

半导体

超配 中性 低配

股票名称 代码 评级 目标价

美光 MU US 买入 US\$148.00

HBM 市场更新：HBM 高端产品供不应求时间或长于预期

- 2024-27 年期间，我们认为 HBM 供不应求局面将持续存在。
- 美光科技或持续提升其在 HBM 市场的市占率。
- 16 层 HBM4 时代，HBM 市场份额争夺或愈发激烈。

HBM 供需分析。与市场观点不同，我们认为 2024-27 年 HBM（高带宽存储器）产品将供不应求。2024-25 年，我们预计 AI 直接拉动的 HBM 需求量或达 14.0 亿/23.6 亿吉字节（GB）。供给侧，2024/25 年底 SK 海力士、美光和三星电子三家合计的 HBM 供应量达 10.1 亿/20.1 亿 GB。展望 2026-27 年，我们预计三家公司 HBM 供给或达 34.1 亿/49.1 亿 GB；而需求量受益于 AI 服务器存储容量提升有望增长到 39.7 亿/51.6 亿 GB。由此可见在未来几年，HBM 市场仍供不应求。

与市场观点不同，我们认为在 16 层 HBM4 时代，SK 海力士的市占率或出现一定程度侵蚀。HBM3E 时代，SK 海力士享有工艺和产能优势，采用先进 MR-MUF 工艺有助于产品出货和产能爬坡。我们预计 2024-25 年 SK 海力士仍是 AI 客户的主要 HBM 供应商之一，但进入 16 层 HBM4 及 HBM5 时代，我们不排除美光和三星电子加快争夺高端 HBM 市场份额的可能性。工艺层面，若 SK 海力士面向 16 层 HBM4E 及更高端产品转向混合键合方案，则其封装技术将与美光和三星电子处于同一起跑线。此外，美国本土客户基于供应链安全和地缘政治因素可能会增加对于美光的采购份额。同时，考虑到美光和三星电子 HBM 新产能有望于 2025-28 年逐步释放，我们不排除 SK 海力士市占率遭受两家公司的挤压。对于美光，基于其 HBM 业务低基数和 2024-26 财年的资本支出计划，我们看好它在 HBM 行业中市占率提升的逻辑。

我们重申对 SK 海力士和美光科技的“买入”评级。两者相比之下，我们更看好美光，主要是基于：1) 中长期而言，美光或在 HBM4 和 HBM4E 时代逐渐追平工艺差异，同时北美客户基于供应安全考虑或逐步加大非韩系厂商供应份额。美光产能释放下，2025-27 财年收入和净利或超出市场预期；2) 基于历史平均估值，美光相较 SK 海力士存在一定估值溢价，其 24 个月的预期市盈率高于 SK 海力士。

风险提示：通用性 AI 需求不及预期，产能利用率和良率不及预期，NAND 和 DRAM 供过于求时间点早于预期。

研究团队

王国哈, 分析师

证书编号: S1680524080001

电话: +86 21 6015 6827

电邮地址: ghwang@huaxingsec.com

财务数据摘要

股票名称	代码	评级	货币	股价	目标价	潜在	市盈率(x)	ROE (%)	EPS CAGR(%)
美光	MU US	买入	US\$	102.76	148.00	+44%	12.3	8.4	17.1
SK 海力士	000660 KS	买入	KRW	176,700.00	259,548.00	+47%	6.7	4.5	26.1

资料来源：，华兴证券预测

并购家 免费行业报告网

IPOIPO.CN 每日更新 不用注册会员 无广告 永久免费

目录

焦点图表.....	3
投资概述.....	5
需求侧：AI 大模型爆发拉动 HBM 需求.....	7
HBM 在 GPU 中得到广泛应用	9
HBM 升级换代伴随带宽和速率提升; 同时伴随 HBM 性能和 HBM 颗数增加	10
SK 海力士、三星电子和美光布局 HBM4，单颗 HBM 容量提升	11
英伟达下一代 GPU 的 HBM 颗数增加	12
英伟达、AMD 及云服务大厂的 AI 芯片出货拉动 HBM 需求	13
供给侧：良率和工艺问题拖累美光和三星电子出货进度	15
HBM3 及 HBM3E，海力士客户验证及出货领先于美光和三星电子	15
良率问题影响美光和三星电子 HBM 出货速度	16
TSV 是制备 HBM 的第一个关键环节	17
凸点在制备 HBM 中必不可少	19
芯片键合是区分 HBM3E 制备良率的关键环节	21
设备端，TCB 设备供应为关键一环	22
TC-NCF 和先进 MR-MUF 均有需求	22
除逻辑芯片先进封装外，HBM 键合是 TCB 设备的又一大使用场景	22
TCB 设备需求有望转向无焊剂 (Fluxless) TCB 方案	23
HBM4 和 HBM5 时代，混合键合技术有望加速推广	23
HBM3E 方面，SK 海力士能实现更高热稳定性和良率	27
三星电子和美光希望借 HBM4 弯道超车	27
SK 海力士 HBM4 优势来自成熟工艺和台积电支持	28
我们认为当前标准对 SK 海力士维护近期市场地位更有利	29
2025-26 年或聚焦 HBM4 市场份额争夺	29

焦点图表

图表 1：2023-27 年 HBM 需求量估算（百万 GB）

需求量估算	2023	2024E	2025E	2026E	2027E
台积电 CoWoS 年末月产能（千片）	15	36	68	92	110
YoY		137%	92%	35%	20%
台积电 CoWoS 年实际产能（千片）	109.592	311	604	972	1221
YoY		184%	94%	61%	26%
产能利用率（%）	100%	100%	97%	97%	96%
台积电 CoWoS 实际年产量（千片）	110	311	586	943	1,172
台积电外溢 CoWoS 实际年产量（千片）	25	50	85	109	131
合计 CoWoS 实际年产量（千片）	135	361	671	1,052	1,303
英伟达占比（%）	46%	51%	51%	48%	49%
英伟达获得的 CoWoS 产能（千片）	62.0	184.0	343.0	504.8	638.3
AMD 及其他 AI 芯片设计厂占比（%）	54%	49%	49%	52%	51%
AMD 及其他 AI 芯片设计厂商获得的 CoWoS 产能（千片）	72.6	177.0	327.8	546.9	664.4
单片 CoWoS 可封装 GPU 颗数	28.0	25.0	18.6	17.0	16.5
YoY		-11%	-26%	-9%	-3%
对应 GPU+ASIC 出货量（百万颗）	3.8	9.0	12.5	17.9	21.5
单 GPU 中平均 HBM 需求个数	6.0	7.0	7.5	7.8	8.0
HBM 堆栈平均层数	8.0	8.5	9.0	9.5	10.0
HBM 单颗携带存储容量（GB）	2.5	2.6	2.8	3.0	3.0
单 GPU 中需求的存储容量（GB）	120.0	154.7	189.0	222.3	240.0
GPU 需求的存储总容量（百万 GB）	452.2	1,396.2	2,357.0	3,974.3	5,158.8

资料来源：TrendForce，华兴证券预测

图表 2：2023-27 年三家原厂 HBM 实际产能估算（晶圆，片）

HBM 实际月产能（当年 12 月）	2023	2024E	2025E	2026E	2027E
三星电子	26,700	44,700	72,200	108,200	150,200
SK 海力士	29,800	68,600	108,200	152,000	197,600
美光	2,800	9,850	25,750	52,450	86,750
合计	59,300	123,150	206,150	312,650	434,550
HBM 实际年产能（）	2023	2024E	2025E	2026E	2027E
三星电子	156,800	437,400	703,700	1,100,400	1,571,400
SK 海力士	188,900	611,000	1,080,600	1,580,400	2,120,400
美光	19,800	66,500	214,900	468,300	849,400
合计	365,500	1,114,900	1,999,200	3,149,100	4,541,200
YoY	123%	205%	79%	58%	44%
三家原厂实际产能占比	2023	2024E	2025E	2026E	2027E
三星电子	43%	39%	35%	35%	35%
SK 海力士	52%	55%	54%	50%	47%
美光	5%	6%	11%	15%	19%
合计	100%	100%	100%	100%	100%

资料来源：TrendForce，华兴证券预测

图表 3：2023-27 年三家原厂 HBM 实际供给量估算

	2023	2024E	2025E	2026E	2027E
三家原厂合计月产能（晶圆，片）	59,300	123,150	206,150	312,650	434,550
三家原厂实际年产能（晶圆，片）	365,500	1,114,900	1,999,200	3,149,100	4,541,200
平均稼动率（%）	87%	87%	88%	89%	89%
三家原厂合计产量（晶圆，片）	319,658	967,990	1,768,530	2,793,216	4,027,684
12 寸晶圆上切割数量（片）	550.0	550.0	550.0	550.0	550.0
出货量/产量比率	0.98	0.98	0.98	0.98	0.98
合格裸片出货量（百万颗）	119.2	387.8	719.1	1,136.2	1,637.7
HBM 堆栈层数	8.0	8.5	9.0	9.5	10.0
HBM 供给量-包括 8 层和 12 层堆栈（百万）	14.9	45.6	79.9	119.6	163.8
HBM 单颗存储容量（GB）	2.5	2.6	2.8	3.0	3.0
三家原厂合计 HBM 供给量（百万 GB）	298	1,008	2,013	3,409	4,913

资料来源：TrendForce，华兴证券预测

图表 4：SK 海力士、三星电子及美光 HBM 各个代际战略

	美光	三星电子	海力士
HBM 3,3E 2024	非导电膜 热压键合 (NCF TC)	非导电膜 热压键合 (NCF TC)	大规模回流 模具底填充 (MR-MUF)
HBM 4,4E 2026	无助焊剂热压 / 混合键合	无助焊剂热压 / 混合键合	高级大规模回流-模具底填充
	这一代 HBM 将采用不同的技术，混合键合技术将首先应用于 16 层堆叠		
HBM 5+	混合键合 最高的触点数、 堆叠高度和散热性能	混合键合 最高的触点数、 堆叠高度和散热性能	混合键合 最高的触点数、 堆叠高度和散热性能

资料来源：Besi 官网，华兴证券整理

投资概述

供需层面：2024-27 年 HBM 行业仍处供不应求，高端 HBM 产品价格有望得到支撑

需求层面，大厂资本支出指引和英伟达及 AMD 的 GPU 出货量增长展望尚未发生明显下调。在 2Q/3Q24 业绩会，北美四大云厂商在 3Q24 业绩会后仍维持对 AI 及服务器投资的正面看法。在 2024 年的基础上，2025-26 年来自于四大云厂商的资本支出规模较 2024 年仍同比提升。顺应下游 AI 芯片出货需求，我们预计台积电 CoWoS 产能在 2024 年 12 月和 2025 年 12 月分别达到 36,000/68,000 片（12 寸晶圆）。结合 DRAM 裸片良率和 HBM 堆栈的良率水平，我们预计 AI 芯片直接对应的 HBM 需求量在 2025/26 年分别为 23.6 亿/39.7 亿 GB。

供给层面，SK 海力士、美光科技和三星电子到 2025 年底实际 HBM 出货量有望达 20.1 亿 GB。结合需求供给情况，我们预计 2024-25 年行业仍供不应求。展望 2026-27 年，下一代 GPU 及 HBM4 产品推出进一步拉高了 HBM 需求。叠加行业供给情况，我们认为 2026-27 年，高端 HBM 产品的供给仍供不应求。

- SK 海力士方面，根据公司官网，其 HBM3E 产品在 2024 年第一季度通过英伟达测试，良率在 2024 年 5 月提升到接近 80%，显示出了在工艺经验和客户方面的优势。除现有 DRAM 产能外，根据 SK 海力士官网，公司于 2024 年 6 月宣布了总额达 750 亿美元的产能扩张计划，预期在 2028 年之前投入新 HBM 产能（包括 M15X 和韩国龙仁工厂）。我们预计 SK 海力士 2024-25 年资本支出分别为 15.2 万亿/21.1 万亿韩元。
- 美光方面，我们认为其擅于采用“跳程”战略发力高端 HBM3E 和 HBM4 产品。美光在 HBM2E 之后，直接跳过 HBM3 产品发力 HBM3E 并于 2024 年第一季度通过英伟达测试。根据美光 4QFY24 业绩会信息，公司预计 2025 年财年资本支出达到 2025 财年收入的 35%，相应我们预计美国 2025 财年资本支出达到 131.4 亿美元。美光在 3QFY24 业绩会上表明新增资本支出将用于 HBM3E 和 HBM4 产能扩张，并将在美国纽约州和爱达荷州计划增加先进 DRAM 产线（2027-28 财年逐步投产），我们认为这一举措有望进一步提升美光先进制程（1 γ ）DRAM 的供应能力。我们预计公司 2025 年（自然年）的 HBM 出货量或达 2.2 亿 GB。
- 三星电子方面，根据半导体行业观察，公司 HBM3E 产品客户验证和出货进度慢于 SK 海力士和美光，一定程度丧失了争夺 HBM3E 市占率的先机。三星电子受困于功耗和散热问题，在客户验证进度方面持续慢于市场预期。进入 HBM4 时代，我们预计三星电子基于自身市场份额和利润率考虑会加大与 SK 海力士和美光的竞争。

良率层面：SK 海力士和美光 HBM3E 产品领先，HBM4 或是行业追赶者的发力机会

工艺路线方面，我们预计 SK 海力士或沿用先进 MR-MUF（批量回流模制底部填充）工艺封装 12 层 HBM3E 和 HBM4 产品。但当 HBM 进入 16 层 HBM4 及更先进产品后，我们不排除 SK 海力士选择混合键合方案可能性。美光采用 TCB（热压键合）设备进行键合，其 HBM3E 良率水平或与 SK 海力士相近；而针对 HBM4 产品，我们认为美光将坚持“TCB 设备和混合键合机”两条腿走路的做法可以避免工艺路线选择错误导致的业务损失。不同于 SK 海力士，三星电子一直坚持采用混合键合方案开发 HBM4 产品。虽然前期采用混合键合的成本可能较 MR-MUF 等工艺更高，但我们认为三星电子或因更早采用混合键合技术而在 HBM4 产品的工艺良率和客户验证方面存在一定的优势。

SK 海力士当前在良率水平和客户供货节奏上处于领先。根据 SK 海力士，SK 海力士 HBM3E 在 2024 年第一季度通过客户验证后，其良率水平持续爬坡，2024 年 5 月已接近 80%。根据 SK 海力士 2Q24 业绩会信息，公司计划于 2025 年下半年出货 HBM4，并采用先进 MR-MUF 工艺提高产品良率。美光当前采用 TCB 设备制造 HBM3E 产品，我们预计其良率仍有一定提升空间。三星电子持续因功耗和散热问题不断推迟客户验证通过时间，其 HBM3E 当前良率水平或低于 SK 海力士和美光。

设备层面：混合键合时代，设备投资强度有望加大

TCB 设备：TC-NCF 和先进 MR-MUF 均有需求

针对美光和三星电子采用的 TC-NCF（非导电薄膜热压缩）工艺，根据半导体行业观察，全球 TCB 核心供应商主要包括 Hanmi 半导体、ASMPT、韩国韩华精密机械等。其中美光和 SK 海力士的

TCB 供应商包括 Hanmi 半导体及 Shibaura 等。三星电子 TCB 供应商主要是其子公司 SEMES、Hanmi 半导体在 TCB 市场上处于领先地位，其三款 TC 键合机（Dual TC Bonder 高级型号 DRAGON、Dual TC Bonder 新型号 TIGER、Dual TC Bonder 超级型号 GRIFFIN）是被 SK 海力士采用的主要设备。

2025-27 年，混合键合有望加速推广，设备投资强度有望加大

混合键合工艺下，混合键合机替换热压键合和真空回流焊键合机，带来单位设备和投资强度的增加。

根据半导体设备制造商 Besi 官网，Besi 公司 8800 Ultra Accurate COW（Chip to Wafer）混合键合设备单价在 250 万美元以上，远高于 Besi 单台 TCB 设备单价（100-120 万美元）。结合 HBM 对应的键合裸片数量及混合键合工艺渗透率情况，我们估算 HBM 对应的混合键合设备需求量在 2027 年或达 **90 台以上**，同时来自逻辑芯片封装端的混合键合设备需求量在 **490 台左右**，即总需求量约为 **580 台**。

ASMPT、Hanmi 半导体、Besi、三星电子子公司 SEMES 是全球头部混合键合设备厂。从客户验证和产品出货角度来看，ASMPT、Besi 和韩国三星电子子公司 SEMES 的进度相对较快。其中，ASMPT 在 2Q24 业绩会上宣布获得了两台混合键合设备的订单。

选股

美光科技（MU US，买入，目标价：148 美元）：我们认为美光擅长采用“跳程”策略推出具备竞争力的产品。公司跳过 HBM3 直接发力 HBM3E 产品，并在 2024 年第一季度通过英伟达客户验证。鉴于美光 HBM 业务的低基数以及其 2024-26 财年的资本支出计划，我们预计美光 HBM 业务的市占率到 2025 年有望提升到 11%。新增资本支出支撑 HBM3E 和 HBM4 产能供应。美国纽约州和爱达荷州先进 DRAM 产线有望提升（1y）DRAM 供应能力。

2024-25 财年经营利润率有望持续提升。基于：1) 我们预计美光 HBM 收入占比有望从 2024 财年的 2%提升到 2026 财年的 17%，从而改善公司利润率水平；2) 与此同时，我们预计服务器 DRAM 价格在 1Q25 有涨价空间。我们给予美光基于分部估值的目标价是 148 美元目标价，主要包括：DRAM 业务给予 2026 财年 11.5x P/E；NAND 业务给予 2026 财年 8.0x P/E；其他业务给予 2026 财年 2.0x P/E；我们在估值中的净现金假设是 90.8 亿美元。

SK 海力士（000660 KS，买入，目标价：**259,548 韩元**）：我们认为 HBM 产能及良率爬坡助力 SK 海力士持续释放 HBM 产量，带来更高收入和利润。我们预计公司 2025 年 HBM 产品出货量或超 4,400 万个堆栈。虽然 2026-27 年可能会面临来自友商竞争而出现可能市占率侵蚀，但我们看好 SK 海力士基于更高的 HBM 出货量持续提升收入和利润水平。

我们给予 SK 海力士买入评级并且预计公司 2024-25 年经营利润率为 35.5%和 40.6%，驱动原因包括：1) 产品结构改善-HBM 占公司整体收入比例有望从 2024 年 20%提升到 2025 年的 33%，且 HBM 利润率高于传统 DRAM 和 NAND，有助于公司整体利润率改善；2) 2024-2025 年 DRAM 和 NAND 利润率回升--SK 海力士或进一步控制自身传统 DRAM 和 NAND 产能，主要依靠现有库存消耗维持市占率；和 3) 我们认为服务器 DRAM 价格在 4Q24 或环比提升个位数百分比。另一方面，我们认为 HBM 产能投放会进一步挤压传统 DRAM 产能，带来 DRAM 均价的进一步上涨。我们给予 SK 海力士基于分部估值的目标价是 **259,548 韩元**，主要包括：DRAM 业务给予 2025-26 年的平均净利润 6.5x P/E；NAND 业务给予 2025-26 年的平均净利润 4.0x P/E；其他业务给予 2025-26 年的平均净利润 1.0x P/E；我们在估值中的净现金假设是 11.5 万亿韩元。

风险提示

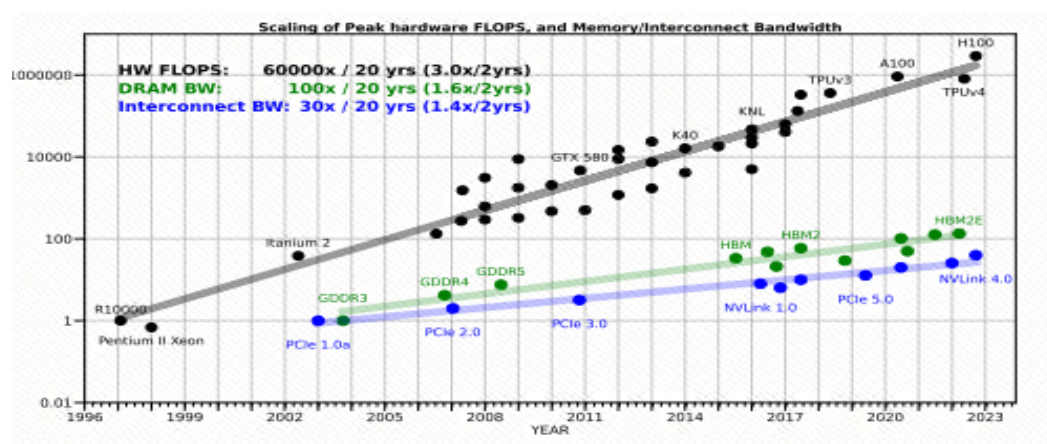
通用性 AI 需求不及预期；美光 HBM3E 产能爬坡和良率水平不及预期；NAND 和 DRAM 供过于求时间点早于预期。

需求侧：AI 大模型爆发拉动 HBM 需求

内存墙问题阻碍人工智能发展。根据《AI and Memory Wall》，过去 20 年中服务器硬件 FLOPS 峰值每 2 年扩大至 3 倍，远超 DRAM 和互连带宽增速，我们认为该差异是限制人工智能应用的瓶颈之一。

HBM 方案解决了内存墙问题，直接受益于 AI 模型需求。GPU 显存一般采用 GDDR（Graphics Double Data Rate）或者 HBM 两种方案，但 HBM 性能远超 GDDR：根据 AMD 发布的 HBM 与 DDR（Double Data Rate）内存参数对比，从时钟频率看，HBM 为 500MHz，远小于 GDDR5 的 1750MHz；从显存带宽来看，HBM 的一个堆栈大于 100GB/s，而 GDDR5 的一颗芯片为 25GB/s。根据海力士官网，其 GDDR6 产品带宽在 56GB/s，而 HBM2 和 HBM2E 的带宽分别达到 256 GB/s 和 368GB/s，远超 GDDR6 产品。

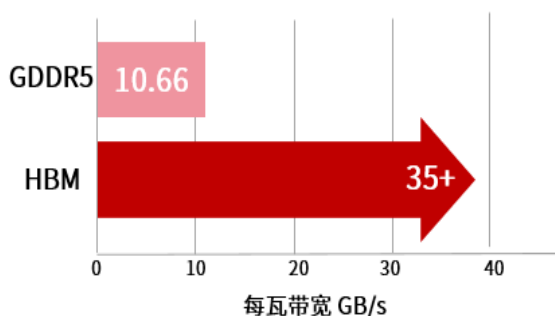
图表 5：相比 GDDR6，HBM2 以及 HBM2E 在带宽和能耗方面均明显提升



资料来源：《AI and Memory Wall》，华兴证券整理

图表 6：存储容量升级拉动英伟达 H100 训练能力提升

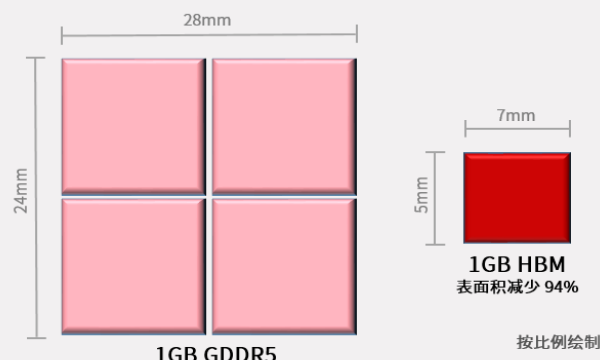
HBM vs GDDR5:
更好的每瓦带宽



资料来源：AMD 公司公告，华兴证券整理

图表 7：存储容量升级拉动英伟达 H100 推理能力提升

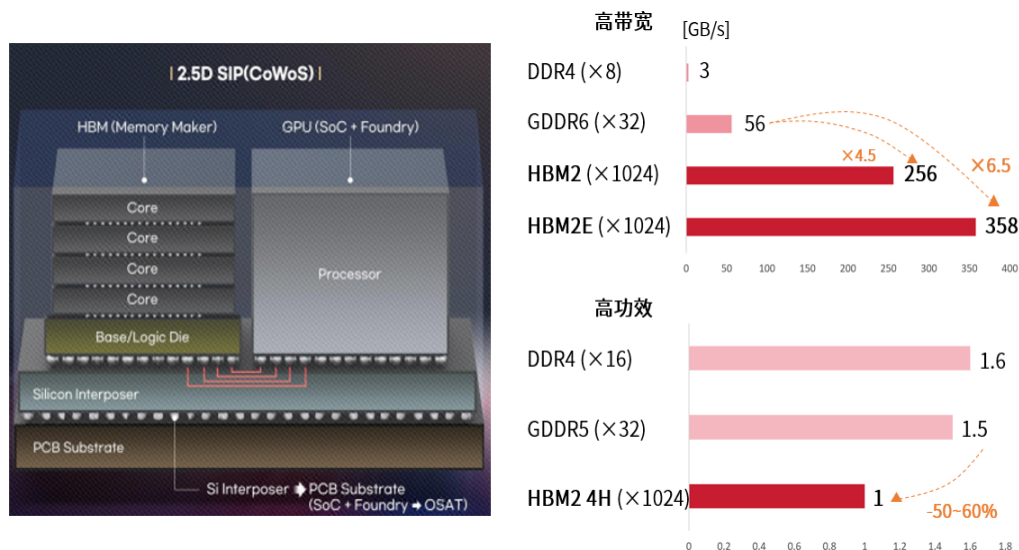
HBM vs GDDR5:
节省大量空间



资料来源：AMD 公司公告，华兴证券整理

-
-
-

图表 8：相比 GDDR6，HBM2 以及 HBM2E 在带宽和能耗方面均明显提升

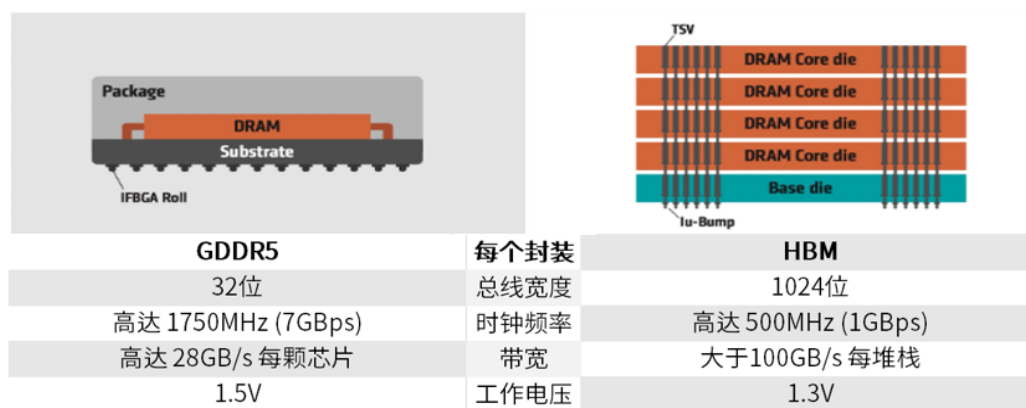


资料来源：SK 海力士公司公告，华兴证券整理

根据《高带宽存储器测试技术研究》，HBM 采用 3D 堆栈技术实现高带宽和大容量存储，通过硅通孔（TSV）堆栈方式将 DRAM 裸片叠加，连同 SoC 裸片和中介层封装在一起，构成了完整 3D 封装芯片。根据《高带宽存储器测试技术研究》，相比传统 DDR，HBM 优势是：1) HBM 具备高带宽优势。通过 TSV 将 DRAM 裸片叠加，HBM 可以达到更高 I/O 数量，助力显存位宽达到 1024 位，远超 GDDR 的 32 位。2) HBM 总带宽上限高于 GDDR5。GDDR 采取大幅提升 I/O 数据速率手段改善总带宽，而 HBM 则利用 TSV 技术提升 I/O 数和单个 I/O 位宽。

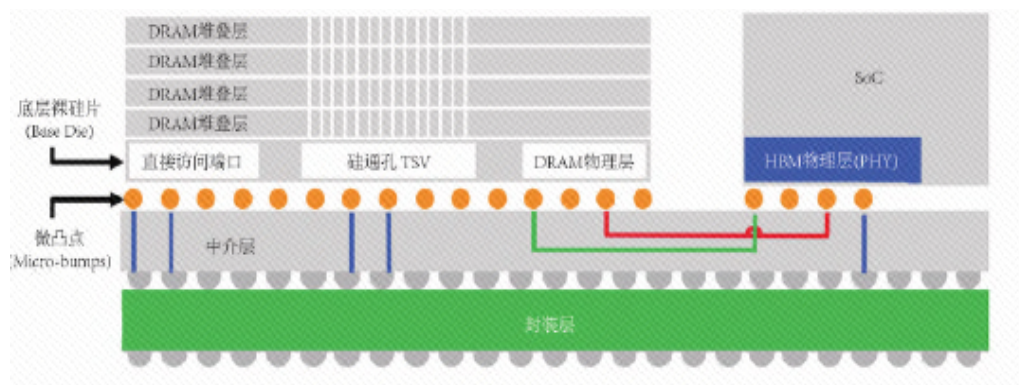
图表 9：GDDR5 和 HBM 产品对比

并排比较



资料来源：AMD 公司公告，华兴证券整理

图表 10：一种工程实现的 HBM3 堆栈结构

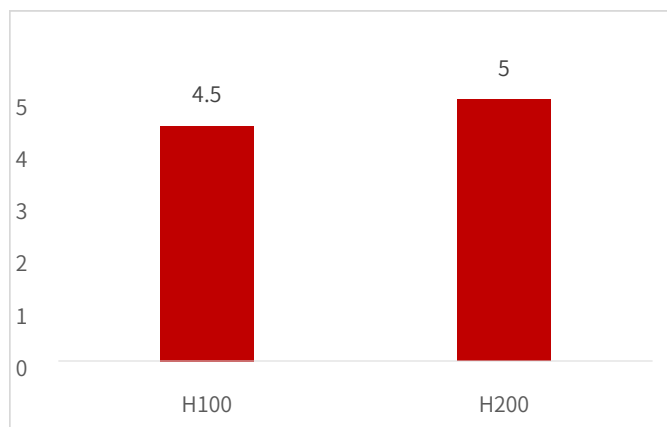


资料来源：《高带宽存储器测试技术研究》，华兴证券整理

HBM 在 GPU 中得到广泛应用

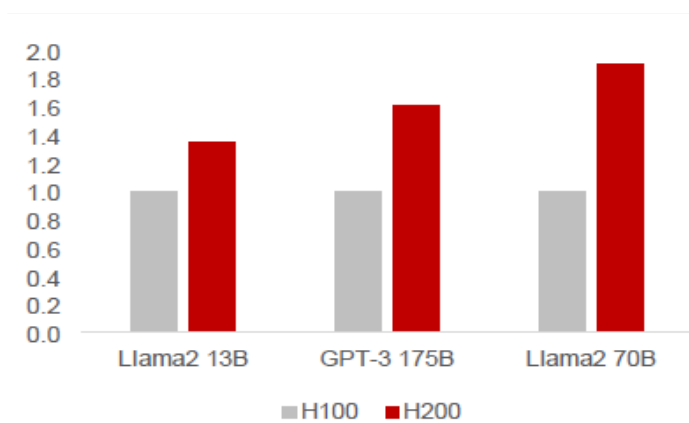
主流数据中心 GPU 均采用 HBM 技术。根据《英伟达 H100 GPU 白皮书》，英伟达 V100、A100 和 H100 产品均采用 HBM 内存，其中 A100 使用 HBM2，H100 产品则升级到 HBM2E 和 HBM3。根据 AMD 官网，其 MI300 芯片采用 HBM3 产品。我们看到英伟达和 AMD 的主流 GPU 产品均采用 HBM2E 和 HBM3 产品，凸显了 HBM 在数据中心 GPU 中的主导地位。

图表 11：存储容量升级拉动英伟达 H200 训练能力提升



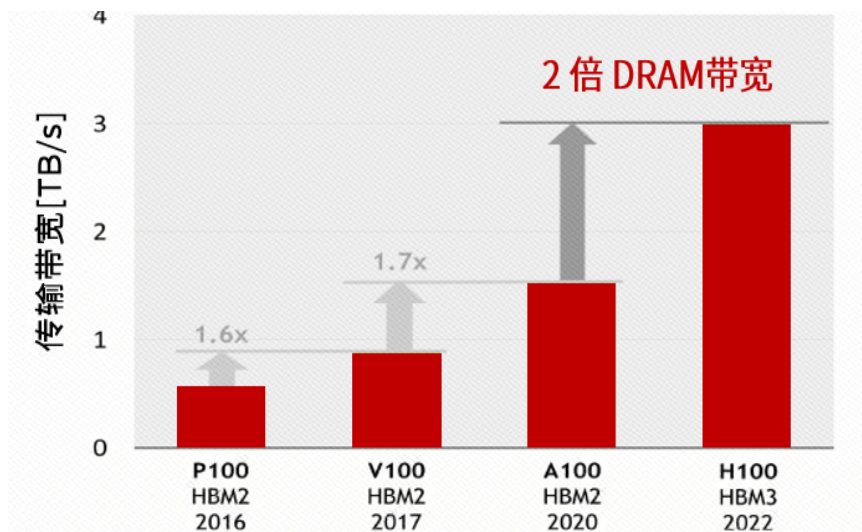
注：单位指的是原文中给出的参考基准的倍数。
资料来源：英伟达公司官网，华兴证券整理

图表 12：存储容量升级拉动英伟达 H200 推理能力提升



注：单位指的是原文中给出的参考基准的倍数。
资料来源：英伟达公司官网，华兴证券整理

图表 13：英伟达历代 GPU 对应 DRAM 带宽比较



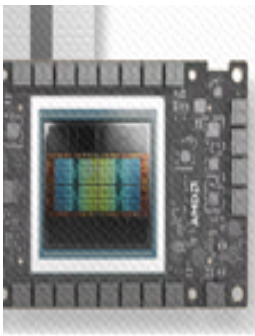
资料来源：《英伟达 H100 GPU 白皮书》，华兴证券整理

图表 14：英伟达 H100 和 A100 GPU 采用 HBM 产品

GPU特征	NVIDIA A100	NVIDIA H100 SXM5	NVIDIA H100 PCIe
GPU架构	NVIDIA Ampere	NVIDIA Hopper	NVIDIA Hopper
GPU主板外形规格	SXM4	SXM5	PCIe 5.0
显存位宽	5120 位 HBM2	5120 位 HBM3	5120 位 HBM2e
显存容量	40 GB	80 GB	80 GB
显存数据速率	1215 MHz DDR	尚未最终确定	尚未最终确定
显存带宽	1555GB/s	3000 GB/s	2000 GB/s
(H100未最终确定)			

资料来源：《英伟达 H100 GPU 白皮书》，华兴证券整理

图表 15：AMD MI300X 加速器采用 HBM3 产品

	显示卡記憶體	
	末級快取記憶體 (LLC)	256 MB
	專用記憶體大小	192 GB
	專用記憶體類型	HBM3

资料来源：AMD 官网，华兴证券整理

HBM 升级换代伴随带宽和速率提升; 同时伴随 HBM 性能和 HBM 颗数增加

HBM 技术伴随下游客户需求不断演进。根据 SK 海力士和半导体行业观察，HBM2 是 SK 海力士在 2018 年发布的产品，其 I/O 速率和带宽分别是 2Gbps 及 356GB/s，能够满足下游客户的传输和存储需求。随着 GPU 及云服务大厂 AI 加速器的推出，三星电子和美光在 2020 年相继发布了 HBM2E。随着英伟达发布 A100 和 H100 系列，由于传统的 HBM 带宽无法满足高速存储数据联通的需求，相应 HBM3 在 2023 年正式问世，带宽为 819GB/s，速度相比第一代 HBM1 提升超过 3 倍。

功耗相同的前提下，HBM 带宽和容量提升带动性能提升。根据半导体行业观察微信公众号，提升 HBM 容量方法包括：1) 增加 DRAM 裸片堆栈层数，例如 HBM3E 常见为 8 层或 12 层 DRAM 裸片堆栈，但进化到 HBM4 时代，可选择的包括 4 层、8 层、12 层和 16 层 TSV 堆栈；2) 增加每片 DRAM 裸片 (Die) 的存储容量，例如单 Die 容量从 2GB 升级到 3GB。单颗 HBM 容量=DRAM 堆栈层数×每片 DRAM 容量。HBM 带宽提升方法包括：1) 提高 I/O 数据传输速率；2) 加大 I/O 总线位宽。

除了单颗 HBM 容量，单颗 GPU 中 HBM 存储总容量还取决于 GPU 中包含的 HBM 颗数。未来进化到 GB200，会出现 2 颗 GPU 搭配 8 颗 12 层 HBM3E (36GB) 的技术方案。

图表 16：HBM 技术演进路线

HBM 代数	HBM	HBM2	HBM2E	HBM3	HBM3E/HBM3P
供应商	SK 海力士	三星电子, SK 海力士	三星电子, SK 海力士, 美光	三星电子, SK 海力士	三星电子, SK 海力士, 美光
产品首次发布年份	2014	2018	2020	2022-2023	2024
每叠典型芯片数 (主要封装方式)	4Hi (TSV & 微凸点)	4-8Hi (TSV & 微凸点)	4-8Hi (TSV & 微凸点)	8-12Hi (TSV & 微凸点)	8-12Hi (TSV & 微凸点)
每叠最大容量	1GB	4-8GB	8-16GB	16-24GB	24-36GB
芯片密度 (典型工艺)	2Gb (2x)	8-16Gb (2γ,2z)	16Gb (1γ,1z)	16Gb (1z)	24Gb (1a,1b/1β)
最大数据速率	1Gbps	2-2.4Gbps	3.2-3.6Gbps	5.6-6.4Gbps	8.0-9.8Gbps
有效总线宽度	1,024	1,024	1,024	1,024	1,024
每叠最大带宽	128GB/s	205-307GB/s	460GB/s	819GB/s	1.2TB/s

资料来源：SK 海力士，美光，三星电子，Besi，华兴证券整理

HBM4 是 HBM3 的进化版本，旨在进一步提高数据处理速率及带宽，并降低功耗和堆栈容量。根据半导体行业观察，相较于 HBM3，HBM4 的提升之处体现在：1) HBM4 把内存堆栈接口从 1024 位扩展至 2048 位，对应的 I/O 引脚数量增加两倍。2048 位内存接口将显著增加通过内存堆栈布线的硅通孔数量。2) 外部芯片接口将凸块间距缩小到 55nm 以下，同时大幅增加微凸点总数。3) DRAM 裸片的堆栈层数有望达 16 层。HBM4 有望在一个 HBM 产品中堆栈 16 层裸片，即 16Hi 堆栈。

图表 17：HBM4 与前代产品比较

	HBM4	HBM3E	HBM3	HBM2
接口宽度	2048 位	1024 位	1024 位	1024
最大堆叠容量	36GB-64GB	36GB	24GB	16GB
数据传输率	待定	9.2 GT/S	6.4 GT/S	3.6 GT
DRAM Ics 每栈	16	NA	NA	8
每栈带宽 (不高于)	1.5 TB/S - 2+ TB/S	1.2 TB/s	819.2 GB/S	460.8

资料来源：半导体行业观察，华兴证券整理

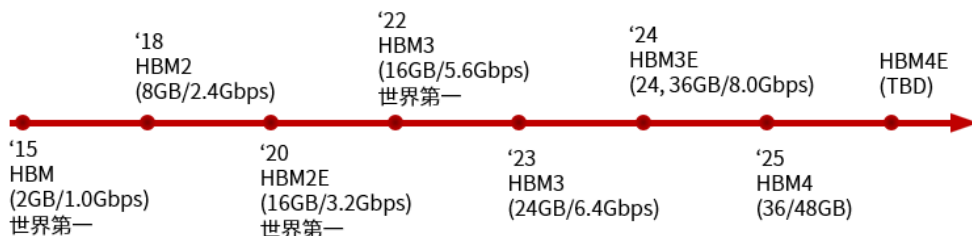
SK 海力士、三星电子和美光布局 HBM4，单颗 HBM 容量提升

我们认为三大原厂明确 HBM4 产品的推出时间有助于提升 HBM 产品容量。1) 根据半导体行业观察，三星电子在其设备解决方案部成立了“HBM 开发团队”，旨在增强其 HBM 竞争力，而这一举措是在副董事长 KyungKyeheun（庆桂显）就任设备解决方案（DS）部门负责人一个月后采取的。2) 根据 SK 海力士 2Q24 业绩会，公司表示将在 2025 年下半年推出下一代 HBM4，并计划在 HBM4 芯片中采用台积电的先进逻辑工艺。3) 根据美光中国官方微信公众号，美光预计 2025-26 年推出 12Hi 和 16Hi 的 HBM4，容量分别是 36GB 和 48GB，两款产品速度均在 1.5TB/s 以上。继 HBM4，HBM4E 和 HBM5 将于 2028 年问世。根据美光中国官方微信公众号，HBM4E 预计将获得更高时钟频率，并将带宽提高到 2TB/s 以上，容量提高到每个堆栈 48GB 到 64GB。

图表 18：海力士明确 HBM4 布局时间图

HBM 产品历史

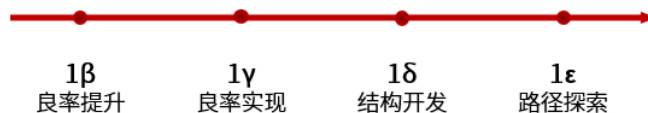
HBM 能够使 AI 系统的数据处理更快、更高效，并将针对各类 AI 客户进行进一步定制。



资料来源：半导体行业观察，华兴证券整理

图表 19：美光明确 DRAM 1 γ 路线时间

- 1 β 工艺的先进CMOS技术
- 1 γ 节点引入极紫外光刻技术
- 延伸平面DRAM技术路线图
- 高数值孔径EUV研发投入
- 持续投资于3D DRAM的研发



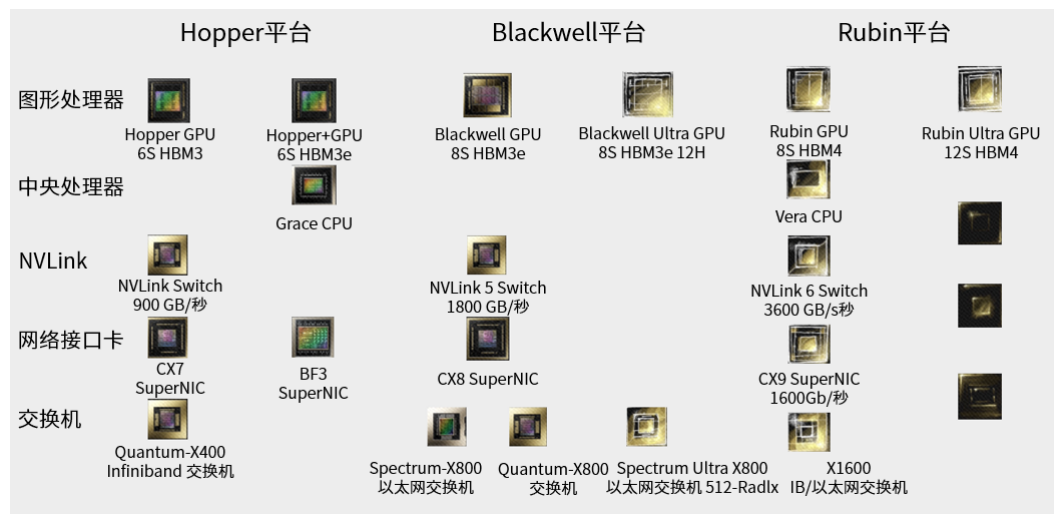
资料来源：半导体行业观察，华兴证券整理

英伟达下一代 GPU 的 HBM 颗数增加

英伟达下一代 Rubin 架构 GPU 或推高 HBM 颗数

- 1) 从 Hopper+ GPU 到 Blackwell 架构 GPU。根据 TrendForce，英伟达采用 HBM3E 产品，其中 HBM 颗数从 6 颗提升到 8 颗，对应超 30%容量提升，以应对更高 GPU 产品的推理和训练性能。
- 2) 从 Blackwell 架构 GPU 到下一代 Rubin Ultra GPU 产品。根据 TrendForce，除了 HBM 产品从 HBM3E 升级到 HBM4 外，英伟达 Rubin Ultra GPU 产品或将采用 12 颗 HBM4 产品。假设一颗 HBM4 容量 36GB，12 颗 HBM4 将赋予英伟达 Rubin Ultra GPU 产品最高 412GB 容量。我们认为下一代 GPU 产品的出货有望拉动对应 HBM 数量和单个产品存储容量的提升。

图表 20：英伟达历代架构对 HBM 颗数的需求量



资料来源：TrendForce，华兴证券整理

图表 21：英伟达和 AMD 相应 AI 芯片 HBM 规格比较

公司	AI 芯片	2022	2023	2024 (F)	2025 (F)
英伟达	H100	HBM3 8H 80GB			
	GH200 (CPU+GPU)			HBM3e 8H 141GB	
	H20			HBM3e 8H 96GB	
	H200			HBM3e 8H 141GB	
	B100			HBM3e 8H 192GB	
	GB200 (CPU+GPU)			HBM3e 8H 192/384GB	
	B200			HBM3e 8H 288GB	
AMD	MI200	HBM2e 8H 128GB			
	MI300X		HBM3 12H 192GB		
	MI300A (CPU+GPU)		HBM3 8H 128GB		
	MI350			HBM3e 12H 288GB	
	MI375 (CPU+GPU)				HBM3e 12H 288GB

资料来源：TrendForce，华兴证券整理

英伟达、AMD 及云服务大厂的 AI 芯片出货拉动 HBM 需求

综上所述 2025-2027 年，我们认为 HBM 升级换代体现在：1) HBM 堆栈层数增加；2) 单个裸片存储容量扩大。根据国际半导体标准组织（JEDEC）最新发布的 HBM4 规范，HBM4 堆栈层数和每个裸片容量均有改变。HBM4 产品将指定 24GB 和 32GB 两种，并支持 4 层、8 层、12 层和 16 层 TSV 堆栈。与此同时，英伟达下一代 Rubin 架构或推高单颗 GPU 中使用的 HBM 颗数。我们认为国际半导体标准组织 JEDEC 的新规范和英伟达未来在产品方面的新举措共同推动 HBM 需求。

英伟达和 AMD 在采购三大原厂 HBM 后，需要通过台积电 CoWoS 产能进行统一封装。我们从 CoWoS 产能推测出 2024-27 年 GPU 产出数量，并以此计算对 HBM 产品的需求量。我们的主要假设包括：

- 1) 台积电 CoWoS 2024-26 年 12 月的月产能分别是 36,000、68,000 及 92,000 片晶圆，2024-26 年实际产能是 311,000、604,000 及 972,000 片晶圆。
- 2) 2024-26 年台积电 CoWoS 产能利用率为 100%、97%及 97%。
- 3) 2024-26 年，单 GPU 中 HBM 的平均需求数分别是 **7.0、7.5 及 7.8 个**；HBM 堆栈层数分别是 **8.5、9.0 及 9.5**；HBM 单颗携带存储容量是 **2.6GB、2.8GB 及 3.0 GB**。
- 4) 2024-2026 年，每 GB 容量的价格为 15-16 美元。

图表 22：2023-27 年 HBM 需求量估算（百万 GB）

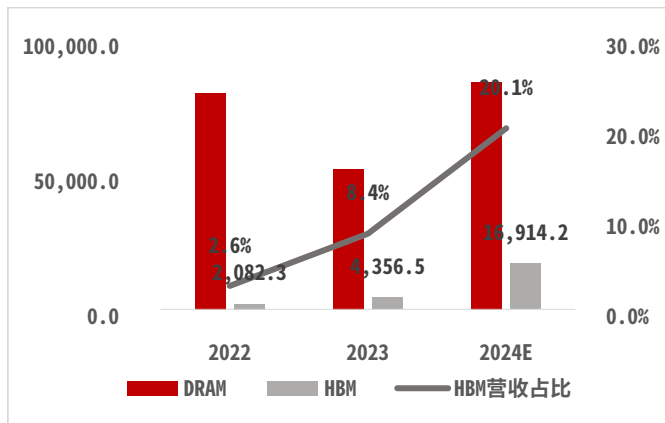
需求量估算	2023	2024E	2025E	2026E	2027E
台积电 CoWoS 年末月产能估算（千片）	15	36	68	92	110
YoY		137%	92%	35%	20%
台积电 CoWoS 年实际产能（千片）	109.592	311	604	972	1221
YoY		184%	94%	61%	26%
产能利用率（%）	100%	100%	97%	97%	96%
台积电 CoWoS 实际年产量（千片）	110	311	586	943	1,172
台积电外溢 CoWoS 实际年产量（千片）	25	50	85	109	131
合计 CoWoS 实际年产量（千片）	135	361	671	1,052	1,303
英伟达占比（%）	46%	51%	51%	48%	49%
英伟达获得 CoWoS 的产能（千片）	62.0	184.0	343.0	504.8	638.3
AMD 及其他 AI 芯片设计厂占比（%）	54%	49%	49%	52%	51%
AMD 及其他 AI 芯片设计厂商获得 CoWoS 产能（千片）	72.6	177.0	327.8	546.9	664.4
单片 CoWoS 可封装 GPU 颗数	28.0	25.0	18.6	17.0	16.5
YoY		-11%	-26%	-9%	-3%
对应 GPU+ASIC 出货量（百万颗）	3.8	9.0	12.5	17.9	21.5
单 GPU 中平均 HBM 需求个数	6.0	7.0	7.5	7.8	8.0
HBM 堆栈平均层数	8.0	8.5	9.0	9.5	10.0
HBM 单颗携带存储容量（GB）	2.5	2.6	2.8	3.0	3.0
单 GPU 中需求的存储容量（GB）	120.0	154.7	189.0	222.3	240.0
GPU 需求的存储总容量（百万 GB）	452.2	1,396.2	2,357.0	3,974.3	5,158.8

资料来源：TrendForce，华兴证券预测

供给测：良率和工艺问题拖累美光和三星电子出货进度

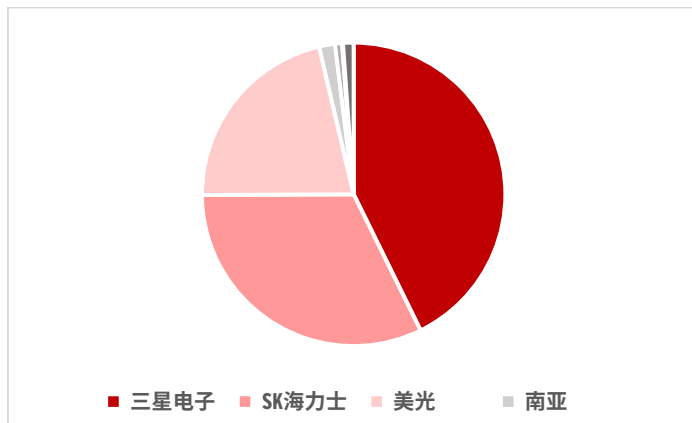
AI 芯片的驱动下，2023-2024 年间，HBM 在 DRAM 行业中的占比一直在上升。根据 TrendForce，截至 2024 年底，全球生产 HBM TSV 产能约 250K/m，占总 DRAM 产能约 14%。我们预计 HBM 在 DRAM 产业的占比有望从 2023 年 8.4%提升到 2024 年底的 20.1%。HBM 方面，三大厂商采取积极扩产战略，其中根据半导体行业观察，SK 海力士计划 2028 年前投资 103 万亿韩元，其中约 82 万亿韩元用于 HBM 投资。美光也在 3QFY24 业绩会上宣布 2025 年自然年其 HBM 市占率将与它的 DRAM 市占率相当，达到约 20-25%。三星电子也在积极扩产。

图表 23：HBM 市场空间（百万美元）及在 DRAM 中渗透率（%）



资料来源：TrendForce 预测，华兴证券整理

图表 24：4Q23 期间 DRAM 市场市占率(%)



资料来源：TrendForce，华兴证券整理

HBM3 及 HBM3E，海力士客户验证及出货领先于美光和三星电子

2024 年市场焦点正从 HBM3 向 HBM3E 切换。虽然主流 GPU 芯片 H100 和 H200 等配套 HBM3，但我们预计 HBM3E 在 2024 年下半年逐步放量，并有望成为三大原厂逐鹿市占率的重要战场。

在 HBM3 方面，根据半导体行业观察，SK 海力士最早于 3Q22 便已通过英伟达测试，目前英伟达 H100 搭配的 HBM3 产品来自 SK 海力士。三星电子凭借与 AMD 的长期合作关系，其 HBM3 产品也通过了 AMD MI300 系列产品验证，包括 8 层和 12 层 HBM 产品，在 1Q24 之后放量。根据半导体行业观察，由于美光始终没有加入 HBM3 供应系列，2024 年 HBM3 产品的主要参与者为 SK 海力士和三星电子。

在 HBM3E 方面，根据 TrendForce，SK 海力士 24GB 产品于 1Q24 率先通过客户验证，美光在 1Q24 底向客户递交 HBM3E 量产产品，并在 2Q24 向英伟达 H200 供应 HBM3E。

图表 25：HBM3E 供货商进展（截止 2024 年 3 月）

	品牌	速度 (Gbps)	技术节点	2022				2023				2024				2025				2026			
				1Q22	2Q22	3Q22	4Q22	1Q23	2Q23	3Q23	4Q23	1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26	4Q26
HBM3e	三星	8-9.8	1 alpha 24Gb									24GB	36GB										
	SK 海力士	8-9.2	1 beta 24Gb									24GB	36GB										
	美光	8-9.2	1beta 24Gb									24GB	36GB										

资料来源：TrendForce，华兴证券整理

我们看到三家厂商把扩充 HBM 产能作为提升市占率和利润率的重要抓手。具体原因有：

美光在 4QFY24 业绩会上表明 HBM 出货量快速爬升。HBM3E 产品收入在 3QFY24 超 1 亿美元，同时拉动公司利润率水平。根据美光产能规划和客户验证导入情况，我们预计 HBM 在 2025 财年为美光贡献超 10 亿美元收入。

根据“Memcon 2024”全球半导体企业大会信息，三星电子 HBM 业务的目标包括：1) 短期内，三星电子在“Memcon 2024”全球半导体企业大会上表示，2024 年 HBM 产量比 2022 年增加 2.9 倍；2) 长期内，三星电子预计 2026 年 HBM 出货量是 2023 年的 13.8 倍，2028 年产量增加到 2023 年的 23.1 倍。

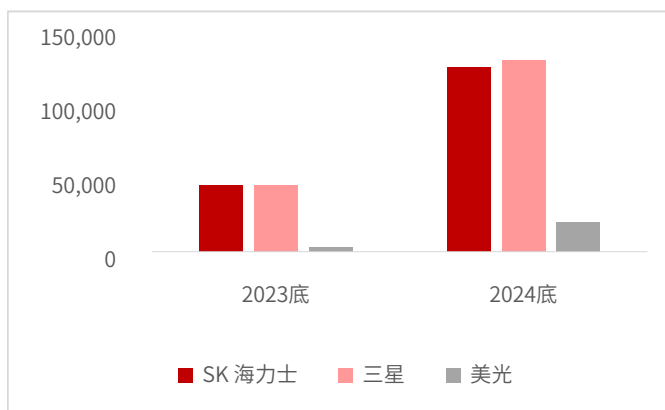
SK 海力士计划对 HBM 重点投入。1) 短期目标：根据 SK 海力士 2Q24 业绩会，公司计划在 2024 年增加资本支出，并将资本开支重心放在 HBM 等存储上，预计 2024 年 HBM 产能比 2023 年增加一倍以上。2) 长期目标：根据路透社信息，SK 集团 6 月 30 日宣布旗下 SK 海力士将投资 103 万亿韩元（约合 747 亿美元），加强人工智能存储芯片业务。其中约 80%（即 82 万亿韩元）投资用于 HBM 芯片，以推动 AI 芯片发展。出货量方面，SK 海力士预计到 2030 年其 HBM 出货量将达到 1 亿颗。

图表 26：英伟达和 AMD 相应 AI 芯片的 HBM 规格比较

	品牌	速度 (Gbps)	技术节点	2022				2023				2024				2025				2026			
				1Q22	2Q22	3Q22	4Q22	1Q23	2Q23	3Q23	4Q23	1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26	4Q26
HBM2e	三星	3.2-3.6	1Y 16Gb	8/16Gb																			
	SK 海力士	3.6	1Y 16Gb	8/16Gb																			
	美光	3.2-3.6	1Z 16Gb	16Gb																			
HBM3	三星	5.6-6.4	1Z 16Gb																				
	SK 海力士	5.6-6.4	1Z 16Gb																				
HBM3e	三星	8-9.2	1 alpha 24Gb																				
	SK 海力士	8-9.2	1 beta 24Gb																				
	美光	8-9.2	1beta 24Gb																				
HBM4	待定	待定	完整规格可能在2024年下半年至2025年发布；客户/服务器（版本）在2026年（发布）																				

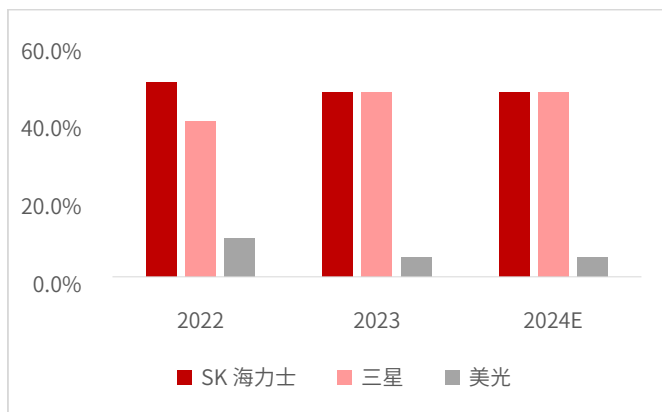
资料来源：TrendForce，华兴证券整理

图表 27：三星电子、SK 海力士和美光年末 12 月产能（晶圆，片）



资料来源：TrendForce，华兴证券预测

图表 28：三星电子、SK 海力士和美光 HBM 市占率 (%)

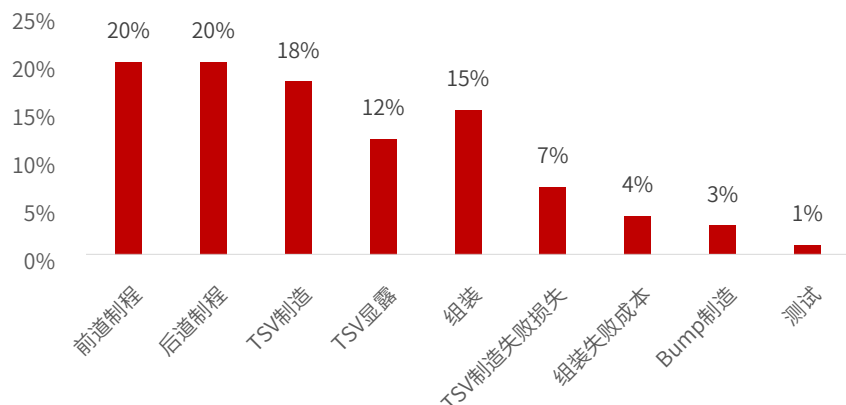


资料来源：TrendForce，华兴证券预测

良率问题影响美光和三星电子 HBM 出货速度

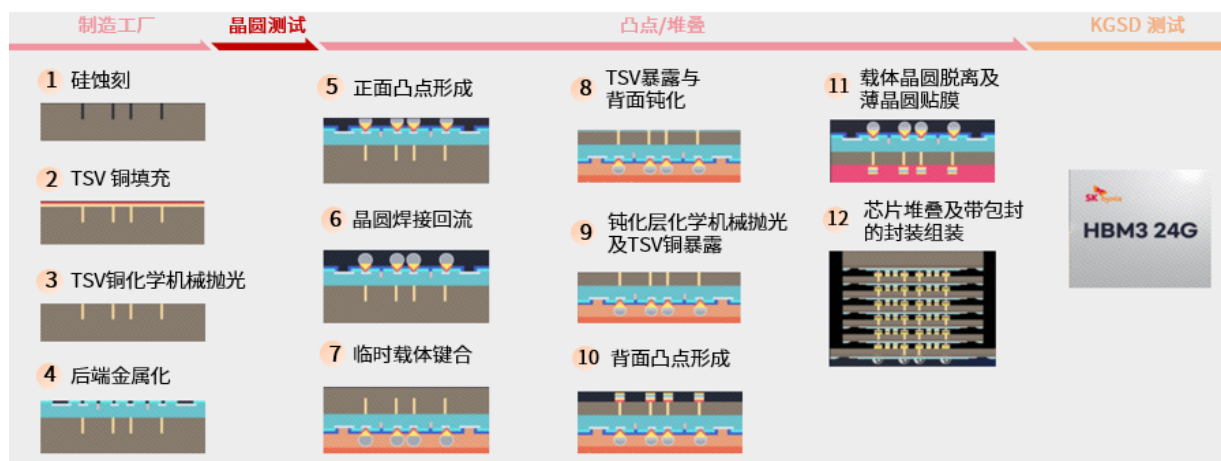
根据 SK 海力士投资者日信息，HBM 工艺流程包括 TSV 制造、TSV 显露、凸块制造及堆栈键合等环节。其中 TSV 制备、凸块制备及堆栈键合是其中最为关键的三个步骤，合计占 HBM 封装成本 40%以上。这三个步骤完成好坏直接影响 HBM 良率。根据半导体行业观察，截至 2024 年 6 月 19 日，三星电子 HBM 由于发热和功耗问题尚未通过英伟达测试。另外美光于 2024 年 4-5 月期间在 HBM3E 供货方面出现不少良率问题，充分体现了 TSV 制备和堆栈键合技术高低直接影响 HBM 良率和供应能力。

图表 29：HBM3D 封装成本构成（99.5%键合良率）



资料来源：3D incities，华兴证券整理

图表 30：HBM 封装全流程

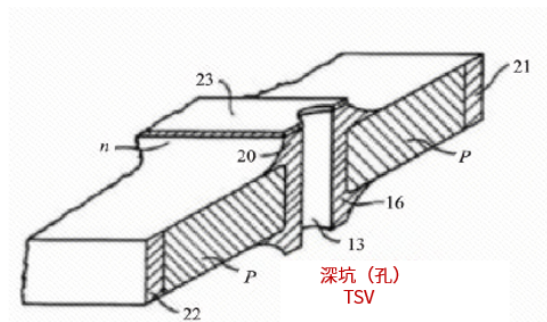


资料来源：SK 海力士投资者日信息，华兴证券整理

TSV 是制备 HBM 的第一个关键环节

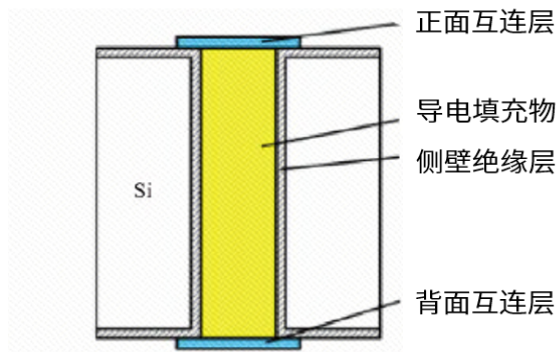
根据名为《硅通孔三维互连与集成技术》的一篇论文（2024 年 6 月发表于《电子与封装》），TSV 是一种连接硅晶圆上下两面并与硅基板和其他通孔绝缘的电信号互连结构。硅通孔起源于 1958 年 William Shockley 申请的专利“半导体晶圆及其等效化方法”。TSV 结构包括可穿透硅基板的导电填充物与侧壁的绝缘层，同时为了实现硅基板上下面的电气互联，还需要正面与背面互连从而实现信号联通。

图表 31: William Shockley 专利中提到的 TSV



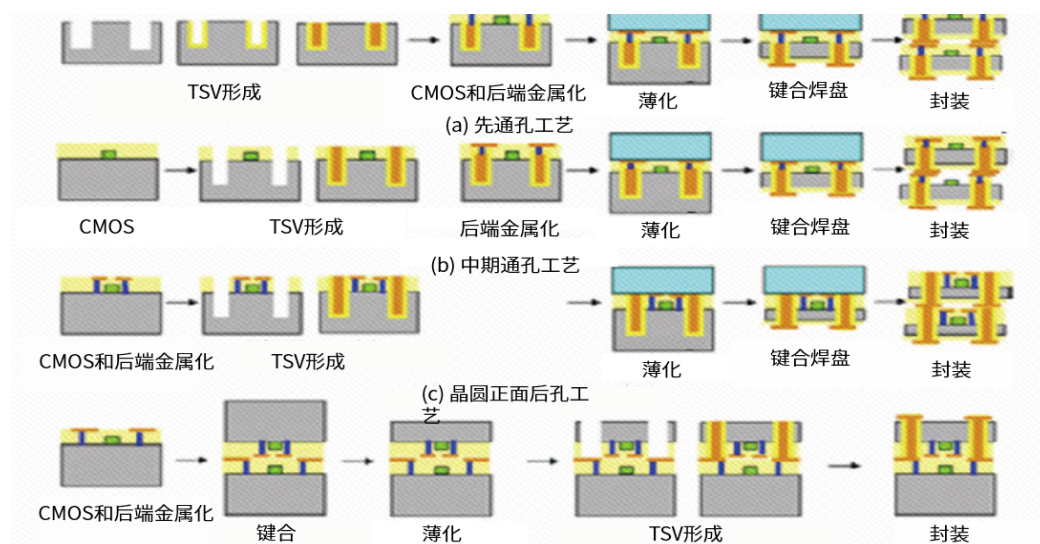
资料来源:《硅通孔三维互连与集成技术》，华兴证券整理

图表 32: TSV 基本结构示意图



资料来源:《硅通孔三维互连与集成技术》，华兴证券整理

图表 33: TSV 工艺流程



资料来源:《硅通孔三维互连与集成技术》，华兴证券整理

根据《硅通孔三维互连与集成技术》介绍的 3D 集成, TSV 工艺方法包括先通孔 (Via-First)、中通孔 (Via-Middle) 及后通孔 (Via-Last)。其中, Via-Last 又分为晶圆正面的后孔 (Front Side Via-Last) 及从晶圆背面的后孔 (Back Side Via-Last) 两种, 是目前 TSV 产业化最成熟的方案之一。

根据《硅通孔三维互连与集成技术》, 当前影响 TSV 应用后电性能的三大核心包括 TSV 刻蚀技术、TSV 侧壁绝缘技术和 TSV 微孔金属化技术。1) TSV 刻蚀: 常见硅刻蚀侧壁缺陷体现为粗糙度大及出现硅缺口 (Notch), 直接影响 TSV 集成后电性能表现。2) TSV 侧壁绝缘: TSV 侧壁需要绝缘, 防止金属和硅短路, 这对器件可靠性至关重要。通常情况下, TSV 介电绝缘层需要良好台阶覆盖和均匀性, 以保证高击穿电压下, 产品不开裂并与工艺温度具备一定相容性。3) TSV 微孔金属化技术: 微孔金属化实现器件信号互连, 是 TSV 核心技术之一。TSV 微孔金属化技术重点是阻挡层、种子层和导电层, 需关注硅通孔内金属填充效果。

根据《硅通孔三维互连与集成技术》, 2.5D 封装中, 在我们观点里, 采用 TSV 有以下优势: 1) 能提供更短电路连接, 大幅提高信号的传输速度; 2) 能实现高密度、高深宽比连接, 拥有更多信号

通道；3）能替代效率低下的引线键合方式，使信号传输速度更快、功耗更少，并保证传递功率一致性；4）能使高密度堆栈成为可能，拥有更高的封装密度，有效降低成本。

图表 34：TSV 通孔工艺研究进展总结

工艺方法	工艺机理	技术难点	研究进展	应用特色
Bosch 刻蚀	刻蚀与钝化循环进行	形成扇形侧壁 微观负载效应 RIE 滞后	✓ 优化射频功率和刻蚀气体流量使扇形侧壁平均深度降低 91%	主要用于 MEMS 适用于极高深宽比的通孔刻蚀 高速率 高垂直度 高选择性
			✓ 利用 KOH 溶液的各向异性刻蚀进行表面抛光消除了扇形粗糙度和锥形效应	
			✓ 减小通孔间距，集中热量以抑制刻蚀滞后	
			✓ 渐进式调整工艺参数形成超高深宽比通孔（深宽比>40）	
			✓ 改良 Bosch 两步/三步刻蚀工艺使滞后降低至 1.5%以下	
激光钻孔	光化学烧蚀和光热烧蚀	严重形变 等离子体屏蔽 热影响区 溅渣沉积	✓ 多脉冲纳秒紫外微光制备的通孔来产生形变皮秒、纳秒脉冲串组合脉冲制备高深宽比通孔来产生形变	主要应用于低通孔密度需求的应用 高效率 高精度 高灵活性 工艺简单
			✓ 施加横向电场和横向气流缓解等离子体屏蔽及热影响区	
			✓ 利用掩膜避免了羽流杂质沉积	
湿法刻蚀	蚀刻剂与硅发生化学反应并腐蚀	蚀刻液污染环境 侧壁不垂直 K ⁺ 污染 CMOS 器件	✓ 在超临界二氧化碳环境中短时间形成高深宽比 TSV	常用于硅悬臂梁或梯形结构的微加工 高速刻蚀 低成本 设备简单 对衬底没有额外损伤
			✓ 结合熔融金属吸附法在厚硅衬底上形成 TSV 使用 HNA 溶液制备大批量高圆度微加工通孔，侧壁光滑	
			✓ 利用 TMAH 溶液进行选择性刻蚀 TMAH 溶液可兼容碱性刻蚀和 CMOS 器件	

资料来源：《三维系统级封装(3D-SiP)中的硅通孔技术研究进展》，华兴证券整理

图表 35：TSV 侧壁涂层工艺研究进展总结

工艺方法	工艺机理	工艺缺陷	研究进展	应用
热氧化	高温氧化作用	温度极高	较薄的 SiO ₂ 沉积厚度	绝缘层
PECVD	等离子体与气体反应	台阶覆盖性差	SiO ₂ 薄膜台阶覆盖率达 16.7%	绝缘层
HDPCVD	高密度混合气体接触硅片	成本高	较薄的 SiO ₂ 沉积厚度	绝缘层
原子层沉积	气-固相反应	要求高纯度基体	超深 TSV 的 TiN 沉积	绝缘层
			Al ₂ O ₃ , TaN, Ru 组合涂层高效阻挡 Cu 扩散	阻隔层 种子层
化学镀	金属离子还原为金属	镀液稳定性差	高深宽比通孔沉积 Ni	阻隔层
			高深宽比结构中 Ni 的覆盖率位 100% CoWB、CoB 薄膜显著阻挡扩散	
溅射	高能粒子轰击靶材	孔口沉积	超薄的 Ti、Cu 沉积厚度	阻隔层
			超深 TSV 的 Cu 沉积	种子层

资料来源：《三维系统级封装(3D-SiP)中的硅通孔技术研究进展》，华兴证券整理

凸点在制备 HBM 中必不可少

TSV 制作完成后，为了实现与上下层芯片或衬底信号连接，需要将 TSV 与 TSV、芯片和衬底相互键合，目前的键合方式是依靠凸点。根据《三维系统级封装(3D-SiP)中的硅通孔技术研究进展》，制造凸点的主要工艺有焊锡凸点电镀和铜柱凸点电镀。通过电镀形成的焊锡凸点在回流焊后无裂缝产生，大小均匀。但目前焊料成分较难调控，并且在回流过程中易发生塌陷问题，从而进一步影响良率。相比于焊锡凸点，电镀铜柱凸点具有良好的散热性和高可靠性，并且不会发生桥接问题，具备高可靠性。

图表 36：TSV 侧壁涂层工艺研究进展总结 TSV 键合工艺研究进展总结

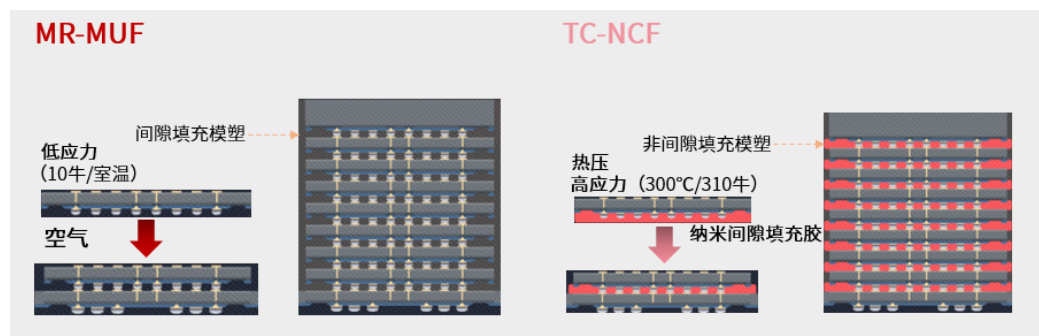
工艺类型	技术难点	研究进展
焊锡凸点	➢ 在回流过程中易发生塌陷和桥接	✓ 在多成分酸性电镀槽中制备 Sn-3.5Ag 焊点，回流焊后无空洞和裂纹
	➢ 焊料成分较难调控	✓ 以非光刻工艺低成本制备 Sn-3.5Ag 焊点
	➢ 镀液稳定性差	✓ 以 Ni-Co 模板制备直径为 100μm 的 SAC305 晶圆级凸点
		✓ 以金属掩膜制备 SAC305 超细间距焊料凸点
铜柱凸点		✓ 通过 PR 模具制备小尺寸细间距凸点，避免桥接
	➢ 金属间化合物影响互连可靠性	✓ 形成直径为 30μm 的凸点，具有细间距、良好散热性能、高机械强度的特点
	➢ 需耐高温 PR 模具	✓ 凸点由 6μm 的铜柱、0.5μm 的镍层、5μm 的锡层组成，呈超细间距高可靠性
		✓ 磁头部件融化焊锡，形成无铅凸点，具有优良的均匀性
混合键合		✓ 通过焊料注头形成 SAC305 凸点，凸点等高
	➢ 需同时兼容 Cu-Cu 与 SiO ₂ -SiO ₂ 两类材料之间的键合	✓ 通过焊料注头制备直径为 20μm 的 SAC305 凸点，焊料填充稳定
	➢ Cu 表面易氧化	✓ 通过 Ar 等离子体表面激活使 Cu 表面改性，强化 Cu-Cu 键合强度
		✓ 以 Ar 和 N ₂ 两步等离子体表面处理实现 Cu-Cu 界面低温键合，键面无 Cu 氧化物
		✓ 通过含 Si 的 Ar 原子束辐射表面处理实现 Cu/SiO ₂ 低温混合键合有高键合强度

资料来源：《三维系统级封装(3D-SiP)中的硅通孔技术研究进展》，华兴证券整理

芯片键合是区分 HBM3E 制备良率的关键环节

目前堆栈键合类型主要包括 SK 海力士采用的 MR-MUF 技术、三星电子和美光采用的 TC-NCF（非导电薄膜热压缩）技术以及未来有希望用于 HBM4 制造的混合键合技术。根据 SK 海力士公开路演 PPT《hynix Tech Seminar（2023）》，MR-MUF 技术采用批量回流焊（MR）融化堆栈芯片之间的凸块，让芯片互相连接。而模塑底部填充（MUF）是在堆栈芯片之间填充保护材料以提升耐久性和散热效果的技术。TC-NCF 则与 MR-MUF 不同，在每次堆栈 DRAM 芯片时，TC-NCF 工艺强调在每层 DRAM 之间放置不导电的粘合膜，用于使芯片彼此绝缘并保护凸点免受撞击。

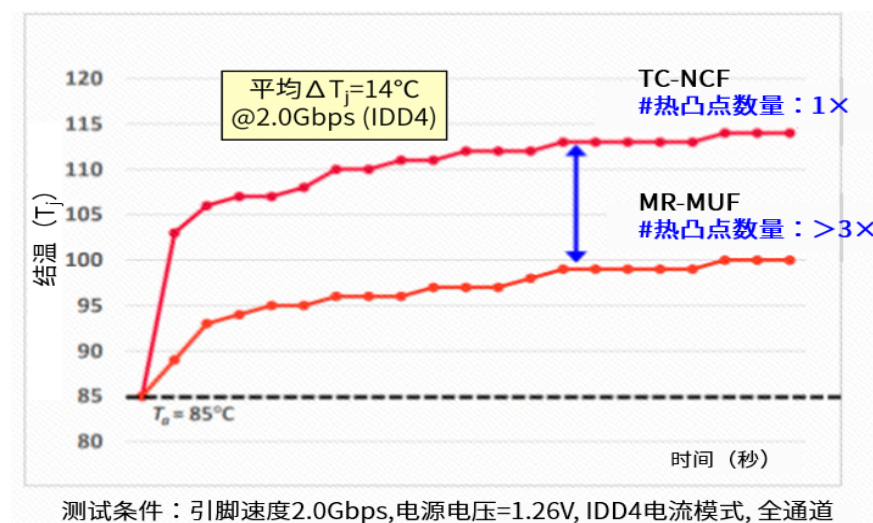
图表 37：SK 海力士 MR-MUF 技术与三星电子 TC-NCF 技术对比



资料来源：SK 海力士公开路演 PPT《hynix Tech Seminar（2023）》，华兴证券整理

根据半导体行业洞察，TC-NCF 相比 MR-MUF 的优势是在堆栈更高层数的 HBM 上更不容易发生翘曲；而 MR-MUF 基于自身结构比 TC-NCF 的散热性能更佳：1) MR-MUF 工艺先通过回流焊连接芯片，再用环氧塑封料填充芯片间隙，而 TC NCF 先用 NCF 非导电薄膜填充芯片间隙，再通过热压键合连接芯片；2) MR-MUF 工艺散热凸点更多，并且使用优良导热性能的环氧塑封料填充芯片与芯片的间隙，完成注塑和底填工艺，具备更好的导热性能。

图表 38：MR-MUF 相比于 TC-NCF 具备更好的散热性



资料来源：半导体行业洞察，华兴证券整理

设备端，TCB 设备供应为关键一环

TC-NCF 和先进 MR-MUF 均有需求

根据半导体行业观察，HBM 制备环节采用 TC-NCF 工艺时，TCB 设备是工艺环节中的核心键合设备，单台价格为人民币 300-700 万元之间，全球核心供应商包括 ASMPT、K&S、Shibaura、Besi 以及 Hanmi 半导体等。分客户看，SK 海力士的核心 TCB 供应商为 Hanmi 半导体、ASMPT、韩国韩华精密机械等。美光的 TCB 供应商包括 Hanmi 半导体及 Shinkawa（新川）等。三星电子的 TCB 供应商主要是其子公司 SEMES 及 Hanmi 半导体。

分供应商看，Hanmi 半导体是 SK 海力士 TCB 设备的核心供应商，并逐步获得美光和三星电子的订单。根据 Hanmi 半导体官网，其在 2024 年 3 月份宣布获得 SK 海力士 215 亿韩元 Dual TC Bonder Griffin 热压键合机订单后随即同月宣布获得美光科技价值 226 亿韩元的订单。

图表 39：TCB 键合下，需求设备及公司汇总

	热压键合机	点胶机	晶圆贴膜机	晶圆刻片机
TC-NCF	√	√	√	√
MR-MUF				
主要厂商	ASM Pacific、Kulicke & Soffa、Besi、Shibaura、Hanmi 半导体、SET	Nordson、NSW Automation、H&S Manufacturing、Graco	Dynatech、Toyo Adtec、Takatori、Teikoku Taping System	DISCO、东京精密、ACCURETECH
	揭膜机	晶圆级真空回流焊键合机	清洗机	塑封设备
TC-NCF	√			
MR-MUF		√	√	√
主要厂商	东洋技术（中国台湾）、至圣集团（中国台湾）、Takatori、Nitto、LINTEC Corporation	Rehm Thermal Systems、Kurtz Ersä、HIRATA、Heller Industries、BTU International	日本 DNS（迪恩士）、东京电子、泛林科技、韩国细美事	TOWA、Besi、ASM Pacific、Yamada、TOWA

资料来源：Besi，华兴证券整理

除逻辑芯片先进封装外，HBM 键合是 TCB 设备的又一大使用场景

2025 年来自 HBM 及逻辑芯片封装的 TCB 设备需求量有望超过 851 台。根据 Besi 官网，Besi 一台 TCB 设备名义小时生产件数（UPH）为 1,000 个裸片左右。但我们认为这是理论值，在实际裸片堆栈键合时，真实 UPH 要远小于理论 UPH。我们在表 40 中估计了 HBM 生产所需的 TCB 设备的需求。考虑到设备需要额外时间进行维护和修理，我们预计单台 TCB 设备的裸片键合能力在每年 420 万片。结合我们对 HBM 裸片的供应量预测，我们认为来自 HBM 键合的 TCB 需求量 2025 年有望达到 340 台。另一方面，TCB 设备需求来自 chip-to-wafer（C2W）和 chip-to-substrate 客户。根据 ASMPT 的 2Q24 业绩会，其 TCB 业务在 2Q24 期间获得了来自 C2W 的 TCB 设备订单。我们看到三星电子、台积电及英特尔的先进封装工艺中存在对 TCB 设备的强烈需求，我们预计来自这一部分逻辑芯片封装的 TCB 设备需求量在 2025 年有望超 510 台。

市占率方面，我们认为 Hanmi 半导体、Besi、ASMPT 和三星电子子公司 SEMES 在 SK 海力士、美光和三星电子 HBM 产线中的 TCB 市占率靠前。根据前文所述，SK 海力士在 HBM3 和 HBM3E 中占据主导地位。

图表 40：HBM 对 TCB 设备需求量估算(台)

	2023	2024E	2025E	2026E	2027E
TCB 设备每小时产量（裸片）	700	700	700	700	700
每日实际工作时间（小时）	20	20	20	20	20
每日加工产量（裸片）	14,000	14,000	14,000	14,000	14,000
一年中维护天数（天）	65	65	65	65	65
年有效工作日（天）	300	300	300	300	300
每台机器年产量（裸片/年）	4,200,000	4,200,000	4,200,000	4,200,000	4,200,000
系数	1.3	1.3	1.3	1.3	1.3
HBM 用 TCB 设备需求量（台）	64	190	340	536	773
YoY		194%	79%	58%	44%
Logic 用 TCB 设备需求量（台）	150	352	511	711	945
合计 TCB 设备需求量（台）	215	542	851	1,247	1,718
YoY		152%	57%	47%	38%

注：每小时生产量=名义吞吐量

资料来源：TrendForce，华兴证券预测

TCB 设备需求有望转向无焊剂 (Fluxless) TCB 方案

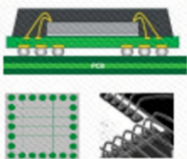
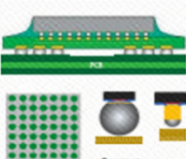
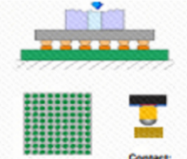
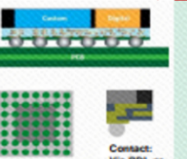
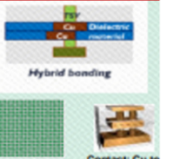
三大原厂均在优化各自现有工艺。根据半导体行业洞察，SK 海力士在面向 12 层 HBM3E 产品时推出了新制程方案 HMR(Heated Mass Reflow)，该工艺注重在稍微加热后再进入回流焊，以解决 MR 时遇到的芯片翘曲问题。根据韩国电子行业新闻媒体 The Elec 报道，在韩国韩华精密机械通过 SK 海力士质量测试后，SK 海力士计划在 2H24 购买 30 多台 12 层 HBM3E 用的 TCB 设备，这是韩华精密机械首次向 SK 海力士供应 TCB 设备，可能打破 Hanmi 半导体的独供市场地位。

Fluxless TCB 设备有望替代传统 TCB 设备。根据 ASMPT 的 2Q24 业绩会，针对逻辑芯片封装客户，公司正在和一家全球性晶圆厂合作开发 Fluxless TCB 设备。针对 HBM 端，公司在 2024 年 7 月为其下一代无焊剂 TCB 赢得了两台设备订单。

HBM4 和 HBM5 时代，混合键合技术有望加速推广

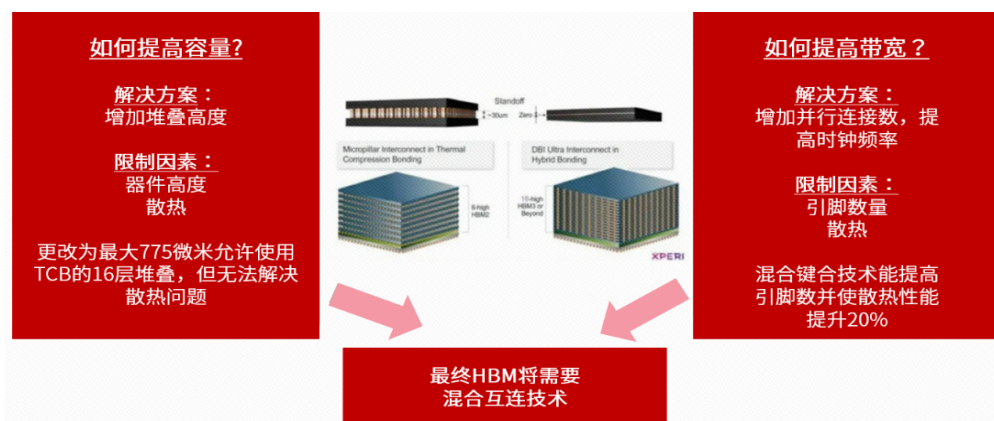
2026-28 年，我们认为混合键合有望提高在 HBM 裸片键合工艺中渗透率。1) 协会规范层面，JEDEC 更新 HBM4 标准后，我们认为采用现有 TCB 方案完成 16 层 HBM 堆栈是可行的。但针对高达 20 层 HBM4E 或者 HBM5 产品，我们认为使用混合键合技术或是大势所趋。2) 技术方面，混合键合技术较 TCB 存在优势。根据 Besi 官网，一方面，其混合键合设备较 TCB 设备具备更高的键合精度，从 TCB 设备的 10-50 μ m 提高到 100-200nm，精度提高至 50-100 倍。另一方面，根据 Besi 官网，热压键合无法从根本上解决 HBM 结构的散热问题。混合键合方案相对热压键合可以更好解决 HBM 结构的散热问题，较热压键合方案的散热效率提高 20%。基于以上几点，我们认为在成本可控前提下，混合键合或有望逐步替代热压键合方案。

图表 41：芯片键合工艺发展历史（引线键合-倒装键合-热压键合-高密度扇出-混合键合）

引线键合 (1975)	倒装芯片 (1995)	TCB键合 (2012)	高密度扇出 (2015)	混合键合 (2018)
				
Density: 5-10mm ²	Density: 25-400mm ²	Density: 156-625mm ²	Density: 500+ mm ²	Density: 10K-1MMmm ²
Contact: wire	Contact: Solder ball/Cu pillar	Contact: Cu pillar	Contact: Via RDL or Cu pillar	Contact: Cu to Cu (no solder)
间距 100-50μm 精度 20-10μm 能耗/比特 10pJ/bit 基板材料 有机/引线框架 产出效率 高 成本 低	200-100μm 10-5μm 0.5pJ/bit 有机/引线框架 高 中	80-40μm 5-1μm 0.1pJ/bit 有机/硅 低 高	100-30μm 5-1μm 0.5pJ/bit 无 中 中	20-1μm 0.5-0.1μm <0.5pJ/bit 无 中 高
环氧裸片键合 引线键合 模型	在裸片上焊凸点/铜柱 倒装芯片放置 回流炉焊接 底填/模型	在裸片上焊铜柱 TCB: 放置, 焊接和键合 无回流炉 底填/模型	在裸片上焊铜柱 放置裸片在扇出载体上 晶圆模型 重布线层	创建铜键合垫 抛光到原子级平整表面 混合键合 (室温) 模型

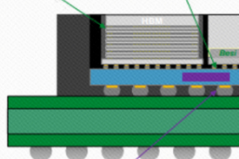
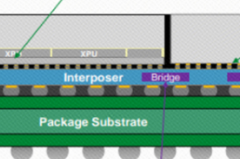
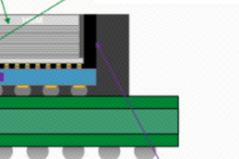
资料来源: Besi 官网, 华兴证券整理

图表 42：面向高容量和带宽 HBM 产品，现有热压键合方案面临散热、功耗、凸点制造等问题，混合键合方案呼之欲出



资料来源: Besi 官网, 华兴证券整理

图表 43：混合键合与热压键合应用场景比较

Chip-to-Wafer (C2W) 热压键合	混合键合	倒装芯片 C2W
HBM 裸片堆叠 ≤16 个内存裸片	3D IC 芯片堆叠 间距≤9 μm	HBM 裸片堆叠 ≥16 个内存裸片
裸片到中介层 10 至 36 μm间距		裸片到中介层 间距>36μm 大规模回流焊
		
中介层接合 (oS)	嵌入式桥接接合	模型
倒装芯片 MR 更薄的中介层 更低的翘曲度	2D 正面桥接 晶圆或面板 精度<1 μm	晶圆级 MUF 薄膜辅助 MUF 用于中介层封装
C2S热压键合 更厚的中介层 更高的翘曲度	带 TSV的3D 桥接 晶圆或面板 C2W热压键合	

资料来源: Besi 官网, 华兴证券整理

伴随混合键合渗透率提升，相关新设备需求有望释放

在 HBM3 和 HBM3E 的制备流程中，三星电子和美光采用 TC-NCF 技术，设备包括：TCB 设备、晶圆贴膜机和划片机等。SK 海力士采用的先进 MR-MUF 设备同样包括 TCB 设备，另外也需要真空回流焊键合机及塑封设备等。根据半导体行业观察和 Besi 官网，与 TCB 工艺不同，混合键合工艺分为三步：1) CMP 抛光，保证晶圆和裸片表面十分光滑；2) 键合过程中需要基于准确位置做介电键成键；3) Cu-Cu 键合，保证芯片键合后电性能稳定。相应的混合键合设备转变为：CVD、刻蚀机、混合键合机、CMP 设备、铜电镀设备、铜互联设备 PVD 等。

2025-26 年逐步采用混合键合后，我们认为设备需求有望发生转变：

- 混合键合机会替换 TCB 和真空回流焊机，带来单位设备价值量增加。根据 Besi 官网，其 8800 Ultra Accurate C2W 混合键合设备可以确保 200 纳米以下的键合精度，远超其 8800 FC 型号。价格方面，当前 Besi 单台 TCB 设备价格在 100-120 万美元，而混合键合设备单价在 250 万美元以上。
- 一些设备不再需要，比如晶圆贴膜机、晶圆划片机、点胶机以及芯片塑封设备。
- 相应新增设备需求包括 CVD、刻蚀机、CMP 设备、铜电镀设备、铜互联设备 PVD。
- TC-NCF、先进 MR-MUF 及混合键合下均需要的设备，例如爱德万和泰瑞达的芯片测试机。

图表 44：混合键合下，新增设备及公司汇总

	晶圆划片机	清洗机	CVD	刻蚀机	铜互联 PVD	铜电镀设备
混合键合-W2W			√	√	√	√
混合键合-D2W	√	√	√	√	√	√
主要厂商	DISCO、东京精密、ACCRETECH	日本 DNS（迪恩士）、东京电子、泛林科技、韩国细美事	AMAT、泛林半导体、东京电子、先晶半导体	泛林半导体、东京电子、应用材料、日立高新	Advanced Energy Industries、Angstrom Engineering、Veeco、AMAT、Platit AG	ACM、AMAT、ASMPT、ClassOne Technology、EBARA
	CMP	混合键合机	退火设备	减薄机	临时键合机/解键合机	测试机
混合键合-W2W	√	√	√	√	√	√
混合键合-D2W	√	√	√			√
主要厂商	AMAT、日本荏原、韩国 KC Tech、华海清科	D2W: Besi、EVG、ASMPT W2W: 华卓清科、拓荆科技	AMAT、Kokusai、东京电子、屹唐、北方华创	DISCO、东京精密、KOYOSEIKO、G & N	TEL、EVG、SUSS、上海微、TAZMO、Tok、ERS、EO Technics、Takatori	爱德万、泰瑞达、精智达、锐芯、长川科技、华峰测控

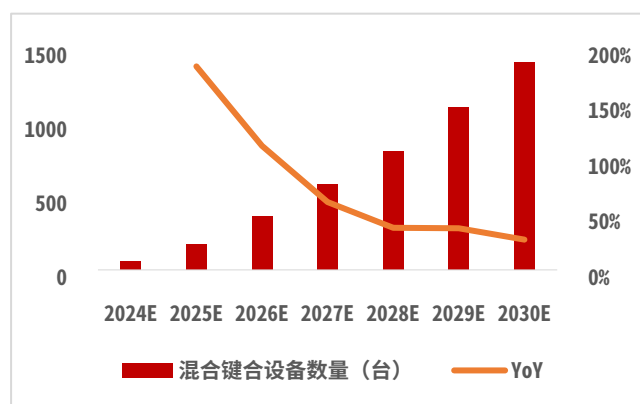
资料来源：Besi 官网，华兴证券整理

混合键合时代，头部键合设备公司有望直接受益

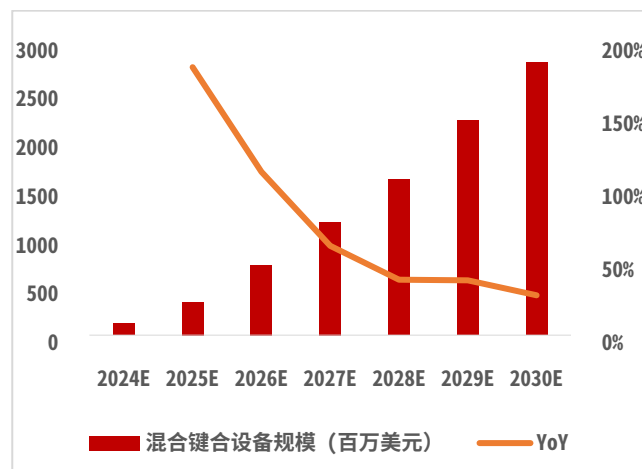
根据 Yole 数据，2020 年全球混合键合机市场价值达到 3.2 亿美元，其中 CoW 键合机为 0.6 亿美元，WoW 键合机为 2.6 亿美元，至 2027 年 CoW 和 WoW 市场空间有望分别达到 2.3 亿和 5.1 亿美元，2020-27 年行业 CAGR 有望分别达到 69%和 16%。

面向 7nm 及以下先进制程键合设备，荷兰 Besi 是主要供应商之一。目前 Besi 公司推出了 8800 Ultra Accurate C2W 混合键合机，较传统热压键合机的键合精度明显提高。另外 ASMPT 也在 2Q24 获得两台混合键合设备订单。

图表 45：混合键合设备出货量（台）



图表 46：混合键合设备市场规模（百万美元）



资料来源：SK 海力士官网 Yole, Besi 官网, 华兴证券预测

资料来源：SK 海力士官网 Yole, Besi 官网, 华兴证券预测

ASMPT、Hanmi 半导体、韩国韩华精密机械、Besi、K&S 及 Shibaura 为主要的混合键合设备供应商

单台混合键合机裸片键合速度或在每年 540 万片。根据 Besi 官网，其典型混合键合机型号（8800 Ultra Accurate C2W）理论 UPH 在 1,500 片左右；而在实际产线中，考虑到和其他设备工作的匹配及额外维护时间，我们估算 Besi 8800 Ultra Accurate C2W 混合键合机的实际 UPH 在 900 片左右。考虑到设备需要定期维护以及工艺参数调试，如图 47 所示，我们估算单台混合键合机的年均裸片键合能力在 540 万片左右。结合 HBM 裸片数量及混合键合工艺渗透率情况，我们预计 2027 年用于 HBM 制备的混合键合设备需求量达到 90 台，另外用于逻辑芯片封装的混合键合设备需求量约为 490 台，两方面加总的需求量约为 580 台。

根据 TrendForce，ASMPT、Hanmi 半导体、Besi、韩国三星电子子公司 SEMES、K&S 及 Shibaura 为主要供应商。ASMPT 在其 2Q24 业绩会上宣布获得混合键合设备订单，为客户测试使用。Besi 不断迭代其混合键合工艺，推出的 8800 Ultra Accurate C2W 混合键合机键合精度达到 200nm，而下一代设备精度有望提升至 100nm。此外，三星电子子公司 SEMES 也正在积极开发混合键合设备。

图表 47：HBM 对混合键合设备需求量估算（台）（逻辑芯片和 HBM 存储）

	2024E	2025E	2026E	2027E
混合键合理论每小时产量（裸片）	1,500	1,500	1,500	1,500
混合键合实际每小时产量（裸片）	900	900	900	900
混合键合实际 每小时生产效率	60%	60%	60%	60%
每日实际工作时间（小时）	20	20	20	20
每日加工产量（裸片）	18,000	18,000	18,000	18,000
维护天数	65	65	65	65
年有效工作日（天）	300	300	300	300
每台机器实际年产量（裸片）	5,400,000	5,400,000	5,400,000	5,400,000
超额比例	1.3	1.3	1.3	1.3
HBM 对应的混合键合设备需求量（台）	1	5	13	90
同比增长%		259%	136%	621%
逻辑芯片对应的混合键合设备需求量（台）	59	165	347	490
同比增长%		181%	111%	41%
混合键合设备需求量（台）	60	170	360	580
同比增长%		183%	112%	61%

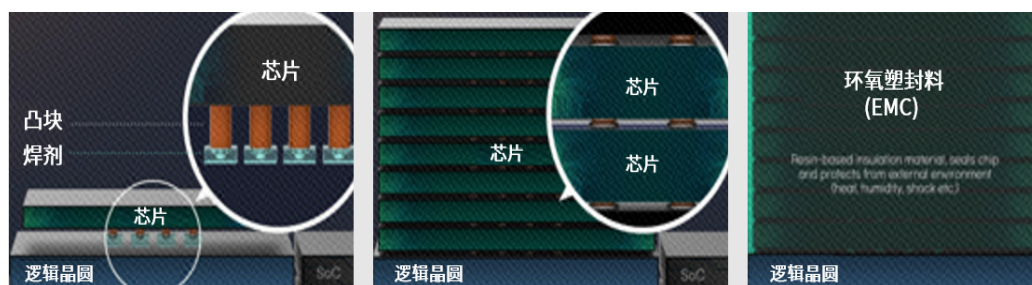
注：每小时生产量=名义吞吐量

资料来源：TrendForce，华兴证券预测

HBM3E 方面，SK 海力士能实现更高热稳定性和良率

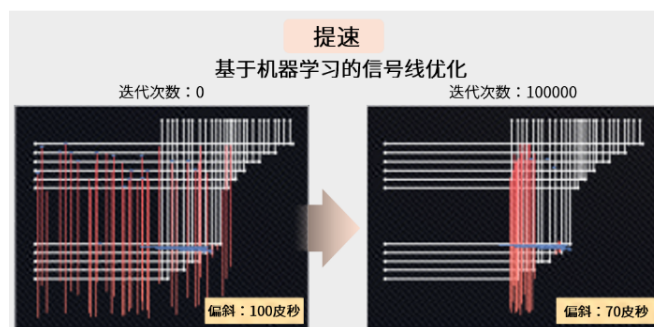
针对 HBM3E，SK 海力士采用先进 MR-MUF 封装工艺，不同于三星电子和美光采用的 TC-NCF 技术。根据 SK 海力士官网，相比 TC-NCF 工艺，SK 海力士采用的 MR-MUF 工艺在试验状态下的导热率比 TC-NCF 高约两倍，可在 HBM 高速运转时满足散热性能要求。根据半导体行业观察，美光在 2024 年 3 月的良率短暂提升到 70%，但 2Q24 大规模量产后良率出现下降。我们认为美光面临的挑战主要是内部预估和实际生产能力不匹配，而其生产能力受限于：1) 量产阶段遇到材料配方和工艺问题；2) 美光来回改用不同供应商设备导致产线的设备精准度出现问题。三星电子当前在 TC-NCF 生产 HBM3E 方面遇到散热和功耗问题，向下游客户批量供货的时间一直在延迟。此外，我们认为 2024 年 7 月三星电子爆发的工会罢工在一定程度上影响了它 HBM3E 产品的出货进度。

图表 48：MR-MUF 一次性融化所有凸点连接芯片与电路，并填充环氧成型化合物进行热保护



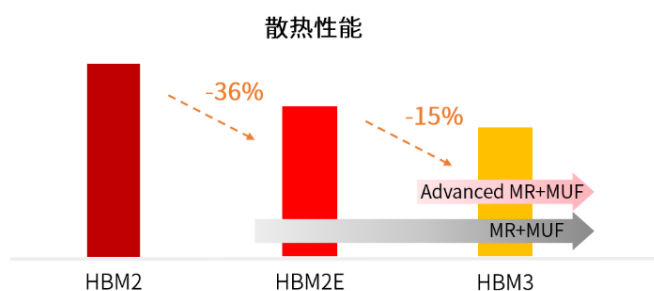
资料来源：SK 海力士，华兴证券整理

图表 49：SK 海力士 HBM 助力芯片加速



资料来源：SK 海力士，华兴证券预测

图表 50：MR-MUF 技术更新推动散热能力进一步提升



资料来源：SK 海力士，华兴证券预测

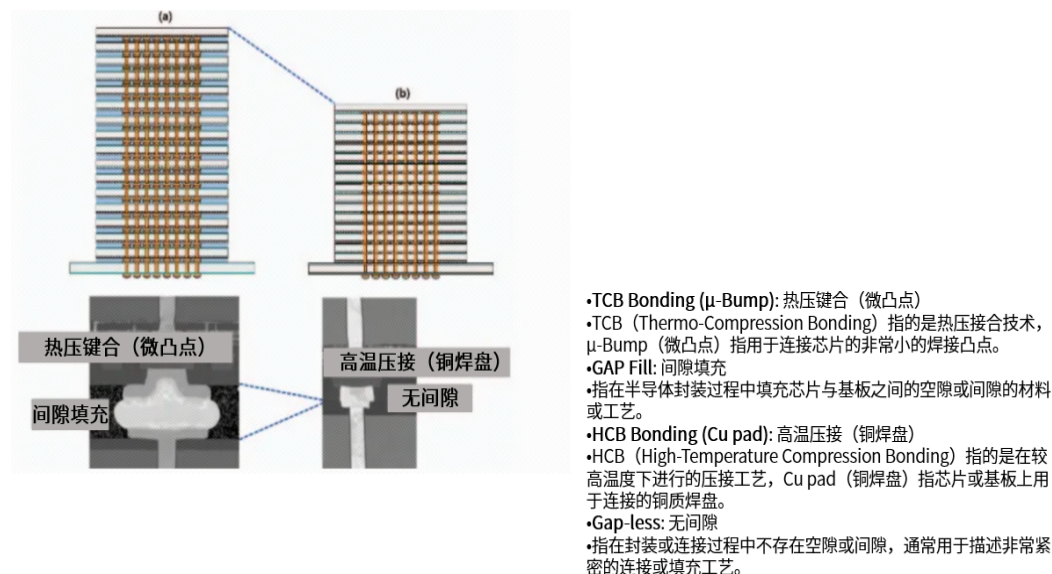
三星电子和美光希望借 HBM4 弯道超车

由于三星电子 TC-NCF 工艺在 HBM3E 良率的爬坡速度不及预期，三星电子或将更多资源投入下一代 HBM4 量产。针对 HBM4，三星电子采用两条腿走路策略，即同时研发 TC-NCF 和混合键合技术。这一做法基于如下考虑：1) 三星电子管理层认为，面向 8 层堆栈 HBM3E 时，MR-MUF 技术工艺生产效率高于 TC-NCF，但一旦 DRAM 堆栈层数达到 12 层以上（12 层或 16 层），TC-NCF 比 MR-MUF 的生产效率更高；2) 三星电子重视混合键合工艺。根据《用于 HBM 堆栈的 D2W（芯片到晶圆）铜键合技术研究》，相较传统的热压键合工艺，混合键合摒弃了凸块步骤，直接通过铜对铜方式连接上下两层的 DRAM 裸片，降低了 DRAM 裸片层间距，提高了信号传输速度，同时提升了功耗和散热。三星电子也认为在 16 层堆栈及以上的 HBM 制造中，混合键合技术是必须的。

与热压键合相比，我们认为混合键合的优势可概括为：1) 省空间，可以在更低的高度键合更多芯片堆栈。2) 强散热，散热效率优于热压键合。3) 数据传输量大，混合键合减少传输距离。

美光也在加快技术追赶，举措包括：1) 根据美光中国官方微信公众号，美光储备下一代 HBM 内存 HBMnext 技术，并预计 HBMNext 将提供 36 GB 和 64 GB 容量，包括 12-Hi 24 Gb 堆栈(36 GB) 或 16-Hi 32Gb 堆栈(64GB)。每个堆栈带宽为 1.5TB/s-2+TB/s，数据传输速率超过 11.5GT/s/pin；2) 美光或通过 HBM-GPU 组合芯片形式获得更快内存访问速度，摒弃将 HBM 和逻辑芯片整合的做法。

图表 51：针对 16 层 DRAM 堆栈键合，混合键合与传统热压键合比较

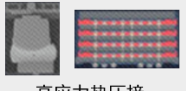
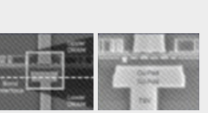


资料来源：《用于 HBM 堆栈的 D2W（芯片到晶圆）铜键合技术研究》，华兴证券整理

SK 海力士 HBM4 优势来自成熟工艺和台积电支持

不同于三星电子，SK 海力士继续采用 MR-MUF 技术实现 HBM 16 层堆栈。与三星电子 IDM 业务模式不同，SK 海力士不参与晶圆代工业务，因而与台积电的合作更加紧密。根据 SK 海力士官网，其在 2024 年 4 月宣布与台积电合作 HBM4 产品，预计于 2026 年投产，具体包括：1) 双方针对 HBM 最底层的基础裸片（Base Die）进行性能改善，为多颗 DRAM 堆栈提供控制；2) 双方将协力优化 SK 海力士的 HBM 产品和台积电的 CoWoS 技术融合，共同应对 HBM 相关客户要求。根据 SK 海力士，其 HBM4 比第五代 HBM3E 速度提升 40%，而耗电量仅为第五代 HBM3E 的 70%。SK 海力士在先进 MR-MUF 技术上经验丰富，相比行业其他厂商，在 DRAM 热隆起控制、间隙高度控制以及热增强材料在全球具备竞争力。

图表 52：SK 海力士 HBM 技术迭代

	HBM2	HBM2E	HBM3	HBM3(12Hi)/HBM3E	HBM4
堆叠技术	TC-NCF	MR-MUF	Advanced MR-MUF		TBD
备注	 高应力热压接  全球首个TSV芯片堆叠	 低应力  空气 低键合力与稳固的接头 更高的凸点部分（热散）	 低力与低热  空气 更高效的散热： 更低的间隙高度与热阻		 Advanced MR-MUF  混合键合
可实现的堆叠高度	4Hi/8Hi	4Hi/8Hi	8Hi/12Hi		12Hi/16Hi
热阻（相对）	(1,0)	(0.65)	(0.55)	(0.5)	(0.4~0.5)

资料来源：SK 海力士官网，华兴证券整理

我们认为当前标准对 SK 海力士维护近期市场地位更有利

最新 HBM 标准或有助于海力士沿用 MR-MUF 技术。国际半导体标准组织（JEDEC）在 2024 年 7 月发布了 HBM4 最新标准，同意将 HBM4 产品标准定为 775 微米，比上一代 720 微米更厚。我们认为最新标准为 SK 海力士使用现有 MR-MUF 技术实现 16 层 DRAM 堆栈奠定了标准基础。SK 海力士在 HBM4 上或仍将采用先进 MR-MUF 工艺进行制备。鉴于 SK 海力士具备丰富的 MR-MUF 改良和制备经验，我们认为其或在新一轮 HBM 竞赛中保持领先地位。

从设备端来看，三星电子提出混合键合技术所需设备与其现有产能设备差异较大。根据半导体行业观察，混合键合工艺涉及许多传统上仅由晶圆代工厂专用的工具，例如 CVD、CMP 和表面离子活化等，DRAM 和 HBM 存储厂商较难在短期内具备大规模生产条件。同时混合键合所需要的设备价格明显高于目前的 TCB 设备，存储厂商在设备配置上或存在一定资金压力。但我们也看到三星电子在混合键合上积极投入研发资源。在美国科罗拉多州举行的 2024 IEEE 第 74 届电子元件和技术会议上，三星电子发表了一篇题为《用于 HBM 堆栈的 D2W（芯片到晶圆）铜键合技术研究》的论文，明确表示对于 16 层以上 DRAM 堆栈的 HBM 产品，混合键合技术为必不可少。三星电子计划在 2025 年制造 HBM4 样品（主要为 16 层堆栈），并于 2026 年进行量产。

综合考虑技术积累、标准修改和设备投入等因素，我们认为 SK 海力士或是三家原厂中最早量产 12 层 HBM4 的厂商。根据 SK 海力士中国官方微信公众号，SK 海力士预计在 2025 年下半年推出采用 12 层 DRAM 堆栈的 HBM4 产品，并于 2026 年发布 16 层 DRAM 堆栈的 HBM4 产品。根据半导体行业观察，美光预计将在 2026 年推出 12 层和 16 层堆栈的 HBM4，带宽超过 1.5TB/s，并且将在 2028 年发布 16 层堆栈的 HBM4E。

图表 53：SK 海力士、三星电子及美光各代 HBM 工艺路径

	美光	三星电子	海力士
HBM 3,3E 2024	非导电膜 热压键合 (NCF TC)	非导电膜 热压键合 (NCF TC)	大规模回流 模具底填充 (MR-MUF)
HBM 4,4E 2026	无助焊剂热压 / 混合键合	无助焊剂热压 / 混合键合	高级大规模回流-模具底填充
	这一代 HBM 将采用不同的技术，混合键合技术将首先应用于 16 层堆叠		
HBM 5+	混合键合 最高的触点数、 堆叠高度和散热性能	混合键合 最高的触点数、 堆叠高度和散热性能	混合键合 最高的触点数、 堆叠高度和散热性能

资料来源：Besi，华兴证券整理

2025-26 年或聚焦 HBM4 市场份额争夺

三大原厂的实际 HBM 产能或与其规划出现一定差异，我们的主要假设如下：

- 1) 根据半导体行业观察，三星电子因为良率和材料问题，其 HBM3E 产品的客户验证进度慢于市场预期。我们认为三星电子 HBM3E 产品或最早于 4Q24 实现大规模量产。三星电子计划于 2025 年更新其 HBM4 技术，并于 2026 年实现量产。
- 2) 我们预计美光的新一代 HBM3E 在 4Q24-1H25 逐渐放量。HBM4 方面，美光在研发和资本支出方面进行了重点投入。我们预计其产品实际出货时间或在 2025 年下半年或 2026 年。
- 3) 根据半导体行业观察，SK 海力士采用先进 MR-MUF 技术研发 HBM3E 产品，并通过了客户验证。在 HBM3E 市场上，SK 海力士 2H24 持续占领高端 HBM 市场。HBM4 方面，SK 海力士或继续采用 MR-MUF 技术研发产品，并有望在 2025-26 年率先出货。但同时，我们也认为美光和三星电子在 HBM4 产品上具备强大的研发能力和产品出货能力，或与 SK 海力士争夺 HBM4 市场份额。
- 4) 产能利用率方面，我们预计三家原厂或有望平衡新增建设产能和现有产能的产能利用率爬坡之间的关系。2024-26 年，我们预计 SK 海力士、美光和三星电子的传统 DRAM 产能利用率在 85-90%区间，HBM 产线稼动率维持在 90%以上。

图表 54：2023-27 年三大原厂 HBM 实际产能估算（晶圆，片）

HBM 实际月产能（当年 12 月）	2023	2024E	2025E	2026E	2027E
三星电子	26,700	44,700	72,200	108,200	150,200
SK 海力士	29,800	68,600	108,200	152,000	197,600
美光	2,800	9,850	25,750	52,450	86,750
合计	59,300	123,150	206,150	312,650	434,550
HBM 实际年产能（年底）	2023	2024E	2025E	2026E	2027E
三星电子	156,800	437,400	703,700	1,100,400	1,571,400
SK 海力士	188,900	611,000	1,080,600	1,580,400	2,120,400
美光	19,800	66,500	214,900	468,300	849,400
合计	365,500	1,114,900	1,999,200	3,149,100	4,541,200
YoY	123%	205%	79%	58%	44%
三家原厂实际产能占比	2023	2024E	2025E	2026E	2027E
三星电子	43%	39%	35%	35%	35%
SK 海力士	52%	55%	54%	50%	47%
美光	5%	6%	11%	15%	19%
合计	100%	100%	100%	100%	100%

资料来源：TrendForce，华兴证券预测

我们预计 2024-26 年三大 HBM 厂商的出货量合计分别是 10.1 亿/20.1 亿/34.1 亿 GB，主要假设是：

- 1) HBM3E 方面，SK 海力士良率高于美光和三星电子。SK 海力士良率在 2024 年 5 月已接近 80%，美光 HBM3E 良率低于 SK 海力士。根据半导体行业观察，三星电子目前仍受限于材料和设备问题，其 HBM3E 产品在 2024 年 7 月底前仍未通过英伟达测试。2025-26 年，随着美光和三星电子逐步改良设备和材料，我们预计两家公司将逐步提升 HBM3E 良率水平，从而拉动整体 HBM 产品组合的良率提升。
- 2) HBM 堆栈层数方面，我们预计 HBM 层数有望从 2023 年的 8 层增长到 2027 年的 10 层。

图表 55：2023-27 年三大原厂 HBM 实际供给量估算

	2023	2024E	2025E	2026E	2027E
三家原厂合计月产能（晶圆，片）	59,300	123,150	206,150	312,650	434,550
三家原厂实际年产能（晶圆，片）	365,500	1,114,900	1,999,200	3,149,100	4,541,200
平均稼动率（%）	87%	87%	88%	89%	89%
三家原厂合计产量（晶圆，片）	319,658	967,990	1,768,530	2,793,216	4,027,684
12 寸晶圆上切割数量（片）	550.0	550.0	550.0	550.0	550.0
出货量/产量比率	0.98	0.98	0.98	0.98	0.98
合格裸片出货量（百万颗）	119.2	387.8	719.1	1,136.2	1,637.7
HBM 堆栈层数	8.0	8.5	9.0	9.5	10.0
HBM 供给量-包括 8 层和 12 层堆栈（百万）	14.9	45.6	79.9	119.6	163.8
HBM 单颗存储容量（GB）	2.5	2.6	2.8	3.0	3.0
三家原厂合计 HBM 供给量（百万 GB）	298	1,008	2,013	3,409	4,913

资料来源：TrendForce，华兴证券预测

附录

【分析师声明】

负责撰写此报告的分析师（一人或多人）就本研究报告确认，本人具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度，独立、客观地出具本报告。本报告清晰地反映了本人的研究观点。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

【法律声明】

一般声明

本报告由华兴证券有限公司（以下简称“本公司”）编制。本公司具有由中国证券监督管理委员会许可的证券投资咨询业务资格。

本公司的特定客户及其他专业人士使用。本公司不会因接收人收到本报告而视其为客户。本公司并不对其他网站和各类媒体转载、摘编的本公司报告负责。

本公司研究报告的信息均来源于公开资料，但本公司对该等信息的准确性和完整性不作任何保证。我们力求报告内容的客观、公正，但该等信息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，仅供投资者参考之用，在任何情况下，报告中的信息或所表述的意见并不构成对任何人的投资建议，本报告的接收人须保持自身的独立判断。在任何情况下，本公司及其雇员不对任何人因使用报告中的任何内容所引致的任何损失负任何责任。

本报告所在的意见、评估及预测仅为本报告出具日的分析师观点和判断，可在不发出通知的情况下作出更改，在不同时期，本公司可发出与该报告所载资料、意见及推测不一致的报告。本公司的销售人员、交易人员以及其他专业人士可能会依据不同的假设和标准、采取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，本公司没有将此意见及建议向报告所有接受者进行更新的义务。本公司的资产管理部门、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。

特别声明

在法律许可的情况下，华兴证券有限公司可能会持有或交易报告中提及公司所发行的证券或投资标的，也可能为这些公司提供或争取建立业务关系或服务关系（包括但不限于提供投资银行业务或财务顾问服务等）。因此，投资者应当考虑到本公司可能存在影响本报告观点客观性的利益冲突。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。

本报告的版权归本公司所有。本公司对本报告保留一切权利。未经本公司书面许可，任何机构和个人均不得以任何形式翻版、复制或转载，否则，本公司将保留追究其法律责任的权利。

【评级说明】

公司评级体系以报告发布日后 6-12 个月的公司股价涨跌幅相对同期公司所在证券市场大盘指数（A 股市场为沪深 300 指数、H 股市场为恒生指数、台股市场为台湾加权指数、美股市场为纳斯达克指数、欧洲市场为 MSCI 欧洲指数）涨跌幅为基准：分析师估测“买入”公司股票相对大盘涨幅在 10% 以上；“持有”公司股票相对大盘涨幅介于 -10% 到 10% 之间；“卖出”公司股票相对大盘涨幅低于 -10%。

行业评级体系以报告发布日后 6-12 个月的行业指数涨跌幅相对同期相关证券市场（A 股市场为沪深 300 指数、H 股市场为恒生指数、台股市场为台湾加权指数、美股市场为纳斯达克指数、欧洲市场为 MSCI 欧洲指数）大盘指数涨跌幅为基准：分析师估测“超配”行业相对大盘涨幅在 10% 以上；“中性”行业相对大盘涨幅介于 -10% 到 10% 之间；“低配”行业相对大盘涨幅低于 -10%。