

AI服务器催化HBM需求爆发，核心工艺变化带来供给端增量

—算力系列报告（一）

太平洋证券电子组 张世杰 博士

分析师登记编号：S1190523020001

太平洋证券电子组 李珏晗

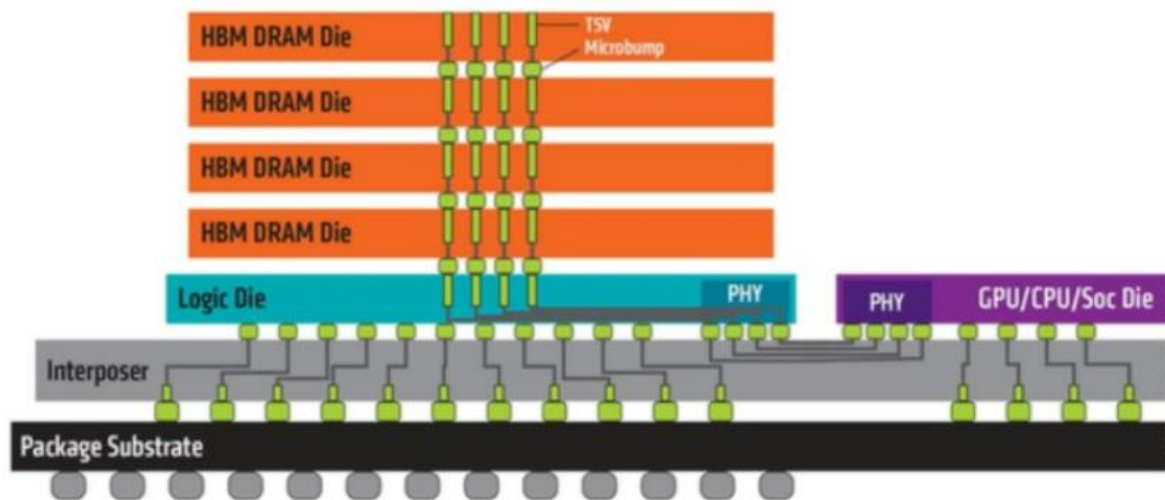
分析师登记编号：S1190523080001

- **HBM助力AI服务器向更高带宽、容量升级。**HBM即高带宽存储，由多层DRAMdie垂直堆叠，每层die通过TSV穿透硅通孔+ μ bumps技术实现与逻辑die连接，使得8层、12层die封装于小体积空间中，从而实现小尺寸于高带宽、高传输速度的兼容，成为高性能AI服务器GPU显存的主流解决方案。目前迭代至HBM3的扩展版本HBM3E，提供高达8Gbps的传输速度和16GB内存，由SK海力士率先发布，将于2024年量。
- **需求端：HBM成为AI服务器标配，催化其市场规模超50%增长。**HBM主要应用场景为AI服务器，最新一代HBM3e搭载于英伟达2023年发布的H200。根据Trendforce数据，2022年AI服务器出货量86万台，预计2026年AI服务器出货量将超过200万台，年复合增速29%。AI服务器出货量增长催化HBM需求爆发，且伴随服务器平均HBM容量增加，经测算，预期25年市场规模约150亿美元，增速超过50%。
- **供给端：三大厂产能供不应求，核心工艺变化带来新增量。**HBM供给厂商主要聚集在SK海力士、三星、美光三大存储原厂，根据Trendforce数据，2023年SK海力士市占率预计为53%，三星市占率38%、美光市占率9%。HBM在工艺上的变化主要在CoWoS和TSV，配套供应链及国产替代厂商有望受益：1) 封测：通富微电、长电科技、太极实业、深科技；2) 设备：赛腾股份、中微公司、拓荆科技；3) 材料：雅克科技、联瑞新材、壹石通、华海诚科。
- **风险提示：下游服务器需求不及预期风险；行业竞争加剧风险；技术发展不及预期风险。**

I	HBM助力AI服务器向更高带宽、容量升级.....	3
II	需求端：HBM成为AI服务器标配，催化其市场规模超50%增长.....	9
III	供给端：三大厂产能供不应求，核心工艺变化带来新增量.....	14
IV	主要公司分析.....	20

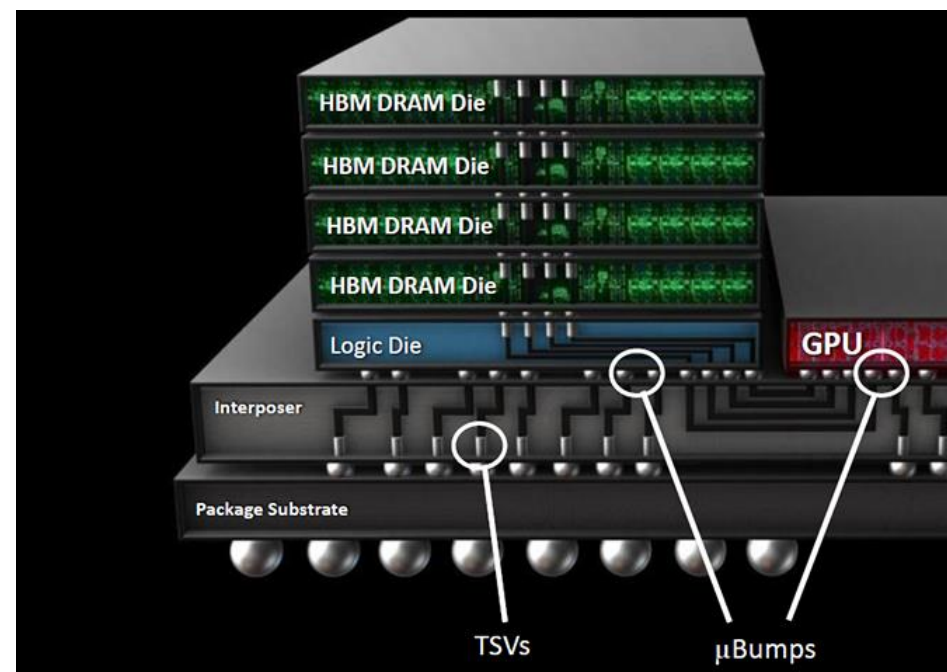
- **HBM (HighBandwidthMemory)** 即高带宽存储器，用于GPU与CPU之间的高速数据传输，主要应用场景集中在数据中心高性能服务器的GPU显存，小部分应用于CPU内存芯片。HBM同样为3D结构，由多层DRAMdie垂直堆叠，每层die通过TSV (Through Silicon Via) 穿透硅通孔+ μ bumps技术实现与逻辑die连接，再通过中阶层与GPU/CPU/SoC连接，使得4层、8层、12层等数量die封装于小体积空间中。

图表：HBM原理示意图6



资料来源：AMD、太平洋证券研究院

图表：HBM原理示意图



资料来源：AMD、太平洋证券研究院

HBM经多次处理速度迭代，现至HBM3的扩展版本HBM3E

- HBM经多次处理速度迭代，现至HBM3的扩展版本HBM3E，助力服务器推理性能提升。**

HBM1最早于2014年由AMD与SK海力士共同推出，作为GDDR竞品，为4层die堆叠，提供128GB/s带宽，4GB内存，显著优于同期GDDR5。HBM2于2016年发布，2018年正式推出，为4层DRAMdie，现在多为8层die，提供256GB/s带宽，2.4Gbps传输速度，和8GB内存；HBM2E于2018年发布，于2020年正式提出，在传输速度和内存等方面均有较大提升，提供3.6Gbps传输速度，和16GB内存。HBM3于2020年发布，2022年正式推出，堆叠层数及管理通道数均有增加，提供6.4Gbps传输速度，传输速度最高可达819GB/s，和16GB内存HBM3E由SK海力士发布HBM3的增强版，提供高达8Gbps的传输速度，24GB容量，计划于2024年大规模量产。

图表：4层Dramdie到8层Dramdie



图表：HBM演进路径

	HBM1	HBM2	HBM2E	HBM3
芯片密度	2GB	8GB	16GB	16GB
带宽	128GB/s	307GB/s	460GB/s	819GB/s
堆叠高度	4层	4层/8层	4层/8层	8层/12层
I/O速率	1Gbps	2.4Gbps	3.6Gbps	6.4Gbps
容量	1GB	4GB/8GB	8GB/16GB	16GB/24GB

资料来源：SK海力士官网，美光官网，太平洋证券研究院

资料来源：Sk海力士、太平洋证券研究院

HBM经多次处理速度迭代，现至HBM3的扩展版本HBM3E

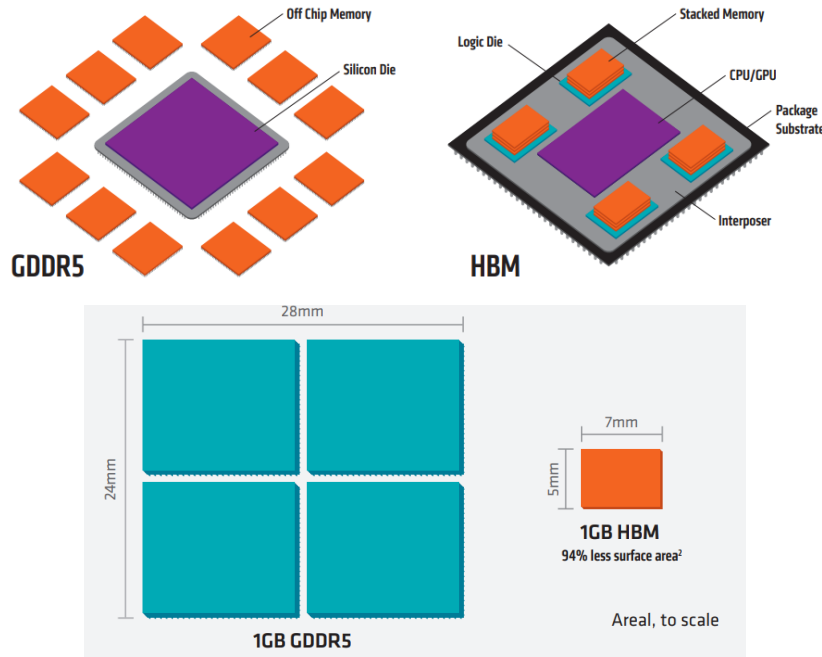
图表：三大存储原厂HBM演进路径

	HBM	HBM2 gen 1	HBM2 gen 2	HBM2E	HBM3	HBM3 gen 2	HBM3E
JEDEC Standard	Oct 13	Jan 16		Late 2018 (307 GB/sec)	Jan 22	July 2022	August 2023
Max dies per stack	4 (4-Hi)	8 (8-Hi)		to 12 (12-Hi)	1 8, 12, 16-hi 12-high 24GB (SK hynix)	8-x24Gb high (Micron). 12-high (Micron)	(12? - SK hynix)
Interface to processor	Interposer or direct stack						
Channels to processor	8 independent, 2 x 128bit per die			8 x 128-bit channels (1024 pins to CPU)	16 x 64-bit channels (1024 pins to CPU)		
Total channel width	1024 bits (with 8-Hi stack)						
Transfer rate	1 GT/sec/pin (1bit/pin)	2 GT/sec/pin		3.2 GT/sec (Samsung Flashbolt - 8 dies/stack) 3.65 GT/sec (SK hynix - 4 dies/stack)	6.4Gbps/pin (SK hynix) and 6.4GT/s	>9.2Gbps/pin (Micron)	8GT/s
Overall package bandwidth	128 GB/sec	256 GB/sec		410 GB/sec (Flashbolt) 460 GB/sec (SK hynix)	819 GBps (SK hynix) 1.075 TBps (Rambus)	1.2TBps (Micron)	1TBPs
Package capacity	4 GB	8 GB		16GB (Samsung Flashbolt).	16GB, 24GB. 24 GB (2GB/die) 12-hi SK hynix.	25GB (Micron Q3 2023) 36GB (Micron Q1 2024)	
Data rate				2.5 Tbit/sec?		1.2TBps (Micron)	1.15TBps (SK hynix)
IO speed				3.6 Gbit/s (SK Hynix).	5.2 Gbps (SK hynix) 8.4 Gbps (Rambus)		8GBps/pin (SK Hhynix)

资料来源：Trendforce，太平洋证券研究院

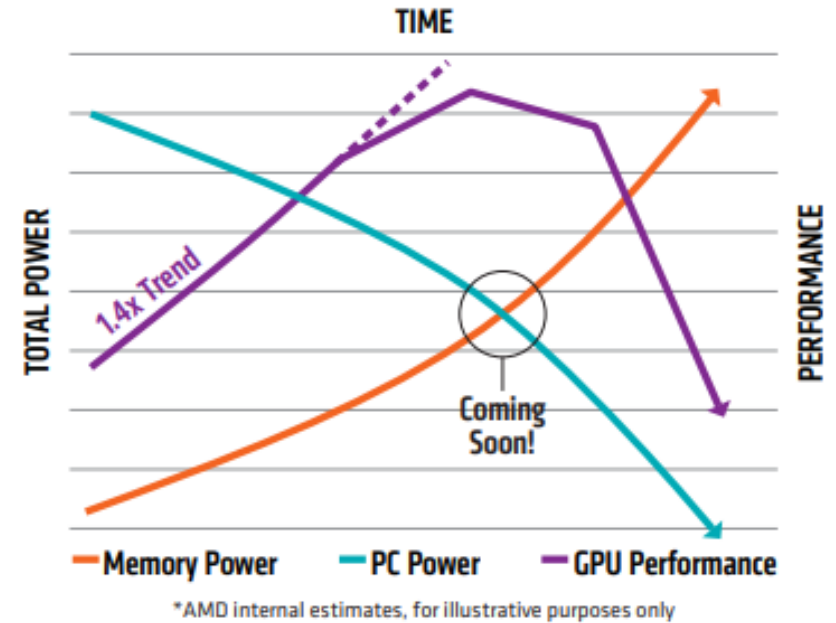
- 目前主流显卡使用的显存为GDDR5，但在应用中仍存在痛点：1) 消耗大量PCB面积：GPU核心周围分布12/16颗GDDR5芯片，而GDDR芯片尺寸无法继续缩小，占据更大空间，同时也需要更大的电压调节模块，2) GDDR5的功耗提升达到了对GPU性能提升产生负面影响的拐点：由于平台和设备需要在逻辑芯片与DRAM之间的实现功耗平衡，而当前GDDR进入功能/性能曲线的低效区，不断增长的内存功耗在未来会阻碍GPU性能提升。

图表：GDDR5与HBM面积对比



资料来源：AMD、太平洋证券研究院

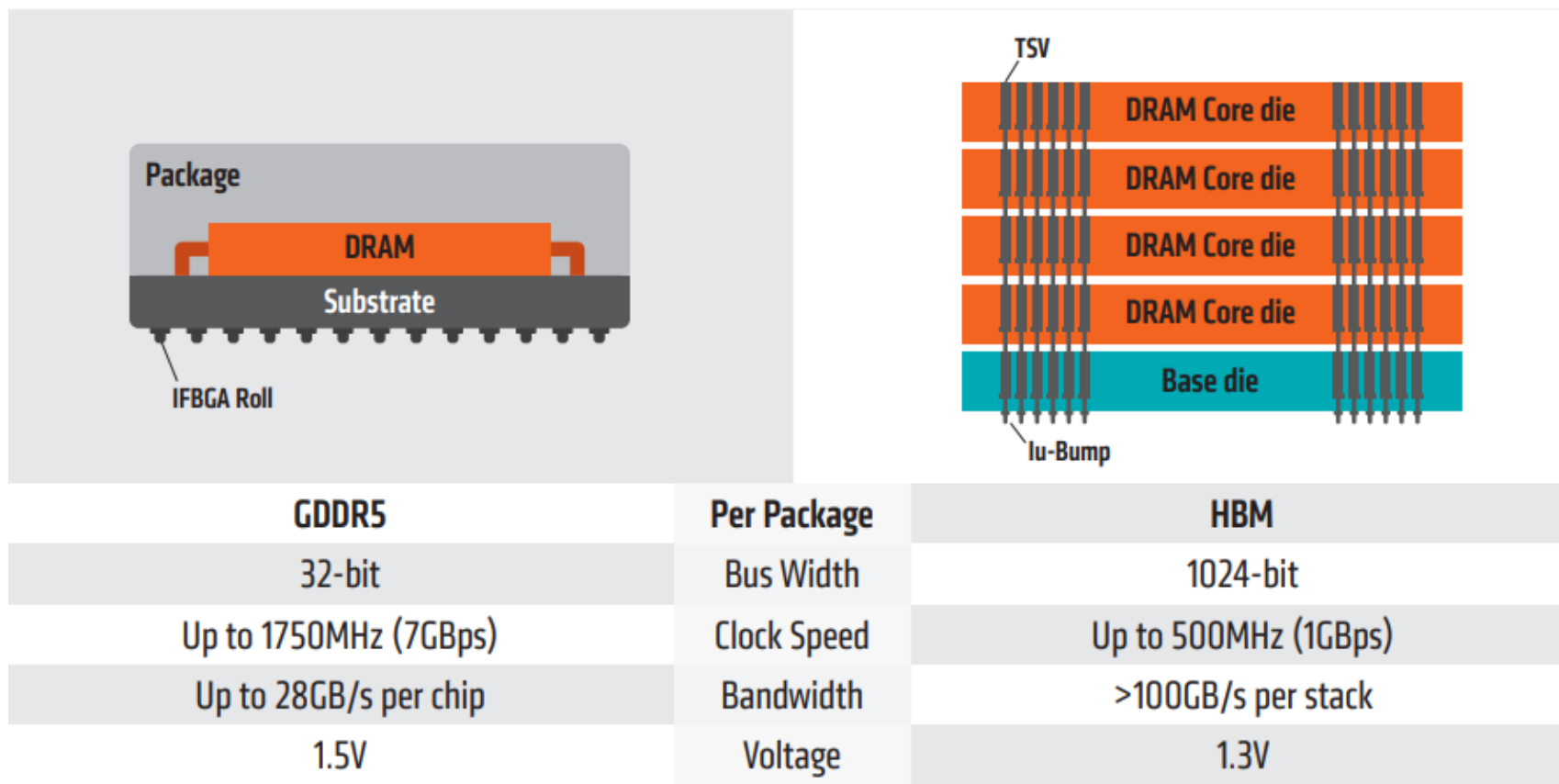
图表：内存能耗与逻辑ic能耗平衡拐点



资料来源：AMD、太平洋证券研究院

- 相较于传统GDDR解决方案，HBM在低功耗、超带宽、小尺寸等方面优势显著。

图表：GDDR5与HBM对比

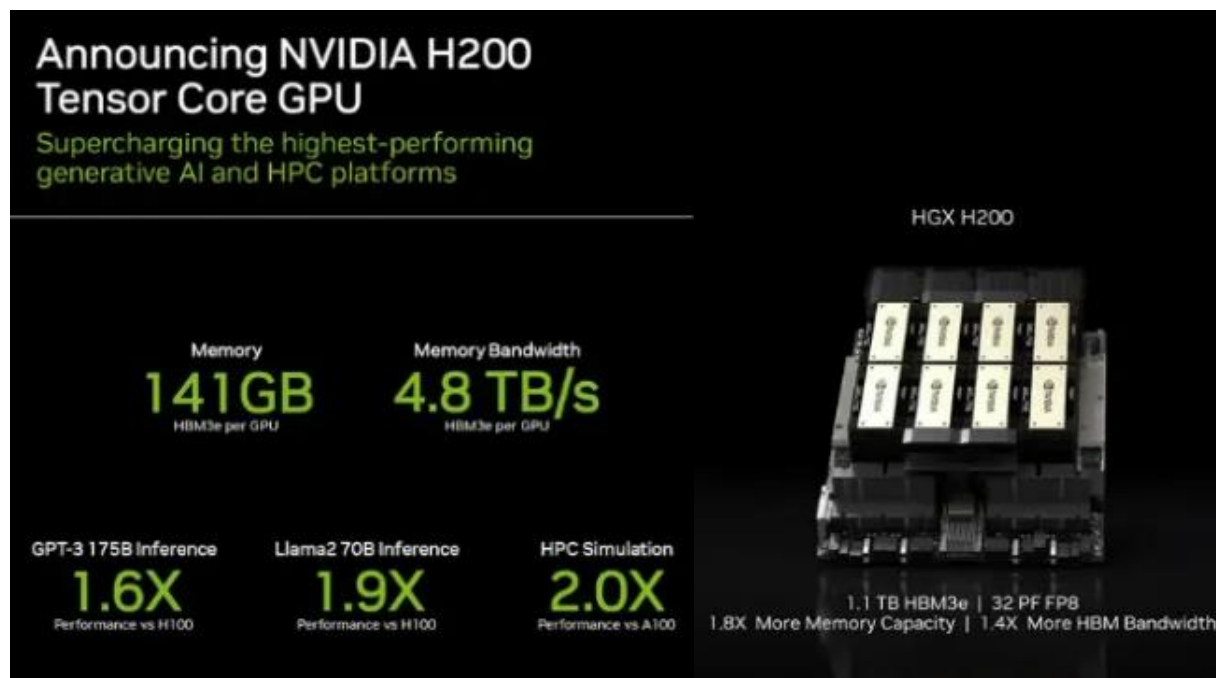


资料来源：AMD，太平洋证券研究院

I	HBM助力AI服务器向更高带宽、容量升级.....	3
II	需求端：HBM成为AI服务器标配，催化其市场规模超50%增长.....	9
III	供给端：三大厂产能供不应求，核心工艺变化带来新增量.....	14
IV	主要公司分析.....	20

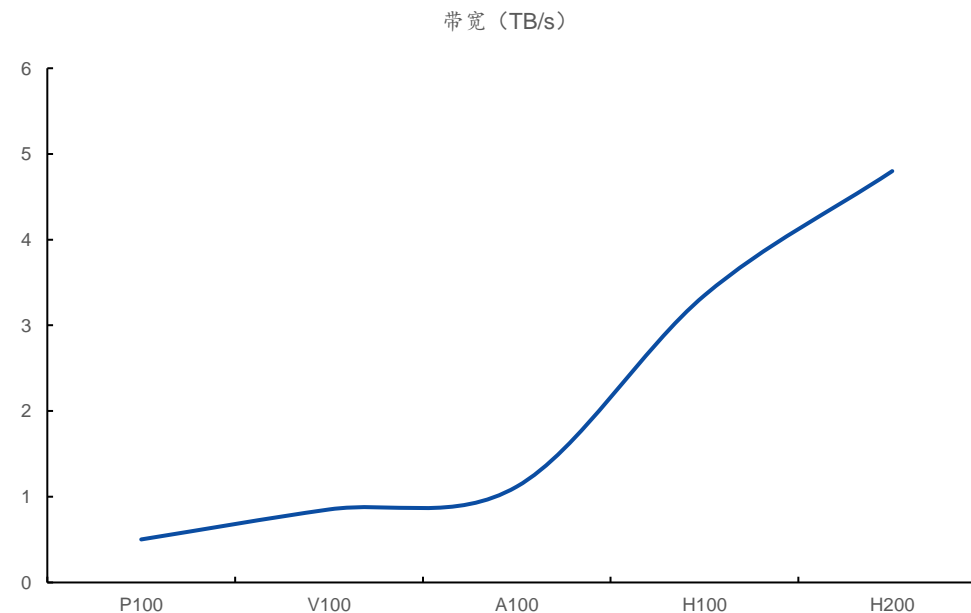
- HBM因其高带宽、低功耗、小体积等特性，广泛应用于AI服务器场景中。HBM的应用主要集中在高性能服务器，最早落地于2016年的NVP100GPU（HBM2）中，后于2017年应用在V100（HBM2）、于2020年应用在A100（HBM2）、于2022年应用在H100（HBM2e/HBM3），最新一代HBM3e搭载于英伟达2023年发布的H200，为服务器提供更快速度及更高容量。

图表：最新一代HBM3E应用于NVH200



资料来源：英伟达官网、太平洋证券研究院

图表：HBM迭代为AI服务器带来的带宽提升



资料来源：英伟达官网、太平洋证券研究院

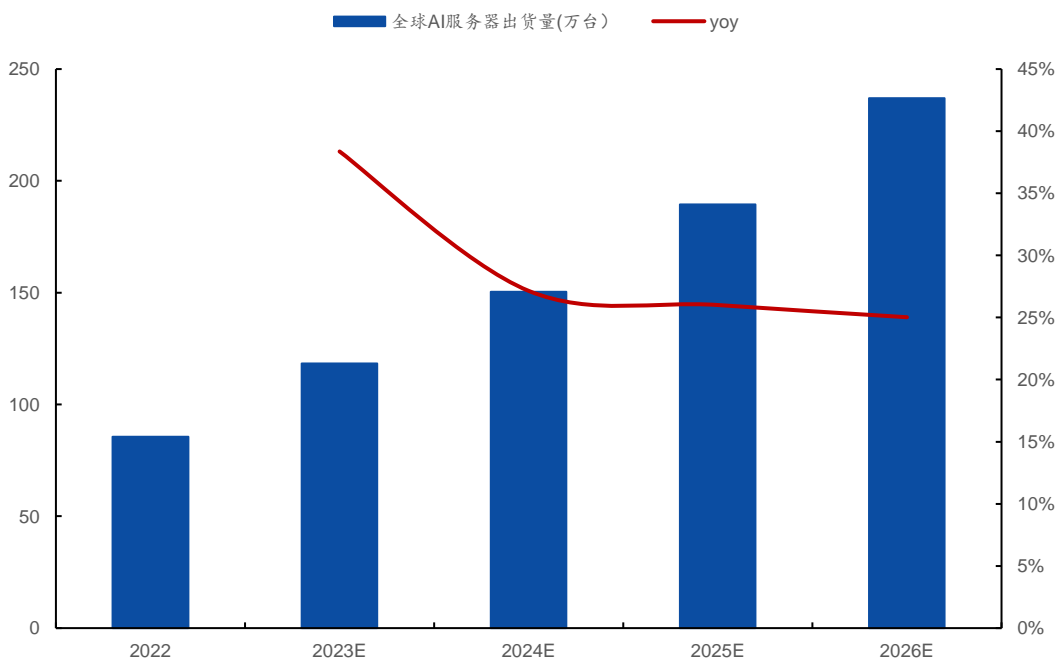
图表： HBM在AI服务器的应用

NVIDIA						
型号	V100	A100	A800	H100	H800	H200
HBM	HBM2	HBM2E	HBM2E	HBM3	HBM3	HBM3E
堆叠层数	8	8	8	12	12	12
带宽	900GB/sec	2TB/sec	2TB/sec	3.35TB/sec	3.35TB/sec	4.8TB/sec
容量	32GB	80G	80G	80GB	80GB	141GB
Google			AMD			
型号	TPUv4	TPUv5i	TPUv5	MI250X	MI300A	MI300X
HBM	HBM2	HBM2E	HBM3	HBM2E	HBM3	HBM3
堆叠层数	8	4	8	8	8	12
带宽	1200GB/s	819GB/s	2662GB/s	3277GB/s	5323GB/s	5734GB/s
容量	32GB	16GB	64GB	128GB	128GB	192GB

资料来源： NVIDIA官网， AMD官网， 太平洋证券研究院

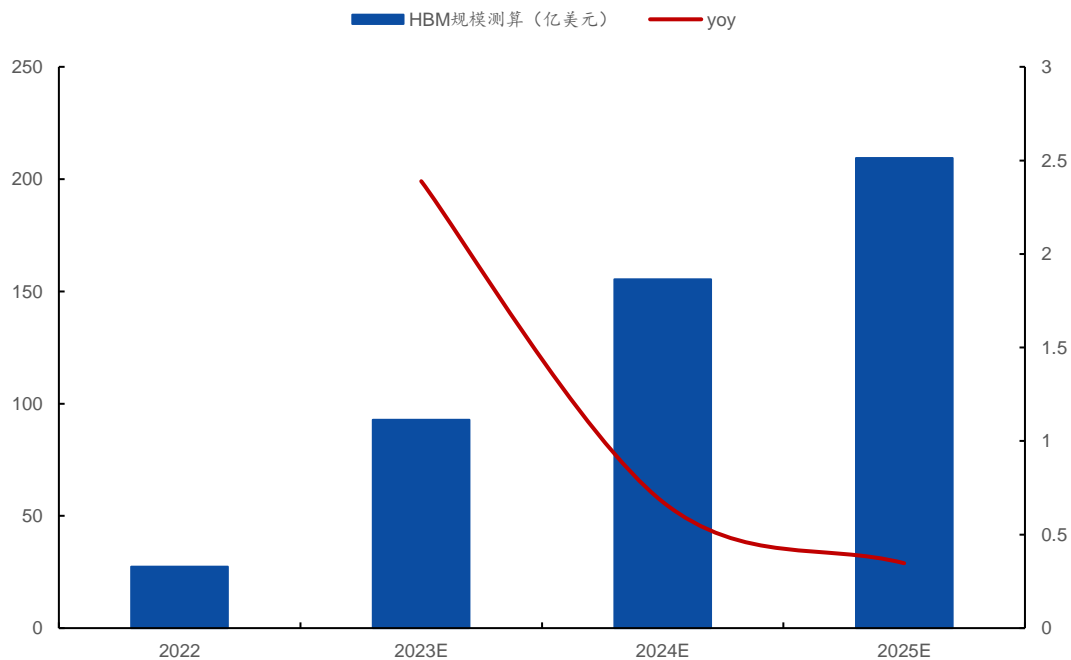
- AI服务器出货量增长催化HBM需求爆发，经测算，HBM 2025年市场规模将超过150亿美元，增速超50%。全球服务器增长较为稳健，但随着大数据模型训练需求AI服务器出货量呈爆发式增长，2022年AI服务器出货量86万台，渗透率不断提升，AI服务器在整体服务器占比10%左右，预计2026年AI服务器出货量将超过200万台，预计年复合增速29%。同时，主流AI服务器HBM容量需求从40GB向80GB、141GB升级，带动平均HBM容量提升。我们测算，2025年HBM全球市场规模将达到150亿美元，同比增速68%。

图表：AI服务器出货量年化增速预计约29%



资料来源：TrendForce、太平洋证券研究院

图表：HBM规模年化增速预计约31%



资料来源：太平洋证券研究院

图表：HBM规模测算

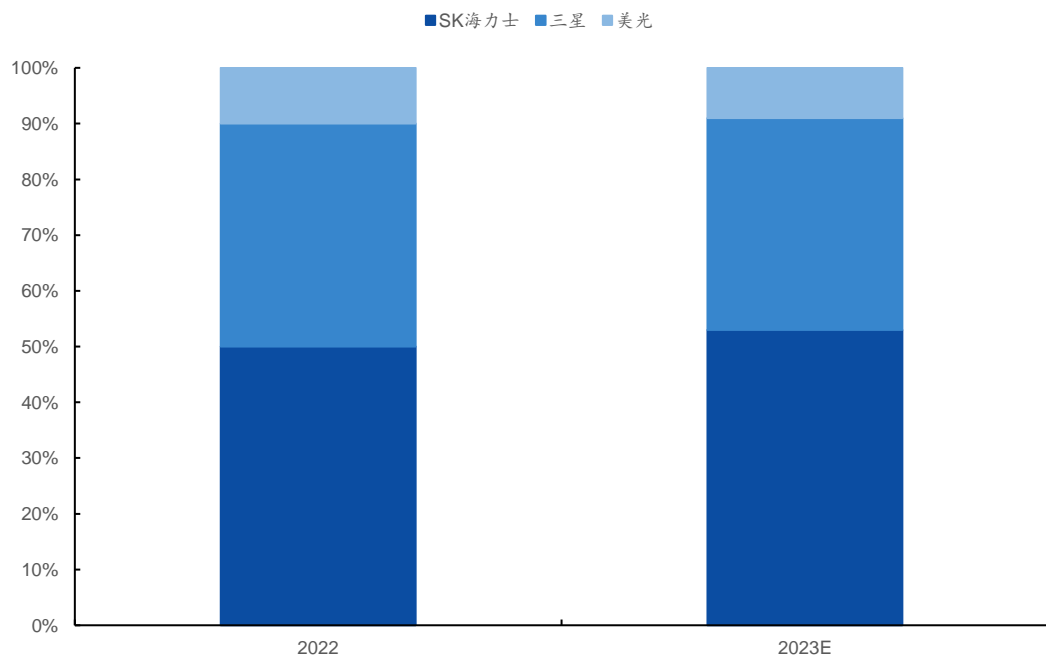
	2022	2023E	2024E	2025E
全球AI服务器出货量（万台）	85.50	118.30	150.40	189.50
yoy		38%	27%	26%
平均HBM容量（GB）	40.0	56.00	86.10	110.50
AI服务器HBM需求增量（GB*万台）	13680.0	26499.2	51797.8	83759.0
yoy		94%	279%	512%
单价（美元/GB）	20.0	35.0	30.0	25.0
HBM规模测算（亿美元）	27.36	92.75	155.39	209.40
yoy		239%	68%	35%

资料来源：太平洋证券研究院

I	HBM助力AI服务器向更高带宽、容量升级.....	3
II	需求端：HBM成为AI服务器标配，催化其市场规模超50%增长.....	9
III	供给端：三大厂产能供不应求，核心工艺变化带来新增量.....	14
IV	主要公司分析.....	20

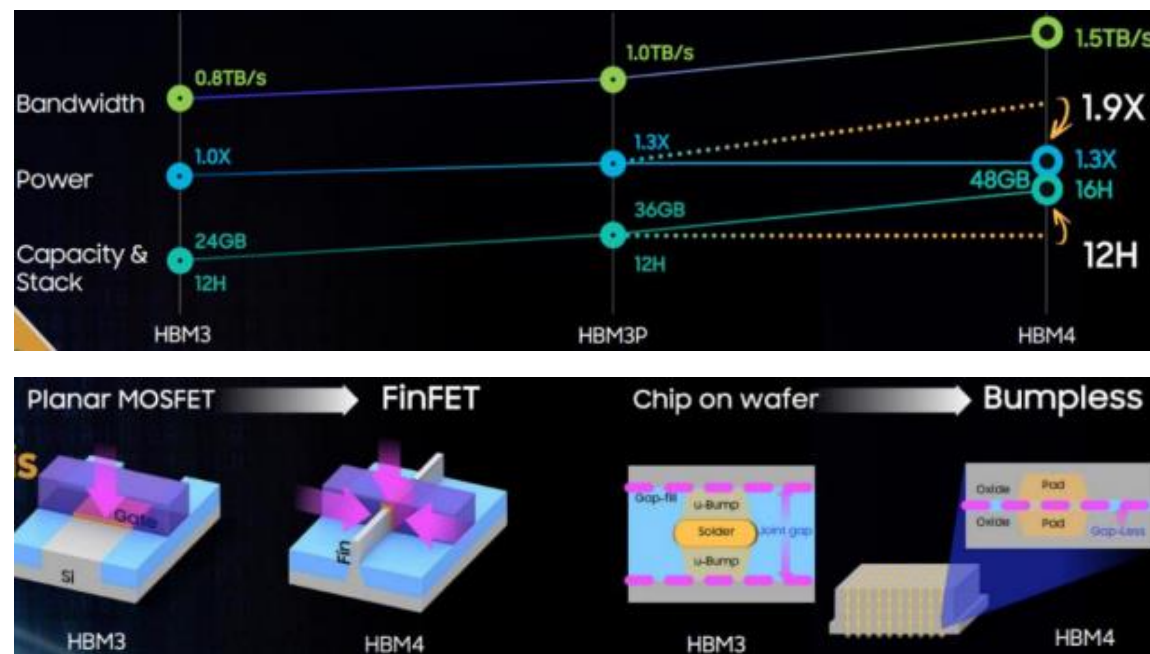
- HBM供给厂商主要聚集在SK海力士、三星、美光三大厂，SK海力士领跑。**三大存储原厂主要承担DRAMDie的生产及堆叠，展开技术升级竞赛，其中SK海力士与AMD合作发布全球首款HBM，23年率先供应新一代HBM3E，先发奠定市场地位，主要供应英伟达，三星供应其他云端厂商，根据TrendForce数据，2022年SK海力士市占率50%、三星市占率40%、美光市占率10%左右，2023年SK海力士市占率预计为53%，三星市占率38%、美光市占率9%。

图表：HBM三大存储原厂市场份额



资料来源：Trendforce、太平洋证券研究院

图表：HBM4封装方式变化将带来格局再演化



资料来源：国家统计局、太平洋证券研究院

供给端：三大存储原厂占据主要市场份额

图表：三大厂HBM规划

	Brand	Speed (Gbps)	Tech Nodes	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	Samsung	3.2-3.6	1Y 16Gb	8/16GB																EOL Time : Not Fixed			
	SK hynix	3.6	1Y 16Gb	8/16GB																EOL Time : Not Fixed			
	Micron	3.2-3.6	1Z 16Gb	16GB																EOL			
HBM3	Samsung	6.4	1Z 16Gb					16GB				24GB											
	SK hynix	5.6-6	1Z 16Gb	16GB				24GB															
HBM3e	Samsung	8	1alpha 24Gb					24GB				36GB											
	SK hynix	8	1beta 24Gb					24GB				36GB											
	Micron	8	1beta 24Gb					24GB				36GB											
HBM4		TBD	TBD	Full spec may be released in 2H24-2025; C/S in 2026																			

资料来源：Trendforce, 太平洋证券研究院

供给端：存储原厂加码HBM产能投放，预计24年提升2.5倍

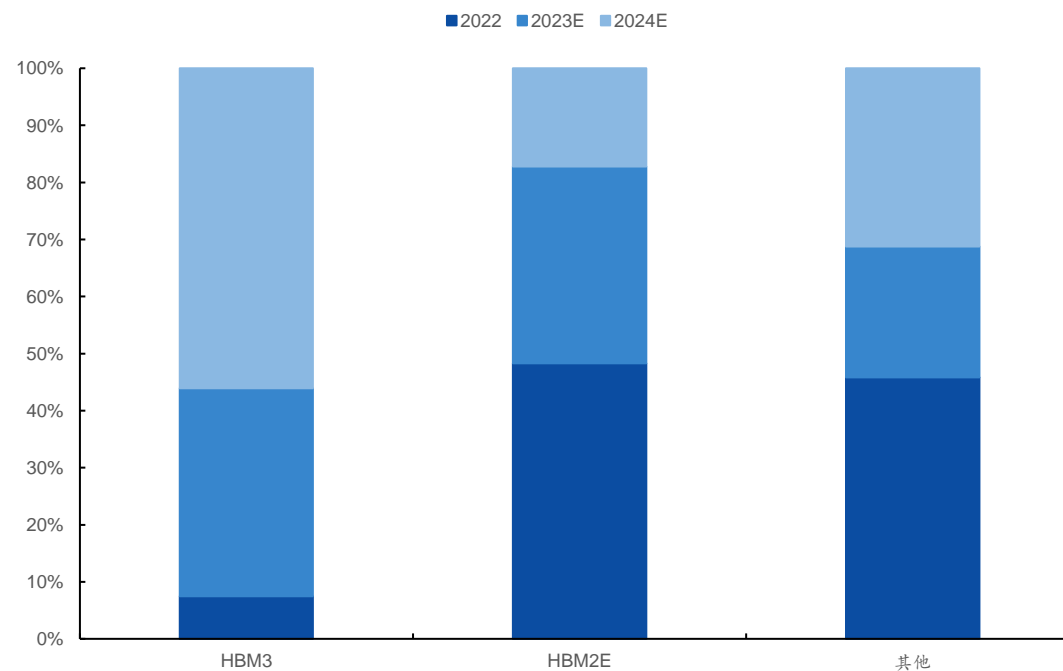
- 存储原厂加码HBM产能，SK海力士24年产能计划翻倍。SK海力士HBM3E将于24年上半年量产，目标24年HBM产能翻倍，24年资本支出规划虽与23年基本持平，但TSV相关投资将同比增加一倍以上。美光HBM3E将于24年初开始量产，预计24年资本支出75-80亿美元，同比略高，主要用于HBM量产。三星规划在天安厂新封装线，用于大规模生产HBM，预计追加投资7亿美元。

图表：三星及海力士计划扩充HBM产能



资料来源：公司公告、太平洋证券研究院

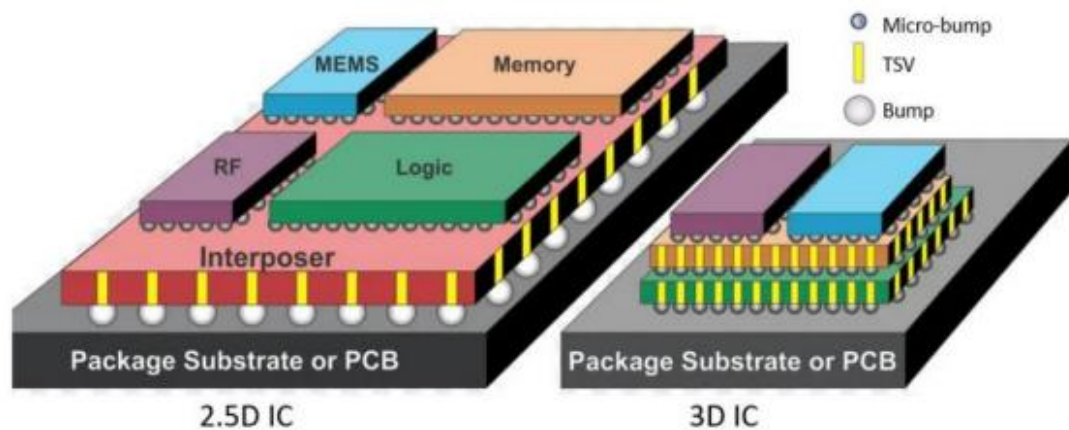
图表：HBM2产能释放用于生产HBM3



资料来源：Trendforce、太平洋证券研究院

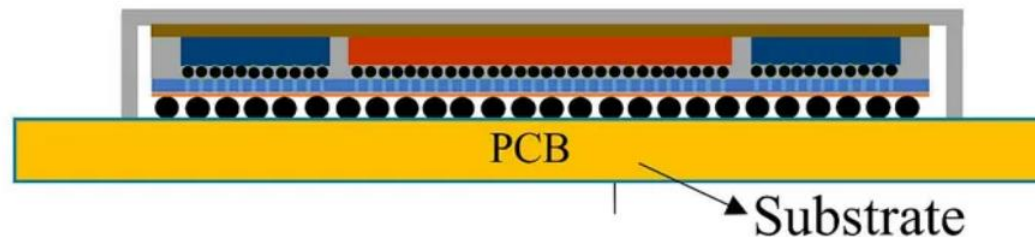
- HBM在封装工艺上的变化主要在CoWoS和TSV。1) CoWoS：是将DRAMDie一同放在硅中介层上，通过过ChiponWafer (CoW) 的封装制程连接至底层基板上，即将芯片通过ChiponWafer (CoW) 的封装制程连接至硅晶圆，再把CoW芯片与基板连接，整合成CoWoS。当前，HBM与GPU集成的主流解决方案为台积电的CoWoS，通过缩短互连长度实现更高速的数据传输，已广泛应用于A100、GH200等算力芯片中。

图表：2.5D与3D封装示意图



资料来源：台积电、太平洋证券研究院

图表：CoW于CoWoS示意图



资料来源：台积电、太平洋证券研究院

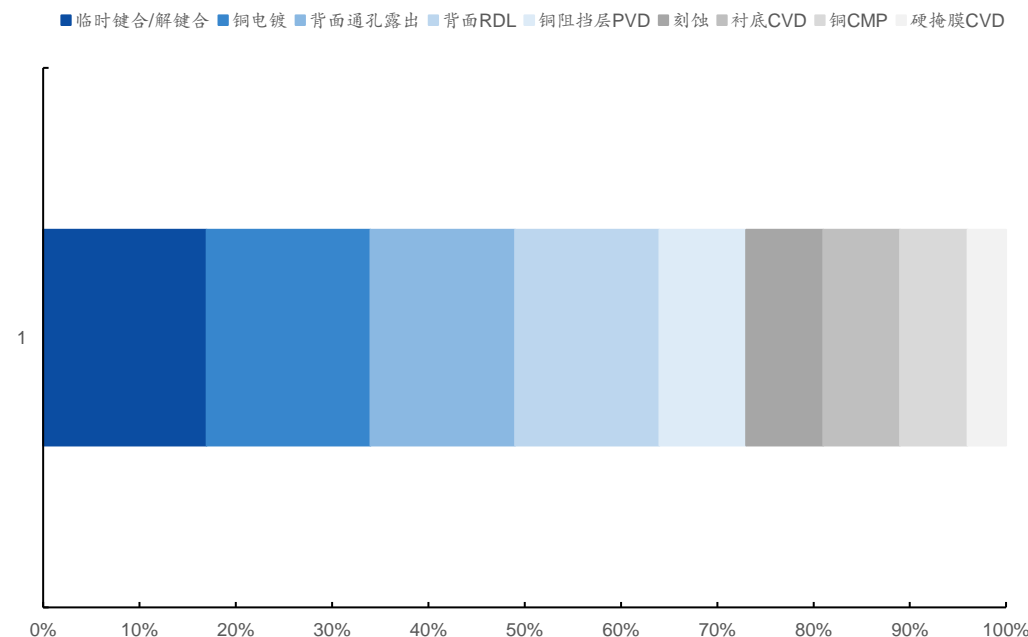
- **HBM在封装工艺上的变化主要在CoWoS和TSV。** 2) **TSV**: TSV硅通孔是实现容量和带宽扩展的核心，通过在整个硅晶圆厚度上打孔，在芯片正面和背面之间形成数千个垂直互连。在HBM中多层DRAMdie堆叠，通过硅通孔和焊接凸点连接，且只有最底部的die能向外连接到存储控制器，其余管芯则通过内部TSV实现互连。

图表：TSV技术实现多层DRAMDie堆叠



资料来源：AMD、太平洋证券研究院

图表：TSV工艺价值量拆解



资料来源：公司公告、太平洋证券研究院

I	HBM助力AI服务器向更高带宽、容量升级.....	3
II	需求端：HBM成为AI服务器标配，催化其市场规模超50%增长.....	9
III	供给端：三大厂占据核心份额，核心工艺变化带来新增量.....	14
IV	主要公司分析.....	20

- 通富微电：**公司具备国内顶级2.5D/3D封装平台及超大尺寸FCBGA研发平台，并且完成高层数再布线技术开发，为客户提供晶圆级和基板级Chiplet封测解决方案，已量产多层堆叠NAND Flash及LPDDR封装，是国内首家完成基于TSV技术的3DS DRAM封装开发的封测厂。公司通过收购AMD苏州及AMD槟城各85%股权，实现与大客户AMD深度绑定，AMD于2023年推出MI300，并于23Q4陆续降幅，预计24年将迎来大幅放量，公司将充分受益。
- 长电科技：**公司与客户共同开发基于高密度Fanout封装技术的2.5DfcBGA产品，TSV异质键合3DSoC的fcBGA通过认证。公司的封测服务覆盖DRAM、Flash等，深耕行业20余年，在16层NANDflash堆叠、35um超薄芯片制程能力、Hybrid异型堆叠等方面行业领先。公司XDF01技术平台布局AI、5G、汽车、工业等领域运用，XDF01 Chiplet量产。
- 太极实业：**公司子公司海太半导体与SK海力士签订5年合作协议，SK海力士持有海太半导体45%股权，海太与SK海力士形成深度绑定，海太为SK海力士提供DRAM封装服务。SK海力士23年占据HBM市场约50%份额，伴随24年HBM出货量爆发，公司有望承接溢出封测需求。
- 深科技：**公司通过收购沛顿科技切入存储封测，沛顿科技专注高端封测，具备DDR5、LPDDR5封测量产能力。沛顿Bumping项目已通过小批量试产，聚焦FC倒装工艺、POPt堆叠封装技术的研发、16层超薄芯片堆叠技术的优化。

- 赛腾股份：**公司通过收购全球领先的晶圆检测设备供应商日本OPTIMA进军晶圆检测装备领域，公司产品涉及固晶设备、分选设备，晶圆包装机、晶圆缺陷检测机、倒角粗糙度量测、晶圆字符检测机、晶圆激光打标机、晶圆激光开槽机等，通过OPTIMA切入三星、SK siltron、sumco等大客户供应链，三星加码24年资本开支，计划24年HBM产线扩产至23年的一倍以上，量测设备需求增加，公司有望受益。
- 中微公司：**公司为国内刻蚀设备龙头，其中ICP刻蚀设备在DRAM、3D NAND多个客户的生产线量产，伴随Nanova VE HP和Nanova LUX推出，ICP刻蚀设备的验证工艺范围持续扩展，在先进逻辑芯片、先进DRAM和3D NAND的ICP验证刻蚀工艺覆盖率有望扩展到50%-70%不等。公司8英寸、12英寸的Primo TSV 200E、Primo TSV 300E在晶圆级先进封装、2.5D封装和微机电系统芯片生产线等订单充沛，在12英寸的3D芯片的硅通孔刻蚀工艺上得到成功验证。
- 拓荆科技：**公司在半导体薄膜设备领域深耕十余年，ALD量产规模逐步扩大，随着存储芯片主流制造工艺已由2D NAND发展为3D NAND结构，结构的复杂化导致对于薄膜沉积设备的需求量逐步增加，同时3D NAND FLASH芯片的堆叠层数不断增高，从32/64层逐步向更多层及更先进工艺发展，对于薄膜沉积设备的需求提升的趋势亦将延续。

- 雅克科技**：公司于2016年通过收购控股UP Chemical，正式切入前驱体行业，UP Chemical自04年成为SK海力士前驱体核心供应商，形成多年深度绑定。SK海力士HBM占据50%市场份额，并且于22Q3起向英伟达独供HBM3。公司产品覆盖硅类前驱体、High-K 前驱体、金属前驱体，伴随HBM需求量增加，SK海力士出货量提升，公司将充分受益。
- 联瑞新材**：公司深耕无机填料和颗粒载体行业近40年，为国内硅微粉龙头。公司持续聚焦高端芯片AI、5G、HPC封装，异构集成先进封装Chiplet、HBM，以及新一代高频高速覆铜板，推出多种规格低CUT点Low微米/亚微米球形硅微粉，低CUT点Low α 微米/亚微米球形氧化铝粉，高频高速覆铜板用低损耗/超低损耗球形硅微粉，新能源电池用高导热微米/亚微米球形氧化铝粉。
- 壹石通**：公司布局记忆体封装用Low- α 高纯石英及Low- α 高纯氧化铝制备技术、Low- α 粉体制备技术、球形化生产工艺等核心技术，已具备Low- α 射线球形氧化铝的产业化能力。公司规划新建的年产200吨高端芯片封装用Low- α 球形氧化铝项目，有望在2023年下半年实现部分投产，目前日韩客户已陆续送样验证，客户初步反馈良好。Low- α 球形氧化铝粉体在EMC或GMC中的体积填充率大约在80%-90%，有望受益HBM出货量增加。
- 华海诚科**：公司为国内环氧塑封材料龙头，在先进封装领域，公司已成功研发了应用于QFN/BGA、FC、SiP、FOWLP/FOPLP 等封装形式的封装材料。GMC颗粒状环氧塑封料可以用于HBM的封装，公司相关产品已通过客户验证，现处于送样阶段，有望受益国产材料替代。

- 一、下游服务器需求不及预期风险；
- 二、行业竞争加剧风险；
- 三、技术发展不及预期风险。

行业评级

看好：我们预计未来6个月内，行业整体回报高于沪深300指数5%以上；
中性：我们预计未来6个月内，行业整体回报介于沪深300指数-5%与5%之间；
看淡：我们预计未来6个月内，行业整体回报低于沪深300指数5%以下。

公司评级

买入：我们预计未来6个月内，个股相对沪深300指数涨幅在15%以上；
增持：我们预计未来6个月内，个股相对沪深300指数涨幅介于5%与15%之间；
持有：我们预计未来6个月内，个股相对沪深300指数涨幅介于-5%与5%之间；
减持：我们预计未来6个月内，个股相对沪深300指数涨幅介于-5%与-15%之间。

重要声明

太平洋证券股份有限公司具有经营证券期货业务许可证，公司统一社会信用代码为91530000757165982D。

本报告信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证。负责准备本报告以及撰写本报告的所有研究分析师或工作人员在此保证，本研究报告中关于任何发行商或证券所发表的观点均如实反映分析人员的个人观点。报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。我公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行业务服务。本报告版权归太平洋证券股份有限公司所有，未经书面许可任何机构和个人不得以任何形式翻版、复制、刊登。任何人使用本报告，视为同意以上声明。

销售团队

职务	姓名	手机	邮箱
全国销售总监	王均丽	13910596682	wangjl@tpyzq.com
华北销售总监	成小勇	18519233712	chengxy@tpyzq.com
华北销售	常新宇	13269957563	changxy@tpyzq.com
华北销售	佟宇婷	13522888135	tongyt@tpyzq.com
华北销售	王辉	18811735399	wanghui@tpyzq.com
华北销售	巩赞阳	18641840513	gongzy@tpyzq.com
华北销售	郭佳佳	18811762128	guojj@tpyzq.com
华东销售总监	杨晶	13851505274	yangjingsh@tpyzq.com
华东销售	李昕蔚	18846036786	lixw@tpyzq.com
华东销售	张国锋	18616165006	zhanggf@tpyzq.com
华东销售	胡平	13122990430	huping@tpyzq.com
华东销售	周许奕	021-58502206	zhouxuyi@tpyzq.com
华东销售	丁锬	13524364874	dingkun@tpyzq.com
华南销售副总监	查方龙	18565481133	zhaf1@tpyzq.com
华南销售	张卓粤	13554982912	zhangzy@tpyzq.com
华南销售	何艺雯	13527560506	heyw@tpyzq.com
华南销售	郑丹璇	15099958914	zhengdx@tpyzq.com

研究院

中国北京100044

北京市西城区北展北街九号

华远·企业号D座

投诉电话：95397

投诉邮箱：kefu@tpyzq.com

