

In-depth long article | Why is the memory product HBM selling like crazy? – January 23, 2024

深度长文 | 内存产品HBM为何卖疯了？

XSY芯师爷

最新更新时间：2024-01-23

阅读数：7553

本报告转载自招商电子，原标题《AI服务器存储量价齐升，算力需求推动HBM市场数倍增长》。报告篇幅较长，建议先收藏，分段阅读。

报告概述

AI算力升级带动服务器的CPU迭代并提升GPU需求，带动AI服务器存储容量和价值量较传统服务器数倍增长。训练型AI服务器中GPU承担大部分算力，算力要求推动了HBM等新型存储器超百亿美元新兴市场，进而提升Bumping、TSV、CoWoS等先进封装工艺需求，并带来减薄、键合、模塑、测试等设备以及EMC、电镀液、PSPI等材料的增量需求。叠加国内自主可控需求持续增长，国内存储及HBM等催生的先进封装产业链发展空间巨大。

AI服务器CPU和GPU随算力需求而升级，对存储器容量和价值量均有数倍拉动。传统服务器以CPU作为算力核心，随着AI训练模型的算力要求不断提升，CPU的核心数、主频、线程数量均不断提升，但仅靠CPU已经无法满足算力需求，需要搭配GPU进行多线程数据处理，主流训练型服务器一般搭配8个GPU。AI服务器用到的主要存储器包括CPU内存、GPU显存和硬盘NAND等，存储器容量和价值量均较普通服务器有数倍提升，

1) DRAM：英伟达训练型AI服务器中的CPU DRAM容量高达2TB，另外单个GPU一般搭载80GB以上的HBM存储器，AI服务器HBM总容量预计超640G，总内存容量相较普通服务器有4-8倍的提升，仅CPU内存价值量预计有5倍的提升，GPU的HBM则为纯增量市场；另外，服务器内存也在不断迭代，目前普通的服务器均多配备DDR4，但最先进的AI服务器已经搭配了DDR5或LPDDR5；

2) NAND：AI服务器的硬盘容量高达30TB，相较传统服务器提升2-4倍，另外传统服务器同时使用机械硬盘和固态硬盘（SSD），但AI服务器基本全部使用SSD，整体价值量较普通服务器预计提升10倍左右。

HBM能够突破训练型AI服务器的GPU带宽极限，2024年增量空间预计超百亿美元。 HBM (High Bandwidth Memory) 即高带宽存储器，是基于2.5/3D封装技术的一种新型CPU/GPU内存芯片，将DRAM Die垂直堆叠，Die之间通过TSV的方式连接。HBM能够以低功耗产生高带宽，因此广泛搭配训练型AI服务器的GPU使用，训练型AI服务器对HBM需求的拉动主要体现在：

- 1) AI服务器搭载GPU数量的提升：由普通服务器的2个提升至目前的8个；
- 2) 单个GPU搭载HBM Stack数量的提升：在HBM1方案中，单个GPU搭载4个HBM1，而在目前HBM2e或HBM3方案中，一般单个GPU搭配6个HBM Stack；
- 3) HBM堆叠的DRAM层数和容量增多：从HBM1到HBM3，单个DRAM Die密度从2Gb提升至16Gb，堆叠高度从4Hi提升至12Hi，单个HBM叠层容量从1GB提升至24GB。Trendforce预计2025年全球服务器出货量为1700万台，当前AI服务器渗透率大概不足2%，假设2024年AI服务器渗透率约4%，按照每个AI服务器搭载8个GPU、每个GPU搭载6个共80GB至100GB及以上的HBM Stack的方案测算，那么2024年AI服务器带来的HBM增量空间预计超百亿美元。

AI服务器的GPU采用2.5D+3D封装工艺，推动TSV、CoWoS等核心封装技术需求。HBM和GPU采用2.5D+3D封装工艺，根据Yole，2021年HBM和Si中介层封装市场规模合计约14亿美元，预计2027年增至35亿美元，其中HBM和硅中介层封装市场分别增至16.3和18.8亿美元。TSV (Through-Silicon Via) 即硅通孔技术，顺应2.5D封装架构而产生，能够以最低的能耗提供极高的带宽和密度，是实现电路小型化、高密度、多功能化的首选解决方案。2.5D TSV技术已经广泛用于AI GPU基板上的HBM中，实现DRAM各层Die之间的连接，以及HBM芯片和下方的金属凸块之间的连接。CoWoS工艺用于将HBM和硅中介层、封装基板等进行整体封装，当前台积电处于领先地位，伴随着谷歌TPU、英伟达GPU、AMD MI300等均导入生成式AI，台积电CoWoS需求自2022年以来翻倍增长，目前持续供不应求，展望2024年将目前CoWoS产能翻倍。

HBM多层堆叠结构提升工序步骤，将带动封装设备和材料需求持续提升。

- 1) 设备：HBM中大量增加前道工序，前道检、量测设备主要增量来自微凸点、TSV、硅中介层等工艺，另外HBM中增加的预键合晶圆级测试和KGSD相关的封装级测试也带动分选机、测试机、探针台等后道测试设备的数量和精度提升；HBM堆叠结构增多，要求晶圆厚度不断降低，进而提升减薄、键合等设备需求；HBM多层堆叠结构要求超薄晶圆和铜-铜混合键合工艺，增加了临时键合/解键合以及混合键合设备需求，各层DRAM die的保护材料也非常关键，对注塑或压塑设备提出较高要求；另外，诸如划片机、固晶机、回流焊机/回流炉等传统设备需求也均受益于HBM封装带来的工艺步骤提升和工艺变革带来的价值量提升；

2) 材料：HBM中芯片间隙采用GMC（颗粒状塑封料）或LMC（液态塑封料）进行填充，GMC最主要原材料为球形硅微粉和球形氧化铝；HBM采用底部填充胶用于FC封装工艺，采用PSPI作为硅中介层中RDL的再钝化层；HBM中的Bumping、RDL、TSV等引入前道工艺，带来电镀液用量提升；另外，HBM也将提升电子粘合剂、封装基板、压敏胶带等其他材料需求。

AI服务器相较普通服务器算力明显提升，存储量价均有数倍增长空间

1、传统服务器采用CPU作为算力来源，一般配置DDR4内存和SATA/SAS等类型SSD

传统服务器以CPU作为算力来源，采用提高核心数来提升算力。传统服务器组件包括CPU、内存、硬盘、RAID卡、网卡等，搭配BMC管理系统，以及电源、主板、机箱等基础硬件。按照外形分类，服务器可分为塔式、刀片、机架和高密度服务器，按照处理器数量，可分为单路、双路及多路。传统服务器以CPU提供算力，擅长逻辑和浮点型计算等，服务器的算力提升主要依靠CPU增加核心数量。

CPU和内存是普通服务器逻辑架构中最重要的部分，等效频率是内存的核心指标。在服务器中，CPU负责对数据进行逻辑运算，随着CPU主频、核心和线程数量的不断提升，CPU运算处理的速度已经远超硬盘的读写速度，因此需要CPU内存作为桥梁来暂时存放CPU中的运算数据，并与硬盘等外部存储器交换数据，CPU和内存之间通过内存总线进行通信。由于内存是能够传输和储存数据的中转站，传输速度和处理性能尤其重要，核心指标为频率。内存的频率分为三种，①核心频率：指内存颗粒核心的时钟频率，核心频率是固定的；②工作频率：指内存颗粒总线的时钟频率；③等效频率：指内存实际的数据传输率，也被称为内存速率。

服务器的DRAM类型伴随CPU而升级，目前服务器标配DDR4或DDR5。SDRAM最早出现，为了与CPU的计时同步而设计，在一个周期内，SDRAM只能在升的时候进行数据传输，即只能读写一次；改进后的DDR能够在升和降两个阶段进行数据传输，所以等效频率相较SDRAM增加1倍；DDR1到DDR3的升级主要在于预读取的能力，预读取从1bit上升至8bit，每进步一代，预读取能力增加一倍，相当于一次读写的数据数量

增加一倍，即工作频率翻倍；相较DDR3，DDR4采用了2-4个Bank Group组的设计，即在同一脉冲工作周期内，DDR4最多可以处理2-4组数据，相当于等效工作频率提升2-4倍；DDR5将预读取数提升至16bit，等效频率再次翻倍。从当前服务器型号来看，传统服务器的CPU内存一般标配DDR4，最先进的AI服务器CPU则配备了DDR5或LPDDR5。

GDDR在DDR的基础上发展形成，主要用于PC端的显卡。GDDR/GDDR2基于原有的内存技术，但无法满足GPU对高性能存储的需求，进而推动了GDDR3作为第一款专为GPU而设计的存储器的诞生。GDDR3采用4-bit预读取，能够实现读写操作的快速切换；GDDR4采用8-bit预读取技术，但由于NVIDIA不愿跟进，最终并未得到广泛应用；GDDR5使用了双DQ总线，实现了双倍的接口带宽，自GDDR5开始，GDDR广泛用于PC显卡中，但由于传统服务器中不配备GPU，因此GDDR并未在传统服务器中得到使用。

服务器硬盘主要包括HDD和SSD，SSD能够大幅提升服务器性能。目前服务器的硬盘主要包括SATA/SAS等机械硬盘（HDD）以及固态硬盘（SSD）等，HDD主要由机械驱动，包括马达、盘片、磁头摇臂等部件，尽管容量不断提升，但磁盘记录方式并未改变，存储系统的发展越来越落后于CPU的发展速度；SSD固态硬盘没有机械结构，主要由主控-内存-固件算法构成，SSD以区块写入和擦除的方式实现读写的功能，具备低功耗、稳定性高、耐震耐低温等特点，可以大幅提升服务器性能，更适合工业自动化、嵌入式移动设备和AI推理/训练服务器等应用。

RAID卡可实现对固态硬盘数据的冗余保护，已成为AI服务器的重要配置。RAID（Redundant Array of Independent Disks），即独立硬盘冗余阵列，意为很多块独立磁盘构成的磁盘组，主要具备如下功能：①通过对硬盘上数据的条带化，实现对数据成块存取，减少硬盘的机械寻道时间，提高数据存取速度；②通过对一个阵列的几块硬盘同时读取，减少硬盘的机械寻道时间，提高数据存取速度；③通过镜像或存储奇偶校验信息的方式，实现对数据的冗余保护。由于硬件RAID是采用集成的阵列卡或专用的阵列卡来控制硬盘驱动器，因此可以极大节省服务器系统CPU和操作系统的资源，目前RAID卡已成为AI服务器的标配。

2、AI服务器中CPU+GPU协同工作，存储容量和价值量相较传统服务器均有数倍提升

AI训练模型算力提升速度突破极限，目前英伟达训练型AI服务器一般配备8个GPU。随着以chatGPT为代表的AI的发展，训练GPT-3、Megatron-Turing NLG 530B等超大语言模型所要求的算力提升速度已经突破了后摩尔定律算力提升速度的极限，尽管CPU不断升级，但CPU制程以及单个CPU和核心数量接近极限，仅依靠CPU无法满足算力需求。

CPU的内核数量大约数十个，但GPU具备成千上万个CUDA核心，因此GPU多个内核决定了其能够在相同的价格和功率范围内，比CPU提供更高的指令吞吐量和内存带宽，GPU能够并行执行成千上万个线程（摊销较慢的单线程性能以实现更大数据吞吐量）。在训练AI模型的过程中，需要同时对所有样本数据执行几乎相同的操作，GPU架构设计能够很好满足AI场景需求。AI服务器相较通用服务器的一个明显差别之一是增加了GPU，通用服务器一般含有1-2个CPU、不含GPU，而当前英伟达训练型AI服务器一般搭载8个GPU。

AI服务器GPU需要CPU来进行指令，模型算力提升带动CPU核心、主频等提升。在AI服务器的GPU模式下，模型训练一般分为4步，①将输入数据从系统内存拷贝到显存；②CPU指示GPU处理数据；③GPU并行完成一系列计算；④将计算结果从显存拷贝到内存。虽然GPU并行能力优异但无法单独工作，必须由CPU进行控制调用，CPU可以独立工作并直接访问内存数据完成计算。

因此在AI服务器中，GPU和CPU需要协同工作，训练模型所需算力升级也将带动CPU技术升级，例如在英伟达DGX-2服务器中，采用英特尔第三代至强处理器8168，主频大约2.7GHz，核心数量为24个；在英伟达DGX H100服务器中，搭载英特尔第四代至强处理器8480C，主频提升至最高3.8GHz，CPU核心数量提升至大约56个。

AI服务器存储器容量伴随CPU/GPU的升级而提升，相较传统服务器有数倍提升。最先进的AI服务器尽管增加了大量GPU需求，但存储器的数据存储方式、总线连接方式均和普通服务器相近，CPU的运行数据写入DRAM中，CPU和GPU产生的数据共同写入NAND中。AI服务器将提升内存、显存的工作频率和带宽等，带动存储容量明显上升。

DRAM：AI服务器CPU+GPU相较普通/高性能服务器对DRAM容量大约有3-8倍的拉动。当前服务器一般配1-2个CPU，每个CPU一般标配DDR4内存，普通双路服务器用于打印、网络管理、数据记录等场景，一般配备1-2个CPU，每个CPU周围搭配大约10个内存插槽，按照单个DDR4 32GB容量计算，每个CPU搭配的内存容量大约不超过320GB，系统容量不超过640GB；根据浪潮官网，高性能服务器用于涉及高密度计算的企业端，例如搭载双路CPU的NF5280M6型号服务器单个CPU插槽数量为16个，按照每个DDR4 32GB、插满内存条计算，系统的内存容量可达1TB；英伟达AI服务器CPU均标配2TB的DDR4/DDR5；另外，AI服务器增加了GPU的数量，每个DGX A100和H100均需要搭配80GB容量的HBM，系统整体需要640GB的容量。综上所述，从容量上看，按照普通服务器CPU内存320-640GB容量计算，AI服务器对DRAM容量的提升大约为4-8倍；按照高性能服务器CPU内存1TB容量计算，AI服务器对DRAM容量的提升大约为2.5-3倍。

NAND：AI服务器相较普通/高性能服务器对NAND容量大约有2-4倍的拉动。由于机械硬盘持续进行成本优化，普通服务器依然会配备较多的机械硬盘，而高性能服务器和AI服务器对于存储速度、准确性等提出更高要求，因此主要配备固态硬盘。从容量上看，传统服务器NP5570M5最多支持16块2.5英寸SAS硬盘或4块3.5英寸SATA硬盘，SAS硬盘支持300GB/600GB/1TB/1.2TB/1.8TB/2.4TB容量，SATA硬盘支持1/2/4/6/8TB容量，按照中间容量测算，系统硬盘容量大约共8-16TB；高性能NF5280M6型号最大支持20块3.5英寸硬盘，系统硬盘容量大约20TB；AI服务器采用的硬盘容量大约30TB，相较传统服务器NAND容量提升大约2-4倍。

AI服务器的GPU和存储价值量均明显提升，GPU价值量占比最高。

1) 基础型服务器：根据IDC，基础型服务器中CPU、内存和硬盘的占比最高，根据英特尔，典型的x86服务器—E5高配服务器的成本中硬盘、CPU和内存成本占比分别为31%、28%和21%。例如浪潮NF5270M5售价大约39000元，采用2颗Xeon Silver 4214 CPU，每颗售价大约1000美金，CPU成本占比大约32%；配置16个32GB DDR4，按单GB成本3美金计算，DRAM成本占比大约26%；最大支持25块2.5英寸硬盘，按照每块硬盘512GB、1TB SATA 40美元计算，硬盘总价值量占比大约20%；

2) AI服务器：以英伟达DGX A100系统为例，售价大约19.5万美金，含有2颗AMD Rome 7742，每颗售价大约7000美金，价值量占比大约7%；含有8颗A100 GPU，每颗GPU售价大约1-1.5万美元，GPU价值量占比大约40-50%；CPU采用2TB的DDR4，价值量占比大约3-4%；GPU配置共640GB HBM2E，假设单GB价值量15-20美金，价值量占比大约5-8%；操作系统配备2块1.9TB SSD，内部配备8块3.84TB SSD，按照1块1.92TB数据中心SSD 650美元、1块3.84TB数据中心SSD 1200美元测算，硬盘价值占比大约6%。

综上所述，从价值量来看，AI服务器中CPU内存和SSD的价值量相较于普通双路服务器预计均提升5倍左右，GPU显存为纯增量市场。

3、存算一体突破存储和功耗“双墙”瓶颈，催生2.5/3D封装及新型存储器需求

传统服务器计算芯片与存储器采用并行架构，“存储墙”和“功耗墙”分别限制数据访问速度提升和数据访问能耗下降。在冯-诺依曼架构中，计算和存储功能分别由CPU和存储器完成，数据从存储单元外的存储器中获取，处理完毕后再写回存储器，计算核心与存储器之间有限的总带宽直接限制了交换数据的速度，处理器速度和访问速度的差异进一步减缓处理速度。一方面，处理器和存储器制程、工艺不同，性能差距逐渐增大，存储器数据访问速度远低于CPU的数据处理速度，即形成“存储墙”；另一方面，数据搬运的能耗比浮点计算高1-2个数量级，DRAM的访问功耗可达芯片内一级功耗的50-100倍，进一步增加了数据访问功耗。

存算一体架构可部分解决“双墙”问题，主要包括近存计算和存内计算。

1) 近存计算：主要包括光互联、2.5/3D封装等形式，其中2.5/3D封装可以提高系统集成度，增大并行带宽或利用串行传输提升存储带宽。典型的产品形态为高带宽内存（High Bandwidth Memory, HBM），将内存颗粒通过硅通孔（TSV）多层堆叠实现存储容量提升，同时基于硅中介板的高速接口与计算单元互联提供高带宽存储服务；

2) 存内计算：主要包括①存储内计算（In-Storage Computing, ISC），即在固态硬盘中等加入FPGA、ARM处理核等计算单元，承担如数据压缩、搜索、视频文件转码等本地处理，减少远端处理器如CPU的负载；②内存内计算（In-Memory Computing, IMC），数据直接在内存中计算后返回，通过将存储层和逻辑层堆叠实现高带宽计算。典型产品形态包括HBM-PIM和PIM-DIMM，在DRAM die中内置AI等处理单元，提供大吞吐低延迟片上处理能力。

存内计算电路可基于易失性和非易失性存储器实现，未来RRAM、MRAM等新型存储器件有望得到更多应用。易失性存储器包含SRAM等，非易失存储器包含NOR Flash、阻变随机存储器（Resistive Random Access Memory, RRAM）、磁性随机存储器（Magnetoresistive Random Access Memory, MRAM）、相变存储器（Phase Change Memory, PCM）等。

SRAM：常用于CPU中的缓存，其读写速度快、无读写次数限制，且工艺成熟，器件的一致性和稳定性较好，缺点是价格相对昂贵、多晶体管单元结构下存储密度较低、需要通电以保持数据，因此芯片面积大、功耗较高；

NOR：基于NOR的存内计算芯片具有存储密度大、工艺成熟、成本低等优点，业内主要利用其多比特存储特性进行模拟存内计算，相关产品在智能终端与边缘场景实现小批量商用。目前NOR的问题在于工艺微缩存在一定瓶颈，可能需要2.5/3D封装工艺配合扩展性能；

阻变随机存储器（RRAM）：基本存储单元一般为金属-绝缘体-金属/半导体结构，通过在电极层施加电压/电流，电阻转变层的电阻值可以实现高阻态和低阻态的切换，并可实现多级电阻状态，从而存储多比特信息。基于RRAM的存内计算芯片制备工艺简单、工艺成本低、支持多比特存储，但目前在器件一致性和准确性等方面需要提高；

磁性存储器（MRAM）：利用磁隧道磁阻效应实现电阻状态改变，从而存储信息，具有非易失、访问速度快、读/写次数高等优点，具备较高的可靠性和稳定性，但目前成熟度较低，功耗、速度和耐久性等指标有待提升；

相变存储器（PCM）：可以介于完全多晶态和完全非晶态之间的多种状态，以此实现多值存储，具备存储密度高、多比特存储、支持3D堆叠等优点，但PCM存在写入功耗较大、擦写次数较少等问题。

目前存内计算主要采用NOR、SRAM、DRAM等成熟存储介质，新型存储器在大算力领域应用尚不成熟。目前HBM等近存计算已经规模商用，但大部分存内计算采用模拟计算或使用DRAM等成熟存储介质，新型存储器落地场景仍相对有限。

在低功耗、低精度和中小算力场景，多使用NOR、SRAM等工艺；在高精度、大算力的AI芯片中，目前RRAM、MRAM等新型存储器技术指标尚在提升，因此新型存储器尚未规模应用，应用方案以HBM、HBM-PIM等为主，例如三星、海力士等均采用2.5D封装形式，将DRAM die堆叠形成HBM，进而提高集成度；三星于2021年推出结合DRAM的高带宽内存-内存内处理（HBM-PIM），将AI计算能力引入内存，通过将经过DRAM优化的AI引擎置于每个存储子单元内，将处理能力直接引入到数据的存储位置，从而实现并行处理并尽可能减少数据移动。

HBM新型存储器突破带宽极限，训练型AI服务器推动超百亿美金增量市场

1、HBM用于高带宽和高容量场景中，已普遍搭配主流AI训练芯片使用

HBM是基于2.5/3D封装技术的新型存储器，满足高带宽、高速度等需求。HBM（High Bandwidth Memory）指高带宽存储器，是一款新型的CPU/GPU内存芯片，基于2.5D/3D封装技术将DRAM Die垂直堆叠，具备高带宽、高速度等特点。DRAM Die之间通过TSV的方式连接，逻辑控制单元对DRAM进行控制，GPU和DRAM之间通过uBump和Interposer（起互联功能的硅片）连通。目前最先进的HBM为第五代HBM3以及第六代HBM3E，封装的DRAM Die层数达到12层。

GPU显存一般采用GDDR或HBM两种方案，HBM能够在实现高带宽和高容量的同时节约芯片面积和功耗，非常适合在功率受限但又需要高带宽的场景中使用。

1) 传统的GDDR方案：主要通过提升GDDR的带宽实现性能的提升，但存在如下问题：①GDDR不断增长的功耗一定程度阻碍了图形性能的表现，未来可能导致GDDR的性能表现无法匹配GPU的性能要求；②为了实现GPU的高带宽，GPU需要搭配大量的GDDR5/6芯片，芯片面积不断增大，同时也需要更高的电压。

2) HBM方案：将原本在PCB板上的DDR内存颗粒和GPU芯片同时集成到SiP封装中，使内存更加靠近GPU，DDR Die采用堆叠方式，通过增加单个DDR容量和层数来实现内存容量和带宽的提高。

HBM具备如下优势，①大大节约芯片面积：根据AMD，1GB HBM相较1GB GDDR5能够节省大约94%的芯片面积；②能够在更低的频率和更少的功耗下提供超越常规内存的带宽：HBM通过同一封装内的硅中介层与SoC集成在一起，能够克服数据I/O封装管脚限制的最大数量，进而突破内存带宽的限制。例如，典型的GDDR5使用32位总线，在1750MHz、1.5V电压下，最高带宽为28GB/s，而一个HBM包使用1024位总线，仅在500MHz、1.3V的电压下，就能达到超过100GB的带宽。

HBM已普遍搭配主流AI训练芯片使用，相较GDDR带宽显著提升。HBM最早于2013年由SK海力士首先制造，第一颗采用HBM存储的GPU是2015年AMD的Radeon R9 Fury X，第一颗应用HBM2的GPU是英伟达的Tesla P100；目前，先进的AI训练GPU芯片均搭载HBM存储芯片，例如英伟达的DGX A100单GPU搭载了总计80GB容量的HBM2E，DGX H100单GPU搭载了总计80GB容量的HBM3存储，单个HBM3存储带宽

最高可达819GB/s，较单个GDDR带宽提升10倍以上，H200更是搭载了共141GB的HBM3e内存，总HBM容量较H100提升76%，总带宽提升约43%。

2、HBM容量、带宽、速度持续升级，训练型AI服务器将带来HBM超百亿美元增量市场

HBM目前最主要搭配AI的GPU使用，训练型AI服务器是最主要的增量市场。根据SK海力士，受益于HPC、AI、CPU等应用，HPC市场复合增长率达到40%左右，由于随着数据量呈指数级增长，以及AI/ML训练等高级工作负载的快速增长，预计AI服务器将是几年内HBM最大的下游市场。

训练型AI服务器对于HBM市场的拉动体现在：①AI服务器渗透率的提升；②单个AI服务器搭载GPU数量的增加；③每个GPU搭载的HBM Stack数量增加；④单个HBM堆叠的DRAM die层数（Hi）增加，以及DRAM die容量增加；⑤HBM代际进步带来单GB价值量的提升。

1) AI训练服务器渗透率提升：根据Trendforce，2022年全球服务器出货大约1430万台，其中训练侧AI服务器占比仅为1%；由于四大CSP陆续下调采购量，Dell及HPE等OEM厂商在2-4月期间下调全年出货量至同比分别减少15%和12%，Trendforce预估2023年全球服务器整机出货量将同比减少2.85%至1383.5万台。同时，考虑到云服务厂商积极投入生成式AI，AI服务器渗透率预计将快速提升；

2) GPU基板搭载HBM数量提升：搭载HBM1的GPU基板上仅采用4个HBM1芯片，而英伟达AI服务器DGX A100和H100的基板搭载了8个GPU，每个GPU搭配6个HBM2e；

3) 单个DRAM die密度增加、HBM堆叠层数和容量提升：根据SK海力士官网信息，从HBM1进步到HBM3，单个DRAM Die密度从2Gb提升至16Gb，堆叠高度从4Hi提升至最高12Hi，单个HBM容量从最高1GB提升至24GB；

4) HBM性能提升：HBM1提升至HBM3，最高速度从1Gbps提高到高于6Gbps，带宽从128GB/s提升至最高4.8TB/s。例如，SK海力士的HBM3将DRAM芯片高度减少至大约30微米，相当于A4纸厚度的1/3，使用TSV技术（在DRAM芯片上打上数千个细微孔并通过垂直贯通的电极连接上下芯片）垂直连接12个芯片，相较于HBM2e，HBM3的速度提高了大约78%，容量提升了1.5倍，并且大大降低了工作温度；

5) 价值量较传统DRAM提升：传统DDR4每GB价格大约数美元，HBM价格大约为其数倍，HBM3、HBM3E相较前几代产品价值量也大大提升。

2024年训练型AI服务器对HBM增量空间预计超百亿美元。由于推理型AI服务器对GPU和HBM需求量相对较少，因此我们仅考虑训练型AI服务器带来的HBM市场增量。假设2024年训练型AI服务器渗透率达4%，单个训练型AI服务器搭载8个GPU，目前DGX A100和H100的每个GPU配6个HBM，总容量为80GB；考虑到改进前的GH200搭载的HBM总容量为96GB，H200搭载的HBM总容量升至141GB，考虑到GPU算力持续提升，因此我们合理假设平均每个GPU搭载HBM容量持续提升；目前HBM持续供不应求，单价快速增长，但目前HBM产品以HBM2及HBM2E为主，考虑到2024年更高价值量的HBM3将持续放量，因此假设2024年价格继续上涨；考虑到2025年HBM产能逐步开出，成本有望持续下降，我们假设2025年价格有所回落。

综上所述，我们预计2024年AI服务器有望带动HBM超百亿美金增量市场。

英伟达在COMPUTEX 2023大会上公布了GH200 Grace Hopper，每个GPU使用96GB的HBM3，搭载的HBM容量继续提升。由NVIDIA GH200 Grace Hopper超级芯片和NVIDIA NVLink Switch System驱动的NVIDIA DGX超级计算机，旨在助力开发面向生成式AI语言应用、推荐系统和数据分析工作负载的巨型、下一代模型。

NVIDIA DGX GH200的超大共享内存空间通过NVLink互连技术以及NVLink Switch System使256个GH200超级芯片相连，使它们能够作为单个GPU运行。其提供1 exaflop的性能和144 TB的共享内存——相较2020年推出的上一代NVIDIA DGX A100内存大了近500倍。H200单GPU总共支持576GB的内存，其中480GB为LPDDR5X，从HBM容量来看，A100和H100单GPU最大支持80GB HBM容量，H200单GPU最大支持96GB容量的HBM3。

英伟达于11月13日上午在“Supercomputing 23”会议上正式发布了全新的H200 GPU和改进后的GH200，每个H200 GPU搭载HBM3e容量进一步提升至141GB。H200依然是建立在现有的HopperH100架构之上，但增加了更多高带宽内存（HBM3e），全新的H200提供了总共高达141GB的HBM3e内存，有效运行速度约为6.25Gbps，六个HBM3e堆栈中每个GPU的总带宽为4.8TB/s。

与上一代的H100（具有80GB HBM3和3.35TB/s带宽）相比，HBM容量提升了超过76%。改进后的GH200预计将包含总计624GB内存，相较改进前的GH200（总计576GB内存），总内存容量提升了8.3%。

长期来看，HBM在AR/VR设备、汽车自动驾驶域等领域也将有更多应用。

1) AR/VR设备：以HTC VIVE pro2为例，其连接GTX1060，当前并不需要HBM等高带宽产品，未来有望伴随显示需求增多逐步产生高带宽需求；

2) 汽车：当前自动驾驶域并不需要HBM的高带宽，比如英伟达Jetson AGX Orin采用256位的LPDDR，带宽最高为204.8GB/s。但长期来看，伴随L3/L4等级别自动驾驶汽车逐渐落地，也将提升HBM需求，例如L3级别自动驾驶带宽要求一般为600GB/s-1TB/s之间，单车可用2个HBM2e或者1个HBM；L4级别自动驾驶带宽要求一般为1-1.5TB/s之间，单车可用3个HBM2e或2个HBM3。

三

HBM采用2.5D+3D封装，核心工艺包括TSV、CoWoS等

1、HBM和Si中介层封装远期市场预计数倍增长，核心封装工艺包括Bumping、TSV、CoWoS等

HBM采用2.5D+3D封装工艺，采用的核心封装工艺包括Bumping、RDL、FC、TSV、CoWoS等。封装工艺主要有四项功能：①保护芯片免受外部冲击或损坏；②将外部电源传输至芯片，保证芯片的正常运行；③为芯片提供线路连接，以便执行信号输入和输出操作；④合理分配芯片产生的热量，确保其稳定运行。HBM采用“2.5+3D”封装工艺，即分别在DRAM die上和硅中介层上制作TSV，主要使用的封装工艺如下：

1) HBM由DRAM Die堆叠而成，Die之间借助TSV（硅通孔）和微凸块（microbump）相连，并借助TC（Thermal Compression，热压）-NCF（non-conductive film，非导电薄膜）或MR-MUF（批量回流模制底部填充）技术，用NCF或LMC（Liquid Molding Compound，液态塑封料）对芯片形成互连和保护；

2) HBM通过TSV和微凸块连接下方的HBM逻辑控制die（Base Die），逻辑控制die再通过凸块（μbump）连接下方的硅中介基板（Si interposer）；

3) HBM Stack通过CoWoS等2.5D封装工艺，和CPU/GPU等并排铺设在硅中介基板上方，CPU/GPU等逻辑die采用倒片封装（FC）形式和硅中介基板连接，存储器和GPU等逻辑芯片之间通过RDL实现通信；

4) 硅中介基板通过2.5D TSV工艺和Cu Bump连接至封装基板（Package Substrate）上；

5) 最后，封装基板通过封装锡球（package balls）和下方的PCB基板（Circuit Board）相连，这一过程采用传统的球栅阵列（Ball Grid Array, BGA）封装工艺。

2021年针对HBM和Si Interposer的高端高性能封装市场规模合计约14亿美元，预计2027年增至约35亿美元。根据Yole，高端高性能封装主要包括UHD FO、嵌入式硅桥、硅中介层（Si Interposer）、3D堆栈存储器和3D SoC，嵌入式硅桥主要解决方案包括台积电的LSI和英特尔的EMIB；硅中介层通常由台积电、三星和联电提供，还包括英特尔的Foveros方案；3D堆栈存储器分为HBM、3DS和3D NAND堆栈。

Yole预计2021年高端高性能封装市场规模约27.4亿美元，其中HBM和硅中介层市场封装规模分别为4.35亿美元和9.64亿美元；到2027年，整体市场规模预计增至78.7亿美元，其中UHD FO、HBM、3DS和硅中介层封装将在2027年占据超50%的市场份额，其中HBM和硅中介层封装预计分别增至16.3亿美元和18.8亿美元，CAGR分别为25%和12%。

2、凸块制造（Bumping）是扇入/扇出、FC等封装必要环节，推动沉积、涂胶显影、电镀等前道工艺需求增长

晶圆级封装采用凸块（Bump）取代键合引线，凸块可以分布在整个芯片表面形成信号触点。凸块工艺即在晶圆切割成单个芯片之前，在晶圆上以整个晶圆的形式形成由焊料组成的“凸块”或“球”，这些凸块是芯片和基板互连形成单个封装的基本互连组件。传统的引线键合工艺中，接合焊盘/引脚放置在外围区域，但用于凸块的I/O焊盘可以分布在芯片的整个表面，每个凸块都是一个信号触点，从而可以缩小芯片尺寸并优化电气路径，因此凸块工艺广泛用于倒装芯片封装、扇入或扇外型封装工艺。

凸块可以由共晶、无铅、高铅材料或晶圆上的铜柱组成，高端应用多采用铜柱作为凸块。含铅焊料作为凸块有多年历史，但已经被锡基无铅焊料取代并用于倒装封装等工艺，而在AI、HPC、基带、高性能存储等

应用场景，铜柱凸块（Cu Pillar Bump, CPB）逐渐取代无铅锡球，主要系CPB的精细节距（Fine Pitch）更小，无铅锡球的Fine Pitch一般为130-250um，铜柱凸块的Fine Pitch可缩小至130-40um左右，采用铜柱凸块的芯片尺寸较锡球能够缩小5-10%，衬底层可从6层减少到4层，基板成本下降约30%；铜柱凸块还具有改善的电子迁移阻力，能够减少功率损耗和信号延迟，同时实现更好的散热性能。无铅锡球整体由锡金属构成，铜柱则分为三层结构，包括铜金属柱、一层薄镍金属作为铜柱的扩散屏蔽层、锡/银金属作为焊帽。

凸块工艺（Bumping）流程为制备UBM和形成凸块，主要使用PVD设备、涂胶显影机、电镀机、刻蚀机、回流炉等设备。在制作凸块前，首先需要在芯片表面需要生长凸块的区域制作凸块下金属（Under Bump Metallization, UBM），用于阻挡凸点金属扩散至下层金属焊盘中；在完成制备UBM后，凸块的制备可以采用三种方式：电镀、印刷锡膏固化和植球方式，电镀方式形成铜柱等凸块，在电镀焊料、去除光刻胶和刻蚀掉溅射金属后，通过回流进行固化；印刷锡膏固化方式通过将含焊锡料的改性锡膏通过贴片、回流焊等加热固化；植球方式则通过植球机将预成型的凸块印制在UBM上。

具体工艺步骤如下：

- 1) 沉积阻挡层、溅射Ti/Cu：采用溅射或其他物理气相沉积方式在晶圆表面沉积一层钛或钛钨作为阻挡层，再沉积一层铜或其他金属作为后续电镀所需的种子层。阻挡层用于阻挡芯片表面焊盘金属与种子层间金属相互扩散。在沉积金属前，晶圆先进入溅射机台的预清洁腔体，用氩气等离子去除焊盘金属表面的氧化层，以提高溅射金属层与芯片表面的结合力；
- 2) 涂胶显影：在晶圆表面旋涂一定厚度的光敏光刻胶（UV胶, Ultraviolet Rays），并应用光刻曝光工艺，对光刻胶进行选择性曝光，光刻系统在UV胶顶部形成预定凸块尺寸的图案；
- 3) 电镀、刻蚀：前一步预定凸块尺寸的图案被刻蚀，形成一个小间隙；使用电化学沉积（ECD）或电镀设备在间隙填充或镀铜；再剥离UV胶并刻蚀结构；
- 4) 回流或加热：最后，形成的凸块结构在回流炉中回流和加热，形成凸块。

3、TSV用于DRAM Die之间及HBM和Interposer之间的互联，为电镀、键合等工艺带来较多增量

TSV技术专为2.5/3D封装而生，能够以最低的能耗提供极高的带宽和密度。TSV（Through-Silicon Via）即**硅通孔技术**，是目前最先进的封装技术之一，与传统的SIP等封装技术相比，TSV的垂直连接可以允许更多数量的连接，因此具备更好的电势能、更低的功耗、更宽的带宽、更高的密度、更小的外形尺寸、更轻的质量等优势，是实现电路小型化、高密度、多功能化的首选解决方案。

2.5/3D TSV技术已经广泛用于AI GPU基板上的HBM中，DRAM各层Die之间的连接采用3D TSV工艺，HBM芯片和下方的金属凸块之间的连接采用2.5D TSV工艺。

TSV工艺价值量在HBM 3D封装工艺中占比最高。3D TSV是HBM堆叠工艺中成本占比最高的工艺，根据3DinCites，考虑4层DRAM die和1层逻辑die堆叠的HBM结构，在99.5%和99%的芯片键合（die bonding）良率下，TSV制造和TSV通孔露出工艺分别占其成本的30%和29%。

TSV工艺主要包括前段的成型制造工艺和中段的露出及金属化工艺。晶圆厂一般负责TSV的成型工艺，主要包括刻蚀和沉积，封装厂一般负责中段制程（MEOL），即**硅通孔露出和背面金属化工艺**，主要包括减薄、钝化和键合等，以及最后的封装工艺。

1) 刻蚀：主要使用深反应离子刻蚀（DRIE，Deep Reactive Ion Etching）法，又称**博世法**，该技术在真空中利用分子气体等离子的诱导化学反应来实现各向异性刻蚀；

2) 沉积：在刻蚀完成后，需要利用**PECVD法**沉积二氧化硅或氮化硅，以形成绝缘层对硅衬底进行电气隔离；之后，需要制作金属扩散阻挡层来防止铜原子扩散、制作种子层方便后续铜的沉积，一般采用**PVD沉积Ti/Ta、采用CVD沉积TiN/TaN等**；

3) 电镀：在沉积完成后，需要电镀方法形成铜，一般采用自下而上的方法（**Bottom-up**）；

4) **减薄抛光**：在电镀完成后，需要对晶圆进行减薄抛光来使TSV背面的铜层快速露出，一般采用机械研磨等方法将晶圆减薄至50μm甚至更薄，并采用CMP等技术露出铜通孔中的铜柱；

5) **晶圆键合**：用来完成通孔金属化和连接端子的晶片之间的互联。

TSV工艺中临时键合/解键合、铜电镀、背面通孔露出成本占比最高。在via-middle的TSV工艺中，临时键合/解键合和铜电镀成本占比均为17%左右，背面通孔露出（主要包括背面减薄和抛光）和背面RDL（主要包括刻蚀、CVD、PVD、电镀等）成本占比约为15%左右，其他关键工艺包括刻蚀、CVD、铜阻挡层PVD等。

4、CoWoS用于英伟达AI GPU的整体封装，生成式AI有望推动2024年CoWoS需求翻倍以上增长

HBM和GPU等芯片采用TSMC CoWoS技术进行整体封装，能够减少封装体积和功耗等。目前立体封装主要包括2.5/3D封装，2.5D封装主要指将存储、记忆体等其他芯片，并排铺设在硅中介板（Silicon Interposer）上，经过微凸块（Micro Bump）连接，以实现硅中介板上不同芯片之间的通信；再通过硅穿孔（TSV）来连接下方的金属凸块，再由导线载板连接外部；3D封装致力于实现芯片如HBM中DRAM die的垂直堆叠。

目前，英伟达AI GPU DGX A100、H100、H200等均采用台积电的CoWoS（Chip on Wafer on Substrate）技术，首先将芯片通过CoW（Chip on Wafer）工艺连接至硅中介层（Interposer），再通过OS（On Substrate）工艺将CoW芯片和基板连接，整合成CoWoS。

按照中介基板（interposer）材料的不同，CoWoS可分为CoWoS-S/R/L三种形式。CoWoS（Chip on Wafer on Substrate，晶圆级封装）是一种高集成度封装技术，能够在中介基板（Interposer）上密集防止多颗芯片。

从工艺流程上来看，已知良品芯片（Know Good Die, KGD）首先通过Chip on Wafer（CoW）的封装制程连接至硅晶圆，再将CoW芯片与基板相连并整合成CoWoS。台积电根据中介基板的差异，将CoWoS技术分为三种类型。

1) CoWoS-S：是基础的CoWoS技术，使用Si衬底作为中介层，可以支持超高集成密度，提供不超过两倍掩膜版尺寸的硅中介层，是HBM等高速高带宽内存芯片使用的工艺；

2) CoWoS-R：引入RDL作为中介层，RDL（Redistribution Layer，重新分配层）是集成电路上形成的额外金属布线层，旨在重新排列I/O焊盘，将焊盘重塑到所需位置，以便于在必要时操作焊盘。这种工艺具备灵活性，增加了封装连接的可靠性，允许新封装可以扩大尺寸以满足更复杂的功能需求，有效支持多个Chiplet之间进行高度可靠互联。例如，芯片中心的凸块阵列可重新分配到靠近芯片边缘的位置。重新分配焊盘有助提高接触密度；

3) CoWoS-L：在CoWoS-S和InFO技术的基础上，引入LSI（Local Silicon Interconnect）技术，LSI芯片在每个产品中可以具有多种连接结构（例如从SoC到SoC、SoC到小芯片、SoC到HBM等），提供更灵活和可复用的多芯片互联架构。

台积电CoWoS需求持续强劲，展望2024年产能将增至目前2倍。台积电在CoWoS工艺处于领先地位，根据Yole，2021年全球领先厂商在高端高性能封装上的资本支出总投入约119亿美元，其中TSMC资本支出为30.5亿美元，主要包括①InFO工艺，用于UHD FO；②CoWoS工艺，用于RDL或硅中介层等；③LSI平台，和英特尔的EMIB直接竞争。由于Google TPU、英伟达GPU和AMD MI300等均导入生成式AI，台积电AIGC订单增多，先进封装需求持续提升，其中英伟达是A100和H100级数据中心CoWoS封装的最大客户，谷歌通过博通成为TPUv4和v5的第二大客户，AMD在MI300上采用CoWoS，亚马逊的Trainium through AI Chip及微软最新AI芯片也使用CoWoS工艺，国内壁仞科技于2022年8月发布首款通用GPU芯片BR100，采用7nm工艺，也使用Chiplet与CoWoS 2.5D封装工艺。台积电总裁魏哲家于6月6日股东大会表示，自去年以来CoWoS需求几乎翻倍增长，明年需求将持续强劲。由于目前市场需求大于产能，台积电表示当前首要任务是增加CoWoS产能，必要情况可能会通过转厂方式生产；台积电于23Q3法说会上表示，2024年CoWoS产能将翻倍。

5、FC工艺用于逻辑die与硅中介层连接，相较传统引线键合方式大大缩短信号路径

倒片封装（Flip Chip，FC）将芯片翻转，在先进封装工艺中取代引线键合。倒装键合（Flip Chip Bonding）指通过在芯片顶部形成凸点，并将芯片倒转过来，实现芯片的有源区域和基板间的电气和机械连接。倒

片封装在先进封装工艺中取代引线键合，主要系①引线键合对于可进行电连接的输入/输出（I/O）引脚的数量和位置有限制，而倒片封装不存在此种限制；②倒片封装的电信号传输路径短于引线键合。

在引线键合方法中，金属焊盘在芯片表面采用一维方式排列，无法出现在芯片边缘或中心位置，而倒片键合方法在键合至基板或形成焊接凸点的过程中不存在任何工艺方面的限制，因此在倒片封装中，金属焊盘可以采用二维方式全部排列在芯片的一个侧面，增加金属焊盘的数量；另外，用于形成凸点的焊盘可以布置在芯片顶部的任何位置，用于供电的焊盘可以布置在靠近需要供电的区域，进一步提升电气性能。

FC封装采用回流焊或热压工艺形成芯片上的凸点与基板上的焊盘连接。倒装封装首先将焊球附着在芯片焊盘上，再将芯片正面朝下放置在基板上，焊球通过回流焊（Reflow）或热压缩（Thermo Compression）工艺实现和基板的连接。

回流焊工艺是倒装封装中最主流的工艺，又称为C4工艺（Controlled Collapse Chip Connection），通过在回流炉的高温下融化结合处的凸点；热压缩工艺通过向接合处施加热量和压力，一般用于I/O密度更高的铜柱凸点。热压工艺的成本高、装片工艺也要求更高的精度，导致目前耗时较长。

四

HBM多层堆叠结构提升前后道工艺价值量，量测、电镀、键合等设备需求倍增

HBM工艺流程包括晶圆测试、中段制造、后段封测等环节，一方面带动部分传统封装工艺环节的设备价值量提升或带来传统封装工艺变革，另一方面，HBM引入凸块、RDL、TSV等先进封装工艺，使得越来越多的检/量测、电镀等前道设备融入中后段工艺中。

1) 前道检/量测：HBM中全部的DRAM die均需要100%检测，并且由于HBM中的铜凸点、铜柱、TSV、UBM（凸点下金属层）的结构极其复杂，其质量影响整个HBM及CoWoS封装的良率，因此需要对每个关键结构的大小、情况、精确程度等进行量测，此环节使用的设备包括前道检测和量测设备；

2) 凸块（Bump）制造：介于前道晶圆制造和后道封装工艺之间，是扇入（Fan-in）封装、芯片级封装（CSP）、系统级封装（SiP）、芯粒封装（Chiplet）等工艺的必要环节，TSV、晶圆级封装（WLP）等均

是凸块制造工艺的延伸。在HBM中，微凸块（uBump）用于HBM之间DRAM Die、HBM和硅中介层之间的连接，C4 Cu凸点用于硅中介层和封装基板之间的连接，实现电气互连和应力缓冲。如前文所述，Bumping工艺主要使用PVD设备、涂胶显影机、电镀机、刻蚀机、植球机、回流炉等设备；

3) TSV制造：如前文所述，TSV工艺主要包括前段的通孔成型工艺和中段的通孔露出和背面金属化工艺，使用的设备主要包括刻蚀机、PECVD、PVD、电镀机、减薄机、CMP、键合机等；

4) 硅中介层（Interposer）制造：主要包括形成TSV、双面RDL和凸块制造、临时键合和解键合、晶圆减薄、切片等工序。RDL（Redistribution Layer，重新分配层）是额外金属布线层，通过重新排列I/O焊盘，满足封装工艺更复杂的功能需求，例如将HBM堆栈与SoC等逻辑芯片相连。凸块工艺可以看做是点与点的连接，RDL工艺则是面阵连接，RDL制备使用的设备与凸块类似，主要包括光刻机、刻蚀机、溅射、CVD、电镀设备等，完整的硅中介层制造还主要需要临时键合/解键合机、背面减薄机等；

5) 后道封装：传统封装工艺流程包括晶圆背面减薄、划片/切割、贴片/固晶、互连、模塑、激光打标、切筋成型等步骤，HBM的多层堆叠结构显著提升晶背减薄、键合等设备的需求，并且对注塑机提出更高要求；

6) 后道测试：在晶圆制造之后、封装开始之前，需要首先使用探针台、测试机等将不良芯片剔除，探针台接触裸芯片上的PAD点，并连接测试机完成测试过程，再将不良芯片进行打点标记，形成晶圆Map图；在封装完成后，需要分选机配合测试机进行成品芯片的终测，该环节主要使用的设备包括探针台、分选机、测试机；

7) 板卡级组装和整机组装：为HBM封装的最后一道工序，主要将引脚通过PCB上的导线和其他器件建立连接，工艺涉及通孔插装技术、表面贴装技术等，为了实现PCB组装工艺，使用的设备主要包括焊膏涂覆设备、丝网印刷机、点胶机、贴片机、回流炉、清洗机、自动光学检测设备等；为了提供PCB基板，使用的设备主要包括真空层压机、钻孔机、通孔电镀设备、涂胶机、光刻机、显影机、刻蚀机、丝网印刷机、电镀铜设备、自动光学检测仪等。

1、测试：用于Bump、TSV、KGSD等环节，HBM工艺步骤增多带来前、后道测试需求提升

(1) 前道检/量测：核心增量来自Bumping、TSV、RDL工艺

检测 (Inspection) 和量测 (Metrology) 两大工艺分别用于检测晶圆异质情况和对晶圆结构做出量化描述。检测指在晶圆表面上或电路结构中，检测其是否出现异质情况，如颗粒污染、表面划伤、开短路等对芯片工艺性能具有不良影响的特征性结构缺陷；量测指对被观测的晶圆电路上的结构尺寸和材料特性做出的量化描述，如薄膜厚度、关键尺寸、刻蚀深度、表面形貌等物理性参数的量测。根据检测类型不同，半导体质量控制设备也相应分为检测设备和量测设备两大类。

2022年全球检测和量测设备市场空间大约125亿美元。根据VLSI Research数据，2016年全球检测和量测设备合计规模为47.6亿美金，2020年增长至76.5亿美金，CAGR达12.6%。根据Gartner，2021年和2022年全球工艺控制设备市场占比大约为11%和12%，结合全球前道设备市场空间，2021和2022年全球半导体检测和量测设备市场规模预计分别为100亿美元和125亿美元。

HBM对前道检/量测设备的增量主要来自Bumping、TSV、RDL工艺。

Ø Bumping工艺中，检/量测设备需要识别凸点的缺失、大小不一致等问题。在HBM中，铜凸点 (Bump)、凸点下金属 (UBM)、RDL、TSV等的结构极其复杂，其形状、质量、精确程度直接影响HBM甚至整体CoWoS封装的良率。例如，Bump用于各DRAM die、HBM和硅中介层、GPU和硅中介层的连接，C4凸点用于硅中介层和封装基板之间的连接，均需要检/量测设备来检测Bump是否缺失、大小是否不一致等问题。

Ø TSV工艺中，刻蚀、衬底/阻挡层沉积、填充和CMP等每一个工艺步骤均需要测量和缺陷检查。例如，TSV的衬底和阻挡层是采用纳米级薄膜沉积工艺制造的，关键尺寸 (CD) 的精确测量和控制需要TSV横截面的高分辨率扫描电子显微镜 (SEM) 图像；对于TSV特定的缺陷机制，包括裂缝、不完全体填充和绝缘体壁上的针孔等，这些纳米级的结构缺陷需要通过高压SEM来进行量测；另外，随着堆叠芯片数量的增加，3D结构中需要更小的CD尺寸，由此将产生D2W或W2W键合、及TSV未对准产生的纳米尺寸缺陷或孔洞，该类量测需求也不断增加。

Ø 针对RDL工艺，更精细的电路增加检/量测设备需求。HBM中硅中介层上排列多层RDL布线，支持极其精细的线路，可以互连芯片上间距为55um甚至40um的微凸点。随着需要互连的凸点增多，RDL互连的凸点间距进一步缩小，增加更加精细和先进的检/量测设备需求。

(2) 后道测试：针对堆叠晶圆增加预键合测试，针对KGSD增加多种封装测试流程

根据SEMI，2021年全球半导体后道测试设备市场规模大约78亿美元，国内市场大约22.5亿美元，其中测试机、分选机、探针台分别占比63%、17.4%、15.2%。目前，测试设备主要被泰瑞达、爱德万等海外厂商垄断，探针台/分选机等主要被东京精密、东京电子等垄断，国产化提升空间较大。

Ø 测试机：主要实现晶圆的功能测试，通过一台工作站来做外部的控制，透过测试程序的定义来产生待测器件上所需要的电压、电流及时序信号等的输入和读取，依次来判断待测器件的好坏；

Ø 探针台：主要用于半导体制造晶圆检测（CP，Circuit Probing）环节，也用于设计验证和成品测试环节。在晶圆检测环节中，探针测试主要对封装前晶圆上的裸芯片进行性能测试，裸芯片上有多个PAD点用于信号传输（即图41绿色方块上多个方形区域，边长约数十微米），测试时探针与之接触并输入、输出测试信号。裸芯片经PAD引出引脚、覆盖封装材料等工序后即成为成品芯片，稳定性与尺寸均大幅提升；

Ø 分选机主要用于IC设计阶段的验证环节和封测阶段的成品测试环节，一般和探针台配合使用。分选机将待检测的芯片自动传送至测试工位（用于将芯片与测试机连接并进行测试的位置），待检测芯片的引脚通过测试工位上的专用连接线与测试机的功能模块进行连接，测试机在进行检测之后将测试结果传送给分选机，分选机根据测试结果将检测过的芯片进行标记、分类、收料。

相较传统DRAM，HBM增加了预键合晶圆级测试流程，并针对KGSD增加多种封装级测试工艺。传统DRAM测试流程包括晶圆级和封装级测试，常见的晶圆级测试包括晶圆老化（WFBI）测试、冷/热测试、存储器矩阵修复等组成，HBM增加了针对堆叠的核心芯片进行预键合测试，以检测MBIST和PHY电路中的缺陷，以确保堆叠的芯片具有相同的质量水平；对于封装级测试，在传统DRAM中，老化测试、热/冷测试和速度测试均在封装级别进行，然而HBM堆叠的KGSD（已知的良好堆叠管芯）具有晶圆的外形尺寸，因此无法使用传统的DRAM封测解决方案，需要增加TSV测试、动态模式老化测试、温度处理无晶圆损坏测试，并通过直接访问焊盘进行高速测试。

Ø 针对基础晶圆（Base die），HBM的基础晶圆为逻辑die，需要增加扫描测试和高速PHY测试；

Ø 针对TSV，需要增加开路/短路测试，确保DRAM和Base die之间数据和命令信号的连接；

Ø 针对KGSD，需要增加高温下的动态和静态应力测试，在预键合阶段，仅进行静态应力测试，以便在堆叠时更好施加应力；另外，KGSD还需要增加晶圆级速度测试环节，这需要特定的探针卡以满足HBM堆叠芯片并行测试需求；

Ø 针对I/O接口，由于HBM中I/O接口高达1000个以上，而直径和间距太小，无法在KGSD期间测试到，还需要增加特定的PHY I/O测试环节。

2、减薄：全球减薄设备市场空间超8亿美元，HBM多层堆叠结构提升设备需求

用来提高晶圆强度的衬底材料90%以上均需要被去除，背面减薄工艺可将晶圆减薄至250um-20um。伴随硅片和晶圆尺寸的增大，其厚度也需要相应增厚来保证硅片或晶圆在制造过程中具有足够的强度，例如6英寸和8英寸的晶圆的厚度分别为625um和725um，12英寸硅片的平均厚度为775um。对于725um厚度的12英寸晶圆，其电路层的有效厚度一般为5-10um，总厚度90%以上的衬底材料是为了保证晶圆在制造、测试和运送过程中具备足够的强度。随着芯片集成化程度提高，晶圆厚度一般越薄越好，因此，在晶圆制造完成之后，需要首先通过晶圆背面减薄（Back Grinding）工艺来降低厚度，再进行划片、贴片等后续封装工艺。对于单芯片封装来说，经过背面减薄的晶圆厚度一般会从700-800um减少到200-250um；对于多层堆叠芯片封装来说，晶圆一般需减薄至70-80um，通过两次研磨工艺甚至可以减薄至20um。

2022年全球晶圆减薄市场约8.2亿美元，主要为DISCO、东京精密等海外厂商垄断。根据QYResearch，2022年全球减薄机市场规模约8.2亿美元，2018-2022年CAGR约为18.7%，预计到2029年将增长至接近13.2亿美元，未来6年CAGR大约6.5%。全球主要减薄机厂商包括Disco、东京精密（TOKYO SEIMITSU）、G&N、Okamoto Semiconductor Equipment Division等，CR3占有85%的份额。

先进封装、Chiplet等技术需要厚度更低的超薄晶圆，将大幅提升对减薄设备的需求。芯片不断追求更高的集成度和更小的体积，3D IC等工艺得以发展，通过硅穿孔（TSV）等技术实现IC堆叠，可以有效减小IC之间互连的长度，将芯片整合成效能最佳、体积最小的状态，目前大部分的3D NAND、背照型CMOS图像传感器、智能手机SoC等先进芯片均使用2.5或3D IC技术。传统的减薄工艺一般只需要将晶圆减薄至100-200um，但在3D封装中，需要将多层芯片进行堆叠，往往需要将晶圆厚度减薄至50-100um甚至50um以下。根据Yole，2019年全球减薄后晶圆出货量为1亿片，预计到2025年上升至1.35亿片，其中100-200um晶圆仍占据主要份额，约为8200万片；30-50um晶圆增速最快，2019-2025年CAGR为98%，在2025年预计出货量约170万片。

减薄抛光是TSV工艺中介于电镀和键合之间的重要工序，帮助实现HBM多层DRAM die堆叠。晶圆厂一般负责TSV的成型工艺，主要包括刻蚀和沉积，封装厂一般负责中段制程（MEOL），即硅通孔露出和背面金属化工艺，主要包括减薄、钝化和键合等，以及最后的封装工艺。在沉积完成后，需要电镀方法形成铜，一般采用自下而上的方法（Bottom-up）；在电镀完成后，需要对晶圆进行减薄抛光来使TSV背面的铜层快速露出，一般采用机械研磨等方法将晶圆减薄至50μm甚至更薄，并采用CMP等技术露出铜通孔中的铜柱。减薄后的晶圆将大大降低TSV阻抗，还会增加数据带宽、降低热阻，最终增加互连密度。在HBM结构中，减薄和键合工艺配合使用，DRAM die之间不再需要导电凸块，芯片的厚度将薄数倍，整体堆叠高度得以降低。经过背面减薄的晶圆厚度一般从700-800um降至70-80um，减薄到十分之一厚度的晶圆能够堆叠4-6层，经过两次减薄工艺，晶圆可以减薄至大约20um，从而实现未来HBM堆叠16-32层。

3、互连：超薄晶圆增加临时键合/解键合需求，混合键合市场有望指数级增长

键合（Bonding）属于后道封装过程，按键合的材料类型分为晶圆-晶圆键合（Wafer-to-Wafer, W2W）和芯片-晶圆键合（Die-to-Wafer, D2W）。键合主要指将两片表面清洁、原子级平整的同质或异质半导体材料经表面清洗和活化处理，在一定条件下直接结合，通过范德华力、分子力甚至原子力使两片半导体材料成为一体的技术。W2W是指通过化学或物理反应将晶圆与晶圆、晶圆与玻璃基板或其他材料圆片永久结合起来的工艺。键合过程为，在外能量的作用下，两个晶圆接合界面上的原子相互反应形成共价键，从而使晶圆接合并达到一定的界面键合强度；D2W指在划片工艺之后，将从晶圆上切割的芯片黏贴在封装基板（引线框架或印刷电路板）上。芯片键合的传统方法为引线键合（Wire Bonding），先进方法采用倒装键合（Flip Chip Bonding）、混合键合（Hybrid Bonding）等。

按键合的方式，键合分为临时键合和永久键合，键合精度、材料选择等是工艺成功的关键。晶圆键合工艺是半导体器件物理、材料物理化学、精密机械设计、高精度自动控制等多学科交叉的领域，对对位精度、键合温度均匀性、键合压力范围及控制精度等要求较高，另外用于固定薄晶圆的键合胶也是工艺成功的关键。

1) 临时键合：超薄晶圆需要临时键合至较厚的载体上，方便后续进行减薄等一些列工艺。柔性、易碎、翘曲是减薄后晶圆的特点，因此需要通过临时键合工艺，用中间材料将薄晶圆键合到较厚的载体片上，再经过背面减薄、TSV开孔、重布互联等工艺后，再输入外界能量（光、电、热、力等）使粘层失效，无损地将晶圆与载体片分离；

2) 永久键合：指通过化学或物理反应将晶圆与晶圆、晶圆与玻璃基板或其他材料圆片永久结合起来的工艺。在外能量的作用下，两个晶圆接合界面上的原子相互反应形成共价键，从而使晶圆接合并达到一定的界面键合强度。在TSV封装工艺中，为了形成良好的电学互联，精度要达到连接柱直径的10%，对连接精度要求很高，TSV连接柱的直径一般为5um以下，那么连接精度就要求在0.5um以下，这时候采用晶圆对晶圆的对准方式就能实现很好的对准精度。

混合键合技术能够实现芯片模块间快速通信，成为最先进的新一代键合技术。在混合键合之前，2D、2.5D、3D封装均使用焊球凸点或微凸点来实现芯片与基板、芯片与中介层之间的连接，而由于凸点间距逐渐缩小，芯片模块间的通信速度不断提升，原来的锡焊球无法满足工艺要求；铜-铜混合键合（Hybrid Bonding）通过精密间隔的铜焊盘垂直连接D2W或W2W，信号丢失率可忽略不计，在高数据量、高性能计算领域优势明显。

HBM的多层堆叠结构带动临时键合、混合键合等需求。在传统工艺中，每一代SoC上集成更多的功能，但芯片缩放越来越困难，而使用Chiplet技术，大型SoC被分割成更小的Dies或IP块，并重新聚合成一个全新的设计，对临时键合、混合键合等需求均有拉动。

1) 临时键合需求：在Chiplet技术中，为了缩小芯片体积、提高芯片散热性能和传导效率等，晶圆减薄工艺会被大量应用，为了不损伤减薄中以及减薄后晶圆，需要将晶圆片与玻璃基板临时键合并在完成后续工艺后最终解键合。同时在Chiplet技术路线下，Fan-out、CoWoS等封装工艺路线都要经过单次或多次的临时键合及解键合工艺来实现芯粒互联；

2) 混合键合需求：随着台积电、英特尔、三星等厂商开始采用3D封装工艺，Chiplet应用范围扩大，混合键合越来越多用于CPU、GPU及高性能计算中。最早实现Hybrid Bonding量产的产品是索尼的图像传感器，下部电路芯片和上部像素芯片之间增加了铜连接焊盘，同时建立物理和电气连接；又如在AMD的一个方案中，AMD堆叠了SRAM和一个处理器芯片，形成了3D封装形式，并在顶部结合了高性能的MPU和高速缓存，使用混合键合连接各个Die。

半导体晶圆键合设备全球市场空间大约9亿美元，混合键合设备细分市场有望迎来指数级增长。根据thebrainyinsights数据，2021年全球半导体键合设备市场空间大约8.74亿美元，当前晶圆键合设备主要为海外厂商垄断，行业龙头包括奥地利的EVG Group、德国的SUSS、日本TEL等，国内尚无多模块集成的晶圆键合设备，技术与国外差距较大，国内主要厂商为拓荆科技、上海微电子（SMEE），苏州芯图（芯睿科技）、华卓精科等，芯源微正在研临时键合机、解键合机等新型设备。伴随着“后摩尔时代”来临，三维集成领域进入成长期，混合键合设备细分市场有望迎来指数级增长。

4、模塑：用于对芯片进行塑封保护，HBM采用注塑或压缩式工艺进行塑封

模塑（Molding）是指一种将芯片或器件模塑料进行保护的封装工艺，传统封装采用注射或传递成型塑封法。模塑使裸露于外界的芯片、器件及连接线路通过外部塑封体得到保护，免受外界环境对半导体器件的侵蚀，避免产品失效。对于传统的TSSOP、DFN、QFN及BGA等封装工艺，半导体塑封主要采用注射成型塑封（Injection Molding）或传递成型塑封（Transfer Molding），过程为：将引线键合连接芯片的基板放在两个模具上，同时将环氧树脂模塑料片放在中间，然后通过注塑头施加热量和压力，使环氧树脂塑料熔化为液态，流入模具并填充间隙。

压缩式成型塑封主要用于存储器等多叠层超薄晶圆的封装，可满足晶圆级等先进封装需求。在压缩式塑封（Compression Molding）工艺中，模具会预先填充环氧树脂模塑料粉末，基板放入模具中后，随后施加热量和压力，模具中填充的环氧树脂模塑料粉末会液化并最终成型。其会即刻熔化为液体，无需流动便可填充间隙。由于塑封料在模具腔体内无流动，可实现无冲线塑封和超薄塑封，也可实现晶圆级等大面积塑封需求，例如LGA、WLCSP等。随着扇外型封装的兴起，塑封起到重新构造晶圆、面板，并增加芯片布线、布球的功能，在主流的晶圆级、面板级扇出封装中，塑封工艺是关键工艺。

SK海力士的HBM采用MR方式将液态LMC注入DRAM die之间。考虑到批量回流工艺的效率、可靠性高，因此SK海力士在其HBM3、HBM3e堆叠中并未采用TC（热压）方式，而是沿用MR（批量回流）工艺进行HBM间模塑材料的填充。具体过程为，直接将液态的LMC材料注入DRAM die之间的空间，并进行固化，从而达到芯片保护的作用。SK海力士采用新技术来控制晶圆变薄，使其不会弯曲；在12层及以上堆叠过程中，瞬时施加强烈热量，以确保连接芯片的凸块均匀拼接；将LMC材料置于真空下，并施加70吨压力来填充芯片之间的狭小空间。

1、环氧塑封料：用于保护芯片，GMC和LMC及原材料均受益于HBM需求增长

EMC起到保护芯片的功能，在传统和先进封装中均广泛应用。环氧塑封料（Epoxy Molding Compound, 简称EMC）全称为环氧树脂模塑料，属于包装材料，是用于半导体封装的一种热固性化学材料，由环氧树脂为基本树脂，以高性能酚醛树脂为固化剂，加入硅微粉等填料，以及添加多种助剂加工而成，主要功能是保护半导体芯片不受外界环境（水汽、温度、污染等）的影响，并实现导热、绝缘、耐湿、耐压、支撑等复合功能。EMC在传统引线键合封装和先进封装中均广泛应用，FOWLP、多层堆叠封装对塑封料提出了更高的性能要求。

EMC主要包括液态塑封料（LMC, Liquid Molding Compound）与颗粒状环氧塑封料（GMC, Granular Molding Compound）。1) GMC指采用均匀撒粉的方式，在预热后变为液态，将带有芯片的承载板浸入到树脂中而成型，具有操作简单、工时较短、成本较低等优势；2) LMC指通过将液态树脂挤压到产品中央，在塑封机温度和压力的作用下增强液态树脂的流动性，从而填满整个晶圆。LMC具备可中低温固化、低翘曲、模塑过程无粉尘、低吸水率及高可靠性等优点，是目前用于晶圆级封装的相对成熟的塑封材料。

EMC的填充料成本占比最高，主要无机填料为球形硅微粉和球形氧化铝。目前常见的环氧塑封料主要组成为填充料（60-90%）、环氧树脂（18%以下）、固化剂（9%以下）、添加剂（约3%）。在先进封装中，环氧塑封料的主要要求为高耐潮、低应力、低 α 射线、耐浸焊和回流焊，需要保证塑封性能好，因此环氧塑封料必须在无机树脂基体内掺杂无机填料，现有的无机填料基本均为二氧化硅球形微粉，具有降低塑封料的线性膨胀系数。

增加热导，降低介电常数，环保、阻燃，减小内应力，防止吸潮，增加塑封料强度，降低封装料成本等作用。另外，针对HBM封装等高导热存储芯片封装领域，颗粒封装材料（GMC）中一般将TOP CUT20um以下球形硅微粉和Low α 球形氧化铝复配混用，散热要求越高的场景，Low- α 球铝的占比会越高。

传统的HBM采用TC-NCF技术堆叠DRAM die并用GMC材料保护，SK海力士独创MR-MUF技术将LMC材料用于塑封HBM3和HBM3E。

1) TC-NCF（Non-Conductive Film, 非导电薄膜）技术即将上下两层die之间通过TC（Thermal Compression, 热压）工艺，用一层薄膜（DAF膜等）连接起来，美光和三星均采用这种方式，SK海力士

在HBM和HBM2E中也采用这种方式。在传统HBM中，**塑封模具中会预先填充GMC颗粒状塑封料**，随后施加热量和压力，使GMC液化并最终成型为塑封料；

2) MR-MUF (Mass reflow bonding with molded underfill, **批量回流模制底部填充**) 为SK海力士独创的技术，即将半导体芯片堆叠后，为保护芯片和芯片之间的电路，在其空间中注入液体形态的保护材料，即LMC (Liquid Molding Compound) 并进行固化。由于MR-MUF效率高、可靠性高，散热性能好 (SK海力士的MR-MUF技术焊接温度比TC-NCF技术低14°C)，因此SK海力士将MR-MUF技术用于HBM3和HBM3E产品中，并凭借此技术迅速占领HBM大多数市场份额。

2021年国内EMC市场规模约66亿元，HBM带动EMC和球形硅微粉等市场增长。根据华海诚科招股书，2021年中国封装材料市场规模为73.6亿元，EMC占比大约90%，即市场规模大约66.24亿元。在传统封装领域，内资EMC厂商份额逐渐提升，在SOP、QFP等领域仍存在一定替代空间；在先进封装领域，市场份额基本被住友电木、蔼司蒂、京瓷等外资厂商占据，内资厂商多处于客户验证阶段，少数厂商产品实现小批量产。

2、电镀液：用于形成HBM铜凸块、TSV等结构，全球市场空间约10亿美元

电镀液及添加剂主要用于铜互联工艺，全球市场空间接近10亿美元。电镀液在IC制造和先进封装中用于铜互联工艺，该工艺贯穿整个芯片制造过程，随着先进封装对镀铜材料需求快速增加，大马士革铜互联、先进封装凸块电镀 (Cu Pillar/Bump/RDL/UBM)、硅通孔 (TSV) 电镀等材料市场不断扩大，在先进芯片中互联材料首尾连接可长达约30英里。根据TECHCET，铜互联材料是电镀材料最大的细分市场，2022年全球半导体用电镀材料市场规模约10.2亿美元，预计到2026年增加至13.8亿美元。

IC制造工艺中金属布线用于连接电子元器件层，铜互联采用电镀方式实现铜的填充。IC最初采用铝作为导体，二氧化硅作为绝缘体来构造互联层，整个互联过程从在晶圆表面沉积铝开始，然后通过选择性刻蚀形成布线图案，沉积氧化物绝缘体，并利用CMP使晶圆表面平坦化；随着器件特征尺寸缩小，越来越薄的铝线无法实现所需的速度和电性能，铜互联结构逐步取代铝互联；然而由于铜不易形成挥发性化合物，因此使用等离子干法刻蚀铜的方法并不可行，因此工程师选择采用大马士革铜互联镶嵌工艺 (即借鉴大马士革的珠宝行业，先在基底金属上刻蚀图案，再将贵金属嵌入图案中)，先沉积和刻蚀电介质材料，再将铜填充到图案之中。为了实现高深宽比图案特征，必须采用电镀而非PVD或CVD来填充铜金属。

HBM引入铜互连工艺，电镀液主要用于形成铜柱凸块、TSV等结构。由于HBM中需要的Bumping、RDL、TSV等先进封装工艺引入前道电镀工艺环节，自然带来电镀液需求提升。

1) 在Bumping工艺中，电镀步骤位于涂胶显影工艺环节之后，用于制备一定厚度的金属层作为UBM；

2) 在RDL工艺中，每层RDL布线均需要电镀来形成铜凸块；

3) TSV技术的核心是在晶圆上打孔，并在硅通孔中进行镀铜填充，从而实现晶圆的互联和堆叠，在无需继续缩小芯片线宽的情况下，提高芯片的集成度和性能。和芯片制造铜互连工艺相比，TSV电镀的尺寸更大，通常需要更长的沉积时间、更高的电镀速率以及多个工艺步骤，铜互连电镀液及添加剂成本占TSV工艺的总成本比重也更高。

3、PSPI：用于HBM的硅中介层，形成RDL的再钝化层

聚酰亚胺具有最高的阻燃等级，高端光敏聚酰亚胺主要用作光敏光刻胶和先进封装树脂。聚酰亚胺（Polyimide, PI）指分子结构主链中含有酰亚胺结构的高分子聚合物，高性能PI的主链大多以芳环和杂环为主要结构单元。

PI具有最高的阻燃等级（UL-94），良好的电气绝缘性能、机械性能、化学稳定性、耐老化性能、耐辐照性能、低节点损耗等，这些性能在很宽的温度范围（-269°C-400°C）内不会发生显著变化。光敏聚酰亚胺（Photosensitive Polyimide, PSPI）是一类在分子链上兼有亚胺环和光敏基团，具备良好的感光性能，主要用于光刻胶和电子封装领域。

∅ 作为光刻胶：在PSPI中添加增感剂、稳定剂等可以得到PSPI光刻胶，相较于传统光刻胶，PSPI光刻胶无需涂覆光阻隔剂；

∅ 作为电子封装材料：PSPI可用于：缓冲涂层、钝化层、α射线屏蔽材料、层间绝缘材料、晶片封装材料等，还可以用于集成电路和多芯片封装件的封装。

在HBM中，PSPI主要用于硅中介层的RDL等。在WLCSP封装中，为了缓解焊球凸点和再布线层对芯片产生的应力，会在芯片表面和再布线层表面涂覆一层高分子薄膜材料，成为再钝化层（Repassivation Layer）。目前，再钝化层材料除了PSPI外，还包括聚苯并咪唑（Polybenzoxazole, PBO）和苯并环丁烯（Benzocyclobutene, BCB）。

4、封装基板：用于HBM硅中介层下方，实现内部芯片与外部电路间的电气连接

封装基板（Package Substrate）可分为有机基板和陶瓷基板。有机基板由有机树脂和玻璃纤维布为主要材料，一般采用铜箔作为导体，有机树脂包括：环氧树脂（FR4）、BT树脂（耐高温双马来酰亚胺三嗪）、PPE树脂（聚苯醚树脂）、PI树脂（聚酰亚胺树脂）等；陶瓷基板具备更好的机械和热性能，通常包括：HTCC、LTCC、氮化铝等。英特尔主导研发了一种以ABF（Ajinomoto Build-up film）作为原材料的基板，相较于BT基板，ABF材质可用于线路较细、高信息传输的IC，比如CPU、GPU等芯片。ABF作为基板的优势在于，铜箔基板上附着ABF增厚薄膜就可以直接电镀铜来形成金属布线，不需要热压过程。

在HBM中，硅中介层一般采用FC-BGA（球栅阵列封装）和下方封装基板相连，采用锡球来代替引线框架，锡球附着于基板底部，引线与基板顶部连接，基板中心位置由名为“芯板（Core）”的材料构成，通过将铜箔与浸渍过有机树脂的玻璃纤维粘合在一起，金属引线在铜箔表面形成，之后在铜箔上涂覆阻焊剂，露出作为保护层的金属焊盘/引脚。

5、CXL：能够突破CPU内存极限，未来将越来越多搭配HBM使用

HBM目前直接搭配CPU的应用较少，主要系访问延迟等限制。由于目前HBM访问延迟较高，在访问某个特定位置的数据后，需要若干周期的时间以后才能抵达该位置并执行控制器发出的指令，同时由于HBM和主芯片在同一系统中进行SiP封装，灵活性欠佳，HBM显存容量基本不可以后续进行扩展。因此，当前HBM在PC等CPU中应用较少，2022年11月，英特尔发布全球首款配备HBM内存的x86 CPU“至强Max”Sapphire Rapids，搭载4个HBM2e Stack供64GB，每个HBM2e Stack容量为16GB。

CXL使CPU内存空间和连接设备上的内存之间保持一致性，进而突破CPU内存极限。AI等应用促进处理的数据指数级增长，服务器越来越多转向异构计算架构，即使用专门的GPU等加速器从CPU卸载专门的工作负载。CXL（Compute Express）是一种新的开放式互联标准，是基于PCIe物理层的高速、低延迟CPU设备互联技术，可以在主机CPU和互联设备（例如加速器和存储器扩展设备）之间提供高效连接，能够简化加速器和内存扩展的互联和可扩展性。

CXL本质特点是①确保直接挂载的CPU存储器与CXL设备上的存储器保持一致，使主机和CXL设备之间可以无缝读取相同的数据；②允许通过CXL接口将各种类型（例如易失性、持久性等）的存储器挂载到主机，这与底层存储器技术无关；③支持交换和内存池，交换功能可实现存储器扩展，池化功能允许动态分配和释放存储器资源，从而提高整体系统效率。

CXL可以作为HBM的补充，未来CXL将越来越多搭配HBM使用。根据CXL联盟，目前确定了CXL互联的三类主要设备：

①智能NIC等加速器通常缺少本地内存，CXL可以实现NIC等设备和主机CPU的DDR内存进行通信；

②GPU、ASIC和FPGA等都配备了DDR或HBM内存，可以使用CXL使主机CPU的内存在本地供加速器使用，加速器的内存也可在本地供CPU使用，并且加速器和CPU的内存位于同一个缓存的一致域中，有助于提升异构工作负载；

③连接内存缓冲区（buffer），为主机CPU提供额外的带宽和容量。

六

海外厂商较早布局HBM，三大存储原厂占据HBM主要市场份额

HBM市场份额集中于三大存储原厂，国内厂商尚不具备HBM DRAM Die生产和堆叠能力。HBM的DRAM Die由存储原厂负责生产和堆叠，目前份额集中于三大原厂SK海力士、三星、美光，根据集邦咨询，2022年三大原厂HBM市占率分别为SK海力士50%、三星约40%、美光约10%，其中SK海力士是HBM3产品的领先生产商，是英伟达AI服务器GPU的主要供应商，三星主要满足其他云端服务商的订单。

1、SK海力士：开发MR-MUF技术，HBM市场份额领先

HBM市场先行者和领军者，全面推进第五代HBM3E。2014年，SK海力士和AMD合作开发第一代硅通孔HBM产品，还联合开发了高带宽三维堆叠存储器技术和相关产品。HBM1带宽高于DDR4和GDDR5产品，同时以较小的外形尺寸消耗较低的功率，更能满足GPU等带宽需求较高的处理器；SK海力士凭借HBM1占

据市场领先地位，并在2018年发布第二代产品——HBM2，其中一项关键的改进是伪通道模式（Pseudo Channel Model），将一个通道分为两个单独的64bit I/O子通道，为每个存储器的读写访问提供128位预取，从而优化内存访问并降低延迟，从而有效提供带宽；2020年，SK海力士发布第三代产品——HBM2E，作为HBM2的扩展版本，HBM2E堆叠8个16Gb芯片，总容量是HBM2的两倍，处理速度高达3.6Gbps；2021年10月，SK海力士成功开发出第四代产品——HBM3，并于2022年6月开始生产；2023年4月，公司宣布，在全球率先研发出12层堆叠的HBM3内存，单颗容量最高可达24GB；2023年8月，公司成功开发出面向AI超高性能DRAM新品HBM3E，预计24H1量产；公司预计2026年量产HBM4。

SK海力士开发MR-MUF等堆叠技术，巩固HBM市场领先地位。SK海力士的CoC（芯片内建芯片）技术将凸块互联（Bump Interconnection）与引线键合（Wire Bonding）相结合，在提高运行速度和降低成本方面实现突破，目前专门用于SK海力士的高密度模块的量产与生产；SK海力士还开发了MR-MUF技术，用于其HBM2E、HBM3等产品中，确保了HBM 10万多个微凸块互连，另外该技术还提供更出色的散热性能；SK海力士还正在开发Cu-to-Cu（Copper-to-Copper，铜-铜）键合工艺替代焊接，Cu-to-Cu可在完全不使用凸块的情况下将间距缩小至10微米及以下；另外，SK海力士还在研究用于Fan-out RDL（扇外型重新分配层）技术等，计划将该技术用于以Chiplet为基础的集成封装，计划到2025年实现1um及以下水平的RDL技术。

MR-MUF技术工艺效率高并且散热性能更好，难点在于热翘曲和芯片偏移问题。对比TC-NCF技术，MR-MUF技术的优点为：1) 提高工艺效率：NCF在每堆叠一个芯片时铺上一层薄膜型材料，MR-MUF在注入LMC材料之后，批量加热每个芯片以实现瞬时互连；2) 散热性能更好，提高凸块互连的质量：MR-MUF由于采用了高导热率的模制底部填充（MUF）材料，散热效果较NCF技术更好，提高了HBM超过10万个微凸块互连的质量。MR-MUF工艺难点在于热翘曲和芯片偏移问题，由于模塑是一个热反应过程，芯片堆叠过程中需要对每个芯片施加瞬时高热，进而熔化凸点实现连接，但由于环氧树脂和晶圆之间的热膨胀系数（CTE）相互不匹配，回流（MR）工艺会导致热翘曲、非接触性断开、局部桥接等问题，进而导致焊接凸点失效；另外，料盘/载具在传输过程中容易受振动影响，回流过程中芯片也会自由偏移，因此芯片的位置也容易发生偏移。

SK海力士对MR-MUF技术进行改进，一定程度上解决热翘曲和芯片偏移问题。SK海力士在先进的HBM中仍沿用MR-MUF工艺，主要系MR-MUF具备可靠性和高效率，并且SK海力士对原始的MR-MUF技术进行改进，主要包括：采用新技术来控制晶圆变薄，使其不会弯曲；在12层及以上堆叠过程中，瞬时施加强烈热

量，以确保连接芯片的凸块均匀拼接；将LMC材料置于真空下，并施加70吨压力来填充芯片之间的狭小空间。SK海力士表示，改进后的MR-MUF技术保留原来工艺的优点，将生产率提高了约3倍，并将散热性能提高了约2.5倍。

2、三星：开创I-Cube及X-cube先进封装工艺，2024年HBM产能将提高2.5倍

24H1 HBM3产量将占HBM总产量一半以上，计划将2024年HBM产能提高2.5倍。2020年，三星推出HBM2；2021年2月，三星推出HBM-PIM（存算一体）芯片；2022年，三星开始量产HBM3，在23Q3量产并开始供应8层和12层产品，计划在23Q4进一步扩大生产规模，预计到24H1，HBM3将占公司HBM总产量的一半以上；公司计划24H1推出HBM3E，24GB 8层样品已经推出，计划24H1量产，36GB 12层产品计划于24Q1提供样品。为了保持行业最高的HBM产能，公司计划将2024年HBM产能提高2.5倍。

三星自研I-Cube技术和X-Cube技术，分别用于2.5D和3D封装。

Ø I-Cube 2.5D封装技术通过并行水平芯片放置，处理热量积存并扩展性能。三星I-Cube技术为2.5D封装工艺的一种，包括I-Cube S和I-Cube E，I-Cube S将逻辑芯片与一组HBM裸片水平放置在硅中介层上，通过调整材料和厚度来控制硅中介层翘曲和热膨胀问题；I-Cube E采用硅嵌入式结构，采用大尺寸、无TSV结构的RDL中介层。目前三星已开发出I-Cube 8（搭载8组HBM）技术，计划2024年量产；

Ø X-Cube 3D封装技术包括微凸块和铜混合键合两种方案。X-Cube技术中，上下层逻辑die通过微凸块（X-Cube TCB方案）或铜混合键合（HCB方案）连接，在铜混合键合方案中，堆叠精度进一步降低，三星正在开发低于4um等更精细的HCB技术。

三星HBM4路线图计划变革HBM封装方式。在HBM4技术路线图中，三星计划使用FinFET节点生产逻辑die，封装方式计划将从基于凸点的CoW（chip on wafer）变为基于Pad连接的Bumpless形式，并且三星正开发针对高温热特性优化的非导电粘合膜（NCF）组装与混合键合（HCB）等技术。

3、美光：计划通过HBM3E实现弯道超车，正开发HBM Next产品

HBM3E产品正在英伟达验证，预计2024财年产生7亿美元收入。美光于2020年开始提供HBM2产品，用于高性能显卡、服务器处理器等领域；美光跳过HBM3直接提供HBM3E样品，在8层die堆叠的情况下，单颗

HBM容量提升至24GB，美光HBM3E采用1 β nm DRAM和TSV技术，正在获得英伟达的产品资格认证，预计2024年初量产，2024财年产生7亿美元收入。

美光预计于2026年推出HBM Next产品。美光正在开发HBM Next产品，预计2026年推出，内存容量预计提升至36-64GB，带宽为1.5-2TB/s。

七

报告综述及赛道活跃企业

1、全球先进封装设备及材料份额主要被海外占据，AI需求激增带来部分厂商创纪录的先进封装订单

根据我们前文所述，半导体先进封装尤其是HBM CoWoS封装核心增量来自Bumping、TSV、RDL等工艺，增量需求最大的设备主要为检/量测、减薄、电镀、键合、模塑机台，并且先进封装也为激光切割机、固晶机、回流炉、后道测试机/分选机/探针台等传统封装设备带来一定需求提升。

用于高端先进封装的设备份额主要为海外厂商占据，例如先进封装检/量测领域的Camtek、ONTO；减薄领域的DISCO；电镀领域的AMAT、ASMPT；键合领域的SUSS、EVG Group；塑封领域的TOWA、YAMADA等；芯片贴装领域的BESI等。由于AI等领域需求旺盛，Camtek、BESI、SUSS等厂商面向先进封装如HBM和异构集成、AI应用的订单均创下历史新高，并仍在持续增长；

用于高端先进封装的材料品类繁多，且更为分散，主要包括IC封装载板、ABF基板层介电材料、制造封装基板核心层材料、环氧树脂固态封装材料、导线架、焊线材、底部填充剂等，市场份额大多被美国、德国、日本等厂商占据，核心供应商包括日本住友化学、日本昭和电工、美国杜邦等。

- Camtek：先进封装检/量测设备龙头，HBM、异构集成等领域订单激增

Camtek是以色列自动光学检测（AOI）和量测设备厂商，涵盖从Wafer、先进封装、PCB的测试。公司产品为Eagle系列，在前端制程领域包括Eagle-i和Eagle-I PLUS型号，可用于大规模2D检测，包括电镀bump、电测针印、划片后的检测等，还可用于CMOS、MEMS、LED等器件的检测；在先进封装领域，公司设备包括Eagle-AP和Golden Eagle型号，主要面向Bump、RDL、TSV等工艺，以及扇外型PLP封装面板等应用。公司设备2D检测精度0.2 μ m，3D检测高度精度0.05 μ m（量测范围2-100 μ m），检测速度为量测单片5000万点的Bump。

Camtek大部分收入来自先进封装和中国地区，大客户覆盖三大存储原厂、台积电、英特尔等。Camtek为先进封装尤其是HBM领域的量测设备龙头，收入从2017年的0.93亿美元增长至2022年3.21亿美元，2022年收入60%以上来自先进封装互连封装应用，其中很大一部分来自HBM和Chiplet，2022年收入44%来自中国地区。公司客户70%以上业务面向Tier1厂商，包括台积电、联电、三星、SK海力士、美光、英飞凌、日月光、安靠等。2023年11月，Camtek以1亿美元完成对德国FormFactor（FRT）公司的收购，FRT是先进封装和SiC市场高精度计量设备领先厂商，具有独特的混合多传感器SurFaceSens技术。

Camtek先进封装订单积压至2024年，指引2024年收入创新高。Camtek自2023年Q3开始不断接收一级制造商客户的订单，大部分来自先进封装领域的异构集成、HBM、扇外型封装等。2023年7月，公司公告称收到了多家一级制造商的42套系统订单，很大一部分用于异构集成的小芯片模块和HBM，这些订单将于23H2交付；8月底，公司公告自2023年8月以来已收到约45套系统的订单，30%用于HBM和Chiplet应用；11月，公司公告收到一家一级制造商的28套系统新订单，用于高带宽的检测、HBM、异构集成（HI）应用程序，进一步增加了将于2024年交付的订单积压量。展望2024年，公司预计整体收入创历史新高，HBM、Chiplet领域收入占比超30%。

- **BESI：芯片固晶机龙头，混合键合设备构筑成长曲线**

BESI聚焦封装固晶/贴装机，具备晶圆级封装和混合键合能力。BESI位于荷兰，是后道封装芯片贴装设备龙头，最大的下游为智能手机，2022年收入79%来自芯片贴装，21%来自封装和组装设备。公司目前具备晶圆级先进封装和混合键合技术能力，23Q1推出混合键合机台8800 CHAMEO ultra plus机台。

BESI是芯片贴装领域龙头，在先进封装贴片领域占比超70%份额。2022年全球封装市场规模大约55亿美元，BESI面向的市场空间大约20亿美元（不包括引线键合、切片等市场），市场份额大约32%，其中Die Attach（芯片贴装）市场空间大约13亿美元，BESI份额大约40%，先进封装芯片贴装市场空间大约4亿美元，BESI占据74%的份额；在芯片封装和组装领域，公司面向的市场为6.83亿美元，占据18%的份额。

混合键合设备构筑公司成长曲线，相关订单持续快速增长。23Q3公司收入和净利润分别为1.233亿欧元和3500万欧元，分别同比下滑24%和33.5%，主要系整体市场下滑和高端智能手机需求疲软；不过，公司23Q3订单环比增长13%，主要系AI及混合键合需求增加，23Q3公司从2家客户收到了新的混合键合订单，其中包括一家领先的客户，以及来自行业领先客户的光子应用的大量订单；23Q3至今，公司收到了来自逻辑和存储客户的更多混合键合订单，预计23Q4订单继续增加；另外，公司在23Q3至今收到了多个来自亚

洲的芯片贴装订单，用于晶圆级芯片和基板的贴装。基于晶圆级封装产品的订单，公司预计23Q4混合键合和其他先进封装设备积压的订单持续发货，指引23Q4收入将环比增长15-25%。

- **SUSS：后道涂胶显影和临时键合/解键合领先厂商，AI需求带来创记录的临时键合机台订单**

SUSS产品组合涵盖后道涂胶显影、键合和光掩模设备，键合和光掩模设备收入强劲增长。SUSS总部位于德国，拥有70多年工程经验，产品矩阵涵盖掩模设备，涂胶显影机、纳米压印光刻机、晶圆键合机等，2023年前三季度，公司实现收入2.023亿欧元，同比增长21.8%，其中光掩模解决方案和键合机分别同比增长61.3%和56.2%，截至三季度末总积压订单为4.147亿欧元。

AI应用为SUSS带来创纪录的临时键合机台订单，中国台湾大客户增加对SUSS键合设备需求。在AI芯片制造过程中，HBM必须磨得尽可能薄，对于研磨过程和进一步加工，必须将晶圆临时键合到第二片晶圆上来加固，之后通过脱键再次释放键合，然后必须清除晶圆上的粘合剂残留物。

SUSS的XBS300和XBC300机台可以为HBM临时键合全流程提供完善的解决方案，23Q3，SUSS共产生1.03亿欧元的订单，其中高端后道设备订单共9370万欧元，键合设备订单高达6210万欧元；从2023年6月到10月底，SUSS已经被预定了价值约1亿欧元的AI应用临时键合解决方案；中国台湾某个世界领先的先进封装代工厂也采用了SUSS的XBS300设备，用于逻辑和存储芯片的耦合，SUSS表示，中国台湾客户产线未来每年将采用多达12台键合机。

- **TOWA：半导体塑封机龙头，高端压塑型号用于HBM和2.5D等封装形式**

TOWA布局半导体后道贴片、植球、塑封成型机等多种设备，是全自动塑封机等设备龙头。TOWA总部位于日本，产品主要面向半导体后道封装，包括贴片机、植球机、焊线机、倒装贴片机、塑封成型机等。公司是半导体先进封装塑封成型机龙头，包括压缩成型和注塑成型两大类设备。根据SEMI，针对中国大陆市场的全自动塑封设备，TOWA每年销量约200台、YAMADA约50台、BESI约50台。

TOWA CPM 1080型号支持晶圆级（WLP）和面板级（PLP）等先进封装形式，FY23H1订单超过FY22全年。CPM 1080是公司面向晶圆级先进封装（WLP）等领域的高端全自动塑封机，可用于HBM的压缩成型。公司表示，生成式AI带来HBM和2.5D封装的压缩塑封成型需求，公司2023财年上半年的订单已经超过2022财年全年。

- **ASMPT**：后道先进封装平台型设备厂商，新增订单来自先进封装和汽车终端

ASMPT设备覆盖几乎先进封装全流程，新增订单来自先进封装和汽车终端。ASMPT总部位于新加坡，为电子制造全流程提供解决方案，公司半导体解决方案包括PVD沉积、电镀、激光划片、热压键合、D2W混合键合，以及测试分选机等设备，先进封装领域重点市场包括WLP、2.5D/3D IC、TSV、扇出、嵌入式芯片等。公司23Q3半导体部门收入受景气度影响同比下滑28.7%至15.7亿港元，亏损1.1亿港元，新增订单同比下滑10.9%至13.3亿港元，但环比增长4.4%，公司表示新增订单主要来自先进封装和汽车终端市场应用，AI服务器需求也有所增长。

- **住友电木**：目前环氧塑封料（EMC）市场份额主要由日本住友电木和日本昭和等占据，日本住友电木主要面向半导体用环氧化合物、电子和电器安装用酚醛树脂及电子器件电路连接用的各向异性导电薄膜和其他化学品。
- **美国杜邦**：半导体领域材料广泛覆盖制造、封装及组装领域，在半导体封装领域，公司产品涵盖电镀光刻胶、铜柱电镀中的锡银帽、RDL层、UBM材料等，组装材料涵盖芯片粘结剂、封装剂、永久键合电介质、热界面材料等。在PI（聚酰亚胺）方面，公司最早开发和生产，市占率全球第一，重视对组分的研究改性，并关注高模量、柔韧性等机械性能。
- **德国汉高**：汉高（Henkel）成立于1876年，其产品在胶粘剂市场占有率全球第一，公司工程胶黏剂、密封剂和表面处理方面的系列产品涵盖了锡膏、厌氧胶、环氧胶、硅胶、瞬干胶、UV胶、PU胶、MS聚合物、清洗剂等八大系列。

2、HBM对封装工艺、设备、材料需求均提升，国内厂商在先进封装领域不断导入

AI对GPU高带宽需求催生HBM百亿美金市场，以HBM为代表的先进封装封测技术、设备和材料等需求有望持续提升。多用于PC显卡的GDDR无法满足AI服务器的高带宽要求，因此HBM等新型存储器应运而生，HBM能够以低功耗实现高带宽，非常适合AI服务器的应用场景。伴随着训练型AI服务器出货量增长、AI服务器渗透率的提升、单GPU搭载HBM叠层的数量提升、以及HBM堆叠DRAM Die数量和DRAM容量的提升，预计到2024年HBM市场空间有望超百亿美元。以HBM为代表的先进封装市场对封测技术、先进封装设备及材料需求持续增长，国内先进封装封测、设备、材料等产业链有望持续受益。

- **设备**：

中科飞测：金属膜厚设备批量出货，纳米图形缺陷检测和关键尺寸量测设备开发顺利。公司目前收入主要来自无图形和图形晶圆缺陷检测设备、三维形貌量测设备，套刻精度量测设备和金属膜厚量测设备成为收入新增长点，套刻精度量测设备在成熟工艺节点实现批量供应，在先进工艺产线已经通过部分客户验证，多台金属膜厚量测设备已实现批量出货。针对2Xnm节点，公司明场和暗场纳米图形晶圆缺陷检测设备、关键尺寸量测设备研发进展顺利；针对1Xnm节点，持续性研发布局多款检测和量测设备。

北方华创：在先进封装领域，针对Flip chip Bumping、Fan-Out、WLCSP、2.5D/3D TSV等技术，公司刻蚀设备、沉积设备、炉管设备等已经实现了在主流先进封装企业的批量生产，并不断获得客户的重复采购订单。①等离子去胶机BMD P300可兼容Fan-out大翘曲晶圆，并具备多种Descum工艺处理能力，如PI curing后，PR显影后电镀前，PR strip后，Ti/Cu刻蚀后，Molding前，Underfill前等；②Polaris B系列PVD，可兼容大翘曲Fan-out圆片，并支持Si, EMC, Glass, Bonding等多种基片的精准高效传输；③12英寸PI胶固化系统（PIQ）SUMERIS AP302C可在更低氧的环境下实现对Polyimide（聚酰亚胺）的固化、精准控温。

中微公司：公司等离子体刻蚀设备可用于先进封装生产线，其深硅刻蚀设备Primo TSV 200E、Primo TSV 300E可用于IC 3D封装、CMPS、MEMS等，在晶圆级封装、2.5D封装和MEMS系统等领域持续获得重复订单，在12英寸3D芯片的TSV刻蚀工艺上成功验证。

拓荆科技：子公司拓荆键科（海宁）产品包括晶圆对晶圆键合（Wafer to Wafer Bonding, W2W）产品和芯片对晶圆键合表面预处理（Die to Wafer Bonding Preparation and Activation, D2W）产品。拓荆科技键合设备主要由控股子公司拓荆键科（海宁）开展，拓荆键科成立于2020年9月30日，母公司持股比例为55%。拓荆键科联合拥有先进晶圆片键合机的技术储备的海宁君鑫科技，利用自身掌握的晶圆键合对准技术，进而开拓晶圆键合设备市场。

1) 晶圆对晶圆常温混合键合（Hybrid Bonding）和熔融键合（Fusion Bonding）：可以实现复杂的12英寸晶圆对晶圆常温共价键合，搭载了晶圆表面活化、清洗、键合和自研的键合精度检测模块，具有对准精度高、产能高、无间隙等性能特点。首台晶圆对晶圆键合产品（Dione 300）已通过客户验收并获得重复订单；

2) 晶圆及切割后芯片的表面活化及清洗：可以实现芯片对晶圆键合前表面预处理工序，包括晶圆及切割后芯片的表面活化及清洗工艺。首台W2W产品Dione300已经出货至客户端验证，并取得突破性进展，D2W产品Pollux完成研发，正在客户端验证。

华海清科：基于CMP技术自研12英寸减薄设备，量产机台已发往龙头客户端并获得先进存储、Chiplet封装等多个订单。2023年5月17日，公司全新一代12英寸超精密晶圆减薄机Versatile-GP300量产机台出机发往集成电路龙头企业，标志着公司自研的国产减薄设备批量进入大生产线。该款设备用于前道晶圆制造背面减薄工艺，是业内首次实现12英寸晶圆超精密磨削和CMP全局平坦化的有机整合集成设备，自主研发的超精密晶圆磨削系统稳定实现12英寸晶圆片内磨削TTV<1um，达到了国内领先和国际先进水平。华海清科创新开发的CMP多区压力智能控制系统，突破传统减薄机的精度限制，实现了减薄工艺全过程的稳定可控。Versatile-GP300机台现已收获包括先进存储、Chiplet封装等技术领域在内的多个订单，近期将陆续出机；另外，公司用于封装领域的12英寸超精密减薄机各项性能指标达到预期目标，已经发往客户端进行验证。

盛美上海：先进封装领域布局湿法类、电镀、涂胶显影等设备，电镀设备有望保持高增长。公司产品线覆盖湿法清洗、电镀、炉管、无应力抛铜、后道先进封装及其他类设备，公司先进封装产品线完整，覆盖电镀、涂胶显影、湿法刻蚀、湿法去胶、金属剥离、无应力抛光先进封装平坦化、清洗设备等，23H1先进封装及其他后道设备收入同比增长47%。公司电镀设备可用于前道和后道，针对先进封装的包括大马士革电镀、TSV电镀、先进封装电镀、第三代半导体电镀设备均开发出来，在工艺上得到验证，公司电镀设备获得较多重复订单，预计明年仍将保持高速增长。其中，公司用于3D TSV和2.5D转接板的三维电镀设备Ultra ECP 3D可为高深宽比（深宽比大于10:1）铜应用提供高性能、无孔洞的镀铜功能。

芯源微：针对Chiplet等先进封装领域开发临时键合机/解键合机，目前正在客户端验证。公司产品线包括涂胶显影机、物理清洗机等，覆盖前道IC、后道先进封装及小尺寸等领域。公司涂胶显影机在国内后道先进封装领域市占率第一，加深和盛合晶微、长电绍兴、上海易卜等国内新兴封装势力的合作关系；针对Chiplet等新兴市场，基于在三维封装工艺的技术储备和前期应用，公司开发了临时键合/解键合机台，目前产品均进入客户验证阶段。

精测电子：覆盖半导体膜厚量测、关键尺寸量测、电子束缺陷检测等设备，持续获得先进制程订单。公司产品面向显示、半导体及新能源检测系统，在半导体领域，公司产品包括检测和量测设备，包括膜厚量测、光学关键尺寸量测、电子束缺陷检测和面向Memory老化、晶圆探测、终测的自动检测设备（ATE）等，其核心产品已覆盖2xnm及以上制程，膜厚产品、OCD设备及电子束缺陷复查设备已取得先进制程订单，23Q3公司先进制程产品订单已实现部分交货且取得重复订单。

芯碁微装：直写光刻机从PCB领域向泛半导体领域延伸，在晶圆级封装的RDL、Bumping、TSV等工艺中优势明显。公司主营直写光刻机，用于PCB和泛半导体领域，公司加快在载板、先进封装、新型显示、掩

膜版制版、功率分立器件等方面的布局。在先进封装领域，公司直写光刻优势主要体现在智能纠偏上，直写光刻在晶圆重构封装中解决偏移问题能力较强。公司设备能够实现再布线，WLP2000系列产品采用多光学引擎并行扫描技术，具备自动套刻、背部对准、智能纠偏、WEE/WEP功能，在RDL、Bumping和TSV等工艺中优势明显。

文一科技：半导体封测领域产品包括集成电路封装模具、自动切筋成型系统、分选机、塑封压机、自动封装系统、芯片封装机器人集成系统、半导体精密备件等。针对先进封装领域，公司正在研发满足晶圆级封装用的模具和设备。

至正股份：23H1子公司苏州桔云纳入公司财务报表，新增半导体专用设备业务，其主要面向半导体后道先进封装，主要产品包括清洗设备、烘箱设备、腐蚀设备、涂胶显影设备、去胶设备、分片设备等。

至正股份：23H1子公司苏州桔云纳入公司财务报表，新增半导体专用设备业务，其主要面向半导体后道先进封装，主要产品包括清洗设备、烘箱设备、腐蚀设备、涂胶显影设备、去胶设备、分片设备等。

新益昌：公司固晶机布局较为完善，焊线设备部分产品通过客户验证，形成小批量出货，先进封装领域的部分高精密设备正在验证，预计短期内设备将通过客户验证。

光力科技：公司半导体封装设备可用于IC、分立器件、光电器件、传感器等多种半导体产品的封装工艺，客户主要为OSAT和IDM厂商，其12英寸全自动划切设备ADT8230实现高端划切设备的国产替代。公司同时开发激光切割划片机、半导体研磨机等新品，计划明年推出。

德龙激光：半导体晶圆激光隐形切割设备可用于硅/砷化镓/碳化硅的晶圆切割加工；晶圆激光开槽设备主要用于半导体40nm及以下线宽的low-k晶圆的表面开槽；晶圆级封装产品综合加工设备兼容晶圆级封装产品的精密加工切割、钻孔、刻蚀、表面处理、开槽。

赛腾股份：收购日本OPTIMA获得晶圆缺陷检测技术。产品用于半导体、光伏、锂电、消费电子、8/12英寸晶圆等，非标准化设备主要用于消费电子和新能源行业，标准化设备主要用于半导体、光伏及锂电领域，包括固晶设备、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、锂电切叠一体机、卷绕机、锂电电芯组装设备以及光伏组件自动化单机及整线等。公司于2019年收购日本OPTIMA，截至23H1持股比例为74%，其晶圆缺陷检测设备和技术全球领先。

耐科装备：公司主营半导体封装设备和模具，塑料挤出成型模具、挤出成型装置及下游设备。公司晶圆级封装设备处于研发过程中，关键装置压机单元在试验阶段。

亚威股份：公司于2021年2月投资苏州芯测电子有限公司，持有其25%的股权，苏州芯测布局高端半导体存储芯片测试设备业务。目前，苏州芯测已完成对韩国GSI 100%的股权收购，GSI成立于2014年，拥有技术难度较高的存储测试机业务，并稳定供货SK海力士、安靠等厂商。

劲拓股份：半导体封装设备主要包含半导体芯片封装炉、Wafer Bumping焊接设备、真空甲酸焊接设备、甩胶机、氮气烤箱、无尘压力烤箱等，广泛应用于各类芯片元器件的封装过程。公司半导体封装炉设备适用于CoWoS其中一种回流焊接工艺。

迈为股份：立足真空、激光、精密设备三大技术平台，在半导体封测领域布局了刀轮切割、激光改质切割、激光开槽设备等。

奥特维：半导体业务主要集中在封测环节，目前布局了划片机、装片机、键合机、AOI等设备，CMP设备处于研发的初始阶段。

长川科技：数字SoC测试机和三温分选机快速放量，探针台逐步起量。公司产品包括测试机、分选机、探针台、AOI设备等，基本覆盖后道测试设备全品类，其测试机具备模拟、功率和数字信号测试机量产能力，正从SoC测试领域向存储等领域延伸；公司分选机包括重力下滑式和平移式分选机等，通过收购马来西亚EXIS获得转塔式分选机产品线，当前三温分选机正在快速放量；探针台产品以全自动晶圆探针台为主，12寸晶圆探针台实现量产出货，正在逐步起量。

华峰测控：公司产品主要为半导体自动化测试系统，用于模拟、数模混合、分立器件和功率模块等，不断拓展氮化镓、碳化硅及IGBT等功率分立器件和模块类测试领域，并推出了面向SoC测试领域的新机型STS8600。

金海通：半导体分选机产品覆盖安靠、长电科技、通富微电等封测客户，三温分选机贡献成长动力。公司聚焦半导体后道测试领域分选机，客户覆盖安靠、联合科技、长电科技、通富微电等封测厂商，博通、瑞萨等IDM企业，以及兴唐通信、澜起科技等IC设计厂商。公司基于自主软件架构和算法开发了集成式三温分选机EXCEED 9000产品，预计将成为长期增长动力。

- 材料：

鼎龙股份：半导体用精抛光垫可用于Grinding（晶背研磨）等制程，部分产品在客户端批量使用，预计明年实现产销量快速增长；多晶硅抛光液、金属铝/钨栅极抛光液导入客户、取得批量订单，有望在第四季度开

始逐步放量，铜和阻挡层抛光液有望在第四季度取得批量订单；拥有YPI、PSPI产品，用于半导体显示领域；封装光刻胶已有两款产品在客户端测试，另有一款新立项的封装光刻胶产品正在客户端送样。

安集科技：CMP铜及铜阻挡层抛光液持续放量，钨抛光液、基于氧化铈磨料的抛光液份额持续提升，介电材料抛光液、衬底抛光液客户进一步拓展；在电镀液及添加剂方面，公司完成了应用于集成电路制造及先进封装领域的产品系列平台的搭建，多款电镀液及添加剂产品在先进封装领域进入量产导入阶段。

雅克科技：公司主营电子材料、LNG保温绝缘板材、阻燃剂业务，公司积极推进国内工厂的电子材料产品在国内客户端的评估，实现部分产品对国内客户的批量供应；公司同时在前驱体材料方面积极推进技术研发，与美光、海力士、台积电、长江存储、合肥长鑫等合作，推进更先进制程产品的研发和验证。考虑到三星、海力士等HBM接单量快速增长，也有望带动前驱体材料需求持续增长。

强力新材：公司主营光刻胶专用电子化学品和绿色光固化材料，其PAG系列光刻胶光引发剂用于i线、KrF线半导体光刻胶和封装材料领域。公司研发生产的PSPI用于封装领域，目前处于下游客户验证阶段。

天承科技：在先进封装领域，公司上海工厂二期项目已经启动，拟投入5000万元用于半导体相关的电镀液等功能性湿电子化学品的生产设备和车间改造，计划明年1月份投产，主要用于先进封装和TSV部分，对晶圆的大马士革电镀后续有产品计划，预计2024年收入上量。

华海诚科：国内环氧树脂塑封料龙头，电子胶黏剂核心供应商。公司专注于半导体封装材料，主要产品为环氧塑封料和电子胶黏剂，公司已于华天科技、通富微电、长电科技、富满微、扬杰科技、气派科技、银河微电等下游知名厂商建立了长期良好的合作关系，相关产品已在上述部分厂商实现对外资厂商产品的替代。公司EMG-900-ACF（GMC颗粒状塑封料）在重庆矽磐微、合肥矽迈、中科芯、通富微电等客户验证，预计在客户A年底会有批量；68系列（LMC液态塑封料）在通富微电验证，模塑性验证合格，在其他客户处进行可靠性考核。

联瑞新材：产品大量用于先进封装EMC、LMC、底部填充等材料，配套供应HBM全球知名GMC供应商所用球硅和Low α 球铝。公司主营工业粉体材料，公司采用火焰熔融法、高温氧化法、液相法三种主流工艺，生产微米级及亚微米级球形硅微粉，产品广泛应用于芯片封装和基板用环氧塑封材料（EMC）、液态塑封材料（LMC）和底部填充材料（Underfill）、印刷电路板用覆铜板（CCL）、积层胶膜、热界面材料（TIM）等领域。产品中Low α 微米级球形硅微粉、Low α 亚微米级球形硅微粉主要应用于存储芯片封装等先进封装领域，Low Df（低介质损耗）球形硅微粉广泛应用于各等级高频高速基板。Low α 球形氧化硅主要用于高

导热存储芯片封装等高端芯片封装领域。公司部分客户是全球知名的GMC供应商，公司配套供应HBM所用球硅和Low α 球铝。

壹石通：规划建设200吨高端芯片封装用Low- α 射线球形氧化铝项目，目前进入产线调试阶段。公司高纯二氧化硅、Low- α 的射线球形氧化铝产品主要用于芯片封装领域，公司高纯二氧化硅粉体材料已为日本雅都玛等日韩企业长期稳定供货，其Low- α 射线球形氧化铝产品打破海外垄断。根据公司2022年定增项目所做的市场调研，Low- α 射线球形氧化铝的存量市场需求大约为1000吨/年，公司定增项目年产200吨高端芯片封装用Low- α 射线球形氧化铝项目已经进入产线调试阶段，目前主要客户集中在日韩。

飞凯材料：公司控股子公司昆山兴凯半导体材料公司主要负责环氧塑封料的产销研，公司通过全资子公司安庆飞凯新材料持股60%，长兴化学工业公司通过长兴投资有限公司持股40%。环氧塑封料的原材料市场竞争关键在于树脂的供应，而长兴化学作为另一股东，能为公司树脂需求提供助力。

德邦科技：主营电子封装材料、导热材料、导电材料、晶圆划片膜、减薄膜等400余种产品，在IC领域，23H1公司UV膜、固晶胶、热界面材料收入分别占比三成、三成、四成，公司底部填充胶、AD胶、固晶胶膜（DAF/CDAF）、芯片级导热界面材料（TIM1）四款芯片级封装材料同时在配合多家设计公司、封测公司推进验证。

兴森科技：珠海FC-BGA封装基板项目拟建设产能200万颗/月的产线，于2022年12月建成并试产，目前处于客户验证阶段，部分大客户的技术评级、体系认证均通过；广州FC-BGA封装基板项目拟分期建设2000万颗/月的产线，一期厂房已于2022年9月完成封顶，目前处于设备安装阶段，预计23Q4完成产线建设。

生益科技：针对服务器需求产品有全系列布局，包括Mid-loss, Low-loss、Very-low loss、Ultra-low loss、Extreme Low-loss及更高级别材料，AI服务器覆铜板表现较好，正开发高效散热封装基板用覆铜板关键技术。

深南电路：广州封装基板项目主要面向FC-BGA、FC-CSP、RF封装基板三类，项目分两期建设，一期已于2023年10月下旬连线，后续进步产能爬坡阶段。FC-BGA封装基板中阶产品目前已在客户端顺利完成认证，部分中高阶产品进入送样阶段，初步建成高阶产品样品试产能力。

神工股份：公司大直径单晶硅材料直接向电极制造商销售，经电极制造商机械加工制程IC刻蚀用硅电极，直接用于芯片制造刻蚀环节，公司客户覆盖日本、韩国等半导体知名硅零部件厂商。公司“集成电路刻蚀设备用硅材料扩产”定增项目积极扩大大直径硅料产能，达产后现有刻蚀用硅材料产能将从500吨/年增长至900吨/年。

上海新阳：公司产品主要包括①晶圆制造及先进封装用电镀液和添加剂系列，包括大马士革铜互联、TSV、Bumping电镀液及配套添加剂；②晶圆制造用清洗液、刻蚀液系列，主要包括铜制程刻蚀后清洗液、铝制程刻蚀后清洗液、氮化硅/钛蚀刻液、CMP后清洗液等；③IC制造用高端光刻胶系列，包括I线、K线、ArF干法和浸没式、底部抗反射膜（BARC）等材料；④CMP抛光液，用于STI、金属钨、金属铜、硅氧化层、多晶硅层等；⑤半导体封装用镀锡化学材料及配套电镀前处理、后处理化学材料等。

华正新材：主要从事覆铜板及粘结片、半导体封装材料、复合材料和膜材料等产品，其覆铜板用于制作PCB，半导体封装材料包括BT封装材料和CBF积层绝缘膜，适用于Chiplet、FC-BGA等先进封装工艺，主要用于Memory、MEMS、RF、ECP嵌入技术即CPU、GPU、FPGA、ASIC等算力芯片的半导体封装。

方邦股份：公司产品包括电磁屏蔽膜、各类铜箔、挠性覆铜板、电阻薄膜、复合铜箔等，其中带载体可剥离超薄铜箔是制备芯片封装基板、HDI板的必需基材，极薄挠性覆铜板是实现高密度互连技术的关键材料之一。

回天新材：公司在芯片封装用胶板块相关产品包括芯片四角邦定胶（edgebond）、芯片底部填充胶（underfill）、SIP屏蔽银浆等，其与H公司在5G通信、消费电子、数字能源、汽车电子等板块均有合作，供应的产品包括环氧底填、PUR胶、三防漆、UV胶等。

国风新材：公司PI薄膜经过深加工后可用于消费电子等设备的柔性电路板、散热材料等领域，半导体封装用PSPI光刻胶研发处于实验室送样检测阶段。

- 封测：

长电科技：公司拥有晶圆级封装、倒装芯片互连、硅通孔(TSV)等多种技术，在2.5D封装领域拥有成熟MEOL TSV集成经验。2022年，公司认证通过TSV异质键合3D SoC的fcBGA技术；2023年1月，公司XDFOI Chiplet高密度多维异构集成工艺实现量产，实现国际客户4nm节点封装产品出货，既有TSV less，也有TSV方案，适用于HBM等芯片。

通富微电：公司多层堆叠NAND Flash及LPDDR封装实现稳定量产，同时在国内首家完成基于TSV技术的3DS DRAM封装开发。公司南通通富工厂三期工程稳步推进，并预计该先进封装生产线建成后，公司将作为国内最先进的2.5D/3D先进封装研发及量产基地，实现国内在HBM（高带宽内存）高性能封装技术领域的突破。

AMD在CES 2023展会上推出了下一代面向数据中心的APU产品Instinct MI300，采用Chiplet设计，由13个小芯片组成，晶体管数量高达1460亿个，超越英特尔的1000亿晶体管的Ponte Vecchio。国内通富微电子于2016年收购AMD苏州和槟城两家工厂，多年来一直和AMD形成合作伙伴关系，承担了AMD主要的封测业务，公司此前于投资者互动平台表示，通富有涉及AMD Instinct MI300的封测项目。考虑到AMD未来将更多导入Chiplet的架构设计，通富微电或将持续受益。

华天科技：公司具备3D、SiP、MEMS、FC、TSV、Bumping、Fan-out、WLP等先进封装技术。2023年3月，公司宣布投资28.58亿元进行“高密度高可靠性先进封测研发及产业化”项目，达产后预计形成Bumping 84万片、WLCSP 48万片、超高密度删除UHDFO 2.6万片的晶圆级封测能力，建设期为5年。

深科技：公司于2015年收购金士顿旗下沛顿科技100%股权，沛顿科技主要从事高端存储芯片的封测，在DRAM封测实力较强，产品包括DDR3、DDR4、DDR5、LPDDR3、LPDDR4、LPDDR5、eMCP4等。技术方面，公司具备先进封装FlipChip/TSV技术（DDR4封装）能力，高端3D TSV等技术持续开发；产能方面，深科技孙公司合肥沛顿存储科技于2020年10月成立于合肥市经济技术开发区空港示范区，是沛顿科技在华东地区的运营基地，为国内主要客户提供封装测试、模组组装等全套服务。

太极实业：公司半导体业务主要涉及IC芯片封装、封装测试、模组装配及测试等，子公司海太公司半导体业务目前主要是为SK海力士的DRAM产品提供后工序服务，拥有完整的封装测试生产线与SK海力士12英寸晶圆生产线紧密配套；子公司太极半导体在传统封装工艺（FC）基础上，开发了高阶混合封装（Hybrid, FC+WB）工艺。

甬矽电子：公司全部产品均为QFN/DFN、WB-LGA、WB-BGA、Hybrid-BGA、FC-LGA等中高端先进封装形式，并在系统级封装（SiP）、高密度细间距凸点倒装产品（FC类产品）、大尺寸/细间距扁平无引脚封装产品（QFN/DFN）等先进封装领域具有技术储备；公司通过开展Bumping项目掌握了RDL及凸点加工能力，后续逐步拓展晶圆级封装、扇出式封装及2.5D/3D封装；公司控股子公司甬矽半导体（宁波）拟投资建设高密度及混合IC封测项目，项目总金额预计不超过21.57亿元，预计可新增年产87000万颗高密度及混合集成电路封装测试，具体投向FC-LGA、FC-CSP、FC-BGA及Hybrid-BGA类产品。

3、AI服务器对存储容量有数倍拉动，行业边际复苏和自主可控趋势带动国内需求增长

AI服务器存储伴随CPU和GPU升级而迭代，对存储容量和价值量均有数倍拉动。传统服务器中，CPU作为算力来源，最终数据储存在硬盘中，CPU内存作为CPU和硬盘之间的桥梁，能够暂时存放CPU的运算数据并与硬盘等外部存储器进行数据交换；传统服务器的CPU主要通过提升核心数进行算力提升，但AI训练模

型算力增长速度已经超过CPU可以满足的算力需求，因此高性能推理服务器或者AI服务器中增加了能够多线程吞吐数据的GPU，GPU由CPU来发出指令调度。伴随着AI算力要求的提升，CPU和GPU均不断升级迭代，CPU显存、服务器SSD的容量和价值量均有数倍增长，另外AI服务器对GPU的需求也催生了GPU显存的纯增量市场。

存储价格持续反弹，美光和SK海力士增加2024年存储资本支出。本轮存储行业下行周期自22Q1开始，存储量价齐跌带来原厂收入和盈利能力快速下滑，在22Q4-23Q1，几大存储原厂均出现巨额亏损；伴随着6-7个季度持续的减产，当前供给侧产能明显收缩，下游终端和原厂库存持续改善，海外原厂收入和利润表现自23Q2以来均有所改善。站在当前时点，美光和SK海力士均对存储行业复苏展望乐观，并适当增加2024年的资本支出；从价格端来看，当前整体价格处于底部，高端HBM、DDR5等价格持续上涨，部分中低端型号价格也有所上升，展望23Q4-2024年，价格有望持续反弹。

美光在华销售产品审查未通过，国产自主可控进程加速。2023年5月21日，据网信办发布，美光公司产品存在较严重网络安全问题隐患，对我国关键信息基础设施供应链造成重大安全风险，影响我国国家安全。为此，网络安全审查办公室依法作出不予通过网络安全审查的结论。按照《网络安全法》等法律法规，我国内关键信息基础设施的运营者应停止采购美光公司产品。美光回应表示，中国国家互联网信息办公室（CAC）审查决定对美光的业务影响目前仍不确定，包括移动设备制造商在内的某些美光客户已接到中国的关键信息基础设施（CII）运营商或政府代表的通知，涉及到未来使用美光产品的事宜。美光总部在中国约50%的业务可能会受到影响，相当于美光全球收入的低两位数百分比，可能会受到影响。尽管美光受影响的一部分市场可以由三星、SK海力士等填补，但将加速长鑫存储等原厂国产DRAM导入，以及存储芯片、模组、封测、主控芯片等产业链标的国产自主可控进程。

建议关注国内存储芯片标的兆易创新、紫光国微、复旦微电、北京君正、聚辰股份、普冉股份、东芯股份、恒烁股份、上海贝岭等；存储模组及主控标的江波龙、佰维存储、朗科科技、德明利、国科微等。

- 存储经销

香农芯创：公司自2020年起进入半导体领域，立足半导体分销业务，在存储器及主控芯片领域，与SK海力士、MTK联发科等具备长期合作关系，并且公司目前重点布局企业级SSD的研发。

雅创电子：拟收购的WE主要代理SK海力士的存储器，上述产品将作为未来布局的重点方向，目前尚未形成销售额。

- AI服务器及HBM配套（Raid、CXL芯片等）：

国芯科技：公司基于C*Core CPU内核C8000的第一代Raid芯片产品，具备多个独立的接口通道、支持连接最多40个机械硬盘或SSD固态存储盘，兼容PCIE标准，目前已经量产流片；另外，目前公司国家重大需求、信息安全以及边缘计算和网络通信等领域有多个SoC芯片正在进行多芯片合封，最多已经实现6颗裸Die的合封，目前正在研究规划合封多HBM内存的2.5D的芯片封装技术。

澜起科技：2022年5月6日，公司宣布发布全球首款CXL内存扩展控制器芯片（MXC），转为内存AIC扩展卡、背板及EDSFF内存模组而设计，可大幅扩展内存容量和带宽，满足HPC、AI等数据密集型应用需求。未来CXL有望搭配HBM使用，HBM需求增长也有望带动CXL用量增长。

创益通：公司主营数据存储互连产品、消费电子互连产品、通讯互连产品和精密结构件，数据存储互连产品主要包括各种型号的高速连接器、高频高速数据线等。

八

产业发展挑战

- 1) AI服务器渗透率提升不及预期。当前全球AI服务器渗透率依旧较低，如果AI服务器渗透率提升不及预期，那么将影响存储市场空间的提升以及HBM等新型存储器、TSV等先进封装技术的发展。
- 2) 存储行业复苏不及预期。存储行业下游手机、服务器、PC等需求仍未明显复苏，如果下游复苏不及预期，那么行业供需平衡状态预计更晚到来，价格涨势或将无法延续。
- 3) 国产替代不及预期。如果存储行业国产替代进程放缓，那么不利于存储原厂、芯片、模组、封测厂商等的国产化导入，进而会对国内存储产业链造成不利影响。
- 4) 研发进展不及预期。国内厂商在DRAM、NAND、HBM等存储器的技术方面仍落后于海外原厂，仍处于持续追赶过程中，长期来看，假如国内存储原厂如合肥长鑫、长江存储，及国内存储芯片、模组、封测等产业链标的研发能力下降，那么将影响国内存储产品发布。

本文转载自招商电子，内容仅供交流学习之用，如有任何疑问，敬请与我们联系info@gsi24.com。

▼ 往期精彩回顾 ▼

[高阶自动驾驶，需要一颗什么样的芯片？](#)

[行业裁员滚滚！这类芯片人才却被疯抢，平均月薪7万](#)

[中国倒爷俄罗斯卖车，富贵险中求](#)

[英伟达的芯片，不够香了](#)

[金士顿，原来你是这样的内存一哥！](#)

[2500亿元！半导体并购案诞生](#)

[小米为啥要造车？](#)

最新有关XSY芯师爷的文章

- [Tier1巨头集体“死磕”最后的堡垒「中国」](#)
- [格创东智AMHS业务正式启动，用AI加速半导体软硬融合](#)
- [武汉冲出一个超级半导体独角兽](#)
- [旋极星源持续拓展核心架构IP领域](#)
- [半导体专题篇十五：功率半导体](#)
- [闪存大会上几个好消息](#)
- [训练AI大模型，有多烧钱？](#)
- [英伟达“算力核弹”强在哪里？](#)
- [从学术讲到商业化，这场电子创新大会不容错过](#)
- [最新！半导体行业又一重磅收购](#)

<http://news.eeworld.com.cn/mp/XSY/a358690.jsp>