



Review

# Photonic Matrix Computing: From Fundamentals to Applications

Junwei Cheng, Hailong Zhou and Jianji Dong \*

Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China; jwcheng@hust.edu.cn (J.C.); hailongzhou@hust.edu.cn (H.Z.)

\* Correspondence: jjdong@hust.edu.cn

**Abstract:** In emerging artificial intelligence applications, massive matrix operations require high computing speed and energy efficiency. Optical computing can realize high-speed parallel information processing with ultra-low energy consumption on photonic integrated platforms or in free space, which can well meet these domain-specific demands. In this review, we firstly introduce the principles of photonic matrix computing implemented by three mainstream schemes, and then review the research progress of optical neural networks (ONNs) based on photonic matrix computing. In addition, we discuss the advantages of optical computing architectures over electronic processors as well as current challenges of optical computing and highlight some promising prospects for the future development.

**Keywords:** photonic matrix computing; photonic accelerators; artificial intelligence; optical neural networks; photonic integrated platform; diffractive planes



**Citation:** Cheng, J.; Zhou, H.; Dong, J. Photonic Matrix Computing: From Fundamentals to Applications. *Nanomaterials* **2021**, *11*, 1683. <https://doi.org/10.3390/nano11071683>

Academic Editors: Carlos Alberto Alonso-Ramos and Linjie Zhou

Received: 29 May 2021  
Accepted: 24 June 2021  
Published: 26 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

With the proliferation of artificial intelligence and the next-generation communication technology, the growing demand for high-performance computing has driven the development of custom hardware to accelerate this specific category of computing. However, processors based on electronic hardware have hit the bottleneck of unsustainable performance growth as the exponential scaling of electronic transistors reaches the physical limit revealed by Moore's law [1]. Photonic processors compute with photons instead of electrons, and therefore optical computing can dramatically accelerate computing speed by overcoming the inherent limitations of electronics. Unlike electrical circuit technologies, photonic circuits have some extraordinary properties such as ultra-wide bandwidth, high frequency, and low energy consumption. Furthermore, light has several dimensions such as wavelength, polarization, and spatial mode to enable parallel data processing, resulting in remarkable acceleration against a conventional von Neumann computer, which makes the optical computing approach a viable and competitive candidate for artificial intelligence accelerators.

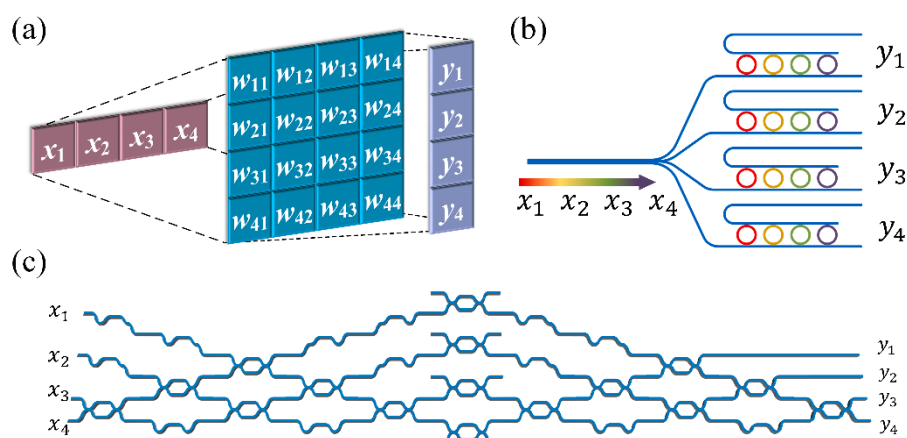
The matrix–vector multiplication (MVM) operation is one of the fundamental mathematical operations widely used in large-scale neuromorphic optoelectronic computing. The weighted interconnections between adjacent photonic neurons in ONNs can be mathematically represented by a matrix whose entries are the weight values, and each entry vector is multiplied by the input signal of a particular synapse [2], which matches the mathematical nature of MVM. To some extent, neuromorphic engineering is an attempt to move computational processes of artificial intelligence algorithms to specific hardware, enabling functions that are difficult to realize with conventional computing hardware. For example, linear optical elements can calculate convolutions, Fourier transforms, random projections, and many other operations through light–matter interaction or light propagation [3–6].

There has been rapid progress in research on photonic matrix computing, and different photonic devices have been successfully used to implement matrix operations. Optical

modulator arrays, such as electro-optic modulated direct-driven LED arrays and acousto-optic Bragg devices, can perform matrix calculations at a much faster rate than existing electronic devices [7–9]. The optical implementation of convolutional neural networks with fast operation speed and high energy efficiency is appealing owing to its outstanding capability of feature extraction [10]. In particular, convolutional processing based on MVM, which is a computationally intensive operation in electronics, occupies over 80% of the total processing time in convolutional neural networks [11], therefore computational acceleration for convolutional neural networks can be achieved by matching hardware and MVM operations. The MVM operation can be mathematically described as

$$Y = WX = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \dots & \dots & \dots & \dots \\ w_{N1} & w_{N2} & \dots & w_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_N \end{bmatrix} \quad (1)$$

where  $X$  is the input vector,  $W$  is the matrix, and  $Y = [y_1, y_2, \dots, y_N]^T$  is the output vector. Optical MVMs can be implemented by three mainstream optical methods, as shown in Figure 1, including the multiple plane light conversion (MPLC) method, the wavelength division multiplexing (WDM) method, and the Mach–Zehnder interferometer (MZI) method.



**Figure 1.** Three categories of optical MVM, including (a) MPLC-MVM method, (b) WDM-MVM method, and (c) MZI-MVM method.

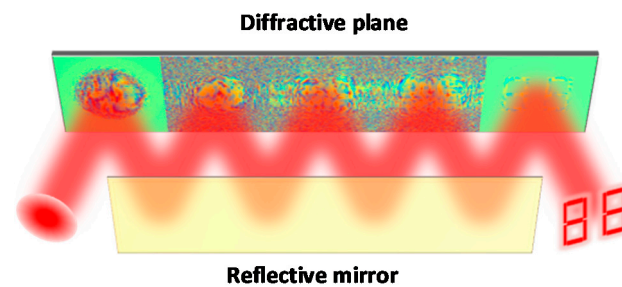
In this review, we firstly survey recent researches and progress in optical MVMs and photonic artificial intelligence hardware. After that, we describe the basic principles of optical MPLC-MVM, WDM-MVM, and MZI-MVM methods. Finally, we discuss the advantages of optical computing architectures over electronic processors as well as current challenges of optical computing, and provide perspectives for further improvement of optical computing architectures.

## 2. MPLC Matrix Core

Unlike integrated schemes such as microring and MZI matrix cores, the MPLC matrix core builds computing capabilities directly above an optical field propagating in free space. Among these three methods, MPLC was the first to be implemented in optical computing [12], and the initially programmable MVM was finished with spatial optical elements [9]. The MPLC matrix core is the only one that can currently support super-large-scale matrix operation, which makes it valuable in pulse shaping [13], mode processing [14–16], and machine learning [5,17–19].

The principle of the MPLC method is shown in Figure 2, and the architecture of the MPLC matrix core consists of a series of planes encoded with amplitude and phase infor-

mation, and reflective mirrors are only used to change the direction of light propagation to reduce the spatial volume of the system. By configuring the parameters of these planes, the light beam irradiated on the plane surface can be modulated to change amplitude and phase. The incident light diffracts in free space and then incident to the first plane, after that, it diffracts in free space and then incident into the second plane, and so on. After passing through all planes, the final optical field will be output and then be detected by the photodetector array.



**Figure 2.** The principle of the MPLC matrix core. The MPLC matrix core is composed of a series of diffractive planes encoded with amplitude and phase information. By configuring the parameters of these diffractive planes, the light beam irradiated on the plane surface can be modulated to change amplitude and phase.

The transmission matrices  $M_i$  ( $i = 1, 2, \dots, N, N + 1$ ) are fixed owing to the propagation distance in free space being fixed. The transmission matrix of each plane is a diagonal matrix, denoted as  $A_i$  ( $i = 1, 2, \dots, N$ ). Therefore, the transmission matrix  $T$  of this MPLC system can be expressed as

$$T = M_{N+1}A_N \dots M_2A_1M_1. \quad (2)$$

The arrangement of pixels in each plane relates to the matrix dimensions. For example, if we assume that the pixels of each plane are  $p \times q$ , then the dimensions of the matrices  $M_i$ ,  $A_i$ , and  $T$  are all  $pq \times pq$ . From Equation (2), the multiplane transformation can be mathematically expressed as the cross product of a series of fixed matrices and configurable diagonal matrices. Theoretically, arbitrary matrix operation can be implemented as long as there are enough planes, and the number of planes required is approximately equal to 2-fold the number of input modes, but in fact, only a few planes are needed to approximately achieve the function of target transmission matrix [20].

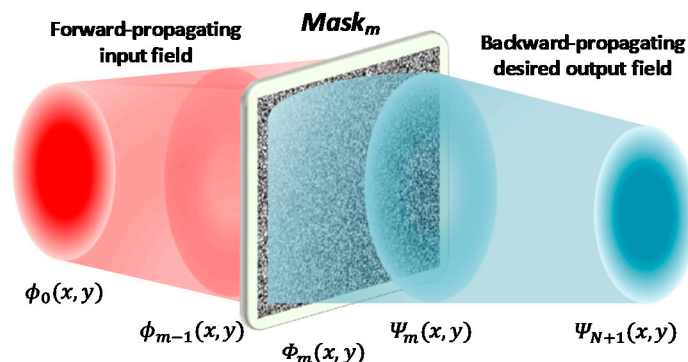
In practice, the transmission matrix  $T$  in Equation (2) is difficult to analyze or measure experimentally. In general, parameters of each phase plane can be obtained by iterating according to the input and target output. One commonly used method is the wavefront matching method [21], as shown in Figure 3, where the input optical field is  $\phi_0(x, y)$ , and after forward propagation, the distribution of the optical field in front of the  $m$ th phase plane is  $\phi_m(x, y)$ . On the other hand,  $\Psi_m(x, y)$ , the distribution of the optical field after the  $m$ th phase plane, can be obtained by the backward propagation of the output optical field  $\Psi_{N+1}(x, y)$ . Hence the theoretical phase distribution of the  $m$ th phase plane should be the phase difference between the two fields, i.e.,  $\phi_m(x, y)$  and  $\Psi_m(x, y)$ :

$$\Phi_m(x, y) = \Lambda[\Psi_m(x, y)\phi_m^*(x, y)] \quad (3)$$

where  $\Lambda$  means the function to get the phase angle. If there are multiple mode pairs of input and output, then the phase value of the plane should be its weighted phase value:

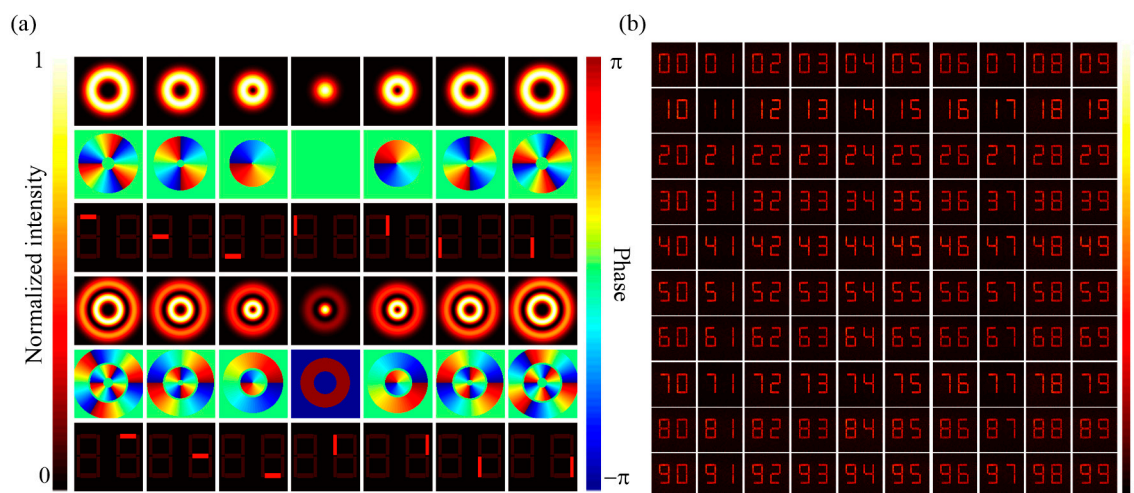
$$\Phi_m(x, y) = \Lambda \left[ \sum_j \Psi_{m,j}(x, y)\phi_{m,j}^*(x, y) \right] \quad (4)$$

where  $j$  represents the  $j^{\text{th}}$  set of input–output mode pairs. By the means of iterating back and forth to update the parameters of phase planes repeatedly, the algorithm will eventually converge. The key advantage of the wavefront matching method is that the iterative speed is very fast, and the whole plane is updated each iterative process, thus the convergence is exponential. Generally, it only needs fewer than 20 iterations back and forth to achieve convergence.



**Figure 3.** Schematic diagram of the wavefront matching method.

Here, in order to visually demonstrate the working mechanism of the MPLC matrix core and verify its capability to implement large-scale optical computing, we theoretically demonstrate a specific example of numeric holographic coding based on the MPLC scheme (Figure 4). We designed a two-digit seven-segment display with 14 segments in total, respectively powered by 14 different Laguerre-Gaussian (LG) modes. The two-digit numbers range from 00 to 99 can be displayed and switched by setting the combination of input modes, as shown in Figure 4b.



**Figure 4.** The numeric holographic coding based on the MPLC matrix core. (a) This figure summarizes the input-to-output mapping relationships between the input LG modes and the output segments of numeric display. (b) Results of the optical-powered numeric display from 00 to 99.

How to efficiently deal with the increasing scale of neural network computing remains a significant problem to be solved. Benefiting from the parallelism and minimal latency of optical systems in free space, the MPLC scheme has the ability to implement large-scale matrix operation, which makes it the potential candidate for large-scale neural networks. In 2018, Lin et al. introduced the diffraction ONNs framework used for all-optical machine learning, i.e.,  $D^2NN$ , and experimentally demonstrated the image classification with Modified National Institute of Standards and Technology (MNIST) handwritten digits

and Fashion-MNIST datasets [5]. The optical D<sup>2</sup>NN architecture shows great potential for machine learning applications, and it would be more complete if it included an optical nonlinear activation function [22]. The image information is encoded in the amplitude or phase channels of the input optical fields, and wave propagation in free space can be mathematically described by Kirchhoff's diffraction integral, which amounts to a convolution operation of the optical field with a trained kernel. In this work, the training process of ONNs is still completed by an electronic computer to update parameters, and each diffractive layer is fabricated by 3D printing technology. In 2021, Rahman et al. applied a pruning algorithm to further improve the image classification accuracy of D<sup>2</sup>NNs [23]. On the image classification of the CIFAR-10 dataset released by Canadian Institute For Advanced Research, whose test images are more complicated than MNIST and Fashion-MNIST, the D<sup>2</sup>NN architecture combined with the pruning algorithm provides an inference improvement of more than 16% compared to the average performance. In neuromorphic optoelectronic computing, the functionality of input nodes and output neurons is implemented with programmable diffractive optical devices, such as the digital micromirror device (DMD) and spatial light modulator (SLM). The DMD provides a high optical contrast for encoding information. It allows encoding the binarized data into the amplitude of coherent optical fields, where a phase SLM subsequently modulates their phase distribution to realize the diffractive modulation. Zhou et al. proposed an optoelectronic fused computing architecture with a reconfigurable diffractive processing unit, which can support different neural networks, and achieved excellent experimental accuracies for image and video recognition over benchmark data sets [19]. It can be seen that the MPLC scheme is of great potential to narrow the gap and surpass in classification performance between optical computing architecture and state-of-the-art electronic computers.

Similar to the MPLC method demonstrated in free space, the MPLC matrix core can also be realized on a chip, whose structure is shown in Figure 5. The integrated MPLC matrix core is composed of multiple layers, including alternately arranged tunable layers and unitary diffractive layers, to implement multichannel transformation described by a unitary matrix. The tunable layer can be achieved by independent phase shifters or time delay units, corresponding a unitary transfer matrix with diagonal form. The unitary diffractive layers describe the interaction between channels, which can be implemented by multimode interference (MMI) couplers or a region of coupled waveguides; therefore, the transfer matrices of unitary diffractive layers are static because the structure of MMI couplers and coupled waveguides cannot be changed once fabricated.

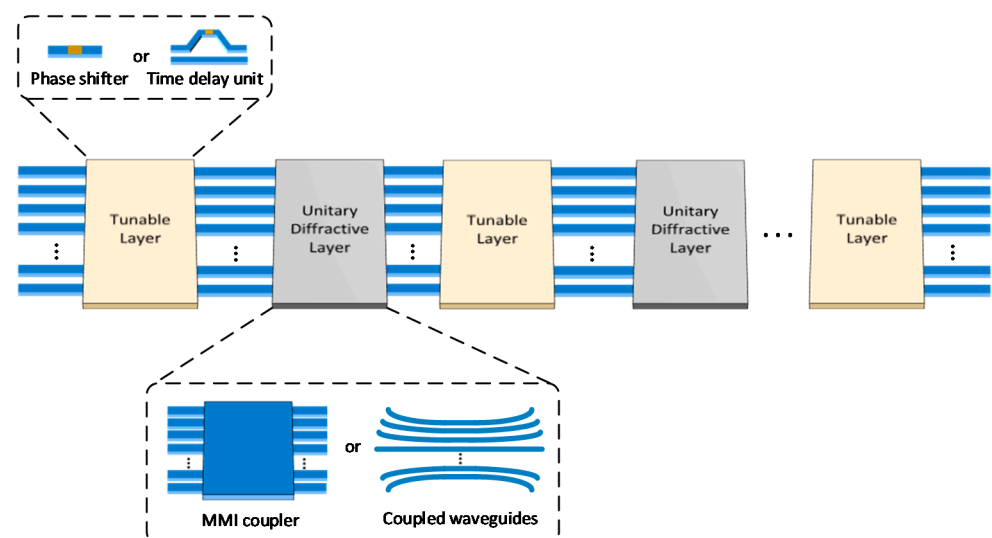


Figure 5. The structure of the integrated MPLC matrix core.

In 2017, Tang et al. proposed a novel integrated architecture consisting of cascaded MMI couplers and phase shifter arrays [24] to realize an  $n \times n$  unitary transfer matrix. The  $n \times n$  unitary transfer matrix can be decomposed as  $M$  stages of unitary matrices with only diagonal elements (i.e., the transfer matrices of phase shifter arrays) and another  $M - 1$  stage of unitary matrices (i.e., the transfer matrices of MMI) when the path-dependent coupling loss is not considered. After that, Saygin et al. made an in-depth analysis of this integrated MPLC architecture, in which the flexibility and robustness of this architecture have been thoroughly discussed. The transfer matrices of the static MMI blocks can be randomly chosen from a continuous class of unitary matrices without sacrificing the quality of approximation for the target unitary transformation, making this scheme insensitive to errors [25]. The integrated MPLC matrix core provides an alternative viable solution to decompose large unitary matrices into small ones with high flexibility and robustness, and thus optical diffractive neural networks are possible to build on a chip based on this method.

### 3. Microring Matrix Core

The MPLC scheme of photonic matrix computing was thoroughly discussed in Section 2, from which we find that the predominant advantage of MPLC schemes is the capability to implement a large-scale matrix operation at present. However, one major drawback cannot be ignored—the MPLC scheme is usually limited by bulky optical instruments, and hence they are difficult to highly integrate on a chip.

The microring has a very compact structure and its radius can be as small as a few microns [26], which means the footprint of photonic devices can be greatly reduced, and thus the integration density can be competitive. The microring has been widely used in on-chip WDM systems [27–29], filtering systems [30–32], etc. In addition, a microring array can also be used in the operations of incoherent matrix computation since each microring can independently configure the transmission coefficient of a wavelength channel. Therefore, the microring matrix core is well suited to implement WDM-MVM operation.

The implementations of matrix computation enabled by a microring array are shown in Figure 6. The  $N \times N$ -sized microring array in the right region of Figure 6 corresponds to an  $N \times N$ -sized matrix  $M$ , alternatively called a microring matrix core, and the input signal  $X$  can be represented by a vector with a length of  $N$ . The input signal  $X$  can be generated off-chip or on-chip. If it is generated off-chip, the microring array in the left region of Figure 6, i.e., the microring front module, only serves as a wavelength division multiplexer for combining multiple input wavelength channels. If it is generated on-chip, the microring front module is used to modulate the input vector  $X$  to different wavelengths. After that, the input signal is divided into multiple beams with equal power after passing through the beam splitter, and each beam is sent to a different row of the microring matrix core in the right region of Figure 6. Each row of the microring matrix core can independently configure the transmission coefficient of each wavelength channel, and the total power is finally detected from each output port by photodetectors.

The add-drop microring structure [33] is widely applied in on-chip optical computing owing to the capability of difference processing. Since the power value is non-negative, early work only utilized the through port, then the transmission matrix and the output vector are non-negative, thus the matrix operation is limited in the non-negative number domain. However, fundamental mathematical operations such as matrix–vector multiplication and matrix–matrix multiplication are usually performed in the real number domain in practice. In order to extend the matrix operation to the full real number domain, the final results need to be obtained via the differential processing between the power values of the drop port and the through port; in this way, the transmission matrix and final output vector are both able to contain negative domain.

The drop transmission coefficient of the microring situated at the  $i^{\text{th}}$  row and the  $j^{\text{th}}$  column of the microring matrix core is represented by  $m_{ij}$ , and the through transmission coefficient is therefore represented by  $1 - m_{ij}$  without considering the loss of the microring.

Each microring situated at the same column merely modulates the optical signal with a specific wavelength, hence the output of each row contains the power of  $N$  different wavelengths, and  $N$  is numerically equal to the number of microrings positioned in the same row. Based on the microring matrix core model consisting of add-drop microring structure, we assume that the input vector is  $X = [x_1, x_2, \dots, x_N]^T$ , thus the output power of drop ports and through ports of each row can be calculated by Equations (5) and (6), respectively.

$$y_i = \sum_{j=1}^N m_{ij} x_j \quad (5)$$

$$y_i = \sum_{j=1}^N (1 - m_{ij}) x_j \quad (6)$$

the output vector of drop ports  $Y_1 = [y_1, y_2, \dots, y_N]^T$  can be mathematically described as

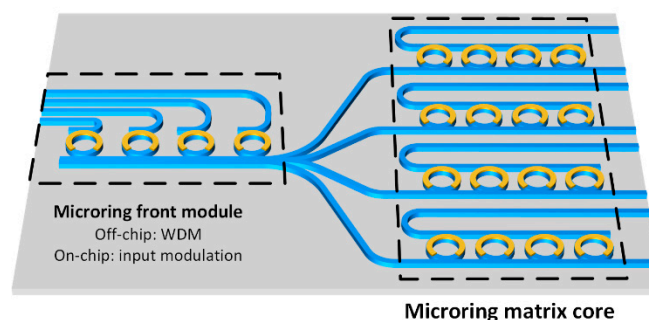
$$Y_1 = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1N} \\ m_{21} & m_{22} & \dots & m_{2N} \\ \dots & \dots & \dots & \dots \\ m_{N1} & m_{N2} & \dots & m_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_N \end{bmatrix} \quad (7)$$

Similarly, the output vector of through ports  $Y_2$  can be written as

$$Y_2 = \begin{bmatrix} 1 - m_{11} & 1 - m_{12} & \dots & 1 - m_{1N} \\ 1 - m_{21} & 1 - m_{22} & \dots & 1 - m_{2N} \\ \dots & \dots & \dots & \dots \\ 1 - m_{N1} & 1 - m_{N2} & \dots & 1 - m_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_N \end{bmatrix} \quad (8)$$

then the final output matrix  $Y$  can be calculated by the differential processing as

$$Y = Y_2 - Y_1 = \begin{bmatrix} 1 - 2m_{11} & 1 - 2m_{12} & \dots & 1 - 2m_{1N} \\ 1 - 2m_{21} & 1 - 2m_{22} & \dots & 1 - 2m_{2N} \\ \dots & \dots & \dots & \dots \\ 1 - 2m_{N1} & 1 - 2m_{N2} & \dots & 1 - 2m_{NN} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_N \end{bmatrix} \quad (9)$$



**Figure 6.** The scheme of WDM-MVM. The microring front module can be positioned off-chip or on-chip, which is designed to serve as a wavelength division multiplexer (off-chip) or to modulate the input vector  $X$  to different wavelengths (on-chip). The microring matrix core is an  $N \times N$ -sized microring array corresponding to an  $N \times N$ -sized matrix.

The microring matrix core was used for on-chip photonic matrix operation early in 2013, when Yang et al. proposed a matrix–vector multiplier based on a microring array and experimentally demonstrated matrix multiplication and weighted interconnection [34]. A “broadcast-and-weight” scheme has been proposed [35,36] and demonstrated [37] to implement large-scale reconfigurable optical interconnections and the integrated optical

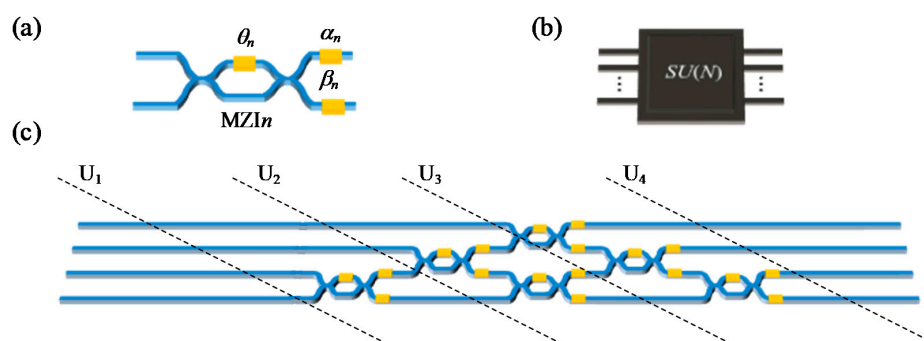
neural network enabled by microring resonators on a silicon photonic chip utilizing WDM technique. In this scheme, each microring plays a role as a tunable filter and only operates at a specific wavelength. Optical signals are modulated in parallel by an array of microrings [38]. The WDM-MVM scheme provides a viable solution to achieve orders of magnitude improvements in both computational speed and energy consumption against existing architectures based on electronic devices. In addition, the microring matrix core can also be used in the linear computation part of the neural networks to achieve the optical acceleration of the ONNs. In 2019, Feldmann et al. presented an all-optical spiking neurosynaptic network successfully realizing pattern recognition directly in the optical domain, in which a microring integrated with phase-change material (PCM) cell is able to control whether to generate an output spike pulse by switching the PCM states to change the optical resonance condition of the microring and its propagation loss [39]. A scalable optical neural network architecture is implemented using the WDM technique. Feldmann et al. improved their architecture and demonstrated an integrated photonic hardware accelerator utilizing optical frequency combs and the WDM technique to achieve parallel convolutional processing [40]. Recently, Xu et al. proposed an optical convolutional neural network architecture based on WDM to accelerate computing speed by utilizing broad optical bandwidth [41], which can be used for various convolutional operations to realize handwritten digits recognition.

#### 4. MZI Matrix Core

As one of the basic photonic devices, MZI has been widely used in optical modulators [42,43], optical communication [44,45], and optical computing [46]. MZI is a natural minimum matrix operation unit, and can be fabricated on a silicon platform to implement the minimum matrix multiplication. The photonic matrix network built by MZIs can be extended to arbitrary matrix multiplication without fundamental loss. In 1994, Reck et al. firstly proposed a general algorithm, the triangular decomposition algorithm, for the design of an experimental realization of any  $n \times n$  unitary matrix [47]. In this case, unitary matrix transforms can be achieved by the architecture consisting of beam splitters, phase shifters, and mirrors arranged according to specific rules.

The structure of MZI is shown in Figure 7a. MZI is composed of two multimode interference couplers and two interference arms [48]. The phase shifts on the internal and external phase shifters of MZI can be expressed as  $\theta_n$ ,  $\alpha_n$ , and  $\beta_n$ , respectively, moreover, these phase shifters can be easily configured and controlled by thermally tuning the heaters. The  $2 \times 2$  unitary transformation matrix of a single MZI can be expressed as a standard  $SU(2)$  rotation matrix:

$$U_{MZIn} = R(n) = \frac{1}{2} \begin{bmatrix} e^{i\alpha_n}(e^{i\theta_n} - 1) & ie^{i\alpha_n}(e^{i\theta_n} + 1) \\ ie^{i\beta_n}(e^{i\theta_n} + 1) & e^{i\beta_n}(1 - e^{i\theta_n}) \end{bmatrix} \quad (10)$$



**Figure 7.** The principle of the MZI matrix core. (a) The structure of one MZI used in an MZI matrix core. (b) The whole MZI mesh consists of MZI matrix cores, and MZIs can be equivalent to a reconfigurable black box. (c) A typical example of a  $4 \times 4$  network structure enabled by an MZI matrix core.



Arbitrary  $n \times n$  unitary transformation matrix  $SU(N)$  can be theoretically decomposed into the product of a series of  $SU(2)$  rotation submatrices, and therefore the whole MZI mesh can be equivalent to a reconfigurable black box as shown in Figure 7b to perform any unitary transformation we desired. A typical example of a  $4 \times 4$  network structure enabled by an MZI matrix core is given in Figure 7c, which is composed of six MZIs, and the matrix transformation relations of each port can be obtained by cutting the plane along the dashed lines. The detailed process can be mathematically described as

$$\begin{aligned} U_2 &= R_{1,1}U_1 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & R(1) \end{bmatrix} U_1 \\ U_3 &= R_{2,1}R_{2,2}U_2 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & R(3) \end{bmatrix} \begin{bmatrix} 1 & & \\ & R(2) & \\ & & 1 \end{bmatrix} U_2 \\ U_4 &= R_{3,1}R_{3,2}R_{3,3}U_3 = \begin{bmatrix} 1 & & \\ & 1 & \\ & & R(6) \end{bmatrix} \begin{bmatrix} 1 & & \\ & R(5) & \\ & & 1 \end{bmatrix} \begin{bmatrix} R(4) & & \\ & 1 & \\ & & 1 \end{bmatrix} U_3 \end{aligned} \quad (11)$$

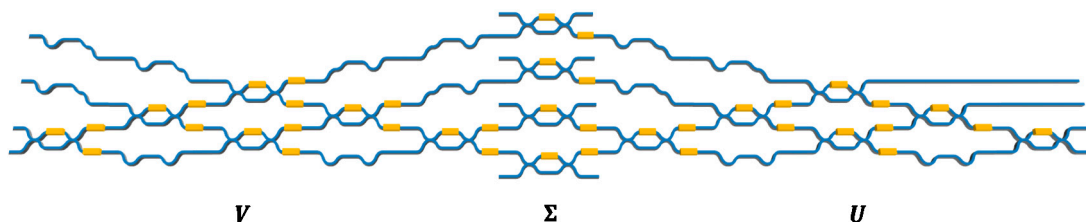
Consequently, the  $4 \times 4$  unitary transformation matrix  $SU(4)$  can be expressed as

$$SU(4) = R_{3,1}R_{3,2}R_{3,3}R_{2,1}R_{2,2}R_{1,1} \quad (12)$$

A similar principle can also be applied to any  $SU(N)$  matrix; in this way, an arbitrary  $n \times n$  unitary matrix can always be decomposed into the product of  $n(n-1)/2$  rotation submatrices

$$SU(N) = R_{N-1,1}R_{N-1,2} \dots R_{N-1,N-1} \dots R_{3,1}R_{3,2}R_{3,3}R_{2,1}R_{2,2}R_{1,1} \quad (13)$$

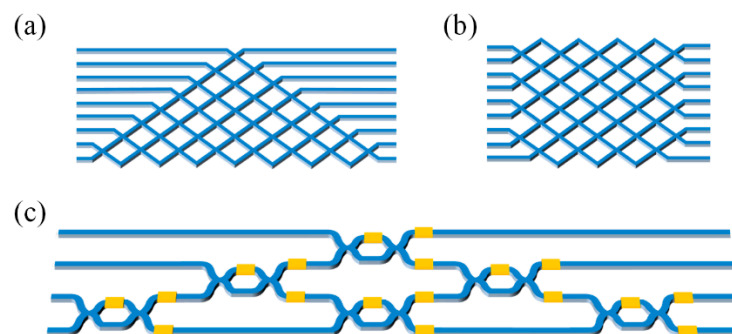
According to Equation (13), we can configure the MZI mesh to mimic the corresponding unitary matrix. Furthermore, when it comes to a general  $n \times n$  matrix, which is not limited in unitary matrix, we know that a general complex-valued matrix  $M$  can be decomposed as  $M = U\Sigma V^\dagger$  utilizing singular value decomposition (SVD) [49], where  $U$  is an  $m \times m$  unitary matrix,  $\Sigma$  is an  $m \times n$  rectangular diagonal matrix, and  $V^\dagger$  is the complex conjugate of the  $n \times n$  unitary matrix  $V$  [50]. Thus, an arbitrary complex-valued matrix network can be decomposed into a unitary MZI mesh, an array of tunable optical attenuators, and another unitary MZI mesh (Figure 8), which can implement  $U$ ,  $\Sigma$ , and  $V^\dagger$ , respectively by tuning phase shifters and attenuators to change the transmission coefficient of each signal channel.



**Figure 8.** MZI mesh can be designed to mimic arbitrary  $n \times n$  complex-valued matrix utilizing SVD.

The above designs are all based on the triangular decomposition algorithm shown in Figure 9a. However, the triangular structure has a large footprint and is not compact enough for highly integrated applications. In 2016, Clements et al. optimized the design on the basis of the triangular decomposition algorithm and proposed a rectangular decomposition scheme (Figure 9b) [51]. The principles of these two schemes are similar for they are both based on rotation submatrices decomposition, but the rectangular scheme is more compact and neater than triangular scheme.

The MZI matrix core has already shown its great potential for accelerating the linear computation part of ONNs and offered the promise to overcome the bottlenecks of state-of-the-art electronics. In 2017, Shen et al. proposed and experimentally demonstrated a coherent optical computing architecture enabled by a cascaded programmable MZI mesh utilizing SVD [50]. This design is capable of remarkably accelerating computing speed and improving power efficiency using coherent light, which makes the MZI matrix core one of the most significant building blocks of ONNs and optical computing acceleration. The following year, Hughes et al. thoroughly discussed the training of ONNs by backpropagation and gradient measurement [52], and this work provided a path toward effectively implementing on-chip training and optimizing reconfigurable integrated optical platforms. By applying on-chip training on the integrated optical platform, Zhou et al. proposed and experimentally demonstrated an all-in-one silicon photonic polarization processor [53,54], a universal matrix computing chip [55], and a self-configuring programmable signal processor [56], etc. Beyond the applications in machine learning, these works may also broaden the access to intelligent optical information processing. In parallel, some recent progress in network structures has been reported, including hexagonal MZI mesh for various filter optical switch signal processing [57–61] and a programmable microwave photonic chip based on a quadrilateral MZI mesh [62], etc. These works further enrich the matrix computation functionalities and versatility of the MZI matrix core. Nevertheless, key issues such as the large footprint of the MZI structure (usually over  $10,000 \mu\text{m}^2$  per interferometer unit) and extra energy consumption of thermo-optic modulation (approximately 10 mW per heater of MZI) limit the application of a large-scale programmable optical neural network to a certain extent.



**Figure 9.** Two decomposition schemes of the MZI matrix core. (a) Triangular [47] and (b) rectangular decomposition schemes [51]. (c) A typical MZI matrix core based on triangular decomposition.

Nano-opto-electro-mechanical systems (NOEMS) are structures designed to maximize both opto-mechanical and electro-mechanical interaction at the nanoscale [63]. Compared to devices based on thermo-optical phase shifters, NOEMS-based devices can work without static power dissipation because mechanical displacements require extra energy only for switching to a different state. Experimental demonstrations of NOEMS-based devices on silicon have been reported using in-plane motion of directional couplers [64], microring [65], and MZI [66], demonstrating the great potential of NOEMS for static and microsecond-scale reconfiguration of integrated photonic circuits and quantum photonic networks. However, NOEMS require higher-precision lithography than conventional devices to ensure the resolution and alignment accuracy of nanophotonic structures, and mechanical systems generally need to be packaged to avoid the impact of the environment.

## 5. Discussion and Outlook

The MVM operations enabled by photonics have a remarkably higher speed and lower energy consumption compared to those of their electronic counter parts, which provides a feasible acceleration solution for the applications of artificial intelligence. On the one hand, on-chip integrated photonic circuits are an ideal platform for artificial

neural networks owing to their high compactness and great potential for competitive integration density. In addition, fast electro-optical modulators and efficient nonlinear optical components built on the LiNbO<sub>3</sub>-on-insulator (LNOI) platform are compatible with silicon photonic circuits [67–69] and provide a promising alternative approach to realize all-optical neural networks on one chip. On the other hand, the MPLC method based on holography can achieve an ultra-large size of MVM operations due to the capability of high parallel processing in free space, and a high model complexity with millions of neurons has already been achieved with the architecture enabled by the MPLC matrix core. The optical AI accelerator provides a hardware platform, which is completely different from the conventional electronic architecture, used to support several universal neural network algorithms to match specific artificial intelligence applications including image recognition, human action recognition, and Google PageRank, etc.

Table 1 summarizes the comparison of different recently demonstrated photonic AI accelerators with well-known analog and digital electronic hardware. The performance parameters of photonic architectures were obtained by theoretically extrapolating from the experimental performance index of a handful of photonic processing units. In complementary metal–oxide–semiconductor (CMOS), MVM operations are typically implemented by systolic arrays [70] or single instruction multiple data (SIMD) units [71]. Due to the properties of electronic components, performing simple operations requires a large number of transistors to work together and an extra scheduler program to coordinate the data movement involved in weights, while MVM operations can be easily implemented by fundamental photonic components such as microring, MZI, and diffractive plane. Therefore, the rate of photonic computing is several orders of magnitude faster than electrons and consumes much less power.

**Table 1.** Comparison of different photonic AI accelerators with well-known analog and digital electronic hardware.

Technology	Computing Density (TMACs/s/mm <sup>2</sup> )	Energy/MAC	Latency	Precision (bits)
MPLC with a reconfigurable diffractive processing unit [19]	-	0.82 fJ/MAC	-	8
Broadcast-and-weight based on WDM [72]	50	2.1 fJ/MAC	<100 ps	5.1+
Photonic WDM/PCM in-memory computing [40] (220 nm SOI platform)	81	17 fJ/MAC	250 ps	5
Optical convolutional accelerator based on WDM [41]	-	1.58 pJ/MAC	-	8
Coherent MZI mesh [50]	0.56	30 fJ/MAC	<100 ps	5.1+
Google TPU (digital) [70]	0.58	0.43 pJ/MAC	1.4 ns	8
Flash (analog) [73]	18	7 fJ/MAC	15 ns	5

However, to turn experimental demonstrations into practical artificial intelligence processors, several key emerging technologies are required to overcome the bottlenecks in computing bandwidth, smart control strategies, and all-optical neural networks, so as to further improve the performance and feasibility of optical computing architectures.

Chip-scale optical frequency combs and high-speed electro-optical modulators are essential when it comes to enlarging computing bandwidth and achieving higher rates. As a type of tailor-made light source, the optical frequency comb provides evenly spaced wavelengths aligned to standardized communication channels. Light has the capability of parallel processing. Combined with WDM technology, an additional dimensional–wavelength can be introduced. In this way, large amounts of data can be independently encoded on different wavelengths and processed in parallel, hence the computing bandwidth of the parallel photonic matrix operation will be increased by dozens of times. LNOI-based electro-optical modulators benefit from outstanding properties such as a strong electro-optical coefficient and their compatibility with silicon photonic devices, offering high modulation frequencies over 100 GHz [69,74] and the on-chip electro-optic modulation desired for nanophotonics.

Currently, electronic technology is already mature for dataflow control, which is difficult for photonic computing. For integrated photonic devices, especially resonant devices, even slight changes in the environment will affect their normal operation, thus the photonic circuits need to be tuned in real time using intelligent control strategies to resist environmental variability, such as temperature and vibration. In addition, a nanostructure fabrication error can lead to random parameter drifts for devices, which cannot be ignored in large-scale array. A commonly used method is to respectively preprogram devices to an ideal default state to compensate for the fabrication errors. Smart algorithms, such as gradient descent and back propagation, are often used to intelligently configure photonic processing units and build some neural network models.

Optical nonlinear activation function and the efficiency of electro-optical conversion are of great significance to construct all-optical neural networks. Nonlinear activation functions enable neural networks to build complex mappings between inputs and outputs. At present, the nonlinear activation function is mainly realized by digital computers, where new optical signals are generated and modulated and then fed to the subsequent layers. Delays and power consumption in electro-optical conversion processes, as well as the rate-constrained I/O ports of traditional processors, result in performance limitations of the ONNs, especially in large matrix dimensions. To solve this challenging problem, optical nonlinear materials, such as 2D materials and PCM, can be developed and integrated with photonic devices to provide a variety of nonlinear responses while avoiding extra latency and loss associated with frequent electro-optical conversion.

In conclusion, optical computing architectures based on integrated photonic circuits and holography have shown great capabilities for high-speed matrix computing and emerging artificial intelligence applications. However, developing general purpose optical computing systems will remain challenging in the foreseeable future, whose high performance can only be achieved through the flexible design combining hardware with software. On the one hand, chip-scale optical frequency combs, high-speed modulators, and new optical materials can be applied to further improve the performance of the hardware, mainly including computing density, speed and latency. On the other hand, intelligent control algorithms are used to solve the challenges of tunability and practicality. At present, optical computing systems have already been used in computer vision, speech recognition, and complex signal processing, and are expected to expand the frontiers of machine learning and information processing applications.

**Author Contributions:** The research topic was initiated by J.D. J.C. worked on it and drafted the review. J.C. and J.D. tabulated the tables and helped with the discussion. H.Z. and J.D. critically reviewed the data and expanded the review with their own experience. Overall supervision of research carried out by J.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was partially supported by the National Natural Science Foundation of China (62075075).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lundstrom, M. Moore's law forever? *Science* **2003**, *299*, 210–211. [[CrossRef](#)] [[PubMed](#)]
2. Shastri, B.J.; Tait, A.N.; Ferreira de Lima, T.; Pernice, W.H.P.; Bhaskaran, H.; Wright, C.D.; Prucnal, P.R. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photonics* **2021**, *15*, 102–114. [[CrossRef](#)]
3. Saade, A.; Caltagirone, F.; Carron, I.; Daudet, L.; Dreameau, A.; Gigan, S.; Krzakala, F. Random projections through multiple optical scattering: Approximating kernels at the speed of light. In Proceedings of the 2016 IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China, 20–25 March 2016; pp. 6215–6219.

4. Liutkus, A.; Martina, D.; Popoff, S.; Chardon, G.; Katz, O.; Lerosey, G.; Gigan, S.; Daudet, L.; Carron, I. Imaging with nature: Compressive imaging using a multiply scattering medium. *Sci. Rep.* **2014**, *4*, 5552. [CrossRef] [PubMed]
5. Lin, X.; Rivenson, Y.; Yardimej, N.T.; Veli, M.; Luo, Y.; Jarrahi, M.; Ozcan, A. All-optical machine learning using diffractive deep neural networks. *Science* **2018**, *361*, 1004–1008. [CrossRef]
6. Chang, J.; Sitzmann, V.; Dun, X.; Heidrich, W.; Wetzstein, G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Sci. Rep.* **2018**, *8*, 12324. [CrossRef] [PubMed]
7. Cartwright, S. New optical matrix-vector multiplier. *Appl. Opt.* **1984**, *23*, 1683–1684. [CrossRef]
8. Bocker, R.P.; Clayton, S.R.; Bromley, K. Electrooptical matrix multiplication using the twos complement arithmetic for improved accuracy. *Appl. Opt.* **1983**, *22*, 2019–2021.
9. Athale, R.A.; Collins, W.C. Optical matrix-matrix multiplier based on outer product decomposition. *Appl. Opt.* **1982**, *21*, 2089–2090. [CrossRef]
10. Shuiying, X.; Yanan, H.; Ziwei, S.; Xingxing, G.; Yahui, Z.; Zhenxing, R.; Suhong, W.; Yuanting, M.; Weiwen, Z.; Bowen, M.; et al. A review: Photonics devices, architectures, and algorithms for optical neural computing. *J. Semicond.* **2021**, *42*, 023105.
11. Li, X.; Zhang, G.; Huang, H.H.; Wang, Z.; Zheng, W. Performance Analysis of GPU-based Convolutional Neural Networks. In Proceedings of the 45th International Conference on Parallel Processing-ICPP, Philadelphia, PA, USA, 16–19 August 2016; pp. 67–76.
12. Goodman, J.W.; Dias, A.R.; Woody, L.M. Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms. *Opt. Lett.* **1978**, *2*, 1–3. [CrossRef]
13. Veli, M.; Mengu, D.; Yardimci, N.T.; Luo, Y.; Li, J.; Rivenson, Y.; Jarrahi, M.; Ozcan, A. Terahertz pulse shaping using diffractive surfaces. *Nat. Comm.* **2021**, *12*, 37. [CrossRef] [PubMed]
14. Wen, H.; Liu, H.; Zhang, Y.; Zhang, P.; Li, G. Mode demultiplexing hybrids for mode-division multiplexing coherent receivers. *Photon. Res.* **2019**, *7*, 917–925. [CrossRef]
15. Wen, H.; Liu, H.; Zhang, Y.; Sampson, R.; Fan, S.; Li, G. Scalable Hermite-Gaussian mode-demultiplexing hybrids. *Opt. Lett.* **2020**, *45*, 2219–2222. [CrossRef] [PubMed]
16. Fontaine, N.K.; Ryf, R.; Chen, H.; Neilson, D.T.; Kim, K.; Carpenter, J. Laguerre-Gaussian mode sorter. *Nat. Comm.* **2019**, *10*, 1865. [CrossRef] [PubMed]
17. Bueno, J.; Maktoobi, S.; Froehly, L.; Fischer, I.; Jacquot, M.; Larger, L.; Brunner, D. Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* **2018**, *5*, 756–760. [CrossRef]
18. Li, J.; Mengu, D.; Luo, Y.; Rivenson, Y.; Ozcan, A. Class-specific differential detection in diffractive optical neural networks improves inference accuracy. *Adv. Photon.* **2019**, *1*, 046001. [CrossRef]
19. Zhou, T.; Lin, X.; Wu, J.; Chen, Y.; Xie, H.; Li, Y.; Fan, J.; Wu, H.; Fang, L.; Dai, Q. Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat. Photonics* **2021**, *15*, 367–373. [CrossRef]
20. Labroille, G.; Denolle, B.; Jian, P.; Genevaux, P.; Treps, N.; Morizur, J.-F. Efficient and mode selective spatial mode multiplexer based on multi-plane light conversion. *Opt. Express* **2014**, *22*, 15599–15607. [CrossRef] [PubMed]
21. Sakamaki, Y.; Hashimoto, T.; Takahashi, H. New optical waveguide design based on wavefront matching method. *J. Lightwave Technol.* **2007**, *25*, 3511–3518. [CrossRef]
22. Wei, H.; Huang, G.; Wei, X.; Sun, Y.; Wang, H. Comment on “All-optical machine learning using diffractive deep neural networks”. *arXiv* **2018**, arXiv:1809.08360. Available online: <https://arxiv.org/abs/1809.08360> (accessed on 22 June 2021).
23. Rahman, M.S.S.; Li, J.X.; Mengu, D.; Rivenson, Y.; Ozcan, A. Ensemble learning of diffractive optical networks. *Light-Sci. Appl.* **2021**, *10*, 14. [CrossRef]
24. Tang, R.; Tanemura, T.; Nakano, Y. Integrated reconfigurable unitary optical mode converter using MMI couplers. *IEEE Photonics Technol. Lett.* **2017**, *29*, 971–974. [CrossRef]
25. Saygin, M.Y.; Kondratyev, I.V.; Dyakonov, I.V.; Mironov, S.A.; Straupe, S.S.; Kulik, S.P. Robust architecture for programmable universal unitaries. *Phys. Rev. Lett.* **2020**, *124*, 010501. [CrossRef] [PubMed]
26. Xu, Q.F.; Schmidt, B.; Pradhan, S.; Lipson, M. Micrometre-scale silicon electro-optic modulator. *Nature* **2005**, *435*, 325–327. [CrossRef] [PubMed]
27. Dong, P.; Qian, W.; Liang, H.; Shafiiha, R.; Feng, N.-N.; Feng, D.; Zheng, X.; Krishnamoorthy, A.V.; Asghari, M. Low power and compact reconfigurable multiplexing devices based on silicon microring resonators. *Opt. Express* **2010**, *18*, 9852–9858. [CrossRef] [PubMed]
28. Xu, S.; Wang, J.; Zou, W. Optical patching scheme for optical convolutional neural networks based on wavelength-division multiplexing and optical delay lines. *Opt. Lett.* **2020**, *45*, 3689–3692. [CrossRef]
29. Xu, S.F.; Wang, J.; Zou, W.W. Optical convolutional neural network with WDM-based optical patching and microring weighting banks. *IEEE Photonics Technol. Lett.* **2021**, *33*, 89–92. [CrossRef]
30. Xu, L.; Hou, J.; Tang, H.; Yu, Y.; Yu, Y.; Shu, X.; Zhang, X. Silicon-on-insulator-based microwave photonic filter with widely adjustable bandwidth. *Photon. Res.* **2019**, *7*, 110–115. [CrossRef]
31. Liu, X.; Yu, Y.; Tang, H.; Xu, L.; Dong, J.; Zhang, X. Silicon-on-insulator-based microwave photonic filter with narrowband and ultrahigh peak rejection. *Opt. Lett.* **2018**, *43*, 1359–1362. [CrossRef] [PubMed]
32. Zhao, Y.; Wang, X.; Gao, D.; Dong, J.; Zhang, X. On-chip programmable pulse processor employing cascaded MZI-MRR structure. *Front. Optoelectron.* **2019**, *12*, 148–156. [CrossRef]

33. Qiang, Z.; Zhou, W.; Soref, R.A. Optical add-drop filters based on photonic crystal ring resonators. *Opt. Express* **2007**, *15*, 1823–1831. [[CrossRef](#)] [[PubMed](#)]
34. Yang, L.; Zhang, L.; Ji, R. On-chip optical matrix-vector multiplier. In *Optics and Photonics for Information Processing VII, Proceeding of the SPIE Optical Engineering + Applications, San Diego, CA, USA, 26 September 2013*; Ifttekharuddin, K.M., Awwal, A.A.S., Márquez, A., Eds.; SPIE: Bellingham, WA, USA, 2013; Volume 8855.
35. Tait, A.N.; Wu, A.X.; Lima, T.F.d.; Zhou, E.; Shastri, B.J.; Nahmias, M.A.; Prucnal, P.R. Microring weight banks. *IEEE J. Sel. Top. Quantum Electron.* **2016**, *22*, 312–325. [[CrossRef](#)]
36. Tait, A.N.; Nahmias, M.A.; Shastri, B.J.; Prucnal, P.R. Broadcast and weight: An integrated network for scalable photonic spike processing. *J. Lightwave Technol.* **2014**, *32*, 4029–4041. [[CrossRef](#)]
37. Tait, A.N.; de Lima, T.F.; Zhou, E.; Wu, A.X.; Nahmias, M.A.; Shastri, B.J.; Prucnal, P.R. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **2017**, *7*, 7430. [[CrossRef](#)]
38. Tait, A.N.; de Lima, T.F.; Nahmias, M.A.; Shastri, B.J.; Prucnal, P.R. Continuous calibration of microring weights for analog optical networks. *IEEE Photonics Technol. Lett.* **2016**, *28*, 887–890. [[CrossRef](#)]
39. Feldmann, J.; Youngblood, N.; Wright, C.D.; Bhaskaran, H.; Pernice, W.H.P. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* **2019**, *569*, 208–214. [[CrossRef](#)] [[PubMed](#)]
40. Feldmann, J.; Youngblood, N.; Karpov, M.; Gehring, H.; Li, X.; Stappers, M.; Le Gallo, M.; Fu, X.; Lukashchuk, A.; Raja, A.S.; et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **2021**, *589*, 52–58. [[CrossRef](#)]
41. Xu, X.; Tan, M.; Corcoran, B.; Wu, J.; Boes, A.; Nguyen, T.G.; Chu, S.T.; Little, B.E.; Hicks, D.G.; Morandotti, R.; et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **2021**, *589*, 44–51. [[CrossRef](#)] [[PubMed](#)]
42. Liu, A.S.; Jones, R.; Liao, L.; Samara-Rubio, D.; Rubin, D.; Cohen, O.; Nicolaescu, R.; Paniccia, M. A high-speed silicon optical modulator based on a metal-oxide-semiconductor capacitor. *Nature* **2004**, *427*, 615–618. [[CrossRef](#)]
43. Liao, L.; Samara-Rubio, D.; Morse, M.; Liu, A.S.; Hodge, D.; Rubin, D.; Keil, U.D.; Franck, T. High speed silicon Mach-Zehnder modulator. *Opt. Express* **2005**, *13*, 3129–3135. [[CrossRef](#)]
44. Su, T.; Zhang, M.; Chen, X.; Zhang, Z.; Liu, M.; Liu, L.; Huang, S. Improved 10-Gbps uplink transmission in WDM-PON with RSOA-based colorless ONUs and MZI-based equalizers. *Opt. Laser Technol.* **2013**, *51*, 90–97. [[CrossRef](#)]
45. Song, J.-I.; Song, H.-J. Simultaneous frequency conversion technique utilizing an SOA-MZI for full-duplex WDM radio over fiber applications. *IEICE Trans. Electron.* **2007**, *E90C*, 351–358. [[CrossRef](#)]
46. Shokraneh, F.; Geoffroy-Gagnon, S.; Nezami, M.S.; Liboiron-Ladouceur, O. A single layer neural network implemented by a  $4 \times 4$  MZI-based optical processor. *IEEE Photonics J.* **2019**, *11*, 1–12. [[CrossRef](#)]
47. Reck, M.; Zeilinger, A.; Bernstein, H.J.; Bertani, P. Experimental realization of any discrete unitary operator. *Phys. Rev. Lett.* **1994**, *73*, 58–61. [[CrossRef](#)] [[PubMed](#)]
48. Thomson, D.J.; Hu, Y.; Reed, G.T.; Fedeli, J.-M. Low Loss MMI Couplers for high performance MZI modulators. *IEEE Photonics Technol. Lett.* **2010**, *22*, 1485–1487. [[CrossRef](#)]
49. Lawson, C.L.; Hanson, R.J. *Solving Least Squares Problems*; SIAM: Philadelphia, PA, USA, 1995.
50. Shen, Y.; Harris, N.C.; Skirlo, S.; Prabhu, M.; Baehr-Jones, T.; Hochberg, M.; Sun, X.; Zhao, S.; Larochelle, H.; Englund, D.; et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **2017**, *11*, 441–446. [[CrossRef](#)]
51. Clements, W.R.; Humphreys, P.C.; Metcalf, B.J.; Kolthammer, W.S.; Walmsley, I.A. Optimal design for universal multiport interferometers. *Optica* **2016**, *3*, 1460–1465. [[CrossRef](#)]
52. Hughes, T.W.; Minkov, M.; Shi, Y.; Fan, S. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* **2018**, *5*, 864–871. [[CrossRef](#)]
53. Zhou, H.; Zhao, Y.; Wei, Y.; Li, F.; Dong, J.; Zhang, X. All-in-one silicon photonic polarization processor. *Nanophotonics* **2019**, *8*, 2257–2267. [[CrossRef](#)]
54. Zhou, H.; Zhao, Y.; Wei, V.; Li, F.; Dong, J.; Zhang, X. Multipurpose photonic polarization processor chip. In Proceedings of the 2019 Asia Communications and Photonics Conference, Chengdu, China, 2–5 November 2019; pp. 1–2.
55. Zhou, H.; Zhao, Y.; Xu, G.; Wang, X.; Tan, Z.; Dong, J.; Zhang, X. Chip-scale optical matrix computation for pagerank algorithm. *IEEE J. Sel. Top. Quantum Electron.* **2020**, *26*, 1–10. [[CrossRef](#)]
56. Zhou, H.; Zhao, Y.; Wang, X.; Gao, D.; Dong, J.; Zhang, X. Self-configuring and reconfigurable silicon photonic signal processor. *ACS Photonics* **2020**, *7*, 792–799. [[CrossRef](#)]
57. Perez-Lopez, D.; Sanchez, E.; Capmany, J. Programmable true time delay lines using integrated waveguide meshes. *J. Lightwave Technol.* **2018**, *36*, 4591–4601. [[CrossRef](#)]
58. Perez, D.; Gasulla, I.; Fraile, F.J.; Crudgington, L.; Thomson, D.J.; Khokhar, A.Z.; Li, K.; Cao, W.; Mashanovich, G.Z.; Capmany, J. Silicon photonics rectangular universal interferometer. *Laser Photonics Rev.* **2017**, *11*, 1700219. [[CrossRef](#)]
59. Perez, D.; Gasulla, I.; Crudgington, L.; Thomson, D.J.; Khokhar, A.Z.; Li, K.; Cao, W.; Mashanovich, G.Z.; Capmany, J. Multipurpose silicon photonics signal processor core. *Nat. Comm.* **2017**, *8*, 636. [[CrossRef](#)]
60. Perez, D.; Gasulla, I.; Capmany, J. Field-programmable photonic arrays. *Opt. Express* **2018**, *26*, 27265–27278. [[CrossRef](#)]
61. Perez, D.; Gasulla, I.; Capmany, J. Programmable multifunctional integrated nanophotonics. *Nanophotonics* **2018**, *7*, 1351–1371. [[CrossRef](#)]
62. Zhuang, L.; Roeloffzen, C.G.H.; Hoekman, M.; Boller, K.-J.; Lowery, A.J. Programmable photonic signal processor chip for radiofrequency applications. *Optica* **2015**, *2*, 854–859. [[CrossRef](#)]

63. Midolo, L.; Schliesser, A.; Fiore, A. Nano-opto-electro-mechanical systems. *Nat. Nanotechnol.* **2018**, *13*, 11–18. [[CrossRef](#)]
64. Akihama, Y.; Hane, K. Single and multiple optical switches that use freestanding silicon nanowire waveguide couplers. *Light-Sci. Appl.* **2012**, *1*, e16. [[CrossRef](#)]
65. Takahashi, K.; Kanamori, Y.; Kokubun, Y.; Hane, K. A wavelength-selective add-drop switch using silicon microring resonator with a submicron-comb electrostatic actuator. *Opt. Express* **2008**, *16*, 14421–14428. [[CrossRef](#)] [[PubMed](#)]
66. Poot, M.; Tang, H.X. Broadband nanoelectromechanical phase shifting of light on a chip. *Appl. Phys. Lett.* **2014**, *104*, 061101. [[CrossRef](#)]
67. Wang, C.; Zhang, M.; Stern, B.; Lipson, M.; Loncar, M. Nanophotonic lithium niobate electro-optic modulators. *Opt. Express* **2018**, *26*, 1547–1555. [[CrossRef](#)] [[PubMed](#)]
68. Mercante, A.J.; Yao, P.; Shi, S.; Schneider, G.; Murakowski, J.; Prather, D.W. 110 GHz CMOS compatible thin film LiNbO<sub>3</sub> modulator on silicon. *Opt. Express* **2016**, *24*, 15590–15595. [[CrossRef](#)] [[PubMed](#)]
69. He, M.; Xu, M.; Ren, Y.; Jian, J.; Ruan, Z.; Xu, Y.; Gao, S.; Sun, S.; Wen, X.; Zhou, L.; et al. High-performance hybrid silicon and lithium niobate Mach-Zehnder modulators for 100 Gbit s<sup>-1</sup> and beyond. *Nat. Photonics* **2019**, *13*, 359–364. [[CrossRef](#)]
70. Jouppi, N.P.; Young, C.; Patil, N.; Patterson, D.; Agrawal, G.; Bajwa, R.; Bates, S.; Bhatia, S.; Boden, N.; Borchers, A.; et al. In-datacenter performance analysis of a tensor processing unit. In Proceedings of the 44th Annual International Symposium on Computer Architecture, Toronto, ON, Canada, 24–28 June 2017; pp. 1–12.
71. Fowers, J.; Ovtcharov, K.; Papamichael, M.; Massengill, T.; Liu, M.; Lo, D.; Alkalay, S.; Haselman, M.; Adams, L.; Ghandi, M.; et al. A Configurable Cloud-Scale DNN Processor for Real-Time AI. In Proceedings of the 2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture, Los Angeles, CA, USA, 1–6 June 2018; pp. 1–14.
72. Tait, A.N.; de Lima, T.F.; Nahmias, M.A.; Miller, H.B.; Peng, H.-T.; Shastri, B.J.; Prucnal, P.R. Silicon photonic modulator neuron. *Phys. Rev. Applied* **2019**, *11*, 064043. [[CrossRef](#)]
73. Mahmoodi, M.R.; Strukov, D. An ultra-low energy internally analog, externally digital vector-matrix multiplier based on NOR flash memory technology. In Proceedings of the 55th Annual Design Automation Conference, San Francisco, CA, USA, 24–29 June 2018; pp. 1–6.
74. Wang, C.; Zhang, M.; Chen, X.; Bertrand, M.; Shams-Ansari, A.; Chandrasekhar, S.; Winzer, P.; Loncar, M. Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages. *Nature* **2018**, *562*, 101–104. [[CrossRef](#)] [[PubMed](#)]