

证券研究报告  
电子/行业深度报告  
2025年3月26日

# 突破存算协同范式下的“存储墙”困境 —— HBM，训练侧/推理侧需求的共同焦点

证券分析师：唐仁杰 S0370524080002

行业评级：增持

# 摘要

- AI 算力倍增背景下，“算力墙”日益凸显。在AI训练和推理中，大量矩阵计算需快速内存访问。当内存无法跟上计算速度时，GPU利用率下降，系统性能无法线性随算力提升。内存带宽不足已成为AI领域亟待解决的关键瓶颈
- HBM (High Bandwidth Memory) 采用3D堆叠DRAM和宽总线并行访问设计，相较于传统平面DRAM、GDDR有着高带宽、低功耗、小封装/集成体积的特点
- 堆栈层数定容量，引脚数量、数据传输速率定带宽。从HBM1至HBM3e，单堆栈堆叠层数由最高8层至16层，单DRAM容量上限由16Gb(2GB)至32Gb(4Gb)，单颗HBM芯片容量由16GB扩展至64GB。相比于容量扩充，存储带宽上限由于总I/O数量扩充至1024，单PIN数据传输速率由1Gbps提升至最高9.6Gbps，单颗HBM存储带宽上限提升至1.2TB/s
- HBM制造核心工艺涉及TSV、微凸点的垂直互连，TSV工艺约占堆叠封装成本的37%。TSV达到通孔高深宽比理论极限，需通过DRIE+BOSCH工艺以形成垂直通孔；镀铜前，需沉积绝缘层及扩散阻挡层，等离子体增强化学气相沉积(PECVD)+正硅酸乙酯气体(TEOS)+硅烷(SiH4)在AR过高(如15:1)时侧壁薄膜变薄或中断，因此业界探索使用原子层沉积(ALD)来获得更佳的膜覆盖。但ALD的缺陷在于沉积速率慢，设备及材料成本较高。刻蚀完成TSV通孔后，需要进行金属填充以形成垂直导通。目前高性能HBM应用多采用电化学镀铜(ECP)填充TSV。最终，需要通过化学机械平坦化抛光(CMP)去除多余铜并露出TSV铜柱顶面，使其与表面平齐。CMP工艺要精确控制露铜(dishing)程度
- 键合工艺实现多层堆叠，成本敏感度较高；传统回流焊+毛细底填充工艺在随着HBM堆叠层数增加和凸点间距减小(芯片间隙可低至20μm)，毛细流动变得困难，容易残留空隙导致应力集中。TCP+NCF通过预先涂布底填充材料来进行键合，由于助焊剂底部清洗难度较大，业界开发出fluxless(无助焊剂)工艺。但热压键合工艺效率较低，无法批量操作。液态模塑底填充使用低粘度、高渗透性的树脂以保证在窄间隙中无死角充填，并且可批量，效率较高或成为主流。混合键合仍处于初步阶段，但高I/O优势显著。D2W方式可以在裸芯良率较低背景下提高封装良率，但效率较低；未来可能出现DRAM层间采用MR-MUF键合工艺，而通过D2W或W2W与基底逻辑Die键合的组合键合方式
- HBM市场现状及市场规模：根据TrendForce数据，从HBM供应市场角度，SK海力士、三星、美光独占HBM市场，且SK海力士、三星的集中度较高，且相关设备国产化率较低。IDTechEX预测，高性能计算(HPC)硬件市场2025-2035期间将保持13.6%的年化增长率，市场规模将超5810亿美元。其中，95%的HPC将使用HBM技术以满足性能需求，HBM单位销售额预计至2035年，将较2024年增长15倍
- 相关公司：拓荆科技(688072.SH)：薄膜沉积设备+混合键合、盛美上海(688082.SH)：电镀设备达到国际水平、华海清科(688120.SH)：磨削、CMP及清洗一体化
- 风险提示：HBM工艺复杂，涉及产业链设备及产业链仍以海外企业为主，地缘影响较大；AI产业链发展不及预期，对高性能存储需求下降。



www.jyzq.cn

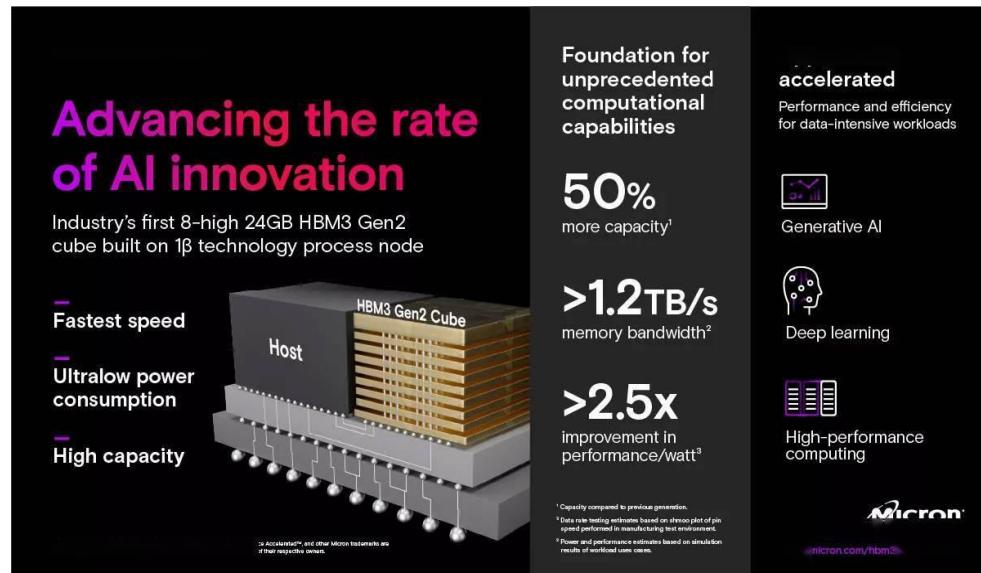


全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 目录



一、“存储墙” - 在算力倍升背景下，存储带宽成为“阿姆达尔短板”

二、HBM: 高带宽/高集成密度内存的核心优势

三、HBM核心要素：垂直互连-TSV及微凸点、UBM工艺

四、键合工艺：实现多层堆叠，多种键合工艺及混合键合

五、HBM现阶段市场规模及工艺趋势

六、相关公司

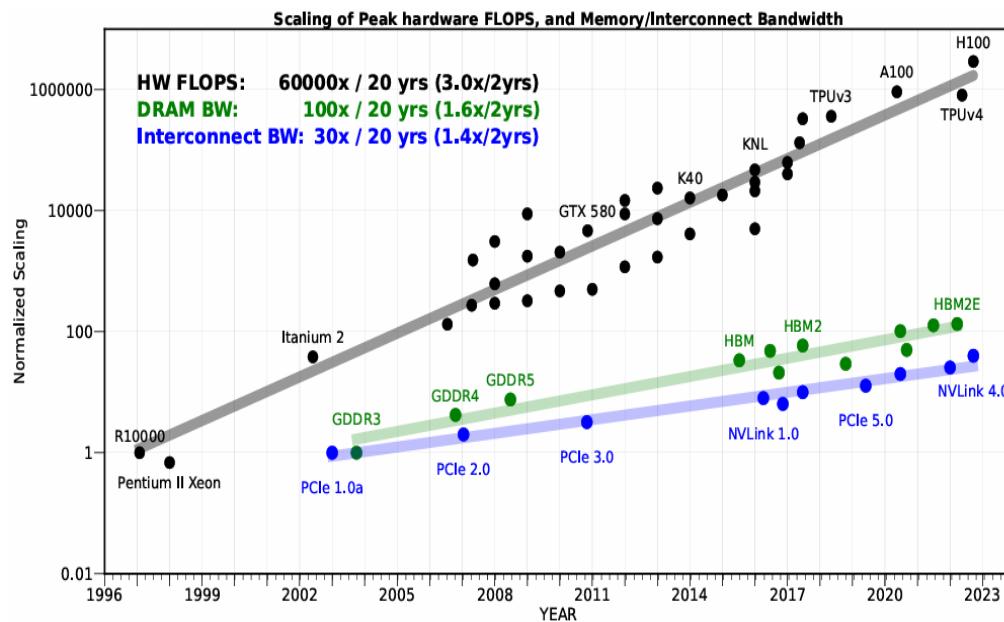
风险提示：HBM工艺复杂，涉及产业链设备及产业链仍以海外企业为主，地缘影响较大；AI产业链发展不及预期，对高性能存储需求下降

# “存储墙” - 在算力倍升背景下，存储带宽成为“阿姆达尔短板”

- 冯·诺依曼架构下，存储器带宽、容量成为了硬件算力效率的“阿姆达尔短板”，处理器性能提升被内存传输速率所制约，内存带宽年增速仅约15%。随着GPU/AI加速器算力猛增，内存无法及时“喂饱”计算单元，导致性能受限。在AI训练和推理中，大量矩阵计算需快速内存访问。当内存无法跟上计算速度时，GPU利用率下降，系统性能无法线性随算力提升。内存带宽不足已成为AI领域亟待解决的关键瓶颈。

图：A100 存储架构与 CPU 类似，均配备私有缓存、共享缓存(HBM)

图：峰值算力 2 年增长 3 倍，而存储带宽及互连带宽仅增长 1.6、1.4 倍



数据来源：《AI and Memory Wall》，金元证券研究所

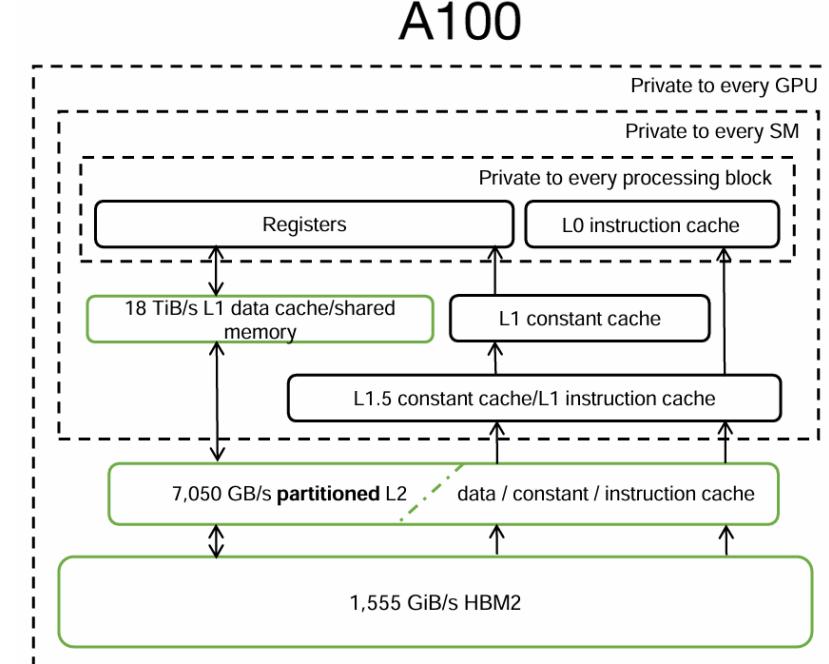


www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。



数据来源：NVIDIA，金元证券研究所

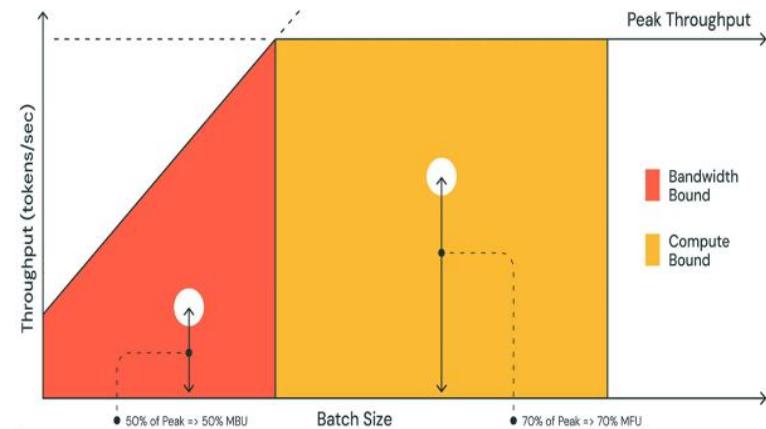


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

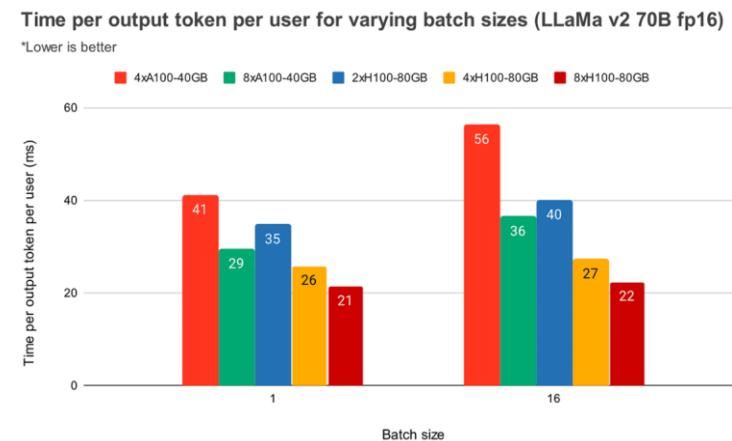
# “存储墙”对AI训练/推理的影响

- 大模型受制于内存容量：百亿/万亿级参数模型训练和推理需要数百GB到数TB内存。当前GPU搭载的高带宽显存容量增长有限，很多模型无法全部加载到单卡内存，不得不拆分到多卡，增加通信开销。
- 带宽瓶颈限制Inference性能：即使GPU峰值算力很高，但若内存带宽不足，实际性能大打折扣。经典屋顶檐模型（Roofline Model）详细阐述了“存储墙”及“算力墙”。Facebook研究人员在Palm论文中引入了模型FLOPs利用率(MFU)来衡量训练中算力的饱和程度，但对于推理，更相关的指标是模型带宽利用率(MBU)。MBU定义为实际消耗的内存带宽除以硬件峰值带宽。在纯memory-bound情况下，MBU接近100%，此时进一步增加算力无济于事，只有提高带宽或减少数据量才能加速。很多大模型推理场景下MBU都很高，而MFU偏低。例如LLama-2 70B在批量1时GPU计算利用较低，但HBM带宽几乎被吃满。工程上会采用批处理等手段提高每次计算处理的token数，从而用额外吞吐换取更高效率。但在极端情况下（如单用户长文本生成），序列步骤无法并行，此时很难摆脱带宽瓶颈。Databricks实测H100集群相比A100在相同模型上Batch Size 1 延迟降低36%，在Batch Size 16延迟降低52%，但简单“堆砌算力”（增加更多卡）效果并不显著。

图：MBU、MFU 的 Roofline 模型



图：4x、8x 系统并不会显著降低延迟，但带宽增大延迟降低明显



数据来源：Databricks、金元证券研究所



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

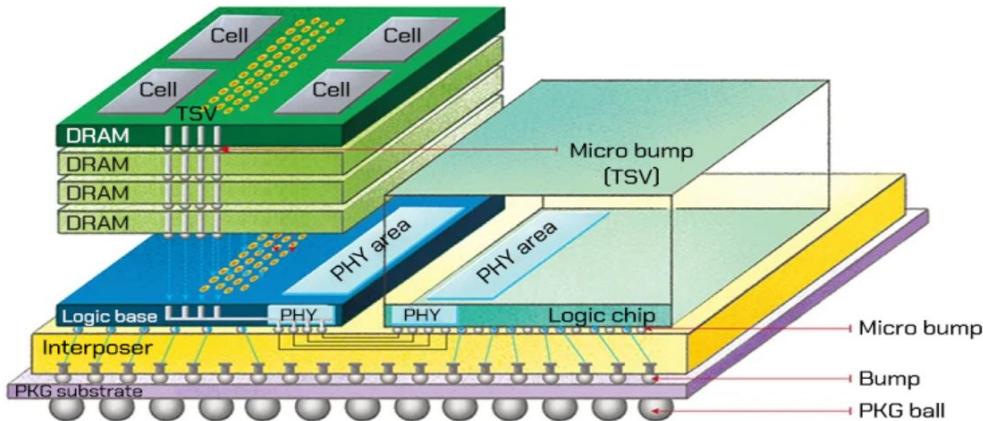


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# HBM:高带宽/高集成密度内存的核心优势

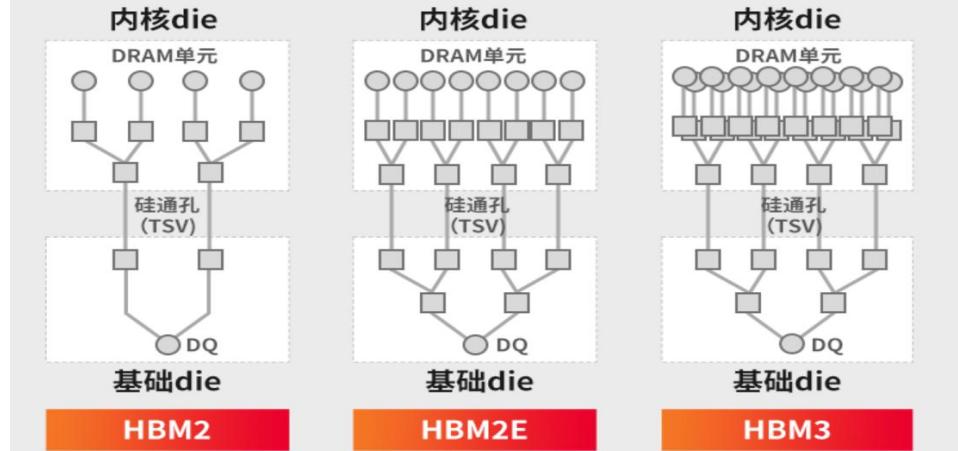
- **高带宽:** HBM (High Bandwidth Memory) 采用3D堆叠DRAM和宽总线并行访问设计，每颗HBM堆叠存储器拥有1024-bit总线，相比传统DDR/GDDR显存带宽大幅提升。例如HBM2提供>256GB/s带宽，HBM3可达819GB/s以上
- **低功耗:** HBM通过降低工作频率、提高总线并行度，实现更高能效。每比特传输能耗显著低于GDDR等显存（约降低30-50%）。这意味着在提供同等带宽下，HBM耗电更少，有利于控制高性能芯片的功耗和发热
- **集成密度提升:** HBM采用硅中介层(Interposer)的2.5D集成，将多个DRAM芯片垂直堆叠并紧贴处理器封装。相比分立显存芯片围绕PCB布局，HBM封装占用空间小、连线距离短，信号延迟低且可靠性更高。这使得在有限封装面积内提供大容量、高带宽存储成为可能。

图：HBM 基础结构及封装体



图：HBM 各代基础数据传输路径

## 各代HBM产品的数据传输路径配置



数据来源：SK Hynix, 金元证券研究所



www.jyzq.cn



全国统一客服电话：95372

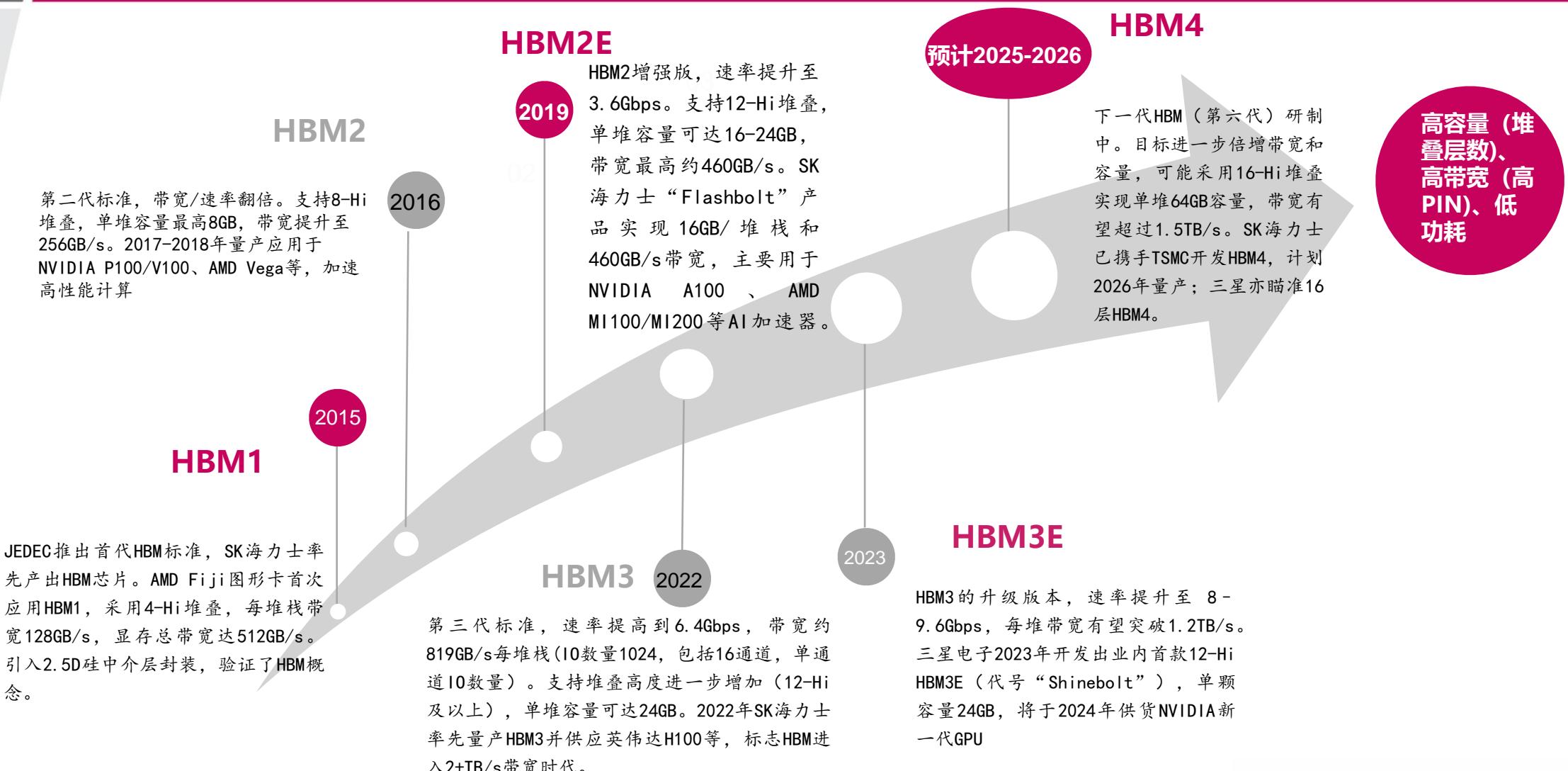
此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

数据来源：金元证券研究所, SK Hynix



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# HBM技术演进历程



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

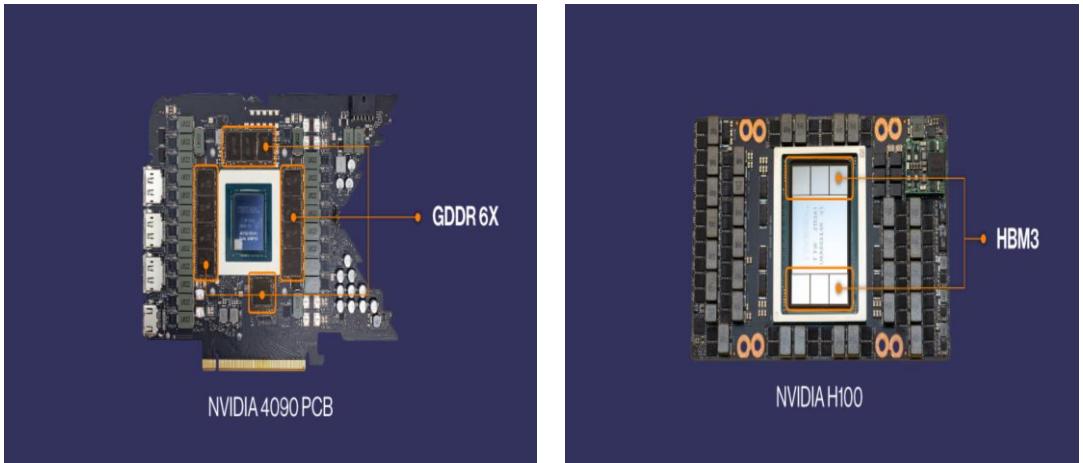


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# HBM vs GDDR：性能、能效、封装体积差异

- 带宽：单颗HBM的带宽虽然工作频率较低，但凭借超宽总线显著领先单颗GDDR。典型GDDR6显存单颗带宽约64GB/s，但需要多颗并行才能达到数百GB/s；而HBM2e单堆栈（8-Hi）带宽已超过460GB/s。HBM3在NVIDIA H100上组成5120-bit总线，整卡内存带宽高达2 TB/s以上。
- 功耗：HBM能在较低频率下提供高带宽，因此每单位带宽功耗大幅低于GDDR。研究表明HBM比GDDR在类似带宽下功耗低30-40%。这对数据中心GPU等功耗受限场景尤为重要，HBM降低了内存系统功耗占比。
- 封装/体积：HBM通过芯片堆叠实现高密度封装，每堆栈可集成多达8-16层DRAM芯片，总容量达数十GB。而GDDR需多颗分散布局，占用PCB空间大且布线复杂。在高性能计算模块中，用HBM替代传统显存可显著缩小体积并提高集成度。

图：GDDR 6X 封装体积较大，且同等数量下带宽远低于 HBM



数据来源：FiberMall，金元证券研究所



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：不同 GPU 在使用 GDDR 与 HBM 的存储性能对比

GPU	Memory Type	Memory Bus Width	Memory Bandwidth
RTX 6000 Ada	GDDR6	384-bits	960 GB/s
GeForce RTX 4090	GDDR6X	384-bits	1008 GB/s (1 TB/s)
NVIDIA L40S	GDDR6	384-bits	864 GB/s
NVIDIA A800 40GB Active	HBM2	5120-bits	1555 GB/s (1.5 TB/s)
NVIDIA H100 80GB PCIe	HBM2e	5120-bits	2039 GB/s (2 TB/s)
NVIDIA H100 80G SXM5	HBM3	5120-bits	3350 GB/s (3.35 TB/s)

数据来源：金元证券研究所

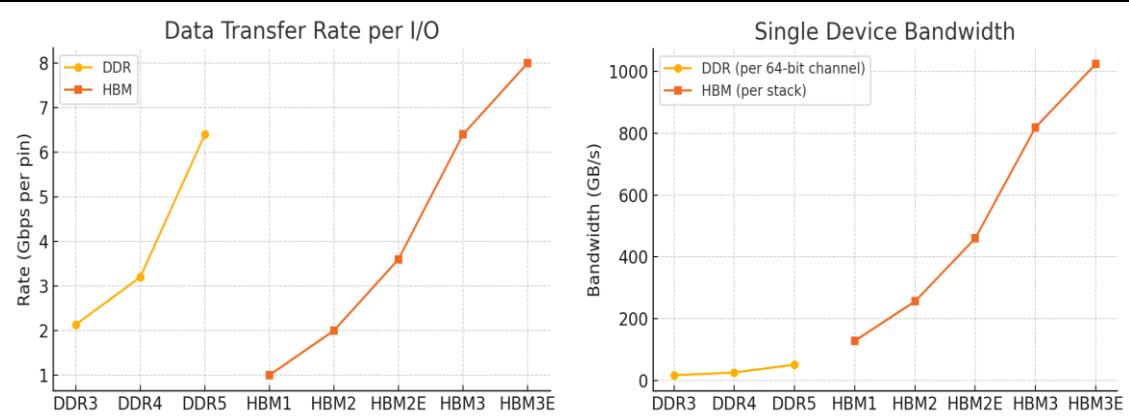


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# HBM vs 平面 DRAM：更多通道、更高带宽

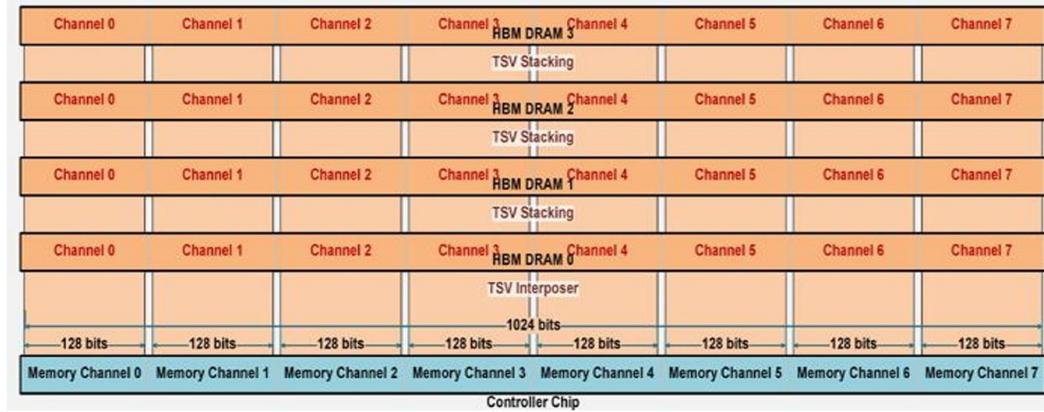
- I/O速率：**HBM每层由裸die (DRAM) 组成，DDR4标准最高支持约3200 MT/s (即每针3.2 Gb/s)；DDR5起步即4800 MT/s (每针4.8 Gb/s)，并规划提升至8400 MT/s左右 (每针8.4 Gb/s)。相比之下，HBM技术单针速率虽起步较低，但提升迅猛：HBM2E约3.6 Gb/s，HBM3提升到6.4 Gb/s；最新HBM3E进一步提高单针速率，上限约8.0~9.6 Gb/s。**HBM自问世以来单针速率从HBM1约1 Gb/s提高到HBM3的6.4 Gb/s，再到HBM3E规划的8+ Gb/s，几乎每代翻倍**
- 总带宽：**DDR系列受限于通道位宽，相同代际下总带宽远低于HBM。单通道64位宽的DDR4-3200带宽约25.6 GB/s，DDR5-6400约51.2 GB/s。HBM采用超宽1024位总线，单颗HBM2E堆栈 (1024位@3.6 Gb/s) 带宽可达460 GB/s；；HBM3提高到819 GB/s (1024位@6.4 Gb/s)，HBM3E则进一步突破，每堆栈带宽可达1 TB/s以上 (1024位@8 Gb/s约=1,024 GB/s；@9.6 Gb/s则≈1.23 TB/s)

图：平面 DRAM vs HBM 数据传输速率、带宽



数据来源：金元证券研究所

图：相比于传统平面 DRAM，HBM 拥有更高 I/O，多通道



数据来源：《A Study on the Effectiveness of Underfill in the High Bandwidth Memory with TSV》，金元证券研究所



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 堆栈层数定容量，引脚数量、数据传输速率定带宽

- HBM容量(Capacity)≈单堆栈层数 × 单层DRAM容量；HBM存储带宽≈总I/O数量位宽(位) × 数据传输速率(DDR, Gbps) / 8
- 从HBM1至HBM3e，单堆栈堆叠层数由最高8层至16层，单DRAM容量上限由16Gb(2GB)至32Gb(4Gb)，单颗HBM芯片容量由16GB扩展至64GB
- 相比于容量扩充，存储带宽上限由于总I/O数量扩充至1024，单PIN数据传输速率由1Gbps提升至最高9.6Gbps，单颗HBM存储带宽上限提升至1.2TB/s

版本	堆栈层数	数据传输速率(每PIN)	总I/O数量(位宽)	通道数	单通道位宽	单DRAM容量上限	单堆栈容量上限	存储带宽(单堆栈)
HBM1	4层或8层	1 Gbps	1024	8通道	128位/通道	16Gb	16GB	128GB/s
HBM2	4层或8层	2 Gbps	1024	8通道	128位/通道	16Gb	16GB	256GB/s
HBM2e	8层或更高	3.6 Gbps	1024	8通道	128位/通道	24Gb	24GB	460.8GB/s
HBM3	12层-16层	6.4 Gbps	1024	16通道	64位/通道或划分位32个pseudo-channels, 单通道32位	32Gb	64GB	819.2GB/s
HBM3e	12层-16层	8-9.6 Gbps	1024	16通道	64位/通道或划分位32个pseudo-channels, 单通道32位	32Gb	64GB	1.2TB/s

数据来源：金元证券研究所



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

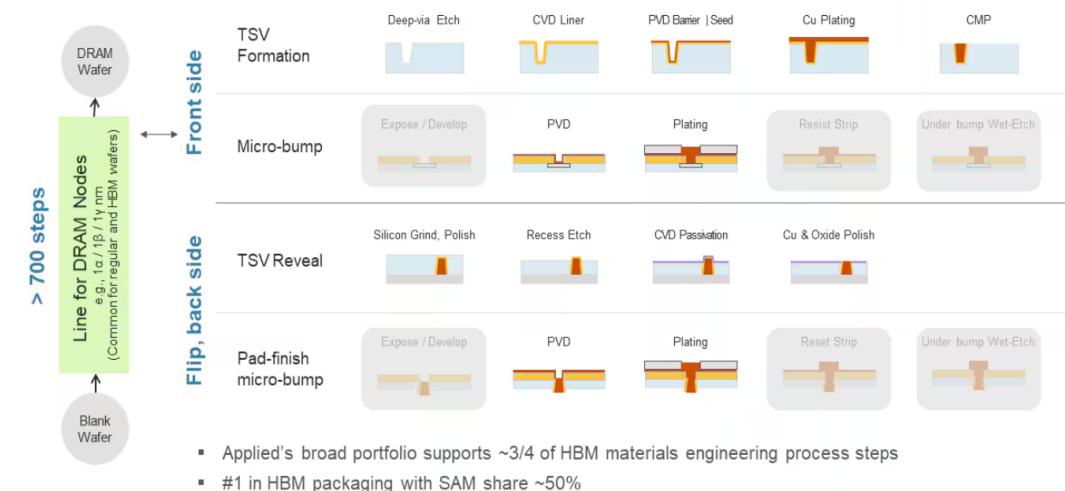


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# HBM核心要素：垂直互连-TSV

- HBM采用硅通孔（TSV）技术将堆叠的DRAM芯片垂直互连，常用的TSV金属填充材料包括铜（Cu，电阻率更低，但填充时易扩散）或钨（W）。工艺需先在硅中刻蚀高深宽比的通孔，再沉积介电衬垫和金属种子层，随后进行电镀填充并CMP抛光。TSV形成使每层芯片通过垂直“铜线”连接成信号通道，实现多芯片堆叠的数据传输。
- 传统堆叠DRAM的封装将导线连接到每个Die的侧面，但随着布线密度提升，寄生参数（如寄生电容、电阻）导致堆叠芯片的延迟增大、功耗上升。采用TSV封装可以有效增加引脚（IO数量，传统DRAM一般为X4或X16，通过TSV堆叠，HBM引脚提升至1,024，单通道128位）的同时，缩短传输路径，提升传输速率，且功耗降低。

图：TSV正面及背面工艺：TSV刻蚀、填充及 CMP



数据来源：Applied Material, 金元证券研究所



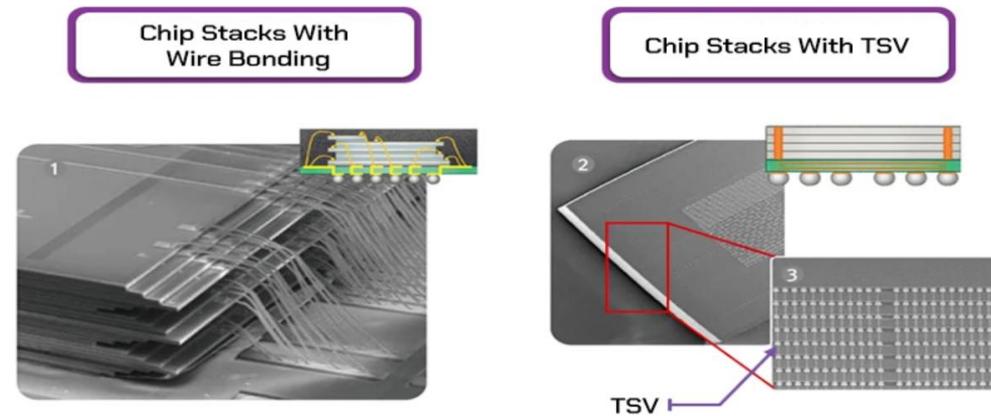
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：传统引线键合堆叠 vs TSV



数据来源：SK Hynix, 金元证券研究所

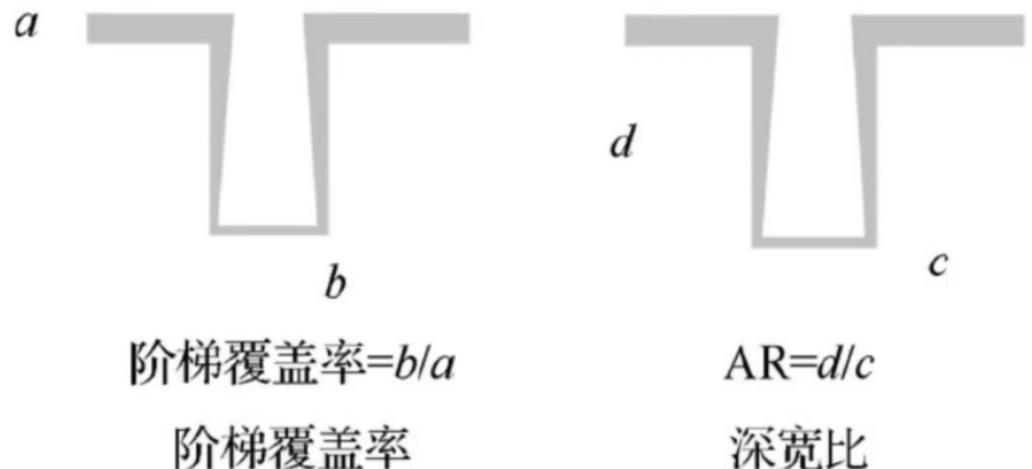


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

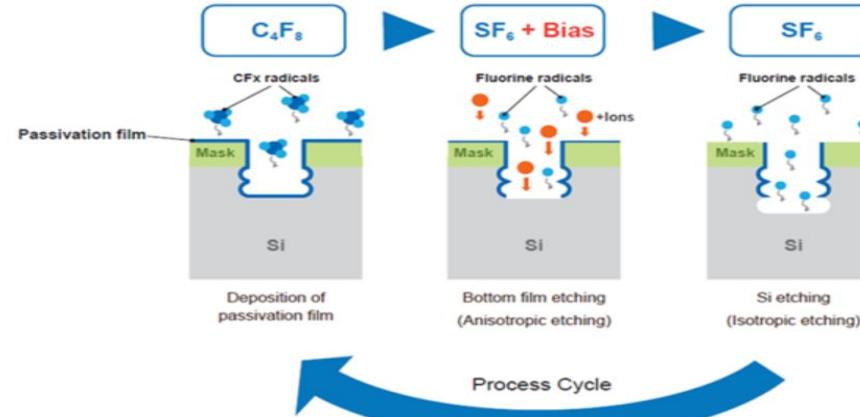
# TSV关键技术：高深宽比刻蚀，铜填充及CMP

- TSV刻蚀及沉积薄膜涉及两项关键指标：深宽比 (Aspect Ratio, TSV深度与直径比值) 及阶梯覆盖率 (Step Coverage, 跨台阶处的膜层厚度与平坦处膜层厚度比值)
- 高深宽比刻蚀：常用工艺包括反应离子刻蚀 (Reactive Ion Etching) 与激光钻孔 (Laser Drill)。激光钻孔具有成本优势，但在精度、批量刻蚀、热预算等方面与RIE差距较大。随着刻蚀工艺发展和TSV密度提升，深反应离子刻蚀 (Deep Reactive Ion Etching, DRIE) 逐步取代RIE。DRIE与RIE的原理基本相同，均基于氟基气体通过化学、物理作用进行刻蚀，但RIE的各向异性不如DRIE。DRIE通过钝化、刻蚀交替 (Bosch) 方式，采用高密度等离子体 (感应耦合等离子体, ICP)，刻蚀速率高达 $20\text{ }\mu\text{m/min}$ 。

图：阶梯覆盖率、深宽比



图：DRIE 刻蚀，基于 Bosch 工艺的高选择比刻蚀



数据来源：金元证券研究所



www.jyzq.cn 全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

数据来源：SAMCO, 金元证券研究所

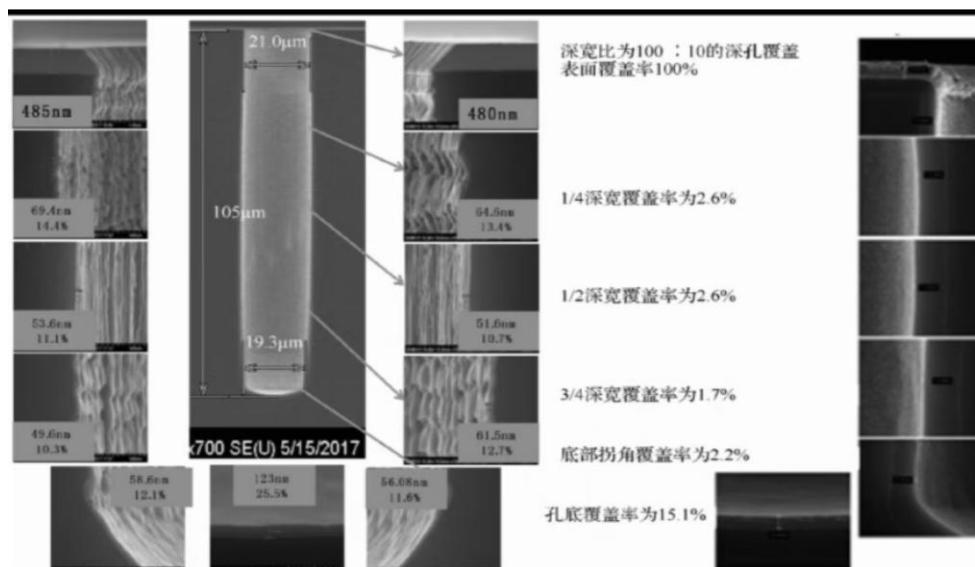


金元证券股份有限公司  
GOLDSTATE SECURITIES CO., LTD.

# TSV关键技术：高深宽比刻蚀，铜填充及CMP

- TSV在铜（或其他金属）填充前需要沉积绝缘层对Si衬底进行完全电气隔离。沉积绝缘层需要考虑热膨胀系数（CTE）匹配后续加工工艺，及台阶覆盖率(step coverage)问题，在高深宽比TSV中，绝缘层(如PECVD氧化硅)和阻挡层(金属薄膜)必须均匀覆盖侧壁，否则后续电镀容易出现空洞。
- 通常绝缘层选择二氧化硅或氮化硅，厚度只有纳米级与微米量级。等离子体增强化学气相沉积(PECVD)+正硅酸乙酯气体(TEOS)+硅烷(SiH<sub>4</sub>)在AR过高(如15:1)时侧壁薄膜变薄或中断，因此业界探索使用原子层沉积(ALD)来获得更佳的膜覆盖。但ALD的缺陷在于沉积速率慢，设备及材料成本较高。

图：随着深宽比增加，孔内覆盖率急剧减少



数据来源：《集成电路系统级封装》，金元证券研究所



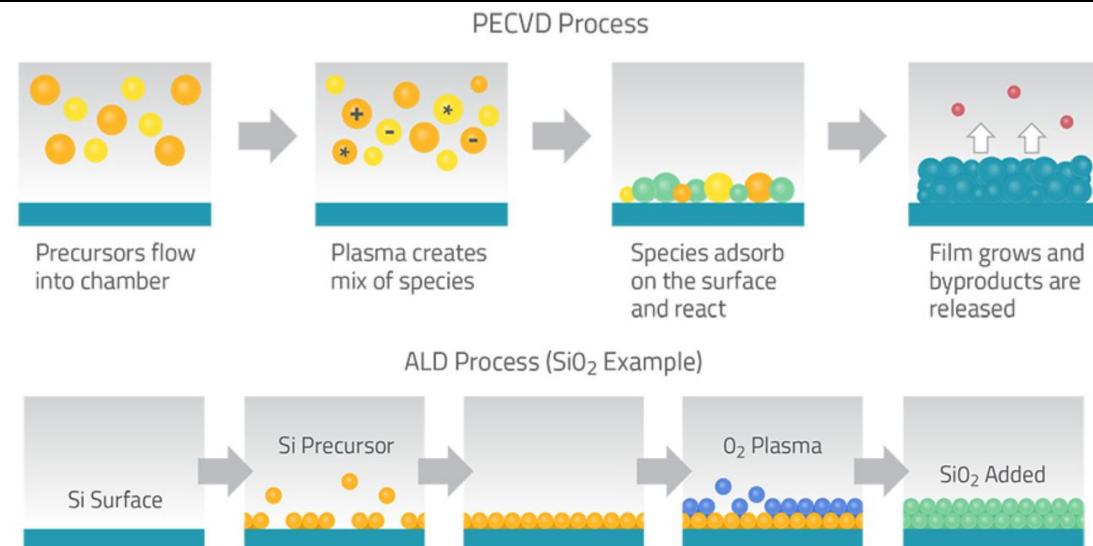
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：PECVD 前驱体、等离子体、副产物和其他分子碎片和物质都在腔室中漂浮，很难控制在“原子级”



数据来源：Lam Research, 金元证券研究所

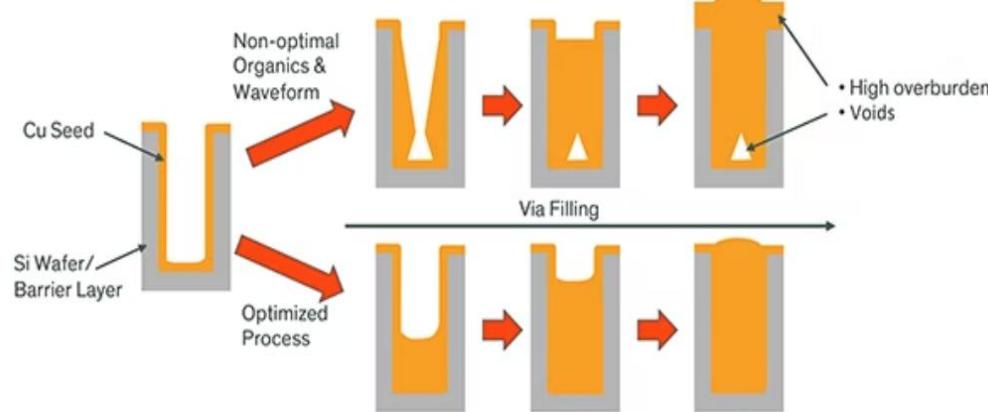


金元证券股份有限公司  
GOLDSTATE SECURITIES CO., LTD.

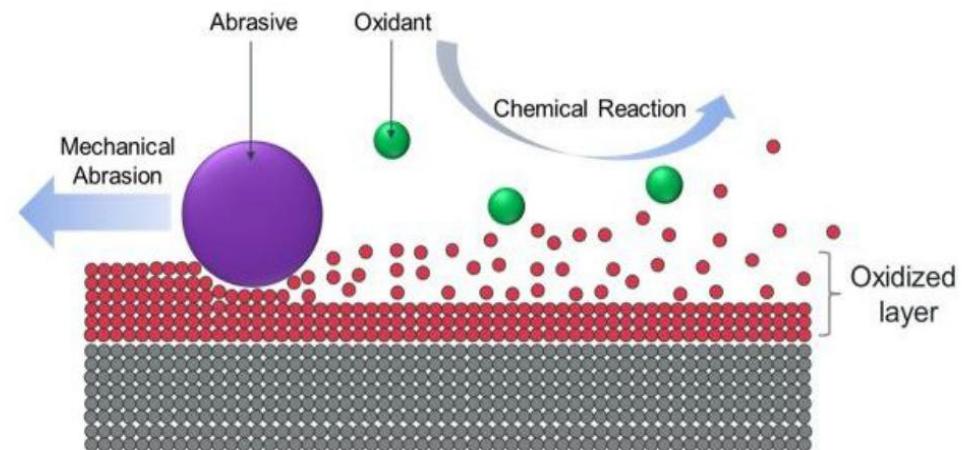
# TSV关键技术：高深宽比刻蚀，铜填充及CMP

- 刻蚀完成TSV通孔后，需要进行金属填充以形成垂直导通。目前高性能HBM应用多采用电化学镀铜(ECP)填充TSV。铜填充工序包括：溅射一层种子铜(seed)作为电镀电极，然后在硫酸铜电镀液中沉积铜直至填满通孔。为避免产生中空或接缝缺陷，必须确保种子层在高深宽比孔内连续覆盖
- 铜填充后，需要通过化学机械平坦化抛光(CMP)去除多余铜并露出TSV铜柱顶面，使其与表面平齐。CMP工艺要精确控制露铜(dishing)程度：既要充分去除过量铜，又避免过度抛光造成凹陷或破坏周围介质层。典型流程是在铜上方有一层阻挡金属(如TaN)，CMP以高选择性停止在阻挡层上，然后再轻微抛光露出铜。关键设备包括CMP抛光机和在线测厚/终点检测系统。

图：传统Cu填充“空洞”问题，及“Bottom-up”填充方法



图：化学机械平坦化抛光(CMP)去除多余铜并露出TSV铜柱顶面



数据来源：Dupont, 金元证券研究所



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

数据来源：Dupont, 金元证券研究所



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# HBM核心要素：垂直互连-微凸点及UBM

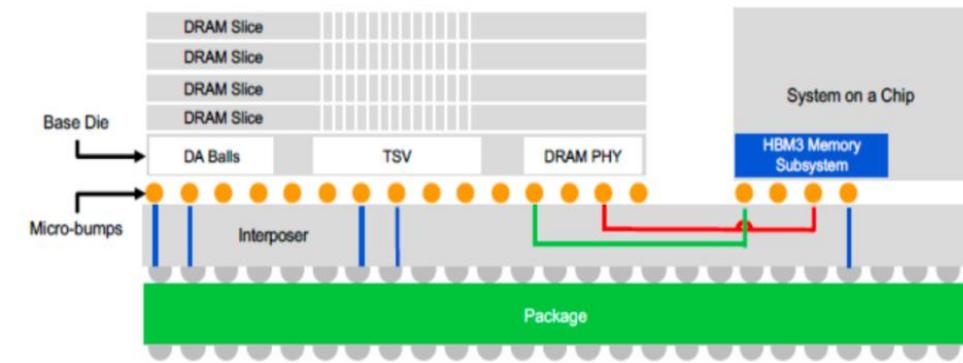
- HBM垂直堆叠另一关键互联是微凸点 (micro bump) 技术，用于实现芯片间、Die到硅中介层 (interposer)间的电气连接。微凸点通常指直径20–50μm的小焊点，相比传统封装焊球（直径数百微米）要小得多，从而支持更紧密的I/O间距。

图：微凸点支撑高 I/O

	HBM1	HBM2	HBM2E	HBM3	HBM3E
Capacity	Die Density	1Gb	8Gb	16Gb	16Gb
	Capacity / Max Stack	1GB (4 Hi)	8GB (8 Hi)	16GB (8 Hi)	24GB (12 Hi)
Bandwidth / Performance	Total IO/Cube	1024 (8Ch x 128 I/O)	1024 (8Ch x 128 I/O)	1024 (8Ch x 128 I/O)	1024 (16Ch x 64 I/O)
	IO Data Speed	1.0Gbps	2.4Gbps	3.6Gbps	5.6Gbps
Package	Max. Bandwidth	128GB/s	307GB/s	460GB/s	717GB/s
	PKG Size (X,Y)	Under 8x12mm <sup>2</sup>	8x12mm <sup>2</sup>	10x11mm <sup>2</sup>	11x11mm <sup>2</sup>
# of Micro Bumps	4942	4942	6303	7775	7775
	Micro Bump Pitch	96 x 55um	96 x 55um	96 x 55um	96 x 110um

数据来源：《Advanced Packaging Technologies in Memory Applications for Future Generative AI Era》，金元证券研究所

图：微凸点用于 HBM 及 xPU 的芯片间、Interposer 电气连接



数据来源：Semiconductor Engineering，金元证券研究所

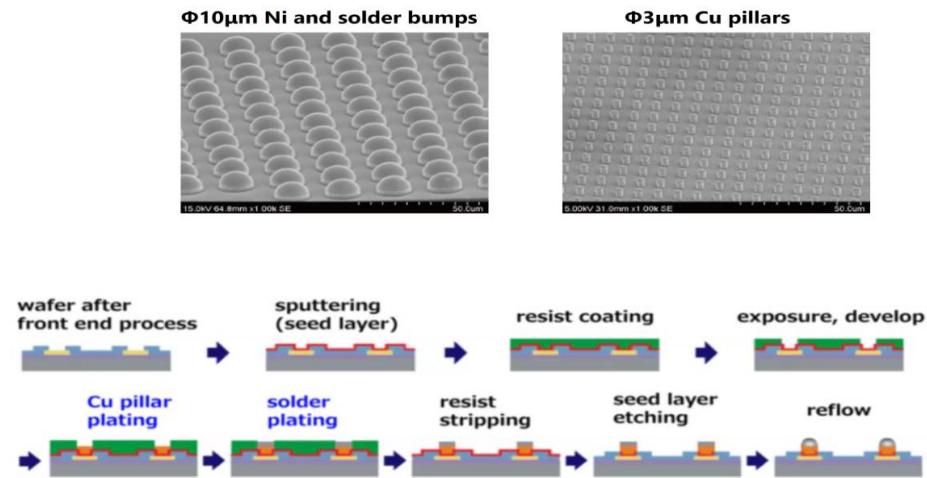
# HBM核心要素：垂直互连-微凸点及UBM

- 在微凸点电镀形成前，需在焊盘上制作UBM金属层作为焊料的结合界面。传统flip-chip中，常用钛/镍/铜/金等多层：钛促进粘附，镍作为主要阻挡/润湿层，表面再镀一层薄金防止氧化。在HBM早期工艺中，一些微凸点采用了化学镀Ni (P)+沉金工艺，其优点是Ni在高温下抑制Sn向硅扩散，并在焊料耗尽Cu后，依然提供可润湿表面，保证连接不断裂。Ni/Au界面可靠性良好，在温循等测试中能减缓焊点金属间化合物生长速度。然而，过厚的Au可能融入焊料引起“金脆”（使焊点变脆易裂）。同时，Ni层本身偏脆，在反复热循环和机械应力下可能产生微裂纹。而且对几十微米 pitch 的成千上万凸点来说，电镀Ni/Au的工艺时间和成本显著，并可能因电流分布不匀导致凸点高度不均
- Cu柱及无Ni方案：为实现更细间距和降低成本，许多先进封装转向Cu柱微凸点。Cu具有良好导电导热特性，可直接作为凸点主要材料，顶部只需很薄的一层Sn或表面处理用于接合。这样UBM可以简化为铜种子层+薄阻挡（例如Ti或TaN）即可，无需厚Ni层
- 微凸点UBM制造的关键设备包括：高精度晶圆溅射机用于沉积UBM (PVD, 可溅射Ti、Cu等薄膜)；膜光刻设备，用于形成几十微米厚度光阻图案（通常使用步进投影机或制程较宽松时用贴膜对准机）；电镀设备，用于高产能地在晶圆上同时电镀高数量级微凸点

图：化学镀 Ni (P)+沉金工艺 vs Cu 柱及无 Ni 方案

bump structures	Ni + Solder (Au)	Cu pillar + Solder (Au)
bump pitch	10μm Min. ~	6μm Min. ~
bump diameter	5μm Min. ~	3μm Min. ~
bump height	aspect ratio ≤1	aspect ratio 1~1.5
solder plating	Sn-Ag-Cu, Sn-Ag, Sn Sn-58Bi, Sn-52In, Sn-Bi-In	Sn-Ag-Cu, Sn-Ag, Sn Sn-58Bi, Sn-52In, Sn-Bi-In

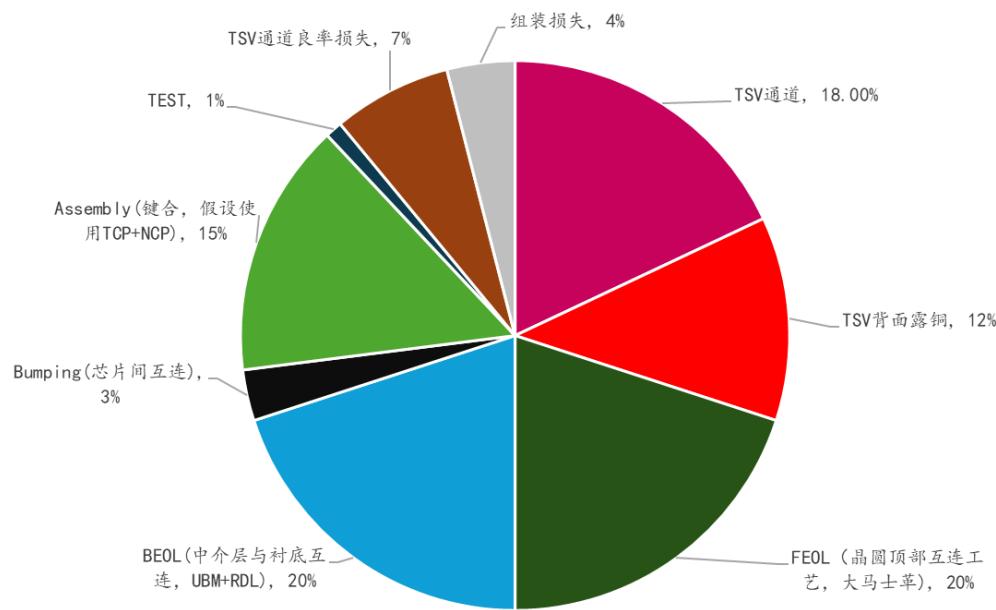
图：化学镀 Ni (P)+沉金工艺 vs Cu 柱及无 Ni 方案



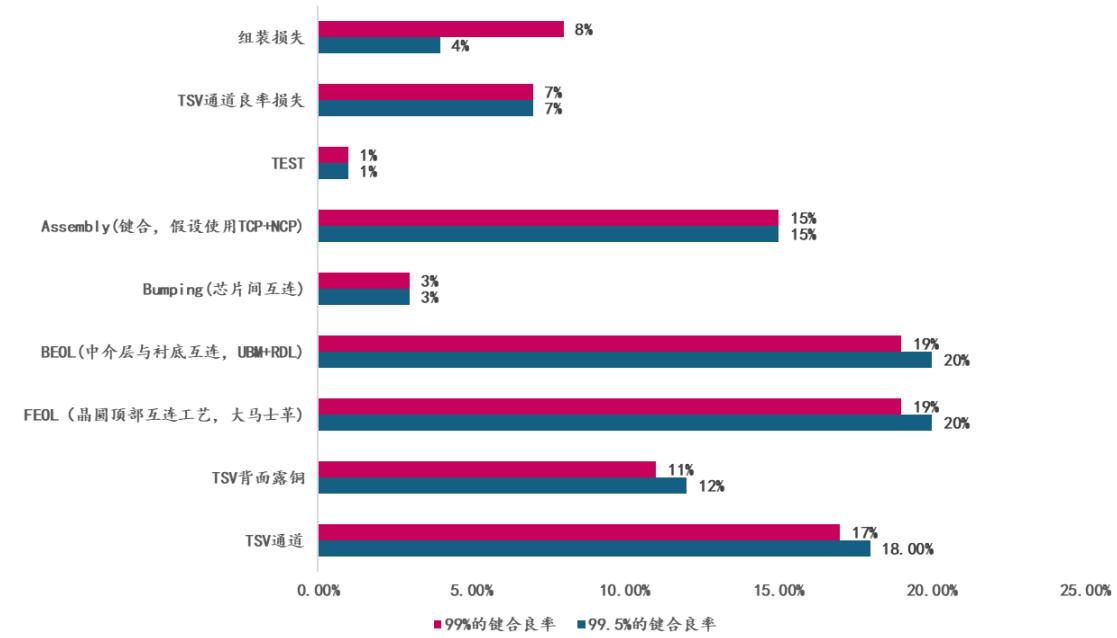
# HBM封装成本：TSV及芯片键合占主导地位

- 假设裸Die为检测合格Die(KGD)，芯片组装良率为99.5%，TSV通道创建良率为98%的前提下，TSV通道+TSV露铜成本占3D堆叠成本的30%，考虑TSV良率损失后，单TSV工艺成本占3D堆叠成本的37%。
- 敏感性分析下，假设芯片组装（键合）良率下降0.5%，3D组装良率损失翻倍，从总成本的4%上升至8%。

图：TSV 通道+TSV 背面露铜+TSV 通道良率损失合计占 3D 堆叠成本 37%



图：99%的键合良率与 99.5%键合良率敏感性分析



数据来源：3DinCities, 金元证券研究所



www.jyzq.cn 全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有,未经许可任何单位或个人不得复制、翻印。

数据来源：3DinCities, 金元证券研究所

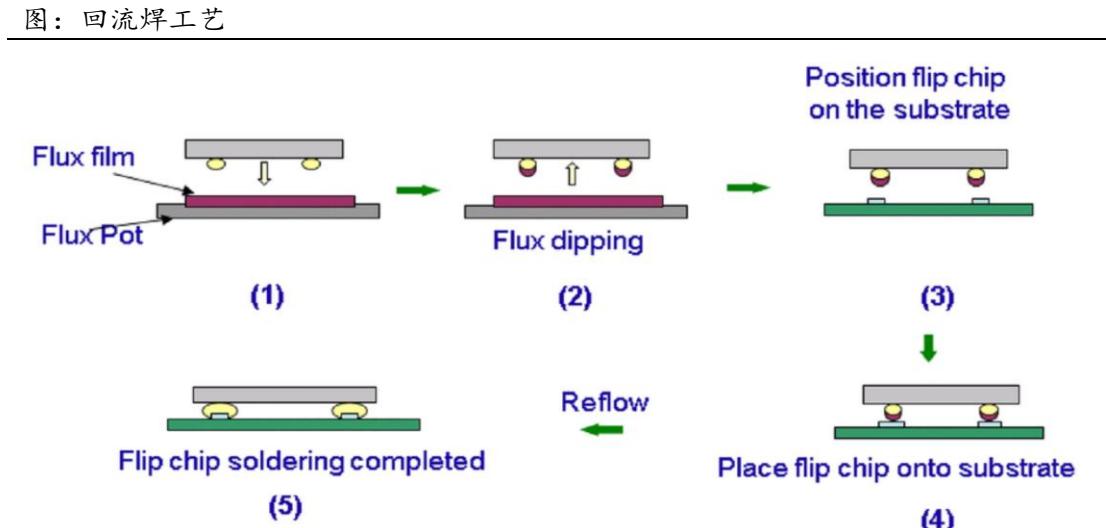


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 键合工艺：实现多层堆叠

- 实现多层堆叠，需要将微凸点成功连接，需要经过芯片对接焊接工艺。常见的两种工艺路径是大批量回流焊接（mass reflow）和逐芯片热压键合（TCB, Thermo-Compression Bonding）
- 回流焊接工艺：传统倒装芯片封装广泛采用的方法。将带有焊料凸点的芯片与另一芯片/基板对准后，一起放入加热炉中，使所有凸点焊料同时熔化并润湿连接，然后冷却凝固形成键合。回流焊允许一次性连接大量凸点，因而效率高、成本低，非常适合凸点间距在 $50\mu\text{m}$ 以上、对准公差较大的封装。传统回流法会遇到全局热膨胀失配的问题，芯片在熔融焊料上发生位移，冷却后可能出现对准偏差和短路失效
- 热压键合工艺：TCB是一种通过局部加热加压来焊接凸点的方法。其过程是使用高精度键合头(BH)逐颗拾取芯片，精确对准其微凸点与目标晶圆/基板上的对应垫，再施加压力和局部加热，使接触的焊料在加压状态下受热熔化并填充分隔，随后冷却形成焊点。这种方法避免了整体回流时的大尺度热应力影响，仅对一个芯片区域进行受控焊接，因而能够更好地解决热胀不匹配导致的翘曲和错位问题

图：TCB 工艺



数据来源：《Optimization of reflow profile for copper pillar with SAC305 solder cap FCCSP》，金元证券研究所

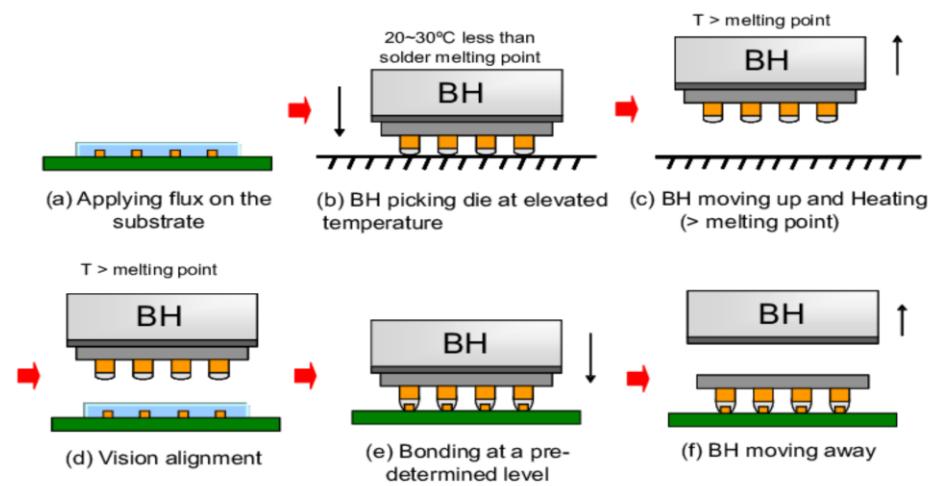


www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。



数据来源：《A high throughput and reliable thermal compression bonding process for advanced interconnections》，金元证券研究所



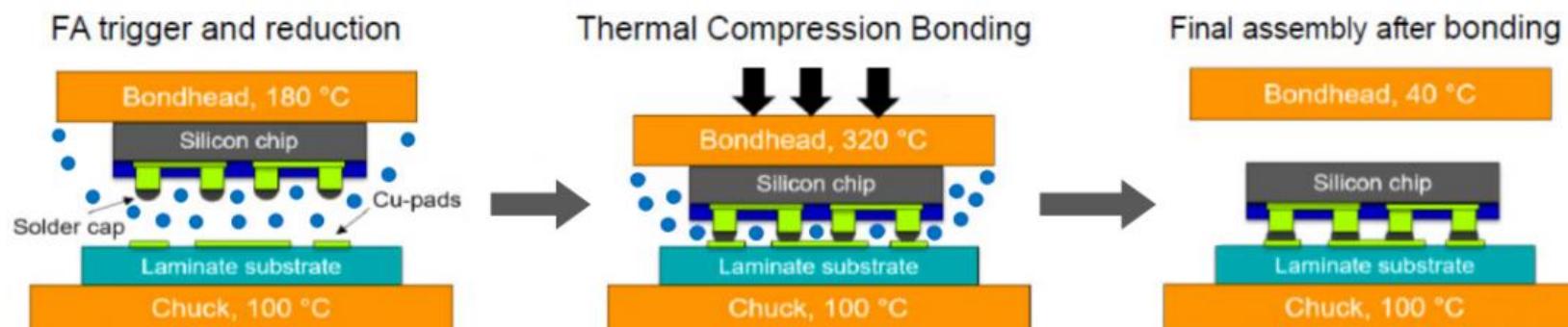
金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 键合工艺：无助焊剂TCB

- TCB工艺缺陷：产能较低，需要逐颗芯片拾取加压，无法像回流那样批量处理；其次，助焊剂残留问题突出，TCB通常需在焊接前给凸点蘸取助焊剂（flux）以去除氧化物。焊接后这些助焊剂残留必须清洗干净，否则会影响焊点可靠性。然而在细间距下，彻底清除残留难度很大。K&S公司开发了在TCB键合头内引入甲酸气氛的方法，可在压焊同时清洁铜垫表面而无需助焊剂，既提高了通孔可靠性又简化了流程

图：Fluxless TCB 工艺流程

## Fluxless TCB 工艺流程

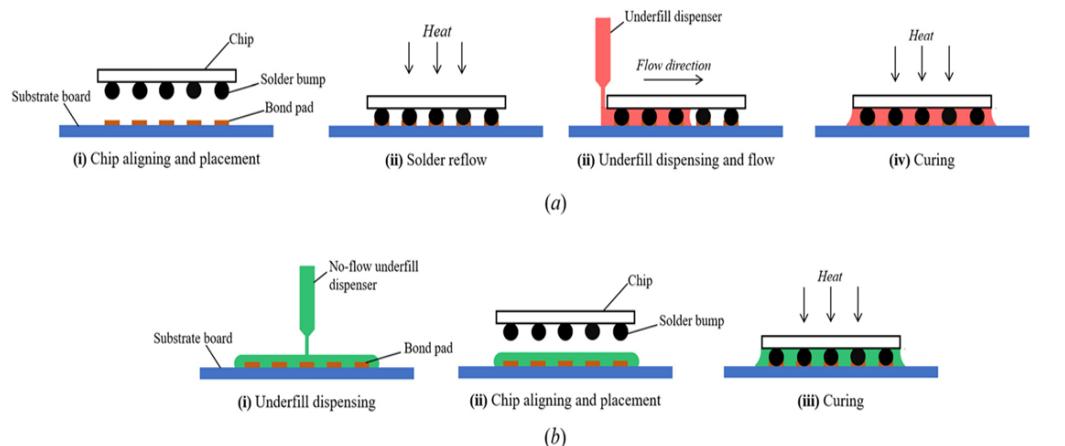


数据来源：普莱信智能，金元证券研究所

# 键合工艺：底部填充材料，从毛细流填充到模塑填充

- **传统毛细底填充(CUF)**：早期HBM封装中，在堆叠和互连完成后，从芯片边缘注入环氧基底填充材料，让其靠毛细作用渗入芯片间隙。这种方法成熟可靠，填充物硬化后支撑起微凸点，缓解热循环和机械应力，对提高冲击可靠性有效。但随着HBM堆叠层数增加和凸点间距减小（芯片间隙可低至 $20\mu m$ ），毛细流动变得困难，容易残留空隙导致应力集中。且逐片注胶过程耗时较长，影响产能。
- **非导电胶/NCP/NCF**：为避免注胶不均，工艺改为在键合前预先涂布底填充材料。例如非导电膏(NCP)和非导电膜(NCF)在芯片贴装前涂于芯片表面或贴附一层薄膜。热压键合时，这些材料在压力和温度作用下熔融充填芯片间隙并固化，相当于边键合边完成底填充。这样无需事后注胶，也免除了助焊剂残留清洗问题。**NCF**在HBM TSV堆叠中应用广泛，它对每个键合界面提供一致的填充，减少了层间空隙和应力。不过，预涂材料需要精确计量涂布厚度，以保证完全填充且不污染凸点表面。

图：传统毛细底部填充（CUF）与非导电胶/NCP/NCF 工艺流程



数据来源：《Underfill Flow in Flip-Chip Encapsulation Process: A Review》，金元证券研究所；

注：上图为 CUF，下图为 NCP/NCF



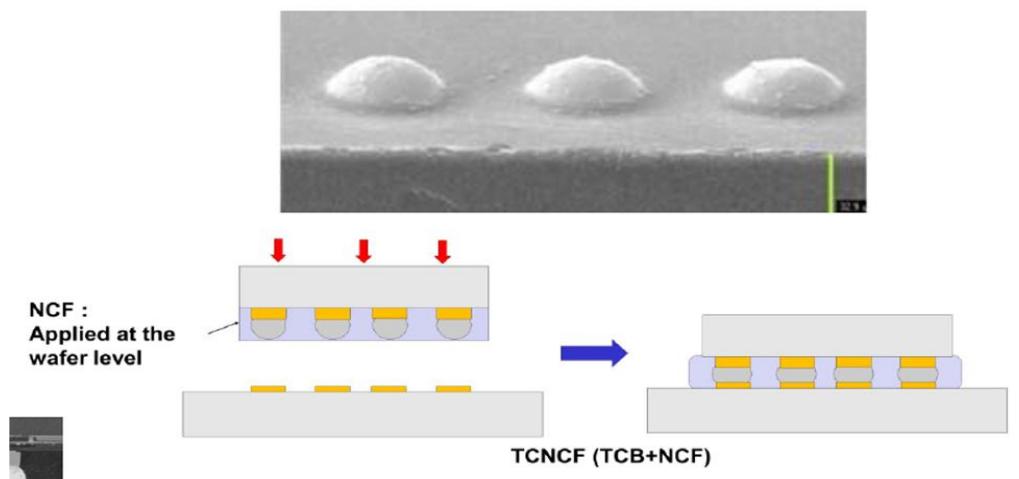
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：TCB+NCF



数据来源：《Reliability and Quality of off-chip Interconnects in Advanced Packages in Perspective of

High-Reliability Space Applications》，金元证券研究所  
首都机场集团  
Capital Airports Holdings

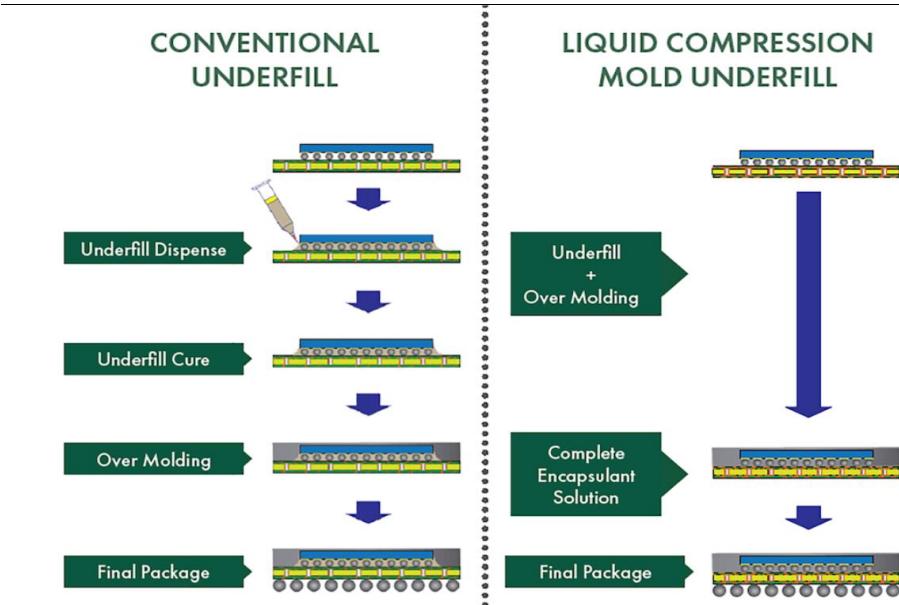


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 键合工艺：底部填充材料，从毛细流填充到模塑底填充

- 模塑底填充(MUF)：最新的发展是液态模塑底填充，一次工序即可完成多芯片的包封和填充。其做法是在多个芯片叠好后，将整个堆叠置于模具中，注入液态环氧模材，在压力下使其流遍芯片四周及间隙，然后固化成型。类似传统塑封工艺，但使用低粘度、高渗透性的树脂以保证在窄间隙中无死角充填。
- 液态MUF的最大优势是速度：同时实现了过模和底填充，相比逐层CUF可缩短约70%的工艺时间。随着HBM堆叠从8-Hi增加到12-Hi甚至16-Hi，一步填充所有层带来的时间收益更加显著。**SK海力士的MR-MUF工艺**即属此类，通过液态模材实现较低键合应力和更优散热。
- 挑战在于：液态模材在窄缝中均匀填充难度较大，可能导致晶圆翘曲。材料开发商如NAMICS等针对HBM开发了粒径极细的填料和低收缩树脂，以减小薄片堆叠的翘曲和应力。为应对更高层堆叠的热散需求，底填充材料还可能掺入高导热填料，兼顾机械支撑和导热功能。总体来看，底部填充技术正从毛细流动走向预涂/模塑的新范式，以提升制程效率和堆叠可靠性。

图：CUF vs Liquid MUF



数据来源：Namics, 3DinCites, 金元证券研究所



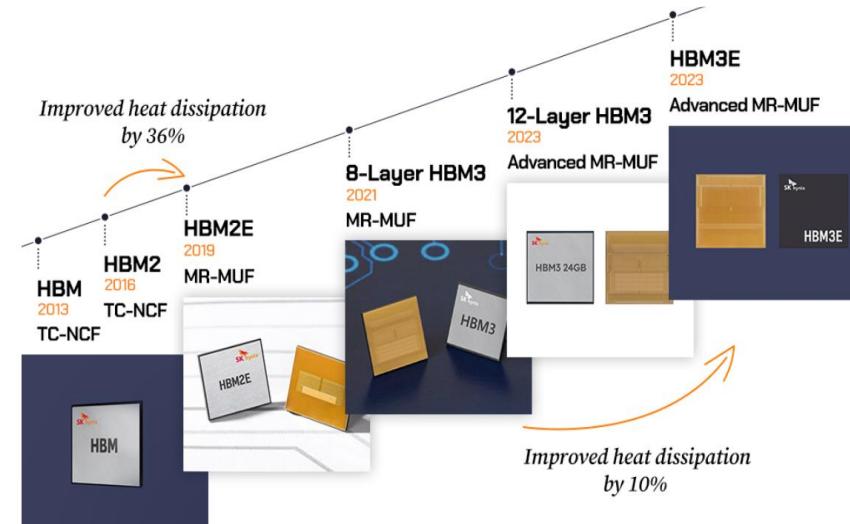
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：SK海力士在 HBM 2e 采用 MR-MUF 底部填充



数据来源：SK Hynix, 金元证券研究所

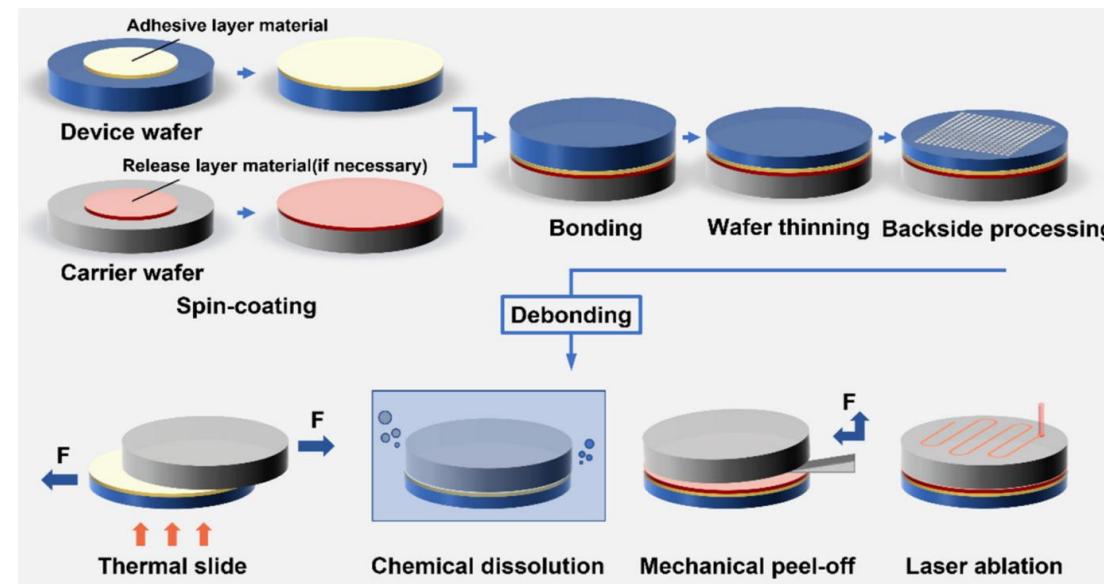


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 键合工艺：临时键合与解键合

- 在3D堆叠和TSV工艺中，经常需要对晶圆进行减薄加工和背面工艺。减薄的超薄晶圆(厚度几十微米)非常脆弱，无法直接在常规设备中处理，因此采用临时键合(Temporary Bonding)技术，该技术在TSV硅通孔工艺、CIS背照式图像传感器制造以及晶圆对晶圆键合后减薄中是必要步骤
- 临时键合：将待加工的薄晶圆临时粘接到一片载体晶圆上，以提供机械支撑和热稳定性。典型流程包括：先在刚性载体(通常是硅或玻璃)上涂敷一层临时粘接材料(如热塑性胶或光敏胶)，然后将待加工晶圆正面朝下贴合在载体上，通过加热或UV固化使两者牢固结合
- 解键合的方法取决于粘接材料类型：对于热塑胶，可加热软化后滑移分离；对于光敏胶，则使用特定波长UV照射使胶失去黏性，然后机械剥离。此外还有激光解键合方案，在透明载体侧用激光照射粘接层使其受热分解，实现分离

图：临时键合及解键合

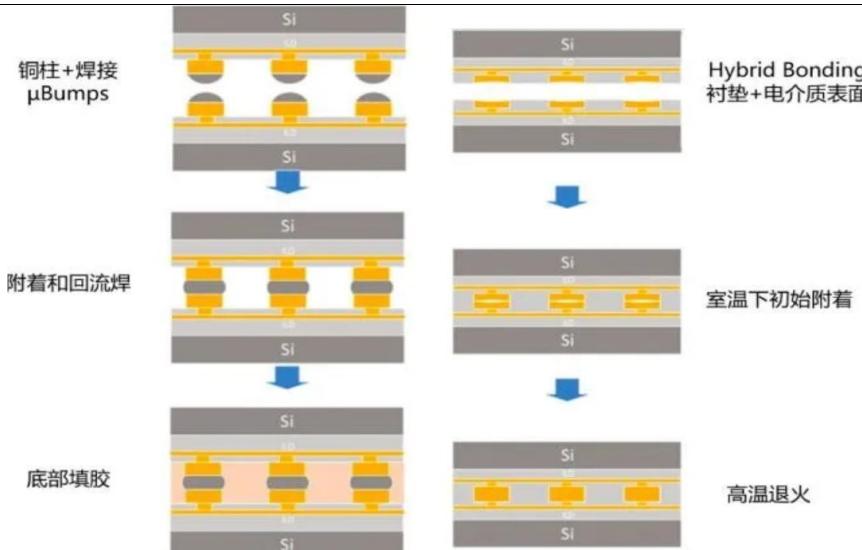


数据来源：MDPI，金元证券研究所

# 键合工艺：混合键合，突破I/O密度极限

- 混合键合：一种不使用中间焊料凸点，而直接通过芯片间表面材料实现连接的3D集成技术。具体而言，在混合键合中，两片待键合表面同时具备金属连接和介质键合：通常是在晶圆顶层形成铜金属垫，并被二氧化硅等介电层包围，然后对两片表面进行超平坦抛光和活化处理，使其氧化物表面可以在接触时形成键合。该工艺同时实现了介质对介质和金属对金属的结合，被称为“混合”键合或直接键合互连（DBI, Direct Bond Interconnect）
- 混合键合最大的特点是彻底取消了凸点焊料：两芯片间没有额外金属填料，只有嵌入在原始晶圆BEOL中的铜垫直接对接，这带来了多重技术优势：首先，键合间距显著缩小。由于不需要凸点焊盘和焊料体积，占位大为减小，可以实现 $10\mu\text{m}$ 及以下pitch的超密互连；第二，电气性能提升：直接的铜-铜连接避免了焊料的高电阻和高延迟，降低了互连阻抗和电容，提高信号完整性和传输速度；第三，功耗和带宽密度优势：混合键合能提供远超凸点的I/O密度和更短互连长度，意味着在相同功耗下可支持更高带宽，或在相同带宽下降低功耗；第四，堆叠高度降低、散热改善：没有焊料凸点后，芯片间距可以做得非常小，使多层堆叠的总高度降低，从而更容易散热

图：回流焊+凸点键合 vs 无凸点混合键合



数据来源：半导体行业观察，金元证券研究所



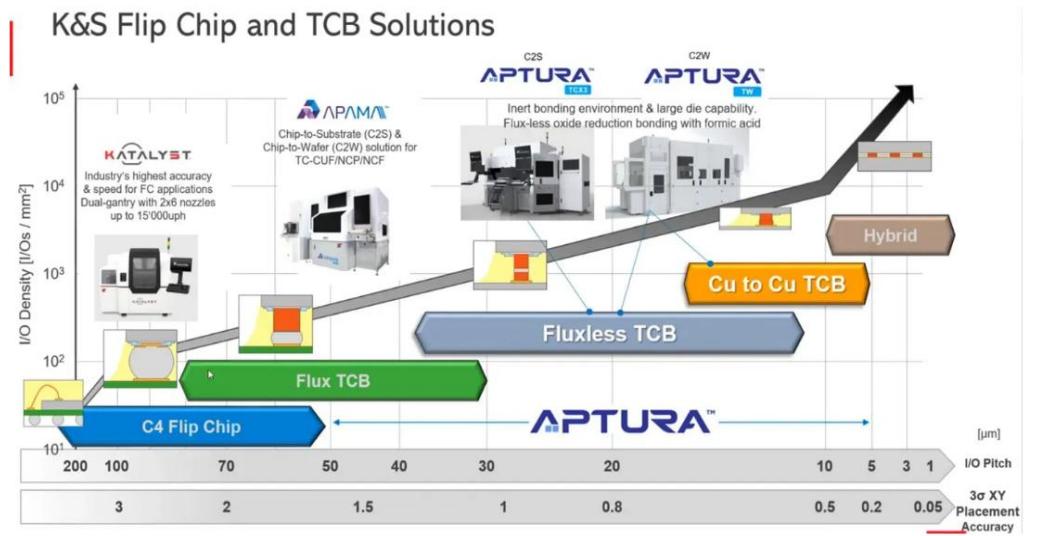
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：10 pitch 步入 10  $\mu\text{m}$  后，将只能采用混合键合技术



数据来源：K&S，金元证券研究所

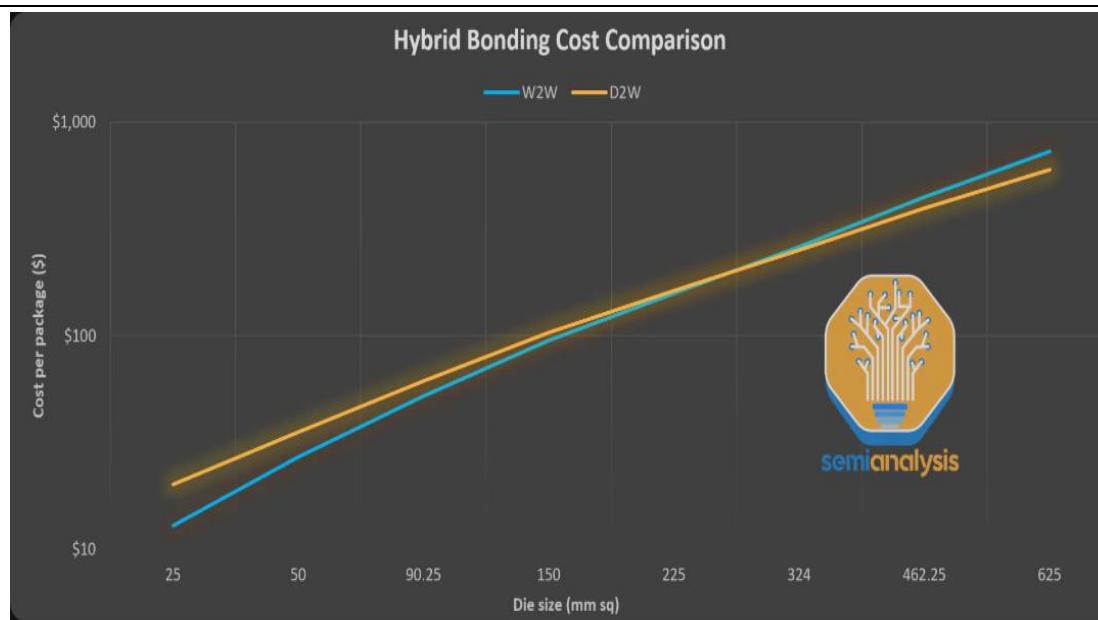


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 键合工艺：D2W、W2W

- 混合键合可按照键合方式分为晶圆对晶圆（Wafer-to-Wafer, W2W）和晶粒对晶圆（Die-to-Wafer, D2W）两种工艺形式。两者关键差异在于：**W2W**是将两片完整晶圆对准后整体键合，而**D2W**则是将经过切割测试的已知良品裸芯片一个一个嵌入并键合到目标晶圆上
- D2W**的优势在于可以“挑好舍坏”：仅转移已知良品（KGD），避免了不良芯片连累，从而支持较大芯片在良率不高时依然实现堆叠。**D2W**能够异构集成不同工艺/不同尺寸的芯片，例如逻辑+存储等，灵活性强。随着芯片面积增大，**W2W**工艺因“牺牲良品”的成本急剧攀升，而**D2W**成本曲线相对平缓。这也是为何早期混合键合产品多采用**D2W**方式：例如AMD 3D V-Cache选择**D2W**，就是考虑到CPU逻辑裸片面积大且良率非完美，需要通过挑选缓存芯片来保证成品率

图：D2W 更适合面积较大芯片 3D 键合，曲线更平滑（以下非真实成本数据）



数据来源：Semianalysis, 金元证券研究所



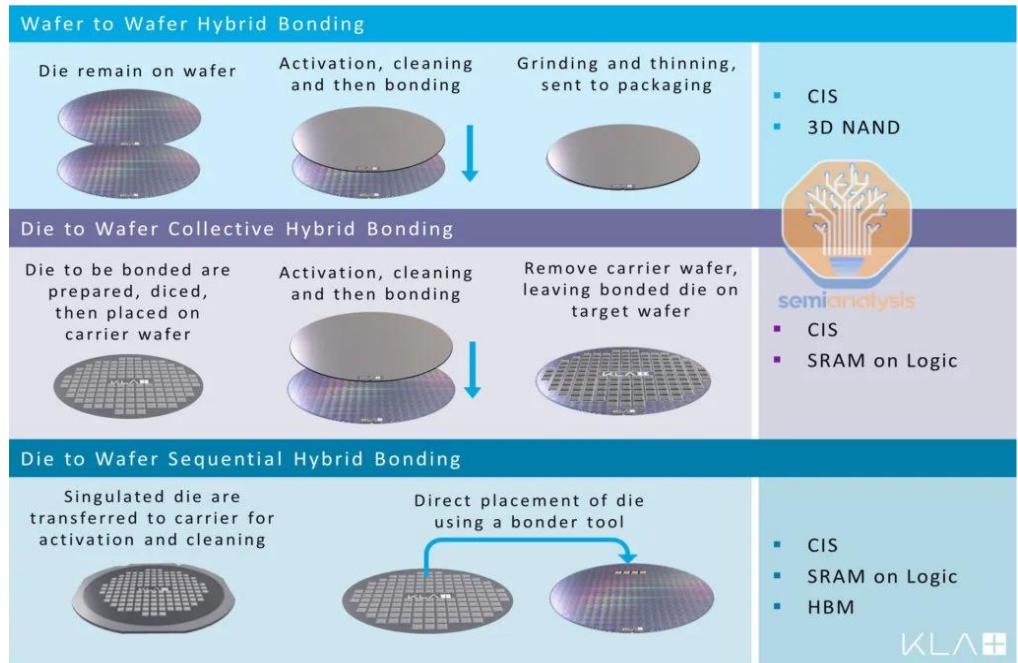
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：W2W vs D2W 及其适用范围



数据来源：Semianalysis, KLA, 金元证券研究所

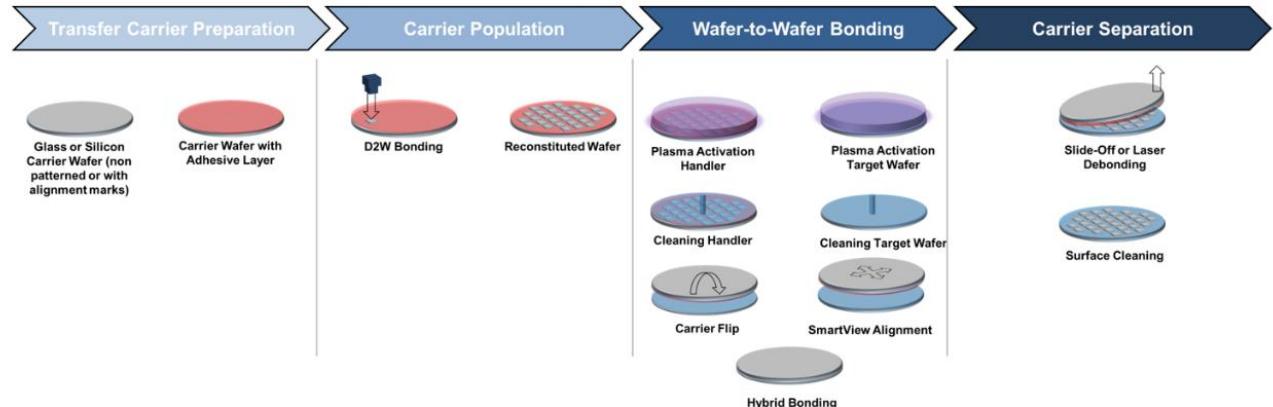


金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 键合工艺：D2W、W2W

- 无论W2W或D2W，其核心设备包括：CMP平坦化设备（确保键合表面原子级平坦度）；等离子活化清洗机（对晶圆表面进行微观粗化和键合作用团激活，提高键合强度）；高精度键合对准机。D2W额外需要高精度拾放机和可能的临时键合载板。
- D2W可以细分为两类：
  - 可采用“集体转移（Collective D2W）”技术，即先将多颗裸片临时键合在载体晶圆上排成阵列，再整体与目标晶圆对准键合，然后移除载体
  - 采用顺序放置“单颗转移（Direct Placement D2W）”，按顺序一颗颗放置到另一个晶圆上对应位置，位置精度提到。

图：Co-D2W vs DP-D2W

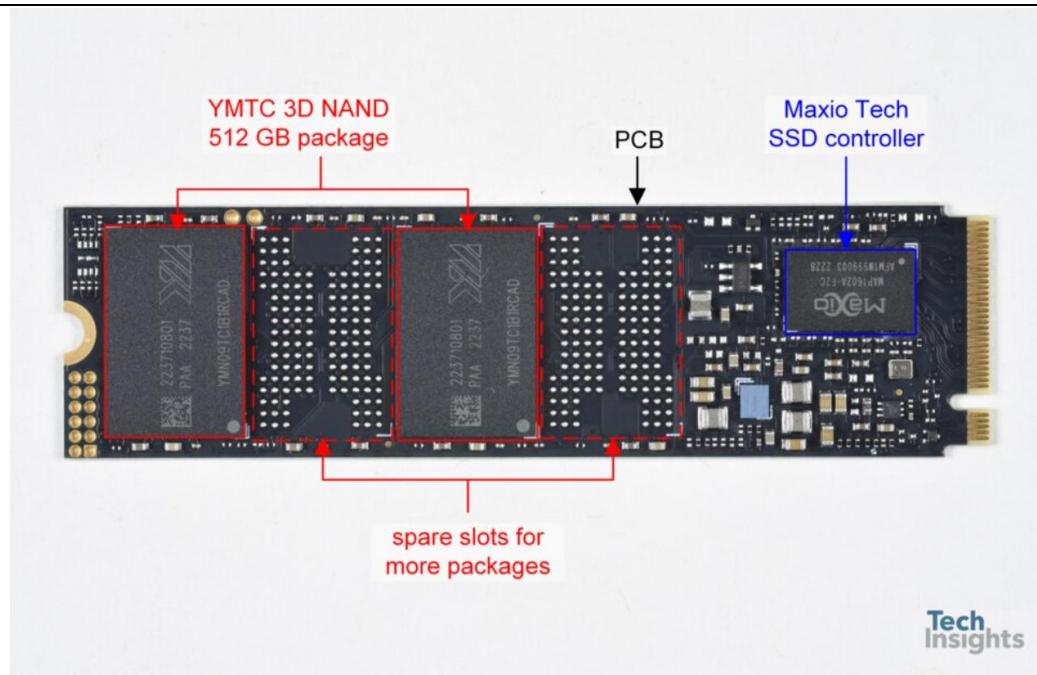


数据来源：EVG, 金元证券研究所

# 混合键合现阶段应用：3D-NAND

- 长江存储独创的“Xtacking”架构在2018年亮相，即通过晶圆键合将NAND存储单元阵列与外围CMOS电路集成。长江存储的第一代Xtacking (Xtacking 1.0) 用于32层和64层3D NAND试产，2019年量产的64层TLC闪存将外围电路和存储阵列分别在两片晶圆上制造，然后W2W键合成一体。2020年底，长江存储发布了128层3D TLC NAND (X2-9060)，采用Xtacking 2.0架构，在国内首次实现了128层闪存量产。进入2022年，长江存储发布第三代Xtacking (Xtacking 3.0)，推出232层堆叠的X3-9070 TLC闪存。这款232层产品堆叠了两个116层阵列板，并引入背侧源极连接(BSSC)等新技术，将I/O速度提升50%、存储密度提高70%

图：128 层 3D NAND 在 TiPlus7100 1TB SSD 应用



数据来源：Tech Insights, 金元证券研究所



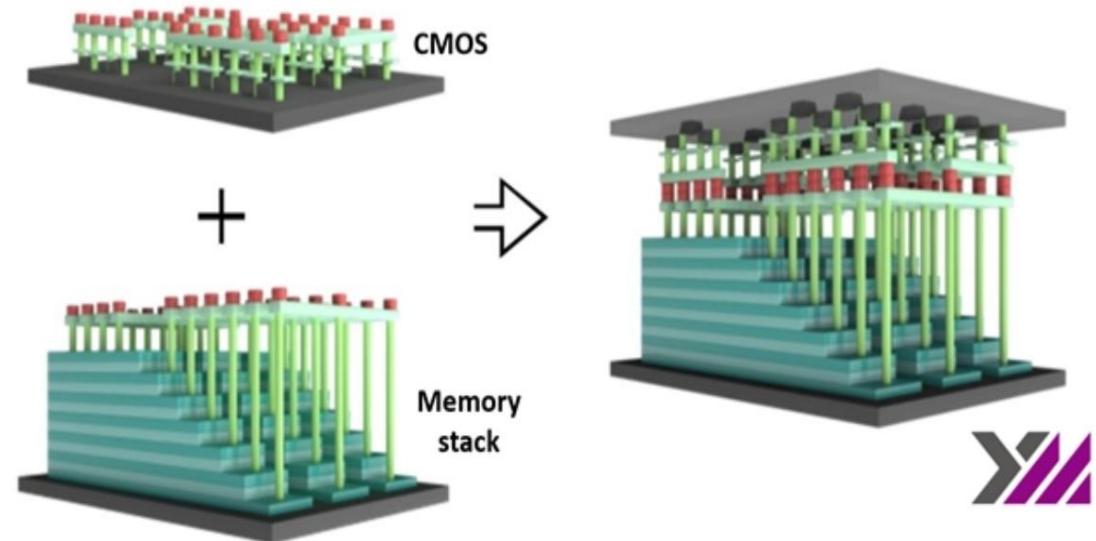
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：通过 Xtacking 技术，将外部电路与 Memory Stack 直接键合



数据来源：YMTC, 金元证券研究所



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

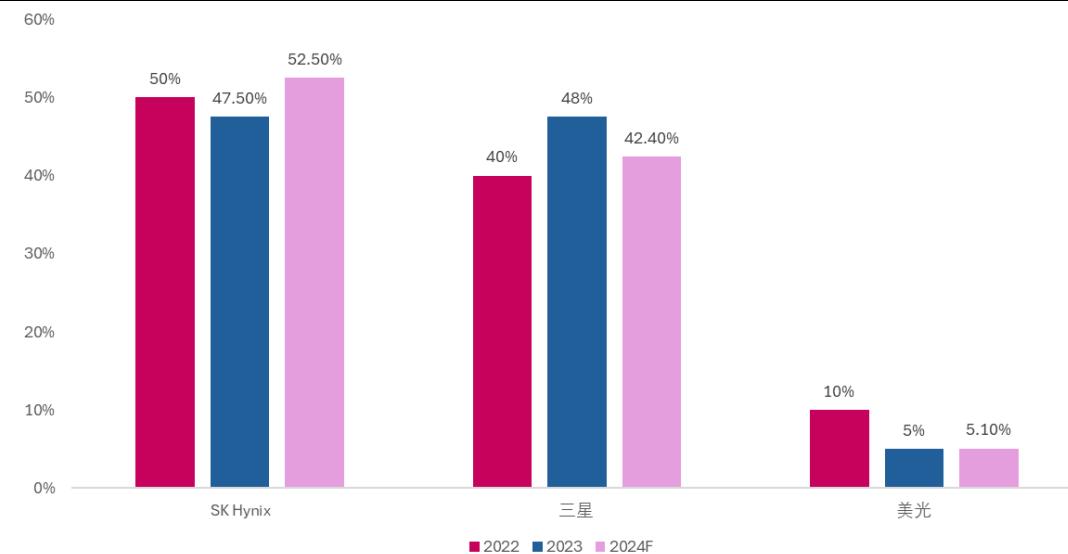
# HBM后道测试挑战

- HBM堆叠包含一个逻辑基底芯片 (Base Die) 和多个DRAM存储芯片。当它们垂直集成后，对各芯片特别是基底逻辑的测试变得复杂。主要难点在于已堆叠状态下基底芯片的可及性。基底die在最底部，上方被DRAM覆盖，其许多功能单元只有通过堆叠内部互连 (I/O 或 TSV) 才能访问。如果等到堆叠完成再测试基底，一旦发现缺陷，将造成整叠报废，损失巨大。因此通常要求在堆叠前对基底芯片进行充分测试，确保其为已知良品 (Known Good Die, KGD)
- 然而基底芯片作为专用HBM接口逻辑，可能在裸片状态下无法完全按照最终工作条件测试——例如它的一些高速 I/O 接口只有连接实际DRAM才能运行全速。这就需要内建自测 (BIST) 和冗余机制：基底die上设计用于裸片测试的特殊模式，或为DRAM接口提供环回 (loopback) 路径，以便在无实际DRAM时进行功能验证。另外，探针在晶圆上直接接触基底die上成千上万微小凸点测试也是难题，通常基底die在晶圆级会保留测试垫 (test pad) 或牺牲焊垫，方便探针卡接触
- HBM接口的另一个挑战是高速 I/O 信号测试。HBM链路工作频率可达每线 6.4Gbps 甚至更高，传统存储测试在 MHz 量级，无法满足此要求。要在ATE上产生和采集上千路高速信号，需要探针卡和测试通道的信号完整性 (SI) 保证。探针卡作为连接ATE和晶圆的桥梁，其设计尤为关键。垂直接触探针卡 (vertical probe card) 常用于高密度场合，即成千上万细针垂直排列接触晶圆上垫点。对于 HBM，通常在基底die周围或顶部设计了一圈测试垫阵列，供探针卡同时接触成百上千信号

# HBM现阶段市场及规模：SK海力士、三星、美光“三分天下”

- 根据TrendForce数据，从HBM供应市场角度，SK海力士、三星、美光独占HBM市场，且SK海力士、三星的集中度较高
- IDTechEx预测，高性能计算（HPC）硬件市场2025-2035期间将保持13.6%的年化增长率，市场规模将超5810亿美元。其中，95%的HPC将使用HBM技术以满足性能需求，HBM单位销售额预计2035年将比2024年增长15倍
- 地缘政治紧张局势和中美贸易紧张局势对内存格局产生了显著影响。国内厂商加速HBM技术突破

图：SK 海力士、三星、美光“三分天下”



数据来源：TrendForce，金元证券研究所



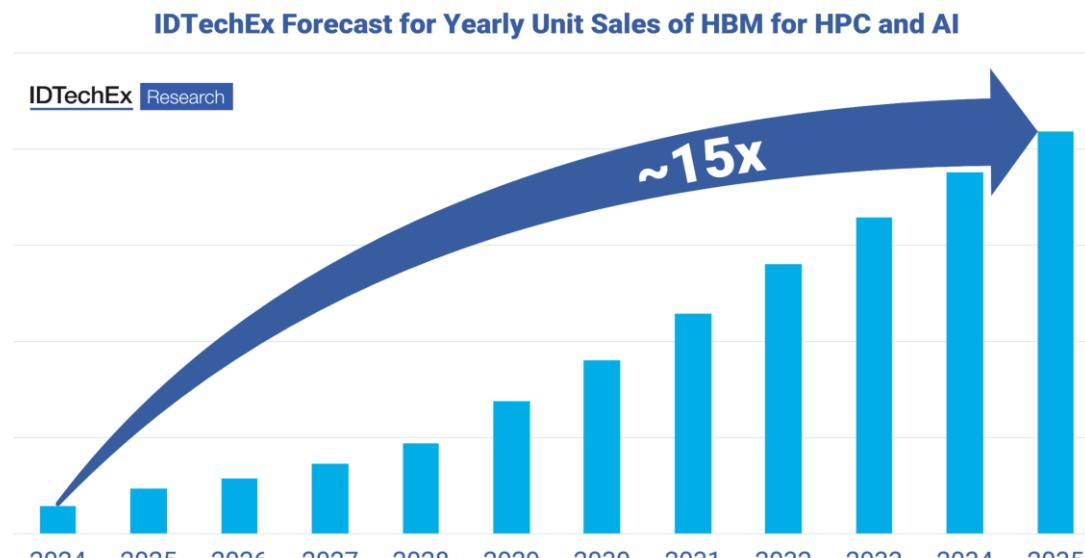
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：IDTECHEx 预测，2035 年 HBM 单位销售额较 2024 年增长 15 倍



数据来源：IDTECHEx Research，金元证券研究所



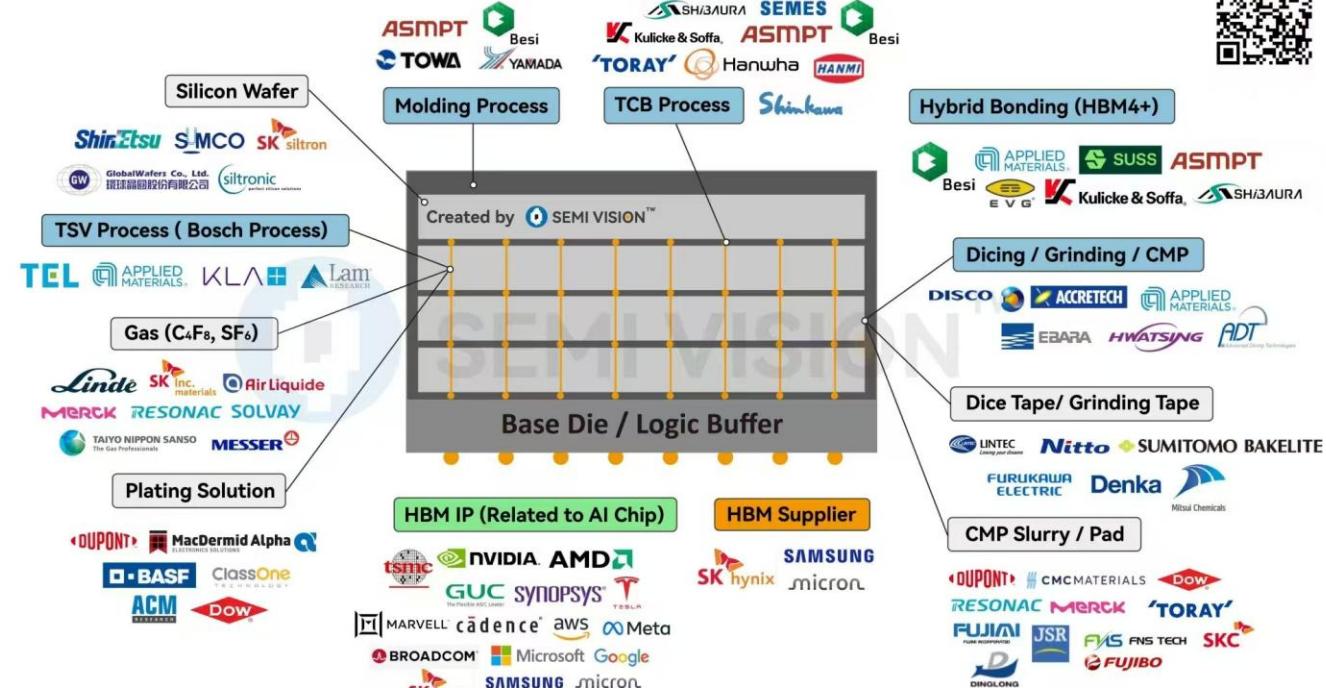
金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# HBM生产设备、材料市场：国产替代需求仍高

- HBM生产闭环，相关生产工艺、材料（包括TSV工艺、镀铜、TCB热压键合、晶圆切割、CMP、CMP抛光液等）主要以欧美、日本企业为主。
- 中国企业在CMP工艺中的华海清科、CMP抛光液的鼎龙股份，但在其他核心工艺中，如TSV刻蚀环节、TCB热压键合及相关材料上仍有一定差距，国产化率仍然较低。

图：HBM 相关工艺设备以欧美、日本企业为主导

## High Bandwidth Memory (HBM) Supply Chain Overview



数据来源：Semi Vision, 金元证券研究所

# HBM工艺趋势：高深宽比TSV、高良率键合

- HBM制造及封装成本角度，TSV通道形成+TSV露铜工艺的占比极高：
  - 高深宽比刻蚀的难度加大，以DRIE+BOSCH工艺将逐步取代RIE工艺。
  - 在绝缘层沉积方面，当前主流方法仍是采用PECVD+TEOS+硅烷气体沉积二氧化硅。随着TSV直径小于5 μm或深宽比超过15: 1时，利用原子层沉积（ALD）更合适，但是ALD工艺设备较为昂贵，且沉积速率较慢。
  - 扩散阻挡层沉积主要采用物理气相沉积（PVD），但时为了提高TSV内部金属覆盖率，常利用直流磁控溅射方式加快沉积速率，降低基材温度。通孔镀铜方面，自下而上镀铜（Bottom Copper Plating）能够尽可能避免空洞（Void）。
  - TSV背面露铜则需要CMP工艺去除芯片表面的铜覆盖层和扩散阻挡层
- 键合工艺在成本敏感性测试中较为重要：
  - 采用传统回流焊+毛细底填充工艺在随着HBM堆叠层数增加和凸点间距减小（芯片间隙可低至20μm），毛细流动变得困难，容易残留空隙导致应力集中。
  - 利用热压键合+NCF的方式，预先涂布底填充材料。由于助焊剂底部清洗难度较大，业界开发出fluxless（无助焊剂）工艺。但热压键合工艺效率较低，无法批量操作。
  - 液态模塑底填充使用低粘度、高渗透性的树脂以保证在窄间隙中无死角充填，并且可批量，效率较高或成为主流。
  - 混合键合仍处于初步阶段，但高I/O优势显著。D2W方式可以在裸芯良率较低背景下提高封装良率，但效率较低；未来可能出现DRAM层间采用MR-MUF键合工艺，而通过D2W或W2W与基底逻辑Die键合的组合键合方式。



www.jyzq.cn



全国统一客服电话：95372

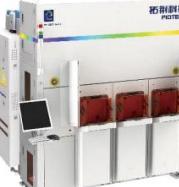
此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 拓荆科技 (688072.SH) : 薄膜沉积设备+混合键合

- 公司具备生产高精度薄膜沉积能力，或成为TSV绝缘层+扩散阻挡层沉积及FOEL核心设备供应商。2024年公司出货超过1000个设备反应腔，创历史年度新高。PECVD、ALD、SACVD、HDPCVD、Flowable CVD及混合键合设备等系列产品量产规模不断扩大，持续获得客户订单。
- 公司不断拓展新产品和新工艺，包括超高纵横比沟槽填充CVD设备（PF-300T Flora，累计出货量超过15腔）、PECVD Bianca工艺设备（晶圆背面薄膜沉积设备，接到超过25个反应腔的订单，部分出货）和键合对准精度测量产品。

类型	主要产品型号	产品图片	应用及特点
PECVD	PF-300T Bianca		主要应用于集成电路逻辑芯片、存储芯片制造领域，可以在晶圆背面沉积 SiN、SiO <sub>2</sub> 等介质薄膜材料，实现对晶圆翘曲的纠正以及晶圆背面的保护。
	PF-300T Flora		主要应用于集成电路逻辑芯片、存储芯片制造领域，已实现产业化应用，可以沉积 SiO <sub>2</sub> 等介质薄膜材料。超高深宽比沟槽填充CVD设备可以在晶圆表面沉积高品质的介电薄膜材料，经过固化及氧化等处理工艺后，可达到完全填充间隙而不会留下孔洞和缝隙的效果。公司自主研发并推出了超高深宽比沟槽填充 CVD 产品
ALD	PF-300T Astra		在集成电路逻辑芯片、存储制造及先进封装领域已实现产业化应用，可以沉积高温、低温、高质量的 SiO <sub>2</sub> 、SiN 等介质薄膜材料。PE-ALD 是利用等离子体增强反应活性，提高反应速率，具有相对较快的薄膜沉积速度、较低的沉积温度等特点，适用于沉积硅基介质薄膜材料

数据来源：拓荆科技半年报，金元证券研究所

类型	主要产品型号	产品图片	应用及特点
D2W混合键合	Dione 300		主要应用于晶圆级三维集成、存储芯片制造领域，已实现产业化应用，可实现 12 寸晶圆对晶圆的混合键合和熔融键合。
	Propus		主要应用于芯片对晶圆三维集成领域，已实现产业化应用，可实现混合键合前晶圆及切割后芯片的表面活化与清洗。
	Crux 300		主要应用于晶圆级三维集成领域，可实现晶圆对晶圆混合键合和芯片对晶圆混合键合后的键合精度量测。

数据来源：拓荆科技半年报，金元证券研究所



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 盛美上海 (688082.SH): 电镀设备达到国际水平

- 公司自主开发针对28nm及以下技术节点的IC前道铜互连镀铜技术Ultra ECP map及三维电镀设备Ultra ECD 3d。公司的多阳极局部电镀技术采用新型的电流控制方法，实现不同阳极之间毫秒级别的快速切换，可在超薄籽晶层上完成无空穴填充，同时通过对不同阳极的电流调整，在无空穴填充后实现更好的沉积铜膜厚的均匀性，可满足各种工艺的镀铜需求。在三维堆叠电镀设备方面，公司应用于填充3d 硅通孔 TSV 和 2.5D 转接板的三维电镀设备Ultra ECP 3d。基于盛美半导体电镀设备的平台，该设备可为高深宽比（深宽比大于 10:1）铜应用提供高性能、无孔洞的镀铜功能。

图：公司专用于3D堆叠电镀设备



数据来源：盛美上海，金元证券研究所



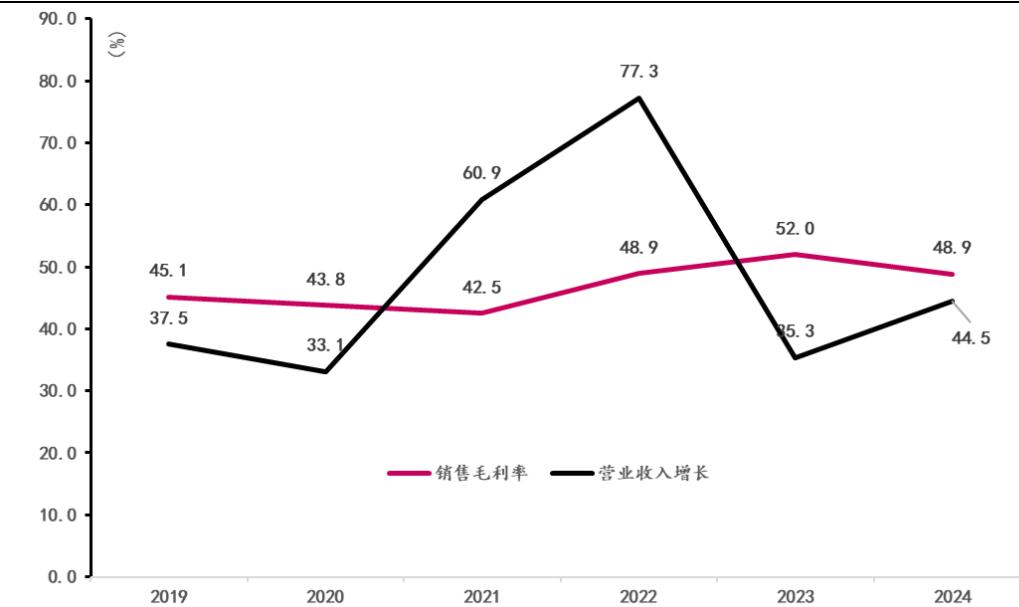
www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。

图：公司营业收入增长及毛利率(%)



数据来源：Wind，金元证券研究所



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 华海清科 (688120.SH): 磨削、CMP及清洗一体化

- 公司产品包括Universal系列CMP设备、Versatile系列减薄设备、HSC系列清洗设备、HSDS/HCDs系列等技术服务，主要产品涵盖CMP设备、减薄设备、划切设备、湿法设备、晶圆再生、关键耗材与维保服务等。公司覆盖多道HBM必备工艺设备，包括减薄、键合，并提供一体化工艺设备

产品类别	产品名称/服务	主要应用领域
CMP设备	Universal系列	适用于逻辑芯片、存储芯片(DRAM、3D NAND)、先进封装、大硅片、第三代半导体、MEMS、Micro LED等制造工艺。
减薄设备	Versatile系列	满足3D IC对超精密磨削、CMP及清洗的一体化工艺需求。
划切设备	Versatile-DT300	满足集成电路、先进封装等制造工艺的12英寸晶圆边缘切割需求。
湿法设备	HSC-S3810、HSC-F3400等	覆盖大硅片清洗和FEOL/BEOL晶圆正背面及边缘清洗工艺。
晶圆再生	再生服务	已实现双线运行，获得多家大生产线批量订单并长期稳定供货。
关键耗材与维保服务	7区抛光头维保服务等	随着CMP设备保有量的增加，耗材零部件、维保服务需求同步提升。



图：Versatile-GP300，应用于3D IC制造、先进封装等领域超精密减薄工艺

数据来源：华海清科公司公告、金元证券研究所

数据来源：华海清科，金元证券研究所



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.

# 风险提示

- HBM工艺复杂，涉及产业链设备及产业链仍以海外企业为主，地缘影响较大；  
当前国内HBM2e仍处于测试阶段，落地或不及预期
- AI产业链发展不及预期，对高性能存储需求下降。
- 市场竞争激烈程度超出预期

# 投资评级说明

## 金元证券行业投资评级标准：

增持：行业股票指数在未来6个月内超越大盘；

中性：行业股票指数在未来6个月内基本与大盘持平；

减持：行业股票指数在未来6个月内明显弱于大盘。

## 金元证券股票投资评级标准：

买入：股票价格在未来6个月内超越大盘15%以上；

增持：股票价格在未来6个月内相对大盘变动幅度为5%~15%；

中性：股票价格在未来6个月内相对大盘变动幅度为-5%~+5%；

减持：股票价格在未来6个月内相对大盘变动幅度为-5%~-15%；。

# 免责声明

本报告由金元证券股份有限公司（已具备中国证监会批复的证券投资咨询业务资格）制作。本报告所载资料的来源及观点的出处皆被金元证券认为可靠，但金元证券不保证其准确性或完整性。该等信息、意见并未考虑到获取本报告人员的具体投资目的、财务状况以及特定需求，在任何时候均不构成对任何人的个人推荐。投资者应当对本报告中的信息和意见进行独立评估，并应同时考量各自的投资目的、财务状况和特定需求，必要时就法律、商业、财务、税收等方面咨询专业财务顾问的意见。对依据或者使用本报告所造成的一切后果，金元证券及/或其关联人员均不承担任何法律责任。投资者需自主作出投资决策并自行承担投资风险，任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

本报告所载的信息、材料或分析工具仅提供给阁下作参考用，不是也不应被视为出售、购买或认购证券或其他金融工具的要约或要约邀请。该等信息、材料及预测无需通知即可随时更改。过往的表现亦不应作为日后表现的预示和担保。在不同时期，金元证券可能会发出与本报告所载意见、评估及预测不一致的研究报告。

金元证券的销售人员、交易人员以及其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。金元证券没有将此意见及建议向报告所有接收者进行更新的义务。金元证券的自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。

在法律许可的情况下，金元证券可能会持有本报告中提及公司所发行的证券头寸并进行交易，也可能为这些公司提供或争取提供投资银行业务服务。因此，投资者应当考虑到金元证券及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。

本报告的版权仅为金元证券所有，未经书面许可任何机构和个人不得以任何形式转发、翻版、复制、刊登、发表或引用。



www.jyzq.cn



全国统一客服电话：95372

此文件版权归金元证券股份有限公司所有，未经许可任何单位或个人不得复制、翻印。



金元证券股份有限公司  
GOLDSTATE SECURITIES CO.,LTD.