

论后摩尔时代的新型集成电路技术 ——光神经网络的挑战和机遇

雷逸鸣¹

(北京大学物理学院, 北京 100871)

摘要: 人工智能的快速发展使其在各个领域被广泛应用, 但由于大数据时代的到来, 传统芯片的运算能力难以达到其算法模型的需求, 于是人们逐渐向光学领域寻找问题的答案。光计算通过利用光的独特性质, 为解决这一瓶颈提供了一个新的视角。我们预计, 大规模集成光学处理器的时代将很快到来, 并以混合光电框架的形式用于实际的人工智能软件中。

关键词: 摩尔时代; 光学芯片; 神经网络; 人工智能; 集成电路

On the new integrated circuit technology in the post-Moore era ——Challenges and opportunities of optical neural networks

Lei Yiming¹

(School of Physics, Peking University, Beijing 100871, China)

Abstract: The rapid development of artificial intelligence facilitates various applications from all areas but also poses great challenges in its hardware implementation in terms of speed and energy because of the explosive growth of data. People gradually turn to Optical areas for further answers. Because of its unique quality, Optical computing provides a distinctive perspective to address this bottleneck. We anticipate that the era of large-scale integrated photonics processors will soon arrive for practical AI applications in the form of hybrid optoelectronic frameworks.

Keywords: Moore's Era; Optical Chip; Neuro Network; Artificial Intelligence; Integrated Circuits

摩尔定律并不是一个被客观证明的定律, 但几十年来, 他给予人们的信心使集成电路业不断繁荣发展^[1]。我们将在总结前人工作的基础上, 阐明芯片未来可能的发展方向, 并结合近期光神经网络热点介绍其面临的挑战和未来发展机遇。

1 集成电路现状概述

在近 60 年里, 人们通过缩小集成电路器件面积的方法, 在相同尺寸的晶圆上制造出了拥有更多电路的芯片, 达到了降低成本和提升性能的目的。现在, 我们已经进入到 5nm 技术节点, 在向着 3nm 节点迈进^[2,3]。

在这些年的不断发展过程中, 人们在金属氧化物半导体场效应晶体管 (MOSFET) 的基础之上, 通过发明 FinFET, 论证了实现 20nm 以下节点硅基集成电路的可行性^[4]。自 FinFET 开始, 人们的目光转向了集成电路在物理上的三个维度。由此, 人们已经发展出了器件级三维集成、电路级三维集成、芯片级三维集成等方向^[5]。

但是, 集成电路业终将面临器件尺寸无法继续缩小的问题^[6]。如今, 集成电路业正从缩小尺寸提高性能的方式向着增大芯片面积来提高集成度的方式逐步转型, 但这同时也带来了如何提高良率的问题^[7]。因此, 集成电路业必然要从传统的 CMOS 大规模集成的形态向其他的技术形态转变, 人们需要寻找新的发展路径、应用方式、使用场景, 这也意味着我们将要进入“后摩尔时代”^[8]。

2 人工智能计算机

传统的计算机一般是基于逻辑电路, 根据人们事先编写好的代码运行特定程序的机器^[9]。随着传统计算机性能的提高, 一套全新的基于神经网络的人工智能算法进入了大众视野。这一算法不断地发展, 其对算力的需求也不断提高, 计算机的算力逐渐成为制约其发展的主要因素。为了满足人工智能算法对算力的需求, 人们开始着手寻找一种能适应人工智能算法的新型计算机。

2.1 基于传统计算机的尝试

人们最先想到的便是基于传统计算机的冯诺依曼体系进行改良。而人工智能算法（神经网络）的特点是大量的并行简单运算和参数的存储与调用，因此改良的核心在于如何在并行计算的基础上尽可能地减少数据存储调用耗费的时间。于是，人们发明了存算一体（存内计算）技术（例如：RRAM、PCM、MRAM、FeRAM）。在过去的 20 年中，存内数字计算一直专注于识别有更低能耗和面积消耗的新型逻辑门概念。使用纳米磁体进行数字运算、量子点等概念已经通过实验证实，二进制电阻开关器件也在数字计算中表现出了多种优势（例如通过互连线路直接接入电路、通过电信号重新配置电路的能力、纳米级的小型化等）^[10]。

与此同时，存内计算也面临着一些挑战。例如：由内存不稳定性带来的计算准确性问题。由于其无定形态的亚稳态性质，电阻会随着时间的推移而漂移^[10]。

种种这些挑战依然等待着人们一一破解。

2.2 基于光的波动性的光神经网络

光的干涉行为的空间构型与神经网络的逻辑构型有很大的相似度。光的波动性给予了我们在其传播过程中进行数据的运算从而提升网络计算性能的可能性。为了能够用光学信号完整的实现神经网络的功能，我们需要在信息调用、运算和输出等方面构造与神经网络模型对应的光路。

3 光神经网络

3.1 信息在光路中的存储

光携带的信息包括光的相位信息和振幅信息。因此，我们可以选择将信息存储在光的相位中或振幅中。对于前者，我们可以使用分振幅反向设计光路，使得在光经过后，所分出的两路光振幅之比恰好与我们希望存储的信息一致；而对于后者，我们则可以通过控制温度，改变不同光路介质的折射率，间接达到调控光的相位的目的。

3.2 信息在光路中的运算

信息在光路中的运算主要通过不同光路的光之间进行相互作用完成，达到这一目的的方法有很多种，例如：

通过在两路光波导间添置耦合环，使两者发生相互作用^[12-14]；通过在两层神经云中添置开放光场，直接让多束光发生干涉^[15]；在光路中加入光电

转化器，通过电学方法在运算中加入激活函数，并达到增强光信号的目的^[16,17]；使用电磁感应透明度（EIT）实现用纯光学元件实现激活函数^[18]（EIT 是一种在介质中发生的相干光学非线性现象，其中一束光的透明度被另一束光控制）；通过反向设计使光在通过元件后光强分布达到预先训练模型的标准^[19]；通过柱面透镜实现光路中的张量运算^[20]。

3.3 信息在光路中的输出

信息在光路中的输出主要通过光电转换器将光信号转化为电信号。

对于将信息存储在振幅中的情况，可以直接将光波聚焦在接收屏上，读取光的强度；对于将信息存储在相位中的情况，我们可以通过在光进入光路前分离半个振幅的光强，并将其投射到接收器处作为背景光强获得光的绝对相位。

4 结 论

在后摩尔时代，光神经网络已在不同的人工智能模型中得到广泛应用，但他仍面临各种挑战，其性能明显优于传统电子处理器的实际应用仍未得到证明。例如，如何改进材料性质及工艺使纯光学激活函数能够集成在光学芯片上；如何在低功耗芯片上实现高速大规模可重构性；如何设置光路结构使得可以在光学芯片上实现反向传播，而不仅仅将以训练好的模型刻蚀在芯片实体上实现。^[21,22]

同时，光神经网络也是一个富有发展潜力的领域，他是一个有望超越冯诺依曼体系实现与人工智能一对一适配的技术。他对于神经元的模拟不仅仅是数值上的，他可以为神经网络提供基于光信号的模拟平台，实现快速、高效的信息处理与传递，满足人工智能对算力的迫切需求。

参考文献：

- [1] Schaller R R. Moore's law: Past, present, and future[J]. IEEE Spectrum, 1997, 34(6): 52-59
- [2] Auth C. 45 nm high-k + metal gate strain-enhanced CMOS transistors[C]//IEEE 2008 Custom Integrated Circuits Conference (CICC). Piscataway: IEEE Press, 2008: 379-386.
- [3] Yadav N, Jadav S, Saini G. Geometrical variability impact on the performance of sub-3 nm gate-all-around stacked nanosheet FET[J]. Silicon, 2022, 14, 10681-10693, Early Access.

- [4] Huang X, Lee W C, Kuo C, et al. Sub 50-nm FinFET: PMOS[C]//1999 IEEE International Devices Meeting (IEDM). Piscataway: IEEE Press, 1999: 67-70
- [5] DONG J, ZHANG X. Prospects of Advanced Integrated Circuit Technologies in Post-Moore Era[J]. *Science and Technology Foresight*, 2022, 1(3): 42-51.
- [6] Williams R S. What's next?[J]. *Computing in Science & Engineering*, 2017, 19(2): 7-13.
- [7] 姜晓鸿, 郝跃. 集成电路局部缺陷及其相关的功能成品率和电迁徙问题的研究[D]. 西安: 西安电子科技大学, 1998.
- [8] 黄如. 后摩尔时代集成电路技术发展探讨[R]. 北京: 中国科学院, 2021.
- [9] Bhavin J. Shastri, Alexander N. Tait, T. Ferreira de Lima, Wolfram H. P. Pernice, Harish Bhaskaran, C. D. Wright & Paul R. Prucnal. Photonics for artificial intelligence and neuromorphic computing[J]. *Nature Photonics*, 2021, 15, 102–114
- [10] Daniele Ielmini & H.-S. Philip Wong. In-memory computing with resistive switching devices[J], *Nature Electronics*, 2018, 1, 333-343
- [11] Liao K, Dai T, Yan Q, Hu X, and Gong Q. Integrated Photonic Neural Networks: Opportunities and Challenges[J]. *ACS Photonics*, 2023, 10, 7, 2001-2010
- [12] Lu, L., Zhao, S., Zhou, L., Li, D., Li, Z., Wang, M., Li, X., Chen, J. 16×16 non-blocking silicon optical switch based on electro-optic mach-zehnder interferometers. *Opt. Express* 2016, 24 (9), 9295–9307.
- [13] Zhou, H., Zhao, Y., Wang, X., Gao, D., Dong, J., Zhang, X. Self-configuring and reconfigurable silicon photonic signal processor. *ACS Photonics* 2020, 7 (3), 792–799.
- [14] Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M, et al. All-optical machine learning using diffractive deep neural networks. *Science* 2018; 361(6406):1004–8.
- [15] Fu, T., Zang, Y., Huang, H., Du, Z., Hu, C., Chen, M., Yang, S., Chen, H. On-chip photonic diffractive optical neural network based on a spatial domain electromagnetic propagation model. *Opt. Express* 2021, 29 (20), 31924–31940.
- [16] Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M, et al. Deep learning with coherent nanophotonic circuits. *Nat Photonics* 2017, 11(7): 441–6.
- [17] George JK, Mehrabian A, Amin R, Meng J, De Lima TF, Tait AN, et al. Neuromorphic photonics with electro-absorption modulators. *Opt Express* 2019, 27(4): 5181–91.
- [18] Zuo Y, Li B, Zhao Y, Jiang Y, Chen YC, Chen P, et al. All-optical neural network with nonlinear activation functions. *Optica* 2019, 6(9): 1132–7
- [19] Khoram E, Chen A, Liu D, Ying L, Wang Q, Yuan M, et al. Nanophotonic media for artificial neural inference. *Photonics Res* 2019, 7(8): 823–7.
- [20] Caulfield HJ, Dolev S. Why future supercomputing requires optics. *Nature Photonics* 2010, 4(5):261–3.
- [21] De Marinis L, Cococcioni M, Castoldi P, Andriolli N. Photonic neural networks: a survey. *IEEE Access* 2019; 7: 175827–41.
- [22] Wu, J., Lin, X., Guo, Y., Liu, J., Fang, L., Jiao, S, Dai, Q. Analog Optical Computing for Artificial Intelligence[J]. *Engineering* 2022, 10: 133-145