

Data is blindfolded, and edge AI is chasing it – April 21, 2022

数据蒙眼狂奔，边缘AI穷追不舍

日期：2022-04-22

来源：与非网eefocus

关键词：**人工智能边缘计算 云端AI**

每一次技术浪潮，
都会产生新的领导型公司。
边缘AI会吗？

随着**人工智能**从云端向边缘扩展，**边缘计算**被视为下一个AI战场。海量的应用场景、庞大的计算需求，不仅吸引了英特尔、英伟达这些巨头加速完善云边端一体化的布局，更吸引了众多AI芯片公司纷纷入局。**云端AI**已经造就了英伟达的巨大成功，如果边缘AI是一次崭新的机会，哪些公司有机会成为赢家？

最适合的芯片路线？

业界对于通用、专用的计算路线讨论已久，与这个话题伴生的其实是多样、碎片的应用场景。场景是碎片的，产品是碎片的，那么芯片呢？也只能是碎片的、专用的，不能是通用的吗？通用计算老大哥CPU、GPU的地位会旁落吗？未来通用计算和专用计算的关系如何发展？

时擎科技总裁于欣告诉<与非网>，针对不同应用的架构设计，是一定需要的。领域专用架构（DSA）处理器和芯片，本质上就是要解决通用性和专用性平衡和取舍的问题。

“这里有两个大前提，一是端侧的应用比较碎片化，第二是对功耗和成本往往有比较高的要求。在这两个前提下，如何既能保证相对于某个场景足够的竞争力，以满足成本和功耗的要求，同时又能兼顾足够的市场空间——这是每一家公司都要面临的挑战，也是对产品定义能力的考验”，他强调。

通用计算芯片固然能够覆盖边缘计算程序所需要的所有运算操作，但在芯片架构伸缩性、性能方面，确实无法及时适应边缘侧需求的快速增长。通用计算和专用计算芯片已经呈现出融合发展的趋势。并且，计算特性决定了边缘芯片和云端芯片的不同，架构设计需要进行优化定制。

灵汐科技副总经理华宝洪表示，二者要相辅相成，在专用计算芯片中会包括通用计算的核，比如Arm或RISC-V等IP内核。新型架构芯片比如神经拟态芯片中，除了包括神经拟态计算核、神经网络计算核，也

可以有通用的ARM核。同时，通用计算芯片比如最新的Arm芯片，也会内置部分针对传统神经网络加速的IP核。

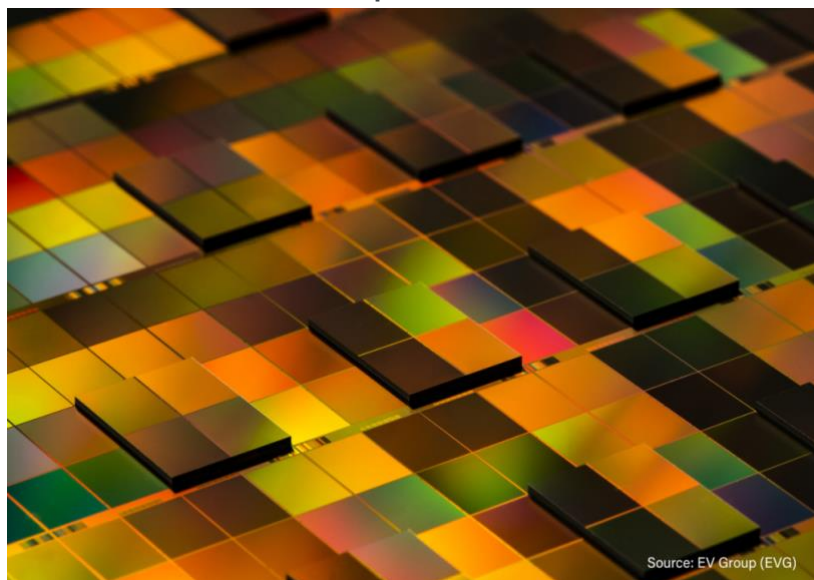
伴随这一趋势，意味着需要把负责加速的专用计算单元移入通用编程模型中，且创建通用处理器的压力始终存在。Imagination计算业务产品管理总监Rob Fisher表示，这主要是从通用处理器编程的易用性角度来考虑的。当任务的规模或所需性能远远超出通用解决方案所能容纳的范围时，这种模式将受到限制。

他指出，GPU就是一个很好的示例。在实际应用中，将图形处理工作负载卸载到GPU上的优势很明显，从而推动了高效图形处理器的独立开发。矢量处理器正越来越多地与CPU架构紧密结合，从而允许对计算任务进行指令级加速。

雪湖科技副总裁赵小吾表示，针对不同场景下功能和性能的不同要求，边缘的需求更复杂，很难用一个通用的架构或者平台来满足大多数的需求，所以会针对不同应用场景进行专门的架构设计。性能要求不高，算法变化快的部分可以使用通用计算芯片，比如CPU；性能要求高，算法相对固化的部分可以使用专用计算芯片，比如ASIC；对性能和算法灵活性都有一定要求的部分可以使用FPGA可编程芯片。

他以智慧交通领域的边缘计算为例谈到，由于基本都是室外场景，环境复杂且恶劣，既要满足AI大算力和低延时，又要满足可靠性和稳定运行，因此目前大部分边缘计算机都无法满足需求。而采用了室外基站同等级的芯片为这种场景定制了专用的大算力计算机，才满足了这些特殊场景的需求。

Chiplet正当红



图源 | EV Group

随着高性能计算和机器学习的兴起，异构处理器必须处理的工作负载急剧增加，在整个半导体行业建立开放的生态合作至关重要。

不久前，英特尔、AMD、Arm、高通、台积电等已经联合成立了Chiplet标准联盟，推出了通用Chiplet的高速互联标准UCIe。在UCIe的框架下，互联接口标准得到统一，各类不同工艺、不同功能的Chiplet芯片，有望通过2D、2.5D、3D等各种封装方式整合在一起，多种形态的处理引擎将共同组成超大规模的复杂芯片系统。

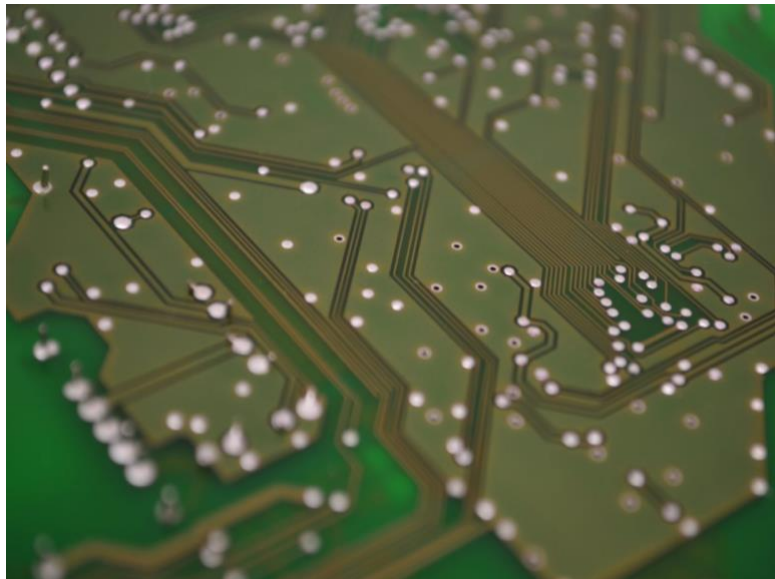
英伟达在上个月的GTC22上，一方面宣布了对UCIe规范的支持，另一方面，宣布为半定制芯片开放其NVLink-C2C互连技术，这是一种支持内存一致性的芯片到芯片、裸片到裸片的互连技术。这一路线已经明确展现了英伟达的异构决心，按照这一规划，理论上甚至可以将英伟达的芯片与竞争对手的芯片放入同一个封装中。

黄仁勋告诉<与非网>，他**第一喜欢PCIe，第二喜欢UCIe，并且预测五年内UCIe的好处会逐渐显现。**至于英伟达自身的NVlink互连技术，他强调优势在于直连能力。UCIe不能直接接入芯片，仍然是一个外设接口；而NVlink的优势在于可以直接连接，几乎就像直接连接到大脑一样。一定程度上，这可能会导致它的组装比较复杂，合作伙伴和客户必须非常了解NVlink。不过，一旦他们能做好这一点，就可以充分利用芯片内部的所有资源，就像这些资源都在同一个芯片上一样。

这一解答既表明了英伟达并不打算自我排除在UCIe联盟之外，同时也展现出了对自身NVLink互连技术的绝对信心，推测**该技术也将成为英伟达构建异构生态的关键。**

新一代“英特尔”、“英伟达”

能否在边缘市场诞生？



图源 | redherring.com

边缘计算市场的巨大潜力，自然也吸引着云端芯片巨头的竞逐，他们正通过异构计算、先进制程、先进封装等方式进行全面布局，加之高筑的生态壁垒，国内AI芯片厂商是否有机会与之一搏？

“能造得了摩天大楼的，不一定擅长雕梁画柱。当然相对于云端目前高度垄断和集中的格局来说，边缘侧还没有确定的格局，大家都有机会，而具有更强技术能力和落地能力的厂家，会有更大的机会在竞争中脱颖而出”，时擎科技于欣表示，“云边端在某些场景融合协同是有道理的，但本身从芯片设计的角度来说，还是会有很大的区别。”

灵汐科技华宝洪则认为，异构计算、先进制程、先进封装等均是手段，不能从根本上解决高能效比、小样本学习、在线学习等问题。在行业导向和市场需求的驱动下，**计算机科学和脑科学的异构融合，将成为驱动边缘计算技术架构创新与未来行业高质量发展的核心引擎。**

一方面，冯·诺依曼架构的芯片依然在走“暴力计算”的美学方向，采用更加先进制程、更加先进封装来提升算力；另一方面，非冯·诺依曼架构优先通过架构创新来满足在生物神经网络、类脑方向和新型混合神经网络等领域的规模化使用。以类脑计算、量子计算为代表的新型计算架构将与传统计算架构深度融合，引领新一轮的技术变革。

雪湖科技赵小吾表示，业界头部厂商都开始用小芯片拼凑大芯片的方式来完成产品布局，以满足不同场景的算力需求。例如苹果和英伟达，都开始采用这种“拼积木”的方式，这是非常明确的趋势。

国内这两年市场很热发展很快，但形成规模和有竞争力的厂商并不多。“芯片还是一个需要积累的产业，且产业链比较长”，赵小吾表示，“**目前国内小而多的形态不利于去和上下游争夺话语权，预计未来1-2年里应该会迎来一波AI芯片厂商淘汰潮。**”

爱芯元智创始人、董事长兼CEO仇肖莘也表达了类似观点，她谈到，现在对国产芯片产业是千载难逢的机会，由于市场需求和国家支持，涌现了许多创业型公司。而从大环境出发，我国的芯片行业还处于初期阶段，正呈现出一种百花齐放的态势，但随着产业的不断发展壮大，随之而来的行业整合也将是必经过程。

她强调，这符合芯片行业在过去几十年发展的规律，**经过这样的整合，业内一定会出现头部企业，这对于国家整体产业发展是非常重要的。**只有这样，中国企业和国际大厂才有同台竞技的机会。

灵汐科技华宝洪表示，**边缘AI芯片市场仍处于开放状态，没有绝对的霸主。**新兴的、多样化的应用场景为国产AI芯片带来了巨大的市场机会，特别是在自动驾驶、智能安防、智能物联网和可穿戴设备等越来越碎片化的市场，国产AI芯片厂商和国际巨头芯片厂商是站在同一起跑线上的，甚至在某些领域更有优势。“国际巨头的AI布局大而全，是优势也是劣势，国内处理器厂商聚焦在某些细分领域进行持续的创新和突破，构建自己的护城河是关键”，他强调。

写在最后

就像计算架构领域一位著名的科学家所说，现在正是芯片体系架构百花齐放的黄金时代，这一时代前所未有。尽管CPU、GPU会持续创新，且在某些计算任务是不可或缺的，但是AI加速计算、数据爆炸等趋势催生的新市场，一定是巨大且多样的，这就给了AI芯片公司带来了新的机会。

从CPU来看，x86架构统治PC和服务器已有松动迹象，Arm一路从手机端和IoT逐渐向上，攻入PC和服务器领域。RISC-V也从物联网设备起步，向更大量的设备进行布局。冯·诺依曼架构和非冯·诺依曼架构的异构融合，正在通往规模化的道路上.....

每一次技术浪潮，都会产生新的领导型公司。边缘AI会吗？



版权声明：本站内容除特别声明的原创文章之外，转载内容只为传递更多信息，并不代表本网站赞同其观点。转载的所有文章、图片、音/视频文件等资料的版权归版权所有人所有。本站采用的非本站原创文章及图片等内容无法一一联系确认版权者。如涉及作品内容、版权和其它问题，请及时通过电子邮件或电话通知我们，以便迅速采取适当措施，避免给双方造成不必要的经济损失。联系电话：010-82306116；邮箱：aet@chinaaet.com。

<http://www.chinaaet.com/article/3000148674>