Silicon microring synapses enable photonic deep learning beyond 9-bit precision

WEIPENG ZHANG¹, CHAORAN HUANG¹,², HSUAN-TUNG PENG¹, SIMON BILODEAU¹, AASHU JHA¹, ERIC BLOW¹, THOMAS FERREIRA DE LIMA¹,⁵, BHAVIN J. SHASTRI³,⁴, AND PAUL PRUCNAL¹

1 Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08540, USA
2 The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China
3 Department of Physics, Engineering Physics and Astronomy, Queen's University, Kingston, Ontario K7L 3N6, Canada
4 Vector Institute, Toronto, Ontario M5G 1M1, Canada
5 Now at NEC Laboratories America, Princeton, NJ 08540, USA

*Corresponding author: weipengz@princeton.edu

1. INTRODUCTION

Deep neural networks (DNN) consist of layers of neurons interconnected by synaptic weights. A high bit-precision in weights is generally required to guarantee high accuracy in many applications. Minimizing error accumulation between layers is also essential when building large-scale networks. Recent demonstrations of photonic neural networks are limited in bit-precision due to crosstalk and the high sensitivity of optical components (e.g., resonators). Here, we experimentally demonstrate a record-high precision of 9 bits with a dithering control scheme for photonic synapses. We then numerically simulated the impact with increased synaptic precision on a wireless signal classification application. This work could help realize the potential of photonic neural networks for many practical, real-world tasks. ©2022 Optica Publishing Group. Users may use, reuse, and build upon the article, or use the article for text or data mining, so long as such uses are for non-commercial purposes and appropriate attribution is maintained. All other rights are reserved. © 2022 Optica Publishing Group. Users may use, reuse, and build upon the article, or use the article for text or data mining, so long as such uses are for non-commercial purposes and appropriate attribution is maintained. All other rights are reserved.

http://dx.doi.org/xx.xxxx/XX.XX.XXXXXX

Resonator-based approaches to optical synapses, such as MRs [18], are appealing due to their compactness, sensitivity, and innate wavelength-division multiplexing (WDM) compatibility. However, this resonant nature results in inter-channel crosstalk (from overlapping spectral filter responses) and susceptibility to spurious environmental fluctuations (from high sensitivity). Many methods have been proposed to improve
weighting control, including feedback methods [19] to address the high sensitivity and a feedforward model [20] to overcome inter-channel crosstalk. For years, however, the control precision of MRR-based photonic synapse has seemingly plateaued at around 7 bits [17, 21].

The limited precision, from our perspective, is due to the lack of monitoring of the entire signal path of the photonic synapse. Previous control methods only monitor drifts of the MRR resonances and are insensitive to other fluctuations. We observed that every component in the input signal path could cause weight drifting and should be considered. For example, the electrical-to-optical modulator is the component that converts the input electrical signal onto the photonic path. If its modulation depth fluctuates (usually caused by polarization drift or temperature drift), the amplitude of the output signal, namely the weight, will drift even if the MRRs remain perfectly stable. This scenario, however, is not necessarily captured by the other feedback methods since the MRR remains on resonance. Thus, the current performance bottleneck is caused by drifting sources other than the MRR synapse. To compensate for the thermal drift of a single microring, a dithering scheme [23–26] was proposed to lock the microring to its resonance frequency in a static fashion. Photonic synapses, however, require a novel implementation of the “dithering” since more than one device needs to be controlled, and each is dynamically tuned for weighting as opposed to being statically stabilized. Also, other elements in synapses, besides the MRR itself, can induce drifts in addition to the thermal drift.

Here, we develop a new dithering scheme to control multiple photonic synapses simultaneously and experimentally achieve a record high precision of 9 bits. This scheme introduces a dithering control signal to inputs, allowing monitoring and stabilizing of the entire optical link comprising the photonic synapses, from the MZM (where the RF signals enter the system) to the BPD (where they exit the system), as shown in Figure 1(b). With this, drifts of weighting caused by environmental changes can be dynamically tracked and compensated along the entire link. More importantly, our scheme can also track and compensate for the drift due to inter-channel crosstalk. Thus, accurate weighting can be guaranteed, especially for high fan-in neurons with multi-
Fig. 2. Procedure of the dithering control. Mat. matrix. The calibration procedure consists of two steps. Step 1: Set current to zero and record the dithering waveforms as the reference of weights equal to one. Step 2: Sweep the current of all the MRRs, record the dithering waveforms and calculate the weights for each MRR. This step generates a weight-current lookup curve for each MRR. The weight control procedure performs a multi-channel binary search to tune the MRRs to the commanded weights. The lookup curves obtained in the calibration procedure provide the initial searching points for this weight control procedure.

channel frontends (multiple MRRs). We tested this method on a two-MRR synaptic system (i.e., a neuron with inputs) and achieved about 9 bits of precision, which, to our knowledge, is the highest recorded bit precision in a photonic neuromorphic architecture. We also simulated a photonic neural network for a real-world wireless signal classification problem, and the result shows that increasing the weighting precision from 5 bits[19] to 9 bits will increase the classification accuracy from only 50% to over 90% in some cases. Such an improvement makes photonics competitive against its analog electronic counterparts, paving the way for practical photonic neural networks for real-world tasks, including but not limited to intelligent signal processing of RF signals [27, 28] and optical communications [29].

2. METHOD AND SETUP

MRRs are WDM-compatible tunable filters where multiple MRRs of different radii can provide parallel weighting to lightwaves of different wavelengths. Accordingly, MRR-based photonic synapses can be realized through the "Broadcast-and-weight" architecture [30]. This architecture establishes connections between a pair of neurons where input signals from each upstream neuron are amplitude-modulated at different carrier wavelengths, then filtered or "weighted" by MRRs, and summed optoelectronically, resulting in an optical output signal that is sent to downstream neurons [12]. In this way, the weighting precision of each synapse is key to the overall performance of the photonic neural network.

A typical MRR-based synaptic system is depicted in Figure 1(b), where the signal path starts at the Mach-Zehnder modulators (MZM) and ends at the photodetector. The optical intensity delivered from the laser sources to the MRR weights is constant in an ideal system. In practice, however, polarization and temperature drifts, unstable optical alignment, or changing electrical parasitics can cause uncompensated power fluctuation resulting in an effective weight error. Fortunately, our dithering control scheme can address all these non-idealities by monitoring the entire signal path, bringing about performance improvement of photonic NNs.

As shown in Figure 1(b), we use an MRR weight bank that can process two signal channels (see detailed description in Ref. [28]). To enable dithering control in the system, we superimpose a dithering signal, a predefined sinusoidal wave (generated by 33220A, Keysight) with a frequency usually much lower (< 10MHz) and an amplitude much smaller (< 100mV) than the input signals, into each signal path. These dithering signals can be separated with a diplexer (ZDPLX-2150-S+, Mini-Circuits) and captured synchronously by the oscilloscope (DPO4032, Tektronix). The oscilloscope is triggered by the sync output of the reference signal generator to align each captured waveform of the dithering signals. Since dithering signals share the same path with the signals being weighted, the variations of the output dithering signals reflect the actual weight value in real-time. Thus, an accurate weight can be continuously monitored regardless of the input data statistics, such as sparsity, variance, etc.

The dithering signal is introduced through the built-in bias tee of the MZMs (see the inset of Figure 1(b)). Instead of applying pure DC voltages to the bias ports, we dither the bias voltages at predefined frequencies while the RF input port remained unchanged. Thus, the only additional components required are a diplexer and a low-end oscillator. Therefore, this method can be seamlessly integrated into many demonstrated optical synaptic setups. Scaling this dithering technique can be done in two ways. For a small system, say less than four microrings in total, new dithering signals of other harmonic frequencies must be added, which can be achieved using a frequency doubler or new generators. For a larger system (> 4 microrings), a more affordable and efficient way is to use an RF switch. In this way, the dithering signal can be applied to each microring at a time in a time-multiplexed manner. Both scenarios require only one diplexer and one scope channel to separate and record the dithering signals, which do not scale with the number of synapses and help maintain a low power budget. We demonstrate the first strategy using a 2 MRR synapse system. This method eliminates the need to directly sense the MRR resonances locally via a dedicated circuit, as with the feedback methods. We can replace the source-measure-units (SMU) with voltages/current sources for driving MRRs, usually of lower cost and smaller footprint, as shown in Figure 1(e). Removing sensing also eases the electrical overhead required for large-scale MRR synapse actuation. We include more details of the PCB design in the supplemental material.

The operational principle of this dithering control method is illustrated in Figure 2. First, a two-step calibration process is done to obtain a weight-current \((W \times I)\) lookup curve for
each channel, providing initial searching points for the subsequent weight control process. During calibration, the first step (refer to step 1 in Figure 2) is to tune all the MRRs to a reference position by applying predefined fixed currents \(i_{ref,m}\). Meanwhile, the MZMs are sequentially dithered, and the output waveforms (captured by the scope) compose the reference dithered signals \(r_m\), for each channel \(m\). Given the sampling length of \(N\) and the total number of channels \(M\), the recorded reference signals can be expressed as \(r_{m}(t)\), where \(t = 0, \Delta T, 2\Delta T, 3\Delta T, \ldots, (N - 1)\Delta T\) and \(m = 1, 2, \ldots, M\) (index of the channel). Then a reference matrix \(A\) can be constructed using all the \(r_m(t)\) as its column (as shown in Eq. 1).

\[
A = [r_1(t), r_2(t), \ldots, r_M(t)]^T
\]

The next step (step 2 in Figure 2) is to turn dithering on for all the channels and do a full-range multidimensional current sweep of all channels. The scope records frames of the dithering signals at each intermediate point in the sweep. For this procedure, each captured waveform, \(S(t) = s(0), s(\Delta T), s((N - 1)\Delta T)\) represents the addition of all the dithering signals. The amplitude of each dithering frequency can be decomposed using Eq. 2.

\[
W = (w_1, w_2, \ldots, w_M) = A^{-1} \times \frac{S}{\sum_{n=1}^{N}}
\]

where \(A^{-1}\) is the pseudo-inverse matrix of matrix \(A\). In the vector \(W = (w_1, w_2, \ldots, w_M)\), each element \(w_m\) represents a correlation between the weighted sum waveform and the reference dithering signals, therefore corresponding to the real-time measurement of the weights. Note that this sweep only needs to be carried out once regardless of the channel count since each captured waveform can decompose the weights for all the channels. Then, we can obtain the \(W \times I\) (weight-current) lookup curves corresponding to the applied current and the output weight for each channel, as step 2 of the calibration procedure shown in Figure 2. In our lab setup, the frequency of the calibration step needs to be done once a day. Whenever a new set of weights needs to be commanded, the dithering will be turned on, allowing the subsequent binary search to precisely tune the microrings to the target weights. The lookup curves obtained in the previous calibration procedure provide the starting points. The binary search iteratively adjusts the current setpoints \(I_{set}\) of the microring according to the measured weight outputs \(W_{get}\), as the weight control procedure shown in Figure 2. It is worth noting that binary searching is carried out with multi-dimensional searching, which means the weights of all the channels are measured and adjusted simultaneously without scaling search time with channel count. The multi-dimensional searching also minimizes inter-channel crosstalk since the applied currents account for inter-channel interference. More details on this weight searching process can be found in section 1 of the Supplemental Information.

3. RESULT

To show the effectiveness of the dithering control method, we evaluated the agreement between the dithering signal and the
actual signal. We apply a 10 MHz dithering signal atop a 120 MHz sine wave representing the synaptic signal. After the calibration steps, we swept the current (roughly from 0.5 to 1.6 mA) applied to the MRR heater to vary the weights in equally spaced increments within a [-1,1] interval as shown in Figure 3(a). Meanwhile, we recorded the weights derived from the output signal waveforms (high-frequency output of the diplexer) and the dithering signals (diplexer low-frequency output) on two oscilloscope channels. Figure 3(a) shows the measured weights for various applied currents. The dashed and solid curves represent the weights (left axis) measured by the actual and dithering signals. The error between the two weight values, calculated as the orange curve in the figure, is within 0.5% across the whole range. This confirms that this dithering method can be used as an accurate proxy of the actual weight.

Next, the performance evaluation was extended to use two MRR synapses. Figure 3(b) shows the result on a mesh-plot. This type of mesh-estimation was used in previous work [17, 20], in which two MRRs are tuned to equidistant grid points of \((w_{1,1}, w_{1,2})\), with \(w_{1,2} \in [-1, 1]\), and the weight accuracy is evaluated at each grid point. The standard deviation of the measured errors at each point (\(\sim 0.0039\)) gives the equivalent bit-precision of the weight control performance. We chose a 9x9 mesh, and for each grid point, we repeated the weight search three times to test the repeatability of the process. Figure 3(c) shows the aggregation of the errors derived from all the tested points. Compared with previous work, the error is significantly reduced, and a precision of 9.0 bits is achieved. This reflects a four-fold (or 2-bit) improvement from the previously best-reported result [21], and over 5-bit improvement from the first reported MRR synapse [31].

Besides the improved accuracy, the multi-dimensional weight searching also alleviates inter-channel crosstalk. Such crosstalk is mainly due to the thermal interference between adjacent photonic devices [32, 33], which causes small weight errors in adjacent MRRs. To evaluate the improvement in crosstalk, we did a comparison test whose results are shown in Figure 4. Here, we first tuned the weights to \((w_{1,1}, w_{1,2}) = (-0.5, -1.0)\), then swept the weight of channel 1 using the dithering weight while keeping a constant current applied to channel 2. This method corresponds to a one-dimensional search in which only channel 1 was considered (refer to the dashed orange curve).

Because of the inter-channel crosstalk, when the current applied to MRR 1 is changed, the applied current to MRR 2 needs to be slightly adjusted. Without that adjustment, the weight of MRR 2 drifts by up to 8.8% (max \(w - \min w = 0.088\)), as indicated by the orange curve in the figure. Next, we performed a two-dimensional weight control at the same command weights (solid line in Fig. 4). In contrast to the one-dimensional case, the multi-dimensional weight control scheme always considers all channels for every command set of weights. It thus can compensate for crosstalk as indicated by the blue curve, reducing the error to 1.5% (max \(w - \min w = 0.013\)). The nature of the multi-dimensional search keeps this performance advantage valid and more significant as the number of MRR channels scales up.

Another benefit of the multi-dimensional search scheme is avoiding long searching time, especially for large-scale MRR synapses, since it does not scale linearly with the number of MRRs. For example, we compared the time it took to find weight(s) for one MRR and two MRRs simultaneously. After averaging 15 trials, we observed an increase of about 12% for adding a new MRR (13s to 16s), rather than the expected 100% increase corresponding to a linear increase. This advantage can be explained by the frequency decomposition of the dithering signals: during one search frame, capturing one waveform can effectively measure the weights of all the MRRs. Accordingly, the currents applied to all the MRRs can be simultaneously adjusted in a single step.

To illustrate the impact of the increased bit resolution in photonic neural networks, we performed a numerical simulation of a photonic NN model under varying weight precision levels. We used PyTorch [34] (in Python) to simulate a model of a continuous-time recurrent neural network (CTRNN) [22] (shown in Figure 1(a)), realized by an MRR-based photonic neural network coupled with an FPGA. This network was recently successfully employed for wireless signal classification [35]. The task is to identify 30 identical ZigBee device transmissions by a photonic recurrent neural network on a wireless signal classification task under SNR of 10dB, 15dB, 20dB, 25dB and 30dB.
ground noise. Assuming a minimum threshold of 80% accuracy, increasing precision from 7- to 9-bits can compensate for an SNR degradation from 30dB to 20dB. Given this result, this improvement of bit-precision by this proposed dithering control method proves to be practically effective for boosting the performance of photonic NNs, which is especially useful for low-SNR inputs.

The supplemental material includes another numerical simulation on a feedforward DNN, a cascading and fully-connected network with two neurons in each layer. By increasing the network size under the conditions of different bit-precision, the simulation explores how the weighting error (the deviation from the expected weight because of a limited bit-precision) accumulates when expanding the network depth under the conditions of varying bit-precision. We find that increased bit-precision reduces the error accumulated in the network, which allows for large networks. Specifically, we find that a 2-bit improvement in precision (from 7 to 9 bit) enables a network of three times more layers (from 5 to 18) by maintaining the signal-to-noise ratio (SNR=20). This result highlights another benefit of our dithering method: the feasibility of building large networks while maintaining signal fidelity, fitting the needs for more applications.

4. CONCLUSION

In summary, we proposed and tested a dithering control method for photonic synapses realized by MRRs, which can boost the accuracy to 9 bits, which is 2 bits higher than the previous best result. The increased bit precision results in increase of prediction accuracy and enables large networks with a reasonable cost. As aforementioned, our proof-of-principle setup uses a benchtop dithering generator, scope, and CPU to complete the control loop. This limits the speed by a few tens of seconds (for each weight), the precision by 9-bit, the ease in scalability, and the overall footprint. A solution is to use an FPGA with high-speed RF I/O coupled with a lock-in circuit, as shown in [23], to replace these bulky instruments and carry the dithering control algorithm.

Our dithering approach has several advantages relative to prior work: it is scalable and can accommodate multi-channel MRR synapses, eliminates the crosstalk issue, and speeds up the searching time (thanks to the multi-dimensional searching strategy), requires less frequent calibration, and can be easily incorporated into other photonic synapses. These advantages have great potential for current and future resonator and non-resonator-based techniques, especially in photonic NN applications. The resulting high accuracy makes it competitive against its electronic counterparts, thus benefiting various silicon photonics applications and narrowing the performance gap between electronic and photonic approaches.

Funding. Defense Advanced Research Projects Agency (HR0011990049). National Science Foundation (ECSS-1642962). The devices were fabricated at the Advanced Micro Foundry (AMF) in Singapore through the support of CMC Microsystems. B. J. Shastri acknowledges support from the Natural Sciences and Engineering Research Council of Canada (NSERC). S. Bilodeau acknowledges funding from the Fonds de recherche du Québec - Nature et technologies.

Disclosures. The authors declare no conflicts of interest.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

1. A. C. Mater and M. L. Coote, “Deep learning in chemistry,” J. chemical information modeling 59, 2545–2559 (2019).
2. C. Angermueller, T. Pärnamaa, L. Parts, and O. Stegle, “Deep learning for computational biology,” Mol. systems biology 12, 878 (2016).
3. J. B. Heaton, N. G. Polson, and J. H. Witte, “Deep learning for finance: deep portfolios,” Appl. Stoch. Model. Bus. Ind. 33, 3–12 (2017).
4. Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang, “Machine learning and deep learning methods for cybersecurity,” Ieee access 6, 35385–35381 (2018).
5. P. Judd, J. Albericio, T. Hetherington, T. Aamodt, N. E. Jerger, R. Urtausen, and A. Moshovos, “Proteus: Exploiting precision variability in deep neural networks,” Parallel Comput. 73, 40–51 (2018).
6. N. Mellempudi, S. Srinivasan, D. Das, and B. Kaul, “Mixed precision training with 8-bit floating point,” arXiv preprint arXiv:1905.12334 (2019).
7. T. S. Nazaré, G. B. P. da Costa, W. A. Contato, and M. Ponti, “Deep convolutional neural networks and noisy images,” in Iberoamerican Congress on Pattern Recognition, (Springer, 2017), pp. 416–424.
8. N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agravat, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers et al., “In-datacenter performance analysis of a tensor processing unit,” in Proceedings of the 44th annual international symposium on computer architecture, (2017), pp. 1–12.
9. T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, “Training of photonic neural networks through in situ backpropagation and gradient measurement,” Optica 5, 864–871 (2018).
10. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. Pernice, “All-optical spiking synapsic networks with self-learning capabilities,” Nature 569, 208–214 (2019).
11. P. R. Prucnal, B. J. Shastr1, and M. C. Teich, Neuromorphic photonics (CRC Press, 2017).
12. B. J. Shastr1, A. N. Tait, T. F. de Lima, W. H. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, “Photonics for artificial intelligence and neuromorphic computing,” Nat. Photonics 15, 102–114 (2021).
13. X. Li, N. Youngblood, C. Rios, Z. Cheng, C. D. Wright, W. H. Pernice, and H. Bhaskaran, “Fast and reliable storage using a 5 bit, nonvolatile photonic memory cell,” Optica 6, 1–6 (2019).
14. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larocheille, D. Englund et al., “Deep learning with coherent nanophotonic circuits,” Nat. Photonics 11, 441–446 (2017).
15. C. Demirkiran, F. Eris, G. Wang, J. Elmhurst, N. Moore, N. C. Harris, A. Basumallik, V. J. Reddi, A. Joshi, and D. Bunandar, “An electrophotonic system for accelerating deep neural networks,” arXiv preprint arXiv:2109.01126 (2021).
16. S. Bandypadhyay, R. Hameiry, and D. Englund, “Hardware error correction for programmable photonics,” arXiv preprint arXiv:2103.04993 (2021).
17. C. Huang, S. Bilodeau, T. Ferreira de Lima, A. N. Tait, P. Y. Ma, E. C. Blow, A. Jha, H.-T. Peng, B. J. Shastr1, and P. R. Prucnal, “Demonstration of scalable microring weight bank control for large-scale photonic integrated circuits,” APL Photonics 5, 040803 (2020).
18. W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, “Silicon microring resonators,” Laser & Photonics Rev. 6, 47–73 (2012).
19. A. N. Tait, H. Jayatilleka, T. F. D. Lima, P. Y. Ma, M. A. Nahmias, B. J. Shastr1, S. Shekhar, L. Chrostowski, and P. R. Prucnal, “Feedback control for microring weight banks,” Opt. Express 26, 26422–26443 (2018).
20. A. N. Tait, T. F. de Lima, M. A. Nahmias, B. J. Shastr1, and P. R. Prucnal, “Multi-channel control for microring weight banks,” Opt. Express 24, 8895–8906 (2016).
21. P. Y. Ma, A. N. Tait, T. F. de Lima, C. Huang, B. J. Shastr1, and P. R. Prucnal, “Photonic independent component analysis using an on-chip microring weight bank,” Opt. Express 28, 1827–1844 (2020).
22. 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. reports 7, 1–10 (2017).

23. K. Padmaraju, J. Chan, L. Chen, M. Lipson, and K. Bergman, “Thermal stabilization of a microring modulator using feedback control,” Opt. express 20, 27999–28008 (2012).

24. K. Padmaraju, D. F. Logan, X. Zhu, J. J. Ackert, A. P. Knights, and K. Bergman, “Integrated thermal stabilization of a microring modulator,” Opt. express 21, 14342–14350 (2013).

25. K. Padmaraju, D. F. Logan, T. Shiraishi, J. J. Ackert, A. P. Knights, and K. Bergman, “Wavelength locking and thermally stabilizing microring resonators using dithering signals,” J. Light. Technol. 32, 505–512 (2013).

26. K. Padmaraju and K. Bergman, “Resolving the thermal challenges for silicon microring resonator devices,” Nanophotonics 3, 269–281 (2014).

27. D. Marpaung, J. Yao, and J. Capmany, “Integrated microwave photonics,” Nat. photonics 13, 80–90 (2019).

28. P. Y. Ma, A. N. Tait, W. Zhang, E. A. Karahan, T. F. de Lima, C. Huang, B. J. Shastr, and P. R. Prucnal, “Blind source separation with integrated photonics and reduced dimensional statistics,” Opt. Lett. 45, 6494–6497 (2020).

29. C. Huang, S. Fujisawa, T. F. De Lima, A. N. Tait, E. Blow, Y. Tian, S. Bilodeau, A. Jha, F. Yaman, H. G. Batshon et al., “Demonstration of photonic neural network for fiber nonlinearity compensation in long-haul transmission systems,” in 2020 Optical Fiber Communications Conference and Exhibition (OFC), (IEEE, 2020), pp. 1–3.

30. A. N. Tait, M. A. Nahmias, B. J. Shastr, and P. R. Prucnal, “Broadcast and weight: an integrated network for scalable photonic spike processing,” J. Light. Technol. 32, 4029–4041 (2014).

31. A. N. Tait, A. X. Wu, T. F. de Lima, E. Zhou, B. J. Shastr, M. A. Nahmias, and P. R. Prucnal, “Microring weight banks,” IEEE J. Sel. Top. Quantum Electron. 22, 312–325 (2016).

32. H. Jayatilleka, K. Murray, M. Caverley, N. A. Jaeger, L. Chrostowski, and S. Shekhar, “Crosstalk in soi microring resonator-based filters,” J. Light. Technol. 34, 2886–2896 (2016).

33. W. Bogaerts, P. De Heyn, T. Van Vaerenbergh, K. De Vos, S. Kumar Selvaraja, T. Claes, P. Dumon, P. Bienstman, D. Van Thourhout, and R. Baets, “Silicon microring resonators,” Laser & Photonics Rev. 6, 47–73 (2012).

34. A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintal, “Pytorch: An imperative style, high-performance deep learning library,” in NeurIPS, (2019).

35. H.-T. Peng, J. Lederman, L. Xu, T. F. de Lima, C. Huang, B. Shastri, D. Rosenbluth, and P. Prucnal, “A photonic-circuits-inspired compact network: Toward real-time wireless signal classification at the edge,” arXiv preprint arXiv:2106.13865 (2021).