

## 激光与光电子学进展

## 光计算的物理架构与应用(特邀)

郭圣炳<sup>1,2</sup>, 刘文哲<sup>1,2,3</sup>, 王佳俊<sup>1,2\*\*\*</sup>, 郑敏嘉<sup>1,2\*\*</sup>, 石磊<sup>1,2,3\*</sup><sup>1</sup>上海量子科学研究中心, 上海 201315;<sup>2</sup>复旦大学物理学系, 应用表面物理全国重点实验室, 微纳光子结构教育部重点实验室, 上海 200433;<sup>3</sup>复旦大学微纳电子器件与量子计算机研究院, 上海 200438

**摘要** 光计算是基于操控光的新型计算架构,通过光进行信息加载、传输与运算。近年来光计算利用了光的偏振、频率、轨道角动量等独立的信息维度,涌现出很多新型架构。同时结合深度学习驱动的结构设计,在矩阵运算、图像处理等任务上表现出卓越的性能。从光计算的物理原理出发,对自由空间衍射光计算、片上集成光计算的主要架构进行总结和讨论,指出了光计算当下发展所面临的紧迫挑战,并对未来的发展趋势进行了展望。

**关键词** 光计算; 光学神经网络; 集成光学; 光子芯片; 光电智能计算

中图分类号 O436

文献标志码 A

DOI: 10.3788/LOP251285

## Physical Architecture and Application for Optical Computing (Invited)

Guo Shengbing<sup>1,2</sup>, Liu Wenzhe<sup>1,2,3</sup>, Wang Jiajun<sup>1,2\*\*\*</sup>, Zheng Minjia<sup>1,2\*\*</sup>, Shi Lei<sup>1,2,3\*</sup><sup>1</sup>Shanghai Research Center for Quantum Sciences, Shanghai 201315, China;<sup>2</sup>State Key Laboratory of Surface Physics, Key Laboratory of Micro- and Nano-Photonic Structures (Ministry of Education), Department of Physics, Fudan University, Shanghai 200433, China;<sup>3</sup>Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200438, China

**Abstract** Optical computing is a novel computational architecture based on the manipulation of photons or optical fields rather than electrons, utilizing light for information encoding, transmitting, and processing. In recent years, optical computing has leveraged independent information dimensions of photons, such as polarization, frequency, and orbital angular momentum, leading to the emergence of numerous novel architectures. Concurrently, the integration of deep-learning-driven structural design has enabled these architectures to demonstrate exceptional performance in tasks including matrix operation and image processing. This paper begins with the physical foundations of optical computing, systematically summarizes and discusses the principal architectures of free-space diffractive and on-chip integrated optical computing. Furthermore, this paper concludes by highlighting the pressing challenges confronting current developments in optical computing and provides perspectives on future trends in this field.

**Key words** optical computing; optical neural network; integrated optics; photonic chip; optoelectronic intelligent computing

## 1 引言

当前阶段,以大语言模型为代表的各种机器学习系统性能进展飞速,计算机视觉<sup>[1]</sup>、大语言模型<sup>[2]</sup>、扩散模型<sup>[3]</sup>等人工智能(AI)工具进入日常生活,在政务、医疗、科研和教育中都发挥着举足轻重的作用<sup>[4]</sup>。随着对AI性能要求的逐渐提高,计算所需能量也不断提升,单位能量和单位时间内执行的计算次数对计算架

构提出了更高的要求。

在2010年之前,以浮点操作数(FLOP)计算,AI系统训练量每21.6个月翻一倍,这与摩尔定律所描绘的每18~24个月处理器性能翻一倍基本相适应。进入深度学习时代后,最尖端大模型计算量每6个月就会翻一倍<sup>[5]</sup>,如图1所示。然而近10年来,集成电路领域的算力发展速度远无法与AI模型的发展速度相比,计算机中央处理器(CPU)和图形处理器(GPU)等器

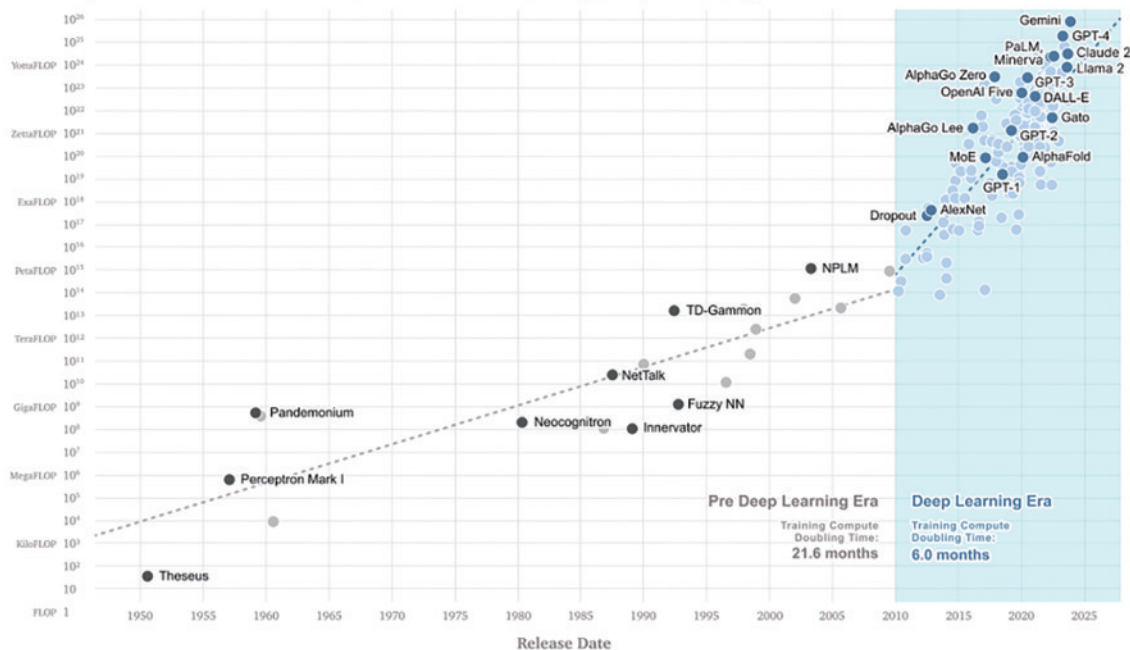
收稿日期: 2025-05-21; 修回日期: 2025-06-23; 录用日期: 2025-06-30; 网络首发日期: 2025-07-02

基金项目: 上海市市级科技重大专项(2019SHZDZX01)、国家重点研发计划(2023YFA1406900, 2022YFA1404800)

通信作者: \*lshi@fudan.edu.cn; \*\*mjzheng@fudan.edu.cn; \*\*\*jiajunwang@fudan.edu.cn

## Compute Used for AI Training Runs

Total compute used to train notable AI models, measured in total FLOP (floating-point operations) | Logarithmic

图1 AI大模型训练所需计算量随时间的演化<sup>[5]</sup>Fig. 1 Evolution of computation required for training large models of AI over time<sup>[5]</sup>

件开发工艺进入数纳米尺度后<sup>[6-7]</sup>,减小晶体管尺寸从而提升单位面积内计算单元数量的方案正面临进一步发展的困境,包括量子隧穿与散热等问题。由此引发的能效比和制程研发、制造成本之间的矛盾日趋严重<sup>[8]</sup>。当下亟须研发兼顾高性能、高能效比与低加工成本的新型计算硬件,发展存算一体的计算架构。在计算架构的革新中,光计算凭借诸多特性脱颖而出:1)不同于电子等费米子,光子作为玻色子不受泡利不兼容原理限制,光子在传播中互不干扰,天然具有高并行性;2)高并行性,包括偏振、频率、强度、轨道角动量等独立信息维度<sup>[9]</sup>;3)光速计算,光场中编码的信息在光速传播的同时完成处理与计算;4)光子传播过程中损耗低,使得光学系统有较高能效比<sup>[10]</sup>。

近些年来,已经有各种光计算架构展现出高算力和高能效比。一类是基于电子计算机的架构,发挥光学优势,完成光学晶体管设计、基于光互连的光电融合计算等数字光计算任务<sup>[9-11]</sup>;另一类则从光学角度出发,基于光子的多维度信息加载能力完成特定的计算任务,如矩阵乘法-累加(MAC)运算<sup>[11]</sup>、全光逻辑运算<sup>[12]</sup>、光学卷积<sup>[13-15]</sup>、全光微分<sup>[16]</sup>等,以模拟光学计算为主。其中的数字光计算体系仍然需要高精度的电子体系来提供控制和输入输出,才能发挥其优势<sup>[17-19]</sup>;模拟光计算则可以利用光的多维度信息<sup>[20]</sup>,不仅实现单一功能的光信息处理,在将来还可能实现高误差容忍度的模糊计算,如光学神经网络(PNN)等计算架构,为AI发展提供计算平台。

模拟光计算发展迅猛,现有探索方向大多聚焦于

光电融合的计算架构。由于存算体系尚不完备、光学非线性实现方式有限等,光计算尚无法完全替代电子计算体系。现阶段的最佳方案是优化AI训练算力需求最大的集成电路模块,从而提高大规模矩阵乘法、卷积的计算速度。在这类任务上,光计算架构避免了串行计算过程中的低速率和高计算开销,可以同时计算所有矩阵元,这种优势随着矩阵运算规模的扩展而越来越突出。因此光电融合计算的优化必然是很长一段时间内的研究重点<sup>[9]</sup>,亟须发展基于此的架构以满足AI大模型,甚至是通用AI(AGI)的算力需求。

本文主要聚焦于模拟光计算,依据其物理架构进行总结分析。首先,讨论了光计算的物理原理,包括自由空间的计算理论、集成光学计算及非线性光学计算的基本原理。其次,对光计算架构的物理实现进行分析和讨论。再次,从信息处理角度,总结各类光计算任务的物理基础。最后,对未来光计算物理架构所面临的挑战和发展趋势进行了展望。

## 2 光计算的物理原理

目前模拟光计算的实现架构主要有两类:一类对在自由空间传播、衍射和干涉的光场施加调制,使其执行特定的计算任务;另一类则是基于片上光学工艺发展的集成光计算架构。下面将对两类架构的计算原理、基本单元和设计方式进行总结。

### 2.1 自由空间衍射式光计算的物理架构

自由空间衍射可以充分利用光的多信息通道优势,实现高复用的信息并行处理,其运算能力和计算规

模可以扩展至与电子计算机相当的水平<sup>[21]</sup>。从调制类型来看,可依据调制的光参数类型进行划分。对振幅进行调制的器件有数字微镜器件(DMD)、振幅型的空间光调制器(SLM)和3D打印或刻蚀制备的掩模板。针对相位进行调制的有相位型SLM、3D打印或刻蚀制备的相位板。此外还有对振幅、相位、频率和偏振等同时调制的超构材料。对于特定的调制类型,要获得实现特定功能的调制参数空间分布,往往采用深度学习驱动算法,搭建“电学训练,光学计算”<sup>[22]</sup>的架构。在讨论上述各种体系的具体物理实现前,将先介绍其基本的物理原理。

### 2.1.1 自由空间衍射中的物理原理

要利用光束在自由空间中的干涉和衍射进行信息处理,需要有高效的模拟计算方式。通常,将自由空间中的麦克斯韦方程组转化为三维波动方程进行求解,在不涉及求解衍射平面附近的光场等问题时,可以忽略电矢量和磁矢量的耦合关系,将电场视为标量场从而描述光的衍射和传播<sup>[23-24]</sup>,以此为前向传播的模拟基础。下面介绍两种基本的衍射计算理论,在其各自的基础上均发展了各种光计算架构。

第一类方法以光学系统的点扩散函数(PSF)为卷积核,将光场传播视为自然卷积过程,对某一空间处的光场进行复数卷积来得到光场在另一空间的分布。其基本公式是瑞利-索末菲衍射积分,将空间中一点的复数场与另一点的复数场联系起来。空间中任意点 $P$ 的复振幅可以由包围 $P$ 的封闭曲面 $\Sigma$ 上各点的复数场 $U$ 及其沿外法线方向的偏导数通过积分式得到,即如下格林定理:

$$U(P_0) = \frac{1}{i\lambda} \iint_{\Sigma} U_0 \frac{\exp(i\vec{k} \cdot \vec{r})}{r} \left( 1 + \frac{i}{kr} \right) \cos(\vec{n}, \vec{r}) dS \quad (1)$$

式中: $P_0$ 为观测点的坐标; $\lambda$ 为波长; $U_0$ 为包围观测点封闭表面上的复数场; $\vec{k}$ 为波矢; $\vec{r}$ 为观测点到面积微元 $S$ 的位矢; $k$ 与 $r$ 为波矢和位矢的大小; $\vec{n}$ 为面积微元的方向矢量。此处复数场视为电场,结合惠更斯-菲涅耳原理,得到表达式:

$$U(P_0) = \iint_{\Sigma} h(P_0, P_1) U(P_1) dS \quad (2)$$

式中: $P_1$ 为封闭表面上点的坐标;PSF  $h(P_0, P_1)$ 的一般情况是

$$h(P_0, P_1) = \frac{1}{2\pi} \left( \frac{1}{r_{01}} - ik \right) \exp(i\vec{k} \cdot \vec{r}_{01}) \cos \theta \quad (3)$$

式中: $\vec{r}_{01}$ 为观测点到面积微元的位矢; $r_{01}$ 为位矢的大小; $\theta$ 为 $\vec{n}$ 和 $\vec{r}_{01}$ 两矢量的夹角。假设 $r_{01}$ 比波长大很多,

可自然简化为 $h(P_0, P_1) = \frac{1}{i\lambda} \frac{\exp(i\vec{k} \cdot \vec{r}_{01})}{r_{01}} \cos \theta$ ,自由空间光场传播的线性性质使得光场的衍射过程总可以视为一个卷积积分,这一结果也可以推广到非单色波,从而利用波长这一信息维度。光学系统对输入信息的调

制视为输入光场与PSF的卷积,由此可以对PSF施加调制,从而实现特定矢量PSF构建<sup>[25]</sup>,如图2(a)所示,并借助PSF实现高精度显微成像<sup>[26-27]</sup>等功能。

第二类方法将光场视为一系列平面波的叠加,不同平面波所对应的复数权重就被称为此光场的角谱<sup>[23]</sup>。光的传播与衍射则是对不同传播方向平面波的截止,即沿任意方向的波矢不能超过波长对应的波矢。又由于衍射前后的平面总是有限的,对应有限的出射/入射角度,其余角度的平面波被自由空间衍射所消除。

假设衍射距离为 $z$ ,设 $U(x, y, 0)$ 和 $U(x, y, z)$ 分别为衍射屏和观察屏上的复数场,经过傅里叶变换之后得到以角频率表示的频谱也即角谱,分别为 $A_0(f_x, f_y)$ 和 $A_z(f_x, f_y)$ 。这两个角谱通过如下公式相联系:

$$A_z(f_x, f_y) = A_0(f_x, f_y) \exp \left[ i \frac{2\pi}{\lambda} z \sqrt{1 - (\lambda f_x)^2 - (\lambda f_y)^2} \right] \quad (4)$$

这表明,自由空间的传播在频域表现为入射屏处的角谱乘以一个与传播方向有关的相位延迟因子。其中频率较大的部分表现为倏逝波而衰减,只有满足自由空间传播条件的波才能到达观察屏处,也就是说自由空间衍射等价于一个半径为 $1/\lambda$ 的低通滤波器。这种计算方式具有普适性,对于实际中采样密度不同的平面,可以使用图2(b)所示的可缩放的角谱法进行计算<sup>[28]</sup>。

由此,基于SLM或3D打印调制材料的模拟光计算方案可以在实空间施加调制进行计算<sup>[29-33]</sup>,也可以在动量空间对角谱施加调制进行处理<sup>[34-36]</sup>,也就是将深度学习算法所获得的调制参数的空间分布和物理实际相对应。

对于基于超构表面的自由空间计算架构,同样需要对相位、振幅等信息维度进行调控,其实现方法是通过设计单个单元结构的几何参数以及单元结构的排列方式来得到特定的调制<sup>[37-40]</sup>。超构表面所执行的任务虽然不可重构,但其高度集成的调控方式使得自由空间的光计算任务可以在微纳尺度执行。

对于相位调制,主要有以下3种类型,实际中往往结合多种类型的相位调制进行设计。1)共振相位。基于单元结构的几何参数变化使得共振频率移动,从而改变特定频率下的相位<sup>[41-44]</sup>,图2(c)展示了利用不同共振频率V形金属结构验证广义斯涅耳定律<sup>[45]</sup>。2)几何相位,也称为PB(Pancharatnam-Berry)相位。其原理在于,不同旋转取向单元结构的交叉极化光场等价于偏振态在庞加莱球上不同的闭合曲面,产生的额外相位调制等于闭合曲面对应立体角的一半<sup>[46-47]</sup>。由于几何相位仅取决于单元结构取向,可以降低微纳工艺公差对器件工作效率的影响,在彩色全息<sup>[48]</sup>、消色差透镜<sup>[49]</sup>等元件的设计上有广泛应用,如图2(d)所示。3)传播相位。基于等效折射率理论,单元结构往往是具有

高宽深比的柱或孔,通过调整其占空比来改变其等效折射率,从而提供特定的相位调制<sup>[50-51]</sup>。

对于振幅调制,则以马吕斯超构表面为主,将单元结构设计为特定的偏振调制方式,依据马吕斯定律对

光场的透反射系数进行调制。如图 2(e)所示,马吕斯超构表面可以在亚波长分辨率下实现 8 bit 位深的振幅调制<sup>[52-53]</sup>,该偏振调制振幅的原理同样适用于扭曲向列(TN)模式的振幅型 SLM。

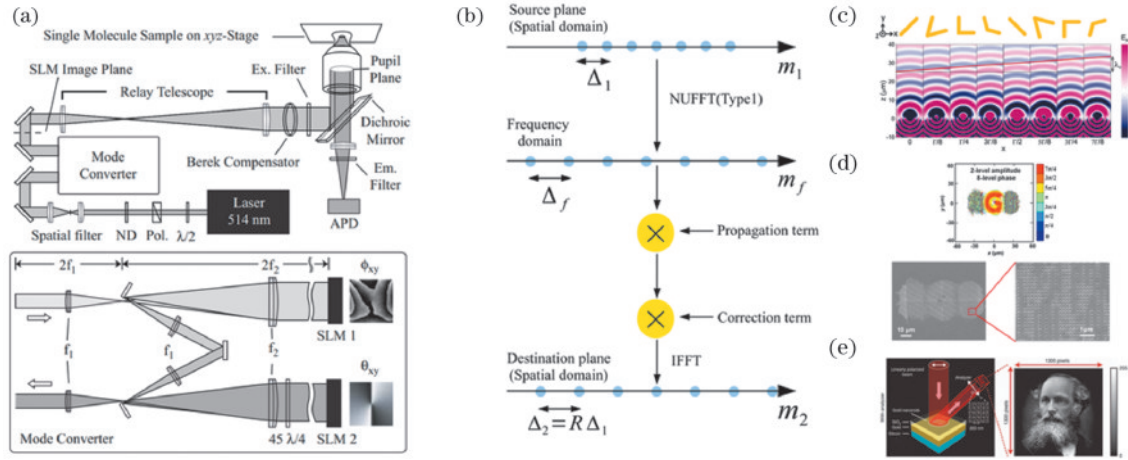


图 2 自由空间衍射式架构的计算原理。(a)基于卷积计算方法构建特定矢量 PSF<sup>[25]</sup>; (b)利用非均匀快速傅里叶变换(NUFFT)的角谱法计算<sup>[28]</sup>; (c)具有不同共振频率的 V 形金属结构验证广义斯涅耳定律<sup>[45]</sup>; (d)基于 8 阶几何相位的彩色全息图<sup>[48]</sup>; (e)反射式马吕斯超构表面构建的 8 bit 深全息图<sup>[53]</sup>

Fig. 2 Computational principles of free-space diffractive architectures. (a) Construction of specific vectorial PSF based on convolutional computation methods<sup>[25]</sup>; (b) angular spectrum calculation employing non-uniform fast Fourier transform (NUFFT) algorithm<sup>[28]</sup>; (c) V-shaped metallic structures with distinct resonant frequencies for validating the generalized Snell's law<sup>[45]</sup>; (d) full-color holograms based on 8th-order geometric phase<sup>[48]</sup>; (e) 8 bit depth hologram construction using reflective Malus meta-surface<sup>[53]</sup>

### 2.1.2 自由空间衍射架构的物理实现

对于自由空间光场中的模拟光计算任务,早期研究一般集中在 MAC 运算任务的光学实现上。早在 1978 年,斯坦福大学团队<sup>[54]</sup>就实现了基于 SLM 的 MAC 运算。2003 年,以色列公司 Lenslet<sup>[55]</sup>开发了基于此架构的商用全光数字信号处理器,计算速度达到  $10^{12} \text{ s}^{-1}$ ,但其执行的任务仅限于 MAC 操作,或特定的离散傅里叶变换等计算任务,并没有利用光学系统的其他信息维度。对当下的深度学习算法<sup>[56-57]</sup>而言,大量计算资源消耗在了矩阵与矩阵的 MAC 运算上。因此除了将电子计算机任务或人工神经网络(ANN)直接同构部署在光学系统中进行 MAC 运算,还可以直接设计基于自由空间传播的深度衍射神经网络(D<sup>2</sup>NN)<sup>[58-59]</sup>。只须依据任务目标设计调制方式,而无须明确其对应的数学形式。

2018 年, Lin 等<sup>[29]</sup>开创了 D<sup>2</sup>NN 作为深度学习的框架,如图 3(a)所示,结合深度学习的训练方法和光学衍射的前向传播算法,通过数层衍射层完成分类任务。受衍射层调制的前向传播模型可以等价为一组复数矩阵算子,作用在入射光场上产生像平面上的复杂光场,如分类任务中的物品标签信息。在光波段,可以使用光刻、微纳加工技术制备衍射层或使用 SLM、DMD 等可调器件,也可以在太赫兹等波段使用 3D 打印等技术制备。他们共同的特点是仅使用衍射和无源器件就可以以光速执行计算任务。对于更复杂的分类任务,

可以采用增加衍射层<sup>[60]</sup>、集成学习<sup>[61]</sup>以及优化菲涅耳数<sup>[32]</sup>等方法来执行,从而在高位深灰度图输入的计算任务<sup>[33]</sup>、CIFAR-10 数据集分类任务<sup>[61]</sup>中实现准确率的显著提升,如图 3(b)~3(c)所示。在利用光场的其他信息维度方面,D<sup>2</sup>NN 还可以复用波长信息维度,如利用宽带输入设计光谱编码的单像素机器视觉系统<sup>[62]</sup>。对于相干光场,输入的物面光场和输出的像面光场可以用衍射表面实现任意的复线性变换<sup>[63]</sup>,因此 D<sup>2</sup>NN 可以在光场的复线性变换中取得良好的效果,包括:1)二维离散傅里叶变换<sup>[63]</sup>;2)任意复数域的西矩阵、非西矩阵和不可逆线性矩阵算子;3)任意二维光场的置换操作<sup>[64]</sup>;4)高通滤波的相干成像。而非相干光场不存在这种映射关系,其振幅和相位随时间任意变换,从而只能将平均光强作为考察的物理量。2023 年, Rahman 等<sup>[65]</sup>利用非相干的单色光实现 D<sup>2</sup>NN。这类非相干光束的自聚焦、自分裂特性及其在复杂环境下的鲁棒性,为自然光条件下的全光信息处理提供了新路径<sup>[66]</sup>。针对非相干光的深入研究有利于在自然光条件下设计实用的全光处理器件。

以上设计均基于光场的衍射进行。从傅里叶光学的角度出发,调制空间频率/动量空间光场中的某一点等价于对物面光场的全局调制。2019 年, Yan 等<sup>[34]</sup>基于此开发了傅里叶 D<sup>2</sup>NN(F-D<sup>2</sup>NN),从理论上证明了这一体系可以处理显著性提取和边缘提取等机器视觉任务,图 3(d)展示了 F-D<sup>2</sup>NN 下的图像显著性提取结

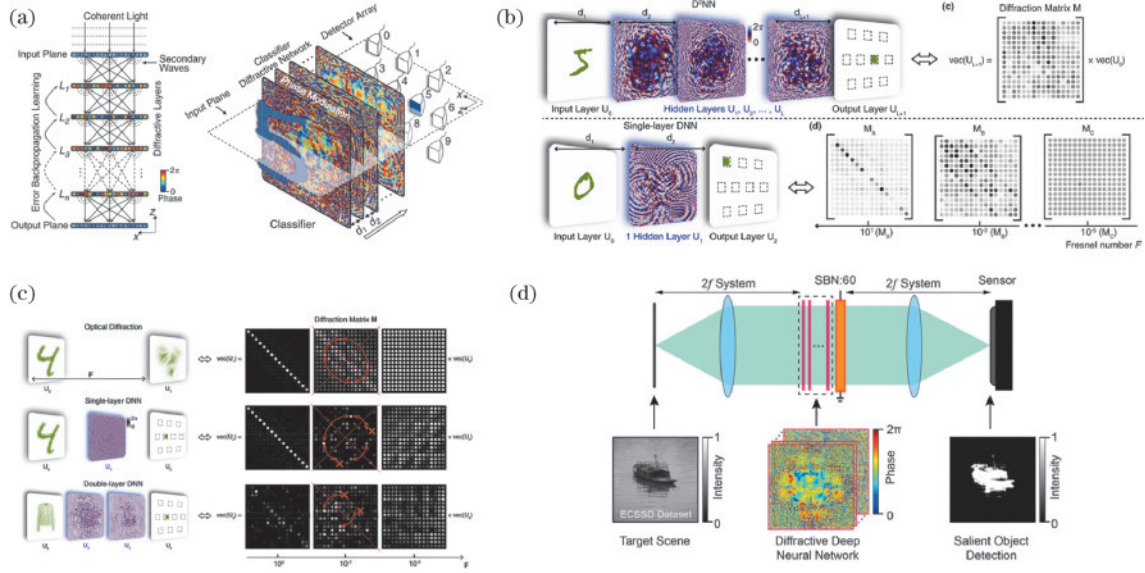


图3 D<sup>2</sup>NN的实现与优化。(a) D<sup>2</sup>NN执行手写数字识别等分类任务<sup>[29]</sup>; (b)通过控制菲涅耳数优化D<sup>2</sup>NN的表达力<sup>[32]</sup>; (c) D<sup>2</sup>NN执行高位深灰度图输入的计算任务<sup>[33]</sup>; (d) F-D<sup>2</sup>NN实现显著性提取<sup>[34]</sup>

Fig. 3 Implementation and optimization of D<sup>2</sup>NN. (a) D<sup>2</sup>NN performing classification tasks including handwritten digit recognition<sup>[29]</sup>; (b) optimization of D<sup>2</sup>NN's expressive capability through Fresnel number control<sup>[32]</sup>; (c) D<sup>2</sup>NN processing high bit-depth grayscale image computational tasks<sup>[33]</sup>; (d) F-D<sup>2</sup>NN achieving saliency extraction<sup>[34]</sup>

果。利用傅里叶变换后的光场,可以建立基于卷积的纯振幅神经网络<sup>[35]</sup>或执行分类任务<sup>[67]</sup>等。与实空间类似,考虑光场的非相干性更符合实际应用场景。近期有研究指出,引入空间非相干性可以在一定程度上提高分类任务的精度<sup>[68]</sup>,这可能得益于非相干性的引入提高了对实验中距离、对准等误差的容忍度。

## 2.2 片上集成光计算的物理架构

与自由空间衍射架构大多采用分立光学器件不同,集成光子学近年来取得的重要突破,使得现有工艺可以制备多种集成平台上的片上集成光学器件,包括硅(Si)、氮化硅(Si<sub>3</sub>N<sub>4</sub>)、铌酸锂(LN)、碳化硅(SiC)和砷化镓(GaAs)等<sup>[69-72]</sup>。结合互补金属氧化物半导体(CMOS)工艺进行设计,可制备微型波导、光电探测器、谐振腔等器件。通过以上器件的级联或集成,可以搭建片上的滤波器、信息处理器,也可以构建神经网络、矩阵计算等集成式光子芯片。

### 2.2.1 可编程单元器件的物理特性

对于单一芯片,通过对器件的调制或编程来实现可重构计算是片上光计算的总体方向,其中的可编程单元可以基于微环、波导、马赫-曾德尔干涉(MZI)结构来实现。为详细总结这类器件的物理架构及处理任务,针对两类常见结构的物理特性及调控方式进行总结<sup>[73-74]</sup>。

第一类器件是基于微环谐振腔(MRR)的单通道或双通道器件。MRR是集成光子学中基本的滤波结构<sup>[75]</sup>,可以对特定频率的光信号进行权重重新分配<sup>[76-77]</sup>,经过算法设计,可以实现基本的运算功能。

单通道 MRR 结构如图 4(a)所示。单个微环和一

个直波导耦合,波导两端对应一个输入端和一个输出端,信号分别为  $E_i$ 、 $E_o$ ,则其传递函数<sup>[78]</sup>为

$$T = E_o/E_i = (t - e^{-\gamma + i\phi(\lambda)}) / (1 - te^{-\gamma + i\phi(\lambda)}) \quad (5)$$

式中: $t$ 代表微环与波导间耦合区域传输率; $\gamma$ 代表微环总体的能量损耗; $\phi(\lambda)$ 代表微环在环路中积累的相位。在谐振波长处, $T$ 的模平方也即透射率达到最低,从而达到滤波的作用。以图 4(c)为图 4(b)中微环结构的 block/pass 模式,波导入射光场的波长为  $\lambda_L$ 。电逻辑信号为“1”时谐振波长为  $\lambda_1$ ,则光场可以透射。电逻辑信号为“0”时,谐振波长  $\lambda_0 \approx \lambda_L$ ,则光无法通过,此时光逻辑输出信号和电逻辑输入信号一致。经过器件调整,也可以实现 pass/block 和 pass/pass 模式,从而实现可调的光开关器件。

双通道 MRR 结构如图 4(d)所示。单个微环和两个直波导耦合,两直波导互相垂直,因而有两个输入端和两个输出端, $E_{i1}$ 、 $E_{i2}$ 、 $E_{o1}$ 、 $E_{o2}$  4 个信号通过传递函数构成的传输矩阵相联系,即  $\begin{bmatrix} E_{o1} \\ E_{o2} \end{bmatrix} = T \begin{bmatrix} E_{i1} \\ E_{i2} \end{bmatrix}$ 。假设各处的传输率  $t$  总相等,则传输矩阵为

$$T = \begin{bmatrix} \frac{\alpha_c t (1 - e^{-\gamma + i\phi(\lambda)})}{1 - t^2 e^{-\gamma + i\phi(\lambda)}} & \frac{-\kappa^2 e^{(-\gamma + i\phi(\lambda))/4}}{1 - t^2 e^{-\gamma + i\phi(\lambda)}} \\ \frac{-\kappa^2 \alpha_c^2 e^{3(-\gamma + i\phi(\lambda))/4}}{1 - t^2 e^{-\gamma + i\phi(\lambda)}} & \frac{\alpha_c t (1 - e^{-\gamma + i\phi(\lambda)})}{1 - t^2 e^{-\gamma + i\phi(\lambda)}} \end{bmatrix} \quad (6)$$

式中:新引入的耦合系数  $\kappa$  与  $t$  的平方之和为 1; $\alpha_c$  代表交叉点处的损耗。在单输入的简单情况下(例如  $E_{i1} = 1$ ,  $E_{i2} = 0$ ),非谐振波长处信号由输出端口 1 输出,谐振波长处则由输出端口 2 输出。基于以上两种谐振环即可构建逻辑计算器件,以较少的结构通过算法设计

构成可编程的光信息处理芯片<sup>[74]</sup>。

第二类器件是片上 MZI 结构,其一般结构如图 4(g) 所示。定向耦合器将两个输入端口分配到上下两个波导中接受调制,波导所积累的相位分别为  $\phi_1(\lambda)$  和  $\phi_2(\lambda)$ ,其传输矩阵的形式<sup>[79]</sup>为

$$\mathbf{T} = \text{ie}^{\text{i}\Delta} \begin{bmatrix} \sin \varphi & \cos \varphi \\ \cos \varphi & -\sin \varphi \end{bmatrix} \gamma_1 \quad (7)$$

式中:  $\Delta = \frac{\phi_1 + \phi_2}{2}$ ,  $\varphi = \frac{\phi_1 - \phi_2}{2}$ ;  $\gamma_1$  为定向耦合器及波导引入的损耗。

对于以上两类器件,要进行可重构计算,最重要的是要调节相位累积  $\phi$  和损耗  $\gamma$ 。相位和损耗主要体现在波导材料折射率的实部和虚部。对于硅基器件而言,可以通过热调控<sup>[80]</sup>或载流子注入<sup>[81]</sup>等方式改变其折射率。由于氮化硅的绝缘性,一般通过微型加热器进行热调控<sup>[82]</sup>。对于其他具有电光、声光或热光效应的材料(如铌酸锂、砷化镓等)制成的片上器件,同样可以进行对应的调控以实现光开关等功能<sup>[83-84]</sup>。

### 2.2.2 片上光计算的物理实现

以上依据片上光计算的基本单元作了简要分类,下面将对基于 MRR、MZI、片上衍射结构及其他架构的光计算进展进行总结。

对于 MRR 结构,由于其对频率具有精确选择性,在逻辑运算上有很大的发挥空间。光开关结构及其级联可以构成逻辑上的乘法。此外,布尔运算的 4 类基本单元是与、或、非和异或逻辑运算单元。通过以上最小 MRR 单元的级联构建逻辑运算单元后,可进行逻辑运算求和从而实现任意布尔运算。仍以图 4(b) 结构为例,多个此类微环直接级联时,block/pass 模式代表光学输出和电学逻辑输入一致,pass/block 模式代表其相反,pass/pass 模式则代表光学输出总是 1,以上 3 个模式通过单通道微环 A、B、C 直接级联,即可构成逻辑运算  $\overline{A}BC + A\overline{B}\overline{C}$ 。

进一步地,要想构成布尔运算,需要在逻辑乘法之外引入逻辑求和,一般的架构将波长作为可用自由度<sup>[81,85]</sup>,如图 4(f) 所示:在同一根直波导上加载无串扰的多波长输入,微环的电学逻辑输入由载流子输入和耗尽的电学调控决定,微环工作模式则由热学调控决定,从而实现布尔运算。对于多波长架构,还可以只使用一根直波导串联起多组微环,每组执行特定任务从而实现可重构逻辑运算<sup>[86]</sup>,或采用深度学习校准的办法来降低误差,提高信息处理精度<sup>[87]</sup>,同时降低体系的能量耗散,如利用薄膜铌酸锂平台制备低损耗的 MRR 器件<sup>[88-89]</sup>。为了避免波长可取范围对维度扩展的限制,已有系列研究在单波长架构中实现了逻辑加法运算,如分段调控单个微环结构实现逻辑编程<sup>[90-91]</sup>、设计新型微环结构避免波导模式串扰<sup>[92]</sup>等。基于以上调控方式,可以在类似架构中实现全加运算<sup>[93]</sup>、MAC 计算<sup>[80]</sup>。MRR

器件易于扩展的性质使得其在同一套器件中可执行众多可重构任务<sup>[94-97]</sup>,如可编程分束器、平顶滤波器、光延时器、脉冲整形器、光学微分器与张量运算器等。

对于基于 MZI 的计算任务,早在 1994 年,Reck 等<sup>[98]</sup>已在理论上证明可以通过级联方式实现任意维的西矩阵算子,并利用 3 个分束器和 3 个移相器在实验上验证了该结论。由于 MZI 结构实际上具有分束器和移相器的全部可调自由度,片上的 MZI 结构同样有非常广阔的应用场景。2016 年,牛津大学的 Clements 等<sup>[99]</sup>提出了基于 MZI 级联的矩阵分解方法,相较于 Reck 等<sup>[98]</sup>提出的架构方式,其损耗更低且结构复杂度是原来的 1/2。同年,根特大学的 Ribeiro 等<sup>[100]</sup>在硅基光子学基础上制造了 4×4 维度的通用线性光子芯片,实现了矩阵重构。2017 年,麻省理工学院的 Shen 等<sup>[101]</sup>扩展了构建维度,通过集成 56 个可独立编程的 MZI 结构实现了 4×4 的全光任意矩阵乘法,在任意矩阵的实现上采取了奇异值分解的方式,将任意矩阵分解为一个对角阵和两个西矩阵(或其复共轭)相乘,其中西矩阵由 MZI 结构实现,对角阵由片上光学衰减器实现。2018 年,Hughes 等<sup>[102]</sup>提出了基于 MZI 结构的原位训练方式,如图 4(h) 所示,该体系基于输出端口的功率测量进行反向传播,从而提高了片上神经网络任务的参数训练速度并提高了对制造误差的容忍度。2021 年,Zhang 等<sup>[103]</sup>在 MZI 级联体系中同时考虑了光场的强度和相位信息,实现了复数光学神经网络,不仅将可调神经元数量翻倍,在收敛速度和分类任务精度上也都有提高。2022 年,Zhu 等<sup>[104]</sup>在 MZI 级联体系中加入衍射层,实现傅里叶变换等功能,对计算进行加速,并将功耗从与芯片面积正相关降低至与线度正相关。2023 年,华东师范大学与上海光机所团队<sup>[105]</sup>在薄膜铌酸锂上实现了由 6 个 MZI 结构实现的任意 SU(4) 变换,插入损耗、总功耗和计算保真度等性能表现优异。同年,董建绩团队<sup>[106]</sup>实现了简化 MZI 结构的实值光学矩阵运算。在后续研究中,通过引入波分复用技术<sup>[107]</sup>、多种结构的混合设计<sup>[108]</sup>,显著提升了以 MZI 结构为基础的光子芯片的计算能力。

在以上两类架构之外,还有研究利用片上的亚波长衍射结构<sup>[109-113]</sup>、二维集成波导<sup>[114]</sup>、相变材料(PCM)体系<sup>[115-117]</sup>、阵列波导光栅<sup>[118-119]</sup>进行 MAC 运算、卷积运算和神经网络构建等。

### 2.3 非线性计算的物理原理

在自由空间衍射计算和片上集成计算的各类任务实现过程中,尤其是神经网络的构建,需要非线性函数来提高架构的表达能力。依据线性层的性质,多层线性运算的直接叠加实际上等效于一层线性变换,在不结合非线性层的情况下增加线性层数量并不能提高神经网络的表达能力。同时对数学结构和算法的研究证明,隐藏层必须结合非线性激活层,才能以任意精度一致逼近任意连续函数<sup>[120-121]</sup>。因此在全光计算尤其是

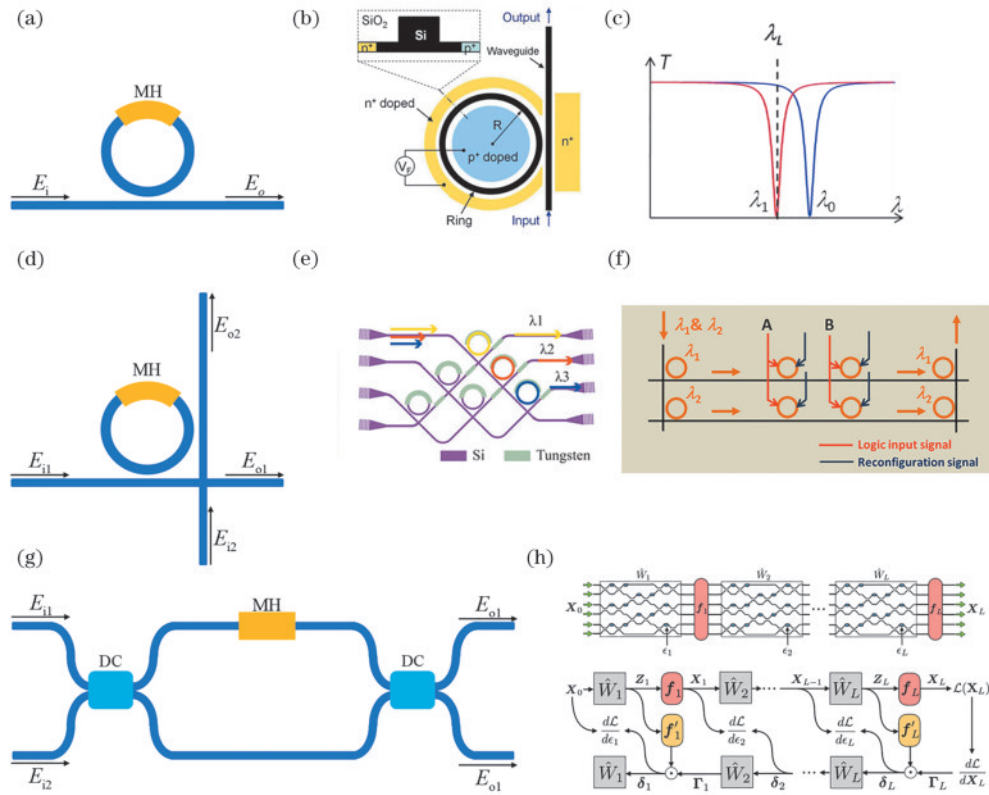


图4 两类典型的片上光计算可编程单元。(a)单通道MRR结构示意图<sup>[78]</sup>;(b)基于MRR的可重构光学开关<sup>[78]</sup>;(c)block/pass模式下的透射光谱<sup>[78]</sup>;(d)双通道MRR结构示意图<sup>[95]</sup>;(e)6个MRR构成的可编程光信息处理芯片<sup>[95]</sup>;(f)基于多波长架构的可重构逻辑计算芯片<sup>[81]</sup>;(g)MZI结构示意图<sup>[102]</sup>;(h)可原位在线训练的片上光学神经网络<sup>[102]</sup>

Fig. 4 Two typical types of programmable units for on-chip optical computing. (a) Schematic of a single-channel MRR structure<sup>[78]</sup>; (b) reconfigurable optical switch based on MRR<sup>[78]</sup>; (c) transmission spectra in block/pass modes<sup>[78]</sup>; (d) schematic of a dual-channel MRR structure<sup>[95]</sup>; (e) programmable optical information processing chip composed of six MRR<sup>[95]</sup>; (f) reconfigurable logic computing chip based on a multi-wavelength architecture<sup>[81]</sup>; (g) schematic of a MZI structure<sup>[102]</sup>; (h) on-chip optical neural network supporting *in situ* online training<sup>[102]</sup>

光学神经网络的构建中非线性激活函数是很必要的。

由于弱光条件下可用的非线性光学器件缺乏,目前主流的实现方案仍然采用光电相结合的方式进行调控。下面从自由空间衍射光计算和片上集成光计算两个方面总结目前的光学非线性函数实现方式。

### 2.3.1 自由空间衍射光计算中的非线性函数

在初期探索阶段,光电相结合的方式主要利用光电探测器<sup>[121]</sup>,但存在转换效率较低、光电转换带来的时间延迟等问题。2023年,康奈尔大学的Wang等<sup>[122]</sup>提出了利用微型通道板实现非线性激活的方案,即利用光电探测阴极和微通道板对光子吸收的饱和效应模拟Sigmoid激活函数,并利用微通道板增强光子。为解决光电探测的响应延迟等问题,Ashtiani等<sup>[123]</sup>提出一种新型光电探测器,如图5(a)所示。探测器基本单元是垂直排列的一系列光电二极管,二极管在吸收光子产生光电子的同时利用热效应促进光电流再生,从而模拟ReLU激活函数,缩短了探测延迟并提高了能量利用率。2024年,Zhang等<sup>[124]</sup>通过透明光电晶体管和液晶的组合形成如图5(b)所示的基本单元,可排列为大规模神经元调制阵列。其非线性源于透明光电晶

体管电阻随入射光强的变化特性,进而调控对应像素处液晶感受到的电压调制,利用液晶的旋光效应改变强度,最终体现为不同光强的透射率不同,实现了弱光下的非线性响应。

在光电结合方式之外,也可以考虑利用新型材料进行全光非线性激活函数的设计,这些材料包括光饱和和吸收体<sup>[125]</sup>、磁致透明效应器件<sup>[126]</sup>、纳米材料<sup>[127]</sup>、多层石墨烯<sup>[128]</sup>、量子点材料<sup>[129]</sup>。2024年,Shi等<sup>[129]</sup>采用钙钛矿量子点薄膜近似实现了ReLU激活函数,并验证了其作为D<sup>2</sup>NN中非线性层对神经网络表达能力的提升作用,图5(c)展示了钙钛矿薄膜在实际工作中的应用形式。2021年,Yu等<sup>[130]</sup>采用等离子体纳米微腔作为结构单元,设计了电调谐超构表面,对图5(d)所示的结构单元在不同调制电压下发射的二次谐波的振幅、相位进行调制,从而以电学频率动态操纵光束。

### 2.3.2 片上集成光计算中的非线性函数

片上集成光子学中早期常使用光学计算-电学输出的形式进行信息采集,将含振幅/相位的光场转换为纯强度电信号,实现非线性。这种方法不易扩展,且损失了光学计算的高能效、高计算速率的优点,对这一问

题的解决方案体现为光-电-光转换方式和片上全光方式<sup>[131]</sup>。

在集成光子学中,利用光电转换进行非线性调制时,常采用两种方式,即电光调制器(EOM)<sup>[132-133]</sup>或者电吸收调制器(EAM)与光电二极管相连,构成非线性元件。2019年,George等<sup>[134]</sup>讨论了5种EAM固有激活函数的性能,并于同年报道了如何将该器件集成到硅光子波导上,同时通过电控铟锡氧化物调控器件特

性<sup>[135]</sup>,如图6(a)所示。在光-电-光转换方面,还有一些其他方向的研究,如对MRR器件进行掺杂,使其对频率的选择性随偏压控制而改变,进而实现对振幅、相位的调制,产生不同的非线性激活函数<sup>[136]</sup>。Chen等<sup>[137]</sup>结合光学芯片和光电二极管构成的电学芯片实现全连接神经网络,由于无需电学读出过程,其计算延迟明显降低。电学部分芯片如图6(b)所示,通过光电效应实现光学非线性。

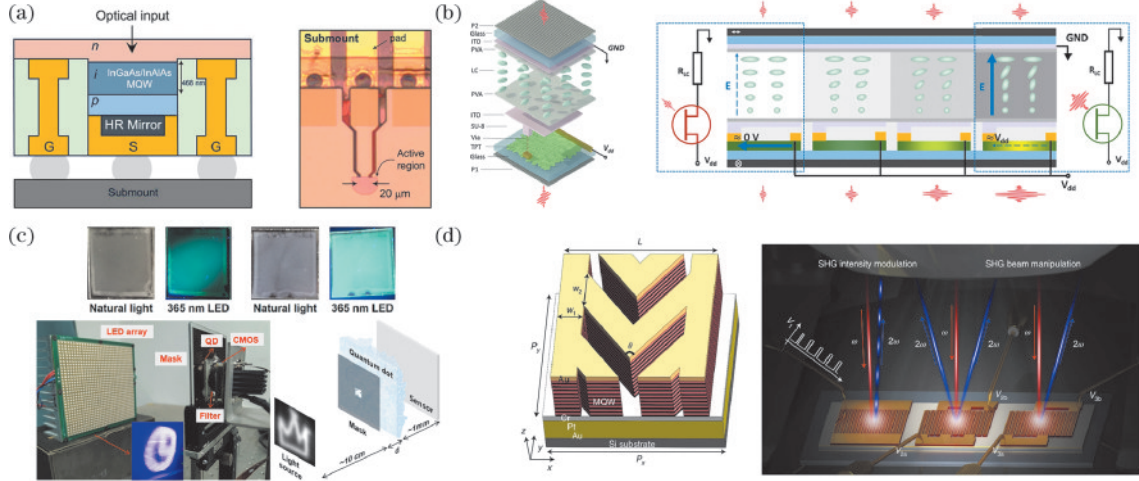


图5 自由空间衍射光计算中的非线性激活方式。(a)表面法向光电探测器的最小单元结构<sup>[123]</sup>; (b)利用透明光电晶体管和液晶调制器实现非相干神经元调制阵列<sup>[124]</sup>; (c)基于钙钛矿量子点薄膜实现非线性激活<sup>[129]</sup>; (d)电可调谐的非线性极化超构表面最小单元及其工作示意图<sup>[130]</sup>

Fig. 5 Implementation of nonlinear activation methods in free-space diffractive optical computing. (a) Unit cell structure of surface-normal photodetector<sup>[123]</sup>; (b) implementation of incoherent neuron modulation array using transparent phototransistor and liquid crystal modulator<sup>[124]</sup>; (c) nonlinear activation implementation based on perovskite quantum dots film<sup>[129]</sup>; (d) minimal unit structure and operation schematic of electrically tunable nonlinear polaritonic meta-surface<sup>[130]</sup>

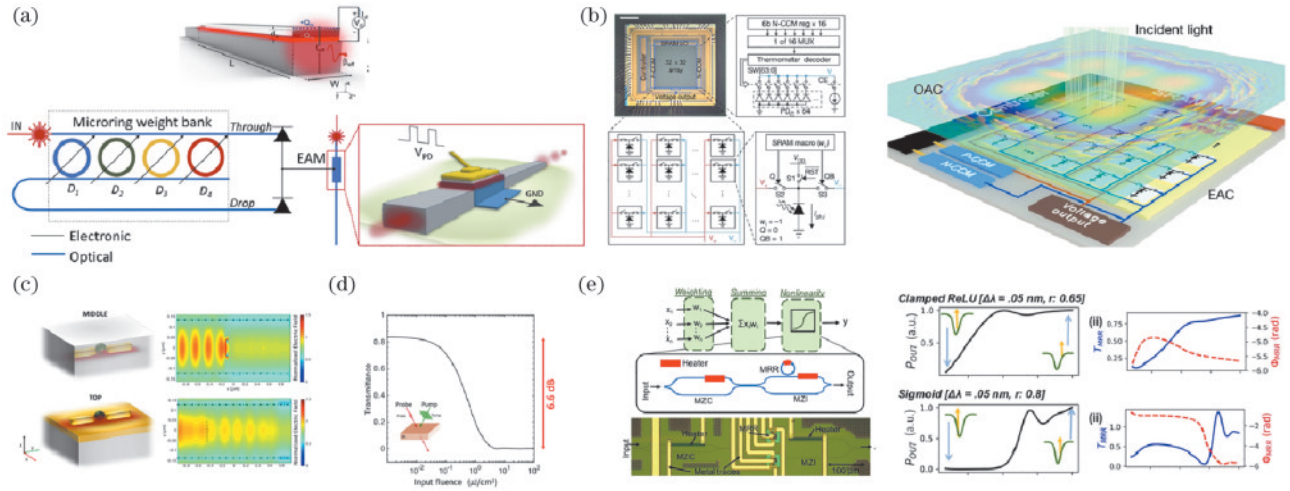


图6 片上集成光计算中的非线性激活方式。(a)集成在硅波导上的EAM<sup>[135]</sup>; (b)光电二极管阵列的电学芯片作为光电融合计算的非线性层<sup>[137]</sup>; 利用电磁诱导透明结构(c)、通过富勒烯实现反饱和和吸收(d)的全光非线性实现方式<sup>[138]</sup>; (e) MRR、MZI复合结构实现Clamped ReLU、Sigmoid等非线性激活函数<sup>[139]</sup>

Fig. 6 Implementation of nonlinear activation methods in on-chip integrated optical computing. (a) EAM integrated on silicon waveguide<sup>[135]</sup>; (b) photodiode array electronic chip serving as nonlinear layer for optoelectronic computing<sup>[137]</sup>; all-optical nonlinear implementation approaches using electromagnetically induced transparency structure (c) and reverse saturable absorption through fullerene (d)<sup>[138]</sup>; (e) hybrid structure with MRR and MZI implementing Clamped ReLU and Sigmoid nonlinear activation functions<sup>[139]</sup>

集成光子学中的全光非线性方案一般通过计算光信号之外的激光对材料进行调控,即基于激光照射下材料光学特性的变化,对计算光信号进行调制。这种调制方式对施加调制的激光功率要求较高,因此研究多集中在更高效灵活的调制方式上。2018 年, Miscuglio 等<sup>[138]</sup>提出了两种实现方案:一种如图 6(c)所示,其核心是在一对金纳米棒中间放入量子点材料,二者在部分频率下发生干涉相消,使得材料对光场的吸收降低,从而增强光在弱光下的非线性效应;另一种如图 6(d)所示,富勒烯的吸收特性随光强增加而增强,以此特性构建反饱和和吸收体实现光学非线性。2020 年, Jha 等<sup>[139]</sup>在硅基平台利用 MRR 和腔载 MZI 结构的设计,实现了不同波长入射下的不同非线性函数,如图 6(e)所示。图 6(e)展示了在入射波长偏离 MRR 谐振波长 0.05 nm、马赫-曾德尔耦合器偏置分别为 0.65 和 0.80 时实现的 Clamped ReLU、Sigmoid 两种激活函数,同时报道了另外两种激活函数 Radial-basis 和 Softplus 的调节方式。2022 年, Guo 等<sup>[140]</sup>利用铌酸锂纳米波导的强瞬时二次非线性实现了高速全光开关并设计了一种非线性分路器。同年, Wu 等<sup>[141]</sup>通过设计硅锗结构 MRR,实现了低功率阈值下的非线性全光激活函数。2023 年, Zhang 等<sup>[142]</sup>利用高维空间的线性操作实现了光学非线性运算。2025 年,宾夕法尼亚大学冯亮团队<sup>[143]</sup>在有源半导体中利用空间调控分布式载流子及其动力学,将光子场可编程能力拓展到非线性领域,从而实现多种非线性激活函数和神经网络原位训练。

综上所述,本章从物理原理出发,阐述了模拟光计算的两大技术路径及其挑战。自由空间架构利用衍射理论,通过 SLM 或超构表面等器件对波前进行调制,实现大规模并行计算;片上集成架构则基于光波导、MZI 和 MRR 等基本单元,在芯片上构建可编程的光路网络以执行计算任务。两种架构都面临共性的挑战,即如何实现全光神经网络所必需的光学非线性激活函数,并使非线性层可调节,以满足网络重构的要求。基于物理原理,两种架构各有其独特的优势和挑战:自由空间衍射光计算以其天然的二维并行处理能力见长,能够直接处理光场波前信息,在神经网络的可扩展性上具有优势;与之相对,片上架构在集成度和鲁棒性上表现更优,能够构建紧凑、稳定的光计算器件,并且其在当前最先进的推理任务上表现优异,但在需要处理视觉信息时,需要额外的光电单元辅助。因此,结合二者优势是未来光计算架构发展的重要方向。在下一章中将依据以上物理原理,对各类光计算任务的前沿成果进行总结介绍。

### 3 各类光计算任务的应用

有了材料、器件和算法的支持,经过适当设计,基于光的计算系统就能够以光速执行计算任务。在任务种类划分上,一类是传统计算任务如 MAC、卷积、微/

积分等,光计算系统提供了备选方案;另一类则是发展迅速的神经网络类计算任务,要兼顾计算并行性、能效比和制造工艺等,探索光电融合的神经网络架构。下面将基于前文所述的物理架构,总结当前各类计算任务中光计算的探索方向和重要成果。

#### 3.1 非神经网络类计算任务

在非神经网络的传统计算任务中,最基本的是 MAC 任务,在各种架构上都具有良好的表现:1) MZI 级联实现的矩阵分解<sup>[99]</sup> [图 7(a)]、矩阵乘法计算<sup>[101]</sup>; 2) 载流子掺杂设计的片上光衰减器实现矩阵运算<sup>[144]</sup>; 3)  $N \times N$  的 MRR 阵列模拟任意  $N \times N$  传输矩阵<sup>[77]</sup>; 4) 基于 PCM 的三维数据处理<sup>[117]</sup> [图 7(b)]等。

矩阵运算之外,还有图像处理等常见二维信息处理任务。在这一任务上,自由空间衍射架构因并行能力更强,分立的光学器件因多通道容量而更具优势。下面介绍几种可实现图像信息处理的光学计算器件。

第一类是微分算子,可以用来对目标图像进行边缘提取,从而有利于后续的主体识别、分类等任务,在自动驾驶、生物医学成像等领域都有广泛的应用。计算机以算法实现的微分难以在处理自然界中光信息的同时保持低能耗、高并行,通过对光计算器件进行合适的设计即可同时满足以上需求<sup>[145]</sup>。从傅里叶光学的角度分析,一阶微分器等价于频域传递函数正比于波矢的平行分量。基于此, Xu 等<sup>[146]</sup>在 2022 年利用几何相位设计的超表面,在频域构造了这一传递函数,实现了一阶微分和对图像的边缘提取。Liu 等<sup>[147]</sup>指出,  $n$  阶微分器的传递函数应当正比于波矢平行分量的  $n$  次幂,据此设计了高阶平面光子芯片,如图 7(c)所示。微分器件的集成化、紧凑化有利于更广泛的应用, Huo 等<sup>[148]</sup>提出了基于贝塞尔涡旋的超透镜,可以在宽带上同步执行多阶微分。最近,还有工作<sup>[149]</sup>实现了宽带消色差的角谱操纵器件,其在光学微分上有良好的效果。光学微分算子不仅可以用于边缘提取,在光场三维重建<sup>[150]</sup>、物体相位成像<sup>[151]</sup> [图 7(d)]等方面也有诸多应用。

第二类是卷积算子,可以提取图像的局部特征用于后续处理,卷积层和全连接层即构成卷积神经网络。惠更斯原理表明,光场的衍射过程可被视作自然的卷积过程,因此可以执行大规模卷积运算,从而利用全光的卷积算子替代神经网络的卷积层。而全连接层可通过计算机实现,从而构建光电融合的卷积神经网络。早在 2016 年, Chen 等<sup>[152]</sup>就利用角度敏感像素实现了卷积操作。2018 年, Chang 等<sup>[153]</sup>利用 4f 系统和掩模板实现了光学卷积,如图 8(a)所示。2021 年, Xu 等<sup>[154]</sup>提出了利用微腔孤子光频梳编码卷积核,可以制作波长、时间、空间复用的光子卷积加速器。2022 年, Shi 等<sup>[155]</sup>提出了利用掩模板和图像传感器实现的光学卷积算子。Xu 等<sup>[156]</sup>基于 MRR 阵列制造了一种高阶光子张量流处理芯片,实现了 20 GBaud 的高速张量流卷积运算,并搭建了卷积神经网络,如图 8(b)所示。2024 年,

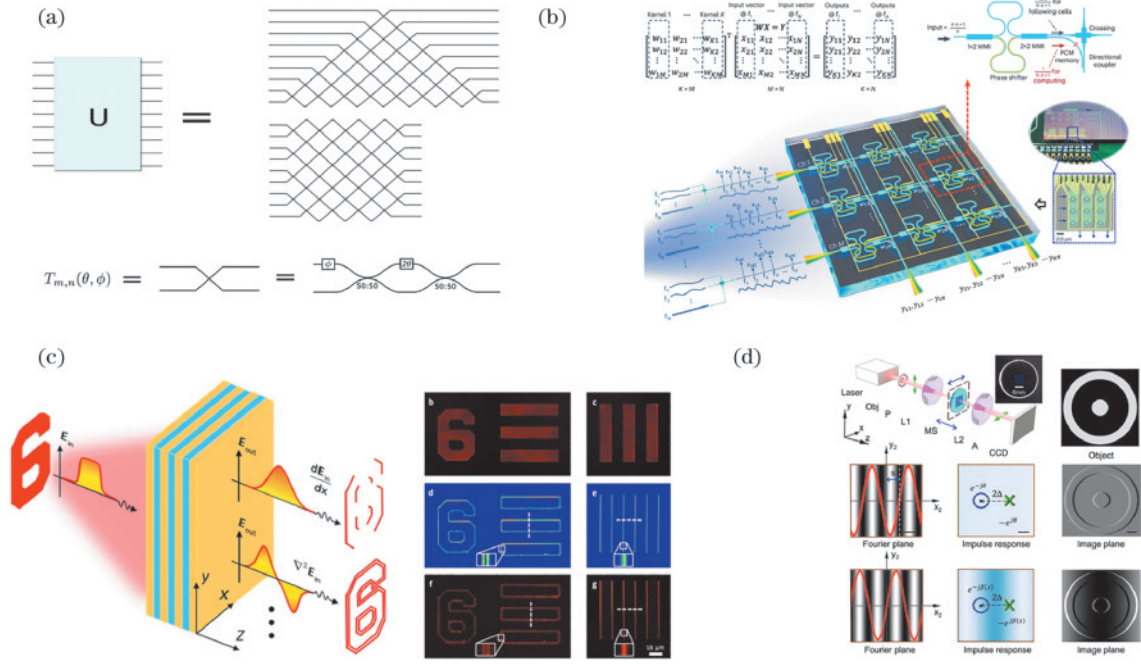


图7 非神经网络类光计算任务中的微分算子。(a)利用分束器和移相器实现复杂度降低一半的西矩阵构建<sup>[99]</sup>；(b)基于PCM的三维数据处理架构<sup>[117]</sup>；(c)可实现高阶光学微分的平面光子芯片<sup>[147]</sup>；(d)基于超表面的定量相位梯度成像<sup>[151]</sup>

Fig. 7 Differential operators in non-neural-network optical computing tasks. (a) Implementation of unitary matrix construction with halved complexity using beam splitters and phase shifters<sup>[99]</sup>; (b) three-dimensional data processing architecture based on PCM<sup>[117]</sup>; (c) planar photonic chip enabling high-order optical differentiation<sup>[147]</sup>; (d) quantitative phase gradient imaging through meta-surface implementation<sup>[151]</sup>

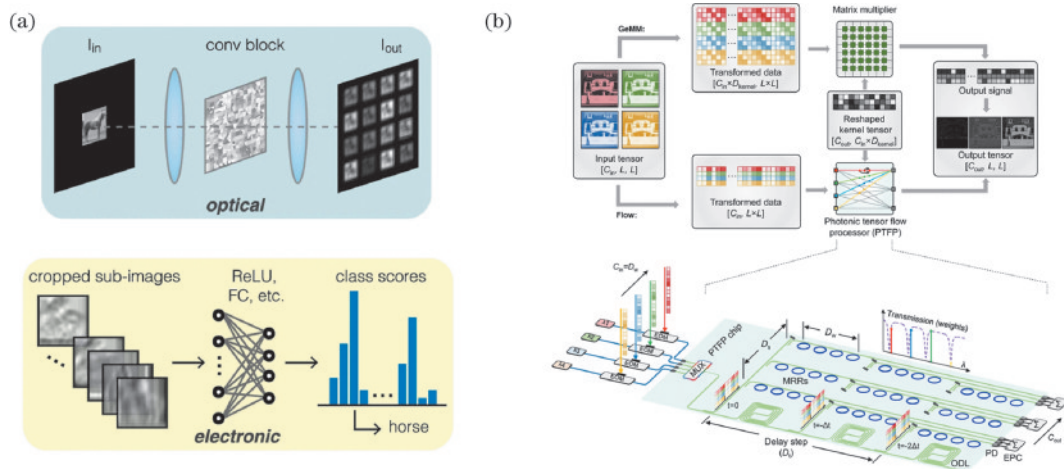


图8 非神经网络类光计算任务中的卷积算子。(a)利用4f系统、掩模板实现光学卷积计算和图像分类<sup>[153]</sup>；(b)高阶光子张量流处理器<sup>[156]</sup>

Fig. 8 Convolution operators in non-neural-network optical computing tasks. (a) Optical convolution computation and image classification achieved via 4f optical system with mask modulation<sup>[153]</sup>; (b) high-order photonic tensor flow processor<sup>[156]</sup>

Ouyang 等<sup>[157]</sup>利用一种并行边缘提取 MRR 芯片架构同时卷积提取 4 个特征,在相同规模和功率驱动下实现 4 倍卷积速度。

与光学神经网络相比,光学算子的结构一般更加简单,其优势在于可直接替代电子计算机的某一计算过程且易于制造。光学卷积、微分算子都被广泛用在神经网络构建中,但其受限于功能的单一性。在 AI 发展迅猛的时代,光学神经网络类的计算任务同样是非常重要的研究方向。

### 3.2 光学神经网络类计算任务

#### 3.2.1 光学神经网络及其新兴应用

在作为自由空间衍射架构的 D<sup>2</sup>NN 提出后,相关研究利用其取得了与电子计算机神经网络相当的效果。这种架构的设计并不限制波长和器件尺寸,可以在各类任务中方便地重构。例如,2025 年, Yu 等<sup>[158]</sup>在理论上采用这一计算方法,在实验中基于制作超构表面的 3D 双光子纳米光刻技术,在直径小于 0.1 mm 的光纤端口处精准制造了多层衍射神经网络,如图 9(a)

所示。这一精确对准的衍射神经网络可以对光纤出射的杂乱振幅和相位信息进行实时解析,尤其是在光纤弯折或扰动时仍可成像,且多模光纤更使得信息容量得到极大扩展。 $D^2NN$ 的易扩展性和泛用性将使其成为极具潜力的光计算范式。

储备池光计算机是一种基于光学芯片的循环神经网络,同时具有人工突触和人工神经元的功能。相对其他有反馈的循环神经网络,具有更快的训练速度和更低的训练成本,在时序信号相关的语言处理、

时序信号处理、回归预测等任务上表现优异<sup>[159-160]</sup>。2023年,Li等<sup>[161]</sup>利用波分复用技术解决了储备池光计算的带宽问题,同年该团队<sup>[162]</sup>实现了具有多个隐藏层的深度储备池光计算机。如图9(b)所示,该储备池光计算机具有4个储备池层,每一层由半导体激光芯片和光学反馈环路构成,每个储备池层的输出通过光注入锁定技术输入下一层,从而实现全光连接。该低能耗、低时延架构在边缘计算等领域有广泛的应用前景。

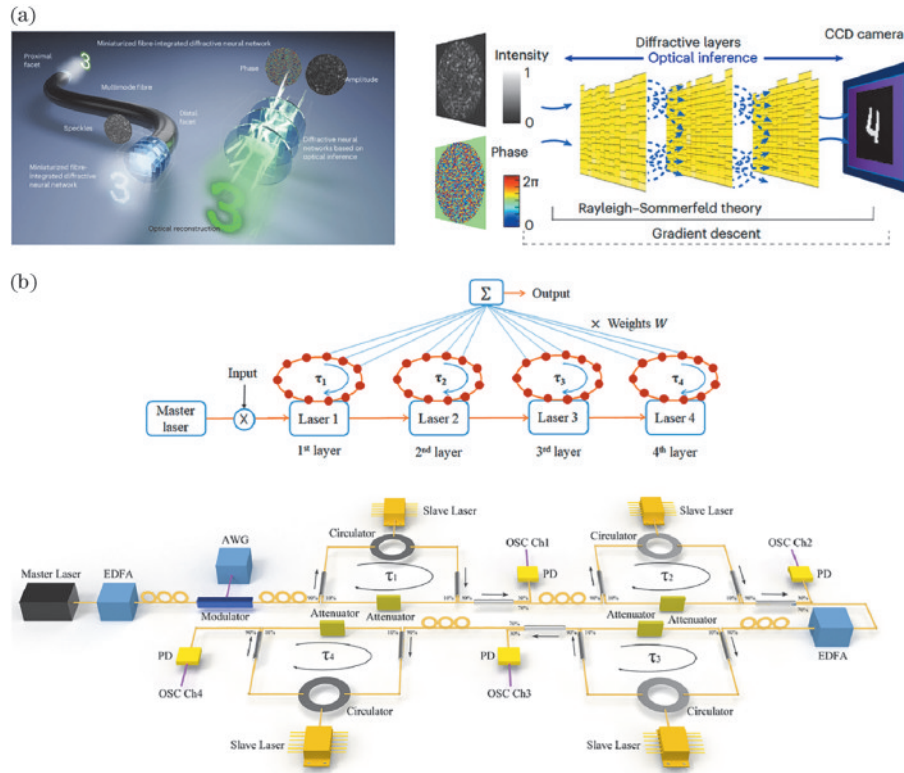


图9 光神经网络计算的物理架构及应用。(a)基于光纤端口集成衍射神经网络的实时抗干扰成像<sup>[158]</sup>;(b)深度储备池计算循环网络<sup>[162]</sup>  
Fig. 9 Various neural network-inspired optical computing architectures and their applications. (a) Real-time anti-interference imaging via a fiber-port-integrated diffractive neural network<sup>[158]</sup>; (b) deep reservoir computing recurrent network<sup>[162]</sup>

上述光学神经网络常用于图像处理、时序信号处理等,组合问题也可以在类似架构中得到解决。组合问题可以映射为复杂伊辛模型,寻找组合问题最优解的过程等价于寻找伊辛模型的基态,可用以解决资源配置、路线规划等大规模组合问题。基于SLM等器件的空间光伊辛机具有规模大、计算快且能效比高的特点。2019年,Pierangeli等<sup>[163]</sup>用SLM实现傅里叶光学系统的空间光伊辛机,如图10(a)所示。SLM相位经二进制编码替代自旋,由振幅调制和电学回馈方案实现自旋的相互作用从而求解基态。为解决多层SLM的对准问题,2021年,Fang等<sup>[164]</sup>提出基于规范变换的单层SLM求解架构,提高了求解稳定性。2023年,Luo等<sup>[165]</sup>提出了基于规范变换并结合波分复用技术的空间光伊辛机,在图分割问题上,该算法能在1500次迭代内以大于92%的概率找到基态,对应于组合问题最优解。2025年,董建绩团队<sup>[166]</sup>构建了基于光电耦合振

荡器的单片集成四自旋伊辛机,在全光物理退火系统的单片集成这一方向迈出了重要一步,大幅减少了伊辛机占地面积、功耗和收敛时间,如图10(b)所示。

### 3.2.2 光学神经网络的重构与在线训练

在以上介绍中,衍射神经网络的衍射层像素单元或片上计算体系中的调控参数一经设计并加载,系统就可以光速执行计算。但这样的系统无法高效地重构从而执行多种任务,同时受到光路系统误差的影响,实验结果往往难以达到理论计算的效果。在具体器件的制备上,体系的精度受限于微纳加工技术,导致系统工作场景难以拓展,即存在电学训练和光学计算不匹配的问题。为解决这一问题,Zhou等<sup>[21]</sup>在2021年提出了规模可扩展、任务可重构的衍射处理单元,并通过开发自适应算法提高了训练精度,如图11(a)所示。

此外,有研究者直接在线训练基于物理体系的神经网络。2022年,Wright等<sup>[167]</sup>提出了可计算反向传播

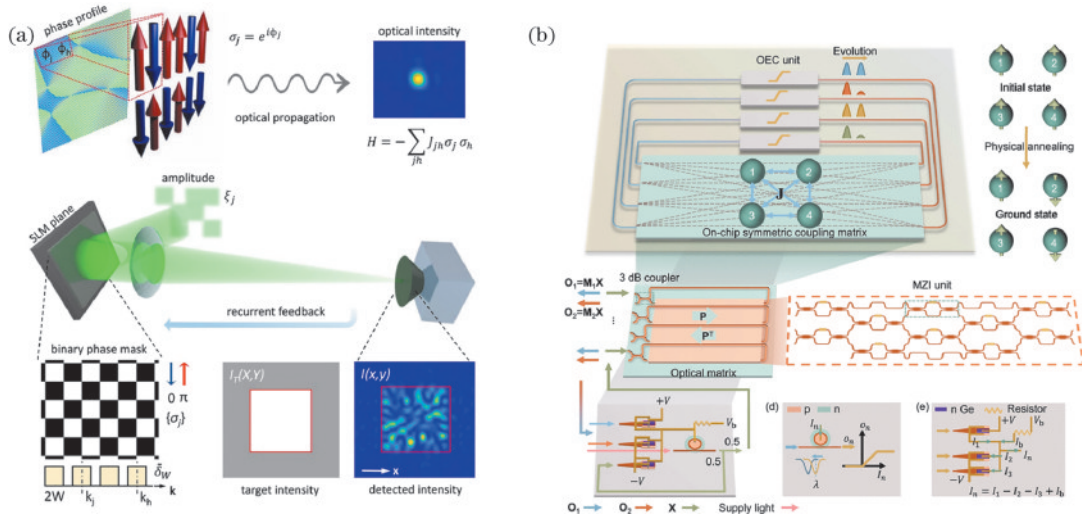


图 10 空间光伊辛机。(a)基于SLM的空间光伊辛机<sup>[163]</sup>;(b)单片集成的四自旋伊辛机<sup>[166]</sup>

Fig. 10 Spatial photonic Ising machine. (a) Spatial photonic Ising machine implemented via SLM<sup>[163]</sup>; (b) monolithically integrated four-spin Ising machine<sup>[166]</sup>

梯度的物理神经网络构建方式,即物理感知训练方案,在物理层和传统的人工神经网络不具有同构关系的情况下即可实现深度神经网络的训练。在这一方法的指导下,该团队在力学、电子学和光学领域均完成音频和图像分类任务,其中光学手段表现优异。2023年,Momeni等<sup>[168]</sup>又提出了无需反向传播的算法,同样在声学、微波、可见光领域均验证了其算法,如图11(b)所示。随着光计算规模的扩大,原位的在线训练将是很重要的发展方向<sup>[169-170]</sup>。

在任务的扩展性上,如图11(c)所示,清华大学戴琼海团队<sup>[108]</sup>在2024年建立干涉、衍射联合设计的传播模型,研制了国际领先的光计算芯片Taichi,在各种分类识别任务、生成任务中表现良好。2025年,清华大学陈宏伟团队<sup>[171]</sup>提出了基于存储式光计算的多任务光子芯片,如图11(d)所示。利用转移学习原理,将大部分参数固定在光学元件中以保持稳定,同时将少量参数分配到可调元件中。该体系在保持多任务处理能力的基础上实现了与电子计算机神经网络相当的准确率,在存储式光计算方向迈出了重要一步。

## 4 总结与展望

在AI大模型发展迅猛的当下,规模、效率与能耗构成的“不可能三角”正严重制约AI发展速度。光计算凭借高并行性、高速与低能耗,成为解决这一问题的有力架构。上述光计算架构的物理基础包括自由空间衍射体系、片上集成体系及非线性光学体系,这些体系均利用光子优异特性构建了各自的计算方案<sup>[172]</sup>。无论是执行固定任务的光计算算子器件,还是光学神经网络,在光电融合发展的趋势下,算力增长速度尚不匹配AI大模型发展速度的背景下,都将为AGI的研发与实现提供坚实的平台<sup>[173]</sup>。

当下,各种光计算的物理架构不断涌现,在商用光子芯片领域表现突出。最近,曦智科技<sup>[174]</sup>、Lightmatter<sup>[175]</sup>各自独立发布了新型的商用光子芯片,能够实现高速矩阵运算、执行AI处理任务,在光子芯片的规模扩展和技术革新上有重大突破。这些突破转化为与通用图形处理器(以NVIDIA H100 PCIe为参照<sup>[176-177]</sup>)相比的数量级优势:在能效比上,NVIDIA H100 PCIe的峰值性能约为0.15 TOPS/W(Tera Operations per Second per Watt),通用光计算加速器可以达到2.38 TOPS/W<sup>[174]</sup>;在计算速度上,电子计算因时钟频率与内存访问限制,延迟通常在微秒级别<sup>[178]</sup>,而光计算利用光速传播完成运算,核心延迟主要取决于光程传播,可低至纳秒级别。

目前,这类光子芯片与普及化、通用化还有一定的距离。模拟光计算是主流发展方向,但其物理系统内部的噪声、电磁结构加工技艺的误差、光学器件的不稳定性及对准问题都会使得实验偏离理论精度。电子计算体系凭借成熟的工业配套和低噪声、高误差容忍度仍然保持优势,因此光电融合计算是目前的一个优秀解决方案。

在光计算的发展方向上,存在3个主要挑战。1)工艺误差、器件不稳定等因素引入的计算误差需要补偿办法。2)用作激活函数的非线性器件尚不能满足需求。使用现场可编程门阵列(FPGA)进行光电转换后离线进行非线性变换的方案会显著增加延迟和能耗,在非线性激活函数之外还需要更加集成化的可调的光学非线性材料以提高光学神经网络的表达能力。3)光计算任务的算法和计算架构需要革新,并发展非易失性光学存储方式<sup>[179-180]</sup>。一方面,现有光计算架构由于物理体系的限制无法跟上电子计算机领域的最新发展,仅能替代简单的卷积、微分与MAC运算,或对图

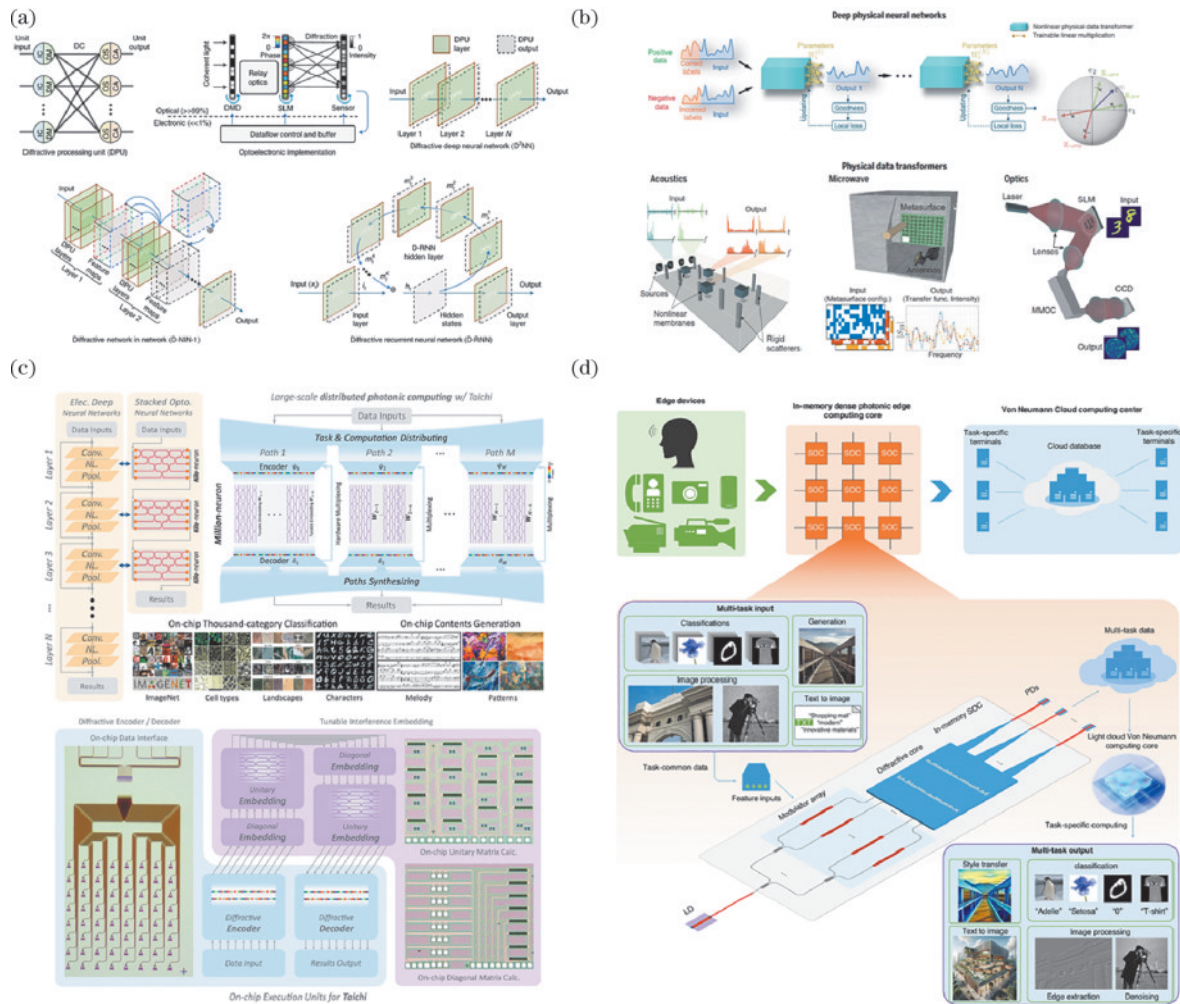


图 11 多任务可重构光神经网络及无反向传播神经网络训练。(a)可重构衍射处理单元构建衍射神经网络<sup>[21]</sup>;(b)无反向传播的深度物理神经网络<sup>[168]</sup>;(c)衍射、干涉联合设计的大规模光子芯片 Taichi<sup>[108]</sup>;(d)基于存储式光学计算的多任务光子芯片<sup>[171]</sup>

Fig. 11 Multi-task reconfigurable optical neural networks and backpropagation-free neural network training. (a) Reconfigurable diffractive processing unit constructing a diffractive neural network<sup>[21]</sup>; (b) backpropagation-free deep physical neural network<sup>[168]</sup>; (c) large-scale photonic chip Taichi with co-designed diffractive and interferometric components<sup>[108]</sup>; (d) multi-task photonic chip based on in-memory optical computing<sup>[171]</sup>

像、视频数据集进行简单处理,无法匹配最新的大模型领域成果,计算架构的革新速度因此受限。另一方面,原位在线训练的方式仍然有待开发,以降低噪声干扰、提高对计算误差的纠偏能力,从而使光计算得到更广泛的应用<sup>[181]</sup>。

在将来,光计算物理架构有以下发展方向。1)利用集成式光计算系统与 CMOS 工艺兼容性好、集成度高的特性,实现基于这类器件的光电融合计算架构,真正服务于计算机前沿研究。2)借助计算机领域的网络架构更新 D<sup>2</sup>NN 架构,同时注重非线性可调器件开发,两个方向协同发展,最终实现兼顾光子全部优良特性的全光神经网络,解决“不可能三角”问题,以光速和极低损耗执行大规模计算任务。3)在计算任务上进行扩展。对分类、识别、视频和图像处理等特定任务进行扩展,使光计算在日常生活的更多领域中得到应用:在物联网边缘计算、智能驾驶、社会科学、金融等领域的复

杂模拟计算中开发算法、执行任务。

全光计算是光计算长期追求的重要目标,光电融合计算是当下核心的科学议题。在发展光计算的过程中,既要聚焦根本科学问题,回答“能力上限”的问题,又要结合实际应用,回答“该怎么用”的问题,解决实际生产中的重点问题。应统筹兼顾、协调发展,构建“前沿研究—实际生产—产业发展”的链路,同时产业发展的需求也将指明前沿发展方向,帮助构建正反馈的产学研协同发展模式,共同助力 AGI 的早日实现。

## 参 考 文 献

- [1] Voulodimos A, Doulamis N, Doulamis A, et al. Deep learning for computer vision: a brief review[J]. Computational Intelligence and Neuroscience, 2018, 2018: 7068349.
- [2] Guo D Y, Yang D J, Zhang H W, et al. Deepseek-R1: incentivizing reasoning capability in LLMs via

- reinforcement learning[EB/OL]. (2025-01-22) [2025-05-01]. <https://arxiv.org/abs/2501.12948>.
- [3] Cheng S N, Chen Y J, Chiu W C, et al. *Adaptively-realistic image generation from stroke and sketch with diffusion model*[C]//2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), January 2-7, 2023, Waikoloa, HI, USA. New York: IEEE Press, 2023: 4043-4051.
  - [4] 喻国明, 苏健威. 生成式人工智能浪潮下的传播革命与媒介生态: 从 ChatGPT 到全面智能化时代的未来[J]. 新疆师范大学学报(哲学社会科学版), 2023, 44(5): 81-90.  
Yu G M, Su J W. Communication revolution and media ecology in the context of generative AI: from ChatGPT to a comprehensively intelligent future[J]. Journal of Xinjiang Normal University (Edition of Philosophy and Social Sciences), 2023, 44(5): 81-90.
  - [5] Sastry G, Heim L, Belfield H, et al. *Computing power and the governance of artificial intelligence*[EB/OL]. (2024-02-13)[2024-12-12]. <https://arxiv.org/abs/2402.08797>.
  - [6] Chhowalla M, Jena D, Zhang H. *Two-dimensional semiconductors for transistors*[J]. Nature Reviews Materials, 2016, 1: 16052.
  - [7] Das S, Sebastian A, Pop E, et al. *Transistors based on two-dimensional materials for future integrated circuits* [J]. Nature Electronics, 2021, 4(11): 786-799.
  - [8] Xu B H, Chen R M, Zhou J R, et al. *Recent progress and challenges regarding carbon nanotube on-chip interconnects*[J]. Micromachines, 2022, 13(7): 1148.
  - [9] 成骏伟, 江雪怡, 周海龙, 等. 光电智能计算研究进展与挑战[J]. 中国激光, 2022, 49(12): 1219001.  
Cheng J W, Jiang X Y, Zhou H L, et al. *Advances and challenges of optoelectronic intelligent computing*[J]. Chinese Journal of Lasers, 2022, 49(12): 1219001.
  - [10] 陈宏伟, 于振明, 张天, 等. 光子神经网络发展与挑战[J]. 中国激光, 2020, 47(5): 0500004.  
Chen H W, Yu Z M, Zhang T, et al. *Advances and challenges of optical neural networks*[J]. Chinese Journal of Lasers, 2020, 47(5): 0500004.
  - [11] Zhou H L, Dong J J, Cheng J W, et al. *Photonic matrix multiplication lights up photonic accelerator and beyond* [J]. Light: Science & Applications, 2022, 11: 30.
  - [12] Zhang X L, Xu J, Dong J J, et al. *All-optical logic gates based on semiconductor optical amplifiers and tunable filters*[M]//Dolev S, Oltean M. Optical supercomputing. Lecture notes in computer science. Berlin: Springer Berlin Heidelberg, 2009, 5882: 19-29.
  - [13] Weaver C S, Goodman J W. *A technique for optically convolving two functions*[J]. Applied Optics, 1966, 5(7): 1248-1249.
  - [14] Meng X Y, Zhang G J, Shi N N, et al. *Compact optical convolution processing unit based on multimode interference*[J]. Nature Communications, 2023, 14: 3000.
  - [15] Fu W W, Zhao D, Li Z Q, et al. *Ultracompact meta-imagers for arbitrary all-optical convolution*[J]. Light: Science & Applications, 2022, 11: 62.
  - [16] Zhu T F, Guo C, Huang J Y, et al. *Topological optical differentiator*[J]. Nature Communications, 2021, 12: 680.
  - [17] 周治平, 许鹏飞, 董晓文. 硅基光电计算[J]. 中国激光, 2020, 47(6): 0600001.  
Zhou Z P, Xu P F, Dong X W. *Computing on silicon photonic platform*[J]. Chinese Journal of Lasers, 2020, 47(6): 0600001.
  - [18] Li C, Zhang X, Li J W, et al. *The challenges of modern computing and new opportunities for optics*[J]. Photonix, 2021, 2(1): 20.
  - [19] Zhou C H, Yu J J, Li G W, et al. *Roadmap of optical computing*[J]. Proceedings of SPIE, 2020, 11898: 118981B.
  - [20] 陈蓓, 张肇阳, 戴庭舸, 等. 光学神经网络及其应用[J]. 激光与光电子学进展, 2023, 60(6): 0600001.  
Chen B, Zhang Z Y, Dai T G, et al. *Photonic neural networks and its applications*[J]. Laser & Optoelectronics Progress, 2023, 60(6): 0600001.
  - [21] Zhou T K, Lin X, Wu J M, et al. *Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit*[J]. Nature Photonics, 2021, 15 (5): 367-373.
  - [22] Buckley S M, Tait A N, McCaughan A N, et al. *Photonic online learning: a perspective*[J]. Nanophotonics, 2023, 12(5): 833-845.
  - [23] Goodman J W. Introduction to Fourier optics [M]. Colorado: Roberts and Company publishers, 2005.
  - [24] Khare K, Butola M, Rajora S. Fourier optics and computational imaging[M]. Beijing: Science Press, 2015.
  - [25] Beversluis M R, Novotny L, Stranick S J. *Programmable vector point-spread function engineering* [J]. Optics Express, 2006, 14(7): 2650-2656.
  - [26] Shechtman Y, Weiss L E, Backer A S, et al. *Multicolour localization microscopy by point-spread-function engineering*[J]. Nature Photonics, 2016, 10(9): 590-594.
  - [27] Wen G, Li S M, Wang L B, et al. *High-fidelity structured illumination microscopy by point-spread-function engineering*[J]. Light: Science & Applications, 2021, 10: 70.
  - [28] Shimobaba T, Matsushima K, Kakue T, et al. *Scaled angular spectrum method*[J]. Optics Letters, 2012, 37 (19): 4128-4130.
  - [29] Lin X, Rivenson Y, Yardimci N T, et al. *All-optical machine learning using diffractive deep neural networks* [J]. Science, 2018, 361(6406): 1004-1008.
  - [30] Luo Y, Meng D, Yardimci N T, et al. *Design of task-specific optical systems using broadband diffractive neural networks*[J]. Light: Science & Applications, 2019, 8: 112.
  - [31] Meng D, Rivenson Y, Ozcan A. *Scale-, shift-, and rotation-invariant diffractive optical networks*[J]. ACS Photonics, 2021, 8(1): 324-334.
  - [32] Zheng M J, Shi L, et al. *Optimize performance of a diffractive neural network by controlling the Fresnel number*[J]. Photonics Research, 2022, 10(11): 2667-2676.
  - [33] Zheng M J, Liu W Z, Shi L, et al. *Diffractive neural*

- networks with improved expressive power for gray-scale image classification[J]. *Photonics Research*, 2024, 12(6): 1159-1166.
- [34] Yan T, Wu J M, Zhou T K, et al. **Fourier-space diffractive deep neural network**[J]. *Physical Review Letters*, 2019, 123(2): 023901.
- [35] Miscuglio M, Hu Z B, Li S R, et al. **Massively parallel amplitude-only Fourier neural network**[J]. *Optica*, 2020, 7(12): 1812-1819.
- [36] Wu Q H, Sui X B, Fei Y H, et al. **Multi-layer optical Fourier neural network based on the convolution theorem** [J]. *AIP Advances*, 2021, 11(5): 055012.
- [37] 郝慧捷, 王新伟, 刘俭, 等. 基于超构表面的光学计算与先进成像(特邀)[J]. *激光与光电子学进展*, 2024, 61(16): 1611008.  
Hao H J, Wang X W, Liu J, et al. Optical computation and advanced imaging based on metasurface (invited)[J]. *Laser & Optoelectronics Progress*, 2024, 61(16): 1611008.
- [38] Sroor H, Huang Y W, Sephton B, et al. **High-purity orbital angular momentum states from a visible metasurface laser**[J]. *Nature Photonics*, 2020, 14(8): 498-503.
- [39] Lin R J, Su V C, Wang S M, et al. **Achromatic metalens array for full-colour light-field imaging**[J]. *Nature Nanotechnology*, 2019, 14(3): 227-231.
- [40] Wang S M, Wu P C, Su V C, et al. **A broadband achromatic metalens in the visible**[J]. *Nature Nanotechnology*, 2018, 13(3): 227-232.
- [41] Decker M, Staude I, Falkner M, et al. **High-efficiency dielectric Huygens' surfaces**[J]. *Advanced Optical Materials*, 2015, 3(6): 813-820.
- [42] Chong K E, Staude I, James A, et al. **Polarization-independent silicon metadevices for efficient optical wavefront control**[J]. *Nano Letters*, 2015, 15(8): 5369-5374.
- [43] Wan X, Jia S L, Cui T J, et al. **Independent modulations of the transmission amplitudes and phases by using Huygens metasurfaces**[J]. *Scientific Reports*, 2016, 6: 25639.
- [44] Kuznetsov A I, Miroshnichenko A E, Brongersma M L, et al. **Optically resonant dielectric nanostructures**[J]. *Science*, 2016, 354(6314): aag2472.
- [45] Yu N F, Genevet P, Kats M A, et al. **Light propagation with phase discontinuities: generalized laws of reflection and refraction**[J]. *Science*, 2011, 334(6054): 333-337.
- [46] Ding X M, Monticone F, Zhang K, et al. **Ultrathin Pancharatnam-Berry metasurface with maximal cross-polarization efficiency**[J]. *Advanced Materials*, 2015, 27(7): 1195-1200.
- [47] Zhang Z Y, Liang H G, He T, et al. **Photonic spin Hall effect based on broadband high-efficiency reflective metasurfaces**[J]. *Applied Optics*, 2020, 59(5): A63-A68.
- [48] Wan W W, Gao J, Yang X D. **Full-color plasmonic metasurface holograms**[J]. *ACS Nano*, 2016, 10(12): 10671-10680.
- [49] Wang S M, Wu P C, Su V C, et al. **Broadband achromatic optical metasurface devices**[J]. *Nature Communications*, 2017, 8: 187.
- [50] Arbabi E, Arbabi A, Kamali S M, et al. **Multiwavelength polarization-insensitive lenses based on dielectric metasurfaces with meta-molecules**[J]. *Optica*, 2016, 3(6): 628-633.
- [51] Arbabi A, Arbabi E, Kamali S M, et al. **Miniature optical planar camera based on a wide-angle metasurface doublet corrected for monochromatic aberrations**[J]. *Nature Communications*, 2016, 7: 13682.
- [52] Zang X F, Dong F L, Yue F Y, et al. **Polarization encoded color image embedded in a dielectric metasurface** [J]. *Advanced Materials*, 2018, 30(21): 1707499.
- [53] Yue F Y, Zhang C M, Zang X F, et al. **High-resolution grayscale image hidden in a laser beam**[J]. *Light: Science & Applications*, 2017, 7(1): 17129.
- [54] Goodman J W, Dias A R, Woody L M. **Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms**[J]. *Optics Letters*, 1978, 2(1): 1-3.
- [55] LensletLabs. **EnLight 256 white paper report**[EB/OL]. [2022-09-01]. <http://besho.narod.ru/reviews/newage/EnLight256.pdf>.
- [56] Young T, Hazarika D, Poria S, et al. **Recent trends in deep learning based natural language processing [review article]**[J]. *IEEE Computational Intelligence Magazine*, 2018, 13(3): 55-75.
- [57] Cong J, Xiao B J. **Minimizing computation in convolutional neural networks**[M]//Wermter S, Weber C, Duch W, et al. *Artificial neural networks and machine learning-ICANN 2014. Lecture notes in computer science*. Cham: Springer International Publishing, 2014, 8681: 281-290.
- [58] Hu J T, Mengü D, Tzarouchis D C, et al. **Diffractive optical computing in free space**[J]. *Nature Communications*, 2024, 15: 1525.
- [59] Fu T Z, Zhang J F, Sun R, et al. **Optical neural networks: progress and challenges**[J]. *Light: Science & Applications*, 2024, 13: 263.
- [60] Kulce O, Mengü D, Rivenson Y, et al. **All-optical information-processing capacity of diffractive surfaces**[J]. *Light: Science & Applications*, 2021, 10: 25.
- [61] Rahman M S S, Li J X, Mengü D, et al. **Ensemble learning of diffractive optical networks**[J]. *Light: Science & Applications*, 2021, 10: 14.
- [62] Li J X, Mengü D, Yardimci N T, et al. **Spectrally encoded single-pixel machine vision using diffractive networks**[J]. *Science Advances*, 2021, 7(13): eabd7690.
- [63] Kulce O, Mengü D, Rivenson Y, et al. **All-optical synthesis of an arbitrary linear transformation using diffractive surfaces**[J]. *Light: Science & Applications*, 2021, 10: 196.
- [64] Mengü D, Zhao Y F, Tabassum A, et al. **Diffractive interconnects: all-optical permutation operation using diffractive networks**[J]. *Nanophotonics*, 2022, 12(5): 905-923.
- [65] Rahman M S S, Yang X L, Li J X, et al. **Universal linear intensity transformations using spatially incoherent**

- diffractive processors[J]. *Light: Science & Applications*, 2023, 12: 195.
- [66] Cai Y J, Chen Y H, Wang F. Generation and propagation of partially coherent beams with nonconventional correlation functions: a review [invited][J]. *Journal of the Optical Society of America A*, 2014, 31(9): 2083-2096.
- [67] Liu Y, Qin J W, Liu Y B, et al. Optical Fourier convolutional neural network with high efficiency in image classification[J]. *Optics Express*, 2024, 32(13): 23575-23583.
- [68] Qin J W, Liu Y B, Liu Y, et al. All-optical Fourier neural network using partially coherent light[J]. *Chip*, 2025, 1: 100140.
- [69] 符庭钊, 孙润, 黄禹尧, 等. 片上集成光学神经网络综述(特邀)[J]. *中国激光*, 2024, 51(1): 0119002.
- Fu T Z, Sun R, Huang Y Y, et al. Review of on-chip integrated optical neural networks (invited)[J]. *Chinese Journal of Lasers*, 2024, 51(1): 0119002.
- [70] 邓莉, 高仁宏, 管江林, 等. 薄膜铌酸锂微纳结构的非线性光学研究进展(特邀)[J]. *激光与光电子学进展*, 2024, 61(11): 1116011.
- Deng L, Gao R H, Guan J L, et al. Research progress in nonlinear optics of thin-film lithium niobate micro/nano structures (invited)[J]. *Laser & Optoelectronics Progress*, 2024, 61(11): 1116011.
- [71] Hu Y W, Zhu D, Lu S Y, et al. Integrated electro-optics on thin-film lithium niobate[J]. *Nature Reviews Physics*, 2025, 7(5): 237-254.
- [72] 程亚. 薄膜铌酸锂光电器件与超大规模光子集成(特邀)[J]. *中国激光*, 2024, 51(1): 0119001.
- Cheng Y. Thin film lithium niobate electro-optic devices and ultralarge-scale photonic integration (invited)[J]. *Chinese Journal of Lasers*, 2024, 51(1): 0119001.
- [73] Xie Y W, Wu J C, Hong S H, et al. Towards large-scale programmable silicon photonic chip for signal processing[J]. *Nanophotonics*, 2024, 13(12): 2051-2073.
- [74] 邱辞源, 马宁, 陈雨, 等. 硅基可编程多功能光子芯片研究进展和展望(特邀)[J]. *激光与光电子学进展*, 2024, 61(19): 1913013.
- Qiu C Y, Ma N, Chen Y, et al. Recent advances in and prospects of programmable silicon multifunctional integrated photonic chips (invited)[J]. *Laser & Optoelectronics Progress*, 2024, 61(19): 1913013.
- [75] Yin Y X, Zhang X P, Yin X J, et al. High-Q-factor tunable silica-based microring resonators[J]. *Photonics*, 2021, 8(7): 256.
- [76] Tait A N, Wu A X, de Lima T F, et al. Microring weight banks[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2016, 22(6): 5900214.
- [77] Tait A N, Jayatilaka H, De Lima T F, et al. Feedback control for microring weight banks[J]. *Optics Express*, 2018, 26(20): 26422-26443.
- [78] Xu Q F, Soref R. Reconfigurable optical directed-logic circuits using microresonator-based optical switches[J]. *Optics Express*, 2011, 19(6): 5244-5259.
- [79] Pérez D, Gasulla I, Capmany J. Programmable multifunctional integrated nanophotonics[J]. *Nanophotonics*, 2018, 7(8): 1351-1371.
- [80] Yang L, Ji R Q, Zhang L, et al. On-chip CMOS-compatible optical signal processor[J]. *Optics Express*, 2012, 20(12): 13560-13565.
- [81] Qiu C Y, Gao W L, Soref R, et al. Reconfigurable electro-optical directed-logic circuit using carrier-depletion micro-ring resonators[J]. *Optics Letters*, 2014, 39(24): 6767-6770.
- [82] Pérez-López D, Gutierrez A M, Sánchez E, et al. Integrated photonic tunable basic units using dual-drive directional couplers[J]. *Optics Express*, 2019, 27(26): 38071-38086.
- [83] Lomonte E, Wolff M A, Beutel F, et al. Single-photon detection and cryogenic reconfigurability in lithium niobate nanophotonic circuits[J]. *Nature Communications*, 2021, 12: 6847.
- [84] 郑远林, 陈险峰. 薄膜铌酸锂集成非线性光学: 走向全光信息时代的新路径[J]. *物理*, 2024, 53(1): 22-32.
- Zheng Y L, Chen X F. Integrated nonlinear photonics on thin-film lithium niobate: a route to an all-optical information era[J]. *Physics*, 2024, 53(1): 22-32.
- [85] Qiu C Y, Ye X, Soref R, et al. Demonstration of reconfigurable electro-optical logic with silicon photonic integrated circuits[J]. *Optics Letters*, 2012, 37(19): 3942-3944.
- [86] Tian Y H, Liu Z L, Xiao H F, et al. Experimental demonstration of a reconfigurable electro-optic directed logic circuit using cascaded carrier-injection micro-ring resonators[J]. *Scientific Reports*, 2017, 7: 6410.
- [87] Yan Q Q, Ouyang H, Tao Z L, et al. Multi-wavelength optical information processing with deep reinforcement learning[J]. *Light: Science & Applications*, 2025, 14: 160.
- [88] Zhang M, Wang C, Cheng R, et al. Monolithic ultra-high-Q lithium niobate microring resonator[J]. *Optica*, 2017, 4(12): 1536-1537.
- [89] Zhang J H, Fang Z W, Lin J T, et al. Fabrication of crystalline microresonators of high quality factors with a controllable wedge angle on lithium niobate on insulator[J]. *Nanomaterials*, 2019, 9(9): 1218.
- [90] Ying Z F, Zhao Z, Feng C H, et al. Automated logic synthesis for electro-optic logic-based integrated optical computing[J]. *Optics Express*, 2018, 26(21): 28002-28012.
- [91] Ying Z F, Feng C H, Zhao Z, et al. Integrated multi-operand electro-optic logic gates for optical computing[J]. *Applied Physics Letters*, 2019, 115(17): 171104.
- [92] Yuan M R, Li Y Y, Xiao H F, et al. Reconfigurable optical directed logic circuits based on mode division multiplexing technology[J]. *IEEE Photonics Journal*, 2023, 15(3): 6601107.
- [93] Ying Z F, Dhar S, Zhao Z, et al. Electro-optic ripple-carry adder in integrated silicon photonics for optical computing[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2018, 24(6): 7600310.
- [94] Zhang W F, Yao J P. Photonic integrated field-programmable disk array signal processor[J]. *Nature*

- Communications, 2020, 11: 406.
- [95] Yi D, Wang Y, Tsang H K. Multi-functional photonic processors using coherent network of micro-ring resonators[J]. APL Photonics, 2021, 6(10): 100801.
  - [96] Yao Y H, Wei Y X, Dong J J, et al. Large-scale reconfigurable integrated circuits for wideband analog photonic computing[J]. Photonics, 2023, 10(3): 300.
  - [97] Ouyang H, Tao Z L, You J, et al. Computing dimension for a reconfigurable photonic tensor processing core based on silicon photonics[J]. Optics Express, 2024, 32(18): 31205-31219.
  - [98] Reck M, Zeilinger A, Bernstein H J, et al. Experimental realization of any discrete unitary operator[J]. Physical Review Letters, 1994, 73(1): 58-61.
  - [99] Clements W R, Humphreys P C, Metcalf B J, et al. Optimal design for universal multiport interferometers[J]. Optica, 2016, 3(12): 1460-1465.
  - [100] Ribeiro A, Ruocco A, Vanacker L, et al. Demonstration of a  $4 \times 4$ -port universal linear circuit[J]. Optica, 2016, 3(12): 1348-1357.
  - [101] Shen Y C, Harris N C, Skirlo S, et al. Deep learning with coherent nanophotonic circuits[J]. Nature Photonics, 2017, 11(7): 441-446.
  - [102] Hughes T W, Minkov M, Shi Y, et al. Training of photonic neural networks through in situ backpropagation and gradient measurement[J]. Optica, 2018, 5(7): 864-871.
  - [103] Zhang H, Gu M, Jiang X D, et al. An optical neural chip for implementing complex-valued neural network[J]. Nature Communications, 2021, 12: 457.
  - [104] Zhu H H, Zou J, Zhang H, et al. Space-efficient optical computing with an integrated chip diffractive neural network[J]. Nature Communications, 2022, 13: 1044.
  - [105] Zheng Y, Zhong H Z, Zhang H S, et al. Electro-optically programmable photonic circuits enabled by wafer-scale integration on thin-film lithium niobate[J]. Physical Review Research, 2023, 5(3): 033206.
  - [106] Wu B, Liu S J, Cheng J W, et al. Real-valued optical matrix computing with simplified MZI mesh[J]. Intelligent Computing, 2023, 2: 47.
  - [107] Wang X Y, Xie P, Chen B H, et al. Chip-based high-dimensional optical neural network[J]. Nano-Micro Letters, 2022, 14(1): 221.
  - [108] Xu Z H, Zhou T K, Ma M Z, et al. Large-scale photonic chiplet Taichi empowers 160-TOPS/W artificial general intelligence[J]. Science, 2024, 384(6692): 202-209.
  - [109] Zarei S, Marzban M R, Khavasi A. Integrated photonic neural network based on silicon metalines[J]. Optics Express, 2020, 28(24): 36668-36684.
  - [110] Fu T Z, Zang Y B, Huang H H, et al. On-chip photonic diffractive optical neural network based on a spatial domain electromagnetic propagation model[J]. Optics Express, 2021, 29(20): 31924-31940.
  - [111] Wang Z, Chang L, Wang F F, et al. Integrated photonic metasystem for image classifications at telecommunication wavelength[J]. Nature Communications, 2022, 13: 2131.
  - [112] Fu T Z, Zang Y B, Huang Y Y, et al. Photonic machine learning with on-chip diffractive optics[J]. Nature Communications, 2023, 14: 70.
  - [113] Liu W C, Fu T Z, Huang Y Y, et al. C-DONN: compact diffractive optical neural network with deep learning regression[J]. Optics Express, 2023, 31(13): 22127-22143.
  - [114] Moughames J, Porte X, Thiel M, et al. Three-dimensional waveguide interconnects for scalable integration of photonic neural networks[J]. Optica, 2020, 7(6): 640-646.
  - [115] Feldmann J, Youngblood N, Karpov M, et al. Parallel convolutional processing using an integrated photonic tensor core[J]. Nature, 2021, 589(7840): 52-58.
  - [116] Zhou W, Dong B W, Farmakidis N, et al. In-memory photonic dot-product engine with electrically programmable weight banks[J]. Nature Communications, 2023, 14: 2887.
  - [117] Dong B W, Aggarwal S, Zhou W, et al. Higher-dimensional processing using a photonic tensor core with continuous-time data[J]. Nature Photonics, 2023, 17(12): 1080-1088.
  - [118] Zhang S J, Zhou H J, Wu B, et al. Redundancy-free integrated optical convolver for optical neural networks based on arrayed waveguide grating[J]. Nanophotonics, 2024, 13(1): 19-28.
  - [119] Wang Z, Fang Z W, Liu Z X, et al. On-chip arrayed waveguide grating fabricated on thin-film lithium niobate [J]. Advanced Photonics Research, 2024, 5(2): 2470005.
  - [120] Dubey S R, Singh S K, Chaudhuri B B. Activation functions in deep learning: a comprehensive survey and benchmark[J]. Neurocomputing, 2022, 503: 92-108.
  - [121] Destras O, Le Beux S, de Magalhães F G, et al. Survey on activation functions for optical neural networks[J]. ACM Computing Surveys, 2024, 56(2): 35.
  - [122] Wang T Y, Sohoni M M, Wright L G, et al. Image sensing with multilayer nonlinear optical neural networks [J]. Nature Photonics, 2023, 17(5): 408-415.
  - [123] Ashtiani F, Idjadi M H, Hu T C, et al. A surface-normal photodetector as nonlinear activation function in diffractive optical neural networks[J]. APL Photonics, 2023, 8(12): 121301.
  - [124] Zhang D H, Xu D, Li Y H, et al. Broadband nonlinear modulation of incoherent light using a transparent optoelectronic neuron array[J]. Nature Communications, 2024, 15: 2433.
  - [125] Guo B, Xiao Q L, Wang S H, et al. 2D layered materials: synthesis, nonlinear optical properties, and device applications[J]. Laser & Photonics Reviews, 2019, 13(12): 1800327.
  - [126] Zuo Y, Li B H, Zhao Y J, et al. All-optical neural network with nonlinear activation functions[J]. Optica, 2019, 6(9): 1132-1137.
  - [127] Ma X G. Research on third-order nonlinearities of nanofunctional materials and their application in fast optical holography[D]. Taiyuan: North University of China, 2024.
  - [128] Sun Y C, Dong M L, Yu M X, et al. Modeling and

- simulation of all-optical diffractive neural network based on nonlinear optical materials[J]. *Optics Letters*, 2022, 47(1): 126-129.
- [129] Shi W X, Jiang X, Zheng H, et al. Lensless optoelectronic neural network with quantum dot nonlinear activation[J]. *Photonics Research*, 2024, 12(4): 682-690.
- [130] Yu J, Park S, Hwang I, et al. Electrically tunable nonlinear polaritonic metasurface[J]. *Nature Photonics*, 2021, 16(1): 72-78.
- [131] 张楠, 黄邳祺, 张子安, 等. 光计算和光电智能计算研究进展[J]. *中国激光*, 2024, 51(18): 1800001.  
Zhang N, Huang Z Q, Zhang Z A, et al. Advances of optical computing and optoelectronic intelligent computing[J]. *Chinese Journal of Lasers*, 2024, 51(18): 1800001.
- [132] He M B, Xu M Y, Ren Y X, et al. High-performance hybrid silicon and lithium niobate Mach-Zehnder modulators for 100 Gbit s<sup>-1</sup> and beyond[J]. *Nature Photonics*, 2019, 13(5): 359-364.
- [133] Xu M Y, Zhu Y T, Pittalà F, et al. Dual-polarization thin-film lithium niobate in-phase quadrature modulators for terabit-per-second transmission[J]. *Optica*, 2022, 9(1): 61-62.
- [134] George J K, Mehrabian A, Amin R, et al. Neuromorphic photonics with electro-absorption modulators[J]. *Optics Express*, 2019, 27(4): 5181-5191.
- [135] Amin R, George J K, Sun S, et al. ITO-based electro-absorption modulator for photonic neural activation function[J]. *APL Materials*, 2019, 7(8): 081112.
- [136] Zhong C Y, Liao K, Dai T X, et al. Graphene/silicon heterojunction for reconfigurable phase-relevant activation function in coherent optical neural networks[J]. *Nature Communications*, 2023, 14: 6939.
- [137] Chen Y T, Nazhamaiti M, Xu H, et al. All-analog photoelectronic chip for high-speed vision tasks[J]. *Nature*, 2023, 623(7985): 48-57.
- [138] Miscuglio M, Mehrabian A, Hu Z B, et al. All-optical nonlinear activation function for photonic neural networks[J]. *Optical Materials Express*, 2018, 8(12): 3851-3863.
- [139] Jha A, Huang C R, Prucnal P R. Reconfigurable all-optical nonlinear activation functions for neuromorphic photonics[J]. *Optics Letters*, 2020, 45(17): 4819-4822.
- [140] Guo Q S, Sekine R, Ledezma L, et al. Femtojoule femtosecond all-optical switching in lithium niobate nanophotonics[J]. *Nature Photonics*, 2022, 16(9): 625-631.
- [141] Wu B, Li H K, Tong W Y, et al. Low-threshold all-optical nonlinear activation function based on a Ge/Si hybrid structure in a microring resonator[J]. *Optical Materials Express*, 2022, 12(3): 970-980.
- [142] Zhang W K, Gu W T, Cheng J W, et al. Performing photonic nonlinear computations by linear operations in a high-dimensional space[J]. *Nanophotonics*, 2023, 12(15): 3189-3197.
- [143] Wu T W, Li Y K, Ge L, et al. Field-programmable photonic nonlinearity[J/OL]. *Nature Photonics*: 1-8[2025-05-06]. <https://www.nature.com/articles/s41566-025-01660-x>.
- [144] Ashtiani F, Geers A J, Aflatouni F. An on-chip photonic deep neural network for image classification[J]. *Nature*, 2022, 606(7914): 501-506.
- [145] Tan S S, Wu Z, Lei L, et al. All-optical computation system for solving differential equations based on optical intensity differentiator[J]. *Optics Express*, 2013, 21(6): 7008-7013.
- [146] Xu D Y, Yang H, Xu W H, et al. Inverse design of Pancharatnam-Berry phase metasurfaces for all-optical image edge detection[J]. *Applied Physics Letters*, 2022, 120(24): 241101.
- [147] Liu Y, Huang M C, Chen Q K, et al. Single planar photonic chip with tailored angular transmission for multiple-order analog spatial differentiator[J]. *Nature Communications*, 2022, 13: 7944.
- [148] Huo P C, Tan L, Jin Y Q, et al. Broadband and parallel multiple-order optical spatial differentiation enabled by Bessel vortex modulated metalens[J]. *Nature Communications*, 2024, 15: 9045.
- [149] Deng M, Cotrufo M, Wang J, et al. Broadband angular spectrum differentiation using dielectric metasurfaces[J]. *Nature Communications*, 2024, 15: 2237.
- [150] Xu D Y, Xu W H, Yang Q, et al. All-optical object identification and three-dimensional reconstruction based on optical computing metasurface[J]. *Opto-Electronic Advances*, 2023, 6(12): 230120.
- [151] Zhou J X, Wu Q Y, Zhao J X, et al. Fourier optical spin splitting microscopy[J]. *Physical Review Letters*, 2022, 129(2): 020801.
- [152] Chen H, Jayasuriya S, Yang J Y, et al. ASP vision: optically computing the first layer of convolutional neural networks using angle sensitive pixels[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 903-912.
- [153] Chang J L, Sitzmann V, Dun X, et al. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification[J]. *Scientific Reports*, 2018, 8: 12324.
- [154] Xu X Y, Tan M X, Corcoran B, et al. 11 TOPS photonic convolutional accelerator for optical neural networks[J]. *Nature*, 2021, 589(7840): 44-51.
- [155] Shi W X, Huang Z, Huang H H, et al. LOEN: lensless optoelectronic neural network empowered machine vision[J]. *Light: Science & Applications*, 2022, 11: 121.
- [156] Xu S F, Wang J, Yi S C, et al. High-order tensor flow processing using integrated photonic circuits[J]. *Nature Communications*, 2022, 13: 7970.
- [157] Ouyang H, Zhao Z Y, Tao Z L, et al. Parallel edge extraction operators on chip speed up photonic convolutional neural networks[J]. *Optics Letters*, 2024, 49(4): 838-841.
- [158] Yu H Y, Huang Z H, Lamon S, et al. All-optical image transportation through a multimode fibre using a miniaturized diffractive neural network on the distal facet[J]. *Nature Photonics*, 2025, 19(5): 486-493.
- [159] Tanaka G, Yamane T, Héroux J B, et al. Recent

- advances in physical reservoir computing: a review[J]. Neural Networks, 2019, 115: 100-123.
- [160] Hasegawa H, Kanno K, Uchida A. Parallel and deep reservoir computing using semiconductor lasers with optical feedback[J]. Nanophotonics, 2022, 12(5): 869-881.
- [161] Li R Q, Shen Y W, Lin B D, et al. Scalable wavelength-multiplexing photonic reservoir computing[J]. APL Machine Learning, 2023, 1(3): 036105.
- [162] Shen Y W, Li R Q, Liu G T, et al. Deep photonic reservoir computing recurrent network[J]. Optica, 2023, 10(12): 1745-1751.
- [163] Pierangeli D, Marcucci G, Conti C. Large-scale photonic Ising machine by spatial light modulation[J]. Physical Review Letters, 2019, 122(21): 213902.
- [164] Fang Y S, Huang J Y, Ruan Z C. Experimental observation of phase transitions in spatial photonic Ising machine[J]. Physical Review Letters, 2021, 127(4): 043902.
- [165] Luo L, Mi Z Y, Huang J Y, et al. Wavelength-division multiplexing optical Ising simulator enabling fully programmable spin couplings and external magnetic fields[J]. Science Advances, 2023, 9(48): eadg6238.
- [166] Wu B, Zhang W K, Zhang S J, et al. A monolithically integrated optical Ising machine[J]. Nature Communications, 2025, 16: 4296.
- [167] Wright L G, Onodera T, Stein M M, et al. Deep physical neural networks trained with backpropagation[J]. Nature, 2022, 601(7894): 549-555.
- [168] Momeni A, Rahmani B, Malléjac M, et al. Backpropagation-free training of deep physical neural networks[J]. Science, 2023, 382(6676): 1297-1303.
- [169] Oguz I, Ke J J, Weng Q F, et al. Forward-forward training of an optical neural network[J]. Optics Letters, 2023, 48(20): 5249-5252.
- [170] Tang S J, Chen C, Qin Q, et al. Reconfigurable integrated photonic unitary neural networks with phase encoding enabled by in situ training[J]. IEEE Photonics Journal, 2024, 16(5): 8800311.
- [171] Liu W C, Huang Y Y, Sun R, et al. Ultra-compact multi-task processor based on in-memory optical computing[J]. Light: Science & Applications, 2025, 14: 134.
- [172] 毕岩峰, 吴星宇, 张璐矾, 等. 模拟光计算的发展与应用[J]. 科学通报, 2024, 69(34): 5028-5041.
- Bi Y F, Wu X Y, Zhang L F, et al. Development and applications of analog optical computing: a review[J]. Chinese Science Bulletin, 2024, 69(34): 5028-5041.
- [173] 马国庆, 周常河, 朱镕威, 等. 光计算的发展趋势: 模拟或数字?[J]. 中国激光, 2023, 50(5): 0500001.
- Ma G Q, Zhou C H, Zhu R W, et al. Future of optical computing: analog or digital? [J]. Chinese Journal of Lasers, 2023, 50(5): 0500001.
- [174] Hua S Y, Divita E, Yu S S, et al. An integrated large-scale photonic accelerator with ultralow latency[J]. Nature, 2025, 640(8058): 361-367.
- [175] Ahmed S R, Baghdadi R, Bernadskiy M, et al. Universal photonic artificial intelligence acceleration[J]. Nature, 2025, 640(8058): 368-374.
- [176] NVIDIA. NVIDIA H100 Tensor Core GPU datasheet [EB/OL]. [2025-04-05]. <https://resources.nvidia.com/en-us-tensor-core/nvidiatensor-core-gpu-datasheet>.
- [177] NVIDIA. NVIDIA H100 Tensor Core GPU architecture overview[EB/OL]. [2025-04-05]. <https://resources.nvidia.com/en-us-tensor-core/gtc22-whitepaper-hopper>.
- [178] Hennessy J L, Patterson D A. Computer architecture: a quantitative approach[M]. Amsterdam: Elsevier, 2011.
- [179] Chen X Z, Xue Y, Sun Y B, et al. Neuromorphic photonic memory devices using ultrafast, non-volatile phase-change materials[J]. Advanced Materials, 2023, 35(37): e2203909.
- [180] Cheng Z G, Rios C A, Youngblood N, et al. On-chip phase-change photonic memory and computing[J]. Proceedings of SPIE, 2017, 10345: 1034519.
- [181] Wu N F, Sun Y X, Hu J T, et al. Intelligent nanophotonics: when machine learning sheds light[J]. eLight, 2025, 5(1): 5.