

半导体产业链调研报告

集成电路四种生产模式与先进封装专题

—— 涵盖 IDM、Fabless、Foundry、OSAT 四种模式 ——

—— 涉及 Intel、SK 海力士、Nvidia、AMD、台积电、ASE 等核心企业 ——

数据截至：2026 年 5 月

摘要

本报告系统梳理半导体产业链的整体结构，并以“集成电路生产的四种主要模式”为主线，深度展开 IDM、Fabless、Foundry、OSAT 四种模式各自的经济模型、典型企业、竞争格局与未来趋势。报告采用 2024-2026 年最新公开数据，结合定量与定性分析，覆盖以下核心内容：

（一）行业整体框架——半导体产品分类、产业链上中下游全貌、四种生产模式的历史演进与对比；

（二）IDM 模式深度——以 Intel 与 SK 海力士为典型案例，剖析两家公司在 AI 时代呈现出的“衰落与崛起”反差；

（三）Fabless 模式深度——五大族群格局，并以 AMD vs Nvidia 的 AI 算力之战作为专题展开；

（四）Foundry 模式深度——以“一超多弱”格局为核心，重点剖析台积电、三星 Foundry、中芯国际，并展开 CoWoS 与先进封装专题；

（五）OSAT 模式深度——ASE、Amkor、长电、通富等头部企业的战略布局，以及 OSAT 与 Foundry 在先进封装领域的“地盘之战”；

（六）A 股先进封装概念股图谱——区分“真受益”与“主题炒作”，覆盖封测三巨头、设备厂商、材料与基板供应商、HBM 国产化主线。

核心判断：摩尔定律放缓后，半导体产业链的“权力中心”正在从晶圆制造向先进封装转移；AI 算力爆发让所有四种模式都受益，但受益程度差异巨大；地缘政治正在重塑全球产业地图，纯粹的全球协作模式正在被区域多元化与本土化模式替代。

第一章 半导体产业链整体框架

1.1 半导体与集成电路的基本概念

半导体是一类导电性介于导体和绝缘体之间的材料，最典型的就硅，可以通过掺杂工艺精确控制它的导电特性。集成电路（IC）则是把大量晶体管、电阻、电容等元件集成到一小块半导体衬底上的电子电路。日常所说的“芯片”，绝大多数指的就是 IC。

广义的“半导体产品”包括四大类：集成电路（占整个半导体市场约 80%）、分立器件（二极管、三极管、IGBT 等）、光电器件（LED、CIS 图像传感器、激光器）、传感器（如 MEMS）。但讨论“半导体行业”时，重心都落在 IC 上。

1.2 产业链全景

完整的半导体产业链可以分为上游、中游、下游三段。

上游（支撑环节）包括两块：一是半导体材料，包括硅片、光刻胶、电子特气、CMP 抛光液、靶材、掩模版、湿电子化学品等；二是半导体设备，最关键的是光刻机（ASML 一家独大，EUV 光刻机全球独家供货），其次是刻蚀设备（泛林、应用材料、东京电子）、薄膜沉积设备、CMP、离子注入机、量测设备等。这一段技术壁垒极高、行业高度集中。

中游（核心制造环节）按工艺分三步：IC 设计——做电路、做版图、用 IP；晶圆制造（前道工艺）——把电路真正“刻”到硅片上；封装测试（后道工艺）——把晶圆切割、封装成芯片并测试。

下游（应用市场）包括智能手机、PC、服务器、汽车电子、工业、通信，以及当下最大的增量“AI 数据中心”。下游的拉动效应非常明显——AI 算力起来之后，直接把 HBM 存储、先进封装（CoWoS）、台积电先进制程产能全拉爆。

1.3 集成电路生产的四种主要模式

20 世纪 80 年代以前只有一种模式——IDM（垂直整合），所有环节都自己做。1987 年台积电成立、首创晶圆代工模式之后，整个行业开始走向专业分工，到今天主要演化出四种模式：

模式	英文全称	业务范围	代表企业
IDM	Integrated Device Manufacturer	设计 + 制造 + 封测全包	Intel、三星、SK 海力士、TI
Fabless	Fabless Design Company	只做设计，制造与封测外包	Nvidia、AMD、Apple、高通
Foundry	Pure-play Foundry	只做晶圆制造	台积电、三星、SMIC、Intel Foundry
OSAT	Outsourced Semiconductor Assembly and Test	只做封装与测试	ASE、Amkor、长电科技、通富微电

四种模式各有优劣：

模式	资本开支	技术门槛	毛利率水平	主要风险
IDM	极高	设计 + 制造双高	中高（因领域而异）	周期波动 + 产能利用率
Fabless	低	设计高	高（50%+）	依赖代工厂、市场竞争
Foundry	极高	制造高	中（40%-55%）	资本周期 + 地缘政治
OSAT	中	中（先进封装变高）	中低（15%-30%）	价格战 + 客户集中

行业大趋势上，纯 IDM 在向“Fab-Lite”过渡（如 Intel 开放代工业务，AMD 早期已完全 Fabless 化）；同时 Fabless + Foundry 这种分工模式因为先进制程门槛太高（3nm 一座工厂超 200 亿美元），越来越占主导。

第二章 IDM 模式：垂直整合的旧王与新王

2.1 IDM 模式的本质：为什么这种“重资产+全链条”模式仍然存活

讲 IDM 不能仅停留在定义层面，要从产业经济学角度看它“为什么”存在、“为什么”还没被 Fabless+Foundry 完全取代。原因主要有四个：

第一是工艺与设计的深度耦合。某些产品的电路特性高度依赖具体工艺参数，比如模拟芯片、功率半导体、CIS 图像传感器，工艺微小波动就会显著影响产品性能，因此设计必须和制造紧密协同，外包给 Foundry 的“标准化工艺”反而做不出最佳性能。

第二是产品标准化 + 巨量产能。存储器（DRAM/NAND）几乎是全球最标准化的半导体产品，规格由 JEDEC 统一定义，比拼的就是制程演进速度、良率、单位成本。这种品类天然适合 IDM——自己投巨资建厂、自己研发制程、自己控制成本曲线。Fabless+Foundry 在存储里完全跑不出成本优势。

第三是产业历史路径依赖。英特尔、TI、英飞凌这些公司都是从 1960-80 年代就开始建厂的，Fab 是它们最核心的战略资产，不会轻易剥离。

第四是地缘政治与供应链安全。这两年特别明显——美、欧、日、韩都在重新强调“本土制造”，IDM 模式天然契合这种安全诉求。美国 CHIPS Act 的钱大头给了 IDM（Intel、Micron、TI 等）。

2.2 全球 IDM 阵营的整体格局

按 2025 年口径，全球 IDM 市场规模约 3274 亿美元。前 10 大 IDM 企业中，三星、Intel、SK 海力士三家就占 IDM 市场约 40% 收入。

排名	企业	主营	大致定位
1	三星电子	存储 + 逻辑 + 代工	全能型
2	Intel	CPU + 代工	逻辑龙头（受困）
3	SK 海力士	存储	DRAM/HBM 王者
4	美光（Micron）	存储	存储第三
5	英飞凌（Infineon）	功率 + 汽车	汽车龙头
6	德州仪器（TI）	模拟	模拟龙头
7	意法半导体（ST）	模拟 + 功率 + MCU	SiC 领先
8	恩智浦（NXP）	汽车 + 边缘 AI	汽车第二梯队
9	索尼	CIS 图像传感器	CIS 全球第一
10	Murata 等	被动器件等	—

我们按主营产品把 IDM 分成三大族群分别讲。

2.3 存储 IDM 三巨头：HBM 重新洗牌

存储是 IDM 模式最纯粹的体现。全球 DRAM 市场三家玩家——三星、SK 海力士、美光，三家市占率合计长期超过 95%；NAND 略多一点（三星、SK 海力士、美光、铠侠、西部数据、长江存储），但前三家也在 70% 以上。

过去三星几乎一直是 DRAM 的王者，但 2025 年发生了一件标志性的事：**SK 海力士首次超越三星，成为全球最大的 DRAM 厂商**，并整体首次登顶全球第一大存储芯片供应商。背后核心原因就是 HBM（高带宽存储）。

HBM 是 AI 时代的“印钞机”。它通过 3D 堆叠把 DRAM 颗粒垂直叠起来、用 TSV（硅通孔）连接，提供超高带宽，是英伟达、AMD AI GPU 的标配。HBM 的单位价值是普通 DRAM 的 5-7 倍，毛利率高得惊人。

HBM 三家在 2025 年的市占率变化非常戏剧：

- 2025 年 Q2：SK 海力士 62%，美光 21%，三星 17%
- 2025 年 Q3：SK 海力士 57%，三星 22%（HBM3E 量产追上来），美光 11%

SK 海力士赢在两点：早押 HBM（2013 年就和 AMD 合作做了第一代），以及绑死英伟达（Nvidia AI GPU 的 HBM 几乎独家来自海力士）。三星因为 HBM3E 良率和散热问题迟迟过不了英伟达认证，掉到了第三，但 2026 年 HBM4 量产后份额预计能回升到 30% 以上。

2.4 逻辑 IDM：Intel 一家独苗

逻辑芯片（CPU/GPU/SoC）领域，**真正的 IDM 现在只剩 Intel 一家**。AMD 早在 2009 年就把晶圆厂剥离成立了格芯，自己变成纯 Fabless；苹果、英伟达、高通、博通从来就是 Fabless；三星虽然有自己的 Foundry，但 Logic 业务严格说是 IDM + Foundry 混合体。

Intel 的故事是过去五年半导体行业最大的戏剧（详见 2.6 节）。

2.5 模拟 / 功率 / 汽车 IDM：最坚固的堡垒

这是 IDM 最不会被颠覆的领域：工艺定制化、品类极度碎片化、单品价值不高但毛利率高、生命周期长。主要玩家及其特点：

- **TI（德州仪器）**：模拟芯片全球第一，约占模拟市场 19% 份额，产品 SKU 上万种。2025 年正在大力扩 12 寸模拟产能。
- **ADI（Analog Devices）**：高性能模拟，2021 年收购 Maxim 之后规模翻倍。
- **Infineon（英飞凌）**：汽车半导体全球第一，强在功率（IGBT、SiC、GaN）、MCU、传感器。
- **NXP（恩智浦）**：汽车第二梯队，强在车规 MCU、雷达、车联网。
- **ST（意法半导体）**：SiC 全球领先，特斯拉是大客户，同时做 MCU（STM32 系列）。

- **Renesas (瑞萨)**：日本汽车 MCU 龙头。2025 年 5 月退出 SiC 业务，战略收缩到 MCU。
- **Bosch (博世)**：汽车 IDM，做 MEMS 传感器、IGBT、SiC，主要供自家。

汽车半导体市场预计从 2024 年的 680 亿美元增长到 2030 年的 1320 亿美元，**Top 5 (Infineon、NXP、ST、TI、Renesas) 合计占据 50% 以上**。模拟/汽车 IDM 毛利率普遍 55-65%，是华尔街最爱的“防御性半导体”。

2.6 案例研究一：Intel —— 一个 IDM 帝国的衰落与自救

2.6.1 Intel 的“黄金时代”

Intel 1968 年由摩尔 (Gordon Moore) 和诺伊斯 (Robert Noyce) 创立，从存储起家。1985 年因日本 DRAM 冲击退出存储，转向 CPU——这是它人生最关键的战略决策。1990s-2010s 这二十多年里，Intel 几乎是 IDM 模式的最佳代言人：“Tick-Tock”模式（一年缩制程、一年改架构）被业界奉为圣经，鼎盛期（2010-2014）Intel 在 22nm、14nm 节点上比台积电、三星至少早 2 年量产，CPU 市场份额 90%+，毛利率 60%+。

2.6.2 转折点：10nm 的灾难（2014-2020）

Intel 衰落的伏笔，埋在 10nm 工艺。按计划 Intel 应在 2015 年量产 10nm，但因为光刻方案（坚持用 DUV 多重曝光）、技术目标过激、内部治理问题等，**10nm 延期了整整 5 年**。

这 5 年是台积电“突袭”的关键窗口期：从 16nm → 10nm → 7nm → 5nm 跨越了三代节点。AMD 借助台积电 7nm 推出 Zen 2/Zen 3，性能能效全面反超 Intel；苹果用台积电 5nm 推出 M1；英伟达用台积电产能扩展 AI GPU。到 2020 年，Intel 已经从“工艺领先 2 年”变成“落后台积电至少 2 年”，IDM 模式的核心壁垒——制程领先——崩塌了。

2.6.3 Pat Gelsinger 时代（2021-2024）：豪赌 IDM 2.0

2021 年 2 月，Intel 把已经离开 12 年的“老将” Pat Gelsinger 请回来当 CEO。基辛格提了一个极其激进的策略——**IDM 2.0**：保留 IDM 身份的同时，开放产能对外做代工（Intel Foundry Services），同时也允许内部产品外包给 TSMC，并提出“四年五个节点”目标。

为了实现这个目标他疯狂建厂——美国亚利桑那、俄亥俄、德国马格德堡、以色列、爱尔兰，**资本开支飙到每年 250 亿美元以上**。但客户拓展极慢、制程节奏不稳、Intel 20A 直接被取消、**2024 年全年净亏损 188 亿美元**。2024 年 12 月基辛格被董事会“提前退休”。

2.6.4 Lip-Bu Tan 时代（2025 年 3 月至今）

2025 年 3 月 12 日，Cadence 前 CEO 陈立武 (Lip-Bu Tan) 就任 Intel CEO。基辛格是技术理想主义者，陈立武则是运营纪律执行者。新策略包括：

- **聚焦 18A:** 放弃 Intel 20A, 把全部精力压到 Intel 18A (约 1.8nm, 引入 RibbonFET + PowerVia)。2025 年 10 月 Intel 18A 进入量产, 搭载它的 Panther Lake 笔记本 CPU 已在 200+ 设计中出货。
- **大幅瘦身:** 2025 年裁员 1.5 万人左右。
- **外部融资输血:** 2025 年 Intel 拿到了软银 20 亿美元、英伟达 50 亿美元、美国政府 89 亿美元 (CHIPS Act)、Silver Lake 44.6 亿美元 (收购 Altera 51% 股份)。
- **代工对外开放:** 18A 和 18A-P 将面向外部客户开放, 微软和 AWS 已确认成为 18A 定制 AI 芯片的锚定客户。

2.6.5 Intel 当前的真实财务状况

Q4 2025 财报最新数据: 全年营收 **529 亿美元** (自 2010 年以来最弱的一年), 全年净亏损 **3 亿美元** (相比 2024 年的 188 亿巨亏接近盈亏平衡); Q4 营收 137 亿美元, 同比下降 4%; Intel Foundry 仍在巨亏: Q4 单季度运营亏损 **25 亿美元**。

解读: 产品业务 (Client + Datacenter) 已经开始触底反弹, 但 Foundry 业务还在失血。

Intel 2025 年的"亏损大幅收窄"主要靠外部融资 (150+ 亿美元的注资) 撑住, 本业还没真正翻身。

2.7 案例研究二: SK 海力士 —— 一场押注 HBM 改写命运的豪赌

2.7.1 从 Hyundai 到 SK 的曲折历史

SK 海力士 1983 年起源于现代电子产业 (Hyundai Electronics), 1996 年起做 DRAM。亚洲金融危机后, 1999 年现代电子合并 LG 半导体改名 Hynix。2001 年又因债务危机面临破产, 差点被美光收购。2012 年韩国 SK 集团以约 30 亿美元收购海力士。当时这笔交易在韩国财经界争议极大——SK 并没有半导体经验, 被认为是"门外汉花高价买了一个负债公司"。

2.7.2 押注 HBM 的"长期主义"

收购后 SK 海力士长期被定位成"全球第二大 DRAM 厂商", 市占率比三星低 10-15 个百分点。但有两件事它做对了:

第一是死守 DRAM 这个核心赛道。不像三星那样四面铺开 (存储、Logic、代工、显示面板都做)。专注让它在每一代 DRAM 制程上都紧跟三星半步。

第二是 2013 年押注 HBM。SK 海力士早在 2013 年就和 AMD 一起做了全球第一代 HBM。当时 HBM 市场极小、价格贵、应用场景窄, 三星和美光当时都把 HBM 当成"边缘产品"对待。但 SK 海力士坚持投入 10 年——HBM2 → HBM2E → HBM3 → HBM3E → HBM4。每代领先半步到一步。

2.7.3 HBM 王者: 被 Nvidia"独家绑定"

2023 年起, Nvidia 数据中心 GPU 爆炸式出货, 每颗 GPU 都需要 6-8 个 HBM。SK 海力士因为 HBM3 良率领先 (独家供货 H100/H200)、最早通过 Nvidia HBM3E 验证、与 Nvidia 联合定义规格, 成了 AI 浪潮里最大的“卖铲子的人”。

数据非常猛: 2025 年 Q2 HBM 市占率 62%、Q3 仍有 57%——一家独占全球 HBM 市场一半以上。HBM 单位价值是普通 DRAM 的 5-7 倍, 毛利率 60-70%。

2.7.4 2025 年的“现金印钞机”

SK 海力士 2025 年财报数据 (创下公司历史所有纪录):

- 全年营收 **97.15 万亿韩元** (约 700 亿美元, 同比增长超 30 万亿韩元)
- 全年运营利润 **47.21 万亿韩元** (约 340 亿美元), 运营利润率 **49%**
- 全年净利润 **42.95 万亿韩元** (约 310 亿美元), 净利润率 44%, 同比 +117%
- 同时宣布了公司历史最高的股东回报

49% 的运营利润率是什么概念? 台积电 2024 年运营利润率约 42%, Apple 约 31%, Microsoft 约 45%。SK 海力士这一年比 Apple、台积电都赚钱——一家被人质疑了二十年的“存储老二”, 靠 HBM 一举成为全球最赚钱的科技公司之一。

2.7.5 HBM4 与下一阶段的考验

HBM4 是下一代标准, 关键升级: 堆叠层数从 8/12 层升到 12/16 层, 带宽从 1.2TB/s 升到 2.0TB/s 以上。SK 海力士 HBM4 进展 (截至 2026 年初):

- 2025 年 9 月, SK 海力士全球第一个完成 HBM4 量产准备
- 已向 Nvidia 提供大量“付费样品”, Nvidia 验证全部通过
- HBM4 实际量产推迟到 2026 年 Q1 末
- 预计 SK 海力士将供给 Nvidia HBM4 需求的约 2/3
- 清州 M15X 新厂 2026 年 5 月完成第一个洁净室, 开始 HBM3E + HBM4 试产

2.8 Intel 与 SK 海力士的对比: IDM 模式的两种命运

维度	Intel	SK 海力士
主营	CPU + Foundry (混合)	DRAM + HBM + NAND
2025 营收	529 亿美元	约 700 亿美元
2025 净利率	-0.6% (接近持平)	44%
IDM 模式状态	转型中 (IDM + Foundry 混合)	纯 IDM, 且暂时无可替代
资本开支	250+ 亿美元/年	~150-200 亿美元/年
工艺地位	落后台积电 1.5-2 代	DRAM 领先半步
客户结构	自己 90% + IFS 外部少量	高度依赖 Nvidia
核心矛盾	工艺执行 + 代工客户拓展	客户集中 + 三星反扑

从这两家公司可以提炼几个关键启示：

第一，IDM 模式不是死的，关键看“工艺领先”是否还在。Intel 因为 10nm 失误丢了工艺领先，整个 IDM 飞轮就转不动了。SK 海力士因为 HBM 上的工艺领先半步，IDM 飞轮转得飞快。

第二，IDM 模式对“押对技术拐点”特别敏感。SK 海力士 2013 年押 HBM 和 Intel 2014 年押 10nm 激进路线，是同一种 IDM 决策——只有 Fab 自有的 IDM 才能做这种长周期、高资本投入的押注。但押对了回报十倍，押错了代价也是十倍。

第三，地缘政治正在重塑 IDM。Intel 拿了美国政府 89 亿美元 + 软银 + Nvidia 战略投资；SK 海力士拿了美国 CHIPS Act 补贴在印第安纳建厂。IDM 模式独有的“国家安全护城河”是 Fabless 公司拿不到的。

第四，纯 IDM 的“完美形态”可能正在消失。Intel 已经不是纯 IDM；SK 海力士的 HBM4 也开始依赖外部代工厂做 Base Die——纯 IDM 在未来 5-10 年可能只剩模拟和功率半导体。

第三章 Fabless 模式：从“反叛”到 AI 时代的最大赢家

3.1 Fabless 的诞生与定义

Fabless（无晶圆厂设计公司）只做芯片设计，把制造（晶圆代工）、封装测试都外包出去，自己只保留设计、销售和品牌环节。但它的历史含义比定义大得多——它代表了半导体行业从“重资本垂直整合”向“轻资本专业分工”的根本转向。

Fabless 模式的“母体”是 1980s 后期的硅谷。当时 IDM 模式如日中天，新创公司想入局必须先建晶圆厂，资本门槛极高。1987 年台积电（TSMC）成立后开创了纯代工模式，彻底打开了“轻资产做芯片”的可能。第一批纯 Fabless 公司在 1980s 末-1990s 初涌现：Xilinx、Altera、Cirrus Logic、Nvidia（1993）、Broadcom（1991）、Marvell（1995）。

到 2025 年，全球 Fabless 市场规模约 **2709 亿美元**，预计 2030 年达到 5300 亿美元，年复合增长率 14.4%——这增速是整个半导体行业平均增速的 2 倍以上。**Top 10 Fabless 公司 2025 年合计营收 3594 亿美元，同比增长 44%**（主要由 AI 拉动）。

3.2 Fabless 的经济模型

Fabless 模式的财务特征非常鲜明：

- **资本开支极低:** Nvidia 一年 capex 大概 30-50 亿美元, 相比之下台积电一年 capex 就要 400+ 亿美元。
- **研发占比高:** Fabless 公司研发占营收比 20-30% (Nvidia 长期 25%+)。
- **毛利率天花板高:** 成功的 Fabless 公司毛利率能做到 50-75%。Nvidia 数据中心 GPU 毛利率超过 75%。
- **增长弹性大:** Nvidia 2023 财年收入 270 亿, 2026 财年 2160 亿, 三年 8 倍。
- **风险也大:** 完全没有自有产能, 先进制程产能全部押给台积电; EDA 工具完全依赖 Synopsys、Cadence; 地缘政治风险被几何放大。

3.3 Fabless 阵营的五大族群

3.3.1 AI 计算 GPU (Nvidia + AMD)

这是 Fabless 阵营当下最热、增长最猛的细分 (详见第四章 AMD vs Nvidia 专题)。

3.3.2 定制 AI 芯片 (Broadcom + Marvell)

云大厂 (Google、AWS、Meta、Microsoft) 都不想被 Nvidia 敲竹杠, 纷纷做自家定制 AI 芯片: Google TPU、AWS Trainium/Inferentia、Meta MTIA、Microsoft Maia。但它们自己只做架构和算法, ASIC 的物理设计、芯片实现、流片量产都外包给 Broadcom 和 Marvell。

Broadcom: 2025 年从全球第三大 Fabless 跃升到第二, 超过 Qualcomm。Google TPU v5/v6 由 Broadcom 主导。2024 年 12 月市值首次突破 1 万亿美元。

Marvell: AWS Trainium2、Microsoft Maia 都和 Marvell 合作。2025 年定制 ASIC 业务成为最大增长引擎。

3.3.3 移动 SoC (Qualcomm + MediaTek)

Qualcomm: 2025 财年营收 389 亿美元 (同比 +12%), 从全球 Fabless 第二跌到第三。主营骁龙 8 Gen 系列 + 5G 调制解调器。正在尝试两条新增长线: 汽车 (Snapdragon Digital Chassis) 和 PC (Snapdragon X Elite)。苹果 5G 基带自研是 Qualcomm 中长期最大风险。

MediaTek (联发科): 2025 年营收 191 亿美元 (同比 +16%), 冲到全球 Fabless 第五。天玑 9500 旗舰 SoC 性能逼近骁龙 8 Gen 旗舰, 价格更便宜, 吃掉了大量中高端 Android 市场。

3.3.4 Apple 的"内部 Fabless"

Apple 是 Fabless 阵营里最特殊的存在。它不卖芯片给任何人, 但是 TSMC 全球最大客户 (占 TSMC 营收近 25%)。A 系列、M 系列芯片设计能力可能是全球 Fabless 公司里最强的。M5 (2025 年发布) 已经在 5nm/3nm 区间领先 Intel、AMD 同代消费 CPU 性能/功耗。

3.3.5 中国 Fabless 阵营

中国 Fabless 是过去 10 年增长最快的细分市场之一，截至 2025 年中国 Fabless 已占全球 Fabless 市场约 30%。但 2018 年中美科技博弈以来，中国 Fabless 走的路非常特殊。

- **华为海思**：2018 年前是中国 Fabless 王者，2020 年被制裁后从国际市场近乎消失。2023 年起靠 SMIC N+2（约 7nm 工艺）重启：麒麟 9000S（搭载 Mate 60 Pro）、昇腾 910C（AI 训练芯片，采用 chiplet 封装）。
- **寒武纪（Cambricon）**：中国 AI 芯片创业明星。2025 年前三季度营收 46 亿元人民币，是其他中国 AI 芯片创业公司的几倍乃至 10 倍。
- **紫光展锐（UNISOC）**：中国第二大移动 SoC Fabless。2024 年营收 17.8 亿美元。
- **韦尔股份（Will Semi）**：中国 CIS 龙头，全球第三（仅次于索尼、三星）。
- **地平线、黑芝麻**：自动驾驶芯片
- **平头哥**：阿里旗下，玄铁 RISC-V

中国 Fabless 整体的“硬伤”：**先进制程产能受限于 SMIC**（SMIC 自己又被光刻机禁运），这是结构性瓶颈，不是设计能力的问题。

3.4 Fabless 模式的核心矛盾与未来

一、**对台积电的高度依赖正在变成系统性风险**。先进制程几乎只有台积电能稳定量产，全球 Fabless 公司都排队抢产能。SEMI 预测 2026 年 AI 芯片将占用全球先进制程产能的 35%。

二、**EDA 工具的“卡脖子”**。所有 Fabless 公司用的电路设计软件来自 Synopsys、Cadence、Siemens EDA 三家（前两家都是美国）。

三、**超大客户“倒戈做 ASIC”的趋势压缩 Fabless 空间**。Google、AWS、Meta、Microsoft 都在做自家定制芯片。

四、**地缘政治带来的“技术分叉”**。未来 Fabless 公司可能要为不同市场做“分叉版本”——一份给国际市场、一份给中国市场。

第四章 专题：AMD vs Nvidia 的 AI 算力之战

4.1 整体力量对比：先把数字摆出来

维度	Nvidia	AMD
公司总营收（2025/FY2026）	2159 亿美元	约 320-340 亿美元
数据中心营收	约 1700+ 亿美元	约 130-140 亿美元

AI 训练 GPU 市场份额	80-95%	5-10%
离散 GPU 市场份额	92%	~7%
数据中心 GPU 毛利率	75%+	50-55%
市值 (2026 年初)	3.5-4 万亿美元	约 3000-4000 亿美元
当家旗舰	Blackwell B300 / Rubin	MI355X / MI400
软件栈	CUDA (20 年生态)	ROCm (加速追赶)

简单一句话：**Nvidia 在数据中心 AI 上的体量是 AMD 的 12-13 倍**。但 2025 年下半年开始，这种差距的“扩大速度”明显放缓——AMD 拿到了几个量级足够大、能改变叙事的客户合同。

4.2 硬件路线图：两家正在打的“产品节奏战”

4.2.1 Nvidia 的“两年一代+一年一改”节奏

- **Hopper 时代 (2022-2024)**：H100 → H200，单卡 4-5 万美元。
- **Blackwell 时代 (2024-2025)**：B100/B200 → GB200 → B300/GB300。GB200 NVL72 机柜 140 万美元一台。Blackwell 占 Nvidia 数据中心计算业务 70%。
- **Rubin 时代 (2026-2027)**：Rubin R100 (336 亿晶体管、288GB HBM4、50 PFLOPS FP4)，2026 年 Q4 采样、2027 Q1 量产。Vera CPU 88 核 ARM。Vera Rubin NVL144 机柜 H2 2026 出货。
- **Rubin Ultra (2027 H2)**：500 亿晶体管、384GB HBM4E、32TB/s 带宽、100 PFLOPS FP4。
- **Feynman (2028+)**：下一代。

4.2.2 AMD 的“小步快跑+整合突袭”策略

- **MI300 时代 (2023-2025)**：MI300X (192GB HBM3)、MI325X (256GB HBM3E)。2024 年 AI 加速器收入约 50 亿美元。
- **MI350 时代 (2025-2026 H1)**：MI350X / MI355X (CDNA 4, 3nm 工艺, 288GB HBM3E)。单价 25,000 美元，对标 Nvidia B200。
- **MI400 时代 (2026 H2 启动)**：CDNA Next 架构, 2nm 工艺, **432GB HBM4、19.6 TB/s 带宽**。Helios 机柜 72 颗 MI400 + Zen 6 EPYC, 对标 Nvidia GB200/GB300 NVL72。
- **MI500 时代 (2027+)**：与 Rubin Ultra 同时代竞争。

4.3 软件之战：CUDA vs ROCm

讲到 AI 算力竞争，硬件只是表象，真正决定胜负的是软件。Nvidia 最深的护城河叫 **CUDA**——一个 2006 年就开始建的并行计算编程模型 + 生态。

CUDA 的护城河有多深：20 年历史；几乎所有 AI 论文、AI 代码、AI 开源项目都基于 CUDA 写；数百万开发者基础；底层手写 kernel 大量存在于 PyTorch、TensorFlow、Triton、vLLM 等核心 AI 框架；cuDNN、NCCL、cuBLAS 等专用库经过 15 年优化。

ROCm 在 2025-2026 的追赶进展:

- PyTorch 官方一级支持 ROCm (Linux 上和 CUDA 平起平坐)。
- vLLM、SGLang 等推理引擎完全支持 ROCm。
- HIP 转译层自动把 90% 的 CUDA 代码转成 ROCm。

性能差距现状 (2026 年 5 月数据): 训练 (PyTorch + 标准 transformer) CUDA 领先 ROCm 约 10-30%; 批量推理高吞吐场景差距缩到 5-10%; 内存带宽密集型场景 ROCm 有时反超; 自定义 CUDA kernel 场景差距大。

简单说: 通用场景下 ROCm 已经追到“可用且只差 10-30%”, 这是历史性突破。但最前沿的 AI 工作 (新算子、新架构、自定义 kernel) 依然只在 CUDA 跑得快。ROCm 的最大瓶颈不是技术, 是生态惯性。

4.4 系统级之战: 互联、机柜、整机能力

Nvidia 的体系 (封闭、垂直整合): NVLink + NVSwitch (Rubin 时代 1.5 PB/s 带宽) → InfiniBand / Spectrum-X 以太网 → NVL72 / NVL144 机柜 (72-144 颗 GPU 当一台逻辑超级机器) → CUDA + 整套软件栈。客户买 Nvidia 不是买 GPU, 是买“端到端 AI 系统”。

AMD 的体系 (开放、联盟、灵活): UALink (GPU 间开放互联, 对标 NVLink) → Ultra Ethernet (开放替代 InfiniBand) → Helios 机柜 (72 卡, 2026 H2) → ROCm + 开源工具链。AMD 的策略是拉拢所有“不想被 Nvidia 绑定”的玩家, 形成“开放标准 vs 封闭生态”的对抗。

4.5 客户动态: 2025 下半年的“分水岭”

4.5.1 OpenAI - AMD 6GW 战略合作 (2025 年 10 月 6 日)

这是 2025 年 AI 行业最重磅的合同。条款:

- **6 GW 算力部署** (够给 450 万户家庭供电)
- 跨多代 AMD Instinct GPU
- 首批 1GW 部署 MI450 系列, 2026 H2 开始
- **累计硬件营收估值超过 900 亿美元**
- AMD 给 OpenAI 1.6 亿股 AMD 认股权证 (约 10% AMD 股权), 行权价仅 0.01 美元, 挂钩业绩和股价里程碑

信号意义大于交易额: Nvidia 不再是唯一选择。

4.5.2 Meta - AMD 6GW 扩展合作 (2026 年 2 月)

Meta 已经在 Llama 3、Llama 4 推理上广泛部署 MI300X。2026 年扩展到 6GW MI400 部署，**双供应商策略明确**。

4.5.3 Oracle - 5 万颗 MI450 集群 (2026 Q3)

Oracle 成为第一家公开提供 AMD MI450 的超级云客户，单一集群 5 万颗 GPU。

4.5.4 其他云大厂

Microsoft Azure: MI300X 已在 Azure 投产，承诺将提供 MI400 实例，但同时押三条腿 (Nvidia + AMD + Maia ASIC)。整体来看，**所有大客户都在多元化**——没有人想“独家依赖 Nvidia”，但也没有人真的“放弃 Nvidia”。

4.6 财务护城河：差距比想象中大

虽然 AMD 拿到了重磅订单，但财务上 Nvidia 的护城河还在持续扩大：

- Nvidia 一年自由现金流 (约 1500 亿美元) 就比 AMD 整个公司的市值还大
- Nvidia 一年研发投入 200+ 亿美元 vs AMD 70 亿
- Nvidia 一年付给台积电 400+ 亿美元 (估算) vs AMD 100 亿

这种“现金流碾压”意味着 Nvidia 可以用低于成本的价格冲击 AMD、可以一次性垄断台积电先进制程产能、可以同时做 5 条产品线。

4.7 本质判断

短期 (2026-2027)：Nvidia 仍然是绝对主角，份额会从 92% 滑到 75-80%，但绝对收入还在涨。AMD 拿到 15-20%，营收三年 5-10 倍。

中期 (2028-2030)：竞争格局可能从“Nvidia 独大 + AMD 追赶”变成“Nvidia 主导 + AMD 第二 + 定制 ASIC 切走 1/3”。CUDA 生态被部分绕过去，但 CUDA 仍然是默认选择。

长期 (2030+)：取决于 AGI 是否真的到来。如果 AI 算力变成“水电煤”级别基础设施，Nvidia 这种封闭生态可能反而成为劣势。

第五章 Foundry 模式：被一家公司主导的全球市场

5.1 Foundry 模式的本质

1987 年台积电由张忠谋创立，开创了纯代工模式——任何人有了设计就能让 TSMC 帮忙生产。这件事的影响是颠覆性的：降低了入行门槛、改变了产业经济学、重塑了创新节奏。

到今天，全球先进制程（5nm 及以下）几乎只有台积电能稳定量产，整个 AI 时代的算力基础设施都建立在这家公司的几座工厂之上。

5.2 Foundry 模式的极端经济学

- **Capex 极重**：一座先进制程晶圆厂的建设成本，2nm Fab 约 280-350 亿美元，A16 Fab 估计 350-400 亿美元
- **Capex 占营收比畸高**：台积电 Capex 占营收比常年 35-45%
- **折旧侵蚀利润、产能利用率决定盈亏**：Foundry 的核心 KPI 是产能利用率，90%+ 是暴利、80% 盈亏平衡、70% 以下亏损
- **工艺领先 = 定价权 = 全部利润**：Foundry 的利润几乎全部来自最先进的 1-2 个制程节点
- **客户高度集中**：台积电 Top 5 客户占营收 60%，Apple 一家占约 25%

5.3 全球 Foundry 格局：一超多弱

2025 年全球 Foundry 市场总规模约 1700+ 亿美元（Q2 单季度 417 亿美元）。市场份额：

排名	公司	总部	市占率	主营节点
1	TSMC（台积电）	中国台湾	~70%	全节点（A16-90nm+）
2	Samsung Foundry	韩国	~7%	2nm-14nm
3	SMIC（中芯国际）	中国大陆	~5.3%	7nm-90nm+
4	UMC（联电）	中国台湾	~4.2%	22nm-150nm+
5	GlobalFoundries（格芯）	美国/阿联酋	~3.6%	12nm-180nm+
6	Hua Hong（华虹）	中国大陆	~1.5%	28nm-180nm+ + 7nm 试产
7	Intel Foundry	美国	~1-2%	Intel 18A-Intel 7

台积电的市占率在持续上升——从 2020 年 Q1 的 54% 到 2025 年的 70%，每一年都在涨。“赢家通吃”效应越来越明显。

5.4 台积电（TSMC）：半导体行业的“中央银行”

5.4.1 体量

- 2025 年营收约 1100 亿美元（同比增长 35-40%）
- 毛利率 55%+，运营利润率 45%+
- 市值 1.2-1.4 万亿美元（2025-2026 年）
- 全球先进制程产能的 90%+

5.4.2 工艺路线图：从 N2 到 A14

N3 / N3E / N3P (3nm 系列)：2022 年量产，目前是主流先进节点。Apple A18/A19、M4、Nvidia Blackwell、AMD MI355X、高通骁龙 8 Gen 4 都在用。

N2 (2nm)：2025 年 Q4 进入量产，引入 GAAFET（环绕栅极晶体管）。客户阵容：**Apple A20/A20 Pro 独占近一半 N2 产能**，AMD EPYC Venice CCD、高通骁龙 8 Gen 5、MediaTek 旗舰。产能：2025 末 40k WPM → 2026 年 100k WPM → 2027 年 200k WPM。**Intel 缺席 N2 客户名单**。

A16 (1.6nm)：2026 年 H2 量产。引入 Backside Power Delivery（背面供电），性能提升 8-10%。**Nvidia 是 A16 首发客户**（用于 Rubin Ultra 后续版本）。Apple 跳过 A16，直接奔 A14。

A14 (1.4nm)：2028 年量产，Apple 先用。

5.4.3 全球建厂：地缘政治避险

- **美国亚利桑那 Fab**：投资 650 亿美元，三座工厂分阶段。Fab 21 (N4) 已 2024 年量产；Fab 22 (N3) 2027 年量产；Fab 23 (N2) 2028+ 量产。
- **日本熊本 Fab**：投资 200+ 亿美元，与索尼、电装合作。
- **德国德累斯顿 Fab**：投资 100 亿欧元，与英飞凌、博世、恩智浦合作。

最先进的节点 (N2、A16) 依然只在台湾——这是台积电最敏感的“硅盾” (Silicon Shield) 筹码。

5.4.4 护城河

- **PDK (工艺设计套件) 成熟度**：客户拿台积电 PDK 设计芯片几乎不会出错
- **Yield Ramp 速度**：从工艺量产到良率达 80% 通常 6-9 个月
- **EDA / IP 生态**：Synopsys、Cadence、ARM 优先适配台积电
- **客户服务文化**：24/7 工程师驻场
- **先进封装能力**：CoWoS、SoIC、InFO 全球领先

5.5 三星 Foundry：苦苦追赶的“代工业务老二”

2018 年市占率 20% (巅峰) → 2022 年 16% → 2025 年 7.2%——和台积电的差距已达到 62.7 个百分点。

核心问题：制程屡次失败 + 客户信任崩塌。5nm 良率仅 30-40%，3nm GAA 良率长期低于 30-40%。高通、AMD、Nvidia 这些“硬核客户”在 2020-2024 年彻底回到台积电。

2nm (SF2) 救命稻草：2025 年良率 55-60%，目标 2026 年 70%。重大客户突破：

- **Tesla 165 亿美元长期合同**：为 Tesla 自动驾驶芯片 AI5、AI6 代工——Samsung Foundry 历史最大单笔订单
- **Qualcomm 不确定**，2026 年 4 月报道称良率不达标，旗舰订单可能仍在台积电
- **1.4nm 节点已推迟到 2029 年**，基本宣告 2027-2028 年三星追不上台积电 A14。

为什么三星总是追不上：IDM + Foundry 角色冲突（三星在为自己的客户竞争对手做代工）；三星集团专注度不足（200+ 个业务线）；工程师文化不以“客户为中心”。

5.6 中芯国际（SMIC）：被动突围的“中国代工龙头”

- 2025 年营收约 93.27 亿美元（同比 +16.2%），全球第三大 Foundry
- N+1 / N+2（7nm 等效）：2022 年首次量产，支撑华为 Mate 60 Pro 的 Kirin 9000S，2024-2025 年量产昇腾 910C
- 5nm 等效：2025 年完成开发，但良率据报仅 33%，单位成本是台积电同节点的 1.5 倍
- 1.4nm 受限：因为 EUV 光刻机被禁运，往下走的路径基本被锁死

中芯不是为了赚钱而存在的——它的真正使命是给中国 Fabless 公司提供“无论如何都能造芯片”的兜底产能。

华虹半导体：中国第二大代工厂，2025 年市占率约 1.5%。主营成熟节点（55-180nm）。2025 年华虹无锡 12 寸 Fab 开始投产 7nm 节点，意味着中国 7nm 代工不再是“中芯单点”。

中国成熟节点的“产能洪水”：2025-2027 年中国成熟节点（28nm-150nm）将出现产能洪水，多家厂商同时扩产。预计 2027 年中国 28nm 及以上成熟节点产能将占全球 30-35%。

5.7 其他 Foundry 玩家

- **联电（UMC）：**中国台湾第二大代工厂，专注成熟节点。2024 年宣布与 Intel 合作做 12nm 工艺。
- **格芯（GlobalFoundries）：**美国本土最大代工厂之一，2018 年放弃 7nm 研发，专注成熟节点。CHIPS Act 给了 15 亿美元补贴。
- **Intel Foundry（IFS）：**新进入者，2025 年市占率不到 2%。Intel 18A 已量产，微软、AWS 已确认为 18A 客户，但 Q4 2025 单季度仍亏损 25 亿美元。

5.8 Foundry 模式的核心矛盾

矛盾一：地缘政治集中风险。全球 90%+ 的先进制程产能在中国台湾。一旦台海发生地缘冲突，全球 AI、智能手机、汽车产业链将全面瘫痪。各国都在用国家力量推动本土化，但本土化建厂成本高 30-50%。

矛盾二：资本开支的“指数增长”威胁。每代节点 capex 涨 30-50%。到 1nm 以下节点，单座 Fab capex 可能突破 500 亿美元——意味着全球能玩得起最先进节点的可能只剩台积电一家。

矛盾三：客户少、议价权失衡。Apple、Nvidia 通过“包下整代产能”获得了反向议价权。

矛盾四：成熟节点的过剩。中国成熟节点产能洪水将让全球 28nm 及以上节点价格继续下行。

第六章 专题：CoWoS 与先进封装——AI 时代新的核心战场

6.1 为什么“封装”突然变成 AI 时代的核心战场

摩尔定律已经放缓 10 年了。从 28nm（2011）到 3nm（2022）的 11 年间，单位晶体管成本几乎不再下降；从 3nm 到 2nm 这一代，制造成本反而上升 25-40%。

这意味着传统的“靠制程缩小”获取 AI 算力的路径不可持续。真正的算力增长动力，从“造更小的晶体管”变成了“把更多的小芯片连起来”——这就是先进封装。

- 封装从过去半导体产业链的“低附加值环节”变成“核心环节”
- 一颗 Nvidia B200 的最终成本里，封装环节占 25-30%（过去封装只占 5-10%）
- 封装产能在 2024-2025 年成为 AI 芯片最大瓶颈，比 2nm 制程还紧

6.2 先进封装的演进：从 2D 到 3D 的“五代”

- 第一代：2D 封装（传统）——倒装芯片、QFN、BGA 等
- 第二代：2.5D 封装——硅中介层（Silicon Interposer），多个芯片通过中介层互联。代表技术：TSMC CoWoS、Intel EMIB
- 第三代：3D 封装——芯片直接垂直堆叠，TSV 垂直连接。代表技术：TSMC SoIC、Intel Foveros、HBM
- 第四代：混合键合（Hybrid Bonding）——铜对铜直接键合，间距小于 10 μm 。应用：HBM4 时代 DRAM 堆叠、AMD 3D V-Cache、TSMC SoIC
- 第五代：异质集成——把不同工艺、功能的芯片（Logic + Memory + RF + Photonics）集成在一个封装内

6.3 CoWoS 的三个变种

CoWoS（Chip-on-Wafer-on-Substrate）是台积电独家技术，2009 年开发，2012 年商用。它现在是 AI 芯片封装的“事实标准”。CoWoS 不是一个技术，而是一个家族：

6.3.1 CoWoS-S（Silicon Interposer，硅中介层）

最经典的版本，2.5D 封装的代表作。多个芯片放在一块大硅片上，大硅片就是“硅中介层”，里面有 TSV（硅通孔）和高密度走线。中介层最大尺寸 3.3 倍光罩面积（约 2700 mm^2 ）。

应用：H100/H200、B100/B200/B300、AMD MI300X/MI355X、Google TPU 等绝大多数主流 AI 芯片。
 特点：性能最高、成本最高、产能最紧张。

6.3.2 CoWoS-R (RDL Interposer)

用有机重布线层替代硅中介层。成本明显低于 CoWoS-S，互联密度比 CoWoS-S 低，但比传统封装高。应用：对成本敏感、对极致性能不那么敏感的 AI 芯片，比如部分车规芯片、网络芯片。

6.3.3 CoWoS-L (Local Silicon Interconnect, 局部硅互联)

2024-2025 年的“明星”，AI 时代的关键技术。不做整块大硅中介层，而是用“小块硅桥”（LSI bridge）局部连接需要高带宽的芯片。能拼出 3000 mm² 以上的超大封装尺寸。

应用：Nvidia Blackwell B200/B300、Rubin——所有“超大尺寸”AI 芯片都用这个。特点：先进性 + 突破光罩极限 + 比 CoWoS-S 良率高。

维度	CoWoS-S	CoWoS-R	CoWoS-L
中介层	硅	有机 RDL	硅桥 + RDL 混合
最大尺寸	3.3× 光罩	类似	超过 3000 mm ² ，可拼接
互联密度	高	中	最高
成本	高	低	中高
良率	中	高	较高
代表客户	AMD、早期 Nvidia	Broadcom、车规	Nvidia Blackwell/Rubin

6.4 CoWoS 之外的主要竞争者

- **TSMC SoIC**: 3D 封装的代表，比 CoWoS 更先进。AMD 3D V-Cache、AMD MI300（CPU+GPU 混合架构）使用。
- **Intel EMIB + Foveros**: EMIB 把硅桥嵌入有机基板里，Foveros 是 3D 堆叠技术。Intel 自家产品（Sapphire Rapids、Meteor Lake、Panther Lake）大量使用。
- **三星 X-Cube、I-Cube、H-Cube**: 自家方案但客户少。
- **SK 海力士 HBM**: HBM 自身就是 3D 封装的极致体现。

关键事实：TSMC 几乎垄断了顶级 AI 芯片的封装。所有用 CoWoS 的客户都直接和台积电签合同，OSAT 厂只能拿“溢出订单”。

6.5 CoWoS 的产能争夺

- 2024 年底产能：35k 片/月
- 2025 年底：约 70k 片/月
- 2026 年底目标：130k 片/月——三年扩 4 倍
- Nvidia 一家锁定了 60% CoWoS 产能

- Nvidia 2026 年已订下 80-85 万片 CoWoS 晶圆

CoWoS 已经超过先进制程本身，成为 AI 芯片产能的最大瓶颈。即使 2nm 产能够，CoWoS 不够，AI 芯片也造不出来。

6.6 A 股先进封装概念股图谱：到底在炒什么

A 股的“先进封装板块”分四圈，越往内圈越接近“真受益”，越往外圈越偏“主题炒作”。

6.6.1 第一圈：封测三巨头（最直接受益）

公司	代码	主营 / 亮点	真假评分
长电科技	600584.SH	中国 OSAT 老大；2025 年前三季度营收 286 亿元；先进封装收入占比 38%+；据传掌握 CoWoS-S 等效工艺，良率 98.5%	★★★★★
通富微电	002156.SZ	2025 年前三季度净利润 +55.74%；AMD 全球独家先进封装合作伙伴；AI 服务器订单排到 2027	★★★★★
华天科技	002185.SZ	2024 年营收 144.62 亿元 (+28%)；具备 5nm 封装能力；Chiplet 已量产	★★★
甬矽电子	688362.SH	SiP 封装新势力，体量较小	★★

6.6.2 第二圈：先进封装专用设备厂商

公司	代码	主营 / 亮点	真假评分
拓荆科技	688072.SH	国产唯一量产级混合键合（Hybrid Bonding）设备厂商；与长鑫合作 HBM 国产化	★★★★
华海清科	688120.SH	国产 CMP 设备龙头；先进封装 TSV 工艺、晶圆减薄关键设备	★★★★
盛美上海	688082.SH	全球电镀设备份额 7%，国产唯一；CoWoS 工艺所需大量电镀	★★★★
长川科技	300604.SZ	测试设备厂商；先进封装产品测试复杂度更高	★★★
中微公司	688012.SH	等离子刻蚀机；TSV 工艺需深硅刻蚀	★★★

6.6.3 第三圈：材料与载板

公司	代码	主营 / 亮点	真假评分
兴森科技	002436.SZ	高端封装基板（FC-BGA、ABF Substrate）国产替代；AI 芯片需求暴涨	★★★★
深南电路	002916.SZ	国内 PCB 龙头；同时做封装基板	★★★★
雅克科技	002409.SZ	半导体前驱体国产替代龙头；HBM 制造工艺所需	★★★★
联瑞新材	688300.SH	HBM 用 Low- α 球铝（环氧塑封料填料）；国内唯一	★★★★
华海诚科	688535.SH	国产环氧塑封料龙头	★★★
飞凯材料 / 圣泉集团 / 中富电路 / 凯华材料 / 壹石通	—	“沾边股”，业务里有先进封装相关材料/PCB 板	★★

6.6.4 第四圈：HBM 国产化主线（间接相关）

- 长鑫存储（未上市）—— 中国 DRAM IDM，目标 HBM 国产化
- 长江存储（未上市）—— 中国 NAND IDM
- 澜起科技（688008.SH）—— DDR5 内存接口芯片，HBM 配套

6.6.5 投资角度的几个判断

长期价值最硬的三家（基本面支撑明确）：长电科技（CoWoS-S 等效工艺 + 头部客户）、通富微电（AMD 独家伙伴 + AI 服务器订单到 2027）、兴森科技（FC-BGA 国产替代 + 量价齐升）。

弹性最大但风险也大（主题催化时涨幅最猛）：拓荆科技、华海清科（设备龙头但估值贵）、雅克科技、联瑞新材（材料细分龙头）。

纯主题跟风：各种小市值“沾边股”。

风险提示：A 股先进封装板块整体估值已经偏高（2024-2025 年 PE 普遍 50-80 倍），炒作充分。需要注意 CoWoS 产能 2026 年扩产 4 倍后 2027 年是否会过剩、国内 OSAT 拿到的“溢出订单”实际占比、AMD/Nvidia AI 投资周期是否会出现拐点等周期性风险。

第七章 OSAT 模式：从“低附加值代工”到 AI 时代新战场

7.1 OSAT 的定义与历史地位

OSAT（Outsourced Semiconductor Assembly and Test）字面意思是“外包的半导体封装与测试”。它把晶圆代工厂送出来的晶圆，做成消费者能用的“芯片成品”，包括：晶圆切割、贴装、键合、塑封、测试、打标和发货——是 IC 制造里“最后一公里”的服务商。

1990s-2010s 它一直是半导体产业链的“低端环节”——技术门槛低、毛利率薄、价格战激烈。整个 OSAT 行业的毛利率长期在 15-25% 之间。但从 2020 年起，三件事改变了 OSAT 的命运：

- 摩尔定律放缓——业界转向 chiplet + 先进封装，封装从“低附加值”变成“决定性能的关键”
- AI 算力爆发——HBM、CoWoS、3D 堆叠需求暴涨
- 地缘政治重塑供应链——OSAT 成为 CHIPS Act 资金的重要受益方

到 2024-2025 年，**头部 OSAT 公司毛利率已经回到 30%+**，先进封装产品毛利率甚至冲到 40%+。

7.2 全球 OSAT 格局：Top 10 占行业 80%

排名	公司	总部	2024 营收	YoY	关键定位
1	ASE Technology (日月)	中国台湾	185.4 亿美元	+13%	OSAT 全球老大，份额 45%

	光)				
2	Amkor Technology	美国/韩国	63.2 亿美元	-2.8%	西方阵营主力
3	长电科技 (JCET)	中国大陆	50 亿美元	+19.3%	中国老大, 全球第三
4	通富微电 (Tongfu)	中国大陆	33.2 亿美元	+5.6%	AMD 独家伙伴
5	力成科技 (PTI)	中国台湾	22.8 亿美元	+1%	内存封测专家
6	京元电子 (KYEC)	中国台湾	~20 亿美元	+25%	测试代工龙头
7	华天科技	中国大陆	~20 亿美元	+28%	中国老三

地理高度集中——中国台湾 + 中国大陆两地 OSAT 占了全球 80%+ 份额。**中国大陆 OSAT 增速最猛**——长电 +19.3%、华天 +28%，中国大陆 OSAT 在全球 Top 10 份额从 22% 提升到 26%。

7.3 ASE Technology (日月光)：OSAT 王者的 AI 时刻

- 1984 年成立于中国台湾高雄，2018 年完成与矽品 (SPIL) 合并。**全球 OSAT 第一**，市占率约 30%
- 2024 年营收 185.4 亿美元；全球员工约 8 万人；6 个国家工厂
- 客户：高通、苹果、AMD、英伟达 (部分)、博通、Marvell、联发科

AI 时代的“次优解”：台积电吃掉了顶级 AI 芯片的封装 (CoWoS-S/L)，ASE 负责吃“溢出订单”和“次顶级订单”。

关键战略动作：

- 先进封装收入：2024 年 11 亿美元 → 2025 年目标 16 亿美元 → 2026 年目标 35 亿美元
- **FOPLP (Fan-Out Panel-Level Packaging) 押注**：投资 2 亿美元建第一条 600×600mm 大尺寸面板级扇出封装产线
- **CoWoP (自研 CoWoS 替代方案)**：产能目标 20,000-25,000 片/月 (2025 年底) ——是 TSMC CoWoS 产能的约 1/4
- 2026 年创纪录 capex 70 亿美元——OSAT 行业历史最大单一年 capex

7.4 Amkor Technology：OSAT 第二、西方阵营的“独苗”

- 1968 年成立于美国，总部亚利桑那州坦佩。早期由韩国资本家 James Kim 控股
 - 2024 年营收 63.2 亿美元 (同比 -2.8%，主要因汽车低迷)；Q4 2025 销售 18.9 亿美元，超预期
 - 全球工厂：韩国 (光州、平泽)、菲律宾、马来西亚 (槟城)、葡萄牙、台湾、越南
- 业务结构升级**：2025 年 Q4，先进产品占总营收 80%+，已基本退出低端封装市场。

亚利桑那 70 亿美元投资——OSAT 行业历史最大单笔：

- 地点：美国亚利桑那州 Peoria；2025 年 10 月动工
- 总投资 70 亿美元，分两期；2027 年中第一期完工，2028 年初投产
- Apple 是亚利桑那项目“第一个、也是最大的客户”，Nvidia 也是核心客户

- 完工后将与台积电亚利桑那 Fab 形成本土“前道+后道”完整链条

7.5 中国 OSAT 三巨头：集体崛起

长电科技、通富微电、华天科技三家合计 2024 年营收约 103 亿美元，全球 OSAT 份额约 25%——比 2018 年的 15% 大幅提升。

崛起三大动力：

- “国产替代”国家战略推动
- AMD 这种特殊客户绑定（通富微电独家）
- 资本投入和产能扩张

中国 OSAT 的瓶颈：高端订单（Nvidia Blackwell、Apple M5）几乎都在 TSMC 内部完成；先进设备（混合键合机、TSV 设备等）国产替代率不到 30%。

7.6 OSAT 与 Foundry 的“封装地盘之战”

TSMC 现在的策略是“客户在我这做晶圆，就要在我这做封装”——CoWoS、SoIC、InFO 全部 TSMC 内部完成。这种“前道+后道一体化”挤压 OSAT 的高端机会。

OSAT 阵营选择三条次优路径：

- 路径 1：吃台积电“溢出产能”——CoWoS 长期供不应求，溢出客户找 OSAT
- 路径 2：做“次顶级”AI 芯片封装——FOWLP / FOPLP / SiP 中端方案
- 路径 3：做 TSMC 不做的“特殊封装”——车规级、CIS、MEMS、晶圆级 CSP 等

一个有趣的迹象：TSMC 开始“放手”。2025 年底业界传出，TSMC 计划从 2026-2027 年起把部分 CoW（Chip-on-Wafer）订单释放给 OSAT 合作伙伴——如果走通，ASE、Amkor、长电、通富都将分到 AI 蛋糕。

7.7 OSAT 模式的核心矛盾

- 矛盾一：上游被 Foundry 挤压，下游被 PCB 厂渗透——TSMC 从上游切入封装高端，PCB 厂从下游切入基板和 RDL
- 矛盾二：客户集中度太高——Top 10 客户占大型 OSAT 营收 60%+
- 矛盾三：地缘政治“双重要求”——OSAT 必须全球分布式建厂，资本压力非常大
- 矛盾四：先进封装设备“卡脖子”——混合键合机、TSV 设备等核心设备的供应商集中度极高
- 矛盾五：AI 周期的“高 beta”风险——如果 AI 投资周期出现拐点，OSAT 行业会瞬间从“产能不足”变成“产能过剩”

第八章 总结与展望

8.1 四种模式的整体回顾

模式	核心定位	代表企业	优势	劣势	当下状态
IDM	设计+制造+封测全包	Intel、三星、SK 海力士、TI、Infineon	工艺协同、技术深度	资本极重、灵活性差	在向 IDM 2.0 / Fab-Lite 演化
Fabless	只做设计	Nvidia、Apple、AMD、高通、Broadcom	轻资产、迭代快、毛利高	依赖 Foundry 产能	AI 时代最大赢家
Foundry	只做制造	TSMC、三星、SMIC、Intel Foundry	规模效应、专业化	资本极重、客户依赖	极度集中 (TSMC 70%)
OSAT	只做封测	ASE、Amkor、长电、通富	灵活性、地缘适应	利润薄、被两端挤压	AI 推动结构升级中

8.2 四种模式的演变趋势

第一，纯模式的边界在模糊。Intel 同时做 IDM 和 Foundry；TSMC 同时做 Foundry 和封装；OSAT 在向“准 Foundry”延伸。未来可能不再有“纯 Fabless”或“纯 Foundry”，混合模式会成为主流。

第二，AI 是当下最大共同因子。无论是哪种模式，AI 算力的爆发都是最重要的增长动力。Nvidia (Fabless)、TSMC (Foundry)、SK 海力士 (IDM)、Amkor (OSAT) 都是 AI 时代的赢家。

第三，地缘政治在重塑产业地图。美国、中国、欧洲、日本都在用国家力量重塑半导体产业链。“全球协作”模式正在被“区域多元化+本土化”模式替代——这会让所有模式的资本开支都升高、效率短期下降。

第四，制程红利衰减、封装红利兴起。摩尔定律放缓后，封装成为“最后的性能增长来源”。这是 OSAT 这种长期被忽视的模式突然成为热门赛道的根本原因。

8.3 关键看点 (2026-2028 年)

未来 2-3 年最值得跟踪的几个核心变量：

- Intel 18A 良率与外部客户拓展——决定纯 IDM 模式能否在逻辑芯片领域复活
- 三星 SF2 良率能否突破 70%——决定 Foundry 阵营是“一超”还是“双强”
- SK 海力士 vs 三星 HBM4 之战——决定存储 IDM 利润分配格局
- Nvidia Rubin vs AMD MI400 量产进度——决定 AI GPU 阵营格局
- OpenAI - AMD 6GW 部署的实际表现——决定 ROCm 生态是否真正立住

- 台积电 N2 / A16 量产爬坡曲线——决定 Apple、AMD、Nvidia 的客户依赖度
- CoWoS 扩产能否跟上 AI 需求——决定 2026-2028 年 AI 芯片供应
- TSMC 是否真的把 CoW 订单释放给 OSAT——决定 OSAT 阵营的“AI 直接受益度”
- 国产 EUV 突破时间表——决定中国先进制程的天花板
- 台海地缘风险——黑天鹅风险，但概率不为零

8.4 本质判断

当我们把四种模式放在一起整体审视，可以提炼几个跨模式的本质判断：

第一，半导体产业链的“权力中心”正在持续漂移。从 1980s 的 IDM (Intel、TI 时代)，到 1990s-2000s 的 Fabless 崛起 (Nvidia、高通)，到 2010s 的 Foundry 主导 (TSMC 一统江湖)，再到 2020s 的“封装与存储”重新成为价值核心 (CoWoS、HBM)。每一次权力转移，都会重塑整个产业的财富分布。

第二，AI 算力是当下产业链最大的“放大器”。它放大了所有头部公司的优势，也放大了所有掉队公司的劣势。Nvidia 的 92% GPU 份额、TSMC 的 70% Foundry 份额、SK 海力士的 62% HBM 份额——这些“赢家通吃”的极端集中度，在历史上从未出现过。

第三，地缘政治是产业链未来 10 年最不可忽视的变量。美国 CHIPS Act、欧盟 Chips Act、中国大基金、日本 Rapidus——各国都在用国家力量重塑半导体地图。这种“半导体国家化”会带来产业效率的下降，但同时也带来新的本地化机会。

第四，中国半导体产业正在走“差异化突围”。在先进制程被卡脖子的情况下，中国在成熟节点扩产、先进封装、HBM 国产化、设备和材料国产替代等多条战线同时展开。这是一场需要 10-15 年才能见分晓的长期战争。

第五，下一个 5 年的关键问题是：当 AI 算力商品化 (commoditization) 发生时，谁能从“卖芯片”转型为“卖系统、卖服务、卖软件”——这才是终极胜负手。Nvidia 已经在做 (DGX Cloud、AI Foundry 服务)；AMD 在追 (收购 ZT Systems 做整机)；TSMC 在尝试 (封装 + 制造一体化)；SK 海力士在尝试 (HBM + 系统集成)。这场转型的成功与否，将决定半导体行业下一个十年的格局。

附录：主要数据来源

A.1 机构与公司财报

- TSMC、Samsung、SK Hynix、Micron、Intel、AMD、Nvidia、Qualcomm、MediaTek、Broadcom、ASE、Amkor 等公司官方季度财报与年报
- IDC、Counterpoint Research、TrendForce、Gartner 等机构 2025-2026 行业报告
- SEMI、WSTS、SIA 等半导体行业协会数据

A.2 核心媒体与分析平台

- Tom's Hardware、TrendForce、DIGITIMES、AnandTech 等技术媒体的产品与产能分析
- Seeking Alpha、Yahoo Finance 财务数据
- Semiwiki、Semiengineering 等半导体专业平台

A.3 中国 A 股市场数据

- 中信证券、华西证券、开源证券、中泰证券等 A 股研究报告
- 东方财富、证券之星、九方智投等 A 股资讯平台

A.4 时间口径

本报告主要数据口径为 2024-2026 年。其中 2024 年数据为完整年度数据；2025 年数据多为前三季度或全年初步财报；2026 年数据为 Q1 实际数据加全年预测。所有市场份额、产能数据均为公开报道与行业机构测算综合结果，特定数据可能存在小幅差异。

A.5 免责声明

本报告所涉及的市场份额、财务数据、产品规格、企业战略等信息均来自公开渠道，仅供研究参考。报告中关于 A 股个股的“真假评分”为基于公开信息的定性判断，不构成任何投资建议。半导体行业波动剧烈、地缘政治变量众多，读者应结合自身情况独立判断。