



# 2024-2025年全球 存储市场趋势白皮书

Global Memory Market  
Trend White Paper

2025年3月



# CATALOGUE

## 目 录

### 第一章 全球存储市场变化和技术发展分析

ANALYSIS OF GLOBAL MEMORY MARKET CHANGES AND TECHNOLOGICAL DEVELOPMENTS

一、高层数3D NAND Flash存储密度持续提升	01
二、服务器NAND和DRAM应用占比持续增长	04
三、存储技术和应用发展展望	06
1、QLC NAND持续改善存储性能和可靠性，适用于读取密集型应用场景	
2、AI PC带动消费类PCIe 4.0/5.0 SSD应用增长	
3、PC品牌及存储厂商积极拓展海外市场	

### 第二章 服务器存储产品应用与市场分析

APPLICATION AND MARKET ANALYSIS OF SERVER MEMORY PRODUCTS

一、全球AI资本支出井喷式增长，AI服务器存储需求火热	12
二、全球HBM 市场规模快速增长，DRAM产能向HBM倾斜	18
三、企业级PCIe 5.0 SSD需求呈现快速增长	22
四、QLC NAND技术发展进入成熟期，带动SSD进入100TB时代	25
五、开源AI大模型加速推动AI与终端应用的融合	27
六、服务器新型存储器LPCAMM/SOCAMM和MRDIMM/MCRDIMM	29

## 第三章 消费类存储产品应用与发展分析

APPLICATION AND DEVELOPMENT ANALYSIS OF CONSUMER MEMORY PRODUCTS

一、PC设备的存储应用与发展	31
1、2025年全球PC出货小幅增长，AI PC渗透率攀升	
2、PCIe 4.0 SSD仍是主流，主控厂商助力PCIe 5.0 SSD加速发展	
3、国产PC产业链锋芒毕露，2025年信创PC市场充满潜力	
二、智能手机终端的存储应用与发展	38
1、全球智能手机销量温和增长，品牌竞争激烈	
2、智能手机通过AI技术加持进行创新，为用户带来新体验	
3、eMMC5.1市场需求趋于稳定，UFS3.1广泛覆盖5G机型，UFS4.0高端机型应用增加	
4、端侧大模型的普及，进一步提升手机单机容量	
5、手机终端踊跃出海，全球化与本地化相结合	

## 第四章 AI消费电子发展与存储应用分析

ANALYSIS OF AI CONSUMER ELECTRONICS DEVELOPMENT AND MEMORY APPLICATIONS

一、AI眼镜产品发展与技术动向	50
二、AI下一代交互入口的技术竞速	51
三、AI重塑硬件产业的未来，打开存储应用想象空间	52

# 第一章 全球存储市场变化和技术发展分析

ANALYSIS OF GLOBAL MEMORY MARKET CHANGES AND TECHNOLOGICAL DEVELOPMENTS

## 一、高层数3D NAND Flash存储密度持续提升

本节要点：

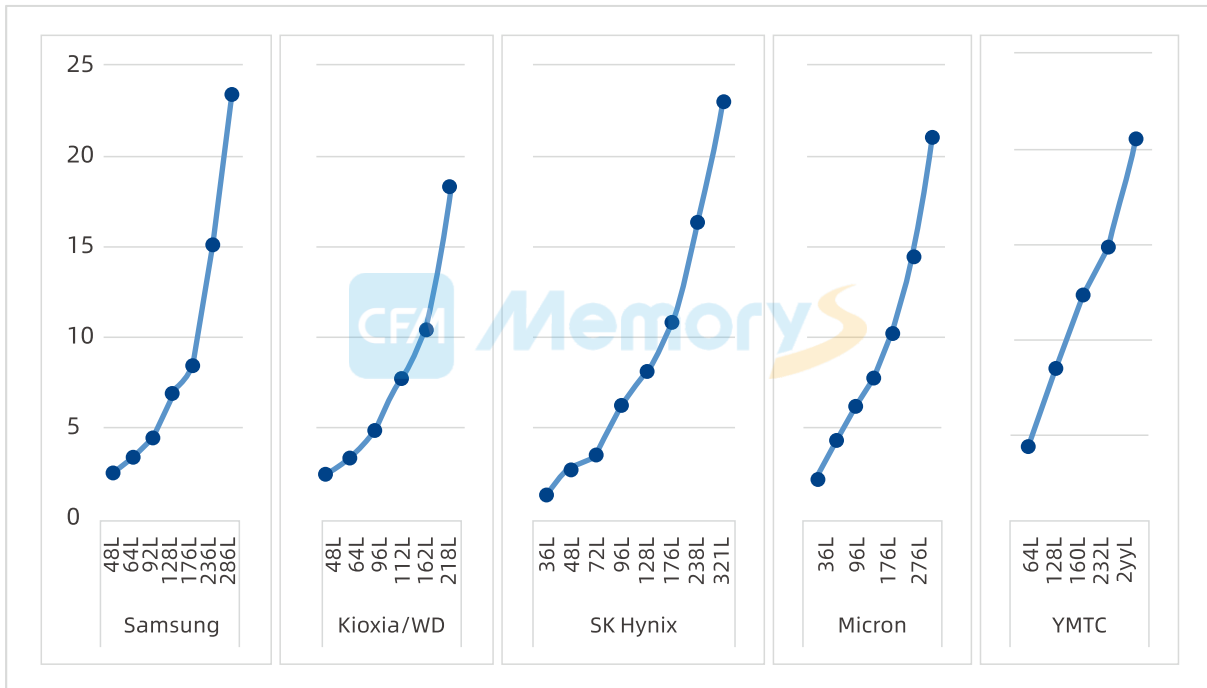


- NAND Flash追求更高的存储密度，NAND产品在读写性能、功耗、耐久、成本等方面持续优化和改善。
- 高层数的NAND Flash将更多地采用双晶圆键合架构，通过将存储单元和外围电路分别在不同晶圆上制造后，再进行键合。

3D NAND通过横向拓展和垂直堆叠以及架构优化，持续改善存储密度。尤其在人工智能、大数据和云存储等应用场景，高密度NAND Flash凭借其高性能、低功耗及更优的成本等综合优势，为高性能运算提供重要的存力支持。在消费类市场，高密度存储顺应电子元器件轻薄化趋势，为智能手机和笔电提供了更薄更轻量化的设计支持，推动多折叠手机及超轻量笔电的发展。

随着NAND Flash进入200层以上超高层时代，高纵横比刻蚀和沉积的工艺复杂性增加。即便在双堆栈架构下，单次也需要刻蚀更深的孔，高层NAND Flash发展面临着高层数易错位和沟道刻蚀偏离的挑战，并且随着多晶硅通道总电阻增加，通道高度也对读取电流构成了阻碍。因此，越来越多存储原厂不再一味追求垂直堆叠的高层NAND，而是通过横向拓展多层存储孔密度、逻辑拓展单元存储比特，以及采用双晶圆键合架构，整体实现更优的NAND存储密度和读写性能。

图1: TLC NAND Flash存储密度 (Gb/mm<sup>2</sup>)



来源: 公开信息

存储原厂最新NAND技术发展概况, 如下

●三星: V9 TLC/QLC NAND单die容量为1Tb, 采用的通道孔蚀刻技术, 能够基于双堆栈架构实现业内最高的单元层数, 为双堆栈的286层NAND Flash。三星成功完成三堆栈的400层NAND技术开发, 已于2024年11月将这项先进技术转移到平泽P1工厂的量产线上, 并持续优化晶圆良率, 预计2025年将准备好大规模量产400层NAND Flash。

●SK海力士: 321层NAND Flash采用3-Plug工艺技术, 分三次进行通孔工艺流程, 随后经过优化的后续工艺将3个通孔进行电气连接, 并引进通孔间自动排列矫正技术。与上代相比生产效率提升了59%, 数据传输速度和读取性能分别提高了12%和13%。

●美光：G9 NAND为276层NAND Flash，为11.5x13.5mm的BGA封装尺寸，可减少28%的PCB面积占用，支持3.6GBps的传输速度，较上一代NAND的存储密度提高了44%。

●铠侠/西部数据\*（西部数据闪存业务于2025年2月21日完成分拆，现为闪迪SanDisk，下文亦同）：BiCS8 218层NAND Flash采用横向微缩技术和CBA（CMOS直接键合到阵列）技术，存储密度提升超过50%，NAND I/O速度超过3.2GB/s，在写入性能和读延迟方面的改善超过20%。

●长江存储：NAND技术迭代至第四代晶栈Xtacking4.0，NAND延续背面源极连接（BSSC），采用混合晶圆键合结构和20孔垂直通道设计，持续改善NAND生产效率并提高产量。

目前，三星、SK海力士均在执行300层以上NAND的量产计划，同时原厂均对W2W（Wafer-to-Wafer）双晶圆键合技术产生较大的应用兴趣。W2W混合键合技术，即在两片独立晶圆上分别加工存储单元和外围电路，再利用垂直互联通道（VIA）将两片晶圆键合在一起。存储单元和CMOS电路的解耦，为材料应用和研发带来更多空间，可选用的材料均一性和可靠性更好，有效克服了传统架构下NAND的高温工艺瓶颈和单元Cell数上升时阈值电压的稳定性问题。同时，键合的方式省去传统芯片连接中所需的“凸点”（Bump），高密度互连有效缩短了电路路径并改善散热。整体来看，NAND Flash混合键合技术的应用，有效提高了NAND Flash单位存储密度、存储性能和可靠性，并在量产环节提高了NAND生产效率，实现了解耦晶圆的模块化量产。

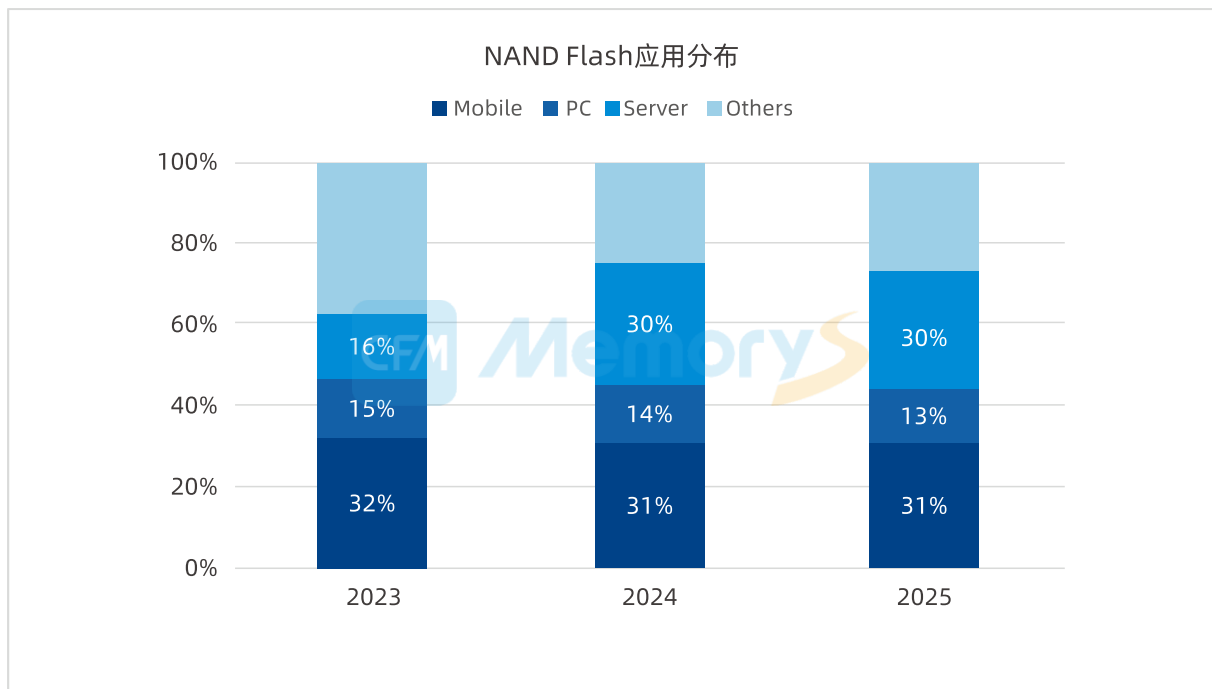
当前，混合键合技术已成为NAND Flash重要的技术发展方向，存储原厂持续通过优化架构和材料，克服超高层NAND Flash的量产挑战。据韩国媒体报道，三星电子与长江存储签署3D NAND混合键合专利的许可协议，计划从2025年下半年量产的V10 NAND开始，借助长江存储的3D NAND混合键合相关专利，将存储单元和控制电路分别在不同晶圆上制造后再键合。NAND Flash技术路径正在形成新的风向，各家存储原厂也需要经过一定时间，逐步攻克技术研发的挑战，从而更好地为数据中心HPC和端侧AI等应用提供高性能高容量的先进存储产品，以满足日益增长的存储需求。

## 二、服务器NAND和DRAM应用占比持续增长

近年来全球数据中心基础设施扩建浪潮升温，大型云服务商适度超前投资AI基础设施，为未来长期发展提前部署算力资源。NAND方面，算力中心更积极地应用超高容量eSSD提升整机性能并有效降低能耗。预计2025年互联网企业对PCIe 5.0 eSSD需求增加，32TB及以上QLC eSSD在服务器市场的渗透率持续增长，令近年来服务器存储位元需求大幅攀升。而在2024年及2025年全球手机及PC出货量持稳的市况下，端侧单机容量增长是消费端存储需求增长的有效推动力，高端化趋势下终端平均售价的提升，为终端和存储厂商提供新的机遇和挑战。

据CFM闪存市场数据显示，在2024年全球存储应用市场中，服务器、手机和PC，在NAND应用占比分别为30%、31%和14%，服务器NAND应用占比从2023年的16%将攀升至2025年的30%。

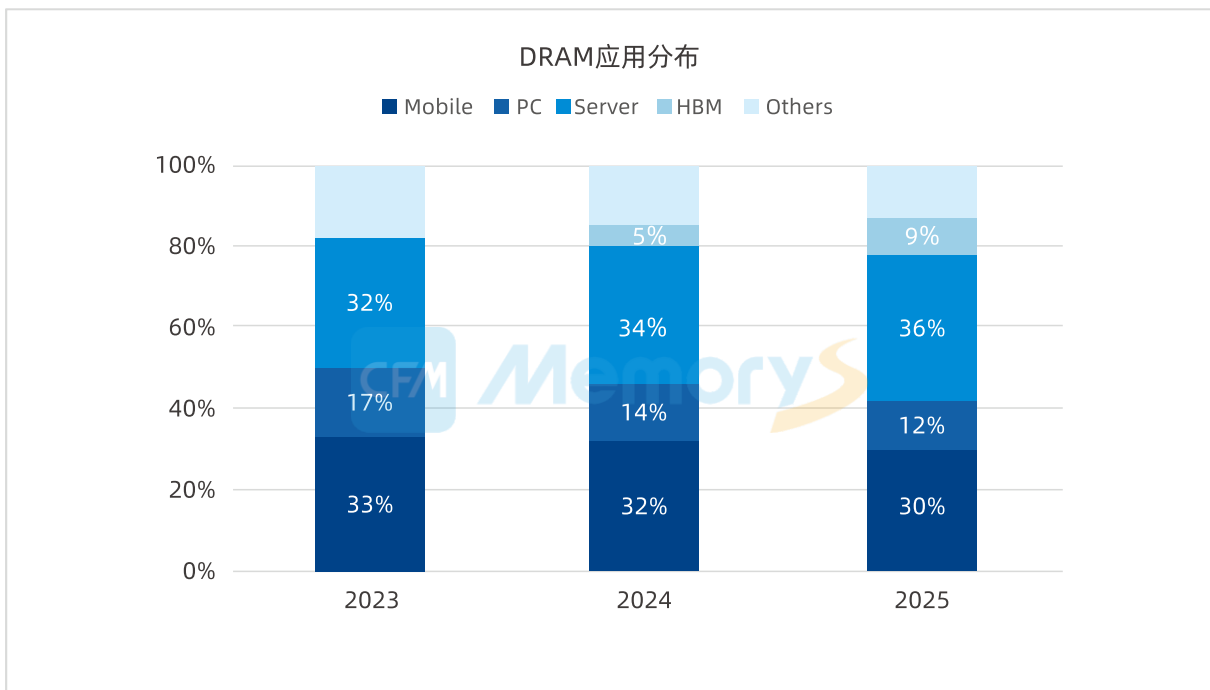
图2：NAND Flash应用分布



数据来源：CFM闪存市场

DRAM方面，随着支持DDR5的Intel EMR和AMD Bergamo处理器平台渗透率增加，以及搭载Nvidia Hopper/Blackwell和AMD MI300等加速芯片的AI服务器出货放量增长，令服务器DDR5及HBM需求快速增长。据CFM闪存市场数据显示，2024年服务器、手机、PC终端在DRAM应用占比分别达34%、32%和14%，其中服务器DRAM应用占比由2023年的32%，预计将增长至2025年的36%。在64GB及以上DDR5及HBM3e 12hi出货增长的带动下，2025年服务器存储位元需求将同比增长超过20%。

图3：DRAM应用分布



数据来源：CFM闪存市场

总体来看，近两年服务器存储需求增幅领涨于各主要应用市场，存储原厂将新增产能和资本支出向高性能eSSD、DDR5及HBM倾斜，AI需求持续推动存储技术和产品迭代发展。CFM闪存市场预计，2025年NAND和DRAM容量需求增长将分别增长12%和15%。



### 三、存储技术和应用发展展望

#### 1、QLC NAND持续改善存储性能和可靠性，适用于读取密集型应用场景

存储原厂持续改进QLC NAND的存储密度和读写性能，令QLC NAND更适配AI、ML和在线分析处理等应用，对高带宽、低延迟和读取优化的性能需求。尤其是算力中心的温热数据多以读取密集型操作为主，且对电力、冷却等产生的运营成本格外看重。QLC NAND通过持续迭代优化性能和可靠性，帮助企业有效提高了存储性能并降低了总拥有成本。

表1：各存储原厂QLC NAND存储密度和I/O速率

	三星	铠侠/西部数据	SK海力士	Solidigm	美光	长江存储
代数	V7 176L QLC	BiCS6 162L QLC	176L QLC	144L QLC	176L QLC	X3 QLC
存储密度	15.3 Gb/mm <sup>2</sup>	15 Gb/mm <sup>2</sup>	14.8 Gb/mm <sup>2</sup>	13.8 Gb/mm <sup>2</sup>	14.7 Gb/mm <sup>2</sup>	20.62 Gb/mm <sup>2</sup>
I/O速率	1.6 GB/s	2.4 GB/s	1.6 GB/s	N/A	1.6 GB/s	2.4 GB/s
代数	V9 290L QLC	BiCS8 218L QLC	N/A	192L QLC	232L QLC	X4 QLC
存储密度	28.5 Gb/mm <sup>2</sup>	23.3 Gb/mm <sup>2</sup>	N/A	18.6 Gb/mm <sup>2</sup>	19.5 Gb/mm <sup>2</sup>	N/A
I/O速率	2.4 GB/s	3.6 GB/s	N/A	N/A	2.4 GB/s	N/A

来源：公开信息

NAND Flash中，每个Plane包含多个Block，每个Block包含多个Page。NAND以Block为最小擦除单元进行擦除，以Page为最小读写单位进行数据写入。在Block擦除前，需要将Page中的有效数据进行搬移并保留，从而触发垃圾回收(GC, Garbage Collection)和写放大(WA, Write Amplification)。NAND Flash每个block的擦写次数是有限的，一般NAND闪存单元可容纳的电荷越多，其最大擦写次数P/E Cycles耐用性越低，即SLC、MLC、TLC到QLC的擦写循环次数呈衰减趋势，取而代之的是进一步提高的存储容量并优化了使用成本。

通常情况下，SLC NAND可以达到100,000 P/E Cycles，MLC NAND多在1,000~10,000之间，TLC NAND和QLC NAND的早期P/E Cycles则分别处于500~3,000和100~1,000左右。当然，随着数据纠错能力的提高，消费类TLC NAND从早期的约1000 P/E Cycles达到3000~5000 P/E Cycles。QLC NAND也从早期的150 P/E Cycles，迭代升级到足以媲美TLC NAND的高耐久度，如长江存储通过独特的Xtacking架构和日渐成熟的技术工艺，其X3-6070 QLC NAND可达到4000 P/E Cycles，是上一代产品

的四倍，读取性能和写入性能均有近100%的提升。三星的QLC NAND则从176L V7跳过236L的V8，直接迭代至290L的V9 QLC NAND，其采用预设模具技术，通过调节单元工作字线(WL)的间距，确保层间及层内单元特性的均匀性，并通过预测编程技术和低功耗设计，减少非必要操作并大幅降低功耗。三星V9 QLC NAND的数据保持性能较前一代提升约20%，写入性能翻倍，数据输入输出速度提升60%，读写数据功耗分别减少30%和50%，计划首先从品牌消费类产品开始，逐步扩展至移动端UFS、PC和服务器SSD应用。

## 2、AI PC带动消费类PCIe 4.0/5.0 SSD应用增长

CFM预计在2025年至2026年，PCIe 4.0 SSD是绝对的主力应用，PCIe 5.0 SSD将逐渐导入高端消费类和企业级市场，企业级导入PCIe 5.0 SSD的进度较消费类市场更快。在PC市场中，AI PC渗透率的持续攀升，加上NAND Flash芯片性能和主控功耗的持续改善，PCIe 4.0/5.0 SSD的应用占比超过75%，PCIe 5.0 SSD以更高性能和更低功耗的优势，有望在AI PC和电竞类PC中率先导入应用。

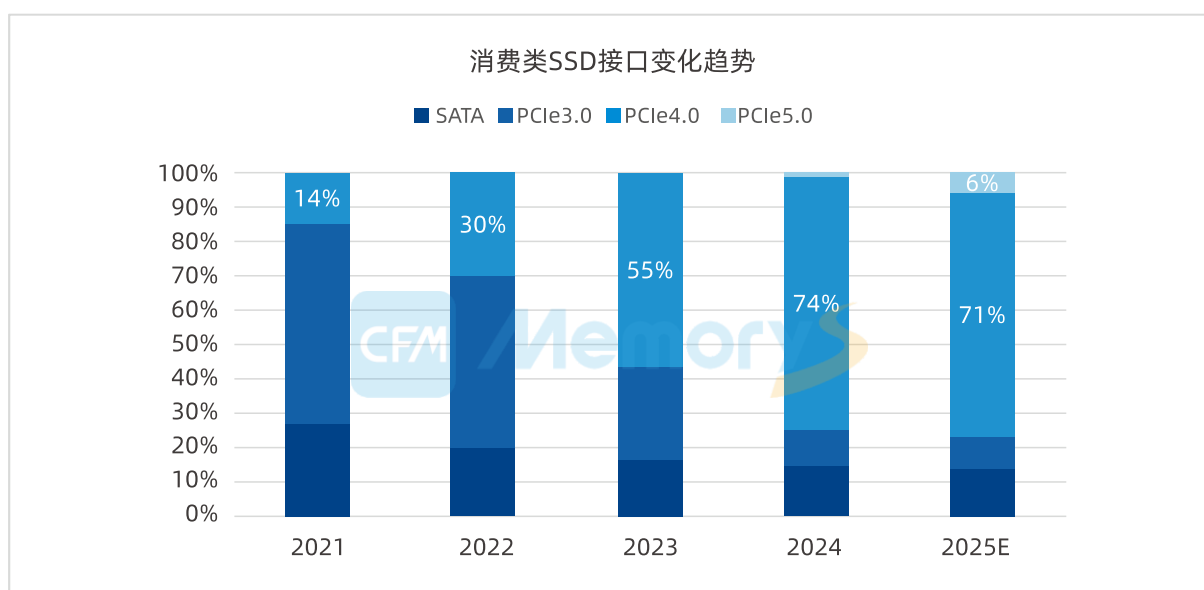
根据英特尔、微软及联想等厂商对AI PC的定义，CFM总结AI PC的基本特性为：①硬件上需集成专用NPU，算力基本在40TOPS以上；②重点强调端侧AI大模型的运行，减少云端数据处理，重视隐私安全性；③存储配置上，基础AI模型需要16GB内存，标准AI模型则需要32GB内存，高级AI模型则要求64GB或更高内存；根据目前已发布的AI PC，x86架构产品所配置的SSD主要在1-2TB，ARM架构在256GB-1TB，Apple Silicon在256GB-8TB。

2024年全球PC出货量达2.53亿台，其中AI PC出货量约4500万台，AI PC占比达18%。尤其自二季度以来，华为发布搭载盘古大模型的MateBook X Pro，联想发布配备AI智能体联想小天的Yoga Book系列，微软官宣新一代Windows为Copilot+PC等，大量AI PC产品的加速问世。AI PC强调本地AI大模型的便捷性和安全性，并有效规避版权风险以及提高工作效率，保障数据信息的隐私性安全性等等的优势，令AI PC将率先在重度内容创作类和开发类用户群体中普及。随着芯片及PC厂商竞相推出AI相关产品，产业链规模效益显现，AI PC将逐渐从初期高价向下缩小与传统PC的价格差距，应用热度有望在2025年至2026年持续升温。2025年，业界普遍以AI PC出货量达1亿台为目标。预计2025年AI PC的渗透率将达到35%，其中AI PC在笔记本电脑中的占比将更高，未来出货超过半数的笔记本电脑将是搭载NPU和本地AI模型的AI PC。

随着PCIe 5.0 SSD功耗的持续优化，目前的消费类PCIe 5.0 SSD较早期产品的功耗已大幅下降，无需添加散热盖顶即可工作。PCIe 5.0 SSD的主控由12nm制程演进至6/7nm的先进制程，采用四通道Dram-less的设计可达到超过10GB/s的连续读取性能，同时功耗低至6~7瓦，空闲状态下的功率更是低

至3.5毫瓦，极大程度地延长笔记本电脑的使用时长和电池寿命。PCIe 5.0 SSD最高可突破14.5GB/s的连续读写速度，加上支持PCIe 5.0的低成本主板逐渐推出，高性能AI PC及电竞类PC将是PCIe5.0 SSD的潜在应用市场。不过，零售端消费需求普遍对价格敏感度极高，令PCIe 5.0 SSD短期在零售市场中的需求难有显著增长，零售端主要为针对性满足少数电竞高性能需求，消费类PCIe 5.0 SSD的主要增长动力主要来自于PC OEM在AI PC整机方案中的应用增长。

图4：消费类SSD接口变化趋势



来源：CFM闪存市场

CFM闪存市场预计，消费类SSD市场中，PCIe 5.0 SSD的出货占比将在2025年增长至6.2%，PCIe 4.0/5.0 SSD占比将达到77%。整体来看，PCIe 5.0 SSD将在2025年中旬逐步导入主力OEM端的AI PC，并有望在2026年通过AI PC整机出货放量。在服务器市场中，高算力数据中心快速发展对存储器性能及功耗提出更高的要求，因此PCIe 5.0 SSD在企业级市场中将迎来更快的增幅。先进的PCIe 5.0 eSSD有助于降低服务器和数据中心的运营成本，大型云服务商及头部互联网企业对PCIe 5.0 eSSD存在一定的应用需求。CFM闪存市场预计，在企业级SSD市场中，2025年PCIe 5.0 eSSD的出货占比有望达到30%，为存储厂商带来新的增长契机。

### 3、PC品牌及存储厂商积极拓展海外市场

从全球市场来看，笔记本电脑的市场需求持续增长，而台式机的市场需求则呈逐渐收缩的趋势，如在成熟市场的日本，笔记本出货量占整体比重约九成。以欧洲、北美、日本为主的发达国家或地区，企业及个人消费者对高质量的新功能产品有更明显的偏好。受益于技术进步和产品更新，PC兼具更好的能效比和高性能。镁合金材料、OLED屏幕、高密度电池和材质更轻、设计更合理的主板等，PC厂商正在逐一攻克超轻薄笔电的痛点，重量仅为1KG的超轻量PC开始进入大众视野。随着成熟市场的通胀降温和经济周期性复苏，欧美日等发达国家及地区的消费力回升，操作系统的更新和AI功能的强化，以及在4-5年前疫情期间采购的设备到达更换的时间节点，将令全球PC销量逐渐回升，其中一部分更新需求将转移至AI PC，未来一两年的PC市场机遇与挑战并存。

虽然智能手机和平板电脑的性能提升，可能在一些应用场景中取代部分PC，但在AI端侧应用增加的趋势下，PC仍是端侧最关键最高效的生产力设备。上游芯片厂商英特尔、AMD、英伟达、高通等积极把部分云端AI能力引入到PC端上，令AI PC的技术进展和生态系统不断完善，联想、惠普、戴尔等PC品牌厂商对AI PC表现出极大的兴趣和支持。微软计划于2025年10月结束对Windows 10的支持，截止2025年1月，目前仍有六成以上的PC用户使用Windows 10，升级至Windows 11的进度较为缓慢，这可能会使得数千万台PC面临包括网络攻击和资料泄露在内的安全风险。基于信息安全的诉求和数字化转型的持续推进，企业及个人需要更高效、更安全的设备，不过考虑到部分用户可能会额外认购官方提供的拓展方案，从而延缓换机速度，因此预计大规模的PC换机浪潮将在Windows 10正式终止后出现。CFM闪存市场预计，2025年下半年尤其是进入四季度以后，商业PC的更新需求，将带动笔电PC出货有更明显的回升。

分地区来看，在欧洲、北美、日韩等成熟市场，2025年PC市场可能会经历一个小幅增长周期，这主要由企业需求驱动，消费者市场的增长更为温和。其中，高性能PC如AI PC的份额占比越来越大，尤其年轻消费者对内容创作和垂直专业的应用，对高性能PC的需求持续增长。而中低端消费市场可能面临更激烈的价格竞争，出海的新兴品牌向市场发起挑战，加剧市场份额的竞争。考虑到全球各地的经济复苏周期和消费支持政策不同，在不同国家和地区的细分表现将有所差异，其中日本PC市场率先出现显著回暖，教育及游戏需求带动2024年日本PC出货量四年来首次增长，预计增长趋势将持续到2025年。

考虑到G20成员国和组织约占全球GDP的85%，占全球贸易的75%以上，约占世界人口的三分之二，对全球消费市场具有极大的影响力。下面以G20及常驻受邀国家的人口数量、人均国民收入和人口年龄中位数为例：

表2: G20成员国/地区的人均国民总收入与人口结构

国家/地区	人均国民总收入(美元)	人口	年龄中位数
美国	80,450	3.35亿	38.5
卡塔尔	79,430	266万	33.5
新加坡	70,590	592万	36.2
澳大利亚	63,150	2666万	38.3
德国	54,800	8330万	45.5
加拿大	54,040	4010万	40.6
阿联酋	49,020	1050万	31.6
英国	47,700	6835万	40.1
法国	45,180	6830万	42.3
日本	39,350	1.25亿	49.8
意大利	37,920	5900万	48.2
韩国	35,490	5171万	45.6
西班牙	32,830	4830万	45.9
沙特阿拉伯	31,860	3330万	29.6
智利	15,800	1966万	36.9
俄罗斯	14,250	1.44亿	40.3
中国大陆	13,390	14.11亿	40.1
阿根廷	12,890	4554万	32.9
墨西哥	11,980	1.30亿	29.6
土耳其	11,730	8533万	33.5
马来西亚	11,710	3513万	31.0
巴西	9,280	2.11亿	34.8
南非	6,480	6321万	28.7
印尼	4,810	2.81亿	30.4
越南	4,110	1.00亿	33.4
埃及	3,840	1.15亿	24.5
印度	2,540	14.38亿	28.8

来源: 公开信息

在东南亚、拉丁美、亚非拉等发展中国家和地区中，传统PC市场发展仍是SSD应用的主要动力。结合新兴国家中产阶级扩张和人口年轻化结构，以及当地基础建设和数字化建设的加速，在东南亚如印尼、越南、马来西亚、菲律宾，非洲如尼日利亚、南非等，拉丁美洲如墨西哥、巴西、阿根廷等国家，经济、技术、人口或特定行业展现出一定增长趋势，消费电子产品具备潜在增长空间，正在经历SSD逐渐取代HDD市场的阶段。

不过在一些地区，可能由于网络覆盖、电力供应等基础设施存在不足，限制了SSD的广泛应用。另外，发展中国家的当地消费者对价格非常敏感，SSD平均容量偏低且对成本要求较高，开拓新兴SSD市场仍然面临着一定挑战，需要品牌厂商持续地正向引导消费市场，提高当地消费者对SSD的价值认知，并持续通过技术创新、品牌建设等方式扩大市场份额和影响力。

## 第二章 服务器存储产品应用与市场分析

APPLICATION AND MARKET ANALYSIS OF SERVER MEMORY PRODUCTS

### 一、全球AI资本支出井喷式增长，AI服务器存储需求火热

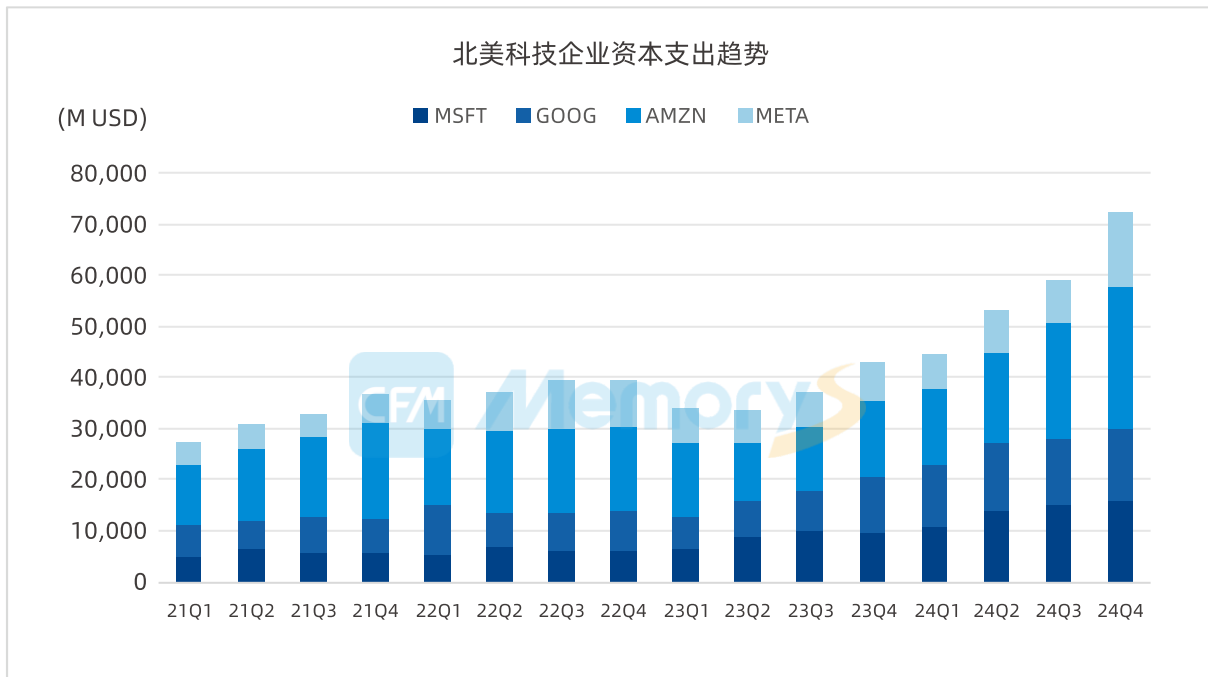
背景



- 数据中心企业级存储需求是推动本轮存储行情发展的“排头兵”以及“中坚力量”。Chat GPT的横空出世将生成式AI技术推向了风口浪尖，一时间全球各大科技与互联网厂商纷纷跟进，加码硬件设备投资。在AI浪潮的助推下，高性能AI服务器对存储设备特性提出了更高的要求，进而推动存储技术和产品迭代升级。然而，AI的发展注定是迂回且波折的，每一种技术路线的尝试都可能带来翻天覆地的变化，这是科技进步的必然经历，存储作为供应链中极为重要的一环，也需随时准备应对变化。

随着服务器及数据中心客户库存调整完毕，以及大型AI超算数据中心兴建浪潮席卷全球，大型云服务商及科技企业的资本支出，主要集中在数据中心建设、云计算和AI基础设施相关的厂房和设备的购建上，直接带动AI服务器需求显著升温。据CFM统计，微软、谷歌、亚马逊和Meta为首的科技企业，2024年全年资本支出分别同比增长58%、63%、57%、37%，四家大型云服务商合计资本支出高达2283亿美元，同比2023年的1474亿美元大幅增长55%。

图5：北美科技企业资本支出趋势



来源：CFM闪存市场

2025年初，美国宣布一项涉及数千亿美元的“星际之门”(Stargate)项目，被称为史上最大的人工智能基础设施投资项目，据悉该项目由OpenAI、软银和甲骨文合资，计划未来4年投资5000亿美元，以共同大力投资AI基础设施，计划建设20个数据中心。软银主要负责财务，OpenAI负责项目运营，初始技术合作伙伴包括Arm、微软、英伟达、甲骨文和OpenAI。OpenAI将陆续宣布更多的建设地点，每个数据中心园区的设计能够提供大约1吉瓦的电力。目前，北美科技企业对AI基础设施的投入，主要为了用于支持未来十余年AI运行的长期资产，因此在北美当地建设了大量基础设施、电力系统和数据中心，从而带动全球AI基建投资竞赛浪潮。据CFM初步统计，2025年微软、谷歌、亚马逊和Meta的资本支出总额将超过3200亿美元，同比增幅将超过40%。

●微软：预计2025财年（截止2025年6月30日）年度将投资高达800亿美元用于建设AI数据中心，以训练AI模型并部署人工智能及基于云端的应用程序，其中约一半以上的投资将投资于美国。对于OpenAI，微软迄今已向OpenAI投资超过130亿美元，保留向其提供云资源的优先权并为其提供云计算基础设施，将OpenAI与微软自己的AI模型一起整合到Windows、Teams等核心产品中。

●谷歌：母公司Alphabet计划2025年资本支出投入约750亿美元，主要用于服务器和数据中心基础设施建设、谷歌云业务拓展、Gemini模型及其配套服务等业务，引入多模态AI增强搜索产品和广告的用户体验，向Google Cloud用户开放其新一代的TPU Trillium芯片，提供从代码生成到数据处理等一系列AI服务。



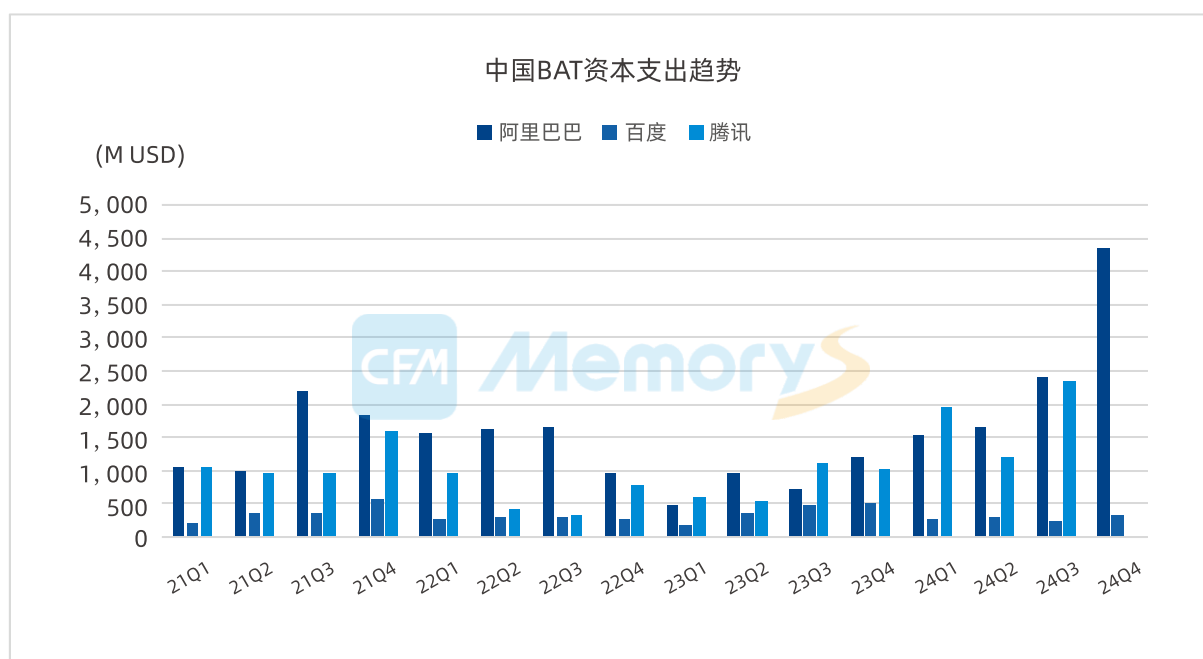
●亚马逊：2025年资本支出提升至1000亿美元，大部分资本支出将用于支持AWS发展的AI技术研发和基础设施建设，已推出Nova多模态模型系列、自研Trainium芯片、AI购物助手Rufus以及Bedrock模型库等，持续加大对数据中心、网络设备及硬件的投资。

●Meta：预计2025年全年资本支出将在600亿美元至650亿美元之间，资本支出中服务器仍然是最大的投资方向，以增加AI计算能力，扩展AI基础设施规模，并更新老旧服务器，从而支持生成式AI和核心业务的增长。

●苹果：计划未来四年，在美国投资超过5000亿美元，为公司成立以来最大规模的支出承诺。苹果计划在德克萨斯州开设AI服务器工厂，占地250,000平方英尺，预计将于2026年投入运行。作为此次新投资计划的一部分，苹果公司将其美国先进制造基金规模从50亿美元翻倍至100亿美元。随着苹果智能系统在全美范围内的推广，苹果还计划继续扩展位于北卡罗来纳州、爱荷华州、俄勒冈州、亚利桑那州和内华达州的数据中心容量。

●特斯拉：2025年资本支出将超过110亿美元，预计2026、2027财年的资本支出同样超过110亿美元，主要用于投资AI和自动驾驶技术，包括部署Dojo超级计算机集群等。特斯拉将在中国推出城市道路Autopilot自动辅助驾驶，正为中国客户分批次更新软件，以提供FSD完全自动驾驶功能。另外，马斯克旗下xAI公司发布了Grok 3模型，并称当Grok 3成熟稳定时，将开源Grok 2。马斯克称，未来将会进一步加大数据中心的建设力度，xAI已着手研究下一个更强大的算力集群，所需电力高达1.2GW。

图6：中国BAT资本支出趋势



数据来源：公开信息

●字节：预计2025年资本开支或高达1600亿元，资本支出主要集中在AI算力采购、IDC基础设施建设、网络设备采购以及海外AI基础设施建设等领域。各类模态基模沿时间线更新迭代，推动文本、语音、图像等模型在端侧上的应用覆盖，并为模型定制推理芯片，可能加速在海外如欧洲、南美、中东等地区自建数据中心。

●阿里：计划在未来三年内投入超过3800亿元，主要用于AI基础设施建设、AI基础模型及原生平台研发、以及现有业务的AI转型。这一投资金额将超过过去十年阿里在云和AI基础设施上的总投入，并创下中国民营企业在该领域的最高纪录。

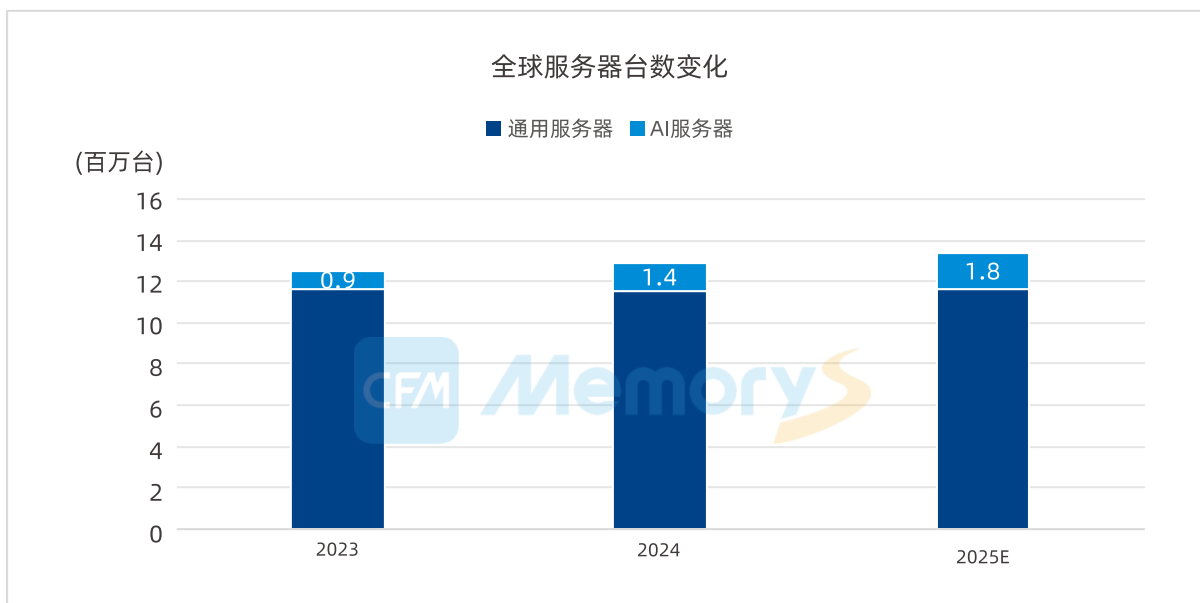
●腾讯：强调持续对AI基础模型和AI创新进行投入，全面接入DeepSeek后加速推动云需求增长，服务器资源利用率和API服务需求激增。腾讯2024年一至三季度的资本支出分别为143.6亿元、87.3亿元和170.9亿元，主要用于GPU芯片和服务器的采购，以支持其混元模型和AI算法的发展。腾讯云将在沙特阿拉伯建设其首个中东数据中心，为中东及周边区域客户提供高质量的弹性计算、存储、安全、AI等云服务。

●百度：以提升AI能力作为长期战略重点进行投资，确保严格的投资回报率管理和有效控制资本支出。通过文心大模型和智能体技术，提升搜索体验和内容生成能力，推进Apollo平台和萝卜快跑的覆盖范围，通过智能云业务、千帆平台和持续迭代的文心大模型系列，加速AI在金融、教育、医疗等垂直行业领域的应用。

●小米：2024年小米研发投入达到240亿元，预计2025年资本支出将进一步增长至300亿元，未来五年内，小米将累计投入超过1050亿元用于研发。主要将用于AI、操作系统以及芯片等关键领域的研发，可能通过云服务商加大AI大模型的投资力度，以推动技术创新提高市场竞争力。

科技企业通过提供丰富的AI模型生态，显著提高了服务器资源的利用率和云服务客户的黏性，对云计算发展和应用形成正反馈作用。如此势头下，AI服务器已然成为驱动全球存储市场增长的重要引擎。据CFM数据显示，2024年全球服务器规模增长3.2%达到1290万台，其中通用服务器台数为1150万台，同比基本持平，AI服务器规模约140万台，同比增长56%。预计2025年全球服务器市场出货量将达到1340万台，同比增长约3.9%，其中通用服务器台数预计约1160万台，与2024年基本持平，AI服务器台数有望达到180万台，同比增长29%。

图7：全球服务器台数变化



数据来源：CFM闪存市场

AI服务器单机存储容量的提升显著增加。以英伟达H100 AI服务器为例，所需的HBM容量可达640GB，所需的DDR5容量可达2TB-4TB，所需的NAND容量可达32TB-132TB。AI服务器市场需求增长确定性较强，在AI大模型训练及推理的应用场景下，智能终端给对存储产品的容量、性能与稳定性提出了更高的要求，带动存储技术迭代和产品应用升级。

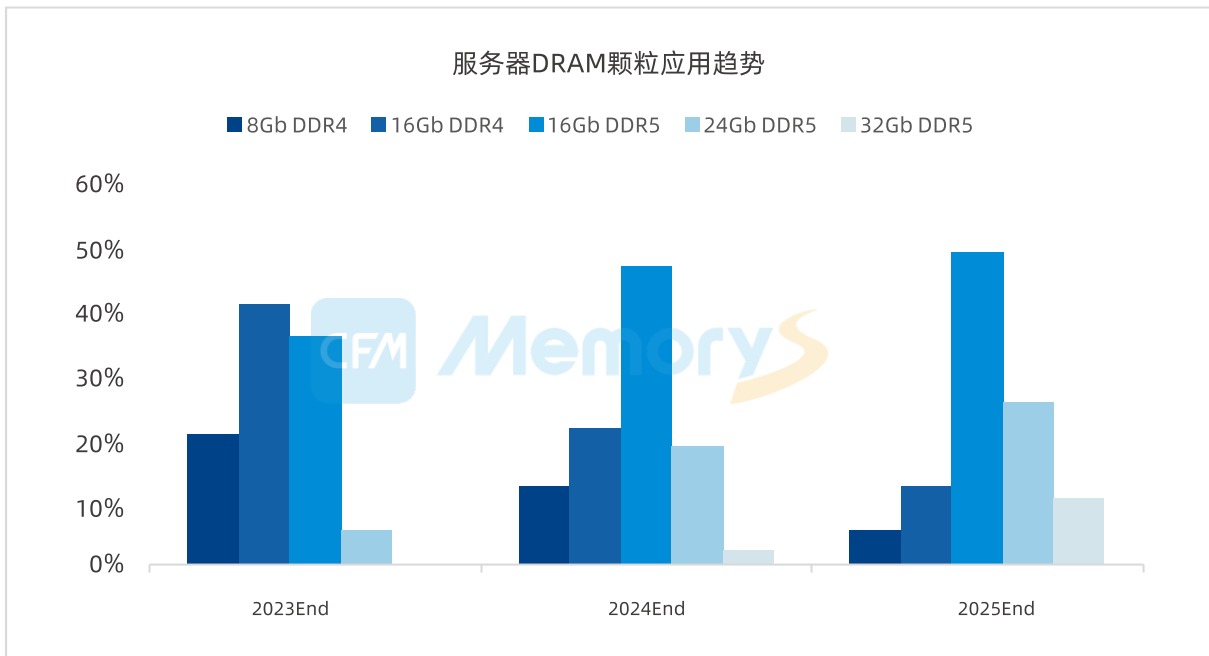
表3：英伟达H100 AI服务器配置

	Memory	单位容量	最大数量	合计容量 (GB)
GPU	HBM3	80GB	8	640GB
系统CPU	DDR5	64-128GB	32	2048-4096GB
内部存储	eSSD	4TB-16TB	8	32-128TB
OS存储	eSSD	2TB	2	4TB

数据来源：公开信息

据CFM数据显示，2024年二季度末，全球服务器市场DRAM出货量中DDR5已然占过半比例，预计2025年末全球服务器市场DRAM出货量中约八成为DDR5，其中单die容量为16Gb的DDR5仍是主力应用，24Gb和32Gb的DDR5出货占比也将大幅提升。目前三大存储原厂均已推出32Gb DDR5，并积极推动在企业级和数据中心客户的验证推广。相较于16Gb DDR5，基于32Gb DDR5单die容量翻倍，无需采用TSV工艺即可量产128GB DDR5内存条，大幅削减高容量DDR5内存条的生产和采购成本，并为更高容量的DDR5内存产品的研发和量产开拓道路。

图8：服务器DRAM颗粒应用趋势

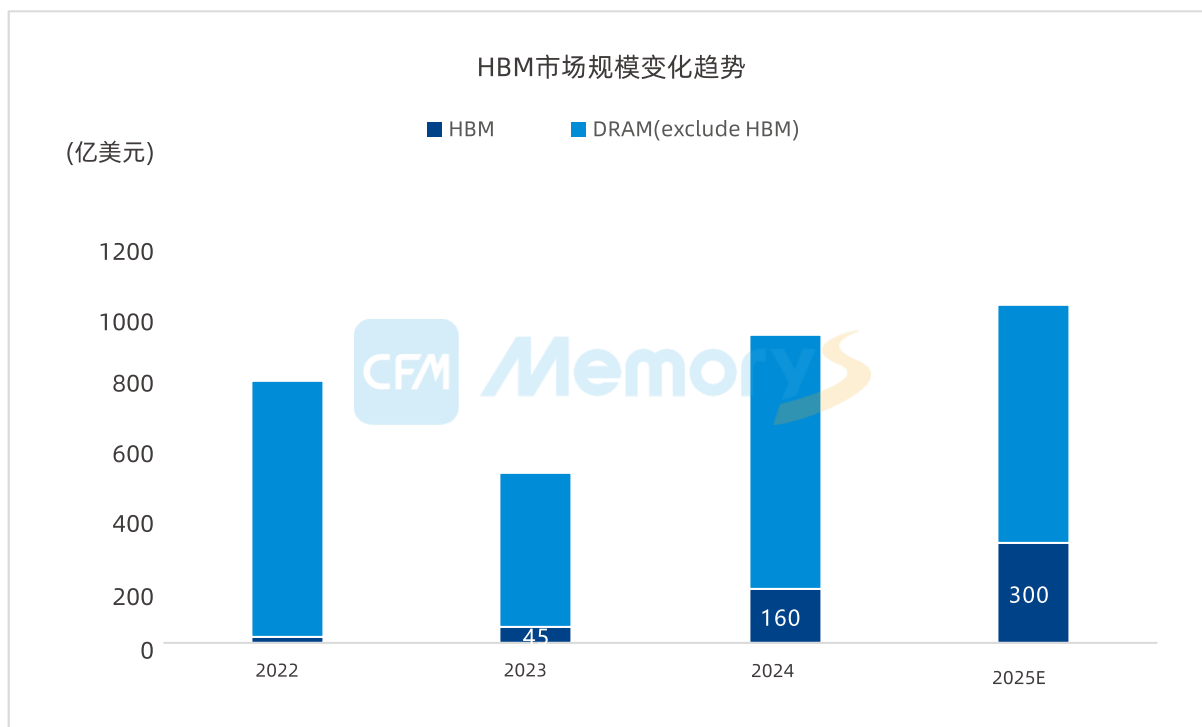


数据来源：CFM闪存市场

## 二、全球HBM 市场规模快速增长，DRAM产能向HBM倾斜

传统的冯诺依曼计算架构下的“存储墙”瓶颈已经阻碍了服务器性能的提升，为解决这个问题，各大厂商推出过非易失性存储（SCM）、存算一体、CXL、GDDR等诸多方案，HBM也是其中一个。HBM，即High Bandwidth Memory，具备高带宽、多I/O数量、低功耗等诸多高性能特性，然而在问世初期却因极高价格的成本劣势被冷落，但是在AI技术发展和应用对存算力提出更高的要求下，HBM市场自2024年起进入爆发增长阶段。据CFM数据显示，2024年全球HBM市场规模达到160亿美元，是2023年市场规模的3倍有余，约占整体DRAM市场的16%，预计2025年，HBM市场规模有望达到300亿美元，约占全球DRAM市场的28%。

图9：HBM市场规模变化趋势



数据来源：CFM闪存市场

HBM技术自2013年列入JEDEC标准以来，已经经过多个版本迭代，芯片堆叠层数、存储容量、带宽也持续增长。截至目前，JEDEC已经陆续披露了HBM4的多个核心参数，HBM4内存将采用4Hi、8Hi、12Hi和16Hi堆栈配置，支持24 Gb和32 Gb的DRAM Die，使得最高容量可达64GB。数据传输速率方面，JEDEC规定HBM4拥有6.4 GT/s，还具有支持高达10 GT/s速度的空间，这使得每个HBM4内存堆栈的内存带宽达到2.56 TB/s，具有2048位内存接口。

表4：HBM技术迭代

	HBM	HBM2	HBM2E	HBM3	HBM3E	HBM4
最多DRAM层数	4-Hi	8-Hi	12-Hi	12-Hi	12-Hi	16-Hi
最大容量	1GB	8GB	16GB	24GB	36GB	64GB
I/O数	1024	1024	1024	1024	1024	2048
I/O传输速率	1Gbps	2.4Gbps	3.6Gbps	6.4Gbps	9.2Gbps	10Gbps
带宽	128GBps	256GBps	480GBps	819GBps	1.2TBps	2.56TBps
制程技术	3x	2x	1y/1z	1z	1a/1b	1b/1c

来源：公开信息

在当前AI芯片市场，即便GPU、ASIC和FPGA三大技术路线仍然有所博弈，但是在高带宽内存上，HBM几乎是各大技术门派一致的选择。根据CFM不完全统计，HBM3/2E是目前主流的AI芯片采用的型号。然而，随着英伟达GB300、AMD MI325X的相继推出，以及其他市场参与者平台迭代，HBM3E市场份额有望得到显著提升。

表5：AI加速卡的HBM配置

供应商	AI芯片	应用领域	芯片名称	制程节点	存储搭配	最大需求容量
英伟达	GPU	AI训练	H100	4nm	HBM3	94GB
			H200	4nm	HBM3e	141GB
			H800	4nm	HBM2e	80GB
			A100	7nm	HBM2/2e	80GB
			A800	7nm	HBM2/2e	80GB
			B300	4NP	HBM3e	288GB
AMD	GPU	AI训练	MI200	6nm	HBM2e	128GB
			MI300	5nm/6nm	HBM3	192/128GB
	FPGA	推理	Virtex UltraScale+HBM	16nm	HBM2	16GB
			Versal HBM系列	7nm	HBM2e	32GB

数据来源：公开信息

续表5：AI加速卡的HBM配置

供应商	AI芯片	应用领域	芯片名称	制程节点	存储搭配	最大需求容量
英特尔	ASIC	训练/推理	Gaudi系列	7nm	HBM2e	128GB
	GPU	训练/推理	Max GPU	Intel 7	HBM2e	128GB
	FPGA	训练/推理	Atera stratix	14nm	HBM2	16GB
谷歌	ASIC	训练/推理	TPU系列	3-4nm	HBM2e	32GB
其他	ASIC	训练/推理	国产芯片为主	7-12nm	HBM2/2e	/

从供应格局上，目前全球HBM市场呈现SK海力士、三星电子和美光三足鼎立的市场格局。其中，SK海力士依然稳坐全球HBM市场第一把交椅，其次是三星和美光。财报显示，2024年SK海力士创下有史以来最佳年度业绩，其中四季度HBM销售额已占DRAM总销售额的40%以上。SK海力士HBM4 12Hi将在2025年完成开发并确保量产，目标是在2025年下半年向客户供应HBM4芯片，并计划在2028年至2030年期间推出HBM5芯片。三星2024年三季度已经量产HBM3e，并积极推动在英伟达的应用测试。美光尽管在HBM领域的进展相对落后于三星和SK海力士，但为了追赶进度直接跳代至HBM3E，并积极扩大产能规模以及下一代产品研发。美光计划在2026年推出 HBM4 内存产品，随后在2027至2028年推出HBM4E。

三星电子和SK海力士等厂商正在积极推进定制化HBM产品，以更好地满足数据中心客户定制化HBM的需求，预计在HBM4量产时实现定制化HBM的商业化。定制化HBM将在性能、功率、PAA方面提供多种选项，并且通过将内存输入/输出（I/O）和控制器功能从加速器转移到HBM基板上，从而在更多的逻辑空间中分配AI功能。

HBM快速发展带来的丰厚利润和规模增长前景，令存储原厂持乐观态度并积极扩大HBM产能。据CFM数据显示，按2025年初HBM的市场价格，HBM per GB ASP约为DDR5 RDIMM的2.5至3倍。AI服务器需求明朗且高性能存储利润充裕，叠加消费类需求复苏缓慢，因此存储原厂积极调整产能，在有限的资本支出下，将旧制程产线升级至先进制程，并将先进产能倾向于高利润的DDR5和HBM。

表6：存储原厂HBM产能调整

厂商	事件
SK海力士	SK海力士计划通过将 HBM3 和 DDR4 DRAM 等现有工艺转换为更先进的制程工艺来应对HBM3E快速增长的需求。
	2024年9月，SK海力士将其位于韩国利川工厂的M10F DRAM产线改造为生产HBM3E，2025年月产能将达到15万/月。
	SK海力士计划在今年3月份开工建设龙仁集群的首座厂房，生产以HBM为代表的面向AI的存储器和新一代DRAM产品。
	2024年以来，SK海力士正在推进5代1b DRAM的增产工作，涉及利川的M16和清州市的M15X。
	为了巩固 HBM 存储芯片市场上的领先地位，SK海力士2024 年投资超过 10 亿美元，扩大在韩国的测试和封装能力。
三星	三星电子将从未来三年内在天安第三综合工业园区三星显示器的28万平方米场地上安装半导体封装加工设备以生产HBM。
	三星电子自2024年以来一直对 HBM 进行大规模投资，2024年底HBM最大产能扩大至17万片/月。
美光	2025年1月8日，美光在其位于新加坡现有工厂旁边开始建造一个新的高带宽存储器 (HBM) 先进封装工厂，计划2026年运营。
	2024年6月，美光在其位于美国爱达荷州博伊西的总部扩建与HBM相关的研发生产设施，此外，美光还考虑在马来西亚建设HBM生产能力，预计产能建设将集中在后端工艺部分。

来源：公开信息



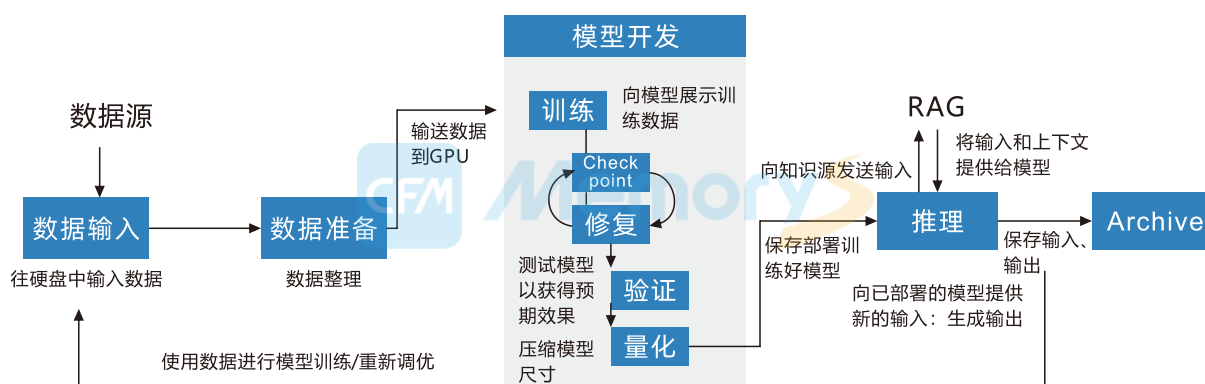
### 三、企业级PCIe 5.0 SSD需求呈现快速增长

近两年AI基础设施建设如火如荼，动辄成千上万的GPU算力集群得到了普遍应用。然而，在高吞吐量、低延迟网络和深度学习模型环境下，如何能更加快速的向算力“巨兽”提供更多的数据，成为存储供应商需要解决的首要难题。

无论是云端或是边端，数据检索和调用在AI应用中越来越频繁，温热数据的抓取、生成和存储的频率显著增加。而擅于存储冷数据的HDD并不擅长处理大数据频繁访问的AI应用，由于HDD的每秒输出操作次数（IOPS）仅为数百次，存在显而易见的性能瓶颈，严重阻碍了AI调用数据的效率。在AI技术突破和应用浪潮下，HDD正在加速退出历史舞台。而SSD的IOPS可以达到数万次，据有效统计，在AI训练和推理阶段，SSD的随机读取性能比HDD高出4500倍，顺序读取能力高约10倍。SSD凭借更快的启动和响应速度，实时更新数据和模型微调结果，为AI高效运行提供了必要的读写性能和低延迟性能。

典型的AI大模型训练和推理过程主要分为五个阶段：数据输入、数据准备、模型训练开发、模型推理部署和归档，了解每个阶段所需的存储特征，可更好地提供针对性的存储解决方案。

图10：AI训练大模型流程图



来源：公开信息

随着模型越来越复杂，训练数据集越来越大，AI工作负载将对存储的需求进一步提高。具体来看，AI训练阶段涉及庞大的数据集和复杂的计算，要求存储系统具备高速读写能力和超大容量。而AI推理阶段虽然数据读取频率高于写入，但同样需要高效的数据访问速度来支持实时决策。

表7：AI运行阶段存储需求

阶段	I/O特性	存储需求	影响
数据输入	海量顺序写入	高顺序写入吞吐量	优化存储意味着数据输入过程更加快速
数据准备	随机读取数据； 按顺序写入预处理项	小规模随机读取低延迟； 高顺序写入吞吐量	优化存储意味着在模型训练过程中能提供更多的数据，使模型准确性提升
模型训练	随机数据读取	多任务性能和容量可扩展性； 优化随机读取；适用于检查点写入的高顺序写入性能	优化存储可以提升昂贵训练资源（GPU,TPU,CPU）利用率
模型部署	混合随机读取和写入	处理组件故障的自修复能力； 不间断的扩展和升级；如果模型持续微调，则应具备与训练过程相同的特征	端侧需要高可用性、良好的服务能力和可靠性
存档	顺序及随机写入	高写入性能	为了合规和审计目的，需要更好的数据保留

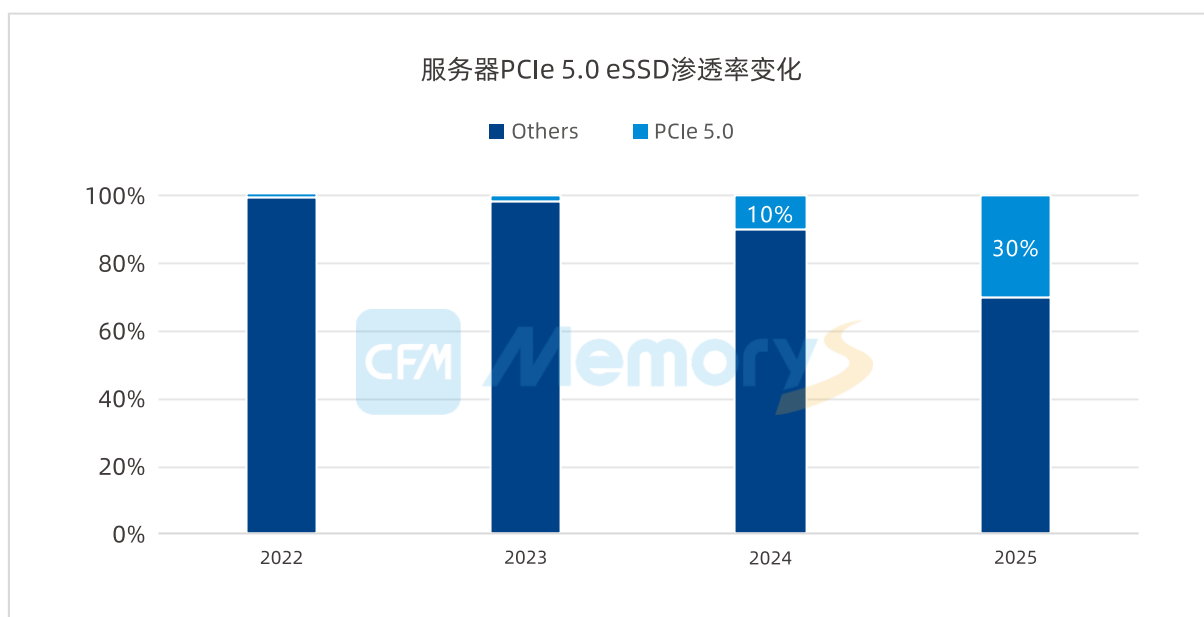
来源：公开信息，CFM闪存市场

为确保模型训练的准确度，所需的数据量不断增长，并且数据的来源已经扩展到文本、音频、图像、视频等多模态。此外，在AI大模型的训练和部署中会频繁使用到Checkpoint检查点环节，以定期保存模型的中间状态，包括模型的权重、偏置、优化器状态等等。Checkpoint能够提高训练效率、增加模型的可靠性，并帮助开发人员进行实验和调优。大模型的Checkpoint通常非常大，如在混合精度训练下GPT-175B模型的一个Checkpoint可达数百GB甚至数TB。而训练过程通常会并发多个Checkpoint，大量的数据需要快速地从存储设备传输到内存或计算单元中，以尽快恢复程序或任务的运行状态。

另外，训练期间的主要工作负载主要由随机读取组成，在触发check point时偶尔会被大量写入打断，因此AI模型训练中，存储需要满足训练负载所需的密集随机访问，即便在检查点写入期间。这种类似“防灾备份”的步骤对模型分析非常有必要，由于在训练过程中累积了更多实时数据，AI服务器对存储设备的性能、容量和可扩展性也提出了更高的要求。

通过对比PCIe 5.0 SSD（以Kioxia CM7为例）与PCIe 4.0 SSD，结果显示，使用高性能的PCIe 5.0 SSD时，机械学习训练时间得到显著减少。以16个GPU并行训练一年为例，使用PCIe 5.0 SSD可以减少约32天的训练时间。而在固定时间内，PCIe 5.0 SSD能够更快地训练多个模型，加速了AI模型的部署。因此，在AI训练场景中，更高性能低延迟的PCIe 5.0 eSSD应用得到青睐。据CFM数据显示，2024年服务器市场中PCIe 5.0 SSD市场份额约为10%，预估2025年该比例有望大幅增长至30%。

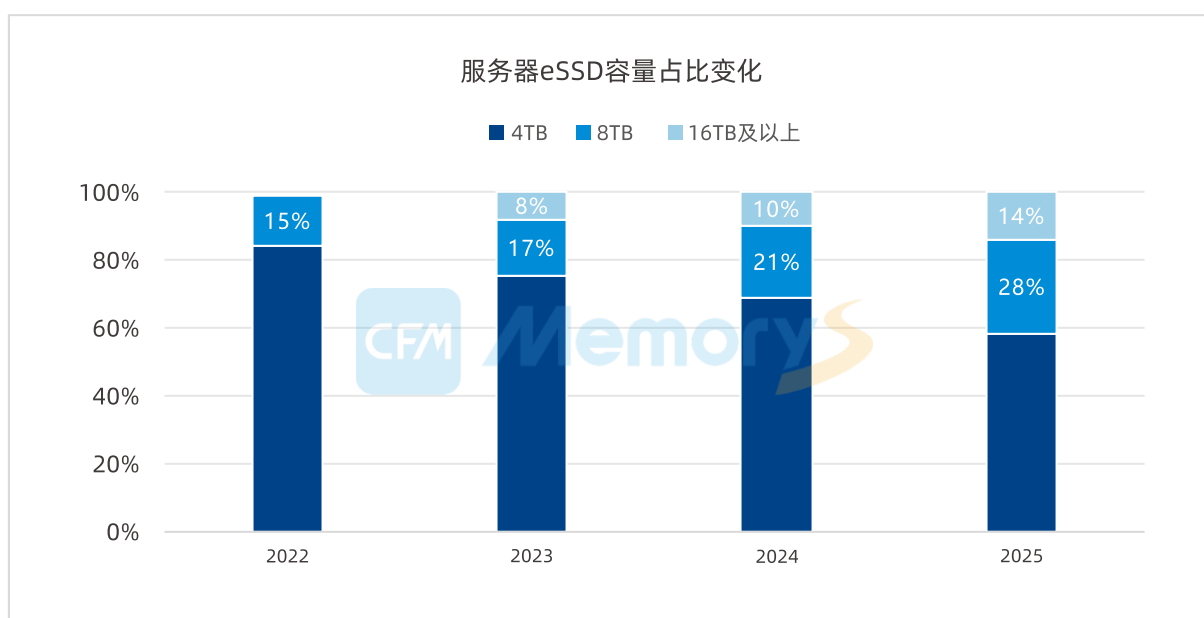
图11：服务器PCIe 5.0 eSSD渗透率变化



数据来源：CFM闪存市场

在eSSD容量方面，8TB/16TB及以上中高容量eSSD应用占比大幅提升。据CFM数据显示，2024年服务器市场8TB eSSD应用占比约21%，16TB及以上大容量SSD占比约10%，预计2025年8TB及以上容量eSSD份额占比将扩大至42%。

图12：服务器eSSD容量占比变化



数据来源：CFM闪存市场

#### 四、QLC NAND技术发展进入成熟期，带动SSD进入100TB时代

为满足AI训练的数据存储需求，自2024年以来，兼顾大容量、低功耗、高性能的QLC SSD成为企业级存储市场的新星，SSD顺势迎来100 TB级别的超高容量时代。存储厂商积极推动超大容量SSD产品研发和应用，Solidigm已经推出122TB SSD，预计在2025年初开始供应。三星电子16TB和32TB QLC SSD已大规模量产，增加64TB和128TB QLC SSD产品线，2024至2026年之间将推出256TB SSD，2027至2029年推出512TB SSD。西部数据128TB企业级SSD采用218层BiCS8 QLC NAND，计划在2027财年推出256TB SSD。另外，长江存储的X3-6070 QLC闪存颗粒已实现4000次PE循环，标志着QLC NAND技术发展进入成熟期，加速QLC NAND在企业级和消费类SSD及嵌入式等全场景应用。

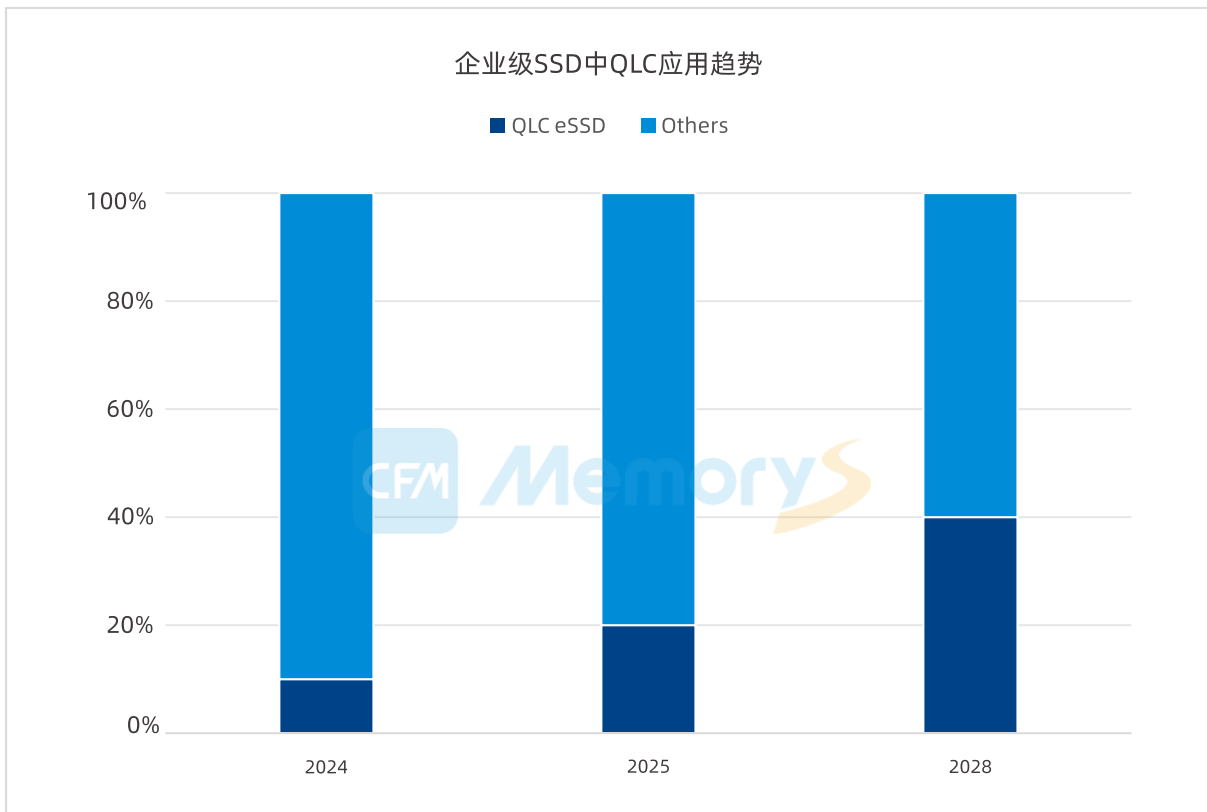
表8：部分QLC SSD产品

厂商	型号	容量	接口	颗粒类型	外形规格
Solidigm	D5-P5430	3.84TB/7.68TB/ 15.36TB/30.72TB	NVMe PCIe 4.0x4	192L QLC	U.2 15mm/E3.S 7.5mm/E1.S 9.5mm
	D5-P5536	7.68TB/15.36TB/ 30.72TB/61.44TB	NVMe PCIe 4.0x4	192L QLC	U.2 15mm/E3.S 7.5mm/E1.L 9.5mm
	D5-P5316	15.36TB/30.72TB	NVMe PCIe 4.0x4	144L QLC	E1.L/U.2 15mm
三星	BM-1743	最高61.44TB	NVMe PCIe 4.0/PCIe 5.0	V7 QLC	包含U.2和E3.S 两种外形规格
美光	7450系列	最高15.36TB	NVMe PCIe 4.0	QLC	2.5英寸 /M.2/E1.S
	7400系列	最高7.68TB	NVMe PCIe 4.0	QLC	2.5英寸/M.2
群联	D205V	最高122.88TB	NVMe PCIe 5.0	QLC	U.2/E3.S/ E3.L
	D200V	最高122.88TB	NVMe PCIe 5.0	QLC	U.2/E3
大普微	J5060系列	最高122.88TB	NVMe PCIe 4.0	QLC	U.2 15mm
忆恒创源	PBlaze7 7340	最高15.36TB	PCIe 5.0	X3-6070 QLC	NA

数据来源：CFM闪存市场，不完全统计

关于QLC技术的分析，CFM在《2023-2024年全球存储市场趋势白皮书》中有过详细的介绍。在QLC SSD取代HDD的进程中，成本已经是最后关卡，而在AI大模型训练场景中，效率、性能与功耗是系统厂商们最为关注的特性。QLC SSD兼具HDD大容量特性，同时以更高的存储密度和低功耗，为超大规模算力中心优化服务器物理空间，并节省大量电力成本，满足绿色环保可持续发展的AI应用趋势。

图13：企业级SSD中QLC应用趋势



数据来源：CFM闪存市场

## 五、开源AI大模型加速推动AI与终端应用的融合

2025开年，“杀手级”AI应用DeepSeek的成功引发全球关注，AI技术突破带来的成本骤降以及开源模式下的巨大潜力，让人们看到了AI普惠以及更多AI应用场景落地的前景。同时，也削弱了此前仅能通过超高算力资源堆砌建立起的算力“护城河”，越来越多厂商意识到AI发展不是单纯的算力竞赛，算法创新、商业模式和生态建设等全方位拓展延伸，成为AI全面发展的重要方向。

与竞相构建越来越庞大参数量的LLM不同，DeepSeek R1所代表的推理模型中采用的混合专家架构（MoE），在训练小体量的模型上，其效率、成本等方面都更具优势。从训练越来越大的通用型LLM，转向训练从大规模LLM中提炼出来的较小参数的推理模型，并使用增强学习技术增强AI在专业领域的专家能力。除此以外，2025年2月底，DeepSeek开源了五个代码库，其中开源的FlashMLA针对可变长度序列进行了优化，在H800上可以实现每秒处理3000GB的数据，每秒执行580万亿次浮点运算，所需的硬件门槛将大幅降低。

因此，DeepSeek模式的成功可能会令科技公司重估ROI和单位成本，部分大型云服务商可能短期会放缓对顶尖加速卡的追捧，取而代之的是将更多中低性能的GPU等闲置算力资源利用起来。短期来看，DeepSeek通过算法创新进一步榨干算力，大幅节约算力并提升模型效率，一定程度上会抑制单位算力需求，然而长期来看，低成本AI训练和开源模式的持续发酵，对AI应用将产生深远的影响。更低的AI使用成本将促使更多企业和个人采用AI技术，这不仅有利于为云服务商带来增量的云需求，同时也促进了端侧AI应用的新增需求进场，利于AI在端侧的应用普及。因此CFM闪存市场认为，凡是能够推动AI技术向前发展都是积极的，AI计算效率的提升虽然短期压制单位算力需求，但长期来看，AI需求的爆发和应用普及将令智能终端的存算力需求不减反增。整体来看，DeepSeek以“开源+算法优化+端侧革命”的三重势能，正在重塑AI行业的底层逻辑，然而AI的发展是极其迅速的，仍需密切关注AI技术的发展动向。

### 部分科技企业对DeepSeek的看法

三星电子：因为其仍向多个客户提供HBM，也在密切关注和监控基于不同场景的行业发展。随着新技术的采用，行业动态总是会发生变化。尽管根据目前掌握的有限信息，现在下结论还为时过早，但确实预计市场将在短期出现风险但仍保持长期的机遇。

苹果CEO蒂姆·库克认为，DeepSeek的人工智能模型代表了“推动效率的创新”。“总的来说，库克认为推动效率的创新是一件好事。”

Meta CEO马克·扎克伯格表示，DeepSeek有很多值得学习的地方，但对于该应用对于AI未来的意义，现在要“真正坚定地作出判断”还为时过早。扎克伯格坦言称，DeepSeek做了一些“新奇的事情”，公司“仍在消化DeepSeek的一些成果”，团队希望能够将其中一些进步应用到自己的AI项目中。

谷歌CEO桑达尔·皮查伊（Sundar Pichai）赞扬了DeepSeek的工作，他认为DeepSeek有一支了不起的团队。

微软 董事长兼CEO萨蒂亚·纳德拉（Satya Nadella）表示，DeepSeek-R1模型目前已可通过微软的AI平台Azure AI Foundry和GitHub获取，并且很快就能在Copilot+电脑上运行。纳德拉认为，DeepSeek“有一些真的创新”，AI成本下降是趋势。

谷歌CEO桑达尔·皮查伊(Sundar Pichai)表示，在全球范围内更具性价比的AI大模型的出现将进一步推动这项技术的采用率，有利于谷歌以及整个行业的发展。未来谷歌也将进一步寻求降低每次查询成本，2025年将是迄今为止搜索创新幅度最大的一年之一。

高通 首席执行官 克里斯蒂安诺·阿蒙表示，DeepSeek-R1的推出是AI产业的一个转折点。AI大模型的预训练将在云端继续，推理将会向端侧迁移。

AMD CEO苏姿丰表示，对于DeepSeek，AMD认为在模型和算法上的创新有利于AI的采用。有新的方法以更少的基础设施来实现训练和推理能力，这实际上是一件好事，因为它使我们能够继续部署AI计算和更广泛的应用领域以及更多的采用。

ARM 也在电话会议表示DeepSeek在其他大模型的基础上做了许多创新，提高了推理的效率，还降低了成本，推动了AI应用场景的多样化，对算力的需求持续增长。

AI is the future，已经成为业界不可置否的共识，然而，到达AGI这一“未知之境”的旅途注定是迂回和波折的，万里长征路中，我们仍处于早期阶段，任何有营养的进步与创新都值得学习，在未知领域探索中，“让子弹再飞一会儿”会显得更加理智。退一步而言，即便市场转变，资源需求也仅是结构性转变，对算力这一硬通货的需求仍然坚挺，这可能也是各大科技厂商维持资本支出计划的根本原因。

随着人工智能的广泛应用，对硬件技术提出了前所未有的挑战和机遇，为了满足AI在数据处理、模型训练、实时响应等方面的高要求，硬件技术创新成为关键。这种以AI需求为导向的硬件创新，不仅推动了硬件技术的快速演进，还促进了整个AI产业生态的繁荣。毫无疑问，AI正在驱动存储在内的硬件进行全面的革新。

## 六、服务器新型存储器LPCAMM/SOCAMM和MRDIMM/MCRDIMM

在AI应用场景中，海量数据并行处理，传输速率的提升确实至关重要，但是引发的能耗问题愈发凸显。市场上除了对像HBM这种不计功耗、只求性能的产品需求之外，对低功耗产品需求也日益增长。与DDR产品相比，LPDDR能够在消耗相对较低的功耗的同时实现高性能，是一种能够应对计算趋势变化的、极具吸引力的解决方案。

近期，关于英伟达下一代重磅产品GB300的最新消息引发了供应链广泛关注。此前GB200产品时期，英伟达提供包括Blackwell GPU、Grace CPU、512GB的LPDDR5X内存、VRM内容等在内的完整Bianca板。但是从GB300开始，英伟达将只提供“SXM Puck”模块上的B300 GPU、BGA封装上的Grace CPU以及交换托盘和铜背板。这就意味着，OEM和ODM将有更多的机会参与计算模块的制造。

其中，在存储部分，在GB200时代，英伟达选择了直接焊接在Bianca板上的512GB LPDDR5X作为主存方案。然而，在GB300的设计中，为了提高灵活性并降低成本，英伟达决定采用可插拔式的LPCAMM模块来替代传统的焊接内存，不仅简化了制造过程，还允许用户根据具体需求灵活配置内存容量和类型。此外，通过减少定制化硬件的需求，降低了整体生产成本，同时也为客户提供了一个更具性价比的选择。

LPCAMM为低功耗压缩附加内存模组，原本是专为轻薄型笔记本和其他紧凑型设备设计的新一代内存，执行JEDEC标准。由于是独立模块，使用户在升级、维修方面提供了更高的灵活性。如今，LPCAMM开始进入服务器应用，并有业界专家大胆预言其甚至有可能取代传统DIMM，由此可见，LPCAMM技术发展前景之广阔。

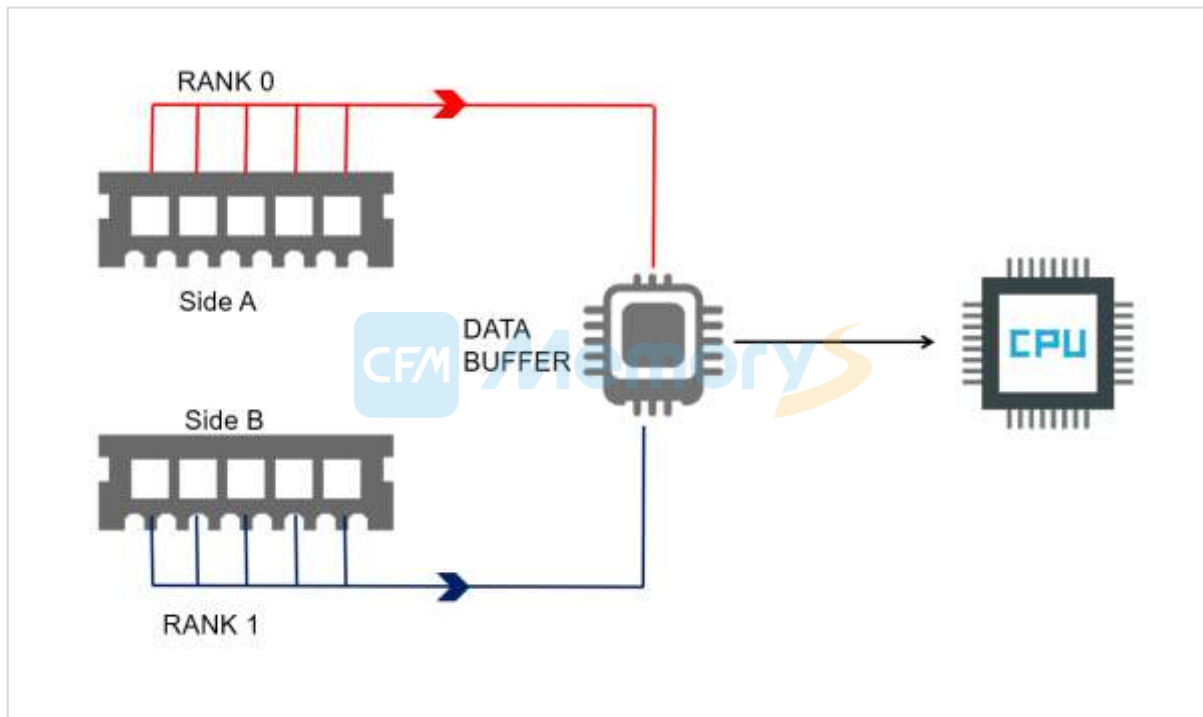
另一方面，英伟达为了解决数据中心空间限制，以及对更高密度内存解决方案的要求，提出了SOCAMM（Space-Optimized CAMM即空间优化型压缩附加内存模组），在存储密度提升、信号传输优化、电源管理创新和散热技术突破等多方面颇具潜力。目前，SOCAMM还处于概念推广和初步应用探索阶段，随着技术的不断成熟，未来有望在超大规模数据中心、边缘计算设备和便携式AI终端进行大规模应用。

MRDIMM（Multiplexer Ranks DIMM），另一种叫法为MCRDIMM（Multiplexer Combined Ranks DIMM）均为多路复用双列直插内存模组，能够提供更高的内存密度和带宽，是一种高性能内存互连解决方案，同样也是为了解决现有计算架构“内存墙”瓶颈而生。

传统RDIMM虽然允许多个内存阵列同时工作，但无法实现真正的并行访问，导致内存带宽无法得到充分利用。MRDIMM技术使用高速多路复用器或数据缓冲器（MDB）来同时读取内存组并将数据传输到CPU完美解决这一问题。目前第一代MRDIMM的速度最高可达8,800 MT/s，预计未来第二代、第三代产品速度可达12,800 MT/s、17,600 MT/s。



图14: MRDIMM示意图



## 第三章 消费类存储产品应用与发展分析

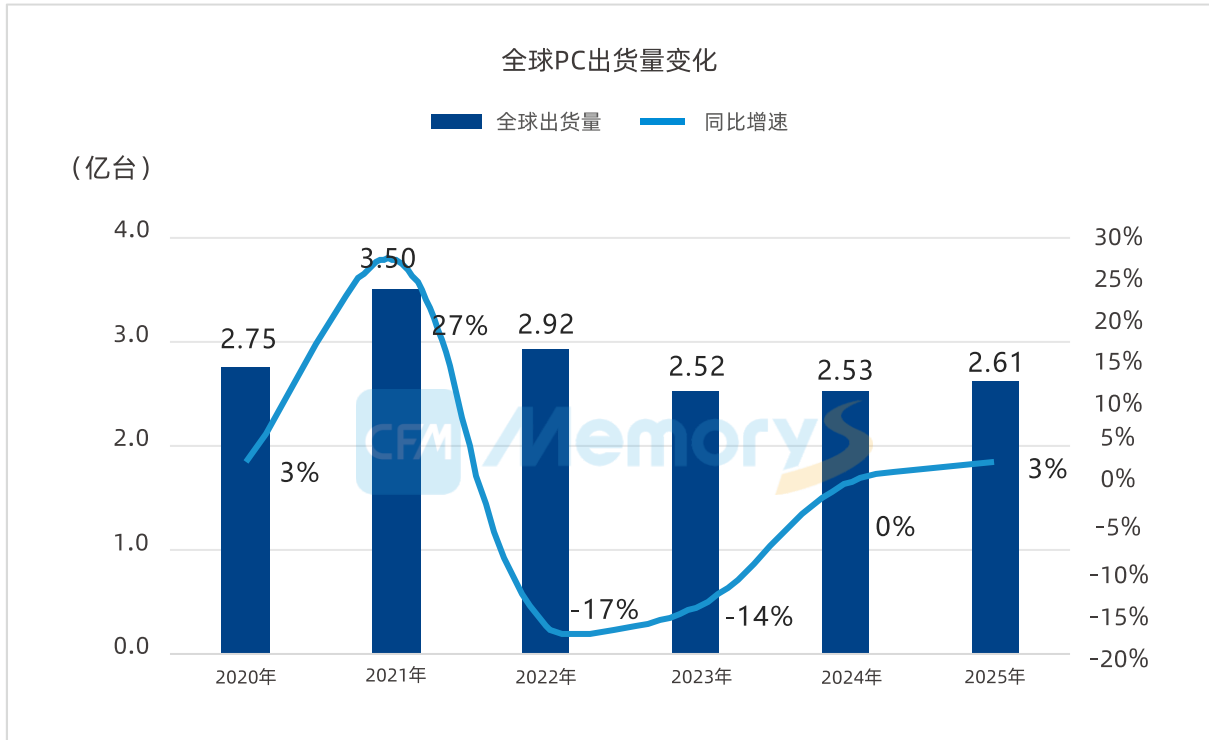
APPLICATION AND DEVELOPMENT ANALYSIS OF CONSUMER MEMORY PRODUCTS

### 一、PC设备的存储应用与发展分析

#### 1、2025年全球PC出货小幅增长，AI PC渗透率攀升

随着2020年年初疫情的爆发，2020-2021年期间，各地掀起了线上教育和远程办公热潮，带动PC市场呈现强劲增长态势。而在经济环境恶化、通货膨胀、疫情影响逐渐消退等因素的催化下，后疫情时代下消费者普遍已拥有较新的PC设备，使得2022年PC市场迅速开启“过冬”模式，连续两年出货量大幅下滑。“至暗”时刻过后的2024年本该迎来可观的增长势头，而PC市场如同按下了“暂停键”，复苏力度非常微弱，基本与2023年持平，据CFM闪存市场分析，预计2025年全球PC出货量将同比小幅增长3%至2.61亿台。

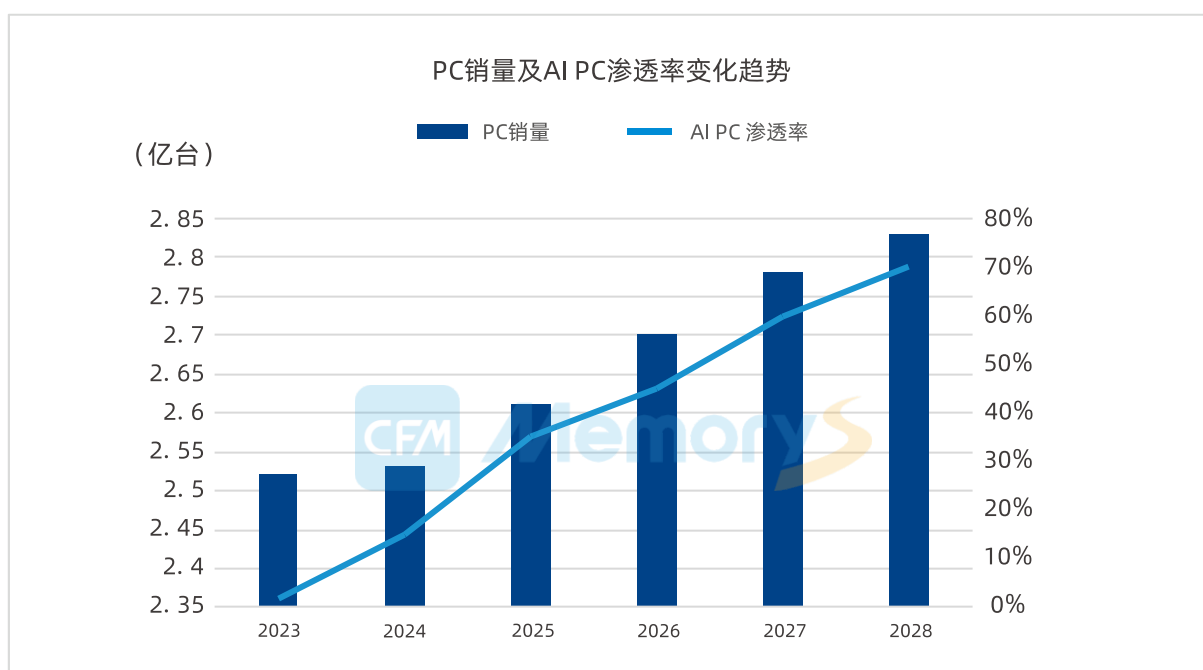
图15：全球PC出货量变化



数据来源：CFM闪存市场

2024以来，微软新推出Copilot+PC，高通、AMD、英特尔等的新一代处理器也陆续发布，正推动着各大PC厂商积极探索全新的AI PC形态。随着苹果、联想、惠普、戴尔、华硕、宏碁、华为等PC厂商的加速布局AI PC，将驱动AI PC的渗透率进一步提升。根据CFM闪存市场数据显示，预计2025年全球AI PC渗透率将达到35%。

图16：PC销量及AI PC渗透率变化趋势



数据来源：CFM闪存市场

相对于传统PC来说，由于AI PC采用的是CPU+GPU+NPU异构方案，新增的NPU令存储容量逐渐提升，功耗和散热要求更加严苛。从微软为Copilot+PC所设定的最低配置要求可看出，目前支持的处理器包括骁龙X Plus和X Elite、AMD Ryzen AI 300系列以及Intel Core Ultra 200V系列，Copilot+PC必须配备一个能够执行40万亿次运算每秒（TOPS）以上的NPU，以支持AI功能的实时处理和计算。除了CPU、GPU和NPU之外，Copilot+PC对内存和存储空间也提出了更高的要求。其中，内存至少需要16GB的DDR5或LPDDR5X内存，而存储容量的最低要求为256GB SSD或UFS存储。为了满足本地化AI模型运行的需求，更高的存储容量和带宽是大势所趋。

在近期推出的AI PC新品中，以联想YOGA Air 14s骁龙AI元启、惠普EliteBook Ultra 14、灵耀16 Air、宏碁Swift 16 AI、苹果MacBook Pro 14英寸为例，除苹果以外，AI PC普遍搭载新一代处理器骁龙X Plus/X Elite、AMD Ryzen AI 300系列以及酷睿Ultra 200V系列，内存配置32GB LPDDR5X，存储容量1TB起步。而苹果MacBook Pro 14英寸也配备苹果最新处理器Apple M4芯片。

表9：部分AI PC产品硬件参数对比

产品名称	联想YOGA Air 14s 骁龙 AI元启	惠普EliteBook Ultra 14	灵耀16 Air	宏碁Swift 16 AI	苹果MacBook Pro 14英寸
CPU	高通骁龙X Plus	酷睿Ultra 9 288V	AMD Ryzen AI 9 HX 370	Intel Core Ultra 9 288V	Apple M4 芯片
内存类型	32GB LPDDR5X 8448MHz	32GB LPDDR5X 8533MHz	32GB LPDDR5X 7500MHz	32GB LPDDR5X 8533MHz	最高32GB
硬盘容量	1TB PCIe 4.0 SSD	1TB PCIe 4.0 SSD	1TB PCIe 4.0 SSD	2 TB PCIe 4.0 SSD	最高2TB SSD
屏幕分辨率	2944×1840	2880×1800	2880×1800	2880 x 1800	3024×1964
显卡类型	Adreno GPU集成显卡	英特尔锐炫 140V核显	AMD Radeon Graphics核显	英特尔锐炫 140V核显	Apple核显
网络	Wi-Fi 7	Wi-Fi 7	Wi-Fi 7	Wi-Fi 7	Wi-Fi 6E
厚度/重量	12.9mm/1.29kg	14.9mm/1.35kg	11mm/1.49kg	1.5kg	22.12 mm/1.55kg
电池容量	70Whr	64Whr	78Whr	75Whr	72.4Whr
散热系统	双风扇、双热管	双风扇、双热管	双风扇加均热板	/	/

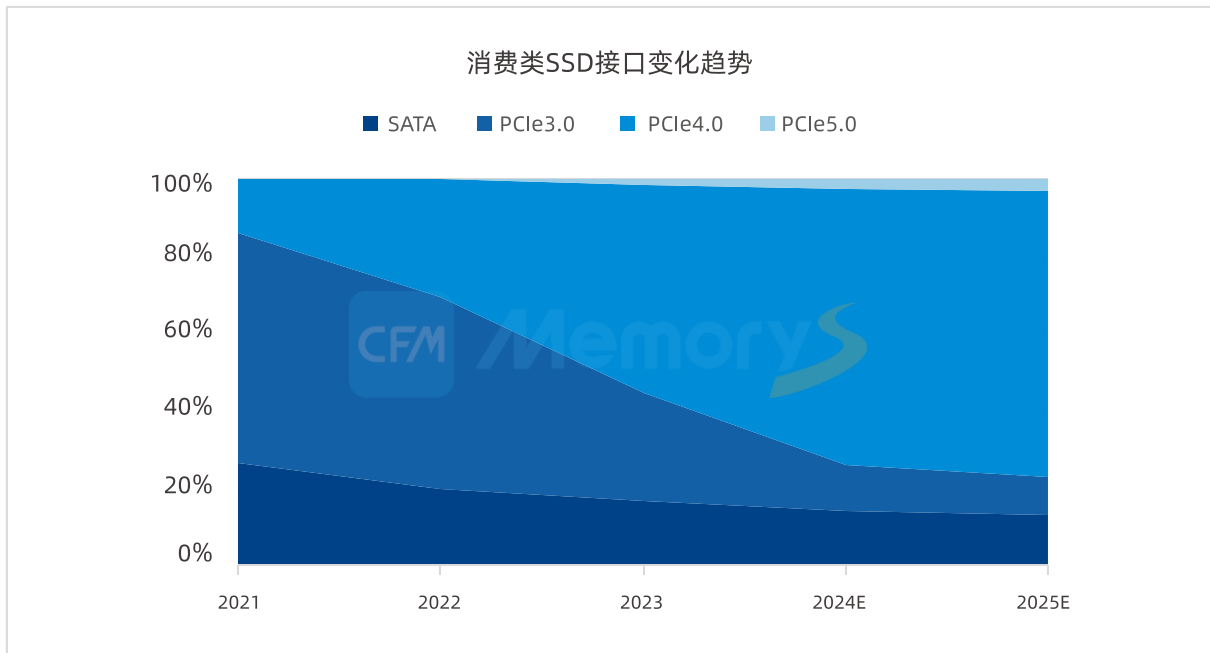
来源：公开信息，CFM闪存市场

近期，随着DeepSeek的热度持续攀升，端侧AI领域掀起了一股本地部署大模型的热潮。1月30日，微软发文宣布为Windows11用户优化DeepSeek R1模型，首批将面向搭载高通骁龙X芯片的Copilot+ PC，而后是酷睿Ultra 200V等平台。英特尔则称，DeepSeek能在英特尔产品上运行，在AI PC上还能离线使用。联想亦宣布个人智能体“小天”已接入DeepSeek，目前，用户已经能够在联想AI PC系列产品之上直接使用DeepSeek。

## 2、PCIe 4.0 SSD仍是主流，主控厂商助力PCIe 5.0 SSD加速发展

2022年消费类PCIe 5.0 SSD首次亮相于人前，酝酿了三年之久，虽在读写性能上比PCIe 4.0 SSD有显著提升，但由于受制于发热问题、价格高昂等未能广泛受到消费者青睐。截至目前，PCIe 4.0 SSD仍是市场主流，2024年起PCIe 4.0 SSD出货占比超过70%。不过，随着存储厂商在消费类PCIe 5.0 SSD的功耗管理上做出显著改进，突破10000MB/s读写性能以及更低功耗的PCIe 5.0主控加持，譬如慧荣SM2508，联芸MAP1806、群联PS5028-E28等，结合AI PC的发展趋势，PCIe 5.0 SSD将得到更广泛的应用。

图17：消费类SSD接口变化趋势



数据来源：CFM闪存市场

表10：部分存储原厂PCIe 5.0 SSD产品对比

企业	三星	美光	SK海力士	铠侠	长江存储
产品名称	PM9E1 PC SSD	Crucial P510 SSD	Platinum P51 SSD	Exceria Plus G4 SSD	致态TiPro 9000
外形	M.2 2280	M.2 2280	M.2 2280	M.2 2280	M.2 2280
主控芯片	Samsung 5nm主控	群联 PS5031-31T 无缓主控	SK hynix Alistar 主控	群联 E31T 系列 无缓主控	慧荣SM2508
NAND Flash	samsung V8 236L 3D TLC NAND Flash	micron G9 276L 3D TLC NAND Flash	SK hynix 238L 3D TLC NAND Flash	218层 BiCS 8 闪存	Xtacking®4.0 3D TLC NAND Flash
DRAM	/	/	SK hynix LPDDR4	/	1GB/2GB LPDDR4
容量	512GB、1TB、2TB、4TB	1TB、2TB	500GB、1TB、2TB	1TB、2TB	1TB、2TB
顺序读取速度	14.5GB/s	11,000 MB/s	14.7GB/s	10000MB/s	14000MB/s
顺序写入速度	13GB/s	9,550 MB/s	13.4GB/s	1TB: 7,900 MB/s; 2TB: 8,200MB/s	12500MB/s
随机读取速度	/	/	2300K IOPS	1300K IOPS	2000K IOPS (2TB版本)
随机写入速度	/	/	2400K IOPS	1400K IOPS	1600K IOPS (2TB版本)
功耗	5W	功耗比前一代降低25%	10W	5.3W	/
耐久度	最大写入量 2400TBW	/	500GB: 300TBW; 1TB: 600TBW; 2TB: 1200TBW	1TB: 600TBW; 2TB: 1200TBW	1TB: 600TBW; 2TB: 1200TBW

来源：公开信息，CFM闪存市场

近两年来，PCIe 5.0 SSD的DRAM less（不带独立缓存）方案正逐渐被一些对于无需在较长运行时间内保持高性能运作的用户作为替代方案，此方案比带独立缓存SSD的价格较为亲民，更低功耗和更大的存储空间，DRAM-less SSD在性能和寿命上的缺陷可通过优化和改进固件和算法等方法来缓解。对于应用在PC上的SSD，DRAM-less SSD是兼顾低功耗和高性价比的选择。上表中美光Crucial P510 SSD和铠侠Exceria Plus G4 SSD均为DRAM less方案。

### 3、国产PC产业链锋芒毕露，2025年信创PC市场充满潜力

近年来，在国产替代持续推进下，国内PC市场见证了诸如联想开天、华为、同方、浪潮计算机、中科可控、雷神科技、神舟电脑、中国长城等一系列本土品牌的迅速崛起与蓬勃发展。根据CFM闪存市场数据显示，2024年国内信创PC出货量超350万台，预计2025年出货量将突破500万台。而从市场份额看，目前信创PC体量较小，占PC整体出货量仍在个位数。随着国产PC生态建设不断完善，上游软硬件技术不断创新，国补政策持续发力，未来信创PC渗透率将不断提高。

在政府消费补贴、以旧换新等多项优惠加持下，2024年双11国产PC品牌打出了一场“漂亮仗”。其中，联想全网销售额突破86亿大关，在京东电脑数码竞速榜中的笔记本、游戏本、轻薄本等多个品类中位列第一，而机械革命、七彩虹、华为、雷神、小米、神舟等大陆品牌均在TOP10榜单中。

表11：2024年双11京东电脑数码竞速榜

2024年双11京东电脑数码竞速榜（11月10日20:00-11月11日24:00）		
序号	总榜（POP销额）	游戏本榜（自营销量）
1	联想	联想
2	ThinkPad	机械革命
3	华硕	华硕
4	机械革命	惠普
5	惠普	七彩虹
6	华为	神舟
7	七彩虹	小米
8	戴尔	微星
9	ROG	雷神
10	HAEWI	ROG

来源：公开资料、CFM闪存市场

当然，国内PC厂商除了深耕本土市场，也在积极通过收购、建立全球研发生产基地、成立合资公司、参展等方式走向国际。联想、机械革命、雷神等国产PC厂商在AI PC和电竞游戏本等领域大展拳脚，国产PC品牌正迎来“出海”浪潮，积极开拓东南亚、欧美、中东非等海外消费者市场。

在信息化和数字化飞速发展的今天，数据安全与可控性议题愈发受到关注，而PC上游硬件和软件研发作为技术壁垒最高的环节，随着国内半导体企业的迅速崛起正逐渐被层层攻破。其中，国内CPU技术路线x86、ARM、MIPS、Alpha等百花齐放，申威、龙芯、鲲鹏、飞腾、兆芯、海光等六大国产CPU为主力军。而长江存储和长鑫存储则作为国产存储芯片的“双星”，在PC半导体供应链中已占据一席之地。

近年来，长江存储在企业用户和个人用户中两手抓，其中，面向企业用户的商用消费级PCIe NVMe固态硬盘包括PC300、PC411、PC41Q等，目前，PC411已在机械革命耀世15 Pro、机械革命无界14 Pro、机械师曙光16等国产PC设备中陆续商用，最大顺序读取写入速度高达7000/6000MB/s，4K随机读写速度达850K/700K IOPS。而面向个人用户的零售品牌——“致态”更是在2024年“双十一”购物节上，首次荣获京东SSD品类（固态硬盘）“双十一大促”交易总额（GMV）及销量双料冠军。值得注意的是，致态也是首次摘得京东SSD品类“双十一”购物节双料冠军的国内存储品牌。具体来看，致态在京东平台的交易总额同比增长43%，总销量同比增长16%。其中，致态TiPlus7100成为京东平台最受欢迎的爆款SSD产品之一。

表12：2024年双11京东电脑数码竞速榜-自营SSD

2024年双11京东电脑数码竞速榜-SSD固态硬盘（自营）	
序号	品牌
1	致态
2	三星
3	宏碁掠夺者
4	梵想
5	金士顿
6	铠侠
7	宏碁
8	SK HYNIX
9	联想
10	雷克沙

来源：公开资料、CFM闪存市场

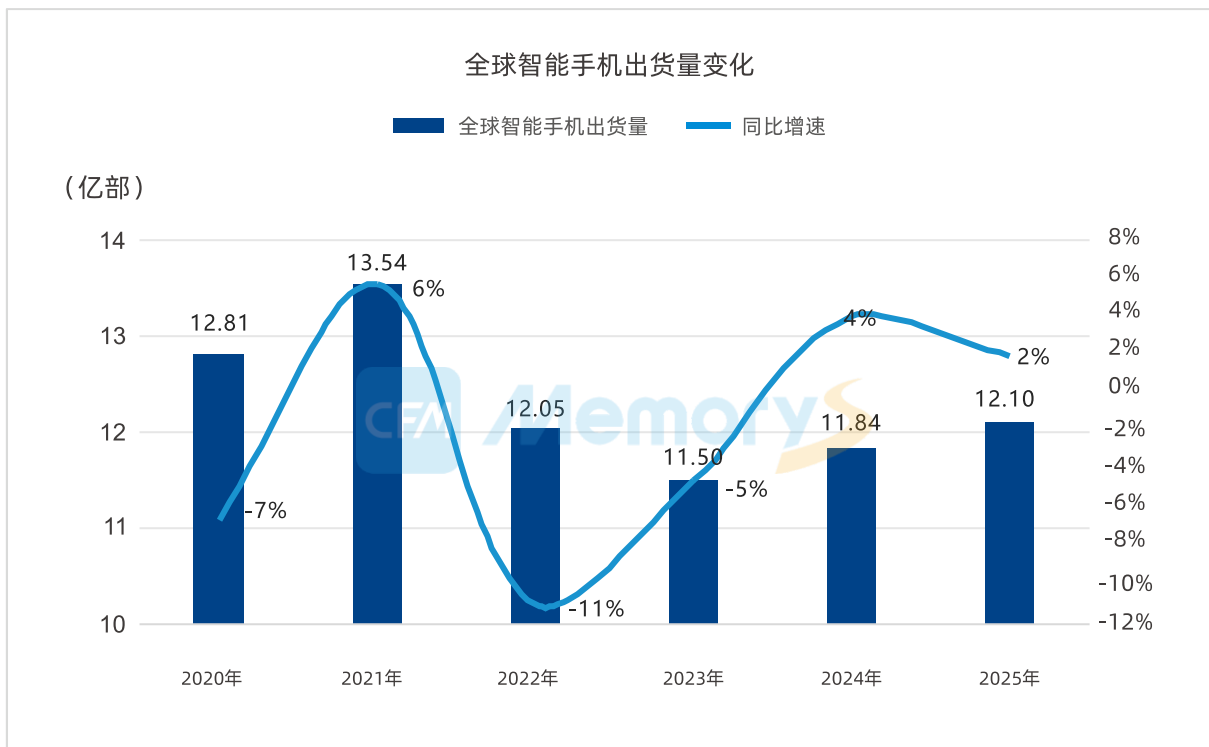


## 二、智能手机的存储应用与发展分析

### 1、全球智能手机销量温和增长，品牌竞争激烈

2024年虽说是智能手机的复苏年，但由于经济环境改善不及预期、消费者购机意愿相对保守、换机周期延长等因素，整体增长力度有限。据CFM数据显示，2024年全球智能手机出货量约达11.84亿部，同比增长4%，预计2025年全球智能手机出货量同比小幅上涨2%至12.10亿部。

图18：全球智能手机出货量变化



数据来源：CFM闪存市场

从国内几大手机品牌对于2025年的销量预估来看都偏保守，基本环比持平或低个位数微增。由于走量的仍是注重性价比的手机，且海外出货以中低端机型为主，国内手机品牌搭载的存储容量普遍以128GB/256GB为主。而在大容量机型的选择中，消费者对512GB容量的手机选择也在逐渐增加。

表13：部分品牌2025年手机出货量预计及主力容量

品牌	预计2025年出货量(百万台)	主力容量
小米	170	128GB/256GB
OPPO	150	128GB/256GB
VIVO	110	128GB/256GB
传音	109	64GB/128GB
荣耀	70	128GB/256GB
华为	60	128GB/256GB

来源：公开资料、CFM闪存市场

随着智能手机不断的技术迭代，消费者得以享受更加出色的硬件性能和流畅的操作系统，即便使用了长达两三年的时间仍然保持良好的运行状态，从而也促使消费者换机周期延长至40个月以上。而随着时间的推移，日积月累的数据信息占用过高的存储空间，容易导致手机运行效率低下，出现卡顿、延迟等问题。因此，人们手机使用周期延长，高存储容量也就成为消费者购机的关键因素。据CFM闪存市场数据显示，目前，虽然手机存储容量256GB及以下仍是主流，但2024年512GB市场占比已高达35%，这一比例在未来还将持续攀升。

## 2、智能手机通过AI技术加持进行创新，为用户带来新体验

AI手机正演变成智能贴心的私人助手，除了AI制图、撰写、生成视频等工具以外，AI手机能够模拟人类操作如点咖啡、订票以及一键问屏、一圈即搜、AI回邮件等功能，手机AI功能日渐丰富。集成AI大模型的AI手机助手，已成为智能手机中不可或缺的一部分。通过语音识别、自然语言处理和机器学习技术，在不依赖于任何第三方应用支持下可自主理解用户意图，捕捉和理解复杂的语音指令等信息，并通过一系列自主操作完成复杂任务。在健康管理、高效办公和便携生活支持等方面，AI手机更加注重提供个性化服务，还能根据用户的习惯和反馈进行自我迭代，令用户体验感进一步提升。据CFM闪存市场数据显示，2025年全球AI手机的渗透率将达到30%，并在2026年达到50%。

从各手机品牌来看，苹果、三星、OPPO、VIVO、华为、荣耀等品牌在新品中持续创新并引入AI技术，自研端侧大模型并量身定做AI语音助手。已上线Apple Intelligence的国家和地区，iPhone 16的销量同比增长显著，但并非在全球各地同步上线，其中也包括中国。2024财年，苹果大中华区营收同比下跌7.7%，其中，第四季度跌幅达11%。对此，苹果公司CEO库克坦言，iPhone缺乏AI功能是重要原因。为苹果扭转在华销量持续下滑的局面，近日，苹果宣布牵手阿里共同开发面向中国市场的AI功能。阿里的AI模型将深度集成至iOS系统，覆盖语音交互、图像生成、个性化推荐等功能模块，预计将在2025年4月随iOS 18.5版本更新向用户开放，首批适配机型为iPhone 16系列及后续新品。除了苹果，三星也宣布与国内头部大模型厂商智谱达成合作，将Agentic GLM应用于Galaxy S25系列手机。华为、荣耀、OPPO等国产手机厂商也纷纷宣布接入DeepSeek-R1，随着端侧AI的不断发展，海外手机厂商加速中国本地化布局，未来中国手机市场的竞争也将愈发激烈。

### 3、eMMC5.1市场需求趋于稳定，UFS3.1广泛覆盖5G机型，UFS4.0高端机型应用增加

据CFM调研显示，手机嵌入式存储需求中，eMMC5.1市场需求趋于稳定，UFS3.1广泛覆盖5G走量机型，市场需求稳定增长，高端及旗舰机型则广泛增加UFS4.0的应用。从入门级机型来看，eMMC5.1市场趋于稳定。尤其在海外市场，非洲、东南亚以及南美洲等地区仍处于4G通讯环境，搭载eMMC5.1和UFS2.2的入门级机型出货居多，人口增速较快的发展中国家和地区，是入门级手机主要的销售市场，支撑eMMC5.1市场需求整体保持稳定。从销量占比最大的5G中低端机型来看，5G SoC平台广泛支持UFS3.1，5G主力机型对UFS3.1的需求庞大，UFS3.1需求稳定增长。在高端及旗舰机型，终端基本全力应用UFS4.0，以满足持续增长的AI手机的存储需求。高性能UFS4.0凭借更高的传输速度和更低的功耗，显著提升智能手机的系统流畅度，提高应用启动速度和文件加载速度，并减少散热问题。同时，高密度NAND Flash能够在收敛的物理空间内存储更多数据，为多折叠、超薄化等机型设计提供更多的灵活性和可能性，长江存储UC420（UFS4.1）的厚度仅为0.8mm 进一步为折叠屏手机和超薄手机做了优化。基于高密度先进NAND Flash的UFS4.0及以上高速嵌入式存储产品，将更好地满足AI时代下智能手机发展的新需求。

高端旗舰机型方面，品牌手机厂商集体“打擂台”，高端SOC处理器+LPDDR5/5X +UFS4.0成为性能铁三角。从2024年底手机厂商发布的旗舰机标准版看，普遍配备高通骁龙8至尊版和天玑9400等最新一代处理器，采用台积电第二代3nm制程工艺，存储配置均搭载LPDDR5/5X+UFS4.0顶配，内存容量除苹果16搭载8GB DRAM以外，三星和国产旗舰机全体以12GB起跳，一加13内存容量最高达24GB。顶配旗舰机型容量普遍高达1TB，高端机型增加了UFS4.0的应用。多款手机配备超声波指纹，更多机型支持无线充电功能。由于处理器和存储芯片等硬件成本的提升，旗舰机普遍掀起了涨价潮，平均上涨100-

500元不等，带动手机平均售价上涨，加快手机品牌高端化进程，预计2025年中国智能手机的平均售价将突破4000元。

表14：主流手机旗舰标准版硬件参数对比

主流手机旗舰标准版对比				
品牌/机型	苹果16	三星Galaxy S25	华为mate70	小米15
系统	iOS 18、灵动岛	OneUI 7	HarmonyOS NEXT	澎湃OS 2.0、HyperCore
处理器	A18芯片	骁龙8至尊版For Galaxy	9010麒麟芯片	骁龙8至尊版
续航	3561mAh	4000mAh	5300mAh	5400mAh
充电	27W有线+15W无线	25W有线+15W无线	66W有线+50W无线	90W有线+50W无线
屏幕	6.1英寸直屏	6.2英寸直屏	6.7英寸直屏	6.36英寸超窄四等边直屏
影像	1200万前摄+4800万主摄1200W超广角(2倍光学变焦)	5000万像素主摄、1200万像素超广角1000万像素3X长焦	5000万像素超光变主摄、4000万像素超广角、1200万像素潜望式长焦、前摄1300万像素、150万多光谱通道红枫原色影像	5000万像素徕卡主摄、5000万像素徕卡60mm浮动长焦、5000万像素徕卡超广角、前摄3200万像素
解锁	Face ID	超声波指纹解锁、人脸识别	侧边指纹解锁、人脸识别	超声波指纹解锁、人脸识别
存储配置	LPDDR5+NVMe存储	LPDDR5X+UFS4.0	LPDDR5+UFS3.1	LPDDR5X+UFS4.0
存储容量/售价	8+128GB：5999元 8+256GB：6999元 8+512GB：8999元	12+256GB：6499元 12+512GB：7499元	12+256GB：5499元 12+512GB：5999元 12+1TB：6999元	12+256GB：4499元 12+512GB：4799元 16+512GB：4999元 16+1TB：5499元

来源：公开信息，CFM闪存市场

表15：主流手机旗舰标准版硬件参数对比（续上表）

主流手机旗舰标准版对比（续上表）					
机型	VIVO X200	OPPO Find X8	荣耀magic7	一加13	iQOO 13
系统	OriginOS 5.0、原子岛	ColorOS 15	MagicOs 9.0	ColorOS 15	OriginOS 5.0
处理器	天玑9400	天玑9400	骁龙8至尊版	骁龙8至尊版	骁龙8至尊版
续航	5800mAh	5630mAh	5650mAh	6000mAh	6150mAh
充电	90W有线(不支持无线)	80W有线+50W无线	100W有线+80W无线	120W有线快充	100W有线快充
屏幕	6.78英寸全等深微四曲屏	6.59英寸极窄边直屏	6.78英寸超窄直屏	6.82英寸全等深柔边直屏	6.82英寸直屏
影像	5000万像素超感光VCS仿生主摄、5000万像素超广角、5000万像素蔡司超级长焦、前摄3200万像素、100倍蔡司超清变焦	5000万像素大底广角、5000万像素哈苏人像、5000万像素超广角、前摄3200万像素哈苏联名、无影抓拍、AI千里长焦	5000万像素荣耀鹰眼主摄、5000万像素长焦、5000万像素超广角、前摄5000万像素	前摄3200万像素、5000万像素主摄像头、5000万像素长焦摄像头、5000万像素超广角摄像头	前摄3200万像素、5000万像素超感光VCS仿生旗舰主摄、5000万像素长焦镜头、5000万像素超广角镜头
解锁	短焦指纹解锁、人脸识别	短焦指纹解锁、人脸识别	超声波指纹解锁、人脸识别	超声波指纹解锁、人脸识别	超声波指纹解锁、人脸识别
存储配置	LPDDR5X+UFS4.0	LPDDR5X+UFS4.0	LPDDR5X+UFS4.0	LPDDR5X+UFS4.0	LPDDR5X+UFS4.0
存储容量/售价	12+256GB: 4299元 12+512GB: 4699元 16+512GB: 4999元 16+1TB: 5499元	12+256GB: 4199元 12+512GB: 4699元 16+256GB: 4399元 16+512GB: 4999元 16+1TB: 5499元	12+256GB: 4499元 12+512GB: 4799元 16+512GB: 4999元 16+1TB: 5499元	12+256GB: 4499元 12+512GB: 4899元 16+512GB: 5299元 24+1TB: 5999元	12+256GB: 3999元 16+256GB: 4299元 12+512GB: 4499元 16+512GB: 4699元 16+1TB: 5199元

来源：公开信息，CFM闪存市场

中端机型方面，1000~2000的千元机型主要搭载UFS2.2/UFS3.1+LPDDR4X，这部分机型对应的市场需求庞大。当然，也有部分中端手机如Redmi K80、iQOO Neo10、一加Ace 3 Pro，在同等价位追求堪比旗舰机型的UFS4.0+LPDDR5X存储配置，反映出存储配置对手机用户体验有着直接影响，终端不断提高对存储容量和性能的追求。

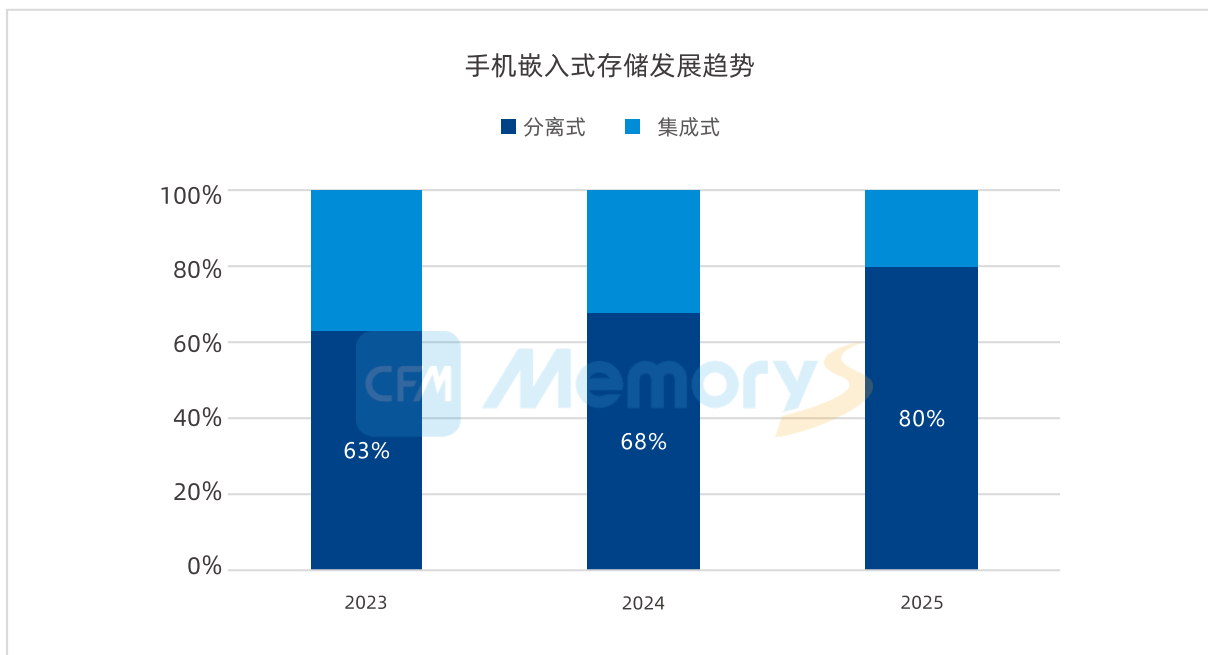
表16：部分中低端手机及其硬件配置

部分中低端手机及其相关配置						
品牌	时间	存储容量	存储配置	处理器	电池容量	当前价格/元
OPPO K12	2024.04	8GB+256GB	LPDDR4X+UFS3.1	骁龙7 Gen 3	5500mAh	1699-1899
		12GB+256GB				
		12GB+512GB				
iQOO Z9	2024.04	8GB+128GB	LPDDR4X+UFS2.2	骁龙7 Gen 3	6000mAh	1499-1999
		8GB+256GB				
		12GB+256GB				
Realme V60	2024.06	6GB+128GB	LPDDR4X+UFS 2.2	天玑6300	5000mAh	1199-1399
		8GB+256GB				
红米Note 14	2024.09	6GB+128GB	LPDDR4X+UFS2.2	天玑7025 Ultra	5110mAh	1099-1599
		8GB+128GB				
		8GB+256GB				
		12GB+256GB				
VIVO Y300 Pro	2024.09	8GB+128GB	LPDDR4X+UFS2.2	骁龙6 Gen 1	6500mAh	1499-2299
		8GB+256GB				
		12GB+256GB				
		12GB+512GB				
Realme13	2024.09	12GB+512GB	LPDDR4X+UFS2.2	天玑6300	5000mAh	2199
Infinix Hot 50	2024.09	4GB+64GB	LPDDR4X+UFS 2.2	天玑6300	5000mAh	845-930
		8GB+128GB				
荣耀X60	2024.10	8GB+128GB	LPDDR4X+UFS3.1	天玑7025 Ultra	5800mAh	1199-1799
		8GB+256GB				
		12GB+256GB				
		12GB+512GB				
华为nova 13	2024.10	12GB+256GB	LPDDR4X+UFS3.1	麒麟8000	5000mAh	2699-3499
		12GB+512GB				
		12GB+1TB				
VIVO S20	2024.11	8GB+256GB	LPDDR4X+UFS2.2	骁龙7 Gen 3	6500mAh	2299-2999
		12GB+256GB				
		12GB+512GB				
		16GB+512GB				
荣耀300	2024.12	8GB+256GB	LPDDR5+UFS 3.1	骁龙7 Gen 3	5300mAh	2299-2999
		12GB+256GB				
		12GB+512GB				
		16GB+512GB				
荣耀Play9T Pro	2024.12	8GB+256GB	LPDDR4X+UFS3.1	天玑 6080	5000mAh	1099-1399
		12GB+256GB				

来源：公开信息，CFM闪存市场

随着用户对手机性能、存储容量、散热等方面的要求不断提高，LPDDR（PoP封装）+分离式eMMC/UFS这种更为高效的存储方案已然成为手机厂商的主流选择，但部分以轻薄著称的手机型号，对集成式嵌入式产品仍有一定需求。通过将LPDDR与处理器进行PoP封装，可以显著减少信号传输路径，提升数据传输速度和能效，在加上分离式eMMC/UFS，可进行更加灵活的存储配置。另外，POP+分离式存储方案在散热和良率上相对于集成式eMCP/uMCP更具优势，使得手机终端能在性能和功耗之间取得更好的平衡。

图19：手机嵌入式存储发展趋势

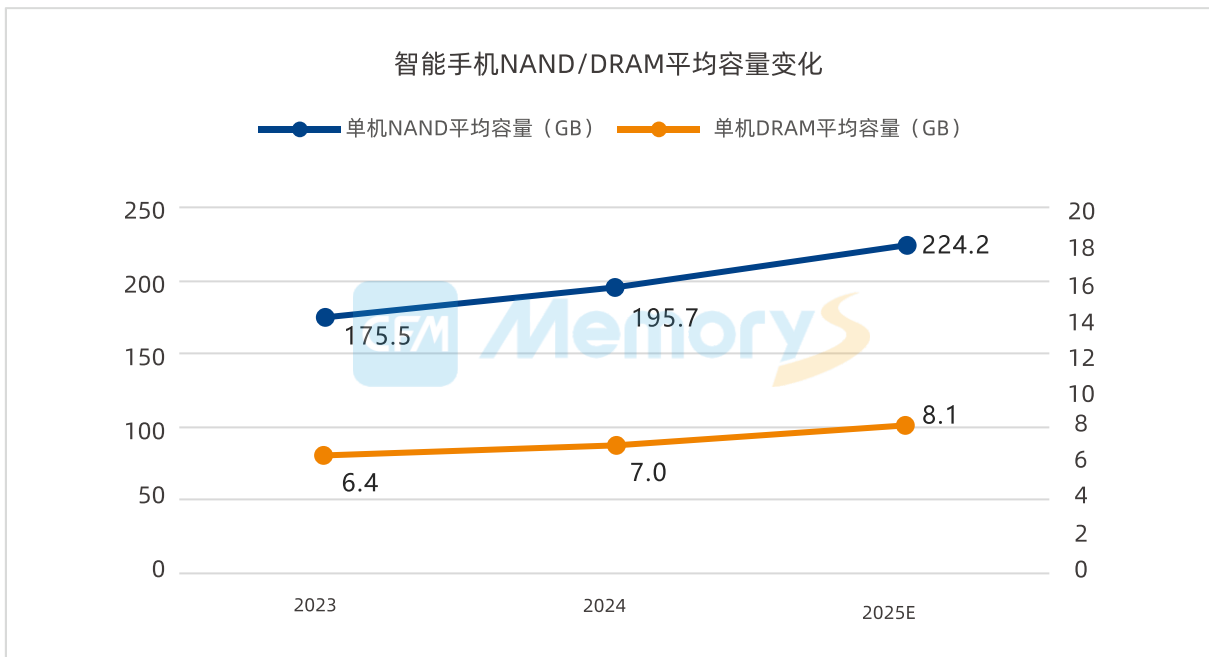


数据来源：CFM闪存市场

#### 4、端侧大模型的普及，进一步提升手机单机容量

2024年是智能手机AI应用的元年，手机厂商纷纷开发各自的端侧大模型，而要想把大模型装进手机内，在算力、内存、存储容量以及电池密度等各方面挑战重重，庞大的参数模型在端侧运行，对手机内存占用量是显而易见的。为实现在手机中成功将端侧大模型内置至系统中，vivo推出尺寸更小的30亿端侧大模型3B，并将其内置在vivo X200的OriginOS 5系统中，内存占用大约1.5GB，相比此前占用户3.5GB内存空间的70亿的蓝心大模型有所下降，而苹果的Apple Intelligence也需要0.7-1.5GB DRAM去运行。至此，除苹果以外，其他手机厂商的新一代旗舰机12GB内存全员上桌，未来也将加速向16GB靠拢，NAND也将从256GB逐渐向512GB、1TB加速渗透。根据CFM闪存市场数据显示，预计2025年智能手机单机NAND容量有望超过220GB，DRAM有望超过8GB。

图20：智能手机NAND/DRAM平均容量变化



数据来源：CFM闪存市场



## 5、手机终端踊跃出海，全球化与本地化相结合

由于国内手机市场竞争日趋激烈，市场饱和度逐渐提高，除始终专做海外市场的传音以外，国产手机厂商扎堆出海，寻求新的增长点，在海外市场凭借差异化策略取得相当不错的成绩。譬如，小米最新财报披露其全球智能手机出货量排名稳居第三，市场份额为13.8%；在全球52个国家和地区的智能手机出货量排名前三，在69个国家和地区的智能手机出货量排名前五。OPPO截至2024年在全球70多个国家设有销售网络，线下售后门店超过3300家，海外出货量占比达到约60%；荣耀去年12月的海外销量占比已突破50%。近年来，国内手机品牌出海的影响力和知名度均较高。根据谷歌携手凯度发布《Google x Kantar BrandZ中国全球化品牌2024》，上榜的50个品牌来自于七大主要品类，其中，消费电子品牌力在所有品类中占比最大，高达25%，但较2023年的28%有所下滑。消费电子类品牌中以手机品牌居多，手机厂商中小米、华为、OPPO进入榜单TOP10。

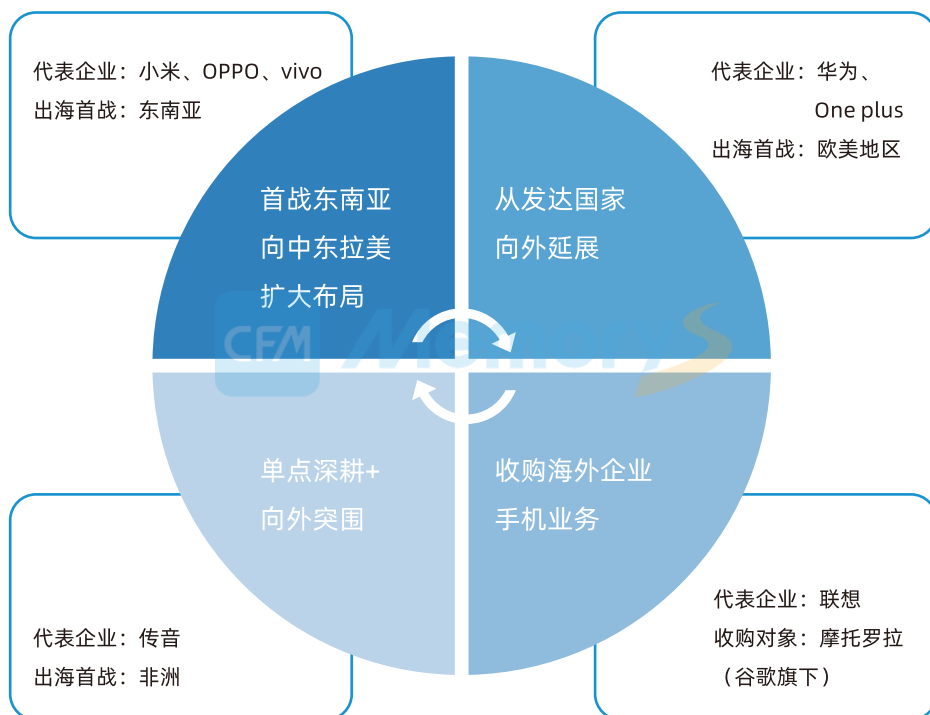
表17：部分手机品牌影响力

	2024年		2023年	
	品牌影响力	排名	品牌影响力	排名
小米	1058	2	1192	2
华为	746	5	828	6
OPPO	635	9	581	7
vivo	445	13	485	10
一加	360	19	344	15
realme	345	21	240	21
荣耀	289	24	215	25
传音	201	44	117	45

来源：Google、Kantar BrandZ，CFM闪存市场

从各家手机厂商出海模式观察发现，小米、OPPO、vivo选择从与中国相邻的东南亚开始踏上出海之旅，向欧洲、中东非、拉美地区扩大布局。华为和一加则选择部分欧美地区中具有区域辐射效应的发达国家作为出海首战。而传音则选择以非洲作为基本盘，向中东和拉美等地延展。联想以收购谷歌旗下的移动设备部门摩托罗拉，加速向全球扩张手机业务。

图21：手机厂商出海模式



来源：公开信息，CFM闪存市场

在市场定位上，小米、vivo、OPPO低端、中端、高端市场全方位布局，华为、荣耀、一加主要倾向于中高端市场，传音主力面向非洲等发展中国家和地区，以中低端机型为主。

表18：主流手机品牌出海机型总结

机型	低端	中端	高端	海外畅销机型(部分)
小米	√	√	√	低端及入门级：Redmi 13C、Redmi A4等。
				中端：小米13T、Redmi Note 13 Pro等。
				高端：MIX Flip、小米14 Ultra等。
OPPO	√	√	√	低端及入门级：OPPO A3、OPPO K12x等。
				中端：OPPO Reno 12 Pro、Reno 12 F等。
				高端：OPPO Find X8系列等。
vivo	√	√	√	低端及入门级：vivo Y03、vivo Y28等。
				中端：vivo X100系列、vivo v30 Pro等。
				高端：vivo X Fold3 Pro、vivo X200系列等。
华为		√	√	中端：nova 12 Pro等。
				高端：Mate 60 Pro、Mate X5、Pura 70 Ultra等。
传音	√	√		低端及入门级：Infinix Smart 8、Hot 40i等。
				中端：TECNO CAMON 30系列等。
荣耀		√	√	中端：Honor 200、Honor X50等。
				高端：Magic V3、Magic 6/7系列等。
oneplus		√	√	中端：OnePlus 12R、Ace 3 Pro等。
				高端：oneplus12/13系列等。

来源：公开信息，CFM闪存市场

在销售渠道上，手机厂商的打法也不太一样。小米注重电商，主推线上+线下相结合，OPPO主攻线下渠道，与当地运营商合作，当然在线上营销也有布局。传音也侧重于线下渠道，不过不同的是，传音以其建立广泛而庞大的经销商体系作为主要的销售渠道。而一加则是注重社区文化，通过建立自己的网络社区，通过社区交流加强品牌与用户之间联系，并通过“海外KOL+海外社媒”的多样化营销策略扩大品牌宣传力度。

表19：部分手机品牌销售渠道侧重点

品牌	销售渠道侧重点
小米	注重电商，“线上+线下”相结合。早前在进入新加坡时，小米借助线上电商平台Lazada等上架新品，在印度尼西亚，与当地经销商合作，举办小型产品体验会。
OPPO	OPPO主攻线下渠道，与当地运营商、零售商合作构建销售网络；采用“厂商一体化”，与代理商合资成立公司。
传音	传音在非洲以由一级国代、二级省代和三级终端门店构成的三级经销商网络为主要销售渠道。
一加	一加注重社交营销，通过打造网络社区，通过社区交流加强品牌与用户之间联系，提高用户对品牌的认同感；投放KOL。联动海外Youtube、Tiktok、Instagram等社媒平台的网红博主，通过测评、功能教程、拆箱等方式扩大宣传。

来源：公开信息，CFM闪存市场

尽管手机厂商们出海的营销打法上存在差异，但在海外供应链策略上却是殊途同归。手机厂商普遍以“长期主义”策略为主，通过建立强大的销售与服务网络，塑造品牌形象，深入当地特色，在供应链、组织管理、产品设计、营销措施上实现本地化。下面以OPPO为例做具体分析。

<b>供应链本地化</b>	生产制造本地化。OPPO在印尼、印度、土耳其、埃及等国家建设了7个制造中心，带动供应链一起出海，打造本土体系化供应网络，降低供应链成本和风险。 物流本地化。打造广泛的配送网络，创办物流公司“极兔”。
<b>组织管理本地化</b>	OPPO在海外培养了大量本地人才，如印尼工厂95%以上管理人员为本地人，非管理人员更是实现了100%本地化，减少文化隔阂，提升管理效率。
<b>产品设计本地化</b>	根据不同市场需求定制产品和营销方案。例如，针对印尼住铁皮屋住户多的特点，增加了抗干扰信号优化；为应对印尼当地雨水较多的天气，特增加了手机防水能力。
<b>营销本地化</b>	OPPO积极融入各区域市场的本地文化。如举办泰国泼水节的品牌活动，OPPO因印尼有咖啡文化盛行在其专卖店内专设咖啡吧。本土明星代言，赞助足球、网球等体育赛事。

来源：公开信息，CFM闪存市场

## 第四章 AI消费电子发展与存储应用分析

ANALYSIS OF AI CONSUMER ELECTRONICS DEVELOPMENT AND MEMORY APPLICATIONS

### 一、AI时代的下一个爆点：AI眼镜

本节要点：



- 人工智能技术的范式突破正引发全球硬件产业的深度变革。基于Transformer架构的大模型与多模态技术的协同进化，推动AI能力从云端向端侧全面渗透。DeepSeek等开源模型的商业化落地，加速了AI算力在终端设备的分布式部署。从AI手机到AI PC，从智能音箱到智能眼镜，从混合现实头显到多功能穿戴设备，AI正在驱动和重构所有硬件，使得硬件形态正在经历从"功能集成"向"智能涌现"的质变，AI定义消费电子时代已然拉开序幕。

我们看到，AI硬件正在成为连接物理世界与数字世界的核心载体，以手机和PC为代表的智能设备依然在加速AI应用中占据重要地位，而同时全球硬科技品牌也都在积极寻找AI时代的下一个爆点，在这场争夺人机交互新入口的产业博弈中，AI眼镜已经崭露头角。

从Google Glass、微软的Hololens，再到苹果的Vision Pro，智能眼镜或者说头戴式设备在AR/VR道路上几乎走到现代技术的极致，极其夸张以及科技感的造型仿佛打通了现实世界与虚拟电影的边界。然而，残酷的现实给过热的预期泼上了冷水，难以减轻的重量以及内容生态的贫瘠令这些智能眼镜或头戴式设备无法快速在消费类市场中扩大应用。

直到，Ray-Ban Meta的横空出世为行业提供出一条全新的思路，激活AI眼镜的热潮。Ray-Ban Meta首先是一副非常时尚的太阳眼镜，其次更是叠加AI属性的强科技感消费电子产品。在与传统太阳眼镜相差无几的价格和重量的基础之上，融入AI技术的Ray-Ban Meta创新地实现第一视角拍摄的核心价值，建立起足够高的社交生态壁垒和差异化竞争优势，而AI应用的增加更是打开了未来的探索空间。Ray-Ban Meta化繁为简，将“相机、耳机、AI助手”三个普通功能做到极致，其48.6g超轻量设计与299美元亲民定价，成功打开消费级市场，令2024年的销量直接突破120万台，重燃智能眼镜行业的战火。

随着AI大模型应用的深入和落地，2025年AI眼镜百镜大战蓄势待发。根据公开信息，截至目前中国已经有至少50家公司正在推进AI眼镜项目，除了雷鸟、影目、Rokid、星纪魅族、XREAL等上一波智能眼镜浪潮中崛起的新兴厂商外，字节跳动、阿里巴巴、百度等互联网企业已经加入，此外小米、华为、OPPO、Vivo、Honor等手机品牌厂商也有计划在未来跟进。字节跳动和阿里巴巴的AI眼镜均分别搭载自家火山大模型和通义大模型，锁定在2025年三季度发布，加入关乎未来的入口争夺战。此外，海外厂商中，苹果、三星、谷歌、亚马逊、索尼等巨头也都在跃跃欲试。

## 二、AI下一代交互入口的技术竞速

在Ray-Ban Meta的带动下，AI眼镜行业迎来快速发展，也走出了两条较为清晰的技术路线，一条是在传统眼镜上做升级，如增加拍摄、语音和AI助手等功能，通过VUI（语音交互界面），以语音、TTS文本转语音以及指示灯反馈进行交互；另一条则是添加显示屏在AR功能上迭代，通过GUI（图像交互界面），以显示、触控、语音以及TTS进行交互。而各大厂商也均在两条路线上有所布局。从最基础的“记录决定性瞬间”，在看到的瞬间下达语音指令即可完成“所见及所得”，发展至今，AI眼镜现在已经融合实时翻译、AR导航、AI大模型、以及还可以实现部分笔记本电脑的功能，一副眼镜、一个键盘即可实现随时随地办公。不同于传统的电子产品，AI眼镜增强了人们对现实世界的掌控感，以轻量化的方式实现了一定程度上的赛博朋克，构建出一种全新的生活方式。

表20：部分品牌AI/AR眼镜技术参数对比

品牌	CPU	存储	重量	售价	交互模式
Ray-Ban Meta	骁龙AR1	32GB	48.6g	299美元	VUI
RayNeo V3	骁龙AR1	32GB	39g	1799元	VUI
闪极AI拍拍镜	紫光展锐W517	/	30g	999元起	VUI
INMO Air3	骁龙AR1	/	/	4999元	GUI
Rokid Glasses	骁龙AR1	32GB	/	2499元	VUI+GUI
StarV View	/	/	74g	2499元	VUI+GUI
XREAL One	自研X1	128GB/256GB	82g	3999元	VUI+GUI
RayNeo X3 Pro	/	/	/	/	VUI+GUI

来源：公开信息，CFM闪存市场

从各智能眼镜的配置来看，AI/AR眼镜常搭载的芯片有高通骁龙AR1和紫光展锐的W517，而高通已经在2022年发布了骁龙AR2，随着性能的提升，预计未来有更多的智能眼镜厂商会选择搭载AR2。在软件系统方面，除了自研系统外，如StarV View内置魅族Flyme XR 2.0空间操作系统，谷歌也在2024年12月推出Android XR系统，旨在为各类VR和AR设备提供统一的操作系统支持，包括头显、眼镜等硬件设备。而在存储方面，目前大部分会采用2GB+32GB的ePOP存储方案，为满足更多大模型的运行和AI应用，128GB/256GB等超薄型定制化ePOP方案也将迎来发展。

尽管目前AI眼镜在续航、散热以及重量等方面还面临许多技术障碍，但随着技术的发展，在5-10年内，AI眼镜有望成为亿级用户的应用平台。据有关统计，AI眼镜成为AI端侧最佳落地场景之一，并定义了人机交互的新方式，其2024年出货量突破200万副，而2025年则有望攀升至1000万，甚至在2030年有望在传统眼镜市场的渗透率达到20%。

### 三、AI重塑硬件产业的未来，打开存储应用想象空间

过去十年，智能手表、智能手环等可穿戴设备风靡一时，它们能够监测心率、记录步数、接收通知，为用户提供便捷的健康管理和信息获取服务。然而，这些设备的功能相对单一，与用户的交互也较为有限。随着自然语言处理、计算机视觉以及机器学习等核心技术不断取得突破，使得AI硬件的功能和性能大幅提升，更加智能化和个性化，并朝着“可融入”的方向进化，能够自然地融入用户的日常生活，无需刻意操作即可提供智能服务。例如，AI眼镜可以实时翻译眼前的外语菜单，AI耳机可以自动过滤环境噪音并增强人声，AI戒指可以识别手势并控制智能家居设备，生成式AI技术的融入也使得智能手表和智能手环从单纯的数据追踪工具转变为全面的健康和健身教练，极大地提升了用户体验。

未来AI硬件将以更强大的感知能力、更深入的场景融合、更紧密的人机协作，不断拓宽在移动穿戴、生活家居、办公工具、运动文娱、情感教育等等场景中的应用，同时更多新形态的硬件设备也将会不断涌现。当AI硬件从“佩戴”走向“融合”，将重新定义人机交互的方式，并催生出新的万亿美元级市场，将为未来存储应用打开想象空间。

## 版权与免责声明

- 1、本报告的版权归深圳市闪存市场资讯有限公司所有。
- 2、本报告是由CFM闪存市场统计分析的成果，所涉及的数据来源于业内厂商、渠道、客户资源以及市场公开数据，由于数据统计的局限性，会存在一定的误差。本报告仅就有关事项提供一般性指导，供客户参考。
- 3、本报告的所有信息不应替代咨询或任何其他专业建议和服务。
- 4、版权所有者可随时更改报告中引用的日期、产品说明、图表和内容。如有更改，将不对外另行通知。
- 5、未经许可，任何人不能以任何形式转载、传输、重制、出版或播送。
- 6、若因本报告造成任何损失、伤害以及纠纷，深圳市闪存市场资讯有限公司不会承担任何责任。
- 7、对报告内容若有异议，请及时与我们联系，Email: [Service@Chinaflashmarket.com](mailto:Service@Chinaflashmarket.com)



微信公众号



闪存市场APP





深圳市闪存市场资讯有限公司

邮箱: [Service@ChinaFlashMarket.com](mailto:Service@ChinaFlashMarket.com)