

评级: 增持 ( 维持 )

分析师: 王芳

执业证书编号: S0740521120002

Email: wangfang02@zts.com.cn

分析师: 杨旭

执业证书编号: S0740521120001

Email: yangxu01@zts.com.cn

分析师: 游凡

执业证书编号: S0740522120002

Email: youfan@zts.com.cn

重点公司基本状况

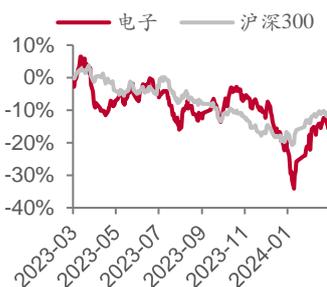
简称	股价(元)	EPS				PE				PEG(24E)	评级
		2022	2023E	2024E	2025E	2022	2023E	2024E	2025E		
通富微电	22.84	0.63	0.33	0.10	0.60	36	69	223	38	-20	买入
赛腾股份	75.90	0.90	1.53	3.37	3.81	85	50	23	20	0	未评级
香农芯创	39.60	0.49	0.69	0.79	0.98	81	58	50	41	2	买入
精智达	70.95	0.72	0.70	1.24	1.71	98	101	57	42	1	未评级
新益昌	71.38	2.27	2.00	0.59	2.58	31	36	121	28	6	未评级
华海诚科	84.12	0.59	0.51	0.41	0.61	143	165	206	139	143	买入
雅克科技	55.26	0.70	1.10	1.40	2.08	79	50	40	27	1	买入
兴森科技	13.01	0.37	0.31	0.15	0.26	35	42	85	49	-5	买入
深南电路	87.00	2.89	3.20	2.73	3.90	30	27	32	22	2	买入
联瑞新材	43.43	0.93	1.01	0.94	1.36	47	43	46	32	2	未评级
长电科技	28.25	1.65	1.81	0.82	1.37	17	16	34	21	-3	买入

备注: 以 2024 年 3 月 25 日收盘价计算, 未评级股票采用 WIND 一致预期

基本状况

上市公司数 480  
行业总市值(百万元) 6,290,396  
行业流通市值(百万元) 3,127,038

行业-市场走势对比



相关报告

《AI 系列之先进封装: 后摩尔时代利器, AI+国产化紧缺赛道》

《香农芯创深度: 海力士云服务存储最大本土代理商, 24 年周期+成长共振》

《AI 系列: 光是通信的必由之路, OCS 已成功应用》

《AI 系列之国产算力: 0-1, 重视产业链历史机遇》

投资要点

- **AI 硬件核心是算力和存力, HBM 高带宽、低功耗优势显著, 是算力性能发挥的关键。** AI 芯片需要处理大量并行数据, 要求高算力和大带宽, 算力越强、每秒处理数据的速度越快, 而带宽越大、每秒可访问的数据越多, 算力强弱主要由 AI 芯片决定, 带宽由存储器决定, 存力是限制 AI 芯片性能的瓶颈之一。AI 芯片需要高带宽、低功耗, 同时在不占用面积的情况下可以扩展容量的存储器。HBM 是 GDDR 的一种, 定位在处理器片上缓存和传统 DRAM 之间, 兼顾带宽和容量, 较其他存储器有高带宽、低功耗、面积小的三大特点, 契合 AI 芯片需求。HBM 不断迭代, 从 HBM1 目前最新到 HBM3E, 迭代方向是提高容量和带宽, 容量可以通过堆叠层数或增加单层容量获得提升, 带宽提升主要是通过提升 I/O 速度。
- **HBM 市场爆发式增长, 海力士和三星垄断市场。** 目前主流 AI 训练芯片均使用 HBM, 一颗 GPU 配多颗 HBM, 如英伟达 1 颗 H100 使用 5 颗 HBM3、容量 80GB, 23 年底发布的 H200 使用 6 颗 HBM3E (全球首颗使用 HBM3E 的 GPU)、容量达 144GB, 3 月 18 日, 英伟达在美国加州圣何塞召开了 GTC2024 大会发布的 B100 和 B200 使用 192GB (8 个 24GB 8 层 HBM3E), 英伟达 GPU HBM 用量提升, 另外 AMD 的 MI300 系列、谷歌的 TPU 系列均使用 HBM。根据我们的测算, 预计 24 年 HBM 市场需求达 150 亿美金, 较 23 年翻倍。HBM 的供应由三星、海力士和美光三大原厂垄断, 22 年海力士/三星/美光份额 50%/40%/10%, 海力士是 HBM 先驱, HBM3 全球领先, 与英伟达强绑定、是英伟达主要 HBM 供应商, 三星紧随其后, 美光因技术路线判断失误份额较低, 目前追赶中, HBM3E 进度直逼海力士。目前 HBM 供不应求, 三大原厂已开启军备竞赛, 三大原厂一方面扩产满足市场需求、抢占份额, 海力士和三星 24 年 HBM 产能均提升 2 倍+, 另外三大原厂加速推进下一代产品 HBM3E 量产以获先发优势, 海力士 3 月宣布已开始量产 8 层 HBM3E, 3 月底开始发货, 美光跳过 HBM3 直接做 HBM3E, 2 月底宣布量产 8 层 HBM3E, 三星 2 月底发布 12 层 HBM3E。
- **先进封装大放异彩, 设备和材料新增量。** HBM 采用 3D 堆叠结构, 多片 HBM DRAM Die 堆叠在 Logic Die 上, Die 之间通过 TSV 和凸点互连, 先进封装技术 TSV、凸点制造、堆叠键合是 HBM 制备的关键, 存储原厂采用不同的堆叠键合方式, 海力士采用 MR-MUF 工艺, 三星和美光采用 TCB 工艺, MR-MUF 工艺较 TCB 工艺效率更高、散热效果更好。HBM 对先进封装材料的需求带动主要体现在 TSV、凸点制造和堆叠键合/底填工艺上, 带来对环氧塑封料、硅微粉、电镀液和前驱体用量等的提升, 在设备端 HBM 带来热压键合机、大规模回流焊机和混合键合机等需求。
- **投资建议:** HBM 海外引领, 核心标的如下: 1) 存储原厂: 海力士/三星/美光; 2) 设备: BESI/ASMPT/Camtek 等。大陆 HBM 产业链相关标的: 1) 存储: 香农芯创/佰维存储/雅创电子等; 2) 设备: 赛腾股份/精智达/新益昌等; 3) 材料: 华海诚科/雅克科技/联瑞新材/兴森科技/深南电路等; 4) 封测: 通富微电/深科技/长电科技等。
- **风险提示:** 行业需求不及预期的风险、大陆厂商技术进步不及预期、技术路线发生分歧、研报使用的信息更新不及时的风险, 计算结果存在与实际情况偏差的风险。

## 内容目录

<b>1.HBM 助力 AI 芯片性能发挥，市场需求爆发式增长</b> .....	<b>5 -</b>
1.1 HBM 高带宽、低功耗优势显著，缓解内存墙问题 .....	5 -
1.2 HBM 随 AI 爆发式增长，2024 年达到百亿美金规模.....	11 -
1.3 三大原厂垄断市场，开启军备竞赛 .....	13 -
<b>2.HBM 制造中先进封装大放异彩</b> .....	<b>18 -</b>
2.1 采用多种先进封装工艺，重点关注堆叠键合方式 .....	18 -
2.2 HBM 三大关键工艺：TSV、Micro bump 和堆叠键合 .....	21 -
<b>3. HBM 驱动先进封装设备和材料需求爆发</b> .....	<b>31 -</b>
3.1 材料端：环氧塑封料、硅微粉、电镀液和前驱体等用量提升.....	31 -
3.2 设备端：热压键合机、大规模回流焊机、混合键合机等需求.....	33 -
3.3 全球/大陆产业链布局情况 .....	38 -
<b>4.投资建议</b> .....	<b>40 -</b>
<b>5.风险提示</b> .....	<b>40 -</b>

## 图表目录

<b>图表 1: 英伟达 GPU 浮点运算和带宽提升速度失衡</b> .....	<b>5 -</b>
<b>图表 2: 处理器和存储器的速度失配</b> .....	<b>5 -</b>
<b>图表 3: 存储的数据的传输速度慢</b> .....	<b>6 -</b>
<b>图表 4: 数据的传输功耗大</b> .....	<b>6 -</b>
<b>图表 5: 存算一体类别</b> .....	<b>7 -</b>
<b>图表 6: HBM 定位在片上存储和普通 DRAM 之间</b> .....	<b>8 -</b>
<b>图表 7: HBM 兼顾带宽和容量</b> .....	<b>8 -</b>
<b>图表 8: HBM 是 DRAM 中的一种</b> .....	<b>9 -</b>
<b>图表 9: 英伟达和 AMD 数据中心 GPU 配套的显存</b> .....	<b>9 -</b>
<b>图表 10: HBM 高带宽</b> .....	<b>10 -</b>
<b>图表 11: HBM 低能耗</b> .....	<b>10 -</b>
<b>图表 12: GDDR 是 2D 结构，HBM 是 3D 结构</b> .....	<b>10 -</b>
<b>图表 13: 256GB/s 带宽对 GDDR6 和 HBM2 的性能要求</b> .....	<b>10 -</b>
<b>图表 14: HBM 的高带宽优势明显</b> .....	<b>10 -</b>
<b>图表 15: HBM 占用的面积小</b> .....	<b>10 -</b>
<b>图表 16: 海力士 HBM 迭代情况</b> .....	<b>11 -</b>
<b>图表 17: HBM 结构</b> .....	<b>12 -</b>
<b>图表 18: H100 封装结构</b> .....	<b>12 -</b>

图表 19: AI 训练芯片的 HBM 使用情况.....	- 12 -
图表 20: HBM 市场需求测算.....	- 13 -
图表 21: Sever DRAM 市场份额.....	- 13 -
图表 22: Mobile DRAM 市场份额.....	- 13 -
图表 23: Consumer Eletronics DRAM (利基 DRAM) 份额.....	- 14 -
图表 24: Graphic DRAM 市场份额.....	- 14 -
图表 25: 三大原厂在不同制程使用 EUV 技术的情况.....	- 14 -
图表 26: DRAM 制程迭代.....	- 15 -
图表 27: HBM 发展历程.....	- 16 -
图表 28: HBM 竞争格局.....	- 16 -
图表 29: HBM 不同世代占比.....	- 16 -
图表 30: HBM 发展历程.....	- 17 -
图表 31: 三大原厂 HBM 开发进度.....	- 17 -
图表 32: 海力士官网的 HBM 产品料号情况.....	- 18 -
图表 33: 三星官网的 HBM 产品料号情况.....	- 18 -
图表 34: 英伟达 P100 芯片供应链情况.....	- 19 -
图表 35: HBM 结构.....	- 19 -
图表 36: HBM 的逻辑芯片的功能区.....	- 19 -
图表 37: 传统 DRAM 与 HBM 制造和封测流程对比.....	- 20 -
图表 38: HBM Stack 制造流程.....	- 21 -
图表 39: 海力士 MR-MUF 工艺.....	- 21 -
图表 40: HBM 核心工艺: TSV、micro bump 和堆叠键合.....	- 21 -
图表 41: HBM (4 层 DRAM+1 层逻辑) 3D 封装成本划分 (99.5%键合良率).....	- 22 -
图表 42: HBM (4 层 DRAM+1 层逻辑) 3D 封装成本划分 (99%键合良率).....	- 22 -
图表 43: 3D TSV 结构.....	- 23 -
图表 44: TSV-Viafirst.....	- 23 -
图表 45: TSV 工艺应用情况.....	- 23 -
图表 46: TSV 工艺流程图.....	- 23 -
图表 47: TSV 工艺设备与材料.....	- 24 -
图表 48: Bump 金属凸点.....	- 24 -
图表 49: 电镀锡球凸点的工艺流程图.....	- 25 -
图表 50: bump 工艺设备与材料.....	- 25 -
图表 51: 不同代际 HBM 的 Bump 间距与互联技术.....	- 26 -
图表 52: 大规模回流焊基本流程.....	- 27 -
图表 53: 热压键合基本流程.....	- 27 -
图表 54: 底填料分类.....	- 28 -

图表 55: TC-NCF 工艺流程图.....	- 28 -
图表 57: TC-NCF 步骤及设备.....	- 28 -
图表 57: MR-MUF 工艺流程图.....	- 29 -
图表 58: MR-MUF 步骤及设备.....	- 29 -
图表 59: MR-MUF 与 TC-NCF 工艺对比.....	- 29 -
图表 60: MR-MUF 具备更高的热导性.....	- 30 -
图表 61: 传统凸点技术与混合键合技术流程对比.....	- 30 -
图表 62: EMC 在 HBM 中的应用.....	- 32 -
图表 63: 硅微粉的应用.....	- 32 -
图表 64: 电镀液在 bump 中的应用.....	- 33 -
图表 65: 前驱体在原子层沉积中的应用.....	- 33 -
图表 66: 三种不同的 HBM 工艺对应三类不同的核心键合设备.....	- 34 -
图表 67: 热压键合机的“热压”过程.....	- 34 -
图表 68: 热压键合机海内外主要厂商简介.....	- 35 -
图表 69: Kurtz Ersa HOTFLOW 型号回流炉示意图.....	- 36 -
图表 70: 回流焊炉海内外主要厂商简介.....	- 36 -
图表 71: 焊球焊接键合与铜铜键合示意图.....	- 37 -
图表 72: 混合键合机海内外厂商简介.....	- 38 -
图表 73: HBM 设备/材料产业链.....	- 39 -
图表 74: 大陆 HBM/先进封装材料产业链公司.....	- 40 -

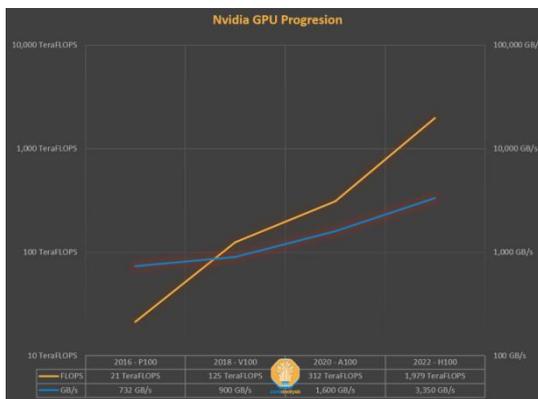
# 1.HBM 助力 AI 芯片性能发挥，市场需求爆发式增长

## 1.1 HBM 高带宽、低功耗优势显著，缓解内存墙问题

- **存储带宽提升速度慢于算力提升速度。**大模型的训练可以简化为 2 步：1) 浮点运算 (FLOPS)，运行矩阵乘法，2) 存储 (带宽)，等待数据或者计算资源。

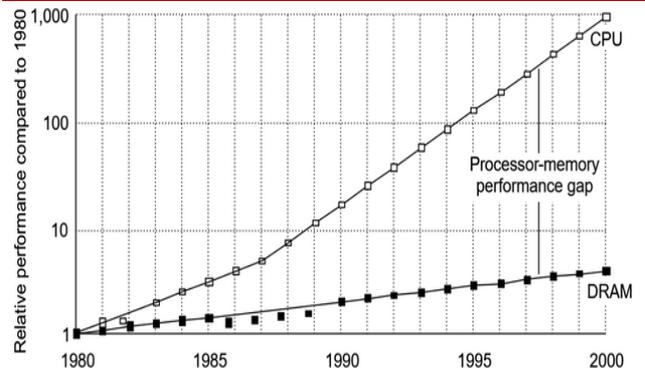
过去大模型的训练时间主要受限于算力，英伟达等 GPU 厂商通过利用摩尔定律缩减制程、改变架构等方式，GPU 的浮点运算能力大幅提高，但对应的存储带宽增速慢于 GPU 算力增长速度。如英伟达 H100 相比 A100，FP32 算力从 19.5 到 66.9tflops 提升 3 倍+，但带宽从 2039GB/s 到 3.35TB/s 仅提升 1.5 倍。

图表 1：英伟达 GPU 浮点运算和带宽提升速度失衡



来源：Semianalysis，中泰证券研究所

图表 2：处理器和存储器的速度失配



来源：知存科技，中泰证券研究所

- **AI 芯片需要处理大量并行数据，要求高算力和大带宽，算力越强、每秒处理数据的速度越快，而带宽越大、每秒可访问的数据越多，算力强弱主要由 AI 芯片决定，带宽由存储器决定，存力是限制 AI 芯片性能的瓶颈之一。AI 芯片需要高带宽、低能耗，同时在不占用面积的情况下可以扩展容量的存储器。**

上世纪 40 年代开始计算机使用冯诺伊曼架构——存算分离，即处理器和存储器相互独立，两者通过总线连接。

1) 存算分离,数据存算间传输造成延迟。处理器从外部存储中调取数据，计算完成后再传输到内存中，一来一回都会造成延迟。

2) 数据在多级存储间传输。为了提升速度，冯诺依曼架构对存储进行分级，越往外的存储介质密度越大、速度越慢，越往内的存储密度越小，速度越快，因此数据需要在多级存储之间搬运，能耗大。通常第一级存储是速度最快、容量低的以 SRAM 为形式的片上缓存，第二级是传统 DDR。

3) 存储制程推进慢于逻辑。目前 DRAM 制程最先进仍在 10-15nm 左右，而逻辑制程已进入 5nm 以下，主要是因存储器制程缩小难度更大。以上三点带来“存储墙”和功耗墙问题，影响处理器性能：1) 数据的传输量小、速度慢，造成延迟，2) 数据的传输功耗大（读取功耗随着存储器密度增大而增大），因此存储器的性能发展方向为更大带宽、更低功耗。

图表 3: 存储的数据传输速度慢



来源: 知存科技, 中泰证券研究所

图表 4: 数据的传输功耗大

Operation	Energy(pJ)
<b>Computation Energy Cost</b>	
Integer Add (32b)	0.1
Integer Multiply (32b)	3.1
Floating Point Add (32b)	0.9
Floating Point Multiply (32b)	3.7
<b>Memory Access Energy Cost</b>	
8KB SRAM (64b)	10
1MB SRAM (64b)	100
DRAM	2000

~650X

来源: 知存科技, 中泰证券研究所

■ **存算一体可有效克服冯诺依曼架构, 迎合 AI 时代提升算力能效的需求。**存算一体是一种新的架构, 其核心理念是将计算和存储融合, 降低“存储墙”问题, 实现计算能效的数量级提升。从广义而言, 存算一体可分为三种: 近存计算 (PNM)、存内处理 (PIM)、存内计算 (CIM), 狭义的存算一体主要指存内计算。

1) **近存计算 PNM**, 通过芯片封装和板卡组装的方式, 将存储和计算单元集成, 提升计算能效。PNM 可分为存储上移以及计算下移。存储上移是指采用先进封装技术将存储器向处理器靠近, 增加计算和存储间的链路数量, 增加带宽, 典型产品为 HBM (高带宽内存), HBM 在英伟达 H200、H100 等 AI 训练 GPU 均有应用。计算下移是指利用板卡集成技术, 在存储设备引入计算引擎, 减少远端处理器负担。

2) **存内处理 PIM**, 核心是存储器具备一定计算能力。芯片制造过程中, 将存储和计算单元集成在同一颗 die 上, 使得存储器本身具备一定计算能力。与近存计算相比, “存”与“算”之间的距离更为紧密。目前的存内处理方案主要通过通过在内存 (DRAM) 芯片中实现部分数据处理, 其中典型的产品形态包括 HBM-PIM 和 PIM-DIMM, HBM-PIM 与 HBM 的不同之处在于 PIM 芯片上的每个存储块内都包含一个内部处理单元。存内处理技术可应用于多种场景, 包括语音识别、数据库索引搜索、基因匹配等。

3) **存内计算 CIM**, 真正意义上实现了同一个晶体管同时具备存储和计算能力。即在芯片设计过程中, 存储电路同时具备存储和计算能力, 使得计算效能实现数量级提升。典型产品为存内计算 (IMC, In-memory Computing) 芯片。典型场景是为 AI 算法提供向量矩阵乘的算子加速, 目前已在神经网络领域开展大量研究, 如卷积神经网络 (Convolutional Neural Network, CNN) 等。

图表 5：存算一体类别

类型	英文全称	英文简称	方法与原理	典型产品	示意图
近存计算	processing near memory	PNM	<p>通过芯片封装和板卡组装，将存储单元和计算单元集成，增加访存带宽、减少数据“搬运”，提升计算效率。</p> <p>又可细分为：</p> <p>→ 存储上移，用先进封装，使得存储器向处理器（CPU、GPU）靠近，增加计算和存储间的链路数量、增加带宽。</p> <p>→ 计算下移，利用板卡集成技术，在存储设备引入计算引擎，承担如数据压缩、搜索、视频转码等本地处理，减少远端处理器负担。</p>	<p>存储上移：HBM（高带宽内存）；</p> <p>计算下移：CSD（可计算存储）；</p>	<p>存储上移</p> <p>计算下移</p>
存内处理	processing in memory	PIM	<p>在芯片制造时，将存和算集成在同一颗晶粒（die）上，存储器本身即具备一定计算能力。比如在DRAM中内置处理单元，提供大吞吐低延迟片上处理能力。</p>	HBM-PIM、PIM-DIMM	<p>内存颗粒 (DRAM Die)</p>
存内计算	computing in memory	CIM	<p>芯片设计过程中，不再区分存储和计算单元，存储电路重新设计后同时具备存储和计算能力，达到计算能效数量级提升</p>	存内计算（IMC）芯片	<p>在存储原位置上实现计算，是真正的存算一体</p>

来源：中国移动研究院《存算一体白皮书（2022）》，中泰证券研究所

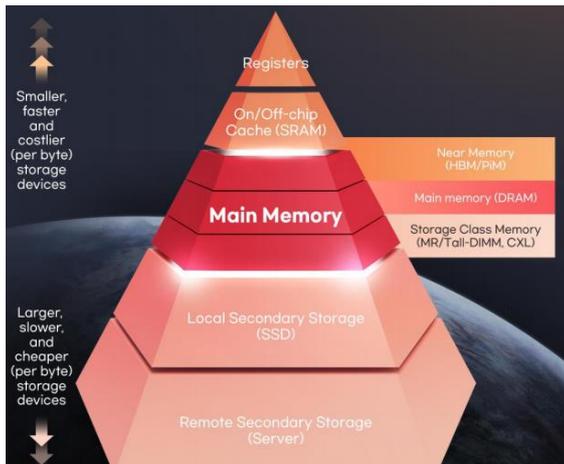
■ **HBM 是近存算一体的一种，定位在片上缓存 LLC 和传统 DDR 中间，弥补带宽缺口，兼顾带宽和容量。**

HBM 定位在 CPU/GPU 片上缓存（Last Level Cache, LLC，通常是 SRAM）和 DRAM 之间，弥补处理器高带宽需求与主存储器最大带宽供应能力之间的带宽缺口，容量大于片上存储、小于传统 DDR，但速度小于片上存储、大于传统 DDR，成本低于片上存储、高于传统 DDR。

以成本为例，1MB SRAM 价值\$5~\$10，1GB HBM 价格\$10-\$20，根据集邦咨询，2 月 1GB DDR4 合约价 \$1.95（历史最高\$4.1），1GB=1024MB，从单位存储成本看，SRAM 的成本是 HBM 的 500 倍+、普通 DRAM 的 1000 倍+，HBM 常备是普通 DRAM 的 5 倍+。

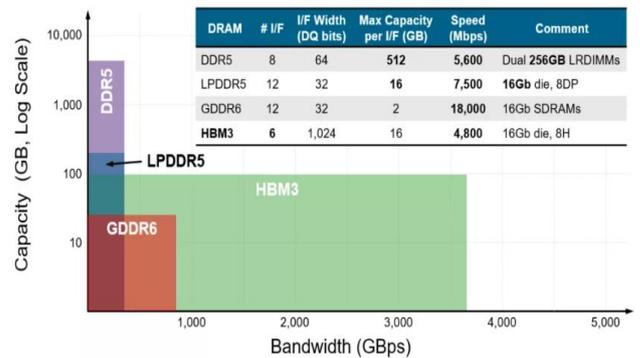
从速度来看，在 AI 应用中，每个 SoC 的带宽需求（尤其是在训练应用中）都会超过几 TB/s，但常规主存储器无法满足这个要求，具有 3200Mbps DDR4 DIMM 的单个主存储器通道只能提供 25.6GB/s 的带宽，具有 4800Mbps DDR5 DIMM 的单个主存储器通道提供 38.4GB/s，即使是具有 8 个存储器通道的最先进的 CPU 平台，DDR4 和 DDR5 对应速度也只能达到 204.8GB/s、307GB/s，围绕单个 SoC 的 4 个 HBM2 堆叠可提供大于 1TB/s 的带宽，因而能够消除带宽差距。

图表 6: HBM 定位在片上存储和普通 DRAM 之间



来源：海力士，中泰证券研究所

图表 7: HBM 兼顾带宽和容量



来源：Synopsys，中泰证券研究所

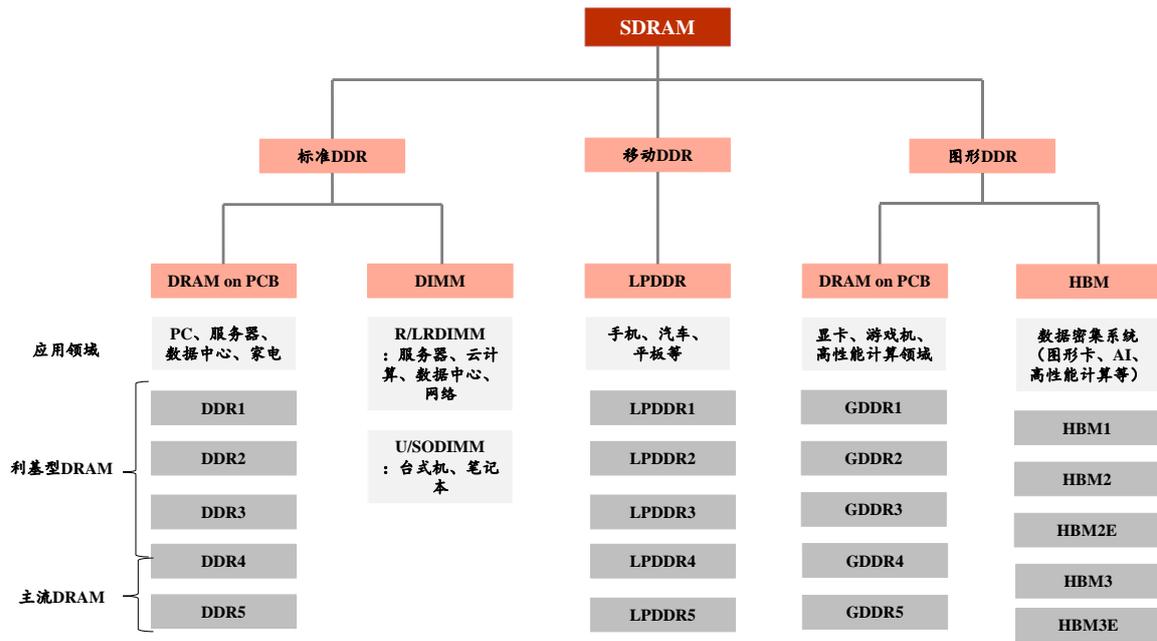
- **HBM 用于 AI 训练卡，GDDR 用于 AI 推理卡。**JEDEC（固态技术协会，微电子产业的领导标准机构）定义并开发了以下三类 DRAM 标准，以帮助设计人员满足其目标应用的功率、性能和尺寸要求。

**1) 标准型 DDR:** Double Data Rate SDRAM，针对服务器、云计算、网络、笔记本电脑、台式机和消费类应用程序，允许更宽的通道宽度、更高的密度和不同的外形尺寸。

**2) LPDDR:** Low Power Double Data Rate SDRAM，LPDDR 具有体积小、功耗低的优势，可以减小无线电子产品的体积，低功耗可延长使用时间，在智能手机、平板等应用中广泛使用。

**3) GDDR:** Graphics Double Data Rate SDRAM，适用于具有高带宽需求的计算领域，通常称为显存，例如图形相关应用程序、数据中心和 AI 等，与 GPU 配套使用，GDDR 又可分为普通 GDDR 和 HBM，普通 GDDR 采用平面结构，HBM 采用 3D 堆叠结构，普通 GDDR 多用在 AI 推理卡中，HBM 用于 AI 训练卡或少数高端的 AI 推理卡。因此，推理芯片如英伟达 A100、H100 通常使用 HBM 配套，而 L4、A16 等均配套 GDDR，高端推理卡可根据要求配套 HBM 或 GDDR，如 A30、A40。

图表 8: HBM 是 DRAM 中的一种



注: 根据DRAMexchange数据, 目前DDR4 4Gb\DDR4 8Gb 512M\*16 属于利基型DRAM

来源: Semiconductor Engineering, 中泰证券研究所

图表 9: 英伟达和 AMD 数据中心 GPU 配套的显存

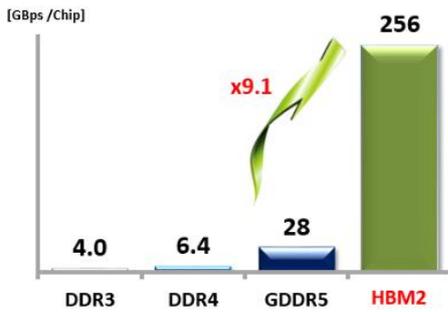
	定位 (训练or推理)	外形规格 (Form Factor)	峰值算力			所用内存	内存容量	内存位宽	峰值内存带宽
			FP8	FP16	FP32				
H200		SXM	3958	1979	67	HBM3E	141GB		4.8TB/s
H100	训练	PCIe(2-slot air cooled)	3026	1513	51.2	HBM2E	80GB	5120bits	2TB/s
		SXM	3958	1978.9	66.9	HBM3	80GB	5120bits	3.35TB/s
		NVL(2 x PCIe: 2-slot air cooled)	7916	3958	134	HBM3	188GB	6144bits	7.8TB/s
L4	推理	PCIe(1-slot, low-profile)	485	242	30.3	GDDR6	24GB	192bits	300GB/s
L40S	推理	4.4" (H) x 10.5" (L) dual slot	1466	733	91.6	GDDR6	48GB	384bits	864GB/s
L40	推理	4.4" (H) x 10.5" (L) dual slot	724	362.1	90.5	GDDR6	48GB	384bits	865GB/s
A100	训练	PCIe(2-slot air cooled/1-slot liquid cooled)		312	19.5	HBM2	80GB	5120bits	1935GB/s
		SXM		624					2039GB/s
A2	推理	PCIe(1-slot, low-profile)		36	4.5	GDDR6	16GB	128bits	200GB/s
A10	推理	FHFL(1-slot)		250	31.2	GDDR6	24GB	384bits	600GB/s
A16	推理	FHFL(2-slot)		71.6	18	GDDR6	64GB	128bits	800GB/s
A30	推理	FHFL(2-slot)		330	10.3	HBM2	24GB	3072bits	933GB/s
A40	-	4.4" (H) x 10.5" (L) dual slot		299.4	37.4	GDDR6	48GB	384bits	696GB/s
M50	推理	PCIe		26.5	13.3	HBM2	16GB	4096bits	1024GB/s
							32GB		
M60	推理	PCIe		29.49	14.7	HBM2	32GB	4096bits	1024GB/s
M100	训练/推理	PCIe		184.6	23.1	HBM2	32GB	4096bits	1.2TB/s
M1250	训练/推理	OAM Module		362.1	45.3	HBM2E	128GB	8192bits	3.2TB/s
M1250X		OAM Module		383	47.87	HBM2E	128GB	8192bits	3.2TB/s
M1210		PCIe		181	22.6	HBM2E	64GB	4096bits	1.6TB/s
M1300A	训练					HBM3	128GB		3.2TB/s
M1300X	训练					HBM3	192GB		5.2TB/s

来源: 英伟达、AMD 官网, 中泰证券研究所

- HBM 相较其他 DRAM 有高带宽、低功耗、面积小的三大特点。一方面通过增加引脚数量达到每颗 HBM 1024bit 内存总线 (即 I/O), 以实现更高带宽, 另一方面缩短数据的传输路径, 降低功耗, HBM 走线长度以毫米为单位, 而 GDDR 和 DDR 均以厘米为单位。
- GDDR VS HBM: 1) HBM 容量易扩展、占用面积小: HBM 采用 TSV

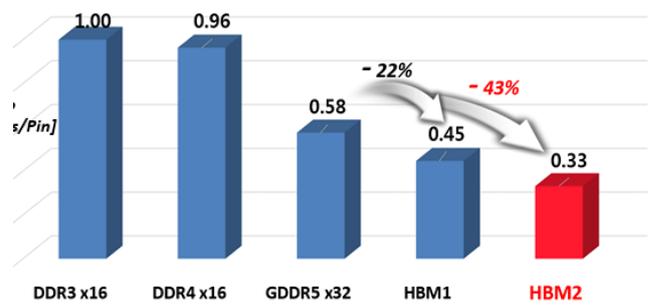
硅通孔技术 3D 堆叠，通过增加层数来扩展容量，GDDR 为 2D 结构，因此 HBM 在实现相同容量下，占用的面积更小。2) HBM 功耗低: GDDR 采用正常 2D 结构，不需要中介层连接，总线位宽小，主要是通过体现数据的传输速率来提升位宽，因为数据的传输速率快，因此功耗也高于 HBM，而 HBM 用 TSV 技术实现走线更短，同时 I/O 数据的传输速度慢，功耗低。3) HBM 带宽大: HBM 通过增加位宽实现高带宽。4) HBM 制造成本高: HBM 采用额外的 TSV、Bumping 等技术，对封装要求高，损耗也更高。

图表 10: HBM 高带宽



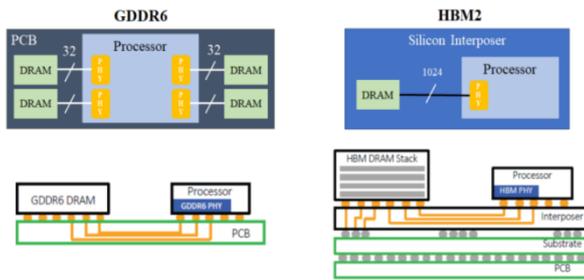
来源: 海力士, 中泰证券研究所

图表 11: HBM 低能耗



来源: 海力士, 中泰证券研究所

图表 12: GDDR 是 2D 结构, HBM 是 3D 结构



来源: Rambus, 中泰证券研究所

图表 13: 256GB/s 带宽对 GDDR6 和 HBM2 的性能要求

	GDDR6	HBM2	
总带宽 (GB/s)	256	256	在带宽相同的情况下
位宽 (bit)	128	1024	HBM2 位宽更大
数据的传输速率 (Gbps)	16	2	HBM2 速度更低
PHY 芯片面积 (相对)	1.5-1.75	1	HBM2 有面积优势
PHY 芯片功耗 (相对)	3.4-4.5	1	HBM2 有功耗优势
中介层	不需求	需要	GDDR6 成本优势
存储	平面, 与 DDR4、LPDDR4 类似	堆叠	GDDR6 成本优势

来源: Rambus, 中泰证券研究所

图表 14: HBM 的高带宽优势明显

Parameter	LPDDR4x	LPDDR5	DDR4	GDDR6	HBM2E
Bandwidth (Gbps)	Low-Medium (135)	Medium (204)	Medium (200)	High (512)	Highest (3686)
Data Rate (Gbps)	4.266	6.4	3.2	16	3.6
Interface width (bits)	32	32	64	32	1024
Board Area / System Design	Large / Medium	Medium / Medium	Large / Easy	Medium / Medium	Small / Complex
Efficiency (mW/Gbps)	High (3)	High (3)	Moderate (10)	Moderate (10)	Highest (2)
Cost (\$)	Medium	Medium	Low	Medium	High
Reliability/yield	Good	Good	Good	Good	Moderate
Applications	Mobile, AI	Mobile, AI	Compute, Network	AI, Graphics, Auto	AI, HPC, Network

来源: Rambus, 中泰证券研究所

图表 15: HBM 占用的面积小



来源: 应用材料, 中泰证券研究所

- **HBM 不断迭代, 迭代方向为增加容量和带宽, 目前最高层数为 12 层。**海力士 2014 年推出全世界第一颗 HBM, 2018 年推出 HBM2, 后续每隔两年推出新一代 HBM, 目前最新量产的是 HBM3, 预计 2024 年量产 HBM3E, 原厂加速研发。从单颗容量看, 堆叠层数和单层 DRAM 容量均有所增加, HBM1 仅堆叠

4层2Gb的DRAM，实现单颗HBM 8Gb（1GB），而HBM3E最高堆叠12层3GB的DRAM，实现单颗HBM 36GB，HBM4可能采用16层堆叠。从I/O数量看（总线位宽），HBM1到HBM3E均保持在1024bit，而数据的传输速率从HBM1的1Gb/s提升到HBM3E的9.2Gb/s，最终实现带宽从HBM1的128GB/s提升至HBM3E的1.2TB/s。HBM4的标准目前未确定，目前普遍预期HBM4最高16层堆叠，2048bit总线位宽。

图表 16：海力士 HBM 迭代情况

	HBM1	HBM2	HBM2E	HBM3	HBM3E	HBM4
年份	2014	2018	2020	2022	2024	2026
堆叠层数	4	4 or 8	4 or 8	8 or 12	8 or 12	12 or 16
单层DRAM容量	2Gb	1GB	2GB	2GB	3GB	3GB
容量	1GB	4GB OR 8GB	8GB OR 16GB	16GB OR 24GB	24GB OR 36GB	48GB或64GB
I/O数量（总线位宽，bit）	1024	1024	1024	1024	1024	2048
I/O速度（数据的传输速率）	1Gbps	2.4Gbps	3.6Gbps	6.4Gbps	9.2 Gbps	?
带宽	128GB/s	307GB/s	460GB/s	819GB/s	1.2TB/s	?
电压	1.2V	1.2V	1.2V	1.1V	?	?

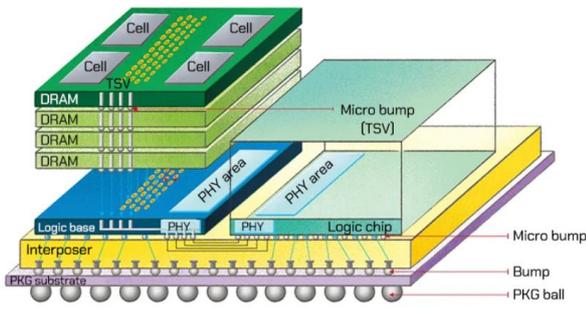
来源：海力士，中泰证券研究所

- 影响 HBM 性能的因素主要是 HBM 带宽、容量、功耗和尺寸规格。**
  - 提升 HBM 容量：**
    - 增加 DRAM Die 的堆叠层数。
    - 增加每片 DRAM Die 的容量
    - 单颗 GPU 增加使用 HBM 的颗数。单颗 HBM 容量=DRAM 堆叠层数×每片 DRAM 容量，GPU 使用 HBM 总容量=单颗 HBM 容量\*HBM 颗数。
  - 提升 HBM 带宽：**
    - 提高 I/O 数据的传输速率。
    - 更宽的 I/O 总线位宽。带宽是处理器与显存之间的数据的传速度，带宽(GB/s)=数据的传输速率（即 I/O 读写速度，Gbps）× 总线位宽 (bit) /8。HBM 主要是通过增加总线位宽的宽度提升带宽，同时数据的传输速率较慢，实现运行功耗更低、面积更小，但因涉及到复杂的堆叠、中介层等技术，成本更高。
  - 功耗：**数据的传输速度越大，功耗越大，数据的传输的距离越远，功耗越大。
- HBM 方案下，GPU 增加带宽的方式主要是增加 HBM 颗数和提升 HBM 性能。**
  - 增加 HBM 颗数：**目前 1 颗 HBM3 可提供 1024bit 总线位宽，增加 1 颗 HBM3，可增加 1024bit 总线位宽。但 HBM 的颗粒必须跟 GPU 对齐和封装在一起，是紧耦合的状态，受限 GPU 面积，HBM 数量不能无限增加，同时还需考虑散热等问题。
  - 提升 HBM 性能：**提高单颗 HBM 的带宽。

## 1.2 HBM 随 AI 爆发式增长，2024 年达到百亿美金规模

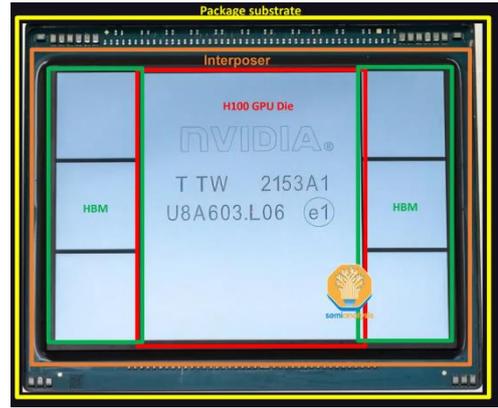
- 目前主流 AI 训练芯片都使用 HBM，一颗 GPU 配多颗 HBM。**以英伟达 H100 为例，1 颗英伟达 H100 PCIe 使用台积电 CoWoS-S 封装技术将 7 颗芯片（1 颗 GPU+6 颗 HBM）封在一起，1 颗 GPU 由 6 颗 HBM 环绕，其中 5 颗是 active HBM，每颗 HBM 提供 1024bit 总线位宽，5 颗共提供 5120bit 总线位宽，另外 1 颗是 non-HBM，可使用硅材料，起到芯片的结构支撑作用。H100 PCIe 的 HBM 总容量 80GB，使用 5 颗 activer HBM2E，每颗 HBM2E 容量 16GB，每颗 HBM2E 是由 8 层 2GB DRAM Die 堆叠组成。

图表 17: HBM 结构



来源: 海力士, 中泰证券研究所

图表 18: H100 封装结构



来源: Semianalysis, 中泰证券研究所

- **英伟达 HBM 用量提升。**英伟达 H100 PCIe 版本使用 80GB HBM2E, H200 提升到 144GB (6 颗 HBM3E), 最新发布的 B100、B200 采用 192GB (8 颗 8 层 HBM3E)。

图表 19: AI 训练芯片的 HBM 使用情况

		HBM用量 (GB)	HBM版本	HBM颗数	每颗HBM容量 (GB)	I/O总线位宽(bit)	传输速度 (GT/s)	带宽(GB/s)	Layers
英伟达	A100 40GB PCIe	40	2	5	8	5120	2.43	1,555	4/8+1
	A100 80GB	80	2E	5	16	5120	3.02	1,935	8+1
	A100 40GBSXM	40	2	5	8	5120	2.43	1,555	4/8+1
	A100 80GBSXM	80	2E	5	16	5120	3.19	2,039	8+1
	H100 PCIe	80	2E	5	16	5120	3.19	2,039	8+1
	H100 SXM	80	3	5	16	5120	5.23	3,350	8+1
	H100 NVL	192	3	12	16	12288	5.08	7,800	8+1
	H100S SXM	120/144	3	5/6	24	5120	5.6	3584/4301	12+1
	H200	144GB	3E	6	24	6144	6.5	48000	8+1
谷歌	B100	192GB	3E	8	24	8192	8	8.19	8+1
	B200	192GB	3E	8	24	8192	8	8.19	8+1
	Google TPUv4i	8	2	2	24	5120	2.29	585	4+1
	Google TPUv4	32	2	4	24	5120	2.34	1,200	8+1
AMD	Google TPUv5i	16	2E	2	24	5120	3.20	819	4+1
	Google TPUv5	64	3	4/6	24	5120	5.20	2662/3993	8+1
	AMD MI250X	128	2E	8	24	5120	3.20	3,277	8+1
	AMD MI300A	128	3	8	24	5120	5.20	5,325	8+1
	AMD MI300X	192	3	8	24	5120	5.60	5,734	12+1

来源: Semianalysis、ANANDTECH, 中泰证券研究所

- **预计 24 年 HBM 市场规模达百亿美金, 较 23 年翻倍。**以搭载 8 颗英伟达 H100 的 AI 服务器为例, H100 配套 5 颗 HBM2E, 单颗 HBM2E 容量 16GB, 8 层堆叠、每层堆叠 2GB, 每颗 H100 需要 80GB HBM, 单颗服务器需要 640GB HBM; HBM 每 GB 单价 10-20 美金。假设单颗 GPU 的 HBM 使用量每年提升 40%左右, 假设 AI 服务器 24-25 年渗透率 12%/13%, 预计 24 年 HBM 市场需求 151 亿美金, 较 23 年基本翻倍。

图表 20: HBM 市场需求测算

	2023	2024E	2025E
单颗GPU的HBM总容量 (GB)	80	112	156.8
yoy		40%	40%
单台AI服务器的GPU使用量 (颗/台)	8	8	8
单台AI服务器的HBM总容量 (GB)	640	896	1254.4
全球服务器出货量 (万台)	1,339	1,365	1,434
yoy	-6%	2%	5%
全球AI服务器出货量 (万台)	120.5	165.2	189.5
AI服务器渗透率	9%	12%	13%
全球AI服务器的HBM需求量 (亿GB)	8	15	24
HBM单价 (美金/GB)	10	10	10
AI服务器HBM市场需求 (亿美金)	76	151	238
DRAM市场规模 (亿美金)	519	822	1,028
yoy		58%	25%
HBM占DRAM市场占比	15%	18%	23%

来源: Trendforce 等, 中泰证券研究所

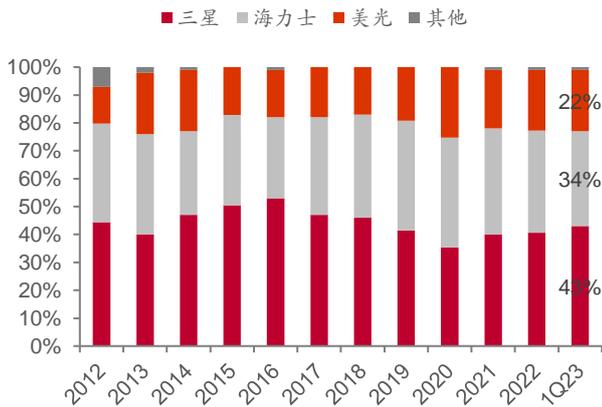
### 1.3 三大原厂垄断市场, 开启军备竞赛

#### ■ 图形和服务器 DRAM 三大家垄断程度高。

1) 分下游三大原厂合计市占率排序 (23Q1): 图形 DRAM (100%) > 服务器 DRAM (99%) > 移动端 DRAM (94%) > 利基 DRAM (71%), 图形、服务器 DRAM 基本全部被三星、海力士和美光垄断, 移动端和利基 DRAM 市场有其他玩家, 尤其是利基 DRAM 市场, 代表厂商有中国台湾厂商华邦、南亚、南亚, 陆厂商长鑫存储、北京君正、东芯股份、兆易创新等。

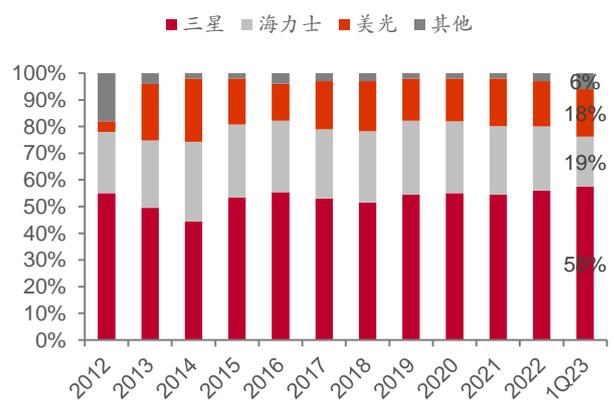
2) 分下游玩家: 三星在每个细分市场均为第一, 图形 DRAM 市场, 三星/海力士/美光份额 35%/31%/34%, 三大原厂份额较为均衡, 在服务器市场, 三星/海力士/美光份额 43%/34%/22%, 移动端市场, 三星/海力士/美光份额 53%/19%/18%, 三星独占一半份额 (三星手机约占据 20% 左右全球份额), 在利基 DRAM 市场, 三星/海力士/美光份额 33%/13%/25%。

图表 21: Sever DRAM 市场份额



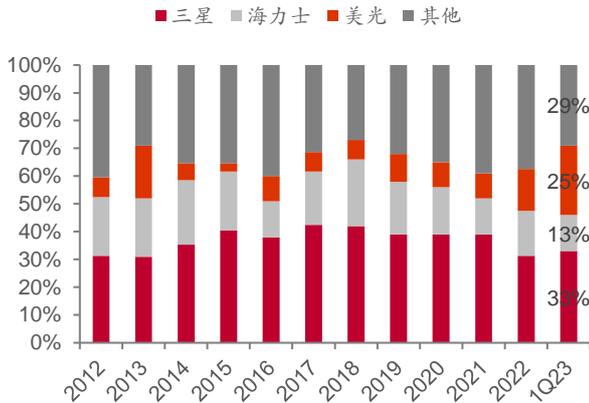
来源: OMDIA, 中泰证券研究所

图表 22: Mobile DRAM 市场份额



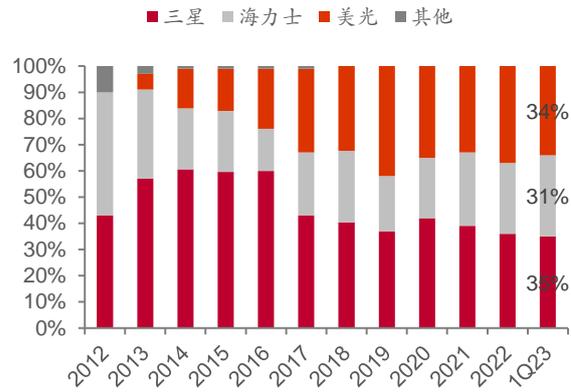
来源: OMDIA, 中泰证券研究所

图表 23: Consumer Eletronics DRAM (利基 DRAM) 份额



来源: OMDIA, 中泰证券研究所

图表 24: Graphic DRAM 市场份额



来源: OMDIA, 中泰证券研究所

■ 未来 DRAM 制程演进需要使用 EUV 光刻技术, 三星最早使用。

1) DRAM 制程: 进入 10nm 级别制程后迭代速度放缓, 使用 1x (16-19nm)、1y (14-16nm)、1z (12-14nm) 等字母表示, 另外三星海力士使用 1a (约 13nm)、1b (10-12nm)、1c (约 10nm), 对应美光 1α、1β、1γ。

2) 三星在 1znm 就已使用 EUV 光刻技术。三星在 2020 年在 1znm DDR5 上采用 1 层 EUV, 2021 年量产采用 5 层 EUV 的 DDR5, 三星是首家在 DRAM 采用 EUV 光刻技术的厂商, 也是在 DRAM 上使用 5 层 EUV 的厂商。而海力士和美光在 1znm 仍然使用 ArF-i 光刻工艺, 2021 年海力士在 1anm 转向使用 EUV, 后续在 1bnm 继续使用 EUV。而美光在 2023 年宣布开始在 1cnm (1γ 制程) 使用 EUV 光刻技术。EUV 技术, 制程更小、单位容量更大, 成本更有优势。

图表 25: 三大原厂在不同制程使用 EUV 技术的情况

			三星	海力士	美光
第一代 10nm 级别	1xnm (D1x)	16-19nm	Test vehicle (试验版本, 未规模量产)		
第二代 10nm 级别	1ynm (D1y)	14-16nm			
第三代 10nm 级别	1znm (D1z)	12-14nm	√ (1 层 EUV)		
第四代 10nm 级别	1anm (D1a)	约 13nm	√ (5 层 EUV)	√	
第五代 10nm 级别	1βnm (D1b)	10-12nm	√	√	
第六代 10nm 级别	1γnm (D1c)	约 10nm, 1β 的增强版	√ (未量产)	√ (未量产)	√ (未量产)

来源: Techinsights, 中泰证券研究所整理

■ 2024 年三大原厂将以迭代量产 1bnm (1β) 为主, 海力士和美光 HBM3E 将直接使用 1bnm, 三星采用 1anm。

目前, DRAM 先进制程技术已发展至第五代 10nm 级别, 美光称之为 1β nm DRAM, 三星和海力士称为 1bnm DRAM。美光最先量产 1bnm 级别 DRAM。

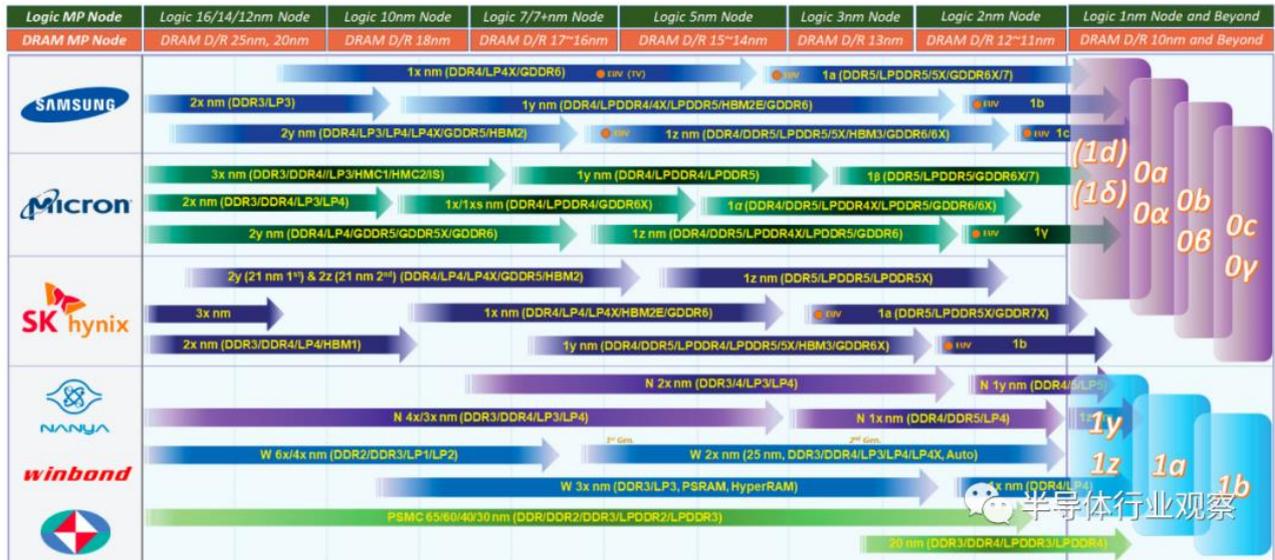
1) 三星: 2023 年 5 月三星量产 12nm 16Gb DDR5 DRAM, 9 月三星开发出基于 12nm 级工艺技术的 32Gb DDR5 DRAM, 将于 2023 年底开始量产。三星透露将于 2026 年推出 DDR6 内存, 2027 年即实现原生 10Gbps

的速度。据披露，三星正在开发行业内领先的 11nm 级 DRAM 芯片。

2) **海力士**：2023 年 1 月，海力士 1anm 服务器 DDR5 适用到英特尔® 第四代至强®可扩展处理器，并在业界首次获得认证。5 月，海力士 1bnm 技术完成研发，采用“HKMG (High-K Metal Gate)”工艺，与 1a nm DDR5 DRAM 相比功耗减少了 20%以上。

3) **美光**：2022 年日本厂开始量产 1bnm (1β) 制程，今年中国台湾厂也开始量产 1bnm (1β) 制程，预计 1cnm (1γ) 制程在 25H1 在中国台湾厂量产，1cnm (1γ) 制程使用 EUV 技术。

图表 26：DRAM 制程迭代

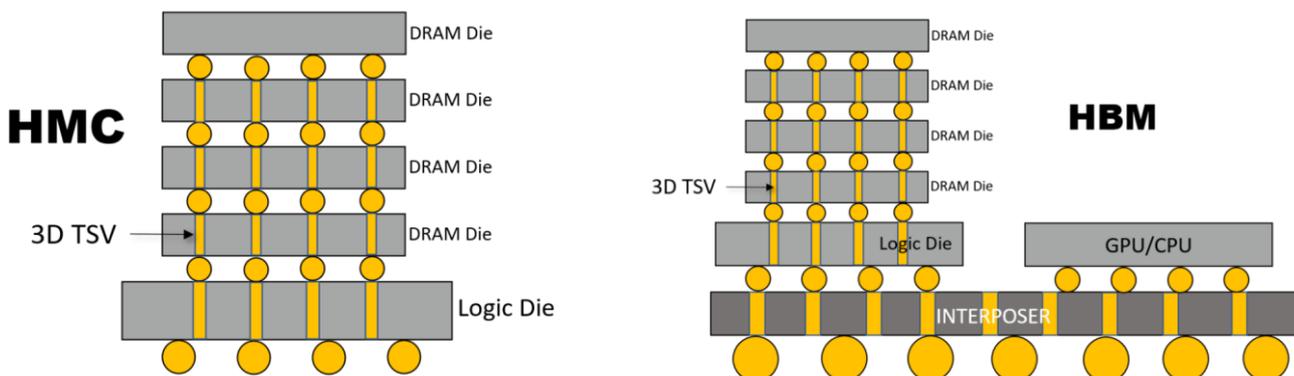


来源：半导体行业观察、Techinsights，中泰证券研究所

- **三星和海力士垄断 9 成 HBM 市场，美光份额落后。**HBM 市场垄断效应更强，2022 年海力士/三星份额为 50%/40%，美光份额仅 10%，海力士 HBM3 产品领先其他原厂，是英伟达 AI 芯片 HBM 的主要供应商，份额最高，而三星着重满足其他云端服务业者的订单，在客户加单下，预计在 HBM3 与海力士的市占率差距会大幅缩小，2023~2024 年三星和海力士市占率预估相当，合计拥 HBM 市场约 95%的市占率，不过因客户组成略有不同，在不同季度的出货表现可能有先后。
- **美光因技术路线判断失误在 HBM 市场份额比较低，在追赶中。**美光此前在 HMC 投入更多资金，HMC (Hybrid Memory Cube, 混合内存立方体) 将 DRAM 堆叠、使用 TSV 硅穿孔技术互连，DRAM 下方是一颗逻辑芯片，从处理器到存储器的通信是通过高速 SERDES 数据链路进行的，该链路会连接到 DRAM 下面的逻辑控制器芯片，但不同于 HBM，HBM 是与 GPU 通过中介层互连。HMC 是与 HBM 竞争的技术，美光 2011 年推出 HMC，海力士 2013 年推出 HBM，HMC 与 HBM 开发时间相近，但 HBM 未被大规模使用，原因如下：1) 相较 HBM 直接与处理器封装在一起，HMC 距离处理器较远，延迟更大。2) HBM 推出不久后就被认证为 JEDEC 标准，而 HMC 推出比 HBM 早 2 年但未被 JEDEC 定为标准，JEDEC 拥有数百家会员公司、奉行一公司一票与三分之二多数的制度，从而降低标准制定被任何一家或一批公司所把控的风险，因此只有大家真正认可，才会最终被推行为正式标准。美光 2018 年由 HMC 转向 HBM。

- 2023 年主流需求自 HBM2E 转向 HBM3，预计 2024 年转向 HBM3 及 3E。随着使用 HBM3 的 AI 训练芯片陆续放量，2024 年市场需求将大幅转往 HBM3，而 2024 年将有望超越 HBM2E，比重预估达 60%，且受惠于其更高的平均销售单价，将带动明年 HBM 营收显著成长。

图表 27: HBM 发展历程



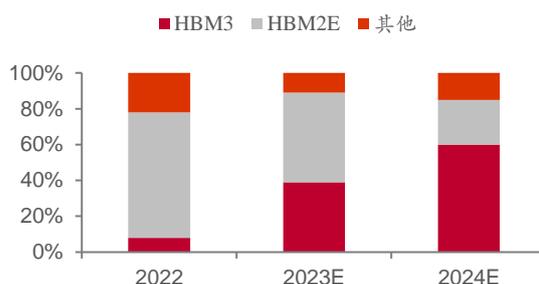
来源：面包板，中泰证券研究所

图表 28: HBM 竞争格局

	2022	2023E	2024E
海力士	50%	46%~49%	47%~49%
三星	40%	46%~49%	47%~49%
美光	10%	4%~6%	3%~5%

来源：Trendforce，中泰证券研究所

图表 29: HBM 不同世代占比



来源：Trendforce，中泰证券研究所

- 海力士是 HBM 先驱、技术最强，最早与 AMD 合作开发，三星紧随其后。

海力士在 2015 年首次为 AMD Fiji 系列游戏 GPU 提供 HBM，该 GPU 由 Amkor 进行 2.5D 封装，随后推出使用 HBM2 的 Vega 系列，但 HBM 对游戏 GPU 性能未产生太大改变，考虑没有明显的性能优势和更高的成本，AMD 在 Vega 之后的游戏 GPU 中重新使用 GDDR，目前英伟达和 AMD 的游戏 GPU 仍然使用更便宜的 GDDR。随着 AI 模组中参数数量的指数级增长，内存墙问题愈加突出，英伟达在 2016 年发布首款 HBM GPU P100，后续英伟达数据中心 GPU 基本都采用海力士 HBM。

海力士 22Q4 量产全球首款 HBM3，而三星由于此前降低 HBM 的投入优先级，HBM3 较海力士晚推出一一年。

- 三大原厂积极扩产 HBM 和推进产品迭代，预计 24H2 HBM3E 量产，未来成为市场主流。

1) 海力士：24 年 Capex 优先保障 HBM 和 TSV 产能，23 年 HBM 产能已出售完、同时持续收到额外需求，预计 DDR5 和 HBM 产线规模将在 24 年增长 2 倍+。公司已从 2023 年 8 月开始提供 HBM3E 样品，2024 年 1 月中旬结束开发，3 月开始量产 8 层 HBM3E，3 月底发货。12 层 HBM3E

已于2月送样。

2) 三星：计划24年HBM产能提高2.5倍。23Q3已量产8层和12层HBM3，计划Q4进一步扩大生产规模，并开始供应8层HBM3E样品，2024年2月底发布12层HBM3E，预计H1量产。

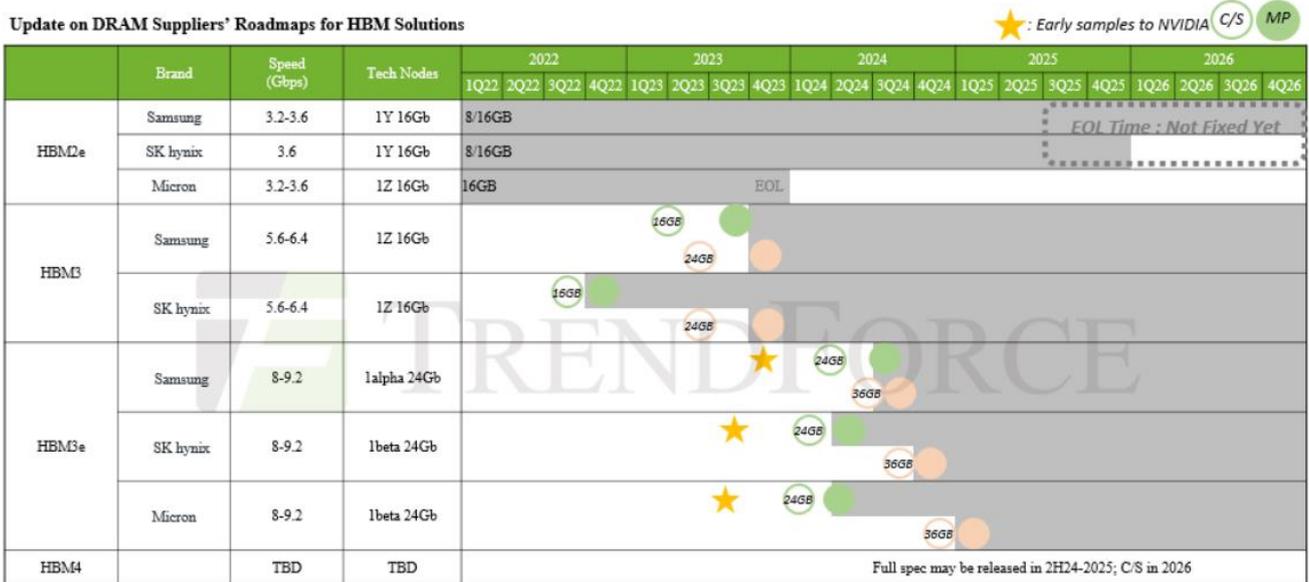
3) 美光：跳过HBM3，直接做HBM3E，2024年3月宣布量产8层HBM3E，将用于英伟达H200，3月送样12层HBM3E。

图表 30: HBM 发展历程

	HBM1		HBM2				HBM2E		HBM3		HBM3E		HBM4
	2013	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024E	2025	2026
HBM原厂	海力士	AMD联合开发全球首款HBM；制程：29nm			发布HBM2；制程：21nm		全球率先量产HBM2E	10月开发出全球首款HBM3	6月率先开始量产HBM3，Q3开始向英伟达	4月宣布已经率先研发出12层HBM3；8月送样8层HBM3E	2月送样12层HBM3E；3月量产8层HBM3E；制程：1b nm		预计量产HBM4
	三星		从HBM2开始布局，率先量产HBM2；制程：20nm				推出HBM2E，后续量产。制程：1ynm	2月推出了HBM-PIM（存算一体）	表示HBM3已量产；制程：1b nm	Q3大规模量产HBM3；8层HBM3E送样	3月12H HBM3E已送样，预计H2量产；制程：14nm		预量产HBM4
	美光			HMC			宣布将会开始提供HBM2产品			7月宣布HBM3E已经开始送样；制程：1P nm	3月8层HBM3E宣布量产		预计量产HBM4
GPU厂商	AMD	AMD Radeon R9 Fury X，第一款搭载HBM的GPU，使用SK海力士的HBM		在Vega架构的GPU上搭载HBM2	M150和M160均搭载HBM2；Radeon Instinct加速器搭载三星HBM2	AMD Radeon VII搭载HBM2	11月发布MI100搭载HBM2	11月发布MI250和MI250X均搭载HBM2E	3月发布MI210 PCIe搭载HBM2E	发布MI300A和MI300X搭载HBM3；Instinct MI100 GPU搭载了三星的HBM-PIM存储器			
	英伟达		NVIDIA Tesla P100，第一款搭载HBM2的GPU，使用三星HBM2				NVIDIA A100 80GB，第一款搭载HBM2E的GPU	升级A100 PCIe GPU，配备80GB HBM2E	H100 PCIe搭载HBM2E	H100 SXM首款使用HBM3的GPU，使用海力士产品，后续H100 NVL和H100S SXM都搭载HBM3；11月发布H200	发布B100、B200		

来源：各公司官网，中泰证券研究所

图表 31: 三大原厂 HBM 开发进度



Source: TrendForce, Nov., 2023

来源：Trendforce，中泰证券研究所

图表 32: 海力士官网的 HBM 产品料号情况

	Part No.	单层容量 (GB) a	堆叠层数 (Hi) b	单颗容量 (GB) a*b	显存数据频率 (Gbps) d	I/O总线位宽 (bit) e	显存带宽 (GB/s) d*e/8	电压	状态
HBM2E	H5WR64ESM4W-N8L	2	4	8	3.6	1024	461	1.2V	大规模量产
	H5WR64ESM4W-N6L	2	4	8	3.2	1024	410	1.2V	大规模量产
	H5WRAGESM8W-N8L	2	8	16	3.6	1024	461	1.2V	大规模量产
	H5WRAGESM8W-N6L	2	8	16	3.2	1024	410	1.2V	大规模量产
HBM3	H5UG7HMD83X020R	2	8	16	5.6	1024	717	1.1V	大规模量产
	H5UG7HME03X020R	2	8	16	6	1024	768	1.1V	商用样品

来源: 海力士官网, 中泰证券研究所

图表 33: 三星官网的 HBM 产品料号情况

	Part Number	Density	Organization	Speed	Refresh	Package	Product Status
HBM2 Flarebolt	KHA843801B-MC12	4 GB	1024	2.0 Gbps	32 ms	MPGA	EOL
	KHA883901B-MC12	8 GB	1024	2.0 Gbps	32 ms	MPGA	EOL
HBM2 aquabolt	KHA844801X-MC12	4 GB	1024	2.0 Gbps	32 ms	MPGA	Mass Production
	KHA844801X-MC13	4 GB	1024	2.4 Gbps	32 ms	MPGA	Mass Production
	KHA844801X-MN12	4 GB	1024	2.0 Gbps	32 ms	MPGA	Mass Production
	KHA844801X-MN13	4 GB	1024	2.4 Gbps	32 ms	MPGA	Mass Production
	KHA884901X-MC12	8 Gb	1024	2.0 Gbps	32 ms	MPGA	Mass Production
	KHA884901X-MC13	8 GB	1024	2.4 Gbps	32 ms	MPGA	Mass Production
	KHA884901X-MN12	8 GB	1024	2.0 Gbps	32 ms	MPGA	Mass Production
2E Flashbolt	KHAA44801B-MC17	8 GB	1024	3.6 Gbps	32 ms	MPGA	Mass Production
	KHAA84901B-MC17	16 GB	1024	3.6 Gbps	32 ms	MPGA	Mass Production
	KHAA44801B-MC16	8 GB	1024	3.2 Gbps	32 ms	MPGA	Mass Production
	KHAA84901B-JC16	16 GB	1024	3.2 Gbps	32 ms	MPGA	Mass Production
	KHAA84901B-JC17	16 GB	1024	3.6 Gbps	32 ms	MPGA	Mass Production
	KHAA84901B-MC16	16 GB	1024	3.2 Gbps	32 ms	MPGA	Mass Production
HBM3 Icebolt pro	KHBA84A03D-MC1H	16 GB	1024	6.4 Gbps	32 ms	MPGA	Mass Production
	KHBAC4A03D-MC1H	24 GB	1024	6.4 Gbps	32 ms	MPGA	Mass Production
	KHBA84A03C-MC1H	16 GB	1024	6.4 Gbps	32 ms	MPGA	Mass Production
	KHBAC4A03C-MC1H	24 GB	1024	6.4 Gbps	32 ms	MPGA	Sample

来源: 三星官网, 中泰证券研究所

注: 美光官网未列示 HBM 料号。

## 2.HBM 制造中先进封装大放异彩

### 2.1 采用多种先进封装工艺, 重点关注堆叠键合方式

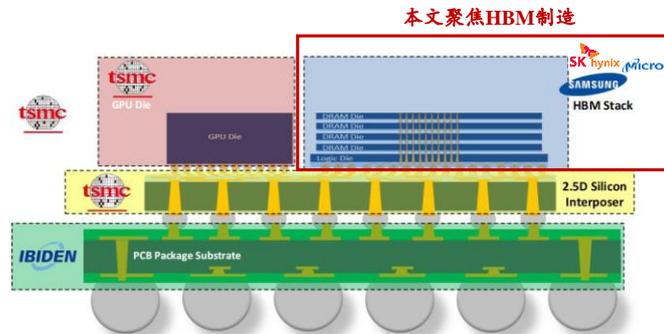
- HBM 颗粒采用 KGSD 封装形式, 由 1 片逻辑芯片+多片 DRAM 芯片组成。**海力士、三星等存储原厂将 HBM 采用晶圆级封装, 以 KGSD (Known Good Die Stack, 已知合格堆叠芯片) 的封装形式交给台积电, 台积电使用 2.5D 封装技术 (包括 CoWoS) 将 HBM 与 SoC (GPU 等) 封装在一起, 本文重点介绍 HBM 制备流程、工艺和相关设备材料, 关于 CoWoS 工艺的具体介绍, 详情见《[AI 系列之先进封装: 后摩尔时代利器, AI+国产化紧缺赛道](#)》。

1 颗 HBM KGSD = N 颗 DRAM 芯片 (也称为 Core Die) + 1 颗逻辑芯片 (也称为 Logic Base Die) 组合而成, 目前 N=4/8/12, 预计 HBM4 将采用 16 颗 DRAM 芯片堆叠。

逻辑芯片主要包括三个功能区, ①用于测试的区域 (DFT Area), ②TSV 区域, TSV 用于给 DRAM 芯片传输信号和电力, ③PHY 芯片区域, HBM 和 SoC 中的存储控制器之间的接口。PHY 芯片区域和 TSV 区域中

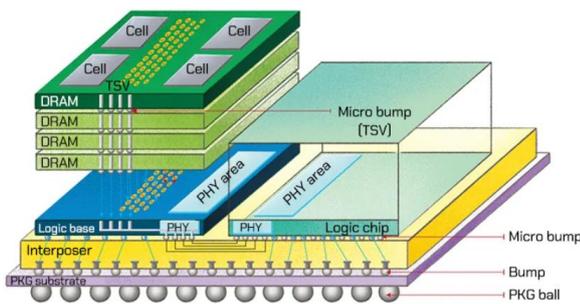
间有 1024 根信号传输线路，对应 1024bit 总线位宽。逻辑芯片的大小通常大于 DRAM 芯片，如海力士 8 层 HBM3 的逻辑芯片大小为 10.8 mm x 9.8 mm，而 DRAM 芯片为 10.5 mm x 9.5 mm，这是为了可以模塑封装（Mode 晶圆模塑，一种扇出型晶圆级芯片封装工艺）以保护晶圆，通常使用环氧树脂模塑料（EMC）作为填充材料。

图表 34：英伟达 P100 芯片供应链情况



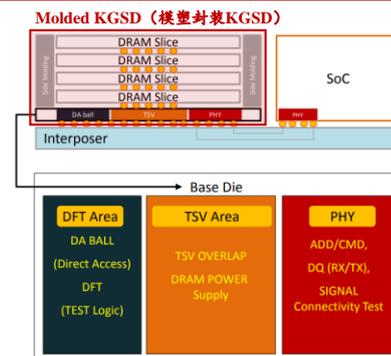
来源：System Plus Consulting，中泰证券研究所

图表 35：HBM 结构



来源：海力士，中泰证券研究所

图表 36：HBM 的逻辑芯片的功能区



来源：海力士，中泰证券研究所

- HBM 使用 TSV 技术、采用 3D 堆叠结构，采用先进封装与 GPU 封装在一起，在不占用面积的前提下，实现容量拓展、有效提升带宽和降低功耗。将多片 HBM DRAM Die 堆叠在一颗 Logic Die，DRAM Die 之间、DRAM 和 Logic Die 均通过硅通孔（TSV）和 Bump（凸点）垂直互连。DRAM 与 Logic Die 放置在 Interposer（中介层）上与 GPU 互联，中介层放置在 ABF 载板上，最后 HBM 与 GPU 使用系统级封装技术封在一起。

HBM 采用的主要技术为：1) TSV、Bumping 技术：有利于 HBM 增加容量，可以通过堆叠更多层数实现扩容，同时实现了较短的信号传输路径，使 HBM 具备更好的内存功耗能效特性，同时也增加了引脚数量，使得每颗 HBM 能有 1024bit 总线位宽。2) 2.5D 封装采用中介层：中介层中连接 GPU 和 DRAM，相较传统 PCB 的电线数量更多，一方面缩短与 GPU 的数据的传输路径，减少能耗，更重要的是实现 GPU 和 DRAM 的高带宽数据的传输。3) 系统级封装：HBM 将原本在 PCB 板上的 DDR 内存颗粒和 GPU 一起全部集成到 SiP 里，因此 HBM 在节省产品空间方面也更具优势。

- HBM 制造流程分为四步，涉及 TSV（硅通孔）、Bumping（凸点制造）、

**堆叠键合等技术。**HBM 从设计、制造和封测方式均与传统 DRAM 有较大区别，相较传统 DRAM，HBM 多了 TSV、逻辑晶圆制备、凸点制造、堆叠键合等工艺，主要差异集中在封装测试部分，HBM KGSD 的制备工艺包括扇外型晶圆级封装、TSV、Microbumping 等先进封装技术。

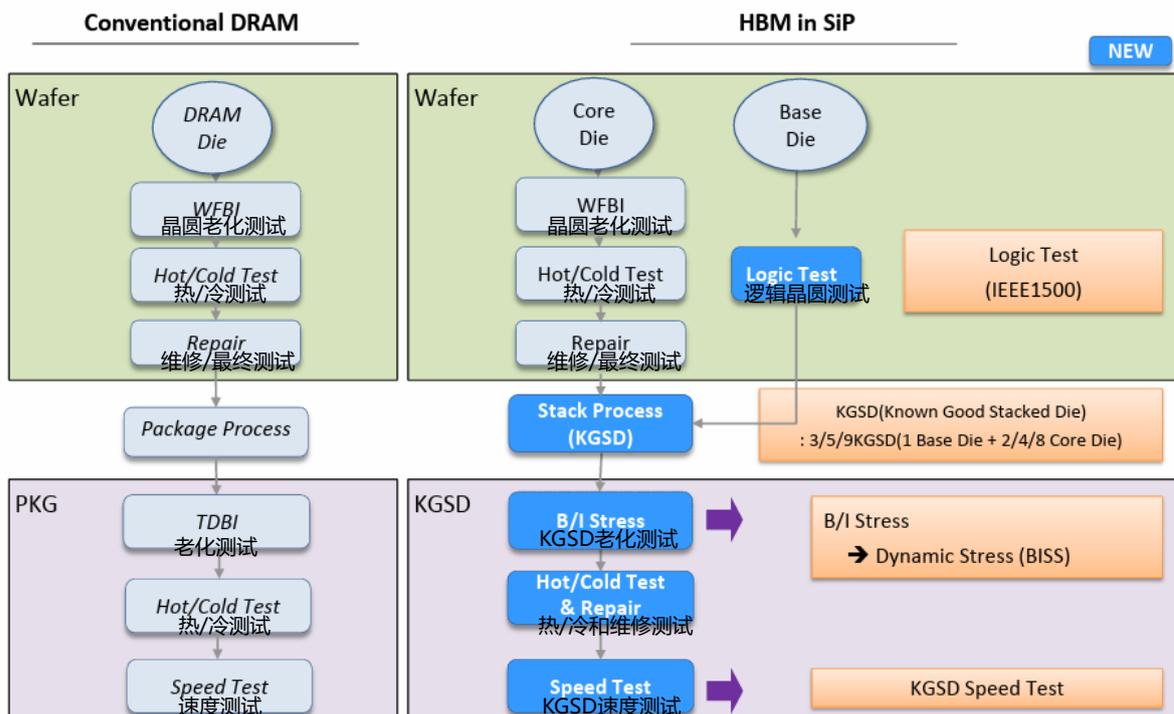
1) **晶圆制造 (包括 TSV):** 分别制造 DRAM 晶圆和逻辑晶圆，同时做好 DRAM 和逻辑晶圆的 TSV 硅通孔，TSV 硅通孔需要晶圆制造工艺，包括深孔刻蚀、气相沉积、铜填充、CMP、晶圆减薄等晶圆制造工艺，此时 DRAM 和逻辑都是处于晶圆阶段，与传统 DRAM 主要差异是 HBM 晶圆需要制造 TSV。

2) **凸点制造 (Microbumping):** 将硅通孔后的 DRAM 晶圆和逻辑晶圆倒装，然后进行减薄，在晶圆背面形成凸点，此时 DRAM 和逻辑都是处于晶圆阶段。

3) **堆叠和键合 (Stacking&Bonding), 主要的差异化环节:** 在进行堆叠前，DRAM 晶圆和逻辑晶圆的 TSV 通孔和凸点均已做好，DRAM 晶圆切割成 DRAM 颗粒，DRAM 颗粒一层一层堆叠在逻辑晶圆上，然后进行键合，再进行晶圆模塑封装，最后获得模塑封装后的 KGSD (Molded KGSD)。海力士和三星/美光主要是在键合工艺上有差异，三星/美光使用较为传统的 TC-NCF (Thermo-Compression Bonding with None Conductive Film, 热压缩-非导电薄膜)，先在有 TSV 和凸点的晶圆上填充 NCF，然后堆叠进行热压键合，后进行模塑封装，而海力士采用独创的 MR-MUF 工艺 (Mass Reflow Bonding with Molded UnderFill, 大规模回流焊-注塑底填充技术)，不使用 NCF，直接先堆叠，然后进行大规模回流焊做凸点的键合，然后使用以液体 EMC 为主要原材料的 MUF 使用模塑方式填充缝隙，工艺具体介绍详见后文。

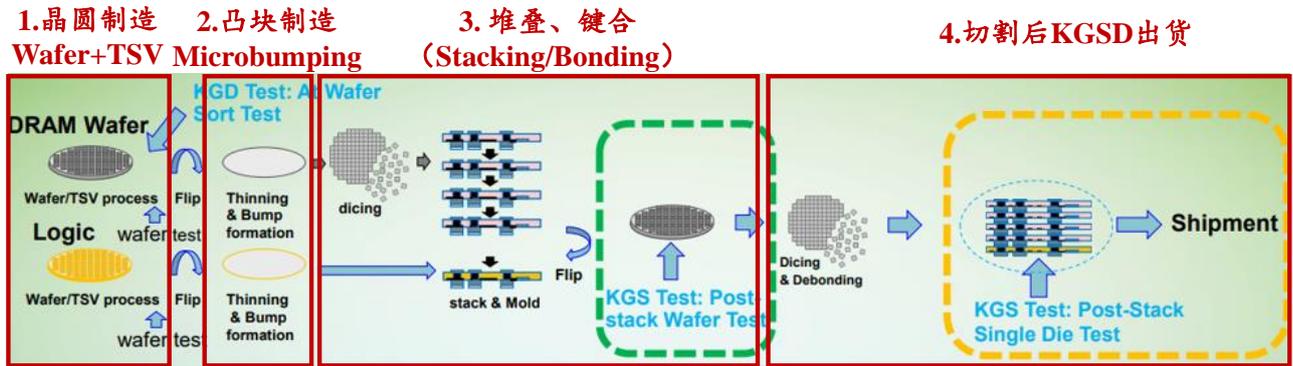
4) **切割 KGSD 晶圆获得 KGSD 颗粒:** 将模塑后的 KGSD 晶圆切割成颗粒，测试完成后出货给台积电继续做 CoWoS 封装。

图表 37: 传统 DRAM 与 HBM 制造和封测流程对比



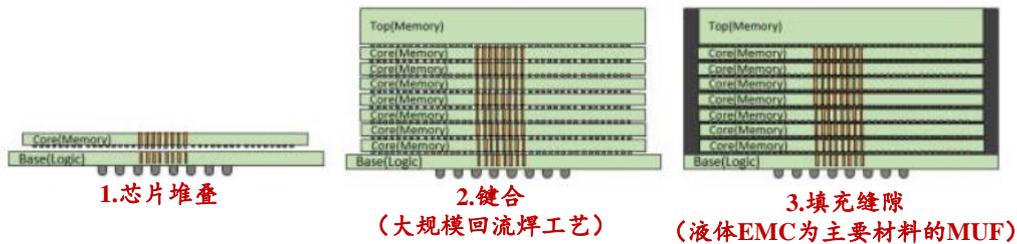
来源: 海力士, 中泰证券研究所

图表 38: HBM Stack 制造流程



来源: 海力士, 中泰证券研究所

图表 39: 海力士 MR-MUF 工艺

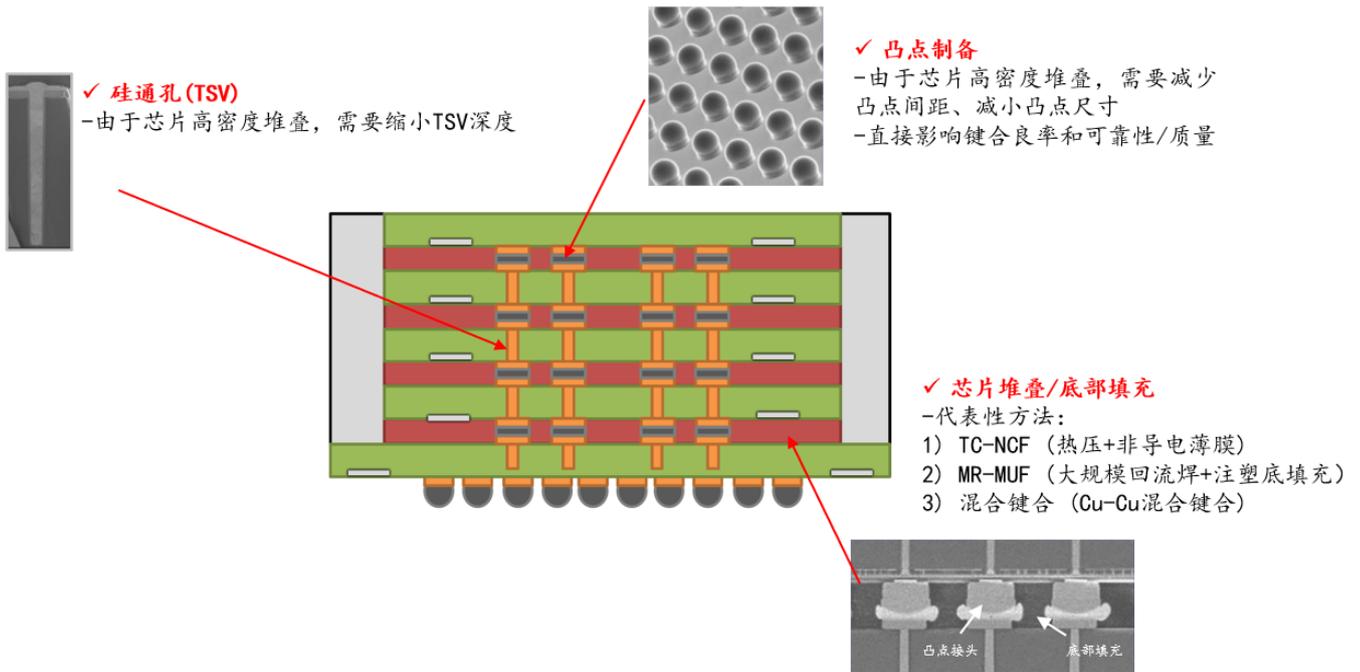


来源: 海力士, 中泰证券研究所

## 2.2 HBM 三大关键工艺: TSV、Micro bump 和堆叠键合

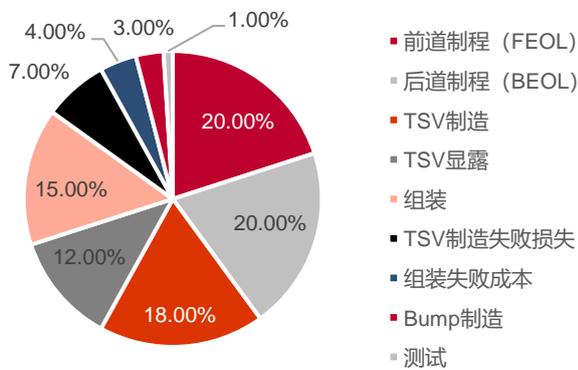
- **TSV 在 HBM 成本中占比最高, 约 30%。** HBM 核心工艺主要是 TSV、micro bump 和堆叠键合, 其中 TSV 工艺是 HBM 中成本占比最高、最核心的工艺, 利用 TSV 才能实现 DRAM 芯片的 3D 堆叠和芯片间的快速传输。根据 3D InCites 2016 年数据, 在 4 层 DRAM 和 1 层逻辑的 HBM 中, 99.5%的键合良率下, TSV 工艺所占的成本比重为 30%, 其中 TSV 制造 (在正常晶圆厚度上制作 TSV 的过程) 为 18%, TSV 显露 (晶圆减薄等工艺使 TSV 触点露出) 为 12%; 在 99%键合良率下, TSV 工艺所占的成本比重为 28%, 其中 TSV 制造为 17%, TSV 显露为 11%。

图表 40: HBM 核心工艺: TSV、micro bump 和堆叠键合



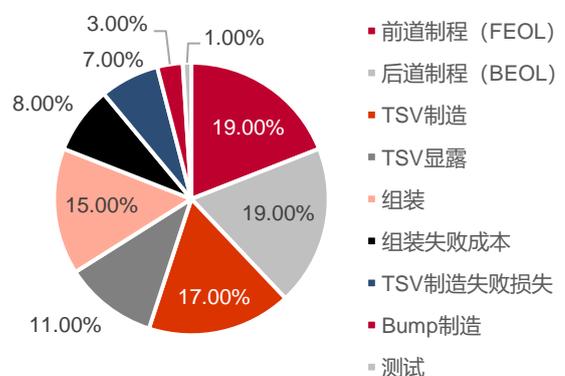
来源: 海力士, 中泰证券研究所整理

**图表 41: HBM(4层 DRAM+1层逻辑)3D 封装成本划分(99.5% 键合良率)**



来源: 3D Incites, 中泰证券研究所

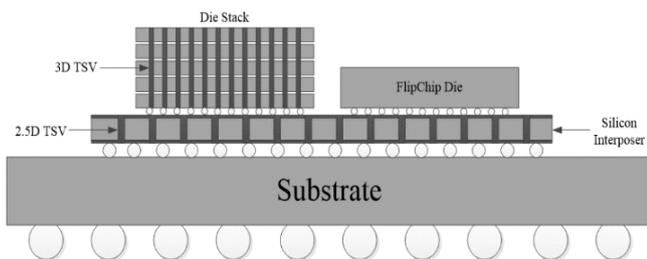
**图表 42: HBM (4层 DRAM+1层逻辑) 3D 封装成本划分 (99%键合良率)**



来源: 3D Incites, 中泰证券研究所

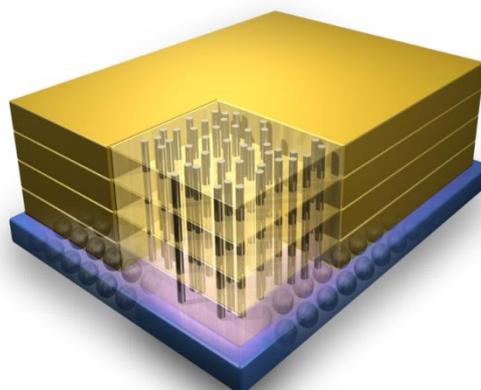
■ **TSV, 即 Through-SiliconVia, 指穿透 Si 晶圆实现各芯片层之间电互连的垂直导电柱。** RDL 主要在 XY 轴上进行电互连, 而 TSV 主要针对 Z 轴方向的电互连, 是唯一的垂直电互连技术。芯片三维堆叠技术需通过 TSV 实现多芯片的短距离高速通信。TSV 有 3 个关键特征: 1) 通过在芯片内部形成孔洞来实现电气互连; 2) 垂直连接芯片的不同层次, 实现多层堆叠结构; 3) TSV 中填充导电材料, 通过孔内材料导电实现电气互连。TSV 主要用于硅转接板、芯片三维堆叠等方面, 典型应用有 cowos、HBM。目前用于三维堆叠的 TSV 直径约为 5~10 μm, 深宽比约为 10:1, 未来先进 TSV 工艺的直径有望达到 1 μm, 深宽比达到 20:1, 实现更高密度的互连。

图表 43: 3D TSV 结构



来源:《先进封装与异构集成》, 中泰证券研究所

图表 44: TSV-Viafirst



来源:《先进封装与异构集成》, 中泰证券研究所

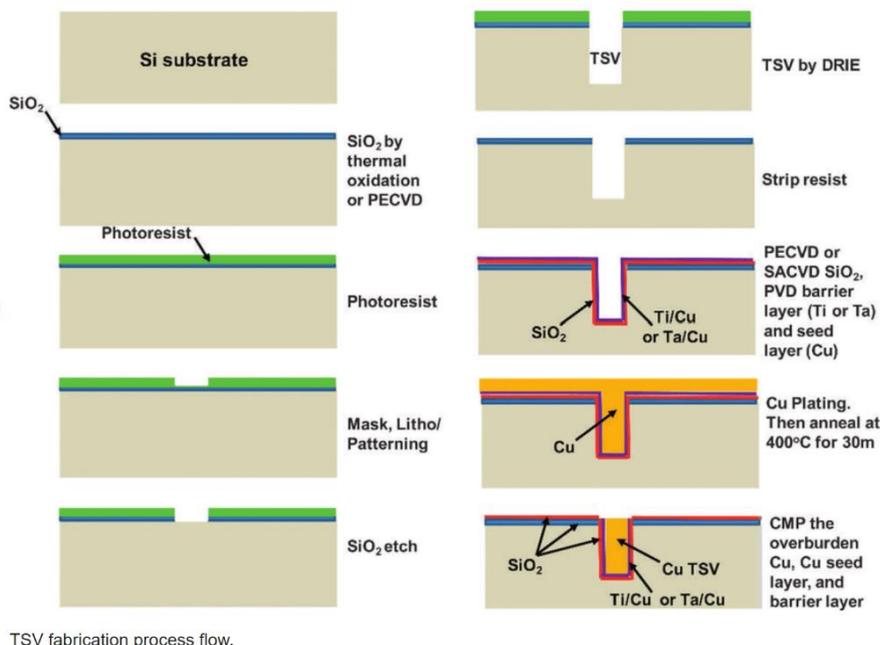
图表 45: TSV 工艺应用情况

产品	TSV Pitch	TSV Size	公司
3D 堆叠 CIS	6.3um	2.5um	Sony 等
3D NAND	-	1um-5um	长江存储等
Si 中介层	< 200um (与 Bump 间距有关)	> 11um	TSMC 等
HBM	35um	5~10um	Samsung/ Micron/SK Hynix 等

来源:《Through-Silicon-Via Design with Clustering Structure and Adaptive Through-Silicon-Via Control for Three-Dimensional Solid-State-Drive Boost Converter System》等, 中泰证券研究所整理

- TSV 制造涉及到深孔刻蚀、气相沉积、铜填充、CMP、晶圆减薄等工序设备, 技术难度高。TSV 制造的主要工艺流程依次为: 深反应离子刻蚀 (DRIE) 行成通孔→通过化学沉积的方法沉积中间介电层、使用物理气相沉积的方法沉积制作阻挡层和种子层→通过电镀或者 PVD 工艺在盲孔中进行铜填充→使用化学和机械抛光 (CMP) 去除多余的铜并对晶圆进行减薄。**从工艺次序角度可分为前通孔、中通孔、后通孔和键合后通孔等几种形式。TSV 技术的工艺难度高: 1) 通常要求晶圆减薄到 50 μm 以下, 须控制好晶圆减薄的水平度, 避免裂片、飞边。2) TSV 工艺对通孔的宽度以及深宽比都有严格要求, 目前首选技术是基于 Bosch 工艺的干法刻蚀, 实现了对腔室内等离子体密度的均匀控制, 满足硅高深宽比刻蚀工艺的要求。**涉及的设备&材料:** 光刻机 (光刻胶)、深孔刻蚀设备 (电子特气)、PVD (靶材)、CVD、电镀设备 (电镀液)、抛光机 (抛光液)、减薄机 (减薄液) 等。此外, 为了满足 TSV 工艺, 晶圆减薄已成为大势所趋, 但超薄晶圆容易产生翘曲, 因此在硅转接板的完整工艺流程中 (报告 3.1 节有流程介绍) 还需要用到**临时键合与解键合工艺:** 采用临时键合材料将完成一面图形制造的晶圆预键合到载片上, 继续进行背面工艺制作, 完成后将晶圆和载板剥离。

图表 46: TSV 工艺流程图



来源：《Redistribution layers (RDLs) for 2.5D/3DIC integration》，中泰证券研究所整理

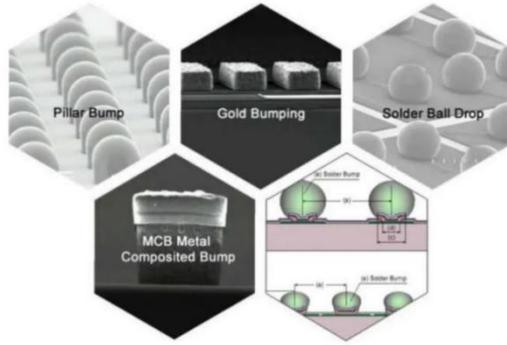
图表 47: TSV 工艺设备与材料

工艺	作用	设备	材料
光刻	对刻蚀区曝光	涂胶显影设备、光刻机	光刻胶
深硅刻蚀	使用深反应离子刻蚀 (DRIE) 法行成通孔	深硅刻蚀设备	电子特气
绝缘层/阻挡层/种子层沉积	使用化学沉积的方法沉积制作绝缘层, 使用物理气相沉积的方法沉积制作阻挡层和种子层		
	▶绝缘层: 防止 Si 导电引起的 TSV 漏电及串扰, 在孔内沉积一层二氧化硅作为绝缘层	气相沉积设备	靶材
	▶阻挡层: 防止二氧化硅与金属层互渗		
	▶种子层: 在阻挡层上沉积一层金属作为之后填孔的种子层		
深孔填充 (TSV 关键工艺)	用电镀方法在盲孔中进行铜填充	铜填充设备	电镀液及添加剂
化学机械抛光	使用化学和机械抛光 (CMP) 法去除多余的铜	CMP 抛光机	抛光液
晶圆背面减薄	打磨晶圆背部, 让电镀铜柱的另一端暴露出来形成通孔	晶圆减薄机	研磨液

来源：《TSV 关键工艺设备特点及国产化展望》，中泰证券研究所整理

- **Micro bump 是芯片倒装的基础。**
- **Bump 技术具备引脚密度高、低成本的特点，是构成倒装技术的基础。**相较于传统打线技术 (Wire Bond) 的“线连接”，Bump 技术“以点代线”，在芯片上制造 Bump，连接芯片与焊盘，此种方法拥有更高的端口密度，缩短了信号传输路径，减少了信号延迟，具备了更优良的热传导性及可靠性，也是进行 FC (Flip Chip) 倒装工艺在内的先进封装工艺的技术基础。

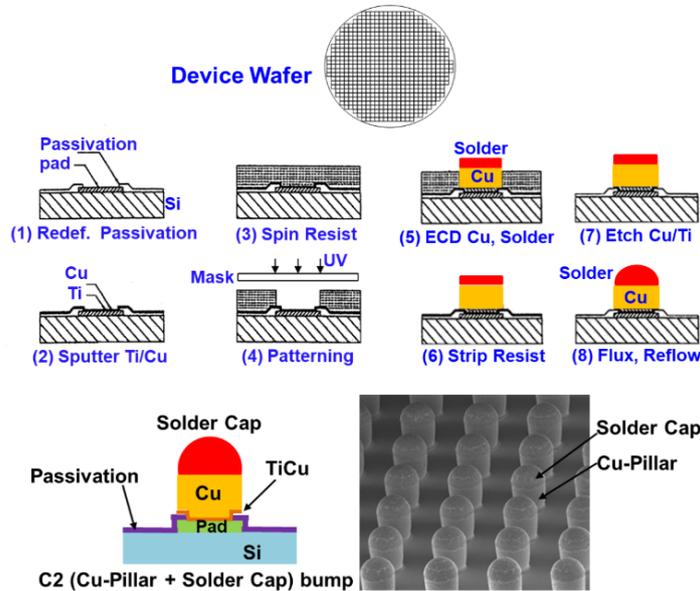
图表 48: Bump 金属凸点



来源：《先进封装与异构集成》，中泰证券研究所整理

- Micro bump** 是铜柱微凸点，主要制备方法是电镀。目前 HBM 的 DRAM 芯片之间主要通过 micro bump（微凸点）互联，micro bump 是电镀形成的铜柱凸点。凸点制作流程为：①首先溅射一层 UBM 层（Under Bump Metallization，凸点下金属层）到整个晶圆的表面，UBM 层作为种子黏附层，可以在电镀时让电流均匀传导到晶圆表面开口的地方，使各处电镀速率尽可能一致。②在 UBM 层上利用光刻胶形成掩膜，仅在需要电镀凸点的区域开口。③通常采用蘑菇头形的电镀，即电镀厚度超过光刻胶厚度，凸点沿着光刻胶表面横向长大，形成蘑菇头形状。④电镀完毕后去胶，并去除凸点外的 UBM 层。⑤最后通过回流形成大小均匀、表面光滑的凸点阵列。整个流程会涉及到的设备&材料：PVD（靶材）、涂胶显影机、光刻机（光刻胶）、电镀设备（金属、焊料）、去胶设备（剥离液）、刻蚀设备（电子特气）、回流焊设备等。

图表 49：电镀锡球凸点的工艺流程图



来源：semi engineering，中泰证券研究所整理

图表 50：bump 工艺设备与材料

工艺	作用	设备	材料
溅射 UBM 层	UBM 作为种子黏附层，可以在电镀时使电流均匀地传导到晶圆表面开口的地方，使各处电镀速率尽可能一致	气相沉积设备	靶材
光刻	在 UBM 上利用光刻胶形成掩膜，在电镀凸点区域开口	涂胶显影设备、光刻机	光刻胶
电镀	电镀需要的凸点金属层	电镀设备	金属、焊料

去胶	去除多余光刻胶及种子黏附层	去胶设备	剥离液
刻蚀	刻蚀种子黏附层	刻蚀设备	电子特气
回流焊	回流形成大小均匀、表面光滑的凸点阵列	回流焊设备	

来源：《集成电路系统级封装》，中泰证券研究所整理

■ **堆叠键合工艺主要包括：NCF、MUF、混合键合。**

■ **HBM2**, Bump pitch (凸点间距) 在 55  $\mu\text{m}$ , 三星和海力士共同使用 TCB (热压合) 技术, 其中海力士采用的是 TCB 的分支 TCB-NCF。

**HBM2/2E/3/3E**, Bump pitch 进展到 25/22  $\mu\text{m}$  水平, 三星继续采用 TCB 技术, 而海力士独家采用 MR-MUF (大规模回流焊-注塑底填充技术)。

**HBM4**, Bump pitch 进展到 20  $\mu\text{m}$  以下, 三星和海力士共同寻求混合键合技术——该技术相比 TCB、MR-MUF 技术最大特点在于, 其为直接键合, 即直接实现上下 die 之间的电气连接, 中间不需要再使用凸点。

三星、海力士之外的另一巨头美光, 此前坚持 HMC (混合存储立方体技术), 于 2022 年底转向 HBM, 并于 2023 年推出 HBM3 Gen2, 技术方式与三星相同, 使用 TCB。

图表 51: 不同代际 HBM 的 Bump 间距与互联技术

	HBM2	HBM2E/3	HBM3 (12层) /3E	HBM4
Bump pitch ( $\mu\text{m}$ )	55	25	22	<20
层数 (Hi)	4/8	4/8	8/12 (HBM3E有8层、12层版本)	12/16
海力士的内部互联封装	TCB-NCF (热压合-非导电薄膜技术)	MR-MUF (大批量回流焊-注塑底填充技术)	Advanced MR-MUF	混合键合
三星的内部互联封装	TCB (热压合)	TCB (热压合)	TCB (热压合)	混合键合

来源：DAUM、NEWSIS, 中泰证券研究所

■ **海力士独创的 MR-MUF 相较 TC-NCF 有更好的散热性能。**

海力士从 TCB 转向独创 MR-MUF, 一方面效率更高, 同时散热效果更好, HBM3 推出时间领先其他原厂 1 年多、占据了先机。

散热是 HBM 产品发展的关键瓶颈之一, MR-MUF 工艺下 HBM 的散热性能更好, 主要是由于: ①散热凸点 (bump) 更多, ②不再使用 NCF, 使用有优良导热性能的塑封料作为间隙填充材料, 相当于 TCB-NCF 下 HBM 需要穿 2 层衣服 (NCF 和 EMC), 而在 MR-MUF 工艺下 HBM 之间仅穿了一层衣服 (EMC)。另外 MR-MUF 工艺也有效率更高、降低 TSV 制造成本等优势。

凸点主要可分为散热类凸点和连接类凸点, 通常凸点越多散热效果越好, 凸点增加了散热路径, 但是在传统的 HBM 堆叠键合 TCB-NCF 工艺下, 考虑到 NCF 的流动性和键合过程中在热压力下的芯片损耗, TCB 难以大规模制备散热凸点, MR-MUF 工艺下一次熔化所有微凸点的焊料然后实现键合、电气互连, 不再使用压力, 因此可以使用更多凸点, 散热性能更佳, 另外采用自己独家研发的液体状 EMC 为主要原材料的底料填充, 散热性能更上新台阶。根据海力士, MR-MUF 工艺确保了 HBM 10 万多个凸点互连的优良质量, 增加了散热凸点数量、实现更好的散热效果, 巩固了海力士在 HBM 市场的地位, 并使 SK 海力士在 HBM3 市场占据领先地位。

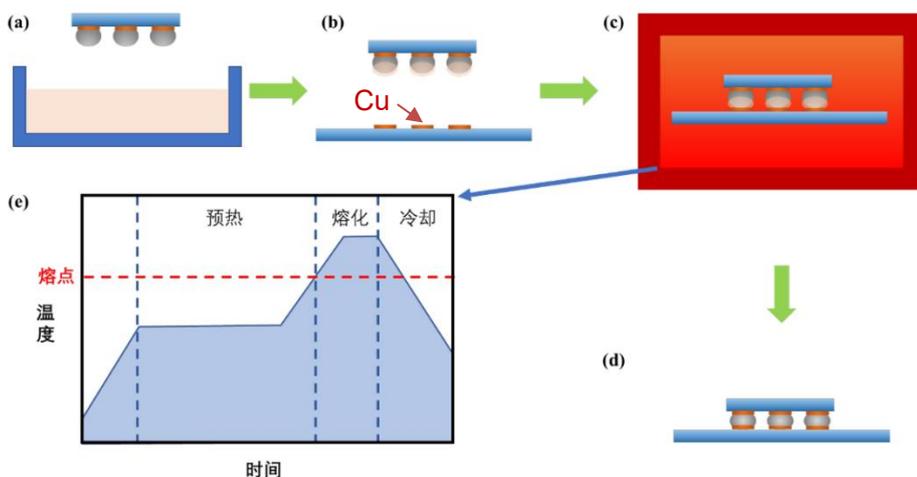
■ **目前 HBM 厂商采用的凸点倒装互连工艺主要分为回流焊和热压键合。**

1) **回流焊**: 加热锡焊料, 熔化的焊料与另一侧凸点金属接触后发生界面

反应，形成互连焊点。缺点是温度变化容易引起翘曲，焊料和金属间的对准存在偏差。**2) 热压键合：**通过加压加热使锡焊料融化与凸点金属接触后发生界面反应形成焊点，与回流焊的区别是：键合时间只有几秒钟（回流焊需要十几分钟），降低翘曲发生率；键合前通过相机对准（回流焊是自对准），精度更高，但是产出效率比回流焊低。因此热压键合更适合微尺寸的互连。

- **底填料是在倒装中起到保护凸点的作用。**在芯片倒装互连过程中，底部填充料是不可或缺的材料，起到保护凸点的作用。目前 HBM 厂商使用的底填充形式主要分为两种：**1) 组装后底部填充技术：**先凸点互联，后底部填充。该技术是传统填充技术，缺点是凸点间填充不完全；**2) 预成型底部填充技术：**先涂覆底填料至芯片，后凸点互联，凸点的互联和底填料固化工艺同时完成。该技术是新型填充技术，优点是简化工艺，填充更完全。随着芯片间 I/O 端口数量的增加，芯片间互联方式正从“回流焊+组装后底部填充”转变到“热压+预成型底部填充”，而凸点间距降低到 10 微米以下后，需要采用混合键合工艺，无需底填料。目前三星和美光的 HBM 使用的工艺是热压+预成型底部填充，而海力士的 HBM 使用的是研发改进后的回流焊+组装后底部填充。

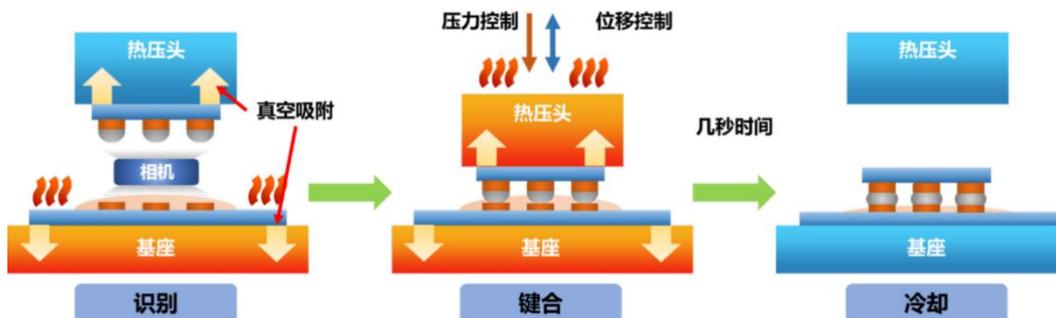
图表 52：大规模回流焊基本流程



大规模回流焊的一般流程(a)蘸取助焊剂；(b)芯片预对准；(c)回流；(d)冷却；(e)一般回流温度变化曲线

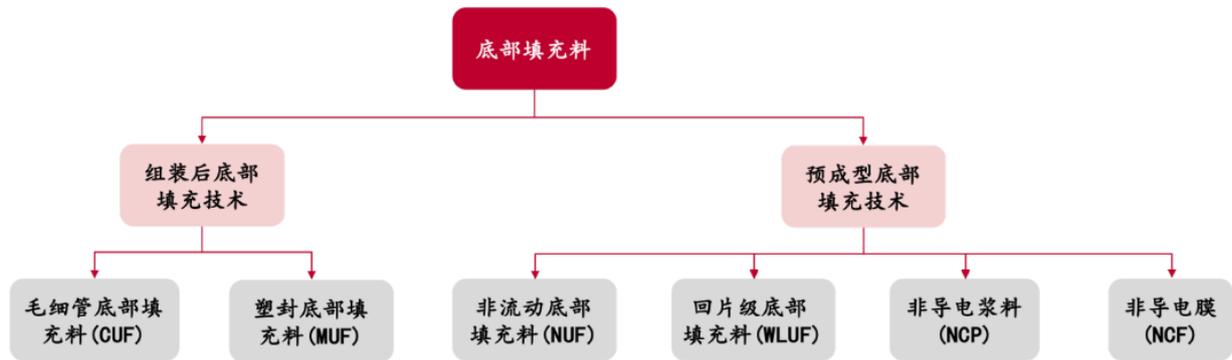
来源：《热压键合微凸点形状与微观组织调控及抗冲击性能研究》，中泰证券研究所整理

图表 53：热压键合基本流程



来源：《热压键合微凸点形状与微观组织调控及抗冲击性能研究》，中泰证券研究所整理

图表 54: 底填料分类

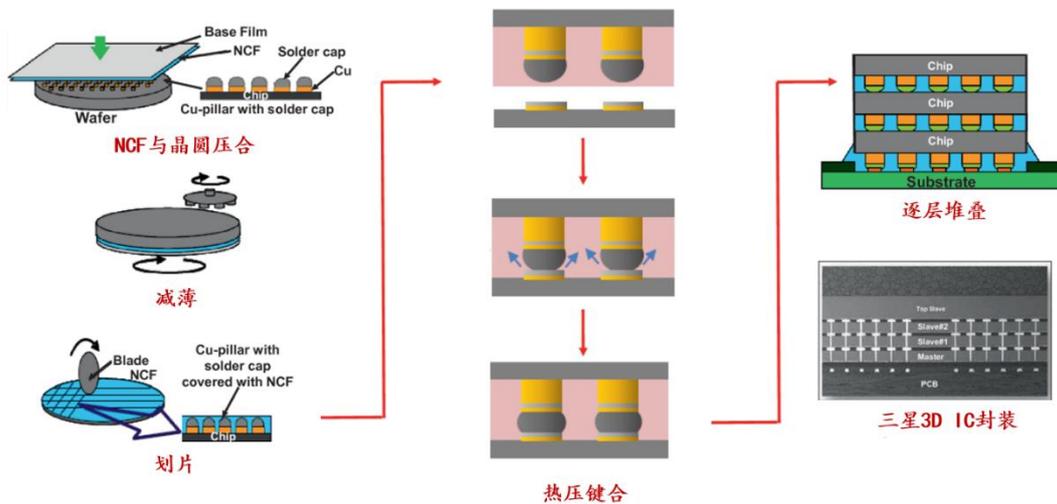


来源:《集成电路先进封装材料》, 中泰证券研究所整理

**TC-NCF**

- TC-NCF (热压键合+非导电薄膜) 工艺先用 NCF 非导电薄膜填充芯片间隙, 再通过热压键合连接芯片。TC-NCF 的工艺流程: 1) 在圆片正面真空层压 NCF。NCF 由丙烯酸和环氧树脂等组成, 是一种底填料, 用于粘合填充。2) 在圆片背面贴划片膜, 接着进行划片分割; 3) 通过热压键合将芯片之间堆叠固定。TC-NCF 具有成本低、操作方便等优点, 缺点是高温易导致芯片翘曲, 影响良率。此外, 它对芯片研磨的要求也很高, 厚度稍有不均, 芯片各部分受到的压力就会变化, 使良率降低。美光和三星从生产 HBM 开始, 一直使用 TC-NCF 工艺, 海力士的 HBM2 使用了 TC-NCF 工艺。

图表 55: TC-NCF 工艺流程图



来源:《Lau, J.H. (2021). System-in-Package (SiP). In: Semiconductor Advanced Packaging》, 中泰证券研究所整理

图表 57: TC-NCF 步骤及设备

步骤	设备
填 NCF	点胶机
背面贴膜	贴膜机
背面减薄	减薄机
揭膜	揭膜机

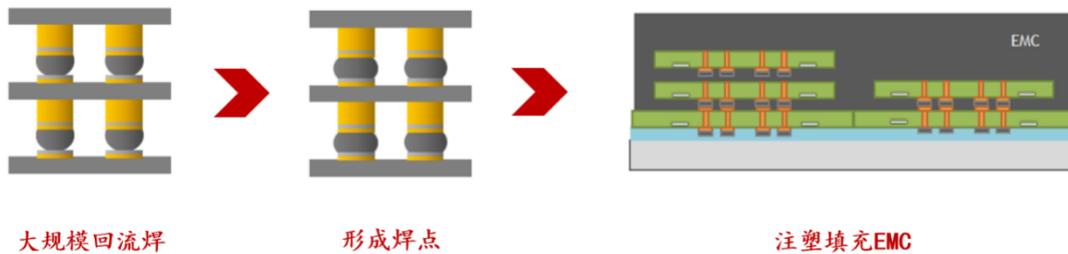
切片	晶圆划片机
热压键合	热压键合机
NCF 固化	热压键合机

来源: Lau, J. H. (2021). System-in-Package (SiP). In: Semiconductor Advanced Packaging, 中泰证券研究所整理

### MR-MUF

- **MR-MUF 是海力士 HBM 核心工艺, MR-MUF (大规模回流焊+注塑底填充) 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。**海力士的 HBM 在市场领先, 依靠的核心技术就是独家的 MR-MUF。MR-MUF 技术壁垒为: 液体环氧塑封料及注塑设备、芯片翘曲控制技术, 均由海力士研发并享有独占权。MR-MUF 流程分为两步骤: 1) 将带有微凸点的芯片堆叠后整体加热, 一次熔化所有微凸点的焊料, 将芯片与电路连接; 2) 用 MUF (主要构成材料是液体环氧塑封料) 填充芯片与芯片的间隙, 同时完成注塑和底填工艺。与 TC-NCF 相比, MR-MUF 的优点: 1) MR-MUF 的键合可以在空隙阶段完成, 提高工艺效率; 2) MUF 具有高导热性, 导热率比 NCF 高出约两倍, 散热性能改善了 10℃ 以上。海力士 HBM2E、HBM3、HBM3E 均使用 MR-MUF 工艺。

图表 57: MR-MUF 工艺流程图



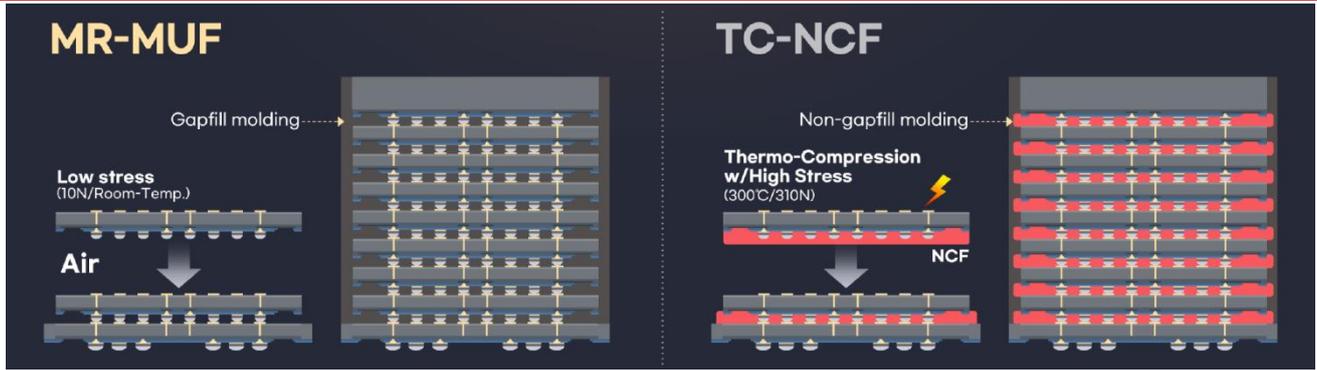
来源: 海力士, 中泰证券研究所整理

图表 58: MR-MUF 步骤及设备

步骤	设备
Bump 沾上助焊剂、对准、接触	晶圆级回流焊键合机
回流焊	晶圆级回流焊键合机
清洗助焊剂	热压键合机、清洗机
模塑底部填充	MUF Molding 设备

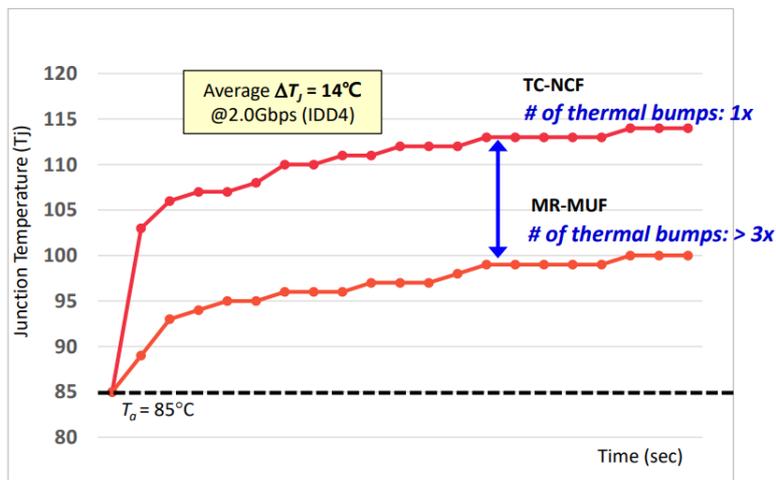
来源: 《热压键合微凸点形状与微观组织调控及抗冲击性能研究》, 中泰证券研究所整理

图表 59: MR-MUF 与 TC-NCF 工艺对比



来源：海力士官网，中泰证券研究所

图表 60: MR-MUF 具备更高的热导性



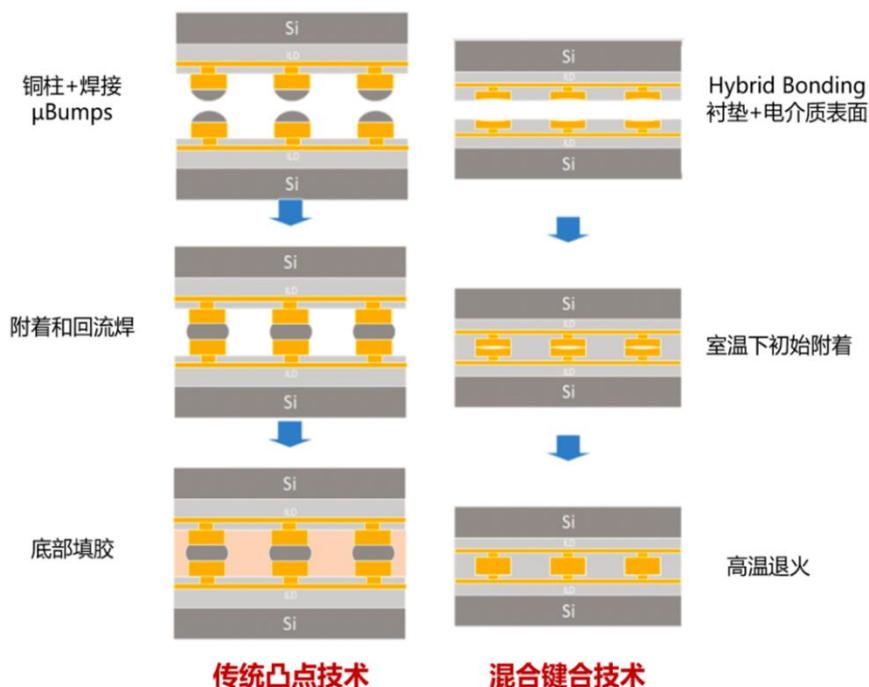
Test condition: Pin speed 2.0Gbps, VDD=1.26V, IDD4, Full Channel

来源：海力士官网，中泰证券研究所整理

### 混合键合

- 三星与 SK 海力士正在研发 HBM4 将使用混合键合技术。混合键合 (hybrid bondind) 是一种将介电键 (SiOx) 与嵌入金属 (Cu) 结合形成互连的工艺技术。热压键合或回流焊工艺通常最小只能做到 10μm 节距，对于细间距微凸点，电镀时凸点之间极小的不均匀也会影响良率和性能，因此 10μm 节距以下只能采用混合键合技术。混合键合与传统凸点焊接不同，其没有凸点，不需要底填料，是一种将介电键 (SiOx) 与嵌入金属 (Cu) 结合形成互连的工艺技术，可以提高芯片间通信速度，节距可以达到 10μm 及以下，未来有望升级至 2μm 及以下，是未来应用于高带宽存储 (HBM) 的理想键合方案，三星与 SK 海力士正在研发 HBM4 预计将使用混合键合技术。

图表 61: 传统凸点技术与混合键合技术流程对比



来源：公众号“SiP与先进封装技术”，中泰证券研究所整理

### 3. HBM 驱动先进封装设备和材料需求爆发

#### 3.1 材料端：环氧塑封料、硅微粉、电镀液和前驱体等用量提升

- 材料端：相较传统 DRAM，HBM 多了 TSV、Microbumping 和堆叠键合等工艺，既使用晶圆制造材料，如光刻胶、靶材等，也使用先进封装材料。
- 在先进封装材料环节的主要增量为：1) 环氧塑封料：在 MR-MUF 方法中，环氧塑封料具备底填料+注塑料的功能，在 TC-NCF 方法中，环氧塑封料具备注塑料的功能；2) 硅微粉：是 HBM 中环氧塑封料最主要原材料，填充比例为 70%-90%，其成本占环氧塑封料原材料成本的 27% 左右。3) 电镀液：TSV 和 bump 工艺在制作过程中，均离不开电镀液，TSV 通孔互联由电镀铜组成，bump 的铜柱由电镀铜组成，此外 bump 中还有电镀锡银等；4) 前驱体：在 TSV 电镀铜前，需要 ALD 沉积形成扩散阻挡层，前驱体是 ALD 过程的一种重要介质。

#### 环氧塑封料

- MR-MUF 的底填+注塑材料为环氧塑封料。环氧塑封料 (Epoxy Molding Compound, 简称 EMC) 是用于半导体封装的一种热固性化学材料。EMC 是由环氧树脂为基体树脂，以高性能酚醛树脂为固化剂，加入硅微粉等填料以及多种助剂加工而成，保护芯片不受外界环境（水汽、温度）的影响，并实现导热、绝缘、耐压等复合功能。在海力士 HBM 的生产中，液体环氧塑封料被用于大规模回流模制底部填充(MR-MUF)封装方法，填充并附着芯片之间的空间，具备底填料+注塑料的功能。目前海外供应商主要有日本住友、日本昭和。根据共研网数据，2022 年我国半导体用环氧塑封料市场规模约 85 亿元，2018-2022 年 CAGR 约 20%。

图表 62: EMC 在 HBM 中的应用

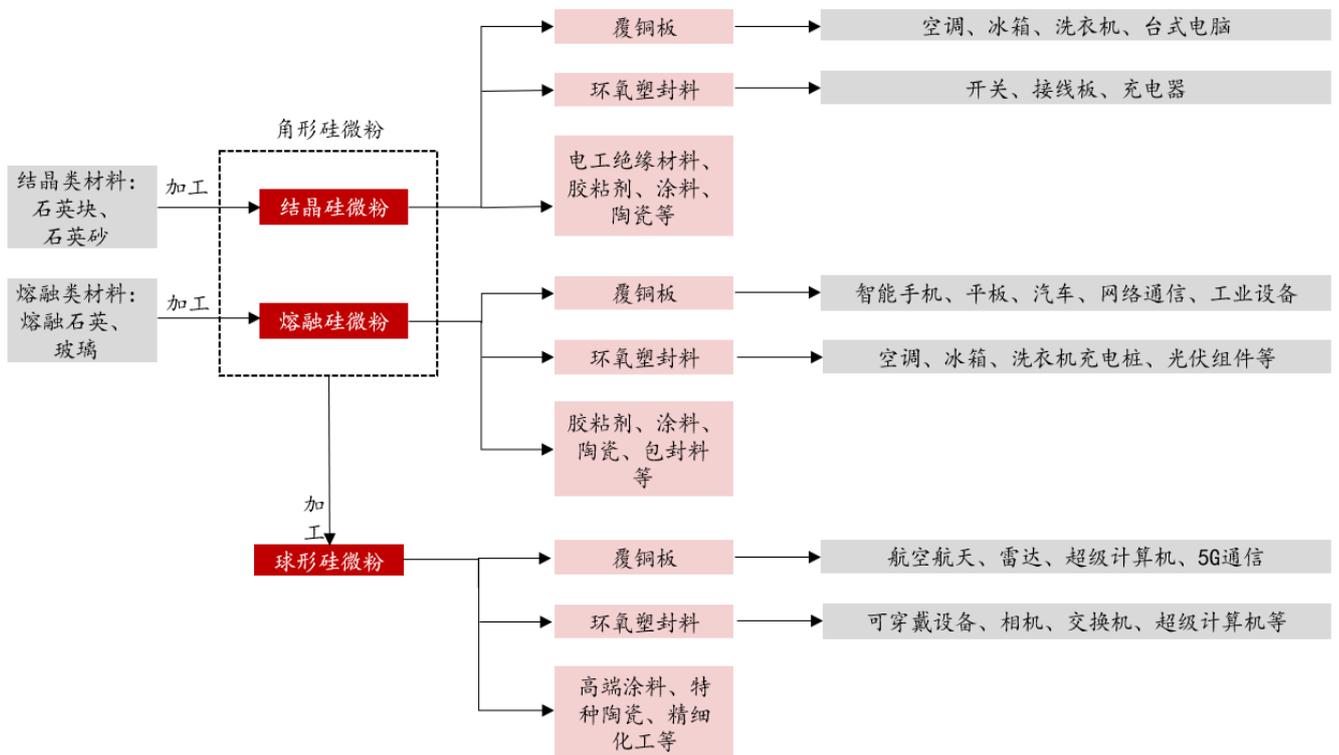


来源: SKhynix, 中泰证券研究所整理

**硅微粉**

- **硅微粉为环氧塑封料主要成分。**硅微粉是以结晶石英、熔融石英等为原料加工而成的二氧化硅粉体材料,是覆铜板、环氧塑封料、电工绝缘材料、胶粘剂等材料的原材料之一。硅微粉分为不同的类型,1)按照结晶特点:结晶、熔融;2)按照颗粒形态:角形、球形。其中球形硅微粉是HBM中环氧塑封料最主要原材料,填充比例为70%-90%,可提高环氧塑封料的硬度、导热系数并减缓震动。目前海外厂商主要有日本电气化学、隆森。根据联瑞新材招股书数据,预计2025年我国环氧塑封料用硅微粉的市场规模约9亿元。

图表 63: 硅微粉的应用



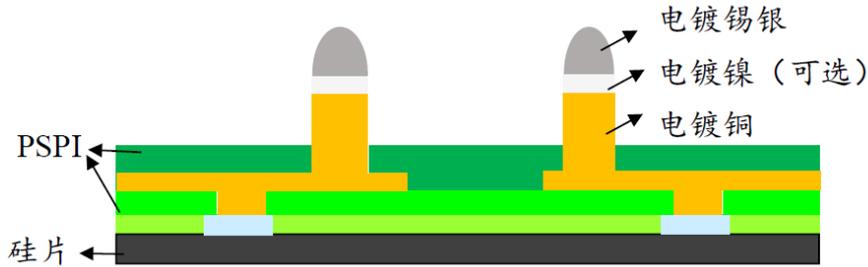
来源: 联瑞新材公司公告, 中泰证券研究所整理

**电镀液**

- **TSV、bump 工艺均需使用电镀液。**电镀是指在芯片制造或先进封装中,将金属离子电镀到晶圆表面形成互连,所需电镀液的种类有铜、镍、金和锡银合金等,其中铜电镀液占据主流,占比在60%以上。电镀液通常由主盐、导电盐、络合剂、各类添加剂及溶剂等构成,其中电镀添加剂

是影响电镀功能的核心组分，配方体系复杂、研发难度大。如前文所述，TSV 和 bump 工艺在制作过程中，均离不开电镀液，TSV 通孔互联由电镀铜组成，bump 的铜柱由电镀铜组成，此外 bump 中还有电镀锡银等。目前海外厂商主要有石原、乐思化学、杜邦。根据 QY Research 数据，2022 年全球高纯电镀液市场规模达到了 5.9 亿美元，预计 2029 年将达到 12.0 亿美元，2022-2029 年 CAGR 为 10.8%。

图表 64：电镀液在 bump 中的应用

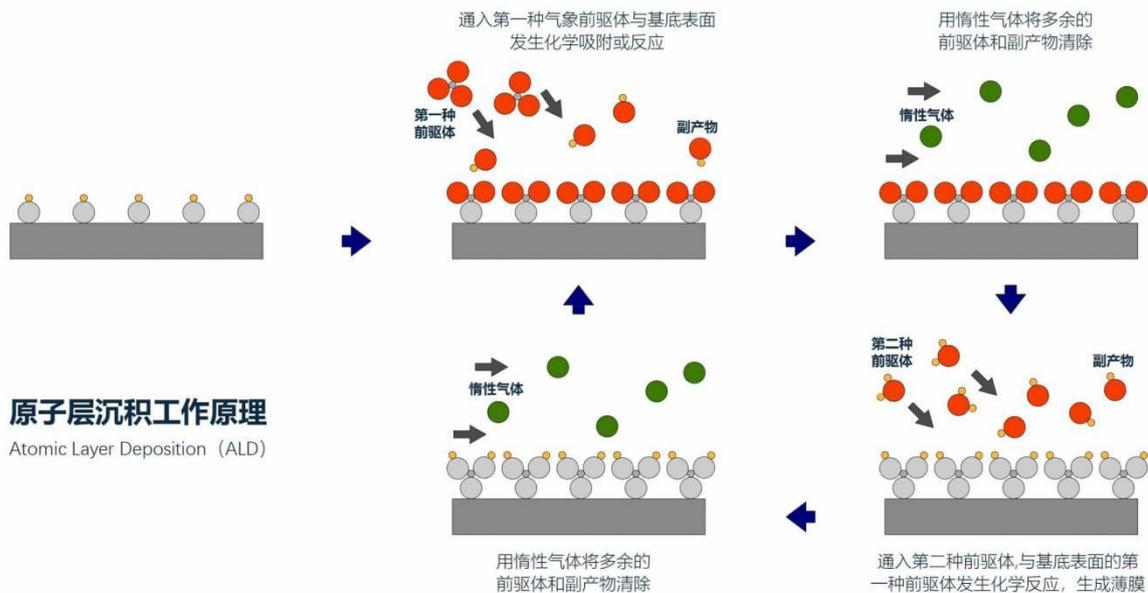


来源：艾森股份招股说明书，中泰证券研究所整理

### 前驱体

- **HBM 多层堆叠带来前驱体用量的加倍提升。**前驱体是原子层薄膜沉积 ALD 使用的一种重要介质。原子层沉积是 TSV 中关键工艺之一，在电镀铜前，需要 ALD 沉积形成扩散阻挡层。ALD 是指将多个化学前驱体交替脉冲通入反应腔体与衬底接触，发生化学吸附反应，形成扩散阻挡层，防止铜的电化学迁移导致物理失效。未来 HBM 将由 8 层堆叠发展到 12 层、16 层堆叠，HBM 的多层 DRAM 堆叠会带来前驱体用量的成倍提升。主要海外厂商有：德国 Merck、法国液化空气。从市场规模来看，2021 年中国半导体用前驱体市场规模达 5.9 亿美元。

图表 65：前驱体在原子层沉积中的应用



原子层沉积工作原理  
Atomic Layer Deposition (ALD)

来源：AccSci，中泰证券研究所整理

### 3.2 设备端：热压键合机、大规模回流焊机、混合键合机等需求

- **设备端：HBM 带来热压键合机、大规模回流焊机、混合键合机的需求。**  
HBM 的不同代际，通常采用不同的 DRAM die 的键合工艺。在海力士技术路线中，除 HBM2 采用 TCB-NCF（热压合-非导电薄膜技术）外，后续 HBM2E/3/3E 均采用自家独创的 MR-MUF（大批量回流焊-注塑底填充技术），而三星则从 HBM2 至 HBM3E 均采用 TCB（热压合技术）。TCB 工艺采用热压键合方式，实现 DRAM die 的上下互联，其核心设备为 TCB 键合机；MR-MUF 采用回流焊实现键合，其核心键合设备为大规模回流焊机，其他核心设备为实现 MUF 工艺的注塑底填充机。而在未来的 HBM4 时代，随着 DRAM die 的 pitch 间距进一步缩小，混合键合有望成为实现 DRAM die 键合的主流工艺，则届时对应的核心设备为混合键合机。

**图表 66：三种不同的 HBM 工艺对应三类不同的核心键合设备**

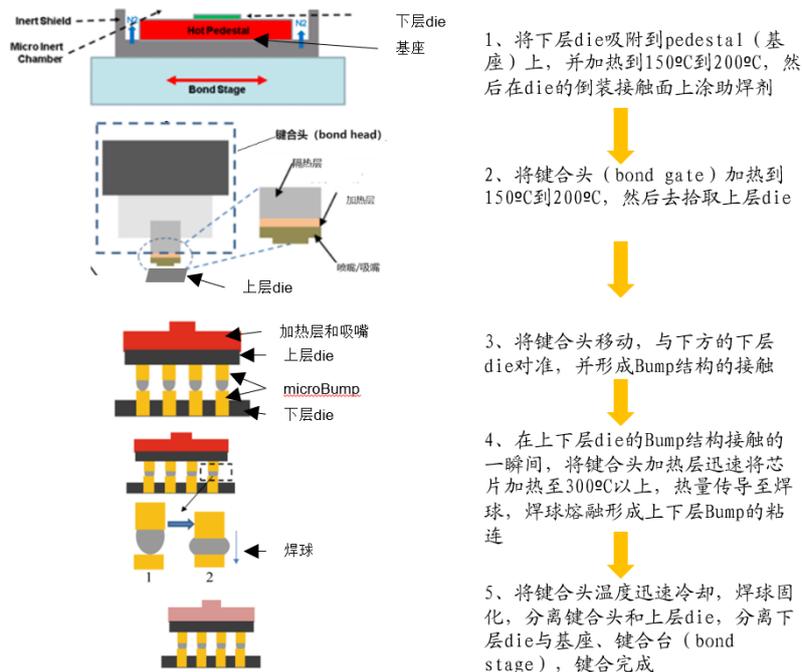
工艺	适用 HBM 代际	相关厂商	DRAM die 的键合工艺	核心键合设备
TCB	HBM2/2E/3/3E	海力士、三星、美光	热压合	热压键合机
MR-MUF	HBM2E/3/3E	海力士	大规模回流焊	大规模回流焊机
混合键合	HBM4 及以上	海力士、三星、美光	混合键合	混合键合机

来源：海力士官网，三星官网，RobbieXChiplet, SW Test Workshop, 中泰证券研究所整理

**TCB-热压键合机**

- **热压键合机通过热压合实现上下层芯片之间的键合。**在具体执行键合的部分，热压键合机一方面将下层芯片放置在基座上，另一方面利用可加热的键合头吸取上层芯片，在上下层芯片的 bump 结构对准并接触后，键合头瞬间将芯片加热到 300℃ 以上并将热量传导至 bump 的焊球，焊球高温熔融后形成上下 bump 的粘连，之后键合头迅速冷却、焊球固化，键合头脱离上层芯片、下层芯片从基座脱离，键合完成。

**图表 67：热压键合机的“热压”过程**



来源: Sangil Lee 《Fundamentals of Thermal Compression Bonding Technology and Process Materials for 2.5/3D Packages》, 中泰证券研究所整理

- 2022 年热压键合机全球市场空间超 12 亿美元，市场由海外厂商垄断，国产厂商正积极布局。**据第三方机构 MarketsGlob，2022 年全球热压键合机市场空间为 12.05 亿美元，至 2030 年有望达 17.80 亿美元，2022-2030 年 CAGR 为 5.7%。该市场主要由海外厂商垄断，重要玩家有 ASMPT、K&S、BESI、Shibaura、Hamni 和 SET，前五大玩家合计市占率高达 88%。国产厂商亦积极布局该领域，如华卓精科在 2022 年招股书中提到其储备了面向 CIS/MEMS/功率器件/射频器件的热压键合设备相关技术，此外，国产厂商唐人制造、新加坡厂商华封科技也涉足热压键合机领域。

**图表 68：热压键合机海内外主要厂商简介**

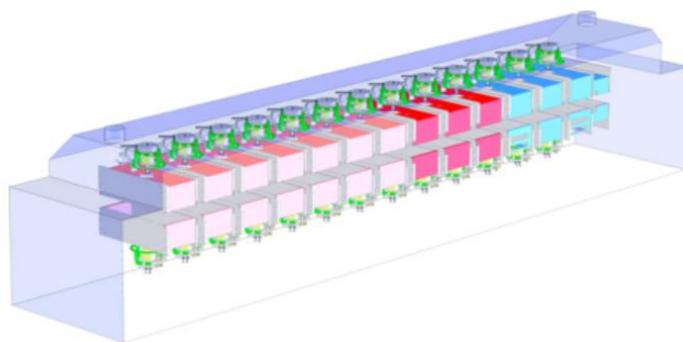
厂商	总部所在地	股票代码	营收 (亿美元)	主营产品	热压键合机发展情况
ASMPT	新加坡	ASMPT	25 (2022 年)	热压键合机、大规模回流焊炉、混合键合机、SMT 贴片机、光电设备	在公司先进封装业务中贡献最大收入和订单的品类，2023 年公司在 TCB 设备上的累计订单超过 1 亿美元，逻辑客户主要为领先的晶圆厂、OSAT 和 IDM 客户，并继续与存储客户合作开发针对下一代 HBM 的相关设备
K&S	美国	KLIC.O	7.4 (2023 财年)	点胶机、芯片传送机、倒装机、TCB、其他各类型键合机	处于市场领先地位，pitch size 可以达到 10 $\mu$ m 以下
Besi	荷兰	BESI.AS	7.7 (2022 年)	键合设备 (79%营收占比，涵盖多模组封装、倒装、嵌入式封装、TCB、混合键合等)、封装&电镀设备 (21%营收占比)	拥有 8800TC (针对 Chip to Substrate)、9800TC (针对 Chip to Wafer) 两大系列
Shibaura	日本	6104.T	9.3 (2022 年)	注塑机 (营收占比近 70%)、机床 (营收占比 22%)、控制技术 (营收占比 7%)、其他 (占比 1%)	全球 TCB 主要玩家之一
Hanmi	韩国	-	-	键合机、切片机、研磨机等	拥有 DRAGON、GRIFFIN、3.0CW (Chip to Wafer)、3.0CS (Chip to Substrate) 四大 TCB 产品系列
SET	前法国公司，后被瑞典 Replisaurus 收购	-	-	倒装键合机、纳米压印光刻机、电动压力机、大气等离子体表面处理系统	ACC $\mu$ RA M、ACC $\mu$ RA Plus、FC150 PLATINUM、FC300 这些型号具备热压合功能
华卓精科	中国	-	0.23 (2020 年)	晶圆键合设备、IGBT 激光退火设备、超精密运动台、隔振产品	在面向 CIS、MEMS、功率器件、射频器件封装等晶圆级的热压键合设备上具有技术储备
华封科技	新加坡	-	-	倒装贴片机、晶圆级贴片机、POP 封装机、层叠半贴片机、面板级贴片机、多晶片贴片机等	从贴片机出发，布局 TCB
唐人制造	中国	-	-	前期主要开发 Fan-Out 高精度模组装片设备，后续开发高速倒装装片设备	专注扇出型、倒装的装片机，布局 TCB

来源：各家公司官网，中泰证券研究所整理

### MR-MUF-大规模回流焊机

- **大规模回流焊机通过回流焊实现上下芯片之间的键合。**大规模回流焊炉，可实现将多颗芯片放入炉中一道加热、通过融化焊料实现上下芯片之间互联的效果。该技术由 2016 年海力士率先采用，将 4 块 50 μm 厚的芯片相互堆叠。搭配高导热性的模制底部填充材料，海力士实现了比竞争产品更强的散热性能，MR-MUR 技术巩固了海力士 HBM 的市场地位。

图表 69: Kurtz Ersa HOTFLOW 型号回流炉示意图



来源：Kurtz Ersa 官网，中泰证券研究所整理

- **2022 年全球回流焊炉市场空间接近 3.7 亿美元，主要市场仍由海外厂商垄断。**据 GIR Global Info Research 调研，2022 年全球 PCB 与半导体用回流焊炉收入约 3.68 亿美元，预计 2029 年达到 4.83 亿美元，2023-2029 年 CAGR 为 3.9%。该市场主要玩家有 Rehm Thermal Systems、Kurtz Ersa 等，国产厂商中涉足真空回流炉的有劲拓股份、中科同志、嘉昊先进等。

图表 70: 回流焊炉海内外主要厂商简介

厂商	总部所在地	股票代码	营收 (亿美元)	主营产品	回流焊机发展情况
Rehm Thermal Systems	德国	-	-	产品覆盖回流焊接、气相焊接、干燥和防护层喷涂系统以及与焊接、涂装和固化相关的各类定制系统	具备真空/非真空/气相三类回流焊技术，为国内真空回流炉的主力品牌之一
Kurtz Ersa	德国	-	-	印刷锡膏机、回流焊接设备、波峰焊接设备、选择性焊接设备、返修工艺设备	该公司的真空回流焊技术能将焊点空洞率降低 99%
劲拓股份	中国	300400.SZ	1.14 (2022 年)	波峰焊、回流焊、智能机器视觉类产品、高速点胶机以及智能全向重载移动平台等	公司半导体热工设备可以广泛应用于各类芯片的封装回流工艺、晶圆级植球焊接工艺等
中科同志	中国	-	-	贴片机、回流焊、波峰焊、丝印机、点胶机等	具备真空/非真空/气相三类回流焊技术，形成了对应的三大回流焊产品系列

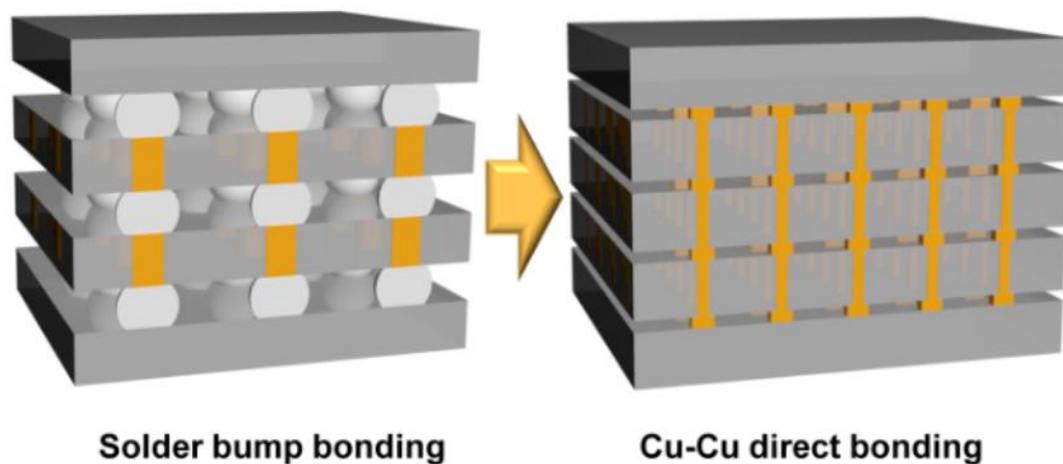
嘉昊先进	中国	-	-	半导体新能源功率模块真空回流焊接设备	22 年底研发出国内首台真空回流焊接炉设备，可在氮气环境中进行焊接，推动气泡率降至 0.5% 以下
------	----	---	---	--------------------	---

来源：各家公司官网，中泰证券研究所整理

### 混合键合-混合键合机

- **混合键合机是实现混合键合工艺的设备，针对连接点间距进一步缩小的需求而生。**当芯片 Bump 间距缩小至  $10\mu\text{m}$  时，焊球减小会增加金属间化合物 (IMC) 形成的风险，导致导电性和机械性能的下降。混合键合机避免完全使用 Bump，之间实现铜对铜的互连，针对更小尺寸的键合提供了新的解决方案。其具体技术原理是，打磨 Cu/SiO<sub>2</sub> 的表面至极度光滑，不同表面间通过范德华力连接，稍微施加压力和高温就可以实现永久键合，其本质是将介电键合 (SiO<sub>x</sub>) 与嵌入式金属 (Cu) 结合起来形成互连形成电介质-电介质和金属-金属键。根据具体工艺过程不同，混合键合机又可以分为 CoW (chip on wafer) 键合机、WoW (Wafer on wafer) 键合机。

图表 71：焊球焊接键合与铜铜键合示意图



来源：RIab，中泰证券研究所整理

- **2020 年全球混合键合机市场达 3.2 亿美元，至 2027 年有望达 7.4 亿美元，增速快。**根据 Yole 数据，2020 年全球混合键合机市场达 3.2 亿美元，其中 CoW 键合机为 0.06 亿美元，WoW 键合机为 2.6 亿美元，至 2027 年二者市场空间分别有望增至 2.3、5.1 亿美元，2020-27 年 CAGR 分别为 69% 和 16%。高增背后的主要驱动力来自于 AI 需求爆发，对先进封装技术尤其是混合键合技术的需求随之爆发。
- **当前混合键合机市场由海外厂商如 Besi、ASMPT、SUSS 等厂商垄断，国产厂商拓荆科技亦积极布局。**全球范围内，较知名的混合键合机厂商有 Besi、EVG、ASMPT、SUSS、TEL 等。鉴于大陆封装工艺现状，当前大陆对混合键合机需求较少，国产厂商拓荆科技 (WoW)、华卓精科 (待上市，WoW)、艾科瑞思 (未上市，CoW)、华封科技 (未上市) 等

已开展前瞻布局。

图表 72：混合键合机海内外厂商简介

厂商	总部所在地	股票代码	营收 (亿美元)	主营产品	混合键合机发展情况
Besi	荷兰	BESI.AS	7.7 (2022 年)	全套的芯片封装键合设备、电镀设备	全球混合键合设备龙头，2021 年即获得台积电、英特尔混合键合设备订单，2023 年产能达 180 台
EVG	奥地利	-	-	光刻机、纳米压印机、键合机、量测设备	EVG 在 WoW 键合机市场处于领先地位，存量设备有数百台，在高端 CIS 混合键合领域与日本 TEL 竞争
ASMPT	新加坡	ASMPT	25 (2022 年)	热压键合机、大规模回流焊炉、混合键合机、SMT 贴片机、光电设备	2021 年与 EVG 联合开发 CoW 混合键合机 LithoBolt™
SUSS	德国	-	-	涂胶显影机、喷墨打印机、键合机、光掩模设备、纳米压印机等	研制的 XBS300 可以通过高度模块化设计实现混合键合的低成本灵活生产，支持 D2W 和 W2W 混合键合，重点关注 3D 存储或 3D SOC 等高端应用
TEL	日本	8035.T	166 (2023 财年)	泛半导体涂胶显影机、刻蚀机、热处理等	坚持 WoW 路线，
华卓精科	中国	-	0.23 (2020 年)	晶圆键合设备、IGBT 激光退火设备、超精密运动台、隔振产品	为以超精密测控技术为基础，开发了 HBS 系列全自动晶圆混合键合系统，可实现室温直接键合
拓荆科技	中国	688072.SH	2.45 (2022 年)	薄膜沉积设备、混合键合机	WoW 产品已通过客户验收并获得重复订单

来源：各家公司官网，未来半导体公众号，中泰证券研究所整理

### 3.3 全球/大陆产业链布局情况

图表 73: HBM 设备/材料产业链

设备/ 材料	具体种类	TCB- NUF	MR- MUF	混合键合- W2W	混合键合- -D2W	主要厂商
	热压键合机	✓				ASM Pacific、Kulicke & Soffa、Besi、Shibaura、韩美半导体、SET
	点胶机	✓				Nordson、NSW Automation、H&S Manufacturing、Graco
	晶圆贴膜机	✓				Dynatech、Toyo Adtec、Takatori、Teikoku Taping System、
	晶圆划片机	✓			✓	OHMIYA INDUSTRY COMPANY、EV Group、Nitto
	揭膜机	✓				DISCO、东京精密、ACCRETECH
	晶圆级真空回流焊键 合机		✓			东洋技术（中国台湾）、至圣集团（中国台湾）、Takatori、Nitto、 LINTEC Corporation
	清洗机		✓		✓	Rehm Thermal Systems、Kurtz Ersas、HIRATA、Heller Industries、 BTU International、INVACU、SMT
	Molding 设备（塑封切 筋成型设备）		✓			日本 DNS（迪恩士）、东京电子、泛林科技和 SEMES（韩国细美 事）
设备	CVD			✓	✓	TOWA、Besi、ASM Pacific、Yamada、TOWA
	刻蚀机			✓	✓	AMAT、泛林半导体、东京电子、先晶半导体
	铜互连 PVD			✓	✓	泛林半导体、东京电子、应用材料、日立高新 Advanced Energy Industries、Angstrom Engineering、Veeco、
	铜电镀设备			✓	✓	AMAT、Platit AG
	CMP			✓	✓	ACM、AMAT、ASMPT、ClassOne Technology、EBARA、 Hitachi、泛林半导体、Besi
	混合键合机			✓	✓	AMAT、日本荏原、韩国 KC Tech、华海清科
	退火设备			✓	✓	D2W: Besi、EVG、ASMPT、SET、Shibaura、SUSS W2W: 华卓清科、拓荆科技
	减薄机			✓		AMAT、Kokusai、东京电子、屹唐、北方华创
	临时键合机/解键合机			✓		DISCO、东京精密、KOYOSEIKO、G&N
	测试机	✓	✓	✓	✓	TEL、EVG、SUSS、上海微、TAZMO、Tok、ERS、EO Technics、Takatori
	光刻胶	✓	✓	✓	✓	爱德万、泰瑞达、精智达、悦芯、长川科技、华峰测控
	掩膜版	✓	✓	✓	✓	JSR、东京应化、信越、杜邦、彤程新材、徐州博康、晶瑞电材、上 海新阳、南大光电、鼎龙股份
	湿电子化学品	✓	✓	✓	✓	福尼克斯、Toppan、路维光电、清溢光电
	前驱体	✓	✓	✓	✓	巴斯夫、霍尼韦尔、住友化学、和光纯药工业
材料	环氧塑封材料	✓	✓	✓	✓	德国默克、法国液空、SouthBrain
	Low- $\alpha$ 球铝	✓	✓	✓	✓	住友电木、蔼司蒂、京瓷、华海诚科、衡所华威
	底填胶	✓				日本电气化学、隆森、admatechs、联瑞新材
	电镀液	✓	✓	✓	✓	汉高、Namics、Delo、Nagase、德邦科技
	临时键合胶			✓		石原、乐思化学、杜邦、强力新材、上海新阳 Brewer Science、Daxin Materials

来源：华海诚科招股书，华经产业研究院，锐观咨询等，中泰证券研究所整理

- **HBM 芯片需求的提升，上游材料厂商有望受益。**目前大陆已有部分材料厂商进入 HBM 产业链，华海诚科的颗粒状环氧塑封料 (GMC) 可以用于 HBM 的封装，相关产品已通过客户验证，现处于送样阶段。雅克科技的前驱体材料已供应三星、海力士等海外龙头厂商。此外，强力新材、艾森股份、天承科技、上海新阳均布局了先进封装电镀液，德邦科技布局了底填胶，飞凯材料布局了环氧塑封料，联瑞新材、壹石通布局了硅微粉。

**图表 74：大陆 HBM/先进封装材料产业链公司**

公司名称	相关业务
华海诚科	公司环氧塑封料 (GMC) 可以用于 HBM 的封装
德邦科技	公司芯片级 underfill 已有型号通过国内部分客户验证，未来能否应用在 HBM 取决于产品性能的匹配、客户供应链的选择等多种因素
飞凯材料	公司 MUF 产品包括液体封装材料 LMC 和 GMC，前者已量产并形成少量销售，后者尚处于研发送样阶段
雅克科技	半导体前驱体材料主要应用在半导体集成电路存储、逻辑芯片制造的薄膜沉积工艺中
唯特偶	微电子材料可以用于 HBM 芯片的堆叠和高速串行的连接
联瑞新材	公司高性能球硅微粉已用于 chiplet 芯片封装用封装材料
壹石通	公司 Low- $\alpha$ 球形氧化铝和 Low- $\alpha$ 球形二氧化硅成为高性能环氧塑封填料的主要选择
强力新材	公司研发生产的光敏性聚酰亚胺 (PSPI) 处于下游客户验证阶段
艾森股份	公司部分产品 (光刻胶、先进封装用电镀液等) 通过认证后还需要经过小批量供应阶段
天承科技	公司成功研发出适用于封装载板 SAP 工艺的沉铜、电镀等专用化学品
上海新阳	公司电镀液及添加剂产品可用于存储器芯片制造企业

来源：wind，深交所互易通等，中泰证券研究所整理

## 4. 投资建议

- HBM 由海外引领，主要核心标的如下：1) 存储原厂：海力士/三星/美光；2) 设备：BESI/ASMPT/Camtek 等。
- 大陆 HBM 产业链相关标的：1) 存储：香农芯创/佰维存储/雅创电子等；2) 设备：赛腾股份/精智达/新益昌等；3) 材料：华海诚科/雅克科技/联瑞新材/兴森科技/深南电路等；4) 封测：通富微电/深科技/长电科技等。

## 5. 风险提示

- 行业需求不及预期的风险；
- 大陆厂商技术进步不及预期；
- HBM 技术路线发生分歧；
- 研报使用的信息更新不及时的风险；
- 计算结果存在与实际情况偏差的风险。

**投资评级说明:**

	评级	说明
股票评级	买入	预期未来 6~12 个月内相对同期基准指数涨幅在 15%以上
	增持	预期未来 6~12 个月内相对同期基准指数涨幅在 5%~15%之间
	持有	预期未来 6~12 个月内相对同期基准指数涨幅在-10%~+5%之间
	减持	预期未来 6~12 个月内相对同期基准指数跌幅在 10%以上
行业评级	增持	预期未来 6~12 个月内对同期基准指数涨幅在 10%以上
	中性	预期未来 6~12 个月内对同期基准指数涨幅在-10%~+10%之间
	减持	预期未来 6~12 个月内对同期基准指数跌幅在 10%以上
备注：评级标准为报告发布日后的 6~12 个月内公司股价（或行业指数）相对同期基准指数的相对市场表现。其中 A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以摩根士丹利中国指数为基准，美股市场以标普 500 指数或纳斯达克综合指数为基准（另有说明的除外）。		

## 重要声明:

中泰证券股份有限公司（以下简称“本公司”）具有中国证券监督管理委员会许可的证券投资咨询业务资格。本公司不会因接收人收到本报告而视其为客户。

本报告基于本公司及其研究人员认为可信的公开资料或实地调研资料，反映了作者的研究观点，力求独立、客观和公正，结论不受任何第三方的授意或影响。本公司力求但不保证这些信息的准确性和完整性，且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，可能会随时调整。本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本报告所载的资料、工具、意见、信息及推测只提供给客户作参考之用，不构成任何投资、法律、会计或税务的最终操作建议，本公司不就报告中的内容对最终操作建议做出任何担保。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。

市场有风险，投资需谨慎。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

投资者应注意，在法律允许的情况下，本公司及其本公司的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供投资银行、财务顾问和金融产品等各种金融服务。本公司及其本公司的关联机构或个人可能在本报告公开发布之前已经使用或了解其中的信息。

本报告版权归“中泰证券股份有限公司”所有。事先未经本公司书面授权，任何机构和个人，不得对本报告进行任何形式的翻版、发布、复制、转载、刊登、篡改，且不得对本报告进行有悖原意的删节或修改。