

“先发+技术”铸就 HBM 护城河，能否继续保持领先？

华泰研究

2024 年 6 月 25 日 | 韩国

首次覆盖

半导体

投资评级(首评):

买入

目标价(韩元):

300,000.00

研究员

何翩翩

SAC No. S0570523020002

purdyho@htsc.com

SFC No. ASI353

+(852) 3658 6000

基本数据

目标价(韩元)	300,000.00
收盘价(韩元 截至 6 月 24 日)	223,000
市值(韩元百万)	153,454,919
6 个月平均日成交额(韩元百万)	650,594
	109,004.98-243,
52 周价格范围(韩元)	000.00
BVPS(韩元)	83,117

股价走势图



资料来源: S&P

HBM 龙头地位有望维持，首次覆盖给予“买入”评级，目标价 30 万韩元

23 年至今 SK 海力士估值高歌猛进，从 23 年 1 月 1 日的 0.8x Forward PB 跃升至 24 年 6 月 21 日的 1.9x，超过 14 年历史峰值，主要系其 HBM3 于 22 年 6 月率先量产独供英伟达 H100，并于 23 年先后发布 12 层 HBM3 和 8 层 HBM3E，坐稳 HBM 龙头。我们认为，24-25 年海力士估值持续突破的关键仍是 HBM：1) AI 浪潮推动 HBM 需求高增，伴随 HBM3/3E 营收占比提升，带动利润率优化及估值上修；2) 传统 DRAM 和 NAND 迎来复苏，营收基本盘或“锦上添花”。对比美光和三星，25 年彭博一致预期为 2.8x 和 1.3x，鉴于 14 年以来美股纳指 PB 约为韩国高科技板块 KOSDAQ 的 2.5 倍，结合其 HBM 龙头地位，海力士有望突破历史高位，我们认为 FY25 PB 约 2.5x 较合理，首次覆盖给予“买入”评级，目标价 30 万韩元。

HBM 战场三足鼎立，海力士制程与键合技术有望保持领先美光和三星

HBM 在内存产品里技术壁垒较高，竞争格局三足鼎立。海力士 1-β 制程和键合技术保持领先。24 年以来，尽管先前“HBM3 独供”格局已被打破，但通过背靠英伟达并持续拓展客户，海力士 HBM 需求基础仍然稳固，同时键合良率已近 80%。相较美光，海力士具有产能和良率稳健双重优势，可更充分受益于下游需求增长；而三星截至 24 年 5 月仍未进入英伟达供应链。

DRAM 业务：“先发+技术”铸就 HBM 护城河，传统内存也巧借 AI 东风

我们预计 FY24/25/26 营收为 44.8/74.3/87.5 万亿韩元，对应同比为 114.7%/65.9%/17.8%。我们认为，HBM 作为 AI 芯片“标配”，GPU+HBM 模式提升芯片 AI 训练和推理表现，ASIC AI 加速器亦能通过增加 HBM 内存有效提升性能，需求能见度高。我们预计 24-25 年 HBM 将维持 3.6%-5.4% 的供给缺口。海力士作为 HBM 龙头，借助 Advanced MR-MUF 键合技术，其 HBM3/3E 的良率爬坡和产线管理更为成熟，因其更早进入英伟达供应链并已积累验证经验，后续 12 层 HBM3E 和 12/16 层 HBM4 出样、验证和量产有望较竞争对手更加顺畅。同时，DDR5 等传统内存产品也能受惠于 AI 设备向端侧渗透，叠加科技巨头资本开支增加带动通用服务器需求的回暖。

NAND 业务：FY24-25 “供销两旺”，协同 Solidigm 面向大容量产品增长

我们预计 FY24/25/26 营收为 21.7/32.9/36.6 万亿韩元，对应同比为 123.1%/51.9%/11.1%。24 年 NAND 复苏有望，核心将是下游消费电子设备，特别是端侧 AI 场景的大容量存储要求带来需求新动能。在实现“更大容量+更高耐用+更低成本”的要求下，海力士 176/238 层 TLC NAND 协同子公司 Solidigm 的 QLC NAND 产品矩阵将受益于各场景需求。

风险提示：AI 落地不及预期；芯片需求不及预期；中美贸易摩擦加剧等。

经营预测指标与估值

会计年度	2022	2023	2024E	2025E	2026E
营业收入(韩元十亿)	44,622	32,766	68,699	109,491	126,407
+/-%	3.78	(26.57)	109.67	59.38	15.45
归属母公司净利润(韩元十亿)	2,230	(9,112)	11,612	25,992	32,000
+/-%	(76.78)	(508.71)	227.43	123.83	23.11
EPS(韩元, 最新摊薄)	3,240	(13,242)	16,875	37,771	46,502
ROE(%)	3.56	(15.61)	19.73	34.13	31.01
PE(倍)	72.22	(17.67)	13.87	6.20	5.03
PB(倍)	2.55	3.01	2.51	1.83	1.36
EV/EBITDA(倍)	9.80	33.64	5.55	3.44	2.48

资料来源：公司公告、华泰研究预测

正文目录

SK 海力士：“先发+技术”铸就 HBM 护城河，能否保持竞争优势	6
HBM：AI 芯片性能催生内存需求，海力士如何稳固龙头优势？	14
AI 芯片存算发展不同步，内存瓶颈催生 HBM 需求	14
不仅仅是英伟达，面向同业 GPU 和 ASIC 新蓝海，HBM 同样蓄势待发	15
HBM3/3E 扩产加速，但供需缺口短期或难以缓解	17
HBM 供需测算：24-25 年需求高速增长，供给缺口短期难以缓解	18
竞争格局：三大厂商瓜分 HBM3/3E 市场，各显神通角逐 HBM4	21
SK 海力士：超越三星的 HBM 领头羊，先发优势、领先制程和键合工艺稳坐头把交椅	23
美光：跳过 HBM3，1-β 制程 HBM3E 加速追赶	27
三星：HBM3 世代落后海力士，HBM4 一站式策略或有望追赶海力士市场份额	28
HBM4：三大厂商方案齐出，混合键合延迟导入对海力士影响几何？	29
DRAM：AI 驱动需求回暖，稳定 24-25 年营收基本盘	32
AI 拉动 PC、手机以及服务器出货量，撬动 DRAM 需求增长	33
PC 市场触底反弹，Windows 系统换代拉动换机潮	33
AI PC 崛起，渗透率未来可期，内存升级需求带动 DRAM 增长	34
AI 手机引领旗舰潮流，中端机种开始全面渗透，拉动 LPDDR5X 需求	36
AI 服务器需求强劲，通用服务器配置新平台引领换机潮	37
面向端侧 AI，海力士 GDDR7、LPDDR5X/5T 和 LPCAMM2 布局完善	39
加速 EUV 导入 1-β 制程，24-25 年加速先进制程爬坡	41
竞争格局：行业高度集中，三足鼎立技术咬合紧密，制程产品各有千秋	41
NAND：AI 如何惠及海力士大容量 NAND？	44
NAND ASP 23Q3 见底回升，24 年 ASP 增长趋平稳	45
HBM 和 DDR5 之外，AI 将如何惠及 NAND Flash？	46
协同 Solidigm，海力士 NAND 用户端和数据中心双管齐下	48
竞争格局：技术趋同，厂商份额差异主要鉴于产能规模	50
盈利预测与估值：HBM 龙头地位有望保持，目标价 30 万韩元	53
风险提示	58

图表目录

图表 1：SK 海力士 Forward PB 估值周期图	7
图表 2：海力士分业务盈利预测（单位：十亿韩元）	12
图表 3：公司分业务营收（CY，万亿韩元，%）	12
图表 4：公司各业务营收占比（CY，%）	12
图表 5：公司 DRAM 营收（CY，万亿韩元，%）	12
图表 6：公司 NAND 营收（CY，万亿韩元，%）	12
图表 7：海力士 DRAM ASP（CY，韩元/Gb，%）	13

图表 8: 海力士 NAND ASP (CY, 韩元/GB, %)	13
图表 9: 公司库存情况 (CY, 十亿韩元, %)	13
图表 10: 公司盈利能力 (CY, %)	13
图表 11: 内存带宽需求伴随模型复杂度快速提升 (GB/s)	14
图表 12: AI 模型复杂度提升和 AI 芯片内存容量提升并不同步	14
图表 13: 主流加速卡算力增长和内存方案/传输带宽增长并不同步	14
图表 14: GPU 内存上限带来“内存墙”问题, 限制模型的复杂度上限	14
图表 15: 主流 AI 芯片参数及内存配置: 伴随芯片迭代, 算力提升、内存扩容和带宽拓展是大势所趋, HBM 成为 AI 芯片标配	16
图表 16: 三大厂商 2024-2025 年 HBM 晶圆投入量	18
图表 17: 根据计算, 单片 12 英寸 Wafer 可切割 HBM3/3E DRAM 约 442 (= 491 x 0.9) 颗	19
图表 18: HBM 各产品产能分布预测 (%)	19
图表 19: 三大厂商 2024-2025 年 HBM 晶圆投入量	19
图表 20: 2022-2025 年 HBM 出货量 (mn GB)	20
图表 21: 2022-2025 年 HBM 市场规模 (十亿美元)	20
图表 22: 2024-2025 HBM 供应端产能预测 (K GB)	21
图表 23: 海力士 2024-2025 年 HBM 营收及占比	21
图表 24: HBM 主要生产商技术路线图 2022-2026	22
图表 25: 三大厂商 HBM3/3E 的性能对比表	22
图表 26: DDR、GDDR 和 HBM 对比: HBM 通过堆叠实现内存容量和带宽倍增	22
图表 27: HBM 进程时间表	23
图表 28: SK 海力士 6 代 HBM 参数对比	24
图表 29: MR-MUF 对比 TC-NCF 温度下降 14%	25
图表 30: Advanced MR-MUF 对比 MR-MUF 散热能力进一步提高	25
图表 31: SK 海力士以 MR-MUF 代替 TC-NCF, 大幅降低 HBM 热阻	25
图表 32: 对比第五代产品 HBM3E, 第六代产品 HBM4 有望实现存储器和逻辑芯片的垂直堆叠	26
图表 33: AMD V-Cache 原理 - 将 L3 缓存置于 CPU 顶部, 两侧硅片用于芯片加固和辅助散热	26
图表 34: 美光 HBM3E 产品展示	27
图表 35: 美光 HBM3E 规格参数	27
图表 36: 美光 HMC 设计概念和性能参数: 与 HBM 高度相仿	27
图表 37: 三星历代 HBM 参数对比	29
图表 38: TC (热压) 键合工艺原理示意	29
图表 39: 微凸点和混合键合横截面对比, 混合键合无焊料球	30
图表 40: 混合键合相较微凸点的 I/O 密度提升	30
图表 41: 混合键合的工艺流程: 电镀-CMP 及清洗-离子活化-对齐并施压贴合-退火使铜接点膨胀并接合-铜离子相互扩散完成键合	30
图表 42: HBM 简要工艺流程及主要设备供应商梳理	30
图表 43: HBM 三大厂商 SWOT 分析	31
图表 44: DRAM 产业结构 (以 PC 为例): 海力士属于 DRAM“原厂”, 即绝大部分 DRAM 颗粒的来源	32
图表 45: 23Q4-24Q1 DRAM 合约价见较大涨幅, 24Q2 DRAM 合约价环比涨幅或逾 20%	33

图表 46: 三大厂商 24 年 DRAM 产能规划	33
图表 47: 全球 PC 出货量: 23H2 以来回温 (百万, %)	34
图表 48: 23 年主流 ODM/OEM 库存基本出清, 24Q1 库存水位略高于疫情前水平	34
图表 49: 各 PC 厂商 AI PC 上市节奏	35
图表 50: 英特尔 AI PC 产品规划参数对比	35
图表 51: 全球智能手机出货量: 23H2 以来回温 (百万, %)	36
图表 52: 三星 Galaxy S24 Ultra 对比 S23 Ultra	36
图表 53: 2019-2024 年全球服务器出货量 (单位: 百万, %)	37
图表 54: 2020-2027 年全球服务器市场规模 (单位: 百万美元, %)	37
图表 55: 四大科技巨头近十年的资本开支 (百万美元, %)	38
图表 56: 2022-2028 DRAM 出货量变化 (以模组和世代区分)	38
图表 57: Intel 和 AMD 采用 DDR5 的服务器 CPU	38
图表 58: GTC 2024 上海力士展出的 GDDR7 样品	39
图表 59: HKMG 实现晶体管栅氧化层厚度减少, 功率降低	40
图表 60: SK 海力士最新 4 款 LPDDR 参数对比	40
图表 61: LPCAMM 模组相较于 SODIMM 减少设备内部空间占用	40
图表 62: 海力士 LPCAMM2 模组结构	41
图表 63: DRAM 市场竞争格局演变	42
图表 64: DRAM 主要玩家的制程和产品路径 (2022 年 6 月发布), 三大厂商 DRAM 产品迭代咬合紧密	43
图表 65: DRAM 三大厂商 SWOT 分析	43
图表 66: 海力士 PS1010/PS1030 eSSD: 176 层 TLC NAND, 支持 PCIe Gen 5	45
图表 67: 海力士 GTC 2024 展出 Platinum P51 SSD: 238 层 TLC NAND, 支持 PCIe Gen 5	45
图表 68: 3D NAND 推出以来, 密度每年以稳定的速度提高 30%	46
图表 69: SLC 向 QLC 等转变提升每单元存储位数, 但牺牲了可靠性	46
图表 70: 22 年底至 23 年半导体厂商 NAND 相继减产, 24Q1 开始逐步回调产能	46
图表 71: 24Q2 NAND 合约价环比涨幅将收敛至 13%-18%	46
图表 72: 加速器举例: eSSD 将存储的数据传输至 CPU 和 DRAM 进行“中转”, 再经由 PCIe 总线至逻辑芯片	47
图表 73: PCIe 迭代带来传输速率的翻倍提升	47
图表 74: 数据中心 eSSD 的性能关注点: 高容量和高耐用性	47
图表 75: UFS 迭代带来传输速率的翻倍提升	47
图表 76: SK 海力士 NAND 研发时间图 (单位: 层)	48
图表 77: SK 海力士 5 代 4D NAND 参数对比	48
图表 78: 海力士 eSSD 代表产品	49
图表 79: 海力士面向用户端的 PC801 和 BC901, 基于 176 层 TLC NAND	49
图表 80: 海力士 UFS 4.0 和 UFS 3.1 参数 (172 层 TLC NAND)	49
图表 81: 海力士 ZUFS 4.0 样品图	49
图表 82: Solidigm 面向 AI 场景的 NAND 矩阵, D5 系列主打高容量、低成本	50
图表 83: NAND 闪存市场竞争格局演变	50
图表 84: NAND 主要生产厂商技术路线图 2014-2024Q1, 3D NAND 层高增长迅速, 头部厂商技术迭代基本同步	51
图表 85: NAND 三大厂商 SWOT 分析	52

图表 86: 十年 PB 对比: 纳斯达克指数和韩国 KOSDAQ 指数, 纳指 P/B 估值享有溢价	53
图表 87: SK 海力士、美光 and 三星历史 ROE (单季度, %)	54
图表 88: SK 海力士、美光 and 三星估值 (截至 2024 年 6 月 21 日)	54
图表 89: SK 海力士、美光 and 三星历史 PB	56
图表 90: SK 海力士、美光 and 三星历史 PS	56
图表 91: 近十年费城半导体指数 (SOX) 和半导体厂商存货 DIO 基本呈相反方向变动	56
图表 92: 海力士分业务盈利预测 (单位: 十亿韩元)	57
图表 93: 全球 DRAM 营收及 YoY (CY, 百万美元)	57
图表 94: 全球 NAND 营收及 YoY (CY, 百万美元)	57
图表 95: SK 海力士 PE-Bands	57
图表 96: SK 海力士 PB-Bands	57

SK 海力士：“先发+技术”铸就 HBM 护城河，能否保持竞争优势

区别于市场观点：我们认为，AI 催动 HBM 需求高增，面对三足鼎立的 HBM 市场，SK 海力士 24-25 年龙头地位有望继续保持，推动估值进一步提升：1) 面向 AI 浪潮催生的内存和带宽瓶颈，“GPU+HBM”及“自研 ASIC+HBM”等设计百花齐放，海力士 HBM3 和 HBM3E 客户基础稳固，良率提升和扩产顺利，其营收占比提升带动利润率优化，引领估值进一步提升；2) 存储行业步入新一轮上行周期，传统 DRAM 和 NAND 的 ASP 回稳及下游，特别是 AI 端侧需求回暖，海力士营收“基本盘”锦上添花。我们认为市场低估了 AI 引领下存储行业（AI 及传统）上行周期的持续性，对 HBM 如何帮助 AI 芯片突破内存和带宽瓶颈，以及对海力士 HBM 龙头地位的保持和受益存在认知差。

市场普遍认知到 DRAM 及 NAND 两种传统存储产品在 24 年将出现量价修复，并认为 PC 出货量将伴随 AI PC 崛起、下游库存出清以及新一轮换机潮回暖，同时手机市场在 24 年也开始复苏。但我们认为，市场忽略了本轮以 AI 驱动的上行周期的持续性与 18 年（云计算数据中心搭建和虚拟货币潮）、21 年（疫情期间在家办公）两轮周期不同，并低估了内存在 AI 计算里起到的重要作用。另外，市场也普遍认为海力士作为 HBM 龙头，市场份额和键合技术均领先于三星和美光，但面向美光 and 三星，海力士在 HBM 市场争夺中的优势能否延续仍存疑虑。而我们认为，AI 芯片的迅速迭代，牵动 HBM 搭载量和技术要求随之提高，24-25 年仍将维持供不应求；而相较美光 and 三星，海力士除却背靠英伟达供应链带来的稳固 HBM 需求，其先发优势和成熟的产能和产线管理、HBM3 世代已有的 8 至 12 层爬坡经验和采用 Advanced MR-MUF 键合技术，带来的高良率和散热性能帮助其 HBM 取得竞争优势，是海力士 HBM 龙头地位维持的核心所在。伴随 M15X 等新 DRAM 生产基地落地，其 HBM 扩产后有望拓展客户至其他 AI 芯片，需求能见度高，进一步提升营收和利润率。

高存储密度和高带宽的特性让 HBM 成为了 AI 场景下内存容量和带宽瓶颈的解决方案，减少算力浪费，在训练和推理场景扮演不同角色：1) AI 训练的大容量“数据中转站”：训练数据需从 SSD 等外部存储器分批次（batch）加载进 HBM，等待加速器调用，同时，训练过程中的模型参数、正向传播的中间状态（activation）、优化器状态以及反向传播的梯度（gradient）也需存入 HBM，方便加速器在计算和优化过程中快速访问和读写。HBM 以其大容量、高带宽的特性，减少加速器访问外部存储器次数，降低数据传输延迟，从而提升 AI 训练效率；2) AI 推理的“模型参数仓库”：训练后的数据本质上是一大堆的参数，HBM 可容纳更为复杂的模型，方便加速器推理时调用，减少数据搬运时间，提升 AI 推理效率。作为 AI 芯片的核心组件之一，我们认为单卡 HBM 堆栈数目增多、层数增高和迭代趋势已初见端倪。GTC 2024 上英伟达发布新一代架构 Blackwell 及 B200 GPU，通过搭载 8 颗 HBM3E 实现 192GB 内存容量和 8TB/s 的带宽，对比上一代 H200（6 颗 HBM3E，内存容量 141GB，带宽 4.8TB/s）在内存/带宽上提升 36.2%/66.67%。除去英伟达，AMD、英特尔和云厂商自研 ASIC 也呈现相同趋势：

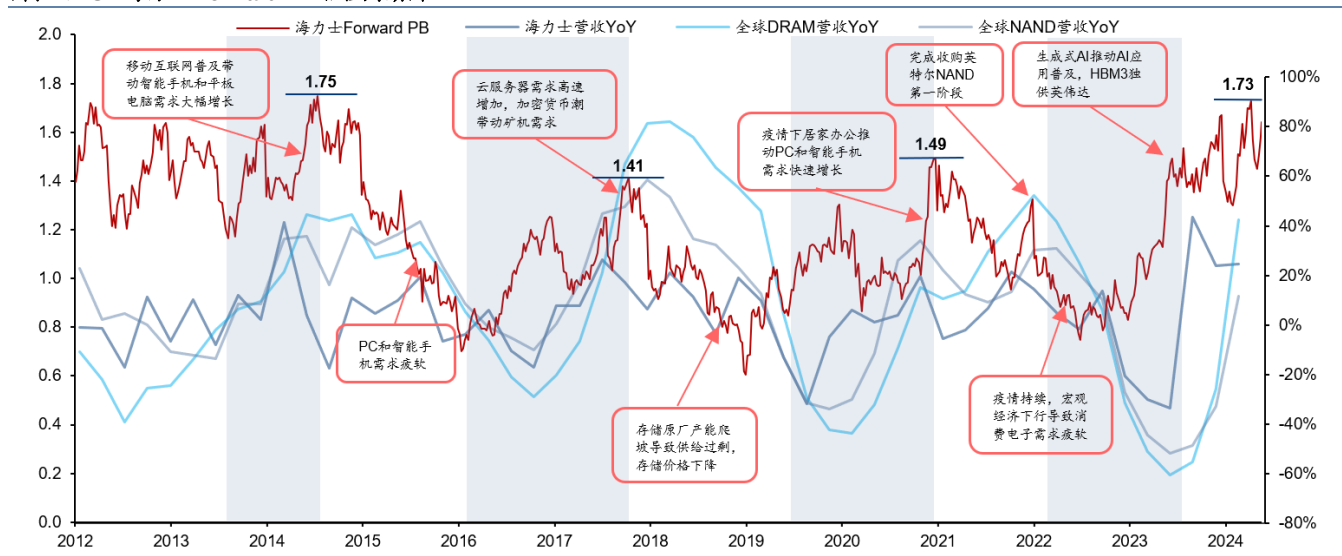
1) AMD: 2023 年 6 月 AMD 发布的 MI300X 搭载 8 颗 12 层 HBM3 实现 192GB 内存和 5.3TB/s 带宽，对比上一代 MI250X（8 颗 HBM2E，内存容量 128GB，带宽 3.2TB/s）在内存/带宽上提升 50%/65.6%，而预计 24H2 量产的 MI350 有望搭载 8 颗 12 层 HBM3E 实现 288GB 内存，对比 MI300X 再次提升 50%；

2) 谷歌: 2024 年 5 月 14 日谷歌 I/O 大会上推出的第六代 TPU Trillium 将 HBM 容量提升为上一代 TPU v5e 的两倍，使单芯片峰值算力对比 TPU v5e 提高 4.7 倍。2023 年 12 月 6 日谷歌公布 TPU v5p 性能参数，配置 95GB HBM3，对比 TPU v5e（16GB HBM2/2E）在 GPT3-175B 模型上训练速度提升近 50%。而作为谷歌 TPU 供应商的博通已于 24Q1 开始对三大厂商的 8 层 HBM3E 样品进行测试；

3) AWS: AWS 于年度大会 re:Invent 2023 公开的 Trainium 2 芯片搭载 4 颗 12 层 24GB HBM3，对比上一代 Trainium 32GB HBM2E 容量提升 200%；

4) 英特尔：2024 年 4 月 9 日英特尔发布 Gaudi 3，搭载 8 颗 16GB HBM2E，实现 3.7TB/s 带宽，对比上一代 Gaudi 2（6 颗 16GB HBM2E）内存容量提升 33%。

图1：SK 海力士 Forward PB 估值周期图



资料来源：公司公告、Bloomberg、TrendForce、Business Quant、华泰研究

截至2024年6月21日海力士估值为1.9x Forward PB，超过上一轮上行周期（21年，1.5x Forward PB）估值峰值。14-22年间海力士估值曾经历三次峰值，分别受益于：移动互联网普及（14年，1.8x Forward PB, 2.5x TTM PB）、上云趋势带来的服务器需求和加密货币潮（17-18年，1.4x Forward PB, 2.1x TTM PB）以及疫情居家办公带来的短期需求激增（21年，1.5x Forward PB, 1.9x TTM PB）：

（1）2014 年：12-14 年移动互联网普及带动智能手机和平板电脑销售量快速增长，根据 Gartner 2015 年数据，11-14 年全球智能手机销售量同比增长均超 28%，CAGR 38.2%。下游终端需求带动全球半导体销售额上行，根据 WSTS 历史数据及 DRAmExchange 11-15 年每季度统计，本轮上行周期历时约 33 个月，全球半导体销售额自 12Q1 谷底波动上升至 14Q3 峰值，12-14 年期间全球半导体销售额涨幅 15%，其中 DRAM 和 NAND 销售额涨幅均超 70%。FY12-14 期间海力士营收增长 68.5%，其中 DRAM 营收领涨 82.5%，NAND 营收增长 33.7%，彼时海力士刚被 SK 集团收购（2012 年）以摆脱资金困境，主要产能集中于 DRAM 业务，14 年海力士 DRAM 营收占比逾 80%，而 NAND 业务增长有限，主要供货苹果智能设备。

（2）2017-2018 年：16 下半年到 18 年，企业上云趋势带动服务器需求高速增长，叠加加密货币潮带动的矿机需求，全球半导体行业迎来新一轮上行周期，DRAM 及 NAND 产品领涨显著。根据 IDC 和 Gartner 16-18 年季度统计数据，16-18 年全球服务器出货量和 ASP 增长带动营收涨幅达 56%，18Q4 出货量近 350 万台，同比增长 8.5%。根据 Coinmarketcap 17 年 12 月 31 日数据，2017 年虚拟货币市场总值达 5724.8 亿美元，全年累计增长 3028%，比特币（BTC）价格上涨 13 倍，一度突破 20089 美元。以太坊（Ethereum）的数字代币以太币 2017 年涨幅达 111%，拉动 GPU 和存储的热度提升。根据 WSTS 历史数据及 DRAmExchange 16-18 年每季度统计，全球半导体销售额 16-18 年涨幅 38%，其中 DRAM 和 NAND 销售额涨幅分别为 145%和 63%，而本轮上行周期历时约 27 个月。FY16-18 海力士增长 49.5%，FY16-18 期间营收增长 135.2%，其中 DRAM 营收领涨 160.2%，NAND 营收增长 70.5%，主要系两年间下游需求使得 DRAM 市场供应紧张，ASP 持续保持高位。

(3) 2021 年：2020 年开始新冠疫情的居家办公场景带来短期消费电子产品（智能手机、PC 等）需求的快速增长。根据 Gartner 2019-2023 年数据，2021 年全球智能手机销售量同比增长 6.1%，为 19-23 年间唯一实现同比增长的年份，而全球 PC 出货量同比增长 20.2%，为 19-23 年间唯一实现两位数增长的年份。根据 WSTS 历史数据及 DRAmEXchange 20-22 年统计，全球半导体销售额自 20Q3 起实现约 20 个月的连续环比增长，期间涨幅 34%，21 年全球 DRAM/NAND 销售额同比增长 42%/21%。FY21 海力士营收同比增长 34.8%，其中 DRAM 营收同比增长 33.3%，NAND 营收同比增长 40.3%。21 年海力士收购英特尔 NAND 业务份额赶超美光，成为仅次于三星电子的 NAND 制造商。

(4) 2023-2024 年：疫情带来的短期需求于 22 年下半年见顶，22 下半年至 23Q3 期间消费电子需求疲软，半导体市场处于下行周期。根据 Gartner 2023 年数据，2023 年全球智能手机出货量同比下降 3.3%，全球 PC 出货量同比下降 13.5%。而 23 年至今海力士估值逆势高歌猛进，从 23 年 1 月 1 日的 0.8x Forward PB 跃升至 24 年 6 月 21 日的 1.9x Forward PB，主要系海力士 HBM3 于 22 年 6 月率先实现量产后独供英伟达 H100，并于 23 年 4 月和 6 月先后发布 12 层 HBM3 和 8 层 HBM3E，坐稳 HBM 头把交椅，背靠高速增长的英伟达 GPU 需求通过高价值 HBM 实现运营利润的强势扭亏，23 年海力士 HBM3 销量同比增长 5 倍。海力士 FY23 营收为 32.8 万亿韩元，同比-27%，超彭博一致预期（31.9 万亿韩元）；23Q4 营收 11.3 万亿韩元，同比+47%，环比+25%，超彭博一致预期（10.4 万亿韩元）。23Q4 运营利润扭亏，达 3,460 亿韩元，超彭博一致预期的亏损 1,699 亿韩元。

我们认为 24-25 年 HBM 仍将是海力士估值持续突破的关键，伴随 AI 芯片的持续迭代，不仅英伟达对 HBM3E 需求增加，AMD 和云厂商自研芯片对 HBM3 和 HBM3E 的需求也不可忽视，海力士作为龙头，将持续受益于高速增长的 HBM 市场，24-25 年维持 50% 左右的市场份额有望；同时，24 年内存行业迈入新一轮上行周期，由 AI 浪潮驱动，并将受益于 AI 设备和 AI 应用普及所带来的中长期增长，跟 14 年的逻辑类似，而并非如同加密货币潮和疫情带来的短期需求激增，因此我们认为海力士借助 AI 东风，估值有望持续突破。我们预计 FY24/25/26 公司营收分别为 68.7/109.5/126.4 万亿韩元，同比为 109.7%/59.4%/15.4%。具体而言：

DRAM 业务：我们预计 DRAM 业务 FY24/25/26 营收为 44.8/74.3/87.5 万亿韩元，对应同比为 114.7%/65.9%/17.8%，主要由 HBM 的量价齐升带动，叠加 AI 手机、AI PC 和 AI 服务器的 DRAM 需求。

1) HBM 受惠于 AI 芯片需求和设计进化，海力士作为龙头，HBM3 和 HBM3E 背靠英伟达供应链，借助更高良率和更好散热性能的 Advanced MR-MUF 键合技术稳固市场份额：我们认为，随着 AI 芯片需求的增加和在设计上的进化，更多的 HBM 将更紧密的搭载在 AI 芯片周边，在 AI 训练和推理时增加内存和降低延时。

通过测算存储厂商 HBM 产线晶圆投入量和台积电先进封装产量，我们认为 24/25 年 HBM3/3E 供需缺口约为 5.4%/3.6%。目前，HBM3/3E 市场已打破“独供”，由 SK 海力士、三星和美光三家割据：SK 海力士原为英伟达 HBM3 独供，率先敲定英伟达 HBM3E 商单，24 年 3 月 19 日宣布开始 8 层 HBM3E 量产，24 年 5 月公司宣布计划 5 月提供 12 层 HBM3E 样品，并将于 24Q3 正式量产，24 年 6 月 COMPUTEX 上海力士展出 12 层 HBM3E 样品。美光并非 HBM3/3E 的先行者，公司跳过 HBM3 直接研发 HBM3E，23 年 7 月，美光宣布与台积电 3D-Fabric 联盟，与海力士并列推出 1- β 制程的 HBM3E，并于 24 年 2 月率先宣布实现量产，确认供货英伟达 H200 并于 24 年 3 月 22 日确认出样 12 层 36GB HBM3E，预期 25 年量产。三星于 23H2 宣布量产 HBM3 Icebolt（8/12 层堆叠）并确认供货 AMD MI300 系列，随后三星加速产品迭代并积极拓展客户，GTC 2024 期间展示其 12 层 36GB HBM3E，英伟达 CEO Jensen Huang 会上于三星 12 层 HBM3E 旁签下“Jensen Approved”，确认该产品正在通过英伟达效能验证，时至 24 年 6 月 COMPUTEX，三星 HBM3 和 HBM3E 尚未进入英伟达供应链，但 HBM3E 24Q2 实现量产计划仍然推进，保持供货 AMD。

海力士作为 HBM 龙头,对比同业,其 HBM3 和 HBM3E 的良率爬坡和产线管理更为成熟,同时,因其更早进入英伟达供应链并已积累产品验证经验,后续 12 层 HBM3E 和 12/16 层 HBM4 出样、验证和量产有望较竞争对手更加顺畅。根据 FY24Q1 业绩会和 24 年 5 月于海力士利川园区举办的记者会,海力士管理层对 HBM 前景乐观,预计 24-25 年 HBM 不会出现“供过于求”,强调海力士 HBM 已确保稳固市占率,导入 EUV 采用 1-b 制程,新增出货也在扩展至英伟达之外的客户,当前 HBM3E 产能按计划提升,24 年 HBM 产能已售罄,25 年也几近售罄,16-24 年 HBM 累计营收将落在 130 亿-170 亿美元之间。

我们认为尽管先前“HBM3 独供”格局已被打破,但通过供应英伟达 H200 和 B 系列,海力士 HBM3E 客户仍然基础稳固,叠加其 Advanced MR-MUF 键合技术带来的更高良率和更好的散热性能,将成为海力士维持 HBM 龙头的“护城河”。若能按照计划在 FY24-25 实现扩产和并进一步扩展至下游客户,同时顺利在 25 年推进 HBM4 量产,则将进一步受惠于高速增长的 HBM 市场,并凭借 HBM 产品更高的 ASP 带来利润率的持续优化。

2)传统 DRAM 供应商库存已降低,以 DDR5 为代表的 AI 相关内存需求则受惠于 AI 手机、AI PC 和 AI 服务器渗透率提升带来的商机:根据 Digitimes 24 年 5 月 14 日报道,HBM 的扩产对 DDR5 产能形成挤兑,每生产 1 位元 (bit) 的 HBM,就有 3 位元 (bit) 的 DDR5 无法生产,DDR5 在 24 年初仍处于供不应求,24Q1 约有 3% 的供给缺口,伴随存储厂商陆续恢复增产,预期 24Q2 起缺口收敛至 1%,24H2 产能将持续开出,但供应端增产到实际出货增加仍需 15 个月左右,25 年 DDR5 整体供给仍将维持紧俏,从而维持 DDR5 ASP 涨势。相较而言,DDR4 和 DDR3 等传统 DDR 方面,DDR4 是 23 年上游原厂减产的重点,23Q3 DDR4 仍处于亏损出货状态,尽管下游库存水平逐渐恢复健康,短期内成熟制程产品(如中低端服务器和消费电子产品)需求增长并不强劲,故海力士等主要供应商仍不倾向恢复产能,叠加 HBM 和 DDR5 需求增长产生 DRAM 产能排挤,整体将对 DDR4 和 DDR3 等传统 DRAM 保持积极的产能控制。

24Q2,短期内受 4 月 3 日中国台湾强震对供应链影响,DRAM ASP 或见较大涨幅。DRAM 三大厂商之一的美光约有 65% 的 DRAM 在台湾生产,根据 Digitimes 24 年 4 月 15 日至 17 日系列报道,该次地震对美光厂房和基础设施未造成永久影响,单季 DRAM 供应影响约为 4%-6%,鉴于 DRAM 厂商原先已有 24Q2 DRAM ASP 调升 10-15% 规划,叠加地震影响,24Q2 美光 DRAM 合约价将调涨 25%,海力士和三星跟随将 DRAM 合约价调涨约 20%。

分应用场景来看,PC 方面,23 年 PC 库存已基本出清,顺应新 CPU 机种逐渐转往 DDR5 的趋势,下游 DDR5 采购量或将继续上升。叠加微软将于 25 年 10 月 14 日开始停止支持 Windows 10 所带动,以及 AI PC 的崛起,PC DRAM 市场有望进一步迎来复苏。根据 IDC 24 年 2 月 7 日预测,AI PC 24 年出货量约为 50 万台,2027 年则有望突破 1670 万台,3 年 CAGR 约为 222%。整体而言,PC 24Q1 出货量已基本恢复至 19Q1 水平,24H1 PC 市况已经逐步好转,虽然需求大规模增长尚未出现,但 24H2 PC 市场可望迎来进一步复苏。

AI 服务器方面,我们认为,硬件设备的规模和性能是 AI 大模型时代的必然要求,CPU+加速器异构带动 DDR5 和 HBM 需求高速增长。鉴于目前生成式 AI 主要以大参数模型路径实行,随着模型数量和所需处理的数据量增长,其训练与推理均需大量的计算能力与存储资源。展望 24 年,AI 基础设施将是重点投入领域,头部云厂商和互联网巨头加大 AI 领域资本开支,将进一步支撑 AI 服务器行情。IDC 23 年 12 月发布数据显示,23 年全球 AI 服务器市场规模达 211 亿美元,预计 2025 年达 317.9 亿美元,2023-2025 年 CAGR 22.7%。

同时，AI 服务器自身需要处理一系列密切相关但要求各异的任务，包括数据预处理、模型训练、推理和后处理等，因此需要 CPU+加速器异构以灵活面对使用场景。AI 服务器搭载的 CPU 负责系统启动和维护以及系统各组件统筹协调等逻辑，在 AI 训练和推理中亦涉及数据预处理等通用计算场景。例如，AI 训练和推理的原始数据往往参差不齐，需要通过数据清洗、转换、特征工程等一系列流程，最终才能为系统所用，而这些任务涉及大量逻辑运算和内存操作，对处理速度和时延要求较高，因此通常由 CPU 承担，需要更大带宽、更大容量的内存作为支撑，一台 AI 服务器的 DRAM 容量是通用服务器的 6-8 倍，因此快速增长的 AI 服务器需求将带动 DRAM 需求的高速增长，推动 DDR4 向 DDR5 迭代；而如前文所述，HBM 在 AI 训练和推理中同样扮演重要角色，AI 服务器的加速器搭配 HBM 已成标配，逻辑芯片频繁读取片外 HBM 进行内存调度。

通用服务器方面，24 年或迎来换机潮，产品组合将向 DDR5 倾斜。根据 Digitimes 24 年 2 月 1 日报道，业界预期 24 上半年服务器需求“淡季不淡”，北美云厂商除了持续采购 AI 服务器，也开始对通用型服务器进行新一轮采购，加速英特尔 Eagle Stream 和 AMD Geona 等新平台导入，其中以谷歌和微软采购力度较大。根据 24 年 2 月 Yole Intelligence 预测，未来五年内，DDR4 份额将进一步被 DDR5 取代。到 2027 年，DDR5 将占 DRAM 总出货量的 80% 以上。

AI 手机方面，Galaxy S24 带动 AI 手机热潮方兴未艾，端侧 AI 功能强化对 LPDDR5 需求。AI 手机相较传统非 AI 旗舰机型多出 50%-100% 的 DRAM 需求，而鉴于手机功耗限制，因此带动 LPDDR5X 需求，IDC 24 年 2 月预测 24 年手机出货量同比增长 3-5%，其中全球 AI 手机出货量或增至 1.7 亿台，占手机市场出货比重达 15%。

面向 DDR5 商机，海力士于 4 月台积电 2024 年技术研讨会和 5 月戴尔全球科技峰会上展示 LPDDR5T 和基于 LPDDR5X 的 LPCAMM2 内存模组。LPCAMM2 相较 DDR5 DIMM 在带宽和能效表现上更具优势，能更有效解决 AI PC 和 AI 手机的内存问题。海力士 LPDDR5T 已于 2023 年 10 月通过高通骁龙 8 Gen 3 和联发科天玑 9300 AP 的性能认证，并在 2023 年 11 月开始量产 16GB 版本。该产品已经交付给 vivo，并配置于其智能手机 X100 和 X100 Pro 中。我们预计公司的 16/24Gb DDR5、LPDDR5T 和 LPCAMM2 作为面向端侧 AI 的内存方案将显著受益。

NAND 业务：我们预计 NAND 业务 FY24/25/26 营收为 21.7/32.9/36.6 万亿韩元，对应营收同比为 123.1%/51.9%/11.1%。

1) 伴随下游客户库存水平正常化，海力士对 FY24 NAND 和 eSSD 趋势较乐观。根据 Bloomberg Intelligence 22-24 年数据统计，NAND ASP 下降 12 个月后至 23Q2 触底，于 23Q3 开始回升。而根据 TrendForce 2024 年 3 月报道，供应商正在推动提高 NAND Flash 的合同价格，24Q1 涨幅约为 23-28%，Q2 NAND Flash 采购量将较 Q1 小幅下滑，但整体市场氛围持续受供应商库存降低，以及减产效应影响，预估 24Q2 NAND Flash 合约价将上涨约 13-18%。结合公司 FY24Q1 业绩会，我们预计海力士 FY24 NAND 整体供给策略维持保守，着力于进一步优化库存，缩短周转天数，关注伴随 AI 设备渗透率提高带来的高层 QLC NAND 和大容量 eSSD 增长机会。

2) 布局数据中心 eSSD、AI PC 和大容量智能手机，Solidigm 补足 QLC NAND 产品矩阵：Digitimes 24 年 4 月 26 日报道提及，高层 QLC NAND 在 PC 中渗透率不断提高，AI PC 搭载 LLM 和图片/影片生成模型，将耗费大量存储容量，同时，由于用户不会立即删除生成的大量图片和视频，带来 AI PC 所需存储空间显著提升，同时出于降低存储成本考量，QLC NAND 导入 PC 成为趋势。同时，业界预期 24 年 QLC NAND 将导入大容量存储的手机新品，随着影像存储需求增加及 QLC NAND 更低的单位存储成本，iPhone 及部分国产手机品牌（如 Oppo）已进入大容量 1TB 机种采用 QLC NAND 的产品验证阶段。



截至 24H1，海力士主流 NAND 技术节点为 176 层 TLC NAND (4D V7 TLC)，基于此，海力士通过 PE9000 系列和 PC801/BC901 分别面向数据中心和 AI PC 场景，并通过 UFS 4.0 布局大容量手机。23 年 6 月海力士宣布开始量产 238 层 4D NAND，但截至 24H1 尚未大规模投片，公司计划首先向智能手机供应该款 NAND，随后扩展至基于 PCIe 5.0 的 PC SSD 和数据中心级 eSSD 等产品。24 年 6 月 COMPUTEX 上海力士展示 PCIe 5.0 Platinum P51 SSD，基于 238 层 3D TLC NAND，各项效能约为前一代 P41 的 2 倍，提供 13500 MB/s 的循序读取速度及 11500 MB/s 的循序写入速度，预期 24H2 开始在韩国销售，并于 25 年向全球发售。

同时，海力士子公司 Solidigm 专注于 QLC NAND 产品线，23Q4 QLC NAND 出货量占其 NAND 总出货的 60%。Solidigm 当前主流技术节点为 192 层 QLC NAND，以其 D5 系列 eSSD 面向数据中心场景，提供 3TB-60TB 的多样化存储方案，同时有 P44 Pro 系列（172 层 NAND）和 P41 Plus 系列（144 层 NAND）面向用户端。

FY24Q1 海力士 NAND 营收为 4.4 万亿韩元，同比+158%，环比+33%，主要系 NAND ASP 回暖，环比增长逾 30%。我们预计 24-25 年海力士和 Solidigm 产品矩阵分工合作，海力士 176 层和 238 层 NAND 先后布局 AI 手机、AI PC 和服务端，其中，AI 手机“DRAM+NAND”配合升级趋势和大容量机种陆续上架，有望带动海力士 ZUFS 4.0 和 UFS 4.0/3.1（176 层 NAND）需求增长。而 Solidigm 则借助大容量 QLC NAND 产品线消化 AI 设备兴起的“更低成本+更高密度”存储需求。

FY24Q1 海力士营收和利润超预期，高端产品带动净利率转正。海力士 FY24Q1 营收 12.4 万亿韩元，同比+144%，环比+10%，超彭博一致预期的 12.1 万亿韩元；Non-GAAP 净利润 1917 十亿韩元，超彭博一致预期的 1129 十亿韩元；Non-GAAP EPS 2788 韩元，超彭博一致预期的 1954 韩元。受益于 ASP 恢复和 HBM、DDR5 等高阶产品销量提升，海力士经调整毛利率 39%，超彭博一致预期的 30.5%，环比+19pct；经调整净利率 15%，实现转正，超彭博一致预期的 9.3%，环比+27pct。

公司预计 24Q2 DRAM 出货量环比中两位数增长，NAND 出货量环比持平。24Q1 海力士 DRAM 营收为 7.6 万亿韩元，同比+160%，环比+3%，出货量环比中两位数上升，ASP 环比上升逾 20%；NAND 营收为 4.4 万亿韩元，同比+158%，环比+33%。公司预计 24Q2 DRAM 出货量环比中两位数增长，NAND 出货量环比持平，同时预期传统设备需求将在 24H2 进一步恢复。

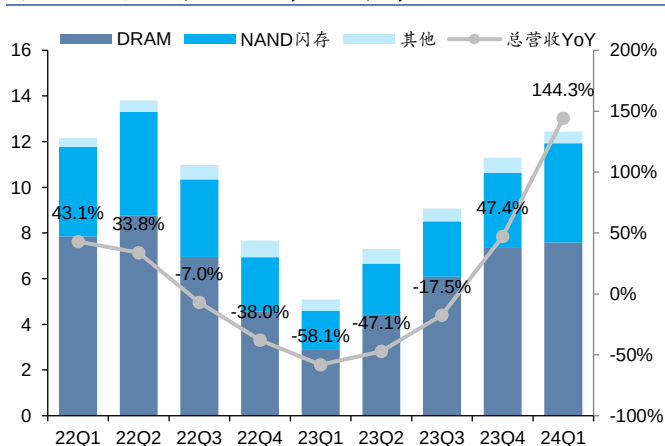
图表2：海力士分业务盈利预测（单位：十亿韩元）

Total Revenue	2019A	2020A	2021A	2022A	2023A	2024E	2025E	2026E
DRAM（十亿韩元）	20,829	23,136	30,836	28,521	20,849	44,765	74,278	87,531
YoY(%)		11.1%	33.3%	-7.5%	-26.9%	114.7%	65.9%	17.8%
位元总供给 (bn Gb)	37.8	46.1	55.4	56.3	64.2	77.4	99.2	112.6
YoY(%)		22.1%	20.1%	1.6%	14.1%	20.6%	28.2%	13.5%
ASP (WON/Gb)	551.6	501.8	556.8	507.0	324.9	578.5	748.8	777.4
YoY(%)		-9.0%	10.9%	-8.9%	-35.9%	78.0%	29.4%	3.8%
HBM（十亿韩元）	-	-	-	-	-	11,424.3	19,919.3	-
位元总供给 (mn GB)	-	-	-	-	-	705.2	1,229.6	-
ASP (WON/GB)	-	-	-	-	-	16,200.0	16,200.0	-
占 DRAM 营收比重	-	-	-	-	-	26%	27%	-
NAND（十亿韩元）	5,201	7,580	10,638	14,169	9,720	21,682	32,940	36,603
YoY(%)		45.7%	40.3%	33.2%	-31.4%	123.1%	51.9%	11.1%
位元总供给 (bn GB)	40.6	57.6	92.7	134.8	159.3	175.8	204.0	234.0
YoY(%)		41.9%	61.0%	45.4%	18.2%	10.3%	16.1%	14.7%
ASP (WON/GB)	128.2	131.6	114.7	105.1	61.0	123.3	161.4	156.4
YoY(%)		2.7%	-12.8%	-8.4%	-42.0%	102.2%	30.9%	-3.1%
Others（十亿韩元）	961	1,184	1,524	1,932	2,197	2,252	2,274	2,274
YoY(%)		23.2%	28.7%	26.8%	13.7%	2.5%	1.0%	0.0%
总收入（十亿韩元）	26,991	31,900	42,998	44,622	32,766	68,699	109,491	126,407
YoY(%)		18.2%	34.8%	3.8%	-26.6%	109.7%	59.4%	15.4%

注：其他收入主要为 NOR

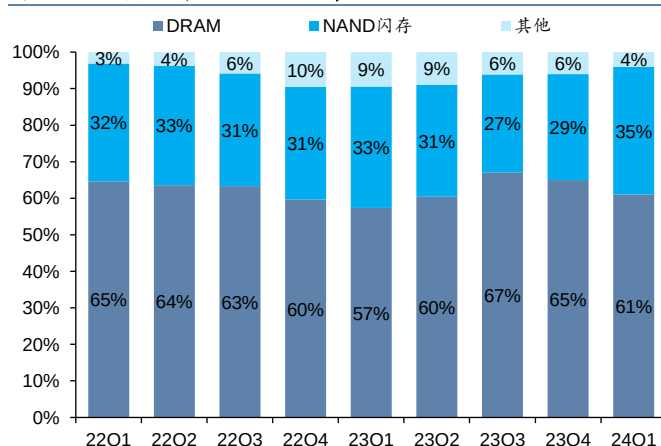
资料来源：公司公告、华泰研究预测

图表3：公司分业务营收（CY，万亿韩元，%）



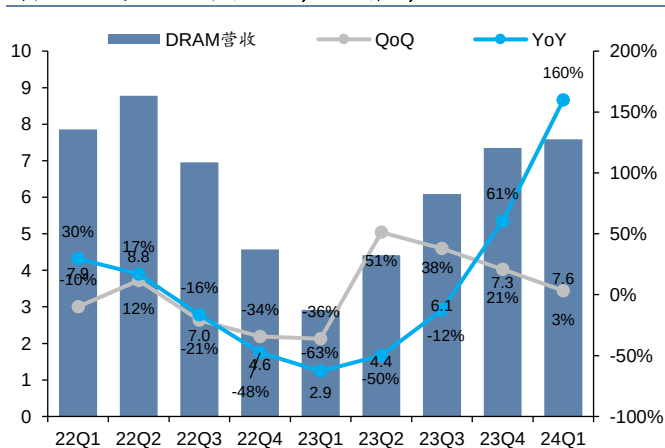
资料来源：公司公告、华泰研究

图表4：公司各业务营收占比（CY，%）



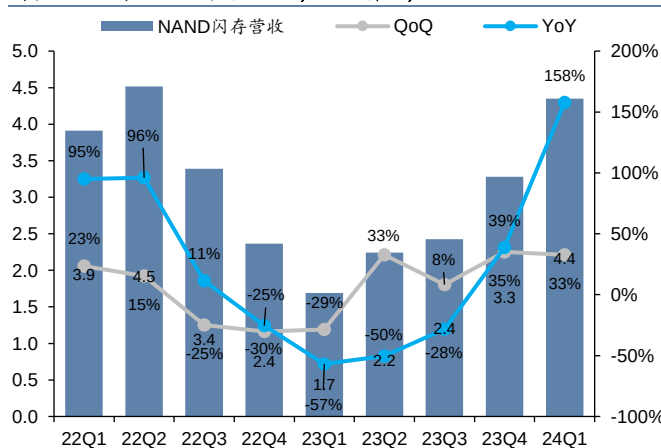
资料来源：公司公告、华泰研究

图表5：公司 DRAM 营收（CY，万亿韩元，%）



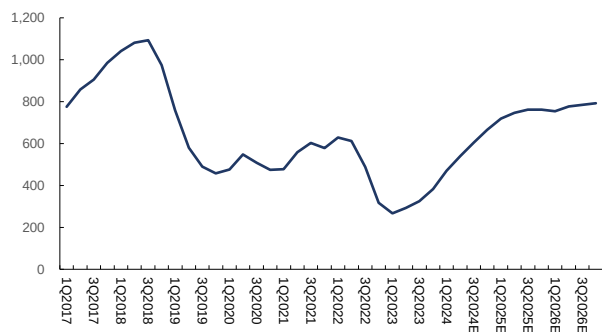
资料来源：公司公告、华泰研究

图表6：公司 NAND 营收（CY，万亿韩元，%）



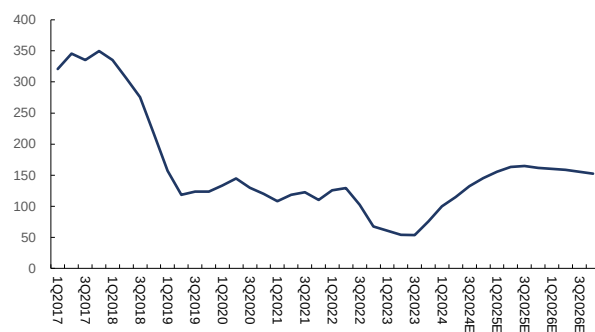
资料来源：公司公告、华泰研究

图表7：海力士 DRAM ASP (CY, 韩元/Gb, %)



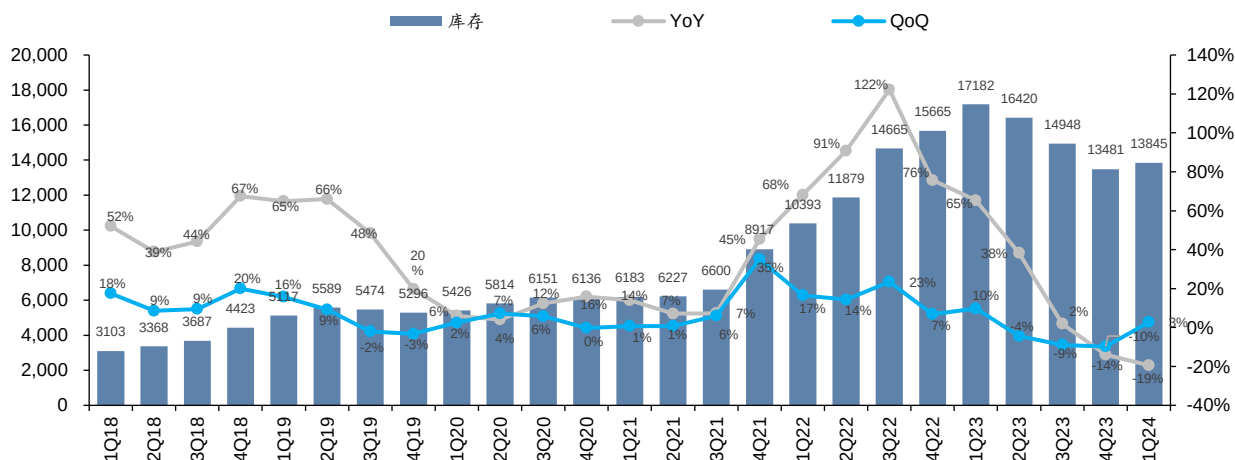
资料来源：公司公告、华泰研究预测

图表8：海力士 NAND ASP (CY, 韩元/GB, %)



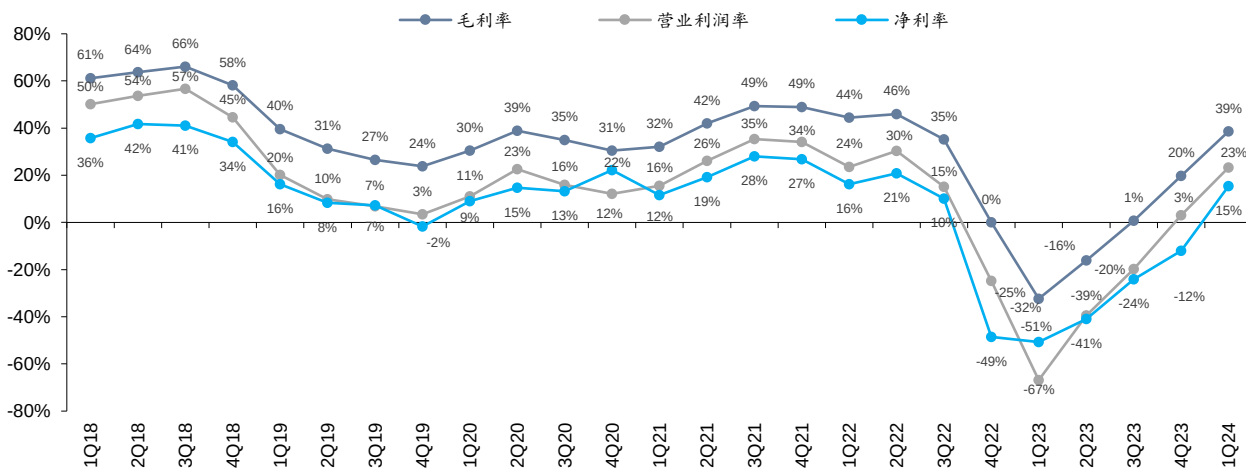
资料来源：公司公告、华泰研究预测

图表9：公司库存情况 (CY, 十亿韩元, %)



资料来源：公司公告、华泰研究

图表10：公司盈利能力 (CY, %)



资料来源：公司公告、华泰研究

HBM：AI 芯片性能催生内存需求，海力士如何稳固龙头优势？

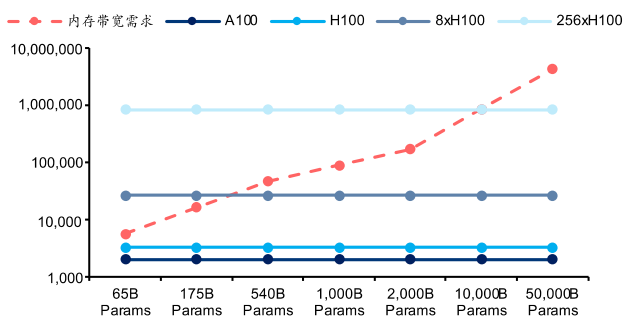
我们认为，在 AI 芯片需求强劲的背景下，伴随着英伟达和 AMD 积极的 AI GPU 迭代蓝图，以及头部云以及互联网厂商考虑削减 TCO、提升研发可控性及集成自身生态圈而积极推进自研芯片（ASIC 为主），HBM 的容量、效能和功耗要求将不断提升，HBM 的需求量将受 AI 芯片的需求增长拉动。我们预测，HBM 产品在 24-25 年将出现需求缺口，叠加 HBM 的复杂工艺需求（TSV 硅穿孔、键合堆叠和后续以 CoWoS 封装与 AI 芯片搭载一起），将带来相对传统 DRAM 更高的 ASP。伴随 HBM 产品在海力士 DRAM 中营收占比提升，我们预计公司营收和利润率将得到优化。

AI 芯片存算发展不同步，内存瓶颈催生 HBM 需求

我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中分析了在 AI 训练和 AI 推理场景下内存扮演的重要角色，指出 AI 芯片“存”和“算”的性能提升并不同步，计算能力与带宽能力之间的差距致使内存容量和带宽难以跟上 AI 硬件的计算速度，成为限制 AI 芯片性能发挥的主要瓶颈，即“内存墙”（Memory Wall）。AI 训练场景下 HBM 作为内存提供更高的带宽，提升存算间通信效率，实现单次更大批量（batch size）的数据传输，从而降低延迟，并提升算力利用率和训练效率。而 AI 推理场景下，加速器搭配大容量 HBM 方案便于训练后的模型整体置入内存，方便加速器推理时调用，减少设备间数据搬运时间，提升 AI 推理效率的同时亦可以借助高带宽满足更多高并发请求（如 Meta 的智能推荐系统）。

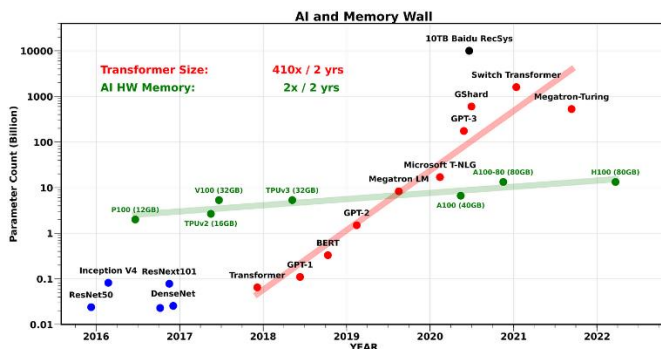
在算力接近的情形下，AI 芯片可以通过扩容内存和带宽提升模型训练和推理性能，而内存迭代对 AI 芯片性能的提升可以英伟达 H200 为例。23 年 11 月全球超算大会（SC23）上，英伟达推出的 H200 搭载 6 颗 HBM3E，内存达 141GB，带宽 4.8TB/S。作为 H100（搭载 6 颗 HBM3，内存达 80GB）的升级款，H200 依然采用 Hopper 架构和台积电 4 纳米工艺。两者的算力指标基本相同，但在 GPU 核心数和频率保持不变的情况下，仅通过内存从 HBM3 向 HBM3E 的迭代，就能实现在 Llama2 和 GPT-3 等大语言模型推理性能 40%-90% 的提升

图表11：内存带宽需求伴随模型复杂度快速提升（GB/s）



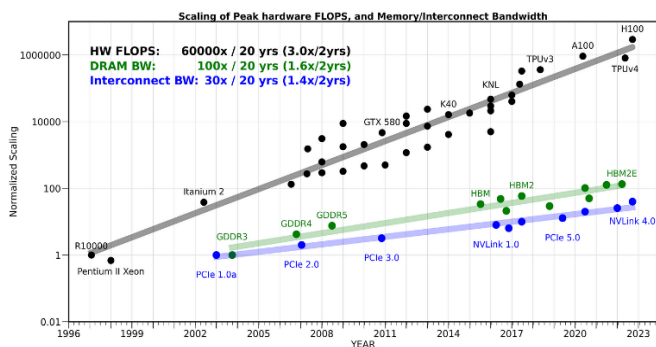
资料来源：Semianalysis、华泰研究

图表12：AI 模型复杂度提升和 AI 芯片内存容量提升并不同步



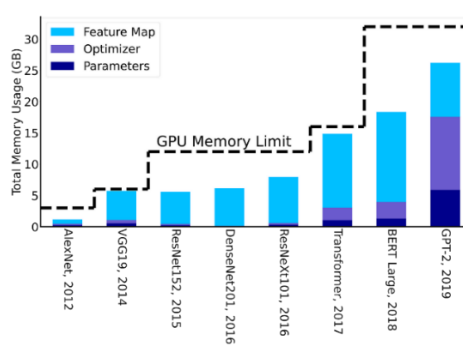
资料来源：Amir Gholami (2024.3):《AI and Memory Wall》、Github、华泰研究

图表13：主流加速卡算力增长和内存方案/传输带宽增长并不同步



资料来源：Amir Gholami (2024.3):《AI and Memory Wall》、Github、华泰研究

图表14：GPU 内存上限带来“内存墙”问题，限制模型的复杂度上限



资料来源：Amir Gholami (2024.3):《AI and Memory Wall》、Github、华泰研究

不仅仅是英伟达，面向同业 GPU 和 ASIC 新蓝海，HBM 同样蓄势待发

作为 AI 芯片的核心组件之一，我们认为单卡 HBM 堆栈数目增多、层数增高和迭代趋势已初见端倪。GTC 2024 上英伟达发布新一代架构 Blackwell 及 B200 GPU，通过搭载 8 颗 HBM3E 实现 192GB 内存容量和 8TB/s 的带宽，对比上一代 H200（6 颗 HBM3E，内存容量 141GB，带宽 4.8TB/s）在内存/带宽上提升 36.2%/66.67%。除去英伟达，AMD 和英特尔 GPU 也同样如此：1) AMD：2023 年 6 月 AMD 发布的 MI300X 搭载 8 颗 12 层 HBM3 实现 192GB 内存和 5.3TB/s 带宽，对比上一代 MI250X（8 颗 HBM2E，内存容量 128GB，带宽 3.2TB/s）在内存/带宽上提升 50%/65.6%，24 年 6 月 COMPUTEX 主题演讲上 AMD 发布 MI325X 和 MI350 系列，分别将于 24Q4 和 25 年上市，两款新品均将搭载 8 颗 12 层 HBM3E 实现 288GB 内存，对比 MI300X 再次提升 50%；2) 英特尔：2024 年 4 月 9 日英特尔发布 Gaudi 3，搭载 8 颗 16GB HBM2E，实现 3.7TB/s 带宽，对比上一代 Gaudi 2（6 颗 16GB HBM2E）内存容量提升 33%。

我们认为，AI 芯片市场竞争日趋白热化，除却 GPU 等通用型芯片，ASIC 专用芯片通过更低的 ASP、在特定场景提供更高效和更低功耗、以及更灵活的设计周期，将是通用型 GPU 的互补或替代。伴随 AI 模型多元化和算法日趋成熟，ASIC 作为“算法的物理呈现”，也将适配各大厂商的差异化模型需求，承接部分算力。头部云及互联网厂商考虑到削减 TCO、提升研发可控性及集成自身生态圈等，均在推进自研芯片。

ASIC 的火热亦可从博通 FY24Q2 业绩略见一斑。此前 24 年 3 月 21 日博通 Investor Day 上宣布除却谷歌 TPU 外新增两位超大规模（Hyperscale）客户，当日博通股价涨幅逾 5%。24 年 6 月 12 日博通于业绩会上表示来自客户的 AI 互联和 ASIC 产品需求强劲，FY24Q2 AI 相关营收同比增长 280%，达到 31 亿美元，上调全年 AI 相关营收预期至 110 亿以上。会上博通 CEO 陈福阳点明博通 ASIC 业务“有利可图”以回应先前 ASIC 芯片利润率较低的市场风声，同时说明博通 ASIC 与英伟达 GPU 并非竞争关系，与通用 GPU 不同，其 ASIC 通过多个 IP 组合帮助客户面向特定的 AI 工作场景，6 月 12 日起截至 6 月 17 日博通股价累计增长 24%。

相较以算力作为核心护城河的 GPU，ASIC 专用芯片并不以算力见长，但随着芯片设计的进化，ASIC 能通过增加 HBM 内存等有效提升性能并降低延时，从而弥补“算力差距”。根据谷歌 23 年 12 月发布的 TPUv5p 参数及英伟达 H100 参数，尽管对比 INT8 算力，TPUv5p 仅为 H100 SXM 的 23%（918TOPS vs. 3958TOPS），但特定场景下（如 GPT-3 175B 训练）二者已可基本对标。其中，H100 SXM 搭载 80GB HBM3，带宽 3.35TB/s；谷歌 TPUv5p 搭载 95GB HBM2e，带宽 2.8TB/s。另外，算法的进步也可有效降低对芯片算力需求，例如特斯拉 FSD 已采用 FP8，而不需采用传统的 FP16，而英伟达 Blackwell 系列 GPU 也已引入 FP4 和 FP6。

2024 年 5 月 14 日谷歌 I/O 大会上推出的第六代 TPU Trillium 将 HBM 容量提升为上一代 TPU v5e 的两倍，使单芯片峰值算力对比 TPU v5e 提高 4.7 倍。2023 年 12 月 6 日谷歌公布 TPU v5p 性能参数，配置 95GB HBM3，对比 TPU v5e（16GB HBM2/2E）在 GPT3-175B 模型上训练速度提升近 50%。而作为谷歌 TPU 供应商的博通已于 24Q1 开始对三大厂商的 8 层 HBM3E 样品进行测试，同时，AWS 于年度大会 re:Invent 2023 公开的 Trainium 2 芯片搭载 4 颗 12 层 24GB HBM3，对比上一代 Trainium 32GB HBM2E 容量提升 200%。

图表15：主流 AI 芯片参数及内存配置：伴随芯片迭代，算力提升、内存扩容和带宽拓展是大势所趋，HBM 成为 AI 芯片标配

产品名称	英伟达										
	A100 PCIe SXM	H100 PCIe SXM NVL	L40S	GH200 HBM3	GH200 HBM3e	H20	H200	B100	B200	Rubin	
芯片类型	GPU	GPU	GPU	GPU	GPU	GPU	GPU	GPU	GPU	GPU	
发布时间	2020.6	2022.9	2023.10	2023.5	2023.8	2024Q1	2023.11	2024.4	2024.3	2024.6	
峰值算力 (TFLOPS/TOPS)	FP16: 312 624	FP8: 3026 3958 7916	FP8: 1466	-	-	FP8: 296	INT8: 1979	FP8: 7000	FP8: 9000	TBC	
	FP32: 19.5	FP16: 1513 1979 3958	FP16: 733			FP16: 148	FP16: 989	FP16: 3500	FP16: 4500		
	FP64: 19.5	FP32: 51 67 134	FP32: 91.6			FP32: 44	FP32: 494	FP32: 1800	FP32: 2200		
		FP64: 51 64 134				FP64: 1	FP64: 34	FP64: 30	FP64: 40		
工艺制程	TSMC 7nm	TSMC 4nm	TSMC	TSMC 4nm	TSMC 4nm	-	TSMC 4nm	TSMC 4nm	TSMC 4nm	TSMC 3nm	
芯片面积	826mm2	814 mm2	-	-	-	-	814 mm2	-	-	-	
晶体管数 (B)	54	80	76.3	200	-	-	80	-	-	-	
内存容量 (GB)	80 GB (HBM2e)	80 80 188 (HBM3)	48GB (GDDR6)	96 (HBM3)	14 (HBM3e)	96 (HBM3)	141 (HBM3e)	192 (HBM3e)	192 (HBM3e)	HBM4	
内存带宽	1935 2039 GB/s	2 3.35 TB/s 7.8TB/s	864GB/s	<=4TB/s	4.9TB/s	4TB/s	4.8TB/s	8TB/s	8TB/s	TBC	
互联	NVLink	NVLink	PCIe Gen4	NVLink	NVLink	PCIe Gen5 128 GB/s	NVLink	PCIe Gen6 256 GB/s	PCIe Gen6 256 GB/s	NVLink	
	600 GB/s	600 900 600 GB/s	64GB/s	900GB/s	900GB/s	NVLink: 900 GB/s	900GB/s	NVLink:1.8 TB/s	NVLink:1.8 TB/s	3.6 TB/s	
TDP (W)	300 400	300-350 700 2x350-400	350	450 - 1000	450 - 1000	400	700	700	1000	TBC	
典型应用场景	AI训练	AI训练	AI训练及推理	AI训练	AI训练	AI训练	AI训练	AI训练	AI训练	AI训练	
产品名称	AMD				谷歌			英特尔			
	MI250X	MI300A	MI300X	MI325X	MI350系列	TPUv4	TPUv5e	TPUv5p	Gaudi 2	Gaudi 3	
芯片类型	GPU	GPU	GPU	GPU	GPU	ASIC	ASIC	ASIC	GPU	GPU	
发布时间	2021.11	2023.1	2023.6	2024.6	2024.6	2021.5	2023.12	2023.12	2022.5	2024.4	
峰值算力 (TFLOPS/TOPS)	INT8: 383 (TOPS)	FP8: 1961.2	FP8: 2614.9	FP8: 2.6 PF		INT8: 275	INT8: 394	INT8: 918	FP8: 865	FP8: 1835	
	FP16: 383	FP16: 980.6	FP16: 1307.4	FP16: 1.3 PF		BF16: 275	BF16: 197	BF16: 459	BF16: 432	FP16: 459	
	FP32/64: 47.9	BF16: 980.6	BF16: 1307.4		TBC					FP32: 229	
		TF32 (Matrix): 490.3	TF32 (Matrix): 653.7							BF16: 1835	
										TF32: 459	
工艺制程	TSMC 6nm	TSMC 5nm	TSMC 5nm	5nm*	3nm	TSMC 7nm	-	-	TSMC 7nm	TSMC 5nm	
芯片面积	724 mm2	1017 mm2	1017 mm2	TBC	TBC	780mm2	-	-	-	-	
晶体管数 (B)	58	146	153	TBC	TBC	31	-	-	-	-	
内存容量 (GB)	128 (HBM2e)	128 (HBM3)	192 (HBM3)	288 (HBM3E)	288 (HBM3E)	32 (HBM2)	16 (HBM2)	95 (HBM2e)	96 (HBM2e)	128 (HBM2e)	
内存带宽	3.3TB/s	5.3TB/s	5.2TB/s	6TB/s	TBC	1228GB/s	820GB/s	2765GB/s	2.45TB/s	3.7TB/s	
互联	Infinity Fabric	Infinity Fabric	Infinity Fabric	Infinity Fabric		3D Mesh	2D Torus	3D Torus	PCIe Gen4	PCIe Gen5	
	>=100GB/s	800GB/s	896GB/s	TBC	TBC	2400Gb/s	1600Gb/s	4800Gb/s	64GB/s	128GB/s	
TDP (W)	500	550-760	750	TBC	TBC	192	-	-	600	900-1200	
典型应用场景	AI训练	AI训练	AI训练	AI训练	AI训练	AI训练及推理	AI训练及推理	AI训练及推理	AI训练	AI训练	
产品名称	亚马逊			微软		Meta		壁仞		燧原科技	Xilinx
	Trainium	Inferentia 2	Trainium 2	Maia 100	MTIA v1	MTIA v2	BR100	BR104	云燧 T20	Versal (HBM)	
芯片类型	ASIC	ASIC	ASIC	ASIC	ASIC	ASIC	GPU	GPU	ASIC	FPGA	
发布时间	2020.12	2022.12	2023.11	2023.11	2023.5	2024.4	2022.8	2023.9	2021.7	2022.3	
峰值算力 (TFLOPS/TOPS)	INT8: 380	INT8: 380	INT8: 861	FP4: 3200	INT8: 102	INT8: 354	INT8: 2048	INT8: 1024	INT8: 256	INT8: 75	
	FP16: 190	FP16: 190	FP16: 431	INT8: 1600	FP16: 51.2	FP16: 177	BF16: 1024	512	FP16: 128		
	FP32: 47.5	FP32: 47.5	FP32: 215				TF32+: 512	TF32+: 256	FP32: 32		
	BF16: 190	BF16: 190	BF16: 431				FP32: 256	FP32: 128			
工艺制程	TSMC 7nm	TSMC 5nm	TSMC 4nm	TSMC 5nm	TSMC 7nm	TSMC 5nm	TSMC 7nm	TSMC 7nm	GF 12nm		
芯片面积	-	-	-	-	373 mm2	421 mm2	1074 mm2	-	-		
晶体管数 (B)	55	54	115	105	-	-	77	-	-		
内存容量 (GB)	32 (HBM2e)	32 (HBM2e)	96 (HBM3)	64 (HBM3)	64 (LPDDR5)	128 (LPDDR5)	64 (HBM2e)	32 (HBM2e)	32 (HBM2e)	32 (HBM2)	
内存带宽	820 GB/s	820 GB/s	4TB/s	1.6TB/s	176GB/s	205GB/s	1.6TB/s	819GB/s	1.6TB/s	819GB/s	
互联	NeuronLink-v2	NeuronLink-v2	NeuronLink-v2	Ethernet	PCIe Gen4	PCIe Gen5	Blink	Blink	PCIe Gen4	PCIe Gen5	
	768GB/s	384GB/s	2043GB/s	4.8TB/s	16GB/s	32GB/s	512GB/s	192GB/s	-	-	
TDP (W)	-	-	-	860	25	90	550	300	300		
典型应用场景	AI训练	AI推理	AI训练	AI训练及推理	AI训练及推理	AI训练及推理	AI训练	AI训练	AI训练	AI推理	

资料来源：各公司 AI 芯片公告、TechPowerUp、Trendforce、NextPlatform、Digitimes、华泰研究

HBM3/3E 扩产加速，但供需缺口短期或难以缓解

市场对 HBM3/3E 反响积极，伴随 AI 芯片迅速迭代，所需 HBM 搭载量和技术要求随之提高，在供需失衡下三大厂商纷纷扩产。据 Tom's Hardware 在 23 年 12 月报道，英伟达向海力士和美光提前支付数亿美元以锁定 HBM 产能，印证了 HBM 产能的迫切需求和对 AI 芯片的重要。英伟达在 GTC 2024 上发布了新一代架构 Blackwell 及 B200 GPU，通过搭载 8 颗 HBM3E 实现 192 GB 内存容量和 8 TB/s 带宽，对比上一代 H200（6 颗 HBM3E，内存容量 141 GB，带宽 4.8 TB/s）在内存/带宽上提升 36.2%/66.67%，而除却 B 系列芯片，英伟达 FY25Q1 业绩会上强调 Hopper 等上一代产品需求仍在增长。

此外，亚马逊 AWS 已计划采购 2 万片 GB200 芯片，以部署高达 27 万亿参数的模型。根据 Digitimes 24 年 3 月的报道，Meta 计划在 24 年底前储备约 35 万颗 H100，加上其采用的其他 AI 芯片，将拥有相当于 60 万颗 H100。Meta 在 3 月 19 日指出，目前 Meta 正在以 2 个 GPU 丛集训练其第三代 Llama 模型，据称每个丛集包含约 2.4 万个 H100 GPU，公司计划导入 Blackwell 训练 Llama 的未来版本。

23 年以来，海力士在 HBM 扩产上动作频频。继 23 年“独供”英伟达 H100 HBM3 之后，海力士积极推动产能向 HBM3E 迭代。公司预计将于利川工厂生产 HBM3E，清州 M15 工厂闲置空间新增 HBM 产线，预期于 24H1 量产 HBM3，同时清州正在兴建的 M15X 厂房也将作为 HBM 生产基地，产能预期于 25H1 开出。根据 FY24Q1 公司业绩会和 24 年 5 月举办的记者会，海力士表示当前 HBM3E 产能按计划提升，但 24 年 HBM 产能已售罄，25 年也几近售罄。公司指引 24 年资本开支高于年初计划，一方面用于投资 M15X 建设，另一方面因 HBM 晶圆消耗相较传统 DRAM 更大，叠加 24 年 DDR5 等传统 DRAM 受 AI 和换机潮带动需求高增，需要相应投资支持扩产。

根据 Digitimes 24 年 4 月 15 日、3 月 11 日和 2 月 7 日报道，23 年以来海力士已从韩美半导体订购逾 2000 亿韩元的 TC 键合设备，包括 Dual TC Bonder Griffin 及 Dual TC Bonder 1.0 Dragon，并同时与 ASM 子公司 ASMPT 讨论采购其 TC Bonder，24 年 6 月 14 日韩美宣布再获规模约 1500 亿韩元的海力士 Dual TC Bonder Griffin 追加订单，为单一产品订单合约最大规模的一次。同时 24 年海力士将于 HBM 投资超 10 亿美元，其中大多新投资将用于改善封装制程，包括 MR-MUF 键合工艺和 TSV 技术，从而提升其 HBM 产品散热能力和生产效率。考量新设备安装后扩大量产仍需时间，**海力士计划 24 年底实现将 HBM 产能增加 2 倍，从 23 年每月 4.5 万片晶圆提升至 24 年底每月 12-12.5 万片晶圆。**

我们也注意到 HBM 另两大厂商美光 and 三星亦在 HBM 扩产上有所动作：

根据美光 FY24Q2 业绩会，管理层对 HBM3E 前景乐观，并宣布 24 年产能已售罄，25 年产能也几近售罄。公司预计 FY24 CAPEX 在 75-80 亿美元，略高于 FY23 水平，主要用于支持 HBM3E 量产和扩产。管理层虽未透露除 CAPEX 之外更多的 HBM 扩产信息，但预期 25 年 HBM 市占率将与公司在 DRAM 市占率平齐，约为 20%。根据 Digitimes 24 年 3 月 12 日和 4 月 15 日报道，美光将逐季提高稼动率，增产重点将以 HBM 等高端 DRAM 为主，并与韩美半导体（Hanmi Semiconductor）签订 HBM TC 键合设备 Dual TC Bonder Tiger 的供应合约，订单规模 225.9 亿韩元，合约期限为 2024 年 4 月 10 日至 7 月 8 日。

根据 Digitimes 24 年 1 月和 Trendforce 24 年 3 月报道，24 年三星决定将 HBM 产能投资增加近 3 倍，从 23 年每月 4.5 万片晶圆提升至 24 年底每月 13 万片晶圆，25 年也将进行类似规模的投资。根据 Digitimes 23 年 12 月汇总信息，23H2 以来三星在设备采购和产线增加上动作积极：1) Theelec 12 月 5 日报道，三星以 5 亿韩元单价从 Shinkawa 订购 16 台键合设备，其中 7 台已交付，推测该批设备将服务于英伟达 GPU 相关的 HBM 和封装业务；2) 韩媒 ZD Net Korea 12 月 1 日报道，韩国设备供应商 YEST 收到三星 123 亿韩元 HBM 加压设备订单，发往三星天安生产线；3) 三星于 CES 2024 宣布计划在 2024 年将 HBM 年产能扩大 2.5 倍以上，并随后投资 105 亿韩元从子公司三星显示（SDC）处收购天安工厂的建筑和设施。根据韩国经济日报 24 年 3 月 28 日报道，三星管理层在 Memcon 2024 宣布上调 24 年 HBM 产能预期至 23 年的 2.9 倍。

HBM 供需测算：24-25 年需求高速增长，供给缺口短期难以缓解

综合 Trendforce、Semianalysis 和 Digitimes 等多方信息，我们对 2024 年 HBM 产能和需求做出预测，我们预计 24/25 年 HBM 需求动态缺口约为产能的 5.4%/3.6%。

供给侧关键假设：

- 产能假设：**以 HBM 与 TSV 产线晶圆投入量为计，我们预期截至 2024 年底海力士、三星和美光将分别投入 12-12.5 万片/月、13 万片/月和 2 万片/月。考虑后道设备交付时间较长，以及新设备安装后扩大量产仍需时间，我们预测 24 年三星、海力士和美光 HBM 产线平均月度晶圆投入量将达 8.5 万片/月、8.5 万片/月和 1.2 万片/月。鉴于海力士和美光 25 年 HBM 产能亦出现售罄迹象，结合前文三星 25 年 HBM 扩产投资计划，我们认为 25 年三大厂商扩产仍将持续，预测 25 年三星、海力士和美光 HBM 产线平均月度晶圆投入量将达 12.5 万片/月、11.5 万片/月和 4.3 万片/月。

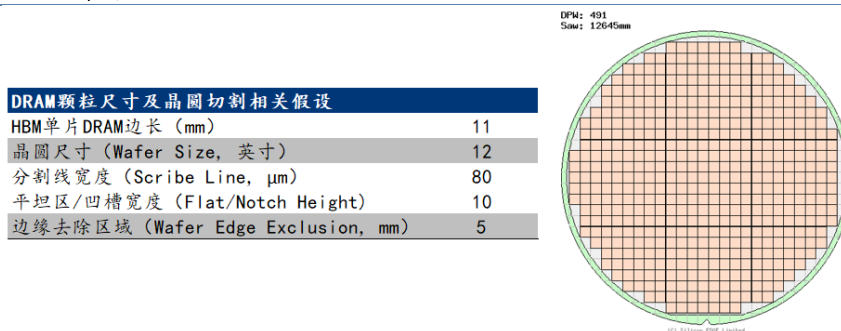
图表16：三大厂商 2024-2025 年 HBM 晶圆投入量

HBM 产线晶圆投入量	三星	海力士	美光
2023 年底（万片/月）	4.5	4.5	0.3
2024 年底（万片/月）	13	12-12.5	2
2024 年平均（万片/月）	8.5	8.5	1.2
2025 年平均（万片/月）	12.5	11.5	4.3

资料来源：Trendforce、华泰研究预测

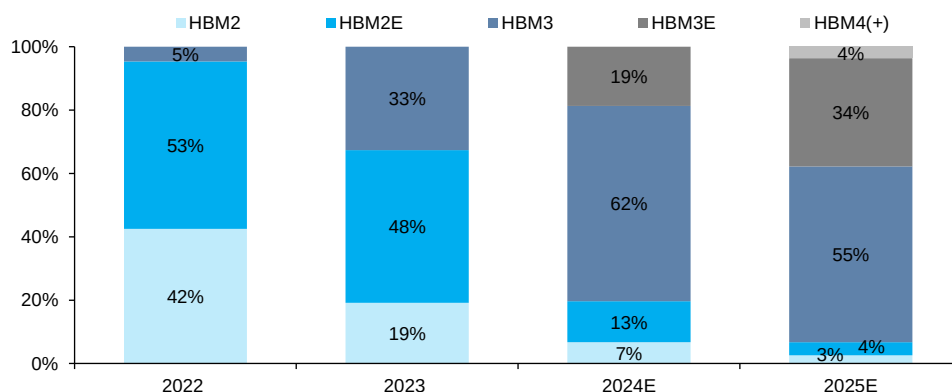
- 生产月份假设：**根据我们前文所述，三大厂商 HBM3E 量产及确认供货时间并不统一，美光 24 年 2 月率先宣布实现量产 HBM3E 并确认供货英伟达 H200；三星 HBM3 和 HBM3E 尚未进入英伟达供应链，由于产品存在发热与能耗方面的问题，24 年 4 月测试的 8 层和 12 层 HBM3E 测试未成功，但 HBM3E 24Q2 实现量产计划仍然推进，HBM3 和 HBM3E 保持供货 AMD；海力士保持英伟达 HBM3 供应同时，24 年 3 月 19 日宣布开始 8 层 HBM3E 量产，12 层 HBM3E 将于 24Q3 正式量产。因此，默认三大厂商均满产能生产 12 个月将带来产能虚高，我们引入“等效生产月份”表示 24 年三大厂商各自供应 HBM 实际时长，我们预测 24 年三星、海力士和美光 HBM 等效生产月份分别为 9/11/11 个月，25 年三家均为 12 个月。
- 切割良率假设：**根据美光官方公布 HBM3E 产品细节，HBM3E 单片 DRAM 尺寸约为 121 平方毫米(11mm x 11mm)，因此单片 12 英寸晶圆理论可切割 HBM3/3E DRAM 491 颗。同时，引用韩国经济日报 21 年 7 月报道，我们估计 DRAM 切割良率约为 90%，故单晶圆可切割 DRAM 442 颗。

此外，根据 Financial Times 24 年 5 月报道引用，海力士 HBM3E TSV+键合工艺良率已近 80%，而海力士采用的 Advanced MR-MUF 键合较三星和美光采用的 TC-NCF 键合在良率上有所优势（采用 TC-NCF 良率约在 40%-60%区间）。同时，台媒科技新报 24 年 3 月报道指出，供应商试图拉高良率会带来产量下降，鉴于 HBM 产能吃紧，因此确保产量优先级更高。结合上述信息，我们认为美光 and 三星 HBM3 量产后，伴随产能爬坡，工艺良率将有所提升并趋于稳定，但使用 TC-NCF 键合所以良率仍将低于海力士，预测 24 年海力士/三星/美光整体 HBM 良率 75%/40%/45%，25 年小有爬升，将是 77%/41%/47%。

图表17：根据计算，单片 12 英寸 Wafer 可切割 HBM3/3E DRAM 约 442 (= 491 x 0.9) 颗


资料来源：Silicon Edge、The Korea Economic Daily、美光官网、华泰研究估算

- 4) **单个 DRAM 容量假设：**伴随 HBM3/3E 产能不断开出，HBM 产品组合将向 HBM3/3E 倾斜，根据 Yole Intelligence 24 年 2 月报告，23 年 HBM3 产量占总产能的 33%，24 年 HBM3/3E 产量占比将跃升至 81%（HBM3：62%，HBM3E：19%），25 年将进一步提升至 89%。我们认为，海力士作为英伟达 HBM3/3E 主供，产能向 HBM3/3E 倾斜将更为激进；与之相比，三星凭借其“一站式策略”承接更多云厂商自研芯片的 HBM 需求，因此 HBM2/2E 仍占部分产能；美光计划通过 HBM3E 追赶市场份额，在产能本不富裕的情形下，扩产 HBM3E 并顺利交付将是首要发力点。因此，我们预测，**24 年海力士/三星/美光 HBM 的单片 DRAM 容量为 2.2/2.1/3.0GB，25 年伴随芯片迭代及 HBM3E 产能开出，其占比将持续扩大，带动 HBM 的单片 DRAM 容量提升至 2.6/2.4/3.0GB。**

图表18：HBM 各产品产能分布预测 (%)


资料来源：Yole Group、华泰研究预测

图表19：三大厂商 2024-2025 年 HBM 晶圆投入量

HBM 单片 DRAM 容量及出货量占比		
海力士	2024E	2025E
HBM3 - 2GB	77%	38%
HBM3E - 3GB	23%	62%
单片 DRAM 容量	2.2	2.6
三星	2024E	2025E
HBM2E - 1/2GB	20%	7%
HBM3 - 2GB	62%	35%
HBM3E - 3GB	19%	57%
单片 DRAM 容量	2.1	2.4
美光	2024E	2025E
HBM3E - 3GB	100%	100%
单片 DRAM 容量	3.0	3.0

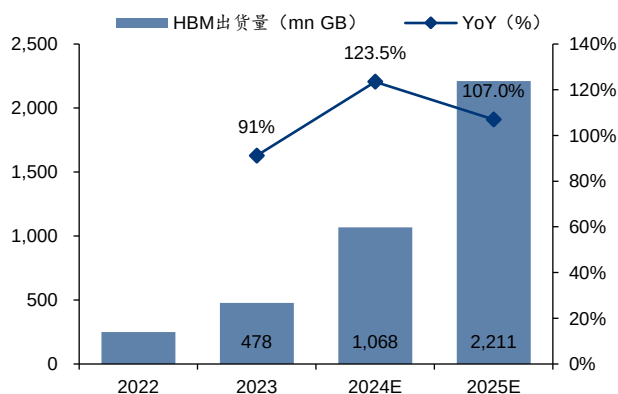
资料来源：Yole Intelligence、华泰研究预测

需求侧关键假设：

- 1) **CoWoS 产能假设：**结合台积电 23 年底 CoWoS 月产能为 1.9 万片，考虑产能爬坡时间，我们预测 24 年台积电动态 CoWoS 产能为 2.8 万片/月，与外溢部分合计 3.3 万片/月，24 年 CoWoS 总产能约 39.6 万片；25 年除去台积电 CoWoS 产能扩张，英特尔 EMIB、Amkor 和 ASE 等外溢产能同样有望进一步开出，预测 25 年 CoWoS 月产能 5.2 万片，总产能 62 万片。
- 2) **各厂商 AI 芯片单晶圆切出量、单卡 HBM 堆栈数量以及单 HBM 堆栈容量：**我们基于各 AI 芯片尺寸计算每片切出芯片数目，并根据 HBM 参数，结合各厂商取得的 CoWoS 产能，以芯片出货量为权重计算三项参数的加权平均。产能分布上，我们认为 24-25 年 HBM 需求仍以英伟达 AI GPU 为主体，英伟达 24-25 年仍将每年取得逾 40% 产能，同时关注到 AMD MI300X/MI325X/MI350、谷歌 TPU v5/Trillium、AWS Trainium 和 Inferentia 以及英特尔 Gaudi 2/3 带来的需求增量；单晶圆切出量上，伴随设计尺寸增大，以及 Blackwell 对比 Hopper 产能占比提升，25 年每片切出的 AI 芯片将随之下降；结合前文，虽然每片切出数目减少，但单卡 HBM 堆栈数目增多、层数增高和容量增大将是大势所趋。因此，我们预测 24/25 年单晶圆平均切出 AI 芯片 39/38 颗、单卡平均 HBM 堆栈数量 5.3/5.9 颗，以及单 HBM 堆栈平均容量 14.0/16.7GB。

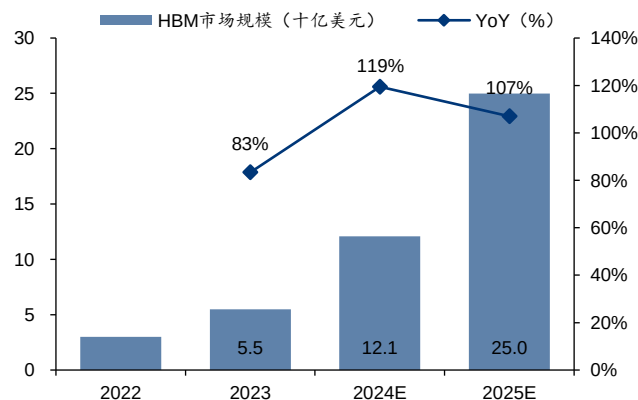
基于以上假设，我们预测 24/25 年 HBM 总供给 1068.1/2210.9 mn GB，同比 +123.5%/107.0%，其中 24 年海力士、三星和美光分别占总供给的 66%、27%和 7%，伴随美光 25 年 HBM 产能开出，海力士、三星和美光 HBM 供给量将分别占总供给的 56%、29%和 15%。同时，我们预测 24/25 年 HBM 总需求 1128.6/2293.0 mn GB，由此测算得 24/25 年 HBM 供给缺口约为产能的 5.4%/3.6%。鉴于三大厂商均提及与下游讨论 25 年 HBM 供应，我们认为 24 年 HBM 扩产后，25 年缺口虽收窄但仍将维持轻微的供给不足，从而使 HBM ASP 维持稳定。以 ASP 11 美元/GB 测算，对应市场空间 120.7/249.8 亿美元，同比 +119.5%/107.0%。

图表20：2022-2025 年 HBM 出货量 (mn GB)



资料来源：Yole Intelligence、华泰研究预测

图表21：2022-2025 年 HBM 市场规模 (十亿美元)



资料来源：Yole Intelligence、华泰研究预测

图表22：2024-2025 HBM 供应端产能预测（K GB）

HBM 供应量	2024E	2025E
SK 海力士		
晶圆投入量（万片/月）	8.5	11.5
单晶圆切割 HBM DRAM 颗粒数目（颗）	442	442
HBM TSV 及键合良率	75%	77%
单个 DRAM die 容量（GB）	2.2	2.6
等效生产月份	11	12
HBM 总产能（K GB）	705,202.8	1,229,583.4
占总产能比重（%）	66%	56%
三星		
晶圆投入量（万片/月）	8.5	12.5
单晶圆切割 HBM DRAM 颗粒数目（颗）	442	442
HBM TSV 及键合良率	40%	41%
单个 DRAM die 容量（GB）	2.1	2.4
等效生产月份	9	12
HBM 总产能（K GB）	285,381.7	659,754.1
占总产能比重（%）	27%	29%
美光		
晶圆投入量（万片/月）	1.2	4.3
单晶圆切割 HBM DRAM 颗粒数目（颗）	442	442
HBM TSV 及键合良率	45%	47%
单个 DRAM die 容量（GB）	3	3
等效生产月份	11	12
HBM 总产能（K GB）	77,541.2	321,581.5
占总产能比重（%）	7%	15%
合计（K GB）	1,068,125.7	2,210,919.0
HBM 需求（以 CoWoS 为计）	2024E	2025E
台积电月产能（万片/月）	3.2	4.4
台积电动态 CoWoS 月产能（万片/月）	2.8	4.2
外溢至 Amkor 等的产能（万片/月）	0.5	1.0
CoWoS 总产能（万片）	39	62
平均单晶圆切出 AI 芯片数（颗）	39	38
单卡平均 HBM 堆栈数量（颗）	5.3	5.9
HBM 堆栈平均容量（GB，综合考量 HBM3/3E 及 HBM2E）	14.0	16.7
HBM 总需求（K GB）	1,128,582.0	2,293,041.2
HBM 供需缺口	-5.4%	-3.6%

资料来源：Digitimes、Trendforce、Silicon Edge、各公司公告、华泰研究预测

图表23：海力士 2024-2025 年 HBM 营收及占比

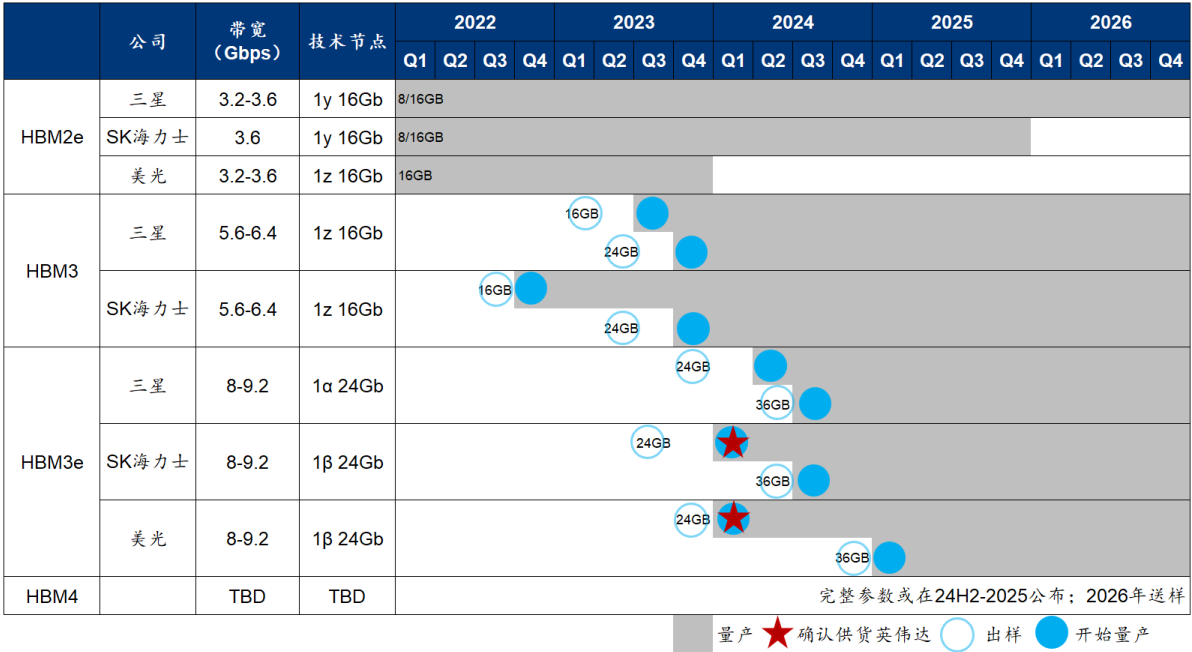
	2024	2025
海力士 HBM 营收（WON bn）	11,424.3	19,919.3
海力士 HBM ASP（WON/GB）	16,200.0	16,200.0
海力士 HBM 产能（K GB）	705,202.8	1,229,583.4
海力士 DRAM 营收（WON bn）	44,765.0	74,277.5
HBM 营收占比	26%	27%

资料来源：华泰研究预测

竞争格局：三大厂商瓜分 HBM3/3E 市场，各显神通角逐 HBM4

在半导体行业，先进制程技术的发展呈现资金和技术壁垒不断提高的趋势，HBM 所需工艺的复杂性和其应用场景，以及对供需双方深度合作的要求，使得行业格局集中于海力士、三星和美光三大厂商。综合 Yole Intelligence 24 年 2 月发布的全球 HBM 市场报告，23 年海力士市场份额为 56%，三星市场份额为 41%，美光市场份额为 3%。

图表24：HBM 主要生产商技术路线图 2022-2026



资料来源：Trendforce、Digitimes、华泰研究

图表25：三大厂商 HBM3/3E 的性能对比表

公司	美光	SK 海力士		三星	
HBM	HBM3E	HBM3	HBM3E	HBM3 Icebolt	HBM3E Shinebolt
量产时间	2024 年 2 月	2022 年 6 月	2024 年 3 月	2023 年下半年	2024 年上半年
容量 (GB)	24/36	16/24	24/36	24	24/36
带宽	1.2TB/s	819GB/s	1.18TB/s	819GB/s	1.2TB/s
单片 DRAM 容量	3GB	2GB	3GB	2GB	3GB
VDD 电压 (V)	-	1.1	1.1	1.1	-
I/O 速率 (Gbps)	9.2	6.4	9.2	6.4	9.8
键合高度 (层)	8/12	8/12	8/12	12	12
技术节点	1β	1z	1β	1z	1α
有效位宽	1024-bit	1024-bit	1024-bit	1024-bit	1024-bit
键合技术	TC-NCF	Advanced MR-MUF	Advanced MR-MUF	TC-NCF	TC-NCF

资料来源：美光官网、SK 海力士官网、三星官网、AnandTech、Digitimes、TrendForce、华泰研究

图表26：DDR、GDDR 和 HBM 对比：HBM 通过堆叠实现内存容量和带宽倍增

DDR 及 HBM 型号	推出时间	带宽	内存容量	典型应用场景
DDR4	2014	25.6 GB/s	2 GB	个人电脑和服务
DDR5	2021	51.2 GB/s	2/3 GB	个人电脑和服务
GDDR6	2018	64 GB/s	2 GB	个人电脑、游戏和 AI
GDDR6X	2020	96 GB/s	2 GB	个人电脑、游戏和 AI
HBM	2013	128 GB/s	2 GB	个人电脑、游戏和 AI
HBM2	2016	307 GB/s	4/8 GB	AI GPU
HBM2E	2020	460 GB/s	4/16 GB	AI GPU
HBM3	2022	819 GB/s	16/24 GB	AI GPU
HBM3E	2023	1.28 TB/s	24/36 GB	AI GPU
HBM4*	2025-2026	1.5 TB/s	64 GB	AI GPU

*注：HBM4 信息综合于美光、三星和海力士 HBM4 路线图

资料来源：美光官网、三星官网、海力士官网、英伟达官网、华泰研究

图表27：HBM 进程时间表

公司	时间	产品	事件
海力士	2013.12	HBM1	与 AMD 联合开发第一代 HBM，和 AMD Fiji GPU 一起应用于 AMD Radeon Fury 系列显卡。
	2017	HBM2	在三星之后推出 HBM2，同年实现量产
	2019.8	HBM2E	率先推出 HBM2E，MR-MUF 键合技术开始应用
	2020.7	HBM2E	HBM2E 实现量产
	2021.10	HBM3	推出业界首款 HBM3，带宽 819GB/s
	2022.6	HBM3	宣布 HBM3 开始供应英伟达
	2023.4	HBM3	推出 24GB 12 层堆叠的 HBM3，并宣布 HBM3 开始使用 Advanced MR-MUF 键合技术
	2023.8	HBM3E	开始出样 8 层 HBM3E，带宽突破至 1.15TB/s
	2024.1	HBM3E	CES 2024 上展示 8 层 24GB HBM3E，带宽 1.18TB/s
	2024.3	HBM3E	宣布 8 层 HBM3E 开始量产，供货英伟达，同时向英伟达送样 12 层 36GB HBM3E
	2024.5	HBM3E	宣布 HBM3E 键合良率已近 80%，并公布 HBM4 和 HBM4E 技术路线图
	2024.6	HBM3E	COMPUTEX 上展出 12 层 HBM3E 样品
	2024Q3	HBM3E	量产 12 层 HBM3E
三星	2016.1	HBM2	与英伟达和台积电联合开发 HBM2，同年实现量产
	2020	HBM2E	推出 HBM2E，同年实现量产
	2023H1	HBM3	先后推出 16GB 和 24GB HBM3
	2023.10	HBM3/3E	确认量产 8 层和 12 层 HBM3 Icebolt，推出 HBM3E Shinebolt 8 层原型版本，带宽 1.225TB/s
	2024.2	HBM3E	宣布 12 层 36GB HBM3E Shinebolt 开发完成，带宽 1.225TB/s
	2024Q1	HBM3	HBM3 通过 AMD 效能验证，供货 MI300 系列
	2024.5	HBM3E	HBM3E 传出英伟达验证因发热和功耗问题受阻，后于 6 月 COMPUTEX 期间被 Jensen 否认，但尚未进入英伟达供应链
	2024Q2	HBM3E	量产 8 层 HBM3E，供货 AMD
美光	2020	HBM2/2E	先后宣布量产 HBM2/2E
	2023.7	HBM3E	宣布与台积电合作推出 HBM3E
	2024.2	HBM3E	8 层 24GB HBM3E 率先量产，供货英伟达 H200
	2024.3	HBM3E	开始出样 12 层 36GB HBM3E
	2024.6	HBM3E	韩媒 ET News 24 年 6 月 3 日报道美光 12 层 HBM3E 功耗表现优异。

资料来源：海力士官网、三星官网、美光官网、Tom's Hardware、TechPowerUp、Digitimes、KED Global、华泰研究

SK 海力士：超越三星的 HBM 领头羊，先发优势、领先制程和键合工艺稳坐头把交椅

我们认为尽管先前“HBM3 独供”格局已被打破，但通过供应英伟达 H200 和 B 系列，海力士 HBM3E 客户仍然基础稳固，叠加其 Advanced MR-MUF 键合带来的更高良率和更好的散热性能，将成为海力士维持 HBM 龙头的“护城河”。若能按照计划在 FY24-25 实现扩产并进一步扩展至下游客户，同时顺利在 25 年推进 HBM4 量产，则将进一步受惠于高速增长的 HBM 市场，并凭借 HBM 产品更高的 ASP 带来利润率的持续优化。海力士管理层于 FY24Q1 业绩会和 24 年 5 月记者会指出，24 年客户需求集中于 8 层 HBM3E，该产线海力士已确保供应稳定性，同时，鉴于 HBM 销售以 1 年为单位捆绑合约供应，从当前合约内容看来，25 年 HBM 预计不会供过于求，从而维持 ASP 稳定。

相较美光和三星，海力士已积累与下游客户合作的丰富经验，如我们前文所述，其效能验证和纠错效率或较竞争对手更高；相较三星，海力士 HBM3 和 HBM3E 已形成成熟供应链，良率稳定且技术较三星更有优势。

SK 海力士的第一代 HBM 与 AMD 联合开发，之后公司多次发布新产品，保持 HBM 领先地位。2013 年 12 月，公司与 AMD 联合开发 HBM1，和 AMD Fiji GPU 一起应用于 AMD Radeon Fury 系列显卡。之后的 7 年间迭代更新 HBM2 和 HBM2E；于 2021 年 10 月推出全球首款 HBM3，2022 年 6 月实现量产后独供英伟达 H100。2023 年 4 月推出 24GB 12 层堆叠的 HBM3，通过将单个 DRAM 的高度磨削到约 30 微米（比 16GB HBM3 单片薄 40%），实现了与 16GB 产品相同的高度。2023 年 8 月发布 HBM3E。海力士在 CES 2024 展示的 HBM3E 提供 1.18TB/s 数据处理速度（发布初为 1.15TB/s），并已经于 24 年 3 月实现 HBM3E 量产，开始向英伟达供货。根据 TechInsights 转载 24 年 2 月国际固态电路研讨会（ISSCC 2024）日程，海力士于会上 Paper 13.4 中公开其 16 层堆叠的 HBM3E 技术，提供 48GB 内存和近 1.3TB/s 的带宽。根据 Digitimes 24 年 2 月 26 日报道，业界估计 24 年海力士 HBM 营收将达 10 兆韩元（约 74.9 亿美元），相当于 23 年 DRAM 总营收的一半。

如今海力士在全球 HBM 市场中份额约五成，和 AI 芯片巨头保持长期合作关系，背靠英伟达供应链的牢固客户基础构筑护城河。据 Business Korea 2023 年 12 月 1 日报道，英伟达计划将新品发布周期从两年缩短到一年，以应对生产式 AI 快速发展对 GPU 性能提升的需求，并已与 SK 海力士签署 HBM3E 的优先供应协议，用于今年较晚推出的 B100。而根据 Digitimes 24 年 2 月 27 日消息并结合 GTC 2024 海力士新闻，海力士于 24 年 1 月中旬正式结束 HBM3E 研发，历时半年的英伟达效能评估顺利结束，3 月获得英伟达的最终产品质量认证，且立即着手量产供应。根据 Digitimes 24 年 2 月报道英伟达已开始与海力士讨论 25 年 HBM 供应量，以确保供应稳定。

公司亦将 24GB HBM3 送样给包括 AMD 在内的其他客户。AMD 在 2023 年 12 月 6 日举行的 Advancing AI 发布会上正式发布搭载 24GB HBM3 的 MI300 系列。根据 Digitimes 23 年 12 月报道，海力士亦为云厂商自研 AI 芯片 AWS Trainium 2 和 Google TPU 的 HBM 供货。24 年 5 月海力士记者会表示，24 年所有 HBM3/3E 产能已卖光，并持续接到额外订单的咨询，25 年产能也几近售罄。

24 年 5 月于海力士利川园区举行的记者会上，海力士公开新的 HBM 发展蓝图。根据计划，海力士将于 24 年 5 月提供 12 层 HBM3E 样品，并于 24Q3 正式量产，第一代 12 层 HBM4 计划于 25 年量产，并于 26 年提供 16 层 HBM4 样品，相较先前计划提前一年。在键合技术方面，海力士计划 12 层 HBM4 仍维持使用 Advanced MR-MUF，16 层堆叠相关技术将发生变化，同时积极投入混合键合 (Hybrid bonding) 技术研发。截至 24 年 6 月 COMPUTEX，海力士 12 层 HBM3E 已经出样，并在会上进行展示。

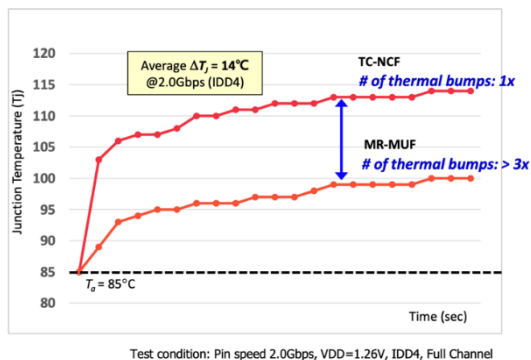
图表28：SK 海力士 6 代 HBM 参数对比

	HBM1	HBM2	HBM2E	HBM3	HBM3E	HBM4 (预测)
发布年份	2013	2017	2019	2021	2023	2025-2026
容量 (GB)	1	4/8	8/16	16/24	24/36	36-48
带宽 (GB/s)	128	307	460	819	1.18TB/s	1.5-2TB/s
VDD 电压 (V)	1.2	1.2	1.2	1.1	1.1	1.1
键合高度 (层)	4	4/8	4/8	8/12	8/12	12/16
键合技术	TC-NCF	TC-NCF	MR-MUF	(Advanced) MR-MUF	Advanced MR-MUF	Advanced MR-MUF Hybrid Bonding
热阻系数	-	1	0.65	0.5/0.55	0.5	0.4-0.5

资料来源：SK 海力士公告、华泰研究

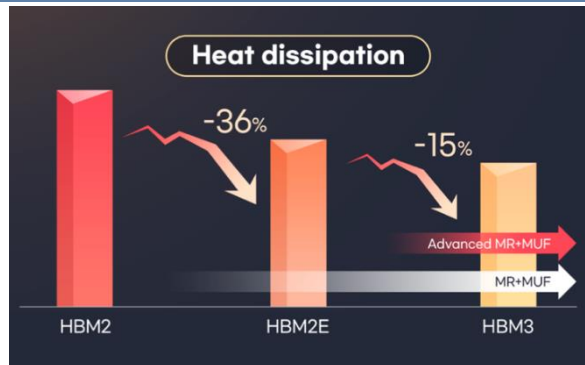
在 HBM3 和 HBM3E 世代，相较三星和美光，海力士采用领先的 HBM 键合工艺，是业内唯一将 MR-MUF 技术应用于 HBM 的公司。海力士的 HBM2 和三星 HBM 键合均采用 TC-NCF（基于热压的非导电薄膜 Thermal Compression - Non Conductive Film）工艺，该过程需要高温高压环境将凸点 (bumps) 推入非导电薄膜，在单个 DRAM 高度减少的环境下更易导致芯片翘曲。海力士在 HBM2E 中首次使用 MR-MUF（批量回流模制底部填充 Mass Reflow Molded Underfill），通过在芯片间注入 EMC（液态环氧树脂模塑料 Epoxy Molding Compound）填充芯片之间或芯片与凸块之间间隙。EMC 本身具备中低温固化、低吸水率等优点，无需借助高温高压，可有效解决芯片翘曲从而提升良率。在设计 12 层 HBM3 时，公司采用了 Advanced MR-MUF 技术，进一步提升工艺良率和散热性能。

图表29: MR-MUF 对比 TC-NCF 温度下降 14%



资料来源: SK 海力士公告、华泰研究

图表30: Advanced MR-MUF 对比 MR-MUF 散热能力进一步提高



资料来源: SK 海力士公告、华泰研究

图表31: SK 海力士以 MR-MUF 代替 TC-NCF, 大幅降低 HBM 热阻

	HBM2	HBM2E	HBM3	HBM3 (12Hi) / HBM3E	HBM4
Stacking Tech.	TC-NCF	MR-MUF	Advanced MR-MUF	Advanced MR-MUF	TBD
Remark	Thermo-compression w/High Stress NCF ✓ World 1st TSV chip stack	Low Stress Air ✓ Low bond force & Robust joints: Higher Bump portion (thermal dissipation ↑)	Low Force & thermal Air ✓ More Enhanced thermal dissipation : Lower gap height & thermal resistance ↓	Advanced MR-MUF	Hybrid Bonding
Achievable Stack Height	4Hi / 8Hi	4Hi / 8Hi	8Hi / 12Hi	8Hi / 12Hi	12Hi / 16Hi
Thermal R (Relative)	○ (1.0)	○ (0.65)	○ (0.55)	⊙ (0.5)	⊙ (0.4 - 0.5)

资料来源: SK 海力士公告、华泰研究

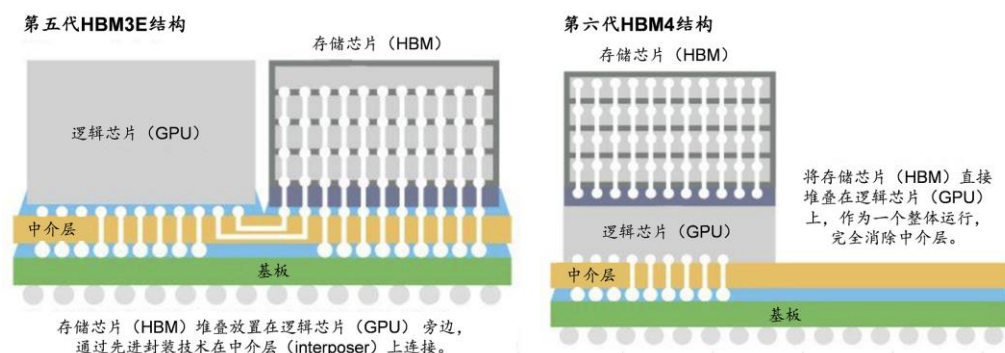
加速 1-β HBM3E 扩产, 海力士持续导入 EUV 设备。根据 Digitimes 24 年 2 月 26 日报道, 海力士将于 24 年投入约 2 兆韩元, 引进 8 台 EUV 设备。自 2021 年首度引进 EUV 以来, 目前海力士已拥有 5 台 EUV 设备, 未来合计将持有 13 台。以海力士 24 年半导体设备投资预算 5.3 兆韩元来看, 等同有将近 40% 规模集中于 EUV 曝光设备采购。我们认为, 海力士扩充 EUV 设备不仅加速其 DRAM 业务向 1-β 制程转换, 也将惠及 HBM3E, 公司通过扩产抢占市场份额的同时, 凭借先进制程取得与三星竞争时更出色的产品性能, 从而巩固其 HBM 龙头地位。

与三星不同, 海力士没有晶圆代工业务, 与台积电合作更加紧密, 二者作为受益于 AI 芯片的利益共同体, 合作推进 HBM4 研发。据 Digitimes 24 年 4 月 19 日报道, 海力士与台积电签署合作备忘录, 合作开发预计在 2026 年投产的 HBM4, 以此与三星的“一站式策略 (Turnkey Strategy)”竞争。该合作主要分为两个方面: **1) 海力士和台积电将首先致力改善 HBM 封装内最底层的基础裸片 (Base Die),** 该片与上方的 DRAM 不同, 对上方的 DRAM 堆栈 (Core Die) 起到控制作用, 同样连接至 GPU。过往包括 HBM3E 在内的海力士 HBM 产品的基础裸片由海力士基于自身制程工艺制造, 但从 HBM4 开始采用台积电先进逻辑制程, 更细微的工艺可以在基础裸片上增加更多功能, 从而满足客户定制化产品在性能和功效上更加多样的需求; **2) 双方将进一步优化海力士 HBM 产品和台积电 CoWoS 技术融合, 改善性能表现。**

Tom's Hardware 于 23 年 11 月也援引韩媒中央日报报导，海力士正在与包括英伟达在内讨论 HBM4 集成设计方案，计划与英伟达合作开发新的 GPU，拟将 HBM4 与逻辑芯片垂直堆叠在一起，通过 3D 封装实现在单芯片上存储器与逻辑芯片的垂直整合。该设计：1) 免去 HBM3E 设计中围绕逻辑芯片堆栈所带来的中介层复杂布线需求，从而简化设计并节约封装成本；2) 同时可进一步提升封装密度（在同样面积上铺设更多数量的元器件），缩小封装尺寸；3) 3D 封装采用更细小、更密集的电，缩短线路之间的传输距离，提升指令的相应速度和降低功耗。

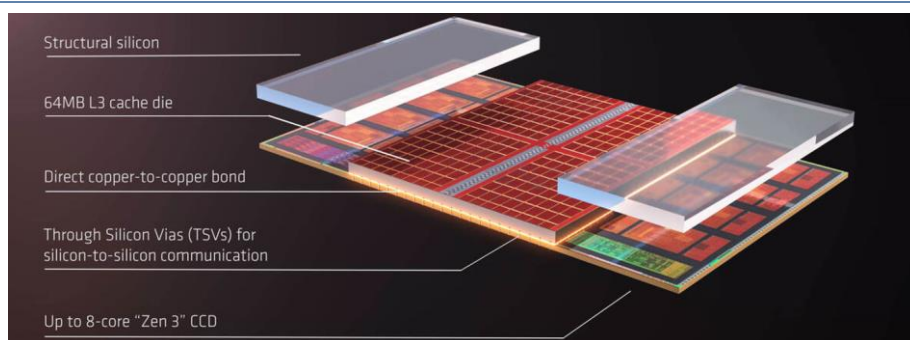
然而 Tom's Hardware 该篇报道亦指出，将存储器与逻辑芯片垂直整合的设计带来一定优越性的同时，因为在较小的封装体内堆叠多个芯粒，对散热也提出更高要求。参考采用 V-Cache（将一块 L3 Cache 置于 CPU 顶部）的 AMD CPU，通过对架构和硅平面的优化来缓解散热问题，例如 AMD 在 SRAM 单元中设计了 3D 64MB SRAM (Static Random Access Memory)，并避免其堆叠以保持低热密度，两侧采用了结构硅从高密度核心中逸出热量，并可通过控制芯片 TDP 和主频以补偿垂直堆叠产生的额外热量，不过，HBM 与 GPU 的功耗远高于 CPU，集成时的散热考量将更为复杂，因此合适的散热方法将是该设计成功的关键。

图表32：对比第五代产品 HBM3E，第六代产品 HBM4 有望实现存储器和逻辑芯片的垂直堆叠



资料来源：Tom's Hardware、华泰研究

图表33：AMD V-Cache 原理 - 将 L3 缓存置于 CPU 顶部，两侧硅片用于芯片加固和辅助散热



资料来源：海力士官网、华泰研究

除却与台积电合作，海力士亦计划在 HBM 中添加诸如基本运算以及高速缓存等功能。24 年 6 月韩媒 ET News 引述韩国业界消息，指出海力士已开始计划在 HBM 中集成 IP，计划从 HBM4 开始在 HBM 上搭载内存控制器，从而减少信号传输延迟和降低功耗，以及从 HBM4E 开始在 HBM 上集成基本运算和高速缓存等功能，将 GPU 和 CPU 的部分工作负载转移至 HBM 完成，从而进一步提升 AI 芯片的运算速度，并服务于客户的“定制化”需求。

我们认为，鉴于 HBM4 的复杂工艺需要，无论分工如何，海力士与台积电本次的合作都将缩短 HBM4 从研发、生产到与逻辑芯片适配的中间环节，并在未来的量产中缩短从存储颗粒制造、与逻辑芯片封装到交付的周期，同时减少海力士寻求第三方封装产能的不确定性。同时，在 HBM 中导入部分 IP 以转移逻辑芯片的工作负载，或将有利于提升延迟和功耗表现，并提升 AI 芯片的空间利用效率。

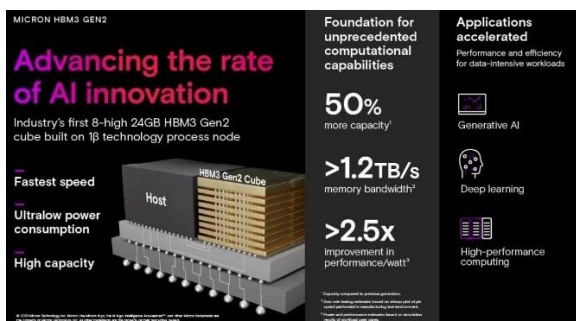
美光：跳过 HBM3，1-β 制程 HBM3E 加速追赶

我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中详细分析了美光在 HBM 市场中的定位：对比海力士和三星，美光并非 HBM 市场的先行者，其跳过 HBM3，意图通过 1-β 制程占领 HBM3E 市场份额。24 年 2 月 26 日，美光率先宣布实现 HBM3E 量产，确认供货英伟达 H200，24 年 3 月美光已宣布出样 12 层 36GB HBM3E，预期 25 年量产。韩媒 ET News 24 年 6 月 3 日报道美光 12 层 HBM3E 功耗表现优异。

我们认为美光将是本轮 HBM 高速增长的主要受惠者之一：1) 中短期内 HBM 市场仍将供不应求，美光 HBM3E 可作为“二供”承接溢出订单，同时扩展下游客户；2) HBM3E 相较传统 DRAM ASP 更高，借助 HBM3E 的营收贡献美光利润率有望迎来优化，同时 HBM 销售多基于年度合同，相较传统 DRAM 大宗商品属性更低，拥有更好的抗周期性；3) 与海力士和三星不同，美光借 HBM3E “重启” HBM 业务几近“从无到有”，2023 年 HBM 市场份额仅为 3%，低基数带来相较前两者更高的成长性；4) 英伟达扶持，受益于英伟达大量 HBM 需求。为保证 AI 芯片按期交付和避免过分依赖海力士，英伟达积极扶持美光作为 HBM “二供”，“货比三家”从而掌握更大的议价主动性。目前海力士 HBM3/3E “独供”格局已被打破，截至 24 年 4 月中旬，海力士和美光 HBM3E 均已通过英伟达效能验证，确认供货 H200。

对比美光，我们认为海力士优势主要有二：1) 产能：24-25 年 HBM 整体供不应求，HBM 产能上限或将决定受益上限，相较美光 HBM 几乎“从无到有”，海力士凭借先发优势，HBM3 和 HBM3E 产线成熟，产能优势较大，从而可以更充分受益于下游需求增长，维持“一供”地位；2) 良率：相较海力士 MR-MUF 键合技术，美光的 TC-NCF 键合技术良率较低，同时海力士在 HBM3 世代已有 8 层至 12 层爬坡经验，12 层 HBM3E 和 HBM4 量产相较美光或将更加顺利。

图表34：美光 HBM3E 产品展示



资料来源：美光官网、NextPlatform、华泰研究

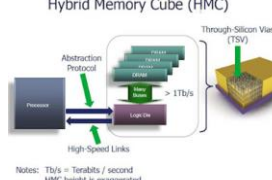
图表35：美光 HBM3E 规格参数

Capacity	24GB, 36GB
Technology	8-high, 12-high HBM3 Gen2
Form factor	11mm x 11mm
Date rate	>9.2Gbps
Burst length	8
Channels	16
Pseudo channels	2
DQ/Pseudo channels	32
Bandwidth	>1.2TB/s
Operating temperature	0°C ≤ T _{OPER} ≤ +105°C

资料来源：美光官网、华泰研究

图表36：美光 HMC 设计概念和性能参数：与 HBM 高度相仿

Hybrid Memory Cube (HMC)



Abstraction Protocol

High-Speed Links

Through-Silicon Vias (TSV)

Notes: Tbit/s = Terabits / second
HMC height is exaggerated

Generation 1 (4 + 1 memory configuration)

Technology	VDD	IDD	BW GB/s	Power (W)	mW/GB/s	pJ/bit	real pJ/bit
SDRAM PC133 1GB Module	3.3	1.50	1.06	4.96	4664.97	583.12	762
DDR-333 1GB Module	2.5	2.19	2.66	5.48	2057.06	257.13	245
DDR2-667 2GB Module	1.8	2.88	5.34	5.18	971.51	121.44	139
DDR3-1333 2GB Module	1.5	3.68	10.66	5.52	517.63	64.70	52
DDR4-2667 4GB Module	1.2	5.50	21.34	6.60	309.34	38.67	39
HMC, 4 DRAM w/ Logic	1.2	9.23	128.00	11.08	86.53	10.82	13.7

资料来源：EE Daily、华泰研究

三星：HBM3 世代落后海力士，HBM4 一站式策略或有望追赶海力士市场份额

三星于 23H2 开始量产 16GB/24GB 1-Z HBM3，并于 GTC 2024 期间展示其 12 层 36GB HBM3E，预期于 24H1 实现量产。英伟达 CEO Jensen Huang 会上于三星 12 层 HBM3E 旁签下“Jensen Approved”，确认该产品正在通过英伟达效能验证，当日（24 年 3 月 20 日）三星电子涨逾 4%，创 24 年 1 月 19 日以来最大单日涨幅。然而，Digitimes 和 Reuters 24 年 5 月 22 日报道三星 HBM3 和 HBM3E 尚未进入英伟达供应链，由于产品存在发热与能耗方面的问题，当周三星电子累计跌幅逾 5%。尽管该原因在 6 月的 COMPUTEX 期间被 Jensen 否认，但 24 年 4 月三星向英伟达测试的 8 层和 12 层 HBM3E 未成功被“实锤”。同时，三星 HBM3E 24Q2 实现量产计划仍然推进，HBM3 和 HBM3E 将保持供货 AMD。在 24 年 5 月举办的三星电子法人说明会上，三星管理层指出，预期 HBM4 将于 25 年开发、26 年量产，并将于 HBM4 开始引入混合键合技术。

我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中复盘了三星与海力士在 HBM 市场的竞争，两者同代产品更新间隔平均在半年之内，技术咬合紧密。进入 HBM3 世代以来，三星因起步较缓，短期内 HBM3 和 HBM3E 市场或难以追赶海力士，中期 HBM4 或存在追赶机会：1) 相较海力士，三星 HBM3 和 HBM3E 市场份额或难以相匹敌，同时截至 24 年 5 月三星 HBM3 和 HBM3E 尚未进入英伟达供应链，与海力士在 HBM3 和 HBM3E 世代的差距或将进一步达拉大；2) 对比海力士和美光，三星的独特之处在于其整合存储芯片和晶圆代工的能力，从而实现“一站式策略（Turnkey Strategy）”。中期来看，三星的“一站式策略”或将服务于 HBM4 的研发，其逻辑类似海力士与台积电的合作，三星存储与封装部门协同或将缩短 HBM4 从研发到生产的中间环节，并在未来的量产中缩短从存储颗粒制造、封装到交付的周期，从而能占得后续产品先机；3) 对比海力士积极推进 HBM3/3E 产能迭代，三星在 HBM2E 市场仍然有着来自 AMD Xilinx、云厂商自研芯片（包括 AWS 和谷歌等）及其他芯片厂商的需求。

对比三星，HBM 工艺将成为海力士 HBM “护城河”的关键：

1) 制程：与美光相同，海力士也在 HBM3E 中采用 1- β nm（第五代 10nm）制程，而三星 HBM3E 仍采用 1- α nm（第四代 10nm），据 Digitimes 24 年 6 月报道其 HBM3E 耗电为海力士的两倍。对于存储芯片，制程工艺的迭代带来内存密度的提升，同时降低芯片的功耗与发热，带来更好的能耗表现。我们前文提及 HBM 的垂直堆叠对芯片散热带来的挑战，而三星在制程上的短板或影响其 HBM 对客户吸引力；

2) 键合：在键合工艺的竞争中，三星 HBM 键合仍采用 TC-NCF 工艺，该过程需要高温高压环境将凸点（bumps）推入非导电薄膜，在 HBM 堆叠层数增加，单个 DRAM 高度减少的情况下更易导致芯片翘曲，进而影响良率。根据 Digitimes 24 年 3 月新闻，三星虽然也曾尝试 MR-MUF 技术，但未能解决翘曲问题，因此至今仍采用 TC-NCF 技术。因此，如何改良或替代现有的 TC-NCF 键合仍将是三星需要攻克的难题。

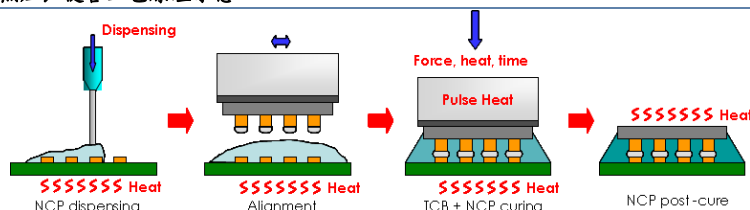
工艺问题和品质管理问题可从三星 HBM3E 产品验证“卡关”略见一斑。24 年 4 月三星向英伟达测试的 8 层和 12 层 HBM3E 未成功，根据韩媒 Alpha Biz 24 年 5 月报道，三星 8 层 HBM3E 产品验证因台积电采用标准问题“卡关”。报道指出，三星 HBM 被判定有问题，主因是负责英伟达 GPU 代工的台积电，在验证三星 8 层 HBM3E 时，采用海力士的标准。由于 SK 海力士 8 层 HBM3E 的生产方式与三星并不相同，导致三星产品未能顺利通过验证。同时，Digitimes 24 年 6 月报道援引韩国业界和成均馆大学教授杨锡俊观点，认为三星 HBM 虽然快速迭代，但一方面存在制程良率问题，且 HBM3E 样品耗电量为 SK 海力士 2 倍以上，另一方面相关业务经验不足，其 8 层 HBM 并未充分进行量产品质管理，因此阻碍产品验证流程，而其 12 层 HBM 产品或也将经历类似难题。

图表37：三星历代 HBM 参数对比

	HBM2 (Flarebolt)	HBM2 (Aquabolt)	HBM2E (Flashbolt)	HBM3 (Icebolt)	HBM3E (Shinebolt)	HBM4 (预测)
发布年份	2016	2018	2020	2023	2023	2025-2026
最大容量 (GB)	8	8	16	16/24	24/36	36-48
带宽 (GB/s)	256	307	460	819	1.225TB/s	1.5-2TB/s
VDD 电压 (V)	-	1.2	1.2	1.1	-	-
堆叠高度 (层)	8	8	8	8/12	8/12	16
堆叠技术	TC-NCF	TC-NCF	TC-NCF	TC-NCF	TC-NCF	Hybrid Bonding
制程	-	2y	1y	1z	1a	1b

资料来源：三星官网、Trendforce、Anand Tech、华泰研究

图表38：TC（热压）键合工艺原理示意



资料来源：Yiu Ming Cheung (2013.10):《Solder Joint Integrity of TC-NCP Thermo-compression Bonding Process》、华泰研究

HBM4：三大厂商方案齐出，混合键合延迟导入对海力士影响几何？

HBM4 是 HBM3E 下一代产品，当前行业标准尚未明确。综合三大厂商的技术路线图，HBM4 提供 36-48GB (v.s. HBM3E 24/36GB) 内存和 1.5-2TB/s 带宽 (v.s. HBM3E 1.2TB/s)，主要基于键合层数的提高 (16 层 v.s. HBM3E 12 层) 及内存接口数目翻倍 (2048 位 v.s. HBM3E 1024 位)，有望于 25 年推出并在 26 年实现量产。

HBM4 的迭代将带来三方面挑战：1) 芯片间键合工艺成为攻克关键，以应对散热和良率两大难题。16 层堆叠在提升堆叠层数的同时需要磨削单片 DRAM 高度，并尽量缩小 DRAM 间距，以控制产品尺寸。TF-NCF 等上一代键合工艺亟待改良，混合键合有望成为主流技术；2) 2048 位内存接口要求中介层 (Interposer) 布线更加复杂，如何连接 HBM 堆栈和逻辑芯片带来不同设计方案，海力士和三星探索存储器和逻辑芯片垂直集成；3) 2048 位内存接口及带宽提升同样需要单位面积上更多的 TSV 穿孔以提升 DRAM 堆栈的垂直互联密度。

我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中详细分析了铜-铜混合键合 (Cu-Cu Hybrid Bonding) 的技术特点，整体而言，混合键合的导入是 HBM4 的“必经之路”，其代替现有的微凸点，提升 I/O 密度、散热性能和能耗表现，在 HBM 层高不断增加的趋势下更适应应用场景。然而，鉴于铜-铜混合键合工艺复杂，其工艺流程涉及许多传统上仅由晶圆代工厂专用的工具，为确保第一代 HBM4 在 26 年落地，混合键合导入 HBM 生产或被推迟。

我们认为，从内存厂商角度看，混合键合的设备要求与三星、海力士和美光现有的键合设备和技术区别较大，ZDNet Korea 24 年 3 月报道指出混合键合设备比现有的 TC 键合机价格高出四倍，因此需要高昂资金成本配置。而初期设备落地后在键合良率上同样面临压力，需要额外时间。从 AI 芯片厂商角度来看，鉴于芯片迭代和出货需求，比起施压存储厂商应用混合键合生产 HBM4，接纳标准放宽以顺利推进芯片迭代和稳定 HBM 供应也是情理之选。

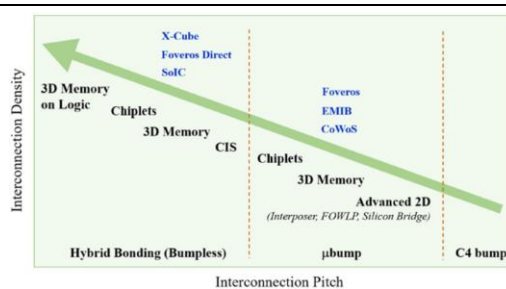
对于海力士而言，混合键合的延迟导入或许是把“双刃剑”，机遇和风险并存。一方面，HBM4 推出伊始，海力士仍可凭借 Advanced MR-MUF 取得相对三星和美光的良率领先，节约的成本亦可帮助其扩产以维持 HBM 市场份额，而另一方面，混合键合的延迟导入将给予美光和三星与海力士在键合工艺上重回“同一起跑线”的机会，三大厂商在 16 层堆叠时均面临技术考验，谁在掌握散热和良率方面能使用更好的材料，将左右今后的竞争力，HBM4 的竞争或许会更加激烈。

图表39：微凸点和混合键合横截面对比，混合键合无焊料球



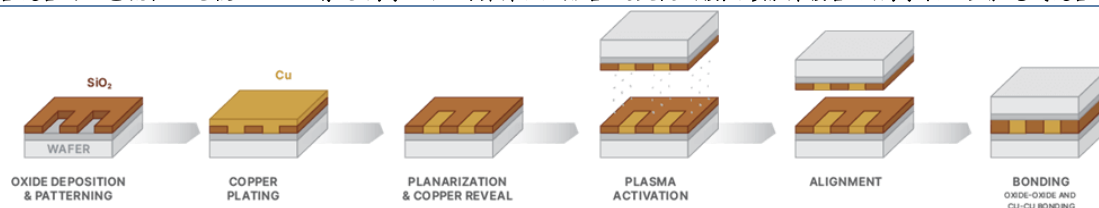
资料来源：Dr. Navid Asadi (2022.10):《Hybrid Bonding for Advanced 3D IC Packaging》、华泰研究

图表40：混合键合相较微凸点的 I/O 密度提升



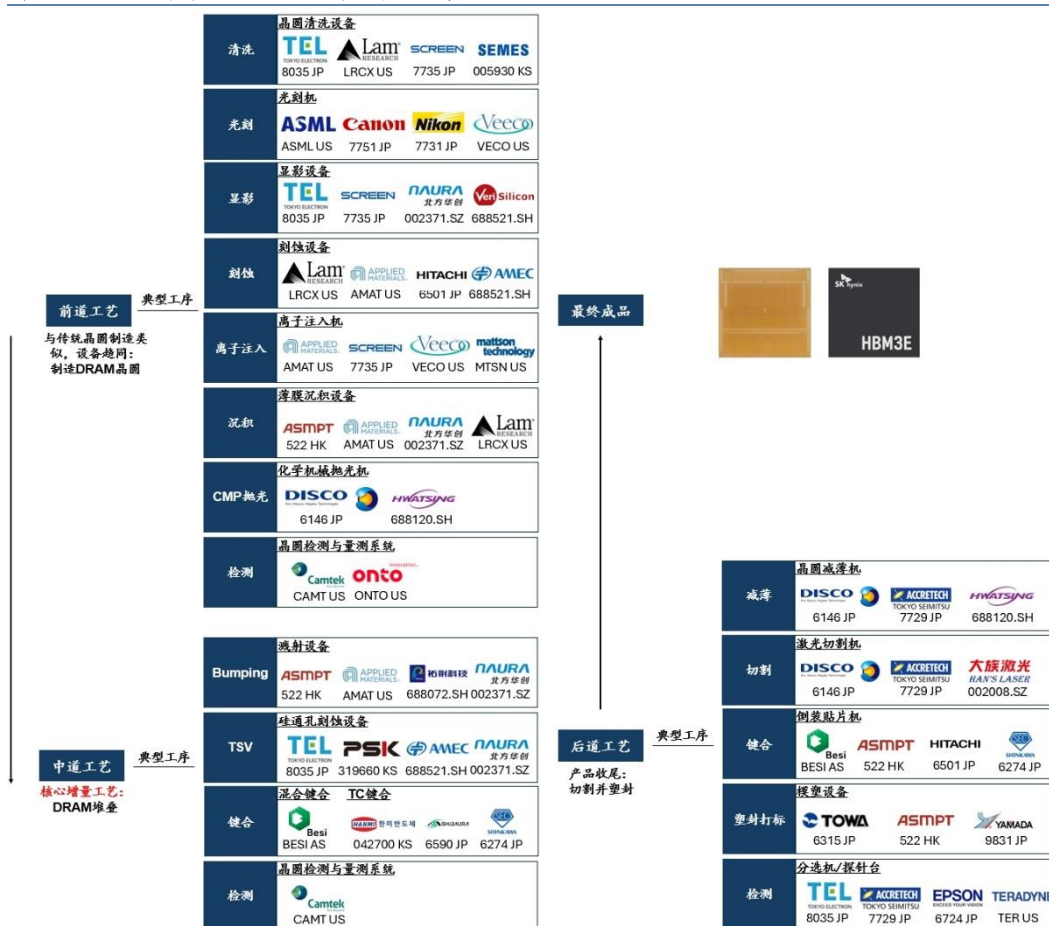
资料来源：Yoon-Gu Lee (2023.4):《Copper Bonding Technology in Heterogeneous Integration》、华泰研究

图表41：混合键合的工艺流程：电镀-CMP及清洗-离子活化-对齐并施压贴合-退火使铜接点膨胀并接合-铜离子相互扩散完成键合



资料来源：3D InCites、Digitimes、华泰研究

图表42：HBM 简要工艺流程及主要设备供应商梳理



资料来源：各公司官网、华泰研究

图表43：HBM 三大厂商 SWOT 分析

	SK 海力士	三星	美光
优势 Strengths	1.HBM 龙头，品牌形象和以英伟达为代表扎实的客户基础 2.研发实力累积，MR-MUF 键合技术相较竞争对手领先 3.与台积电和英伟达不构成竞争，率先合作开启 HBM4 定制化进程	1.资金雄厚，HBM 扩产激进，产能追赶节奏快，承接 AMD MI300/MI325X/MI350 系列 HBM 需求 2.集存储、AI 芯片设计、晶圆代工和封装为一体，承接自研芯片市场的 HBM 需求 3.晶圆代工业务技术反哺 HBM，混合键合导入或更加迅速	1.英伟达扶持作为“二供”，承接大量 HBM 需求，并以此有望进一步拓展客户 2.HBM 业务近乎“从无到有”，受益 HBM 市场迅速增长带来更高增长弹性 3.布局中国台湾，与中国台湾供应链合作紧密，同时亦与台积电和英伟达不构成竞争
劣势 Weakness	1.依赖 AI 带动 HBM 关键市场，传统 DRAM 和 NAND 竞争力较三星有限 2.资金较三星更少，HBM 未来扩产步伐或不及三星 3.依赖台积电 CoWoS 产能，供应链管理存在不确定性	1.与台积电和英伟达部分竞争，与二者合作不及竞争对手紧密 2.制程短板影响 HBM 性能，或影响对大客户吸引力 3.TC-NCF 键合良率低，单位 HBM 的晶圆损耗大，蚕食其余 DRAM 产能	1.HBM 起步晚，技术积累不及竞争对手 2.当前客户单一，仅确认供货英伟达，多元化有限 3.TC-NCF 键合良率低，单位 HBM 的晶圆损耗大，蚕食其余 DRAM 产能
机会 Opportunities	1.生成式 AI 牵动大模型浪潮，膨胀的内存需求带来 HBM 高速增长的 TAM 2. AI 数据中心对加速计算应用的扩张，以及多模态推理需求增加 3.HBM 短期仍供不应求，ASP 较传统 DRAM 更高，带来利润率优化空间 4.需求不拘泥 GPU，ASIC 市场对 HBM 同样需求广阔		
威胁 Threats	1.重大技术转变，如光通信和量子计算，或对需求带来颠覆式影响 2. HBM 高定价带来客户定位狭窄，除却高端 HBM 或难以铺开，同时扩产后供给缺口不再，未来存在被压价可能 3.存在其余 DRAM 厂商进入市场分食市场份额的可能		

资料来源：华泰研究

DRAM: AI 驱动需求回暖，稳定 24-25 年营收基本盘

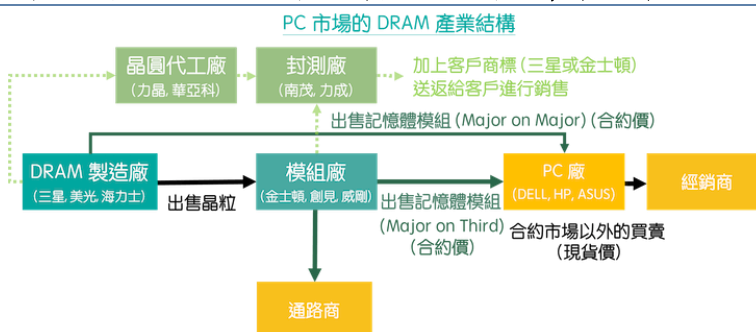
我们预计海力士 DRAM 业务 FY24/25/26 营收为 44.8/74.3/87.5 万亿韩元，对应营收同比为 114.7%/65.9%/17.8%。海力士 DRAM 业务 FY23 营收同比下降 27%，主要受疫情初期带来的电子产品需求见顶后的高基数影响。展望 24 年，各终端应用与半导体市场以 AI 需求为核心驱动：**1) 智能手机** 24H1 为传统淡季，呈现缓步复苏趋势，24Q1 AI 手机（如三星 Galaxy S24）表现亮眼，Oppo Reno 11 等 AI 手机新机型加速导入；**2) PC** 24Q1 出货量已基本恢复至 19Q1 水平，24H1 PC 市况已经逐步好转，虽然需求大规模增长尚未出现，但伴随换机潮和 AI PC 逐步落地，24H2 至 25 年 PC 市场可望迎来进一步复苏；**3) 传统服务器** 逐渐回温，订单能见度从原本 4-6 周提高至 8 周，**AI 服务器** 订单受北美需求拉动仍然高涨，24Q2 起英特尔 Eagle Stream 和 AMD 服务器 DDR5 新平台渗透率持续提升。24 年 5 月 22 日英伟达 FY25Q1 业绩会称 24Q1 北美云厂商持续加码 AI 算力基础设施，其数据中心业务的强劲增长（yoy +427%）得益于北美 CSP（Cloud Service Provider）规模部署和升级，FY25Q1 约贡献约 45% 数据中心收入。

我们预计 AI 驱动下 24 年存储行业将逐步迈入新一轮上行周期，DRAM ASP 和需求周期性回暖，带动供应端产能逐渐恢复，海力士 DRAM 将受益于在 24H2 至 25 年明朗的“供需两旺”趋势。根据 Digitimes 24 年 4 月报道，主流 DRAM 厂商自 24Q2 起持续扩大 DRAM 晶圆投片量，预计 24Q3 和 24Q4 稼动率或提升至 100%。鉴于 DRAM 晶圆投片到产品出货存在间隔，24Q2 起主流厂商扩大投片，应当是预期 24H2 需求稳固，24H2 产能恢复后将自 24 年底起至 25 年持续对 DRAM 业绩带来积极影响。值得一提的是，HBM 的扩产对 DDR5 产能形成挤兑，每生产 1 位元 (bit) 的 HBM，就有 3 位元 (bit) 的 DDR5 无法生产，DDR5 在 24 年初仍处于供不应求，24Q1 约有 3% 的供给缺口，伴随存储厂商陆续恢复增产，预期 24Q2 起缺口收敛至 1%，24H2 产能将持续开出，但考量供应端增产到实际出货增加仍需 15 个月左右，25 年 DDR5 整体供给仍将维持紧俏，从而维持 DDR5 ASP 涨势。

回到 24Q2，短期内受 4 月 3 日中国台湾强震对供应链影响，DRAM ASP 或见较大涨幅，DRAM 三大厂商之一的美光约有 65% 的 DRAM 在台湾生产，根据 Digitimes 24 年 4 月 15 日至 17 日系列报道，该次地震对美光厂房和基础设施未造成永久影响，单季 DRAM 供应影响约为 4%-6%，鉴于 DRAM 厂商原先已有 24Q2 DRAM ASP 调升 10-15% 规划，叠加地震影响，24Q2 美光 DRAM 合约价将调涨 25%，海力士和三星跟随将 DRAM 合约价调涨约 20%。

我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中详细分析了美光的传统 DRAM 产品矩阵，回顾上一轮 DRAM 周期的同时对本轮周期做出展望，海力士作为 DRAM 三大厂商之一，各产品线布局与美光相仿，除了 HBM 龙头产品，其传统 DRAM 产品亦覆盖通用内存 (DDR) 及对应内存模组 (DIMM)、低功耗内存 (LPDDR) 和显存 (GDDR) 等不同场景。鉴于传统 DRAM 大宗商品属性较强，标准化程度高，美光报告中分析的 DRAM ASP 变动同样适用于海力士。

图表44：DRAM 产业结构（以 PC 为例）：海力士属于 DRAM “原厂”，即绝大部分 DRAM 颗粒的来源



资料来源：ISCA2016 Samsung、华泰研究

图表45： 23Q4-24Q1 DRAM 合约价见较大涨幅，24Q2 DRAM 合约价环比涨幅或逾 20%

DRAM	2023Q4	2024Q1	2024Q2
PC DRAM	DDR4: 涨价 8-13%	DDR4: 涨价 15-20%	
	DDR5: 涨价 10-15%	DDR5: 涨价 15-20%	-
	平均售价: 涨价 10-15%	平均售价: 涨价 10-15%	
服务器 DRAM	DDR4: 涨价 5-10%	DDR4: 涨价约 20%;	
	DDR5: 涨价 0-5%	DDR5: 涨价 15-20%;	-
	平均售价: 涨价 8-13%	平均售价: 涨价 15-20%	
移动 DRAM	涨价 18-23%	涨价 18-23%	-
GDDR	涨价 8-13%	涨价 13-18%	-
利基型 DRAM	涨价 8-13%	DDR3: 涨价 8-13%	-
		DDR4: 涨价 10-15%	
		平均售价: 涨价 10-15%	
总 DRAM	涨价 13-18%	涨价 20%	涨价逾 20%

资料来源：Trendforce、Digitimes、华泰研究

图表46： 三大厂商 24 年 DRAM 产能规划

	三星	SK 海力士	美光
DRAM 产能计划	23Q4 DRAM 产能触底后，稼动率预计自 23Q4 的 65%-70%，恢复至 24Q1 的 81%，24Q2 回升至 89%，预计 24Q3-Q4 稼动率提升至 100%	24Q1 起将视 PC 和手机市场 ASP 和客户 需求修复情况，对 DRAM 减产计划进行 调整，24H2 稼动率回升至减产前水平	将针对性调升 DDR5 等高端产品产能，但整 体维持谨慎，FY24 年底晶圆投入量将相较 22 年峰值降低 11-13%。
资本支出	增加半导体设备投资和晶圆投入量，约 27 兆韩元（包括 HBM 和 NAND），集中量产 DDR5，扩大 HBM3/3E 比重。	24 年资本开支高于年初计划，一方面投 资 M15X 建设，另一方面因 HBM 晶圆消 耗相较传统 DRAM 更大，叠加 24 年 DDR5 等传统 DRAM 受 AI 和换机潮带动 需求高增，需要相应投资支持扩产。	资本支出 75-80 亿美元，因 HBM 晶圆消耗 较传统 DRAM 更大，24 年投入量同比下 降。
产量预估	24H2 开始出货量同比增加，24Q3 同比增长 11%， 24Q4 同比增长 41%，恢复至减产前水平。全年出 货量同比增长约 2%。	24Q1 业绩会指引 24 年 DRAM 产量同比 增长有限。	FY24Q2 业绩会指引 DRAM 供给增长将低于 需求。
行业产量预估	DRAM 供给同比增长中两位数，低于需求增长		

资料来源：Digitimes、台湾经济日报、首尔经济日报、Omdia、三星官网、海力士官网、美光官网、华泰研究

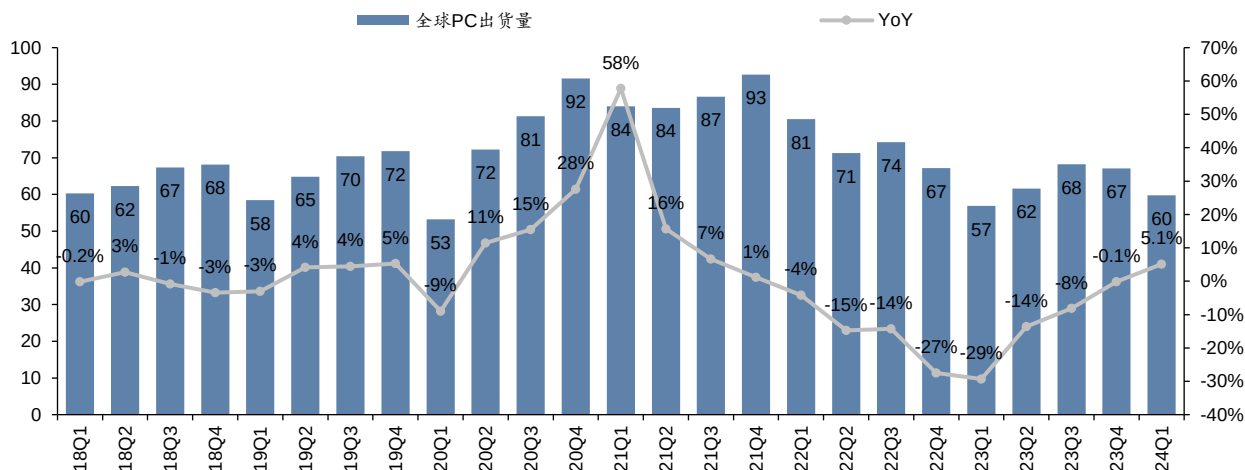
AI 拉动 PC、手机以及服务器出货量，撬动 DRAM 需求增长

PC 市场触底反弹，Windows 系统换代拉动换机潮

我们认为，在这波 AI 端侧设备的崛起过程中，AI PC 或率先成为主力，鉴于其在散热和应用场景均具备优势，而 AI 手机将紧随其后。热传导和热对流为目前手机及电脑的主流散热方案，而手机相对电脑散热面积小，热传导及热对流效率较低，因此在散热性上较逊色，且手机便携性需求高，设计以轻薄为首也限制了散热系统的部署。

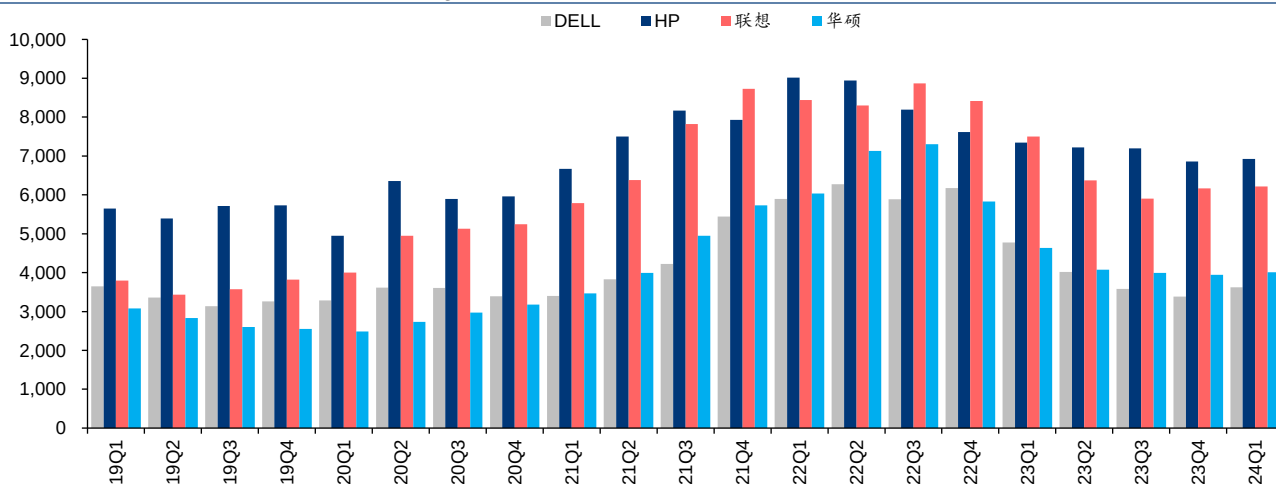
PC 市场 24H2 或开始复苏，根据 IDC 数据，23Q4 全球个人电脑出货量为 6710 万台，同比下降 0.1%，环比下降 1.4%；24Q1 全球个人电脑出货量为 5980 万台，同比上升 5.1%，环比下降 10.9%，恢复至 19Q1 水平。我们认为目前随着疫情后 PC 出货量已由高基数回落，加上 23H2 PC 库存开始恢复正常水平，且 Windows 系统换代及 AI PC 崛起，均将带动换机潮出现，但 PC 需求回暖和 AI PC 普及并非一蹴而就，因此需等到 24H2 才会出现更明显的回暖信号。DRAM 作为 PC 供应链上游，尽管 PC 需求总量增长较小，AI PC 渗透率尚待提升，但换机（PC 规格升级）和 AI PC 布局准备将带动下游厂商提前备货及硬件调试，使得在未来相关端侧 AI 应用上线之后，相关硬件设备能即时兼容，因此海力士端侧 AI 相关的 DRAM 产品（如 DDR5、LPDDR5T 和 LPCAMM2）将显著受益。

图表47：全球 PC 出货量：23H2 以来回温（百万，%）



资料来源：IDC、华泰研究

图表48：23 年主流 ODM/OEM 库存基本出清，24Q1 库存水位略高于疫情前水平



资料来源：Bloomberg、华泰研究

Windows 10 即将终止服务及 Windows 12 和 co-pilot 的推出将对 PC 需求有拉动作用。微软已确定将于 2025 年 10 月 14 日终止对 Windows 10 的支持，其后用户虽仍可使用 Windows 10，但微软不会再为其推出安全更新，这意味着该版本将会暴露于各种安全风险之中，我们认为这将触发换机需求，拉动 PC 出货量。另外，根据 PC Gamer 官网 2023 年 10 月 6 日报道，Windows 12 或将在 2024 年推出，我们认为新一代操作系统叠加 co-pilot 的推出，将有望进一步拉动 PC 需求。

AI PC 崛起，渗透率未来可期，内存升级需求带动 DRAM 增长

24 年 5 月 21 日微软 Build 开发者大会明确 AI PC 定义“Copilot+PC”，即 AI PC 搭载 NPU 算力须达至少 40 TOPS、支持全天候续航、支持大语言模型 (LLM)+小语言模型 (SLM) 并用以及云端和边端并行等特点。根据 Counterpoint 官网 2023 年 10 月 17 日预测，AI PC 的出货量在 2020~2030 年的 CAGR 将达到 50% 以上，并将从 2026 年起占据市场主导地位。Digitimes 24 年 4 月 11 日援引 Canalys 预测，2024 年将有 5000 万台具有 AI 运算功能的 PC 出货。

当然，AI PC 渗透并非一蹴而就，根据 Digitimes 24 年 3 月和 5 月采访 PC ODM 观点，预期 AI PC 24 年渗透率约 5%。报道指出，当前仍处于 AI PC 起步阶段，各机型以高阶机种为主，24H1 各厂商仍在观望定价 1000 美元以上的中高阶机种销售状况，预期 24 年整体 AI PC 将是维持“试水阶段”，25 年才会出现明显动能。但同时，报道称业界对 AI PC 相关芯片出货看法相对乐观，鉴于 AI PC 的硬件端升级要早于相关应用落地，24 年 PC 整



体规格升级需求将有显著提升，许多原定在 2023 年导入的新技术延期至 2024 年，带动 ODM 和 OEM 提前备货建立库存。

因此，展望 2025，各大芯片和 PC 厂商纷纷进军 AI PC，促进商业化落地节奏，25 年 AI PC 或将逐步在个人电脑市场占据重要份额。24 年 5 月 21 日微软 Copilot+PC 发布后，惠普、戴尔和联想等厂商均将自 6 月起陆续推出搭载高通骁龙 X Elite 和 X Plus 处理器的 AI PC。华硕同日发布的 Vivobook S15 已经支持 Copilot+PC，NPU 算力 45 TOPS，总算力 75 TOPS；英特尔 5 月 21 日宣布满足 Copilot+PC AI 性能的 Lunar Lake 处理器将于 24Q3 推出，NPU 算力 45 TOPS，该芯片支持逾 20 家 PC 制造商的 80+全新机种；24 年 5 月 7 日苹果发布 M4 芯片（38 TOPS），首发搭载于 Ipad Pro，为后续 MacBook 的更新铺垫。24 年 6 月英伟达于 COMPUTEX 发布 4 款 AI PC 及 RTX AI 套件，协助在 Windows PC 上部署 AI 模型。

值得注意的是，华硕 Vivobook S15 发布会提及，相较传统 PC，AI PC 的内存需求从 8GB 提升至 16GB 以上。因此我们预计 AI PC 潮将带动 PC 内存规格升级，拉动 DRAM 需求量增长同时利好海力士对 DDR5 及 LPDDR5T/5X 营收增长。

图表49：各 PC 厂商 AI PC 上市节奏

厂商	AI PC 上市节奏
联想	2023 年 10 月 14 日，联想展示 AI PC，能创建个性化的本地知识库，通过模型压缩技术运行个人大模型，实现 AI 自然交互；联想 AI PC 预计在 2024 年 9 月后正式上市
苹果	24 年 5 月 7 日苹果发布 M4 芯片（38 TOPS），首发搭载于 Ipad pro 为后续 MacBook 的更新进行铺垫
惠普	惠普正与关键软件供应商、关键芯片提供商展开紧密的合作，探索和重新设计 PC 的架构，最早会在 2024 年推出全新 PC
宏碁	宏碁目前已经与 CPU 厂商展开合作，预计将把 AIGC 或其他 AI 应用导入到终端设备上，相关 AI 笔记本方案会在 2024、2025 年陆续推出
戴尔	戴尔宣布与英伟达合作推出新的生成式 AI 解决方案，帮助客户在本地快速安全地构建生成式 AI（GenAI）模型；此外将发布内置 AI 功能的小型节能新电脑

资料来源：各公司公告、华泰研究

图表50：英特尔 AI PC 产品规划参数对比

系列名称	Meteor Lake	Lunar Lake	Arrow Lake	Panther Lake
（预计）发布时间	2023.12.14	2024.6.6	2H 2024	2025
CPU 单元制程节点	Intel 4	TSMC N3B	Intel 20A	Intel 18A
GPU 单元制程节点	TSMC 5nm	TSMC 3nm	TSMC 3nm	TBC
CPU 微架构	Redwood Cove (P-Core) Crestmont (E-Core)	Lion Cove (P-Core) Skymont (E-Core)	Lion Cove (P-Core) Skymont (E-Core)	Cougar Cove (P-Core) Darkmont (E-Core)
最大核心数	14	8	24	TBC
GPU 微架构	Xe-LPG	Xe2	Xe2	Xe3-LPG
内存支持	DDR5 / LPDDR5X / LPDDR5T			
技术要点	英特尔首次使用 EUV 制造芯片；首次在 CPU 使用 chiplet 架构，提升了制造的灵活性和成本效益；首次加入了 AI 加速引擎 NPU；引入新低功耗 E-Core。	使用台积电 N3B 和 N6 工艺混合制造，面向 AI 搭载 48 TOPS NPU；首次采用封装内存，将 LPDDR5X 与逻辑芯片统一封装进一步缩短数据传输延迟和能量损耗	首款具有人工智能功能的 PC 游戏处理器；Intel 20A 使用 PowerVia 和 RibbonFET 两项突破性技术：RibbonFET 是英特尔基于 GAA 的技术，能有效的控制电流流通，解决漏电问题并提高能源效率；PowerVia 是一种背面供电方案，通过将电力传输与信号发送分离，使得信号互联层空间增加，带来效率和性能的提升。	CPU 架构更新：采用全新 Cougar Cove P 核和 Darkmont E 核；iGPU 架构更新：由 Lunar Lake 的 Xe2 升级至 Xe3 "Celestial" 微架构。
AI PC 方案架构	CPU+GPU+NPU			
NPU 算力	-		>40 TOPS	
整体算力 (TOPS)	34	120	TBC	TBC

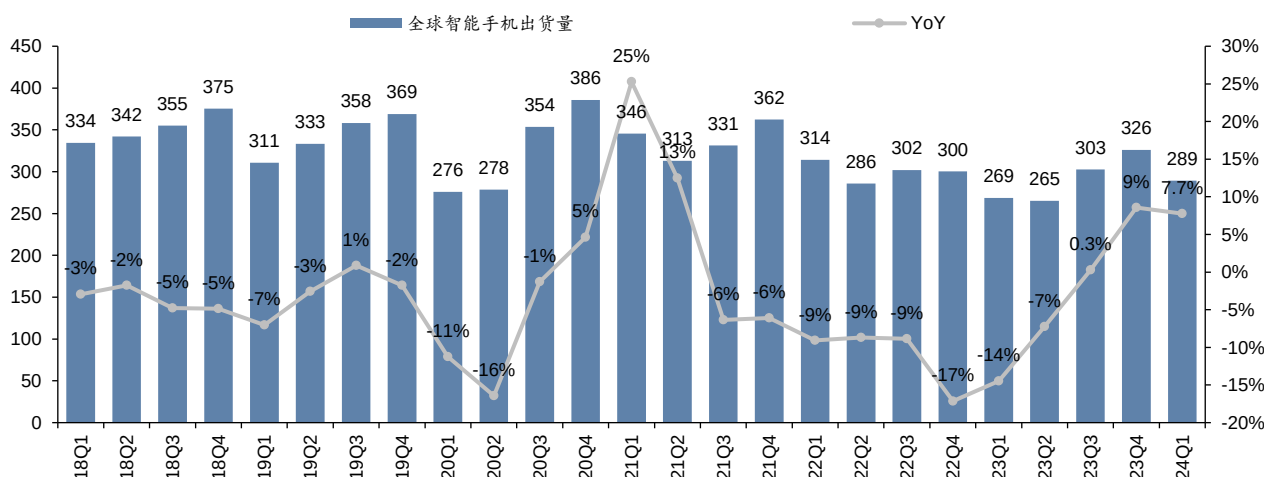
资料来源：英特尔官网，Anandtech 官网，Videocardz 官网，华泰研究

AI 手机引领旗舰潮流，中端机种开始全面渗透，拉动 LPDDR5X 需求

我们认为截至 24Q1，手机市场已经开始逐步复苏。根据 IDC 24 年 4 月数据，23Q3 至 24Q1 连续 3 个季度呈现同比上涨趋势（23Q3/23Q4/24Q1 同比增长 0.3%/8.6%/7.7%），2024 年出货量预计同比增长 2.8%。根据 Digitimes 23 年 3 月报道，2024 年全球 AI 手机出货量或增至 1.7 亿台，占手机市场出货比重达 14%-15%。Digitimes 24 年 3 月 12 日报道指出，三星的 S24 手机 AI 应用仍相对“轻量化”，主要集中于照片编辑、翻译搜寻、写作助理等，而想在手机端顺畅运转 70 亿参数的 LLM 模型需要 7-8 GB DRAM 容量，相当于当前主流手机（8/12GB）的大部分内存，为维持其他功能正常运作，AI 手机的理想 DRAM 容量需达到 24GB。AI 手机相较传统非 AI 旗舰机型多出 50%-100% 的 DRAM 需求，而鉴于手机功耗限制，因此带动 LPDDR5X 需求。

同时，伴随中国手机品牌入局 AI 手机，AI 功能开始全面渗透进入中端手机市场。华为、小米、Vivo、Oppo 和荣耀等品牌同样加速 AI 手机产线的布局力度。2024 年 3 月底，一加发布 Ace 3V，搭载 Snapdragon 7+ Gen 3 平台，售价为人民币 1,999 元起，开启 AI 进入中端手机的大门；2024 年 4 月 9 日，努比亚举行 2024 年春季新品发布会，将 AI 手机价格一举压到百元机款位阶；2024 年 4 月 10 日，红米（Redmi）也在新款手机搭配 AI 功能，Redmi Turbo 3 搭载 Snapdragon 8s Gen 3 旗舰平台，搭配 LPDDR5x+UFS4.0 内存架构，起售价为 1,999 元。从目前中端机 AI 功能来看，以 Redmi Turbo 3 为例，该机型提供 AI 通话、AI 隔空手势交互、智慧扩图、魔法消除 Pro、AI 相册等功能；其小爱大模型还配文本创作、文章摘要、知识问答、角色扮演等功能。

图表51：全球智能手机出货量：23H2 以来回温（百万，%）



资料来源：IDC、华泰研究

图表52：三星 Galaxy S24 Ultra 对比 S23 Ultra

	三星 Galaxy S24 Ultra	三星 Galaxy S23 Ultra
处理器	高通骁龙 8 Gen 3	高通骁龙 8 Gen 2
内存	12GB LPDDR5X	8/12GB LPDDR5X
存储	256/512GB/1TB UFS4.0	256/512GB/1TB UFS4.0
电池	5000mAh	5000mAh
连接	NFC、WiFi 7、蓝牙 5.3、UWB	NFC、WiFi 6E、蓝牙 5.3、UWB

资料来源：三星官网、华泰研究

AI 服务器需求强劲，通用服务器配置新平台引领换机潮

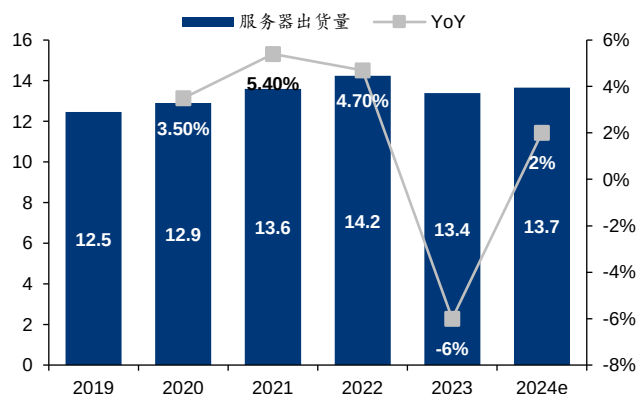
受益于通用服务器换机潮和 AI 服务器需求上升，我们预期 24-25 年 DRAM 产品组合向 DDR5 和 LPDDR5X 转变，带动 ASP 提升。

AI 服务器 24 年需求维持强劲，一台 AI 服务器的 DRAM 容量是通用服务器的 6-8 倍，因此快速增长的 AI 服务器需求将带动 DRAM 需求的高速增长。Trendforce 24 年 2 月估计 24 年 AI 服务器出货量同比两位数增长，占整体服务器出货的比重也将达 12.1%。AI 服务器动能同样可从台积电 24Q1 法说会预期略见一斑，台积电 24 年 4 月 18 日法说会预期 24 年服务器 AI 处理器（用于 AI 训练和推理，不包含边缘设备）营收贡献增长超 1 倍，24 年起未来 5 年营收 CAGR 逾 50%，将是台积电 HPC 平台成长和整体营收增长的主要驱动力。24 年 5 月 22 日英伟达 FY25Q1 业绩会称 24Q1 北美云厂商持续加码 AI 算力基础设施，其数据中心业务的强劲增长（yoy +427%）得益于北美 CSP 大规模部署和升级，FY25Q1 约贡献约 45% 数据中心收入。

同时，伴随数据中心基础设施建设需求，云服务提供商的支本开支增加，2024 年的通用服务器市场亦将表现回升态势。23 年由于 IT 资本开支预算有限，AI 服务器支出的迅速增加导致通用服务器的需求受挤兑暂时下降，但通用服务器的投资仍是发展 AI 设施的必要条件之一。根据 Digitimes 24 年 2 月 1 日和 4 月 18 日报道，传统服务器逐渐回温，24 年传统服务器订单能见度从原本的 4-6 周提高至 8 周，北美云厂商除了持续采购 AI 服务器，也开始对通用型服务器进行新一轮采购，加速英特尔 Eagle Stream 及 AMD Genoa 等采用 DDR5 的新平台布局。

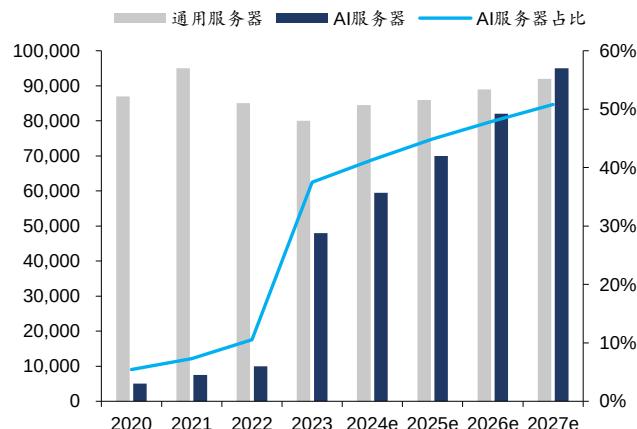
值得一提的是，随着数据中心及 HPC 领域能耗飙升，低能耗成为了服务器的一大关注重点，众多厂商纷纷加码低能耗产品以求突破，与 ARM 相仿，LPDDR5X 也凭借低功耗优势从移动端向服务器 DRAM 渗透，2023 年发布的英伟达 Grace CPU Superchip 即为 ARM+LPDDR5X 产品，搭配 ARM Neoverse V2 和 960GB LPDDR5X，带宽 1TB/s。Digitimes 24 年 4 月 23 日报道三星成功研发带宽 10.7 Gbps 的 LPDDR5X，计划同时面向移动端和 AI 服务器市场，该报道同时援引 Omdia 观点，表示 LPDDR 应用领域有望持续扩大至 HPC 和车载运算等场景。

图表 53：2019-2024 年全球服务器出货量（单位：百万，%）



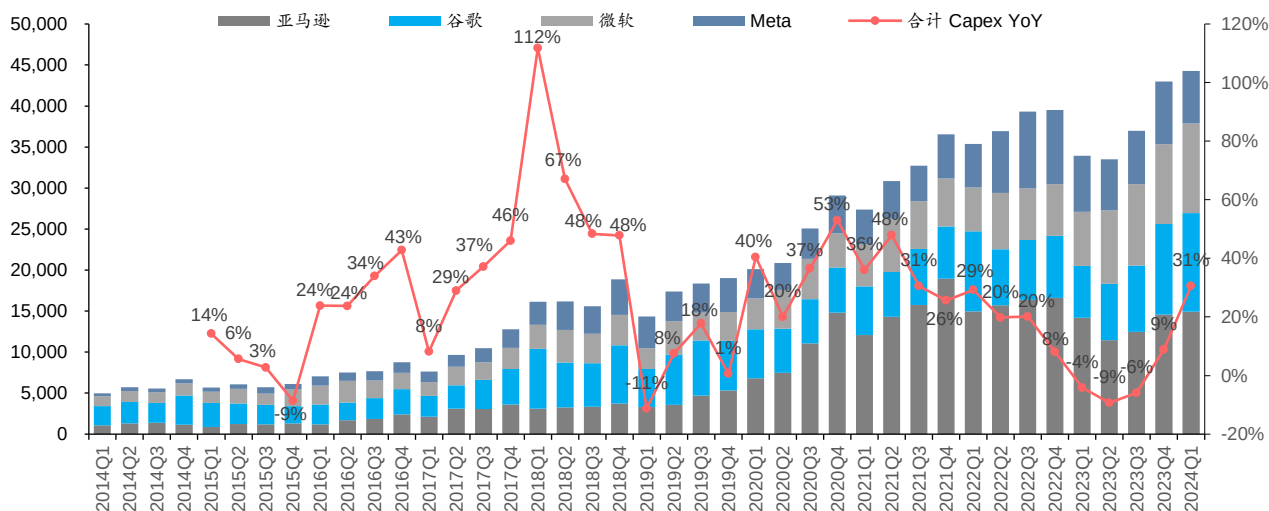
资料来源：TrendForce、华泰研究

图表 54：2020-2027 年全球服务器市场规模（单位：百万美元，%）



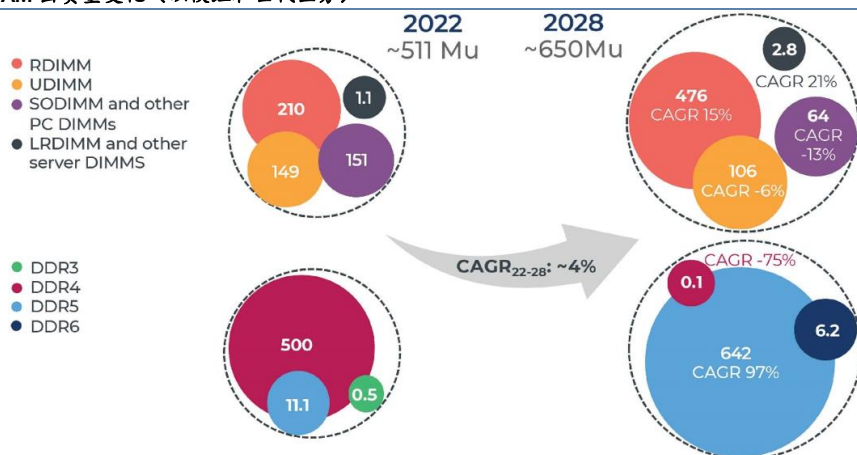
资料来源：IDC、The Next Platform、华泰研究

图表55：四大科技巨头近十年的资本开支（百万美元，%）



资料来源：各公司财报、华泰研究

图表56：2022-2028 DRAM 出货量变化（以模组和世代区分）



资料来源：Yole Intelligence、华泰研究

图表57：Intel 和 AMD 采用 DDR5 的服务器 CPU

公司	英特尔					AMD							
系列名称	Sapphire Rapids	Emerald Rapids	Sierra Forest	Granite Rapids	Clearwater Forest	Genoa	Genoa-X	Bergamo	Siena	Turin	Turin-X	Turin-Dense	
发布时间	2023.1	2023.12	2024	2024	2025	2022.10	2023.1	2023.1	2023.9	2024	2025*	2025*	
平台名称	Intel Eagle Stream	Intel Eagle Stream	Intel Mountain/Birch Stream	Intel Mountain/Birch	Intel Mountain/Birch	EPYC 9004 Zen 4	EPYC 9004 Zen 4 V-	EPYC 9004 Zen 4C	EPYC 8004 Zen 4	EPYC Zen 5	EPYC Zen 5	EPYC Zen 5C	
工艺节点	Intel 7	Intel 7	Intel 3	Intel 3	Intel 18A	5nm TSMC	5nm TSMC	4nm TSMC	5nm TSMC	4nm TSMC	4nm TSMC	3nm TSMC*	
插槽	LGA 4677	LGA 4677	LGA 4677/7529	LGA 4677/7529	LGA 4677/7529	LGA 6096	LGA 6096	LGA 6096	LGA 4844	LGA 6096	LGA 6096	LGA 6096	
最大核心数	60 P	64 P	288 E	136 P*	288 E*	96	96	128	64	128	128	192	
最大线程数	120	128	288	272*	288*	192	192	256	128	256	256	384	
L3 Cache (MB)	105	320	108	480*	-	384	1024	256	128	384	1536	384	
内存	DDR5-4800	DDR5-5600	DDR5-6400	DDR5-6400	-	DDR5-4800	DDR5-4800	DDR5-4800	DDR5-4800	DDR5-6000*	DDR5-6000*	DDR5-6000*	
内存通道	8通道	8通道	8通道	12通道	-	12通道	12通道	12通道	6通道	12通道	12通道	12通道	
PCIe	80 Gen 5	80 Gen 5	68 Gen 5	136 Gen 5	-	128 Gen 5	128 Gen 5	160 Gen 5	96 Gen 5	-	-	-	
最大功耗 (W)	350	350	350	500	-	400	400	320 (cTDP 400)	225	480 (cTDP 600)	480 (cTDP 600)	480 (cTDP 600)	

*注：具体参数官方尚未确认

资料来源：AMD 官网、Wccitech、the Register、Hardware Times、Tom's Hardware、华泰研究

面向端侧 AI，海力士 GDDR7、LPDDR5X/5T 和 LPCAMM2 布局完善

我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中指出，端侧 AI 将成为 AI 计算的下一个发展重心。端侧 AI 部署是一种将 AI 模型（数据和计算）和应用均置于个人或移动设备上运行的方式，并在处理器中新增了 NPU 模块（AI 加速器），提供更强大的 AI 算力，相较纯云端在成本、延时、隐私安全和定制化等方面具有优势。

海力士于 17 年 5 月的英伟达 GTC 上展示第一代 8Gb GDDR6，提供 16Gbps 带宽，之后逐步更新 16Gb 版本，将带宽提升至 20 Gbps，并将于 24H1 发布 GDDR7。根据 JEDEC 24 年 3 月 5 日发布的第一代 GDDR7 标准，其对比 GDDR6 世代进一步提升内存密度和带宽，可提供 16-24Gb 的内存容量和 32 Gbps 的带宽，以 2 通道（每通道 16 bit）为计，可以为显卡提供最高 128GB/s 的带宽，显卡性能表现相较前代提升 60%。

海力士在 GTC 2024 展示 16Gb (2GB) GDDR7，提供最大 40 Gbps 带宽，高于 JEDEC 标准。根据 Tom's Hardware 24 年 3 月 25 日对美光、三星和海力士三大存储厂商的采访，搭载 GDDR7 的显卡预期将于 24 年底面世，例如英伟达 Blackwell 架构游戏显卡（GB202，台积电 4nm，512 位显存位宽，32GB VRAM）和 AMD RDNA 4 GPU。该篇采访同时指出，当前量产的 GDDR7 为 16Gb 版本，而 24Gb (3GB) 版本预期将于 25 年面世。鉴于 12GB 显存已成主流 3A 游戏开启最高性能（如 4K 画质）的最低配置要求，更多显存在 12GB 及以上的显卡面世将成为大势所趋。

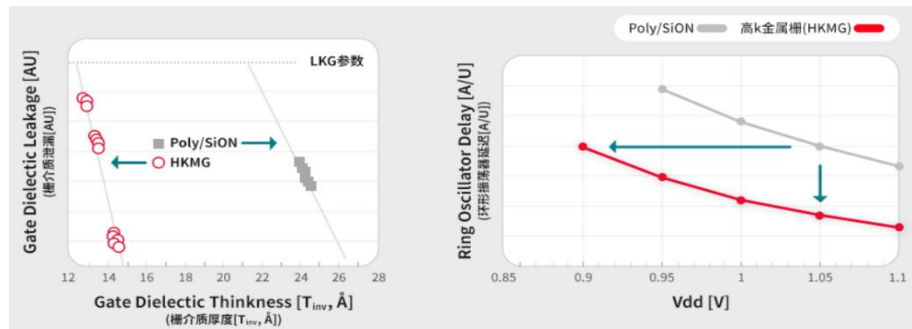
图表58：GTC 2024 上海力士展出的 GDDR7 样品



资料来源：Tom's Hardware、华泰研究

21-23 年海力士先后推出 LPDDR5、LPDDR5X 和 LPDDR5T，分别提供 6.4 Gbps、8.5 Gbps 和 9.6 Gbps 带宽。 1) **LPDDR5X**：22 年 11 月海力士率先在 LPDDR5 基础上引入 HKMG（High-k/Metal Gate）工艺，研发 LPDDR5X，对比前代 LPDDR5 功耗降低 25%，23 年 8 月 24GB LPDDR5X 实现量产，搭载在 2023 年 8 月发布的 OPPO“一加 Ace2Pro”智能手机上。其中，HKMG 工艺通过引入一层高 k 薄膜替代 SiON 栅氧化层，实现晶体管栅氧化层厚度减少同时防止漏电。更薄的栅氧化层支持晶体管进一步微缩，源极和漏极间距离缩短，栅极电压降低，从而节约功耗。2) **LPDDR5T**：海力士 LPDDR5T 于 2023 年 1 月推出，采用 1-α 制程和 HKMG 工艺，首发容量 16GB，相较前代 LPDDR5X，LPDDR5T 实现峰值带宽提升 13%。2023 年 10 月 LPDDR5T 通过高通骁龙 8 Gen3 和联发科天玑 9300 平台验证，并在 2023 年 11 月实现量产，首发搭载于 vivo X100 和 X100 Pro，24 年旗舰机型或将见到更多搭载 LPDDR5T 的设计。根据公司 FY23Q4 业绩电话会，24GB LPDDR5T 量产已经提上规划。

图表59: HKMG 实现晶体管栅氧化层厚度减少, 功率降低



资料来源: SK 海力士公告、华泰研究

图表60: SK 海力士最新 4 款 LPDDR 参数对比

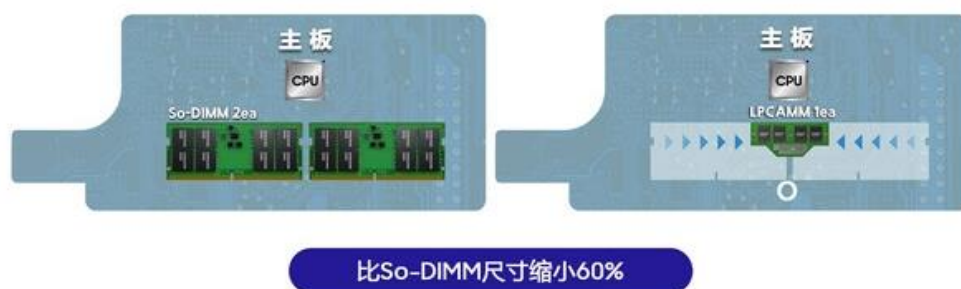
	LPDDR4X	LPDDR5	LPDDR5X	LPDDR5T
发布年份	2017	2021	2022	2023
Die 密度 (GB)	2/4/8/12	8/12/16/24	8/12/16/24	16/24
传输速率 (Gbps)	4.2	6.4	8.5	9.6/10.5
电压 (V)	1.8/1.1/0.6 (VDD1/VDD2/VDDQ)	1.8/1.05/0.5 (VDD1/VDD2/VDDQ)	1.8/1.05/0.5 (VDD1/VDD2/VDDQ)	1.01~1.12
I/O 架构	1ch x16	1ch x16	1ch x16	-
封装技术	POP / MCP / Discrete	POP / MCP / Discrete	PoP / MCP / Discrete	-

注: VDDQ 为 I/O 端口供电的电源, 降低 VDDQ 能减少功耗

资料来源: SK 海力士公告、华泰研究

LPCAMM 是一种可拆卸内存模组, 可有效克服 LPDDR 和 So-DIMM (small outline dual in-line memory module 小型双内嵌式内存模组) 的缺陷。个人 PC 和笔记本电脑通常采用传统的 LPDDR 或基于 DDR 的 So-DIMM 模组。但是, 由于结构上的限制, LPDDR 需直接安装在设备的主板上, 这使得在维修或进行升级时难以替换。与此相反, So-DIMM 虽然在安装或拆卸上更为便捷, 但在性能、功耗以及其他物理属性方面仍有许多限制。随着行业对更高效、更小巧设备的需求日益增长, LPCAMM 作为搭载 LPDDR 的模组, 其兼具 LPDDR 的能效表现和 So-DIMM 的易修性, 同时其封装尺寸更小, 在制造过程中为 PC ODM/OEM 提供了更大的灵活性。据韩媒 ZDNet Korea 24 年 1 月消息, LPCAMM 或将是 26 多年来首度取代 So-DIMM 的产品, 相较 So-DIMM, 其低能耗和小型封装尺寸可以使得笔记本设计更加轻薄。

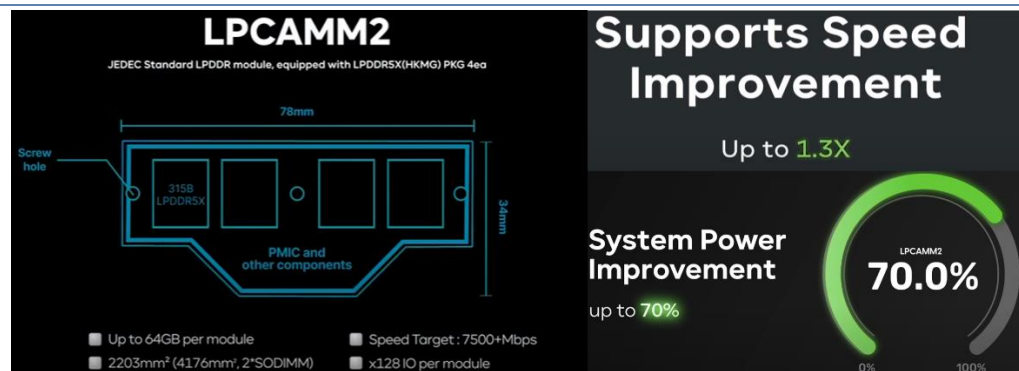
图表61: LPCAMM 模组相较于 SODIMM 减少设备内部空间占用



资料来源: 三星官网、华泰研究

海力士于 23 年 11 月的 SK 科技峰会上首度公开 LPCAMM2, 并在 CES 2024 和 24 年 5 月的戴尔全球科技峰会上展示样品。根据 Digitimes 24 年 2 月 1 日报道和美光官网 1 月数据, AI PC 所需 DRAM 容量是普通 PC 的两倍以上, 以 Llama2-70B (INT4) 模型为例, 其在端侧载和推理需要大约 42GB 的内存, 输出速度为 1.4 token/秒, 这是一般个人 PC 无法提供的内存规格。海力士展示的 LPCAMM2 模块计划 24 年内实现量产, 容量最高可达 64GB, 传输速率最高可达 9.6 Gbps, 相较其 SO-DIMM 功耗降低 47%, 效能提升 50%, 单 LPCAMM2 模块相当于 2 个 DDR5 SO-DIMM。

图表62：海力士 LPCAMM2 模组结构



资料来源：海力士产品页面、华泰研究

加速 EUV 导入 1-β 制程，24-25 年加速先进制程爬坡

对于存储芯片，制程工艺的迭代是缩减晶体管尺寸，带来单位面积上实现更小、更紧密的晶体管布局，提升内存密度，实现更高性能的同时达到更好的功耗表现。Trendforce 24 年 3 月报告指出，DRAM 供应端大幅转进至先进制程生产 DDR5 将带来成本优化，使得供应端利润明显改善。

2023 年 5 月 SK 海力士紧随美光跟进 1-β 节点，24 年着重 1-α 和 1-β 节点产能爬坡。根据 Digitimes 23 年 12 月报道，海力士规划 24 年于利川 M16、M14 工厂扩充 1-α 和 1-β 的产能，并将从 1-α DRAM 开始着手导入 EUV 制程，计划在 2024 年底将 1-α 和 1-β 产量增加到 DRAM 总产量的一半以上。报道同时指出，相较于美光在 HBM3E 和 LPDDR5X 等产品广泛铺开 1-β 制程，海力士优先将 1-β 制程应用于 24 年 3 月量产的 HBM3E，预期 1-β 制程量产初期月产能 1.5 万-2 万片，对比海力士整体 DRAM 月产能逾 40 万片，1-β 比重仍然较低。Digitimes 24 年 4 月援引韩国业界预期，称 24Q4 海力士 1-β DRAM 产能将约占总产能的 11%。

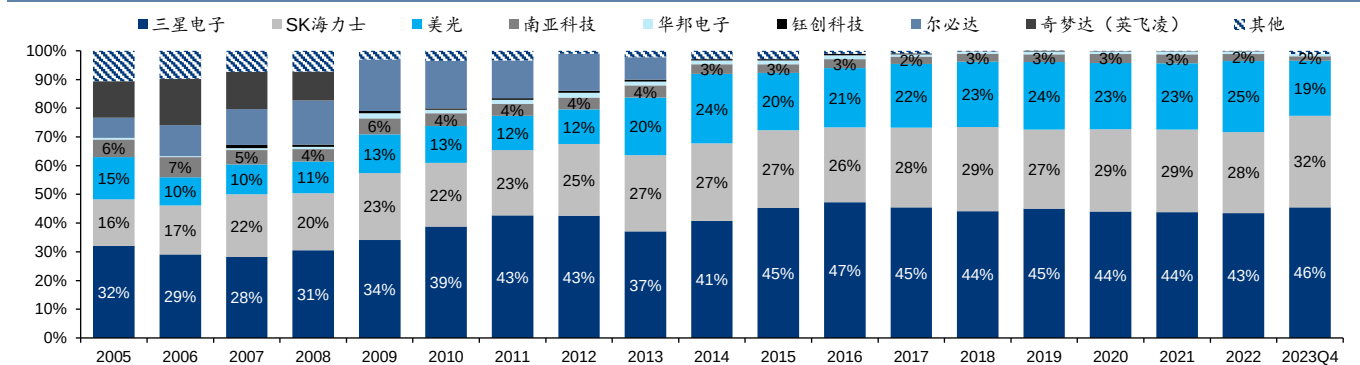
海力士无锡工厂 DRAM 产线制程亟待升级，但长期看来无锡产能或将缩减。海力士无锡工厂承接海力士 40% 的 DRAM 产能，主要量产 1-y 和 1-z 等成熟制程产品。韩媒首尔经济 24 年 1 月援引韩国业界消息指出，海力士计划 24 年内将无锡工厂 C2 产线转换为 1-α 制程，而鉴于 19 年起美国对 ASML EUV 光刻机进入中国的限制，海力士或难以将无锡工厂转换为 1-α 以上的制程。因此，考虑 24-25 年竞争对手美光制程已经基本转向 1-α 和 1-β 并广泛应用于 DDR5 和 HBM3E 生产，同时 1-γ 制程导入已经提上日程，海力士或考虑逐步减少无锡园区产能，将更多产能转回韩国园区，并通过新建 M15X 等新 DRAM 产线实现先进制程产能爬坡，由此带来的资本开支增加值得关注。

竞争格局：行业高度集中，三足鼎立技术咬合紧密，制程产品各有千秋

传统 DRAM 受 JEDEC 组织规范，标准化程度较高，相较于 HBM 等高端产线有着更强的大宗商品属性，同代产品间差异较小。DRAM 行业高度集中，早在 2014 年已确立三寡头格局。奇梦达 (Qimonda) 于 2006 年 5 月 1 日由英飞凌的内存部门拆分而成。奇梦达于 2009 年 1 月 23 日宣布破产清算，尔必达在 2013 年 7 月 31 日被美光收购，自此，DRAM 市场的三足鼎立格局确立。

以营收为计，23Q4 三星/海力士/美光的 DRAM 市占率分别为 46%/32%/19%，其中：1) 三星 23Q4 DRAM 营收 80 亿美元，环比+53%，主要系年底 IT 采购高峰拉动，根据 Digitimes 24 年 3 月报道，三星 DRAM 产能约为海力士 1.5-2 倍，得以支持更大的订单量；2) 海力士 23Q4 DRAM 营收 55 亿美元，环比+19%，受益于 HBM3 独供英伟达，海力士 23 年市场份额从 23Q1 的 24% 跃升至 23Q4 的 32%；3) 美光 FY24Q2 DRAM 营收 42 亿美元，环比+21%，由于 HBM 布局较晚，美光市场份额从 23Q1 的 27% 回落至 23Q4 的 19%。

图表63：DRAM 市场竞争格局演变



资料来源：Bloomberg Intelligence、TrendForce、华泰研究

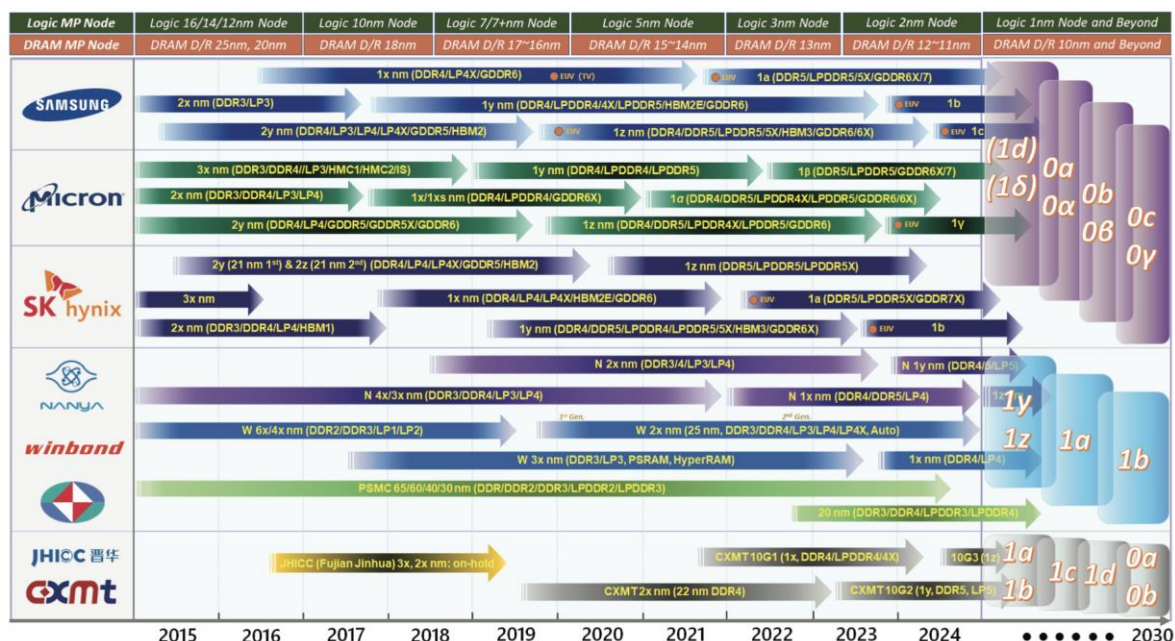
我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中详细分析了三大厂商的先进制程爬坡进度和产品矩阵：

制程上：1) 美光：相较海力士 23 年 5 月导入 1-β 制程，美光自 2021 和 2022 年分别率先量产 1-α 和 1-β 制程 DRAM，取得制程领先，历经三年，DRAM 产能已基本完成转换，计划 25 年量产 1-γ 制程的 DRAM；**2) 海力士：**如我们前文提到，海力士规划 24 年于利川 M16、M14 工厂扩充 1-α 和 1-β 的产能，并将从 1-α DRAM 开始着手导入 EUV 制程，计划在 2024 年底将 1-α 和 1-β 产量增加到 DRAM 总产量的一半以上；**3) 三星：**受自 1z nm 开始过早导入 EUV 拖累，丢失 DRAM 制程优势，时至今日，尽管业界已先后采用了 EUV 制造 DRAM，但三星在密度和性能竞赛中仍然落后，三星 24 年同样着眼先进制程产能爬坡。

1-γ 制程成“兵家必争”，大规模 EUV 导入或重塑格局。1) 海力士：根据 Digitimes 24 年 4 月报道，海力士内部 DRAM 技术路线规划 24Q3 开始试产 1-γ，并随后向英特尔等客户验证产品效能，报道指出效能认证通常持续 1-2 个月，完成之后可着手进行量产，海力士正通过利川厂安装 6-7 台 EUV 设备，将在日后服务于 1-γ 量产。同时，韩媒 The Elec 24 年 5 月报道海力士计划在 1-γ 制程上引入 Inpria MOR 光刻胶，其 1-γ DRAM 将有 5 层 EUV 曝光，其中一层采用 MOR 技术绘制，有助于提升 DRAM 性能并降低成本；**2) 美光：**将领先竞争对手率先转向 1-γ 制程，将在 1-γ 制程中开始导入 EUV，截至美光 FY24Q2 业绩会，1-γ 制程已开始在台中试产，并将于 25H1 先行在台中厂量产；**3) 三星：**根据 Digitimes 24 年 4 月报道，三星在 Memcon 2024 上公布 DRAM 量产路线图，根据计划将在 24 年底跟进 1-γ 节点。报道指出鉴于三星在 EUV 上导入更早，因此产线上估计已使用至少 30 余台 EUV 设备。

我们认为，EUV 导入 1-γ 制程对于三家厂商均是风险和机遇并存，量产时间和产能均值得关注。三星和海力士导入 EUV 较美光更早，前期积累经验或有助于两家在转向 EUV 生产 1-γ 过程中更加顺畅，然而鉴于二者 24 年仍将着重于进行 1-α 和 1-β 产能爬坡，1-γ 试产成功后何时推出 1-γ DRAM 产品并顺利进行产能爬坡值得关注；同时，美光采用 DUV 的 1-α 和 1-β 制程已经成熟，但 EUV 导入 1-γ 仍是首次，能否攻克工程问题，实现良率和产能爬坡将是其工艺优势维持关键。

图表64：DRAM 主要玩家的制程和产品路径（2022 年 6 月发布），三大厂商 DRAM 产品迭代咬合紧密



资料来源：TechInsights、华泰研究

在 GDDR 和 LPDDR 方面，三大厂商同样进度相仿：1) GDDR：与海力士规划相同，美光和三星也将于 24H1 发布并量产 GDDR7，向英伟达 RTX50 系显卡供货。Tom's Hardware 24 年 4 月预测 AMD 和英伟达将于 24 年底或 25 年初发布下一代 GPU，考虑到“黑五”和圣诞节为传统购物旺季，英伟达在年底购物季前后发布“RTX 50 系列”显卡的可能性较大。我们认为关于搭载 GDDR7 的显卡将于 24H2 有更多进展信息。2) LPDDR：美光 23 年量产的 LPDDR5X 已率先采用 1-β 制程，提供与海力士 LPDDR5T 相同的 9.6 Gbps 带宽，同时，24 年 4 月 23 日三星宣布 10.7 Gbps LPDDR5X 研发成功。根据 Tom's Hardware 24 年 1 月报道，该 LPDDR5X 将为 Intel Lunar Lake MX 平台供货，搭载于轻薄型笔记本。而该款 10.7 Gbps LPDDR5X 相较前代产品效能与容量分别提高 25%及 30%以上，有望在 AI PC 和 AI 加速器中得到应用。

图表65：DRAM 三大厂商 SWOT 分析

	SK 海力士	三星	美光
优势 Strengths	1. 相较三星，EUV 导入循序渐进，相较美光 EUV 技术积累更加丰富，1-γ 制程有望 24Q3 取得进展 2. 制程优势，DRAM 密度和成本结构领先三星	1. 资金雄厚，产能优势坐稳 DRAM 龙头 2. 集存储、芯片设计、晶圆代工为一体，10.7 Gbps LPDDR5X 率先研发凸显技术积累 3. EUV 导入经验丰富，1-γ 制程转换有望更平稳	1. 1-α 和 1-β 制程成熟，覆盖 3/4 DRAM 产能 2. 率先转向 1-γ，工艺有望维持领先
劣势 Weakness	1-β 节点尚在爬坡，资金限制先进产能扩张或不及三星	1. 制程竞争落后，DRAM 密度和成本结构不及海力士和美光 2. 1-α 和 1-β 爬坡尚在进行时，或影响 1-γ 节点量产节奏	1. 产能较三星、海力士更小，HBM 扩产进一步挤占 DRAM 产能 2. EUV 导入经验不及海力士和三星，1-γ 节点转换和良率面临风险
机会 Opportunities	1. AI 手机、AI PC 和 AI 服务器等新需求带来以 DDR5、LPDDR5X 和 LPCAMM2 为主导的 DRAM 市场增长新动能 2. 新一轮通用服务器布局叠加换机潮，传统设备需求或在 24H2 至 25 年明朗 3. 三大厂商各产品线布局完善，AI 需求风口、传统需求回暖机遇两不误		
威胁 Threats	1. EUV 导入 1-γ 节点，或影响 DRAM 竞争格局 2. 短期增长依赖 AI 驱动，传统设备需求前景仍有不确定性 3. 传统 DRAM 高度标准化，抗周期能力较弱，24Q2 起三大厂商逐步扩大投产量，25 年 DRAM 或再现供过于求		

资料来源：华泰研究

NAND：AI 如何惠及海力士大容量 NAND？

我们预计海力士 NAND 业务 FY24/25/26 营收为 21.7/32.9/36.6 万亿韩元，对应营收同比为 123.1%/51.9%/11.1%。疫情以来高基数以及下游 NAND 库存已基本在历史平均水平附近，我们认为 AI 端侧设备渗透率提高带动 24 年 NAND 复苏有望。结合公司 FY24Q1 业绩会，我们预计海力士 FY24 NAND 整体供给策略维持保守，着力于进一步优化库存，缩短周转天数，关注伴随 AI 设备渗透率提高带来的高层 QLC NAND 和大容量 eSSD 增长机会。

展望 24 和 25 年，我们认为 NAND 增长核心在于下游设备需求拉动：

1) 用户端 SSD (PC 为主)：根据 Digitimes 24 年 3 月报道，23 年 NAND ASP 大幅下降带来大容量 SSD 导入 PC 的“价格甜点”，PCIe 4.0+1TB 成为 PC 市场的主流配置。同时，伴随 24H1 PC OEM/ODM 库存回到正常水平、25 年微软停止支持 Win10，以及 Win12 带动换机潮流，叠加 AI PC 逐渐放量，24H2 将为 PC SSD 带来增量。

值得注意的是，Digitimes 24 年 4 月 26 日报道提及，高层 QLC NAND 在 PC 中渗透率不断提高，报道援引微软 AI PC 观点和业界对 QLC NAND 看法，指出 AI PC 搭载 LLM 和图片/影片生成模型，将耗费大量存储容量，同时，由于用户不会立即删除生成的大量图片和视频，带来 AI PC 所需存储空间显著提升，而伴随主流 NAND 厂商 NAND 堆叠层数均超过 200 层，为了持续提高密度及降低存储成本，QLC NAND 发展成为趋势，目前 PC OEM 采用 QLC NAND 比重约为 20%-25%，24 年渗透率将随着 PC OEM 加速导入 QLC NAND 而高速增长。

2) eSSD (企业级，服务器为主)：与 DRAM 逻辑类似，AI 服务器对高规格设备硬件需求以及通用服务器换机潮将带动大容量 SSD 需求。根据 Digitimes 24 年 3 月报道，24 年云端服务器采用 PCIe 5.0 SSD 的渗透率将较 23 年翻倍，更多 8TB/16TB 或以上的 SSD 应用需求将增加。同时，鉴于服务器成本控制，企业客户对 NAND 需求在于更高密度且成本不变(HDD/SSD 等存储设备约占通用服务器成本的 15%)，因此更高层数 NAND 将更受青睐；

3) 手机 SSD：根据 Digitimes 24 年 3 月报道，24 年高阶手机将进入 512GB 以及 TB 时代，预计全年手机平均容量将超过 200GB。同时，业界预期 24 年 QLC NAND 将导入大容量存储的手机新品，随着影像存储需求增加及 QLC NAND 更低的单位存储成本，iPhone 及部分国产手机品牌（如 Oppo）已进入大容量 1TB 机种采用 QLC NAND 的产品验证阶段，并可能增加存储容量作为营销卖点。

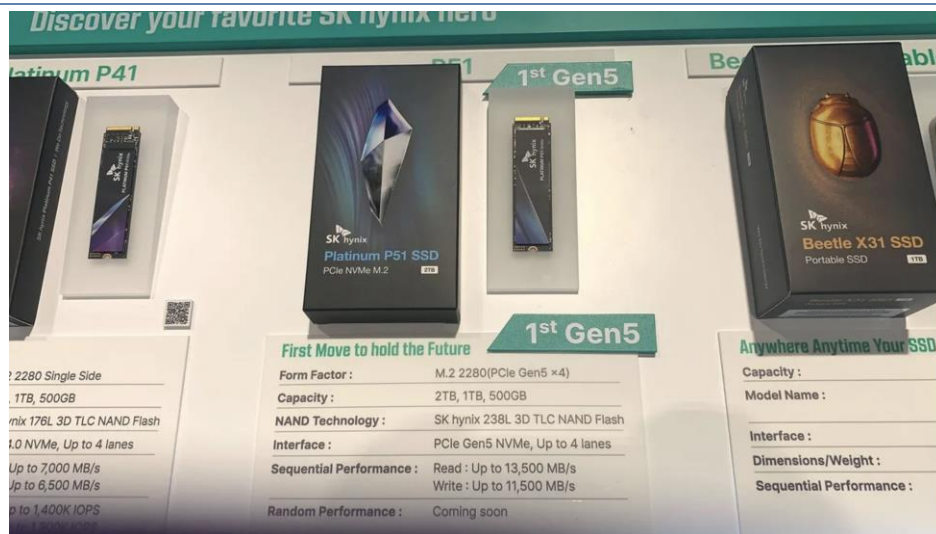
我们认为 24-25 年海力士和其子公司 Solidigm 分工合作，海力士 176 层和 238 层 TLC NAND 持续布局 AI 手机、AI PC 和服务端，更加着重于用户端，而 Solidigm 则借助大容量 QLC NAND 产品线消化 AI 设备和数据中心兴起的“更低成本+更高密度”存储需求：**1) QLC NAND：**Solidigm 专注于 QLC NAND 产品线，23Q4 QLC NAND 出货量占其 NAND 总出货的 60%，D5 系列 eSSD 基于其 192 层 QLC NAND，向数据中心场景，提供 3TB-60TB 的多样化存储方案；**2) TLC NAND：**海力士主流 NAND 技术节点为 176 层 TLC NAND (4D V7 TLC)，基于此，海力士通过 PE9000 系列和 PC801/BC901 分别面向数据中心和 AI PC 场景，并通过 UFS 4.0 布局大容量手机，同时关注 24-25 年海力士 238 层 TLC NAND 产能爬坡和相关产品逐步导入市场。24 年 6 月 COMPUTEX 上海力士展示 PCIe 5.0 Platinum P51 SSD，基于 238 层 3D TLC NAND，各项效能约为前一代 P41 的 2 倍，提供 13500 MB/s 的循序读取速度及 11500 MB/s 的循序写入速度，预期 24H2 开始在韩国销售，并于 25 年向全球发售。FY24Q1 海力士 NAND 营收为 4.4 万亿韩元，同比+158%，环比+33%，主要系 NAND ASP 回暖，环比增长逾 30%。

图66：海力士 PS1010/PS1030 eSSD：176 层 TLC NAND，支持 PCIe Gen 5



资料来源：海力士官网、华泰研究

图67：海力士 GTC 2024 展出 Platinum P51 SSD：238 层 TLC NAND，支持 PCIe Gen 5



资料来源：Tom's Hardware、华泰研究

NAND ASP 23Q3 见底回升，24 年 ASP 增长趋平稳

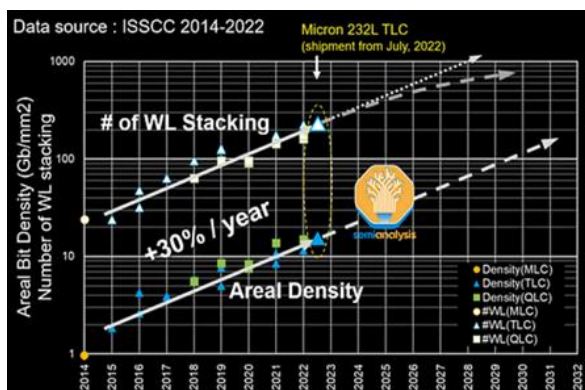
我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中复盘了 NAND 与 DRAM 区别所在，并对 TLC 和 QLC 等 3D NAND 分类进行介绍，同时回顾 23Q2 以来 NAND ASP 和产能恢复走势。

整体而言，NAND 作为外部存储器保存设备运算所需的海量数据，同时，随着每个单元存储的位数增加（如 QLC 相比 TLC），存储密度提高，位元成本降低，但读写速度和耐用性会相应减少。

23 年半导体供给侧厂商相继加速减产、去库存，以稳定下游产品定价，23Q3 以来 NAND ASP 增势显著。自 22Q4 起，海力士和美光主动减产，调整库存；三星从 23Q2 加入减产阵营，将其 NAND Flash 减产幅度扩大至总产能的 50%，集中减产 128 层以下的中低端产品。积极的产能控制后，NAND ASP 于 23Q3 见底回升，24Q1 合同价环比增长逾 20%。当前价格已趋近 20 年平均水平，理论上供应商已有获利空间，距离价格高点已不远。

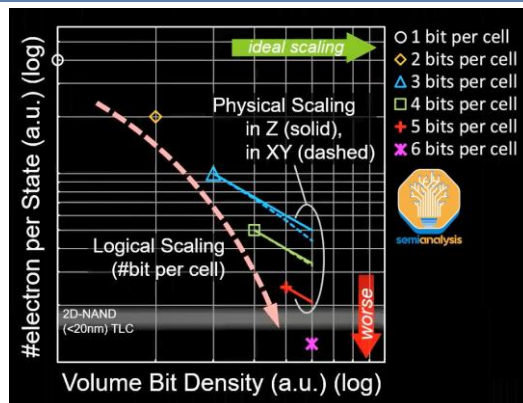
我们预计 24Q2 价格上涨或仍将持续，但伴随 QLC NAND 导入，24H2 持续上涨空间或有限，预期 NAND ASP 趋势渐趋平稳。根据 Digitimes 24 年 3 月和 4 月报道，24Q1 终端需求持续增温，手机和 PC 端 NAND 需求领涨，部分厂商调升产能，增加出货量以弥补 23 年 NAND 亏损，24Q2 需求状况或仍会保持，但涨幅将会开始收敛，NAND 市场完全恢复可能在 2025 年。伴随市场需求消化上游新增产能，以及下游市场在 24H1 陆续补库完成，24H2 NAND ASP 增长或回到低个位数幅度。

图表68：3D NAND 推出以来，密度每年以稳定的速度提高 30%



资料来源：Semianalysis、华泰研究

图表69：SLC 向 QLC 等转变提升每单元存储位数，但牺牲了可靠性



资料来源：Semianalysis、华泰研究

图表70：22 年底至 23 年半导体厂商 NAND 相继减产，24Q1 开始逐步回调产能

公司	减产计划
美光	22 年 11 月宣布减产 20%，23 年 3 月宣布 NAND 的投产量减少至 30%，24 年 NAND 整体维持低产，主要产能分配给高阶 NAND 产品（176 层和 232 层）
SK 海力士	22 年宣布 23 年资本开支减半，23 年 7 月宣布进一步降低 NAND 产能，24 年维持整体减产策略，各季度稼动率维持 50%-60%，针对性增加 eSSD 产能
三星	23Q2 减少 NAND FLASH 产能约 25-30%，23H2 减产幅度进一步扩大，24Q1 西安 NAND 厂产能利用率恢复至 70%，24 年 4 月整体 NAND 稼动率已达 90%
铠侠	22 年 10 月宣布减产 30% NAND 产能，24Q1 调升 NAND 产能
西部数据	23 年 1 月起宣布减产 30% 晶圆投产量，23 年资本开支预计减少 25%，24Q1 调升 NAND 产能

资料来源：各公司公告、Digitimes、华泰研究

图表71：24Q2 NAND 合约价环比涨幅将收敛至 13%-18%

NAND	2023Q4	2024Q1	2024Q2
智能手机：UFS/eMMC	涨价 10-15%	涨价 25-30%	涨价 10-15%
数据中心：企业级 SSD	涨价 10-15%	涨价 23-28%	涨价 20-25%
PC：用户端 SSD	涨价 13-18%	涨价 23-28%	涨价 10-15%
3D NAND 晶圆：TLC 和 QLC	涨价 35-40%	涨价 23-28%	涨价 5-10%
总 NAND	涨价 13-18%	涨价 23-28%	涨价 13-18%

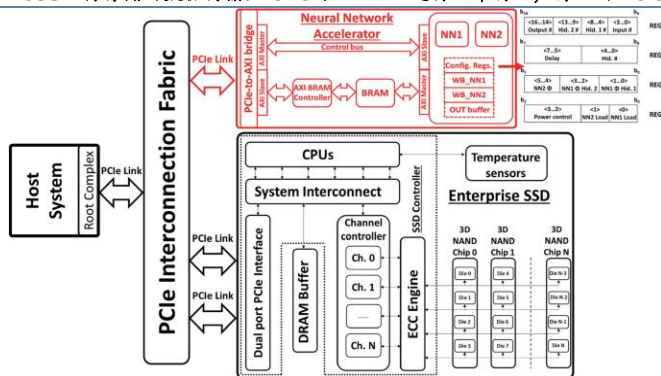
资料来源：Trendforce、华泰研究

HBM 和 DDR5 之外，AI 将如何惠及 NAND Flash？

我们认为，NAND 在 AI 训练和 AI 推理中角色类似，从数据传输路径来看，不同于逻辑芯片的“计算核心”和内存的“中转站”作用，NAND 承担的主要角色是存储训练和推理所需的大量数据，因此要求在于以更低成本实现“更大容量+更高耐用+更快读取速度”。同时，与 DDR5 受益于 AI 换机潮逻辑类似，AI 将带动 PC、手机及服务器的需求增加，驱动 NAND 供应商提供更大容量、更高速的产品。

NAND 对芯片性能影响可从 iPad Pro 的性能表现略见一斑。根据 Digitimes 24 年 5 月报道，256GB、13 英寸的 iPad Pro 仅搭载 1 个 NAND 晶片，而 1TB 版本的表现比 256GB 表现更好。同时，因为配备较少 NAND，搭载苹果 M2 芯片的 MacBook 的 SSD 读写能力曾一度逊于搭载 M1 芯片的版本。

24 年 3 月 21 日举办的 CFMS & Memory 2024 论坛上，针对 SSD “更快读取速度”，三大厂商均有新品推出：1) 三星：推出 4 通道 UFS 4.0，相较于传统 2 通道 UFS 4.0 有望实现读写速度翻倍（从约 4GB/s 提升至约 8GB/s），以适配 AI 手机本地运行 LLM 的数据读取需求，预期 24 年夏天出货工程样品；2) 海力士：宣布 PCIe 5.0 eSSD 23 年完成客户认证，当前准备进入量产和出货阶段，预期 24 年客户数量持续提升；3) 美光：透露近期将同样发布 PCIe 5.0 SSD，指出 SSD 升级可有效缩短 AI 训练的系统性延迟，美光 PCIe 5.0 SSD 通过加速技术可将部分模型训练时间缩短 50%。

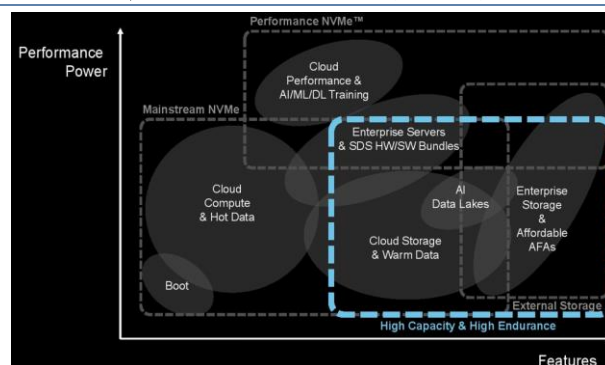
图表72：加速器举例：eSSD 将存储的数据传输至 CPU 和 DRAM 进行“中转”，再经由 PCIe 总线至逻辑芯片


资料来源：Cristian Zambelli (2022.5):《Machine Learning for 3D NAND Flash and Solid State Drives Reliability/Performance Optimization》、华泰研究

图表73：PCIe 迭代带来传输速率的翻倍提升

PCI Express	Bandwidth	
Version	×4	×16
1.0	1.0 GB/s	4.0 GB/s
2.0	2.0 GB/s	8.0 GB/s
3.0	3.94 GB/s	15.8 GB/s
4.0	7.88 GB/s	31.5 GB/s
5.0	15.75 GB/s	63.0 GB/s
6.0	31.51 GB/s	126.03 GB/s

资料来源：MSI 官网、华泰研究

图表74：数据中心 eSSD 的性能关注点：高容量和高耐用性


资料来源：美光官网、Tom's Hardware、华泰研究

图表75：UFS 迭代带来传输速率的翻倍提升

2013	2014	2016	2019	2020	2022
Industry's 1st release*	Industry's 1st release*	Started mass production	Industry's 1st release*	Released in the market	Started sample shipments
UFS 1.1	UFS 2.0	UFS 2.1	UFS 3.0	UFS 3.1	UFS 4.0
290 MB/s	1,160 MB/s	1,160 MB/s	2,320 MB/s	2,320 MB/s	4,640 MB/s
* KIOXIA Corporation first sample shipment, as of February 8, 2013	* KIOXIA Corporation first sample shipment, as of April 25, 2014		* KIOXIA Corporation first sample shipment, as of January 23, 2019		

资料来源：Kioxia 官网、华泰研究

根据 Digitimes 24 年 3 月援引铠侠观点，与从 DRAM 衍生的 HBM 不同，Gen AI 并不直接对 NAND 有设计上的特殊要求，而是 AI 将带动 PC、手机及服务器的需求增加，驱动 NAND 供应商提供更大容量、更高速的产品，并以更低成本满足客户需求。AI 服务器场景下，DRAM 与 SSD 同步搭配方能实现系统层面的低延迟，因此 Gen AI 浪潮将同样带来 SSD 领域的投资意愿提升。23 年 NAND 和 SSD 价格下跌使得其与传统 HDD (Hard Disk Drive 机械硬盘) 价差缩小，有助于市场从 HDD 转移至 SSD，尤其是低成本、高容量的 QLC SSD 需求迅速增长。对比 HDD 和 SSD，鉴于 HDD 读写延迟较 SSD 均更高，服务器“换芯不换存”相当于浪费逻辑芯片和 DRAM 性能，而 SSD 有助减少逻辑芯片和 DRAM 的性能浪费。

协同 Solidigm，海力士 NAND 用户端和数据中心双管齐下

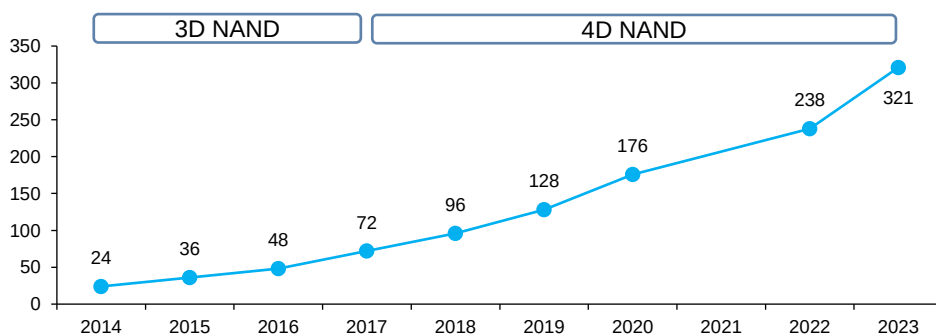
TLC 和 QLC 是当前面向数据中心高端 SSD 市场的主流方案，TLC 和 QLC 之间的取舍尚有争议，但共同目的在于提高存储密度同时降低每 Gb 成本，兼顾提升读写性能，伴随技术迭代，TLC 和 QLC 壁垒逐渐模糊。

海力士自身的 NAND 产能以 TLC 为主，复盘其制程精进过程，公司在 2018 年 11 月首次将 4D NAND 技术运用在 NAND 闪存，推出全球首款 96 层 512Gb 4D NAND 闪存。与 72 层 512Gb 3D NAND 相比，4D NAND 芯片尺寸减少 30% 以上。随后，海力士在 2019、2020 和 2022 年陆续开发出 128、176 和 238 层 NAND，数据传输速度伴随迭代提高。2023 年 8 月，公司在 2023 闪存峰会 (FMS) 上展示全球首款超 300 层的 NAND 产品，321 层 1Tb TLC 4D NAND，并计划于 2025H1 开始量产。

海力士主流 NAND 技术节点为 176 层 TLC NAND (4D V7 TLC)，基于此，海力士通过 PE9000 系列和 PC801/BC901 分别面向数据中心和 AI PC 场景，并通过 UFS 4.0 布局大容量手机。虽然技术研发进展迅速，但高层 NAND 产能爬坡并非一蹴而就。23 年 6 月海力士宣布开始量产 238 层 4D NAND，但截至 24H1 尚未大规模投片，23 年 11 月 238 层 NAND 投片量仅 5000 片 (Digitimes 24 年 4 月报道透露海力士 NAND 50%-60% 稼动率时月产能约 20 万片)，公司计划首先向智能手机供应该款 NAND，随后扩展至基于 PCIe 5.0 的 PC SSD 和数据中心级 eSSD 等产品，相关产品 (如前文提到 PCIe 5.0 Platinum P51 SSD) 将于近期及 24H2 逐渐流入市场。

24 年 5 月 9 日海力士宣布在现有 UFS 4.0 产品基础上面向 AI 手机开发出 ZUFS 4.0，预期 24Q3 开始量产。2022 年 8 月海力士宣布开始基于 238 层 TLC NAND 开发 UFS4.0，尽管本次 ZUFS 4.0 工艺节点尚未确认，但相较于现有的 UFS 产品，海力士 ZUFS 根据数据各自特性 (用途和使用频率等) 对智能手机生成的数据进行区管理，从而提高手机的运行速度和存储设备的数据管理效率。根据公司官网新闻，在长期使用环境下，使用 ZUFS 的手机应用程序加载时间相较现有 UFS 缩短 45%，同时延长 40% 的产品寿命。

图表76：SK 海力士 NAND 研发时间图 (单位：层)



资料来源：SK 海力士公告、华泰研究

图表77：SK 海力士 5 代 4D NAND 参数对比

	第一代	第二代	第三代	第四代	第五代
发布时间	2018 年 11 月	2019 年 6 月	2020 年 12 月	2022 年 8 月	2023 年 8 月
堆叠层数	96 层	128 层	176 层	238 层	321 层
存储容量	512Gb	1Tb	512Gb	512Gb	1Tb
单元层数	TLC	TLC	TLC	TLC	TLC
传输速度	1200Mbps	1400Mbps	1.6Gbps	2.4Gbps	-

资料来源：SK 海力士公告、华泰研究

图表78: 海力士 eSSD 代表产品

产品型号	SE5110/5130	PE8110/PE8130	PE8111	PE9110	PE9010	PS1010/PS1030
技术节点	128 层 TLC	128 层 TLC	128 层 TLC	128 层 TLC	176 层 TLC	176 层 TLC
存储容量	480GB-3.84TB	800GB-7.68TB	15.36TB	1.92-7.68TB	960GB-3.84TB	1.6-15.36TB
传输接口	SATA3	PCIe Gen4x4	PCIe Gen3x4	PCIe Gen4x4	PCIe Gen3x4	PCIe Gen5x4
平均失效时间 (mn hrs)	2	2	2.5	2.5	2.5	2.5
循序读取速率 (MB/s)	530/555	6500	3400	6600	3670	14500
循序写入速率 (MB/s)	490/530	4200	3000	6300	2200	9300
随机读取速率 (K IOPS)	83/96	1100	700	1250	800	2750
随机写入速率 (K IOPS)	39/84	400	100	185	83	360/800

资料来源: SK 海力士产品公告、华泰研究

图表79: 海力士面向用户端的 PC801 和 BC901, 基于 176 层 TLC NAND

PC801/BC901 			
PC801 Model Name	M.2 2280 Form Factor	PCIe Gen4x4 Interface	512GB-2TB Capacity
BC901 Model Name	M.2 2230/42/80 Form Factor	PCIe Gen4x4 Interface	256GB-1TB Capacity
Application  PC  Server - Best-in-class performance as PCIe Gen4 NVMe Client SSD - Faster and larger storage for various applications in changing PC client environment			

资料来源: 海力士官网、华泰研究

图表80: 海力士 UFS 4.0 和 UFS 3.1 参数 (172 层 TLC NAND)

UFS4.0		UFS3.1	
Interface	UFS4.0	Interface	UFS3.1
NAND Tech	V7 512Gb	NAND Tech	V7 512Gb
Capacity	256GB - 1TB	Capacity	128GB - 1TB
PKG size	11x13	PKG size	11x13
Applications  Mobile Phone  Tablet/Laptop - The first in-house P Controller - Support Max Bandwidth (~4300MB/s) - Competitiveness of Performance Sustainability & Latency		Applications  Mobile Phone  Tablet/Laptop - Provide Best-in-class UFS3.1 Performance - Competitiveness of Performance Sustainability & Latency	

资料来源: 海力士产品官网、华泰研究

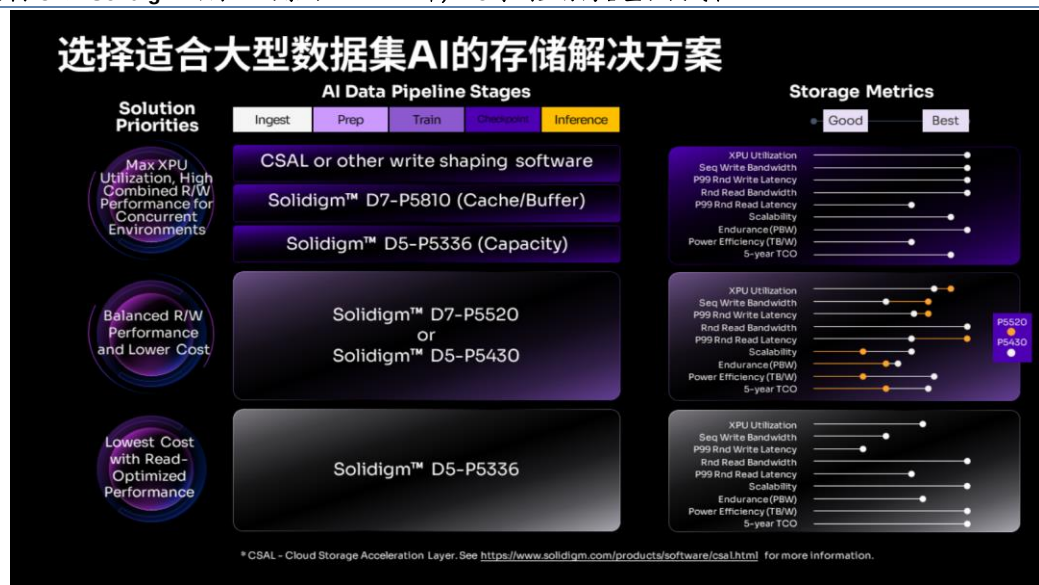
图表81: 海力士 ZUFS 4.0 样品图



资料来源: 海力士产品官网、华泰研究

SK 海力士于 2020 年 10 月 20 日宣布收购英特尔的 NAND 闪存及存储业务, 并在美国新设立子公司 Solidigm 用于运营英特尔 SSD 业务。Solidigm 专注于 QLC NAND 闪存产品线, 23Q4 QLC NAND 出货量占其 NAND 总出货的 60%, 面向 QLC NAND 导入端侧设备趋势有望与海力士 TLC NAND 产生协同效应。其中, D7 系列 eSSD 基于 144 层 SLC 和 TLC NAND, 侧重面对 AI 场景的高速传输, 产品多为 800GB-15TB 的中小型容量。而 D5 系列 eSSD 基于其 192 层 QLC NAND, 主打大容量、低成本, 面向数据中心场景, 提供 3TB-60TB 的多样化存储方。根据 Solidigm 测试, 以 10PB 规模的 AI 数据流存储为例, HDD 方案采用 24TB, 70%利用率, SSD 方案采用 D5-P5336, 61.44TB, 95%利用率, 虽然单 GB 价格仍有数倍差异, 但整体方案 SSD 可实现 46%的 TCO 节省。我们预计海力士将借助 Solidigm 大容量 QLC NAND 产品线消化数据中心和 AI 设备兴起的“更低成本+更高密度”存储需求。

图表82: Solidigm 面向 AI 场景的 NAND 矩阵, D5 系列主打高容量、低成本

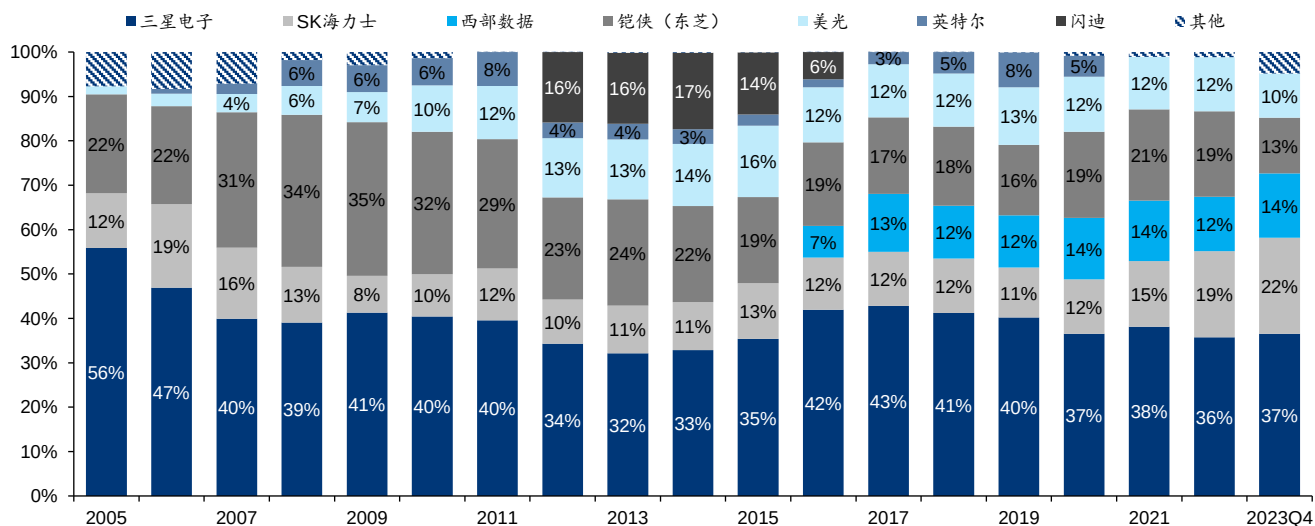


资料来源: 海力士官网、华泰研究

竞争格局: 技术趋同, 厂商份额差异主要鉴于产能规模

相较于 14 年以来“三足鼎立”的 DRAM 市场, NAND 市场更为分散, 23Q3 CR3 为 68%, 对比 2021 年和 2022 年的 74% 有所下降, 23Q4 CR5 约为 95%, 行业竞争更加激烈。2016 年, 闪迪 (SanDisk) 通过与东芝合资加快 NAND 布局, 而后在 2016 年被西部数据收购。至此, 五大 NAND 闪存供应商格局形成, 分别为三星电子、东芝、西部数据、SK 海力士和美光, 2017 年分别占市场规模的 43%/17%/13%/12%/12%。2022 年, SK 海力士通过收购英特尔 NAND 业务完成对铠侠和美光的赶超, 成为仅次于三星电子的 NAND 闪存制造商。

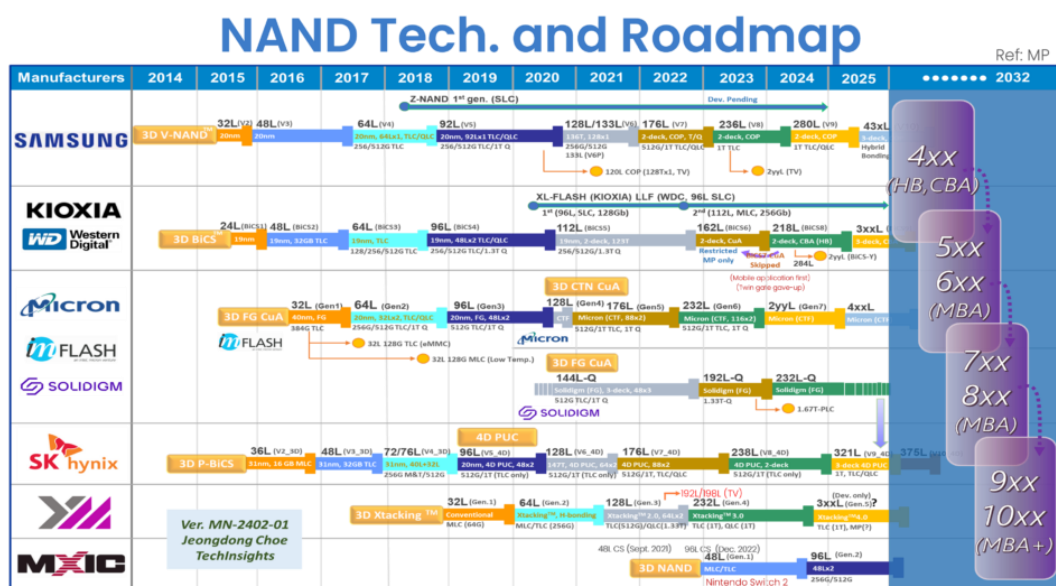
图表83: NAND 闪存市场竞争格局演变



资料来源: Bloomberg Intelligence、TrendForce、华泰研究

复盘 NAND 主流厂商技术路线图，NAND 堆叠层数的提升带来更高的存储密度，同时亦提升了工艺难度，各大厂商技术进度咬合紧密，2023 年至今推出的 3D NAND 层数均为 218 层及以上。以三大厂商为例：1) 三星：根据 Digitimes 24 年 4 月和 5 月报道，三星于 4 月开始量产 290 层 1Tb TLC NAND (v9)，据报道，与三星上一代 236 层 TLC NAND (v8) 相比，存储密度提升 50%，功耗改善 10%，但三星产能仍主要是 128 层 NAND，23-24 年通过投入资金将产能向 236 层倾斜（跳过 176 层节点），同时为迎合 AI 端侧设备渗透带来的 QLC NAND 需求，三星计划 24H2 推出相应 QLC NAND 新品；2) 海力士：海力士 23 年 8 月公开 321 层 TLC NAND 样品，预计 25H1 实现量产；3) 美光：TechInsights 24 年 3 月报道透露，美光在 232 层节点后着手开发下一代“近 300 层”节点，该节点类似三星 280 层 NAND，同时，25 年美光或跳过 300 层直接公布 400 层技术节点开发情况。

图表84：NAND 主要生产厂商技术路线图 2014-2024Q1，3D NAND 层高增长迅速，头部厂商技术迭代基本同步



*注：图中节点为量产时间

资料来源：TechInsights、华泰研究

我们在 2024 年 5 月 14 日的《美光科技：HBM 自我突破在 AI 浪潮中引领重估》的报告中分析并指出：头部厂商技术迭代基本同步，结合千禧年以来主流厂商的并购时间和产能信息，我们认为厂商间 NAND 技术发展和成本差异较小，而市场份额差异主要鉴于产能规模：譬如，2021 年之前，美光 NAND 市场份额与海力士相近，21 年 12 月海力士完成收购英特尔 NAND 及 SSD 业务第一阶段后，22 年市场份额与美光拉开差距。同时，西部数据和铠侠 24 年初曾希望重启 NAND 合并计划，但报道同时援引市场观点，认为海力士仍将维持反对。报道指出，铠侠和西部数据合并后 NAND 产能和市占率将超越海力士，进而影响海力士在本轮 NAND 上升周期凭借市占率优势的获利空间，因此海力士作为铠侠 15% 股权的间接投资方，将维持反对态度，NAND 产能对于市场份额重要性也可由此略见一斑。

展望未来几个季度，我们认为 NAND 主要厂商的市场份额将维持稳定，24Q1 各厂商均有部分产能调升，但整体投产仍维持审慎，具体波动将根据 24H2 主要厂商扩产计划和下游需求恢复情况。

图表85: NAND 三大厂商 SWOT 分析

	SK 海力士	美光	三星
优势 Strengths	1. 2022 年收购英特尔 NAND 业务完成对铠侠和美光的产能赶超 2. 子公司 Solidigm 专注 QLC NAND, 有望受益于 QLC NAND 导入 AI PC 和 AI 手机趋势	1. 200 层以上 TLC 和 QLC NAND 量产时程取得领先, 产能爬坡先于竞争对手 2. QLC NAND 领域起步较竞争对手更早, 有望受益于 QLC NAND 导入 AI PC 和 AI 手机趋势	1. 资金雄厚, 产能优势坐稳 NAND 龙头 2. 集存储和消费电子产业于一体, 上下游供需衔接紧密, NAND 扩产产能自身可消化
劣势 Weakness	产能较三星更小, 同时并非自带消费电子下游需求, 投产增产相较三星需更加审慎		NAND 工艺爬坡落后, 高层 NAND 产能替换速度不及海力士和美光
机会 Opportunities	1. 24-25 年 AI 设备需求增长及通用设备需求回暖驱动 NAND“供需两旺” 2. 三大厂商均已积极布局高层 NAND, 面向更高密度、更低成本的 NAND 需求趋势均有望受益		
威胁 Threats	1. 高层 NAND 工艺爬坡、量产和产能爬坡均耗时更久, 厂商间技术差距逐渐缩小, 竞争日趋激烈 2. 部分中企厂商技术实力不逊传统龙头, 低成本 QLC NAND 冲击市场		

资料来源: 华泰研究

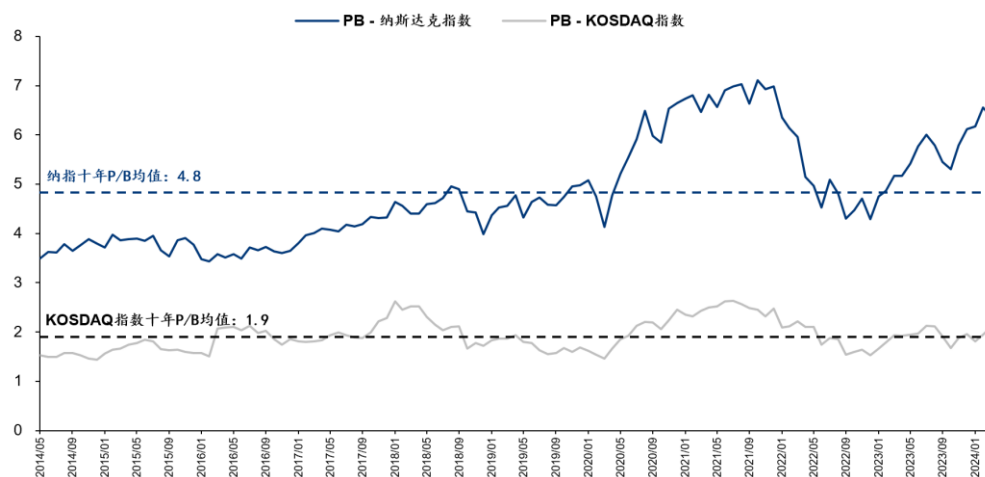
盈利预测与估值：HBM 龙头地位有望保持，目标价 30 万韩元

2014 年-2022 年海力士估值经历三次峰值，分别受益于移动互联网普及（2014 年，1.75x Forward PB，简称为 FPB）、上云趋势带来的服务器需求和加密货币潮（2017-2018 年，1.4x FPB）以及疫情居家办公带来的短期需求激增（2021 年，1.5x FPB）。而 23 年至今海力士估值逆势高歌猛进，从 23 年 1 月 1 日的 0.8x FPB 跃升至 24 年 6 月 21 日的 1.9x FPB，主要系海力士 HBM3 于 22 年 6 月率先实现量产后独供英伟达 H100，并于 23 年 4 月和 6 月先后发布 12 层 HBM3 和 8 层 HBM3E，坐稳 HBM 头把交椅，背靠高速增长的英伟达 GPU 需求通过高价值 HBM 实现利润率和估值的双重提升。

截至 2024 年 6 月 21 日海力士估值为 1.9x FPB，逼近 14 年历史峰值。我们认为市场或完全反应存储行业上行周期的持续性，以及对海力士面向美光和三星，在 HBM 市场争夺中的优势尚存疑虑。展望 2025，我们认为海力士估值的持续突破或将是双轮驱动：1) 相较美光和三星，海力士 HBM “先发+技术”形成护城河，除了背靠英伟达供应链带来的稳固 HBM 需求，其成熟的产能和产线管理、HBM3 世代已有的 8 至 12 层爬坡经验和采用 Advanced MR-MUF 键合技术带来的高良率和散热性能，帮助海力士 HBM 取得竞争优势。伴随 M15X 等新 DRAM 生产基地落地，其 HBM 扩产后有望拓展客户至其他 AI 芯片，需求能见度提高，有望进一步提升营收和利润率；2) 存储行业步入新一轮上行周期，而本轮以 AI 驱动的上行周期的持续性与 18 年（云计算数据中心搭建和虚拟货币潮）、21 年（疫情期间在家办公）两轮周期不同，内存在 AI 设备里起到的重要作用支撑传统 DRAM 和 NAND 的需求增长，带动 ASP 修复，或帮助海力士营收基本盘“锦上添花”。因此，我们认为海力士的中长期增长，跟 2014 年的逻辑类似，并非如同加密货币潮和疫情带来的短期需求激增，借助 AI 和 HBM 东风，海力士估值或可往 2014 年水平靠拢，甚至有望突破。我们认为 FY25 海力士的 PB 估值约为 2.5 倍较合理，FY25 BVPS 12.1 万韩元，对应目标价为 30 万韩元。

我们也认为，作为市场中仅三家的 HBM 供应商之一，海力士业务与美光科技（MU US）和三星电子（005930 KS）业务重叠度较高。但其中，美光与海力士所处市场不同，对比过往 10 年纳斯达克综合指数和韩国 KOSDAQ P/B 走势，纳指的 P/B 均值约为 KOSDAQ 的 2.5 倍。截至 24 年 6 月 21 日，美光和三星 25 年彭博一致预期 PB 为 2.8x 和 1.3x，而在 HBM 方面，海力士相较美光在产能、客户基础和键合技术领域均有优势，而三星电子在存储设备业务之外，还兼有消费电子（显示屏、PC 和手机等），业务不纯粹，考虑到海力士作为行业龙头具有显著壁垒和高成长性，以及高利润率的 HBM 占比不断上升，我们认为应享受估值溢价。

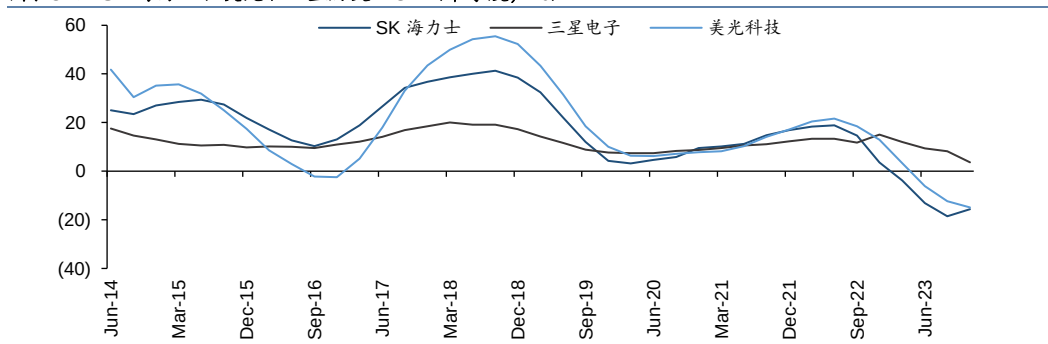
图表86：十年 PB 对比：纳斯达克指数和韩国 KOSDAQ 指数，纳指 P/B 估值享有溢价



注：韩国 KOSDAQ 指数是隶属韩国交易所的创业板市场，与纳斯达克类似，主要面向 IT 及高新技术股。

资料来源：Bloomberg、华泰研究

图表87: SK 海力士、美光和三星历史 ROE (单季度, %)



资料来源: Bloomberg、华泰研究

图表88: SK 海力士、美光和三星估值 (截至 2024 年 6 月 21 日)

		PB			PS		
		2024E	2025E	2026E	2024E	2025E	2026E
美光	MU US	3.40	2.79	2.33	6.23	4.14	3.89
SK 海力士	000660 KS	2.48	1.89	1.53	2.57	1.99	1.86
三星	005930 KS	1.42	1.27	1.13	1.54	1.37	1.29

注: 估值数据为彭博一致预期

资料来源: Bloomberg、华泰研究

我们预计 FY24/25/26 海力士总营收分别为 68.7/109.5/126.4 万亿韩元, 同比为 109.7%/59.4%/15.4%。具体而言, 基于报告前序章节对海力士 HBM、DRAM 和 NAND 业务现状和前景分析, 以及对三项业务市场态势的判断, 我们按海力士业务, 分别对其 DRAM 和 NAND 板块进行盈利预测:

DRAM 业务: 我们预计 FY24/25/26 营收为 44.8/74.3/87.5 万亿韩元, 对应同比为 114.7%/65.9%/17.8%, 其中 FY24/25/26 海力士 DRAM 位元出货量 (bit shipments) 为 774/992/1126 亿 Gb, 对应同比为 20.6%/28.2%/13.5%; ASP 为每 Gb 578.5/748.8/777.4 韩元, 对应同比为 78.0%/29.4%/3.8%, 主要由 HBM 的量价齐升带动, 叠加 AI 手机、AI PC 和 AI 服务器的 DRAM 需求推动。海力士作为 HBM 龙头, HBM 业务将受惠于 AI 芯片需求和设计进化, 并通过自身优势维持市场份额; 同时, 以 DDR5 为代表的 AI 相关内存需求则受惠于 AI 手机、AI PC 和 AI 服务器渗透率提升, 为海力士营收锦上添花。另外, 经过 23 年供应端主动减产以及下游对高端产品补库需求, 传统 DRAM ASP 和需求周期性回暖, 带动供应端产能逐渐恢复, “供销两旺” 趋势或在 24H2 至 25 年明朗。

展望 FY24-25, 我们认为海力士 DRAM 业务:

1) 以 HBM 为代表的新兴 DRAM 产品, 24-25 年将维持供不应求, 供需缺口分别为 5.4% 和 3.6%, 受益于 AI 芯片需求火热, 全球 HBM TAM 将快速增长到 24 年的 120.7 亿和 25 年的 249.8 亿美元。海力士凭借 HBM3 和 HBM3E 稳定供应英伟达, 相较 23 年约 50% 的市场份额, 我们预计凭借前文所述竞争优势, 公司 HBM 业务在 24 及 25 年的份额将增长至 66% 和 56%, 对应 HBM 营收 11.4 万亿和 19.9 万亿韩元, 占海力士 DRAM 业务总营收的 26% 和 27%。

海力士 HBM 需求能见度, 公司称 24 年 HBM 产能已售罄, 25 年产能也几近售罄, 若能按照计划在 FY24-FY25 实现 HBM3E 扩产和客户拓展, 将进一步受惠于高速增长 HBM 市场, 并凭借 HBM 产品更高的 ASP 带来利润率的优化;

2) 以 DDR5 为代表的 DRAM 产品将受 AI 相关需求拉动, AI 手机、AI PC 和 AI 服务器渗透率提升带来新商机: 大模型和生成式 AI 逐渐向端侧扩展, 数据中心 AI 服务器需求蓬勃同时, AI PC 和 AI 手机也开始放量。我们认为海力士面向 AI 的产品布局完善, 作为面向端侧 AI 场景的内存方案将显著受益;

3) 通用服务器需求或见回暖, Windows 10 终止服务及 Windows 12 的推出将对 PC 需求有拉动作用, 带动传统 DRAM 逐步需求修复: 经过 23 年减产, 我们认为下游库存或已出清, 通用服务器和 PC 迎来换机潮。北美云厂商继 17-18 年大规模配置服务器扩展云服务后, 24-25 年计划对搭载 DDR5 新平台通用服务器进行新一轮采购, 同时, 微软停止支持 Windows 10 带动 PC 换机潮。根据 IDC 24 年 1 月和 4 月数据, 23Q4 全球个人电脑出货量为 6710 万台, 同比下降 0.1%; 24Q1 全球个人电脑出货量为 5980 万台, 同比上升 5.1%, 恢复至 19Q1 水平。我们认为 24H1 PC 市况已经逐步好转, 虽然需求大规模增长尚未出现, 但 24H2 伴随 Windows 系统换代及 AI PC 逐渐流入市场, PC 市场可望迎来进一步复苏, DRAM 需求“基本盘”步入上行周期。

NAND 业务: 我们预计 FY24/25/26 营收为 21.7/32.9/36.6 万亿韩元, 对应营收同比为 123.1%/51.9%/11.1%, 其中 FY24/25/26 海力士 NAND 位元出货量 (bit shipments) 为 1758/2040/2340 亿 GB, 对应同比为 10.3%/16.1%/14.7%; ASP 为每 GB 123.3/161.4/156.4 韩元, 对应同比为 102.2%/30.9%/-3.1%, 26 年 ASP 下滑主要系高容量、低成本的 QLC NAND 在 AI 端侧设备和数据中心的持续导入。

展望 FY24-25, 我们认为 NAND 有望“供销两旺”, 增长核心在于下游设备需求拉动, 同时 ASP 稳步增长:

1) 伴随下游客户库存水平正常化, FY24 NAND 和 SSD 趋势较乐观。23Q3 以来 NAND ASP 逐渐回升, 部分主流厂商释放 24 年产能利用率回升积极信号, 重点放在基于高层 NAND 的大容量 SSD。疫情以来高基数影响以及下游 NAND 库存已基本在历史平均水平附近, 我们认为 24 年 NAND 逐步复苏已成共识;

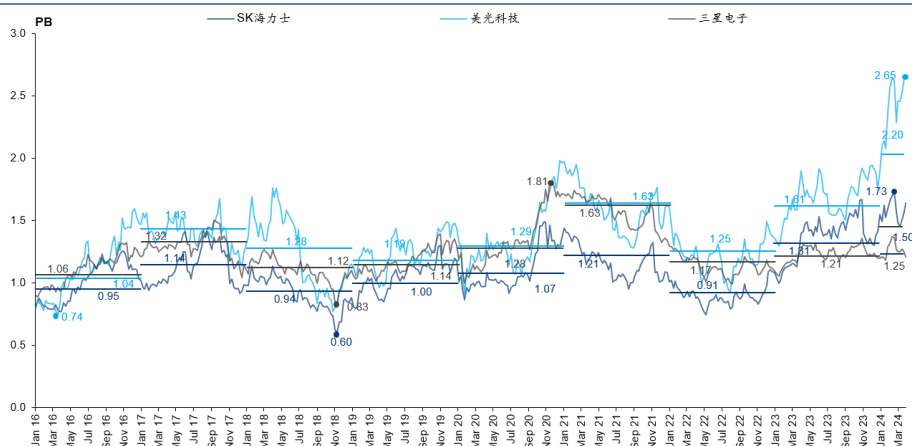
2) AI 浪潮惠及 NAND Flash, 带动 PC、手机及服务器的需求增加, 驱动 NAND 供应商提供更大容量、更高速的产品: 鉴于 HDD 读写延迟较 SSD 均更高, “换芯不换存”相当于浪费逻辑芯片和 DRAM 性能, 而 SSD 有助减少逻辑芯片和 DRAM 的性能浪费。海力士 176 层和 238 层 TLC NAND 面向端侧 AI, 并协同 Solidigm 布局大容量 QLC NAND, 其 NAND 产品矩阵有望受益于 AI 服务器对高规格设备硬件需求以及通用服务器换机潮将带动大容量 SSD 需求, 以及未来 AI PC 和 AI 手机放量, 端侧运行 AI 模型生成图像影片等场景带来的用户端 SSD 和 UFS 4.0 大容量存储需求。

其他业务: 我们预计 FY24/25/26 营收为 2.25/2.27/2.27 万亿韩元, 对应营收同比为 2.5%/1.0%/0.0%。海力士其他业务以 NOR 闪存为主, 鉴于 NOR 闪存的应用场景主要是智能可穿戴设备和车载芯片, 经过 23 年减产, 疫情带来高基数影响逐渐消退。且考虑到厂商库存积压情况逐渐得到缓解, 消费电子端设备出货量将实现稳定修复, 我们认为公司其他业务将逐渐步入正常发展轨道, 但并非公司重心所在。

费率和利润率: 毛利率方面, 我们认为 HBM 和 DDR5 营收贡献将带来海力士毛利率持续优化, 同时, DRAM 和 NAND ASP 上行在 24H2 和 25 年将持续, 但幅度或不及 23Q3 至今, 26 年考虑到周期因素 ASP 或有下跌, 结合海力士 FY23Q4/FY24Q1 毛利率为 20%/39%, 考量公司 24-25 年 HBM 按计划扩产并出货, 我们预计 FY24/25/26 年毛利率分别为 45%/50%/50%。费用率方面, 我们认为公司将集中资金于先进制程的产能爬升和 M15X DRAM 基地建设, 因此削减其他非核心业务的运营开支, 预测 FY24/25/26 销售及管理费率分别为 9%/8.5%/8.5%。我们认为为保证 12 层 36GB HBM3E 量产, HBM4 研发和 EUV 顺利导入, 研发费用将稳中有升, 而研发费用率则随营收规模的增长而逐年下降, 预测 FY24/25/26 研发费率为 6.0%/5.5%/5.4%。

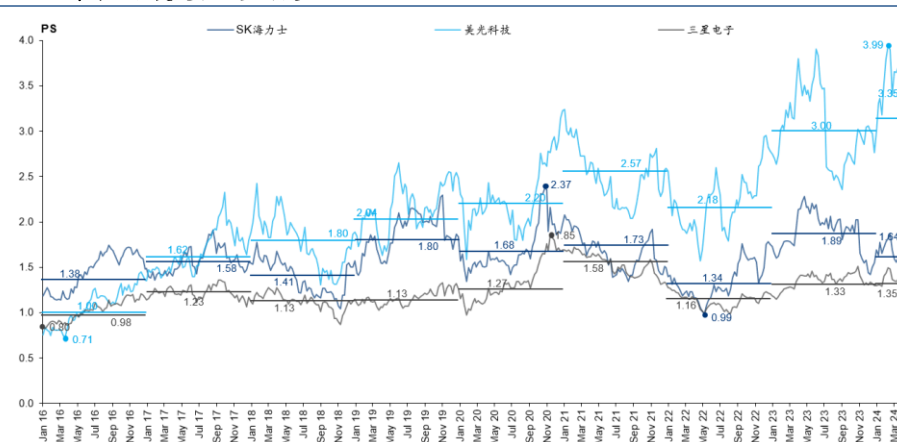
海力士 HBM3/3E 背靠英伟达供应链、伴随存储市场复苏以及 AI 设备（PC、手机和服务器等）崛起，我们预计海力士将维持高速增长。鉴于海力士所在存储行业的强周期性，营收及盈利水平随着周期波动，而作为重资产、自行投片生产存储颗粒的“原厂”，海力士净资产相对波动较小，因此我们主要基于 PB 估值，考虑到本轮上行并非如同加密货币潮和疫情带来的短期需求激增，而是受益于 AI 芯片及 AI 设备普及有望实现中长期增长，借助 HBM 龙头地位，其估值上修可期，并有望突破历史高位，我们预测海力士 FY24/25 净资产 64.2/88.1 万亿韩元，对应每股净资产 8.8/12.1 万韩元，同时，我们认为海力士 25 年约 2.5x PB 较为合理，目标价 30 万韩元，首次覆盖给予“买入”评级。

图表89：SK 海力士、美光和三星历史 PB



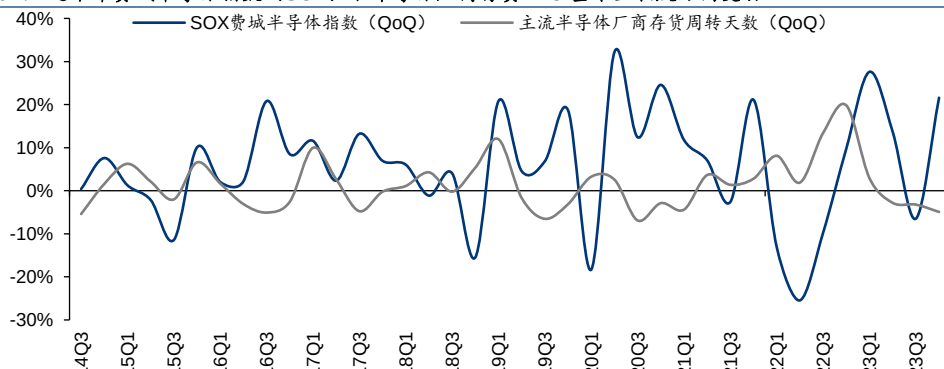
资料来源：Bloomberg、华泰研究

图表90：SK 海力士、美光和三星历史 PS



资料来源：Bloomberg、华泰研究

图表91：近十年费城半导体指数（SOX）和半导体厂商存货 DIO 基本呈相反方向变动



资料来源：Bloomberg、华泰研究

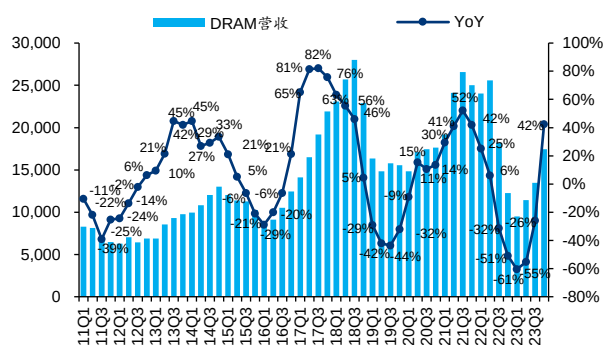
图表92：海力士分业务盈利预测（单位：十亿韩元）

Total Revenue	2019A	2020A	2021A	2022A	2023A	2024E	2025E	2026E
DRAM（十亿韩元）	20,829	23,136	30,836	28,521	20,849	44,765	74,278	87,531
YoY(%)		11.1%	33.3%	-7.5%	-26.9%	114.7%	65.9%	17.8%
位元总供给 (bn Gb)	37.8	46.1	55.4	56.3	64.2	77.4	99.2	112.6
YoY(%)		22.1%	20.1%	1.6%	14.1%	20.6%	28.2%	13.5%
ASP (WON/Gb)	551.6	501.8	556.8	507.0	324.9	578.5	748.8	777.4
YoY(%)		-9.0%	10.9%	-8.9%	-35.9%	78.0%	29.4%	3.8%
HBM（十亿韩元）	-	-	-	-	-	11,424.3	19,919.3	-
位元总供给 (mn GB)	-	-	-	-	-	705.2	1,229.6	-
ASP (WON/GB)	-	-	-	-	-	16,200.0	16,200.0	-
占 DRAM 营收比重	-	-	-	-	-	26%	27%	-
NAND（十亿韩元）	5,201	7,580	10,638	14,169	9,720	21,682	32,940	36,603
YoY(%)		45.7%	40.3%	33.2%	-31.4%	123.1%	51.9%	11.1%
位元总供给 (bn GB)	40.6	57.6	92.7	134.8	159.3	175.8	204.0	234.0
YoY(%)		41.9%	61.0%	45.4%	18.2%	10.3%	16.1%	14.7%
ASP (WON/GB)	128.2	131.6	114.7	105.1	61.0	123.3	161.4	156.4
YoY(%)		2.7%	-12.8%	-8.4%	-42.0%	102.2%	30.9%	-3.1%
Others（十亿韩元）	961	1,184	1,524	1,932	2,197	2,252	2,274	2,274
YoY(%)		23.2%	28.7%	26.8%	13.7%	2.5%	1.0%	0.0%
总收入（十亿韩元）	26,991	31,900	42,998	44,622	32,766	68,699	109,491	126,407
YoY(%)		18.2%	34.8%	3.8%	-26.6%	109.7%	59.4%	15.4%

注：其他收入主要为 NOR

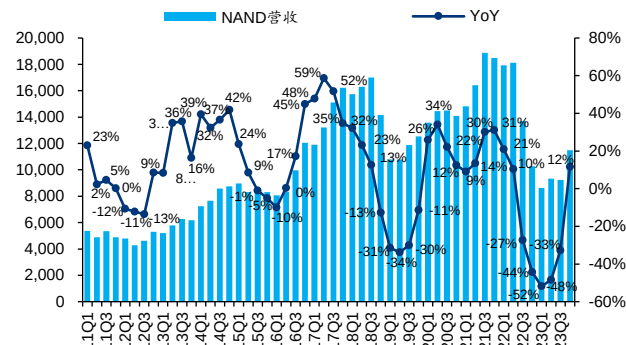
资料来源：公司公告、华泰研究预测

图表93：全球 DRAM 营收及 YoY (CY, 百万美元)



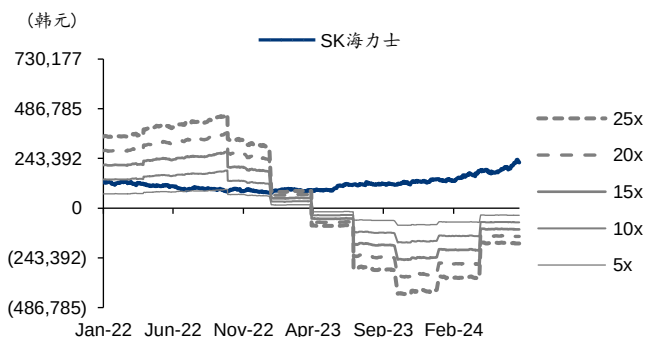
资料来源：TrendForce、华泰研究

图表94：全球 NAND 营收及 YoY (CY, 百万美元)



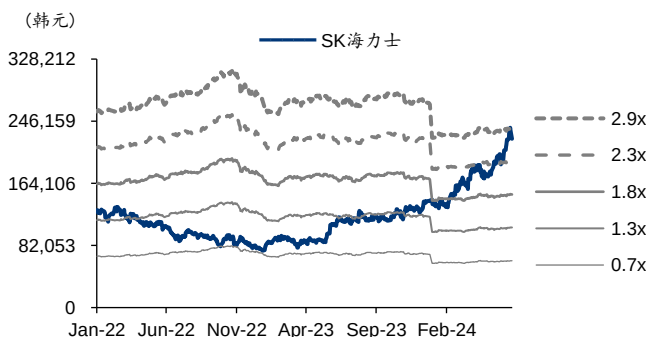
资料来源：TrendForce、Business Quant、华泰研究

图表95：SK 海力士 PE-Bands



资料来源：S&P、华泰研究

图表96：SK 海力士 PB-Bands



资料来源：S&P、华泰研究

风险提示

AI 落地不及预期

自 ChatGPT 落地应用并取得一定成功，各科技巨头均加快 和加大力度布局 AIGC 领域，如 Meta 于 23 年年初建立 AIGC 团队、微软也在其 Azure、Bing 等多项自有业务进一步整合 AI 技术。由于人工智能属于高新技术，需投入较大前期研发成本和时间，后续 AI 技术落地可能会受企业投入、宏观经济、政策和舆论等多方面影响，致使研发进度不及预期。

芯片需求不及预期

存储芯片需求规模可能受下游应用端需求影响不及预期，影响存储行业营收及利润。

中美贸易摩擦加剧

贸易摩擦加剧将对行业投资、运营和海力士产品销售产生影响。

盈利预测

利润表

会计年度 (韩元十亿)	2022	2023	2024E	2025E	2026E
营业收入	44,622	32,766	68,699	109,491	126,407
销售成本	(28,994)	(33,299)	(37,784)	(54,746)	(63,204)
毛利润	15,628	(533)	30,914	54,746	63,204
销售及管理费用	(4,344)	(3,446)	(6,183)	(9,307)	(10,745)
研发费用	(4,475)	(3,751)	(4,122)	(6,022)	(6,826)
其他收入/支出	(1,561)	(111)	(579)	(750)	(480)
财务成本净额	(1,377)	(3,831)	(3,185)	(256)	(407)
应占联营公司利润及亏损	131	15	103	83	67
税前利润	4,003	(11,658)	16,949	38,494	44,813
税费开支	(1,761)	2,520	(5,307)	(12,437)	(12,733)
少数股东损益	12	(25)	29	65	80
归母净利润	2,230	(9,112)	11,612	25,992	32,000
折旧和摊销	(14,151)	(13,674)	(14,515)	(14,768)	(15,182)
EBITDA	19,532	5,847	34,648	53,518	60,401
EPS (韩元, 基本)	3,241	(13,245)	16,878	37,779	46,511

资产负债表

会计年度 (韩元十亿)	2022	2023	2024E	2025E	2026E
存货	15,665	13,481	13,808	24,210	17,926
应收账款和票据	5,459	6,958	14,797	19,571	20,036
现金及现金等价物	4,977	7,587	13,326	24,275	58,244
其他流动资产	2,633	2,442	2,998	2,998	2,998
总流动资产	28,733	30,468	44,929	71,054	99,204
固定资产	60,229	52,705	52,831	54,696	56,127
无形资产	3,512	3,835	4,048	3,733	3,803
其他长期资产	11,398	13,323	13,863	14,457	15,145
总长期资产	75,138	69,862	70,742	72,887	75,076
总资产	103,872	100,330	115,670	143,941	174,280
应付账款	5,211	3,534	6,752	8,456	8,750
短期借款	7,704	10,489	10,839	11,190	11,541
其他负债	6,929	6,985	6,985	6,985	6,985
总流动负债	19,844	21,008	24,576	26,630	27,275
长期债务	17,088	22,010	22,010	22,010	22,010
其他长期债务	3,650	3,809	4,868	7,107	6,513
总长期负债	20,737	25,819	26,878	29,117	28,523
股本	3,658	3,658	3,658	3,658	3,658
储备/其他项目	59,609	49,847	60,530	84,443	114,650
股东权益	63,266	53,505	64,188	88,100	118,308
少数股东权益	24	(1)	28	93	174
总权益	63,291	53,504	64,216	88,194	118,481

估值指标

会计年度 (倍)	2022	2023	2024E	2025E	2026E
PE	72.22	(17.67)	13.87	6.20	5.03
PB	2.55	3.01	2.51	1.83	1.36
EV EBITDA	9.80	33.64	5.55	3.44	2.48
股息率 (%)	1.04	0.51	0.58	1.29	1.11
自由现金流收益率 (%)	(2.97)	(1.39)	5.03	6.64	22.58

资料来源：公司公告、华泰研究预测

现金流量表

会计年度 (韩元十亿)	2022	2023	2024E	2025E	2026E
EBITDA	19,532	5,847	34,648	53,518	60,401
融资成本	1,377	3,831	3,185	256	407
营运资本变动	(2,195)	(992)	(4,949)	(13,473)	6,114
税费	(1,761)	2,520	(5,307)	(12,437)	(12,733)
其他	(2,172)	(6,929)	(3,844)	(339)	(474)
经营活动现金流	14,781	4,278	23,734	27,525	53,715
CAPEX	(19,749)	(8,780)	(15,291)	(16,830)	(17,304)
其他投资活动	1,865	1,445	1,060	2,238	(594)
投资活动现金流	(17,884)	(7,335)	(14,232)	(14,592)	(17,898)
债务增加量	5	8	351	351	351
权益增加量	0	0	0	0	0
派发股息	(1,681)	(826)	(929)	(2,079)	(1,792)
其他融资活动现金流	4,497	6,515	(3,185)	(256)	(407)
融资活动现金流	2,822	5,697	(3,764)	(1,984)	(1,849)
现金变动	(281)	2,640	5,742	10,952	33,972
年初现金	5,058	4,977	7,587	13,326	24,275
汇率波动影响	200	(3)	(3)	(3)	(3)
年末现金	4,977	7,587	13,326	24,275	58,244

业绩指标

会计年度 (倍)	2022	2023	2024E	2025E	2026E
增长率 (%)					
营业收入	3.78	(26.57)	109.67	59.38	15.45
毛利润	(17.54)	(103.41)	(5,895)	77.09	15.45
营业利润	(45.13)	(213.52)	(366.61)	91.25	15.77
净利润	(76.78)	(508.71)	(227.43)	123.83	23.11
EPS	(76.84)	(508.71)	(227.43)	123.83	23.11
盈利能力比率 (%)					
毛利润率	35.02	(1.63)	45.00	50.00	50.00
EBITDA	43.77	17.85	50.44	48.88	47.78
净利润率	5.00	(27.81)	16.90	23.74	25.31
ROE	3.56	(15.61)	19.73	34.13	31.01
ROA	2.23	(8.92)	10.75	20.02	20.11
偿债能力 (倍)					
净负债比率 (%)	31.32	46.56	30.41	10.13	(20.87)
流动比率	1.45	1.45	1.83	2.67	3.64
速动比率	0.66	0.81	1.27	1.76	2.98
营运能力 (天)					
总资产周转率 (次)	0.45	0.32	0.64	0.84	0.79
应收账款周转天数	56.06	68.21	57.00	56.50	56.40
应付账款周转天数	54.93	47.27	49.00	50.00	49.00
存货周转天数	152.61	157.55	130.00	125.00	120.00
现金转换周期	153.73	178.49	138.00	131.50	127.40
每股指标 (韩元)					
EPS	3,241	(13,245)	16,878	37,779	46,511
每股净资产	91,938	77,753	93,278	128,027	171,924

免责声明

分析师声明

本人，何翩翩，兹证明本报告所表达的观点准确地反映了分析师对标的证券或发行人的个人意见；彼以往、现在或未来并未就其研究报告所提供的具体建议或所表达的意见直接或间接收取任何报酬。

一般声明及披露

本报告由华泰证券股份有限公司（已具备中国证监会批准的证券投资咨询业务资格，以下简称“本公司”）制作。本报告所载资料是仅供接收人的严格保密资料。本报告仅供本公司及其客户和其关联机构使用。本公司不因接收人收到本报告而视其为客户。

本报告基于本公司认为可靠的、已公开的信息编制，但本公司及其关联机构（以下统称为“华泰”）对该等信息的准确性及完整性不作任何保证。

本报告所载的意见、评估及预测仅反映报告发布当日的观点和判断。在不同时期，华泰可能会发出与本报告所载意见、评估及预测不一致的研究报告。同时，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。以往表现并不能指引未来，未来回报并不能得到保证，并存在损失本金的可能。华泰不保证本报告所含信息保持在最新状态。华泰对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。

本公司不是 FINRA 的注册会员，其研究分析师亦没有注册为 FINRA 的研究分析师/不具有 FINRA 分析师的注册资格。

华泰力求报告内容客观、公正，但本报告所载的观点、结论和建议仅供参考，不构成购买或出售所述证券的要约或招揽。该等观点、建议并未考虑到个别投资者的具体投资目的、财务状况以及特定需求，在任何时候均不构成对客户私人投资建议。投资者应当充分考虑自身特定状况，并完整理解和使用本报告内容，不应视本报告为做出投资决策的唯一因素。对依据或者使用本报告所造成的一切后果，华泰及作者均不承担任何法律责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

除非另行说明，本报告中所引用的关于业绩的数据代表过往表现，过往的业绩表现不应作为日后回报的预示。华泰不承诺也不保证任何预示的回报会得以实现，分析中所做的预测可能是基于相应的假设，任何假设的变化可能会显著影响所预测的回报。

华泰及作者在自身所知情的范围内，与本报告所指的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，华泰可能会持有报告中提到的公司所发行的证券头寸并进行交易，为该公司提供投资银行、财务顾问或者金融产品等相关服务或向该公司招揽业务。

华泰的销售人员、交易人员或其他专业人士可能会依据不同假设和标准、采用不同的分析方法而口头或书面发表与本报告意见及建议不一致的市场评论和/或交易观点。华泰没有将此意见及建议向报告所有接收者进行更新的义务。华泰的资产管理部门、自营部门以及其他投资业务部门可能独立做出与本报告中的意见或建议不一致的投资决策。投资者应当考虑到华泰及/或其相关人员可能存在影响本报告观点客观性的潜在利益冲突。投资者请勿将本报告视为投资或其他决定的唯一信赖依据。有关该方面的具体披露请参照本报告尾部。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布的机构或人员，也并非意图发送、发布给因可得到、使用本报告的行为而使华泰违反或受制于当地法律或监管规则的机构或人员。

本报告版权仅为本公司所有。未经本公司书面许可，任何机构或个人不得以翻版、复制、发表、引用或再次分发他人（无论整份或部分）等任何形式侵犯本公司版权。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并需在使用前获取独立的法律意见，以确定该引用、刊发符合当地适用法规的要求，同时注明出处为“华泰证券研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。本公司保留追究相关责任的权利。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

中国香港

本报告由华泰证券股份有限公司制作，在香港由华泰金融控股（香港）有限公司向符合《证券及期货条例》及其附属法律规定的机构投资者和专业投资者的客户进行分发。华泰金融控股（香港）有限公司受香港证券及期货事务监察委员会监管，是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。在香港获得本报告的人员若有任何有关本报告的问题，请与华泰金融控股（香港）有限公司联系。

香港-重要监管披露

- 华泰金融控股（香港）有限公司的雇员或其关联人士没有担任本报告中提及的公司或发行人的高级人员。
- 有关重要的披露信息，请参华泰金融控股（香港）有限公司的网页 https://www.htsc.com.hk/stock_disclosure 其他信息请参见下方 “美国-重要监管披露”。

美国

在美国本报告由华泰证券（美国）有限公司向符合美国监管规定的机构投资者进行发表与分发。华泰证券（美国）有限公司是美国注册经纪商和美国金融业监管局（FINRA）的注册会员。对于其在美国分发的研究报告，华泰证券（美国）有限公司根据《1934 年证券交易法》（修订版）第 15a-6 条规定以及美国证券交易委员会人员解释，对本研究报告内容负责。华泰证券（美国）有限公司联营公司的分析师不具有美国金融监管（FINRA）分析师的注册资格，可能不属于华泰证券（美国）有限公司的关联人员，因此可能不受 FINRA 关于分析师与标的公司沟通、公开露面和所持交易证券的限制。华泰证券（美国）有限公司是华泰国际金融控股有限公司的全资子公司，后者为华泰证券股份有限公司的全资子公司。任何直接从华泰证券（美国）有限公司收到此报告并希望就本报告所述任何证券进行交易的人士，应通过华泰证券（美国）有限公司进行交易。

美国-重要监管披露

- 分析师何翩翩本人及相关人士并不担任本报告所提及的标的证券或发行人的高级人员、董事或顾问。分析师及相关人士与本报告所提及的标的证券或发行人并无任何相关财务利益。本披露中所提及的“相关人士”包括 FINRA 定义下分析师的家庭成员。分析师根据华泰证券的整体收入和盈利能力获得薪酬，包括源自公司投资银行业务的收入。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或不时会以自身或代理形式向客户出售及购买华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）华泰证券研究所覆盖公司的证券/衍生工具，包括股票及债券（包括衍生品）。
- 华泰证券股份有限公司、其子公司和/或其联营公司，及/或其高级管理层、董事和雇员可能会持有本报告中所提到的任何证券（或任何相关投资）头寸，并可能不时进行增持或减持该证券（或投资）。因此，投资者应该意识到可能存在利益冲突。

新加坡

华泰证券（新加坡）有限公司持有新加坡金融管理局颁发的资本市场服务许可证，可从事资本市场产品交易，包括证券、集体投资计划中的单位、交易所交易的衍生品合约和场外衍生品合约，并且是《财务顾问法》规定的豁免财务顾问，就投资产品向他人提供建议，包括发布或公布研究分析或研究报告。华泰证券（新加坡）有限公司可能会根据《财务顾问条例》第 32C 条的规定分发其在华泰内的外国附属公司各自制作的信息/研究。本报告仅供认可投资者、专家投资者或机构投资者使用，华泰证券（新加坡）有限公司不对本报告内容承担法律责任。如果您是非预期接收者，请您立即通知并直接将本报告返回给华泰证券（新加坡）有限公司。本报告的新加坡接收者应联系您的华泰证券（新加坡）有限公司关系经理或客户主管，了解来自或所分发的信息相关的事宜。

评级说明

投资评级基于分析师对报告发布日后 6 至 12 个月内行业或公司回报潜力（含此期间的股息回报）相对基准表现的预期（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数，台湾市场基准为台湾加权指数，日本市场基准为日经 225 指数，新加坡市场基准为海峡时报指数，韩国市场基准为韩国有价证券指数），具体如下：

行业评级

增持：预计行业股票指数超越基准

中性：预计行业股票指数基本与基准持平

减持：预计行业股票指数明显弱于基准

公司评级

买入：预计股价超越基准 15% 以上

增持：预计股价超越基准 5%~15%

持有：预计股价相对基准波动在-15%~5%之间

卖出：预计股价弱于基准 15% 以上

暂停评级：已暂停评级、目标价及预测，以遵守适用法规及/或公司政策

无评级：股票不在常规研究覆盖范围内。投资者不应期待华泰提供该等证券及/或公司相关的持续或补充信息

法律实体披露

中国: 华泰证券股份有限公司具有中国证监会核准的“证券投资咨询”业务资格, 经营许可证编号为: 91320000704041011J

香港: 华泰金融控股(香港)有限公司具有香港证监会核准的“就证券提供意见”业务资格, 经营许可证编号为: AOK809

美国: 华泰证券(美国)有限公司为美国金融业监管局(FINRA)成员, 具有在美国开展经纪交易商业业务的资格, 经营业务许可编号为: CRD#:298809/SEC#:8-70231

新加坡: 华泰证券(新加坡)有限公司具有新加坡金融管理局颁发的资本市场服务许可证, 并且是豁免财务顾问。公司注册号: 202233398E

华泰证券股份有限公司**南京**

南京市建邺区江东中路228号华泰证券广场1号楼/邮政编码: 210019

电话: 86 25 83389999/传真: 86 25 83387521

电子邮件: ht-rd@htsc.com

深圳

深圳市福田区益田路5999号基金大厦10楼/邮政编码: 518017

电话: 86 755 82493932/传真: 86 755 82492062

电子邮件: ht-rd@htsc.com

北京

北京市西城区太平桥大街丰盛胡同28号太平洋保险大厦A座18层/

邮政编码: 100032

电话: 86 10 63211166/传真: 86 10 63211275

电子邮件: ht-rd@htsc.com

上海

上海市浦东新区东方路18号保利广场E栋23楼/邮政编码: 200120

电话: 86 21 28972098/传真: 86 21 28972068

电子邮件: ht-rd@htsc.com

华泰金融控股(香港)有限公司

香港中环皇后大道中99号中环中心53楼

电话: +852-3658-6000/传真: +852-2567-6123

电子邮件: research@htsc.com

http://www.htsc.com.hk

华泰证券(美国)有限公司

美国纽约公园大道280号21楼东(纽约10017)

电话: +212-763-8160/传真: +917-725-9702

电子邮件: Huatai@htsc-us.com

http://www.htsc-us.com

华泰证券(新加坡)有限公司

滨海湾金融中心1号大厦, #08-02, 新加坡 018981

电话: +65 68603600

传真: +65 65091183

©版权所有2024年华泰证券股份有限公司