



ELSEVIER

Contents lists available at ScienceDirect

Engineering

journal homepage: www.elsevier.com/locate/eng



Research
Artificial Intelligence—Review

模拟光计算推动人工智能发展

吴嘉敏^{a,b,c,d,#}, 林星^{b,d,e,#}, 郭雨晨^{b,c,#}, 刘军伟^f, 方璐^{b,c,e,*}, 焦述铭^{g,*}, 戴琼海^{a,b,c,d,*}

^a Department of Automation, Tsinghua University, Beijing 100084, China

^b Institute for Brain and Cognitive Science, Tsinghua University, Beijing 100084, China

^c Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

^d Beijing Innovation Center for Future Chips, Tsinghua University, Beijing 100084, China

^e Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

^f Department of Physics, The Hong Kong University of Science and Technology, Hong Kong 999077, China

^g Peng Cheng Laboratory, Shenzhen 518055, China

ARTICLE INFO

Article history:

Received 23 November 2020

Revised 19 June 2021

Accepted 20 June 2021

Available online 21 August 2021

关键词

人工智能

光计算

光电架构

神经网络

神经形态计算

蓄水池计算

光电处理器

摘要

人工智能(AI)技术正在飞速发展并已在各行各业得到了广泛的应用,但大数据的爆炸式增长使其在硬件的处理速度和功耗方面面临了前所未有的严峻挑战。而光计算恰好利用光子的特性,包括宽带、低延迟和高能效,为解决这一瓶颈提供了一个独特的视角。本文介绍了不同AI模型的光计算的最新研究进展,包括前馈神经网络、蓄水池计算和脉冲神经网络(SNN)。集成光子器件的最新进展和AI的兴起为光计算在实际应用中的再次崛起创造了良好的发展机遇。这项浩大的工程需要不同领域多学科的交叉来实现。本文综述了近年来该领域的最前沿研究成果,讨论了目前相关技术的可用性,并指出了从不同方面推进该领域发展所面临的挑战。我们预计实际AI应用所需的大规模集成光电子处理器时代,将很快以光电混合框架的形式到来。

© 2021 THE AUTHORS. Published by Elsevier LTD on behalf of Chinese Academy of Engineering and Higher Education Press Limited Company. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. 引言

近十年以来,随着深度学习的广泛应用,人工智能(AI)在机器视觉[1]、自动驾驶[2]、棋类游戏[3–4]以及临床诊断[5–6]等很多领域取得了举世瞩目的成功。AI在理论和应用方面都有了快速的发展[7],对国际前沿AI模型的训练和执行所需要的算力远比由著名的摩尔定律刻画的集成电路的发展更快,后者的发展正在放缓[8]。在即将到来的第5代移动通讯技术(5G)时代,大量的传感器和联

网设备每秒以爆炸式的增长速率生成大数据。而AI需要尽可能快速地处理这些数据。神经网络是由通过成千上万个或者数百万个相互连接的多层神经元组成的用于学习多级抽象数据的典型的AI模型,其变得越来越复杂。目前的AI模型已经能满足很多应用的需求,但在未来的发展中其将面临计算硬件的速度和功耗的基本瓶颈限制[9–10]。

为了解决这个公认的问题,近年来研究人员在电子计算架构创新方面进行了各种努力以加速低能耗下的AI推理和训练[11–12]。比如,专用集成电路(ASIC)可以减

* Corresponding authors.

E-mail addresses: fanglu@mail.tsinghua.edu.cn (L. Fang), jiaoshm@pcl.ac.cn (S. Jiao), qh dai@tsinghua.edu.cn (Q. Dai).

These authors contributed equally to this work.

少不必要的操作而被工业界广泛接受[13]。脑启发（类脑形态）计算模拟通过内存和处理器的结合模仿了大脑的工作原理，这可以解决传统的冯·诺伊曼结构中著名的“内存墙”问题，其能量利用率非常高（低能耗）[14]。最近研究发现，脉冲神经网络（SNN）[14]和忆阻器交叉阵列[15]很有可能解决集成芯片的存算一体问题。不过，这些方法的功耗和速度仍然受限于电子器件的散热、电容和射频串扰[16]。

通过利用光的独特性质，光计算提供了另一种解决方案[17–18]。由于光的带宽更宽，信息可以在不同波长上高度并行处理，而相邻频率（不同波长上）的信息互不干扰[19–20]。此外其他维度，如偏振和轨道角动量也可以增加带宽[21–22]。利用光频梳对不同波长的数据和权重进行编码，这一操作在一个时间步长内自然而然可以被同时运用到成百上千或者数百万的频道上。此外，光可以在相位维度上以低能耗的形式进行调制[23]。光纤和波导的发展使得光可以以前所未有的光速传播很长的距离，且损耗极小[24]。所有这些因素都使得光子比电子更适合作为信息载体，尽管光子是不易被操控的。远程通信就是一个典型的例子，通过光纤传输的数据比传统电缆传输的数据流量更大而能耗更低。此外，由于高带宽密度，在云计算中应用芯片对芯片的光学互连变得越来越流行[25]。

尽管光计算因其可拓展的架构有着悠久的历史[26–29]，但将光子应用于计算来实现与前几十年最先进的电子处理器相媲美的性能仍然是一个巨大的挑战，主要是由于缺乏合适的集成架构和集成光子器件，包括高效、高速的光电调制器[30–31]、探测器[32]、低功耗光频梳[33]和非线性非易失性光学材料[34]。近年来，各种集成光子器件和新型材料的发展以及人工智能的兴起，似乎为光计算的复兴提供了一个巨大的机遇，同时，人们提出不同类型的光学神经网络（ONN）和光子处理器，来解决AI计算能力所遇到的瓶颈。这项研究比以往任何时候都更加实用，因为我们不需要完全取代电子计算机，而是努力在特定的任务或操作中找到一个光计算的切入点。作为未来全光计算机的过渡阶段，也许一个光电混合框架可以促进电子和光子的互补，使其在人工智能特定应用[35]上发生数量级的提升。

因此，AI的特定光计算是一个相对较新的领域，但是其已经在过去五年里特别受到人们的青睐并具有很强的竞争力，几乎在各个领域都有很大发展。由于集成光子处理器的发展必然需要多学科共同努力，因此，这个令人兴奋的领域必然会聚集越来越多的研究。本文首先简要回顾了类光计算在不同AI模型中的最新突破，以及它们在

解决各种不同应用问题中的独特优势和在实际实现中仍然存在的挑战，接着分别概述了不同类型的AI模型。第2部分主要介绍了前馈ONN，分别描述了用于不同数学运算的不同光学方法，第3部分介绍了光学蓄水池计算（RC），第4部分简要介绍了其他AI模型，最后，在第5部分总结了几个有待解决的重要挑战。

2. 前馈光学神经网络

在前馈人工神经网络中，相邻层的所有神经元以不同的突触权值相互连接。对于每个神经元，首先进行线性加权求和运算。假设从上一层传递进来的信号共有 N 个，用列向量 $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ 表示（其中， x_N 是第 N 个信号强度； N 是总的输入信号； T 是转置矩阵）。其对应的权重系数由另一个行向量 $\mathbf{w} = [w_1, w_2, \dots, w_N]$ 表示（其中， w_N 是第 N 个权重系数； N 是加权系数的总数）。单个神经元的线性求和 $\mathbf{y} = \sum_{n=1}^{n=N} \mathbf{x}_n \mathbf{w}_n$ ，将触发非线性激活函数 $f(\mathbf{y})$ 。这个神经元的最终输出是 $f(\sum_{n=1}^{n=N} \mathbf{x}_n \mathbf{w}_n)$ 。如果考虑所有 M 个神经元在同一层，则整个线性操作就是向量矩阵的乘法， $\mathbf{y}' = \mathbf{w}' \mathbf{x}$ ，其中， \mathbf{y}' 是由 M 个不同神经元的线性加权求和组成的向量， \mathbf{w}' 指 $M \times N$ 是由所有权重系数组成的矩阵。与人工神经网络的电子实现相比，光学实现可以以更低能耗光速进行各个神经元的并行计算[36–38]。

2.1. 光学线性加权求和

对于全连接神经网络的光学实现，需要光学地进行加权求和运算或向量矩阵乘法运算。到目前为止，已经有多种不同光学加权求和的实验实现，见下文。

2.1.1. 级联的马赫-曾德尔干涉仪

文献[36,39–50]中，输入和输出矢量由不同端口的多个相干光信号强度表示。作为矢量矩阵乘法的光计算装置，其由众多的级联马赫-曾德尔干涉仪（MZI）组成一个相互连接的网络。级联MZI的基本概念可以追溯到1994年的早期工作[39]，最近的工作[36]提出了一个包含56个MZI的硅光子集成电路的可编程纳米光子处理器。

数学上，通过奇异值分解（SVD），一个普通的实矩阵 \mathbf{w}' 可以被分解为 $\mathbf{w}' = \mathbf{U} \mathbf{\Sigma} \mathbf{V}$ ，其中， \mathbf{U} 和 \mathbf{V} 分别是 $M \times M$ 和 $N \times N$ 的幺正矩阵， $\mathbf{\Sigma}$ 是 $M \times N$ 的非负定对角矩阵。 \mathbf{U} 和 \mathbf{V} 都可以利用MZI的网格用光学实现， $\mathbf{\Sigma}$ 可以通过光衰减器实现。作为系统的基本单元，MZI有两个分束器和两个可调谐的移相器，如图1所示[47]。需要注意的是，一个MZI并不等同于人工神经网络中的一个神经元，但是一个级联

MZI系统可以在一个神经网络中的两层神经元之间同时执行相同的线性向量矩阵乘法运算。MZI网络设计时需要优化所有MZI的互连拓扑和每个移相器的移相角。对于一个典型的级联MZI系统，一个 $N \times N$ 的矩阵乘法至少需要 $N \times (N-1)/2$ 个分束器[41]。对于向量矩阵计算，MZI网络的构建方法并不是唯一的[41–42]。例如，可以为相同的向量矩阵计算设计两种不同的MZI网络，分别具有更可调或更容错的特点[42]。人们已经研究了MZI网络对实际不完美光子元件对应的工艺误差和不精确度的鲁棒性[42–43]。通过三维（3D）打印光波导，级联MZI系统可以从二维（2D）扩展到3D，这可以实现更高的计算能力[44–45]。复数运算还可以有效地在该架构实现[46]。人们已经尝试将级联MZI用于各种AI任务，如语音识别[36]、花卉数据分类[47]和葡萄酒数据分类[47]。

2.1.2. 深度衍射神经网络

在深度衍射神经网络（D²NN）结构中[51]，级联衍射光学元件（DOE）垂直于光的传播方向并互相之间有一定空间分隔距离，如图2所示[51]。输入和输出平面上的空间光强分布分别对应输入和输出向量。入射光场在自由空间中向前传播，并由每个DOE依次调制。类似于深度学习中的对应算法，所有DOE的像素值（仅相位、仅幅值或复振幅值）是通过误差反向传播算法进行优化的。优化后的系统从输入光场到输出光场通过线性变换实现向量矩阵乘法。D²NN系统的信息处理能力取决于衍射层的数目[52]。该体系结构可以充当一个线性分类器，并已被证明能够以适当的精度[51]对MNIST数据库中的数字图像和Fashion-MNIST数据库中的服饰产品图像进行光学分类。后来的相关研究在不同方面改进D²NN结构[53–68]。原始的D²NN是由太赫兹源驱动的，一个带有红外源的体积更小的系统在参考文献[53]中被提出。D²NN的实现不局限于单色相干光，宽带D²NN还可以用于光谱滤波和波长解复用应用[54]。D²NN可以与数字处理器[55–57]或数字神经网络[58]相结合来提高推断能力。一种类似于残差

深度学习网络的残差D²NN结构已经被提出，其可以通过多个反射镜简单地实现输入输出之间的直接快捷连接[59]。无透镜自由空间传播可以用透镜的光学傅里叶变换来代替，从而提高其预测精度[60]。通过对训练图像进行几何变换，可以增强D²NN系统的平移、旋转和尺度不变的识别能力[61]。仿真结果表明，D²NN系统能够解决更多样化的计算机视觉任务，比如，图像分割[60]、特征检测[60]和图像超分辨率[59]以及目标分类[51]。

除了传统的机器学习任务外，D²NN还可以应用于其他的光计算和光信号处理任务，包括激光雷达的智能波束控制[62]、图像加密[63]、光学逻辑门[64]、脉冲整形[65]以及模分复用光纤通信中的模式识别/复用/解复用[66–68]。

2.1.3. 基于空间光调制器和透镜的光计算

与前两种使用相干光的架构不同，这种基于空间光调制器（SLM）和透镜的架构不受限于相干光和非相干光照明[69–74]。照亮一个SLM平面的不同像素（或部分）的入射光强度分布携带输入矢量值。SLM的不同像素通过不同加权系数进行编码。其中，SLM放置在透镜的后焦平面，而光汇聚到焦点。探测器放置在透镜的焦点，然后收集SLM平面上的总光强作为输入向量和加权系数矢量的内积。图3（a）所示的装置类似于光学成像中的单像素成像（或鬼成像）概念[70–71]。

通过使用不同类型的透镜，有许多方法可以实现完全的向量矩阵乘法。通常，可以使用傅里叶透镜实现沿同一方向输出光的叠加，沿不同方向平行光的不同焦点对应于输出向量的不同分量[69]。另一种选择是使用柱面透镜[73–74]。柱面透镜只在水平方向或垂直方向的平行光和会聚光之间进行转换。输入向量由沿水平像素化阵列的光强分布表示。两个柱面透镜在SLM平面上实现一维（1D）输入阵列的扇出复制和1D加权求和，如图3（b）所示[74]。最后，用输出焦平面上沿垂直阵列的光强分布表示乘法运算结果。此外，向量矩阵乘法可以通过简单地用一个傅里叶透镜多次重复加权求和操作来实现。

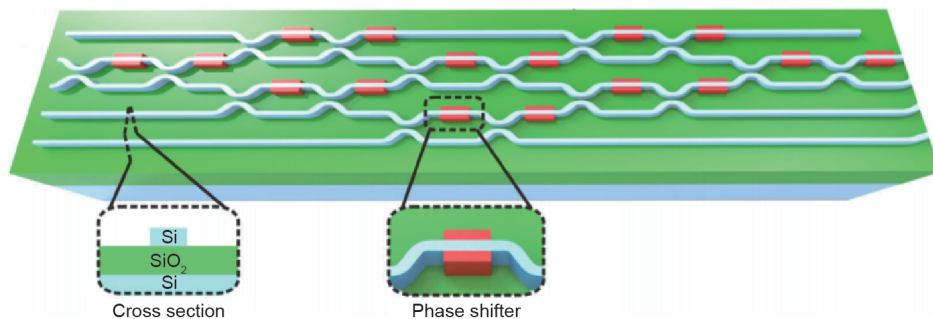


图1. 线性向量矩阵乘法的级联MZI的系统。Si: 硅材料；SiO₂: 二氧化硅材料。经 The Optical Society of America, ©2019 许可转载自文献[47]。

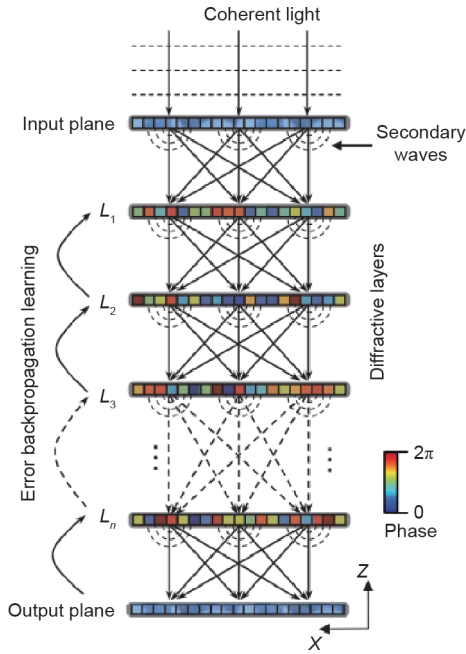
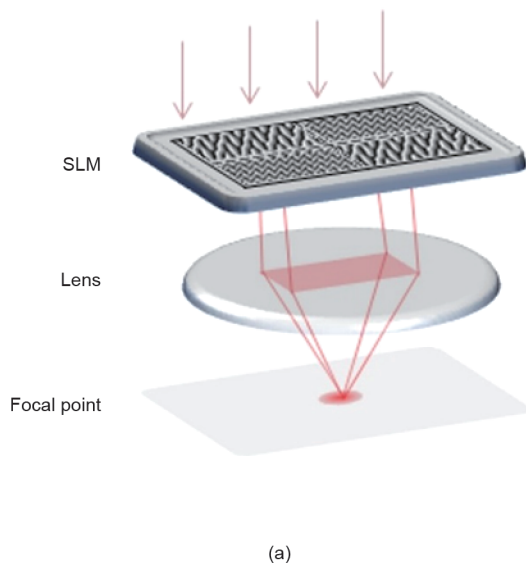


图 2. D^2NN 的光学系统。 L_1 、 L_2 、 L_3 和 L_n ：第一、第二、第三和第 n 个衍射层。经 The American Association for the Advancement of Science, ©2018 许可转载自文献[51]。

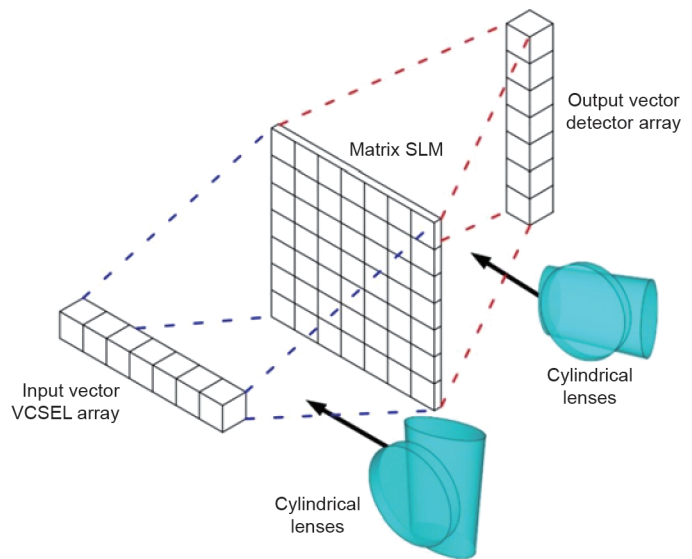
基于 SLM 和透镜的体系结构可以很容易地与冷原子系统相结合，实现具有线性和非线性变换的全光学深度神经网络[69,72]。基于该结构可以构建包含 174 个光学神经元的大规模可编程全光深度神经网络[72]。与其他结构相比，其可实现对不同任务进行可重复编程，而无需改变物理装置。

2.1.4. 波分多路复用

该方法[34,75–81]采用波分多路复用（WDM）方式实



(a)



(b)

图 3. (a) SLM 和傅里叶透镜的线性加权求和运算[69]；(b) 用 SLM 和柱面透镜进行的向量矩阵乘法[74]。(a) 经 The Optical Society of America, © 2019 许可转载自文献[69]。

现线性加权求和。输入矢量中的每个元素都由一个特定频率（或波长）的光波表示，然后通过微环谐振器（MRR）构建的光子权重组对其进行不同的光谱滤波；从而对输入信号进行加权，如图 4 所示[75]。平衡光电二极管（BPD）通过采集不同波段所有信号的总光强来实现加权线性求和。这种结构被认为有可能与主流的硅光子器件平台兼容。参考文献[78]提出了一种简化的微环权重组反馈控制方法，以产生良好的滤波响应。改进后的 MRR 控制方法可以兼容大规模光子集成电路[80]。该结构已用于主成分分析计算[81]。在最近的工作[19,38,82]中，WDM 架构与基于光子芯片的微频梳相结合，这可以显著提高数据处理速度和容量。另外可以将非易失性相变材料（PCM）集成到波导上来实现芯片上局部存储加权值[38]。文献[82]中的光矢量卷积加速器每秒可执行超过 10 万亿次运算。

2.1.5. 其他方案

矢量矩阵乘法也可以基于光在纳米光子介质中的传播实现，如图 5 所示[83]。进入介质的输入光强分布表示输入图像模式。光场通过介质后的输出强度分布表示计算结果，比如图像分类结果。介质由不同光学系数的基底材料和夹杂物材料组成，夹杂物材料可以产生更强的散射光。通过适当优化夹杂物的位置和形状，设计介质内部光场的散射，实现其与权重矩阵的矢量乘法。在最近的一项研究中[84]，为了降低设计的复杂性，连续介质被优化带状的多层超表面所取代。

在基于相干矩阵乘法器的架构中[85]，输入矢量和权重值都被编码为多通道的相干光信号。零差探测器采用量

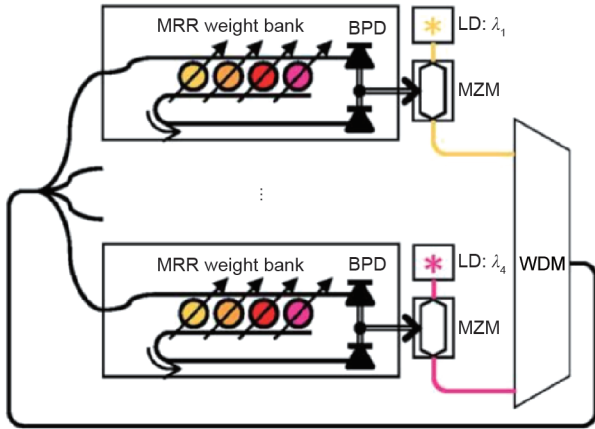


图 4. 带有 WDM 和 MRR 的光计算。 λ_1 和 λ_4 : 权重组的第一和第四个 MRR; LD: 激光二极管; MZM: Mach-Zehnder 调制器。经 Springer Nature, ©2017 许可转载自文献[75]。

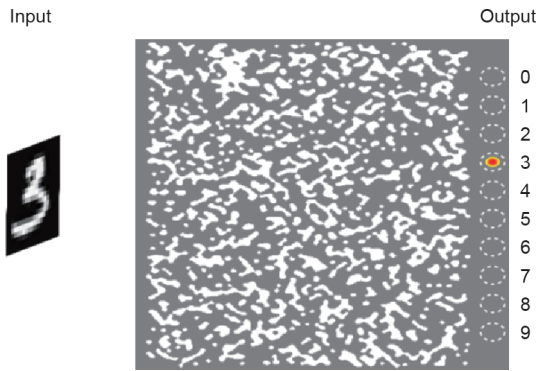


图 5. 用于光计算的带有夹杂物的纳米光子介质。灰色: 主体基底材料; 白: 夹杂物。经 China Laser Press, ©2019 许可转载自文献[83]。

子光电倍增过程实现线性加权求和。相干矩阵乘法器在实现大规模、快速可编程的光子神经网络方面具有一些潜在的优势。

此外, ONN 可以在多芯光纤中实现。每个输入、输出和加权值都由多芯光纤中每个芯的光强表示。表示权重的芯经过掺铒, 可作为可控放大器。光信号通过光耦合在芯间横向传输来模拟神经网络中信号的前向传播。参考文献[86–87]中的仿真结果表明, 多芯光纤可以构建三层神经网络。

2.2. 光学线性卷积

与全连接神经网络相比, 卷积神经网络 (CNN) 中的神经元连接更为稀疏。此外, 多个连接具有相同的权重值。相邻两层神经元之间的卷积运算本质上是一种简化的向量矩阵乘法运算。

数学上, 输入图像与核函数的卷积等价于输入图像的傅里叶滤波。在光学中, 包含图像信息的输入光场的傅里叶变换和反傅里叶变换可以很容易地通过 $4f$ 双透镜装置来实现。根据卷积核设计的滤波掩模可以放置在 $4f$ 装置

的傅里叶平面里。之前的工作[37,88–90]证明了这种光学卷积操作可以显著降低 CNN 的计算成本。文献[91]进一步实现了在傅里叶域进行卷积的广义光子张量运算处理器。此外, 卷积运算也可以通过其他方式实现, 比如, 级联声光调制器阵列[92]、WDW 加光延迟线[93]或色散诱导延迟[82]和 3D 打印光子波导电路[44]。

2.3. 光学非线性激活函数

如果一个神经网络是完全线性的, 没有任何非线性激活函数, 那么在物理上可以有多层线性变换, 但由于多个矩阵的乘法结果仍然是一个矩阵, 因此有效的计算总是等价于一层。为了实现全光深度神经网络, 非线性激活函数的实现是必不可少的。不幸的是, 光学中的非线性行为并不容易通过实验实现, 这使得实现可行的非线性激活函数成为 ONN 研究中最具挑战性的问题之一。在许多 ONN 研究中, 只在光学上实现线性运算, 而非线性运算通常是通过模拟实现的[36–37,58,60,89,94]。

一般情况下, 非线性激活函数可以由电子元件和光学元件组成的混合系统来实现[36,55–56,95–96], 但这种方式不可避免地会由于光信号和电信号之间的相互转换而降低光计算性能。理想的方案是使用纯光学元件实现非线性激活函数, 最近通过使用电磁感应透明 (EIT) 证明了这是可能的[69,72]。

EIT 是在一束光的透明度受另一束光控制的介质中的相干光学非线性效应, 如图 6 所示[69]。理论上, EIT 可以发生在任何具有三重态的真实材料系统中, 其中不同态之间的跃迁概率幅值取决于它们的粒子分布。在参考文献[69,72]中, Zuo 等在 2D 磁光阱 (MOT) 中使用冷原子系统的非线性激活函数构造了一个完整的全光实现的 ONN, 并利用其对 Ising 模型的不同相位进行了分类。

此外, 还有其他实现非线性激活函数的方法[94], 比如, 反向饱和吸收 (RSA) 等, 但到目前为止, 很少有方法在实验中被证明是可行的。

2.4. 在线光学训练

深度学习网络通常首先通过大量的样本进行训练, 以优化所有的权重系数。经过训练后, 网络可以对测试样本有效地执行某一推理任务。在许多 ONN 相关研究中, 比如参考文献[36,51], 训练在计算机上完全数字化离线进行, 只有 ONN 的推断是通过光学进行的。但是, 研究者在一定程度上也尝试了 ONN 的在线训练。

MZI 网络的在线训练已经在一些工作中进行了研究

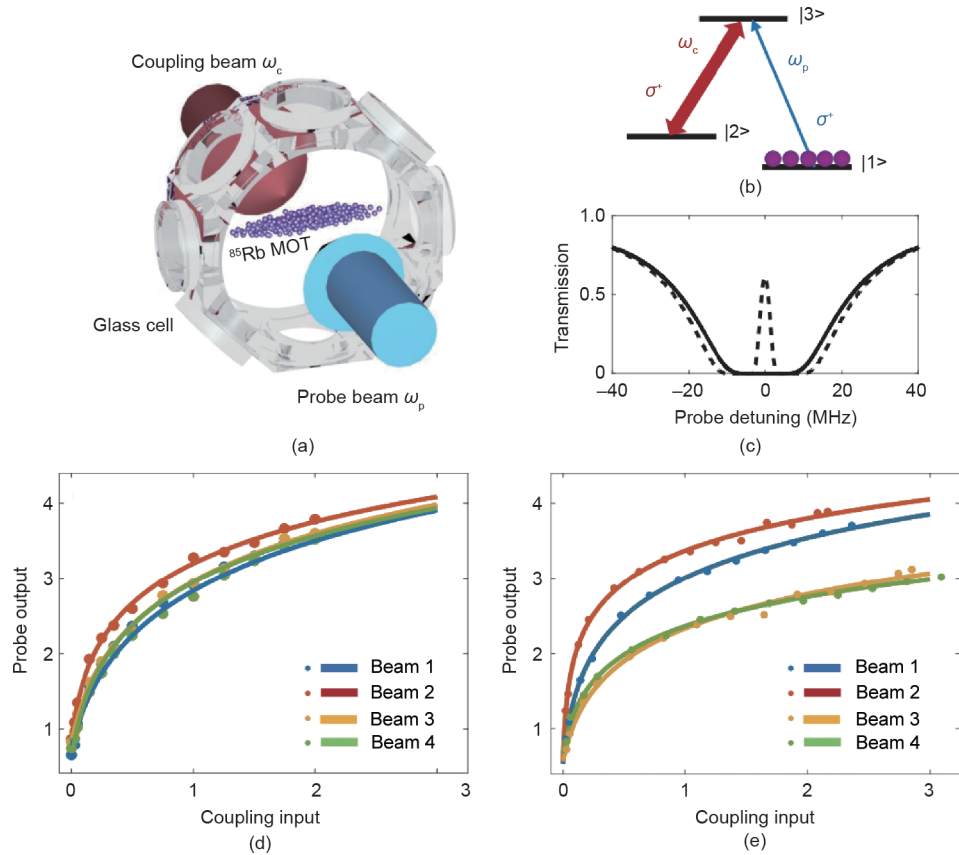


图6. 利用EIT实现非线性激活函数。 ω_c : 耦合束的频率; ω_p : 探头光束频率; σ^+ : 圆偏振光。经 The Optical Society of America, ©2019 许可转载自文献[69]。

[36,47–48]。误差反向传播是数字神经网络中应用最广泛的训练算法。用伴随变量法可以推导出复杂的MZI网络对应的反向传播算法的光学模拟[48]。在训练中，计算中的梯度可以通过MZI网络中不同节点的光强度来测量。另外，可以使用仅需正向传播而无需反向传播的有限差分方法对级联MZI系统进行片上训练[36]。

遗传算法和粒子群算法是对MZI网络的参数进行优化的两种神经演化算法，其可用于在线训练[47]。这些全局优化算法是无梯度的，并且可以避免局部极小值问题。

参考文献[56]已经研究过D²NN结构的在线训练方法。DNN的标准训练算法是基于误差反向传播，而光互易和相位关联原理是通过测量向前和向后传播光场得到的原位梯度。可以使用SLM替代每一层的静态DOE来进行动态训练和参数更新。实际输出光场与目标光场之间的残差是由一个用于光场反向传播的复杂场产生模块(CFGM)生成的。仿真结果表明，与离线计算机训练相比，该在线训练方法可以节省大量的计算成本。

如果存在合适的优化算法、精确的实验测量和动态编码的光学元件，则其他ONN架构也可以实现片上在线训练。

3. 光学蓄水池计算

光学蓄水池计算(RC)是一种来自递归神经网络中神经形态的计算框架，其允许快速学习和各种硬件系统的简单实现。由于其在处理暂存数据方面的特殊优势，该结构在执行时序信号预测和语音识别等时序相关任务时可以得到最佳优化效果。RC结构通常由一个固定的非线性机制(即储存节点)组成，它允许输入信号在更高维空间中转换成时空态。通过训练一个储存态的读取器来获取蓄水池动态，从而得到储存的时间输出数据。本小节回顾了RC在集成光学系统或空间光学衍射系统中全光学或光电混合的实现方式。

3.1. 全光学RC

RC的全光实现和光电实现之间的差异取决于不同类型的蓄水池和输入层。接下来，我们将全光结构从空间分布的RC引入到基于时延的RC。超高速计算和低功耗是全光RC系统的主要优点。

3.1.1. 基于衍射耦合的空间分布RC

空间分布光子RC是利用DOE实现光学RC的一种方

法。2015年，Brunner和Fischer [97]提出了一种垂直腔表面发射激光器（VCSEL）网络。图7为普林斯顿光电公司的芯片示意图[97]。

该RC结构已通过DOE实现了激光耦合和衍射复用。图7展示了实验实现过程，其介绍了由VCSEL组成的图像结构。此处，VCSEL阵列晶格间距与成像透镜焦距的组合可以在相邻激光器的主射线之间形成一个角度，这个角度可以通过透镜的焦距来调节。

固有的生产过程使得位于阵列上的激光器受到整个激光集成电路参数变化的影响。因此，所展示的与周期DOE的衍射耦合易受这些差异的影响。除了网络耦合外，衍射成像还允许多个激光器进行平行光调制，而激光耦合的主要局限是由图像的光学像差决定的。该结构具有相当大的可扩展性[98–99]，其可扩展到一个由数百个节点组成的网络，且节点的放置面积小于 1 mm^2 。对于技术相关的实现，首先必须确保其引入方法的小型化。特别是由于

注入锁定现象，其需要在波长上均匀分布[100]。

3.1.2. 基于延迟的RC

为了简化复杂的光学硬件，Appeltant等[99]以及Paquot等[101]引入了一种新型的RC，即基于延迟线RC。他们将一个反馈结构分成一系列虚拟节点和一个非线性节点形成的一个循环，从而形成延迟反馈。与空间分布的网络自由度高的RC相比，这类储存是固定的，其可以被用于更传统的光通信硬件。基于延迟线的RC示意图如图8(a)所示[99]，下面我们展示了基于延迟线RC中的两种典型的延迟。

我们首先讨论了第一种基于延迟的RC，其光输入是全光蓄水池。这种RC包含几种基于半导体激光器[102]、半导体光学放大器（SOA）[103]或无源光腔[104]的实现。Duport等[103]和Brunner等[102]利用有源器件引入了两种不同类型的非线性响应：SOA和半导体激光器。在输入端注入调制光场，检测后通过输出层。这个RC范式

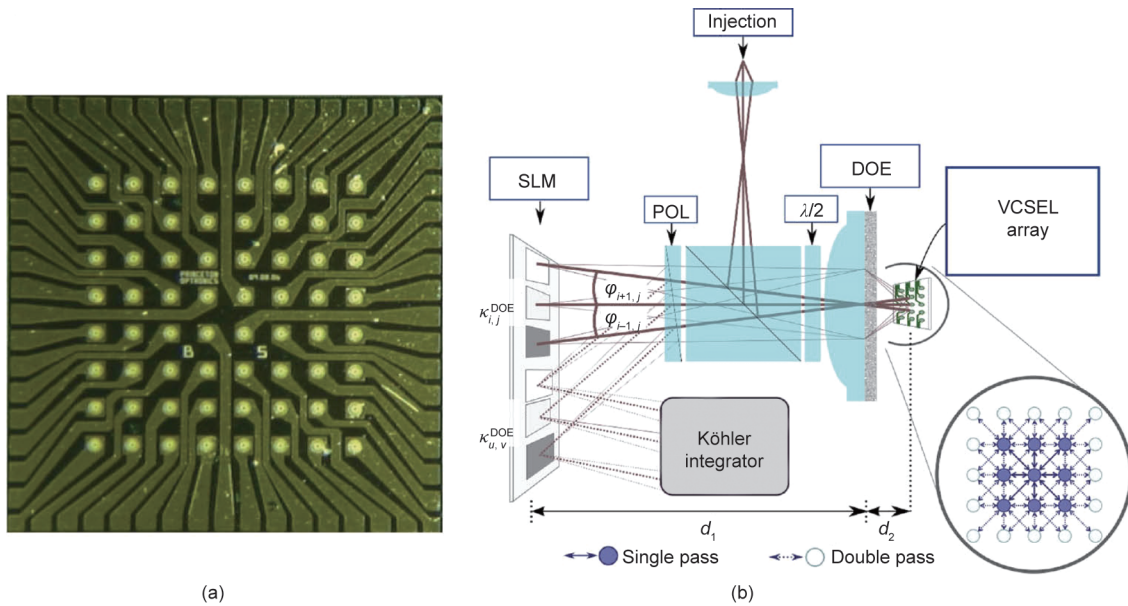


图7. (a) VCSEL激光二极管阵列；(b) 通过实现所设计的衍射谐振器来实现激光间的耦合。 d_1 : 像距； d_2 : 物距； λ : 波长； φ : 偏置角； κ^{DOE} : DOE的效率；POL: 偏振片[97]。经 The Optical Society of America, ©2015 许可转载自文献[97]。

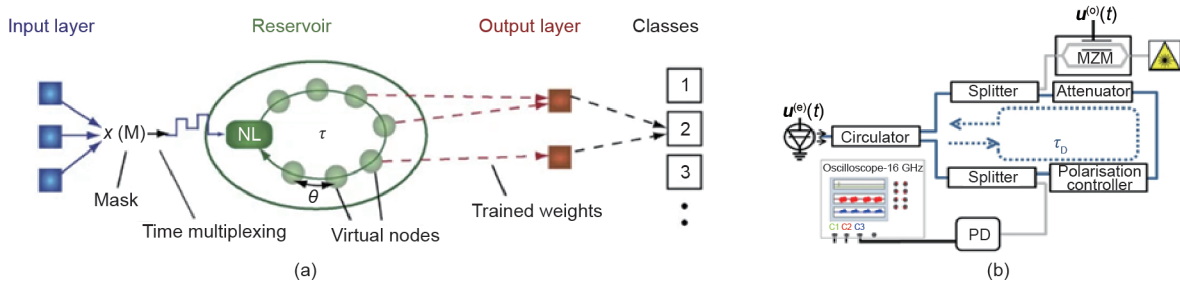


图8. (a) 基于延迟线的RC框架。延迟回路由许多间隔隔开，形成蓄水池。蓄水池包含沿延迟线的虚拟节点[99]。(b) 基于半导体激光器的蓄水池框架；它利用一系列光学器件建立RC。NL: 非线性节点； τ : 反馈回路中的延迟；PD: 光电二极管； $U^{(e)}(t)$: 电气信息输入； $U^{(o)}(t)$: 光学信息输入； τ_D : 反馈延迟[102]。(a) 经 Springer Nature, ©2011 许可转载自文献[99]；(b) 经 Springer Nature, ©2013 许可转载自文献[102]。

具有处理复杂计算任务的潜力。此外，光学蓄水池能够以相对较高的速率处理信息，速率达每秒千兆字节。图8(b)为该系统的原理图[102]。通过半导体和光反馈注入的外部输入层可用于形成光学蓄水池。Hicke等[105]和Nakayama等[106]认为，他们的方法可以提高其处理能力。

此外，基于延迟的光学蓄水池可以在芯片上实现。Nguimdo等[107]证明，必要的光学偏置注入可以增加半导体激光器的光学调制带宽，允许比Brunner等[102]的系统更短的延迟时间。基于光反馈的片上半导体环形激光器可以同时解决两个不同的任务，即分类任务和时间序列预测任务[108]。通过Nguimdo等[107]的偏置注入方法可以增加半导体激光器的带宽，从而得到比Brunner等[102]的系统更短的延迟时间。分类和预测任务[108]可以通过一个基于光学反馈的片上半导体激光器来解决。如果对读出层稍作调制[106]，使用短的外腔可以避免相位灵敏度[109]。无源器件的发展必然要求开发低功耗、高效率的光学计算机。Dejonckheere等[110]将半导体可饱和吸收镜(SESAM)置于环形光学腔中。Vinckier等[104]研究了一种相干驱动无源腔下的光子RC。如果我们在蓄水池计算机中进行非线性输出，就有可能使用线性光纤腔，其可以很容易地完成基于语音信息的数字识别等任务。

3.2. 光电 RC

基于光电系统的RC在语音识别[111–113]、混沌时间预测[111,114–115]以及雷达信号预测[116–117]等多个领域得到了广泛的应用并扮演了很重要的角色。目前光电RC的处理速度已经实现了兆赫速度，在可预见的未来，其可能发展到千兆赫速度[118]。

3.2.1. 芯片级别的RC

2008年，Vandoorne等[119]宣布了第一个基于绝缘体

上硅的硬件RC，其由波导、光组合器和光学分束器组成。研究发现，通过标准的快速光电二极管检测可以解决非线性问题。然而，由于该系统依赖于探测器中的光电转换，因此不能完全采用光学操作[119–121]。

使用无源器件的优点是其宽带宽，甚至可以同步发送多个波长到系统中，从而实现离散波长的并行操作系统。新的学习技术可以用来适应长期的相移。其缺点之一是随着芯片扩展到更多的节点，对应的光学损耗可能是相当大的，很难实现对所有节点的并行测量。在无源器件网络中，输入信号时钟和储存时间的频率依赖于离散节点之间的传播延迟，这需要高达每秒数百Gbit的高注入频率[122]。

通过各种任务对RC的计算性能进行了估计。为此，研究人员需要通过在某一点将光学信号注入蓄水池来简化软件RC中的随机输入连接性。当处理较长时间序列数据时，需要更大的芯片。如图9所示，实验结果仅为数值，其可以构造为一个集成滤波器[122]。在实验和系统的数字化应用中，还需要进一步评估语音数字的识别等任务。在所有测试中，系统都输出了足够的结果。此外，还可以在硅芯片上实现类似池田环形光腔[123–124]的光电实现，从而构建基于延迟的RC。Larger等[111][图10(a)]和Paquot等[112]分别研制了这种RC的光电实现。

光电RC的大多数硬件实现中其输入和输出层是在一个标准计算机上的离线仿真，其中计算机在硬件上的模拟RC可以实现大多数输入和输出层。然而，通过在模拟硬件上实现的三层RC已经建立了带有光电器件的单层蓄水池计算机[116]。输入层可以通过具有两种不同正弦频率的掩模实现。双输出Mach-Zehnder调制器(MZM)包含从光学蓄水池获取的信号以及RLC滤波器对BPD输出信号的滤波，如图10(b)所示[116]。整个模拟系统的输出成为蓄水池计算机的终端输出。其读出层如图10(b)

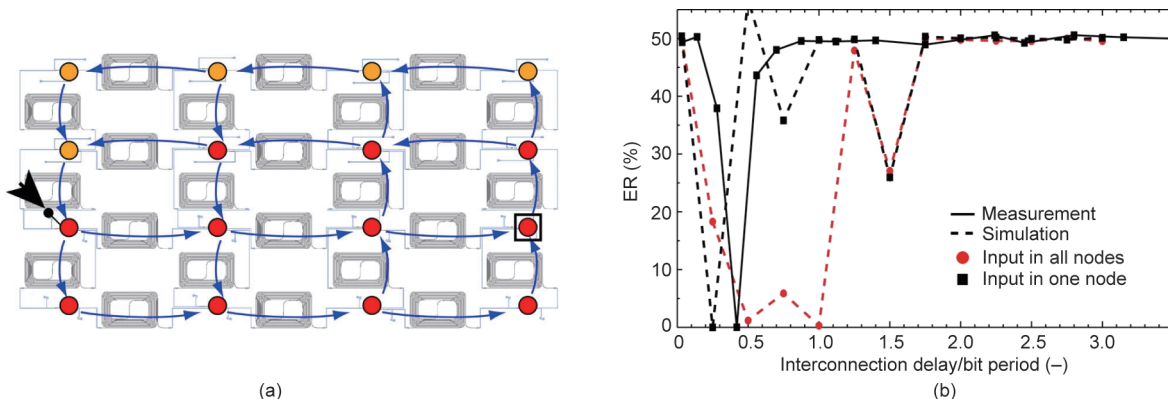


图9. 基于无源硅片的蓄层。ER: 出错率; “-”表示前一个比特编号。经 Springer Nature, ©2014 许可转载自文献[122]。

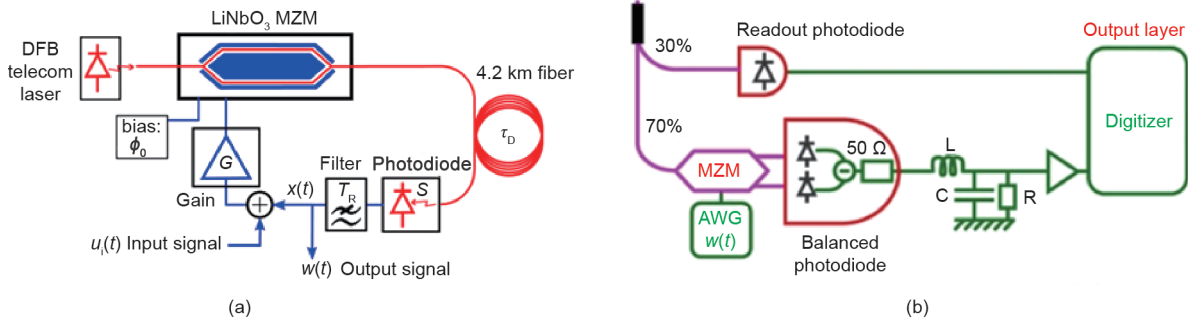


图 10. (a) 光电蓄水池计算机架构。光学路径用红色表示，电子路径用蓝色表示。(b) 模拟读出层架构。DFB: 分布式反馈; $x(t)$: 延迟信号; τ_D : 延迟时间; $w(t)$: 连续时间加权函数; AWG: 任意波形发生器; R、L、C: 电阻器、电感、电容。(a) 经 The Optical Society of America, ©2012 许可转载自文献[111]; (b) 经 Springer Nature, ©2016 许可转载自文献[116]。

所示。

虽然人们很少考虑到训练时延 RC 的读出权重，但 Antonik 等[125]提出使用可编程门阵列 (FPGA) 作为一种新的在线训练权值的专用硬件。这种光电实现已经很好地应用于神经信息处理[125]和混沌系统仿真[126]。FPGA 可以以数百兆赫的速度处理信息。此外，传统的训练算法，比如遗传算法和梯度下降算法可以用于上述提到的在线训练。其还可以实现多个延迟来提高光子延迟 RC 的性能[113,127–128]。

3.2.2. 自由空间光学的 RC

RC 光子实现的另一种方法是利用特殊光学元件的可激性。基于脉冲 (spike) 模型的神经形态可激光器件可以通过半导体技术实现[77,129–133]。因此，这种可激发的纳米光学器件的网络实现类似于神经形态的激发，这大大减小了功耗[134]。当光强增加到一个稳定的阈值内时，激光器被转化为一个激发态系统，如 spike。2016 年，Shastri 等[132]用两个可激激光器完成了其对应的数字实验。他们评估了一个基于半导体激光器的电光激发的更大网络[135]。利用 WDM 技术就可以将系统中的多个节点集成在一个芯片上。2011 年，Barbay 等[129]建立了另一种结构，即单片半导体微柱激光器的可激神经元样脉冲行为。该结构具有不应期[136]。微环和圆盘激光器也表现出了光学可激性[133–134]。与以往的设计相比，这些同步激光器件的机制取代了饱和吸收体，并表现出内部对称性破缺的特性。到目前为止，最新的研究进展表明，光学技术可以以空前速度实现 SNN。

最近，人们提出了一种基于空间可扩展的数字微镜器件 (DMD) 和 SLM 光子库的新方法[137–141]。2018 年，Bueno 等[137]证明了一个带有 SLM、DOE 和相机的光学系统可以存储多达 2500 个衍射耦合光子节点的网络。Dong 等[138]通过二进制方式调制光强，并用 DMD 对输

入信息和蓄水池进行编码。后来，他们采用了相同的方法，使用纯相位 SLM 制得了一个大规模的光蓄水池网络[139]。借助空间相位分布，蓄水池网络可以引入输入和蓄水池信息。2020 年，Rafayelyan 等[140]发现，强散射介质在光学网络中起着关键作用，其保证了众多光学节点之间的随机耦合权重以及网络中的并行处理。多维大混沌系统的预测任务已在其大系统中得到了验证，其具有较高的预测精度、较高的速度以及较低的功耗。该网络具有良好的可扩展性和处理大数据集的能力。

3.3. 展望

随着纳米光学与 RC 复合结构的出现，一种新的光学信息处理方法正在引领这一领域的发展。与传统方法相比，该方法具有低功耗、超高速等优点。对于光学 RC 未来的发展潜力，超高速光学非线性和超低功耗的光学器件在该领域可能会有广阔的发展前景，其可以在速度和功耗方面大大提高蓄水池性能[142–143]。此外，多节点库光学芯片的可扩展性也是未来研究的需要。

4. 脉冲神经网络

4.1. 国内外研究现状

脉冲神经网络 (SNN) 被认为是生物学和神经科学的交叉学科[144]。与前馈神经网络和递归神经网络相比，SNN 在神经形态计算中的应用更为广泛。除了 meta 模型外，SNN 中的神经元只有在膜电位达到阈值时才被激活。当神经元被激活时，就会产生信号并将其传递到级联神经元的膜电位。在 SNN 中，当前神经元的激活程度通常被建模为一个特定的微分方程。达到刺激峰值后，被激活程度将会增大，并持续一段时间，然后逐渐下降。

SNN 增强了时空数据的处理能力。一方面，SNN 中的神经元只与附近的神经元相连并单独处理，其增强了空

间信息的处理能力。另一方面，由于训练依赖于脉冲的间隔和二进制码中信息的缺失，其可以从脉冲的时间信息中找回，从而增加对时间信息的处理能力。事实证明，与传统的人工神经元相比，脉冲神经元是更好的计算单元。然而，由于SNN在训练和物理实现上的困难，其还没有得到广泛的应用。目前对SNN的研究大多集中在简单结构的理论分析和验证上。然而，越来越多关于SNN的研究正在关注其训练算法和硬件（光学）实现。

2016年，普林斯顿大学Prucnal研究小组提出了一种基于可激活石墨烯光纤激光器的脉冲处理系统[132]。该系统包含一个掺铒光纤（增益部分）、一个石墨烯饱和吸收体（损耗部分）、一个980 nm激光的泵浦源和一个1480 nm的激发系统的脉冲刺激信号，其可以产生基于带泄漏整合发放模型的脉冲神经元响应。2018年，研究团队提出了一种基于分布式反馈（DFB）的激光结构类脑形态的集成光路[145]。此外，研究小组还讨论了构建可编程级联光学神经网络的可行方案，包括传播-加权（B&W）网络原型[132]和相干光学架构。其中，B&W网络原型是一种支持大规模光学脉冲神经元并行互连的网络架构。下面SNN的训练方法主要采用监督学习算法。这些监督学习算法包括：①SpikeProp算法[146]，该算法采用梯度下降法，通常用于多层前馈人工神经网络；②Tempotron算法[147]，利用脉冲序列输出与期望序列的差值来改变网络权重；③基于突触可塑性的算法，如Hebbian学习算法[148–152]；④远程监督学习算法，如ReSuMe算法[149]；⑤基于脉冲序列卷积的监督学习算法，比如，脉冲模式关联神经元（SPAN）算法[150]和精确脉冲驱动（PSD）算法[151]。还有基于VCSEL [152]并使用垂直腔半导体光学放大器（VCSOA）实现脉冲时间依赖可塑性（STDP）的监督和/或无监督学习算法[153]。

2019年Feldmann等[34]提出了另一种基于PCM的架构，并利用该架构对光学神经元进行非线性变换。光控PCM是一种工作状态。材料在晶体和非晶体材料之间的工作状态由输入光强控制。当输入光强低于其阈值时，PCM处于晶体态，其可以吸收大量光强。当输入光强高于阈值时，PCM处于非晶态，大部分光强可以通过PCM。因此，将材料集成到光学传输介质中，可以根据输入的光强改变材料的光透射性。这样其就可以作为光学神经元的激活函数。

4.2. 神经形态光子学与电子学的比较

与类似TrueNorth [154]、Neurogrid [155]和SpiNNaker [156]等的SNN类脑形态的电子实现相比，类脑形态光

学利用光处理信息，其能够实现更高的带宽和数据吞吐量，以及更低的延迟。已有研究表明，采用硅/III-V混合平台的光电SNN可以达到20 GHz的处理速率，这比纯电子SNN高出至少6个数量级[157]。计算所需能量主要由激光源消耗，而无源过滤器损耗的电流很低，硅/III-V混合平台每次操作的能源效率是0.26 pJ，这比Neurogrid对应的能源效率高大约3个数量级，比SpiNNaker对应的能源效率高至少6个数量级。利用优化的亚波长结构（如光子晶体）和波长复用技术可以进一步提高类脑形态光学平台的计算性能。

5. 结论

虽然光计算在不同的AI模型中得到了广泛的应用，但由于各种挑战，其实际应用仍未得到明显优于传统电子处理器的性能证明。比如，如何在响应时间短、对探测光功耗要求低、数据并行的情况下获得显著的光学非线性特性[158]？如何在不同架构中优化非线性表征？如何在低功耗的芯片上实现高速大规模可重构？如何将不同的光学器件集成到单个芯片上，特别是与外部器件连接？如何为不同的任务设计具有自动化软件的ONN？尽管未来仍有许多问题有待解决，但目前的光学模拟计算技术在速度、数据并行和功耗方面已经表现出了光的独特潜力[158]。

下一步，我们认为需要投入更多的努力来克服光的关键缺点，并在不同的实际应用中展示光计算相对于电子学计算的优越性。如上所述，不同的架构可能适合于相应的特定任务。光计算在自由空间中对部分相干光场的传播和调制模型的建立有助于计算机视觉中对数据的高速处理。由于自动驾驶等自然场景中的大部分光线是非相干的，因此基于衍射神经网络的相干光模型还难以直接应用。由于非线性特性在深度神经网络中起着至关重要的作用，因此采用低功耗、低延迟的光学方法实现非线性激活函数可以显著提高当前ONN的精度。存储和访问是扩展类脑形态处理器所需的基础技术。然而，构建具有极快读写速度的光学随机存储器一直是一个挑战，这为光计算和存储的特殊材料的发展提供了巨大机遇。尽管系统体积较大，但自由空间的光计算可能会加速不需要便携式系统的各种数据中心的云计算。我们期待更先进的具有更高迭代速度（高达兆赫和千兆像素数）的SLM。高速、低功耗光电转换在目前的过渡阶段变得越来越重要，其在实际应用中需要光子和电子。所有的挑战和应用表明，在不久的将来，光计算会有越来越多的突破[159]。

综上所述，我们回顾了在过去五年中用于AI特定硬

件实现的光计算的发展。我们描述了几个在不同架构下具有代表性和独特优势的AI模型。尽管大多数方法仍处于概念验证的初级阶段，但我们预期这些架构的实际应用在未来发展中将在速度或功耗方面取得数量级的提高。我们相信，随着光子学、电子学、材料学、制造学、计算机科学和生物学等不同领域的不断努力，加速AI训练和推理的光电混合计算机将很快成为最终实现通用全光计算机的过渡阶段产品。

致谢

本工作得到了国家自然科学基金(61927802、61722209、61805145)、北京市科学技术委员会(Z181100003118014)和国家重点研发计划(2020AAA0130000)的支持。J. W.感谢国家博士后创新人才计划和清华大学水木学者计划的支持。J. L.感谢香港研究资助局(16306220)的支持。

Compliance with ethics guidelines

Jiamin Wu, Xing Lin, Yuchen Guo, Junwei Liu, Lu Fang, Shuming Jiao and Qionghai Dai declare that they have no conflict of interest or financial conflicts to disclose.

References

- [1] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing humanlevel performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision; 2015 Dec 7–13; Santiago, Chile. New York: IEEE; 2014. p. 1026–34.
- [2] Wu B, Iandola F, Jin PH, Keutzer K. Squeezedet: unified, small, low power fully convolutional neural networks for real-time object detection for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2017 Jul 21–6. Honolulu, HI, USA. New York: IEEE; 2017. p. 446–54.
- [3] Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G, et al. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016;529(7587):484–9.
- [4] Singh S, Okun A, Jackson A. Artificial intelligence: learning to play Go from scratch. *Nature* 2017;550(7676):336–7.
- [5] Wang Y, Ji M, Jiang S, Wang X, Wu J, Duan F, et al. Augmenting vascular disease diagnosis by vasculature-aware unsupervised learning. *Nat Mach Intell* 2020;2:337–46.
- [6] De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* 2018;24(9):1342–50.
- [7] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436–44.
- [8] Zhang W, Gao B, Tang J, Yao P, Yu S, Chang MF, et al. Neuro-inspired computing chips. *Nat Electron* 2020;3(7):371–82.
- [9] Adam GC, Khiat A, Prodromakis T. Challenges hindering memristive neuromorphic hardware from going mainstream. *Nat Commun* 2018;9(1):1–4.
- [10] Wu F, Lu C, Zhu M, Chen H, Zhu J, Yu K, et al. Towards a new generation of artificial intelligence in China. *Nat Mach Intell* 2020;2(6):312–6.
- [11] Huang C, Barbera SL, Small C. Building brain-inspired computing. *Nat Commun* 2019;10(1):1–6.
- [12] Zhang Y, Qu P, Ji Y, Zhang W, Gao G, Wang G, et al. A system hierarchy for brain-inspired computing. *Nature* 2020;586(7829):378–84.
- [13] Jouppi NP, Young C, Patil N, Patterson D, Agrawal G, Bajwa R, et al. Indatcenter performance analysis of a tensor processing unit. In: Proceedings of the 44th Annual International Symposium on Computer Architecture; 2017 Jun 24–28; Toronto, ON, Canada. New York: ACM; 2017. p. 1–12.
- [14] Pei J, Deng L, Song S, Zhao M, Zhang Y, Wu S, et al. Towards artificial general intelligence with hybrid Tianjic chip architecture. *Nature* 2019;572(7767):106–11.
- [15] Yao P, WuH, GaoB, Tang J, ZhangQ, ZhangW, et al. Fullyhardware-implemented memristor convolutional neural network. *Nature* 2020;577(7792):641–6.
- [16] Miller DAB. Attojoule optoelectronics for low-energy information processing and communications. *J Light Technol* 2017;35(3):346–96.
- [17] Burr GW. A role for optics in AI hardware. *Nature* 2019;569(7755):199–200.
- [18] Wetzstein G, Ozcan A, Gigan S, Fan S, Englund D, Soljačić M, et al. Inference in artificial intelligence with deep optics and photonics. *Nature* 2020;588:39–47.
- [19] Xu X, Tan M, Corcoran B, Wu J, Nguyen TG, Boes A, et al. Photonic perceptron based on a Kerr microcomb for high-speed, scalable, optical neural networks. *Laser Photon Rev* 2020;14(10):2000070.
- [20] Le ST, Aref V, Buelow H. Nonlinear signal multiplexing for communication beyond the Kerr nonlinearity limit. *Nat Photonics* 2017;11(9):570–6.
- [21] Devlin RC, Ambrosio A, Rubin NA, Mueller JPB, Capasso F. Arbitrary spin-to-orbital angular momentum conversion of light. *Science* 2017;358(6365):896–901.
- [22] Ji Z, Liu W, Krylyuk S, Fan X, Zhang Z, Pan A, et al. Photocurrent detection of the orbital angular momentum of light. *Science* 2020;368(6492):763–7.
- [23] Han JH, Boeuf F, Fujikata J, Takahashi S, Takagi S, Takenaka M. Efficient lowloss InGaAsP/Si hybrid MOS optical modulator. *Nat Photonics* 2017;11(8):486–90.
- [24] Ferrera M, Razzari L, Duchesne D, Morandotti R, Yang Z, Liscidini M, et al. Low-power continuous-wave nonlinear optics in doped silica glass integrated waveguide structures. *Nat Photonics* 2008;2(12):737–40.
- [25] Alexoudi T, Terzenidis N, Pitris S, Moralis-Pegios M, Maniotis P, Vagionas C, et al. Optics in computing: from photonic network-on-chip to chip-to-chip interconnects and disintegrated architectures. *J Light Technol* 2019;37(2):363–79.
- [26] Sawchuk AA, Strand TC. Digital optical computing. *Proc IEEE* 1984;72(7):758–79.
- [27] Wang W, Jin G, Yan Y, Wu M. Image feature extraction with the optical Haar wavelet transform. *Opt Eng* 1995;34(4):1238–42.
- [28] Solli DR, Jalali B. Analog optical computing. *Nat Photonics* 2015;9(11):704–6.
- [29] Farhat NH, Psaltis D, Prata A, Paek E. Optical implementation of the Hopfield model. *Appl Opt* 1985;24(10):1469–75.
- [30] Nozaki K, Matsuo S, Fujii T, Takeda K, Shinya A, Kuramochi E, et al. Femtofarad optoelectronic integration demonstrating energy-saving signal conversion and nonlinear functions. *Nat Photonics* 2019;13(7):454–9.
- [31] Li M, Ling J, He Y, Javid UA, Xue S, Lin Q. Lithium niobate photonic-crystal electro-optic modulator. *Nat Commun* 2020;11(4123):1–8.
- [32] Mauthe S, Baumgartner Y, Sousa M, Ding Q, Rossell MD, Schenk A, et al. Highspeed III–V nanowire photodetector monolithically integrated on Si. *Nat Commun* 2020;11(4565):1–7.
- [33] Stern B, Ji X, Okawachi Y, Gaeta AL, Lipson M. Battery-operated integrated frequency comb generator. *Nature* 2018;562(7727):401–5.
- [34] Feldmann J, Youngblood N, Wright CD, Bhaskaran H, Pernice WHP. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* 2019;569(7755):208–14.
- [35] Böhm F, Verschaffelt G, Van der Sande G. A poor man’s coherent Ising machine based on opto-electronic feedback systems for solving optimization problems. *Nat Commun* 2019;10(3538):1–9.
- [36] Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics* 2017;11(7):441–6.
- [37] Miscuglio M, Hu Z, Li S, George JK, Capanna R, Dalir H, et al. Massively parallel amplitude-only Fourier neural network. *Optica* 2020;7(12):1812–9.
- [38] Feldmann J, Youngblood N, Karpov M, Gehring H, Li X, Stappers M. Parallel convolutional processing using an integrated photonic tensor core. *Nature* 2021; 589:52–8.
- [39] Reck M, Zeilinger A, Bernstein HJ, Bertani P. Experimental realization of any

- discrete unitary operator. *Phys Rev Lett* 1994;73(1):58–61.
- [40] Carolan J, Harrold C, Sparrow C, Martín-López E, Russell NJ, Silverstone JW, et al. Universal linear optics. *Science* 2015;349(6249):711–6.
- [41] Clements WR, Humphreys PC, Metcalfe BJ, Kolthammer WS, Walmsley IA. Optimal design for universal multiport interferometers. *Optica* 2016; 3(12):1460–5.
- [42] Fang MYS, Manipatruni S, Wierzyński C, Khosrowshahi A, DeWeese MR. Design of optical neural networks with component imprecisions. *Opt Express* 2019;27(10):14009–29.
- [43] Miller DAB. Perfect optics with imperfect components. *Optica* 2015;2(8):747–50.
- [44] Moughames J, Porte X, Thiel M, Ulliac G, Larger L, Jacquot M, et al. Three-dimensional waveguide interconnects for scalable integration of photonic neural networks. *Optica* 2020;7(6):640–6.
- [45] Dinc NU, Psaltis D, Brunner D. Optical neural networks: the 3D connection. *Photonics* 2020;104:34–8.
- [46] Zhang H, Gu M, Jiang XD, Thompson J, Cai H, Paesani S, et al. An optical neural chip for implementing complex-valued neural network. *Nat Commun* 2021;12(457):1–11.
- [47] Zhang T, Wang J, Dan Y, Lanqiu Y, Dai J, Han X, et al. Efficient training and design of photonic neural network through neuroevolution. *Opt Express* 2019; 27(26):37150–63.
- [48] Hughes TW, Minkov M, Shi Y, Fan S. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* 2018;5(7):864–71.
- [49] Harris NC, Carolan J, Bunandar D, Prabhu M, Hochberg M, Baehr-Jones T, et al. Linear programmable nanophotonic processors. *Optica* 2018;5(12):1623–31.
- [50] Bogaerts W, Pérez D, Capmany J, Miller DAB, Poon J, Englund D, et al. Programmable photonic circuits. *Nature* 2020;586(7828):207–16.
- [51] Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M, et al. All-optical machine learning using diffractive deep neural networks. *Science* 2018; 361(6406):1004–8.
- [52] Kulce O, Mengü D, Rivenson Y, Ozcan A. All-optical information-processing capacity of diffractive surfaces. *Light Sci Appl* 2021;10(1):1–17.
- [53] Lu L, Zhu L, Zhang Q, Zhu B, Yao Q, Yu M, et al. Miniaturized diffraction grating design and processing for deep neural network. *IEEE Photonics Technol Lett* 2019;31(24):1952–5.
- [54] Luo Y, Mengü D, Yardimci NT, Rivenson Y, Veli M, Jarrahi M, et al. Design of task-specific optical systems using broadband diffractive neural networks. *Light Sci Appl* 2019;8(112):1–14.
- [55] Li J, Mengü D, Luo Y, Rivenson Y, Ozcan A. Class-specific differential detection in diffractive optical neural networks improves inference accuracy. *Adv Photonics* 2019;1(4):46001.
- [56] Zhou T, Fang L, Yan T, Wu J, Li Y, Fan J, et al. *In situ* optical backpropagation training of diffractive optical neural networks. *Photonics Res* 2020;8(6):940–53.
- [57] Rahman MSS, Li J, Mengü D, Rivenson Y, Ozcan A. Ensemble learning of diffractive optical networks. *Light Sci Appl* 2021;10(14):1–13.
- [58] Mengü D, Luo Y, Rivenson Y, Ozcan A. Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J Sel Top Quantum Electron* 2019;26(1):1–14.
- [59] Dou H, Deng Y, Yan T, Wu H, Lin X, Dai Q. Residual D²NN: training diffractive deep neural networks via learnable light shortcuts. *Opt Lett* 2020;45(10):2688–91.
- [60] Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J, et al. Fourier-space diffractive deep neural network. *Phys Rev Lett* 2019;123(2):023901.
- [61] Mengü D, Rivenson Y, Ozcan A. Scale-, shift-, and rotation-invariant diffractive optical networks. *ACS Photonics* 2021;8(1):324–34.
- [62] Idehenre IU, Mills MS. Multi-directional beam steering using diffractive neural networks. *Opt Express* 2020;28(18):25915–34.
- [63] Gao Y, Jiao S, Fang J, Lei T, Xie Z, Yuan X. Multiple-image encryption and hiding with an optical diffractive neural network. *Opt Commun* 2020; 463:125476.
- [64] Qian C, Lin X, Lin X, Xu J, Sun Y, Li E, et al. Performing optical logic operations by a diffractive neural network. *Light Sci Appl* 2020;9(59):1–7.
- [65] Veli M, Mengü D, Yardimci NT, Luo Y, Li J, Rivenson Y, et al. Terahertz pulse shaping using diffractive surfaces. *Nat Commun* 2021;12(37):1–13.
- [66] Zhao Q, Hao S, Wang Y, Wang L, Xu C. Orbital angular momentum detection based on diffractive deep neural network. *Opt Commun* 2019;443:245–9.
- [67] Fontaine NK, Ryf R, Chen H, Neilson DT, Kim K, Carpenter J. Laguerre–Gaussian mode sorter. *Nat Commun* 1865;2019(10):1–7.
- [68] Huang Z, Wang P, Liu J, Xiong W, He Y, Xiao J, et al. All-optical signal processing of vortex beams with diffractive deep neural networks. *Phys Rev Appl* 2021;15(1):014037.
- [69] Zuo Y, Li B, Zhao Y, Jiang Y, Chen YC, Chen P, et al. All-optical neural network with nonlinear activation functions. *Optica* 2019;6(9):1132–7.
- [70] Jiao S, Feng J, Gao Y, Lei T, Xie Z, Yuan X. Optical machine learning with incoherent light and a single-pixel detector. *Opt Lett* 2019;44(21):5186–9.
- [71] Limbacher B, Schönhuber S, Wenclawiak M, Kainz MA, Andrews AM, Strasser G, et al. Terahertz optical machine learning for object recognition. *APL Photonics* 2020;5(12):126103.
- [72] Zuo Y, Zhao Y, Chen YC, Du S, Liu J. Scalability of all-optical neural networks based on spatial light modulators. 2021. arXiv:2102.09722.
- [73] Tamir DE, Shaked NT, Wilson PJ, Dolev S. High-speed and low-power electrooptical DSP coprocessor. *J Soc Am A* 2009;26(8):A11–20.
- [74] Caulfield HJ, Dolev S. Why future supercomputing requires optics. *Nat Photonics* 2010;4(5):261–3.
- [75] Tait AN, De Lima TF, Zhou E, Wu AX, Nahmias MA, Shastri BJ, et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci Rep* 2017;7(7430):1–10.
- [76] Tait AN, De Lima TF, Nahmias MA, Shastri BJ, Prucnal PR. Multi-channel control for microring weight banks. *Opt Express* 2016;24(8):8895–906.
- [77] Tait AN, Nahmias MA, Shastri BJ, Prucnal PR. Broadcast and weight: an integrated network for scalable photonic spike processing. *J Light Technol* 2014;32(21):4029–41.
- [78] Tait AN, Jayatilaka H, De Lima TF, Ma PY, Nahmias MA, Shastri BJ, et al. Feedback control for microring weight banks. *Opt Express* 2018;26(20):26422–43.
- [79] Mehrabian A, Al-Kabani Y, Sorger VJ, El-Ghazawi T. PCNNA: a photonic convolutional neural network accelerator. In: *IEEE International System-on-Chip Conference*; 2018 Sept 4–7; Arlington, VA, USA. New York: IEEE; 2018. p. 169–73.
- [80] Huang C, Bilodeau S, De Lima TF, Tait AN, Ma PY, Blow EC, et al. Demonstration of scalable microring weight bank control for large-scale photonic integrated circuits. *APL Photonics* 2020;5(4):040803.
- [81] Ma PY, Tait AN, De Lima TF, Abbaslou S, Shastri BJ, Prucnal PR. Photonic principal component analysis using an on-chip microring weight bank. *Opt Express* 2019;27(13):18329–42.
- [82] Xu X, Tan M, Corcoran B, Wu J, Boes A, Nguyen TG, et al. 11 Tops photonic convolutional accelerator for optical neural networks. *Nature* 2021;589:44–51.
- [83] Khoram E, Chen A, Liu D, Ying L, Wang Q, Yuan M, et al. Nanophotonic media for artificial neural inference. *Photonics Res* 2019;7(8):823–7.
- [84] Wu Z, Zhou M, Khoram E, Liu B, Yu Z. Neuromorphic metasurface. *Photonics Res* 2020;8(1):46–50.
- [85] Hamerly R, Bernstein L, Sludds A, Soljačić M, Englund D. Large-scale optical neural networks based on photoelectric multiplication. *Phys Rev X* 2019;9(2):21032.
- [86] Cohen E, Malka D, Shemer A, Shahmoon A, Zalevsky Z, London M. Neural networks within multi-core optic fibers. *Sci Rep* 2016;6(29080):1–14.
- [87] Shabairou N, Cohen E, Wagner O, Malka D, Zalevsky Z. Color image identification and reconstruction using artificial neural networks on multimode fiber images: towards an all-optical design. *Opt Lett* 2018;43(22):5603–6.
- [88] Chang J, Sitzmann V, Dun X, Heidrich W, Wetzstein G. Hybrid optoelectronic convolutional neural networks with optimized diffractive optics for image classification. *Sci Rep* 2018;8(12324):1–10.
- [89] Pad P, Narduzzi S, Kundig C, Turetken E, Bigdeli SA, Dunbar LA. Efficient neural vision systems based on convolutional image acquisition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2020 Jun 13–19; Seattle, WA, USA. New York: IEEE; 2020. p. 12285–94.
- [90] Colburn S, Chu Y, Shilzerman E, Majumdar A. Optical frontend for a convolutional neural network. *Appl Opt* 2019;58(12):3179–86.
- [91] Ahmed M, Al-Hadeethi Y, Bakry A, Dalir H, Sorger VJ. Integrated photonic FFT for optical convolutions towards efficient and high-speed neural networks. 2020. arXiv:2002.01308.
- [92] Xu S, Wang J, Wang R, Chen J, Zou W. High-accuracy optical convolution unit architecture for convolutional neural networks by cascaded acousto-optical modulator arrays. *Opt Express* 2020;28(15):21854.
- [93] 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(13):3689–92.
- [94] Miscuglio M, Mehrabian A, Hu Z, Azzam SI, George JK, Kildishev AV, et al. All-optical nonlinear activation function for photonic neural networks. *Opt Mater Express* 2018;8(12):3851–63.
- [95] George JK, Mehrabian A, Amin R, Meng J, De Lima TF, Tait AN, et al. Neuromorphic photonics with electro-absorption modulators. *Opt Express* 2019;

- 27(4):5181–91.
- [96] Fard MMP, Williamson IAD, Edwards M, Liu K, Pai S, Bartlett B, et al. Experimental realization of arbitrary activation functions for optical neural networks. *Opt Express* 2020;28(8):12138–48.
- [97] Brunner D, Fischer I. Reconfigurable semiconductor laser networks based on diffractive coupling. *Opt Lett* 2015;40(16):3854–7.
- [98] Maktoobi S, Froehly L, Andreoli L, Porte X, Jacquot M, Larger L, et al. Diffractive coupling for photonic networks: how big can we go? *IEEE J Sel Top Quant* 2020;26(1):1–8.
- [99] Appeltant L, Soriano MC, Van der Sande G, Danckaert J, Massar S, Dambre J, et al. Information processing using a single dynamical node as complex system. *Nat Commun* 2011;2(468):1–6.
- [100] Heuser T, Pflüger M, Fischer I, Lott JA, Brunner D, Reitzenstein S. Developing a photonic hardware platform for brain-inspired computing based on 5×5 VCSEL arrays. *J Phys Photonics* 2020;2(4):044002.
- [101] Paquot Y, Dambre J, Schrauwen B, Haelterman M, Massar S. Reservoir computing: a photonic neural network for information processing. In: *Proceedings of Nonlinear Optics and Applications IV*; 2010 Apr 12–16; Brussels, Belgium. SPIE; 2010. p. 77280.
- [102] Brunner D, Soriano MC, Mirasso CR, Fischer I. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat Commun* 2013;4(1364):1–7.
- [103] Duport F, Schneider B, Smerieri A, Haelterman M, Massar S. All-optical reservoir computing. *Opt Express* 2012;20(20):22783–95.
- [104] Vinckier Q, Duport F, Smerieri A, Vandoorne K, Bienstman P, Haelterman M, et al. High-performance photonic reservoir computer based on a coherently driven passive cavity. *Optica* 2015;2(5):438–46.
- [105] Hicke K, Escalona-Morán MA, Brunner D, Soriano MC, Fischer I, Mirasso CR. Information processing using transient dynamics of semiconductor lasers subject to delayed feedback. *IEEE J Sel Top Quant* 2013;19(4):1501610.
- [106] Nakayama J, Kanno K, Uchida A. Laser dynamical reservoir computing with consistency: an approach of a chaos mask signal. *Opt Express* 2016;24(8):8679–92.
- [107] Nguimdo RM, Verschaffelt G, Danckaert J, Van der Sande G. Fast photonic information processing using semiconductor lasers with delayed optical feedback: role of phase dynamics. *Opt Express* 2014;22(7):8672–86.
- [108] Nguimdo RM, Verschaffelt G, Danckaert J, Van der Sande G. Simultaneous computation of two independent tasks using reservoir computing based on a single photonic nonlinear node with optical feedback. *IEEE Trans Neural Netw Learn Syst* 2015;26(12):3301–7.
- [109] Nguimdo RM, Verschaffelt G, Danckaert J, Van der Sande G. Reducing the phase sensitivity of laser-based optical reservoir computing systems. *Opt Express* 2016;24(2):1238–52.
- [110] Dejonckheere A, Duport F, Smerieri A, Fang L, Oudar JL, Haelterman M, et al. All-optical reservoir computer based on saturation of absorption. *Opt Express* 2014;22(9):10868–81.
- [111] Larger L, Soriano MC, Brunner D, Appeltant L, Gutierrez M, Pesquera L, et al. Photonic information processing beyond turning: an optoelectronic implementation of reservoir computing. *Opt Express* 2012;20(3):3241–9.
- [112] Paquot Y, Duport F, Smerieri A, Dambre J, Schrauwen B, Haelterman M, et al. Optoelectronic reservoir computing. *Sci Rep* 2012;2(287):1–6.
- [113] Martinenghi R, Rybalko S, Jacquot M, Chembo YK, Larger L. Photonic nonlinear transient computing with multiple-delay wavelength dynamics. *Phys Rev Lett* 2012;108(24):244101.
- [114] Soriano MC, Ortín S, Brunner D, Larger L, Mirasso CR, Fischer I, et al. Optoelectronic reservoir computing: tackling noise-induced performance degradation. *Opt Express* 2013;21(1):12–20.
- [115] Ortín S, Soriano MC, Pesquera L, Brunner D, San-Marín D, Fischer CR, et al. A unified framework for reservoir computing and extreme learning machines based on a single time-delayed neuron. *Sci Rep* 2015;5(14945):1–11.
- [116] Duport F, Smerieri A, Akrouf A, Haelterman M, Massar S. Fully analogue photonic reservoir computer. *Sci Rep* 2016;6(22381):1–12.
- [117] Duport F, Smerieri A, Akrouf A, Haelterman M, Massar S. Virtualization of a photonic reservoir computer. *J Lightwave Technol* 2016;34(9):2085–91.
- [118] Lavrov R, Jacquot M, Larger L. Nonlocal nonlinear electrooptic phase dynamics demonstrating 10 Gb/s chaos communications. *IEEE J Quantum Elect* 2010;46(10):1430–5.
- [119] Vandoorne K, Dierckx W, Schrauwen B, Verstraeten D, Baets R, Bienstman P, et al. Toward optical signal processing using photonic reservoir computing. *Opt Express* 2008;16(15):11182–92.
- [120] Vandoorne K, Dambre J, Verstraeten D, Schrauwen B, Bienstman P. Parallel reservoir computing using optical amplifiers. *IEEE Trans Neural Netw* 2011;22(9):1469–81.
- [121] Salehi MR, Dehyadegari L. Optical signal processing using photonic reservoir computing. *J Mod Opt* 2014;61(17):1442–51.
- [122] Vandoorne K, Mechet P, Van Vaerenbergh T, Fiers M, Morthier G, Verstraeten D, et al. Experimental demonstration of reservoir computing on a silicon photonics chip. *Nat Commun* 2014;5(3541):1–6.
- [123] Ikeda K, Daido H, Akimoto O. Optical turbulence: chaotic behavior of transmitted light from a ring cavity. *Phys Rev Lett* 1980;45(9):709–12.
- [124] Goedgebuer JP, Larger L, Porte H, Delorme F. Chaos in wavelength with a feedback tunable laser diode. *Phys Rev E* 1998;57(3):2795–8.
- [125] Antonik P, Dupont F, Hermans M, Smerieri A, Haelterman M, Massar S. Online training of an opto-electronic reservoir computer applied to real-time channel equalization. *IEEE Trans Neural Netw Learn Syst* 2016;28(11):2686–98.
- [126] Antonik P, Haelterman M, Massar S. Brain-inspired photonic signal processor for generating periodic patterns and emulating chaotic systems. *Phys Rev Appl* 2017;7(5):054014.
- [127] Larger L, Baylón-Fuentes A, Martinenghi R, Udaltsov VS, Chembo YK, Jacquot M. High-speed photonic reservoir computing using a time-delay-based architecture: million words per second classification. *Phys Rev X* 2017;7(1):011015.
- [128] Hou Y, Xia G, Yang W, Wang D, Jayaprasath E, Jiang Z, et al. Prediction performance of reservoir computing system based on a semiconductor laser subject to double optical feedback and optical injection. *Opt Express* 2018; 26(8):10211–9.
- [129] Barbay S, Kuszelewicz R, Yacomotti AM. Excitability in a semiconductor laser with saturable absorber. *Opt Lett* 2011;36(23):4476–8.
- [130] Hurtado A, Schires K, Henning ID, Adams MJ. Investigation of vertical cavity surface emitting laser dynamics for neuromorphic photonic systems. *Appl Phys Lett* 2012;100(10):103703.
- [131] Nahmias MA, Tait AN, Shastri BJ, De Lima TF, Prucnal PR. Excitable laser processing network node in hybrid silicon: analysis and simulation. *Opt Express* 2015;23(20):26800–13.
- [132] Shastri BJ, Nahmias MA, Tait AN, Rodriguez AW, Wu B, Prucnal PR. Spike processing with a graphene excitable laser. *Sci Rep* 2016;6(19126):1–12.
- [133] Coomans W, Geleens L, Beri S, Danckaert J, Van der Sande G. Solitary and coupled semiconductor ring lasers as optical spiking neurons. *Phys Rev E Stat Nonlin Soft Matter Phys* 2011;84:036209.
- [134] Boahen K. A neuromorph's prospectus. *Comput Sci Eng* 2017;19(2):14–28.
- [135] Vaerenbergh TV, Fiers M, Mechet P, Spuesens T, Kumar R, Morthier G, et al. Cascadable excitability in microrings. *Opt Express* 2012;20(18):20292–308.
- [136] Selmi F, Braive R, Beaudoin G, Sagnes I, Kuszelewicz R, Barbay S. Relative refractory period in an excitable semiconductor laser. *Phys Rev Lett* 2014; 112(18):183902.
- [137] Bueno J, Maktoobi S, Froehly L, Fischer I, Jacquot M, Larger L, et al. Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* 2018;5(6):756–60.
- [138] Davidson NS, Richards RW, Geissler E. Scaling laws and polystyrene networks: a quasi-elastic light scattering study. *Polymer* 1985;26(11):1643–6.
- [139] Dong J, Rafayelyan M, Krzakala F, Gigan S. Optical reservoir computing using multiple light scattering for chaotic systems prediction. *IEEE J Sel Top Quant* 2020;26(1):1–12.
- [140] Rafayelyan M, Dong J, Tan Y, Krzakala F, Gigan S. Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction. 2020. arXiv: 2001.09131.
- [141] Bao X. Research progress in optoelectronic reservoir computing system. *Laser Optoelectron Prog* 2015;3:5.
- [142] Zhao Q, Yin H. Research progress of reservoir computing using chaotic laser. *Laser Optoelectron Prog* 2013;50(3):030003.
- [143] Zhang H, Feng X, Li B, Wang Y, Cui K, Liu F, et al. Integrated photonic reservoir computing based on hierarchical time-multiplexing structure. *Opt Express* 2014;22(25):31356–70.
- [144] Maass W. Networks of spiking neurons: the third generation of neural network models. *Neural Networks* 1997;10(9):1659–71.
- [145] Peng HT, Nahmias MA, De Lima TF, Tait AN, Shastri BJ. Neuromorphic photonic integrated circuits. *IEEE J Sel Top Quantum Electron* 2018;24(6): 1–15.
- [146] Bohte SM, Kok JN, La Poutre H. Error-backpropagation in temporally encoded networks of spiking neurons. *Neurocomputing* 2002;48(1–4):17–37.
- [147] Gütig R, Sompolinsky H. The tempotron: a neuron that learns spike timing-based decisions. *Nat Neurosci* 2006;9(3):420–8.
- [148] Song S, Miller KD, Abbott LF. Competitive Hebbian learning through spiketiming-dependent synaptic plasticity. *Nat Neurosci* 2000;3(9):919–26.

- [149] Ponulak F, Kasiński A. Supervised learning in spiking neural networks with ReSuMe: sequence learning, classification, and spike shifting. *Neural Comput* 2010;22(2):467–510.
- [150] Mohammed A, Schliebs S, Matsuda S, Kasabov N. SPAN: spike pattern association neuron for learning spatio-temporal spike patterns. *Int J Neural Syst* 2012;22(4):1250012.
- [151] Yu Q, Tang H, Tan KC, Li H. Precise-spike-driven synaptic plasticity: learning hetero-association of spatiotemporal spike patterns. *PLoS ONE* 2013; 8(11): e78318.
- [152] Xiang S, Ren Z, Song Z, Zhang Y, Guo X, Han G, et al. Computing primitive of fully-VCSELS-based all-optical spiking neural network for supervised learning and pattern classification. *IEEE Trans Neural Netw Learn Syst* 2020;32(6):2494–505.
- [153] Xiang S, Zhang Y, Gong J, Guo X, Lin L, Hao Y. STDP-based unsupervised spike pattern learning in a photonic spiking neural network with VCSELS and VCSOAs. *IEEE J Sel Top Quant* 2019;25(6):1700109.
- [154] Merolla PA, Arthur JV, Alvarez-Icaza R, Cassidy AS, Sawada J, Akopyan F, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* 2014;345(6197):668–73.
- [155] Benjamin BV, Gao P, McQuinn E, Choudhary S, Chandrasekaran AR, Bussat JM, et al. Neurogrid: a mixed-analog-digital multichip system for large-scale neural simulations. *Proc IEEE* 2014;102(5):699–716.
- [156] Furber SB, Galluppi F, Temple S, Plana LA. The spinnaker project. *Proc IEEE* 2014;102(5):652–65.
- [157] Prucnal PR, Shastri BJ, Teich MC. *Neuromorphic photonics*. Boca Raton: CRC Press; 2017.
- [158] De Marinis L, Cococcioni M, Castoldi P, Andriolli N. Photonic neural networks: a survey. *IEEE Access* 2019;7:175827–41.
- [159] Shastri BJ, Tait AN, De Lima TF, Pernice WHP, Bhaskaran H, Wright CD, et al. Photonics for artificial intelligence and neuromorphic computing. *Nat Photonics* 2021;15:102–14.