Research Article | Vol. 28, No. 8 / 13 April 2020 / Optics Express 12138
Optics EXPRESS

Experimental realization of arbitrary activation functions for optical neural networks

MONIREH MOAYEDI POUR FARD,[1] IAN A. D. WILLIAMSON,[2] MATTHEW EDWARDS,[1] KE LIU,[1] SUNIL PAI,[2] BEN BARTLETT,[3] MOMCHIL MINKOV,[2] TYLER W. HUGHES,[3] SHANHUI FAN,[2] AND THIEN-AN NGUYEN[1,*]

[1]GenXComm Incorporated, 10000 Metric Blvd. Ste 200, Austin, TX 78758, USA
[2]Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA
[3]Department of Applied Physics, Stanford University, Stanford, California 94305, USA
[*]tnguyen@genxcomm.com

Abstract: We experimentally demonstrate an on-chip electro-optic circuit for realizing arbitrary nonlinear activation functions for optical neural networks (ONNs). The circuit operates by converting a small portion of the input optical signal into an electrical signal and modulating the intensity of the remaining optical signal. Electrical signal processing allows the activation function circuit to realize any optical-to-optical nonlinearity that does not require amplification. Such line shapes are not constrained to those of conventional optical nonlinearities. Through numerical simulations, we demonstrate that the activation function improves the performance of an ONN on the MNIST image classification task. Moreover, the activation circuit allows for the realization of nonlinearities with far lower optical signal attenuation, paving the way for much deeper ONNs.

1. Introduction

Machine learning and artificial neural networks are playing an increasingly important role in a number of important application areas, ranging from health care to self-driving cars. Optics provides a unique platform for high-speed, low latency, and energy-efficient matrix-vector multiplications, the computational bottleneck in artificial neural networks. In the late 1980s, optical neural networks (ONNs) were originally proposed based on free-space optical setups consisting of lenses and holographic components [1,2].

Over the last several decades, the development of integrated photonic technologies has opened up new research directions into large-scale optical information processing in programmable microwave photonics and ONNs [3–7]. Particularly for ONNs, integrated interferometer meshes implemented in a silicon photonics platform have been proposed for accelerating matrix-vector multiplications [4,8], with a constant computational time scaling over the dimension of the matrix. Such scaling is a significant advantage over digital processors, where the computational time cost of matrix-vector multiplications scale quadratically with the matrix dimension. Moreover, the large modulation rates and signal bandwidths available in photonic platforms allows for the realization of neural networks with much larger effective clock rates than in electronic circuits, which are limited by thermal effects.

Nonlinear activation functions are an indispensable component of artificial neural networks, allowing them to learn complex nonlinear relationships between their inputs and outputs. However, implementing nonlinear activation functions in large-scale integrated ONNs remains a significant challenge. One of the main limitations in realizing an on-chip optical nonlinear function comes from the relatively weak nonlinearities available in photonic platforms. As a result, implementing strong nonlinear activation functions would require very long waveguide interaction lengths and

#391473 https://doi.org/10.1364/OE.391473
Journal © 2020 Received 25 Feb 2020; revised 31 Mar 2020; accepted 31 Mar 2020; published 8 Apr 2020

Optics EXPRESS

high optical signal powers that undesirably increase the footprint and power consumption of the ONN. While the optical nonlinearity can be enhanced by resonant structures [9], this comes at the cost of an unavoidable trade-off in bandwidth and the added requirement of calibration circuitry for the optical resonators [10–12]. Another challenge with using optical nonlinearities in ONNs is that their nonlinear responses are static, largely being determined during device fabrication through lithographically defined components. This limits the flexibility of such activation functions to generate different responses and to adapt to different machine learning tasks. Moreover, for a fixed nonlinear response with a relatively large activation threshold, the performance of very deep ONNs with many layers is limited by the high optical loss of many nonlinear activation layers. For instance, the activation threshold of an optical saturable absorption from 2D-materials is fixed by the choice of material in the range of 1-10 mW [13–15]. Therefore, as light propagates through many layers of an ONN, the strength of the nonlinear response becomes continuously weaker, and the signal power may drop below the activation threshold. In contrast, electro-optics opens up the possibility for much stronger and more tunable nonlinearities.

In this work, we fabricate and experimentally demonstrate a recently proposed electro-optic architecture for realizing optical-to-optical activation functions [16]. Related proposals of realizing optical nonlinearities using electro-optics have been proposed in Refs. [17–20]. In our scheme, rather than using traditional optical nonlinearities we fabricate structures on a photonic integrated circuit to measure a small portion of the incoming optical signal power and use electro-optic modulators to modulate the original optical signal. Our prototype relies on thermo-optic modulation but, in principle, the demonstrated activation can use fast modulation mechanisms to enable ONNs operating with GHz-rate computational speeds. This activation circuit allows for the realization of strong nonlinearities without the requirement of having additional optical sources between each layer of the network [20]. We also demonstrate an extension of the circuit capabilities originally proposed in Ref. [16] to realize arbitrary nonlinearities via electrical signal processing with ultra-low activation thresholds. In this work, we focus on the implementation of the activation function that does not use optical gain elements.

The remainder of this paper is organized as follows. In section 2 we review the activation circuit from Ref. [16] and discuss its experimental realization in a silicon nitride (SiN) platform. In section 3, we report the measured activation circuit response and demonstrate that the fabricated device can generate a range of nonlinear optical transfer functions. In section 4, using numerical ONN simulations, we demonstrate that the measured activation functions support low optical transmission losses and high inference accuracy.

2. Nonlinear activation function

In this section, we briefly review the electro-optic activation function architecture proposed in Ref. [16]. We then present an experimental realization of this activation function circuit which is fabricated in a SiN technology platform.

2.1. Optical-to-optical activation function circuit

In this work, we focus on the ONN architecture shown in Fig. 1(a) based on waveguide Mach-Zehnder interferometer (MZI) meshes and reprogrammable phase shifters [4]. A single layer of the ONN architecture consists of one optical interference unit, which implements a unitary matrix-vector multiplication, followed by the element-wise application of the nonlinear activation function circuit, corresponding to the green boxes in Fig. 1(a). The optical-to-optical nonlinear activation function is achieved by converting a small portion of the optical input signal into an electrical signal. The electrical signal is then applied to an intensity modulator which controls the intensity of the remaining portion of the original optical signal. For an input signal with field amplitude z, the resulting nonlinear optical activation function $f(z)$ is a result of both the intensity

Optics EXPRESS

modulator response as well as the components in the electrical signal pathway. The schematic of the proposed nonlinear activation function circuit is shown in Fig. 1(b). The input signal first enters a directional coupler with coupling coefficient of α, which routes a portion, $\alpha|z|^2$, of the input optical power to a photodetector. The photodetector converts the received optical power into an electrical current, $I_{\text{pd}} = \Re \cdot \alpha|z|^2$, where $\Re$ is the photodetector responsivity. A transimpedance amplifier with gain G converts the current into a voltage $V_G = G \cdot \Re \cdot \alpha|z|^2$. The output voltage of the optical-to-electrical conversion circuit is then transformed by a nonlinear signal conditioner with transfer function $H(\cdot)$. Finally, the conditioned voltage signal, $H(V_G)$ is combined with a static bias voltage V_b to generate the modulating signal $V_m = V_b + H\left(G\Re\alpha|z|^2\right)$. This signal modulates the optical signal routed through the intensity modulator, thus implementing an activation function. The transfer function of the activation can be written as

$$f(z) = z\sqrt{1-\alpha}\exp\left(-j2\pi\nu\tau_{\text{opt}}\right)\cdot t_m(z), \tag{1}$$

where ν is the frequency of the optical signal, $t_m(z)$ is the transfer function of the intensity modulator, τ_{opt} is the optical delay time accounting for the contributions from the group delay of the optical-to-electrical conversion stage (τ_{oe}), the delay associated with the nonlinear signal conditioner (τ_{nl}), and the RC time constant of the phase modulator (τ_{rc}) [16]. In practice, the activation function could be realized using a number of different modulator designs, including Mach-Zehnder interferometer (MZI) modulators, ring modulators, and electro-absorption modulators, as shown in the inset of Fig. 1(b). Here we focus on a fabricated device that uses an MZI, which has a transfer function given by

$$t_{\text{MZI}} = j\exp\left(-j\left(\phi_0 + \frac{\Delta\phi}{2}\right)\right)\cos\left(\frac{\Delta\phi}{2}\right), \tag{2}$$

where ϕ_0 is the phase accumulated by the light propagating through the MZI when no voltage is applied to the phase shifter. $\Delta\phi$ is the phase shift induced by the modulating signal V_m and is given by

$$\Delta\phi(V_m) = \frac{\pi}{V_\pi}V_m = \frac{\pi}{V_\pi}\left(V_b + H\left(G\Re\alpha|z|^2\right)\right). \tag{3}$$

V_π is the voltage that the phase shifter requires to produce a π phase shift. The activation function of the proposed circuit with MZI modulator is calculated by substituting Eq. (2) and Eq. (3) in Eq. (1) to obtain:

$$\begin{aligned} f_{\text{MZI}}(z) =& jz\sqrt{1-\alpha}\cos\left(\pi\frac{V_b + H\left(G\Re\alpha|z|^2\right)}{2V_\pi}\right) \\ &\cdot\exp\left(-j\left[2\pi\nu\tau_{\text{opt}} + \phi_0 + \pi\frac{V_b + H\left(G\Re\alpha|z|^2\right)}{2V_\pi}\right]\right). \end{aligned} \tag{4}$$

Equation 4 clearly shows that the activation function exhibits a highly nonlinear response. The strength of the activation function's nonlinearity can be increased by increasing either the TIA gain G or the photodiode responsivity $\Re$. The nonlinearity can also be increased through increasing the directional coupler coefficient α. However, tapping out more optical power undesirably increases the *linear* insertion loss of the circuit. From Eq. (4), the electrical biasing of the activation phase shifter, given by V_b, is an important degree of freedom for controlling the shape of the nonlinear response.

2.2. Fabricated device

Figure 1(c) shows a micrograph of the on-chip nonlinear activation function circuit using a SiN waveguide technology. The circuit consists of a 1:99 directional coupler (DC) and a MZI with a

Optics EXPRESS

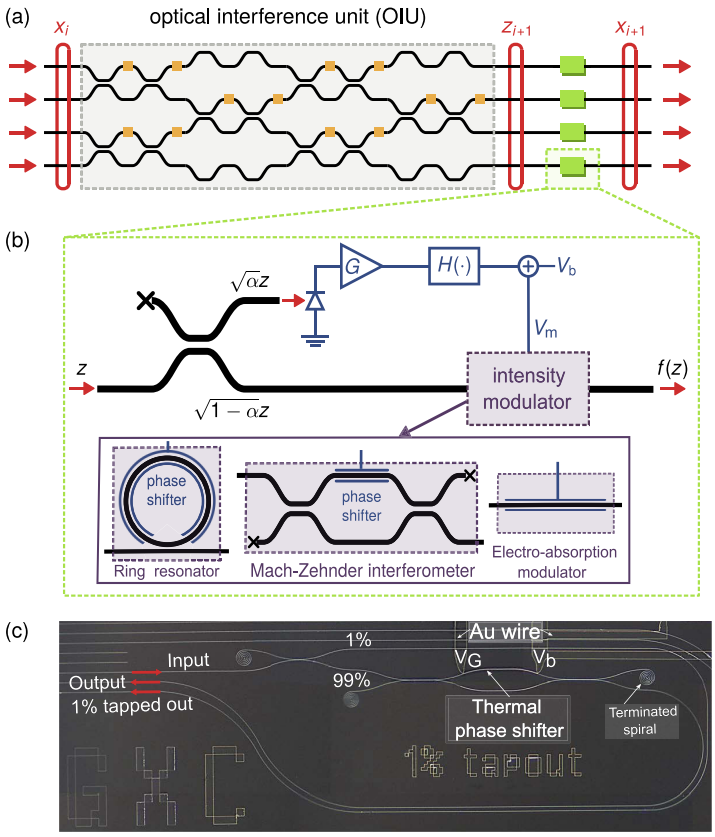

Fig. 1. (a) Schematic of the optical interferometer mesh implementation of a single layer of the feedforward neural network. (b) Schematic of the proposed optical-to-optical activation function circuit. The black and blue lines represent optical waveguides and electrical signal pathways respectively. (c) Optical image of the fabricated activation function circuit with Mach-Zehnder interferometer modulator.

top metal thermal phase shifter. Note that the low speed of the thermal phase shifter limits the operational speed of the prototype device. This limited operational speed is not concerning since the main purpose of this experiment is to demonstrate the capabilities of the proposed circuit to generate nonlinear activation functions. Fabricating the proposed circuit in technologies that provide high-speed intensity modulation, such as silicon photonics, can provide fast operational speed [21–23]. The 1% tapped out port of the DC and the cross-port of the MZI are routed to the edge of the die for edge coupling, while the unused ports are terminated by small spirals, which scatter light due to the small bend radius of the spirals and prevent signal reflection.Fig. 2 shows the measured MZI output for various applied voltages to the phase shifter. The thermal phase shifter requires 12.8V for a π-phase shift and MZI cross-port shows an extinction ratio larger than 40 dB. The insertion loss of the device excluding two 3.25 dB fiber-to-waveguide coupling loss is around 1 dB.

The phase shift in a high-speed intensity modulator typically follows the applied modulating voltage linearly as in Eq. (3). However, thermal phase shifters induce a phase shift proportional to the square of the applied modulating voltage given by

$$\Delta\phi_{\text{thermal}} = \pi \left(\frac{V_m}{V_\pi}\right)^2. \tag{5}$$

Optics EXPRESS

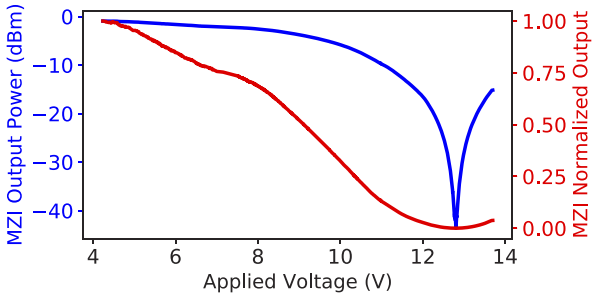

Fig. 2. The output of the MZI for various voltages applied to the phase shift for 0dBm input power to the directional coupler. The blue line shows the output power of the MZI in dBm, and the red line shows the corresponding normalized MZI output.

3. Experimental results

This section details two sets of experiments demonstrating the capabilities of the proposed activation function circuit. Figure 3(a) depicts the measurement setup, while Figs. 3(b) and 3(c) present the block diagram of the two measurement setups. In the first experiment, the tapped out power is converted to a voltage signal and amplified by an optical receiver circuit and directly used to modulate the thermal phase shifter [Fig. 3(b)]. This setup implements a limited number of activation functions. In the second setup, the direct controller is replaced by a re-configurable lookup table that generates arbitrary nonlinear activation functions. The microcontroller in Fig. 3(c) implements the lookup table.

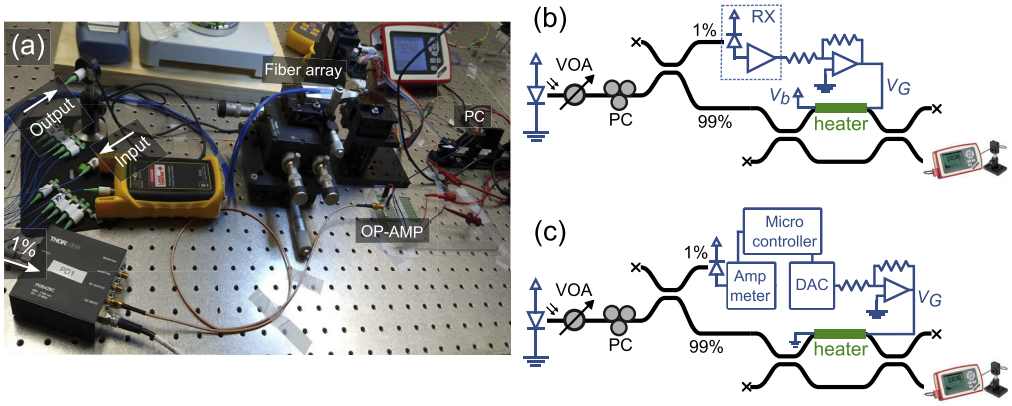

Fig. 3. (a) A photo of the measurement setup, (b) the block diagram of the measurement setup with direct controller, (c) the block diagram of the measurement setup with the lookup table.

To perform the experiments, we first beam a 1550 nm laser through a variable optical attenuator (VOA). The VOA allows us to control the amplitude of the input optical signal. Next, the signal is sent through a polarization controller. Integrated waveguides are polarization sensitive, so the polarization controller is used to minimize the coupling loss. The signal is then sent to the on-chip nonlinear activation function circuit through a fiber array. Finally, the activated signal exits through the fiber array. An electrical probe card is used to control the on-chip thermal phase shifter.

Optics EXPRESS

3.1. Direct controller

We first consider the *direct controller* experimental setup, as shown in Fig. 3(b), which routes the 1% tapped out signal from the directional coupler to a 75 MHz Thorlabs photoreceiver (PDB420C) with a conversion gain of 250K V/W. The maximum input power to the device is limited to -8 dBm to ensure linear photoreceiver operation. The output of the photoreceiver [RX in Fig. 3(b)] is amplified by an operational amplifier [OP-AMP in Fig. 3(a)] and is connected to the thermal phase shifter on the top MZI arm. The opposite end of the thermal heater is connected to a power supply for controlling the initial MZI bias. With this biasing configuration, the effective modulating voltage equals the difference of the bias voltage and the tapped out photo-generated voltage: $V_m = V_G - V_b$.

Figure 4 plots the normalized output power $|f(z)|^2$ as a function of normalized input power $|z|^2$ and compares it with the simulation result, at four different bias voltages applied to the thermal phase shifter. In modeling the performance of the device, we assume that no nonlinear signal conditioning was applied to the electrical signal pathway, i.e. $V_G = G\Re\alpha|z|^2$. We observe excellent agreement between the measured and simulated activation function response, as shown in Fig. 4. The small difference between the measurements and simulation results could be due to the nonlinear response of the photoreceiver. Figures 4(b), 4(c), and 4(d), corresponding to V_b = 12.8V, 14V, and 16V, exhibit a response which is similar to the `ReLU` activation function: optical signal transmission is low for small input values and high for large input values. For the bias of V_b = 14V and 16V, transmission at low input power values is slightly increased compared to the response at V_b = 12.8V. Unlike the ideal `ReLU` response, the activation at V_b = 14V and 16V is not entirely monotonic because transmission first goes to zero before increasing [16]. The response shown in Fig. 4(a), corresponding to V_b = 0.0V, is quite different. It demonstrates a saturation response in which the output is suppressed for higher input values but enhanced for lower input values. As shown in Fig. 4, the bias voltage changes the activation response. The same control circuitry which programs linear interferometer meshes can control the activation response through the bias voltage. The resulting device is a programmable ONN that can implement a range of activation functions.

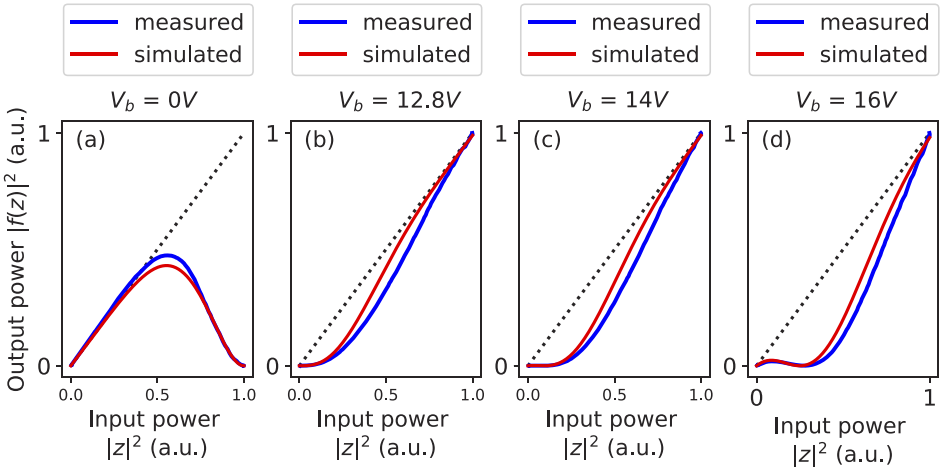

Fig. 4. The normalized output power ($|f(z)|^2$) of nonlinear activation function circuit for various the normalized input power ($|z|^2$) at several bias point. In all of the figures, the normalized output power is relative to the maximum achievable output power of the circuit, which relates to minimum MZI attenuation.

Optics EXPRESS

A fully integrated ONN in a high-speed photonic platform, such as silicon photonics [22,23] would include on-chip high-speed modulators and detectors [24] to modulate and detect the sequences of input vectors to the input layer of the ONN and output vectors of the output layer of the ONN, respectively. The same high-speed detector and modulator elements could also be integrated between the optical interference unit to provide the activation function circuit. State of the art integrated transimpedance amplifiers operate at speeds comparable to the optical modulator and detector rates, which are on the order of 50 - 100 GHz [25,26]. Therefore, the proposed activation function circuit would not be a limiting factor in the speed of the ONN.

3.2. Lookup table controller

We now consider the *lookup table controller* experimental setup, which uses a voltage lookup table to implement the nonlinear electrical signal transformation, H. Specifically, the lookup table maps the tapped out photogenerated current to a modulating voltage applied to the MZI phase shifter. To produce the lookup table, two traces of MZI normalized output as a function of applied voltage to the phase shifter and photogenerated current as a function of optical input power P_{in} are used. We linearly combine these two traces to produce a 2-dimensional map of the optical output power of the activation function circuit as a function of input power to the circuit (P_{in}) and normalized output of MZI. The lookup table is determined by overlaying the target activation function on the map; it is then implemented by a microcontroller. Figure 3(c) shows the block diagram of the test setup with a lookup table controller. The 1% tapped output of the DC is connected to a photodetector with a responsivity $\Re$ of 1 A/W. The photogenerated current of the photodetector is measured by a B&K Precision 393 ammeter. As expected, the measured current is proportional to the optical input power $I_{\text{pd}} = \alpha \Re P_{\text{in}}$. The digital output of the ammeter is sent to a microcontroller to specify the modulating voltage for controlling the phase shifter. The voltage of the phase shifter is set using a lookup table for a specific activation function. Figures 5(a) and 5(b) demonstrate two activation function of `sigmoid` and `modReLU` [27] overlaid on the 2D power throughput map of the activation function circuit, respectively. Figures 5(c) and 5(d) compare the target `sigmoid` function and target `modReLU` function with

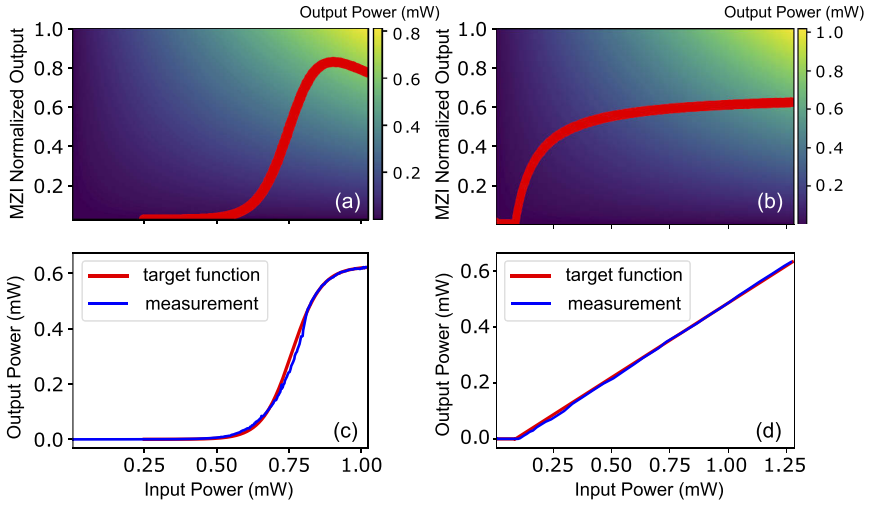

Fig. 5. (a) Target `sigmoid` function overlaid on the 2D power throughput map of the nonlinear activation function, (b) target `modReLU` function overlaid on the 2D power throughput map of the nonlinear activation function, (c) measurement result compared its target `sigmoid` function, (d) measurement result compared to its target `modReLU` function.

Optics EXPRESS

their measurement result respectively. Both measured responses agree very well with the target functions.

Using the lookup table to control the activation response provides a tool to heuristically select an activation function response or to directly optimize the activation function using a training routine. This realization of a controllable optical-to-optical nonlinearity allows ONNs to be applied to a broader classes of machine learning tasks [28]. However, implementing the lookup table on a microcontroller limits the operation speed of the activation function circuit to sub-GHz range. For a specific ONN application, one can use a moderate-speed flexible lookup table implemented on a microcontroller or field-programmable gate array to optimize the activation function. The associated transfer function can then be related to the optimized lookup table, and a piecewise linear approximation can synthesize the optimized transfer function. In a high-speed (GHz) implementation, the circuitry of the piecewise linear function can consist of an application-specific integrated circuit in a high-speed analog/RF circuit platform. A number of technologies with high transit frequencies can be utilized for this purpose. Examples include SiGe BiCMOS, i.e. combination of bipolar and complementary metal–oxide–semiconductor (CMOS) technology, III-V technologies, and advanced CMOS technologies provide high-speed platforms for implementing over 50 GHz bandwidth analog/RF circuits [25,26,29].

4. Machine learning tasks

In this section, we numerically characterize the performance of the activation function on the benchmark machine learning task of classifying images from the MNIST dataset, which consists of 60,000 images of handwritten digits ranging from 0 to 9. The ONN setup is shown schematically in Fig. 6(a), and consists of a sequence of linear layers, corresponding to interferometer meshes [4], and nonlinear activation layers. The last layer is a drop layer that reduces the vector to a length of 10 elements, suitable for one-hot detection across the 10 digit classes. After the drop layer, the optical intensity is detected and passed through a softmax function. As in Ref. [16], before entering the ONN, the images undergo a pre-processing stage consisting of a Fourier transform step and a cropping step. These operations reduce the total size of the input data from $28 \times 28 = 784$ real-space pixels to 16 complex Fourier coefficients. We found that an ONN with 16 inputs resulted in reasonably high classification performance, but was still feasible to simulate and train numerically. In practice, the Fourier transformation and cropping steps could be experimentally achieved completely passively with a Fourier optics setup [30].

We now compare the classification performance of the ONN on the digit recognition task for several nonlinearity settings and quantify the optical transmission through the ONN. Figures 6(b)–6(c) show the classification accuracy of the ONN on the test dataset and the optical transmission through the ONN as a function of the network depth. The transmission shown in Fig. 6(c) is calculated as the mean over the transmission for all samples in the training dataset. These simulations were performed using `TensorFlow` [31] and the `neurophox` ONN modeling framework [32,33], which implements a physical model of the ONN by parameterizing the linear layers in terms of MZI interferometers and phase shifters and complex-valued field quantities. The ONN is trained using the Adam optimizer [34] for 400 epochs with a batch size of 512.

In our comparison, we consider several variants of the ONN in Fig. 6(a): a linear ONN with no activation, an electro-optic activation that uses settings similar to those in Ref. [16], and an electro-optic activation implementing the complex `modReLU` function [27] corresponding to the lookup table implementation measured from our prototype in Fig. 5. Unsurprisingly, we observe in Fig. 6 that the linear ONN does not benefit from an increase in the network depth because a sequence of linear transformations is also a linear transformation. In other words, additional linear layers without intermediate nonlinearities do not meaningfully increase the learning capacity of the ONN. The linear ONN achieves a test accuracy of 82% and, because

Optics EXPRESS

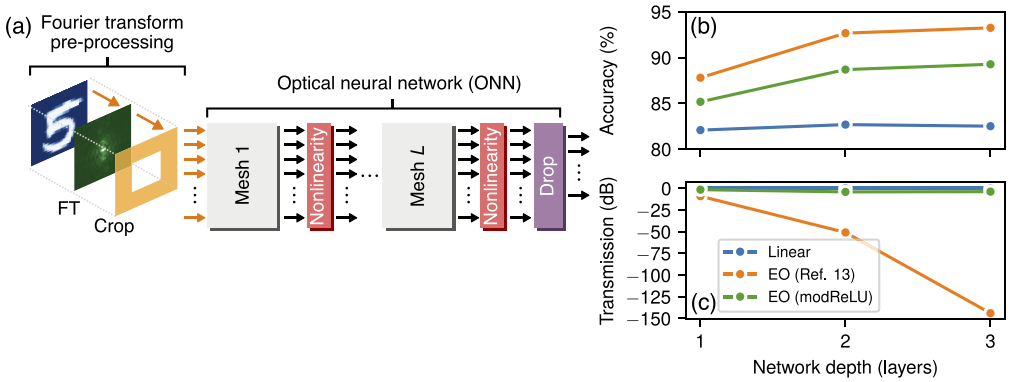

Fig. 6. (a) Schematic of the optical neural network configuration for classifying handwritten digits from the MNIST dataset. The 28×28 pixel images are first Fourier transformed and cropped before being fed into the ONN. (b) Classification accuracy achieved for various ONN configurations. (c) Optical signal transmission through the neural network, averaged over all samples in the training dataset.

we have assumed lossless interferometer meshes, it exhibits ideal optical transmission which is independent of the number of layers.

In contrast, the electro-optic activation function with settings similar to those used in Ref. [16] increases its classification accuracy substantially with additional layers. This ONN achieves a test accuracy of 93% with three layers. However, this relatively high accuracy comes with a high cost in terms of optical signal attenuation. Although nonlinear amplitude responses inherently involve signal attenuation, this activation configuration results in an optical transmission of -144 dB for the network with three layers. In practice, such loss could be prohibitively high due to the finite dynamic range of optical detectors at the output of the ONN.

However, by configuring the electro-optic activation lookup table to synthesize the `modReLU` function, the optical transmission can be increased significantly. We observe that the `modReLU` response results in an optical transmission of -4 dB for the 3 layer network, which is 140 dB larger than the transmission through the network with the electro-optic activation settings from Ref. [16]. However, we note that the ONN with the `modReLU` activation does have a classification accuracy that is reduced by 5% from the activation in Ref. [16]. However, the ONN with the `modReLU` activation still outperforms the linear ONN. The performance of the ONN with the `modReLU` activation could potentially be improved by adjusting (or directly training) the activation threshold. We emphasize that the ability to synthesize the `modReLU` activation is a unique capability of this electro-optic activation function architecture and is an important degree of freedom over all-optical nonlinearities. We note that constraining the ONN to $N = 16$ Fourier coefficients from each input image does somewhat limit accuracy of the MNIST task. Other works have demonstrated that increasing N can lead to an increased classification accuracy in ONNs [35], approaching the performance of conventional artificial neural networks.

5. Conclusion

In this work, we have presented the experimental results of an on-chip optical-to-optical nonlinear activation function circuit fabricated on a SiN waveguide technology platform. The capabilities of the circuit were demonstrated through two experimental setups. In the first experiment, only the nonlinear response of the Mach-Zehnder modulator was used to generate the nonlinear activation function. In this setup, a limited set of activation functions could be realized by varying the bias of the phase shifter. In the second experiment, a lookup table was used to apply a

Optics EXPRESS

nonlinear modulation signal to the phase shifter which allowed realization of arbitrary nonlinear responses. While the prototype demonstrated in this work relied on thermo-optic modulation, the activation architecture can be readily implemented using much faster modulation mechanisms that are widely used in GHz-rate optical communications [21–23]. Faster modulation will allow an ONN using this activation to achieve higher computational speeds and lower latencies than conventional digital processors.

Using numerical ONN simulations, we demonstrated that the measured activation functions improve the accuracy of optical neural networks on the benchmark task of classifying images from the MNIST dataset. Our simulations revealed that the ability to generate arbitrary nonlinear optical transfer functions provides a powerful tool to achieve high performance while maintaining a low optical transmission loss. Compared to a linear ONN with depth of three layers, using the activation from Ref. [16] improves the accuracy of the classification task by more than 11% but at the cost of over 140 dB optical transmission loss. However, by configuring the lookup table to generate the `modReLU` activation [27] instead of the response from Ref. [16], the optical transmission is improved by more than 140 dB with only 5% degradation in classification accuracy. The ability of this activation to tailor the loss of optical nonlinearities may be very useful for much deeper neural networks where signal attenuation is a significant concern. Future work could consider the effect of including optical loss from the activation as a penalty term into the objective function during training to balance inference performance and the output signal-to-noise ratio.

Funding

Air Force Office of Scientific Research (FA9550-17-1-0002, MURI Project).

Disclosures

The authors declare no conflicts of interest.

References

1. Y. S. Abu-Mostafa and D. Psaltis, "Optical Neural Computers," Sci. Am. **256**(3), 88–95 (1987).
2. D. Psaltis, D. Brady, X.-G. Gu, and S. Lin, "Holography in artificial neural networks," Nature **343**(6256), 325–330 (1990).
3. A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," Sci. Rep. **7**(1), 7430–7439 (2017).
4. 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. Photonics **11**(7), 441–446 (2017).
5. D. Pérez, I. Gasulla, L. Crudgington, D. J. Thomson, A. Z. Khokhar, K. Li, W. Cao, G. Z. Mashanovich, and J. Capmany, "Multipurpose silicon photonics signal processor core," Nat. Commun. **8**(1), 636–644 (2017).
6. D. Marpaung, J. Yao, and J. Capmany, "Integrated microwave photonics," Nat. Photonics **13**(2), 80–90 (2019).
7. R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," Phys. Rev. X **9**(2), 021032 (2019).
8. D. A. B. Miller, "Self-configuring universal linear optical component," Photonics Res. **1**(1), 1–15 (2013).
9. M. W. Puckett, J. Wang, D. Bose, G. M. Brodnik, J. Wu, K. Nelson, and D. J. Blumenthal, "Silicon nitride ring resonators with 0.123 dB/m loss and Q-factors of 216 million for nonlinear optical applications," in *2019 Conference on Lasers and Electro-Optics Europe and European Quantum Electronics Conference*, (Optical Society of America, 2019), pp. ce–11–3.
10. L. Zhou, K. Okamoto, and S. J. B. Yoo, "Athermalizing and trimming of slotted silicon microring resonators with uv-sensitive pmma upper-cladding," IEEE Photonics Technol. Lett. **21**(17), 1175–1177 (2009).
11. Y. Zhang, Y. Li, S. Feng, and A. W. Poon, "Towards adaptively tuned silicon microring resonators for optical networks-on-chip applications," IEEE J. Sel. Top. Quantum Electron. **20**(4), 136–149 (2014).
12. M. Radulaski, R. Bose, T. Tran, T. Van Vaerenbergh, D. Kielpinski, and R. G. Beausoleil, "Thermally Tunable Hybrid Photonic Architecture for Nonlinear Optical Circuits," ACS Photonics **5**(11), 4323–4329 (2018).
13. Q. Bao, H. Zhang, Z. Ni, Y. Wang, L. Polavarapu, Z. Shen, Q.-H. Xu, D. Tang, and K. P. Loh, "Monolayer graphene as a saturable absorber in a mode-locked laser," Nano Res. **4**(3), 297–307 (2011).

Optics EXPRESS

14. N. H. Park, H. Jeong, S. Y. Choi, M. H. Kim, F. Rotermund, and D.-I. Yeom, "Monolayer graphene saturable absorbers with strongly enhanced evanescent-field interaction for ultrafast fiber laser mode-locking," Opt. Express **23**(15), 19806–19812 (2015).
15. X. Jiang, S. Gross, M. J. Withford, H. Zhang, D.-I. Yeom, F. Rotermund, and A. Fuerbach, "Low-dimensional nanomaterial saturable absorbers for ultrashort-pulsed waveguide lasers," Opt. Mater. Express **8**(10), 3055–3071 (2018).
16. 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. Top. Quantum Electron. **26**(1), 1–12 (2020).
17. A. L. Lentine and D. A. B. Miller, "Evolution of the SEED technology: Bistable logic gates to optoelectronic smart pixels," IEEE J. Quantum Electron. **29**(2), 655–669 (1993).
18. A. Majumdar and A. Rundquist, "Cavity-enabled self-electro-optic bistability in silicon photonics," Opt. Lett. **39**(13), 3864–3867 (2014).
19. N. C. Harris, "Programmable nanophotonics for quantum information processing and artificial intelligence," Thesis, Massachusetts Institute of Technology (2017).
20. A. N. Tait, T. Ferreira de Lima, M. A. Nahmias, H. B. Miller, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, "Silicon Photonic Modulator Neuron," Phys. Rev. Appl. **11**(6), 064043 (2019).
21. J. E. Roth, O. Fidaner, R. K. Schaevitz, Y.-H. Kuo, T. I. Kamins, J. S. Harris, and D. A. B. Miller, "Optical modulator on silicon employing germanium quantum wells," Opt. Express **15**(9), 5851–5859 (2007).
22. T. Shi, T.-I. Su, N. Zhang, C. yin Hong, and D. Pan, "Silicon Photonics Platform for 400G Data Center Applications," in *Optical Fiber Communication Conference*, (Optical Society of America, 2018), p. M3F.4.
23. T. Baehr-Jones, R. Ding, A. Ayazi, T. Pinguet, M. Streshinsky, N. Harris, J. Li, L. He, M. Gould, Y. Zhang, A. Eu-Jin Lim, T.-Y. Liow, S. Hwee-Gee Teo, G.-Q. Lo, and M. Hochberg, "A 25 Gb/s Silicon Photonics Platform," arXiv e-prints arXiv:1203.0767 (2012).
24. M. M. P. Fard, G. Cowan, and O. Liboiron-Ladouceur, "Responsivity optimization of a high-speed germanium-on-silicon photodetector," Opt. Express **24**(24), 27738–27752 (2016).
25. G. Yu, X. Zou, L. Zhang, Q. Zou, M. Zheng, and J. Zhong, "A low-noise high-gain transimpedance amplifier with high dynamic range in 0.13ìm CMOS," in *2012 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, (2012), pp. 37–40.
26. M. N. Ahmed, J. Chong, and D. S. Ha, "A 100 Gb/s transimpedance amplifier in 65 nm CMOS technology for optical communications," in *2014 IEEE International Symposium on Circuits and Systems (ISCAS)*, (2014), pp. 1885–1888.
27. C. Trabelsi, O. Bilaniuk, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep complex networks," CoRR **abs/1705.09792** (2017).
28. C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of trends in Practice and Research for Deep Learning," arXiv e-prints arXiv:1811.03378 (2018).
29. I. García López, A. Awny, P. Rito, M. Ko, A. C. Ulusoy, and D. Kissinger, "100 gb/s differential linear tias with less than 10 pa/ $\sqrt{\text{hz}}$ in 130-nm sige:c bicmos," IEEE J. Solid-State Circuits **53**(2), 458–469 (2018).
30. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company Publishers, 2005).
31. A. Agrawal, A. N. Modi, A. Passos, A. Lavoie, A. Agarwal, A. Shankar, I. Ganichev, J. Levenberg, M. Hong, R. Monga, and S. Cai, "TensorFlow Eager: A Multi-Stage, Python-Embedded DSL for Machine Learning," arXiv:1903.01855 [cs] (2019).
32. "Neurophox: A simulation framework for unitary neural networks and photonic devices," https://github.com/solgaardlab/neurophox.
33. S. Pai, B. Bartlett, O. Solgaard, and D. A. B. Miller, "Matrix Optimization on Universal Unitary Photonic Devices," Phys. Rev. Appl. **11**(6), 064044 (2019).
34. D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," arXiv:1412.6980 [cs] (2014).
35. S. Pai, I. A. D. Williamson, T. W. Hughes, M. Minkov, O. Solgaard, S. Fan, and D. A. B. Miller, "Parallel fault-tolerant programming of an arbitrary feedforward photonic network," arXiv:1909.06179 [physics] (2019).