中国激光

光计算的发展趋势：模拟或数字？

马国庆[1,2]，周常河[3*]，朱镕威[1,2]，郑奉禄[1,2]，余俊杰[1,2**]，司徒国海[1,2***]

[1]中国科学院上海光学精密机械研究所信息光学与光电技术实验室，上海 201800；

[2]中国科学院大学材料与光电学院，北京 100049；

[3]暨南大学光子技术研究院，广东 广州 510632

摘要 受益于光子独特的优势，光计算技术在构建高速、高算力和高能效比的专用计算加速器方面被寄予厚望，目前已经涌现出了许多极具吸引力的方案。特别是对于涉及运算量巨大的二维矩阵-矩阵乘加操作的专用场景，光计算有望在算力和能效比等方面实现超越当前最先进电子计算机几个数量级的性能提升。不同于电子计算通过构建逻辑门实现通用数字计算，主要受深度学习驱动而复兴的光计算更倾向于模拟计算。本文从模拟和数字光计算的角度出发对主流的光计算架构进行分析和讨论，指出了目前光计算技术发展面临的瓶颈，并对光计算未来的发展趋势进行了展望。

关键词 光计算；模拟光计算；数字光计算；光计算架构；光学矩阵计算；光学神经网络；光电智能计算；光学信号处理

中图分类号 TN45 **文献标志码** A **DOI：** 10.3788/CJL221209

1 引 言

深度学习已经成为智能时代的强大驱动力，在计算机视觉[1-2]、语音识别[3]、自然语言处理[4]等方面得到了广泛应用。然而，深度学习算法[5-6]中80%以上的计算都是矩阵-矩阵乘加操作(MM-MAC)[7]，大规模的MM-MAC运算转换为CPU可执行代码时会导致大量访存需求。受限于电子计算机“存算分离”架构体系和芯片上铜线互连极限的物理制约（如图1所示），深度神经网络(DNNs)的训练效率和速度受到严重影响。根据de Lima等[8]的报道，训练最先进的DNNs所需的算力约每3.5个月就要翻一番，远远超出了遵循摩尔定律发展的电子集成电路(EIC)的算力供给。

目前，解决“后摩尔时代”电子计算机在算力和能耗方面瓶颈的主要技术路径可分为三类。第一类通过先进制程技术进一步缩小电子逻辑器件的尺寸从而延续摩尔定律，包括极紫外(EUV)光刻机、鳍式场效应型晶体管(FinFET)、环绕式栅极晶体管(GAA)以及可实现1 nm工艺的二维材料晶体管[9-10]等技术；第二类通过高级封装方案将多个芯片异质集成到一起以提高系统的整体性能，如光互连[11]、2.5D/3D封装、小芯片及其互连协议UCIe等方法；第三类是超越传统CMOS技术开发的具有高算力和高能效比的新型计算体系，如光（电）子计算、碳基计算[12]和通用量子计算[13]等。

尽管图形处理器(GPU)、张量处理器(TPU)、专用芯片(ASIC)和感存算一体等电子硬件加速方案在深度学习中已经得到广泛应用，但前两条路径仍然是在传统电子芯片体系内进行改进，且受限于有限的互连密度，难以实现单位面积内算力的数量级提升。根据Amdahl's定律[14]，随着并行计算单元数目的不断增加，系统加速比将仅取决于问题规模中串行分量的比例。因此，CPU固定的时钟速率和“存算分离”架构从根本上限制着电子硬件加速器的计算效率。近年来，高频集成电路和微波光子技术的结合有望实现高频信号处理，但是阻抗匹配和速度匹配的要求进一步增加了系统设计的难度和系统复杂性。第三种路径则代表了更为革新、更具前景的技术途径。总的来看，碳基计算或通用量子计算离真正实用化还有很长的路要走[15-17]，而光子计算或光电混合计算则是当前最有望解决算力供给和低功耗数据处理等难题并得到实际应用的技术途径。

相比于传统的电子计算，光计算得益于光子本身高度并行、高速低功耗等特点有望构建具有高算力和高能效比的深度学习加速器[19-20]。目前，各种光计算架构[21-22]已经初步展示了它们在高算力和高能效比方

收稿日期：2022-09-02；**修回日期**：2022-09-26；**录用日期**：2022-10-11；**网络首发日期**：2022-10-21

基金项目：中国科学院前沿科学重点项目(QYZDJ-SSW-JSC014)、上海市自然科学基金(19JC1415400、19DZ2291102、20ZR1464700)

通信作者：*chazhou@mail.shcnc.ac.cn；**Junjiey@siom.ac.cn；***ghsitu@siom.ac.cn

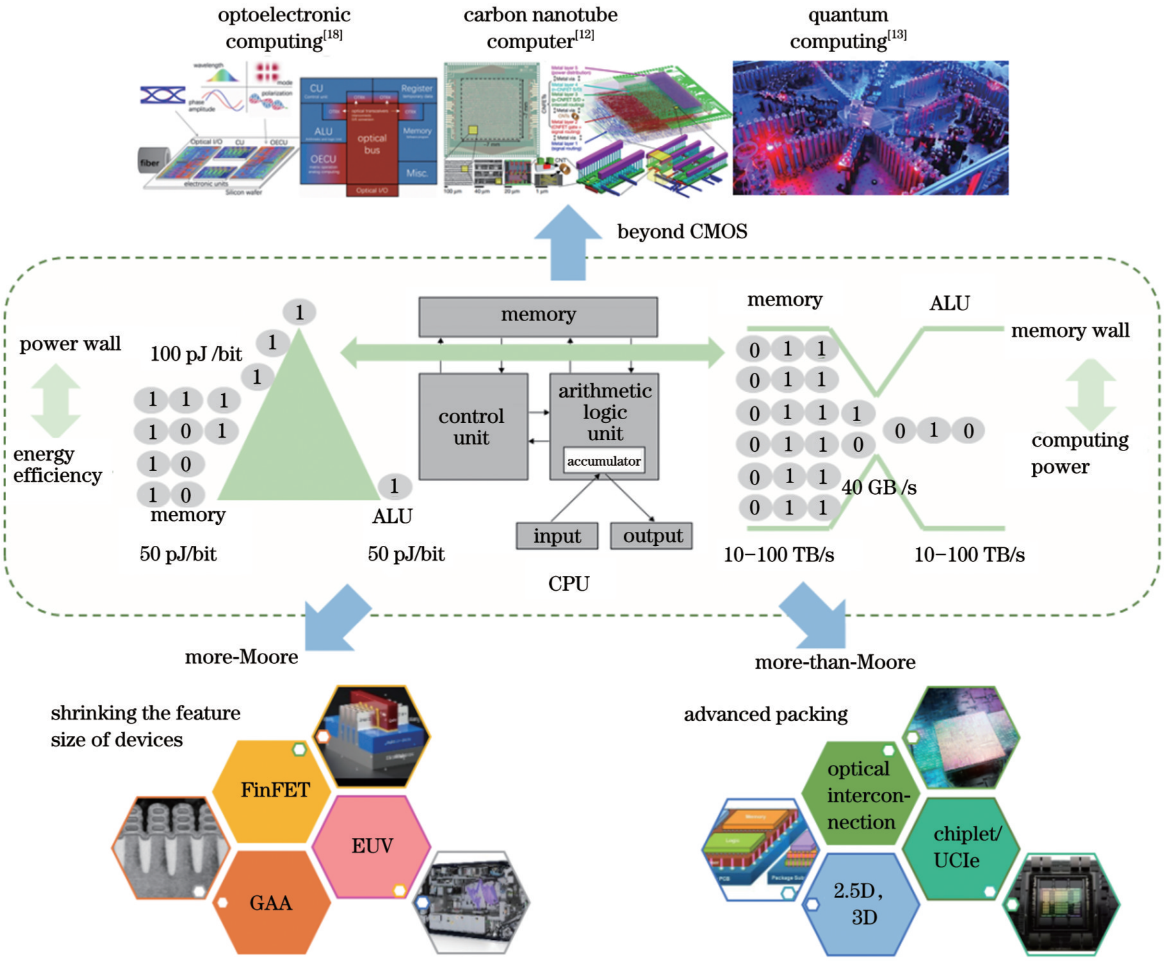

图 1　“后摩尔时代”提升算力和能效比的解决路径(图中 FinFET 为场效应型晶体管,EUV 为极紫外光刻机,GAA 为环绕式栅极晶体管,UCIe 为芯片互连协议,chiplet 为芯粒)

Fig. 1　Methods to improve computing power and energy efficiency ratio in the “post-Moore era” (FinFET is field effect transistor, EUV is extreme ultraviolet lithography, GAA is gate-all-around FET, UCIe is universal chiplet interconnect express, and chiplet is an integrated circuit block)

面的优势,其发展路线大致可以分为两种(如图 2 所示):一种是从光学角度出发基于多维光场信号调制实现某种专用的光学信息处理,如乘加运算(MAC)[23]、卷积/相关[24]、微分/积分[25-26]、傅里叶变换[27]、光学神经网络(ONNs)[28-32]等,通常以模拟光学计算为主;另一种则是从计算机角度出发借鉴电子计算机的概念实现数字光学计算[33-34],如光学晶体管[35]、光学逻辑器件[36-37]、导向逻辑[38-39]、时空编码[40]、三值光计算机(TOC)[41]等。除此之外,光互连[42]、光电共封装[43]、光电异质集成[44]、三维堆叠[45]等支撑技术已经被广泛用于提升电子计算机的性能[46]。

由于缺乏高效可靠的弱光非线性相互作用和光学逻辑器件,目前的光计算技术主要以线性模拟计算为主。此外,由于缺乏真正意义上的光子信息存储手段,全光信号处理仍然难以实现,即光子不能独立完成存算/算存的完整过程。因此,从电子信息存储到光子信息加载或从光子信息加载到电子信息存储,仍然需要高精度、高速的并行电子控制和模数转换电路配合才能充分发挥光计算技术的全部优势[18,47-49]。模拟光学计算的计算精度虽然仍受限于系统的线性度、动态范围等因素,但已足以用于构建实用的深度学习加速器。同时,通过开发合适的编码方案、并行算法和架构体系,进一步充分利用光学各个维度的并行特性,光子有望在相同的单位面积内实现超越电子的算力密度和能效比[50],并可在对误差敏感的光子系统中实现较高的计算精度。虽然二值电子逻辑计算可以通过足够高的计算精度模拟各种实际的物理场景,但不断增加的计算量将显著增加系统功耗。光子并行计算有望在有限的计算精度下构建类似于人脑的高性能、高能效比的模糊并行计算系统。本文首先总结了模拟和数字光计算技术各自的特点,从模拟和数字光计算角度出发回顾并探讨了光计算在不同发展时期所取得的主要进展及其代表性成果;然后对主要的模拟和数字光计算技术进行分析和讨论,指出了目前光计算发展面临的困境;最后对光计算未来的发展趋势和方向进行了展望。

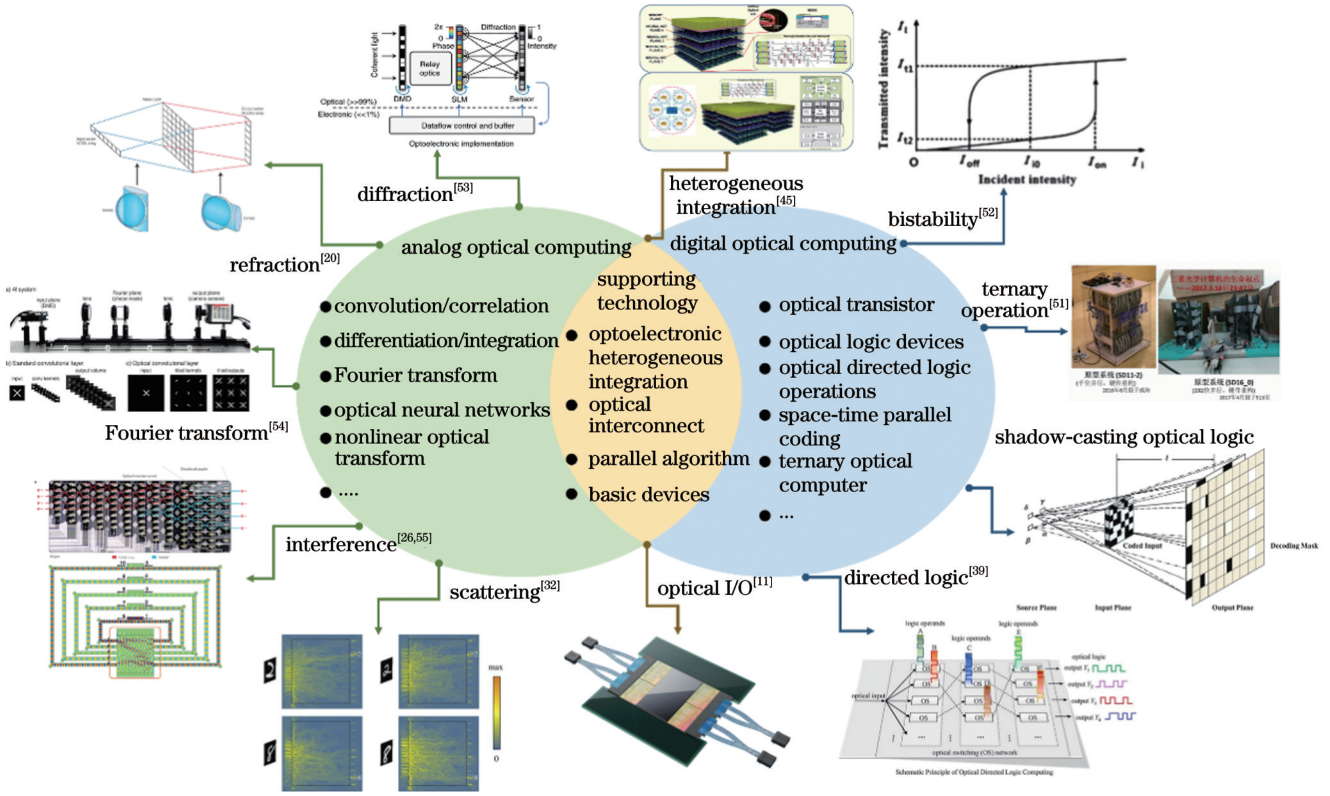

图 2　光计算发展模式及主要技术概览

Fig. 2　Overview of development model and main technologies of optical computing

2　主流光计算技术研究进展及讨论

2.1　模拟和数字光计算的概念及特点

光子计算和电子计算一样都可以分为模拟计算和数字计算两大类。模拟计算是指输入量(自变量)和输出量(计算结果)均以连续信号的形式表示,即通过某种物理效应直接对信号进行特定变换完成某种专用计算,如放大/衰减、相关、卷积和 Hough 变换等线性变换,或者平方、对数、微/积分和阈值处理(ReLu、Sigmoid 等)等非线性变换,最后将处理后的模拟结果输出到探测器中。对于光学模拟计算[56-59]而言,在将输入模拟信号转换为光信号后即可利用光电材料本身优异的光学效应或者微纳结构的高维光学调控能力对输入光场的各维度参数(如振幅、相位、偏振、轨道角动量等)直接进行调制,然后通过探测器测量调制后的结果。

数字计算是指输入量(自变量)和输出量(计算结果)均以离散信号的形式表示,即通过高性能模数转换器将输入信号采样、量化、编码为一定进制的离散量,然后使用一系列多值逻辑器件对离散信号进行处理,最后把处理后的结果输入探测器中。因此,为了提高计算的准确率,减小系统的误码率,电子数字计算机广泛采用二进制,即高低(0 和 1)电平,对应的多值逻辑器件是电子晶体管。光学系统天然具有多态特性,如衍射光学元件具有不同角谱的衍射级次以及不同的偏振特性、不同的轨道角动量等。因此,对于光学数字计算[33-34]可以考虑构建高进制计算系统,如三值光计算[51],即将光强度与偏振方向结合起来表示三值信息,通过液晶的旋光效应和偏振器实现三种光学状态的相互转换;也可以采用二进制,如光学双稳态[52],即一定范围的输入信号对应着两种输出结果,输入信号高于某个值,输出信号就会跳变为高,输入信号低于某个值,输出信号就会跳变为低。对于数字光学计算而言[60],为了消除逻辑计算中频繁出现的进位和移位操作,必须采用更适合并行计算的编码方案和相应算法[61-63],如在数字光计算研究早期提出的符号替换[64]、余数算法[65]、有符号数(MSD)[66-67]、三进制有符号数(TSD)[68]等。数字光学计算采用的进制越高,其算力和计算效率越高,计算步骤越简单,但同时对光学计算系统信噪比的要求越高;采用的进制越低,则编码位占用的空间或时间等维度的资源越多,算力和计算效率越低,计算步骤越复杂。

在实际工程应用中,为了提高终端设备的处理速度和能效比,在对计算精度要求不高的场合中模拟计算更具优势[69],特别是在人工智能应用场景中[70]。很多研究已经表明神经网络的预测准确性在一定程度上不因计算精度的下降而降低[71],这允许研究人员以较低的成本将光计算设备部署到边缘端系统中,如随处可见的摄像头[72]或小型无人机[73]。虽然数字光计算可以通过编码在一定程度上减弱系统噪声对计算结果的影响,提高系统容差能力和计算精度,但在对探测器获得的计算结果进行解码时,由于没有进位操作,探测器

上得到的编码结果通常具有比输入编码信号更高的进制，因此必须确保探测器计算结果的每一位编码位的值经光学计算系统后都能得到正确的计算结果。否则，最后解码结果必然会出现错误，而且越高位的编码位对解码结果的影响越大。模拟光计算的计算精度受系统整体级联噪声水平的影响比较严重，因而模拟光计算对光学系统、调制器和探测器的线性度和动态范围的要求更高。表 1 列出了模拟和数字光计算的优点、缺点及代表性技术。可以看到，模拟和数字光计算共同的难题是难以充分利用光学系统巨大的信道容量，同时，它们因缺乏高速、宽带、高精度的并行电子控制系统而难以发挥光计算技术的全部优势。

表 1　模拟和数字光计算的特点及代表性技术

Table 1　Characteristics and typical techniques of analog and digital optical computing

Optical computing	Advantage	Disadvantage	Typical technique
Analog optical computing	● High information density, no quantization error ● The design freedom of optoelectronic materials and micro-nano structures is high ● Fast computing speed ● Match with deep learning algorithm	● Signals are greatly affected by cascaded system noise, and new fault tolerance methods need to be developed ● Dedicated information processing ● High computing accuracy requires high linearity and dynamic range of system	● Optical Fourier transform ● Smart metamaterials ● ONNs ● Optical convolutional filtering
Digital optical computing	● No cascading noise, strong anti-interference ability ● The computing results can be easily transplanted into existing electronic digital devices for storage and processing ● Under the same system, higher computing accuracy can be obtained through coding	● The resolution of time and space is reduced, and the computing scale is limited ● The complexity of signal processing algorithm is high ● Each coded bit must be calculated accurately, which requires high signal-to-noise ratio (SNR) of the system	● Optical logic device ● Optical coding methods ● Directed logic ● TOC

2.2　早期光计算技术

20 世纪八九十年代，人们意识到光学技术在信息处理方面的巨大优势后掀起了光计算研究的热潮。同时，由于当时集成电路技术的发展并不成熟，其集成度远低于今天的商用电子计算芯片，因此，电子计算机处理大型科学工程所需的计算时间很长，难以满足国防、科研、大型工程等的算力需求。在这种背景下，一大批信息光学领域的专家开始对光计算技术进行研究。其中，合成孔径雷达[74]、联合变换[24]、小波变换[75]、连续空间机（CSM）[76-77]等模拟光计算技术得到了广泛应用。然而，早期光电设备的有限性能导致的系统噪声和不断累积的计算误差严重影响了计算精度。1990 年前后人们转而效仿数字电子计算机发展数字光计算技术[33-34,78]，这方面的研究成果包括逻辑器件[79-82]、逻辑门阵列[83]、逻辑计算[84]、时空编码[40]以及数字并行算法[61-67,85]等。其中，逻辑器件包括光学晶体管[86-88]、光学双稳态器件[52]、空间光调制器（液晶光阀[89]、磁光/声光调制器[90]、可变形镜[91]）等。虽然目前基于低维功能材料的非线性光学效应可以实现超快（百飞秒）、低功耗（低至 fJ/bit）的全光逻辑器件[92]，但这些全光逻辑器件难以实现高的调制深度，且无法进行级联。

之后，基于空间编码技术构建并行逻辑计算系统成为主流，阴影投递法[93]、空间滤波逻辑[94]、符号替换[64,95-97]等技术被相继提出。同时，为了尽可能利用光信号不同维度的信息，许多数字光学并行算法也被提出，这些算法主要包括余数算法[65]、多值逻辑[98]、查找表法[85]等。其中：前两种技术主要被用于解决二进制算数运算时产生的进位问题，而且余数算法可以将计算分解为计算复杂度更小的子计算，在最终解码过程以前的各子计算过程都可以并行完成而无须考虑进位和互连；查找表法通过提前建立输入输出信号之间的关系表来实现任意逻辑功能。在上述技术的基础上，人们搭建了完整的数字光计算原理样机，包括数字光学蜂窝图像处理器（DOCIP）[99]、光学细胞逻辑图像处理器（O-CLIP）[100]、光学并行逻辑系统（H-OPALS、P-OPALS）[101-103]、光电数字计算系统（SPE-4k）[104]、智能像素系统[105-107]等，相关研究成果如图 3 所示。

可以看到，20 世纪八九十年代研究人员就已经在光计算技术的器件、算法、架构、系统和实际应用[108]等方面取得了一系列激动人心的进展，特别是美国贝尔实验室于 1990 年初成功研制了世界上第一台光计算机。然而，受限于半导体制造工艺，早期的光计算原型机主要基于自由空间互连式光计算架构进行设计，这就导致仿照电子计算机构建的数字光计算机体积庞大、结构复杂并且稳定性有限。此外，飞速发展的集成电路和电子计算机使得早期数字光计算原型机的优势不明显，并于 2000 年后逐渐陷入发展低谷。不过，早期光计算的研究经验和相关成果对于今天更好地推动光计算技术的发展仍具有积极的借鉴意义。从早期的光计算发展历史来看，光子很难在逻辑计算领域超越

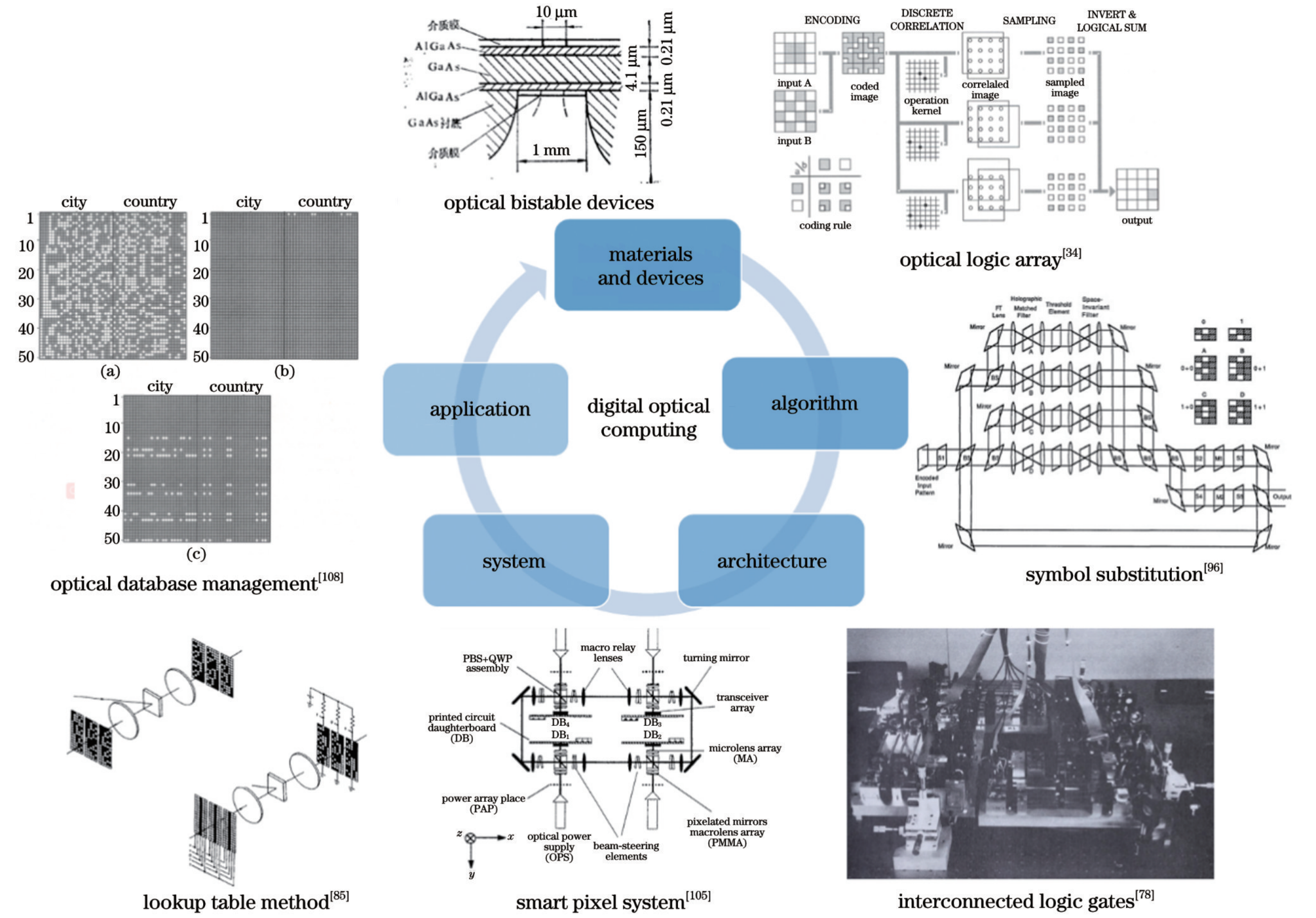

图 3　早期数字光学计算技术的代表性成果

Fig. 3　Representative achievements of early digital optical computing technologies

电子实现通用计算，只有在某些专用场合充分发挥光子的并行特性才能与电子计算形成互补优势。当前，以DNNs为代表的人工智能技术的兴起，使得海量精度不高的线性MM-MAC的算力需求刚好与并行光计算的优势相契合，这也正是光计算再次兴起的核心驱动力。未来，随着集成光路工艺的不断成熟和光学信号调控手段的不断丰富，这种面向专用场合的光计算或光电混合计算有望逐渐走向实用。

2.3　近年的主要光计算技术

目前光计算的主流研究方案[19,109]可分为平面集成式和空间互连式两种架构。其中：平面集成式光计算[55,110-112]主要基于马赫-曾德尔干涉仪阵列（MZIs）[55]、微环谐振器阵列（MRRs）[110-111]、波导调制器阵列（WMA）[112]等基本单元器件（TBBs）实现矢量-矩阵乘法（VVM-MAC）或ONNs，图4展示了平面集成式光计算包含的基本单元及相关进展；另外一条路线即空间互连式光计算[113-122]，主要包括VM-MAC[111]、光学衍射神经网络（D^{2}NNs）[53,114]、傅里叶变换光学滤波系统[54,115]、智能超材料[116-121]、阴影投递法[122]等技术，图5展示了空间互连式模拟光计算的基本原理及相关成果。

图4（a）展示了平面集成式模拟光计算的主要方案，其中，基于MZIs和MRRs的方案可以实现VM-MAC运算和小卷积核功能，并且已经成功演示了算力高达11 TOPS[110]和10^{12} MAC/s的光子卷积计算（OCC）[111]。除此之外，光学储备池神经网络[123]、脉冲神经网络[124]、光子伊辛机[125-126]等新型神经形态光子计算架构也都展现出诱人的应用前景。最近，美国宾夕法尼亚大学提出了一种“感算一体”的光学神经网络芯片[112]，该芯片通过调制波导损耗实现光信号乘法运算[如图4（b）所示]，用探测器将光信号转换为电信号进行求和，然后根据求和后的电信号结果调节微环谐振器响应，实现非线性运算。相比于MZIs和MRRs方案，该架构不需要精确校正和调节每个光子器件的响应特性，减少了激光器的使用数量，降低了对外部高性能电子控制设备的依赖以及系统整体的静态功耗，进一步提高了平面集成光计算的集成度和部署能力。但是，该方案将MM-MAC运算分解为多次光信号乘法和电信号求和操作，这将产生数量庞大的波导损耗调制和光电转换需求，对于芯片散热和速率提升是不利的。除此以外，平面集成式光计算必须采用高速、高带宽、高精度的电子设备才能完成光芯片每个通道的性能测试，从而严重制约了平面集成光计算方案的实际应用。

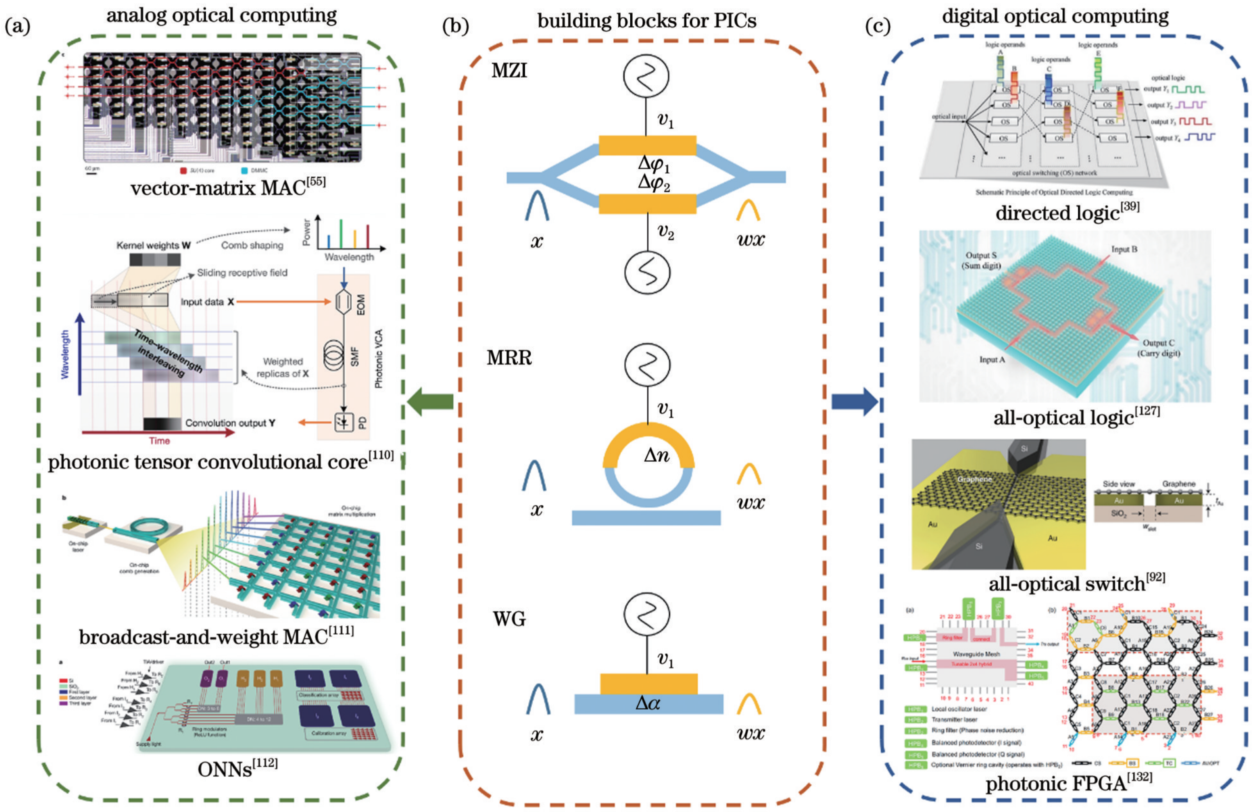

图 4　主要平面集成式光计算基本单元及相关技术方案。(a)平面集成式模拟光计算的主要技术方案；(b)平面集成式光计算技术所需的基本构造单元；(c)平面集成式数字光计算的主要技术方案

Fig. 4　Fundamental building blocks and releated technical solutions of major optical computing based on photonic integrated circuits (PICs). (a) Main technical schemes of analog optical computing for PICs; (b) basic building blocks required for optical computing based on PICs; (c) main technical schemes of digital optical computing for PICs

平面集成式数字光计算的相关进展如图 4(c)所示，基于低维功能材料和等离激元波导[92]的全光开关有望实现能耗极低、速度极快的逻辑操作。然而，这些全光开关具有较大损耗，必须通过高功率脉冲激光进行泵浦，并且难以级联成更大的阵列。相比于全光逻辑[127]，导向逻辑[38-39]不需要依赖非线性光学效应，而是基于电信号加载逻辑操作数并使用光信号实现逻辑结果的互连。这种架构继承了光通信技术的优点，有望实现数字电子-光子一体化的片上异构计算。然而，大多数导向逻辑器件是根据逻辑表达式进行设计的，其功能单一且集成度有限。随着高速度、低功耗光子器件的进步，近年来发展迅猛的片上可重构光子模块[128-129]有望实现大规模的光子可编程逻辑阵列(photonic FPGA)[130-132]。总而言之，与集成电路相比，平面集成式光计算受限于波导对光的有限限制能力以及由制造工艺的高敏感度引起的器件损耗，如光纤和波导耦合损耗、波导侧壁损耗、弯曲漏泄损耗等，这些损耗都将限制平面集成式数字光计算的精度和规模。

对于空间互连式光计算而言，目前的主流方案是模拟光学计算，如图 5 所示。如图 5(b)所示，斯坦福大学最早提出的 VM-MAC 基于柱面镜以光学信号扇入和扇出的方式实现光学 VM-MAC 计算，基于此架构开发的商用光计算系统的最大计算规模已达到 256 维，计算速度高达 10^{12} MAC/s[133]。然而，该架构只能实现 VM-MAC 运算，空间光学系统的高度并行性未得到充分利用。之后，傅里叶变换光学系统[54,115]被广泛用于在频谱面上通过设计合适的掩模实现对输入图像的滤波、卷积、相关等操作。然而，受限于物面和频谱面之间傅里叶变换关系的互相制约，该架构难以实现大规模光学矩阵的卷积计算。如图 5(a)、(b)所示，近年来随着超表面技术兴起而发展起来的格林函数法[134]通过光学超材料或微纳结构调控光学系统的传递函数来实现一阶或高阶微分、积分及方程求解等模拟计算。该方案可以有效减小整个光计算系统的体积，但其角度响应有限，只能实现某种特殊的计算过程，缺乏灵活调整的能力。如图 5(c)所示，基于衍射或散射效应直接构建的 ONNs 相比基于柱面镜的网络具有更大规模的神经元数量，已经展现出了堪比电子神经网络的性能[53,114]。只要将各衍射或散射层的权重系数训练好，整个系统即可以光速完成相关的识别任务。然而，衍射或散射层也需要复杂的计算过程才能确定调制平面上各个像元的振幅和相位等参数，同时整个系统的计算精度仍然受限于电磁结构的加工工艺而难以提高。为了解决这一难题，清华大学提出了一

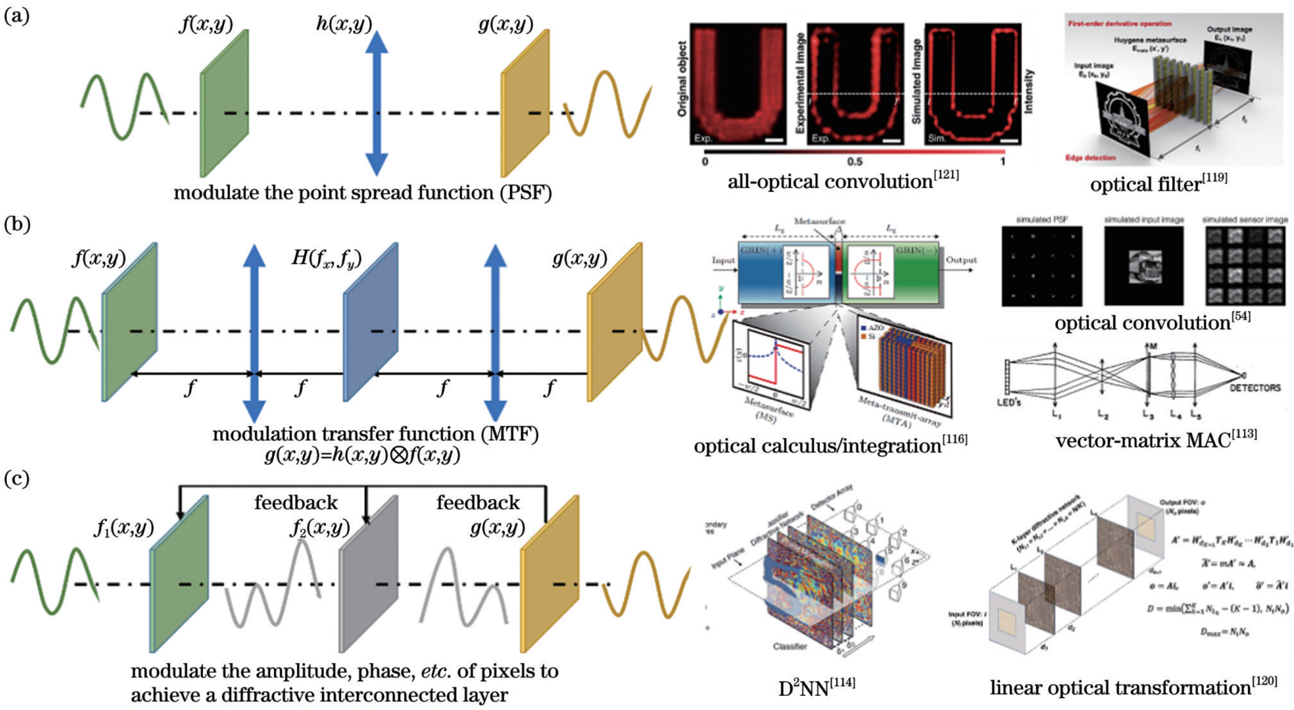

图5　自由空间模拟光学计算方案的主要进展。(a)通过调制点扩散函数实现图像滤波和卷积等操作；(b)通过在频谱面调制系统的调制传递函数实现光学微分/积分和VM-MAC等运算；(c)通过直接调制每一衍射层或散射层中各个像元的参数实现光学互连和光学变换

Fig. 5　Major advances in free-space analog optical computing. (a) Image filtering and convolution operations are achieved by modulating point spread function (PSF); (b) operations such as optical differentiation/integration and VM-MAC are achieved by modulating MTF (modulation transfer function) of the system on the spectral plane; (c) modulate parameters of each pixel in each diffractive or scattered layer to realize optical interconnection and optical transformation

种可构建大规模和复杂神经网络的可重构衍射处理单元（DPU），并且通过开发自适应训练算法实现了可以媲美电子计算机的识别精度[53]，是极具前景的光电混合计算范式。

2.4　多成像投影架构

除了上述主流光计算架构以外，早在20世纪八九十年代，中国科学院上海光学精密机械研究所的王之江、刘立人和周常河等[135]就基于阴影投递架构实现了光学VM-MAC和Hopfield神经网络[136]，并在此基础上提出了二进制补码编码、混合负二进制编码等多种空间编码方案[137]。阴影投递架构具有实现大规模矩阵乘法的潜力，同时通过合适的编码方案可以利用多个空间像元来编码矩阵元素从而提高计算精度。然而，阴影投递法不可避免地会受到光信号在两个矩阵平面传播时引入的衍射误差的影响，其计算精度和计算规模相互制约。近期，周常河、余俊杰等[138]在阴影投递架构中引入高性能衍射分束元件——达曼光栅（DG），巧妙地构建了一种可以实现大规模矩阵-矩阵卷积计算的光学多成像投影（OMica）架构，其原理如图6所示。OMcia通过物像共轭关系减弱了传统阴影投递法中衍射效应对计算结果的干扰，同时利用达曼光栅实现卷积核矩阵在输入矩阵上的并行滑动，因此有望构建具有极高算力的大规模光学矩阵-矩阵卷积计算系统[133]。同时，考虑到实际的深度卷积神经网络

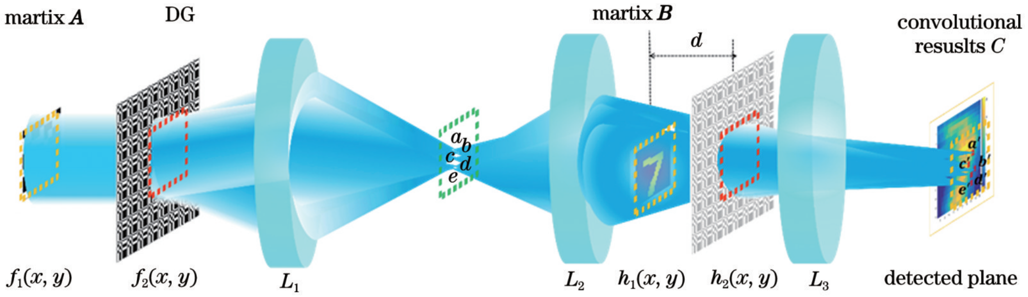

图6　基于多成像投影架构的光学矩阵卷积计算系统[138]

Fig. 6　Optical matrix convolution computing system based on multi-imaging-casting architecture[138]

中的卷积核大小一般 3×3，为了充分发挥系统的算力，笔者课题组提出了多通道卷积核矩阵和多通道输入矩阵并行卷积的空间编码方案，以实现光子张量卷积。需要指出的是，空间互连式光计算方案相比于片上集成式光计算方案充分利用了光学信号的二维并行特性，其单次算力远高于片上集成式光计算。

除此之外，笔者课题组还提出了一种 TE 光垂直入射条件下在 C 和 C+L 波段负一级平均衍射效率分别为 97.3% 和 92.8% 的双脊光栅耦合器[139]，该光栅耦合器可在宽光谱范围内将输入光信号高效地耦合到平板波导中。因此，相比于其他光计算架构，通过设计微光学投影系统和折射、衍射光学元件可以构造紧凑的具有高带宽、复消色差能力的集成式几何波导光电智能计算系统，从而有望推动自然光照明下智能相机和智能光学引擎等第三代智能终端的实现[140-142]。然而，目前面向显示光学应用发展的有源空间光学器件的刷新频率较低，而且线性度和动态范围不够高，仍然需要开发相应的针对光计算技术的专用单元器件。其中，深度学习中急需的非线性光学器件的发展尚不成熟，虽然香港科技大学的研究人员已经基于磁致透明效应演示了全光神经网络[143]，但高效非线性光学层的缺乏对于构建全光神经网络仍然具有一定制约。

3　结束语

由于光子之间缺乏相互作用，光计算目前仍然不能像电子计算一样通过级联多个晶体管并使其稳定工作在指定状态而得到规模更大的电路系统。然而，正因为光子之间缺乏相互作用，才使得不同维度的光信号得以并行复用，如波分复用[144]、频分复用[145]、模分复用[146]等，从而使得信息容量倍增。也就是说，电子计算是在固定的芯片面积内通过增加晶体管密度实现算力翻番，而光计算则是在集成度受限的前提下在固定的芯片面积内通过复用多维度光场信息实现超越电子计算的算力密度。但是，光计算和光通信的技术要求有所不同。光通信技术要求在发射端按照时间序列依次调制光信号，然后将不同维度调制后的光信号复用同一条光学信道进行并行传输，最后在接收端解复用并解调信号，其主要目的是以光信号为载波不失真地远距离、大容量、高速传输信息。光计算系统在每次并行计算时都需要对所有参与运算的信号同时进行调制，调制的信号路数越多，单次计算算力就越高。然而，当光计算系统进行不同维度光信号的并行计算时，必须先对不同维度的光信号进行调制并使其携带所需计算的信号，这就增加了光计算系统对调制器数量的要求。因此，对于光计算硬件系统而言，算力拓展的一大瓶颈在于如何对光信号的不同维度信息进行并行调制以及如何通过共用权重来减少调制器的数量，即将不同维度的光场信息复用到相同的时/空序列中经由同一个调制器进行调制。

综上所述，本文回顾并对比了模拟和数字光计算技术的相关进展。可以看到：第一，数字光计算仍然受到光学逻辑器件的制约，效仿数字电子计算机构造光子逻辑通用计算机的路线存在相当大的难度。至少从短期来看，光计算尤其是满足人工智能算力需求的光计算应该以模拟光计算为主，且关键在于开发高维光场的并行调制/解调技术。第二，光计算系统是一种并行计算系统，其并行度越高，对调制器数量的需求也就越高，对并行电子控制系统和并行信号加载的要求也越高，即没有高速并行电子控制系统就不可能制造实用的光计算系统。第三，光计算应当像量子计算一样，开发可以发挥光计算并行特点的专用算法[58-64]、编码方案[39]、编译器[147-148]及相应的计算理论[149]等，以便充分发挥光计算架构的潜力，比如对网络结构进行预处理后可以避免光学非线性计算[150]，这将有助于全光智能系统的研发。第四，利用光子器件的高带宽实现并行计算不仅要提高光电器件的制造水平和性能，还要进一步开发将模拟和数字光电处理技术优点相结合的新型并行计算架构。第五，由于光计算技术涉及材料、物理、光学、机械、电子、数学、计算机等诸多学科，除了要关注光计算架构的进展外，也应同步开展核心单元器件和配套电子控制系统的研发，加强光电系统的混合设计、集成、封装和测试能力。最后，应明确各种光计算架构可以彰显光计算优越性的应用场景，如光电超算中心、光电处理板卡、神经拟态计算、智能视觉终端[151]、激光雷达和生物医疗传感等。从更长远的角度来看，“碳中和”的战略目标与处理海量数据产生的巨大能耗之间的矛盾会不断加剧，在通用量子计算成熟之前，发展节能、高速和大算力的光电融合计算系统是必经之路[152]。目前来看，光计算发展的重心应该瞄准那些能够充分发挥不同光计算架构并行优势，同时电子计算难以解决的专用应用场景，开发能结合光子和电子各自优势的光电融合计算芯片，最终与电子计算一起构建具有极高算力和能效比的计算系统，进而推动人类社会加速驶向智能时代。

参　考　文　献

[1] Minar M R, Naher J. Recent advances in deep learning: an overview[EB/OL]. (2018-07-21)[2022-09-01]. https://arxiv.org/abs/1807.08169.

[2] Wang X Z, Zhao Y X, Pourpanah F. Recent advances in deep learning[J]. International Journal of Machine Learning and Cybernetics, 2020, 11(4): 747-750.

[3] Tian Y C, Pei K X, Jana S, et al. DeepTest: automated testing of deep-neural-network-driven autonomous cars[EB/OL]. (2017-08-28)[2022-09-01]. https://arxiv.org/abs/1708.08559.

[4] Wang P Y. Research and design of smart home speech recognition system based on deep learning[C]//2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), June 10-12, 2020, Chongqing, China. New York: IEEE Press, 2020: 218-221.

[5] Young T, Hazarika D, Poria S, et al. Recent trends in deep learning based natural language processing[J]. IEEE Computational

Intelligence Magazine, 2018, 13(3): 55-75.

[6] Torfi A, Shirvani R A, Keneshloo Y, et al. Natural language processing advancements by deep learning: a survey[EB/OL]. (2020-03-02)[2022-09-01]. https://arxiv.org/abs/2003.01200.

[7] 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, 2014, 8681: 281-290.

[8] de Lima T F, Peng H T, Tait A N, et al. Machine learning with neuromorphic photonics[J]. Journal of Lightwave Technology, 2019, 37(5): 1515-1534.

[9] Chhowalla M, Jena D, Zhang H. Two-dimensional semiconductors for transistors[J]. Nature Reviews Materials, 2016, 1: 16052.

[10] 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.

[11] AyarOLabs. Technical brief: optical I/O chiplets eliminate bottlenecks to unleash innovation[EB/OL]. (2019-11-15)[2022-09-01]. https://ayarlabs.com/technical-brief-optical-i-o-chiplets-eliminate-bottlenecks-to-unleash-innovation/.

[12] Hills G, Lau C, Wright A, et al. Modern microprocessor built from complementary carbon nanotube transistors[J]. Nature, 2019, 572(7771): 595-602.

[13] Zhong H S, Deng Y H, Qin J, et al. Phase-programmable Gaussian boson sampling using stimulated squeezed light[J]. Physical Review Letters, 2021, 127(18): 180502.

[14] Amdahl's law[EB/OL]. (2022-08-27)[2022-09-01]. https://en.wikipedia.org/wiki/Amdahl%27s_law.

[15] 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.

[16] 刘一凡，张志勇．后摩尔时代的碳基电子技术：进展、应用与挑战[J]．物理学报，2022，71(6)：068503.

Liu Y F, Zhang Z Y. Carbon based electronic technology in post-Moore era: progress, applications and challenges[J]. Acta Physica Sinica, 2022, 71(6): 068503.

[17] 郭光灿．量子信息技术研究现状与未来[J]．中国科学：信息科学，2020，50(9)：1395-1406.

Guo G C. Research status and future of quantum information technology[J]. Scientia Sinica (Informationis), 2020, 50(9): 1395-1406.

[18] 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.

[19] Wetzstein G, Ozcan A, Gigan S, et al. Inference in artificial intelligence with deep optics and photonics[J]. Nature, 2020, 588(7836): 39-47.

[20] Caulfield H J, Dolev S. Why future supercomputing requires optics[J]. Nature Photonics, 2010, 4(5): 261-263.

[21] Tooley F A P, Wherrett B S. Optical computing[M]. Boca Raton: CRC Press, 1989.

[22] Ambs P. Optical computing: a 60-year adventure[J]. Advances in Optical Technologies, 2010, 2010: 372652.

[23] 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(1): 30.

[24] Weaver C S, Goodman J W. A technique for optically convolving two functions[J]. Applied Optics, 1966, 5(7): 1248-1249.

[25] Zhu T F, Zhou Y H, Lou Y J, et al. Plasmonic computing of spatial differentiation[J]. Nature Communications, 2017, 8: 15391.

[26] Estakhri N M, Edwards B, Engheta N. Inverse-designed metastructures that solve equations[J]. Science, 2019, 363(6433): 1333-1338.

[27] Goodman J W. Introduction to Fourier optics[M]. 2nd ed. New York: McGraw-Hill, 1996.

[28] Sui X B, Wu Q H, Liu J, et al. A review of optical neural networks[J]. IEEE Access, 2020, 8: 70773-70783.

[29] Liu J, Wu Q H, Sui X B, et al. Research progress in optical neural networks: theory, applications and developments[J]. PhotoniX, 2021, 2(1): 1-39.

[30] Brunner D, Soriano M C, van der Sande G. Photonic reservoir computing: optical recurrent neural networks[M]. Boston: De Gruyter, 2019.

[31] El Srouji L, Krishnan A, Ravichandran R, et al. Photonic and optoelectronic neuromorphic computing[J]. APL Photonics, 2022, 7(5): 051101.

[32] Khoram E, Chen A, Liu D J, et al. Nanophotonic media for artificial neural inference[J]. Photonics Research, 2019, 7(8): 823-827.

[33] Sawchuk A A, Strand T C. Digital optical computing[J]. Proceedings of the IEEE, 1984, 72(7): 758-779.

[34] Tanida J, Ichioka Y. II Digital optical computing[M]//Progress in optics. Amsterdam: Elsevier, 2000: 77-114.

[35] Zasedatelev A V, Baranikov A V, Sannikov D, et al. Single-photon nonlinearity at room temperature[J]. Nature, 2021, 597: 493-497.

[36] Minzioni P, Lacava C, Tanabe T, et al. Roadmap on all-optical processing[J]. Journal of Optics, 2019, 21(6): 063001.

[37] Yang X Y, Hu X Y, Yang H, et al. Ultracompact all-optical logic gates based on nonlinear plasmonic nanocavities[J]. Nanophotonics, 2017, 6(1): 365-376.

[38] Hardy J, Shamir J. Optics inspired logic architecture[J]. Optics Express, 2007, 15(1): 150-165.

[39] Qiu C Y, Xiao H F, Wang L H, et al. Recent advances in integrated optical directed logic operations for high performance optical computing: a review[J]. Frontiers of Optoelectronics, 2022, 15(1): 17.

[40] Tanida J, Iwata M, Ichioka Y. Extended coding for optical array logic[J]. Applied Optics, 1994, 33(17): 3663-3669.

[41] Jin Y, He H C, Lu Y T. Ternary optical computer principle[J]. Science China Information Sciences, 2003, 46(2): 145-150.

[42] Taubenblatt M A. Optical interconnects for high-performance computing[J]. Journal of Lightwave Technology, 2012, 30(4): 448-457.

[43] Mahajan R, Li X Q, Fryman J, et al. Co-packaged photonics for high performance computing: status, challenges and opportunities[J]. Journal of Lightwave Technology, 2022, 40(2): 379-392.

[44] Hao Y, Xiang S Y, Han G Q, et al. Recent progress of integrated circuits and optoelectronic chips[J]. Science China Information Sciences, 2021, 64(10): 201401.

[45] Zhang Y, Samanta A, Shang K P, et al. Scalable 3D silicon photonic electronic integrated circuits and their applications[J]. IEEE Journal of Selected Topics in Quantum Electronics, 2020, 26(2): 8201510.

[46] Pasricha S, Nicolescu G, Seyedi A, et al. Silicon photonics for high-performance computing and beyond[M]. Boca Raton: CRC Press, 2021.

[47] 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.

[48] 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.

[49] Zhou C H, Yu J J, Li G W, et al. Roadmap of optical computing[J]. Proceedings of SPIE, 2020, 11898: 118981B.

[50] Nahmias M A, de Lima T F, Tait A N, et al. Photonic multiply-accumulate operations for neural networks[J]. IEEE Journal of Selected Topics in Quantum Electronics, 2020, 26(1): 7701518.

[51] 金翊，王哲河，刘玉静，等．三值光学计算机[J]．自然杂志，2019，41(3)：207-218.

Jin Y, Wang Z H, Liu Y J, et al. Ternary optical computer[J]. Chinese Journal of Nature, 2019, 41(3): 207-218.

[52] 李淳飞. 光学双稳态研究20年[J]. 物理, 1996, 25(5): 267-272.
Li C F. Optical bistability research for 20 years[J]. Physics, 1996, 25(5): 267-272.

[53] 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.

[54] 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.

[55] Shen Y C, Harris N C, Skirlo S, et al. Deep learning with coherent nanophotonic circuits[J]. Nature Photonics, 2017, 11(7): 441-446.

[56] Solli D R, Jalali B. Analog optical computing[J]. Nature Photonics, 2015, 9(11): 704-706.

[57] Abdollahramezani S, Hemmatyar O, Adibi A. Meta-optics for spatial optical analog computing[J]. Nanophotonics, 2020, 9(13): 4075-4095.

[58] Wu J M, Lin X, Guo Y C, et al. Analog optical computing for artificial intelligence[J]. Engineering, 2022, 10: 133-145.

[59] Vafa A P M Q, Karimi P, Khavasi A. Recent advances in spatial analog optical computing[C]∥2018 Fifth International Conference on Millimeter-Wave and Terahertz Technologies (MMWaTT), December 18-20, 2018, Tehran, Iran. New York: IEEE Press, 2018: 6-11.

[60] Prise M E, Streibl N, Downs M M. Optical considerations in the design of digital optical computers[J]. Optical and Quantum Electronics, 1988, 20(1): 49-77.

[61] Reif J H, Tyagi A. Efficient parallel algorithms for optical computing with the discrete Fourier transform (DFT) primitive[J]. Applied Optics, 1997, 36(29): 7327-7340.

[62] Barakat R, Reif J. Polynomial convolution algorithm for matrix multiplication with application for optical computing[J]. Applied Optics, 1987, 26(14): 2707-2711.

[63] Timmel A N, Daly J T. Multiplication with Fourier optics simulating 16-bit modular multiplication[C] ∥2018 IEEE International Conference on Rebooting Computing, November 7-9, 2018, McLean, VA, USA. New York: IEEE Press, 2018.

[64] Brenner K H, Huang A, Streibl N. Digital optical computing with symbolic substitution[J]. Applied Optics, 1986, 25(18): 3054-3060.

[65] Huang A, Tsunoda Y, Goodman J W, et al. Optical computation using residue arithmetic[J]. Applied Optics, 1979, 18(2): 149-162.

[66] Avizienis A. Signed-digit number representations for fast parallel arithmetic[J]. IRE Transactions on Electronic Computers, 1961, EC-10(3): 389-400.

[67] Hwang K, Louri A. Optical multiplication and division using modified-signed-digit symbolic substitution[J]. Optical Engineering, 1989, 28(4): 284364.

[68] Cherri A K, Habib M K, Alam M S. Optoelectronic recoded and nonrecoded trinary signed-digit adder that uses optical correlation [J]. Applied Optics, 1998, 37(11): 2153-2163.

[69] Köppel S, Ulmann B, Heimann L, et al. Using analog computers in today' s largest computational challenges[J]. Advances in Radio Science, 2021, 19: 105-116.

[70] Haensch W, Gokmen T, Puri R. The next generation of deep learning hardware: analog computing[J]. Proceedings of the IEEE, 2019, 107(1): 108-122.

[71] Hubara I, Courbariaux M, Soudry D, et al. Quantized neural networks: training neural networks with low precision weights and activations[EB/OL]. (2016-09-22)[2022-09-01]. https:∥arxiv.org/abs/1609.07061.

[72] Chen W L, Zhang Z, Liu G. Retinomorphic optoelectronic devices for intelligent machine vision[J]. iScience, 2022, 25(1): 103729.

[73] Yang X, Chen J Y, Dang Y J, et al. Fast depth prediction and obstacle avoidance on a monocular drone using probabilistic convolutional neural network[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(1): 156-167.

[74] Leith E N. Optical processing techniques for simultaneous pulse compression and beam sharpening[J]. IEEE Transactions on Aerospace and Electronic Systems, 1968, AES-4(6): 879-885.

[75] Sheng Y L, Roberge D, Szu H H. Optical wavelet transform[J]. Optical Engineering, 1992, 31(9): 1840.

[76] Naughton T J. Continuous-space model of computation is Turing universal[J]. Proceedings of SPIE, 2000, 4109: 121-128.

[77] Naughton T J. A model of computation for Fourier optical processors[J]. Proceedings of SPIE, 2000, 4089: 386820.

[78] Murdocca M. A digital design methodology for optical computing [M]. Cambridge: MIT Press, 1990.

[79] Gibbs H M, McCall S L, Venkatesan T N C. Optical bistable devices: the basic components of all-optical systems? [J]. Optical Engineering, 1980, 19(4): 463-468.

[80] Miller D A B, Chemla D S, Damen T C, et al. Band-edge electroabsorption in quantum well structures: the quantum-confined stark effect[J]. Physical Review Letters, 1984, 53(22): 2173-2176.

[81] Boyd G D, Fox A M, Milleret D A B, et al. 33 ps optical switching of symmetric self-electro-optic effect devices[J]. Applied Physics Letters, 1990, 59(21):2631-2633.

[82] Ogura I, Tashiro Y, Kawai S, et al. Reconfigurable optical interconnection using a two-dimensional vertical to surface transmission electrophotonic device array[J]. Applied Physics Letters, 1990, 57(6): 540-542.

[83] Tanida J, Ichioka Y. Programming of optical array logic. 1: image data processing[J]. Applied Optics, 1988, 27(14): 2926-2930.

[84] Jenkins B K, Sawchuk A A, Strand T C, et al. Sequential optical logic implementation[J]. Applied Optics, 1984, 23(19): 3455-3464.

[85] Guest C C, Gaylord T K. Truth-table look-up optical processing utilizing binary and residue arithmetic[J]. Applied Optics, 1980, 19 (7): 1201-1207.

[86] Jain K, Pratt G W. Optical transistor[J]. Applied Physics Letters, 1976, 28(12): 719-721.

[87] Yanik M F, Fan S H, Soljacić M, et al. All-optical transistor action with bistable switching in a photonic crystal cross-waveguide geometry[J]. Optics Letters, 2003, 28(24): 2506-2508.

[88] Jewell J L, Rushford M C, Gibbs H M, et al. Single-etalon optical logic gales[C]∥Conference on Lasers and Electro-Optics, June 19-22, 1984, Anaheim, California. Washington, D. C.: Optica Publishing Group, 1984: THJ2.

[89] Mukohzaka N, Yoshida N, Toyoda H, et al. Diffraction efficiency analysis of a parallel-aligned nematic-liquid-crystal spatial light modulator[J]. Applied Optics, 1994, 33(14): 2804-2811.

[90] Farhat N H, Shae Z Y. Scheme for enhancing the frame rate of magnetooptic spatial light modulators[J]. Applied Optics, 1989, 28 (22): 4792-4800.

[91] Pape D R, Hornbeck L J. Characteristics of the deformable mirror device for optical information processing[J]. Optical Engineering, 1983, 22(6): 226675.

[92] Ono M, Hata M, Tsunekawa M, et al. Ultrafast and energy-efficient all-optical switching with graphene-loaded deep-subwavelength plasmonic waveguides[J]. Nature Photonics, 2020, 14(1): 37-43.

[93] Tanida J, Ichioka Y. Optical logic array processor using shadowgrams[J]. Journal of the Optical Society of America, 1983, 73(6): 800-809.

[94] Bartelt H, Lohmann A W, Sicre E E. Optical logical processing in parallel with theta modulation[J]. Journal of the Optical Society of America A, 1984, 1(9): 944-951.

[95] Huang A. Architectural considerations involved in the design of an optical digital computer[J]. Proceedings of the IEEE, 1984, 72(7): 780-786.

[96] Brenner K H, Huang A. An optical processor based on symbolic substitution[C] ∥Proceeding of the Topical Meeting on Optical Computing, March 18, 1985. [S.l.:s.n.], 1985: WA4.1-4.3.

[97] Jeon H I, Abushagur M A, Sawchuk A A, et al. Digital optical

processor based on symbolic substitution using holographic matched filtering[J]. Applied Optics, 1990, 29(14): 2113-2125.

[98] Abraham G. Multiple-valued logic for optoelectronics[J]. Optical Engineering, 1986, 25(1): 250103.

[99] Huang K S, Sawchuk A A, Jenkins B K, et al. Digital optical cellular image processor (DOCIP): experimental implementation [J]. Applied Optics, 1993, 32(2): 166-173.

[100] Wherrett B S, Walker A C, Craig R G A, et al. The implementation of a programmable digital optical processor[C]∥Conference on Lasers and Electro-Optics 1991, May 12-17, 1991, Baltimore, Maryland, USA. Washington, D. C.: Optica Publishing Group, 1991: CTuD4.

[101] Tanida J, Ichioka Y. OPALS: optical parallel array logic system [J]. Applied Optics, 1986, 25(10): 1565-1570.

[102] Tanida J, Miyazaki D, Ichioka Y. H-OPALS: hybrid optical parallel array logic system[J]. Proceedings of SPIE, 1993, 1806: 568-574.

[103] Tanida J, Konishi T, Ichioka Y. P-OPALS: pure optical-parallel array logic system[J]. Proceedings of the IEEE, 1994, 82(11): 1668-1677.

[104] Ishikawa M, Morita A, Takayanagi N. Massively parallel processing system with an architecture for optoelectronic computing [C]∥Optical Computing 1993, March 16-19, 1993, California, United States. Washington, D.C.: Optica Publishing Group, 1993: OThD.3.

[105] Lentine A L, Reiley D J, Novotny R A, et al. Asynchronous transfer mode distribution network by use of an optoelectronic VLSI switching chip[J]. Applied Optics, 1997, 36(8): 1804-1814.

[106] Desmulliez M P, Tooley F A, Dines J A, et al. Perfect-shuffle interconnected bitonic sorter: optoelectronic design[J]. Applied Optics, 1995, 34(23): 5077-5090.

[107] Liu Y, Robertson B, Boisset G C, et al. Design, implementation, and characterization of a hybrid optical interconnect for a four-stage free-space optical backplane demonstrator[J]. Applied Optics, 1998, 37(14): 2895-2914.

[108] Iwata M, Tanida J, Ichioka Y. Database management using optical array logic[J]. Applied Optics, 1993, 32(11): 1987-1995.

[109] Shastri B J, Tait A N, de Lima T F, et al. Photonics for artificial intelligence and neuromorphic computing[J]. Nature Photonics, 2021, 15(2): 102-114.

[110] 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.

[111] Feldmann J, Youngblood N, Karpov M, et al. Parallel convolutional processing using an integrated photonic tensor core [J]. Nature, 2021, 589(7840): 52-58.

[112] Ashtiani F, Geers A J, Aflatouni F. An on-chip photonic deep neural network for image classification[J]. Nature, 2022, 606 (7914): 501-506.

[113] 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.

[114] 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.

[115] Miscuglio M, Hu Z B, Li S R, et al. Massively-parallel amplitude-only Fourier neural network[J]. Optica, 2020, 7(12): 1812-1819.

[116] Silva A, Monticone F, Castaldi G, et al. Performing mathematical operations with metamaterials[J]. Science, 2014, 343(6167): 160-163.

[117] Pors A, Nielsen M G, Bozhevolnyi S I. Analog computing using reflective plasmonic metasurfaces[J]. Nano Letters, 2015, 15(1): 791-797.

[118] Li L L, Zhao H T, Liu C, et al. Intelligent metasurfaces: control, communication and computing[J]. eLight, 2022, 2(1): 7.

[119] Wang Z C, Hu G W, Wang X W, et al. Single-layer spatial analog meta-processor for imaging processing[J]. Nature Communications, 2022, 13: 2188.

[120] Kulce O, Mengu D, Rivenson Y, et al. All-optical synthesis of an arbitrary linear transformation using diffractive surfaces[J]. Light: Science & Applications, 2021, 10: 196.

[121] 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.

[122] Wang T Y, Ma S Y, Wright L G, et al. An optical neural network using less than 1 photon per multiplication[J]. Nature Communications, 2022, 13(1): 123.

[123] Tanaka G, Yamane T, Héroux J B, et al. Recent advances in physical reservoir computing: a review[J]. Neural Networks, 2019, 115: 100-123.

[124] Feldmann J, Youngblood N, Wright C D, et al. All-optical spiking neurosynaptic networks with self-learning capabilities[J]. Nature, 2019, 569(7755): 208-214.

[125] Pierangeli D, Marcucci G, Conti C. Large-scale photonic Ising machine by spatial light modulation[J]. Physical Review Letters, 2019, 122(21): 213902.

[126] Fabre C. The optical Ising machine[J]. Nature Photonics, 2014, 8 (12): 883-884.

[127] Lu C H, Zhu B, Zhu C Y, et al. All-optical logic gates and a half-adder based on lithium niobate photonic crystal micro-cavities[J]. Chinese Optics Letters, 2019, 17(7): 072301.

[128] Bogaerts W, Pérez D, Capmany J, et al. Programmable photonic circuits[J]. Nature, 2020, 586(7828): 207-216.

[129] Capmany J, Pérez D. Programmable integrated photonics[M]. Oxford: Oxford University Press, 2020.

[130] Zhang W F, Yao J P. Photonic integrated field-programmable disk array signal processor[J]. Nature Communications, 2020, 11: 406.

[131] Pérez-López D, López A, DasMahapatra P, et al. Multipurpose self-configuration of programmable photonic circuits[J]. Nature Communications, 2020, 11: 6359.

[132] Perez-Lopez D, López-Hernandez A, Macho A, et al. Towards field-programmable photonic gate arrays[EB/OL]. (2020-02-22) [2022-09-01]. https://arxiv.org/abs/2002.09681.

[133] LabsLenslet. Enlight256 white paper report[EB/OL]. [2022-09-01]. http://besho.narod.ru/reviews/newage/EnLight256.pdf.

[134] 周毅，陈瑞，陈雯洁，等. 空域模拟光学计算器件的研究进展[J]. 物理学报，2020，69(15): 157803.
Zhou Y, Chen R, Chen W J, et al. Advances in spatial analog optical computing devices[J]. Acta Physica Sinica, 2020, 69(15): 157803.

[135] Zhou C H, Liu L R, Wang Z J. Binary-encoded vector-matrix multiplication architecture[J]. Optics Letters, 1992, 17(24): 1800-1802.

[136] 周常河. Hopfield光学神经网络[D]. 上海：中国科学院上海光学精密机械研究所，1995.
Zhou C H. Hopfield optical neural network[D]. Shanghai: Shanghai Institute of Optics and Fine Mechanics, The Chinese Academy of Sciences, 1995.

[137] Liu L R, Li G Q, Yin Y Z. Optical complex matrix-vector multiplication with negative binary inner products[J]. Optics Letters, 1994, 19(21): 1759-1761.

[138] 周常河，余俊杰，马国庆. 基于多成像投影架构的光学卷积计算系统及方法：CN202110742313.4[P]. 2021-07-01.
Zhou C H, Yu J J, Ma G Q. Optical convolution computing system and method based on multi imaging projection architecture: CN202110742313.4[P]. 2021-07-01.

[139] Ma G Q, Zhou C H, Xie Y F, et al. Double-groove rectangular gratings for high-efficiency wideband vertical coupling in planar-integrated optical systems[J]. Chinese Optics Letters, 2022, 20(9): 090501.

[140] Gruber M, Jahns J, Sinzinger S. Planar-integrated optical vector-matrix multiplier[J]. Applied Optics, 2000, 39(29): 5367-5373.

[141] Jahns J, Huang A. Planar integration of free-space optical components[J]. Applied Optics, 1989, 28(9): 1602-1605.

[142] Hofmann M, Hauguth-Frank S, Lebedev V, et al. Sapphire-GaN-based planar integrated free-space optical system[J]. Applied Optics, 2008, 47(16): 2950-2955.

[143] 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.

[144] Ishio H, Minowa J, Nosu K. Review and status of wavelength-division-multiplexing technology and its application[J]. Journal of Lightwave Technology, 1984, 2(4): 448-463.

[145] Gatto A, Parolari P, Boffi P. Frequency division multiplexing for very high capacity transmission in bandwidth-limited systems[C]//Optical Fiber Communication Conference, March 19-23, 2017, Los Angeles, California. Washington, D.C.: Optica Publishing Group, 2017: W1K.1.

[146] Zeb K, Zhang X P, Lu Z G. High capacity mode division multiplexing based MIMO enabled all-optical analog millimeter-wave over fiber fronthaul architecture for 5G and beyond[J]. IEEE Access, 2019, 7: 89522-89533.

[147] Woods D. Computational complexity of an optical model of computation[D]. Maynooth: National University of Ireland, 2005.

[148] de Lima T F, Tait A N, Mehrabian A, et al. Primer on silicon neuromorphic photonic processors: architecture and compiler[J]. Nanophotonics, 2020, 9(13): 4055-4073.

[149] Dolev S, Fitoussi H. The traveling beams optical solutions for bounded NP-complete problems[M]//Crescenzi P, Prencipe G, Pucci G. Fun with algorithms. Lecture notes in computer science. Heidelberg: Springer, 2007, 4475: 120-134.

[150] Xiang J L, Colburn S, Majumdar A, et al. Knowledge distillation circumvents nonlinearity for optical convolutional neural networks[J]. Applied Optics, 2022, 61(9): 2173-2183.

[151] Bai B J, Luo Y, Gan T Y, et al. To image, or not to image: class-specific diffractive cameras with all-optical erasure of undesired objects[J]. eLight, 2022, 2(1): 14.

[152] Hennessy J L, Patterson D A. A new golden age for computer architecture[J]. Communications of the ACM, 2019, 62(2): 48-60.

Future of Optical Computing: Analog or Digital?

Ma Guoqing[1,2], Zhou Changhe[3*], Zhu Rongwei[1,2], Zheng Fenglu[1,2], Yu Junjie[1,2**], Situ Guohai[1,2***]

[1]*Laboratory of Information Optics and Optoelectronic Technology, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai* 201800, *China*;

[2]*Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing* 100049, *China*;

[3]*The Institute of Photonics Technology, Jinan University, Guangzhou* 510632, *Guangdong, China*

Abstract

Significance Deep learning (DL) has become a powerful driving force in the era of intelligence and has been widely used in computer vision, speech recognition, natural language processing, *etc*. However, more than 80% of calculations in DL are matrix-matrix multiply-accumulate (MM-MAC) operations. Large-scale MM-MAC operations result in a large number of memory access requirements when the algorithm is converted into CPU-executable code. Limited by the von Neumann architecture of electronic computers and the physical constraints of the interconnection limit of copper wires on a chip, the training efficiency and speed of deep neural networks (DNNs) are severely restricted. According to the research of de Lima *et al.*, the computing power required to train state-of-the-art DNNs doubles approximately every 3.5 months, far exceeding the computing power supply of electronic integrated circuits (EICs) that follow Moore's Law.

Compared to traditional electronic computing, optical computing is expected to build artificial intelligence (AI) accelerators with high computing power and energy efficiency ratio owing to the high parallelism, high speed, and low power consumption of photons. Currently, various optical computing architectures have demonstrated advantages in terms of high computing power and energy efficiency ratio, and their development routes can be divided into two types. One route is to realize dedicated optical information processing based on multidimensional optical signal modulation and to primarily focus on analog optical computing, such as multiply-accumulate (MAC) operations, convolutions and correlations, differentiation and integration, Fourier transform, and optical neural networks (ONNs). The other route is to use the conception of electronic computers to design digital optical computers, such as optical transistors, optical logic devices, optical directed logic operations, space-time parallel coding, and ternary optical computers. Additionally, some important supporting technologies such as optical interconnections, optoelectronic copackaging, optoelectronic heterogeneous integration, and three-dimensional advanced packing have been widely used to improve the performance of electronic computers.

In general, owing to the lack of efficient and reliable weak-light nonlinear optical effects and optical logic devices, it is difficult for photons to realize general digital logic computers like electrons. In addition, the technology that uses photons to store information has not been proven effectively, which implies that photons cannot independently complete the entire process between memory and computing, and all-optical signal processing is still challenging to achieve. Therefore, from electronic information storage to photonic information loading or from photonic information loading to electronic information storage, high-precision and high-speed parallel electronic control systems and analog-to-digital conversion circuits are still required to fully utilize the parallelism of optical computing. Currently, optical computing is primarily based on linear analog computing, and its computing accuracy is sufficient to build practical high-performance AI accelerators. Moreover, by developing suitable coding schemes, parallel algorithms, and architectures to further fully utilize the parallelism of each dimension of photons, photons are expected to provide a computing power density and energy efficiency ratio that exceed those of electrons on the same footprint. Furthermore, a high computing accuracy can be realized even with error-sensitive photonic devices and optical systems. Although binary electronic logic computing can simulate various practical physical scenarios with a sufficiently high computational accuracy, the ever-increasing computational load will significantly increase power consumption. Optical computing is expected to build high-performance and high-energy-efficiency fuzzy parallel computing systems similar to the human brain with limited computing precision.

Progress This review analyzes and discusses mainstream optical computing technologies from the perspective of analog and digital optical computing, aiming to guide the development of optical computing. First, we introduce three main technical paths for solving the bottleneck of computing power supply and power consumption in the post-Moore era (Fig. 1). Additionally, we indicate that optoelectronic computing or all-optical computing is the most promising method for building the next generation of human-like fuzzy parallel computing systems with high computing power and energy efficiency (Fig. 2). We then summarize the advantages and disadvantages of analog and digital optical computing (Table 1) and discuss the main progress and representative achievements of optical computing at different stages, including early optical computing (Fig. 3), integrated optical computing (Fig. 4), free-space interconnected optical computing (Fig.5), and multi-imaging-casting architecture (Fig. 6). On this basis, we describe the limitations and key technical bottlenecks facing the further development of optical computing. Finally, we discuss future trends and directions of optical computing.

Conclusions and Prospects The route of imitating the digital electronic computer to construct a general computer for photonic logic is severely limited by optical logic devices. Currently, the focus of optical computing should be on special application scenarios that take full advantage of optical parallelism and are challenging to solve by electronic computing. From a long-term perspective, in addition to AI technology, the contradiction between the strategic goal of carbon neutrality and the significant energy consumption of data centers that provide high computing power for the rapidly increasing data processing requirements will continue to aggravate. Before the advent of practical quantum computers, the development of energy-saving, highly efficient, and high computing power optoelectronic intelligent computing is the most promising solution.

Key words optical computing; analog optical computing; digital optical computing; architecture of optical computing; optical matrices multiply-accumulate; optical neural networks; optoelectronic intelligent computing; optical signal processing