDAR signals are a special case that makes this procedure much easier. The following is a general

equation for the inverse of a 2x2 matrix.

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{21}a_{12}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} \quad (6)$$

This form shows that a scaled version of s_1 can be recovered using only amplitude information about s_2 , which can be measured directly while no LiDAR pulse is being transmitted. Because the sensor has control over the behavior of s_1 and the intent of this system is to discard s_2 after the statistical measurements are complete, measuring a_{12} and a_{21} is easily possible and sufficient to separate the desired signals without requiring any time-consuming statistical analysis. To de-mix the signals physically, one can simply align the two signals in time and implement the described de-mixing matrix somewhere along the signal path. This can easily be done with fiber-bound optical signals using tunable optical delays and attenuators without changing the throughput of the system at all.

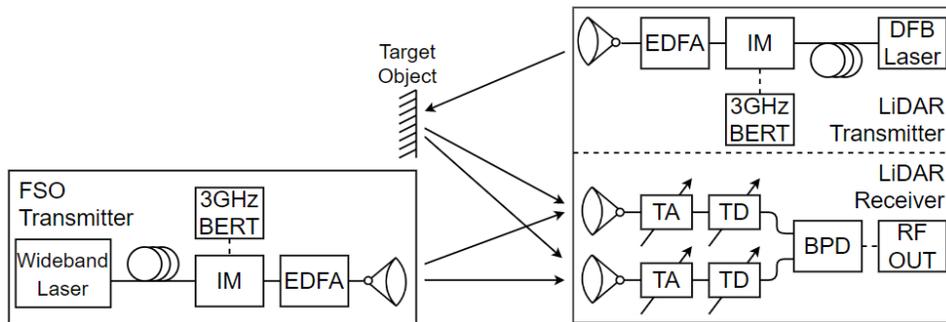
2.2.2 *Experimental Setup*

Figure 6 shows the experimental setup used to test this system. The LiDAR transmitter is made up of a Distributed Feedback (DFB) laser tuned to 1550nm, a polarization controller, Erbium-Doped Fiber Amplifier (EDFA), and an optical intensity modulator and a collimator lens that decouples the light from fiber. The intensity modulator is driven by a 3GHz Bit-Error Rate tester (BERT) that was configured to generate a pulse with a width of 0.33ns and a repetition rate of 3MHz to simulate a LiDAR sensor with a distance resolution of 48mm. The modulated light is decoupled at an optical power of 16.7dBm and directed towards a mirror that simulates the target object for the LiDAR sensor, and reflected back to the dual receiver for use in de-mixing. The interference generator is set up similarly, with the only change being the light source. For this experiment, a wideband laser was chosen to

generate interference signals, in order to simulate the broad spectral contents of signals that tend to interfere with LiDAR sensors. The optical power of the interference transmitter was measured to be 16.5dBm. The dual receiver is made up of a pair of paths consisting of parallel collimator lenses, tunable attenuators, and tunable delays, terminating in one of the two inputs of a balanced photodetector. The tunable delays and attenuators in these paths can be controlled by the signal processing computer automatically, allowing for the implementation of one of the multiply-accumulate operations that make up the de-mixing transform. This enables the sensor to separate the LiDAR pulses from the interference signals before they are used to infer distance. The use of tunable attenuators to implement the transform coefficients also provides immunity to photodiode saturation by reducing saturation-level optical signals to a usable amplitude prior to analysis.

Figure 6

System Setup for Testing the Interference Cancelling LiDAR Receiver

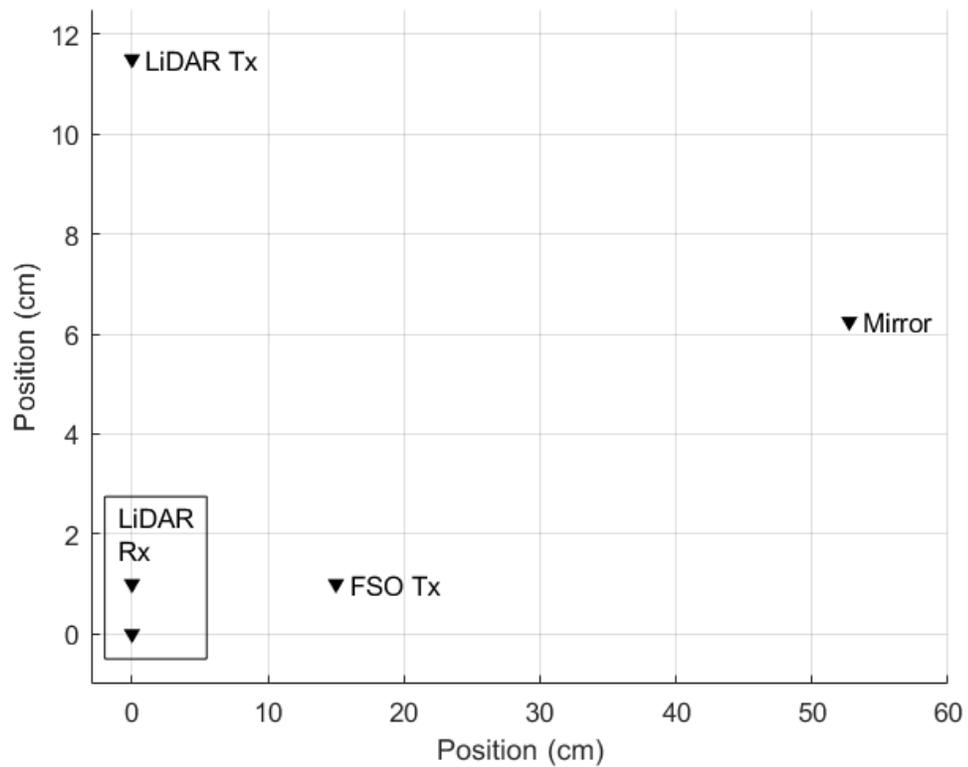


Note. ASE: Amplified Spontaneous Emission source, BERT: Bit Error Rate Tester, IM: Intensity Modulator, EDFA; Erbium-Doped Fiber Amplifier, DFB: Distributed Feedback Laser, TA: Tunable Attenuator, TD: Tunable Delay, BPD: Bipolar Photodiode.

For the purposes of this analysis, the separation was computed in software. To do this, the collimator lenses and mirror in the setup were positioned with the spacings shown in Figure 7. The one-way distance to the target is 53cm. The dual receiver ports were positioned close together and aimed parallel to one another, both to simulate the size constraints of building a sensor, as well as to ensure that drastically different optical powers are received by the reflection from the mirror. The result of this design decision is a mixing matrix that is easily separable via the statistical blind-source-separation system detailed in [12]. The rest of the positions in the diagram were fairly arbitrary.

Figure 7

Position of Optical Lenses and Mirrors in the Experimental Setup



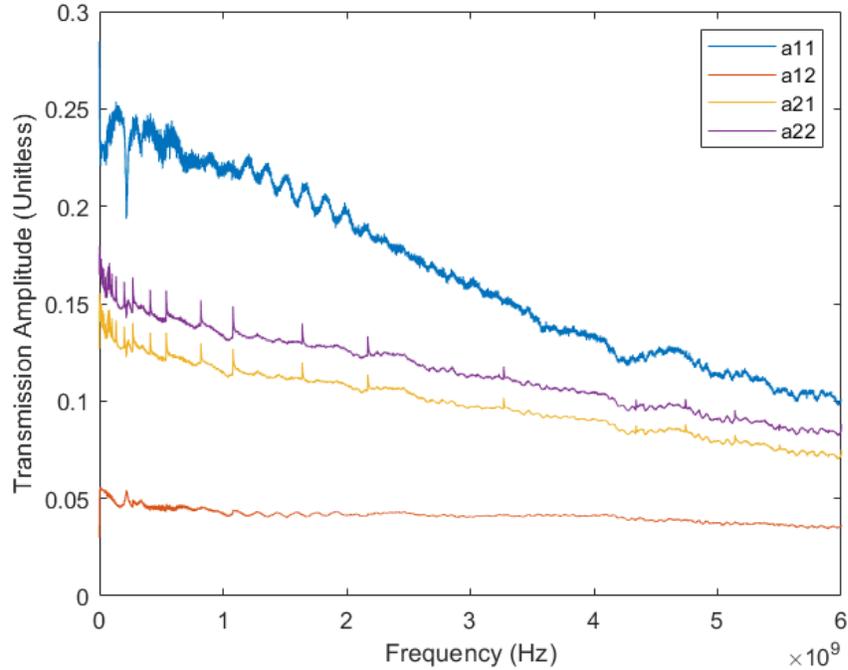
In its correct position, Receiver 1 received -15.8dBm of optical power from the LiDAR transmitter and -17.8dBm from the interference transmitter. Receiver 2 received -21.8dBm from the LiDAR transmitter and -17.4dBm from the interference transmitter. Once the lenses were positioned, the receiver paths were replaced with photodetectors, and the transmission spectra of the four paths were measured over the bandwidth from DC to 6GHz with a precision of 3 significant figures. This would allow simulated data to be given a realistic frequency response that could be used to perform the simulated BSS system without the need to repeat experiments or change out equipment. These spectra could also be scaled as necessary in order to simulate different transmission distances and scenarios

2.3 Results and Discussion

Measuring the four transmission paths resulted in the spectra shown in Figure 8. This figure illustrates the distinct differences between each spectrum. While the interference signal spectra are similar in amplitude and shape, the LiDAR spectra are drastically different in amplitude due to the directionality of the parallel receivers. The interference signal spectra are also differently shaped from the LiDAR spectra, because of the difference in spectral contents of the laser sources used to carry the signals. These differences increase the likelihood of separability of the incoming signals.

Figure 8

Transmission Spectrum for Each Channel

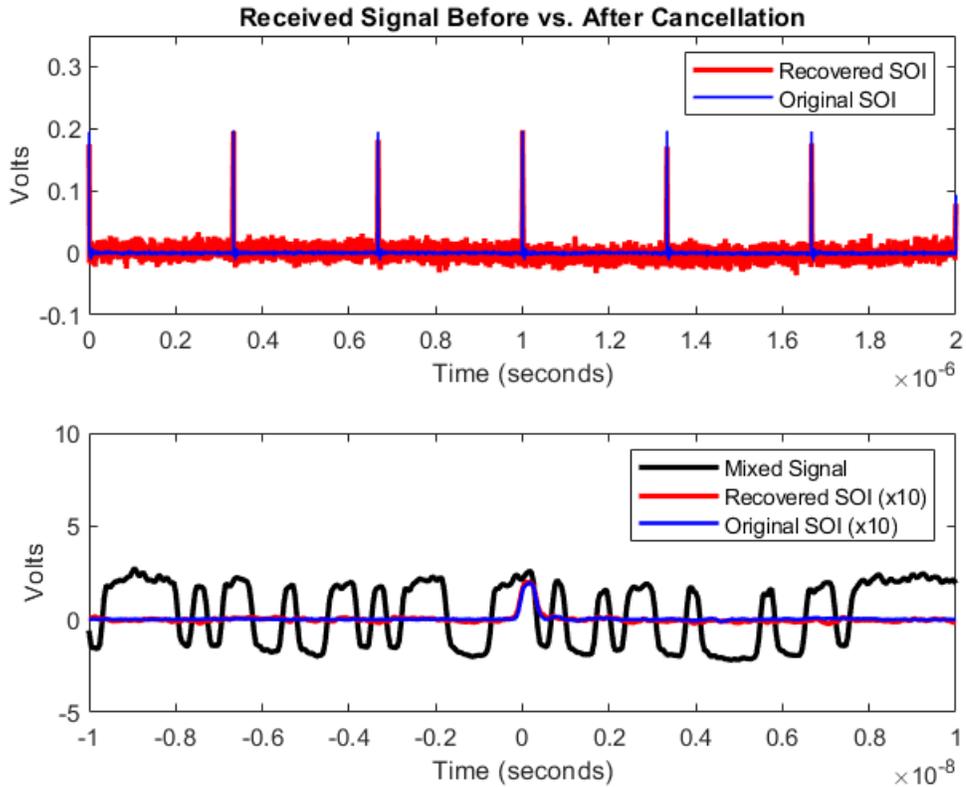


To test the ability of the BSS system to de-noise the LiDAR signal, two different types of interference were tested. For the case of interference from wide-band LiDAR sensors, adversarial devices and/or free-space optical communication devices, a pseudo-random binary sequence (PRBS) was used to simulate interference. To ensure overlap of the signals' spectra, the bit rate of the PRBS signal needed to have a similar bit width to the pulse width of the LiDAR signal; however, the widths cannot be the same, or the two signals will be coherent and the separation of the signals will be trivial. To satisfy both of these conditions, the bit rate of the PRBS signal was chosen to be 2.9GBPS. Once the signals were sampled, they were rescaled with the spectra sampled from the system, put through the BSS system, and rescaled such that the peak amplitudes of the SOI and post-cancellation signal were the same. To demonstrate that this system is compatible with

all types of LiDAR, the method proposed by [12] was used to compute the de-mixing transform rather than directly measuring the interference amplitudes. The resulting signals before and after separation are shown with the noisy signal in Figure 9. This figure shows the signal after cancellation very closely resembling the ideal SOI. The comparison plot with the noisy signal shows that, while pulses could not be detected in the mixed signal using a threshold, they can easily be detected this way in the signal after cancellation. This separation was calculated to give an average interference rejection ratio of 41.7dBm, increasing the SIR by 46.5dB and resulting in a post-cancellation SIR of -2.7dB. Time jitter analysis shows that the distance measurement error caused by this residual interference is several orders of magnitude below the simulated precision of the sensor.

Figure 9

Results of De-Mixing With a PRBS Interference Signal

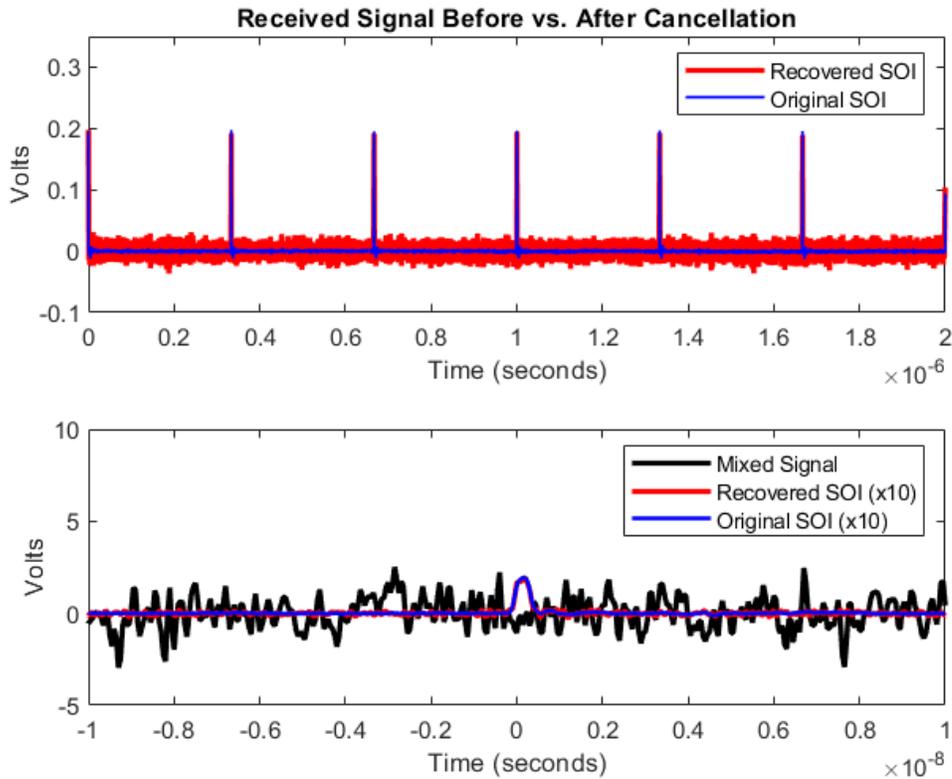


Note. (a) Signal of interest before mixing and after de-mixing. (b) Original, mixed and de-mixed signals, with SOI's amplified by 10dB for visibility.

To test the case of environmental interference, a Gaussian distributed noise signal with a 20GHz sample rate was chosen as the interference signal, and the same test was repeated. The results of this test are shown in Figure 10. As with the previous test, the signals were able to be separated sufficiently to allow for threshold detection of the LiDAR pulses. This separation yielded an average interference rejection ratio of 41.4dB, increasing the SIR by 42.4dB and resulting in a post-cancellation SIR of -2.4dB. This separation also yielded negligible distance measurement errors due to residual interference.

Figure 10

Results of De-Mixing With a Gaussian Noise Interference Signal



Note. (a) Signal of interest before mixing and after de-mixing. (b) Original, mixed and de-mixed signals, with SOI's amplified by 10dB for visibility.

The interference rejection ratios of these separations are mainly limited by the frequency responses used to mix the signals. The BSS algorithm can only provide scalars for de-mixing coefficients, meaning that dissimilar mixing spectra for the same source signal will result in residual interference in the frequency components that cannot be perfectly matched. In physical implementations, noise will be accumulated that cannot be cancelled, further decreasing the maximum rejection ratio and corrupting measurements that could lead to more inaccurate de-mixing matrix estimation. These limitations scale in severity with the condition number of the mixing matrix, as demonstrated by [47], though this sys-

tem is designed to minimize condition number as much as possible. Hardware experiments using this algorithm, such as the one detailed in [10], report achievable interference rejection ratios in the order of 30dB.

2.4 Conclusion

We propose a multi-input LiDAR receiver that is capable of cancelling optical interference. This system uses blind source separation to remove interference signals from LiDAR pulses. The proposed system implements the interference cancellation calculations in photonic analog circuitry, meaning that the system does not appreciably reduce the sensor's throughput. This system was tested against wideband interference signals, simulating both digital and analog interference signals, and showing the potential to reject as much as 40dB of interference. This cancellation allows for threshold detection in scenarios where it would not be possible otherwise.

Chapter 3

Variational Neural Networks

The material in this chapter is used with permission from *Frontiers in Optics / Laser Science* from content awaiting review.

3.1 Introduction

Neural networks are universal function approximators that can be optimized to accomplish a variety of tasks using training data sampled from the approximated data distribution. While it is common to assume that this approximation will hold during deployment if proper care is taken during training, Out-of-Distribution (OoD) samples and adversarial attacks have been shown to significantly degrade the performance of models that perform well on the controlled validation datasets that are used to evaluate them [57]. Bayesian Neural Networks (BNNs) present a solution to this problem by way of introducing Gaussian-distributed uncertainty in the weight tensors of the models, which is trained to model the epistemic uncertainty of the dataset through a modified form of backpropagation. [58]. In addition to preventing overfitting in a manner comparable to Monte-Carlo Dropout [58, 59, 60], this method provides an estimation of the quality of the prediction in the form of the variance of the output distribution [61]. This technique also expands the applicability of a given model, granting the ability to prune redundant weights from the model tensor for more time sensitive applications[62], as well as the ability to highlight highly sensitive features of a decision for applications where explainability is necessary[63].

One factor limiting the prevalence of BNNs in machine learning is the difficulty involved with propagating variances through layers of the network. While matrix-vector multiplications can be solved empirically using random variable algebra, nonlinearities in the network are not trivial to propagate Gaussian distributions through. Solving this

problem robustly requires the propagation of discrete points, either sampled from the distribution[59] or selected via the unscented transform[61], through the nonlinearities at a rate of $2N+1$, where N is the number of elements in the input vector. This process can be approximated with random variable algebra using first-order Taylor expansions[64], but these approximations break down with the use of highly nonlinear activation functions[59]. Even in the best case, execution time is at least doubled that of deterministic neural networks, which already struggle to meet the needs of real-time applications like autonomous driving[6], communication signal processing[11], and robotic control [15].

Analog neural accelerators are an emerging technology dedicated to improving the throughput and power efficiency of neural network computation. They function by broadcasting a vector of inputs, encoded as some type of electrical or optical signal, across an array of tunable attenuators to perform Hadamard products, the results of which are summed in the circuit to produce a full matrix-vector-multiplication. Different circuit design paradigms have emerged in recent years, each with its own set of advantages, drawbacks, architectures, and level of maturity in the current state of the art. Electronic implementations tend to use variable resistance to perform signal weighting, converting voltage-encoded inputs to currents for summation and sampling. Memristors are a common and relatively successful variable resistor used for this task, inspiring architectures like Intel's Loihi chip[19]. Other common variable resistors include electronic phase change material resistors[21] and floating gate transistors[20], with the latter even seeing some success in industry applications[43]. Electronic implementations boast high parallelizability, high precision, compatibility with digital electronics, and extremely low power in comparison to digital computational circuitry. Photonic implementations encode inputs as optical amplitudes, which are weighted with variable optical amplifiers or attenuators, and typically summed using a broad-wavelength bipolar photodetector. Most implementations use microring resonators (MRRs) as the variable attenuator[40, 26, 28], though earlier implementations that focus on temporal pattern recognition utilized silicon optical amplifiers[40, 11]

and emerging technologies have begun utilizing photonic phase change material attenuators or delays[27, 65]. Photonic implementations boast higher throughputs and lower latency than their electrical counterparts, without requiring impedance matching as inference rate increases into the RF bands[40, 26], while sacrificing parallelizability and computational precision[28, 27]. Both paradigms suffer from the stochasticity of physical layer computing, commonly leading researchers to report model performance relative to the deterministic version of a given model when presenting on a new innovation in analog neural acceleration.

While inconvenient for approximating deterministic neural networks, the inherent stochasticity of analog computational components shows promise for implementing BNNs. Because analog noise exists in continuous time and across the frequency spectrum, it is possible to perform an approximated time-average to measure the variability of a hidden state in the same amount of time it would take to compute the hidden state while assuming determinism. To the knowledge of the authors, only one attempt to accelerate a BNN with analog circuitry has been made[66], but the authors achieve this function by using two separate photonic matrix multipliers, computing variance with zero mean and computing mean while assuming zero variance. This technique still leaves the circuit vulnerable to computational inaccuracies in the mean calculation, and leaves the variance as a simulation of what the variability of the neurons would look like if they were treated as stochastic. In addition, the activation functions are not considered in the sampling of these signals, leaving the most computationally intensive parts of the model to digital circuitry. This work proposes a novel detector for analog bayesian neurons which separates the central tendency of a computational signal from its variance and detects both from the same circuit. Because the distribution is reparametrized from the same signal, this method is compatible with analog activation functions, and can be used after an arbitrary number of layers to consider non-diagonal elements of the covariance matrix of a hidden layer when sampling the diagonal elements of the output covariance matrix. A method of injecting

programmable noises into the circuit is then proposed, and tested in a simulated photonic neural network using the PyTorch deep learning framework[67] to illustrate the functions of the method.

3.2 Principle and System Setup

3.2.1 Principle: Variational Density Propagation

Instead of a tensor of constants, BNNs interpret the parameter space as a tensor of random variables with a prior distribution $p(\mathcal{W})$. Observing the training data $\mathcal{D} = \{X^{(i)}, y^{(i)}\}$ enables us to infer the posterior distribution $p(\mathcal{W}|\mathcal{D})$. With this distribution, we can predict the distribution of any new datapoint \tilde{X} as the following

$$p(\hat{y}|\mathbf{X}, \mathcal{D}) = \int p(\hat{y}|\mathbf{X}, \mathcal{W})p(\mathcal{W}|\mathcal{D})d\mathcal{W} \quad (7)$$

Where the mean of this distribution represents the prediction of the network, and the variance represents the epistemic uncertainty of the model’s prediction of the datapoint. Unfortunately, this integral is not feasible to compute for high-dimensional nonlinear models like deep neural networks, so the results can instead be approximated.

Variational inference is a method of training a parametrized variational approximation $q_{\theta}(\mathcal{W})$ of the posterior distribution. This training method involves minimizing the Kulback-Leibler (KL) divergence between the variational approximation and the true posterior. This is done through the Evidence Lower Bound (ELBO) loss, which is a refactoring of the full KL divergence using bayes rule. The quantity can be calculated as

$$\mathcal{L} = -\mathbb{E}_{q_{\theta}(\mathcal{W})} [\log(p(\mathcal{D}|\mathcal{W}))] + KL[p(\mathcal{W}), q_{\theta}(\mathcal{W})] \quad (8)$$

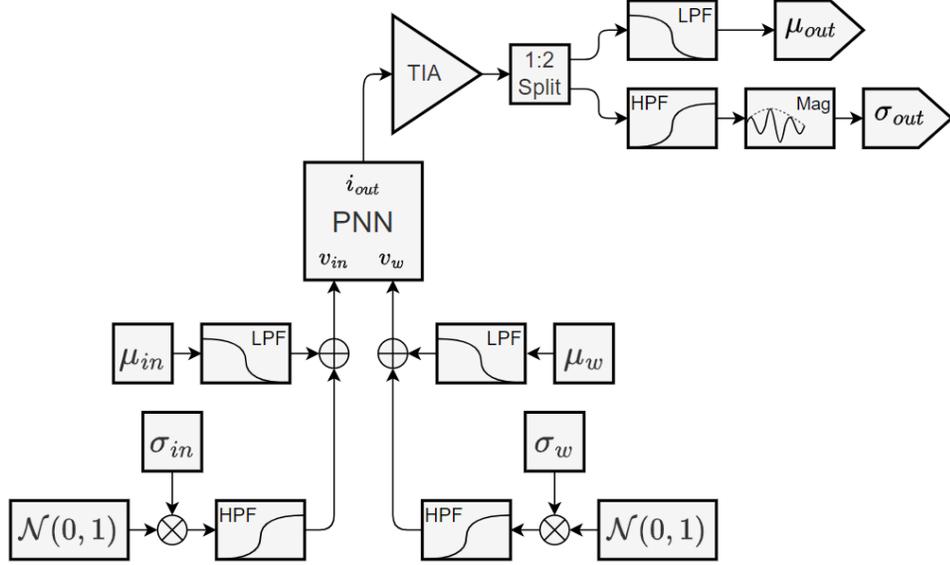
Where the first term is the negative-log-likelihood of the observed dataset appearing in the prediction output distribution, and the second term is the KL-divergence between the variational approximation parameters and the prior distribution. This prior distribu-

tion is a design parameter of the training pipeline, with a typical choice being the standard normal distribution $\mathcal{N}(0, 1)$ [58, 61, 60].

Modern methods of implementing BNNs in software involve the propagation of the first and second-order moments of the variational distribution through nonlinear activation functions through approximations of the corresponding integrals[64, 59, 61]. These approximation techniques are commonly referred to as Variational Density Propagation (VDP) techniques. In the continuous-time domain, signal central tendency can be approximated using a low-pass filter, and variation of a signal with zero mean can be approximated using a magnitude detector. These approximations will be valid for any variation that occupies the bandwidth above the cutoff frequency of the low-pass filters in the relevant circuits. To propagate the variation through a layer of an analog neural network, both the weight and input signals should be composed of a mean signal that occupies the bandwidth below the low-pass filter, and a source of blue Gaussian noise whose signal power is mostly concentrated above the cutoff of the low-pass filter, preferably with an empty transition band to avoid frequency smearing due to the time-domain multiplication corresponding to a frequency-domain convolution. Photonic neural networks (PNNs) are particularly well-equipped to represent such signals, as the inference bandwidth is largely bottlenecked by the electrical circuitry[26], allowing for a wide bandwidth of operation above the cutoff of the anti-aliasing filters used in the sampling circuitry. A system-level diagram illustrating the proposed operation is shown in Figure 11.

Figure 11

Variational Density Propagation System Schematic



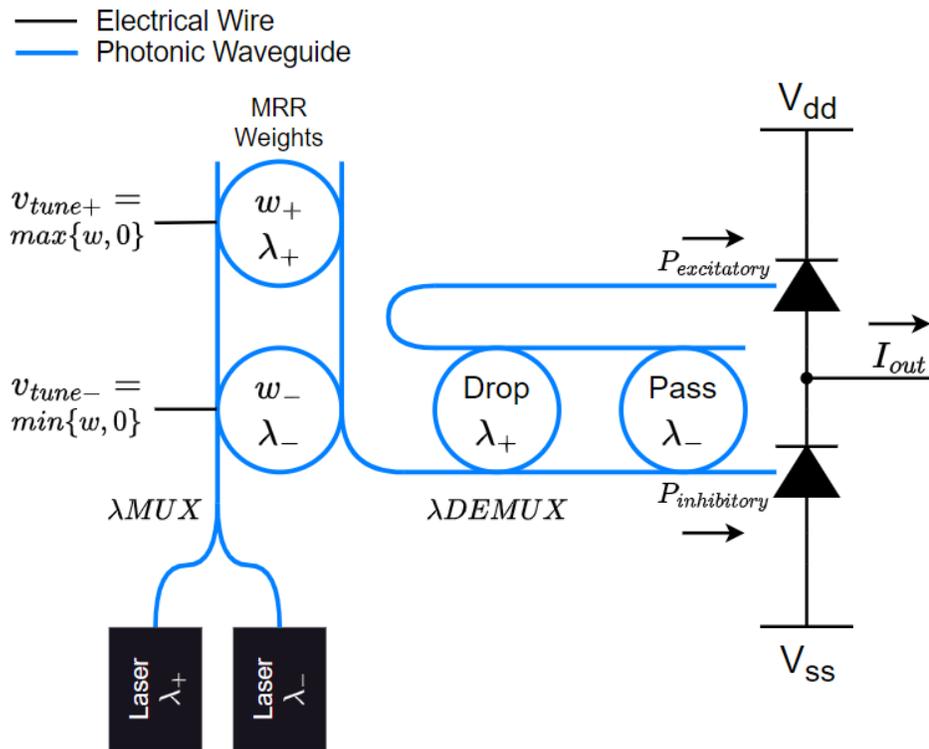
3.2.2 System Setup

The simulated photonic neural network represents input signals as a pair of optical amplitudes at different wavelengths, corresponding to positively and negatively weighed inputs. Each input is weighed by a separate MRR with a PIN junction modulator, allowing for an approximate 3dB modulation bandwidth of 5GHz and an approximate 10dB bandwidth of 10GHz. These wavelengths are multiplexed onto one waveguide, and demultiplexed into excitatory and inhibitory inputs by a pair of thermally tuned MRRs in a through-drop resonator configuration tuned to weights of 1 and -1 respectively. For simulation purposes, the electrical modulation bandwidth of the input rings were applied to digitally synthesized input signals using FIR filters which were fitted to the measured bandwidths from [29] using MSE minimization. The electro-optical amplitude transfer function was assumed to be applied before the modulation bandwidth, such that the simulation had direct control over the weights, rather than controlling the tuning voltages themselves. The input optical inten-

sities were assumed to be modulated off chip by optical intensity modulators with 20GHz modulation bandwidths. The represented architecture is depicted in Figure 12.

Figure 12

Architecture of the Simulated Two-Quadrant Photonic Multiplier Circuit



Note. This schematic was adapted with permission from [29]

The neural network architecture being accelerated consists of three fully-connected layers, separated by ReLU activation functions to allow for the use of a two-quadrant multiplier. The input is a 784-element vector, fitting to the data shape of the MNIST Handwritten Digits dataset[68]. The first hidden layer had 128 neurons, and the second had 64, before a

10-neuron classification head. For this experiment, the multiplications were performed using the low-pass and magnitude detection system depicted in Figure 11, and the activation functions were propagated through using extended VDP [64]. For comparison, a control model was trained using random variable algebra for the vector-matrix multiplication. For the remainder of this work, these models will be labeled "Photonic VDP" and "ExVDP", respectively. The models were made to minimize the cross-entropy of the classification, with added penalties for the KL divergence between the weight tensor distributions and the standard tensor normal distribution, as well as for the log-determinant of the output covariance matrix to encourage minimal output entropy. Both models were trained in batches of 32 images on the MNIST Handwritten Digits dataset[68] using the Adam optimizer[69] with a learn rate of 0.001 and clamped such that, with 99% confidence, the weight signals would not exceed the interval $(-1,1)$. Both models were trained for 25 epochs, and the weight tensors that performed best on the validation dataset were saved for evaluation. Once trained, the models were subjected to three sets of robustness tests to examine the behaviors of the variances in each model in response to various corruptions. Because these tests were made to assess the anomaly detection ability of this method of variational density propagation, these tests were modeled after [61].

The first test was an assessment of covariate shift. This assessment is performed by perturbing the test images by a small amount and measuring the difference in average output variance. This test examines random gaussian noise, as well as the Fast-Gradient-Sign-Method (FGSM) adversarial attack. The severity of these attacks was measured using the signal-to-noise ratio (SNR) in dB, where the variance of the pixel values was considered to be the signal amplitude, and the variance of the perturbation was considered to be the noise amplitude. This SNR was swept for each perturbation, and the average predictive variance of each perturbation was measured for all samples, those classified correctly by the network, and those classified incorrectly due to the perturbation at that SNR.

The next test was intended to examine the ability of the propagated variance signal

to predict misclassification due to various real-world corruptions that are commonly observed in image processing. To assess this behavior, the network was presented with handwritten digits corrupted by 8 different corruptions from the MNIST-C robustness benchmark [70]. The 8 corruptions used in this assessment were: 1) Dotted Line, 2) Scale, 3) Spatter, 4) Shear, 5) Shot Noise, 6) Translate, 7) Rotate, 8) Impulse Noise.

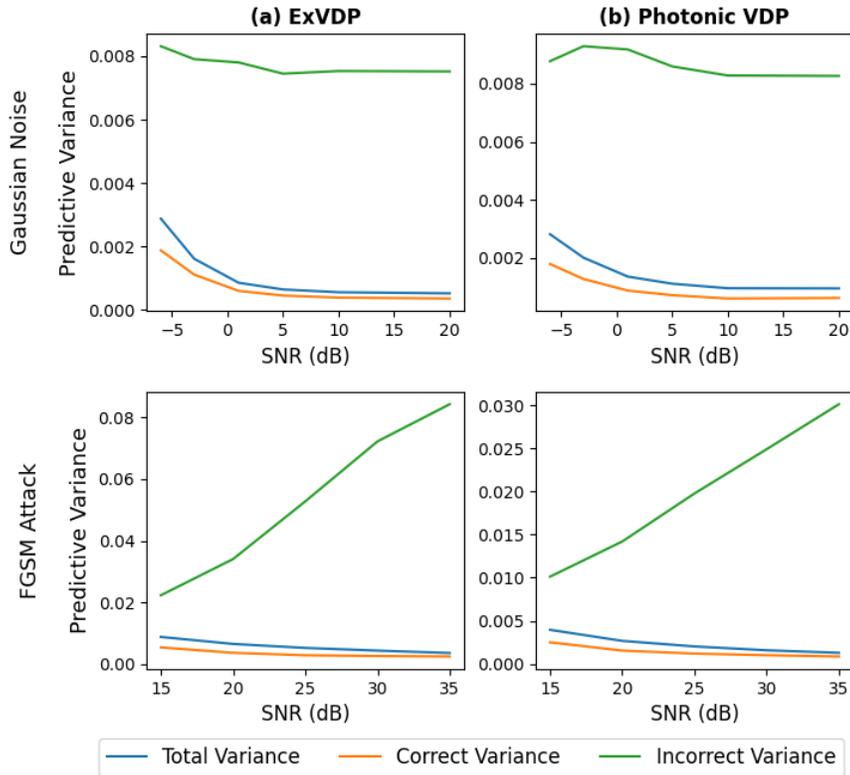
The last test assesses the ability of the predictive variance to detect semantic shift. For this test, the average variance of correctly classified unperturbed samples for each model was used as a static threshold, as described in [61]. The testing set from the Fashion MNIST dataset[71] was used as an out of distribution set of samples, which should ideally be detected as anomalous. The detection behavior was evaluated in terms of the true positive detections, which are defined as Fashion MNIST samples which were correctly detected as anomalous, and false positive detections, which are defined MNIST handwritten digit samples as that is the intended distribution that the network should generalize to.

3.3 Results and Discussion

Photonic VDP was trained to optimality after 14 epochs, reaching a validation accuracy of 95.70% on unperturbed handwritten digits and exhibiting a static variance threshold of $1.1708 * 10^{-4}$. ExVDP was trained to optimality after 15 epochs, reaching a validation accuracy of 96.41% and exhibiting a static variance threshold of $3.5736 * 10^{-4}$. These models were subjected to the three robustness evaluations to examine the behaviors of their predictive variances. The results of the covariate shift experiment are shown in Figure 13.

Figure 13

Average Predictive Variance Versus Signal-to-Noise Ratio for Gaussian Noise and Adversarial Attacks



Note. The depicted curves represent the predictive variance for all predicted samples (blue), correctly classified samples (orange) and misclassified samples (green). a) Extended VDP, as proposed by [64]. b) The proposed photonic VDP method.

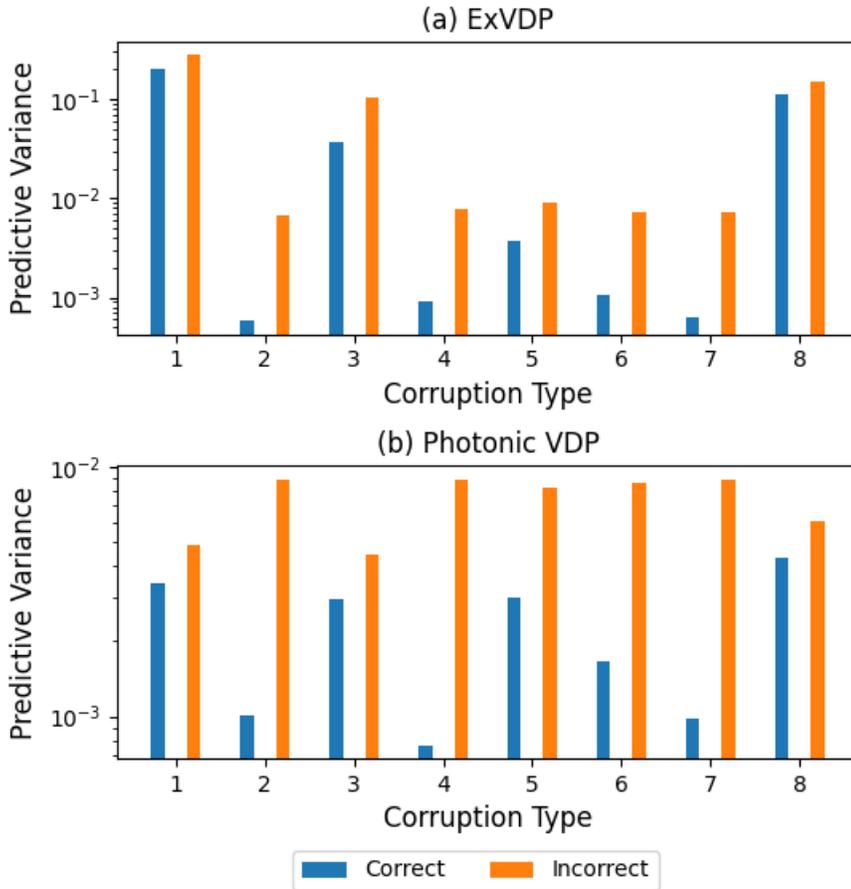
These figures illustrate the correctly classified samples in all cases to have a lower average predictive variance than incorrectly classified samples, with a very clear distinction between the groups. Comparing the models, they are shown to be comparably effective at indicating incorrect classification in the case of Gaussian noise perturbation. As expected in this case, the more severe perturbations, which stray further from the intended distribution, exhibit a higher variance. In contrast, the FGSM perturbations actually increase in variance

with increasing SNR, which seems to run in contrast to the hypothesis of the experiment. That said, high-severity adversarial attacks are typically able to be detected by other means like supervised anomaly classification or distribution modeling [72], so the lower severity attacks are more important to be detected with high predictive variance. Nonetheless, the incorrectly classified samples did exhibit a higher predictive variance, thus proving that the anomaly detection behavior works as intended.

The results of the real-world corruptions experiment are shown in Figure 14. This experiment shows that, in all cases for both models, the predictive variance of correctly classified samples is lesser than that of incorrectly classified samples, indicating the certainty of the model in correctly predicting the class. While the average overall predictive variance of classifications is shown to be smaller with Photonic VDP than the ExVDP baseline, the Photonic VDP experiment actually was shown to be more effective at cleanly discriminating incorrectly classified samples from correctly classified ones on average. This is indicated by the fact that there are less correct average variances that are higher than another category's incorrect average variance. This evidence would suggest that a thresholding method might be able to cleanly separate samples for which the model incorrectly predicts the class from those which are correctly predicted, regardless of corruption type, whereas ExVDP varies from type-to-type with corruptions much more, not allowing a single threshold to be defined as the variance of an incorrect classification.

Figure 14

Average Predictive Variance for 8 Different Corruptions from MNIST-C



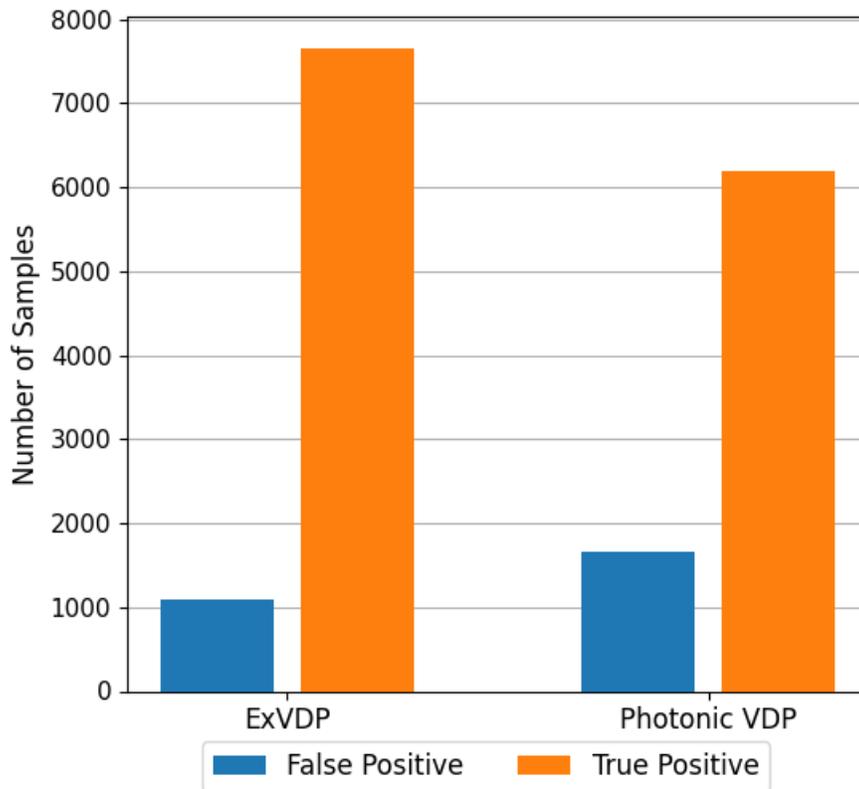
Note. The depicted curves represent the predictive variance for correctly classified samples (blue) and misclassified samples (orange). a) Extended VDP, as proposed by [64]. b) The proposed photonic VDP method.

The results of the semantic shift experiment are shown in Figure 15. As with the previous tests, this figure shows both methods to be capable of discriminating in-distribution from out-of-distribution samples. With perfect behavior, a network would have no false positive detections and detect all 10,000 samples in Fashion MNIST as truly anomalous. Using these criteria, Photonic VDP is shown to slightly under-perform com-

pared to the ExVDP baseline. This is likely due to the low average predictive variance as observed in the previous tests, both leading to anomalous samples being less dramatically uncertain, as well as causing the static threshold for detection to be lower, thus collecting more false positive detections.

Figure 15

Evaluation Under Semantic Shift



Note. True positive detections are Fashion MNIST samples that were correctly detected as anomalous. False positive detections are samples from the MNIST Handwritten Digits dataset that were incorrectly detected as anomalous.

Low predictive variance is a common problem between experiments for the Photonic implementation. It is possible that the uncertainty is learned to be small so as to ensure that the measured mean more accurately represents the theoretical mean of the propagation. This hypothesis is supported by findings comparing ExVDP to other variational density propagation methods [61], as the quality of the probability density estimate is shown to directly affect the allowable predictive variance during training. Findings from this work also observe that estimation methods can be improved by increasing the number of samples considered for each random variable. In the time-continuous equivalent of time-invariant discrete variational density propagation, an increase in sample size corresponds to an increase in electrical bandwidth of the operating devices and/or the whitening of naturally pink electrical noise. With the direct-digital-synthesis method used in this simulation, increasing the number of samples would violate the Nyquist sampling limit, but time-continuous noise sources like diode shot noise, ASE beating noise, or Johnson-Nyquist noise in resistive devices can be whitened and programmed simultaneously with an active high-pass filter. Using these time-continuous noise sources would likely result in a better estimate of the variational dynamics in practice, and also drastically reduce the computational complexity of the algorithm by removing random number generation, frequency filtration and time domain multiplication from the supporting software of the control system.

The major benefits of the proposed VDP method come in the form of execution speed. Deterministic convolutional neural networks implemented using digital electronic and analog photonic neural accelerators have been shown to out-pace early GPU technologies by speed-up factors between 1.4x and 7x[26], with improvements still to be made [73]. Key to this design is the encoding of parameter noise in frequency bands above the cutoff of the anti-aliasing filters of the ADC bottlenecks, which should allow for variational inference with no decrease in throughput compared to the deterministic method of inference, so long as mean and variance can be sampled in parallel. This means that implementing the proposed VDP method in a photonic convolution core would allow for variational in-

ference at a faster rate than the benchmark GPUs could perform deterministic inference. In exchange for this speed, the sampling circuitry increases in complexity, and the number of necessary sampling resources doubles. In addition, some method of generating and programming noise must be added, doubling the number of DACs necessary. Digital control signal synthesis consumes the majority of power in photonic convolution cores, so this increase in hardware demand results in an increase in estimated power consumption of roughly 1.6x (150W-183W from 95W-112W), using power calculations from [26] and assuming comparatively negligible costs due to the actual noise generation circuitry. Even with this increase in power consumption, the proposed method still consumes less power than the benchmark GPUs.

3.4 Conclusion

We proposed and simulated a method of leveraging high speed modulators in photonic neural accelerators to perform variational inference, considering device noises in the computation of central tendency and variation. This method was shown to exhibit the anomaly detection behaviors of variational inference in a comparable manner to state-of-the-art digital variational density propagation methods while theoretically operating at speeds faster than parallelized digital deterministic inference. The added functionality comes at the cost of a roughly 1.6x increase in power consumption over deterministic photonic inference, which remains less than the power consumption of traditional GPU architectures.

Chapter 4

Conclusions

This thesis thoroughly reviewed technologies from analog computing in general, and photonic computing in particular. The reviewed contributions demonstrate novel solutions for solving complex matrix and tensor processing tasks, and verify their validity through software simulation. These solutions leverage the natural advantages of integrated photonics, allowing for elegance of design and the use of resources that are not available in other computing paradigms, as advocated for in Chapter 1.

Chapter 2 addresses the problem of interference in LiDAR sensors. The prevalence of this problem has as of yet required commercial LiDAR sensors to operate at undesirably high powers to overcome the noise floor of the environments in which they are intended to operate. LiDAR sensors are also found to be vulnerable to jamming adversarial attacks, and incompatible with free space optical communication links or ranging pulses from other LiDAR sensors in the area. By changing LiDAR receivers to be multi-input, it became possible to strongly reject optical interference through the use of physical layer blind-source separation. This operation can be implemented on a photonic circuit positioned at the input of the LiDAR sensor, operating on the received signals before they are down-converted to electrical signals, thus adding effectively no latency to the system. This process can be accomplished with independent component analysis, though it was observed that time-of-flight LiDAR sensors need not perform such expensive statistical analyses as they can simply minimize interference in the absence of a ranging pulse before transmitting one.

Chapter 3 focuses on the implementation of emerging Bayesian neural networks using photonic neural accelerators. These networks are notoriously costly to implement digitally due to the time complexity of digitally approximating probability distributions,

leading to the under-exploration of their advantages in the machine learning industry. By leveraging unused bandwidth in common photonic neural accelerator architectures, these expensive distribution approximation integrals can be performed as high-speed analog operations. This not only eliminates the slow down incurred by considering uncertainty in neural network inference, but actually stands to accelerate inference over that of deterministic digital inference by using an already faster neural network co-processor. This benefit is made possible at the cost of an increase in power consumption over deterministic photonic inference, but this increased power consumption still leaves the neural accelerator more power efficient than a digital matrix processor in addition to performing faster.

References

- [1] S. M. Kuo, B. H. Lee, and W. Tian, *Real-time Digital Signal Processing: Fundamentals, implementations and applications*. John Wiley & Sons, 2017.
- [2] Q. Huang, Z. Huang, P. Werstein, and M. Purvis, “Gpu as a general purpose computing resource,” in *2008 Ninth International Conference on Parallel and Distributed Computing, Applications and Technologies*, 2008, pp. 151–158. DOI: 10.1109/PDCAT.2008.38.
- [3] G. Zhou, J. Zhou, and H. Lin, “Research on nvidia deep learning accelerator,” in *2018 12th IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, 2018, pp. 192–195. DOI: 10.1109/ICASID.2018.8693202.
- [4] W. N. H. E. and D. M. Harris, *CMOS VLSI Design: A circuits and systems perspective*. Pearson India, 2015.
- [5] L. Couch, *Digital and Analog Communication Systems* (Prentice-Hall international editions). Pearson, 2013, ISBN: 9780132915380. [Online]. Available: <https://books.google.com/books?id=j1UjygAACAAJ>.
- [6] Y. Li *et al.*, “Deep learning for lidar point clouds in autonomous driving: A review,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3412–3432, 2021. DOI: 10.1109/TNNLS.2020.3015992.
- [7] J. Garofolo, Y. Qi, T. Shi, and B. Wu, “Jamming-resilient lidar based on photonic blind-source separation,” in *Frontiers in Optics + Laser Science 2022 (FIO, LS)*, Optica Publishing Group, 2022, JW5A.47. DOI: 10.1364/FIO.2022.JW5A.47. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=FiO-2022-JW5A.47>.
- [8] J. Garofolo, Y. Qi, T. Shi, G. Tian, and B. Wu, “Photonic interference cancellation for lidar sensors,” *IEEE Photonics Technology Letters*, vol. 35, no. 23, pp. 1279–1282, 2023. DOI: 10.1109/LPT.2023.3316592.
- [9] J. Garofolo, T. Shi, T. F. de Lima, P. Prucnal, and B. Wu, “Digital electronic and analog photonic acceleration for point cloud classifiers,” in *Frontiers in Optics + Laser Science 2023 (FiO, LS)*, Optica Publishing Group, 2023, JM7A.129. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=LS-2023-JM7A.129>.
- [10] Y. Qi, T. Shi, J. Garofolo, G. Tian, and B. Wu, “Wideband interference management for free space optical communication based on photonic signal processing,” in *Frontiers in Optics + Laser Science 2022 (FIO, LS)*, Optica Publishing Group, 2022, JTU4A.47. DOI: 10.1364/FIO.2022.JTU4A.47. [Online]. Available: <https://opg.optica.org/abstract.cfm?URI=FiO-2022-JTU4A.47>.

- [11] C. Huang *et al.*, “Demonstration of photonic neural network for fiber nonlinearity compensation in long-haul transmission systems,” in *2020 Optical Fiber Communications Conference and Exhibition (OFC)*, 2020, pp. 1–3.
- [12] A. N. Tait *et al.*, “Demonstration of multivariate photonics: Blind dimensionality reduction with integrated photonics,” *Journal of Lightwave Technology*, vol. 37, no. 24, pp. 5996–6006, 2019. DOI: 10.1109/JLT.2019.2945017.
- [13] H. Haas *et al.*, “Introduction to indoor networking concepts and challenges in lifi,” *Journal of Optical Communications and Networking*, vol. 12, no. 2, A190–A203, 2020. DOI: 10.1364/JOCN.12.00A190.
- [14] J. Liu, Q. Sun, Z. Fan, and Y. Jia, “Tof lidar development in autonomous vehicle,” in *2018 IEEE 3rd Optoelectronics Global Conference (OGC)*, 2018, pp. 185–190. DOI: 10.1109/OGC.2018.8529992.
- [15] P. Wu, A. Escontrela, D. Hafner, K. Goldberg, and P. Abbeel, *Daydreamer: World models for physical robot learning*, 2022. arXiv: 2206.14176 [cs.R0].
- [16] A. Behagi and M. Ghanevati, *Fundamentals of RF and Microwave Circuit Design: Practical Analysis and Design Tools*. CreateSpace Independent Publishing Platform, 2018, ISBN: 9781722777418. [Online]. Available: <https://books.google.com/books?id=aJgougEACAAJ>.
- [17] J. Hasler and S. Shah, “An soc fpaa based programmable, ladder-filter based, linear-phase analog filter,” *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 2, pp. 592–602, 2021. DOI: 10.1109/TCSI.2020.3038360.
- [18] K. Atkinson and W. Han, *Elementary Numerical Analysis*. Wiley, 2004, ISBN: 9780471452263. [Online]. Available: https://books.google.com/books?id=F_4gPwAACAAJ.
- [19] M. Davies *et al.*, “Loihi: A neuromorphic manycore processor with on-chip learning,” *IEEE Micro*, vol. 38, no. 1, pp. 82–99, 2018. DOI: 10.1109/MM.2018.112130359.
- [20] A. Ige and J. Hasler, “Efficient implementation of a fully analog neural network on a reconfigurable platform,” in *2023 IEEE 66th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2023, pp. 454–457. DOI: 10.1109/MWSCAS57524.2023.10405875.
- [21] S. Ambrogio *et al.*, *An analog-ai chip for energy-efficient speech recognition and transcription*, Aug. 2023. [Online]. Available: <https://www.nature.com/articles/s41586-023-06337-5>.

- [22] S. Qiao, S. Moran, D. Srinivas, S. Pamarti, and S. S. Iyer, “Demonstration of analog compute-in-memory using the charge-trap transistor in 22 fdx technology,” in *2022 International Electron Devices Meeting (IEDM)*, 2022, pp. 2.5.1–2.5.4. DOI: 10.1109/IEDM45625.2022.10019527.
- [23] Q. Xia and J. J. Yang, *Publisher correction: Memristive crossbar arrays for brain-inspired computing*, Apr. 2019. [Online]. Available: <https://www.nature.com/articles/s41563-019-0357-9>.
- [24] G. P. Agrawal, *Fiber-optic communication systems*. Wiley, 2022.
- [25] Y. Du *et al.*, “Implementation of optical neural network based on mach–zehnder interferometer array,” *IET Optoelectronics*, vol. 17, no. 1, pp. 1–11, Nov. 2022. DOI: 10.1049/ote2.12086.
- [26] V. Bangari *et al.*, “Digital electronics and analog photonics for convolutional neural networks (deap-cnns),” *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–13, 2020. DOI: 10.1109/JSTQE.2019.2945540.
- [27] I. Chakraborty, G. Saha, and K. Roy, “Photonic in-memory computing primitive for spiking neural networks using phase-change materials,” *Physical Review Applied*, vol. 11, no. 1, Jan. 2019, ISSN: 2331-7019. DOI: 10.1103/physrevapplied.11.014063. [Online]. Available: <http://dx.doi.org/10.1103/PhysRevApplied.11.014063>.
- [28] W. Zhang *et al.*, “Silicon microring synapses enable photonic deep learning beyond 9-bit precision,” *Optica*, vol. 9, no. 5, pp. 579–584, May 2022. DOI: 10.1364/OPTICA.446100. [Online]. Available: <https://opg.optica.org/optica/abstract.cfm?URI=optica-9-5-579>.
- [29] W. Zhang *et al.*, “A system-on-chip microwave photonic processor solves dynamic rf interference in real time with picosecond latency,” *Light: Science & Applications*, vol. 13, no. 1, Jan. 2024. DOI: 10.1038/s41377-023-01362-5.
- [30] J. Garofolo and B. Wu, “Photonic analog signal processing and neuromorphic computing [invited],” *Chin. Opt. Lett.*, vol. 22, no. 3, p. 032 501, Mar. 2024. [Online]. Available: <https://opg.optica.org/col/abstract.cfm?URI=col-22-3-032501>.
- [31] A. Monpapassorn, “High frequency/low voltage cmos adder,” *Science & Technology Asia*, pp. 48–52, 2000.
- [32] Monika and P. Mittal, “Different current mirror topologies at multiple technology nodes: Performance comparison and parameters extraction,” in *2021 International Conference on Simulation, Automation & Smart Manufacturing (SASM)*, 2021, pp. 1–7. DOI: 10.1109/SASM51857.2021.9841209.

- [33] F. Khoeini, B. Hadidian, K. Zhang, and E. Afshari, "A transimpedance-to-noise optimized analog front-end with high psrr for pulsed tof lidar receivers," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 9, pp. 3642–3655, 2021. DOI: 10.1109/TCSI.2021.3089098.
- [34] J. Xu and Z. Lu, "A 10 gb/s cmos transimpedance amplifier with adaptive gain," in *2011 International Conference on Electric Information and Control Engineering*, 2011, pp. 5–7. DOI: 10.1109/ICEICE.2011.5777268.
- [35] A. Fish, V. Milrud, and O. Yadid-Pecht, "High-speed and high-precision current winner-take-all circuit," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 52, no. 3, pp. 131–135, 2005. DOI: 10.1109/TCSII.2004.842062.
- [36] T. F. de Lima *et al.*, "Noise analysis of photonic modulator neurons," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–9, 2020. DOI: 10.1109/JSTQE.2019.2931252.
- [37] S. Jung *et al.*, *A crossbar array of magnetoresistive memory devices for in-memory computing*, Jan. 2022. [Online]. Available: <https://www.nature.com/articles/s41586-021-04196-6>.
- [38] S. George *et al.*, "A programmable and configurable mixed-mode fpaa soc," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 6, pp. 2253–2261, 2016. DOI: 10.1109/TVLSI.2015.2504119.
- [39] J. Feldmann *et al.*, *Parallel convolutional processing using an integrated photonic tensor core*, Jan. 2021. [Online]. Available: <https://www.nature.com/articles/s41586-020-03070-1>.
- [40] P. R. Prucnal, B. J. Shastri, and M. C. Teich, *Neuromorphic Photonics*. CRC Press, Taylor & Francis Group, 2017.
- [41] L. Fick, M. El - Chammas, S. Skrzyniarz, and D. Fick, *System and methods for mixed-signal computing*, Apr. 2019.
- [42] S. Kim, S. Shah, and J. Hasler, "Calibration of floating-gate soc fpaa system," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 9, pp. 2649–2657, 2017. DOI: 10.1109/TVLSI.2017.2710020.
- [43] B. D. Rumberg and D. W. Graham, *Analog voice activity detection*, Oct. 2018.
- [44] M. Fuhrman and B. D. Rumberg, *Motion detection based on analog video stream*, Dec. 2023.

- [45] C. Ríos *et al.*, “In-memory computing on a photonic platform,” *Science Advances*, vol. 5, no. 2, eaau5759, 2019. DOI: 10.1126/sciadv.aau5759. eprint: <https://www.science.org/doi/pdf/10.1126/sciadv.aau5759>. [Online]. Available: <https://www.science.org/doi/abs/10.1126/sciadv.aau5759>.
- [46] T. Baehr-Jones *et al.*, “Monolithically integrated 112 gbps pam4 optical transmitter and receiver in a 45nm cmos-silicon photonics process,” *Opt. Express*, vol. 31, no. 15, pp. 24 926–24 938, Jul. 2023. DOI: 10.1364/OE.495246. [Online]. Available: <https://opg.optica.org/oe/abstract.cfm?URI=oe-31-15-24926>.
- [47] T. Shi, Y. Qi, and B. Wu, “Hybrid free space optical communication and radio frequency mimo system for photonic interference separation,” *IEEE Photonics Technology Letters*, vol. 34, no. 3, pp. 149–152, 2022. DOI: 10.1109/LPT.2022.3140364.
- [48] G. Kim, J. Eom, and Y. Park, “An experiment of mutual interference between automotive lidar scanners,” in *2015 12th International Conference on Information Technology - New Generations*, 2015, pp. 680–685. DOI: 10.1109/ITNG.2015.113.
- [49] Y. Cao *et al.*, *Adversarial sensor attack on lidar-based perception in autonomous driving*, Aug. 2019. [Online]. Available: <https://arxiv.org/abs/1907.06826>.
- [50] K. M. A. Alheeti, A. Alzahrani, and D. Al Dosary, “Lidar spoofing attack detection in autonomous vehicles,” in *2022 IEEE International Conference on Consumer Electronics (ICCE)*, 2022, pp. 1–2. DOI: 10.1109/ICCE53296.2022.9730540.
- [51] Z.-J. Han, X. Tang, Z.-M. Wu, and G.-Q. Xia, “A real-time and anti-interference lidar based on field programmable gate array,” in *2021 19th International Conference on Optical Communications and Networks (ICOON)*, 2021, pp. 1–3. DOI: 10.1109/ICOON53177.2021.9563654.
- [52] S. Yin and W. Wang, “Lidar signal denoising based on wavelet domain spatial filtering,” in *2006 CIE International Conference on Radar*, 2006, pp. 1–3. DOI: 10.1109/ICR.2006.343207.
- [53] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar*. Scitech Publishing Inc, 2010, vol. 1.
- [54] Q. Liu, B. Gily, and M. P. Fok, “Adaptive photonic microwave instantaneous frequency estimation using machine learning,” *IEEE Photonics Technology Letters*, vol. 33, no. 24, pp. 1511–1514, 2021. DOI: 10.1109/LPT.2021.3128867.
- [55] P. Y. Ma *et al.*, “Blind source separation with integrated photonics and reduced dimensional statistics,” *Optics Letters*, vol. 45, no. 23, p. 6494, 2020. DOI: 10.1364/ol.409474.

- [56] T. Shi, Y. Qi, W. Zhang, P. Prucnal, J. Li, and B. Wu, “Sub-nyquist optical pulse sampling for photonic blind source separation,” *Optics Express*, vol. 30, pp. 19 300–19 310, May 2022.
- [57] G. Ditzler, M. Roveri, C. Alippi, and R. Polikar, “Learning in nonstationary environments: A survey,” *IEEE Computational Intelligence Magazine*, vol. 10, no. 4, pp. 12–25, 2015. DOI: 10.1109/MCI.2015.2471196.
- [58] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, *Weight uncertainty in neural networks*, 2015. arXiv: 1505.05424 [stat.ML].
- [59] G. Carannante, D. Dera, G. Rasool, N. C. Bouaynaya, and L. Mihaylova, “Robust learning via ensemble density propagation in deep neural networks,” in *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2020, pp. 1–6. DOI: 10.1109/MLSP49062.2020.9231635.
- [60] C. F. Angelini, N. C. Bouaynaya, and G. Rasool, “Variational density propagation continual learning,” in *2023 International Symposium on Image and Signal Processing and Analysis (ISPA)*, 2023, pp. 1–6. DOI: 10.1109/ISPA58351.2023.10279186.
- [61] G. Carannante, D. Dera, O. Aminul, N. C. Bouaynaya, and G. Rasool, “Self-assessment and robust anomaly detection with bayesian deep learning,” in *2022 25th International Conference on Information Fusion (FUSION)*, 2022, pp. 1–8. DOI: 10.23919/FUSION49751.2022.9841358.
- [62] G. Carannante, D. Dera, G. Rasool, and N. C. Bouaynaya, “Self-compression in bayesian neural networks,” in *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2020, pp. 1–6. DOI: 10.1109/MLSP49062.2020.9231550.
- [63] I. E. Nielsen, R. P. Ramachandran, N. Bouaynaya, H. M. Fathallah-Shaykh, and G. Rasool, “Evalattai: A holistic approach to evaluating attribution maps in robust and non-robust models,” *IEEE Access*, vol. 11, pp. 82 556–82 569, 2023. DOI: 10.1109/ACCESS.2023.3300242.
- [64] D. Dera, G. Rasool, and N. Bouaynaya, “Extended variational inference for propagating uncertainty in convolutional neural networks,” in *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2019, pp. 1–6. DOI: 10.1109/MLSP.2019.8918747.
- [65] J. Feldmann *et al.*, *Parallel convolutional processing using an integrated photonic tensor core*, Jan. 2021. [Online]. Available: <https://www.nature.com/articles/s41586-020-03070-1>.

- [66] C. Wu, X. Yang, Y. Chen, and M. Li, "Photonic bayesian neural network using programmed optical noises," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 29, no. 2: Optical Computing, pp. 1–6, 2023. DOI: 10.1109/JSTQE.2022.3217819.
- [67] A. Paszke *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *CoRR*, vol. abs/1912.01703, 2019. arXiv: 1912.01703. [Online]. Available: <http://arxiv.org/abs/1912.01703>.
- [68] L. Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 141–142, 2012. DOI: 10.1109/MSP.2012.2211477.
- [69] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, 2017. arXiv: 1412.6980 [cs.LG].
- [70] N. Mu and J. Gilmer, *Mnist-c: A robustness benchmark for computer vision*, 2019. arXiv: 1906.02337 [cs.CV].
- [71] H. Xiao, K. Rasul, and R. Vollgraf, *Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms*, 2017. arXiv: 1708.07747 [cs.LG].
- [72] S. Bulusu, B. Kailkhura, B. Li, P. K. Varshney, and D. Song, "Anomalous example detection in deep learning: A survey," *IEEE Access*, vol. 8, pp. 132 330–132 347, 2020. DOI: 10.1109/ACCESS.2020.3010274.
- [73] S. Xu, J. Wang, and W. Zou, "Optical convolutional neural network with wdm-based optical patching and microring weighting banks," *IEEE Photonics Technology Letters*, vol. 33, no. 2, pp. 89–92, 2021. DOI: 10.1109/LPT.2020.3045478.