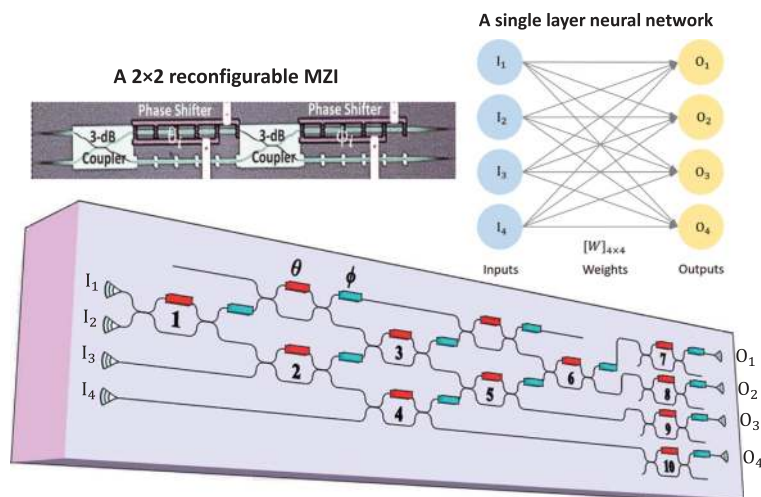


# A Single Layer Neural Network Implemented by a $4 \times 4$ MZI-Based Optical Processor

Volume 11, Number 6, December 2019

Farhad Shokraneh  
Simon Geoffroy-Gagnon  
Mohammadreza Sanadgol Nezami  
Odile Liboiron-Ladouceur, *Senior Member, IEEE*



A  $4 \times 4$  MZI-based linear optical processor implementing the weight matrix  $[W]_{4 \times 4}$  of a single layer neural network.

DOI: 10.1109/JPHOT.2019.2952562

# A Single Layer Neural Network Implemented by a $4 \times 4$ MZI-Based Optical Processor

Farhad Shokraneh , Simon Geoffroy-Gagnon ,  
Mohammadreza Sanadgol Nezami ,  
and Odile Liboiron-Ladouceur , *Senior Member, IEEE*

Department of Electrical and Computer Engineering, McGill University, Montréal,  
QC H3A 0G4, Canada

DOI:10.1109/JPHOT.2019.2952562

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see  
<https://creativecommons.org/licenses/by/4.0/>

Manuscript received October 25, 2019; accepted November 5, 2019. Date of publication November 11, 2019; date of current version December 17, 2019. This work was supported in part by the Canadian Microelectronic Corporation for the subsidized MPW fabrication through Advanced Micro Foundry, and in part by the partial Canada Research Chairs program. Corresponding author: Farhad Shokraneh (e-mail: farhad.shokraneh@mail.mcgill.ca).

**Abstract:** Implementing any linear transformation matrix through the optical channels of an on-chip reconfigurable multiport interferometer has been emerging as a promising technique for various fields of study, such as information processing and optical communication systems. Recently, the use of multiport optical interferometric-based linear structures in neural networks has attracted a great deal of attention. Optical neural networks have proven to be promising in terms of computational speed and power efficiency, allowing for the increasingly large neural networks that are being created today. This paper demonstrates the experimental analysis of programming a  $4 \times 4$  reconfigurable optical processor using a unitary transformation matrix implemented by a single layer neural network. To this end, the Mach-Zehnder interferometers (MZIs) in the structure are first experimentally calibrated to circumvent the random phase errors originating from fabrication process variations. The linear transformation matrix of the given application can be implemented by the successive multiplications of the unitary transformation matrices of the constituent MZIs in the optical structure. The required phase shifts to construct the linear transformation matrix by means of the optical processor are determined theoretically. Using this method, a single layer neural network is trained to classify a synthetic linearly separable multivariate Gaussian dataset on a conventional computer using a stochastic optimization algorithm. Additionally, the effect of the phase errors and uncertainties caused by the experimental equipment inaccuracies and the device components imperfections is also analyzed and simulated. Finally, the optical processor is experimentally programmed by applying the obtained phase shifts from the matrix decomposition process to the corresponding phase shifters in the device. The experimental results show that the optical processor achieves 72% classification accuracy compared to the 98.9% of the simulated optical neural network on a digital computer.

**Index Terms:** Programming a reconfigurable linear optical processor, optical neural networks, optical matrix multiplication structures, neuromorphic photonics.

## 1. Introduction

Neural networks (NNs) have been impressively emerging since they have proven to be promising in solving complex computational functions [1]–[3]. The growing number of research in this domain confirms that photonic implementation of neural networks has made it possible to perform fast and

power efficient functionalities and matrix multiplications. For example, the implemented optical NN in [3] operates at 20 MHz with  $10 \mu\text{W}/\text{mm}^2$  power density whereas most modern graphics processing units (GPUs), which are presently used in most NNs, have a power density of 0.3 to  $0.408 \text{ W}/\text{mm}^2$  [4], [5]. In this regard, on-chip multiport reconfigurable interferometers as optical processors with small footprint can implement linear operations between several optical channels. They can be used to efficiently perform complex matrix multiplications in NNs by exploiting the inherent parallelism of optics which provides a linear time complexity as compared to digital NNs which scale matrix multiplications by polynomial time complexity [6]. Such structures have been employed in different applications, such as optical networking [7], quantum photonics [8], [9] and microwave photonics [10]. Multiport programmable MZI-based interferometers which implement a unitary transformation matrix between the  $N$  input and  $N$  output ports are used as a new programming method in various applications [11]–[14]. Their reconfigurability allows for performing complex and precise linear optical functions in information processing applications, such as optical neural networks [2], [14], [15]. Each MZI in a multiport reconfigurable interferometric structure is constructed by two 3-dB directional couplers and one phase shifter between them, on one of their connecting arms, and another phase shifter on one of the external arms of the second directional coupler [16]. Consequently, the structure is a mesh of reconfigurable MZIs each of which representing a special unitary group of degree two (SU2) experimentally programmed as a multiport reconfigurable optical processor [9], [17], [18]. The linear transformation matrix of a reconfigurable multiport optical processor is obtained by the product of the unitary matrices of its constituent MZIs. The linear transformation matrix of a given application can also be decomposed into the unitary matrices of the MZIs in the optical structure and the related phase shifts can be extracted. These phase shifts can be used for programming the optical processor [16].

NNs extensively exploit matrix multiplications to compute forward propagation of a system, i.e., the multiplication of the weight matrix by the input vector of each layer [19]. An integrated optical processor being able to compute an  $N \times N$  vector-matrix multiplication can therefore lend itself well to applications in neural networks, such as image recognition [20], teaching robots to accurately throw objects [21], controlling self-driving cars [22], and protein-folding [23]. This paper presents the theoretical and experimental configuration and programming process of a  $4 \times 4$  reconfigurable MZI-based optical processor to implement fast and energy efficient matrix multiplications in a single layer NN. Using a stochastic optimization method, the NN is trained to classify a synthetic linearly separable multivariate Gaussian dataset by generating a  $4 \times 4$  weight matrix on a digital computer which is then implemented by the fabricated  $4 \times 4$  optical processor. The experimental implementation is done by programming the MZIs using the calculated phase shifts from the optimization process. This work demonstrates how the classification accuracy of a single layer NN implemented by the optical processor can be affected by the phase shifts precision and the insertion loss of the constituent MZIs. The results show that the classification performance of the device can be degraded by various sources such as phase errors, thermal crosstalk between the MZIs, bias voltage inaccuracy, and the optical losses in the device. The experimental results demonstrate that the optical processor achieves 72% accuracy in classifying 50 data samples which were correctly classified through simulation.

## 2. Background

An ideal  $N \times N$  multiport reconfigurable MZI-based interferometer represents the so-called special unitary group of degree  $N$  ( $\text{SU}(N)$ ). It consists of  $n$  MZIs within  $N$  optical channels making up a unitary transformation matrix  $[T_{\text{SU}(N)}]$ . Each MZI in the structure is composed of two 3-dB directional couplers with a thermal-based phase shifter ( $\theta$ ) on the upper internal arm of the MZI and another one ( $\phi$ ) on its upper output. The internal phase shifter adjusts the coupling ratio at the output of the MZI. The second phase shifter controls the relative phase of the MZI outputs. The unitary transformation matrix of the  $2 \times 2$  reconfigurable MZI,  $[D_{\text{MZI}}]$ , can be determined by the product of

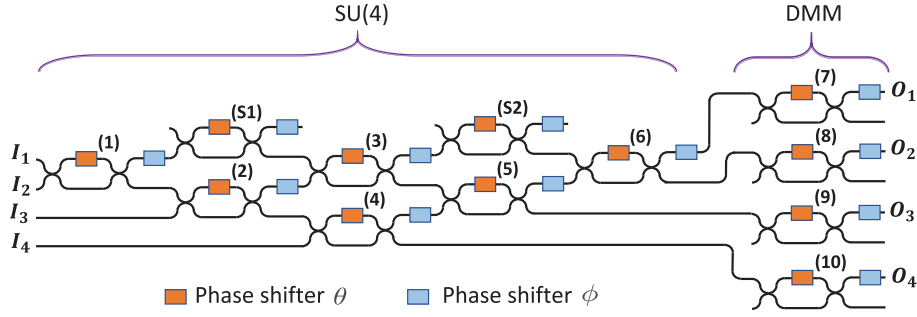


Fig. 1. Illustration of the  $4 \times 4$  MZI-based reconfigurable linear optical processor with four input ( $I_1$ ,  $I_2$ ,  $I_3$  and  $I_4$ ) and four output ports ( $O_1$ ,  $O_2$ ,  $O_3$  and  $O_4$ ). The SU(4) consists of MZIs (1) to (6) which can be extended to a large structure by the MZIs S1 and S2. The DMM section of the  $4 \times 4$  structure is composed of MZIs (7) to (10).

the transformation matrices of its directional couplers and phase-shifters [16]. Thus,

$$[D_{MZI}] = \begin{pmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{pmatrix} = j e^{j\left(\frac{\theta}{2}\right)} \begin{pmatrix} e^{j\phi} \sin\left(\frac{\theta}{2}\right) & e^{j\phi} \cos\left(\frac{\theta}{2}\right) \\ \cos\left(\frac{\theta}{2}\right) & -\sin\left(\frac{\theta}{2}\right) \end{pmatrix}, \quad (1)$$

where  $u_{pq}$ , ( $p$  and  $q \in \{1, 2\}$ ) represent the field transmission between the input and output ports of the  $2 \times 2$  reconfigurable MZI. To construct an SU( $N$ ) matrix, the unitary matrix of each MZI is presented on a two-dimensional subspace within an  $N$ -dimensional Hilbert space ( $H_{N \times N}$ ) [14], [16]. Each MZI is experimentally configurable to set a certain power splitting ratio and a relative phase at its outputs. The unitary transformation matrix of an SU( $N$ ) can be determined by the product of the transformation matrices of its  $n$  constituent MZIs, represented by  $[D_{MZI}^{(n)}]_{H_{N \times N}}$  where  $n = N(N-1)/2$  [16].

Fig. 1 depicts the schematic of the designed  $4 \times 4$  optical processor. The structure is composed of an SU(4) section followed by a diagonal matrix multiplication (DMM) section for controlling the optical power of the output ports of the device. Depending on the application, the DMM section can also be used for extending the  $4 \times 4$  optical processor to a larger structure through cascading the structure to a following section for a complete matrix singular value decomposition [13].

As shown in Fig. 1, the SU(4) section contains the MZIs labelled (1) to (6) constructing the unitary transformation matrix  $[T_{SU(4)}]$ , while the latter section (DMM) consists of MZIs (7) to (10) implementing a non-unitary diagonal matrix  $[\Sigma]$ . The structure performs a linear transformation  $[W]_{4 \times 4}$  between an input matrix  $[I]_{4 \times 1}$  and output matrix  $[O]_{4 \times 1}$  based on the linear optical wave interactions in the physical device [13]. The linear transformation matrix of the structure is determined by

$$[W]_{4 \times 4} = [\Sigma] \cdot [T_{SU(4)}] = \begin{pmatrix} u_{11}^{(7)} & 0 & 0 & 0 \\ 0 & u_{11}^{(8)} & 0 & 0 \\ 0 & 0 & u_{11}^{(9)} & 0 \\ 0 & 0 & 0 & u_{11}^{(10)} \end{pmatrix} \cdot \begin{pmatrix} U_{11} & U_{12} & U_{13} & U_{14} \\ U_{21} & U_{22} & U_{23} & U_{24} \\ U_{31} & U_{32} & U_{33} & U_{34} \\ U_{41} & U_{42} & U_{43} & U_{44} \end{pmatrix}, \quad (2)$$

where  $U_{kl}$  ( $k$  and  $l \in \{1, 2, 3, 4\}$ ) are elements of  $[T_{SU(4)}]$  which can be determined by the product of unitary matrices in a four-dimensional Hilbert space ( $[D^{(n)}]_{4 \times 4}$ ) for  $n \in \{1, 2, \dots, 6\}$  as the following

$$[T_{SU(4)}] = [D^{(6)}]_{4 \times 4} \cdot [D^{(5)}]_{4 \times 4} \cdot [D^{(4)}]_{4 \times 4} \cdot [D^{(3)}]_{4 \times 4} \cdot [D^{(2)}]_{4 \times 4} \cdot [D^{(1)}]_{4 \times 4}. \quad (3)$$

To find the required phase shifts for a given application, the weight matrix is decomposed into a unitary matrix  $[T_{SU(4)}]$  and a complex diagonal matrix  $[\Sigma]$  from which the required phase shifts can be determined. One way is to decompose  $[T_{SU(4)}]$  through its successive multiplications by

$[D^{(n)}]_{4 \times 4}^{-1}$ . In this method, each multiplication step sets a specific off-diagonal element to zero in the resultant matrix [16]. By making an off-diagonal element zero at each step of the successive multiplications, the required phase shifts in the inverse transformation matrix of the corresponding MZI can be calculated [14], [16]. However, this method may result in complex phase values, which are not implementable experimentally. To address this issue, a stochastic optimization algorithm is proposed in the next section, which provides the required real valued phases to optically implement the weight matrix.

### 3. Application of the $4 \times 4$ Optical Processor in a Single Layer Neural Network

Programming the optical processor requires high accuracy in the controlling strategy of the phase shifters in all MZIs. The weight matrix of the optical NN can be defined by the successive multiplications of the rotational matrices implemented by the MZIs in the structure. Therefore, one can conclude that the generated digital weight matrix should be precisely converted such that the resultant matrix can be implemented by the optical processor using the phase shifters of the MZIs. However, converting an arbitrary linear matrix to successive product of a limited number of rotational matrices, i.e., the unitary transformation matrices of the MZIs in the optical device, leads to inaccuracies. One way to tackle the precision issue in converting the digital weight matrix  $[W]_{4 \times 4}$  to the linear transformation matrix of the optical processor is to use a complex structure with more MZIs [11], [15]. However, this strategy will increase the optical channel losses, the footprint, and the complexity of programming the device. For instance, an  $8 \times 8$  structure requires twenty eight MZIs in its SU(8) section. Another approach used in this work is to employ a stochastic optimization algorithm to train the NN to classify a linearly separable multivariate Gaussian dataset to obtain the required phase shifts for implementing the optical NN experimentally [24]. According to equation (3),  $[D^{(n)}]_{4 \times 4}$  matrices are multiplied together to obtain the  $4 \times 4$  unitary transformation matrix  $[T_{SU(4)}]$ . Consequently, the different elements in the resultant matrix are related to one another, which needs to be considered in the optimization algorithm. In this sense, a single layer NN is implemented by the  $4 \times 4$  optical processor as shown in Fig. 2a. It is essential to note that a single layer NN yields the same maximum value whether or not a non-linear function is implemented if a monotonic non-linear function is used. The monotonicity criterion helps the NN to converge more easily into a more accurate digital NN [25]. Thus, a non-linear activation function after the matrix multiplication is not used in this research work. However, in a complex optical NN with several hidden layers, using non-linear activation functions is indispensable. A multi layer NN without non-linear activation functions can always be reduced to a single layer NN which degrades the performance of the whole system [26]. A non-linear activation function can be implemented either analytically by a computer [15], or experimentally in optics [27], [28]. To evaluate the optical processor in terms of classification accuracy, a synthetic data set is created with four features distributed across  $[0, 1]$  within the four classes. Each class is represented by a differently colored Gaussian distribution, each of which populated with a set of four-dimensional positive real valued points represented by  $[\mathbf{I}] \in \mathbb{R}_+^4$ , as illustrated in Fig. 2b.

The proposed stochastic optimization algorithm for the single layer NN is based on the topology of the device which can implement a unitary or sub-unitary weight matrix. This method is different from the stochastic gradient descent algorithm which is commonly used in neural networks with several layers [29]. In the case of large  $N$  values, the stochastic optimization algorithm takes longer to converge compared to a gradient descent algorithm, but still gives a high classification accuracy. In this regard, the algorithm starts by assigning random phases to the phase shifters and calculating the generated weight matrix through equation (2). The optical weight matrix is constructed using the phase shifts  $\theta_i$  and  $\phi_i$  for  $i = 1, 2, 3, \dots, 10$ , forming the interferometric structure. To process the forward propagation of a single sample in this NN, the input sample  $[\mathbf{I}] \in \mathbb{R}_+^4$  representing input optical power levels of the device is multiplied by the optical weight matrix  $[W]_{4 \times 4}$ . The absolute

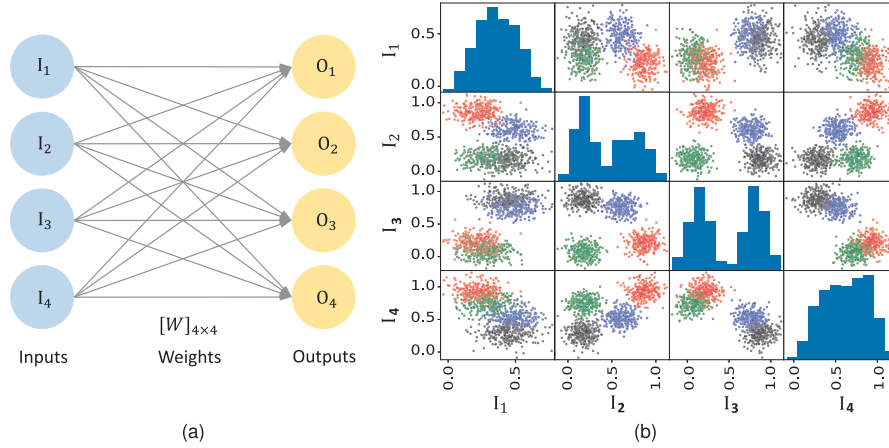


Fig. 2. Simulated NN on a digital computer: (a) Illustration of a single layer neural network. (b) Scatter matrix of the synthetic dataset consisting of four classes each represented by differently colored Gaussian distributions with four features.

TABLE 1

Calculated Phase Shifts of the Phase Shifters for Programming the  $4 \times 4$  Optical Processor

MZI	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$\theta_i$ (Rad)	0.7156	2.0627	2.9221	0.1850	2.2975	0.7777	2.1495	1.7414	1.3713	1.3382
$\phi_i$ (Rad)	1.3105	5.0834	3.9101	4.5238	1.6935	0.6633	5.5455	5.9179	2.2976	2.9633

value of the resultant matrix yields an output optical power levels vector  $[\mathbf{O}] \in \mathbb{R}_+^4$  that is used to predict the classification category. The output element with the highest optical power designates the predicted class of the data sample. The classes of the separate Gaussian distributions are one-hot encoded such that the ground truth vector for a single sample is  $[\mathbf{O}_{\text{true}}] \in \mathbb{R}_+^4$ , with a different value set to 1 for each class [30]. The classification of a sample is carried out based on which element of the output vector  $[\mathbf{O}]$  has the maximum value. A sample is correctly classified if the index of the maximum element in the prediction vector is equal to the index of the 1 value in the one-hot encoded ground truth vector  $[\mathbf{O}_{\text{true}}]$  (i.e., if  $\text{argmax}([\mathbf{O}]) == \text{argmax}([\mathbf{O}_{\text{true}}])$ , the classification is correct). For instance,  $[\mathbf{O}] = [0, 0, 0, 1]$  is the correct classification if the input vector  $[\mathbf{I}]$  belongs to the fourth class. To mitigate the effect of experimental imperfections which will be discussed later, an additional condition is defined to record correct classifications. This condition ignores a classification in which the difference between a maximum output value  $O_{\text{max}}$  and its second maximum value  $O_{\text{secondmax}}$  is less than a decision threshold  $\xi$ . Thus, a classification is counted as correct if and only if:

$$O_{\text{max}} - O_{\text{secondmax}} > \xi. \quad (4)$$

Before properly adjusting the phase values of the optical processor, the classification accuracy initially tends to be quite poor. By stochastically changing the phase values, it is tested whether the new optical weight matrix achieves a higher accuracy than the previous one. If it is higher, the applied phase is stored, otherwise, it is reverted to the previous phase value. Over many iterations, the optimal phases  $\theta_i$  and  $\phi_i$  can be determined. Consequently, an arbitrary four-feature-four-target dataset can be classified using the  $4 \times 4$  optical processor. The dataset is well classified with a single layer NN implemented on a conventional computer, which means that the dataset is linearly separable. This allows for the evaluation of the optical NN performance compared to the simulated digital NN in terms of classification accuracy and the impacting experimental parameters on it. Table 1 shows the resultant phases  $\theta_i$  and  $\phi_i$  to implement the weight matrix  $[W]_{4 \times 4}$  from the simulated NN.

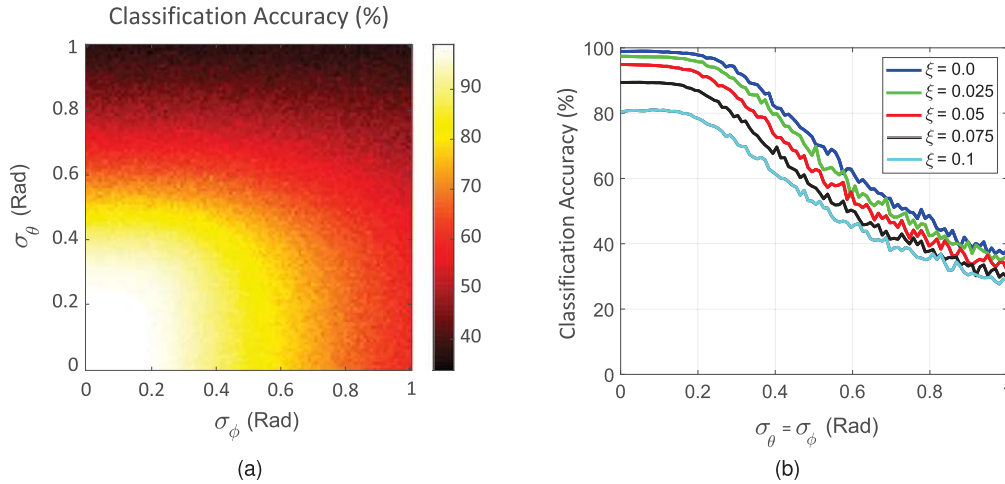


Fig. 3. Simulation results of the optical processor classification accuracy when classifying four classes: (a) Classification accuracy for different phase error standard deviations of the phase shifters,  $\sigma_\theta$  and  $\sigma_\phi$ , for  $\xi = 0$ . (b) classification accuracy as a function of phase error standard deviation of the phase shifters,  $\sigma_\theta = \sigma_\phi$  for different threshold values  $\xi$ . The results were created by taking 200 noisy phase samples and calculating classification accuracy of the resultant matrices for different  $\xi$  values.

The resultant matrix from the simulated NN corresponding to linear transformation matrix of the  $4 \times 4$  structure ( $[W]_{4 \times 4}$  is given by

$$[W]_{4 \times 4} = \begin{pmatrix} 0.4537 + 0.0634i & -0.5334 + 0.2092i & -0.1614 - 0.3148i & -0.3219 + 0.0813i \\ -0.1616 + 0.4619i & 0.0177 + 0.5501i & 0.1671 + 0.0253i & 0.0566 - 0.1039i \\ -0.1107 - 0.1541i & 0.0583 + 0.0828i & -0.1295 + 0.0843i & -0.3601 - 0.4485i \\ 0.1162 - 0.2735i & -0.0434 + 0.1023i & 0.5294 + 0.0232i & -0.0478 - 0.0316i \end{pmatrix} \quad (5)$$

After training, the implementation of the NN using the simulated optical matrix with the dataset on a computer and setting  $\xi = 0$  as an example results in 98.9% classifying accuracy. Applying the phases to the phase shifters with exact accuracy is a challenging task in the measurements due to the fluctuation errors from the voltage sources and the thermal crosstalk between the MZIs, and fabrication process variations leading to undetermined phase offsets. Additionally, the losses in optical channels and MZIs challenges the signal-to-noise ratio (SNR) at the receiver. As a result, the classification accuracy of the optical processor is degraded. Fig. 3a demonstrates the accuracy as a function of different phase error standard deviations,  $\sigma_\theta$  and  $\sigma_\phi$ . Fig. 3b shows the simulation results of the degradation in the classification accuracy of the weight matrix as a function of the same standard deviation in the phase noise, using 200 separate noisy phase samples for different  $\xi$  values. It can be inferred from the figures that how the phase errors and the threshold value  $\xi$  affect the classification accuracy of the weight matrix which highlights the importance of accurate and stable experimental setup and the related factors for testing the device. Fig. 4 illustrates the simulation results of the classification accuracy as a function of phase error standard deviation for different IL values of each MZI in the optical processor.

According to this figure, the IL of every MZI in the device also plays a determining role in the classification accuracy. For instance, in the case of 0.5 dB loss for each MZI, the IL of the structure ranges from 1 dB to 3 dB, the classification accuracy is reduced to approximately 75% with no phase error.

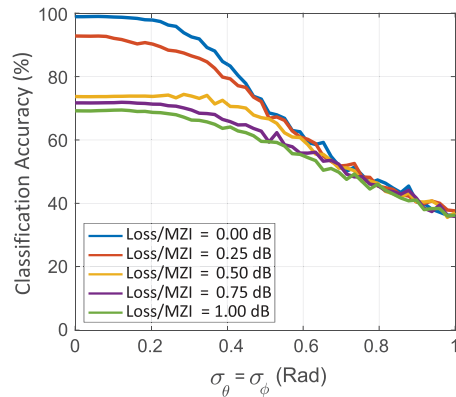


Fig. 4. Simulation results of classification accuracy as a function of phase error standard deviations of the phase shifters,  $\sigma_\theta = \sigma_\phi$  with  $\xi = 0$  for different IL values of every MZI in the structure.

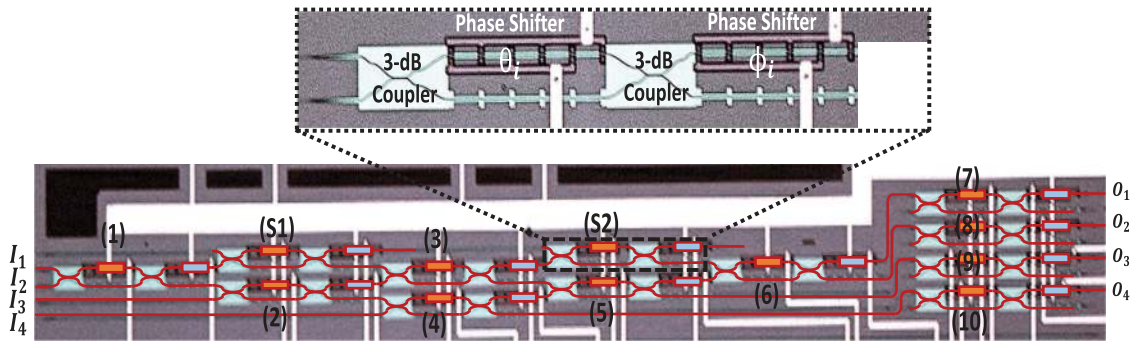


Fig. 5. Microscope image of the fabricated  $4 \times 4$  MZI-based linear optical processor. Inset shows one of the reconfigurable MZIs in the structure. MZIs (1) to (6) implement the unitary transformation matrix  $[T_{SU(4)}]$ , whereas MZIs (7) to (10) construct  $[\Sigma]$ . The MZIs (S1) and (S2) in the SU(4) section are in bar states for the  $4 \times 4$  optical processor.

#### 4. Experimental Programming of the $4 \times 4$ Optical Processor

The device in this work is designed for operating at 1310 nm of wavelength and exploiting a  $220 \text{ nm} \times 420 \text{ nm}$  cross-sectional area SOI ridge waveguide with a 90 nm slab fabricated using 193 nm DUV lithography. As explained earlier, the reconfigurable MZI-based optical processor is a mesh of  $2 \times 2$  tunable MZIs, each of which having two phase shifters  $\theta_i$  and  $\phi_i$  to control the power and the relative phase of its outputs, respectively. Fig. 5 is a microscope image of the fabricated  $4 \times 4$  reconfigurable linear optical processor. The device can be reconfigured by applying the required DC voltages to the phase shifters of the MZIs. An off-chip VCU is used to adjust the required DC voltages for the phase shifters. As shown in Fig. 1, the structure can be developed to a larger multiport reconfigurable MZI-based processor using the MZIs (S1) and (S2). These two MZIs in the  $4 \times 4$  structure are tuned to be in their bar states functioning as simple waveguides.

To program the device experimentally based on the calculated phase shifts given in Table 1, it is essential to determine the required DC voltages to be applied to the corresponding phase shifters. For the external phase shifters with the phase shifts  $\phi_i$ , an optical vector analyzer (LUNA OVA 5013) is used to determine the required DC voltages  $V_{prog, \phi_i}$ . In the case of the internal phase shifters of the MZIs,  $\theta_i$ , they control the optical power splitting ratio (transmission) at the outputs of the MZIs.

Fig. 6 illustrates the schematic of the experimental setup used to assess the prediction accuracy of the fabricated optical processor. The continuous wave (CW) at 1310 nm is generated by a tunable O-band laser and passes through an O-band booster optical amplifier (BOA) which amplifies the



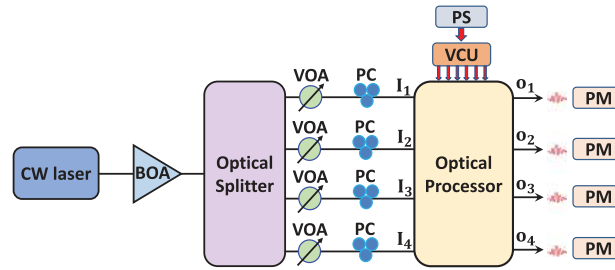


Fig. 6. Schematic of the experimental setup for sending an input optical power vector  $[I]$  from the dataset into the optical processor to predict the class of the output vector  $[O]$ . CW: tunable O-band laser; BOA: O-band booster optical amplifier; VOA: variable optical attenuators; PC: polarization controller; PS: power supply; VCU: electrical DC voltage controlling unit; PM: optical power meter.

input power such that its output optical power is set to 20 dBm. The optical signal is then split into four channels using a  $1 \times 4$  optical splitter. The optical signal in each channel passes through a variable optical attenuator (VOA) which adjusts the amplitude of the input optical signal based on the requirements of the data samples being tested. Each channel employs a polarization controller (PC) to optimize the state of polarization to the TE mode required by the device. An optical fiber array couples the light onto the input vertical grating couplers of the device. For tuning the phase shifters of the MZIs, a power supply (PS) is used to provide the electrical DC voltage for the off-chip VCU which regulates the voltage of the phase shifters. The optical signals from the four output vertical grating couplers are coupled to the fiber array and then monitored using optical power meters (PM).

Before programming the device for a given application, all MZIs need to be characterized for calibration purposes to mitigate the effects of input phase errors and fabrication process errors [12], [13]. The calibration process is carried out based on the topology of the structure, particularly, the SU(4) section in the device and the experimental setup from the input ports to the outputs of the device. The simplest path on which each MZI is located is chosen for its configuration and the corresponding calibrated  $\phi_i$  is applied to the external phase shifter of the MZI. The phase shift  $\theta$  determines the transmission at the bar port of the MZI i.e.,  $0 \leq T_{BP} \leq 1$ . The experimental configuration of an MZI in the device represents the power transmission at its bar port as a function of the applied DC bias voltage. The calibration scheme is based on the topology of the SU(4) section choosing the simplest path for configuration of each MZI. It starts from MZI (4) on the path  $I_4-O_4$  of the structure shown in Fig. 1. The configuration of MZI (4) in its CS allows for calibrating MZIs (5) and (6) on the paths  $I_4-O_3$  and  $I_4-O_2$ , respectively. At this point, it is also possible to configure MZI (7), (8), (9) and (10) by setting MZIs (4), (5), and (6) which are previously configured to the required states on the corresponding paths. In the next step MZIs (2), (3), and  $S_2$  on the paths  $I_3-O_3$ ,  $I_3-O_2$  and  $I_3-O_1$  are configured, respectively, while the related MZIs are in the required states. Eventually, MZIs (1) and  $S_1$  on the paths  $I_2-O_2$  and  $I_2-O_1$  are configured in a similar way. Fig. 7a illustrates bar state (BS) and cross state (CS) of a single MZI. Fig. 7b shows the measured optical power levels of a single MZI for the bar and cross states,  $P_{BS}$  and  $P_{CS}$ , as a function of the bias voltage and the respective phase shift  $\theta$ , along with the extinction ratio (ER).

In the  $4 \times 4$  optical processor shown in Fig. 5, the measured optical power levels in BS and CS of every MZI through its configuration, is exploited to program the MZI by obtaining the target optical power  $P_{prog,\theta_i}$  at its bar port as the following

$$P_{prog,\theta_i} = P_{CS,i} + T_{BP,i} \cdot ER_i, \quad (6)$$

where  $P_{prog,\theta_i}$  denotes the required optical power levels for the target phase values in programming the device shown in Table 1.  $P_{CS,i}$  is the transmitted optical power in cross state of a given MZI measured at the corresponding output port of the  $4 \times 4$  structure.  $T_{BP,i} = \sin^2(\frac{\theta_i}{2})$  expresses the optical power transmission at the bar port of the MZIs.  $ER_i = P_{BS,i} - P_{CS,i}$  represents the extinction

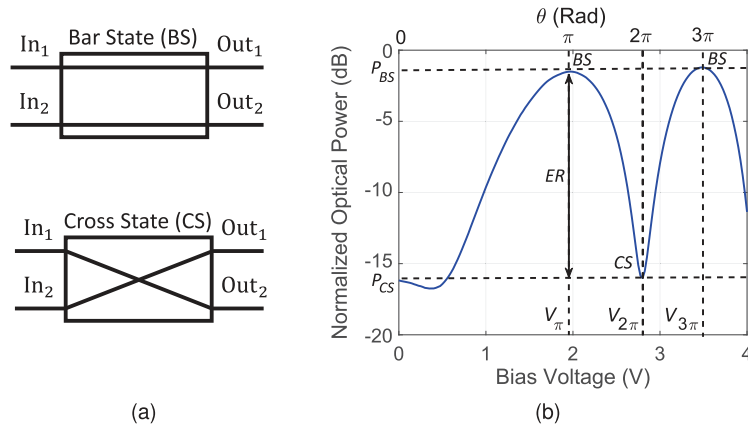


Fig. 7. (a) Schematic of BS and CS of a  $2 \times 2$  MZI. (b) Measured optical power levels of a single MZI in BS and CS represented by  $P_{BS}$  and  $P_{CS}$ , as a function of the bias voltage and the respective phase shift  $\theta$ , along with the extinction ratio (ER).

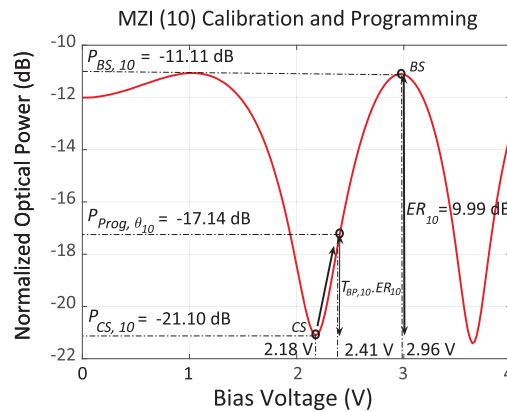


Fig. 8. Experimental calibration and programming process of MZI (10) on the path  $I_4-O_4$  when MZI (4) is set in its BS. The transmission at the bar port of the MZI as a function of bias voltage is exploited to determine the target power level ( $P_{prog, \theta_{10}}$ ) at the related output port and the required bias voltage.

ratio of the MZIs, where  $P_{BS,i}$  is the optical transmitted power in BS measured at the related output power of the device, as shown in Fig. 7. By setting the  $i_{th}$  MZI in its CS and increasing the DC voltage applied to  $\theta_i$  corresponding to  $T_{BP,i}$ , the optical power of the related output port increases to the calculated  $P_{prog, \theta_i}$ . The measured DC voltage that adjusts this power level in the bar port of the MZI is recorded as  $V_{prog, \theta_i}$ . Fig. 8 depicts the measurement results used for the calibration and programming of MZI (10) on the path  $I_4-O_4$  while MZI (4) is in BS.

Similarly, all MZIs are programmed using the configuration protocol explained earlier. Table 2 lists the corresponding DC bias voltages of the phase shifters  $\theta_i$  for CS of the MZIs represented by  $V_{CS,i}$ , and the corresponding measured optical power levels  $P_{CS,i}$  along with the required transmission  $T_{BS,i}$  at the bar port of each MZI measured at the respective output of the device.

The variation in the voltage values are associated with the random phase offsets of the MZIs due to fabrication process variations. The various optical power levels in  $P_{CS,i}$  are attributed to the difference in losses between the optical paths on which a given MZI is located. For instance, MZI (4) on the input to output path  $I_4-O_4$  has higher power levels due to fewer MZIs leading to lower losses on the path, as shown in Fig. 1. On the other hand, MZIs (2) and (3) on the input to output paths  $I_3-O_3$  and  $I_3-O_2$ , respectively, experience much lower power levels as a result of higher losses and more MZIs located on the corresponding optical paths. Table 3 summarizes the

TABLE 2  
Measured Parameters Through the Experimental Configuration of the MZIs in the  $4 \times 4$  Optical Processor to Determined the Required  $P_{CS,i}$  Values

MZI	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$V_{CS,i}$ (V)	2.52	1.06	2.22	2.30	2.72	2.80	2.50	2.68	2.80	2.18
$P_{CS,i}$ (dBm)	-37.45	-46.21	-47.77	-13.18	-23.41	-28.29	-19.76	-29.03	-17.00	-21.10
$T_{BS,i}$	0.123	0.736	0.988	0.009	0.832	0.144	0.774	0.585	0.401	0.385
$ER_i$ (dB)	14.75	21.46	24.24	9.42	9.93	8.77	9.78	9.94	10.02	9.99
$P_{prog,\theta_i}$ (dBm)	-35.64	-30.41	-23.82	-13.00	-15.15	-27.03	-12.20	-23.22	-12.98	-17.14

TABLE 3  
Measured DC Bias Voltages and the Corresponding Current Values of the Phase Shifters for Programming the  $4 \times 4$  Optical Processor Using the NN Dataset

MZI	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$V_{prog,\theta_i}$ (V)	2.60	1.51	2.91	2.38	2.98	2.76	2.86	2.96	3.20	2.41
$I_{prog,\theta_i}$ (mA)	14.01	8.65	15.37	12.99	15.65	14.71	15.16	15.58	16.57	13.96
$V_{prog,\phi_i}$ (V)	1.26	2.49	2.18	2.35	1.44	0.89	2.60	2.69	1.68	1.90
$I_{prog,\phi_i}$ (mA)	7.26	13.50	12.03	12.85	8.23	5.18	13.82	14.41	9.51	10.64

measured bias voltages and the corresponding current values of the phase shifts  $\theta_i$  and  $\phi_i$  for programming the  $4 \times 4$  optical processor to implement the weight matrix  $[W]_{4 \times 4}$  of the simulated NN given by equation (5). According to Table 3, the total power consumption of the programmed optical processor for this application is approximately 609 mW.

To experimentally construct the linear transformation matrix of the NN application, the required DC bias voltages given in Table 3 are applied to the corresponding phase shifters. After programming the device for the application, 50 correctly classified data samples through simulation were applied to the optical processor. The 50 data samples were chosen such that all possible paths from the inputs to the outputs of the device were covered, which allows for the evaluation of the device performance in terms of classification accuracy, loss and phase errors. The generated input values by the application, range from zero to one with increments of 0.1. However, from experimental point of view, it is essential to quantize these values to the possible range of the optical input power. The use of the BOA made it possible to linearly quantize the four input power levels from  $-5$  dBm to 4 dBm with increments of 1 dB.

Fig. 9 shows the experimental results of the optical processor used in implementing a single layer NN for classifying the data samples compared to that of generated by a computer. According to the experimental results, the programmed  $4 \times 4$  optical processor could identify 36 out of 50 classification data samples correctly (72% accuracy) compared to that of simulated NN on a digital computer with no phase errors and imperfections. The experimental misclassification of the four classes shown in Fig. 9 are associated with the random phase errors caused by different sources and the IL of the MZIs in the device. According to Fig. 3b, the classification accuracy will remain at approximately 98.8% as long as the phase error standard deviation remains below 0.1 Rad. The precision of the VCU used in this work is 10 mV. From our previous work in [16], a voltage inaccuracy below 10 mV in a phase shifter corresponds to a phase deviation (the worst-case scenario) of approximately 0.032 Rad. Therefore, the accumulated phase error due to bias voltage noises of the phase shifters in the programmed optical processor is approximately 0.67 Rad. For the phase error less than 0.1 Rad, the precision of the voltage regulators must be higher than 1.56 mV [16]. There is also thermal crosstalk between the individual MZIs (e.g., between MZIs (3) and (4)) which leads to temperature fluctuations in the order of 1.4 Kelvin [31], which corresponds

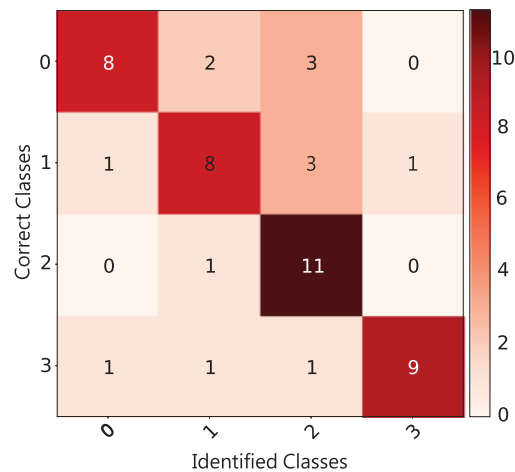


Fig. 9. Experimental classification results of 50 data samples compared to the correctly classified data samples through simulation. The identified (measured) classes achieve a classification accuracy of 72% compared to the correct ones simulated by a digital computer.

to a phase error of 0.176 Rad [16]. Considering all MZIs and their phase shifters, the thermal crosstalk phase error is higher, degrading the classification accuracy of the optical processor. Furthermore, the degradation in the classification accuracy is also attributed to the IL of each MZI in the structure which is in accordance with the simulation results shown in Fig. 4. An improvement for lowering thermal crosstalk is to use deep trenches to isolate the phase shifters of the MZIs [32], [33]. Additionally, designing a structure with MMI based MZIs would make it possible to reduce the loss and enhances the robustness of the device against the possible errors. Finally, a better fabrication process can also improve the classification accuracy by lowering the optical loss and less fabrication process variations.

## 5. Conclusion

A  $4 \times 4$  MZI-based optical processor is investigated, both theoretically and experimentally. The analytical implementation of an arbitrary unitary matrix by means of the optical processor is demonstrated through a stochastic optimization algorithm to determine the required phase shifts in the constituent MZIs. Furthermore, the simulation results of a single layer optical NN could achieve 98.9% classification accuracy in a linearly separable, synthetic dataset. The integrated  $4 \times 4$  MZI-based optical processor with a compact footprint can be embedded within a computer architecture as an accelerator to compute matrix multiplications. It was investigated that the classification accuracy of the device can be degraded by experimental and fabrication imperfections causing phase errors and optical losses. The experimental results of the optical processor show 72% classification accuracy.

## References

- [1] M. Prezioso, F. Merrikh-Bayat, B. D. Hoskins, G. C. Adam, K. K. Likharev, and D. B. Strukov, "Training and operation of an integrated neuromorphic network based on metal-oxide memristors," *Nature*, vol. 521, no. 7550, May 2015, Art. no. 61.
- [2] T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, "Training of photonic neural networks through in situ back-propagation and gradient measurement," *Optica*, vol. 5, no. 7, pp. 864–871, Jul. 2018. [Online]. Available: <http://www.osapublishing.org/optica/abstract.cfm?URI=optica-5-7-864>
- [3] J. M. Shainline, S. M. Buckley, R. P. Mirin, and S. W. Nam, "Superconducting optoelectronic circuits for neuromorphic computing," *Phys. Rev. Appl.*, vol. 7, no. 3, Mar. 2017, Art. no. 034013.

- [4] J. Y. Chen, "GPU technology trends and future requirements," in *Proc. IEEE Int. Electron Devices Meeting*, Dec. 2009, pp. 1–6.
- [5] T. NVIDIA, "V100 GPU Architecture," 2017.
- [6] F. Le Gall, "Powers of tensors and fast matrix multiplication," in *Proc. 39th Int. Symp. Symbolic Algebr. Comput.*, 2014, pp. 296–303.
- [7] R. Stabile, A. Albores-Mejia, A. Rohit, and K. A. Williams, "Integrated optical switch matrices for packet data networks," *Microsyst. Nanoeng.*, vol. 2, Jan. 2016, Art. no. 15042.
- [8] J. Carolan *et al.*, "Universal Linear optics," *Science*, vol. 349, no. 6249, pp. 711–716, Aug. 2015.
- [9] N. C. Harris *et al.*, "Quantum Transport simulations in a programmable nanophotonic processor," *Nature Photon.*, vol. 11, no. 7, pp. 447–452, Jun. 2017.
- [10] D. Perez, E. S. Gomariz, and J. Capmany, "Programmable true-time delay lines using integrated waveguide meshes," *J. Lightw. Technol.*, vol. 36, no. 19, pp. 4591–4601, Oct. 2018.
- [11] R. Burgwal *et al.*, "Using an imperfect photonic network to implement random unitaries," *Opt. Exp.*, vol. 25, no. 23, pp. 28236–28245, Nov. 2017. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-25-23-28236>
- [12] D. A. B. Miller, "Self-aligning universal beam coupler," *Opt. Exp.*, vol. 21, no. 5, pp. 6360–6370, Mar. 2013.
- [13] D. A. B. Miller, "Self-configuring universal linear optical component," *Photon. Res.*, vol. 1, no. 1, pp. 1–15, Jun. 2013.
- [14] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," *Phys. Rev. Lett.*, vol. 73, pp. 58–61, Jul. 1994. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.73.58>.
- [15] Y. Shen *et al.*, "Deep learning with coherent nanophotonic circuits," *Nature Photon.*, vol. 11, no. 7, pp. 441–446, Jul. 2017.
- [16] F. Shokraneh, M. S. Nezami, and O. Liboiron-Ladouceur, "A  $4 \times 4$  reconfigurable optical processor," in *Proc. Asia Commun. Photon. Conf.*, Oct. 2018, pp. 1–3. [Online]. Available: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8596132>
- [17] D. A. B. Miller, "Establishing optimal wave communication channels automatically," *J. Lightw. Technol.*, vol. 31, no. 24, pp. 3987–3994, Dec. 2013.
- [18] D. A. B. Miller, "How complicated must an optical component be?" *J. Opt. Soc. Amer. A*, vol. 30, no. 2, pp. 238–251, Feb. 2013. [Online]. Available: <http://josaa.osa.org/abstract.cfm>.
- [19] M. A. Nielsen, *Neural Networks and Deep Learning*. Determination Press, 2018. [Online]. Available: <http://neuralnetworksanddeeplearning.com/>
- [20] F. N.andola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1 MB model size," 2016. [Online]. Available: <http://arxiv.org/abs/1602.07360>
- [21] A. Zeng, S. Song, J. Lee, A. Rodriguez, and T. A. Funkhouser, "TossingBot: Learning to throw arbitrary objects with residual physics," 2019. [Online]. Available: <http://arxiv.org/abs/1903.11239>
- [22] M. Bojarski *et al.*, "End to end learning for self-driving cars," 2016. [Online]. Available: <http://arxiv.org/abs/1604.07316>
- [23] T. Jo, J. Hou, J. Eickholt, and J. Cheng, "Improving protein fold recognition by deep learning networks," *Scientific Rep.*, vol. 5, Dec. 2015, Art. no. 17573.
- [24] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951. [Online]. Available: <http://www.jstor.org/stable/2236626>
- [25] H. Wu, "Stability analysis for periodic solution of neural networks with discontinuous neuron activations," *Nonlinear Anal., Real World Appl.*, vol. 10, no. 3, pp. 1717–1729, 2009.
- [26] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Netw.*, vol. 4, no. 2, pp. 251–257, 1991.
- [27] I. A. D. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai, and S. Fan, "Reprogrammable electro-optic nonlinear activation functions for optical neural networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–12, 2020.
- [28] A. N. Tait *et al.*, "Silicon photonic modulator neuron," *Phys. Rev. Appl.*, vol. 11, Jun. 2019, Art. no. 064043. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevApplied.11.064043>
- [29] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [30] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
- [31] A. Al-hetar, A. Supa'at, A. Mohammad, and I. Yulianti, "Crosstalk improvement of a thermo-optic polymer waveguide MZIMMI Switch," *Opt. Commun.*, vol. 281, pp. 5764–5767, Dec. 2008.
- [32] Z. Lu, K. Murray, H. Jayatilaka, and L. Chrostowski, "Michelson interferometer thermo-optic switch on SOI with a 50- $\mu$ W power consumption," *IEEE Photon. Technol. Lett.*, vol. 27, no. 22, pp. 2319–2322, Nov. 2015.
- [33] L. Chrostowski *et al.*, "Silicon photonic circuit design using rapid prototyping foundry process design kits," *IEEE J. Sel. Topics Quantum Electron.*, vol. 25, no. 5, Sep./Oct. 2019, Art. no. 8201326.