采用多态电光电路与架构实现高速、节能的非二进制计算

Ishan Thakkar 电子与计算机工程系 肯塔基大学 美国肯塔基州列克星敦
igthakkar@uky.edu

Sairam Sri Vatsavai 电子与计算机工程系 肯塔基大学 美国肯塔基州列克星敦
ssr226@uky.edu

Venkata Sai Praneeth Karempudi 电子与计算机工程系 肯塔基大学 美国肯塔基州列克星敦
kvspraneeth@uky.edu

arXiv:2304.07608v1 [cs.AR] 15 Apr 2023

摘要
本文介绍基于微环谐振器（MRR）的多态电光电路与架构，可用于高速、节能的非二进制可重构计算。我们的多态电光电路可动态编程，在不同时间实现不同的逻辑和算术函数。它们能够实现紧凑化与多态性，从而改进操作数处理、减少空闲时间，并提高面积和静态功耗开销的摊销效率。结合具有固有能力、可在原位累积大量光脉冲的灵活光电探测器后，我们的电路可支持对随机/一元格式和高维储备池格式等非二进制格式数据进行节能处理。此外，我们的多态电光电路可构建可配置的电光计算加速器架构，用于处理二值化和整数定点量化卷积神经网络（CNNs）。我们从面积、延迟和能耗方面，将所设计的多态电光电路与架构同以往研究中的若干电路与架构进行了比较。

CCS 概念
• **计算机系统组织 → 神经网络**；**可重构计算**；**光计算**。

关键词
电光多态电路、非二进制计算、微环谐振器（MRRs）

ACM 参考格式：
Ishan Thakkar、Sairam Sri Vatsavai 和 Venkata Sai Praneeth Karempudi。2023年。利用多态电光电路与架构实现高速节能非二进制计算。载于《*2023*年五大湖超大规模集成电路研讨会论文集》（*GLSVLSI '23*），*2023*年*6*月*5*日至*7*日，美国田纳西州诺克斯维尔。ACM，美国纽约州纽约，6页。https://doi.org/10.1145/3583781. 3590258

允许免费制作本作品全部或部分内容的数字副本或纸质副本，用于个人或课堂用途，但不得为营利或商业利益而制作或分发副本，且副本首页必须附有本声明及完整引文。对于本作品中归他人而非作者所有的部分，必须尊重其版权。注明出处后，允许摘录。若要以其他方式复制、转载、发布到服务器或重新分发至列表，则需事先获得明确许可和/或支付费用。请向 permissions@acm.org 申请许可。*GLSVLSI '23*，*2023*年*6*月*5*日至*7*日，美国田纳西州诺克斯维尔© 2023 版权归所有者/作者所有。出版权已授权给 ACM。ACM ISBN 979-8-4007-0125-2/23/06...$15.00
https://doi.org/10.1145/3583781.3590258

1 引言
近年来，由于晶体管尺寸极小，纳米制造技术正面临严重限制，摩尔定律也因此遭遇严峻挑战 [1]。幸运的是，随着摩尔定律日渐式微，集成电光（E-O）计算电路展现出革命性潜力，有望为计算提供速度更快、效率更高的硬件。此前的研究已展示了用于计算的电光电路（例如 [15, 19, 21, 22, 33–35]），它们通常用于实现以下四类逻辑和算术功能：**（I）基本逻辑门功能** [21, 22, 35]，使用两个二进制输入操作数来加速神经网络。**（II）任意组合逻辑功能** [19, 34]，可作为现场可编程门阵列（FPGA）的直接光学替代方案。**（III）双操作数算术函数** [15, 33]，用于累加，并支持定制精度和全精度算术运算。**（IV）多操作数线性算术函数** [7, 17, 21, 35]，用于实现乘加（MAC）和向量点积（VDP）运算，以处理深度学习工作负载。然而，正如 [16], 所详述，这些电光电路存在以下不足：（*i*）空闲时间长，导致面积开销和静态功耗开销高且无法摊销；（*ii*）波长并行性与可实现位精度之间存在明显的权衡。

为缓解这些不足，本文的贡献主要有两方面：**（i）**提出一种多态电光电路（PEOC），以及**（ii）**设计一种可配置电光计算加速器（CEONA）。我们展示了 PEOC 可通过重新配置，在不同时间实现不同的算术和逻辑函数。我们在 CEONA 中以波分复用（WDM）的方式采用此类 PEOC，从而灵活支持对具有不同位精度的卷积神经网络（CNNs）进行加速。此外，CEONA 还能够加速基于延迟反馈储备池计算（DFRC）的应用。我们将设计的 PEOC 和 CEONA 与先前工作中的若干电路和架构进行了比较，并展示了它们在面积、延迟和能耗方面的优势。为帮助读者在深入阅读本文前了解相关基础知识，我们建议读者先阅读有关微环谐振器（MRRs）[8] 和光计算架构 [7, 20]的教程。

2 多态电光电路
图1(a)展示了我们的多态电光（E-O）电路（PEOC）的结构。该电路由一个有源微环谐振器组成，既可用作微环调制器（MRM），也可用作多态电光逻辑门（MRR-PEOLG）[16]。将有源微环谐振器用作微环调制器，可实现

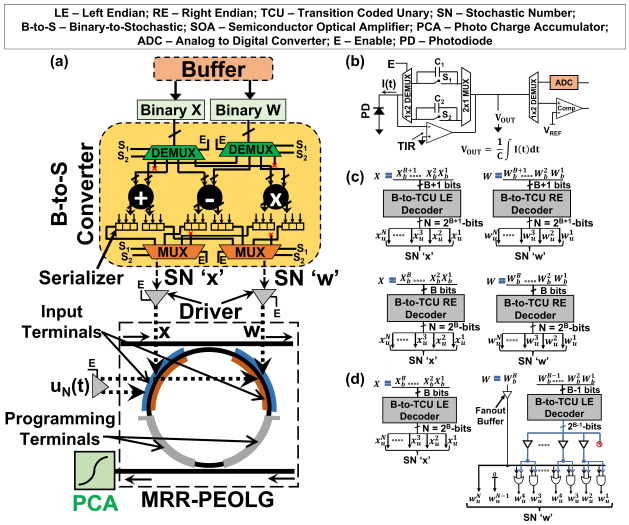


图1：(a) 我们的基于微环谐振器的多态电光电路（PEOC）的结构，(b) 光电荷累加器（PCA）。用于加法（(c)-上）、减法（(c)-下）和乘法（d）运算的二进制到随机数转换电路。

将输入的电时序信号 $u_N(t)$ 调制到输出光信号 [32] 上。此外，当有源微环谐振器用作 MRR-PEOLG 时，可以在不同时间重新配置，以实现不同的逻辑函数 [16]。与、或、异或、与非、或非和同或等所有逻辑函数均已得到演示 [16]。此外，在图1中，还将二进制到随机（B-to-S）转换电路和光电荷累加器（PCA）与 MRR-PEOLG 集成，从而将 PEOC 转变为多态二进制算术单元（PBAU）。PBAU 利用 MRR-PEOLG 的或、与和异或函数，实现基于随机计算（SC）的近似加法（ADD）、乘法（MUL）和减法（SUB）运算 [3]。关于我们的 MRR-PEOLG、PCA 和 PBAU 的结构与运行机制，后续小节将提供更多详细信息。

表1：电光电路的性能比较。

指标	同或计数		位串行乘法器	
	[35]	MRR-PEOLG	[22]	MRR-PEOLG
A (mm^2)	0.013	0.011 (1.16×)	0.023	0.011 (2.08×)
E（nJ）	0.05	0.032 (1.53×)	0.327	0.033 (9.89×)
L（ns）	0.02	0.025 (0.8×)	0.1	0.025 (4×)
A*E*L	1.3e-5	0.9e-5 (1.44×)	75.2e-5	0.91e-5 (82.6×)

2.1 基于微环谐振器的多态电光逻辑门

我们提出的 **MRR-PEOLG** 处于编程状态 [16]。实现特定的逻辑门函数，通过向 MRR-PEOLG 的编程端子施加电压，将 MRR 与操作数无关的谐振位置 'κ'（图2中的洋红色通带）调整到相对于输入波长 'λ_{in}' 和 MRR 初始谐振位置 'η' 的特定光谱位置（见图1(a)中的端子）。随后，将电输入操作数施加到基于 PN结的 MRR 输入端子（图1和图2中的(x,w)）。进行此操作后 MRR 的谐振会根据所施加输入操作数的组合向较短波长方向偏移。将输入操作数位施加到输入端子后，我们的 MRR-PEOLG 在下载端口和直通端口的光学响应便会遵循其所编程实现的逻辑门函数真值表。借此，我们的 MRR-PEOLG 可以在不同时间执行不同的逻辑函数（图2）。为验证 MRR-PEOLG 的这一多态功能，我们还使用 Ansys/Lumerical 套件中的 INTERCONNECT 工具进行了时域（瞬态）分析 [4]。为此，我们向 MRR-PEOLG 的输入端子施加了两个电脉冲（图3(a)和图3(b)），并收集了对应不同逻辑函数的输出脉冲模式：来自 MRR-PEOLG 下载端口的输出脉冲模式（图3(c)、3(e)和3(g)）以及来自直通端口的输出脉冲模式（图3(d)、3(f)和3(h)）。显然，输出信号遵循相应逻辑函数的逐脉冲真值表，这证明了我们的 MRR-PEOLG 实现不同逻辑函数的能力。此外，我们还评估了采用 MRR-PEOLG 对先前研究中的两个电光电路的面积、延迟和能耗的改进效果 [35] 和 [22]。结果见表1。

2.2 光电荷累加器

Fig. 1(b) 展示了我们的光电荷累加器（PCA）。其设计综合借鉴了 [24] 中的时间积分接收器（TIR）和 [9] 中基于光电探测器（PD）的光脉冲累加器。因此，我们的 PCA 采用一个 PD、一个 TIR 和两个电容器。根据预期用例，模数转换器或比较器与 TIR 相连。PD 具备相干与非相干双重叠加及光电荷累积的能力 [9]。如果 PD 的带宽倒数为t=（1/符号速率），那么 PD 在每个t时间间隔内产生的自由载流子数量与吸收的光子数量成正比，因此，该时间间隔内的输出光电流与此期间入射到 PD 上的所有相干及非相干光脉冲的平均光功率之和成正比 [9]。我们的 PCA 中的 TIR 收集该输出光电流，并在其中一个电容器上产生相应的电压。该输出光电流以及电容器上累积的电压，给出了t [24]时间间隔内入射到 PD 上的所有光脉冲的累积结果。TIR 电路允许在 PCA 电路饱和之前，连续累积总计γ 个t时间间隔。γ称为 PCA 的累积容量。在经过γ个时间间隔的光脉冲累积后，需要对活动电容器（例如 C1）放电，以便为下一次累积做好准备。电容器 C1 放电期间，电容器 C2 通过允许继续进行另一次并行累积，来减轻放电延迟。表2报告了我们的 PCA 在 [30] 中不同符号率下的累积容量γ。在 50 千兆符号/秒下，我们的 PCA 具有γ=8503 （表2），高于大多数现代卷积神经网络中每个神经元所需的累积次数。我们的 PCA 如此大的γ，无需将每个输出神经元所需的累积操作分解为多个部分和 [30]。

2.3 多态二进制算术单元

根据第 2 节和图1，将我们的 MRR-PEOLG 与二进制到符号转换电路及 PCA 集成，可将 PEOC 转变为

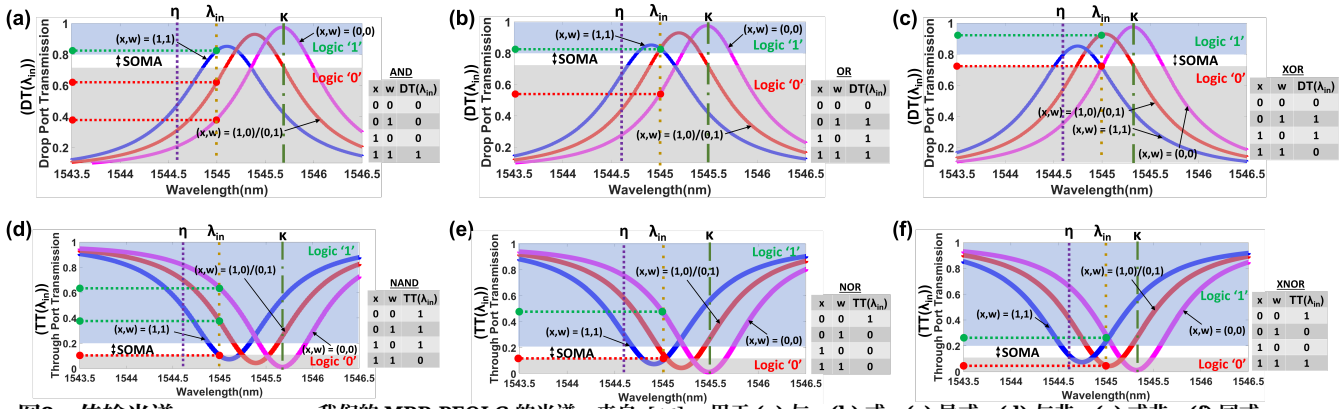
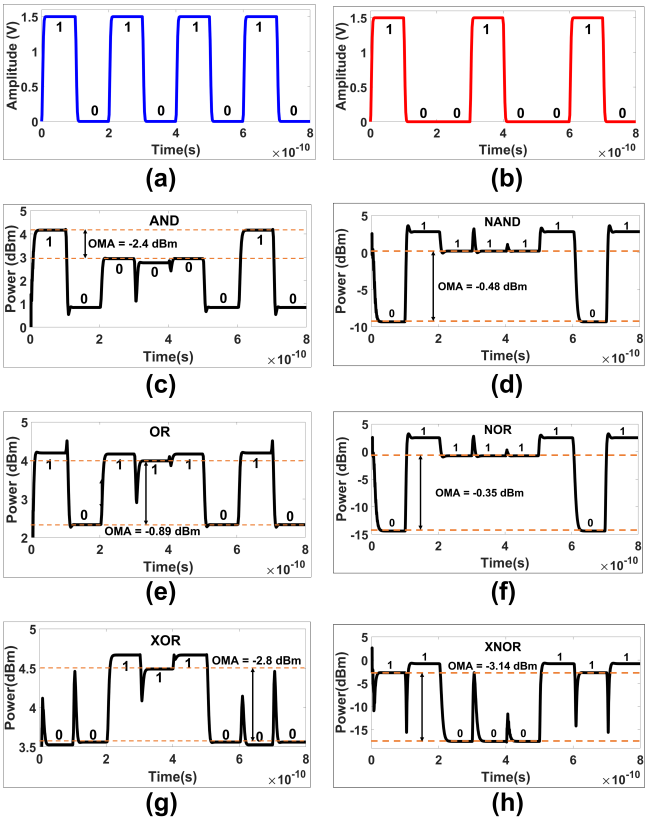


图2：传输光谱 我们的 MRR-PEOLG 的光谱，来自 [16]，用于 (a) 与、(b) 或、(c) 异或、(d) 与非、(e) 或非、(f) 同或 .



我们的 MRR-PEOLG 的瞬态分析结果，来自 [16]。(a)、(b) 输入电脉冲。(c) 与、(d) 与非、(e) 或、(f) 或非、(g) 异或、(h) 同或的输出光脉冲。

表2：不同符号速率 (SR) (千兆符号/秒) 下的 PCA 累积容量 (γ) [30]。

SR	3	5	10	20	30	40	50
γ	39682	29761	19841	14880	10822	9920	8503

多态二进制算术单元（PBAU）。我们的 PBAU 结构可分为三个阶段。首先，B-to-S 外围电路有助于将输入的二进制操作数（N 位）转换为随机比特流（2^N 位）。简而言之，B-to-S 转换通过位并行二进制到跃迁编码一元（B-to-TCU）解码器 [26] 实现。随后，使用高速串行器将 B-to-TCU 解码器的位并行输出转换为比特流。针对加法、乘法和减法函数定制设计了 B-to-TCU 解码器，以确保输入的随机比特流（“x”和“w”）具有适当的相关性，从而尽可能减少结果中的误差 [3]。为使随机比特流之间具有适当的相关性，我们借鉴了 [3]，为特定函数的 B-to-S 电路（及其组成部分 B-to-TCU 解码器）设置适当的字节序和比特流长度。例如，对于加法函数，比特流 x 和 w 的字节序相反；而对于减法和乘法函数，它们的字节序相同（均为右字节序）（图 1(c) 和图 1(d)）。此外，对于乘法和减法函数，B-to-S 转换电路生成的随机比特流长度为 2^N 位；对于加法函数，长度则为 2^{N+1} 位 [3]。此外，我们用于乘法函数的 B-to-S 转换电路（图 1(d)）通过确保条件概率 P(w/x) 等于边缘概率 P(x)，最大限度地减少结果中的相关性误差 [26]。

由 B-to-S 转换阶段生成的随机比特流作为输入送至第二阶段，该阶段由我们的 MRR-PEOLG [16] 组成。如第 2.1 节所述，我们的 MRR-PEOLG 实现与、或和异或逻辑函数，并以逐位方式将这些函数应用于随机比特流，分别实现目标乘法、加法和减法函数。PBAU 的第三阶段由我们的 PCA 组成，用于将 MRR-PEOLG 输出的随机比特流转换为二进制格式。

表3：我们的 PBAU 执行加法、减法和乘法的性能。MAE=平均绝对误差。

位精度	6位			8位		
	加法	减法	乘法	加法	减法	乘法
延迟（纳秒）	5.32	2.74	2.76	20.51	10.27	10.29
能耗（皮焦耳）	16.1	6.8	10.2	60.1	23.6	36.2
MAE	0	0	0.03	0	0	0.04

在表3中，我们评估了 PBAU 在 6位和8位二进制（整数）操作数精度下执行乘法、减法和加法函数时，每次操作的延迟、能耗及平均绝对误差（MAE）值。显然，PBAU 执行减法

表4：PBAU与先前工作的电光算术电路在延迟、能耗和面积方面的比较。

	Area	能耗	面积*延迟
8位PBAU	0.0012mm^2	36.2pJ	3.312mm^2.ps
8位PoNALU [15]	0.6 mm^2	31.25nJ	201.3mm^2.ps
8位 EPALU [33]	1.4 mm^2	37.5nJ	523.5 mm^2.ps
8位 PIXEL [21]	0.00359 mm^2	51.2pJ	36.9mm^2.ps

和加法函数时不会产生误差，乘法函数的 MAE 值也低得可以忽略不计。同样，表4对比了我们的 PBAU 与先前研究中的 E-O算术电路在延迟、能耗和面积方面的表现。从表4可以看出，我们的 PBAU 能耗显著更低，占用面积也更小。

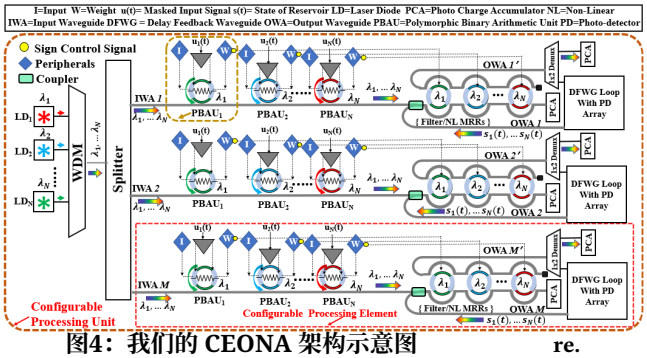


图4：我们的 CEONA 架构示意图 re.

3 可配置电光计算加速器

3.1 概述

多个 PBAU 以波分复用（WDM）方式排列成阵列，构成我们可配置电光计算加速器（CEONA）架构的主处理单元。该单元称为可配置处理单元（CoPU），如图 4 所示。一个 CoPU 由梳状激光源组成，[14] 该光源在N个不同波长（即从 λ_1 到 λ_N）上发射光功率。每个波长上的光功率被分到M条输入波导（IWA）中，每条输入波导都连接到一个可配置处理元件（CoPE）。每个 CoPE 包含N个以 WDM 方式排列的 PB AU 阵列，旁边设有一组微环谐振器，这些微环谐振器可用作滤波器或非线性有源微环谐振器（图 4）。每组微环谐振器连接至 PCA 或延迟反馈环路波导（DFWG）。根据各 PBAU 和微环谐振器组的配置，以及 CoPE 与 PCA 或 DFWG 的连接方式，我们的 CEONA 加速器可用于两种应用场景。

3.2 案例 I：神经网络加速器

CEONA 可配置为执行二值神经网络（BNN）的推理，以及整数定点量化卷积神经网络的推理。在推理过程中，CEONA 接收权重和输入操作数：对于二值神经网络，它们为 1 位（二进制）；对于整数定点量化卷积神经网络，它们为 8 位。

采用二值化操作数的 CEONA：采用二值化操作数的 CEONA 称为 CEONA-B。二值神经网络的推理需要进行同或-位计数运算 [30]。因此，对于二值化操作数，CEONA 会将每个 CoPE 的 PBAU 动态配置为如第 2 节所述的同或门。PBAU 对二值化的 I 和 W 执行同或运算（图 4）。同或门的光学输出被发送至底部的 OWA（图 4）。微环谐振器组关闭，使所有同或输出位都能到达底部的 PCA（位于 OWA 上）。PCA 对来自 PBAU 的输入光学位进行计数，以执行位计数运算并生成最终累加结果。因此，如 [30]所详述，每个 CoPE 无需部分和存储或归约即可生成二值神经网络输出张量中的一个值。我们采用的评估配置见 [30]。

图 5(a) 和图 5(b) 比较了 CEONA-B 与先前加速器在多种二值神经网络上实现的 FPS（每秒帧数，即吞吐量）和 FPS/W（能效）值。总体而言，CEONA-B_5 （符号率为=5 千兆符号/秒）和 CEONA-B_50 （符号率为=50 千兆符号/秒）的吞吐量和能效均优于其他加速器。在这些二值神经网络上的几何平均值方面，我们的 CEONA-B_50 比 ROBIN_EO、ROBIN_PO 和 LIGHTBULB [35], 的 FPS 分别高出 52×、 7×和 7×。在这些二值神经网络上的几何平均值方面，我们的 CEONA-B_5的 FPS/W 分别比 ROBIN_EO、ROBIN_PO 和 LIGHTBULB 高 2.6×、3.3×和 1.7× 。从 [30], 起，CEONA-B 通过增大N来提升吞吐量，从而提高并行度。能耗优势主要来自无需存储和归约部分和，因为 PCA 本身具备原位累加能力。

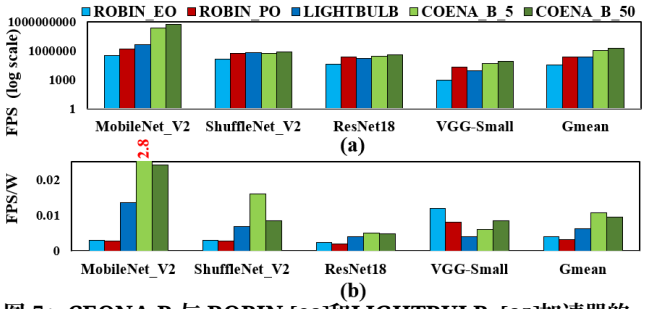


图 5：CEONA-B 与 ROBIN [28]和LIGHTBULB [35]加速器的（a）每秒帧数（对数刻度），（b）每瓦每秒帧数。

带整数操作数的 CEONA：带整数操作数的 CEONA 称为 CEONA-I。卷积神经网络的推理需要进行点积运算 [27]。如第 2 节所述，在随机计算中，乘法可以通过与门实现。因此，对于卷积神经网络的推理，CEONA-I 将 PBAU 配置为与门，以对 I 和 W 进行逐点乘法 [31]（图 4）。微环谐振器组作为滤波器组运行，相应 PBAU 发出的符号控制信号则开启或关闭滤波器，从而使 PCA 能够进行有符号累加。 [31]中对 CEONA-I 架构及其所采用的评估配置进行了全面说明。

图 6(a)、6(b) 和 6(c) 比较了 CEONA-I 与此前加速器在各种卷积神经网络上的 FPS（吞吐量）、FPS/W（能效）和 FPS/W/mm^2（面积效率）。从图 6(a) 可见，在所有卷积神经网络的几何平均值方面，CEONA-I 的表现显著优于模拟光学加速器 MAW（HOLYLIGHT）[17] 和 AMW（DEAPCNN） [7] ，分别高出 66.5×和 146.4×。从图 6(b) 可见，在所有卷积神经网络的几何平均值方面，CEONA-I 的 FPS/W 分别比模拟MAW（HOLYLIGHT）和 AMW（DEAPCNN）高出 90× 和 183× 。从图 6(c) 可见，CEONA-I 的表现高出 91× 和

184× FPS/W/mm^2，分别优于模拟MAW（HOLYLIGHT）和AMW（DEAPCNN），在所有卷积神经网络上的几何平均值方面如此。吞吐量优势主要归因于 CEONA-I 相较于模拟光学加速器更出色的N [31]。此外，使用 PCA 消除了存储和累加部分和的需要，从而带来吞吐量、能耗和面积方面的优势。总体而言，与此前的模拟加速器相比，CEONA-I 显著提升了吞吐量、能效和面积效率。

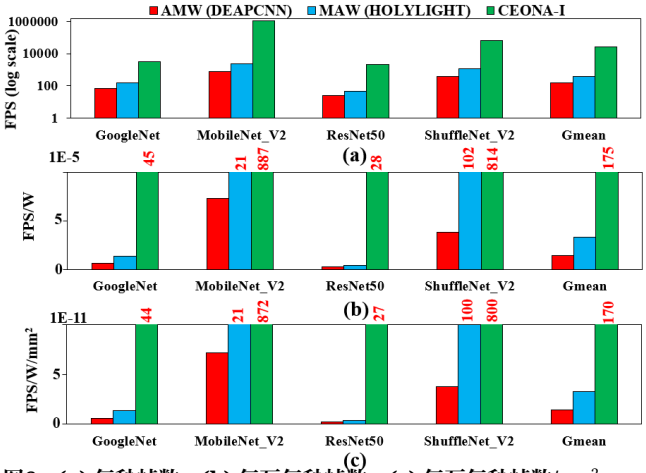


图6：(a) 每秒帧数，(b) 每瓦每秒帧数，(c) 每瓦每秒帧数/mm^2，比较 CEONA-I 与 MAW 和 AMW 加速器在 8位整数精度下的表现。

CEONA-I 的可扩展性：为确定我们的 CEONA-I CoPU 在不同整数精度级别 (B) 下可实现的规模 N，我们采用 [2] 和 [27]中的可扩展性分析方程（方程 1、方程 2 和方程 3）。这些方程所用参数的定义及其取值列于 [31]。在方程 1 中，对于 AMW 和 MAW 架构，数据速率 (DR) 等于符号速率 (SR)，且 $n_{i/p}$=B。但是，对于 CEONA-I 架构，DR=(SR/2^B)，且 $n_{i/p}$=1 [31]。我们令 M=N，并采用 [31]中的方法求解方程。图 7 展示了 CEONA-I、AMW 和 MAW 架构在不同符号速率SR下、不同 B 级别所能实现的 N。如图 7 所示，在不同 SR 下的所有位精度级别中，与 AMW 和 MAW 相比，我们的 CEONA-I 均可支持更大的 N值。例如，在 1 GS/s 下，CEONA-I 在 4 位精度时实现的 N=192 大于 AMW 和 MAW，后两者分别实现 N=31 和 N=44。这是因为 CEONA-I 的 PCA 具有原位累加能力 [9]（见第 2.2 节），使 PCA 能够以较低的 DR=(SR/2^B)运行，从而在更大的 B下显著提高 N。相比之下，随着 B [27] 增加，AMW 和 MAW 所支持的 N 会减少。此外，可实现的 N也受波长间距限制。我们采用较为乐观的自由光谱范围 (FSR)=50nm。对于 AMW 和 MAW，波长间距设为 0.8nm [27]；对于 CEONA-I，则可设为 0.25nm [31]。因此，在图 7 中，AMW/MAW 和 CEONA-I 的 N 值上限分别为 62（=FSR/0.8nm）和 200（=FSR/0.25nm）。

$$n_{i/p} = \frac{1}{6.02}\left[20log_{10}\left(\frac{R \times P_{PD-opt}}{\beta\sqrt{\frac{DR}{\sqrt{2}}}} - 1.76\right)\right] \quad (1)$$

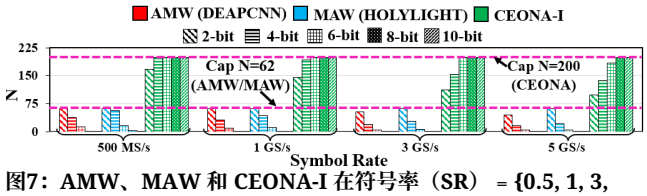


图7：AMW、MAW 和 CEONA-I 在符号率（SR）={0.5, 1, 3, 5}千兆符号/秒下，位精度为 ={2, 4, 6, 8, 10}位时所支持的 CoPE 大小 N。

$$\beta = \sqrt{2q(RP_{PD-opt} + I_d) + \frac{4kT}{R_L} + R^2P^2_{PD-opt}RIN} \quad (2)$$

$$P_{Laser} = \frac{10^{\frac{\eta_{WG}(dB)[N(d_{OSM})]}{10}}M}{\eta_{SMF}\eta_{EC}IL_{i/p-OSM}} \times \frac{P_{PD-opt}}{\eta_{WPE}IL_{MRR}} \times \frac{1}{(OBL_{OSM})^{N-1}(EL_{splitter})^{log_2M}} \times \frac{1}{(OBL_{MRR})^{N-1}(IL_{penalty})} \quad (3)$$

3.3 案例 II：储备池计算

CEONA 还可配置为延迟反馈储备池计算加速器 (CEONA-DFRC)，用于时间序列任务的训练和推理。为此，CEONA-DFRC 将 PBAU 配置为传统调制器，以调制输入掩码信号 $u_i(t)$（图 4）。滤波器组中的每个 MRR 均配置为充当储备池的非线性节点 [5]。

有源微环谐振器的下载端口传输特性呈现出丰富的非线性响应，这是由双光子吸收（TPA）[6]导致的。非线性程度取决于微环谐振腔的光子寿命（τ_{ph}）。对于微环谐振器，τ_{ph} 取决于其品质因数。因此，可通过调节微环谐振器的品质因数（进而调节τ_{ph}）来控制其非线性。为实现对微环谐振器品质因数（进而对τ_{ph}）的控制，我们采用了图4中[18, 23] 所示的非线性微环谐振器设计。

在图 4 中，非线性微环谐振器与延迟反馈波导环路共同构成一个储备池 [32]。微环谐振器的品质因数经过调节 [18, 29]，以根据任务设定非线性程度。微环谐振器通过对调制后的 $u_i(t)$ 进行非线性变换来生成储备池的状态。储备池的状态 $s_i(t)$ 被发送至延迟反馈波导环路，在此利用光电探测器阵列进行捕获和存储；部分 $s_i(t)$ 信号则作为反馈，进一步馈入非线性微环谐振器，并与调制后的掩码输入 $u_i(t+1)$ 一同输入。CEONA-DFRC 的运作方式和架构已在 [32] 中进行了详细讨论。

图 8(a)、8(b) 和 8(c) 比较了各种延迟反馈储备池计算加速器在不同时间序列任务中的符号错误率（SER）、归一化均方根误差（NRMSE）和训练时间。从图 8(a) 可以看出，在不同目标信噪比下取平均后，CEONA-DFRC 在非线性信道均衡任务 [13]上的 SER 比全光学（MZI）[11] 低 58.8%。同样，图 8(b) 显示，对于 NARMA [12] 和 SantaFe [25],，CEONA-DFRC 的 NRMSE 比全光学（MZI）_低 35%，且表现与电子式（MG）[5]相当。CEONA-DFRC 的主要优势体现在图 8(c) 中；它显著缩短了训练时间 98×

与全光学_（MZI）和电子式（MG）相比，平均分别快 93× 。CEONA-DFRC 利用有源微环谐振器丰富的非线性特性，在储备池层中实现非线性节点。此外，与全光学 MZI 和电子式 MG 储备池相比，CEONA-DFRC 的基于微环谐振器的储备池转换掩码输入信号所需时间分别少168× 和 2*10^5× 。此外，CEONA-DFRC 使用光波导作为延迟反馈环路，进一步缩短训练时间 [32]。总体而言，与以往的光学和电子式深度反馈储备池计算相比，CEONA-DFRC 显著缩短了时间序列任务的训练时间，同时取得更低或相当的误差。

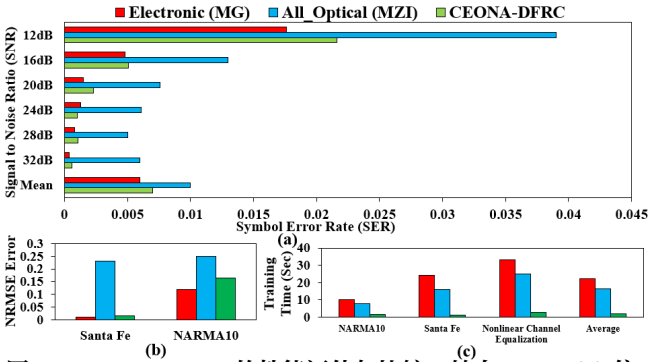


图8：CEONA-DFRC 的性能评估与比较，转自 [32]。(a) 信道均衡任务在不同目标信噪比下的符号错误率，(b) 圣菲和 NARMA 序列预测的归一化均方误差，以及 (c) 所考虑任务的训练时间比较。

4 总结与未来展望

本文展示了我们发明的 PEOC 和 CEONA。我们证明，PEOC 可重新配置，以便在不同时间实现不同的逻辑和算术功能。同样，我们证明，CEONA 可用于灵活支持卷积神经网络和深度反馈储备池计算应用的加速。我们还展示了 CEONA 可重新配置，以处理二值化卷积神经网络和整数定点量化卷积神经网络。在这两种情况下，与既有工作相比，CEONA 在吞吐量、能效和面积效率方面均具有显著优势。此外，我们还展示了 CEONA 在加速各种时间序列任务（如 NARMA10、SantaFe 和非线性信道均衡）的推理与训练方面，能够显著降低延迟。

我们设想，CEONA 可扩展以支持混合精度卷积神经网络的加速 [10]。此外，我们的 CEONA 架构中的每个 CoPE 都可以通过波分复用方式加速多个时间序列任务，从而显著提高加速器的吞吐量。总而言之，这种灵活性可使 CEONA 同时处理来自机器学习和人工智能应用的多个工作负载。

致谢

本研究由美国国家科学基金会的资助项目（CNS-2139167）支持。

参考文献

[1] A.Ganguly 等2022。《面向 DNA、量子、存内和光计算的互连：小组讨论的见解》。《*IEEE Micro*》42, 3 (2022), 40–49。

[2] M. A. Al-Qadasi 等。2022。硅光子加速器的扩展：挑战与机遇。《APL Photonics》(2022)。[3] Armin Alaghi 和 John P Hayes。2013。在随机电路设计中利用相关性。载于《2013 IEEE第31届计算机设计国际会议》(ICCD)。IEEE，39–46。[4] ANSYS。2003。Lumerical。http://www.lumerical.com/products [5] Lennert Appeltant 等。2011。将单个动力学节点用作复杂系统进行信息处理。《Nature Communications》2 (2011)。[6] Meisam Bahadori 等。2017。面向高性能计算的硅光子互连网络能耗与性能优化设计。载于 DATE。[7] Viraj Bangari 等。2020。面向卷积神经网络的数字电子与模拟光子学（DEAP-CNN）。《JSTQE》(2020)。[8] W. Bogaerts 等。2012。硅微环谐振器。《Laser & Photonics Reviews》6, 1 (2012), 47–73。[9] Frank Brückerhoff-Plückelmann 等。2022。基于电荷累积的大规模光子矩阵处理器。《Nanophotonics》(2022)。[10] Weihan Chen 等。2021。通过约束优化实现神经网络的混合精度量化。载于 ICCV。[11] François Duport 等。2016。全模拟光子储备池计算机。《Scientific Reports》(2016)。[12] Herbert Jaeger。2002。利用回声状态网络进行自适应非线性系统辨识。载于 NIPS。[13] Herbert Jaeger 等。2004。利用非线性：预测混沌系统并降低无线通信能耗。《Science》(2004)。[14] Venkata Sai Praneeth Karempudi 等。2022。采用多电平信号传输的光子片上网络：跨层比较研究。《JETCS》(2022)。[15] Venkata Sai Praneeth Karempudi、Shreyan Datta 和 Ishan G Thakkar。2021。CMOS集成、可变形态纳米光子算术逻辑单元的设计探索与可扩展性分析。载于《第19届ACM嵌入式网络传感器系统会议论文集》。628–634。[16] V. Karempudi 等。2023。用于高速可重构计算电路的可变形态电光逻辑门。arXiv 预印本 arXiv:2301.13626 (2023)。[17] Weichen Liu 等。2019。HolyLight：面向数据中心深度学习的纳米光子加速器。载于 DATE。[18] Enxiao Luan 等。2023。利用强度调制微环和光子线键合构建高密度光子张量核心。《Scientific Reports》(2023)。[19] C. Qiu 等。2012。利用硅光子集成电路实现可重构电光逻辑。《Optics Letters》37, 19 (2012), 3942–3944。[20] B.J. Shastri 等。2021。面向人工智能和神经形态计算的光子学。《Nature Photonics》15 (2021), 102–114。[21] Kyle Shiflett 等。2020。PIXEL：光子神经网络加速器。载于 HPCA。[22] K. Shiflett 等。2021。利用硅光子学实现按位神经网络加速。载于 GLSVLSI。9–14。[23] Hossam Shoman 等。2018。具有可调消光比和宽自由光谱范围的紧凑型硅微环调制器。载于 OFC。[24] Alexander Sludds 等。2022。互联网边缘的分布式光子深度学习。《Science》(2022)。[25] Andrew R. Solow。1994。预测未来，理解过去。《Science》(1994)。[26] Sairam Sri Vatsavai 和 Ishan Thakkar。2023。位并行确定性随机乘法器。arXiv 电子印本 (2023), arXiv–2302。[27] Sairam Sri Vatsavai 和 Ishan G. Thakkar。2022。用于混合尺寸张量卷积神经网络高效推理的光子可重构加速器。《TCAD》(2022)。[28] Febin P. Sunny 等。2021。ROBIN：鲁棒光学二值神经网络加速器。《ACM嵌入式计算系统汇刊》(2021)。[29] Sairam Sri Vatsavai 等。2020。PROTEUS：光子片上网络中的基于规则的自适应，实现激光功率与性能的损耗感知协同管理。载于 NOCS。[30] Sairam Sri Vatsavai 等。2023。用于二值神经网络高效推理的基于光学同或位计数的加速器。arXiv: 2302 06405 [ ]。[31] 2023 cs.AR Sairam Sri Vatsavai 等。《SCONNA：用于整数定点量化卷积神经网络超高速、低能耗推理的基于随机计算的光学加速器》。arXiv: 2302 07036 [ ]。cs.AR Sairam Sri Vatsavai 和 Ishan Thakkar。[32] 2021。用于延迟反馈储备池计算的基于硅光子微环的芯片级加速器。载于 VLSID。[33] Zhoufeng Ying、Chenghao Feng、Zheng Zhao、Shounak Dhar、Hamed Dalir、Jiaqi Gu、Yue Cheng、Richard Soref、David Z Pan 和 Ray T Chen。2020。用于高速计算的电子-光子算术逻辑单元。《Nature Communications》11, 1 (2020), 2154。[34] Z Ying 等。2019。用于光计算的集成式多操作数电光逻辑门。《APL》(2019)。[35] F. Zoakee 等。2020。LightBulb：基于光子非易失性存储器的二值化卷积神经网络加速器。载于《2020 DATE》。IEEE，1438–1443。