人工智能存储系列报告一：

AI拉动需求增长，存储大周期方兴未艾

行业研究 · 行业专题

计算机 · 人工智能

投资评级：优于大市（维持评级）

证券分析师：熊莉
021-61761067
xiongli1@guosen.com.cn
S0980519030002

证券分析师：艾宪
0755-22941051
aixian@guosen.com.cn
S0980524090001

摘要

国信证券 GUOSEN SECURITIES

- **存储系统：系统构成与分类。1）系统构成：**AI存储系统主要分为网络端存储和本地端存储，其中网络端存储主要存放冷数据，主要由HDD和SSD存储；本地端存储主要存放热数据和温数据，主要由HBM、DRAM、本地SSD存储。**2）存储分类：**存储主要可以分为“易失性”的Memory和“非易失性”的Storage，其中Memory主要包括DRAM和HBM，优势为速度快；Storage主要包括SSD和HDD，优势为容量大、成本低。

- **市场与技术趋势：HDD、SSD、NAND、DRAM、HBM。1）HDD：**HDD即硬盘，主要通过提升面密度提升HDD容量，近期HAMR（热辅助磁记录）技术可以大幅提升单碟容量，市场格局呈现双寡头垄断，希捷科技、西部数据为主要参与者；**2）SSD：**SSD即固态硬盘，一种以NAND Flash为介质的存储设备，NAND堆叠层数持续增长，单元架构逐步从SLC转化为MLC、QLC，市场参与者主要为三星、海力士（包括Solidigm）、美光、闪迪、铠侠；**3）DRAM：**DRAM具备功耗小，集成度高，成本低等优势，逐步从DRAM迭代至SDRAM、DDR，目前头部厂商已经开始对DDR6研发，下游需求主要在手机、PC、服务器领域，主要参与者包括三星、海力士、美光；**4）HBM：**多层DRAM芯片堆叠，通过TSV实现垂直方向的互联，进而具有更高的存储密度和更大的带宽，目前主要应用于AI领域，主要参与者为海力士、美光、三星；**5）NAND：**存储单元表达的bit数持续增长，堆叠层数持续增长，下游需求主要为SSD和手机，主要参与者为三星、海力士、铠侠、美光、闪迪。

- **需求测算：AI训练、推理拉动存储需求增长。**AI大模型推理拉动存储需求快速增长，根据我们测算结果，2026年AI推理对DRAM、NAND需求分别为23.0EB、593.5EB，短期供不应求，存储价格有望持续提升，存储大周期方兴未艾。

- **公司梳理：全球存储公司业务重心：**三星电子、海力士在DRAM、HBM、NAND、SSD等领域市占率均较高，为全球存储龙头公司；其次为美光，产品矩阵全面，但市占率略低于三星、海力士；闪迪、铠侠聚焦于NAND、SSD领域，西部数据、希捷科技聚焦于HDD领域。

- **风险提示：**厂商DRAM、NAND扩产，进而导致产品价格下降风险；互联网大厂资本开支不及预期风险；AI应用活跃用户数增长不及预期风险；AI大模型方案优化，进而减少对存储需求风险等。

目录

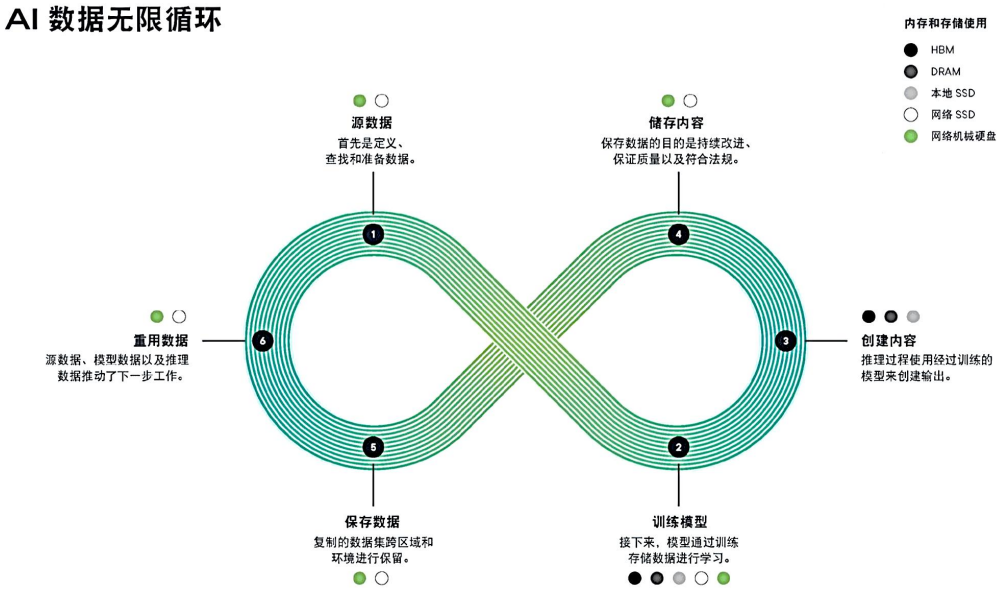
存储系统：AI工作负载在不同阶段需要不同存力支撑

国信证券 GUOSEN SECURITIES

■ AI工作负载在不同阶段需要不同的存力支撑。

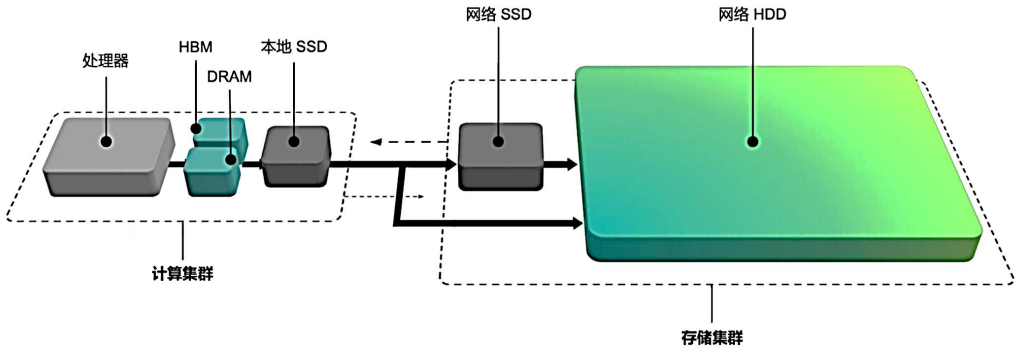
- **源数据（网络SSD和HDD）**：硬盘（HDD）能够长期保存原始数据并提供数据保护；固态盘（SSD）作为即时访问的数据层；
- **训练模型（HBM、DRAM、本地SSD、网络SSD和HDD）**：数据快速从存储加载到HBM、DRAM以及本地固态盘，供后续计算密集型操作使用；其中，网络HDD和SSD存储检查点（CheckPoint），以保护和优化模型训练；
- **创建内容（HBM、DRAM、本地SSD）**：推理过程中的内容创建主要依靠HBM、DRAM和本地SSD或HDD完成；
- **存储内容（网络SSD和HDD）**：存储内容以便后续优化，硬盘用于存储并保护内容的副本；
- **保留数据（网络SSD和HDD）**：复制的数据集跨区域和环境进行保留；
- **重用数据（网络SSD和HDD）**：元数据、模型数据以及推理数据推动了下一步工作。

图1：AI数据在不同处理环节需要使用内存和存储



资料来源：Seagate，国信证券经济研究所整理

图2：存储系统可以分为本地存储和网络存储



资料来源：益企研究院，国信证券经济研究所整理

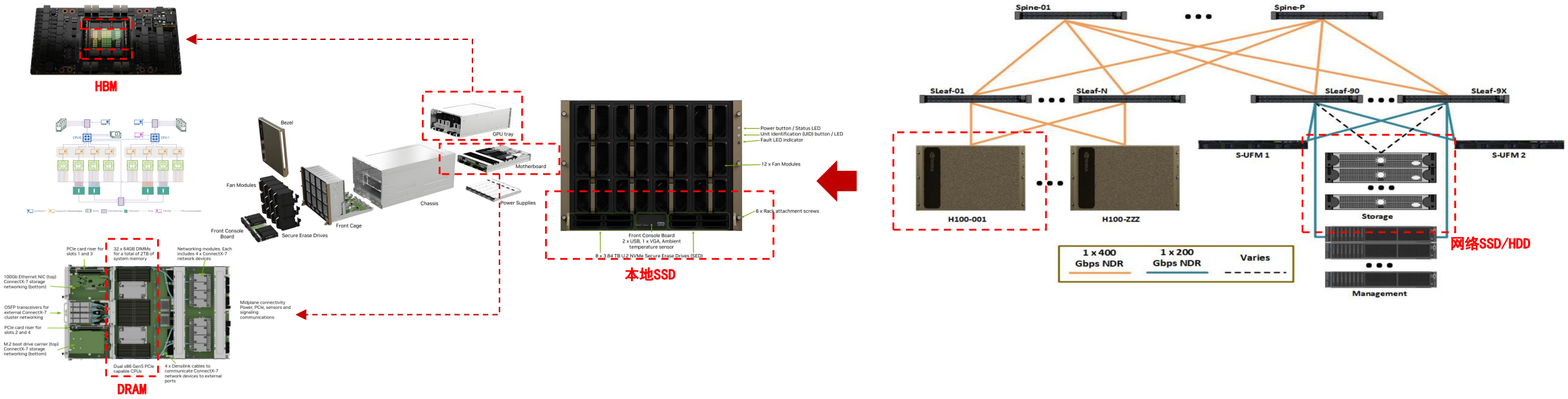
存储系统：英伟达H100存储系统拆解

国信证券 GUOSEN SECURITIES

■ 英伟达H100存储系统拆解：

- **HBM**：以英伟达H100为例，单颗H100配置了6颗HBM3堆栈（单堆栈为8个堆叠，单一堆叠为2GB，总共96GB）；
- **DRAM**：以英特尔Sapphire Rapids为例，单颗CPU支持8个内存通道，每通道支持2个DIMM；英伟达H100单台服务器配置两颗CPU，共计32个DIMM插槽，配置32个64GB DRAM（共计2TB）作为系统内存；
- **本地SSD**：英伟达H100服务器配置8个3.84TB的SSD（使用NVMe协议）。

图3：英伟达H100存储系统拆解



资料来源：英伟达，国信证券经济研究所整理

存储系统：英伟达存储系统比例关系

国信证券 GUOSEN SECURITIES

- **英伟达DGX B200（8卡服务器）为例：**根据英伟达披露数据，单颗B200配置HBM3E（180GB/s），则单台服务器（8颗B200）合计1.4TB/s；主板DRAM为2TB（可拓展至4TB，即System Memory），服务器本地SSD为8个3.84TB（合计30.72TB）；根据IBM披露数据，如果英伟达DGX B200服务器搭配IBM存储系统，4U（约127个计算节点）需要配置一个IBM Storage Scale 6000存储系统（加装9个硬盘HDD，对应3.4PB存储），则单台服务器对应27.4TB HDD，随着多模态模型的发展，以及逐步从训练侧转向推理侧，大量生成的图片、视频数据需要存储，网络存储需求有望持续提升。

图4：IBM存储系统（Scale 6000）针对英伟达4 SU DGX SuperPOS(B200）参考配置

Table 2-2 NVIDIA DGX SuperPOD 4 SU rack space requirements

Type of rack	No. of Racks	No. of Units	Component
Compute	32	127	NVIDIA DGX systems
		96	Raritan PDUs
Network	6	32 compute leaf switches	NVIDIA Quantum QM9700
		16 compute spine switches	NVIDIA Quantum QM9700
		16 storage switches	NVIDIA Quantum QM9700
		8 in-band management switches	NVIDIA SN5600
		8 out-of-band management switches	NVIDIA SN2201 (48x1Gbps ports)
		4 appliances	NVIDIA Unified Fabric Manager Appliance 3.1
		4 BCM management servers	NVIDIA Base Command Manager
		12	Raritan PDUs
Storage	1	1 IBM Storage Scale 6000 building block with 9 disk enclosures	IBM Storage Scale 6000

资料来源：IBM，国信证券经济研究所整理

图5：IBM Storage Scale System

Here are the key features of the IBM Storage Scale System 6000:

- A single 4U node with active-active controllers and redundant hardware to maximize uptime.
- Up to 310 GBps throughput with low latency.
- Up to 13 millions IOPS by using NVMeoF.
- Up to 3.4 PBe (effective capacity) in a standard 4U rack space.
- Supports up to forty-eight 3.84 TB, 7.68 TB, 15.36 TB, or 30 TB 2.5” Non-Volatile Memory Express (NVMe) flash drives.
- Supports 19.2 TB and 38.4 TB FlashCore Module 4 NVMe drives.

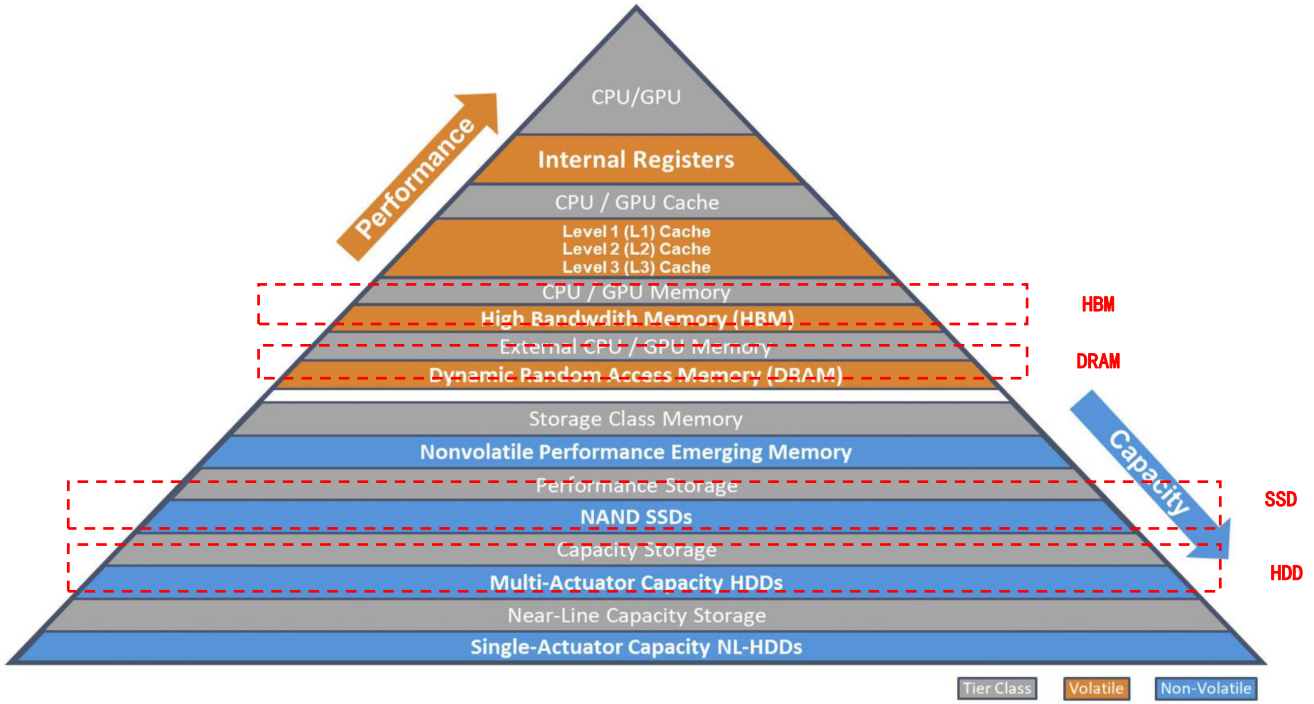
资料来源：IBM，国信证券经济研究所整理

存储分类：经典的存储金字塔层级

国信证券 GUOSEN SECURITIES

- **存储金字塔：**存储主要可以分为“易失性”的内存（Memory）和“非易失性”的存储（Storage），存储金字塔自上而下，性能逐级下降、容量逐级递减、成本逐级递减。
  - **内存（Memory）：**主要包括DRAM和HBM，属于“易失性”介质，断电后就会丢失信息，优势为速度快，劣势为成本高、容量有限，通常访问频繁或者随时变更的数据会保留在较高的存储层；
  - **存储（Storage）：**主要包括SSD和HDD，属于“非易失性”介质，优势为容量大、成本低，劣势为性能较弱，通常访问较不频繁或需要长期保存的数据将移动到较低的存储层。

图6：经典的存储金字塔层级



资料来源：希捷科技，国信证券经济研究所整理

目录

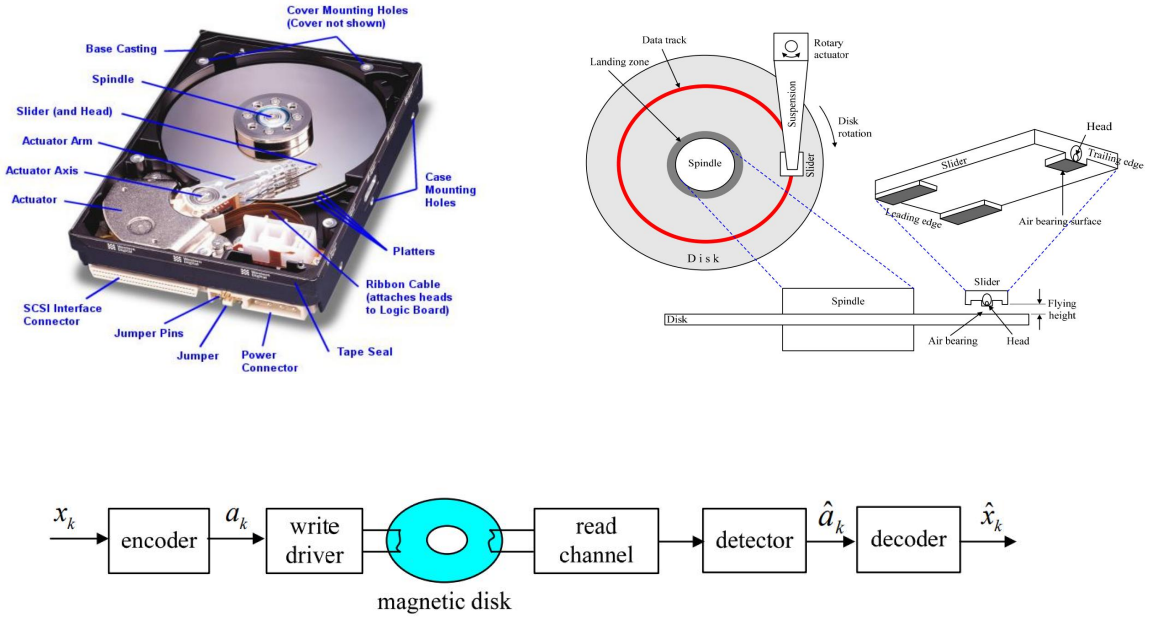
硬盘驱动器（HDD）：面密度提升是容量提升的核心

国信证券 GUOSEN SECURITIES

- **HDD架构**：通常包括机械部分和电子部分。
  - **机械部分**：主要包括底座、主轴电机、盘片、音圈电机、磁头组、顶盖等部分，所有盘片平行地安装在同一个转轴上，盘片的两面分别对应一个磁头，所有磁头关联在同一个磁头组上，磁头组尾部有一个音圈电机，驱动整个磁头组围绕同一个轴承旋转摆动；
  - **电子部分**：主要包括主控SOC、电机驱动芯片、RV传感器、Shock传感器、DRAM、Flash ROM等器件；

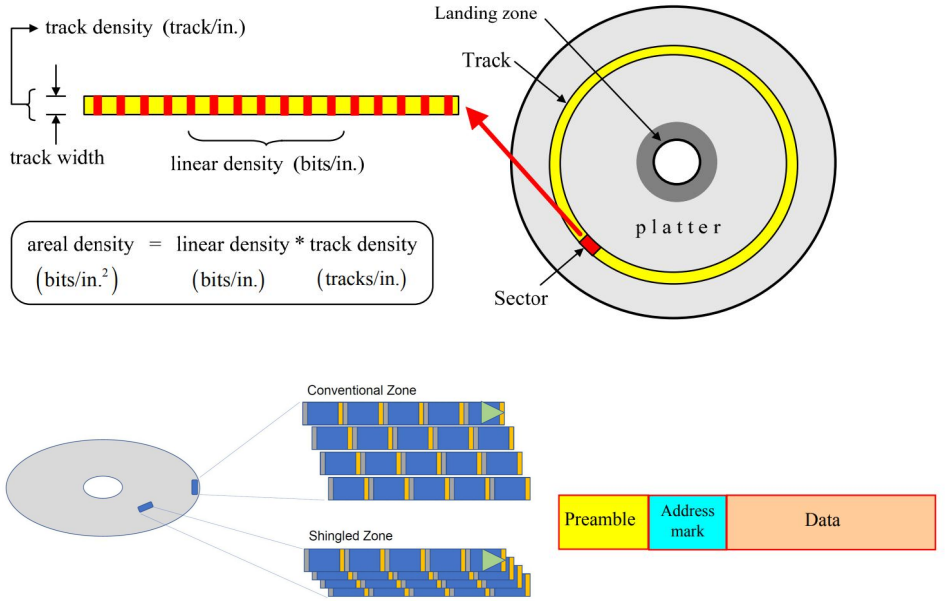
- **面密度提升是HDD容量提升的核心**：面密度可以分解为两个相互垂直的分量，圆周方向的记录密度（Linear Density，沿着单个磁道上单位长度可以存储的数据位数）和磁道密度（Track Density，每英寸磁盘表面可以容纳的磁道数量。传统磁记录技术（CMR），磁道间是独立、有间隙的，没有依赖关系；叠瓦式磁记录（SMR）允许相邻磁道部分重叠，提高了磁道密度，能存储更多的数据。

图7：HDD磁盘内部构造



资料来源：Nakhon Pathom Rajabhat University，国信证券经济研究所整理

图8：HDD磁盘面密度受记录密度和磁道密度影响



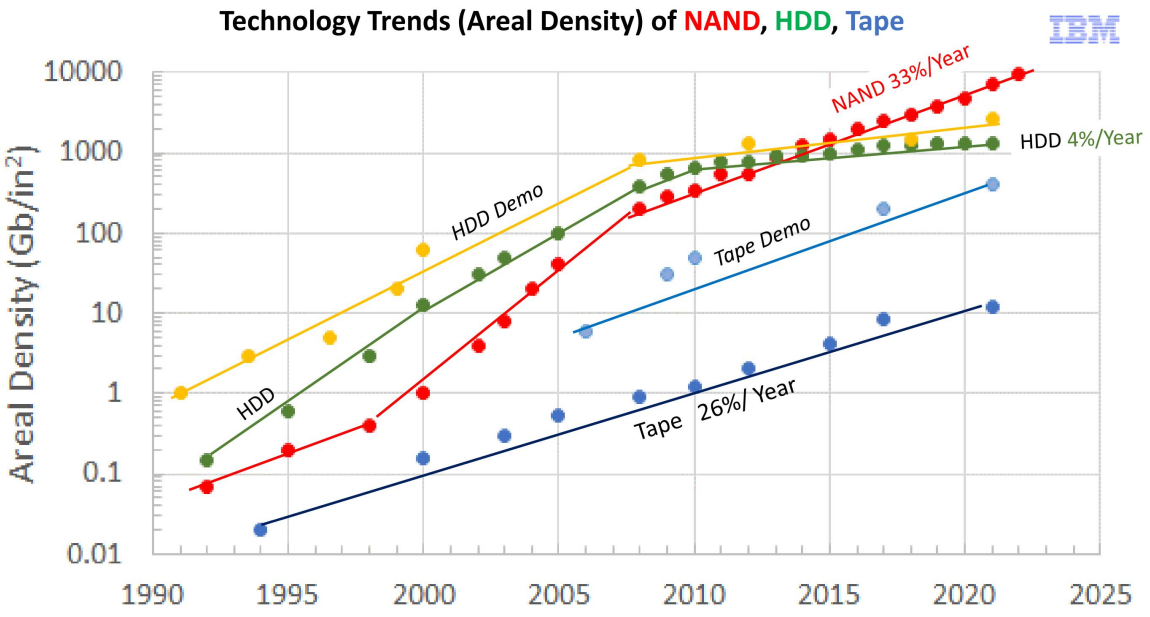
资料来源：Nakhon Pathom Rajabhat University、微软，国信证券经济研究所整理

硬盘驱动器（HDD）：面密度持续提升，成本持续下降

国信证券 GUOSEN SECURITIES

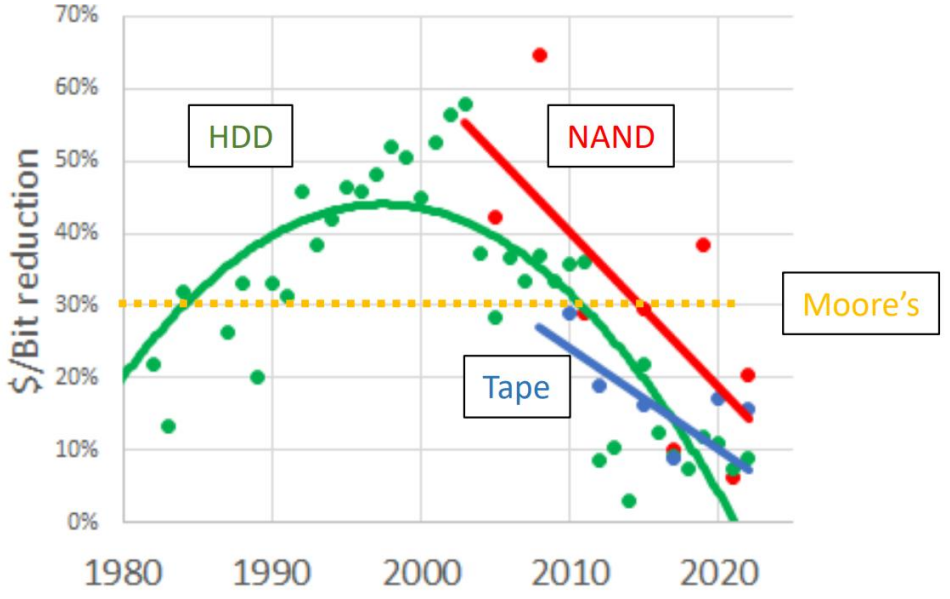
- **HDD磁盘面密度持续提升：**根据IBM统计数据，HDD磁盘面密度在2008年之前提升速度较快，2008年之后面密度提升斜率放缓，2023年仅4%的同比提升，未来有望在新技术的驱动下加速提升。
- **HDD磁盘每Bit成本持续下降：**根据IBM统计数据，HDD磁盘每Bit成本持续下降，2010年之后年化成本下降幅度已经低于30%，且持续放缓。

图9：HDD磁盘面密度持续提升，但增速放缓



资料来源：IBM，国信证券经济研究所整理

图10：HDD磁盘每Bit成本持续下降，但下降幅度放缓



资料来源：IBM，国信证券经济研究所整理

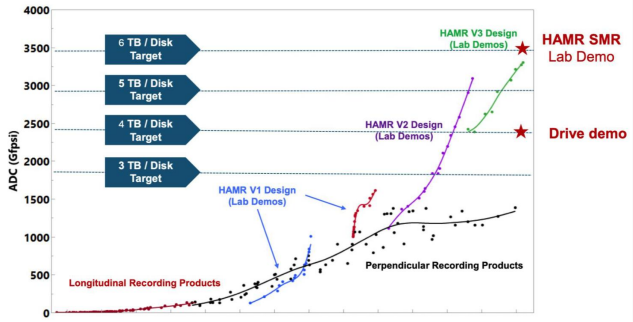
硬盘驱动器（HDD）技术趋势：HAMR提升单碟容量

国信证券 GUOSEN SECURITIES

- **HAMR（Heat-Assisted Magnetic Recording，热辅助磁记录）技术发展：**希捷在HAMR技术具备领先优势，通过HAMR技术，目前量产产品达到单碟3TB、全盘30TB，此外单碟4TB、5TB已经在路线图中。25年1月希捷宣布，基于HAMR技术的魔彩盒3+平台加持下，可在十碟片中提供36TB的容量点，未来有望实现单碟容量10TB。

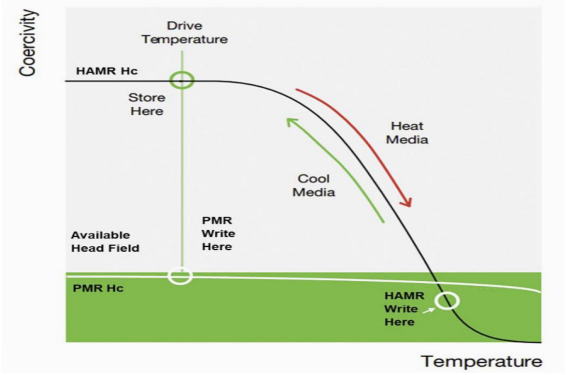
- **HAMR提升单碟容量：**随着时间的发展，HDD的盘片数量和盘片面积已经相对固定，提升磁盘的面密度成为主要的技术路径。面密度的提升会导致放置单位比特信息所占用的磁性颗粒面积变小，因为导致颗粒间相互磁影响越来越大，为保持信息稳定，需要使用高矫顽力颗粒，因而需要磁头施加刚强大的磁场变化，进而导致更长的操作时间和更多的干扰。HAMR技术通过等离子写入器精准地加热目标区域的超晶格铂合金介质，瞬间升温400℃以上，临时降低矫顽力以辅助写入，且在不到2ns的时间内迅速冷却。

图11：HAMR技术发展



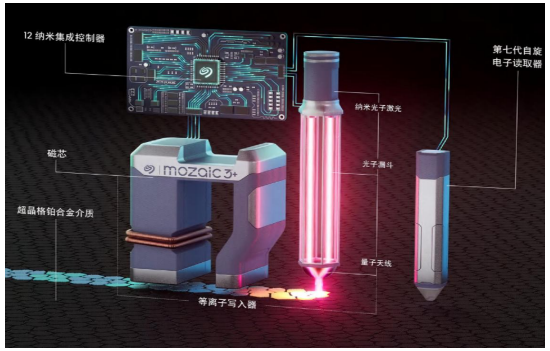
资料来源：益企研究院，国信证券经济研究所整理

图12：HAMR工作原理说明



资料来源：益企研究院，国信证券经济研究所整理

图13：HAMR技术关键组件



资料来源：希捷，国信证券经济研究所整理

图14：M30TB和传统的X22企业盘的参数对比

	Exos M 30TB	Exos X22
格式化容量	30 TB	22 TB
单碟容量	3 TB	2.2 TB
最大传输率	275 MB/s	285 MB/s
随机读 / 写 IOPS（4KB QD16）	170 / 350 IOPS	168 / 550 IOPS
平均时延	4.16 ms	4.16 ms
最大运行功耗（随机读 4KB QD16）	9.5 W	9.4 W
运行温度	10~60 °C	10~60 °C
运行震动	30 Gs	40 Gs

资料来源：希捷，国信证券经济研究所整理

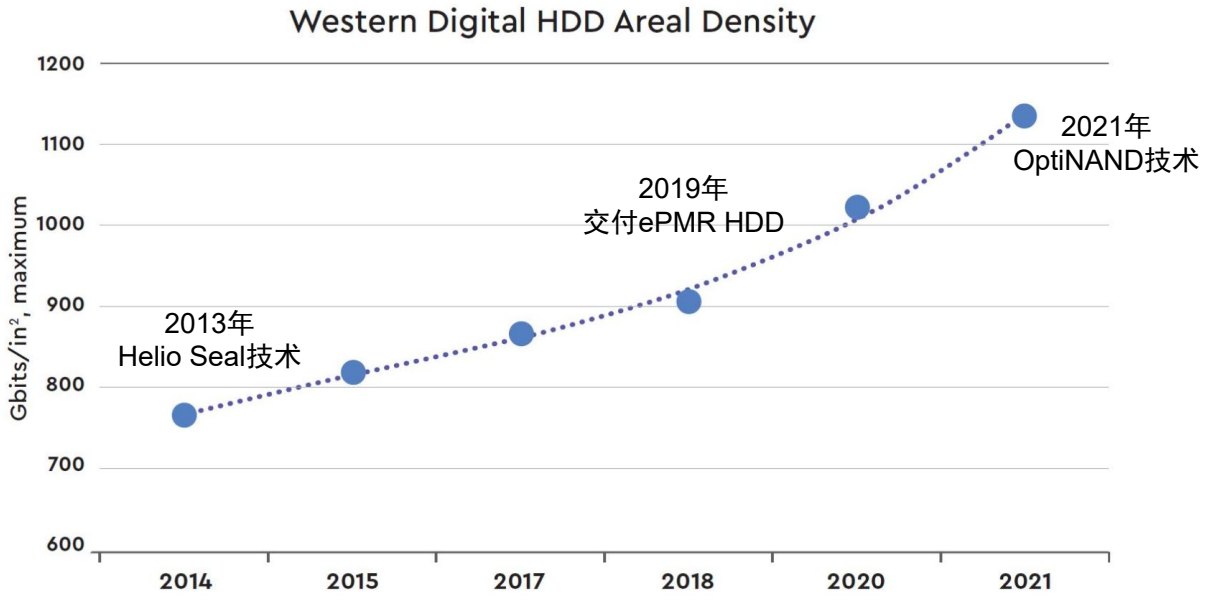
硬盘驱动器（HDD）技术趋势：西部数据OptiNAND技术

国信证券 GUOSEN SECURITIES

- **西部数据HDD磁盘面密度持续提升：**西部数据2013年率先对HDD进行密封封装，2013年发布氦气HDD（Helio Seal技术）；2019年率先交付ePMR HDD（能量辅助垂直磁记录）产品；2021年发布OptiNAND技术，进一步提升HDD磁盘面密度。

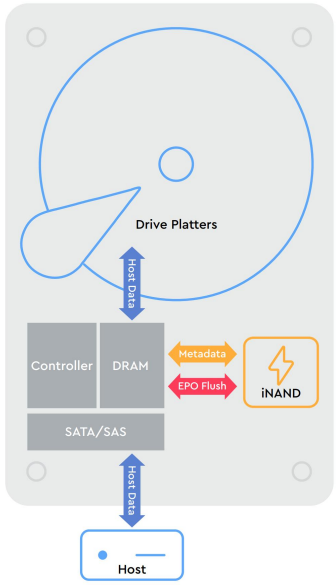
- **OptiNAND提升HDD面密度：**集成了iNAND UFS EFD、EFD（嵌入式闪存盘）和旋转型磁碟介质，同时对固件算法和SoC进行了革新，将寻道数据、定位数据等元数据记录在iNAND中（元数据量过大，无法以成本效益保存在DRAM中，若从磁盘检索又会干扰主机操作和性能）；与采用TSA（Triple Stage Actuator）技术相结合，实现更大TPI（每英寸磁道数量）和面密度（上一代HDD写入操作是以磁道为单位，因为刷新是对整个磁道进行，而OptisNAND技术在iNAND中记录了扇区级别的写入操作，则仅刷新扇区即可，通过消除过多的刷新，相邻磁道可以更靠近），在不需要增加碟片数量和磁头数量的情况下，实现容量的提升；此外，原来放置在碟片中的元数据，放置在iNAND中，腾出了碟片空间。

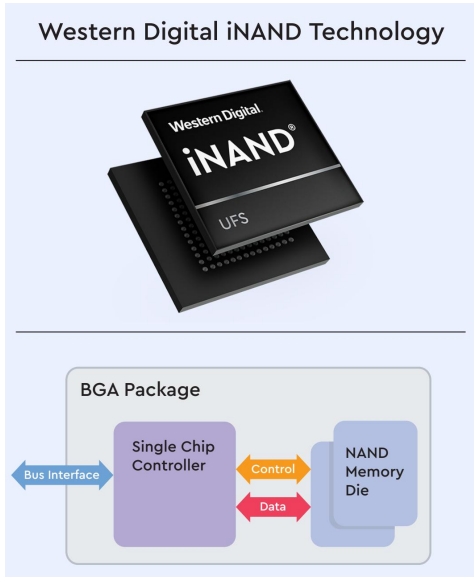
图15：西部数据HDD面密度持续提升



资料来源：西部数据，国信证券经济研究所整理

图16：西部数据OptisNAND技术





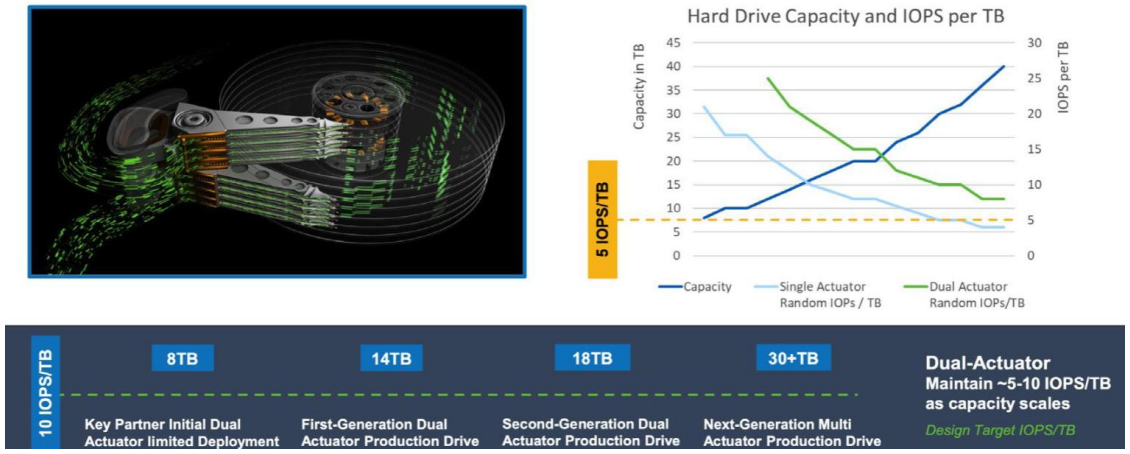
资料来源：西部数据，国信证券经济研究所整理

硬盘驱动器（HDD）技术趋势：多磁臂技术与生态统一

- **多磁臂技术：**随着硬盘容量的增加，单位容量的IOPS持续下降，而对于分布式存储来说，规模越大，低IOPS的危害也就越大，尤其是写入操作的过大延迟会拖累这个集群的响应能力；以希捷的双磁臂（MACH.2）为例，两组磁臂上下堆叠，共用一个枢轴，每组磁头除了拥有独立的音圈马达及驱动芯片，也对应各自独立的主控、缓存等（MACH.2在逻辑上就是2个硬盘，通过SAS协议的逻辑单元（LUN）功能（SATA接口版本不可以），在操作系统中显示为两个容量减半LUN），其可以让传输速率翻倍、读IPOS也接近翻倍，同时单位成本低于使用两块较小容量的硬盘。

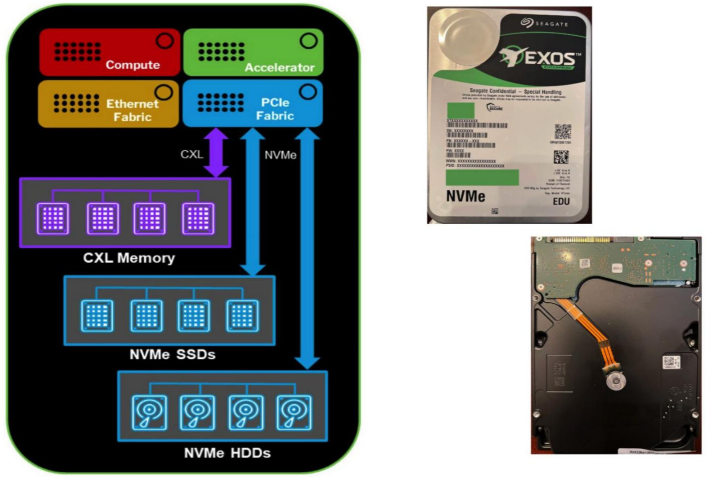
- **生态统一，使用NVMe协议：**目前SATA规范已经停止演进，SAS生态前景亦不如NVMe，统一到NVMe生态利于HDD发展。NVMe本为一种转为非易失性存储器设计的高性能、低延迟接口协议，主要用于使用PCIe总线的SSD，随着多磁臂的出现，硬盘最大传输率已经接近SATA接口的上限，且将SSD和硬盘接口协议统一，有更高的总线利用率，简化了存储的拓扑结构，利于与高性能存储网络解决方案的整合。

图17：HDD双磁臂可以扭转硬盘单位容量IOPS随容量提升的下降态势



资料来源：希捷，国信证券经济研究所整理

图18：NVMe协议的HDD



资料来源：益企研究院，国信证券经济研究所整理

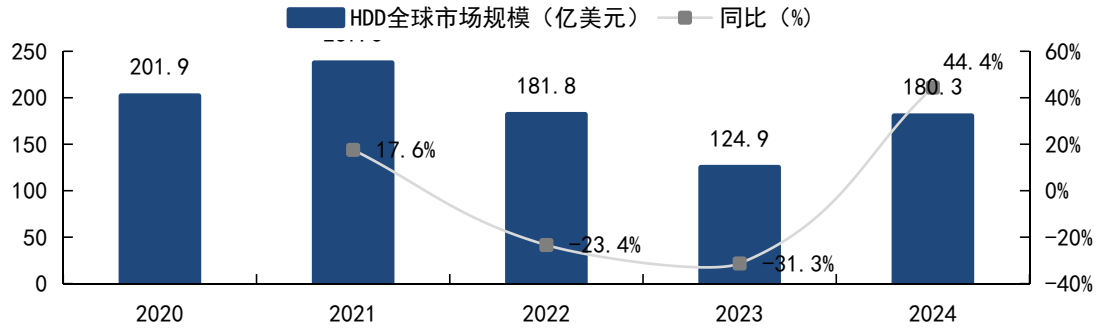
硬盘驱动器（HDD）供给侧：行业呈周期性波动，双寡头垄断

国信证券 GUOSEN SECURITIES

- **HDD行业呈现周期性波动：**受供求关系及下游囤货影响，HDD行业呈现周期波动，2022年行业开始进入下行周期，2023年见底，2024年开始企稳复苏，2025年受益AI基础设施建设驱动，需求端爆发式增长，短期供不应求，产品价格有望持续上行；根据IDC披露数据，2024年全球HDD行业市场规模为180.3亿美元，同比+44.4%；全球HDD出货量为1.24亿颗，同比+1.1%。

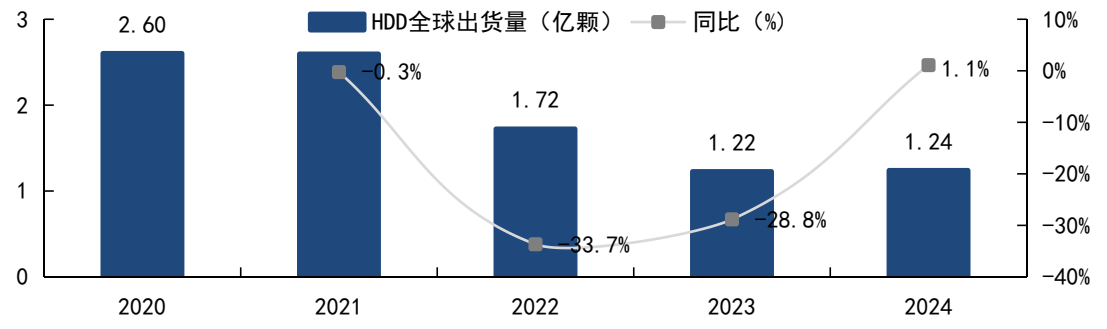
- **行业呈现双寡头垄断格局：**根据IDC披露数据，2024年，希捷、西部数据、东芝市占率分别为40.8%、40.0%、19.2%，其中希捷、西部数据市占率合计达80.8%，呈现双寡头垄断格局。

图19：HDD行业市场规模及增速（亿美元）



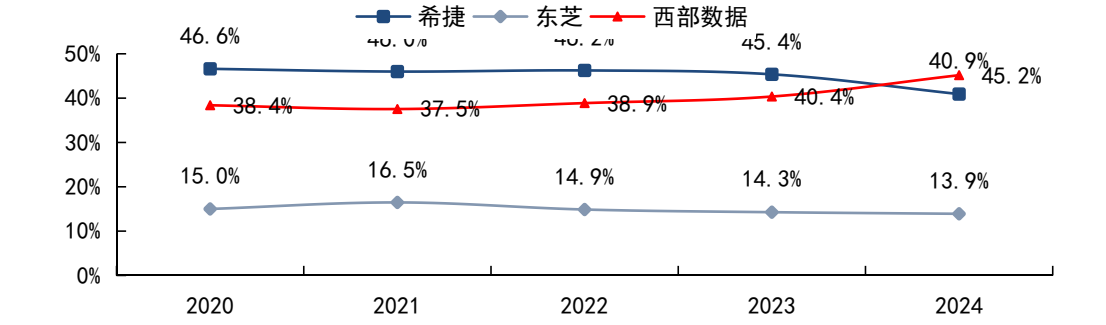
资料来源：IDC，国信证券经济研究所整理

图20：HDD行业出货量及增速（亿颗）



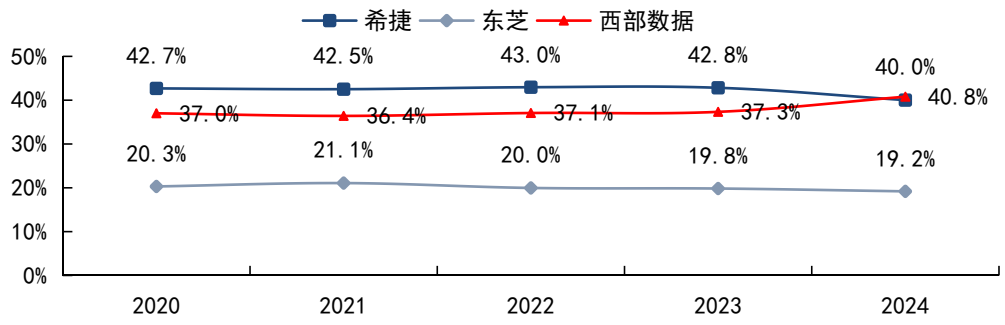
资料来源：IDC，国信证券经济研究所整理

图21：HDD行业竞争格局（按收入）



资料来源：IDC，国信证券经济研究所整理

图22：HDD行业竞争格局（按出货量）



资料来源：IDC，国信证券经济研究所整理

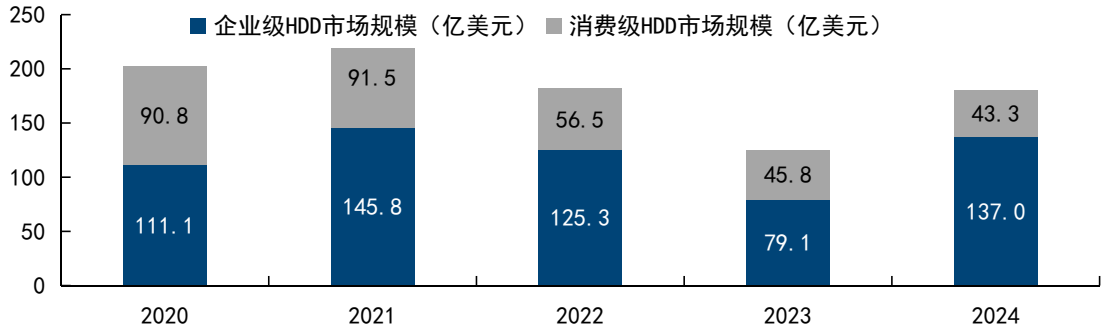
硬盘驱动器（HDD）供给侧：企业级HDD占比持续提升

国信证券 GUOSEN SECURITIES

- **企业级HDD占比持续提升：**根据IDC披露数据，2024年企业级HDD市场规模为137亿美金，占比约76%，相较于2020年，企业级HDD占比提升11个pct，占比持续提升，主要由于短视频等新兴媒体的快速发展以及AI需求的推动，企业级HDD需求提升。

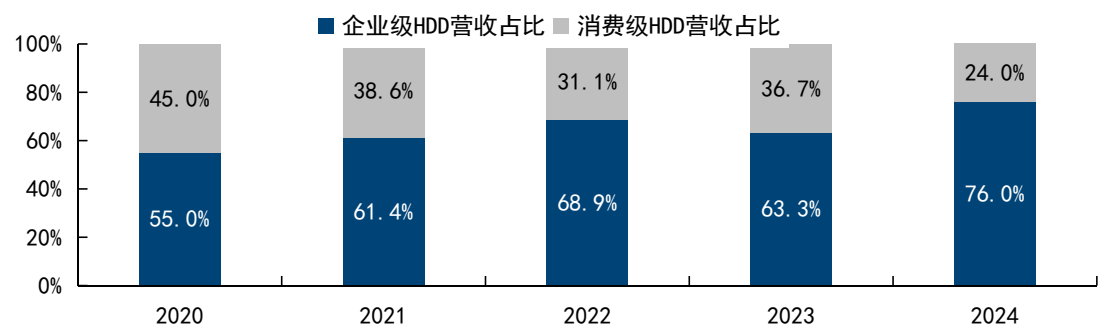
- **西部数据企业级HDD市占率持续提升：**根据IDC披露数据，2024年，希捷、西部数据、东芝市占率分别为40.7%、46.5%、12.9%，相较于2023年，分别-7.0、+7.5、-0.3个pct，西部数据企业级HDD市占率持续提升；从各家产品结构来看，2024年希捷、西部数据、东芝企业级HDD收入占比分别为75.5%、78.2%、70.3%，贡献公司主要的HDD收入。

图23：企业级HDD和消费级HDD市场规模（亿美元）



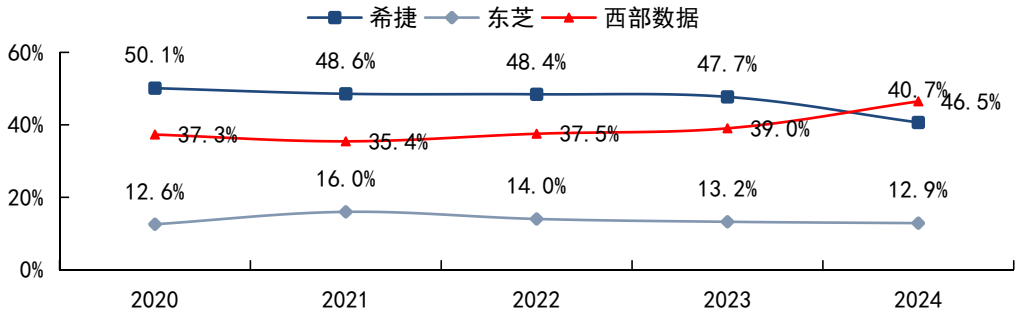
资料来源：IDC，国信证券经济研究所整理

图24：企业级HDD占比持续提升



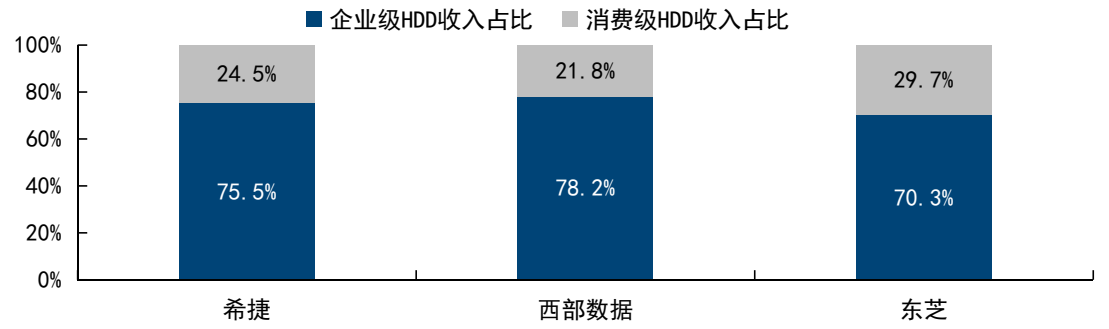
资料来源：IDC，国信证券经济研究所整理

图25：企业级HDD市占率情况



资料来源：IDC，国信证券经济研究所整理

图26：希捷、西部数据、东芝企业级HDD收入占比



资料来源：IDC，国信证券经济研究所整理

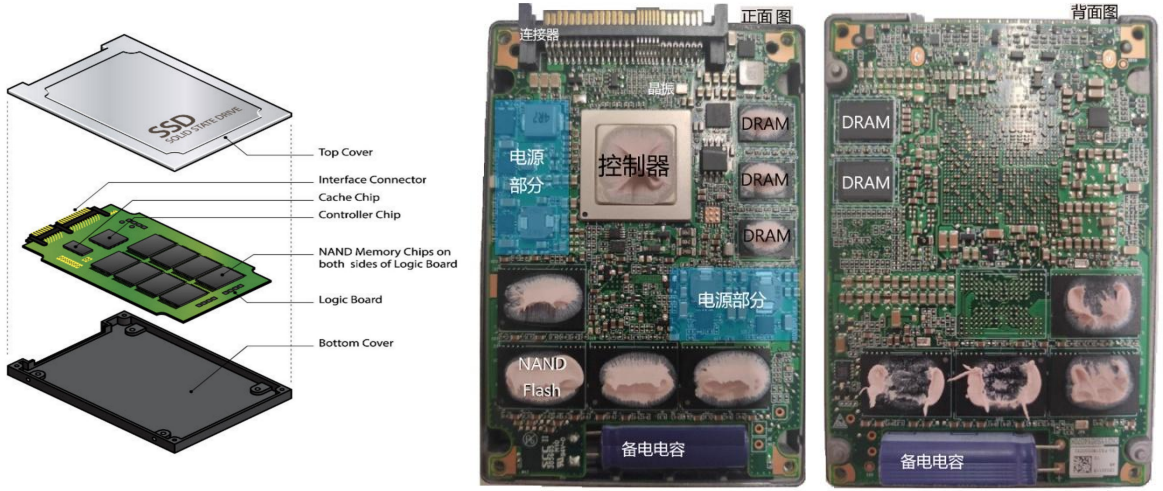
固态硬盘（SSD）：以NAND Flash为介质的存储设备

国信证券 GUOSEN SECURITIES

- **固态硬盘（SSD）结构**：SSD是一种以NAND Flash为介质的存储设备，其与HDD不同，SSD以半导体存储数据，用纯电子电路实现，其主要包括主控（Controller Chip）、闪存（NAND）、缓存芯片（Cache Chip，部分SSD只配置SRAM，未配置DRAM）、PCB（电源芯片、电阻、电容等）、接口（SATA、SAS、PCIe等）。

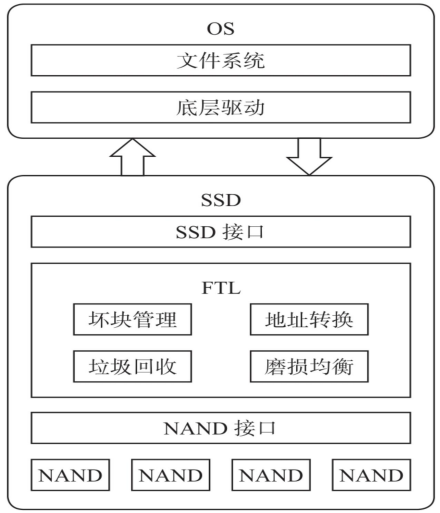
- **固态硬盘（SSD）工作原理**：从主机开始，用户从操作系统应用层面对SSD发出请求，文件系统将读写请求经驱动转化为相应的符合协议的读写和其他命令，SSD收到命令后执行相应操作，然后输出结果。**SSD的输入是命令（Command），输出的是数据（Data）和命令状态（Command Status）**。
  - **前端通信和相关协议模块**：接受主机发来的命令和数据，命令经SSD处理后，交由前端将命令状态或数据返回主机，SSD通过SATA、SAS和PCIe等接口与主机相连；
  - **中间的FTL（Flash Translation Layer）模块**：SSD收到命令后执行命令，并接收主机要写入的数据，数据一般会先缓存在SSD内部的RAM中，FTL会为每个逻辑数据块分配一个闪存地址，当数据到达一定量后，FTL便会给后端发送写闪存情况；
  - **后端和闪存通信模块**：接受FTL写闪存请求后，会将缓存中的数据写到对应的闪存空间，由于闪存不能覆盖写，所以闪存块需要擦除才能写入，主机发来的某个数据块不是写在闪存固定位置，SSD可以为其分配任何可能的闪存空间供其写入，所以需要FTL完成逻辑数据块到闪存物理空间的转换或映射。

图27：SSD结构



资料来源：Stephanie等著-《Object Storage，Persistent Memory，and Data Infrastructure for HPC Materials Informatics》-arXiv（2022）-P8、舒继武著-《数据存储架构与技术》-人民邮电出版社（2024年）-P37，国信证券经济研究所整理

图28：SSD工作原理



资料来源：SSDFans等著-《深入浅出SSD：固态存储技术核心、原理与实战》-机械工业出版社（2023年）-P40，国信证券经济研究所整理

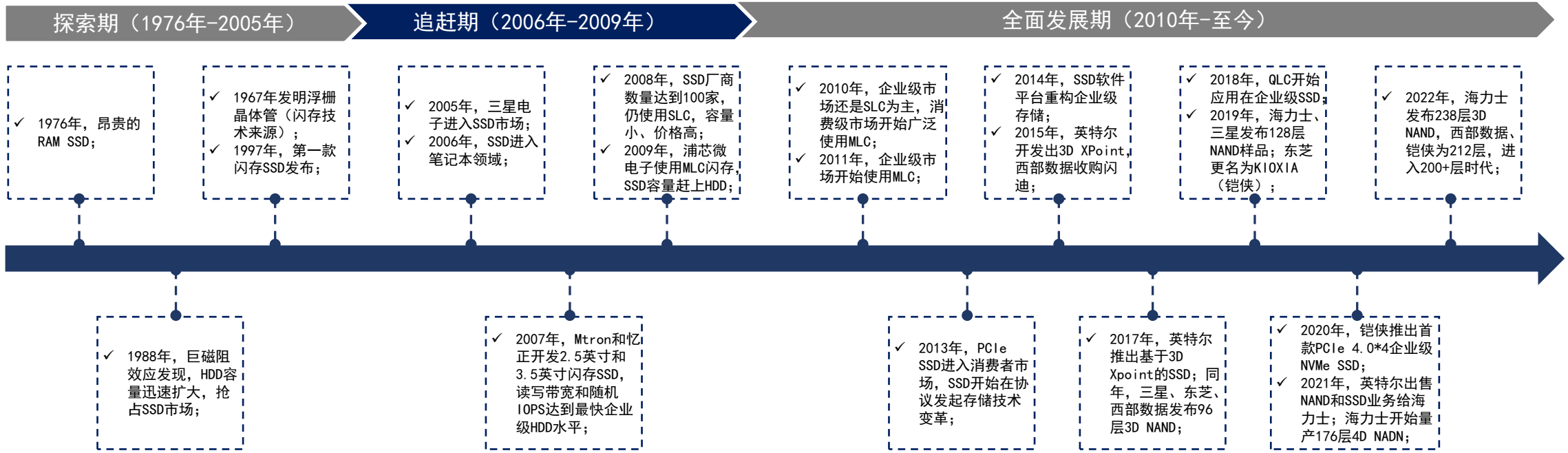
固态硬盘（SSD）发展历史：从追赶HDD到自我迭代升级

国信证券 GUOSEN SECURITIES

■ **固态硬盘（SSD）发展历史：**固态硬盘（SSD）发展主要经过探索期、追赶期、全面发展期三个阶段。

- **探索期（1976年-2005年）**：1976年第一款RAM SSD发布，价格昂贵；后随着闪存技术的发展，1997年第一款闪存SSD发布；
- **追赶期（2006年-2009年）**：2007年SSD在读写带宽和随机IOPS达到HDD水平；2009年通过MLC闪存技术，SSD容量赶上HDD；
- **全面发展期（2010年-至今）**：1）单元架构：从SLC逐步转化为MLC、QLC；2）协议：从最初和HDD共用SATA协议，2013年率先在消费者市场发布PCIe SSD（即NVMe协议）；3）NAND层数持续增长：2019年NAND层数突破100层，2022年SSD开始进入200+层时代。

图29：SSD发展历史



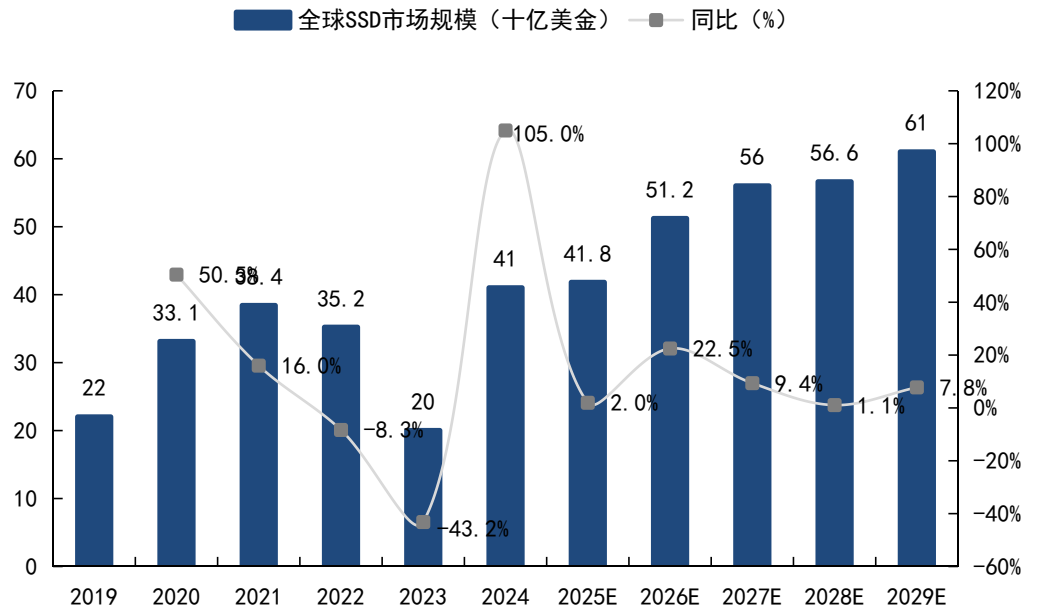
资料来源：SSDFans等著-《深入浅出SSD：固态存储技术核心、原理与实战》-机械工业出版社（2023年）-P26，国信证券经济研究所整理

固态硬盘（SSD）市场规模：410亿美金市场，市场规模稳步增长

国信证券 GUOSEN SECURITIES

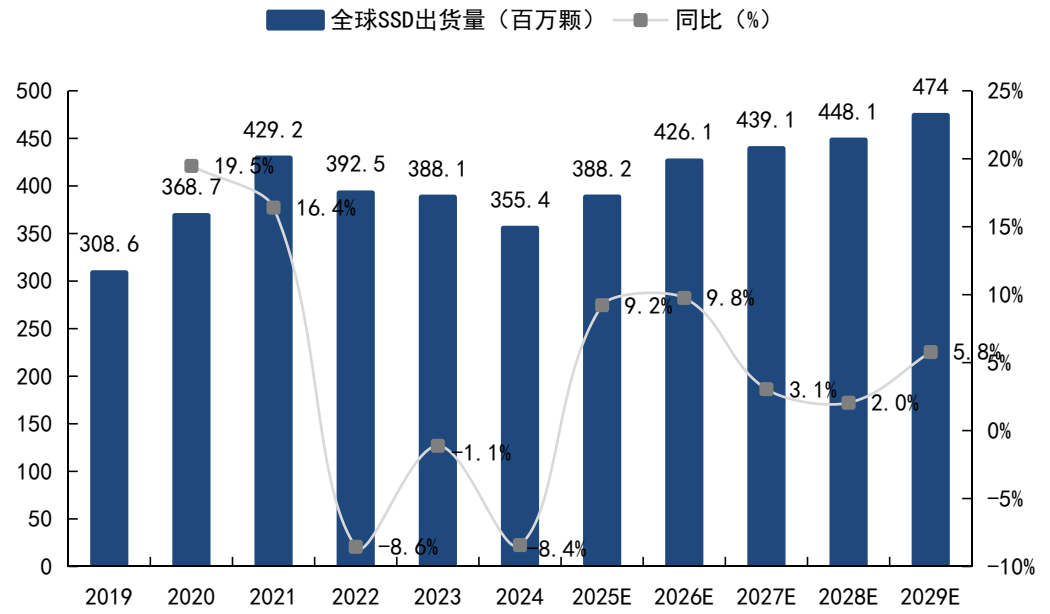
- **固态硬盘（SSD）市场规模：**根据IDC披露数据，2024年全球SSD市场规模为410亿美金，预计2029年增长至610亿美金，对应24-29年CAGR为8.3%。
- **固态硬盘（SSD）出货量情况：**根据IDC披露数据，2024年全球SSD市场出货量为3.55亿个，预计2029年增长至4.74亿个，对应24-29年CAGR为5.93%。

图30：全球SSD市场规模及增速情况



资料来源：IDC，国信证券经济研究所整理

图31：全球SSD出货量及增速情况



资料来源：IDC，国信证券经济研究所整理

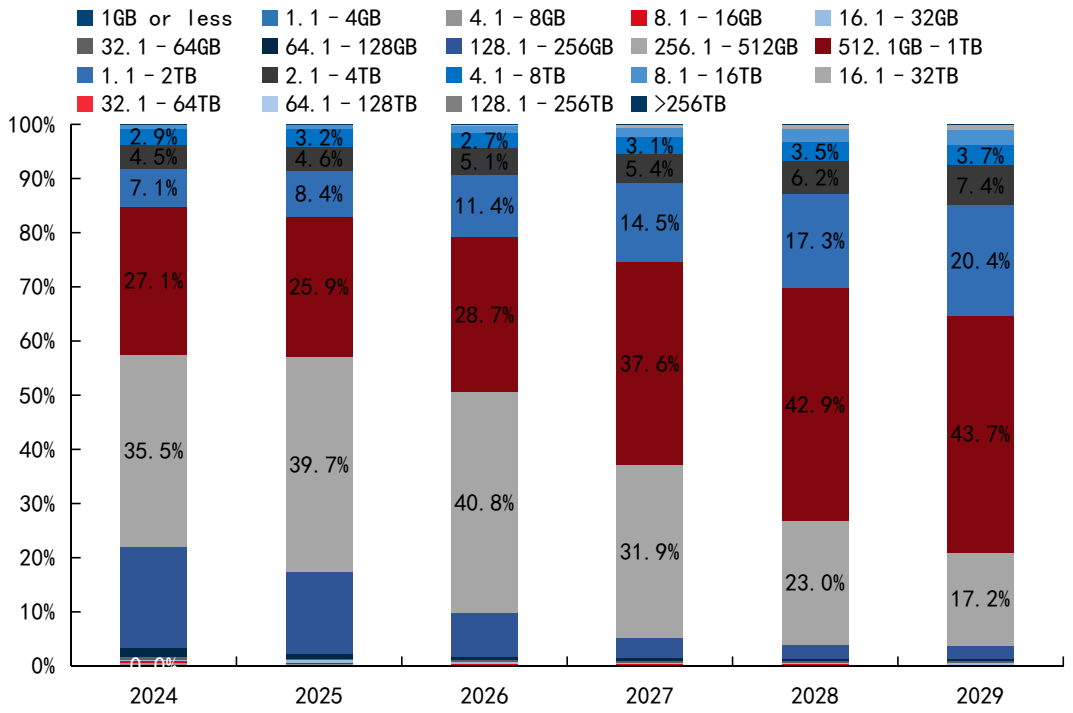
固态硬盘（SSD）：高容量产品占比提升，单位GB价格持续下降

国信证券 GUOSEN SECURITIES

- **固态硬盘（SSD）产品容量：**根据IDC披露数据，2024年SSD产品主要以256GB-512GB、512GB-1TB、1TB-2TB产品为主，占比分别为35.5%、27.1%、7.1%，合计占比为69.7%；未来，高容量SSD出货量占比持续提升，预计2029年上述三类产品占比分别为17.2%、43.7%、20.4%，分别-18.3、+16.6、+13.3个pct。

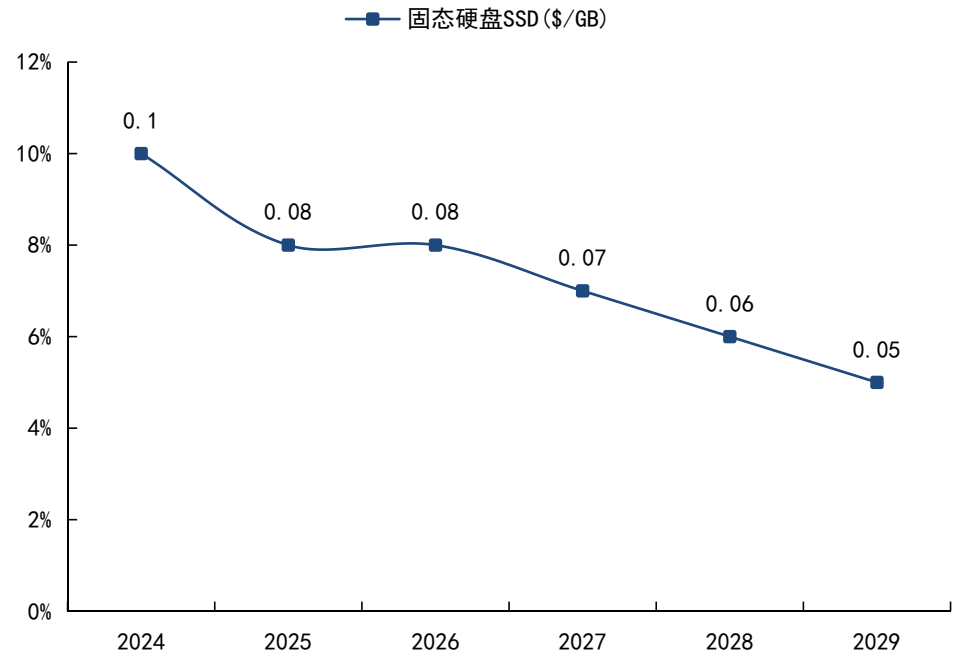
- **固态硬盘（SSD）单位GB价格：**根据IDC披露数据，2024年固态硬盘SSD单位GB价格为0.1美金，预计2029年将下降为0.05美金，单位GB价格持续下降。

图32：固态硬盘(SSD）高容量产品出货量占比持续提升



资料来源：IDC，国信证券经济研究所整理

图33：固态硬盘(SSD）单位GB价格持续下降



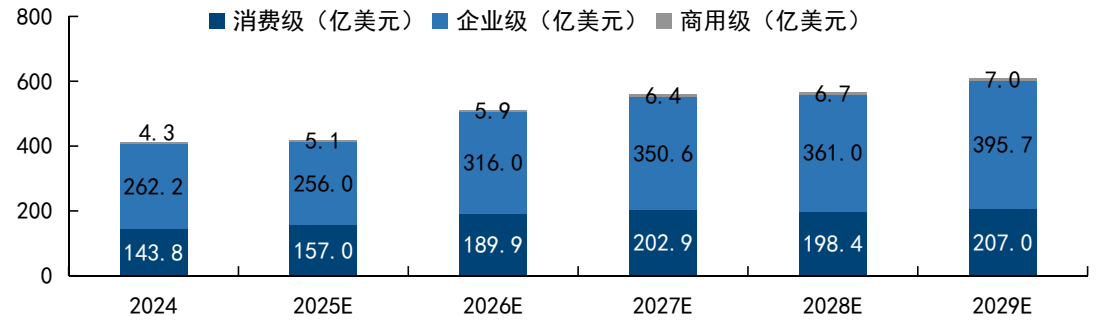
资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）：企业SSD市场占比较高，消费SSD出货量较大

国信证券 GUOSEN SECURITIES

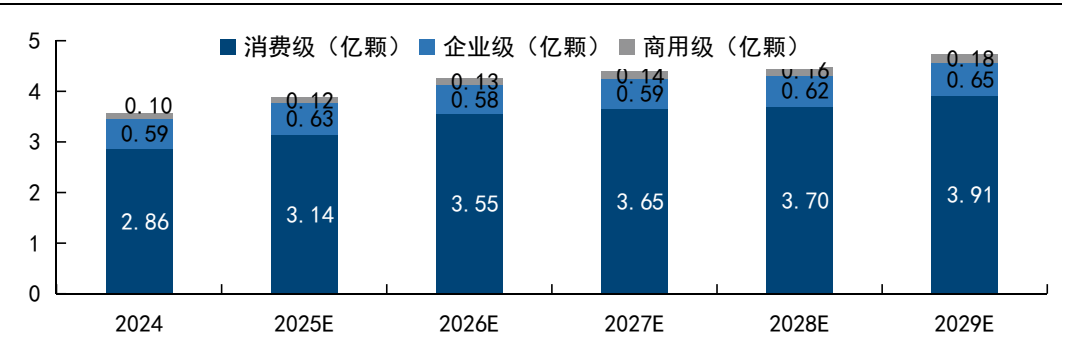
- **固态硬盘（SSD）细分市场情况：**根据IDC披露数据，2024年消费级、企业级、商用级市场规模分别为143.8、262.2、4.3亿美金，占比分别为35%、64%、1%，企业级SSD市场占比较高；从出货量来看，2024年消费级、企业级、商用级SSD出货量分别为81%、17%、3%，消费级SSD出货量较大，但单价较低。
  - **消费级SSD：**主要包括便携式PC、台式PC和消费电子用SSD市场，主要通过分销、零售渠道出货，增长动力来自PC从机械硬盘向SSD转换；
  - **企业级SSD：**主要包括服务器、外部存储系统用SSD，人工智能对存储的需求拉动企业级SSD需求增长；
  - **商用级SSD：**涵盖工业设备、工厂自动化、医疗设备、军事、航空航天等领域对SSD的需求。

图34：固态硬盘（SSD）细分市场收入情况（亿美元）



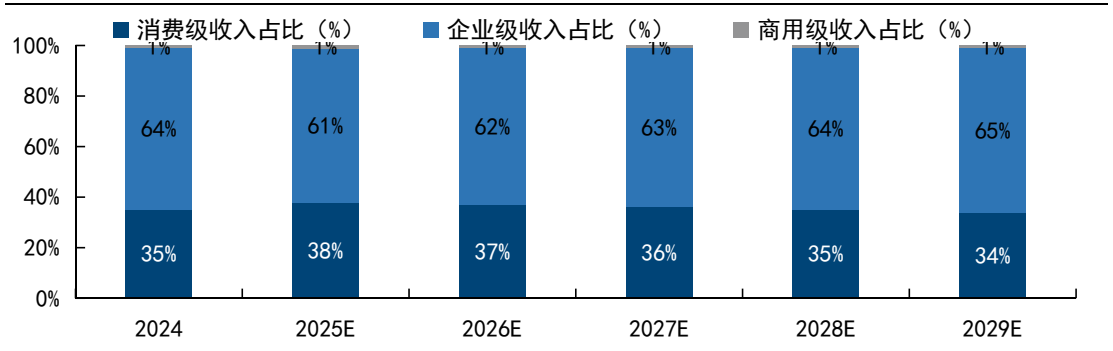
资料来源：IDC，国信证券经济研究所整理

图35：固态硬盘（SSD）细分市场出货量情况（亿颗）



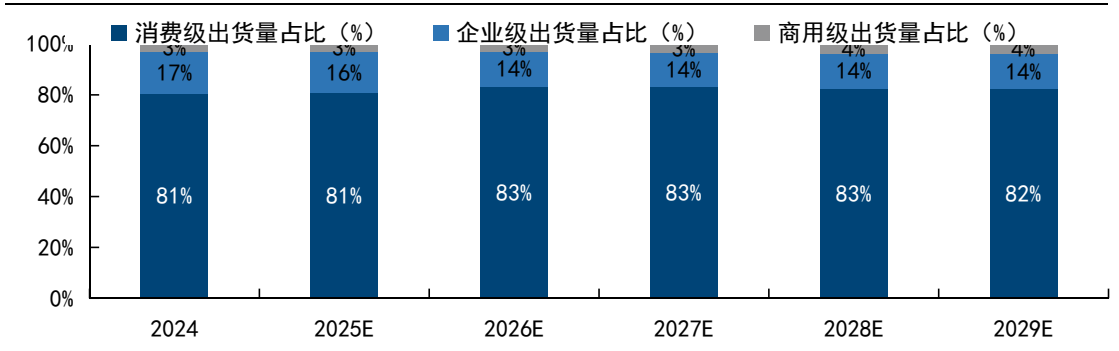
资料来源：IDC，国信证券经济研究所整理

图36：固态硬盘（SSD）细分市场收入占比情况



资料来源：IDC，国信证券经济研究所整理

图37：固态硬盘（SSD）细分市场出货量占比情况



资料来源：IDC，国信证券经济研究所整理

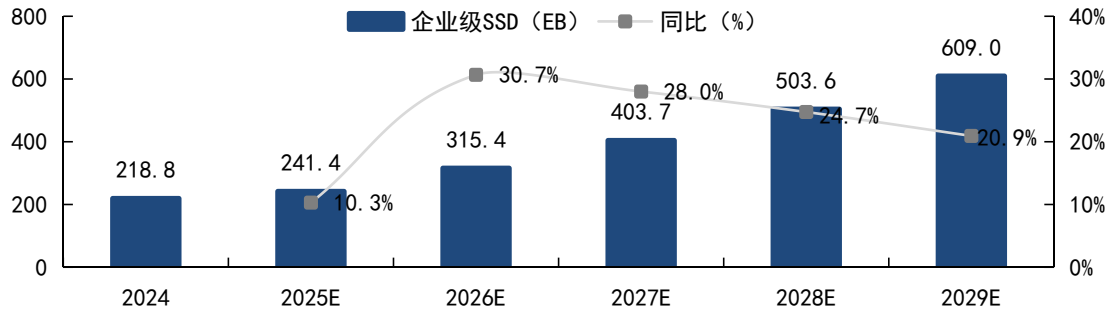
固态硬盘（SSD）-企业级：高容量SSD出货量占比持续提升

国信证券 GUOSEN SECURITIES

- **高容量企业级固态硬盘（SSD）占比持续提升：**根据IDC披露数据，2024年企业级SSD出货容量为218.8EB，预计2029年提升至609.0EB，对应24-29年CAGR为22.7%；2024年512GB-1TB、1TB-2TB、2TB-4TB、4TB-8TB出货量占比较高，分别为1507、941、1369、973万颗，占比分别为25.4%、15.9%、23.1%、16.4%，合计占比为80.8%，未来高容量SSD出货量占比持续提升，8T-16T版本SSD出货占比有望从2024年的4.1%提升至2029年19.9%。

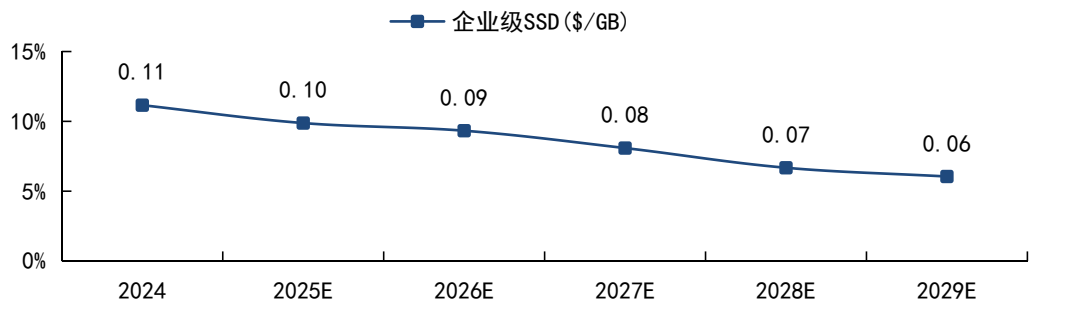
- **企业级固态硬盘(SSD）单GB价格持续下滑：**随着技术的迭代，单颗SSD硬盘的容量持续提升，则单GB对应的价格将持续下滑，根据IDC数据，2024年企业级SSD单GB价格为0.11美元/GB，预计2029年下降至0.06美元/GB。

图40：企业级固态硬盘（SSD）容量出货量情况（单位：EB）



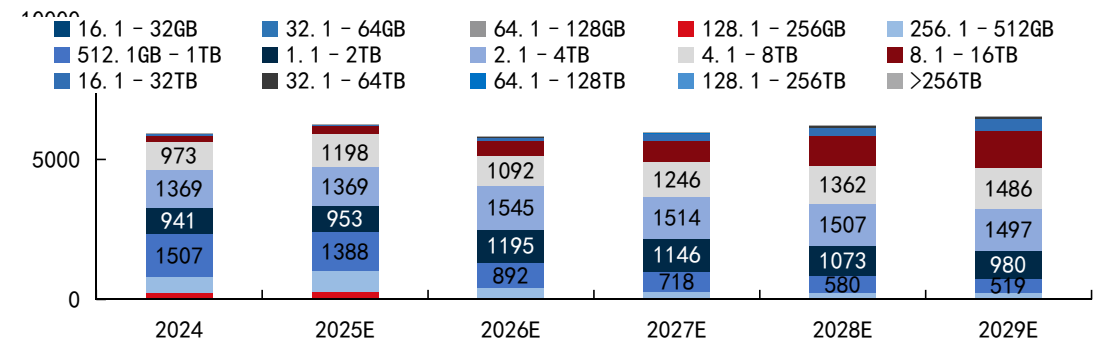
资料来源：IDC，国信证券经济研究所整理

图41：企业级固态硬盘（SSD）单GB价格持续下滑



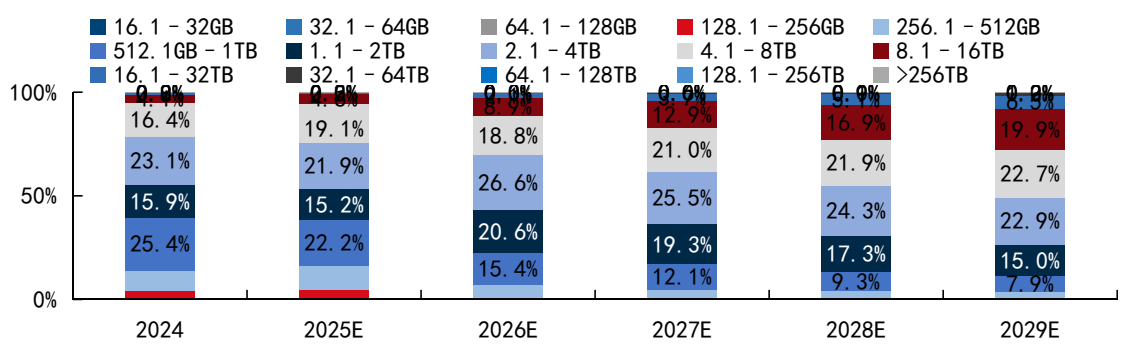
资料来源：IDC，国信证券经济研究所整理

图42：企业级固态硬盘（SSD）各容量出货情况



资料来源：IDC，国信证券经济研究所整理

图43：企业级固态硬盘（SSD）各容量出货占比情况



资料来源：IDC，国信证券经济研究所整理

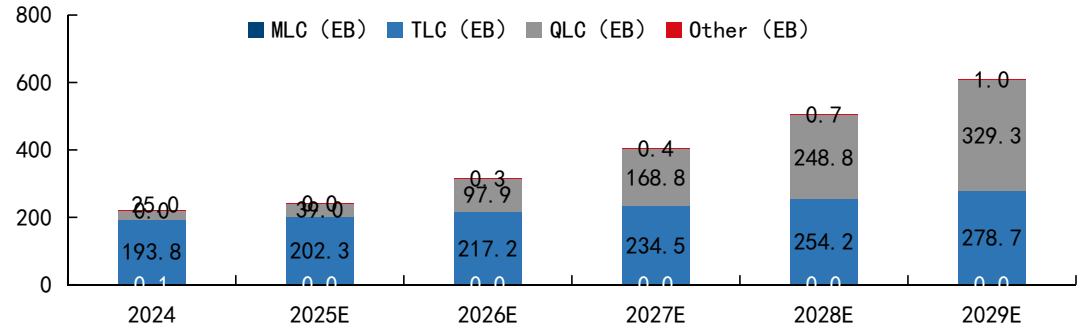
固态硬盘（SSD）-企业级：QLC出货容量快速提升，PCIe接口成为主流

国信证券 GUOSEN SECURITIES

- **QLC版本企业级固态硬盘（SSD）出货容量快速提升：**根据IDC披露数据，2024年TLC、QLC版本SSD出货容量分别为193.8、25.0EB，占比分别为88.5%、11.4%；在存储密度持续提升的大背景下，存储4位数据的QLC（Quad Level Cell）出货量占比有望快速提升，根据IDC预测数据，2029年QLC出货容量有望达到329.3EB，占比提升至54.1%。

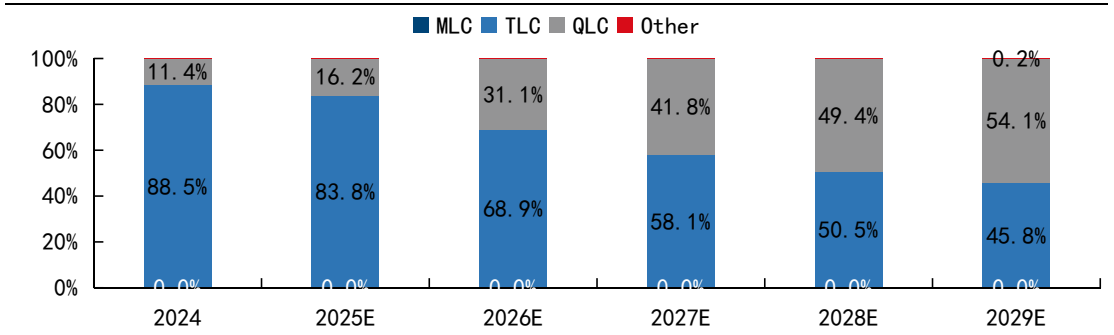
- **PCIe接口固态硬盘(SSD）成为主流：**传统的SATA接口限制了SSD带宽，以NVMe协议为代表的PCIe接口SSD逐步成为主流，根据IDC披露数据，2024年PCIe版本企业级SSD出货量达4365万颗，占比为73.7%，预计2029年出货量达6200万颗，占比提升至94.8%。

图44：QLC企业级固态硬盘（SSD）出货容量将快速提升（单位：EB）



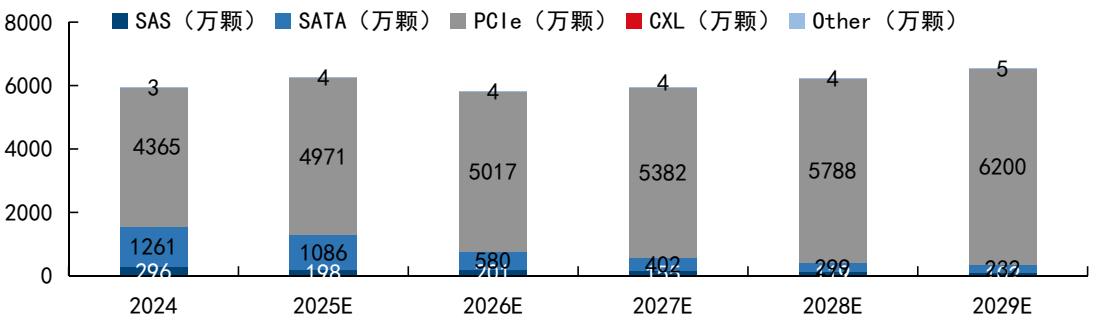
资料来源：IDC，国信证券经济研究所整理

图45：QLC企业级固态硬盘（SSD）出货容量占比将快速提升



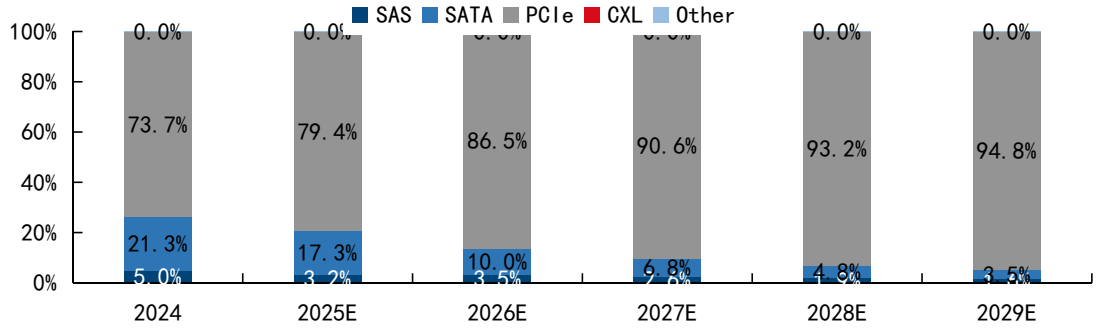
资料来源：IDC，国信证券经济研究所整理

图46：企业级固态硬盘（SSD）PCIe接口成为主流



资料来源：IDC，国信证券经济研究所整理

图47：企业级固态硬盘（SSD）PCIe接口出货量占比持续提升



资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）vs硬盘驱动器（HDD）：SSD性能占优，HDD高性价比

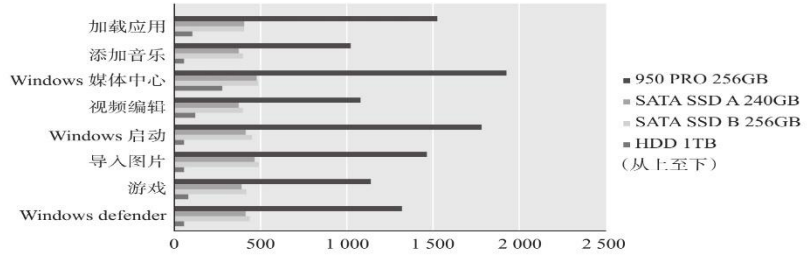
国信证券 GUOSEN SECURITIES

- **固态硬盘（SSD）在性能、功耗、抗震防摔、噪声、尺寸等维度具备优势。**
  - **性能**：SSD在连续读写、随机读写速度显著快速HDD；同时，SSD数据访问时间大幅低于HDD；
  - **功耗**：通常使用功耗/IOPS（即单位IOPS的功耗输出）衡量存储器的功耗水平，由于SSD性能较高（IOPS值高），SSD的功耗/IOPS较低，功耗表现出色；
  - **抗震防摔**：SSD内部不存在机械部件，相比于HDD更加抗震；
  - **噪声**：SSD内部结构无马达，不存在旋转噪音；
  - **尺寸**：HDD一般只有2.5、3.5英寸形式，SSD具有更小的尺寸，例如可以贴在主板上的M.2；
- **硬盘驱动器（HDD）在价格维度存在优势。**从单GB价格来看，根据IDC统计数据，2024年企业级SSD单GB价格是HDD的9.0x，HDD价格优势明显。

图48：HDD与SSD性能对比

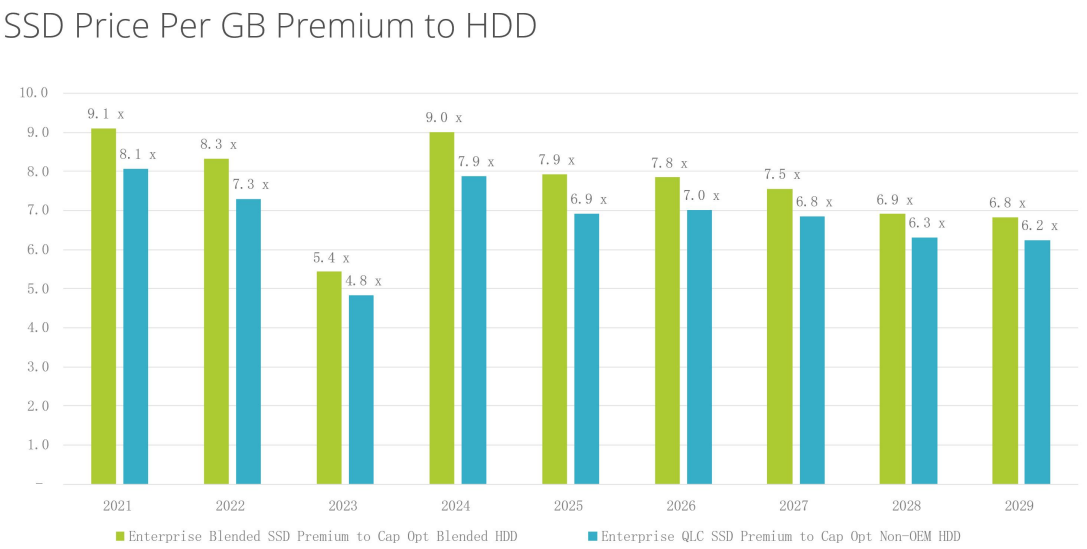
对比项	SATA SSD（500GB）	SATA HDD（500GB 7 200rpm）	差　异
介质	闪存	磁盘	—
连续读写/(MB/s)	540/330①	160/60	3 倍/6 倍
随机读写/IOPS	98 000/70 000	450/400	217 倍/175 倍
数据访问时间/ms	0.1	10 ～ 12	100 ～ 120 倍
性能得分（基于 PCMark）	78 700	5 600	14 倍

对比项	峰值功耗/W	读写功耗/W	睡眠功耗/mW	深度睡眠功耗/mW
HDD	8	6	500	不支持
SATA SSD	6	5	100	5
PCIe SSD	25	15	200	10



资料来源：SSDFans等著-《深入浅出SSD：固态存储技术核心、原理与实战》-机械工业出版社（2023年）-P23，国信证券经济研究所整理

图49：HDD与SSD价格倍数关系（SSD/HDD-价格/GB）



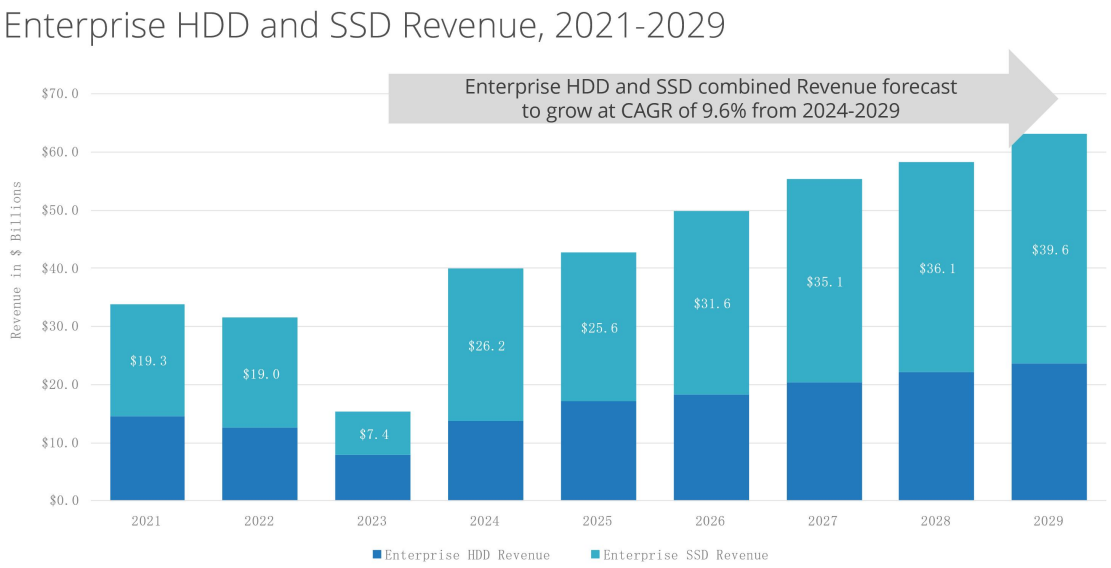
资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）vs硬盘驱动器（HDD）：SSD占比稳步提升

国信证券 GUOSEN SECURITIES

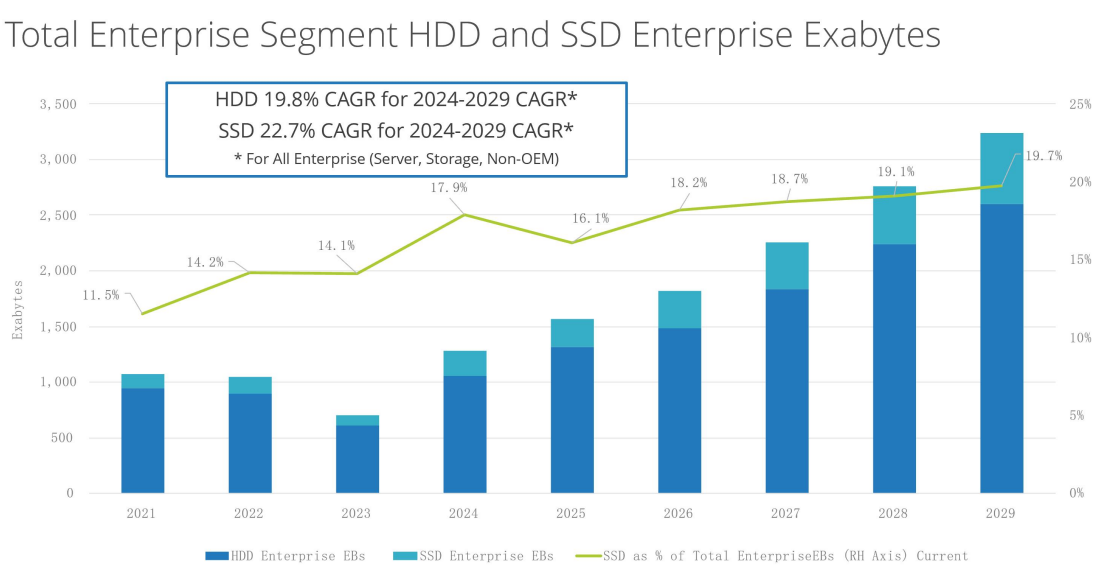
- **企业级固态硬盘（SSD）市场规模：**根据IDC披露数据，2024年企业级SSD市场规模为262亿美金，预计2029年增长至396亿美金，对应24-29年CAGR为8.6%。
- **企业级SSD占比持续提升：**根据IDC披露数据，2024年企业级SSD（以Exabytes计）占比为17.9%，预计2029年提升至19.7%，稳步提升。

图50：企业级SSD市场规模（十亿美金）和增速



资料来源：IDC，国信证券经济研究所整理

图51：企业级SSD占比稳步提升



资料来源：IDC，国信证券经济研究所整理

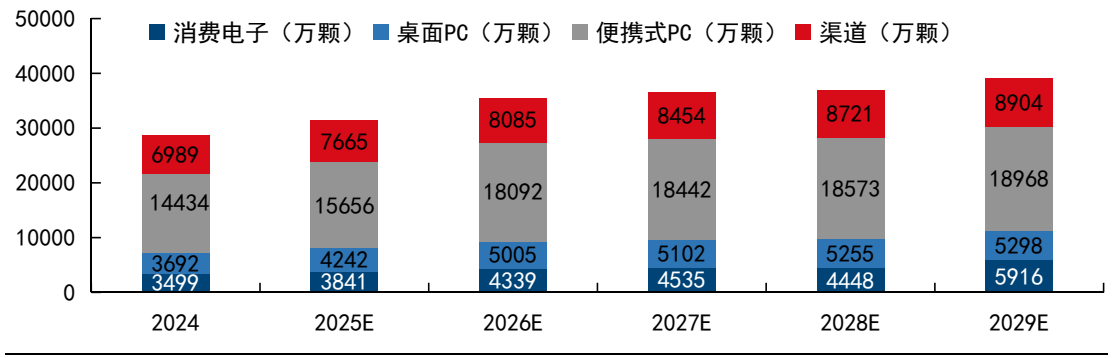
固态硬盘（SSD）-消费级：24年PC用SSD收入占比约81%

国信证券 GUOSEN SECURITIES

■ **消费级SSD细分市场情况：**消费级SSD下游市场主要包括PC和消费电子，其中PC又可细分为桌面PC、便携式PC和渠道。

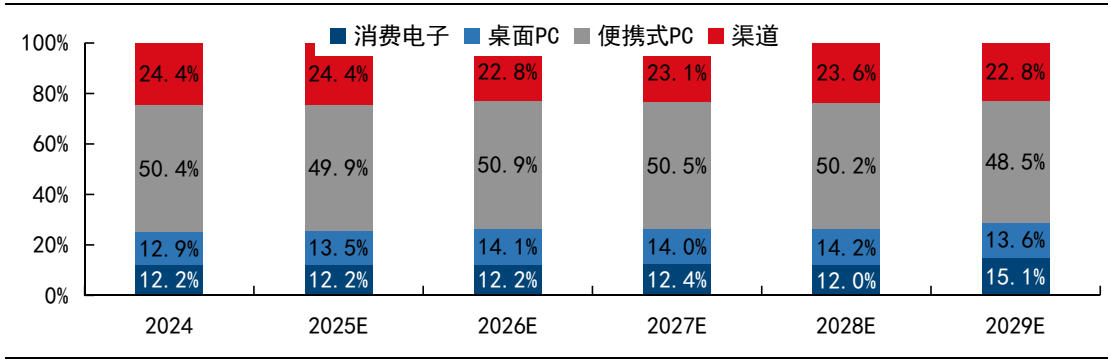
- **出货量维度：**根据IDC披露数据，2024年消费级SSD出货量为2.86亿颗，消费电子、桌面PC、便携式PC、渠道SSD出货量分别为3499、3692、14434、6989万颗，占比分别为12.2%、12.9%、50.4%、24.4%；
- **收入维度：**2024年PC用SSD市场规模约为117.1亿美金，消费电子用SSD市场规模约26.6亿美金，占比分别为81%、19%，PC用SSD仍占据主要市场份额；
- **变化趋势：**随着主机游戏市场存储和个人存储市场需求的增长，预计消费电子用SSD收入占比稳步提升。

图52：消费级SSD细分领域出货情况（单位：万颗）



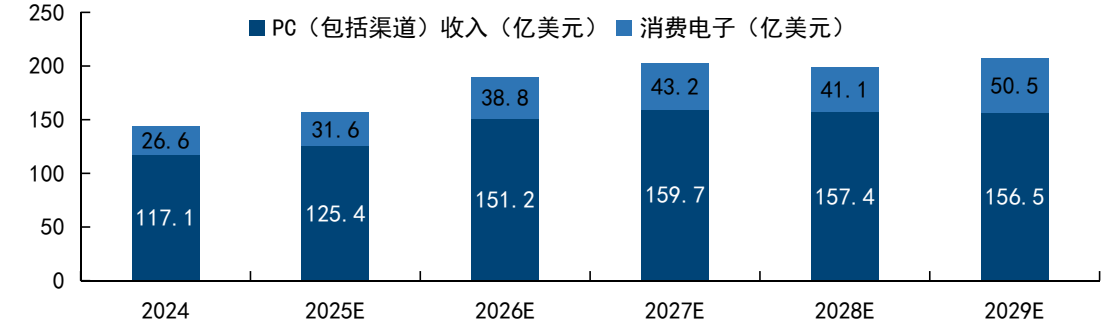
资料来源：IDC，国信证券经济研究所整理

图53：消费级SSD细分领域出货占比情况



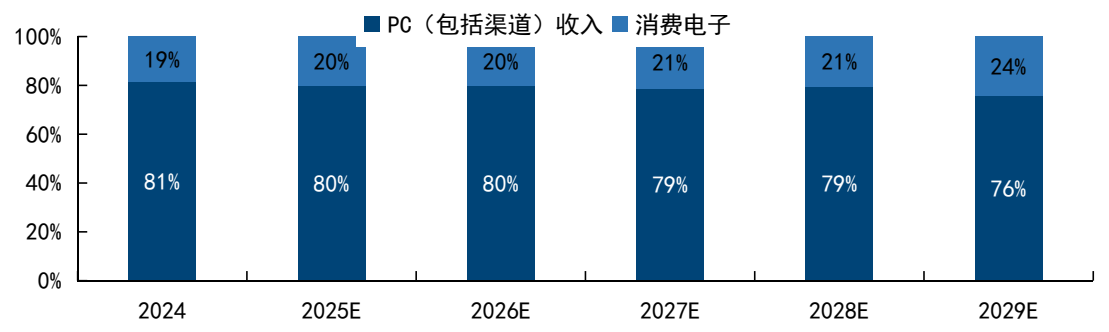
资料来源：IDC，国信证券经济研究所整理

图54：消费级SSD细分领域市场规模情况（单位：亿美元）



资料来源：IDC，国信证券经济研究所整理

图55：消费级SSD细分领域收入占比情况



资料来源：IDC，国信证券经济研究所整理

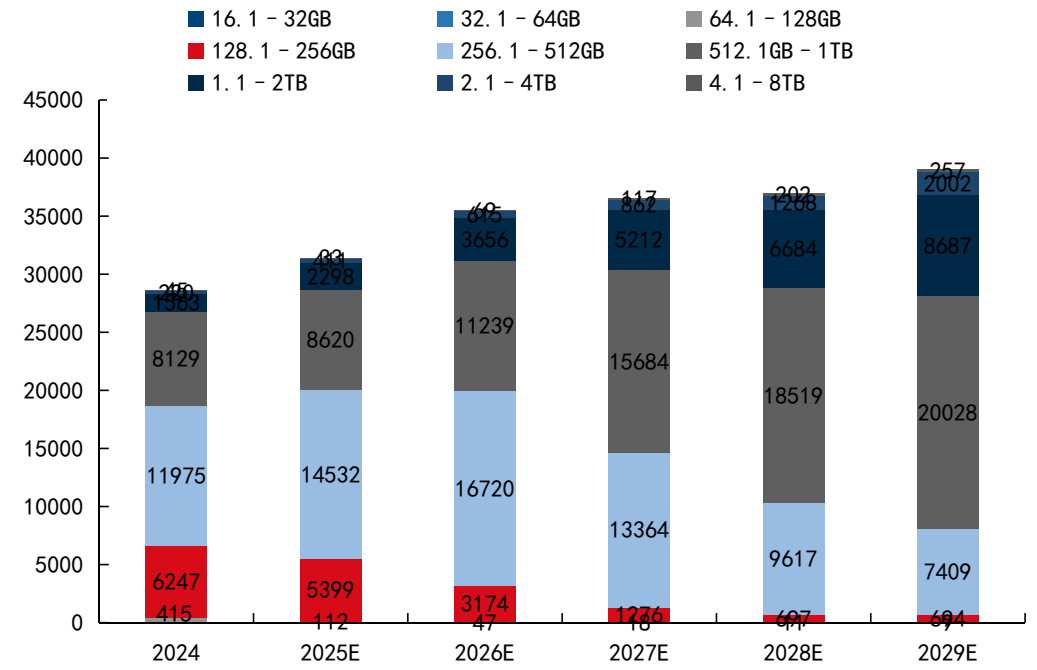
固态硬盘（SSD）-消费级：出货容量集中在128GB-1TB区间，高容量占比提升

国信证券 GUOSEN SECURITIES

- **消费级SSD出货容量主要集中在128GB-1TB区间：**根据IDC披露数据，2024年128GB-256GB、256GB-512GB、512GB-1TB的消费级SSD出货量分别为0.62、1.20、0.81亿颗，出货量占比分别为21.8%、41.9%、28.4%，合计占比为92.1%。

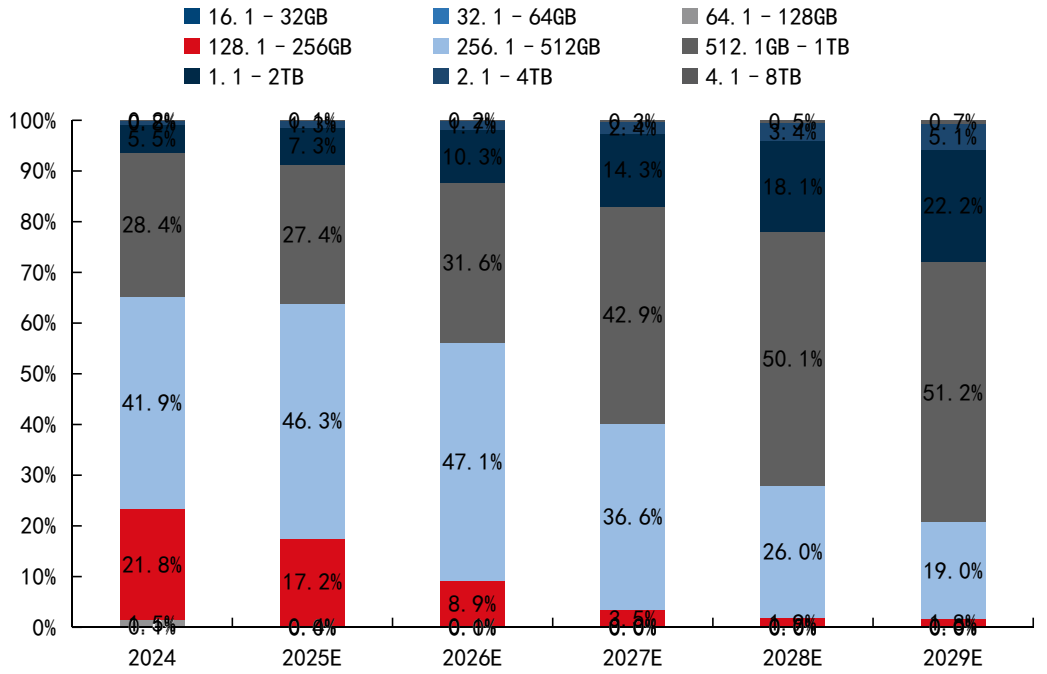
- **消费级SSD整体向高容量方向发展：**根据IDC预测数据，128GB-256GB消费级SSD出货量占比，预计从2024年的21.8%下降至2029年1.8%；同时，1TB-2TB高容量版本SSD出货量占比，预计从2024年的5.5%提升至2029年22.2%。

图56：消费级SSD各容量出货情况（单位：万颗）



资料来源：IDC，国信证券经济研究所整理

图57：消费级SSD各容量出货占比情况



资料来源：IDC，国信证券经济研究所整理

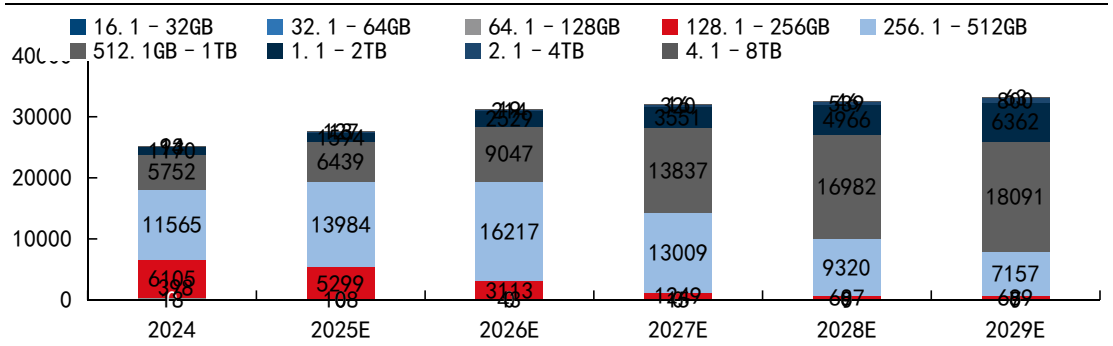
固态硬盘（SSD）-消费级：消费电子用SSD平均容量高于PC

国信证券 GUOSEN SECURITIES

■ **消费级SSD细分市场容量出货情况：**消费级电子用SSD平均容量高于PC用SSD。

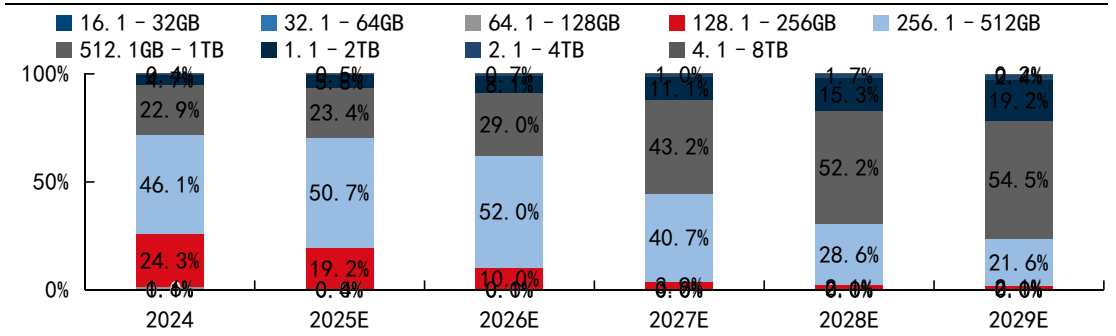
- **消费级-PC用SSD：**根据IDC披露数据，2024年PC用SSD出货量主要集中在128GB-256GB、256GB-512GB、512GB-1TB，出货量分别为0.61、1.16、0.58亿颗，占比分别为24.3%、46.1%、22.9%，基于单位GB价格下降和存储数据量的增长，高容量1TB-2TB的占比逐步提升；
- **消费级-消费电子用SSD：**根据IDC披露数据，2024年消费电子用SSD出货量主要集中在512GB-1TB、1TB-2TB，出货量粉笔为0.24、0.04亿颗，占比分别为67.9%、11.2%，消费电子用SSD平均容量高于PC用SSD，预计未来1TB-2TB、2TB-4TB容量SSD成为消费电子主流需求。

图58：消费级-PC用SSD各容量出货情况（单位：万颗）



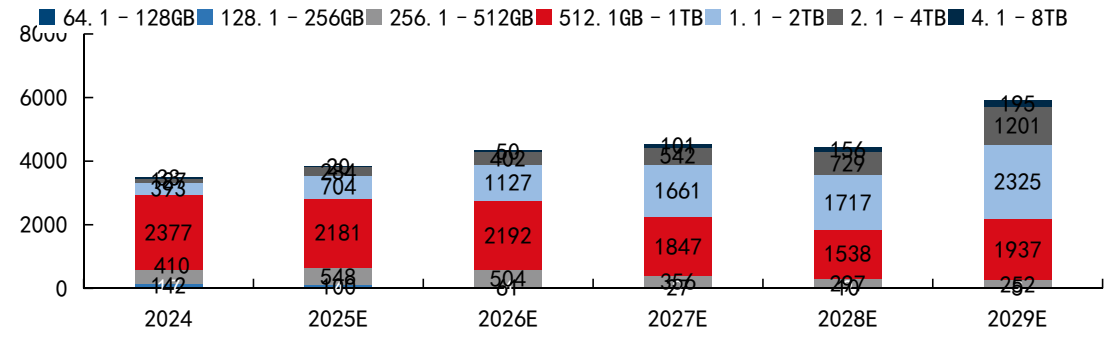
资料来源：IDC，国信证券经济研究所整理

图59：消费级-PC用SSD各容量占比情况



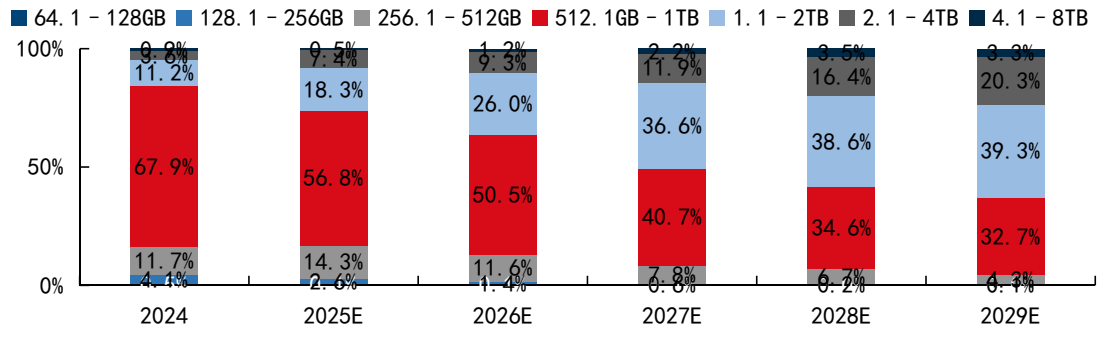
资料来源：IDC，国信证券经济研究所整理

图60：消费级-消费电子用SSD各容量出货情况（单位：万颗）



资料来源：IDC，国信证券经济研究所整理

图61：消费级-消费电子用SSD各容量占比情况

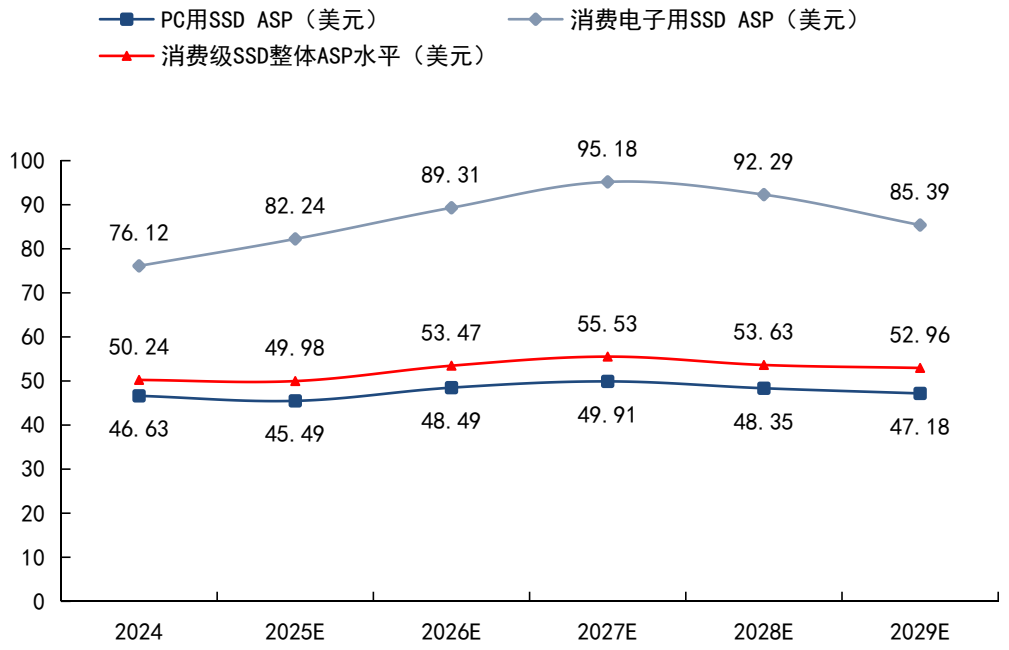


资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）-消费级：ASP稳中有升，单GB价格持续下降

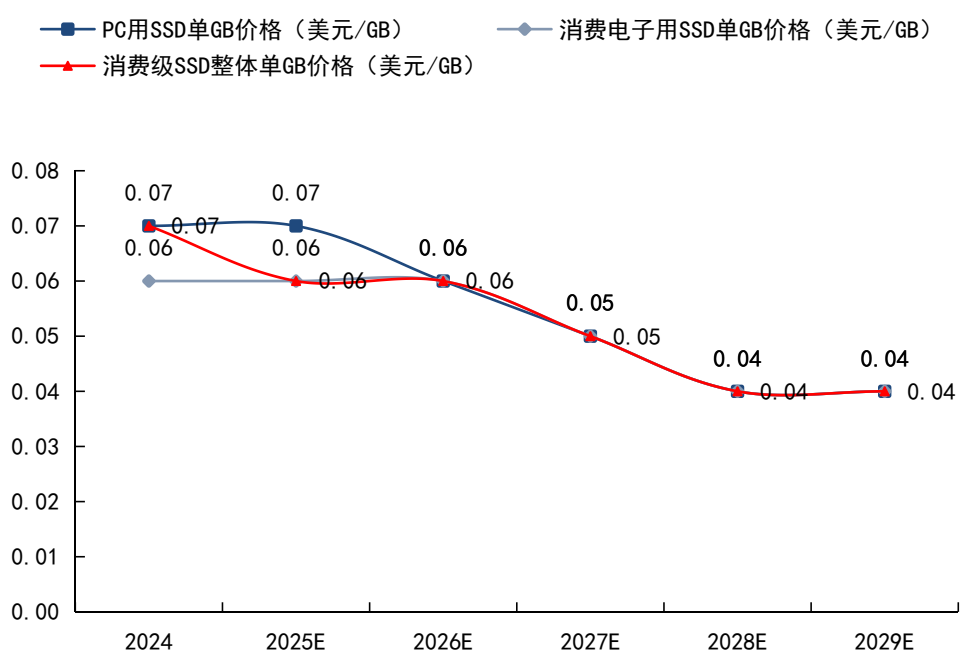
- **消费级SSD ASP价格稳中有升。**根据IDC披露数据，2024年消费级SSD整体ASP水平为50.24美元，预计2029年逐步提升至52.96美元，稳步上升；从细分领域来看，消费电子用SSD ASP高于PC用SSD价格。
- **消费级SSD单GB价格持续下降。**根据IDC预测数据，2024年消费级SSD整体单GB价格为0.07美元/GB，预计2029年将下降至0.04美元/GB，持续下降；从细分领域来看，目前PC用SSD单GB价格略高于消费电子用SSD单GB价格。

图62：消费级SSD整体及细分领域ASP情况（单位：美元）



资料来源：IDC，国信证券经济研究所整理

图63：消费级SSD整体及细分领域单GB价格情况（单位：美元/GB）



资料来源：IDC，国信证券经济研究所整理

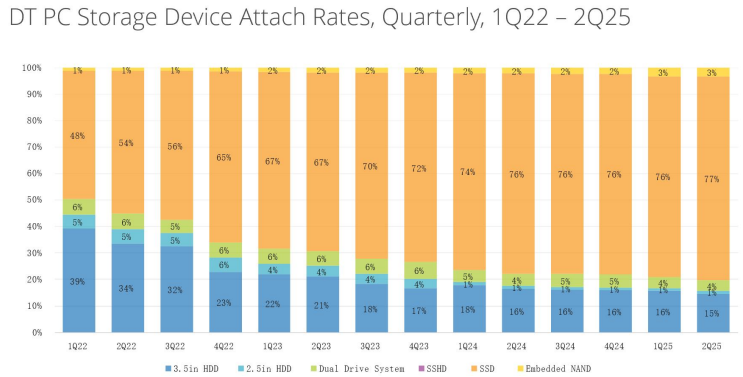
固态硬盘（SSD）-消费级：桌面PC领域替代HDD，便携式PC领域相对稳定

国信证券 GUOSEN SECURITIES

- **桌面PC领域，SSD有望逐步替代HDD。**根据IDC披露数据，桌面PC领域主要为SSD和HDD，其中SSD占比接近80%，为主流存储设备；同时，SSD以其高读写速率、低功耗、抗震防摔、无噪声等特性，预计在桌面PC领域将进一步替代HDD。

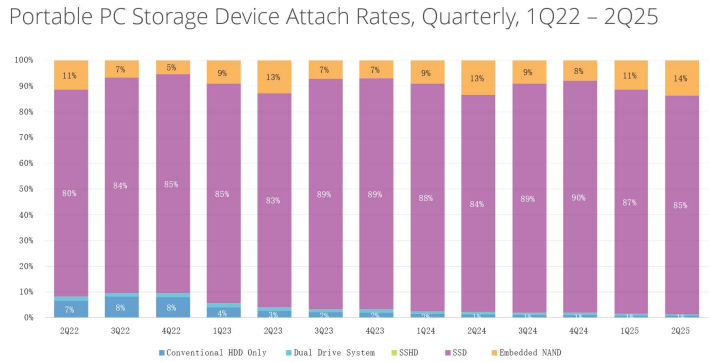
- **便携式PC领域，SSD份额预计相对稳定。**根据IDC披露数据，便携式PC领域主要为SSD和Embedded NAND，其中SSD占比近90%，但由于Embedded NAND具备超低功耗、体积小等特点，预计SSD和Embedded NAND将长期共存，双方市占率维持稳定。

图64：桌面PC领域存储器占比情况



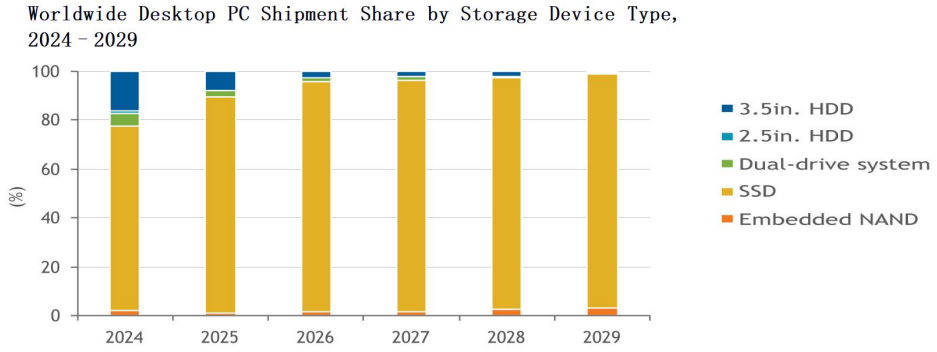
资料来源：IDC，国信证券经济研究所整理

图65：便携式PC存储器占比情况



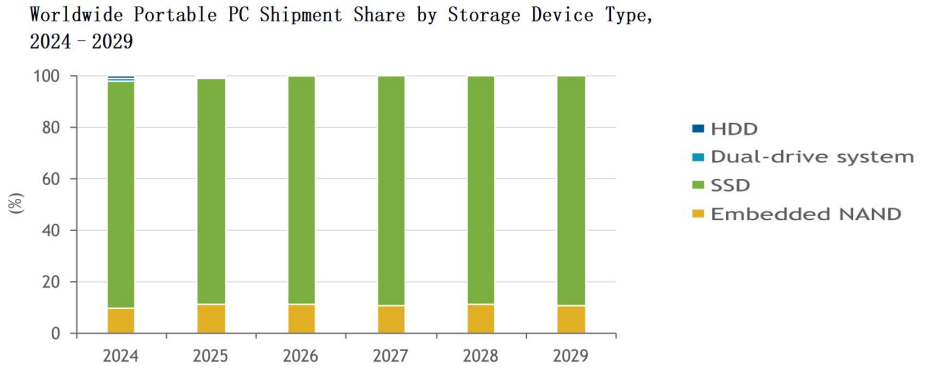
资料来源：IDC，国信证券经济研究所整理

图66：桌面PC领域，SSD有望逐步替代HDD



资料来源：IDC，国信证券经济研究所整理

图67：便携式PC领域，SSD市占率相对稳定



资料来源：IDC，国信证券经济研究所整理

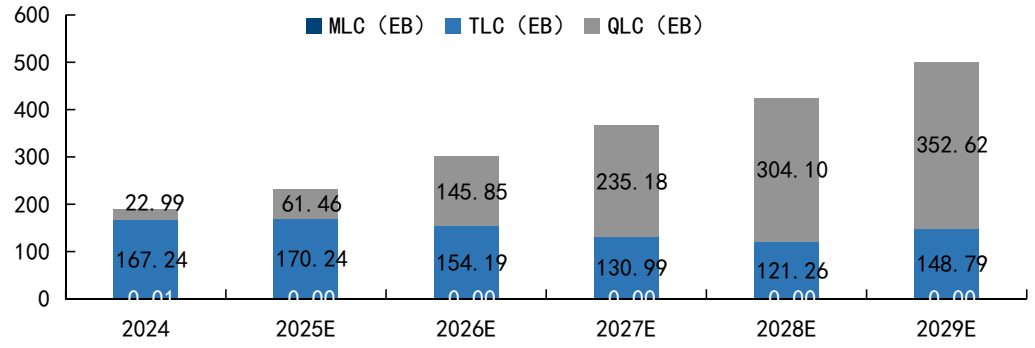
固态硬盘（SSD）-消费级：QLC占比持续提升，PCIe接口为主流

国信证券 GUOSEN SECURITIES

- **消费级SSD中QLC占比持续提升：**根据IDC披露数据，2024年MLC、TLC、QLC出货容量分别为0.01、167.24、22.99EB，占比分别为0.0%、87.9%、12.1%，TLC为主流；在存储密度持续提升的大背景下，存储4位数据的QLC（Quad Level Cell）出货量占比有望快速提升，根据IDC预测数据，预计2029年消费级SSD中QLC的占比达到70.3%，成为消费级SSD的主流。

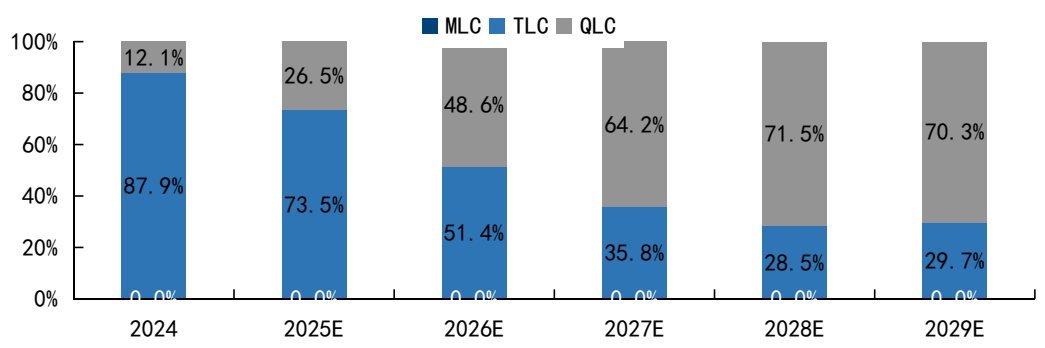
- **消费级SSD中PCIe接口为主流：**根据IDC披露数据，2024年PCIe、SATA接口版本消费级SSD出货量分别为166.2、24.0EB，占比分别为87.4%，PCIe版本为主流。

图68：消费级SSD中MLC、TLC、QLC出货容量情况（单位：EB）



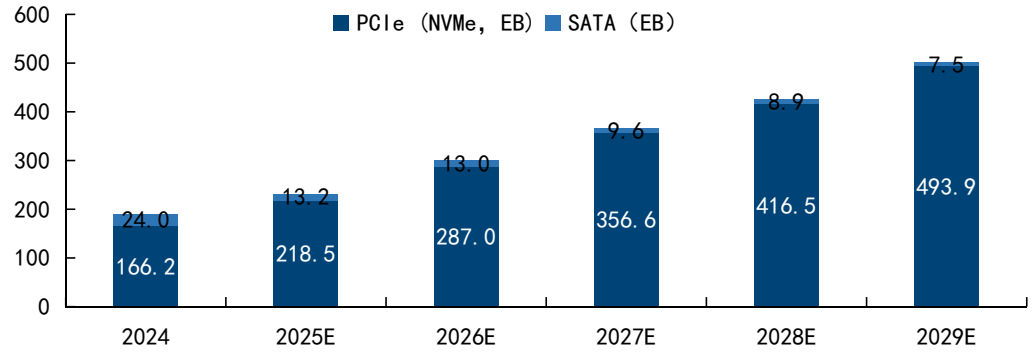
资料来源：IDC，国信证券经济研究所整理

图69：消费级SSD中MLC、TLC、QLC出货容量占比情况



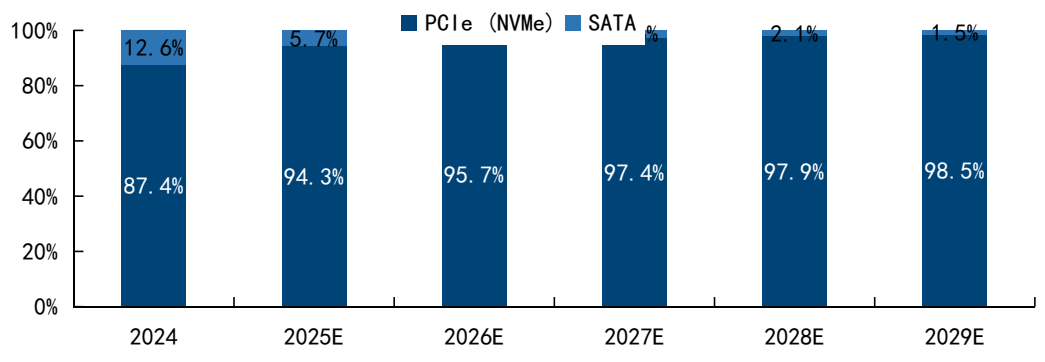
资料来源：IDC，国信证券经济研究所整理

图70：消费级SSD中PCIe、SATA出货容量情况（单位：EB）



资料来源：IDC，国信证券经济研究所整理

图71：消费级SSD中PCIe、SATA出货容量占比情况

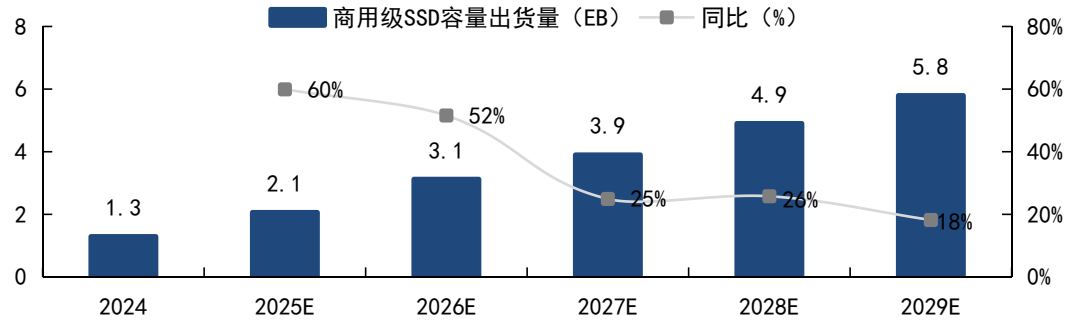


资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）-商用级：单颗容量持续提升，单GB价格持续下降

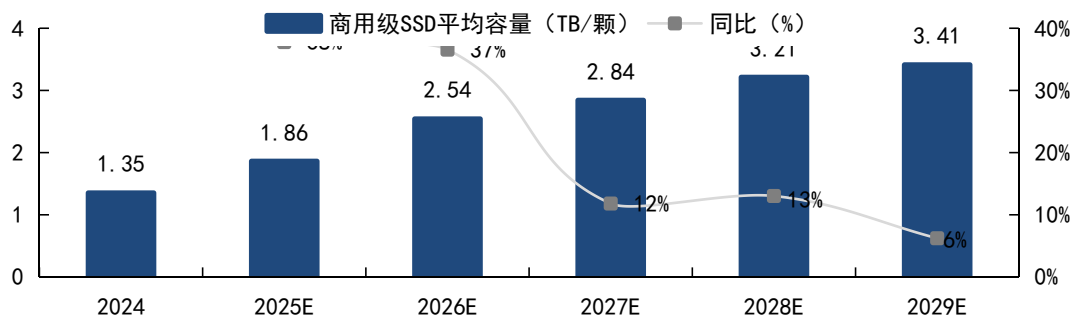
- **商用级固态硬盘（SSD）出货容量稳步增长，平均容量持续提升：**根据IDC披露数据，2024年商用级SSD容量出货量为1.3EB，预计2029年增长至5.8EB，对应24-29年CAGR为35.1%；2024年商用级SSD平均容量为1.35TB，预计2029年增长至3.41TB。
- **商用级固态硬盘（SSD）单GB价格持续下降：**根据IDC披露数据，2024年商用级SSD ASP为43.01美元/颗，预计2027年开始下降；2024年商用级SSD单GB价格为0.32美元/GB，持续下降，预计2029年下降至0.11美元/GB。

图72：商用级SSD容量出货量（单位：EB）



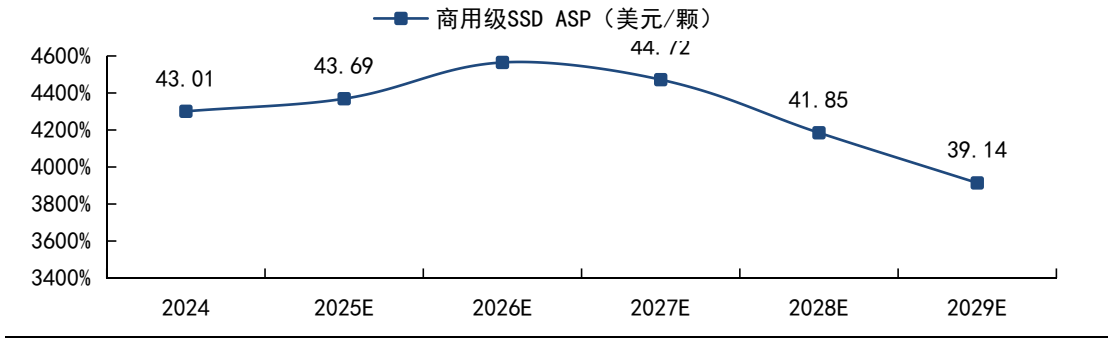
资料来源：IDC，国信证券经济研究所整理

图73：商用级SSD平均容量（TB/颗）



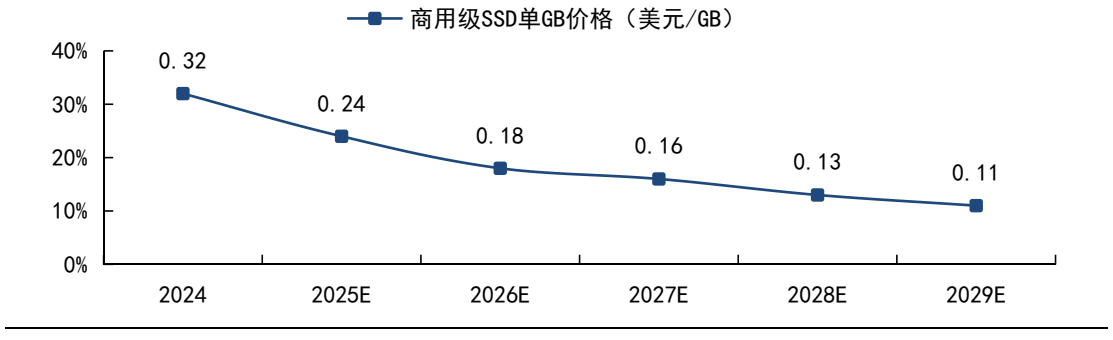
资料来源：IDC，国信证券经济研究所整理

图74：商用级SSD ASP情况（单位：美元/颗）



资料来源：IDC，国信证券经济研究所整理

图75：商用级SSD单GB价格情况（单位：美元/GB）

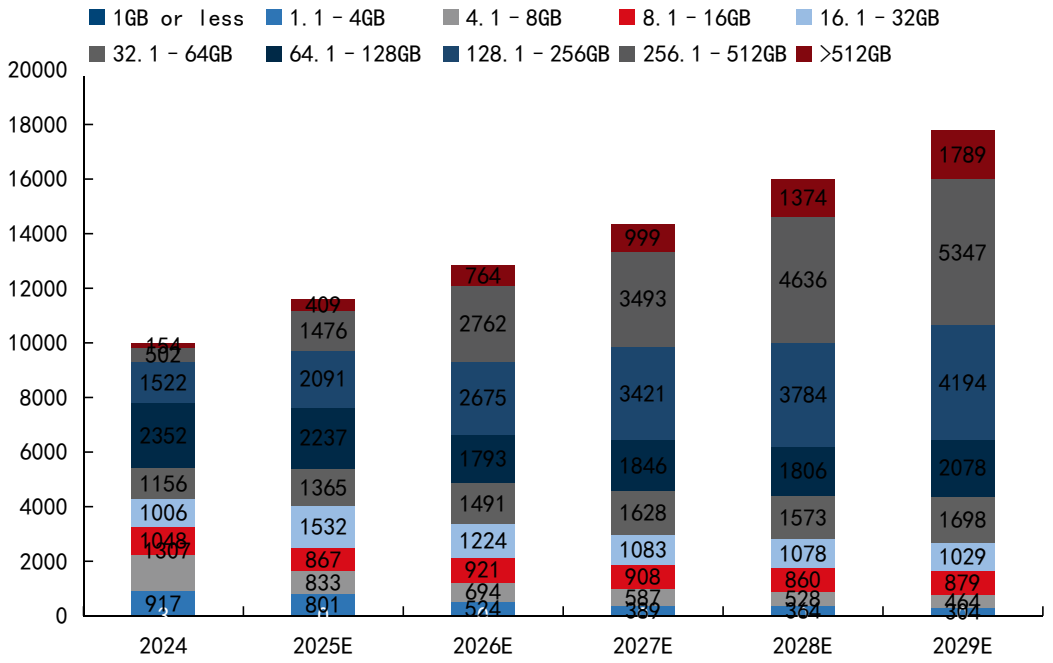


资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）-商用级：容量集中在4GB-512GB，高容量占比逐步提升

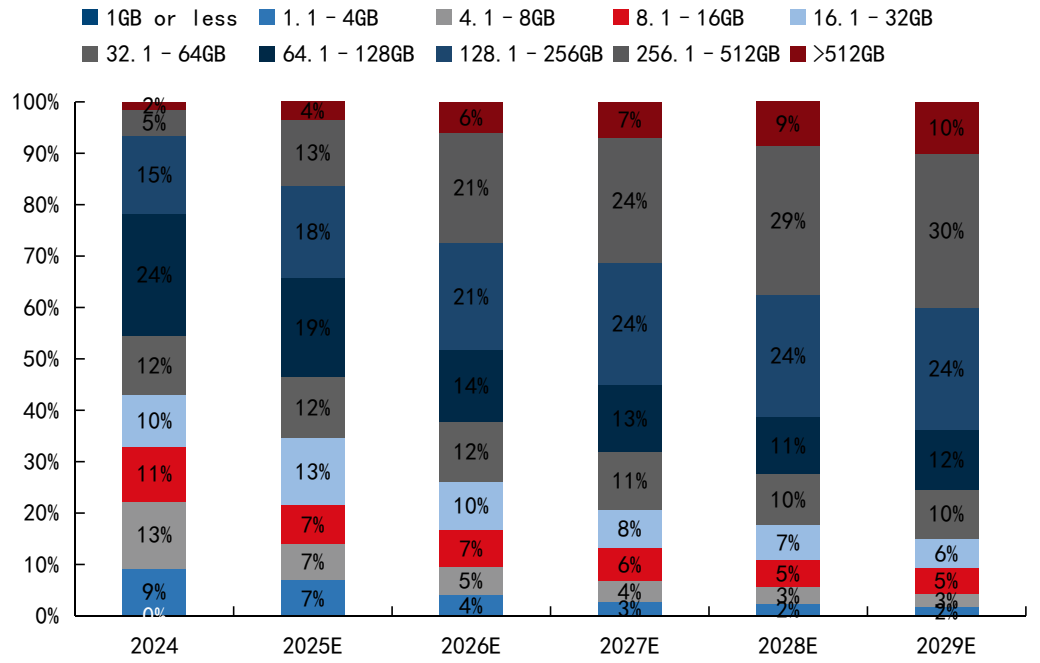
- **商用级固态硬盘（SSD）容量主要集中在4GB-512GB，高容量占比逐步提升：**根据IDC披露数据，2024年商用级SSD容量主要集中在4GB-512GB，整体容量区间低于消费级和企业级，其中2024年4GB-8GB、8GB-16GB、16GB-32GB、32GB-64GB、64GB-128GB、128GB-256GB、256GB-512GB出货量分别为130.7、104.8、100.6、115.6、235.2、152.2、50.2万颗，占比分别为13%、11%、10%、12%、24%、15%、5%；未来，高容量商用级SSD占比逐步提升，其中256GB-512GB、512GB以上版本SSD占比预计从2024年5%、2%提升至2029年的30%、10%。

图76：商用级SSD各容量出货情况（单位：千颗）



资料来源：IDC，国信证券经济研究所整理

图77：商用级SSD各容量出货占比情况



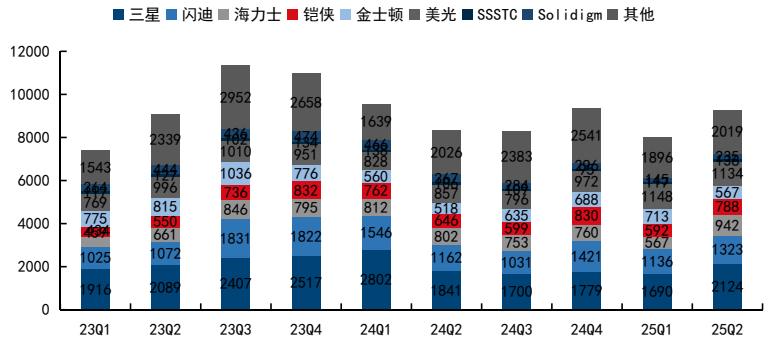
资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）供给侧（出货量）：三星为行业龙头，CR5为68.1%

国信证券 GUOSEN SECURITIES

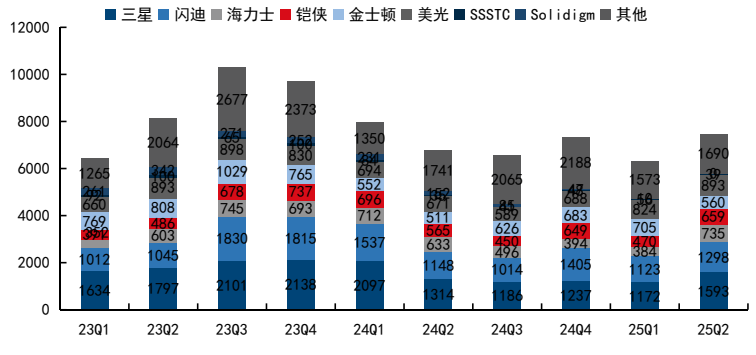
- **固态硬盘（SSD）各家出货量维度：**根据IDC披露数据，25Q2三星、闪迪、美光、海力士、铠侠市占率分别为22.9%、14.3%、12.2%、10.2%、8.5%，合计占比68.1%。
- **消费级SSD各家出货量维度：**根据IDC披露数据，25Q2消费级SSD市场，三星、闪迪、美光、海力士、铠侠市占率分别为21.3%、17.4%、12.0%、9.8%、8.8%，合计占比69.4%。
- **企业级SSD各家出货量维度：**根据IDC披露数据，25Q2企业级SSD市场，三星、美光、Solidigm、海力士、铠侠市占率分别为34.1%、15.4%、15.1%、13.3%、8.3%，合计占比86.2%。

图78：各SSD厂商出货量情况（单位：万颗）



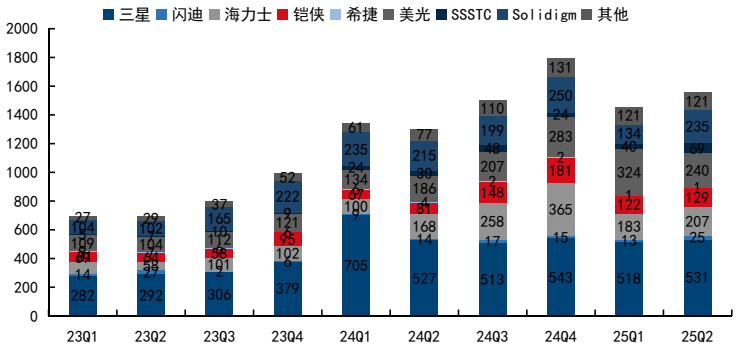
资料来源：IDC，国信证券经济研究所整理

图79：消费级SSD各厂商出货量情况（单位：万颗）



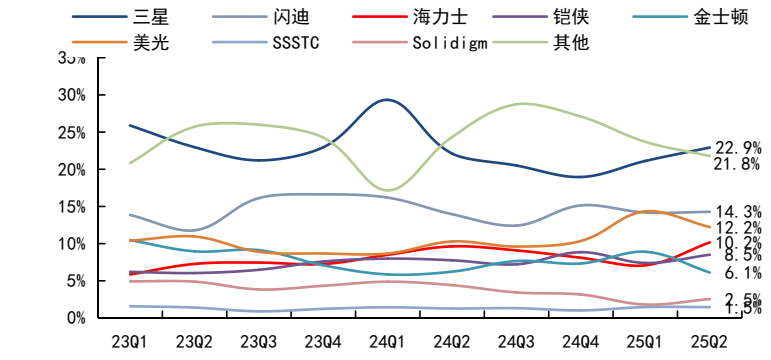
资料来源：IDC，国信证券经济研究所整理

图80：企业级SSD各厂商出货量情况（单位：万颗）



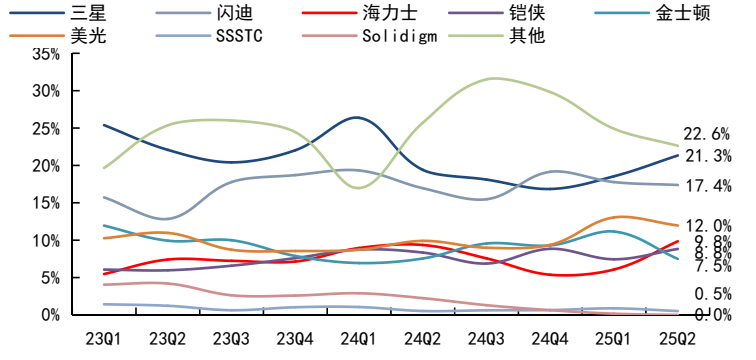
资料来源：IDC，国信证券经济研究所整理

图81：各SSD厂商市占率情况（出货量维度）



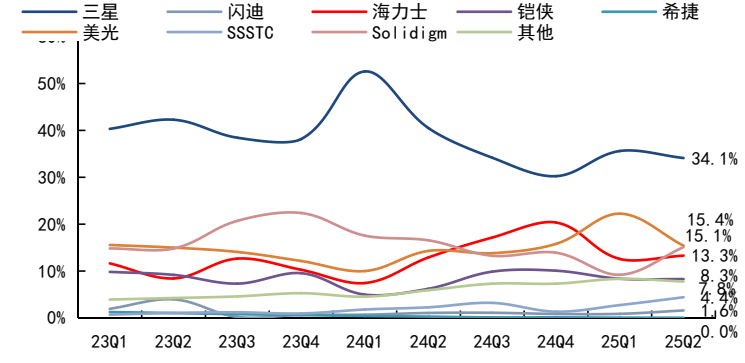
资料来源：IDC，国信证券经济研究所整理

图82：消费级SSD各厂商市场率情况（出货量维度）



资料来源：IDC，国信证券经济研究所整理

图83：企业级SSD各厂商市场率情况（出货量维度）



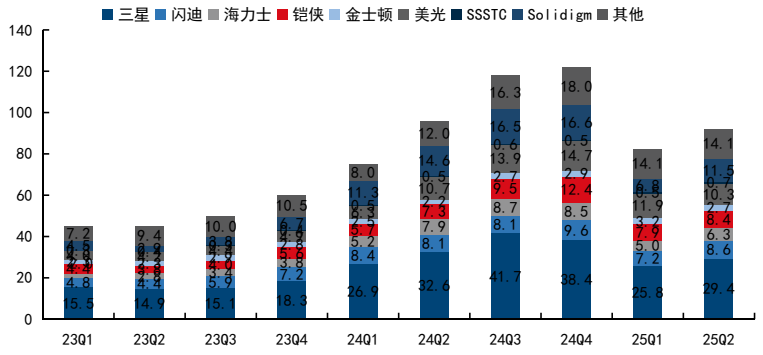
资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）供给侧（收入）：三星为行业龙头，CR5为74.1%

国信证券 GUOSEN SECURITIES

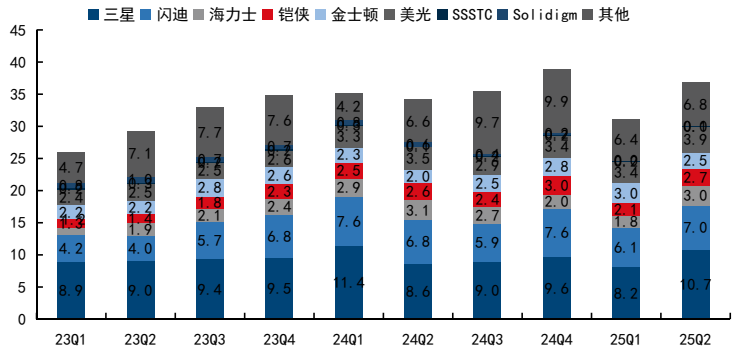
- **固态硬盘（SSD）各家收入维度：**根据IDC披露数据，25Q2三星、Solidigm、美光、闪迪、铠侠市占率分别为32.0%、12.5%、11.1%、9.4%、9.1%，合计占比74.1%。
- **消费级SSD各家收入维度：**根据IDC披露数据，25Q2消费级SSD市场，三星、闪迪、美光、海力士、铠侠市占率分别为29.1%、19.0%、10.6%、8.3%、7.3%，合计占比74.3%。
- **企业级SSD各家出货量维度：**根据IDC披露数据，25Q2企业级SSD市场，三星、Solidigm、美光、铠侠、海力士市占率分别为34.6%、21.3%、11.8%、10.6%、6.0%，合计占比84.3%。

图84：各SSD厂商营收情况（单位：亿美元）



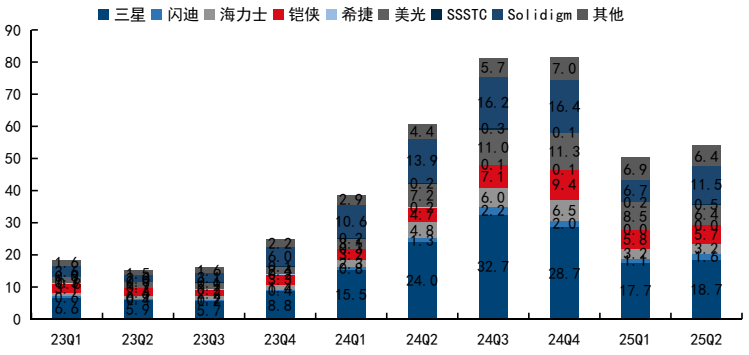
资料来源：IDC，国信证券经济研究所整理

图85：消费级SSD各厂商营收情况（单位：亿美元）



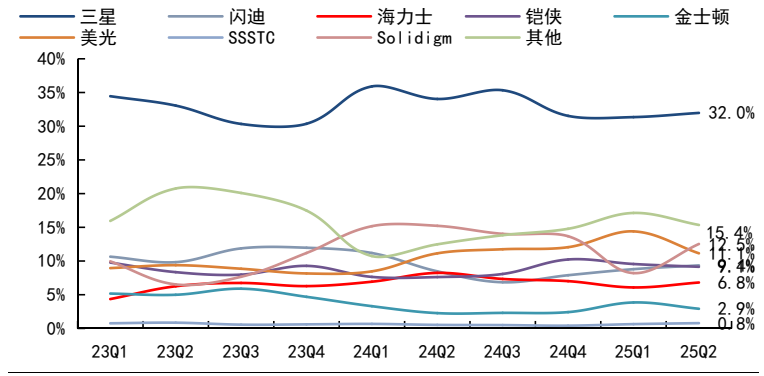
资料来源：IDC，国信证券经济研究所整理

图86：企业级SSD各厂商营收情况（单位：亿美元）



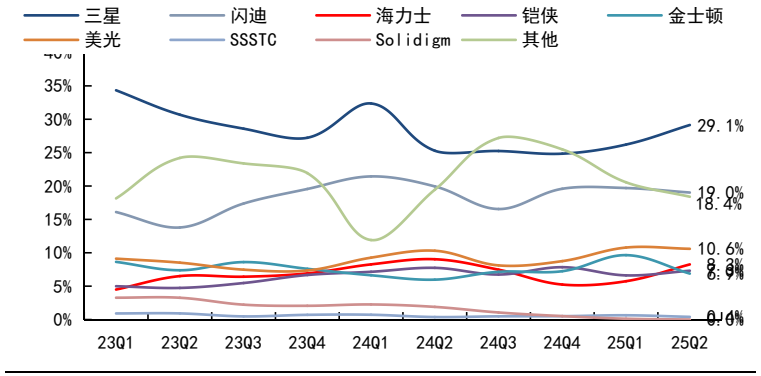
资料来源：IDC，国信证券经济研究所整理

图87：各SSD厂商市占率情况（收入维度）



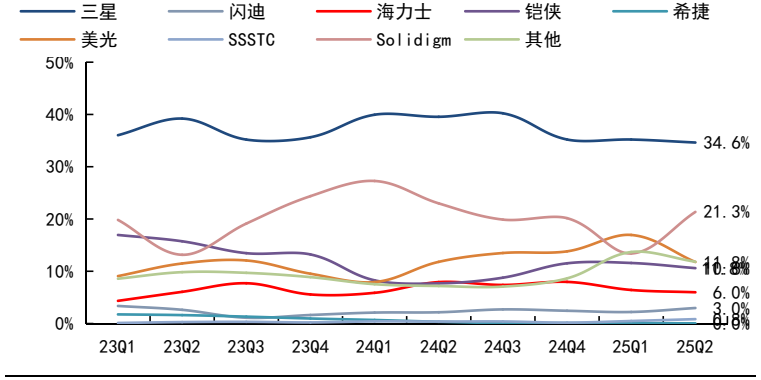
资料来源：IDC，国信证券经济研究所整理

图88：消费级SSD各厂商市场率情况（收入维度）



资料来源：IDC，国信证券经济研究所整理

图89：企业级SSD各厂商市场率情况（收入维度）



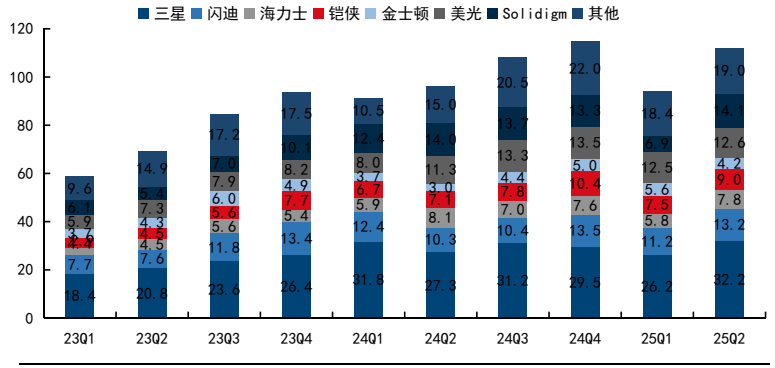
资料来源：IDC，国信证券经济研究所整理

固态硬盘（SSD）供给侧（容量）：三星为行业龙头，CR5为68.1%

国信证券 GUOSEN SECURITIES

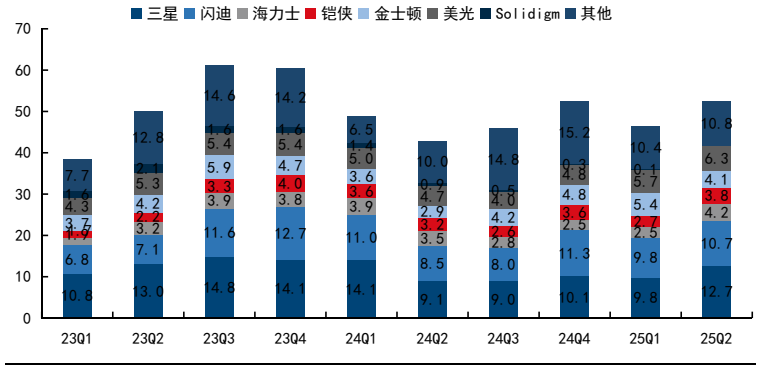
- **固态硬盘（SSD）各家容量出货量维度：**根据IDC披露数据，25Q2三星、Solidigm、闪迪、美光、铠侠市占率分别为28.7%、12.6%、11.8%、11.2%、8.1%，合计占比72.3%。
- **消费级SSD各家容量出货量维度：**根据IDC数据，25Q2消费级SSD市场，三星、闪迪、美光、海力士、金士顿市占率分别为24.2%、20.4%、11.9%、8.0%、7.7%，合计为72.3%。
- **企业级SSD各家容量出货量维度：**根据IDC数据，25Q2企业级SSD市场，三星、Solidigm、美光、铠侠、海力士市占率分别为32.8%、23.8%、10.7%、8.8%、6.1%，合计为86.2%。

图90：各SSD厂商容量出货量情况（单位：EB）



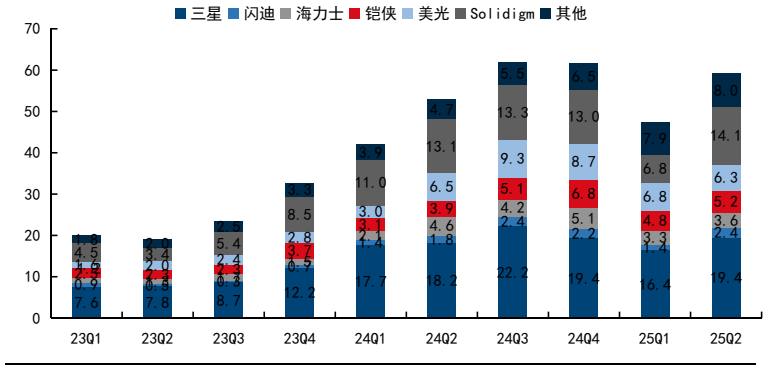
资料来源：IDC，国信证券经济研究所整理

图91：消费级SSD各厂商容量出货量情况（单位：EB）



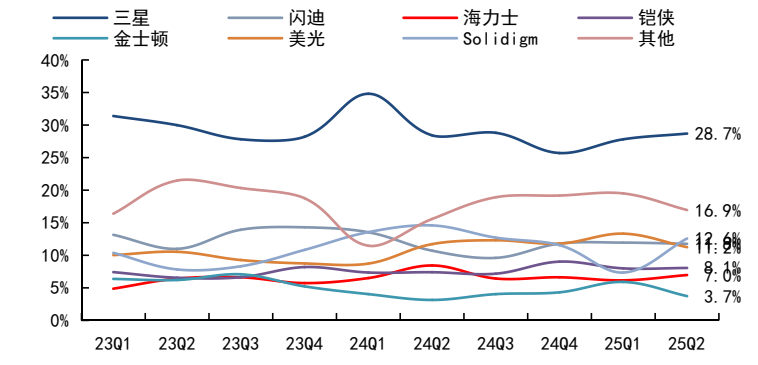
资料来源：IDC，国信证券经济研究所整理

图92：企业级SSD各厂商容量出货量情况（单位：EB）



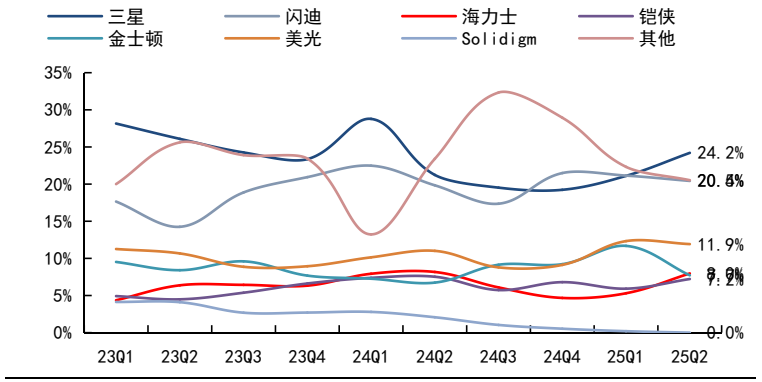
资料来源：IDC，国信证券经济研究所整理

图93：各SSD厂商市占率情况（容量维度）



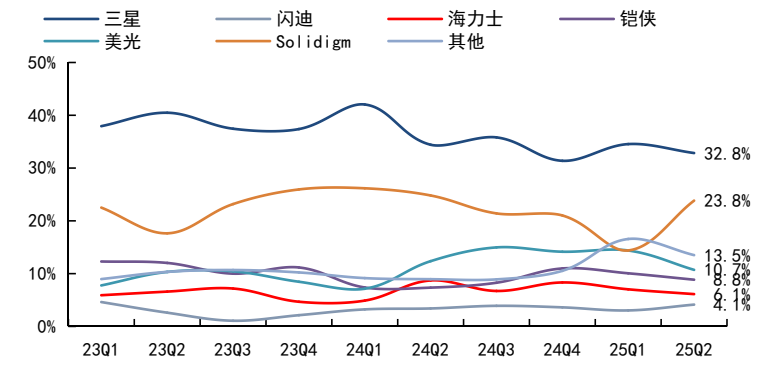
资料来源：IDC，国信证券经济研究所整理

图94：消费级SSD各厂商市场率情况（容量维度）



资料来源：IDC，国信证券经济研究所整理

图95：企业级SSD各厂商市场率情况（容量维度）



资料来源：IDC，国信证券经济研究所整理

存储器：主要包括DRAM和Flash存储

国信证券 GUOSEN SECURITIES

■ **存储器：**根据存储器传统定义，可分为只读存储器（ROM）和随机存储器（RAM）。

➢ **只读存储器（ROM）：**ROM是指将内容写在芯片上，只能读取，不能随意更改内容的一种存储器，一般用于存放固定的程序，例如BIOS、ROM中内容不会因掉电而丢。

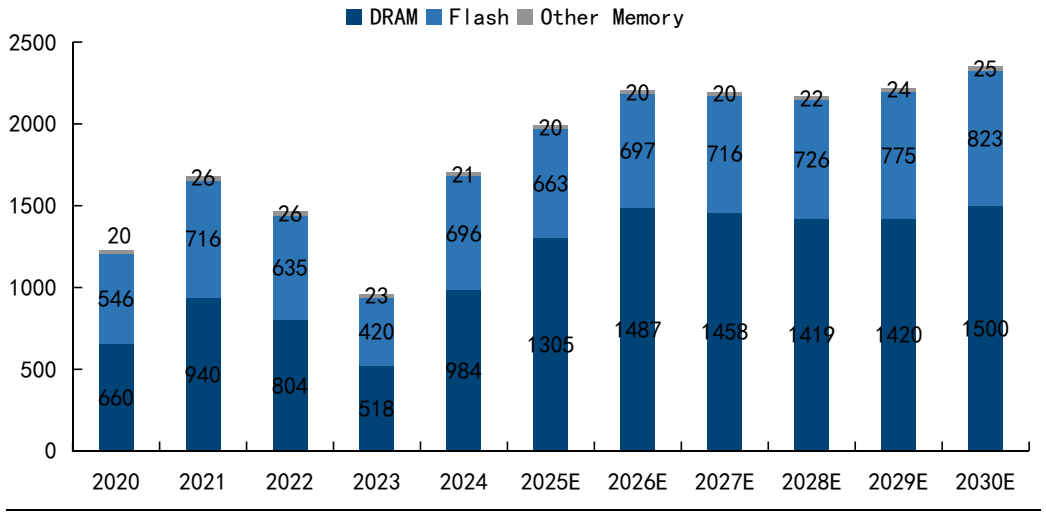
- ERPOM：可编程ROM存储器，芯片上有一个透明窗口，用特殊的装置向芯片写完毕后，用不透明的标签贴住，如果要擦除ERPOM中的内容，揭掉标签，用紫外线照射EPROM的窗口，EPROM中的内容就会丢失；
- OTPROM：一次编程ROM，写入原理类似ERPOM，但为了节省成本，编程写入后不再抹除，因此不设置透明窗；
- EEPROM：电擦除可编程ROM，内部信息可以抹去，也可以写入新的数据，用电对其进行改写，而不需要紫外线；
- **Flash Memory：主要特点为在不加电的情况下能长期保存存储信息，其既有ROM特点，又有很高的存储速度，且易于擦除和重写，功耗很小，属于EEPROM的升级版，Flash按扇区（Block）进行工作，而EEPROM按照字节操作，且Flash结构相对简单，成本较低。**

➢ **随机存储器（RAM）：**RAM仅能暂时存放程序和数据，一旦关闭电源或发生断电，其中的数据就会丢失，按照制造原理不同，通常可分为静态RAM（SRAM）和动态RAM（DRAM）。

- 静态RAM（SRAM）：基本结构为一个双稳态电路，由于读、写的转换被写电路控制，所以只要写电路不工作，电路有电，开关就保持现状，不需要刷新，因此SRAM又叫静态RAM；这里的开关实际上由晶体管替代，晶体管转换时间一般小于20ns，所以SRAM读写速度非常快，一般比DRAM快2-3倍；但这种开关电路元件较多（一个存储单元需要4个晶体管和2个电阻），一方面降低了SRAM的集成度，另一方面增加了生产成本；
- **动态RAM（DRAM）：内部存储的数据需要不断刷新，因为一个DRAM单元由一个晶体管和一个小电容组成，晶体管通过小电容的电压来保持断开、接通状态，当小电容有电时，晶体管接通表示1，当小电容没电时，晶体管断开表示0；但是充电后的小电容上的电荷会很快丢失，所以需要不断地刷新。同时，由于电容充、放电需要时间，所以DRAM的读写时间远远慢于SRAM，其平均读写时间为60-120ns；但由于其结构简单，实际生产时集成度很高，成本相对较低。**

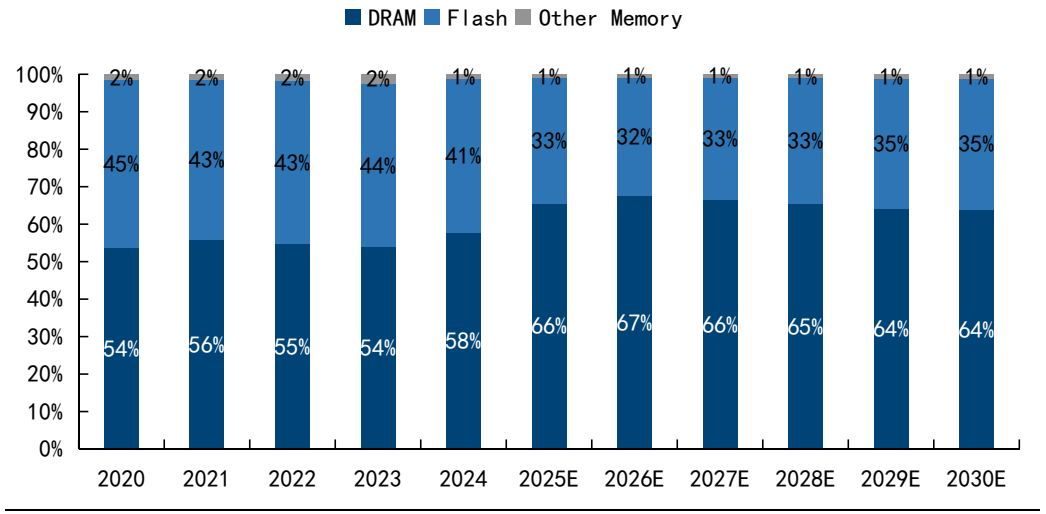
➢ **目前，存储器市场以DRAM和Flash为主：**根据IDC披露数据，2024年全球存储器市场为1701亿美金，其中DRAM、Flash市场规模分别为984、696亿美金，占比分别为58%、41%。

图96：全球存储器市场规模（单位：亿美元）



资料来源：IDC，国信证券经济研究所整理

图97：全球存储器以DRAM和Flash为主



资料来源：IDC，国信证券经济研究所整理

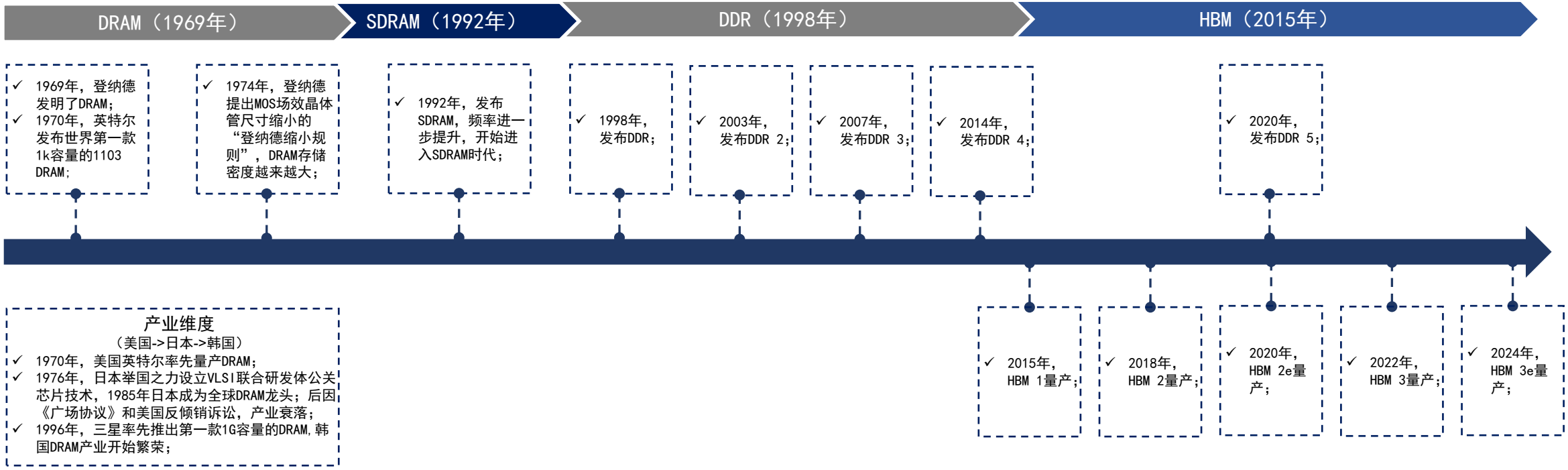
DRAM发展历史：DRAM -> SDRAM -> DDR -> HBM

国信证券 GUOSEN SECURITIES

■ **DRAM发展历史：**DRAM产品形态主要经过DRAM、SDRAM、DDR、HBM。

- **DRAM（1969年）：**1969年登纳德发明了DRAM，1970年英特尔发布世界第一款1k容量的1103 DRAM；产业最初起源于美国，1985年日本逐步成为全球DRAM龙头，后因《广场协议》和美国反倾销诉求，产业衰落，之后韩国接替美国成为全球DRAM龙头；
- **SDRAM（1992年）：**1992年，发布SDRAM，频率进一步提升，开始进入SDRAM时代；
- **DDR（1998年）：**1998年发布第一代DDR，随后持续迭代，2020年发布第五代DDR；
- **HBM（2015年）：**2015年第一代HBM开始量产，随后持续迭代，2024年HBM3E开始量产。

图98：DRAM发展历史



资料来源：汪波著-《芯片简史》-浙江教育出版社（2023年）-P205、袁春风等著-《计算机系统：基于x86+Linux平台》-机械工业出版社（2024年）-P428、三星、海力士，国信证券经济研究所整理

DRAM：功耗小，集成度高，成本低，但读写速度慢于SRAM

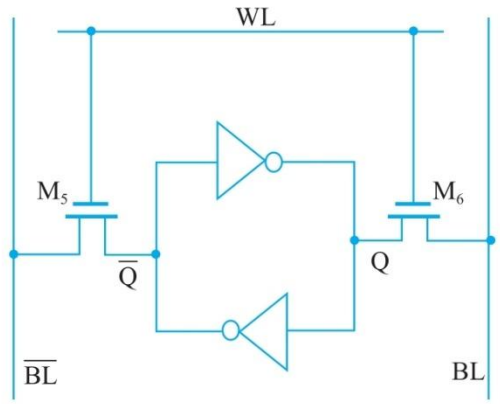
- **SRAM**：使用6个MOS管组成一个存储元件，其中一个反向器由2个MOS管构成，两个反相器反向连接构成1位锁位器，用于存储信息Q，若Q电为高电平，则存储状态为1，否则为0；读写时需向门控管M5和M6加高电平使其导通。
  - **信息的保持**：字选择线WL加低电平时，M5和M6截止，锁存器与外界隔离，保持原有信息不变；
  - **读出**：在两侧位线加高电平，当字选择线WL上加高电平时，M5和M6导通，由于锁位器两侧电平相反，可通过在位线上检测电平变化来区分读出的是0还是1；
  - **写入**：当字选择线WL上加高电平，M5与M6导通；若要写0，则在右侧位线BL上加低电平，使Q点电位下降，将0写入锁存器；同理，若写1，则左侧位线加低电平。
- **DRAM**：利用MOS管和电容Cs保存信息，T管为为字选门控管，在信息保持状态下，T管截止，存储元件中没有电流流动，因而可以减少功耗。
  - **读出**：若原存为1，则Cs上电荷通过T管在数据线上产生电流；若原存为0，则无电流；由于读出时Cs电荷放电，电位下降，因此是破坏性读出，读后应有重写操作，称为原生；
  - **写入**：写1时，在数据线上加高电平，经T管对Cs充电；写0时，加低电平，使Cs充分放电；
  - **刷新**：由于电容Cs上的电荷会缓慢放电，超过一定时间，就会丢失信息，因此必须定时给电容Cs充电。

图99：SRAM和DRAM对比

存储器	单位晶体管数目	相对访问时间	是否需要刷新	是否对抗干扰敏感	相对花费	应用场景
SRAM	6-8	1x	是	否	100x	高速缓存
DRAM	1	10x	否	是	1x	主存，帧缓存

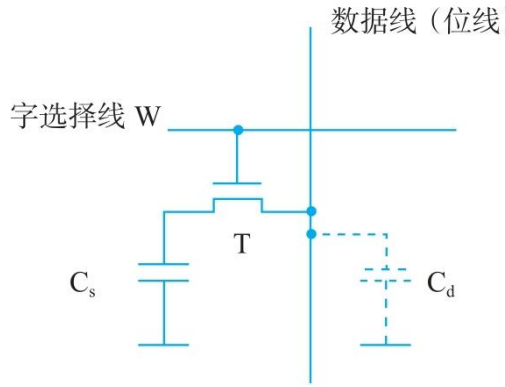
资料来源：赵欢著-《计算机系统：从应用程序到底层实现》-机械工业出版社（2024年）-P260，国信证券经济研究所整理

图100：六管静态存储元件



资料来源：袁春风等著-《计算机系统：基于x86+Linux平台》-机械工业出版社（2024年）-P420，国信证券经济研究所整理

图101：单管动态存储元件



资料来源：袁春风等著-《计算机系统：基于x86+Linux平台》-机械工业出版社（2024年）-P421，国信证券经济研究所整理

SDRAM技术：与DRAM控制器之间采用同步方式交换数据

国信证券 GUOSEN SECURITIES

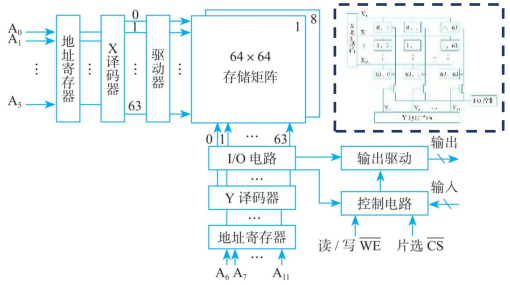
- **传统的DRAM技术：**传统DRAM芯片与DRAM控制器之间采用异步通信方式交换数据，DRAM控制器发出地址和控制信号后，经过一段延迟时间才读出或写入数据，但异步通信表示这些信号可能会在任意时刻达到DRAM芯片的引脚，为了保证时序的正确性，DRAM芯片的最高频率不宜过高，随着CPU主频的提升，异步DRAM不能满足需求。

- **SDRAM技术：**SDRAM芯片与DRAM控制器之间采用同步方式交换数据，其读写受存储器总线时钟控制，因此信号到达DRAM芯片引脚的时刻是可预测的，从而可以实现更高频率。
  - **突发传输（Burst）方式：**SDRAM芯片的每一步操作都是在外部存储器总线时钟控制下进行，SDRAM控制器只要在第一次存取时给出首地址，SDRAM芯片内部的一个列地址计数器会在每次访问数据后自动递增，因为无须发送后续地址即可连续快速访问存储体中的一连串数据。内部的模式寄存器可用于设置传送数据长度（即突发长度，BL）和从收到读命令到开始传送数据的延迟时间（即CAS潜伏期，CL），根据设定的BL和CL，SDRAM可以确定何时开始从存储器总线上取数以及连续取多少个数据。

图102：DRAM芯片结构及工作原理

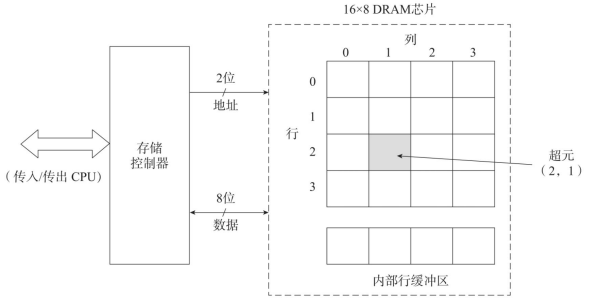
存储器芯片结构

- **存储矩阵：**如下图所示，4096个存储单元被排成64*64的存储阵列，成为位平面；
- **地址编码器：**用来将地址转换为输出线上的高电平，以便驱动相应的读写电路，目前DRAM主要用二维译码，分为行、列方向两个地址译码器（如下图X译码器、Y译码器）；图中4096个单元，需要12根地址线A0~A11，其中A0~A5送X地址译码器，有64条译码输出线X0~X63，各连接存储矩阵中相应一行所有记忆单元的字选择线；A6~A11送Y地址译码器，亦有64条译码输出线Y0~Y63，分别控制一列单元的位线控制门。（例如A0A1….A11为000001000000，则X1为高电平，与其相连的64个存储单元字选择线为高电平，同时Y0为高电平，则存储矩阵中（1，0）被选中；
- **驱动器：**双译码结构中，一条X方向的选择线要控制在其上的各个存储单元的字选择线，负载较大，需要在译码器输出后加驱动器；
- **I/O控制电路：**用于控制被选中单元的读出或写入，具有放大信息的作用；
- **片选控制电路：**单个芯片容量太小，需要将一定数量的芯片按特定方式连接成一个完整的存储器，在访问某字时，必须选中该字所在芯片，而其他芯片不被选中，需要片选控制信号，由DRAM控制器产生。



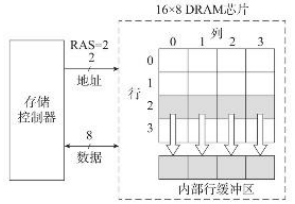
16*8的DRAM芯片内部结构示意图

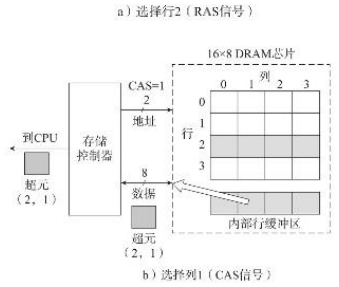
- **引脚：**信息通过引脚（PIN）的外部连接器传入和传出芯片，每个引脚通过高低电平可以传输一位二进制的信号，引脚又分为地址引脚和数据引脚，地址引脚用来定位芯片内单元地址，数据引脚传输数据；所以如果数据引脚数量为n，则一次可以从DRAM芯片中传送n位二进制数据；
- **超元（Supercell）：**为数据传输的基本单位，DRAM被分成s个超元，一个s*n的DRAM芯片可以存储s*n位信息，超元按照r行*c列组成一个二维矩阵（s=r*n），每个超元地址用（i，j）表示；
- **16*8的DRAM芯片举例：**“16”表示超元数量，即s=16，则矩阵可为4*4，“8”表示每个超元的位数，即n=8；有两组引脚（2位地址引脚和8位数据引脚），内部行缓冲区（用来缓存指定行中每一列的数据，一般用SRAM实现）。



DRAM芯片读写过程

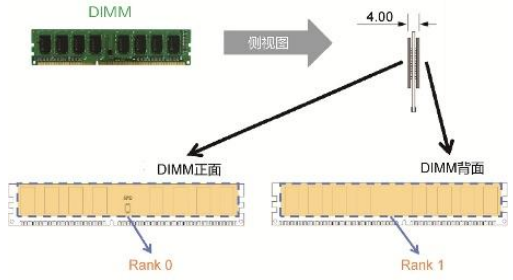
- **读数据：**例如需要读出超元（2，1）数据，存储控制器通过地址线（引脚）首先发送行地址2，DRAM收到RAS（行地址选通）信号2后，将行2的整个内容都复制到内部行缓冲区；接下来，存储控制器发送列地址1，DRAM收到CAS（列地址选通）信号1后，从行缓冲区复制出超元（2，1）中8位数据，通过数据线（引脚）传送到存储控制器；





DIMM结构

- 内存条通过基板上的引脚（金手指）与主板连接，在内存条的正反两面都有金手指；
- 常见的存储器模块封装有SIMM（单列直插式存储模块）和DIMM（双列直插式存储模块）；
- SIMM：提供32位数据通道，有72位引脚，两侧金手指互通，都提供相同的信号；
- DIMM：提供64位数据通道，有168个引脚，两侧金手指各自独立传输信号。



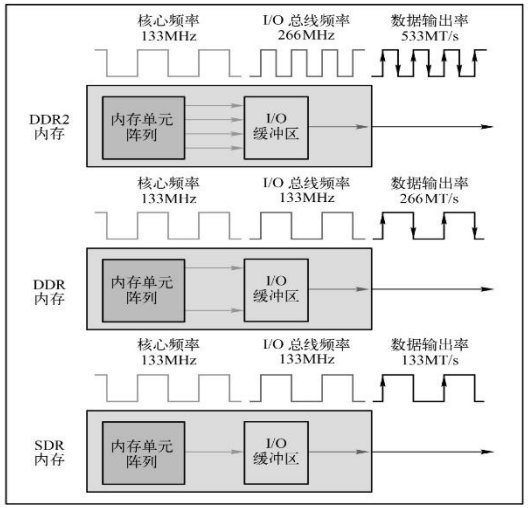
资料来源：袁春风等著-《计算机系统：基于x86+Linux平台》-机械工业出版社（2024年）-P422、赵欢著-《计算机系统：从应用程序到底层实现》-机械工业出版社（2024年）-P262、舒继武著-《数据存储架构与技术》-人民邮电出版社（2024年）-P53，国信证券经济研究所整理

DDR：DDR持续迭代，大厂开始布局DDR6

国信证券 GUOSEN SECURITIES

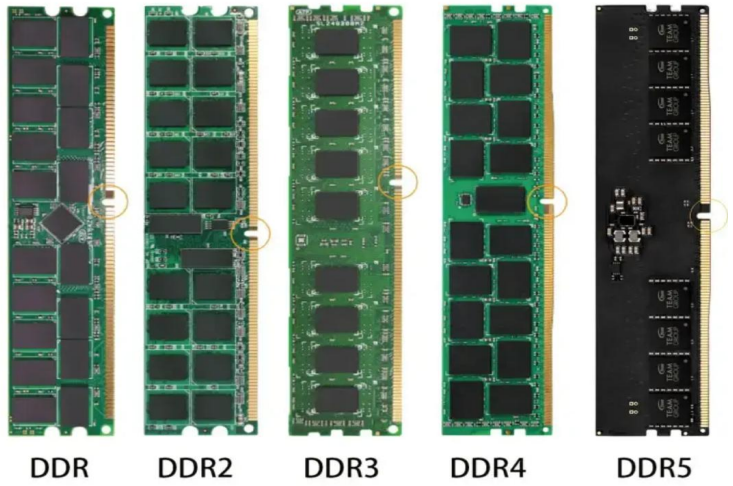
- **DDR**：1998年，发布第一代DDR（Double Data Rate） SDRAM芯片，改进了标准的SDRAM设计，通过芯片内部的预取缓冲区提供双字预取功能，并利用存储总线上的时钟信号的上升沿与下降沿，实现一个时钟内传送两个存储字功能（即时钟信号的上升沿和下降沿都可以读/写数据）；通常用“等效频率”衡量内存条的数据传输率（即1秒钟内完成的数据传输次数，MT/s），DDR的等效频率为核心频率的2倍，所以常见的DDR-200、DDR-266等的核心频率分别为100MHz、133MHz，传输带宽=核心频率*2*64/8MB/s=100MHz*2*64/8MB/s=1.6GB/s（以DDR-200为例）。
- **DDR2**：2003年，发布第二代DDR SDRAM芯片，采用与DDR类似的技术，利用芯片内部的预取缓冲区可以进行4字预取，同时通过改进接口电气特性、简化存储器总线协议等技术，其内部I/O总线频率为核心频率的2倍，结合起来，DDR2的等效频率为核心频率的4倍，完成“4位预取”（4-bit Prefetch）技术，传输带宽=核心频率*2（I/O总线频率倍增）*2（双倍速率）*64/8MB/s。
- **DDR3**：2007年，发布第三代DDR SDRAM芯片，首先，其能耗大约下降了30%（工作电压从1.8V下降到1.5V）；其次，通过频率倍增，DDR3内存的I/O总线频率为核心频率的4倍，达到了“8位预取”，其等效频率为核心频率的8倍，传输带宽=核心频率*4（I/O总线频率倍增）*2（双倍速率）*64/8MB/s。
- **DDR4**：2014年，发布第四代DDR SDRAM芯片，首先，其能耗进一步下降25%-40%（工作电压下降到1.2V）；其次，由于“16位预取”技术较难实现（与Catch Line有关），DDR4仍沿用DDR3的“8位预取”所以主要提升点在于核心频率，同时引入Bank Group概念，每个Bank Group具备独立启动操作读、写等动作特性，DDR4根据列宽可以有2个或4个Bank Group，当分为2个Bank Group时且每个Group都进行“8位预取”时，相当于“16位预取”，传输带宽=2*核心频率*4（I/O总线频率倍增）*2（双倍速率）*64/8MB/s。
- **DDR5**：2020年，发布第五代DDR SDRAM芯片，首先，其能耗进一步下降（工作电压下降到1.1V）；其次，DDR5每个模块有两块独立的32位通道，使得并行操作数量翻倍，Banks数量从16增长至32，Bank Group从2和4翻倍到4和8，使数据总线更有效率；突发长度从8翻到16，从“8位预取”翻到“16位预取”，数据传输率和总线效率更高，传输带宽=2*核心频率*8（I/O总线频率倍增）*2（双倍速率）*64/8MB/s。
- **DDR6**：头部厂商已经开始对DDR6研发，其中三星计划在CES 2026上展出全球首款LPDDR6内存，同时海力士在2025年SK AI峰会上展示产品路线图，预计2026-2028年会推出LPDDR6内存。

图103：SDR、DDR、DDR2性能对比



资料来源：黄勤主编-《微型计算机原理及接口技术（第二版）》-机械工业出版社（2023年）-P290，国信证券经济研究所整理

图104：DDR外观形态



资料来源：爱集微，国信证券经济研究所整理

DDR：与LPDDR、GDDR参数区分

国信证券 GUOSEN SECURITIES

- **LPDDR：**LPDDR（Low Power Double Data Rate SDRAM）为移动设备专用的低功耗内存，以低电压、高密度封装为核心特征，主要用于智能手机、平板电脑、汽车电子等领域，其与DDR相比，容量、电压、功耗低于DDR，但最大数据传输速率、价格高于DDR。

- **GDDR：**GDDR（Graphics Double Data Rate SDRAM）指图形的双倍数据传输率，其针对显卡设计，具备两个主要特点：1）高密度寻址能力，颗粒的容量大，可以满足显卡对内容容量的要求和显卡有限的PCB板面积设计要求；2）性能，显存带宽必须满足高速传输大量纹理和贴图的能力。与传统DDR相比，GDDR通常具有较高的延迟和更大的带宽，因为GDDR主要用于GPU和显卡，架构在追求更高带宽时，对延迟的要求相对较低。

图105：DDR4与GDDR5、GDDR6参数对比

Product 产品	Clock Period (tCK) 时钟周期延迟（ns）		Data Rate (Gb/s) 数据速率（Gb/s）		Density 颗粒 存储密度（Gb）	Prefetch (Burst Length) 预取长度 （每次预取操作可访问的位数）	Number of Banks 板块数量
	Max	Min	Min	Max			
DDR4	1.25	0.625	1.6	3.2	4~16	8 n	8、16
GDDR5	20	1.00	2	8	4~8	8 n	16
GDDR6	20	0.571	2	14	8~6	16 n	16

资料来源：濮元恺著-《算力芯片：高性能CPU/GPU/NPU微架构分析》-电子工业出版社（2024年）-P576，国信证券经济研究所整理

图106：DDR4与LPDDR4、GDDR5参数对比

Memory	DDR4	LPDDR4	GDDR5
最大数据传输速率	3.2Gb/s	4.267Gb/s	8Gb/s
接口位宽	64+8bits	双16-bit通道	多个32-bit通道
最大容量	128GB	2GB	1GB
电压	1.2V	1.1V	1.6V
安装方式	Surface or DIMM	Surface or Module	Surface
价格	较低	中	高

资料来源：杨剑等著-《边缘计算系统设计与实践》-北京大学出版社（2023年）-P30，国信证券经济研究所整理

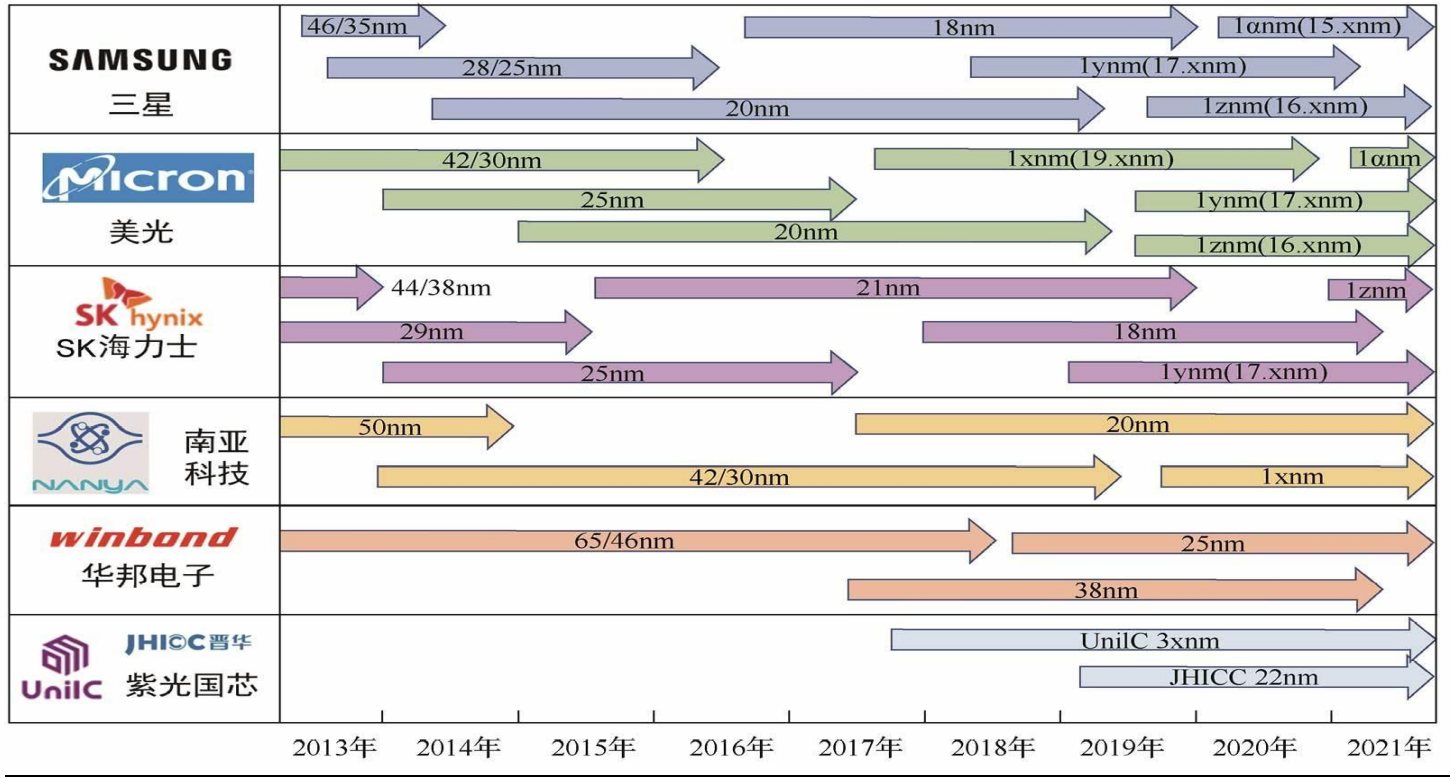
DRAM工艺：未来有望攻克9nm生产工艺

国信证券 GUOSEN SECURITIES

■ **DRAM工艺：进入1znm节点后，制造复杂度急剧增加，DRAM单元微缩变得非常困难。**

- **目前停留在1znm节点：**2008年-2016年，DRAM工艺逐步从4xnm级（49-40nm）进步到1xnm，但2016年至今，DRAM工艺仍然停留在1xnm级；这一阶段，不同厂商将不同技术命名为1xnm（19-18nm）、1ynm（17nm）、1znm（16nm）、1αnm、1βnm等，随着工艺节点的进步，DRAM制造工艺愈加复杂，且电容的制造成本也越来越高；
- **未来有望攻克9nm生产工艺：**三星等头部厂商已经将“9nm级”纳入其长期技术路径。

图107：DRAM生产工艺演进



资料来源：赵巍胜等著-《“芯”制造——集成电路制造技术链》-人民邮电出版社（2024年）-P209，国信证券经济研究所整理

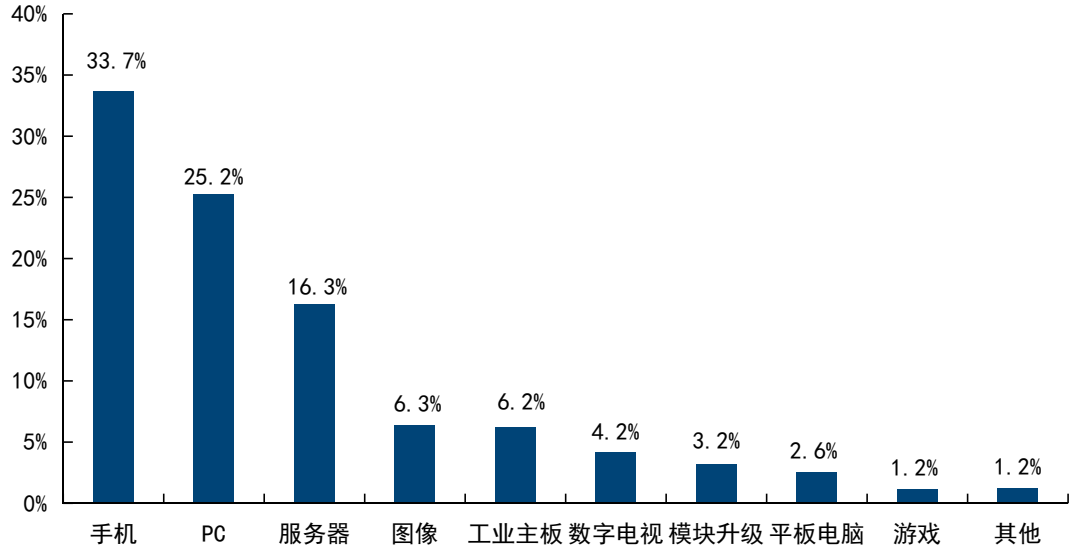
DRAM下游需求：手机、PC、服务器合计占比为75.2%

国信证券 GUOSEN SECURITIES

■ DRAM下游需求：手机、PC、服务器合计占比为75.2%。

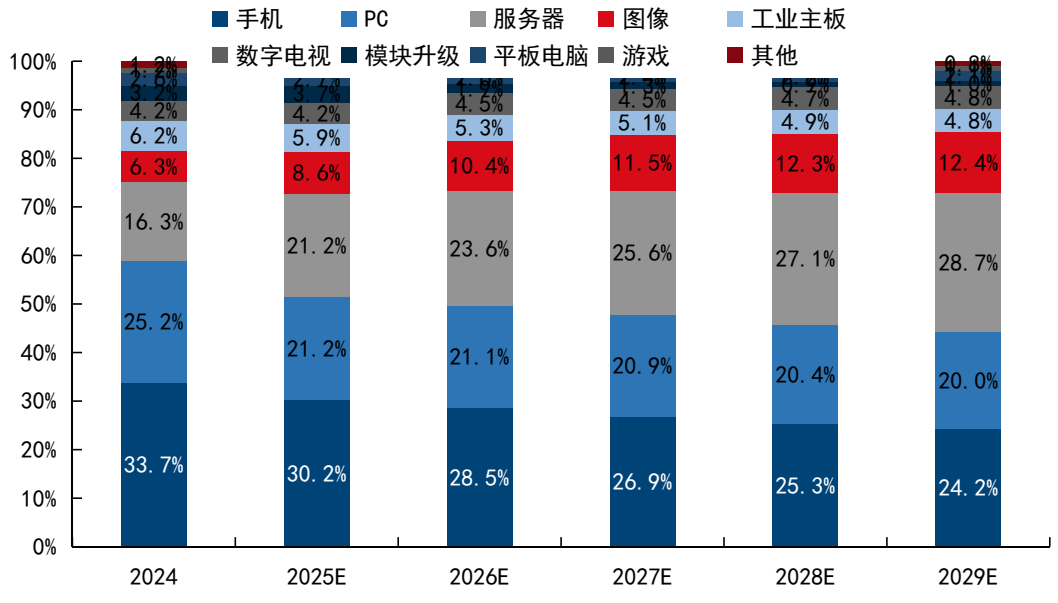
- **DRAM下游需求情况：**根据IDC披露数据，2024年DRAM下游需求领域主要包括手机、PC、服务器、图像、工业主板、数字电视、模块升级、平板电脑、游戏、其他等，其中，手机、PC、服务器占比分别为33.7%、25.2%、16.3%，合计为75.2%，为DRAM主要需求领域；
- **DRAM下游需求变动情况：**手机、PC行业销量增速放缓，预计占比持续下滑，根据IDC披露数据，预计2029年占比分别为24.2%、20.0%，相较于2024年分别-9.4、-5.2个pct；此外，随着AI服务器快速放量，大模型需要高容量的DRAM作为主存储器，服务器用DRAM、图像用DRAM（包括HBM和游戏主机GDDR）占比有望持续增长，预计2029年占比将分别提升至28.7%、12.4%。

图108：2024年DRAM下游需求分布



资料来源：IDC，国信证券经济研究所整理

图109：DRAM下游需求占比及预测



资料来源：IDCv，国信证券经济研究所整理

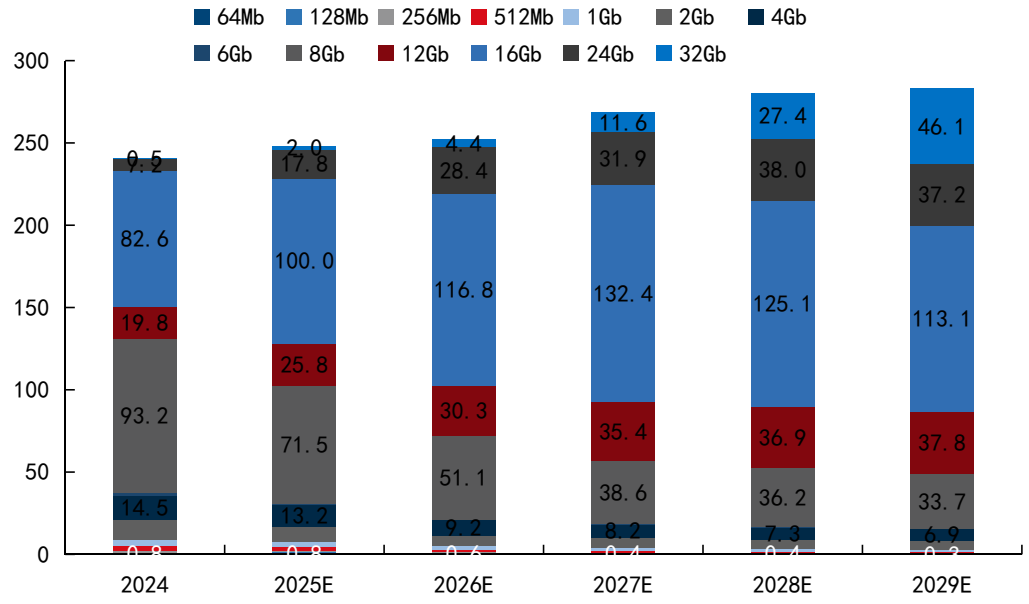
DRAM下游需求：8GB和16GB为主，高容量占比持续提升

国信证券 GUOSEN SECURITIES

■ **DRAM下游需求：8GB和16GB为主，高容量占比持续提升。**

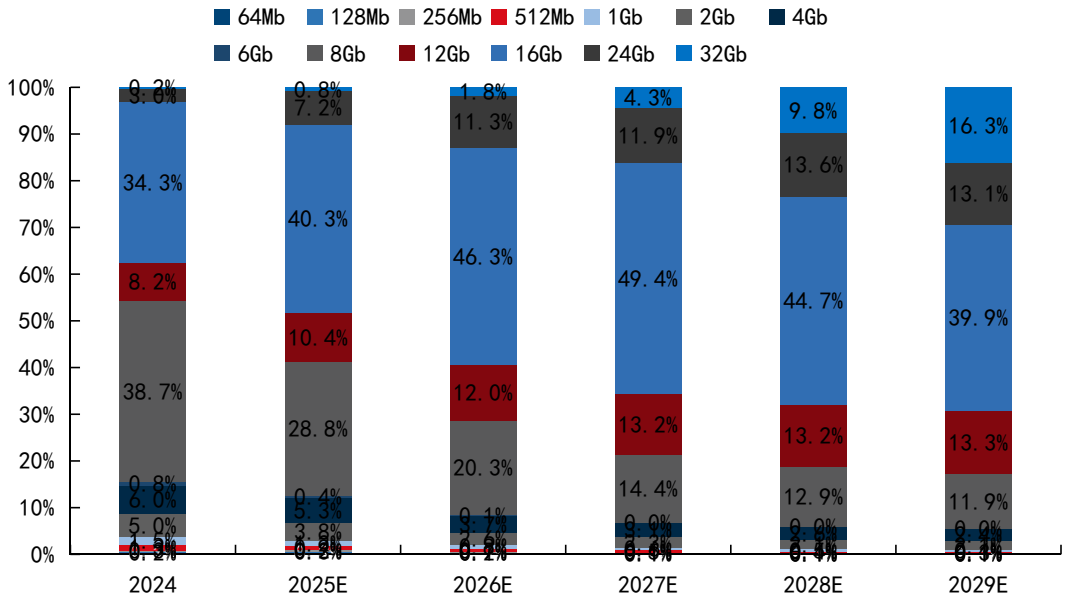
- **DRAM各容量出货量情况：**根据IDC披露数据，2024年1GB以下DRAM出货量合计为5.3亿颗，占比为2.2%；1GB-6GB DRAM出货量为32.2亿颗，占比为13.4%；**8-16GB出货量195.6亿颗，占比为81.2%，为主力出货容量，其中8GB、12GB、16GB出货量分别为93.2、19.8、82.6亿颗，占比分别为38.7%、8.2%、34.3%**；16GB以上出货量为7.7亿颗，占比为3.2%；
- **DRAM高容量出货量占比持续提升：**根据IDC预测数据，预计16GB、24GB、32GB等高容量出货量占比将持续提升，预计2029年分别提升至39.9%、13.1%、16.3%，分别+5.6、+10.1、16.0个pct。

图110：DRAM各容量出货量情况及预测（单位：亿颗）



资料来源：IDC，国信证券经济研究所整理

图111：DRAM各容量出货量占比及预测



资料来源：IDC，国信证券经济研究所整理

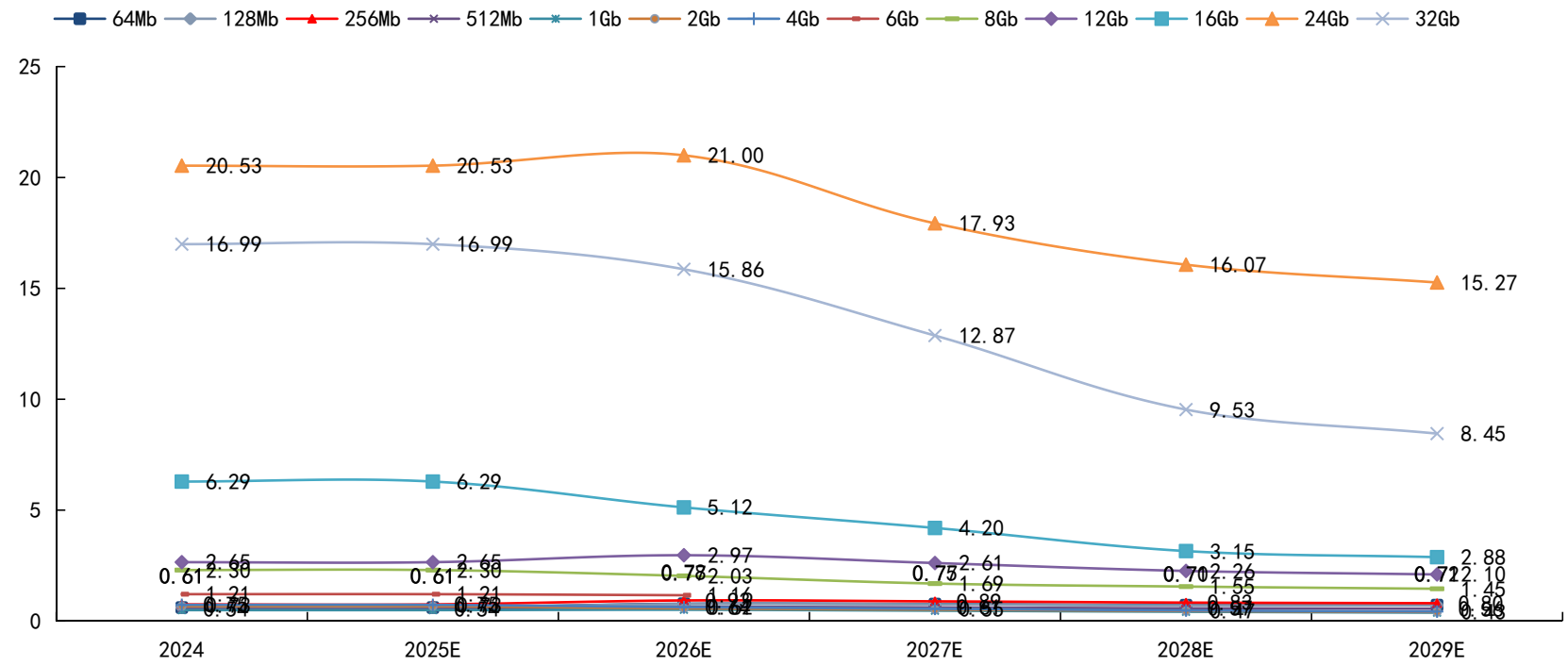
DRAM价格：ASP呈下降趋势

国信证券 GUOSEN SECURITIES

■ **DRAM价格：ASP呈下降趋势，高容量DRAM 27年开始加速下滑。**

- **随着容量提升，ASP提升，且高容量DRAM ASP明显高于低容量产品：**根据IDC披露数据，2024年64Mb-4GB价格区间基本在0.61-0.73美元/颗之间，基本随着容量的提升，ASP随之提升；6GB ASP提升至1.21美元/颗，8GB、12GB ASP分别为2.30、2.65美元/颗，16GB、24GB、32GB ASP分别为6.29、20.53、16.99美元/颗，ASP大幅提升；
- **高容量DRAM ASP 27年开始加速下滑：**根据IDC披露数据，短期由于AI驱动，需求强劲，且供给侧控制产能，预计短期价格维稳或持续上涨，预计本轮周期27年达到顶峰，后DRAM ASP随之下滑。

图112：DRAM各容量ASP情况（单位：美元/颗）



资料来源：IDC，国信证券经济研究所整理

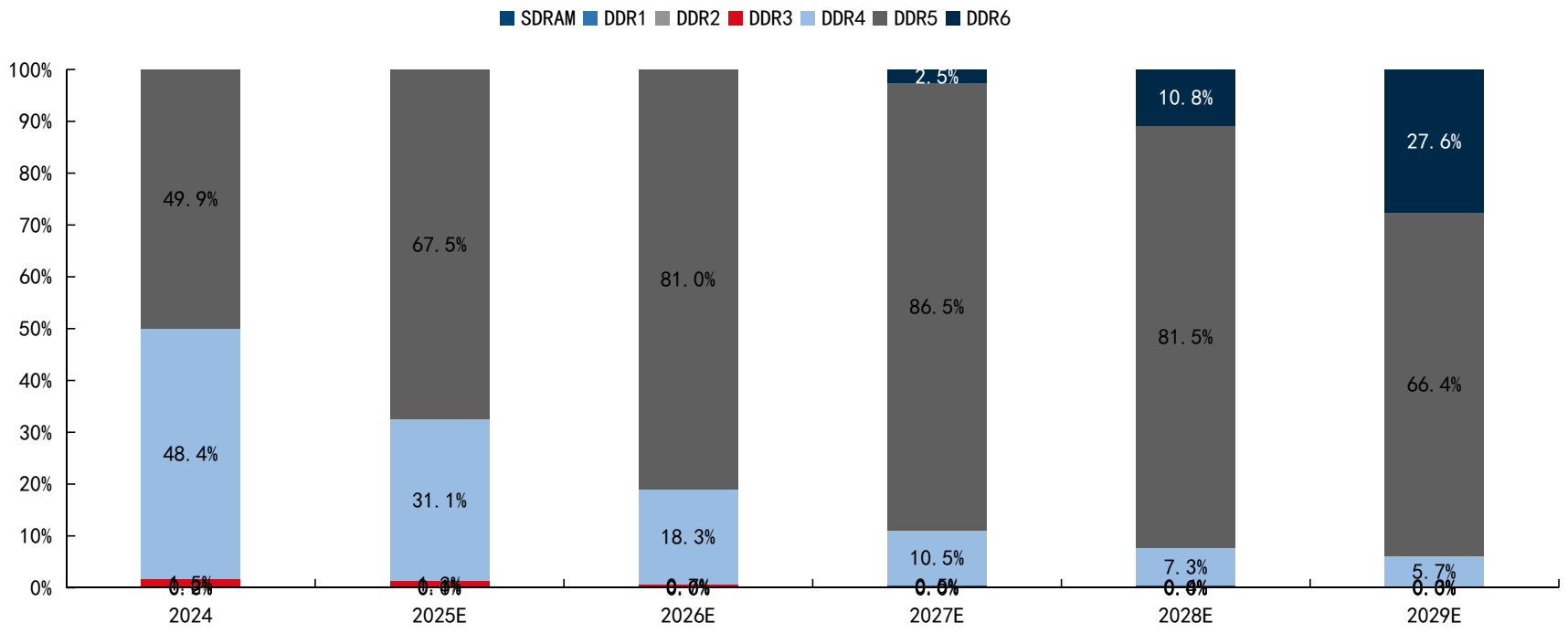
DRAM各类型出货量：DDR4、DDR5为出货主力

国信证券 GUOSEN SECURITIES

■ DRAM各类型出货量：DDR4、DDR5为出货主力。

- **DDR4、DDR5为出货主力：**根据IDC披露数据，2024年DDR4、DDR5出货量占比分别为48.4%、49.9%，合计为98.3%，为出货主力；
- **DDR6 28年开始加速渗透：**根据IDC披露数据，随着DDR6技术的成熟，2027年开始逐步渗透，预计出货量占比为2.5%，2028开始加速渗透，预计2028年、2029年出货量占比占比分别为10.8%、27.6%。

图113：DRAM各类型出货量占比



资料来源：IDC，国信证券经济研究所整理

DRAM各类型出货量（按容量）：

■ **DRAM各类型出货量占比（按容量）：**根据IDC披露数据，128MB-512MB DRAM主要以SDRAM、DDR1、DDR2为主，1GB-2GB DRAM主要以DDR3为主，4GB-8GB DRAM主要以DDR4为主，12GB-32GB DRAM主要以DDR5为主，其中16GB、32GB将逐步切换为DDR6。

图114：DRAM各类型出货量占比（按容量）

		2024	2025E	2026E	2027E	2028E	2029E
128MB	SDRAM	54.9%	68.7%	65.8%	58.3%	52.5%	46.9%
	DDR 1	19.5%	7.7%	9.4%	12.4%	14.8%	16.7%
	DDR 2	25.6%	23.6%	24.8%	29.3%	32.7%	36.4%
256MB	SDRAM	26.1%	36.8%	30.2%	25.3%	22.2%	17.9%
	DDR 1	48.1%	29.2%	31.0%	33.3%	33.4%	32.9%
	DDR 2	25.8%	34.0%	38.9%	41.4%	44.5%	49.2%
512MB	SDRAM	0.5%	0.9%	0.5%	0.0%	0.0%	0.0%
	DDR 1	2.0%	2.3%	2.0%	1.7%	1.3%	0.8%
	DDR 2	96.8%	95.3%	94.7%	93.5%	92.4%	89.7%
	DDR 3	0.8%	1.5%	2.9%	4.8%	6.3%	9.4%
1GB	SDRAM	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%
	DDR 1	0.8%	0.8%	0.6%	0.4%	0.1%	0.0%
	DDR 2	37.3%	27.5%	19.4%	15.3%	14.5%	12.8%
	DDR 3	60.3%	67.5%	75.1%	79.7%	81.1%	84.6%
	DDR 4	1.5%	4.0%	4.9%	4.6%	4.3%	2.6%
2GB	DDR 1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	DDR 2	5.0%	6.4%	2.1%	0.4%	0.2%	0.1%
	DDR 3	37.4%	49.0%	73.1%	82.1%	92.1%	99.4%
	DDR 4	57.6%	44.6%	24.8%	17.5%	7.7%	0.5%
	XDR				0.0%	0.0%	0.0%
4GB	DDR 2	1.5%	0.7%	0.6%	0.1%	0.0%	0.0%
	DDR 3	43.1%	49.5%	32.0%	21.6%	17.0%	9.0%
	DDR 4	55.5%	49.8%	67.4%	78.3%	83.0%	91.0%
6GB	DDR 2						
	DDR 3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	DDR 4	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
8GB	DDR 3	0.3%	0.3%	0.3%	0.0%	0.0%	0.0%
	DDR 4	95.5%	96.4%	97.0%	97.6%	98.1%	98.4%
	DDR 5	4.2%	3.3%	2.7%	2.4%	1.9%	1.6%
12GB	DDR 4	44.2%	23.0%	4.1%	1.0%	0.0%	0.0%
	DDR 5	55.8%	77.0%	95.9%	99.0%	100.0%	100.0%
16GB	DDR 4	29.1%	18.0%	10.7%	4.2%	1.4%	0.2%
	DDR 5	70.9%	82.0%	89.2%	92.8%	87.7%	72.1%
	DDR 6			0.1%	3.0%	10.9%	27.8%
24GB	DDR 5	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
	DDR 6	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
32GB	DDR 5	100.0%	100.0%	99.9%	89.5%	69.4%	45.0%
	DDR 6			0.1%	10.5%	30.6%	55.0%

资料来源：IDC，国信证券经济研究所整理

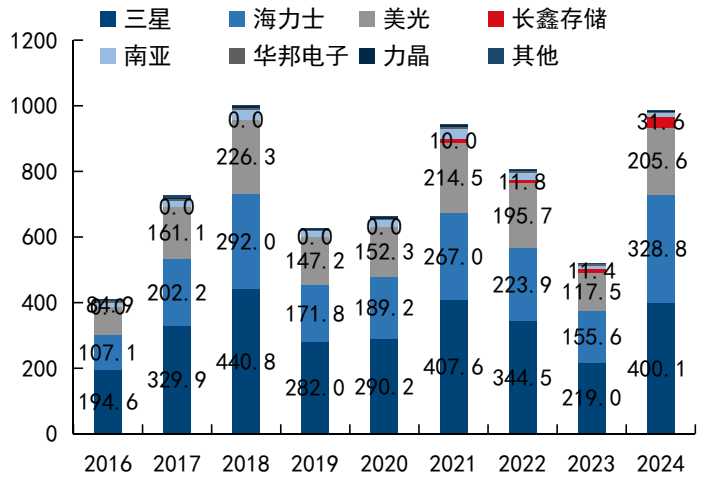
DRAM供给侧：三星、海力士、美光合计占比95.1%

国信证券 GUOSEN SECURITIES

■ DRAM供给侧：三星、海力士、美光为行业龙头。

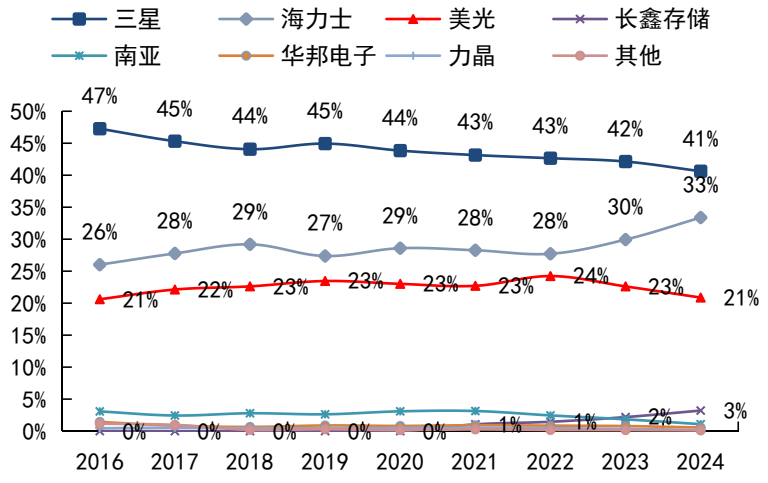
- **市占率情况：**根据IDC披露数据，25Q2全球DRAM竞争格局，海力士、三星、美光占比分别为39.1%、33.2%、22.8%，合计占比为95.1%，为行业龙头，其中三星、海力士为韩国企业，美光为美国企业；我国长鑫存储市占率持续提升，25Q2达到3.1%；
- **市占率变化情况：**根据IDC披露数据，三星市占率略有下降，海力士的市占率在持续提升，美光市占率基本稳定。

图115：DRAM各公司收入情况



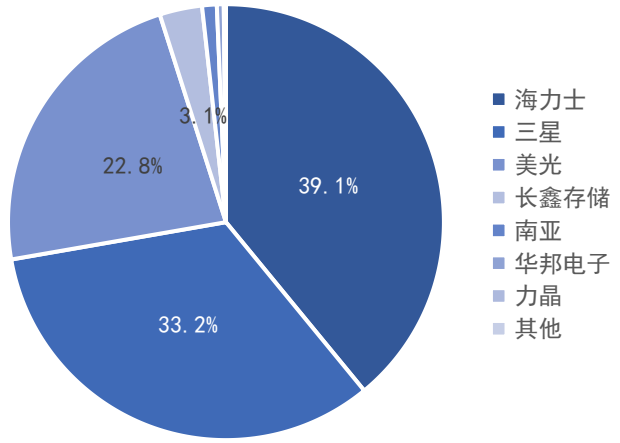
资料来源：IDC，国信证券经济研究所整理

图116：DRAM各公司市占率变化



资料来源：IDC，国信证券经济研究所整理

图117：25Q2 DRAM竞争格局



资料来源：IDC，国信证券经济研究所整理

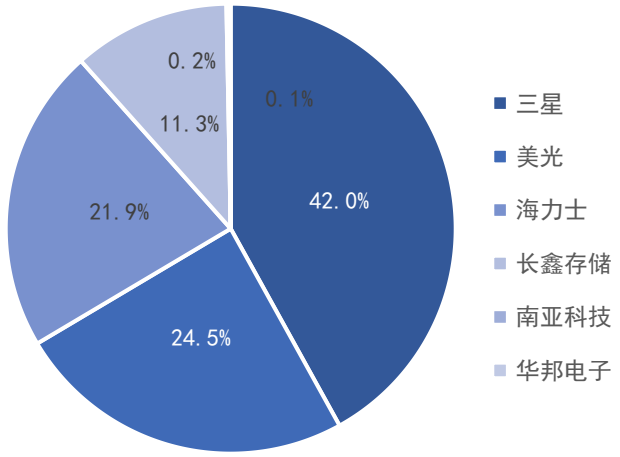
DRAM下游细分供给：三星、海力士、美光为服务器、图像用DRAM主要供应商

国信证券 GUOSEN SECURITIES

■ **DRAM下细分供给：三星、海力士、美光为服务器、图像用DRAM主要供应商。**

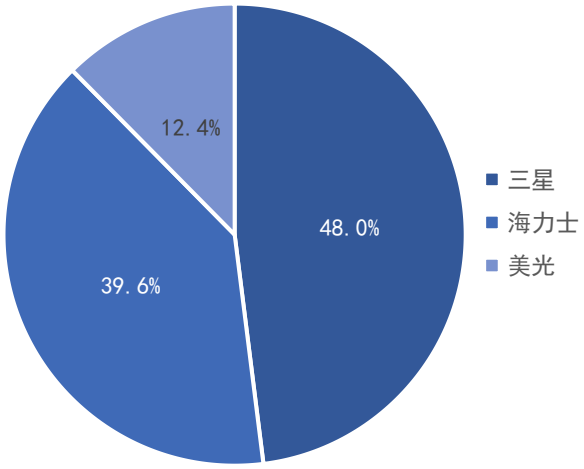
- **手机领域：**根据IDC披露数据，2024年手机领域DRAM供应商主要有三星、美光、海力士、长鑫存储，占比分别为42.0%、24.5%、21.9%、11.3%，合计占比为99.7%；
- **图像领域：**根据IDC披露数据，图像用DRAM供应商主要为三星、海力士、美光，占比分别为48.0%、39.6%、12.4%；
- **服务器领域：**根据IDC披露数据，服务器领域供应商分别为三星、海力士、美光，占比分别为42.5%、36.0%、21.4%。

图118：手机领域DRAM供应商格局（2024年）



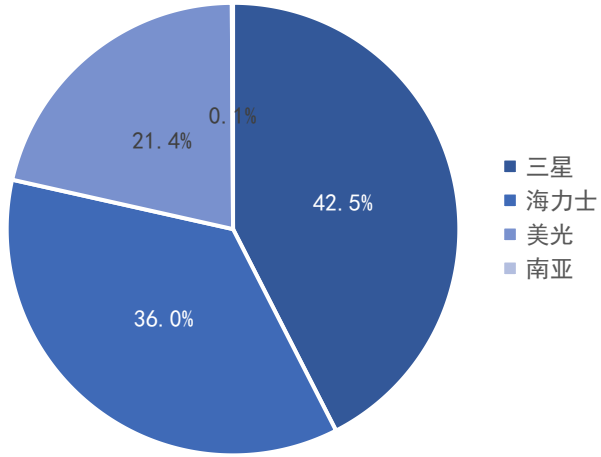
资料来源：IDC，国信证券经济研究所整理

图119：图像领域DRAM供应商格局（2024年）



资料来源：IDC，国信证券经济研究所整理

图120：服务器领域DRAM供应商格局（2024年）



资料来源：IDC，国信证券经济研究所整理

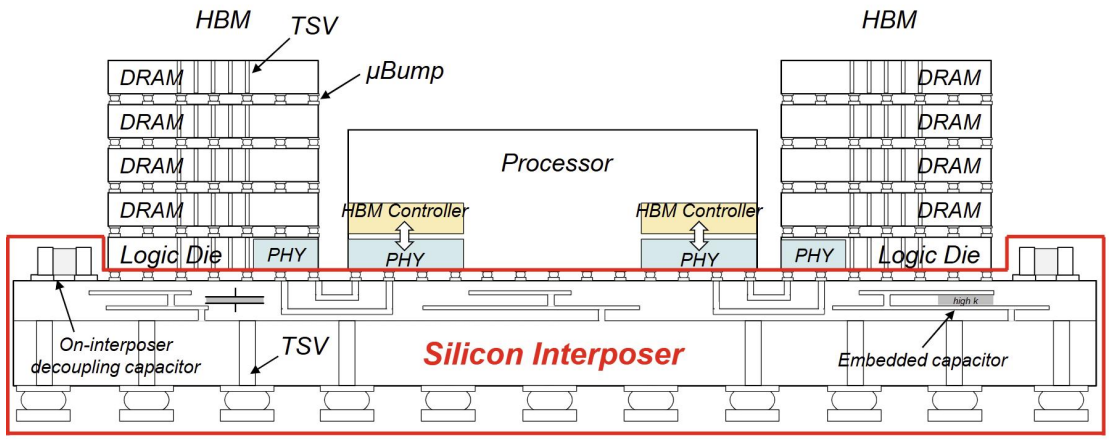
HBM：更高的显存位宽和带宽

国信证券 GUOSEN SECURITIES

- **HBM（High-Bandwidth Memory，高带宽内存）：**多层DRAM芯片堆叠，通过TSV实现垂直方向的互联，即HBM通过TSV（硅通孔，在数个DRAM芯片搭配数千个细微孔并通过垂直贯通的电极连接上下芯片的技术，通过贯通所有芯片层的柱状通道传输信号、指令、电流）及微凸点将4层DRAM芯片层与1层基本逻辑控制芯片实现三维堆叠，进而具有更高的存储密度和更大的带宽，以及更大规模的I/O端口；其核心价值在于提供了更多的I/O数量，通过增加位宽的方式尽可能降低了GPU访问DRAM的延迟。此外，GPU和DRAM先通过Bump（微凸点）和Interposer（互联功能硅片）连通，Interposer通过Bump和Substrate（封装基板）连通到BALL，最后由BGA BALL连接到PCB。

- **HBM具备更高的显存位宽和带宽：**根据SemiAnalysis披露数据，同DDR5、LPDDR5、GDDR6X相比，HBM3在显存位宽（Bus Width）大幅提升（通过TSV技术），进而带宽（Bandwidth，传输速率*位宽）大幅提升。

图121：HBM架构图



资料来源：KAIST，国信证券经济研究所整理

图122：DDR4与LPDDR4、GDDR5参数对比

Memory	Data Rate（Gbps）	Bus Width（bits）	Bandwidth（GB/s）
DDR5	8.8	64	70.4
LPDDR5	8.5	32	34.1
GDDR6X	24	32	96
HBM3	6.4	1024	819.2

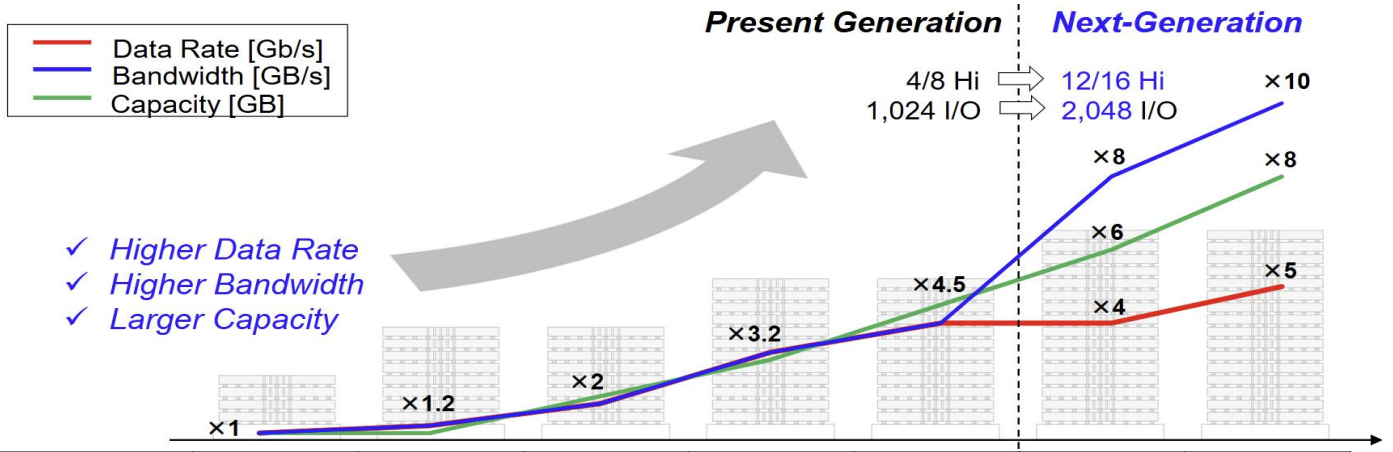
资料来源：SemiAnalysis，国信证券经济研究所整理

HBM：速率、带宽、容量、叠层数持续提升

国信证券 GUOSEN SECURITIES

■ **HBM的速率、带宽、容量、叠层数持续提升。**HBM逐步从4/8Hi（叠层）提升至12/16Hi，同时I/O数量有望从1024提升至2048，进而HBM带宽、容量快速提升。

图123：HBM演进历史



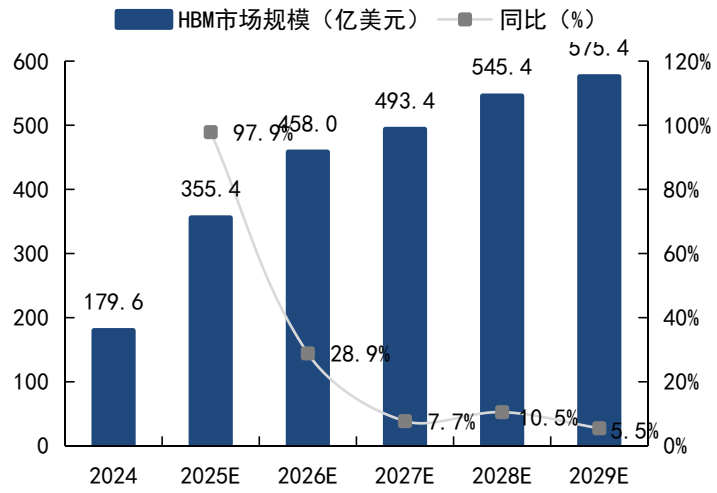
	HBM (2015)	HBM2 (2018)	HBM2E (2020)	HBM3 (2022)	HBM3E (2024)	HBM4 (2026)	HBM4E (*2028)
Data Rate	2 Gbps	2.4 Gbps	3.6 Gbps	6.4 Gbps	8 Gbps	8 Gbps	10 Gbps
# of I/O	1,024					2,048	
Bandwidth	256 GB/s	307 GB/s	461 GB/s	819 GB/s	1.0 TB/s	2.0 TB/s	2.5 TB/s
Capacity/die	8 Gb		16 Gb		24 Gb		32 Gb
# of stack die	4/8-Hi			8/12-Hi		12/16-Hi	
Capacity/HBM	4/8 GB		8/16 GB	16/24 GB	24/36 GB	36/48 GB	48/64 GB
Power/HBM	4 W	10 W	19 W	25 W	32 W	43 / 75 W	48 / 80 W
Cooling Method	Thermo-Electric Cooling (TEC) w/ Heatsink		Direct-to-Chip (D2C) Liquid Cooling				

资料来源：KAIST，国信证券经济研究所整理

HBM市场规模：2024年全球市场规模为179.62亿美金

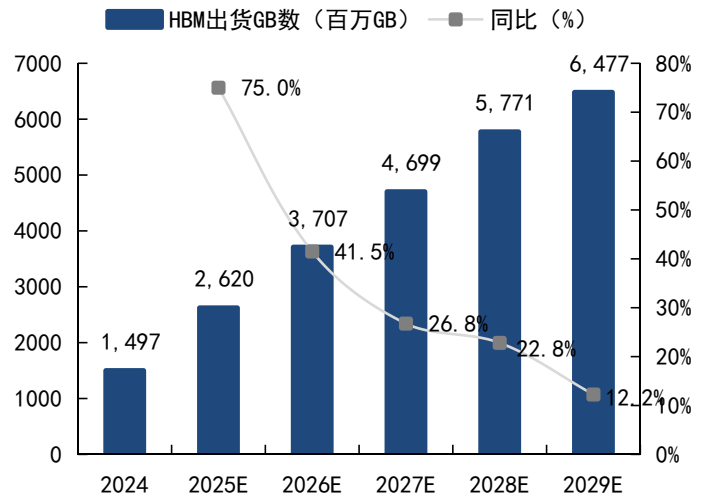
- **HBM市场规模：**根据IDC披露数据，2024年全球HBM市场规模为179.62亿美金，受益于全球AI芯片对HBM的需求，预计2025、2026年需求加速增长，对应24-26年CAGR为59.7%；远期来看，预计2029年HBM市场规模将达到575.4亿美金，对应24-29年CAGR为26.2%；
- **HBM全球出货量：**根据IDC披露数据，2024年全球HBM出货量为14.97亿GB，预计2029年将达到64.77亿GB，对应24-29年CAGR为34.0%；
- **HBM单GB价格逐步下滑：**根据IDC披露数据，2024年HBM单价为12.00美元/GB，随着堆叠等技术的发展，预计HBM单GB价格将逐步下滑；

图124：HBM市场规模及预测（单位：亿美元）



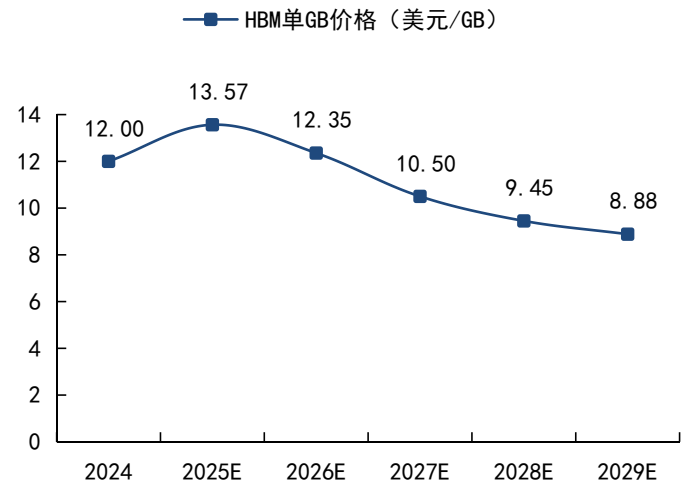
资料来源：IDC，国信证券经济研究所整理

图125：HBM出货量情况（百万GB）



资料来源：IDC，国信证券经济研究所整理

图126：HBM单GB价格情况（美元/GB）



资料来源：IDC，国信证券经济研究所整理

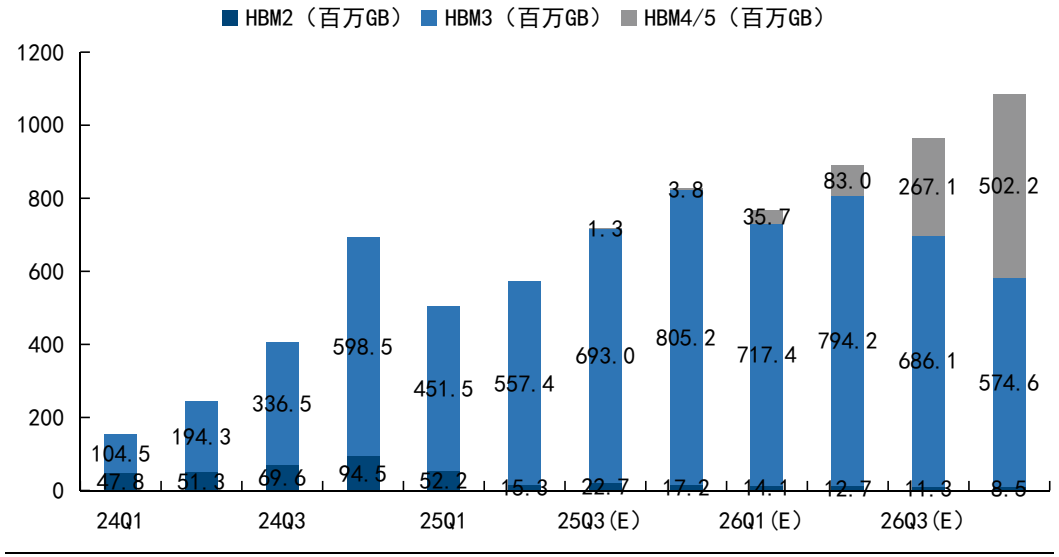
HBM市场规模：2025年HBM3为出货主力，HBM4将逐步起量

国信证券 GUOSEN SECURITIES

■ HBM分类型出货量情况：2025年HBM3为出货主力，HBM4将逐步起量。

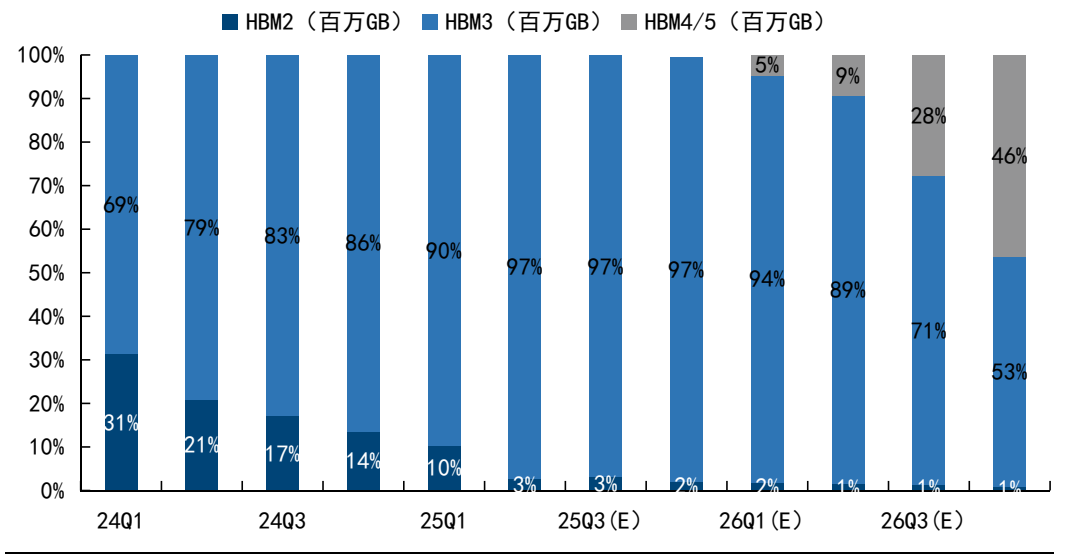
- 根据IDC披露数据，2024年HBM2、HBM3出货量分别为2.63、12.34亿GB，占比分别为17.6%、82.4%；
- 2025年底HBM4开始少量出货，HBM2、HBM3、HBM4出货量分别为1.07、25.07、0.05亿GB，占比分别为4.1%、95.7%、0.2%，HBM3仍为出货主力；
- 2026年HBM4出货量快速增长，预计26Q4 HBM4出货量基本追平HBM3，预计全年HBM2、HBM3、HBM4出货量分别为0.47、27.72、8.88亿GB，占比分别为1.3%、74.8%、24.0%。

图127：HBM分类型出货量情况（单位：百万GB）



资料来源：IDC，国信证券经济研究所整理

图128：HBM分类型出货量占比

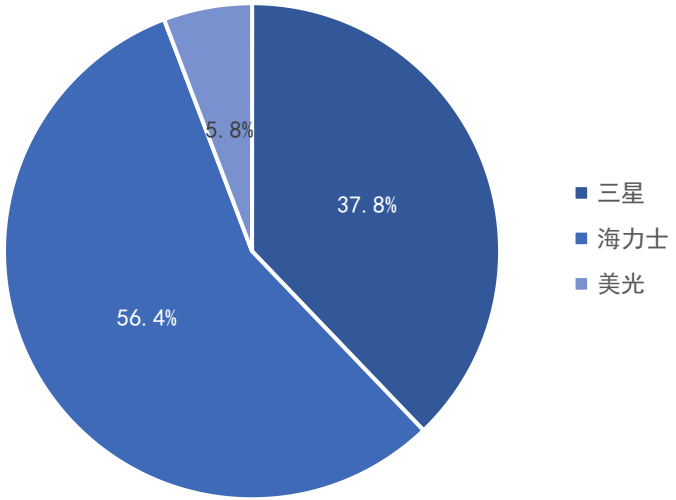


资料来源：IDC，国信证券经济研究所整理

HBM供给侧：海力士为全球龙头

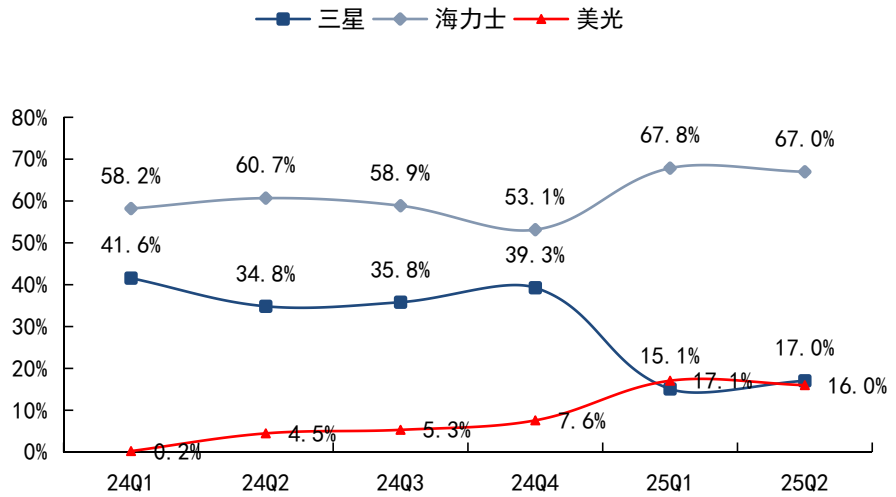
- **全球HBM市场竞争格局：**根据IDC披露数据，2024年全球HBM供应商主要有海力士、三星、美光，其占比分别为56.4%、37.8%、5.8%，海力士、三星占据主要市场份额；
- **全球HBM厂商市占率变化：**根据IDC披露数据，海力士HBM市占率基本维稳，三星市占率2025年开始快速下降，美光市占率从24Q1的0.2%提升至25Q2的16.0%，市占率快速提升。

图129：2024年全球HBM市场竞争格局



资料来源：IDC，国信证券经济研究所整理

图130：HBM厂商市占率变化情况



资料来源：IDC，国信证券经济研究所整理

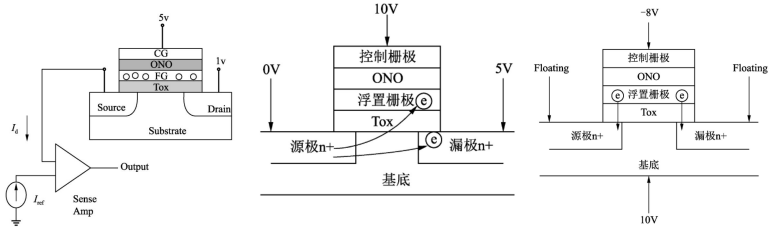
Flash存储：NOR Flash和NAND Flash

国信证券 GUOSEN SECURITIES

- **Flash存储：**使用浮栅场效应管（Floating Gate FET）来存储数据，其共有4个端电极，分别为源极(S)、漏极(D)、控制栅极(Control Gate)、浮置栅极(Floating Gate)，前三个端电极与普通MOSFET相同，区别在于浮置栅极，Flash利用浮置栅极是否存储电荷表表征数字“0”和“1”（即浮置栅极注入电荷后，D极和S极存在导电通道，从D极读到“0”）。
  - **NOR Flash：1）密度低，容量小：**NOR为“非或”，则位线是并联的，进而金属导线占用了很大面积，所以存储密度较低，因而适合存储代码，不适合存储数据；**2）读效率高：**带有SRAM接口，有足够的地址引脚来寻址，很容易地存取其内部每一字节；同时，并联结构决定了其存储单元可独立寻址，并随机访问，读取效率高，适合代码存储，程序可以直接在NOR Flash中运行（RAM特性）；**3）写操作效率低：**采用热电子注入方式，效率较低，不适用于频繁擦除/写入场合；**4）擦除效率低：**要求在擦除前先要将目标块内所有地位都写为0；
  - **NAND Flash：**典型结构是具有两个多晶硅栅极，其中一个有外接电路连接（即控制栅极，其上的连线为字线），另外一个没有外引线的栅极，其被完全包括在一层氧化物介质层里面，进而使浮空的，成为浮置栅极。

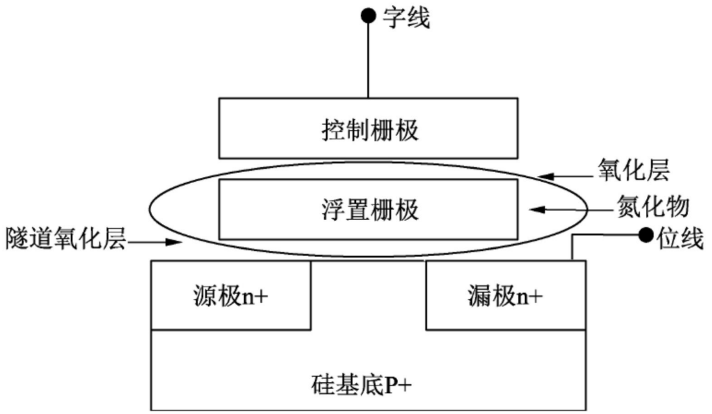
图131：NOR Flash写操作和擦除操作

NOR Flash读操作	NOR Flash写操作	NOR Flash擦除操作
在控制栅极（CG）施加一个电压，漏极（Drain）加上一个小的正电压，通过比较Drain电流和参考电流的大小关系来判断存储单元存储的是“1”或“0”；	热电子注入方式，漏极施加电压，电子加速向漏极方向移动，成为热电子，同时控制栅极加上高电压，热电子就会通过绝缘层（Tox）进入浮置栅极；	隧穿方式，在控制栅极加负电压，在阱上加正电压，而源极、漏极悬空，进而浮置栅极的电子穿过绝缘层（Tox）被拉下来；



资料来源：夏鲁宁等著-《固态存储：原理、架构与数据安全》-机械工业出版社（2017年）-P164，国信证券经济研究所整理

图132：NAND Flash架构



资料来源：夏鲁宁等著-《固态存储：原理、架构与数据安全》-机械工业出版社（2017年）-P169，国信证券经济研究所整理

图133：NAND Flash与NOR Flash对比

项目	NOR Flash	NAND Flash
连接方式	存储单元之间并联	存储单元之间串联
读性能	较NAND Flash更快	较NOR Flash更慢
写性能	较NAND Flash更慢	较NOR Flash更快
擦除性能	较NAND Flash更慢	较NOR Flash更快
可靠性	可靠性高	可靠性相对较低
容量	密度低，容量小	密度高，容量大
成本	高	低
适用范围	适用于存储代码	适用于存储数据

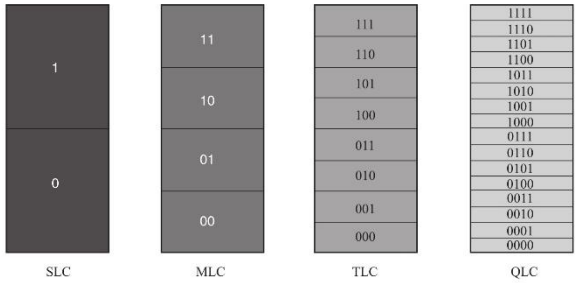
资料来源：夏鲁宁等著-《固态存储：原理、架构与数据安全》-机械工业出版社（2017年）-P156，国信证券经济研究所整理

Flash：NAND持续提升存储密度，存储单元表达的bit数持续增长

国信证券 GUOSEN SECURITIES

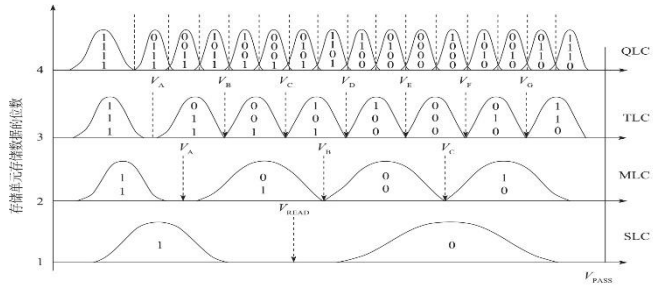
■ **通过存储单元存储更多位的数据来提升存储密度：**先后出现了SLC（Single Level Cell）、MLC（Multiple Level Cell）、TLC（Triple Level Cell）、QLC（Quad Level Cell），可分别存储1位、2位、3位、4位数据，单位存储单元表达的bit数持续增长，但其读写性能、稳定性、寿命持续下降，同时成本大幅降低。目前，SSD主流存储介质是TLC，SLC和MLC成本太高（SLC基本在对成本不敏感且对可靠性要求高的场合，例如航空航天），QLC寿命有限，目前应用在消费级或者读密集型企业级应用，未来随着技术的发展，QLC有望成为主流存储介质。

图134：不同闪存类型的状态划分和编码示例



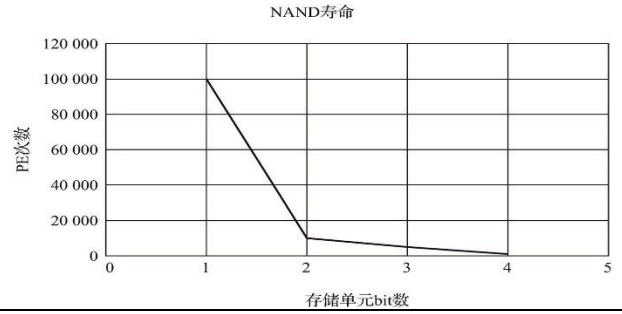
资料来源：SSDFans等著-《深入浅出SSD：固态存储技术核心、原理与实战》-机械工业出版社（2023年）-P208，国信证券经济研究所整理

图135：不同闪存类型的阀值电压分布示意图



资料来源：SSDFans等著-《深入浅出SSD：固态存储技术核心、原理与实战》-机械工业出版社（2023年）-P210，国信证券经济研究所整理

图136：不同闪存类型寿命示意图



资料来源：阿伦等著-《深入浅出SSD测试：固态存储测试流程、方法与工具》-机械工业出版社（2025年）-P29，国信证券经济研究所整理

图137：不同闪存类型参数对比

Memory	SLC	MLC	TLC	QLC
每单元容量	1bit	2bits	3bits	4bits
读延迟	25μs	50μs	75μs	>100μs
写入速度	200~300μs	600~900μs	900~1350μs	>1500μs
擦除时间	1.2~2ms	3ms	5ms	>6ms
P/E次数	100000	3000	1000	100
价格	极高	高	中	低
典型容量	1~32Gbit	32~128Gbit	128~256Gbit	>256Gbit

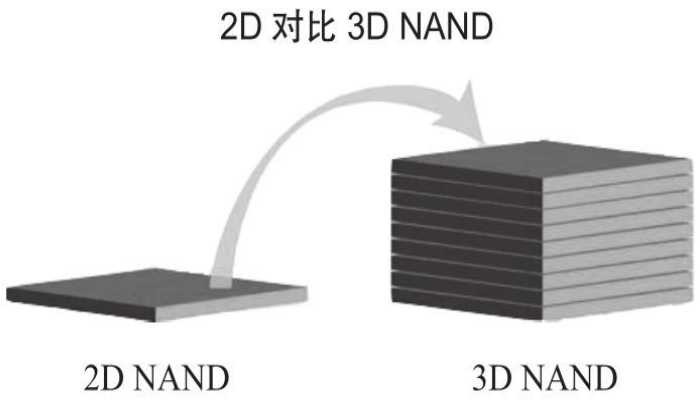
资料来源：杨剑等著-《边缘计算系统设计与实践》-北京大学出版社（2023年）-P30，国信证券经济研究所整理

Flash：2D NAND->3D NAND，堆叠层数持续增长

国信证券 GUOSEN SECURITIES

■ **2D NAND -> 3D NAND，堆叠层数持续增长：**为了提升存储密度，除了通过存储单元存储更多位的数据外，亦可以通过堆叠等方式，大幅提升存储密度；以2016年三星发布的48层3D V-NAND为例，其每平方毫米能生产出2600Mb的数据，是2D NAND的3倍，进而每GB成本大幅降低；后续，各厂商持续增加堆叠层数。

图138：2D NAND转向3D NAND



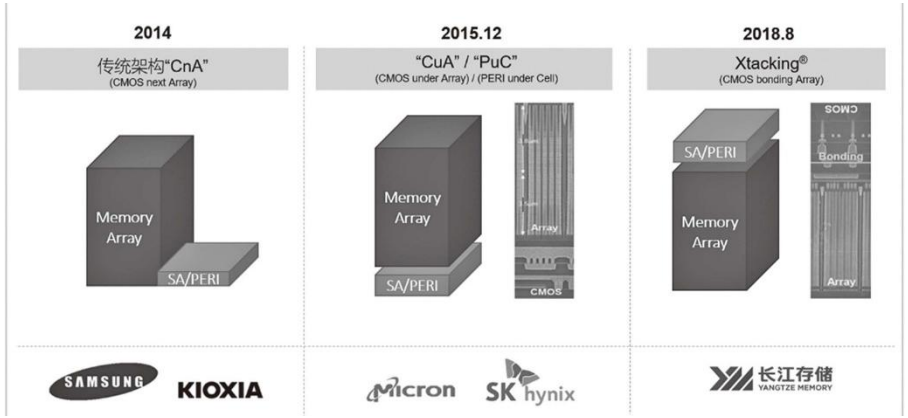
资料来源：SSDFans等著-《深入浅出SSD：固态存储技术核心、原理与实战》-机械工业出版社（2023年）-P48，国信证券经济研究所整理

图139：3D NAND密度大幅提升

对比项	海力士 16nm	三星 16nm	三星 48L V-NAND
年份	2014	2015	2016
制程节点/nm	16	16	21
Die 容量/Gb	64	64	256
Die 面积/mm^2	93	86.4	99
密度/(Mb/mm^2)	690	740	2 600

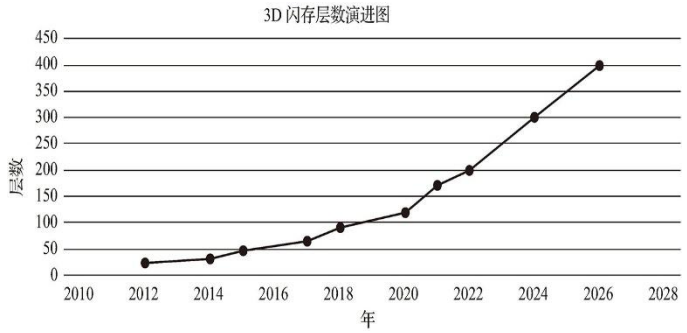
资料来源：SSDFans等著-《深入浅出SSD：固态存储技术核心、原理与实战》-机械工业出版社（2023年）-P48，国信证券经济研究所整理

图140：业界3D闪存架构对比



资料来源：TechInsight，国信证券经济研究所整理

图141：2D NAND转向3D NAND



资料来源：FMS 2022，国信证券经济研究所整理

图142：3D NAND拓展的几种方式

1	Z Scaling	增加3D闪存层数
2	XY Scaling	2D平面增加位密度
3	Architecture Scaling	架构扩展
4	Logical Scaling	存储单元比特位扩展

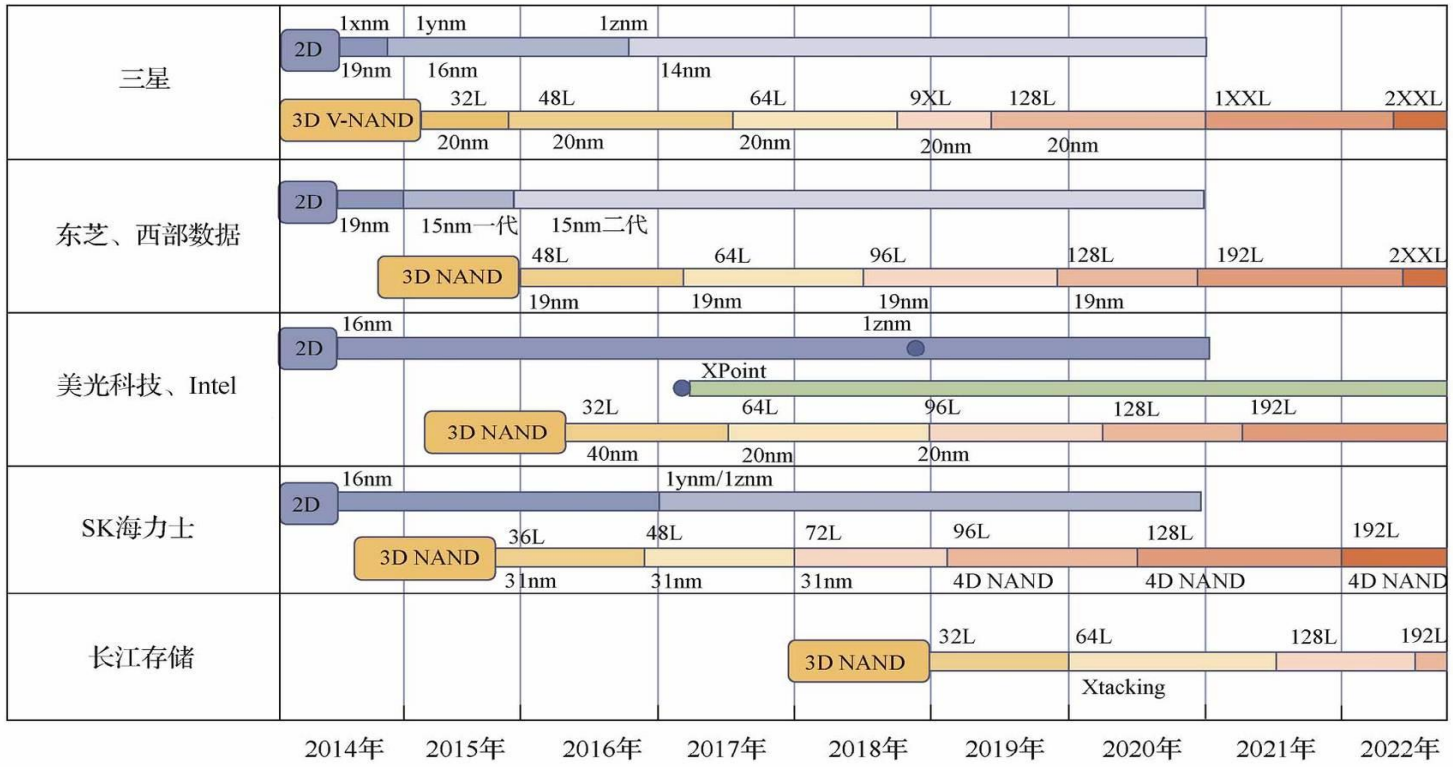
资料来源：阿伦等著-《深入浅出SSD测试：固态存储测试流程、方法与工具》-机械工业出版社（2025年）-P409，国信证券经济研究所整理

Flash工艺：三星400层NAND有望量产

国信证券 GUOSEN SECURITIES

■ **NAND工艺：堆叠层数持续增长，25年400层NAND有望量产。**3D NAND阶段，各家大厂致力于增加堆叠层数，2019年开始进入100层+堆叠NAND，2022年进入200层+堆叠NAND，2024年开始量产300层+堆叠NAND，2025年400层+堆叠NAND有望量产。

图143：Flash生产工艺演进



资料来源：赵巍胜等著-《“芯”制造——集成电路制造技术链》-人民邮电出版社（2024年）-P214，国信证券经济研究所整理

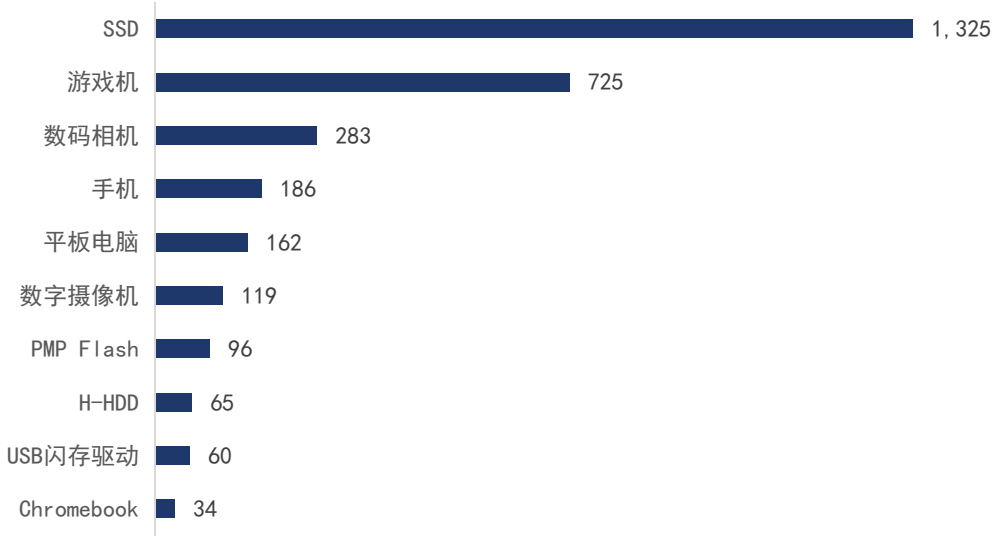
NAND不同领域单系统需求情况：SSD单系统需求持续增长

国信证券 GUOSEN SECURITIES

- **单系统NAND需求情况：**根据IDC披露数据，SSD单系统NAND需求为1325Gbytes，其次为游戏机725Gbytes，数码相机、数字摄像机、手机、平板电脑单系统NAND需求基本在100-300Gbytes，PMP Flash、H-HDD、USB闪存驱动、Chromebook单系统NAND需求较低，基本在100Gbytes以下；

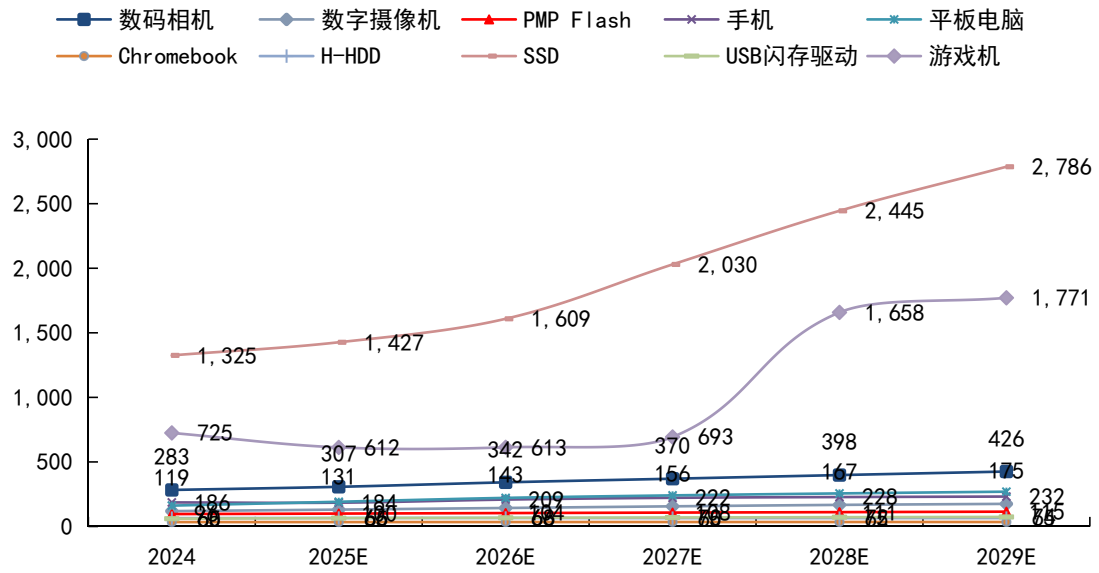
- **SSD、手机、游戏机单系统NAND需求预计将持续提升：**1）SSD：随着PC端1TB SSD渗透率提升以及企业需求复苏，单系统NAND需求预计持续提升；2）手机：随着AI功能的丰富，可能通过NAND存储弥补DRAM性能；3）游戏：新版Xbox已经开始使用2TB存储，次世代主机有望延续该趋势，单系统NAND有望持续增长。

图144：单系统NAND需求情况（单位：Gbytes/系统）



资料来源：IDC，国信证券经济研究所整理

图145：单系统NAND需求预测（单位：Gbytes/系统）



资料来源：IDC，国信证券经济研究所整理

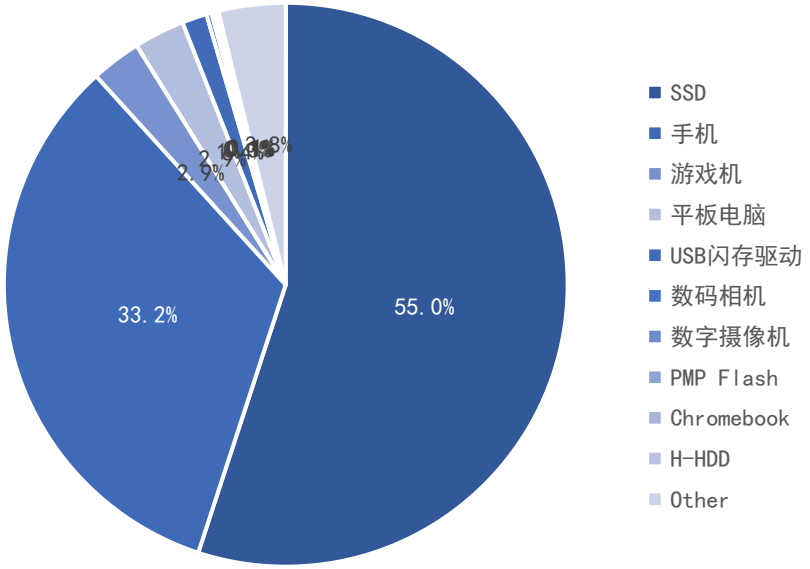
NAND下游需求情况（按Bit）：SSD需求占比持续提升

国信证券 GUOSEN SECURITIES

- **NAND下游需求情况：**根据IDC披露数据，2024年SSD、手机领域NAND需求占比分别为55.0%、33.2%，为NAND下游主要应用领域，合计占比为88.2%；其次为游戏机、平板电脑、USB闪存驱动、数码相机、数字摄像机、PMP Flash、Chromebook、H-HDD领域，占比分别为2.9%、2.9%、1.4%、0.3%、0.1%、0.1%、0.1%、0.1%。

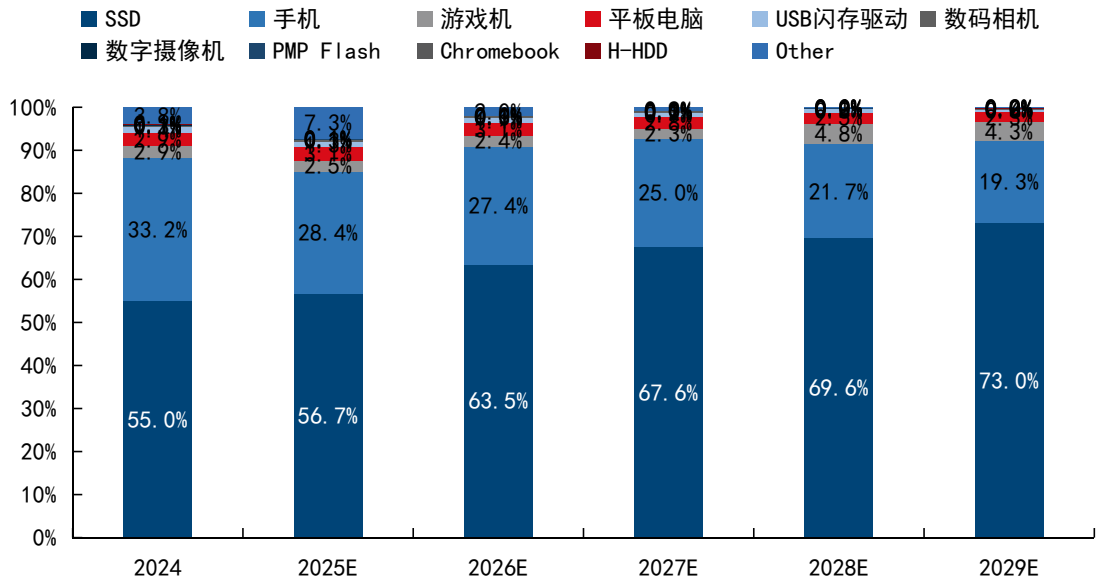
- **SSD需求占比持续提升：**一方面，PC用SSD的容量持续提升，拉动SSD用NAND增长；另一方面，随着AI推理爆发，SSD作为服务器的本地存储，需求强劲，拉动SSD出货容量增长。根据IDC披露数据，SSD需求占比从2024年55.0%将提升至2029年73.0%。

图146：2024年NAND下游需求情况（按Bit）



资料来源：IDC，国信证券经济研究所整理

图147：NAND下游需求及预测（按Bit）

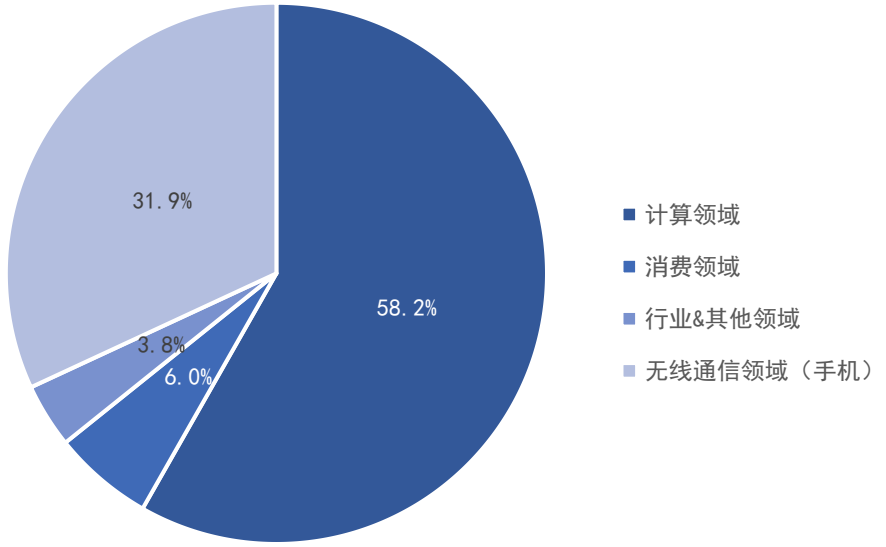


资料来源：IDC，国信证券经济研究所整理

NAND分领域需求情况：计算领域需求持续提升

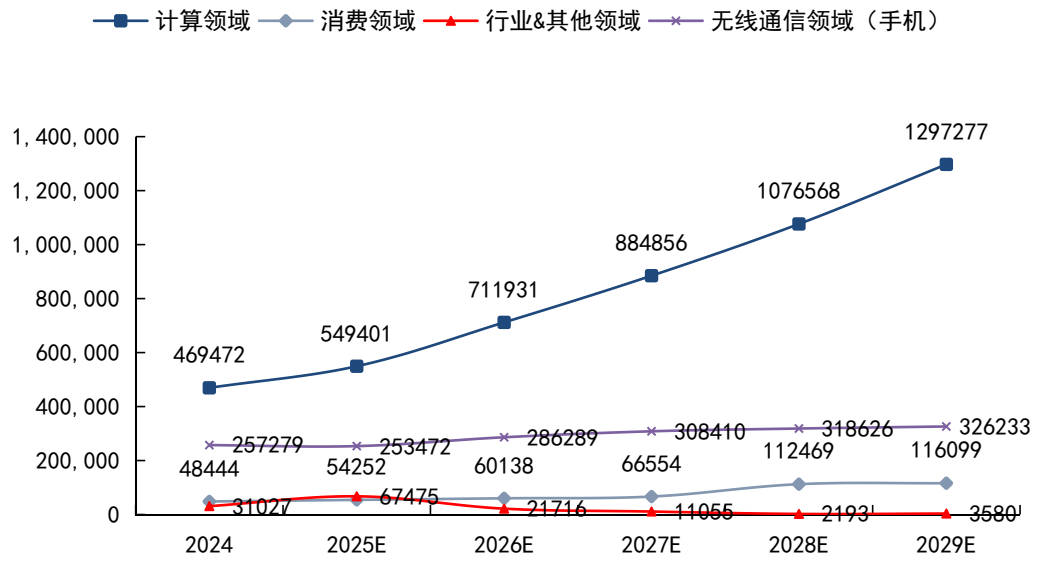
- **计算领域、无线通信领域NAND需求占比较高：**根据IDC披露数据，2024年计算领域、消费领域、行业&其他领域、无线通信领域NAND需求占比分别为58.2%、6.0%、3.8%、31.9%，计算、无线通信领域合计占比为90.1%；其中，计算领域包括存储（SSD、HDD、闪存、USB驱动）和个人计算（Chromebook），无线通信领域主要为手机，消费领域包括数码相机、游戏机、平板、PMP Flash。
- **计算领域需求快速提升：**受益于SSD需求快速增长，根据IDC预测数据，计算领域NAND需求快速增长，预计2029年计算领域NAND需求达到1,297,277 Gbytes in MU，对应24-29年CAGR为22.5%。

图148：2024年NAND分领域需求情况



资料来源：IDC，国信证券经济研究所整理

图149：计算领域NAND需求持续提升（单位：Gbytes in MU）



资料来源：IDC，国信证券经济研究所整理

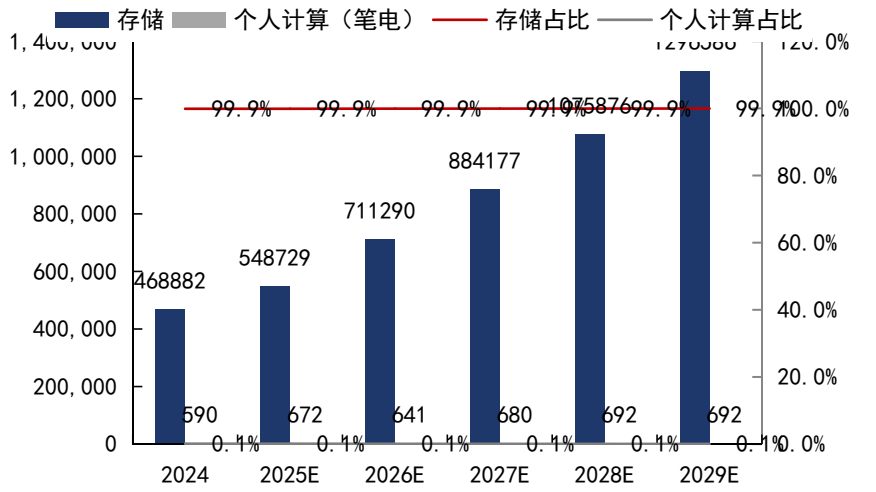
NAND——计算领域：SSD驱动NAND需求增长

国信证券 GUOSEN SECURITIES

■ **计算领域NAND需求主要分为：存储NAND需求和个人计算NAND需求。**根据IDC披露数据，2024年存储NAND需求、个人计算NAND需求分别为468,882、590 Gbytes in MU,占比分别为99.9%和0.1%。

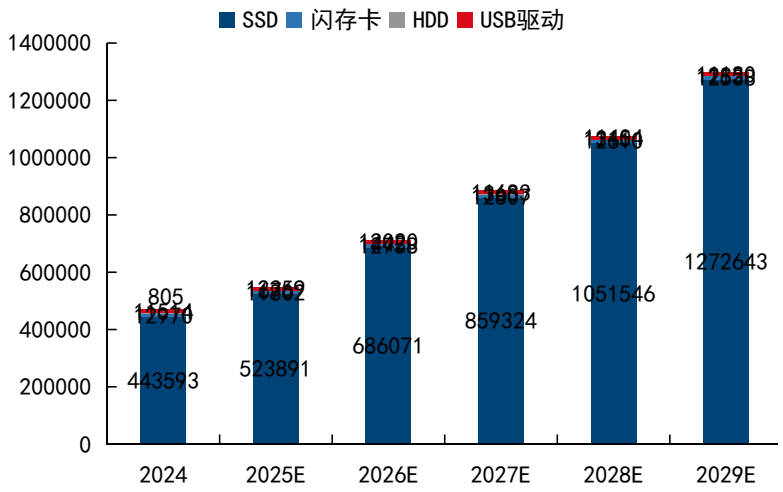
- **存储NAND需求：**存储NAND需求主要包括SSD、闪存卡、HDD、USB驱动，根据IDC披露数据，2024年出货量分别为443,593、12,970、805、11,514 Gbytes in MU，占比分别为94.6%、2.8%、0.2%、2.5%；未来，受PC和AI服务器驱动，SSD需求快速增长，占比持续提升，根据IDC预测数据，预计2028年提升至97.7%，相较于2024年提升3.1个pct；
- **个人计算NAND需求：**主要为Chromebook的NAND需求，需求基本维稳。

图150：存储和个人计算需求及占比（单位：Gbytes in MU）



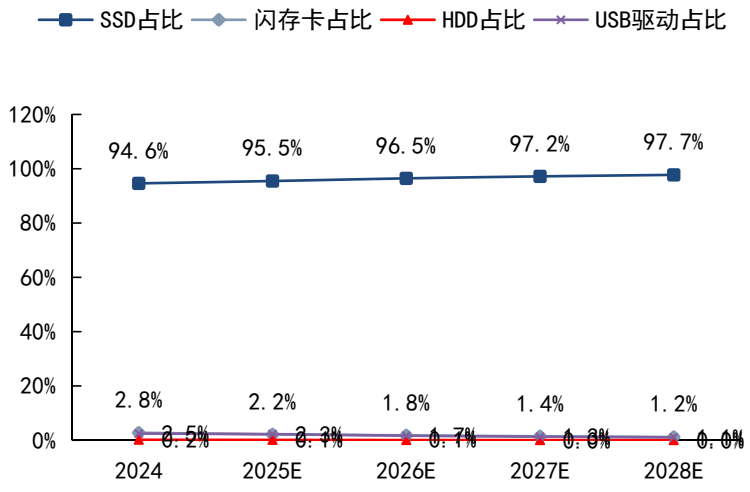
资料来源：IDC，国信证券经济研究所整理

图151：存储中各产品需求情况（单位：Gbytes in MU）



资料来源：IDC，国信证券经济研究所整理

图152：存储中各产品需求占比情况



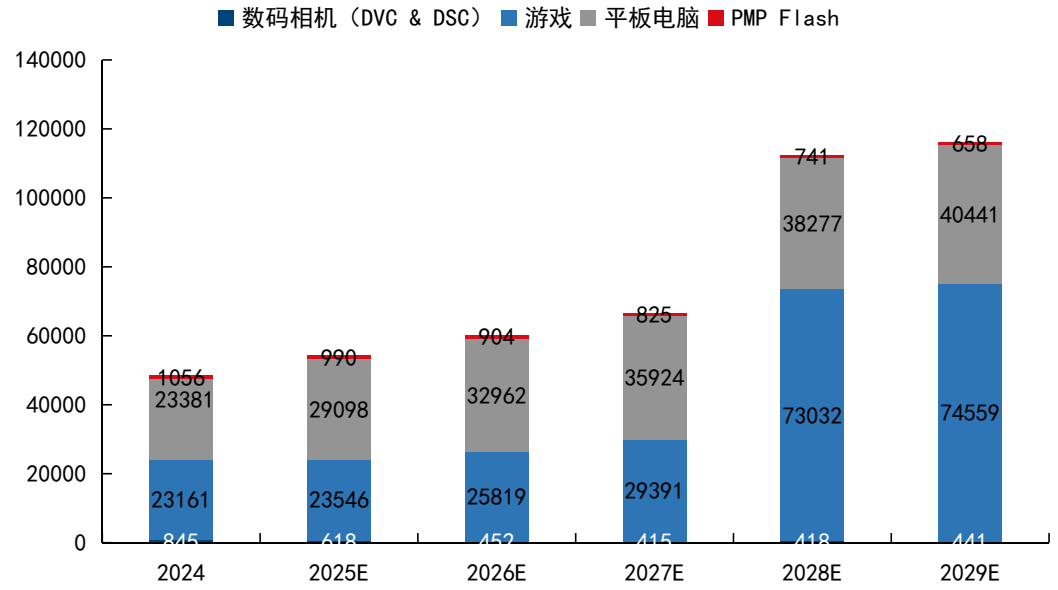
资料来源：IDC，国信证券经济研究所整理

NAND——消费领域：游戏机驱动NAND需求增长

国信证券 GUOSEN SECURITIES

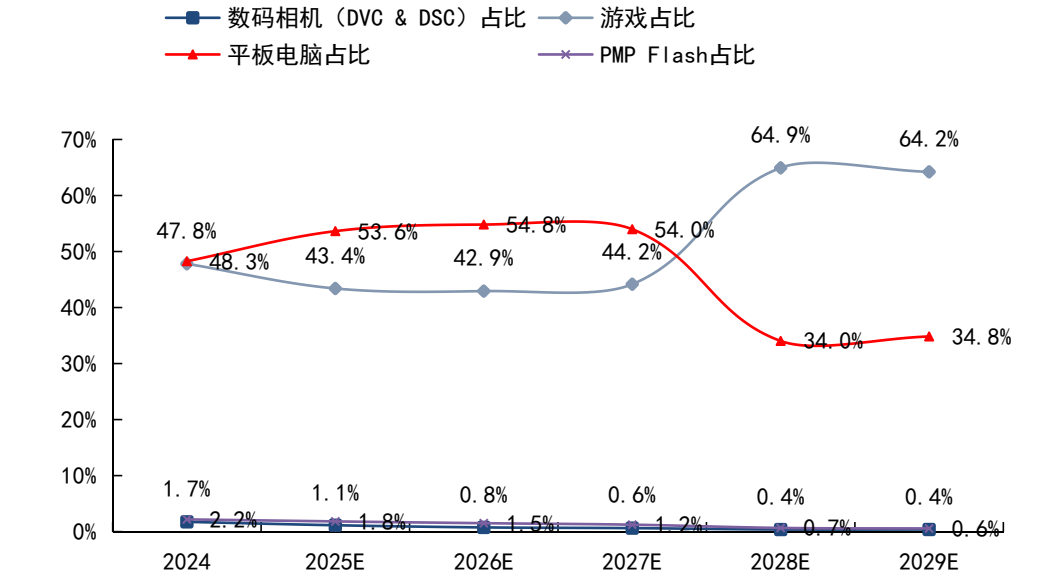
■ **消费领域NAND需求：**主要包括数码相机（DVC & DSC）、游戏、平板电脑、PMP Flash，根据IDC披露数据2024年NAND需求分别为845、23,161、23,381、1,056，占比分别为1.7%、47.8%、48.3%、2.2%，游戏和平板电脑需求占比较高；此外，随着Xbox游戏机逐步使用2TB存储，预计未来游戏机NAND需求占比将进一步提升，预计2029年提升至64.2%，相较于2024年+16.4个pct。

图153：消费领域各产品NAND需求情况（单位：Gbytes in MU）



资料来源：IDC，国信证券经济研究所整理

图154：消费领域各产品NAND需求占比情况



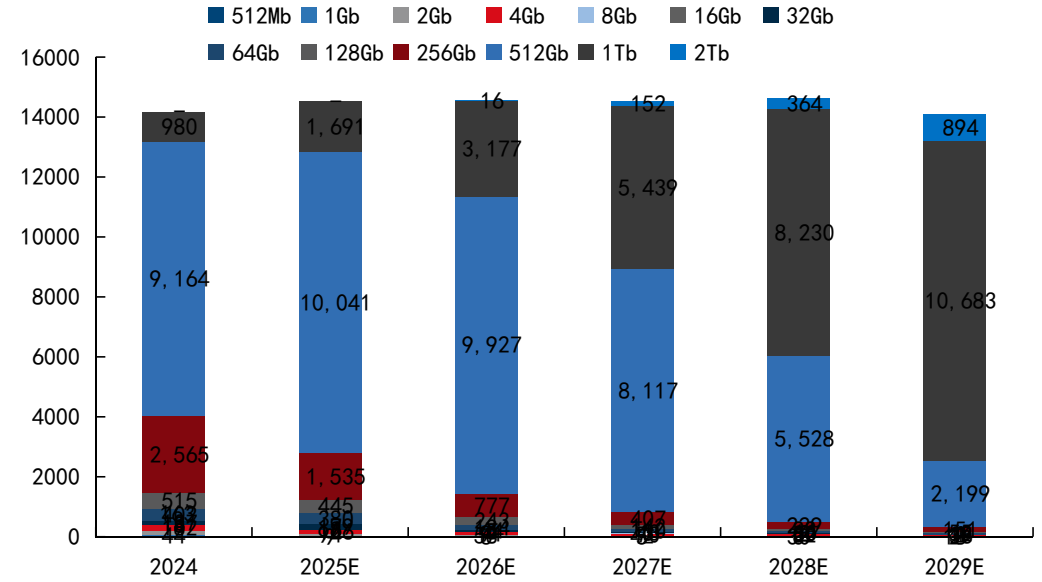
资料来源：IDC，国信证券经济研究所整理

NAND按容量出货情况：512GB为出货主力，1TB占比快速提升

国信证券 GUOSEN SECURITIES

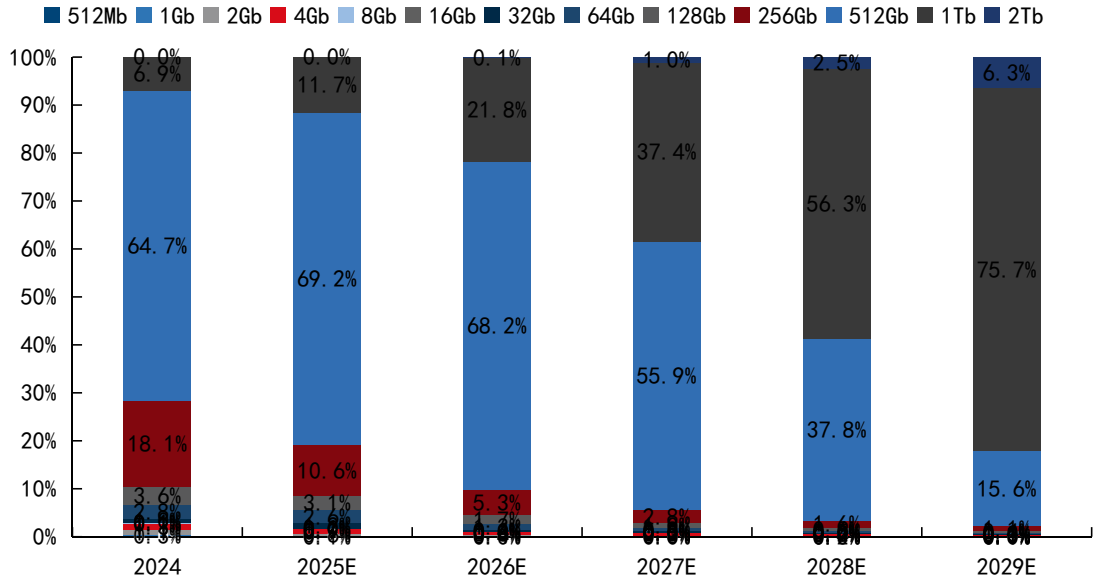
■ **各容量NAND出货量情况：**根据IDC披露数据，2024年256GB、512GB NAND为出货主力，分别为25.65、91.64亿颗，占比分别为82.8%，占据主要出货份额；未来，随着PC、AI服务器用SSD持续向高容量迭代，高容量NAND出货量有望持续提升，根据IDC预测数据，预计2029年1TB、2TB NAND出货量占比分别提升至75.7%、6.3%，相较于2024年分别+68.8、+6.3个pct。

图155：NAND按容量出货情况（单位：百万颗）



资料来源：IDC，国信证券经济研究所整理

图156：NAND不同容量产品出货占比



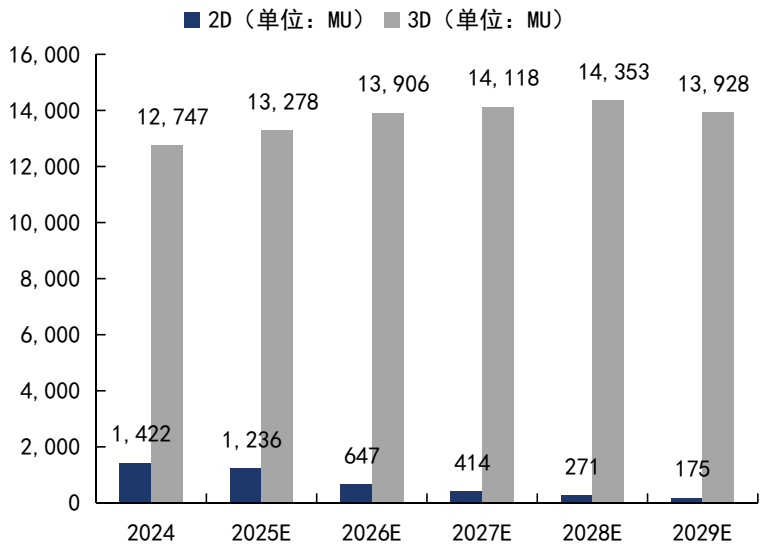
资料来源：IDC，国信证券经济研究所整理

NAND按架构出货情况：3D架构为出货主力，TLC占比最高

国信证券 GUOSEN SECURITIES

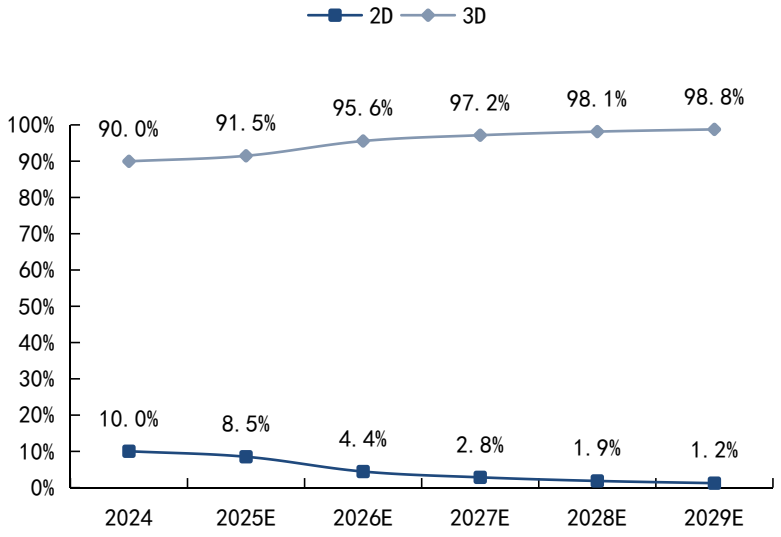
- **3D架构为出货主力，占比持续提升：**根据IDC披露数据，2024年2D、3D NAND出货量分别为1,422、12,747 MU，占比分别为10.0%、90.0%，预计3D架构NAND占比将持续提升，预计2029年提升至98.8%，相较于2024年+8.8个pct。
- **3D架构中，TLC为出货主力，QLC占比持续提升：**根据IDC披露数据，2024年3D架构中，TLC、QLC、MLC占比分别为95.9%、4.0%、0.1%，TLC为出货主力；随着QLC技术逐步成熟，预计QLC出货量占比将快速提升，根据IDC预测数据，2029年QLC出货量占比有望提升至42.5%，相比于2024年+38.5个pct。

图157：2D及3D NAND出货量情况（单位：MU）



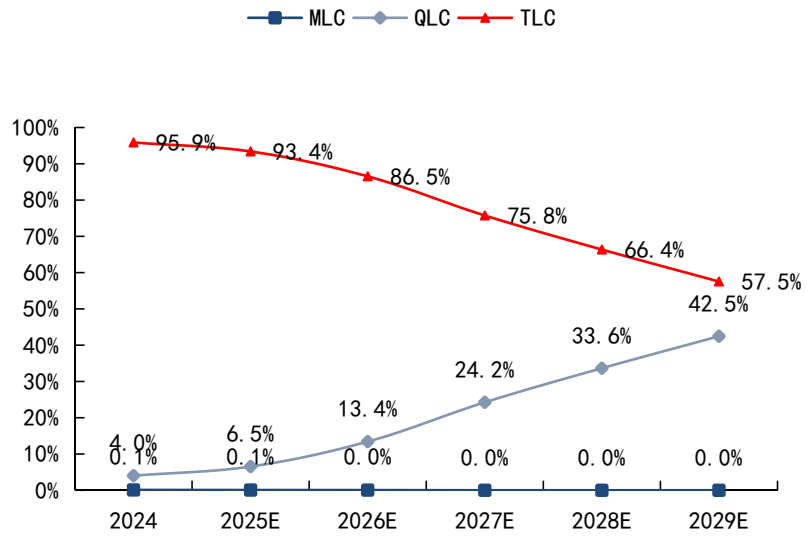
资料来源：IDC，国信证券经济研究所整理

图158：2D及3D NAND出货量占比



资料来源：IDC，国信证券经济研究所整理

图159：3D NAND中不同接口出货量占比



资料来源：IDC，国信证券经济研究所整理

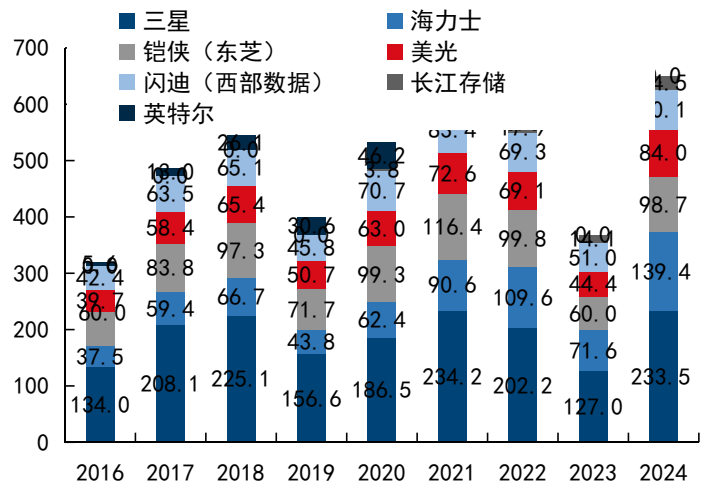
NAND供给侧：三星、海力士占比较高，铠侠、美光、闪迪为第二梯队

国信证券 GUOSEN SECURITIES

■ **NAND供给侧：三星、海力士占比较高，铠侠、美光、闪迪为第二梯队。**

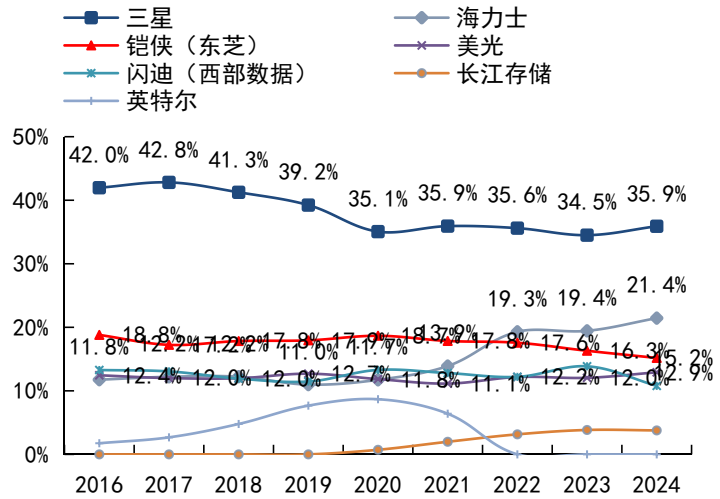
- **市占率情况：**根据IDC披露数据，25Q2全球NAND竞争格局，三星、海力士占比分别为34.0%、22.0%，合计占比为56.0%，为行业龙头；此外，铠侠、美光、闪迪占比分别为13.9%、13.8%、12.1%，为第二梯队，NAND市场CR5为95.8%，相对集中；
- **市占率变化情况：**1）整体趋势，三星NAND市场规模基本维稳，海力士市占率持续提升（2020年收购英特尔的NAND业务），铠侠、美光、闪迪市占率基本维稳；2）历史收并购情况，NAND行业历史收并购数目较多，2017年贝恩收购东芝存储部分股权，2019年东芝存储正式改名为铠侠；2016年西部数据收购闪迪，2023年西部数据开始分拆闪迪，2025年正式完成分拆，闪迪重新在纳斯达克上市；2020年，英特尔出售NAND业务给海力士。

图160：NAND各公司收入情况



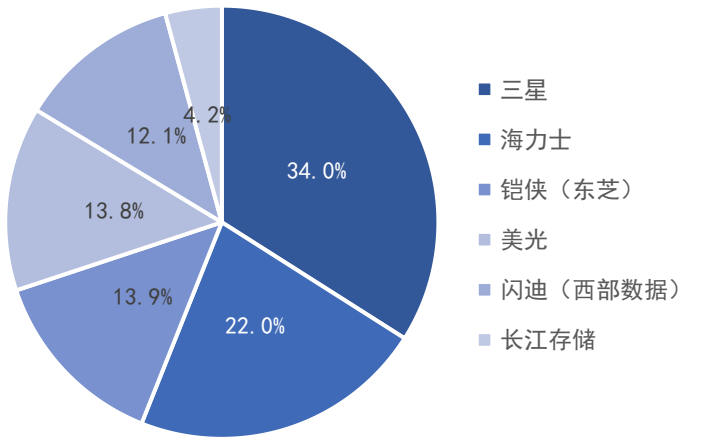
资料来源：IDC，国信证券经济研究所整理

图161：NAND各公司市占率变化



资料来源：IDC，国信证券经济研究所整理

图162：25Q2 NAND竞争格局



资料来源：IDC，国信证券经济研究所整理

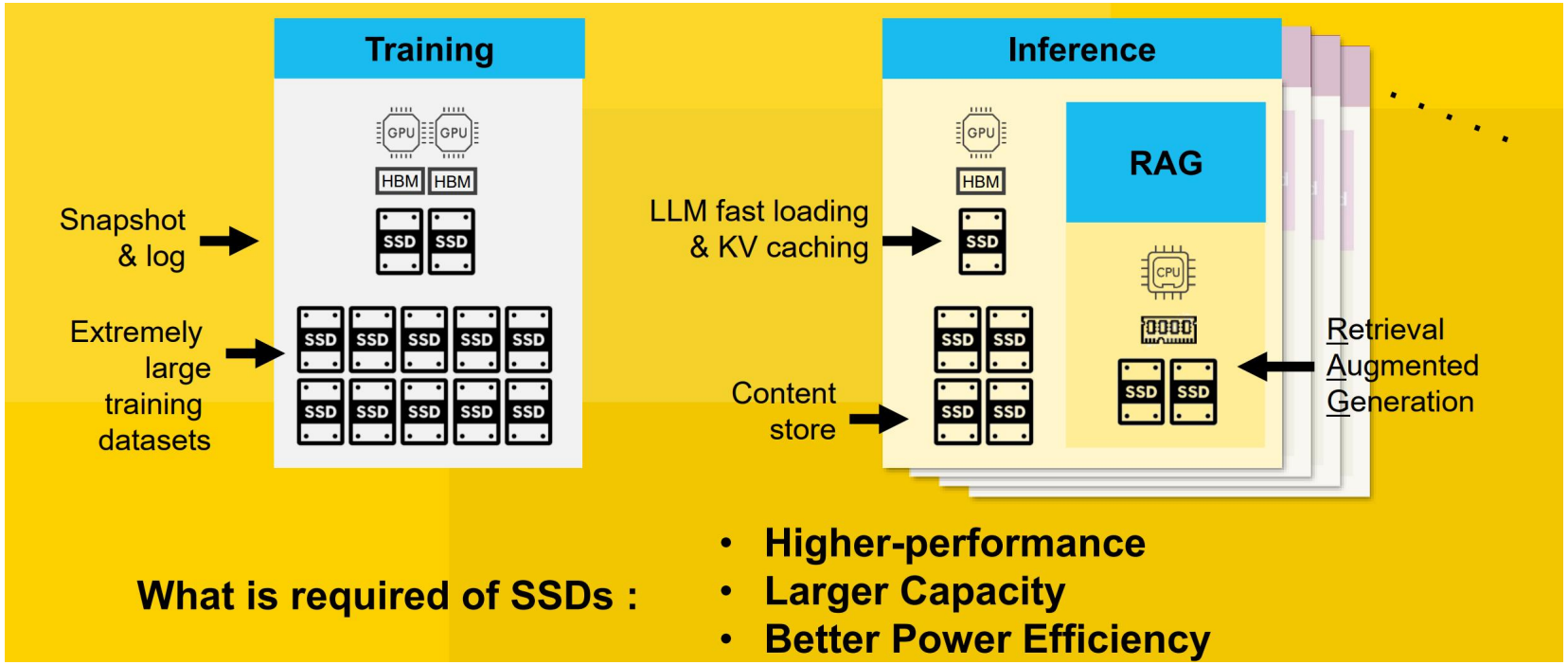
目录

大模型对存储的需求：训练&推理

- **大模型训练：**原始模型参数权重、训练数据等存储在网络存储（HDD/SSD），然后传输到本地SSD，进而读取到CPU内存（DRAM），然后加载到GPU显存（HBM）进行计算。

- **大模型推理：**通常从本地SSD读取模型权重文件到HBM/DRAM中，并处理输入Token在HBM/DRAM中生成初始的KV Cache（若启用了检索增强功能，通常向量数据库存储在NVMe SSD中），之后大模型开始逐token生成输出序列（每生成一个新的token，KV Cache随之增长，如果超过HBM/DRAM可用容量，则会offload到NVMe SSD中，便于后续推理再读回），最终模型响应和交互元数据会写入永久性存储，亦放入NVMe SSD中。

图163：大模型训练&推理对存储的需求



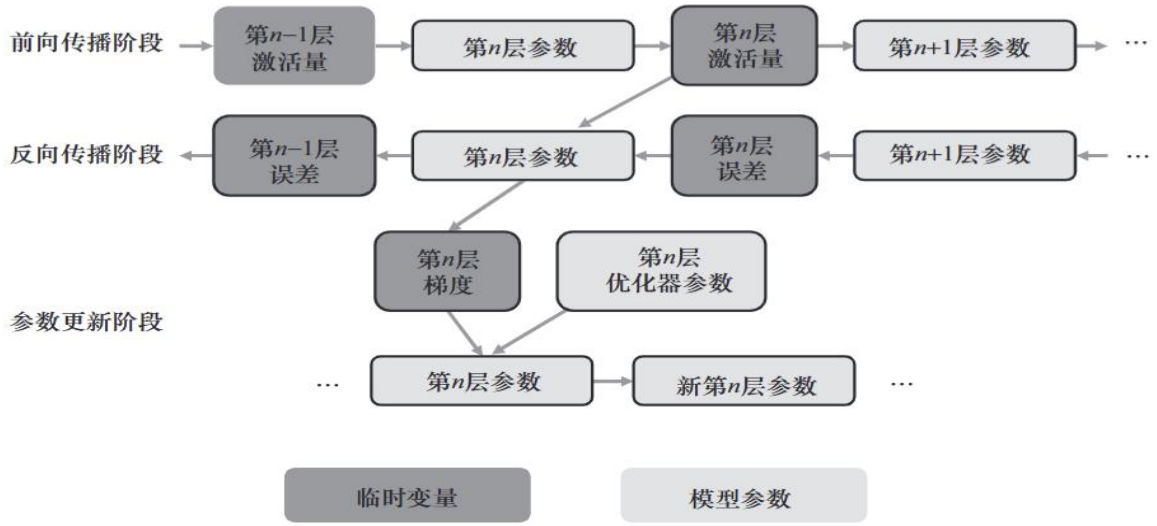
资料来源：铠侠，国信证券经济研究所整理

大模型对存储的需求：训练侧-对HBM的需求测算

国信证券 GUOSEN SECURITIES

- **大模型训练对HBM需求：**大模型训练主要包括前向传播、反向传播和参数更新三个阶段；1）**前向传播：**训练数据依次通过模型的各层进行计算，并产生激活量（不仅作为下一层的输入，还会被保存以供反向传播使用）；2）**反向传播：**前向传播结束后，计算结果与训练数据的标签比对，生成Loss（误差），Loss与各层参数以及保存在该层的激活量进行计算，进而产生每一层参数对应的梯度；（此处，每一层梯度计算完后，其激活量可以释放掉，而梯度数据需要被保留用于参数更新）；3）**参数更新：**优化器根据保存的梯度以及其自身的参数（如动量等）更新模型参数，参数更新完成后，上一版本的模型参数、梯度数据以及优化器参数被释放。
  - **模型参数：**以GPT-3为例，参数量为1750亿，考虑全精度（FP32），单个参数占用4个字节，则1750亿参数占用空间为175*10^9*4=700GB；
  - **梯度参数：**反向传播阶段，每个参数对应一个梯度值，则占用空间=模型参数占用空间=700GB；
  - **优化器状态参数：**通常采用Adam优化器，对每个参数存储动量（Momentum）和方差（Variance），则占用空间为700*2=1400GB；
  - **激活值：**激活值计算相对复杂，则根据模型训练设定不同，呈现出较大的变化；通常情况下激活值=Batch Size（单次请求批量大小）*Sequence Length（序列长度）*Hidden Size（模型隐藏层）*（34+5*Attention(头数）*Sequence Length（序列长度） / Hidden Size（模型隐藏层））*Layers（模型Transformer层数）*精度=1*2048*12288*（34+5*96*2048/12288）*96*（4/1024/1024/1024）=1026GB（此处假设Batch Size=1，随着Batch Size的增长，激活量所需存储空间将进一步提升）；
  - **加总：**700+700+1400+1026=3826GB。

图164：大模型训练中各阶段数据依赖



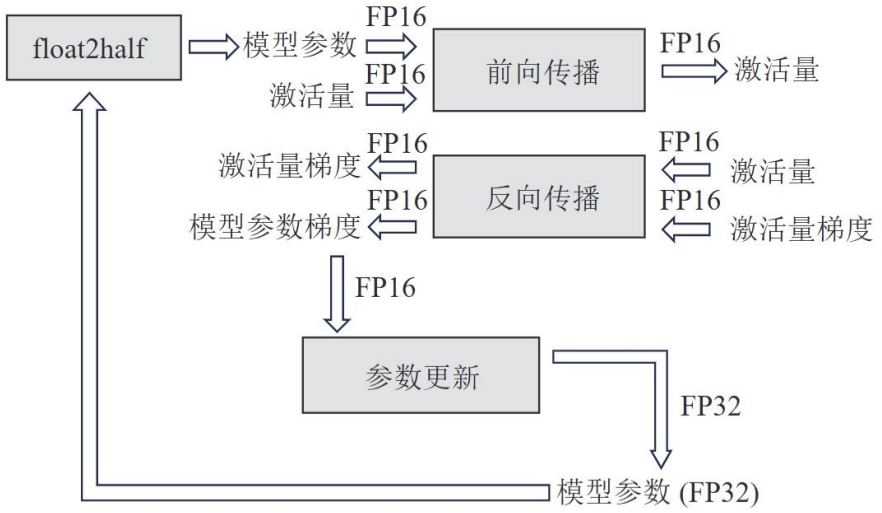
资料来源：冯杨洋等著-《大模型时代下的存储系统挑战与技术发展》-大数据（2025年）-P81，国信证券经济研究所整理

大模型对存储的需求：训练侧-对HBM的需求测算（混合精度训练）

国信证券 GUOSEN SECURITIES

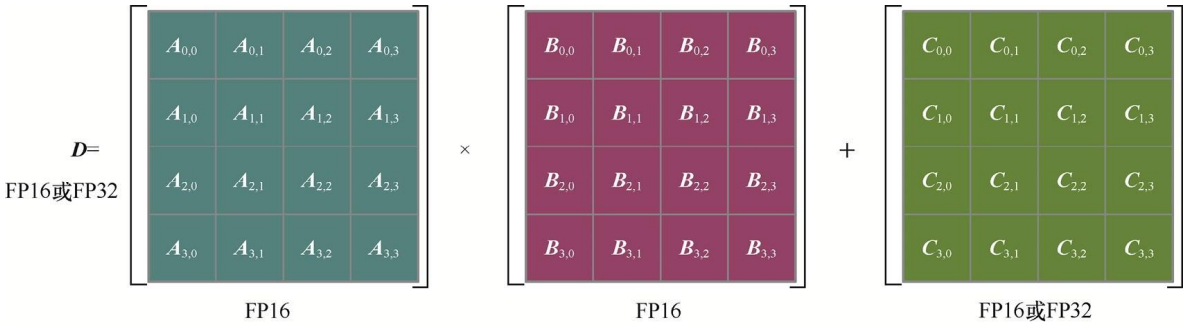
- **混合精度训练可降低大模型对HBM的需求：**模型参数、激活量、梯度均以半精度的格式存储（FP16），除此之外，还要保存一份全精度格式的模型参数。在模型训练过程中使用半精度的模型参数进行计算，生成半精度的梯度数据；在进行参数更新时，半精度的梯度数据与全精度版本的参数相加得到全精度的新模型参数，新的全精度模型参数在下一轮训练前会先转化为半精度版本。（精度累加：FP16进行矩阵乘法运算，FP32进行矩阵乘法中间的累加，可大幅降低训练的精度损失）
  - **模型参数**：以GPT-3为例，参数量为1750亿，考虑半精度（FP16），单个参数占用2个字节，则1750亿参数占用空间为175*10⁹*2=350GB；**备份全精度模型参数**，175*10⁹*4=700GB；
  - **梯度参数**：反向传播阶段，每个参数对应一个梯度值，则占用空间=模型参数占用空间=350GB；
  - **优化器状态参数**：通常采用Adam优化器，对每个参数存储动量（Momentum）和方差（Variance），则占用空间为350*2=700GB；
  - **激活值**：激活值计算相对复杂，则根据模型训练设定不同，呈现出较大的变化；通常情况下激活值=Batch Size（单次请求批量大小）*Sequence Length（序列长度）*Hidden Size（模型隐藏层）*（34+5*Attention(头数)*Sequence Length（序列长度） / Hidden Size（模型隐藏层））*Layers（模型Transformer层数）*精度=1*2048*12288*（34+5*96*2048/12288）*96*（2/1024/1024/1024）=513GB（此处假设Batch Size=1，随着Batch Size的增长，激活量所需存储空间将进一步提升）；
  - **加总**：350+700+350+700+513=2513GB，采用混合精度训练比全精度显存需求下降。

图165：大模型混合精度训练示意图



资料来源：冯杨洋等著-《从 BERT 到 ChatGPT：大模型训练中的存储系统挑战与技术发展》-计算机研究与发展（2024年）-P819，国信证券经济研究所整理

图166：精度累加操作



资料来源：文亮等著-《揭秘大模型：从原理到实战》-人民邮电出版社（2025年）-P324，国信证券经济研究所整理

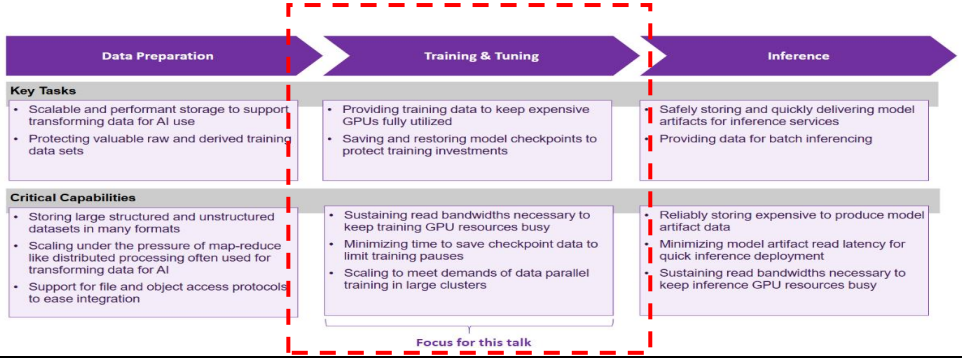
大模型对存储的需求：训练侧-对SSD的需求测算

国信证券 GUOSEN SECURITIES

■ **大模型训练对SSD需求：**大模型训练阶段，本地SSD主要用于存储训练数据和Checkpoint（检查点）。

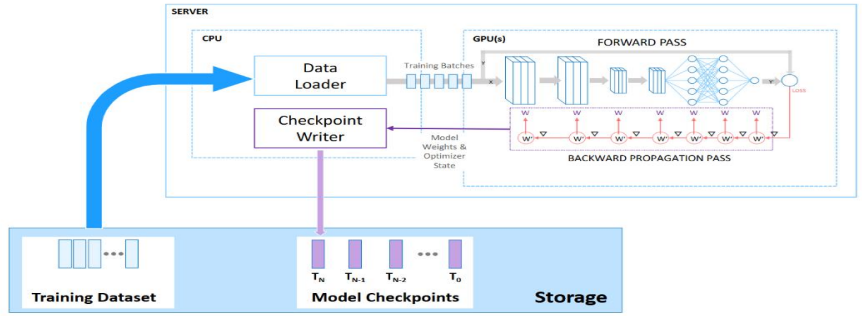
- **训练数据**：预处理好的训练放到SSD存储，随后根据需求加载到CPU的DRAM中，转换为Batches交由GPU进行计算；根据OpenAI在《Language Models are Few-Shot Learners》中披露，原始数据为45TB，进行清洗和去重后，保留了570GB（对应300 billion tokens）数据用于模型训练；
- **Checkpoint（检查点）**：Checkpoint是指训练过程中某个特定时间点保存的模型的快照，其包括模型参数、优化器状态、训练元数据等信息，如果大模型训练中止，可以从上次保存的Checkpoint中加载权重进而恢复训练。根据Dell披露数据，1750亿参数模型，如果允许在2小时训练中允许5%的时间（360秒）用于生成Checkpoint，对应存储写带宽为6.8GB/s，对应的Checkpoint存储空间为6.8*360=2448GB。通常采用检查点管理策略，例如保留最新的5-10个检查点，旧的检查点进行释放。

图167：存储在AI全生命周期发挥重要作用



资料来源：DELL，国信证券经济研究所整理

图168：大模型训练过程中产生Checkpoint



资料来源：DELL，国信证券经济研究所整理

图169：GPT-3的训练数据量

Model Name	n_{params}	n_{layers}	d_{model}	n_{heads}	d_{head}	Batch Size	Learning Rate
GPT-3 Small	125M	12	768	12	64	0.5M	6.0×10^{-4}
GPT-3 Medium	350M	24	1024	16	64	0.5M	3.0×10^{-4}
GPT-3 Large	760M	24	1536	16	96	0.5M	2.5×10^{-4}
GPT-3 XL	1.3B	24	2048	24	128	1M	2.0×10^{-4}
GPT-3 2.7B	2.7B	32	2560	32	80	1M	1.6×10^{-4}
GPT-3 6.7B	6.7B	32	4096	32	128	2M	1.2×10^{-4}
GPT-3 13B	13.0B	40	5140	40	128	2M	1.0×10^{-4}
GPT-3 175B or "GPT-3"	175.0B	96	12288	96	128	3.2M	0.6×10^{-4}

Table 2.1: Sizes, architectures, and learning hyper-parameters (batch size in tokens and learning rate) of the models which we trained. All models were trained for a total of 300 billion tokens.

资料来源：OpenAI著-《Language Models are Few-Shot Learners》-arXiv（2020）-P8，国信证券经济研究所整理

图170：GPT-3的Checkpoint所需时间

Model Parameters (B)	Checkpoint Size (GB)	Total Read BW (GBps) Needed to Restore Checkpoint within 5 Minutes # Model Instances (Data Parallelism) 1	8	16	32	64	128
3	42	0.002	0.02	0.04	0.07	0.15	0.30
7	98	0.01	0.04	0.09	0.17	0.35	0.70
13	182	0.01	0.08	0.16	0.32	0.65	1.29
33	462	0.03	0.21	0.41	0.82	1.64	3.29
70	980	0.05	0.44	0.87	1.74	3.48	6.97
140	1960	0.11	0.87	1.74	3.48	6.97	13.94
175	2450	0.14	1.09	2.18	4.36	8.71	17.42
530	7420	0.41	3.30	6.60	13.19	26.38	52.76

资料来源：DELL，国信证券经济研究所整理

大模型对存储的需求：推理侧-对HBM的需求测算

国信证券 GUOSEN SECURITIES

■ **大模型推理对HBM的需求主要包括两部分，承载模型权重和KV Cache，其中，在MOE架构中（当前主流大模型架构），HBM仅存放稠密部分（例如路由器等）和活跃专家模型。**

- **用户数及需处理的Token数量：**我们以GPT-5为例，根据25年10月OpenAI的DevDay开发者大会披露数据，目前ChatGPT的周活用户已经达到8亿人，我们假设（WAU/MAU）=70%，则ChatGPT的月活用户约11.4亿人，假设单MAU每日消耗2万个token，则全部MAU每月消耗Token的总量为685.7万亿；同时，OpenAI披露其API约60亿/分钟调用，则每月OpenAI API处理token量为259.2万亿；加总，ChatGPT每月Token处理总量为944.9万亿，对应每秒处理3.65亿token。此外，根据RiteshAI披露数据，输入token和输出token的比例大约为7.5:1，则我们假设输出token数量/（输入token数量+输出token数量）=17%，对token总量进行了进一步划分，对每秒钟输出token的需求为6067万token/s；
- **大模型权重对应HBM需求：**目前主流大模型为MOE架构（多专家混合模型），我们假设专家模型参数占比为90%，稠密部分占比为10%，目前HBM主要承载大模型稠密部分（包括路由器等）和专家模型中激活的部分。1）稠密部分：假设GPT-5模型参数量为2万亿，则稠密部分参数量为0.2万亿，FP16下单参数占2bit空间，对应0.40TB显存需求；2）激活专家部分：专家模型参数量为1.8万亿，假设单次激活专家参数量占比为3%，则单次激活参数量为0.054万亿，对应0.10TB显存需求。综上，运行单个模型推理需要0.4+0.10=0.50TB显存。假设单个模型副本每秒可以输出1000个token，而GPT-5的用户需求为6067万token/s，则需求副本数量为60758个，对应HBM总需求为60758*0.50TB=29.4PB。

图171：大模型输入Token、输出Token比例关系

Cost of AI: Inference

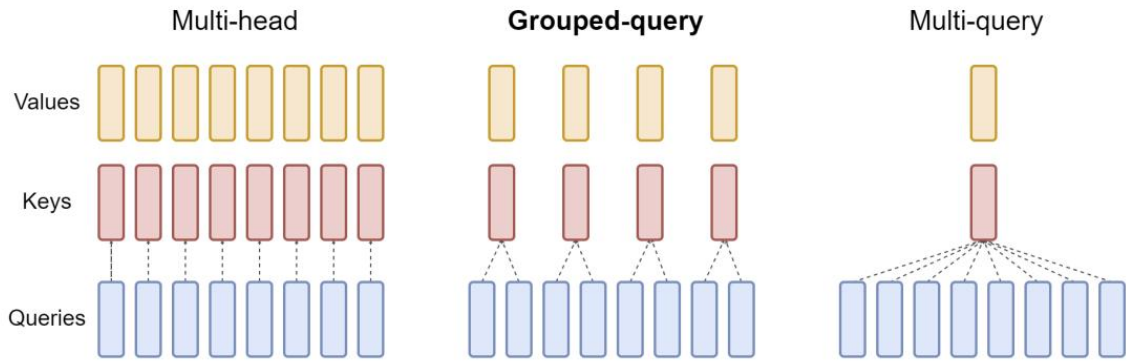
Estimated cost to do AI summarization

Provider	input/output price per 1M tokens	input tokens	7,449,600,000	7.5 billion
		output tokens	1,489,920,000	1.5 billion
		input price	output price	total price
OpenAI gpt-4-turbo-2024-04-09	$10/$30	$ 74,496	$ 44,698	$ 119,194
Claude Opus (Bedrock)	$15/$75	$ 111,744	$ 111,744	$ 223,488
Claude Sonnet (Bedrock)	$3/$15	$ 22,349	$ 22,349	$ 44,698
Llama 3 instruct 70B (Bedrock)	$2.65/$3.50	$ 19,741	$ 5,215	$ 24,956
Llama 3 instruct 70B (Groq)	$0.59/$0.79	$ 4,395	$ 1,177	$ 5,572
Mixtral 8x7B (Bedrock)	$0.45/$0.70	$ 3,352	$ 1,043	$ 4,395
Mixtral 8x7B SMoE (Groq)	$0.27/0.27	$ 2,011	$ 402	$ 2,414

RiteshAI.com

资料来源：RiteshAI，国信证券经济研究所整理

图172：KV Cache处理方法（Multi-head、Grouped-query、Multi-query）



资料来源：Joshua等著-《GQA: Training Generalized Multi-Query Transformer Models from Multi-Head Checkpoints》-arXiv（2023年）-P2，国信证券经济研究所整理

大模型对存储的需求：推理侧-对HBM的需求测算

国信证券 GUOSEN SECURITIES

- **KV Cache对HBM的需求：1）Multi-Head方法：**非主流方法，对HBM需求较大，其KV size=2*Hidden*模型层数*精度=7680KB/Token，假设单位用户对Token输出速率要求为20token/s，则所需并发数为304万个，假设平均输出长度为200token，输入长度为1000token，则单次并发KV Cache对显存的需求为8.79GB，面对304万个并发，对应显存总需求为25.5TB；**2）GQA方法：**目前主流方法，简单来说就是多个Query共用一个KV，假设模型注意力头数为128头，分组数常为注意力头数的公约数（此处可取8、16），此处分为8组，则其KV Size为480 KB/Token，对应显存总需求为1.6PB。如果采用GQA方法，加总模型权重和KV Cache，则对HBM总需求为31.0PB。

图173：大模型推理对HBM的需求测算

用户数及需处理的Token量		激活模型参数所需HBM空间		KV Cache所需HBM空间			
				Multi-Head方法		GQA方法（主流方法）	
ChatGPT 周活（亿人，25年10月）	8	GPT-5模型参数量（万亿）	2	模型的注意力头数（个）	128		
ChatGPT 周活/月活	70%	稠密部分（路由等）占比	10%	模型层数（层）	120		
ChatGPT 月活（亿人）	11.4	稠密部分参数量（万亿）	0.2	模型的Hidden Size	16384	分组数（组）	8
MAU每日Token消耗量（万）	2	专家模型占比	90%	参数精度（Bit，假设FP16）	2	对应NV注意力头数	8
MAU每月Token消耗量（万）	60	专家模型参数量（万亿）	1.8	KV Size（KB/Token）	7680	KV Size（KB/Token）	480
月活用户每月Token处理量（万亿）	685.7	单次激活专家数/专家总数（MOE架构）	3%	注： 1、Multi-Head方法：KV Size=2*Hidden Size*模型层数*精度 2、GQA方法：KV Size=2*KV注意力头数*Head Size*模型层数*精度=2*KV注意力头数*模型的Hidden Size/模型的注意力头数*精度			
		单次激活专家参数量（万亿）	0.05				
API Token数 （含峰值缓冲，亿/分钟，25年10月）	60	GPT-5模型稠密部分+单次激活专家参数量（万亿）	0.25				
每月API Token数（万亿/月）	259.2			单位用户要求Token输出速率（Token/秒）	20		
		单个参数占用空间（Bit，假设FP16）	2	所需并发数（万个）	304		
ChatGPT每月Token处理量（加总，万亿）	944.9	单模型推理所需HBM空间（TB）	0.50				
ChatGPT平均每秒Token处理量（亿Token/秒）	3.65			假设：平均输出长度（Token）	200		
		假设单一模型副本每秒输出Token量（个）	1000	平均输入长度（Token）	1000		
输出Token/（输入Token+输出Token）	17%	满足输出需求所许副本数量（个）	60758				
输出Token（万Token/秒）	6076			每个并发的Token数量（Token）	1200		
		满足MAU所需HBM空间（PB）——存储模型权重	29.4	总的并发token数量（亿Token）	36.5		
				满足MAU所需HBM空间（PB）——存储KV Cache（Multi-Head方法）	25.5	满足MAU所需HBM空间（PB）——存储KV Cache（GQA方法）	1.6
				加总：推理阶段对HBM需求（PB）	54.9	加总：推理阶段对HBM需求（PB）	31.0

资料来源：OpenAI，国信证券经济研究所测算

大模型对存储的需求：推理侧-对DRAM的需求测算

国信证券 GUOSEN SECURITIES

■ **大模型推理对DRAM的需求主要包括两部分，承载非活跃专家模型权重和KV Cache。**

- **大模型权重对应DRAM需求：**我们假设专家模型参数占比为90%，专家模型参数量为1.8万亿，假设单次激活专家参数量占比为3%，非激活激活专家参数量占比为97%，对应1.74万亿参数，对应3.4TB DRAM需求，假设单个模型副本每秒可以输出1000个token，而GPT-5的用户需求为6067万token/s，则需求副本数量为60758个，对应HBM总需求为60758*3.4TB=0.2EB。
- **KV Cache对应DRAM需求：**假设采用GQA方法，如上文所述KV Size为480KB/token，考虑KV Cache命中率，由于用户工作规律，大部分重用发生在24小时（一天）内，则每日Token处理量为944.9/30=31.5万亿，对应KV Cache空间为13.4EB，此处假设KV Cache命中率为50%,即需要留存在DRAM里的KV Cache为6.7EB，加总大模型推理对DRAM需求为6.9EB。

图174：大模型推理对HBM的需求测算

用户数及需处理的Token量		激活模型参数所需DRAM空间		KV Cache所需DRAM空间			
				Multi-Head方法		GQA方法（主流方法）	
ChatGPT 周活（亿人，25年10月）	8	GPT-5模型参数量（万亿）	2	模型的注意力头数（个）	128		
ChatGPT 周活/月活	70%	专家模型占比	90%	模型层数（层）	120		
ChatGPT 月活（亿人）	11.4	专家模型参数量（万亿）	1.8	模型的Hidden Size	16384	分组数（组）	8
MAU每日Token消耗量（万）	2	单次激活专家数/专家总数（MOE架构）	3%	参数精度（Bit，假设FP16）	2	对应NV注意力头数	8
MAU每月Token消耗量（万）	60	单次激活专家参数量（万亿）	0.05	KV Size（KB/Token）	7680	KV Size（KB/Token）	480
月活用户每月Token处理量（万亿）	685.7	非激活专家参数量	1.74	注： 1、Multi-Head方法：KV Size=2*Hidden Size*模型层数*精度 2、GQA方法：KV Size=2*KV注意力头数*Head Size*模型层数*精度=2*KV注意力头数*模型的Hidden Size/模型的注意力头数*精度			
API Token数（含峰值缓冲，亿/分钟，25年10月）	60						
每月API Token数（万亿/月）	259.2			此处考虑KV Cache命中率，考虑到用户对上下文的重用基本上约束在一天（24小时）内			
		单个参数占用空间（Bit，假设FP16）	2				
ChatGPT每月Token处理量（加总，万亿）	944.9	单模型推理所需DRAM空间（TB）	3.4	每日Token处理量（万亿）	31.5	每日Token处理量（万亿）	31.5
ChatGPT平均每秒Token处理量（亿Token/秒）	3.65			对应KV Cache的DRAM空间需求（EB）	214.8	对应KV Cache的DRAM空间需求（EB）	13.4
		假设单一模型副本每秒输出Token量（个）	1000				
输出Token/（输入Token+输出Token）	17%	满足输出需求所许副本数量（个）	60758	假设：每日（24小时）KV Cache命中率	50%	假设：每日（24小时）KV Cache命中率	50%
输出Token（万Token/秒）	6076						
		满足MAU所需DRAM空间（EB）——存储模型权重	0.2	满足MAU的KV Cache所需DRAM空间（EB）——（Multi-Head方法）	107.4	满足MAU的KV Cache所需DRAM空间（EB）——（GQA方法）	6.7
				加总：推理阶段对DRAM需求（EB）	107.6	加总：推理阶段对DRAM需求（EB）	6.9

资料来源：OpenAI，国信证券经济研究所测算

大模型对存储的需求：推理侧-对本地NAND的需求测算

国信证券 GUOSEN SECURITIES

■ 大模型推理对本地NAND的需求主要为RAG存储。

- **测算RGA NAND需求/原始文本数据的比例：**我们原始文本数据为100GB，以512个Token作为一个切块，则可转化为5452.6万条向量，此处选择768维的向量维度（主流选择）、FP32量化精度，则对应向量化数据空间为156GB；此处采用HNSW索引，通常情况下，HNSW索引空间占向量化数据空间的40%，则对应索引存储需求为62.4GB，加总为218.4GB；此外，考虑道地理复制倍数（实现灾难恢复，通常为2-3x）、纠删码预留空间（防止机架或节点级别的故障，对于冷数据/超大规模向量库，推荐数据块数：校验块数为12：2，即（12+2）/12=1.17x）、系统开销预留空间（文件系统、元数据等功能占用的额外空间，通常1.1x）、扩容缓冲预留（通常1.3x），则RAG NAND总需求为1093.4GB，则RAG NAND总需求/原始文本数据比例关系为10.93：1；
- **总NAND需求测算：**根据前文所述，ChatGPT月活为11.4亿人，假设单用户每天上传1GB数据，则每年上传数据总量达到12.8EB，根据第一步求得的比例关系，可得RAG NAND总需求为139.7EB；此外，假设RAG存储占总NAND存储比例为70%，则总NAND需求为199.5EB。

图175：大模型推理对本地NAND的需求测算

RAG存储需求测算				总NAND需求测算	
假设：原始文本数据（GB）	100.0	地理复制倍数	3	ChatGPT 周活（亿人，25年10月）	8
对应字符数（英文UTF-8，亿字符）	1073.7	对应RAG NAND需求量（GB，考虑地理复制）	655.2	ChatGPT 周活/月活	70%
平均每个英文单词包含的英文字母数量（个）	5			ChatGPT 月活（亿人）	11.4
对应英文单词数量（亿个）	214.7	纠删码预留空间占比	16.7%		
Token/单词	1.3	对应RAG NAND存储倍数	1.17	假设：单用户每月上传数据量规模（GB）	1
对应Token数量（亿个）	279.2				
切块：单向量包含Token数量（个）	512	系统开销预留空间占比	10%	单用户每月上传原始文本数据总量（EB）	1.1
对应向量数量（万条）	5452.6	对应RAG NAND存储倍数	1.1	单用户全年上传原始文本数据总量（EB）	12.8
向量维度（以通用BERT、Sentence-BERT为例）	768	扩容缓冲预留占比	30%	对应RAG NAND总需求为（EB）	139.7
量化精度（以FP32为例）	4	对应RAG NAND存储倍数	1.3		
				假设：RAG NAND/总NAND	70%
对应向量化数据空间（GB）	156				
		加总：RAG NAND总需求（GB）	1093.4		
索引存储占比（此处使用HNSW）	40%	RAG NAND总需求/原始文本数据比例	10.93	NAND总需求（EB）	199.5
索引存储数据空间（GB）	62.4				
向量化数据空间+索引存储对应NAND需求（GB）	218.4				

资料来源：OpenAI，国信证券经济研究所测算

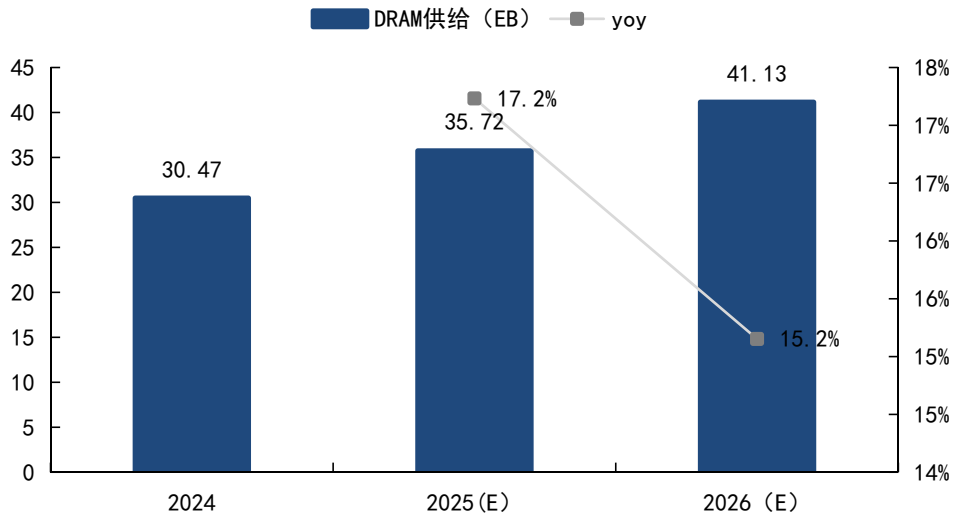
大模型对存储的供求关系：供给侧低双位数增长

国信证券 GUOSEN SECURITIES

■ **全球存储(DRAM和NAND）供给侧呈现低双位数增长，基本维持15%-20%同比增速。**

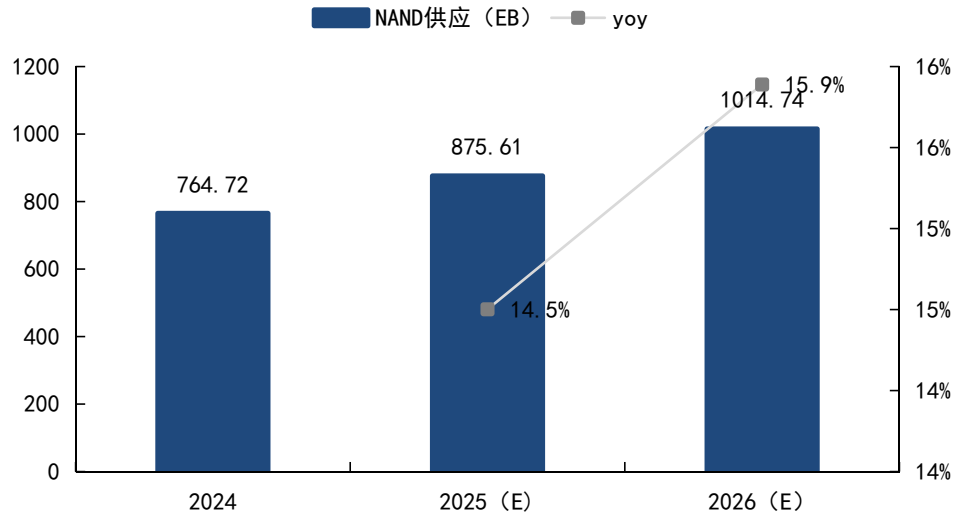
- **全球DRAM供给侧情况：**根据IDC披露数据，2024年全球DRAM供给量为30.47EB，预计2025、2026年分别增长至35.72EB、41.13EB，对应同比增速分别为17.2%、15.2%；
- **全球NAND供给侧情况：**根据IDC披露数据，2024年全球NAND供应量为764.72EB，预计2025、2026年分别增长至875.61、1014.74EB，对应同比增速分别为14.5%、15.9%。

图176：全球DRAM供给侧情况及预测（单位：EB）



资料来源：IDC，国信证券经济研究所整理

图177：全球NAND供给侧情况及预测（单位：EB）



资料来源：IDC，国信证券经济研究所整理

大模型对存储的供求关系：短期供不应求，驱动价格持续增长（DRAM）

国信证券 GUOSEN SECURITIES

■ **DRAM短期供不应求，驱动价格持续增长。**

- **2026年AI对DRAM总需求预测：**根据前文所述，预计2026年ChatGPT日活人数同比+30%；随着大模型向全模态方向快速发展，单位用户每月消耗的token数量同比+70%，则ChatGPT每月处理的Token量为1515.4万亿Token，此外假设API Token量同比+100%，对应518.4万亿token/月，则ChatGPT对DRAM（包括HBM）需求合计为14.9EB。同时，假设ChatGPT占全球AI DRAM需求的65%，则全球AI对DRAM总需求为23EB；
- **DRAM短期供不应求：**我们仅考虑AI对DRAM需求的拉动，根据前文测算2025、2026年全球DRAM总需求将达到40.1、53.2EB，而全球DRAM供给为35.7、41.1EB，2025、2026年全球DRAM总需求/总供给分别为112%、129%，短期供不应求，有望驱动价格持续增长。

图178：全球DRAM供求情况预测

	2024	2025(E)	2026(E)
ChatGPT 月活（亿人）		11.4	14.9
			30%
单位MAU每月Token消耗量（万）		60	102.0
			70%
ChatGPT每月Token处理量（万亿）		685.7	1515.4
			121%
API Token数（含峰值缓冲，亿/分钟）		60	120
			100%
API Token数（万亿/月）		259.2	518.4
ChatGPT每月Token处理量（加总，万亿）		944.9	2033.8
ChatGPT平均每秒Token处理量（亿Token/秒）		3.6	7.8
输出Token/（输入Token+输出Token）		17%	17%
输出Token（万Token/秒）		6076	13078
推理阶段对HBM需求（PB）		31.0	66.8
推理阶段对DRAM需求（EB）		6.9	14.9
加总：ChatGPT对DRAM需求（EB）		6.9	14.9
假设：ChatGPT占AI对DRAM总需求的比例		70%	65%
AI对DRAM总需求（EB）		9.9	23.0
DRAM总需求（EB，25、26年仅考虑AI拉动）	30.2	40.1	53.2
DRAM总供给（EB）	30.5	35.7	41.1
DRAM总需求/总供给	99%	112%	129%

资料来源：IDC，国信证券经济研究所测算

大模型对存储的供求关系：短期供不应求，驱动价格持续增长（NAND）

国信证券 GUOSEN SECURITIES

■ **NAND短期供不应求，驱动价格持续增长。**

- **2026年AI对NAND总需求预测：**根据前文所述，预计2026年ChatGPT日活人数同比+30%；随着大模型向全模态方向快速发展，单位用户上传数据量同比+70%，则ChatGPT用户全年上传原始文本数据总量为28.2EB，引用前文计算的换算比例，则得ChatGPT对应的RAG NAND总需求为308.6EB，随着RAG NAND需求快速增长，预计RAG NAND/总NAND提升至80%，则ChatGPT NAND总需求为385.8EB；此外，随着谷歌Gemini、Claude等大模型发力，假设ChatGPT占AI对NAND总需求比例下降至65%，则2026年AI对NAND总需求为593.5EB。
- **NAND短期供不应求：**我们仅考虑AI对NAND需求的拉动，根据前文测算，2025、2026年全球NAND总需求将达到1035.9、1344.4EB，而全球NAND供给为875.6、1014.7EB，2025、2026年全球NAND总需求/总供给分别为118%、132%，短期供不应求，有望驱动价格持续增长。

图179：全球NAND供求情况预测

	2024	2025(E)	2026(E)
ChatGPT 月活（亿人）		11.4	14.9
			30%
假设：单用户每月上传数据量规模（GB）		1	1.7
			70%
ChatGPT用户每月上传原始文本数据总量（EB）		1.1	2.4
ChatGPT用户全年上传原始文本数据总量（EB）		12.8	28.2
ChatGPT RAG NAND总需求/原始文本数据比例		10.9	10.9
ChatGPT对应RAG NAND总需求为（EB）		139.7	308.6
假设：RAG NAND/总NAND		70%	80%
ChatGPT NAND总需求（EB）		199.5	385.8
假设：ChatGPT占AI对NAND总需求的比例		70%	65%
AI对NAND总需求（EB）		285.0	593.5
NAND总需求（EB，25、26年仅考虑AI拉动）	750.9	1035.9	1344.4
NAND总供给（EB）	764.7	875.6	1014.7
NAND总需求/总供给	98%	118%	132%

资料来源：IDC，国信证券经济研究所测算

目录

全球存储公司梳理：总览（按市占率）

国信证券 GUOSEN SECURITIES

- 三星电子、海力士在DRAM、HBM、NAND、SSD等领域市占率均较高，为全球存储龙头公司；其次为美光，产品矩阵全面，但市占率略低于三星、海力士；
- 闪迪、铠侠聚焦于NAND、SSD领域，西部数据、希捷科技聚焦于HDD领域；
- 从地域来看，全球存储公司主要集中在韩国、美国、中国（包括中国台湾）、日本。

图180：全球存储公司总览（25Q2各产品市占率情况）

	DRAM	HBM	NAND	SSD	HDD	国家
三星电子	33.2%	17.0%	34.0%	32.0%		韩国
海力士	39.1%	67.0%	22.0%	6.8%		韩国
美光	22.8%	16.0%	13.8%	11.2%		美国
闪迪			12.1%	9.4%		美国
西部数据					46.5%	美国
希捷科技					41.5%	美国
铠侠			13.9%	9.1%		日本
合肥长鑫	3.1%					中国大陆
长江存储			4.2%			中国大陆
南亚	1.1%					中国台湾
华邦电子	0.5%					中国台湾
力晶	0.1%					中国台湾
Solidigm（海力士旗下）				12.5%		美国
东芝					12.0%	日本
金士顿				2.9%		美国
SSSTC				0.8%		中国台湾
其他				15.4%		

资料来源：IDC，国信证券经济研究所整理

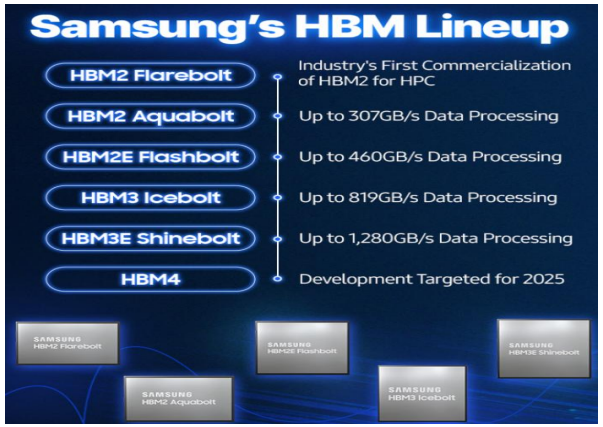
全球存储公司梳理：三星电子

国信证券 GUOSEN SECURITIES

■ **三星电子：全球NAND、DRAM领先企业。**三星电子业务包括消费电子产品、存储产品（NAND&DRAM等）、OLED手机面板以及Harman（音响产品）等。

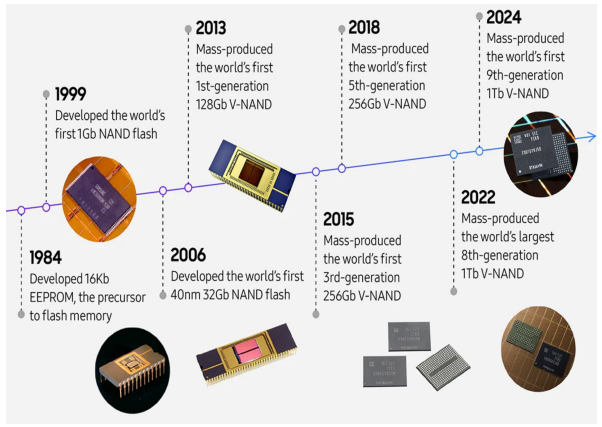
- 从收入结构来看：根据公司财报披露数据，3QFY2025公司DX部门（TV、手机、空调等）、DS部门（DRAM、NAND、移动Aps等）、SDC（OLED手机面板）、Harman（智能座舱、汽车音响等）收入占比分别为55%、33%、8%、4%，存储业务为公司第二大业务，仅次于消费电子业务。
- 从竞争格局来看：根据IDC披露数据，25Q2全球DRAM市场，三星电子市占率为33.2%，全球第二；25Q2全球NAND市场，海力士市占率34.0%，全球第一；25Q2全球HBM市场，三星电子市占率17.0%，全球第二；25Q2全球SSD市场三星电子市占率为32.0%，全球第一，在消费级、企业级SSD均具备竞争优势。
- 从发展历史来看：1）公司发展历史：公司1969年成立，最初主要生产电子电器类产品，1874年通过收购韩国半导体进军半导体业务；1983年开始发力DRAM领域，1984年研发成功64k DRAM；1988年公司在韩国推出第一部手机；1998年开发出第一款128MB Flash闪存，2004年公司研发出全球首款8GB NAND闪存，2005年开始进入晶圆代工行业，2011年出售HDD业务，专注发展存储芯片，2016年收购Harman，强化自身在汽车芯片、车载娱乐领域优势。2）从DRAM产品发展历史来看：1984年研发成功64k DRAM，1996年首次提出DDR概念，1998年发布全球第一个商用DDR芯片，2003年量产DDR2，2008年开始DDR3，2013年量产DDR4，2021年量产DDR5。3）从HBM产品发展历史来看：2016年量产HBM2，2020年量产HBM2E，2023年量产HBM3，2024年量产HBM3E。3）从NAND产品发展历史来看：1998年开发出第一款128NM Flash闪存，2013年开始进入3D V-NAND阶段（24层），后堆叠层数持续增长，2024年第九代V-NAND（300+层）量产。

图183：三星电子HBM产品lineup



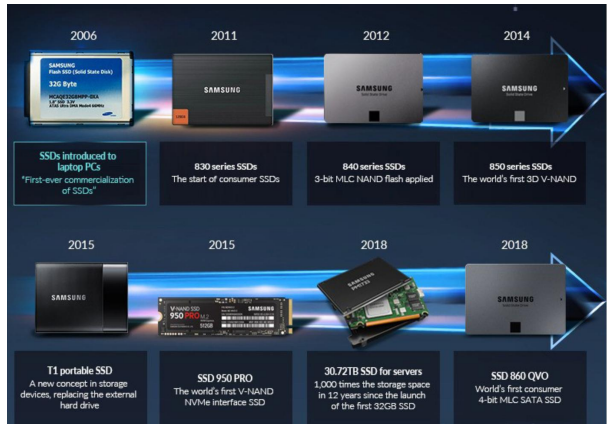
资料来源：三星电子，国信证券经济研究所整理

图184：三星电子NAND产品发展



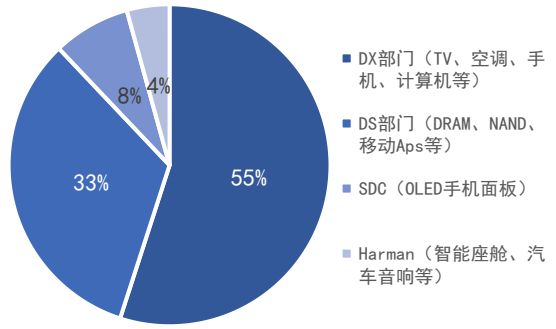
资料来源：三星电子，国信证券经济研究所整理

图185：三星电子SSD产品发展



资料来源：三星电子，国信证券经济研究所整理

图186：三星电子收入结构（25Q3）



资料来源：三星电子，国信证券经济研究所整理

全球存储公司梳理：海力士

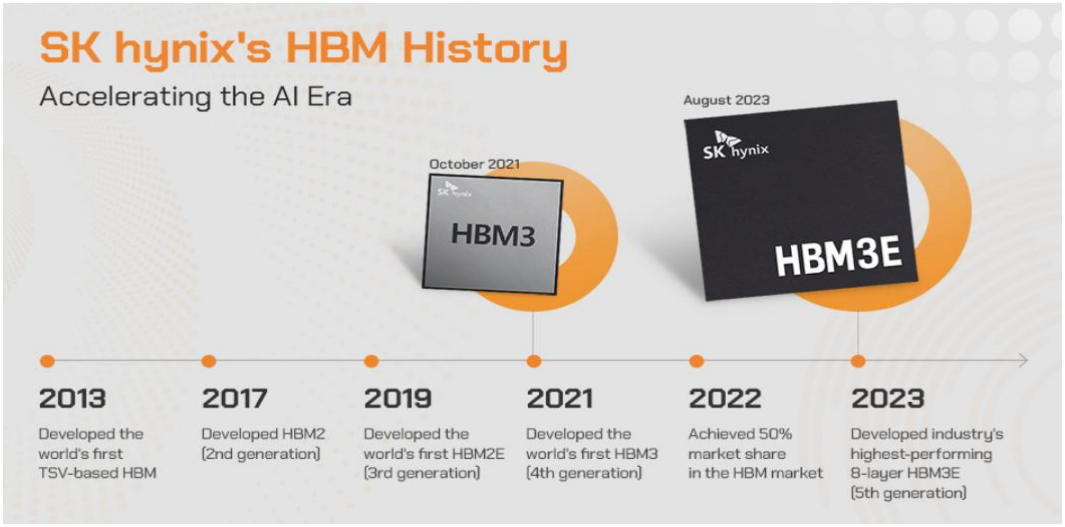
■ **海力士：全球NAND、DRAM领先企业。**海力士主要从事NAND、DRAM的开发、生产和销售，产品覆盖DRAM、NAND、HBM、SSD等。

- 从收入结构来看：根据公司财报披露数据，3QFY2025公司DRAM（包括HBM）、NAND（包括SSD）、其他业务收入占比分别为78%、20%、2%，DRAM、NAND业务为公司主营业务。
- 从竞争格局来看：根据IDC披露数据，25Q2全球DRAM市场，海力士市占率为39.1%，全球第一；25Q2全球NAND市场，海力士市占率22.0%，全球第二；25Q2全球HBM市场，海力士市占率67.0%，全球第一；25Q2全球SSD市场海力士市占率为19.6%（包括Solidigm，收购的英特尔的NAND业务），其中海力士优势在消费级SSD领域，市占率为9.15%，全球第四，Solidigm优势在企业级SSD领域，市占率23.18%，全球第二。
- 从发展历史来看：1）公司发展历史：公司1983年成立，最初名为现代电子，以DRAM的生产制造为主；1998年收购LG半导体，2001年改名为海力士，2003年从现代集团完成剥离，2004年出售非存储业务，将重心转向半导体存储；公司于2012年并入SK集团，改名为SK海力士；2021年SK海力士收购英特尔的NAND业务，成立了Solidigm；2023年开发并大规模生产全球最高规格的HBM3E。2）从DRAM产品发展历史来看：1985年，公司256kb DRAM量产，整合LG半导体后于2001年推出128MB DDR SDRAM；2003年底512MB DDR2通过英特尔认证，2004年开始量产；2007年开发出业内首款DDR3并获得英特尔认证；2011年成功研发2GB DDR4；2018年完成DDR5研发，2020年正式量产；2013年与AMD联合开发全球首款HBM1，2017年发布HBM2，2019年发布HBM2E，2021年发布HBM3，2023年量产HBM3E。3）从NAND产品发展历史来看：2004年开始进军闪存领域，开发了512MB NAND闪存，2014年正式进军3D NAND领域，推出24层3D NAND芯片，2018年业内率先实现4D NAND商业化突破，发布96层4D NAND闪存，2024年开始量产全球首款321层4D NAND闪存。

图187：海力士NAND产品发展

资料来源：海力士，国信证券经济研究所整理

图188：海力士HBM产品发展



资料来源：海力士，国信证券经济研究所整理

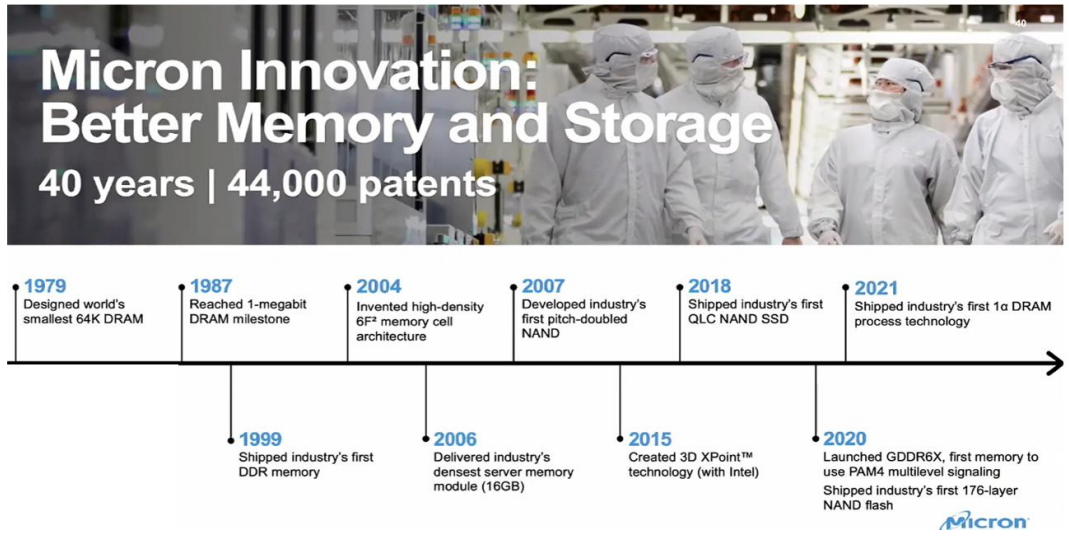
全球存储公司梳理：美光

国信证券 GUOSEN SECURITIES

■ **美光：全球DRAM、NAND领先企业。**美光主要从事DRAM、NAND的开发、生产和销售，产品覆盖DRAM、NAND、HBM等。

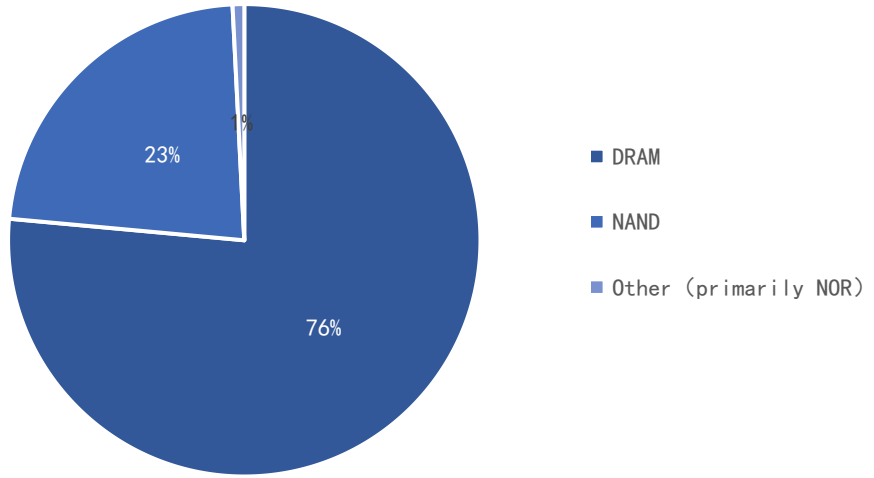
- 从收入结构来看：根据公司财报披露数据，FY2025公司DRAM（包括HBM）、NAND、其他业务收入占比分别为76%、23%、1%，DRAM、NAND业务为公司主营业务。
- 从竞争格局来看：根据IDC披露数据，25Q2全球DRAM市场，美光市占率为22.8%，全球第三；25Q2全球NAND市场，美光市占率13.8%，全球第四；25Q2全球HBM市场，美光市占率16.0%，全球第三；25Q2全球SSD市场美光市占率为11.1%，全球第三。
- 从发展历史来看：1）公司发展历史：公司1978年成立，以DRAM的生产制造为主，1981年量产64KB DRAM；1998年收购TI（德州仪器）全球存储业务，跻身全球存储前列；同年，并购Rendition，进入3D图形芯片领域；1999年发布产业内第一款DDR芯片；2002年美光收购日本东芝在美国的DRAM工厂，生产能力进一步提升；2005年美光与英特尔合资成立IM Flash Technologies，开进发力NAND领域，2006年又收购了Lexar Media，进一步开拓消费级存储卡和SSD市场；2009年收购奇梦达持有的华亚科技全部股份，2015年收购华亚科技全部股份，改名为台湾美光；2013年美光完成对尔必达的收购，改名为美光日本；2015年美光与英特尔联合发布3D NAND技术；2018年美光放弃HMC技术，开始转向HBM产品的研发，2020年美光开始提供HBM2产品。2）从DRAM产品发展历史来看：1981年，公司64KB DRAM量产，通过收购TI存储业务、华亚科技、尔必达，DRAM产品能力持续提升；1991年推出DDR1，2003年推出DDR2，2007年推出DDR3，2014年推出DDR4，2020年推出DDR5；3）从HBM产品发展历史来看：美光最初专注于HMC技术，2011年发布第一代HMC，2018年放弃HMC，开始转向HBM技术，2020年推出HBM2,后持续迭代；4）从NAND产品发展历史来看：2005年通过和英特尔合资公司进军闪存领域，2015年和英特尔联合发布3D NAND技术，2024年232层NAND已经大批量量产。

图189：美光发展历史



资料来源：美光，国信证券经济研究所整理

图190：美光收入结构（FY2025）



资料来源：美光，国信证券经济研究所整理

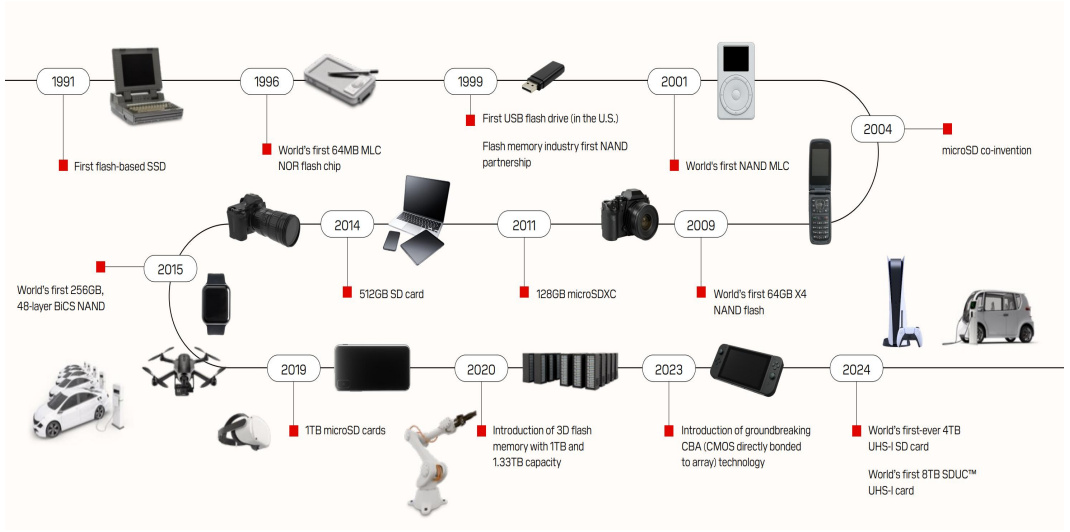
全球存储公司梳理：闪迪

国信证券 GUOSEN SECURITIES

■ **闪迪：全球消费级NAND领先企业。**闪迪主要从事闪存及固定硬盘（SSD）的开发、生产和销售，在消费级存储市场具有领先优势。

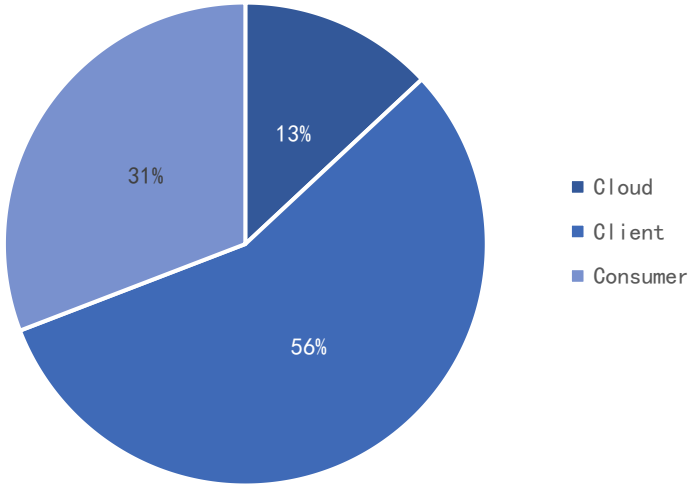
- 从收入结构来看：根据公司财报披露数据，公司产品主要包括存储卡、USB闪存驱动器、SSD，FY2025公司消费端、商业端、云端产品收入占比分别为56%、31%、13%。
- 从竞争格局来看：根据IDC披露数据，从全球NAND竞争格局来看，25Q2闪迪市占率为12.1%，全球市占率第五；从全球SSD市场来看，25Q2闪迪市占率为11.8%，全球市占率第三名；从消费级SSD市场，25Q2闪迪市占率分别为19.0%，市占率全球第二。
- 从发展历史来看：1）公司发展历史：1988年公司成立，1991年推出首款基于闪存的固态硬盘，2016年被西部数据收购，2025年西部数据剥离闪存业务，独立上市；2）从NAND和SSD产品发展历史来看：1991年闪迪为IBM生产了首款基于闪存的SSD；1992年推出FlashDisk，可以插入笔记本电脑的拓展槽中；1994年发布CF卡，成为早期数码相机和便携式电子设备的主流存储方案；2000年，公司正式推出SD存储卡格式；2001年，公司与东芝联合发布全球首款商业化的NAND MLC闪存技术；2004年，作为核心推动者参与制定microSD存储卡标准；2005年发布首款基于闪存的MP3播放器SanDisk Sansa e100，开始进入数字音频播放器领域；2011年，公司收购固态硬盘制造商Pliant Technology；2012年，发布首款SATA SSD产品，发力消费级SSD市场；2013年，公司收购面向企业市场的SSD制造商SMART Storage Systems；2014年，公司收购企业数据中心闪存制造商Fusion-io，开始布局企业级SSD市场；2019年闪迪主导发布SD Express标准，通过PCIe NVMe将接口速度大幅提升；2024年闪迪发布microSD Express存储卡。

图181：闪迪产品发展历史



资料来源：闪迪，国信证券经济研究所整理

图182：闪迪收入结构（FY2025）



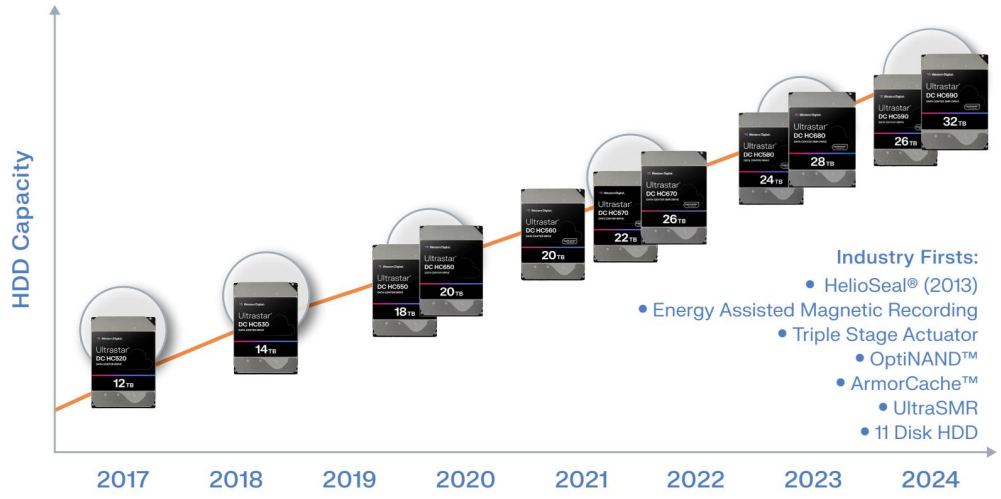
资料来源：闪迪，国信证券经济研究所整理

全球存储公司梳理：西部数据

■ **西部数据：全球HDD领先企业。**西部数据主要从事硬盘（HDD）的开发、生产和销售，覆盖数据中心级、商业级、消费级HDD产品。

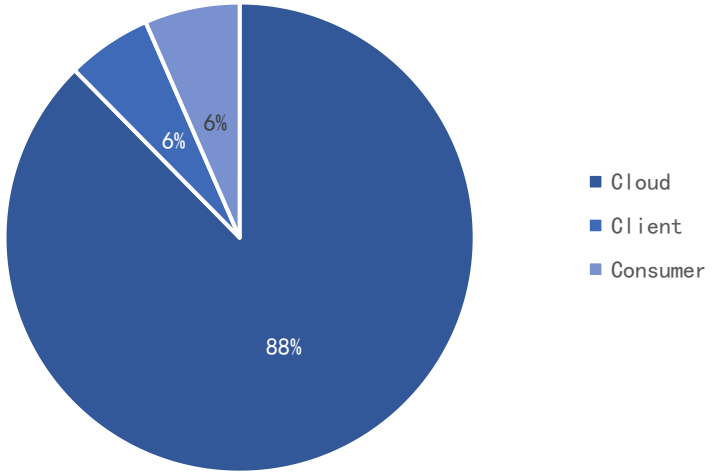
- 从收入结构来看：根据西部数据财报披露数据，HDD为其核心产品，FY2025财年西部数据Cloud（云）、Client（商业级）、Consumer（消费级）收入占比分别为88%、6%、6%，数据中心级别HDD贡献主要收入。
- 从竞争格局来看：根据IDC披露数据，2024年希捷、西部数据、东芝市占率分别为40.8%、40.0%、19.2%，其中希捷、西部数据市占率合计达80.8%，呈现双寡头垄断格局。
- 从发展历史来看：1）公司发展历史：1970年西部数据成立，为全球知名的存储解决方案提供商，1990年开始确立“硬盘主业”战略，2011年收购日立环球存储科技，扩张企业级HDD市场，2016年收购闪迪，确立“HDD+闪存”双核心业务，2025年完成对闪迪分拆，正式剥离NAND闪存业务，目前全面聚焦于HDD领域；2）从HDD产品发展历史来看：1988年收购Tandon，切入硬盘制造领域；1992年发布Caviar驱动器，获得市场认可；1999年，同IBM合作，推出Expert系列驱动器；2001年，西部数据成为第一家提供8MB磁盘缓冲区的主流ATA硬盘驱动器制造商；2003年收购磁头制造商Read-Rite，推出首款10,000转SATA硬盘WD Raptor；2005年，开始进入笔记本HDD市场，推出Scorpio系列硬盘；2006年推出采用垂直磁记录（PMR）技术的硬盘；2009年，推出全球首款2TB硬盘WD Caviar Green；2012年，完成对HGST收购，成为全球最大的传统硬盘制造商，并获得HGST在企业级HDD和氦气硬盘技术优势；2013年，发布全球首款商用氦气硬盘Ultrastar He；2017年，展示MAMR（微波辅助磁记录）技术原型；2020年，发布ePMR技术(Energy-Assist PMR)。

图191：西部数据产品发展



资料来源：西部数据，国信证券经济研究所整理

图192：西部数据收入结构（FY2025）



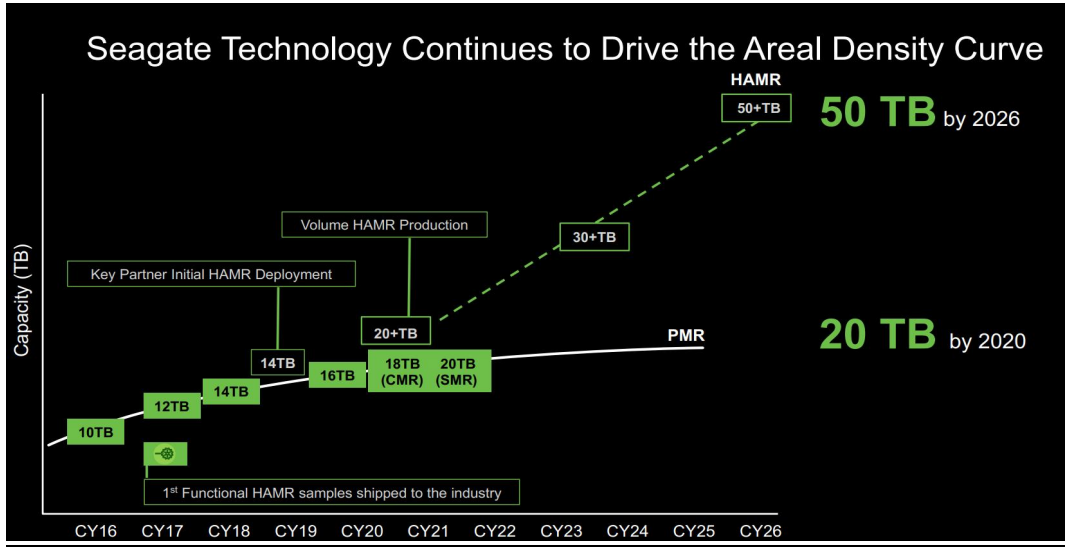
资料来源：西部数据，国信证券经济研究所整理

全球存储公司梳理：希捷科技

■ **希捷科技：全球HDD领先企业。**希捷科技主要从事硬盘及存储产品的开发、生产和销售，核心产品为HDD硬盘，此外亦覆盖系统、SSD等产品。

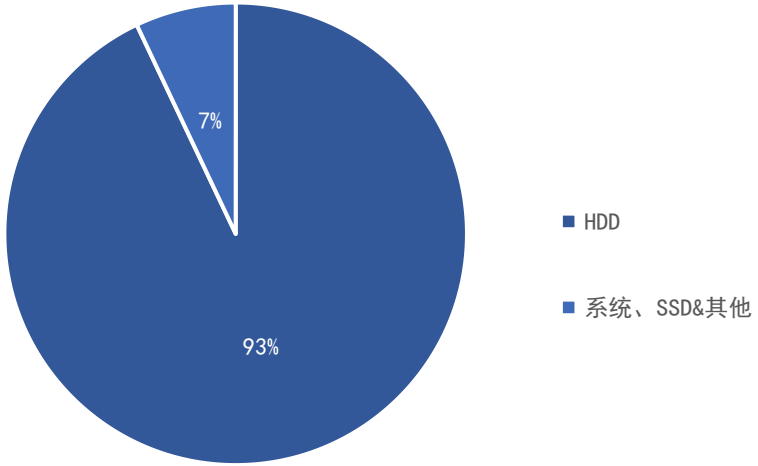
- 从收入结构来看：根据公司财报披露数据，FY2025公司HHD、系统&SSD&其他产品收入占比分别为93%、7%，HDD产品业务为公司主营业务。
- 从竞争格局来看：根据IDC披露数据，2024年HDD市场希捷、西部数据、东芝市占率分别为40.8%、40.0%、19.2%，希捷为全球龙头；SSD产品全球占比较低。
- 从发展历史来看：1）公司发展历史：公司成立于1978年，1980年发布全球首款5.25英寸硬盘ST-506，后成为IBM PC主要供应商，1989年收购CDC旗下的存储部门Imprimis Technology，进一步扩大企业级HDD市场份额；1996年，公司与Conner Peripherals合并，成为当时全球最大的独立存储设备制造商；2003年发布Momentus系列产品，重回笔记本HDD市场；2006年收购Maxtor，整合其消费级产品线，市场份额进一步提升；2011年，为应对西部数据收购HSFT，希捷收购三星旗下硬盘业务；2013年布局SSD市场，2014年收购LSI闪存业务；2020年公司HAMR技术开始商业化，HDD容量大幅提升。2）从HDD产品发展历史来看：1980年公司发布全球首款5.25英寸硬盘ST-506，为个人电脑提供了首个标准存储解决方案，1981年发布10MB版本ST-412，以此获得IBM主要OEM供应商合同，后出货量快速增长；1989年收购CDC旗下的存储部门Imprimis Technology，进一步扩大企业级HDD市场份额；1992年发布全球首款7200RPM硬盘BarraCuda（酷鱼），1996年发布全球首款10000RPM硬盘Cheetah系列，巩固其在企业级存储领域地位；2003年发布Momentus系列硬盘，重回笔记本硬盘领域；2005年收购Maxtor，整合其消费级产品线FreeAgent等，扩大消费级市场份额；2011年收购三星旗下硬盘业务，进一步提升市场份额；2020年HAMR技术开始商业化，2022年推出基于HAMR技术的Exos M系列硬盘，2025年推动第二代HAMR技术开发，目标容量提升到30TB以上。

图193：通过HAMR技术，希捷产品容量持续提升



资料来源：希捷科技，国信证券经济研究所整理

图194：希捷科技收入结构（FY2025）



资料来源：希捷科技，国信证券经济研究所整理

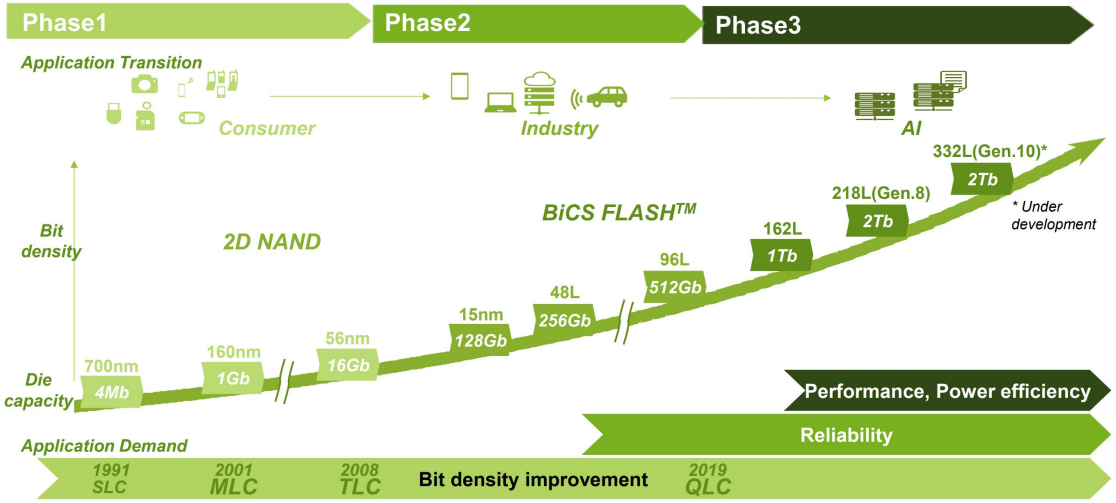
全球存储公司梳理：铠侠

国信证券 GUOSEN SECURITIES

■ **铠侠：全球NAND闪存领先企业。**铠侠主要从事闪存及固定硬盘（SSD）的开发、生产和销售，覆盖企业存储产品、个人消费存储产品。

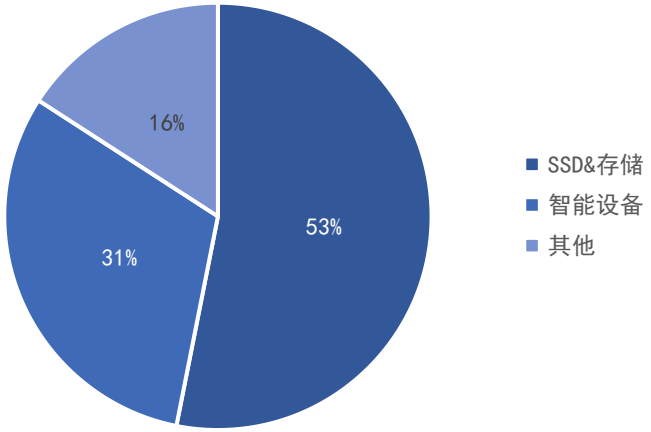
- 从收入结构来看：根据公司财报披露数据，2025H1公司SSD&存储、智能设备、其他业务收入占比分别为53%、31%、16%，SSD&存储业务为公司主营业务。
- 从竞争格局来看：根据IDC披露数据，从全球NAND竞争格局来看，25Q2三星、海力士、铠侠、美光、闪迪占比分别为34.0%、22.0%、13.9%、13.9%、12.1%，铠侠仅次于三星、海力士；从全球SSD市场来看，25Q2三星、Solidigm、闪迪、美光、铠侠市占率分别为28.7%、12.6%、11.8%、11.2%、8.1%，全球市占率第五名。
- 从发展历史来看：1）公司发展历史：1987年东芝公司发明NAND闪存，2017年铠侠前身东芝存储器集团从东芝公司剥离，2019年公司改名为铠侠；2）从NAND产品发展历史来看：1987年公司发明NAND闪存，1991年发布全球首个4MB NAND闪存（全球第一个商业化NAND产品），2001年铠侠和闪迪联合开发，推出首款1GB MLC NAND产品（首次在NAND中使用MLC技术，实现每单元存储2位数据），2007年率先发布三维(3D)闪存层叠技术，2014年发布全球首款15nm工艺NAND产品，2015年发布全球首款256GB（48层）TLC BICS FLASH，实现了3D NAND技术的商业化突破，2019年全球首次实现量产96层QLC 3D NAND产品，后续堆叠层数、容量持续提升；3）从SSD产品发展历史来看：2001年发布首个PATA SSD，2007年发布第一代SATA接口SSD（128GB），2010年首款SAS SLC企业级SSD发布，2012年发布首款SAS MLC企业级SSD+QSBC技术（错误矫正技术），2016年面向消费级市场推出SATA接口TLC SSD产品，2018年发布全球首款采购96层BiCS4 FLASH的NVMe SSD。

图195：铠侠产品发展



资料来源：铠侠，国信证券经济研究所整理

图196：铠侠收入结构（2025H1）



资料来源：铠侠，国信证券经济研究所整理

目录

风险提示

- 厂商DRAM、NAND扩产，进而导致产品价格下降风险；
- 互联网大厂资本开支不及预期风险；
- AI应用活跃用户数增长不及预期风险；
- AI大模型方案优化，进而减少对存储需求风险；

免责声明

国信证券 GUOSEN SECURITIES

国信证券投资评级

投资评级标准	类别	级别	说明
报告中投资建议所涉及的评级（如有）分为股票评级和行业评级（另有说明的除外）。评级标准为报告发布日后6到12个月内的相对市场表现，也即报告发布日后的6到12个月内公司股价（或行业指数）相对同期相关证券市场代表性指数的涨跌幅作为基准。A股市场以沪深300指数（000300.SH）作为基准；新三板市场以三板成指（899001.CSI）为基准；香港市场以恒生指数(HSI.HI)作为基准；美国市场以标普500指数(SPX.GI)或纳斯达克指数（IXIC.GI）为基准。	股票投资评级	优于大市	股价表现优于市场代表性指数10%以上
		中性	股价表现介于市场代表性指数±10%之间
		弱于大市	股价表现弱于市场代表性指数10%以上
		无评级	股价与市场代表性指数相比无明确观点
	行业投资评级	优于大市	行业指数表现优于市场代表性指数10%以上
		中性	行业指数表现介于市场代表性指数±10%之间
		弱于大市	行业指数表现弱于市场代表性指数10%以上

分析师承诺

作者保证报告所采用的数据均来自合规渠道；分析逻辑基于作者的职业理解，通过合理判断并得出结论，力求独立、客观、公正，结论不受任何第三方的授意或影响；作者在过去、现在或未来未就其研究报告所提供的具体建议或所表述的意见直接或间接收取任何报酬，特此声明。

重要声明

本报告由国信证券股份有限公司（已具备中国证监会许可的证券投资咨询业务资格）制作；报告版权归国信证券股份有限公司（以下简称“我公司”）所有。本报告仅供我公司客户使用，本公司不会因接收人收到本报告而视其为客户。未经书面许可，任何机构和个人不得以任何形式使用、复制或传播。任何有关本报告的摘要或节选都不代表本报告正式完整的观点，一切须以我公司向客户发布的本报告完整版本为准。

本报告基于已公开的资料或信息撰写，但我公司不保证该资料及信息的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映我公司于本报告公开发布当日的判断，在不同时期，我公司可能撰写并发布与本报告所载资料、建议及推测不一致的报告。我公司不保证本报告所含信息及资料处于最新状态；我公司可能随时补充、更新和修订有关信息及资料，投资者应当自行关注相关更新和修订内容。我公司或关联机构可能会持有本报告中所提到的公司所发行的证券并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或金融产品等相关服务。本公司的资产管理部门、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意见或建议不一致的投资决策。

本报告仅供参考之用，不构成出售或购买证券或其他投资标的要约或邀请。在任何情况下，本报告中的信息和意见均不构成对任何个人的投资建议。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。投资者应结合自己的投资目标和财务状况自行判断是否采用本报告所载内容和信息并自行承担风险，我公司及雇员对投资者使用本报告及其内容而造成的一切后果不承担任何法律责任。

证券投资咨询业务的说明

本公司具备中国证监会核准的证券投资咨询业务资格。证券投资咨询，是指从事证券投资咨询业务的机构及其投资咨询人员以下列形式为证券投资人或者客户提供证券投资分析、预测或者建议等直接或者间接有偿咨询服务的活动：接受投资人或者客户委托，提供证券投资咨询服务；举办有关证券投资咨询的讲座、报告会、分析会等；在报刊上发表证券投资咨询的文章、评论、报告，以及通过电台、电视台等公众传播媒体提供证券投资咨询服务；通过电话、传真、电脑网络等电信设备系统，提供证券投资咨询服务；中国证监会认定的其他形式。

发布证券研究报告是证券投资咨询业务的一种基本形式，指证券公司、证券投资咨询机构对证券及证券相关产品的价值、市场走势或者相关影响因素进行分析，形成证券估值、投资评级等投资分析意见，制作证券研究报告，并向客户发布的行为。

国信证券
GUOSEN SECURITIES

国信证券经济研究所

深圳

深圳市福田区福华一路125号国信金融大厦36层

邮编：518046　　总机：0755-82130833

上海

上海浦东民生路1199弄证大五道口广场1号楼12楼

邮编：200135

北京

北京西城区金融大街兴盛街6号国信证券9层

邮编：100032