

## 电子行业深度报告

# 先进封装赋能 AI 计算，国内龙头加速布局

增持（维持）

2024 年 03 月 06 日

证券分析师 马天翼

执业证书：S0600522090001

maty@dwzq.com.cn

证券分析师 周高鼎

执业证书：S0600523030003

zhougd@dwzq.com.cn

### 投资要点

- **先进封装本质目的是增加触点连接，以代替制程提升。**量子隧穿效应导致先进制程的研发制造成本过高，而良率过低，先进封装技术能够弥补制程提升的困难。先进封装技术的本质为提升连接效率。其中，重布线层技术（RDL）重新布局裸片 I/O 触点，支持更多、更密引脚，广泛用于晶圆级封装（WLP）；硅通孔技术（TSV）通过将芯片的焊点打穿，在通孔里填充金属材料实现芯片与芯片、芯片与基板的垂直连接，是 2.5D 和 3D 封装的关键解决方案；凸块技术使用凸点（bump）代替传统引线，增加触点、缩小传输距离和电阻；混合键合技术（Hybrid Bonding）通过将芯片或晶圆平面上的铜触点抛光后进行退火处理，使得连接平面完全贴合，以无凸点（Bumpless）的方式缩减连接距离和散热能力。先进封装对制造设备精度、无尘环境、测试精度要求极高。技术升级方向为增加连接效率（如使用玻璃基板代替有机基板）和降低成本（如使用“硅桥”代替硅中介层）。
- **先进封装赋能高速计算，算力需求提升，先进封装产能供不应求。**先进封装主要通过两方面提升逻辑芯片的算力：一、提升处理器集成度，从而提升性能；二、提升处理器和存储器间的连接带宽、减小连接功耗，从而解决“内存墙”和“功耗墙”，提升芯片算力。随着 AI 大语言模型市场的发展，模型训练和推理应用所需算力不断提升；国内新入局 AI 企业众多，智算芯片需求旺盛。根据 IDC，至 2026 年，国内智算规模可达 2023 年的 3 倍。与此同时，供给端高性能 GPU 产能明显不足，先进封装产能成为主要瓶颈。2023 年 8 月，英伟达表示计划 2024 年将 H100 产能拉高至少 3 倍。2023 年 9 月，台积电表示 CoWoS 产能只能尽量满足客户 80% 的需求。先进封装发展前景、国产替代空间广阔。
- **先进封装行业壁垒高，专业封测厂商不具优势；海外龙头加速扩产，国内企业追赶。**先进封装行业壁垒高，且相比 OSAT 厂，Fab 厂和 IDM 厂更具优势，主要原因有二：第一，技术精度高，且高度依赖晶圆制造技术、与芯片设计环节的协同，例如重布线层(RDL)、硅通孔(TSV)、混合键合(HB)需要在裸晶本体上进行线路设计、刻蚀、电镀，晶圆厂在技术和硬件方面更有优势；第二，晶圆厂主导了先进封装领域的技术路线和订单分配，封装厂需要与上游厂商密切合作以获取订单。面对高增需求，海外龙头加大扩产力度，但扩产难度大、周期长。台积电、三星、英特尔、日月光纷纷增加先进封装产线，但由于上游设备供应不足等原因，扩产周期普遍达 2-3 年。与此同时，国内龙头积极布局先进封装领域。长电科技聚焦 XDFOI 新技术、2.5D/3D 技术的量产；通富微电聚焦消化高端 CPU、GPU 封装产能，现已涉及 AMD MI300 的封装；甬矽电子积极研发 Fan-in/Fan-out、2.5/3D 晶圆级封装相关技术，并大力建厂扩产，未来营收增长空间广阔。
- **风险提示：**AI 算力需求增长不及预期；先进封装技术进展缓慢；国产替代不及预期。

### 行业走势



### 相关研究

《AI 拉动景气度向上叠加业绩拐点，存储板块成长动能充足》

2024-03-05

《整车制造降价提速，一体化嵌塑集成蓄势待发》

2024-02-20

## 内容目录

1. 先进封装的基本逻辑是增加触点连接，解决摩尔上限 .....	5
1.1. 纳米制程因量子隧穿效应及高成本低良率，提升困难 .....	5
1.2. 封装技术的迭代规律提升本质是提高连接效率 .....	7
1.2.1. 技术难点主要在于精度不足 .....	13
1.2.2. 升级逻辑为增加连接效率、降低制造成本 .....	14
2. 算力需求提升导致先进封装产能供不应求 .....	15
2.1. 逻辑芯片为主要需求点，先进封装赋能高速计算 .....	15
2.2. 算力供不应求，拉动先进封装需求增长 .....	17
3. 国内外封装厂企业加速扩产 .....	21
3.1. 技术及订单承接方面晶圆厂具有较强优势 .....	21
3.2. 海外厂商积极扩产，扩产周期约 2-3 年 .....	23
3.3. 国内公司加速布局先进封装，关注长电、通富、甬矽等 .....	24
4. 风险提示 .....	26

## 图表目录

图 1: 台积电芯片制程发明年份.....	5
图 2: MOSFET 晶体管横截面图.....	6
图 3: 量子隧穿效应示意图.....	6
图 4: 各制程芯片设计成本预测.....	6
图 5: ASML 历年 EUV 光刻机单价.....	6
图 6: 将 Chiplet 紧密连接以替代整块芯片的先进封装技术.....	7
图 7: 封装技术发展历程.....	7
图 8: RDL 技术示意图.....	8
图 9: 传统封装及晶圆级封装.....	9
图 10: 扇入型及扇外型晶圆级封装.....	9
图 11: InFO 技术示意图.....	9
图 12: TSV 技术示意图.....	10
图 13: 平面互连与垂直互连对比图.....	10
图 14: 引线互连与 TSV 对比图.....	10
图 15: 台积电 CoWoS 结构示意图.....	11
图 16: 各键合技术的触点密度.....	11
图 17: 热压键合技术流程示意图.....	11
图 18: 混合键合技术流程示意图.....	12
图 19: Wafer-to-WaferHybridBonding 技术示意图.....	12
图 20: 各凸点技术信号传输流失程度与频率的关系.....	12
图 21: 代表厂商混合键合技术应用及性能提升.....	13
图 22: 封装失效分析的挑战.....	14
图 23: 玻璃封装基板的优势.....	14
图 24: EMIB 技术示意图.....	15
图 25: 采用台积电 LSI 技术的 AppleM1Ultra 芯片.....	15
图 26: AppleM1Ultra 带宽与 16 核 CPU 性能对比图.....	16
图 27: AppleM1Ultra 带宽与 16 核 CPU 功耗对比图.....	16
图 28: 冯·诺伊曼架构示意图.....	16
图 29: HBM 和 GDDR5 对比示意图.....	17
图 30: HBM 和 GDDR5 性能对比.....	17
图 31: Transformer 原始模型结构.....	17
图 32: 参数量对语言模型(LM)的性能有积极影响.....	17
图 33: 部分大语言模型参数量.....	18
图 34: 部分大语言模型数据集规模及来源.....	18
图 35: 2023 年数据集规模排名.....	18
图 36: 训练各 Transformer 架构的模型所需算力增速远超摩尔定律.....	19
图 37: 2023 年国内 AIGC 算力产业全景图.....	20
图 38: 中国智能算力发展情况及预测.....	20
图 39: 2023 年英伟达 H100 主要客户购买量估测.....	20
图 40: 先进封装市场份额.....	21
图 41: 主流封装设备国产化率及预测.....	21
图 42: 2019-2029 先进封装触点间距预测.....	22

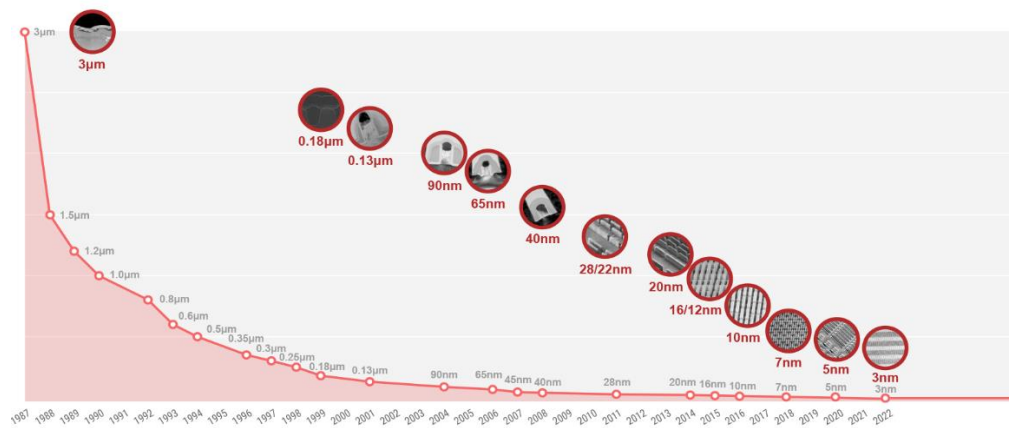
图 43: 台积电 3DFabric 及前段后段封装技术 .....	22
图 44: 英特尔、三星、台积电先进封装技术布局 .....	22
图 45: IDM 及 Foundry 厂商进入先进封装环节 .....	23
表 1: 海外先进封装厂商龙头扩产规划 .....	24
表 2: 国内先进封装厂商主要产品及产业布局 .....	25

## 1. 先进封装的基本逻辑是增加触点连接，解决摩尔上限

### 1.1. 纳米制程因量子隧穿效应及低成本低良率，提升困难

半导体制程工艺因量子隧穿效应及低成本低良率，难以继续提升。半导体及芯片发明以来，主流的发展方向是对摩尔定律的延伸，不断缩小晶体管的制程。缩小制程能够缩小芯片尺寸、提升芯片承载晶体管数，从而提升芯片算力、速度及性能、减小功耗、降低成本。随着制程工艺进入纳米级别，制程的提升越发艰难。主要的阻碍来自两方面。

图1：台积电芯片制程发明年份

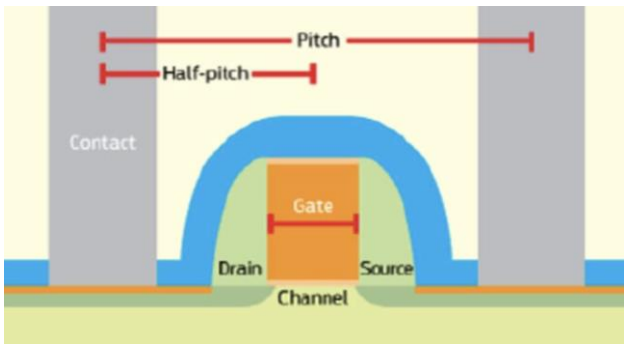


数据来源：公司官网，东吴证券研究所

**量子隧穿效应（短沟道效应的一种）使得晶体管漏电、芯片发热，导致芯片性能下降、功耗增加。**晶体管的工作原理是通过栅极(Gate)施加电压，控制沟道中的自由电荷，以控制从源极(Source)到漏极(Drain)的电流通断，从而传递出0或1的电信号。当栅极两端电压为0时，源极和漏极之间没有自由电荷，无法通电；当栅极两端存在方向正确的电压时，源极和漏极间便存在被吸引来的自由电荷，电路导通。制程，即为栅极或沟道的等效宽度。当缩小沟道宽度至一定数值时，即使栅极两端电压为0，由于微观电荷的随机运动，源极和漏极间依旧有电流流过，栅极不再能够控制电流的通断，晶体管性能大打折扣，因“漏电”产生的热量亦增加了功耗。虽然该技术难题已在部分实验室利用碳化硅等新材料取得小规模突破，但尚未发展至可商业化的程度。

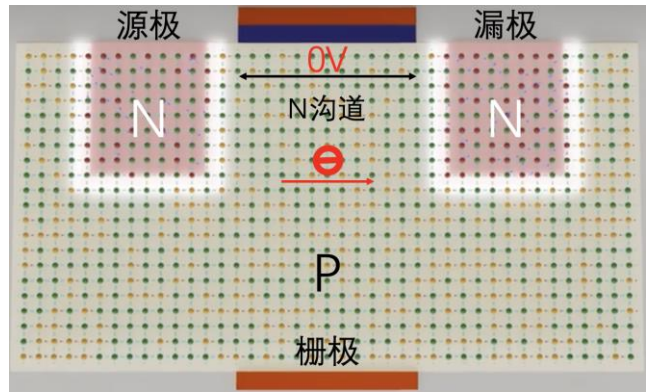


图2: MOSFET 晶体管横截面图



数据来源: 3DInCites, 东吴证券研究所

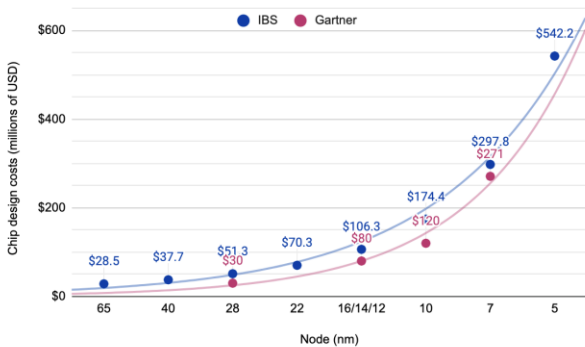
图3: 量子隧穿效应示意图



数据来源: 哔哩哔哩, 东吴证券研究所

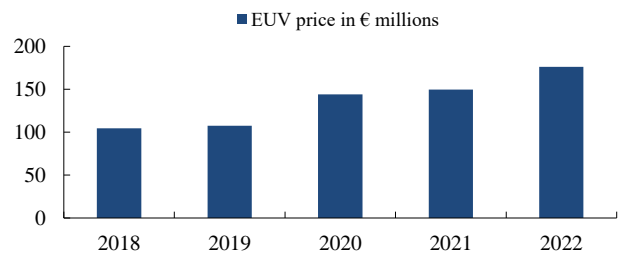
先进制程芯片的研发和制造成本高居不下，良率却越来越低。根据 IBS 和 Gartner 的预测，自进入先进制程时代以来，芯片的设计成本不断飙升，5nm 的总设计成本高达约 5 亿美元。EUV 光刻机、掩膜等价格随技术要求提升也不断增加，拉高了芯片代工成本。与此同时，小制程芯片的良率却越来越低。2023 年，有消息称台积电的 3nm 良率在 55% 左右，腰斩的良率也使苹果为其 iPhone15Pro 中内置的 A17 处理器芯片谈下了更便宜的价格。根据韩国媒体 Chosunbiz 的消息，三星和台积电的 3nm 半导体良率均难以超过 60%。低良率显然增加了芯片的制造成本和销售压力。基于此，主流厂商转而企图利用先进的封装技术，以期在不牺牲小制程芯片高性能、小体积、低功耗的基础上降低成本，弥补先进制程前进的困难。

图4: 各制程芯片设计成本预测



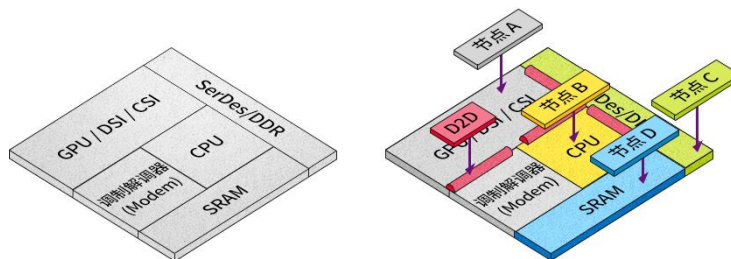
数据来源: AI Chips: What They Are and Why They Matter, 东吴证券研究所

图5: ASML 历年 EUV 光刻机单价



数据来源: ASML 官网, 东吴证券研究所

图6: 将 Chiplet 紧密连接以替代整块芯片的先进封装技术



如果工艺节点是必不可少的，则应使用A、B、C和D等节点。

**单块芯片 (Monolithic Chip)**

单一芯片

**芯粒 (Chiplet)**

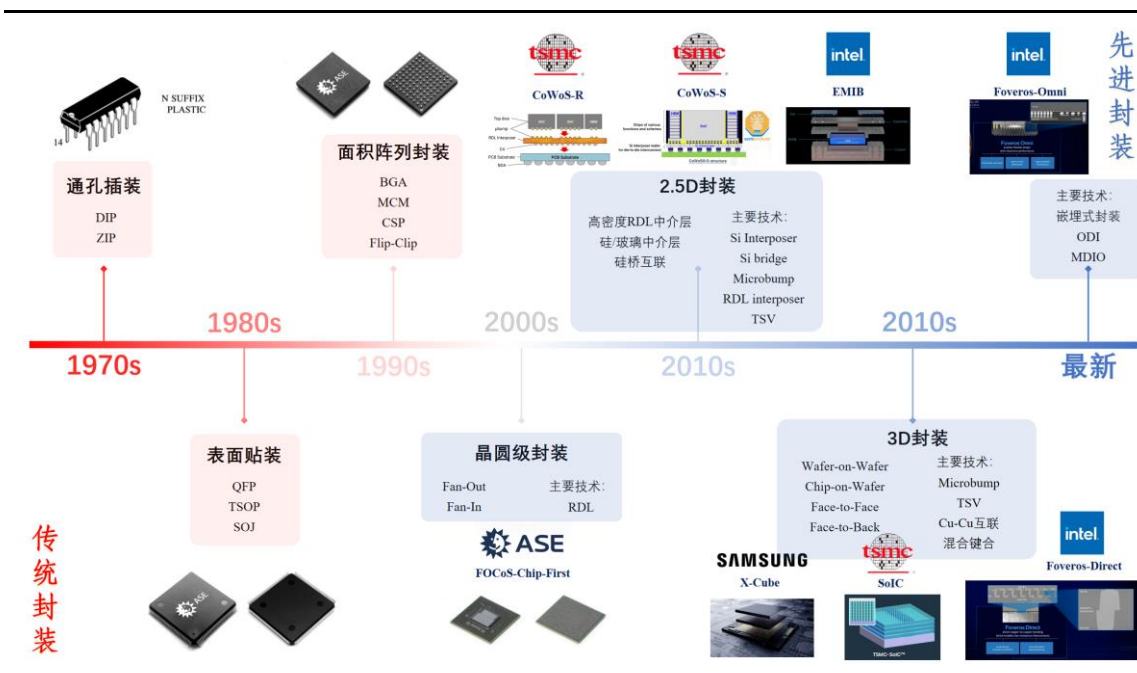
一种按照功能对现有逻辑芯片进行拆分并使用硅穿孔堆叠实现互连的技术

数据来源: SKhynixNEWSROOM, 东吴证券研究所

1.2. 封装技术的迭代规律提升本质是提高连接效率

封装技术的迭代过程实际是连接效率不断提高的过程。传统封装时代，封装采用通孔插装、表面贴装方式，引脚密度小、传输距离长、带宽小、电阻大，传输效率低下。20世纪90年代，焊球连接技术取代引线，增加了接触面积和引脚密度，减小了传输距离和电阻，芯片尺寸因此得到缩减。2000年以后，正式进入先进封装时代。晶圆级封装缩小了芯片尺寸，RDL、中介层、TSV等技术提升了芯片的堆栈密度和性能，Microbump和混合键合技术进一步缩小了触点间距和堆栈高度。封装技术的迭代增加了引脚密度和带宽、缩小了传输距离和电阻，实际上提高了连接效率。

图7: 封装技术发展历程



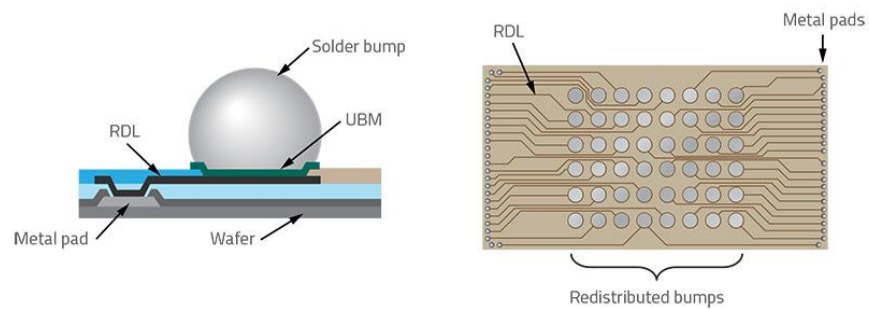
数据来源: 各公司官网, 东吴证券研究所制图

先进封装技术的本质目的为**进一步增加连接效率**。为实现小制程芯片通过将小体积晶体管高密度排布获得的多功能、大算力，先进封装技术聚焦增加芯片间的连接效率(减小信号延迟、功耗，提升传输速率)，提升连接的紧密度。增加连接效率，一般通过缩小触点距离以增加触点密度，以及缩小芯片与芯片、芯片与基板间的距离两方面实现。

先进封装技术主要有：

**重布线层技术(RDL)**。芯片的 I/O 触点通常分布在边缘或四周，直接进行封装会因缺少引线或引线过于密集而导致连接受限。RDL 技术能够将裸片的触点重新布局到空间较为宽松的芯片中间，并使得接口处凸点面积更大、数量更多。当下的 RDL 技术能够将线距缩小至 1-10 $\mu\text{m}$  的范围。

图8: RDL 技术示意图



数据来源：LamResearch，东吴证券研究所

**RDL 技术使芯片在封装后支持更多的引脚，以增加芯片的算力、芯片间的连接。**该优势广泛体现在晶圆级封装(WaferLevelPackage)中。晶圆级封装主要分为扇入型晶圆级封装(Fan-inWLP)和扇出型晶圆级封装(Fan-outWLP)，扇入型晶圆级封装利用 RDL 在芯片原有区域增加了触点，扇出型晶圆级封装则使用环氧塑封材料适当拓展芯片面积，同时利用 RDL 进行触点的二维延伸。



图9：传统封装及晶圆级封装

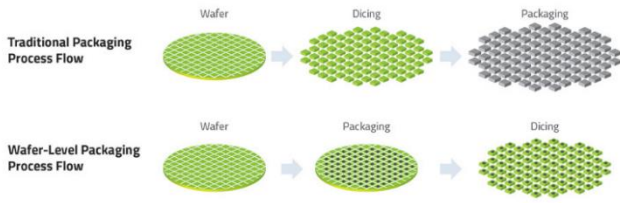


图10：扇入型及扇外型晶圆级封装

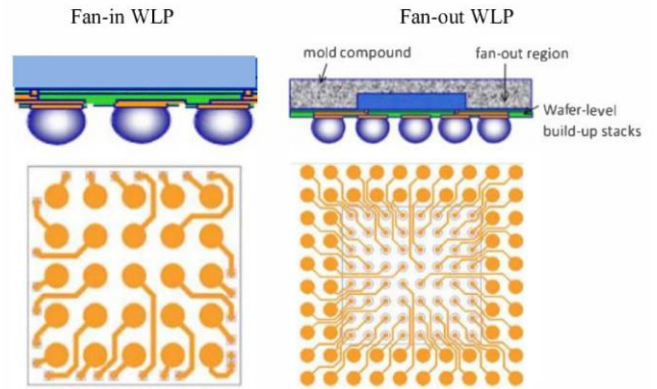


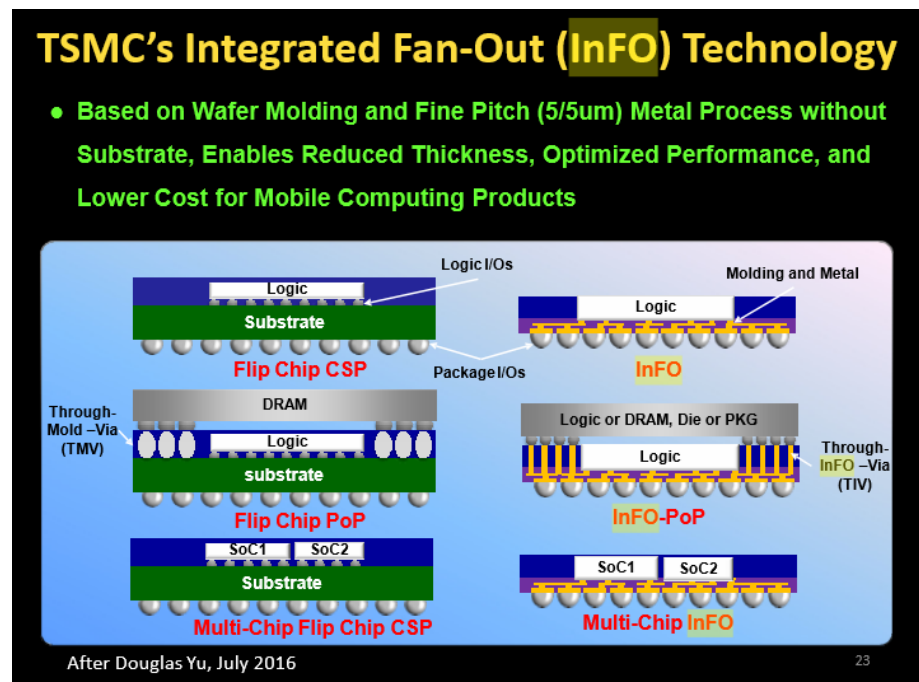
Fig. 1. Fan-in WLP and Fan-out WLP

数据来源：SEMICONDUCTOR ENGINEERING，东吴证券研究所

数据来源：Investigation on wafer warpage evolution and wafer asymmetrical deformation in fan-out wafer level packaging processes，东吴证券研究所

**RDL 技术能够代替中介层，从而缩小连接距离，提升传输速率。**该技术能够在垂直堆叠封装时直接连接芯片和基板，为封装系统缩小减薄，提高集成度。台积电的 InFO(Integrated Fan-out)系列封装技术即体现了该优势。与传统的垂直堆叠先进封装技术（如 PoP 等）不同，InFO 没有使用硅中介层，而是在最底层逻辑芯片上进行了扇出塑封，并利用 RDL 技术在塑封区域布局上下连通的电路，以连接上层芯片和基板。该连接方式被称为 TIV(Through-InFO-Via)。InFO 首用于 iPhone7，并助力台积电收获苹果 A10 芯片的全部订单。

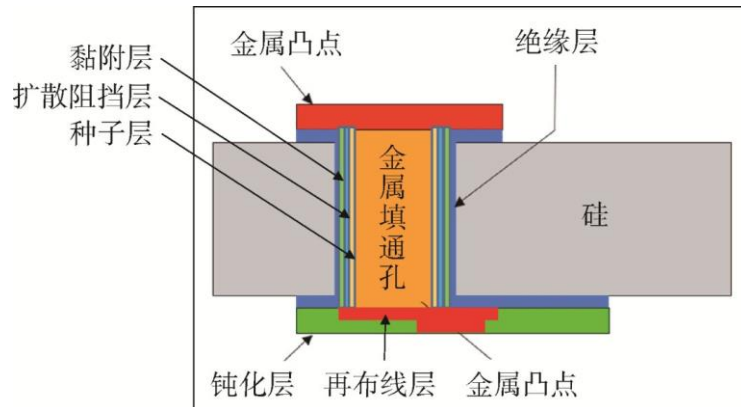
图11：InFO 技术示意图



数据来源：EE|Times，东吴证券研究所

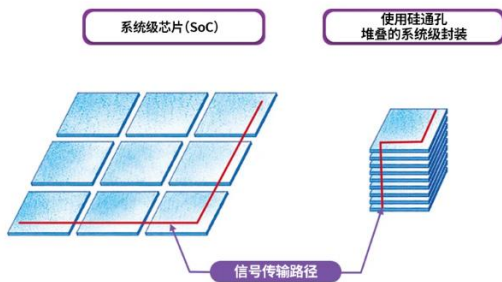
**硅通孔技术(TSV)**。为了缩小传输距离，人们使用堆叠芯片的方式进行封装。硅通孔技术通过将芯片的焊点打穿，并在通孔里填充金属材料(主要为铜)，使芯片与芯片、芯片与基板实现垂直互连。比起传统的平铺芯片或者引线互连堆叠芯片，利用 TSV 的先进封装能够大幅缩小连接距离、提升连接效率。

图12: TSV 技术示意图



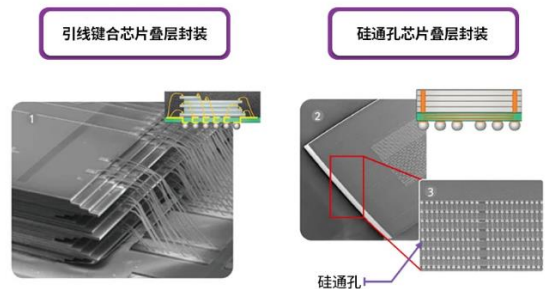
数据来源:《前瞻科技》, 东吴证券研究所

图13: 平面互连与垂直互连对比图



数据来源: SKhynixNEWSROOM, 东吴证券研究所

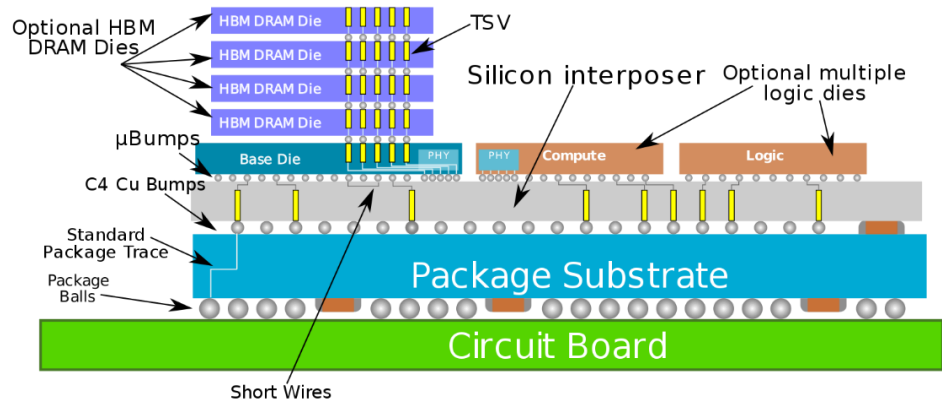
图14: 引线互连与 TSV 对比图



数据来源: SKhynixNEWSROOM, 东吴证券研究所

**硅通孔技术是实现 2.5D 及 3D 封装的关键解决方案**。台积电的 CoWoS 封装中采用了大量 TSV 技术，其传输的高速和可靠性使之成为了 AI (如英伟达 A100、H100, AMDMI300) 等高性能芯片的主流选择。

图15: 台积电 CoWoS 结构示意图

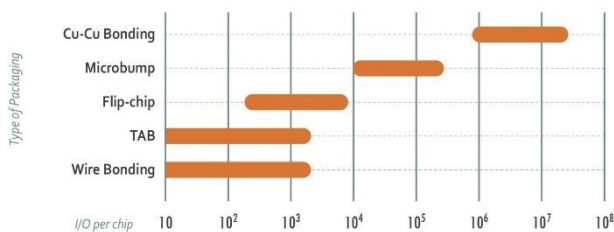


数据来源: WikiChip, 东吴证券研究所

**凸块技术(bumping)。**该技术使用凸点(bump)代替传统引线，能够增加 I/O 触点密度，缩短传输距离。不同于要求焊盘分布于芯片四周的引线键合技术，面分布的凸点阵列允许 I/O 触点分布于芯片中间，大幅提高空间利用率和触点密度；利用倒装技术(Flip-Clip)和凸点垂直连接各芯片，也比引线键合的电路距离更短。

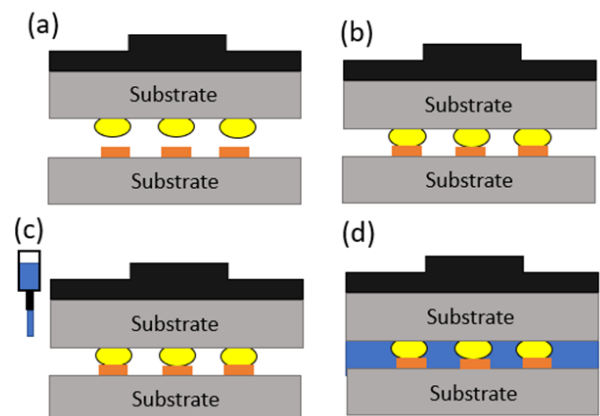
凸块技术主要分为球栅阵列焊球(Ball-Grid-Array Solder Ball, BGA ball, 直径 0.25-0.76mm)；倒装凸点(Flip-Chip Solder Bump, FC Bump)，也被称为可控塌陷芯片焊点(Controlled Collapse Chip Connections solder joint, C4 solder joint, 直径 100-150 $\mu$ m)；微凸点(microbump, 直径可小至 2 $\mu$ m)。连接凸点时通常利用热压键合技术(Thermal Compressive Bonding)融化焊球并使之冷却融合，并填入底部填充剂提高芯片机械性质。如今，微凸块的直径和间距仍在不断缩小。

图16: 各键合技术的触点密度



数据来源: 闾康科技, 东吴证券研究所

图17: 热压键合技术流程示意图



数据来源: 闾康科技, 东吴证券研究所

**混合键合技术(Hybrid Bonding)**能够解决接点间距(Pitch)缩小时出现的问题，进一步提升接点密度、提升连接效率。当接点间距微缩至 10 微米左右时，焊锡球尺寸过小，

容易在加热熔化过程中完全反应变质，降低导电性能；植球回流过程中两相邻焊锡球容易碰触在一起，导致芯片失效。混合键合技术通过将芯片或晶圆平面上抛光后凹陷的CuBump进行退火处理，使得Cu略微膨胀，两平面完全贴合，以无凸点(Bumpless)的方式缩减连接距离、提升接点密度、散热能力、信号传输准确度，从而降低能耗、提升效率。相比微凸点，混合键合技术能使I/O引脚密度增加5-10倍。当下，混合键合技术主要用于晶圆级封装，在晶圆制造环节即设计铜触点连接两片晶圆，切割后成为一体化的封装模块。

图18：混合键合技术流程示意图

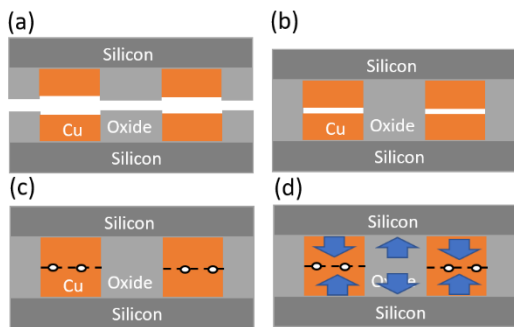
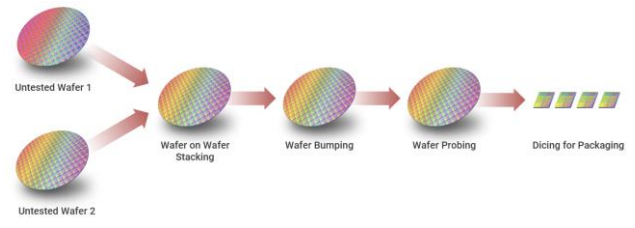


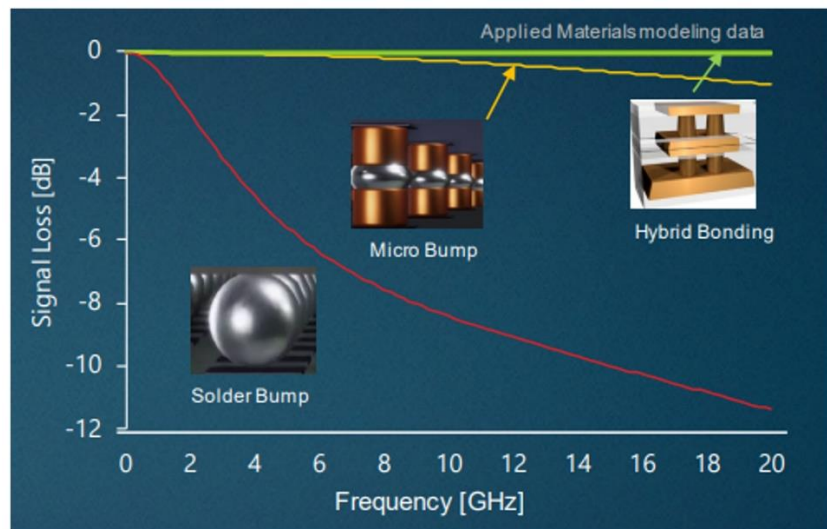
图19：Wafer-to-Wafer Hybrid Bonding 技术示意图



数据来源：闵康科技，东吴证券研究所

数据来源：台积电官网，东吴证券研究所

图20：各凸点技术信号传输流失程度与频率的关系

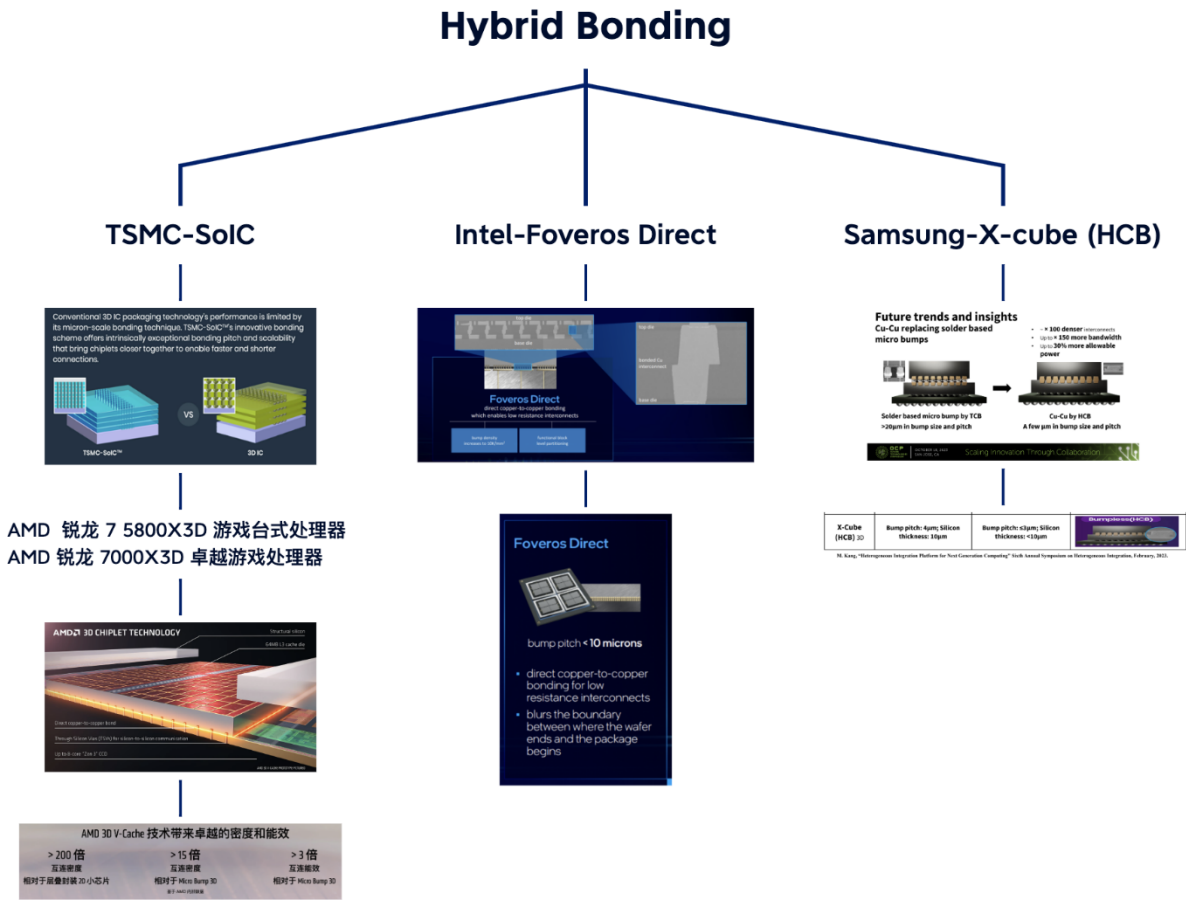


数据来源：SEMICONDUCTORENGINEERING，东吴证券研究所

台积电、三星、英特尔领衔发展混合键合技术。当前，台积电的SoIC技术、三星的X-Cube技术、英特尔的FoverosDirect技术均运用了铜对铜直接键合的方式。使用SoIC的AMD锐龙75800X3D游戏台式处理器和锐龙7000X3D卓越游戏处理器率先实现量产。



图21：代表厂商混合键合技术应用及性能提升



数据来源：公司官网，OCP，WikiChipFuse，东吴证券研究所制图

### 1.2.1. 技术难点主要在于精度不足

随着 I/O 触点密度的不断提升，先进封装对技术精密度的需求也不断提高。以 1.2 中三项核心技术为例，RDL 重布线层中电线距离、TSV 通孔距离和粗细、凸点直径和间距缩小，要求设计制造技术和设备的不断精进。以混合键合技术为例，由于触点平面需要完全接触，且触点距离极小（一般在 10 微米内），在设计环节即需要保证两平面触点的严格对应（误差不得超过 1 微米）。在制造环节，在铜材料沉积至裸片后，对其进行化学机械抛光(CMP)需要保证铜表面非常光滑并适当凹陷，以保证平面在加热熔化后绝对贴合。在贴装环节，布满细密铜触点的裸片必须精确地瞄准，且不能掺杂灰尘颗粒。

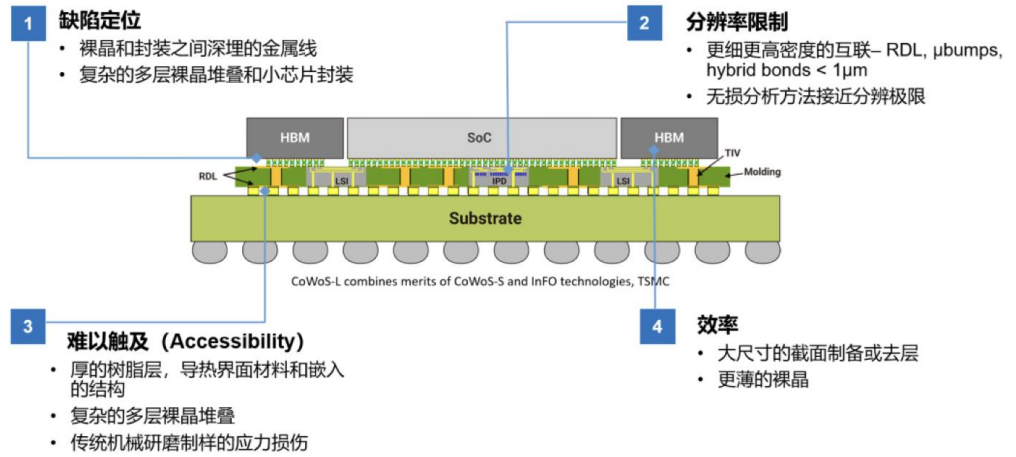
微型化、复杂化的芯片对封装过程中测试技术的精度要求也随之提升。芯片的结构愈发复杂，垂直堆叠封装使得芯片集成度提高，如果无法在封装期间及时检测出缺陷，会导致成本升高、良率降低。然而，多层结构、隐蔽电路、脆弱结构使得检测机器难以



触达、保持元件的完好。下图列举的先进封装失效分析的挑战能够反映缺陷检验的困难。

图22: 封装失效分析的挑战

封装失效分析的挑战

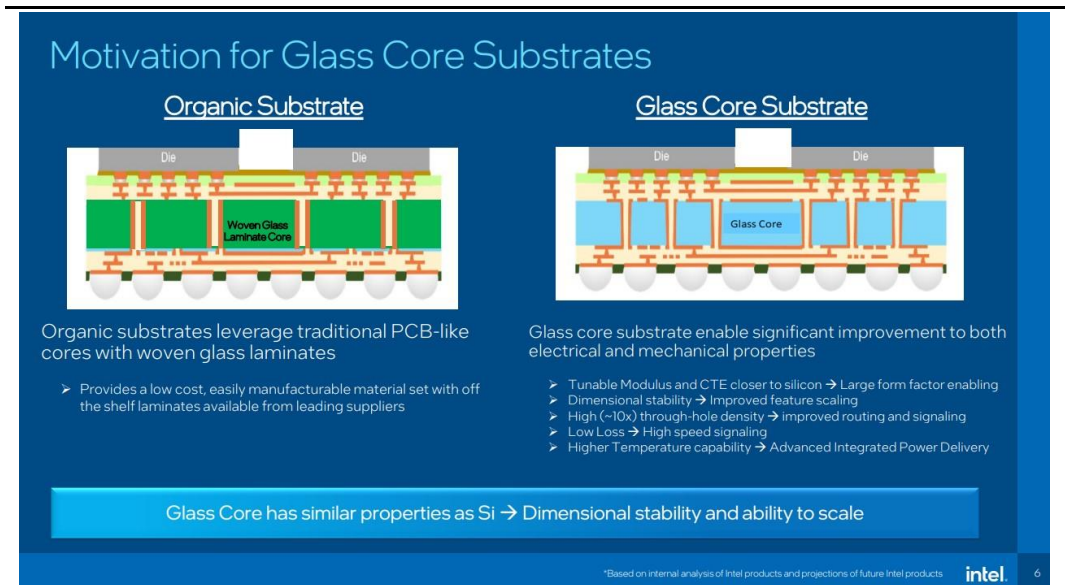


数据来源: 无锡灵恩机电设备有限公司, 东吴证券研究所

1.2.2. 升级逻辑为增加连接效率、降低制造成本

先进封装技术升级改良方向主要有二: 第一, 通过增加连接密度、缩小连接距离或改善材料相关属性以提升连接效率。对于提升连接效率, 增加连接密度、缩小连接距离的主要技术进展已在 1.2 讨论。当下改善触点及电路电器属性的代表技术有玻璃基板。相比有机基板, 玻璃基板能够承受更高的工作温度, 热膨胀系数与硅接近; 有更高的平整度和结构稳定性; 有更好的光学属性, 有利于精细光刻。英特尔预计, 玻璃基板在封装中的使用能够使连接密度增加 10 倍。

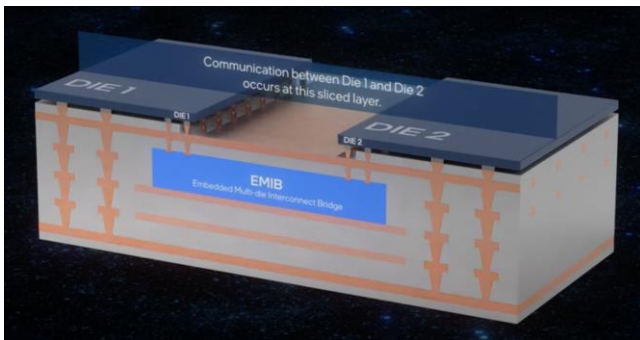
图23: 玻璃封装基板的优势



数据来源: Intel, 东吴证券研究所

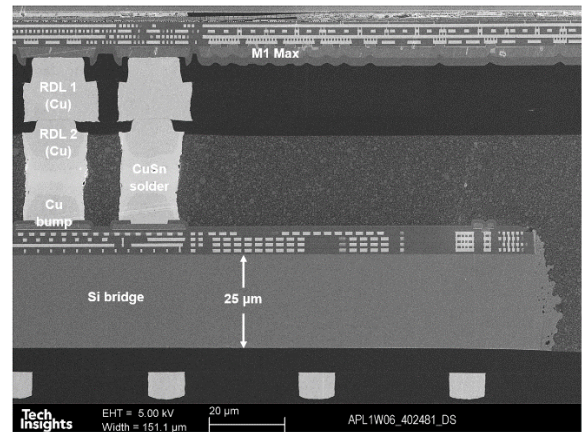
先进封装的另一技术发展方向为降低成本。先进封装对设计精度、制造设备等要求较高，故成本昂贵。为了提高先进封装的普及率，各厂商尝试减少材料用量、使用成本较低的材料及工艺以缓解成本压力。降低封装成本的代表技术有英特尔的 EMIB（嵌入式多芯片互连桥）。EMIB 属于 2.5D 封装，与传统 2.5D 工艺相比，EMIB 舍弃使用整块硅板作为硅中介层，代之以“硅桥”，仅在小芯片的堆栈之间嵌入硅制“桥梁”，减少硅用量；并利用基板中的垂直铜柱向上层芯片供电，减少昂贵的 TSV 技术的使用。当前，台积电的 InFO-LSI 和 CoWoS-L 技术也采取了相似方案。

图24: EMIB 技术示意图



数据来源：英特尔官网，东吴证券研究所

图25: 采用台积电 LSI 技术的 AppleM1Ultra 芯片



数据来源：TechInsights，东吴证券研究所

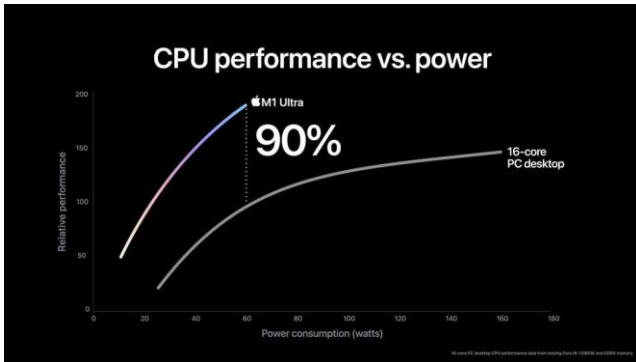
## 2. 算力需求提升导致先进封装产能供不应求

### 2.1. 逻辑芯片为主要需求点，先进封装赋能高速计算

先进封装技术主要应用于提升逻辑芯片的算力。算力通常指芯片每秒可执行的计算任务数量。先进封装主要通过两方面提升逻辑芯片的算力。

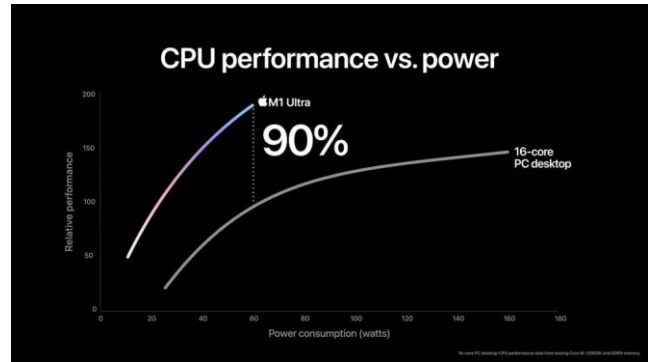
第一，提升处理器集成度，从而提升处理器性能。先进封装使得多处理器间的连接更加快速紧密，因而并行处理数据或进行复杂计算的能力得以加强。比如 AppleM1Ultra 芯片采用硅中介层(Silicon Interposer)和硅桥(Silicon Bridge)技术连接两块 AppleM1Max 芯片，性能提升幅度较大。

图26: AppleM1Ultra 带宽与 16 核 CPU 性能对比图



数据来源: Apple 官网, 东吴证券研究所

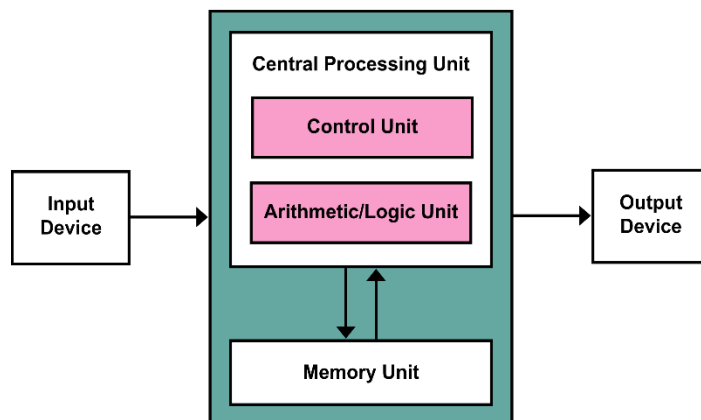
图27: AppleM1Ultra 带宽与 16 核 CPU 功耗对比图



数据来源: Apple 官网, 东吴证券研究所

**第二, 通过解决“内存墙”和“功耗墙”, 以提升计算机算力。**当下主流的计算机设备采用的是冯·诺伊曼架构, 即中央处理器、存储器、控制器等互相分离、各司其职。指令和数据需要从同一存储器存取, 并经由同一总线完成与处理器之间的传输。控制器发布运算指令后, 计算机要先从存储器中调用所需数据至处理器, 运算完成时再将结果传输至存储器。在冯·诺伊曼架构下, 数据在存储器与处理器间反复读取, 造成“内存墙”和“功耗墙”。一方面, 当数据量较多时, 处理器与存储器间的总线带宽(bandwidth)有限, 总线拥挤, 数据传输延迟增大; 另一方面, 数据的反复传输会消耗大量功耗。英特尔的研究表明, 当半导体工艺达到 7nm 时, 数据搬运功耗高达 35pJ/bit, 占总功耗的 63.7%。先进封装通过缩短处理器和存储器间的连接距离、提升连接效率, 能够增加连接带宽, 减小传输功耗。比如 AMD、海力士、英伟达主推的 HBM 内存技术, 利用 TSV 和硅中介层等工艺垂直堆叠 DRAM 芯片, 并将 CPU/GPU 与存储单元封装在一起。和传统显存 GDDR5 相比, HBM 带宽更高、面积更小(与 GPU 总面积减少超过 50%)、功耗更小(HBM2 的功耗减少超过 20%), 因而性能更强, 已成为先进高性能计算芯片的首选内存方案。

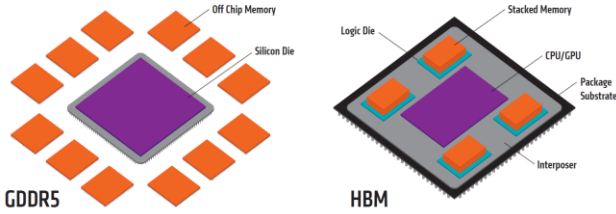
图28: 冯·诺伊曼架构示意图



数据来源: 维基百科, 东吴证券研究所

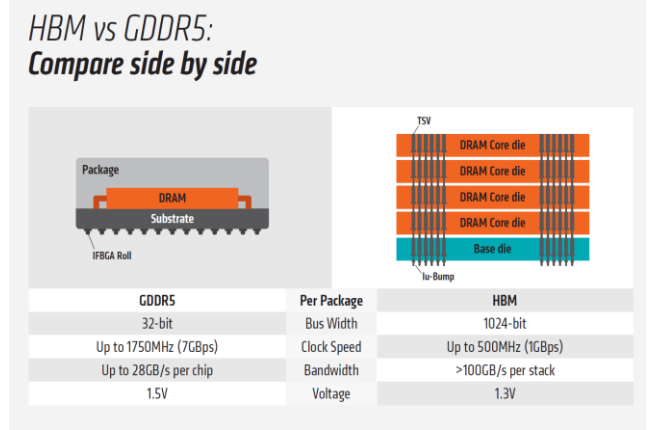
图29: HBM 和 GDDR5 对比示意图

HBM vs GDDR5:  
HBM shortens your information commute



数据来源: AMD 官网, 东吴证券研究所

图30: HBM 和 GDDR5 性能对比

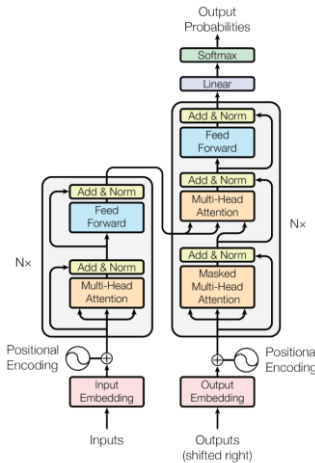


数据来源: AMD 官网, 东吴证券研究所

## 2.2. 算力供不应求, 拉动先进封装需求增长

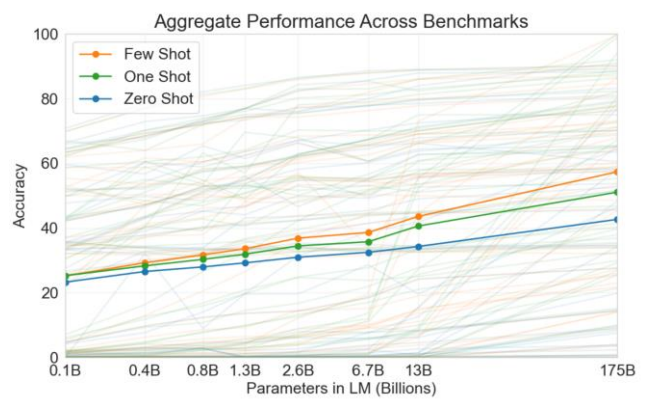
AI 大语言模型对算力的需求呈指数级增长。全球 AI 大预言模型主要采用 Transformer 模型架构。Transformer 模型是一种非串行的神经网络架构, 最初被用于执行基于上下文的机器翻译任务。Transformer 标志性地采用了“注意力层”(AttentionLayers) 结构, 以词嵌入向量叠加位置编码作为输入, 能够跟踪上下文位置的文本间关系, 从而根据输入端文本及文本语料库预测出下文文本, 具有能够并行运算、关注上下文信息、表达能力强等优势。

图31: Transformer 原始模型结构



数据来源: Attentionisallyouneed, 东吴证券研究所

图32: 参数量对语言模型(LM)的性能有积极影响



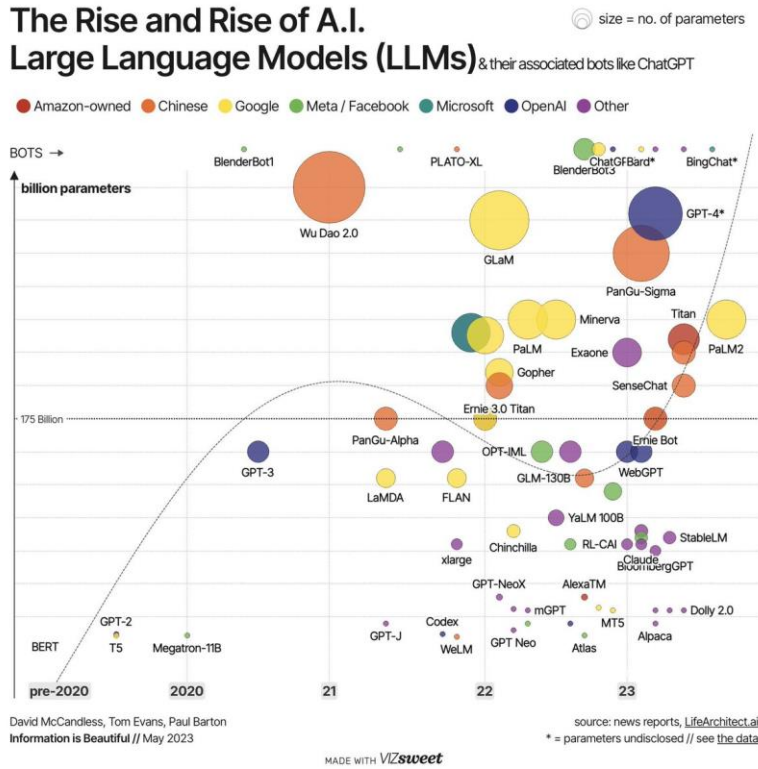
数据来源: LanguageModelsaareFew-ShotLearners, 东吴证券研究所

训练 Transformer 架构的大语言模型需要大规模的参数以及数据库、文本语料库。根据 OpenAI 发表的论文 LanguageModelsaareFew-ShotLearners, 参数量对语言模型的预



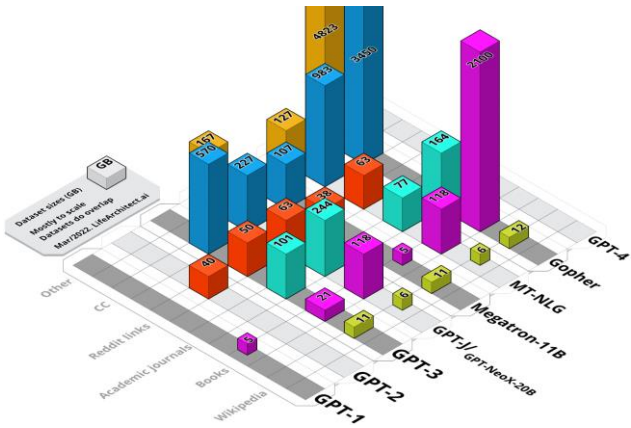
测准确程度能够起到明显积极作用。在当下主流的大语言模型中，GPT-3 和 3.5 的参数量可达 175B，部分国内模型的参数量已远超此数。与此同时，训练模型利用的数据集规模也不断增长。根据 LifeArchitect 的估计数据，截至 2023 年 1 月，全球最大的大语言模型数据集前两名分别是 Pipermonorepo(Google)和 GPT-4(OpenAI)，规模为 86TB 和 40TB。

图33：部分大语言模型参数量



数据来源：LifeArchitect, 东吴证券研究所

图34：部分大语言模型数据集规模及来源



数据来源：LifeArchitect, 东吴证券研究所

图35：2023 年数据集规模排名

#	Lab	Dataset	Size (TB)	Tokens (trillion)	Notes
1	Google	Piper monorepo	86TB	37.9T	DIDACT, code only. From 2016 paper.
2	OpenAI	GPT-4	40TB	20T	1T model .: 20T tokens. gdb said 40TB.
3	TTI	RefinedWeb	23.2TB	5.0T	CC-only dataset prepared by UAE.
4	DeepMind	MassiveText (ml)	20TB	5.0T	From Retro paper.
5	Google	PaLM 2	13TB	3.6T	From PaLM 2 CNBC report.
6	Google	Infiniset	12.6TB	2.8T	From LaMDA paper.

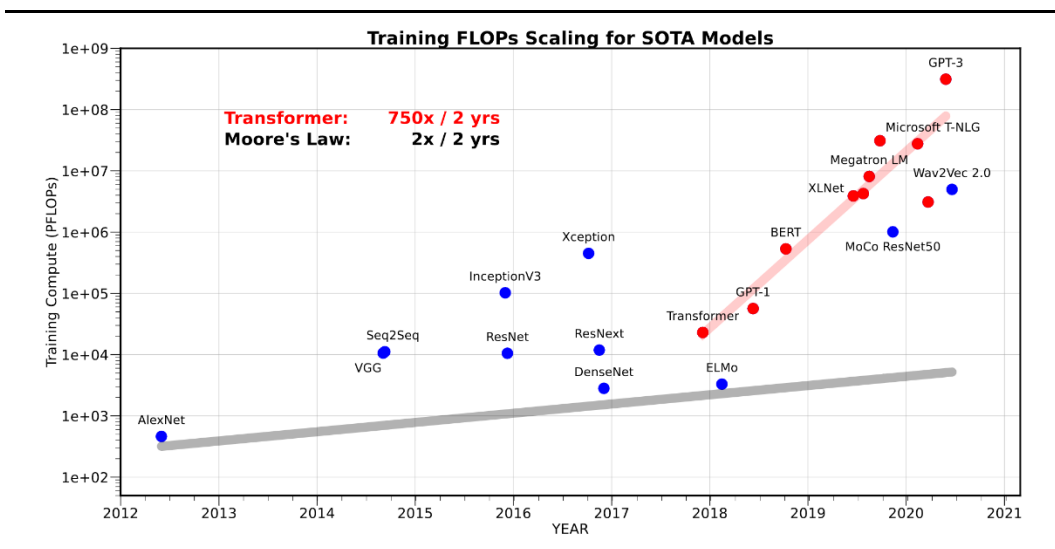
Table. 2023 largest dataset estimates to Jun/2023. Rounded. Disclosed in bold. Determined in italics. For similar models, see my What's in my AI paper.

数据来源：LifeArchitect, 东吴证券研究所



大语言模型的前期训练和后期推理应用所需算力与参数量及数据集规模成正相关关系，随参数和数据集规模迅猛增长。根据 OpenAI 2012 年的测算，2012 年以来，最大的 AI 训练运行所使用的算力呈指数增长，每 3-4 个月增长一倍。2012-2018 年间，AI 训练运行所使用的算力已增长超 30 万倍（摩尔定律预测仅增长 7 倍）。2020 年的数据显示，训练一次 1746 亿参数的 GPT-3 模型需要的算力约为 3640PFlops/天，总训练成本为 1200 万美元，需要连续使用 1024 块英伟达 A100 芯片一个月，后续推理（日常运营）阶段所需算力更大。商汤科技联合创始人杨帆表示，ChatGPT5.0 的训练量对应到算力，约等效于 5 万块的英伟达 H100 芯片。

图36：训练各 Transformer 架构的模型所需算力增速远超摩尔定律



数据来源：UCBerkeleyRISELab, 东吴证券研究所

**国内 AI 模型不断涌现，智能算力需求持续提升，短期内 AI 算力仍旧供不应求。**

需求方面，根据量子位智库的估测，一万枚英伟达 A100 芯片是做好 AI 大模型的算力门槛，国内新入局 AI 行业企业众多（AI 大模型、MaaS、云计算等），而具备此量级的企业最多只有 1 家，GPU 芯片持有量超过一万枚的企业不超过 5 家，企业增购高性能 GPU 需求旺盛。《智能计算中心创新发展指南》指出，在智算中心实现 80% 应用水平的情况下，城市/地区对智算中心的投资可带动人工智能核心产业增长约 2.9-3.4 倍，带动相关产业增长约 36-42 倍，发展弹性明显。IDC 估计，至 2026 年，国内智能算力规模可达 1271.4EFLOPS，约为 2023 年的 3 倍。

供给方面，GPU 主要厂商英伟达产能明显不足以满足市场需求。根据 36 氪在 2023 年 8 月的数据，在各巨头的抢购下，英伟达 H100 在非官方渠道上的溢价高达 25%，23Q3 订购的 H100 需要 24Q1 或 Q2 才能提货。2023 年 8 月，根据外媒报道，英伟达计划将 H100 加速卡（GH100 芯片）的产能拉高至少 3 倍，预测 2024 年出货量将介于 150-200 万颗之间。

智算芯片供应缺口明显，瓶颈在于先进封装产能，国产化替代先进封装产业链空间广阔。目前，提高芯片算力高度依赖 CoWoS、HBM 等先进封装技术。根据

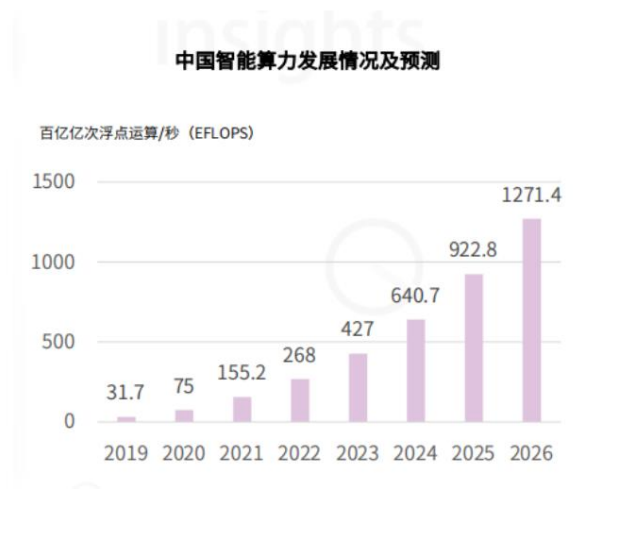
TheInsightPartners 的估测，2020 年，先进封装占据半导体封装市场 40%的份额，到 2030 年，该份额将达到 60%（整理主流 AI 芯片使用到的先进封装技术）。先进封装产能紧张直接限制高性能芯片出货量。2023 年 9 月，台积电董事长刘德音表示，过去一年内 CoWoS 的需求量增加了 2 倍，台积电只能尽量满足客户 80%的需求。根据 Quartz 的报道，CoWoS 的产能缺口成为 H100 的供应瓶颈，一些服务器制造商需要等待六个月才能提货。另一方面，先进封装设备国产化率较低。根据 MIRDATABANK 的数据，2021 年中国大陆主流半导体封装设备划片机、贴片机、引线键合机国产化率均不足 5%，产业链国产替代空间广阔。

图37：2023 年国内 AIGC 算力产业全景图



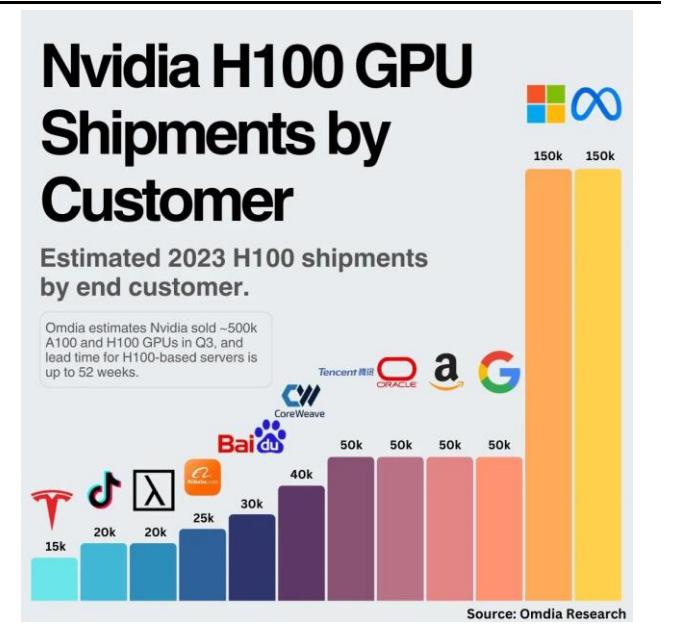
数据来源：量子位智库《AIGC 算力全景与趋势报告》，东吴证券研究所

图38：中国智能算力发展情况及预测



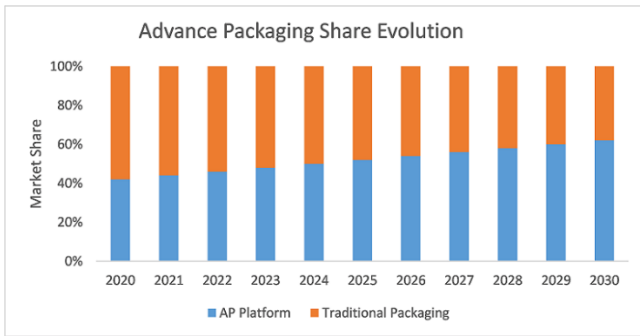
数据来源：量子位智库《AIGC 算力全景与趋势报告》，东吴证券研究所

图39：2023 年英伟达 H100 主要客户购买量估测



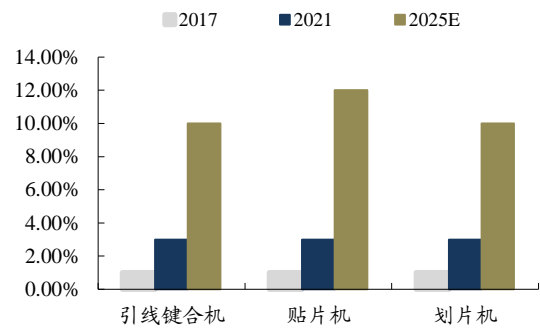
数据来源：OmdiaResearch, 东吴证券研究所

图40: 先进封装市场份额



数据来源: TheInsightPartners, 东吴证券研究所

图41: 主流封装设备国产化率及预测



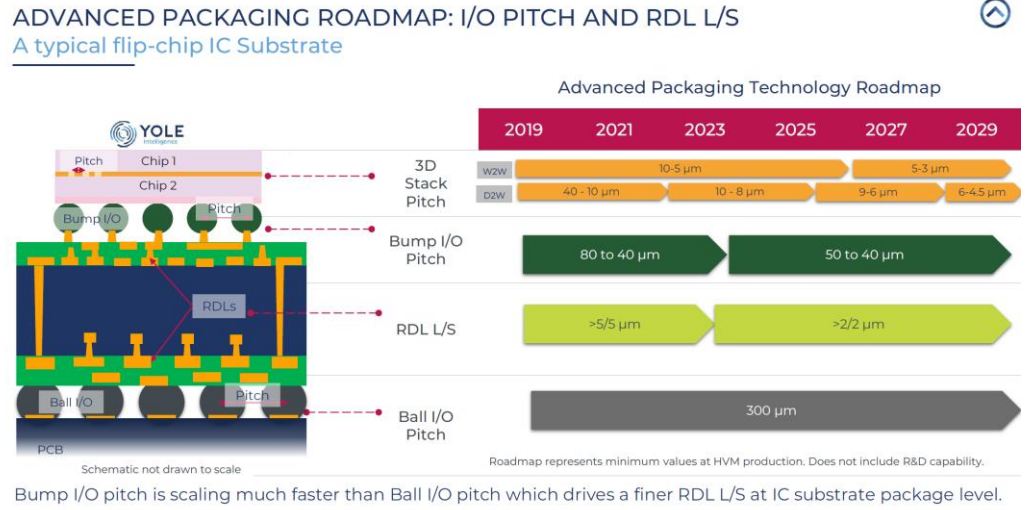
数据来源: MIRDATABANK, 东吴证券研究所制图

### 3. 国内外封装厂企业加速扩产

#### 3.1. 技术及订单承接方面晶圆厂具有较强优势

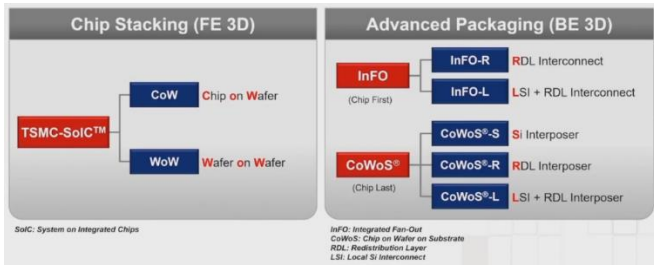
先进封装对精度要求高, 高度依赖晶圆制造技术、与芯片设计环节的协同, 故行业壁垒高, 晶圆代工厂或 IDM 厂比封测厂商更具优势。为了提高连接效率, 先进封装的元件连接非常紧密, 晶圆级封装的 3D 堆叠触点间距可达 10 微米。2021 年, 台积电将其 2.5D 和 3D 封装产品合并为品牌 3DFabric, 并将先进封装进一步划分为“前段”(芯片堆栈技术, 如 CoW 和 WoW)和“后段”(外部导线分布技术)技术, 模糊了晶圆制造和封装环节的界线。这是因为先进封装偏向于前段晶圆制造(光刻、刻蚀、沉积、减薄等)技术, 且需要从芯片设计开始紧密沟通协作, 相比 OSAT 厂, Fab 厂和 IDM 厂更有技术和硬件优势。例如, 重布线层(RDL)、硅通孔(TSV)、混合键合(HB)需要在裸晶本体上进行线路设计、刻蚀、电镀; 硅桥(SiBridge)、硅中介层(interposer)需要利用硅质原件连接裸晶或芯片; 晶圆级封装(WLP)需要先精确对应两块晶圆位置、封装整块晶圆, 再切割成芯片。因此, 龙头 IDM 或晶圆代工厂均依靠自身技术积累开拓先进封测业务, 推进制造封测一体化, 代表有台积电、英特尔、三星。

图42: 2019-2029 先进封装触点间距预测



数据来源: Yole, 东吴证券研究所

图43: 台积电 3DFabric 及前段后段封装技术



数据来源: EE|Times, 东吴证券研究所

图44: 英特尔、三星、台积电先进封装技术布局

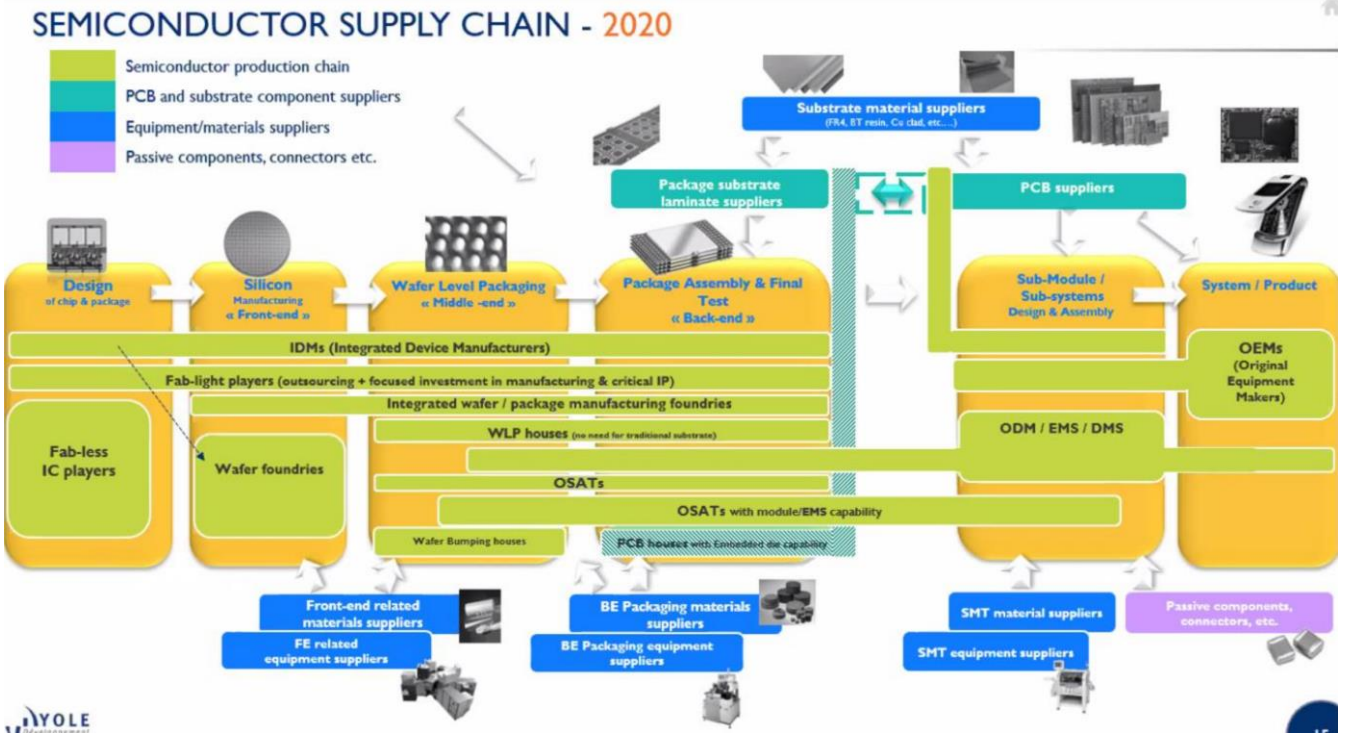
公司	类型	封装名称	Interconnect			
			RDL	Interposer	Stack	Bump
台积电/ TSMC	2D	InFO	InFO_B	RDL		
			InFO_PoP	RDL		
			InFO_SoIS	RDL		Cu/C4
			InFO_SoW	RDL		Cu/C4
			InFO-R/oS	RDL		Cu/C4
	InFO-L/LSI	LSI(TSV)+RDL		Cu/C4		
	2.5D	CoWoS	CoWoS-S		Si	Cu/C4
CoWoS-R			RDL		Cu/C4	
3D	SoIC	CoWoS-L	LSI(TSV)+RDL		Cu/C4	
		SoIC-CoW			CoW DCB	
		SoIC-WoW			WoW DCB	
三星/ Samsung	2.5D	I-Cube	I-Cube S			DCB
			I-Cube E		Si	Bump
	3D	X-Cube	X-Cube Bump	Si Bridge+RDL		CoW Bump
			X-Cube HCB	Si+ABF		CoW DCB
英特尔/ Intel	2.5D	EMIB		Si Bridge+RDL		Bump
			Foveros			CoC Cu
			Foveros Omni			CoC Cu
3D	Co-EMIB	Foveros Direct			CoC DCB	

数据来源: Yole, 东吴证券研究所

与上游晶圆厂的合作紧密程度成为先进封装行业除技术之外的壁垒之一。英特尔、台积电、三星凭借其晶圆厂在前道环节的优势，主导了先进封装领域的技术路线和订单分配，OSAT 工厂只能分担此类厂商的产能压力或供应先进封装的低端环节。因此，除技术外，OSAT 厂商需要与上游厂商密切合作以获取先进封装订单。



图45: IDM 及 Foundry 厂商进入先进封装环节



数据来源: Yole: Status of the Advanced Packaging 2023, 东吴证券研究所

### 3.2. 海外厂商积极扩产，扩产周期约 2-3 年

目前，先进封装需求高增，产能紧缺，各海外龙头加大扩产力度，但扩产普遍难度大、周期长。以台积电为代表的晶圆代工厂，英特尔、三星为代表的 IDM 厂商，以及以日月光为主的 OSAT 厂商纷纷增加先进封装产线。先进封装上游设备供应不足等原因导致扩产速度较慢，新建工厂普遍需要 2-3 年才能量产。台积电采购的 CoWoS 设备需要超过 6 个月才能交付，先进封装七厂预计至 2027 年 Q3 才能量产。短期内先进封装产能缺口无法解决，将持续制约高算力芯片出货量。



表1: 海外先进封装厂商龙头扩产规划

企业	扩产规划
台积电	目前已建成6家先进封测厂。应众多客户要求,台积电于2023年Q2开始紧急为CoWoS购进设备、配置产能。23年底台积电CoWoS月产能约为15000片晶圆,追加设备进驻后,月产能预计可达20000片以上,并逐季增加。2023年7月25日,台积电对外表示规划于铜锣科学园区设立第七个先进封装厂。然而,新CoWoS设备的交付时间超过6个月,搭建一条完整CoWoS生产线需要6-9个月;先进封测七厂预计至2027年Q3才能开始量产。
英特尔	目前在美国奥勒冈州和新墨西哥州建成2座先进封装厂,2021年5月宣布斥资35亿美元扩充新墨西哥州先进封装产能。2023年8月宣布在马来西亚槟城建立先进封装新厂,预计2024年底到2025年完工投产,该厂将成为英特尔最大的3D先进封装基地。英特尔规划2025年3DFoveros封装产能达2023年水平的4倍。
三星	2023年计划在韩国天安厂区建立一条HBM所需的新封装线,用于供应高性能芯片厂商,并计划于2024年将HBM产能提升为当前的2.5倍。三星的HBM3已通过英伟达和AMD的质量检测,即将成为供应商。
日月光	2023年12月,日月光租用福雷电子高雄楠梓厂房,扩充高雄厂先进封装产能,以满足AI芯片所需。日月光和晶圆厂合作先进封装中介层(interposer)相关技术,研发CoWoS解决方案,预计即将跨入认证CoW环节,分担台积电CoWoS产能压力。

数据来源: 各公司官网及公告, 东吴证券研究所整理

### 3.3. 国内公司加速布局先进封装, 关注长电、通富、甬矽等

国内龙头正在积极布局先进封装领域, 代表有长电科技、通富微电等。国内先进封装产业起步较晚, 技术较为落后, 主要承接高性能芯片封装的后道工艺。近年高性能芯片封装产能缺口加大, 国内封测厂纷纷布局先进封装。国内龙头长电科技聚焦XDFOI新技术、2.5D/3D技术的量产; 通富微电利用与AMD的密切关系及自身Chiplet技术优势扩产消化高端CPU、GPU封装产能, 现已涉及AMD MI300的封装; 甬矽电子积极研发Bumping、RDL等技术, 展望Fan-in/Fan-out、2.5/3D晶圆级封装, 并大幅建厂扩产, 营收增长空间广阔。先进封装国产替代发展空间大, 看好国内龙头。

表2: 国内先进封装厂商主要产品及产业布局

企业	主要封装产品	先进封装产业布局
长电科技	SiP、TSV 晶圆级封装 Flip-Clip MEMS	<p>全球第三大专业封测厂。公司在先进封装领域起步较早，与中芯国际合作紧密，有望在前道工艺达成合作，形成相对其他 OSAT 厂商的差异化优势。</p> <ul style="list-style-type: none"> <li>● 2015 年，收购全球第四大 OSAT 星科金朋，获得 SiP、FOWLP 等先进技术，跻身国际顶级封测梯队。</li> <li>● 2016 年设立长电韩国，拓展海外业务。</li> <li>● 2021 年发布 XDFOIChiplet 高密度多维异构集成系列工艺，并于 2023 年 1 月进入稳定量产阶段。</li> <li>● 2022 年设立长电微电子晶圆级微系统集成高端制造项目，项目一期将于 2024 年初竣工并投产，聚焦 2.5D/3D 技术，面向高算力芯片。</li> <li>● 2022 年设立上海创新中心，加速搭建先进封测技术研发服务平台。</li> <li>● 2023 年计划资本开支 65 亿元，先进封装占比超 80%。</li> </ul>
通富微电	FCBGA、FCPGA、 FCLGA、MCM 等 倒装技术， FOWLP，SiP， Chiplet 封装解决方 案	<p>全球第四大专业封测厂，是 AMD 封测的最大供应商，与 AMD 合作紧密。</p> <ul style="list-style-type: none"> <li>● 2016 年，收购 AMD 苏州和槟城封测厂各 85% 股权，获得国际先进技术。</li> <li>● 公司持续积极扩产。2021 年开始建设通富同科南通工厂、通富超威槟城工厂，2023 年均已竣工投产。</li> <li>● 拥有国内最完善的 Chiplet 封装解决方案，自建 2.5D/3D 产线，面向以 AMD 为主的高端 CPU、GPU 产品封装。</li> <li>● 公司是国内研发费用投入最多的封测厂，2022 年投入 13.23 亿，研发费用率达 6.17%，研发占比稳健。</li> </ul>
甬矽电子	OFN/DFN、Flip- Clip、SiP、Hybrid- BGA、MEMS	<p>成立于 2017 年，2022 年 11 月在上交所上市，2023 年国内市场份额第六，专注于中高端封装，技术水平位于行业前列。</p> <ul style="list-style-type: none"> <li>● 2018 年后逐步实现多种尖端产品及技术的量产，包括大尺寸/细间距无引脚封装（QFN/DFN）、焊线类 BGA 和系统级封装（SiP）等，目前封装技术以 SiP 为主。</li> <li>● 近年来积极研发 Fan-in/Fan-out、2.5/3D 晶圆级封装相关技术。2023 年，公司在 Bumping、RDL 及大尺寸 FC-BGA 领域取得技术突破，具备了“Bumping+CP+FC+FT”的一站式交付能力，并尝试开发 TSV 等 3D 工艺。</li> <li>● 公司持续向车规级、工控级和 5G、物联网等应用市场扩张拓展，调整产品结构。</li> <li>● 2021 年 3 月开始，为扩大先进封装产能，公司分阶段投入建设二期工厂，预计总投资逾 110 亿元，满产将达年产 130 亿颗芯片，聚焦先进晶圆级封装、PoP、TSV 等，2023 年 9 月工厂完工，部分产线已投入使用。</li> </ul>

兴森科技  
PCB、IC 封装基  
板、FPC、半导体  
测试板

- 2022 年 IPO 募投项目包括高密度 SiP 射频模块封测扩产项目，2024 年初达产后每月将新增 14500 万颗产能。

成立于 1999 年，2010 年于深圳交易所中小企业板上市，是全球电子电路方案数字制造提供商，产品布局覆盖电子硬件三级封装领域，为客户提供从设计到测试交付的高价值整体解决方案。

- 公司以收购、投资等方式获取技术、拓展业务版图。2013 年兴森香港收购 Exception PCB Solutions Limited，获取欧洲市场中高端样板、快件制造的业务；2015 年，兴森香港收购 Xcerra Corporation 半导体测试板相关业务，开拓美国市场；2015-2021 年，兴森香港持续增资国际领先的 PCB 供应商 Finline，2021 年以合计 100% 的股权控股。
- 2020 年，珠海兴科半导体与国家“大基金”项目合资，启动封装基板项目，预计 2024 年一季度进入小批量生产阶段。
- 2022 年，公司宣布进入 FCBGA 封装基板领域，广州兴森半导体 FCBGA 封装基板项目正式动工，2023 年年底完成产线建设。

成立于 2005 年，2014 年于上海证券交易所上市，收入主要来源于 TSV 和 3DIC 技术板块。拥有 WLCSP、TSV 等先进封装技术，专注高端封装，CMOS 影像传感器晶圆级封装技术全球领先。

晶方科技  
WLCSP、TSV、  
3DIC

- 公司以并购方式实现业务布局扩张。2014 年公司收购智瑞达电子，完善了安防、车用影像传感器布局；2019 年收购 Anteryon 公司，拓展光电传感系统业务；2021-2023 年投资并控股以色列氮化镓设计公司 VisIC，布局车用高功率氮化镓技术。
- 2005-2008 年，引进以色列 Shellcase 技术，填补国内晶圆级尺寸封装技术的空白。
- 2009-2011 年，自主创新开发 THINPAC 技术，并在硅谷建立研发中心。
- 2012-2014 年，在国内成功建成全球首条 12 英寸晶圆级硅通孔封装量产线，自主开发生物身份识别技术。
- 2015-2019 年，推出针对高端产品领域的 Fan-out 技术。
- 2020-2023 年，成立车规半导体产业技术研究所，在新加坡投资设立全资子公司，建立公司海外业务中心、研发工程中心与投融资平台，同时布局全球化的生产与制造基地，进一步推进公司的国际化发展战略。

数据来源：各公司官网及公告，上海证券交易所，东吴证券研究所

## 4. 风险提示

AI 算力需求增长不及预期。先进封装行业下游以 AI 智能运算及大模型为主，若需求增长节奏较慢，下游行业发展低于预期，先进封装行业的业绩增长也会受到影响。

先进封装技术进展缓慢。如果先进封装技术的发展未能如预期成功降本增效、提高产能，市场规模增速和渗透率可能不及预期。

国产替代不及预期。先进封装对技术和工艺的要求极高，国内本身的技术和硬件积淀不足以满足要求，需要先进生产设备和研发资金的不断投入，成本显著提高，如未能准确把握市场需求或成功取得技术突破，容易在与海外龙头的竞争中落后，国产替代进程不及预期。

## 免责声明

东吴证券股份有限公司经中国证券监督管理委员会批准,已具备证券投资咨询业务资格。

本研究报告仅供东吴证券股份有限公司(以下简称“本公司”)的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下,本报告中的信息或所表述的意见并不构成对任何人的投资建议,本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。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。

在法律许可的情况下,东吴证券及其所属关联机构可能会持有报告中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投资银行服务或其他服务。

市场有风险,投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息,本公司力求但不保证这些信息的准确性和完整性,也不保证文中观点或陈述不会发生任何变更,在不同时期,本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有,未经书面许可,任何机构和个人不得以任何形式翻版、复制和发布。经授权刊载、转发本报告或者摘要的,应当注明出处为东吴证券研究所,并注明本报告发布人和发布日期,提示使用本报告的风险,且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的,应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

## 东吴证券投资评级标准

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力相对基准表现的预期(A 股市场基准为沪深 300 指数,香港市场基准为恒生指数,美国市场基准为标普 500 指数,新三板基准指数为三板成指(针对协议转让标的)或三板做市指数(针对做市转让标的),北交所基准指数为北证 50 指数),具体如下:

公司投资评级:

- 买入: 预期未来 6 个月个股涨跌幅相对基准在 15%以上;
- 增持: 预期未来 6 个月个股涨跌幅相对基准介于 5%与 15%之间;
- 中性: 预期未来 6 个月个股涨跌幅相对基准介于-5%与 5%之间;
- 减持: 预期未来 6 个月个股涨跌幅相对基准介于-15%与-5%之间;
- 卖出: 预期未来 6 个月个股涨跌幅相对基准在-15%以下。

行业投资评级:

- 增持: 预期未来 6 个月内,行业指数相对强于基准 5%以上;
- 中性: 预期未来 6 个月内,行业指数相对基准-5%与 5%;
- 减持: 预期未来 6 个月内,行业指数相对弱于基准 5%以上。

我们在此提醒您,不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系,表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况,如具体投资目的、财务状况以及特定需求等,并完整理解和使用本报告内容,不应视本报告为做出投资决策的唯一因素。

东吴证券研究所  
苏州工业园区星阳街 5 号  
邮政编码: 215021  
传真: (0512) 62938527  
公司网址: <http://www.dwzq.com.cn>