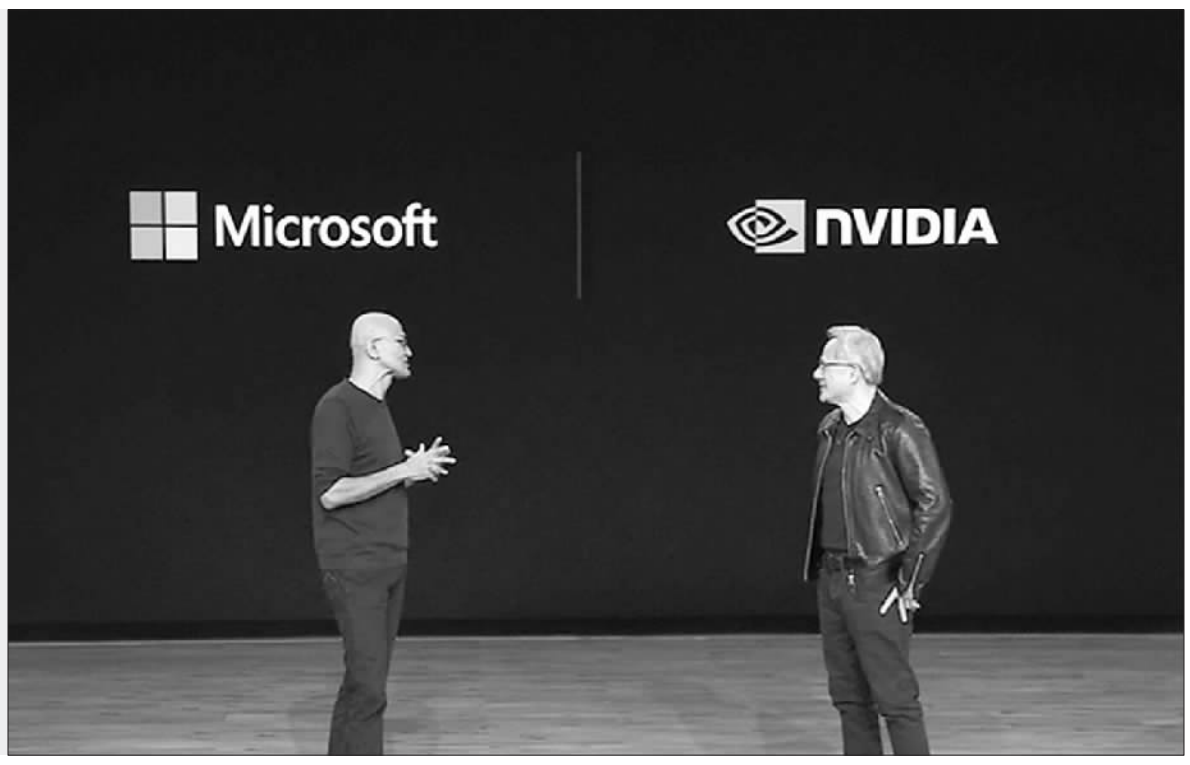
微软造芯尘埃落定

本报记者 姬晓婷 张心怡

北京时间11月16日凌晨，“微软造芯”的传言终于尘埃落定——微软CEO萨提亚·纳德拉在Ignite2023开发者大会上发布了两款芯片，一款CPU、一款AI加速器，均用于云服务，分别命名为Azure Cobalt 100和Azure Maia 100。

芯片用于云服务，切中了此前网友的猜测。而此次发布会还有两点超出了网友原本的预料：第一，没发布NPU，反而推出了一款CPU；第二，英伟达CEO黄仁勋被请到现场，大谈两家公司在AI领域的合作成果，坊间猜测的“微软挑战英伟达”说法在一定程度上被消解。

果然，科技圈没有永远的市场争夺，只有慕强是唯一真理。



微软为何要做CPU？

相比造芯，对于芯片的迭代和调优过程，也就是如何让芯片走向能用和好用，才是最考验微软的地方。

“体量够大。”这是记者提问“为什么微软要做CPU时”，业界专家给出的答案。

对于微软来说，云计算是当前最重要的业务板块，也是最大的盈利来源。在算力需求日益旺盛的当下，由购置处理器带来的成本已相当可观。自研处理器，很有可能是基于实现企业利益最大化的考量。

芯谋研究企业服务部总监王笑龙在接受《中国电子报》记者采访时表示，靠自研芯片提升企业品牌价值可能是一些小企业的思路，对于微软这样体量的企业来说，自研芯片肯定是为优化企业业务服务的。

对于云计算厂商来说，传统的服务器通常由CPU和GPU两类处理器组成。其中CPU的主要供应商是英特尔和AMD，GPU的主要供应商是英伟达和AMD。而微软此次推出的两款芯片，恰恰对应这两大类型：Azure Cobalt 100是CPU，Azure Maia 100作为一款AI加速器，主要对应的是当前GPU的功能。萨提亚·纳德拉表示，这两款产品将先自用，再逐步对外供应。

另外，现有处理器不是最适合AI的，几乎已经成为算力芯片供应商的共识。这也是在大模型浪潮的推动下，NPU、APU、TPU等AI专用处理器类型纷繁迭出的底层逻辑。在AMD、英特尔等传统处理器企业之外，芯片设计已经吸引了亚马逊、谷歌、腾讯、阿里等云服务商入场。作为服务器芯片最大的客户，以微软为代表的云服务商最了解业内对芯片的需求。

从微软的角度来看，一面是可见的日益走高的处理器购置费用，另一面是选择更适配AI硬件的需求推动。拿出一些成本，给自身业务的硬件配置提供新选择，何乐而不为？

选择做CPU，而不仅仅是做AI加速卡，恐怕也是基于未来增长可能性的考量：在自己已有相当的市场容量的基础上，采用功耗更低的Arm设计CPU，也可以在一定程度上实现风险对冲。

那么微软造芯，真能做得好吗？

所谓微软“造芯”，实际上是指入局芯片设计行业。

从芯片的全生命周期来看，芯片设计的门槛正在降低。一方面，EDA等工具链正在逐渐完善；另一方面，Arm提供的IP内核，也为芯片设计者提供了很多预设。业界专家告诉《中国电子报》记者，当前芯片设计的难度正在逐步降低，相较于五六年前已经大打折扣。相比造芯，对于芯片的迭代和调优过程，也就是如何让芯片走向能用和好用，才是最考验微软的地方。

黄仁勋为何站台？

关于微软自研AI芯片是否会抢占英伟达GPU市场的话题，引发了产业界的热议。

对于很多“蹲”微软发布会的观众来说，英伟达CEO黄仁勋的出现点燃了发布会的一波高潮。自10月微软造芯的消息传出后，网络上就流传着很多关于“微软撬动英伟达”的声音。关于微软自研AI芯片是否会抢占英伟达GPU市场的话题，引发了产业界的热议。

而黄仁勋此次到微软发布会捧场，看似使微软与英伟达进行市场争夺的猜测不攻自破，而他发布的内容，也展示了两家企业在AI业务领域的新谋划。

黄仁勋表示，企业使用AI能力主要基于三种模式：一是基于ChatGPT等公有云服务，二是基于Windows等内嵌AI应用的操作系统，三是基于自己的数据和规则定制化创建大模型。

而这第三点，正是AI未来的潜力所在。11月7日，OpenAI发布了名为GPT-4 Turbo的新模型，可支持用户实现“自定义模型”，即通过给模型专有数据，使其可以处理个别细分领域的任务。此次发布会上还发布了ChatGPT的自定义版本GPTs，支持用户的定制化需求。微软Ignite大会上，支持定制化AI服务同样是发布重点。当“人人都可训练模型”，未来的增量市场就相当可观。

接下来黄仁勋的发言，给定制化AI模型提供了一个具象化的实现方式：“我们将成为AI模型的代工厂。”

所谓“代工厂”，也就是英伟达基于微软Azure提供的生成式AI Foundry（AI代工）服务，这项服务整合了英伟达的AI基础模型、NeMo云原生框架和工具，以及NVIDIA DGX云端AI超算服务，面向企业提供创建自定义AIGC模型的端到端解决方案，支持企业定制模型以更高效地支撑AIGC应用。

在Ignite2023开发者大会现场，黄仁勋回顾了英伟达与微软团队在过去一年的合作成果，包括双方共同打造的AI超级计算机已经成为全球速度最快的AI超级计算机和全球第三快的超级计算机，以及“英伟达GPU+Windows PC”的合作模式构建了将大模型从云端推广到PC端和工作站的安装基础等。双方还透露除了AI Foundry服务，英伟达还将产业数字化平台Omniverse托管到了Azure，以及英伟达已经获得了微软AI助手Copilot的全站许可。

两大科技头部企业CEO同台，针锋相对的意思没看到，倒是看出了很多在AI浪潮推动下“慕强”的味道。

AI颠覆一切？

生成式人工智能的出现，给算力市场带来的影响是颠覆性的。算力的需求急遽攀升，供给出现严重不足。

就像黄仁勋在大会现场所说的：“过去四十年，（计算产业）还没有发生过（影响力）如此大的事情（生成式人工智能）。它比PC大，比手机还大，这是计算机行业历史上最大规模的TAM（可触达市场）扩张。”

生成式人工智能的出现，给算力市场带来的影响是颠覆性的。在生成式人工智能的推动下，算力的需求急遽攀升，供给出现严重不足。原有软硬件、云服务商等市场参与者之间泾渭分明的界限，也因为算力需求的攀升而被打破。

在满足算力需求的过程中，各种技术路线正呈现融合趋势。例如CPU和NPU正在从松散的耦合走向异构融合等。

在这样的趋势下，微软与英伟达的持续合作就变得容易理解。在AI上云的初期，微软利用英伟达GPU给Azure的虚拟机加速，英伟达通过微软Azure扩张在云计算的市场。当ChatGPT横空出世，微软与英伟达共同打造AI超级计算机探索AIGC的发展。英伟达还受邀将软件栈托管到Azure，让英伟达CUDA平台的开发者成为微软Azure的客户。本次推出的AI Foundry服务，将使双方在AIGC的合作更加紧密。

AI时代，生态为王——微软和英伟达的合作一直基于这个逻辑，从简单的软硬件结合走向了系统软件级别的合作，接下来也将向生态级别升维。在生成式AI的浪潮下，把握机会找准合作对象占领市场最为要紧，芯片供应是否存在变化，也不再是头部厂商最关注的点。

可能黄仁勋也会觉得，最终市场归谁所有，时间自会给出答案。

深圳算力微电子产业联盟启动

本报讯 日前，在第二十五届高交会高性能芯片设计与制造高峰论坛暨重大项目签约仪式上，深圳算力微电子产业联盟启动，深圳理工大学（筹）算力微电子学院也正式成立，为深圳、粤港澳大湾区乃至全国的算力微电子发展注入新动力。

据了解，深圳算力微电子产业联盟由深圳理工大学（筹）（以下简称“深理工”）牵头，联合中国科学院深圳先进技术研究院（以下简称“深圳先进院”）、海光信息、曙光信息、中科寒武纪等国内知名集成电路企业共同设立，联盟汇聚深圳、粤港澳大湾区乃至全国集成电路领域的各方力量与创新资源，以深圳集成电路与半导体产业集群建设发展的重大需求为导向，为深圳集成电路高新技术企业供给即用型人才，促进基础技术创新，通过产教融合实现关键核心技术突破。

会上发布的算力微电子（深圳）宣言指出，深圳作为电子信息产业重镇，将充分发挥完备的产业配套优势，探索构建“教育、科研、技术平台、产业孵化、产业联盟、产业基金”六位一体的生态系统。在人才培养上，创新高校芯片专业培养模式，培养贯通计算机科学与集成电路学科知识“即用实战型”拔尖创新人才；在科技创新上，集聚力量进行原创性引领性科技攻关，突破产业共性关键底层技术；在产业发展上，促进产业上下游融通，构建高效的产业生态，提升产业整体供给能力。

算力微电子学院是深理工的第7个学院，同时也是中国首个以“算力微电子”命名的学院，聘请曾经的龙芯CPU、海光CPU创始人之一的唐志敏为院长，专注于算力与微电子交叉集成发展，瞄准世界科技前沿，突破关键核心技术，培养一流科学家和卓越工程师。

此外，深圳算力微电子工业级教研与公共服务平台、算力微电子楼上楼下科创综合体也在当日揭牌，与深圳算力微电子产业联盟、深理工算力微电子学院共同助力深圳集成电路与半导体产业集群建设发展。

在重大项目签约环节，深理工与曙光信息、海光信息、中科寒武纪、朗科科技、博瑞晶芯等14家算力微电子信息相关企业签订合作协议，一起推动科技产业融合发展。（张 赢）

服务高质量发展

中国光学光电子行业协会测试共享平台升级上线

本报讯 记者**卢梦琪**报道：记者从中国光学光电子行业协会了解到，近期，中国光学光电子行业网光电测试服务资源信息共享平台（以下简称“共享平台”）网站专栏已完成换版升级。该共享平台旨在汇聚行业内测试相关资源构建信息枢纽，为供需双方构建快捷高效的信息渠道，解决企业在科技创新过程中关于测试、试验的痛点问题。

据悉，共享平台由中国光学光电子行业协会与中国电子科技集团公司第十一研究所联合发起，5家行业内具有代表性、权威性的检测机构、高校、企业作为联合创始单位共同打造，是国内光学光电子行业首个专业测试服务资源信息线上共享平台。

共享平台以专栏形式依托中国光学光电子行业网，面向全行业提供光电测试服务资源的信息共享服务，目前已汇聚中国电子科技集团公司第十一研究所、国家半导体器件质量检验检测中心、国家红外及工业电热产品质量检验检测中心、国防科技工业光电子一级计量站、宁波大学高等技术研究院、杭州远方光电信息股份有限公司等多家国内光电专业检测权威机构的数百台（套）专业仪器设备资源信息，可提供涵盖光学、激光、红外、LCD、LED等细分专业领域的各类参数测试技术服务。

中国光学光电子行业协会秘书长姚大虎表示，共享平台的建设与公益运行是协会服务行业中小企业高质量发展的重要举措之一，通过汇聚行业资源信息，构建专业信息枢纽，搭建供需沟通桥梁，坚守服务行业的理念，将为我国光学光电子行业高质量发展贡献力量。

阿里巴巴发布三款玄铁RISC-V处理器

与众多操作系统深度融合

本报讯 记者**宋婧**报道：11月21日，阿里巴巴玄铁RISC-V上新了三款处理器：首次实现AI矩阵扩展的C907、满足Vector1.0标准的C920，以及实时处理器R910。

大模型带来了AI算力的爆发，在端侧及边缘侧，业界正在探索加速计算的高性能、低功耗处理器新方案，玄铁C920应运而生。基于软硬协同优化，C920升级支持最新的Vector1.0标准，可实现更精准、稳定地分配任务，进而提升整体性能。C920较上代提升了高至3.8倍的AI性能，可跑Transformer模型，特别适合机器学习、自动驾驶等领域。

另一款新处理器玄铁C907，则首次实现了独立矩阵运算（Matrix）扩展，极大地提高了计算密度和计算并行能力，较传统方案可提速15倍。玄铁C907充分挖掘出RISC-V架构的AI原生优势，为加速运算提供了新选择。

面对高可靠、高实时性的计算新需求，新推出的玄铁R910支持Cache以及TCM存储架构，可选配快速及一致性的外设接口，在保持高可靠的基础上大幅提升系统实时性。R910可应用于存储控制、网络通信、自动驾驶等领域，平头哥日前新发布的SSD主控芯片镇岳510就采用了R910。

至此，玄铁RISC-V处理器家族更新增至9款，已广泛应用于计算机视觉、存储解决方案、工业互联、网络通信、智能设备终端等领域。

目前，玄铁RISC-V已完成与安卓、Debian、Fedora、Gentoo、Ubuntu、龙蜥、统信、openKylin、创维酷开系统、RTT等操作系统的深度融合。在日前举办的2023 RISC-V国际峰会期间，玄铁展示了由酷开科技研发的首个基于RISC-V的智能大屏WebOS“Coolita”，在TV端实现了主流网站视频播放、多屏互动、游戏、唱吧影音娱乐、钉钉在线会议等丰富功能。

联发科推出5G芯片天玑8300

支持生成式AI

本报讯 记者**沈丛**报道：11月21日，联发科在北京举办天玑8300 5G生成式AI移动芯片新品发布会。

据悉，相比较于此前联发科发布的新款旗舰芯片天玑9300，天玑8300的定位是次旗舰芯片。据了解，天玑8300与天玑9300系列芯片均采用台积电4nm工艺。

值得注意的是，联发科今年新近推出的“4+4”CPU架构首次应用于天玑8300。其中，八核CPU包含4个Cortex-A715性能核心和4个Cortex-A510能效核心，能够使其CPU峰值性能较上一代提升20%，功耗降低30%。此外，天玑8300还搭载6核GPU Mali-G615，GPU峰值性能较上一代提升60%，功耗降低55%。

联发科无线通信事业部副总经理李彦辑在会上表示，与同级别手机芯片相比，天玑8300是率先支持生成式AI的5G芯片。据了解，搭载生成式AI技术的5G芯片通过学习大量数据来自动生成文本、图像、音频等内容，最终实现人工智能化的应用和服务。

在发布会上，小米集团总裁卢伟冰表示，红米将与联发科联合定制天玑8300 Ultra，红米K70E将首发搭载天玑8300 Ultra处理器。对此，联发科官方表示，采用MediaTek天玑8300移动芯片的智能手机预计于2023年年底上市。

西门子收购Insight EDA

完善电路可靠性验证领域技术路线图

本报讯 西门子数字化工业软件近日宣布完成对Insight EDA公司的收购，后者能够为全球集成电路（IC）设计团队提供突破性的电路可靠性解决方案。

Insight EDA致力于为全球范围的客户提供模拟/混合信号和晶体管级别的定制化数字设计流程。Insight EDA能够提供高效的电路可靠性分析使用模型，发现设计特定的潜在电路可靠性故障区域，并且协助解决相关问题，助力工程师实现一次投片成功。

电路可靠性的需求在IC设计行业增长迅速，西门子的Calibre PERC致力于打造可靠性sign-off软件，可提供传统验证工具无法实现的检查功能，Insight EDA的加入将帮助西门子为芯片设计人员提供端到端的电路可靠性解决方案。

西门子数字化工业软件Calibre设计解决方案产品管理副总裁Michael Buehler-Garcia表示：“对Insight EDA的收购能够进一步完善西门子在电路可靠性验证领域的技术路线图。通过将Insight EDA工具添加到西门子的Calibre PERC产品线，设计工程师现在可以更加轻松地执行特定设计的可靠性检查和分析。”（微 文）