High-Capacity Space-Division Multiplexing Communications With Silicon Photonic Blind Source Separation

Chaoran Huang, Dongliang Wang, Weipeng Zhang, *Graduate Student Member, IEEE*, Benshan Wang, Alexander N. Tait, Thomas Ferreira de Lima, Bhavin J. Shastri, and Paul R. Prucnal, *Life Fellow, IEEE*

(Invited Paper)

***Abstract*—Space-division multiplexing is a widely used technique to improve data-carrying capacities in both wireless and optical communication systems. However, tightly packed spatial channels cause severe crosstalk. High data rates and large channel counts impose severe constraints on resolving the crosstalk using traditional digital signal processing algorithms and electronic circuits. In order to solve these issues, this paper presents a silicon photonic system combining high-speed silicon photonic devices with a novel blind source separation (BSS) algorithm. We first demonstrate using photonic BSS to undo modal crosstalk in a short-reach multimode optical fiber interconnect for intra-data-center communications. The proposed photonic BSS system inherits the advantages of photonic matrix processor and the "blindness" of BSS, leading to superior energy and cost efficiency and reduced latency, while allowing to recover the signals using a sub-Nyquist sampling rate and in a free-running mode, and offering unmatched agility in signal format and data rate. The feasibility of using photonic processors for mode crosstalk equalization has been recently demonstrated, assisted with training sequences. Our approach, photonic BSS, in contrast, can tackle the more difficult problem of making the receiver transparent to any data rate and modulation format, and workable with slow and cost-effective electronics. In addition, we find that photonic BSS has a much better scaling law for space-division multiplexing (SDM)-based communication systems than digital signal processing (DSP). When compared to state-of-the-art DSP, photonic BSS can reduce system power consumption, speed, and latency by several orders of magnitude, particularly for high-capacity communications with high data rates per channel and a large number of channels. Photonic BSS has the added advantages of being agnostic to transmission content, making it exceptional at protecting communication privacy. This paper also discusses our previous work in demonstrating photonic BSS for privacy protection in wireless multiple-in multiple-out (MIMO) communications using silicon photonic micoring resonator (MRR) weight banks.**

***Index Terms*—Machine learning, optical signal processing, silicon photonics, space-division multiplexing.**

Manuscript received August 31, 2021; revised December 22, 2021; accepted February 4, 2022. Date of publication February 16, 2022; date of current version March 16, 2022. This work was supported by CUHK Direct Grant 170257018 and CUHK startup fund. The devices were fabricated at the IME ASTAR foundry in Singapore. Fabrication support was provided via the Natural Sciences and Engineering Research Council of Canada (NSERC), Silicon Electronic-Photonic Integrated Circuits (SiEPIC) Program and the Canadian Microelectronics Corporation (CMC). *(Corresponding author: Chaoran Huang.)*

Chaoran Huang was with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08540 USA. She is now with Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: crhuang@ee.cuhk.edu.hk).

Dongliang Wang and Benshan Wang are with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong (e-mail: 1155167615@link.cuhk.edu.hk; bswang@link.cuhk.edu.hk).

Weipeng Zhang, Thomas Ferreira de Lima, and Paul R. Prucnal are with the Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08540 USA (e-mail: weipengz@princeton.edu; tlima@princeton.edu; prucnal@princeton.edu).

Alexander N. Tait is with the Department of Electrical and Computer Engineering, Queen's University, Kingston, ON KL7 3N6, Canada (e-mail: atait@ieee.org).

Bhavin J. Shastri is with the Department of Physics, Engineering Physics and Astronomy, Queen's University, Kingston, ON KL7 3N6, Canada (e-mail: bhavin.shastri@queensu.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2022.3152027.

Digital Object Identifier 10.1109/JLT.2022.3152027

I. Introduction

TODAY'S world is witnessing an explosion of network traffic [1]. Numerous emerging services and applications, such as cloud service, video streaming, and Internet of Things, are accelerating the demand for high-capacity communications even further. Space-division multiplexing (SDM) is a multiplexing technology used to increase data transmission capacity by exploring the new degrees of freedom in space to establish parallel data stream pipelines. As a result, SDM increases the communication channel throughput without increasing bandwidth usage or transmission power.

SDM based technologies have benefited both optical and wireless communications. In optical communications, SDM has also been used in optical communication systems using multimode or multi-core optical fibers (i.e. photonic Multiple-In Multiple-Out MIMO) [2]–[5] (Fig. 1(a)). SDM promises to overcome the problem of nonlinear Shannon limit encountered in single mode fibers, resulting in a massive increase in data transmission capacity [6]. Similarly, in wireless communications, SDM is a well-known Multiple-In Multiple-Out (MIMO) technique which uses multiple antennas to send and receive independent and separately encoded wireless signals (Fig. 1(b)), and is the key enabler in many modern wireless technologies such as 4G long-term evolution (LTE) and 5G [7], [8].

0733-8724 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

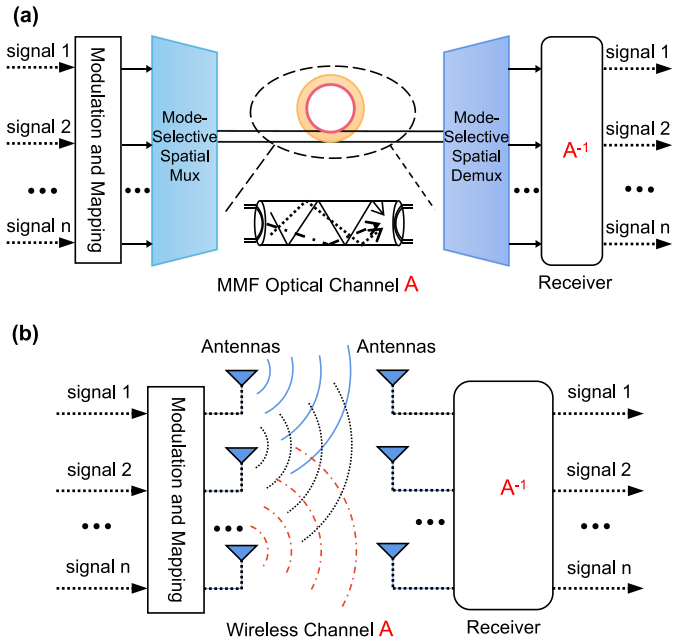


Fig. 1. Spatial division multiplexing (SDM) for (a) multi-mode fiber (MMF) optical fiber communication system. (b) multi-antenna wireless communication system. $\boldsymbol{A}$ is the channel state for each system. Signal processing is needed at the receiver to undo the channel induced signal distortions.

A. Advantages of Blind Source Separation (BSS)

Dense SDM is always accompanied by crosstalks caused by deleterious effects in transmission channels. Signal processing techniques are needed to retrieve the transmitted data streams. To do this, the receiver needs to estimate the channel state that characterized the combined transmission effects (i.e., channel estimation), and then inverts the distortions (i.e., channel equalization). Least squares (LS) and minimum mean square error (MMSE) and their numerous variations are popular channel estimation techniques [9], [10]. However, LS and MMSE need prior knowledge of transmitted signals or the incorporation of training symbols [11]. The training signals obviously decrease communications throughput, especially for time-varying channels.

Unlike previous methods, blind source separation (BSS) is an unsupervised learning technique that requires minimal knowledge of either source or channel characteristics. Instead of relying on training symbols, BSS only relies on signal's statistical features and estimate channels state via maximization of signal independence [12], [13]. For this reason, unlike most algorithms, a significant advantage of BSS is it allows to sample signals at a far slower frequency than the Nyquist frequency, and the sampling does not need to synchronize with the signal, both resulting in significantly lower power consumption and implementation cost. In addition, BSS is completely transparent to modulation format, data rate, and waveforms, offering unrivaled agility in communications systems. Algorithm details will be discussed in Section II-A and Appendix A.

B. Photonics versus DSP

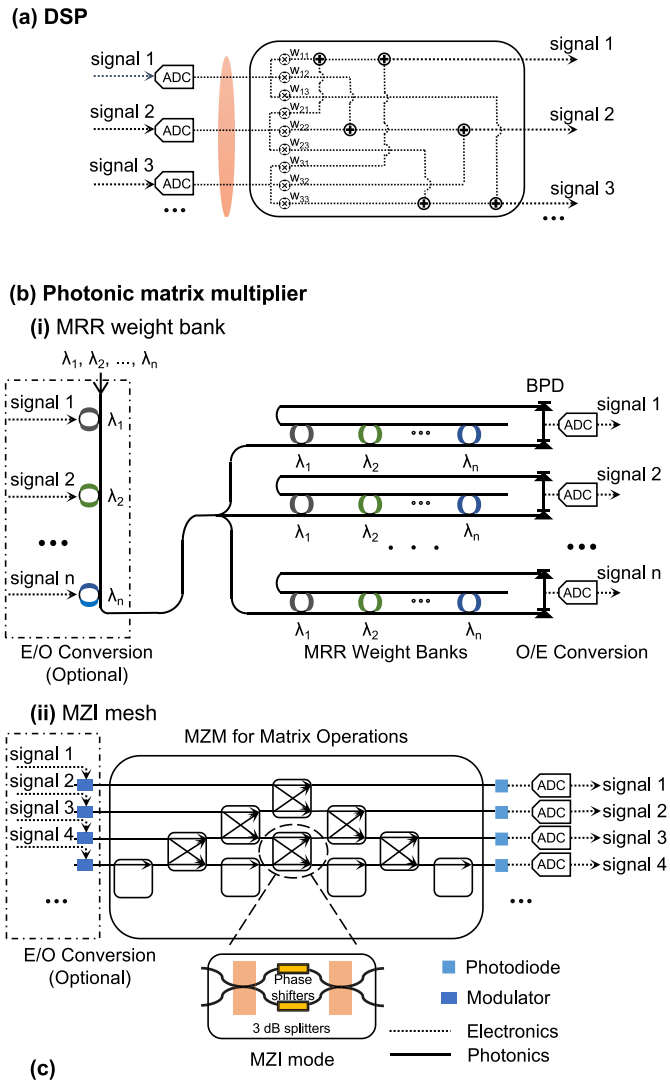
Despite the existence of many mature algorithms for SDM-based communication system, their hardware implementation remains a challenging problem, because SDM-based communication systems are expected to process high-dimensional and high-speed signals in a short period of time (i.e., low latency) and with low energy usage. However, digital electronics is facing harsh tradeoffs between the number of spatial channels, energy use, and bandwidth [14]. A typical receiver for SDM based communication systems is composed of an array of analog-to-digital converters (ADCs) followed by a digital signal processing (DSP) backend as shown in Fig. 2(a). In this case, the energy use E grows in proportions with the signal bandwidth B_s and quadratically with channel number N [15]. The processing latency is ultimately bounded by the clock rate f_{clock}, even with the efforts on parallelism [16], which causes additional interconnect problem [17], [18]. The processing latency of digital electronics also increase quadratically with the channel number. Current communication systems already have to operate at Peta (10^{15}) multiply–accumulate (MAC) operations per second [19]. As a result, as data throughput increases in the future, DSP will confront formidable challenges in terms of processing speed and will result in an explosion in energy consumption.

Photonics provides a completely different scaling physics that can solve many fundamental constrains of digital electronics [16], [20], [21], as shown in Fig. 2(c). Photonic processors deal with signals in the optical domain (i.e., photonic front-end), prior to digitization, breaking the ADC imposed limitations in a fundamental way [21] (Fig. 2(b)). Photonic processors have a nearly flat frequency response to the signals over 5-THz window. As a result, the energy use of photonic processors is not affected by the signal bandwidth B_s, and they only increase in proportion with N, in contrast to N^2 in digital electronics [16]. Furthermore, the processing latency is only limited by the bandwidths of the optical components on the signal pathway, leading photonic processors to have much higher speeds than digital devices. For this reason, photonics processors can offer high bandwidth with low energies and have been used in many applications including intelligent signal processing ([22] and this paper), neuromorphic computing [17], [18], [23], [24], quantum computing [25]–[27], and optimization problems such as Ising machine [28], [29]. These application systems have two major features in common: 1) the same task (especially matrix operations) needs to be done repeatedly and needs to be done quickly; and 2) the signals to be processed are already in the analog domain (optical, wireless) [17], [18], [22].

C. Paper Contributions and Organizations

To address the energy and latency challenges in SDM communication systems, this paper introduces photonic BSS [33], [34], a multivariate photonic processor with high bandwidth and low latency that handles high-dimension signal processing at the physical layer. We propose the use of photonic BSS for multi-mode optical fiber interconnects to solve DSP constraints and enable high-capacity and low-power data-center interconnects in Section II. We demonstrate that, without inserting training

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.



(c)

Parameters	Electronics	Photonics
Energy	$\propto N^2 \cdot B_s$	$\propto N^2$ (or N)
Latency	$\propto N^2 / f_{clock}$	$\propto 1/B_{MOD,PD}$

Fig. 2. Building blocks of (a) DSP, (b) photonic processor for matrix multiplications. (b.i) photonic matrix operator using microring resonator (MRR) weight banks [30] and multiport Mach-Zender interferometers (MZI) meshes [31], [32]. If the inputs are electrical signals, the signals can be modulated to the optical domain using an array of optical modulators before processing. Balanced photodetector (BPD) if the inputs are optical signals, they can be coupled to the following optical devices directly, and (c) scaling law of photonic and electronic devices in terms of channel number N, signal frequency f_{signal}, electronic device clock f_{clock}, optical device bandwidth of modulator B_MOD, and bandwidth of photodetector B_PD.

symbols, our approach can recover multi-channel data streams using a sub-Nyquist sampling rate running in a free-running mode and unscramble random modal crosstalk varying with millisecond-timescale.

In this section, we first present the principle, algorithm, and hardware that bridge BSS with analog photonic integrated hardware. We then demonstrate to recover 50 Gbaud/s per channel pulse amplitude modulation 4-level (PAM4) data streams using a sub-Nyquist sampling rate of only 2 GSample/s in a free-running mode. In Section III, we provide comprehensive analysis showing that photonic BSS has distinct benefits over DSP in terms of energy consumption and bandwidth as the channel number and data rate increases, showing our approach is especially important for high-throughput interconnects. Section IV experimentally demonstrates another application of photonic BSS for privacy protection in wireless MIMO communications, as a review of our previous work [20], [33]–[35]. Using these examples, we show that the benefits of photonic BSS system from the combination of a modified hardware-friendly BSS algorithm and a reconfigurable photonic matrix operator: photonic BSS inherits the advantages from both hardware and algorithm prospects, and, as a result, can relieve some fundamental limitations of digital electronics such as latency, bandwidth, energy efficiency, agility, and throughput, in particular for high bandwidth and high-throughput telecommunication systems.

II. Photonic BSS for Multimode Optical Fiber Interconnects

Data centers play a critical part of today information technology (IT) services, computing and communications. Data center traffic has increased 10 times in the past decade and this trend would continue in the future. On the other hand, data center is gobbling up the world's electrical power. Data-centre is likely to take up 8% of the global electricity use by 2030 [36]. Therefore, a paradigm shift in optical communication technologies is needed to improve data throughout and energy efficiency of data center interconnects.

The use of mode multiplexing in multi-mode fibers can scale the information capacity of an optical transmission link with the number of spatial modes. However, modal crosstalk increases with the number of spatial channels, limiting the achievable data rate and transmission distance. In long-haul communication systems, power-hungry DSP is widely deployed to address modal crosstalk using well-developed MIMO algorithms implemented with digital electronics [37]. However, optical interconnects for data center applications are highly sensitive to latency, energy consumption, and cost. For this reason, DSP is typically unacceptable due to its high latency, power consumption, and upgrading cost (the application-specific integrated circuit (ASIC) needs to be re-designed when the data rate and channel number increase).

Here, we focus on solving the modal crosstalk problem of SDM optical fiber interconnects for intra-data-center applications. Intra-data-center interconnects connect one server to another within a data center, which constitutes more than 70% volume among all market segments in data center interconnects. A typical SDM based optical fiber interconnect for short-reach transmissions is shown in Fig. 3. At the transmitter, the single-wavelength laser source is split into N branches, and then each branch is modulated by independent data. The modulated signals are then combined by a spatial mode multiplexer which converts the signal into orthogonal eigenmodes. The multimode fiber can ideally carry orthogonal modes without interference, however, random perturbations along the fiber would cause mode coupling

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

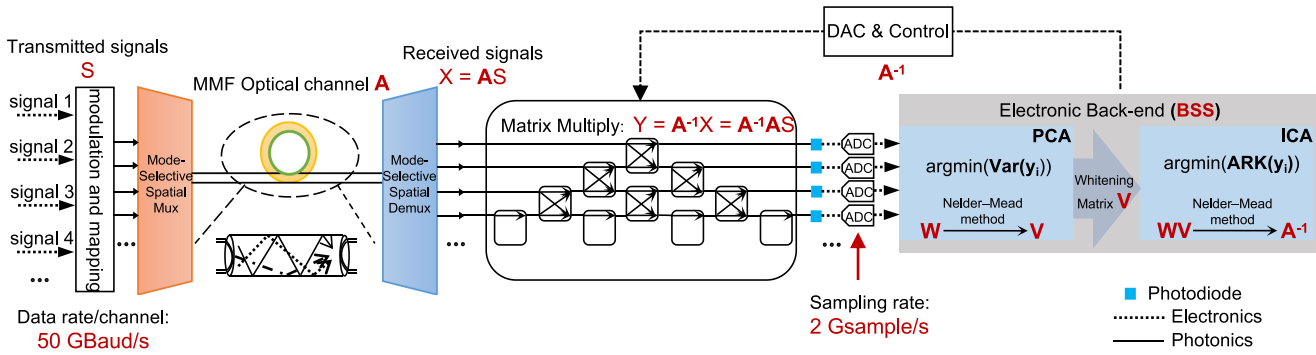


Fig. 3. Multimode fiber link with photonic BSS. The link is composed of the multimode optical fiber communication system, MZM and the electronic back-end block. The photonic BSS algorithm would be realized by MZM and electronic back-end block. The data rate of transmitted signals is 50 GBaud/s and the sampling rate of ADCs is 2 GSample/s.

across different modes. Within this distance of intra-data-center links, chromatic and modal dispersion can be neglected. In this case, modal crosstalk can be modeled as an $N \times N$ unitary matrix with complex-valued elements, where N is the number of spatial modes [38]. The element values in the unitary matrix randomly drift with the perturbations along with the fiber and have millisecond-timescale dynamics [39]. At the receiver, the signals are demultiplexed by another mode demultiplexer [40] and then processed to equalize transmission channel induced distortions.

Our approach combines silicon photonic processors with BSS to mitigate modal crosstalk, in order to address the challenges of the high cost and power consumption in high-speed DSP MIMO processors. The system is illustrated in Fig. 3. The feasibility of using photonic processors for mode crosstalk equalization has been recently demonstrated [38], [41], assisted with training sequences. Our approach, photonic BSS, in contrast, can tackle the more difficult problem of making the receiver transparent to any data rate and modulation format, and workable with slow and cost-effective electronics.

In following two subsections, we first provide a succinct exposition of the BSS algorithm and then present the proposed photonic hardware to implement the algorithm.

A. BSS Algorithm: Independent Component Analysis (ICA)

A number of BSS algorithms have been reviewed in [13], [42]. Depending on the assumption made to the sources, the algorithms are classified into ICA, sparse component analysis (SCA), non-negative matrix factorization (NMF), and bounded component analysis (BCA), as reviewed in [42]. The most commonly used blind separation method is ICA [12], [43], which is based on two generally accepted assumptions: 1) the source signals are unrelated to one another; 2) each source signal has non-Gaussian distribution. These two assumptions hold true for communication signals encoded in orthogonal SDM channels. Under these assumptions, ICA separates signals using the central limit theorem, which says that the sum of independent sources is more Gaussian than any individual sources. So ICA separates signal by maximizing the relative distance from the Gaussian distribution of each independent components.

1) Theory: Consider a group of independent time-series signal $S_i(t)\,(i = 1, ..., N)$ form a transmitter signal matrix $\boldsymbol{S}$. These signals are mixed in the transmission channel and result in the observed mixtures $\boldsymbol{X}$ at the receiver. The transmission channel model can be described by

$$\boldsymbol{X} = \boldsymbol{AS} + \boldsymbol{C}, \tag{1}$$

where $\boldsymbol{A}$ is the matrix containing the mixing coefficient, and $\boldsymbol{C}$ is the noise in the transmission channel. The goal of ICA is to figure out the inverse matrix of $\boldsymbol{A}$, known as demixing matrix $\boldsymbol{A}^{-1}$, so that we can retrieve the transmitter signals by multiplying them with the received mixtures $\bar{\boldsymbol{S}} = \boldsymbol{A}^{-1}\boldsymbol{X}$.

ICA searches for $\boldsymbol{A}^{-1}$ by maximizing the relative distance from the Gaussian distribution, that is, the non-Gaussianity of each independent component, measured by kurtosis, a fourth-order moment,

$$kurt(\boldsymbol{y}) = \frac{E[\boldsymbol{y} - \bar{\boldsymbol{y}}]^4}{(E[(\boldsymbol{y} - \bar{\boldsymbol{y}})^2])^2} - 3, \tag{2}$$

where $\boldsymbol{y}$ is a given IC estimation, and $\bar{\boldsymbol{y}}$ is the mean value of $\boldsymbol{y}$. Directly searching for $\boldsymbol{A}^{-1}$ has N^2 degrees of freedom, and is therefore difficult to calculate. Therefore, instead of searching for $\boldsymbol{A}^{-1}$ from the received signal $\boldsymbol{X}$ directly, the ICA algorithm first applies a whitening matrix $\boldsymbol{V}$ to $\boldsymbol{X}$ such that the matrix after transformation $\boldsymbol{Z} = \boldsymbol{VX}$ is white, i.e., expectation $E(\boldsymbol{ZZ^T}) = \mathbf{I}$, where $\mathbf{I}$ is the identify matrix. It is commonly assumed that independence still holds for the estimated independent components, i.e., $E(\bar{\boldsymbol{S}}\bar{\boldsymbol{S}}^{\boldsymbol{T}}) = \mathbf{I}$. In this case, we consider a new matrix $\boldsymbol{Q}$ that can transfer $\boldsymbol{Z}$ to $\bar{\boldsymbol{S}}$, i.e.,

$$\bar{\boldsymbol{S}} = \boldsymbol{QZ} = \boldsymbol{QVX}. \tag{3}$$

The new transformation matrix $\boldsymbol{Q}$ naturally becomes an orthogonal matrix, making it considerably simpler to find $\boldsymbol{Q}$ since orthogonal matrices have degrees of freedom of $N(N-1)/2$ in contrast to N^2. Whitening matrix $\boldsymbol{V}$ can be derived from principal component analysis (PCA) using the formula $\boldsymbol{V} = \boldsymbol{U\Sigma^{-1/2}U^T}$, where $\boldsymbol{U}$ contains principal component (PC) vectors of $\boldsymbol{X}$ as its columns while diagonal matrix $\boldsymbol{\Sigma}$ has PC variances as its diagonal entries. As a result, the BSS algorithm consists of two consecutive steps: (1) **PCA** searching for $\boldsymbol{U}$, and (2) **ICA** searching for $\boldsymbol{Q}$.

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

2) PCA: PCA is the oldest and the most widely used technique in multivariate analysis. PCA uses a linear matrix $\boldsymbol{W}$ to transfers data $\boldsymbol{X}$ to a new coordinate system $\boldsymbol{U}$. Such transformation causes the projection of the data on the first coordinate to have the largest variance (referred to as first principal component), on the second coordinate to have the second greatest variance, and so on. In this way, PCA produces an array of decorrelated signals ordered in decreasing order of statistical variance. Thus, the goal of PCA is to find the projection matrix $\boldsymbol{U}$ that maximizes the variance σ^2, i.e., the 2nd-order (central) moment, of the projected data. Traditionally, PCA can be obtained analytically from singular value decomposition of $\boldsymbol{X}$. However, the runtime and the memory usage can be unbearable for communication systems. Instead, we use Nelder-Mead searching method [44] to continuously update the estimated projection vectors while looking for $\boldsymbol{W}$ until the projected signal yields the largest variance. This method can take the best advantages of the front-end photonic matrix operator described in the following section in accelerating the convergence speed [35]. We present the detailed implementation procedures of Nelder-Mead for PCA in Appendix-A and Appendix-B. PCA produces the whitening matrix $\boldsymbol{V}$ to simplify the subsequent ICA.

3) ICA: Whitening matrix $\boldsymbol{V}$ is introduced as an input argument to the ICA algorithm. According to Equation 3, ICA then searches for matrix $\boldsymbol{Q}$ that yields the largest output kurtosis. Similar to PCA, The Nelder-Mead method is used to continuously update the values of $\boldsymbol{Q}$ until it converges to the target IC vectors (details see Appendix A-C).

B. Hardware Implementation

From the theory, BSS consists of two major of operations: 1) matrix operations which multiply the input signal $\boldsymbol{X}$ with the estimated transformation matrix $\boldsymbol{V}$ at the PCA stage and $\boldsymbol{Q}$ at the ICA stage in each iteration until the algorithm converges to its solutions, followed by 2) statistical moment fitting which measures the second- and fourth-order statistical moments (σ^2 and $kurt$ respectively) of the transformed signals. To perform these functions, as shown in Fig. 3, the photonic processor is composed of a number of building blocks: 1) a photonic front-end for matrix operations; 2) an electronic subsystem performing moment fitting and analog-to-digital conversion (ADC) to digitize the photonic operator output and digital-to-analog conversion (DAC) to generate driving voltages in order to control the matrix, and 3) signal interface consisting of data input/output (I/O) ports and data domain conversions depending on the applications.

1) Integrated Photonic Matrix Operator: Matrix operations carried out in the optical domain only need one time step by simply allowing the signal to travel across the photonic system. The characteristics of photonic devices allow them to operate at considerably faster rates than digital devices, resulting in a less than 100 ps delay [23], [30]. In other words, the whole matrix may be calculated in less than one digital electronic clock cycle. Matrix operations can be performed with a variety of photonic systems using tunable photonic components [30]–[32], [45]. Their operation principles can be categorized into two primary mechanisms: *coherent* or *incoherent*, as defined in [16] and illustrated in Fig. 2(b).

We use the *coherent mechanism* as shown in Fig. 3, which uses a multiport interferometer with cascaded Mach-Zehnder interferometers (MZIs). Each MZI is made of two 50:50 beamsplitters and two tunable shifters. By controlling the two phase shifters, MZI is capable of executing a 2×2 complex unitary matrix operation. An $N \times N$ arbitrary unitary transformations can be accomplished using $N \times (N-1)/2$ MZIs arranged in a triangular architecture proposed by Reck *et al.* [31], a rectangular architecture by Clements *et al.* [32] or several other variants. The inputs of the multiport interferometer are N spatial optical modes separated from a single wavelength laser source. Phase shifters change the relative power and phase levels (i.e. complex values) of the output optical modes by controlling the light propagation via constructive and destructive interference. The multiport interferometer show in Fig. 3 can only perform unitary matrix transformations, and thus can only unscramble modal coupling. Adding programmable attenuators before or after the multiport interferometer can address mode-dependent loss.

Photonic matrix operators can be fabricated in a variety of technology platforms, and monolithically integrate with the multi-mode integrated receivers [5]. As shown in Fig. 3, a mode demultiplexer splits the overlapped and orthogonal modes into separate waveguides. Because the signals in the two waveguides are generated from the same laser source, they can be directly coupled into a $N \times N$ MZI and get processed by the photonic front-end prior to optical-electrical (O/E) conversion.

To integrate a large number of building blocks, these platforms must be able to offer compact and low loss components [46]. Silicon photonics has high refractive index contrast and low loss, thus providing an unprecedented platform to produce large-scale integrated optical systems. Phase shifters are the key components to control and program the photonic matrix, which is accomplished by shifting the effective refractive index of the silicon photonic waveguide. Phase shifters are expected to offer low-power, high-speed, and low-loss tuning capabilities. So far, phase shifters are mostly accomplished using two convenient effects in silicon: thermo-optic effect or plasma dispersion effect. The former involves shifting the temperature through the use of heaters [47], while the latter involves changing the carrier density via a biased p-n junction [48]. Heaters are simple to operate and are most widely used for reconfigurable photonic circuits, although they dissipate several milliwatts of electrical power when tuning and holding the states and have a relatively slow response time of 10–100 μs. Compared to heaters, the plasma-dispersion effect offers much better tuning speed and lower-power consumption; nevertheless, it requires doping, which introduces a few dB insertion loss to each component, preventing them from scaling to large-scale photonic systems. There are many less mature but better or alternative approaches for phase tuning, for example, micro-electro-mechanical systems (MEMS) [49], Si/III-V hybrid waveguide [50], lithium niobate for electrooptic effect [51], single-layer-graphene (SLG) [52], [53], and non-volatile phase change material [54]. We direct interested readers to the review article on various phase tuning methods and cutting-edge devices [46].

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

2) Electronic Subsystems: BSS algorithm only needs to measure signals' statistic properties, rather than the full waveform. In communication systems, signal statistics change on timescales related to environmental and channel fluctuations, which is much lower than the signals themselves. For this reason, unlike other DSP algorithms, BSS allows ADCs sampling at a frequency much slower than the Nyquist frequency, and the sampling does not need to synchronize with the signal's clock. As a result, the BSS algorithm offers benefits in terms of hardware cost, complexity, and energy efficiency, because they all increase with the samplers' frequency.

One of the primary functions of the electronic subsystem is to measure the second and fourth statistic moment and update the weight matrix following the Nelder-Mead method. As shown in Fig. 3, DACs convert the weight matrix to a group of voltages and apply the voltages to the photonic matrix operator via the phase shifters. Then the photonic matrix operator would compute an updated PC or IC from the optical domain "on-the-fly". The photonic processor repeats the operations until all the weight vectors pointing in nearly the same direction, at which point the algorithm is converged. The photonic matrix operator plays a crucial role in accelerating the convergence speed by two orders of magnitude.

In addition, a control system must precisely control a large number of phase shifters in order to conduct precise matrix computations. The control system consists of a circuit to implement the control algorithm and monitor components that provide insight into the internal distribution of light in MZI meshes or the partial transmission of individual MRRs. Controlling a large number of optical components can be a challenging problem because it requires a large amount of dedicated electrical I/Os for control and monitor purposes. N-doped photoconductive heater is a solution to solve this challenge in I/Os [55], [56]. The N-doped heater can actuate the phase shifter by thermally tuning, and at the same time, can monitor the transmission from the photoabsorption-induced change in the heater resistance. As a result, N-doped photoconductive heaters allow accurate control without the need for direct access to device output and the need to layout electrical and optical I/Os specific for monitor purposes [57]. We demonstrated the control of multi-channel MRR weight banks [57] and achieved 8.5-bit precision [58], outperforming the 8-bit precision utilized in DSP systems for processing communication systems.

C. Results

We first consider a short-reach optical link carrying two spatial modes modulated by two pulse amplitude modulation 4-level (PAM4) data streams. The signals on the two modes are independent of each other, but have an identical modulation format and an identical data rate, making it a challenging problem to demixing them blindly. Within the intra-data-center distance, the value of crosstalk is around $=-5$ dB and fluctuate ± 1 dB at a time scale of a few milliseconds due to the environmental fluctuations [59].

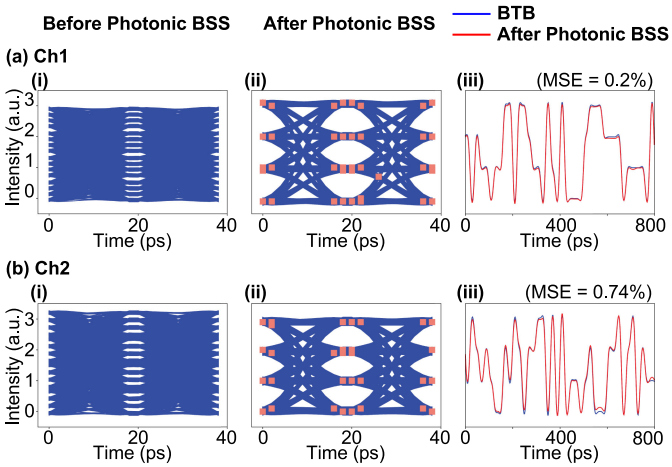


Fig. 4. (a.i)(b.i) Eye diagrams of PAM-4 signals on x and y polarization after transmission; (a.ii)(b.ii) eye diagram after being recovered by photonic BSS; (a.iii)(b.iii) signal waveforms comparison between back-to-back signals and after photonic BSS.

The transmission channel with two orthogonal modes produces a 2×2 effective mixing matrix given by

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{4}$$

where b/a denotes the crosstalk from mode x to y (defined as modal coupling coefficient K_{x-y}), and c/d denotes the crosstalk from mode y to x (denoted as K_{y-x}). The mixing of two modal states is because the mode degeneracy is broken by random variations in core shape and stress-induced anisotropy along the optical fiber [39], [60]. Under such crosstalk condition, we randomly generate an exemplary time-varying matrix to emulate the transmission channel as shown below,

$$\boldsymbol{A} = \begin{bmatrix} 0.82 & 0.23 \\ 0.18 & 0.77 \end{bmatrix} + \boldsymbol{N}(t)(0, \sigma^2) \tag{5}$$

where $\boldsymbol{N}(t)(0, \sigma^2)$ is the ambient noise which fluctuates at a frequency of 1 kHz.

At the receiver, a mode demultiplexer splits the two overlapped and orthogonal modes into two separate waveguides and directly coupled into a 2×2 MZI.

1) Results: After the mixture matrix, the eye diagrams of the received signals are plotted in Fig. 4(a.i) and (b.i), showing nearly closed eye diagrams. To determine the demixing matrix $\bar{\boldsymbol{A}}$, we apply the BSS algorithm described in Section II. The signals do not insert any training sequence are sampled at a sub-Nyquist rate of 2 Gsample/s in a free-running mode. BSS applies the transformation matrix according to the Nelder–Mead rules until the solutions for PCA and ICA converge. PCA requires 24 iterations and ICA requires 35 iterations for convergence. As shown in Fig. 4(a.ii) and (b.ii), BSS fully recovers the signals from the mixtures, resulting in clearly opened eye diagrams for both polarization channels. The samplers used to plot the eye diagrams have a time length of one second, which is considerably longer than the time scale of the matrix fluctuation, so the result demonstrates that BSS can adapt to ambient noise. We also evaluate the accuracy of BSS by comparing the recovered signal

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

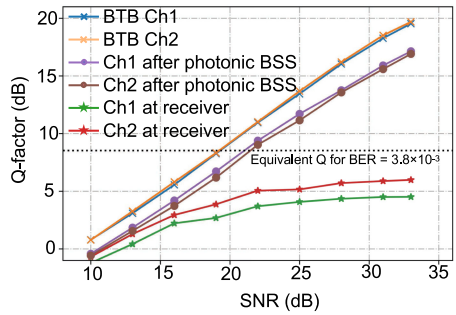


Fig. 5. System performance Q-factors at different SNRs.

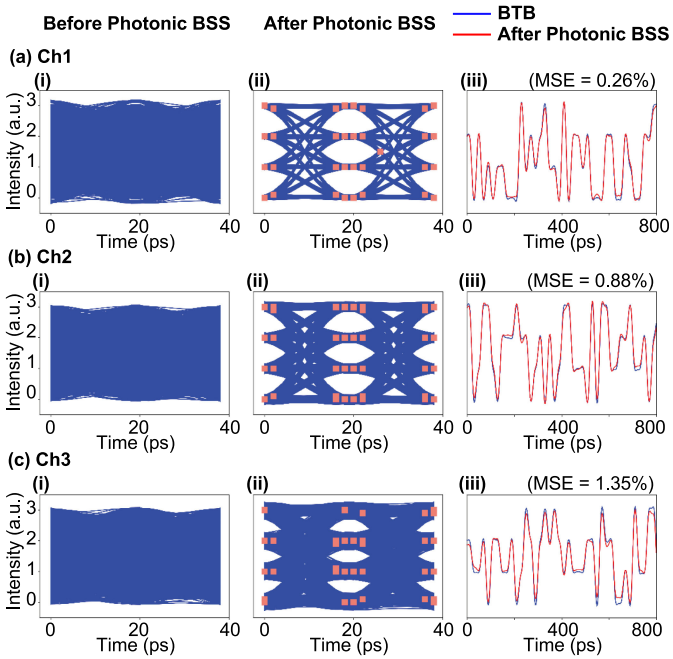


Fig. 6. (a.i)(b.i)(c.i) Eye diagrams of PAM-4 signals on three spatial modes after transmission. (a.ii)(b.ii)(c.ii) eye diagram after being recovered by photonic BSS. (a.iii)(b.iii)(c.iii) signal waveforms comparison between back to back signals and after photonic BSS.

TABLE I
MSE OF 2- AND 3-CHANNEL BSS

Channel number	IC1	IC2	IC3
2	0.20%	0.74%	
3	0.26%	0.88%	1.35%

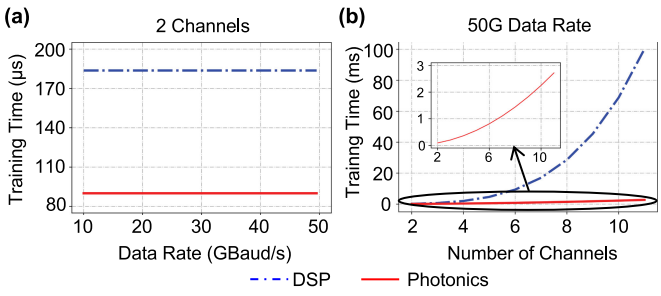


Fig. 7. The training time of DSP and photonics (a) at different data rates and (b) with different number of channels.

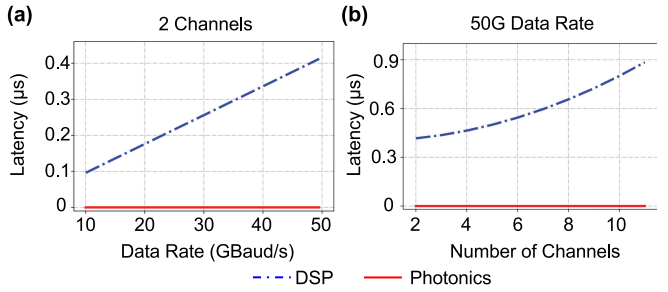


Fig. 8. During the execution phase, the latency in processing each symbol using DSP and photonics. Comparison (a) under different data rates and (b) under different number of channels.

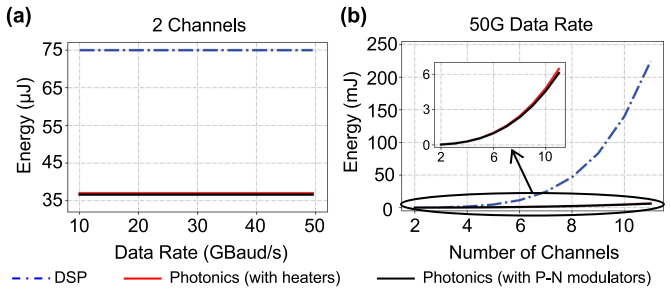


Fig. 9. During the training, the total energy use of photonics and DSP (a) at different data rates and (b) at different number of channels.

waveforms with back-to-back signals. We run the algorithm 40 times, all with newly-initialized random weight vectors, and find that the average mean square error (MSE) of the two signals is 0.20% and 0.74%, respectively.

We next evaluate the system performance (indicated by signal's Q-factor), taking the noise in the transmission channel into consideration. To do so, we assume a channel with additive white Gaussian noise, and add the noises with zero mean and varying variance to emulate a wide range of signal-to-noise ratios (SNRs). Q-factor is estimated from the signal's noise variance using the equations in [61] shown in Appendix B. The result is shown in Fig 5. The received signals are severely contaminated by the modal crosstalk; as a result, their Q factors are always below the equivalent Q factor for hard-decision forward error correction (FEC) threshold (BER $= 3.8 \times 10^{-3}$). After BSS, the system can achieve the FEC threshold at an SNR of approximately 24 dB, which has a 3 dB SNR penalty compared to the back-to-back signals.

2) Channel Scalability: Here we investigate how photonic BSS performs when it is extended to a higher number of spatial channels. To this end, we apply photonic BSS to a three-channel SDM system characterized by a 3×3 mixture matrix. As shown in Fig. 6(a)(i)(ii) and (iii), the eye diagrams of the three received signals are completely closed due to the spatial mode coupling. After being processed by photonic BSS, opened eye diagrams are observed in all three channels. However, we also observe the signal quality is degrading as we move from low-order IC (channel 1) to higher-order IC (channel 3), and the corresponding MSE increases from 0.26% to 1.35%. Table I summarizes the MSE of two-channel (dual-polarization case as discussed earlier) and three-channel BSS. The main source of scaling

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

TABLE II
ENERGY USE IN TRAINING PHASE

Buliding Blocks	DSP	Photonics
PD	$N \times R \times V_{bias} Prect_{training,dsp}$	$N \times R \times V_{bias} Prect_{training,opt}$
ADC	$N \times F_{adc} n_{adc} f_{sample} t_{training,dsp}$	$N \times F_{adc} n_{adc} f_{sample} t_{training,opt}$
MVM	$(N(N-1)E_{op,A} + N^2 E_{op,M}) N_{sample} N_{update}$	$N_{heater} P_{heater} t_{training,opt}$ [a] $N_{Mod} E_{Mod} N_{update}$ [b]
DAC	N.A.	$N_{DAC} F_{dac} n_{dac} \frac{f_{D,sample}}{f_{actuator}} N_{update}$

[a]Heater
[b]P-N modulator

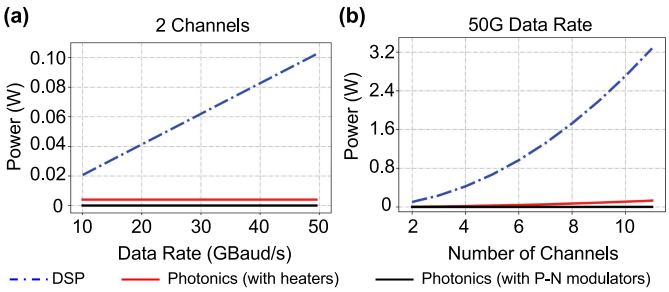


Fig. 10. During the execution phase, the average power in processing each symbol using DSP and photonics. Comparison (a) under different data rates and (b) under different number of channels.

limitation comes from the sequential learning aspect of the proposed photonic BSS approach, in which a higher-order IC is obtained after the lower-order IC is extracted. In this case, the errors generated in the lower-order ICs will accumulate and cascade to the higher-order learning process, limiting the number of channels that can be faithfully processed. To solve this issue, a straightforward solution is to improve the accuracy of individual IC by enforcing a more stringent convergence condition. This, however, comes with the cost of a longer training time and may not be applicable to applications where the mixture matrix is time-varying. Another direction is to find an alternative algorithm that is not based on sequential searching or can reduce the error transfer from lower-order to higher-order ICs. Several global optimization algorithms have been proposed [62]–[65] to solve other searching problems. Our future work will investigate the implementation of the improved optimization algorithm with the proposed photonic hardware.

III. ENERGY AND LATENCY COMPARISON BETWEEN PHOTONICS AND DSP

The photonic BSS system benefits the data center interconnect in two ways: *energy consumption* and *latency*. We are interested in how the two approaches compare when the SDM systems scale to larger data rates and a greater number of channels. The components under evaluation and their corresponding energy consumption and latency are summarized by Table II to Table V in Appendix C. We make our best efforts in including all the possible elements in the two receiver systems (photonic and DSP) in our comparision.

Even although we discuss the DSP performance under different data rate, it should be noted that once a DSP chip has been built for a certain data rate, it cannot self-adapt to a higher data rate. Instead, the DSP chip must be redesigned to include more transistors in order to execute more operations. In contrast, photonics has a flat response to the input signals, so the system doesn't need to be changed as the data rate increases.

We model photonics and DSP in running the same BSS algorithm and compare their energy consumption and latency. Like other machine learning techniques, BSS for channel equalization can be divided into two phases. The first phase is training, during which the transformation matrix needs to be constantly updated until the receiver learns the correct transformation matrix for equalizing the channel crosstalk. During the training stage, we evaluate the energy use $E_{training}$ and training time $t_{training}$ required to learn the transformation matrix. Following training, the system enters the execution phase, during which the receiver applies the transformation matrix to correct distorted signals. The transformation matrix is fixed at the execution phase. We focus on the average power $P_{execution}$ required to perform the transformation matrix to the input signals and the delay required to compute the matrix $t_{latency}$.

A. Training Phase

The training phase consists of two main operations, which includes matrix-vector multiplication (which may be done using photonics or DSP) and variance or kurtosis computation (which is done in the electronic subsystem). As a result, the difference of energy use $E_{training}$ in photonics and DSP is a multiplication of two factors: the energy used in each transformation matrix update and the training time $t_{training}$. Photonics has shown exceptional capabilities in accelerating matrix computing while delivering significant energy savings, particularly when the signal is already analog, so power-intensive and high-speed DACs can be avoided at the system input. However, the DACs that translate the transformation matrix values to the driving voltages of each phase shifter add to the expense of the photonic system.

We first evaluate the training time $t_{training}$ in the photonic and DSP systems, respectively. We consider the best scenario for DSP where DSP hardware can always operate at line rate speed via massive parallelism, so the delay in serial to parallel (s/p) conversion is included. The training time is spent on matrix multiplication, moment fittings, and s/p conversions

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

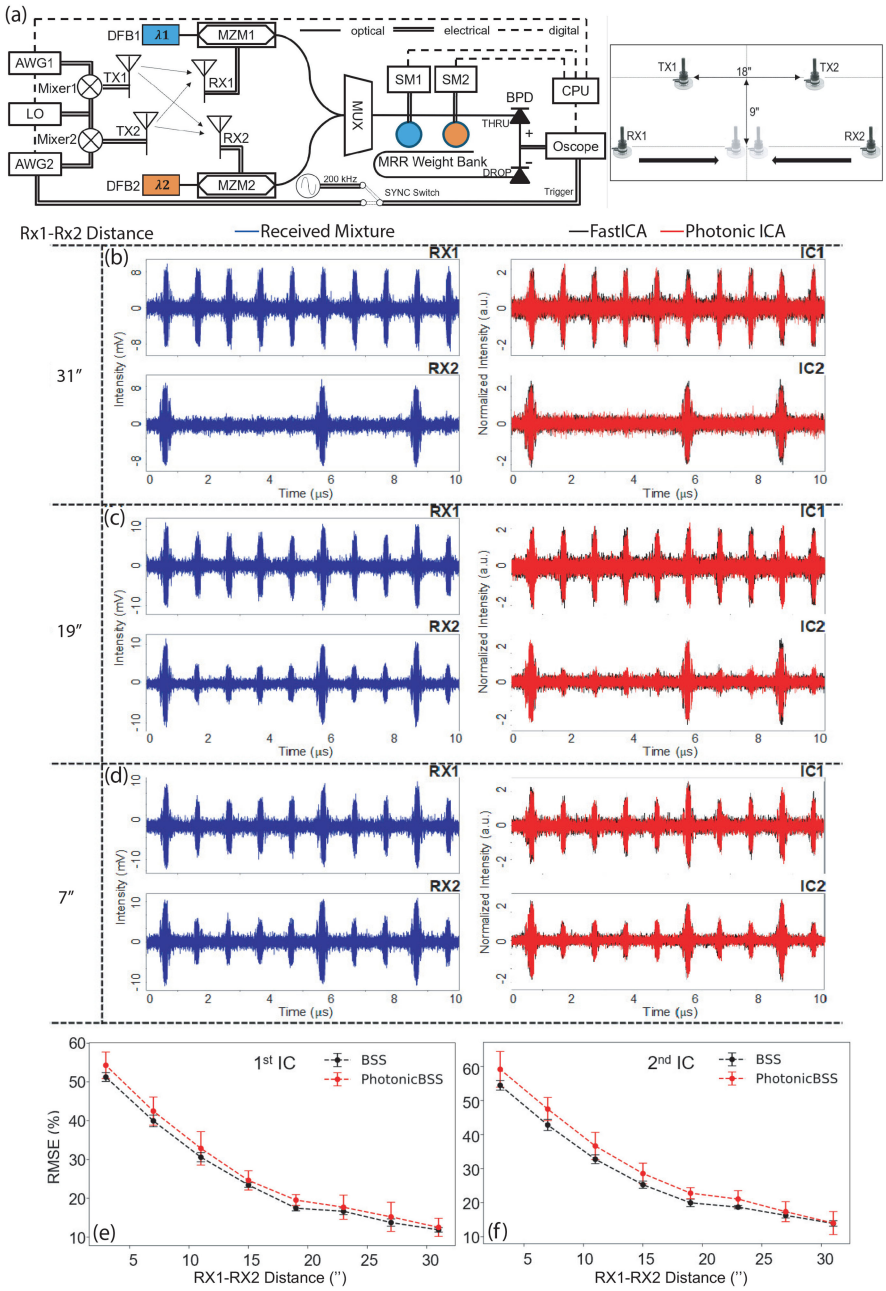


Fig. 11. Experimental results of wireless MIMO system. (a) Experimental setup of photonic BSS for wireless communications. (b)(c)(d) Signal waveforms with distance 31 inches, 19 inches, and 9 inches, respectively. blue line: received mixture signals, black line: simulation waveforms after BSS, red line: experimental waveforms after photonic BSS. and (e)(f) RMSE under different receiver distance.

(see Table IV). Using photonics, matrix multiplication can be done at a single time step, so its speed is limited only by the bandwidth of photodetectors, which is only tens of picoseconds. In this case, the training time $t_{training}$ is determined by $\max\left[1/B_{pd}, 1/f_{ADC,sampling}\right]$, where f_{sample} is the sampling rate of ADC. BSS can operate at any sampling rate; here we make $f_{sample} = 10$˜GHz as a tradeoff between energy consumption and convergence speed. The training time of DSP, on the other hand, is bounded by its clock frequency f_{clock} which is set a few hundreds of MHz in most DSP systems. The results of the training time comparison are shown in Fig. 7. We assume the channel number is two in Fig. 7(a), and the results show that the training time remains constant when the signal data rate increases, owing to the BSS statistic feature. The photonic system reduces the training time by $\sim$90 μs, as a result of its high-speed matrix multiplier. We then evaluate the training time as a function of the number of channels and fix the signal data rate per channel to 50 GBaud/s. Since the Nelder-Mead method finds out the IC/PCs channel by channel, so the number of overall iterations increases in proportion to the number of channels. In addition, the operations required to computing an N-channel matrix grow quadratically with the channel number. As a result, as shown in Fig. 7(b), the training time of DSP increases rapidly with the channel number and becomes too

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

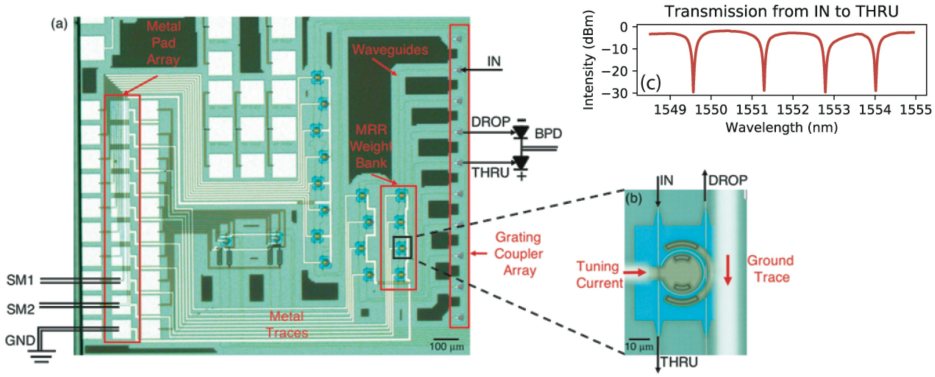


Fig. 12. (a) Optical image of silicon microring (MRR) weight bank with four MRRs. (b) Zoom-in of an MRR. (c) Transmission from the IN port to the THRU port. Two of the total four MRRs are used to implement BSS of two received signals, with their resonance locating around 1549.6 nm and 1554 nm, respectively. On top of these MRRs are dedicated metal heaters, which allow for independent tuning of each MRR. All four MRRs share a common ground, as do two additional signaling traces. As a result, three DC pads are used and wire-bonded for testing.

long to capture the environment vibrations. In contrast, the relation between the training time of the photonic system and the number of channels is approximately linear. As we shall see later, the shortened training period in the photonic system will also minimize energy consumption during the training phase. The equations and parameters used in the calculations are shown in Appendix C.

We then evaluate the energy use in the training phase $E_{training}$. Fig. 9(a) shows how energy changes at different data rates in the training process. In DSP, the energy depends on the training time derived above and the total number of operations during the training phase. The energy of individual operations is closely linked to the CMOS process technology. Here we assume the DSP employs the cutting-edge 7 nm process. Using photonics, the energy cost comes from the phase shifters and the DACs used to control the phase shifters. The energy use of individual phase shifters depends on the type of tuning mechanisms. Here we analyze two mechanisms, micro-heaters and PN junction-based modulators. The primary distinction is that micro-heaters waste energy even while the actuation voltages stay constant, but PN modulators, being a kind of capacitive device, have virtually little static power. The same DSP hardware is used to compute the statistical moments in the photonic BSS, and its energy consumption is also included in our study (see Appendix C, Table II). As shown in Fig. 9, because the actual sample rate may stay constant owing to BSS, the energy consumption throughout training period has a flat response to the data rate. The photonic system with heaters and PN-modulators has similar energy use because the ADCs dominate overall energy consumption. As the channel number rises, the total number of operations grows proportional to the third power of the channel number. As a result, we observe two orders of magnitude energy reduction in the photonic systems at large channel numbers.

B. Execution Phase

The main operation in the execution phase is to execute matrix-vector multiplications on the input signals using the transformation matrix learned during training. Similarly to the training phase, we begin by calculating the time delay $t_{latency}$ in processing each data symbol in the incoming data stream. In DSP, as the data rate increases, more s/p conversions are required to produce more parallel operations in order to compute the matrix at the signal's line rate. This comes at the expense of extra $t_{latency}$ delay, as shown in Fig. 8. In comparison, $t_{latency}$ in phonics systems is independent to the signal data rate. For the similar reason, $t_{latency}$ grows quadratically with the channel number N in DSP, while $t_{latency}$ remains a constant in the photonic system as N increases.

We now consider the average power consumption $P_{execution}$ in different types of systems during the execution phase when a single matrix-vector operation, consisting of N input channels and N output channels with a fixed transformation matrix, is calculated. In DSP, the average power is a function of multiplying number of logic operations in a unit time by the energy per operation. As a result, the power consumption of DSP grows approximately linear with the signal data rate and quadratically with the channel number, as shown in Fig. 10. The photonic system has a completely different picture. When the transformation matrix becomes fixed, the DACs operate in static mode, consuming very little static power. PN modulator-based actuators, like DACs, have minimal power consumption. As a result, using PN modulators as actuators, the processor's power consumption is nearly zero, independent of the signal data rate or channel number. On the other hand, micro-heaters have a constant power, and the overall power consumption does not vary with the signal data rate, but increases quadratically with the channel number. Nonetheless, the power consumption is still much lower than that of DSP.

C. Remarks About Photonics and DSP Comparison

In contrast to DSP, the photonic BSS front-end has a significantly more favorable scaling law for SDM-based optical fiber interconnects, which is especially important for high-throughput interconnects with high data rates per channel and large number of channels. For example, in a 1 Tbit/s optical interconnect consisting of 10 spatial channels and 50 Gbaud PAM4 per channel, photonics would reduce the training time by closely

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

TABLE III
AVERAGE POWER IN EXECUTION PHASE (WE ONLY CALCULATE THE POWER USED IN SIGNAL PROCESSING WHILE IGNORING THE POWER USED IN PD AND ADC)

Buliding Blocks	**DSP**	**Photonics**
MVM	$(N(N-1)E_{op,A}+N^2E_{op,M})B_{data}$	$N_{heater}P_{heater}$ [a] ~ 0 [b]
DAC	N.A.	$N_{DAC}P_{dac,static}$

[a]Heater
[b]P-N modulator

TABLE IV
TRAINING TIME $t_{training}$ (DEFINED AS THE TIME TO FIND THE TRANSFORMATION MATRIX)

Buliding Blocks	**DSP**	**Photonics**
MVM	$2N^2N_{sample}N_{update}/f_{sample}$	$N_{sample}N_{update}/f_{sample}$
Moment fitting	$(3N_{sample}N_{PCA}+7N_{sample}N_{ICA})/f_{sample}$	$(3N_{sample}N_{PCA}+7N_{sample}N_{ICA})/f_{sample}$
s/p conversion	f_{sample}/f_{clock}^2	f_{sample}/f_{clock}^2

TABLE V
EXECUTION LATENCY ($T_{latency}$)

Buliding Blocks	**DSP**	**Photonics**
MVM	$2N^2/f_{clock}$	$1/B_{PD}$
s/p conversion	$2*B_{data}/f_{clock}^2$	0

to 30 times from 69 ms to 2.3 ms and reduce the energy use of training by 30 times from 140 mJ to 4.6 mJ. In the execution phase, photonics lowers latency by 32 times, from 0.8 μs to 25 ps, and average power by 270 times, from 2.7 W to 1 mW.

IV. PRIVACY PROTECTION FOR WIRELESS MIMO AND COGNITIVE RADIO

This section reviews the application and experimental demonstrations of photonic BSS for wireless communications [33]–[35]. The accelerating demands on the efficient utilization of spectrum resources are pushing the implementation of cognitive radio, in which a transceiver can intelligently detect the spectrum usage and moves into a vacant spectrum while avoiding the used spectrum. However, cognitive ratio requires monitoring opportunistic users in real-time, which is challenging in MIMO systems due to the existence of a large number of unpredictability and transience of time-frequency transmission blocks. Fortunately, BSS, which can distill the salient information sharing the same spectrum and time, can effectively address potential spectrum collision problems. In addition, BSS is agnostic to signal format and, most importantly, blind to the content of transmissions, which makes BSS outstanding in protecting users' privacy [66].

While the conventional radio-frequency (RF) electronics for wireless communication systems [67], has dominated for decades, they could hardly offer the broadband frequency agility, which is a prerequisite for achieving the full potential of BSS. And the RF electronics have to come up with burdensome solutions which require large banks of sub-circuits (each designed for a specific frequency band) for such a broad spectrum coverage [68], and sophisticated switching technology to ensure smooth transitions across frequency bands. These practical difficulties motivate photonic BSS that provides a completely different set of signal processing properties [69].

Here we experimentally showcase using silicon photonic BSS chip for wireless MIMO communications. The experimental setup is shown in Fig. 11(a). The system consists of a photonic RF front-end for a 2×2 wireless MIMO system. The two channel wireless systems carry signals with two modulation formats: Binary Phase Shift Keying (BPSK) and Amplitude Shift Keying (ASK) respectively. In this system, two optical modulators first convert the received wireless mixtures onto optical carriers with different optical wavelengths. Then the two optical signals are coupled into a silicon photonic device fabricated on a silicon-on-insulator (SOI) wafer through optical grating couplers. The silicon photonic device is an MRR weight bank composed of two MRRs, each having its resonances aligned with the wavelengths of two optical carriers, as shown in Fig. 12. In addition, each MRR embeds an in-ring N-doped photoconductive heater to realize precise MRR control [56]–[58]. The MRR weight bank provides matrix operations to the input signals in the optical domain, and then the signals are converted to electrical format by the balanced photodetector. By effectively upconverting the wireless signals to a 193 THz intermediate frequency, photonic processors are nearly flat frequency response. Even gigahertz wireless signals (e.g., millimeter-wave

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
MODELING PARAMETERS

Symbol	Quantity	Value
N	Signal channel number	N
N_{heater}	Heater number	N^2+N-2 [38]
N_{Mod}	P-N modulator number	N^2+N-2
N_{DAC}	DAC number	N^2+N-2
N_{sample}	Sample number for statistic analysis per weight update	600
$f_{actuator}$	Change frequency of MZI mesh	f_{sample}/N_{sample}
B_{data}	Signal baud rate	50 GBaud/s
f_{sample}	Sample rate of ADC in a free-running mode	10 GHz
F_{dac}	DAC figures of merit	1.25×10^{-12} J/conv-step
n_{dac}	DAC resolution	8
F_{adc}	ADC figures of merit	2.5×10^{-12} J/conv-step
n_{adc}	ADC resolution	8
N_{update}	Number of weight update for training	223 (for $N=2$) 223*($N^2/4$) (for $N>2$)
N_{PCA}	Number of weight update for training of PCA	71 (for $N=2$) 71*($N^2/4$) (for $N>2$)
N_{ICA}	Number of weight update for training of ICA	152 (for $N=2$) 152*($N^2/4$) (for $N>2$)
$F_{A,sample}$	Sample frequency at ADC	$n_{sa}B_{data}$
n_{sa}	Oversampling ratio	2
P_{heater}	Heater power	1 mW
$P_{dac,static}$	Static power of DAC	10 μW [72]
E_{Mod}	Energy consumed to change the voltage on the modulator	$C_{Mod}V_r^2$
C_{Mod}	Modulator capacitance	30 fF
V_r	Modulator voltage swing	4.8 V
$E_{op,A}$	Average energy per n_b-bit adder op.	$2.57n_bp_tV^2$ fJ
$E_{op,M}$	Average energy per n_b-bit multiplexer op.	$2.57n_b^2p_tV^2$ fJ
n_b	Average bit resolution of the DSP module	8
p_t	CMOS process technology feature size	7 nm
V	CMOS supply voltage	0.65 V
R	Photodiode responsivity	1 A/W
V_{bias}	Photodiode bias voltage	3.3 V
P_{rec}	Received optical power	0 dBm
B_{Mod}	Modulator bandwidth	40 GHz
B_{PD}	Photodiode bandwidth	40 GHz
f_{clock}	Clock frequency of DSP	500 MHz

signals) are considered narrowband on the optical spectrum since this optical path gives a flat frequency response over a 5 THz window [70]. Between the E/O modulation and O/E detection, optical signals can be processed with tunable optical devices, which may benefit from both the unconventional utilization of standard photonic integrated devices and the enormous information density made possible through wavelength-division multiplexing (WDM) [18], [30].

Fig. 11(b) presents the experimental results of two-channel photonic BSS, in which we used the photonic weight bank to find the PC/IC vectors and then applied the vectors at the RF photonic front-end to recover the sources. The experiment was done when the antennas were configured at different locations to evaluate the system's performance at different transmission conditions. To do so, we fixed the transmitter antennas and moved to adjust the distance between the receiver antennas. By doing so, we could test this photonic BSS system dealing with mixtures with different mixing ratios. As shown in Fig. 11(b)–(d), the top mixture (RX1) is received by the antenna that is closer to the BPSK transmitter, while the bottom mixture (RX2) is received by the antenna that is closer to the ASK transmitter. The photonic BSS pipeline demonstrates its effectiveness by successfully separating these two sources at a distance of 31," resulting in the top estimated source (IC1) being the BPSK and the bottom estimated source (IC2) being the ASK. The other two trails, which represent the distance of 19" (Fig. 11(c)) and 17" (Fig. 11(d)) experience more difficulty in source estimation because the mixing ratios are very close to 0.5. However, we demonstrated that photonic BSS still manages to achieve perceivably effective source separation.

V. CONCLUSION

Because of the intrinsic scaling constraints of digital systems, DSP has faced significant energy and speed challenges as the data volumes required in communication systems has increased rapidly. Photonics provides a solution, alleviating the energy use and processing delays in matrix-vector multiplications, an operation ubiquitously used in different signal processing algorithms. In this work, we introduce a novel photonic front-end that synthesizes integrated silicon photonic devices with a novel BSS algorithm. We demonstrate its various applications in space-division multiplexing based communication systems, which includes addressing the modal crosstalk in multi-mode optical fiber interconnect in data centers and preserving user privacy for wireless MIMO and cognitive radios, enabled by the "blindness" in BSS. We extensively compare our photonic system with DSP in terms of energy use and processing latency, demonstrating that photonics has distinct benefits at high data rates and large channel numbers. The findings indicate that photonics can solve growing constraints in DSP and pave ways for future high-speed, low-energy communication systems.

ACKNOWLEDGMENT

The authors would like to thank Prof. HonKi Tsang for his fruitful discussions.

APPENDIX A
BSS ALGORITHM

A. Constrained Nelder–Mead Search

The constrained N-M method aims to find the minimum or maximum of an objective function in a multidimensional space. Here we make all the weight vectors $\boldsymbol{w_i}$ have unit norm [35] to make it implementable with photonic devices. Nelder–Mead search is applied to both PCA and ICA to find the maximum second- and fourth-moment, respectively, by searching for the updated vectors $\boldsymbol{w_{max}}$ and $\boldsymbol{w_{min}}$. Nelder–Mead search first initializes $N+1$ random weight vectors distributed on a unit

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

N-dimension space. We define the centroid of all weight vectors as $\boldsymbol{w_{cen}} = \frac{1}{N}\sum_{k=1}^{n} \boldsymbol{w_i}$, and define unit vectors $\boldsymbol{e_{max}}$, $\boldsymbol{e_{min}}$ representing the 1st and Nth PC (or IC) vector directions, respectively. Then, the weight vector $\boldsymbol{w_{min}}$ associated with the minimum output variance σ^2 (or kurtosis $kurt$) is updated until all the weight vectors pointing in nearly the same direction at which point the algorithm is converged.

Each iteration consists of four type of main operations: *reflection, expansion, contraction, and shrinkage*. Reflection improves $\boldsymbol{w_{min}}$ by simply reversing its direction with respect to $\boldsymbol{w_{cen}}$. We get many $\boldsymbol{w_{min}^{new}}$ in the directions of continuous reflections $\boldsymbol{w_{ref}} = \boldsymbol{w_{cen}} + \beta(\boldsymbol{w_{cen}} - \boldsymbol{w_{min}})$, which can make w_i converge into a cone near $\boldsymbol{e_{max}}$. If reflection does not lead to a better weight vector direction than current w_{min} in some special cases, we need to continue apply expansion, contraction, and shrinkage operations. Expansion $\boldsymbol{w_{exp}} = \boldsymbol{w_{cen}} + \gamma(\boldsymbol{w_{ref}} - \boldsymbol{w_{cen}})$ along the reflection direction may reach an even better $\boldsymbol{w_{min}^{new}}$ direction. Contraction $\boldsymbol{w_{con}} = \boldsymbol{w_{cen}} + \rho(\boldsymbol{w_{min}} - \boldsymbol{w_{cen}})$ can instead find a better alternative $\boldsymbol{w_{min}^{new}}$ direction when reflection does not lead to a better weight vector direction than current $\boldsymbol{w_{min}}$. Finally, shrinkage with $\boldsymbol{w_i} = \boldsymbol{w_{max}} + \sigma(\boldsymbol{w_i} - \boldsymbol{w_{max}})$ often has an impact when $\boldsymbol{w_{cen}}$ is negligible, it will change $\boldsymbol{w_i}$ to further create better $\boldsymbol{w_{min}^{new}}$ direction. In addition, standard values of above coefficients are $\beta = 1$, $\gamma = 2$, $\rho = 0.5$, $\sigma = 0.5$, respectively. Good convergence is achieved when the different between final $\boldsymbol{w_{max}}$ and $\boldsymbol{w_{min}}$ is smaller than convergence tolerance ϵ.

B. PCA Algorithm

Input: Weighted addition output y, 1^{st},..., $(k-1)^{th}$ PC vectors e_1,..., e_{k-1};

Parameters: Convergence tolerance ϵ, reflection coefficient β, expansion coefficient γ, contraction coefficient ρ, shrinkage coefficient σ;

Output: k_{th} PC vector e_k;

Initialize n random weight vectors $\boldsymbol{w_i}(i = 1, \ldots, n)$;
for $i = 1 : n$ **do**
  if $k > 1$ **then**
    Orthogonalize $\boldsymbol{w_i}$ such that $\boldsymbol{w_i} = \boldsymbol{w_i} - \sum_{j=1}^{k-1}(\boldsymbol{w_i}^T \boldsymbol{e_j})\boldsymbol{e_j}$;
  end if
  Normalize $\boldsymbol{w_i}$ such that $|\boldsymbol{w_i}| = 1$;
  Apply $\boldsymbol{w_i}$ to the weight bank and obtain the resulted $\boldsymbol{y}(\boldsymbol{w_i})$;
  Compute the variance of $\boldsymbol{y}$: $f(\boldsymbol{w_i}) = E(\boldsymbol{y}^2)$;
end for
Among $f(\boldsymbol{w_i})(i = 1, \ldots, n)$, find $f_{max} = max(f(\boldsymbol{w_i}))$, $f_{min} = min(f(\boldsymbol{w_i}))$;
Set $\boldsymbol{w_{max}} = argmax(f(\boldsymbol{w_i}))$, $\boldsymbol{w_{min}} = argmin(f(\boldsymbol{w_i}))$;
if $|f_{max} - f_{min}| \leq \epsilon$ **then**
  return $e_k = w_{max}$;
end if
while $|f_{max} - fmin| > \epsilon$ **do**
  Compute the centroid of all weight vectors except $\boldsymbol{w_{min}}$: $\boldsymbol{w_{cen}} = E(\boldsymbol{w_i})$(for $\boldsymbol{w_i} \neq \boldsymbol{w_{min}}$);
  Compute the reflected weight vector: $\boldsymbol{w_{ref}} = \boldsymbol{w_{cen}} + \beta(\boldsymbol{w_{cen}} - \boldsymbol{w_{min}})$;
  Normalize $\boldsymbol{w_{ref}}$ such that $|\boldsymbol{w_{ref}}| = 1$;
  if $f_{min} < f(\boldsymbol{w_{ref}}) < f_{max}$ **then**
    $\boldsymbol{w_{min}} = \boldsymbol{w_{ref}}$ and update f_{min};
  else if $f(\boldsymbol{w_{ref}}) \geq f_{max}$ **then**
    Compute the expanded weight vector: $\boldsymbol{w_{exp}} = \boldsymbol{w_{cen}} + \gamma(\boldsymbol{w_{ref}} - \boldsymbol{w_{cen}})$;
    Normalize $\boldsymbol{w_{exp}}$ such that $|\boldsymbol{w_{exp}}| = 1$;
    if $f(\boldsymbol{w_{ref}}) < f(\boldsymbol{w_{exp}})$ **then**
      $\boldsymbol{w_{max}} = \boldsymbol{w_{exp}}$ and update $f_{max} = f(\boldsymbol{w_{exp}})$;
    else
      $\boldsymbol{w_{max}} = \boldsymbol{w_{ref}}$ and update $f_{max} = f(\boldsymbol{w_{ref}})$;
      if $f(\boldsymbol{w_{exp}}) > f_{min}$ **then**
        $\boldsymbol{w_{min}} = \boldsymbol{w_{exp}}$ and update f_{min}, $\boldsymbol{w_{min}}$;
      end if
    end if
  else if $f(\boldsymbol{w_{ref}}) \leq f_{min}$ **then**
    Compute the contracted weight vector: $\boldsymbol{w_{con}} = \boldsymbol{w_{cen}} + \rho(\boldsymbol{w_{min}} - \boldsymbol{w_{cen}})$;
    Normalize $\boldsymbol{w_{con}}$ such that $|\boldsymbol{w_{con}}| = 1$;
    if $f_{min} < f(\boldsymbol{w_{con}}) < f_(max)$ **then**
      $\boldsymbol{w_{min}} = \boldsymbol{w_{con}}$ and update f_{min}
    else if $f(\boldsymbol{w_{con}}) > f_{max}$ then
      $\boldsymbol{w_{max}} = \boldsymbol{w_{con}}$ and update f_{max}
    else
      for $i = 1 : n$ **do**
        Replace all weight vectors with $\boldsymbol{w_i} = \boldsymbol{w_{max}} + \sigma(\boldsymbol{w_i} - \boldsymbol{w_{max}})$;
        Normalize w_i such that $|\boldsymbol{w_i}| = 1$;
      end for
      Update f_{max}, f_{min}, and $\boldsymbol{w_{max}}$, $\boldsymbol{w_{min}}$;
    end if
  end if
end while
return $\boldsymbol{e_k} = \boldsymbol{w_{max}}$.

C. ICA Algorithm

Input: Weighted addition output y, 1^{st},..., $(k-1)^{th}$ IC vectors e_1,..., e_{k-1}; whitening matrix $/bmV$;

Parameters: Convergence tolerance ϵ, reflection coefficient β, expansion coefficient γ, contraction coefficient ρ, shrinkage coefficient σ;

Output: k_{th} IC vector e_k

Initialize n random weight vectors $\boldsymbol{w_i}(i = 1, \ldots, n)$;
for $i = 1 : n$ **do**
  if $k > 1$ **then**
    Orthogonalize $\boldsymbol{w_i}$ such that $\boldsymbol{w_i} = \boldsymbol{w_i} - \sum_{j=1}^{k-1}(\boldsymbol{w_i}^T \boldsymbol{e_j})\boldsymbol{e_j}$;
  end if
  Normalize w_i such that $|\boldsymbol{w_i}| = 1$;
  compute the normalized whitened weight vectors $\boldsymbol{v_i} = \boldsymbol{V}\boldsymbol{w_i}$;
  apply $\boldsymbol{v_i}$ to the weight bank and obtain the resulted $\boldsymbol{y}(\boldsymbol{v_i})$;
  compute the objective function, i.e., the kurtosis of $\boldsymbol{y}$: $f(\boldsymbol{w_i}) = kurt(\boldsymbol{y}^4(\boldsymbol{v_i}))$;

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

end for
Among $f(\boldsymbol{w_i})(i = 1, \ldots, n)$, find $f_{max} = max(f(\boldsymbol{w_i}))$, $f_{min} = min(f(\boldsymbol{w_i}))$;
  Set $\boldsymbol{w_{max}} = argmax(f(\boldsymbol{w_i}))$, $\boldsymbol{w_{min}} = argmin(f(\boldsymbol{w_i}))$;
if $|f_{max} - f_{min}| \leq \epsilon$ **then**
  return $\boldsymbol{e_1} = \boldsymbol{w_{max}}$, $\boldsymbol{e_2} = \boldsymbol{e_1^T}$;
end if
while $|f_{max} - fmin| > \epsilon$ **do**
  Compute the centroid of all weight vectors: $\boldsymbol{w_{cen}} = E(\boldsymbol{w_i})$;
  Compute the normalized reflected weight vector: $\boldsymbol{w_{ref}} = \boldsymbol{w_{cen}} + \beta(\boldsymbol{w_{cen}} - \boldsymbol{w_{min}})$;
  Compute the normalized whitened reflected weight vector: $\boldsymbol{V_{ref}} = \boldsymbol{V}\boldsymbol{w_{ref}}$
  if $f_{min} < f(\boldsymbol{w_{ref}}) = kurt(\boldsymbol{y}^4(\boldsymbol{v_{ref}})) < f_{max}$ **then**
    $\boldsymbol{w_{min}} = \boldsymbol{w_{ref}}$ and update f_{min};
  else if $f(\boldsymbol{w_{ref}}) \geq f_{max}$ **then**
    Compute the whitened expanded weight vector: $\boldsymbol{w_{exp}} = \boldsymbol{w_{cen}} + \gamma(\boldsymbol{w_{ref}} - \boldsymbol{w_{cen}})$;
    Compute the expanded weight vector: $\boldsymbol{V_{exp}} = \boldsymbol{V}\boldsymbol{w_{exp}}$
    if $f(\boldsymbol{w_{ref}}) < f(\boldsymbol{w_{exp}}) = kurt(\boldsymbol{y}^4(\boldsymbol{v_{exp}}))$ **then**
      $\boldsymbol{w_{max}} = \boldsymbol{w_{exp}}$ and update $f_{max} = f(\boldsymbol{w_{exp}})$;
    else
      $\boldsymbol{w_{max}} = \boldsymbol{w_{ref}}$ and update $f_{max} = f(\boldsymbol{w_{ref}})$;
      if $f(\boldsymbol{w_{exp}}) > f_{min}$ **then**
        $\boldsymbol{w_{min}} = \boldsymbol{w_{exp}}$ and update f_{min}, $\boldsymbol{w_{min}}$;
      end if
    end if
  else if $f(\boldsymbol{w_{ref}}) \leq f_{min}$ **then**
    Compute the contracted weight vector: $\boldsymbol{w_{con}} = \boldsymbol{w_{cen}} + \rho(\boldsymbol{w_{min}} - \boldsymbol{w_{cen}})$;
    Compute the whitened contracted weight vector: $\boldsymbol{V_{con}} = \boldsymbol{V}\boldsymbol{w_{con}}$
    if $f_{min} < f(\boldsymbol{w_{con}}) = kurt(\boldsymbol{y}^4(\boldsymbol{v_{con}})) < f_{max})$ **then**
      $\boldsymbol{w_{min}} = \boldsymbol{w_{con}}$ and update f_{min}
    else if $f(\boldsymbol{w_{con}}) = kurt(\boldsymbol{y}^4(\boldsymbol{v_{con}})) > f_{max}$ then
      $\boldsymbol{w_{max}} = \boldsymbol{w_{con}}$ and update f_{max}
    else
      for $i = 1 : n$ **do**
        Replace all weight vectors with $\boldsymbol{w_i} = \boldsymbol{w_{max}} + \sigma(\boldsymbol{w_i} - \boldsymbol{w_{max}})$;
        Compute the whitened weight vectors: $\boldsymbol{V_i} = \boldsymbol{V}\boldsymbol{w_i}$
      end for
      Update f_{max}, f_{min}, and $\boldsymbol{w_{max}}$, $\boldsymbol{w_{min}}$;
    end if
  end if
end while
return $\boldsymbol{e_k} = \boldsymbol{w_{max}}$.

APPENDIX B
Q-FACTOR CALCULATION

The Q-factor is calculated using equations 6,7, and 8 [61]: where μ_i is the mean amplitude of the received symbols at level i, and σ_i^2 is their variance.

APPENDIX C
COMPARISONS OF PHOTONICS AND DSP

Here, we model photonic and DSP systems using the formulas in Table II to V. The four Tables show the time and energy used by each building block at various system phases, providing the scaling law of photonics and DSP as the system data rate and channel count rise. The detailed modeling parameters can be found in Table VI. Below, we explore 1. energy consumption and training time when training the photonic BSS model; 2. power consumption and latency when executing photonic BSS tasks. By modeling the processes, excellent performances can be found in photonics BSS over the DSP counterpart. We show the comparison results in Section III.

REFERENCES

[1] Cisco, "Cisco annual internet report," *White Paper*, 2020. Accessed: Apr. 23, 2020. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html

[2] R.-J. Essiambre, R. Ryf, N. Fontaine, and S. Randel, "Breakthroughs in photonics 2012: Space-division multiplexing in multimode and multicore fibers for high-capacity optical communication," *IEEE Photon. J.*, vol. 5, no. 2, Apr. 2013, Art. no.0701307.

[3] G. Li, N. Bai, N. Zhao, and C. Xia, "Space-division multiplexing: The next frontier in optical communication," *Adv. Opt. Photon.*, vol. 6, no. 4, pp. 413–487, 2014.

[4] A. R. Shah, R. C. Hsu, A. Tarighat, A. H. Sayed, and B. Jalali, "Coherent optical MIMO (COMIMO)," *J. Lightw. Technol.*, vol. 23, no. 8, pp. 2410–2419, 2005.

[5] X. Wu, C. Huang, K. Xu, C. Shu, and H. K. Tsang, "Mode-division multiplexing for silicon photonic network-on-chip," *J. Lightw. Technol.*, vol. 35, no. 15, pp. 3223–3228, 2017.

[6] P. J. Winzer and G. J. Foschini, "MIMO capacities and outage probabilities in spatially multiplexed optical transport systems," *Opt. Exp.*, vol. 19, no. 17, pp. 16680–16696, 2011.

[7] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[8] D. Gesbert, M. Shafi, D. S. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 281–302, Apr. 2003.

$$I_{i-high,low}^{th} = \frac{\mu_i\sigma_{i+1,i-1}^2 - \mu_{i+1,i-1}\sigma_i^2 + \sigma_{i+1}\sigma i\sqrt{(\mu_{i+1,i-1} - \mu_i)^2 + 2(\sigma_{i+1,i-1}^2 - \sigma_i^2)ln(\sigma_{i+1,i-1}/\sigma_i)}}{\sigma_{i+1,i-1}^2 - \sigma_i^2} \tag{6}$$

$$BER = \frac{1}{N\log_2(N)}\frac{1}{2}\sum_{i=1}^{N}\left(erfc\left(\frac{|\mu_i - I_{i-low}^{th}|}{\sqrt{2}\sigma_i}\right) + erfc\left(\frac{|\mu_i - I_{i-high}^{th}|}{\sqrt{2}\sigma_i}\right)\right. \tag{7}$$

$$Q = 20\log_{10}(N\sqrt{2}erfcinv(2BER)) \tag{8}$$

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

[9] Å. Björck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA, USA: SIAM, 1996.

[10] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Hoboken, NJ, USA: Prentice-Hall, 1993.

[11] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, Mar. 2006.

[12] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Netw.*, vol. 13, no. 4/5, pp. 411–430, 2000.

[13] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[14] J. Mo and R. W. Heath, "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Proc. Inf. Theory Appl. Workshop*, 2014, pp. 1–5.

[15] 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, 2014.

[16] M. A. Nahmias, T. F. De Lima, A. N. Tait, H.-T. Peng, B. J. Shastri, and P. R.Prucnal, "Photonic multiply-accumulate operations for neural networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–18, Jan./Feb. 2020.

[17] B. J. Shastri *et al.*, "Photonics for artificial intelligence and neuromorphic computing," *Nature Photon.*, vol. 15, no. 2, pp. 102–114, 2021.

[18] C. Huang *et al.*, "Prospects and applications of photonic neural networks," *Adv. Physics: X*, vol. 7, no. 1, pp. 1981155, Jan. 2022.

[19] "Building the adaptive network starting with silicon - Ciena." [Online]. Available: http://www.ciena.com/insights/articles/building-the-adaptive-network--starting-with-silicon-coherent-dsp-design.html

[20] A. N. Tait *et al.*, "Demonstration of multivariate photonics: Blind dimensionality reduction with integrated photonics," *J. Lightw. Technol.*, vol. 37, no. 24, pp. 5996–6006, 2019.

[21] J. Capmany and D. Novak, "Microwave photonics combines two worlds," *Nature Photon.*, vol. 1, no. 6, 2007, Art. no. 319.

[22] C. Huang *et al.*, "A silicon photonic-electronic neural network for fibre nonlinearity compensation," *Nature Electron.*, vol. 4, pp. 837–844, 2021.

[23] Y. Shen *et al.*, "Deep learning with coherent nanophotonic circuits," *Nature Photon.*, vol. 11, no. 7, pp. 441–446, 2017.

[24] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, 2019.

[25] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, "Linear optical quantum computing with photonic qubits," *Rev. Modern Phys.*, vol. 79, no. 1, 2007, Art. no. 135.

[26] T. Rudolph, "Why I am optimistic about the silicon-photonic route to quantum computing," *APL Photon.*, vol. 2, no. 3, 2017, Art. no. 030901.

[27] X. Qiang *et al.*, "Large-scale silicon quantum photonics implementing arbitrary two-qubit processing," *Nature Photon.*, vol. 12, no. 9, pp. 534–539, 2018.

[28] M. Prabhu *et al.*, "A recurrent ising machine in a photonic integrated circuit," 2019, *arXiv:1909.13877*.

[29] P. L. McMahon *et al.*, "A fully programmable 100-spin coherent ising machine with all-to-all connections," *Science*, vol. 354, no. 6312, pp. 614–617, 2016.

[30] 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. 4029–4041, 2014.

[31] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," *Phys. Rev. Lett.*, vol. 73, no. 1, 1994, Art. no. 58.

[32] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, "Optimal design for universal multiport interferometers," *Optica*, vol. 3, no. 12, pp. 1460–1465, 2016.

[33] P. Y. Ma *et al.*, "Blind source separation with integrated photonics and reduced dimensional statistics," *Opt. Lett.*, vol. 45, no. 23, pp. 6494–6497, 2020.

[34] A. N. Tait *et al.*, "Blind source separation in the physical layer," in *Proc. 52nd Annu. Conf. Inf. Sci. Syst.*, 2018, pp. 1–6.

[35] P. Y. Ma *et al.*, "Photonic independent component analysis using an on-chip microring weight bank," *Opt. Exp.*, vol. 28, no. 2, pp. 1827–1844, 2020.

[36] N. Jones, "How to stop data centres from gobbling up the world's electricity," *Nature*, vol. 561, no. 7722, pp. 163–167, 2018.

[37] K. Shibahara *et al.*, "DMD-unmanaged long-haul SDM transmission over 2500-km 12-core× 3-mode MC-FMF and 6300-km 3-mode FMF employing intermodal interference canceling technique," *J. Lightw. Technol.*, vol. 37, no. 1, pp. 138–147, 2018.

[38] K. Choutagunta, I. Roberts, D. A. Miller, and J. M. Kahn, "Adapting Mach-Zehnder mesh equalizers in direct-detection mode-division-multiplexed links," *J. Lightw. Technol.*, vol. 38, no. 4, pp. 723–735, 2019.

[39] K. Choutagunta and J. M. Kahn, "Dynamic channel modeling for mode-division multiplexing," *J. Lightw. Technol.*, vol. 35, no. 12, pp. 2451–2463, 2017.

[40] J. Wang, S. He, and D. Dai, "On-chip silicon 8-channel hybrid (de) multiplexer enabling simultaneous mode-and polarization-division-multiplexing," *Laser Photon. Rev.*, vol. 8, no. 2, pp. L18–L22, 2014.

[41] A. Annoni *et al.*, "Unscrambling light-automatically undoing strong mixing between modes," *Light: Sci. Appl.*, vol. 6, no. 12, 2017, Art. no.e17110.

[42] Z. Luo, C. Li, and L. Zhu, "A comprehensive survey on blind source separation for wireless adaptive processing: Principles, perspectives, challenges and new research directions," *IEEE Access*, vol. 6, pp. 66685–66708, 2018.

[43] P. Comon, "Independent component analysis, A new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[44] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, no. 4, pp. 308–313, 1965.

[45] Q. Xu and R. Soref, "Reconfigurable optical directed-logic circuits using microresonator-based optical switches," *Opt. Exp.*, vol. 19, no. 6, pp. 5244–5259, 2011.

[46] W. Bogaerts *et al.*, "Programmable photonic circuits," *Nature*, vol. 586, no. 7828, pp. 207–216, 2020.

[47] J. Komma, C. Schwarz, G. Hofmann, D. Heinert, and R. Nawrodt, "Thermo-optic coefficient of silicon at 1550 nm and cryogenic temperatures," *Appl. Phys. Lett.*, vol. 101, no. 4, 2012, Art. no. 041905.

[48] G. T. Reed, G. Mashanovich, F. Y. Gardes, and D. Thomson, "Silicon optical modulators," *Nature Photon.*, vol. 4, no. 8, pp. 518–526, 2010.

[49] L. Midolo, A. Schliesser, and A. Fiore, "Nano-opto-electro-mechanical systems," *Nature Nanotechnol.*, vol. 13, no. 1, pp. 11–18, 2018.

[50] T. Komljenovic *et al.*, "Heterogeneous silicon photonic integrated circuits," *J. Lightw. Technol.*, vol. 34, no. 1, pp. 20–35, 2016.

[51] C. Wang *et al.*, "Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages," *Nature*, vol. 562, no. 7725, pp. 101–104, 2018.

[52] V. Sorianello *et al.*, "Graphene-silicon phase modulators with gigahertz bandwidth," *Nature Photon.*, vol. 12, no. 1, pp. 40–44, 2018.

[53] M. Liu *et al.*, "A graphene-based broadband optical modulator," *Nature*, vol. 474, no. 7349, pp. 64–67, 2011.

[54] I. Chakraborty, G. Saha, A. Sengupta, and K. Roy, "Toward fast neural computing using all-photonic phase change spiking neurons," *Sci. Rep.*, vol. 8, no. 1, pp. 1–9, 2018.

[55] H. Jayatilleka *et al.*, "Wavelength tuning and stabilization of microring-based filters using silicon in-resonator photoconductive heaters," *Opt. Exp.*, vol. 23, no. 19, 2015, Art. no. 25084.

[56] A. N. Tait *et al.*, "Feedback control for microring weight banks," *Opt. Exp.*, vol. 26, no. 20, pp. 26422–26443, 2018.

[57] 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.

[58] W. Zhang *et al.*, "Microring weight banks control beyond 8.5-bits accuracy," 2021, *arXiv:2104.01164*.

[59] T. Mizuno, H. Takara, M. Oguma, T. Kobayashi, and Y. Miyamoto, "Modal crosstalk measurement based on intensity tone for few-mode fiber transmission systems," in *Proc. OFC*, 2014, pp. 1–3.

[60] G. P. Agrawal, "Nonlinear fiber optics," in *Nonlinear Science at the Dawn of the 21st Century*. Berlin, Germany: Springer, 2000.

[61] M. Chagnon *et al.*, "Experimental study of 112 Gb/s short reach transmission employing PAM formats and SIP intensity modulator at 1.3 μm," *Opt. Exp.*, vol. 22, no. 17, pp. 21018–21036, 2014.

[62] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[63] R. S. Parpinelli, H. S. Lopes, and A. A. Freitas, "Data mining with an ant colony optimization algorithm," *IEEE Trans. Evol. Comput.*, vol. 6, no. 4, pp. 321–332, Aug. 2002.

[64] A. L.-S. Chua, N. A. Benedek, L. Chen, M. W. Finnis, and A. P. Sutton, "A genetic algorithm for predicting the structures of interfaces in multicomponent systems," *Nature Mater.*, vol. 9, no. 5, pp. 418–422, 2010.

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.

[65] J. Hu and X. Xing, "Sports image detection based on FPGA hardware system and particle swarm algorithm," *Microprocessors Microsyst.*, vol. 80, 2021, Art. no. 103348.
[66] J. L. Burbank, "Security in cognitive radio networks: The required evolution in approaches to wireless network security," in *Proc. 3rd Int. Conf. Cogn. Radio Oriented Wireless Netw. Commun.*, 2008, pp. 1–7.
[67] T. H. Lee, *The Design of CMOS Radio-Frequency Integrated Circuits*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.
[68] E. Lourandakis, R. Weigel, H. Mextorf, and R. Knoechel, "Circuit agility," *IEEE Microw. Mag.*, vol. 13, no. 1, pp. 111–121, Jan./Feb. 2012.
[69] J. Capmany, J. Mora, I. Gasulla, J. Sancho, J. Lloret, and S. Sales, "Microwave photonic signal processing," *J. Lightw. Technol.*, vol. 31, no. 4, pp. 571–586, 2013.
[70] G. P. Agrawal, *Fiber-Optic Communication Systems*. Hoboken, NJ, USA: Wiley, 2012.
[71] G. Bertotti *et al.*, "An 8 bit current steering DAC for offset compensation purposes in sensor arrays," *Adv. Radio Sci.*, vol. 10, pp. 201–206, 2012.

Authorized licensed use limited to: Queen's University. Downloaded on March 12,2022 at 02:39:35 UTC from IEEE Xplore. Restrictions apply.