

# Multi-Wavelength Photonic Neuromorphic Computing for Intra and Inter-Channel Distortion Compensations in WDM Optical Communication Systems

Benshan Wang<sup>1</sup>, Thomas Ferreira de Lima<sup>2</sup>, Bhavin J Shastri<sup>3</sup>, *Senior Member, IEEE*, Paul R Prucnal, *Life Fellow, IEEE*, and Chaoran Huang<sup>4</sup>, *Member, IEEE*

(Invited Paper)

**Abstract**—Digital signal processing (DSP) has been widely applied in optical communication systems to mitigate various signal distortions and has become one of the key technologies that have sustained data traffic growth over the past decade. However, the strict energy budget of application-specific integrated circuit-based DSP chips has prevented the deployment of some powerful but computationally costly DSP algorithms in real applications. As a result, fiber nonlinearity-induced signal distortions impede fiber communication systems, especially in wavelength-division multiplexed (WDM) transmission systems. To solve these challenges in DSP, there has been a surge of interest in implementing neural networks-based signal processing using photonics hardware (i.e., photonic neural networks). Photonic neural networks promise to break performance limitations in electronics and gain advantages in bandwidth, latency, and power consumption in solving intellectual tasks that are unreachable by conventional digital electronic platforms. This work proposes a photonic recurrent neural network (RNN) capable of simultaneously resolving dispersion and both intra and inter-channel fiber nonlinearities in multiple WDM channels in the photonic domain, for the first time to our best knowledge. Furthermore, our photonic RNN can directly process optical

WDM signals in the photonic domain, avoiding prohibitive energy consumption and speed overhead in analog to digital converters (ADCs). Our proposed photonic RNN is fully compatible with mature silicon photonic fabrications. We demonstrate in simulation that our photonic RNN can process multiple WDM channels simultaneously and achieve a reduced bit error rate compared to typical DSP algorithms for all WDM channels in a pulse-amplitude modulation 4-level (PAM4) transmission system, thanks to its unique capability to address inter-channel fiber nonlinearities. In addition to signal quality performance, the proposed system also promises to significantly reduce the power consumption and the latency compared to the state-of-the-art DSP chips, according to our power and latency analysis.

**Index Terms**—Digital signal processing, neuromorphic computing, nonlinear optics, optical fiber communication, photonic neural network, signal equalization, wavelength-division multiplexing.

## I. INTRODUCTION

GLOBAL internet traffic is growing exponentially, as driven by bandwidth-hungry applications, such as cloud computing, virtual/augmented reality, and high-definition video streaming [1]. To meet the demands for capacity growth, today's optical communication systems widely deploy densely spaced wavelength-division multiplexed (WDM) channels and high-order modulation formats [2], which inevitably leads to increased signal distortions. The signal distortions originate from the combined effects of chromatic dispersion (CD) and Kerr nonlinearity in optical fibers. In WDM systems, Kerr nonlinearity causes more severe impairments due to the cross-phase modulation (XPM) and four-wave mixing (FWM) among different wavelength channels.

Digital signal processing (DSP) technology is widely employed in today's transport systems to compensate for signal distortions. It has successfully supported fiber capacity growth over the past decade [3], [4], [5], [6]. However, in handling the data rates in state-of-the-art optical communication systems, the power dissipation of 7 nm complementary metal-oxide-semiconductor (CMOS) DSP chips for 800-Gigabit Ethernet has already approached the maximum thermal dissipation capacity

Manuscript received 3 July 2022; revised 13 September 2022; accepted 3 October 2022. Date of publication 10 October 2022; date of current version 24 October 2022. This work was supported in part by CUHK Direct under Grant 170257018, in part by the Shun Hing Institute of Advanced Engineering, CUHK, under Grant RNE-p4-22, and in part by CUHK startup fund. The devices were fabricated at the IME ASTAR foundry in Singapore. Fabrication was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), in part by Silicon Electronic-Photonic Integrated Circuits (SiEPIC) Program, and in part by the Canadian Microelectronics Corporation (CMC). (Corresponding author: Chaoran Huang.)

Benshan Wang and Chaoran Huang are with the Department of Electronic Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong (e-mail: bswang@link.cuhk.edu.hk; crhuang@ee.cuhk.edu.hk).

Thomas Ferreira de Lima was with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08540 USA. He is now with the NEC Laboratories America, Princeton, NJ 08540 USA (e-mail: research@tlima.me).

Bhavin J Shastri is with the Department of Physics, Engineering Physics & Astronomy, Queen's University, Kingston, ON K7L 3N6, Canada (e-mail: bhavin.shastri@queensu.ca).

Paul R Prucnal is with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08540 USA (e-mail: prucnal@princeton.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JSTQE.2022.3213172>.

Digital Object Identifier 10.1109/JSTQE.2022.3213172

of modern packaging technologies [7], [8]. The problem of surpassing the thermal capacity has necessitated a careful compromise in the complexity of DSP algorithms to ensure that power consumption is within an acceptable limit. As a result, some powerful but computationally intensive DSP algorithms have not been employed in practice. A prominent algorithm for fiber nonlinearity compensation is the digital backpropagation (DBP) algorithm [4], [9], [10], [11], [12], [13], which is capable of solving both intra and inter-channel distortions by inverting the coupled multi-channel nonlinear Schrödinger equation using the received WDM signals as the inputs to the equation. However, the consequence is that DBP must process several high-speed and large bandwidth WDM channels concurrently, resulting in prohibitively large bandwidth and data communication overhead while being too complicated to implement on an application-specific integrated circuits (ASIC) chip [14]. As a result, nonlinear distortions in fiber communication systems remain an impediment, particularly in WDM transmission systems.

More fundamentally, a critical aspect of increasing fiber capacity is that it relies on consistently reducing the power consumption per bit of the DSP chip as the data rate increases. A survey article by Frey et al. [15] compared the scaling of integrated circuit node size and DSP ASICs for optical communications. The result showed that DSP ASICs closely follow the current CMOS generations. This implies that increases in the fiber capacity over the past decade heavily rely on the node size scaling promised by Moore's Law. However, CMOS node scaling will inevitably slow down and finally halt [16], [17], suggesting that signal processing hardware needs to be redefined to maintain the explosive internet traffic growth in the future.

The field of neuromorphic photonics aims to build practical photonic neural networks (PNNs) to solve these challenges in DSP [18], [19], [20], [21], [22], [23], [24], [25], [26]. PNNs emulate the biological or artificial neural network models using high-speed devices and large-bandwidth photonic waveguides, thus allowing PNNs to execute neural network algorithms (e.g., deep learning) with unmatched speed. PNNs, like other machine learning algorithms, can learn optical fiber transmission channel characteristics from abundant transmission data without explicit channel knowledge and then invert various transmission impairments using the learned transmission channel model [26], [27]. However, unlike digital platforms such as ASICs, which heavily rely on parallel computing to keep up with the ever-increasing data rate (resulting in significant power consumption overhead), PNNs leverage photonic devices that are originally designed for optical communications, and thus, if appropriately designed, can always offer a processing speed matching the fiber communication line rate in the future.

Photonic reservoir computing (RC) [28], [29], [30], [31], [32], a subclass of recurrent neural networks (RNNs), was first proposed and demonstrated to process optical communication systems fast and efficiently [33], [34], [35], [36], [37], [38], [39]. RC consists a reservoir of randomly linked neurons followed by a readout layer. The weight connections are fixed inside the reservoir. Only the readout layer's weights are trained using linear regression. Photonic RC for optical communications has attracted significant research interest due to their ease of implementation. They have been implemented using both optical

fibers and integrated photonic platforms, and have demonstrated many functions, including dispersion and nonlinearity compensations for both IM/DD and coherent optical systems. The reader is directed to the most recent review articles in this field [40], [41]. Since only the readout layer of RC can be trained, RC faces the challenges of limited expressivity and thus sometimes cannot guarantee convergence to desired behaviors.

As proposed by some authors [42], a different method of overcoming RC's constraint is creating a rigorous mapping between physical RC's models of optoelectronic systems and abstract models of neural networks. Consequently, PNNs can then use standard machine learning methods (such as back-propagation) to train every parameter in the PNN. We used this approach to demonstrate a silicon photonic feed-forward neural network (FNN) to model and compensate for fiber nonlinearities in a 10,080 km trans-pacific transmission system [43]. The PNN functions as a photonic accelerator to replace the power-intensive nonlinear compensation module in DSP. However, the power analysis reveals that the digital to analog converters (DACs) used to convert digital signals to analog photonic neural network parameters contribute to most of the power consumption in the system, undermining the benefits of PNN in terms of energy efficiency [26].

To address these challenges, this paper proposes a WDM-based photonic recurrent neural network (depicted in Fig. 1(b)) [21] as a front-end processor, capable of simultaneously resolving dispersions and both intra and inter-channel fiber nonlinearities in multiple WDM channels in the photonic domain, for the first time to our best knowledge. Furthermore, our photonic RNN can directly interface and process optical WDM signals in the photonic domain, eliminating prohibitive energy consumption overhead and speed reduction in ADCs. The WDM-based photonic neural network encodes the information to different wavelengths as inputs, and thus it can directly interface with the WDM optical communication systems. The proposed photonic neural network composes microring resonator (MRR) banks for synaptic weighting, photodetector-modulator neurons as activation function, and a feedback waveguide providing all-to-all recurrency connecting all WDM channels. The proposed photonic RNN can be analogous to optical fiber transmission systems in that their linear neuron-to-neuron connections with the internal feedback waveguide are analogous to dispersive memory. Meanwhile, the neuron nonlinearity can mimic all forms of nonlinear effects (not limited to fiber nonlinearities) in a fiber transmission system [27]. Thanks to these analogies, the photonic RNN can be trained to resemble the fiber transmission channels and invert the various channel impairments. More interestingly, PNN provides wideband weighted addition among WDM channels, allowing addressing inter-channel nonlinearity problems in WDM systems that DSP built on ASIC chips lacks the bandwidth or power to process.

To evaluate performance of our proposed front-end neuromorphic processor, we apply the proposed photonic RNN to a WDM pulse-amplitude modulation 4-level (PAM4) system and evaluate the signals' bit error rate (BER) reduction after being processed by the photonic RNN in the numerically simulated transmission system. We compare the signal performances with various DSP approaches, demonstrating that our photonic RNN

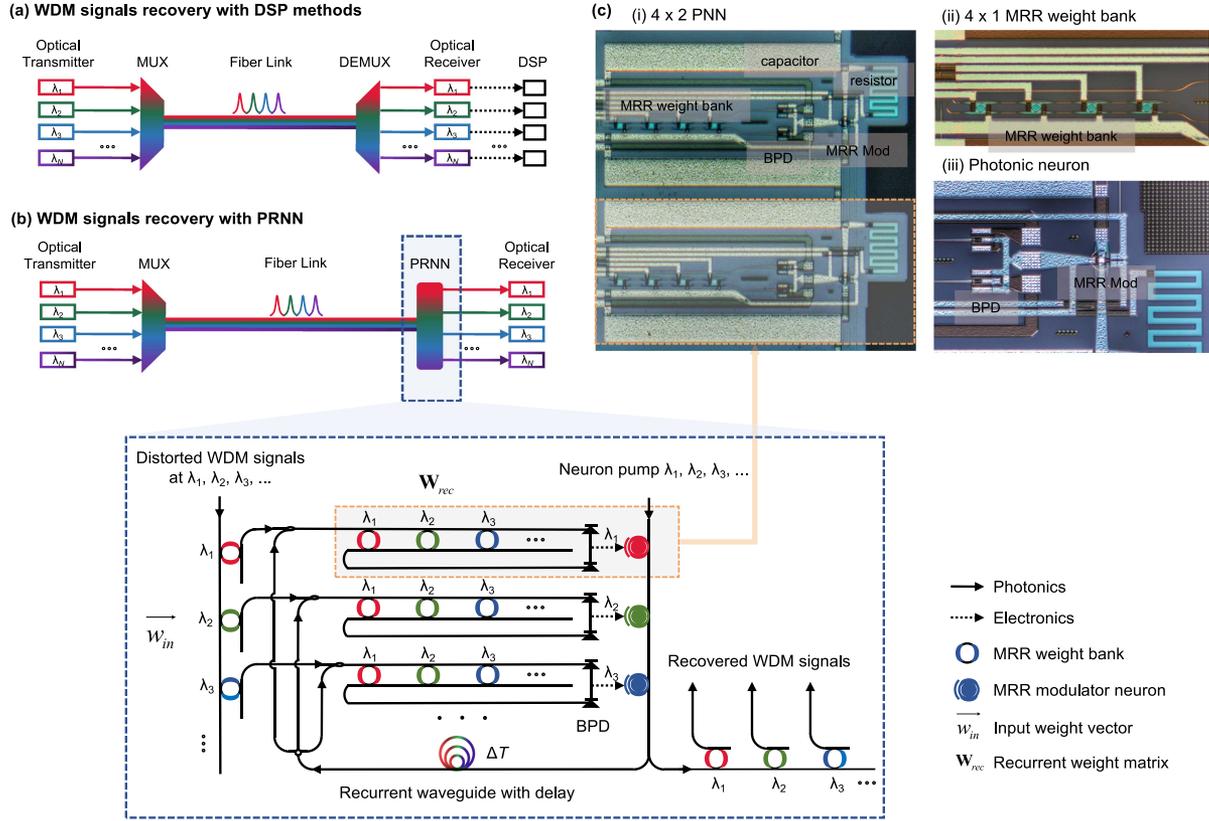


Fig. 1. (a) Conventional WDM transmission system with DSP. (b) Proposed WDM transmission system where signal recovery is performed in the optical domain using the proposed photonic recurrent neural network before the optical receiver. (c) A false-colour confocal micrograph of the silicon photonic neuron device. (c) is reproduced from Huang et al., *Nat. Electron.* **28**, 8, 2021, ref. [43], Springer Nature Ltd. WDM: wavelength-division multiplexing, DSP: digital signal processor, (DE)MUX: (de)multiplexing, PRNN: photonic recurrent neural network, MRR: microring resonator, BPD: balanced photodetector.

has smaller BERs for all wavelength channels. We verify that the performance improvement comes from the unique capability of compensating for inter-channel nonlinearities. Our power consumption and latency analysis show that our photonic RNN can significantly reduce the power consumption and latency compared to the state-of-the-art DSP chips. We hope our proposed photonic RNN, which is fully compatible with commercial silicon photonic platforms, can provide and inspire new solutions to the future high-speed low-energy fiber communication systems.

## II. PRINCIPLES AND SYSTEM DESCRIPTION

Fig. 1(a) shows a typical  $N$ -channel WDM fiber transmission system. At the transmitter, independent information streams are modulated on different wavelengths of lasers

and then multiplexed and transmitted simultaneously in a single-mode fiber. The signals experience chromatic dispersion and Kerr effect induced nonlinear distortions in the fiber link as described by the coupled nonlinear Schrödinger equation [44] in (1) shown at the bottom of this page, where  $u_{\lambda_i}(t, z)$  is the optical field of the signal at wavelength  $\lambda_i$ ,  $\beta_2$  is the group-velocity dispersion governing linear impairment, and  $\gamma$  is the fiber nonlinear coefficient.  $\Delta k_{mnlk} = k_m + k_n - k_l - k_k$  represents the phase mismatch.  $f_R$  represents the fractional Raman contribution,  $h_R(t)$  is Raman response function for stimulated Raman scattering,  $\Omega$  is the Stokes shift [44]. The first term on the right hand side of the NLSE is self-phase modulation which introduces a nonlinear phase related to the signal power, and the second term is cross-phase modulation which describes how the signal in one wavelength is

$$\begin{aligned}
 \frac{\partial u_{\lambda_k}}{\partial z} + i \frac{\beta_2}{2} \frac{\partial^2 u_{\lambda_k}}{\partial t^2} = & \underbrace{i\gamma |u_{\lambda_k}|^2 u_{\lambda_k}}_{\text{Self-phase modulation}} + \underbrace{i2\gamma \sum_{i \neq k}^N |u_{\lambda_i}|^2 u_{\lambda_k}}_{\text{Cross-phase modulation}} + \underbrace{i2\gamma \sum_{k=m+n-l}^N u_{\lambda_m} u_{\lambda_n} u_{\lambda_l}^* \exp(i\Delta k_{mnlk}z)}_{\text{Four-wave mixing}} \\
 & + \underbrace{i\gamma f_R u_{\lambda_k} \sum_{i \neq k}^N \int_{-\infty}^t h_R(t-t') \left[ |u_{\lambda_k}(z, t')|^2 + |u_{\lambda_i}(z, t')|^2 \right] + u_{\lambda_k}(z, t') u_{\lambda_i}^*(z, t') \exp[\pm i\Omega(t-t')]}_{\text{Stimulated Raman scattering}} dt' \quad (1)
 \end{aligned}$$

affected by the nonlinear phase introduced by other wavelength signals. The third term is four-wave mixing which involves nonlinear interaction among four optical waves, and the fourth term represents stimulated Raman scattering where one signal acts as a pump and generates the frequency-shift radiation to another signal. The inter-channel stimulated Raman scattering is significant only if in ultra-wideband WDM transmission [45]. The XPM and FWM terms indicate that fully compensating for the inter-channel nonlinearity requires the receiver to be able to detect and process multiple wavelength channels simultaneously. At the receiver of a conventional WDM fiber transmission system shown in Fig. 1(a), the WDM signals are first demultiplexed and then detected by the photodetectors and processed by DSPs separately. In principle, the fiber-induced distortions can be compensated by DBP, which solves the  $z$ -reversed NLSE digitally, leaving only non-deterministic noises [4]. However, full compensation of inter-channel effects using DBP is too computationally expensive to implement in DSP because it requires first high oversampling of the signal and then solving a set of coupled NLSE with a fine step size [9]. As a result, inter-channel nonlinearity remains one of the major limitations to increasing fiber capacity.

Our proposed photonic RNN solves this problem from a new perspective as shown in Fig. 1(b). The photonic RNN's input is an array of MRRs, each with a resonance frequency corresponding to the central frequency of the signals in the WDM transmission system. The input MRR array serves as the input layer of the photonic RNN, accepting the WDM signals from the transmission link and providing a tunable weight vector  $w_{in}$  to the WDM signals at the optical domain. After the input layer, the WDM signals are broadcast to and weighted by the reconfigurable, continuous-valued MRR weight banks [47] consisting of  $N \times N$  MRRs, where  $N$  is the number of WDM channels to be processed. Next, the WDM signals are summed by the followed balanced photodetector which converts the total optical power from each MRR output into photocurrent, resulting in a complementary  $-1$  to  $+1$  continuous weight range. The weight matrix given by the  $N \times N$  weight bank is denoted as  $\mathbf{W}_{rec}$ . The photocurrent modulates the transmission of the MRR modulator (that is, photonic neuron) and hence modulates the optical power of a continuous-wave optical carrier (labeled as a "neuron pump"). The MRR modulator exhibits nonlinear electrical-to-optical transfer functions, producing the nonlinear activation function in the photonic neural network. The outputs of the photonic neurons are partially connected back to the  $N \times N$  MRRs' inputs using a single feedback waveguide.

The "Broadcast-and-weight" protocol [18] is compatible with the mainstream silicon photonics platform. Fig. 1(c) shows the micrograph of a  $4 \times 2$  photonic neural network device fabricated by the commercial silicon photonic foundry as an example, which comprises two arrays of MRR weight banks connected to two photonic neurons (i.e., germanium-silicon photodetector–MRR PN-junction modulator pair). The experimental demonstration of MRR weight banks was first published by Tait et al. [18], followed by several efforts to improve the weighting precision in the network [48], [49], [50]. The most recent work demonstrated 9-bit precision, beyond the 8-bit precision widely used in DSP chips for optical communications [51]. The first full

integration of the photonic neural network comprising MRR weight bank and photonic neuron was demonstrated in [52]. Current silicon photonic foundry chips lack on-chip lasers and optical amplification to support neuron-to-neuron cascability. The current demonstration uses externally coupled lasers as neuron pumps and passive on-chip front-end impedance connecting the photodetector and MRR modulator to provide neuron-neuron gain and compensate for the circuit losses, at the cost of speed reduction and power consumption increase caused by the on-chip impedance [43], [53], [54]. Other advanced but less mature technologies promise to solve the tradeoffs. As discussed in [55], integrating a transimpedance amplifier with a silicon photonic chip can provide a significant electrical gain to retain cascability without losing speed. Meanwhile, III/V and heterogeneous III/V-silicon integration platforms promise on-chip lasers and optical amplification [46], [56], [57], [58], [59]. The processing latency of this network is determined by the light traveling time in the recurrent loop, which includes the delays at the MRR weight banks, the PD, and the MRR modulator neuron caused by their limited bandwidth and the delay in the waveguide. By using high-speed optoelectronic devices, the processing latency of the photonic RNN is usually smaller than a clock cycle of a DSP chip [60]. In addition, photonic RNNs don't always suffer from the constant data movement, memory fetch and clock distribution operations required in digital RNNs. As a result, the photonic RNN has much lower latency than DSP.

### III. TRAINING PHOTONIC RNN

#### A. Principle

The principle of training the photonic RNN is to digitize the continuous photonic RNN model to conventional RNN model, and then the photonic RNN can be trained using the standard deep learning algorithm, as shown in Fig. 2. The photonic RNN can be modelled by a set of coupled ordinary differential equations modified from [21], [46]:

$$\frac{ds(t)}{dt} = \mathbf{W}_{rec}\mathbf{y}(t - \Delta T) + w_{in}\mathbf{u}(t) + \mathbf{b} - \frac{\mathbf{s}(t)}{\tau} \quad (2a)$$

$$\mathbf{y}(t) = \sigma(\mathbf{s}(t)) \quad (2b)$$

where  $\mathbf{u}$  and  $\mathbf{y}$  are the input and output WDM signals of the photonic RNN. In our case, both  $\mathbf{u}$  and  $\mathbf{y}$  have  $N$  elements, each representing the time series of the distorted and recovered signal, respectively, at a particular wavelength.  $N$  is the channel number of the WDM transmission system.  $\mathbf{s}$  and  $\mathbf{b}$  are vectors, representing the neuron state and bias respectively,  $\tau$  is the time constant of the photonic neuron.  $\Delta T$  is the time delay caused by the feedback optical waveguide. In this paper, we choose  $\Delta T = 1/B_{data}$ , where  $B_{data}$  is the signal's baud rate, to ensure that the memory of the photonic RNN is approximate to the ratio between dispersion-induced pulse spreading and the original pulse width.  $\sigma(\cdot)$  is the nonlinear activation function of the photonic MRR neuron, which is typically a Lorenz-shape function [52]. Here we use  $\sigma(x) = -0.28/((x - 0.5)^2 + 0.49^2) + 1.35$  experimentally characterized from our silicon photonic neuron

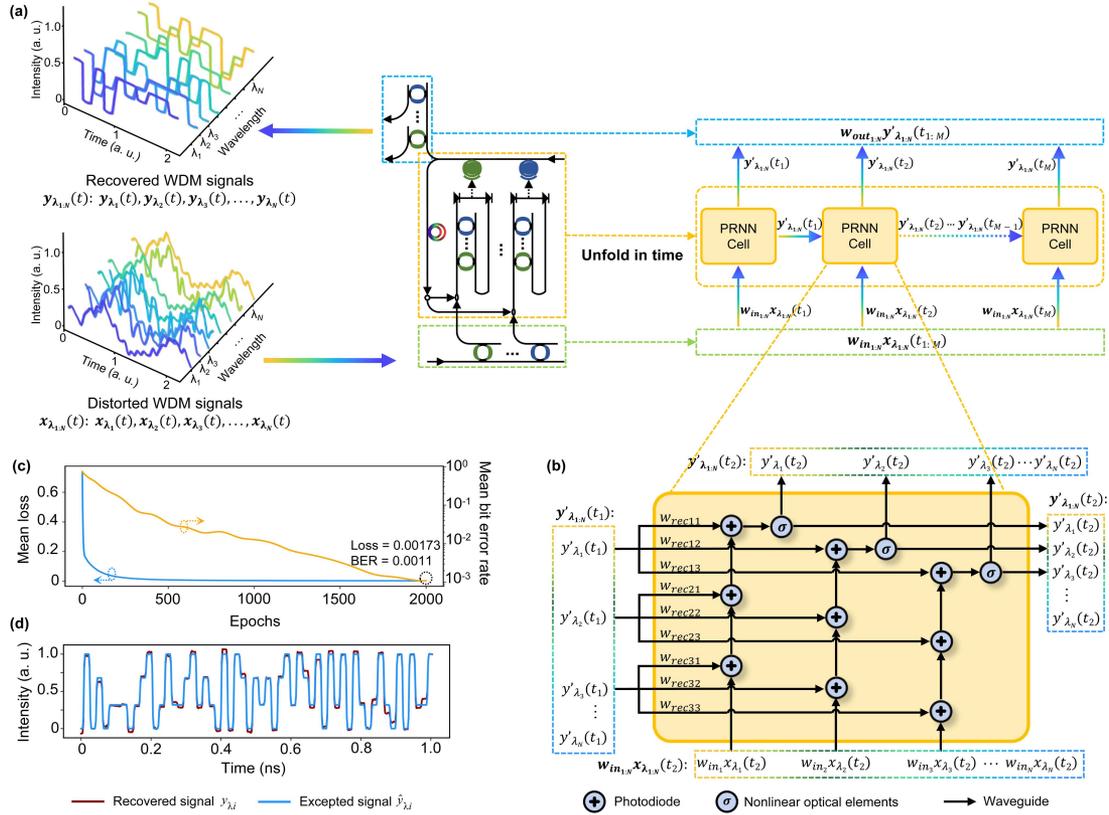


Fig. 2. Framework and training of the photonic RNN model. (a) Photonic RNN for  $N$ -channel ( $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_N$ ) WDM PAM4 optical signals equalization. The distorted WDM signals  $x_{\lambda_{1:N}}$  are regarded as inputs of photonic RNN and recovered to original signals  $y_{\lambda_{1:N}}$ . Every channel has the same signal sequence length of  $M$ . The signals are inputted and outputted via the MRR arrays,  $w_{ini}$  and  $w_{out}$  are the tunable weight vectors for different WDM signals  $\lambda_i$  in the input and output MRR array, respectively. Reproduce from Peng et al., *arXiv:2106.13865* (2021) Ref. [46]. (b) Schematic diagram of the discrete photonic RNN cell.  $\sigma$ : activation function. (c) The mean training loss and mean BER for 2-channel  $\times$  56 GBaud/s/lane 20 km transmission system. (d) Comparison of the expected signal (blue) and the recovery output (red) of the photonic RNN model.

device [43]. Other activation functions can be assembled using coupled-cavity devices [61], [62], [63].  $W_{rec}$  is the photonic recurrent weight matrix, and  $w_{in}$  is the input weight vector.

The task of training the photonic RNN is to learn the optimized  $W_{rec}$ ,  $w_{in}$  and  $b$  such that the outputs  $y$  are the recovered WDM signals. To train the analog photonic RNN, we first discretize (2) using the forward-Euler method, as proposed by [46]. The discretized photonic RNN model now can be represented by a series of RNN cells similar to software-based RNNs as shown in Fig. 2(a), consisting to a series of photonic RNN cells illustrated in Fig. 2(b). Fig. 2(b) shows a  $3 \times 3$  recurrent structure as an example.  $N \times N$  photonic RNN can be achieved using the same architecture with an  $N \times N$  network. In this way, the photonic RNN can be constructed by Pytorch nn.module [64] and trained using the back propagation through time (BPTT) algorithm [65], similar to software-based RNNs.

### B. Training Photonic RNN for WDM Transmission System

We use VPI Transmission Maker [66] to simulate the WDM transmission system and generate the dataset. Here we investigate a WDM PAM4 direct-detection system for short-reach transmission. In our simulation, we assume the standard

single-mode fiber (SSMF) with a dispersion parameter  $D = 16$  ps/nm/km, a dispersion slope of  $D_s = 0.08$  ps/nm<sup>2</sup>/km, a nonlinear index  $= 2.6 \times 10^{-20}$  m<sup>2</sup>/W, an attenuation loss = 0.2 dB/km. The fiber loss is compensated by a pre-amplifier with a noise figure of 4 dB. We first look at a 2-channel 56 GBaud transmission system with the signal carrier frequency at 193.1 and 193.2 THz respectively, over a 20 km transmission distance. We generate 102,400 symbols for each channel at the transmitter using pseudorandom bit sequences (PRBSs) with different seeds and obtain 102,400 received symbols from the VPI simulation. 80% of the symbols are used for training and validation, and 20% are used as the testing dataset. For  $N$ -channel WDM system with training sequence length of  $M$  per channel, the training objective is to minimize the average mean square error (MSE) loss between the obtained and the expected output of the training input sequences, defined as  $\frac{1}{N} \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^M [y_{\lambda_i}(t_j) - \hat{y}_{\lambda_i}(t_j)]^2$ . We use Kaiming initialization [67] to initialize the parameters of the PRNN layer and Xavier initialization [68] for output linear regression layer. Adam optimizer is set with an initial learning rate of 0.001, which decays every 1,000 iterations by 0.5 [69]. To validate that our network is reliable, we use Monte-Carlo cross-validation with 5 iterations [70]. With this method, we obtain 5 bit error rate values by counting the number of bit errors in total in each iteration. We then obtain the average BER

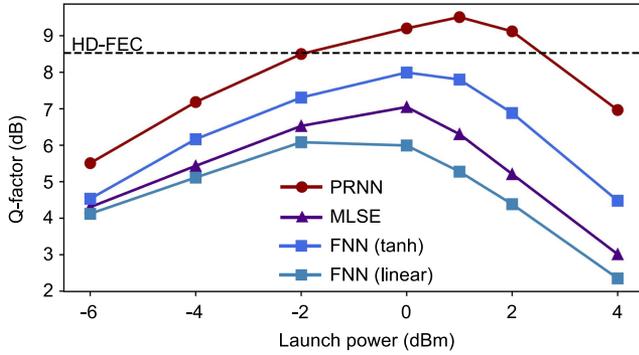


Fig. 3. Comparison of PRNN and various DSP methods in 2-channel 56 GBaud/λ WDM PAM4 signals over 20 km transmission link. PRNN: photonic recurrent neural network; FNN: feed-forward neural network; MLSE: maximum likelihood sequence estimation. tanh: hyperbolic tangent activation function.

to evaluate the equalization performance. Fig. 2(c) shows the mean loss (i.e., MSE) and BER during the training stage. The results are evaluated using the mean BER of the test dataset in different channels.

#### IV. RESULTS

##### A. Equalization Performance

After training, we test the photonic RNN's ability to resolve fiber-induced distortions, particularly fiber nonlinearity, in a 2-channel 56 GBaud/channel transmission link over 20 km distance. We compare the photonic RNN with different DSP methods including, maximum likelihood sequence estimation (MLSE) and feed-forward neural network equalizer (FNN), as shown in Fig. 3. MLSE is a linear equalizer commonly used for dispersion compensation [71]. The MLSE used in the comparison is operated with two samples per symbol and a memory length of 8 symbols according to the ratio between dispersion-induced pulse spreading and the original pulse width. The coefficients of MLSE are updated to minimize the MSE. The FNN design is similar to that published in [72], having an input layer accepting 8 time-delayed symbols, one hidden layer with 16 neurons, and an output layer with 2 neurons outputting the recovered signals. The input size is approximate to the dispersion-induced pulse spreading divided by the original pulse width. To study the performance of linear and nonlinear equalizers, we include two FNNs in our comparison, one with the hyperbolic tangent (*tanh*) activation function and the other without activation functions (i.e., a linear network). It is worth noting that the DSP algorithms can only be implemented on a channel-by-channel basis, while the photonic RNN is able to process all WDM channels simultaneously.

Fig. 3 shows the results of system Q-factor performance as a function of optical powers launched into the fiber input under different compensation schemes. The Q-factor value is calculated by [73]:

$$Q = 20 \cdot \log_{10}(\sqrt{2} \operatorname{erfc}^{-1}(2\operatorname{BER})) \quad (3)$$

where BER is the average BER over all WDM channels. As shown in Fig. 3, we see that the photonic RNN significantly outperforms all other DSP-based compensation algorithms due

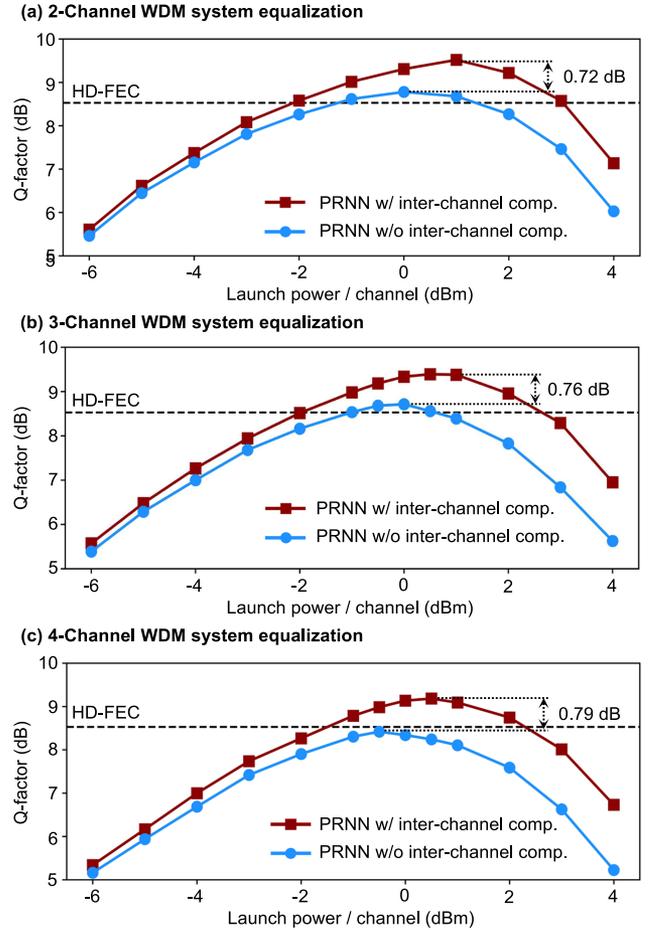


Fig. 4. Equalization performance for (a) 2, (b) 3, and (c) 4 channels WDM systems respectively. The WDM systems have 56 GBaud/λ PAM4 signals up to 20 km transmission with different launch power per channel. The red and blue curves represent the average equalization performance of all channels using PRNN with and without inter-channel nonlinearity compensation, respectively. PRNN: photonic recurrent neural network.

to its capability of processing both intra and inter-channel distortions. The optimum launched power is 1 dBm with a BER of  $1.4 \times 10^{-3}$  (corresponding to 9.51 dB Q-factor). FNN with *tanh* activation functions is only capable of mitigating intra-channel distortions including dispersion and self-phase modulation, thus the optimum BER after compensation is  $6.2 \times 10^{-3}$  (7.96 dB Q-factor) at a launch power of 0 dBm, still worse than the hard-decision forward error correction (HD-FEC) threshold ( $3.8 \times 10^{-3}$ ). The FNN without nonlinear activation functions and MLSE are linear equalizers, so they can only compensate for fiber dispersions and have the worst BER performances.

To further confirm that the photonic RNN can compensate for inter-channel nonlinearity, we conduct the following simulations based on the same photonic RNN. In the following two cases, the transmission link contains  $2 \times 56$  GBaud PAM4 signals. In one case shown in the blue curve in Fig. 4(a), we train the neural network using the training dataset of channel 1 and evaluate the BER performance of channel 1. Due to lacking the data from the second channel, this case can only address intra-channel distortions similar to the FNN. In the other case, the photonic RNN is trained using the training datasets from both channels, as

shown by the red curve in Fig. 4(a). In this case, since the neural network can learn the response of all WDM channels rather than only a single channel, the inter-channel crosstalk can be compensated. As a result, we observe a 0.72 dB improvement in the Q-factor and a 1 dB improvement in optimum launch power than those in the first case.

### B. Channel Scalability

We investigate the performance of the proposed photonic RNN with more WDM channels. The signal speed is 56 GBaud/s/channel and the transmission distance is 20 km. The channel space is 100 GHz. The results of 3 and 4 channels is shown in Fig. 4(b), (c). Similar to Fig. 4(a), the red curve shows the results of training all channels simultaneously to compensate for inter-channel crosstalk, whereas the blue curve shows the results of training only one channel. In the 3-channel system, the optimal launch power increases from 0 dBm to 1 dBm, and the optimal Q-factor increase from 8.63 dB to 9.39 dB. In the 4-channel system, the optimal launch power increases from -0.5 dBm to 0.5 dBm, and the optimal Q-factor increase from 8.38 dB to 9.17 dB.

We find it becomes more challenging to train more channels because the current objective function is difficult to render optimized BER performances for every channel simultaneously. The WDM signal in one channel is more likely to be influenced by the signals from its adjacent channels than by the channels farther away. Thus in principle, the weights between neighboring channels should be more significant than distant channels. However, the current training objective is the MSE averaged across all channels. The consequence of this non-optimized objective function is the trained weights between distant channels are larger than they should be. A potential solution to this problem is to assign a dedicated objective function to every individual channel [74].

### C. Data Rate-Distance Product Comparison With Other Published Work

To further evaluate the capability of our proposed photonic RNN, we simulate the transmission performance under different single-lane data rates to find out the maximum transmission distance at  $\text{BER} = 3.8 \times 10^{-3}$ . Fig. 5 shows the data rate-distance product and the comparison between our work (red dots) and other published transmission system demonstrations with PAM4 signals. We have included both DSP (green dots) and photonic reservoir computing (blue dots) in the comparison. The gray curves are the contour lines given by the product of the data rate and the distance. When only considering a single channel, our work can achieve a data rate-distance product of 1200 GBaud · km, outperforming most of the other approaches except [81]. The algorithm used in [81] requires a feedforward equalizer and MLSE. Its problem is that in the system with more distinct dispersion-induced pulse broadening, the computational complexity increases enormously. In contrast, our approach only requires a very simple photonic RNN with one photonic neuron in the single-channel transmission system. When there are two channels in the transmission system, we see that our works' data rate-distance products (red squares) are approximately doubled

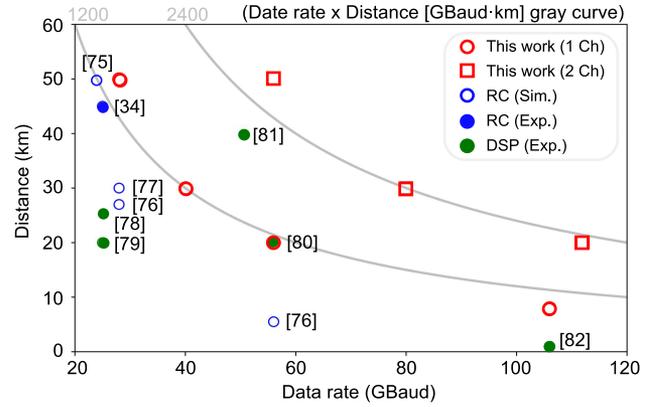


Fig. 5. Comparison of different PAM4 equalization systems with reservoir computing (RC) [34], [75], [76], [77] and DSP [78], [79], [80], [81], [82] in terms of data rate and distance. Ch: channel; Sim: simulation; Exp: experiment.

to 2400 GBaud · km, implying that our approach can extend the transmission performance by compensating for the inter-channel crosstalk in multi-channel WDM system. Other works under comparison only consider single transmission channel.

## V. POWER CONSUMPTION AND LATENCY ANALYSIS

This section will show that photonic RNNs not only exceed DSP in channel equalization performance but they also have the potential to drastically reduce power consumption and latency in optical communication systems. When estimating the power consumption and latency of PRNN, we assume that the system uses devices accessible in commercial silicon photonics foundries. For the DSP, we assume that it uses 5 nm CMOS technology. We realize the challenges of quantifying the power consumption and latency of DSP at the chip level. Thus, we only focus on the basic multiplication and addition operations when analyzing the DSP. When analyzing the photonic RNN, we have included fundamental sources of on-chip energy dissipation that are needed for a photonic neural network to operate (such as weight banks, photonic neurons, pump lasers, as well as most of the peripheral circuits (DACs)).

### A. Power Consumption

1) *Photonic RNN With Silicon Photonic Platform:* We follow the outline of power consumption of photonic neural network derived in previous publications [43], [52], [60], [83]. The power consumption of a  $N \times N$  photonic RNN is given as:

$$\begin{aligned}
 P_{\text{Photonics}} &= \underbrace{[(E_{dac} + E_{mrr}) \cdot B_{data} + P_{driver} + P_{heater}] \cdot N^2}_{\text{Weighting power}} \\
 &+ \underbrace{[(E_{rec} + E_{mod}) \cdot B_{data} + P_{pump} + P_{heater}] \cdot N}_{\text{Neurons power}} \quad (4)
 \end{aligned}$$

The first term in (4) is the power consumption cost by photonic weights, where  $B_{data}$  is the baud rate,  $E_{dac}$  and  $E_{mrr}$  are the energy consumption spent on digital-to-analog converters and optical modulators respectively, to actuate weights. High-speed DACs based on 7-nm FinFET [84] are used to convert the digital

weights to analog voltages.  $P_{driver}$  is the power consumption of the MRR driver. Microheaters are used to tune the MRR resonance to match the pump wavelength. Thus,  $P_{heater}$  is the power consumption of the microheater and its control circuit for resonance alignment. The second term is the power consumption of photonic neurons. In the physical RNN, the signal power ratio between the input and output in the recurrent loop must be greater than unity [52]. For photonic neurons, physical cascability is expressed as the input and output being both optical, and their modulation depth and optical power remain unity. Many factors account for losses inside the closed loop, including (1) optical losses of optical devices and waveguides, which can be compensated by an external and powerful enough optical pump, and (2) electrical losses due to inefficient modulation depth. In our analysis, we assume a receiver with photodetector following transimpedance amplifier (TIA) [85]. The electrical loss is assumed to be compensated by a TIA between the O/E and E/O stages of photonic neurons. The TIA amplifies the photocurrent to drive the following modulator in full swing while sidestepping the gain-bandwidth trade-off in the neuron unit.  $E_{rec}$  and  $E_{mod}$  represent the energy consumption of electrical receiver (including the photodetector and TIA) and the MRR modulator neuron, respectively. The optical pump not only serves as an optical carrier of neuron output, but it also injects extra optical energy into the neuron to compensate for optical losses due to insertion losses and limited extinction ratio. The network needs to ensure greater-than-unity O/E/O efficiency. Thus, the neuron pump must provide sufficient photocurrent (after detection) to drive the neuron modulator in full swing. It also needs to compensate for various losses in the chip, such as coupling loss and device insertion losses. The  $P_{pump}$  used in our calculation is modified from [52] by further considering the losses in the peripheral circuits. The value of power consumption would depend on different technology developments [43]. Similar to [43], [83], in this paper, we assume that the system uses the silicon PN modulators and Ge-on-Si photodetectors available in commercial silicon photonics foundries. Table I in Appendix A summarizes the parameter values used to calculate the power consumption of the photonic neural network.

2) *DSP*: We follow the paper by Pillai et al. [86] to calculate the power consumption of DSP. Generally, the power consumption in DSP is the number of channel  $N$  times the sum of the energy of the required operations  $E_{op}$  multiplied by how many operations are required in a second  $N_{op}$ , as shown in (5).

$$P_{DSP} = N \cdot \sum E_{op} N_{op} \quad (5)$$

For simplicity, we exclude the power consumption of serializer/deserializer, which although is a highly challenging and power consuming module in DSP, particularly for high speed signals. We look at two algorithms, FNN and MLSE, and focus on their number of operations  $N_{op}$ , especially multiplication  $N_M$  and addition  $N_A$  which account for most power consumption.

We first consider the  $I \times H \times O$  3-layer delayed FNN system for single channel compensation, where  $I$ ,  $H$ ,  $O$  represent the number of neurons in each layer. Assuming that the FNN processes data sequentially, the number of operations completely

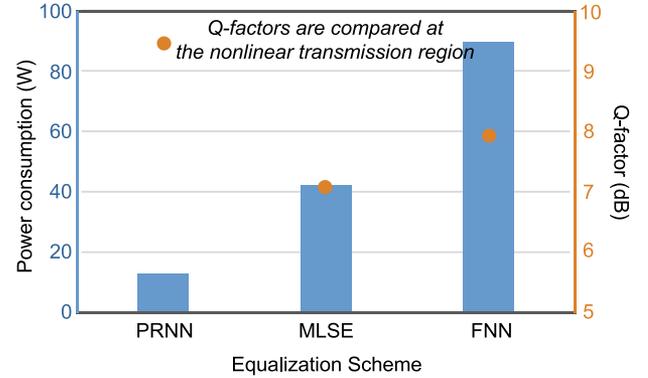


Fig. 6. Power consumption of 1.6 T transceiver module and Q-factor of 2-channel 56 GBaud/λ WDM PAM4 signals over a 20 km transmission link with different equalization schemes. PRNN: photonic recurrent neural network; MLSE: maximum likelihood sequence estimation; FNN: feed-forward neural network. In calculating the power consumption of PRNN, we assume that the system uses the silicon PN modulators and Ge-on-Si photodetectors available in commercial silicon photonics foundries. MLSE and FNN are DSP approach using 5 nm CMOS node.

depends on the network size and channel number. The number of multiplication  $N_{M,fnn}$  and addition  $N_{A,fnn}$  operations is as follows:

$$N_{M,fnn} = (I \cdot H + H \cdot O) \cdot B_{data} \quad (6a)$$

$$N_{A,fnn} = ((I - 1) \cdot H + (H - 1) \cdot O) \cdot B_{data} \quad (6b)$$

We then adapt MLSE structure from [81] to derive the operations number of multiplication  $N_{M,mlse}$  and addition  $N_{A,mlse}$ . In order to search for the most probable signal sequence using MLSE, the memory length must scale with the dispersion-induced pulse spreading length which leads to the exponentially increased operation number [81], as shown in follows:

$$N_{M,mlse} = L \cdot P^{L+1} \cdot B_{data} \quad (7a)$$

$$N_{A,mlse} = (L + 2) \cdot P^{L+1} \cdot B_{data} \quad (7b)$$

where  $L = 4$  is the truncated channel length and  $P = 2$  is the number of symbol to construct a reduced-state trellis [81].

Fig. 6 plots the predicted power consumption of the next generation 1.6 T transceiver for data center communications with 8 WDM channels  $\times$  200 Gbit/s/lane.  $N = 8$  in (5). We assume 5 nm CMOS node is used for DSP chips. The operation energy consumption of multiplication  $E_{op,M}$  and addition  $E_{op,A}$  per information bit are 347.5 fJ/b and 43.4 fJ/b, respectively, according to [86]. We can see from Fig. 6 that the PRNN power consumption is predicted to be 12.7 W which is lower than other DSP algorithms while achieving the best signal quality. In comparison, the estimated power consumption of MLSE is 42.3 W, and the FNN-based equalizer has a power consumption of over 90.0 W, both are beyond the thermal management capacity of today's semiconductor packaging technology [7]. The parameters used to generate Fig. 6 are summarized in Table I of Appendix A.

## B. Latency

The processing latency of this network is determined by the light traveling time in the recurrent loop, which includes the

delays at the MRR weight bank, PDs, and MRR modulator neurons caused by their limited bandwidth and the delay in the waveguide. The MRRs in the weight bank have a low Q-factor of only a few thousand, and thus the weights are computed in a single time step within a few picoseconds. The delay in the photodetector and MRR modulator neuron is approximately equal to the inverse of their total bandwidth (i.e.,  $1/B_{mod} + 1/B_{pd}$ ). The delay in the recurrent waveguide equals one symbol duration (i.e.,  $1/B_{data}$ ). To compensate for the dispersion, the signals need to recirculate in the loop for at least  $M$  times.  $M$  is approximately the ratio between dispersion-induced pulse spreading and the original pulse width. Therefore, the total latency of PRNN is  $M(1/B_{mod} + 1/B_{pd}) + (M - 1)(1/B_{data})$  [60].  $M = 8$  in the 1.6 T transmission system with 8 PAM4 WDM channels and 100 Gbaud/lane. Therefore, the total latency of a photonic RNN is approximately 470 ps.

A significant difference between DSP and photonic systems is that the latency of DSP systems is fundamentally bounded by the system clock  $f_{clock}$ . The DSP clock  $f_{clock}$  is usually set at a few hundreds of MHz in most DSP systems for the best energy efficiency [86]. It is challenging to estimate the latency in DSP because the latency is not only determined by the required operations but also the exact circuit layout. Thus for simplicity, we consider the best scenario for DSP where DSP hardware can always operate at line rate speed via massive parallelism, so only the delay in serial to parallel (s/p) conversion is included. The latency of s/p conversion is  $2B_{data}/f_{clock}^2$  [83]. With the parameters values in Table I, we can obtain the minimum latency of DSP processor's latency of 1.6  $\mu$ s, which is more than 3,400 times longer than 470 ps in photonic RNN.

## VI. CONCLUSION

In this paper, we present a photonic recurrent neural network (RNN) compatible with the silicon photonic platform for intra and inter-channel impairments compensation in WDM systems. We demonstrate in simulation that our photonic RNN equalizer can outperform digital MLSE and FNN equalizer in a 56 Gbaud WDM PAM4 transmission system in terms of optimum Q-factor and launch power. A higher baud rate can be achieved by replacing the optical modulators and photodetectors in the photonic RNN with those with matched speeds. Due to inter-channel nonlinearity compensation, our photonic RNN improves the signals' Q-factor by 0.72, 0.76, and 0.79 dB in 2, 3, and 4 channels WDM systems, respectively, compared to other schemes without inter-channel compensation capability. Furthermore, we show that our photonic RNN has a better data rate-distance product than most other channel equalization using photonic reservoir computing and DSP, except for some computationally expensive DSP algorithms. In addition to the advantages in BER performance, our photonic RNN has a unique capability of processing WDM communication signals of multiple channels directly in the optical domain without ADCs. This capability ensures the photonic RNN has much greater energy efficiency compared to other photonic neural network and DSP approaches. Our proposed photonic RNN has shown significantly reduced power consumption and processing latency compared to the

conventional DSP hardware when using the mature silicon photonic foundry technologies. Prior research has suggested that the power consumption can be further reduced by two orders of magnitude with advanced nanophotonics [60].

## APPENDIX A CALCULATION PARAMETERS VALUES

Related values of involved parameters for calculation are listed in Table 1.

TABLE I  
MODELING PARAMETERS

Symbol	Quantity	Value
$N$	Signal channel number	8
$B_{data}$	Signal baud rate	100 GBaud/s
$E_{dac}$	Energy consumption of 7-nm FinFET DAC	1.05 pJ/b [85]
$E_{mrr}, E_{mod}$	Energy consumed to change the voltage on the modulator	$C_{mod}V_r^2/4$
$C_{mod}$	Modulator capacitance	30 fF
$V_r$	Modulator voltage swing	4.8 V
$P_{driver}$	Power consumed to drive modulator	61 mW [88]
$P_{heater}$	The power consumption of heater and control circuits for each MRR to resonance alignment	4.6 mW [89]
$E_{rec}$	Optical receiver with TIA	550 fJ/b [86]
$P_{pump}$	Pump laser power	$2V_r/(\pi\eta_{wall-plug}\alpha\beta R_{pd}R_{load})$ [53] + $P_{tec}$
$\eta_{wall-plug}$	Wall-plug efficiency of pump lasers	0.3
$\alpha$	Coupling efficiency between laser and photonic chip	0.8 [90]
$\beta$	WDM multiplexer efficiency	0.83 [91]
$R_{pd}$	Photodiode responsivity	1 A/W
$R_{load}$	Resistance of the receiver junction	3 k $\Omega$
$P_{tec}$	TEC power for each laser	1.3 mW [92]
$E_{op,A}$	Average energy per $n_b$ -bit adder op.	$2.57n_b p_t V^2$ fJ/b
$E_{op,M}$	Average energy per $n_b$ -bit multiplier op.	$2.57n_b^2 p_t V^2$ fJ/b
$n_b$	Average bit resolution of the DSP module	8
$p_t$	CMOS process technology feature size	5 nm
$V$	CMOS supply voltage	0.65 V
$I, H, O$	FNN size of different layers	16, 16, 2
$N_{fnn,M}$	The number of multiplication in FNN	$288 \cdot B_{data}$
$N_{fnn,A}$	The number of addition in FNN	$270 \cdot B_{data}$
$N_{mlse,M}$	The number of multiplication in MLSE	$L P^{L+1} B_{data}$
$N_{mlse,A}$	The number of addition in MLSE	$(L + 2) P^{L+1} B_{data}$
$L$	The truncated channel of length	4 [82]
$P$	The number of symbol to construct a reduced-state trellis	2 [82]
$B_{mod}$	Modulator bandwidth	40 GHz
$B_{pd}$	Photodiode bandwidth	40 GHz
$f_{clock}$	Clock frequency of DSP	500 MHz

## REFERENCES

- [1] U. Cisco, "Cisco annual internet report (2018–2023) white paper. 2020," *Accessado em*, vol. 10, no. 01, 2021.
- [2] P. J. Winzer and R.-J. Essiambre, "Advanced optical modulation formats," in *Optical Fiber Telecommunications VB*. Amsterdam, The Netherlands: Elsevier, 2008, pp. 23–93.
- [3] K. Kikuchi, "Fundamentals of coherent optical fiber communications," *J. Lightw. Technol.*, vol. 34, no. 1, pp. 157–179, Jan. 2016.
- [4] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416–3425, Oct. 2008.
- [5] E. Ip, A. P. T. Lau, D. J. Barros, and J. M. Kahn, "Coherent detection in optical fiber systems," *Opt. Exp.*, vol. 16, no. 2, pp. 753–791, 2008.
- [6] S. J. Savory, "Digital filters for coherent optical receivers," *Opt. Exp.*, vol. 16, no. 2, pp. 804–817, 2008.
- [7] "Qsfp-dd whitepaper, optimizing qsfp-dd systems to achieve at least 25 watt thermal port performance." [Online]. Available: <http://www.qsfp-dd.com/wp-content/uploads/2021/01/2021-QSFP-DD-MSA-Thermal-Whitepaper-Final.pdf>
- [8] K. V. Valavala, K. D. Coulson, M. C. Rajagopal, D. Gelda, and S. Sinha, "Thermal engineering at the limits of the CMOS Era," in *Handbook of Thin Film Deposition*. Amsterdam, The Netherlands: Elsevier, 2018, pp. 63–101.
- [9] E. F. Mateo, F. Yaman, and G. Li, "Efficient compensation of inter-channel nonlinear effects via digital backward propagation in WDM optical transmission," *Opt. Exp.*, vol. 18, no. 14, pp. 15144–15154, 2010.
- [10] R. Dar, M. Feder, A. Mecozzi, and M. Shtaif, "Inter-channel nonlinear interference noise in WDM systems: Modeling and mitigation," *J. Lightw. Technol.*, vol. 33, no. 5, pp. 1044–1053, Mar. 2015.
- [11] D. Tang, Z. Wu, Z. Sun, X. Tang, and Y. Qiao, "Joint intra and inter-channel nonlinearity compensation based on interpretable neural network for long-haul coherent systems," *Opt. Exp.*, vol. 29, no. 22, pp. 36242–36256, 2021.
- [12] F. Zhang, X. Fang, and X. Chen, "Neural network-based fiber nonlinearity mitigation in high-speed coherent optical transmission systems," in *Proc. Opt. Fiber Commun. Conf.* Optica Publishing Group, 2022, pp. MIH–1.
- [13] T. Inoue, R. Matsumoto, and S. Namiki, "Learning-based digital back propagation to compensate for fiber nonlinearity considering self-phase and cross-phase modulation for wavelength-division multiplexed systems," *Opt. Exp.*, vol. 30, no. 9, pp. 14851–14872, 2022.
- [14] S. Zhang et al., "Field and lab experimental demonstration of nonlinear impairment compensation using neural networks," *Nature Commun.*, vol. 10, no. 1, pp. 1–8, 2019.
- [15] F. Frey, R. Elschner, and J. K. Fischer, "Estimation of trends for coherent dsp asic power dissipation for different bitrates and transmission reaches," in *Proc. Photonic Netw.; 18. ITG-Symp.* VDE, 2017, pp. 1–8.
- [16] J. Shalf, "The future of computing beyond Moore's law," *Philos. Trans. Roy. Soc. A*, vol. 378, no. 2166, 2020, Art. no. 20190061.
- [17] M. M. Waldrop, "The chips are down for Moore's law," *Nature News*, vol. 530, no. 7589, pp. 144–147, 2016.
- [18] A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," *J. Lightw. Technol.*, vol. 32, no. 21, pp. 3427–3439, Nov. 2014.
- [19] Y. Shen et al., "Deep learning with coherent nanophotonic circuits," *Nature Photon.*, vol. 11, no. 7, pp. 441–446, 2017.
- [20] Z. Cheng, C. Ríos, W. H. Pernice, C. D. Wright, and H. Bhaskaran, "On-chip photonic synapse," *Sci. Adv.*, vol. 3, no. 9, 2017, Art. no. e1700160.
- [21] A. N. Tait et al., "Neuromorphic photonic networks using silicon photonic weight banks," *Sci. Rep.*, vol. 7, no. 1, pp. 1–10, 2017.
- [22] X. Lin et al., "All-optical machine learning using diffractive deep neural networks," *Sci.*, vol. 361, no. 6406, pp. 1004–1008, 2018.
- [23] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, 2019.
- [24] J. Feldmann et al., "Parallel convolutional processing using an integrated photonic tensor core," *Nature*, vol. 589, no. 7840, pp. 52–58, 2021.
- [25] B. J. Shastri et al., "Photonics for artificial intelligence and neuromorphic computing," *Nature Photon.*, vol. 15, no. 2, pp. 102–114, 2021.
- [26] C. Huang et al., "Prospects and applications of photonic neural networks," *Adv. Physics: X*, vol. 7, no. 1, 2022, Art. no. 1981155.
- [27] C. Huang et al., "Neuromorphic photonics for optical communication systems," in *Proc. Opt. Fiber Commun. Conf.*, 2021, Paper M5B.1. [Online]. Available: <http://opg.optica.org/abstract.cfm?URI=OFC-2021-M5B.1>
- [28] L. Appeltant et al., "Information processing using a single dynamical node as complex system," *Nature Commun.*, vol. 2, no. 1, pp. 1–6, 2011.
- [29] K. Vandoorne et al., "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nature Commun.*, vol. 5, no. 1, pp. 1–6, 2014.
- [30] Q. Vinckier et al., "High-performance photonic reservoir computer based on a coherently driven passive cavity," *Optica*, vol. 2, no. 5, pp. 438–446, 2015.
- [31] L. Larger et al., "High-speed photonic reservoir computing using a time-delay-based architecture: Million words per second classification," *Phys. Rev. X*, vol. 7, no. 1, 2017, Art. no. 011015.
- [32] M. Nakajima, K. Tanaka, and T. Hashimoto, "Scalable reservoir computing on coherent linear photonic processor," *Commun. Phys.*, vol. 4, no. 1, pp. 1–12, 2021.
- [33] P. Antonik et al., "Online training of an opto-electronic reservoir computer applied to real-time channel equalization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 11, pp. 2686–2698, Nov. 2017.
- [34] A. Argyris, J. Bueno, and I. Fischer, "Photonic machine learning implementation for signal recovery in optical communications," *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, 2018.
- [35] F. Denis-Le Coarer et al., "All-optical reservoir computing on a photonic chip using silicon-based ring resonators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 24, no. 6, pp. 1–8, Nov./Dec. 2018.
- [36] A. Argyris et al., "Comparison of photonic reservoir computing systems for fiber transmission equalization," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–9, Feb. 2020.
- [37] F. Da Ros, S. M. Ranzini, H. Bülow, and D. Zibar, "Reservoir-computing based equalization with optical pre-processing for short-reach optical transmission," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 5, pp. 1–12, Sep./Oct. 2020.
- [38] S. Sackesyn, C. Ma, J. Dambre, and P. Bienstman, "Experimental realization of integrated photonic reservoir computing for nonlinear fiber distortion compensation," *Opt. Exp.*, vol. 29, no. 20, pp. 30991–30997, 2021.
- [39] G. Donati, C. R. Mirasso, M. Mancinelli, L. Pavesi, and A. Argyris, "Microring resonators with external optical feedback for time delay reservoir computing," *Opt. Exp.*, vol. 30, no. 1, pp. 522–537, 2022.
- [40] K. Nakajima and I. Fischer, *Reservoir Computing*. Berlin, Germany: Springer, 2021.
- [41] D. Brunner et al., "Tutorial: Photonic neural networks in delay systems," *J. Appl. Phys.*, vol. 124, no. 15, 2018, Art. no. 152004.
- [42] T. F. De Lima et al., "Primer on silicon neuromorphic photonic processors: Architecture and compiler," *Nanophotonics*, vol. 9, no. 13, pp. 4055–4073, 2020.
- [43] C. Huang et al., "A silicon photonic–electronic neural network for fibre nonlinearity compensation," *Nature Electron.*, vol. 4, no. 11, pp. 837–844, 2021.
- [44] G. P. Agrawal, "Nonlinear fiber optics," in *Nonlinear Science at the Dawn of the 21st Century*. Berlin, Germany: Springer, 2000, pp. 195–211.
- [45] D. Semrau, R. I. Killey, and P. Bayvel, "The Gaussian noise model in the presence of inter-channel stimulated Raman scattering," *J. Lightw. Technol.*, vol. 36, no. 14, pp. 3046–3055, Jul. 2018.
- [46] H.-T. Peng et al., "A photonic-circuits-inspired compact network: Toward real-time wireless signal classification at the edge," 2021, *arXiv:2106.13865*.
- [47] A. N. Tait et al., "Microring weight banks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 22, no. 6, pp. 312–325, Nov./Dec. 2016.
- [48] A. N. Tait et al., "Feedback control for microring weight banks," *Opt. Exp.*, vol. 26, no. 20, pp. 26422–26443, 2018.
- [49] P. Y. Ma et al., "Photonic principal component analysis using an on-chip microring weight bank," *Opt. Exp.*, vol. 27, no. 13, pp. 18329–18342, 2019.
- [50] C. Huang et al., "Demonstration of scalable microring weight bank control for large-scale photonic integrated circuits," *APL Photon.*, vol. 5, no. 4, 2020, Art. no. 040803.
- [51] W. Zhang et al., "Silicon microring synapses enable photonic deep learning beyond 9-bit precision," *Optica*, vol. 9, no. 5, pp. 579–584, 2022.
- [52] A. N. Tait et al., "Silicon photonic modulator neuron," *Phys. Rev. Appl.*, vol. 11, no. 6, 2019, Art. no. 064043.
- [53] F. Ashtiani, A. J. Geers, and F. Aflatouni, "An on-chip photonic deep neural network for image classification," *Nature*, vol. 606, pp. 501–506, 2022.
- [54] S. Bandyopadhyay et al., "Single chip photonic deep neural network with accelerated training," 2022, *arXiv:2208.01623*.
- [55] T. F. de Lima et al., "Noise analysis of photonic modulator neurons," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–9, Jan./Feb. 2020.
- [56] G. Roelkens et al., "Iii-v/silicon photonics for on-chip and intra-chip optical interconnects," *Laser Photon. Rev.*, vol. 4, no. 6, pp. 751–779, 2010.

- [57] R. Stabile, "Neuromorphic photonics: 2D or not 2D?," *J. Appl. Phys.*, vol. 129, no. 20, 2021, Art. no. 200901.
- [58] X. Xiao et al., "Large-scale and energy-efficient tensorized optical neural networks on III-V-on-silicon MOSCAP platform," *APL Photon.*, vol. 6, no. 12, 2021, Art. no. 126107.
- [59] M. A. Nahmias et al., "A laser spiking neuron in a photonic integrated circuit," 2020, *arXiv:2012.08516*.
- [60] M. A. Nahmias et al., "Photonic multiply-accumulate operations for neural networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–18, 2019.
- [61] C. Huang et al., "On-chip programmable nonlinear optical signal processor and its applications," *IEEE J. Sel. Topics Quantum Electron.*, vol. 27, no. 2, pp. 1–11, Mar./Apr. 2021.
- [62] A. Jha, C. Huang, and P. R. Prucnal, "Reconfigurable all-optical nonlinear activation functions for neuromorphic photonics," *Opt. Lett.*, vol. 45, no. 17, pp. 4819–4822, 2020.
- [63] J. R. Campo and D. Pérez-López, "Reconfigurable activation functions in integrated optical neural networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 28, no. 4, pp. 1–13, Jul./Aug. 2022.
- [64] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," *Adv. Neural Inf. Process. Syst.*, vol. 32, 2019.
- [65] P. J. Werbos, "Backpropagation through time: What it does and how to do it," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct. 1990.
- [66] V. T. Maker, "VPI component maker, user's manual, photonic modules reference manuals," Accessed: 6 Jan. 2022. [Online]. Available: <http://www.vpiphotonics.com>
- [67] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.
- [68] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. thirteenth Int. Conf. Artif. Intell. Statist. JMLR Workshop Conf. Proc.*, 2010, pp. 249–256.
- [69] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [70] Q.-S. Xu and Y.-Z. Liang, "Monte Carlo cross validation," *Chemometrics Intell. Lab. Syst.*, vol. 56, no. 1, pp. 1–11, 2001.
- [71] O. E. Agazzi, M. R. Hueda, H. S. Carrer, and D. E. Crivelli, "Maximum-likelihood sequence estimation in dispersive optical channels," *J. Lightw. Technol.*, vol. 23, no. 2, pp. 749–763, Feb. 2005.
- [72] S. M. Ranzini, R. Dischler, F. Da Ros, H. Bülow, and D. Zibar, "Experimental investigation of optoelectronic receiver with reservoir computing in short reach optical fiber communications," *J. Lightw. Technol.*, vol. 39, no. 8, pp. 2460–2467, Apr. 2021.
- [73] W. Freude et al., "Quality metrics for optical signals: Eye diagram, Q-factor, OSNR, EVM and BER," in *Proc. IEEE 14th Int. Conf. Transparent Opt. Netw.*, 2012, pp. 1–4.
- [74] Y. Chen et al., "Photonic unsupervised learning processor for secure and high-throughput optical fiber communication," 2022, *arXiv:2203.03807*.
- [75] A. Bogris, C. Mesaritakis, S. Deligiannidis, and P. Li, "Fabry-perot lasers as enablers for parallel reservoir computing," *IEEE J. Sel. Topics Quantum Electron.*, vol. 27, no. 2, pp. 1–7, Mar./Apr. 2021.
- [76] A. Argyris, J. Bueno, and I. Fischer, "Pam-4 transmission at 1550 nm using photonic reservoir computing post-processing," *IEEE Access*, vol. 7, pp. 37017–37025, 2019.
- [77] X. Yu and F. Zhang, "Reservoir computing based signal recovery for 56 Gb/s PAM4 system," in *Proc. Optoelectron. Commun. Conf. Optical Society of America*, 2021, pp. S3A–5.
- [78] Z. Xu, C. Sun, J. H. Manton, and W. Shieh, "Joint equalization of linear and nonlinear impairments for pam4 short-reach direct detection systems," *IEEE Photon. Technol. Lett.*, vol. 33, no. 9, pp. 425–428, May 2021.
- [79] Z. Xu, C. Sun, T. Ji, J. H. Manton, and W. Shieh, "Feedforward and recurrent neural network-based transfer learning for nonlinear equalization in short-reach optical links," *J. Lightw. Technol.*, vol. 39, no. 2, pp. 475–480, Jan. 2021.
- [80] X. Tang, Y. Qiao, Y.-W. Chen, Y. Lu, and G.-K. Chang, "Digital pre- and post-equalization for c-band 112-gb/s PAM4 short-reach transport systems," *J. Lightw. Technol.*, vol. 38, no. 17, pp. 4683–4690, Sep. 2020.
- [81] Y. Yu, Y. Che, T. Bo, D. Kim, and H. Kim, "Reduced-state MLSE for an IM/DD system using PAM modulation," *Opt. Exp.*, vol. 28, no. 26, pp. 38505–38515, 2020.
- [82] B. Sang et al., "Multi-symbol output long short-term memory neural network equalizer for 200 Gbps IM/DD system," in *Proc. IEEE Eur. Conf. Opt. Commun. (ECOC)*, 2021, pp. 1–4.
- [83] C. Huang et al., "High-capacity space-division multiplexing communications with silicon photonic blind source separation," *J. Lightw. Technol.*, vol. 40, no. 6, pp. 1617–1632, Mar. 2022.
- [84] E. Groen et al., "10-to-112-gb/s DSP-DAC-based transmitter in 7-nm FinFET with flex clocking architecture," *IEEE J. Solid-State Circuits*, vol. 56, no. 1, pp. 30–42, Jan. 2021.
- [85] K. Li et al., "Co-design of a differential transimpedance amplifier and balanced photodetector for a sub-pj/bit silicon photonics receiver," *Opt. Exp.*, vol. 28, no. 9, pp. 14038–14054, 2020.
- [86] B. S. G. Pillai et al., "End-to-end energy modeling and analysis of long-haul coherent transmission systems," *J. Lightw. Technol.*, vol. 32, no. 18, pp. 3093–3111, Sep. 2014.
- [87] H. Ramon et al., "70 Gb/s low-power dc-coupled NRZ differential electro-absorption modulator driver in 55 nm SiGe BICMOS," *J. Lightw. Technol.*, vol. 37, no. 5, pp. 1504–1514, 2019.
- [88] H. Jayatilaka, H. Shoman, L. Chrostowski, and S. Shekhar, "Photoconductive heaters enable control of large-scale silicon photonic ring resonator circuits," *Optica*, vol. 6, no. 1, pp. 84–91, 2019.
- [89] S. Y. Siew et al., "Review of silicon photonics technology and platform development," *J. Lightw. Technol.*, vol. 39, no. 13, pp. 4374–4389, Jul. 2021.
- [90] D. T. H. Tan et al., "Wide bandwidth, low loss 1 by 4 wavelength division multiplexer on silicon for optical interconnects," *Opt. Exp.*, vol. 19, no. 3, pp. 2401–2409, 2011.
- [91] J. Gao, X. Han, X. Lei, and Y. Yu, "Tec power consumption in laser array packaging," *Opt. Quantum Electron.*, vol. 49, no. 4, pp. 1–13, 2017.

**Benshan Wang** received the B.Eng. degree in electrical information engineering from Wuhan University, Wuhan, China, in 2020. He is currently working towards the Ph.D. degree at the Chinese University of Hong Kong, Hong Kong. His research interests include neuromorphic photonics, photonic integrated circuits, optical signal processing, and nonlinear dynamics of photonic systems for neuromorphic applications.



**Thomas Ferreira de Lima** received the B.Sc. and the Ingénieur Polytechnique master's degrees from Ecole Polytechnique, Palaiseau, France, in 2016 with a focus on physics for optics and nanosciences, and the Ph.D. degree in electrical engineering in 2022 with the Lightwave Communications Research Laboratory, Department of Electrical Engineering, Princeton University, Princeton, NJ, USA. He is currently a Researcher with the NEC Laboratories America, Inc., Princeton, NJ. His research interests include real-time integrated photonic systems, nonlinear signal processing with photonic devices, spike-timing based processing, ultrafast cognitive computing, and dynamical light-matter neuro-inspired learning, and computing.

**Bhavin J Shastri** (Senior Member, IEEE) received the Honors B.Eng. (with distinction), M.Eng., and Ph.D. degrees in electrical engineering (photonics) from McGill University, Montreal, QC, Canada, in 2005, 2007, and 2012, respectively. He is currently an Assistant Professor of engineering physics with Queen's University, Kingston, ON, Canada, and a Faculty Affiliate with the Vector Institute for Artificial Intelligence, Canada. He was an NSERC and Banting Postdoctoral Fellow (2012–2016) and an Associate Research Scholar (2016–2018) with Princeton University. He has authored and coauthored more than 70 journal articles and 100 conference proceedings, seven book chapters, and given more than 60 invited talks and lectures including five keynotes and three tutorials. His research interests include silicon photonics, photonic integrated circuits, neuromorphic computing, and machine learning. He is a coauthor of the book, *Neuromorphic Photonics* (Taylor Francis, CRC Press, 2017), a term he helped to coin. He was the recipient of the 2022 SPIE Early Career Achievement Award and the 2020 IUPAP Young Scientist Prize in Optics for his pioneering contributions to neuromorphic photonics from the ICO. He is a Senior Member of Optica (formerly OSA), 2014 Banting Postdoctoral Fellowship from the Government of Canada, the 2012 D. W. Ambridge Prize for the top graduating Ph.D. student at McGill, an IEEE Photonics Society 2011 Graduate Student Fellowship, a 2011 NSERC Postdoctoral Fellowship, a 2011 SPIE Scholarship in Optics and Photonics, a 2008 NSERC Alexander Graham Bell Canada Graduate Scholarship, including the Best Student Paper Awards at the 2014 IEEE Photonics Conference, 2010 IEEE Midwest Symposium on Circuits and Systems, the 2004 IEEE Computer Society Lance Stafford Larson Outstanding Student Award, and the 2003 IEEE Canada Life Member Award.

**Paul R. Prucnal** (Life Fellow, IEEE) received the A.B. degree (summa cum laud) in mathematics and physics from Bowdoin College, the M.S., M.Phil., and Ph.D. degrees in electrical engineering from Columbia University, New York, NY, USA. After his doctorate, he joined the Faculty with Columbia University in 1979. He is currently a Professor of electrical engineering with Princeton University, Princeton, NJ, USA. He is best known for his seminal work in Neuromorphic Photonics, optical code division multiple access (OCDMA) and the invention of the terahertz optical asymmetric demultiplexer (TOAD). As a Member of the Columbia Radiation Laboratory, he performed groundbreaking work in Optical CDMA, which initiated a new research field in which more than 1000 papers have since been published, exploring applications ranging from information security to communication speed and bandwidth. In 1988, he joined the faculty at Princeton University. He is the Author of the book, *Neuromorphic Photonics*, and the Editor of the book, *Optical Code Division Multiple Access: Fundamentals and Applications*. He has authored or coauthored more than 300 journal articles, 390 conference papers and 34 book chapters, and holds 28 U.S. patents. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE), the Optical Society of America (OSA) and the National Academy of Inventors (NAI), and a Member of Phi Beta Kappa and Sigma Xi. He was the recipient of the Gold Medal from the Faculty of Mathematics, Physics, and Informatics at Comenius University, for leadership in the field of optics and numerous teaching awards at Princeton, including the President's Award for Distinguished Teaching, the Distinguished Teacher Award, School of Engineering and Applied Science, the Engineering Council Lifetime Achievement Award for Excellence in Teaching, the Graduate Mentoring Award, the Walter Curtis Johnson Prize for Teaching Excellence.



**Chaoran Huang** (Member, IEEE) received the B.Eng. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2012, and the Ph.D. degree from the Chinese University of Hong Kong, Hong Kong, in 2016. She is currently an Assistant Professor with the Chinese University of Hong Kong (CUHK), Hong Kong. Before joining CUHK, she was a Postdoctoral Research Fellow with Princeton University, Princeton, NJ, USA, from 2017 to 2021. She has authored more than 40 peer-reviewed research papers, three book chapters, and one US patent, and has given more than 15 invited talks at different international conferences. Her research interests include photonic neuromorphic computing, optical computing, silicon photonics, photonic integrated circuits, and optical communications. Her current research interest is to develop neuromorphic photonic platforms, including novel devices, photonic integrated circuits, and complementary algorithms for high-performance AI computing and information processing. She has been the TPC Member of several international conferences and is a frequent reviewer for different journals in IEEE, OSA, and the Nature Publishing Group. She was the recipient of 2019 Rising Stars Women in Engineering Asia, and was nominated by Princeton University to compete for Blavatnik Regional Awards for Young Scientists.