



硬件是 Token 经济学下核心投资方向

—— Token 经济研究系列之电子行业篇

2026 年 05 月 20 日

核心观点

- Token 成核心生产要素，数据中心化身 Token “工厂”**：Token 是 AI 大模型处理信息的最小单元，可精准计量算力、显存与电力消耗，作为 AI 算力的价值载体，堪称 AI 时代“水电煤”与核心生产要素，推动数据中心成为 Token 规模化生产的智能工厂。据此我们构建了完整的 Token 经济学分析框架，覆盖供给、需求、定价与资源配置四大环节。以 Token 为核心的算力基建决定 AI 时代经济新形态，上游硬件作为“卖铲人”，决定 Token 供给效率、成本上限与质量，并约束中游算力规模；中游算力将硬件性能转化为标准化算力供给，支撑下游 Token 应用爆发，形成“上游硬件约束、中游算力承载、下游需求牵引”的闭环产业逻辑。
- Token 生成速度与硬件强相关，硬件产业迎发展机遇**：Token 生成速度与硬件性能高度相关，硬件成为 Token 经济增长的核心驱动力。全球日均 Token 消耗量已突破 360 万亿，中国市场达 180 万亿，带动 AI 硬件需求全面爆发。2026 年全球八大云服务商资本开支预计同比增长 24% 至 5200 亿美元以上，AI 服务器出货量持续攀升；AI 服务器对 DRAM、NAND 及 HBM 需求数倍增长，推动存储芯片价格上行，全球算力芯片形成“一超多强”格局。中国 AI 芯片市场规模预计 2026 年突破 1600 亿元，国产替代在政策与供应链驱动下加速推进，硬件产业链迎来高确定性发展机遇。
- 电子器件是 Token 经济的核心骨架**：AI 驱动下数据中心向 Token 超级工厂转型，带动 PCB、被动元件等核心元器件全面升级。2025 年全球 PCB 市场规模达 849 亿美元，同比增长 15.4%，其中高多层板与 HDI 板增速显著，中国以 19.2% 的增速成为全球增长最快市场，并持续稳固全球 PCB 制造中心地位；AI 服务器对高端 MLCC、芯片电感、钽电容需求呈指数级增长，单机 MLCC 用量达传统服务器的 2-3 倍，一体成型电感用量大幅提升，叠加液冷普及与高速信号传输需求，推动电子元器件向高端化、定制化演进，国产厂商在高端市场份额持续提升，共同支撑 Token 生成效率提升与成本下行。
- 投资建议**：硬件产业是 Token 经济学下的核心投资方向，我们看好半导体和电子器件方向，推荐兆易创新、中芯国际、生益科技、澜起科技、水晶光电。
- 风险提示**：AI 技术迭代不及预期；硬件产能扩张速度低于需求；国产芯片替代进度不及预期；全球 AI 算力投资增速放缓。

重点公司盈利预测与估值

股票代码	股票名称	EPS			PE			投资评级
		2025A	2026E	2027E	2025A	2026E	2027E	
688981.SH	中芯国际	0.63	0.82	1.19	212.40	163.18	112.45	推荐
603986.SH	兆易创新	2.47	4.24	5.37	180.17	104.96	82.87	推荐
600183.SH	生益科技	1.37	2.32	3.25	73.40	43.34	30.94	推荐
002273.SZ	水晶光电	0.84	1.03	1.29	47.26	38.54	30.78	推荐
688008.SH	澜起科技	1.95	2.50	3.20	131.95	102.92	80.41	推荐

资料来源：、中国银河证券研究院。备注：以 2026 年 5 月 20 日收盘价为基准计算。

电子行业

推荐 维持评级

分析师

高峰

☎：010-80927671

✉：gaofeng_yj@chinastock.com.cn

分析师登记编码：S0130522040001

王子路

☎：010-80927632

✉：wangzilu_yj@chinastock.com.cn

分析师登记编码：S0130522050001

钱德胜

☎：021-20252665

✉：qiandesheng_yj@chinastock.com.cn

分析师登记编码：S0130521070001

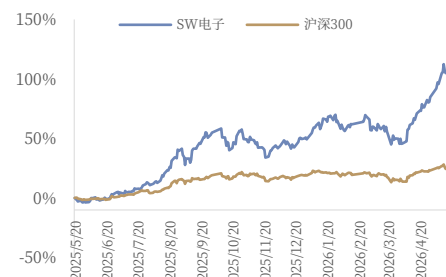
刘来珍

☎：021-20252647

✉：liulaizhen_yj@chinastock.com.cn

分析师登记编码：S0130523040001

相对沪深 300 表现图 2026 年 05 月 19 日



资料来源：中国银河证券研究院

相关研究

- 【银河电子】半导体行情周报_板块行情向增量、瓶颈、周期反转环节扩散
- 【银河电子】半导体行情周报_跟随海外大涨，存储板块最强
- 【银河电子】半导体行情周报_板块持续上行，半导体设备表现突出

目录

Catalog

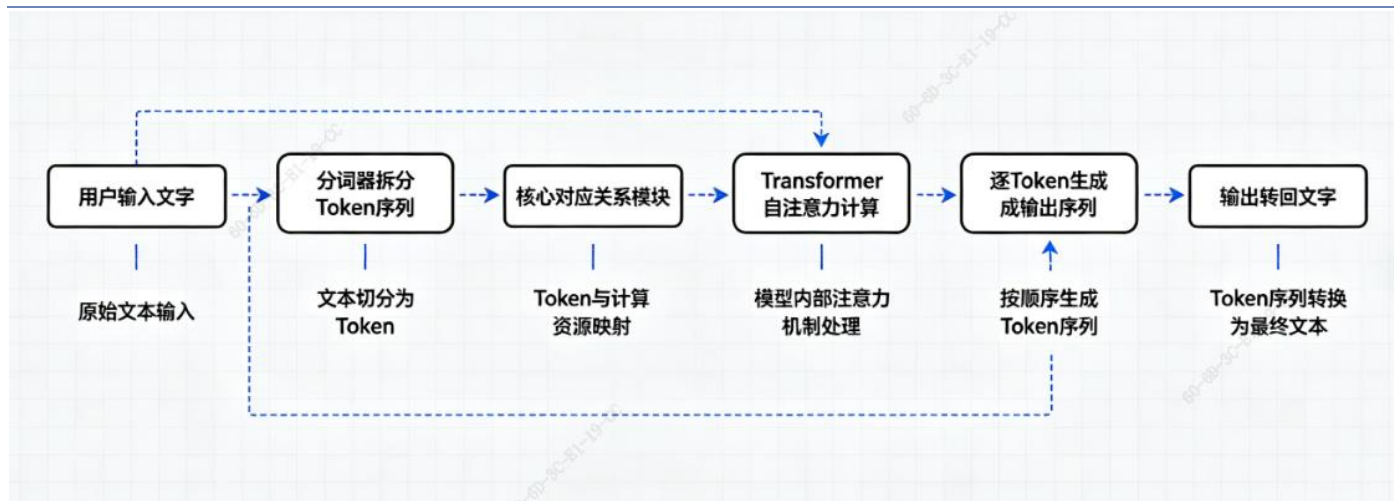
一、Token 成核心要素，数据中心化身 Token 工厂	3
(一) Token 是什么?	3
(二) Token 经济学框架	3
(三) Token 需求端爆发式增长	4
(四) 上游硬件是 Token 生产的“物理底座”	5
二、Token 速度与硬件强相关，硬件产业迎发展机遇	7
(一) 硬件决定 Token 生成速率，全产业链受益明显	7
(二) Token 生成需求爆发，对上游硬件需求持续提升	11
(三) “一超多强”算力芯片格局确定，关注国产芯片替代机遇	17
三、电子器件是 Token 经济的核心骨架	20
(一) AI 驱动电子元器件升级迭代	20
(二) PCB: Token 时代 PCB 重要性提升，中国制造中心地位稳固	22
(三) 被动元件：高容、芯片电感和钽电容需求有望加速增长	25
四、投资建议	27
五、风险提示	28

一、Token 成核心要素，数据中心化身 Token 工厂

（一）Token 是什么？

Token（词元）是 AI 大模型处理文本、语音、图像等各类信息的**最小离散处理单元**，其产生原理是：先通过分词器（Tokenizer）借助 BPE 等子词拆分算法，将文本切分为字、词、标点或子词片段并编码为唯一数字 ID，完成从自然语言到模型可识别序列的转换。这些 Token ID 会进一步被送入 Transformer 架构，通过嵌入层转化为带有位置信息的向量，让模型既能识别语义又能理解语序。Transformer 利用自注意力机制对所有 Token 进行全局关联计算，捕捉序列间的依赖关系，从而理解完整上下文含义。模型在生成内容时，同样以 Token 为基本单位进行预测输出，文本类结果会将 Token 还原为文字，多模态场景下还可进一步解码为图像、音频、代码等形式，最终拼接组合成人类可理解的连贯内容。

图1: Token 的产生机制



资料来源：中国银河证券研究院

（二）Token 经济学框架

从以上 Token 的产生机制来看，它是大模型输入输出的基本计量单位，能够精准对应模型训练和推理过程中的真实算力、显存与电力消耗。每一次交互产生的 Token 数量，本质上都是 AI 资源投入的量化体现，这使得 Token 天然具备了价值衡量的基础。正因为 Token 与实际资源消耗高度绑定，它才具备了成为价值载体的核心条件。

Token 由此可以作为 AI 算力经济的价值载体，在 AI 生态中承担类似实体经济里“货币”的角色，实现算力服务的计量、交换与价值流转。它不再只是技术层面的统计单元，而是串联起算力供给、模型服务与终端需求的关键纽带，让原本抽象的 AI 算力变成可度量、可定价、可分配的标准化合约要素。基于这一核心定位，我们可以为 AI 时代构建一套完整的 Token 经济学分析框架。

该框架以大模型输入输出 Token 为核心，系统搭建起 AI 算力经济的经济学体系。它从供给端阐释 Token 的生成逻辑、成本结构与算力约束条件，从需求端分析 Token 的使用场景、价值来源与各类主体的使用需求。依托供需关系形成合理的定价机制，实现 Token 价值与成本的匹配，并通过资源配置机制优化分配效率，减少无效消耗与算力浪费，最终形成供给、需求、定价、配置协同运转的高效 AI 经济运行模式。

图2: AI 时代 Token 经济学框架

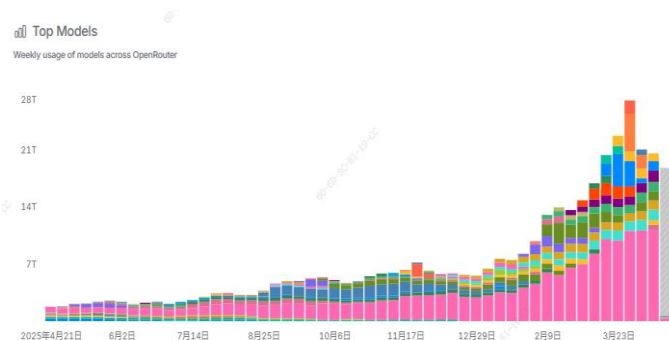


资料来源: 中国银河证券研究院

(三) Token 需求端爆发式增长

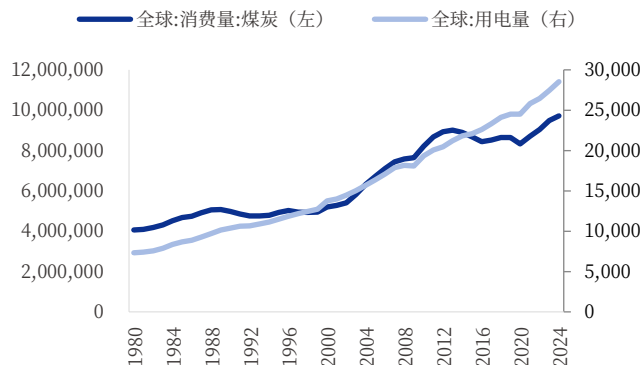
根据 OpenRouter 的统计, 全球大模型 Token 调用量从 2024 年 9 月约 0.25T 增长至目前日均约 20T (2026 年 4 月 13 日 20.6T), 年化增长率约 1144%。**指数级增长的背后, 是 AI 模型训练、推理、交互等全场景对 Token 资源的高度依赖。**Token 在 AI 体系中承担着价值计量、资源分配与服务结算的核心功能, 支撑着整个 AI 生态运转。Token 如同工业时代的电力、水源与燃气一样, 成为 AI 运行不可或缺的基础能源与生产资料。正是这种刚性需求与指数级增长, 让 Token 真正成为驱动 AI 时代发展的核心“水电煤”。

图3: 全球大模型 Token 调用量



资料来源: Open Router, 中国银河证券研究院

图4: 全球煤炭消费量 (千短吨)、用电量 (十亿千瓦时)



资料来源: EIA, 中国银河证券研究院

(四) 上游硬件是 Token 生产的“物理底座”

在 AI Token 经济的产业分工体系中，上游硬件属于典型的基础性生产要素与核心资本品，具备技术壁垒高、产能周期长、供给刚性强的特征，是 AI 产业发展进程中确定性极强的“卖铲人”环节。从经济学生产函数视角来看，上游芯片、存储、互联及配套硬件的工艺水平、架构设计与良率产能，直接决定 Token 生成环节的要素投入效率、边际成本上限与标准化供给质量，构成整个 Token 市场的底层技术约束与成本约束；同时上游硬件的迭代速率、产能释放节奏与供应链垄断格局，会形成强外溢效应，自上而下锁定中游算力行业的产能扩张空间、单位算力成本曲线与供给弹性边界，成为制约中游算力规模能否持续扩容的前置核心变量。

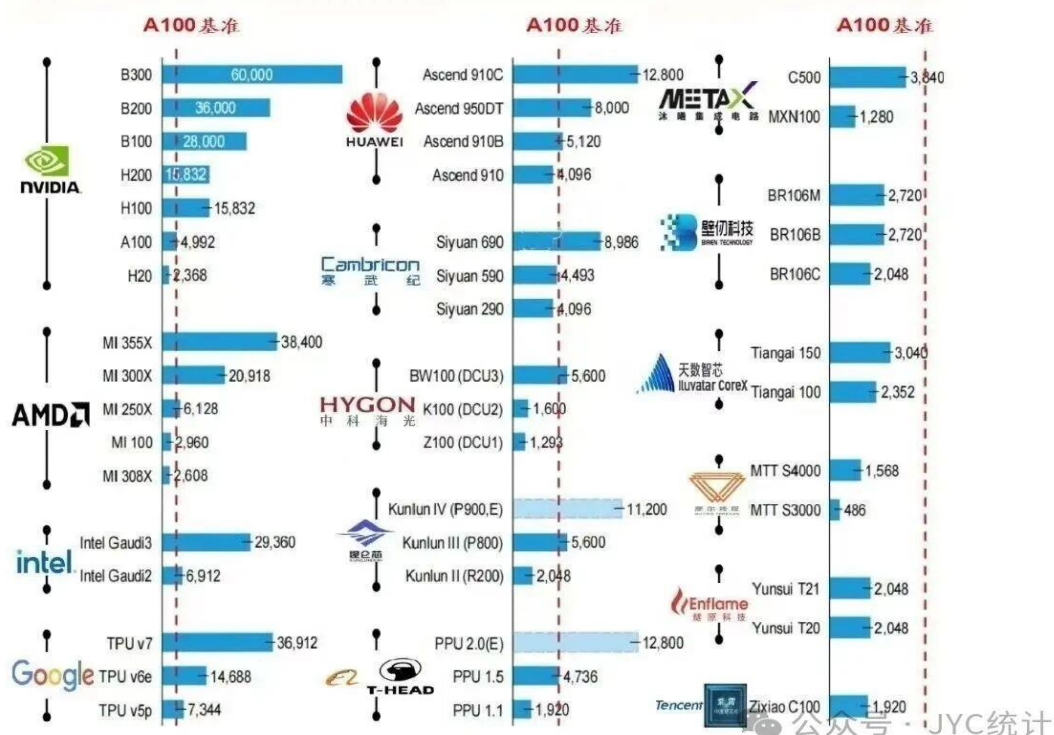
图5：不同模型生成 Token 的成本对比



资料来源：Open Router，中国银河证券研究院

中游算力市场是上游硬件价值落地的中间转化环节，作为典型的产业中间品供给端，承担着将底层硬件性能转化为可调度、可商品化算力资源的枢纽功能。下游大模型推理、生成式 AI 应用、智能服务生态的规模化落地，本质是市场化 Token 服务的终端需求持续扩容，而这一需求的释放高度依赖中游算力形成足量、低价、稳定、可弹性调度的供给底座；由此构建起完整的产业传导逻辑：上游硬件决定算力供给的底层约束，中游算力完成资源整合与规模化供给匹配，下游应用端以 Token 需求反哺拉动中上游资本开支与产能扩张，形成“硬件约束-算力承载-需求牵引”的正向循环，也意味着上游硬件的技术突破与产能释放，是打破算力供需瓶颈、激活下游 Token 商业价值、推动全产业链实现供需均衡与长期价值增值的关键先决条件。

图6：当前主流 AI 芯片算力对比图



资料来源: Bernstein, 中国银河证券研究院

二、Token 速度与硬件强相关，硬件产业迎发展机遇

（一）硬件决定 Token 生成速率，全产业链受益明显

2026 年，全球人工智能产业正式确立了以“Token（词元）”为核心计量单位的经济体系。随着生成式 AI 从简单的聊天机器人向 AI Agent 演进, AI 产业链的价值分配逻辑发生了根本性变化。过去，产业关注的是“训练出多大的模型”；现在，产业关注的是“在单位能耗下能生成多少有价值的 Token”。根据华为在 2026 年中国合作伙伴大会上的披露，当前全球日均 Token 消耗量已突破 360 万亿，而中国市场的日均消耗量也跃升至 180 万亿，且平均每 1.5 天就有一个新模型诞生。这种爆发式的增长使得 AI 基础设施不再仅仅是数据存储与处理的中心，而是转变为生产“智能”的工厂。

图7：国内外代表性大模型每百万 Token 价格

厂商	模型名称	输入价格 (M tokens)	输出价格 (M tokens)
月之暗面	Kimi K2.5	¥4	¥21
智谱AI	GLM-5	¥4-¥6	¥18-¥22
字节跳动	豆包2.0 Lite	¥3.2-¥9.6	¥16-¥48
阿里巴巴	Qwen3.5-27B	\$0.3	\$2.4
DeepSeek	DeepSeek-R1	\$0.55	\$2.19
OpenAI	GPT-5.4	\$2.50	\$15.00
Anthropic	Claude Opus 4.5	\$5-\$10	\$25-\$37.5
Google	Gemini 3.1 Pro	\$2-\$4	\$12-\$18

电力是词元生成的核心成本支撑，电费占算力运营成本比重高，全球电价差异直接塑造词元价格格局。中国依托西部低成本绿电、完善电力基建与特高压调度优势，形成显著电力成本壁垒。国内主流大模型依托成本优势，实现分级定价、高性价比供给，在全球算力通胀环境中保持竞争力，为词元规模化应用与出海奠定坚实基础。电力与词元深度耦合，既是产业发展的关键约束，也是中国词元经济的核心竞争支点。

国内	区域/主体	电力成本	海外	区域/主体	电力成本
	中国西部	0.2-0.3 元/每度		美国数据中心	0.86-1.09元/每度
	中国东部	0.6-0.8 元/每度		欧洲主要国家	1.15-1.50元/每度

资料来源：iMedia Research，中国银河证券研究院

在这一背景下，AI 产业链呈现出清晰的三层金字塔结构，每一层都承担着将“电力”转化为“智能”的不同职责，整个产业链分为三层，上游来看包括 AI 芯片和信号传输芯片为硬件支撑，这是 Token 经济的物理基础，决定了算力的“天花板”。包括半导体材料、制造设备、AI 芯片（GPU/NPU/LPU）、高速存储（HBM）以及各类电子元器件。这一层直接决定了 Token 的生成速率（TPS）和能效比。中游产业链包括算力整合，这是连接硬件与应用的枢纽。主要包括智算中心、算力租赁平台、云服务提供商以及算力调度软件。中游负责将上游的硬件芯片集成为服务器集群，并通过虚拟化技术将算力转化为标准化的服务；下游为应用变现厂商，这是 Token 的最终消费端。包括大模型应用、AI 智能体（Agent）、智能驾驶、智慧医疗等场景。下游的繁荣直接拉动了对上游硬件的需求，形成“下游调用量→中游扩容→上游缺货”的传导链条。

图8: Token 产业链情况



资料来源：致网科技，中国银河证券研究院

Token 生产与发电厂类似，核心是算力基座，由计算芯片、服务器与网络、智算中心三部分支撑。计算芯片是 Token 生产效率的决定者，正从通用 GPU 向专用 NPU 演进；服务器与网络是算力的物理载体，通过 NVLink、RoCE 高速互联技术实现万卡集群协同；智算中心则是算力的发电厂，提供大规模标准化算力租赁与托管服务。此外，图下方总结了该领域重资产与高壁垒的商业特征，体现在三方面：一是高准入门槛，先进工艺制程稀缺，芯片研发与流片成本极高；二是规模效应，智算中心规模越大，单位 Token 生产成本越低；三是强能源依赖，算力本质是对电力的消耗，低能耗技术成为核心竞争力。整体来看，上游算力基建是 Token 生产的底层支撑，其技术迭代与商业特性共同决定了 Token 供应的效率与成本结构。

图9: Token 生产上游产业链情况



资料来源：致网科技，中国银河证券研究院

产业链中游为 Token 的加工厂，将上游的算力资源转化为可商用的语义能力。这一层级主要由三类厂商构成，它们分工明确又相互协作：首先是拥有核心技术的大模型厂商，它们直接定义 Token 的质量标准，通过自研顶尖算法确立智能水平的上限；其次是云厂商，它们凭借强大的云生态和工具链，提供一站式开发平台，降低了企业调用和使用模型的门槛；最后是电信运营商，它们发挥网络与信任优势，重点解决跨区域数据传输的时延问题以及政企客户的合规准入难题。这三类角色共同完成了从原始算力到智能服务的转化，将无形的算力封装成通过 API、私有化部署等形式对外输出的产品。因此，中游模型层是整个产业链的价值放大器，它决定了 Token 最终的服务形态、应用便捷性以及商业化落地的广度，是连接底层硬件与上层应用的关键桥梁。

图10: Token 产业链中游参与方



资料来源：致网科技，中国银河证券研究院

产业链下游为 Token 消费，并将 AI 能力落地于具体业务场景，推动各行业数字化转型与智能化升级，场景包括智能办公实现智能写作、代码生成、数据分析及会议纪要自动化处理，提升办公效率；金融服务通过智能投顾、风险控制模型、反欺诈监测与智能客服机器人优化业务流程；医疗健康借助医学影像分析、药物研发加速、辅助诊断与个性化健康管理提升诊疗水平；教育培训支持个性化学习路径、智能辅导答疑、教育评估与教学内容创作，赋能因材施教；智能制造聚焦生产流程优化、智能质量检测、预测性设备维护与供应链管理，提高工业效能；内容创作则覆盖营销文案生成、视频脚本撰写、游戏剧情设计与数字艺术创作，释放创意生产力。底部强调核心升级趋势：Token 正迅速渗透各行各业，成为数字化转型与智能化升级的关键驱动力，不仅提升现有业务效率，更催生全新商业模式与用户体验。下游用户通过消耗 Token 获取 AI 服务，实现降本增效、创新突破与体验升级，标志着 AI 从技术概念走向规模化产业落地，形成完整的价值闭环。

图11: Token 产业链下游消费场景

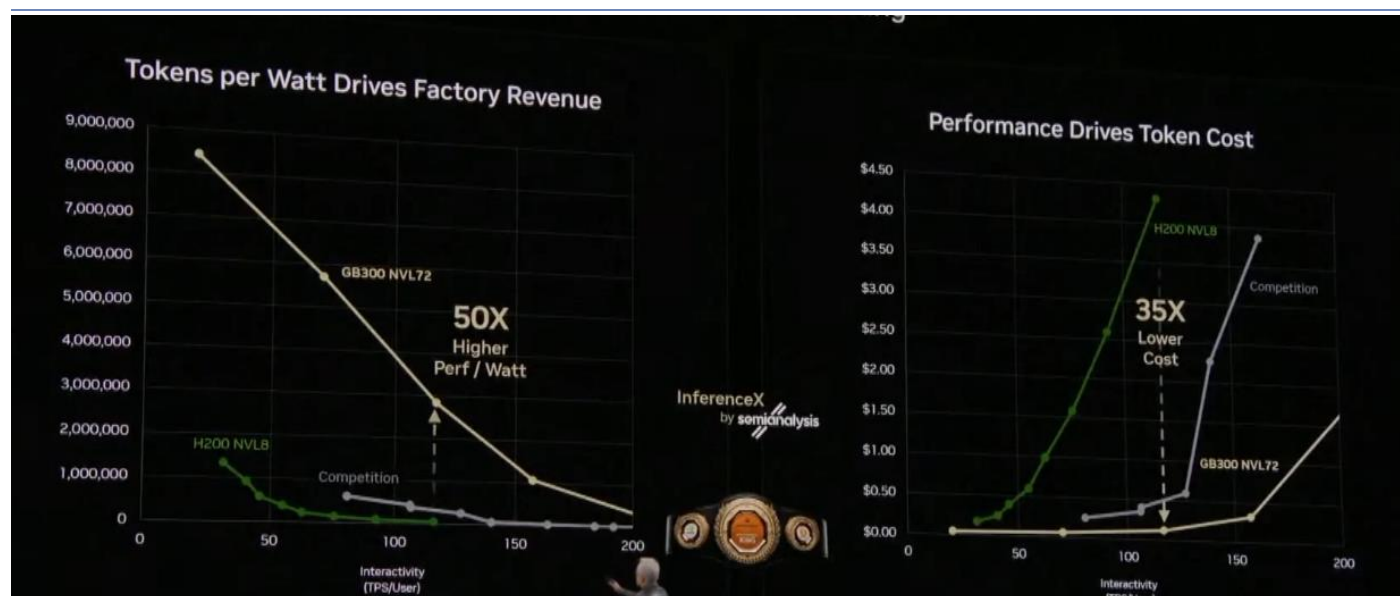
核心价值：消费Token，将AI能力落地于具体业务场景。



资料来源：致网科技，中国银河证券研究院

Token 生成的成本快速降低，重新定义生产效率。在构建大规模 AI 数据中心时，芯片与电子器件的选择直接决定了 Token 的生成速率与每瓦吞吐量，进而决定了企业的营收天花板与运营成本。根据 GTC 2026 上数据显示，通过 Semi Analysis 的评测数据论证其新一代 GB300 NVL72 系统如何重新定义 Token 生产的成本与效率，GB300 NVL72 在单位功耗下生成的 Token 数量远超竞争对手，达到惊人的 50 倍提升，这意味着同样的电力预算能产生 5 倍的商业价值。从经济性角度切入，揭示了随着用户交互频率增加，传统竞品的成本曲线呈指数级飙升，而 GB300 NVL72 的成本却维持在极低水平，实现了 35 倍的单 Token 成本下降。这种性能飞跃并非单纯堆砌硬件，而是源于 NVIDIA 提出的极限协同设计理念，通过将 GPU、网络 and 机架级系统进行全栈优化，成功打破了 AI 工厂的物理瓶颈。

图12: NVIDIA 在 AI 推理领域的极致性能优势

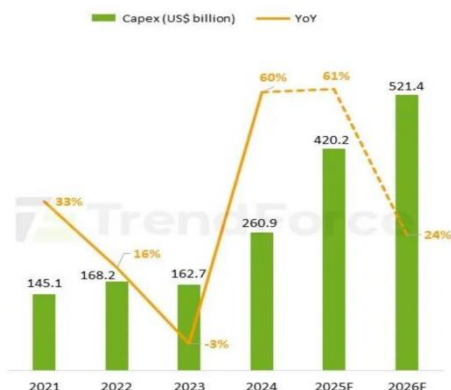


资料来源：GTC2026，中国银河证券研究院

(二) Token 生成需求爆发，对上游硬件需求持续提升

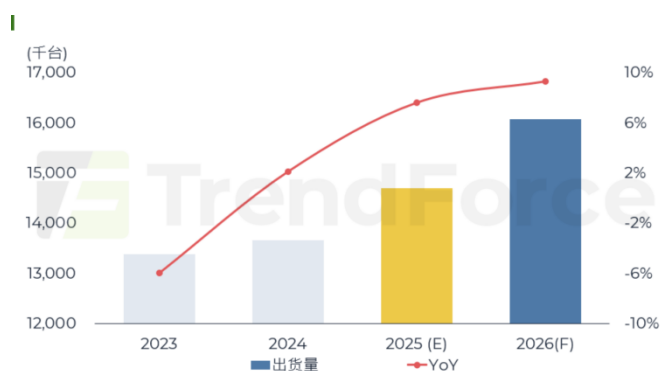
AI 大模型训练对于 Token 需求持续爆发，已成为驱动人工智能算力需求增长的关键力量。为应对这一趋势，全球主要云端服务提供商（CSP）正持续加大在计算基础设施方面的投入。根据 TrendForce 集邦咨询的研究，预计 2026 年在 GB/VR 等 AI 机柜方案持续放量下，八大 CSP 的总资本支出有望再创新高，年增 24%，来到 5,200 亿美元以上。支出结构已从能直接创造收益的设备，转向 Server、GPU 等资产，意味着巩固中长期竞争力与市占率优先于改善短期获利。

图13: CSP 预期资本开支情况



资料来源: Trendforce, 中国银河证券研究院

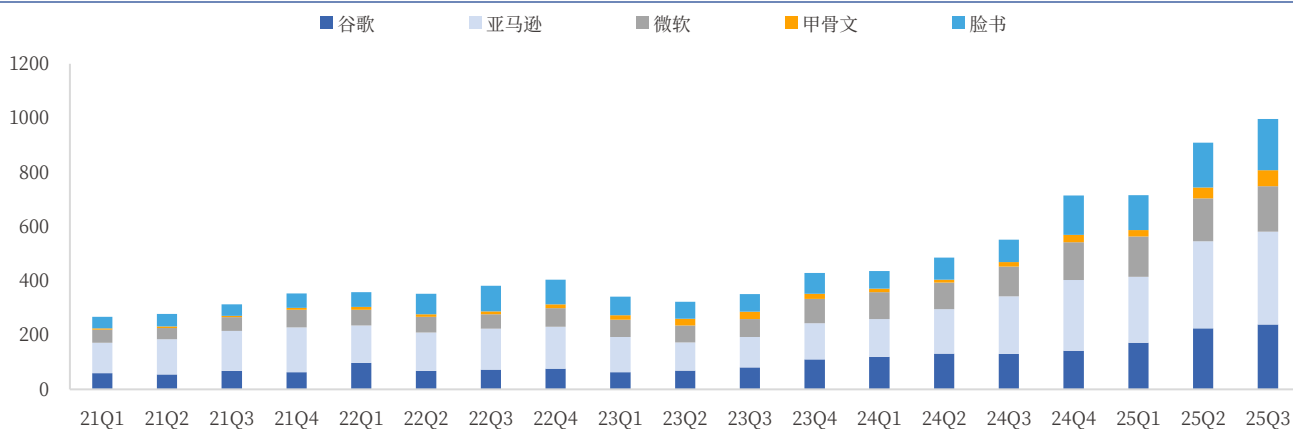
图14: 2023-2026 年 AI 服务器出货量预估



资料来源: Trendforce, 中国银河证券研究院

谷歌、亚马逊、Meta、微软、甲骨文这五大 CSP 厂商的资本开支呈现爆发式增长态势。具体来看，微软本季度资本支出达到 349 亿美元，同比增长 74%，主要用于 AI 数据中心和 TPU 算力基础设施建设；谷歌资本支出为 239.53 亿美元，同比增长 83%，其中 60% 投向服务器，40% 投向数据中心和网络设备；亚马逊资本支出 342 亿美元，同比增长 61%，主要用于 AWS 云服务的基础设施扩建；Meta 资本支出 194 亿美元，同比增长 111%，主要投向 AI 基础设施和生成式 AI 研发；甲骨文资本支出 120 亿美元，主要用于数据中心扩建以支持与 OpenAI 等签署的 AI 合作协议。这五大厂商合计资本支出超过 1244 亿美元，同比增长普遍在 60%-110% 之间，AI 算力军备竞赛逐步激烈。

图15: 海外 CSP 厂商资本开支投入情况 (单位: 亿美元)



资料来源: , 中国银河证券研究院

大模型的快速发展，带动了 AI 算力的强劲需求。OpenAI 已承诺从甲骨文、英伟达、AMD 及博通等供应商处获取超过 26GW 的算力资源，未来十年相关成本可能远超 1 万亿美元规模。

表1: OPEN AI 和互联网大厂的合作情况

合作厂商	时间	合作内容与规划
甲骨文	2025 年 9 月	计划在约 5 年内采购价值 3000 亿美元的计算资源。 · 采购启动: 2027 年 · 电力需求: 合同执行需 4.5 GW 电力容量。
英伟达	2025 年 10 月	为部署下一代 AI 基础设施, 计划部署至少 10 GW 的 NVIDIA 系统, 用于训练和运行下一代模型 (超级智能)。 · 投资规模: NVIDIA 计划逐步投资高达 1000 亿美元, 首阶段上线: 2026 年下半年, 基于 NVIDIA Vera Rubin 平台
AMD	2025 年 10 月	达成总规模 6 GW 的多世代 AMD Instinct GPU 合作协议。 · 首阶段部署: 2026 年下半年, 启动 1 GW 的 AMD Instinct MI450 · 股权安排: AMD 向 OpenAI 发行认股权证, 可行权购买至多 1.6 亿股 AMD 股票 (行权价 0.01 美元/股)。
博通	2025 年 10 月	共同开发并部署总规模 10 GW 的定制 AI 加速器及网络系统。 · 启动时间: 2026 年下半年部署首批机架 · 完成时间: 2029 年底前完成全部 10 GW 容量落地。

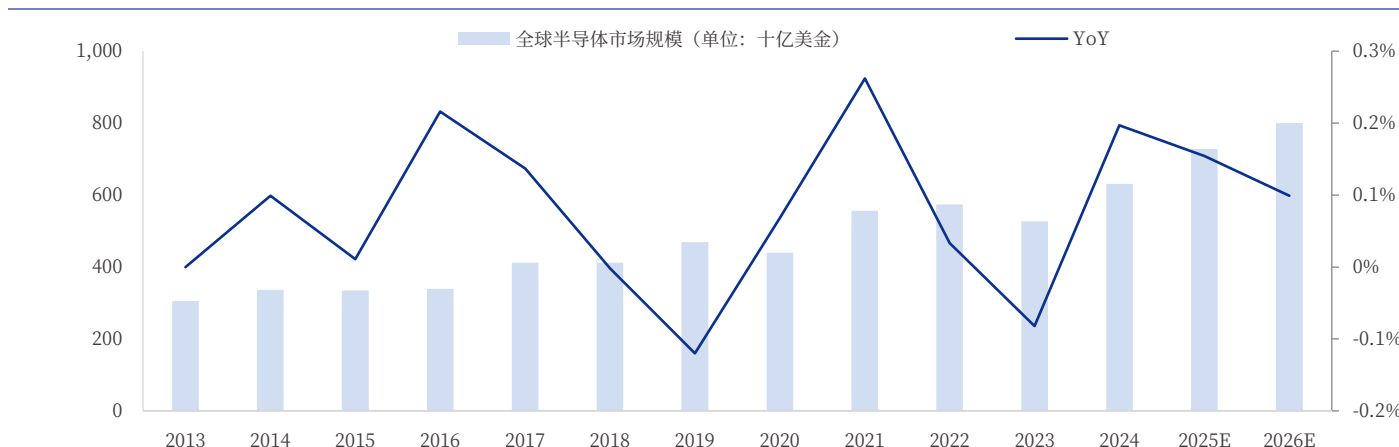
资料来源: 甲骨文, 英伟达, AMD, 博通, 中国银河证券研究院整理

存储行业的发展呈现出明显的周期性波动，目前已进入由人工智能驱动新一轮上行周期，本轮周期的核心驱动力发生了根本性转变。

需求端的结构性变革是当前周期的突出特征。AI 服务器对存储芯片的容量和性能要求远超传统服务器，其 DRAM 用量约为普通服务器的 8 倍，NAND Flash 用量约为 3 倍。这直接引爆了对高带宽内存（HBM）、DDR5 内存及企业级固态硬盘（eSSD）等高性能产品的“吞噬式”需求。特别是 HBM，因其在 AI 训练和推理中的关键作用，需求呈现井喷之势，预计到 2027 年其在 DRAM 市场总价值中的占比将高达 43%。

供给端的结构性失衡随之加剧。为追求更高利润，三星、SK 海力士等主要厂商将先进产能优先转向 HBM 和 DDR5 等高端产品，这大幅挤压了 DDR4 等传统存储芯片的产能。然而，大量存量服务器和新兴 AI 应用仍广泛需要 DDR4，导致其供应严重短缺，价格出现飙升，甚至出现了 DDR4 价格反超 DDR5 的异常现象。此外，HBM 本身因技术复杂、产能扩张缓慢，供应持续紧张，巨头的远期产能已被云服务商通过长期协议提前锁定。

图16: 全球半导体市场规模及增长



资料来源: WSTS, 中国银河证券研究院

受人工智能需求强劲拉动, 预计 2025 年第四季度存储芯片市场预计将持续上扬。主要驱动力在于, 三星、SK 海力士与美光三大原厂持续将先进产能优先分配给高阶服务器 DRAM 和 HBM (高带宽内存), 这直接导致面向普通消费市场的通用型 DRAM 产能受限, 供应紧张。据 TrendForce 预测, 本季度通用 DRAM 合约价有望环比上涨 8-13%, 若将价值更高的 HBM 一并计算, 整体 DRAM 涨幅将进一步扩大至 13-18%。预测 NAND 产品价格将在 25Q4 涨价 5-10%。

图17: 预测 DRAM 价格趋势

3Q25-4Q25DRAM价格预测

	3Q25E	4Q25F
Total DRAM	Conventional DRAM: up 10~15% HBM Blended: Up15%~20% (HBM Penetration: 8%)	Conventional DRAM: up 8~13% HBM Blended: Up13%~18% (HBM Penetration: 11%)

Source: TrendForce, Sept. 2025

TrendForce

资料来源: Trendforce, 中国银河证券研究院

图18: 预测 NAND 价格趋势

3Q25-4Q25 NAND Flash 价格预测

	3Q25E	4Q25F
Total NAND Flash	up 3~8%	up 5~10%

Source: TrendForce, Sept. 2025

TrendForce

资料来源: Trendforce, 中国银河证券研究院

供给与需求共振, Token 需求快速增长对于高性能产品需求快速增长

本轮周期来看, 供给端收缩 6 季度, 从 24Q1-25Q2, 面对前期下行周期, 原厂持续减产, 叠加渠道库存已降至低位, 供给端收缩效应逐步显现。同期, 在政策刺激下智能手机、PC 等传统需求温和回暖, 共同驱动存储价格从低位企稳并进入回升通道, 为后续上涨夯实了基础。25Q3-至今, 需求结构切换, 驱动价格加速上行。核心驱动力已转向 AI 带来的结构性需求。一方面, AI 服务器出货量快速攀升, 其配备的企业级 SSD 需求是传统服务器的数倍; 另一方面, 原厂为优先保障 HBM、DDR5 等高附加值产品, 进一步挤压了通用型 NAND Flash 和 DDR4 的产能。供需两侧的剧变, 导致市场从整体平衡迅速转为结构性短缺。

表2: 2025 年存储芯片涨价节奏

时间节点	关键事件与市场动态	价格走势标志
4Q24	1. 原厂延续减产策略; 2. 智能手机去库存, AI 及 DeepSeek 相关需求开始释放	价格止跌回升
1Q25	市场价格触底, 供需结构初步改善	价格触底
2Q25	1. 闪迪首次全面涨价 (涨幅>10%); 2. 三星、SK 海力士、美光等跟进调价	价格上涨

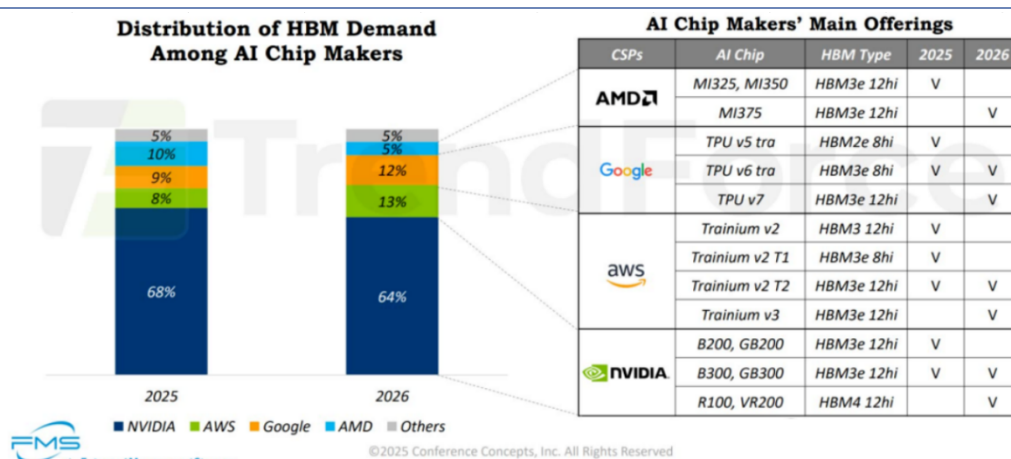
3Q25	1. 闪迪二次涨价 (>10%)；美光暂停报价；2. HDD 短缺推动 CSP 转向 QLC eSSD	涨幅持续
4Q25	1. 闪迪第三次涨价（合约价上调 50%）；2. 三星、海力士跟进涨价 20%-30%；3. 原厂保持产能紧缺策略	涨幅高点

资料来源：中国银河证券研究院

需求端：AI Token 需求激增，HBM/DRAM/NAND 均受益本轮周期爆发

2026 年 AI 芯片市场将呈现明显的技术分层格局。NVIDIA 凭借 Rubin 架构率先采用 HBM4（12-Hi 堆叠），成为 HBM4 时代的主要消费驱动者，其 Vera Rubin 平台预计在 2026 年第三或第四季度量产，FP4 推理算力达 3.6 Exaflops，相比 GB300 提升约 3.3 倍。与此同时，AMD、Google、AWS 等厂商则聚焦 HBM3e 市场：AMD 的 MI350 系列采用 HBM3e（12-Hi），容量达 288GB；Google 的 TPU 产品线从 HBM2e 升级至 HBM3e，堆叠层数从 8-Hi 向 12-Hi 演进；AWS 的 Trainium3 同样采用 HBM3e（12-Hi），单芯片提供 144GB 显存。这一技术分层反映了市场对总拥有成本（TCO）的考量，以及 AI 推理需求的持续增长，推动云服务提供商自研 ASIC 芯片在 2026 年展现出更强的出货势头。

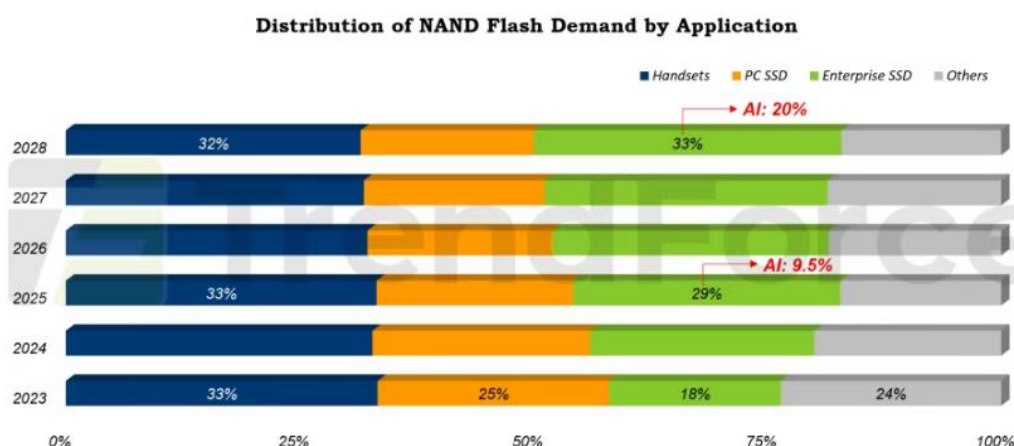
图19：AI 芯片对于 HBM 的需求情况



资料来源：Trendforce，中国银河证券研究院

AI 及服务器成 NAND 产品需求增长最快领域。AI 浪潮下，未来 AI 服务器对 SSD 特别是企业级 SSD 的需求有很高的增长，手机、PC 和服务器，服务器部分对 SSD（也是 NAND）的需求从 2025 占比 29%增长到 2028 年占比 33%，PC 的占比明显在缩小。谁战略投入企业级 SSD，谁可能赢得了未来，特别是在 AI server 上，除了传统的 TLC 和 QLC SSD 产品，也包括未来的大容量 QLC SSD 的机会。

图20: NAND Flash 在不同应用领域占比



资料来源: Trendforce, 中国银河证券研究院

供给端：HBM 产品成产业争夺价值高地，产能扩张计划的差异将直接决定未来的市场格局

HBM（高带宽内存）已成为全球存储行业资本开支和产能扩张的绝对焦点。整个行业的主要增量资源将集中投向 HBM 领域，推动 HBM 在整体产能中的占比实现显著提升。在这一轮技术竞赛中，SK 海力士展现出最为激进的扩张姿态，计划在 2026 年将高达三分之一的总产能用于 HBM 生产，这一战略决策旨在巩固并扩大其在 HBM 市场的领先优势。面对 SK 海力士的强势领跑，美光通过近乎翻倍的 HBM 产能扩张，显示出其绝不甘心在这一高增长市场落后的决心，正全力追赶三星和 SK 海力士两大巨头。

HBM 竞争格局高度集中，主要在三巨头——三星、SK 海力士和美光之间展开，这将进一步加剧内存市场的技术和盈利分化。值得注意的是，HBM4 已成为 2026 年的决胜关键，各大厂商不仅在积极扩充现有 HBM3/HBM3E 的产能，更在为下一代 HBM4 产品的量产进行技术和产能储备，预示着下一轮技术竞赛的序幕已经拉开。

图21: HBM 将成为 2026 年存储器行业产能投资的焦点



资料来源: Trendforce, 中国银河证券研究院

全球 DRAM 技术演进与市场竞争呈现鲜明格局。技术层面，主流供应商正全力聚焦于更先进的

1b/1beta 和 1c/1gamma 制程 DDR5 产品的开发与量产，其中 6400 MT/s 已成为 DDR5 的主流速度，并向 7200-8800 MT/s 迈进。三星计划将 1c 制程产能优先分配给下一代 HBM4 产品，凸显了 AI 驱动下高端存储的战略重心转移。同时中国厂商取得了显著进展，长鑫存储已于 2025Q1 推出其 G4 制程 DDR5 产品，性能已接近国际大厂的上一代主流水平，正在努力缩小技术代差。

图22: DRAM 在 DDR5 的技术变化路径

Density	Nodes (nm)	Company	1Q25	2Q25	3Q25	4Q25	1Q26	2Q26	3Q26	4Q26
16Gb	1Y	SAMSUNG	4800-5600 MT/s				EOL			
		SK hynix	4800 MT/s				EOL			
	G4	exmt								
	1Z	micron	4800-5600 MT/s	EOL						4800-6400 MT/s
	1a/1alpha	SAMSUNG								
		SK hynix								
	1b/1beta	SAMSUNG	• C/S: 5600-8000 MT/s							5600-6400 MT/s
		SK hynix	C/S							7200 MT/s
		micron	• C/S: 7200 MT/s							5600-6400 MT/s
		SK hynix	• C/S: 7200 MT/s							5600-6400 MT/s
1c/1gamma		SK hynix	• C/S: 8000 MT/s							5600-6400 MT/s
		micron	• C/S: 8000 MT/s							5600-7200 MT/s

资料来源: Trendforce, 中国银河证券研究院

从 NAND 产品技术路线图来看，各家会聚焦大容量 QLC NAND 和高速接口高性能 NAND 产品，以应对未来 AI 时代对 NAND 的各种需求。

图23: NAND 厂商技术路径

Vendors	2023		2024		2025		2026	
	1H	2H	1H	2H	1H	2H	1H	2H
SAMSUNG	14nm (MLC/TLC)							
	V7, 176L (TLC/QLC)	V8, 236L (TLC) ★			V9, 286L (TLC/QLC)			
SK hynix	14nm (MLC/TLC)							
	V7,176L(TLC/QLC) ★	V8, 238L (TLC/QLC)			V9, 321L (TLC/QLC)			V10, 375L
SOLIDIGM	G5, 192L FG (QLC)						G6, 240L (TLC/QLC)	
KIOXIA SANDISK™	15nm (MLC/TLC)							
	★ BiCS5, 112L (TLC/QLC)	BiCS6, 162L (TLC/QLC)			BiCS8, 218L (TLC/QLC)		BiCS9, 2YYL (TLC/QLC)	
micron.	16nm (MLC/TLC)							
	G8, 232L (TLC/QLC) ★		G9, 276L (TLC/QLC)					
MXIC	19nm (SLC/MLC) ★							
	96L (TLC)	192L (TLC)						

资料来源: Trendforce, 中国银河证券研究院

全球存储芯片供给仍偏谨慎，以高密度、高容量的技术路线转型为主。以三星、SK 海力士、美光为代表的巨头普遍对 NAND 的资本性扩张持高度谨慎态度，其投资重心明显向需求更旺盛的 DRAM（特别是 HBM）领域倾斜。在 NAND 业务上，策略核心从“扩产”转向“提质”，即通过向更高层数（如 1c/1γ、BiCS 8/10 等）的技术节点转型升级，来提升单位面积的存储密度和晶圆产

出比特量，从而实现以更少的晶圆投入驱动更高的位元增长。例如，铠侠与西部数据通过合资企业专注于 BiCS 技术的迭代，其新建产能也主要用于先进制程的转换。这种战略共识旨在严格控制供给端，通过技术升级而非产能堆叠来满足市场需求。

表3：2025 年存储芯片厂商扩产节奏

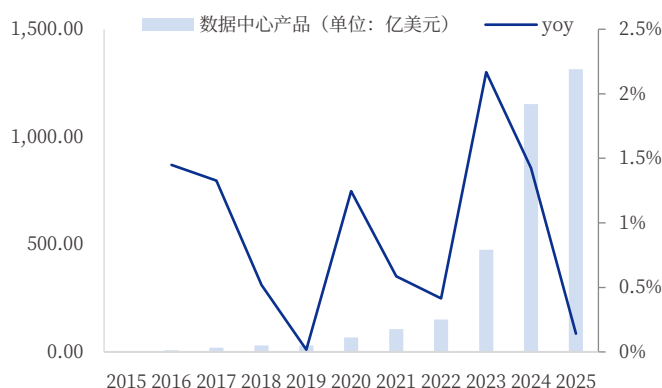
厂商	扩产计划与战略重点
三星电子	将部分 HBM3E 产能转向利润率更高的通用 DRAM（如 DDR5），并计划在平泽 P4 工厂扩产 1c 纳米制程的 DRAM 和 HBM4，目标是在 2026 年第二季度将相关月产能提升至 14 万片。
SK 海力士	聚焦巩固 HBM 市场领导地位，已开始向客户送样 HBM4，并计划在 2026 年第二季度大幅扩产。同时，计划将 1c 纳米 DRAM 的月产能从约 2 万片大幅提升至 16-19 万片。
美光科技	战略性退出消费级品牌业务，将全部资源集中于 HBM 和数据中心等高附加值领域。正在加速美国爱达荷州和纽约州新晶圆厂的建设，以满足激增的 HBM 需求，其 2026 年产能已售罄。
铠侠	专注于 NAND 闪存技术升级与产能扩张。北上 K2 工厂已于 2025 年秋季投产，并计划扩大基于高层数（如 321 层）技术的产品供应，以应对 AI 服务器带来的需求。
台积电/三星（先进制程）	将高端产能（如 2 纳米、HBM）的扩张重心重新锚定在亚洲（中国台湾、韩国），同时放缓美国建厂计划，主要由于成本、人才和供应链效率的考量。
国产/利基型厂商	抓住三大原厂聚焦先进制程留下的市场缺口，积极扩产 DDR4 等成熟制程产品，并实现技术突破（如长鑫发布 DDR5 产品），市场份额和盈利能力显著提升。

资料来源：三星，海力士，美光，台积电，铠侠，中国银河证券研究院

（三）“一超多强”算力芯片格局确定，关注国产芯片替代机遇

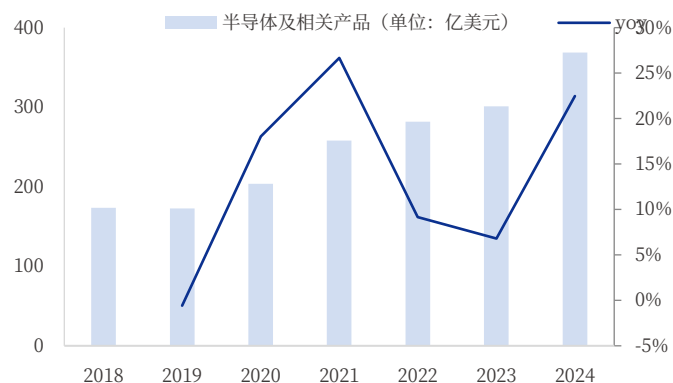
英伟达数据中心业务和博通 AI 芯片业务的持续高增长，确实印证了全球 AI 算力需求的强劲态势。英伟达凭借其在 GPU 硬件和 CUDA 生态的绝对优势，持续主导 AI 训练市场。其数据中心业务自 2023 年第二季度以来保持超高速增长，例如在 2026 财年第二季度，该业务营收达到 411 亿美元，同比增长 56.4%，这主要得益于市场对 AI 和高性能计算的强劲需求推动其 GPU 销售，尤其是 Blackwell 等架构产品的旺盛需求。与此同时，博通则通过定制化 AI 芯片（ASIC）路线异军突起，精准抓住了云厂商为优化总体拥有成本（TCO）而寻求差异化解决方案的趋势。其 AI 业务表现尤为抢眼，2025 财年第三财季 AI 半导体收入同比增长 63%至 52 亿美元，第四财季更是同比增长 74%，并预计实现连续多个季度的增长，这反映了大型科技公司如谷歌、Meta 以及 OpenAI 等对专用 AI 加速器的需求激增。

图24：英伟达数据中心业务持续高增长



资料来源：，中国银河证券研究院

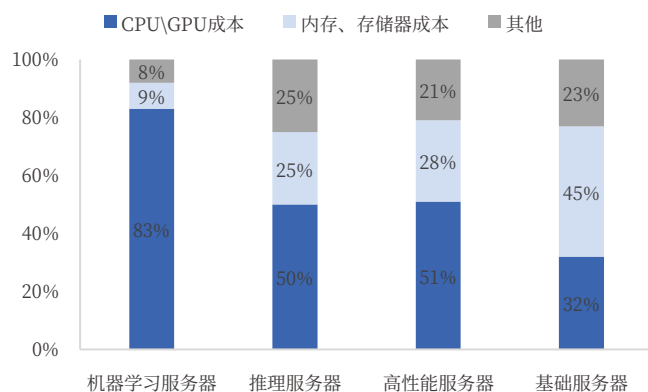
图25：博通 AI 芯片业务高增长



资料来源：，中国银河证券研究院

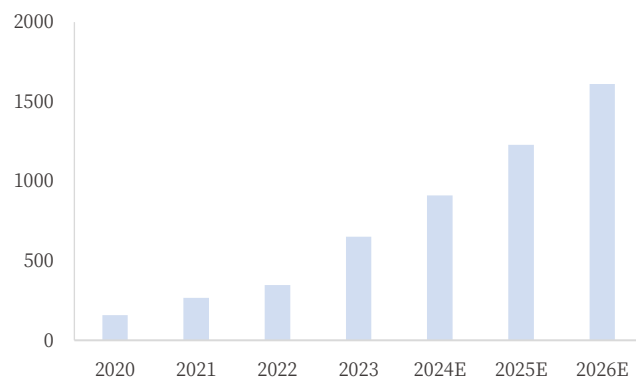
AI 芯片是算力产业的核心基础，为智能算法与各类应用场景提供关键计算支持。在服务器成本结构中，以 GPU 为代表的核心芯片占比高达 80% 以上，实现 AI 芯片技术的自主可控对智算产业的长远发展具有战略意义，随着人工智能应用场景的持续扩展，市场对高性能 AI 芯片的需求显著提升。存算一体架构、光通信等前沿技术的突破，进一步推动了该领域的技术迭代与市场增长。2023 年中国 AI 芯片市场规模约为 652 亿元，预计到 2026 年将突破 1600 亿元，行业快速增长

图26：服务器成本构成



资料来源：甲子光年，中国银河证券研究院

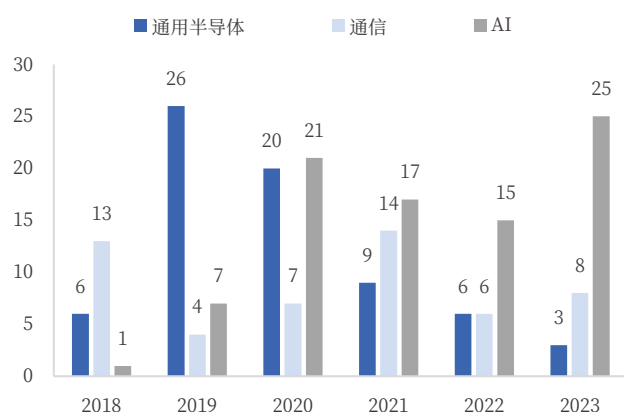
图27：2020-2026 年中国 AI 芯片市场规模(单位:亿人民币)



资料来源：甲子光年，中国银河证券研究院

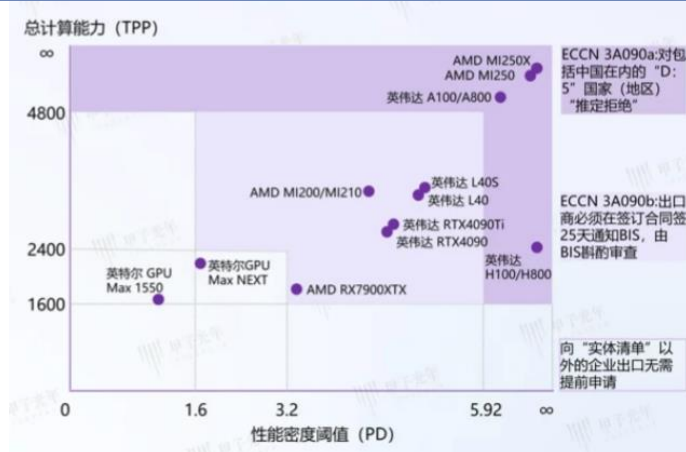
美国对华持续封锁，关注 AI 国产化进程。在美国对华 AI 芯片技术封锁的背景下，中国通过政策扶持、技术创新和产业链协同加速国产替代进程。华为昇腾、寒武纪等企业通过架构创新（如存算一体、Chiplet 技术）提升成熟制程芯片性能，昇腾 910B 已接近国际主流 GPU 水平。政策层面，国家将 AI 芯片列为科技战略重点，预计 2025 年自给率有望从 34% 提升至 82%。未来，中国将聚焦云端训练、边缘计算及类脑芯片等方向，通过开源生态（如 RISC-V）和技术标准制定，构建自主产业链。

图28：被列入“实体清单”的 AI 芯片企业数量



资料来源：甲子光年，中国银河证券研究院

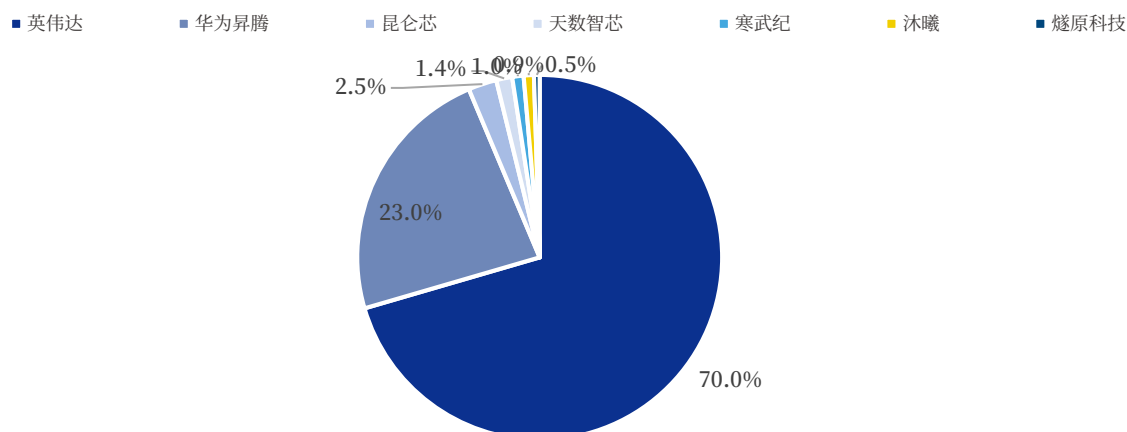
图29：美国商务部工业和安全局对华高算力芯片出口限制



资料来源：甲子光年，中国银河证券研究院

2024 年，在 AI 加速芯片领域，GPU 卡以 80% 的份额主导市场，但国产芯片表现亮眼：华为昇腾出货 64 万张（占比 23%），仅次于英伟达（190 万张，70%），昆仑芯、天数智芯、寒武纪分列三至五名。国产芯片在推理场景占比达 61%，生态适配逐步完善。根据 IDC 预测，到 2028 年中国加速服务器市场规模将达 253 亿美元，非 GPU 占比或接近 50%。当前，美国技术管制虽限制进口，但刺激了本土研发，华为昇腾等芯片已成为国产替代主力，尤其在运营商智算中心项目中占据主导。

图30：2024 年中国 AI 芯片出货量统计



资料来源：半导体产业联盟，IDC，深芯盟，中国银河证券研究院

国家高度重视算力发展，自 2022 年 8 月至 2024 年 10 月，科技部、工信部、发改委、国务院等多部门陆续出台一系列政策，旨在系统推动算力基础设施建设与产业生态培育。这些政策聚焦于构建全国一体化算力体系，通过“东数西算”工程优化布局，强化智能算力占比，并超前部署数字基础设施。同时，政策着力完善数据标准体系，推动公共数据资源开发利用，鼓励技术创新与场景应用，并通过税收优惠、绿电比例要求等具体措施，促进算力产业的普惠、安全、绿色与高质量发展，为数字经济和人工智能创新奠定坚实基础。

表4：国内市场对于 AI 支持政策情况

时间	发文部门	文件名称	主要内容概要
2022 年 8 月	科技部、财政部	《企业技术创新能力提升行动方案（2022-2023 年）》	推动国家超算中心、智能计算中心等面向企业提供低成本算力服务，支持建设重大示范应用场景，释放场景合作机会。
2023 年 3 月	科技部	《关于开展国家新一代人工智能公共算力开放创新平台申报工作的通知》	启动“国家新一代人工智能公共算力开放创新平台”建设申报，满足我国人工智能技术与算力基础设施建设需求。
2023 年 10 月	工业和信息化部	《算力基础设施高质量发展行动计划》	推动算力结构多元配置，提升智能算力占比，实现智能算力与通用算力协同发展，满足不同业务需求。
2023 年 12 月	国家发展改革委员会	《关于深入实施“东数西算”工程加快构建全国一体化算力网的实施意见》	提出到 2025 年初步构建普惠、绿色、安全的算力基础设施体系，推动算力资源协同调度，提高智能算力占比，降低用能成本和网络传输成本。
2024 年 3 月	中央人民政府	《政府工作报告》	适度超前建设数字基础设施，加快形成全国一体化算力体系，培育算力产业生态。
2024 年 9 月	国务院办公厅	《国务院办公厅关于加快公共数据资源开发利用的意见》	繁荣数据产业发展生态，将数据产业纳入产业结构调整指导目录，支持技术创新，落实税收优惠，推动数据资源开发利用。
2024 年 10 月	国家发展改革委员会	《国家数据标准体系建设指南》	强化基础设施互联互通、算力保障和流通利用标准建设，为数据资源、技术、流通与应用提供系统标准支撑。

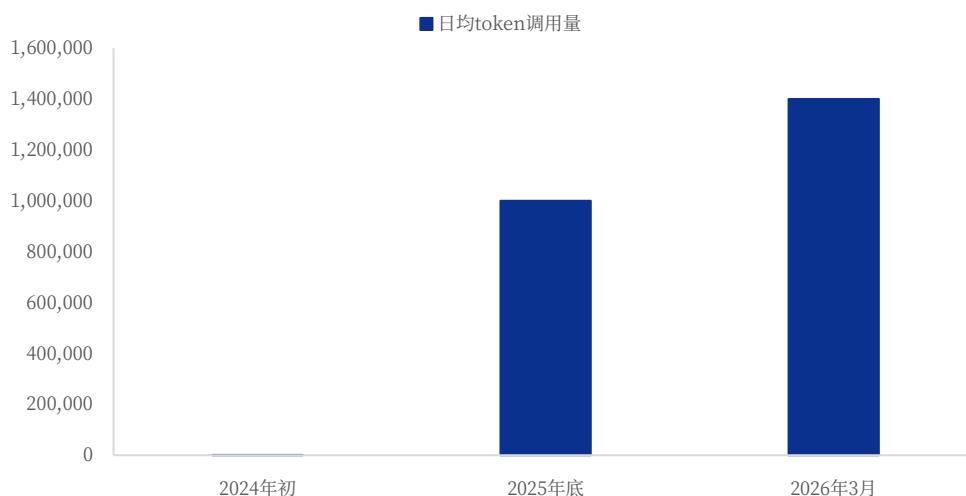
资料来源：甲子光年，中国银河证券研究院整理

三、电子器件是 Token 经济的核心骨架

（一）AI 驱动电子元器件升级迭代

数据中心正演变为处理海量 Token 生成的超级工厂。Token 经济时代，AI 服务器承载 Token 生成的重要功能，相当于 Token 经济时代的基础设施。为支撑超大规模 AI 模型的训练与推理效能，底层基础设施需在高速互联、极致散热与信号完整性三个维度实现系统性突破，以匹配 Token 工厂的高通量生产需求。其中，随着 GPU 集群规模迈向万卡级别，传统以太网已难以承载庞大的参数同步流量。采用高带宽互连技术，可实现计算单元间纳秒级低延迟通信，确保万亿参数模型训练时的梯度同步效率，避免因网络拥塞导致的算力空转。AI 算力密度的激增推动单机柜功率从数千瓦跃升至数十千瓦，风冷散热已逼近物理极限，液冷技术凭借其约为空气 1000 倍的热传导系数，成为高算力场景的必要选择。在 112Gbps 及更高速率的 SerDes 接口标准下，信号传输质量直接决定了数据流转的上限。通过 PCB 板材升级、差分走线阻抗匹配及光模块的抗干扰设计，可以确保信号在长距离传输中不失真，满足工厂数据吞吐的稳定性与可靠性。

图31：中国日均 Token 调用量快速增长



资料来源：国家数据局，中国银河证券研究院

信号传输质量直接决定算力释放效率。面对 112Gbps 及以上速率的信号传输需求，传统 PCB 面临趋肤效应、介质损耗与串扰加剧等挑战。AI 服务器 PCB 需采用 Ultra-Low Loss 材料体系，将介电损耗降至 0.002 以下，确保信号在长距离传输中衰减可控。同时，通过优化层叠结构设计，采用多阶 HDI 与极细线路，缩短高速信号传输路径，降低寄生电容与电感干扰。

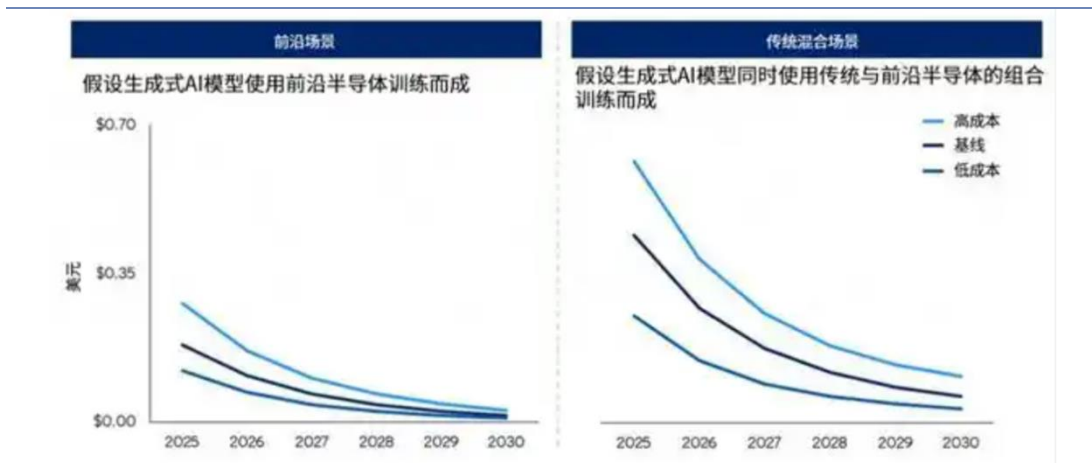
被动元件是保障 GPU/TPU 核心稳定运行的基石。随着单颗芯片功耗突破 700W，电源系统需在毫秒级时间内响应百安级电流波动，MLCC 与钽电容的配置密度呈指数级增长。特别是大容量、耐高温的高端 MLCC，被广泛部署在 CPU 插座周围，利用超低 ESR 特性滤除电源纹波，防止瞬态电压跌落引发的算力宕机。为应对开关频率提升带来的损耗，电感元件从传统的铁氧体向金属合金粉芯升级，显著提升饱和电流与效率。随着 Chiplet 技术普及，供电方案从板级下沉至封装级，催生了对芯片电感等特殊被动元件的需求，以支撑更精准的电压调节。被动元件正从通用标准件向高性能、定制化方向演进，支撑 AI 服务器高效运算。

数据中心正从风冷向液冷加速转换。AI 服务器单机柜功率密度从传统的 5kW 跃升至 40kW 甚至 100kW 以上，传统风冷系统因空气导热系数低、换热路径长，已逼近物理散热极限，极易导致芯片因过热而降频，严重制约算力输出效率。液冷技术的核心价值在于突破了热传导的瓶颈。液体的比热容是空气的千倍以上，能够以更少的能耗带走更多的热量。通过冷板式或浸没式液冷方案，可将数据中心 PUE 值从风冷的 1.5 以上大幅压降至 1.1 以内，液冷能为 AI 工厂节省近 40% 的制冷能耗，直接转化为运营成本的显著下降。液冷方案使得数据中心摆脱了对庞大风扇与空调系统的依赖，服务器节点可高密度堆叠，大幅提升数据中心单位建筑面积的算力产出。

电子元器件产业的规模化生产与技术优化带动 Token 生成成本下行。AI 算力需求井喷的背景下，数据中心建设成本与运营支出居高不下，成为制约 Token 大规模商业化落地的核心瓶颈。随着 PCB、被动元件及光模块的制造工艺走向成熟，供应链规模效应开始显现。高端服务器所需的超低损耗覆铜板、高容高温 MLCC 等器件，在产能爬坡后单位成本显著摊薄，直接降低了 AI 服务器的整机物料成本，使得数据中心能以更低的资本开支部署同等规模的算力集群。元器件的精细化升级带来了能效红利。通过采用更高集成度的 HDI 板替代传统背板，以及引入更低 ESL 的电容与电感，

系统供电效率与信号完整性大幅提升，有效降低了无效功耗。硬件层面的降本增效叠加液冷等散热技术，显著压低了数据中心的 PUE 值与电力支出。单位 Token 的生成成本随着随元器件迭代而持续下移，为 AI 应用的普惠化提供坚实的经济基础。**Gartner 预测，到 2030 年，生成式 AI 提供商对 1 万亿参数大语言模型（LLM）的推理成本将比 2025 年降低 90% 以上。**

图32：生成式 AI 推理成本场景预测

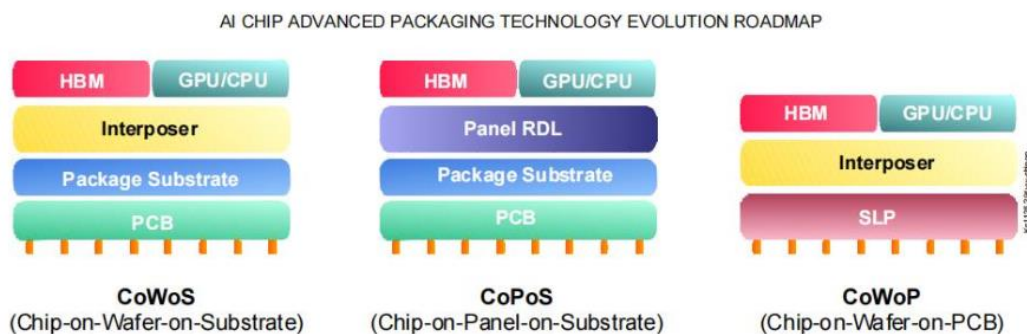


资料来源: Gartner, 中国银河证券研究院

（二）PCB：Token 时代 PCB 重要性提升，中国制造中心地位稳固

Token 经济时代，PCB 重要性显著提升。随着系统架构向更高密度、更极致能效、更大规模协同设计的跃迁，PCB 技术难度和性能要求全方位升维，以适配更高标准的信号传输、散热及集成需求。数据通讯应用领域高规格硬件正驱动超高层（HLC）与高密度互连（HDI）等工艺走向前所未有的深度融合。为了支撑 224G SerDes 等极速传输需求，在 HLC、高频高速、HDI 及高通流等维度对 PCB 提出了严苛要求。传统 PCB 与先进封装之间的技术边界也正在快速模糊，催生出 CoWoP 等新的前沿架构模式，将先进封装技术延伸至 PCB 级别，通过将芯片直接集成在 PCB 上实现更短的信号路径。**PCB 技术核心已从传统的电路支撑演变为高速系统互联与电源管理的集成平台，对布线密度、信号完整性及电源分配网络（PDN）提出了严苛要求。**PCB 技术突破的重心已全面跨越单一的制造维度，系统性地转向对信号完整性与电源完整性的协同优化、高密复杂结构的设计落地、严苛工况下的产品可靠性保障，以及针对超低损耗、极低损耗及更高级别材料的先进加工工艺探索。

图33: AI 芯片的封装技术路径



资料来源: Prismark, 中国银河证券研究院

在 AI 产业的推动下, 2025 年全球 PCB 市场规模达到 849 亿美元, 同比增长 15.4%, 增速较 2024 年进一步提升。其中第四季度市场规模达到 228 亿美元, 同比增长 18%。结构上看, HDI 板和高多层板成为这一轮增长的主力军, 其中 HDI 板市场规模为 157 亿美元, 同比增加 25.6%; 高多层板 (18 层以上) 市场规模 45 亿美元, 同比增加 85.5%。此外, 封装基板市场亦有较高增长, 市场规模达到 147 亿美元, 同比增长 16.9%。

图34: AI 推动全球 PCB 行业规模持续高增



资料来源: Prismark, 中国银河证券研究院

国内 PCB 企业份额提升。面向数据通讯应用领域 PCB 需求的爆发式增长, 行业内公司调整战略, 将资源向该领域倾斜, 以图通过激进的资本开支与产能扩张切入市场并取得一定的市场份额, 随着行业产能的持续扩张, 新增产能的逐步落地, 成熟技术平台的准入门槛将被摊薄, 未来的竞争或将趋向同质化与白热化, 利润空间或将面临结构性挤压; 同时, PCB 日渐成为数据通讯应用领域高阶硬件的重要赋能者, 其技术跃迁正逐步重构 PCB 行业的竞争边界与价值分布格局, 行业或将呈现“入场者多、通关者少”的格局, 并驱使高阶 PCB 产品市场加速向具备全球化产能协同与顶尖研发积淀的行业头部梯队集中, 呈现明显的结构性分化。2025 年中国 PCB 市场尽管已是全球规模最

大的市场，但其产值增速预计仍为全球最快，同比增长约 19.2%，其在 AI 相关的高多层 PCB、高密度互连（HDI）PCB 及封装基板领域的增长尤为突出。

表5：2024-2025 全球 PCB 产值增长率估算

产值增长率	多层板			HDI	封装基板	柔性板	其他	总计
	4-6 层	8-16 层	18 层以上					
美洲	5.50%	6.20%	10.70%	14.20%	5.90%	7.70%	3.30%	7.50%
欧洲	13.20%	16.70%	13.50%	15.30%	807.70%	4.50%	6.90%	13.80%
日本	5.80%	9.20%	16.80%	6.60%	18.10%	1.00%	7.10%	11.30%
中国	5.90%	28.40%	119.20%	33.50%	21.70%	5.60%	6.70%	19.20%
亚洲	7.00%	19.60%	44.00%	14.10%	15.90%	0.10%	4.10%	11.80%
总计	6.20%	22.10%	72.80%	26.00%	18.20%	3.20%	6.20%	15.80%

资料来源：Prismark，中国银河证券研究院

未来中国将继续保持全球 PCB 产业生态的核心制造中心地位。据 Prismark 预测，受 AI 基础设施、供应链重构与终端智能化等驱动，2030 年全球 PCB 市场产值将突破 1,230 亿美元，2025-2030 年五年复合年增长率约为 7.7%，其中 AI 服务器及高速网络系统，汽车电子（电动汽车及高级驾驶辅助系统领域）以及从长期来看具备先进人工智能功能的便携式智能消费电子设备，是其最重要增长驱动力，相应 2025-2030 年增长最强劲的领域也将是 HLC（18+）、封装基板及 HDI。全球供应链多元化布局（“中国+N”模式）并未对中国的 PCB 生产规模造成实质性削弱。东南亚、印度及部分西方国家的新增产能仍在持续释放，但中国 PCB 制造商通过在国内外同步扩张做出了应对，利用海外工厂服务敏感市场，同时在国内保持技术领先地位和规模优势。尽管地缘政治和贸易摩擦不断加剧，中国仍是全球 PCB 产业生态的核心制造中心。

表6：2025-2030 全球 PCB 产值增长率估算

产值增长率	多层板			HDI	封装基板	柔性板	其他	总计
	4-6 层	8-16 层	18 层以上					
美洲	4.40%	4.20%	7.70%	5.40%	17.40%	3.40%	2.20%	4.70%
欧洲	4.20%	5.50%	8.10%	6.60%	15.30%	1.60%	2.10%	4.40%
日本	2.50%	4.60%	9.30%	5.20%	11.80%	3.10%	1.50%	7.80%
中国	3.20%	6.20%	21.60%	9.10%	9.70%	3.90%	2.20%	7.00%
亚洲	3.50%	9.10%	29.10%	10.30%	11.10%	3.90%	6.40%	9.80%
总计	3.30%	6.50%	21.70%	9.20%	10.90%	3.80%	2.80%	7.70%

资料来源：Prismark，中国银河证券研究院

国产算力崛起进一步驱动中国 PCB 企业市占率提升。随着华为昇腾、寒武纪等国产 AI 芯片性能逐步提升，国内头部互联网厂商与运营商纷纷转向国产算力集群建设，为本土 PCB 企业创造了天然的近场优势。中国 PCB 企业凭借与客户的研发协同优势，能够迅速响应客户对高多层板及高阶 HDI 的定制化需求，实现了从跟随配套到联合定义的角色跃迁。

（三）被动元件：高容、芯片电感和钽电容需求有望加速增长

在 Token 经济爆发式增长的驱动下，AI 服务器与大模型训练集群正朝着超高算力密度方向演进，对被动元件提出了“量增质升”双重要求。量增指随着 GPU/TPU 模组数量激增，单台 AI 服务器的被动元件用量呈指数级上升。高端服务器中 MLCC 的搭载量较传统服务器增长一倍以上，以满足数百安培级电流的瞬态响应与滤波需求。质升则体现在性能指标的跨越式升级。面对芯片工作频率的提升与供电电压的微缩，普通被动元件已无法满足稳定性要求。高算力场景下，必须采用高容值、耐高温、低 ESR/ESL 特性的高端 MLCC，以及金属合金材质的电感，以应对纳秒级的电流波动与电磁干扰。

AI 服务器正成为高端 MLCC 增长的核心引擎。相比通用服务器，AI 服务器因搭载多颗 GPU/ASIC 及高功耗 CPU，电源管理复杂度大幅提升，对 MLCC 在数量、性能与可靠性上的要求显著提高。单机 MLCC 用量呈倍数级增长。AI 服务器整机 MLCC 用量可达普通服务器的 2-3 倍甚至更高，尤其是 CPU/GPU 周边供电、电源模块（VRM）、去耦与滤波电路，均需大量高容值、低 ESR、小尺寸 MLCC 以保证电源稳定与信号完整性。技术规格持续升级：为满足高频开关与高算力负载，厂商倾向采用高可靠性介质，并推进 0201、01005 等微型化封装，以适配高密板级布局；同时，耐高温、长寿命特性也成为关键指标。从市场结构看，村田、三星电机、TDK 等头部厂商主导高端 AI 服务器 MLCC 供应。AI 服务器的高功率密度与高可靠性需求，正在重塑 MLCC 的技术路线与市场格局，推动行业向高端化、定制化方向加速演进。

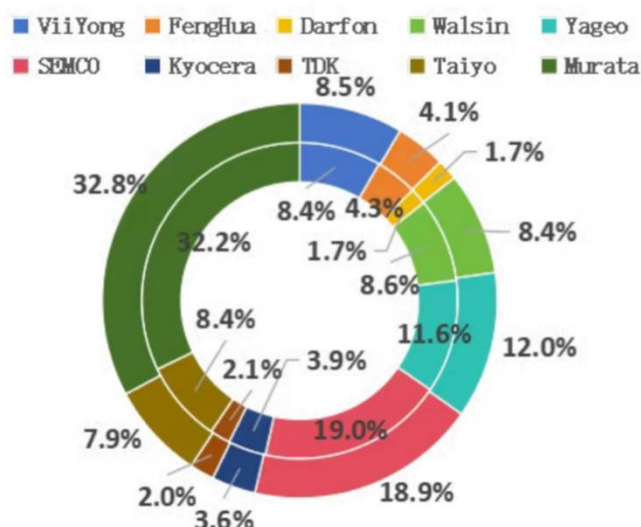
表7：全球主要 MLCC 应用产品年度需求总量趋势（单位：十亿颗）

终端目录	2024	2025	2026F	24/25	25/26 (F)
智能手机	1730	1760	1765	1.7%	0.3%
汽车	660	675	690	2.3%	2.2%
5G 基站	116	118	120	1.7%	1.7%
服务器（含 AI）	115	125	143	8.7%	14.4%
笔记本电脑	175	178	177	1.7%	-0.6%
TV/控制板	59	58	59	-1.7%	1.7%
游戏机	116	119	122	2.6%	2.5%
其他消费品	1360	1377	1380	1.3%	25.0%
总计	4331	4410	4456	1.80%	1.00%

资料来源：Trendforce，中国银河证券研究院

中国大陆企业在高容市场份额逐渐提升。在日韩巨头长期主导的高端市场中，国产 MLCC 厂商表现亮眼。根据 TrendForce2025 年第四季度排名数据，微容科技和风华高科的全球市占率稳居前十，在高端市场成为不可忽视的中国力量，与村田、三星电机、太阳诱电等全球主要厂商同台竞技。

图35: 2025Q3（内环）、Q4（外环）全球 MLCC 厂商市占率



资料来源: Trendforce, 中国银河证券研究院

AI 服务器功率快速提升, 带动新型电感需求。AI 服务器功率相较于普通服务器提升了 6~8 倍, 前期主流的芯片电感主要采用铁氧体材质, 但随着电源模块的小型化、低电压、大电流的发展趋势, 铁氧体材料受限于其饱和磁通密度低等条件制约, 已经很难满足后续发展需求, 而基于金属软磁材料开发的芯片电感具有低电压、大电流、小体积的优势, 更加符合未来大算力的应用需求。一体成型电感在服务器中主要作用是调压稳压, 给电源稳定供电提供有效保障, 在 CPU、GPU 及其他电源模块, 通过电磁转换, 实现电压调节, 确保不同负载获得稳定的电源供应。单台传统服务器的一体成型电感用量在 30-50 个之间, AI 服务器 (以配置 4 路 GPU 和 2 路 CPU 为例) 的一体成型电感用量在 80-120 个, 其中, 核心增量部分来自一体成型电感 (用于 PCIe/USB4 滤波) 和大电流电感 (用于 GPU 供电的芯片电感)。以英伟达 GB300 NVL72 服务器为例, 单台服务器使用一体成型电感数量在 4800-5000 个。

表8: AI 服务器各功能模块一体成型电感使用情况

功能模块	一体成型电感数量 (个)
CPU VRM	16-24
GPU	6-10
内存 DDR5	4-8
PCIe 插槽	6-12
芯片组/PCH	2-4
存储接口	4-6

资料来源: 田村 (中国) 企业管理有限公司, 中国银河证券研究院

钽电容凭借极高的体积效率和优异的阳极自愈特性, 成为 AI 服务器中 GPU/ASIC 供电电路的关键元件。随着 AI 服务器出货量的爆发, 钽电容市场正迎来量价齐升, 特别是在高算力、大电流、低电压的应用场景下, 其不可替代性日益凸显。其特点包括: 1、相比 MLCC 和铝电解电容, 钽电容在相同体积下能提供更大的电容量。这对于 AI 服务器中高密度的 GPU 板卡至关重要, 能有效节省 PCB 空间。2、钽电容具有较低的等效串联电阻, 在高频滤波和去耦方面表现优异, 能够快速响

应 GPU 瞬态负载变化，确保电源稳定性。3、固态钽电容没有液态电解质干涸的问题，且钽阳极具备一定的自愈功能，氧化膜受损处可重新形成，非常适合 AI 服务器 7x24 小时不间断运行的高可靠性要求。4、在较宽的温度范围内，其电容值和 ESR 变化较小，适应服务器机房可能出现的温度波动。

聚合物钽电容正加速取代传统二氧化锰钽电容，国产品牌迎来结构性机遇。相较于传统体系，导电聚合物电解质彻底消除了“热致击穿”引发的燃烧风险，完美契合 AI 数据中心对设备安全的严苛要求；同时，其极低的等效串联电阻与优异的高频响应特性，能够精准应对 GPU/ASIC 等大电流芯片在瞬态负载下的电压波动，显著提升电源完整性。长期以来，高端钽电容市场被美国基美（Kemet）、威世（Vishay）等巨头垄断，地缘政治与供应链安全考量正加速国产替代进程。国内头部企业凭借在军用领域积累的高可靠性技术底蕴，快速向民用高端市场渗透。

四、投资建议

硬件产业是 Token 经济学下的核心投资方向，我们看好半导体和电子器件方向，推荐兆易创新、中芯国际、生益科技、澜起科技、水晶光电。

表 9：重点公司盈利预测与估值

股票代码	股票名称	EPS			PE			投资评级
		2025A	2026E	2027E	2025A	2026E	2027E	
688981.SH	中芯国际	0.63	0.82	1.19	212.40	163.18	112.45	推荐
603986.SH	兆易创新	2.47	4.24	5.37	180.17	104.96	82.87	推荐
600183.SH	生益科技	1.37	2.32	3.25	73.40	43.34	30.94	推荐
002273.SZ	水晶光电	0.84	1.03	1.29	47.26	38.54	30.78	推荐
688008.SH	澜起科技	1.95	2.50	3.20	131.95	102.92	80.41	推荐

资料来源： 、中国银河证券研究院。备注：以 2026 年 5 月 20 日收盘价为基准计算。

五、风险提示

AI 技术迭代不及预期的风险。AI 大模型算法、推理效率与多模态能力的迭代进度若低于市场预期，会直接影响 Token 生成效率与应用落地节奏，拖累 AI 商业化普及速度，进而削弱全产业链的增长动能。

硬件产能扩张速度低于需求的风险。AI 芯片、存储、服务器等核心硬件产能扩张若跟不上指数级增长的 Token 需求，将引发供应链紧张、成本高企、交付延迟等问题，制约算力基础设施建设与下游 AI 应用规模化落地。

国产芯片替代进度不及预期的风险。AI 芯片在性能、生态适配与规模化应用上的突破若慢于预期，会影响国内算力供应链自主可控进程，延缓国产替代节奏，对国内 AI 算力产业安全与长期发展构成制约。

全球 AI 算力投资增速放缓的风险。全球云服务商与科技企业对 AI 算力基建的资本开支若出现增速回落，将直接降低上游硬件需求，导致行业景气度下行，对产业链出货、盈利与扩张空间形成明显压制。

图表目录

图 1: Token 的产生机制.....	3
图 2: AI 时代 Token 经济学框架.....	4
图 3: 全球大模型 Token 调用量.....	4
图 4: 全球煤炭消费量（千短吨）、用电量（十亿千瓦时）.....	4
图 5: 不同模型生成 Token 的成本对比.....	5
图 6: 当前主流 AI 芯片算力对比图.....	6
图 7: 国内外代表性大模型每百万 Token 价格.....	7
图 8: Token 产业链情况.....	8
图 9: Token 生产上游产业链情况.....	8
图 10: Token 产业链中游参与方.....	9
图 11: Token 产业链下游消费场景.....	10
图 12: NVIDIA 在 AI 推理领域的极致性能优势.....	10
图 13: CSP 预期资本开支情况.....	11
图 14: 2023-2026 年 AI 服务器出货量预估.....	11
图 15: 海外 CSP 厂商资本开支投入情况（单位：亿美元）.....	11
图 16: 全球半导体市场规模及增长.....	13
图 17: 预测 DRAM 价格趋势.....	13
图 18: 预测 NAND 价格趋势.....	13
图 19: AI 芯片对于 HBM 的需求情况.....	14
图 20: NAND Flash 在不同应用领域占比.....	15
图 21: HBM 将成为 2026 年存储器行业产能投资的焦点.....	15
图 22: DRAM 在 DDR5 的技术变化路径.....	16
图 23: NAND 厂商技术路径.....	16
图 24: 英伟达数据中心业务持续高增长.....	18
图 25: 博通 AI 芯片业务高增长.....	18
图 26: 服务器成本构成.....	18
图 27: 2020-2026 年中国 AI 芯片市场规模(单位:亿人民币).....	18
图 28: 被列入“实体清单”的 AI 芯片企业数量.....	19
图 29: 美国商务部工业和安全局对华高算力芯片出口限制.....	19
图 30: 2024 年中国 AI 芯片出货量统计.....	19
图 31: 中国日均 Token 调用量快速增长.....	21
图 32: 生成式 AI 推理成本场景预测.....	22
图 33: AI 芯片的封装技术路径.....	23
图 34: AI 推动全球 PCB 行业规模持续高增.....	23
图 35: 2025Q3（内环）、Q4（外环）全球 MLCC 厂商市占率.....	26

表 1: OPEN AI 和互联网大厂的合作情况.....	12
表 2: 2025 年存储芯片涨价节奏	13
表 3: 2025 年存储芯片厂商扩产节奏	17
表 4: 国内市场对于 AI 支持政策情况	20
表 5: 2024-2025 全球 PCB 产值增长率估算.....	24
表 6: 2025-2030 全球 PCB 产值增长率估算.....	24
表 7: 全球主要 MLCC 应用产品年度需求总量趋势（单位：十亿颗）	25
表 8: AI 服务器各功能模块一体成型电感使用情况	26
表 9: 重点公司盈利预测与估值	26

分析师承诺及简介

本人承诺以勤勉的执业态度，独立、客观地出具本报告，本报告清晰准确地反映本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与、未来也将不会与本报告的具体推荐或观点直接或间接相关。

高峰：电子行业首席分析师，北京邮电大学电子与通信工程硕士，吉林大学工学学士。2 年电子实业工作经验，6 年证券从业经验，曾就职于渤海证券、国信证券、北京信托证券部。2022 年加入中国银河证券研究院，担任电子团队组长，主要从事硬科技方向研究。

电子行业分析师：王子路、钱德胜、刘来珍。

免责声明

本报告由中国银河证券股份有限公司（以下简称银河证券）向其客户提供。银河证券无需因接收人收到本报告而视其为客户。若您并非银河证券客户中的专业投资者，为保证服务质量、控制投资风险、应首先联系银河证券机构销售部门或客户经理，完成投资者适当性匹配，并充分了解该项服务的性质、特点、使用的注意事项以及若不当使用可能带来的风险或损失。

本报告所载的全部内容只提供给客户做参考之用，并不构成对客户的投资咨询建议，并非作为买卖、认购证券或其它金融工具的邀请或保证。客户不应单纯依靠本报告而取代自我独立判断。投资者应自主作出投资决策并自行承担投资风险，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。银河证券认为本报告资料来源是可靠的，所载内容及观点客观公正，但不担保其准确性或完整性。本报告所载内容反映的是银河证券在最初发表本报告日期当日的判断，银河证券可发出其它与本报告所载内容不一致或有不同结论的报告，但银河证券没有义务和责任去及时更新本报告涉及的内容并通知客户。银河证券不对因客户使用本报告而导致的损失负任何责任。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的银河证券网站以外的地址或超级链接，银河证券不对其内容负责。链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

银河证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。银河证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

银河证券已具备中国证监会批复的证券投资咨询业务资格。除非另有说明，所有本报告的版权属于银河证券。未经银河证券书面授权许可，任何机构或个人不得以任何形式转发、转载、翻版或传播本报告。特提醒客户及公众投资者慎重使用公众媒体刊载的证券研究报告。

本报告版权归银河证券所有并保留最终解释权。

评级标准

评级标准	评级	说明
评级标准为报告发布日后的 6 到 12 个月行业指数（或公司股价）相对市场表现，其中：A 股市场以沪深 300 指数为基准，新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准，北交所市场以北证 50 指数为基准，香港市场以恒生指数为基准。	行业评级	推荐：相对基准指数涨幅 10% 以上
		中性：相对基准指数涨幅在 -5%~10% 之间
		回避：相对基准指数跌幅 5% 以上
	公司评级	推荐：相对基准指数涨幅 20% 以上
		谨慎推荐：相对基准指数涨幅在 5%~20% 之间
		中性：相对基准指数涨幅在 -5%~5% 之间
		回避：相对基准指数跌幅 5% 以上

联系

中国银河证券股份有限公司 研究院	机构请致电：	
深圳市福田区金田路 3088 号中洲大厦 20 层	深广地区：	苏一耘 0755-83479312 suyiyun_yj@chinastock.com.cn 程曦 0755-83471683 chengxi_yj@chinastock.com.cn
上海浦东新区富城路 99 号震旦大厦 31 层	上海地区：	林程 021-60387901 lincheng_yj@chinastock.com.cn 李洋洋 021-20252671 liyangyang_yj@chinastock.com.cn
北京市丰台区西营街 8 号院 1 号楼青海金融大厦	北京地区：	田薇 010-80927721 tianwei@chinastock.com.cn
公司网址：www.chinastock.com.cn		