



下一代HBM离不开台积电了？

本报记者 许子皓

近期，台积电生态系统和联盟管理负责人Dan Kochpatcharin公开表示，台积电将与三星联手研发下一代HBM产品HBM4。这也是三星与台积电首次在HBM领域的合作。而HBM领域的另一位巨头SK海力士同样在此前发布公告称，将与台积电合作开发HBM4，预计在2026年投产。两大HBM供应商接连送上橄榄枝，台积电掌握了哪些“关键法宝”？

迭代越来越快 技术越来越难

HBM即高带宽存储，由多层DRAM Die垂直堆叠，每层Die通过硅通孔(TSV)技术实现与逻辑Die连接，使得8层、12层Die封装于小体积空间中，从而实现小尺寸存与高带宽、高传输速度的兼容，优秀的特性使其成为高性能AI服务器GPU显存的主流解决方案。

当前，随着AI市场持续增长，AI服务器对于HBM的内存容量和传输带宽提出了更高要求，推动了HBM技术的迭代升级。每一次技术的突破，都会让HBM在带宽、容量、能效、架构以及市场应用等方面取得显著提升，但所需的制造工艺也越来越复杂。

HBM初代产品于2014年由AMD与SK海力士共同推出。作为GDDR5产品的竞品，HBM采用了硅通孔(TSV)技术，实现4层die堆叠，可以提供128GB/s带宽和1GB内存，其当时的带宽就已经高于DDR4和

GDDR5产品，同时以较小的外形尺寸消耗较低的功率，能更好地满足GPU等带宽需求较高的处理器。性能、尺寸和功耗上的种种优势，让HBM在业内崭露头角。

第二代的HBM于2016年由后来者三星“越级”抢先发布。作为旗下首款HBM产品，三星将每个HBM堆栈容量提升至8GB，使用硅通孔和微凸块(microbump)技术，可以提供256GB/s的带宽，是GDDR5产品带宽的7倍之多。SK海力士则是在2018年正式推出堆叠高度为4层和8层的HBM2，并且引入了伪通道模式，将一个通道分为两个独立的64位I/O子通道，从而降低延迟，提高带宽。

HBM2的加强版HBM2E于2019年再次由三星抢先发布。据了解，这款HBM2E单颗最大容量为16GB，最高可提供3.2Gbps的数据传输速度。而SK海力士则是于2020

制程和封装是两大“拦路虎”

随着每一代产品在性能上的飞跃式提升，HBM对制程和封装技术的要求也更加严苛。

TrendForce集邦咨询分析师许家源表示：“HBM1至HBM2e的堆栈层数仅为4层和8层，制程难度相对较低。但到了HBM3，堆栈层数增加至12层，制程难度大幅提升，HBM4更是要将堆栈层数增加至16

层，复杂程度和制造难度都将进一步提升。”

据了解，更先进的制程节点有助于提高HBM的能效比和性能。因此，从HBM1到HBM3E，所采用的制程节点从28纳米进化到10纳米。三星和SK海力士近期发布的信息显示，HBM4将采用5纳米甚至更小的工艺节点，进入先进制程范畴。

先进封装技术也是同理。除了更先进

台积电掌握核心技术

台积电作为全球领先的半导体制造企业，可以说是当下先进制程的天花板，并且良率较为稳定，可靠性和安全性在业内首屈一指，这都是研发HBM4的必备条件。

据了解，台积电计划在HBM4生产中采用N12FFC+和N5两种制程技术。N5是一种领先的5纳米制程技术，能够提供更小的晶体管尺寸，实现更高的性能和更低的能耗。这种先进的制程技术能够优化HBM4的逻辑芯片，提升整体的计算效率和能效比。

在封装技术方面，半导体行业专家告诉记者，台积电拥有诸如InFO和CoWoS等先进封装解决方案，能够在芯片之间构建更快速的数据交换路径，有助于解决热管理和信号完整性问题。通过先进的封装技术，台积电能够为HBM4提供更大的带宽和更低的延迟。

这些优势正是SK海力士所渴望的。SK海力士缺乏先进制程的制造能力，和三星又是HBM领域的直接竞争对手，再加上三星的先进制程良率一直是个大问题，双方合作的可能性很低。因此，SK海力士最好的选择就是尽快与台积电达成合作。

于是，在今年4月，SK海力士宣布与台积电签署一份谅解备忘录，双方将合作生产下一代HBM，并通过先进的封装技术提高逻辑和HBM的集成度。

据了解，两家公司将首先对HBM封装

内最底层的基础裸片进行性能改善。SK海力士以往的HBM产品都是基于公司自身工艺制造基础裸片，但从HBM4产品开始计划采用台积电的先进逻辑工艺。若在基础裸片采用超细微工艺，就可以增加更多的功能。SK海力士计划在2025年下半年推出12层DRAM堆叠的首批HBM4产品，于2026年推出16层堆叠的HBM4E产品，内存带宽将是HBM4的1.4倍。产品面市时间将与英伟达AI加速器发布周期保持一致。SK海力士封装开发副总裁李康旭在“2024异构集成全球峰会”上详细阐述了这一合作的重要性。他指出，通过将逻辑工艺应用于HBM4的基片，可以显著改善性能和能效。

近期，英伟达、台积电和SK海力士又宣布组建“三角联盟”，面向AI需求共同推进HBM4等下一代技术。可以看出，SK海力士想与台积电进行深度绑定。据了解，SK海力士和台积电决定在原计划的12纳米制程基础上，增加5纳米工艺来生产HBM4。

在封装技术方面，SK海力士表示，SK海力士独有的先进MR-MUF技术以及混合键合等新一代封装技术，将与台积电独有的CoWoS技术相融合，这对于下一代HBM的发展至关重要。

而三星最初的想法是继续押注自家的4纳米工艺。三星的一位高管表示：“我们与台

随着AI市场持续增长，AI服务器对于HBM的内存容量和传输带宽提出了更高要求。

年率先开始批量生产HBM2E产品，性能比三星的产品更强，在容量同样达到16GB的同时，带宽达到了460GB/s。此外，SK海力士还首次采用了批量回流底部模制填充(MR-MUF)技术，提升了散热性能，使得这款产品的散热性能比上一代的HBM2高出36%。

随后的HBM3和其强化版HBM3E都由SK海力士占据主导权。SK海力士的HBM3产品于2022年正式推出，其堆叠层数及管理通道数均有增加，可以提供6.4Gbps传输速度，带宽可达819GB/s，容量提升到24GB。而SK海力士最新的HBM3E产品将使用其1b纳米制造技术(10纳米级的DRAM节点)来生产，并将MR-MUF技术升级为先进的MR-MUF4技术，数据传输速度有望提高到8.0GT/s，带宽提高到1TB/s，SK海力士计划今年实施量产HBM3E产品。

除了更先进的TSV技术，HBM还需要MR-MUF技术、混合键合等新一代封装技术加持。

的TSV技术，HBM还需要MR-MUF技术、混合键合等新一代封装技术加持。

HBM4对先进制程工艺和先进封装技术的需求，让如何保持高良率成为一个难题。在制造过程中，任何一层die出现问题，都可能导致整个堆叠失效。这一系列的难题让三星这个全球第二的代工厂商都感到棘手，联合先进制程的霸主台积电似乎成了唯一选择。

SK海力士和台积电决定在原计划的12纳米制程基础上，增加5纳米工艺来生产HBM4。

积电和SK海力士不同，我们的芯片设计人员直接参与HBM4生产，这是我们的独特优势。”目前，三星已在其设备解决方案部门新设“HBM开发组”，专注于推进HBM4技术。

此外，三星先进封装团队高管Dae Woo Kim在2024年度韩国微电子与封装学会年会上表示，三星成功制造了基于混合键合技术的16层堆叠HBM3内存，该内存样品工作正常，未来16层堆叠混合键合技术将用于HBM4内存量产。

相较现有键合工艺，混合键合无须在DRAM内存层间添加凸块，而是将上下两层直接铜对铜连接，可显著提高信号传输速率，更适合AI计算对高带宽的需求，还可降低DRAM层间距，减少HBM模块的整体高度。

但三星近期宣布将与台积电合作研发HBM4，这一“打脸”举动，也侧面反映出三星自身的先进制程工艺和封装技术尚不足以支撑其HBM4研发。

专家告诉记者，三星与台积电将共同开发一种特殊版本的无缓冲HBM4，技术路线与SK海力士的HBM4并不相同，这款产品取消了用于分配电压和防止电气问题的缓冲器，预计能提高40%的功率效率，并减少10%的延迟。三星计划在年底前完成HBM4的流片，为2025年年底的大规模生产奠定基础。

未来3年300mm晶圆厂设备投资将达4000亿美元

本报讯 9月26日，SEMI发布《300mm晶圆厂2027年展望报告》指出，从2025年到2027年，全球300mm晶圆厂设备支出预计将达到创纪录的4000亿美元。强劲的支出是由半导体晶圆厂的区域化以及数据中心和边缘设备对人工智能(AI)芯片日益增长的需求推动的。

2024年，全球300mm晶圆厂设备支出预计将增长4%，达到993亿美元，到2025年将进一步增长24%，首次突破1000亿美元，达到1232亿美元。预计2026年支出将增长11%，达到1362亿美元；2027年将增长3%，达到1408亿美元。

SEMI总裁兼首席执行官Ajit Manocha表示：“2025年全球300mm晶圆厂设备支出的预期增长幅度为创纪录的3年半导体制造投资奠定了基础。全球芯片需求正在推动设备支出，包括人工智能应用的前沿技术和由汽车与物联网应用驱动的成熟技术。”

预计到2027年，中国大陆将保持其作为全球300mm设备支出第一的地位，未来3年投资将超过1000亿美元。然而，预计投资将从2024年的450亿美元峰值逐渐减少到2027年的310亿美元。

预计韩国将排名第二，未来3年将投资810亿美元，以进一步巩固其在DRAM、HBM和3D NAND等存储领域的主导地位。预计中国台湾地区未来3年将投资750亿美元，位居第三，因为该地区的芯片制造商将在海外建造一些新的晶圆厂，3nm以下是其投资的主要驱动力。

从2025年到2027年，美洲地区预计投资630亿美元，而日本、欧洲和中东以及东南亚预计将在3年内分别投资320亿美

元、270亿美元和130亿美元。值得注意的是，由于旨在缓解对关键半导体供应担忧的政策激励措施，预计2027年这些地区的设备投资将比2024年增加1倍以上。

2025年至2027年间，Foundry设备支出预计将达到约2300亿美元，这得益于对先进节点的投资以及对成熟节点的持续支出。对2nm工艺的投资和2nm关键技术的开发，如全环绕栅极(GAA)晶体管结构和背面功率传输技术，对于满足未来高性能和节能计算需求至关重要，特别是对于人工智能应用。由于对汽车电子和物联网应用的需求不断增加，在成本效益高的22nm和28nm工艺上有望实现增长。

Logic和Micro领域预计将在未来3年率先扩大设备支出，预计总投资为1730亿美元。Memory位居第二，预计同期将贡献超过1200亿美元的支出，标志着另一个细分市场增长周期的开始。在Memory领域，DRAM相关设备的投资预计将超过750亿美元，而3D NAND的投资预计将达到450亿美元。

Power相关领域排名第三，预计未来3年投资将超过300亿美元，其中化合物半导体项目投资约140亿美元。同期，模拟和混合信号领域预计将达到230亿美元，其次是光电/传感器，为128亿美元。

作为SEMI Fab Forecast数据库的一部分，SEMI《300mm晶圆厂2027年展望报告》列出了全球420座设施和生产线，其中包括预计高概率将在2024年开始的未来4年内开始运营的79座设施。该报告反映了自上次2024年6月发布以来的169次更新和9个新的晶圆厂/生产线项目。

(文)

英特尔发布至强6性能核处理器

本报讯 记者姬晓婷报道：9月26日，英特尔正式发布面向数据中心市场的至强6性能核处理器，具备3个计算芯片单元，多达128个内核，相较于前代产品内核数实现大幅提升。在性能提升的同时，平均TCO相较于前代产品降低30%。其中至强6900P系列处理器(代号Granite Rapids-AP)宣布同日上市。

面向云计算领域，至强6处理器在每路核心数、单核性能、每瓦性能方面，相比上一代处理器均有提升，可帮助云服务提供商降低计算成本。与第五代英特尔至强可扩展处理器相比，至强6处理器拥有多达2倍的每路核心数，平均单核性能提升高达1.2倍，平均每瓦性能提升高达1.6倍，同等性能水平下平均节省30%的TCO。

面对数据中心对数据传输效率提升的要求，同日宣布上市的至强6900P系列处理器进行了一系列性能更新。该处理器最高配备128个内核，支持高达每秒6400MT的DDR5内存、每秒8800MT的MRDIMM内存、6条UPI 2.0链路(速率高达每秒

24GT)，96条PCIe 5.0或64条CXL 2.0通道、504MB的L3缓存，支持FP16数据格式的英特尔高级矩阵扩展(英特尔AMX)，可为AI和科学计算等内存带宽敏感型工作负载提供MRDIMM选择，且新增对CXL 2.0的支持。

在此次发布会上，英特尔一众生态伙伴分享了其基于至强6性能核产品的最新解决方案。其中，阿里云方升架构协同至强处理器持续迭代演进，推出最新一代磐久计算型服务器，以及基于至强6性能核的第九代阿里云英特尔平台企业级计算实例产品；超聚变带来了其首款支持CXL2.0+的内存池，可通过灵活分配与共享内存以充分提升系统效能；火山引擎基于至强6性能核处理器进行了架构和性能优化，CPU核心数翻倍并采用双单路架构设计，使其第四代云服务器实例在保障稳定性的同时大幅提升了计算密度和应用性能。

据了解，至强6处理器包括至强6700P、至强6500P等在内的多款产品，将于2025年第一季度发布。

佳能交付首台新型纳米压印光刻机

本报讯 日本佳能公司近日宣布，已成功交付位于美国得克萨斯州的半导体联盟得克萨斯电子研究所(TIE)交付了其最先进的纳米压印光刻(NIL)系统FPA-1200NZ2C。这标志着纳米压印光刻技术在半导体制造领域迈出了重要一步，为未来的芯片生产开辟了新的可能性。

佳能此次交付的FPA-1200NZ2C纳米压印光刻机，是该公司与铠侠(Kioxia)和大日本印刷公司合作开发的最新成果。该系统自去年10月推出以来，便以其独特的纳米压印技术和卓越的性能表现，吸引了业界的广泛关注。据佳能方面介绍，FPA-1200NZ2C能够实现最小14纳米的线宽图案化，支持5纳米制程逻辑半导体的生产，为芯片制造商提供了更为高效、精确的制造工具。

与传统光刻机通过光学成像将电路图案投影到晶圆上不同，佳能的纳米压印光刻技术采用物理手段，将已经设计好的电路图案模板通过机械加压直接“复印”到晶圆上的光刻胶中。这种“盖印章”式的生产方式省去了复杂的光学曝光系统，不仅降低了设备成本，还显著提高了生产效率。据佳能产品负责人透露，纳米压印光刻机的价格将比传统的极紫外光

(EUV)光刻机少一位数，且耗电量仅为后者的十分之一。

此次交付的FPA-1200NZ2C将被用于得克萨斯电子研究所的先进半导体研发和原型生产工作。该研究所由得克萨斯大学奥斯汀分校支持，成员包括英特尔、恩智浦、三星等全球领先的芯片公司，以及公共部门和学术组织。这些机构将共同利用这一先进设备，探索纳米压印光刻技术在半导体制造中的潜力和应用前景。

佳能光学产品副总裁岩本一典表示，公司对未来纳米压印光刻技术的发展充满信心，并计划在接下来的3~5年内，每年销售约10~20台此类设备。他强调，佳能将继续加大研发投入，与全球合作伙伴紧密合作，共同推动纳米压印光刻技术的商业化进程，为全球半导体产业的进步贡献力量。

尽管纳米压印光刻技术在成本和生产效率方面具有显著优势，但在实际应用中仍面临诸多挑战。例如，如何有效防止尘埃颗粒对芯片制造的影响，以及如何开发出与纳米压印技术相兼容的材料等。佳能方面表示，将积极应对这些挑战，与产业链上下游企业通力合作，共同推动纳米压印光刻技术的完善和发展。

(佳讯)