

## 光学学报

硅光集成与光电融合:后摩尔时代的关键使能技术  
(特邀)周林杰<sup>1,2\*</sup>, 冉诗环<sup>1</sup>, 原旗旗<sup>1</sup>, 吴月<sup>1</sup>, 陆梁军<sup>1,2</sup>, 李雨<sup>1,2</sup>, 郭宇耀<sup>1,2</sup>, 陈建平<sup>1,2</sup><sup>1</sup>上海交通大学信息与电子工程学院光子传输与通信全国重点实验室, 上海 200240;<sup>2</sup>上海交大-平湖智能光电研究院, 浙江 平湖 314200

**摘要** 随着摩尔定律放缓和传统电子技术逼近物理极限,硅光集成技术作为突破信息系统性能瓶颈的关键路径,正引领信息技术向光电融合新范式转型。系统回顾硅基光电子规模化集成在光通信、光感知和光计算三大核心应用领域的技术演进历程,深入分析其在后摩尔时代的战略价值。在光通信领域,硅基光电子技术实现调制带宽突破,通过多通道并行和密集波分复用技术,大幅提升单芯片数据传输容量,为数据中心互连、高性能计算互连等应用提供核心技术支撑;在光感知领域,基于硅光平台的激光雷达技术在集成度和性能方面实现突破,为自动驾驶等前沿应用提供核心技术支撑;在光计算领域,光电协同架构展现出超越传统电子计算的能效优势,通过光子并行处理与电子精确控制的深度融合,为人工智能大算力需求提供新的解决方案,引领计算架构向光电一体化方向演进。然而,规模化集成仍面临材料异质集成、阵列调控、系统封装与热管理等关键挑战,需要从材料、器件和系统层面协同攻关。硅光集成与光电融合不仅是延续信息产业发展的技术选择,更是重塑通信、感知和计算基础架构的战略性使能技术,三大应用领域的协同发展正推动硅光技术从单点突破向系统级光电融合的战略转型,为构建高效、智能、可持续的数字社会提供核心技术支撑。

**关键词** 集成光学; 光电融合; 光通信; 光感知; 光计算

中图分类号 O439

文献标志码 A

DOI: 10.3788/AOS251225

## 1 引言

光电子技术凭借其超快速度、高带宽、高能效和多维物理特性,正逐渐成为现代信息社会的重要组成部分。它不仅在通信、计算、传感和显示等传统领域推动了创新,也在量子信息和智能光子技术等前沿领域展现出巨大潜力,成为引领未来信息技术发展的重要力量。

在数据通信方面,光电子技术已成为全球互联网基础设施的基石。依托光纤网络和光子集成芯片(PIC),该技术实现了超高速数据传输,大幅提升电信和数据中心的性能,降低数据传输延迟,为云计算和大规模数据处理提供有力支撑。与传统电气互连相比,光学传输在能耗控制上具有重大优势,对构建可持续的数据基础设施具有战略价值。在计算领域,光子神经网络和量子光电系统展现出在复杂计算和安全通信方面的巨大潜力。光子加速器突破了传统电子硬件在带宽和能耗上的限制,为深度神经网络的训练和推理提供更高效率的计算模式。光电子技术利

用并行处理和低延时的优势,正在重构计算架构的基本范式。此外,在感知领域,光电子技术的应用也日益广泛,从生物传感和气体检测到自动驾驶汽车的激光雷达系统,它提供高精度和低功耗的感知解决方案。中红外光子技术的持续发展,也为环境监测和精准医疗开辟了新的技术维度,展现出跨学科创新的巨大潜力。

硅基光电子技术作为光电子领域的革命性技术,正在重塑信息技术的发展边界<sup>[1-3]</sup>。这一技术的核心理念是在硅基衬底上直接整合光学与电学元件,实现光电功能的无缝融合。硅材料独特的光学特性,尤其是其高折射率,为亚微米尺度的光传输与调控提供坚实基础。在技术发展历程中,基于绝缘体上硅(SOI)晶圆的亚微米尺寸单模光波导传播损耗已降至 0.1 dB/cm 以下<sup>[1]</sup>。通过深紫外(DUV) 193 nm 光刻技术,可实现低至 100 nm 的特征尺寸,为多种硅基微纳器件提供高可靠的加工能力。基于硅基光电子技术的光互连在米级以上距离能耗显著低于传统电互连,每比特能耗可低至 1 pJ,仅为电互

收稿日期: 2025-06-05; 修回日期: 2025-06-24; 录用日期: 2025-06-25; 网络首发日期: 2025-06-25

基金项目: 国家重点研发计划(2022YFB2804502)、国家自然科学基金(62305212, 62090052, 62135010)、上海市 2024 年度“科技创新行动计划”集成电路领域基础研究项目(24JD1401600)

通信作者: \*ljzhou@sjtu.edu.cn

连能耗的 1/10。英特尔的光电共封装(CPO)技术通过整合硅光收发器<sup>[4]</sup>,使功耗降低 50%,对超大规模数据中心具有重大意义。硅光技术还具有高带宽和可扩展的优势。密级波分复用(DWDM)技术能够将波长通道间距压缩至 25 GHz,可轻松实现 Tbit/s 级传输容量。比利时微电子研究中心(IMEC)开发的 8 通道微环收发器在 5 mm<sup>2</sup>面积内实现 400 Gbit/s 的数据传输,并可通过增加通道数扩展至 1.6 Tbit/s<sup>[5]</sup>。硅基异质集成工艺的突破进一步拓展硅光的能力<sup>[6]</sup>。集成锗-锡(Ge-Sn)材料的光电探测器(PD)可将探测范围扩展至中红外波段<sup>[7]</sup>,而集成薄膜铌酸锂(TFLN)的调制器则可实现 200 GHz 以上的调制

带宽<sup>[8]</sup>。

硅基光电子为光电融合技术的发展提供天然的平台(图 1)。光电融合通过将光子器件与电子器件集成在同一硅光平台,实现了光通信的高速和大带宽与电子系统的高精度和高灵活性的完美结合。这种融合不仅突破了传统微电子技术在速度、带宽和功耗等方面的物理瓶颈,为数据中心、高性能计算和人工智能(AI)等领域提供前所未有的算力和能效提升,也为后摩尔时代的信息产业提供可持续发展的技术路径,推动从“以电为主”到“光电协同”的历史性跨越。光电融合将深刻影响未来计算、通信和感知系统的架构,是实现智能社会和数字经济的核心基础。

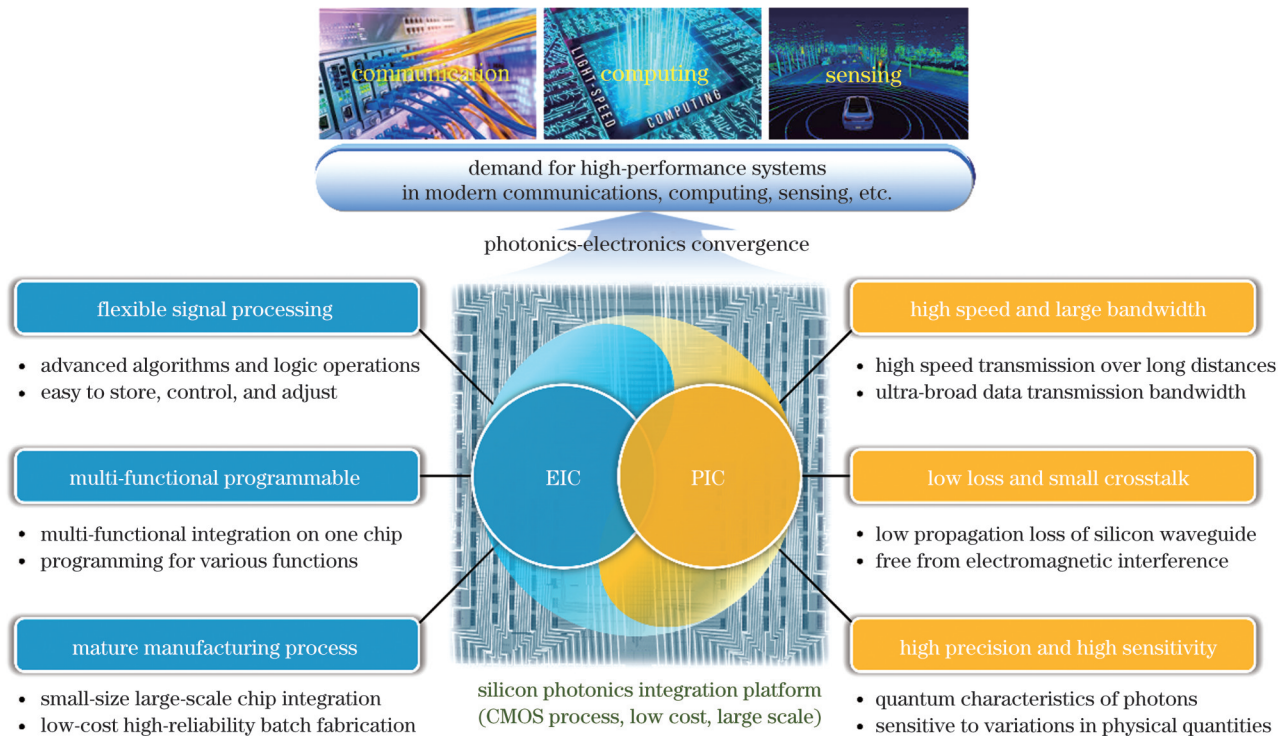


图 1 硅基光电融合

Fig. 1 Silicon-based photonics-electronics convergence

本文围绕硅基光电子规模化集成这一关键技术,从理论基础到技术应用、从现状分析到未来展望,系统地构建硅光的技术框架体系。全文共 7 个部分,每个部分既相对独立又有机关联,呈现硅基光电子技术的科学内涵与战略价值。第 1 部分概括论述硅基光电子技术在后摩尔时代的战略定位,分析其在突破传统电子技术瓶颈中的关键作用,明确光电融合发展的重要意义与研究价值。第 2 部分梳理硅基光电子中的关键技术,包括无源器件、有源器件和集成平台技术,分析当前技术发展水平与面临的核心挑战。第 3 部分聚焦光通信应用领域,剖析硅基光电子技术在高速数据传输、多通道密集波分复用、片上光互连等方面的技术突破与应用价值,展现其在通信基础设施变革中的重

要作用。第 4 部分专注于光感知技术创新,探讨硅基光电子在激光雷达系统中的关键技术,特别是调频连续波(FMCW)与光学相控阵(OPA)技术的集成应用,揭示其在智能感知领域的变革潜力。第 5 部分立足于光计算这一新兴领域,阐述硅基光电子在 AI 计算架构中的独特优势,探讨光电协同计算范式对传统计算模式的突破与重构。第 6 部分总结硅光规模化集成过程中的关键问题,包括材料异质集成的复杂性、热管理与性能稳定性、高密度封装技术等核心瓶颈,为技术突破指明方向。第 7 部分作为总结,回顾硅基光电子规模化集成的主要进展与技术成就,分析当前瓶颈与挑战,展望未来发展趋势与潜在应用前景,为光电融合技术的持续发展提供战略指引。

## 2 硅基光电子发展现状

### 2.1 硅基光电子的核心器件

无源器件作为光电子集成的基础,在光信号的路由、分配、滤波和转换中发挥着核心作用<sup>[9-10]</sup>。当前,无源器件的发展主要聚焦于减小尺寸、提升性能、提高工艺容忍度等方面。典型无源器件包括波导、功率分路/合路器、耦合器、滤波器、波分(模分)复用/解复用器、偏振操控器件以及波导模斑转换器等。为追求高性能紧凑型无源器件,学术界正积极推进计算优化与器件设计的深度融合<sup>[11]</sup>。逆向设计与亚波长光栅/超表面结构的结合,已成为突破传统设计范式的重要路径<sup>[12]</sup>。在芯片与光纤耦合效率方面,边缘耦合器和光栅耦合器的性能不断提升,在大带宽范围内保持低插入损耗( $<0.5$  dB)<sup>[13-15]</sup>。同时,为克服硅材料的局限性,人们也开始探索多材料集成。硅-氮化硅(Si-SiN)多层波导集成平台应运而生,其代表光电子集成的创新方向。相较于传统硅材料,氮化硅具有特有优势<sup>[16]</sup>:热光系数仅为 $2.5 \times 10^{-5} \text{ K}^{-1}$ ,远低于硅的 $1.8 \times 10^{-4} \text{ K}^{-1}$ ,这意味着更优异的热稳定性和高功率处理能力。硅光无源器件也正加速向多元化应用场景拓展。除传统的电信和数据通信领域,这些器件在光束操纵、生物传感、量子信息处理、光谱学和陀螺仪等新兴领域展现出巨大潜力。紧凑的器件尺寸、低损耗特性以及成熟的互补金属氧化物半导体器件(CMOS)加工工艺,为其广泛应用提供坚实基础。

硅材料固有的间接带隙特性,严重制约硅基光源的发展,这一技术瓶颈推动了硅基混合集成和异质集成的发展<sup>[17-19]</sup>。磷化铟(InP)和砷化镓(GaAs)等直接带隙材料,因其高效发光特性,已成为硅基光源研究的重点方向。当前实现高性能激光器制备的主要方式包括片间混合集成、片上倒装焊、片上键合、外延生长等,它们具有不同的技术成熟度和发展潜力。量子点(QD)因其对晶体缺陷的高容忍度、低温度敏感性和低阈值电流密度等独特优势,成为当前研究的热点。III-V族量子点激光器在硅衬底上的外延生长技术取得的突破性进展,不仅实现与硅波导的单片集成,还显著提升器件的集成密度和温度稳定性。通过精细的掺杂调控和位错过滤层技术,量子点激光器在性能和可靠性方面展现出巨大潜力。尽管锗本身是间接带隙半导体,但通过能带工程,已成功实现室温下工作的锗激光器。锗锡合金的引入进一步拓展带隙调控空间,为硅基光源的性能提升开辟新的技术路径。

作为光信号调制的核心组件,调制器的性能直接决定光电系统的信息传输效率。硅基高速电光调制器根据其光学结构分为马赫-曾德尔调制器(MZM)和微环调制器(MRM)。近年来这两类硅基电光调制器均取得较大进展,如图2所示。硅基MZM的电光带宽主要取决于光信号与电信号的群速率匹配、行波电极的

阻抗匹配和行波电极的微波损耗。2015年,麦吉尔大学研究团队采用慢波电极结构对光信号与电信号的群速率匹配以及行波电极的阻抗匹配进行优化,获得带宽为41 GHz的MZM,支持112 Gbit/s四电平幅度调制(PAM-4)信号的调制<sup>[20]</sup>,如图2(a)所示。2018年,华中科技大学研究团队通过挖空衬底来降低微波损耗,将MZM的带宽提高至50 GHz以上<sup>[21]</sup>,如图2(b)所示。拉瓦尔大学研究团队于2022年和2024年先后通过电极分段、集成分布式微电容等方式降低微波损耗,获得带宽大于60 GHz的MZM<sup>[22-23]</sup>,如图2(c)和(d)所示。受限于硅的弱载流子色散效应,传统MZM器件的调制效率、带宽与传输损耗相互制约。为了打破这一瓶颈,2023年,北京大学研究团队采用慢光波导实现110 GHz带宽的MZM<sup>[24]</sup>,如图2(e)所示。慢光波导结构通过引入谐振来减小MZM的尺寸,可在增强MZM调制效率的同时提高MZM的带宽。2024年,日本横滨国立大学研究团队采用慢光波导、慢波电极相结合的方式,获得尺寸仅为 $0.3 \text{ mm}^2$ 的MZM,其带宽为31 GHz<sup>[25]</sup>,如图2(f)所示。近年来,研究发现采用时间-频率均衡方案可以进一步拓展MZM的带宽,如图2(g)~(i)所示。采用时间-频率均衡的MZM的调制区域可分为3段,分别是主调制区、同相调制区和反相调制区,同相调制区与反相调制区之间还有一段电光信号的速度失配区。在低频情况下,同相调制区与反相调制区所产生的调制信号相互抵消;而在高频情况下,由于射频(RF)信号在行波电极上传输的衰减,同相调制区所产生的调制信号大于反相调制区的调制信号,使得总的调制信号增加,从而提高了高频情况下的MZM响应,增加MZM带宽。2024年,日本国家信息与通信技术研究所研究团队采用该方案获得带宽为60 GHz的MZM<sup>[26]</sup>,如图2(g)所示;浙江大学(ZJU)研究团队采用该方案将MZM的带宽提高至110 GHz,其支持140 Gbit/s开关键控(OOK)信号调制<sup>[27]</sup>,如图2(h)所示;2025年,新加坡AMF(Advanced Micro Foundry)公司进一步采用该方案进行优化,获得带宽为90 GHz,其支持200 Gbit/s PAM-4信号调制的硅基MZM<sup>[28]</sup>,如图2(i)所示。

传统硅基高速电光调制器调制带宽通常为40~50 GHz<sup>[29-31]</sup>,虽然可以通过结构优化将其带宽拓展至110 GHz以上,但调制效率和带宽依然具有固有局限性。通过在硅光平台上引入新型电光材料,可突破该性能限制。其中,石墨烯材料凭借其高载流子迁移率和电吸收性质,为高速调制开辟新途径。通过在锗和锗-硅材料中引入法兰兹-卡尔迪西(FK)效应和量子限制斯塔克(QCS)效应,可实现带宽超过100 GHz的紧凑型低功耗电吸收调制器<sup>[32]</sup>。通过在硅上键合或外延生长III-V半导体材料,可以获得较大的折射率变化、高电子迁移率及低吸收损耗,使调制效率较全硅器件提升3~5倍<sup>[33-34]</sup>。钛酸钡(BTO)和有机电光聚合

物在硅波导或等离子-有机混合(POH)调制器中的应用,实现了大于百 GHz 的超高带宽和紧凑的器件尺寸<sup>[35-37]</sup>。

微环谐振器(MRR)结构也常用于高速光互连和数据传输领域<sup>[38-40]</sup>,近年来硅基微环高速调制器在调制效率、带宽、能耗和集成能力等方面取得长足进展。在器件设计创新方面,研究人员通过设计异形PN结,如L形PN结、垂直PN结、Z形PN结等,有效提升MRM的调制效率,目前基于异形PN结的MRM的调制效率大多在 $0.5\text{ V}\cdot\text{cm}$ 左右<sup>[38]</sup>。2019年,美国英特尔公司采用L形PN结获得调制效率为 $0.52\text{ V}\cdot\text{cm}$ 、带宽为50 GHz、支持128 Gbit/s PAM-4信号发射的硅基MRM<sup>[41]</sup>,如图2(j)所示。根特大学(Ghent U.)在2019年报道了其基于垂直PN结的MRM,进一步将MRM的调制效率提高至 $0.37\text{ V}\cdot\text{cm}$ ,该MRM支持100 Gbit/s PAM-4信号调制<sup>[42]</sup>,如图2(k)所示。此外,利用光学峰值效应可突破MRM光学带宽的限制,有效提高电光带宽<sup>[43-44]</sup>。2020年,国家信息光子创新中心利用光学峰值效应获得带宽大于67 GHz的MRM,该MRM采用水平PN结进行调制,支持200 Gbit/s PAM-4信号发射<sup>[45]</sup>,如图2(l)所示。采用异形PN结虽然能够提高MRM的调制效率,但也带来了较大的结电容,限制MRM的电学带宽。常见的解决方案是减小MRM的半径,从而减小结电容。2022年,英特尔公司将基于L形PN结的MRM半径从 $6\text{ }\mu\text{m}$ 缩小至 $4\text{ }\mu\text{m}$ ,该MRM在获得 $0.53\text{ V}\cdot\text{cm}$ 高调制效率的同时具有62 GHz的大带宽,支持240 Gbit/s PAM-4信号调制<sup>[46]</sup>,如图2(m)所示。2024年,根特大学在垂直PN结的基础上通过将微环半径减小至 $2.5\text{ }\mu\text{m}$ 以提高其数据传输能力,在仅为 $0.8\text{ V}_{\text{pp}}$ (峰峰值)的驱动电压下实现40 Gbit/s OOK信号调制<sup>[47]</sup>,如图2(o)所示。值得注意的是,减小MRM半径不仅能够提升带宽,也能获得更大的自由光谱范围(FSR)以容纳更多波分复用(WDM)通道数,从而提升总通信容量。2024年,美国Advanced Micro Devices公司通过在垂直PN结的基础上集成片上电感以引入峰值效应,实现 $44\text{ pm/V}$ 的调制效率和41 GHz的带宽,相比于未集成片上电感的参考MRM,其带宽提高11 GHz<sup>[48]</sup>,如图2(n)所示。惠普实验室则采用Z形PN结分段的方案来提高MRM的带宽,获得调制效率为 $0.6\text{ V}\cdot\text{cm}$ 、带宽为58.1 GHz、支持单通道200 Gbit/s的PAM-4信号调制<sup>[38]</sup>,如图2(p)所示。

PD和雪崩光电探测器(APD)是硅基光子中用于光电转换和探测的主要器件,近年来PD和APD的研究进展如图3所示。在锗-硅PD方面,新加坡国立大学和华中科技大学通过集成电感获得大于60 GHz的光电带宽<sup>[49-50]</sup>,分别如图3(a)和(b)所示;莱布尼兹高性能微电子研究所通过鳍状结构设计将锗层嵌入互

补的原位掺杂硅层之间,在将探测器3 dB带宽提升至265 GHz的同时获得 $0.3\text{ A/W}$ 的响应度<sup>[51]</sup>,如图3(c)所示。针对锗-硅外延的晶格失配问题,研究人员开发了缺陷钝化技术,包括选择性生长、退火处理和界面工程,以有效减少晶格失配引入的缺陷,在降低暗电流的同时提高载流子寿命和收集效率。麻省理工学院(MIT)通过低温(低于 $450\text{ }^{\circ}\text{C}$ )锗层生长技术结合退火处理(如 $500\text{ }^{\circ}\text{C}$ 退火3 h),使得探测器内部量子效率从38%跃升至57%,展现出良好的CMOS兼容性,适合后端集成工艺<sup>[52]</sup>,如图3(d)所示。华中科技大学研究团队近年来采用光诱捕[图3(e)]、锗材料侧壁掺杂[图3(f)]、反射结构[图3(g)]等方案,有效缓和PD的带宽和响应度之间的矛盾<sup>[53-55]</sup>。此外,华中科技大学研究团队还采用U形电极优化PD的寄生参数,获得带宽大于100 GHz、支持200 Gbit/s PAM-4信号探测的PD<sup>[56]</sup>,如图3(h)所示。

APD主要可以分为全硅APD和锗-硅APD。对于锗-硅APD,可以采用横向分离吸收、电荷倍增(SACM)结构,在低偏置电压下实现高效工作<sup>[57]</sup>。通过电场分布的精确控制,可以在锗中维持 $7\text{ kV/cm}$ 以上电场以实现载流子饱和速度,同时在锗-硅界面保持电场低于 $150\text{ kV/cm}$ ,从而有效抑制电离。这一设计使器件带宽超过50 GHz,在1310 nm波长下响应度达约 $3.5\text{ A/W}$ 。波导APD中的槽状结构进一步改善微弱耦合,响应度提升至 $5.26\text{ A/W}$ ,带宽约为29.5 GHz,增益带宽积(GBP)达260 GHz<sup>[58]</sup>。此外,将超薄硅倍增层(如65 nm)与锗吸收层结合,利用负微分电阻效应改善雪崩状态下的带宽制约,可以进一步将增益带宽积提升至460 GHz,增益超过15,带宽接近30 GHz,支持50 Gbit/s高速探测<sup>[59]</sup>。与PD相似,通过在APD中引入光学反射结构也可以有效缓解其带宽与响应度之间的矛盾<sup>[60]</sup>,如图3(k)所示。华中科技大学研究团队在2024年通过引入片上电感有效拓展APD的光电带宽,提高APD的增益带宽积,且其支持200 Gbit/s PAM-4信号探测<sup>[61]</sup>,如图3(l)所示。近年来,基于全硅材料的APD也得到深入研究,相比于锗-硅APD,全硅APD有效降低40%的芯片制造成本<sup>[62]</sup>。通过采用异形PN结增加PN结耗尽区与光场的交叠面积,可以有效增加全硅材料APD的响应度<sup>[63-64]</sup>。此外,利用MRR的谐振增强效应提高MRR的品质因子(Q值),全硅材料APD的响应度可以大于 $65\text{ A/W}$ <sup>[65]</sup>,如图3(m)所示。惠普实验室在2024年报道了基于全硅APD的 $8\times 160\text{ Gbit/s}$ 高速信号探测芯片,其每一个探测通道均由双环构成,以增加APD的响应度并降低通道间串扰<sup>[62]</sup>,如图3(n)所示。在集成应用方面,平衡锗-硅APD已成功集成到硅光相干接收器中。该接收器具备 $0.84\text{ A/W}$ 的高响应度和约289的增益,能够在较低的本地光功率( $-11\text{ dBm}$ )下探测25 GBaud的正交相移键控(QPSK)信号<sup>[66]</sup>。

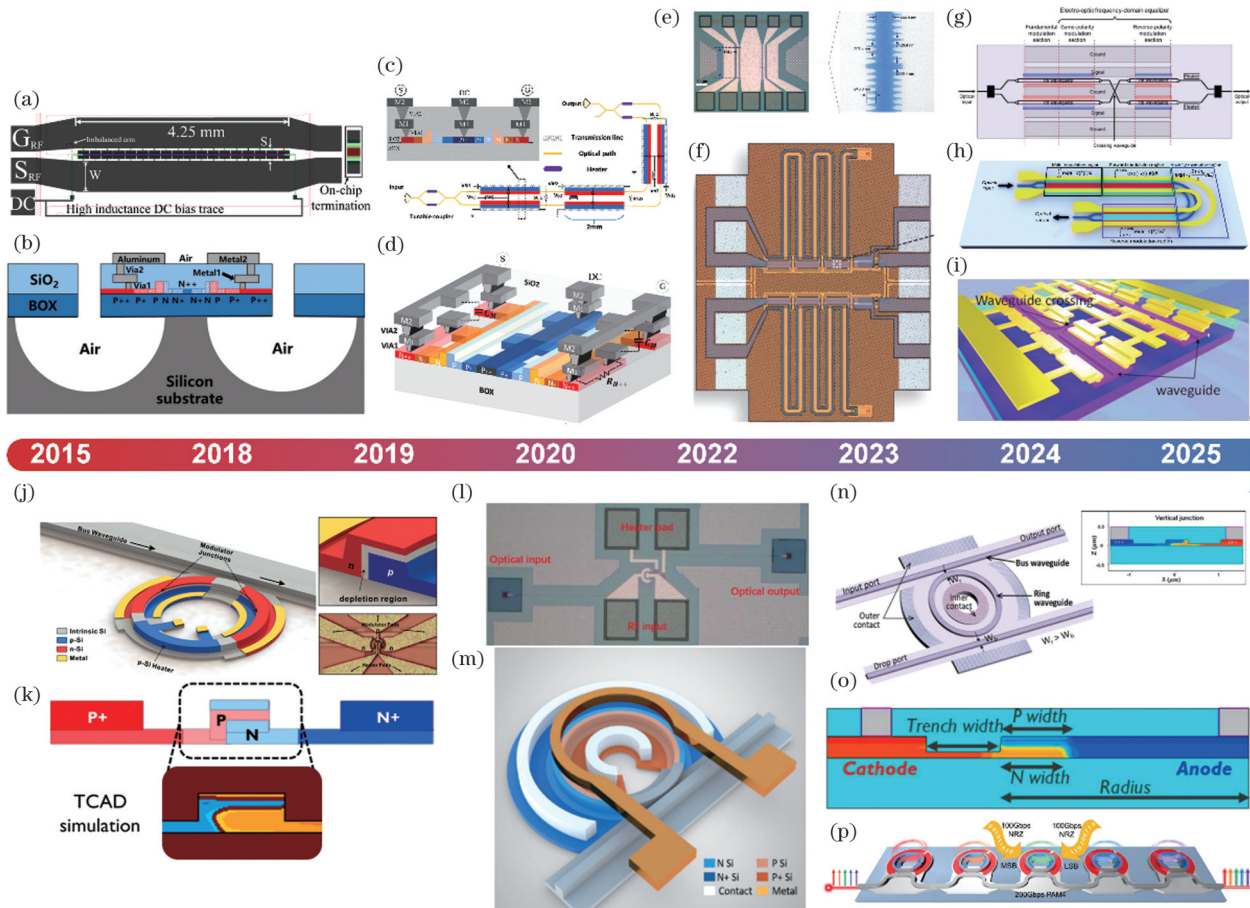


图2 硅基高速光电调制器研究进展。(a)麦吉尔大学<sup>[20]</sup>;(b)华中科技大学<sup>[21]</sup>;(c)(d)拉瓦尔大学<sup>[22-23]</sup>;(e)北京大学<sup>[24]</sup>;(f)横滨国立大学<sup>[25]</sup>;(g)日本国家信息与通信技术研究所<sup>[26]</sup>;(h)浙江大学<sup>[27]</sup>;(i)新加坡 Advanced Micro Foundry<sup>[28]</sup>;(j)英特尔公司<sup>[41]</sup>;(k)根特大学<sup>[42]</sup>;(l)国家信息光电子创新中心<sup>[45]</sup>;(m)英特尔公司<sup>[46]</sup>;(n)美国 Advanced Micro Devices 公司<sup>[48]</sup>;(o)根特大学<sup>[47]</sup>;(p)惠普实验室<sup>[38]</sup>

Fig. 2 Research progress in silicon high-speed electro-optic modulators. (a) McGill University<sup>[20]</sup>; (b) Huazhong University of Science and Technology<sup>[21]</sup>; (c)–(d) Laval University<sup>[22-23]</sup>; (e) Peking University<sup>[24]</sup>; (f) Yokohama National University<sup>[25]</sup>; (g) National Institute of Information and Communication Technology<sup>[26]</sup>; (h) Zhejiang University<sup>[27]</sup>; (i) Advanced Micro Foundry<sup>[28]</sup>; (j) Intel Corporation<sup>[41]</sup>; (k) Ghent University<sup>[42]</sup>; (l) National Information Optoelectronics Innovation Center<sup>[45]</sup>; (m) Intel Corporation<sup>[46]</sup>; (n) Advanced Micro Devices<sup>[48]</sup>; (o) Ghent University<sup>[47]</sup>; (p) Hewlett Packard Laboratories<sup>[38]</sup>

硅材料也可以和二维材料[如石墨烯、过渡金属二硫化物(TMDC)、黑磷]相结合实现探测。其凭借可调带隙、高载流子迁移率和低暗电流等优势,有望突破传统硅探测器 $1.1\ \mu\text{m}$ 的光谱响应限制,实现从近红外到中红外波段的广谱探测<sup>[67-68]</sup>。

## 2.2 硅光工艺与光电共封

硅光采用与传统 CMOS 集成电路兼容的微纳加工工艺,在硅晶圆上制备光子器件(如波导、调制器、探测器等)与电子器件(如驱动和控制电路),从而实现系统级光电集成。光电融合集成技术的快速发展经历多个阶段,主要包括混合集成、异质集成和单片集成的演进。尽管单片集成是未来技术发展的终极目标,但目前的主流方法仍以混合和异质集成为主。工艺制备是硅光集成的核心环节,涵盖材料沉积、光刻图形转移、刻蚀和表面处理等复杂步骤。单一复杂器件的工

艺流程可能超过 200 个步骤。通过工艺优化,可以在同一硅片上实现调制、探测、分束和耦合等关键光学功能的器件集成。图 4 为国外硅光流片厂家格芯(Global Foundry)的 45 nm 单片集成硅光工艺截面示意图,单片集成工艺由电学元器件部分和光学元器件部分构成,光学元器件部分包含硅和氮化硅波导层、锗探测器层、氮化钛(TiN)加热层以及双层金属连接层等。

随着硅基光电子技术从实验室走向商业化,晶圆加工平台正迈向大规模集成的关键阶段。比如,台湾积体电路制造股份有限公司(台积电)采用 65 nm 工艺节点制造硅光子集成器件,而电子集成电路(EIC)则采用更先进的 6 nm 与 4 nm 制程。通过这种工艺节点的解耦,既保证了光电子组件的高可靠性(因光子器件对工艺微缩不敏感,65 nm 已足够),又利用更先进节

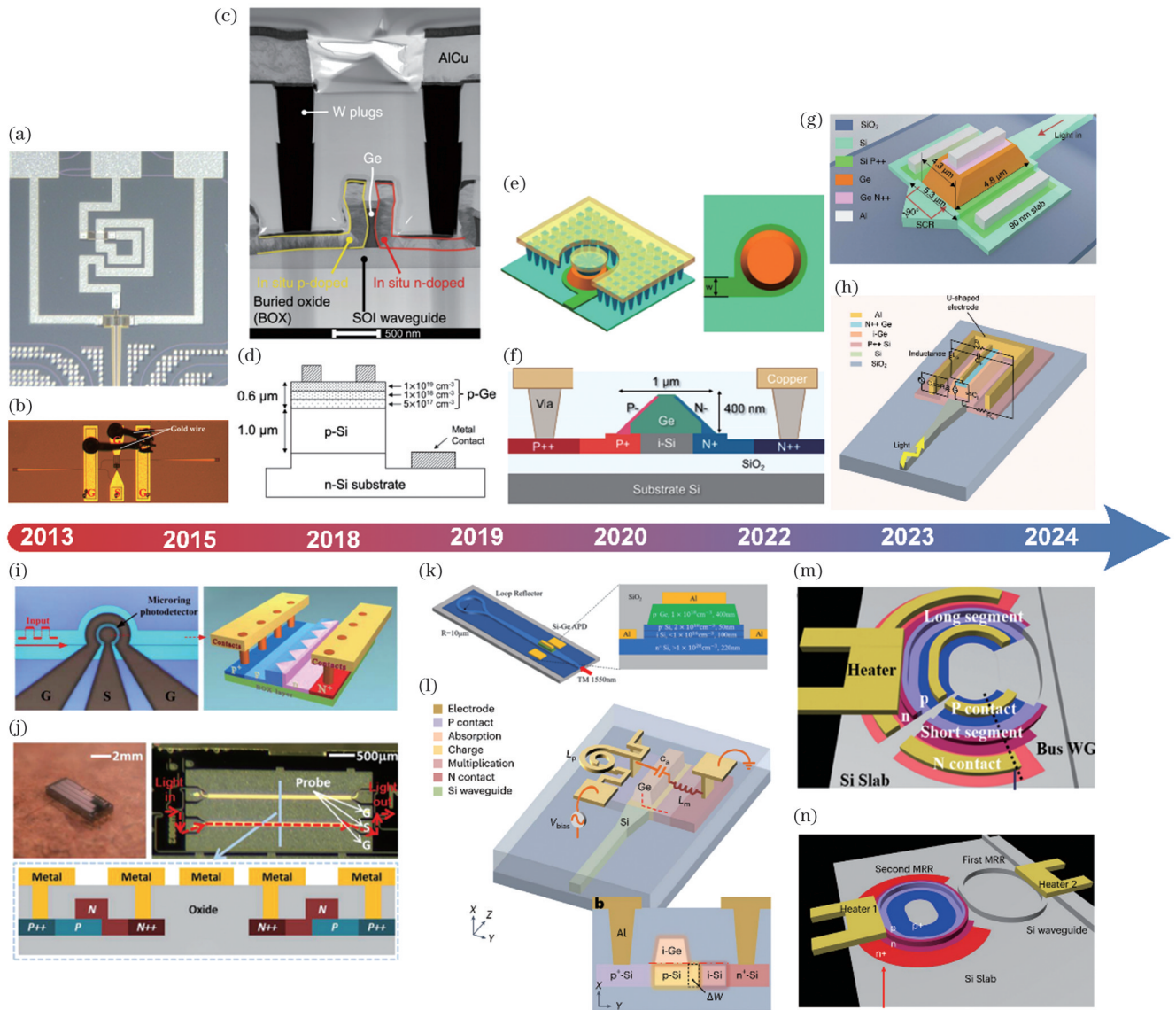


图 3 硅/锗-硅 PD 研究进展。(a)新加坡国立大学<sup>[49]</sup>;(b)华中科技大学<sup>[50]</sup>;(c)莱布尼兹高性能微电子研究所<sup>[51]</sup>;(d)麻省理工学院<sup>[52]</sup>;(e)~(i)华中科技大学<sup>[53-56,63]</sup>;(j)日本富士通公司<sup>[64]</sup>;(k)惠普实验室<sup>[60]</sup>;(l)华中科技大学<sup>[61]</sup>;(m)~(n)惠普实验室<sup>[61,65]</sup>  
 Fig. 3 Research progress in Si/Si-Ge (avalanche) PDs. (a) National University of Singapore<sup>[49]</sup>; (b) Huazhong University of Science and Technology<sup>[50]</sup>; (c) Leibniz Institute for High-Performance Microelectronics<sup>[51]</sup>; (d) Massachusetts Institute of Technology<sup>[52]</sup>; (e)~(i) Huazhong University of Science and Technology<sup>[53-56,63]</sup>; (j) Fujitsu<sup>[64]</sup>; (k) Hewlett-Packard Laboratories<sup>[60]</sup>; (l) Huazhong University of Science and Technology<sup>[61]</sup>; (m)~(n) Hewlett-Packard Laboratories<sup>[61,65]</sup>

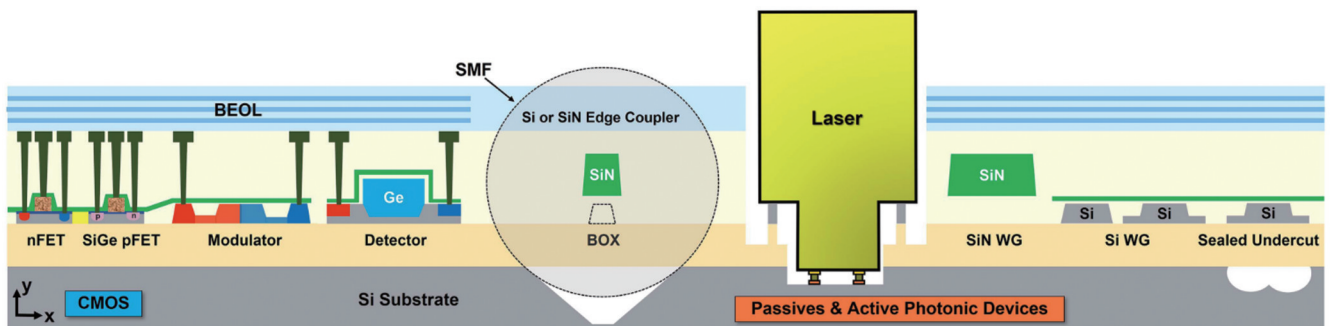


图 4 格芯 45 nm 单片集成工艺截面示意图<sup>[69]</sup>  
 Fig. 4 Schematic cross-section of Global Foundries 45 nm monolithic integration process<sup>[69]</sup>

点提升微电子电路的性能和能效。此举推动了硅光技术在高性能计算和AI领域的落地与发展。在规模化集成发展上,异质集成日益受到重视。通过直接晶圆键合技术可以将III-V族材料与硅衬底结合,有效绕过外延生长中的晶格失配问题<sup>[70]</sup>。

CPO是一种先进的光电子集成技术,核心思想是将光芯片(光引擎)与电芯片(如交换芯片)在同一封装体内实现紧密集成,形成共封装系统。传统光模块通常是光学器件和电芯片分开封装,通过电缆或印刷电路板(PCB)连接,信号传输路径长,功耗和延迟较高。CPO通过缩短光模块与电芯片之间的物理距离,极大降低信号传输损耗和功耗,同时提升互连带宽密度和系统性能。

从可插拔光模块向CPO的转变是光电融合的一个关键步骤。日本电报电话公司(NTT)为数字相干光收发器设计的第二代光电子器件利用了硅光技术,与分立解决方案相比,将尺寸和功耗降低40%<sup>[71]</sup>。台积电推出的紧凑型通用光子引擎(COUPÉ)<sup>[72]</sup>,通过3D混合键合实现电芯片与光芯片的高密度集成。COUPÉ的第一代模块单颗芯片可实现1.6 Tbit/s的高速光互连带宽。台积电进一步利用CoWoS(Chip-on-Wafer-on-Substrate)中介层集成技术,将多颗COUPÉ模块集成于同一封装中,目标带宽提升至12.8 Tbit/s。

光电深度融合有望突破当前通信和计算系统的性能瓶颈。通过持续的技术创新和跨界协作,人们正逐步走向一个更加高效、智能和可持续的信息技术新时代。这一技术演进不仅是工程技术的突破,更是人类信息处理能力的重大飞跃。

### 3 硅基光通信:从器件到系统

#### 3.1 光通信对硅基光电子的需求

近年来,全球数字化进程的加速催生数据规模的爆炸性增长。截至2024年,互联网用户规模已突破50亿<sup>[73]</sup>,物联网终端接入量超过188亿节点<sup>[74]</sup>,加之AI大模型训练所需的数据吞吐量呈指数级上升<sup>[75-76]</sup>,推动全球数据流量以24.4%的年复合增长率持续攀升,国际数据公司(IDC)最新预测显示<sup>[77]</sup>,2028年全球数据总量将达到393.8 ZB,较2023年将实现倍增,如图5(a)和(b)所示。为应对如此庞大的数据洪流,光纤通信系统通过波分复用、偏振复用、空分复用、波段扩展及高阶调制等技术革新,单光纤传输容量已从几十Gbit/s持续增长到数百Tbit/s,在实验室研究中更是突破Pbit/s量级<sup>[78]</sup>。此外,随着通信需求的不断提升,全球通信技术的耗电量也不断增加,预计到2030年,将消耗全球电力的20%<sup>[79]</sup>,如图5(c)所示,其中,数据中心到2030年预计将消耗全球电力的3%~13%<sup>[80]</sup>。在此背景下,光通信技术正面临两方面需求:一方面需满足数据中心内部Tbit/s级互连带

宽与 $\mu\text{s}$ 级时延要求,另一方面需实现城域/骨干网在百公里至千公里级传输距离下数百Tbit/s至Pbit/s量级的长距离大容量通信需求。这种双重需求为硅基光电子技术的产业化应用提供历史性机遇。硅基光电子技术依托其高集成度和与CMOS工艺相兼容的优势,可在单一芯片上实现多通道多功能光电器件的集成,突破传统分立式光模块与InP光模块在集成密度与功耗等方面的技术瓶颈,如图5(d)所示。通过集成低损耗波导、高速调制单元、高速探测单元及其他所需的无源器件,结合异质键合等先进工艺,硅基光电子芯片有望同时满足数据中心应用场景和城域及长距离传输领域的互连需求,降低多维复用系统的体积和成本,提升多维度复用系统的稳定性,降低能耗。目前,基于硅基光电子芯片的光模块产能正在持续提升中,未来有望成为光互连领域的主流解决方案<sup>[81]</sup>。此外,硅基光电子技术的发展还为CPO等新型架构奠定基础,有力支撑高带宽、低时延、可重构的新一代智能光网络建设<sup>[82-83]</sup>。

#### 3.2 多通道/密集波分复用应用进展

在光通信系统中,单个器件的带宽受限于物理机制,难以独立支撑快速增长的数据传输需求。为此,硅基光电子器件的小型化与阵列化设计成为突破容量限制的核心路径:通过亚微米级波导工艺与高密度集成技术,将调制单元、探测单元及所需无源器件等压缩至毫米级芯片,提升单位面积带宽密度,同时结合低功耗相移器与片上驱动电路,保持较低的功耗。基于光电子技术的面向高吞吐量的片上光互连为数据中心、算力网络等场景提供高密度、低功耗的光互连基础架构。因此,近年来,基于硅基光电子技术的多通道复用技术成为研究热点,如表1所示。最为直接的多通道方案是并行单模光纤(PSM),该方案无需引入额外的通道复用器件,可以有效提升通信容量。北京大学Shi等<sup>[84]</sup>在2023年开发了基于PSM的16通道硅基MZM发射机,该发射机采用引线键合集成电驱动芯片,总速率达800 Gbit/s,功耗为5.35 pJ/bit,如图6(a)所示。美国Nubis Communications公司在2025年开发了基于PSM的1.6 Tbit/s线性驱动光引擎,该光引擎省去了数字信号处理(DSP)芯片,有效降低能耗<sup>[85]</sup>,如图6(b)所示。虽然PSM方案可以有效提升通信容量,但随着传输距离的逐渐增加,PSM方案中光纤资源的成本会逐渐上升,为降低成本,波分复用技术得到广泛研究。哥伦比亚大学的Daudlin等<sup>[86]</sup>在2025报道了基于3D集成的80通道硅基PSM-WDM收发机,其总速率为800 Gbit/s,能耗为120 fJ/bit,如图6(c)所示。天津工业大学Zhang等<sup>[87]</sup>在2024年报道了8通道Lan(Local Area Network)-WDM收发器,实现10 km光纤1.35 Tbit/s PAM-4信号传输,如图6(d)所示。宾夕法尼亚大学Omirezakhov等<sup>[88]</sup>在2024年报道单片集成8通道波分复用硅基MRM发射机,采用金属-氧化

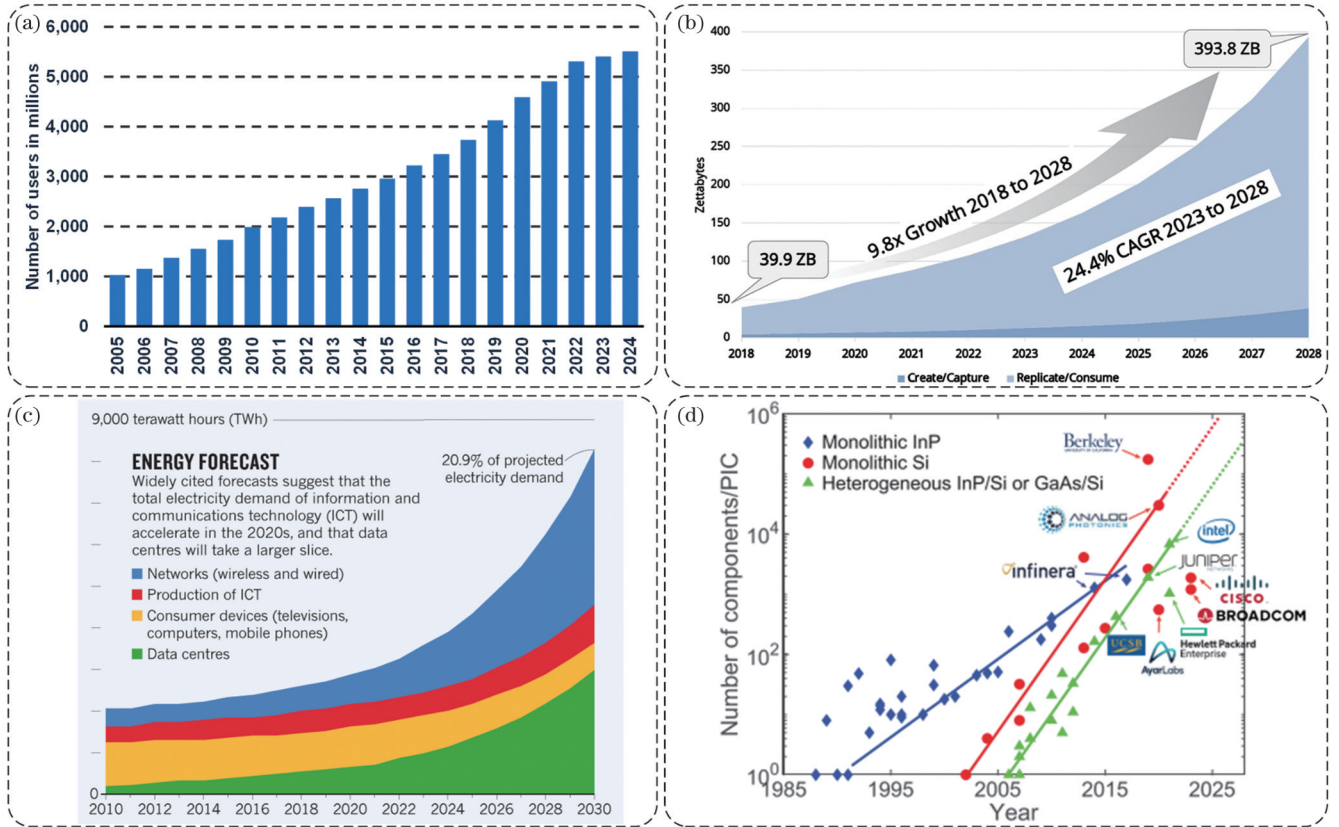


图5 硅基光电子在光通信中的应用。(a)全球互联网用户增长图<sup>[73]</sup>;(b)全球IDC数据领域预测<sup>[77]</sup>;(c)通信技术耗电量持续增长<sup>[79]</sup>;(d)不同材料集成光学平台集成器件数演进<sup>[82]</sup>

Fig. 5 Applications of silicon photonics in optical communications. (a) Global internet user growth<sup>[73]</sup>; (b) worldwide IDC global DataSphere forecast<sup>[77]</sup>; (c) continued growth of power consumption in communication technologies<sup>[79]</sup>; (d) evolution of integrated device count across different material photonic integration platform<sup>[82]</sup>

物-半导体(MOS)电容进行MRM的谐振波长接近零功耗调节,单通道速率为32 Gbit/s,带宽密度为3.3 Tbit/(s·mm<sup>2</sup>),总能效为328 fJ/bit,如图6(e)所示。中国科学院西安光学精密机械研究所Xue等<sup>[89]</sup>在2024年成功研制了4×112 Gbit/s偏振不敏感波分复用接收机,该接收机采用偏振分束光栅耦合器与双环滤波器,实现偏振不敏感操作,单通道速率达到

112 Gbit/s,在100 Gbit/s PAM-4信号测试中,系统灵敏度达到-5.6 dBm,如图6(f)所示。哥伦比亚大学的Rizzo等<sup>[90]</sup>在2023利用Kerr频率梳与级联微环架构实现512 Gbit/s的链路(32×16 Gbit/s),能效低于1 pJ/bit,如图6(g)所示。基于硅基MRM的多通道发射机具有更低的功耗和更高的带宽密度,为数据中心与AI集群提供低功耗、高兼容性的光互连解决方案。

表1 多通道/密集波分复用研究进展

Table 1 Research progress in multi-channel/dense wavelength division multiplexing

Reference	Scheme	Number of channels	Single-channel rate/(Gbit/s)	Aggregate rate/(Gbit/s)	Bandwidth density	Power consumption/(pJ/bit)	Integration method
[84]	PSM	16	50	800	-	5.35	Wire bonding
[85]	PSM	16	100	1600	0.246 Tbit/(s·mm)	3.9	Flip chip
[86]	PSM+WDM	80	10	800	5.3 Tbit/(s·mm <sup>2</sup> )	0.12	Flip chip
[87]	WDM	8	-	1350	-	2.13	None
[88]	WDM	8	32	256	3.3 Tbit/(s·mm <sup>2</sup> )	0.328	Monolithic
[89]	WDM	4	112	448	-	1.34	Wire bonding
[90]	WDM	32	16	512	0.465 Tbit/(s·mm)	<1	None

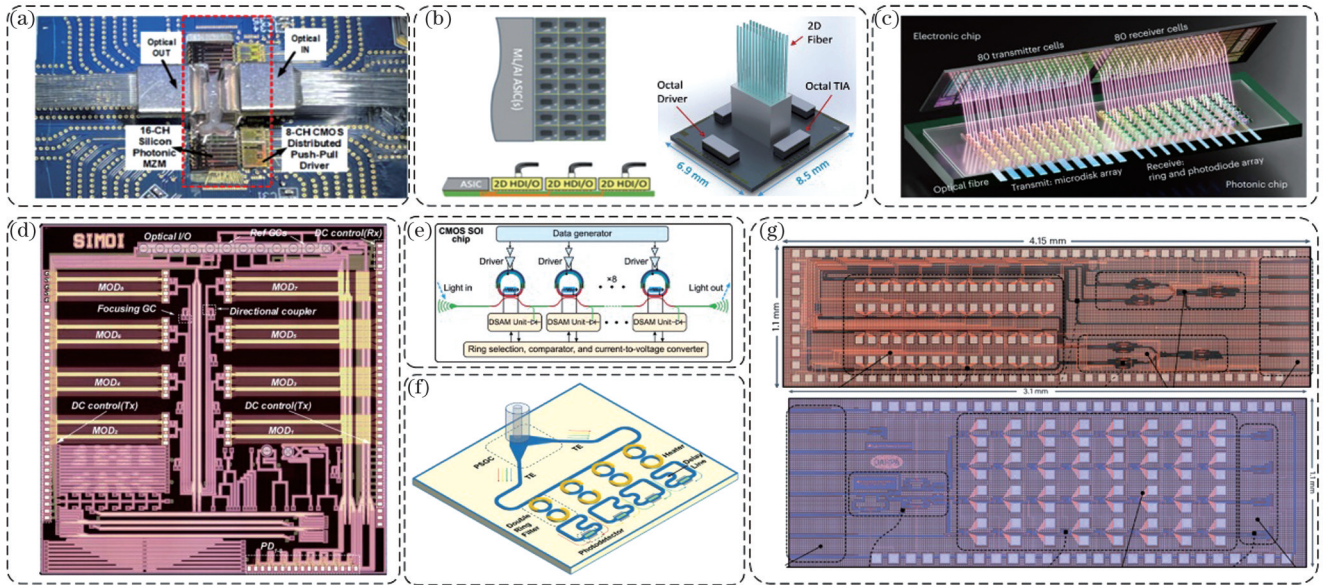


图 6 硅基多通道光发射和接收芯片研究进展。(a)低功耗 16 通道并行硅基 MZM 发射机<sup>[84]</sup>；(b)线性驱动的高密度 16 通道并行硅基 MZM 发射机<sup>[85]</sup>；(c)3D 集成的低功耗大带宽 80 通道波分复用硅基 MRM 发射机<sup>[86]</sup>；(d)基于光栅的 8 通道 Lan-WDM 硅基收发机<sup>[87]</sup>；(e)单片集成的 8 通道波分复用硅基 MRM 发射机<sup>[88]</sup>；(f)基于双环滤波器的 4 通道偏振不敏感波分复用硅基接收机<sup>[89]</sup>；(g)克尔光梳驱动的 32 通道波分复用硅基 MRM 收发机<sup>[90]</sup>

Fig. 6 Research progress in silicon photonic multi-channel optical transmitter and receiver chips. (a) Low-power 16-channel parallel silicon MZM transmitter<sup>[84]</sup>; (b) linearly driven high-density 16-channel parallel silicon-based MZM transmitter<sup>[85]</sup>; (c) 3D-integrated low-power and high-bandwidth 80-channel WDM silicon-based MRM transmitter<sup>[86]</sup>; (d) grating-based 8-channel Lan-WDM silicon-based photonic transceiver<sup>[87]</sup>; (e) monolithic 8-channel WDM transmitter based on silicon-based MRM<sup>[88]</sup>; (f) 4-channel polarization-insensitive WDM silicon-based receiver based on dual-ring filters<sup>[89]</sup>; (g) Kerr comb-driven 32-channel WDM silicon-based MRM transceiver<sup>[90]</sup>

### 3.3 应用实例与典型系统

当前,全球每年生成的数据中超过 70% 的流量产生于数据中心内部<sup>[91-92]</sup>,这一趋势驱动硅基光通信技术成为数据中心互连的核心解决方案。现代超大规模数据中心普遍采用叶脊网络架构<sup>[91]</sup>,该架构通过多层交换机实现服务器集群的高效互连,具备可扩展性强、传输延迟低等特点,如图 7(a)所示。在此架构中,服务器与叶交换机、叶交换机与脊交换机之间通过高速光模块构建光互连通道<sup>[93]</sup>。目前主流商用光模块已从 400 Gbit/s 向 800 Gbit/s 演进,2025 年 Acacia、博通、海信宽带、新易盛 (Eoptolink)、光迅科技、讯石科技、立讯技术等光模块供应商均在 OFC (Optical Fiber Communication) 会议上展示了 1.6 Tbit/s 光模块原型<sup>[94]</sup>,而 3.2 Tbit/s 模块研发也进入关键阶段。硅基光模块产业链已形成完整生态:上游制造由高意 (Coherent)、旭创 (InnoLight)、新易盛、Lumentum 及 Fabrinet 等企业主导,下游应用则集中于微软、Meta、英伟达、亚马逊服务 (AWS)、谷歌等云计算巨头<sup>[95]</sup>,如图 7(b)所示。

数据中心内部的交换机带宽从 2010 年的 640 Gbit/s 增加至 2024 年的 102.4 Tbit/s,带宽增长 160 倍<sup>[83]</sup>,如图 8(a)所示。随着交换机带宽的不断提高,传统可插拔光模块面临功耗与信号完整性的双重瓶颈,为此,

可以有效降低功耗、提高信号完整性的硅光芯片与交换机芯片 CPO 技术将成为破局方向,如图 8(b)所示。博通 (Broadcom) 于 2021 年推出首款 CPO 交换机 Tomahawk 4-Humboldt,带宽达 26.5 Tbit/s,推动 CPO 的早期技术探索。基于此,博通随后又发布了带宽为 51.2 Tbit/s 的 CPO 交换机 Tomahawk 5,其单通道速率为 100 Gbit/s,紧接着,在 2025 年又发布了第三代单波 200 Gbit/s 的 CPO 交换机,其 CPO 交换机架构<sup>[93, 96]</sup>如图 8(c)所示。英伟达 2025 年发布的 Quantum-X 系列 CPO 交换机采用先进的 3D 异构封装技术,实现高达 115.2 Tbit/s 的总交换容量<sup>[4]</sup>,如图 8(d)所示。该交换机集成 4 颗基于台积电 4 nm 工艺的交换芯片,以及 18 个硅光子引擎,每个引擎包含 8 个支持 200 Gbit/s 信号调制的 MRM。与传统可插拔光模块方案相比,Quantum-X 交换机在多项关键性能指标上取得重大突破:能源效率提升 3.5 倍,信号完整性提升 63 倍,大规模组网的可靠性提升 10 倍,部署速度提升 1.3 倍。此外,该交换机通过将光子引擎与专用集成电路 (ASIC) 芯片直接 3D 封装,大幅减少所需的光纤数量和外部激光器数量 (减少 75%),并采用液冷设计以应对高密度集成带来的散热挑战。整体架构标志着数据中心从传统“电交换、光互连”向“光电深度融合”新范式的转变。

随着全球数据流量的激增,光通信的容量也需要

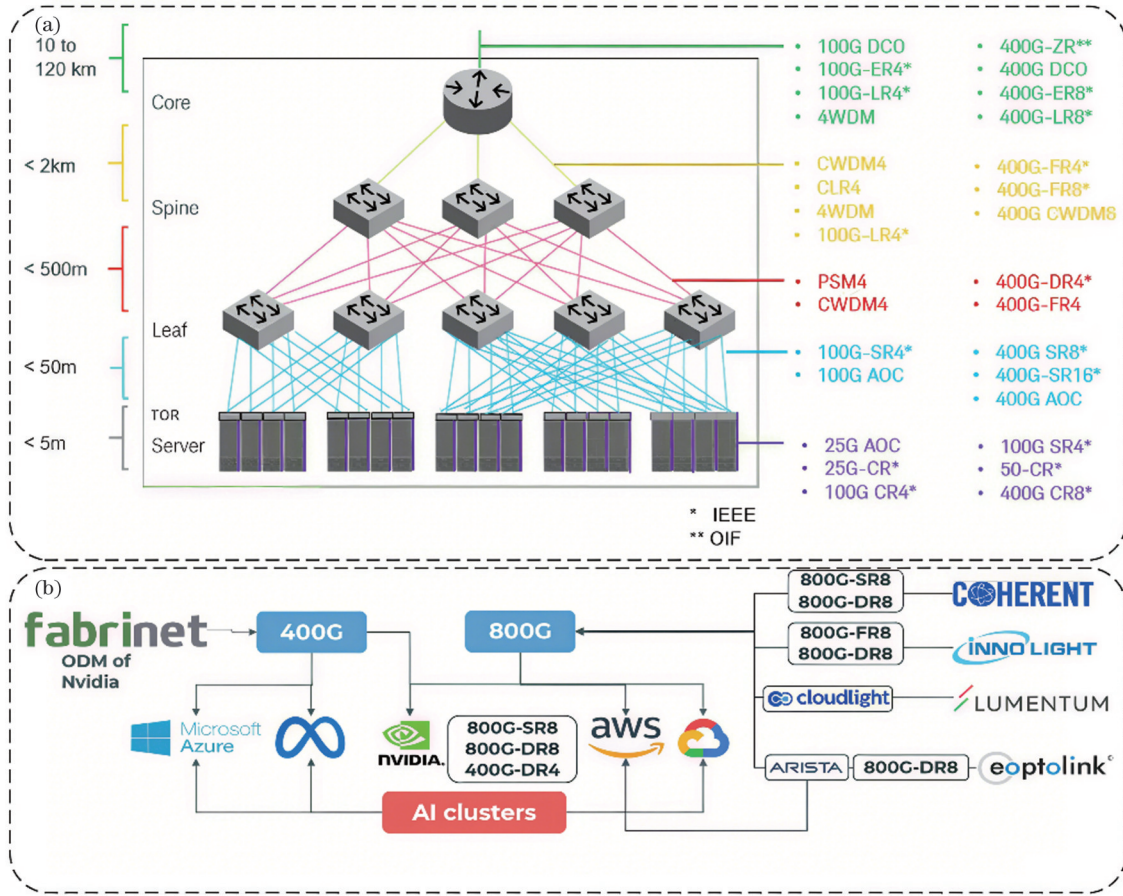


图 7 大规模数据中心内互连。(a)典型的超大规模数据中心网络架构<sup>[91]</sup>; (b)2024年光模块供应链<sup>[95]</sup>

Fig. 7 Interconnection within large-scale data centers. (a) Typical hyperscale data center network architecture<sup>[91]</sup>; (b) optical module supply chain in 2024<sup>[95]</sup>

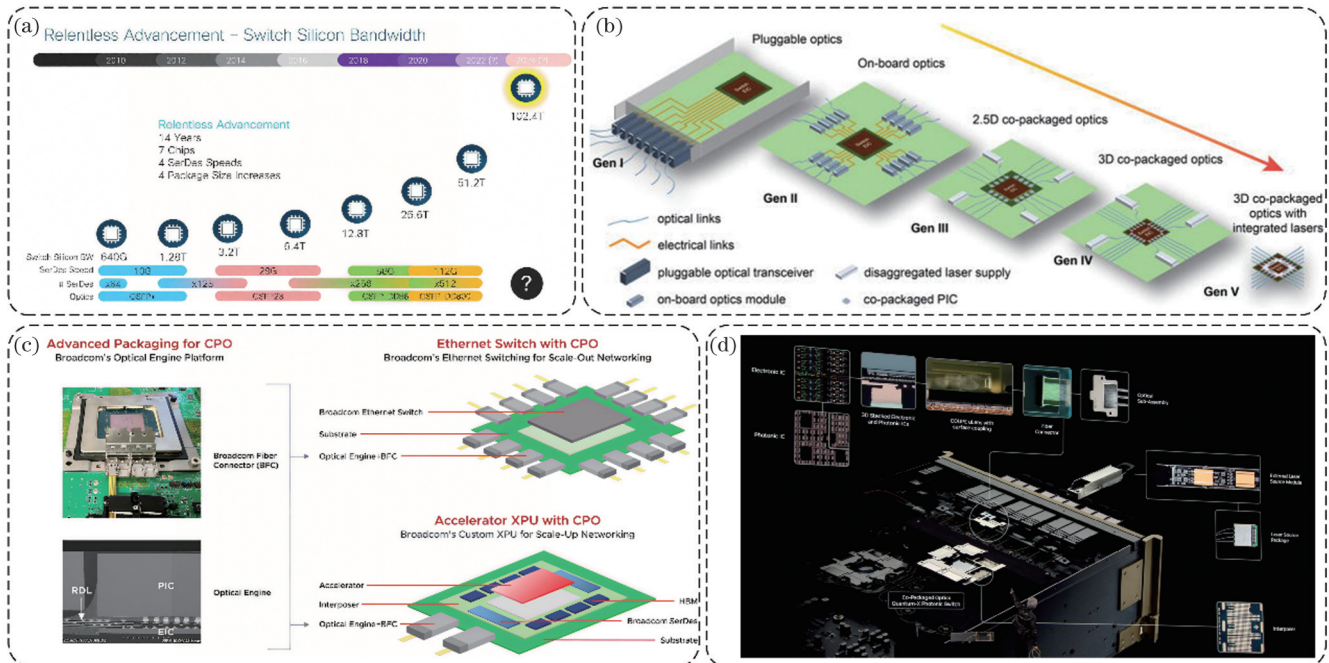


图 8 CPO 行业进展。(a)数据中心交换机带宽演进图<sup>[83]</sup>; (b)光互连技术演进图<sup>[82]</sup>; (c)博通 CPO 交换机<sup>[96]</sup>; (c)英伟达 CPO 交换机<sup>[4]</sup>

Fig. 8 Advances in CPO. (a) Bandwidth evolution trends in data center switches<sup>[83]</sup>; (b) progression of optical interconnect technologies<sup>[82]</sup>; (c) Broadcom CPO switches<sup>[96]</sup>; (c) NVIDIA CPO switches<sup>[4]</sup>

持续提升,这需要持续提升单个器件的带宽。采用频谱效率更高的调制格式,如 PAM-6、PAM-8 等强度调制直接检测格式和 16 QAM 至 128 QAM 等相干调制格式,并通过波长、模式、空间、偏振等进行多维度复用以满足光通信对于容量的需求。在光模块中采用偏振复用技术,可以有效实现通信容量的倍增。Aloe 半导体公司在 2025 年提出 425 Gbit/(s·λ)(λ 为波长)双偏振强度调制直接检测 (IMDD) 硅基收发机,通过

Stokes 参数偏振跟踪算法实现 2×212.5 Gbit/s PAM-4 信号稳定传输的 1 km 光纤,为短距互连提供新的大容量低成本方案<sup>[97]</sup>,如图 9(a)所示。英属哥伦比亚大学在 2023 年开发双偏振相干硅基接收器,支持单波 528 Gbit/s 16 QAM 调制信号接收<sup>[98]</sup>,如图 9(b)所示。这些成果从调制格式优化、算法增强、多维复用及器件集成等维度,为下一代超高速数据中心互连提供多样化的技术路径与实验基准。

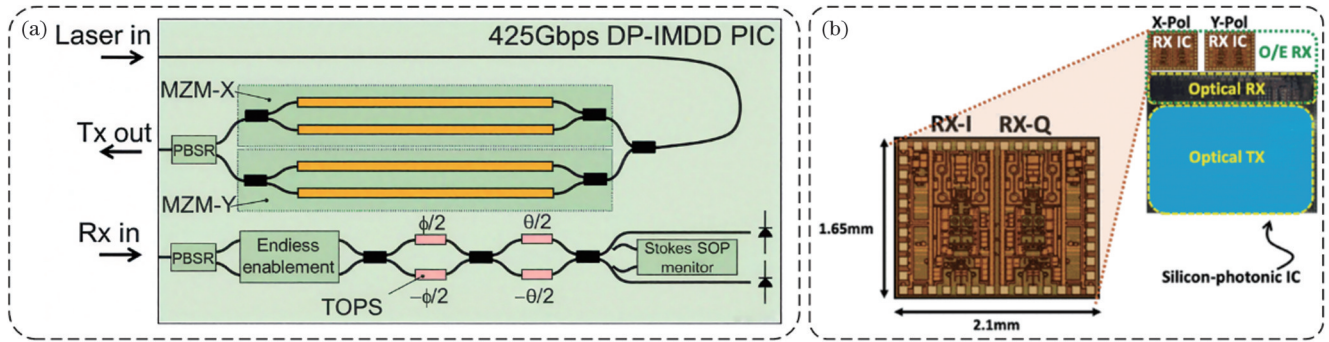


图 9 新一代 PAM/相干通信实验系统。(a) Aloe 半导体公司基于双偏振的单波 425 Gbit/s PAM-4 硅基发射机<sup>[97]</sup>; (b) 英属哥伦比亚大学单波 528 Gbit/s 双偏振相干接收器<sup>[98]</sup>

Fig. 9 Next-generation PAM/coherent communications experimental systems. (a) Aloe Semiconductor's single-wave 425 Gbit/s dual-polarization PAM-4 silicon transmitters<sup>[97]</sup>; (b) University of British Columbia's single-wave 528 Gbit/s dual-polarization coherent receiver<sup>[98]</sup>

### 3.4 未来发展趋势

硅基光电子技术在光通信领域正迎来高速发展阶段,朝着“更高、更低、更密”的方向演进,如图 10 所示。这一技术革新与光电融合的深度结合,正在重塑数据传输与处理的基础架构,为后摩尔时代信息技术提供

新的增长动力。

在传输速率方面,电光调制器的带宽已经从传统几十 GHz 增加至 100 GHz 以上。结合多通道波分复用技术和高阶调制格式,单根光纤的传输容量正持续突破新的极限。最新一代高速光模块通过先进制程的

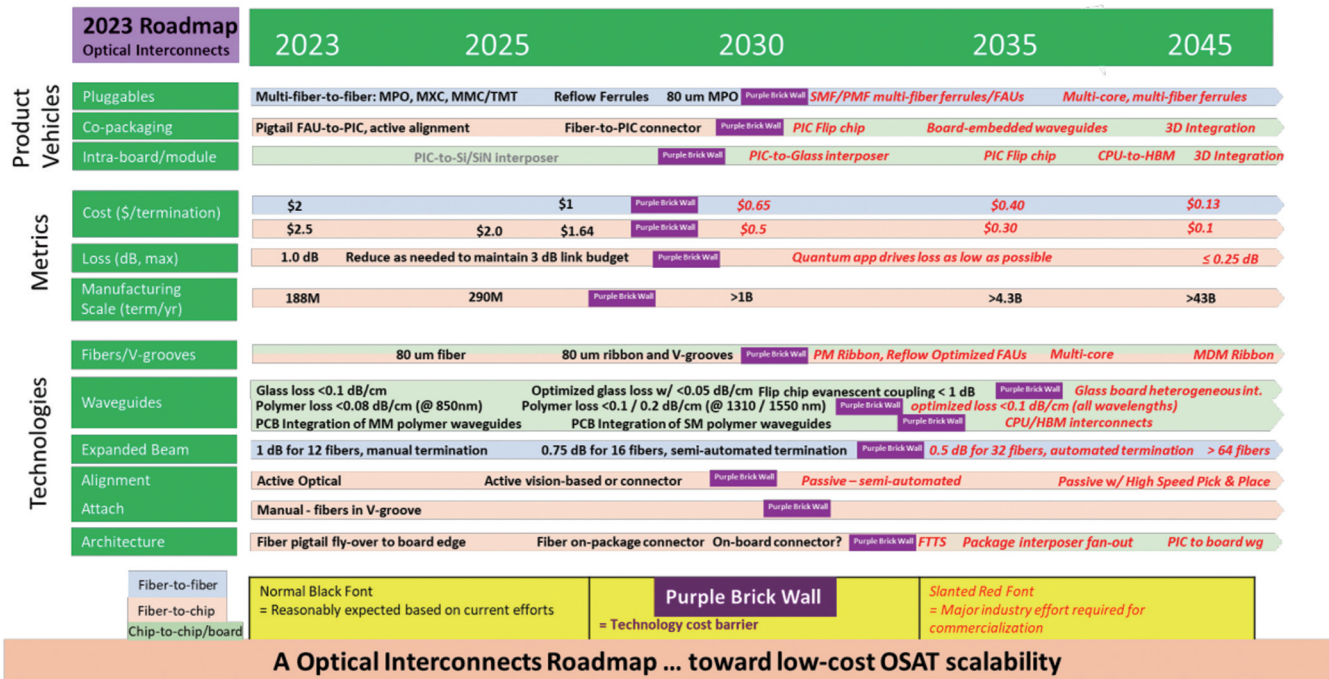


图 10 光互连发展路线图<sup>[99]</sup>

Fig. 10 Optical interconnect development roadmap<sup>[99]</sup>

DSP 芯片与硅光技术的深度融合,实现更高的单波传输速率,并在功耗方面较前代产品有显著降低,有效支撑 AI 算力集群的扩展需求。在功耗与集成度优化方面,CPO 技术发挥着核心作用。通过将光引擎与电芯片直接集成封装,系统能效得到大幅提升。光电协同设计推动驱动电压降低至 1 V 以下,配合线性驱动架构的普及,在保证高速传输的同时实现更高的能效。硅光技术显著提升带宽密度,同时大幅降低功耗,相较传统电互连方式具有明显优势。NTT 研究预测,光电融合技术的广泛应用到 2030 年将实现 40% 以上的能源节约。材料与制造工艺创新同样发展迅速,低损耗玻璃波导( $<0.05$  dB/cm)和柔性聚合物波导为板级和短距离光互连提供坚实基础。光束连接器通过透镜阵列实现低于 0.5 dB 的插入损耗,提升连接效率和系统稳定性。低温制造工艺与自动化贴片技术的应用有效降低制造成本,使光互连成本逼近每端口 1 美元,为大规模商业化铺平道路。

数据中心内部高速互连已成为最迫切需要光电融合的应用场景。随着 AI 大模型训练对算力需求的指数级增长,传统电互连在带宽密度和功耗方面接近物理极限。CPO 技术通过将交换 ASIC 芯片与硅光引擎在同一高速主板上协同封装,实现光学收发器与处理芯片的近距离集成。2025 年,英伟达计划推出基于 CPO 技术的 Spectrum-X 和 Quantum-X 光子网络交换机,标志着光电融合从概念走向大规模商用。在芯片间互连环节,光电融合通过 chiplet 架构实现异构集成,将光子互连引入芯片内部和芯片间通信。台积电的 EPIC-BOE 平台采用 300 mm 晶圆工艺,集成 N65 及更先进节点的光子集成电路与 N7 及更先进节点的电子集成电路,为下一代 AI 加速器提供高带宽、低延迟的片间互连解决方案。嵌入式波导技术实现 CPU 与高带宽存储器之间的超高速互连,满足高性能计算对低延迟、高带宽互连的需求。在集成度提升方面,光电融合打破了传统分立式光模块的空间限制。通过将激光器、调制器、探测器等光电器件与 DSP、跨阻放大器(TIA)、驱动器等电子芯片集成在同一硅基平台上,系统体积大幅缩减,不仅提高系统可靠性,还通过减少器件间互连损耗提升整体性能。这种高度集成使构建百万图形处理单元(GPU)级别的 AI 工厂成为可能,为大规模并行计算提供物理基础。

展望未来,光电融合技术正朝着更高集成度、更低功耗和更广应用的方向发展。三维集成技术通过垂直堆叠实现光电器件的立体布局,有望突破平面集成的密度限制;基于 chiplet 的光电协同设计将实现更灵活的系统架构。随着英伟达、英特尔、台积电等巨头深度参与,光电融合生态系统日趋完善。OSFP-XD、COBO 等行业联盟积极推动高速小型化模块和 CPO 接口的兼容,促进产业链上下游协同发展,加速技术产业化。随着技术不断成熟和产业链日益完善,光电融合将成为

推动后摩尔时代信息技术革新的核心驱动力,为构建高效、智能、可持续的数字基础设施奠定坚实基础。

## 4 硅基光感知:从相干激光雷达到多模态传感

### 4.1 激光雷达需求与挑战

近年来,以无人驾驶、智能机器人、增强/虚拟现实(AR/VR)为代表的智能技术革命,正深刻改变着人类与物理世界的交互方式。这些新兴技术的核心需求之一,是系统必须能够构建对复杂动态环境的高保真数字孪生,从而实现精准感知与实时响应。这种需求推动感知系统突破传统的分辨率与精度物理极限,催生激光雷达等高性能传感技术的快速发展。

权威市场研究机构 Yole Développement 的最新分析报告显示<sup>[100]</sup>,全球激光雷达市场规模预计将以 38% 的年复合增长率持续扩张,到 2029 年市场规模有望达到 363.2 亿美元。这一数据不仅反映无人驾驶、智能机器人、AR/VR 等领域对高精度感知能力的强烈需求,也说明激光雷达技术正成为智能感知系统的关键支撑。激光雷达通过发射激光束并接收反射信号,能够实现对环境的三维高精度扫描,生成丰富的空间信息,为自动驾驶车辆提供实时环境感知,保障行车安全,同时也为机器人导航、增强现实场景构建等应用提供基础数据。

然而,要提升激光雷达的感知能力,尤其是在分辨率和精度方面,技术路线面临诸多挑战。以基于光学相控阵技术的激光雷达为例,其感知性能与相控阵的维度密切相关。具体来说,探测点数与阵元数量呈正相关,阵元数量越多,激光雷达的分辨率和扫描精度越高,能够捕捉到更细微的环境细节。但这也意味着需要更大规模的芯片集成,集成更多的光学元件和电子控制单元。大规模芯片集成不仅对半导体制造工艺提出更高要求,还带来成本控制的巨大挑战。芯片尺寸的扩大和复杂度的提升,往往导致制造良率下降和生产成本上升,这在一定程度上限制激光雷达的普及和应用推广。此外,光学相控阵技术本身对材料、散热、功耗管理等方面也提出严苛要求,需要跨学科的技术创新和系统优化。

为应对这些挑战,行业内正在积极探索多种技术路径,包括采用硅光技术、优化芯片设计架构、提升制造工艺精度,以及结合 AI 算法提升数据处理效率等。同时,模块化设计和标准化生产也成为降低成本、加速市场应用的重要手段。未来,随着技术的不断成熟和产业链的完善,激光雷达有望实现更高集成度、更低成本和更强性能,推动智能感知系统迈向新的高度。

### 4.2 调频连续波与光学相控阵

典型的 FMCW 激光雷达系统主要由扫频光源模块、光束偏转模块、相干接收模块以及数据采集与分析模块构成。扫频光源模块通常由波形发生器和可调谐

激光器组成,用于产生线性调频的连续波光信号;光束偏转模块则负责将激光束精确地扫描至待测目标,实现对空间目标的成像;相干接收模块通过平衡 PD,将本振光与目标反射回来的回波信号进行相干接收,获取包含目标距离和速度信息的拍频信号;数据采集与分析模块则对拍频信号进行频谱分析,从中提取目标的距离和速度信息。其基本原理是:系统发射线性 FMCW,经过目标反射后返回接收端,由于光的传播存在时间延迟,本振光与回波信号之间产生频率差,通过拍频处理和 PD 提取该频率差即可获得目标的距离信息,同时结合多普勒效应可进一步测量目标速度。线性 FMCW 的稳定性和高效的相干检测技术,是 FMCW 激光雷达实现高精度测距与测速的关键。

与目前商用激光雷达中常用的脉冲飞行时间 (TOF) 方案相比, FMCW 方案具有以下优点<sup>[101]</sup>: 由于 FMCW 基于相干检测,该传感系统不会受到附近其他激光雷达系统的干扰,并且只需较低的峰值光功率;除距离信息外,还能通过多普勒效应获取速度信息;此外,该方案的测距精度主要取决于拍频频率,而拍频频率是通过多次采样并进行快速傅里叶变换获得的,因此 FMCW 更易实现高精度测距。

光学相控阵是一种新兴的固态光束偏转技术,其核心原理与传统微波相控阵相似。光学相控阵利用集成波导结构引导光波传播,通过精确调节阵列中各发

射单元的相位,实现光束的相干叠加与远场方向的动态控制,从而实现光束的成形与偏转。其数学基础是相控阵中各单元波前的幅度和相位分布与自由空间远场辐射角谱之间的傅里叶变换关系,这类似于 DSP 中时域与频域的对对应关系。理论上,只需将多个相干光源的相位精确对齐,发射的光波便会在远场产生特定角度的相长干涉,实现定向光束输出。虽然无线电频段的相控阵技术已有百年以上的发展历史,但光学相控阵技术直到 20 世纪 80 年代末才首次实现光束相位调制的偏转,这主要受限于当时的微纳制造工艺和集成技术<sup>[102]</sup>。

近年来,随着微纳加工技术的进步和光子集成技术的发展,光学相控阵在激光雷达、空间光通信和光学开关等领域展现出快速发展的态势,具备响应速度快、结构紧凑、控制灵活等优势,成为高性能光束控制的重要技术路径。如图 11 所示,典型光学相控阵的核心功能模块包括激光的产生或输入、功率分配、光场的幅度与相位调控,以及激光的相干发射。光学相控阵的关键特性在于能够对各相干阵元的相位进行可编程控制,实现光束的精确成形与偏转。实现相位编程的主要技术手段包括:无源色散阵列中的波长调谐,即通过调节激光波长实现相位变化;有源阵列中的光场调控方法,如热光效应、电光效应和液晶调制等,直接调节光场的相位或幅度。根据相控阵的空间维度(即一维

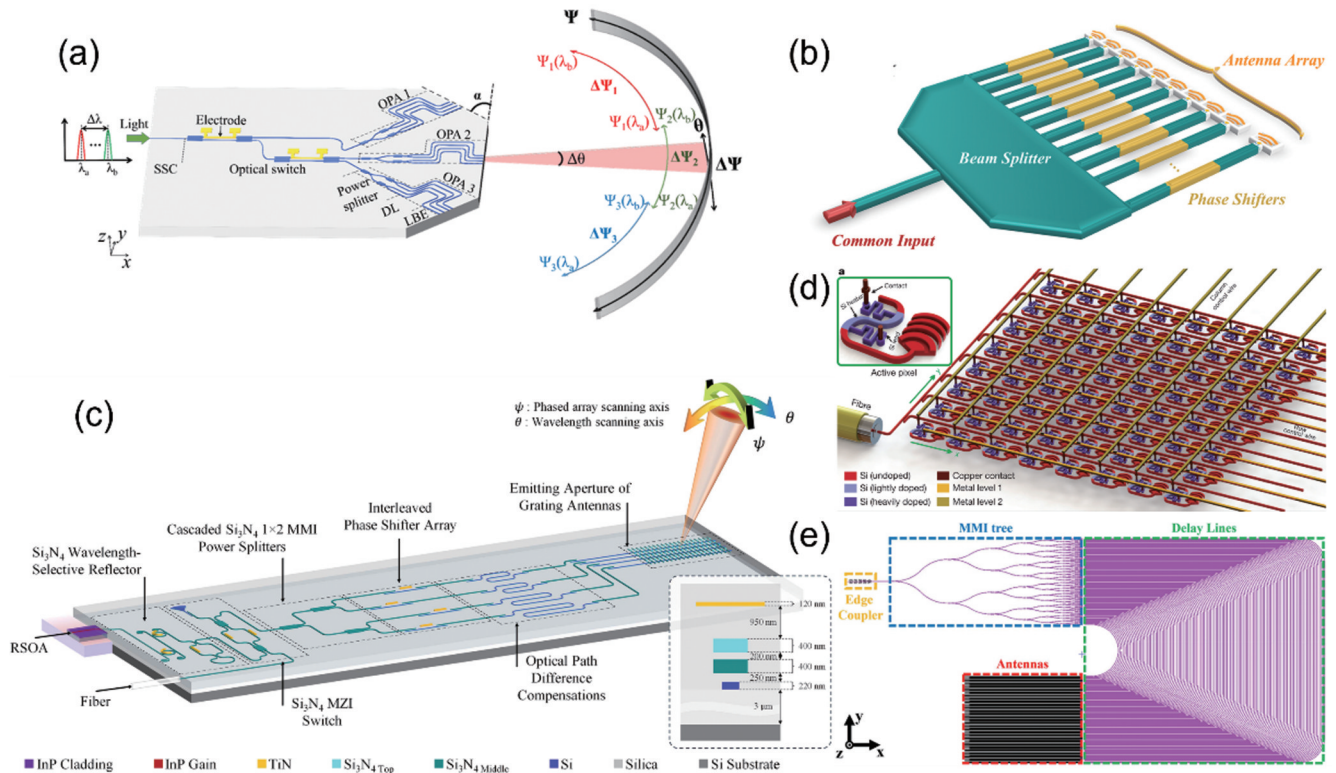


图 11 五种主流光学相控阵。(a)一维色散辅助光学相控阵<sup>[103]</sup>; (b)一维相扫光学相控阵<sup>[104]</sup>; (c)一维色散辅助与一维相扫相结合的光学相控阵<sup>[105]</sup>; (d)二维相扫光学相控阵<sup>[106]</sup>; (e)二维色散辅助光学相控阵<sup>[107]</sup>

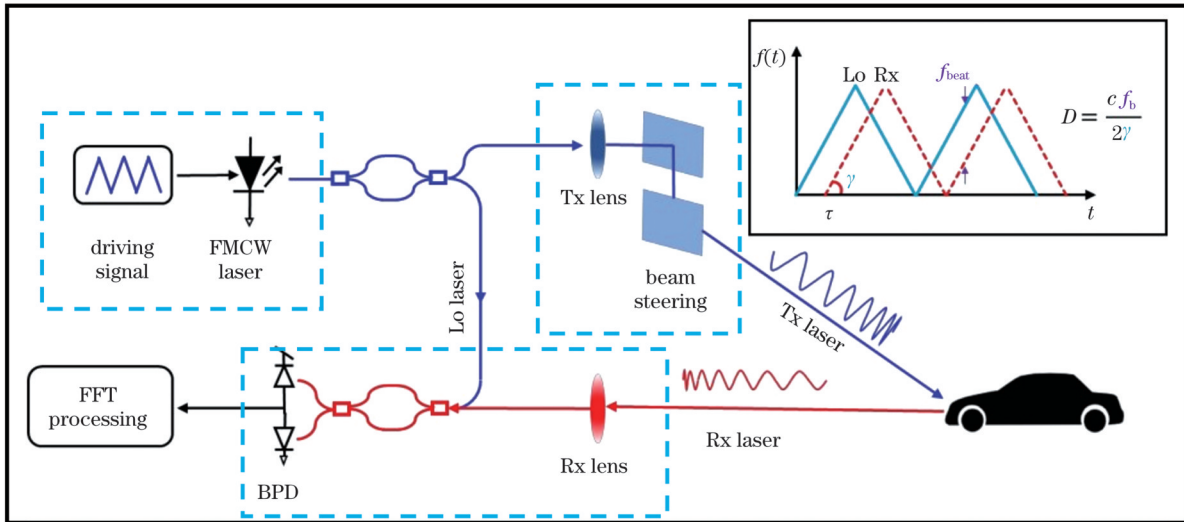
Fig. 11 Five types of mainstream OPAs. (a) 1D dispersion-assisted OPA<sup>[103]</sup>; (b) 1D phase-sweep OPA<sup>[104]</sup>; (c) combined 1D dispersion-assisted and 1D phase-sweep OPA<sup>[105]</sup>; (d) 2D phase-sweep OPA<sup>[106]</sup>; (e) 2D dispersion-assisted OPA<sup>[107]</sup>

或二维)及相位编程方式的不同,目前主流光学相控阵可分为五大类:一维色散辅助光学相控阵、一维相扫光学相控阵、一维色散辅助与一维相扫相结合的光学相控阵、二维色散辅助光学相控阵、二维相扫光学相控阵。这 5 种方案各有优势与局限,适用于不同的应用场景和性能需求,例如,色散辅助方案通常具有较宽的扫描角度范围,而相扫方案则在响应速度和控制精

度方面表现更优。图 11 展示了近年来这 5 类典型光学相控阵方案的代表性实验与应用成果,直观反映各方案的发展趋势和技术特点。

### 4.3 硅基集成全固态激光雷达系统

图 12 为典型的相干激光雷达测距系统架构。按照功能模块以及技术难点,可以将其划分为四大部分:线性调频光源、波束偏转、相干接收、控制与系统集成。



FFT: fast Fourier transform BPD: balanced photodetector Lo: local oscillator Rx: receiver Tx: transmitter  $f_{\text{beat}}/f_b$ : beat frequency  $D$ : distance  $c$ : light speed in vacuum  $\gamma$ : slope of frequency modulation

图 12 FMCW-OPA 测距系统架构

Fig. 12 FMCW-OPA ranging system architecture

窄线宽、高功率、大调谐范围激光源以及高线性度 FMCW 的产生是光源部分的研究重点。图 13(a)为上海交通大学(SJTU)设计的一款同时具有大波长调谐范围、大扫频带宽、窄线宽以及高扫频线性度的 FMCW 激光雷达光源<sup>[108]</sup>,其为基于游标微环结构的 InP/Si<sub>3</sub>N<sub>4</sub> 混合集成外腔激光器(ECL)。借助游标微环效应,可大幅扩展光源的波长扫描范围至 56 nm;又得益于 Si<sub>3</sub>N<sub>4</sub> 外腔芯片中较低的腔内传输损耗和更长的腔长,激光器的本征线宽被压缩至 2.9 kHz,扫频带宽为 7.68 GHz。结合复合预校正算法,在 1 kHz 重复频率下,可以将扫频线性度提升至 99.96%。理论上,借助该光源可以实现 2.47 cm 的距离分辨率。

波束偏转的核心在于设计具备低损耗、小发散角和大扫描范围的光学相控阵系统。其中,相控阵天线单元与波导阵列的集成是实现高性能波束控制的主要技术难点。与传统电磁波域的天线阵列不同,基于介质波导平台的光学相控阵目前仍受制于波导的衍射极限,具体表现为波导及其光学元件的尺寸受限,尤其是阵元的最小尺寸和阵元间距,通常保持在工作波长的 2~3 倍以上。阵元间的尺寸限制显著限制相控阵的最大扫描角度,影响其应用的广泛性和灵活性。为突破上述限制,当前研究的主要方向是阵列几何结构的稀疏化设计。不同于传统的均匀矩形网格

排布,稀疏阵列通过非均匀的波前采样方式,有效抑制主波束之外的混淆波束的相干叠加,提升波束质量。采用非均匀波前采样方式,包括如图 13(b)所示的圆环阵列<sup>[109]</sup>、费马螺线阵列<sup>[110]</sup>、科斯塔斯阵列(Costas array)<sup>[111]</sup>以及数值优化的稀疏阵列<sup>[112]</sup>等典型稀疏阵列布局,来抑制主波束之外其他混淆波束的相干相长。上述稀疏阵列设计显著降低阵列的冗余度,增强二维阵列的实际可实现性。同时,这些设计在提高旁瓣抑制比(SLSR)和扩大无混淆扫描范围(AFSR)方面取得了很好的效果。通过优化阵列布局,不仅提升波束质量,还推动二维相控阵理论分析和数值仿真技术的发展。

当前,相干接收与探测领域的主流发展方向集中于通过片上集成技术提升系统效率并降低制造成本。图 13(c)为一种片上相干接收探测架构,采用 CMOS 兼容工艺实现关键光电器件的高度集成。将接收光栅、平衡 PD 以及跨阻放大器集成于同一芯片上,可以大幅减少电学连接数量,有效降低信号传输过程中的损耗和干扰。该架构具备良好的可扩展性,能够通过增加阵元数量实现大规模阵列扩展<sup>[101]</sup>。

随着性能需求的提升,系统封装的要求也越来越高,传统的封装技术逐渐无法满足要求,需要采用更高密度的先进封装技术。图 13(d)为 Analog Photonics

公司(AP)2023年提出的一种借助3D CPO技术完成的9216阵元的光学相控阵激光雷达,其每个阵元间距为 $1.7\ \mu\text{m}$ ,有效发射孔径尺寸高达 $94\ \text{mm}^2$ ,扫描范围

为 $50^\circ \times 11^\circ$ <sup>[113]</sup>。所采用的大规模光学相控阵对制造工艺提出极高要求,也使制造成本急剧增加;同时,高密度集成封装也给散热造成巨大压力。

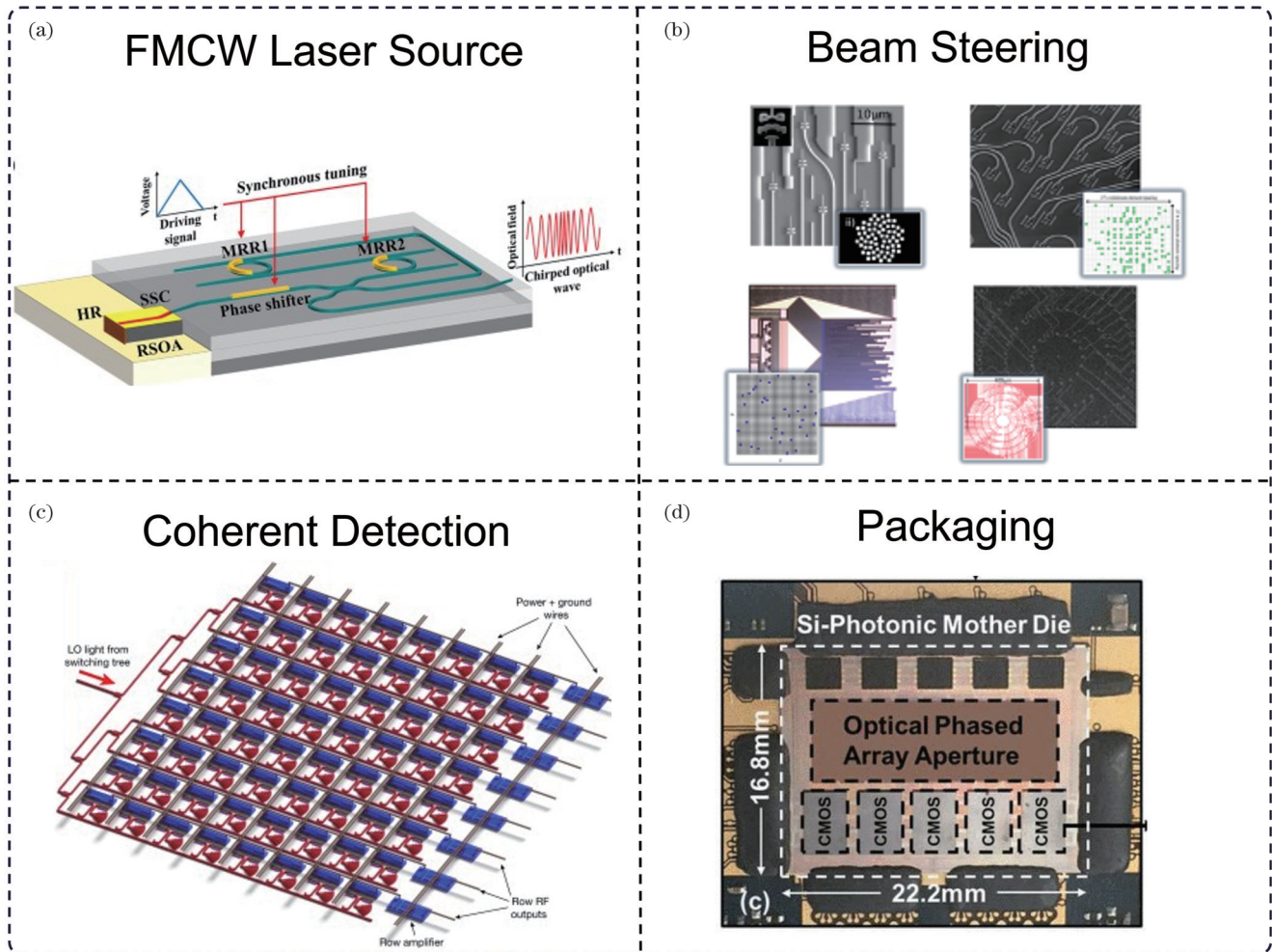


图 13 硅平台 FMCW-OPA 方案核心部件。(a)外腔激光器<sup>[108]</sup>; (b)片上光束偏转二维稀疏化方案<sup>[109-112]</sup>; (c)片上相干接收方案<sup>[101]</sup>; (d) 3D CPO 后的光学相控阵芯片<sup>[113]</sup>

Fig. 13 Core components of FMCW-OPA scheme on silicon platforms. (a) External cavity laser<sup>[108]</sup>; (b) on-chip 2D sparse matrix-based beam steering scheme<sup>[109-112]</sup>; (c) on-chip coherent reception scheme<sup>[101]</sup>; (d) 3D CPO OPA chip<sup>[113]</sup>

#### 4.4 硅基激光雷达最新研究进展

##### 1) 单片集成光学相控阵与高分辨率成像提升

光学相控阵的远场波束特征与其激励分布满足傅里叶变换关系,即波束的精细度与阵列孔径大小呈反比关系。在固定工作波长条件下,增大阵列孔径是提升远场分辨率的关键途径。如图 14 所示,自 2009 年以来,光学相控阵的规模和性能持续提升。2009 年,根特大学首次实现基于硅平台的 16 通道二维相控阵,结合相控与色散复合技术,构建有效数值孔径约为  $90\ \mu\text{m}^2$  的阵列,远场发散角约为  $2.5^\circ \times 2.7^\circ$ ,该系统采用三角形加热区实现单驱动波束转向控制<sup>[114]</sup>;2016 年,英特尔通过稀疏化阵列设计,利用 128 根波导实现约为  $0.9\ \text{mm}^2$  的发射面积,成功实现高达  $80^\circ$  的无混叠扫描范围<sup>[115]</sup>;2017 年,美国麻省理工学院采用 1024 个均匀排列的阵元,阵元间距为  $4\ \mu\text{m}$  (受限于波导间耦合),实现

$4\ \text{mm} \times 4\ \text{mm}$  的有效发射孔径,极小发散角达  $0.021^\circ \times 0.021^\circ$ <sup>[116]</sup>;2020—2023 年,Analog Photonics 公司先后开发出包含 8192 和 9216 阵元的光学相控阵,分别对应有有效数值孔径为  $8.2\ \text{mm} \times 5\ \text{mm}$  和  $15.667\ \text{mm} \times 6\ \text{mm}$ ,提升阵列的远场分辨率和扫描能力<sup>[113,117]</sup>。尽管制造工艺的进步使得大规模、高分辨率光学相控阵的实现成为可能,但阵元数量的增加也带来诸多挑战:驱动复杂度和功耗增加,导致系统控制难度提升;存在散热管理难题,影响器件稳定性和寿命;制造成本高和良率压力大,限制大规模量产。这些因素共同构成当前光学相控阵商业化推广的主要瓶颈,亟需在设计、材料和工艺等方面持续创新以实现突破。

##### 2) 高线性度 FMCW 光源优化

在 FMCW 激光雷达系统中,如何实现大扫频带宽、高重复频率及高线性度的线性 FMCW 光源,是提

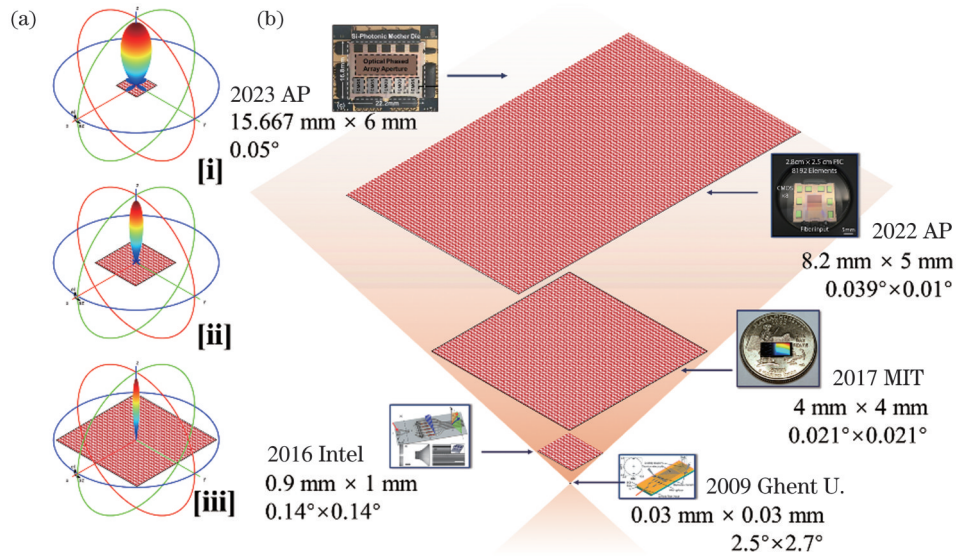


图 14 阵列孔径与波束精细度之间的关系。(a)不同孔径的亚波长间隔阵列主波束精度;(b)2009年以来在孔径尺寸上具有代表性的工作<sup>[113-117]</sup>

Fig. 14 Aperture-beam angular resolution relationship. (a) Main beam accuracy of subwavelength-spaced arrays with different apertures; (b) survey of notable works on aperture scaling since 2009<sup>[113-117]</sup>

升系统性能的核心技术难题。当前主流的实现方案主要包括以下三种:外调光源方案、啁啾脉冲光源方案、内调光源方案。外调光源方案通过外部调制器对激光光源的频率进行调制,实现线性调频。扫频带宽、重复频率和线性度等性能指标主要受限于外部调制器的性能。外调光源方案的优势在于显著降低光源本身的复杂度,便于系统设计与维护。然而,其缺点也较为明显,需要高性能、高带宽的调制器支持,且由于仅利用调制产生的单边带信号,必然引入至少 3 dB 的插入损耗,降低系统整体效率。啁啾脉冲光源方案采用宽谱激光器发射周期性脉冲光,通过色散器件使不同频率成分产生不同延时。当延时与光场频率呈线性关系时,周期脉冲光即被转换为线性 FMCW。该方法可实现较大的调频带宽,调谐速度快,适合高速测距应用。但色散器件在频率调制过程中会引入非线性效应,导致调频线性度受限,限制其在高精度 FMCW 系统中的应用。内调光源方案通过调控激光腔内的谐振条件,直接改变激光输出频率,实现线性调频。该方法输出功率高,满足激光雷达对信号强度的需求,同时系统结构相对简单,降低整体复杂度,也具备良好的片上集成潜力,有利于实现大规模量产。因此,内调光源方案被认为是推动 FMCW 激光雷达量产化的最具前景的技术路径。本文将重点围绕内调光源方案展开,对其进行详细介绍与分析。

在内调光源方案中,扫频线性度是影响 FMCW 激光雷达测距精度和准确性的核心技术指标。高线性度的扫频过程能够有效降低测距误差,提高系统性能,因此成为研究人员重点攻关的方向。目前,提升 FMCW 激光雷达扫频线性度的主要技术手段包括预校正算法、

环路反馈校正算法,以及重采样算法等。预校正算法通过对激光频率调制过程中的非线性特征进行建模和补偿,实现扫频信号的线性化。该方法依赖准确的系统参数识别,但对动态变化的系统环境适应性有限。环路反馈校正算法利用闭环控制系统实时监测和调整激光频率,动态抑制非线性误差。该方法能够显著提升线性度,但增加了系统复杂度和控制难度。重采样算法核心思想是基于参考拍频信号,获取扫频过程中相位间隔对应的准确时间间隔,进而对测距系统采集的拍频信号进行时间轴上的重采样处理。该方法无需事先进行预校正,也不依赖复杂的电路设计,能够实现线性度高且相位噪声抑制良好的扫频过程。但是其存在系统难以小型化、后端处理复杂以及参考光纤长度限制探测距离等问题。由于上述缺点,重采样算法近年来的研究和应用相对减少,更多研究倾向于结合预校正和环路反馈等方法,以实现性能与系统复杂度的平衡。

预校正算法利用时频分析提取调频过程中的非线性情况,结合校正算法得到驱动信号的预校正量,最后将预校正量加载到驱动信号中,产生线性的扫频过程。根据算法的不同,预校正算法可以分为传递函数法和线性迭代法,前者以台湾交通大学<sup>[118]</sup>的工作为代表。对于线性迭代算法,目前国内外都进行了广泛的研究。如表 2 所示,2019 年,美国加利福尼亚大学伯克利分校对分布式反馈(DFB)激光器和垂直腔面发射激光器(VCSEL)这两款扫频光源采用线性迭代算法<sup>[119]</sup>,分析不同校正系数对预校正算法收敛速度的影响。在此基础上,上海交通大学在 2021 年将线性迭代过程的反馈变量更改为扫频过程的斜率<sup>[120]</sup>,即斜率越大,校正的强度越大,从而加快预校正算法的收敛速度。同年,

清华大学基于比例-积分-微分控制(PID)算法的思想<sup>[121]</sup>,在预校正过程中引入微分项,以 1 kHz 的重复频率实现  $5.3 \times 10^{-9}$  的非线性度。2022 年,上海交通大学借助外腔激光器,在 7.68 GHz 调频带宽下,以 1 kHz 的重复频率实现  $1.02 \times 10^{-7}$  的非线性度<sup>[10]</sup>。同年,华中科技大学展示了一个高输出功率的分布式布拉格反射器(DBR)与半导体光放大器(SOA)集成的 FMCW 光探测和测距系统<sup>[122]</sup>。

此外,在预校正算法的基础上,搭建可以实时控制扫频线性度的环路反馈校正系统。主流的环路反馈校正系统采用光电锁相环(EO-PLL)的形式,这种方法参考了电锁相环技术,将激光器视为外部压控振荡电路,实现在锁相带宽内的反馈校正。2019 年,以加州大学伯克利分校为代表的研究机构结合 EO-PLL 技术<sup>[119]</sup>,在 36 GHz 调频带宽和 4 kHz 重复频率的条件下实现环路锁定。2021 年,奥地利维也纳技术大学在

60 GHz 调频带宽下以 0.53 kHz 的重复频率实现  $8.65 \times 10^{-9}$  的非线性度<sup>[123]</sup>。在国内,上海交通大学于 2015 年通过搭建混合 EO-PLL,实现 0.18% 的调频线性度<sup>[124]</sup>。

2023 年,瑞士洛桑联邦理工学院利用异质集成技术,即晶圆级别的直接键合方案将氮化硅和铌酸锂(LN)集成在一起,通过自注入锁定方案实现高达 10 MHz 的重复频率<sup>[125]</sup>。2024 年,上海交通大学使用 III-V/Si<sub>3</sub>N<sub>4</sub> 混合集成外腔激光器,在 68 nm 的长调谐范围内产生 FMCW 信号,并借助预失真算法与增益可调 EO-PLL 系统,抑制多个波长的线性调频和频率噪声,从而在 200 m 的距离下实现 4.2 cm 的测距精度<sup>[126]</sup>。2025 年,该团队基于双层 Si-SiN 平台,借助氮化硅来压缩激光器线宽以及硅基中等离子色散效应,实现相位的快速调节相结合,最高实现 2 MHz 的重复频率,为 FMCW 激光雷达的实际应用提供可行性验证<sup>[127]</sup>。

表 2 FMCW 激光器典型研究进展

Table 2 Typical research progress of FMCW lasers

Year	Reference	Laser type	Bandwidth	Repetition rate	Nonlinear regression coefficient	Scheme
2015	[124]	DFB	50 GHz	10 Hz	-	Optoelectronic phase-locked loop
2019	[119]	DFB	36 GHz	4 kHz	$1.8 \times 10^{-8}$	Optoelectronic phase-locked loop
2020	[118]	DFB	5.5 GHz	10 kHz	-	Pre-calibration algorithm
2021	[123]	DFB	60 GHz	0.53 kHz	$8.65 \times 10^{-9}$	Optoelectronic phase-locked loop
2021	[120]	DFB	26 GHz	1 kHz	$5.19 \times 10^{-8}$	Pre-calibration algorithm
2021	[121]	Self-injection DFB	5.56 GHz	1 kHz	$5.29 \times 10^{-9}$	Pre-calibration algorithm
2022	[122]	DBR	30 GHz	10 kHz	$5.07 \times 10^{-7}$	Pre-calibration algorithm
2022	[10]	ECL	7.68 GHz	1 kHz	$1.02 \times 10^{-7}$	Pre-calibration algorithm
2023	[125]	DFB	-	10 MHz	$1.2 \times 10^{-3}$	-
2024	[126]	ECL	0.7 GHz	1 kHz	$1.19 \times 10^{-6}$	Optoelectronic phase-locked loop and pre-calibration algorithm
2025	[127]	Self-injection DFB	1.05 GHz	2 MHz	$4.3 \times 10^{-6}$ @100 kHz	Pre-calibration algorithm

### 3) 国内外代表性研究工作

基于 FMCW-OPA 的激光雷达研究在国外起步相对较早。2017—2019 年,美国麻省理工学院完成了诸多开创性的工作,如:2017 年,其借助硅光平台首次利用 FMCW-OPA 系统实现纯固态激光雷达的测距、测速功能,其扫描视场为  $46^\circ \times 36^\circ$ ,对应发散角为  $0.85^\circ \times 0.18^\circ$ ,测试距离最远为 2 m,测距分辨率为  $20 \text{ mm}$ <sup>[128]</sup>;2019 年,该团队借助 512 通道的光学相控阵架构,展示其更高性能的测距测速能力,最远可实现 185 m 的探测距离,同时横向发散角也压缩至  $0.04^\circ$ <sup>[129]</sup>;同年,该团队联合加州大学伯克利分校完成基于硅基的两路光学相控阵收发和 CMOS 驱动电路的全集成全固态激光雷达,其在最远 50 cm 的探测距离内实现 3.3 cm 的测距分辨率<sup>[130]</sup>。

2021 年,美国南加州大学(USC)借助  $1.4 \mu\text{m}$  天线间距的 256 通道光学相控阵实现  $67^\circ \times 5^\circ$  的视场角<sup>[131]</sup>。

同年,国内吉林大学(JLU)借助稀疏化光栅间距+链式结构天线实现 128 通道的  $140^\circ \times 19.23^\circ$  的视场角,对应发散角压缩至  $0.021^\circ \times 0.1^\circ$ 。其借助 FMCW 系统实现 100 m 的测距<sup>[132]</sup>。2022 年,Analog Photonics 公司发表其基于 8192 通道光学相控阵的 FMCW 激光雷达系统,其收发孔径高达  $8 \text{ mm} \times 5 \text{ mm}$ ,在最远 35 m 的距离下实现  $154 \times 20$  的点云密度<sup>[117]</sup>。次年,该团队借助 3D CPO 技术制造一套 9216 阵元的光学相控阵激光雷达,其每个阵元间距为  $1.7 \mu\text{m}$ ,有效发射孔径高达  $94 \text{ mm}^2$ ,扫描范围为  $50^\circ \times 11^\circ$ 。同时该系统可实现 50 m 的探测距离,每秒可探测点数在 10000 以上,标志着高密度点云输出成为可能<sup>[113]</sup>。

从 2022 年起,国内的研究发展势头开始超过国外。吉林大学基于 Si-SiN 工艺平台制作 64 通道光学相控阵芯片,主瓣输出功率达到 690 mW,实现 100 m 测距<sup>[133]</sup>。同年,中国科学院半导体研究所(ISCAS)借助氮化硅

工艺平台以及啁啾光栅天线实现近似均匀的近场光强分布,得益于该啁啾设计,主瓣发散角被抑制在 $0.04^\circ \times 0.05^\circ$ 内。另外,得益于啁啾设计,该系统实现了 100 m 的收发测距(而均匀光栅天线由于有效孔径相对较小,只能实现 50 m 测距)<sup>[134]</sup>。2023 年,中国科学院西安光学精密机械研究所(XIOPM)借助稀疏化技术将两组光学相控阵光栅交错分布在同一片区域,构成同轴全固态 FMCW 激光雷达系统<sup>[135]</sup>。2024 年,上海交通大学将外腔激光器和光学相控阵芯片集成在一起,形成全固态集成激光雷达。借助稀疏化技术,该芯片实现 $0.051^\circ \times 0.016^\circ$ 的超小发散角。该芯片的外腔激光器可以在实现 0.9 GHz 调频带宽、1 kHz 调频速率的同时,调频信号线性度高达 99.74%<sup>[136]</sup>。同年,浙江大学借助 32

通道的色散光学相控阵结构实现信噪比为 36.8 dB 条件下的 10 m 距离探测<sup>[137]</sup>。2025 年,浙江大学利用宽波导和连续谱中的束缚态效应获得超长的光栅天线阵列,并借助 FMCW 展示了单光子微腔光梳结合光学相控阵在相干三维成像中的应用,实现纯固态高帧率相干激光雷达的并行收发<sup>[138]</sup>。同年,吉林大学在 OFC 会议上发表其最新研究成果,其实现了 40 kHz 的激光雷达扫点频率,并展示了速率高达每秒 30 帧的实时成像<sup>[139]</sup>。比利时微电子研究中心在硅光子平台上展示了一种单片集成的双轴激光雷达,其结合了具有快速片上校准的光学相控阵发射器,以及具有相干像素接收器的焦平面阵列(FPA)。借助 FMCW 系统,其实现了距离为 1.4 m、信噪比为 21 dB 的测距性能<sup>[140]</sup>。

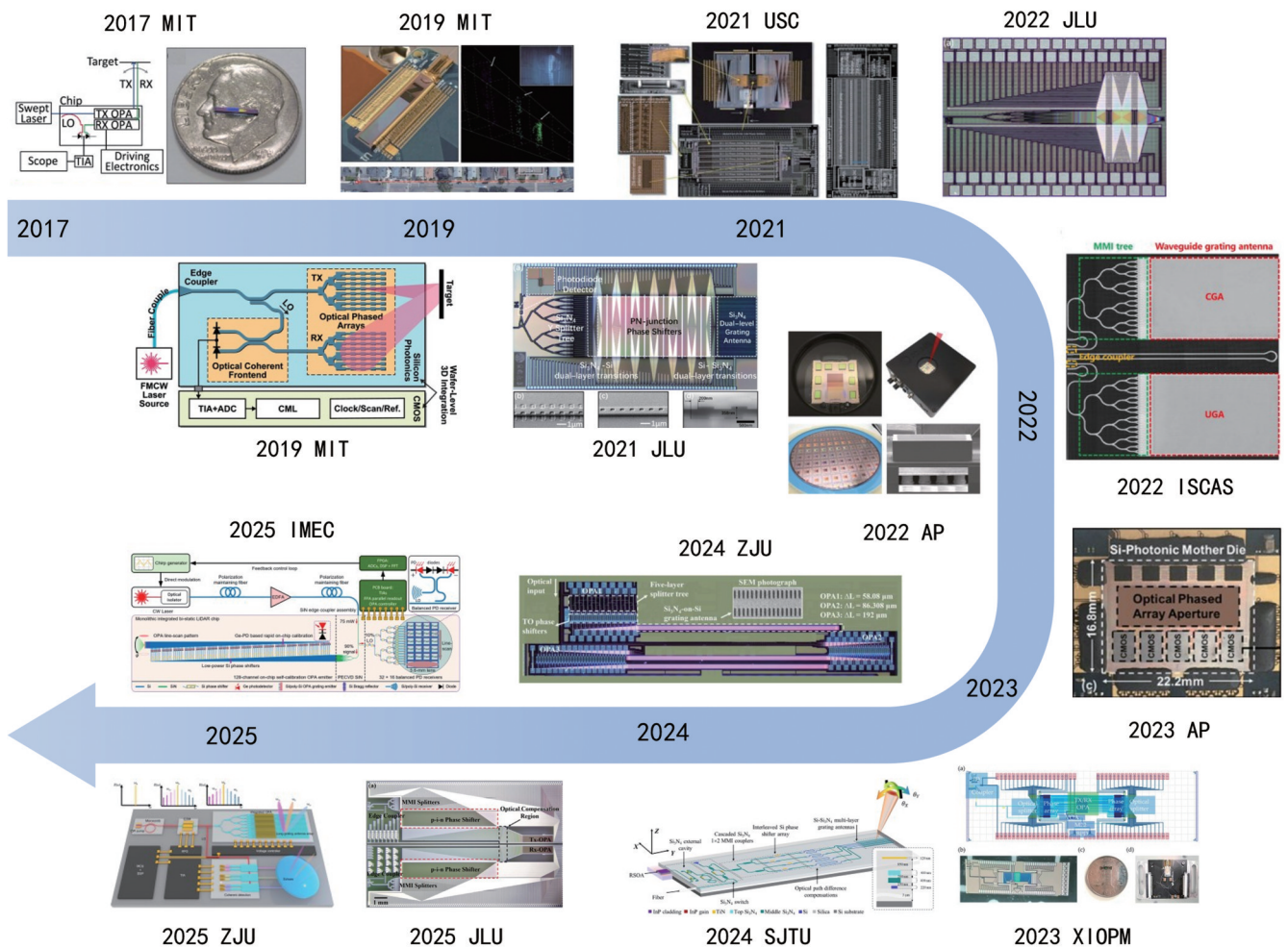


图 15 近年硅基集成 FMCW-OPA 激光雷达典型研究进展<sup>[113,117,128-140]</sup>

Fig. 15 Recent typical research progress of silicon FMCW-OPA lidar<sup>[113,117,128-140]</sup>

#### 4.5 未来发展趋势

激光雷达在无人驾驶、工业机器人、AR/VR 等领域的应用展现出其独特优势,如图 16 所示。在无人驾驶领域,激光雷达是实现安全冗余的关键传感器,随着固态化与成本下降,高级别自动驾驶商业化成为可能。在工业检测、安防监控与机器人导航领域,激光雷达因

其非接触式高精度测量能力、复杂场景适应性成为标配。在 AR/VR 应用中,激光雷达为虚实融合提供底层技术支持,iPad Pro 中的 AR 特效即为典型应用案例。

激光雷达正从机械扫描激光雷达向硅基芯片化相干激光雷达转型。光电融合技术通过光学器件与电子

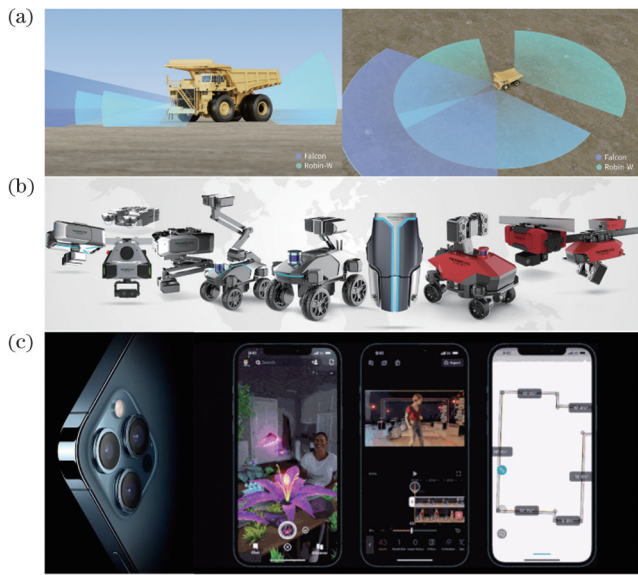


图 16 激光雷达的典型应用场景。(a)无人驾驶中环境感知与障碍物检测<sup>[141]</sup>；(b)工业检测、安防监控与机器人导航<sup>[142]</sup>；(c)AR/VR场景下的深度感知及交互增强<sup>[143]</sup>

Fig. 16 Typical application scenarios of lidar. (a) Environment sensing and obstacle detection in autonomous driving<sup>[141]</sup>；(b) industrial inspection, security monitoring, and robot navigation<sup>[142]</sup>；(c) depth perception and interaction enhancement in AR/VR scenarios<sup>[143]</sup>

器件在硅基平台的深度集成,实现光信号产生、调制、传输与电信号处理、控制、计算的统一,为自动驾驶、智能机器人等应用提供高精度、低成本感知方案。光电融合技术打破传统光学器件与电子器件分离设计的界限,实现光域与电域功能协同优化。在硅基光感知系统中,这不仅是器件物理集成,更是系统架构重构。芯片级激光雷达展现了光电融合技术价值。片上集成 FMCW-OPA 激光雷达具备高精度动态四维感知能力,在低光和复杂光照条件下表现出优异的环境适应性,同时具备集成化、小型化、规模化潜力。基于 FMCW 技术的相干激光雷达将激光发射、相位调制、相干探测和 DSP 功能集成在单一硅基平台上。光电融合消除了光学子系统与电子子系统间的接口瓶颈,使 FMCW 技术与硅光集成芯片协同工作,在保持高精度测距的同时降低系统复杂性和制造成本。光学相控阵技术同样受益于光电融合。硅基光学相控阵将数百至数千个光学天线单元与相位控制电路在硅基衬底上协同集成,通过光电协同控制实现激光束电子扫描。光电器件在同一硅基衬底上集成,消除分立器件间的接口损耗和信号延迟,提升系统性能密度和能效比。光电融合使激光雷达从“系统级集成”向“芯片级集成”演进,将复杂光机电系统功能压缩到单一芯片上。

利用光电融合技术重构系统架构,也可推动多模态传感发展。微波雷达穿透力强但分辨率较低;毫米波雷达兼具穿透性和中短距探测能力但角分辨率受

限;激光雷达分辨率高但受天气影响较大。光电融合技术能在单一芯片上实现多种感知模态协同工作,通过光域与电域深度融合,突破单一传感技术性能瓶颈。激光雷达与毫米波雷达协同感知通过光电微波融合技术实现全天候感知能力飞跃。硅基光电子平台通过异构集成技术,将光学相干探测单元与毫米波收发前端在同一芯片上融合,实现光学感知与射频感知在信号处理层面的协同优化。光学激光雷达提供高精度距离和角度信息,毫米波雷达在恶劣天气条件下保持稳定探测性能,两者实现数据层面实时融合和算法层面协同处理。激光雷达与图像传感器融合展现了光电融合在多维信息获取方面的优势。通过在同一硅基平台上集成激光雷达发射接收单元、CMOS 图像传感器和图像信号处理器,实现距离信息与视觉信息的像素级融合。这种深度集成消除传统多传感器系统中的时空标定复杂性,通过共享光学系统和信号处理资源大幅降低系统成本和功耗。

多模态雷达融合技术通过互补感知能力与算法优化,推动智能感知从单一功能向全场景覆盖演进。边缘智能与光电算协同成为多模态感知系统重要发展趋势。随着固态化、边缘计算与 AI 算法结合,光电融合技术应用边界持续拓展。通过将深度学习算法直接集成到硅基光电子芯片中,实现“感知-融合-决策”一体化设计。

## 5 硅基光计算:从加速器到智能系统

### 5.1 光计算对硅基光电子的需求

以人工神经网络(ANN)为代表的 AI 算法在图像处理<sup>[144]</sup>、自动驾驶<sup>[145]</sup>、生命科学<sup>[146]</sup>、自然语言处理<sup>[147]</sup>等诸多领域展现出强大的处理能力。随着 AI 模型规模的指数级增长<sup>[148]</sup>,传统的基于冯·诺依曼架构的数字计算机计算单元与存储单元在物理上相互独立,频繁的数据交互传输导致功耗急剧增加,无法满足指数增长的算力需求<sup>[149]</sup>。尽管现场可编程门阵列(FPGA)、ASIC、张量处理器(TPU)、忆阻器(memristor)等电子计算方案在提升计算效率以及计算吞吐量上取得进展,但受限于电子充放电的过程,其处理速度限制在 GHz 以内,难以获得数量级的算力提升。此外,焦耳热、电磁串扰等问题限制其进一步发展。光具有超高带宽、低延迟、低传输损耗等天然优势,具有波长、偏振态、模式、空间、时间等多个调控并行度,采用光计算有望实现数量级的算力提升以及计算能耗降低。

光计算按照集成度可划分为基于体光学元件的非集成光计算和基于集成光器件的集成光子计算。相较于非集成光计算方案,集成光子计算在系统小型化、稳定度、大规模制造等方面具有明显的优势。在材料平台方面,得益于硅基集成光子工艺、器件、先进封装的快速发展,基于硅基光电子平台的集成硅基光子计算得到广泛研究。

## 5.2 光计算核心器件

ANN 通过模拟人脑神经元之间的连接通信方式搭建等效数据模型以解决复杂任务,如图 17(a)所示。当前神经元接收来自上层与其相连的多个神经元的输入信号  $x_i (i=1, 2, \dots, n)$ , 接收信号总强度表示为累加和  $\sum_{i=1}^n x_i w_i$ , 其中,  $w_i (i=1, 2, \dots, n)$  表示神经元之间的连接作用强度。考虑到输入信号需要在一定强度的刺激下才能激发输出, 引入非线性激活函数  $f(\cdot)$  表示输入与输出关系。由此, 以多层感知机模型为例的单层

$$\text{网络输出模型表示为 } \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = f \left( \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \right),$$

包括线性矩阵向量乘法 (MVM) 计算与非线性激活 (nonlinear activation) 计算, 其中向量  $\mathbf{X} = [x_1 \ \cdots \ x_n]^T$  表示来自上层神经元的输入信号, 向量  $\mathbf{Y} = [y_1 \ \cdots \ y_n]^T$  表示当前层神经元的输出信号, 权重矩阵  $\mathbf{W} = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nn} \end{bmatrix}$  表示当前层与前层神经元之间的相互作用。ANN 中存在大量 MVM 计算, 通用型光子计算加速器通过实现 MVM 计算加速从而加速 ANN。

硅基光计算加速器通过高速光电调制器将输入信号  $X$  编码到光域, 利用硅光单元器件搭建权重矩阵  $\mathbf{W}$ , 利用高速探测器实现输出光信号  $Y$  的累加和探测。随着光信号流经整个光计算芯片, 线性 MVM 计算即可完成。硅基光计算加速器根据光信号的作用方式, 大致可分为相干光方案以及非相干光方案。相干光矩阵计算加速器方案基于不同计算单元之间光相位调制, 从而完成权重矩阵的表达, 主要包括基于奇异值分解理论的马赫-曾德尔干涉器网格 (MZI mesh) 拓扑架构 [图 17(b)] 以及基于超表面衍射 (metasurface diffraction) 的多层衍射神经元架构 [图 17(e)]。非相干光矩阵计算加速器方案中, 单个计算单元实现单点数据的乘法计算, 通过计算单元的拓扑完成 MVM 计算, 其架构主要包括基于波分复用技术的 MRR 权重库 (WB) 架构 [图 17(c)] 以及基于衰减器的强度调制阵列 (IM array) 架构 [图 17(d)]。MZI 网格架构的核心器件为 MZI, 根据奇异值分解理论, 将任意复数矩阵  $\mathbf{M}$  分解为酉矩阵  $\mathbf{U}$ 、对角矩阵  $\mathbf{D}$  和另一个酉矩阵  $\mathbf{V}$  的乘积 ( $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{V}$ ) 并用硅基分束器和移相器表征, 通过器件拓扑, 即可在光计算芯片上完成权重矩阵的映射 [150-152]。基于波分复用技术的 MRR 权重库架构的核心器件为 MRR, 输入信号  $X$  编码到独立的光波长, 级联 MRR 在不同的工作波长下实现与对应波长处输入信号的乘法计算, 通过平衡探测器监控比较 MRR 总线直通端波导和下载端波导的输出光功率, 从而实现

MVM 计算 [153]。基于强度调制阵列架构方案的核心器件为光强度调制器, 可使用 MZI、MRR、光衰减器等实现。输入信号  $X$  经强度调制器件实现乘法计算, 不同路径的加权光信号通过 PD 接收并转换成光电流在电域实现累加, 通过阵列拓扑实现 MVM 计算 [154]。基于超表面衍射方案的核心单元为超表面衍射单元, 基于惠更斯-菲涅耳原理, 每个神经元可视为一个二次波源, 二次波的振幅和相对相位由神经元的输入波与其传输系数的乘积决定, 数学描述表示为  $n_{i,p}^m = w_{i,p}^m \cdot u_i^m \cdot T_i^m$ , 其中  $T_i^m$  表示当前第  $m$  层网络中第  $i$  个衍射神经元对输入光振幅和相位的调制作用,  $u_i^m$  表示上层其他神经元对当前神经元的输入,  $w_{i,p}^m$  表示当前神经元输出到下一层第  $p$  个神经元的光衍射过程 [155]。

为了进一步提升光计算芯片的计算速度, 除了基于标准硅光工艺平台的硅基光电子高速调制器, 基于硅衬底的 TFLN、BTO、薄膜钽酸锂 (TFLT) 等新材料平台也具有较强的电光效应, 有望进一步提升调制器性能。此外, 传统的热光、电光调制需要大量的能量消耗以维持权重数值, 具有断电数据易失性。为了进一步降低片上计算能耗, 解决光域数据非易失存储难题, 结合非易失性材料, 如相变材料 (PCM)、磁电材料、铁电材料等, 可实现近零静态功耗的非易失光子存内计算。

## 5.3 片上光学加速器研究进展

基于 MZI 网格的硅基光子计算加速器, 其进展如图 18(a) 所示。2017 年, 麻省理工学院在 SOI 硅片上级联 56 个 MZI, 搭建一个  $4 \times 4$  的矩阵, 应用于 ANN 元音识别的推理任务中, 展开基于 MZI 网格方案的片上光子计算加速器研究的新篇章 [152]。针对片上 MZI 网格矩阵难以训练的问题, 2018 年, 斯坦福大学提出一种原位训练方法, 测量片内光强度, 利用伴随变量推导求解反向传播算法中的权重梯度, 成功训练出一个 2 层  $3 \times 3$  MZI 网格光子神经网络, 实现非线性的“异或” (XOR) 逻辑门功能 [156]。2021 年, 南洋理工大学提出一种同时进行光振幅和相位信息编码以搭建复值神经网络的方法, 该方案进一步释放光子计算的优势, 在逻辑门、鸮尾花数据集分类、非线性数据集分类以及手写字识别等实验任务中, 光复值神经网络展现出比实值神经网络更强的学习能力 [157]。2023 年, 斯坦福大学基于原位训练算法实验训练出一个具有可编程移相器和光功率监测功能的 3 层 4 端口硅基光子神经网络 [158]。该网络实现双向光传播, 通过干扰前向和后向传播的光信号, 测量移相器电压的反向传播梯度模拟光子神经网络的原位反向传播, 在 MNIST 数据集上实现与数字计算机相当的计算效果。2024 年, 麻省理工学院单片集成基于 MZI 网格的线性矩阵与基于探测器的非线性激活单元, 搭建一个具有 6 个神经元的 3 层神经网络, 其计算延迟仅为 410 ps [159]。为了进一步提升神经网络计算

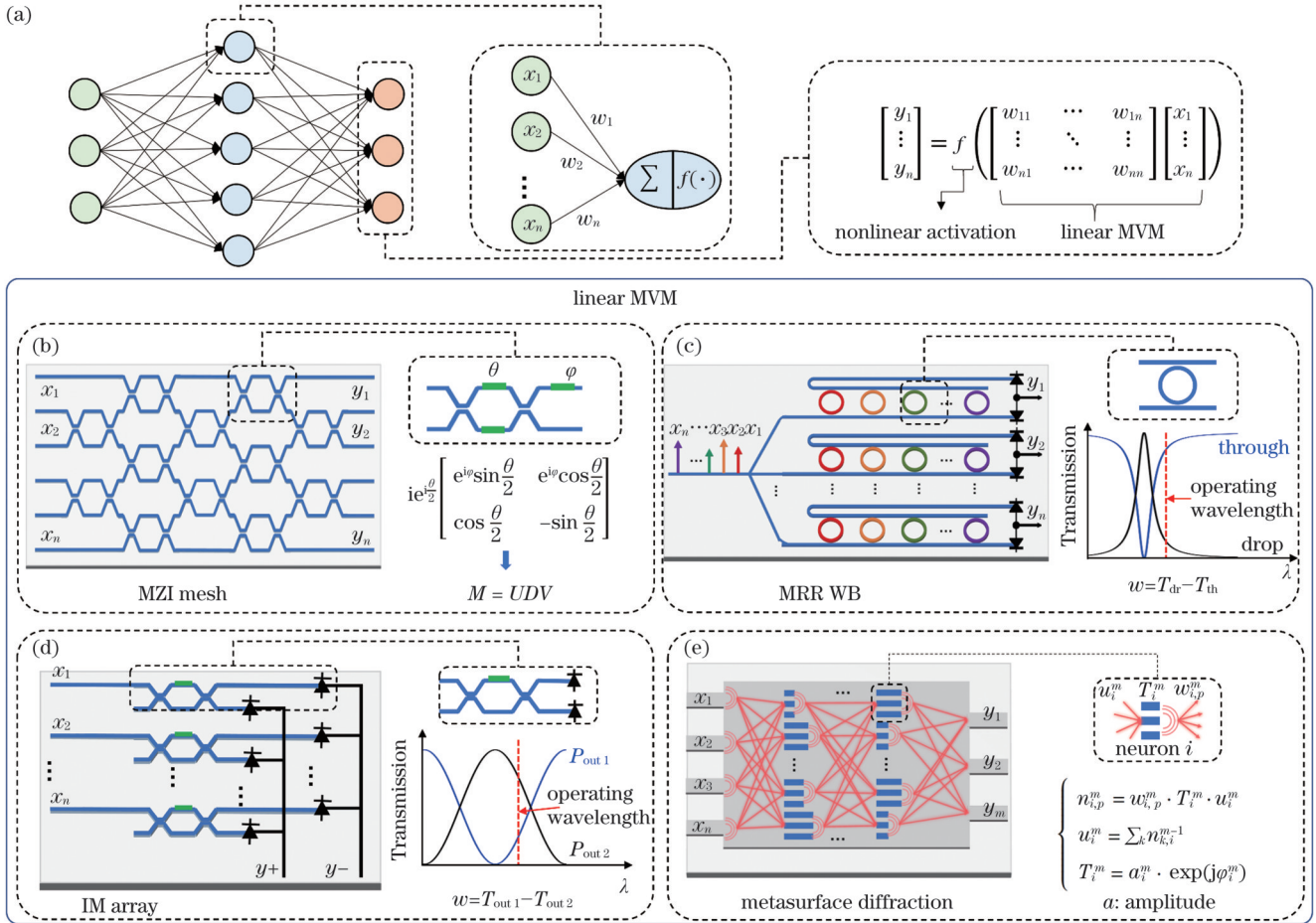


图 17 ANN 等效数学模型与硅基光子计算加速器架构。(a)ANN 等效数学模型包括线性矩阵向量乘积计算与非线性激活函数；硅基光子计算加速器架构包括：(b)基于 MZI 网格架构；(c)基于 MRR 架构；(d)基于强度调制阵列的架构；(e)基于超表面衍射架构

Fig. 17 ANN equivalent mathematical model and silicon-based photonic computing accelerator architectures. (a) ANN equivalent mathematical model comprising linear matrix-vector multiplication computation and nonlinear activation functions; silicon-based photonic computing accelerator architectures including: (b) MZI mesh-based architecture; (c) MRR-based architecture; (d) intensity modulation array-based architecture; (e) metasurface diffraction-based architecture

容量和可拓展性,2024年,清华大学基于硅基集成平面衍射和 MZI 网格,混合设计一款分布式架构的太极光子芯片<sup>[160]</sup>,该芯片具有数百万个神经元,能量效率达到 160 万亿次运算每秒瓦特(TOPS/W)。

基于 MRR 权重库的硅基光子计算加速器进展如图 18(b)所示。2014年,普林斯顿大学提出一种基于波分复用技术的传播-加权的片上神经网络架构<sup>[161]</sup>。该架构采用 MRR 作为权重调控单元,通过对微环谐振波长的控制实现滤波与加权,利用 PD 实现加权光信号的累加并驱动下一级的神经元实现非线性激活。基于该架构,2017年,普林斯顿大学在硅基平台制备 4 个 MRR 级联的网络结构,提出适用于硅光子平台的“调制器类”神经元架构,利用 PD 输出直接驱动调制器神经元实现非线性激活。实验中,通过求解微分方程问题实验验证了该神经网络架构在计算加速方面的潜力<sup>[153]</sup>。随后几年,针对 MRR 权重高精度调控以实现高精度光计算的研究也获得广泛关注。2022年,

普林斯顿大学提出一种新颖的“抖动”控制方案,通过在输入信号上叠加一个低频、小振幅的“抖动”信号,实现了对从调制器到探测器的整个光子突触链路的实时监控和稳定,系统加权精度得以提升至 9 bit<sup>[162]</sup>。2021年,普林斯顿大学基于硅基集成了具有精确加权、求和以及非线性激活功能的所有光子组件,针对跨太平洋传输数据链路中的光纤非线性实现光域补偿,该光子神经网络能够成功预测并补偿 10080 km 海底光纤链路中的非线性失真,将信号的 Q 值提升 0.60 dB,缓解传统 DSP 补偿方案的复杂度和对速度的要求<sup>[163]</sup>。在权重调控的基础上,2021年,斯威本理工大学引入时间调制,提出基于光频梳的波长-时间交织调控的光计算加速器,该方案采用分立的光源、调制器、色散光纤等单元器件,实现 11 TOPS 的算力演示<sup>[164]</sup>。2023年,北京大学将调制器、光波导延迟线、MRR 权重库等核心器件集成到单片硅基芯片上,采用 AlGaAsOI 微梳光源,在 MNIST 数据集上进行手写字识别的实验演

示<sup>[165]</sup>。以上方案大多采用固定的权重映射到光计算芯片,难以实现快速的权重更新与在线训练。2024年,普林斯顿大学演示了一种在模拟域盲源分离解决动态干扰的光子处理器<sup>[166]</sup>,该处理器采用FPGA进行外围电路控制以实现完全的片上信号通路,具有精确的抖动重量控制,核心处理延迟低于15 ps,即使考虑到外围电路的延迟,总延迟也不超过200 ps,比传统电子方案快3个数量级以上。

基于光强度调制阵列的方案及其进展如图18(c)所示。2022年,宾夕法尼亚大学提出一种端到端的集成光子深度神经网络(PDNN),通过直接处理在片上像素阵列上传播的光波,执行亚纳秒级的图像分类,实验中该PDNN的端到端分类时间经测量为570 ps<sup>[154]</sup>。神经元的线性计算通过衰减器进行强度调制,其多路乘积结果经PD阵列探测后在电域实现累加计算与非线性激活。2025年,Lightelligence公司报道了一款大规模集成光子加速器<sup>[167]</sup>,采用2.5D混合封装集成光芯片和电芯片,其光芯片采用MZI器件实现光强度调制并采用PD阵列将光乘法计算结果转换到电域,实现电域累加计算。光计算矩阵规模为 $64 \times 64$ ,集成超过16000个光子组件,实现纳秒级的单周期计算延迟。在求解Ising问题的实验中,最低延迟可达5 ns,比在商用高端GPU(NVIDIA A10)上运行相同任务的单周期延迟( $>2300$  ns)快了近500倍。其计算吞吐量约为8.19 TOPS,计算能效在包含激光器时为2.38 TOPS/W,不包含激光器时为4.21 TOPS/W。同年,Lightmatter公司推出一款光子AI处理器<sup>[168]</sup>,采用3D堆叠封装电芯片与光芯片,其光芯片包括128个10位光子矢量单元和 $128 \times 128$ 个7位权重单元,输入与输出数据流为256 GB/s,权重输入数据流为1024 GB/s,达到65.5 TOPS的16位自适应块浮点(ABFP)运算速度,总功耗为79.6 W(其中电功耗为78 W,光功耗为1.6 W),展示了光子处理器在残差神经网络(ResNet)和基于Transformer的双向编码器表示模型(BERT)等高级AI模型中与数字计算相当的计算结果。

基于超表面衍射的方案及其进展如图18(d)所示。2021年,清华大学提出一种基于SOI的集成衍射神经网络<sup>[169]</sup>,采用微型槽作为衍射神经元,并进行仿真验证。2022年,特拉华大学在 $0.135 \text{ mm}^2$ 芯片上制备103个亚波长结构移相器作为权重单元,在空间模式分类、波长解复用和高光谱图像分类等应用方向展现出多任务能力<sup>[170]</sup>。随后在2023年,清华大学在SOI平台制备了具有1个和3个隐藏层的片上衍射神经网络<sup>[155]</sup>,并在Iris数据集上进行分类实验,其准确度分别达到86.7%和90%。针对传统片上光学衍射神经网络参数不可调的问题,2024年,华中科技大学展示了一款包括1个输入层、5个隐藏层和1个输出层的衍射光学神经网络芯片<sup>[171]</sup>,不同于物理刻蚀的固定的衍射单元,该工作采用电热控制衍射单元产生相位变化,在

视觉、音频和触摸多任务中进行演示。2025年,清华大学在建模层面提出多任务处理网络架构<sup>[172]</sup>,利用深度回归算法对物理传播过程进行建模以处理不同的任务,制作每集成参数超过 $60000/\text{mm}^2$ 的超小型衍射芯片,完成多层面的分类和回归任务。

结合非易失材料实现光存内计算加速的方案及其进展如图18(e)所示。2021年,明斯特大学提出一款结合相变材料的集成光子计算加速器(张量核心)<sup>[173]</sup>,采用相变材料(如 $\text{Ge}_2\text{Sb}_2\text{Te}_5$ , GST)与Si异质集成为权重单元,利用光脉冲对相变材料进行激励以改变其结晶态,从而实现输出光强度调制,利用片上产生的耗散克尔孤子光频梳作为多波长光源,通过波分复用技术,将多个输入向量并行地编码在不同的频率(波长)上实现大规模并行非易失存内计算。2023年,牛津大学实验制备了一款 $1 \times 4$ 电编程光子存内计算芯片<sup>[174]</sup>,权重调控单元采用GST-Si异质集成波导,使用掺杂硅波导微加热器实现电脉冲调控,其电可编程相变材料存储单元在4.13 dB光调制深度下实现4 bit(18个可分辨能级)的权重编码精度。为了提升光计算并发度,2023年,牛津大学在传统光子计算利用的空间和波长维度之外引入射频调制来增加计算并行性,采用基于GST-Si异质集成波导的光计算芯片,实现了并行度高达100(2个波长通道、50个射频分量)的高纬度光子存内计算,在心电图识别任务中达到93.5%的准确率<sup>[175]</sup>。此外,为了减轻波分复用对光源的需求,2024年,牛津大学提出利用部分相干光作为光源的方案<sup>[176]</sup>,在不牺牲计算精度的前提下,有望实现更大尺寸的光子张量核心。在基于GST-Si异质波导的光子张量核心上执行帕金森病步态分类任务,部分相干光作为光源的识别准确率(92.2%)与使用相干光的对照组(92.6%)表现相当。

在标准硅基光平台的基础上集成新型材料有望实现光器件性能的进一步提升,基于硅基集成其他材料平台的光计算发展及其进展如图18(f)所示。2024年,香港城市大学展示了一款基于4 inch(1 inch=2.54 cm)晶圆TFLN平台的集成微波光子处理引擎<sup>[177]</sup>,在CMOS兼容电压下可实现高达67 GHz的处理带宽,在图像边缘检测应用中,处理一个 $250 \times 250$ 像素的图像仅需244 ns,比在通用个人计算机上运行的传统算法(380  $\mu\text{s}$ )快了3个数量级以上。此外,该光子处理引擎还展现了求解常微分方程、生成超宽带信号等多任务能力。2025年,苏黎世联邦理工学院测量了100 MHz~330 GHz连续频率范围内LN和BTO中的Pockels系数和介电常数<sup>[178]</sup>,实验结果表明,与LN相比,BTO具有更高的线性电光系数。2025年,希腊塞萨洛尼基亚里士多德大学介绍了一种基于 $\text{Si}_3\text{N}_4$ -BaTiO<sub>3</sub>电光移相器的高能效光子MZI调制器<sup>[179]</sup>,实验表明其 $\pi$ 相移功耗仅为121 nW,有望应用于低功耗光计算加速。2025年,

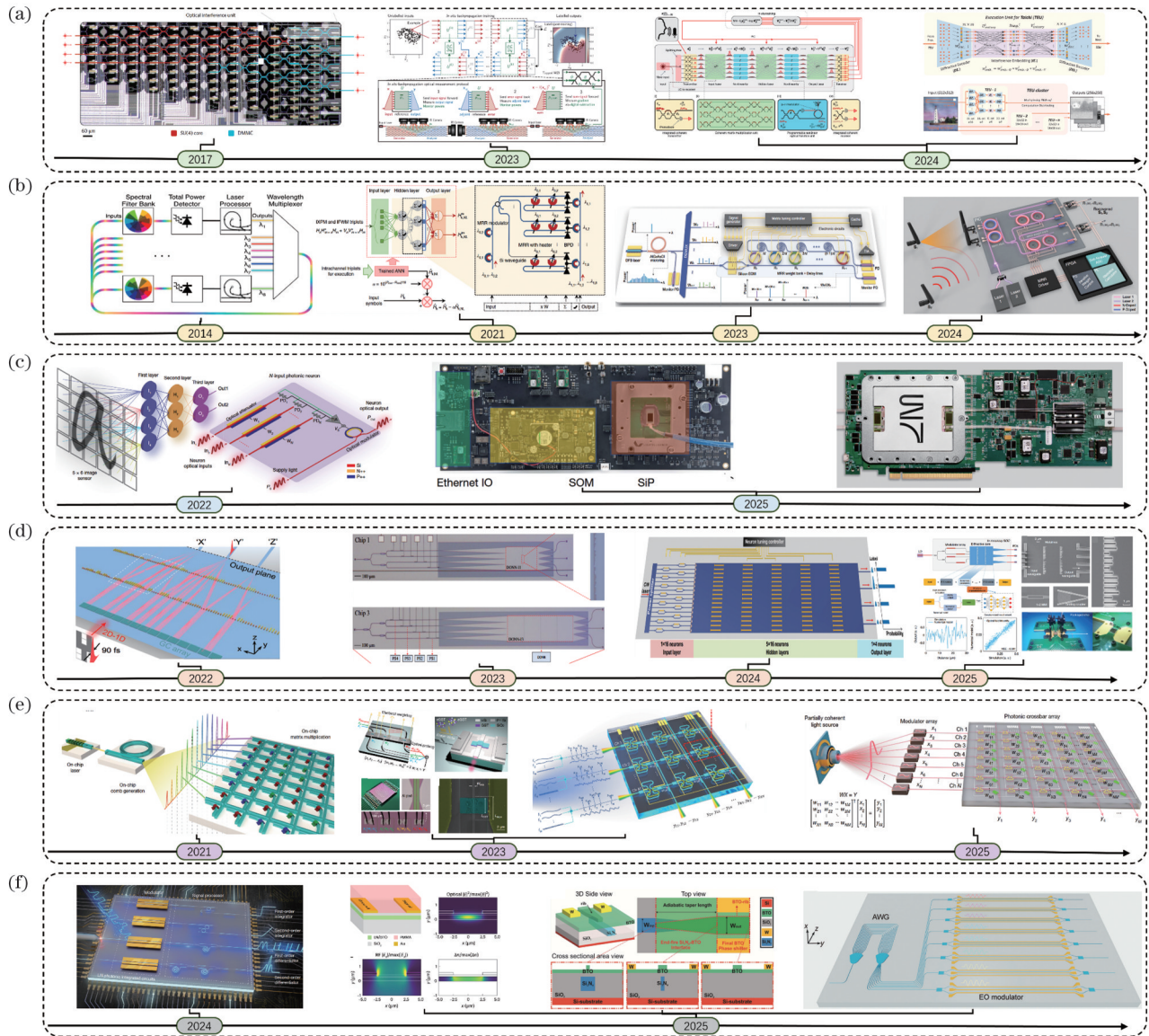


图 18 基于硅基的片上光学加速器发展时间线。(a)基于 MZI 网格<sup>[152,158-160]</sup>；(b)基于 MRR 权重库<sup>[161,163,165-166]</sup>；(c)基于强度调制阵列<sup>[154,167-168]</sup>；(d)基于片上衍射<sup>[155,170-172]</sup>；(e)基于非易失相变材料<sup>[173-176]</sup>；(f)其他材料<sup>[177-180]</sup>

Fig. 18 Development timeline of silicon-based on-chip optical computing accelerators. (a) Based on MZI meshes<sup>[152, 158-160]</sup>; (b) based on MRR weight banks<sup>[161, 163, 165-166]</sup>; (c) based on intensity-modulator arrays<sup>[154, 167-168]</sup>; (d) based on on-chip diffractions<sup>[155, 170-172]</sup>; (e) based on non-volatile phase change materials<sup>[173-176]</sup>; (f) other materials<sup>[177-180]</sup>

浙江大学提出并演示了基于绝缘体上钽酸锂 (LTOI) 的低双折射光子集成回路<sup>[180]</sup>, 通过单片集成 8 通道阵列波导光栅无源光子器件和 8 路高性能电光调制器, 实现数据速率为 1.6 Tbit/s 的波分复用光发射机。

表 3 给出了 4 种基础方案架构的优缺点对比。基于 MZI 网络的方案可实现任意实数与复数矩阵计算, 但是其基于级联 MZI 相位调控, 控制复杂度较高, 芯片尺寸大, 相较而言计算密度较低, 比较适合通用矩阵计算、量子模拟计算等。基于 MRR 权重库的方案能实现实数域矩阵计算, 芯片尺寸小, 但是波分复用技术需要  $N$  个波长的光源以及波分复用与解复用器件, 其系统复杂度增加。此外, 微环工艺的鲁棒性以及调制线性度均低于 MZI 调制器, 比较适合多通道的稀疏矩

阵计算。基于 IM array 的方案可以实现实数域矩阵计算, 具有较高的调制线性度, 控制复杂度低, 可拓展性强, 工艺鲁棒性高, 但是其计算密度有待进一步提升, 比较适合实数域矩阵加权求和计算。基于超表面衍射的方案能实现快速傅里叶变换、卷积计算等, 具有超高的计算密度, 但是其工艺鲁棒性较低, 权重不可调, 难以适应多变的任务计算需求, 比较适合固定任务的边缘计算以及全光信号预处理。在以上 4 种基本架构的基础上, 引入相变材料能有效弥补光域存储的空白, 减少静态计算功耗。此外, 通过开发新材料平台, 能进一步提升调制器的调制带宽以提升光计算的算力, 实现超低损耗的单元器件以降低片上损耗, 从而实现芯片计算规模的进一步拓展。

表 3 光学线性 MVM 加速器性能对比

Table 3 Performance comparison of optical linear MVMs

Performance	MZI mesh	MRR WB	IM array	Metasurface diffraction
Laser count	1	$N$	1	1
Wavelength division multiplexer	None	✓	None	None
Programmable	✓	✓	✓	None
Control complexity	High	Medium	Low	None
Linearity	Medium	Medium	High	Low
Scalability	High	High	High	Fixed
Computing capability	Real/complex	Real	Real	FFT
Computing density	Low	Medium	Medium	High
Process robustness	High	Low	High	Low

#### 5.4 光神经网络加速器应用场景

基于硅基的集成光神经网络加速器具有超高的计算吞吐量、较低的计算功耗,在自动驾驶、医疗诊断、金融分析、工业自动化、航空航天、内容生成、光纤通信等多个领域展现出广阔的应用前景,如图 19 所示。

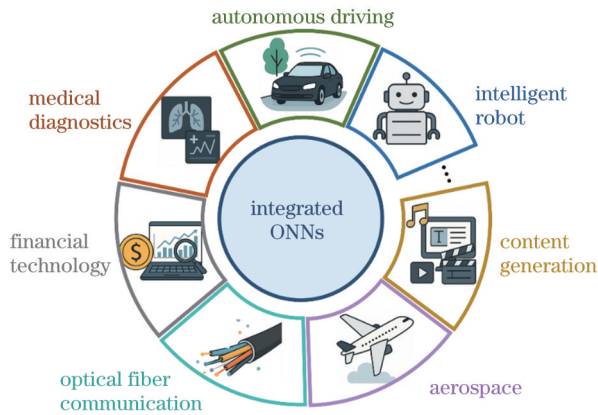


图 19 集成光计算芯片应用场景

Fig. 19 Application scenarios of integrated optical computing chips

目前,光子计算仍处于早期的发展阶段,但是已有的研究工作表明了其在特定场景计算以及通用计算中的应用前景。光通信系统中,光纤非线性是限制长距离光纤传输系统容量提升的主要障碍,而传统的采用 DSP 技术进行光纤非线性补偿的方案受限于电子计算速度且具有较大的计算复杂度。2021 年,普林斯顿大学提出一种硅基光子计算解决方案<sup>[163]</sup>,在 10080 km 的海底光纤通信系统中对光纤非线性进行模拟光域补偿,从而大大放松传统 DSP 电路对复杂度和速度的严格要求。边缘计算通常受限于电子处理速度和内存,使先进的机器学习模型无法在智能传感器和无人机等边缘设备上运行。2022 年,麻省理工学院提出一种跨网络的离域机器学习推理方案<sup>[181]</sup>,通过云端的“智能收发器”将数据流加权计算后发送到边缘设备,实现了超高效的光子推理。毫米波可以同时广播多个空间分离的数据流以提高数据传输率,但是容易受到数据流

之间的相互干扰。2023 年,普林斯顿大学提出一种使用 FPGA-光子计算芯片执行零校准控制的实时光子盲源干扰消除方案<sup>[182]</sup>,与数字电子方案相比,该基于光子的干扰消除方法可将数字化和信号恢复所需的功率降至 1/74。同年,普林斯顿大学展示了在 19.2 GHz 处理带宽下高能效的波分复用可扩展盲源分离<sup>[166]</sup>。

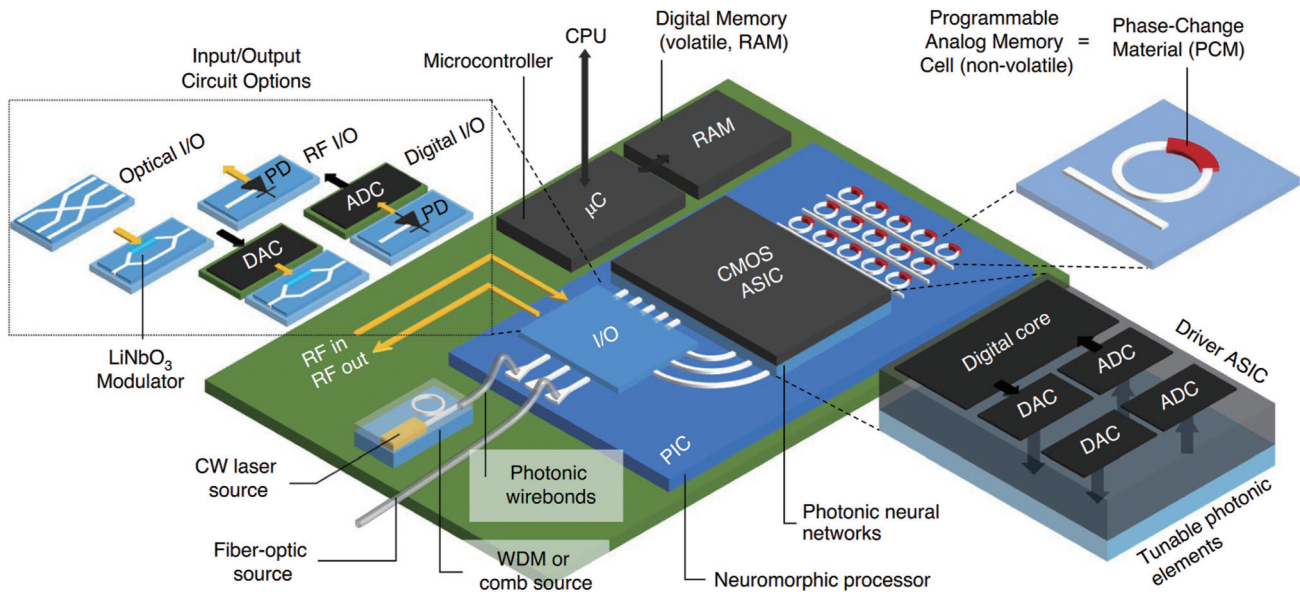
此外,光子矩阵乘法计算加速器具有计算通用性,将部分神经网络计算任务部署到光子计算芯片上,可实现图像识别分类<sup>[173]</sup>、语音识别<sup>[152]</sup>、人类表情识别<sup>[183]</sup>、医疗诊断<sup>[175]</sup>等神经网络应用场景计算加速。

#### 5.5 未来发展趋势

利用光高带宽、低延迟、低损耗的优势,将部分计算任务转移到集成光芯片以及通过硅光芯片实现电子计算芯片互连,是针对 AI 大模型算力需求的有力解决方案。图 20 为硅基光计算架构示意图,包括光计算芯片及光电协同控制电路和配套控制算法、光存储芯片、光互连芯片、电子数字内存芯片、电控制芯片、电计算芯片等。

光计算芯片通过主流的硅光子平台制备提供,包括调制器、探测器等有源器件以及波导、光栅、功分器、波分器等无源器件,用于搭建线性矩阵计算以及非线性函数计算等特定的计算网络。光计算系统对器件性能极为敏感,尤其在大规模集成时,器件损耗、相位稳定性、耦合效率等直接影响计算精度与可靠性,因此,需要开发超低损耗波导、高效相位调控器件、抗温漂的光学滤波器等关键技术。除了传统的硅基光电子器件,采用硅基集成 TFLN、BTO 和 TFLT 等具有强烈电光效应的材料有望进一步提升调制器性能。光源方面,开发在硅或兼容材料上产生受激辐射的替代光源是新兴的解决方案之一,但是目前其发光效率有待进一步提升。光频梳具有丰富的波长资源,尽管其出光功率有待进一步提升,结合片上光放大技术的发展,有望在密集光计算中发挥巨大的作用。

光存储芯片以光作为信息载体,通过光子器件来完成信息的写入、存储、读取或擦除操作,以期弥补光

图 20 硅基光计算架构示意图<sup>[184]</sup>Fig. 20 Schematic diagram of silicon optical computing architecture<sup>[184]</sup>

域存储的空白。通过硅基异质集成相变材料等非易失性材料,有望实现片上光域近零功耗信息存储,减少光计算芯片计算能耗,从而减少数据搬移带来的计算复杂度、计算能耗与计算耗时。

光电协同控制是实现光计算的关键技术之一。光电协同硬件方面,光计算芯片往往需要匹配对应的控制电路、微控制芯片以及数字存储芯片。其控制电路包括直流偏置电路、高速驱动电路、放大电路、模数转换电路、数模转换电路以及其温度控制电路等。微控制芯片用于光电协同控制的逻辑控制与调度,数字存储芯片用于缓存计算数据。与传统的光通信芯片相比,光计算芯片的集成光器件数量与控制难度呈指数上升,如何设计优化光电协同控制是实现大规模光计算的必要条件之一。高密度集成光计算集成芯片中光电协同优化设计难点主要体现在以下几个方面:1) 高速光调制器与驱动电路之间的低驱动电压和高速阻抗匹配设计优化;2) 高速探测器与放大电路之间的电流放大与带宽优化;3) 温度不敏感有源和无源器件设计与优化;4) 大规模集成超低损耗光电器件设计与制备;5) 片上非线性效应抑制以提升线性光模拟计算精度;6) 光电芯片间多计算通道的信号同步控制;7) 大规模光计算芯片中MRM等单元器件的自适应工作点锁定电路以及控制算法;8) 光电协同标准化接口的制定。此外,在传统电子模数转换器/数模转换器(ADC/DAC)逐渐遇到带宽和功耗瓶颈的背景下,基于硅光的ADC/DAC正成为下一代高速信号处理的关键方案。光ADC/DAC技术正朝着“硅光+CMOS深度融合的高集成芯片”发展,与光计算、神经网络、边缘AI平台结合潜力巨大,有望在未来类脑计算芯片中成为核心组件之一。

摩尔定律放缓和存储墙问题促使人们寻找新的计算范式,光计算凭借其天然优势在大规模并行计算和AI领域展现出巨大潜力。为了实现这一潜力,硅基光电子技术必须不断突破集成密度、互连效率、器件性能与系统可编程性等方面的限制。光计算与硅光技术的深度融合,标志着信息处理方式可能迎来一次根本性变革,未来可期。

## 6 硅光集成和光电融合面临的挑战

### 6.1 材料异质集成的复杂性

III-V族化合物半导体(如InP和GaAs)在构建有源光电器件方面具有不可替代的作用,这使得材料异质集成技术必不可少。尽管学术界和产业界已经发展出亲水芯片-晶圆键合等直接键合技术,并成功制备了通过2500 h老化测试的混合激光器<sup>[185]</sup>,但材料间固有的物理差异仍构成重大集成障碍。材料异质集成的根本复杂性源于晶格常数和热膨胀系数的显著失配。例如,GaAs与硅之间的晶格失配高达4.1%,热膨胀失配更是显著,这种失配会在界面处引入大量位错和机械应力,严重降低光学性能。以英特尔的异质集成方法为例,其要求在原子尺度实现精确对准,以最小化界面损耗<sup>[186]</sup>。然而,这种极高精度的对准过程大幅增加了大规模生产成本。

氮化硅因其超低光学损耗和与CMOS工艺的兼容性,成为一种很有前景的无源波导材料平台,推动了非线性光学和量子计算等应用的发展<sup>[187]</sup>。然而,将氮化硅与集成电路单片集成时,需要采用新型沉积技术,以避免在后端工艺过程中因应力导致的材料开裂问题。

硅材料具有中心反演对称的晶体结构,不具备线

性电光效应(泡克尔效应),因此传统硅基电光调制器主要依赖载流子色散效应来实现电光调制功能。然而,这种调制机制本身较为微弱,导致硅基高速电光调制器普遍存在插入损耗高、电光调制带宽受限等关键技术瓶颈,已难以充分满足现代高速光通信系统日益严苛的性能要求。为了突破上述技术局限并进一步拓展电光调制器的工作带宽,可以在硅光平台上异质集成高性能铁电材料,如 TFLN、BTO 等。然而,将高电光系数的铁电材料异质集成到硅光平台上仍面临以下几个方面的挑战:首先是需要克服硅基底与铁电材料之间巨大的材料物理特性差异,包括晶格常数失配、热膨胀系数不匹配等;其次是实现低温、低损伤的工艺兼容性,确保制备过程不会破坏材料的本征特性;第三是解决因光学模式失配所引起的高耦合损耗问题;最后是保证异质集成后材料的高质量以及电光器件的长期性能稳定性和可靠性。

硅光平台上集成难度更高的是磁光材料(如钇铁石榴石, Bi: YIG)的单片集成。传统的结晶过程需要超过 700 °C 的高温,这会破坏 CMOS 电子器件。通过使用局部微加热在 400 °C 退火,可以实现后端工艺兼容集成<sup>[188]</sup>,但同时也引入额外的热管理层(用于快速传导和分散局部加热产生的热量),进一步增加设计复杂性。面对这些技术难题,需要开发低热预算的集成技术。局部激光退火和等离子增强化学气相沉积(PECVD)等方法有望将热预算降低 50%,为在硅平台上集成温度敏感材料开辟新的技术路径。

## 6.2 热管理与性能稳定性

热管理是硅基光电子芯片面临的核心挑战之一。即便是微小的温度梯度,也会对系统性能产生显著影响。比如在光学相控阵中,仅 0.6 °C 的温度差就可能导光学相控阵的相位误差,严重降低波束转向精度<sup>[189]</sup>。在 CPO 架构中,单个 3.2 Tbit/s 的光学引擎可产生高达 290 W 的热量。这种极高的热密度要求采用先进的液冷解决方案,以确保器件结温保持在 51 °C 以下<sup>[190]</sup>。微流体冷却通过在互连器件中嵌入精密设计的微流体通道,可以实现局部精确的热管理。嵌入式微流体通道可有效耗散热量,但同时增加 3D 封装的复杂性。被动温度稳定技术,如无热波导设计,通常以牺牲带宽为代价换取热稳定性,极大地限制器件的实用性。

## 6.3 高密度封装技术

先进的 3D 异质集成技术是实现硅基光电子系统大规模、高性能集成的核心。以 Cisco 的 3.2 Tbit/s 光学引擎为例,其采用扇出重分布层(RDL)技术,成功实现 2 μm 线宽的高密度互连,将 16 个光学芯片高效连接到单个 ASIC 上<sup>[191]</sup>。这种技术大幅提升系统的带宽和集成度。然而,高密度互连带来了严峻的材料和工艺挑战,尤其是硅基芯片与有机基底之间的热膨胀系数(CTE)不匹配问题。在封装和键合过程中,CTE

失配会导致材料翘曲、应力集中,进而影响互连的机械稳定性和电性能,甚至引发失效。为此,需要开发低应力的封装材料、优化键合工艺以及采用先进的应力缓解设计,确保异质集成的可靠性和良率。

光纤耦合效率直接决定硅基光电子系统的整体性能。目前主流的耦合方式包括边缘耦合器和光栅耦合器,但其插入损耗普遍在 1~3 dB 之间,主要因光纤模场与硅波导模式尺寸和形状的不匹配导致光学能量传输效率降低。为降低耦合损耗,近年来出现了光子引线键合技术<sup>[192]</sup>,利用 3D 打印技术制造聚合物波导,实现光纤与芯片之间的渐进式模式匹配,插入损耗可降至 0.5 dB 以下。这种方法灵活且适应性强,能够显著提升耦合效率。尽管光子线键合技术在损耗控制上表现优异,但其在潮湿环境中的长期可靠性仍存在隐患。聚合物材料易受环境湿度影响,可能导致性能退化或机械失效。因此,提升材料的环境稳定性和封装保护技术是当前研究重点。

## 7 结语与展望

硅基光电子规模化集成技术的发展历程昭示着信息技术正在经历一场深刻变革。从最初的单个器件探索到如今的系统级集成,从概念验证到产业化应用,硅光技术已经成为后摩尔时代技术发展的核心驱动力。本文系统梳理的光通信、光感知和光计算三大应用领域的技术进展,不仅展现了硅光技术的多元化发展轨迹,更揭示了光电融合作为新技术范式的战略价值。

硅光技术引发的信息技术产业的变革,其影响远超技术本身,将重塑整个数字社会的基础设施。在光通信领域,硅基光电子技术已将单芯片数据传输容量推至 Tbit/s 级别,为数据中心和高性能计算系统提供前所未有的带宽密度。CPO 技术的兴起标志着光电融合从器件级向系统级的关键跨越,为构建百万 GPU 规模的 AI 工厂提供核心支撑。硅光技术也将成为 6G 网络和卫星互联网的技术支柱,通过实现超高速、低延迟的光学互连,构建覆盖全球的高速数据传输网络,为万物互连提供基础设施保障。在光感知领域,基于硅光平台的相控阵激光雷达技术能重新定义智能感知系统的性能边界。高精度、低成本的光学传感器使智能感知技术真正走向普及化应用,在自动驾驶、机器人导航、智能制造和智慧城市等领域推动技术革命。这种基于硅光的智能感知系统不仅提升了感知精度和响应速度,更重要的是重塑了人类与物理世界的交互方式,为智能产业的发展模式转变提供技术支撑。在光计算领域,光电协同架构展现出的能效优势和并行处理能力,为突破传统冯·诺依曼架构的限制提供全新路径。光子的超高速传输特性与电子的精确控制能力完美结合,推动计算架构从传统模式向分布式、并行化的新型计算范式转变。这种光电融合技术将为大规模 AI 应用提供前所未有的算力支撑,推动 AI 从现有的深度学

习模式向更加复杂的认知计算模式演进,有望从根本上重塑 AI 计算的基础架构。

尽管硅基光电子技术取得快速发展并不断拓展应用领域,但其规模化集成仍面临严峻挑战,能否解决这些问题将决定光电融合技术能否实现产业化突破。材料异质集成的复杂性构成最根本的挑战,不同材料的晶格失配和热膨胀差异产生的界面缺陷和机械应力既影响器件光学性能又制约大规模生产,需从材料科学基础理论出发来发展新的界面工程和应力缓解技术;大规模器件阵列的精确调控与良率提升是另一关键挑战,随着集成密度的提升,工艺偏差对器件性能影响呈指数级放大,传统统计工艺控制已无法满足要求,需从器件物理层面理解偏差传播机制并发展基于机器学习的智能化工艺控制技术;系统级封装与热管理挑战则源于光电融合系统的固有复杂性,光子器件的温度敏感性与电子器件发热特性形成天然矛盾,需在材料、结构和系统设计层面协同创新以实现高性能前提下的有效热管理。

未来十年,硅基光电子技术将站在一个历史性的转折点上。随着摩尔定律放缓和数据爆炸式增长的双重压力,传统的平面集成技术路径已接近物理极限。在这一背景下,三维集成、芯粒化设计和 AI 融合三大技术路径的协同演进,不仅是技术发展的自然延续,更是产业转型的战略必然。这种多维度的技术融合体现了现代信息技术发展的一个重要特征:单一技术突破已难以支撑产业的跃升,只有通过多技术域的深度协同,才能实现真正的范式变革。

三维集成技术的核心价值在于突破传统二维平面的空间限制,通过垂直堆叠实现更高的集成密度和更短的互连距离。然而,这一技术路径同样面临着严峻挑战:垂直堆叠界面的可靠性、高密度互连的信号完整性,以及三维结构内部的热管理,每一个问题都可能成为整体性能提升的瓶颈。从更深层次来看,三维集成代表从“面积驱动”向“体积驱动”的设计哲学转变,这要求研究人员重新审视传统的热设计、电磁兼容和机械可靠性等理念。硅中介层技术的发展和微流道冷却系统的集成,正是对这一设计哲学转变的具体响应。芯粒化设计通过将复杂系统分解为功能明确的标准化模块,大大降低设计复杂度和开发成本,也为产业生态的水平分工创造条件。芯粒化的成功关键在于接口标准的建立和生态互操作性的实现。高速光电接口标准制定,不仅是技术规范的统一,更是产业话语权的重新分配。AI 与硅基光电子技术的深度融合,标志着这一领域从经验驱动向数据驱动的根本转变。深度学习逆向设计技术能够突破传统正向设计的局限,通过物理约束下的智能优化,实现超紧凑硅光器件结构的快速生成。这不仅提升设计效率,也拓展设计空间的边界,能实现许多传统方法无法实现的器件结构。同时,AI 在硅光制造工艺优化中的应用,通过实时监测和闭环

控制,能够有效补偿制程偏差,提高良率和一致性。

这三大技术路径的真正价值并非在于各自的独立发展,而在于它们之间的深度协同和相互赋能。三维集成成为芯粒化提供更灵活的空间布局方案,芯粒化标准为异质异构集成创造规模化的应用场景,而 AI 技术则为这两者的优化设计和可靠制造提供强大工具。这种协同演进模式将重新定义硅光产业的竞争格局:传统的单点技术优势将被系统性的协同能力所取代,产业价值将更多地集中在标准制定、生态构建和系统集成等高附加值环节。对于参与者而言,这既是机遇也是挑战——那些能够在多技术域实现协同突破的团队,将在未来竞争中取得决定性优势。

硅基光电子规模化集成技术作为后摩尔时代的关键使能技术,正在从通信基础设施、智能感知应用到计算架构等核心领域推动根本性变革,为数字社会的未来发展描绘全新蓝图。这一技术的影响远超技术性能本身的提升,正在深刻改变产业分工模式、商业生态结构以及人类社会的数字化生存方式。光电融合不仅是技术演进的必然选择,更是人类信息处理能力实现历史性跃升的重要机遇。通过持续的技术创新和产业协同,硅基光电子技术必将推动信息技术向更高效、更智能的方向发展,成为开启智能时代新篇章的关键钥匙,为构建智能化的未来社会提供强有力的技术支撑。

#### 参 考 文 献

- [1] Shekhar S, Bogaerts W, Chrostowski L, et al. Roadmapping the next generation of silicon photonics[J]. Nature Communications, 2024, 15: 751.
- [2] Poon J K S, Govdeli A, Sharma A, et al. Silicon photonics for the visible and near-infrared spectrum[J]. Advances in Optics and Photonics, 2024, 16(1): 1-59.
- [3] Xiao Z A, Liu W X, Xu S Y, et al. Recent progress in silicon-based photonic integrated circuits and emerging applications[J]. Advanced Optical Materials, 2023, 11(20): 2301028.
- [4] NVIDIA. NVIDIA announces spectrum-X photonics, co-packaged optics networking switches to scale AI factories millions of GPUs[EB/OL]. [2025-06-10]. <https://nvidianews.nvidia.com/news/nvidia-spectrum-x-co-packaged-optics-networking-switches-ai-factories>.
- [5] Ferraro F J, de Heyn P, Kim M, et al. Imec silicon photonics platforms: performance overview and roadmap[J]. Proceedings of SPIE, 2023, 12429: 1242909.
- [6] Zhang Y, Guo X H, Ji X C, et al. What can be integrated on the silicon photonics platform and how?[J]. APL Photonics, 2024, 9(9): 090902.
- [7] Mitchell C J, Hu T H, Sun S Y, et al. Mid-infrared silicon photonics: from benchtop to real-world applications[J]. APL Photonics, 2024, 9(8): 080901.
- [8] Namiki S. Digital infrastructure pivots on silicon photonics: an aspect from the past, present, and future of optical communications[C]//2024 IEEE Silicon Photonics Conference (SiPhotonics), April 15-18, 2024, Tokyo Bay, Japan. New York: IEEE Press, 2024: 1-2.
- [9] Amanti F, Andrini G, Armani F, et al. Integrated photonic passive building blocks on silicon-on-insulator platform[J]. Photonics, 2024, 11(6): 494.
- [10] Zhang L, Hong S H, Wang Y, et al. Ultralow-loss silicon

- photonics beyond the singlemode regime[J]. *Laser & Photonics Reviews*, 2022, 16(4): 2100292.
- [11] Li Z A, Zhou Z P, Qiu C, et al. The intelligent design of silicon photonic devices[J]. *Advanced Optical Materials*, 2024, 12(7): 2301337.
- [12] Guo J S, Yu L W, Xiang H T, et al. Realization of advanced passive silicon photonic devices with subwavelength grating structures developed by efficient inverse design[J]. *Advanced Photonics Nexus*, 2023, 2(2): 026005.
- [13] Wang H R, Zuo Y, Yin X F, et al. Ultralow-loss optical interconnect enabled by topological unidirectional guided resonance[J]. *Science Advances*, 2024, 10(12): eadn4372.
- [14] Sánchez-Postigo A, Ginel-Moreno P, Schmid J H, et al. Ultra-efficient surface grating couplers for chip-to-fiber and chip-to-free-space coupling based on single-beam radiation[J]. *Optics Express*, 2024, 32(23): 41156-41172.
- [15] Yi X L, Zhao W K, Zhang L, et al. Polarization-insensitive and low-loss O-band edge coupler for silicon photonics[J]. *Optics Letters*, 2025, 50(5): 1699-1702.
- [16] Buzaverov K A, Baburin A S, Sergeev E V, et al. Silicon nitride integrated photonics from visible to mid-infrared spectra [J]. *Laser & Photonics Reviews*, 2024, 18(12): 2400508.
- [17] Tran M A, Zhang C, Morin T J, et al. Extending the spectrum of fully integrated photonics to submicrometre wavelengths[J]. *Nature*, 2022, 610(7930): 54-60.
- [18] Xiang C, Jin W, Huang D N, et al. High-performance silicon photonics using heterogeneous integration[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2022, 28(3): 8200515.
- [19] Liang D, Bowers J E. Recent progress in heterogeneous III-V-on-silicon photonic integration[J]. *Light: Advanced Manufacturing*, 2021, 2(1): 59-83.
- [20] Patel D, Ghosh S, Chagnon M, et al. Design, analysis, and transmission system performance of a 41 GHz silicon photonic modulator[J]. *Optics Express*, 2015, 23(11): 14263-14287.
- [21] Li M F, Wang L, Li X, et al. Silicon intensity Mach-Zehnder modulator for single lane 100 Gb/s applications[J]. *Photonics Research*, 2018, 6(2): 109-116.
- [22] Mohammadi A, Zheng Z B, Lin J C, et al. Segmented silicon photonic modulator with a 67-GHz bandwidth for high-speed signaling[C]//2022 Optical Fiber Communications Conference and Exhibition (OFC), March 6-10, 2022, San Diego, CA, USA. New York: IEEE Press, 2022: 1-3.
- [23] Mohammadi A, Safarnejadian A, Zheng Z B, et al. High-bandwidth silicon traveling-wave modulator using distributed micro-capacitors[J]. *Journal of Lightwave Technology*, 2024, 42(17): 5942-5948.
- [24] Han C H, Zheng Z, Shu H W, et al. Slow-light silicon modulator with 110-GHz bandwidth[J]. *Science Advances*, 2023, 9(42): eadi5339.
- [25] Kawahara K, Tsuchizawa T, Yamamoto N, et al. High-speed, low-voltage, low-bit-energy silicon photonic crystal slow-light modulator with impedance-engineered distributed electrodes[J]. *Optica*, 2024, 11(9): 1212-1219.
- [26] Yamaguchi Y, Matsumoto A, Dat P T, et al. 60-GHz-bandwidth silicon optical modulator utilizing electro-optic frequency-domain equalizer[C]//ECOC 2024; 50th European Conference on Optical Communication, September 22-26, 2024, Frankfurt, Germany. London: VDE, 2024: 695-697.
- [27] Yue H S, Fu J B, Zhang H W, et al. Silicon modulator exceeding 110 GHz using tunable time-frequency equalization[J]. *Optica*, 2025, 12(2): 203-215.
- [28] Wang H B, Tan J Y, Tang H Z, et al. 90 GHz silicon Mach-Zehnder modulator with integrated equalizer for 1.6 Tbps (200 G/λ) IMDD transceivers[C]//Optical Fiber Communication Conference (OFC) 2025, March 30-April 3, 2025, San Francisco, California. Washington, DC: Optica Publishing Group, 2025: M2K.1.
- [29] Li K, Thomson D J, Zhou L J, et al. Beyond 300 Gb/s from an integrated single-channel silicon photonics modulator driver combination[C]//2024 IEEE Silicon Photonics Conference (SiPhotonics), April 15-18, 2024, Tokyo Bay, Japan. New York: IEEE Press, 2024: 1-2.
- [30] Gao J, Zhu K H, Wu H. A 50-GHz bandwidth traveling-wave Mach-Zehnder modulator with built-in feedback equalization[J]. *Journal of Lightwave Technology*, 2022, 40(12): 3872-3881.
- [31] Alam M S, Li X Y, Jacques M, et al. Net 220 Gbps/λ IM/DD transmission in O-band and C-band with silicon photonic traveling-wave MZM[J]. *Journal of Lightwave Technology*, 2021, 39(13): 4270-4278.
- [32] Hu X, Wu D Y, Chen D G, et al. 280 Gbit/s PAM-4 Ge/Si electro-absorption modulator with 3-dB bandwidth beyond 110 GHz[C]//2023 Optical Fiber Communications Conference and Exhibition (OFC), March 5-9, 2023, San Diego, CA, USA. New York: IEEE Press, 2023: 1-3.
- [33] Wakita H, Nagatani M, Ogiso Y, et al. InP-based on-board coherent tx front-end with 100-GHz bandwidth utilizing flip-chip technology enabling 2-Tb/s/λ optical transmission[J]. *Journal of Lightwave Technology*, 2025, 43(4): 1819-1825.
- [34] Ozaki J, Ogiso Y, Yamazaki H, et al. C+L-band InP-based coherent driver modulator enabled net-1.8 Tbps/λ transmission [J]. *Journal of Lightwave Technology*, 2025, 43(4): 1972-1978.
- [35] Kohli M, Chelladurai D, Messner A, et al. Plasmonic ferroelectric modulator monolithically integrated on SiN for 216 GBd data transmission[J]. *Journal of Lightwave Technology*, 2023, 41(12): 3825-3831.
- [36] Moor D, Winiger J, Ma P, et al. >150 GHz hybrid-plasmonic BaTiO<sub>3</sub>-on-SOI modulator for CMOS foundry integration[C]//Frontiers in Optics + Laser Science 2021, November 1-4, 2021, Washington, DC. Washington, DC: Optica Publishing Group, 2021: FTh4D.2.
- [37] Rahim A, Hermans A, Wohlfeil B, et al. Taking silicon photonics modulators to a higher performance level: state-of-the-art and a review of new technologies[J]. *Advanced Photonics*, 2021, 3(2): 024003.
- [38] Yuan Y, Peng Y W, Sorin W V, et al. A 5×200 Gbps microring modulator silicon chip empowered by two-segment Z-shape junctions[J]. *Nature Communications*, 2024, 15: 818.
- [39] Hsu W C, Nujhat N, Kupp B, et al. Sub-volt high-speed silicon MOSCAP microring modulator driven by high-mobility conductive oxide[J]. *Nature Communications*, 2024, 15: 826.
- [40] Geravand A, Weckenmann E, Zheng Z B, et al. Micrometer-scale silicon modulator for O-band coherent interconnects beyond 100 GBaud[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2025, 31(3): 6100209.
- [41] Sun J, Kumar R, Sakib M, et al. A 128 Gb/s PAM4 silicon microring modulator with integrated thermo-optic resonance tuning[J]. *Journal of Lightwave Technology*, 2019, 37(1): 110-115.
- [42] Ban Y, Verbist J, Vanhooecke M, et al. Low-voltage 60 Gb/s NRZ and 100 Gb/s PAM4 O-band silicon ring modulator[C]//2019 IEEE Optical Interconnects Conference (OI), April 24-26, 2019, Santa Fe, NM, USA. New York: IEEE Press, 2019: 1-2.
- [43] Yu H, Ying D Q, Pantouvaki M, et al. Trade-off between optical modulation amplitude and modulation bandwidth of silicon micro-ring modulators[J]. *Optics Express*, 2014, 22(12): 15178-15189.
- [44] Müller J, Merget F, Azadeh S S, et al. Optical peaking enhancement in high-speed ring modulators[J]. *Scientific Reports*, 2014, 4: 6310.
- [45] Zhang Y G, Zhang H G, Li M F, et al. 200 Gbit/s optical PAM4 modulation based on silicon microring modulator[C]//2020 European Conference on Optical Communications (ECOC), December 6-10, 2020, Brussels, Belgium. New

- York: IEEE Press, 2020: 1-4.
- [46] Sakib M, Kumar R, Ma C X, et al. **A 240 Gb/s PAM4 silicon micro-ring optical modulator**[C]//Optical Fiber Communication Conference (OFC) 2022, March 6–10, 2022, San Diego, California. Washington, DC: Optica Publishing Group, 2022: M2D.4.
- [47] Kim M, Rahimi J, Lepage G, et al. **A 40 Gb/s NRZ O-band silicon disk modulator with 5.4 THz FSR and 60 GHz/mW heater efficiency**[C]//ECOC 2024; 50th European Conference on Optical Communication, September 22–26, 2024, Frankfurt, Germany. London: VDE, 2024: 1980-1983.
- [48] Xie C, Raj M, Joshi A, et al. **A 64 Gb/s NRZ O-band ring modulator with 3.2 THz FSR for DWDM applications**[C]//2024 Optical Fiber Communications Conference and Exhibition (OFC), March 24–28, 2024, San Diego, CA, USA. New York: IEEE Press, 2024: 1-3.
- [49] Novack A, Gould M, Yang Y S, et al. **Germanium photodetector with 60 GHz bandwidth using inductive gain peaking**[J]. *Optics Express*, 2013, 21(23): 28387-28393.
- [50] Chen G Y, Yu Y, Deng S P, et al. **Bandwidth improvement for germanium photodetector using wire bonding technology**[J]. *Optics Express*, 2015, 23(20): 25700-25706.
- [51] Lischke S, Peczek A, Morgan J S, et al. **Ultra-fast germanium photodiode with 3-dB bandwidth of 265 GHz**[J]. *Nature Photonics*, 2021, 15(12): 925-931.
- [52] Marzen S, Postelnicu E, Michel J, et al. **High performance germanium on silicon photodiodes for back-end-of-line photonic integration**[J]. *Applied Physics Letters*, 2023, 123(11): 111105.
- [53] Chen D G, Zhang H G, Liu M, et al. **67 GHz light-trapping-structure germanium photodetector supporting 240 Gb/s PAM-4 transmission**[J]. *Photonics Research*, 2022, 10(9): 2165-2171.
- [54] Hu X, Wu D Y, Liu Y, et al. **408 Gbit/s PAM-8 sidewall-doped germanium-silicon photodetector**[J]. *Photonics Research*, 2023, 11(6): 961-967.
- [55] Chen W, Shi Y, Wang Y L, et al. **High-performance compact vertical germanium photodiodes enabled by silicon corner reflectors**[J]. *Photonics Research*, 2024, 12(12): 2983-2988.
- [56] Shi Y, Li X, Zou M J, et al. **103 GHz germanium-on-silicon photodiode enabled by an optimized U-shaped electrode**[J]. *Photonics Research*, 2024, 12(1): 1-6.
- [57] Wang B H, Mu J F. **High-speed Si-Ge avalanche photodiodes**[J]. *Photonix*, 2022, 3(1): 8.
- [58] Liu T H, Yang G H, et al. **High-speed avalanche photodiodes for optical communication**[J]. *Photonics Research*, 2025, 13(6): 1438-1451.
- [59] Kim G, Kim S, Kim S A, et al. **NDR-effect vertical-illumination-type Ge-on-Si avalanche photodetector**[J]. *Optics Letters*, 2018, 43(22): 5583-5586.
- [60] Yuan Y, Huang Z H, Zeng X G, et al. **High responsivity Si-Ge waveguide avalanche photodiodes enhanced by loop reflector**[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2022, 28(2): 3800508.
- [61] Shi Y, Li X, Chen G Y, et al. **Avalanche photodiode with ultrahigh gain - bandwidth product of 1, 033 GHz**[J]. *Nature Photonics*, 2024, 18(6): 610-616.
- [62] Peng Y W, Yuan Y, Sorin W V, et al. **An 8×160 Gb s<sup>-1</sup> all-silicon avalanche photodiode chip**[J]. *Nature Photonics*, 2024, 18(9): 928-934.
- [63] Li X Y, Li Z Y, Xiao X, et al. **40 Gb/s all-silicon photodetector based on microring resonators**[J]. *IEEE Photonics Technology Letters*, 2015, 27(7): 729-732.
- [64] Zhu H K, Goi K, Ogawa K. **All-silicon waveguide photodetection for low-bias power monitoring and 20-km 28-Gb/s NRZ-OOK signal transmission**[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2018, 24(2): 4400207.
- [65] Peng Y W, Yuan Y, Sorin W V, et al. **All-silicon microring avalanche photodiodes with a >65 A/W response**[J]. *Optics Letters*, 2023, 48(5): 1315-1318.
- [66] Sun W C, Cao H Z, Xie J, et al. **Ultralow-LO-power silicon photonic coherent receiver with balanced Ge/Si avalanche photodiodes**[J]. *ACS Photonics*, 2025, 12(4): 1793-1801.
- [67] Wu Q C, Qian J, Wang Y C, et al. **Waveguide-integrated twisted bilayer graphene photodetectors**[J]. *Nature Communications*, 2024, 15: 3688.
- [68] Yu L W, Guo J S, Xiang H T, et al. **High-bandwidth zero-biased waveguide-integrated p-n homojunction graphene photodetectors on silicon for a wavelength band of 1.55 μm and beyond**[J]. *ACS Photonics*, 2023, 10(10): 3621-3628.
- [69] Bian Y S, Ramachandran K, Wu Z J, et al. **3D integrated laser attach technology on a 300-mm monolithic CMOS silicon photonics platform**[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2023, 29(3): 8200519.
- [70] Nallamothu N S, Muduli M, Xia Y K, et al. **Heterogeneous integration of GaAsSb and silicon via direct wafer bonding for high-speed avalanche photodiodes**[J]. *Proceedings of SPIE*, 2025, 13372: 133720B.
- [71] Kamei S, Ishii Y. **Photonics-electronics convergence devices enabling IOWN: development of second- and third-generation devices**[J]. *NTT Technical Review*, 2024, 22(1): 73-79.
- [72] Lu C H, Lin C C, Tang T C, et al. **High bandwidth and energy efficient electrical-optical system integration using COUPE technology**[C]//2024 IEEE 74th Electronic Components and Technology Conference (ECTC), May 28–31, 2024, Denver, CO, USA. New York: IEEE Press, 2024: 893-897.
- [73] Statista. **Number of internet users worldwide from 2005 to 2024**[EB/OL]. [2024-12-12]. <https://www.statista.com/statistics/273018/number-of-internet-users-worldwide/>.
- [74] Sinha S. **State of IoT 2024: number of connected IoT devices growing 13% to 18.8 billion globally**[EB/OL]. [2024-12-12]. <https://thespicyinvestments.com/2024/09/04/state-of-iot-2024-number-of-connected-iot-devices-growing-13-to-18-8-billion-globally/>.
- [75] Ossieur P, Moeneclaey B, Coudyzer G, et al. **Integrated photonics and electronics for optical transceivers supporting AI/ML applications**[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2025, 31(3): 6000116.
- [76] Bi J, Wang Z Q, Yuan H T, et al. **Large AI models and their applications: classification, limitations, and potential solutions**[J]. *Software: Practice and Experience*, 2025, 55(6): 1003-1017.
- [77] IDC. **Worldwide IDC Global DataSphere Forecast, 2024 - 2028: AI Everywhere, But Upsurge in Data Will Take Time**[EB/OL]. [2024-12-12]. <https://my.idc.com/getdoc.jsp?containerId=US52076424>.
- [78] Winzer P J. **The future of communications is massively parallel**[J]. *Journal of Optical Communications and Networking*, 2023, 15(10): 783-787.
- [79] Jones N. **The information factories**[J]. *Nature*, 2018, 561(7722): 163-166.
- [80] Andrae A, Edler T. **On global electricity usage of communication technology: trends to 2030**[J]. *Challenges*, 2015, 6(1): 117-157.
- [81] GroupYole. **Silicon photonics: from transceivers to speed-of-light AI**[EB/OL]. [2024-12-12]. <https://www.yolegroup.com/strategy-insights/silicon-photonics-from-transceivers-to-speed-of-light-ai/>.
- [82] Margalit N, Xiang C, Bowers S M, et al. **Perspective on the future of silicon photonics and electronics**[J]. *Applied Physics Letters*, 2021, 118(22): 220501.
- [83] Razdan S, De Dobbelaere P, Xue J, et al. **Advanced 2.5D and 3D packaging technologies for next generation Silicon Photonics in high performance networking applications**[C]//2022 IEEE 72nd Electronic Components and Technology Conference (ECTC), May 31–June 3, 2022, San Diego, CA, USA. New York: IEEE Press, 2022: 428-435.
- [84] Shi J B, Jin M, Yang T, et al. **16-channel photonic - electric co-**

- designed silicon transmitter with ultra-low power consumption [J]. *Photonics Research*, 2023, 11(2): 143-149.
- [85] de Valicourt G, Pupalaiakis P, Giles R, et al. 1.6-Tbps low-power linear-drive high-density optical interface for machine learning/artificial intelligence[J]. *Optics Express*, 2025, 33(7): 15338-15354.
- [86] Daudlin S, Rizzo A, Lee S, et al. Three-dimensional photonic integration for ultra-low-energy, high-bandwidth interchip data links[J]. *Nature Photonics*, 2025, 19(5): 502-509.
- [87] Zhang Z Y, Li M X, Lv X Q, et al. Grating-based eight-channel Lan-WDM silicon photonic transceiver for Tb/s applications[J]. *Laser & Photonics Reviews*, 2025, 19(4): 2401249.
- [88] Omirzakhov K, Hao H, Pirmoradi A, et al. Energy efficient monolithically integrated 256 Gb/s optical transmitter with autonomous wavelength stabilization in 45 nm CMOS SOI[J]. *IEEE Journal of Solid-State Circuits*, 2024, 60: 1-10.
- [89] Xue J T, Wang B H, Chen Y H, et al. A  $4 \times 112$  Gb/s ultra-compact polarization-insensitive silicon photonics WDM receiver with CMOS TIA for co-packaged optics and optical I/O[J]. *Journal of Lightwave Technology*, 2024, 42(17): 6028-6035.
- [90] Rizzo A, Novick A, Gopal V, et al. Massively scalable Kerr comb-driven silicon photonic link[J]. *Nature Photonics*, 2023, 17(9): 781-790.
- [91] Michel S. Data center Ethernet on the move to 224 Gbps[EB/OL]. [2024-12-12]. <https://www.keysight.com/blogs/en/inds/2021/12/01/data-center-ethernet-on-the-move-to-224-gbps>.
- [92] Bernal S, Dumont M, Berikaa E, et al. 12.1 terabit/second data center interconnects using O-band coherent transmission with QD-MLL frequency combs[J]. *Nature Communications*, 2024, 15: 7741.
- [93] Lau J H. Co-packaged optics: heterogeneous integration of photonic integrated circuits and electronic integrated circuits[J]. *Journal of Electronic Packaging*, 2025, 147(1): 011004.
- [94] 讯石光通讯网. 1.6T 光模块预计 2026 年起量 关键技术百花齐放 [EB/OL]. [2025-02-05]. <http://www.iccsz.com/site/cn/News/2025/04/21/20250421163516336622.htm>.  
ICCSZ. 1.6T optical module is expected to start production in 2026, with a variety of key technologies emerging[EB/OL]. [2025-02-05]. <http://www.iccsz.com/site/cn/News/2025/04/21/20250421163516336622.htm>.
- [95] Group Yole. The AI-driven optical transceiver market in datacom will see phenomenal growth of 45% in 2024 (year over year) [EB/OL]. [2024-12-12]. <https://www.yolegroup.com/press-release/yg-press-news-ai-driven-optical-transceiver-market-towards-the-belle-epoque/>.
- [96] BROADCOM. Broadcom announces third-generation co-packaged optics (CPO) technology with 200G/lane capability [EB/OL]. [2025-02-05]. <https://investors.broadcom.com/news-releases/news-release-details/broadcom-announces-third-generation-co-packaged-optics-cpo>.
- [97] Zhao Y, Doerr C, Liu F Y, et al. 425-Gbps/ $\lambda$  dual-polarization IMDD transceiver[C]//Optical Fiber Communication Conference (OFC) Postdeadline Papers 2025, March 30–April 3, 2025, San Francisco, California. Washington, DC: Optica Publishing Group, 2025: Th4B.5.
- [98] Ahmed A H, Vera L, Iotti L, et al. A dual-polarization silicon-photonic coherent receiver front-end supporting 528 Gb/s/wavelength[J]. *IEEE Journal of Solid-State Circuits*, 2023, 58(8): 2202-2213.
- [99] PhotonDelta. Integrated photonics systems roadmap 2023[EB/OL]. [2025-06-10]. <https://www.photondelta.com/downloads/>.
- [100] Yole Group. Automotive LiDAR deployment ramps up in 2024 [EB/OL]. [2025-06-10]. <https://www.yolegroup.com/strategy-insights/automotive-lidar-deployment-ramps-up-in-2024/>.
- [101] Rogers C, Piggott A Y, Thomson D J, et al. A universal 3D imaging sensor on a silicon photonics platform[J]. *Nature*, 2021, 590(7845): 256-261.
- [102] Takeuchi N, Sugimoto N, Baba H, et al. Random modulation cw lidar[J]. *Applied Optics*, 1983, 22(9): 1382-1386.
- [103] Wang C X, Lee W B, Sunwoo Y H, et al. Angularly offset multiline dispersive optical phased array enabling large field of view and plateau envelope[J]. *Optics Express*, 2024, 32(8): 13048-13064.
- [104] Xu W H, Zhou L J, Lu L J, et al. Aliasing-free optical phased array beam-steering with a plateau envelope[J]. *Optics Express*, 2019, 27(3): 3354-3368.
- [105] Xu W H, Guo Y Y, Li X H, et al. Fully integrated solid-state LiDAR transmitter on a multi-layer silicon-nitride-on-silicon photonic platform[J]. *Journal of Lightwave Technology*, 2023, 41(3): 832-840.
- [106] Sun J, Timurdogan E, Yaacobi A, et al. Large-scale nanophotonic phased array[J]. *Nature*, 2013, 493(7431): 195-199.
- [107] Yu L, Wang P F, Ma P F, et al. Two-dimensional beam scanning of passive optical phased array based on silicon nitride delay line[J]. *Journal of Lightwave Technology*, 2023, 41(9): 2756-2764.
- [108] Liu C X, Lu L J, Guo Y Y, et al. Hybrid integrated frequency-modulated continuous-wave laser with synchronous tuning[J]. *Journal of Lightwave Technology*, 2022, 40(16): 5636-5645.
- [109] Pita J L, Aldaya I, Santana O J S, et al. Side-lobe level reduction in bio-inspired optical phased-array antennas[J]. *Optics Express*, 2017, 25(24): 30105-30114.
- [110] Fatemi R, Khachaturian A, Hajimiri A. A nonuniform sparse 2-D large-FOV optical phased array with a low-power PWM drive[J]. *IEEE Journal of Solid-State Circuits*, 2019, 54(5): 1200-1215.
- [111] Fukui T, Tanomura R, Komatsu K, et al. Non-redundant optical phased array[J]. *Optica*, 2021, 8(10): 1350-1358.
- [112] Khachaturian A, Fatemi R, Darbinian A, et al. Discretization of annular-ring diffraction pattern for large-scale photonics beamforming[J]. *Photonics Research*, 2022, 10(5): 1177-1186.
- [113] Moss B R, Poulton C V, Byrd M J, et al. A 2048-channel, 125  $\mu$ W/ch DAC controlling a 9, 216-element optical phased array coherent solid-state LiDAR[C]//2023 IEEE Symposium on VLSI Technology and Circuits (VLSI Technology and Circuits), June 11–16, 2023, Kyoto, Japan. New York: IEEE Press, 2023: 1-2.
- [114] Van Acoleyen K, Bogaerts W, Jágerská J, et al. Off-chip beam steering with a one-dimensional optical phased array on silicon-on-insulator[J]. *Optics Letters*, 2009, 34(9): 1477-1479.
- [115] Hutchison D N, Sun J, Doylend J K, et al. High-resolution aliasing-free optical beam steering[J]. *Optica*, 2016, 3(8): 887-890.
- [116] Poulton C V, Byrd M J, Raval M, et al. Large-scale silicon nitride nanophotonic phased arrays at infrared and visible wavelengths[J]. *Optics Letters*, 2017, 42(1): 21-24.
- [117] Poulton C V, Byrd M J, Russo P, et al. Coherent LiDAR with an 8, 192-element optical phased array and driving laser[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2022, 28(5): 6100508.
- [118] Chen T H, Huang C Y, Shia T K, et al. A frequency digital pre-distortion compensation method for FMCW LiDAR system [C]//2020 Optical Fiber Communications Conference and Exhibition (OFC), March 8–12, 2020, San Diego, CA, USA. New York: IEEE Press, 2020: 1-3.
- [119] Zhang X S, Pouls J, Wu M C. Laser frequency sweep linearization by iterative learning pre-distortion for FMCW LiDAR[J]. *Optics Express*, 2019, 27(7): 9965-9974.
- [120] Cao X Y, Wu K, Li C, et al. Highly efficient iteration algorithm for a linear frequency-sweep distributed feedback laser in frequency-modulated continuous wave lidar applications[J]. *Journal of the Optical Society of America B Optical Physics*, 2021, 38(10): D8-D14.

- [121] Tang L W, Jia H X, Shao S, et al. Hybrid integrated low-noise linear chirp frequency-modulated continuous-wave laser source based on self-injection to an external cavity[J]. *Photonics Research*, 2021, 9(10): 1948-1957.
- [122] Zhang G, Ding Z H, Wang K K, et al. Demonstration of high output power DBR laser integrated with SOA for the FMCW LiDAR system[J]. *Optics Express*, 2022, 30(2): 2599-2609.
- [123] Hauser M, Hofbauer M. FPGA-based EO-PLL with repetitive control for highly linear laser frequency tuning in FMCW LIDAR applications[J]. *IEEE Photonics Journal*, 2022, 14(1): 6808608.
- [124] Qin J, Zhou Q, Xie W L, et al. Coherence enhancement of a chirped DFB laser for frequency-modulated continuous-wave reflectometry using a composite feedback loop[J]. *Optics Letters*, 2015, 40(19): 4500-4503.
- [125] Snigirev V, Riedhauser A, Lihachev G, et al. Ultrafast tunable lasers using lithium niobate integrated photonics[J]. *Nature*, 2023, 615(7952): 411-417.
- [126] Liu C X, Guo Y Y, Xu R Y, et al. Highly-linear and wavelength-tunable frequency-modulated continuous-wave hybrid-integrated laser[J]. *Laser & Photonics Reviews*, 2024, 18(7): 2300882.
- [127] Liu C X, Guo Y Y, Zhou Y Y, et al. Fast-tuning and narrow-linewidth hybrid laser for FMCW ranging[J]. *Laser & Photonics Reviews*, 2025, 19(8): 2401338.
- [128] Poulton C V, Yaacobi A, Cole D B, et al. Coherent solid-state LIDAR with silicon photonic optical phased arrays[J]. *Optics Letters*, 2017, 42(20): 4091-4094.
- [129] Poulton C V, Byrd M J, Russo P, et al. Long-range LiDAR and free-space data communication with high-performance optical phased arrays[J]. *IEEE Journal of Selected Topics in Quantum Electronics*, 2019, 25(5): 7700108.
- [130] Bhargava P, Kim T, Poulton C V, et al. Fully integrated coherent LiDAR in 3D-integrated silicon photonics/65 nm CMOS[C]//2019 Symposium on VLSI Circuits, June 9–14, 2019, Kyoto, Japan. New York: IEEE Press, 2019: C262-C263.
- [131] Chung S, Nakai M, Idres S, et al. 19.1 optical phased-array FMCW LiDAR with on-chip calibration[C]//2021 IEEE International Solid – State Circuits Conference (ISSCC), February 13–22, 2021, San Francisco, CA, USA. New York: IEEE Press, 2021: 286-288.
- [132] Li Y Z, Chen B S, Na Q X, et al. Wide-steering-angle high-resolution optical phased array[J]. *Photonics Research*, 2021, 9(12): 2511-2518.
- [133] Zhang L X, Li Y Z, Chen B S, et al. Two-dimensional multi-layered SiN-on-SOI optical phased array with wide-scanning and long-distance ranging[J]. *Optics Express*, 2022, 30(4): 5008-5018.
- [134] Yu L, Ma P F, Luo G Z, et al. Adoption of large aperture chirped grating antennas in optical phase array for long distance ranging[J]. *Optics Express*, 2022, 30(15): 28112-28120.
- [135] Lei Y F, Zhang L X, Yu Z Y, et al. Si photonics FMCW LiDAR chip with solid-state beam steering by interleaved coaxial optical phased array[J]. *Micromachines*, 2023, 14(5): 1001.
- [136] Lu L J, Xu W H, Guo Y Y, et al. Large-scale optical phased array based on a multi-layer silicon-nitride-on-silicon photonic platform[C]//2024 Optical Fiber Communications Conference and Exhibition (OFC), March 24–28, 2024, San Diego, CA, USA. New York: IEEE Press, 2024: 1-3.
- [137] Jiang X Y, Zhang Z Y, Huang Q K, et al. Beam-steering based on dispersive optical phased array for FMCW LiDAR application [C]//2024 Optical Fiber Communications Conference and Exhibition (OFC), March 24–28, 2024, San Diego, CA, USA. New York: IEEE Press, 2024: 1-3.
- [138] Chen J Y, Li W L, Kang Z, et al. Single soliton microcomb combined with optical phased array for parallel FMCW LiDAR [J]. *Nature Communications*, 2025, 16: 1056.
- [139] Chen B S, Li Y Z, Na Q X, et al. Demonstrated high-frame-rate real-time 4D imaging of optical phased array solid-state LiDAR system[C]//Optical Fiber Communication Conference (OFC) 2025, March 30–April 3, 2025, San Francisco, California. Washington, DC: Optica Publishing Group, 2025: M2J.1.
- [140] Prost M, Qiu H Q, Croes G, et al. Monolithically integrated silicon photonic biaxial FMCW LiDAR for scalable 3D sensing [C]//Optical Fiber Communication Conference (OFC) Postdeadline Papers 2025, March 30–April 3, 2025, San Francisco, California. Optica Publishing Group, 2025: Th4A.5.
- [141] Wang J, Zhou J Y, Zhu L K, et al. Frequency- and time-domain modeling and characterization of PN phase shifters in all-silicon carrier-depletion modulators[J]. *Journal of Lightwave Technology*, 2020, 38(16): 4462-4469.
- [142] Mishra D, Sonkar R K. Performance assessment of a silicon traveling-wave Mach–Zehnder modulator employing quasi-TM mode propagation[J]. *Optical Engineering*, 2020, 59(12): 125104.
- [143] Misra A, Hosseini R, Dev S, et al. Flexible Nyquist pulse sequence generation from an integrated slow-light silicon modulator for elastic network applications[C]//Photonic Networks; 21th ITG-Symposium, November 24–25, 2020, Online. London: VDE, 2020: 1-5.
- [144] LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision[C]//Proceedings of 2010 IEEE International Symposium on Circuits and Systems, May 30-June 2, 2010, Paris, France. New York: IEEE Press, 2010: 253-256.
- [145] Li Z C, Wang F, Wang N Y. LiDAR R-CNN: an efficient and universal 3D object detector[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 20–25, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 7542-7551.
- [146] Bermejo-Peláez D, Ash S Y, Washko G R, et al. Classification of interstitial lung abnormality patterns with an ensemble of deep convolutional neural networks[J]. *Scientific Reports*, 2020, 10: 338.
- [147] Wolf T, Debut L, Sanh V, et al. Transformers: state-of-the-art natural language processing[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, November 16–20, 2020, Online. Stroudsburg: ACL, 2020: 38-45.
- [148] Shoman H, Jayatilleka H, Park A H K, et al. Compact wavelength- and bandwidth-tunable microring modulator[J]. *Optics Express*, 2019, 27(19): 26661-26675.
- [149] Kitayama K I, Notomi M, Naruse M, et al. Novel frontier of photonics for data processing: Photonic accelerator[J]. *APL Photonics*, 2019, 4(9): 090901.
- [150] 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.
- [151] Miller D A B. Perfect optics with imperfect components[J]. *Optica*, 2015, 2(8): 747-750.
- [152] Shen Y C, Harris N C, Skirlo S, et al. Deep learning with coherent nanophotonic circuits[J]. *Nature Photonics*, 2017, 11(7): 441-446.
- [153] Tait A N, de Lima T F, Zhou E, et al. Neuromorphic photonic networks using silicon photonic weight banks[J]. *Scientific Reports*, 2017, 7: 7430.
- [154] Ashtiani F, Geers A J, Aflatouni F. An on-chip photonic deep neural network for image classification[J]. *Nature*, 2022, 606(7914): 501-506.
- [155] 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.

- [156] 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.
- [157] 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.
- [158] Pai S, Sun Z H, Hughes T W, et al. Experimentally realized in situ backpropagation for deep learning in photonic neural networks[J]. *Science*, 2023, 380(6643): 398-404.
- [159] Bandyopadhyay S, Sludds A, Krastanov S, et al. Single-chip photonic deep neural network with forward-only training[J]. *Nature Photonics*, 2024, 18(12): 1335-1343.
- [160] 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.
- [161] Tait A N, Nahmias M A, Shastri B J, et al. Broadcast and weight: an integrated network for scalable photonic spike processing[J]. *Journal of Lightwave Technology*, 2014, 32(21): 4029-4041.
- [162] Zhang W P, Huang C R, Peng H T, et al. Silicon microring synapses enable photonic deep learning beyond 9-bit precision[J]. *Optica*, 2022, 9(5): 579-584.
- [163] Huang C R, Fujisawa S, de Lima T F, et al. A silicon photonic-electronic neural network for fibre nonlinearity compensation[J]. *Nature Electronics*, 2021, 4(11): 837-844.
- [164] 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.
- [165] Bai B W, Yang Q P, Shu H W, et al. Microcomb-based integrated photonic processing unit[J]. *Nature Communications*, 2023, 14: 66.
- [166] Zhang W P, Lederman J C, Ferreira de Lima T, et al. A system-on-chip microwave photonic processor solves dynamic RF interference in real time with picosecond latency[J]. *Light: Science & Applications*, 2024, 13: 14.
- [167] 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.
- [168] Ahmed S R, Baghdadi R, Bernadskiy M, et al. Universal photonic artificial intelligence acceleration[J]. *Nature*, 2025, 640(8058): 368-374.
- [169] 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.
- [170] Wang Z, Chang L, Wang F F, et al. Integrated photonic metasystem for image classifications at telecommunication wavelength[J]. *Nature Communications*, 2022, 13: 2131.
- [171] Cheng J W, Huang C R, Zhang J L, et al. Multimodal deep learning using on-chip diffractive optics with in situ training capability[J]. *Nature Communications*, 2024, 15: 6189.
- [172] 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.
- [173] Feldmann J, Youngblood N, Karpov M, et al. Parallel convolutional processing using an integrated photonic tensor core[J]. *Nature*, 2021, 589(7840): 52-58.
- [174] 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.
- [175] 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.
- [176] Dong B W, Brückhoff-Plückelmann F, Meyer L, et al. Partial coherence enhances parallelized photonic computing[J]. *Nature*, 2024, 632(8023): 55-62.
- [177] Feng H K, Ge T, Guo X Q, et al. Integrated lithium niobate microwave photonic processing engine[J]. *Nature*, 2024, 627(8002): 80-87.
- [178] Chelladurai D, Kohli M, Winiger J, et al. Barium titanate and lithium niobate permittivity and Pockels coefficients from megahertz to sub-terahertz frequencies[J]. *Nature Materials*, 2025, 24(6): 868-875.
- [179] Chrysostomidis T, Chatzitheocharis D, Eltes F, et al. Ultra-efficient  $\text{Si}_3\text{N}_4$  MZIs with  $\text{BaTiO}_3$  as weight elements for neuromorphic photonics[J]. *Journal of Lightwave Technology*, 2025, 43(9): 4404-4415.
- [180] Huang F, Shen X W, Wang S Y, et al. Toward large-scale photonic chips using low-anisotropy thin-film lithium-tantalate[J]. *Advanced Science*, 2025, 12(9): e2410345.
- [181] Sludds A, Bandyopadhyay S, Chen Z J, et al. Delocalized photonic deep learning on the Internet's edge[J]. *Science*, 2022, 378(6617): 270-276.
- [182] Lederman J C, Zhang W P, de Lima T F, et al. Real-time photonic blind interference cancellation[J]. *Nature Communications*, 2023, 14: 8197.
- [183] Cheng J W, Xie Y Z, Liu Y, et al. Human emotion recognition with a microcomb-enabled integrated optical neural network[J]. *Nanophotonics*, 2023, 12(20): 3883-3894.
- [184] Shastri B J, Tait A N, Ferreira de Lima T, et al. Photonics for artificial intelligence and neuromorphic computing[J]. *Nature Photonics*, 2021, 15(2): 102-114.
- [185] Kikuchi T, Kurokawa M, Fujiwara N, et al. Monolithic integration of various-type III-V epitaxial structures on silicon-photonics platform using chip-on-wafer hydrophilic bonding process[C]//2024 IEEE Silicon Photonics Conference (SiPhotonics), April 15-18, 2024, Tokyo Bay, Japan. New York: IEEE Press, 2024: 1-2.
- [186] Yao W M, Liu X, Matters-Kammerer M K, et al. Towards the integration of InP photonics with silicon electronics: design and technology challenges[J]. *Journal of Lightwave Technology*, 2021, 39(4): 999-1009.
- [187] Ye Z C, Jia H Y, Huang Z J, et al. Foundry manufacturing of tight-confinement, dispersion-engineered, ultralow-loss silicon nitride photonic integrated circuits[J]. *Photonics Research*, 2023, 11(4): 558-568.
- [188] Dao K, Marquez Chavez J, Hayashi K, et al. Low-thermal-budget monolithic integration of optical isolators for silicon photonics[J]. *Proceedings of SPIE*, 2025, 13369: 133690G.
- [189] Yepez P A K, Scholz U, Caspers J N, et al. Novel measures for thermal management of silicon photonic optical phased arrays[J]. *IEEE Photonics Journal*, 2019, 11(4): 6602415.
- [190] Wu S H, Wen S, Xue H Y. Liquid-cooled heat dissipation technology for co-packaged optics over 12.8 Tbps[C]//2024 25th International Conference on Electronic Packaging Technology (ICEPT), August 7-9, 2024, Tianjin, China. New York: IEEE Press, 2024: 1-4.
- [191] Prasad A, Muzio C, Ton P, et al. Advanced 3D packaging of 3.2Tbs optical engine for co-packaged optics (CPO) in hyperscale data center networks[C]//2024 IEEE 74th Electronic Components and Technology Conference (ECTC), May 28-31, 2024, Denver, CO, USA. New York: IEEE Press, 2024: 101-106.
- [192] Billal M R, Blaicher M, Hoose T, et al. Hybrid integration of silicon photonics circuits and InP lasers by photonic wire bonding[J]. *Optica*, 2018, 5(7): 876-883.

# Silicon Photonic Integration and Photonics-Electronics Convergence: Key Enabling Technologies for the Post-Moore Era (Invited)

Zhou Linjie<sup>1,2\*</sup>, Ran Shihuan<sup>1</sup>, Yuan Qiqi<sup>1</sup>, Wu Yue<sup>1</sup>, Lu Liangjun<sup>1,2</sup>, Li Yu<sup>1,2</sup>,  
Guo Yuyao<sup>1,2</sup>, Chen Jianping<sup>1,2</sup>

<sup>1</sup>State Key Laboratory of Photonics and Communications, School of Information and Electronic Engineering,  
Shanghai Jiao Tong University, Shanghai 200240, China;

<sup>2</sup>SJTU-Pinghu Institute of Intelligent Optoelectronics, Pinghu 314200, Zhejiang, China

## Abstract

**Significance** As Moore's law approaches its physical limits, conventional electronic systems encounter fundamental bottlenecks in data transmission bandwidth, computational efficiency, and energy consumption. Silicon photonics emerges as a transformative solution that enables seamless photonics-electronics convergence, fundamentally addressing these limitations that constrain the future of information technology. By synergistically combining the inherent parallelism and high bandwidth capabilities of light with the mature processing infrastructure of silicon electronics, this convergence creates unified platforms that transcend traditional performance boundaries and establish new benchmarks across critical application domains. The strategic significance of silicon photonics lies in its ability to provide integrated photonic-electronic solutions across three foundational pillars of modern information infrastructure: optical communication, optical sensing, and optical computing. This convergence enables each domain to leverage the complementary strengths of both photonic and electronic technologies, creating synergistic capabilities that neither could achieve independently. Through this unified approach, silicon photonic systems deliver exceptional energy efficiencies via hybrid light-electronic processing, positioning photonics-electronics convergence as an essential enabler for artificial intelligence acceleration, edge computing applications, and next-generation technological advancement.

The comprehensive review of silicon photonics presented in this article plays a significant academic role during the ongoing shift from traditional electronic architectures to photonic-electronic integrated systems. The significance of such systematic reviews extends beyond documenting existing knowledge to providing strategic guidance for a field undergoing fundamental technological transformation. The timing of focused research synthesis is particularly relevant as the semiconductor industry and broader technology ecosystem address the approaching limits of Moore's law scaling. At this transition point, the research community and industry benefit from authoritative consolidation of knowledge that has developed across multiple disciplines. Comprehensive reviews help bridge knowledge gaps by integrating insights from materials science, device physics, system integration, and application domains into unified analytical frameworks. Such consolidation supports informed decision-making regarding research priorities and technological roadmaps in the post-Moore era.

Systematic reviews in silicon photonics also establish intellectual coherence within this rapidly evolving, interdisciplinary field. Silicon photonics research spans from fundamental physics to commercial applications, often creating compartmentalized knowledge development. By examining technological evolution across optical communication, sensing, and computing domains, comprehensive reviews facilitate cross-pollination of ideas and identify synergistic opportunities that might otherwise remain unrecognized. This synthetic approach accelerates innovation by enabling researchers to leverage insights from adjacent areas. Such reviews provide valuable intellectual infrastructure for coordinated advancement toward photonic-electronic integration, transforming distributed research efforts into coherent strategic vision supporting the post-Moore era transition.

**Progress** The exponential growth of global internet traffic and artificial intelligence (AI)-driven applications has created an insatiable demand for data transmission capacity, rendering conventional copper-based interconnects inadequate. Silicon photonic integration addresses this critical bottleneck by enabling multi-channel parallelism and dense wavelength division multiplexing (DWDM), achieving single-chip data transmission rates exceeding terabits per second (Fig. 6). Significant advances in modulation technologies have been achieved through innovations in Mach-Zehnder modulators (MZMs) and microring modulators (MRMs), with modulation bandwidths surpassing 110 GHz. These breakthroughs have enabled the development of energy-efficient co-packaged optics (CPO), which deliver substantial power reductions compared to traditional pluggable transceiver modules. This technology has become indispensable for next-generation hyperscale data

centers and exascale high-performance computing infrastructures (Fig. 7). Silicon photonics has also revolutionized optical sensing applications, particularly in lidar systems for autonomous vehicles and robotics (Figs. 14 and 15). The technology enables unprecedented integration density and performance capabilities through frequency-modulated continuous-wave (FMCW) lidar systems integrated with optical phased arrays (OPAs), demonstrating high-resolution beam steering and long-range detection capabilities. Remarkable scaling progress has been achieved from early proof-of-concept systems to large-scale optical phased array implementations, while monolithic integration breakthroughs have enabled solid-state FMCW lidar systems with real-time multi-dimensional imaging capabilities and enhanced ranging performance. These advances have culminated in the demonstration of parallel processing capabilities through advanced microcomb integration and sophisticated opto-electronic co-packaging technologies, positioning silicon photonic lidar as a commercially viable solution for autonomous vehicles and advanced sensing applications. Photonic-electronic synergistic architectures represent a paradigm shift beyond the traditional von Neumann bottleneck. Photonic neural networks and matrix accelerators exploit the inherent parallelism and high bandwidth of light, achieving energy efficiencies several orders of magnitude higher than conventional electronic systems. Advanced architectures including Mach-Zehnder interferometer (MZI) meshes, microring weight banks, intensity modulation arrays, and metasurface diffractive optical networks have demonstrated energy efficiency exceeding 100 TOPS/W (TOPS: tera operations per second). These systems enable real-time inference on complex AI models such as ResNet and bidirectional encoder representations from Transformers (BERT) (Figs. 17 and 18). The integration of non-volatile phase-change materials and plasmonic modulators further enhances computational speed and efficiency, positioning silicon photonics as a disruptive technology in high-performance computing and edge AI applications.

Despite substantial technological advances, achieving industrial-scale viability in silicon photonics requires addressing fundamental scientific barriers to large-scale integration. The foremost challenge stems from inherent material compatibility issues, where significant lattice mismatches and thermal expansion coefficient differentials between dissimilar materials generate interfacial defects and thermomechanical stress, critically degrading optical performance parameters and compromising long-term reliability. Equally critical is achieving nanoscale precision control and high manufacturing yield across massively integrated photonic-electronic arrays, where increasing integration densities exponentially amplify the performance impact of process variations. System-level packaging and thermal management present additional complexities arising from the contrasting requirements of temperature-sensitive photonic components and power-dissipating electronic circuits.

**Conclusions and Prospects** Transformative pathways forward include three-dimensional heterogeneous integration utilizing through-silicon vias and micro-bump technologies, transcending planar density limitations through vertical stacking and functional layer co-optimization. The emergence of chiplet-based modular design paradigms enables standardized photonic input/output cores, enhancing system configurability while reducing development cycles. Most significantly, the deep convergence of artificial intelligence with photonics will unlock transformative capabilities through deep learning-enabled inverse design, enabling discovery of novel micro/nano-photonic structures beyond empirical intuition, while AI-driven process optimization and autonomous device tuning will dramatically improve manufacturing yield and operational reliability. These advances will catalyze profound transformations across the information technology landscape. Silicon photonics will underpin terabit-per-second optical backbones essential for 6G networks, enable miniaturized high-sensitivity sensors for autonomous mobility and smart infrastructure, and drive photonic-electronic convergence toward distributed computing paradigms that harness the inherent parallelism of light for unprecedented computational efficiency in exascale artificial intelligence applications.

As the pivotal enabler for the post-Moore era, silicon photonics represents both an evolutionary imperative and a defining opportunity to augment human cognitive capabilities. Through sustained cross-disciplinary innovation, this technology will establish the indispensable technological foundation for an intelligently interconnected future, marking a transformative chapter in information technology evolution.

**Key words** integrated optics; photonics-electronics convergence; optical communication; optical sensing; optical computing