存储专题

AI 发展驱动 HBM 高带宽存储器放量

超配

核心观点

HBM 是当前 GPU 存储单元理想解决方案，AI 发展驱动 HBM 放量。HBM（高带宽存储器，High Bandwidth Memory）是由 AMD 和 SK Hynix 发起的基于 3D 堆栈工艺的高性能 DRAM，适用于高存储器带宽需求的应用场合。AI 大模型的数据计算量激增，需要应用并行处理数据的 GPU 作为核心处理器，而“内存墙”的存在限制了 GPU 数据处理能力，HBM 突破了内存容量与带宽瓶颈，可以为 GPU 提供更快的并行数据处理速度，打破“内存墙”对算力提升的桎梏，被视为 GPU 存储单元理想解决方案，将在 AI 发展中持续收益。

TSV 技术是 HBM 的核心技术之一，中微公司是 TSV 设备主要供应商。硅通孔技术（TSV）为连接硅晶圆两面并与硅衬底和其他通孔绝缘的电互连结构，可以穿过硅基板实现硅片内部垂直电互联，是实现 2.5D、3D 先进封装的关键技术之一，主要用于硅转接板、芯片三维堆叠等方面。中微公司在 2010 年就推出了首台 TSV 深孔硅刻蚀设备 Primo TSV®，提供的 8 英寸和 12 英寸硅通孔刻蚀设备，均可刻蚀孔径从低至 1 微米以下到几百微米的孔洞，并具有工艺协调性。

ALD 沉积在 HBM 工艺中不可或缺，雅克科技是 ALD 前驱体核心供应商，拓荆科技是 ALD 设备核心供应商。由于 ALD 设备可以实现高深宽比、极窄沟槽开口的优异台阶覆盖率及精确薄膜厚度控制，在 HBM 中先进 DRAM 加工工艺和 TSV 加工工艺中是必不可少的工艺环节。雅克科技是国内 ALD 沉积主要材料前驱体供应商，公司前驱体产品供应 HBM 核心厂商 SK 海力士，High-K、硅金属前驱体产品覆盖先进 1bDRAM、200 层以上 3DNAND 以及 3nm 先进逻辑电路等。拓荆科技是国内 ALD 设备的主要供应商之一，公司 PEALD 产品用于沉积 SiO_2、SiN 等介质薄膜，在客户端验证顺利；Thermal-ALD 产品已完成研发，主要用于沉积 Al_2O_3 等金属化合物薄膜。

HBM 主要应用 2.5D+3D 先进集成，IC 载板是转接板核心材料。HBM 借助 TSV 技术实现 2.5D+3D 先进集成，而 IC 载板是集成电路先进封装环节的关键载体，建立 IC 芯片与 PCB 板之间的讯号连接。在目前应用较广的 2.5D+3D 的先进封装集成电路中，都采用 IC 载板作为承载芯片的转接板，如 AMD2015 年推出的 Radeon R9 Fury X GPU 中使用了 64nm 的 TSV IC 载板作为转接板，NVIDIA 的 Pascal 100 GPU 基于台积电 16nm 工艺技术，连接在台积电 64nm CoWoS-2 转接板上，然后封装在 PCB 板上完成搭建。

相关公司：中微公司、雅克科技、拓荆科技、兆易创新、北京君正。

风险提示：HBM 下游需求不及预期，产业链相关企业发展进度不及预期。

重点公司盈利预测及投资评级

公司代码	公司名称	投资评级	昨收盘（元）	总市值（亿元）	EPS 2023E	EPS 2024E	PE 2023E	PE 2024E
002409	雅克科技	买入	70.5	335.5	1.78	2.47	38.1	27.6
688012	中微公司	买入	169.9	1050.3	2.26	2.84	72.6	57.6
688072	拓荆科技	买入	428	541.3	4.51	6.51	90.8	62.9
603986	兆易创新	买入	111.88	746.3	3.38	4.53	32.1	23.9
300223	北京君正	买入	93.36	449.6	1.82	2.45	48.5	36.1

资料来源：Wind、国信证券经济研究所预测

行业研究·行业专题

电子·半导体

超配·维持评级

证券分析师：胡剑
021-60893306
hujian1@guosen.com.cn
S0980521080001

证券分析师：胡慧
021-60871321
huhui2@guosen.com.cn
S0980521080002

证券分析师：周靖翔
021-60375402
zhoujingxiang@guosen.com.cn
S0980522100001

证券分析师：李梓澎
0755-81981181
lizipeng@guosen.com.cn
S0980522090001

证券分析师：叶子
0755-81982153
yezi3@guosen.com.cn
S0980522100003

联系人：詹浏洋
010-88005307
zhanliuyang@guosen.com.cn

联系人：李书颖
0755-81982362
lishuying@guosen.com.cn

市场走势

资料来源：Wind、国信证券经济研究所整理

相关研究报告

《射频电源行业专题-等离子体加工设备核心零部件，实现设备自主可控的必要条件》 ——2023-04-19
《半导体 5 月投资策略及英伟达复盘-半导体周期已触底，3 月全球销售额环比微增》 ——2023-05-14
《半导体行业一季报业绩综述：基金重仓股变化显著，半导体周期已触底》 ——2023-05-07
《半导体 4 月投资策略及英特尔复盘-AI+开启半导体新周期，看好设备国产化提速及服务器产业链》 ——2023-04-17
《半导体 3 月投资策略及美光科技复盘-继续推荐封测龙头及产品、客户拓展顺利的设计企业》 ——2023-03-06

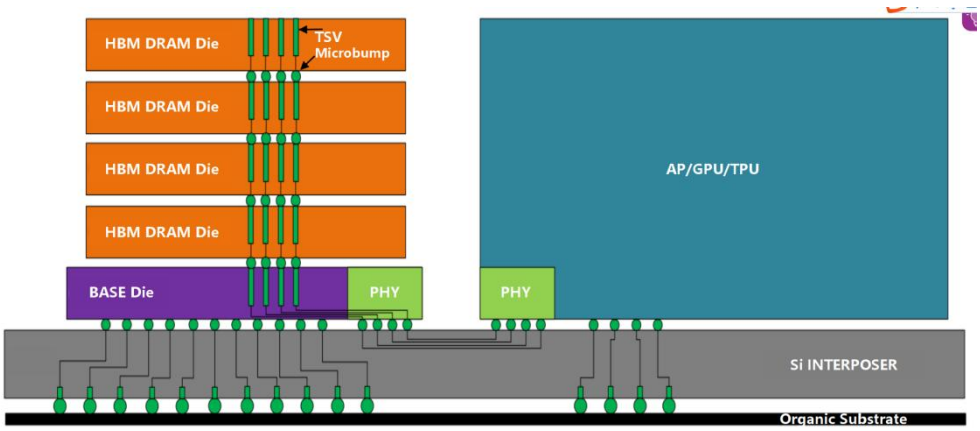
内容目录

图表目录

HBM：高带宽 DRAM，GPU 理想存储解决方案

HBM（高带宽存储器，High Bandwidth Memory）是一款新型的 CPU/GPU 内存芯片，是由 AMD 和 SK Hynix 发起的基于 3D 堆栈工艺的高性能 DRAM，适用于高存储器带宽需求的应用场合。HBM 以位元计算，通过增加带宽，扩展内存容量，让更大的模型、更多的参数留在离核心计算更近的地方，从而减少内存和存储解决方案带来的延迟，目的实现大容量，高位宽的 DDR 组合阵列，目前 HBM 占整个 DRAM 市场比重约 1.5%，为新型高性能存储产品。

图1：HBM 主要以 TSV 技术垂直堆叠芯片，达到缩减体积、降低能耗的目的

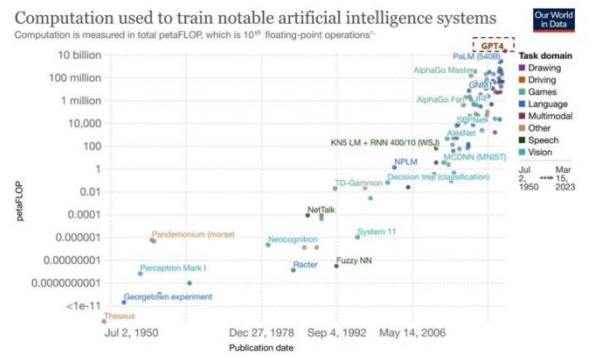


资料来源：Micron，国信证券经济研究所整理

AI 大模型催动 DRAM 需求

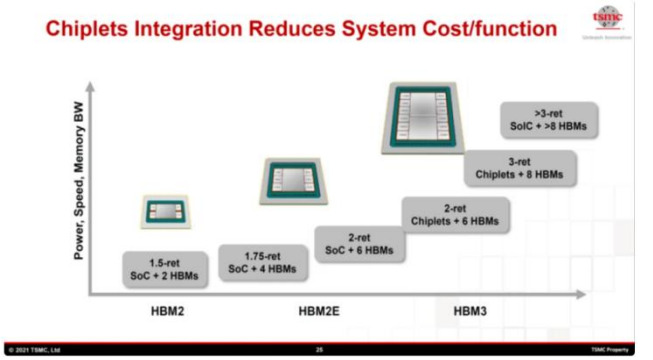
AI 大模型处理数据的吞吐量呈指数级增长，对内存的提出更高的带宽需求，HBM 迎来发展机遇。AI 大模型的数据计算量激增，需要应用并行处理数据的 GPU 作为核心处理器，GPU 搭载的内存芯片带宽关联 GPU 数据处理能力，高带宽的内存芯片可以为 GPU 提供更快的并行数据处理速度，对 GPU 的性能起到了决定性作用。

图2：AI 模型计算量增长迅猛



资料来源：Our World in Data，国信证券经济研究所整理

图3：HBM 提供更快的数据处理速度



资料来源：TSMC，国信证券经济研究所整理

动态内存能力对大模型训练至关重要。内存方面，大模型训练的内存可以大致理解为参数、优化器状态、激活、梯度四部分的和。它们大致分为两类：静态内存和动态内存。参数、优化器状态较为固定，属于静态内存，激活和梯度等中间变

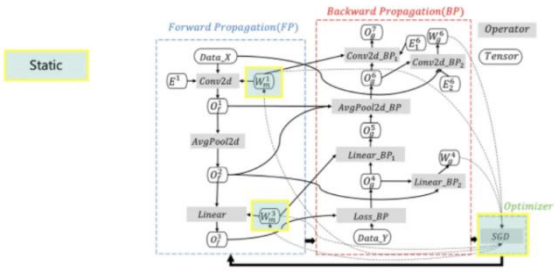
量属于动态内存，是最主要的内存占用原因，动态内存通常是静态内存的数倍。

图4：大模型语言计算对应内存需求

	公式	注释
模型参数	• 混合精度（fp16/bf16 和 fp32），$\text{memory}_{\text{model}} = (2\ \text{bytes/param}) \cdot (\text{No. params})$	
优化器内存	• 对于vanilla AdamW，$\text{memory}_{\text{optimizer}} = (12\ \text{bytes/param}) \cdot (\text{No. params})$ ▪ fp32参数副本：4字节/参数 ▪ 动量：4字节/参数 ▪ 方差：4字节/参数 • 对于像bitsandbytes这样的8位优化器，$\text{memory}_{\text{optimizer}} = (6\ \text{bytes/param}) \cdot (\text{No. params})$ ▪ fp32参数副本：4字节/参数 ▪ 动量：1字节/参数 ▪ 方差：1字节/参数 • 对于具有动量的类SGD优化器，$\text{memory}_{\text{optimizer}} = (8\ \text{bytes/param}) \cdot (\text{No. params})$ ▪ fp32参数副本：4字节/参数 ▪ 动量：4字节/参数	
梯度内存	• fp32，$\text{memory}_{\text{gradients}} = (4\ \text{bytes/param}) \cdot (\text{No. params})$ • fp16，$\text{memory}_{\text{gradients}} = (2\ \text{bytes/param}) \cdot (\text{No. params})$	
激活重计算	$\text{memory}_{\text{activations}}^{\text{No Recomputation}} = sbhL(10 + \frac{24}{t} + 5\frac{a \cdot s}{h \cdot t})$ bytes $\text{memory}_{\text{activations}}^{\text{Selective Recomputation}} = sbhL(10 + \frac{24}{t})$ bytes $\text{memory}_{\text{activations}}^{\text{Full Recomputation}} = 2 \cdot sbhL$ bytes	a number of attention heads; b microbatch size; h hidden dimension size; L number of transformer layers; p pipeline parallel size; s sequence length; t tensor parallel size; v vocabulary size
模型训练内存需求	$\text{Total Memory}_{\text{Training}} = \text{Model Memory} + \text{Optimiser Memory} + \text{Activation Memory} + \text{Gradient Memory}$	
模型推理内存需求	$\text{Total Memory}_{\text{Inference}} \approx (1.2) \times \text{Model Memory}$	

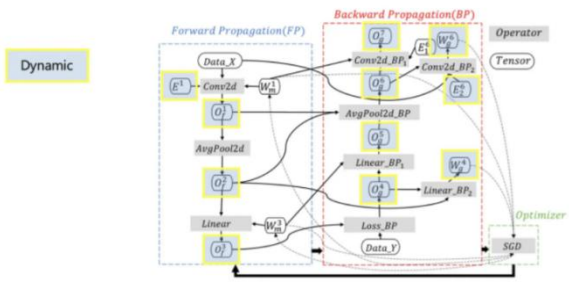
资料来源：Eleutheral，国信证券经济研究所整理

图5：静态内存参数、优化器状态较为固定



资料来源：知乎，国信证券经济研究所整理

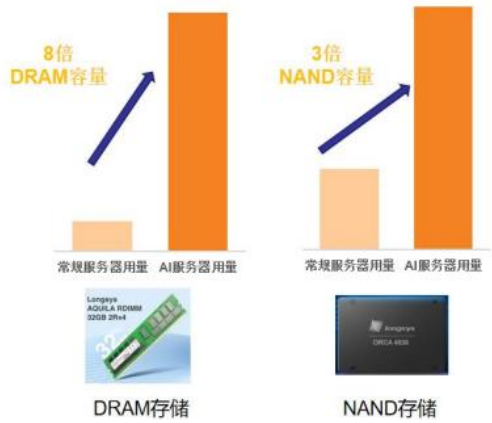
图6：动态内存通常是静态内存的数倍



资料来源：知乎，国信证券经济研究所整理

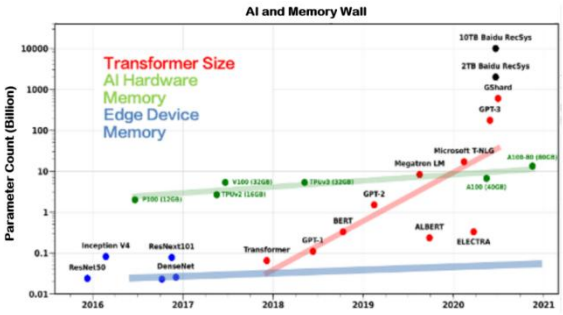
训练 1750 亿参数的 GPT3 所需内存，大约需要 3.2TB 以上。静态内存方面，大多数 Transformer 都是以混合精度训练的，如 FP16+FP32，以减少训练模型内存，则一个参数占 2 个字节，参数和优化器状态合计占用内存 1635G。而动态内存，根据不同的批量大小、并行技术等结果相差较大，通常是静态内存的数倍。更简洁的估算方法，可以假设典型的 LLM 训练中，优化器状态、梯度和参数所需的内存为 20N 字节，其中 N 是模型参数数量，则 1750 亿参数的 GPT3 大概需要 3.2TB 内存。推理所需内存则较小，假设以 FP16 存储，175B 参数的 GPT3 推理大约需要内存 327G，则对应 4 张 80G A100，如果以 FP32 运算，则需要 10 张。

图7：AI 服务器提升存储器需求



资料来源：闪存市场，国信证券经济研究所整理

图8：模型越大需要设备内存越大

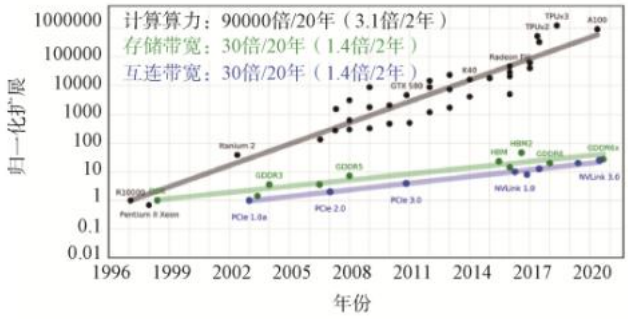


资料来源：NVIDIA，国信证券经济研究所整理

3D DRAM 解决“内存墙”问题

“内存墙”是处理器算力超过存储芯片存取能力，内存墙的存在导致综合算力被存储器制约。据行业预计，处理器的峰值算力每两年增长 3.1 倍，而动态存储器(DRAM)的带宽每两年增长 1.4 倍，存储器的发展速度远落后于处理器，相差 1.7 倍。由于处理器处理数据过程同样需要动态存储器的支持，“内存墙”的存在制约了处理器的算力提升速度。

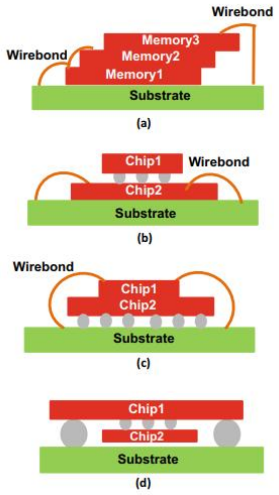
图9：存储带宽落后于算力成长速度形成“内存墙”



资料来源：曹立强、侯峰泽，《先进封装技术的发展与机遇》，前瞻科技杂志，2022 年第 3 期"集成电路科学与工程专刊”，前瞻科技杂志，国信证券经济研究所整理

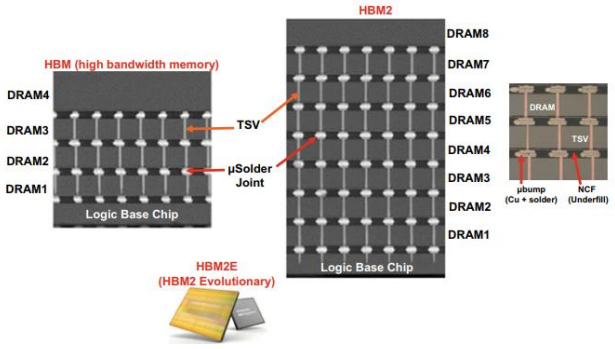
将 DRAM 3D 化是解决内存墙的主要方法。将 DRAM 从传统 2D 转变为立体 3D，借助 TSV 等技术实现内存芯片在 3D 维度进行堆叠，充分利用空间提升内存芯片密度，缩小芯片表面积，契合半导体行业小型化、集成化的发展趋势。3D DRAM 的发展也有堆叠引线键合、倒装混合引线键合等多种实现方式，HBM 是 3D DRAM 的一种形式，相较于其他 DRAM 的集成方式，HBM 存储单元外的导线长度最短，数据传递速度最快，损耗最小，是目前最理想化的 3D DRAM 形式。HBM 突破了内存容量与带宽瓶颈，打破了“内存墙”对算力提升的桎梏，被视为新一代 DRAM 解决方案，是未来 DRAM 重要发展路径。

图10：3D DRAM 几种实现方式



资料来源：Lau. J，《Chip Design and Heterogeneous Integration Packaging》，2023 版，140-145 页，国信证券经济研究所整理

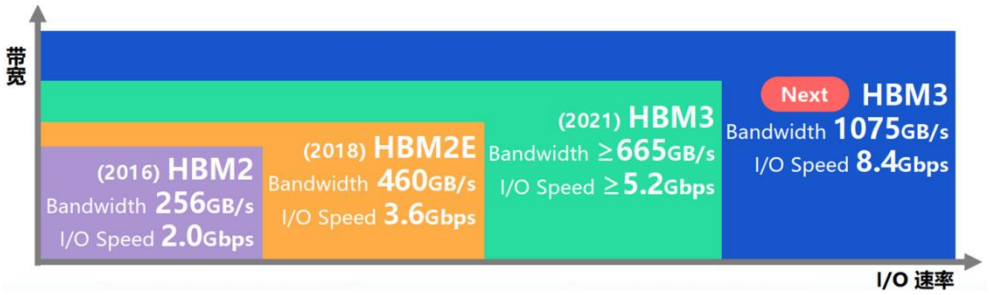
图11：HBM 每个 DRAM 单元间引线最短



资料来源：Lau. J，《Chip Design and Heterogeneous Integration Packaging》，2023 版，140-145 页，国信证券经济研究所整理

据集邦咨询数据，存储巨头 SK 海力士是目前 HBM 最大的供应商，占据 50%的市场份额。SK 海力士在 2013 年推出了首款 HBM 存储器，共包含 4 个 DRAM 单元，后续海力士陆续推出了 HBM2、HBM2e 和 HBM3，带宽和 I/O 速度进一步提升。除海力士外，三星、美光占据了 HBM 其余市场。由于 HBM 主要和 GPU 搭载使用，封装主要以 TSV 3D 封装进行，所以通常在晶圆厂内完成，当前台积电、格芯等也在发力 HBM 技术的研究与制造。当前 SK 海力士已经实现了 HBM3 的量产，搭载在 NVIDIA GPU H100 之中，其带宽在 HBM2 460 GB/s 的基础上提升了 78%，达到了 819 GB/s，随着 GPU 算力的不断提升，HBM 在速度、密度、功耗、占板空间方面也将持续提升。

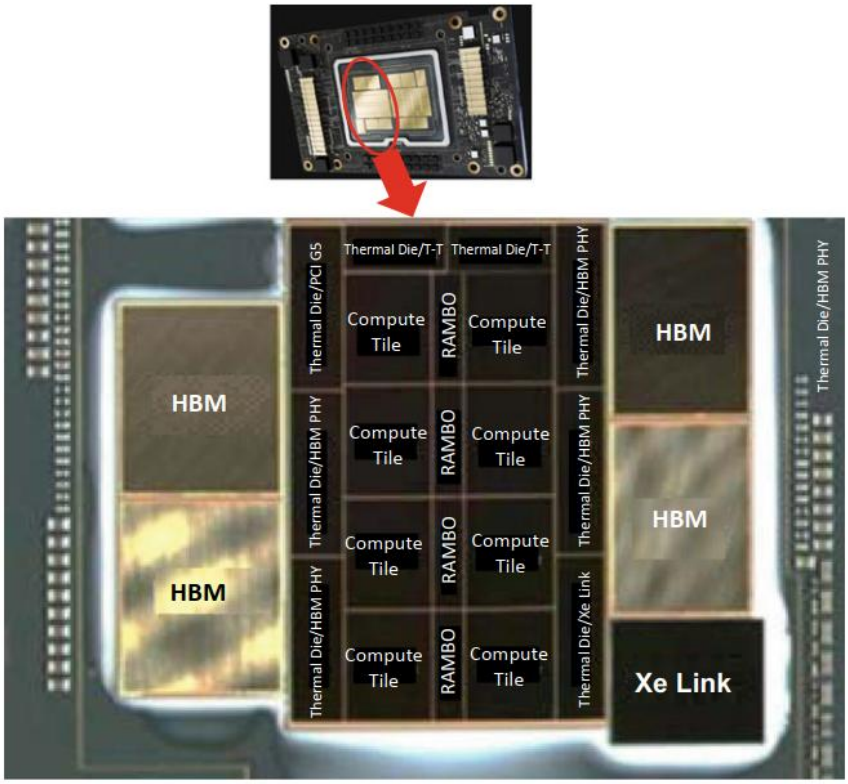
图12：HBM3 带宽进一步提升



资料来源：SK 海力士，国信证券经济研究所整理

当前高端 GPU 已搭载高端 HBM 作为先进封装存储单元的解决方案。NVIDIA 高端 GPU H100、A100 主采 HBM2e、HBM3，H100 GPU 上主要搭载 HBM3 内存。此外，AMD 的 MI200、MI300 以及 Google 自研 TPU 等均将搭载高带宽的 HBM 提升内存能力，Trend Force 集邦咨询预估 2023 年 HBM 需求量将年增 58%，2024 年有望再增长 30%。

图13：Chiplet 搭载 HBM 作为存储单元解决方案



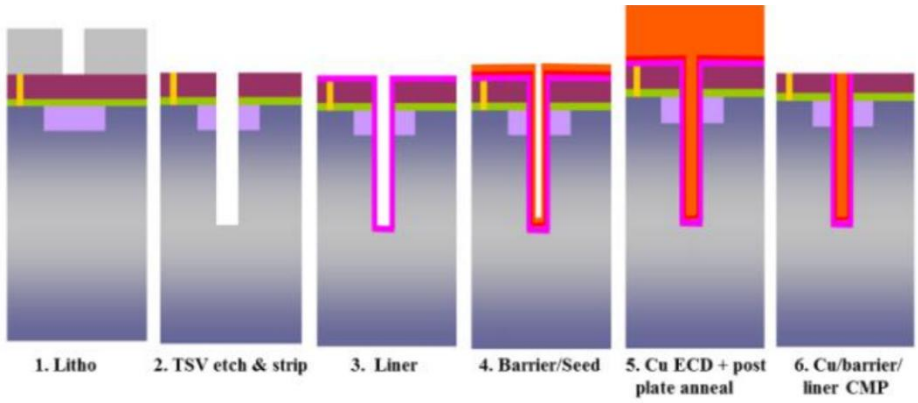
资料来源：电子发烧友网，国信证券经济研究所整理

关键技术助力 HBM 发展

HBM 关键技术#1：硅通孔技术(TSV)

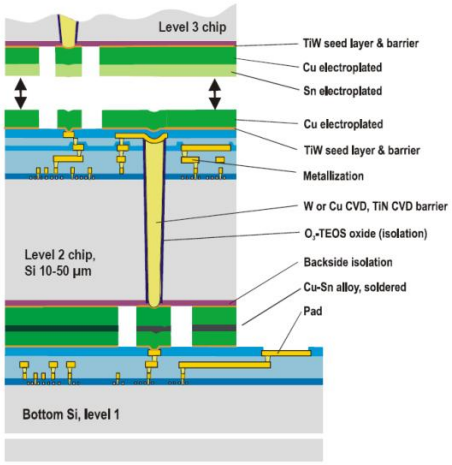
硅通孔技术（TSV，Through Silicon Via）为连接硅晶圆两面并与硅衬底和其他通孔绝缘的电互连结构，可以穿过硅基板实现硅片内部垂直电互联，这项技术是目前唯一的垂直电互联技术，是实现 2.5D、3D 先进封装的关键技术之一，主要用于硅转接板、芯片三维堆叠等方面。TSV 的尺寸多为 10μm×100μm 和 30μm×200μm，开口率介于 0.1%-1%。相比平面互连，TSV 可减小互连长度和信号延迟，降低寄生电容和电感，实现芯片间的低功耗和高速通信，增加宽带和封装小型化。在有源芯片中，当前 TSV 开孔一般在 10um 左右，深宽比约为 10:1，微凸点互联间距在 40-50um，由于 TSV 本身占据面积较大，且会形成一定应力影响区，发展方向向 5um 以下、深宽比 10 以上发展，实现更小的体积和更低的成本。

图14：硅通孔技术流程



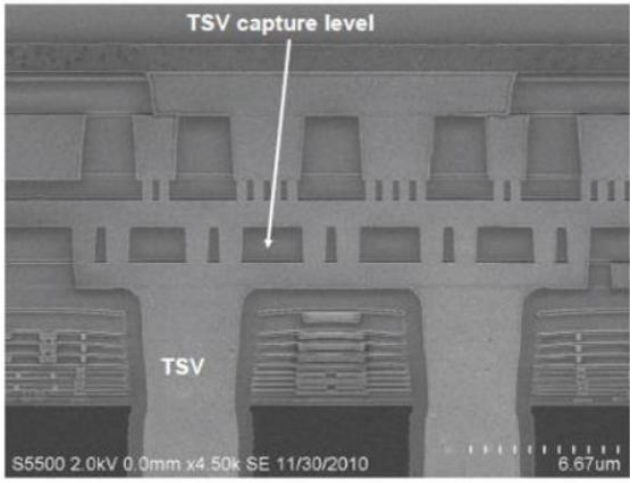
资料来源：《新时代先进封装》，于大全，2021 版，8-10 页，国信证券经济研究所整理

图15：TSV 当前深宽比约在 10：1



资料来源：《新时代先进封装》，于大全，2021 版，8-10 页，国信证券经济研究所整理

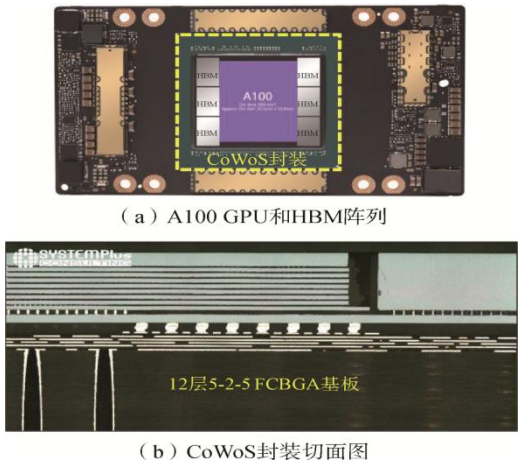
图16：TSV 目前开孔约在 10um



资料来源：《新时代先进封装》，于大全，2021 版，8-10 页，国信证券经济研究所整理

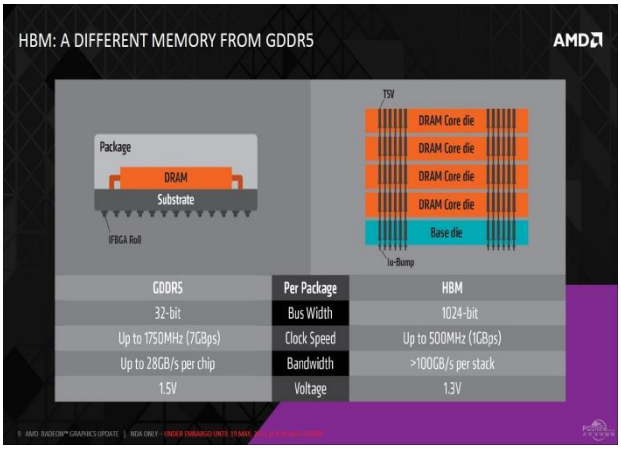
HBM 是借助 TSV 技术实现多个 DRAM 之间的连通堆叠。借助 TSV 技术，多个 HBM 单元可以以 3D 形式集成在同一个转接板上。英伟达采用台积电第 4 代 CoWoS 技术封装了 A100 GPU，实现一颗 A100 GPU 和 6 个三星 HBM2 集成为一颗芯片。该技术将多颗芯片键合至硅基转接板晶圆上(Si Interposer)，形成逻辑 SoC 芯片和 HBM 阵列，通过 RDL 和 TSV 形成互联并连接硅基转接板晶圆凸点。英特尔 Foveros 技术（3D Face to Face Chip Stack for heterogeneous integration）亦通过 3D TSV 实现 3D 堆叠异构封装技术。

图17：英伟达 A100 GPU CoWoS 封装



资料来源：曹立强、侯峰泽，《先进封装技术的发展与机遇》，前瞻科技杂志，2022 年第 3 期"集成电路科学与工程专刊"，前瞻科技杂志，国信证券经济研究所整理

图18：基于 TSV 技术实现堆叠 HBM



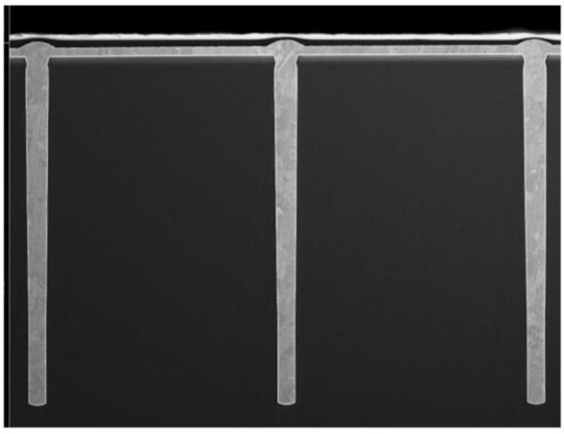
资料来源：AMD，国信证券经济研究所整理

关键技术#2：ALD 沉积

原子层沉积（ALD）是将原子逐层沉积在衬底材料上的工艺，通过将两种或多种前驱体交替通过衬底表面，发生化学吸附反应逐层沉积在衬底表面，能对复杂形貌基底表面全覆盖成膜。由于 ALD 设备可以实现高深宽比、极窄沟槽开口的优异台阶覆盖率及精确薄膜厚度控制，实现了芯片制造工艺中关键尺寸的精度控制。HBM 先进 DRAM 加工工艺和 TSV 加工工艺两个环节中，ALD 是必不可少的核心设备之一。

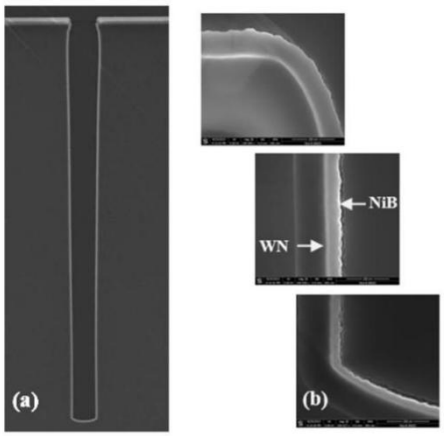
ALD 沉积 TSV 扩散阻挡层。TSV 深孔制作完成后，需要进行电化学镀铜来完成金属沉积形成导线，由于铜化学性质活泼，在电镀前需要以 ALD 方式沉积 WN 形成扩散阻挡层，防止铜的电化学迁移导致物理失效。IMEC 基于 via middle 的 TSV 制造工艺中，硅通孔采用了高保型 ALD 氧化层绝缘，厚度为 125nm，获得了 100%覆盖率。在单纯热工艺下，按顺序地驱动多种前驱体和反应体沉积 WN 作为扩散阻挡层，沉积温度 375℃，覆盖率大于 90%。

图19：IMEC TSV 工艺示意图



资料来源：Electrochimica Acta，国信证券经济研究所整理

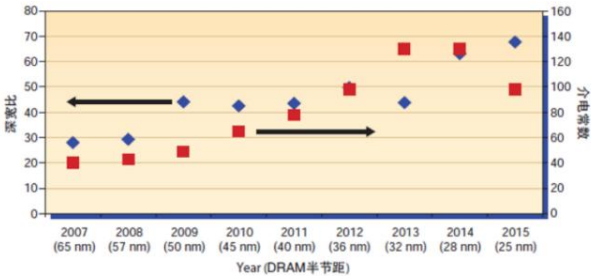
图20：ALD 形成扩散阻挡层



资料来源：Electrochimica Acta，国信证券经济研究所整理

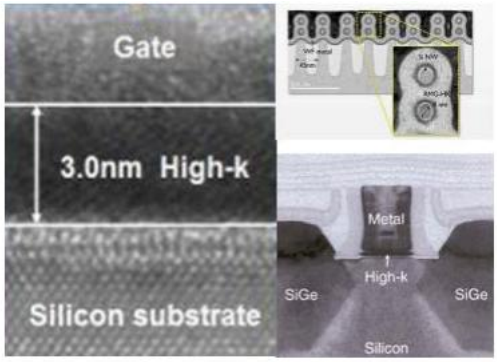
ALD 沉积 DRAM High-K Metal Gate。随着晶体管体积持续缩小，传统 SiO_2 栅极电介质受介电性能达到极限，在 45nm 内先进制程芯片中会产生隧穿现象从而导致漏电，从而造成晶体管可靠性下降。High-K 材料相比传统 SiO_2 具有更强介电常数，可使栅极漏电流减少 10 倍左右，同时降低工作电压，提高材料理论性能。

图21：先进 DRAM 需要更高介电常数材料



资料来源：CNKI，国信证券经济研究所整理

图22：ALD 形成 High-K Metal Gate

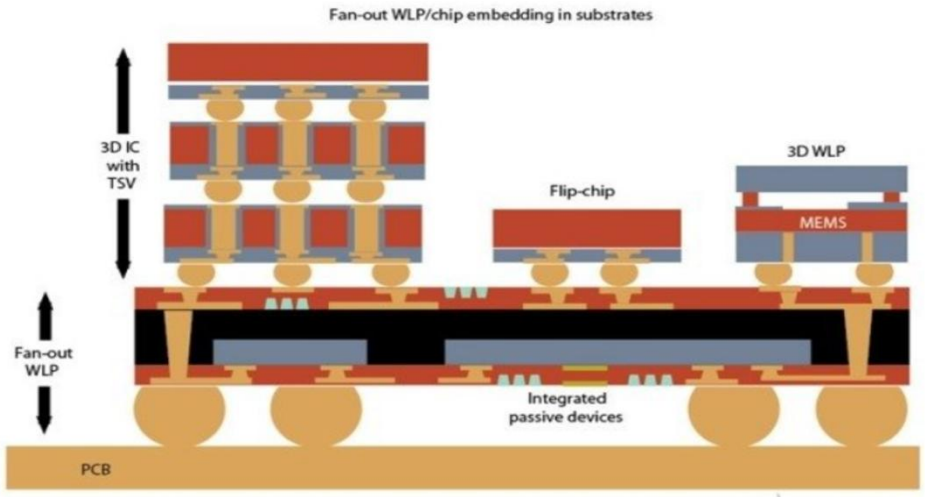


资料来源：微导纳米路演资料，国信证券经济研究所整理

关键技术#3：2.5D+3D 集成

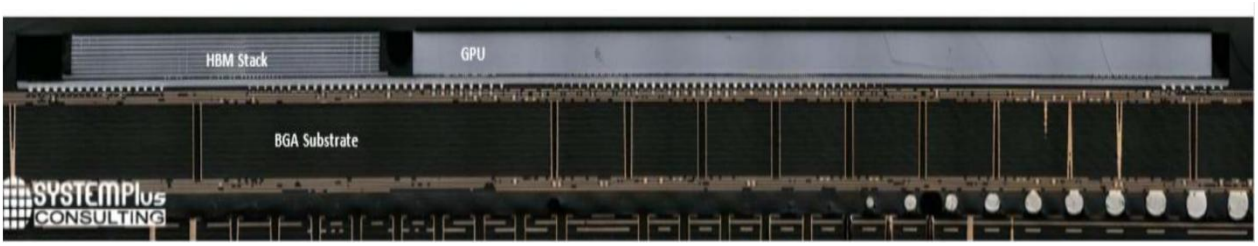
借助 TSV 可以实现 2.5D+3D 的集成技术，HBM 借助 TSV 技术实现多个 DRAM 之间的连通堆叠。2013 年 10 月，HBM 成为了 JEDEC 通过的工业标准，第二代 HBM2 于 2016 年 1 月成为了工业标准，NVIDIA 的 Tesla 运算加速卡 Tesla P100、AMD 的 Radeon Vega、Intel 的 Knight landing 都采用了 HBM2。AMD Radeon Vega GPU 中使用的 HBM2，由 8 个 8Gb 芯片和一个逻辑芯片通过 TSV 和微凸点垂直互连， 每个芯片内包含 5000 个 TSV，在一个 HBM2 中，超过 40000 个 TSV 通孔。

图23：2.5D+3D 先进封装集成



资料来源：电子发烧友网，国信证券经济研究所整理

图24：AMD Radeon Vega GPU & HBM2 横截面

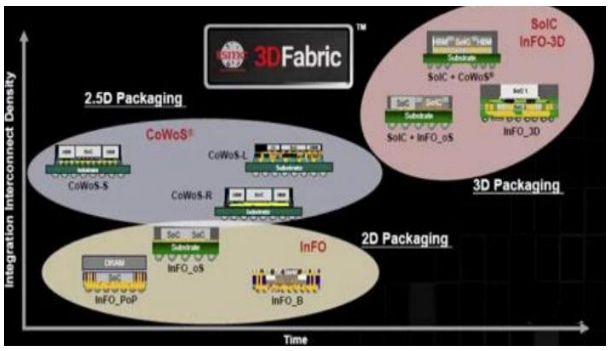


资料来源：AMD，国信证券经济研究所整理

台积电借 2.5D+3D 封装技术推出“3D Fabric”先进封装平台。台积电将 2.5D 和 3D 先进封装技术整合为“3D Fabric”平台，在 2.5D 层面推出了 CoWoS 及 InFO 等技术，在 3D 层面推出了 3D SoIC 技术。其中前段技术包含 3D 的整合芯片系统（SoIC InFO-3D），后端组装测试相关技术包含 2D/2.5D 的整合型扇出（InFO）以及 2.5D 的 CoWoS 系列。目前最新的第五代 CoWoS-S 封装技术，将增加 3 倍的中介层面积、8 个 HBM2e 堆栈（容量高达 128GB）、全新的硅通孔（TSV）解决方案等，有望将晶体管数量翻至第 3 代封装解决方案的 20 倍。

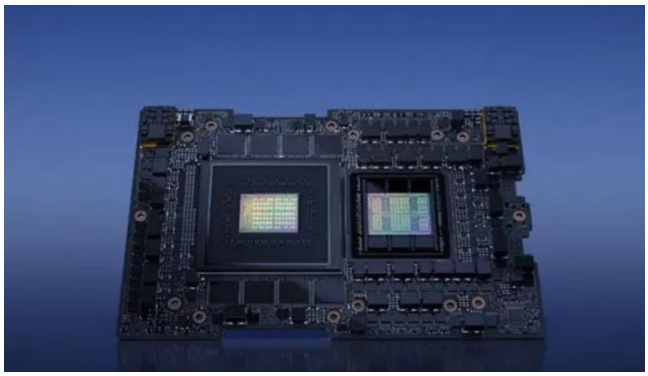
英伟达 GH200 Grace Hopper 超级芯片采用 2.5D+3D 封装技术。2023 年 5 月，英伟达正式发布了全新的 GH200 Grace Hopper 超级芯片，以及基于 NVIDIA NVLink Switch System 驱动的拥有 256 个 GH200 超级芯片的 NVIDIA DGX GH200 超级计算机。Grace Hopper 超级芯片使用 NVIDIA NVLink®-C2C 互连技术，DGX GH200 将基于 Arm 的 NVIDIA Grace CPU 和 Hopper GPU 架构互联，通过 Chiplet 工艺将 72 核的 Grace CPU、Hopper GPU、96GB 的 HBM3 和 512 GB 的 LPDDR5X 集成在同一个封装中。

图25：台积电“3D Fabric”平台使用 8 个 HBM2e 堆栈



资料来源：芯智讯，国信证券经济研究所整理

图26：NVIDIA GH200 Grace Hopper 芯片中使用 96GB HBM3 堆栈



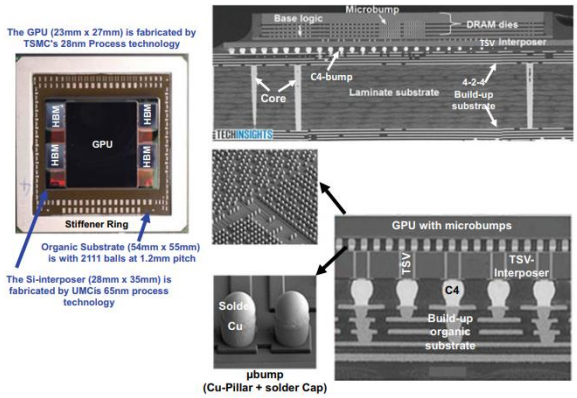
资料来源：NVIDIA，国信证券经济研究所整理

先进封装技术发展推动 IC 载板需求。IC 封装基板又称 IC 载板，是集成电路先进封装环节的关键载体和材料，建立 IC 芯片与 PCB 板之间的讯号连接，可为芯片提供电连接、保护、制成、散热等功效，相较于普通 PCB，IC 载板具有高密度、高脚数、高性能、小型化及轻薄化特点。

在目前应用较广的 2.5D+3D 的先进封装集成电路中，都采用 IC 载板作为承载芯片的转接板，如 AMD2015 年推出的 Radeon R9 Fury X GPU 中使用了 64nm 的 TSV IC 载板作为转接板，以 ubump 的形式连接上方 GPU 和 HBM，再以 C4 bump 连接下方

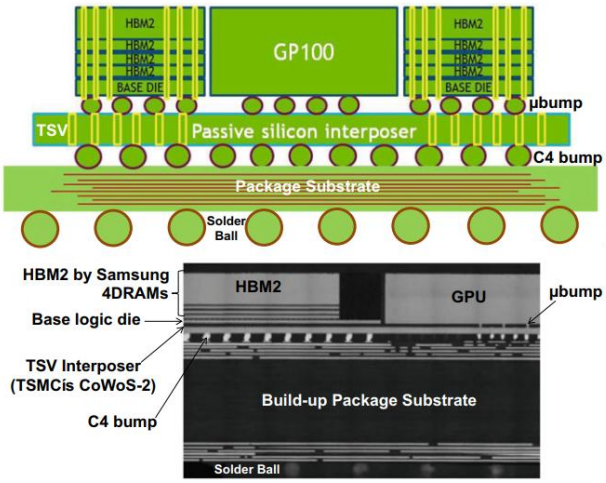
PCB 板，NVIDIA 的 Pascal 100 GPU 基于台积电 16nm 工艺技术构建，以 ubump 的形式和 HBM 以 ubump 形式连接在台积电 64nm CoWoS-2 转接板上，然后封装在 PCB 板上完成搭建。

图27：AMD/UMC 2.5D+3D 集成示意图



资料来源：Lau. J，《Chip Design and Heterogeneous Integration Packaging》，2023 版，152-157 页，国信证券经济研究所整理

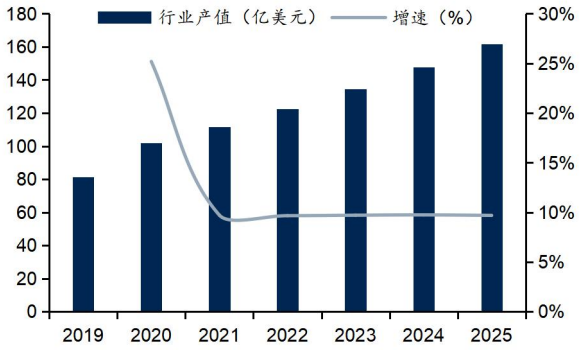
图28：NVIDIA/TSMC 2.5D+3D 集成示意图



资料来源：Lau. J，《Chip Design and Heterogeneous Integration Packaging》，2023 版，152-157 页，国信证券经济研究所整理

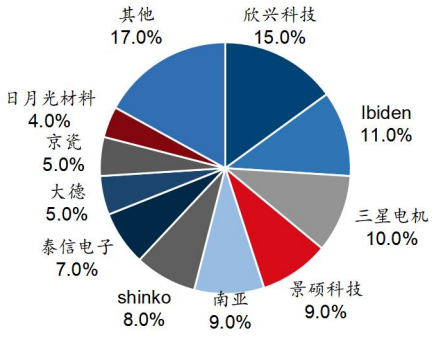
IC 封装基板是 PCB 行业增速最快的细分领域。IC 封装载板下游应用广泛，根据 Prismark 数据，2020 年 IC 载板行业产值已达到 102 亿美元，预计将在 2025 年达到 162 亿美元的产值，2021-2026 年年均复合增长率为 9.7%，高于 PCB 行业 5.8%年均复合增长率，是 PCB 行业增速最快的细分领域。IC 载板已经成为封装工艺价值量最大的材料，在传统封装中约占 40%-50%，在先进封装中约占 70%-80%。据 Yole 数据，预计全球先进封装将从 2019 年的 290 亿美元，增长到 2025 年的 420 亿美元，占封装市场整体份额的 49.4%，复合增速达 6.6%。

图29：2019-2025 全球封装基板行业产值及增速



资料来源：Prismark，国信证券经济研究所整理

图30：全球 IC 载板市场格局



资料来源：Yole，国信证券经济研究所整理

相关企业

雅克科技（002409.SZ）：雅克科技是国内 ALD 沉积主要材料前驱体供应商，**也**是全球领先的前驱体供应商之一，公司前驱体产品供应 HBM 核心厂商 SK 海力士，High-K、硅金属前驱体产品覆盖先进 1bDRAM、200 层以上 3DNAND 以及 3nm 先进逻辑电路等，应用于 AI 服务器的 HBM3 中堆叠 8 至 12 个 DRAM 裸片，AI 发展驱动 HBM 市场扩容，对应 ALD 前驱体需求将大幅增长，公司有望在 HBM 行业发展下持续受益。

中微公司（688012.SH）：中微公司是国内 TSV 设备主要供应商，研发能力强，技术经验丰富，早在 2010 年就推出了首台 TSV 深孔硅刻蚀设备 Primo TSV®，每台系统可配置多达三个双反应台的反应腔。每个反应腔可同时加工两片晶圆。中微提供的 8 英寸和 12 英寸硅通孔刻蚀设备，均可刻蚀孔径从低至 1 微米以下到几百微米、深度可达几百微米的孔洞，并具有工艺协调性，可根据客户的需求产生不同的刻蚀形状（例如垂直、圆锥形和锥形等）。Primo TSV 还具有多种新颖的功能，诸如预热反应台、晶圆边缘保护环、低频射频脉冲、侧引入气体均匀化技术等，为 TSV 应用提供所需的高技术、灵活性和生产能力。公司也是国内等离子体刻蚀设备的核心供应商。

拓荆科技（688072.SH）：拓荆科技是国内 ALD 设备的主要供应商之一，公司 PEALD 产品用于沉积 SiO_2、SiN 等介质薄膜，用于填孔、侧墙、衬垫层等工艺，其 PF-300TAstra 完成产业化验证，NF-300HAstra 在客户端验证顺利；Thermal-ALD 产品（PF-300TAltair、TS-300Altair）已完成研发，目前出货至不同客户端进行验证，主要用于沉积 Al_2O_3 等金属化合物薄膜，ALD 设备有望成为公司新的业绩增长点。

兴森科技（002436.SZ）：公司是国内 IC 载板行业龙头，公司 2012 年开始布局 IC 载板，2018 年成为三星在国内唯一 IC 载板供应商。目前拥有 CSP 封装基板产能 3.5 万平方米/月，其中，广州基地 2 万平方米/月的产能满产满销，盈利能力稳定；珠海兴科与大基金合作项目规划 4.5 万平方米/月新产能，于 2022 年二季度末已建成 1.5w 平方米/月新产能。公司在 2022 年分别斥资 60 亿元和 12 亿元人民币在广州和珠海投资建设 FCBGA 封装基板生产和研发项目，广州拟建设设月产能为 2000 万颗的 FCBGA 封装基板智能化工厂，占地 8 万平方米，预计最早 2025 年达产。珠海拟建设产能 200 万颗/月（约 6000 平方米/月）的 FCBGA 封装基板产线，目前进展顺利，预期珠海 FCBGA 项目于 22 年底完成产线建设，23 年完成量产出货。

深科技（000021.SZ）：深科技于 2015 年 6 月以 1.1 亿美元价格收购沛顿科技进入存储芯片封测领域，沛顿科技是全球第一大独立内存制造商美国金士顿科技公司于国内投资的外商独资企业，专门从事动态随机存储（DRAM）芯片封装和测试业务。公司积极布局高端封测工艺，规划建设凸块（Bumping）项目，目前同步进行净化间施工和首线设备采购。同时持续优化倒装工艺（Flip-chip）、POPt 堆叠封装技术的研发及 16 层超薄芯片堆叠技术，是国内唯一通过 Intel CPU 架构存储认证的企业，所有测试过的存储芯片产品可直接配套 Intel 服务器。

兆易创新（603986.SH）：公司 NOR Flash 市场排名全球第三，国内第一，并且持续在 NOR Flash 市场发力，中大容量 NOR Flash 客户群和覆盖面不断扩大，需求持续稳定；同时公司积极切入 DRAM 存储器利基市场，同长鑫存储关系密切，并已推出 DDR4、DDR3L 等产品，产品覆盖消费电子、工业安防、网络通信等多个领域。

北京君正（300223.SZ）：公司是国内领先的 IC 设计厂商，以“计算+存储+模拟”

平台化发展。目前在嵌入式 CPU 技术、视频编解码技术、影像信号处理技术、神经网络处理器技术、AI 算法技术、高性能存储器技术、模拟互联技术、车规级芯片设计技术等领域形成了多项核心技术。2020 年并购北京矽成切入存储和车规级市场。公司目前形成“计算+存储+模拟”的技术和产品格局，在汽车电子、工业、医疗、安防监控、IOT 等重点应用领域均有积极布局。

风险提示

1、宏观 AI 应用推广不及预期。AI 技术在应用推广的过程可能面临各种挑战，比如：（1）AI 技术需要更多的时间来研发和调试，而且在应用过程中可能会受到数据质量、资源限制和技术能力等因素的制约；（2）AI 技术的实施需要更多的资源和资金支持；（3）市场竞争可能也会影响企业在 AI 应用推广方面的表现。因此，投资者应审慎评估相关企业的技术实力、资金实力以及管理能力，相关企业的 AI 应用存在推广进度不及预期的风险。

2、AI 投资规模低于预期。尽管 AI 技术在过去几年中受到广泛关注，但 AI 相关领域的企业投资回报并不总是符合预期。部分企业在 AI 领域可能缺乏足够的经验和资源，难以把握市场机会。此外，市场竞争也可能会影响企业的投资力度。因此，存在 AI 领域投资规模低于预期，导致企业相关业务销售收入不及预期的风险。

3、HBM 渗透率提升低于预期。虽然 HBM 的应用已经较为广泛，但 HBM 渗透率提升的速度存在低于预期的风险，这与企业对 HBM 的投资意愿有关，也可能与市场需求和技术进展的速度有关。

免责声明

分析师声明

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

国信证券投资评级

类别	级别	说明
股票投资评级	买入	股价表现优于市场指数 20%以上
	增持	股价表现优于市场指数 10%-20%之间
	中性	股价表现介于市场指数 ±10%之间
	卖出	股价表现弱于市场指数 10%以上
行业投资评级	超配	行业指数表现优于市场指数 10%以上
	中性	行业指数表现介于市场指数 ±10%之间
	低配	行业指数表现弱于市场指数 10%以上

重要声明

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。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意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制

作证券研究报告，并向客户发布的行为。

国信证券经济研究所

深圳
深圳市福田区福华一路 125 号国信金融大厦 36 层
邮编：518046　总机：0755-82130833

上海
上海浦东民生路 1199 弄证大五道口广场 1 号楼 12 层
邮编：200135

北京
北京西城区金融大街兴盛街 6 号国信证券 9 层
邮编：100032