
北京大学团队领衔发表存内计算技术全谱

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/24950.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

北京大学团队领衔发表存内计算技术全谱。北京时间2023年11月14日，北京大学集成电路学院/人工智能研究院研究团队联合国内外专家学者，在《自然—电子学》（Nature Electronics）在线发表了存内计算技术全谱论文（A full spectrum of computing-in-memory technologies）。该论文提出存内计算技术全谱概念，对所有类型存内计算技术进行了原理性分类，从而为比较每种不同技术的性能提供了一个平台，同时有望启发新型的存内计算技术。

论文的第一作者与通讯作者是北京大学孙仲研究员，共同作者包括以色列理工学院Shahar Kvatinisky副教授，东南大学司鑫副教授，伦敦大学学院Adnan Mehonic副教授，以及北京大学蔡一茂教授、黄如院士。

过去几十年间，在摩尔定律的驱动下，处理器的性能得到了飞跃式的发展。然而，传统计算机采用冯·诺依曼架构，数据的处理与存储在物理上是分离的，二者之间的数据传输引起了极大的计算时延与能耗。此外，尽管传统逻辑门进行计算时具有通用性和鲁棒性，但其效率较低，在乘法、加法和非线性函数等运算中消耗大量的硬件与时间资源。在物联网和大数据时代，我们迫切需要进行使计算范式从冯·诺依曼架构中转移出来的技术革命。

存内计算技术是解决这一瓶颈的潜在解决方案，其基本思想是将数据计算移动到存储器中，从而实现原位计算，消除带宽限制和数据传输成本。它的基本原理是利用物理定律，如存储器阵列中的基尔霍夫电流定律和电荷共享机制等，高效地实现诸如逻辑门操作和乘累加（MAC）运算等计算原语。

目前，所有类型存储器都已被报道用来实现存内计算，包括易失性存储器（如成熟的SRAM、DRAM技术）和非易失性存储器（如成熟的Flash技术，以及新兴的阻变存储器RRAM、PCM等）。在传统冯·诺依曼计算机中，CMOS晶体管构成的逻辑门完备集是基础，它们被用来构建处理核进行标量MAC运算等算术操作，再通过串行处理或多核并行处理实现向量/矩阵运算，进而实现各种算法（如神经网络）。存内计算架构则采用截然不同的路径，它的基础是基于物理定律的MAC运算，利用存储器阵列的并行化进行向量/矩阵算术操作，并同样用于实现各种算法，再通过神经网络的阈值逻辑概念来执行逻辑门。可以看到，在两种计算架构中，各种计算原语形成了一种衔尾蛇（Ouroboros）的结构（图1）。

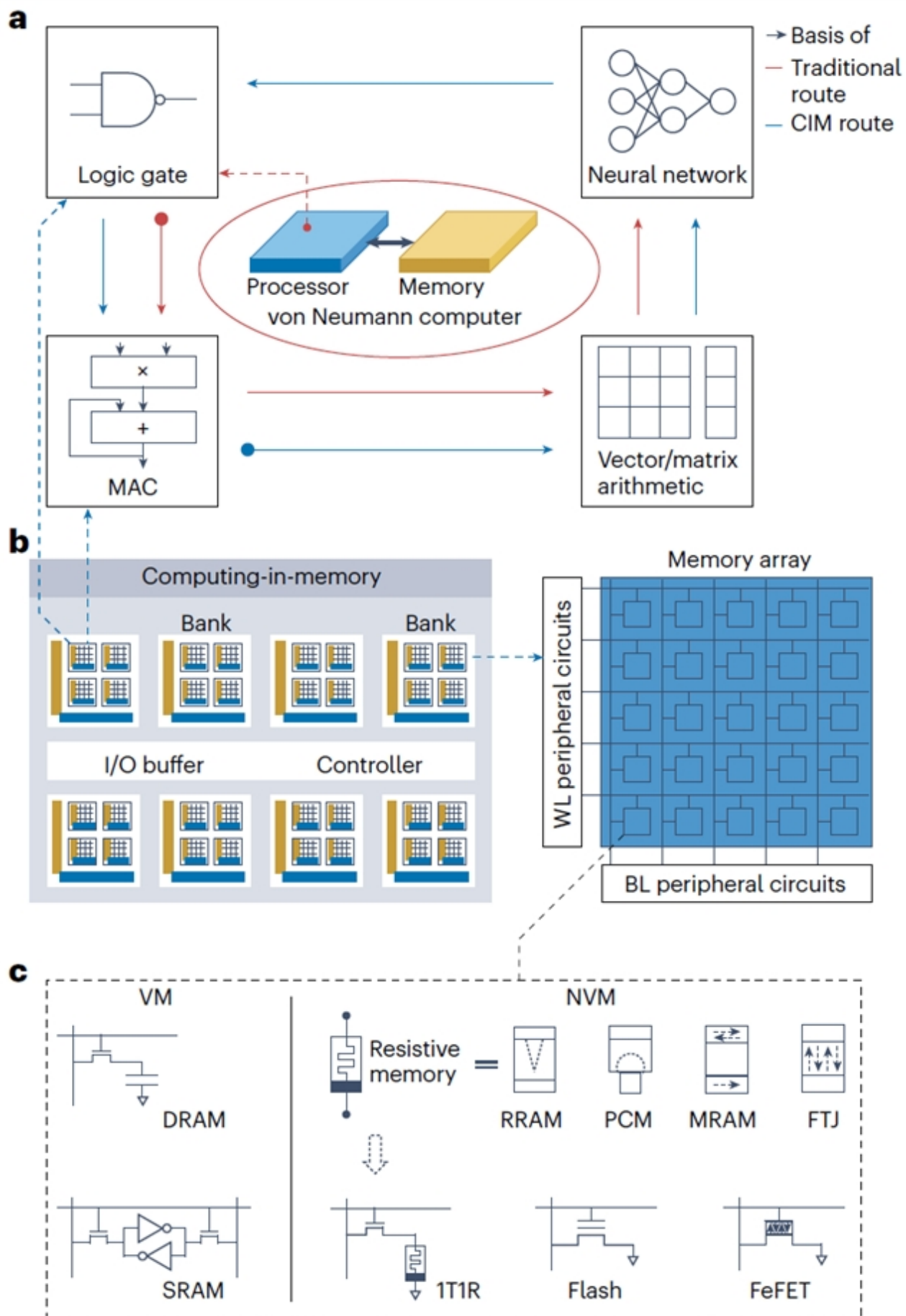


图1. 存内计算架构与计算原语

逻辑门和MAC运算是存内计算架构中两个基本的计算原语，它们在不同类型存内计算技术中的实现方式存在原理性不同，具体取决于：（1）输入操作数是否由存储单元原位提供；（2）计算输出是否原位回存至存储单元；（3）输入/输出数据是易失性的还是非易失性的；（4）输入/输出数据是否以相同的物理变量表示。这些差异使得人们很难对不同存内计算技术形成包容性的全面了解，也阻碍了学术界和工业界不同领域（如半导体器件、IC设计、计算机体系结构等）之间的理解与沟通。本论文以此为出发点，提出了一个存内计算技术全谱，对其进行了全面的分类与评估。

作者将计算原语抽象为 $Z = X \square Y$ 的形式，其中X和Y可以是标量或矢量的输入，Z是标量的输出， $\square$ 为表示存内计算操作的符号。根据输入X和Y是否由存储单元提供，以及完成计算时输出Z是否重新存储在存储单元中，将所有的存内计算技术分类为六种（XYZ-CIM、XZ-CIM、Z-CIM、XY-CIM、X-CIM和O-CIM），且每种类型都可以用几种常见的易失性或非易失性存储器实现，并用于执行逻辑门操作或并行MAC运算（图2）。作者还进一步阐述了各种存内计算技术的实现方式，总结了其主要特点、优缺点和部分应用，以及这些技术在器件可靠性、计算复杂性、延迟和能源效率等多个方面所面临的挑战。

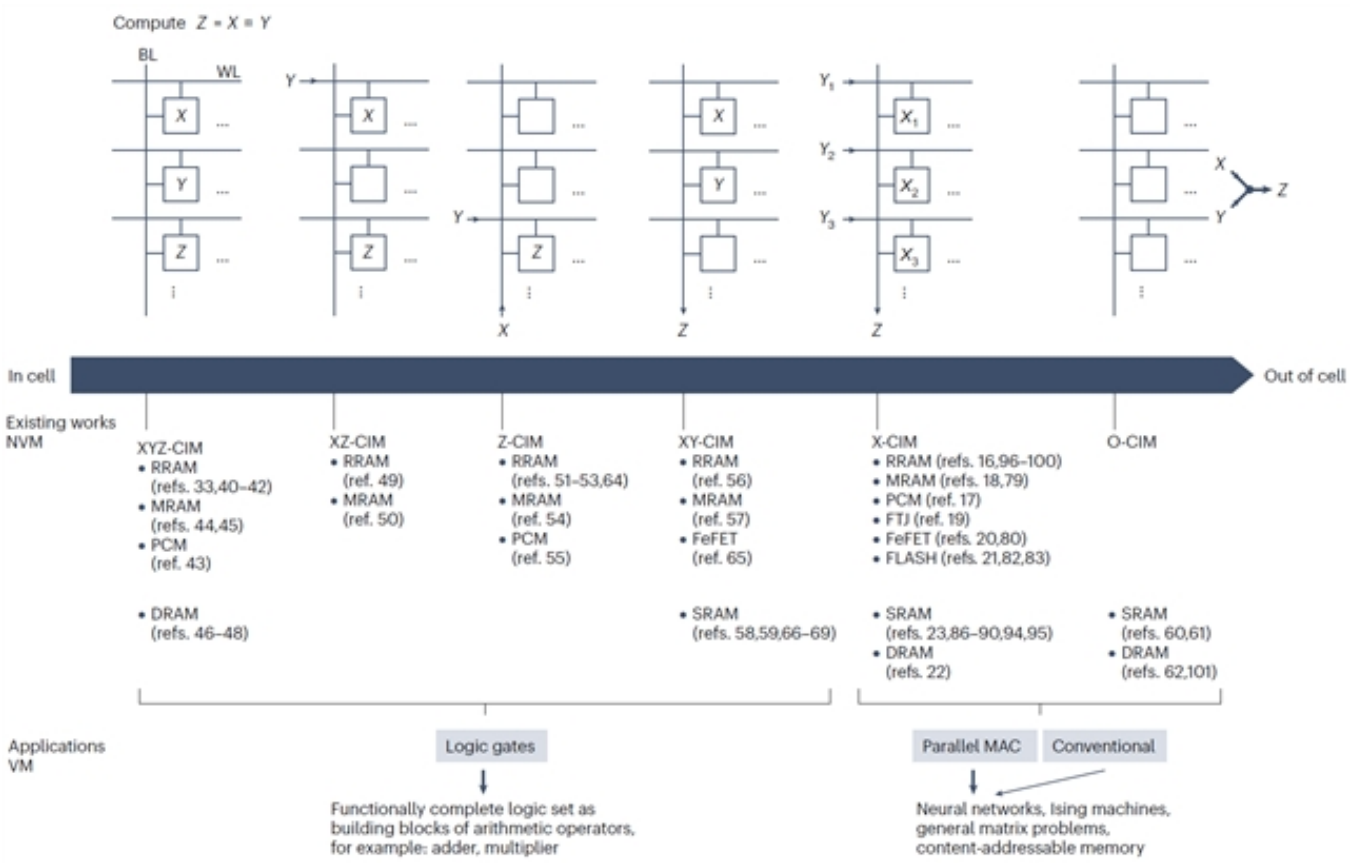


图2. 存内计算技术全谱

总而言之，该全谱将所有存内计算技术统一为连续的整体，明晰了全谱范围内标准存内计算技术的基本原理，并提供了一个可以在各方面评估不同技术的统一标准。结合该全谱，可以基于相同的存储技术对不同类型的存内计算技术进行集成来实现取长补短，还有可能用于开发其它的存内

计算技术，这对所有存内计算技术的发展都具有重要意义。

该项研究工作获得了中国科技部、国家自然科学基金、111计划，以色列NSF-BSF基金，英国皇家工程院等项目的支持。（来源：科学网）

相关论文信息：<https://doi.org/10.1038/s41928-023-01053-4>

作者：孙仲等 来源：《自然—电子学》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发