

三星专家

请问三星晶圆代工业务目前在中国市场的客户构成、订单规模以及工艺节点是怎样的？

三星的晶圆代工业务在中国没有设立工厂，所有产能均在韩国。目前，三星主要为中国的先进算力客户提供服务。这些客户在 5 纳米和 4 纳米工艺节点的月度订单总量合计约为 8,700 片晶圆。此外，三星也为非算力客户提供代工服务，例如飞腾的 CPU 和盛科通信的交换机芯片。

在 5 纳米工艺下，每片晶圆大约能切割出多少颗 GPU 芯片？

在 5 纳米工艺下，以目前 600 至 700 平方毫米的 die size 计算，每片晶圆平均能切割出约 45 颗 GPU 芯片。这个平均值考虑了多种因素。首先，理论计算基于晶圆面积和 die size，并计入约 90% 的可用面积（扣除边角料等损失）和约 70% 的工艺良率。按此计算，可得到约 60 至 70 颗 die。然而，实际产出中存在性能分级的情况。在这 60 至 70 颗 die 中，大约 70% 属于性能达标的 Bin1 产品，这部分会交付给客户。剩余约 30% 为 Bin2 产品，虽然性能稍弱，并非完全损坏的 bad die，但通常不作为正式产品出货。这些 Bin2 芯片，客户（如芯原、字节、阿里、燧原科技）可能会用于内部验证等工作。因此，综合考虑良率和性能分级后，最终交付给客户的平均数量约为 45 颗。

目前这些中国客户的订单是否已进入稳定量产阶段？

目前，在这些客户中，只有百度的昆仑芯属于大批量稳定下单，每月都有固定订单量。其他客户则采用分批下单的模式，其订单量级尚未完全释放，仍处于验证或逐步上量的阶段。

三星在 4 纳米和 5 纳米制程上还有多少产能余量可供国内企业使用？

目前三星 4 纳米制程的月产能为 3.7 万片，5 纳米为 5.8 万片。三星自身占据约 30% 的产能，其余由高通、AMD、GoOgle、特斯拉等客户分配。扣除这些以及国内现有客户的用量后，预计仍有约五六千片的产能空间可以分配给国内企业。尽管当前稼动率已达到 90% 左右，但仍存在一定的可利用空间。

目前三星 2 纳米和 3 纳米制程的进展如何，以及与 4 纳米、5 纳米在成本和应用上的差异是什么？

2 纳米制程的产能目前非常低，月产能仅为 6,000 余片。三星在该节点采用 GAA(Gate-All-Around)技术，早于台积电应用，但目前双方的 2 纳米良率均不高。当前市场广泛应用的主流制程仍是 5 纳米和 4 纳米，主要原因是性能成熟且成本更优。3 纳米和 2 纳米的晶圆价格过于昂贵，导致应用受限。

考虑到三星晶圆代工业务已实现扭亏为盈，明年(2027 年)在先进逻辑制程方面有怎样的扩产计划？

预计到 2027 年，三星的 4 纳米和 5 纳米制程将分别扩产约 2 万片/月的规模。此轮扩产的动因在于，公司晶圆代工业务在 2025 年左右实现扭亏为盈，且当前稼动率已达 90% 左右，有必要升级产能结构。此外，三星正在推行一种捆绑策略，向其存储产品的客户推广晶圆代工服务，提供一站式解决方案，预期这将带动其晶圆代工业务的增长。

三星 4 纳米和 5 纳米的扩产是否都采用 EUV 技术，以及每万片产能需要配置多少台 EUV 光刻机？

是的，4 纳米和 5 纳米的扩产肯定会采用 EUV 技术。根据工艺节点的需求，每新增 1 万片月产能大约需要配置 2 到 3 台 EUV 光刻机。这一配置是根据光刻、刻蚀、离子注入、

清洗、扩散、量测等一系列工艺流程的匹配需求来确定的，台积电也遵循类似的配置标准。

目前三星 EUV 光刻机的实际产能利用率是多少，以及 4 纳米和 5 纳米工艺分别需要多少道光刻工序？

当前 EUV 光刻机的实际产能利用率大约在 80%以上。在光刻工序方面，4 纳米工艺大约需要 14 道，而 5 纳米工艺则需要约 10 道。

在存储芯片方面，每万片 DRAM 产能需要配置多少台 EUV 光刻机，其工艺流程需要多少道光刻工序？

对于 DRAM，每万片月产能大约需要配置 2 台 EUV 光刻机。以 1a(1-alpha)工艺为例，其光刻工序大约为 7 道。尽管从工序数量上看似乎 1 台即可，但实际生产中为避免光刻环节成为瓶颈，并确保与后续刻蚀、离子注入等工序的顺畅匹配，通常需要配置 2 台。

目前三星传统的 DRAM 产能与 HBM 产能分别是多少，以及全球 HBM 市场的产能格局是怎样的？

目前三星的传统 DRAM 月产能约为 66 万片，这不包含 HBM。全球传统 DRAM 总月产能约为 186 万片，其中三星占 66 万片，海力士和美光各约 50 万片，长鑫存储约 20 万片。HBM 的产能是独立计算的，三星当前的 HBM 月产能约为 14 万片。全球 HBM 总月产能为 32.6 万片，其中三星占 14 万片，海力士为 15 万片，美光为 3.6 万片。

HBM3E 每一万片月产能需要配置多少台光刻机？

对于 HBM3E，每一万片的月产能大约需要配置两台左右的光刻机。其所需的光刻工艺与 DRAM 基本一致，因为 HBM 的底层逻辑晶圆(Base Die)同样是基于 DRAM 晶圆进行制造。

三星的 HBM 生产及封装工厂主要分布在哪些地点，其具体的工厂代码是什么？

三星的 HBM 生产工厂主要位于韩国的华城市和平泽市，具体的工厂代码为华城市的 H2 至 H4，以及平泽的 P1 和 P2。其封装业务则在天安工厂进行。

三星在 2026 年针对 DRAM 和 HBM 的扩产计划分别是多少？

2026 年，三星计划扩产 DRAM 约 6 万片月产能，同时扩产 HBM4 万片月产能。

国产半导体设备在刻蚀、薄膜沉积以及图形晶圆检测领域的进展如何？

在刻蚀设备领域，目前主要由华创和中微供应。在 PVD 和 CVD 等薄膜沉积设备方面，则以拓荆科技为主。在量测检测环节，对于有图案晶圆的检测，市场仍由 KLA 主导，特别是在明场和暗场检测设备领域，国产设备尚未取得显著进展。

当前存储新产品导入工作呈现出哪些新特点和挑战？

当前新产品导入工作的主要特点是速度要求越来越快，这源于产品更新迭代加速和生命周期缩短。同时，由于近期存储市场需求旺盛，订单量饱满，导致生产线上的设备资源紧张，为新产品导入借用设备带来了一定的困难。

近期存储市场的新产品需求主要集中在哪些领域，产能分配有何变化？

目前新产品的需求主要集中在服务器端，特别是用于 RDIMM 和 LRDIMM 的 DDR 内存产品。相比之下，移动消费电子领域的 LPDDR 和 UFS 产品需求保持稳定，甚至因价格过高出现需求反噬的迹象。因此，许多产能正从移动端转向服务器端 DDR 产品。这一趋势的背后，是 AWS、Google、Microsoft、Meta 等大型云服务商开始签订长期供货协议，导致服务器端产品的订单量持续处于高位。

目前与主要客户签订的长期供货协议有哪些具体条款，尤其是在定价机制方面？

当前签订的 LTA 订单，合同期限通常为两年或三年。与以往周期不同，本轮协议不仅锁定供应量，同时也锁定价格。客户需要预付约 40% 的款项。在定价方面，合同价格基于签订时所在季度的合约价，上浮 64% 至 70%。例如，若协议在 2026 年第一季度签订，则最终价格会在该季度合约价的基础上上涨 64% 至 70%。

目前 Google 的 V7、V8 以及 Trainium V3 服务器的 DRAM 配置情况如何？

当前 Google V7 服务器的 DRAM 配置主要看 RDIMM 插槽数量。以六个插槽的配置为例，每个插槽通常配置 24GB 或 32GB 的容量，其中 24GB 是大部分情况下的配置。对于 Amazon 的 Trainium 方案，其 DRAM 配套容量则相对更高，大约在 48GB 到 64GB 的水平。

关于近期 NVIDIA Rubin 平台计划将每个插槽的内存容量减半的传闻，其背后的原因是什么？

此举主要是为了节省成本，通过采用 LP (Low Power) 方案来平衡性能与成本。由于存储在整体成本中的占比较高，NVIDIA 选择在出厂时不预先配满内存，而是将插槽空间预留出来。这意味着终端客户需要根据自身需求自行采购并配置 SOCAMM 等内存模组，而非 NVIDIA 强制捆绑销售。

NVIDIA 与存储原厂关于 SOCAMM 和 HBM 的长协价格谈判进展如何？预计价格涨幅会是多少？

NVIDIA 的 SOCAMM 长协价格谈判预计要到 2026 年 6 月底才能有结果，价格涨幅预计至少在 50% 以上。对于 HBM，由于其当前价格已处于高位，预计价格涨幅相对温和，大约在 20% 至 30% 之间。

2026 年 HBM 市场的供需状况和主要厂商的扩产计划是怎样的？

2026 年 HBM 市场将持续处于紧缺状态。为应对需求，主要厂商均有扩产计划。在目前约 32.6 万片月产能的基础上，预计总共将扩产约 15 万片，产能增幅约为 47%。具体到各家，三星和海力士计划分别扩产 5 万片和 4 万片，美光计划扩产约 7 万片。尽管有如此规模的扩产，市场供应依然紧张，这也是推动价格上涨的原因，但其涨价幅度预计将小于 DRAM。

市场传闻三星希望 HBM 价格上涨 20%至 30%，该消息的来源是什么？这一涨幅的计算基准是否为 2025 年签订的年度长协价格？

该消息并非无端猜测，而是三星向客户释放的信号。涨价 20%至 30%的计算基准确实是基于 2025 年签订的年度长协价格。

Google 采用的 CXL 内存池化技术将如何影响对 HBM 和 DRAM 的需求？如果通过该方案减少 HBM 的使用，例如减少 96GB 的 HBM，需要增加多少 DRAM 来进行补偿？

CXL 内存池化技术(内存共享池)的概念已提出多年，目前开始进入实施阶段。该技术对内存接口芯片(如 Montage.Rambus、Renesas 的 MXC 芯片)的带动作用大于对存储芯片本身。虽然它会降低对存储的需求，但主要是起到池化作用，大幅降低需求的实际空间有限。减少 HBM 用量与增加 DRAM 用量之间没有固定的比例关系，具体取决于减少 HBM 的规模。若以具体数值估算，减少 96GB 的 HBM，大约需要增加 110GB 的传统 DRAM 来覆盖相应的缓存量。

用 110GB 的 DRAM 替代 96GB 的 HBM，在缓存速度上会慢很多，这种性能上的差异是否会影响 CXL 内存池化技术的广泛应用？

速度确实会变慢，这也是内存池化技术多年来未能广泛应用的原因之一。当前该技术开始步入正轨，正是因为业界开始着力解决速度问题，并沿着提升传输速度的思路进行技术演进。

Google 采用 CXL 内存池化方案的根本原因是什么?是因为无法获得足够的 HBM 供应，还是认为该方案本身在技术架构上更优越?

Google 采用该方案是基于技术架构的考量，即认为通过提升传输速度可以打通技术瓶颈，而非因为 HBM 的供应链管理问题。企业通常不会因为供应链受限而改变自身的核心架构思路。

从存储行业的周期来看，本轮价格上涨预计将持续到何时?

目前尚未看到价格见顶的迹象，价格仍在持续上涨。预计到 2027 年中旬之前，价格的绝对值将继续上行，但上涨的幅度可能会开始收敛。在当前的可视范围内，没有看到价格触顶的明确信号。

在三星当前约 66 万片的 DRAM 月产能中，数据中心客户的占比是多少?其中，不同制程工艺(如 1c、1b、D1A、D1Z)的产能分布情况如何?

在三星 66 万片的 DRAM 月产能中，数据中心客户的占比约为 37%。从制程工艺来看，1c(1c) 工艺的产能全部供应给服务器市场。在整体产能中，1b (1b)工艺约占 5%,D1A(1-alpha)工艺约占 20%，剩余部分为 D1Z(1z) 工艺。服务器客户使用的产品基本不涉及 D1A 和 D1Z 工艺，主要集中在 1c 和 1b 工艺，两者大致各占一半。

在当前价格上涨的背景下，服务器客户是否会考虑使用成本更低的 1b 工艺替代 1c 工艺以控制成本?

目前市场上几乎没有客户采取这种做法。客户普遍能够接受当前的价格水平，不会为了成本而牺牲产品性能。

2026 年三星计划扩产的 8 万片 DRAM 产能，是否全部用于服务器市场？

是的，2026 年计划扩产的 8 万片 DRAM 产能将全部用于服务器市场。

长鑫存储目前的产能规模和技术进展如何？

长鑫存储当前的月产能接近 20 万片，约为 19.8 万片。在技术方面，其最高制程已达到 D1A 级别，即 14 纳米左右的工艺。作为对比，其在 2025 年时主要生产 15.8 纳米的 D1Z 产品。由于无法获得 EUV 光刻机，其技术发展受到了限制。

在 DRAM 制造工艺方面，使用 DUV 光刻机生产 1b 制程的技术可行性与良率表现如何？

目前来看，实现 1b 制程存在一定挑战。行业内其他几家厂商主要采用 EUV 光刻机进行生产。虽然理论上也可以使用 DUV 光刻机，但这会导致显著的良率损失，从而使得生产的性价比变得非常低。

关于 HBM 的工艺进展，目前能做到哪个节点？1A 和 1Z 节点分别对应 HBM 的哪一代产品？

210

目前 HBM 产品还处在开发阶段，当前仅能生产 2E 产品，该产品对应的是 1Z(15.8 纳米的迭代)工艺节点。正在开发的 HBM3 产品则对应 1A 工艺节点，由于 1A 制程尚未使用 EUV 光刻，因此开发上没有根本性障碍。然而，若要进一步发展到 HBM3E，则必须依赖 EUV 光刻技术，否则无法实现生产。

从技术原理上看，为何逻辑芯片可以采用多重曝光(mult-patterning)技术，而 DRAM 则难以应用该技术来替代 EUV？

两者的制造逻辑存在根本差异。逻辑芯片关注的是金属栅极的宽度，而 DRAM 存储单元则关注金属氧化物薄膜的厚度。这种在基础制造工艺上的不同，导致了多重曝光技术在 DRAM 上不适用。DRAM 的厚度较大，在后续生产 HBM 时还需要进行专门的减薄处理。在技术上，是否可以用多颗 1a 级别的芯片来替代 1b 或 1c 级别的芯片，以应对高端芯片供应不足的问题？

目前没有这样的技术方案。采用性能较弱的 1a 芯片替代 1b 或 1c 芯片，必然会导致整体性能的折损，无法满足原有的应用场景需求。业内不存在通过降级使用来弥补高端产品空缺的做法。

目前已签订的 DRAM 长期供货协议(长协)占总产能的比例是多少?从产能和金额两个维度看，其具体情况如何？

从产能口径看，目前月产能为 60 多万片，其中已签订的长协量约占总产能的 40%。未来预期长协的占比能达到 60%以上。从金额口径看，由于长协主要覆盖价格更高的服务器级产品，而剩余部分主要是价格较低的消费级产品(如 LPDDR、UFS、eMMC 等)，因此长协在总销售额中的占比会更高。

2026 年和 2027 年的 DRAM 总产量及增速预计是多少?已签订的长期供货协议是否会覆盖未来的增量产能？

2026 年的产能增量为 8 万片/月，2027 年预计增加 4-5 万片/月，两年总计增加约 13 万片/月。已签订的长期供货协议是基于一个固定的量，不会自动覆盖未来新增的产能部分。

DRAM 产品结构将如何调整?2026 年新增的 8 万片和 2027 年新增的 5 万片产能是否全部为 1c?是否还包含由 1b、1a 或 D1Z 转产的 1c 产能？

新增产能确实全部是 1c。此外，每年还会有约 1-2 万片规模的 1a 和 1b 产能通过迭代升级转向 1c，这部分也需要纳入考量。

考虑到海力士、美光、三星这三家主要厂商的产能调整逻辑相似，是否意味着 1b、1a 和 D1Z 的总产能每年都在下降，而长鑫存储的扩产主要是在抢占这部分市场？

是的，这三家厂商的 1b、1a 和 D1Z 产能确实在逐年下降。长鑫存储的扩产主要针对的就是 1a 和 D1Z 这两个市场。

如果长鑫存储通过快速扩产，导致 D1Z 和 1a 市场供过于求、价格大幅下降，这对 1b、1c 和 HBM 等高端产品的价格会产生什么影响？市场是否会因此分层，价差拉大？

对 1b、1c 和 HBM 的价格基本没有影响。由于 D1Z 和 1a 产品无法反向替代更高规格的服务器端芯片，即便其价格因扩产而降低，也不会传导至高端市场。因此，市场会进一步分层，不同等级产品间的价差将会拉大。

考虑到三星等厂商曾公开表示到 2030 年产能要翻倍，而当前规划的每年 8 万片、5 万片的扩产规模似乎无法满足需求，应如何看待未来的产能规划？

所谓的“产能翻倍”更多是宣传口号。实际的扩产决策需要综合考虑海力士、美光等所有主要厂商的行动，市场并非由一家独占。如果三家厂商合计增加二三十万片产能，将对市场造成巨大冲击。因此，实际的扩产规模不会达到宣传的水平。

从晶圆产出的角度看，1a、1b 和 1c 三种制程在单片晶圆上可切割的颗粒数量、以及单个颗粒的容量分别是多少？

在单片晶圆产出方面：

1a 制程可切割约 810 个颗粒，单个颗粒容量约 3.2GB。

1b 制程可切割约 800 个颗粒，单个颗粒容量在 3.8GB 到 3.9GB 之间。

1c 制程可切割约 790 个颗粒，单个颗粒容量约 4.7GB。

从 1a 到 1c，为何纳米制程进步的同时，单片晶圆可切割的颗粒数量反而略有减少？容量提升为何是比切割数量更关键的参数？

从 1a 到 1b 再到 1c，虽然纳米制程进步会提升存储密度，但由于单颗粒容量持续增大，导致 die size 也相应变大，因此单片晶圆可切割出的颗粒数量会略微减少。容量是更关键的参数，因为它与纳米制程的进步强相关，代表了存储密度的核心提升。虽然切割数量变化不大，但容量的增长幅度远大于切割数量的降幅。

NAND 业务未来几年是否有扩产计划？技术迭代将如何进行？

目前来看，NAND 业务在 2026 至 2029 年间没有扩产计划。正因为不进行产能扩张，所以技术迭代，即增加堆叠层数，是必须推进的方向。技术会持续向 V8、v9、V10 升级。

在不扩产的背景下，NAND 业务中负责堆叠层数技术研发的部门是否会变得边缘化？

不会。提高堆叠层数的技术难度很大，例如需要从过去的 COP 技术转向类似长江存储的 Xtacking 技术，这需要持续的研发投入。此外，尽管 NAND 的单位 GB 价格低于 DRAM，但其总容量基数庞大，因此在业务中依然占据重要地位。

请确认三星存储业务当前的月产能规划以及未来两年的扩产计划？

目前三星存储的月产能为 60 万片。近两年的计划是维持这一产能水平，同时进行研发投入以提升堆叠层数，并维持必要的资本开支。未来的扩产计划会根据市场情况进行调整，但近两年内将保持在每月 60 万片的水平。

从产品线能力角度看，长江存储与三星相比如何？在先进封装技术方面，两家公司是否存在差距？

从技术角度来看，长江存储的产品线能力确实更强，但这并不意味着其最终产品的性能必然优于三星，部分原因可能与设备国产化有关。在先进封装领域，存储芯片方面的差距不大，例如 HBM 或传统 DRAM 的先进封装技术差异有限，因为封装在整个半导体制造流程中相对简单。然而，在逻辑芯片封装技术上，如 CoWoS 这类技术，则存在明显差距。

对于 HBM4 和 HBM5 这类需要采用 Hybrid Bonding 技术的产品，其封装流程是如何分工的？

存储厂商负责利用 Hybrid Bonding(混合键合)技术完成存储颗粒的堆叠，形成完整的 HBM 产品。之后，这些堆叠好的 HBM 产品会交付给台积电等晶圆代工厂，由他们进行后续的 CoWoS 封装。

在 HBM 技术路线中，是否会为了简化工艺而放弃部分容量，以及未来的主流方案是什么？

确实存在不采用 bonding 技术、牺牲部分容量的方案。不过，未来的主流趋势是通过 bonding 技术将 die 做得更薄，从而在有限空间内堆叠更多层，例如塞入 12 颗或 16 颗 die，以实现更高的容量和性能。

请详细说明三星在逻辑、DRAM 及 HBM 领域的扩产计划，特别是 HBM4 的具体规格？

逻辑芯片方面，将继续扩产 4 纳米和 5 纳米制程，主要为了满足中国客户的需求，并作为台积电之外的备选供应商服务于美国客户。DRAM 和 HBM 的扩产将集中于 1c 制程和 HBM 产品。从 2026 年开始，扩产的 HBM 主要是 HBM4。具体到规格，扩产的 HBM4 将采用 16-H 的堆叠方式，即堆叠 16 颗 DRAM die。考虑到冗余，每颗 die 的有效容量约为 4GB，因此一个 16-H 的 HBM4 堆叠容量约为 64GB。以英伟达的方案为例，若配置 6 颗这样的 HBM4，总显存容量将达到 384GB。

2026 年 1c 制程的产能增速为何会超过 HBM?这一产能安排背后的需求驱动因素是什么?

2026 年 1c 制程的产能将从 2025 年的 12 万片/月增加 8 万片至 20 万片/月, 若考虑转产因素, 增速将超过 80%。1c 制程增速快于 HBM, 主要是因为其需求最为强劲。除了用于 HBM, 1c 作为先进制程, 还广泛应用于其他传统 DRAM 领域, 例如 SOCAMM 等产品。此外, 先进制程产能也可以灵活转用于生产成熟制程产品, 具备更高的灵活性。

1c 制程在主流 AI 芯片和服务器 CPU 中的应用情况是怎样的?

1c 制程的应用非常广泛。在 AI 芯片领域, 英伟达从 H100 开始就采用了 1c 制程。Google 的 TPU V7 和 V8 也使用 1c 制程, 而其早期的 V6 则使用 1a 和 1b 制程。亚马逊的 Trainium 和 Inferentia 体系目前也正全面转向 1c 制程。在服务器 CPU 方面, 配套的内存插槽同时支持 1b 和 1c 制程的 DRAM, AMD 和 Intel 的至强平台均采用这种混合配置。