

Implementing Optical Analog Computing and Electrooptic Hopfield Network by Silicon Photonic Circuits

Guangwei CONG^{†a)}, Noritsugu YAMAMOTO[†], *Nonmembers*, Takashi INOUE[†], Yuriko MAEGAMI[†], *Members*, Morifumi OHNO[†], *Nonmember*, Shota KITA^{††}, *Member*, Rai KOU[†], *Nonmember*, Shu NAMIKI[†], *Member*, and Koji YAMADA[†], *Senior Member*

SUMMARY Wide deployment of artificial intelligence (AI) is inducing exponentially growing energy consumption. Traditional digital platforms are becoming difficult to fulfill such ever-growing demands on energy efficiency as well as computing latency, which necessitates the development of high efficiency analog hardware platforms for AI. Recently, optical and electrooptic hybrid computing is reactivated as a promising analog hardware alternative because it can accelerate the information processing in an energy-efficient way. Integrated photonic circuits offer such an analog hardware solution for implementing photonic AI and machine learning. For this purpose, we proposed a photonic analog of support vector machine and experimentally demonstrated low-latency and low-energy classification computing, which evidences the latency and energy advantages of optical analog computing over traditional digital computing. We also proposed an electrooptic Hopfield network for classifying and recognizing time-series data. This paper will review our work on implementing classification computing and Hopfield network by leveraging silicon photonic circuits.

key words: *photonic neural networks, support vector machine, integrated photonic circuits, silicon photonics, machine learning, optical AI*

1. Introduction

1.1 Optical Neural Networks

Nowadays, super-big artificial intelligence (AI) systems built on various neural network models have been widely deployed for advanced applications such as image generation and natural language understanding [1]. Their scales are continuously increasing at a rate of almost ten times per year, inducing exponential growth in energy consumption as well as computing latency [2]. Such ever-growing demands on high energy efficiency and low latency are difficult to be fulfilled by traditional digital processors with the CMOS technology approaching to physical limits. Under such a circumstance, the brain-like neuromorphic systems showing unparalleled energy efficiency have intrigued a wide interest to develop physical analog hardware for accelerating AI in an energy efficient way [3]–[5].

Optics/photonics provide a promising platform for such a hardware, which can accelerate AI by implementing neural networks in optical systems such as integrated photonics [6], [7], diffractive optics [8], Fourier optics [9], and fiber optics [10]. The main reason that these optical systems can accelerate neural network computing lies

in that the vector-matrix multiplication (VMM) in heavily used fully connected linear layers in neural networks can be completed just by optical propagation at the light speed in media. Optical propagation itself does not consume additional energy and is not influenced by parasitic RC effects. Thus, this optical VMM process in linear layers of neural networks can be accelerated in a very high energy efficiency. Meanwhile, high speed optical modulation and detection have been well established and already deployed for high-capacity optical interconnects in data centers. Bandwidth as high as 10~100 GHz is achievable for all peripheral components such as drivers, modulators, detectors, and transimpedance amplifiers, which enable high speed data input and output (I/O). Similarly, optical/photonic neural networks can leverage these components to achieve high-throughput computing. Therefore, optical implementation of neural networks offers advantages of high speed, low energy consumption, low latency, and high throughput [11].

1.2 Issues and Perspectives

For general AI applications, the neural network models such as multiple-layer perceptron, convolutional neural network (CNN), and recurrent neural network (RNN) must use a large quantity of learnable parameters, nonlinear activation functions, and deep layer to achieve high performance. For an example of CNN for recognizing the Japanese katakana characters, we had been able to achieve > 99% test accuracy with larger than 250 thousand of parameters and cascaded convolutional layers. Obviously, so far, it is unable to realize such a scale for integrated photonic neural networks (PNN).

Besides the small scale, there are also several issues that need to be addressed for applications of PNN.

(1) Lack of efficient all-optical nonlinear activation functions. Currently, the layers of nonlinear neurons usually adopt the optical-electrical-optical (OEO) conversion scheme, where the electrical part can be a digital processor [6], [12] or an analog circuit [13]. PNN with such a hybrid scenario has been proven able to show energy efficiency and latency advantages over pure electronic ones [12].

(2) Latency of post-convolution data processing for optically implementing CNN. The convolution computing for multiple kernels is usually done by leveraging both time and wavelength domains [14]. The convolutional part can be optically accelerated; however, the heavy use of post-

Manuscript received November 9, 2023.

Manuscript publicized January 5, 2024.

[†]The authors are with AIST, Tsukuba-shi, 305-8569 Japan.

^{††}The author is with NTT, Atsugi-shi, 243-0198 Japan.

a) E-mail: gw-cong@aist.go.jp

DOI: 10.1587/transfun.2023GCI0001

convolution processing causes a non-negligible delay and power consumption. Since the subsequent part (usually having a form of MLP) requires the flatten input of all convolutional results, the results being sampled in time domain must be waited to complete for all kernel sizes, saved in memory, flattened, and post-processed before being sent to the next layer, which induces delays and consumes energy and may cancel out those saved in optical convolution. It is still challenging to optically connect the convolutional part and its subsequent layer.

(3) Training compatibility. Training the hardware neural network cannot be performed in the same way as training the software one. The latter one usually adopts backpropagation training algorithm for which all calculation can be done in a tensor-based operation. Recently, analog backpropagation training was demonstrated for PNN by adopting additional efforts including in-situ gradient evaluation using cameras and optical error vector propagation [12]. This analog backpropagation is different from the digital one; thus, training compatibility will be a problem for general AI applications heterogeneously incorporating both hardware and software.

(4) Learning needs large quantities of samples. The learning processing of current neural network models is substantially different from the learning of our brains. The neural network learns in a statistic manner to cover sufficient features from a large quantity of samples. This learning method is not efficient for hardware implementation of neural networks because it increases the power consumption and training time. Training with small-number samples (i.e., few-shot learning) will be an important issue to be considered for training PNN hardware [15].

Due to above issues and challenges, PNN could be a promising solution more suitable for specific AI applications instead of general ones. For specific tasks, we could explore specific machine learning algorithms that could avoid above challenges and meanwhile offer power and latency advantages. Since the electronics is indispensable for the data I/O at the least, PNN must be of an OE-hybrid form. For an ideal architecture, the electronics remains as data I/O only and the photonics completes computing only by optical propagation without being broken by electronic modules. This schematic could offer the best latency and power efficiency.

For other hybrid architectures toward general AI, the boundary between electronics and photonics should be optimized to achieve system advantages. For learning with small-number samples, the traditional Hopfield network is an important recurrent neural network to model the brain-like learning process. It does not need large numbers of training samples and once its network parameters are set according to few samples, it can correctly recognize even partially forgotten or deviated patterns due to the associative memory effect that shows high similarity to brain-like learning. Establishing analog Hopfield network on photonics platforms can greatly save training cost and enable robust recognition by just one-time sampling for temporally coded

signals.

Based on above perspectives, we proposed to implement support vector machine (SVM) like principle [16] and EO-Hopfield network [17] by silicon photonic circuits for classification computing, both of which are specific machine learning models that even have profound applications inside general AI models. This paper will review our works on implementing these two specific machine learning models on silicon photonics platform from principle, architecture, algorithm, to applications.

2. Principle and Architecture

2.1 Projection-Based Classification

Classifying data is a common task of machine learning. In machine learning, besides neural network-based algorithms (MLP, CNN, etc), SVM as an important supervised learning algorithm is also widely used to analyze data for classification and regression [18]. It has high robustness and generalization ability and can be incorporated into neural network as a classifier to form hybrid models [19]. At the beginning, SVM is a binary linear classifier, but now in addition to performing linear classification, SVMs can efficiently perform a non-linear classification using the kernel trick by mapping their inputs into high-dimensional or infinite feature spaces to seek a hyper-plane to separate the data of different labels.

This nonlinear projection principle can be implemented by several mapping ways, either treating the input data using a nonlinear function or performing mutual computing between different elements of the input data. This principle is different from that of the MLP. MLP can be expressed as $y = wf(\dots(wf(wx + b) + b))$, linear combinations of nonlinear activation functions f (sigmoid, ReLU, etc), where w and b are learnable parameters of matrix transformation. In contrast, this principle can be expressed as $y = wG(x) + b$, where G is the nonlinear mapping function, and it is not necessary to keep the same dimension as the original data. Figure 1(a) shows a schematic to illustrate this principle for a one-dimensional data x of two labels with a periodic feature. If adopting the MLP model to classify the data, the classification is like to train the equation of $y = wf(\dots(wf(wx + b) + b))$ to approximate a sinusoidal function, which is obviously not a simple way. However, if we directly map the data by sinusoidal function to a two-dimensional space, the new generated data x' can be easily separated by a line. The separating line is not unique, but the optimized line is given by the maximum margin, i.e., the maximum distances of the worst data points to the line. Similarly, this principle can be extended to higher dimensional data classification.

Compared to the MLP, this principle has advantages in photonic implementation because it does not need activation functions to interleave fully connected layers and deep-layer structures. Direct nonlinear mapping and inter-element multiplication on the original input data can leverage the EO nonlinearity of various kinds of photonics devices such as

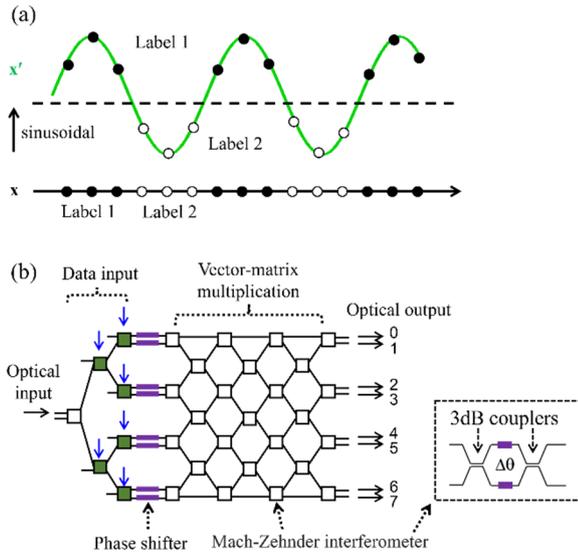


Fig. 1 (a) Schematic of nonlinear projection enabled classification. (b) A photonic topology to implement the nonlinear projection-based classification principle utilizing MZI's EO nonlinearity. Inset: a MZI unit (revised from [16]).

Mach-Zehnder interferometers (MZI), ring resonators, and p - i - n waveguide attenuators. Thus, the whole photonic circuit can be realized by only passive waveguides, without necessities to integrate optical nonlinearity such as optical amplifiers and all-optical nonlinear devices, offering an easy way for implementing machine learning tasks in photonic circuits.

To demonstrate the photonic implementation of this principle, we proposed a photonic classifier network (PCN) in Fig. 1(b) [16]. The phase shifter is a device which can change the optical phase for the light passing through when a voltage or current is applied to it. A MZI is an interferometer device consisting of two 3-dB couplers and two phase shifters. When a phase difference is set between two phase shifters, the optical output amplitude of MZI can be arbitrarily adjusted. The data to be analyzed is originally an electrical signal and input as voltage or current to one phase shifter of a MZI, which is indicated as the “data input”. Then, we utilize the EO nonlinearity of MZI to map the input data into the 8-dimensional optical complex amplitude space. The mapping functions can be constructed in various ways by inputting the data into the MZI networks. For example, we input two bits of XOR into two MZIs of the left column of MZIs indicated by arrows in Fig. 1(b) and input four parameters of Iris dataset into the four MZIs of the right column of MZIs. A subsequent VMM consisting of MZI meshes of Clement's topology completes the linear separation by constructing the plane in 8-dimensional complex space. The results (i.e., labels) are marked out by the optical power distributions at the output ports. Therefore, the classification computing is completed only by optical propagation inside the photonic circuit without being interrupted by intermediate OEO conversion, offering low latency and low power consumption. This PCN can offer high efficiency classifica-

tion and equivalent performance for several machine learning tasks even with fewer learnable parameters compared to the conventional MLP models [16].

2.2 Principle Explanation by XOR

We exemplify the abovementioned principle by XOR which is a linearly inseparable problem and usually used to verify the nonlinear classification capability for neural network models [20]. XOR has four data patterns and two labels as expressed by $\{x = [[0, 0], [0, 1], [1, 0], [1, 1]], \text{label} = [0, 1, 1, 0]\}$, a two-parameter and two-target problem. Obviously, these four patterns cannot be linearly divided into two groups in accordance with their labels. For explaining the nonlinear projection principle described above, we simply use two MZIs (2×2 type, two inputs with only one port having light input) to accept the input of its two bits (in a unit of π) as optical phase, respectively. As a result, this maps the two bits into a complex space and after mapping, the data x becomes $x' = [[0, i, 0, i], [0, i, -1, 0], [-1, 0, 0, i], [-1, 0, -1, 0]]$, according to the equation of MZI [16]. For the projected data x' , we can easily find a linear transformation in complex space having the coefficients w and b , as shown in Eqs. (1) and (2) respectively, to separate its labels according to $y = |wx^T + b|^2$ using optical power detection. The label here is using a one-hot vector to present 0 and 1; thus, y has two complementary columns as shown in Eq. (3). This XOR example explains the nonlinear projection enabled classification.

$$w = \begin{bmatrix} -0.678 - 0.215i & -0.381 + 0.116i & -0.579 - 0.526i & -0.095 + 0.211i \\ -0.630 - 0.189i & -0.570 + 0.046i & 0.463 - 0.228i & 0.990 - 0.159i \end{bmatrix} \quad (1)$$

$$b = [-0.476 - 0.126i \quad -0.142 - 0.410i] \quad (2)$$

$$y^T = \begin{bmatrix} 1.0 & 5.4e-4 & 1.2e-4 & 1.0 \\ 1e-3 & 1.0 & 1.0 & 6.8e-4 \end{bmatrix} \quad (3)$$

2.3 EO-Hopfield Network

Hopfield network is a single layer recurrent network and accepts one-shot trigger input into the neurons. It can be a discrete or continuous type, depending on using a binary neuron function or a continuous one. Hopfield network has one important unique feature, associative memory, which is a brain-like learning behavior. The pattern is remembered by the network weights, and it can be recalled by recursively updating the neurons even when inputting a partial or damaged pattern. Thus, it does not need a large quantity of training data.

We proposed to input time-series data into Hopfield network and leverage the associative memory effect for feature extraction and recognition for temporal analog signals. For this purpose, we extend the photonic topology in Fig. 1(b) to EO-Hopfield network by adding electrical feedback to the MZIs that are used as the neurons [17]. As shown in Fig. 2, the optical outputs are converted to electrical signals by photodetectors (PD) which are feedbacked

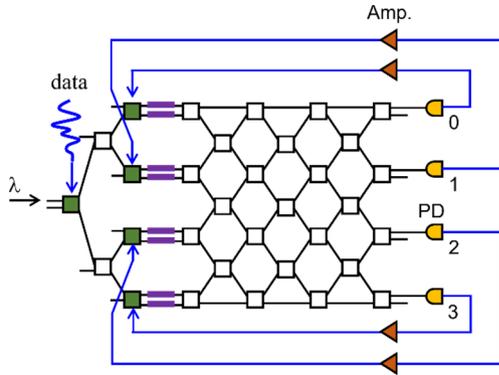


Fig. 2 Architecture of the EO-Hopfield network utilizing electrical feedback to the MZI (revised from [17]).

to four MZIs via amplifiers (Amp), forming OE loops. This is continuous Hopfield-like recurrent network by leveraging the MZI’s OE nonlinearity. The analog data is input to an MZI, and the output of each MZI is connected to other three MZIs and itself, depending on the weight state given by the MZI mesh. The architecture can be trained to remember four waveforms by using four corresponding spatial feature vectors of the optical outputs which are sampled just one time at the end of signal input. For some simple tasks, this feature vector can directly give out the results of classification, while for complex tasks, this vector can be used as an input feature for subsequent processing by linear transformation as did in photonic reservoir computing [21], [22]. Once being trained, even though the input waveform deviates from its learned one, it can be correctly recognized.

3. Experiment and Discussion

3.1 Photonic Device and Measurement

We fabricated a silicon photonic device for the topology in Fig. 1(b) based on the AIST SCR 12-inch silicon photonic platform [23]. Figure 3(a) shows the fabricated device on a 220-nm silicon-on-insulator wafer with a 3- μm buried oxide layer. The chip is of 5 mm in length and 1.3 mm in width and consists of fully etched silicon wire waveguide of 430 nm in width. All phase shifters and MZIs are thermo-optic ones using TiN micro-heaters. All these fundamental devices are the standardized ones in our silicon photonic PDK (process design kits). For all MZIs, only the top arm was fabricated with a heater. The average π -shift power for these thermo-optical phase shifters is about 17 mW at 1.53 μm . The blue and red arrows indicate the MZIs for inputting the data of 2-bit Boolean logic and 4-parameter Iris dataset [24], respectively. All ground pads are indicated by the letter “G”. The red square shows the MZI unit device. This chip works only for TE polarization and a *p-i-n* phase shifter of rib waveguide was inserted in each output waveguide for easy polarization adjustment, which is not shown in Fig. 3(a).

The chip in Fig. 3(a) was packaged into a module with coupled fibers and electrical connectors via wiring bonding.

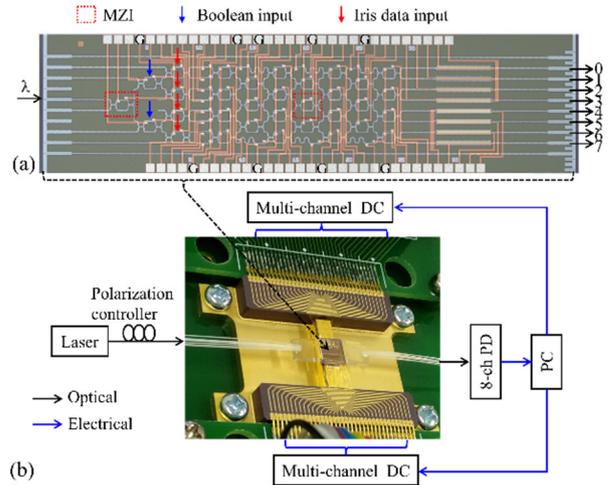


Fig. 3 (a) Fabricated silicon photonic chip. (b) Packaged module and experimental setup (revised from [16]).

After packaging, the fiber-to-fiber loss is about 4.5 dB for the referenced straight waveguide. For measurement, a laser at 1.53 μm was input after being tuned to TE polarization. An 8-channel optical power meter was used to measure the optical powers at eight output ports. The computer read the optical powers and ran the training algorithms to update the voltages for all on-chip phase shifters via two multi-channel direct-current sources. Thus, after training, the learned state is indicated by the voltage distribution among all heaters. For verification, we just keep the learning voltage distribution and input the new data into the MZIs.

3.2 Training Method

On-chip training is a challenge for PNN hardware. Neural networks built on computer can be trained by back-propagation (BP) algorithm which is very highly efficient because of tensor-based gradient evaluation for both fully connected linear layer and nonlinear activation layer. For PNN, the weight information is in electrical domain, while the gradient information is in optical domain. Due to the lack of efficient protocol for evaluating the gradient in optical domain and updating it from optical to electrical domain, the same back-propagation algorithm as used in software cannot be applied to the PNN if without taking additional measures. So far, there are four ways for setting weight for PNN.

(1) Training the computer model of PNN. A circuit analogy is first built on computer and then it can be trained by the same BP algorithm as used in software. After training, the learned weight parameters will be deployed to the PNN for optical implementation of the learned model. This method means that the PNN only implements the inference for a computer model of PNN, leaving the training to digital processors. An example of this method is the diffractive optical neural network [8].

(2) Forward differential method. For this method, the gradient for each parameter is evaluated by two-time for-

ward propagation (FP). For all parameters, one propagation is performed with the current value and the other propagation is done by adding a small variation to the current value. The cost function is evaluated using the optical outputs of each propagation and then the gradient is calculated using two costs. After the gradients are evaluated for all parameters, an optimizer (Adam, RMSprop, etc) can be used to update the parameters, similar as that used in back-propagation algorithm. This method is mentioned in [6] for on-chip training, while it was not experimentally demonstrated yet. In this study, we experimentally demonstrated this FP algorithm for on-chip training. We used the mean squared error (MSE) as the cost function.

(3) Global optimization algorithm. This algorithm treats the chip as a black box and determines the parameters by global optimization process. We established such an algorithm named bacterial foraging optimization (BFO) based on stochastic process [16], [25]–[28] for on-chip training and reconfiguring photonic chip. Genetic algorithm was also demonstrated for on-chip training a photonic chip [29]. If using only the optical input and output ports of the chip, abovementioned FP and global optimization algorithms are the only possible ways for on-chip training although they are not as highly efficient as the BP. We developed a Python-based tool based on Pytorch [30] for both on-chip training experiment and PNN simulation with both FP and BFO algorithms incorporated.

(4) Analog optical back-propagation training. Recently, an in-situ optical BP training was demonstrated by measuring the gradient through infrared camera and preparing optical error back propagation [12]. On-chip power monitoring, local feedback circuits from optical to electrical domains, and optical error preparation are required for really deploying this method.

3.3 Boolean Logic Classification

We demonstrated Boolean logic classification based on both FP and BFO algorithms. Here we show the results of BFO training. The logical values (0 and 1) of the AND and OR can be separated linearly, so they are regarded as the linearly separable problems. In contrast, the logical values of XOR cannot be separated linearly; thus, XOR is regarded as a linearly inseparable problem that is usually used to test the nonlinear classification capability for a neural network model. Based on the principle explained above in the Sect. 2, we normalize the two bits (0 and 1) of Boolean logics as 0 and π , and input them into the MZIs as indicated in Fig. 3(a). Here we demonstrate classification, simultaneously for both XOR and AND. For XOR, we assign the ports 1 and 3 to present the logic values 0 and 1, respectively. At the same time, for AND, we assign the ports 5 and 7 to present the logic values of 0 and 1, respectively.

Before training, the optical output powers (not shown) are random for all bit patterns ([00, 01, 10, 11]). After training, as shown in Fig. 4, the power map features the XOR-like pattern at the ports 1 and 3, while the AND-like pattern

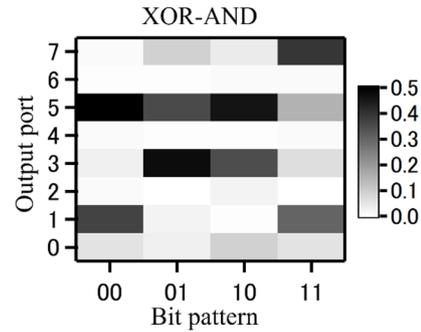


Fig. 4 Optical output power distribution after BFO training, which shows the XOR and AND classification result (cited from [16]).

at the ports 5 and 7. In other words, when we input a bit pattern, the maximum optical powers occur at the ports according to its assigned logic values. For example, when the bit pattern is 01, the port 3 has a higher power than the port 1, and the port 5 higher than the port 7, indicating the logic value of 1 for XOR and 0 for AND. Therefore, the high or low powers at the ports 3 and 7 correspond to the logic values, while the high or low powers at the ports 1 and 5 indicate the bar states of their logic values, for XOR and AND, respectively. For actual applications, balanced detection can be adopted between the two ports assigned to stand for different logic values, from which the binary output can be achieved. In essence, this is optical analog computing enabled by optical interference, although the results can be interpreted as digital logics through comparators when reading the optical results into electrical domain.

3.4 Iris Dataset Classification

We demonstrated another machine learning benchmark, the Iris dataset classification [16]. This task is to judge the Iris flower species from the statistic data of four flower sizes [7], [24]. This dataset has three kinds of Iris flowers (Setosa, Versicolor, and Virginica) and 150 samples in total (50 samples for each species). Thus, it is a four-parameter and three-label classification task. Similar as the XOR classification as shown in the Sect. 3.3, we need assign three ports to stand for each species for the Iris classification. At the port that stands for a species, the optical output power is to be maximized by the training when the parameters of a flower belonging to this species are input.

All parameters in the dataset were first normalized to 2π (min-max normalized) and then input into the four MZIs as indicated in Fig. 3(a). Before training, the optical power distribution is random-like pattern (not shown), showing no information related to the Iris classification. After training, as shown in Fig. 5(a), the maximum-power ports are clearly separated into three groups for 90 training samples. When the input parameters belong to Setosa, Versicolor, or Virginica, the maximum power occurs at the port 1, 3, or 5, respectively, as we assigned. If the maximum power port is not the assigned one, the classification is accounted as the wrong recognition. The training accuracy is about 94.44%.

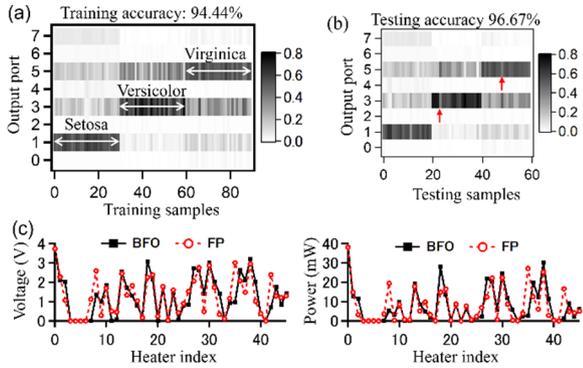


Fig. 5 Optical output power distribution (a) after BFO training for 90 Iris samples and (b) for model test for 60 samples. (c) Learned voltages and powers for all phase shifters by FP and BFO algorithms (cited from [16]).

After training, we kept the learned voltages at all phase shifters (i.e., the learned on-chip phase distribution is remained as it is) and input another 60 samples that were not included in the train samples to verify the classification. As shown in Fig. 5(b), the test accuracy is about 96.67%, verifying the training effectiveness and validation of our photonic chip in classification computing. The learned voltages and corresponding powers are shown in Fig. 5(c) for all phase shifters (heaters). The BFO and FP algorithms show a similar overall profile of voltage distribution, presenting a mutual confirmation. These electrical powers are required to maintain the optical phases to implement Iris classification. The total on-chip electrical power is about 360 mW. In addition, for our photonic chip, the computing latency is the optical propagation time from the data input to output, which is estimated to be less than 100 ps. This demonstration shows the potential of our projection based PNN in low latency and low power classification without using traditional nonlinear activation functions.

3.5 Bit Pattern Recognition by EO-Hopfield Network

Here we use the EO-Hopfield network architecture in Fig. 2 to recognize the bit sequences by simulation to understand its function [17]. We consider four 4-bit patterns, all-one (1111), all-zero (0000), half-zero/half-one (0011), and half-one/half-zero (1100) for waveform recognition. Each pattern will be presented by an output power distribution (feature vector) at four output ports. For simpleness, the maximum-power position will be used present each pattern. We input the bit in a unit of π and set the gain coefficient (convert optical powers to optical phase by a pre-amplified circuits supplied with a normalized voltage) to 1.5π for the amplifier. The loop delay can be a tunable parameter and one bit delay is assumed here (other delay lengths are also applicable, which are not discussed in this work).

Obviously, if without OE feedback, we cannot distinguish all-one and half-zero/half-one, and all-zero and half-one/half-zero by only sampling the last output because their last bits are same. We train the EO-Hopfield architecture to

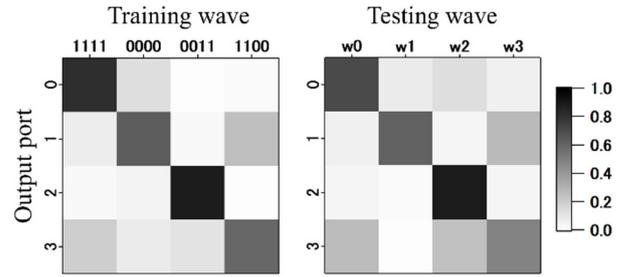


Fig. 6 Normalized optical power sampled at the last bit for training and testing waves ($w_0 = [0.9, 0.9, 0.9, 0.9]$, $w_1 = [0.1, 0.1, 0.1, 0.1]$, $w_2 = [0.1, 0.1, 0.9, 0.9]$, $w_3 = [0.9, 0.9, 0.1, 0.1]$).

distinguish above four patterns by just sampling the optical output at the last bit. The architecture can be described as a recursive equation $y_t = w \cdot (s \otimes f(x_t)) \cdot f(gy_{t-1})$, where x_t and y_t are the input data and output optical power at a time, g is the amplifier's gain parameter, f is the EO nonlinearity of MZI neuron, s is an eight-dimensional optical complex vector to be input into the MZI neurons, and w is a matrix containing the network parameters which describe the neuron connections and save the memorized patterns. w and s are training parameters. The final spatial vector y depends on the interaction between these two parameters and the history of input data. In other words, this system converts the time-series data to a unique spatial vector.

As shown in Fig. 6, after training, each pattern can be characterized by a different port of maximum power. The maximum power occurs at the respective ports of 0, 1, 2, 3 at the last bit for different patterns, indicating the pattern separation. After training the network using these four ideal bit patterns, we input the damaged bit patterns deviated by 10% for all bits (as denoted by w_0 , w_1 , w_2 , and w_3 in Fig. 6) and these damaged analog patterns can also be correctly distinguished. Thus, training the EO-Hopfield network does not need a large quantity of training samples. We perform the training using only four samples, but the trained system shows wider applicability to much more samples deviated from the training ones, offering robust recognition [17]. It is still an open question to figure out the range of deviation without influencing recognition accuracy. Inputting data into recurrent Hopfield network has not been proposed previously, as far as we know, especially for hardware implementation. Thus, our proposal offers a novel application for Hopfield network.

4. Conclusions

We reviewed our work on implementing machine learning tasks and EO-Hopfield network by leveraging silicon photonic circuits. We experimentally demonstrated the machine learning benchmarks by adopting nonlinear projection-based principle which can be realized only by passive silicon waveguides and can avoid the difficulty of integrating optical nonlinear device. Computing was completed only by on-chip optical propagation. We verified in principle that our proposed EO-Hopfield network was an

applicable analog hardware for classifying and recognizing time-series data. Our work evidences the potential of silicon photonic circuits for low latency and low power classification computing.

Acknowledgments

This work was partly supported by Japan Science and Technology Agency (JST), CREST Grant Number JP-MJCR15N4 and JPMJCR21C3, and JSPS KAKENHI Grant Number JP23H01885. The authors thank all staffs at the AIST-SCR station for device fabrication.

References

- [1] Y. Liu, T. Han, S. Ma, J. Zhang, Y. Yang, J. Tian, H. He, A. Li, M. He, Z. Liu, Z. Wu, L. Zhao, D. Zhu, X. Li, N. Qiang, D. Shen, T. Liu, and B. Ge, “Summary of ChatGPT-related research and perspective towards the future of large language models,” arXiv:2304.01852, 2023.
- [2] I. Shumailov, Y. Zhao, D. Bates, N. Papernot, R. Mullins, and R. Anderson, “Sponge examples: Energy-latency attacks on neural networks,” arXiv:2006.03463v2, 2021.
- [3] T.P. Xiao, C.H. Bennett, B. Feinberg, S. Agarwal, and M.J. Marinella, “Analog architectures for neural network acceleration based on non-volatile memory,” *Appl. Phys. Rev.*, vol.7, no.3, p.031301, 2020.
- [4] K. Kitayama, M. Notomi, M. Naruse, K. Inoue, S. Kawakami, and A. Uchida, “Novel frontier of photonics for data processing — Photonic accelerator,” *APL Photon.*, vol.4, no.9, p.090901, 2019.
- [5] L. De Marinis, M. Cococcioni, P. Castoldi, and N. Andriolli, “Photonic neural networks: A survey,” *IEEE Access*, vol.7, pp.175827–175841, 2019.
- [6] Y. Shen, N.C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, “Deep learning with coherent nanophotonic circuits,” *Nat. Photon.*, vol.11, pp.441–446, 2017.
- [7] H. Zhang, M. Gu, X.D. Jiang, J. Thompson, H. Cai, S. Paesani, R. Santagati, A. Laing, Y. Zhang, M.H. Yung, Y.Z. Shi, F.K. Muhammad, G.Q. Lo, X.S. Luo, B. Dong, D.L. Kwong, L.C. Kwek, and A.Q. Liu, “An optical neural chip for implementing complex-valued neural network,” *Nat. Commun.*, vol.12, p.457, 2021.
- [8] X. Lin, Y. Rivenson, N.T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” *Science*, vol.361, no.6406, pp.1004–1008, 2018.
- [9] M. Miscuglio, Z. Hu, S. Li, J.K. George, R. Capanna, H. Dalir, P.M. Bardet, P. Gupta, and V.J. Sorger, “Massively parallel amplitude-only Fourier neural network,” *Optica*, vol.7, no.12, pp.1812–1819, 2020.
- [10] J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M.L. Gallo, X. Fu, A. Lukashchuk, A.S. Raja, J. Liu, C.D. Wright, A. Sebastian, T.J. Kippenberg, W.H.P. Pernice, and H. Bhaskaran, “Parallel convolutional processing using an integrated photonic tensor core,” *Nature*, vol.589, pp.52–58, 2021.
- [11] X. Xiao, M.B. On, T. Van Vaerenbergh, D. Liang, R.G. Beausoleil, and S.J.B. Yoo, “Large-scale and energy-efficient tensorized optical neural networks on III–V-on-silicon MOSCAP platform,” *APL Photon.*, vol.6, no.12, p.126107, 2021.
- [12] S. Pai, Z. Sun, T.W. Hughes, T. Park, B. Bartlett, I.A.D. Williamson, M. Minkov, M. Milanizadeh, N. Abebe, F. Morichetti, A. Melloni, S. Fan, O. Solgaard, and D.A.B. Miller, “Experimentally realized in situ backpropagation for deep learning in photonic neural networks,” *Science*, vol.380, no.6643, pp.398–404, 2023.
- [13] S. Bandyopadhyay, A. Sludds, S. Krastanov, R. Hamerly, N. Harris, D. Bunandar, M. Streshinsky, M. Hochberg, and D. Englund, “Single chip photonic deep neural network with accelerated training,” arXiv:2208.01623, 2022.
- [14] X.Y. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T.G. Nguyen, S.T. Chu, B.E. Little, D.G. Hicks, R. Morandotti, A. Mitchell, and D.J. Moss, “11 TOPS photonic convolutional accelerator for optical neural networks,” *Nature*, vol.589, pp.44–51, 2021.
- [15] A. Parnami and M. Lee, “Learning from few examples: A summary of approaches to few-shot learning,” arXiv:2203.04291, 2022.
- [16] G. Cong, N. Yamamoto, T. Inoue, Y. Maegami, M. Ohno, S. Kita, S. Namiki, and K. Yamada, “On-chip bacterial foraging training in silicon photonic circuits for projection-enabled nonlinear classification,” *Nat. Commun.*, vol.13, p.3261, 2022.
- [17] G. Cong, N. Yamamoto, R. Kou, Y. Maegami, M. Ohno, and K. Yamada, “Silicon photonic Hopfield-like electro-optical recurrent network for time-series data processing and recognition,” *Proc. OFC2023, W3G.2*, 2023.
- [18] B. Schölkopf and A.J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, Ch. 1&2, MIT Press, London, 2002.
- [19] D. Keerthana, V. Venugopal, M.K. Nath, and M. Mishra, “Hybrid convolutional neural networks with SVM classifier for classification of skin cancer,” *Biomedical Engineering Advances*, vol.5, no.100069, 2023.
- [20] T.W. Hughes, M. Minkov, Y. Shi, and S. Fan, “Training of photonic neural networks through in situ backpropagation and gradient measurement,” *Optica*, vol.5, no.7, pp.864–871, 2018.
- [21] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, “Experimental demonstration of reservoir computing on a silicon photonics chip,” *Nat. Commun.*, vol.5, p.3541, 2014.
- [22] H. Hasegawa, K. Kanno, and A. Uchida, “Parallel and deep reservoir computing using semiconductor lasers with optical feedback,” *Nanophotonics*, vol.12, no.5, pp.869–881, 2022.
- [23] K. Yamada, T. Horikawa, M. Okano, G. Cong, Y. Maegami, M. Ohno, N. Yamamoto, K. Suzuki, K. Tanizawa, S. Suda, H. Matsuura, K. Koshino, N. Yokoyama, M. Ohtsuka, M. Seki, K. Matsumaro, T. Narushima, K. Ikeda, H. Kawashima, S. Namiki, and M. Mori, “A 300-mm-wafer silicon photonics technology for ultra-low-energy optical network systems,” *Proc. ACP2017, S4H.3*, 2017.
- [24] D. Dua and C. Graff, *UCI Machine Learning Repository* [<http://archive.ics.uci.edu/ml/>] (University of California, School of Informatics and Computer Science, Irvine, CA, 2019).
- [25] G. Cong, N. Yamamoto, T. Inoue, M. Okano, Y. Maegami, M. Ohno, and K. Yamada, “Arbitrary reconfiguration of universal silicon photonic circuits by bacterial foraging algorithm to achieve reconfigurable photonic digital-to-analog conversion,” *Opt. Express*, vol.27, no.18, p.24914, 2019.
- [26] G. Cong, N. Yamamoto, T. Inoue, M. Okano, Y. Maegami, M. Ohno, and K. Yamada, “High-efficient black-box calibration of large-scale silicon photonics switches by bacterial foraging algorithm,” *OFC2019, M3B.3*.
- [27] G. Cong, N. Yamamoto, T. Inoue, Y. Maegami, M. Ohno, S. Kita, S. Namiki, and K. Yamada, “Experimental demonstration of XOR separation by on-chip training a linear silicon photonic circuit,” *OFC2021, Th4I.3*, 2021.
- [28] G. Cong, N. Yamamoto, Y. Maegami, M. Ohno, and K. Yamada, “Experimental demonstration of automatic reconfiguration and failure recovery of silicon photonics circuits,” *ECOC2021, We4D.3*, 2021.
- [29] H. Zhang, J. Thompson, M. Gu3, X.D. Jiang, H. Cai, P.Y. Liu, Y. Shi, Y. Zhang, M.F. Karim, G.Q. Lo, X. Luo, B. Dong, L.C. Kwek, and A.Q. Liu, “Efficient on-chip training of optical neural networks using genetic algorithm,” *ACS Photon.*, vol.8, no.6, pp.1662–1672, 2021.
- [30] PyTorch at <https://pytorch.org/>



Guangwei Cong received the Ph.D. degree from the Institute of Semiconductors, Chinese Academy of Sciences, in 2006. In the same year, he joined the National Institute of Advanced Industrial Science and Technology (AIST), Japan, as a Postdoc and JSPS researcher, where he did research on ultrafast all-optical switches using intersubband transitions in quantum wells. Since 2010, he has been a permanent researcher and senior researcher with AIST. Currently he is a Chief Senior Researcher of AIST and involved in developing silicon photonics platform technology for communication and computing applications. He is a member of the IEEE Photonics Society and Optica.

involved in developing silicon photonics platform technology for communication and computing applications. He is a member of the IEEE Photonics Society and Optica.



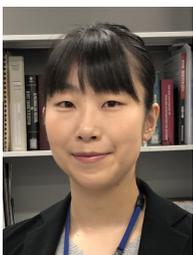
Noritsugu Yamamoto received the B.E., M.E., and Ph.D. degrees in Electronics from Kyoto University in 1994, 1996 and 1999, respectively. In 1999, he joined Electrotechnical Laboratory, which was reorganized into the National Institute of Advanced Industrial Science and Technology (AIST) in 2001. Currently, he is a Research Manager of the AIST Platform Photonics Research Center. His current research field is silicon photonics. He is member of the Japan Society of Applied Physics (JSAP) and

IEEE Photonics Society.



Takashi Inoue received his Ph.D. degree in communications engineering from Osaka University, Osaka, Japan, in 2002. He joined Furukawa Electric Co. Ltd., Ichihara, Japan, in 2002, where he developed optical signal processing devices based on nonlinear fiber optics, and silica-based planar lightwave circuits. In 2011, he joined National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan, where he is currently working on digital coherent transmission systems, signal processing techniques, and optical networks. He is a member of the IEEE Photonics Society and the IEICE Communication Society.

terms, signal processing techniques, and optical networks. He is a member of the IEEE Photonics Society and the IEICE Communication Society.



Yuriko Maegami received her M.E. and Ph.D. degrees of the Department of Mechanical Systems Engineering, Tohoku University, Sendai, Japan, in 2009 and 2013, respectively. She joined the National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan, in 2013. Currently, she is studying silicon photonics devices using CMOS-compatible materials such as silicon nitride. She is a member of IEEE Photonics Society, the Institute of Electronics, Information and Communication Engineers (IEICE), and the Japan Society of Applied Physics (JSAP).

neers (IEICE), and the Japan Society of Applied Physics (JSAP).



Morifumi Ohno received the M.S. degree in electrical engineering from the Tokyo University of Agriculture and Technology, Tokyo, Japan, in 1983, and the Ph.D. degree in electronic materials science from the Graduate School of Electronic Science and Technology, Shizuoka University, Shizuoka, Japan, in 1986. He had joined Oki Electric Industry Company, Ltd. (present, OKI Semiconductor Company, Ltd.), where he engaged in the R&D of advanced NVRAM (FeRAM) and FD-SOI devices. In 2009, he joined, as an Invited Researcher, with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan, where he was involved in the design, simulation, and characterization of nano-electronic devices and silicon photonics devices, including modulators and other novel devices. Currently, he is an Invited Researcher, of the AIST Platform Photonics Research Center. He is a member of the IEEE Photonics Society.

ices. In 2009, he joined, as an Invited Researcher, with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan, where he was involved in the design, simulation, and characterization of nano-electronic devices and silicon photonics devices, including modulators and other novel devices. Currently, he is an Invited Researcher, of the AIST Platform Photonics Research Center. He is a member of the IEEE Photonics Society.



Shota Kita is a Senior Researcher, Photonic Nanostructure Research Group of NTT Basic Research Laboratories and NTT Nanophotonics Center. He received a Ph.D. in engineering from Yokohama National University in 2012. He was a postdoc researcher in John A. Polson Harvard University, USA, for 3 years. His interests are in silicon photonic-based photoelectronic converged processing technologies. He received the Poster Presentation Award at the International Nano-Optoelectronics Workshop (iNOW) 2009 and Young Scientist Presentation Award from the Japan Society of Applied Physics (JSAP) in 2010. He is a member of JSAP, IEICE, and Optical Society (OSA).

and Young Scientist Presentation Award from the Japan Society of Applied Physics (JSAP) in 2010. He is a member of JSAP, IEICE, and Optical Society (OSA).



Rai Kou received the B.E., M.E. and Ph.D. degrees in Applied Physics from Waseda University, Tokyo, Japan, in 2007, 2009, and 2014, respectively. From 2009 to 2015, he was a researcher at NTT Microsystem Integration Laboratories, and Nanophotonics Center. He was a postdoctoral fellow at the Department of Physics, University of California, Berkeley in 2015, and joined the AIST in 2017. His current research interests include heterogenous integration on silicon photonics circuits, and advanced nonlinear materials. Dr. Kou is a member of the Japan Society of Applied Physics (JSAP), the Laser Society of Japan (LSJ), and OPTICA.

nonlinear materials. Dr. Kou is a member of the Japan Society of Applied Physics (JSAP), the Laser Society of Japan (LSJ), and OPTICA.



Shu Namiki received M.S. and Dr. Sci. degrees in applied physics from Waseda University, Tokyo, Japan, in 1988 and 1998, respectively. From 1988 to 2005, he was working for Furukawa Electric Co., Ltd., where he developed award-winning high-power pump lasers, and patented multi-wavelength-pumped fiber Raman amplifiers. From 1994 to 1997, he was a Visiting Scientist at the Massachusetts Institute of Technology, Cambridge, where he studied mode-locked fiber lasers and ultra-short

pulses in optical fiber. In 2005, he moved to the National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan, where he has served as Executive Committee Chair of a ten-year national project called “Vertically Integrated Center for Technologies of Optical Routing toward Ideal Energy Savings (VICTORIES)” in collaboration with ten telecom-related companies, and is currently Research Director of Platform Photonics Research Center in AIST. He has authored or coauthored more than 680 conference presentations, papers, book chapters, and patents. His current research interests include software defined dynamic optical path networking and their enabling devices such as nonlinear fiber-optics and silicon photonics. He was an Associate Editor and Advisory Editor of Optics Express and the Co-Editor-in-Chief of the IEICE Transactions on Communications. He was on the technical committee for OFC, ECOC, CLEO, OECC, OAA, etc. and served as Program Co-Chair of OFC 2015, and General Co-Chair of OFC 2017. Dr. Namiki is a Fellow of Optica and IEEE, and a member of IEICE and JSAP. He is a recipient of IEICE Achievement Award in 2020.



Koji Yamada received his B.E., M.E. and Ph.D. degrees in nuclear engineering from Kyushu University, Japan, in 1986, 1988 and 2003, respectively. Currently, he is a Principal Research Manager in National Institute of Advanced Industrial Science and Technology (AIST), Japan, and is also a leader of Silicon Photonics Team in the institute. From 1988 to 2015, in NTT laboratories, he was engaged in studies on particle accelerator physics/engineering for synchrotron light

sources and studies on silicon-based photonic platform. Since joining AIST in 2015, he continues studying silicon-based photonic platform, and has established the first public silicon photonics foundry service in Japan. He is a Fellow of IEEE, a senior member of IEICE, and a member of the Japan Society of Applied Physics. He is a recipient of IEICE Achievement Award in 2023.