
新计算架构，迎来从学术到产业实践重大机遇

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/38328.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！

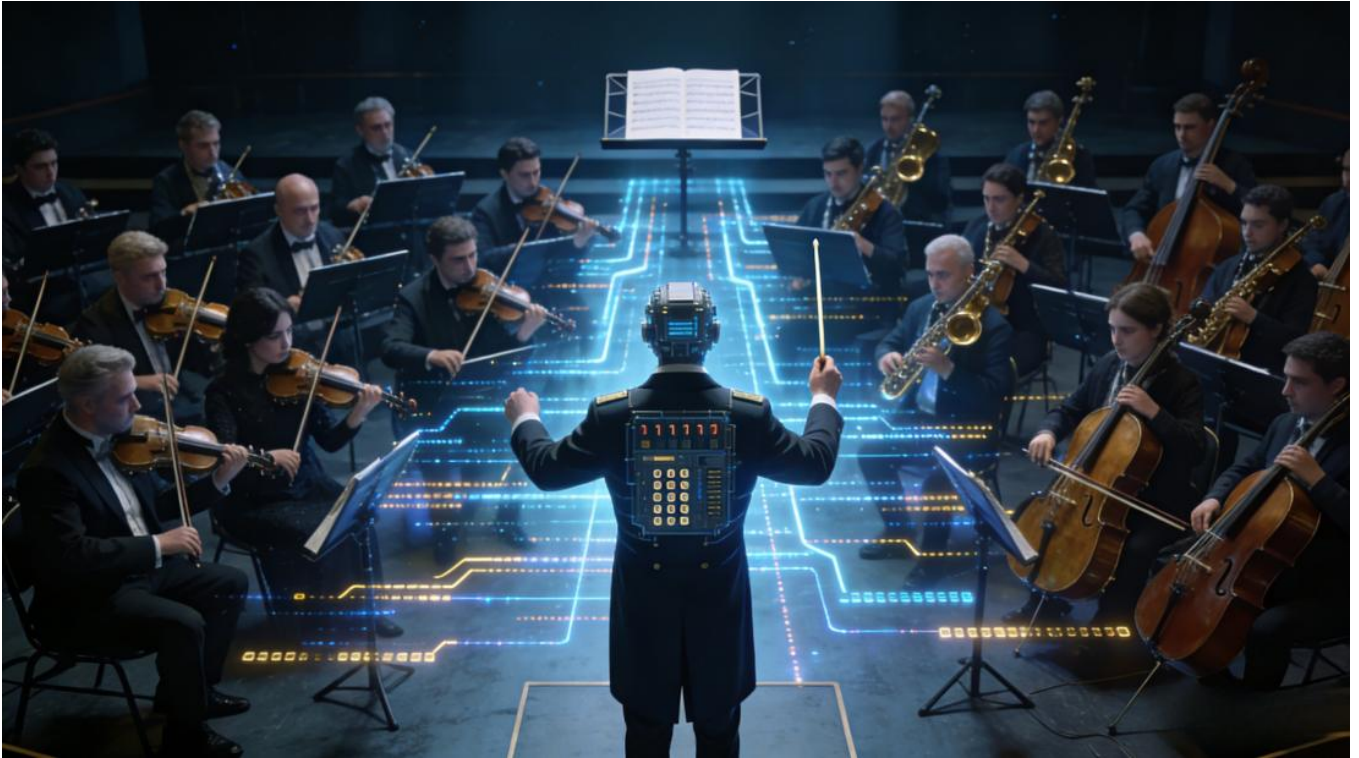
新计算架构，迎来从学术到产业实践重大机遇

。近期，英伟达以200亿美元天价拿下美国人工智能芯片独角兽Groq的核心资产，英特尔拟以16亿美元将美国“数据流架构”芯片公司SambaNova收入囊中。这两笔合计超210亿美元的芯片领域并购大案，并非寻常的商业布局，而是对下一代计算架构的集体押注——数据流架构，这个沉寂数十年的技术方向，正从学术殿堂走向产业中央。

Groq的语言处理芯片（LPU）在AI推理任务中实现了约10倍于GPU的性能突破，SambaNova的数据流芯片（RDU）芯片更迭了AI计算范式，它们的技术内核指向同一个答案：在冯·诺依曼架构面临核心瓶颈的今天，数据流正在开启一场计算架构突破的序幕。正如GPU曾让英伟达超越英特尔登顶芯片王座，数据流架构正为全球企业提供改写行业格局的历史性机遇。

架构之变：打破冯·诺依曼瓶颈的技术革命

自1945年冯·诺依曼提出存储程序架构以来，CPU与GPU始终遵循着程序计数器+共享内存的核心逻辑。这种控制流架构如同严谨的交响乐团，所有指令必须在程序计数器的指挥下按部就班执行，数据与指令共享内存空间，通过总线频繁搬运。



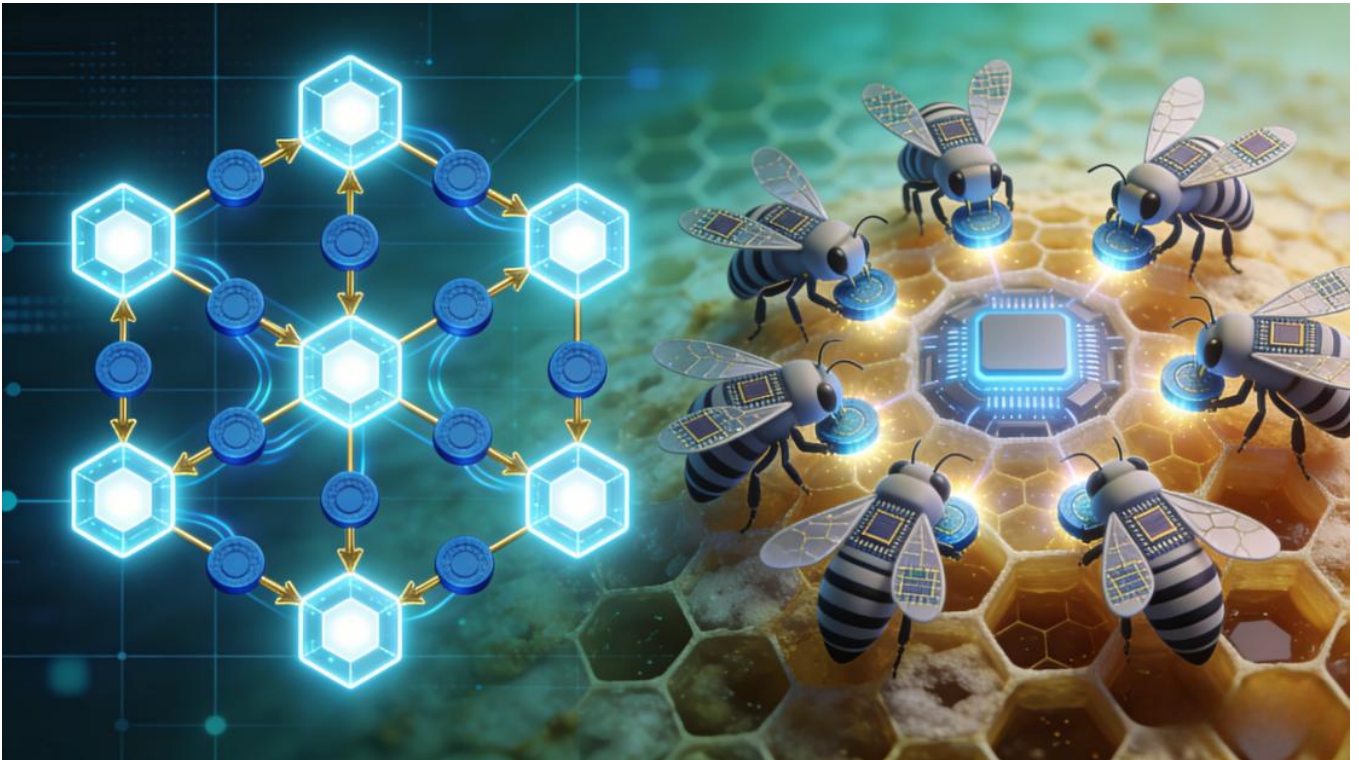
CPU+GPU的控制流架构形象图。图片由豆包AI生成

尽管冯·诺依曼架构至今仍是全球计算设备的主流选择，在通用计算场景中具备不可替代的兼容性与成熟度，但在AI大模型与科学计算的浪潮中，其固有的架构局限逐渐凸显：现代处理器的计算速度远超数据访问速度，处理器大部分的时间在等待数据，如同一群运动健将被困在狭窄的独木桥前。当AI大模型需要每秒处理万亿次浮点运算，当气候模拟需要实时调度PB级数据，传统控制流架构的存储墙与能效墙显著制约了算力提升的上限。

数据流架构的出现，为破解这一瓶颈提供了全新路径。

数据流架构摒弃了程序计数器的集中控制，将计算的主导权交还给数据本身，其核心原理可概括为数据就绪即执行，数据流过即计算：程序被拆解为节点（计算操作）与有向边（数据依赖）构成的有向无环图，如同精密的工厂流水线；数据以令牌形式在节点间流动，当节点收集齐所有输入令牌即被点火执行，无需等待指令调度；所有就绪节点可同步执行，实现指令级、线程级、任务级的全维度并行，硬件利用率大幅提升。

这种架构如同蜂群协作，没有统一指挥，每个工蜂（计算单元）收到任务（数据）后立即行动，协作效率远超传统架构的“单兵作战”。



蜂群协作式的数据流架构。图片由豆包AI生成

数据流架构计算有四大核心优势。

一是能极致释放并行性。传统控制流结构在执行过程中虽然也允许乱序执行和多发射执行，但是仍然只能在指令窗口内选取可执行的指令执行，并行度受限；而在数据流结构中，程序被完全展开成数据流图的形式，所有指令的执行没有强制顺序，操作数到齐即可发射执行，可以在整体程序范围充分挖掘指令级并行。

二是有着突出的低延迟特性。数据直接流动、无集中式指令调度的设计，可大幅减少指令等待与数据搬运的时间损耗，端到端延迟较传统架构大幅降低，尤其适配边缘AI推理、高频交易等对响应速度敏感的场景——这也是通用并行图形处理器（GPGPU）难以企及的核心优势，更是英伟达斥资200亿美元收购Groq的关键考量，旨在弥补其在低延迟计算领域的短板。

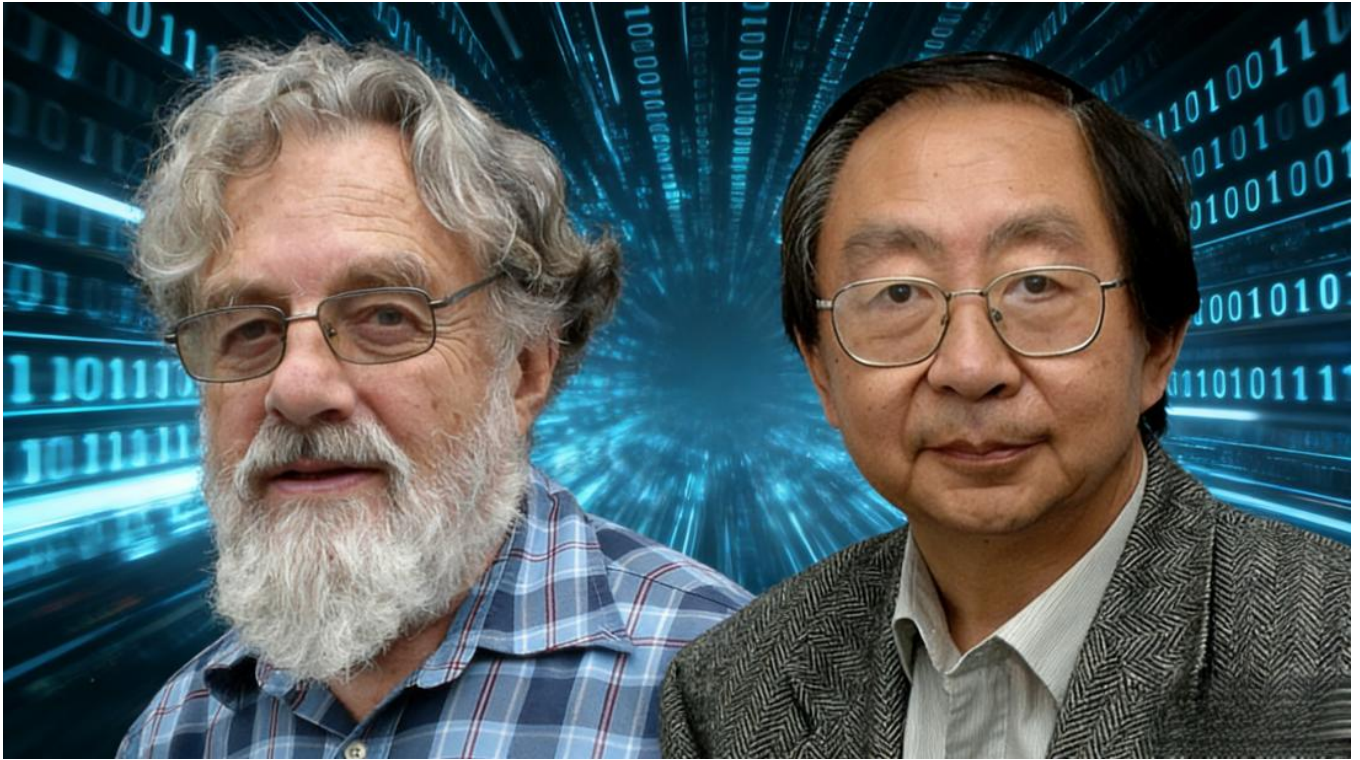
三是访存效率更高。采用分布式局部存储，数据在计算单元间直接流动，访存能耗可大幅降低，存储墙难题也可大大缓解。

四是能效比显著提升。省去分支预测、重命名寄存器等复杂控制逻辑，同等面积下算力密度更高，能效比优化明显。

如今，全球头部芯片厂商已纷纷入局数据流赛道：谷歌TPU使用了的专用数据流架构，英伟达Tensor Core本质上是为矩阵运算特殊固化的数据流计算单元，AMD则推出集成XDNA架构的Ryzen AI处理器——其自适应数据流架构通过专用AI引擎实现数据驱动的并行计算，既保留了冯·诺依曼架构的通用计算能力，又通过数据流模块实现低延迟的AI任务加速，在笔记本等消费级边缘设备中落地，体现了两种架构的协同价值。

源流溯往：数据流架构的薪火与传承

谈及数据流计算的起源，MIT教授杰克·B·戴尼斯（Jack B. Dennis）是无可争议的奠基人。他于20世纪60年代首次系统提出数据流计算模型，摒弃程序计数器的集中控制，确立数据依赖驱动执行的核心思想，为非冯·诺依曼架构开辟了全新道路。而将这一理论推向工程化、并构建起全球化学术传承脉络的关键人物，正是他的弟子、中国学者高光荣教授。



杰克·B·戴尼斯（左）、高光荣（右）。作者供图

1980年代，高光荣赴MIT攻读博士学位，师从戴尼斯，成为其核心研究团队的重要成员。在戴尼斯的指导下，高光荣深耕数据流理论，不仅掌握了非冯·诺依曼计算模型的精髓，更敏锐发现纯数据流架构难以兼容复杂控制逻辑的行业痛点。作为戴尼斯学术思想的直接传承者与创新者，他创新性地提出了EARTH（Efficient Architecture for Running Threads）模型——这一创新既保留了数据流的天然并行优势，又通过线程级同步机制兼容传统程序语义，成为连接理论与实践的关键桥梁。

两人的学术羁绊并未止步于师生关系：高光荣毕业后，与戴尼斯保持长期深度合作，戴尼斯提出的基础数据流模型，经高光荣的工程化拓展，最终落地为IBM Cyclops64超级计算机的核心架构，2005年即实现了每秒超过100万亿次运算的巅峰性能，成为数据流技术从实验室走向产业的里程碑。

戴尼斯与高光荣这种提出理论—工程转化—全球推广的传承脉络，构成了数据流架构发展的核心主线。而高光荣对数据流工程化的开拓，也点燃了其在中国发展的星星之火。中国科学院计算技术研究所、华中科技大学、清华大学、上海交通大学的相关研究团队，始终深受其学术思想的启迪，在基于和融合数据流思想的处理器体系结构研究与芯片设计方面开展了长期的研究工作。从数据流与控制流的融合理念，到高效并行执行的核心逻辑，高光荣的研究成果为这些团队提供了

坚实的学术养分与技术基石。

中国科学院计算技术研究所研究员、中科睿芯创始人范东睿在美国访学期间，跟随高光荣教授深度研习数据流理论，深受其思想熏陶，回国后深耕十余年，将高通量计算技术产业化。中科睿芯主导研发的数据流芯片已应用于安防、自动驾驶等领域。

从戴尼斯的理论奠基，到高光荣的工程化突破，再到范东睿等学者的产业转化，数据流架构的学术薪火跨越国界，在中国落地生根并结出硕果。

时代之机：AI浪潮下的产业重构与历史机遇

AI大模型的爆发式增长，为数据流芯片创造了绝佳的产业化窗口。根据Global Market Insights的最新报告，2024年全球数据流AI处理器市场规模达到52亿美元，预计2025年将增长至57亿美元，到2034年将达到147亿美元，年复合增长率为11.1%。

数据流处理器的应用场景，也正从AI推理向全领域渗透。AI与机器学习领域，神经网络的矩阵运算天然适配数据流图，谷歌TPU在大模型推理方面实现了显著的成本降低；在科学计算中，流体力学、量子模拟等场景中，数据流芯片可将计算效率大幅提升；边缘计算方面，低功耗、低延迟的双重优势使数据流芯片成为边缘AI的理想选择；而在金融科技行业，高频交易、风险建模等场景中，数据流的低延迟特性可大幅减少决策响应时间。

市场需求催生技术迭代，谷歌2025年发布的Ironwood TPU、SambaNova与Hume AI合作的多语言模型加速器，均印证了数据流架构的产业化潜力。正如GPU曾让英伟达超越英特尔，数据流架构为后发企业提供了改写行业格局的契机。

笔者认为，当前全球数据流处理器产业仍处于发展初期，生态尚未固化，中国企业已具备三大核心优势。

一是技术积累先发优势。经过二十余年深耕，中国已构建完整的数据流技术体系：基础研究层面，中国科学院计算技术研究所、华中科技大学、清华大学、上海交通大学均长期开展相关研究工作，为数据流技术的持续发展提供了技术与人才基础；硬件层面，有中科睿芯研发的Godson-T、SPU、DPU等多款芯片；软件层面，中科院计算所提出的COStream编程框架已初步具备商业化潜力；应用层面，在科学计算、图计算等领域形成特色优势。

二是应用场景规模优势。中国拥有全球最大的AI应用市场，短视频、自动驾驶、智慧城市等场景产生海量数据流计算需求。科学智能（AI4Science）的应用特征需要同时对AI与科学计算优化的体系结构，为数据流处理器提供了新的重要场景。反过来，这些应用场景为技术迭代提供了宝贵的数据反馈。

三是政策与资本双重加持。国家东数西算工程明确提出要加快突破先进计算架构、高性能芯片等关键核心技术，提升算力供给质量和效率；集成电路产业专项基金（大基金）三期将芯片产业列为重点投资方向。政策与资本的合力，正在加速技术产业化进程。

上世纪90年代，英伟达凭借GPU架构抓住图形计算浪潮，后来伴随AI时代的到来，超越英特尔成为行业霸主；今天，数据流架构引发的计算架构突破，为中国企业提供了相似的历史契机。当英伟达、英特尔、AMD三大芯片巨头纷纷通过收购的方式布局数据流技术时，中国企业已在这条

赛道上积累了二十年技术沉淀——这正是实现弯道超车的最佳时机。

数据流重塑计算未来

当英伟达的200亿美元收购案尘埃落定，当英特尔、AMD加速布局数据流赛道，一个清晰的信号已然浮现：计算架构的突破正在发生。数据流架构不仅是缓解冯·诺依曼瓶颈的技术方案，更是全球科技竞争的产业关键点。

从高光荣在麻省理工学院的理论探索，到国内多个团队专家学者的技术攻坚，再到国产芯片的量产应用，中国在数据流赛道上的深耕，为全球芯片业提供了中国方案。正如GPU成就了英伟达，数据流架构正在孕育新的行业巨头——而这一次，中国企业有望占据产业链的核心位置。

在AI与算力需求爆发的时代，数据流不仅是数据流动的路径，更是技术创新的脉络、产业升级的动脉。当数据成为数字经济的核心生产要素，以数据流架构为核心的计算革命，有望重塑全球科技格局。而那些抓住这一历史机遇的企业，也有潜力成为下一个计算时代的领航者。

（作者系清华大学计算机科学与技术系教授、中国计算机学会会士，《中国科学报》记者赵广立编辑整理）

作者：陈文光 来源：中国科学报

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