

异构计算：算力突破新抓手

本报记者 许子皓

高性能计算类应用的发展,驱动算力需求不断攀升,但目前单一计算类型和架构的处理器已经无法处理更复杂、更多的数据。数据中心如何在增强算力和性能的同时,具备应对多类型任务的处理能力,成为全球性的技术难题。在计算领域芯片企业的不断探索和研究中,异构计算成为公认的算力突破“抓手”。

芯片厂商发力异构计算

异构计算可以提高算力和性能,降低功耗和成本,又具备多类型任务的处理能力,有望主导未来的高性能计算市场。英伟达、英特尔、AMD作为行业头部企业,近两年在异构计算方面的新产品层出不穷。

英特尔在最近的投资者会议中提出了一款将x86和Xe GPU 整合到一个Xe-on插槽中的新架构Falcon Shores。英特尔中国研究院院长宋继强向《中国电子报》记者介绍,将x86的主芯片和GPU的性能整合在一起,这是一个创新。在性能上,Falcon Shores 将提供超过5倍的每瓦性能、计算密度以及内存容量和带宽。

英伟达在今年的GTC2022上,宣布推出首款面向AI基础设施和高性能计算的数据中心专属CPU——Grace CPU超级芯片。该芯片专为AI、HPC、云计算和超大规模应用而设计,由两个CPU芯片组成,两者通过NVLink-C2C进行互连。

AMD完成对赛灵思的合并后,将其CPU与赛灵思的FPGA结合为CPU+FPGA的异构模式。借助赛灵思在5G、通信、自动驾驶和行业领域的资源,AMD能够将高性能计算能力带入更多领域。

异构计算瓶颈待突破

异构的优势显而易见,与此同时暴露出的技术难题也越来越多。一是异构计算产品需要面对不同的系统架构、指令集和编程模型,需要降低多样计算带给软件开发者的难度;二是异构计算芯片产品除了要在芯片设计层面实现突破之外,还需要解决在芯片制造和封装过程中不同结构之间的适配和升级问题;三是异构计算要实现性能的多样性合一,使其同时满足人工智能训练、推理、

图像视频处理等各种不同的需求。

燧原科技创始人兼CEO赵立东在接受《中国电子报》记者采访时表示,从产业规范与标准层面看,异构计算是指在完成一个计算任务时,采用一种以上的硬件计算单元、互联协议、差异化架构、软件接口等。由于不同的硬件设备、协议、应用二进制接口、软件应用层接口等都存在着巨大的差异,如何构建一个高效、系统化的协调统一的异构计算系统,成为最关键的技术问题。

宋继强同样表示,异构计算在技术方面、互连方面和软件方面都面临瓶颈。一是技术设计流程的协同问题。要保证与其他不同厂商的芯粒一起互联互通、协同工作时不出现问题。二是需要统一各厂商芯粒之间的互连标准。目前各厂商之间的互连标准并不统一,英特尔携手业界部分企业,

推出了UCIe标准,未来有望把不同厂商的芯粒连到一起,形成一个符合要求的、大的、封装级的集成芯片。三是软件层面要过关。不同厂商的计算芯片之间要做出适配的I/O、内存通道。

“东数西算”带来新风口

当下正值我国“东数西算”工程建设稳步推进之际。宋继强指出,异构计算是“东数西算”的底层支撑。“东数西算”所需的大型计算中心内部,CPU、GPU、FPGA、DPU、专用加速器等对处理不同类型的工作数据各有优势与劣势,也有运营成本高低之分,所以未来一定要综合部署,多种架构可能都需要涵盖,而且要有有机联合起来。

赵立东认为,异构计算对于“东数西算”

专家观点

英特尔中国研究院院长宋继强：

建立更加完整的

异构计算生态至关重要

异构计算已经成为全球新的竞争点。现在主流的芯片供应商,都想把异构的布局变得更加完整。

想要在同异构竞争中脱颖而出,一是要有不同架构的积累。二是在不同的工艺节点上有不同的资源可以去利用。三是提前布局 and 规划封装级别的技术。异构需要布局架构、生产工艺以及封装技术,这都是为了建立未来更加完整的异构生态。因为生态里会包含不同架构的IP模块或芯粒模块,在不同的制程工艺上,能以最好的性价比生产

这些模块,还要能够用最标准的方法把它们串连起来。四是需要有一套方便且好用的软件,只需上层应用者指定功能需求,下层就可以随着异构变化自动迁移。

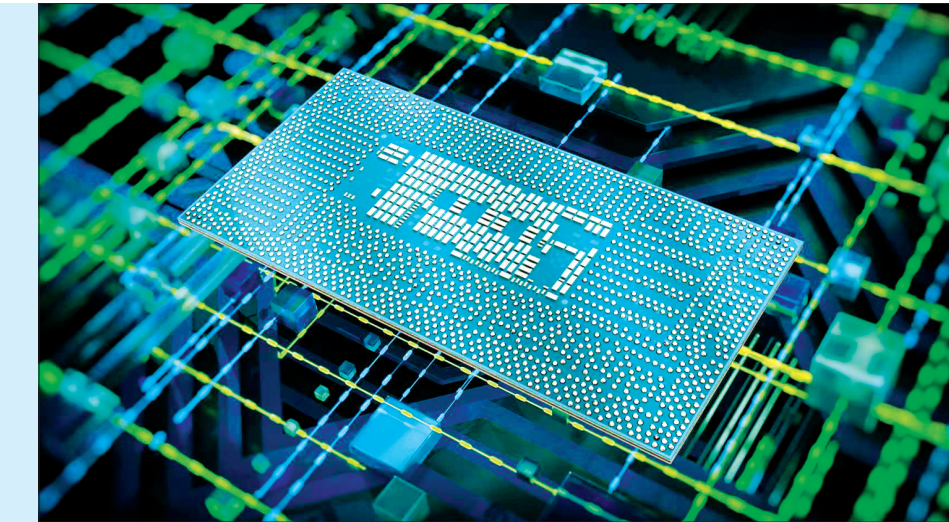
燧原科技创始人兼CEO赵立东：

异构计算将成为

国家技术竞争力

异构计算对未来以数据为核心的智能产业起到支撑和“底座”作用,加上芯片产业本身高门槛、高技术、长周期的特性,使其成为国家技术竞争力的关键组成部分之一。

当下,异构计算的竞争方向是在相应产业规范与标准指导下,以关键技术与产品创新为突破口的产业链完善以及相应的产业



来说是刚需。从半导体集成电路“赢家通吃”的发展特性来说,不管是从内部的产品业务布局,还是外部的生态与商业布局出发,单个数据中心或智算中心项目在落地时,会考虑异构计算的复杂性及发展不完备性所带来的巨大的额外部署成本,在一定程度上会偏好选择以单一厂商为核心的现成产品组合,尽量避免技术适配的投入。而置身于“东数西算”背景下,最终要实现的是全国算力中心的大一统,从而尽可能地优化利用资源,在多个数据中心或智算中心互联互通的场景下,异构计算的复杂性会是前所未有的。因此,在异构算力溯源、调度与管理平台的建设上,需要尽早布局与投入。

算力网络的构建不只依赖于异构计算,通用计算也将为“东数西算”提供更多助力。近两年,国内涌现出一批通用计算芯片

初创企业,以FPGA和ASIC类型芯片作为切入点,部分企业购买国外公司的IP授权开发通用GPU芯片,少数企业自主研发通用GPU核心IP,实现了国内通用计算的突破。

天数智芯副总裁郭为表示,通用计算所构建的算力平台通常并行处理能力特别强,计算能效比高,并且有很大的存储带宽,能够释放非常可观的算力。

郭为建议,强化顶层谋划,将算力网络上升为国家战略,从全局层面整体谋划战略定位、发展目标,加快构建新型算力网络格局。此外,应成立国家重点实验室,打造重大原创技术策源地。优化异构计算产业布局,发挥龙头领军企业的“链长”作用,协同上下游企业,制定产业链供应链图谱及协同发展计划,推动异构计算产业链本土化、供应链多元化,确保异构计算产业安全稳定。

集聚和规模化发展。

中国集成电路产业正处于蓬勃发展阶段,无论是横向的异构计算芯片种类,还是纵向的垂直赛道厂商,都可以说是百花齐放。而每一家厂商的底层硬件架构和编程模型等核心技术都各有千秋,需要更多地从规范和标准层面去引领市场的走向。

天数智芯副总裁郭为：

积极构建

国内通用计算芯片产业生态

要推动异构计算产业链本土化、供应链多元化,确保产业安全稳定。

“东数西算”中的“数”,指的是数据,“算”指的是算力,即对数据处理的能力。数

据中心包括计算、存储和网络三大元素。因此,这三大核心功能所需要的芯片,成为“东数西算”战略下产业关注的焦点,包括通用GPU在内的通用计算产品将为基础建设提供强大助推力。

我国通用计算产业还有两方面待提升:一是产业链供应链存在风险。通用GPU市场被英伟达、AMD垄断,芯片设计工具依赖国外企业,晶圆代工无法自给自足,封装基板等关键零部件依靠进口,我国通用计算芯片产业链、供应链面临受制于人的风险。

二是产业生态尚未有效构建。国内通用计算芯片起步较晚,大多是近几年才成立的创新型企 业,规模较小,实力有限,软件栈、演算法及工具链开发难度大、周期长,已经问世的计算芯片产品大多处于适配验证、批量出货阶段,有待大规模市场验证,基于自主芯片的产业生态尚未形成。

算力需求激增,天数智芯把握GPU新风口

本报记者 许子皓

近日,国内通用GPU企业天数智芯宣布,其首款7纳米通用GPU推理产品——智钺100成功点亮,天数智芯也因此成为同时拥有GPU架构下云端训练+推理完整解决方案的硬科技公司。

我国自“东数西算”这一国家级工程正式启动以来,全国上下都在努力编制这张无形的“算力之网”。这一宏观战略对于数据中心、计算、存储、数据通信、光传输等产品都有着较高建设需求,从而带动了我国计算领域的整体发展,形成了诸多新风口,其中数据中心必备的通用GPU成了国内企业争相突破的新方向。

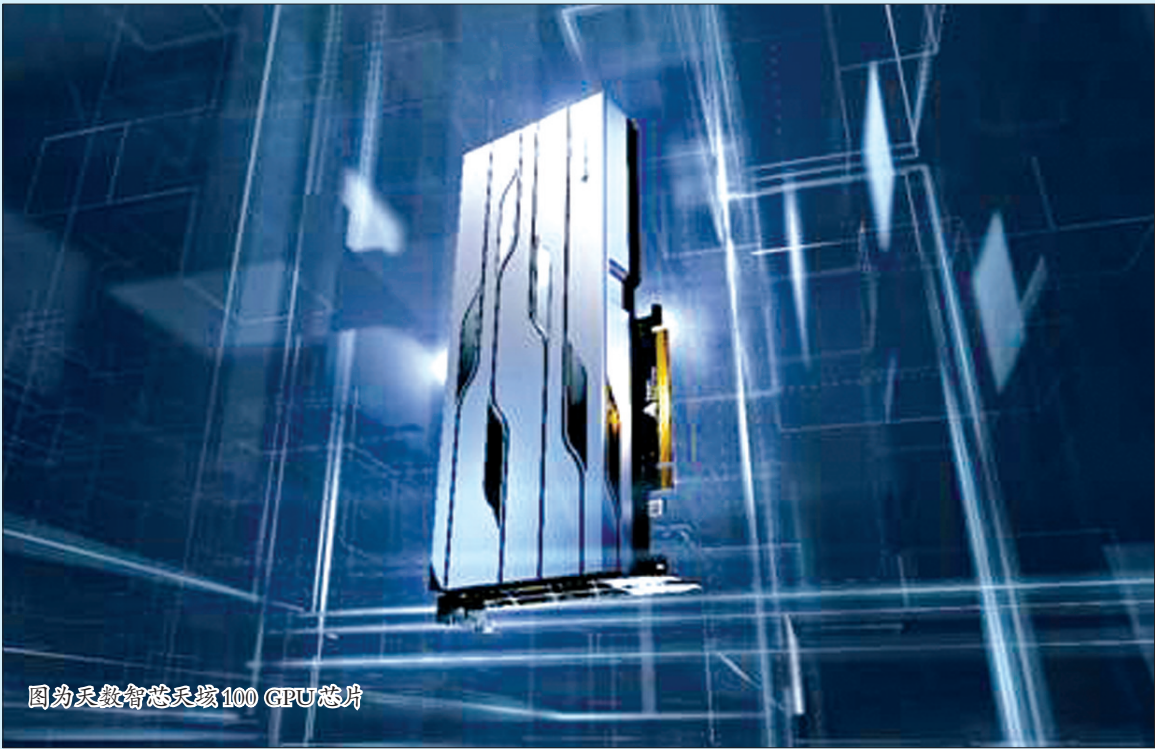
天数通用GPU全自研

助力“东数西算”

随着物联网、大数据的发展,联网设备数量不断提升,网络间产生的数据量势必将得到巨大提升,对算力的需求也将上升到全新的高度。

根据赛迪研究院测算,2020年我国通用算力规模为77EFLOPS,AI算力规模为56.23EFLOPS。预计到2025年,我国基础算力规模和AI算力规模将分别超过300EFLOPS和1800EFLOPS。

GPU芯片具有高算力、高画面处理能力的特点,应用领域广泛,全球需求量逐渐攀升。市场研究公



图为天数智芯天垓100 GPU芯片

司Jon PeddieResearch的报告显示,2021年第一季度全球GPU出货量高达1.19亿颗,同比增长38.78%。日益庞大且高速增长的GPU市场,使GPU成为了全球集成电路增速最快的三大产品之一,显示出了市场对高算力持续且高涨的需求。数据中心、PC、移动端、游戏主机构成GPU最主要的四大市场。

在全球领先的数据中心供应商戴尔、HPE、浪潮、联想、IBM中,HPE、联想、IBM已推出搭载GPU的产品。由于GPU具有良好的并行运算能力,GPU数据中心已成为AI加速方案的优选,当前市面上的GPU数据中心多配备八核GPU。据IDC预测,到2024年,中国GPU数据中心市场规模有望达到64亿

美元。

因此,国内各大厂商大力推动GPU方面的研究。天数智芯作为国内唯一实现通用GPU量产的硬科技企业,更是稳稳地抓住了此次机会,取得了突破性进展,助力“东数西算”工程实施。

“东数西算”工程的实施基础在于建立完整的算力网络,并且要保证数据传输的安全性。天数智芯从底层硬件到上层软件均独立设计开发,不走购买国外GPU IP的捷径,独立研发高性能计算IP核和系统架构,确保拥有完整的知识产权,打破了国内企业长期作为国外IP代理的局面,做到了核心IP全自研。

其旗舰产品天垓100是国内第一款自研通用GPU芯片,实现了从

0到1的突破,可以从物理底层锁定数据安全,也从根本上保障了客户使用安全、信息安全。

要想客户反馈好

实力才是硬道理

天垓100仅用一年多时间就完成了流片、点亮、测试、客户适配、稳定供货等阶段,成功量产并实现规模应用。基于独立研发的芯片架构及软件栈产品,天垓100实现了多维度技术创新,具备独立可控、高性能、通用性、灵活性等特点,其应用覆盖广、性能可预期、开发易迁移以及全栈可定制,能够支持国内外主流AI生态和各种深

度学习框架,并通过标准化的软硬件生态接口为行业解决产品使用难、开发平台迁移成本大等痛点,大幅缩短适配验证周期。

目前,天垓100已能够运行包括悟道等在内的一批新的算法大模型,还适用于2022年国际CVPR大会上的清华、北大、智源等国内研究机构联合发表的最先进算法,如定制算子的Point-BERT、量子启发的Wave-MLP、结合自注意力与卷积的ACMix等。

在过去一年里,天垓100已支撑近百个客户在人工智能领域进行超过两百个不同种类的模型训练,ResNet50、SSD、BERT等骨干网络模型的性能接近国际市场主流产品。此外,天垓100还广泛支持传统机器学习、数学运算、加解密及数字信号处理等领域,也是唯一一家已适配X86、ARM、MIPS等各种类型CPU架构的通用GPU产品。

近日,天数智芯正在陆续与国内主要数据中心厂商完成天垓100产品引入并进入其供应商目录,主流数据中心厂商近期将会陆续发布搭载天垓100的服务器产品进行销售,天数智芯已与新华三集团等行业合作伙伴达成战略合作协议,开启全方位合作。

天数智芯目前的产品广泛适用于互联网、运营商、智能安防、生物医疗、教育科研、智算中心等不同行业众多应用场景。在生物医疗领域,天垓100能够支持基因检测、X光图像识别、远程肠胃镜图像识别等应用场景;在智能安防领域,天垓100能够支持视频图像分析、监控物体识别、跨街区事件分析、物品检测等应用场景;在智慧教育领域,天垓100能够满足姿态识别及纠正、智能教学等应用场景的需要。目前,天垓100产品累计订单金额已经接近2亿元。

在国内应建立能够

深层合作的大联盟

全球GPU市场虽然目前仍被国际厂商所占据,但随着国内市场需求量不断攀升,大量的国内厂商纷纷加入战局。在国内建立完整的生态链是立足之本,这对于国内厂商来说,既是机遇也是挑战。

我国GPU产业未来该如何发展?天数智芯首席技术官吕坚平表示,一是要与国内客户合作,做出适合国内市场的DSA通用化来保持通用优势。二是以计算图形化提升通用计算效能。三是要与生态伙伴合作,借由图形计算化,让专精于张量计算的通用GPU在图形领域大显身手。

“中国需要的,或许是个在技术深度及广度上,足以与国外龙头企业相比拟,以深度合作和良性竞争为基调的大联盟。”吕坚平这样说。

天数智芯预测,图形持续走向通用,AI与图形必然融合,图形人机界面必定移动,内容生成及图形渲染必在云端。遵循图形走向通用计算的潮流趋势,绕过传统桌面显卡市场,讲求“云-移动端”协同内容生成与图形渲染,达到元宇宙、数字孪生所需的AI与图形融合。

天数智芯表示,未来将以DSA通用化、图形计算化、计算图形化、硬件微分化为目标,在天垓100基础上加速技术创新步伐,不断提升产品通用优势、图形处理能力、计算效能,以天垓系列产品加速AI计算与图形渲染融合,探索走出一条通用GPU赶超发展道路,打造世界一流的算力引擎,支持中国特有的云-移动端协同内容生成与图形渲染,推动城市智能化、工业自动化、载具无人化以及娱乐置入化,开启中国引领世界走向元宇宙、数字孪生的崭新一页。