

Received March 15, 2020, accepted April 6, 2020, date of publication April 13, 2020, date of current version April 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2987333

A Review of Optical Neural Networks

XIUBAO SUI¹, QIUHAO WU¹, JIA LIU¹, QIAN CHEN¹, AND GUOHUA GU¹

School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Corresponding author: Xiubao Sui (sxbhandsome@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 11773018 and Grant 61727802, in part by the Key Research and Development programs in Jiangsu China under Grant BE2018126, in part by the Fundamental Research Funds for the Central Universities under Grant 30919011401 and Grant 30920010001, and in part by the Leading Technology of Jiangsu Basic Research Plan under Grant BK20192003.

ABSTRACT Optical neural network can process information in parallel by using the technology based on free-space and integrated platform. Over the last half century, the development of integrated circuits has been limited by Moore's law. We know that neural network is based on the digital computer for successive calculation, most of which cannot be made into real-time processing system. Therefore, it is necessary to develop ONN for real-time processing and device miniaturization. In this paper, we review the progress of optical neural networks. Firstly, based on the principle of artificial neural networks, we elaborate the essence of optical matrix multiplier for linear operation. Then we introduce the optical neural network achieved by free-space optical interconnection and waveguide optical interconnection. Finally we talk about the nonlinearity in optical neural networks. With the gradual maturity of nanotechnology and the rapid advancement of silicon photonic integrated circuits, the progress of integrated photonic neural network has been promoted. Therefore, the construction of optical neural network on the future integrated photonic platform has potential application value.

INDEX TERMS Optical neural networks, optical waveguides, free-space optical interconnection, optical non-linearities, optical elements, photonic integrated circuits.

I. INTRODUCTION

The concept of artificial neural networks (ANNs) originated in 1943, when Walter Pitts [1] first created a model based on mathematics and algorithms that simulated the principles and processes of biological nerve cells [2], which proved to be able to perform logical functions, opening the era of ANNs' research. Since the development of more than half a century, neural network has made great progress and achieved extensive application, with the abilities of self-learning [3], associative storage [4], high-speed searching for optimal solutions [5], strong nonlinear fitting [6], [7], and mapping arbitrarily complex nonlinear relations [7]. However, as a model of computational network inspired by brain signal processing, ANNs are based on von neumann architecture and implemented electronically [8], which has some inherent defects. For example, electronic signals are easy to interfere with each other [9], which brings certain difficulties to neural networks that need high-density connections. In addition, the demand of energy is too high, resulting in large computing costs

[10]. Such flaws in electronics have led to a shift in focus from computers to so-called "light brains", which replace electrons or electricity with light to process large amounts of information at high speeds. Its basic component is a spatial light modulator (SLM) [11], which can convert incoherent light into coherent light. In the optical processing of images, coherent light is easier to be modulated. In addition, the optical computer adopts the optical interconnection technology [12], [13] which connects the part of computing and storage, breaking through the traditional architecture of connecting the arithmetic unit, memory, input and output devices by bus [14]. What we're going to talk about is a hot topic in optical computing, building neural networks using optical methods. Refers to optical neural network (ONN), it has a large number of linear layers which interconnect complexly. Optical interconnection is highly parallel, the beam can cross in the space without crosstalk, and the propagation velocity of light is very high, so the time delay and dispersion are negligible [15], [16]. In 1978, J.W. Goodman of Stanford university in the United States firstly proposed the theoretical model of optical vector-matrix multiplier [17]. This became an important step of optical computing [18], [19] which promoted the

The associate editor coordinating the review of this manuscript and approving it for publication was Vivek Kumar Sehgal¹.

research on optical matrix multiplier (OMM) [20], [21] and the development of photonic neural networks. With the rapid advancement of integrated photonic technology [22], [23], and the hardware implementation of nano-photon processors [24] building neural networks based on integrated photonic platform has become more and more popular.

In this review, we describe the progress of ONN over the last half century. The structure of this review is as follows. In Sec. II, we start from the basic theory of ANNs, explain the principle of OMM and optical interconnection, and emphasize the significance of nonlinear threshold operation. We know, optical interconnection is divided into free-space optical interconnection (FSOI) and waveguide optical interconnection (WOI). Thus, in Sec. III, we introduce the ONN realized by SLM, micro-lens array (MLA) and holographic element (HOE) from the perspective of FSOI, and then in terms of WOI, we introduce the all-optical neural network (AONN) based on fiber optics and integrated photonic technology in Sec. IV. Furthermore, we elaborate the optical nonlinearity in Sec. V. Finally, Sec. VI gives the summary and outlook of ONN.

II. THE PRINCIPLE OF ONN

A. THE WORKING MECHANISM OF ANNs

ANNs are constructed by the inspiration of biological neural networks. Biological neurons have four basic units: dendrites, axons, cell bodies and synapses [25], as shown in Fig. 1(a). A neuron usually has multiple dendrites that receive incoming information when it receives a stimulus. Axons carry information from this neuron to others. There is only one axon with a number of terminals at its end, which can carry information to multiple other neurons. The axon terminals connect to the dendrites of other neurons, and then transmit signals. As shown in Fig. 1(b), similar to the biological neurons, artificial neurons are information processing unit, which consist of the input signals, the weight W_i , bias b , linear summation, and threshold function Φ [26]. The interconnection of signals represents axons and dendrites (input is the dendrite, output is the axon), the threshold and linear summation are equal to the activation of the cell body, weights and memories correspond to synapses. As shown in Fig. 1(c), in feed-forward neural network [27], the input x is sent to the artificial neurons, each corresponding to a weight, then we weight and sum the input, perform the nonlinear activation, finally we obtain the output S . The purpose of training is to get the output close to the ground truth. It constantly updated to optimize W and b in the process of training until the network is converged. In the above statement, the form of S is as follows:

$$S = \Phi\left(\sum_{i=1}^n W_i * X_i + b_i\right). \quad (1)$$

B. OPTICAL REALIZATION OF LINEAR OPERATION

Linear operations in neural networks involve a large amount of matrix multiplications. The design of ONN needs to con-

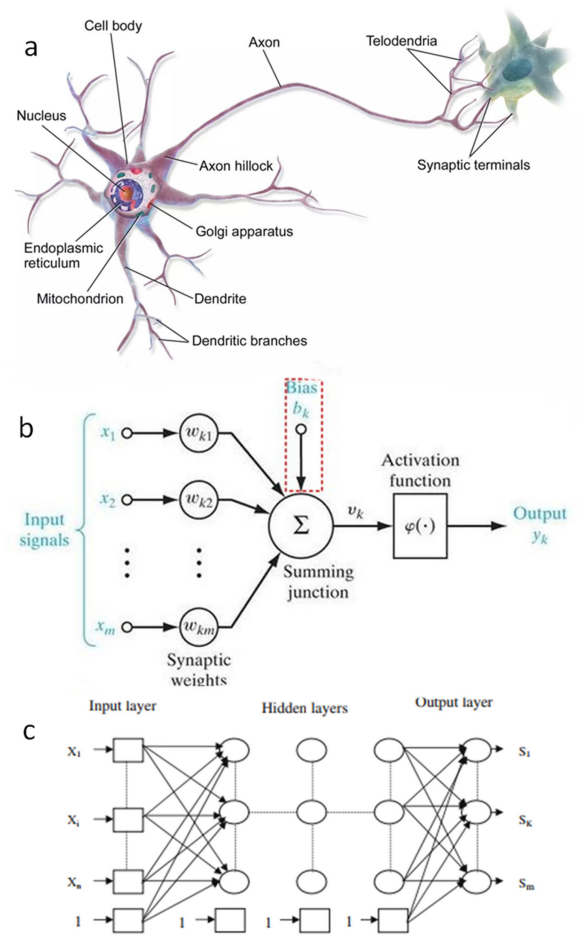


FIGURE 1. Mechanism of ANNs. (a) Structure of a biological neuron [25]. (b) Structure of artificial neuron [26]. (c) The model of feed-forward neural network [27].

sider using optical methods to achieve such linear operations. OMM can quickly realize optical interconnection. Next, we will explain the principle of matrix-vector and matrix-matrix multiplication.

In Fig. 2(a), optical vector-matrix multiplier is composed of an array of LED light source, collimating lens L , column lens $CL1$, SLM, column lens $CL2$ and CCD [28]. Its working principle is as follows: suppose the vector $b=(1,0,2)$, matrix

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \text{ By controlling the luminous intensity of the}$$

LED light source to represent vector b , and matrix A is loaded by SLM, beams pass through the collimating lens L and shine parallel light on the $CL1$, then reach the SLM, complete multiplication with matrix A , again after lens $CL2$, focus on the probe CCD which is vertically placed. The intensity on the CCD can be detected to determine whether the obtained vector is $(3,2,3)$.

Principle of matrix - matrix multiplier:

Suppose the magnitude of matrix A is m by p , and the magnitude of matrix B is p by n , then the magnitude of matrix

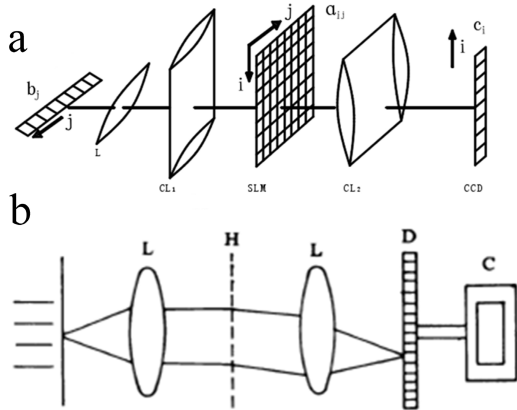


FIGURE 2. Optical linear operation. (a) An implementation of OMM [28].(b) Optical 4f system [29].

C is m by n .

$$\begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & \vdots & \vdots \\ a_{m1} & \cdots & a_{mp} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots \\ b_{p1} & \cdots & b_{pn} \end{pmatrix} = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \vdots & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} \quad (2)$$

The essence of multiplying A and B is to take the inner product of all the row vectors of A and all the column vectors of B. According to the convolution theorem [30], considered separately two vector convolution between alpha and beta, we can regard the calculation process as one of the vector alpha which is centered on the vertical axis rotate a degree of 180, and then moving in the direction of beta in turn. For each unit we move, we multiply and add the entries of vectors until the two vectors no longer intersect. The essence of multiplying matrices by matrices is that if you transpose all the vectors of one matrix, invert them, and then convolve them with another matrix, from the result of the convolution, you get the mathematical result of multiplying matrices. Take any row vector α_i in matrix A convolved with matrix B to get D as an example: transpose and invert the row vector of A.

$$\begin{pmatrix} a_{ip} \\ \vdots \\ a_{i1} \end{pmatrix} \cdot \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots \\ b_{p1} & \cdots & b_{pn} \end{pmatrix} = \begin{pmatrix} d_{11} & \cdots & d_{1n} \\ \vdots & \vdots & \vdots \\ d_{p1} & \cdots & d_{pn} \\ \vdots & \vdots & \vdots \\ d_{(2p-1)1} & \cdots & d_{(2p-1)n} \end{pmatrix} \quad (3)$$

The row number of matrix D is $2p-1$, and the column number is n . It can be found that the most middle row element of matrix D is the mathematical result of vector α_i and matrix B. The matrix A is multiplied by the matrix B, and the matrix A can be segmented to obtain m row vectors, which are transposed and disordered respectively, and the 0's padding of sufficient columns are added to obtain m submatrices, each of which is $\begin{pmatrix} a_{ip} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & 0 & \cdots & 0 \end{pmatrix}$. The m submatrix is arranged

horizontally in turn to form a new matrix to convolute with B, and the result is $(2p-1)$ rows and $(mn+n-1)$ columns. Take the former mn column elements of the middle row of the result matrix and arrange them into matrix $C_{m \times n}$ in turn, and the resulting matrix C is the mathematical result of multiplying matrix A and B.

According to the above principle, the optical 4f system can be built to realize the matrix multiplier [29]. The matrix B is put at the object surface in the form of input image, and the matrix A is constructed into an optical matrix multiplication filter $h(x,y)$, which is converted to the frequency domain by Fourier transform to obtain $H(u,v)$. $H(u,v)$ is placed at the frequency spectrum plane, the image at the object plane is multiplied by the matrix A, after passing through the Fourier transform lens L1. The result of matrix convolution can be obtained by inverting the Fourier transform, and the mathematical result of matrix multiplication can be obtained by taking part of the elements in the middle row.

C. THE NECESSITY OF NONLINEAR ACTIVATION

The deep neural network can be realized by adding hidden layers. In fact, if deep neural networks do not have nonlinear activation, hidden layers in the network are equivalent to a single linear layer, so it is impossible to learn and identify the nonlinear model, therefore the nonlinearity in neural networks is crucial. The lack of direct and effective nonlinearity in optics seriously limits its role in deep learning calculation. In recent years, studies on optical nonlinearity activation function emerge endlessly. We will discuss it in Sec. V.

So, the implementation of ONN needs to consider optical linear operation and nonlinear activation. Optical linear operation can be realized by FSOI and WOI, among which the weight of optical interconnection can be realized by SLM, holographic element, 4f system, phase change material (PCM) [31],etc. Optical nonlinear activation is indispensable, which can be achieved by using saturable absorber, Kerr medium, and nonlinear optical thermal effect [32].

III. USE FSOI TO IMPLEMENT ONN

FSOI refers to the interconnection mode in which the optical signal propagates in free space after being emitted from the emitter, and finally reaches the receiver after changing the optical path and controlling the beam through some optical components such as lens and gratings. Linear operations in ONN can be implemented through FSOI. The common elements in the FSOI to realize ONN are SLM, MLA and HOE.

A. USE SLM TO BUILD ONN

It is well known that ONN originated in the 1980s, and its mathematical model was first proposed by Hopfield in 1982 [33]. During this period, emerging optical technologies such as optical bistability can perform threshold operations, which were necessary for the realization of this model. SLM is a device that modulates the spatial distribution of light waves and can modulate the beam in real time [11]. It is a

key component of real-time optical information processing, optical computing and other systems. Next, we will introduce how to use SLM to realize optical interconnection and build ONN.

In 1985, D.Pasltis and H.Farhat first reported the Hopfield neural network (HNN) based on optical vector-matrix multiplier, namely the external product optical associative storage ONN system [34], which could improve the accuracy and robustness of optical information processing. However, such a system has problems with synapse saturation in the hardware associated with the connection matrix, and the network converges slowly. In the following year, H.K.Liu *et al.* introduced the inner product optical array processor [35], which did not need to calculate the outer product interconnection matrix in advance, and the dimension of the matrix was reduced in the process of calculation, so the convergent speed of the network was faster than that of the outer product. Neither the outer product nor the inner product ONN is too large in spacial size. Although it shows certain advantages, it is not conducive to its large-scale promotion. Benefit from the initial development of integrated optics, in 1989, Ohta *et al.* designed a miniaturized integrated chip based on ONN, using dynamic integration of SLM mask instead of the static mask [36]. It can simulate a system with 32 neurons, preliminary realized the associative memory based on neural network. This is the first time that integrated optical neural networks have been developed. Furthermore, Ying Zuo *et al.* demonstrates an AONN with adjustable linear operations using SLM and Fourier lenses to achieve such linear process [37]. In the process of linear operation, the power of incident light in different regions of SLM represents different input nodes. Through SLM and Fourier lens, all the diffracted beams are superimposed in the same direction and converge to a point on its focal plane to achieve linear summation. Based on this, AONN can obtain linear matrix elements through certain iterative feedback algorithm, and achieve a suitable precision.

B. USE MLA TO BUILD ONN

MLA is an array of lens with aperture of light and relief depth of micron. A complete laser wavefront can be divided into many tiny parts in space, and each part is focused by the corresponding small lens. On the focal plane, a series of microlens can obtain a plane composed of a series of focal points. Based on this, the modulation of incident light wave can be realized, and the spots in different directions can be convergent into points to achieve linear summation. Next, we will introduce the utilization of MLA to realize optical interconnection and ONN.

In 1993, Yasunori Kuratomi proposed an ONN with vector feature extraction, and innovated vector feature extraction equipment (FEOND) [38]. A 2×2 small lens array is used in the network to image the same four regions of the reticular pattern and focus these images on the FEOND, fulfilling the function of the first hidden layer. The network can correctly recognize handwritten letters, which is easy to implement

in optics, but it has neither scale invariance nor shift invariance, and has limit of linear separability. If the position or size of input letters changes greatly, the vector's characteristics will also change. In addition, Taiwei Lu introduced a two-dimensional hybrid ONN using a high-resolution video monitor as programmable associative memory [39], which improved the Hopfield model, realized optical interconnection and memory storage by using MLA, and had strong self-adaptiveness. However, due to the limitations of imaging aberration and light detection, the number of neurons placed in the structure is severely limited. Based on this, Jianwen Yang's team proposed an ONN that utilized lens arrays and liquid crystal light valves (LCLV) to subtract in parallel, the subtraction between the positive and negative of the simulated neuron is transformed into the addition of the normal and antistate of LCLV [40]. It is difficult to process the state of 2d input neuron because the dynamic threshold related to inputs is established on a single mask. In order to solve the problem of aberration and process the input neurons in two dimensions, Yang's team put forward an experimental device with 32×32 neurons [41]. The interconnection of large scale ONN is realized by coaxial lens array, aberration and the light efficiency are obviously improved [42].

C. BUILD ONN WITH HOE

HOE is an optical element made according to the principle of holography, which is usually made on a photosensitive thin film material [43], [44]. It acts on the principle of diffraction and is a diffractive optical element (DOE).

In 1990, D.Pasltis *et al.* proposed holography in the neural network [45], which uses the conjugate mirror to increase the threshold, feedback and gain of the pattern in the hologram reading process to realize the model of the neural network, as shown in Fig. 3(a). The information is encoded and distributed in the recording medium. Two - dimensional input and output planes are connected by interference fringes, and the pixels on the input and output plane are equivalent to neurons. In 2004, Sheng L. Yeh proposed a new optical method to realize HNN by using matrix gratings. Fig. 3(b) shows the structure of matrix gratings. The experimental results prove the effectiveness of this method and the stability of optical efficiency of different optical elements in ONN [46]. In 2018, Lin Xing *et al.* innovatively invented the diffraction optical neural network [47], and the team made an in-depth analysis of this architecture [48], as shown in Fig. 3(c), which consists of an input layer, a number of diffraction layers and an output layer. The input layer is illuminated by a terahertz light source, and the information can be encoded into the amplitude channel or phase channel of the input surface. The beam irradiated on the input surface was diffracted, and the coherent superposition of wavelet was used to change the amplitude and phase of the input wave. The photodetector array placed in the output detects the intensity of output light, and handwritten digits are recognized according to the difference of intensity in 10 different regions. Each diffractive layer is 3D printed by modeling the updated phase

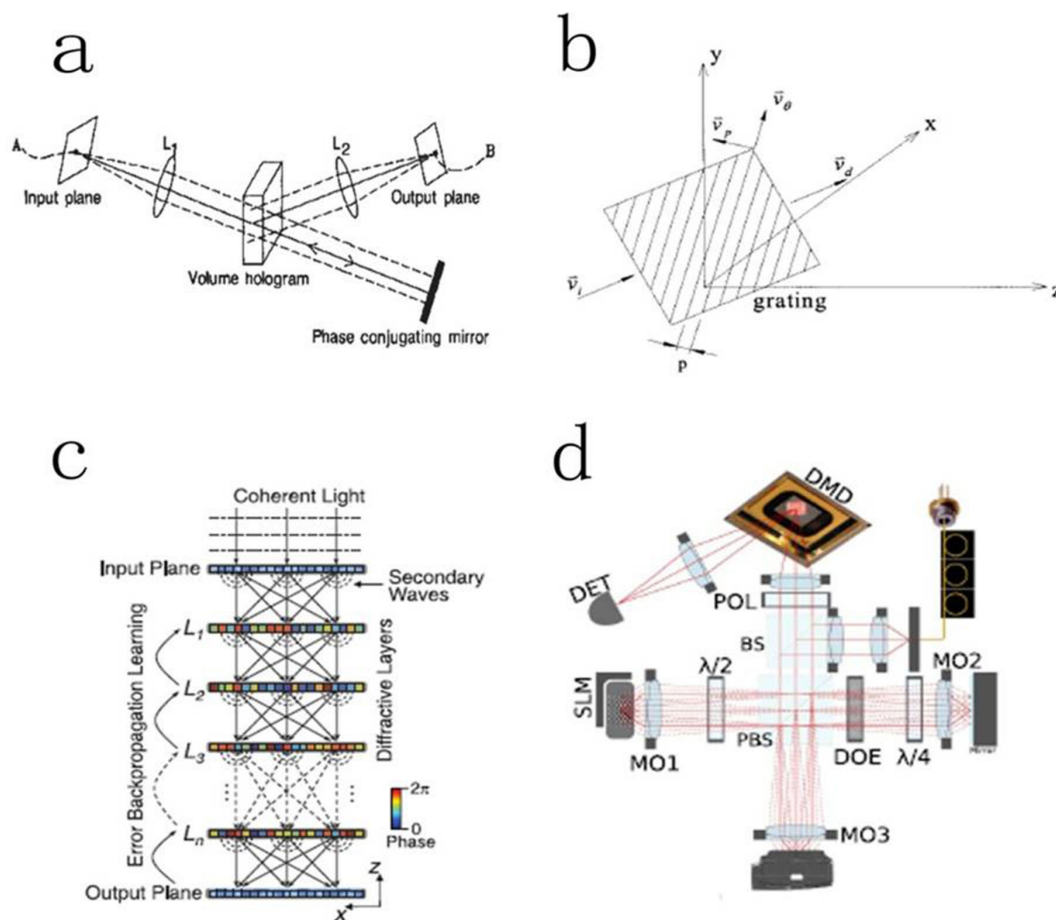


FIGURE 3. ONN. (a) Optical neuron realized by holographic grating based on diffraction principle [45]. (b) Matrix grating structure [46]. (c) Diffraction optical neural network composed of a number of diffraction layers [47]. (d) Neural networks implemented by DOE and DMD [51].

parameters. Training of this ONN, however, is still done by an electronic computer. It is impossible to realize fast real-time programming by parameterizing and 3D printing the virtual diffraction layers. Besides, the experimental environment is also limited by the use of terahertz light source. In 2019, their team proposed a wideband diffraction neural network based on the above architecture [49]. The requirement of the model for light source is no longer limited to monochromatic coherent light, and the application scope of ONN realized by this architecture is expanded, but the above problems still exist. From the perspective of free space, the architecture of this diffraction neural network is highly innovative. Based on the diffraction neural network, a team from Tianjin University creatively developed matrix gratings to replace 3D printed diffraction layers and obtained higher recognition accuracy [50]. They use a carbon dioxide laser tube to emit infrared light of 10.6 μm and the HgCdTe detector array. The model parameters trained by the computer are converted into the parameter distribution of the matrix grating. The advantage of using an infrared light source is that the size of a single neuron can be reduced to 5 microns, so that a 1mm by 1mm

matrix grating can contain 200 by 200 neurons. This kind of miniaturized matrix grating will be very beneficial to the integration of silicon photonic platform and obtain a wider range of applications.

In addition, J.Bueno *et al.* introduced a network of up to 2,025 diffractive photonic nodes, forming a large-scale recursive photonic neural network [51]. As shown in Fig. 3(d), the digital micro-mirror device (DMD) is used to realize reinforcement learning, and the results are effectively converge. In a network consisting of 2,025 nonlinear network nodes, each of which is a SLM pixel, the DOE is used to realize complex network connections. Because of the large number of neurons in this architecture, it has stimulated the researchers' thinking on the extensibility of the diffraction coupled photonic network. Furthermore, Sheler Maktoobi *et al.* demonstrated that a diffraction coupled photonic network can contain 30,000 photons and described its scalability in detail [52]. It is pointed out that the diffraction itself is not the basic limit of coupling, the vignetting is the fundamental cause, and the aberration as well as distortion are the limits of the actual optical system. If the diffraction order and the

position of the periodic arrangement of photon nodes deviate obviously, the coupling is not complete. The coupling efficiency of diffraction can be solved by changing the position of the single-mode fiber and adjusting the tilt angle of DOE.

The size of optical element in the optical neuron implemented by FSOI above is too large, resulting in a limited number of neurons in the network and a limited number of layers. Moreover, energy cannot be fully utilized when the beam propagated in free space. Due to the small number of layers in the network, large-scale complex nonlinear models cannot be learned and identified. In addition, the ONN realized by FSOI has strict requirements on the experimental environment, the existence of optical diffraction, interference and other phenomena also have different degrees of influence on the experimental results. However, the optical fiber communication interconnection technology, nanotechnology is gradually mature, PCM for neural computation and silicon photonic integrated circuits has been developed rapidly. It is possible to use optical waveguide technology to realize optical interconnection. In the next chapter we will introduce ONN based on optical waveguide technology and integrated photonic technology.

IV. USE WOI TO IMPLEMENT ONN

Optical waveguides are devices that guide the propagation of light waves and are divided into two categories. One is cylindrical optical waveguide, often called optical fibre. The other is the integrated optical waveguide, including the planar optical waveguides and the stripe medium optical waveguides.

A. ONN: BASED ON FIBER INTERCONNECTION

In 2012, F. Duport *et al.* proposed an all-optical implementation of reservoir computing based on a semiconductor optical amplifier array on a chip [53]. The off-line training is carried out by using the fiber delay switching of single nonlinear node, in which the nonlinearity is realized by using the saturation gain effect of the semiconductor optical amplifier [54]. On this basis, T. Cheng *et al.* proposed a scheme of an ONN based on reservoir computing in 2019 [55]. The soma of optical neurons can use directional couplers, fibers and amplifiers. Meanwhile, researchers use amplifiers to realize threshold, which is equivalent to nonlinear activation. The results indicate ONN can provide better performance and recognize the input signal waveform. We believe that the advantage of optical fibre is that it can be combined with the optical network in the 5G era, which is an engineering technology that can be further studied.

B. ONN: BASED ON INTEGRATED OPTICAL WAVEGUIDE PLATFORM

The integration of optical elements can greatly reduce the space occupied by neurons, which brings great changes to the further development of ONN. Special materials such as PCM have experienced the development of nearly ten years. In 2011, C. David Wright introduced the use of PCM for

arithmetic and bio-inspired calculation [56], and provided the principle experimental proof of “processor” based on PCM for the first time, demonstrating the four basic operations of addition, multiplication, division and subtraction, and storing the results at the same time. In the same year, D. Kuzum reported new nanoscale electronic synapses based on PCM for optical data storage and non-volatile storage [57]. Continuous resistance transitions in PCM are used to simulate the properties of biological synapses so as to realize synaptic learning rules. According to the review of integrated optical quantum technology [58], the integrated photonic platforms and devices have experienced rapid development from 2008 to 2018 and have become more and more mature. Under such a background, ONNs based on integrated optical waveguide interconnection technology have emerged.

In 2017, Alexander N. Tait of Princeton university published a paper referred to neuromorphic silicon photonics, introducing the world’s first integrated photonic neural network [59], as shown in Fig. 4(a). It uses a neural compiler to program a silicon photonic neural network with 49 nodes, each node operates at a specific wavelength, light from each node is detected and summed before it is fed into the laser, then the output will be feedback to create a feedback loop with nonlinear characteristics. Tait *et al.* simulated traditional neural networks, demonstrated how photonic neural networks can solve differential equations, and found that photonic neural networks using silicon photonic platforms can be connected to ultra-fast information processing environments for radio control and scientific computation.

In the same year, Shen. Y experimentally demonstrated the new architecture of AONN with a programmable nanophotonic processor (PNP) in a cascade array of 56 programmable Mach-Zehnder interferometers (MZIs) [60], as shown in Fig. 4(b). We know any real matrix M can be decomposed into $M = U \Sigma V^+$. U , V^+ can be realized by using optical beam-splitter and phase shifter, Σ can be realized by optical attenuator. By tuning the phase shifter integrated in MZIs, it is possible to compute any size of the input. Such nanophotonic integrated circuits hold out the promise of AONN capable of meeting the upcoming big data and deep learning challenges, but directional couplers and phase modulators take up too much space for ONN with more than 1,000 neurons. So far, the goal of a large-scale and fast programmable photonic neural network has not been achieved.

In 2019, Feldmann J *et al.* proposed a fully optical synaptic system capable of both supervised and unsupervised learning [61]. By using wavelength division multiplexing (WDM), an extensible photonic neural network circuit is realized and the pattern recognition in optical field is successfully carried out. This photonic synaptic network promises to achieve the high speed and bandwidth inherent in optical systems, enabling direct processing of optical communications and visual data. From Fig. 4(c) we can see that the red rectangular block at the input represents PCM, four small circle is equivalent to multiplexers. If the sum of the four input pulses is insufficient to switch the state of PCM, the probe pulse resonates with

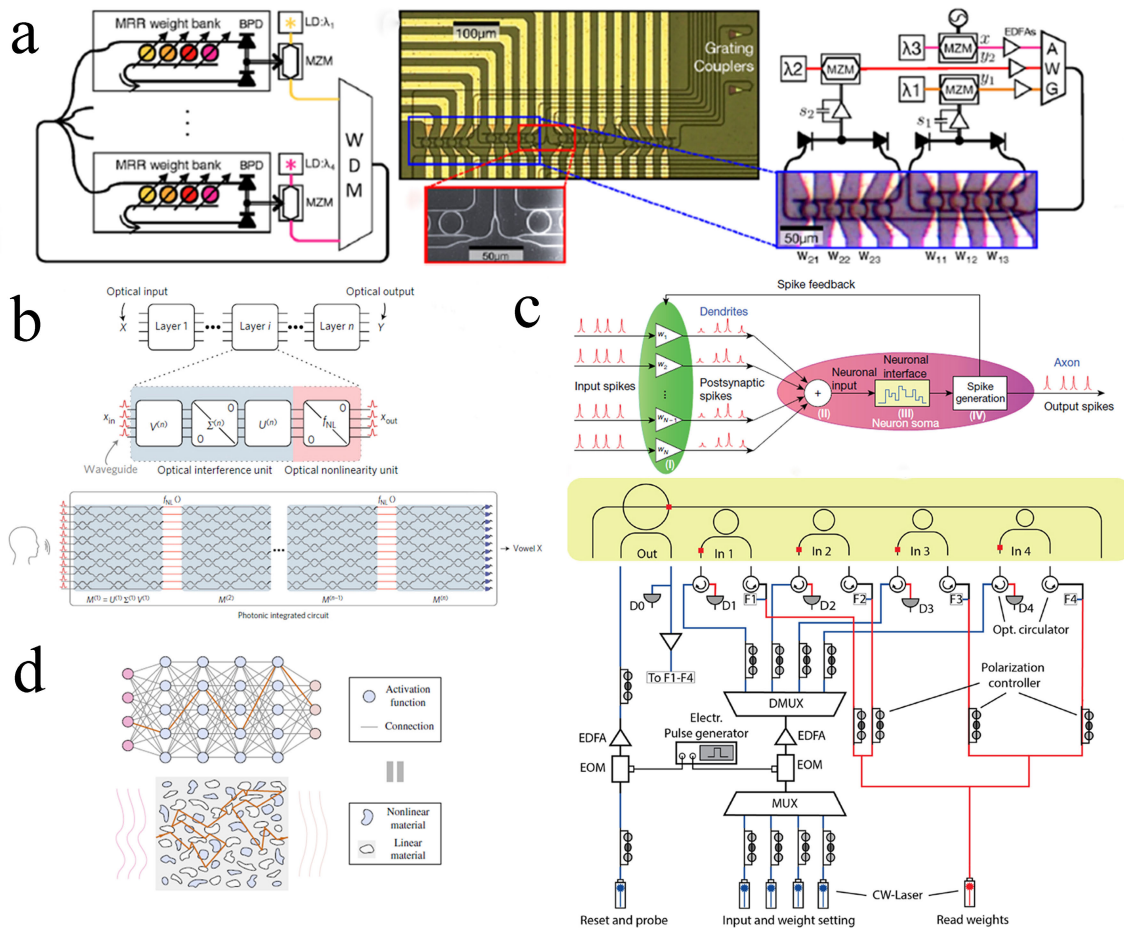


FIGURE 4. ONN based on integrated photon technology. (a) Concept of broadcast weighted network with modulator as neuron [59]. (b) Basic structure of ONN based on programmable photon processor [60]. (c) All-optical peak neural network based on WDM [61]. (d) Structure of nanomaterials and comparison with artificial neurons [62].

the ring resonator. The weight can be modified by adjusting the pulse power of the laser to change the state of PCM, and then generate the output. It is worth learning to integrate PCM or other materials with optical nonlinearity into a ring resonator.

In addition, E.Khoram introduced a nanophotonic medium that can be used for neural inference [62], as shown in Fig. 4(d). The major material of the medium is silicon dioxide, which contains dopants with different refractive index from the major material, and there are linear and nonlinear particles in the dopants. The ONN based on nanophotonic neural medium has achieved high accuracy in handwriting recognition. This new-type of nanoscale neural medium replaces the traditional hierarchical network and can be used to realize photonic memristor and study synaptic photonics [63].

V. OPTICAL NONLINEARITY AND ITS IMPLEMENTATION

Nonlinear optics studies the generation mechanism of new phenomena in the interaction between coherent light and matter. The optical properties of a material change under the

irradiation of a laser beam, which in turn affects the properties of the beam. The nonlinear optical effect originates from the nonlinear polarization of molecules and materials. If there is an energy exchange between light and matter, this effect is called the active optical nonlinear effect. It includes: nonlinear absorption, nonlinear refraction, nonlinear scattering, optical bistability, etc.

A. OPTICAL KERR EFFECT AND ITS APPLICATION

Kerr nonlinearity belongs to active optical nonlinearity. When the incident light is strong enough, the refractive index of the wave polarized along the direction of parallel and perpendicular to the light field will have a difference Δn , which is proportional to the light intensity of the laser beam acting in the medium. The first and third order effects are only considered when a high power laser with frequency ω is incident on an isotropic medium.

Suppose the first order polarizability: $\chi^{(1)} = \chi^{(1)'} + i \times \chi^{(1)''}$, the third is: $\chi^{(3)} = \chi^{(3)'} + i \times \chi^{(3)''}$. Refractive index is only related to the real part of the polarizability. Linear refractive index: $n_0^2 = 1 + \chi(\omega)^{(1)'}$, effective third

order polarizability: $\chi_{\epsilon}^{(3)'} = 3\chi(\omega)^{(3)'}$. From $D = \epsilon_0 E + P$, $D = \epsilon E$, we can know $\epsilon E = \epsilon_0 E + P$, so total dielectric coefficient is $\epsilon = \epsilon_0(1 + \chi^{(1)'} + \chi_{\epsilon}^{(3)'}|E(\omega)|^2)$. The refractive index of frequency ω is:

$$\begin{aligned} n(\omega) &= \sqrt{\epsilon(\omega)/\epsilon_0} \\ &= \sqrt{1 + \chi(\omega)^{(1)} + 3\chi(\omega; \omega, -\omega)|E(\omega)|^2} \\ &= n_0 \sqrt{1 + \frac{\chi_{\epsilon}^{(3)}}{n_0^2}|E(\omega)|^2} \\ &\approx n_0 + \frac{\chi_{\epsilon}^{(3)}}{2n} |E(\omega)|^2 \end{aligned} \quad (4)$$

The nonlinear refractive index is: $\Delta n = \frac{\chi_{\epsilon}^{(3)}}{2n} |E(\omega)|^2$. We know that the incident light intensity I can be expressed as $I = \frac{1}{2}\epsilon_0 c n_0 |E(\omega)|^2$, so the nonlinear refractive index is $\Delta n = \frac{\chi_{\epsilon}^{(3)}}{\epsilon_0 c n_0^2} I = n_2 I$, n_2 is nonlinear refraction coefficient. Finally, the total refractive index of Kerr medium is $n = n_0 + n_2 I$.

One application of Kerr's nonlinearity is to produce optical bistability, which is produced by two conditions: nonlinearity and positive feedback [64]. The simplest optical bistable device is to place a nonlinear medium in an F-P optical cavity. Kerr medium is a dispersive medium. According to the formula $n = n_0 + n_2 I$, the refractive index of the medium is non-linear depending on the intensity in the cavity, so the resonant wavelength of the cavity is related to I . The difference between the resonant wavelength and the incident laser wavelength can affect the transmission coefficient of the incident light and make the intensity change in the cavity. Therefore, this is a feedback process, and the optical non-linear activation generated by optical bistable devices can be added to the optical path.

In view of the nonlinearity of Kerr medium, AONN is constructed by using the thin layer of Kerr material separated by free space to realize weighted connection and nonlinear processing, which is discussed in [65]. Nonlinear processing is caused by the Kerr effect (self-focusing or self-defocusing) and interference of the optical signal in the device. The output error of the light is propagated back to produce the error signal, and then the intensity is changed on the nonlinear layer. Compared with other ONNs, which require independent linear and nonlinear operation, AONN constructed by Kerr nonlinear materials can achieve the above functions simultaneously and with fast processing speed. Similarly, single mode Kerr interaction is used to realize single photon coherent nonlinearity [66]. We previously mentioned that active optical nonlinearity includes nonlinear absorption, one of which is two-photon absorption. Considering the fast speed of the Kerr nonlinear material, researchers combined the Kerr effect with two-photon absorption to build a nonlinear mechanism, and used it in conjunction with the ring resonator to achieve all-optical reservoir calculation [67]. This is similar to [31], [61] in the implementation of optical nonlinear activation.

B. OTHER NONLINEAR OPTICAL ACTIVATION

We discussed the importance of optical nonlinearity in Sec. II-C, many researchers have introduced nonlinear optical effects in recent years. R. Amin and J. George *et al.* pointed out that electro-optical absorption modulation can be used to realize nonlinear modulation of light waves [68]–[70]. The method of mapping the nonlinear activation function to the transfer function of the electro-optic modulator is discussed. It is pointed out that s-type function, ReLu function or other nonlinear activation function can be realized by using different electro-optic materials. For example, the ReLu function can be realized by using the inverted state filling light absorption mechanism of quantum dots (QD) [71]. At present, the mechanism of the nonlinear activation function unit first converts the optical signal into an electrical signal detection mechanism, and then converts it into an optical signal. In optical reservoir computing [72], the most commonly used optical nonlinear unit is graphene saturable absorber or is implemented based on the two-photon absorption [73], other research on nonlinear is based on the bistable switch and ring-resonators [74], [75]. However, these methods do not reach an ideal expectation in efficiency and speed. The optical non-linearity designed by M. Miscuglio *et al.* depends on the reversible transparent induction caused by the fano resonance in the plasma oscillator subsystem, and the anti-saturation absorption response of buckyball (C60) membrane to the non-linearity [76], achieving fast and efficient all-optical nonlinearity, improving the throughput of neural network, reducing latency and power consumption. In addition, the team from Hong Kong designed an AONN with nonlinear activation function [37]. Based on electromagnetic induction transparency (EIT), a kind of optical quantum interference effect between atomic transitions, AONN can accurately identify the Ising model.

C. IMPLEMENTATION: FREE SPACE AND INTEGRATED WAVEGUIDE

To sum up, SLM, MLA and HOE can be used to realize optical linear interconnection in free space. There are two ways to realize the nonlinearity: (1) the optical field information after optical linear interconnection is recorded by CCD, being sent to the personal computer to add nonlinear threshold, and then feedback to SLM. The Gerchberg-Saxton algorithm is used for feedback iteration, and the weight on SLM is updated in real time [32], [77]. (2) develop fully optical nonlinear activation materials. Optical nonlinearity can be divided into active and passive, and the difference lies in whether there is energy exchange between light and medium. Dyes and graphene with saturable absorption effects can be developed according to active optical nonlinear refraction and nonlinear absorption, and the above-mentioned optical bistable devices can also be added.

The free space interconnection technology of ONN often occupies a large volume of space. Moreover, the energy loss of light is large, and there is distortion in the light path.

The ONN realized by integrated optical interconnect technology can integrate excellent nonlinear materials such as black phosphorus and graphene well, so that the volume of ONN can be greatly reduced and the network can have more neurons and depth. The most important is that the integrated photonic neural chip produced by nanoscale technology can be programmable in real time, and can be used for training and recognition of more complex nonlinear models.

VI. SUMMARY AND OUTLOOK

In this review, we introduce the implementation of ONN in free space and integrated waveguides, and emphasize the significance of optical nonlinear activation. Here, we point out that compared with ONN realized in free space, the integrated photonic neural network is smaller in size and programmable, which has potential application value. Therefore, future research can focus on exploring nano-optical elements and nonlinear materials that can be integrated in optical waveguides, such as black phosphorus [78] and nanolasers [79], and designing integrated photonic circuits with higher density [80]. Of course, exploring the implementation of ONN in free space is a prerequisite for its integration (we can call this “experience”). In addition, the growing maturity of nanotechnology allows metamaterials to be applied to the ONN [81], [82], and the network can be applied to image recognition, speech processing, target tracking and other fields. We believe that in the future super-large scale and programmable photonic neural networks can be realized.

REFERENCES

- [1] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *Bull. Math. Biophys.*, vol. 5, no. 4, pp. 115–133, Dec. 1943.
- [2] R. F. Thompson, “The neurobiology of learning and memory,” *Science*, vol. 233, no. 4767, pp. 941–947, 1986.
- [3] H.-J. Zhang and J. Zhao, “A new self-learning algorithm of neural networks and its application,” in *Proc. IEEE Int. Conf. Ind. Technol.*, Apr. 2008, pp. 1–6.
- [4] W. Lin and G. Chen, “Large memory capacity in chaotic artificial neural networks: A view of the anti-integrable limit,” *IEEE Trans. Neural Netw.*, vol. 20, no. 8, pp. 1340–1351, Aug. 2009.
- [5] K. T. Sun and H. C. Fu, “A hybrid neural network model for solving optimization problems,” *IEEE Trans. Comput.*, vol. 42, no. 2, pp. 218–227, 1993.
- [6] L. Jin-Yue and Z. Bao-Ling, “Research on the non-linear function fitting of RBF neural network,” in *Proc. Int. Conf. Comput. Inf. Sci.*, Jun. 2013, pp. 842–845.
- [7] S.-C. Huang and B.-H. Do, “Radial basis function based neural network for motion detection in dynamic scenes,” *IEEE Trans. Cybern.*, vol. 44, no. 1, pp. 114–125, Jan. 2014.
- [8] M. L. Hines and N. T. Carnevale, “The NEURON simulation environment,” *Neural Comput.*, vol. 9, no. 6, pp. 1179–1209, Aug. 1997.
- [9] R. J. Schwabe, S. Zelinger, T. S. Key, and K. O. Phipps, “Electronic lighting interference,” *IEEE Ind. Appl. Mag.*, vol. 4, no. 4, pp. 43–48, Jul./Aug. 1998.
- [10] H. Markram et al., “Reconstruction and simulation of neocortical micro-circuitry,” *Cell*, vol. 163, no. 2, pp. 456–492, 2015.
- [11] D. Armitage, W. Anderson, and T. Karr, “High-speed spatial light modulator,” *IEEE J. Quantum Electron.*, vol. QE-21, no. 8, pp. 1241–1248, Aug. 1985.
- [12] P. Chaisakul, D. Marris-Morini, J. Frigerio, D. Chrastina, M.-S. Rouifed, S. Cecchi, P. Crozat, G. Isella, and L. Viviani, “Integrated germanium optical interconnects on silicon substrates,” *Nature Photon.*, vol. 8, no. 6, pp. 482–488, Jun. 2014.
- [13] J. K. Doylend and A. P. Knights, “The evolution of silicon photonics as an enabling technology for optical interconnection,” *Laser Photon. Rev.*, vol. 6, no. 4, pp. 504–525, Jul. 2012.
- [14] M. J. Murdocca and V. P. Heuring, *Principles of Computer Architecture*, 1st ed. Reading, MA, USA: Addison-Wesley, 1999.
- [15] F.-C.-F. Tsai, C. J. O’Brien, N. S. Petrović, and A. D. Rakić, “Analysis of optical channel cross talk for free-space optical interconnects in the presence of higher-order transverse modes,” *Appl. Opt.*, vol. 44, no. 30, pp. 6380–6387, Oct. 2005.
- [16] W. Hu, X. Li, J. Yang, and D. Kong, “Crosstalk analysis of aligned and misaligned free-space optical interconnect systems,” *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 27, no. 2, pp. 200–205, Feb. 2010.
- [17] J. W. Goodman, A. R. Dias, and L. M. Woody, “Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms,” *Opt. Lett.*, vol. 2, no. 1, pp. 1–3, 1978.
- [18] X. Hu, A. Wang, M. Zeng, Y. Long, L. Zhu, C. Gui, L. Fu, L. Shen, and J. Wang, “Graphene-assisted multiple-input high-base optical computing,” in *Proc. 15th Int. Conf. Opt. Commun. Netw. (ICOON)*, Sep. 2016, pp. 1–3.
- [19] H. J. Caulfield and S. Dolev, “Why future supercomputing requires optics,” *Nature Photon.*, vol. 4, no. 5, pp. 261–263, May 2010.
- [20] E. P. Mosca, R. D. Griffin, F. P. Pursell, and J. N. Lee, “Acoustooptical matrix-vector product processor: Implementation issues,” *Appl. Opt.*, vol. 28, no. 18, pp. 3843–3851, Sep. 1989.
- [21] C. C. Sun, M. W. Chang, and K. Y. Hsu, “Matrix-matrix multiplication by using anisotropic self-diffraction in batio3,” *Appl. Opt.*, vol. 33, no. 20, pp. 4501–4507, 1994.
- [22] J. O’Brien, A. Furusawa, and J. Vuckovic, “Photonic quantum technologies,” *Nature Photon.*, vol. 3, no. 12, pp. 687–695, 2009.
- [23] M. Kues, C. Reimer, P. Roztockı, L. R. Cortés, S. Sciara, B. Wetzel, Y. Zhang, A. Cino, S. T. Chu, B. E. Little, D. J. Moss, L. Caspani, J. Azaña, and R. Morandotti, “On-chip generation of high-dimensional entangled quantum states and their coherent control,” *Nature*, vol. 546, no. 7660, pp. 622–626, Jun. 2017.
- [24] N. C. Harris, G. R. Steinbrecher, M. Prabhu, Y. Lahini, J. Mower, D. Bunandar, C. Chen, F. N. C. Wong, T. Baehr-Jones, M. Hochberg, S. Lloyd, and D. Englund, “Quantum transport simulations in a programmable nanophotonic processor,” *Nature Photon.*, vol. 11, no. 7, pp. 447–452, Jul. 2017.
- [25] R. E. Uhrig, “Introduction to artificial neural networks,” in *Proc. 21st Annu. Conf. IEEE Ind. Electron. (IECON)*, Nov. 1995, pp. 33–37.
- [26] T. F. de Lima, B. J. Shastri, A. N. Tait, M. A. Nahmias, and P. R. Prucnal, “Progress in neuromorphic photonics,” *Nanophotonics*, vol. 6, no. 3, pp. 577–599, Jan. 2017.
- [27] M. B. Nasr and M. Chtourou, “A hybrid training algorithm for feedforward neural networks,” *Neural Process. Lett.*, vol. 24, no. 2, pp. 107–117, Oct. 2006.
- [28] M. Li, “Research on optical vector matrix multiplication,” *Comput. Technol. Develop.*, vol. 22, no. 2, pp. 198–201, 2012.
- [29] Y. Chen, “4f-type optical system for matrix multiplication,” *Opt. Eng.*, vol. 32, no. 1, pp. 77–79, 1993.
- [30] B. Hunt, “A matrix theory proof of the discrete convolution theorem,” *IEEE Trans. Audio Electroacoust.*, vol. AU-19, no. 4, pp. 285–288, Dec. 1971.
- [31] Z. Cheng, C. Ríos, W. H. P. Pernice, C. D. Wright, and H. Bhaskaran, “On-chip photonic synapse,” *Sci. Adv.*, vol. 3, no. 9, Sep. 2017, Art. no. e1700160.
- [32] D. Pierangeli, G. Marcucci, C. Moriconi, G. Perini, M. De Spirito, M. Papi, and C. Conti, “Deep optical neural network by living tumour brain cells,” 2018, *arXiv:1812.09311*. [Online]. Available: <https://arxiv.org/abs/1812.09311>
- [33] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proc. Nat. Acad. Sci. USA*, vol. 79, no. 8, pp. 2554–2558, 1982.
- [34] N. H. Farhat, D. Psaltis, A. Prata, and E. Paek, “Optical implementation of the hopfield model,” *Appl. Opt.*, vol. 24, no. 10, p. 1469, May 1985.
- [35] S. Y. Kung and H. K. Liu, “An optical inner-product array processor for associative retrieval,” *Proc. SPIE*, vol. 613, pp. 214–219, Jun. 1986.
- [36] J. Ohta, M. Takahashi, Y. Nitta, S. Tai, K. Mitsunaga, and K. Kyuma, “A new approach to a GaAs/AlGaAs optical neurochip with three layered structure,” in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 1989, pp. 477–482.
- [37] Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y.-C. Chen, P. Chen, G.-B. Jo, J. Liu, and S. Du, “All-optical neural network with nonlinear activation functions,” *Optica*, vol. 6, no. 9, pp. 1132–1137, Sep. 2019.

- [38] Y. Kuratomi, A. Takimoto, K. Akiyama, and H. Ogawa, "Optical neural network using vector-feature extraction," *Appl. Opt.*, vol. 32, no. 29, pp. 5750–5758, Oct. 1993.
- [39] L. Taiwei, W. Shudong, X. Xu, and F. T. S. Yu, "Two-dimensional programmable optical neural network," *Appl. Opt.*, vol. 28, no. 22, pp. 4908–4913, 1989.
- [40] Y. Jianwen, W. Bin, T. Weijian, B. Zhengkang, and L. Zukang, "Parallel subtraction using lclv and lenslet array in optical neural networks," *Chin. J. Lasers*, vol. 6, no. 2, pp. 169–176, 1997.
- [41] S. Gao, J. Yang, Z. Feng, and Y. Zhang, "Implementation of a large-scale optical neural network by use of a coaxial lenslet array for interconnection," *Appl. Opt.*, vol. 36, no. 20, pp. 4779–4783, Jul. 1997.
- [42] S. Gao, Y. Zhang, J. Yang, and G. Mu, "Coaxial architecture of an optical neural network with a lenslet array," *Opt. Lett.*, vol. 19, no. 24, pp. 2155–2157, 1994.
- [43] A. Zanutta, E. Orsellì, T. Fäcke, and A. Bianco, "Photopolymeric films with highly tunable refractive index modulation for high precision diffractive optics," *Opt. Mater. Express*, vol. 6, no. 1, pp. 252–263, 2015.
- [44] R. Pashaie and N. H. Farhat, "Optical realization of bioinspired spiking neurons in the electron trapping material thin film," *Appl. Opt.*, vol. 46, no. 35, pp. 8411–8418, 2007.
- [45] D. Psaltis, D. Brady, X.-G. Gu, and S. Lin, "Holography in artificial neural networks," *Nature*, vol. 343, no. 6256, pp. 325–330, 1990.
- [46] S. L. Yeh, R. C. Lo, and C. Y. Shi, "Optical implementation of the hopfield neural network with matrix gratings," *Appl. Opt.*, vol. 43, no. 4, pp. 858–865, 2004.
- [47] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," *Science*, vol. 361, no. 6406, pp. 1004–1008, Sep. 2018.
- [48] D. Mengü, Y. Luo, Y. Rivenson, and A. Ozcan, "Analysis of diffractive optical neural networks and their integration with electronic neural networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–14, Jan. 2020.
- [49] Y. Luo, D. Mengü, N. T. Yardimci, Y. Rivenson, M. Veli, M. Jarrahi, and A. Ozcan, "Design of task-specific optical systems using broadband diffractive neural networks," *Light, Sci. Appl.*, vol. 8, no. 1, pp. 1–14, Dec. 2019.
- [50] L. Lu, Z. Zeng, L. Zhu, Q. Zhang, B. Zhu, Q. Yao, M. Yu, H. Niu, M. Dong, and G. Zhong, "Miniaturized diffraction grating design and processing for deep neural network," *IEEE Photon. Technol. Lett.*, vol. 31, no. 24, pp. 1952–1955, Dec. 15, 2019.
- [51] J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, "Reinforcement learning in a large-scale photonic recurrent neural network," *Optica*, vol. 5, no. 6, pp. 756–760, 2018.
- [52] S. Maktoobi, L. Froehly, L. Andreoli, X. Porte, M. Jacquot, L. Larger, and D. Brunner, "Diffractive coupling for photonic networks: How big can we go?" *IEEE J. Sel. Topics Quantum Electron.*, vol. 26, no. 1, pp. 1–8, Jan. 2020.
- [53] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, "All-optical reservoir computing," *Opt. Express*, vol. 20, no. 20, pp. 22783–22795, 2012.
- [54] D. Rosenbluth, K. Kravtsov, M. P. Fok, and P. R. Prucnal, "A high performance photonic pulse processing device," *Opt. Express*, vol. 17, no. 25, pp. 22767–22772, 2009.
- [55] T.-Y. Cheng, D.-Y. Chou, C.-C. Liu, Y.-J. Chang, and C.-C. Chen, "Optical neural networks based on optical fiber-communication system," *Neuro-computing*, vol. 364, pp. 239–244, Oct. 2019.
- [56] C. D. Wright, Y. Liu, K. I. Kohary, M. M. Aziz, and R. J. Hicken, "Arithmetic and biologically-inspired computing using phase-change materials," *Adv. Mater.*, vol. 23, no. 30, pp. 3408–3413, Aug. 2011.
- [57] D. Kuzum, R. G. D. Jeyasingh, B. Lee, and H.-S.-P. Wong, "Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing," *Nano Lett.*, vol. 12, no. 5, pp. 2179–2186, May 2012.
- [58] J. Wang, F. Sciarino, A. Laing, and M. G. Thompson, "Integrated photonic quantum technologies," *Nature Photon.*, Oct. 2019.
- [59] A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," *Sci. Rep.*, vol. 7, no. 1, Dec. 2017, Art. no. 7430.
- [60] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," *Nature Photon.*, vol. 11, no. 7, pp. 441–446, Jul. 2017.
- [61] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, no. 7755, pp. 208–214, May 2019.
- [62] E. Khoram, A. Chen, D. Liu, L. Ying, Q. Wang, M. Yuan, and Z. Yu, "Nanophotonic media for artificial neural inference," *Photon. Res.*, vol. 7, no. 8, pp. 823–827, 2019.
- [63] X. Feng, X. Liu, and K.-W. Ang, "2D photonic memristor beyond graphene: Progress and prospects," *Nanophotonics*, Mar. 2020, doi: 10.1515/nanoph-2019-0543.
- [64] S. Li and X. Cai, "High-contrast all optical bistable switching in coupled nonlinear photonic crystal microcavities," *Appl. Phys. Lett.*, vol. 96, no. 13, Mar. 2010, Art. no. 131114.
- [65] S. R. Skinner, J. E. Steck, and E. C. Behrman, "Optical neural network using kerr-type nonlinear materials," in *Proc. 4th Int. Conf. Microelectron. for Neural Netw. Fuzzy Syst.*, 1994, pp. 12–15.
- [66] G. R. Steinbrecher, J. P. Olson, D. Englund, and J. Carolan, "Quantum optical neural networks," *npj Quantum Inf.*, vol. 5, no. 1, p. 60, Dec. 2019.
- [67] C. Mesaritakis and D. Syvridis, "All-optical reservoir computing system based on ingaasp ring resonators for high-speed identification and optical routing in optical networks," *Proc. SPIE*, vol. 9370, no. 6, pp. 1269–1277, 2015.
- [68] R. Amin, J. George, J. Khurgin, T. El-Ghazawi, and V. J. Sorger, "Attojoule modulators for photonic neuromorphic computing," in *Proc. CLEO, Appl. Technol.*, Jan. 2018, Paper ATH1Q.4. [Online]. Available: https://www.osapublishing.org/abstract.cfm?URI=CLEO_AT-2018-ATH1Q.4, doi: 10.1364/CLEO_AT.2018.ATH1Q.4.
- [69] R. Amin, S. Khan, C. J. Lee, H. Dalir, and V. J. Sorger, "110 Attojoule-per-bit efficient graphene-based plasmon modulator on silicon," in *Proc. Conf. Lasers Electro-Opt.*, 2018, pp. 13–18.
- [70] J. K. George, A. Mehrabian, R. Amin, J. Meng, T. F. de Lima, A. N. Tait, B. J. Shastri, T. El-Ghazawi, P. R. Prucnal, and V. J. Sorger, "Neuromorphic photonics with electro-absorption modulators," *Opt. Express*, vol. 27, no. 4, pp. 5181–5191, Feb. 2019.
- [71] J. George, R. Amin, A. Mehrabian, J. Khurgin, T. El-Ghazawi, P. R. Prucnal, and V. J. Sorger, "Electrooptic nonlinear activation functions for vector matrix multiplications in optical neural networks," in *Proc. Adv. Photon.*, 2018, pp. 2–5.
- [72] A. Dejonckheere, F. Duport, A. Smerieri, L. Fang, J.-L. Oudar, M. Haelterman, and S. Massar, "All-optical reservoir computer based on saturation of absorption," *Opt. Express*, vol. 22, no. 9, pp. 10868–10881, May 2014.
- [73] Z. Cheng, H. K. Tsang, X. Wang, K. Xu, and J.-B. Xu, "In-plane optical absorption and free carrier absorption in graphene-on-silicon waveguides," *IEEE J. Sel. Topics Quantum Electron.*, vol. 20, no. 1, pp. 43–48, Jan. 2014.
- [74] M. Soljačić, M. Ibanescu, S. G. Johnson, Y. Fink, and J. D. Joannopoulos, "Optimal bistable switching in nonlinear photonic crystals," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 66, no. 5, Nov. 2002, Art. no. 055601.
- [75] F. D.-L. Coarer, M. Sciamanna, A. Katumba, M. Freiberger, J. Dambre, P. Bienstman, and D. Rontani, "All-optical reservoir computing on a photonic chip using silicon-based ring resonators," *IEEE J. Sel. Topics Quantum Electron.*, vol. 24, no. 6, pp. 1–8, Nov. 2018.
- [76] M. Mario, M. Armin, H. Zibo, A. Shaimaa, G. Jonathan, A. Kildishev, P. Matthew, and S. Volker, "All-optical nonlinear activation function for photonic neural networks," *Opt. Mater. Express*, vol. 8, no. 12, pp. 3851–3863, 2018.
- [77] G. Z. Yang, B. Z. Dong, B. Y. Gu, J. Y. Zhuang, and O. K. Ersoy, "Gerchberg–saxton and yang–gu algorithms for phase retrieval in a nonunitary transform system: A comparison," *Appl. Opt.*, vol. 33, no. 2, pp. 209–218, 1994.
- [78] Z. Cheng, R. Cao, J. Guo, Y. Yao, K. Wei, S. Gao, Y. Wang, J. Dong, and H. Zhang, "Phosphorene-assisted silicon photonic modulator with fast response time," *Nanophotonics*, Feb. 2020, doi: 10.1515/nanoph-2019-0510.
- [79] A. A. Vyshnevyy and D. Y. Fedyanin, "Lasing threshold of thresholdless and non-thresholdless metal-semiconductor nanolasers," *Opt. Express*, vol. 26, no. 25, pp. 33473–33483, Dec. 2018.
- [80] A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," *J. Lightw. Technol.*, vol. 32, no. 21, pp. 4029–4041, Nov. 1, 2014.
- [81] S. Colburn, Y. Chu, E. Shilzerman, and A. Majumdar, "Optical frontend for a convolutional neural network," *Appl. Opt.*, vol. 58, no. 12, pp. 3179–3186, 2019.

[82] P. Hougne, M. F. Imani, A. V. Diebold, R. Horstmeyer, and D. R. Smith, "Learned integrated sensing pipeline: Reconfigurable metasurface transceivers as trainable physical layer in an artificial neural network," *Adv. Sci.*, vol. 7, no. 3, Feb. 2020, Art. no. 1901913. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/advs.201901913>



XIUBAO SUI received the B.S., M.S., and Ph.D. degrees in optical engineering from the Nanjing University of Science and Technology, Nanjing, China. He is currently a Professor with the Nanjing University of Science and Technology. So far, he has published 50 SCI papers. His current research interests include machine learning, image processing, and computer vision.



QIUHAOWU received the B.S. degree in photoelectric information science and engineering from Nantong University, Nantong, China, in 2019. He is currently pursuing the M.S. degree in optical engineering with the Nanjing University of Science and Technology, Nanjing, China. His current research direction is optical deep learning.



JIA LIU received the B.S. degree in electronic science and technology from the Nanjing University of Science and Technology, Nanjing, China, in 2019, where she is currently pursuing the M.S. degree in optical engineering. Her current research direction is optical deep learning.



QIAN CHEN received the B.S., M.S., and Ph.D. degrees in optical engineering from the Nanjing University of Science and Technology, Nanjing, China.

He is currently a Professor with the Nanjing University of Science and Technology. His current research interests include machine learning, image processing, and computer vision. He presided over more than 50 scientific research tasks, including major national projects, models, key pre-research and fund projects, had 43 authorized invention patents and published a monograph, nearly 200 academic articles of which more than 100 have collected by SCI and EI. He had won a second prize of national science and technology progress award, third first prizes of provincial and ministerial level science and technology award and 12 other science and technology awards. In addition, he is the member of the American Optical Society (OSA) and the International Society of Optical Engineering (SPIE), the Standing Director of the Chinese Institute of Electronics and the Chinese Optical Engineering Society, the Vice Chairman of the Optoelectronic Professional Committee of the Chinese Ordnance Society, the Vice Chairman of the Microscopic Instrument Branch of the Chinese Instrumentation Society, the Vice Chairman of the Jiangsu Optical Society, and the Chairman of the Optoelectronic Technology Professional Committee, Jiangsu Province.



GUOHUA GU received the B.S., M.S., and Ph.D. degrees in optical engineering from the Nanjing University of Science and Technology, Nanjing, China. He is currently a Professor with the Nanjing University of Science and Technology. From 2015 to 2019, he has published 29 articles and 37 patents. His current research interests include machine learning, image processing, and computer vision.

...