中国移动研究院完成业界首次忆阻器存算一体芯片的端到端技术验证

2023-05-31 16:00 本文共 1088 字 阅读完需 4.5 分钟

5月29日,中国移动联合清华大学完成业界首次忆阻器存算一体芯片的端 到端技术验证,成功打通芯片、软件、算法、应用全流程,标志着基于 新型存储器的存算一体技术向产业化迈出了关键一步,为拓展存算一体 芯片在端边侧应用、构建存算一体产业生态奠定了重要基础。

存算一体作为先进计算技术,从架构上突破冯·诺依曼架构存储和计算分离导致的存储墙和功耗墙问题,可大幅提升算力和能效水平,在AI时代的重要性日益公显。

中国移动自2021年起开展存算一体技术研究,并与高校和产业伙伴联合推动存算一体多技术路线探索和研发。2022年提出"如何实现存算一体芯片工程化和产业化"提案,并获评为中国科协"10个对产业发展具有引领作用的产业技术问题"之一;同年,联合清华大学研发业界首款基于忆阻器的110nm制程存算一体SoC芯片,单芯片忆阻器集成规模突破300万,算力能效较相同工艺下的主流GPU提升2个数量级,达业内领先水平。但从产业维度来看存算一体目前仍处于初级发展阶段,主要厂商多聚焦在芯片研发层面,而在产业链其他环节难以集中力量开展布局,导致缺乏成熟的软件工具链和模型适配算法,缺少典型性的应用引领及相关的测试验证标准平台,产业生态构建存在诸多困难。

面对存算一体产业发展的挑战,中国移动联合清华大学等高校和智存科技等头部企业,围绕软件、算法、应用环节的短板精准发力:**在软件方面**,研发了适配存算一体芯片的半自动化软件计算引擎,支持面向存算一体计算架构的AI模型编译、算子编排、模型部署和推理以及芯片性能

模拟等功能,有效降低用户开发和部署门槛,开发调试效率提升3倍以上。**在算法方面**,针对存算一体器件规模受限和模拟计算误差的问题,提出面向存算一体的模型压缩、误差补偿和加噪训练等适配算法,在保证计算准确性的同时将模型规模缩减至1/16,实现AI模型在存算一体芯片的高效、准确和鲁棒运行。**在应用验证方面**,实现存算一体芯片与自研的5G工业网关的硬件系统集成,并通过自研的软件计算引擎辅助实现了多种人工智能模型的自动部署和推理,在卫星图像分类、PCB板质检等场景完成应用验证:其中卫星图像识别准确率达96%以上,PCB板质检实现5种常见元件的精准识别和定位,并支持3种以上微小瑕疵的自动检测。

本次端对端技术验证是存算一体由技术研究迈向产业应用的关键环节,中国移动作为移动信息产业链链长,将持续发挥产业引领作用,推进产学研用合作,积极构建自主可控的存算一体新型算力生态。

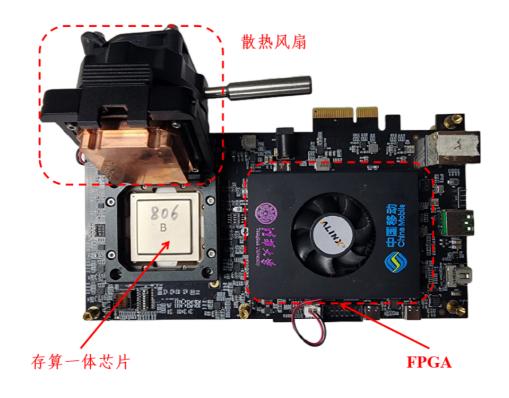


图1 忆阻器存算一体芯片A111及存算一体计算板卡



图2 基于存算一体芯片的5G工业网关



图3基于存算一体芯片的边缘智能示范演示平台