Design Technology for Scalable and Robust Photonic Integrated Circuits

(Invited Paper)

Zheng Zhao[1], Jiaqi Gu[1], Zhoufeng Ying[1], Chenghao Feng[1], Ray T. Chen[1,2], David Z. Pan[1]
[1]Department of Electrical and Computer Engineering, University of Texas at Austin
[2]Department of Materials Science and Engineering, University of Texas at Austin

***Abstract*—Photonic integrated circuit (PIC), as a promising alternative to traditional CMOS circuit, has demonstrated the potential to accomplish on-chip optical signal transmission and computations in ultra-high speed and/or low power consumption. One of the critical challenges of PIC, however, is that its scalability and robustness are limited by cascaded optical power loss and noise error. In this paper, we analyze the scalability and noise robustness challenges facing photonic integrated circuits, for two representative PIC applications: logic computing and neural networks. Automated design algorithms and learning methodologies are proposed to resolve these issues.**

I. Introduction

Computing on photonic integrated circuit (PIC) has been reignited as a promising alternative to traditional CMOS electronics as Moore's law winds down. By leveraging the property of light to process information, where the information is in the form of optical signals sourced by optical lasers and detected by photo-detectors, photonics has demonstrated the potential to accomplish ultra-high speed and/or low power consumption for on-chip signal transmission and computation [1], [2]. Compared with computing, optical interconnects have been more intensively investigated, which manifests the advantages over metal interconnects especially in intra- and inter-chip communications [2]–[5].

To catch up with the advancement with optical interconnects, previous works on optical computing have demonstrated upon two computing paradigms: digital and analog computing. Digital optical computing performs boolean logic, where optical switches serve as the core of this paradigm. Analog optical computing, on the other hand, interprets light signals as continuous values in the real or complex domain and performs analog-style computing using linear optics. As for optical logic applications, concentrated study has been performed on basic bitwise operations such as (N)AND, (N)OR and X(N)OR gates [6], [7], algebraic functions such as 1-bit half and full adders [8]–[10] and switchers [11]–[13]. In order to implement general and larger-scale logic functions and pave the way for design-space exploration, automated design methods are proposed based on various schemes: from virtual gates [14], to more recently, and-inverter gate (AIG) [15] and binary decision diagram (BDD) [14], [16]–[19]. However, as optical devices lack the capability of logic-level restoration and signal isolation as CMOS transistors, functional cascadability turns out to be very limited. For example, each splitter of virtual gate-based schemes and Y-branch combiner of AIG and BDD-based schemes produce a $-3dB$ loss which is cascaded and inevitably leads to an extremely weak output signal indistinguishable from noises. The power loss is thus a key reason for low signal-to-noise ratio (SNR). As the integration advances, crosstalk noise also becomes a critical issue facing signal integrity. The crosstalk noise has already been revealed in large-dimensional optical routers [20] and it would be important to revisit the solutions in the new context of computing. Last but not least, the garbage outputs mentioned in [16] also leads to a degraded SNR.

As for analog computing applications, research efforts have been made on matrix multiplication [21]–[24], and optical neural networks (ONNs), evolving from the former [25], [26]. ONN distinguishes itself by directly exploiting linear optics to perform neuromorphic operations and demonstrated both speed and power efficiency moving beyond von Neumann architecture. For electronics, matrix multiplication, the core and performance-critical computation of neural networks, is a computationally expensive operation; while with optics, it can be performed with near-zero energy using Mach-Zehnder interferometers (MZIs), as successfully demonstrated on chip in [25]. Furthermore, as optical signals can transports in the same channel in parallel via wavelength-division multiplexing (WDM), this also brings the potential of scaling the computation bandwidth by tens of times. However, same as logic computing PICs, ONNs bear on the challenges of scalability and robustness. On one hand, scalability is inevitably impeded by the inherent size of optical devices such as MZIs. This problem deteriorates as the scale of neural network models keeps increasing to accommodate ever more complex applications. On the other hand, robustness also becomes more and more critical due to the scale-up. Specifically, since the phase of each MZI is highly impacted by environmental change, thermal crosstalk, and imperfect manufacturing, the phase error is cascaded throughout the computation. As discussed in [25], the accuracy could be aggravated by 20% for small ONN applications. Preliminary research has studied a slimmed architecture to reduce the size by reducing the number of MZIs [26]. It is also interesting to notice that, when applied phase noise on each MZI, a smaller ONN also show better robustness.

As can be concluded, robustness and scalability have been the two major obstacles of building large scale and

978-1-7281-2350-9/19/$31.00

practical optical computing circuits. They are also deeply correlated in the context of optical computing due to the cascaded power loss and/or noise error, engendered by the lack of signal restoration and isolation in optics. This paper studies the two issues for two representative optical computing paradigms. The paradigms have shared problems while demonstrating their unique features that require different ways to approach. As for optical logic circuit, we discuss automated design techniques by introducing restoration using OE/EO converters. For optical neural network, we studied the noise sensibility under different situations. Automated learning and co-design methodologies are presented to compensate for both issues.

The remainder of this paper is organized as follows. Section II discusses the scalability and robustness issues of optical logic circuit and proposes an automated design method using OE/EO converters. Section III focuses on optical neural networks, providing a wide range of characterizations and potential solutions. The paper is finally concluded in Section IV.

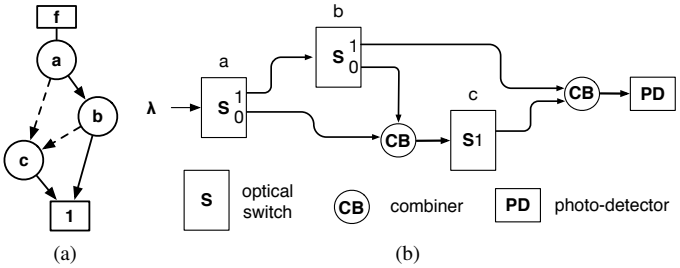


Figure 1: BDD and the corresponding optical implementation.

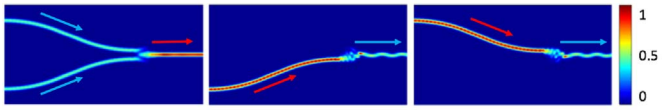


Figure 2: Power distribution of a typical optical Y-branch combiner.

II. Optical Logic Circuits

In this section, we study the first application of optical computing: optical boolean logic. The scalability and robustness issues of optical logic circuits are mainly due to optical power depletion. is a major obstacle to build complicated systems. As demonstrated in the previous synthesis methodologies [10], [14], [16], [17], optical devices lack the capability of signal restoration and input-output isolation as its CMOS counterparts, optical signal inevitably diminishes throughout the computation and become indistinguishable from environmental noise. One solution is to redistribute the power by logic rewriting [17], however, the method has limited effectiveness in terms of the potential to improve both the scalability and robustness, even if the overhead is not constrained. In this section, we highlight another more method which introduces optical OE/EO converter into the synthesis flow for signal restoration. The current state-of-the-art integrated OE/EO converter can achieve highly efficient [12], [27], [28] signal conversion. Distinguished to the previous method [17], this method guarantees the improvement of both criteria with the increase of the overhead budget. In intuition, consider the extreme case if we apply the restoration at every waveguide between any two devices to compensate for the loss, the resultant power depletion can be minimized to 0.

A. Optical Power Depletion and Noise Robustness

We start with the background of the classic optical synthesis method based on binary decision diagram (BDD) [16], [17], [19], [29]. A BDD is a directed acyclic graph that can represent a boolean function. As an example in Figure 1a, BDD has two types of nodes, the terminal node and decision node. A 1-terminal node, representing the functional output evaluation to be logic 1. A decision node is functionally a 1×2 crossbar switch, which is controlled by a decision variable. The solid (dashed) edges correspond to an assignment of the variable to be 1 (0). The classic BDD-based synthesis is demonstrated in Figure 1b. The light (λ) from a laser source (or from the output of the previous optical network) is streamed from the BDD top node to the 1-terminal, where a photodetector (PD) (or optical amplifier to the next computation stage) is located. The synthesis replaces each BDD node by an optical 1×2 crossbar, each controlled by a primary input. Waveguides and combiners are used to connect the crossbars. When there are multiple inputs to a crossbar, optical combiners (CB) are used to merge the inputs. The output of the optical network is a logical 1, if the PD detects light at the 1-terminal, otherwise it is a logical 0.

The classic BDD-based implementation suffers from cascading optical power loss as the single input caused intrinsic combiner loss and is thus prohibited from building larger-scale functionalities. As shown in the simulated power distribution in Figure 2, the Y-branch takes light from the two ports on the right side and passes light to the left. In the first figure, if two input ports have light, the output power doubles each of the input power and there is almost zero loss. In the second and third figures, if only one input has light, due to the mode mismatch, half of the light escape from the waveguide to the free space. Therefore, there will be a -3dB (50%) power loss at the output. The latter is the only case in BDD-based implementation. The loss cascaded inevitably leads to an extremely weak output signal indistinguishable from noises. Other sources of optical power loss of optical logic circuits include optical switch loss, waveguide propagation, and crossing loss. As the latter is dependent on final physical placement and routing, in this work, we focus on the optical loss induced by optical switches and combiners. The proposed method can be adapted to other various loss sources. The type of

switches and choice of platforms is also not restricted.

We use L to denote the absolute optical power loss in dB. In a BDD-based optical network, L can be defined for the node v as an optical switch (L_v), edge as an input branch of a combiner ($L_{(u,v)}$), path as a sub-network ($L_{v \to u}$), or the whole network (L_{net}). The path loss is calculated by adding the loss of all the components including combiners, couplers and switches, along the path. The network loss is defined as the greatest loss of all the paths from the network input (BDD top node) to the network output (BDD terminal node). Our goal of the synthesis is to improve the network loss by using OE/EO converter to provide boosted optical power at selected locations of a network.

B. Optical Power Restoration

Given a certain optical power loss goal $G > 0$ in dB, the optical loss restoration problem is defined as:

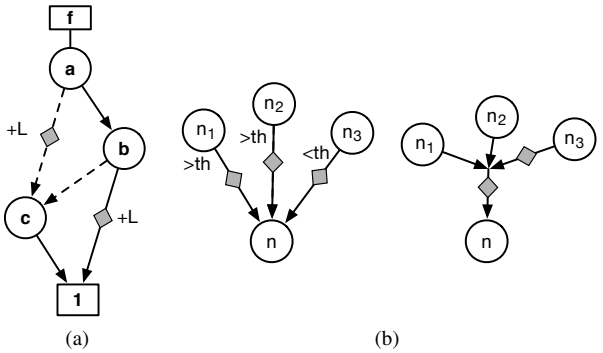
$$Minimize\ R = \sum_{i=1}^{|E|} x_i \tag{1}$$

$$\text{s.t. } L_{net} < G \tag{2}$$

$$x_i \in \{0, 1\}, \forall i \tag{3}$$

Each boolean variable x_i represents an assignment of restoration for a critical edge e_i. If x_i is 1, then an OE/EO converter is assigned to this edge; otherwise, the edge is not assigned any converter. The objective R is the number of OE/EO converters for restoration. Equation (2) states the power loss of the whole BDD network, which is defined by the minimum of all the path efficiency factor, is smaller than the given target G. Note that each converter has a determined detection threshold th measured in dB. As shown in Figure 3a, the insertion of converters at two edges boosts each power to the EO converter source power. The boosting value of a converter inserted at an edge is equal to the path loss ending at this edge. The converter threshold is determined by both the device detection limitation and the environmental noise. If the power loss from the top node to some point is greater than th, the converter is not able to be applied.

We propose a simple algorithm to this problem detailed in Algorithm 1. As a start, the critical paths are computed based on their power loss and the target G [30]. Then we check each edge on the critical path, from top to bottom, whether the threshold condition can be satisfied. If it is the case, we apply the OE/EO converter at this edge for restoration. Upon this change, an update of all the power loss of downstream nodes is then required, so that certain previously non-restorable edges can be restored due to this move. At any point, if the terminal power loss $L_{top \to 1-terminal}$ meets the target loss, we finish the loop. Finally, the merge operation is performed to further reduced redundant converters. The merge operation merges the assigned converters of edges $(n_i \to n)$ with a single converter at the output of the combiner connected to the node n. The merge operation is only effective if both



(a) (b)

Figure 3: (a) OE/EO converter inserted at two edges $(a \to c)$ and $(b \to 1 - terminal)$ (b) Merge operation example.

Algorithm 1 Optical Power Restoration.

Require: Target BDD function and and target loss G.

```
Compute critical path P ← {p_i : L_{p_i} > G}
for Each critical paths p_i sorted from top L_{p_i} do
    for Each critical edge e_j in p_i do
        n_in, n_out ← input and output node of e_j
        if L_{top→n_in} ≤ th then        ▷ e_j can be restored.
            Apply restoration at edge e_j
            for n_k ∈ {downstream of n_in} do
                Update L_{top→n_k}
        if Inputs of n_out is traversed then
            Merge converter if condition is satisfied.
        if L_{top→1−terminal} < G then
            break
```

conditions are satisfied: (1) the merged point meets the noise threshold, and (2) the number of converters after the merge becomes smaller. Note that the first condition may not always satisfy during the merge operation; otherwise, one can continue the operation till reaching the 1-terminal, resulting in one single converter. Figure 3b shows the example where two edge $(n_1 \to n)$ and $(n_2 \to n)$ meets the threshold before using any converters, and their respective converters are merged to be 1; while $(n_3 \to n)$ does not meet the second criterion so the converter cannot be merged.

The simulation results are shown in Table I. The first three columns summarized the benchmark name and the number of primary outputs and the number of optical switches based on the given ordered BDD. Column 4-6 show the number of OE/EO converters under corresponding to each relative improvement, varying from 10dB, 12dB to 16dB. The threshold is set to be a relatively conservative 25dB. We can notice the trend that in order to improve more power, the number of OE/EO converters also needs to increase. Bigger designs do not necessarily need more converters as the number of critical paths is not necessarily greater and how they distribute is uncertain. As was calculated in the last row, the average converter numbers are 3.57% , 5.27%, 11.04% of the total number of optical

TABLE I: Simulation results with different optical power target: 10dB, 12dB and 16dB.

benchmark	#PO	#sw	10dB	12dB	16dB
dalu	16	1692	20	39	194
apex7	37	458	37	53	111
stpmotor	29	491	31	52	104
k2	43	2113	10	16	56
cps	102	2224	45	49	94
i5	66	672	44	78	152
x3	99	851	19	24	100
frg2	139	1981	30	47	58
pdc	40	960	43	75	66
spla	46	977	45	54	67
vda	39	1117	55	71	222
apex5	85	1410	89	118	166
simple_spi	144	1473	70	100	177
x4	71	602	91	147	254
i2c	140	1836	44	71	260
example2	66	645	14	48	98
average	73.1	1257.1	44.9	66.3	138.7
ratio to #sw			3.57%	5.27%	11.04%

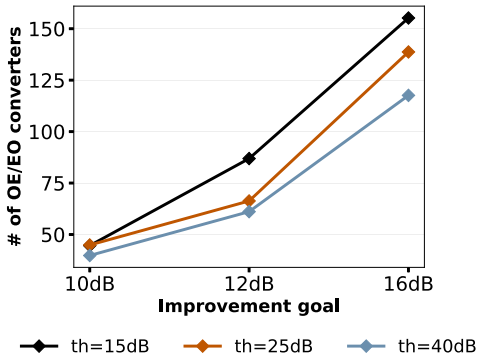


Figure 4: Number of converters under different improvement goals and noise threshold.

switches, respectively.

When the noise threshold is the dominating factor of th, we perform the second set of experiments. Figure 4 plots the number of converters with respect to different th of 15dB, 25dB, and 40dB as well as different improvement goals. In general, the trend is super-linear to the power target. Smaller thresholds reflect more noisy environment. For example, in the noisiest environment, the threshold is also the smallest 15dB for the signal to be detected correctly. Intuitively, the number of converters increases when the noise threshold becomes smaller.

III. Optical Neural Networks

In this section, we focus on the robustness and scalability issues of recently proposed integrated ONN architectures. We will demonstrate several software-hardware co-design methodologies that can enable more noise-robust and scalable ONN implementations.

A. Phase Noise Robustness

In the classical integrated ONN architecture, Mach-Zehnder Interferometer (MZI) arrays are constructed to realize MLP inference. Due to environmental changes, thermal crosstalk, and manufacturing imperfactions, noise exists in each MZI such that the phases of the output light signals will be perturbed. Thus we refer to this noise as phase noise. In each fully-connected layer of the MLP, matrix-vector multiplication is performed. The optical implementation of matrix multiplication is shown in Figure 5. Specifically, consider an n-input channel, m-output channel layer, the weight matrix $\boldsymbol{W} \in \mathbb{R}^{m\times n}$ is first decomposed using singular value decomposition $\boldsymbol{W} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^*$, where $\boldsymbol{U}$, $\boldsymbol{V}^*$ are $m \times m$ and $n \times n$ unitary matrices, respectively, and $\boldsymbol{\Sigma}$ is an $m \times n$ diagonal matrix. Each of the unitary matrices $\boldsymbol{U}$ and $\boldsymbol{V}^*$ can be further parametrized into the product of a series of planar rotations,

$$\boldsymbol{U}(n) = \boldsymbol{D}\prod_{i=n}^{2}\prod_{j=1}^{i-1}\boldsymbol{R}_{ij}, \tag{4}$$

where $\boldsymbol{D}$ is an $n \times n$ diagonal matrix that only contains 1 or -1, $\boldsymbol{R}_{ij}$ is an $n \times n$ identity matrix except for the four entries at (i,i), (i,j), (j,i) and (j,j), which are replaced by $\cos\phi$, $\sin\phi$, -$\sin\phi$, and $\cos\phi$. This planar rotation $\boldsymbol{R}_{ij}$ can be implemented with a 2×2 MZI, and its transfer function is given by,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}. \tag{5}$$

Therefore, for an arbitary $m \times n$ matrix $\boldsymbol{W}$, we can use total $n(n-1)/2 + m(m-1)/2$ MZIs to build two MZI arrays for each of the decomposed unitary matrices $\boldsymbol{U}$ and $\boldsymbol{V}^*$. The diagonal matrix $\boldsymbol{\Sigma}$ can be simply realized by attenuators/optical amplifiers.

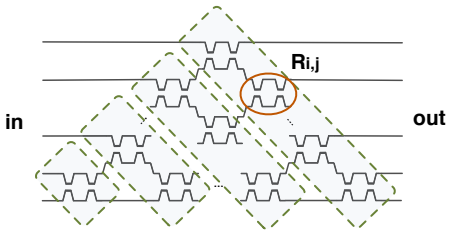


Figure 5: MZI array for unitary matrix.

The phase of the phase shifter on each MZI is theoretically given by $\phi = \gamma v^2$, where γ is the device- and temperature-related coefficient and v is the voltage control of the thermal-optic phase shifter on the MZI. Thus, the actual phase lag caused by a phase shifter is perturbed by multiple noise sources. This phase noise can be approximately modeled as a gaussian noise $\mathcal{N}(0,\sigma^2)$. To evaluate the error caused by phase noise, we first demonstrate the ℓ_2 distance between random unitary matrices without and with phase noise injected $||\boldsymbol{U} - \boldsymbol{U}_n||_2^2$. Figure 6 shows that larger phase noise and larger unitary matrix size will both contribute to larger error (ℓ_2 distance) of the unitary matrix. This matrix size-related error caused by phase noise could limit the ONN scale, as it may cause significant accuracy

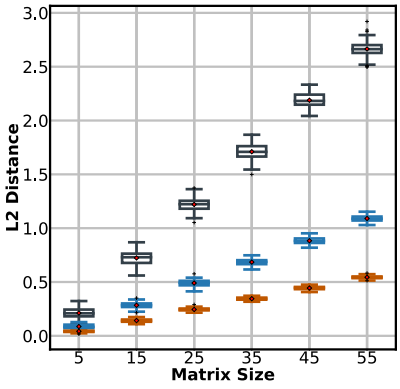


Figure 6: ℓ_2 difference of various-size unitary matrices under different phase noise standard deviations. Black, blue, red boxes represent noise standard deviation of 0.05, 0.02, and 0.01, respectively.

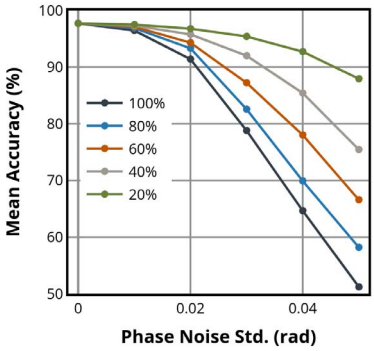


Figure 7: Model accuracy on MNIST dataset with different exposure rate and phase noise standard deviations. The three-layer MLP setup is (10×10)-64-64-10.

degradation when implementing relatively large matrices. To investigate the relation between model accuracy and phase noise, we inject phase noise with different σ into a pre-trained three-layer MLP. The model accuracy is tested on a downsampled MNIST dataset. We expose different portions of MZIs to noise to demonstrate the impact of phase perturbation on ONN performance. Figure 7 illustrates noise robustness of a three-layer MLP to different phase noises. When the phase noise std. is smaller than 0.01, the accuracy degradation is negligible (<1%), but the accuracy drops drastically as larger noise is injected. We also notice that when fewer MZIs are exposed to phase noise, higher testing accuracy the model will achieve. This enables a possible methodology to improve robustness by cutting down the number of components used in the hardware implementation.

As different layers in the MLP will extract different levels of features, the sensitivity to phase noise will also vary from layer to layer. We individually inject phase noise

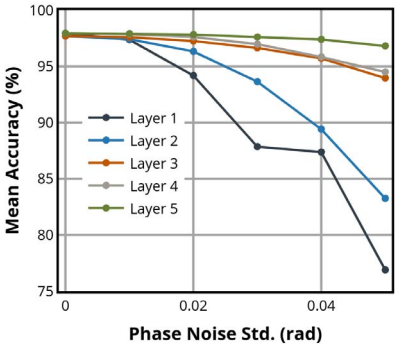


Figure 8: Model accuracy on MNIST dataset when injecting phase noise into different layers. The five-layer MLP setup is (10×10)-64-64-64-64-10.

into each layer in a five-layer MLP and show its resultant inference accuracy in Fig. 8. The first several layers are more sensitive to phase noise compared with deeper layers, because shallow layers are closer to inputs and responsible to extract low-level features, which are of vital significance to the model expressivity. In addition, the first layer often consists of a large input feature dimension, which will be more sensitive to noise according to the above discussion on unitary matrix size. Therefore, selective protection to MZIs in the shallow layers would be a good strategy to mitigate noise robustness issues of ONNs.

This set of experiments intends to demonstrate that by decreasing the number of optical components, the neural network robustness can also be improved. As depicted in the box plots of Figure 9, there are three random noise amplitude settings imposed upon the phases of MZI: 0.020, 0.025 and 0.050. Each conforms with a truncated norm distribution. For each noise setting, we generate 20 noisy samples for both the previous architecture (Figure 9a) and the slimmed architecture (Figure 9b). Taking (14×14)-150-150-10 as an example, it can be seen that the accuracy distribution of the slimmed architecture not only has higher average and geometric means but also a smaller variation range between the best and worst among all the samples.

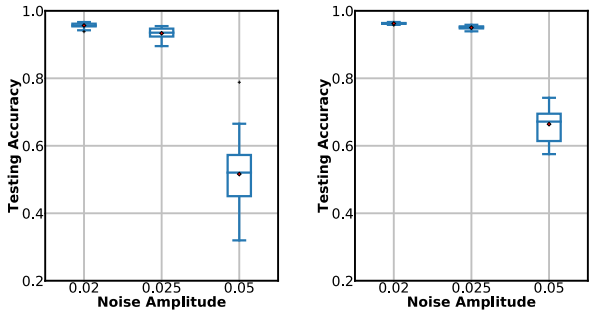
Another approach to improving ONN robustness is to add weight regularization term in the learning objective,

$$L = L_{base} + \sum ||\boldsymbol{W}||_2^2, \tag{6}$$

where L_{base} is the basic loss function. ℓ_2 regularization penalizes weights with large norms and mitigates overfitting problems, which could lead to a more smooth solution space and thus less sensitivity to noise perturbation. This regularization term is equivalent to performing weight decay in the optimization step,

$$\boldsymbol{W}^{t+1} = \boldsymbol{W}^t - \eta(\nabla L_{base} + \lambda \sum \boldsymbol{W}^t), \tag{7}$$

where η is the learning rate and λ is the weight decay rate. We train a tree-layer MLP with various weight decay rates, and plot the testing accuracy under different phase



(a) (b)

Figure 9: Noise robustness of (a) the classic architecture (b) the proposed architecture with the (14×14)-150-150-10 setup.

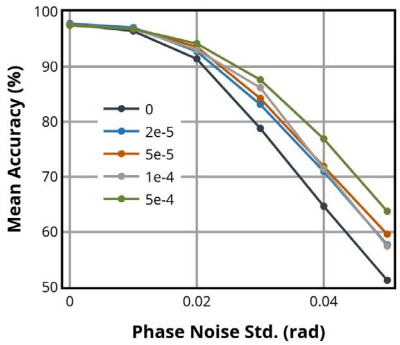


Figure 10: Model accuracy on MNIST dataset when usign different weight decay rates λ. The three-layer MLP setup is (10×10)-64-64-10.

noise standard deviations in Fig. 10. The experimental results show that training with weight decay can regularize the neural networks and achieve higher testing accuracy than the baseline without weight decay. This regularization approach introduces minimum training overhead and can effectively improve the robustness of ONNs to phase noise.

B. ONN Scalability

The scalability issue of ONNs attributes to two aspects: Area cost and model performance. We will discuss these two aspects individually together with corresponding design methodologies.

In the classical integrated ONN architecture, total $m(m-1)/2+n(n-1)/2+\max(m,n)$ MZIs will be used to build an $m\times n$ weight matrix. This hardware complexity can limit the actually implementation scale of ONNs, especially when the size of an MZI reaches up to $\sim$100 μm. For instance, a typical on-chip MZI array will have approximately 100 MZIs at most, which therefore requires to partition the matrix multiplication into blocks with extra scheduling overhead and larger latencies. To improve the area efficiency

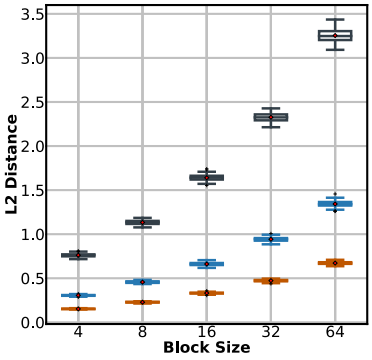


Figure 11: ℓ_2 distance between a 64×64 weight matrix with and without phase noise injected. The matrix is partitioned into various size of blocks. Black, blue, red boxes represent phase noise standard deviation of 0.05, 0.02, and 0.01, respectively.

and thus scalability of ONN, novel architectures can be adopted to cut down its area cost. The proposed slimmed ONN architecture can cut down 15-38% optical components and thus improve the area cost of ONNs. This architecture adopts a software-hardware co-design methodology to substitute the original SVD with a $TU\Sigma$ decomposition method. In this way, one of the area-expensive unitary blocks $\boldsymbol{V}^*$ is replaced by a sparse tree network $\boldsymbol{T}$. The sparse tree network $\boldsymbol{T}$ adapts the difference between input channel m and output channel n, with merely $\mathcal{O}(n)$ MZIs adopted. Consider that the theoretical hardware cost of this slimmed architecture is $n(n+1)/2$ MZIs, the scalability of ONNs is thus improved, enabling the implementation of a more compact integrated ONN.

Another scalability issue that faces ONNs is related to the aforementioned robustness issue. As we discussed above, larger unitary matrices will accumulate larger errors when phase noise is injected. Hence, directly mapping a large neural network into MZI arrays in a flattened way leads to dramatically decreasing signal-to-noise ratio, which will severely harm the inference performance. Similarly, we can adopt the slimmed architecture and network pruning technique to cut down the component utilization and reduce the number of noise sources to improve the scalability. Also, environmental noises can be modeled and considered in the training flow with weight regularization strategy to obtain a more robust solution. This noise-aware training method has the potential to train a fault-tolerant ONN with better scalability.

Another possible approach to resolving this scalability issue is to strike a balance between efficiency and performance through tiled matrix multiplication algorithm. If the matrix multiplication can be performed in a tiled way, each small sub-matrix multiplication can be mapped to a small-scale MZI array such that the area cost is well-controlled

and the noise error will be constrained in an acceptable range as the phase error will only impact one particular sub-matrix. We partition a 64×64 weight matrix $\boldsymbol{W}$ with different sizes of sub-matrices, from 4 to 64, and Fig. 11 plots the ℓ_2 distance between the original matrix $\boldsymbol{W}$ and $\boldsymbol{W}_n$ with phase noise injected in each block. As can be seen, by using small blocks, the total error caused by phase noise reduces accordingly, which offers a good reason to adopt blocking matrix multiplication for better scalability of ONNs.

Even though this blocking matrix multiplication method requires to perform the partial product accumulation in electronics with extra overhead, this could still benefit the overall performance and throughput if the optical computing part can offer orders-of-magnitude faster matrix multiplications with reasonable fidelity.

IV. Conclusion

Robustness and scalability have been two major obstacles to building large scale and practical optical computing circuits. This paper discusses the two issues for two representative optical computing paradigms: logic computing and neural networks. As discussed in the previous sections, they are highly correlated due to the cascaded power loss or noise error, both a result of the lack of signal restoration and isolation in optics. Logic computing and neural networks have shared properties while still demonstrating unique features that require specific ways to approach. In this paper, for optical logic circuit, we discuss automated design techniques by introducing the restoration into optics using OE/EO converters. For optical neural networks, we have studied the noise sensitivity under different situations. Automated learning and co-design methodologies are presented to compensate for both issues.

Acknowledgement

The authors acknowledge the Multidisciplinary University Research Initiative (MURI) program through the Air Force Office of Scientific Research (AFOSR), contract No.FA 9550-17-1-0071, monitored by Dr. Gernot S. Pomrenke.

References

[1] D. A. Miller, "Attojoule optoelectronics for low-energy information processing and communications," *Journal of Lightwave Technology*, 2017.

[2] C. Sun, M. T. Wade, Y. Lee, J. S. Orcutt, L. Alloatti, M. S. Georgas, A. S. Waterman, J. M. Shainline, R. R. Avizienis, and S. Lin, "Single-chip microprocessor that communicates directly using light," *Nature*, 2015.

[3] Y. Ye *et al.*, "3-D Mesh-based Optical Network-on-Chip for Multiprocessor System-on-Chip," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2013.

[4] D. Ding, B. Yu, and D. Z. Pan, "GLOW: A Global Router for Low-power Thermal-Reliable Interconnect Synthesis using Photonic Wavelength Multiplexing," in *Design Automation Conference (ASP-DAC), 2012 17th Asia and South Pacific*. IEEE, 2012.

[5] D. Liu, Z. Zhao, Z. Wang, Z. Ying, R. T. Chen, and D. Z. Pan, "Operon: optical-electrical power-efficient route synthesis for on-chip signals," in *Proceedings of the 55th Annual Design Automation Conference*. ACM, 2018.

[6] P. Zhou, L. Zhang, Y. Tian, and L. Yang, "10 ghz electro-optical or/nor directed logic device based on silicon micro-ring resonators," *Optics letters*, 2014.

[7] L. Yang, L. Zhang, C. Guo, and J. Ding, "Xor and xnor operations at 12.5 gb/s using cascaded carrier-depletion microring resonators," *Optics express*, 2014.

[8] Z. Ying, Z. Wang, S. Dhar, Z. Zhao, D. Z. Pan, and R. T. Chen, "On-chip microring resonator based electro-optic full adder for optical computing," in *CLEO: QELS_Fundamental Science*. Optical Society of America, 2017.

[9] Y. Tian, L. Zhang, J. Ding, and L. Yang, "Demonstration of electro-optic half-adder using silicon photonic integrated circuits," *Optics express*, 2014.

[10] Z. Ying, Z. Wang, Z. Zhao, S. Dhar, D. Z. Pan, R. Soref, and R. T. Chen, "Silicon microdisk-based full adders for optical computing," *Optics letters*, 2018.

[11] ——, "Comparison of microrings and microdisks for high-speed optical modulation in silicon photonics." AIP Publishing, 2018.

[12] E. Timurdogan *et al.*, "An ultralow power athermal silicon modulator," *Nature communications*, 2014.

[13] Z. Wang, X. Xu, D. Fan, Y. Wang, and R. T. Chen, "High quality factor subwavelength grating waveguide micro-ring resonator based on trapezoidal silicon pillars," *Optics letters*, 2016.

[14] C. Condrat, P. Kalla, and S. Blair, "Logic Synthesis for Integrated Optics," in *Proceedings of the 21st edition of the great lakes symposium on Great lakes symposium on VLSI*. ACM, 2011.

[15] Z. Ying, Z. Zhao, C. Feng, R. Mital, S. Dhar, D. Z. Pan, and R. T. Chen, "Automated logic synthesis for electro-optic computing in integrated photonics," in *Optical Interconnects XIX*. International Society for Optics and Photonics, 2019.

[16] R. Wille, O. Keszocze, C. Hopfmuller, and R. Drechsler, "Reverse BDD-based Synthesis for Splitter-free Optical Circuits," in *Design Automation Conference (ASP-DAC), 2015 20th Asia and South Pacific*. IEEE, 2015.

[17] Z. Zhao, Z. Wang, Z. Ying, S. Dhar, R. T. Chen, and D. Z. Pan, "Logic synthesis for energy-efficient photonic integrated circuits," in *Proceedings of the 23rd Asia and South Pacific Design Automation Conference*. IEEE Press, 2018.

[18] R. Matsuo, J. Shiomi, T. Ishihara, H. Onodera, A. Shinya, and M. Notomi, "Bdd-based synthesis of optical logic circuits exploiting wavelength division multiplexing," in *Proceedings of the 24th Asia and South Pacific Design Automation Conference*. ACM, 2019, pp. 203–209.

[19] Z. Zhao, D. Liu, Z. Ying, B. Xu, R. T. Chen, and D. Z. Pan, "Exploiting wavelength division multiplexing for optical logic synthesis," in *Proceedings of the 2019 Design, Automation and Test in Europe Conference (DATE)*. IEEE, 2019.

[20] Y. Xie, J. Xu, J. Zhang, Z. Wu, and G. Xia, "Crosstalk noise analysis and optimization in 5× 5 hitless silicon-based optical router for optical networks-on-chip (ONoC)," *Journal of Lightwave Technology*, 2012.

[21] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," *Physical review letters*, 1994.

[22] A. Ribeiro, A. Ruocco, L. Vanacker, and W. Bogaerts, "Demonstration of a 4 × 4-port universal linear circuit," *Optica*, 2016.

[23] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, "Optimal design for universal multiport interferometers," *Optica*, 2016.

[24] D. A. Miller, "Perfect optics with imperfect components," *Optica*, 2015.

[25] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund *et al.*, "Deep learning with coherent nanophotonic circuits," *Nature Photonics*, 2017.

[26] Z. Zhao, D. Liu, M. Li, Z. Ying, L. Zhang, B. Xu, B. Yu, R. T. Chen, and D. Z. Pan, "Hardware-software co-design of slimmed optical neural networks," in *Proceedings of the 24th Asia and South Pacific Design Automation Conference*. ACM, 2019.

[27] L. Vivien *et al.*, "Zero-bias 40gbit/s germanium waveguide photodetector on silicon," *Optics express*, 2012.

[28] C. Wang *et al.*, "Integrated lithium niobate electro-optic modulators operating at cmos-compatible voltages," *Nature*, 2018.

[29] R. E. Bryant, "Symbolic Boolean Manipulation with Ordered Binary-Decision Diagrams," *ACM Computing Surveys (CSUR)*, 1992.

[30] J. Y. Yen, "Finding the k shortest loopless paths in a network," *management Science*, 1971.