

# 算力革命将深刻影响芯片未来发展方向

本报记者 张依依

## 数据结构变化是 芯片行业最大驱动力

**胡春民:**近两年,芯片供应短缺现象很严重。2021年即将翻页,你认为明年芯片短缺问题会得到缓解吗?

**吴雄昂:**芯片供应短缺现象使得芯片价格出现上涨,不过明年这一现象可能会有所缓解。具体来说,明年,某些领域的芯片短缺现象可能会慢慢舒缓,比如汽车芯片的缺货状况就不会那么严重了,因为马来西亚等地的工厂已经恢复运行。当然,某些技术节点的芯片短缺现象可能还是要持续到明年年底,甚至到后年年底,但紧缺程度肯定会有所缓解。

**胡春民:**明年芯片产业的市场热点有哪些?

**吴雄昂:**虽然产业中不断涌现出新概念,但市场需求非常高涨的行业还是目前的5G、人工智能、物联网和新能源汽车等领域。

**胡春民:**手机、PC已经进入存量时代。人工智能、物联网、5G甚至新能源汽车领域的兴起,对芯片产业的发展是不是一个机会?

**吴雄昂:**回顾整个芯片或者IT革命的历史,我们会发现,很多行业的发展都是通过芯片这一载体,把算力和数据结合起来,通过自动化来提高效率,提升用户体验。

现阶段,智能化时代的大幕已经拉开,设备多元、数据爆炸、应用多元是这个时代的特征。智能摄像头等各种设备会产生大量数据,通过手机、PC办公也会产生大量数据。这些数据不仅量大,产生后还需要进行实时处理。在这种情况下,制程、架构、软件、硬件等都需要做出改变,才能将算力提升10~100倍,实现完全的自动化、智能化。

以物联网领域为例,近年来,物联网设备的智能化程度越来越高,而这种智能化程度的提升其实是基于人工智能计算场景的极大丰富。

在自动驾驶领域,摄像头的智能化程度已经非常高了,并且成本也下降得非常快。除摄像头外,自动驾驶领域还有很多新型传感器和激光雷达等产品搭载在车上,会产生各种各样的新数据流。一个智能手机,可能一个月内只能产生5G~10G的流量,但是一辆L4级别的自动驾驶车辆在路上仅仅行驶1个小时,就可以产生1T左右的巨大数据量。这些产生的数据需要通过5G或者传感器等传输到后端进行处理,所以芯片行业发展的最大驱动力之一是来自于整个数据结构的变化。

**胡春民:**2021年是元宇宙大火的一年。您对元宇宙概念有什么看法?

**吴雄昂:**我认为元宇宙的一大特点是互动性。在一个平行的数字世界里面,我们可以复制现有环境,在这个数字世界中完成我们在现实中做得到和做不到的事情。

如果说物联网时代是通过各种传感器和芯片,把现有世界的一部分感知进行数据化,那么元宇宙则是将“数据化”更全面地铺开,也就是把整个世界“数据化”,然后在这个基础上,通过“数据化”创造一个平行世界。在这个平行世界里,实现更深度的交流。

未来在元宇宙的世界中,我们将在“数据化”的进程中越走越远,产生越来越多的数据。这种数据化进程的加快将对算力与芯片提出越来越多的要求。

这些要求非常多,但可以分几个层次。第一个层次是架构。在数据化进程中是不是需要新的架构,新的架构是什么,该如何解决进程中产生的种种问题。

第二个层次是制造。也就是说,是不是要研发出新的制造技术,因为无论是芯片制造技术还是先进封装技术,实际上都需要进一步提升。

第三个层次是软件。如果没有软件对数据进行高效优化,产业整体的算力提升会非常困难。在加大算力提升方面,软件优化是其中最重要的环节之一。

**胡春民:**中国数字经济正在高速发展,你对这一波数字化进程有什么建议?

**吴雄昂:**回顾过往,国内整个数字产业的发展经历了几个阶段,第一个阶段是学习能力建设阶段。这个时期,计算架构比较单一,主要是以学习和模仿国外的计算架构为主。但在这一阶段,国内在商业模式上已有所创新。

第二个阶段,以华为服务器的出现和5G应用场景拓展等为代表,我们看到了越来越

**对话人:**安谋科技董事长兼CEO 吴雄昂

中国电子报总编辑 胡春民

**对话时间:**2021年12月17日

**对话地点:**北京赛迪大厦



多计算架构方面的创新。可以说,国内产业发展到现在,在设计方面与国外差距不大,但在设计的具体细节和一些细分领域还有提升的空间,需要通过快速迭代来解决这些问题。

所以,国内企业要继续完善产业配套,在实现核心技术架构进步的同时,还要不断推动配套工艺和软件的发展。此外,要积极推动大中小企业的发展,在企业层面实现“百花齐放”,产业发展既需要大企业起到带动作用,也需要小企业发挥助力作用。

## 未来芯片架构 将呈现多元格局

**胡春民:**在算力革命下,很多产业将出现新的趋势和特征。未来新计算的发展方向是什么,计算主流架构的发展情况如何?

**吴雄昂:**根据以往经验,我们看到的肯定是多样化架构不断涌现,而不是单一化的计算架构。为什么这么说?因为在大量数据的产生、传输与处理过程中,实际上对性能、功耗和成本的要求非常多样化,也就是说不同的领域可能会用到不同的架构。

比如,现在GPU很火,但GPU一开始的功能主要是为了游戏。随着GPU通用化以后,算力的特性就发挥出来了。但是说到自动驾驶领域的GPU,至少在国内市场英伟达还是比较主流的,它是以GPU为主的算力结构。特斯拉作为最先进的自动驾驶公司之一,却使用NPU架构来实现自动驾驶功能,在某种程度上牺牲了一定的通用性,换取了性能上的提升。

所以,针对架构技术,我们不可一概而论哪个架构一定会赢。在架构多样化的过程中,我们要针对某些特定的应用场景,用最优的架构去解决其中的问题。如果某项技术市场中大量应用、成熟落地,那项技术一定是性价比最优的,而不一定是性能最好的。

**胡春民:**x86、Arm架构和最近火热的RISC-V架构可以说是目前市场中存在的三大架构。尤其是在中国,很多企业都在尝试使用RISC-V架构。你怎么看待这三大架构的未来发展格局?RISC-V的出现对Arm架构是否是一种冲击?

**吴雄昂:**时代的演进需要不同类型的计算。从PC时代的CPU到移动互联网时代的CPU+GPU,再到以NPU为首的全新算力单元开始在数据驱动的智能计算时代崭露头角,架构的演进总是推动更有效率的架构逐步取代通用性架构。

对于任何一个成功的架构而言,生态规模非常重要。目前,围绕Arm架构的生态系统中,有全球合作伙伴1000多家,中国250多家;全球出货量累计已经突破2000亿,其

中国本土出货量超过250亿。在生态系统的规模方面,Arm架构的优势还是非常明显的,特别是在操作系统和高性能方面。

更重要的是,未来计算更深层次的挑战是以海量的智能数据流为主,而应对这一挑战的主力将不再是CPU而是NPU。因此,今年7月,安谋科技联合产业龙头企业和组织发起成立了“智能计算产业技术创新联合体”(ONIA),共同打造全球首个开源的神经网络处理器指令集架构(NPU ISA),同心协力建设“中国首发,全球开源”的NPU生态系统。我们相信ONIA以及这一开源NPU ISA将在未来很长一段时间内,在构建由中国本土发起、以全球领先技术为标准的智能计算产业生态过程中,在实现NPU处理器创新和智能计算的持续演进方面发挥重要的作用。

**胡春民:**计算架构百花齐放,软硬件协同发展同样是整个信息社会发展的主流趋势。谷歌、英特尔都认为软硬件协同是新时代下的必然选择,英伟达甚至想成为一家软件公司。对此您怎么看?安谋科技在软硬件协同创新方面又是怎么做的?

**吴雄昂:**软硬件结合是必然趋势。在算力提升时,仅靠架构和芯片,软件却跟不上,一样很难实现性能提升。

安谋科技目前有很多工程师在进行各种软件研发工作,这些工作覆盖从开源软件到服务器的软件支持。我们发起成立的开源联盟ONIA也是为了让大家进行软件资源投入,包括开源工具链与开源社区支持。

除了软硬件结合这个趋势,现在芯片、系统和软件厂商的业务界限越来越模糊。随着新的数据出现,计算架构一直在演变,算力需求无法满足,所以系统厂商想做自己的芯片,针对自己的负载或数据流来做优化。反过来说,芯片厂商会发现所生产的芯片算力无法得到充分发挥,是因为软件性能不够优良,所以只能着手自己做软件。

技术范式的转移会造成双方的差异化。未来软硬件肯定是并行模式,但没有软件就好像汽车没有发动机,算力的提升必须依靠芯片加软件的优化。

**胡春民:**现阶段,智能手机市场趋于饱和,AI等新兴领域的发展引发了新的技术革命。您认为未来AI领域的发展热点是什么?安谋科技又将在其中扮演什么角色?

**吴雄昂:**新领域的发展对算力与芯片的要求会越来越多。正如前面所说,我们需要研发新架构来解决算力提升过程中出现的各种问题;要推出新的制造技术,让芯片制造和先进封装技术都实现进一步发展;还需要研发出性能更优秀的软件来适配硬件,以满足算力需求。

总体来说,Arm是算力革命过程中的重要一员,未来也会扮演一个重要角色。因为

就像x86代表了计算架构的一个体系一样,Arm代表的也是一个体系,我们做的东西就好像“跑车的引擎”。

下一个阶段,我们可能就是一个电动机而不是内燃机了。

## Arm生态体系 助力中国产业发展

**胡春民:**Arm将英国技术插到了美国的创新土壤中,通过开放授权的方式长成参天大树。Arm在中国生根、壮大,与中国伙伴是如何协同成长的?

**吴雄昂:**安谋科技源于Arm,但是已经生根于中国,因此从合资公司成立的第一天起,我们就持续发力自主可控的本土研发业务,我们要带着Arm的科技养分开把安谋科技种成一颗本土的科技大树。

现阶段,我们通过开源这种新模式和中国产业进行联合。回顾过往,我们在2006年之前是通过单个产品打入市场,实现单点突破的。不管是服务行业头部企业,还是服务一些体量较小的创业公司,我们都通过平台化的SoC构建自己的计算平台能力,再加上软件能力,不断进入各个垂直市场,为中国产业提供技术支持。

2020年,在网络处理器、高性能服务器方面,中国合作伙伴的出货量全球第一。Arm生态在高性能服务器领域对中国整个产业提供了非常多的帮助。

**胡春民:**Arm在开源社区建设方面做了非常多的工作。Linaro、TARS、欧拉和龙蜥等开源社区都能看到Arm作为引领者的身影。在国内外开源社区建设中,Arm更像一个桥梁,把国外优秀的开源项目带到国内落地,同时把中国的开源场景带到国外。安谋科技作为脱胎于Arm的中外合资公司,您是否愿意分享一下,安谋科技针对开源社区有什么发展计划?未来将怎样更好地推动开源生态系统的建设?

**吴雄昂:**多年来,安谋科技一直通过安创空间、安创加速器 etc 等创新服务平台,致力于帮助国内技术驱动型创新创业企业精准对接生态资源、投资机构、销售渠道、宣传渠道等,提供一站式深度加速服务,扶持本土企业成长。

在智能计算时代里,随着算力越来越多地由CPU、GPU转为由NPU来提供,在NPU这个架构上,今年7月,安谋科技联合产业龙头企业和组织发起成立了开源联盟ONIA,并宣布全球首个开源的NPU ISA,真正做到了“中国首发,全球开源”。安谋科技通过共建共享、开放创新的模式,希望形成产学研进的联动,推动自主核心技术的落地与演进。

此外,安谋科技还为今年所有加入

ONIA的会员提供“周易”NPU的免费授权,并且联合安创加速器,孵化10家基于开源架构和XPU的创业企业,为他们提供技术支持、产业对接、场地、资金投入等。目前,安谋科技已经成立了一个研发团队,把我们的工具链及软件开发做成开源模式,这样,我们的芯片合作伙伴可以减少很多开发成本与时间,对上游的应用厂商或者算法厂商来说,也会减少很多与硬件适配的时间,加快研发速度、减少研发障碍。在这个过程中,安谋科技其实在扮演一个“抛砖引玉”的角色。

未来,我们的软件将继续支持开源体系,这个体系涉及各种各样的基础软件。现阶段,我们有10多个工程师提供各种各样的开源软件支持,为开源社区的构建、开源模式的推广保驾护航。

## “A+X 双轮驱动”战略 建立生态大树

**胡春民:**安谋科技今年推出“A+X 双轮驱动”战略,如何解读?XPU融合计算平台有哪些优势?

**吴雄昂:**今年8月,安谋科技推出新业务战略,也就是兼容Arm架构的CPU生态系统和自主架构XPU生态系统的“双轮驱动”战略。过去,我们把Arm这一全球标准和技术带到了国内,赋能了中国半导体产业,也打造了完整的生态链;将来,我们不仅要在CPU方面继续开发兼容Arm架构的CPU,更重要的是,在新架构方面,我们开发了自主的XPU架构,以XPU自主架构为核心,我们将打造下一代融合计算平台。

对安谋科技来说,在多元化的基础上,“X”还有三方面含义:一是可定制化、多样化的结构,包括多域的支持;二是可支持多样化算力的扩展堆叠;三是针对各个场景更精准的优化和效能的提升。过去,从CPU到NPU、VPU再到ISP,每个计算单元都是独立的,合作伙伴往往需要从多个供应商那里把这些IP根据场景重新定义、二次开发、裁剪和优化,在新的高密度数据流计算环境当中,这是一个巨大的挑战。针对这样的挑战,我们根据场景,重新定义融合计算架构、指令集、处理器IP、系统软件、设计服务等,优化数据搬运,降低功耗,提升效能,给合作伙伴提供一个完整的交付方案,XPU正是在此背景下应运而生的。

基于超域架构(xDSA)的XPU可以针对不同应用,将处理AI、视频、图像等功能的计算单元组成不同的解决方案,打造融合计算架构平台,以解决海量智能数据流的处理效率和功耗问题,满足客户多元化的需求。

**胡春民:**“核芯动力”的理念有什么具体内涵?对于其他企业有何可借鉴之处?

**吴雄昂:**选择此时推出“核芯动力”新业务品牌,是安谋科技在洞察智能计算时代发生的深刻变革之后做出的重大决策,以应对未来多样化算力堆叠和多计算域的需求。“核芯动力”的推出,就是为了打造开放的智能数据流融合计算平台,更好地以定制化的方式来解决不同行业的痛点。它的logo看起来像一个涡轮发动机,象征着为中国芯片产业持续提供源源动力。从品牌理念上来讲,也代表了创新(Innovation)、智能(Intelligence)以及无限可能(Infinity)。

“核芯动力”的发布,体现了安谋科技“承前启后”的战略定位。承,是对Arm CPU生态的传承;启,是开启自主架构XPU新时代的使命。我们希望通过强大的自研XPU产品和专业的一站式服务,赋能中国客户和产业生态,在已有成熟繁荣生态的基础上,携手产业合作伙伴,打造领先的、共赢的XPU智能计算生态。

**胡春民:**未来的芯片和供应链正在向区域化方向发展,中国现在是否有必要自主研发类似的核心系统架构?

**吴雄昂:**不管架构如何演进,我们的目的都是要掌握核心技术。为什么要采取“双轮驱动”策略?因为从CPU角度出发,其将涉及全球安卓软件的兼容性,而Arm架构在CPU和控制层上的优势,是全球领先的,所以在CPU这块,安谋科技会继续开发兼容Arm架构的CPU;此外,随着AI、5G、IoT等技术的快速发展,单靠CPU已经无法满足海量、多元化的计算需求,所以我们“双轮”的另一个轮子,就是开发自主的XPU架构,并以此为核心打造下一代融合计算平台。这就是“存量国际兼容,增量以我为主”。在中国这个半导体产业发展最快的市场上,如何以有效且符合中国速度的方式打造这个生态?现在要补足的是生态架构的领先性。在下一代XPU新架构生态上开放合作,争取中国首发,全球开源。