

RISC-V一定能成为AI时代的原生计算架构

——访奕斯伟集团董事长王东升



本报记者 张心怡

“5年前,我们说AI时代的原生计算架构可能是RISC-V,现在我们相信,RISC-V一定能成为AI时代的原生计算架构。”奕斯伟集团董事长王东升对《中国电子报》记者说道。这是王东升自卸任京东方董事长以来,第一次接受媒体专访。2019年,王东升从京东方卸任后,应邀加入了奕斯伟,在集成电路领域开启二次创业。基于对集成电路领域的长期思考,他以基础材料和计算芯片为着眼点,希望助力产业破解制造端和产品端的难点;从事电子级12英寸大硅片的奕斯伟材料公司主攻基础材料研发,目前出货量已居国内头部;奕斯伟计算公司则瞄准RISC-V新一代计算架构,主攻底层技术和计算芯片。而这一次采访的主题正是RISC-V。这一问世于2010年的年轻芯片架构,在短短12年的时间就迈过了出货百亿颗内核的门槛。但如此体量的出货,却因为缺乏体系化、成规模的应用场景,尚难以被外界确切地感知。

面向RISC-V产业的发展阶段和从业者的普遍诉求,王东升提出了RDI(RISC-V数字基础设施)的产业概念,并呼吁产业链上下游企业共建RDI生态。在他看来,RDI既是RISC-V从隐形走向显性、从嵌入式系统走向更大市场的契机,也是将RISC-V市场蛋糕做大,帮助产业高效发展的关键。

半导体产业的两个难题

2014年,国务院印发《国家集成电路产业发展推进纲要》,明确了要着力发展集成电路设计业、加速发展集成电路制造业、提升先进封装测试业发展水平、突破集成电路关键装备和材料四个任务。彼时,京东方已经进入全球显示产业的第一梯队,此时王东升在行业内喊响了“半导体显示”的概念,因为在他看来,显示就是半导体与光电的结合,与集成电路同属半导体大类。

“在观察和思考中,我发现中国集成电路产业面临的难题有两个。”王东升向《中国电子报》记者表示,一个是制造端的精密度问题,这是一个涉及晶圆制造本身、装备、材料的系统性问题;另一个是产品端的底层技术问题,关乎产业链发展和数据流通的效率与安全。

带着“不能为了做而做,而要为了解决难题而做”的目标,2019年6月底王东升从京东方卸任,7月份加入奕斯伟团队,在集成电路领域开始二次创业。面向半导体产业的“两个难题”,王东升选择硅片和计算芯片作为奕斯伟的两个主营业务。

硅片作为半导体占比最大的基础材料,其质量直接关乎晶圆制造和芯片成品的良率,也就是王东升提到的制造端精密度。巧合的是,王东升毕业之后实习的第一个车间,就是硅片的拉晶车间。在他加入奕斯伟的2019年,距离那次车间实习已经过去了30多年,国内硅片产业在主流的12英寸硅片正片中几乎仍处于空白阶段。这也坚定了奕斯伟团队发力硅片领域的决心。

2019年9月正式成立的奕斯伟计算,锚定了产品端的底层技术问题,在创立之初就以RISC-V为攻关方向。彼时,RISC-V这一新架构刚刚开始在全球启动产品化商业化探索,远没有现在的关注度,王东升做出了一个颇具预见性的选择。

RISC-V——AI时代的原生计算架构

在成立奕斯伟计算之前,王东升思考着一个问题:什么是AI时代的原生架构。

2016年3月,人工智能围棋软件AlphaGo击败顶尖职业棋手李世石,引发全球对AI技术风潮和未来发展的热烈探讨。

如果PC时代的芯片架构是x86,移动时代的芯片架构是Arm,那么人工智能时代的原生架构是什么?抱着这样的疑问,王东升与国内外的专家广泛探讨,也曾与当时奕斯伟计算的

核心团队进行深入而充分的探讨,最终他坚定选择了一个“年轻而充满无限可能”的计算架构——RISC-V。

RISC-V是诞生于2010年的芯片架构。那一年,计算产业已经有了源于CISC(复杂指令集)的x86架构和源于RISC(精简指令集)的Arm、Power、MIPS、SPARC等架构。但在RISC阵营中,除了Arm通过主打低功耗、高性能、小体积与x86架构错位竞争,在移动和嵌入式处理器市场站稳脚跟外,其他芯片架构由于力量分散,难以对抗由英特尔与微软组成的Wintel联盟,逐渐在市场式微。

而RISC-V作为RISC阵营的最新成员,凭借开放、精简高效、模块化、可拓展等优势,走出了与封闭式的x86和授权模式的Arm不同的发展路径,让产业界看到了通过更加简洁的方式进行架构创新的机遇,也让王东升看到了铺设AI时代算力底座의切入口。

“那时候选择RISC-V这个方向是需要勇气的,但在反复思考和探讨之后,我们在当时认为AI时代的原生架构可能是RISC-V。一方面,它是绿色、开放的,不断演进变化的,可以更好地适配AI时代软件与算法模型不断优化的需求。另一方面,它既能通过移植适配,与原有架构的生态兼容,也可以作为原生架构,支撑不断涌现的新应用、新场景。这意味着它是可以落地的,不是一个悬空的无人区,所以我们要把RISC-V作为攻关方向。”王东升说道。

2019年9月,定位为“AI时代以RISC-V为基础的新一代计算架构芯片与方案提供商”的奕斯伟计算在北京成立。在王东升看来,RISC-V不仅能作为国内产业提供绿色安全的算力底座,也能助力解决全球在AI时代面临的挑战。“在AI时代,整个人类社会发展到第四次产业革命之后,算力底座对于解决功耗和资源浪费问题至关重要。RISC-V能助力解决AI带来的诸多挑战,对全球产业发展乃至人类文明的进步做出贡献。”

五年过去,RISC-V在AI领域的潜力和势头更加清晰。一方面,随着生成式人工智能到来,以GPU为代表的AI加速器正在给全球数据中心的能耗和运维带来挑战,这凸显了低功耗架构的重要性。波士顿咨询集团(BCG)的数据显示,到2030年,美国数据中心在美国总用电量的占比将达到7.5%,是2022年的3倍,将近美国家庭用电量的1/3,主要原因是大模型训练和不断增加的人工智能查询。而RISC-V指令设计更加精简,没有向前兼容的历史包袱,能够按需裁剪和定制,相比已有架构进一步降低了功耗。

另一方面,RISC-V已经具备了支撑人工智能发展的雏形。除了持续提升在嵌入式处理器领域的市场占比,RISC-V在AI加速、边缘计算、智能终端等领域的应用也开始崭露头角,通过矢量和张量等指令扩展方式,不断完善对AI计算的支撑。在软件生态层面,RISC-V基本完成了与主流操作系统的适配,RISC-V原生的软件 and 工具持续涌现。

“5年前,我们说AI时代原生的计算架构可能是RISC-V,现在我们相信,RISC-V一定能成为AI时代的原生计算架构。”王东升对《中国电子报》记者说道。

RISC-V亟待从隐形走向显性

虽然RISC-V取得了可喜的进展,但这种进展似乎并没有被外界明确地感知到。其实,2022年,RISC-V内核出货量就迈过了百亿颗大关,可这些内核较为零散地分布在各类芯片中,难以形成规模化、体系化的应用场景。

于是,如何让RISC-V从隐形走向显性,并走向嵌入式系统以外的更多领域,成为王东升和奕斯伟计算团队思考的又一个课题。基于RISC-V产业的发展阶段,以及5年来的技术研发与工程实践,王东升提出了RDI的产业概念,即RISC-V Digital Infrastructure(RISC-V数字基础设施)。

所谓RISC-V数字基础设施,是包含采用RISC-V架构的底层芯片、硬件设备、软件系统、应用场景等多层级的数字基础设施创新体系。

在王东升看来,RDI对产业有两个利好:一是将RISC-V的蛋糕做大,二是帮助所有从业者开拓更广阔的应用场景。“RDI就像海上的冰山,海平面以上的部分是透明、开放的,海平面以下的部分是企业各自的能力。对于海平面以上的部分,我们要完善生态,把蛋糕做大。海平面以下的部分,则需要IP层到系统层的生态链伙伴持续提升各自的技术实力,再一起反哺RISC-V生态。”

对于“海平面以上的部分”,奕斯伟正在建设包含IP层、芯粒层、芯片层、板卡层和系统层的开放平台,通过开放的软硬件接口,让更多的企业参与RISC-V产业生态建设,并在开放合作的过程中获得商业回报。

RDI生态的创新和繁荣,需要RISC-V全产业链企业的参与,也需要与垂直行业的充分对接。2023年8月,元石智算、金融街资本、东软、云从、天数智芯等公司与奕斯伟计算发起成立了聚力开源生态技术有限公司(简称“聚力联盟”),推进RDI生态是该联盟的主要目标之一。在此基础上,2023年10月,武汉数据集团、武汉光谷金控集团、聚力开源公司,共同组建了武汉计算生态技术有限公司,即武汉RISC-V生态创新中心,其任务是携手RISC-V生态链伙伴遴选垂直行业场景,通过组建产品矩阵、形成系统方案、开展测试验证、实现应用示范及推广,推动RDI生态建设和产业发展。

与此同时,RDI的理念也受到了国际社会的认可和支持。在今年6月的RISC-V欧洲峰会期间,王东升向RISC-V的发明人之一、Sifive联合创始人Krstje Asanovi等国际伙伴介绍了RDI理念,Krstje听后立刻竖起大拇指,表示愿意与中国的产业伙伴一起在全球推广这一理念。

“我们希望推动产业一起发展RDI。在这个过程中,芯片企业不一定是获利最多的,按照营收金额来说,整机厂商可能会有更好的表现。但我们和联盟的其他成员企业,都获得了推广产品和验证场景的机会,在国内跑通这套商业模式之后,还可以向全球推广。”王东升说道。

未来六年先进封装市场年复合增长率有望达到12.9%

本报讯 咨询公司Yole Group近日在其最新报告中预测,受到高性能计算(HPC)和生成式人工智能(AIGC)技术发展的推动,先进封装行业在未来六年将迎来显著增长,预计年复合增长率(CAGR)将达到12.9%。这一趋势将促使行业规模从2023年的392亿美元增长至2029年的811亿美元。

尽管2024年第一季度的先进封装收入出现了季节性的下滑,降至102亿美元,但这一数字仍然超过了2023年的同期水平。第二季度预计收入将有所反弹,达到107亿美元,这表明尽管市场整体面临挑战,但该行业仍然呈现出复苏的迹象,并且预计下半年的表现会更加稳健。

从资本支出来看,2023年整个行业对先进封装的投资总额约为99亿美元,相较于2022年下降了21%。然而,

SK海力士投建韩国龙仁半导体集群首座工厂

本报讯 近日,全球第二大内存芯片制造商SK海力士在其官网上宣布,公司通过董事会决议,以约9.4万韩元(约合人民币492.56亿元)投资建设韩国龙仁半导体集群的首座厂房和业务设施。SK海力士称,此举是为了夯实公司未来发展基础,并及时应对日益剧增的面向AI的半导体存储器需求。

记者了解到,龙仁半导体集群位于韩国京畿道龙仁市远三面,其占地面积达415万平方米,拥有专门提供给半导体材料、零部件和设备公司的267万平方米专用工业用地。SK海力士将在此建造生产新一代半导体产品的四座先进厂房,携手全球50多家材料、零部件和设备企业构建半导体合作园区。

SK海力士表示,公司计划在龙仁首座工厂生产以HBM为代表的面向

随着市场需求回暖和技术进步,2024年的资本支出有望实现20%的增长,显示出业界对于未来发展的信心。

Yole Group的报告还强调了HPC和生成式AI领域的快速增长如何成为推动先进封装行业发展的重要驱动力。随着这些技术的应用越来越广泛,对更高性能和更高效能的需求也在不断增加,进而促进了封装技术的进步与创新。为了满足不断增长的需求,封装厂商正在加大研发投入,以开发出能够适应新兴应用的新一代封装解决方案。

此外,由于封装技术的进步能够显著提升芯片性能并降低功耗,因此在诸如数据中心、自动驾驶汽车以及消费电子等领域,先进封装技术的应用前景十分广阔。随着这些领域的持续发展,可以预见先进封装市场将在未来几年迎来显著的增长机会。

(文 编)

SK海力士投建韩国龙仁半导体集群首座工厂

AI的存储器和新一代DRAM产品,也将根据竣工时的市场需求,做好生产另外产品的准备。与此同时,公司还计划在首座厂房内建造“迷你工厂”(Mini Fab,具备300毫米晶圆工艺设备),以支持韩国材料、零部件和设备公司进行技术研发、验证和评估。

据悉,按照原定日程,SK海力士将于明年3月开工建设龙仁集群的首座厂房,将于2027年5月竣工。首座厂房建设完成后,将依次推进剩余的三座厂房建设,将龙仁集群发展成为“全球人工智能半导体生产据点”。

SK海力士副总裁、制造技术负责人金永式表示,龙仁集群将成为SK海力士中长期发展的基础,也将是与合作伙伴公司携手打造的创新和共赢的空间。公司将通过顺利完成大规模产业园区建设,加强韩国半导体技术和生态系统的竞争力。

(杨鹏岳)

信越化学营业利润六个季度以来首度增长

本报讯 近日,全球PVC(聚氯乙烯)及半导体硅片龙头企业信越化学公布了2024年第二季度财报,其整体业绩优于公司此前预期,营业利润6个季度以来首度恢复增长。信越化学表示,半导体硅片市场即将迎来触底反弹,12英寸产品出货量在AI需求的推动下,将在第三季度以后逐步增加。

财报显示,虽然信越化学第二季度的PVC销售萎缩,但硅片等半导体材料销售增长,合并营收较去年同期下滑0.2%至5979亿日元;合并营业利润同比增长0.1%至1910亿日元,虽然增长幅度较小,但这也是近6个季度以来,信越化学的利润首度呈现增长;合并净利润同比下滑6.3%至1440亿日元,连续第6个季度陷入萎缩。不过,与信越化学上一季度的预期(营业利润1650亿日元、净利润1200亿日元)相比,第二季度业绩优于预期。

从各主要业务来看,第二季度信越化学以PVC为核心的“生活环境基础材料(还包含苛性钠等产品)”业务营收

较去年同期减少3.3%至2325亿日元,营业利润下滑15.2%至679亿日元。信越化学指出,由于中国PVC厂商的出口额增加,PVC市场的竞争更加激烈,为销售带来了一定压力,不过以北美为中心调涨了PVC价格。

信越化学第二季度以半导体硅片为核心的“电子材料(还包含稀土磁铁、光刻胶、合成石英等)”业务营收增长3.0%至2270亿日元,营业利润增长11.8%至895亿日元,获利金额居所有业务之首。

信越化学表示,在半导体市场上,自2022年秋天以后开始的调整局面已大致触底,复苏情况依用途、领域而有所差异。在此情况下,致力于按照原定计划出货半导体硅片、光刻胶等半导体材料。

信越化学在此前公布的2024年财报(2024年4月—2025年3月)预测中乐观表示,其合并营收将年增3.5%至2.5万亿日元,合并营业利润将年增4.8%至7350亿日元,合并净利润后将年增2.5%至5330亿日元。

(许子皓)

武汉RISC-V生态创新中心在东湖高新区揭牌

本报讯 武汉RISC-V生态创新中心近日在东湖高新区揭牌。据悉,该中心由武汉数据集团、武汉光谷金控集团、聚力联盟共同组建,通过制定规范、测试验证、工程示范和应用推广等手段,致力于共同提升RISC-V技术、产品和系统能力,推动RISC-V数字基础设施生态建设、创新和产业发展。

在当天举行的“RDI生态·武汉创新论坛·2024”上,众多行业专家、企业家及RISC-V领域上下游产业链伙伴,共同探讨RISC-V数字基础设施发展新机遇,旨在促进RISC-V在千行百业的应用落地,打造面向未来的RISC-V创新生态。

据悉,RISC-V生态体系正处于技术与产业高速发展的窗口期。作为一个全新指令集架构,RISC-V具有精简、开放、低功耗、模块化、可扩展等优势,其应用已延伸到物联网、汽车电子、数据中心、人工智能等领域,未来应用场景还将更加多样化。

武汉RISC-V生态创新中心聚焦基于RISC-V CPU和操作系统的数字基础设施创新体系的规划、建设、运营、技术服务及生态链创新企业的引进和孵化支持。旨在夯实RISC-V CPU+OS及软件的数字基础设施,构建数字基础设施创新体系及数字安全屏障,助推RISC-V应用生态发展。

(锐 思)