首款神经动力学芯片问世

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/40750.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！

首款神经动力学芯片问世

。本报北京7月3日电（记者晋浩天）大脑皮层表面那层复杂的沟回褶皱，想要在计算机里实时重建，过去需要昂贵的大型计算设备离线运算良久。如今，这一局面被一颗拇指大小的芯片改写。北京大学集成电路学院教授杨玉超团队联合中国科学院上海微系统与信息技术研究所研究员宋志棠团队，成功研制出全球首款基于相变忆阻器的神经动力学系统芯片，首次将这类复杂运算的单步时延压缩至2.12毫秒，在脑皮层重建等任务中较目前先进图形处理器（GPU）提速达50至478倍，一举突破了制约神经动力学长达半个世纪的实时计算瓶颈。相关成果3日凌晨发表于《科学》。

杨玉超告诉记者，要让机器像大脑那样实时建模和理解物理世界，需要一种将神经网络与微分方程相结合的“神经动力学系统”。它能在不完整、带噪声的数据中重建出平滑精确的三维脑结构，应用潜力巨大。然而，传统计算架构存在一个核心瓶颈：存储与计算分离，求解过程中海量的中间变量在内存和处理器之间反复“奔波”，如同一个庞大的数据工厂，大量时间被浪费在搬运路途中，不仅延迟巨大，功耗也居高不下。

面对这一难题，研究团队从忆阻器本身的物理特性里找到了破局答案。他们利用相变存储器独特的“电导漂移”现象——在一定时间窗口内，其电导变化是可预测、可精准调控的。基于此，团队提出“可控存内计算”新范式，通俗地讲，原本需要复杂数字电路反复执行的运算、缓存访问、数据搬运等工作，现在交给了器件自身的物理规律去“跑”。

更值得关注的是，团队还将神经网络权重映射到相变存储器的多级电导态上，在同一个阵列内同步完成矩阵乘加运算。两大核心计算任务由此被统一集成在总面积仅0.28平方毫米的存算阵列中。这颗采用40纳米工艺的芯片实现了2.12毫秒的单次迭代时延，首次将神经动力学硬件推入毫秒时代。

“性能表现令人振奋。”杨玉超表示，在同等运算下，该芯片较当前最先进的专用加速器速度提升3.82至36.27倍，功耗大幅降低；在脑皮层表面高保真重建任务中，该芯片较国外先进GPU提速达478.18倍。重建出的脑皮层网格平滑、拓扑一致，能精准刻画复杂的褶皱结构，并有效抑制传统方法中的伪影和自相交缺陷。

杨玉超说，这一突破为脑机接口和脑疾病诊疗开启了全新想象空间。未来，个体化、动态化的脑数字孪生成为可能，术中神经导航、阿尔茨海默病早筛及个性化干预等，将获得可实时运行的硬件底座。

●新知解码

什么是“可控存内计算”

本报记者晋浩天

如果把传统计算机比作一间办公室，处理器是坐在中央的“计算员”，存储器则是满墙的“档案柜”。每次计算，计算员都得起身取数据，算完再回去——时间都花在了路上。这就是著名的“冯·诺依曼瓶颈”：存储与计算分离，数据搬运拖垮效率。

“存内计算”思路直接：让档案柜自己学会算账。数据不再需要搬来搬去，而是在存储单元内部原地完成计算。这听起来完美，但实现起来困难重重——存储单元天生只会“记”，让它同时“算”，还得算得准、算得稳，已是很大的挑战。

更大的难题在于“可控”。计算时很多任务需要动态调整、相机判断。如何让一堆物理器件具备“临场应变”能力，是存内计算走向现实的关键门槛。

突破口来自一个“反科学直觉”的思路：利用器件电导会规律性漂移这一曾被视作“缺陷”的特性。如果摸清它的变化轨道，这种漂移就能被驯化为计算能力——不再用数字电路反复读写比较，而是让物理过程本身完成运算。

这便是“可控存内计算”的核心思想：让存储单元在“记”的同时，按设计者设定的方式、在可约束的范围内完成“算”。存储即计算，且整个过程精准可控。基于这一范式研制的芯片，能将复杂运算压缩到毫秒级别，能效提升数十甚至数百倍。

作者：晋浩天 来源：光明日报

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转发