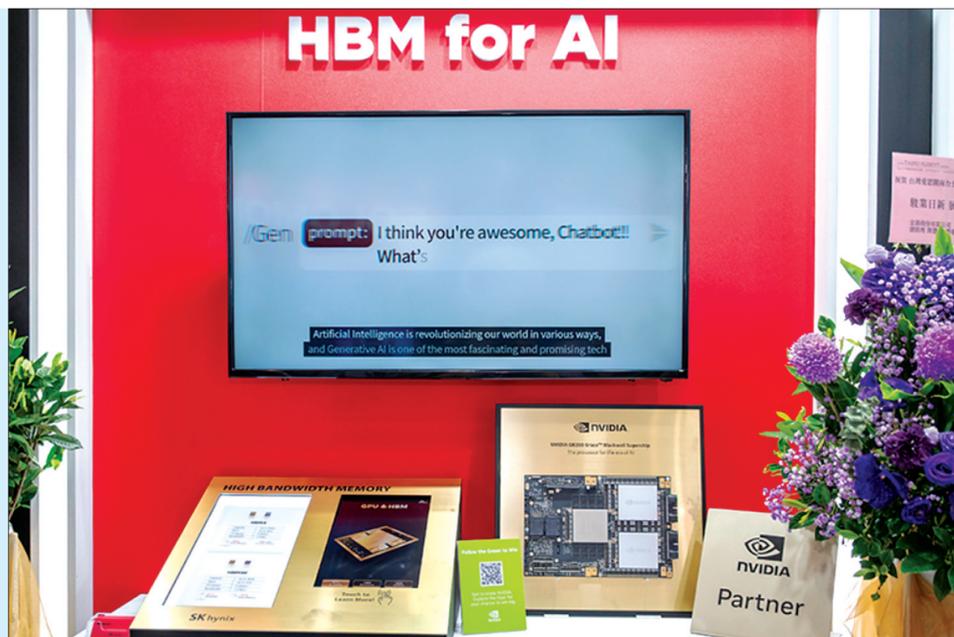


更深更广更热

HBM产业呈现三大特点



本报记者 许子皓

HBM正从高性能计算领域逐渐向消费电子等更广泛的市场渗透,其影响力正不断扩大,各大HBM制造企业在这一领域展开的技术博弈也将更加激烈。

近日,苹果公司表示,正为20周年版iPhone研发多项创新技术,其中HBM被视为关键发展方向之一,苹果已与三星和SK海力士等主要内存供应商展开讨论。集邦咨询在最新报告中提出,HBM技术迭代迅速,预计HBM3e将在2025年占据超90%的出货份额,而HBM4也已蓄势待发,将于2026年第二季度投入量产。

技术发展更“深”

HBM的性能飞跃源于多项关键技术协同创新,其中硅通孔(TSV)技术是实现垂直互联的基础。

TSV的本质是用立体互联打破平面的限制,通过在硅片内部构建导电通道,将多层DRAM芯片堆叠成数据洪流的高速通道。因此,该技术的每一次技术突破都直接推动着HBM性能的代际跃升。韩美半导体董事长 Kwak Dong-shin 表示:“TSV的精度每提升1微米,HBM的带宽就能增加100GB/s。”

HBM1采用的第一代TSV技术,深宽比仅1:10,单堆栈最多堆叠4层DRAM芯片,带宽仅128GB/s;到HBM2时代,TSV深宽比提升至1:15,8层堆叠实现256GB/s带宽;而HBM3e的第三代TSV技术将深宽比拓展至1:20,配合12层堆叠设计,带宽跃升至2.8TB/s。

当前,头部企业的TSV技术路线呈现出差异化竞争态势。

SK海力士选择“精度优先”策略,其HBM3e产品采用的第三代TSV技术,通过“先通孔后研磨”工艺,将产品的深宽比提升至1:20,显著缩短了信号传输路径,使信号延迟较前代降低15%,为2.8TB/s的带宽提供了支撑。这项技术使其产品在英伟达H100的认证测试中,信号完整性达到99.99%。

三星主攻“密度突破”,在16层堆叠样品中采用“阶梯式TSV”设计,将不同层的通孔错开排列,减少应力集中,使堆栈高度控制在300微米以内,较同规格产品降低15%,为薄型化封装创造条件。

美光则在边缘计算专用HBM中采用“混合TSV”方案,将关键信号通道的通孔直径设为7微米,普通通道放宽至10微米,在保证性能的同时降低20%的制造成本,赢得亚马逊AWS的边缘服务器订单。

随着各家企业开始进军HBM4,TSV技术的演进正面临物理极限的考验。专家表示,当TSV的深宽比需突破1:25,会导致蚀刻过程中的电荷积累不均,可能使孔壁出现纳米级裂纹。IMEC(比利时微电子研究中心)的研究显示,采用新型硅锆合金衬底可减少蚀刻损伤,使深宽比1:30的TSV漏电率降低50%。同时,随着通孔数量突破百万级,信号串扰问题越发突出。

但台积电在近期表示,可以在CoWoS中介层中引入TSV屏蔽技术,通过在通孔周围设置接地环,将串扰噪声控制在-80dB以下,为HBM4的4096位接口提供

稳定保障。

此外,在TSV的蚀刻环节,等离子体蚀刻需要在硅片上打出深度均匀的孔洞,孔壁粗糙度需控制在2纳米以内,相当于头发丝直径的十万分之一。TEL提供的TSV专用蚀刻设备,通过实时监测等离子体密度,将孔深偏差控制在±0.5微米内,为SK海力士HBM3e的良率提升至85%提供了支撑。

而通孔填充环节更堪称微观世界的“无缝焊接”,应用材料公司开发的电化学沉积系统,能在5微米的孔洞内实现铜原子的均匀堆积,避免产生气泡或缝隙,其填充速率达到每秒10纳米,确保HBM堆栈的垂直电阻低于50毫欧。

与TSV技术共同成长的还有ALD技术和混合键合技术。在TSV加工环节,ALD技术负责沉积5-10纳米的氧化铝绝缘层,将铜柱与硅基底的漏电通道彻底阻断。在DRAM单元制造中,ALD生长的高介电常数薄膜使存储电容在面积缩小30%的情况下保持容量不变。SK海力士在HBM3e中采用的“ALD+等离子体处理”复合工艺,使TSV孔壁的粗糙度降至0.5纳米以下,信号传输损耗较前代降低15%。

混合键合作为先进的三维集成技术,也在进一步发挥作用。SK海力士在HBM3e中率先应用铜-铜直接键合,取消传统凸点设计,使互联密度从每平方毫米10万个提升至100万个,英伟达H100因此实现1.8TB/s的内存带宽。三星则在混合键验证中遇挫,但其16层堆叠样品通过优化键合压力分布,将界面电阻控制在50毫欧以下,为HBM4的量产打下基础。

应用领域更“广”

HBM技术的飞速发展,使其不止在高性能计算(HPC)和人工智能领域大放异彩,也在消费电子、智能汽车、物联网等新领域崭露头角。消费电子领域对高性能、低功耗组件的追求从未停歇,HBM技术的特性恰好契合这一需求。

由于AI摄影、AI语音助手等功能需要在手机端运行大模型,智能手机需要在短时间内处理大量的数据。传统的LPDDR内存难以满足端侧的高性能需求,而HBM技术的出现为智能手机性能的提升带来了新的可能,各大手机厂商已开始布局。

有消息称,苹果公司已与三星和SK海力士等厂商商讨,计划在iPhone中采用HBM产品;三星也计划于2028年推出首款搭载LPW DRAM内存(移动HBM)的移动产品,这款产品通过堆叠LPDDR DRAM,显著增加了I/O接口数量,不仅降低了功耗,还将设备性能提升至全新高度,其带宽可达200GB/s以上,较现有的LPDDR5x提升了166%。

不仅是手机,在笔记本电脑领域,也有企业跃跃欲试。有消息称,SK海力士推出的“Mini HBM”方案,通过减少堆栈层数至4层,将单颗芯片容量控制在8GB,成本降至

标准HBM3的1/3,却仍保持1024GB/s的带宽,这一方案有望应用于高端游戏本。

随着汽车“新四化”趋势的推进,汽车对数据处理能力的要求呈指数级增长。例如,先进辅助驾驶需要实时处理来自摄像头、雷达等传感器的大量数据,包括车辆周围的环境信息、其他车辆和行人的位置信息等,每小时产生的数据量可达数TB。基于HBM,汽车科技企业能够为汽车的数据处理提供更多方案。

目前,SK海力士的车规级HBM2E已被应用于谷歌旗下的Waymo自动驾驶汽车,这款产品内存为8GB,传输速度达到3.2Gb-ops,实现了410GB/s的带宽,为自动驾驶汽车的实时数据处理、高分辨率图像处理等提供了支持。SK海力士还积极与英伟达、特斯拉等行业巨头拓展合作,致力于将HBM技术更广泛地应用于自动驾驶汽车中。

HBM在物联网设备数据处理和存储中同样具有广阔的应用前景。HBM技术能够为物联网设备提供高带宽、低延迟的内存支持,满足边缘计算在本地快速处理数据的要求。比如西门子部署的智能工厂里,装配机器人搭载的边缘计算单元集成美光HBM3,通过实时处理32路高清摄像头的数据流,实现了0.1毫米级的装配精度,错误率较传统方案降低90%,提高了生产效率和产品质量。

在智能家居方面,各种智能设备如智能摄像头、智能音箱、智能家电等都需要进行数据处理和存储,HBM可以为智能家电提供更强大的数据处理能力,使其更快速地响应用户指令,实现设备之间的智能联动。

尽管HBM在应用场景拓展方面展现出了巨大的潜力,但在实际推广和应用过程中,仍然面临着诸多挑战。这些挑战涵盖了成本、性能、散热、体积、兼容性和安全性等多个领域。

为了应对不同的应用场景,定制化也成为了HBM未来发展的关键方向之一。例如,针对手机、平板电脑等紧凑型设备,定制化的HBM需要优化散热设计,采用高效的散热模块和热管等技术,并尽可能轻量化。

市场竞争更“热”

市场研究机构Yole预测,HBM的市场规模将从2022年的27亿美元增至2029年的377亿美元,年复合增长率将达到38%,到2030年HBM市场规模有望突破1000亿美元。如此大的“蛋糕”,促使各大HBM制造企业之间的竞争越来越激烈。

SK海力士作为HBM的缔造者,在HBM市场份额中独占鳌头。今年第一季度,SK海力士以36.7%的市场份额超越三星,首次在全球DRAM市场夺冠,其在HBM市场的优势进一步扩大,市场份额已达到三星的两倍以上。

与英伟达的紧密合作是SK海力士巩固市场份额的重要策略。这

种紧密的合作关系,不仅为SK海力士带来了稳定的订单和收入,还使得SK海力士能够深入了解英伟达的技术需求和发展方向,从而提前布局,研发出更符合英伟达需求的HBM产品。

在产品技术方面,SK海力士不仅率先推出面向AI的12层HBM4内存样品,并向主要客户出样,还与台积电签署合作备忘录,共同生产下一代HBM,通过先进封装技术提高逻辑和HBM的集成度。SK海力士计划在今年下半年量产12层HBM4产品,并且在2026年推出16层堆叠的HBM4E产品,内存带宽将是HBM4的1.4倍。

三星作为存储领域的老大哥,纵横江湖多年,在HBM市场的发展历程却是一波三折,其HBM3e产品始终无法得到英伟达等厂商的验证,主要原因是散热、功耗和良率问题,也严重影响了三星的收入和市场份额。

为了扭转局势,三星内部进行了一系列的调整。在人员调配方面,三星将部分员工从半导体代工部门调配至HBM业务,加强该业务的研发和生产能力。

在技术研发方向上,三星启动了“双轨计划”。一方面,短期冲刺验证,调整1a DRAM的信号完整性设计,优化TSV(硅通孔)封装良率,目标在今年第三季度前满足英伟达要求。另一方面,长期押注HBM4,在设备解决方案部门新设“HBM开发组”,专注推进HBM4技术。三星计划在HBM4设计中采用4纳米工艺,联合AMD、高通等客户共同研究3D堆叠与混合键合技术,计划2026年实现HBM4与CPU/GPU的协同设计,构建差异化竞争力。

美光在HBM市场虽然起步相对较晚,但凭借其先进的制程技术,正展现出惊人的追赶速度。美光近期宣布,其采用1β(1-beta)DRAM制程、经过验证的12层先进封装技术及内存内建自测试(MBIST)功能生产的12层堆叠36GB HBM4内存,已向多家主要客户送样,预计将于2026年量产,以配合客户下一代AI平台的扩产进度。

在产能扩充方面,美光计划在全球范围内扩展其生产能力,包括在日本广岛市新建一家DRAM芯片工厂,预计最快将于2027年投入运营。这家新工厂将专注于DRAM生产,尤其是HBM产品的生产,届时将大幅提升美光的HBM产能。市场拓展方面,目前,美光已经与多家下游客户就2025年的HBM供货达成协议。

美光预计,2025年其HBM市场占有率将提高到20%~25%之间,到2027~2028年,发布12层和16层堆叠的HBM4E产品,带宽达2TB/s以上。此外,美光还与美光电子等合作,为下一代XPU定制HBM方案,以满足不同客户的多样化需求。

HBM作为顺应时代发展的产物,仍处于半导体行业变革的核心。随着技术瓶颈的逐步突破和市场需求的持续增长,HBM有望在未来十年内成为内存市场的主流,继续改变计算架构和数据处理模式。

7月11日,新紫光集团迎来战略重组三周年。作为横跨芯片、ICT、云网等多个关键产业,拥有200多家下属企业的综合性集团,新紫光的“求新”之路与“焕新”之道,一直受到产业界和资本市场的强烈关注。三周年之际,新紫光集团交出了怎样的答卷?

三周年
新紫光集团“交卷”!

本报记者 张心怡

为大于其细

大而不强、全而不精,是许多大型综合性企业的共性问题,也是我国产业体系面临的挑战。新紫光作为大型高科技企业集团,也曾陷入体系庞大但缺乏纵向串联、下属企业众多但缺乏横向联动的局面。新紫光集团成立以来,通过重构治理框架、梳理业务板块,在高度聚焦“智能科技”主业的同时增强资源协同,推动集团业务从“大而全”走向“精而强”。

在新紫光的业务版图中,以半导体、ICT、云网为代表的高技术板块,不仅是公司的高附加值业务,也是当前产业数字化转型所需的底座平台。据悉,新紫光集团整合了自身及股东控股的数十家核心科技企业,组成“新紫光体系”,形成覆盖芯片设计、封测、设备、材料和模组的半导体产业链,以及ICT设备、云服务及数字化解决方案的数字经济全产业链生态。可以说,“新紫光体系”不仅塑造了新紫光集团的“全产业链体系”,也是中国智能科技产业创新的“突击队”。

与此同时,新紫光从集团层面进行资源协同,加强下属企业之间

的联动,形成合力。以汽车电子业务为例,新紫光体系下的紫光展锐、紫光同志、紫光国芯、国芯晶源、睿睿、瑞能半导体等企业均在布局汽车电子,若彼此之间缺乏联动,就只能以单点产品各自突围,无法形成系统性、平台化的解决方案。

集团层面抓总,使新紫光的汽车电子业务完成了从“单点”到“体系”的蜕变——不仅实现了对车载计算、控制、存储、传感、功率等环节的全覆盖,还推动了新紫光系产品和服务的跨域技术共享、跨端场景融合。今年5月,中国一汽与新紫光集团签署战略合作协议,以“总对总”的合作模式应对汽车EE架构向“域集中式”“中央集中式”转变的趋势,共同助力本土车规芯片生态圈的成长壮大。

“为大于其细”——综合型集团在扩大版图的同时,也需精细化管理,结合市场需求、产业需求与企业专长,找到“立身之本”,并基于资源协同提升发展效益与创新能力的,成为合作伙伴提质增效、重点产业转型升级的“助推器”。

焕新驱智擎

以生成式人工智能、智能汽车、新一代通信技术为代表的新技术,正在重塑经济增长引擎。要持续发挥新兴技术的引领作用,一方面要通过创新研发,加速突破技术瓶颈;另一方面要推动技术的应用落地,使其真正普惠千行百业和广大消费者,构建“技术研发—量产转化—应用落地”的创新闭环。

技术研发层面,新紫光集团近三年累计专利申请量增加25%达30000余项。2024年,新华三作为主要参与单位分别获得国家科技进步一等奖及二等奖。

量产转化层面,新紫光集团纵向贯通产业链,基于全球30余家生产基地、60多个国家和地区的分支机构,构建了全球范围的生产能力。

应用落地层面,新紫光集团从消费市场、行业场景、区域发展三个维度,推动AI等新兴技术的普惠发展。

比如,在新型消费终端方面,紫光展锐在今年6月推出UNISOC端侧AI平台化解决方案。其单芯片AI解决方案可实现大模型生成速度较上一代异构方案提升20%以上、大模型推理功耗降低60%。异构分布式AI解决方案支持从1T至100T的算力配置,适配30多种不同参数模型。



图为新紫光集团车规级动力域高端MCU芯片THA6系列