

中电标协 RISC-V 工委轮值会长孟建熠:

RISC-V 在 AI 时代的“下半场”即将开启

本报记者 许子皓

近日,第二十一届中国国际半导体博览会(IC China 2024)在北京国家会议中心举行。在同期举办的第六届全球 IC 企业家大会上,中电标协 RISC-V 工委轮值会长、阿里达摩院首席科学家、知合计算 CEO 孟建熠分享了 RISC-V 在 AI 时代的机遇和挑战,介绍了 RISC-V 工委如何以 RISC-V 架构为技术底座,通过统一标准建设促进生态有序发展,推广城市级应用场景,打造 RISC-V 应用场景试验田、关键和共性技术孵化基地。



第二十一届中国国际半导体博览会
The 21st China International Semiconductor Expo

RISC-V 的变革力量来自其开源的特殊性,所有企业都可以根据它的开源标准进行创新。

RISC-V 有望成为 AI 时代技术底座

搭上 AI 时代快车的 RISC-V 计算架构,是新一代数字基础设施算力底座的优选,既能满足高效、高性能的需求,又能兼备安全可靠、高性价比、可拓展等特点。因此,越来越多的厂商开始使用 RISC-V 架构,使其应用场景不断丰富,从物联网设备、边缘计算逐渐向 AI 计算、高性能计算等领域迈进,发展速度迅猛。

孟建熠表示,对比 ARM 架构、x86 架构和 RISC-V 架构在 GitHub 上的开源项目的数量,RISC-V 架构项目在过去几年里飞速增长,数

量达到 1 万多个,与另外两家的差距持续拉近,极具潜力。从 RISC-V 基金会会员数量的角度来看,目前全球已有 4400 多家会员,预计到 2030 年,RISC-V 架构产品出货量的年复合增长率会达到 40% 以上,在整个产业的渗透率将进一步加深。标准制定方面,RISC-V 基金会在过去的两年已经完成 40 项标准制定的工作,包括 123 项 RISC-V 扩展,标准体系正在逐渐完善。

为什么 RISC-V 架构会发展得如此之快?孟建熠指出,是因为

RISC-V 正在改变“芯片架构”的生产关系。在半导体发展的历史中,类似的变革总是能推动产业效率的跃升。例如,代工模式改变了“芯片制造”的生产关系,让更多企业只需专注于自己的设计,不再需要兼顾复杂的生产制造环节;ARM 授权模式改变了“芯片设计”的生产关系,企业只需要获得 ARM 一家公司提供的架构,就可以制作各种各样不同功能的芯片。

RISC-V 的变革力量来自其开源的特殊性,所有企业都可以根据它的开源标准进行创新,从而更自

由地在各种应用场景中,开发出最具性价比和计算效率的产品,这点在 AI 时代尤为重要,所以,当前很多 AI 芯片的底座都基于 RISC-V 架构。

“RISC-V 有望成为 AI 时代计算架构变革的技术底座。”孟建熠表示,“行业最初只认为 RISC-V 是 CPU 的一种架构,但随着不断实践探索,我们发现 RISC-V 可以实现多样化的计算形态。比如 AI 计算、图形计算、隐私计算和科学计算等等,为整个行业提供了新的机遇。”

RISC-V 需要一个满足市场所需性能的标杆产品,并以此为中心完善软件生态。

标杆产品与生态发展互相促进

有机遇就有挑战,孟建熠认为 RISC-V 在 AI 时代主要面临三个挑战:已规模落地产品以中低性能为主,软件生态尚处于发展阶段,以及高性能产品尚未形成迭代。

与 ARM 架构的发展历程相似,RISC-V 想要在未来成为主流架构,就必须在算力上超越行业对它需要的阈值,才能被广泛使

用。因此,孟建熠认为,RISC-V 需要一个满足市场所需性能的标杆产品,并以此为中心完善软件生态。

目前,RISC-V 在高性能通用计算领域,已经拥有了媲美 x86、ARM 高端产品的性能,在通用性、可扩展性等方面具有优势。因此,孟建熠表示,在不远的未来,RISC-V 的标杆产品将突破通

用计算算力阈值,真正成为主流。软件生态方面,2024 年,RISC-V 通用计算生态已经逐步成熟,正式进入商用软件移植阶段,商业化的迭代正式开始,标杆产品与生态发展将互相促进形成“正循环”。

此外,RISC-V 在 AI 时代的“下半场”也即将开启。孟建熠表示,RISC-V 将有四个主要的发

展方向:一是针对 AI 的架构创新,RISC-V 具有计算架构的定制化能力,可以适配各种最新的应用需求;二是 RISC-V 处理器的产品性能将超过阈值,达到主流性能需求;三是 RISC-V 通用 CPU 的功能将更加完善;四是基于 Vector 和 Matrix, RISC-V 在 AI 加速以及各类应用中的扩展将更加有力。

当前,我国正在积极推动 RISC-V 产业发展,完善 RISC-V 软硬件生态建设。

以标准促进 RISC-V 生态建设

当前,我国正在积极推动 RISC-V 产业发展,完善 RISC-V 软硬件生态建设,RISC-V 工委在其中起到了引导作用。

RISC-V 工委由中国电子工业标准化技术协会牵头组建,主要围绕 RISC-V 产业发展开展标准研制、符合性评估、知识产权保护、人才培养、产业研究等方面工

作,并与 RISC-V 国际基金会协同合作,引导产业实现优势互补、资源共享、协同推进,推动 RISC-V 生态的繁荣。

据孟建熠介绍,RISC-V 工委正在推进三项工作:一是标准体系建设。以 RISC-V 国际基金会的标准文档为基础,结合中国应用需求,建立国内统一的 RISC-V 标准

体系,推进生态建设。二是推动应用图谱建设。结合国内市场需求,对不同应用的 RISC-V 参考产品进行图谱分类,并推动应用示范。三是软件社区的组建。RISC-V 工委联合软件所、国内企业构建完整的 RISC-V 软件生态平台。

2024 年,RISC-V 工委会在指令集、安全可靠、基础软件、测试评

估和产业链五个工作部推进团体标准立项工作。与此同时,RISC-V 工委正在推动 RISC-V 的落地应用。以城市级应用场景推广为主,打造 RISC-V 应用场景试验田、关键和共性技术孵化基地、应用人才培养基地;结合各地的产业与技术发展趋势,聚焦包括 AI 应用在内的具体方向,建立应用示范专项基地。

2024 年第四季度表现疲软。英飞凌、德州仪器、恩智浦半导体和瑞萨电子均由于汽车市场疲软和库存减少,预计 2024 年第四季度收入将下降。意法半导体也提到了这些因素,但预计收入将增长 2.1%。高度依赖智能手机的公司收入预期各异,高通预计增长 7.2%,而联发科预计下降 1.0%。2024 年第四季度与第三季度相比的加权平均指引为增长 3%。然而,各公司的指引差异很大,从美光的同比增长 12% 到英飞凌的下降 18% 和瑞萨电子的下降 19%。(吉文)

台积电有望 2027 年推出超大版 CoWoS 封装

本报讯 在近期召开的欧洲开放创新平台(OIP)论坛上,台积电宣布计划在 2027 年推出超大尺寸晶圆级封装(CoWoS)技术,最高实现 9 倍光罩尺寸(reticle sizes)和 12 个 HBM4 内存堆叠,预计它将在 2027 年至 2028 年被超高端 AI 处理器采用。

据悉,这一全新封装方法将解决性能要求最高的应用,并让 AI (人工智能)和 HPC(高性能计算)芯片设计人员能够构建手掌大小的处理器。

台积电每年都会推出新的工艺技术,尽最大努力满足客户对功耗、

性能和面积(PPA)改进的需求。对于有更高性能需求的客户来说,EUV 光刻工具现有的光罩尺寸(858 平方毫米)是不够的。台积电于 2016 年推出的初代 CoWoS 封装方案,支持大约 1.5 倍光罩尺寸,而目前已经提升到 3.3 倍,可以封装 8 个 HBM3 堆栈。

台积电承诺在 2025—2026 年期间,支持 5.5 倍光罩尺寸和最高 12 个 HBM4 内存堆叠,并计划在 2027 年推出“Super Carrier”9 倍光罩尺寸的 CoWoS 封装方案,为芯粒(chiplets)和内存提供高达 7722 平方毫米的空间。(文编)

德州仪器新型 MCU 可实现边缘 AI 计算

本报讯 记者许子皓报道:近日,德州仪器推出其首款集成神经处理单元(NPU)的实时 MCU 产品——TMS320F28P55x 系列 MCU。德州仪器表示,这款 MCU 借助边缘 AI 的计算能力,可以实现高精度、低延迟的故障检测,故障检测准确率达到 99%。

当前,在智能物联、智能工控、汽车电子等新兴领域的快速发展下,MCU 的应用场景不断拓宽,市场需求不断提升,根据 Yole 的研究报告,2023 年全球 MCU 市场规模约为 282 亿美元,预计至 2028 年,将以 5.5% 的年复合增长率增长,达到 388 亿美元。

但随着这些新兴领域的高速发展,对于 MCU 的性能需求也在激增,通常需要新的 MCU 产品具备低功耗、低成本、实时性和高可靠性的特点。而将边缘 AI 技术融入 MCU 就成了厂商解决这一难题的秘诀。

德州仪器全球高级副总裁、嵌入式处理产品负责人 Amichai Ron 表示:“我们在 C2000 产品系列中引入了增强型实时性能和边缘 AI 技术,这将赋能工程师解决更复杂的问题,实现更高水平的系统效率、安全性和可持续性。”

德州仪器同期推出了 F29H85x 系列中的新型 64 位 C29 内核,该产品符合国际标准化组织 26262 和国际电工委员会 61508 汽车和工业安全标准,级别达到 ASIL D 和 SIL 3。据了解,这款产品基于 64 位架构设计,其实时信号链性能相较于 C28 内核实现了翻倍的增长,同时具有全方位的诊断和故障检查机制以及网络安全功能。

德州仪器中国区技术支持总监师英指出:“当前汽车行业与工业领域正加速迈向高效、快速决策系统的时代。德州仪器关注到这一趋势,并通过不断的技术创新推动实现单芯片实时控制功能,并融入功能安全与信息安全机制,以应对日益复杂的系统挑战。”

英飞凌发布首款太空级 NOR Flash

本报讯 英飞凌近日宣布推出业界首款用于太空和极端环境应用的 512 Mbit 抗辐射加固设计 QSPI NOR Flash。

据介绍,这款 NOR Flash 采用快速四串行外设接口(133 MHz),具有极高的密度、辐射和单次事件效应(SEE)性能,是一款完全通过 QML 认证的非易失性存储器,可与太空级 FPGA 和微处理器配合使用。

这款新器件由美国空军研究实验室太空飞行器局(AFRL)资助,并与 Microelectronics Research Development Corporation(以下简称“Micro-RDC”)共同开发完成。它基于英飞凌经过实际验证的 SONOS(硅衬底-隧穿氧化层-电荷存储层氮化硅-阻挡氧化层-多晶硅栅极)电荷栅极技术,运行速度较低密度替代品提高 30%。

AFRL 太空电子技术项目经理 Richard Marquez 表示:“下一代太空级系统的设计者对高可靠性、高密度存储器的需求不断增长。我们与英飞凌、Micro-RDC 等行业领导者合作,共同开发一种集高密度、高数据传输速率与优于替代品的辐射性能于一身的技术解决方案。”Micro-RDC 总裁 Joseph Cuchiaro 表示:“英飞凌的抗辐射加固设计 NOR Flash 很好地补充了 Micro-RDC 的极端应用环境解决方案系列。随着 512 Mbit 密度器件的推出,设计者能够设计出性能卓越的系统,以满足比以往更广泛任务类型的严格要求。”

英飞凌的 SONOS 技术具有高达 10000 P/E 的出色耐用性和长达 10 年的数据保存期。该产品的 133 MHz QSPI 接口为太空级 FP-GA 和处理器提供了高数据传输速率,并采用占板面积 1 英寸 × 1 英寸的陶瓷 QFP(QML-V),以及占板面积更小的 0.5 英寸 × 0.8 英寸塑料 TQFP(QML-P)两种封装。此外,该器件还为太空 FPGA 引导代码解决方案提供了最高密度的 TID/SEE 性能组合,其 QML-V/P 封装获得 DLAM 认证,能够满足最严格的行业资格认证要求。

该器件的典型应用包括太空级 FPGA 的配置映像存储和太空级多核处理器的独立启动代码存储。(英讯)

世界半导体贸易统计组织:第三季度半导体市场增长至 1660 亿美元

本报讯 近日,世界半导体贸易统计组织(WSTS)发布报告称,2024 年第三季度半导体市场增长至 1660 亿美元,较 2024 年第二季度增长 10.7%。2024 年第三季度的增长是自 2016 年第三季度 11.6% 以来的最高季度环比增长。

从具体公司来看,位居第一的英伟达凭借其在 AI GPU 领域的实力,在 2024 年第三季度仍是全球最大的半导体公司,收入达 351 亿美元。英伟达将其 AI GPU 作为模块销售,其中包括由 SK 海力士、美光科技和三星提供的内存,以及其他

外部供应商提供的组件。

三星半导体以 220 亿美元的收入位居第二,AI 服务器的内存是其主要的收入来源。

博通位列第三位,其 2024 年第三季度的收入为 140 亿美元。博通表示,AI 半导体是其主要收入来源。

英特尔和 SK 海力士位列第四和第五名。位列第六到第十的公司分别为高通、美光、AMD、英飞凌、德州仪器。

报告显示,2024 年第三季度,大多数主要半导体公司表现强劲。存储公司 SK 海力士、美光科技和铠

侠的报告都显示,2024 年第三季度与 2024 年第二季度相比,收入均增长了两位数。

在排名前 15 的公司中,唯一收入下降的公司是瑞萨电子,由于汽车市场疲软和库存减少,其收入下降了 3.8%。

展望 2024 年第四季度,在人工智能的推动下,数据中心市场预计将为英伟达、美光和 AMD 带来可观的收入增长。三星半导体和 SK 海力士未提供具体的 2024 年第四季度收入指引,但都指出 AI 服务器需求强劲。依赖汽车行业的公司预计