

# 半导体行业专题： AI 创新与周期向上共振，半导体开启新一轮成长

行业研究 · 行业专题

电子 · 半导体

投资评级：优于大市（维持）

证券分析师：胡剑

021-60893306

hujian1@guosen.com.cn

S0980521080001

证券分析师：胡慧

021-60871321

huhui2@guosen.com.cn

S0980521080002

证券分析师：周靖翔

021-60375402

zhoujingxiang@guosen.com.cn

S0980522100001

证券分析师：叶子

0755-81982153

yezi3@guosen.com.cn

S0980522100003

- 算力、存力是AI的基础，先进封装为其助力

**算力：**Transformer类AI大模型过去几年所需算力平均每2年增长750倍，对AI计算芯片提出了极高的要求，目前AI芯片主要包括GPU、FPGA、ASIC，Gartner预计全球AI芯片市场规模将从2022年的442亿美元增至2027年的1194亿美元。根据TrendForce的数据，目前AI服务器加速芯片以英伟达GPU为主，占比高达6-7成，但CSP业者有扩大自研ASIC的趋势。

**存力：**随着数据量和模型参数的增加，算力环节和应用环节所需存储容量均在持续增加。根据美光科技的预测，2021-2025年DRAM容量的CAGR为15-20%，NAND容量的CAGR为25%-30%，其中数据中心、工业、汽车均超过整体增速。HBM可以满足算力芯片对内存带宽的需求，部分解决内存墙问题，TrendForce预计HBM在DRAM总产能和总产值中的占比将快速提升，2025年分别超过10%和30%，主要由SK海力士、三星、美光主导。

**先进封装：**集成电路制程已接近物理尺寸的极限，进入“后摩尔时代”，先进封装接力“摩尔定律”助力算力和存力芯片，Chiplet异构集成和TSV等是关键技术，比如HBM需要TSV进行堆叠，英伟达H100和AMD MI300需要通过CoWoS将GPU Die和HBM Die集成在一起。

- AI应用落地对半导体的推动将更为长远

2023年英伟达CEO黄仁勋提出“我们正处于AI的iPhone时刻”；2024年又提出AI已到“Tipping Point”。在AI基建和大模型算法的良性循环下，我们认为AI的应用将逐渐落地，从特定人群使用走向千家万户。通讯技术催生出了手机，互联网技术催生出了电脑和智能手机，AI除赋能现有终端推出AI手机、AI PC、AIoT外，更可能使期待已久的人形机器人、自动驾驶汽车成为现实，或者催生出未知终端，赋能基础研究。

## ● 半导体周期进入上行阶段

根据SIA的数据，全球半导体季度销售额同比增速在1Q23触底，之后跌幅收窄，4Q23同比转正，1Q24全球实现销售额为1377亿美元，同比增长15.2%，环比减少5.7%。存储芯片是半导体产品中波动性最大的类别，存储大厂美光科技和SK海力士的季度收入和毛利率均已拐头向上。多家机构预计2024年全球半导体销售额恢复增长，增速在10%-25%之间，其中TechInsights在3月的最新预测中上调2024年全球半导体销售额增速至24%（前次预测值为16%），将超过6500亿美元，并预计2025/2026年分别增至超过8000亿美元和接近9000亿美元。

## ● 投资策略：AI创新与周期向上共振，半导体开启新一轮成长

根据SIA的数据，1Q24全球半导体销售额为1377亿美元，同比增长15.2%，环比减少5.7%；中国销售额为424亿美元，同比增长27.4%，环比减少6.6%。WSTS等多个机构均预计2024年全年将恢复10%以上增长，我们认为本轮半导体周期进入上行阶段。同时，AI创新正在从算力基础设施建设，扩展至AI手机、AI PC、AIoT等AI终端，更有望加速人形机器人、自动驾驶汽车等落地，AI的“Tipping Point”已到，预计将为半导体带来新一轮的成长。

## ● 风险提示

国产替代进程不及预期；下游需求不及预期；行业竞争加剧的风险；国际关系发生不利变化的风险。

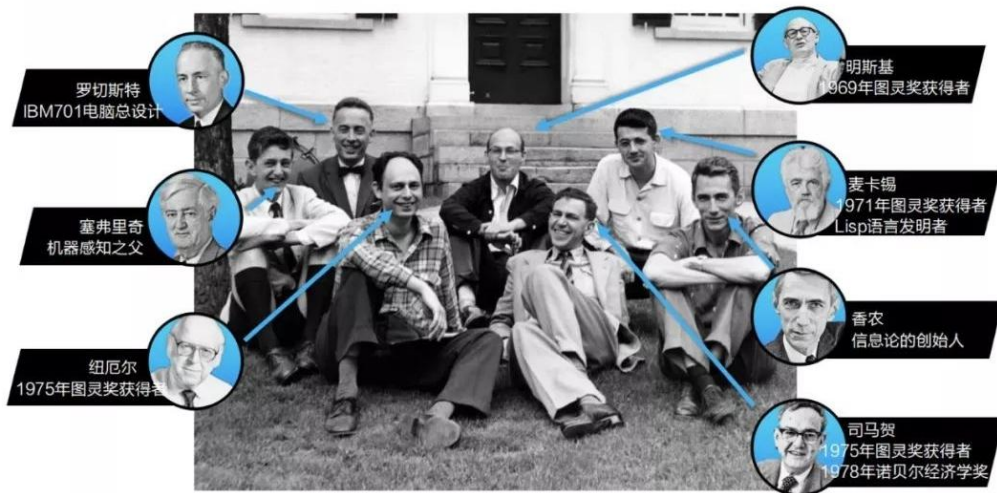
- [ 01 ] AI 真的“狼来了”
- [ 02 ] 算力、存力是AI的基础，先进封装为其助力
- [ 03 ] AI应用落地对半导体的推动将更为长远
- [ 04 ] 半导体周期进入上行阶段
- [ 05 ] 投资策略

## AI真的“狼来了”

# AI的发展历程

- **AI的诞生。**1950年，艾伦·图灵（Alan Turin）写了论文《计算机器与智能》，提出**图灵测试**用来检测机器智能水平；1956年，美国达特茅斯学院举行了历史上第一次人工智能研讨会，被认为是人工智能诞生的标志。
- AI诞生后经历了三次浪潮，我们认为，前两次主要由政府和科研单位推动，第三次主要由产业界推动。

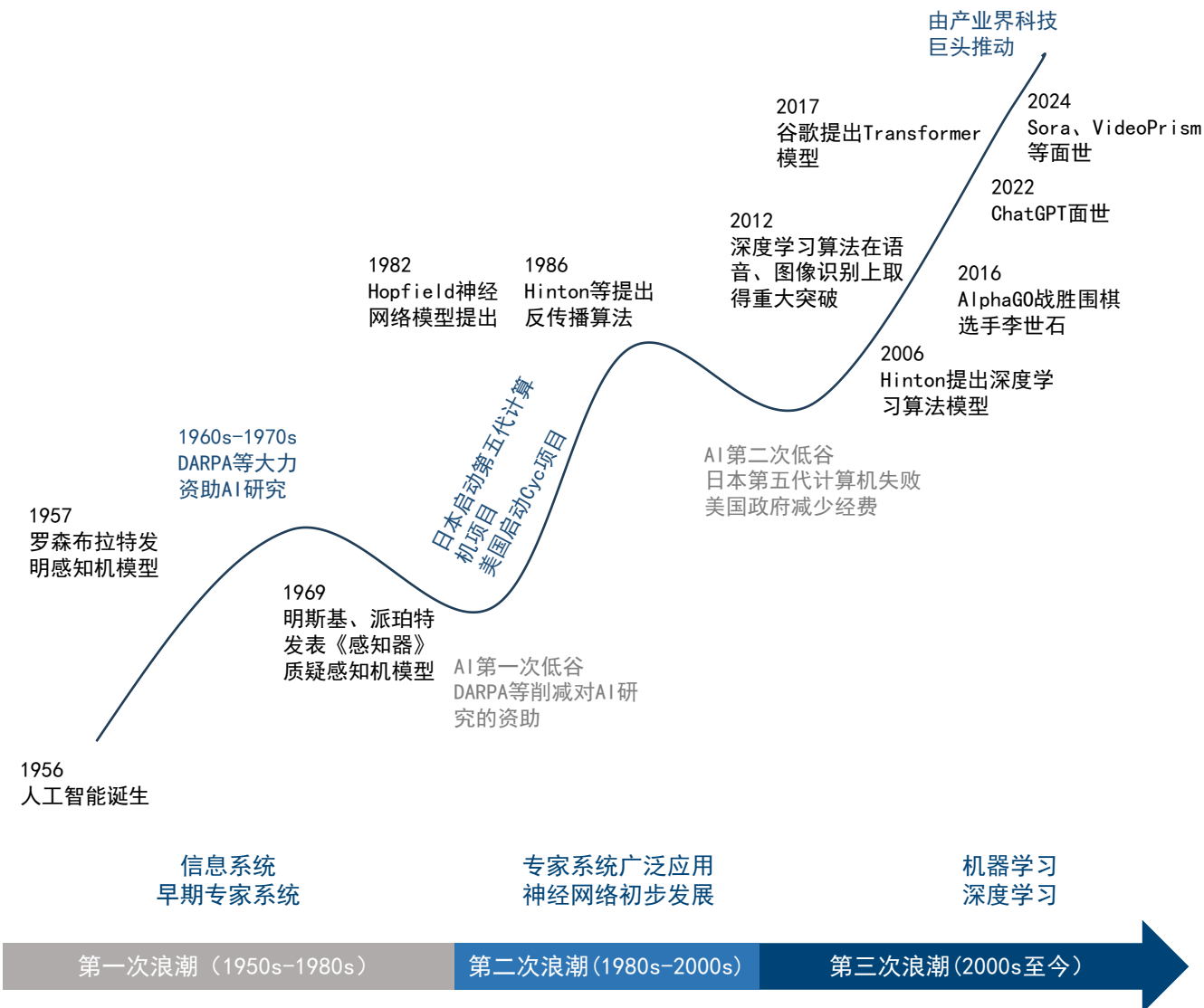
图：达特茅斯会议——历史上第一次AI研讨会



资料来源：数据侠，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：人工智能的发展中的三次浪潮



资料来源：新华三，国信证券经济研究所整理

谷歌



- 2011年成立谷歌大脑 (Google Brain)
- 2017年战略从Mobile First到AI First; 2023年将谷歌大脑和DeepMind (2014收购) 合并为Google DeepMind
- 2015年发布TensorFlow; 2017年推出Transformer架构; 2018年推出BERT模型; 2021年发布机器学习平台Vertex AI; 2022年推出PaLM模型; 2023年发布Gemini模型; 2024年发布Gemma模型
- 2016年推出TPU; 2021年发布手机自研芯片Tensor, 可在手机上处理AI和ML模型

微软



- 2017年成立人工智能实验室
- 2019年投资OpenAI 10亿美元, OpenAI在微软Azure云平台开发AI技术并将微软作为商业化的首要合作伙伴, 到目前微软累计对OpenAI投资了130亿美元
- 2018年发布人工智能开发项目平台AI Lab; 2021年推出自动化机器学习框架FLAML; 2023年发布小模型phi-1和Orca; 2023年推出多Agents框架Autogen

Meta



- 2013年成立人工智能研究院 (FAIR)
- 2017年开源框架PyTorch; 2019年发布模型DLRM; 2023年发布模型LLaMA; 2023年发布视觉模型SAM

百度

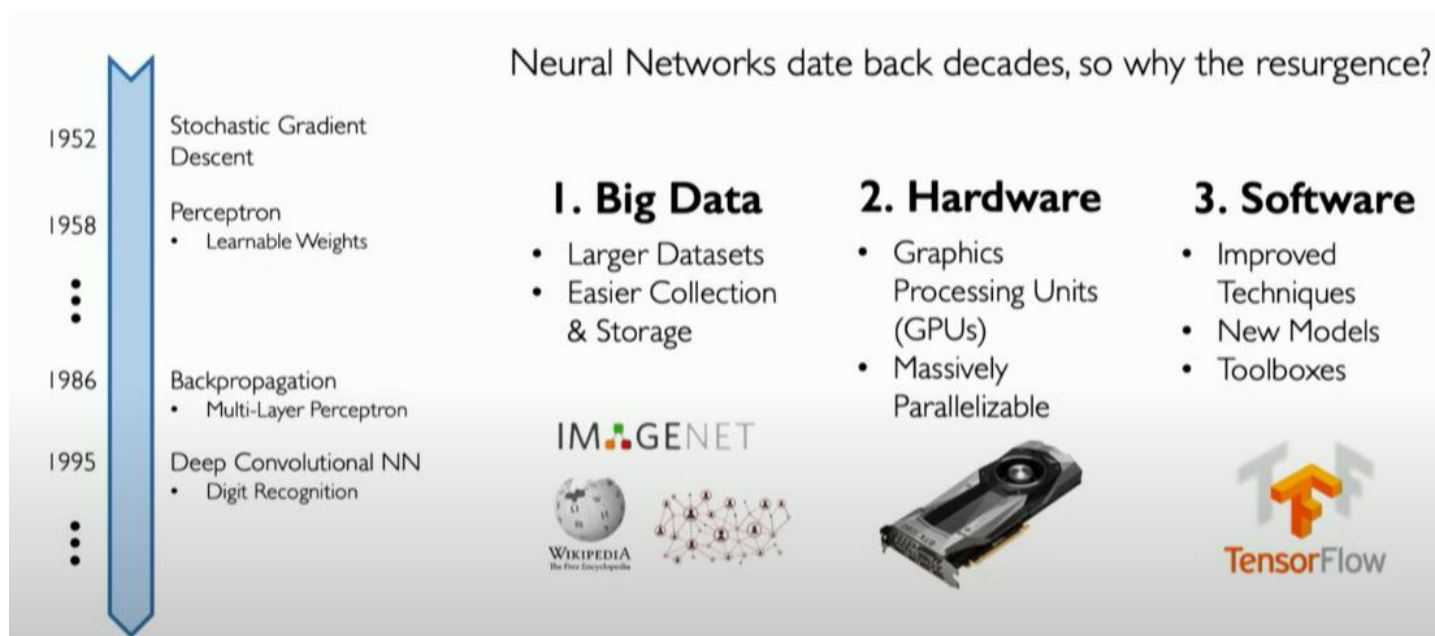


- 2013年深度学习研究院 (IDL) 成立
- 2014年硅谷人工智能实验室正式成立
- 2019年发布文心大模型1.0; 2023年发布大模型文心一言

## ● 为什么是现在？量变带来质变，AI已到临界点。

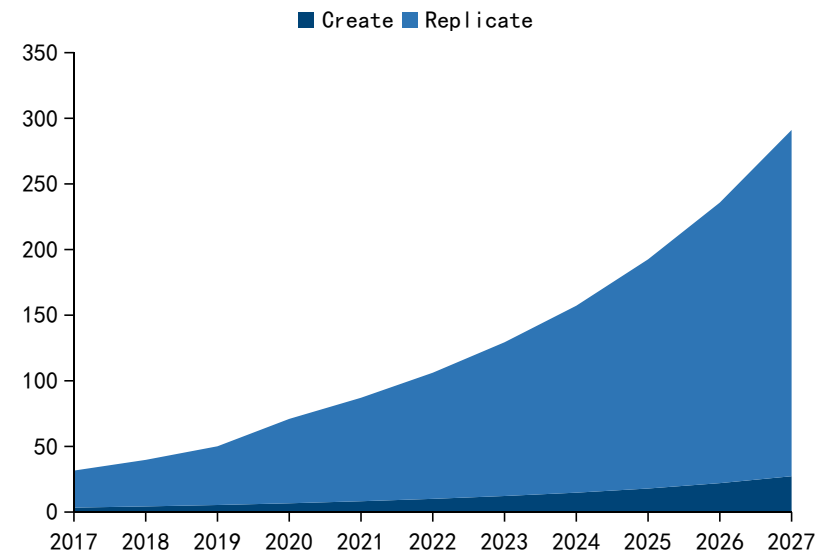
- **数据量：**移动互联网时代，数据量快速增长。根据IDC的数据，2010年底全球数据量为1.2ZB，2022年突破100ZB，增长了近百倍。
- **算力：**2006年，英伟达发布首个通用GPU架构Tesla，采用CUDA架构，GPU从专用图形处理器转为通用数据并行处理器；2012年AlexNet用两块GTX 580 GPU赢得了ImageNet图象识别大赛，英伟达日益加大CUDA投入；2022年英伟达推出的H100 SXM5 FP32单精度算力为60TFlops。
- **算法：**各类深度学习框架和模型推出。

图：数据、算力、算法共同推动AI兴起



资料来源：MIT，国信证券经济研究所整理

图：全球数据量（ZB）



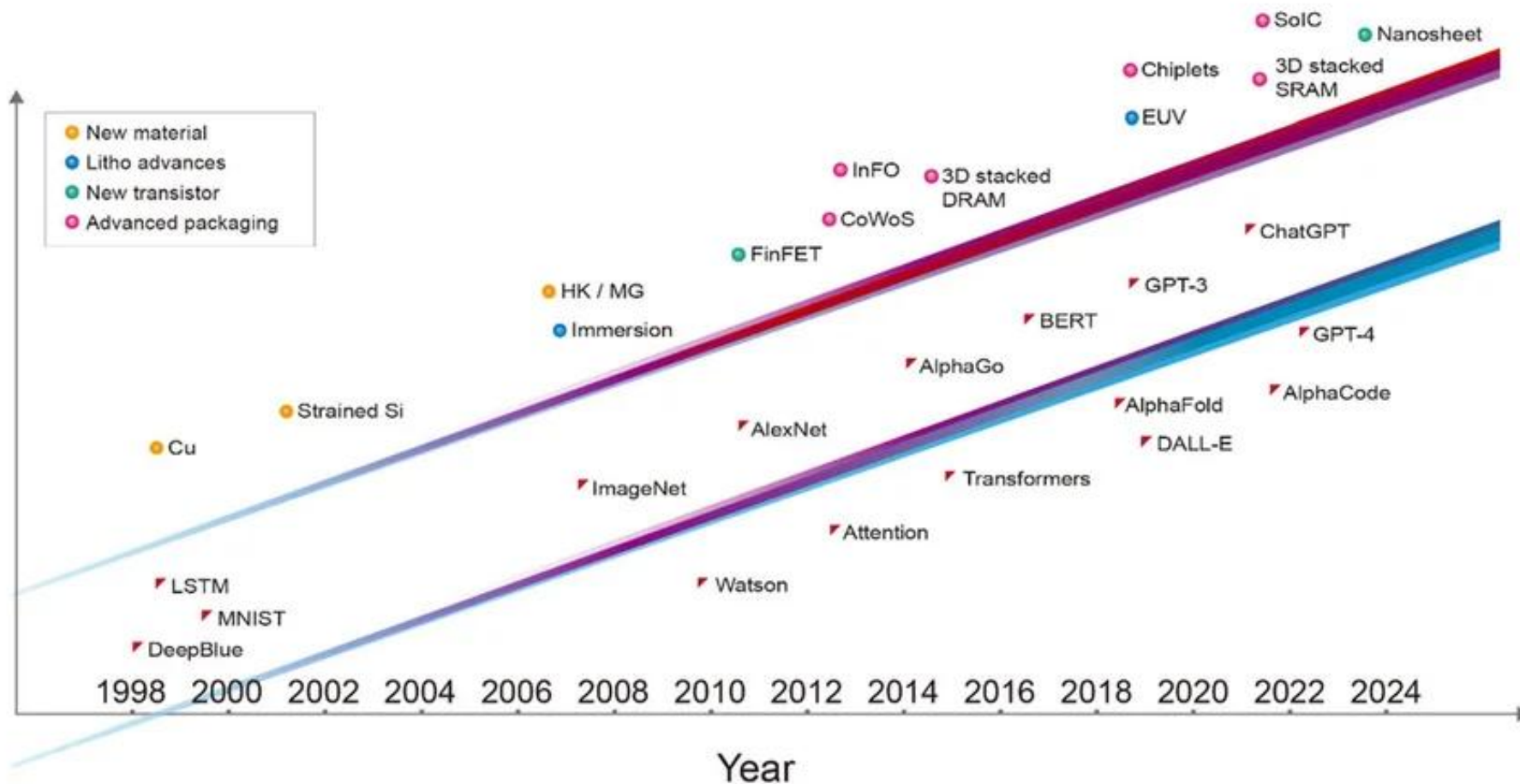
资料来源：IDC，国信证券经济研究所整理



# AI真的“狼来了”

● 半导体技术是AI背后的功臣。过去三十年AI的重大里程碑都是由当时的先进半导体技术实现的，比如赢得ImageNet竞赛的GPU使用了40nm技术，AlphaGo 征服围棋游戏使用28nm技术，ChatGPT的初始版本采用5nm技术。

图：半导体技术是AI背后的功臣



---

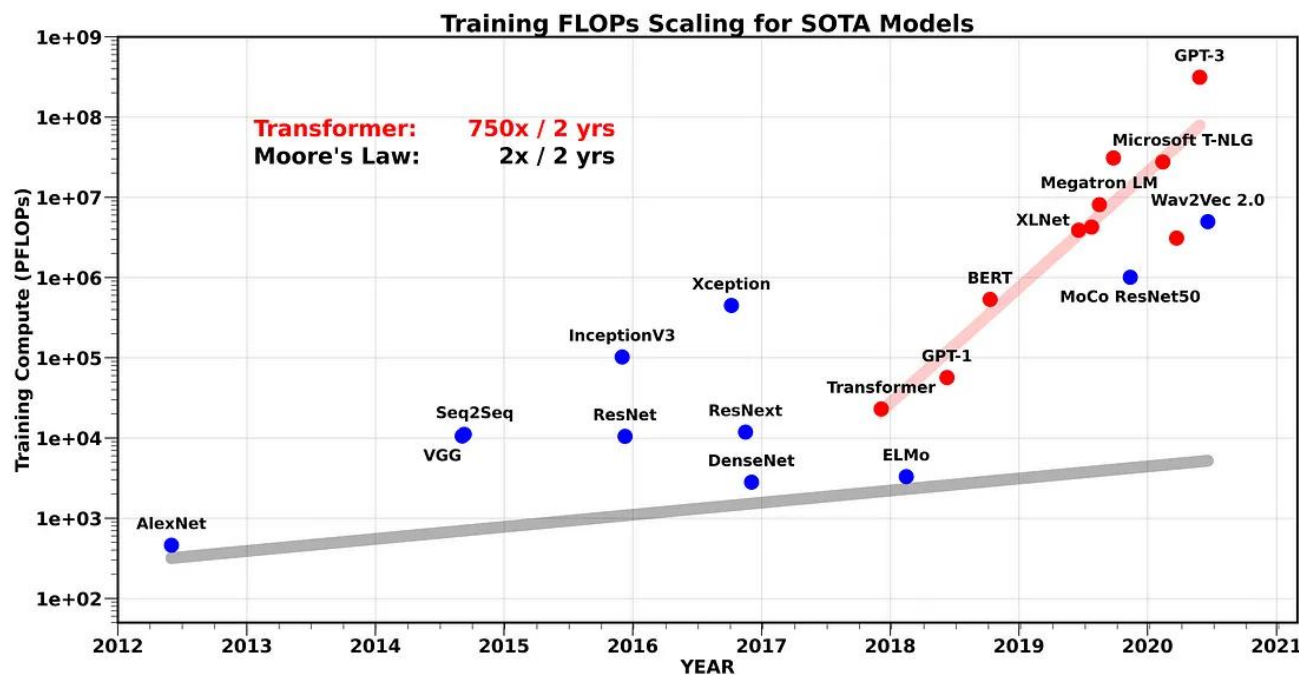
算力、存力是AI的基础，先进封装为其助力

---

# AI大模型所需算力大幅增加

- Transformer类AI大模型所需算力平均每2年增长750倍，对计算芯片的算力提出了极高的要求。
- 智能算力是面向人工智能应用，提供人工智能算法模型训练与模型运行服务的计算机系统能力。
  - 根据中国信通院数据，2022年全球算力总规模达到906 EFLOPS，其中智能算力（FP32）为451 EFLOPS。
  - 根据IDC的数据，2022年中国智能算力（FP16）规模达259.9 EFLOPS，预计2027年将增加到1117.4 EFLOPS，CAGR达34%。

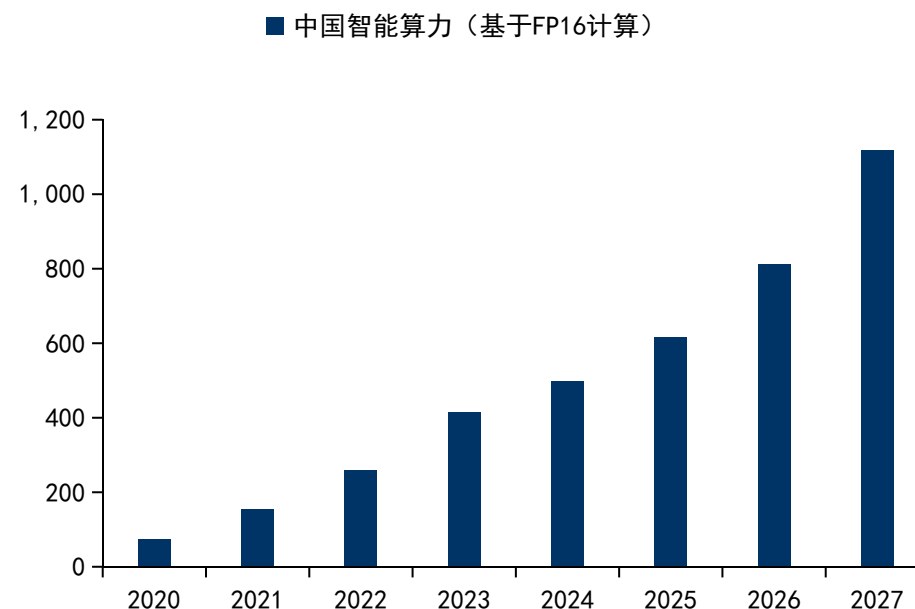
图：Transformer大模型对算力需求快速增长



资料来源：riselab，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：中国智能算力规模预测（单位：EFLOPS）

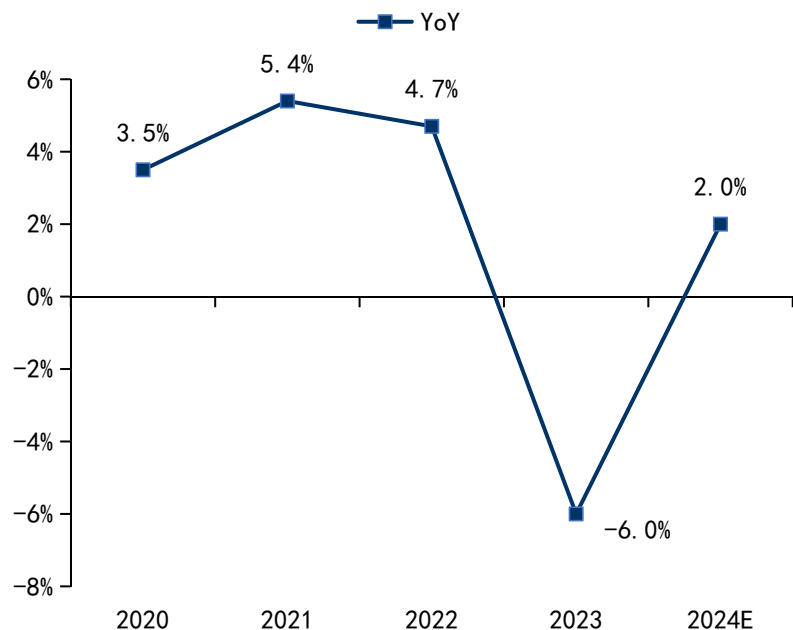


资料来源：IDC，国信证券经济研究所整理

# 2024年AI服务器出货量165万台

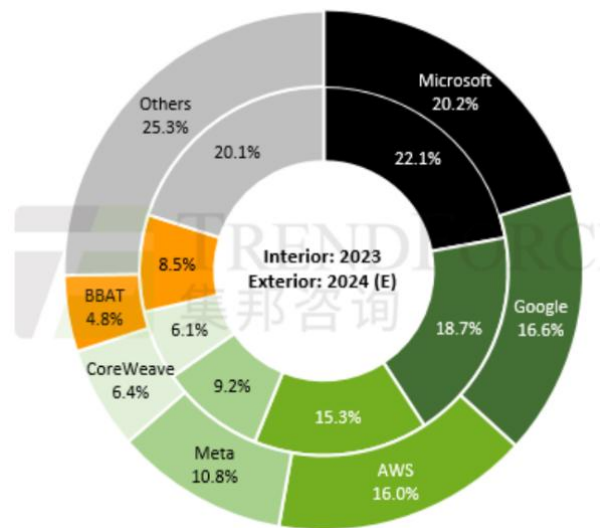
- 预计2024年AI服务器出货量165万台，美国四大CSP需求超6成。根据TrendForce的预测，2024年全球服务器整机出货量将增长2.05%至1365.4万台，其中AI服务器出货占比将提高至12.1%，出货量165万台。美国CSP主导对AI服务器需求，Microsoft、Google、AWS、Meta合计需求超6成。
- AI服务器加速芯片以英伟达GPU为主，占比高达6-7成，CSP业者有扩大自研ASIC的趋势。

图：全球服务器整机出货量YoY



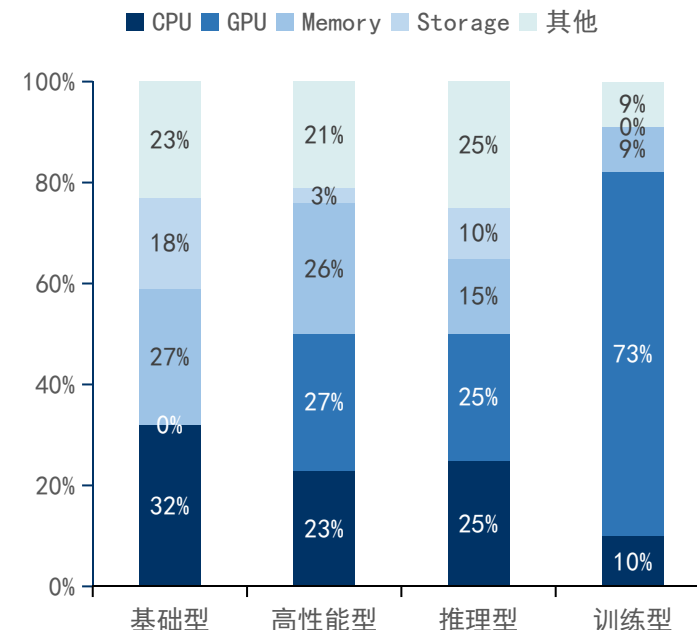
资料来源：TrendForce，国信证券经济研究所整理

图：AI服务器需求来源



资料来源：TrendForce，国信证券经济研究所整理

图：各类服务器成本结构分拆



资料来源：IDC，集微网，国信证券经济研究所整理

# 2027年AI芯片市场规模达1194亿美元

- AI芯片主要包括GPU、FPGA、ASIC。AI芯片也被称为AI加速器或计算卡，是专门用于处理人工智能应用中的大量计算任务的模块，主要包括图形处理器（GPU）、现场可编程门阵列（FPGA）、专用集成电路（ASIC）等。
- 预计2027年AI芯片市场规模达1194亿美元。根据Gartner的数据，2022年全球AI芯片市场规模为442亿美元，预计2027年增长至1194亿美元。

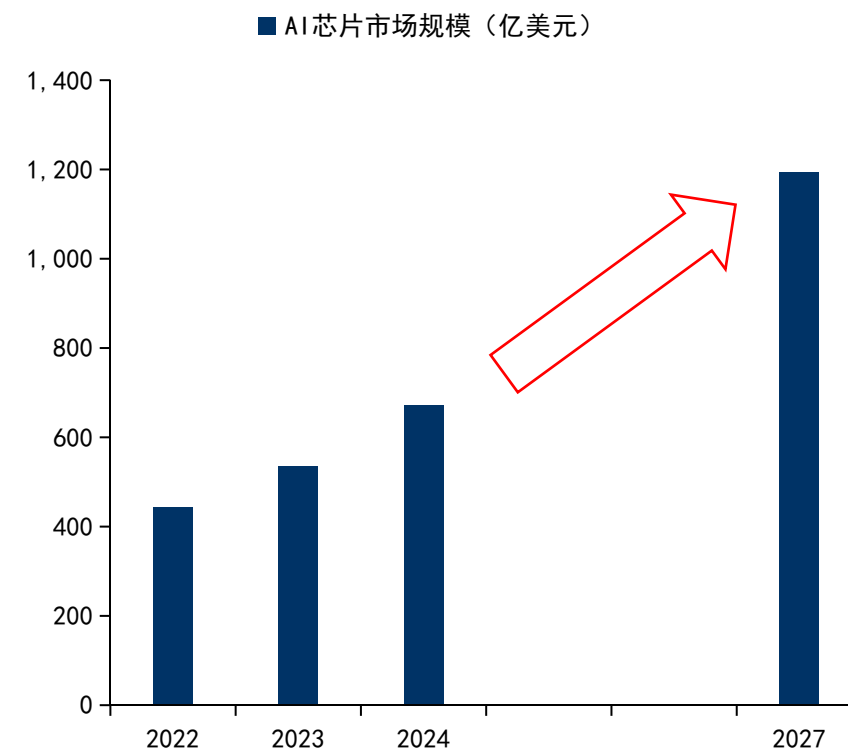
图：AI芯片对比

特点	GPU	FPGA	ASIC
定制化程度	通用型	半定制化	定制化
延迟	较高	低	低
优势	擅长并行计算，浮点数据计算能力强，软硬件体系一致	可进行数据并行和流水线并行，可编程，灵活度高	AI运算效率高，功耗低，体积小
劣势	面积大，功耗高，由于通用性要求难以专一面对某一模型深度优化	开发周期长，复杂算法开发难度大	灵活性差，算法支持有限，算法迭代后需重新开发
算力	中	高	高
价格	高	中	低
应用场景	训练和推理	主要用于推理	训练和推理
代表厂商	英伟达、AMD	赛灵思、Altera	谷歌TPU

资料来源：集微网，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：全球AI芯片市场规模

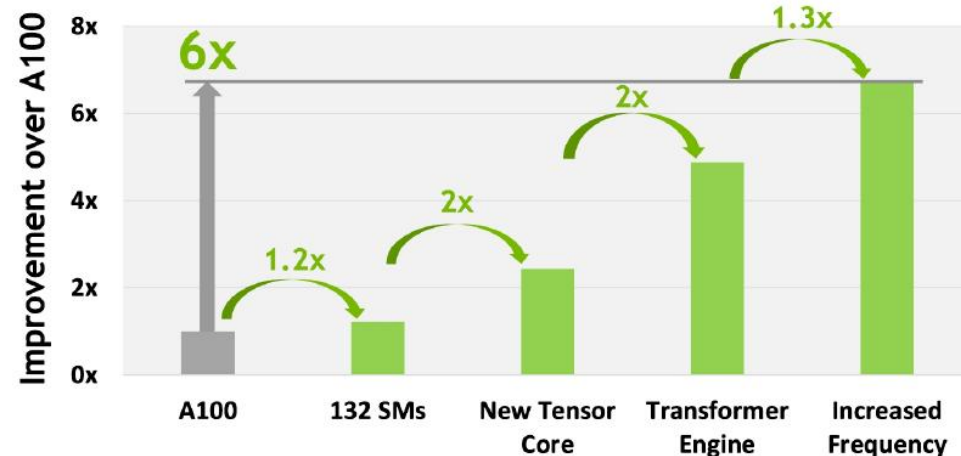


资料来源：Gartner，国信证券经济研究所整理

# GPU架构持续演进以满足算力需求

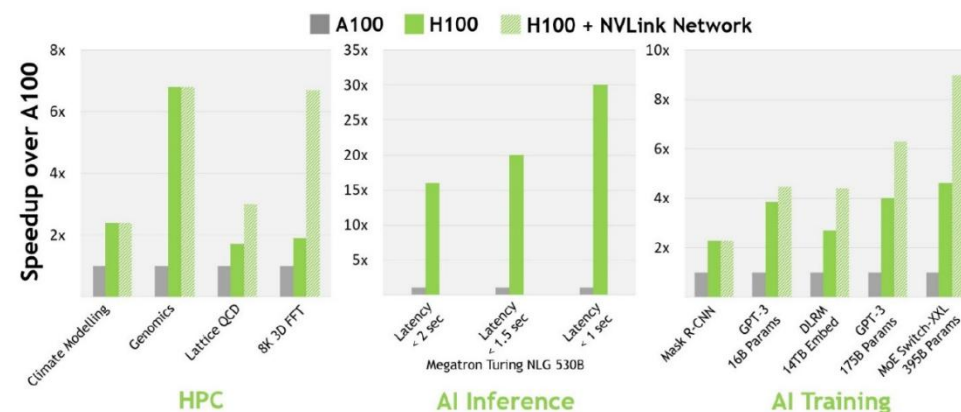
- GPU起于图形处理，兴于通用计算。GPU的推出最初是为了解决CPU在图形处理领域性能不足的问题；之后演变出GPGPU，增加专用向量、张量、矩阵运算指令，提高浮点运算的精度和性能，以满足AI等通用计算需求。
- GPU架构持续演进。英伟达H100计算性能较A100提升约6倍；预计2024年将推出的基于Blackwell架构的B200算力是H100的5倍。

图：英伟达H100相比A100计算性能提升约6倍



图：GPU技术演变

时间	类型	相关标准	代表产品	基本特征	意义
20世纪80年代	图形显示	CGA, VGA	IBM 5150	光栅生成器	最早图形显示控制器
20世纪80年代末	2D加速	GDI, DirectFB	S386C911	2D图元加速	开启2D图形硬件加速
20世纪90初	部分3D加速		3DLabs Glint300SX	硬件T&L	第一颗用于PC的3D图形加速芯片
20世纪90年代后期	固定管线	OpenGL (1.1~4.1), DirectX (6.0~11)	NVIDIA GeForce256	shader功能固定	首次提出GPU概念
2004-2010	统一渲染		NVIDIA G80	多功能shader	CUDA与G80一同发布
2011至今	通用计算 GPGPU	CUDA, Open CL 1.2~2.0	NVIDA TESLA	完成与图形处理无关的科学计算	NVIDA正式将用于计算的GPU产品线独立出来










资料来源：集微网，国信证券经济研究所整理

资料来源：英伟达，国信证券经济研究所整理

# 存储容量长期呈上升趋势

- 各领域存储容量均将增长，其中数据中心、工业、汽车增速较高。根据美光科技的预测，2021-2025年DRAM容量的CAGR为14-19%，其中数据中心、工业、汽车分别为28%、24%、40%；NAND容量的CAGR为26%-29%，其中数据中心、工业、汽车分别为33%、45%、49%。

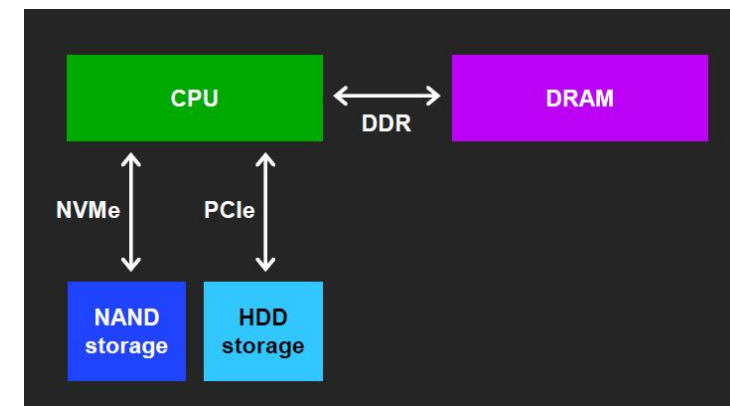
图：各领域存储容量增速预测

	 Data center	 Mobile	 PC	 Industrial	 Auto	 Others	 Total
Market size CY21	\$50B	\$46B	\$34B	\$5B	\$4B	\$21B	\$161B
DRAM bit demand CAGR CY21-25	28%	14%	9%	24%	40%	16%	Mid to high teens %
NAND bit demand CAGR CY21-25	33%	24%	21%	45%	49%	35%	High 20s %

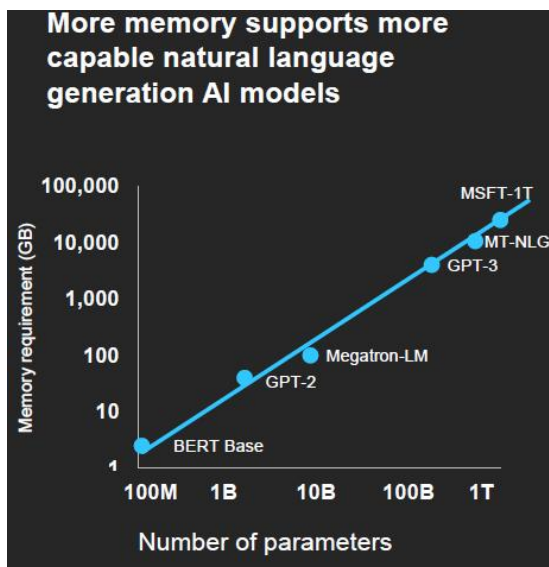
# 异构数据中心推动存储容量增加

- 异构数据中心存储容量数倍增长，预计2021-2030年数据中心存储TAM的CAGR为14%。异构计算系统是由CPU加上GPU/ASIC/FPGA等各种加速器构成的计算系统，以更优、更经济的解决计算任务，包括AI大模型。美光预计2025年服务器所需要的NAND和DRAM容量分别是2021年的3倍和2倍，2021-2030年总市场规模CAGR为14%。
- 所需存储容量和AI模型参数量正相关。

图：数据中心转向异构计算

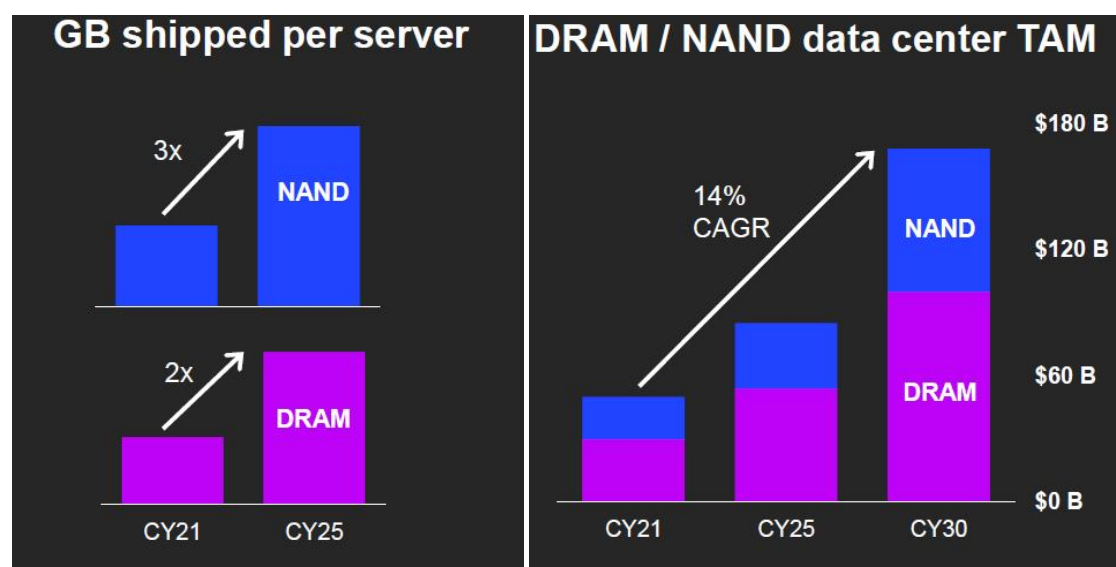


图：存储容量和AI模型参数量正相关

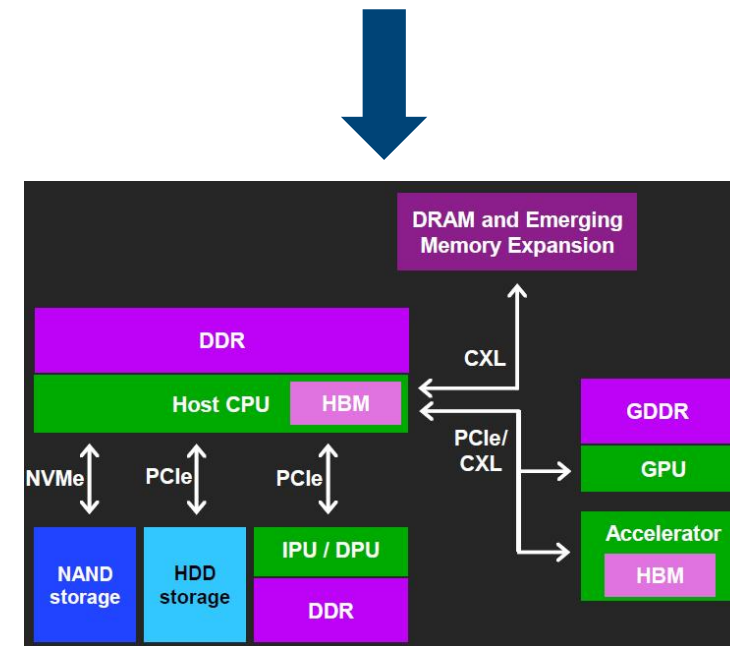


资料来源：美光科技，国信证券经济研究所整理

图：数据中心存储容量和TAM增加



资料来源：美光科技，国信证券经济研究所整理

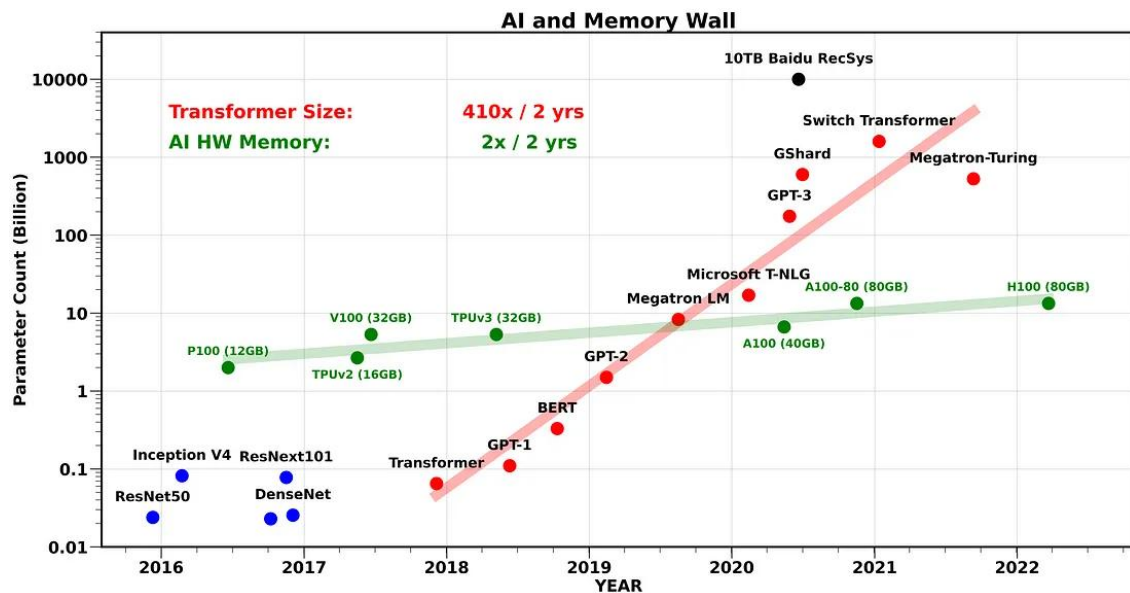


资料来源：美光科技，国信证券经济研究所整理

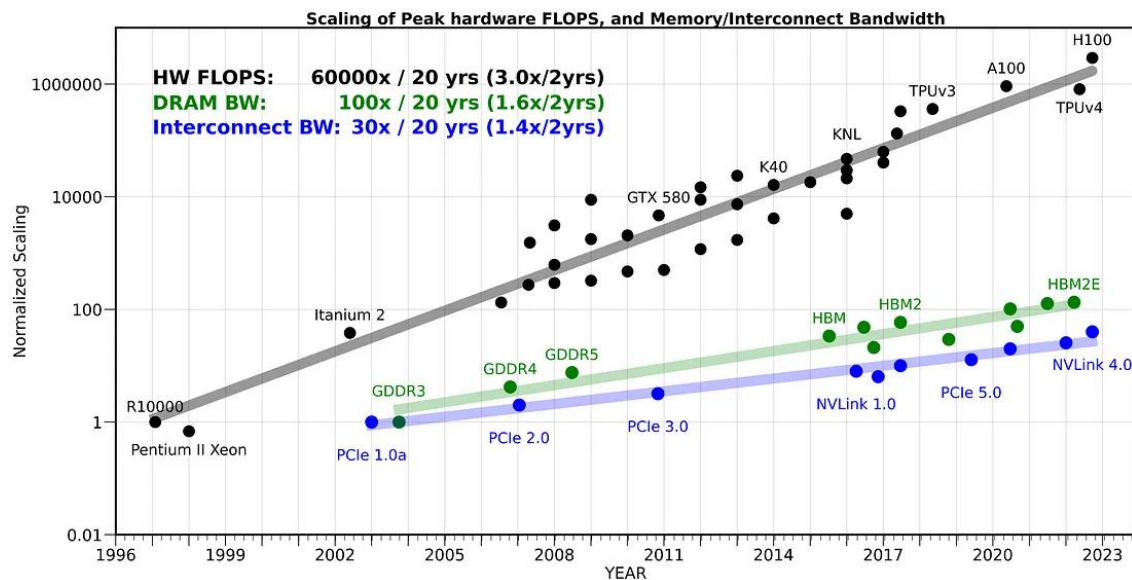


# HBM可解决部分“内存墙”问题

- AI发展中面临“内存墙”问题。内存墙问题既包括内存容量大小不足，也包括内存的传输带宽不足。
  - 大语言模型Transformer参数量平均每2年增长410倍，但加速器DRAM内存每2年仅增长2倍。
  - 过去20年峰值算力提高了90000倍（平均每2年3.0倍），DRAM带宽仅提高了100倍（平均每2年1.6倍），互联带宽仅提高了30倍（平均每2年1.4倍）
- HBM可以满足算力芯片对内存带宽的需求。高带宽内存（High Bandwidth Memory, HBM）是通过TVS技术将DRAM裸片垂直堆叠的3D DRAM，与GPU一起封装以实现更高的传输带宽。相较于其他DRAM的集成方式，HBM存储单元外的导线短，数据传递速度快，损耗小。



资料来源: riselab, 国信证券经济研究所整理

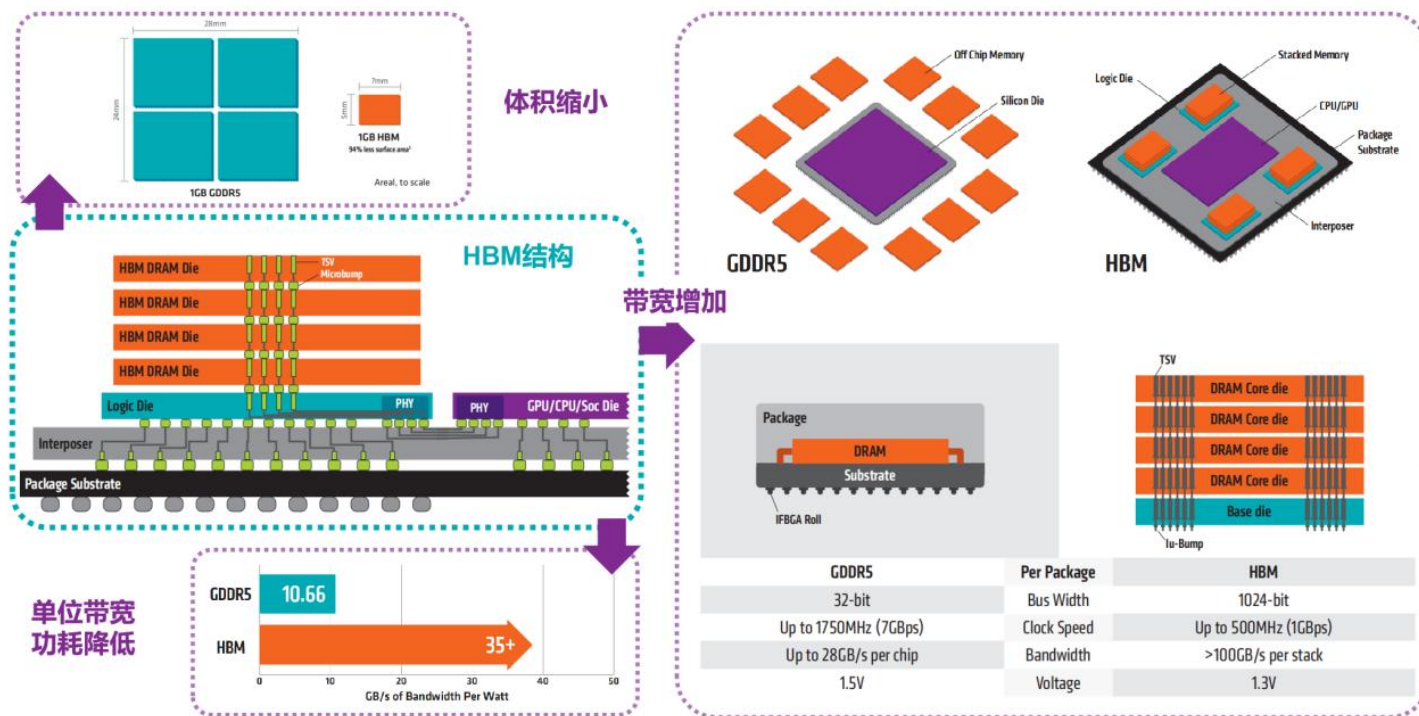


资料来源: riselab, 国信证券经济研究所整理

# HBM可解决部分“内存墙”问题

● HBM于2015年商用，技术不断迭代。2013年HBM被半导体行业标准机构JEDEC正式采纳，2015年AMD Radeon Fury系列率先搭载SK海力士生产的全球首款HBM，之后HBM沿着HBM1、HBM2、HBM2e、HBM3演进，容量和带宽均明显提升。HBM3于2022年1月发布，并在H100中率先搭载。最新的HBM3e预计将在2024年下半年逐季放量。

图：HBM的优势



图：HBM技术迭代

最高性能	HBM1	HBM2	HBM2e	HBM3
Die密度	2Gb	8Gb	16Gb	32Gb
最高堆叠	4-Hi	8-Hi	12-Hi	16-Hi
容量	1GB	8GB	24GB	64GB
带宽 (GB/s)	128	307	410	819

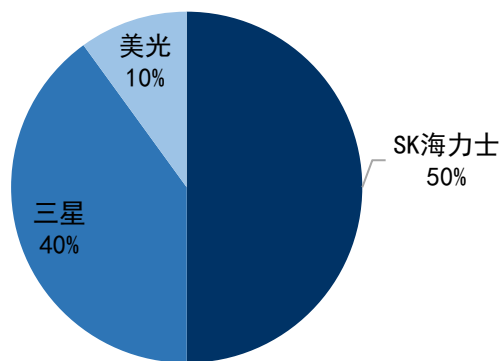
资料来源：AMD，国信证券经济研究所整理

资料来源：JEDEC，CFM，国信证券经济研究所整理

# HBM由SK海力士、三星、美光主导

- HBM在DRAM产能和产值中的占比快速提升。根据TrendForce预计，2024年HBM需求位元增长近2倍，2025年有望再翻倍；预计2025年HBM占DRAM总产能的比例将超过10%；由于HBM单价高，预计2025年其占DRAM总产值的比例将超过30%。
- HBM由SK海力士、三星、美光主导。根据TrendForce数据，2022年SK海力士、三星、美光HBM市占率分别为50%、40%、10%，集中度相对较高。截至2024年底，预计三星、SK海力士、美光HBM的产能分别为13万片/月、12+万片/月、2万片/月。

图：HBM的主要参与者



	三星	SK海力士	美光
2023年底HBM TSV产能	4.5万片/月	4.5万片/月	3000片/月
2024年底HBM TSV产能	13万片/月	12-12.5万片/月	2万片/月

资料来源：TrendForce，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：HBM在DRAM中的占比

	2023	2024E	2025F
HBM位元占比	2%	5%	超过10%
HBM产值占比	8%	21%	超过30%

图：英伟达和AMD AI芯片发展进程及HBM规格

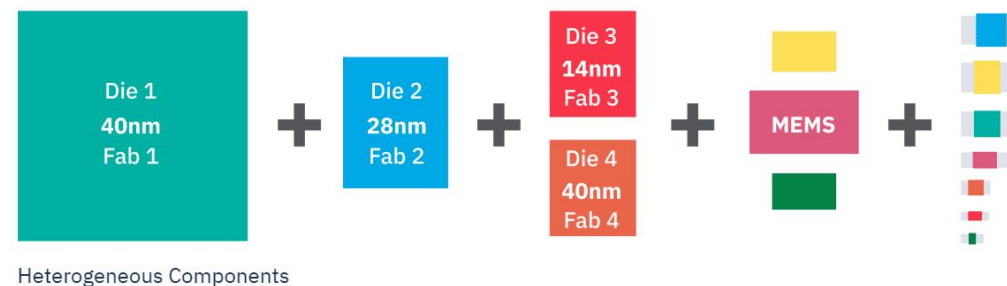
Company	AI Chips	2022	2023				2024F				2025F			
			1Q23	2Q23	3Q23	4Q23	1Q24	2Q24	3Q24	4Q24	1Q25	2Q25	3Q25	4Q25
NVIDIA	H100	HBM3 8hi 80GB												
	GH200 (CPU+GPU)						HBM3e 8hi 141GB							
	H20						HBM3 8hi 96GB							
	H200						HBM3e 8hi 141GB							
	B100										HBM3e 8hi 192GB			
	GB200 (CPU+GPU)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

资料来源：TrendForce，国信证券经济研究所整理

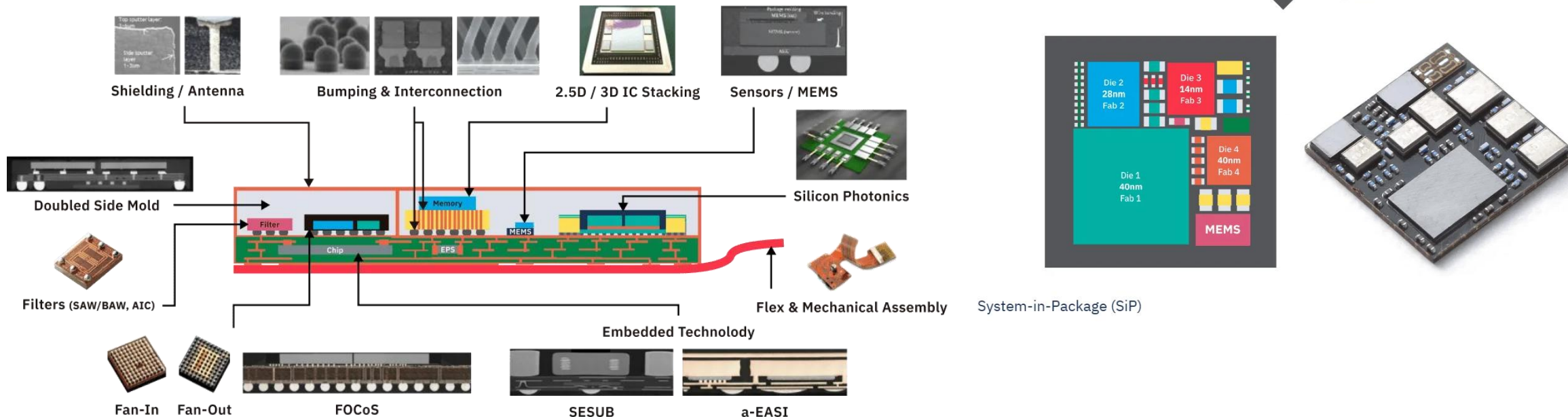
# 先进封装接力摩尔定律为算力和存力芯片助力

- 集成电路行业进入了“后摩尔时代”。集成电路制程工艺已接近物理尺寸的极限，进入了“后摩尔时代”。
- 从集成器件到集成小芯片。通过芯片异构集成，将传感、存储、计算、通信等不同功能的元器件集成在一起，可以突破光罩限制，进一步提高集成度，平衡计算性能、功耗、成本。比如台积电的CoWoS（Chip-on-Wafer-on-Substrate）技术可以容纳六个掩模版区域的计算芯片，及十几个高带宽内存（HBM）芯片。

图：异构集成技术



Integration

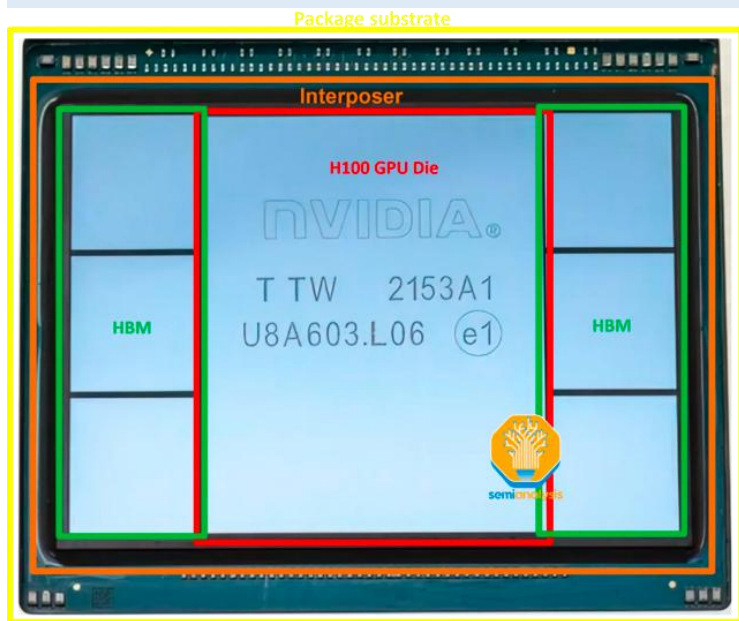


资料来源：日月光，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

- CoWoS成为AI芯片的主要封装形式。台积电2008年底成立集成互连与封装技术整合部门，2011年推出CoWoS封装技术，2013年在赛灵思28nm的FPGA上量产，之后随着AI的发展被大量采用，包括英伟达H100和AMD MI300，通过CoWoS将GPU Die和HBM Die集成在一起。
- TSV技术是HBM的核心。HBM通过3D TSV（硅通孔）技术将多层DRAM Die堆叠起来。
- 台积电扩产CoWoS。台积电表示，2023-2026年公司CoWoS产能将以60%的CAGR增长。TrendForce预计2024年底台积电CoWoS月产能将接近40K，2025年规划总产能接近翻倍。

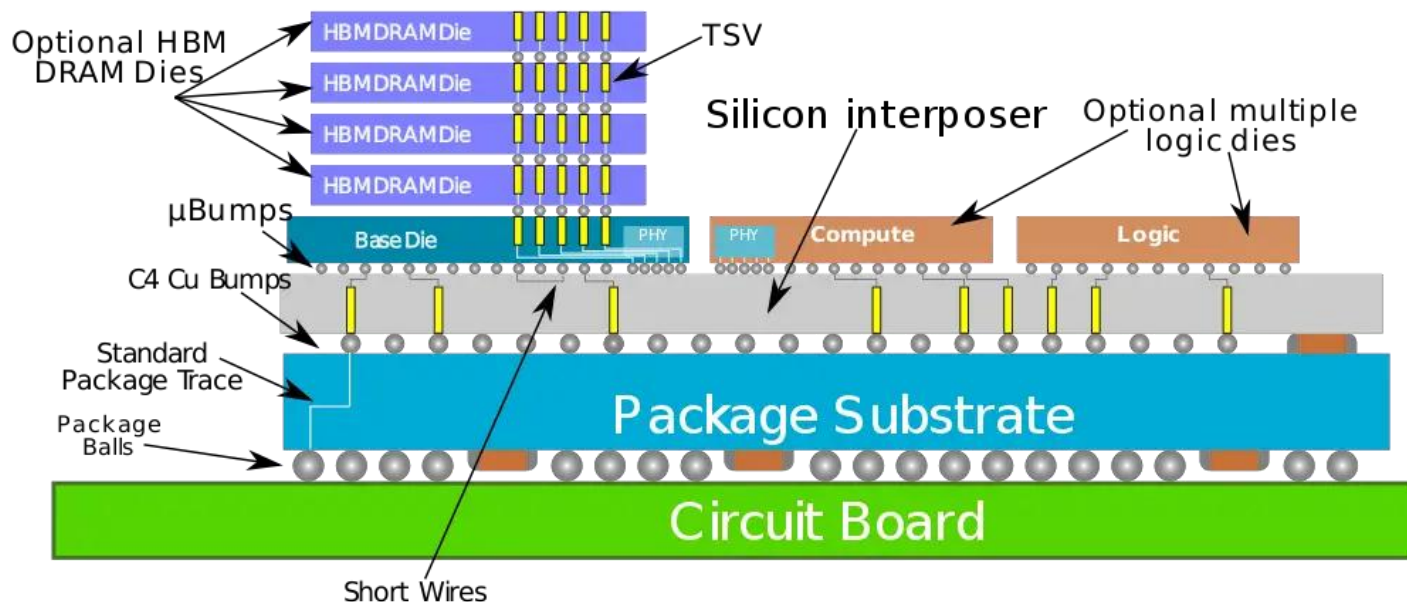
图：H100通过CoWoS集成GPU和HBM



资料来源：英伟达，semianalysis，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：CoWoS和TSV示意图

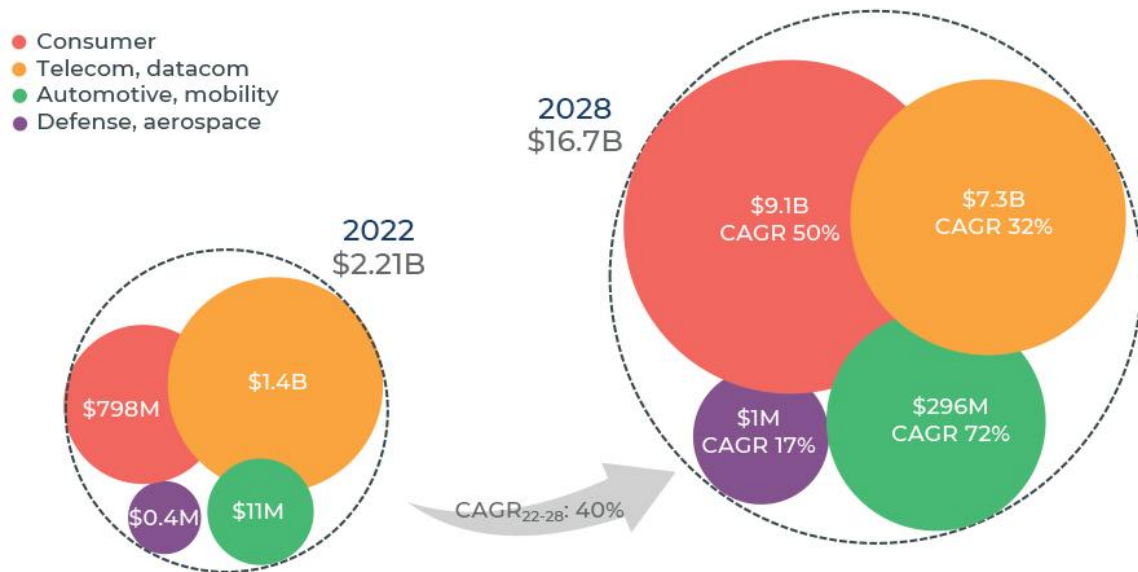


资料来源：wikichip，国信证券经济研究所整理

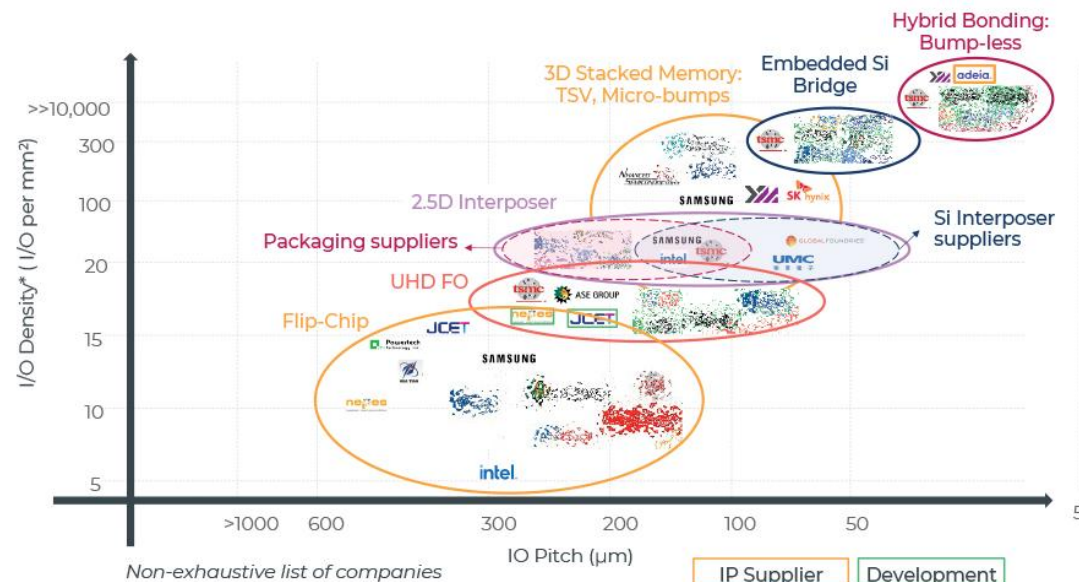
# 先进封装接力摩尔定律为算力和存力芯片助力

- 预计2022-2028年高性能封装市场规模CAGR达40%。根据Yole的预测，高性能封装市场规模将从2022年的22.1亿美元增长至2028年的167亿美元，CAGR达40%。
  - 按终端市场来看，2022年最大的下游是通讯/数据中心，占比超过60%；随着消费电子（含手机）以50%的CAGR增长，到2028年将成为第一大下游；2022-2028年CAGR最高的下游是汽车，达72%。
  - 按技术来看，3D堆叠存储（包括HBM，3DS，3D NAND）占比最大，预计2028年贡献70%；2022-2028年增速较高的是3D SoC、有源硅中介层、嵌入式硅桥和3D NAND堆栈。

图：2022-2028年高性能封装各应用市场规模预测



图：2023年各封装技术玩家



## ● 算力

- SoC设计：海光信息、寒武纪、龙芯中科
- IP授权：芯原股份
- 算力辅助芯片：杰华特、晶丰明源、芯朋微

## ● 存力

- 运力芯片：澜起科技
- 存储模组：江波龙、德明利、佰维存储
- 存储设计：兆易创新、北京君正
- 存储生产链：中微公司、北方华创、拓荆科技、雅克科技、鼎龙股份、沪硅产业、深科技

## ● 先进封装

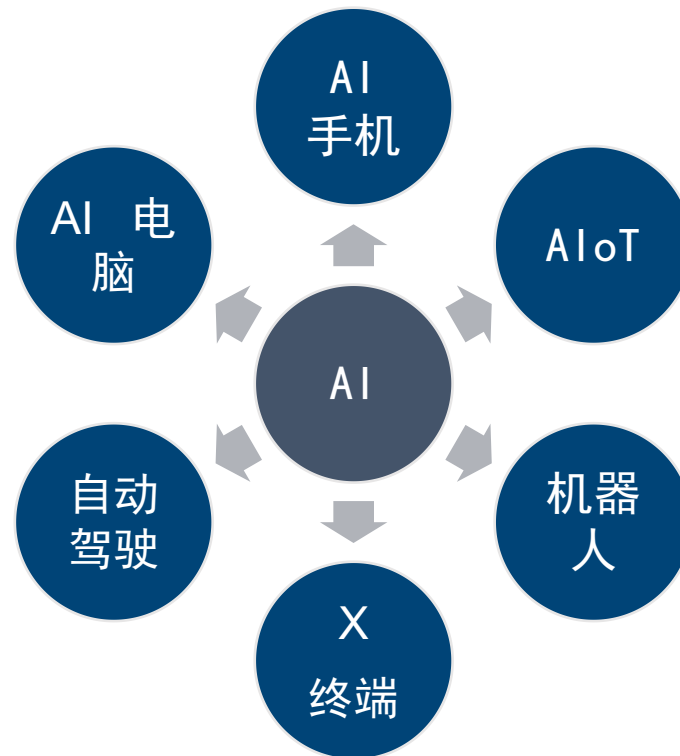
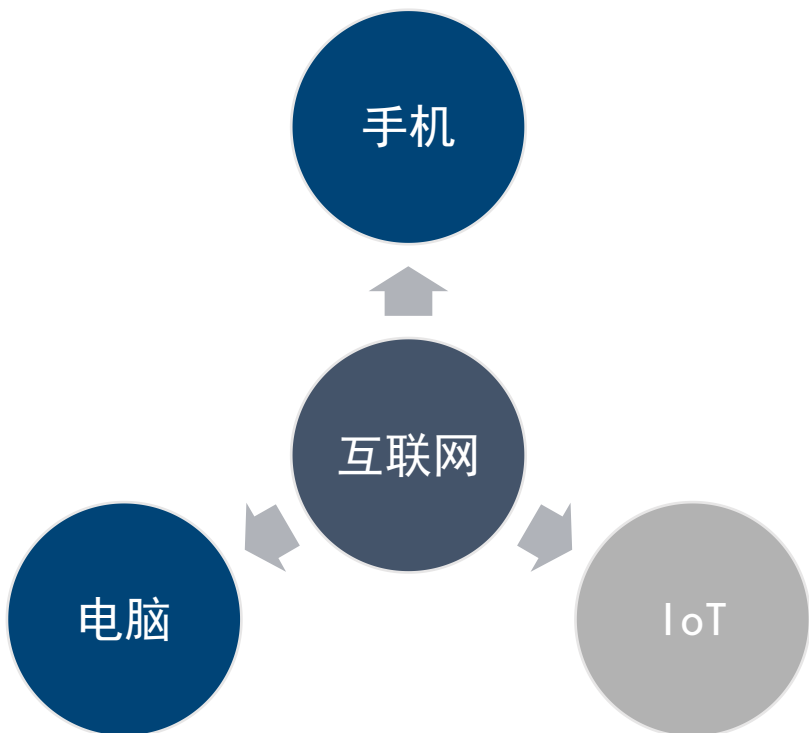
- 封测厂：长电科技、通富微电、华天科技
- IP授权：芯原股份

## AI应用落地对半导体的推动将更为长远



# AI应用落地对半导体的推动将更为长远

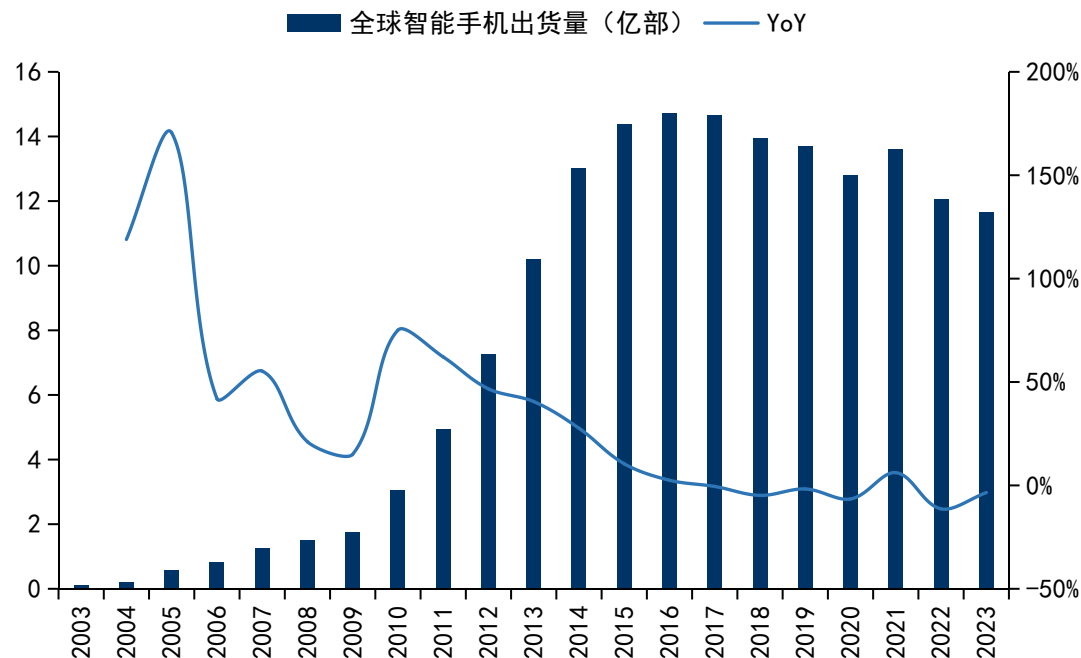
- 2023年英伟达CEO黄仁勋提出“我们正处于AI的iPhone时刻”；2024年又提出AI已到“Tipping Point”。
- 我们认为，AI对人类的影响更甚于过去的通讯技术、互联网技术。在AI基建和大模型算法的良性循环下，我们认为AI的应用将逐渐落地，从特定人群使用走向千家万户。通讯技术催生出了手机，互联网技术催生出了电脑和智能手机，AI除了赋能现有终端外，更可能使期待已久的人形机器人、自动驾驶汽车成为现实，或者催生未知终端，赋能基础研究。



# 2023年全球智能手机出货量11.6亿部

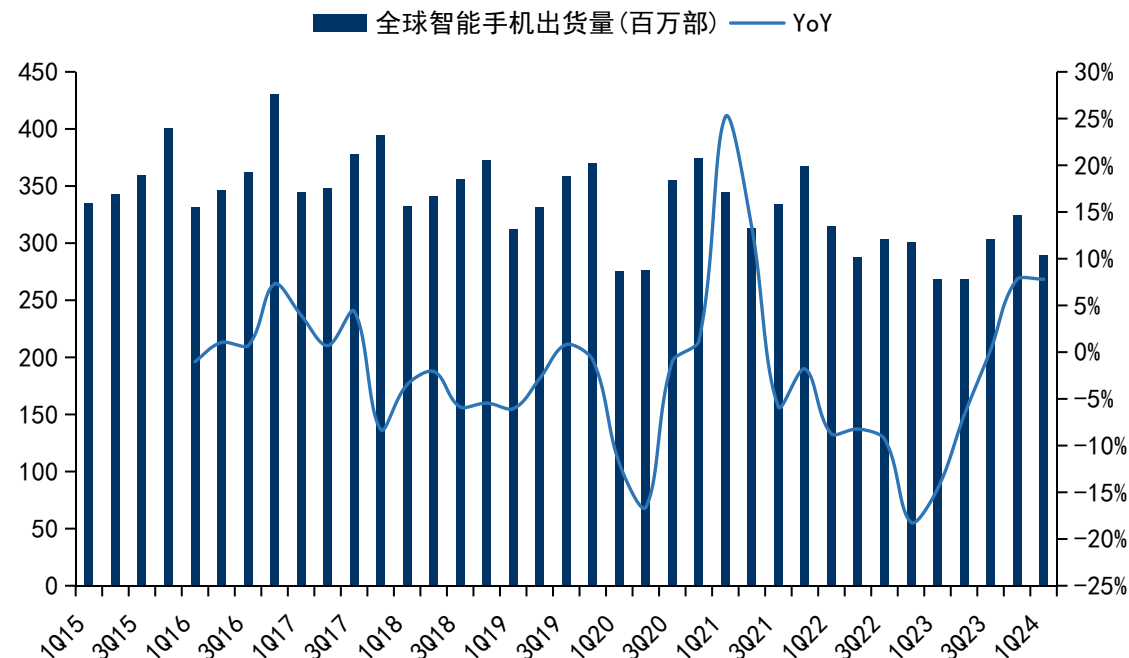
- 2023年全球智能手机出货量11.6亿部，同比减少3%。根据IDC的数据，全球智能手机在2015年后结束高增长，2017-2023年间仅2021年实现同比6%的增长，2023年同比减少3%至11.6亿部。
- 1Q24全球智能手机出货量同比增长8%至2.89亿部。从短期季度数据来看，经过4Q22和1Q23两位数同比减少后，2Q23同比降幅明显收窄，3Q23至1Q24连续三个季度同比增长，1Q24出货量为2.89亿部，同比增长8%，环比减少11%。

图：全球智能手机年出货量及增速



资料来源：IDC，国信证券经济研究所整理

图：全球智能手机季度出货量及增速



资料来源：IDC，国信证券经济研究所整理

# 2024年是AI手机元年

- 2011年10月，苹果在iPhone 4S上推出第一个人工智能作品Siri。之后，AI在手机中的应用主要集中在语音助手、人脸识别、拍照美化等，模型参数量通常不超过1000万。
- 2023年大模型在手机端运行成为可能。2022年发布的ChatGPT参数量达17.5B，远超手机的能力。2023年，可在手机端运行的模型出现，比如Meta 2月发布的大语言模型LLaMA包括7B/13B/33B/65B 4种参数量；谷歌5月发布的大语言模型PaLM2也包含4种不同参数量，其中参数量最小的Gecko模型可以在手机上运行，即使不联网也能正常工作，且运行速度足够快；3月，高通利用骁龙平台的AI软件栈，首次在Android智能手机内运行了参数超过10亿的Stable Diffusion模型。
- 2024年是AI手机元年。OPPO首席产品官表示“AI手机将是继功能机、智能机之后，手机行业的第三个重大的变革阶段”。集成大模型的手机将打破传统应用壁垒，成为用户真正的智能助手。

图：2024年发布的部分AI手机

厂商	产品名称	参考图片	处理器/内存参数	主要AI功能	发布时间
三星	Galaxy S24系列		处理器：骁龙8Gen3 for Galaxy 内存：8GB/12GB	三星AI大模型加持下的首个手机系列，采用“端云结合”的方式，所有的非本地化处理的数据，都会在上云之前，通过端侧模型的辅助，剔除个人信息后再交由云端处理，保证用户隐私的安全。 具有通话实时翻译、即圈即搜、图生图、音生文等能力。	2024. 1. 18
小米	小米14 Ultra		处理器：骁龙8Gen3 内存：12GB/16GB	接入首个AI大模型计算摄影平台Xiaomi AISP，通过全面整合CPU、GPU、NPU和ISP算力，该平台可实现60TOPS的计算能力，提供超级抓拍、超级底片、超级变焦功能。	2024. 2. 23
荣耀	荣耀Magic6系列		处理器：骁龙8Gen3 内存：12GB/16GB	首发搭载自研70亿参数端侧平台级AI大模型“魔法大模型”，提供智慧成片、一语查图、任意门的“一拖日程”等功能。	2024. 1. 11
OPPO	OPPO Find X7系列		处理器：天玑9300/骁龙8Gen3 内存：12GB/16GB	搭载全球首个端侧应用的70亿参数安第斯智能大模型，提供AIGC消除、AI大模型通话摘要等功能。	2024. 1. 8

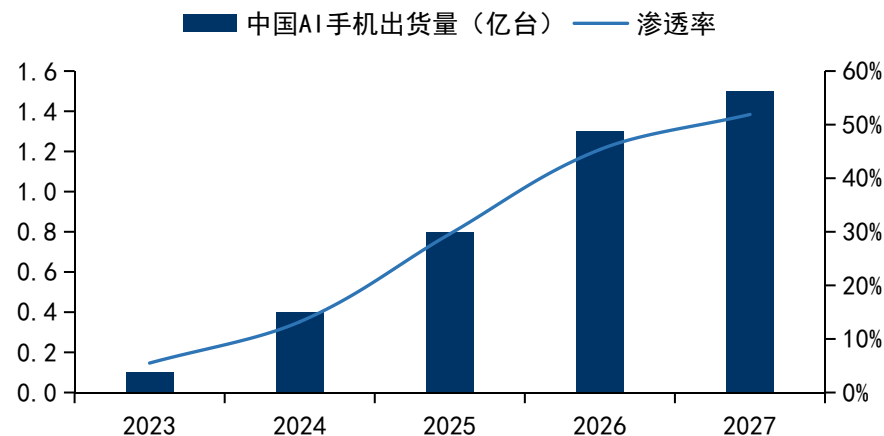
资料来源：各公司官网，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

# AI手机的渗透率将快速提升

- AI手机的用户价值：自在交互、智能随心、专属陪伴、安全可信的个人化助理。
- 预计AI手机渗透率将快速提升。IDC 预计2024年全球新一代AI手机的出货量将达到1.7亿部，约占智能手机整体出货量的15%。其中中国市场新一代AI手机在2024年的出货量将达到3700万台，2027年达到1.5亿台，所占市场份额超过50%。

图：中国AI手机市场预测



资料来源：IDC，国信证券经济研究所整理

图：AI手机的用户价值

## 自在交互：图文多模态的能力与全域知识

智能机：提供信息供给 ⊙ AI手机：提供知识和能力的供给  
用户价值：获得此时此刻唯一正确的答案，交互更自然、更直接

## 安全可信：内容安全和隐私保护

智能机：强调隐私安全 ⊙ AI手机：除了隐私安全还强调伦理价值观对齐、幻觉消除  
用户价值：个人数据被保护，回答贴切可信赖



## 智能随心：即时意图理解和服务响应

智能机：闲聊 ⊙ AI手机：提供服务  
用户价值：一键/一句话直达服务调度

## 专属陪伴：个性化的模型微调和知识增强

智能机：基于搜索式AI的信息平台 ⊙ AI手机：基于个人知识增强的生成  
用户价值：通过学习用户使用习惯、陪伴用户成长，是可成长的AI手机

资料来源：IDC，OPPO，国信证券经济研究所整理

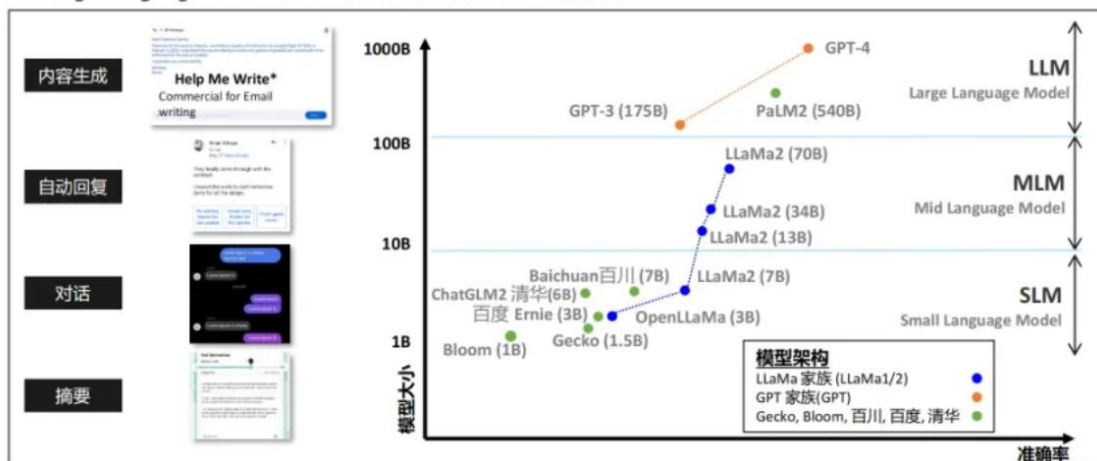
请务必阅读正文之后的免责声明及其项下所有内容

## ● AI手机为半导体带来哪些变化？

- **处理器性能提高。**手机是多模态设备，除了处理文本外，也需要处理图像、视频，算力需求会增加。2023年10月高通发布的骁龙8Gen3支持运行100亿参数端侧大模型；2023年11月联发科发布的天玑9300支持运行10亿至330亿参数的端侧大模型。
- **内存容量和带宽提高。**在端侧跑13B大模型需要13GB的内存，再加上操作系统需要的内存，预计未来AI手机的基础配置将达到16GB或更高；另外，端侧运行13B大模型需要的内存带宽达130GB/s，比现有旗舰机的配置翻倍不止。因此，内存的容量和性能均需要提升。
- **功耗管理需求增加。**虽然目前AI手机的电源管理芯片没有看到明显的需求变化，但我们认为，随着AI功能的使用增加，大模型运算会导致功耗变大，对负责能耗管理/充电管理的电源管理芯片在数量或性能方面也将提出新的需求。

图：大模型趋势

- Small Language Model (<10B)：端侧执行特定领域任务
- Mid Language Model (10B~100B)：正在端侧计算平台上兴起，用于高级人工智能（多模态，智能体）
- Large Language Model (>100B)：云端执行通用人工智能任务



资料来源：联发科，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

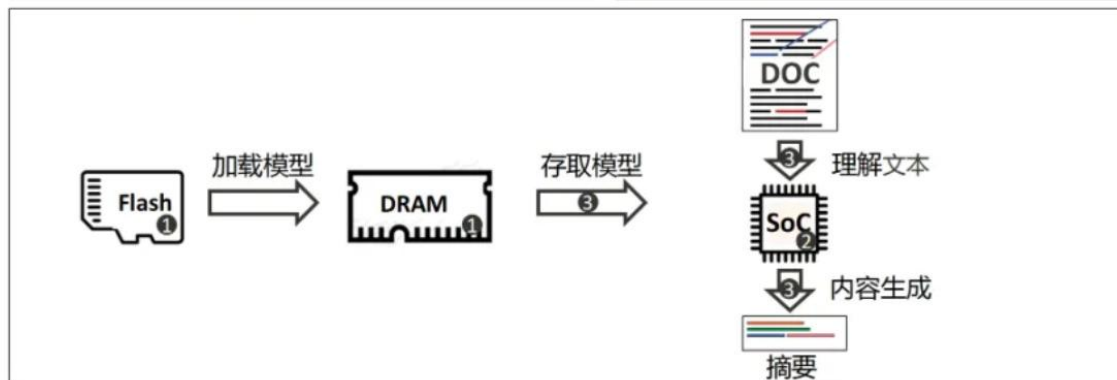
图：端侧大模型的挑战

### 端侧 13B 大模型需求

- ① 内存占用：13GB
- ② 推理算力：70TOPS for 200ms latency of 1024 words
- ③ 内存带宽：130GB/sec for 10 words/sec

### 2023旗舰智能手机规格

- ① 内存：16GB
- ② 推理算力：50 TOPS
- ③ 内存带宽：50 GB/sec

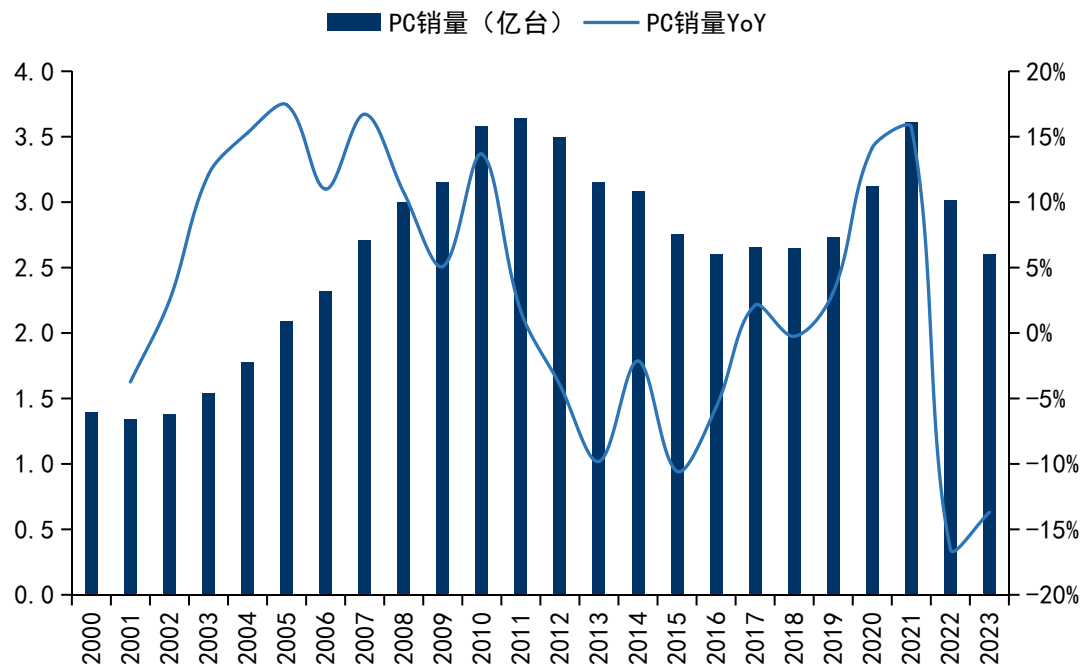


资料来源：联发科，国信证券经济研究所整理

# 2023年全球PC出货量2.6亿台

- 2023年全球PC出货量2.6亿台，同比减少14%。根据IDC数据，全球PC出货量的高点在2011年，达3.6亿台；2012-2023年仅2020/2021年由于在线办公学习需求增速达两位数；2023年同比减少14%至2.6亿台，回到了2012年以来的最底部。
- 1Q24全球PC出货量同比增长2%至0.60亿台。从季度数据来看，全球PC出货量自1Q22以来持续同比下降，1Q24同比转正，同比增长2%至0.60亿台，环比减少12%。

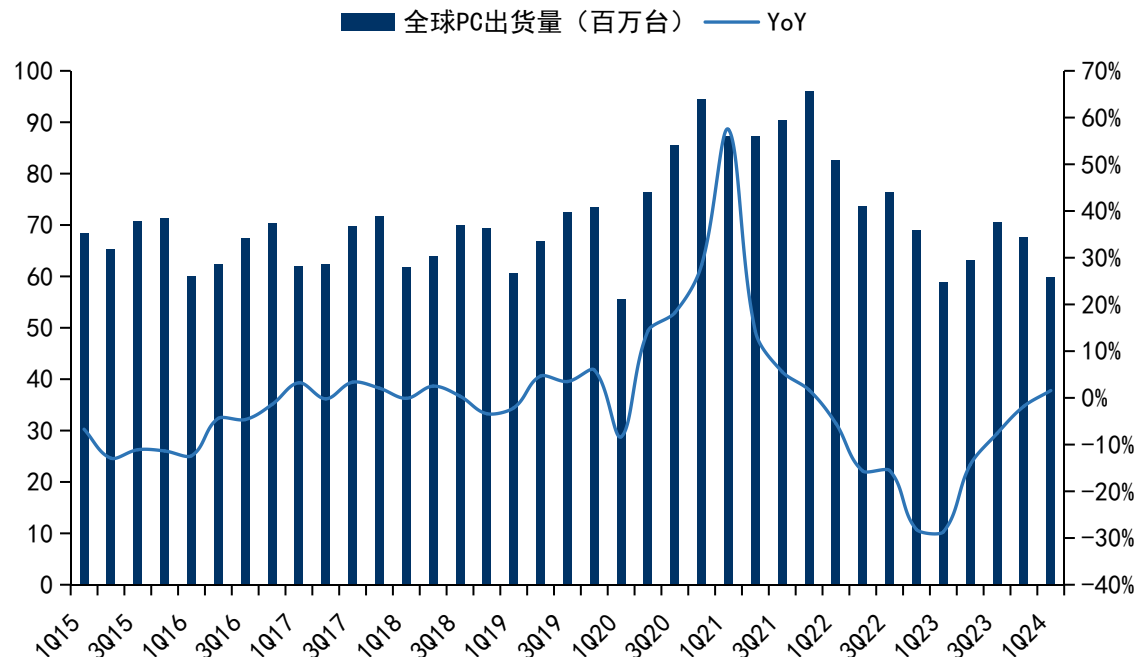
图：全球PC年出货量及增速



资料来源：IDC，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：全球PC季度出货量及增速



资料来源：IDC，国信证券经济研究所整理

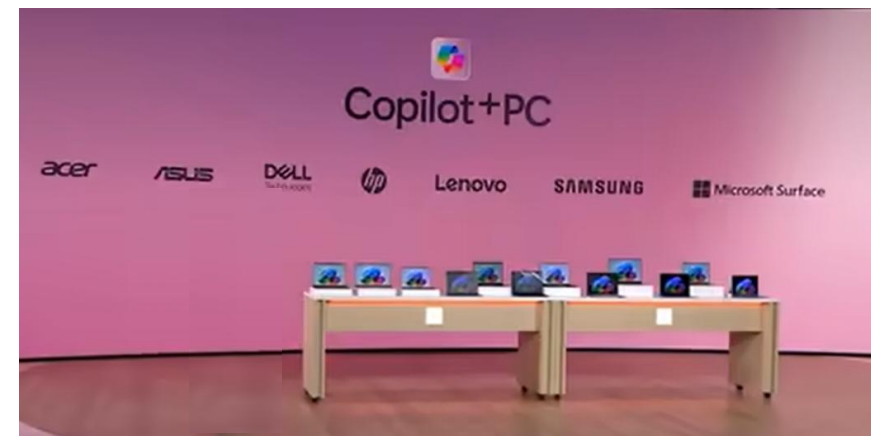
- 2023年英特尔启动AI PC加速计划，发布为AI PC而生的酷睿Ultra处理器。2023年10月，英特尔启动AI PC加速计划，将在2025年前为超过1亿台PC带来AI特性；2023年12月，英特尔在北京举办以“AI无处不在，创芯无所不及”为主题的2023英特尔新品发布会上发布了酷睿Ultra处理器，这是英特尔首款支持端侧部署AI模型的消费级处理器，除内置CPU和GPU外，还内置了英特尔专为AI处理需求研发的专业NPU模块。最大可支持20B量级的AI模型本地部署，对话、资料整理、答疑、文生图等功能都可以在不通过云端AI模型的情况下，完全由本地NPU进行运算生成。
- 2023年10月，高通推出骁龙X Elite处理器，可在本地支持13B的大模型；2024年4月推出骁龙X Plus处理器。
- 2024年也是AI PC元年。各PC厂商基于英特尔酷睿Ultra处理器、AMD锐龙处理器和高通骁龙X发布AI PC；微软在2024年5月的新品发布会上推出Windows 11 AI PC（Copilot+PC）；2024年5月苹果发布M4芯片，搭载该AI芯片的AI PC有望年底前亮相。

图：各大厂商推出AI PC

联想	荣耀	惠普	戴尔
<ul style="list-style-type: none"><li>• 2023年12月发布两款AI Ready的AI PC产品</li><li>• CES 2024上亮相10余款AI PC产品</li><li>• MWC 2024上发布新一代商务AI PC</li></ul>	<ul style="list-style-type: none"><li>• MWC 2024上发布公司首款AI PC MagicBook Pro 16。借助平台级AI能力，可以实现跨设备、跨系统、跨应用的智慧互联</li></ul>	<ul style="list-style-type: none"><li>• 2024年1月发布惠普星Book Pro 14 AI轻薄战力本</li><li>• 2024年5月推出多款商用AI PC新品</li></ul>	<ul style="list-style-type: none"><li>• 2024年1月发布三款AI PC：XPS 16、XPS 14 和 XPS 13</li></ul>

资料来源：各公司官网，国信证券经济研究所整理

图：Copilot+PC



资料来源：微软，国信证券经济研究所整理

● 人工智能的需求正在爆发，AI PC是大模型普惠于人的首选终端。用户不仅需要公共的大模型服务，更需要专属自己的个人大模型，个人大模型的普惠要求和PC的承载优势完美契合，显示着PC将再一次承载技术普惠的历史使命，成为AI普惠的首选终端。

## ● 什么是AI PC?

- 英特尔和微软合作定义的AI PC：英特尔的酷睿Ultra处理器，该处理器内置神经处理器单元（NPU），用于处理AI工作负载；微软的人工智能聊天机器人Copilot；以及微软设置的“Copilot key”键，用来唤醒Copilot体验。
- 联想提出的AI PC五大核心特征：1) 内嵌个人智能体，实现多模态自然语言交互；2) 内嵌由本地大模型和个性化本地知识库构成的个人大模型；3) 全面标配CPU+GPU+NPU的本地混合AI算力；4) 开放连接丰富的AI应用生态；5) 提供设备级的本地化个人数据和隐私保护方案。

图：英特尔和微软合作定义的AI PC



资料来源：英特尔，微软，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：AI PC五大核心特征



资料来源：联想，IDC，国信证券经济研究所整理



# AI PC提高生产力，渗透率快速提升

- **AI PC的用户价值：**通用场景下的个性化服务，提供即时、可靠的服务响应，更低的大模型使用成本，可信、安全的个人数据和隐私保障。相比于手机，PC在工作中承担了更重要的角色，AI PC可以显著提高生产力，比如AI PC能够更准确的理解用户的创作意图，掌握用户的历史创作习惯及个人风格，并通过调用多种模型和应用、互联网公共资源等方式共同完成任务创作。
- **作为生产力工具，AI PC的渗透将快于AI手机。**IDC预计AI PC在中国PC市场中新机的装配比例将在未来几年中快速攀升，2024年超过50%，2027年达到85%。

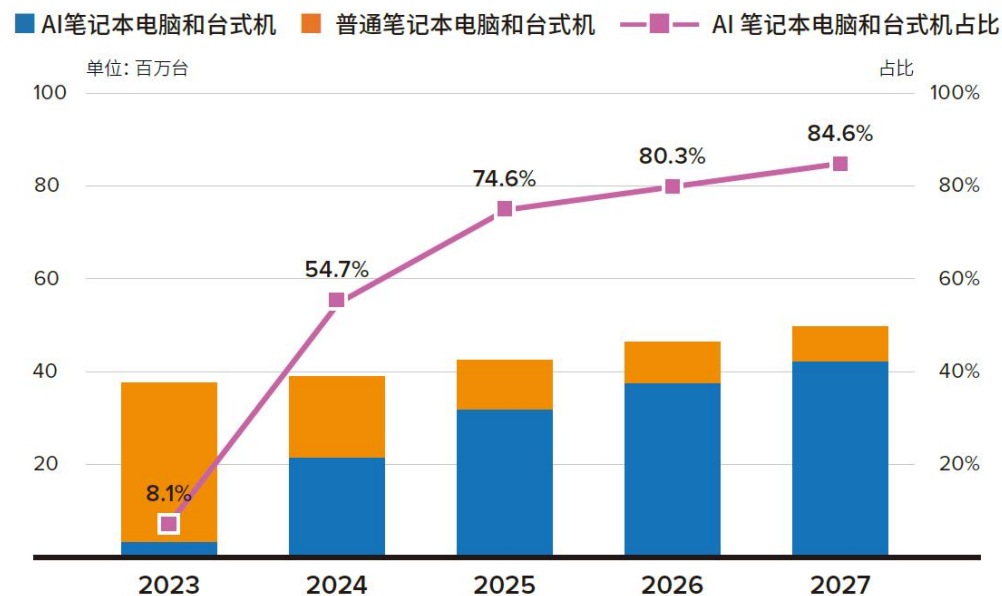
图：AI PC通用场景下的个性化服务

	工作	学习	生活
个性创作	<ul style="list-style-type: none"><li>会议材料准备</li><li>会议总结和纪要</li><li>专业 PPT/Word/Excel...</li></ul>	<ul style="list-style-type: none"><li>AI 课堂笔记和记录</li><li>文献翻译和总结</li><li>...</li></ul>	<ul style="list-style-type: none"><li>游戏攻略</li><li>AI 游记</li><li>...</li></ul>
秘书服务	<ul style="list-style-type: none"><li>个人日程表</li><li>同声传译</li><li>...</li></ul>	<ul style="list-style-type: none"><li>个人课程表</li><li>选课和提醒</li><li>...</li></ul>	<ul style="list-style-type: none"><li>AI 旅行计划</li><li>AI 实时游戏指导</li><li>...</li></ul>
设备管家	<ul style="list-style-type: none"><li>主动调优</li><li>专业模式</li><li>...</li></ul>	<ul style="list-style-type: none"><li>智能防护</li><li>学习模式</li><li>...</li></ul>	<ul style="list-style-type: none"><li>智能互联</li><li>游戏模式</li><li>...</li></ul>

资料来源：IDC，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：中国AI PC市场规模及占比预测

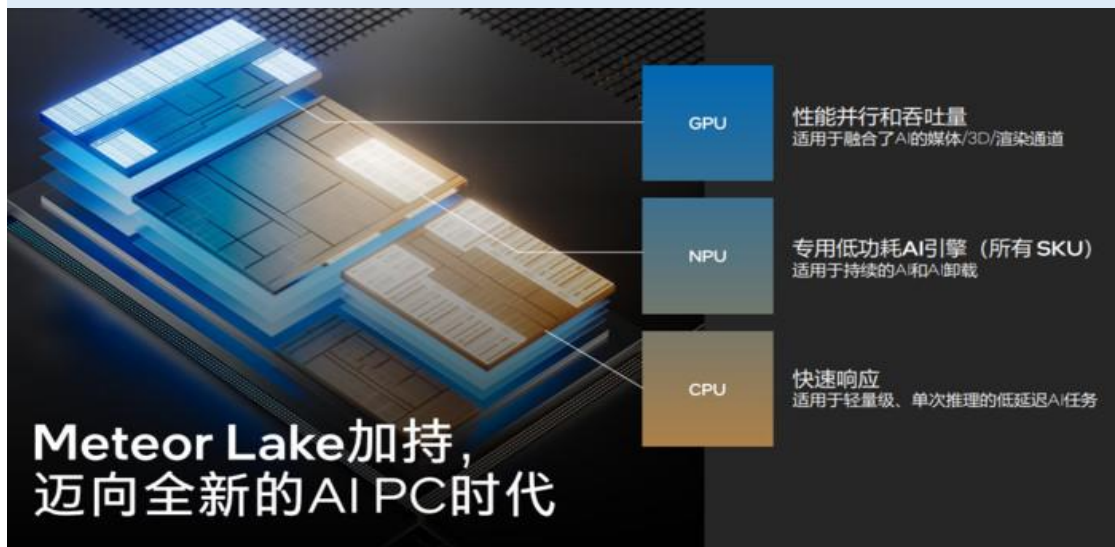


资料来源：IDC，国信证券经济研究所整理

## ● AI PC为半导体带来哪些变化？

- **处理器集成CPU、GPU、AI引擎。**为了满足AI端侧应用，目前AI PC处理器基本采用CPU+GPU+NPU（AI引擎）架构，其中CPU具有快速响应优势，适合运行较小工作负载并实现低延迟；GPU具有高吞吐量优势，适合需要并行吞吐量的大型工作负载；NPU具有高能效比优势，能够以低功耗处理持续存在、大量使用的AI工作负载，提高效率。预计未来处理器架构将随着需求而演变，性能、价格、能耗是重点。
- **存储容量和带宽提高。**大模型运行需要更大的内存和带宽，而个性化需求和独立性需要更多的本地数据支撑，预计DRAM和NAND的需求均将明显增加。
- **功耗管理需求增加。**随着AI PC处理器的算力提高，对应的大电流DCDC芯片需求将增加；为了控制整体的能耗，提高使用时间，对其他配套电源管理芯片在数量或性能方面也将提出新的需求。

图：英特尔AI PC酷睿处理器



资料来源：英特尔，国信证券经济研究所整理

图：AMD AI PC锐龙处理器



资料来源：AMD，国信证券经济研究所整理

# 真正智能的AIoT有望激发需求

- 大模型加速AI发展，真正智能的AIoT有望激发消费者需求。2017年“万物智能·新纪元AIoT未来峰会”在国内首次正式提出AIoT概念，天猫精灵、小爱同学等智能音箱也在同年推出并畅销一时。但由于智能程度有限，最终普及和使用未达最初预期。随着大模型的发展，真正智能的AIoT将提升消费者体验，有望重新激发需求。2023年11月华为发布的全屋智能5.0便是AIoT应用场景之一。
- AIoT终端虽然单机半导体价值含量可能不高，但当普及终端数量足够多时，也将对半导体带来不少增量。

图：华为全屋智能5.0



资料来源：华为，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

## ● AI手机

- 基带芯片：翱捷科技
- 射频芯片：卓胜微、唯捷创芯、慧智微
- 电源管理芯片：艾为电子、南芯科技、力芯微、希荻微、圣邦股份、帝奥微、美芯晟
- CIS：韦尔股份、格科微、思特威
- 显示驱动芯片：天德钰、新相微、韦尔股份、格科微、晶合集成（代工）、颀中科技（封测）、汇成股份（封测）
- 存储：江波龙、德明利、佰维存储、聚辰股份、普冉股份

## ● AI PC

- 处理器：海光信息、龙芯中科
- 存储：江波龙、德明利、佰维存储
- 内存接口芯片：澜起科技
- 模拟芯片：杰华特、帝奥微、艾为电子、圣邦股份、赛微微电、希荻微
- 显示驱动芯片：天德钰、新相微、格科微、晶合集成（代工）、颀中科技（封测）、汇成股份（封测）

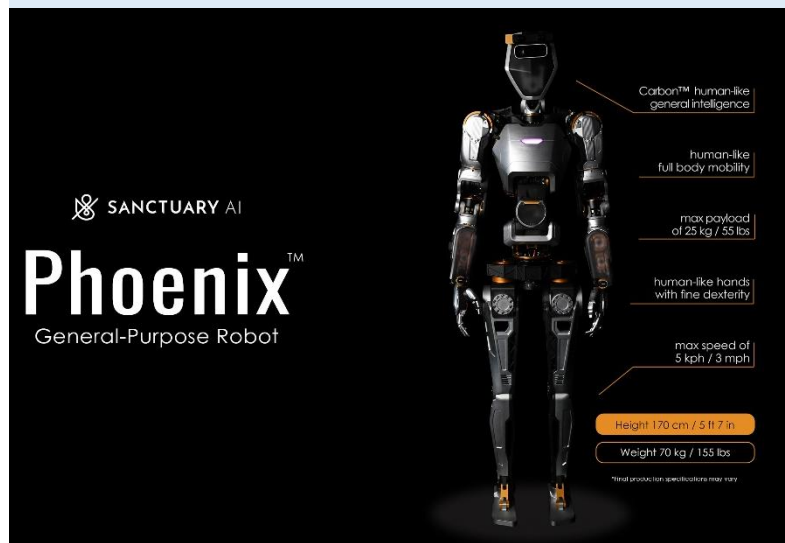
## ● AIoT

- SoC：晶晨股份、瑞芯微、全志科技、恒玄科技、中科蓝讯
- 蓝牙芯片：恒玄科技、中科蓝讯
- MCU：乐鑫科技、芯海科技、兆易创新、中颖电子
- 物联网芯片：翱捷科技
- 存储芯片：兆易创新、东芯股份、普冉股份
- 模拟芯片：圣邦股份、艾为电子、帝奥微、南芯科技、希荻微、英集芯、赛微微电
- 显示驱动芯片：天德钰、新相微、格科微、晶合集成（代工）、顾中科技（封测）、汇成股份（封测）

● 具身智能有望成为AI发展的下一个浪潮。具身智能是指一种智能系统或机器能够通过感知和交互与环境进行实时互动的能力，而一个能理解、推理并与物理世界互动的智能系统需要一具“肉身”，人形机器人是目前普遍认可的载体。在AI大模型赋能下，人形机器人商业化落地速度进一步提升。

- 2023年5月Sanctuary AI发布了世界上第一个能以人类的速率自主完成任务的机器人Phoenix，Phoenix由AI系统Carbon驱动，可以将自然语言转化为现实世界中的行动，像人一样思考和行动，能够完成十几个不同行业的数百项任务。
- 特斯拉2024年2月发布的Optimus人形机器人视频显示，其与2023年12月相比，步伐更加稳健，步速达到0.6m/s，提高了30%以上。
- 2024年3月Figure发布了自己第一个OpenAI大模型加持的机器人demo，Figure 01可以听从人类命令递给人类苹果等。

图：机器人Phoenix



资料来源：Sanctuary AI，国信证券经济研究所整理

图：机器人Optimus



资料来源：特斯拉，国信证券经济研究所整理

图：机器人Figure 01



资料来源：Figure，国信证券经济研究所整理

# 工信部印发《人形机器人创新发展指导意见》

● 2023年10月工信部印发《人形机器人创新发展指导意见》，指出人形机器人集成人工智能、高端制造、新材料等先进技术，将深刻变革人类生产生活方式，重塑全球产业发展格局。

- 到2025年，人形机器人创新体系初步建立，“大脑、小脑、肢体”等一批关键技术取得突破，**确保核心部组件安全有效供给**。整机产品达到国际先进水平，并实现批量生产，在特种、制造、民生服务等场景得到示范应用，探索形成有效的治理机制和手段。培育2-3家有全球影响力的生态型企业 and 一批专精特新中小企业，打造2-3个产业发展集聚区，孕育开拓一批新业务、新模式、新业态。
- 到2027年，人形机器人技术创新能力显著提升，形成安全可靠的产业链供应链体系，构建具有国际竞争力的产业生态，综合实力达到世界先进水平。产业加速实现规模化发展，应用场景更加丰富，相关产品深度融入实体经济，成为重要的经济增长新引擎。
- 突破关键技术：打造人形机器人“大脑”和“小脑”；突破“肢体”关键技术；健全技术创新体系。

图：关键技术攻关



资料来源：工信部，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

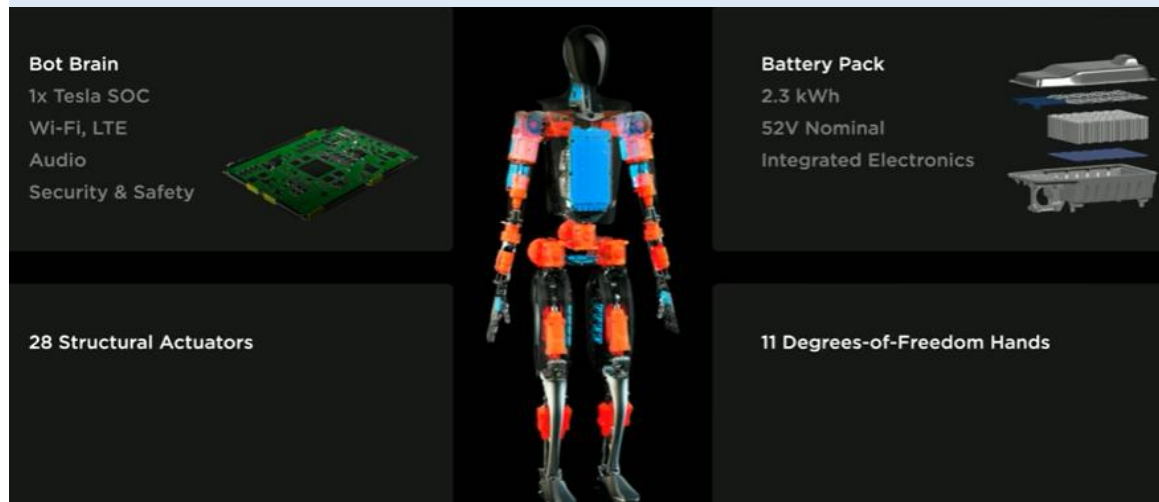
图：重点产品和部组件攻关



资料来源：工信部，国信证券经济研究所整理

- 人形机器人中的半导体产品。人形机器人与现实世界交互的方式与人类相似。
  - 机器人**大脑（人脑）**。运行AI算法的SoC为人形机器人提供智力，随着大模型的发展，人形机器人的智力将发展迅速。Optimus使用的是**特斯拉SoC**，沿用**FSD算法**。
  - **通过语音、视觉感知外界（人眼/耳）**。人形机器人靠声音和画面获得外界信息，因此需要**视频芯片、音频芯片**。另外，人形机器人也有**联网需求**，因此需要**Wi-Fi、蜂窝联网芯片**。
  - **依靠执行器作用于外界（人关节/触觉）**。人形机器人最终的目的是要帮助人类完成工作或服务于人类，需要完成行走、抓取、搬运等动作，这需要依靠执行器帮助完成。Optimus（2022）拥有28个执行器，手部有11个自由度，手指可以独立运动。执行器需要**力传感器、位置传感器等各类传感器**来获得运动信息等，也需要**主控芯片、驱动芯片、模拟芯片、功率器件**等以快速精确完成任务。

图：Optimus



资料来源：特斯拉，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

图：Optimus-Gen 2的每个手指上都有触觉传感器



资料来源：特斯拉，国信证券经济研究所整理



# AI加速自动驾驶的到来

## 国外自动驾驶发展历程



资料来源：英伟达，特斯拉，谷歌，国信证券经济研究所整理

请务必阅读正文之后的免责声明及其项下所有内容

# AI加速自动驾驶的到来

## 国内自动驾驶发展历程

1980 国家立项“遥控驾驶的防核化侦察车”项目

1990s 八五期间，ATB-1无人车研制成功，这是我国第一辆能够自主行驶的测试样车，行驶速度21km/h

2009 首届中国“智能车未来挑战赛”在西安举行



2006 一汽红旗HQ3无人驾驶车研制成功，2011年完成从长沙至武汉286km高速全程无人驾驶试验

2006



“军交猛狮III号”完成114km京津高速行驶，车顶安装复杂的**视听感知系统**，车内装有两台计算机和一台备用计算机组成的**执行系统**来处理视听感知系统获得的信息



2012

2023



华为问界M5智驾版车型上首次推出高阶智能驾驶系统ADS 2.0（ADS 1.0版本在2021年推出）；  
**基于盘古大模型3.0推出自动驾驶行业大模型**

2021



发布全新升级的自动驾驶出行服务平台“萝卜快跑”

2017



推出自动驾驶开放平台Apollo

2015



国内首辆无人驾驶客车路测完成

2013

**百度**组建自动驾驶研发团队

# AI加速自动驾驶的到来

● 自动驾驶等级分为L0-L5。根据自动驾驶标准权威机构SAE发布的自动驾驶相关标准，自动驾驶从L0到L5共六个等级。其中L5是完全自动驾驶，也是自动驾驶的终极形态。目前许多汽车已经达到L2级，AI大模型的发展有望加速L5的到来。

图：代表车型



图：SAE自动驾驶等级（SAE J3016）

	SAE LEVEL 0	SAE LEVEL 1	SAE LEVEL 2	SAE LEVEL 3	SAE LEVEL 4	SAE LEVEL 5
驾驶员坐在驾驶座上需要做些什么？	无论何时使用驾驶辅助功能，您必须处于驾驶状态 即使双脚离开踏板，也没有控制方向盘			当使用自动驾驶功能，您无需驾驶汽车 您仅仅是坐在“驾驶座”上		
	您必须时刻观察各种情况 您需要主动制动、加速或者转向，确保安全			当功能请求时， 您必须驾驶汽车	这些自动驾驶功能 不需要您接管驾驶	
具有哪些功能特征？	以下是辅助驾驶功能			以下是自动驾驶功能		
	仅提供警告 以及瞬时辅助	能够制动、 加速 或 转向， 辅助驾驶	能够制动、 加速 和 转向， 辅助驾驶	可以在有限的条件下 驾驶车辆， 除非满足所有条件， 否则不会运行		可以在 任何条件下 驾驶车辆
功能示例	<ul style="list-style-type: none"> <li>自动紧急制动</li> <li>视觉盲点提醒</li> <li>车身稳定系统</li> </ul>	<ul style="list-style-type: none"> <li>车道偏离修正 或</li> <li>自适应巡航</li> </ul>	<b>同时进行</b> <ul style="list-style-type: none"> <li>车道偏离修正</li> <li>自适应巡航</li> </ul>	<ul style="list-style-type: none"> <li>在交通拥堵的情况下自动驾驶</li> </ul>	<ul style="list-style-type: none"> <li>城市中“机器人出租车”</li> <li>踏板、转向装置可能无需安装</li> </ul>	<ul style="list-style-type: none"> <li>与L4相似，但是可以在任何条件下进行驾驶</li> </ul>

资料来源：华为，特斯拉，奥迪，国信证券经济研究所整理

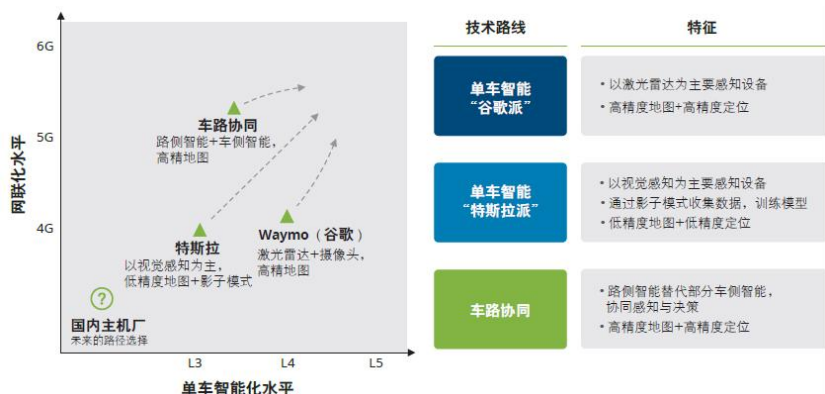
资料来源：SAE，国信证券经济研究所整理

# 自动驾驶为半导体带来增量

● 预计2030年全球自动驾驶新车渗透率将超过10%，为半导体带来增量。华为预测到2030年，中国自动驾驶新车渗透率将超20%，全球自动驾驶新车渗透率将超过10%。自动驾驶汽车作为大型移动智能终端，核心部件由传统的体现动力和操控的传动系统转向体现自动驾驶技术水平的算法和处理器芯片，智能座舱也成为差异化焦点。

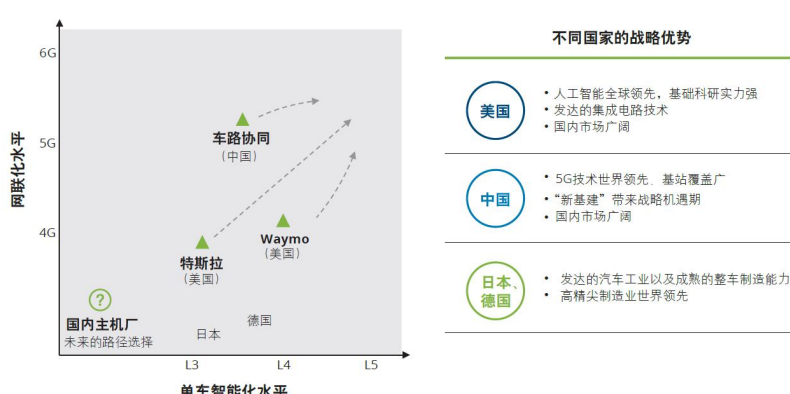
- 自动驾驶决策需要强大的算力。高性能自动驾驶芯片为自动驾驶汽车提供算力支持，并且保持低功耗，承担来自雷达系统和视觉系统的数据处理工作，在短时间内做出相应的驾驶决策。NVIDIA DRIVE Orin算力达254 TOPS。
- 通过摄像头、雷达等传感器获知外界信息。环境感知是自动驾驶的前提，摄像头、雷达等传感器的用量将增加。
- 智能座舱是自动驾驶差异化焦点。座舱是人机交互的中心，华为预计到2030年智能座舱的渗透率将超过90%，一个新的移动式智能终端生态系统逐步构建完成，为各类半导体带来需求，比如音频芯片、视频芯片、模拟芯片等。
- 数据量的增加需要更多的存储容量。根据美光的预计，L5级汽车所需DRAM和NAND分别超过L0的30倍和100倍。
- 若采用车路协同实现自动驾驶，路侧智能化将为各类半导体带来更多增量。

图：自动驾驶的技术路线



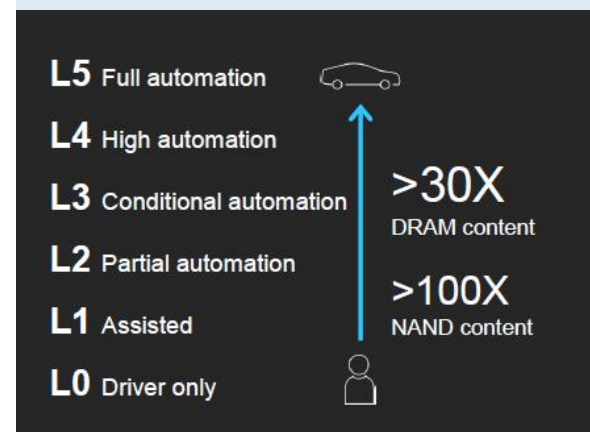
资料来源：德勤，国信证券经济研究所整理

图：不同国家的自动驾驶路径



资料来源：德勤，国信证券经济研究所整理

图：自动驾驶所需存储容量增加



资料来源：美光，国信证券经济研究所整理

## ● 人形机器人/自动驾驶汽车

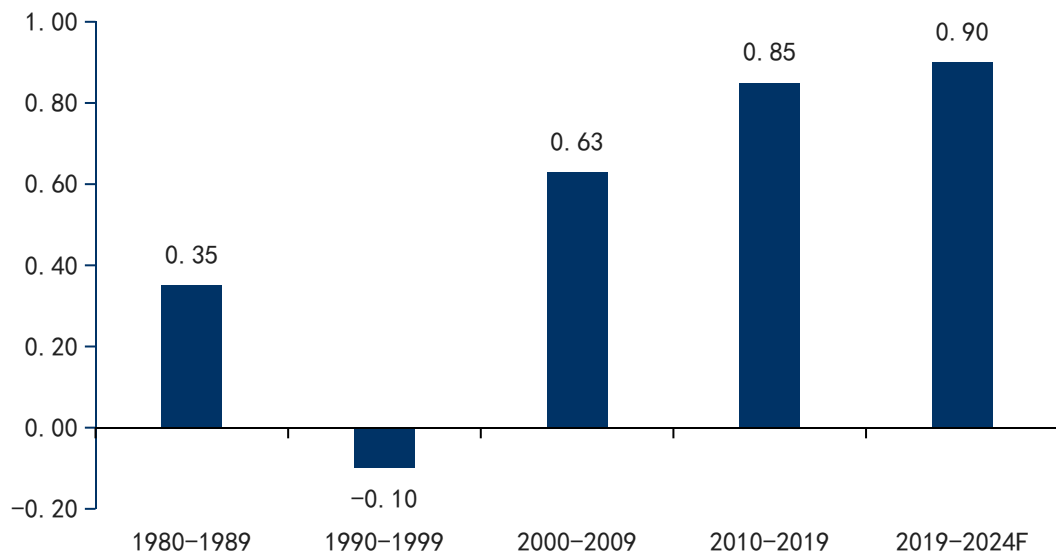
- MCU：兆易创新、峰昭科技、乐鑫科技、中颖电子、芯海科技
- 存储：北京君正、兆易创新、东芯股份、普冉股份、江波龙、德明利、佰维存储
- 模拟芯片：圣邦股份、纳芯微、思瑞浦、杰华特、芯朋微、南芯科技、艾为电子、帝奥微、希荻微
- 传感器：芯动联科、敏芯股份、纳芯微、灿瑞科技、赛微电子（代工）
- CIS：韦尔股份、格科微、思特威
- 功率器件：华润微、士兰微、时代电气、斯达半导、扬杰科技、捷捷微电、新洁能、东微半导
- 通信芯片：翱捷科技
- 射频芯片：卓胜微、唯捷创芯、慧智微

## 半导体周期进入上行阶段

# 半导体行业具有明显的周期性

- 全球半导体增速与GDP增速的相关性提高，预计2019-2024年的相关系数为0.90。根据IC Insights的数据，2000年后全球半导体增速与GDP增速的相关系数逐渐提高，预计2019-2024年将达到0.90。
- 从2011年后的历史数据来看，全球半导体销售额同比增速从峰顶到谷底一般需要4-6个季度。1Q13-4Q14景气上行（补库存，增速平稳），1Q15-2Q16景气下行（PC需求疲软，2015年PC销量减少8%）；3Q16-2Q18景气上行（手机、服务器存储容量升级等），3Q18-3Q19景气下行（存储产能大幅释放，去库存）；4Q19-4Q21景气上行（疫情增加了半导体需求，同时影响了部分供给，半导体缺货涨价），1Q22-1Q23景气下行（PC、手机需求疲软，去库存）。

图：全球GDP和IC增速的相关系数



资料来源：IC Insights，国信证券经济研究所整理

图：全球半导体季度销售额同比增速

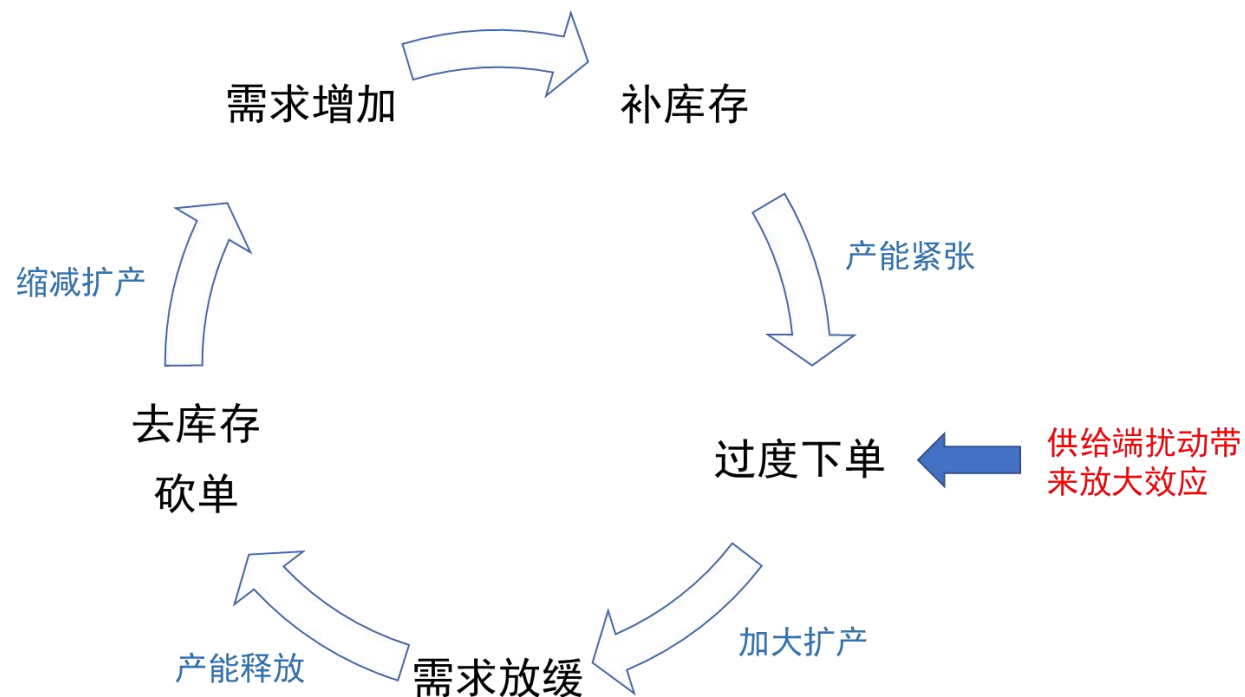


资料来源：SIA，国信证券经济研究所整理

# 半导体行业具有明显的周期性

● 半导体的周期性主要由于供需失衡，若产能出现意外受限会放大行业的波动。晶圆厂扩建一般需要1-2年才能释放产能，因此在需求增加时无法即时进行响应，同时由于晶圆产能相对刚性，在需求减少时也无法进行收缩，因此半导体行业会由于供需失衡而呈现出周期性：下游终端客户在需求增加时会进行补库存，带动晶圆产能紧张，为了保证供应抢占市场份额，终端客户存在过度下单的可能；为了满足客户的需求，晶圆厂会加大扩产；经过一两年的产能建设，下游需求可能出现放缓，客户会通过砍单去库存；晶圆产能供过于求，缩减投资。在此过程中，原有晶圆产能若由于外部因素出现供给减少，比如2016-2017年存储厂制程转换不及预期或2020年疫情导致的停产，将会放大下游重复性订单需求，从而放大行业的波动。

图：半导体行业的周期循环



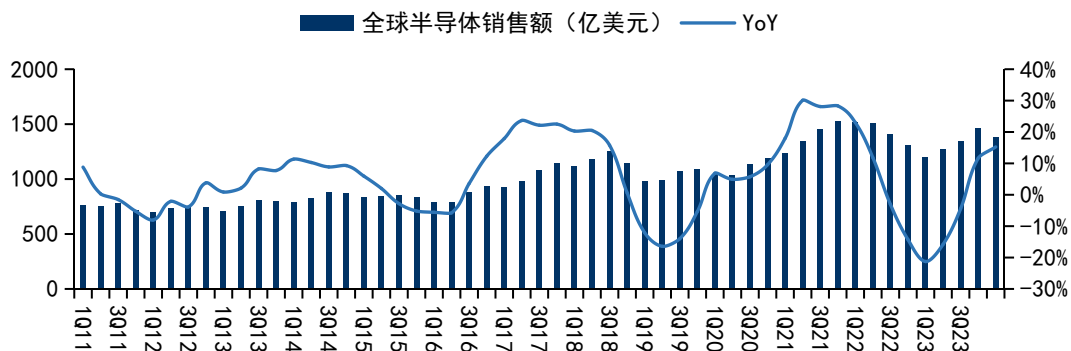
资料来源：SIA，国信证券经济研究所整理



# 全球半导体周期已进入上行阶段

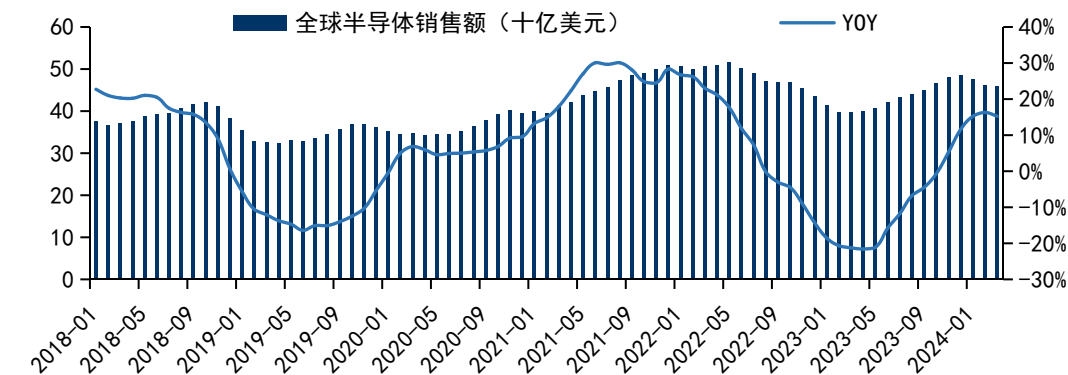
● 1Q24全球半导体销售额同比增长15%。根据SIA的数据，全球半导体季度销售额同比增速在1Q23触底，之后跌幅收窄，4Q23同比转正。1Q24全球实现销售额为1377亿美元，同比增长15.2%，环比减少5.7%；中国实现销售额为424亿美元，占全球的31%，同比增长27.4%，环比减少6.6%。

图：全球半导体季度销售额及同比增速



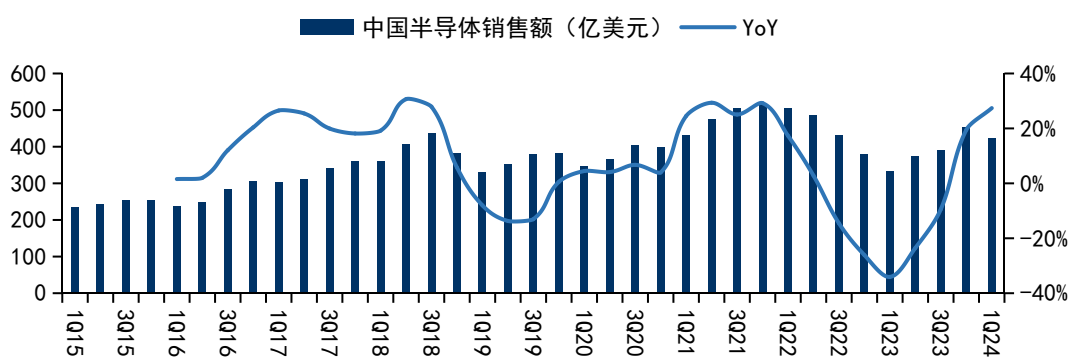
资料来源：SIA，国信证券经济研究所整理

图：全球半导体月度销售额及同比增速



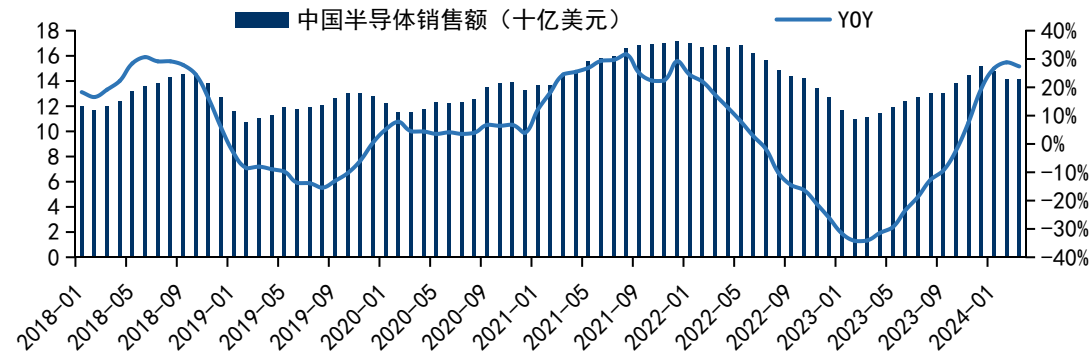
资料来源：SIA，国信证券经济研究所整理

图：中国半导体季度销售额及同比增速



资料来源：SIA，国信证券经济研究所整理

图：中国半导体月度销售额及同比增速

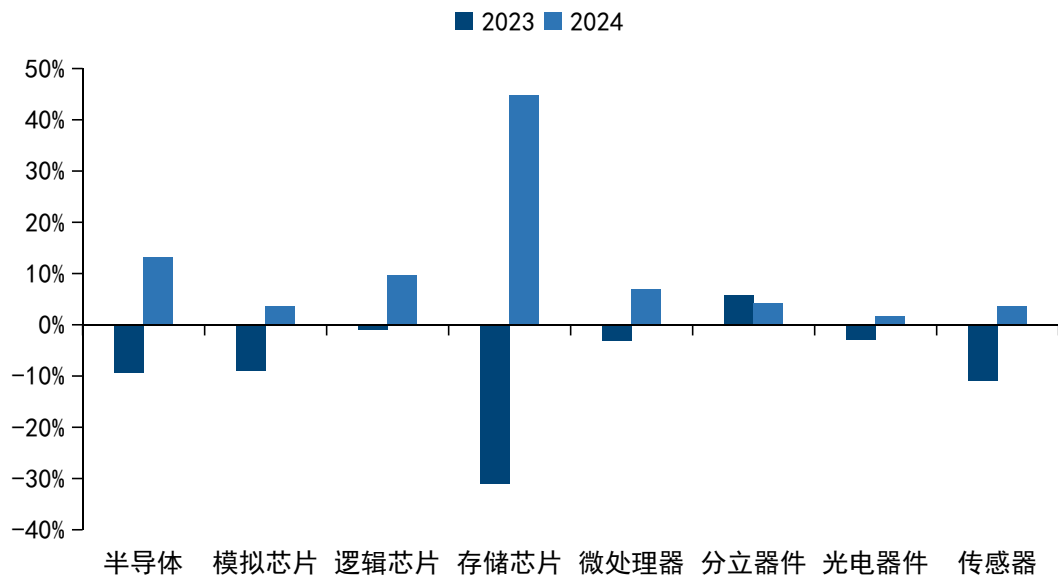


资料来源：SIA，国信证券经济研究所整理

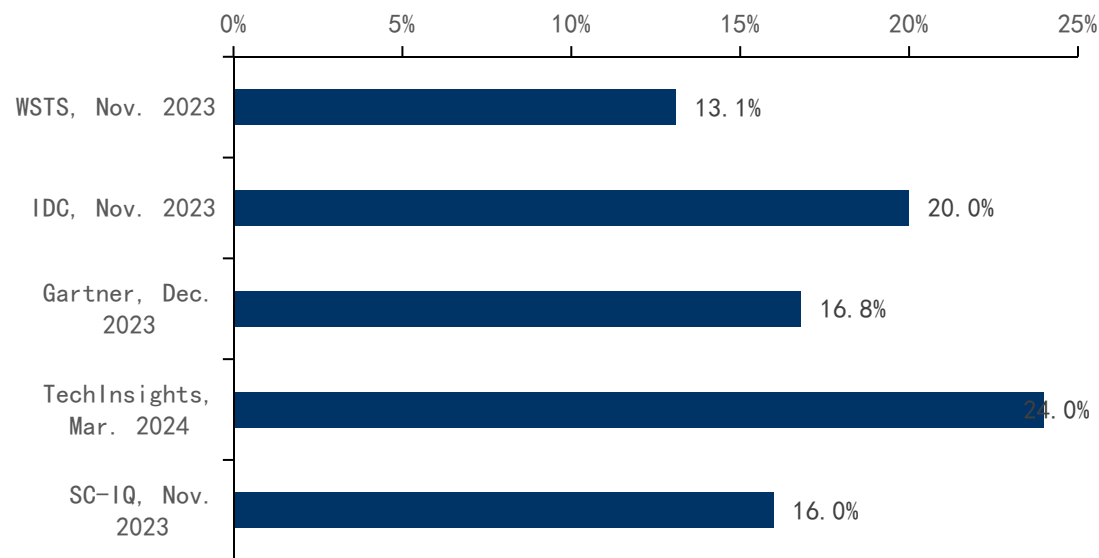
# 全球半导体周期已进入上行阶段

- **WSTS提高全球半导体销售额预测值，预计2024年将恢复增长13.1%。**根据WSTS 11月的最新预测，2023年全球半导体销售额为5201亿美元(前次预测值为5151亿美元)，同比减少9.4%；2024年将恢复增长，同比增长13.1%至5884亿美元(前次预测值为5760亿美元)。具体来看，2024年所有地区和产品线的销售额均将正增长。
- **TechInsights调高全球半导体市场展望，多家机构预计2024年增速超过10%。**除WSTS外，其他机构也预计2024年全球半导体销售额恢复增长，预计增速在10%-25%之间。其中TechInsights在3月的最新预测中上调2024年全球半导体销售额增速至24%（前次预测值为16%），将超过6500亿美元，并预计2025/2026年分别增至超过8000亿美元和接近9000亿美元。

图：2024年各半导体产品增速预测



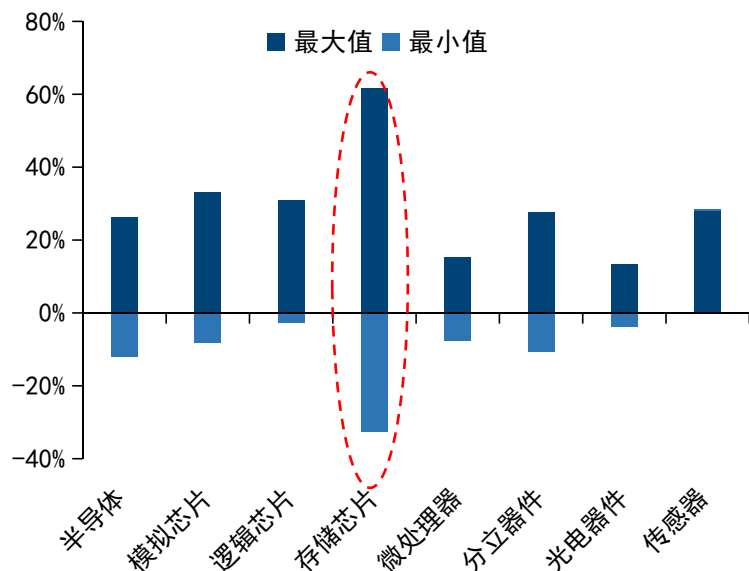
图：多家机构预计2024年全球半导体销售额增速超过10%



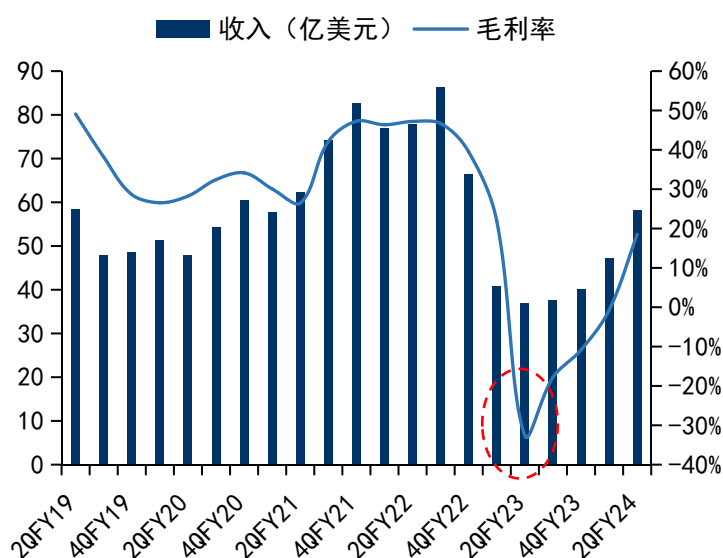
# 存储是半导体周期中弹性最大的品种

- 存储芯片是波动性最大的产品类别。根据WSTS的数据，2011-2022年全球半导体销售额增速的最大值为2021年的26.2%，最小值为2019年的-12.0%。在各子行业中，存储芯片由于其大宗产品属性明显，价格波动大，增速变动范围最大，2011-2022年增速的最大值为2017年的61.5%，最小值为2019年的-32.6%。
- 美光科技和SK海力士的季度收入和毛利率均拐头向上。SK海力士的收入和毛利率均在1Q23见底，3Q23毛利率由负转正。美光科技的收入和毛利率在2QFY23（截至2023年3月2日）见底，2QFY24（截至2024年2月29日）实现收入58亿美元（YoY +58%，QoQ +23%），毛利率回升至+18.5%。

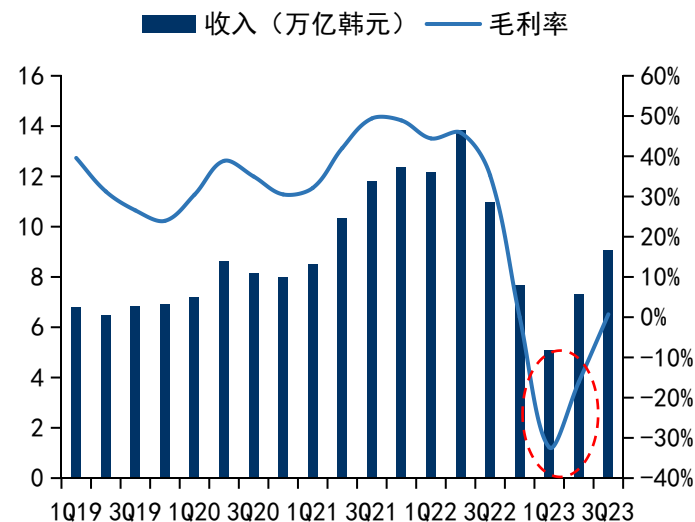
图：半导体及各子行业2011-2022年增速最大值和最小值



图：美光科技季度收入和毛利率



图：SK海力士季度收入和毛利率



资料来源：WSTS，国信证券经济研究所整理

资料来源：美光科技，国信证券经济研究所整理

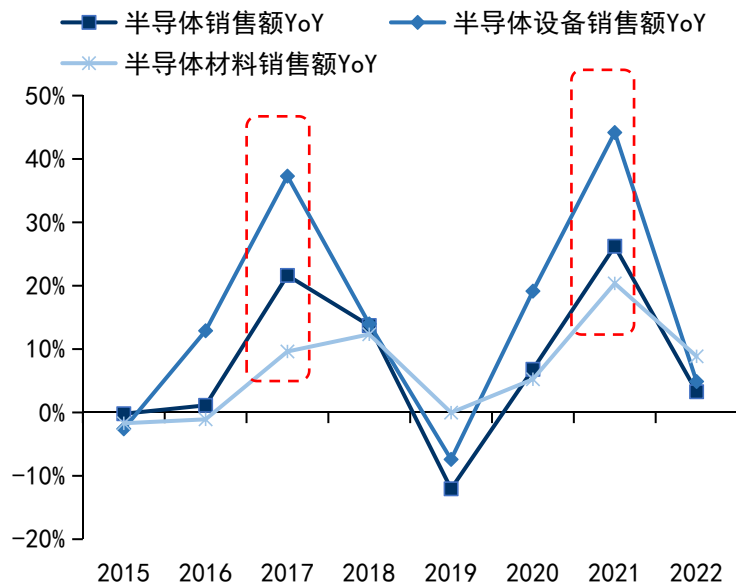
资料来源：SK海力士，国信证券经济研究所整理

# 半导体设备波动较大，半导体材料波动较小

● 从产业链环节来看，半导体设备销售额波动较大，半导体材料销售额波动较小。从2015-2022年来看，半导体设备销售额的增速波动较大，2017、2021年的增速分别为37.3%、44.2%，远高于半导体销售额增速21.6%、26.2%和半导体材料销售额增速9.6%、15.9%。从增速变化来看，半导体、半导体设备、半导体材料整体保持一致，不过半导体材料销售额波动幅度较小。

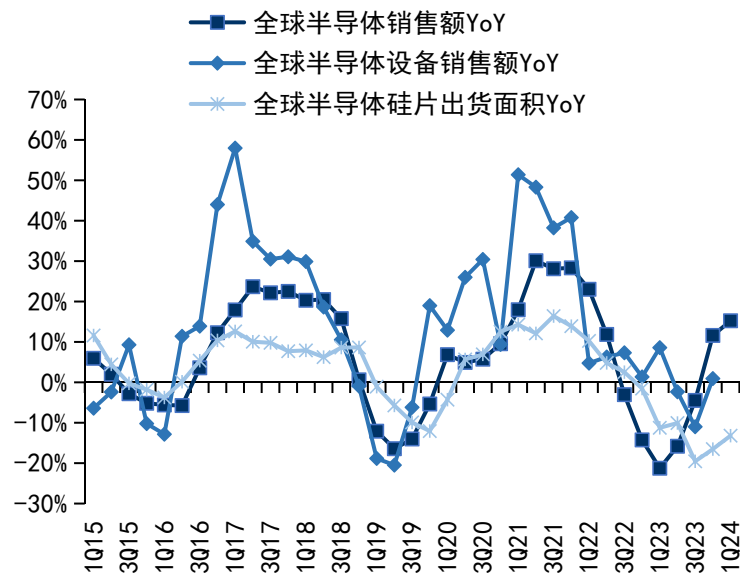
● 半导体材料跟随晶圆厂产能利用率波动，周期时间相对靠后。

图：半导体、半导体设备、半导体材料年增速



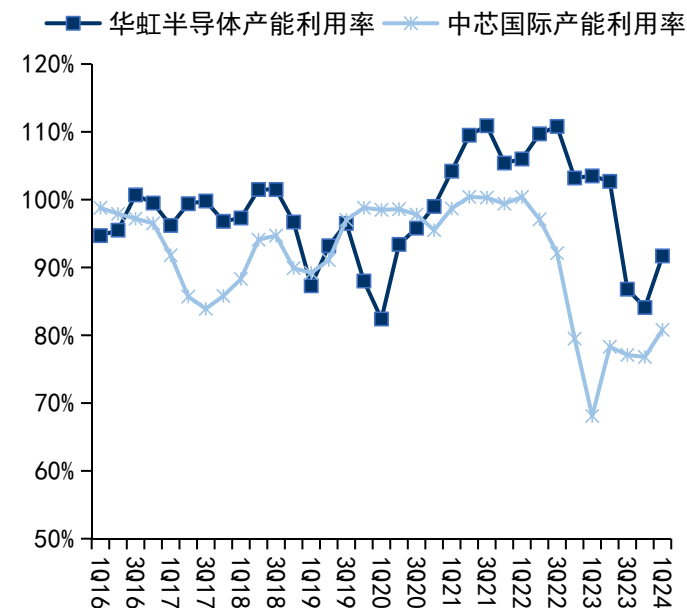
资料来源：SIA, SEMI, 国信证券经济研究所整理

图：半导体、半导体设备、半导体硅片季度同比增速



资料来源：SIA, SEMI, 国信证券经济研究所整理

图：中芯国际、华虹半导体季度产能利用率



资料来源：SK海力士, 国信证券经济研究所整理

# 费城半导体指数相对收益与半导体月销售额同比增速复盘

图：2002-2011年半导体月销售额同比增速和费城半导体指数走势复盘



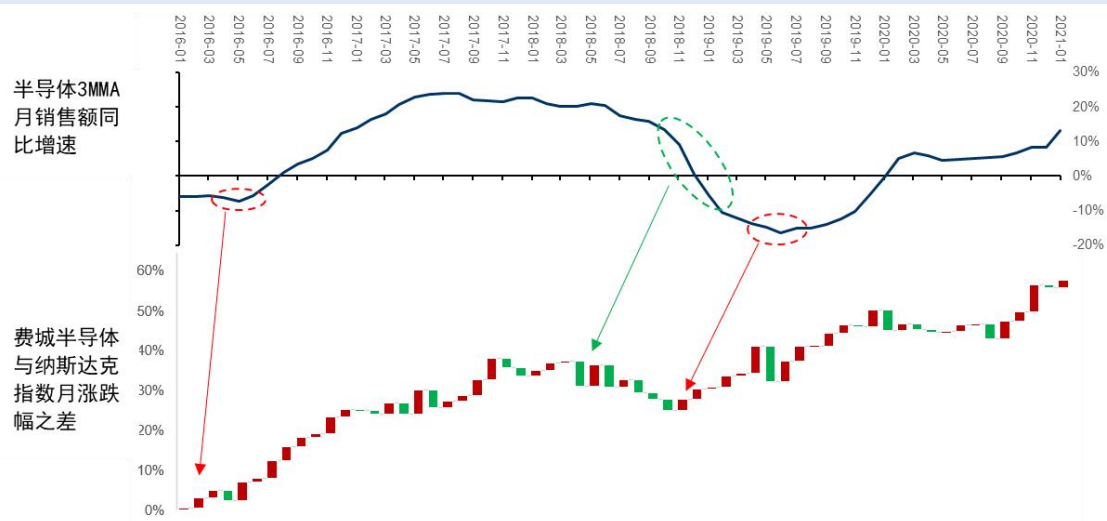
资料来源：SIA, Wind, 国信证券经济研究所整理

图：2012-2015年半导体月销售额同比增速和费城半导体指数走势复盘



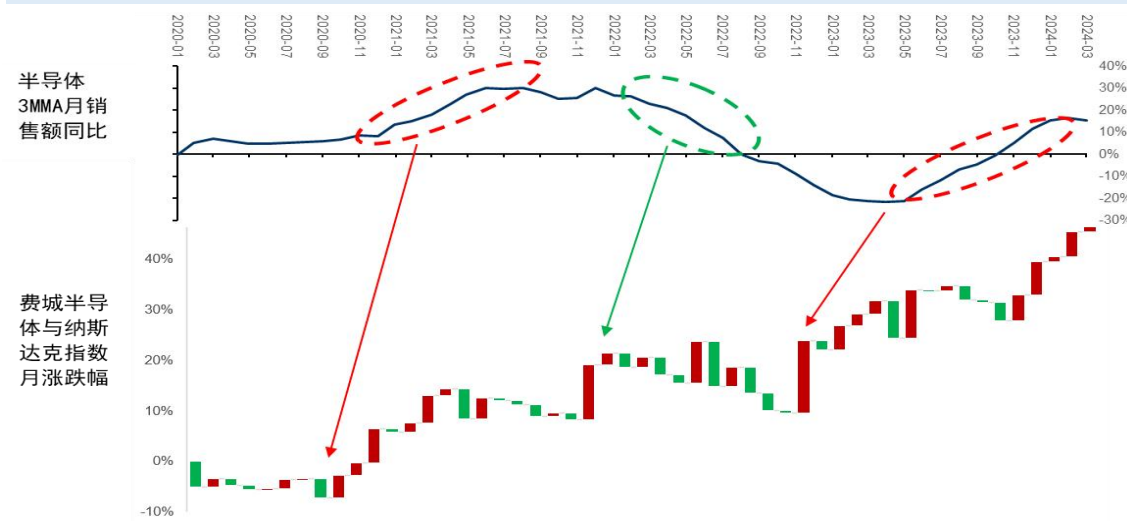
资料来源：SIA, Wind, 国信证券经济研究所整理

图：2016-2019年半导体月销售额同比增速和费城半导体指数走势复盘



资料来源：SIA, Wind, 国信证券经济研究所整理

图：2020-2023年半导体月销售额同比增速和费城半导体指数走势复盘



资料来源：SIA, Wind, 国信证券经济研究所整理

## 投资策略

# 投资建议：AI创新与周期向上共振，半导体开启新一轮成长



● AI创新叠加周期向上，半导体开启新一轮成长。根据SIA的数据，全球半导体季度销售额同比增速在1Q23触底，之后跌幅收窄，4Q23同比转正。1Q24全球实现销售额为1377亿美元，同比增长15.2%，环比减少5.7%；中国实现销售额为424亿美元，同比增长27.4%，环比减少6.6%。WSTS等多个机构均预计2024年全年将恢复10%以上增长，我们认为本轮半导体周期进入上行阶段。同时，AI创新正在从算力基础设施建设，扩展至AI手机、AI PC、AIoT等AI终端，更有望加速人形机器人、自动驾驶汽车等落地，AI的“Tipping Point”已到，预计将为半导体带来新一轮的成长。建议关注：

- **存储**：半导体产品中周期性最强的品种，上行阶段弹性较大，同时受益于AI带来的存力需求，相关标的江波龙、德明利、佰维存储等。
- **先进封装**：算力和存力均依赖的技术方向，相关标的长电科技、通富微电、芯原股份等。
- **晶圆代工**：中芯国际、华虹半导体、赛微电子等。
- **处理器**：龙芯中科等。
- **端侧芯片**：晶晨股份、乐鑫科技、恒玄科技、兆易创新等。
- **工业/汽车芯片**：纳芯微、思瑞浦、圣邦股份、北京君正等。
- **家电芯片**：芯朋微、晶丰明源等。
- **消费电子芯片**：韦尔股份、卓胜微、唯捷创芯、艾为电子、南芯科技、力芯微、天德钰、圣邦股份、帝奥微、杰华特、兆易创新等。
- **分立器件**：士兰微、华润微、新洁能、扬杰科技、斯达半导、时代电气、闻泰科技、捷捷微电等。
- **设备材料**：北方华创、中微公司、拓荆科技、鼎龙股份、天岳先进、沪硅产业等。

**一、国产替代进程不及预期。**国内半导体企业相比海外半导体大厂起步较晚，在技术和人才等方面存在差距，在国产替代过程中产品研发和客户导入进程可能不及预期。

**二、下游需求不及预期。**在地缘政治和全球经济疲软背景下，全球电子产品等终端需求可能不及预期，从而导致对半导体产品需求量减少。

**三、行业竞争加剧的风险。**在政策和资本支持下，国内半导体企业数量较多，在部分细分市场可能出现竞争加剧的风险，从而影响企业盈利能力。

**四、国际关系发生不利变化的风险。**我国半导体产业链在部分环节需要依赖海外厂商，若未来国际关系发生不利变化，可能对半导体产业链运营产生重大影响。



## 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.GSI）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