

数据中心“第三颗主力芯片” DPU 打响卡位战

本报记者 张心怡

面对后摩尔时代和异构计算时代的需求,DPU(数据处理器)应运而生,并迅速点燃了芯片产业的热情。近日,英伟达宣布其DPU将服务器和存储之间的IOPS(每秒进行读写操作的次数)性能提升至4100万以上,超过之前世界纪录的4倍,再次提升了数据中心存储访问速度的上限。在投融资领域,国内DPU芯片设计企业中科驭数在2021年12月宣布完成数亿元规模A+轮融资。芯启源、星云智能等国内DPU企业,也在2021年获得数亿元融资。在国际市场,英伟达、英特尔等企业正在积极卡位DPU产业。在国内市场,广阔的应用空间为DPU提供了良好的发展条件。

先进计算产业的新机遇

随着摩尔定律放缓,通用CPU性能增长的边际成本迅速上升。数据显示,CPU的性能年化增长为3%左右。而IDC数据显示,全球数据量在过去10年的年复合增长率接近50%,每四个月对于算力的需求就会翻一倍。

作为信息时代的算力基础设施,数据中心正在面临极大的算力和带宽挑战。在这种背景下,DPU应运而生。

DPU最直接的效果就是为CPU减负,承担网络虚拟化、IO虚拟化、存储虚拟化等占用CPU算力但并非直接用于计算的基础设施功能,从而将CPU的算力释放给应用程序。

由于DPU是从智能网卡发展而来,因而更强调网络处理能力。相比CPU间接

支持网络IO,DPU能够直接与网络连接,其IO带宽可以与网络带宽等同。这就意味着CPU用于网络协议处理的算力也可以卸载到DPU上。

芯启源电子科技有限公司战略发展副总裁王鹏向《中国电子报》记者指出,DPU的本质作用在于承载网络侧专用性的网络堆栈算法和传输协议运算转移,核心效用在于释放CPU算力资源,助力其他计算模块高效处理业务数据。未来,DPU将致力于解决网络协议处理、数据安全、算法加速等问题,而这些问题有着“CPU做不好,GPU做不了”的特点。

“DPU能够更好地执行网络传输的协议栈,同时降低CPU负荷,让CPU更有效地处理业务数据。CPU、GPU和DPU的

DPU能够更好地执行网络传输的协议栈,同时降低CPU负荷,让CPU更有效地处理业务数据。

组合是协同处理的下一个飞跃,它将利用革命性的硬件加速技术和软件定义的可编程技术,来应对以数据为中心和边缘计算架构的挑战,开启先进计算产业新的发展机遇。”王鹏告诉记者。

作为面向数据的专用处理器,DPU也是异构计算的一块新拼图,推动数据中心加速芯片进一步向专业化发展。

“DPU的出现是算力从通用走向专用异构计算的一个阶段性标志,它采用软件定义技术路线支撑基础设施层的资源池化,可以更加高效地处理基础设施层的网络、存储、安全、服务质量管理等服务,被业界定位为数据中心中继CPU和GPU之后的‘第三颗主力芯片’。”中科驭数高级副总裁张宇向《中国电子报》记者表示。

目前国际厂商已经在着手布局DPU市场,加速建立技术标准,构建各自的产品生态。

持规模化的云端处理能力。

在英特尔IPU背后,是面向异构计算的ONE API软件框架。ONE API为开发者提供了统一的、基于标准的编程模型。开发者也可以通过ONE API的性能库和移植、调试工具,将在其他软件平台上的代码移植到ONE API。该软件框架有助于开发者开发运行在IPU上的应用,也能让IPU更好地融入英特尔的异构生态。

在部署软硬件的基础上,国际大厂也积极开展生态建设。英伟达宣布华硕、戴尔、富士通、技嘉、浪潮、联想等厂商已经计划集成其DPU。英特尔也在2021年12月宣布与浪潮、锐捷网络、Silicom携手合作,共同设计并开发全新的IPU解决方案。

国内厂商需要尽早在目标领域进行技术应用落地和产品服务打磨,打造完备的DPU应用生态圈。

外厂商都处于起步阶段。国内市场需求体量巨大,而DPU芯片的设计又偏重于应用驱动的体系结构。因此,国内DPU的发展正面临可观的市场机遇,厂商需要尽早在目标领域进行技术应用落地和产品服务打磨,打造完备的DPU应用生态圈。”张宇说。据悉,中科驭数正在积极拓展DPU在多个行业的应用,为电信、金融、能源、互联网等行业的数据中心研发DPU和智能网卡产品,以及基于DPU技术加速云原生生态的解决方案,已经开展在5G MEC和边缘云场景的布局。

王鹏表示,芯启源将着力推进基于自研DPU芯片的智能网卡等系列产品在云计算、5G和6G、金融、交通等行业场景下的应用拓展,并加快第二代DPU芯片的研发和产业化工作,在通信与计算融合发展的创新节点上,与上下游合作伙伴一起走出具有中国特色的DPU发展之路。

“我国应该充分利用DPU这一换道机会,树立国产品牌,积极参与制定产业技术标准,加紧研发和生态布局,打造自主产业链。新一代DPU核心芯片完全有机会打破国际企业在计算芯片领域的行业垄断。”王鹏表示。

自动驾驶感知精度愈发关键 毫米波雷达担当重任

本报记者 张依依

全自动驾驶时代已经微露曙光,人们对自动驾驶的安全性越发关注。现阶段,高级驾驶辅助系统(ADAS)正在成为汽车的标准配置,ADAS技术应用在车辆中的成本也在逐渐降低,开始在众多车型中成为主流。在这一背景下,利用毫米波雷达提升车辆的感知能力,为车辆安装“敏锐”的眼睛就显得十分关键。基于此,很多公司都开始在毫米波雷达领域展开布局。

自动驾驶对感知提出更高要求

在刚刚过去的2021年,大量汽车厂商都推出了自己的自动驾驶和辅助驾驶车型,汽车市场出现了百家争鸣、百花齐放的局面。

自动驾驶可以分为三个部分。德州仪器(TI)中国区汽车业务部总经理蔡征向《中国电子报》记者表示,第一个部分是感知,即通过雷达和传感器技术去感应外围道路的情况;第二个部分是决策,即通过处理器对数据的分析,对自动驾驶的决策机制给出控制策略;第三个部分是通过主机控制汽车,完成整个车体的控制,起到规避风险和辅助驾驶的作用。

新兴传感技术在自动驾驶领域的重要性不言而喻,自动驾驶最核心的一点就是感知。针对传感器技术的发展趋势,德州仪器(TI)中国区嵌入式产品系统与应用总监蒋宏对记者表示,随着自动驾驶等级的不断提高,传感器要能够全天候工作,探测更远的距离,感知更多的信息维度,感知的精确度也要提高。毫米波雷达就体现出了这样的发展方向。

“自动驾驶对感知的精度和准确度提出了更高的要求,推动77GHz毫米波雷达和60GHz毫米波雷达在DMS(驾驶员监测系统)和自动辅助泊车方面广泛使用。”蔡征对记者谈道,为了更好地满足自动驾驶时代对传感器等技术的高标准与严要求,TI推出了AWR2944毫米波雷达传感器。相比于前代产品,其分辨率提升了33%,探测距离提升了40%,表明在自动驾驶时代,感知的精确度正变得愈发关键。

L2或L2+级是当前市场主流

L4级别自动驾驶功能的实现是目前业界正在努力的方向。相比前几级自动驾驶功能,L4级对算力的要求更高,如何均衡相关系统的成本也成为业界关注的话题。

针对算力与成本,蔡征向记者表示,目前硬件已经不是L4等级自动驾驶推广和普及的瓶颈。考虑到最新处理器的运算速度,现在很多厂商都已经推出了能够达到L4等级的自动驾驶产品。但是,L4级别的自动驾驶产品能否得到推广和应用,其实取决于两个因素。

一是国家法规。车辆上路之后不仅需要完善车辆自身的自动驾驶功能,还需要考虑

到往来的行人和复杂的交通情况。二是成本问题。符合L4等级的核心芯片不仅自身成本较高,它的辅助供电系统、散热系统和整个系统成本也在提升。蔡征认为,这些成本的不断提升可能是未来限制L4级别自动驾驶产品市场拓展率的最主要因素。

目前来看,要想实现L4级别的自动驾驶功能,产业界还需要走很长的一段路。聚焦当下,L2级别或者L2+级别的自动驾驶功能是目前市场上的主流。

适配自动驾驶场景需要精准预测

很多公司都针对L2级别或者L2+级别的自动驾驶功能推出了相关产品。为了更好地适配这些自动驾驶场景,产品自身需要在性能和硬件等诸多方面进行提升。

精准预测是最需要提升的性能之一。以TI推出的毫米波雷达传感器AWR2944为例,AWR2944会做前角雷达,甚至有时还会做前向雷达。蒋宏向记者解释道,因为该款产品是4收4发,当它的FOV(视场角,镜头所能覆盖的范围)进行“收敛”时,就可以很精准地对前向目标进行探测。

而当整个自动驾驶系统提升到L3、L4级别,相关产品需要有更多雷达或前角雷达。比如,L3级别以上的自动驾驶产品可能有8个雷达或者更多前角雷达,所以能做前角雷达,甚至会做前向雷达的AWR2944可以很好地发挥作用。

从第一代到第二代毫米波雷达,RF(射频)性能和性价比都在提升,同时一些硬件也在不断提升。毫米波雷达的持续进阶对设计工艺提出了更高需求,也让生产难度有所提升。

77GHz的毫米波雷达是目前自动驾驶领域的主流产品,能够更好地为驾驶安全保驾护航。以77GHz的毫米波雷达为例,因为77GHz的毫米波雷达体积比较小,所以它的线路板面积很小。这很容易让毫米波雷达射频线路的设计难度变高,同时降低产品的成品率。

在这种情况下,该怎么解决射频线路的设计问题和产品的成品率问题?蒋宏认为,射频线路和成品率问题需要从多方面进行考量。由于天线设计是非专业人士的一个领域,该领域不光涉及一些理论问题,还需要相关技术积累。蒋宏表示,线路仿真测试、材料、工艺、做板等方面都需要非常专业的技术积累。可喜的是,目前整个产业生态都在这些领域实现了快速进步。

蒋宏表示,毫米波雷达传感器产品成品率提升背后的“秘诀”,其实源自芯片内部具备的非常多的自校准模块。

如果一个雷达安装在车里,它的工作温度范围其实变化很大。冬季的温度在零下20℃甚至更低温度,而夏季气温通常可达到六七十摄氏度。在这种情况下,片上的自校准、开环的自校准等环节需要做到环环相扣,以确保产品可以实现稳定工作。这一点对于客户终端产品成品率的提升至关重要。

台积电做出440亿美元预算 八成用于先进工艺

本报记者 沈丛

台积电于1月13日召开法人说明会,公布了2021年财报、2022年首季与全年展望、资本支出规划、半导体市场情况等重点方向。其中,先进制程的营收情况,以及未来在先进制程的资本支出,成为了本次法说会最受关注的亮点。

5nm营收趋近于7nm

成为第二大营收来源

受到大规模芯片短缺影响,市场对芯片的需求持续飙升,台积电在2021年取得了超预期的利润。同时,随着技术的不断演进,市场对于5nm先进制程的需求越来越高,2021年台积电5nm工艺的营收一跃成为了其第二大营收来源。

在此次说明会上,台积电表示,他们在2021年第四季度营收达157.4亿美元,同比增长21.2%,连续第6个季度创纪录。2021年台积电全年营收达568.2亿美元,同比增长24.9%。

随着量产时间的延长,5nm工艺与7nm工艺营收的差距已明显缩小。5nm工艺在台积电2020年营收中所占的比重仅为8%,远低于7nm工艺的33%,也不及占比17%的16nm工艺和占比13%的28nm工艺。而在2021年,5nm工艺营收占比高达19%,仅次于7nm的31%,成为台积电第二大营收来源。在2021年第四季度的营收中,5nm工艺所占的比重为23%,越来越趋近占比27%的7nm工艺。预计在不久的将来,5nm将很快取代7nm,成为台积电的第一大营收工艺。

在收入来源方面,台积电2021年主要收

入来源依旧来自智能手机,收入占比达44%,同比增长了8%。高性能计算是台积电2021年第二大收入来源,收入占比37%,同比增长了34%。而随着智能汽车产业的愈发火热,台积电2021在汽车产业的营收有了大幅度提升,成为了2021年营收增长最高的产业,同比增长了51%。

致力于先进制程研发

台积电将迎来下一个增长年

在此次法说会上,台积电公布了2022年展望及资本支出计划。台积电表示,预计2022年第一季度销售额达166亿美元至172亿美元,比分析师平均预估高出至少5%。据了解,台积电在2022年首季的营收情况将与大客户苹果的动作息息相关,若苹果智能手机在2022年第一季度市场需求相对乐观,台积电首季营收有望季增0.5%至3%,再次挑战单季历史新高。然而,若苹果手机销量一般,台积电第一季度的营收将与前一季度持平或有小幅减少。

此外,在此次说明会上,台积电表示,2022年资本支出将达400亿~440亿美元,比去年增加约100亿美元,其中,70%~80%的资本预算将用于先进工艺技术的研发,包括2纳米、3纳米、5纳米和7纳米。并预计未来几年营收将平均增长15%~20%,长期毛利率有望达到53%或更高。

台积电认为,2022年芯片产能将持续吃紧,同时,随着5G和高性能计算相关应用的飞速发展,半导体产业的需求将出现长期结构性增长趋势。因此,台积电表示,2022年将成为台积电的另一个增长年。