

存储行业巨擘，AI 时代大有可为

核心观点

美光科技是全球存储行业巨头，已在 DRAM 和 NAND Flash 领域建立了显著的市场份额和技术优势，是全球仅有的三大 HBM 原厂之一。AI 时代下，HBM 成为整个存储领域的重要增长点，美光科技跳过 HBM3 直接成功研发 HBM3E 实现弯道超车，进度追平海力士，先于三星电子获得英伟达验证与订单。本轮全球半导体周期于 2023 年触底，2024 年开始有所回升，AI 需求红利仍将延续，虽然短期内受 PC 和智能手机端需求疲软和暂时的买方去库存策略影响，DRAM 涨幅受限、NAND 略有下调，但中长期来看，随着 AI 应用在端侧设备中的升级以及大模型逐步在端侧进行本地化部署，单位设备 DRAM 与 NAND 容量将持续增长，同时端侧设备的 AI 升级也将加速新一轮换机潮到来。

摘要

全球存储行业巨擘，AI 时代 HBM 大有可为。美光科技 1978 年成立于美国爱达荷州，专注于存储芯片这一半导体细分领域，通过多年并购，成功将业务版图扩展至 DRAM、NAND 闪存、CMOS 图像传感器等多样化领域，市场份额也稳坐存储市场前三把交椅，DRAM 与 NAND Flash 为其核心业务。AI 时代下，匹配高算力的高带宽存储 HBM 成为整个存储领域的重要增长点，目前仅有海力士、美光和三星三家厂商能够供给。2023 年美光直接跳过 HBM3，与台积电合作研发 HBM3E，2024 年一季度，其率先宣布 HBM3E 8Hi 量产并供货英伟达 H200 产品，与海力士节奏同步，领先三星。三家 HBM 产品特性各有不同，形成互补且竞争的局面，其中美光的产品在功耗和广泛兼容性上表现突出，对 HBM 在 AI、HPC 和图形处理等领域的应用有独到见解。

短期传统 DRAM 整体偏中性，NAND 略有疲软。短期来看，传统 DRAM 虽然产能受 HBM 挤出导致供给有所改善，但传统需求 PC 和智能手机依旧疲软，叠加买方去库存策略，整体价格改善不明显，未来的价格回升一定程度依赖于 PC 和手机市场的回暖。同时 DDR5 替代 DDR4 的进程正处于加速阶段，DDR5 的供给厂商更为集中，议价能力更强，预计也将对美光 DRAM 整体价格有所支撑。NAND 受到消费电子疲软的冲击更大，除了受 AI 服务器推动的企业级 SSD 之外，其他价格略有下降。

美光科技 (MU.O)

首次评级

买入

崔世峰

cuishifeng@csc.com.cn

SAC 编号:S1440521100004

SFC 编号:BUI663

于伯韬

yubotao@csc.com.cn

SAC 编号:S1440520110001

SFC 编号:BRR519

发布日期：2024 年 11 月 04 日

当前股价：99.73 美元

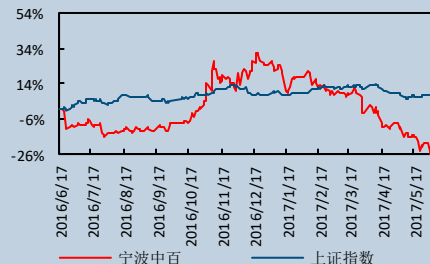
目标价格 6 个月：142.75 美元

主要数据

股票价格绝对/相对市场表现 (%)

1 个月	3 个月	12 个月
25.27/20.38	-6.99/-9.72	61.09/24.47
12 月最高/最低价 (美元)		153.45/64.53
总股本 (万股)		110,874.27
流通股本 (万股)		110,874.27
总市值 (亿美元)		1,232.37
流通市值 (亿美元)		1,232.37
近 3 月日均成交量 (万)		2508.59
主要股东		
The Vanguard Group, Inc.		8.30%

股价表现



相关研究报告

24.08.17 【中信建投海外研究】台积电 (TSM): 半导体行业复苏，大象起舞

的

HBM 渗透率正在提升，AI 需求保持强劲，但对远期供需持谨慎态度。AI 模型对高带宽存力的需求导致 HBM 需求旺盛，但受制于 TSV 产能，HBM 供不应求，维持高溢价，2025 年 HBM3e 产能已经大量售出。尽管三大厂商都有扩产，但新加产能最早 2025 年第二季度才能释放。由于量产 HBM3e 12hi 的学习曲线长，三大厂商能否如期落实 HBM3e 产能还有待验证。同时中长期来看，受制于超大规模企业资本支出的不确定性，以及模型从训练转入推理阶段之后，AI 芯片对 HBM 的高速拉动能否维持还有待验证，厂商需要在短期扩产以满足需求和中长期过度建设导致供给过剩之间取得平衡。

本轮半导体存储周期尚未结束，端侧传统需求可能成为未来市场的主要驱动力。尽管存在上述担忧，但中期而言，半导体存储的前景依然光明。首先，随着厂商减产计划的完成和 HBM 的挤出效应，整体库存水平将逐渐回复到健康状态，同时本次短期的买方去库存行为难以长期持续，本轮半导体存储周期尚未结束。其次，随着 AI 设备渗透率的逐渐上升，以及端侧设备（PC 和智能手机）在 LLM 中的角色从客户端转向模型推理本地化，单位设备中 DRAM 和 NAND 的容量需求将持续提升，同时 AIPC 和 AI 手机也将带动整体需求量的增长。因此 AI 时代下，端侧传统需求在未来存储市场可能依然是主要驱动力。

盈利预测与估值：我们预计 2025-2026 财年美光营业收入分别为 382.97 亿和 458.62 亿美元，同比增长 52.51% 和 19.75%，Non-GAAP 毛利率分别为 42.64% 和 48.40%，Non-GAAP 净利润分别为 100.54 亿和 139.85 亿美元，同比增长 577% 和 39%，对应净利率分别为 26.25% 和 30.49%。考虑到美光在全球存储以及 HBM 领域的优势地位，按照 PB 估值法，给予其 2025 财年 3X PB，对应目标价 140.42 美元/ADS，按照 PE 估值法，给予其 2025 财年 16X PE，对应目标价 145.08 美元/ADS。首次覆盖，给予“买入”评级，综合 PB 和 PE 估值，给予目标价 142.75 美元/ADS。

风险提示：全球宏观经济超预期下行；地缘政治风险；生成式 AI 发展低预期，AI 算力芯片需求低预期，Mag7 资本开支低预期；AI PC 以及 AI 手机发布低预期，PC 及手机换机进度低预期；汽车 SoC 需求持续低预期，终端 Tier1 库存去化低预期；存储行业供给超预期增长，存储价格超预期下行；行业竞争加剧，客户流失，来自三星、海力士的竞争；HBM 制程迭代低预期，新品发布低预期；美联储降息进程低预期；其他影响半导体行业的宏观和中观风险。

目录

一、全球存储行业巨擘，产业并购实现逆势扩张	1
（一）存储作为半导体行业风向标，是最标准化的科技大宗品	1
（二）产业并购造就美光行业龙头地位	3
二、DRAM 及 HBM 呈三足鼎立之势，NAND 格局更为分散	5
（一）DRAM：DDR5 渗透率不断提升，美光 25 年即将量产 1- γ 制程	5
（二）NAND：堆叠层数不断提升，行业格局更为分散	8
（三）HBM：AI 时代大有可为，供给受限于 TSV 产能释放情况	11
（四）竞争格局：三星 VS 海力士 VS 美光	15
三、AI 引领下游需求增长，消费电子及其他端侧产品需求有待接棒	19
（一）AI 算力催化 HBM 需求	19
（二）PC、手机需求短期仍有待提振	21
（1）PC 市场：AIPC 加速换机周期	21
（2）智能手机：AI 赋能是主要驱动力	23
（三）智能汽车市场远期潜力巨大	26
（四）AI 算力需求景气支撑 HBM 价格，消费电子景气疲弱令传统 DRAM 及 NAND 价格短期承压	27
四、盈利预测和估值	33
风险提示	35

图目录

图 1: 半导体存储分类图	2
图 2: 存储是半导体第二大细分赛道	2
图 3: 存储行业波动性大于半导体行业整体水平	2
图 4: DRAM 市占率美光排名第三（2023）	3
图 5: NAND 市占率美光排名第五（2023）	3
图 6: 美光历史沿革	4
图 7: 美光收入构成（by technology）	5
图 8: F24Q4 美光收入构成（by technology）	5
图 9: DDR4 依然是当前 DRAM 市场主流	6
图 10: 从 DDR1 到 DDR5 的演变：电压更低，速率翻倍，容量翻倍	6
图 11: 头部存储厂商 DRAM 制程演进图	7
图 12: LPDDR2/3 和 LPDDR4 原理比较	7
图 13: GDDR 制程演进比较	7
图 14: 头部存储厂商 CAPEX 情况	8
图 15: 头部厂商近年来 DRAM 收入变化情况	8
图 16: NAND 从 2D 到 3D 结构演进	9
图 17: 2016-2023 年全球 SSD 出货量结构	9

图 18:各原厂 NAND 制程演进图	9
图 19:2021-2024 年 NAND 产量（按堆叠层数划分）	10
图 20:24Q2 美光 NAND 市占率升至第四	11
图 21: NAND 容量在迅速增长	11
图 22:HBM 与 GDDR 原理比较	12
图 23:HBM 性能相比 GDDR、LPDDR 等有明显提升	12
图 24:HBM 主要厂商技术路线演进图	13
图 25: HBM 各产品产能分布预测	13
图 26:HBM 市场呈现三足鼎立格局	13
图 27:TSV 原理示意图	14
图 28:HBM 出货量&市场规模统计及预测	15
图 29:美光 HBM 及 GDDR 路线演进图	16
图 30:海力士 6-phase 和 4-phase RDQS（HBM3）方案比较	17
图 31:三星对于存储技术未来趋势和展望	18
图 32: CNBU 和 MBU 部门贡献美光主要收入	19
图 33: NVIDIA 及 AMD AI 芯片发展进程及 HBM 规格比较	20
图 34: 全球服务器出货量及预测	20
图 35: AI 服务器出货量及预测	20
图 36: 微软 CAPAX 及增速	21
图 37: 谷歌 CAPAX 及增速	21
图 38: meta CAPAX 及增速	21
图 39: 亚马逊 CAPAX 及增速	21
图 40: 全球 PC 季度出货量及增速	22
图 41: 全球 PC 年度出货量及增速	22
图 42: 全球 PC 市场出货量（百万台）	22
图 43: 全球 PC 市占率（%）	22
图 44:PC 对于存储配置的需求与日俱增	23
图 45: 全球智能手机年度出货量	23
图 46: 全球智能手机厂商市场份额变化	23
图 47: Galaxy S24 的主要 AI 功能	24
图 48: Apple Intelligence	24
图 49:AIGC 对手机存储的影响	25
图 50: 智驾行业内存技术演进	26
图 51: 常见智能驾驶 soc 芯片内存	27
图 52: 全球主要云厂商资本支出（市场需求波动剧烈）	28
图 53:2024-2025 年 DRAM 各领域出货预测	28
图 54:2024-2025 年 DDR4 价格和供需缺口预测	29
图 55: 3Q24 与 4Q24 名 DRAM 产品价格预测	29
图 56: 2024-2025 年 SSD 出货量预测	30
图 57: 2024-2025 年 NAND 价格和供需缺口预测	30
图 58: 3Q24 与 4Q24 名 NAND 产品价格预测	31

图 59: 本轮半导体存储周期尚未结束.....	32
图 60:美光 PB.....	32
图 61:三星、海力士、美光 PB 比较	32
图 62:美光分业务收入预测	33

表目录

表 1:HBM/TSV 产能预测	14
表 2: 2024-2028 年全球 AI PC 出货量预测.....	22
表 3: Apple Intelligence 带来的功能更新	24
表 4:美光盈利预测（单位：百万美元）	34
表 5:公司营收敏感性测算	35

一、全球存储行业巨擘，产业并购实现逆势扩张

（一）存储作为半导体行业风向标，是最标准化的科技大宗品

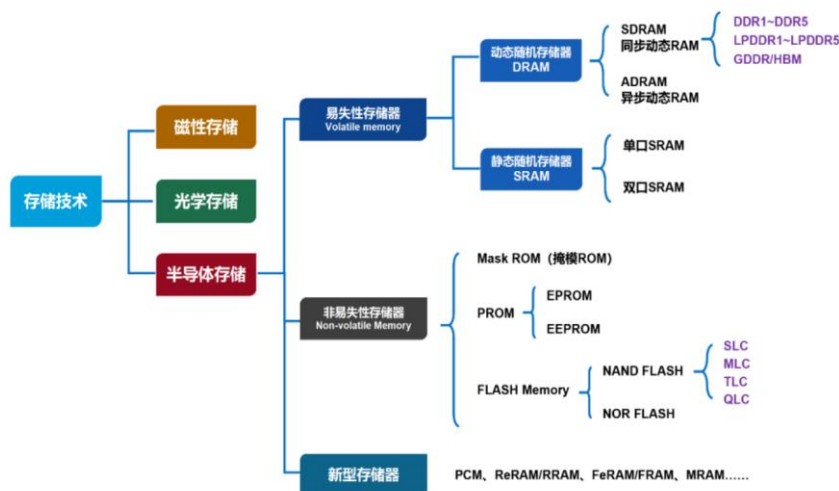
存储器是计算机系统的核心组件之一，对整体性能至关重要。存储器是计算机中接收和存储数据，并能根据控制命令提供所存数据的基本硬件，存储器的特性是影响整个计算机系统最大吞吐量的核心因素。按照用途，存储器可分为主存储器（内存）和辅助存储器（外存）。主存储器与 CPU 一起构成主机，用于存放计算机运行时随时需要使用的程序和数据，一切数据要被 CPU 操作都必须先装入主存，主存的工作速度较快，存储容量较小，主要应用类型为动态随机存储器（DRAM）；辅助存储器则用于长期存放 CPU 运行时暂时不用的大量程序和数据。存储器中存在大量存储元，每个存储元以两种稳定的物理或化学状态的变化表示二进制信息，一组存储元组成一个存储单元，一个存储单元通常可以存放一个字节并具备一个唯一的地址以供数据存取。

半导体存储是现代存储的核心。按照存储介质，现代存储器可以分为磁性存储器、光学存储器与半导体存储。与传统机械式存储介质（如磁带、硬盘）相比，半导体存储器具有访问速度快、功耗低、易携带等优点，是现代计算机中存储设备的核心。半导体存储可以分为易失性存储（RAM）与非易失性存储（ROM）。易失性存储访问速度快，便于 CPU 高效地从存储器的任意存储单元读取或写入数据，断电后会丢失所存储数据。易失性存储分为动态随机存储器（DRAM）和静态随机存储器（SRAM），DRAM 是目前计算机、手机内存的主流方案，需要在数据改变或者断电前进行周期性充电，否则将丢失数据；SRAM 在电力持续供应的状态下，不需要定期刷新，但功耗大、集成度低且价格高昂，主要用于 CPU 缓存与辅助缓存。非易失性存储是一种只读存储器，在电源不供电的状态下仍能保存，其中的数据不能被常规方式修改或写入。非易失性存储根据其特性和可编程性可以分为掩模 ROM、ERROM 和 EEPROM 等，Flash 是一种特殊的 EEPROM，具有更快的擦写速度和更大的存储容量，其主流产品为 NOR Flash 和 NAND Flash。NOR Flash 是代码型闪存芯片，应用程序不必把代码读到系统 RAM 中，而可以直接在 Flash 闪存内运行，但相较 NAND 擦写速度慢且体积更大，市场占比较小，用于家用电器、穿戴式设备等。NAND Flash 为数据型闪存芯片，可以实现大容量存储，市场占比非常高。新兴非易失性存储技术方面，PCM、MRAM、RRAM 等正在开发中，旨在提供比传统存储技术更高的性能和更低的功耗。

在存储行业中，DRAM 市场规模占比最大，2023 年占比约 56%。DRAM 是存储器细分产品中的营收占比最大的产品，主要分为 DDR、LPDDR、GDDR 及新型存储 HBM。DRAM 下游应用以移动电子产品为首，服务器次之，个人计算机也占有一定比例。DDR 为目前 DRAM 占比最高的品类。DDR 主要应用于个人计算机、服务器与云计算领域，现已经发展到 DDR5，从 DDR1 到 DDR6 主要在传输速率与电压方面进行迭代，现行主流为 DDR4 和 DDR5。

NAND Flash 是存储市场的另一核心，2023 年市场规模约 41%。根据存储密度、耐久性与成本效益等技术类型，NAND Flash 可以分为 SLC（单层存储单元）、MLC（双层存储单元）、TLC（三层存储单元）、QLC（四层存储单元），目前市场主流产品为 TLC 和 QLC。根据产品应用划分，NAND Flash 广泛用于固态硬盘（SSD）、嵌入式存储产品（eMMC）、SD 卡和 U 盘中，主要应用领域包括嵌入式存储产品（eMMC）、计算机的消费级固态硬盘（SSD）及服务器市场的企业级或数据中心级固态硬盘（SSD），其合计规模占 NAND 市场 85% 左右。

图 1: 半导体存储分类图



数据来源：芯师爷，中信建投证券

存储行业是半导体第二大细分赛道，是半导体行业风向标。根据世界半导体贸易统计组织（WSTS）数据，2023 年存储行业市场规模为 922.88 亿美元，在整个半导体产业中占比排名第二，仅次于逻辑芯片的 1785.89 亿美元。同时，存储行业与半导体都表现出明显的价格周期，且在波动周期上二者表现出同步性，但存储行业市场规模的波动水平高于整个半导体行业，弹性更强，是整个半导体行业的风向标。

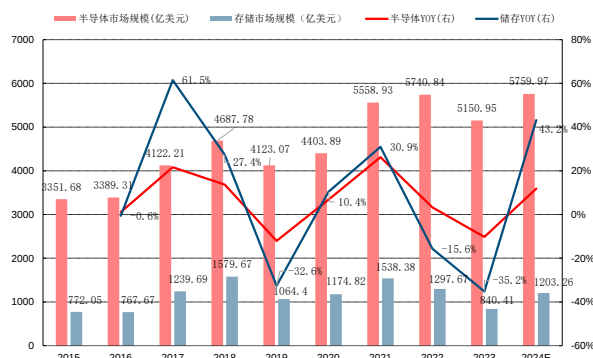
存储具备大宗商品特性，价格反映其强周期性。存储芯片具备科技大宗产品特性，标准化程度高，产品间可替代性强，价格透明度高；同时供给厂商呈现寡头竞争格局，在产能规划方面倾向于选择相互跟随策略，放大了周期性的扩产-去库存现象。存储芯片的大宗商品特性也决定其将表现出更强的买方市场特性，价格能有效反映供需关系尤其是需求端的变动。存储价格有合约价格和现货价格两种。合约价指一线 OEM、二线 OEM 和原厂代理商向三星等原厂订货的价格，通常每月在 5-10 号和 20-25 号分别更新一次，能反映大型 OEM 和原厂对于后市价格与供求关系的预期协调。现货价是指存储器现货市场上流通的价格，一般是各类现货由二线 OEM 和代理商流入市场，加之各通路炒卖而形成现货市场。现货价格受市场供求关系变化、市场预期等因素影响，变化较大，可以即时反映市场变化。基于 2015-2023 的 WSTS 数据，存储行业一轮价格周期约为 3-4 年，本轮周期在 2023 年触底，2024 年开始逐渐修复。

图 2: 存储是半导体第二大细分赛道

Spring 2024	Amounts in US\$M			Year on Year Growth in %		
	2023	2024	2025	2023	2024	2025
Americas	134,377	168,062	192,941	-4.8	25.1	14.8
Europe	55,763	56,038	60,901	3.5	0.5	8.7
Japan	46,751	46,254	50,578	-2.9	-1.1	9.3
Asia Pacific	289,994	340,877	382,961	-12.4	17.5	12.3
Total World - \$M	526,885	611,231	687,380	-8.2	16.0	12.5
Discrete Semiconductors	35,530	32,773	35,310	4.5	-7.8	7.7
Optoelectronics	43,184	42,736	44,232	-1.6	-1.0	3.5
Sensors	19,730	18,265	19,414	-9.4	-7.4	6.3
Integrated Circuits	428,442	517,457	588,425	-9.7	20.8	13.7
Analog	81,225	79,058	84,344	-8.7	-2.7	6.7
Micro	76,340	77,590	81,611	-3.5	1.6	5.2
Logic	178,589	197,656	218,189	1.1	10.7	10.4
Memory	92,288	163,153	204,281	-28.9	76.8	25.2
Total Products - \$M	526,885	611,231	687,380	-8.2	16.0	12.5

数据来源：WSTS，中信建投证券

图 3: 存储行业波动性大于半导体行业整体水平



数据来源：WSTS，中信建投证券

（二）产业并购造就美光行业龙头地位

美光科技是专注于存储行业的半导体龙头。美光科技成立于 1978 年，总部位于美国爱达荷州，1984 年在纳斯达克证券交易所上市。美光科技是全球领先的半导体解决方案供应商，负责设计、制造和销售包括 DRAM、NAND 闪存、CMOS 图像传感器、其他半导体组件和存储模块在内的多种产品。不同于三星等其他半导体龙头，美光科技专注于存储芯片这一细分领域，DRAM 与 NAND Flash 为其核心业务，2023 年 DRAM 和 NAND 营收额分别为 109.8 亿美元和 42.1 亿美元，占总营收比重分别为 71%、27%，即 DRAM 与 NAND 两项业务加总贡献了美光约 98% 的收入。金士顿、惠普与英特尔是美光大客户，金士顿与惠普主要购买 DRAM 产品，占销售额比例分别为 10% 与 9%，英特尔主要购买通过 IM Flash 销售的 NAND 产品，占销售额比例为 8%。美光科技的产品广泛应用于计算机、消费电子、服务器、移动设备、汽车和工业等多个领域。按照销售额，其销售下游领域分布为：30% 计算机和图形（包括台式电脑、笔记本电脑和工作站）；20% 智能手机；20% SSD 和其他存储；10% 服务器；10% 为汽车、工业、医疗和其他嵌入式芯片。

美光科技在 DRAM 与 NAND 市场的供给端占据龙头地位。DRAM 与 NAND 都是集中度非常高的市场，DRAM 市场尤甚。DRAM 市场供给侧竞争格局稳定，基本被三星、海力士与美光垄断，三者合计市占率近 95%。其中三星占比最高，份额稳定在 40% 以上，2023 年海力士市占率为 28.3%；而美光持续排名第三，2023 年市场占有率占比为 23.1%。相较 DRAM 而言，NAND 市场竞争更为激烈、集中度相对较低，但就竞争格局而言，层次也较为分明。NAND 市场中三星依旧稳坐市占率第一的位置，市占率持续领先其他厂商，2023 年市占率为 34.1%，其次为铠侠，2023 年市占率为 18.8%；而美光、海力士与西部数据的市占率形成第三梯队，三者市占率总体相差较小，均在 10%-15% 的范围中波动。

图 4:DRAM 市占率美光排名第三（2023）

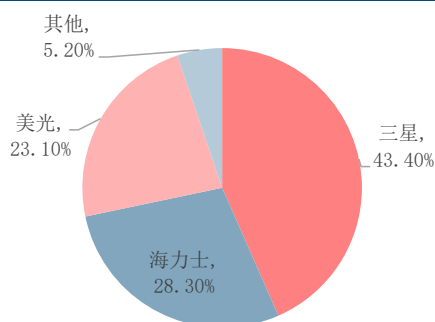
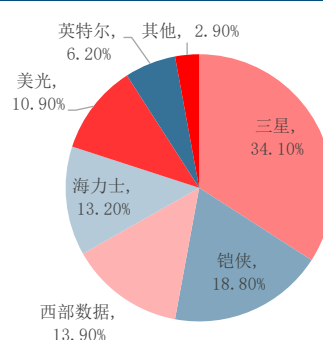


图 5:NAND 市占率美光排名第五（2023）



数据来源：HDIN，中信建投证券

数据来源：HDIN，中信建投证券

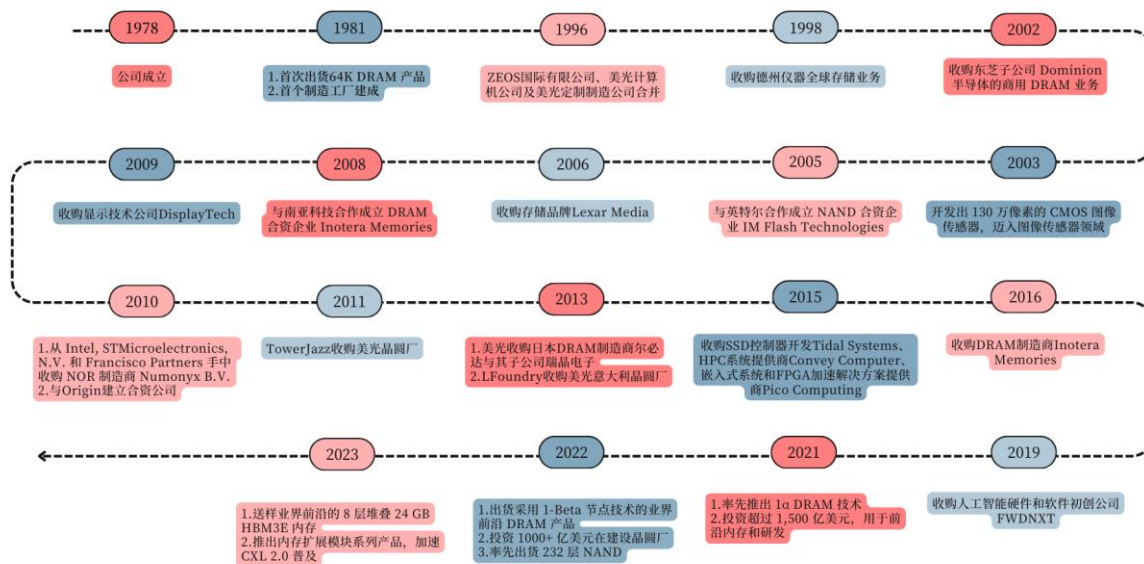
美光通过一系列战略性收购和技术创新，确立了其在全球半导体存储器领域的领导地位。美光科技从初创公司发展为半导体存储龙头企业，整体发展历史可以分为三个阶段，1978-1980 年是公司发展的早期，1990-2005 年是公司快速扩展与多元化时期，而 2005 年之后公司着力于技术创新，与行业巨头竞争市场。

美光发展早期即奠定 DRAM 领域技术领先的基调。美光科技由沃德·帕金森、乔·帕金森、丹尼斯·威尔逊和道格·皮特曼于 1978 年在爱达荷州博伊西市创立，最初专注于开发和制造 DRAM。1980 年公司即建成了第一个制造工厂，并推出了当时世界上最小的 256K DRAM，在 DRAM 领域取得了重大技术突破。

之后公司通过不断合并与收购拓展业务版图与提升市场份额。1996 年，美光科技通过与 ZEOS 国际有限公司、美光计算机公司及美光定制制造公司的合并，进入计算机系统集成领域，增强了其在 PC 市场的影响力。1998 年，公司收购了德州仪器的全球存储业务，2002 年收购了东芝子公司 Dominion 半导体的商用 DRAM 业务，大大提高了其在存储芯片市场的市场份额和技术实力。2003 年，美光科技推出了 130 万像素的 CMOS 图像传感器，正式进入图像传感器领域，开启了多元化发展战略。2005 年，美光科技与英特尔合作成立了 NAND 合资企业 IM Flash Technologies，双方共享技术和资源，推动其在 NAND Flash 技术与市场的发展。2006 年，公司收购了存储品牌 Lexar Media，扩展了其在消费存储市场的产品线。2008 年，公司与南亚科技合作成立 DRAM 合资企业 Inotera Memories，进一步加强了在 DRAM 领域的市场影响力。2010 年，从 Intel、STMicroelectronics、N.V. 和 Francisco Partners 手中收购 NOR 制造商 Numonyx B.V.，扩大了其在非易失性存储器市场的份额。2013 年，美光科技收购日本 DRAM 制造商尔必达及其子公司瑞晶电子这一重大举措，显著增强了其在全球 DRAM 市场的份额。2015 年，美光科技连续收购了 SSD 控制器开发商 Tidal Systems、HPC 系统提供商 Convey Computer 和嵌入式系统及 FPGA 加速解决方案提供商 Pico Computing，进一步增强了美光科技在 SSD 控制器、高性能计算和嵌入式系统领域的技术优势。2016 年，收购 Inotera Memories，进一步巩固了美光科技在 DRAM 市场的领导地位。

近年来美光科技着力于前沿领域与技术创新。2019 年，美光科技收购人工智能硬件和软件初创公司 FWDNXT，布局 AI 领域。2021 年，推出 1- α DRAM 技术，并宣布投资超过 1,500 亿美元用于前沿内存和研发。2022 年，出货采用 1- β 节点技术的 DRAM 产品，并投资 1000+亿美元建设晶圆厂，同时率先出货 232 层 NAND。2023 年，送样英伟达 8 层堆叠 24GB HBM3E，并推出内存扩展模块系列产品，加速 CXL2.0 普及。

图 6:美光历史沿革



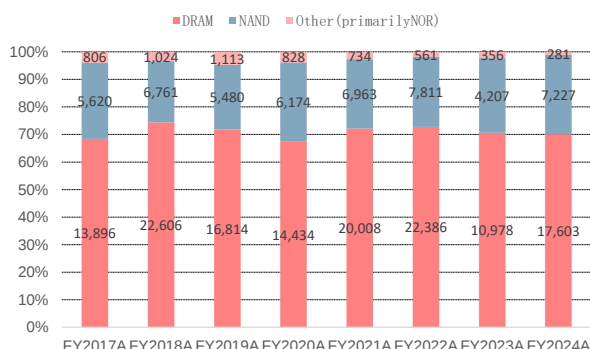
数据来源：公司官网，中信建投证券

请务必阅读正文之后的免责条款和声明。

二、DRAM 及 HBM 呈三足鼎立之势，NAND 格局更为分散

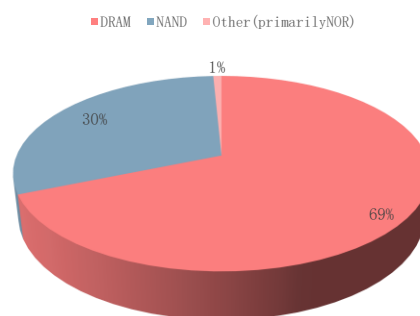
美光科技专注于存储领域，收入构成稳定，DRAM 与 NAND Flash 为其核心业务，分别贡献近七成与三成的收入，仅余 1%-2% 收入来自于 NOR Flash、CMOS 图像传感器与其他存储产品。

图 7: 美光收入构成 (by technology)



数据来源: Wind, 中信建投证券

图 8: F24Q4 美光收入构成 (by technology)



数据来源: Wind, 中信建投证券

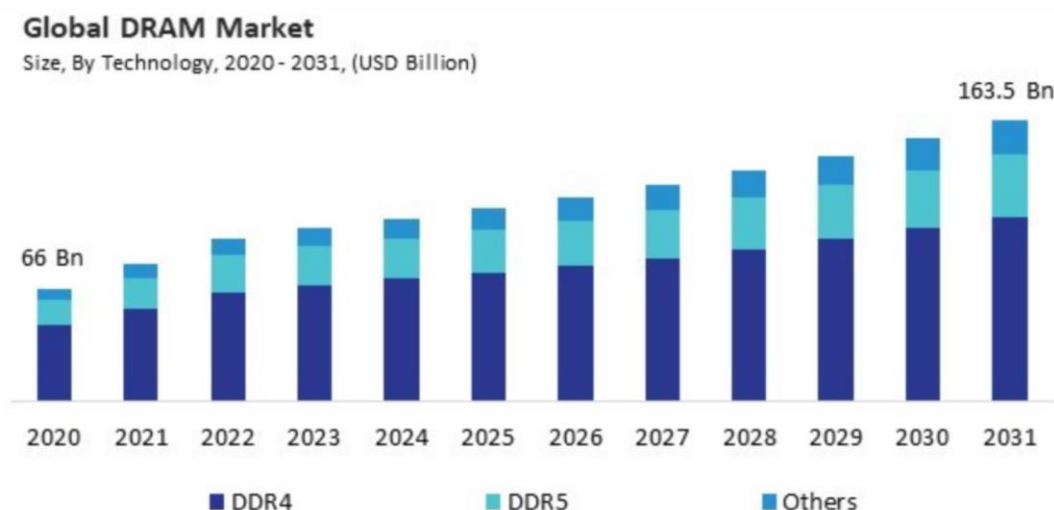
(一) DRAM: DDR5 渗透率不断提升，美光 25 年即将量产 1- γ 制程

目前主流的 DRAM 产品为通用 DDR，DDR 也从 DDR1 向 DDR6 不断演进。DDR 全称为 DDR SDRAM，即双倍数据率同步动态随机存取存储器，特点是走线数量多，速度快，是一种性能与成本的平衡解决方案，其发展趋向是逐步提高频率、增强带宽。DDR 是从 SDRAM 发展而来，其核心技术点为双沿传输和预取 Prefetch。双沿传输是指允许在每个时钟周期的上升沿和下降沿读出数据，从而使数据传输速率相对 SDRAM 翻倍，地址与控制信号则在时钟上升沿进行数据判断。预取 Prefetch 是指在内存控制器发出读取命令后，DRAM 芯片内部会预先读取多个数据单元，并将它们存储在内部缓存中，以便后续的读取操作可以更快地访问这些数据。预取技术减少了内存访问的延迟，提高了内存的整体性能。从 DDR1 到 DDR5，电压、预取深度与频率不断迭代，DDR1 采用了 SSTL2 标准的 2.5V 电压，预取为 2 个数据单元，频率为 200-400MHz，容量在 64M~1GB，随后每一任的迭代都在上一代的基础上降低电压，翻倍预取深度与频率，容量也随之出现倍数级的增长。其中值得注意的是，从 DDR1 到 DDR2 舍弃了传统 TSOP，开启了内存 FBGA 封装，缓解了寄生电容和阻抗匹配问题，增强稳定性；从 DDR3 到 DDR4 引入了差分信号技术、三维堆叠封装技术、温度补偿自动刷新和数据总线倒置技术，增大了单位芯片容量、显著降低了功耗。当下市场上最新一代的产品是 DDR5，其改善了 DIMM 的电源架构与双通道架构，使得其频率达到 3.2-6.4GT/s；电压降低至 1.1V，命令/地址 (CA) 信号从 SSTL 变为 PODL，其使引脚处于高电平状态时会消耗静态功率；预取深度达到 16，极大地提高了并发性。同时以三星为代表，主要厂商也在开发 DDR6，或将在 2025 年前后发布，DDR6 传输速率将从 8.8 Gbps 起步，最高可达 17.6 Gbps，带宽也将是 DDR5 的两倍。

DDR 现正经历 DDR5 对 DDR4 的替代过程。DDR 现正经历 DDR5 对 DDR4 的替代过程。据 Yole 分析，两代内存之间的过渡时间大概需要两年，尽管目前 DDR4 仍是主流，随着支持 DDR5 的 Intel EMR 和 AMD Bergamo 处理器平台渗透率的增加，预计 DDR5 在服务器上的渗透率也将逐渐提高，同时考虑到各大厂商在逐步削减 DDR4 产能转向 DDR5，DDR5 逐步替代 DDR4 也将是市场趋向。根据 TrendForce 在 2024 年 7 月的预测

数据，2024 年 DDR5 在服务器 DRAM 位出货量中的占比将达到 40%，2025 年将增加到 60-65%，DDR5 对 DDR4 的替代正处于加速阶段。而美光科技在 2023 年第三季度业绩会上说明，其将在 2024 年第一季度实现 DDR5 出货量高于 DDR4，领先行业平均进度，因此预测美光 DRAM 中的产品结构也已经从 DDR4 向以 DDR5 为主转变。因此，可以判断即使 2025 年国内厂商大幅扩产，也主要将加剧 DDR4 的价格竞争，而对 DDR5 的价格与出货量影响较小，DDR5 将对未来 DRAM 整体价位起到更强的支撑作用。综合 DDR5 在服务器市场的渗透率与美光产品结构而言，受惠于 DDR5 对 DDR4 的替代，美光未来 DRAM 收入受到来自国内厂商的冲击较小。

图 9:DDR4 依然是当前 DRAM 市场主流



数据来源: kbvresearch, 中信建投证券

图 10: 从 DDR1 到 DDR5 的演变: 电压更低, 速率翻倍, 容量翻倍

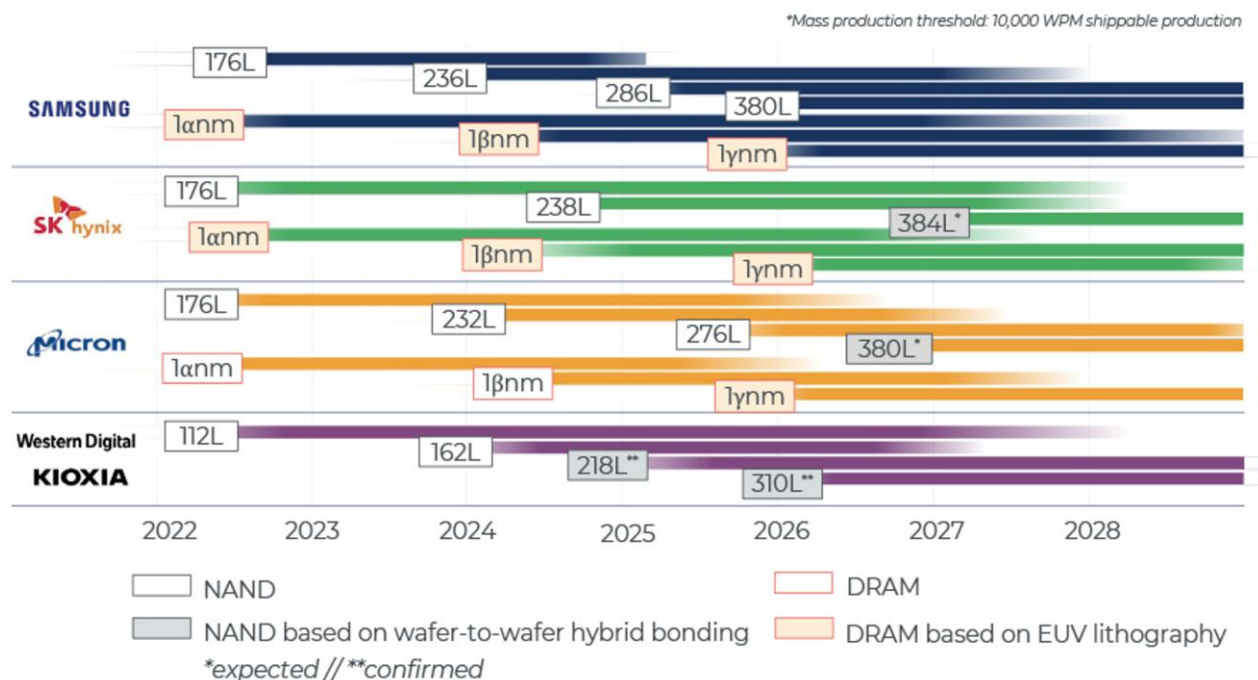
Gens	Release Year	Voltage	Data Rate	Density	Number of Banks	Prefetch Architecture	Supported Burst Length
DDR1	2000	2.5/2.6 V	200 MT/s ~ 400 MT/s	64 Mb ~ 1 Gb	4	2n-prefetch	2, 4, 8
DDR2	2003	1.8 V	400 MT/s ~ 800 MT/s	256 Mb ~ 4 Gb	4, 8	4n-prefetch	4, 8
DDR3	2007	1.5/1.35 V	800 MT/s ~ 2133 MT/s	512 Mb ~ 8 Gb	8	8n-prefetch	BC4, BL8
DDR4	2014	1.2 V	1600 MT/s ~ 3200 MT/s	2 Gb ~ 16 Gb	8, 16	8n-prefetch + bank grouping	BC4, BL8
DDR5	2020	1.1 V	3200 MT/s ~ 6400 MT/s	8 Gb ~ 32 Gb	8, 16, 32	16n-prefetch + bank grouping	BC8, BL16, BL32

数据来源: 启威测实验室, 中信建投证券

平面 DRAM 升级主要依赖于制程进步，美光在 1- α 和 1- β 制程领先。三大厂商制程都向着 10nm 演进，但随着晶体管尺寸的减小，电子泄漏与其他量子效应问题将对存储性能和稳定性产生严重影响，因此都暂未突破 10nm 以下。2021 年三家厂商都开始生产基于 1- α 制程节点的产品，但三星与海力士已经开始引入 EUV 光刻机技术；2023-2024 年，三家厂商开始量产 1- β 制程节点的 DRAM 产品；目前三家厂商都在向 1- γ 制程节点演进，美光科技率先公布 1- γ 制程节点相关计划，其已经在广岛试产采用 EUV 技术的 1- γ 制程节点 DRAM 产品，并

计划于 2025 年大规模量产，而三星与海力士暂未公开宣布具体时间。

图 11:头部存储厂商 DRAM 制程演进图



数据来源: YOLE, 中信建投证券

除了通用 DDR 之外, DRAM 还包括 LPDDR、GDDR 及新型存储 HBM 等产品。LPDDR 相比于同代 DDR 产品功耗更低、体积更小, 广泛应用于移动式电子产品, 目前主流配置为 LPDDR4, 三星与海力士已经分别开发出 LPDDR5X 和 LPDDR5T 两款最新一代 LPDDR 产品。GDDR 为专门适配需要极高吞吐量的应用程序设计的高性能存储器, 主要应用在显卡领域。

图 12:LPDDR2/3 和 LPDDR4 原理比较

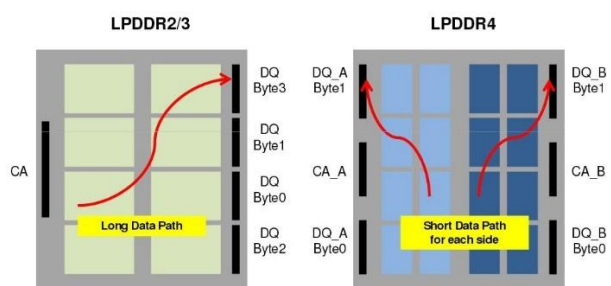


图 13:GDDR 制程演进比较

	GDDR7	GDDR6X (Non-JEDEC)	GDDR6
B/W Per Pin	32 Gbps (Gen 1) 48 Gbps (Spec Max)	24 Gbps (Shipping)	24 Gbps (Sampling)
Chip Density	2 GB (16 Gb)	2 GB (16 Gb)	2 GB (16 Gb)
Total B/W (256-bit bus)	1024 GB/sec	768 GB/sec	768 GB/sec
DRAM Voltage	1.2 V	1.35 V	1.35 V
Data Rate	QDR	QDR	QDR
Signaling	PAM-3	PAM-4	NRZ (Binary)
Maximum Density	64 Gb	32 Gb	32 Gb
Packaging	266 FBGA	180 FBGA	180 FBGA

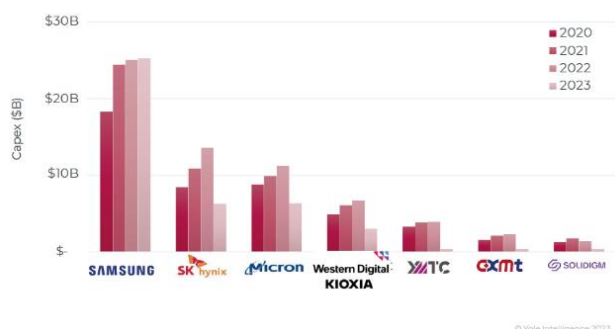
数据来源: 《Lpddr Ram Types Understanding》, 中信建投证券

数据来源: AnandTech, 中信建投证券

2022 年底以来多家厂商统一减产, DRAM 价格迎来修复。2021 年四季度开始 DRAM 平均价格连续 8 季下跌, 2022 年底多家厂商都做出减产决策, 希望通过减少供给进而恢复供需平衡。2023 年存储厂商资本开支都有所削减, 美光资本开支从 2022 年的 121 亿美元削减至 77 亿美元, 降幅达到 36%, 海力士资本开支削减从 2022 年的 19.7 万亿韩元至 8.8 万亿韩元, 同比下降 55%。三星和海力士在 2023 年第三季度的产量相较于第一季度分别下降了 15% 和 12%, 主动减产效应明显。2023 年下半年开始, DRAM 价格逐渐修复, 2024 年第二季度改善明显, 三星、美光和海力士的 DRAM 收入都明显回升。

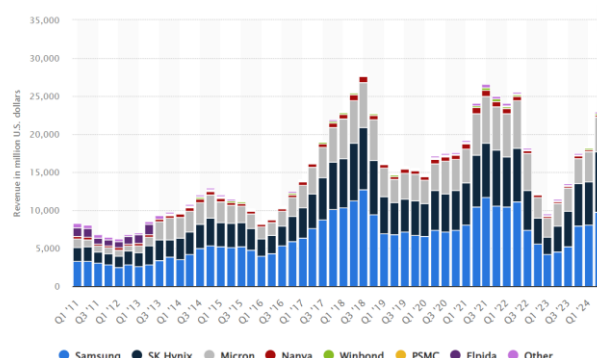
请务必阅读正文之后的免责条款和声明。

图 14:头部存储厂商 CAPEX 情况



数据来源: YOLE, 中信建投证券

图 15:头部厂商近年来 DRAM 收入变化情况



数据来源: statista, 中信建投证券

（二）NAND：堆叠层数不断提升，行业格局更为分散

根据存储单元，NAND Flash可分为 QLC、TLC、MLC 和 SLC。根据存储密度、耐久性与成本效益等技术类型，NAND Flash 可以分为 SLC（单层存储单元）、MLC（双层存储单元）、TLC（三层存储单元）、QLC（四层存储单元），依次对应每个存储单元存储的数据从 1 位到 4 位。从 SLC 到 QLC，存储密度逐步提升，单位比特成本降低，但性能、功耗、可靠性与寿命也会下降。目前市场主流产品为 TLC 和 QLC。

3D NAND 架构的突破显著增加了储存容量。3D NAND 自 2015 年后成为主流，该架构突破了 2D NAND 在容量上所受到的限制，其将多层记忆存储格垂直堆叠、通过微小的垂直通道实现层间互联，能在不破坏数据完整性的同时，建立更大的储存容量，降低单位位元成本，能实现更高密度、更低功耗与更快的读写速度。3D NAND 以 3D 堆叠层数作为主要发展方向。多家厂商已开发出五种主流 3D NAND 架构。2013 年，三星率先推出了 V-NAND 架构，但其单层蚀刻技术最多至 128 层，而 128 层之后的 COP 结构有部分外围设备仍位于单元外部，则需要减少单元阵列与外围区域以减小尺寸，存在一定局限；美光/英特尔采用在芯片的外围逻辑上构建其 3D NAND 阵列的方法实现 3D 架构，称为 CuA，现已基于此推出 232 层 NAND 产品，可实现每平方毫米最高的 TLC 密度(14.6 Gb/mm²)；铠侠和西部数据联合开发 BiCS 架构，其存储单元不是一层一层地堆叠，而是先堆叠板状电极，然后在之间开孔连接电极，可有效降低制造成本；海力士在 3D 的基础上，推出 4D PUC，提升了 CTF NAND 阵列下的外围电路，从而在芯片上节省空间提高密度并降低成本，但缺点是堆叠技术可能在未来达到极限，海力士计划以多站点电池来克服这一问题；长江存储采用 Xtacking 技术，把存储阵列和外围电路分开在两个独立晶圆上加工，然后通过铜混合键合技术堆叠并连接，但成本依旧较高。

图 16:NAND 从 2D 到 3D 结构演进

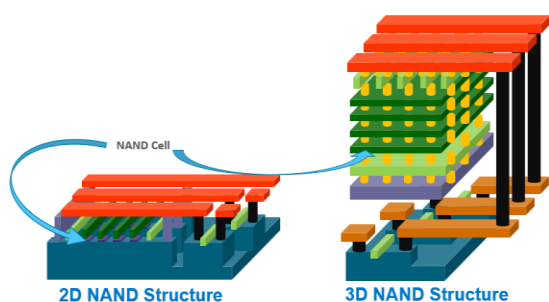
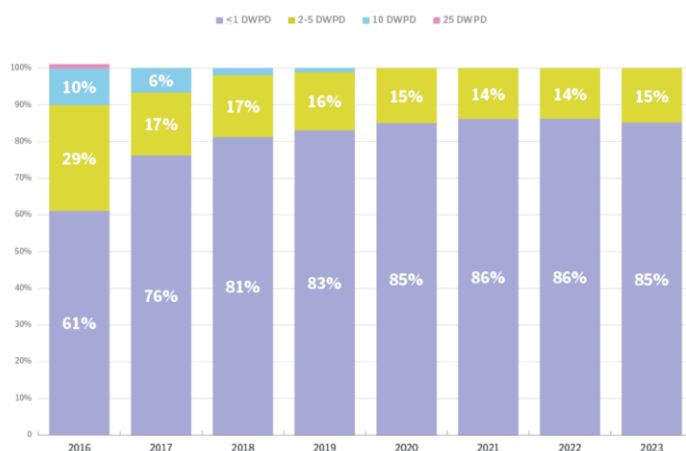


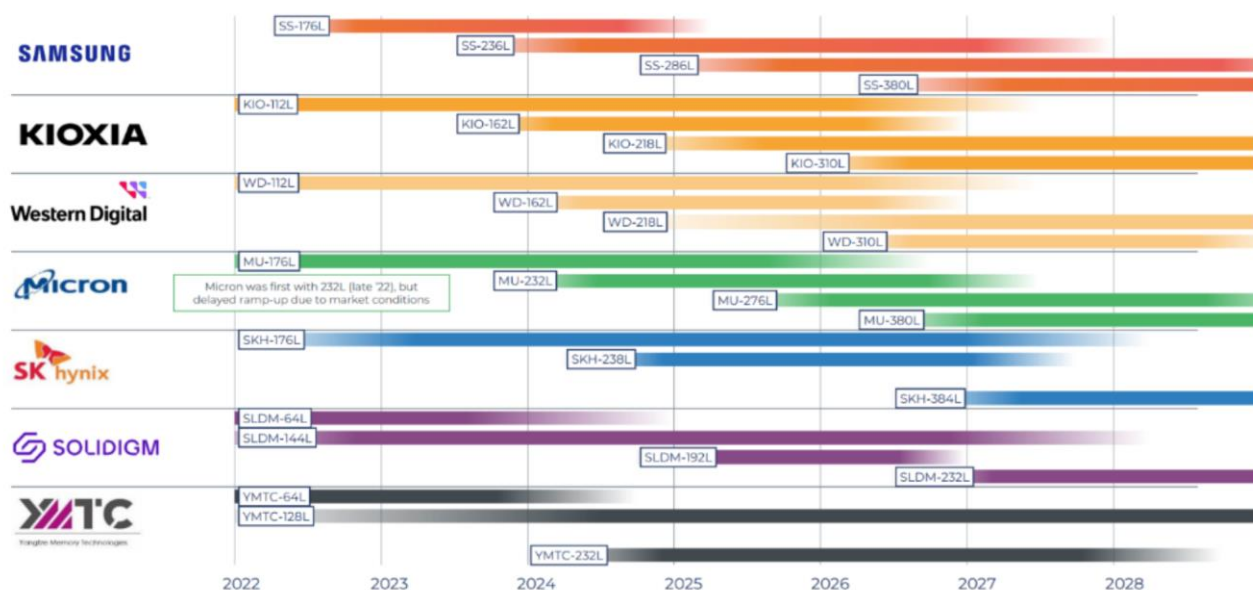
图 17: 2016-2023 年全球 SSD 出货量结构



数据来源: Simms, 中信建投证券

数据来源: techtarget, 中信建投证券

图 18:各原厂 NAND 制程演进图

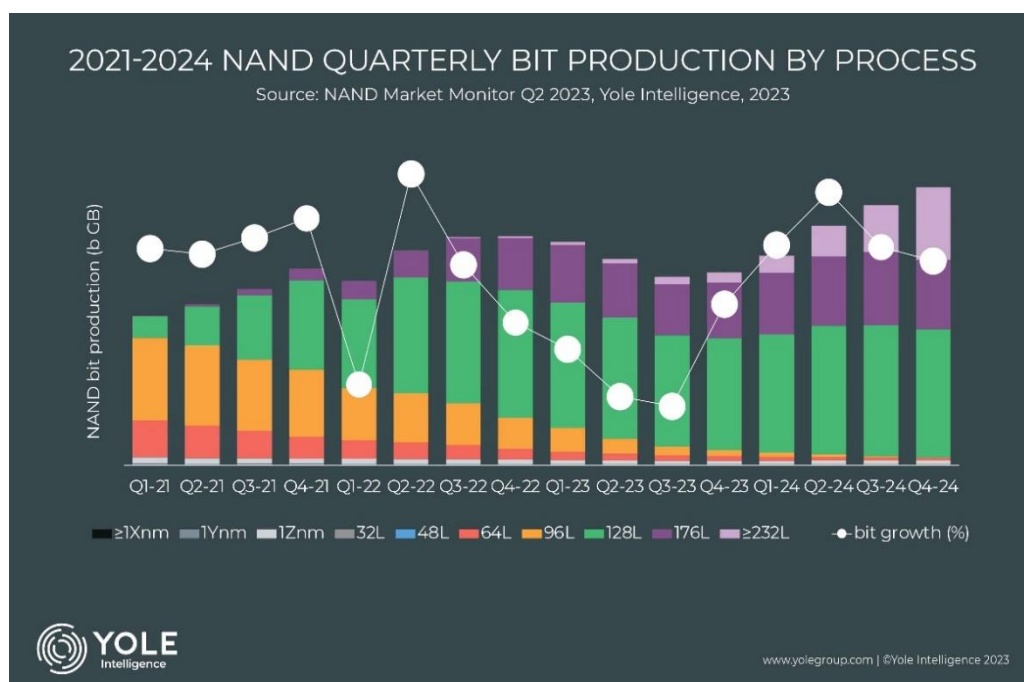


数据来源: YOLE, 中信建投证券

美光科技是首个推出 232 层 TLC NAND 芯片的厂商。3D NAND 闪存中垂直堆叠的存储单元层数的增加能够在相同的晶圆面积上容纳更多的存储单元，从而显著提高存储密度，提高读写速度和数据传输速率，因此层数增加是 3D NAND 的主要发展方向。根据 Yole 预测，232 层以上的 NAND 在 2024 年开始快速增长，并成为 NAND 市场的重要增量。美光科技最先推出突破 200 层的 232 层 NAND 产品，是市场上层数最多的 3D NAND 产品之一，与 176 层 3D NAND 相比读取带宽提升了 75% 以上，具备高密度、高性能和低功耗的特点，适用于高性能计算、数据中心和移动设备等多种应用场景。今年 4 月，美光宣布 232 层 QLC NAND 已经量产，并在部分 Crucial 英睿达固态硬盘（SSD）中出货。2022 年，美光就混合键合 IP 与 Xperi-Adeia 签订协议，可能用于突破 300 层以上 3D NAND。

请务必阅读正文之后的免责条款和声明。

图 19:2021-2024 年 NAND 产量（按堆叠层数划分）

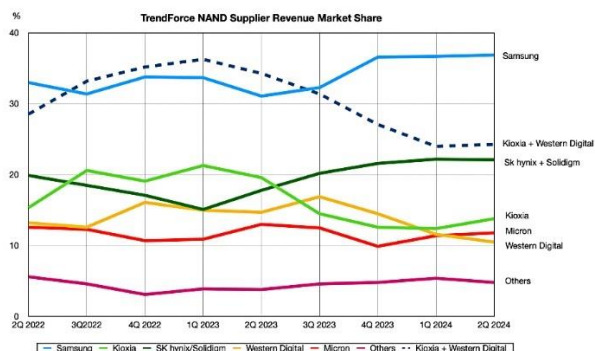


数据来源: YOLE, 中信建投证券

NAND 市场竞争更为充分, 格局更为分散。尽管都属于存储行业, NAND 在市场结构和技术门槛上与 DRAM 存在显著差异。DRAM 生产高度依赖于先进的制程技术尤其是 EUV 技术, 设备和工艺的复杂度更高, 同时其高生产成本仅有少数公司能够负担, 因此表现出三寡头的高集中度市场竞争格局。而 NAND 产品相对而言, 对技术难度生产投入的要求更低, 因此市场竞争更为充分、产品更为标准化、市场竞争格局也更为分散。2024 年以来, 零售市场需求下滑明显, PC 及智能手机领域市场需求未达预期, 同时西部数据面临企业级 SSD 验证时间较长的困境, 因此出现显著业绩下滑; 而美光 NAND 业绩稳中略进, 成功超越西部数据, 市占率升至第四。

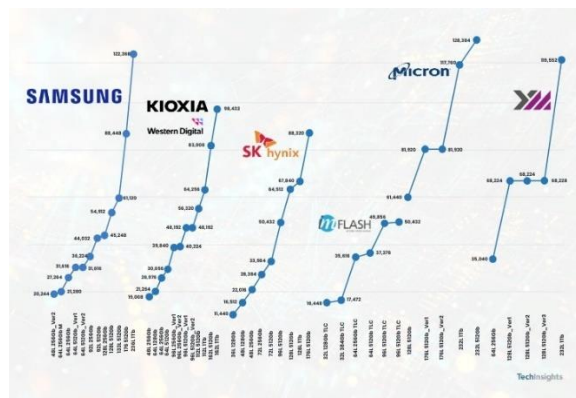
NAND 库存水位上升, 供应商调整产能。2024 年二季度以来手机终端需求走弱, 仅北美服务器 QLC SSD 需求走强但承载产能有限, 随着原厂新制程产能释放, 库存水平上升, NAND ASP 正在见顶。三星继 2024 年上半年 NAND 减产扩大 40%-50 后, 可能在 2024 年四季度对 NAND 继续减产; 海力士最新减产措施持续至 2024 年上半年; 铠侠也可能在 2024 年四季度减产 NAND。

图 20:24Q2 美光 NAND 市占率升至第四



数据来源: trendforce, 中信建投证券

图 21: NAND 容量在迅速增长



数据来源: Techinsights, 中信建投证券 (纵轴为最大 GB/晶圆)

(三) HBM: AI 时代大有可为, 供给受限于 TSV 产能释放情况

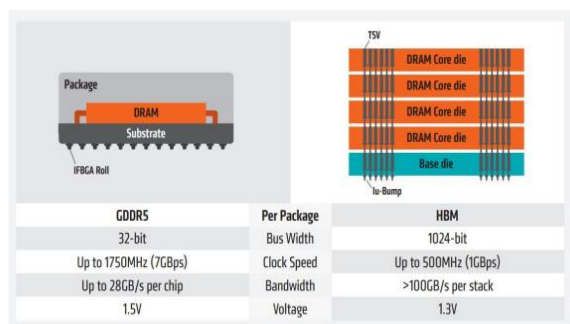
AI时代, 传统存储性能制约高性能计算。目前绝大部分计算系统都采用冯·诺依曼计算机体系, 而该体系下存储与计算单元分离, 需要通过总线传输数据, 因此存储的带宽将限制算力的利用效率。当下大模型训练需要处理大量并行数据, 需要算力与带宽的同步提升, 而现在算力已经提高三倍至 66.9tflops, 但存储带宽慢或容量不足限制了每秒可访问的数据 (即“内存墙”问题), 因此存储带宽限制了 AI 芯片性能。目前解决这一问题的思路是采用存算一体的新架构, 具体包括近存计算 (PNM)、存内处理 (PIM)、存内计算 (CIM) 三种。

HBM 的高带宽特性满足 AI 芯片需求。HBM 即高带宽内存, 是一种基于 3D 堆栈工艺的高性能 DRAM, 是近存算一体的一种, 可与高性能图形处理器、网络交换及转发设备、高性能数据中心的 AI 特殊应用集成电路结合使用, 在 CPU 中用作包内高速缓存, 在 CPU 和 FPGA 中用作包内 RAM。HBM 由多颗 DRAM 3D 堆叠而成并由 TSV 技术实现信号互连与分配, 而每颗 DRAM 包含多个通道, 每个通道包含多个 I/O 口, 最终高 I/O 数量显著提升了位宽, 现 HBM3 的总位宽高达 1024bit, 带宽 819GB/s, 远远超过 DDR5 的 51.2GB/s 和 GDDR6 的 64GB/s。

HBM 相比 GDDR 有更高的带宽、更优的能效比与更高的存储密度。HBM 与 GDDR 对应着两种提高总带宽的思路。总带宽=I/O 数据速率*位宽/8, 位宽=I/O 数*单 I/O 位宽, 因此总带宽的提升有两种重要思路, 一是提高 I/O 数, 即 HBM 的思路; 二是提高 I/O 数据速率, 即 GDDR 的思路。HBM 利用 TSV 技术互连多颗 DRAM, 而更高的 DRAM 数量提升 I/O 数, 而不改变单 I/O 数据速率, 使总带宽提升明显; 而 GDDR 大幅提升单 I/O 数据速率, 最高可达 16Gb/s, 远超 HBM3 的 6.4 Gb/s, 但由 PCB 与计算核心互连的架构最终限制其总位宽在 32bit, 导致最终总带宽提升幅度受限, 最高 28GB/s。而 HBM 采用了 TSV、微凸块等技术, DRAM 裸片与计算核心信号传输路径较短、I/O 电压更低, 使其能效更高。HBM 的 3D 堆叠结构能够节省 94% 的面积 (以 1GB HBM1 为例), 因此有效节省了面积空间。

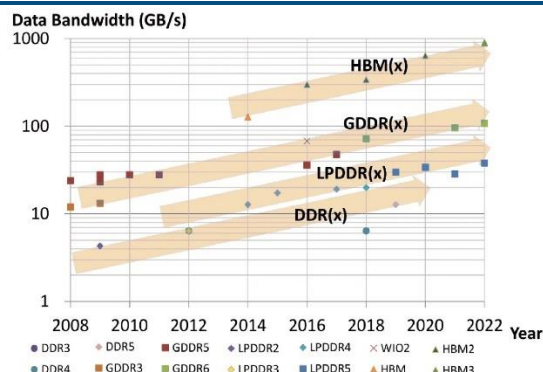
目前只有海力士、美光和三星有能力生产兼容高性能 AI 计算系统的 HBM 芯片。HBM 制备中, TSV 硅通孔制造依托于晶圆制造工艺, 包括深孔刻蚀、气相沉积、铜填充、CMP、晶圆减薄等; 同时 HBM 制备还需要复杂的堆叠键合技术, 技术难度、投入成本与供应链稳定形成了 HBM 的入行门槛。

图 22:HBM 与 GDDR 原理比较



数据来源: techbang, 中信建投证券

图 23:HBM 性能相比 GDDR、LPDDR 等有明显提升

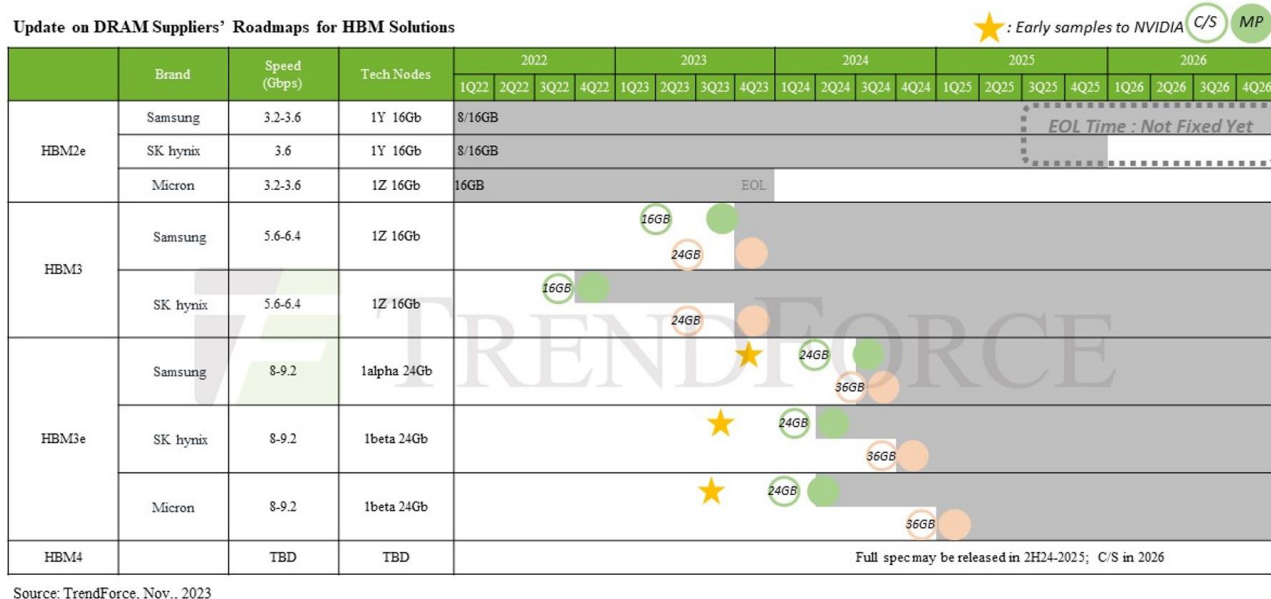


数据来源: springer, 中信建投证券

HBM 经历从 HBM1-HBM3E 的演变迭代, 主要从提升容量与带宽的方向升级。单颗 HBM 容量取决于 DRAM 堆叠层数与每片 DRAM 容量, 而带宽的升级主要通过提升 I/O 数据传输速度与 I/O 总线位宽。第一代 HBM 由 AMD 与海力士合作研发, 采用 4 层芯片堆叠, 可提供 128GB/s 带宽和 4GB 单层容量, 但成本较高。HBM2 则由英伟达、三星和台积电合作研发, 于 2018 年上市, 由 8 层 DRAM 芯片堆叠而成, 传输速度与容量大幅提升至接口速度 2.4Gbps, 单层容量 16GB, 总容量 256Gb。HBM2E 是 HBM2 的扩展版, 增加了每堆叠层的容量和带宽, 接口速度达到 3.6Gbps, 峰值带宽为 410GB/s。HBM3 则于 2022 年正式上市, 堆叠层数和管理通道数增加, 由 12 层 DRAM 芯片组成, 单层容量为 32Gb, 总容量为 640Gb, 接口速度为 6.4Gbps, 峰值带宽翻倍为 819GB/s。而 2023 最新发布的 HBM3E, 作为 HBM3 的增强版, 将接口速度提高到 8Gbps, 峰值带宽达到 1TB/s。

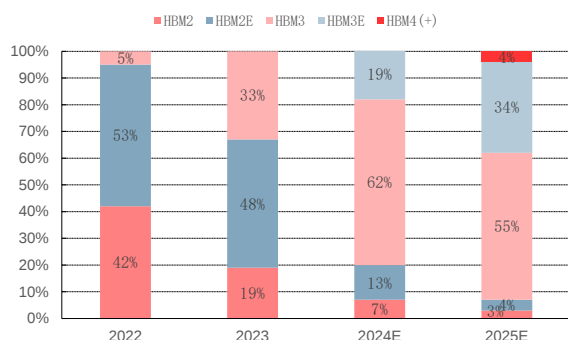
HBM3 与 HBM3E 将替代 HBM2E 成为 HBM 市场的主流。2022-2023 年全球 HBM 市场主要以 HBM2E 为主, 而 HBM3 与 HBM3E 契合了 AI 服务器需求, 2024 年来持续增长, 随着其发布与持续放量, 2024 年及之后市场需求将向 HBM3 以及更高版本转移, HBM3 与 HBM3E 将成为市场主流 HBM 产品, 市占率有望由 2023 年的 33%急速提升至 2024 年的 80%以上。截至目前, 三大厂商的 HBM3E 8Hi 都已通过英伟达检验, 美光与海力士已经收获英伟达订单并实现量产, 同时三大厂商也已送样最新一代产品 HBM3E 12Hi。

图 24:HBM 主要厂商技术路线演进图



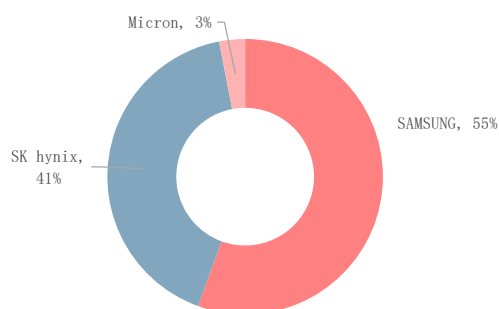
数据来源: trendforce, 中信建投证券

图 25: HBM 各产品产能分布预测



数据来源: YOLE, 中信建投证券

图 26:HBM 市场呈现三足鼎立格局



数据来源: YOLE, 中信建投证券

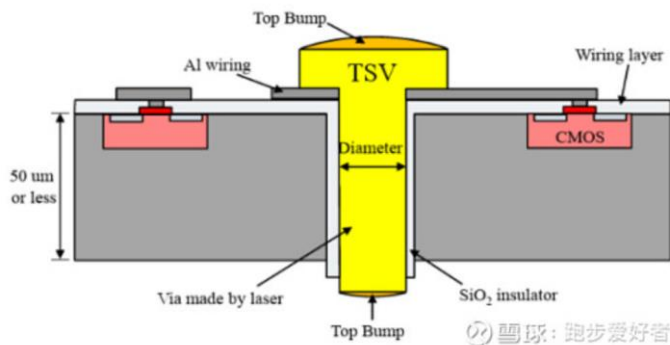
HBM 供给受限 TSV 产能。在 HBM 制备中,HBM 的多层 DRAM 颗粒需要堆叠在一颗逻辑晶圆上,DRAM 颗粒之间、DRAM 和逻辑晶圆间均通过硅通孔 (TSV) 和凸点垂直互连,因此 HBM 的产能与 TSV 的产能直接相关而 TSV 工艺的复杂性和技术要求极高,包括深孔刻蚀、气相沉积、铜填充、CMP、晶圆减薄等。目前 HBM 的市场需求激增,据 TrendForce 集邦咨询预估,2024 年底 HBM 产值占 DRAM 的比重将扩大至 20.1%,而整体 DRAM 产业规划生产 HBM TSV 的产能约为 250K/m, 占总 DRAM 产能 (约 1,800K/m) 约 14%。因此,即使 TSV 产能显著增长,也依旧难以满足 HBM 的需求,HBM 的供给能力在很大程度上受限 TSV 的产能。

HBM 工艺流程中,堆叠键合与封装是重要工艺。HBM 工艺流程主要分为五步:一是制造 DRAM 晶圆、逻辑晶圆与 TSV;二是制造凸点,将硅通孔后的 DRAM 晶圆和逻辑晶圆倒装,然后进行减薄,在晶圆背面形成凸点;三则是堆叠键合,这一步是存在差异化的环节,三星/美光使用较为传统的 TC-NCF 工艺,海力士采用独创的 MR-MUF 工艺能减少芯片键合产生的应力且散热效率更高。未来在 HBM4 当中,可能不再使用凸点,而是使用铜与铜互连。四则是将模塑后的 KGSD 晶圆切割成颗粒。五则是将 HBM 与 SoC (GPU 等) 封装在

请务必阅读正文之后的免责条款和声明。

一起，这一步的主流封装技术为台积电的 2.5D 封装技术（CoWoS）。

图 27:TSV 原理示意图



数据来源：雪球，中信建投证券

AI 芯片强劲需求下，三大厂商都在积极推进 HBM 扩产。海力士预计 2024 年 HBM 产能同比翻倍增长，从 2023 年每月 4.5 万片晶圆提升至 2024 年底每月 12-12.5 万片晶圆，同时 2024 年保持 10 万亿韩元的新增资本开支以用于改善封装制程，包括 MR-MUF 键合工艺和 TSV 技术进而支持 HBM 产能增长。海力士还计划与铠侠在日本共同生产 HBM。三星预计 2024 年 HBM 芯片产量将比 2023 年增长 2.5 倍，2025 年将再次翻倍，截至 2024 年底预计每月生产 13 万片晶圆，三星现已收购韩国天安厂区内部分建筑及设备，用于 HBM 生产。美光在 2023 年 11 月 6 日在台中开设新工厂，致力于大规模生产 HBM3E 以及其他产品，并计划资本开支约 75-80 亿美元，主要用于 HBM 生产。根据 Trendforce 预测，美光在 2024 年底，有望实现每月 2 万片晶圆的产能。

AI 催化 HBM 产量与市场规模爆发。相较于传统服务器，AI 服务器中的 DRAM 需求量提升了 8 倍，同时随着 GPU 的迭代升级，单个 GPU 使用的 HBM 数量也会增加。根据 Yole 预测，2022 年至 2025 年，HBM 出货量将以 90% 的复合增速快速增长，其中 2024 年存在显著跃迁，同比增长约 150%，至 2025 年，HBM 出货量将达到 17 亿 GB；营收方面，也存在类似趋势，HBM 整体市场规模将以 94% 的复合增速快速增长，至 2025 年达到 1990 万美元，其中 2024 年同比增长超过 150%，达到 1410 万美元。

表 1:HBM/TSV 产能预测

	三星	海力士	美光
HBM TSv capacityin End of 2023	45K/m	45K/m	3K/m
HBM TSv Capacityin End of 2024	130k/m	120-125K/m	20K/m

资料来源：trendforce，中信建投证券

图 28:HBM 出货量&市场规模统计及预测



数据来源: YOLE, 中信建投证券

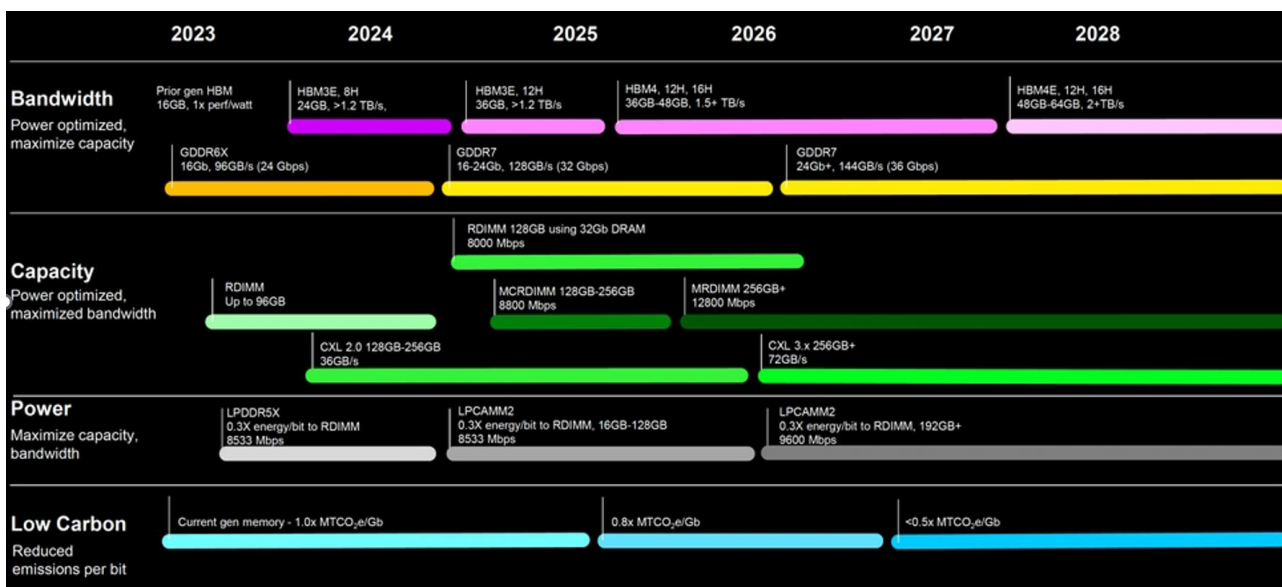
（四）竞争格局：三星 VS 海力士 VS 美光

美光直接开发 HBM3E, 弯道超车先于三星获得英伟达订单验证。不同于海力士与三星一开始就是 HBM 的先导者, 一开始美光选择的技术路径为 HMC 而非 HBM, 但由于成本高昂与技术缺陷等问题, 市场最终选择了 HBM。美光首次量产的 HBM 产品为 HBM2 与 HBM2E, 但由于失去先发优势, 整体节奏落后三星与海力士, 根据 Yole 数据, 美光 2023 年市场份额仅为 3%。为了追赶海力士和三星, 美光 2023 年美光直接跳过 HBM3, 与台积电合作研发 HBM3E, 2024 年一季度, 其率先宣布 HBM3E 8Hi 量产并供货英伟达 H200, 与海力士节奏同步, 而三星电子则是在 2024 年 9 月获得英伟达验证并正式出货。至此, 美光弯道超车三星, 追平 HBM 进度, 并获得重要大客户英伟达的验证与支持。根据美光业绩会报道, 其 2024 年 HBM 订单已全部订满。

美光在功耗上存在性能优势。美光 HBM3E 相较于竞争对手而言能实现功耗降低约 30%。在制程节点方面, 美光与海力士 HBM3E 都采用 1-β 制程节点制造, 领先于三星的 1-α 制程节点。但在键合技术方面, 美光 and 三星都采用的是较为传统的 TC-NCF (热压非导电膜) 技术, 主要是通过芯片间涂抹薄膜状物质来堆叠芯片, 并通过加热和加压使这种物质融化, 从而将芯片粘在一起, 该过程需要高温高压环境将凸点推入非导电薄膜, 在单个 DRAM 高度减少的环境下更易导致芯片翘曲, 因此技术要求与限制相对海力士采用的 MR-MUF 技术将更加严格。

美光规划于 2026 年量产 HBM4。美光对 2024-2028 年的 HBM 发展节奏做出了系统性规划, 公司预计于 2025 年实现 HBM3E 12Hi 量产, 容量达到 36GB, 带宽达到 1.2TB/s; 同时于 2025 年末-2027 年量产 HBM4 12Hi 与 HBM4 16 Hi, 容量最高可达到 48GB, 带宽达到 1.5 TB/s; 2028 年及其之后公司将迭代至 HBM4E, 开始量产 12Hi 与 16Hi, 最高容量达到 64GB, 带宽超过 2TB/s。

图 29:美光 HBM 及 GDDR 路线演进图

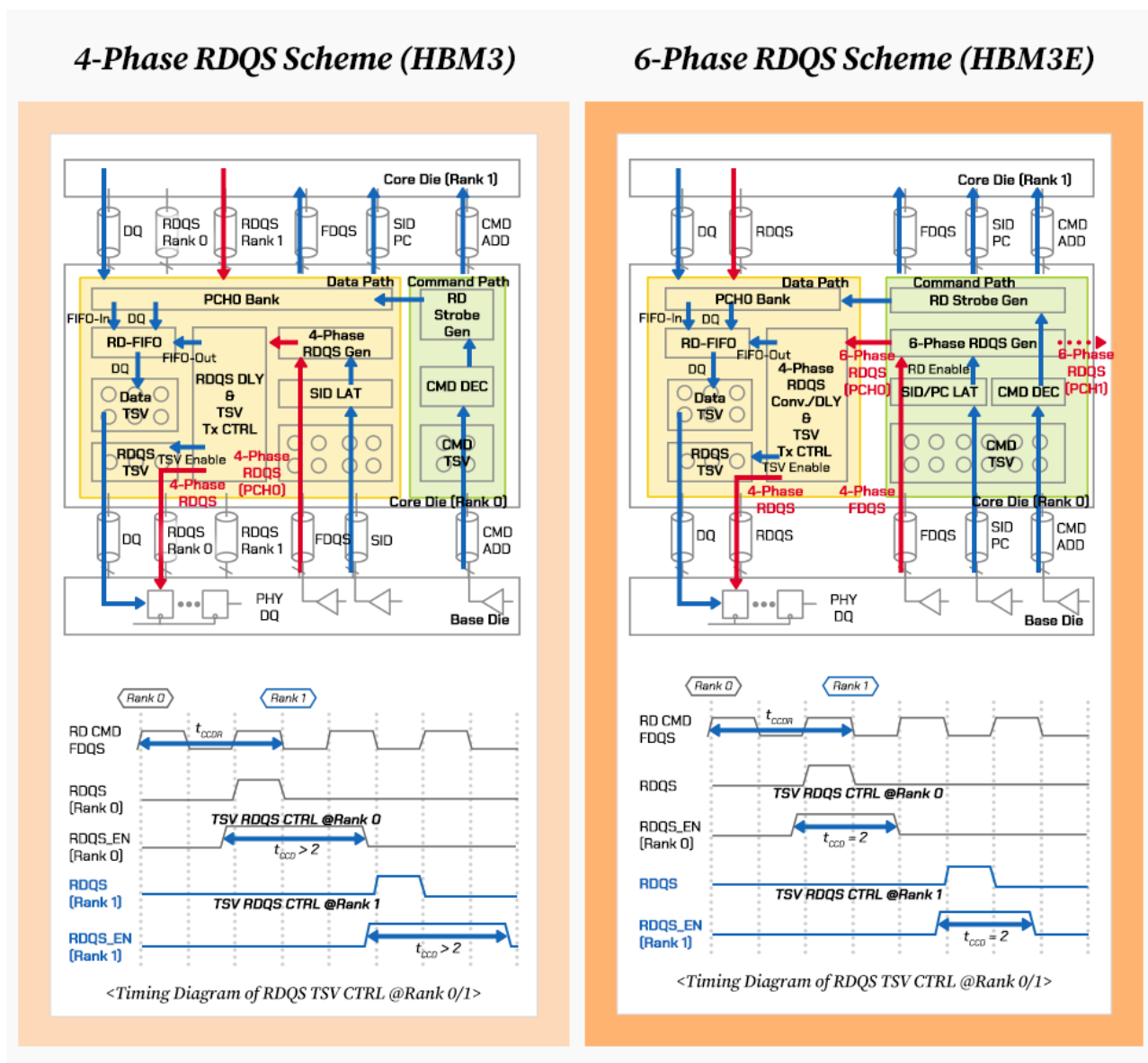


数据来源：美光，中信建投证券

海力士先发优势明显，是 HBM 市场的领导者。2023 年海力士市场份额占比第一，超过一半，达到 55%。海力士是 HBM 领域的先行者也是最先开始研发 HBM 的存储企业，技术持续领先，目前其制程节点为 1-β，领先三星；同时 HBM 生产中的键合技术为其独有的 MR-MUF（批量回流模制底部填充）技术，即将多个芯片放置在下层基板上，通过回流焊一次性粘合，然后用模塑料填充芯片之间或芯片与基板之间间隙，该方法不依靠高温和高压，减少了芯片键合产生的应力，可有效解决芯片翘曲从而提升良率，据海力士披露数据，其 HBM3E 良品率已达 80%，领先竞争对手；同时使用的环氧树脂模塑料的导热系数是非导电薄膜的 2~4 倍，散热效率更高。客户方面，海力士是英伟达 AI 芯片 HBM 的主要供应商，市场份额最高。产品方面，海力士率先量产 HBM3 12Hi，单个 DRAM 芯片制造得比以前薄 40%，因此 12 颗 3GB DRAM 芯片在厚度与现有 8 层产品一致的情况下，容量增加了 50%，运行速度为现有内存的最高速度 9.6Gbps，实现每秒可读取 35 次 700 亿个整体参数。

海力士计划 2025 年实现 HBM4 量产。由于海力士在晶圆代工厂上的缺乏，公司与台积电签署合作备忘录，合作开发 HBM4。海力士预计 2025 年提供 HBM4 样品并于当年实现 HBM4 12 Hi 的量产，在 2026 年实现 HBM4 12 Hi 的量产。在 HBM4 上，海力士预计仍将采用 Advanced MR-MUF 工艺，还同时对 Fan-out RDL（扇出型重新分配层）及混合键合（Hybrid bonding）等下一代先进封装技术有所开发。同时针对 HBM4，海力士考虑从而将存储芯片和逻辑半导体集成在同一芯片上。

图 30:海力士 6-phase 和 4-phase RDQS (HBM3) 方案比较

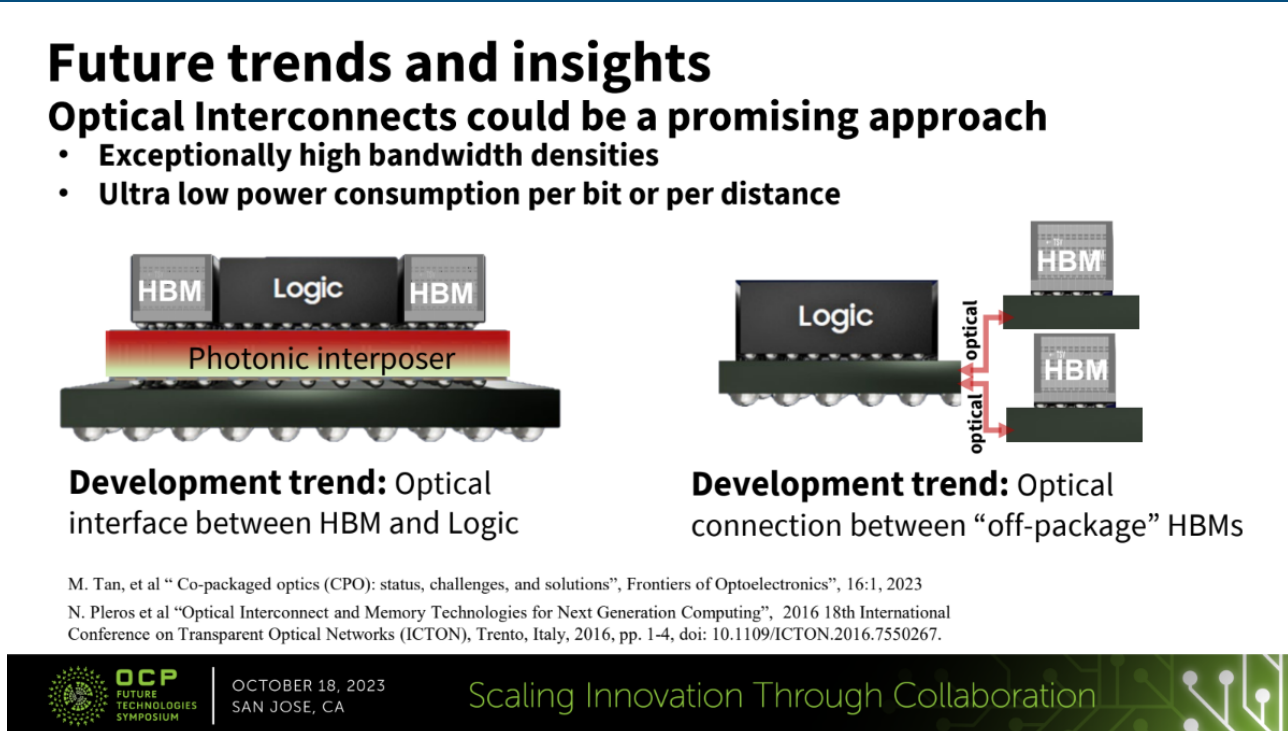


数据来源: SK 海力士, 中信建投证券

三星短期有所落后,但中期“一站式策略”存在竞争优势。技术方面,三星采用 1- α 制程节点构建 HBM3E,而美光与海力士已经采用 1- β 技术,三家厂商都计划在 HBM4 及 HBM4E 中采用 1- γ 制程节点,可能对三星而言是更大的挑战。在键合技术方面,三星目前采用 TC-NCF 键合技术,良率相比于海力士 MR-MUF 技术落后较多,但三星计划在 HBM4 键合技术上采取双轨策略,同时开发混合键合技术和传统的 TC-NCF 技术,再考虑到 HBM 高毛利的特点,三星或许可凭借技术转变与规模优势消除潜在的良率的负面影响。三星作为存储厂商和晶圆代工厂,能提供集存储、AI 芯片设计、晶圆代工和封装为一体的一站式服务,相比起海力士与台积电合作可能出现的产能错配问题,三星部门间协同效率可能高于海力士与台积电合作可能大大缩短从芯片研发到制造到封装到交付的周期。同时,在台积电产能不足的情况下,三星的先进封装技术可能成为台积电 CoWoS 的有效替代,有利于其争夺市场。

三星对光互连技术与应用提出了两种构想。三星认为光互联技术将对未来存储带来巨大提升。光互连是指对单个光子（光的粒子/波）信息进行编码的技术，其能大幅降低功耗，带宽密度大幅提升，使处理延迟达到飞秒级。同时光学 I/O 也有两种方式：一是在 HBM 和逻辑电路之间使用光接口，以提高数据传输效率，具体是将光子介质层安置在基础封装层和顶层(CPU 和 HBM)之间，使其通过光子介质层进行通信，但这个方案的成本较高。二是在逻辑模块和存储模块之间使用光连接，而不是传统的电气连接，进而减少信号传输的延迟和功耗。具体而言是将 HBM 模组从芯片本体分离出来，并通过光学技术与逻辑处理单元相连，进而降低 HBM 和逻辑单元的芯片制造与封装成本，且能避免复杂的数字-光学信号转换。

图 31:三星对于存储技术未来趋势和展望



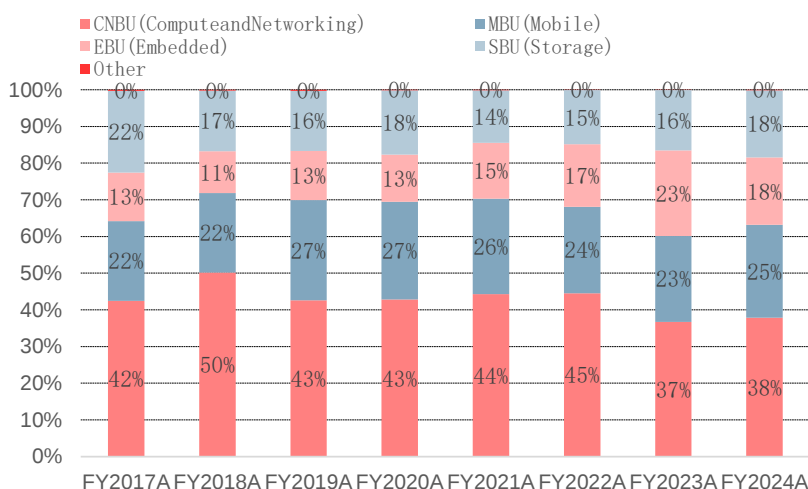
数据来源：OCP，三星，中信建投证券

三家供应商的 HBM 产品竞争与互补。三星的 HBM 产品在性能和速度上表现优异，特别是数据传输速度上具有很大优势，能支持多个行业的 AI 和高性能计算应用。海力士 HBM 产品的功耗管理和散热设计具有技术领先性，特别适合用于高效能计算的应用场景，与 AI 芯片霸主英伟达有更广更久的合作。美光 HBM 产品的耐久性和广泛兼容性表现突出，对 HBM 在 AI、HPC 和图形处理等领域的应用有独到见解。

三、AI 引领下游需求增长，消费电子及其他端侧产品需求有待接棒

根据需求，美光收入主要分为 CNBU（计算和网络）、MBU（移动设备相关业务）、SBU（存储产品）、EBU（嵌入式系统业务）四个板块。其中计算与网络是公司的需求的大头，占比 40%左右，近两年占比略有下降；移动设备相关产品是传统需求，也是公司第二大收入来源板块，占比近 25%；嵌入式相关业务近年占比有所上升，由 2021 年前近 13%的平均收入占比，到 2022-2024 财年平均收入占比 19%；存储产品占比平均在 17%。

图 32: CNBU 和 MBU 部门贡献美光主要收入



数据来源: Wind, 中信建投证券

（一）AI 算力催化 HBM 需求

AI 服务器需求快速增长，在服务器出货量中的渗透率不断提高。根据 Yole 和 FMS 统计及预测，AI 服务器的需求从 2022 年开始显著增长，在 2025 年之后预计增长将加速，2025 年 AI 服务器领域的出货量预计同比增长将达到 28%，预计在 2022-2029 年将实现 43% 的复合增速，远高于传统服务器增速，AI 服务器在全球服务器出货量的占比不断提升。英伟达仍然是主流 AI 服务器芯片的主要供应商，占比近 50%。但相对 2023 年及之前英伟达一家独大的场面而言，2024 年 AMD 和其他厂商的市场份额也有所扩大。英伟达 2024 年的出货量集中于 H 系列，而 B 系列预计将在 2025 年推出以满足大型数据中心的需求。

HBM 是 AI 时代下的核心存力。HBM 采用了近存算一体的架构，通过中介层连接信号处理器芯片，节省了数据传输的时间与耗能，解决了传统内存存在的“内存墙”问题，同时堆栈技术的应用也减少了其对空间的占用。随着 AI 时代的到来，AI 服务器渗透率显著提升，处理大量并行数据的需求对存储带宽提出了更高的要求，而 HBM 最新一代产品 HBM3E 每个堆栈的速度达到近 1.2 TB/s，远远超过传统内存。其高带宽、低功耗、高效传输速率等性能使其允许 AI 和 ML 模型实现快速数据访问、大量并行数据处理，因此将成为 AI 时代的核心存力。不仅如此，HBM 能够实现在小空间内的高带宽稳定数据流，支持更高的分辨率与帧率，因此也与图形应用程序非常适配。同时 HBM 的高带宽、低延迟和低能效也将有效优化数据中心的应用程序。

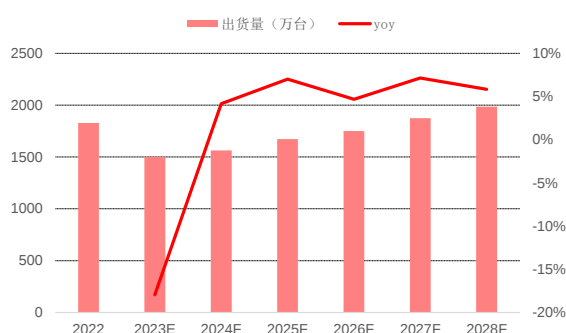
图 33: NVIDIA 及 AMD AI 芯片发展进程及 HBM 规格比较

Company	AI Chips	2022	2023				2024F				2025F			
			1Q23	2Q23	3Q23	4Q23	1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25
NVIDIA	H100	HBM3 8hi 80GB												
	GH200 (CPU+GPU)						HBM3e 8hi 141GB							
	H20						HBM3 8hi 96GB							
	H200						HBM3e 8hi 141GB							
	B100										HBM3e 8hi 192GB			
	GB200 (CPU+GPU)										HBM3e 8hi 192/384GB			
	B200											HBM3e 12hi 288GB		
AMD	MI200	HBM2e 8hi 128GB												
	MI300X					HBM3 12hi 192GB								
	MI300A (CPU+GPU)					HBM3 8hi 128GB								
	MI350										HBM3e 12hi 288GB			
	MI375 (CPU+GPU)												HBM3e 12hi 288GB	

数据来源: trendforce, 中信建投证券

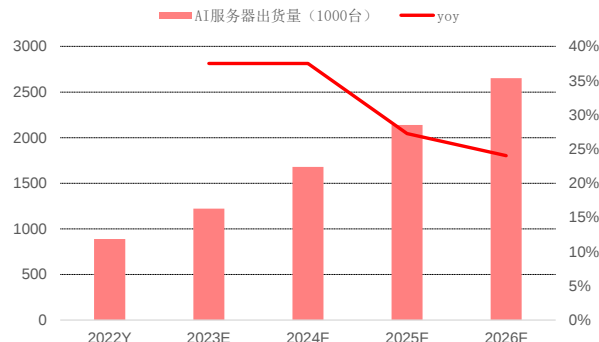
除了 HBM 外, AI 服务器也将显著拉动传统 DRAM 和 NAND Flash 的需求增长。根据 TrendForce 和 FMS 统计及预测, DRAM 容量在服务器中的平均系统容量在 2025 年将实现 18.8% 的增长, 从 718GB 增加到 853GB; 其中存储系统、流媒体、AI/HPC 这三个场景的 DRAM 容量增长显著, 其中 AI/HPC 的 DRAM 容量将增长 17.1%, 从 1792GB 增长至 2098GB, 同时流媒体中的 DRAM 容量也将增长 19.4%, 从 1228GB 增长至 1466GB。同时服务器中的内存容量正在向高容量芯片转变, DDR5 24Gb 产品将占据重要地位, 同时 HBM 的产量也将增加到 10%。NAND 方面, 根据《中国 NAND 闪存行业发展趋势研究与投资前景分析报告》, AI 服务器的 DRAM 需求量是常规服务器的 8 倍, NAND 需求量是常规服务器 3 倍。根据 TrendForce 和 FMS, 受数据中心和云计算需求增加的驱动, 服务器中平均每台设备内 NAND 容量增长率达到 27%, 从 2023 年的 3851GB 大幅增长到 2024 年的 4879GB。

图 34: 全球服务器出货量及预测



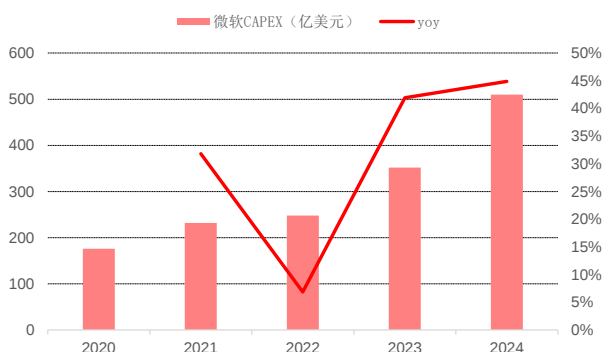
数据来源: DIGITIMES, 中信建投证券

图 35: AI 服务器出货量及预测



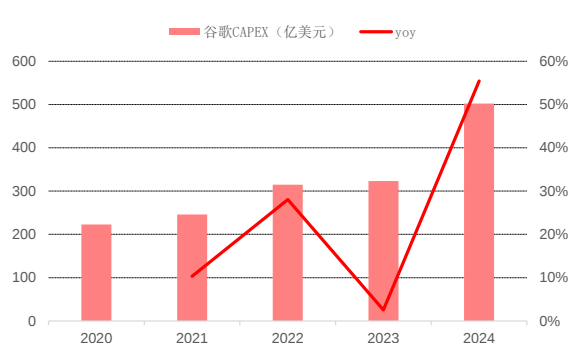
数据来源: Trendforce, 中信建投证券

图 36: 微软 CAPAX 及增速



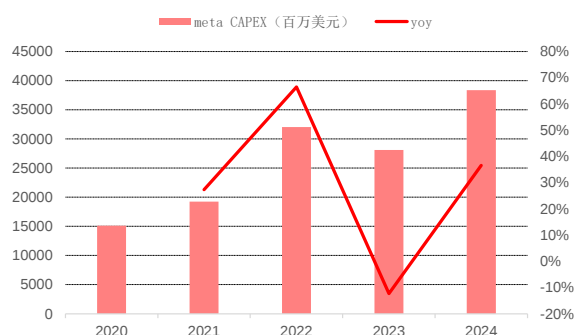
数据来源：公司公告，中信建投证券

图 37: 谷歌 CAPAX 及增速



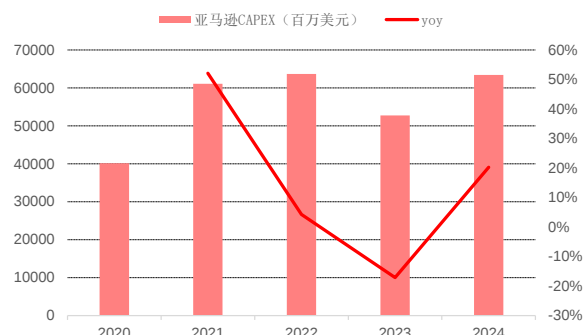
数据来源：公司公告，中信建投证券

图 38: meta CAPAX 及增速



数据来源：公司公告，中信建投证券

图 39: 亚马逊 CAPAX 及增速



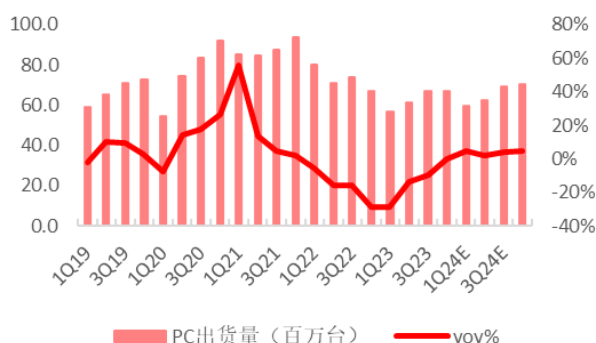
数据来源：公司公告，中信建投证券

（二）PC、手机需求短期仍有待提振

（1）PC 市场：AIPC 加速换机周期

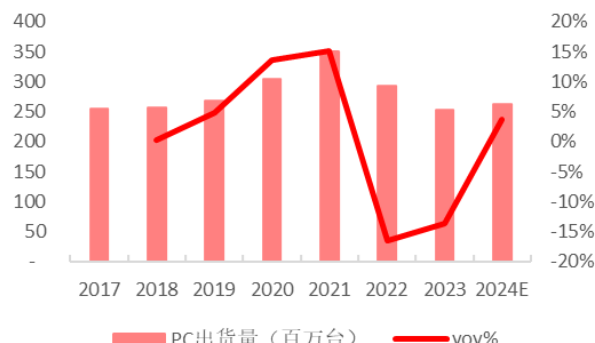
产业进入换机周期，技术革新成为 PC 市场重要推动力。PC 市场较为成熟，整体趋于稳定，2020-2021 年疫情推动线上化需求释放，此后市场出现疲软，考虑到目前 PC 的换机周期普遍为 5 年，疫情售卖的 PC 已进入换机周期，市场收缩几近触底，预计 2024 年将实现正增长。一方面，系统停更或升级将成为换机潮重要动力。微软官方宣布 Win10 系统将于 2025 年 10 月 14 日停止更新，所有版本将全部终止支持，根据 statcounter，截至 2023 年 9 月，Win10 的市场份额高达 71.6%，Win11 的市场份额仅为 23.6%，在 Win10 停更背景下，PC 有较大的换机空间。此外 Win12 也将于 2024 年发布，消费者可能会因为 Windows 的新版本而升级他们的个人电脑，系统更新或将助推换机潮的到来。另一方面，AIPC 通过 AI 能力的本地化部署，有望进一步提升交互体验与工作效率，看好 AIPC 成为 PC 市场下一波增长的重要推动力。Canalys 预计 2024 年 AI PC 的渗透率会达到接近 20%（对应 5200 万台出货量），得益于换机动能和全新本地体验，2027 年 AI PC 将成为主流化的 PC 产品类型，渗透率达到 60% 以上。比起消费者，企业将更积极地采用 AI 技术，2027 年预计 AI PC 出货量的 60% 将应用到商用领域。

图 40: 全球 PC 季度出货量及增速



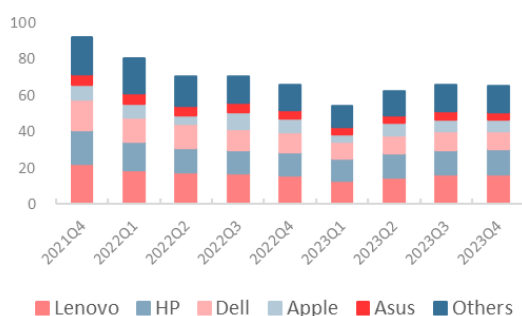
数据来源: IDC, 中信建投证券

图 41: 全球 PC 年度出货量及增速



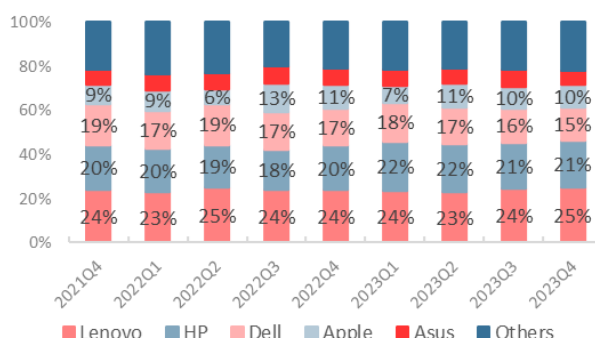
数据来源: IDC, 中信建投证券

图 42: 全球 PC 市场出货量 (百万台)



数据来源: Canlys, 中信建投证券

图 43: 全球 PC 市占率 (%)



数据来源: Canlys, 中信建投证券

表 2: 2024-2028 年全球 AIPC 出货量预测

	2024E	2025E	2026E	2027E	2028E
全球 PC 出货量 (百万台)	261	270	281	292	301
yoy%	4%	3%	4%	4%	3%
AIPC 渗透率预测	17%	25%	35%	50%	60%
全球 AIPC 出货量 (百万台)	44	68	98	146	181

资料来源: IDC, 中信建投证券预测, (注: PC 出货量预测来源于 IDC)

长期来看, PC 端在 LLM 中的角色转型对 DRAM 的需求潜力巨大。目前端侧设备 (如智能手机、PC) 在 LLM 的工作流中, 仅充当客户端角色; 根据 Yole 和 FMS 数据, 如果模型本地化推理在端侧成功推行, 以 10b 参数以下的 AI 模型为例, LLM 就将对 DRAM 的需求从 10GB 提升到 20GB (int8)。同时按照趋势而言, 对 DRAM 的拉动主要依靠于标配 DRAM 容量增加。

短期 PC 需求略有疲软, 但单位设备 DRAM 容量增加趋势不改。尽管 AIPC 时代已悄然拉开序幕, 但仅看 2024 年实际增长情况, PC 需求并未立即显著改善, 根据 Yole 和 FMS 统计及预测, 今年全球 PC 出货量或仅为小个位数增速。但 PC 设备中的 DRAM 含量仍在增加, 2023 年 8 及 8GB 以下是主流, 预计到 2029 年, 近 75%

请务必阅读正文之后的免责条款和声明。

的单位设备 DRAM 含量将达到 16GB 及以上。而根据 FMS 和 Tendforce 数据，2023 年笔记本电脑中平均每台设备 NAND 容量为 612.91GB，预计 2024 年将达到 671.12GB，同比增长 9.5%。

图 44:PC 对于存储配置的需求与日俱增

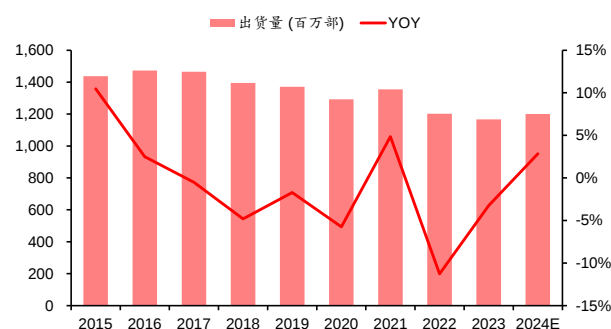


数据来源: YOLE, 中信建投证券

(2) 智能手机: AI 赋能是主要驱动力

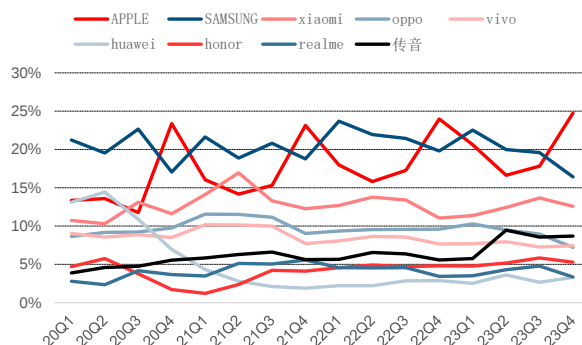
2023 年全球智能手机市场跌幅收窄，2024 年有望实现反弹。全球智能手机 2023 年出货量约 11.7 亿台，同比下降 3.2%。随着消费逐步复苏，折叠屏等新品起量以及 AI 赋能，2024 年全球智能手机市场或将回归增长趋势，根据 IDC 预测，2024 年全球智能手机出货量将恢复到约 12 亿台，同比增长 2.8%。重大创新是手机换机潮的核心驱动力，若 AI 手机实现使用体验的革命性创新，将复刻智能手机取代功能手机的高速增长。通过融入 AI 大模型，新一代 AI 手机有望改善用户体验、创造差异化竞争优势，成为缩短手机换机周期和加速市场复苏的关键驱动力。

图 45: 全球智能手机年度出货量



数据来源: IDC, 中信建投证券

图 46: 全球智能手机厂商市场份额变化



数据来源: IDC, 中信建投证券

三星打响 AI 手机第一枪，苹果 Apple Intelligence 初版落地 iOS 18.1 开发者测试版。2024 年 1 月，三星发布旗舰手机 Galaxy S24 系列，采用谷歌云平台 Vertex AI 上配置大模型 Gemini Pro（概括理解多模态信息）和 Imagen 2（文生图、照片编辑），支持端侧模型 Gemini Nano（处理端侧任务效率最高模型）。Galaxy S24 的 AI 功能主要聚焦三大高频场景：（1）翻译/笔记（端侧 AI 双向实时翻译，自动摘要并生成会议纪要）；（2）

搜索（圈选即搜）；（3）影像（自动分析照片并建议编辑工具，照片优化/视频补帧）。2024WWDC 苹果发布 Apple Intelligence，作为系统级的 AI 集成，Apple Intelligence 具备强大的理解能力，支持多轮对话、屏幕内容感知、应用智能交互等。Apple Intelligence 采用设备端+云端处理，端侧结合 A/M 系列芯片的硬件支持以及苹果内置的大语言模型处理一般问题；云端侧，苹果正式宣布了与 OpenAI 合作，苹果将 GPT-4o 支持下的 ChatGPT 集成到 iOS、iPadOS 和 macOS 系统中，端侧+云端可以实现无缝切换。7 月 29 日，苹果 Apple Intelligence 初版落地 iOS 18.1 开发者测试版，该版本暂未融入 ChatGPT，首批推出的功能包括：writing tools、Siri 的新功能、邮件的全文摘要等新功能、专注模式、可用日常语言查找照片、语音转录等，其余功能有望在明年上线。目前 Apple Intelligence 仅支持 A17 Pro 芯片以及 M 系列芯片及后续迭代的产品机型，其中对 iPad/Mac 产品要求不高，2020 年后的产品基本满足要求；但在手机端，目前仅有 iPhone15 Pro 和 iPhone15 Pro Max 支持 Apple Intelligence，AI 终端生态加速成熟+硬件机型限制，有望刺激用户换机需求。

图 47: Galaxy S24 的主要 AI 功能



数据来源：三星，中信建投证券

图 48: Apple Intelligence



数据来源：苹果 WWDC2024 开发者大会，中信建投证券

表 3: Apple Intelligence 带来的功能更新

应用	新增功能
照片	- 支持智能图像搜索能力 - 按相册自动组织相关事件照片，提供精准的背景音乐推荐 - 改进的编辑工具可进行更复杂的图像处理，例如精确去除照片中任何特定元素
邮件	- 邮件内容根据优先级自动排序并置顶 - 提供了改写重写、写作建议、文本总结等功能，并根据不同的语气场景生成不同版的文本
通话	允许通话录音，支持语音转成文字并写成总结
备忘录	- 引入智能笔记功能（Smart Script），这一功能使得编辑手写文本变得像编辑键入文本一样简单和便捷。利用端侧机器学习自动清理用户的书写内容，输入编辑键入文本，输出手写文本 - 笔记中的内容直接生成可视化图像
Siri	- 拥有更加强大的理解能力，即使语法不完整的语音指令，它也能准确理解并且可以进行更深层次的多轮对话 - 支持更多自然语言交互形式，如文本和语音混合输入 - 能够执行更广泛的应用内及跨应用操作，例如从日历应用获取会议时间后，自动在备忘录中设置提醒，或者在地图应用中根据电子邮件中提及的地址导航
Writing Tools	引入了先进的语言处理技术和智能建议功能，如改写重写、文本总结和文本结构优化，或者根

请务必阅读正文之后的免责条款和声明。

据文本内容提供符合语境的修改建议，帮助用户提升写作质量

- 能运用到备忘录、Pages、Keynote 及其他应用程序中
- 用户可以根据不同的图像主题和风格，要求创建符合心意的图像，提供高度定制化的视觉内容
- Genmoji、Image Playground 和 Image Wand

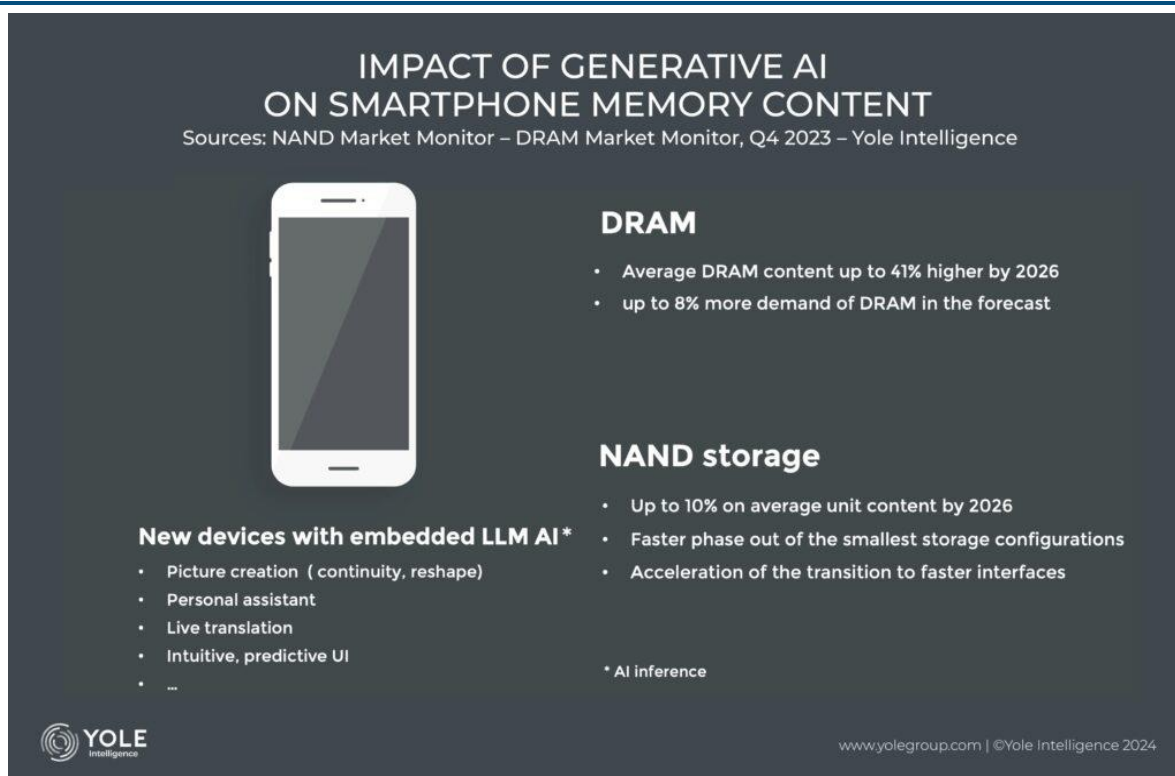
图像工具

资料来源：苹果 WWDC2024 开发者大会，中信建投证券

AI 手机渗透率提升叠加大模型本地化部署将助力 DRAM 需求增长。随着苹果与安卓系手机厂商纷纷加入 AI 手机布局，AI 手机渗透率预计将快速提升。根据 Canlys 统计及预测，2023 年全球 AI 手机渗透率仅 5%，但预计到 2027 年将快速攀升至 47%，2026 年全球 AI 手机累计出货将超过 10 亿部。根据 IDC 统计及预测，中国 AI 手机出货量预计在 2027 年达到 1.5 亿部，市场渗透率超 50%。AI 手机的处理需求也将要求更高的单位设备 DRAM 容量，满足更高的处理需求，如果在手机上推行模型本地化推理，根据 Yole 推算，以 10b 参数为例，LLM 所需的 DRAM 容量将从 1GB 提升到 7GB(INT8)。随着 AI 手机渗透率的增加，智能手机单位设备平均 DRAM 容量将从 2023 年的 9GB 增长至 2026 年的 12GB。

AI 手机也将推动 NAND 向更高容量更快接口转变。随着 LLM 的嵌入，智能手机对内存的需求增加，且 NAND 方面将加速淘汰最小的存储配置并加速过渡到更快的接口，到 2026 智能手机中 NAND 存储的平均单位内容容量预计将增加高达 10%。而根据 FMS 和 Tendforce 统计及预测，2023 年智能手机中平均每台设备 NAND 容量为 215GB，预计 2024 年增长到 233GB，涨幅 8.5%。

图 49:AIGC 对手机存储的影响



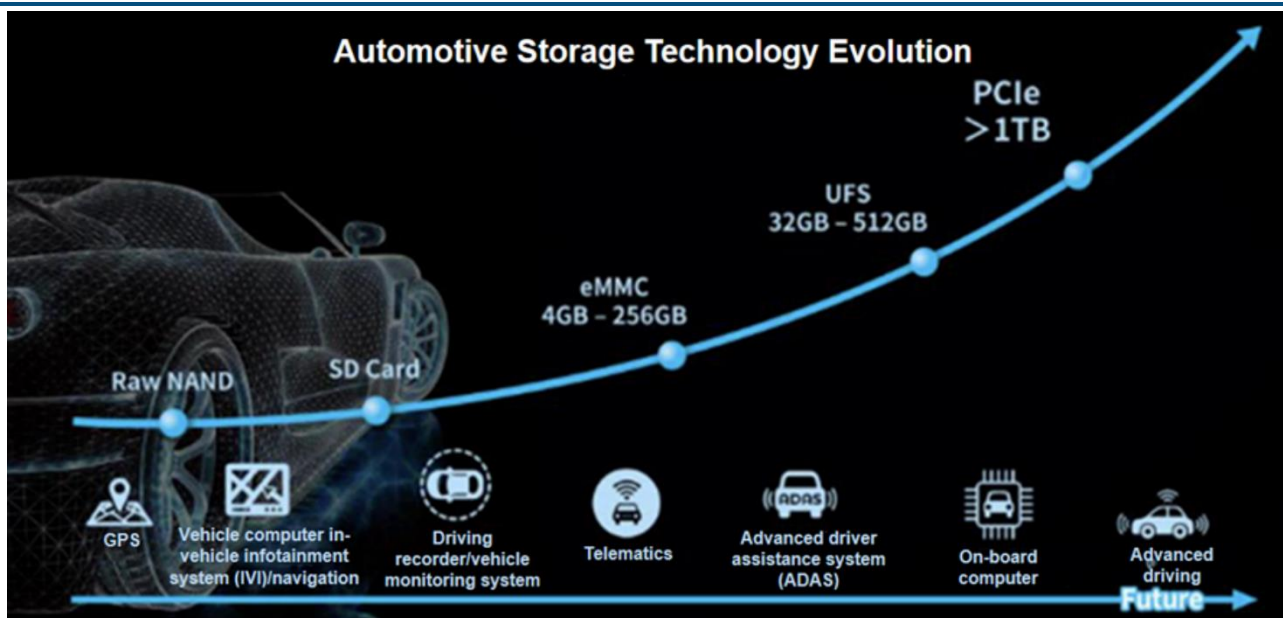
数据来源：YOLE，中信建投证券

（三）智能汽车市场远期潜力巨大

智能驾驶导致车规级 DRAM 需求向高带宽大容量方向转变。自动驾驶 ADAS 系统依赖于摄像头、雷达、GPS 等各种传感器来感知车辆周围环境并收集数据，然后通过 AI 训练精细信息分析、实现静态和动态物体的辨识追踪并向油门、刹车等系统下达执行指令，整个过程将产生大量数据，同时需要更高的读取速度支持算法进行，因此对于 DRAM 的带宽和容量将有更高的需求。根据美光科技统计及预测，L3 级自动驾驶汽车对内存带宽需求为 200GB/s；L1/2 级对 DRAM 的平均容量需求约 8GB，L3 和 L5 则将分别增长至 16GB 和 74GB。

汽车智能化也对 NAND 容量提出了更高的要求。汽车智能化水平的提升对于车载芯片的数据存储量有更高的要求。根据 Yole 数据，2020-2026 年，NAND 平均存储容量 CAGR 将达到 57%。同时，UFS 读写效率、延时、和容量方面相比更有 e-MMC 优势，未来 UFS 将对 e-MMC 实现逐步替代，更远来看，随着技术与成本的不断优化，能实现 1TB 以上容量的 PCIe 也将逐步成为车载 NAND 的主流选择。

图 50: 智驾行业内存技术演进



数据来源: Silicon Motion, 中信建投证券

图 51: 常见智能驾驶 soc 芯片内存

Vendor	Model	Maximum memory supported	Memory width	Memory bandwidth
Nvidia	Orin	LPDDR5x	256bit	204.8GB/s
Nvidia	Xavier	LPDDR4X	256bit	137GB/s
Qualcomm	SA8155P	LPDDR4X	128bit	68GB/s
Qualcomm	SA8295P	LPDDR4X	256bit	137GB/s
Qualcomm	SA8255P	LPDDR5		
SiEngine Technology	Longying No.1	LPDDR5		51.2GB/s
Huawei	Ascend 310	LPDDR4X	128bit	59.8GB/s
Huawei	Ascend 610	LPDDR4X	256bit (estimated)	137GB/s (estimated)
Tesla	First-generation FSD	LPDDR4	128bit	34GB/s
Tesla	Second-generation FSD	GDDR6	256bit (estimated)	448GB/s (estimated)
Renesas	V4H	LPDDR5	64bit	51.2GB/s
Mobileye	EyeQ5	LPDDR4	32bit	8.5GB/s (estimated)
Mobileye	EyeQ6L	LPDDR4X	64bit (estimated)	17GB/s (estimated)
Mobileye	EyeQ6H	LPDDR5	128bit (estimated)	102.4GB/s (estimated)
Mobileye	EyeQ6 ULTRA	LPDDR5X	128bit (estimated)	136.2GB/s (estimated)
Nvidia	RTX4090	GDDR6X	384bit	1008GB/s
Nvidia	RTX3050	GDDR6	128bit	224GB/s

数据来源: ResearchinChina, 中信建投证券

（四）AI 算力需求景气支撑 HBM 价格，消费电子景气疲弱令传统 DRAM 及 NAND 价格短期承压

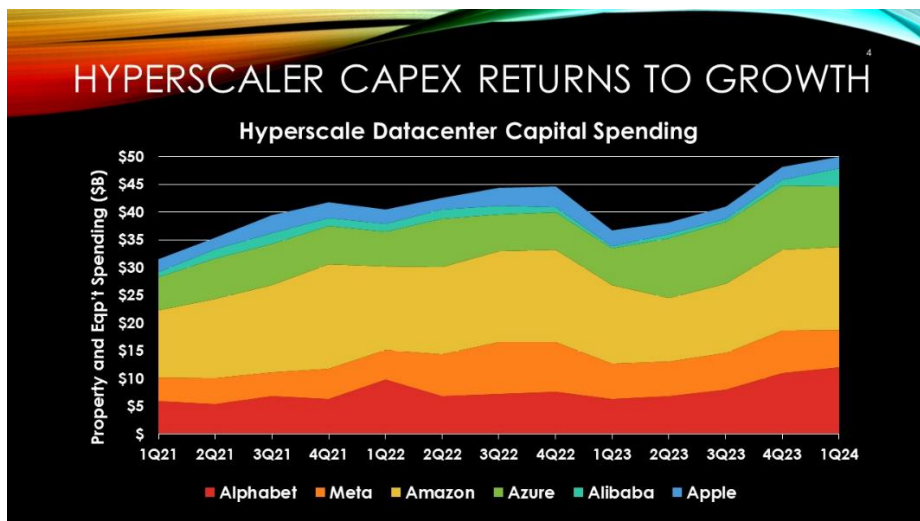
短期内 HBM 需求强劲，供不应求。由于 HBM 在 AI 时代的性能优势突出，数据中心与 AI 服务器对 HBM 需求强劲，买方积极采购但受限于厂商 HBM 产能制约，目前仍处于供不应求的状态，2025 年客户就 HBM 数量和价格谈判已基本完成。三大厂商面对英伟达和 CSP 自研芯片的加单下，加开 TSV 产线以扩大 HBM 产能，预计 HBM 供给位元量将年增 105%，但受制于 HBM 的长交货期，可能直至 2025 年第二季度才有望释放。2025 年 HBM 需求位元将有超过 80% 是 HBM3e，其中 12hi 的占比将超过一半，其次则是 8hi。因此，可能对 HBM2e 和 HBM3 等上一代产品的价格有所冲击，但 HBM3e 的价格仍将维持高位。

中长期看，对 HBM 过度投资存在风险。HBM 目前溢价较高，HBM 目前价格为 DDR5 的 5-6 倍，成本大约为 DDR5 的 2 到 3 倍，溢价主要是由单一来源采购和短缺导致。而 2023~2024 年是 AI 建设爆发期，大量需求集中在 AI 训练，相关芯片拉动 HBM 需求高涨，转向推理阶段之后，对 HBM 需求将略为收敛；同时相关芯片需求与超大规模公司的资本支出有较高的依赖度，目前主要 CSP 资本投入增长较快，尤其是 Alphabet、Amazon 和 Azure，而超大规模公司的资本支出面临一定的不确定性，如果现在的超高速增长放缓，叠加 HBM 的长交货期，将对三大厂商在产量和库存的把控上带来较大的挑战。现在三大厂商纷纷扩产以满足客户需求，年增速达

请务必阅读正文之后的免责条款和声明。

到 105%，但目前的需求可能存在买方预期 HBM 可能缺货因此超额下单的风险，如果出现供给过剩、竞争加剧的情况，将对 HBM 的价格带来较大的负面影响。因此，厂商需要在短期内扩产满足用户需求和过度扩产导致供过于求之间做出平衡。

图 52: 全球主要云厂商资本支出（市场需求波动剧烈）



数据来源: FMS-2024 闪存峰会, 中信建投证券

传统 DRAM 产品买方优先去库存导致价格涨幅收敛，HBM 拉动总体收入。DRAM 生产在 2024 年下半年开始恢复，预计 2025 年的总体产出水平将保持在 2024 年季度水平之上。PC 端 DRAM 由于 PC OEM 在第三季传统销售旺季未达预期，加上 DRAM 采购成本升高，将有更强的 DRAM 去库存意愿，第四季采购量将减少；同时 HBM 挤出 DRAM 产能，供给减少，因此价格大致持平。LPDDR 方面，智能手机品牌厂可能延迟采购以争取有利的合约价，同时 LPDDR4X 供给增加，供过于求导致价格下降，LPDDR5X 则供给与库存健康价格稳定。消费类 DRAM 需求依旧疲软，同时部分供应商冲刺出货目标，因此总体呈现价格下降风险。服务器 DRAM(非 HBM)方面，旺季可能对 DDR5 采购需求有所拉升，但同时 DDR5 与 DDR4 的价差较大可能导致部分需求从 DDR5 转向 DDR4。整体而言，传统 DRAM 受到买方去库存以及 PC 和智能手机消费疲软的需求，价格涨幅受限；HBM 供不应求，维持较高溢价，因此 HBM 渗透率将达到 7%并对总体 DRAM 收入有较大的拉动。

图 53: 2024-2025 年 DRAM 各领域出货预测



Source: Company Data, inSpectrum estimates, Sep. 2024

数据来源: ADATA, inSpectrum, 中信建投证券

请务必阅读正文之后的免责条款和声明。

图 54:2024-2025 年 DDR4 价格和供需缺口预测



数据来源: ADATA, inSpectrum, 中信建投证券

图 55:3Q24 与 4Q24 名 DRAM 产品价格预测

	3Q24	4Q24E
PC DRAM	DDR4: up 8~13% DDR5: up 8~13% Blended: up 8~13%	DDR4: mostly flat DDR5: mostly flat Blended: mostly flat
Server DRAM	DDR4: up 8~13% DDR5: up 13~18% Blended: up 13~18%	DDR4: mostly flat DDR5: up 3~8% Blended: up 0~5%
Mobile DRAM	mostly flat	LPDDR4X: down 5~10% LPDDR5X: mostly flat
Graphics DRAM	up 3~8%	mostly flat
Consumer DRAM	DDR3: mostly flat DDR4: up 3~8%	DDR3: down 0~5% DDR4: mostly flat
Total DRAM	Conventional DRAM: up 8~13% HBM Blended: up 10~15% (HBM Penetration: 6%)	Conventional DRAM: up 0~5% HBM Blended: up 8~13% (HBM Penetration: 7%)

数据来源: trendforce, 中信建投证券

供给增加和需求疲软导致 NAND 价格压力较大。第三季度消费市场的主要需求智能手机和 PC 都未出现好转;第四季度 PC 消费疲软、现货与合约价差过大导致 PCNAND 合约价降低 5%-10%, 智能手机方面由于 iPhone 16、华为三折叠与其他安卓系的新机发布需求有所改善, 但厂商库存增加, 买方避免库存压力压低价格, 最终 eMMC 合约价降低 8%-13%。企业级 SSD 的采购需求动能依旧高于其他 NAND 产品, 但 CSP 采购高峰已过, 整体需求有所下降, 同时厂商增产导致供过于求, 最终导致增长幅度在 5% 以下。整体而言, 受需求疲软与供给增加的影响, NAND 价格承压较大, 仅企业级 SSD 受 AI 服务器建设需求推动, 尚能保持增长。

请务必阅读正文之后的免责条款和声明。

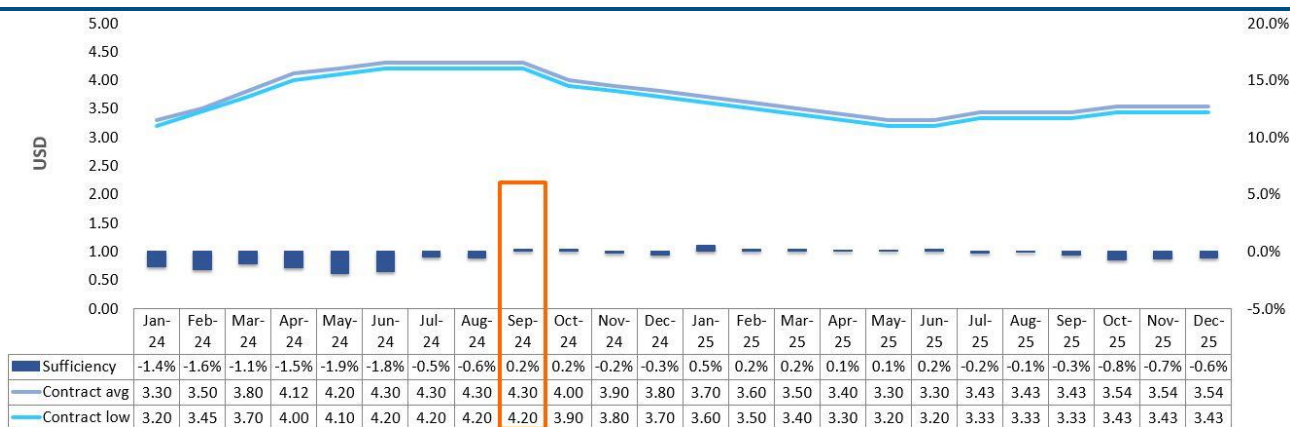
图 56: 2024-2025 年 SSD 出货量预测



Source: Company Data, inSpectrum estimates, Sep. 2024

数据来源: ADATA, inSpectrum, 中信建投证券

图 57: 2024-2025 年 NAND 价格和供需缺口预测



Source: Company Data, inSpectrum estimates, Sep. 2024

数据来源: ADATA, inSpectrum, 中信建投证券

图 58: 3Q24 与 4Q24 名 NAND 产品价格预测

	3Q24	4Q24E
eMMC UFS	mostly flat	down 8~13%
Enterprise SSD	up 15~20%	up 0~5%
Client SSD	up 3~8%	down 5~10%
3D NAND Wafers (TLC & QLC)	down 3~8%	down 10~15%
Total NAND Flash	up 5~10%	down 3~8%

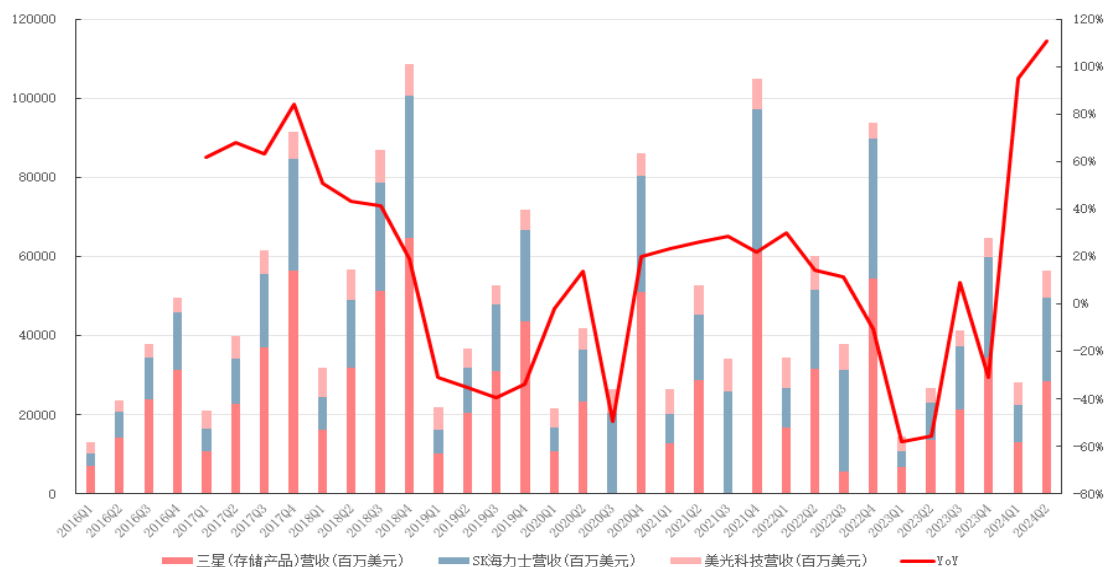
数据来源: trendforce, 中信建投证券

短期传统 DRAM 整体偏中性, NAND 略有疲软。短期来看, 传统 DRAM 虽然产能受 HBM 挤出导致供给有所改善, 但传统需求 PC 和智能手机依旧疲软, 叠加买方去库存策略, 整体价格改善不明显, 未来的价格回升一定程度依赖于 PC 和手机市场的回暖。同时 DDR5 替代 DDR4 的进程正处于加速阶段, DDR5 的供给厂商更为集中, 议价能力更强, 预计也将对美光 DRAM 整体价格有所支撑。NAND 受到消费电子疲软的冲击更大, 除了受 AI 服务器推动的企业级 SSD 之外, 其他价格略有下降。

HBM 渗透率正在提升, AI 需求保持强劲, 但对远期供需持谨慎态度。AI 模型对高带宽存力的需求导致 HBM 需求旺盛, 但受制于 TSV 产能, HBM 供不应求, 维持高溢价, 2025 年 HBM3e 产能已经大量售出。尽管三大厂商都有扩产, 但新加产能最早 2025 年第二季度才能释放。由于量产 HBM3e 12hi 的学习曲线长, 三大厂商能否如期落实 HBM3e 产能还有待验证。同时中长期来看, 受制于超大规模企业资本支出的不确定性, 以及模型从训练转入推理阶段之后, AI 芯片对 HBM 的高速拉动能否维持还有待验证, 厂商需要在短期扩产以满足需求和中长期过度建设导致供给过剩之间取得平衡。

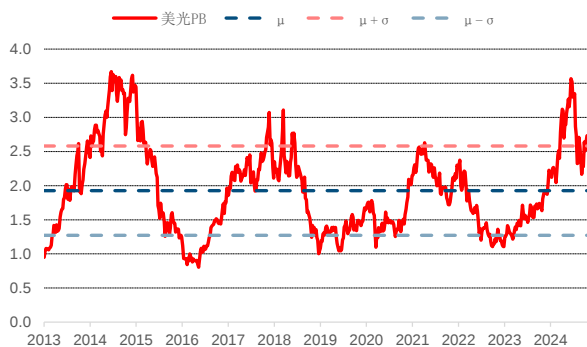
中期前景依然看好, 本轮半导体存储周期尚未结束, 端侧传统需求可能成为未来市场的主要驱动力。尽管存在上述担忧, 但中期而言, 半导体存储的前景依然光明。首先, 随着厂商减产计划的完成和 HBM 的挤出效应, 整体库存水平将逐渐回复到健康状态, 同时本次短期的买方去库存行为难以长期持续, 本轮半导体存储周期尚未结束。其次, 随着 AI 设备渗透率的逐渐上升, 以及端侧设备 (PC 和智能手机) 在 LLM 中的角色从客户端转向模型推理本地化, 单位设备中 DRAM 和 NAND 的容量需求将持续提升, 同时 AIPC 和 AI 手机也将带动整体需求量的增长。因此 AI 时代下, 端侧传统需求在未来存储市场可能依然是主要驱动力。

图 59: 本轮半导体存储周期尚未结束



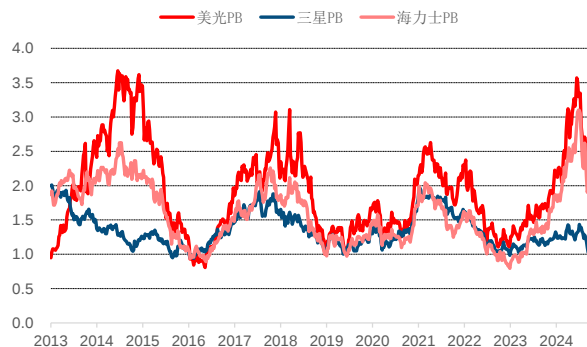
数据来源: Wind, 中信建投证券

图 60: 美光 PB



数据来源: Wind, 中信建投证券

图 61: 三星、海力士、美光 PB 比较



数据来源: Wind, 中信建投证券

四、盈利预测和估值

我们预计 2025-2026 财年美光营业收入分别为 382.97 亿和 458.62 亿美元，同比增长 52.51%和 19.75%，其中 DRAM 收入分别为 272.68 亿和 333.22 亿美元，同比增长 55%和 22%，NAND 收入分别为 107.59 亿和 122.74 亿美元，同比增长 49%和 14%。预计 2025-2026 财年公司 Non-GAAP 毛利率分别为 42.64%和 48.40%，Non-GAAP 净利润分别为 100.54 亿和 139.85 亿美元，同比增长 577%和 39%，对应净利率分别为 26.25%和 30.49%。

考虑到美光在全球存储以及 HBM 领域的优势地位，按照 PB 估值法，给予其 2025 财年 3X PB，对应目标价 140.42 美元/ADS，按照 PE 估值法，给予其 2025 财年 16X PE，对应目标价 145.08 美元/ADS。首次覆盖，给予“买入”评级，综合 PB 和 PE 估值，给予目标价 142.75 美元/ADS。

图 62:美光分业务收入预测

单位:百万美元	FY2018A	FY2019A	FY2020A	FY2021A	FY2022A	FY2023A	FY2024A	FY2025E	FY2026E	FY2027E
总收入	30,391	23,406	21,435	27,705	30,758	15,540	25,111	38,297	45,862	46,696
bytechnology										
DRAM	22,606	16,814	14,434	20,008	22,386	10,978	17,603	27,268	33,322	33,690
NAND	6,761	5,480	6,174	6,963	7,811	4,207	7,227	10,759	12,274	12,740
Other(primarilyNOR)	1,024	1,113	828	734	561	356	281	270	267	266
收入增速YoY	50%	-23%	-8%	29%	11%	-49%	62%	53%	20%	2%
bytechnology										
DRAM	63%	-26%	-14%	39%	12%	-51%	60%	55%	22%	1%
NAND	20%	-19%	13%	13%	12%	-46%	72%	49%	14%	4%
Other(primarilyNOR)	27%	9%	-26%	-11%	-24%	-37%	-21%	-4%	-1%	0%
收入占比										
bytechnology										
DRAM	74%	72%	67%	72%	73%	71%	70%	71%	73%	72%
NAND	22%	23%	29%	25%	25%	27%	29%	28%	27%	27%
Other(primarilyNOR)	3%	5%	4%	3%	2%	2%	1%	1%	1%	1%

数据来源: Bloomberg, 中信建投证券

表 4:美光盈利预测（单位：百万美元）

资产负债表	FY2023A	FY2024A	FY2025E	FY2026E	利润表	FY2023A	FY2024A	FY2025E	FY2026E
总流动资产	21,244	24,372	26,361	33,367	营业收入	15,540	25,111	38,297	45,862
总非流动资产	43,010	45,044	49,733	52,814	营业成本	-16,956	-19,498	-22,325	-24,021
总资产	64,254	69,416	76,093	86,180	毛利	-1,416	5,613	15,973	21,841
总流动负债	4,765	9,248	9,160	9,402	研发费用	-3,114	-3,430	-3,626	-4,011
总非流动负债	15,369	15,037	15,037	15,037	SG&A 费用	-920	-1,129	-1,226	-1,468
总负债	20,134	24,285	24,197	24,439	重组费用资产减值	-171	-1	0	0
普通股	124	125	125	125	其他收入/（支出）	-124	251	0	0
股本溢价	11,036	12,115	12,948	13,781	净额-经营				
留存收益	40,824	40,877	46,810	55,822	EBIT	-5,745	1,304	11,121	16,363
减:库存股	7,552	7,852	7,852	7,852	利息收入（支出）	80	-33	-20	-20
其他综合收益	-312	-134	-134	-134	其他收益/（支出）	7	-31	0	0
归属母公司股东权益	44,120	45,131	51,897	61,742	净额-非经营				
少数股东权益	0	0	0	0	EBT	-5,658	1,240	11,101	16,343
股东权益合计	44,120	45,131	51,897	61,742	所得税	-177	-451	-2,775	-4,086
总权益及负债	64,254	69,416	76,093	86,180	权益法下长期股权投资损益	2	-11	0	0
现金流量表					GAAP 净利润（含少数股东权益）	-5,833	778	8,326	12,257
经营活动所得					Non-GAAP 净利润	-4,863	1,484	10,054	13,985
现金净额	1,559	8,507	15,849	20,060	关键比率				
投资活动所得	-6,191	-8,309	-13,184	-12,428	营业收入 YoY	-49.48%	61.59%	52.51%	19.75%
融资活动所得	4,983	-1,842	-3,254	-4,214	毛利率	-9.11%	22.35%	41.71%	47.62%
现金净额					GAAP 净利润 YoY	-167.15%	113.34%	970.13%	47.22%
汇率变动对现金的影响	-34	40	0	0	Non-GAAP 净利润 YoY	-151.35%	130.52%	577.46%	39.11%
现金及现金等价物净增加额	317	-1,604	-588	3,418	净利率	-31.29%	5.91%	26.25%	30.49%
期初现金及现金等价物余额	8,339	8,656	7,052	6,464					
期末现金及现金等价物余额	8,656	7,052	6,464	9,882					

资料来源: Bloomberg, 中信建投证券

风险提示

全球宏观经济超预期下行；地缘政治风险；生成式 AI 发展低预期，AI 算力芯片需求低预期，Mag7 资本开支低预期；AI PC 以及 AI 手机发布低预期，PC 及手机换机进度低预期；汽车 SoC 需求持续低预期，终端 Tier1 库存去化低预期；存储行业供给超预期增长，存储价格超预期下行；行业竞争加剧，客户流失，来自三星、海力士的竞争；HBM 制程迭代低预期，新品发布低预期；美联储降息进程低预期；其他影响半导体行业的宏观和中观风险。

敏感性分析：

由于全球宏观经济发展的不确定性，公司 DRAM 和 NAND 收入可能出现和预期不一致的情况。我们预计 2025 财年公司 DRAM 收入增速在 30%-65% 之间，NAND 收入增速在 30%-60% 之间，其他参数和预期基本一致，对公司 2025 财年总营收进行如下敏感性测算：

表 5: 公司营收敏感性测算

2025 财年公司收入（百万美元）		2025 财年 DRAM 收入增速				
		30.00%	50.00%	54.90%	60.00%	65.00%
2025 财年 NAND 收入增速	30.00%	32549	36070	36933	37830	38710
	45.00%	33633	37154	38017	38914	39794
	48.88%	33913	37434	38297	39194	40074
	55.00%	34356	37876	38740	39637	40517
	60.00%	34717	38238	39101	39998	40878

资料来源：Bloomberg，中信建投证券

分析师介绍

崔世峰

海外研究首席分析师，南京大学硕士，7 年买方及卖方复合从业经历，专注于互联网及科技公司研究，擅长游戏行业研究。2022-2023 年新财富港股及海外最佳研究团队入围，2019-2020 年新财富传媒最佳研究团队第二名核心成员。

于伯韬

FRM，香港大学金融学硕士，武汉大学经济学学士，6 年互联网科技及港股策略卖方从业经历，2021、2020 年新财富港股及海外方向第五名，2022 年新浪金麒麟港股及海外市场最佳分析师第三名，2020 年新浪金麒麟港股及海外市场新锐分析师第一名。

评级说明

投资评级标准		评级	说明
报告中投资建议涉及的评级标准为报告发布日后的 6 个月内的相对市场表现，也即报告发布日后的 6 个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A 股市场以沪深 300 指数作为基准；新三板市场以三板成指为基准；香港市场以恒生指数作为基准；美国市场以标普 500 指数为基准。	股票评级	买入	相对涨幅 15% 以上
		增持	相对涨幅 5%—15%
		中性	相对涨幅-5%—5%之间
		减持	相对跌幅 5%—15%
		卖出	相对跌幅 15% 以上
	行业评级	强于大市	相对涨幅 10%以上
		中性	相对涨幅-10—10%之间
		弱于大市	相对跌幅 10%以上

分析师声明

本报告署名分析师在此声明：（i）以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，结论不受任何第三方的授意或影响。（ii）本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接收到任何形式的补偿。

法律主体说明

本报告由中信建投证券股份有限公司及/或其附属机构（以下合称“中信建投”）制作，由中信建投证券股份有限公司在中华人民共和国（仅为本报告目的，不包括香港、澳门、台湾）提供。中信建投证券股份有限公司具有中国证监会许可的投资咨询业务资格，本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格证书编号已披露在报告首页。

在遵守适用的法律法规情况下，本报告亦可能由中信建投（国际）证券有限公司在香港提供。本报告作者所持香港证监会牌照的中央编号已披露在报告首页。

一般性声明

本报告由中信建投制作。发送本报告不构成任何合同或承诺的基础，不因接收者收到本报告而视其为中信建投客户。

本报告的信息均来源于中信建投认为可靠的公开资料，但中信建投对这些信息的准确性及完整性不作任何保证。本报告所载观点、评估和预测仅反映本报告出具日该分析师的判断，该等观点、评估和预测可能在不发出通知的情况下有所变更，亦有可能因使用不同假设和标准或者采用不同分析方法而与中信建投其他部门、人员口头或书面表达的意见不同或相反。本报告所引证券或其他金融工具的过往业绩不代表其未来表现。报告中所含任何具有预测性质的内容皆基于相应的假设条件，而任何假设条件都可能随时发生变化并影响实际投资收益。中信建投不承诺、不保证本报告所含具有预测性质的内容必然得以实现。

本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况，报告接收者应当独立评估本报告所含信息，基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。中信建投建议所有投资者应就任何潜在投资向其税务、会计或法律顾问咨询。不论报告接收者是否根据本报告做出投资决策，中信建投都不对该等投资决策提供任何形式的担保，亦不以任何形式分享投资收益或者分担投资损失。中信建投不对使用本报告所产生的任何直接或间接损失承担责任。

在法律法规及监管规定允许的范围内，中信建投可能持有并交易本报告中所提公司的股份或其他财产权益，也可能在过去 12 个月、目前或者将来为本报中所提公司提供或者争取为其提供投资银行、做市交易、财务顾问或其他金融服务。本报告内容真实、准确、完整地反映了署名分析师的观点，分析师的薪酬无论过去、现在或未来都不会直接或间接与其所撰写报告中的具体观点相联系，分析师亦不会因撰写本报告而获取不当利益。

本报告为中信建投所有。未经中信建投事先书面许可，任何机构和/或个人不得以任何形式转发、翻版、复制、发布或引用本报告全部或部分内容，亦不得从未经中信建投书面授权的任何机构、个人或其运营的媒体平台接收、翻版、复制或引用本报告全部或部分内容。版权所有，违者必究。

中信建投证券研究发展部

北京
朝阳区景辉街 16 号院 1 号楼 18 层
电话：（8610）56135088
联系人：李祉瑶
邮箱：lizhiyao@csc.com.cn

上海
上海浦东新区浦东南路 528 号南塔 2103 室
电话：（8621）6882-1600
联系人：翁起帆
邮箱：wengqifan@csc.com.cn

深圳
福田区福中三路与鹏程一路交汇处广电金融中心 35 楼
电话：（86755）8252-1369
联系人：曹莹
邮箱：caoying@csc.com.cn

中信建投（国际）

香港
中环交易广场 2 期 18 楼
电话：（852）3465-5600
联系人：刘泓麟
邮箱：charleneliu@csci.hk