Photonic dot-product engine for optical signal and information processing

Enxiao Luan, Mahsa Salmani and Armaghan Eshaghi

Huawei Technologies Canada Co., Ltd, 19 Allstate Parkway, Markham, Ontario L3R 5A4, Canada
* Author to whom any correspondence should be addressed.
E-mail: armaghan.eshaghi@huawei.com

Keywords: photonic dot product, optical signal and information processing, Bragg grating modulator, electro-absorption modulator, Mach-Zehnder interferometer

Abstract

In this paper, a novel 2 × 2 Mach–Zehnder-interferometer (MZI)-based dot-product calculator is proposed and analyzed in the silicon-on-insulator (SOI) platform. To calculate the dot product, a phase-shifted Bragg grating (PSBG) modulator is placed in each arm of the MZI for the phase modulation at the resonant wavelength, followed by a 3 dB 2 × 2 directional coupler (DC) as the output to convert the phase difference into the intensity distribution across bar and cross ports. Moreover, an electro-absorption modulator (EAM) is implemented between the PSBG and the DC to change the intensity of the passing light in both arms. Theoretical modeling shows that by adjusting the phase difference and absorption strength individually, multiplication of two input values can be achieved using the proposed design. Numerical analysis over 10 000 dot-product operations with 7 bit precision for input values reveals a mean squared error (MSE) of $2.67 \times 10^{-5}$. By cascading multiple proposed designs operated at different wavelengths, vector-by-vector multiplication can be realized in parallel, leveraging the wavelength-division multiplexing (WDM) scheme and the Bragg reflection mechanism, with results superior to the current and prior MZI-based processors with coherent light sources. This design paves the way for large-scale optical information processing systems with high throughput.

1. Introduction

To address the growing demand for faster computation, computing processors driven by electronics, such as central processing units (CPUs), graphics processing units (GPUs), as well as tensor processing units (TPUs), have been extensively developed over the past decade [1]. However, Moore’s law, the principle that has driven the information-technology revolution since the 1960s, is approaching its growth limit and slowing down the speed of the shrinking transistor dimensions and other data-processing-related improvements. Light has been established as a communication medium for telecommunications and data centers for years, but has not been widely utilized in information processing and computing [2].

Photonic integrated circuits (PICs), which manipulate light signals using on-chip optical waveguides, beam couplers, and splitters, electro-optic modulators, photodetectors, and lasers, etc, have been demonstrated as the platform for a new class of information processing machines to deal with the limits of Moore’s law. Leveraging photons instead of electrons for computation, optical processors can provide high-throughput, power-efficient, and low-latency computing performance by overcoming the inherent limitations of electronics [3]. Many application-specific optical processors have been exploited for solving mathematical [4, 5] and signal processing [6–8] tasks with performance beyond those existing electronic counterparts by orders of magnitude.

Optoelectronic components on a PIC platform have flourished due to the capability of transducing signals between light and electricity. Among those, electro-optic modulators, serving as one of the core components in non-digital computing systems, have been commercially deployed with outstanding...
Figure 1. Schematic of the proposed phase-shifted Bragg grating (BG)-assisted Mach–Zehnder modulator. Input light is equally split between the two arms of the MZM, and phase changed according to the PN-PSBG modulation in each arm. Then, both lights pass through Ge EAMs, which attenuate the intensity to the same level. A 3 dB 2×2 DC after the MZM recombines the light and distributes it between the two output ports according to the phase difference ($I_{in}$ represents the input intensity at resonant wavelength, $\gamma_{EAM}$ represents the power transmittance of the EAM, and $\Delta \phi$ represents the phase difference between the two arms).

performance [9]. Leveraging the thermo-optic effect, or the plasma dispersion effect, a variety of planar modulators have been proposed and demonstrated, using microring resonators (MRRs) [10, 11], microdisk resonators (MDRs) [12], Bragg gratings (BGs) [13], and Mach–Zehnder interferometers (MZIs) [14]. Compared with MZIs, the miniaturized size of MRRs makes them a better candidate for large-scale photonic systems since they allow dense on-chip integration for reducing the footprint, power consumption, and cost [11] as well as parallel operations with incoherent light sources.

Recently, the ‘Broadcast-and-Weight’ (BW) protocol was proposed for implementing neuromorphic optical processors using a bank of tunable MRRs [15]. In this protocol, each MRR is assigned a unique wavelength carrier that is wavelength-division multiplexed and broadcasted. By modulating individual MRRs, incoming WDM signals are weighted and multiplied accordingly, and the outputs of these MRR banks are then accumulated by total power detection at the readout. The number of wavelengths that can be utilized in the ‘BW’ protocol is said to be theoretically limited to 108, depending on both the finesse of the resonator and the channel spacing in linewidth-normalized units [16]. The optical signal modulation is realized by drifting the resonance peak of the MRR modulator. When a bias voltage is applied, the resonance condition changes according to the refractive index (RI) change of the core waveguide, thus resulting in the transmission intensity variation. Wavelength-drift modulation may cause optical crosstalk when one of the resonance peaks moves towards an adjacent one. This happens particularly in large-scale systems with multiple cascaded MRRs, where the optical modulation amplitude (OMA) of the adjacent peak is suppressed proportional to the distance between the two peaks. Therefore, even with fewer than 108 wavelengths in the system, the optical crosstalk becomes a severe issue.

Integration of various electro-refractive and electro-absorptive materials on the PIC platform has recently opened up additional routes toward performant optical processing systems [9]. Electro-absorption modulators (EAMs), based on the Franz–Keldysh (FK) effect or the quantum-confined stark effect, are one of the appealing candidates for enabling high-speed modulation by changing the absorption coefficient within one picosecond. EAMs based on indium tin oxide (ITO) [17], germanium/silicon–germanium alloy (Ge/SiGe) [18], and graphene [19] have been investigated and demonstrated in large-scale optical systems, and present the potential to achieve high-density and low-power consumption modulation. More importantly, modulation schemes based on the absorption coefficient change, intensity modulation, may provide a solution for eliminating optical crosstalk [20].

In this paper, a novel silicon photonic modulator using the intensity modulation scheme is proposed to solve the aforementioned crosstalk issue. As shown in figure 1, by introducing the PN-junction-based phase-shifted Bragg grating (PN-PSBG) and the Ge-EAM to a symmetric Mach–Zehnder modulator (MZM), transmission intensity modulation on resonance can be achieved to realize dot-product operations. Although PSBG-assisted MZI architectures have been investigated for applications of high-efficiency modulation [13, 21] and optical biosensing [22], to the best of our knowledge, this work is the first proposed
architecture for optical signal and information processing leveraging direct intensity modulation. Furthermore, compared to the architectures that are based on coherent optical signal processing [23, 24], our proposed design operates with incoherent input signals benefiting from wavelength-division multiplexing (WDM) technology thanks to the wavelength filtration mechanism of BGs. The PSBG serves as the band-pass filter that can select the resonant wavelength. The PN-junction modulates the output intensity at the resonant wavelength by changing the RI difference between the two MZM arms, and the EAM can achieve the intensity modulation by changing the absorption coefficient of Ge. It is worth noting that these two components do not operate independently. Modulating the PN junction will change the propagation loss and biasing the EAM will generate extra phase change. However, it has been reported that the absorption losses due to the free carrier distribution are limited to 5 dB cm$^{-1}$ when the PN junction is reverse-biased from 0 to 10 V [21].

Considering the PN-PSBGs proposed in our design with the length of approximately 40 μm, the intensity variation due to different biases will be negligible. Furthermore, since the EAM generates identical absorption in both arms, the phase responses in these two arms are the same. Since the proposed design operates at a fixed resonant wavelength for output intensity modulation, the modulation efficiency is highly enhanced due to the round-trip cavity, and the impact of optical crosstalk is suppressed. In addition, compared with the MRR-based optical processing system [16], our proposed design can realize dot-product computation within one device. Therefore, it is more tolerant to the fabrication variations in large-scale systems. Finally, BG resonators present an FSR-free transmission spectrum within the stop-band. Thus, as opposed to MRRs, the number of fan-in wavelength channels is not limited by the finesse [16]. These features facilitate the implementation of large-scale multi-wavelength BW architectures within the proposed design to realize high-density optical processors with low crosstalk.

2. Design and modeling

2.1. Phase-shifted BGs

The BG waveguide is the structure with periodic modulations of the effective index in the propagation direction, realized by varying the RI of materials or physical waveguide dimensions. BG-waveguides have been widely utilized in telecommunications as a part of filters, modulators, or semiconductor lasers for years. When a phase-shifted configuration is added in the center, gratings on each side can be treated as pseudo-mirrors which form a first-order Fabry–Perot (FP) resonant cavity and generate a single resonance peak in the middle of the stop-band (as shown in figure 2(b)) [22]. If the phase-shifted cavity has a length of $\Lambda/2$, where $\Lambda$ is the grating period, the phase-shifted section will introduce a $\pi$-phase shift at the center wavelength of the stop-band.

Figure 2(a) presents the schematics of the cross-section and top view of PN-PSBG in a rib waveguide. By incorporating a lateral PN junction into the PSBG, the RI of the waveguide and the phase response change according to the reverse bias applied to the PN-junction in a free-carrier plasma-depletion mode. Using Lumerical MODE Solutions [25], 2.5D-FDTD simulations were performed to obtain the spectral response (figure 2(b)) and group delays (figure 2(c)), respectively. The simulated spectral result shows that the resonance peak localized at the center wavelength passes through the PSBG, while the rest of the wavelengths are reflected within the stop-band. A group delay spike of 9 ps is observed at the resonant wavelength in figure 2(c). This indicates that the slow-light on resonance propagates a longer optical length. Hence, a larger phase change is obtained. It has been experimentally demonstrated that a PSBG-based modulation structure provides an enhancement factor of 7 in the modulation efficiency compared with the conventional rib waveguide-based counterpart [13].

It is worth mentioning that the width of the stop-band and center wavelength can be selected by tailoring the geometry of the PSBG, namely adjusting the period ($\Lambda$), grating total length ($L$), and the corrugation width ($\Delta W$) as shown in figure 2(a). The center wavelength ($\lambda_c$) is given as

$$\lambda_c = 2\Delta n_{\text{eff}},$$

where $n_{\text{eff}}$ is the effective index of the BG structure. The stop-band ($\Delta \lambda$) is determined by [26]

$$\Delta \lambda = \frac{\lambda^2}{\pi n_g} \sqrt{\kappa^2 + (\pi/L)^2},$$

where $\kappa$ is the coupling coefficient of the gratings, which can be interpreted as the amount of the reflection per unit length, and $n_g$ is the group index. For a stepwise effective index variation ($\Delta n = n_{\text{eff}} - n_{\text{eff}}$ as shown in figure 2(a)), the reflection at each interface can be written as $\Delta n/2n_{\text{eff}}$ according to Fresnel
equations. Each grating period contributes to two reflections, and thus the coupling coefficient ($\kappa$) can be extracted using \[\kappa = \frac{2\Delta n}{\lambda_c}.\] (3)

Therefore, by physically adjusting the corrugation width and the number of gratings on each side, the stop-band can be enlarged, symmetrical to the center wavelength that enables more wavelength channels operated with decent spacing. More details will be discussed in section 3.3.

2.2. Electro-absorption modulators

Research in Ge or GeSi EAMs has flourished on SOI platforms due to their pseudo-direct gap behavior and the compatibility with Si-CMOS technology [28]. Even so, most Ge EAMs require chemical-mechanical polish of Ge, extra poly-Si tapers and several precise Ge etchings, which enhance the manufacturing complexity and cost [29]. A novel evanescent-coupled high-speed Ge waveguide EAM on a 220-nm-thick SOI platform was recently realized [29]. It includes an asymmetric PIN junction employed to change the electric field intensity in the Ge layer for the FK effect. The proposed EAM shows a high electro-optic (EO) bandwidth of 36 GHz at $-1$ V, and clear open 56 Gbps eye diagrams were observed at 1610 nm with a dynamic modulation depth and power consumption equal to 2.7 dB and 45 fJ/bit, respectively [29].

The foundry-compatible Ge EAM in this work is designed according to [29], with the asymmetric PIN junction in Ge formed by N$^{++}$-Ge, i-Ge, and P-Si (figure 3(a)). The cross-section of the proposed Ge EAM is exhibited in figure 3(a), along with the electric and optical field distributions simulated by Lumerical CHARGE Solver [25]. The electric field distribution is obtained at the reverse bias voltage of $-2$ V. Since the optical field is confined in the center of Ge, it can be observed that the electric field and optical field in the EAM overlap. As a result of the FK effect, the proposed Ge EAM offers a highly efficient modulation and maintains a low insertion loss (IL) at wavelengths higher than 1600 nm, presented in figure 3(b).

To shift the operation wavelength to 1550 nm, a small amount of Si can be added to tensile stained Ge to increase the direct bandgap. The absorption contrast, i.e., the relative change in absorption coefficient when an electric field is applied, can reach up to 3 at 1550 nm with the silicon content of 0.8% [28].

2.3. PSBG-assisted electro-absorption MZM

In a conventional MZM, the output intensity is periodically oscillated based on the phase difference between the arms, applied by the phase shifters. Due to the relatively low modulation efficiency of the rib
waveguide-based phase shifter, the footprint is typically on the order of millimeters, which results in high power consumption [30]. A combination of resonators and MZM structures has been utilized to overcome the low modulation efficiency [31, 32]. Instead of using MRRs to assist the MZM, employing PSBGs with strong corrugation in each period exhibits a larger optical bandwidth (no FSR) and smaller footprint, supporting stronger robustness to fabrication variations. To mitigate the effect of fabrication imperfections, a thermal heater is required to be placed on top of the device to cover both arms of the MZM for pre-calibration of the resonance peak to the desired wavelength.

The basic scheme of the proposed PSBG-assisted electro-absorption MZM is depicted in figure 4. By introducing the phase-change component into a symmetric MZI, the output intensity changes according to the phase difference between the two arms. An intensity-absorption element after the phase-change introduces the phase-change component into a symmetric MZI, the output intensity changes according to the low modulation efficiency [31]. Electric fields at the output ports of the output coupler are expressed as:

\[ E_3 = e^{-\alpha/2} \cdot j e^{j \frac{\Delta \phi}{2}} \cdot E_1 \sin \left( \frac{\Delta \phi}{2} \right), \]

\[ E_4 = e^{-\alpha/2} \cdot j e^{j \frac{\Delta \phi}{2}} \cdot E_1 \cos \left( \frac{\Delta \phi}{2} \right). \]

Figure 4. Simplified diagram of the proposed PSBG-assisted electro-absorption MZM, which contains two $2 \times 2$ couplers at the input and output, and a symmetric MZM with one phase shifter in each arm for the phase change, following a shared absorber for the intensity attenuation.
According to equations (7) and (8), output intensities at the two ports of the output coupler follow squared sinusoidal functions with respect to the phase difference and are described as

\[ I_3 = A \cdot \sin^2 \left( \frac{\Delta \phi}{2} \right) \cdot I_1 \propto |E_3|^2, \]  
\[ I_4 = A \cdot \cos^2 \left( \frac{\Delta \phi}{2} \right) \cdot I_1 \propto |E_4|^2, \]

where \( A = e^{-\alpha} \) is the intensity attenuation by the absorber. Therefore, \( I_4 - I_3 = A \cdot [I_1 \cos(\Delta \phi)] \), which correlates to the absorber attenuation and the phase difference in two arms.

### 2.4. Dot-product operation

As described in section 2.3, the proposed PSBG-assisted electro-absorption MZM performs two individual intensity modulations within a single device that can be exploited for carrying out element-wise dot-product operation. Considering the simplest case of scalar product \([X] \times [Y]\), where each matrix only contains one element, element \( X \) can be represented by the intensity variation due to the phase difference \((X = I_{0a}\cos(\Delta \phi))\), where \( I_{0a} \) is the intensity of the input light, and element \( Y \) can be represented by the attenuation of the absorber \((Y = A = e^{-\alpha})\). Therefore, the dot product of \([X] \times [Y]\) is performed using the intensity modulation and represented by the intensity difference at the two output ports:

\[ [X] \times [Y] = I_4 - I_3 = I_{0a}\cos(\Delta \phi) \cdot e^{-\alpha}. \]

Equation (11) implies that element \( X \) can range from \(-1\) to \(1\) due to the subtraction. However, element \( Y \) can only possess positive values.

In order to support any arbitrary, namely positive and negative, values for both matrices, an intermediate absorption status of the EAM is defined and exploited in the dot-product operation. To do so, the righthand-side (RHS) matrix, \( Y \) can be written as \( Y = Y_{\text{real}} - Y_{\text{null}} \), where both \( Y_{\text{real}} \) and \( Y_{\text{null}} \) are positive-valued entities. First, the median of the attenuation of the EAM is selected so as to represent \( Y_{\text{null}} \). Then, the multiplication of elements \( X \) and \( Y_{\text{null}} \) is performed and preserved. Finally, \([X] \times [Y_{\text{real}}]\) is performed and subtracted from the previous result \([X] \times [Y_{\text{null}}]\) \((Y_{\text{real}} - Y_{\text{null}} \in [-1, +1])\). Thus, by applying a two-step dot-product operation, both lefthand-side (LHS) and RHS matrices can support values ranging from \(-1\) to \(1\).

### 3. Simulations

#### 3.1. Model characterization

To demonstrate the dot-product operation, a compact model that represents the PSBG-assisted electro-absorption MZM is developed using Luemrical INTERCONNECT [25], which includes DC components for the optical input/output, PN-PSBG components (introduced in section 2.1) for phase shifting, and Ge EAM components (introduced in section 2.2) for the intensity attenuation. Simulation results of the compact model are shown in figure 5, where the resonant wavelength is located at 1600 nm for the best OMA.

By independently applying reverse biases to the PN-PSBG and the Ge-EAM in the model, resonance peak intensities at the two output ports vary accordingly. Figure 5(a) shows that changing the phase difference \((\Delta \phi)\) between two arms, achieved by modulating only one of the two PN-PSBGs in the MZM, can gradually transfer the output intensity from the in-phase port (fully transmitted when \(\Delta \phi = 0\)) to the out-of-phase port (fully transmitted when \(\Delta \phi = \pi\)) of the output coupler. Therefore, by subtracting the intensities at in-phase and out-of-phase ports, the plot of the weight versus the reverse bias (figure 5(b)) is achieved. To maintain a balanced weight range from \(-1\) to \(1\), a normalization step is required that selects the maximum absolute value that both negative and positive weights can achieve as the normalization parameter (shown in the red region in figure 5(b)). Figure 5(c) depicts the normalized weight plotted considering a 7-bit precision for the reverse bias voltage. Similarly, in figure 5(d), when the reverse bias voltages are simultaneously applied to Ge EAMs in both arms, no additional phase difference occurs between two arms. Hence, only the intensity at the in-phase port drops due to the absorption coefficient change. Figures 5(e) and (f) present the obtained weight values, containing only positive values and the normalized weight values after the intermediate-level absorption subtraction, respectively.

The resonance peak drift in the aforementioned model that happens as a result of applying a reverse bias voltage to the PN-PSBG is ignored for simplicity. The reason is that the resonant wavelength shift \((\Delta \lambda)\),
caused by the index change ($\Delta n$), is independent of the size of the resonator and is equal to $\Delta \lambda / \Delta n = \lambda / n g$ while the phase change ($\Delta \phi$), caused by the index change within the MZI, is related to its optical length ($l$) and is equal to $\Delta \phi / \Delta n = 2\pi l / \lambda$ [35]. Therefore, in our proposed PSBG-assisted electro-absorption MZM architecture, the phase response of the MZI benefits from the optical length enhancement caused by the BG resonator, while the resonance peak drift does not.

3.2. Dot-product demonstration

Balanced photodetectors (BPDs) have been widely used in optical processing systems to serve as the output intensity subtractor, multi-wavelength signal accumulator, and O/E signal converter [15, 36]. In our proposed architecture, a BPD model, with the responsivity of $1 \text{ A/W}$ and the noise spectral density of $1 \times 10^{-12} \text{ A/}\sqrt{\text{Hz}}$, is employed at the readout for dot-product calculations. Although the responsivity is wavelength-dependent at C-band, it is weakly dependent on the carrier wavelength. So, this dependence is not considered, and we assume that the output current is linear with respect to the input power in simulations at operational wavelengths, where the input power requires to be higher than the noise-equivalent power (NEP) and lower than the saturation power of the PD.

In order to obtain the result of the dot product, the mapping between the photocurrent generated by the BPD and the correlated numerical value needs to be investigated. Here, a look-up table approach has been adopted. The total number of points in the point-to-point look-up table increases exponentially with the precision bits of the input values. Hence, generating a complete look-up table, which contains the results of all the possible dot-product operations, is challenging in many cases, particularly for more complex calculations such as vector-to-vector or matrix-to-matrix multiplications. Our preliminary observation from simulations indicates that the BPD output photocurrent scales almost linearly with the actual value of the multiplication result. Accordingly, the energy and the time required for generating a complete look-up table can be significantly reduced by employing a prediction scheme based on finding the correlation between the BPD output photocurrent and the actual dot-product result over a limited number of sample points, and then generalizing it to the other possible values using linear regression approach.

It is worth mentioning that the mapping accuracy using the predicted look-up table is highly dependent on the precision of input values and the number of sample points for predicting the look-up table. In our preliminary simulations, we noticed that using 49 points to calculate the linear mapping results in a mapping error with a standard deviation of 1.698 when the input data has a precision of 7 bits. Therefore, in this paper, 49 sample points are utilized to predict the look-up table.
Our demonstration performs 100 scalar-to-scalar dot-product operations by using the proposed PSBG-assisted electro-absorption MZM model. The scalar weight values are randomly chosen based on the plots in figures 5(c) and (f), which exhibit the normalized weight versus applied voltage for the PN-PSBG and the Ge EAM, respectively. Then, by measuring the photocurrent values after the BPD’s subtraction, the correlation between the actual dot-product results of the considered points and their corresponding photocurrent values can be obtained, as plotted in figure 6(a). As can be seen, the output photocurrent of these 100 random dot-product calculations has an almost linear relationship with respect to the actual result, and it can be perfectly fit through a first-order linear fitting. Accordingly, the complete look-up table can be efficiently generated as depicted by the red line in figure 6(a).

To evaluate the accuracy of the proposed approach for generating the look-up table, 10,000 dot-product operations with 7-bit precision are calculated, mapped, and compared with the actual dot-product results, where the input values lie within $[-1, 1]$ and include one decimal digit. The evaluation result, i.e. predicted dot-product results versus the actual ones, is shown in figure 6(b), with the inset showing a histogram of the error with an MSE of $2.67 \times 10^{-5}$ and a standard deviation of 0.0051. Furthermore, we investigated the impact of the precision of the modulator driver on the calculation accuracy. As shown in figure 6(c), changing the precision of the reverse bias voltage from 5 to 7 bits results in a significant reduction in the MSE and the standard deviation of the error, while for higher precisions, i.e. 8 to 10 bits, the changes in the MSE and the standard deviation are no longer significant.

### 3.3. Vector-by-vector dot product

In this work, to realize vector-by-vector dot product, cascading multiple PSBG-assisted electro-absorption MZM units are proposed along with the WDM scheme. Multiplexed broadcast signals in the waveguide need to be de-multiplexed for each of these units. Thanks to the wavelength-filtering feature of the PSBG resonator, the resonant wavelength is trapped in the FP cavity and passes through the gratings, while the rest of the light in the stop-band is reflected. The width of the stop-band can be adjusted according to the number of wavelength channels and their spacing when necessary. As described in figure 7, three PSBG-assisted electro-absorption MZM units are cascaded for $3 \times 3$ vector-by-vector dot-product calculation, while both input and output ports are implemented using 3 dB $2 \times 2$ DCs. These three devices work at different resonant wavelengths at $\lambda_1$, $\lambda_2$, and $\lambda_3$, respectively. When the multiplexed signal is imported into the system through the upper port of the input DC of Device 1, the signal at $\lambda_1$ will be transmitted for a $1 \times 1$ dot-product calculation within Device 1. The rest of the signals, at $\lambda_2$ and $\lambda_3$, will be reflected, since they are within the stop-band of Device 1. These signals will be delivered to the next device (Device 2) through the bottom port of the input-DC of Device 1, which is connected to the upper port of the input-DC of Device 2. Accordingly, broadcast signals at different wavelengths are de-multiplexed and operated individually to perform $1 \times 1$ dot product in each unit of the proposed system. By adding the BPD after each computing unit, the output powers from two output ports of the DC are O/E converted and subtracted. Ultimately, the vector-by-vector dot product is achieved throughout the whole system by accumulating the photocurrent from each computing unit.

It is worth mentioning that the wavelength on resonance passes through the PN-PSBG multiple times, and this generates a strong phase change. Hence, the phase variation of the reflected light, caused by RI change, has a negligible impact. Moreover, the intensity drop of the reflected light can be treated as the system insertion loss. Based on the simulation results in figure 7, signals at three different wavelengths can be operated independently, using the intensity-modulation scheme, with no optical crosstalk.
Figure 7. Schematic of the proposed PSBG-assisted electro-absorption MZM-based optical information processing system, with broadcast signals multiplexed at three different resonant wavelengths. Each signal is de-multiplexed and operated in each device with an intensity-modulation scheme using the PN-PSBG and EAM to represent $X_i$ and $Y_j$, and the rest of the signals are reflected and delivered to the next device. A BPD is employed after each device to realize the O/E conversion and the subtraction. Then, the accumulation of the obtained photocurrent from each BPD is mapped to the dot-product result.

Table 1. Performance comparison between the design proposed in this work and MRM-based systems.

|                              | This Work (in theory) | MRM-based Systems |
|------------------------------|-----------------------|-------------------|
| Footprint                   | $L_{PSBG} = \sim 40 \mu m$ | Radius = $\sim 10 \mu m$ |
|                             | $L_{DC} = \sim 50 \mu m$  | Total size with 2 MRMs = $\sim 126 \mu m^2$ |
|                             | Total size = $\sim 180 \mu m^2$ |                   |
| EO-bandwidth                | $BW_{PN-PSBG} = 28 \text{GHz}$ [13] | $175 \text{kHz}$ [11] |
|                             | $BW_{Ge\ EAM} = 56 \text{GHz}$ [29] | $50 \text{GHz}$ [38] |
| Power consumption           | 84 fl/bit for PN-PSBGs [13] | 274 nl/bit per MRM [11] |
|                             | 45 fl/bit for Ge EAMs [29] | 70 fl/bit per MRM [38] |
| Weight range                | LHS: $[-1, +1]$ | LHS: $[0, +1]$ |
|                             | RHS: $[-1, +1]$ | RHS: $[-1, +1]$ [15] |
| Optical crosstalk           | No (intensity modulation) | Yes (wavelength modulation) |
| Channel density             | FSR-free in the stop-band | FSR limited |
| Compute density             | $15.6 \text{TOPS mm}^{-2}$ | $2.8 \times 10^{-4} \text{TOPS mm}^{-2}$ |
|                             |                        | $79.5 \text{TOPS mm}^{-2}$ |

a The footprint is defined as the total architecture size that can realize $1 \times 1$ dot product. This work only requires one device, while the systems leveraging MRMs require at least two individual modulators.
b The compute density describes the computation speed of the unit area of a processor under the highest achievable EO-bandwidth of the system. We assume that the total required size for each processor system is 10 times the footprint of the core computing device specified in the first row. It is worth noting that in this work, a two-step computing scheme is used to realize arbitrary values for both RHS and LHS matrices, and this reduces the compute density (section 2.4) while, in the MRM-based system, only positive values are realizable for the LHS matrix.

3 × 3 vector-by-vector dot-product simulations were carried out with 7-bit precision for the input data. 10 000 calculations indicate an MSE of $3.06 \times 10^{-5}$.

It is known that when the light is injected into one of the input ports of the 3-dB 2 × 2 DC, output light at the 2 output ports of the coupler is equal in amplitude but with a $\pi/2$ phase difference. For the reflected light within the stop-band, the 2 × 2 DC that originally served as the input splitter will act as a combiner. The original $\pi/2$ phase difference caused by the DC will force the reflected light in the two arms to recombine and obtain a total reflection and delivery from Device 1 to Device 2 (shown in figure 7).

In table 1, PSBG-assisted electro-absorption MZM design is compared to other optical modulation systems reported in the literature, which are based on MRR modulators (MRMs) using thermal or carrier density modulations. It can be observed that the design proposed in this work has a slightly larger footprint compared with the MRM-based system that uses at least two individual MRMs to realize $1 \times 1$ dot-product operations. Thanks to the low power consumption of the EAM element, this design has a relatively smaller energy usage per bit. Moreover, the proposed design supports arbitrary values for both LHS and RHS matrices, while only positive values can be realized in the LHS matrix of the MRM-based system.
Furthermore, since the PSBG resonator is not subject to FSR limitations, our design has the potential of multi-wavelength operations with higher channel density, and no optical crosstalk is present due to the intensity-modulation scheme. Finally, a compute density of 15.6 TOPS mm$^{-2}$ (trillions of operations per second per square millimeter) is estimated based on the highest reported EO-bandwidth of PN-PSBGs [13] and considering a two-step operation scheme for the dot-product computation discussed in section 2.4.

4. Conclusion

In this paper, a novel optical processor is proposed that performs vector-by-vector dot products. Using the PN-junction-based PSBG resonator implemented in a symmetric MZI, the phase difference between the two arms caused by the RI modulation can be enhanced at the resonant wavelength (around sevenfold that of the rib waveguide-based counterpart) and transferred to the intensity difference at the two output ports. A BPD is employed at the readout for O/E conversion to realize an electrical signal ranging from negative to positive values after subtraction. A broadband EAM absorption component is added to the MZI to modulate the transmittance by changing the absorption coefficient. This allows an independent intensity modulation in both arms for the dot-product operation. An intermediate level of absorption is predefined for the absorption element. By applying a 2-step operation for the single $1 \times 1$ dot product, the absorption-based modulation can achieve negative and positive values. It is worth noticing that the absorption element can also be replaced by phase-change materials (PCMs) operated in the non-volatile regime for in-memory computing.

Simulation results show that for $1 \times 1$ dot product with 7-bit precision for the input values and modulation speed up to 28 GHz, the MSE is on the order of $10^{-5}$ with the standard deviation of 0.0051. This demonstrates the capability of the proposed system to perform high-throughput and high-accuracy computing. Compared with the system reported in [15], which uses MRMs as the weight bank, this design offers lower power consumption and a larger range for input values. No optical crosstalk and FSR limitations are observed in this design. The proposed PSBG-assisted electro-absorption MZM design supports the BW protocol for multi-wavelength operations in parallel to facilitate the calculation of the dot product for vectors of bigger size.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

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

Conflict of interest

The authors declare no conflicts of interest.

ORCID iDs

Enxiao Luan  https://orcid.org/0000-0001-6404-5433
Mahsa Salmani  https://orcid.org/0000-0002-9810-4435

References

[1] Zhou T, Lin X, Wu J, Chen Y, Xie H, Li Y, Fan J, Wu H, Fang L and Dai Q 2021 Large–scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit Nat. Photon. 15 367–73
[2] Shastri B J, Tait A N, de Lima T F, Pernice W H, Bhaskaran H, Wright C D and Prucnal P R 2021 Photonics for artificial intelligence and neuromorphic computing Nat. Photon. 15 102–14
[3] Zhang Q, Yu H, Barbiero M, Wang B and Gu M 2019 Artificial neural networks enabled by nanophotonics Light Sci. Appl. 8 1–14
[4] Estakhri N M, Edwards B and Engheta N 2019 Inverse–designed metastructures that solve equations Science 363 1333–8
[5] Xu X, Huang X, Li Z, Gao J, Jiao Z, Wang Y, Ren R, Zhang H and Jin X 2020 A scalable photonic computer solving the subset sum problem Sci. Adv. 6 eaay5853
[6] Wang J and Long Y 2018 On–chip silicon photonic signaling and processing: a review Sci. Bull. 63 1267–310
[7] Zhang W and Yao J 2020 Photonic integrated field–programmable disk array signal processor Nat. Commun. 11 1–9
[8] Salmani M, Eshaghi A, Luan E and Saha S 2021 Photonic computing to accelerate data processing in wireless communications Opt. Express 29 22299–314
[9] Rahim A, Hermans A, Wohlfel B, Petousi D, Kuyken B, Van Thourhout D and Baets R 2021 Taking silicon photonics modulators to a higher performance level: state–of–the–art and a review of new technologies Adv. Photonics 3 024003
[10] Pantouvaki M, Verheyen P, De Coster J, Lepage G, Absil P and Van Campenhout J 2015 56 Gb/s ring modulator on a 300 mm silicon photonics platform 2015 European Conf. on Optical Communication (IEEE) pp 1–3
[11] Jayatilleka H, Shoman H, Chrostowski L and Shekhar S 2019 Photoconductive heaters enable control of large–scale silicon photonic ring resonator circuits Optica 6 84–91.
[12] Timurdogan E, Sorace–Agaskar C M, Sun J, Hosseini E S, Biberman A and Watts M R 2014 An ultralow power a thermal silicon modulator Nat. Commun. 5 1–11.
[13] Jafari O, Shi W and LaRochelle S 2020 Mach–Zehnder silicon photonic modulator assisted by phase–shifted Bragg gratings IEEE Photonics Technol. Lett. 32 445–8.
[14] Liao L, Liu A, Rubin D, Basak J, Chetrit Y, Nguyen H, Cohen R, Izhaky N and Paniccia M 2007 40 Gbit/s silicon optical modulator for high–speed applications Electron. Lett. 43 1196–7.
[15] Bangari V, Marquez B A, Miller H, Tait A N, Nahmias M A, De Lima T F, Peng H T, Prucnal P R and Shastri B J 2019 Digital electronics and analog photonics for convolutional neural networks (DEAP-CNNs) IEEE J. Sel. Top. Quantum Electron. 26 1–13.
[16] Tait A N 2018 Silicon photonic neural networks PhD Dissertation Princeton University.
[17] Amin R et al 2019 ITO–based electro–absorption modulator for photonic neural activation function APL Mater. 7 081112.
[18] Female S A et al 2015 56 Gb/s germanium waveguide electro–absorption modulator J. Lightwave Technol. 34 419–24.
[19] Giambra M A, Sorianello V, Miseikis V, Marconi S, Montanaro A, Galli P, Pezzini S, Coletti C and Romagnoli M 2019 High–speed double layer graphene electro–absorption modulator on SOI waveguide Opt. Express 27 20145–55.
[20] Marquez B A, Morison H, Guo Z, Filipovich M, Prucnal P R and Shastri B J 2020 Graphene–based photonic synapse for multi wavelength neural networks MRS Adv. 5 1909–17.
[21] Jafari O, Sephehrian H, Shi W and LaRochelle S 2019 High–efficiency silicon photonic modulator using coupled Bragg grating resonators J. Lightwave Technol. 37 2065–75.
[22] Luan E, Yun H, Lin S H, Cheung K C and Chrostowski L 2020 Phase–shifted Mach–Zehnder interferometer sensor using an intensity interrogation scheme Optical Fiber Communication Conf. (Optical Society of America) p M1C–5.
[23] Shen Y et al 2017 Deep learning with coherent nanophotonic circuits Nat. Photon. 11 441–6.
[24] De Marinis L, Sorel M, Klitis C, Contestabile G and Andriolli N 2021 Silicon photonic filter–based dot product engine for convolutional neural networks Signal Processing in Photonic Communications (26 July 2021) (Optical Society of America) SpM5C–3.
[25] Ansys-Lumerical 2021 Lumerical (available at: www.lumerical.com) (Accessed 21 July 2021).
[26] Buus J, Amann M–C and Blumenthal D J 2005 Tunable Laser Diodes and Related Optical Sources (Hoboken, NJ: Wiley).
[27] Wang X 2013 Silicon photonic waveguide Bragg gratings PhD Dissertation University of British Columbia.
[28] Liu J 2014 Monolithically integrated Ge-on-Si active photonics Photonics 1 162–97.
[29] Liu Z, Li X, Niu C, Zheng J, Xue C, Zuo Y and Cheng B 2020 56 Gbps high–speed Ge electro–absorption modulator Photon. Res. 8 1648–52.
[30] Shrestinsky M, Ayazi A, Xuan Z, Lim A E–J, Lo G–Q, Baehr–Jones T and Hochberg M 2013 Highly linear silicon traveling wave Mach–Zehnder carrier depletion modulator based on differential drive Opt. Express 21 3818–25.
[31] Li R, Patel D, Samani A, El-Fiky E, Xing Z, Morsy–Osman M and Plant D V 2017 Silicon photonic ring–assisted MZI for 50 Gb/s DAC–less and DSP–free PAM–4 transmission IEEE Photonics Technol. Lett. 29 1046–9.
[32] Heebner J, Grover R and Ibrahim T 2008 Optical Microresonator Theory (Berlin: Springer).
[33] Le T–T and Cahill L 2013 Generation of two Fano resonances using 4 × 4 multimode interference structures on silicon waveguides Opt. Commun. 301 100–5.
[34] Lu X et al 2021 11 TOPS photonic convolutional accelerator for optical neural networks Nature 589 44–51.
[35] Sun J, Sakib M, Driscoll J, Kumar R, Jayatilleka H, Chetrit Y and Rong H 2018 A 128 Gb/s PAM4 silicon microring modulator Optical Fiber Conf. (Optical Society of America) p Th4A–7.