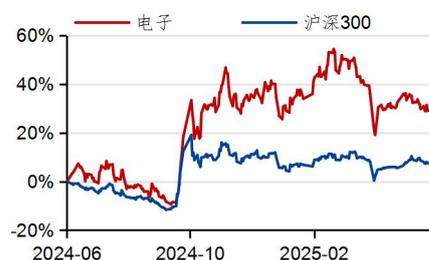


## GPU企业加速上市进程，国产算力有望迎来加速发展

### 投资要点

- ◆ **国内算力快速发展，增速显著提升。**根据国际数据公司(IDC)与浪潮信息日前联合发布《中国人工智能算力发展评估报告》显示，2024年中国智能算力规模达725.3百亿亿次/秒(EFLOPS)，同比增长74.1%，增幅是同期通用算力增幅(20.6%)的3倍以上；市场规模为190亿美元，同比增长86.9%。预计2025年，中国智能算力规模将达到1037.3EFLOPS，较2024年增长43%，中国人工智能算力市场规模将达到259亿美元，较2024年增长36.2%。
- ◆ **华为发布CloudMatrix 384 AI超节点，算力性能进一步提升。**2025年6月20日，华为在开发者大会发布CloudMatrix384AI超节点，将384颗昇腾NPU和192鲲鹏CPU通过全新高速网络MatrixLink对等互连形成了一个整体AI“服务器”。算力已从单台的6.4pFLOPS提升至超节点300pFLOPS，算力提升50倍。单卡的推理吞吐量从每秒600token提升至每秒2300token，提升近4倍。支持混合专家MoE大模型推理，实现1卡1专家，支持384个专家并行推理，极大提升效率。推理时延也从之前100ms降低至50ms以下。
- ◆ **龙芯中科发布新一代CPU及GPGPU。**龙芯中科一直致力于研究除ARM架构和X86架构之后的第三套自研信息技术体系及产业生态。2025年龙芯产品发布会上，龙芯中科发布最新一代3C6000系列处理器。本次产品主要打造了5种服务器主板方案(3C6000/S单路、双路;3C6000/D双路;3C6000/Q双路;3C6000/D四路)主要针对不同应用群体，服务器性能全面对标Intel第三代至强服务器系列，基本达到2023年市场主流产品水平，实测中64核心双路和32核心四路产品整机性能比较Intel的8380有小幅优势。GPU9A1000定位入门级的显卡，支持AI加速，AI算力预计在40TOPS，相较于公司上一代2K3000性能提升5倍以上，产品目前已经在研发尾声，即将进入流片。
- ◆ **国内多家GPU企业发布招股说明，国内GPU算力发展脚步加快。**摩尔线程，沐曦集成电路(上海)股份有限公司两家公司于2025年6月30日发布招股说明书。摩尔线程自主研发的全功能GPU，沐曦股份是国内高性能通用GPU产品的主要领军企业之一。
- ◆ **国内政策利好算力市场发展。**根据2023年10月工业和信息化部等六部门关于印发《算力基础设施高质量发展行动计划》文件，国家对算力发展保持高度重视，文件中指出到2025年，算力方面目标：算力规模超过300EFLOPS，智能算力占比达到35%，东西部算力平衡协调发展。
- ◆ **投资建议：**我国算力市场未来具备广阔发展空间，未来持续看好国内算力市场发展。建议关注进入各国内产品供应链或有相关技术储备厂商。如，寒武纪、海光信息、景嘉微、摩尔线程(未上市)、沐曦股份(未上市)。
- ◆ **风险提示：**国内算力不及预期风险；AI发展不及预期影响算力基建投资；国际市场竞争加剧风险；供应链变动风险，国际制裁风险。

投资评级

**领先大市(维持)**
**首选股票**
**评级**
**一年行业表现**


资料来源：聚源

升幅%	1M	3M	12M
相对收益	3.25	0.89	17.19
绝对收益	6.4	7.93	32.81

分析师

熊军

 SAC执业证书编号：S0910525050001  
 xiongjun@huajinsec.com

**相关报告**

电子：如何看待全球及国内算力的景气度？-行业周报\_电子\_如何看待全球及国内算力的景气度？——\_20250609-06015 2025.6.15



## 内容目录

1、国内智能算力快速增长，算力市场进入高质量发展时期 .....	4
1.1 华为 CloudMatrix 384 AI 超节点发布 .....	6
1.2 龙芯中科发布新一代 CPU 及 GPGPU .....	7
1.3 国内多家 GPU 企业发布招股说明，国内 GPU 算力发展脚步加快 .....	9
1.4 国内政策利好算力市场发展 .....	12
2、行情回顾 .....	12
2.1 周涨跌幅排名 .....	12
2.2 全球行情一览 .....	13
3、行业高频数据跟踪 .....	15
3.1 面板价格 .....	15
3.2 存储器价格 .....	17
4、风险提示 .....	19

## 图表目录

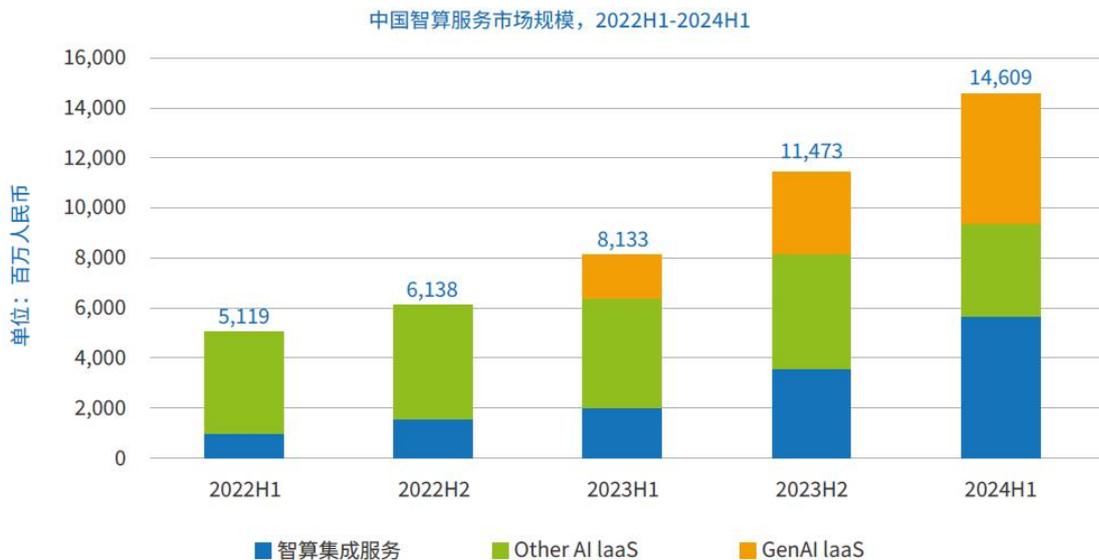
图 1: CloudMatrix 架构与传统架构对比 .....	6
图 2: CloudMatrix 架构创新 .....	6
图 3: DeepSeek-R1 的总体预填充吞吐量（每个加速器） .....	7
图 4: DeepSeek-R1 的总体解码吞吐量（每个加速器） .....	7
图 5: 龙芯 3C6000 服务器主板方案 .....	8
图 6: 龙芯 3C6000 服务器主板方案对比 Intel .....	8
图 7: 龙芯 2K3000/3B6000N .....	8
图 8: 龙芯首款 GPU 芯片 9A1000 .....	8
图 9: 摩尔线程主要产品分类 .....	11
图 10: MUSA 架构示意图 .....	11
图 11: 沐曦股份典型的 GPU 芯片基本组成架构 .....	12
图 12: MXMACA 软件栈框架结构 .....	12
图 13: 2025/07/07-2025/07/11 申万一级各行业涨跌幅（%） .....	13
图 14: 2025/07/07-2025/07/11 电子各行业涨跌幅（%） .....	13
图 15: 费城半导体指数（短时间维度） .....	14
图 16: 费城半导体指数（长时间维度） .....	14
图 17: 2025/04/28-2025/07/11 各类 DRAM 颗粒现货价格（美元） .....	17
图 18: 近两周 DDR5 16G（2Gx8）4800/5600 颗粒现货价格（\$） .....	18
图 19: 近两周 DDR4 16Gb（1Gx16）3200 颗粒现货价格（\$） .....	18
图 20: 近两周 DDR4 16Gb（2Gx8）3200 颗粒现货价格（\$） .....	18
图 21: 近两周 DDR4 8Gb（1Gx8）3200 颗粒现货价格（\$） .....	18
图 22: 近两周 DDR4 8Gb（512Mx16）3200 颗粒现货价格（\$） .....	18
图 23: 近两周 DDR3 4Gb（512Mx8）1600/1866 颗粒现货价格（\$） .....	18
表 1: 2025/07/07-2025/07/11 全球半导体龙头公司行情 .....	15

表 2: 2025 年 7 月 TV 面板价格风向标.....	16
表 3: 2025 年 7 月 IT 面板价格趋势.....	16

## 1、国内智能算力快速增长，算力市场进入高质量发展时期

国内算力增速显著，算力市场进入高质量发展时期。根据国际数据公司(IDC)与浪潮信息日前联合发布《中国人工智能算力发展评估报告》显示，2024年中国智能算力规模达725.3百亿亿次/秒(EFLOPS)，同比增长74.1%，增幅是同期通用算力增幅(20.6%)的3倍以上；市场规模为190亿美元，同比增长86.9%。预计2025年，中国智能算力规模将达到1037.3EFLOPS，较2024年增长43%，中国人工智能算力市场规模将达到259亿美元，较2024年增长36.2%。

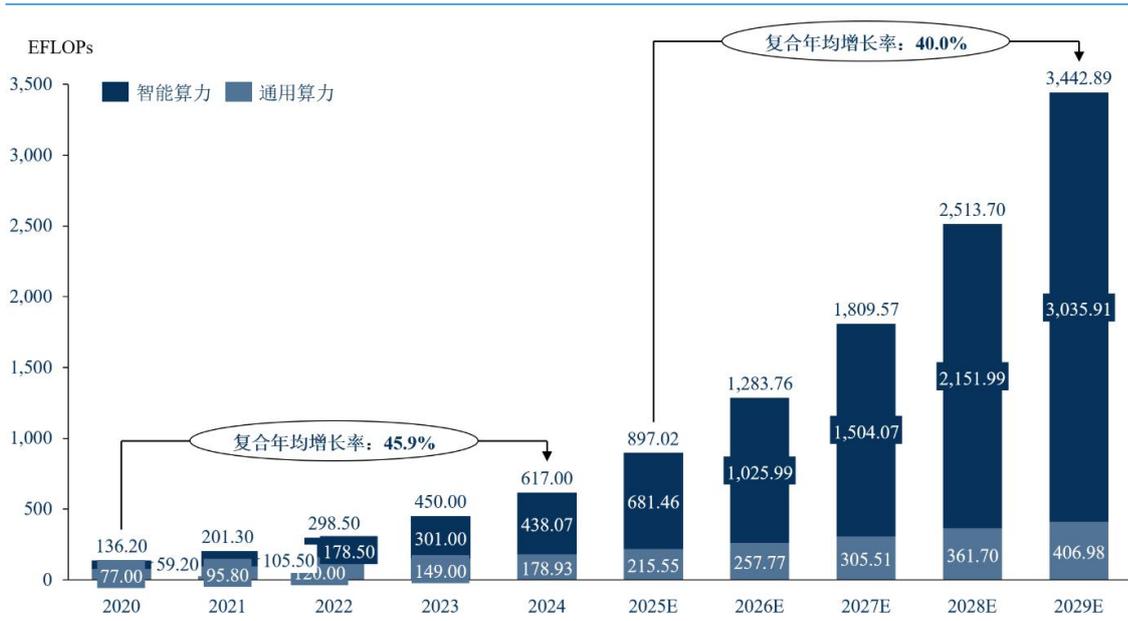
图 1：中国智能算力服务市场规模 2022H1-2024H1



资料来源：IDC2025、华金证券研究所

根据弗若斯特沙利文预测，随着AI和大数据技术的广泛应用，中国算力规模呈现快速增长态势，整体规模从2020年的136.20EFLOPs增长至2024年的617.00EFLOPs，期间年均复合增长率为45.9%；预计到2029年中国算力总规模将达到3,442.89EFLOPs，预测期年均复合增长率达40.0%。其中，智能算力是引领算力规模指数级增长的核心，其规模从2020年59.20EFLOPs增长至2024年的438.07EFLOPs，期间年均复合增长率高达64.9%，预计在2025年至2029年期间，智能算力将以45.3%的年均复合增长率增长至3,035.91EFLOPs。

图 1：2020-2029 年中国算力总规模

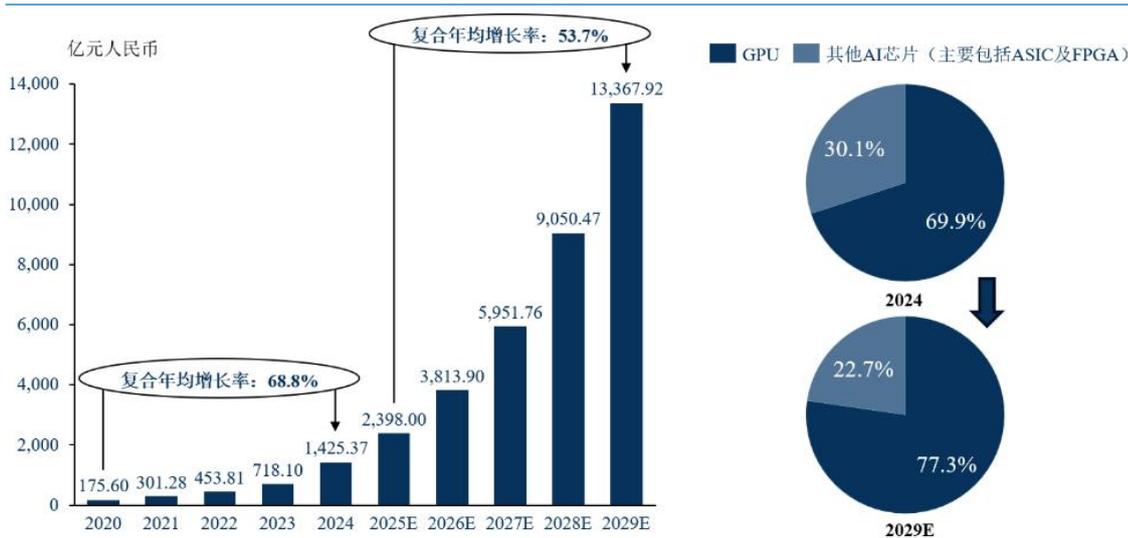


资料来源：弗若斯特沙利文、华金证券研究所

目前, GPU 依然是 AI 市场的主导芯片。不过以 ASIC 和 FPGA 为代表的其他类型芯片也已实现商业化, 并在市场中占据一定比例, 我国企业华为生产的昇腾 NPU 在市场上有一定客户群体。

根据弗若斯特沙利文预测, 到 2029 年, 中国的 AI 芯片市场规模将从 2024 年的 1,425.37 亿元激增至 13,367.92 亿元, 2025 年至 2029 年期间年均复合增长率为 53.7%。从细分市场上看, GPU 的市场增长速度最快, 其市场份额预计将从 2024 年的 69.9% 上升至 2029 年的 77.3%。

图 1：2020-2029 年中国 AI 计算加速芯片市场规模



资料来源：弗若斯特沙利文、华金证券研究所

**国内单卡性能显著提升, 国产芯片加速推进。**景嘉微、海光信息、龙芯中科等国内芯片上市公司发展迅速, 其快速发展对国内芯片行业迅速崛起发挥重要作用。与此同时, 新一批国产芯片研发企业不断涌现, 天数智芯、燧原科技、壁仞科技等非上市公司在产品研发方面也取得了突破性进展, 为我国芯片产业的多元化发展注入了强劲动力。现阶段, 国内厂商在设计、制程、封装

和测试等各个环节均表现突出，国产芯片性能显著提升。据 BernsteinResearch 数据显示，华为自研的昇腾 910B、寒武纪的思元 590、百度的昆仑芯 II-R200、腾讯的紫霄 C100 在一定程度上表现出了性能优势。

## 1.1 华为 CloudMatrix 384 AI 超节点发布

2025 年 6 月 20 日，华为在开发者大会发布 CloudMatrix384AI 超节点，将 384 颗昇腾 NPU 和 192 鲲鹏 CPU 通过全新高速网络 MatrixLink 对等互连形成了一个整体 AI “服务器”。算力已从单台的 6.4pFLOPS 提升至超节点 300pFLOPS，算力提升 50 倍。单卡的推理吞吐量从每秒 600tokens 提升至每秒 2300tokens，提升近 4 倍。

超节点架构相较于单节点能够更好支持混合专家 MoE 大模型推理，实现 1 卡 1 专家，支持 384 个专家并行推理，极大提升效率。推理时延也从之前 100ms 降低至 50ms 以下。在应用上新浪微博已经使用搭载 CloudMatrix384AI 超节点服务器进行日常 AI 推理模型应用，中科院基于 CloudMatrix384 超节点，打造了自己的模型训练框架，帮助中科院快速构建 AlforScience 科研大模型。

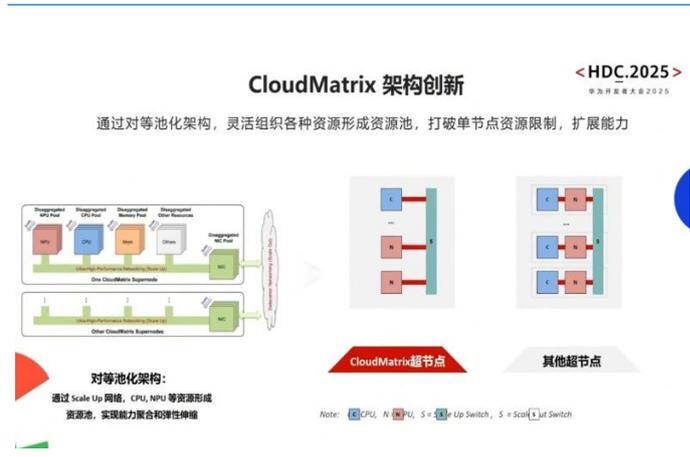
CloudMatrix384 超节点与其他万卡链接的区别在于它使用对等池化架构，通过 ScaleUp 网络（纵向扩展），打破单节点边界资源限制，扩展能力。384NPU+192CPU 组成“一台计算机”，将各种资源的通信打开，节点内和节点间可以以相同的带宽通信，直接将 NPU 和 CPU 的通信打开无需中转。

图 1: CloudMatrix 架构与传统架构对比



资料来源：华为开发者大会 2025、华金证券研究所

图 2: CloudMatrix 架构创新



资料来源：华为开发者大会 2025、华金证券研究所

根据华为联合硅基流动发布论文《Serving Large Language Model on Huawei CloudMatrix384》中可以看出，在 DeepSeek-R1 模型的测试中，使用 CloudMatrix384 的昇腾 910CNPU 与英伟达 H800GPU 在计算效率上各有优劣。

实际测试中在默认配置下，总体预填充吞吐量的测试 CloudMatrix-Infer 每个 NPU 每秒处理 5655 tokens。鉴于每个昇腾 910CNPU 的计算能力为 1504TFLOPS (INT8)，对应的计算效率为 3.76 tokens/s per TFLOPS。这比 SGLang 在英伟达 H100 上的默认配置 (3.18 tokens/s per TFLOPS) 效率要高得多，尽管后者的原始吞吐量略高。

解码吞吐量的测试中，CloudMatrix-Infer 配置了每个 NPU 的批处理大小为 96，KV 缓存长度为 4,096 个 tokens，可实现 49.4 毫秒的出色 TPOT。在绝对系统吞吐量方面，CloudMatrix-Infer 每个 NPU 产生 1,943 个 tokens/秒，其批处理大小为 96。这高于 DeepSeek ((Blog)) H800 基线 (每个 GPU1,850 个 tokens/秒)。虽然在数值上低于 H800 上的 DeepSeek (Profile) (每个 GPU2,325 个 tokens/秒) 和 H100 上的 SGLang (每个 GPU2,172 个 tokens/秒)，但后者的基准测试具有更大的批处理大小，为 128。

截止 6 月 20 日开发者大会，使用昇腾 AI 云服务客户从 2024 年 321 个跃升至 2025 年上半年 1346 个，客户数量增长明显。

图 3: DeepSeek-R1 的总体预填充吞吐量 (每个加速器)

Method	Batch Size	Input Length	Hardware TFLOPS (Precision)	Throughput (tokens/s)	Throughput per TFLOPS
DeepSeek on H800 (Blog)	N/A	N/A	1979 (FP8)	4,026	2.03
SGLang on H100 (Default)	16,384	4,096	1979 (FP8)	6,288	3.18
CloudMatrix-Infer (Default)	16,384	4,096	1504 (INT8)	5,655	3.76
DeepSeek on H800 (Profile)	16,384	4,096	1979 (FP8)	<b>7,839</b>	3.96
SGLang on H100 (Perfect EPLB)	16,384	4,096	1979 (FP8)	7,417	3.75
CloudMatrix-Infer (Perfect EPLB)	16,384	4,096	1504 (INT8)	6,688	<b>4.45</b>

资料来源: 《Serving Large Language Models on HuaweiCloudMatrix384》、  
 华金证券研究所

图 4: DeepSeek-R1 的总体解码吞吐量 (每个加速器)

Method	Batch Size	KV Cache Length	Hardware TFLOPS	TPOT (ms)	Throughput (tokens/s)	Throughput per TFLOPS
DeepSeek (Blog) on H800	N/A	4,989	1979 (FP8)	~50.0	1,850	0.93
DeepSeek (Profile) on H800	128	4,096	1979 (FP8)	~50.2	<b>2,325</b>	1.17
SGLang (Simu. MTP) on H100	128	4,000	1979 (FP8)	~55.6	2,172	1.10
CloudMatrix-Infer	96	4,096	1504 (INT8)	49.4	1,943	<b>1.29</b>

资料来源: 《Serving Large Language Models on HuaweiCloudMatrix384》、  
 华金证券研究所

## 1.2 龙芯中科发布新一代 CPU 及 GPGPU

龙芯中科一直致力于研究除 ARM 架构和 X86 架构之后的第三套自研信息技术体系及产业生态。2025 年龙芯产品发布会上，龙芯中科发布最新一代 3C6000 系列处理器。

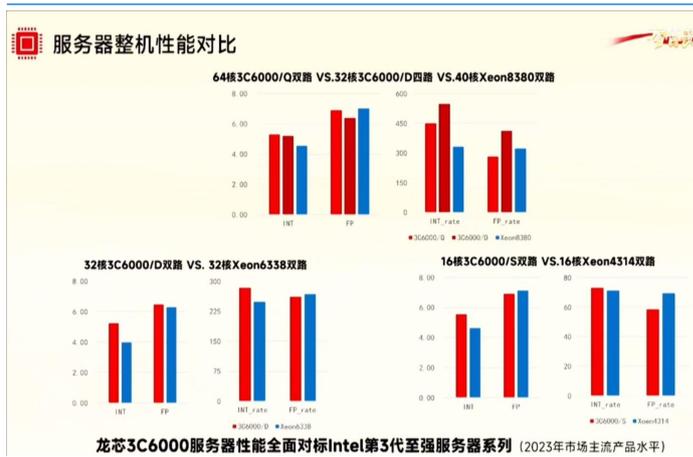
龙芯 3C6000 系列处理器具备高性能，高可靠，高安全，高能效的特点。本次产品主要打造了 5 种服务器主板方案 (3C6000/S 单路、双路;3C6000/D 双路;3C6000/Q 双路;3C6000/D 四路) 主要针对不同应用群体，服务器性能全面对标 Intel 第三代至强服务器系列，基本达到 2023 年市场主流产品水平，实测中 64 核心双路和 32 核心四路产品整机性能比较 Intel 的 8380 有小幅优势。

图 5：龙芯 3C6000 服务器主板方案



资料来源：龙芯中科产品发布会、华金证券研究所

图 6：龙芯 3C6000 服务器主板方案对比 Intel

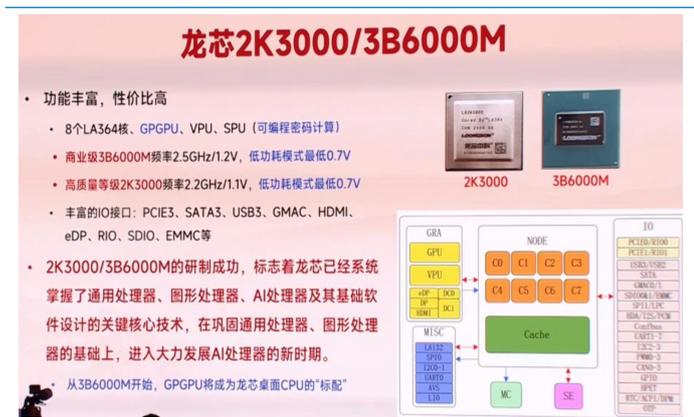


资料来源：龙芯中科产品发布会、华金证券研究所

在工控领域本次发布会发布了龙芯 2K3000：采用 8 核 SoC 芯片，龙芯自主指令系统（龙架构），基于自主研发的 LA364E 处理器核，SPEC CPU2006Base 单线程/浮点峰值性能均超过 10 分/GHz，与龙芯 3A5000 处理器使用的 LA464 核性能相当；集成第二代自研 GPGPU 核心 LG200，图形性能成倍提升，支持通用计算和 AI 加速；集成丰富的 I/O 接口，可广泛应用于各种场景。

公司针对未来 GPU 芯片发展方向，发布首款 GPU 芯片 9A1000：其定位是入门级的显卡并且支持 AI 加速，AI 算力预计在 40TOPS，相较于公司上一代 2K3000 性能提升 5 倍以上，产品目前已经在研发尾声，即将进入流片。下一代 9A2000 单精度浮点算力可达到 5TFlops，INT8AI 算力达到 160TOPS，内存带宽 256BG/s，并且支持双片互联拓展。

图 7：龙芯 2K3000/3B6000M



资料来源：龙芯中科产品发布会、华金证券研究所

图 8：龙芯首款 GPU 芯片 9A1000



资料来源：龙芯中科产品发布会、华金证券研究所

## 1.3 国内多家 GPU 企业发布招股说明，国内 GPU 算力发展脚步加快

我国目前多家 GPU 生产企业将于科创板上市，国内 GPU 芯片算力厂商进步明显。摩尔线程，沐曦集成电路（上海）股份有限公司两家公司于 2025 年 6 月 30 日发布招股说明书。摩尔线程自主研发的全功能 GPU，沐曦股份是国内高性能通用 GPU 产品的主要领军企业之一。

表 1：国际龙头企业及国内重点企业 GPU 市场对比图

公司名称	产品结构	盈利模式	研发实力	产业化程度	应用场景落地	技术成熟度	生态竞争能力
英伟达	基于 GPU 的云端和边缘端人工智能芯片	销售芯片、加速卡以及整机系统	2024 年研发投入 129.14 亿美元，研发费用率 9.90%	产业化程度极高，市场占有率全球领先	主要面向云端和边缘端人工智能、游戏、专业图形、汽车等应用场景	技术成熟度全球领先，且不断迭代中	生态竞争能力全球领先，CUDA 生态引领国际市场
AMD	数据中心、客户端、游戏及嵌入式四大业务板块，包括 CPU、GPU、FPGA、DPU、SoC 等芯片产品	销售芯片、加速卡以及整机系统	2024 年研发投入 64.56 亿美元，研发费用率 23.23%	产业化程度高，在全球市场占有率较高	数据中心、客户端、游戏及嵌入式等多领域	成熟且不断迭代中	生态竞争能力较强，ROCm 生态在专业计算等领域表现较为优异
N 公司	云端、边缘端和终端人工智能芯片	销售芯片、加速卡及整机	2024 年研发投入 1,797 亿元(集团整体口径)，研发费用率 20.84%	产业化程度较高，并已在国内市场实现规模化营收	云边端的视觉、语音、自然语言、搜索推荐等	较为成熟且不断迭代中	依托于 N 公司整体生态优势
海光信息	基于 x86 指令框架的海光 CP 基于 GPGPU 架构 DCU 芯片	销售 CPU 和 DCU 产品	2024 年研发投入 34.46 亿元，研发投入率 37.61%；截至 2024 年底研发人员数量 2,157 人，占比 90.18%	产业化程度较高，并已在国内市场实现规模化营收	电信、金融、互联网等领域	较为成熟且不断迭代中	积极参与开源社区，与众多厂商合作，构建软硬件生态
寒武纪	云端、边缘端、终端通用智能芯片和 IP	销售 IP 授权、芯片、加速卡以及智能计算集群系统	2024 年研发投入 10.72 亿元，研发费用率 91.30%；截至 2024 年底研发人员数量 741 人，占比 75.61%	产业化程度较高，并已在国内市场实现规模化营收	云边端的视觉、语音、自然语言、搜索推荐等	较为成熟且不断迭代中	具备云边端一体化开发环境，支持统一的软件生态
景嘉微	GPU 芯片、图形显控模块和加固类产品、小型专用化雷达等	销售 GPU 芯片、图形显控模块和加固类产品、小型专用化雷达等	2024 年研发投入 2.81 亿元，研发费用率 60.18%；截至 2024 年底研发人员数量 586 人，占比 61.55%	产业化程度较高，并已在国内市场实现规模化营收	应用于图形渲染等领域，产品在图形处理和可视化等场景中广泛应用	较为成熟且不断迭代中	产品以图形渲染为主，应用场景明确

公司名称	产品结构	盈利模式	研发实力	产业化程度	应用场景落地	技术成熟度	生态竞争能力
摩尔线程	基于全功能 GPU 的板卡、一体机及集群设备，覆盖云端、边缘端和终端场景	销售 GPU 板卡、一体机和集群设备	2024 年研发投入 13.59 亿元，研发费用率 309.88%；截至 2024 年底研发人员数量 886 人，占比 78.69%	量产，并已在国内市场实现商业化	AI 智算集群、专业图形加速、桌面级图形加速，以及端侧智能设备	较为成熟且不断迭代中	依托自研 MUSA 架构搭建国产计算加速生态

资料来源：摩尔线程招股说明书、华金证券研究所

摩尔线程智能科技（北京）股份有限公司自 2020 年成立以来，以自主研发的全功能 GPU 为核心，致力于为 AI、数字孪生、科学计算等高性能计算领域提供计算加速平台。公司已成功推出四代 GPU 架构，并拓展出覆盖 AI 智算、云计算和个人智算等应用领域的计算加速产品矩阵。

摩尔线程核心产品主要包括四类：（1）AI 智算产品：主要负责 AI 训练智算卡、AI 推理卡、AI 超节点服务器及夸娥（KUA E）智算集群等，为 AI 计算中心、云服务平台等打造，满足从大模型预训练及后训练、推理部署到 GPU 云服务等场景应用需求。（2）专业图形加速产品：主要应用于工业设计、高清视频编辑、数字孪生、AI 云电脑等高端场景的 GPU 及相关产品系列，涵盖 MTT S3000/S2000/S1000 等系列显卡，以及基于前述显卡打造的一体机等解决方案。（3）桌面级图形加速产品：主要应用于 AIPC、游戏 PC 及办公 PC 等场景的 GPU 及相关产品，包括 MTT S80/S70/S50/S30/S10/X300/X100 等系列显卡，以及基于前述显卡打造的工作站等。该产品支持 Windows、麒麟 KylinOS、统信 UOS、openEuler 等多款国内外主流操作系统，以及 Intel、AMD、海光、飞腾、鲲鹏等多款国内外主流 CPU 平台。（4）智能 SoC 类产品：主要应用于 AIPC、边缘智能、具身智能、智能汽车和低空经济等众多场景，包括基于 SoC 芯片的 AI 算力本-A140、AI 模组-E300 等产品。该产品可以广泛服务于 C 端和 B 端客户，满足上述行业对于端侧和边缘类 AI 场景的需求，同时可与公司的 AI 智算产品结合，形成云-边-端一体化解决方案，赋能客户实现 AI 的训练-推理需求。

MUSA 架构是公司自主研发的融合 GPU 硬件和软件的全功能 GPU 计算加速统一系统架构。公司 MUSA 架构本身并不作为单独产品对外销售，而是作为公司全功能 GPU 产品的核心技术支撑。该架构涵盖统一的芯片架构、指令集、编程模型、软件运行库及驱动程序框架等关键要素，旨在为各类并行计算场景提供高性能计算能力。开发人员可借助 C/C++、Triton 等编程语言，在该架构下编写并行计算程序，且同一代码能够在公司不同 GPU 产品及系统上运行，具有良好的灵活性与可扩展性。此外，MUSA 架构具备与由英伟达主导的国际主流 GPU 生态的兼容性，使得开发者能够以较低成本充分利用目前国际主流生态下的代码资源。基于 MUSA 架构开发的应用程序不仅具有广泛的可移植性，还能够同时在云端及边缘的众多计算平台上运行，其应用领域广泛，涵盖 AI、图形处理、科学计算等多个重要方向。

公司第四代 GPU 芯片“平湖”于 2024 年发布成功，该芯片增加了 FP8 精度支持，大幅提升 AI 算力，公司基于该芯片支撑面向 DeepSeek 类前沿大模型预训练的万卡集群智算中心解决方案。规格上，“平湖”最大频率为 2.0GHz，拥有 8,192 着色核心，片间互连带宽 800GB/s，最大显存容量 80GB。公司 MTT S80 显卡的 FP32 算力性能接近英伟达 RTX3060；公司 MTT S5000 千卡智算集群效率超过同等规模国外同代系 GPU 训练集群计算效率。

图 9: 摩尔线程主要产品分类

分类		芯片	板卡/模组	一体机	集群设备
服务器级	AI 智算	企业级	第四代 GPU “平湖” S5000	D800 X1/X	KUAE2
			第三代 GPU “曲院” S4000		KUAE1
桌面级图形加速	专业图形加速	企业级	第二代 GPU “春晓” S3000	D200/D400/ D800	MCCX
			第一代 GPU “苏堤” S1000/S2000		
		消费级	第二代 GPU “春晓” S70/S80		
		企业级	第二代 GPU “春晓” X300		
			第一代 GPU “苏堤” S10/S30/S50 /X100		
智能 SoC 类	企业级	第一代 SoC “长江”	AI 模组 -E300	AI 算力本 -A140	
示意图					

资料来源: 摩尔线程招股说明书、华金证券研究所

图 10: MUSA 架构示意图



资料来源: 摩尔线程招股说明书、华金证券研究所

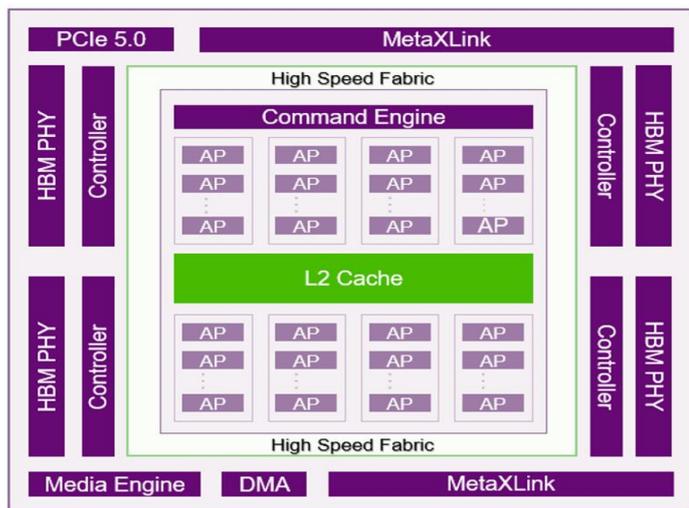
沐曦集成电路(上海)股份有限公司是国内高性能通用 GPU 产品的主要领军企业之一, 致力于自主研发全栈高性能 GPU 芯片及计算平台, 持续为国家人工智能公共算力平台、运营商智算平台、商业化智算中心等提供基础算力底座, 推动人工智能赋能千行百业并重点布局教科研、金融、交通、能源、医疗健康、大文娱等行业应用场景。

沐曦股份核心产品主要包括三类: (1) 训练推理一体 GPU (曦云 C 系列): 产品针对云端计算场景, 具有强大的多精度混合算力、高带宽和大容量存储, 结合其自研的 MetaXLink 高速互连技术, 能够满足大规模计算集群扩展需求、支持千亿参数以上的 AI 大模型训练, 可以大幅增加集群算力、缩短大模型计算时间。(2) 智算推理 GPU (曦思 N 系列): 曦思 N100 产品是公司 2022 年推出的首款产品, 是面向云端和边端多种传统 AI 应用场景的智算推理 GPU, 提供强大的推理算力和视频编解码能力, 可广泛应用于智慧城市、智慧交通、智慧教育、智能视频处理等。后续曦思 N260、曦思 N300(在研)产品针对生成式人工智能下的云端人工智能推理场景, 拥有强大的多精度混合算力, 配以大容量显存和新一代高速 IO 接口, 支持主流深度学习开发框架, 可为内容生成式应用和大语言模型等智能应用提供端到端的加速服务。(3) 图形渲染 GPU (曦彩 G 系列): 产品采用公司自研的 XCORE2.0 架构及指令集, 内置性能强大的 GPUIP 和硬件单元, 具备卓越的图形处理能力, 可广泛应用于云渲染、游戏、数字孪生、影视动画制作、专业制图等场景且支持光线追踪技术。截至 2025 年 6 月 30 日曦彩 G 系列相关产品已完成 GPUIP 设计和验证。

公司自研的 MXMACA 软件栈系提高 GPU 易用性的核心, 该软件栈是基于公司自研指令集以及 GPU 并行计算引擎, 集成了主流算法框架、运算库、通信库、操作系统、编程语言、调试和运维管理工具等, 为 GPU 产品提供了一套统一、完整且高效的全栈式软件工具链, 涵盖应用开发、功能调试和性能调优等核心环节。按照抽象层次由高到低, 可分为四个层级①算法应用层②数学库及通信库层③编程语言层④软件驱动层。

MXMACA 软件栈具有较强的易用性、可编程性和可扩展性, 功能丰富全面。通过提供简易编程语言来大幅降低用户应用门槛。

图 11: 沐曦股份典型的 GPU 芯片基本组成架构



资料来源: 沐曦股份招股说明书、华金证券研究所

图 12: MXMACA 软件栈框架结构



资料来源: 沐曦股份招股说明书、华金证券研究所

## 1.4 国内政策利好算力市场发展

根据 2023 年 10 月工业和信息化部等六部门关于印发《算力基础设施高质量发展行动计划》通知, 国家对算力发展保持高度重视, 算力是集信息计算力、网络运载力、数据存储力于一体的新型生产力, 需要保持高质量发展, 提出了“完善算力综合供给体系”、“提升算力高效运载能力”、“强化存力高效灵活保障”、“深化算力赋能行业应用”、“促进绿色低碳算力发展”、“加强安全保障能力建设”等重点任务, 从顶层明确了未来三年算力基础设施的发展路径, 为我国算力产业链建设提供宏观指引。文件中指出到 2025 年, 计算力方面目标: 算力规模超过 300EFLOPS, 智能算力占比达到 35%, 东西部算力平衡协调发展。

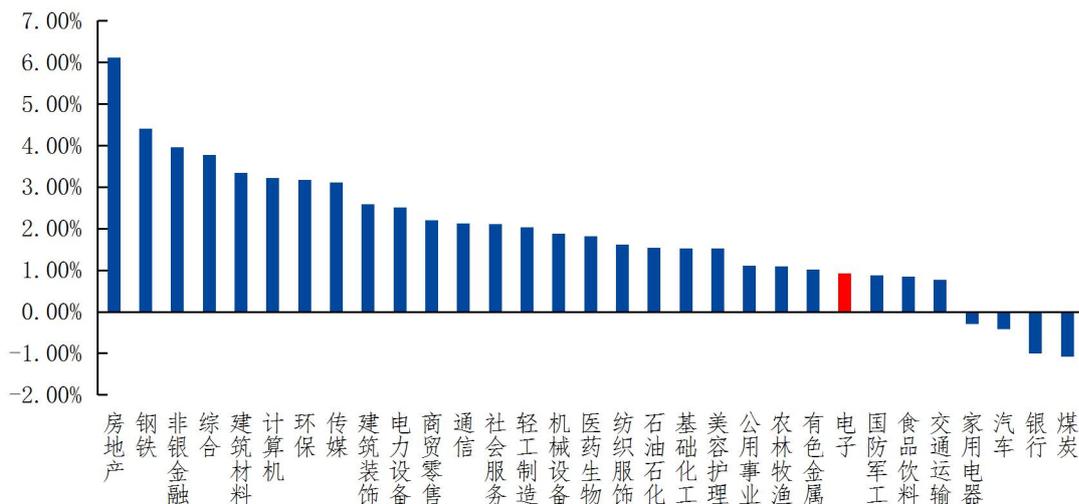
我们认为, 我国算力市场未来具备广阔发展空间, 未来持续看好国内算力市场发展。

## 2、行情回顾

### 2.1 周涨跌幅排名

7 月 7 日至 7 月 11 日当周电子行业涨幅为 0.93%。7 月 7 日至 7 月 11 日, 申万一级行业涨多跌少。其中房地产板块涨幅最大, 上涨 6.12%, 其次是钢铁板块及非银金融板块, 分别上涨 4.41%/3.96%。煤炭板块跌幅最大, 下跌 1.08%, 其次是银行板块及汽车板块, 分别下跌 1.00%/0.41%。

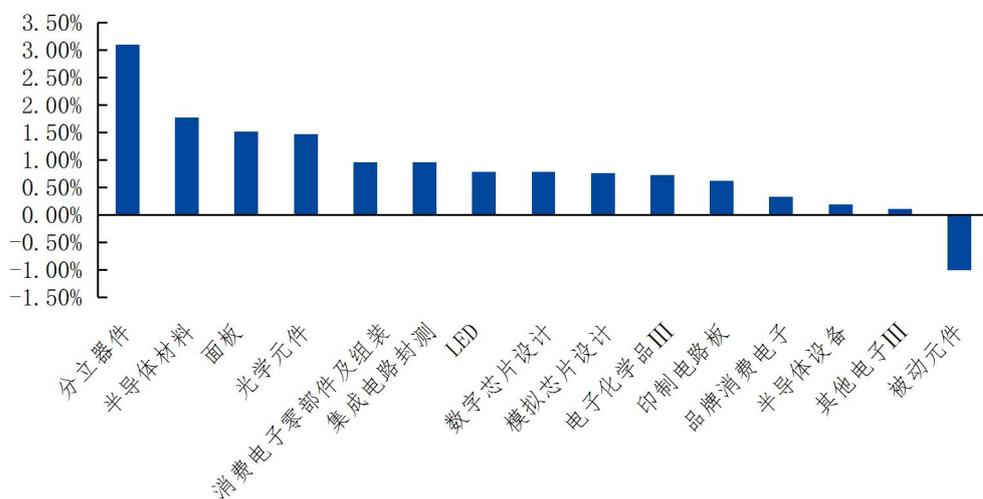
图 13: 2025/07/07-2025/07/11 申万一级各行业涨跌幅 (%)



资料来源: Wind、华金证券研究所

电子板块细分比较, 7月7日至7月11日, 申万三级电子行业细分板块涨块多跌少。其中, 分立器件板块涨幅最大, 上涨 3.10%, 其次是半导体材料板块及面板板块, 分别上涨 1.77%/1.52%; 跌幅最大的是被动元件板块, 跌幅为 1.01%, 其他电子III板块及半导体设备板块涨幅最小, 涨幅分别为 0.11%/0.19%。

图 14: 2025/07/07-2025/07/11 电子各行业涨跌幅 (%)



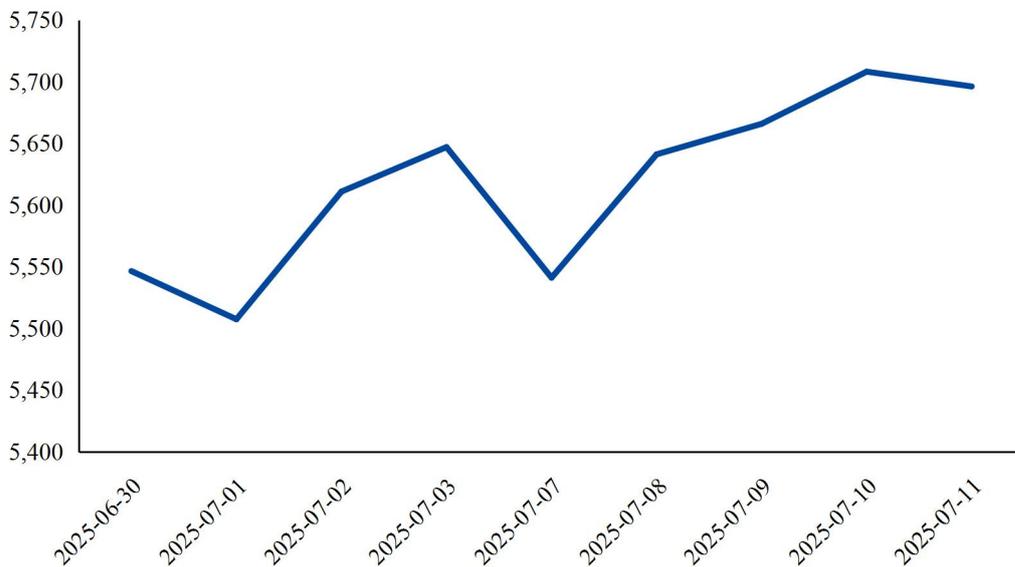
资料来源: Wind、华金证券研究所

注: 申万三级暂无集成电路制造指数

## 2.2 全球行情一览

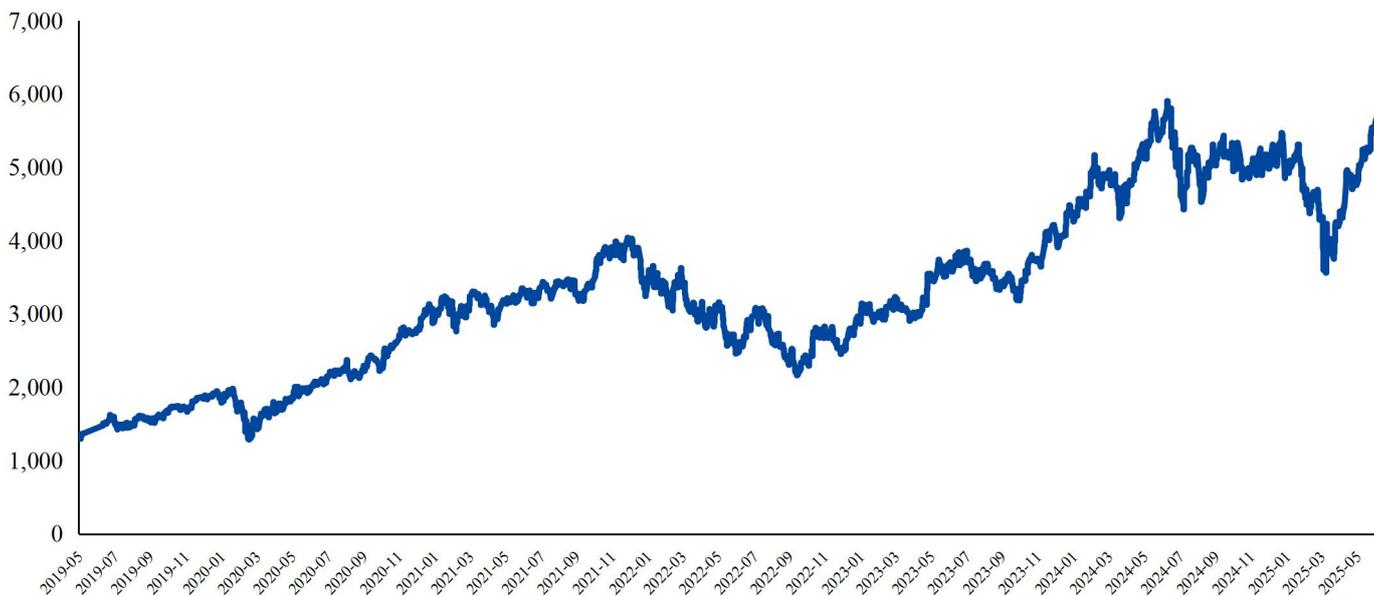
7月7日至7月11日当周，费城半导体指数整体呈现上涨走势，由7月7日的5,541.20点上涨至7月11日的5,696.29点。从更长的时间维度上看，指数仍处于2025年4月以来的反弹通道中。

图 15: 费城半导体指数（短时间维度）



资料来源: Wind、华金证券研究所

图 16: 费城半导体指数（长时间维度）



资料来源: Wind、华金证券研究所

7月7日至7月11日，全球半导体龙头股涨多跌少。SK海力士涨幅最大，为8.87%；思佳讯跌幅最大，跌幅为4.24%。

表 1: 2025/07/07-2025/07/11 全球半导体龙头公司行情

分类	股票代码	公司简称	市值 (亿元)	PE(LYR)	PB(MRQ)	周涨跌幅
处理器	INTC.O	英特尔	7,305	60.51	1.02	4.18%
	QCOM.O	高通	12,357	17.05	6.24	-2.93%
	NVDA.O	英伟达	287,619	55.21	48.00	3.50%
	AMD.O	超威半导体	16,969	277.99	4.10	6.17%
存储	MU.O	美光	9,961	179.13	2.75	1.93%
	000660.KS	SK 海力士	11,187	10.83	2.63	8.87%
模拟	TXN.O	德州仪器	14,366	41.88	12.25	2.42%
	NXPI.O	恩智浦	4,134	23.04	6.20	-1.37%
射频	SWKS.O	思佳讯	809	19.00	1.91	-4.24%
	QRVO.O	QORVO	587	147.55	2.42	-0.62%
功率半导体	STM.N	意法	2,063	18.54	1.63	-0.06%
	0KED.L	英飞凌	4,130	37.87	2.88	2.53%
光学	3008.TW	大立光	748	11.77	1.68	1.33%
半导体设备	AMAT.O	AMAT	11,353	22.13	8.38	3.60%
	ASML.O	阿斯麦	22,574	39.84	16.65	0.94%
硅片	6488.TWO	环球晶圆	373	15.44	1.66	3.75%
晶圆代工	2330.TW	台积电	85,411	24.63	6.25	1.38%
封测	ASX.N	日月光	1,643	22.82	2.29	0.39%
	AMKR.O	安靠	386	15.26	1.30	-1.66%
	2449.TW	京元电子	302	15.95	2.59	2.53%
化合物半导体	3105.TWO	稳懋	87	46.36	0.98	0.00%
分销	ARW.N	艾睿电子	454	16.15	1.07	0.08%
	AVT.O	安富利	304	8.49	0.87	-0.98%

资料来源: Wind、华金证券研究所

注:数据截至当周最后一个交易日

### 3、行业高频数据跟踪

#### 3.1 面板价格

**TV 面板:** 品牌采购需求环比增加叠加代工厂商旺季备货需求, 预计电视面板需求呈现环比温和恢复。根据群智咨询数据: (1) 32", 需求承压且库存偏高, 6 月价格小幅下降 1 美金, 预计 7 月均价下降 2 美金。(2) 50", 供应分散, 且需求恢复力度有限, 6 月均价将下降 2 美金, 预计 7 月均价下降 4 美金。(3) 55", 供应趋于集中, 而需求分化, 6 月均价将下降 2 美金, 预计 7 月均价下降 4 美金。(4) 大尺寸方面, 终端市场需求降温影响品牌库存压力依然较大, 6 月均价下降 2 美金, 预计 7 月均价下降 4 美金。

表 2: 2025 年 7 月 TV 面板价格风向标

Application	Size	Resolution	OC/LCM	Range	Jun'25	Jul'25(E)	Jul. VS Jun. Change
TV	32"	1366x768	OC	High	36.0	34.0	(2.0)
				Typical	35.0	33.0	(2.0)
				Low	34.0	32.0	(2.0)
	50"	3840x2160	OC	High	99.0	95.0	(4.0)
				Typical	98.0	94.0	(4.0)
				Low	91.0	87.0	(4.0)
	55"	3840x2160	OC	High	126.0	122.0	(4.0)
				Typical	125.0	121.0	(4.0)
				Low	118.0	114.0	(4.0)
	65"	3840x2160	OC	High	180.0	176.0	(4.0)
				Typical	179.0	175.0	(4.0)
				Low	173.0	169.0	(4.0)
75"	3840x2160	OC	High	241.0	237.0	(4.0)	
			Typical	240.0	236.0	(4.0)	
			Low	233.0	229.0	(4.0)	

资料来源: 群智咨询、华金证券研究所

**Monitor 面板:** 7 月 Monitor OC 及 LCM 面板主流规格价格仍有望持平。根据群智咨询数据: (1) 21.5"FHD, 7 月 Open cell & LCM 面板价格环比持平; (2) 23.8"FHD, 7 月 Open cell & LCM 面板价格环比持平; (3) 27"FHD, 7 月 Open cell & LCM 面板价格环比持平;

**Notebook 面板:** 由于激烈的技术竞争中高阶面板价格则呈现一定的分化。(1) 低端 HD TN: 7 月主流 TN LCM 均价预计持平; (2) IPS FHD&FHD+ 产品: 7 月 16:9 主流规格面板价格保持稳定。16:10 主流规格面板价格小幅下跌, 中高规格面板价格则继续呈分化态势。

表 3: 2025 年 7 月 IT 面板价格趋势

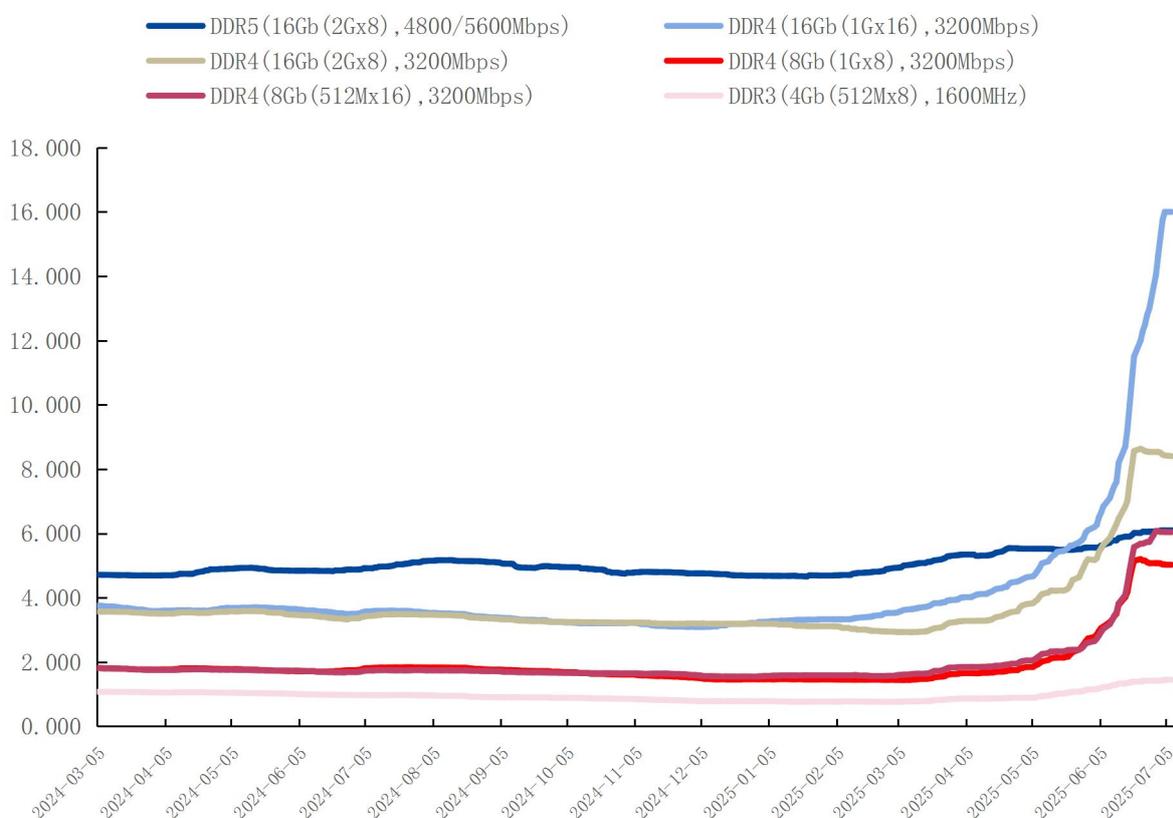
Application	Size	Resolution	OC/LCM	Spec(T)	Jun'25 (A)	Jul'25(E)	Jul. VS Jun. Change
Monitor	21.5"	1920x1080	Module	TN	34.2	34.2	0.0
	23.8"	1920x1080	Module	IPS	41.5	41.5	0.0
	23.8"	1920x1080	Open Cell	IPS	29.0	29.0	0.0
	23.8"	1920x1080	Open Cell	VA	28.2	28.2	0.0
	27"	1920x1080	Module	IPS	50.9	50.9	0.0
Notebook	14.0"	1366x768	Module	TN	26.5	26.5	0.0
	14.0"	1920x1080	Module	Value-added IPS	46.3	46.3	0.0
	14.0"	1920x1200	Module	Entry-level IPS	38.1	38.0	(0.1)
	15.6"	1920x1080	Module	Entry-level IPS	38.0	38.0	0.0

资料来源: 群智咨询、华金证券研究所

### 3.2 存储器价格

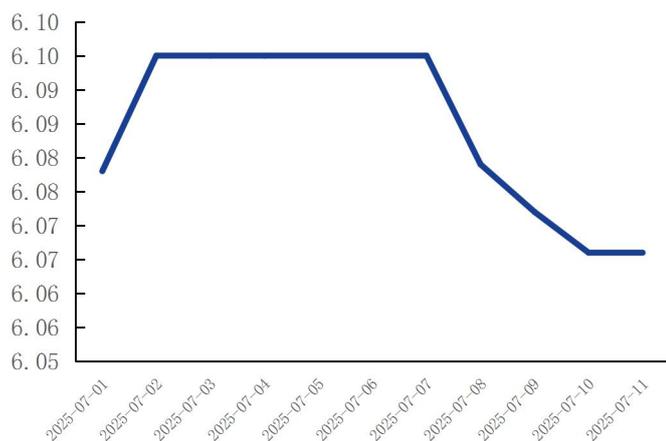
7月7日至7月11日，(1) DDR5 16G (2Gx8)4800/5600: 价格呈现下跌趋势，7月7日的6.095美元下跌至7月11日的6.066美元。(2) DDR4 16Gb(1Gx16)3200: 价格呈现下跌趋势，由7月7日的16.000美元下跌至7月11日的15.913美元。(3) DDR4 16Gb(2Gx8)3200: 价格呈现上涨趋势，由7月7日的8.404美元上涨至7月11日的8.524美元。(4) DDR4 8Gb (1Gx8) 3200: 价格呈现下跌趋势，由7月7日的5.020美元下跌至7月11日的4.968美元。(5) DDR4 8Gb (512Mx16)3200: 价格呈现上涨趋势，由7月7日的6.030美元上涨至7月11日的6.061美元。(6) DDR3 4Gb (512Mx8) 1600/1866: 价格呈现上涨趋势，由7月7日的1.464美元上涨至7月11日的1.508美元。

图 17: 2025/04/28-2025/07/11 各类 DRAM 颗粒现货价格 (美元)



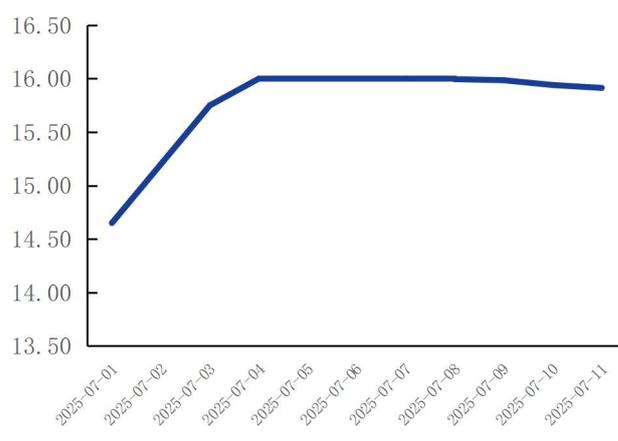
资料来源: 集邦存储市场、华金证券研究所

图 18: 近两周 DDR5 16G (2Gx8) 4800/5600 颗粒现货价格 (\$)



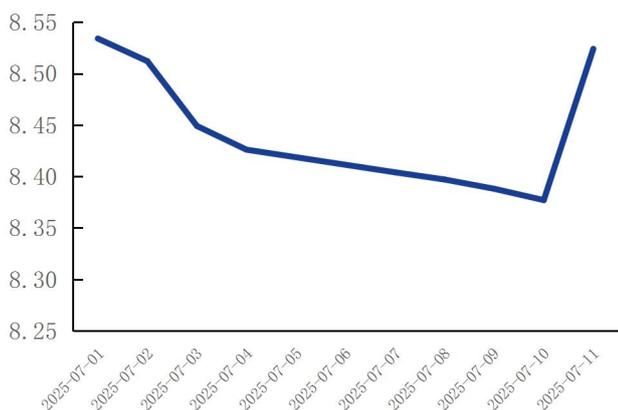
资料来源: 集邦存储市场、华金证券研究所

图 19: 近两周 DDR4 16Gb (1Gx16) 3200 颗粒现货价格 (\$)



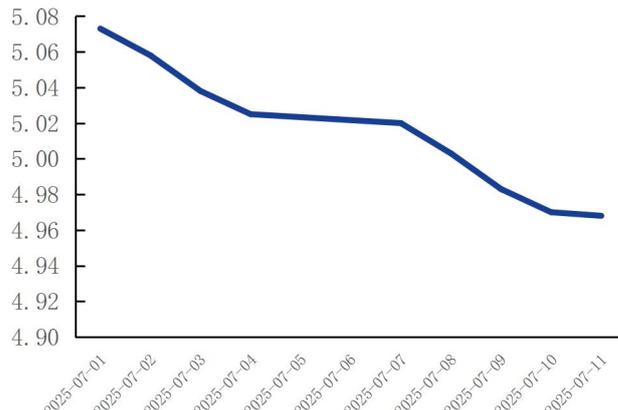
资料来源: 集邦存储市场、华金证券研究所

图 20: 近两周 DDR4 16Gb (2Gx8) 3200 颗粒现货价格 (\$)



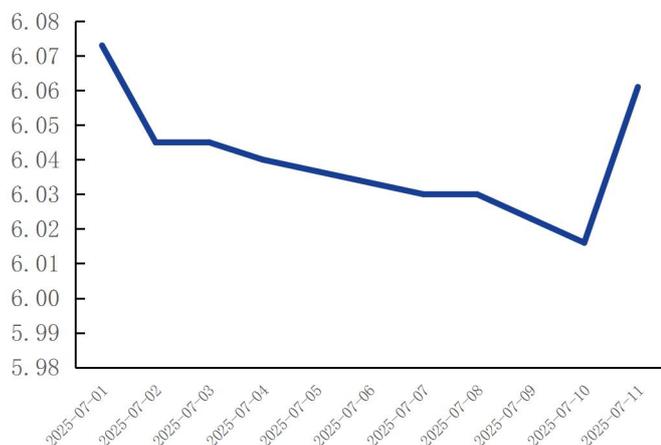
资料来源: 集邦存储市场、华金证券研究所

图 21: 近两周 DDR4 8Gb (1Gx8) 3200 颗粒现货价格 (\$)



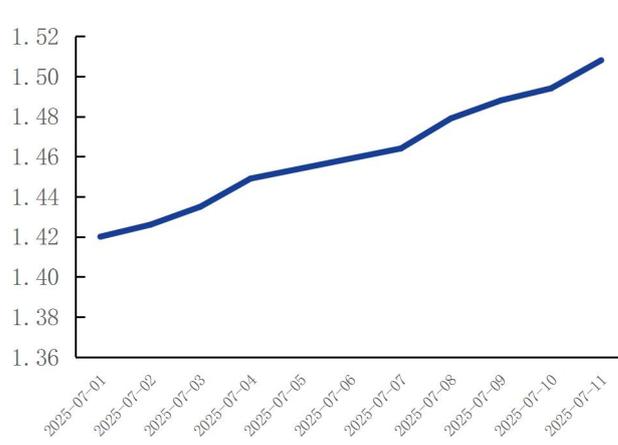
资料来源: 集邦存储市场、华金证券研究所

图 22: 近两周 DDR4 8Gb (512Mx16) 3200 颗粒现货价格 (\$)



资料来源: 集邦存储市场、华金证券研究所

图 23: 近两周 DDR3 4Gb (512Mx8) 1600/1866 颗粒现货价格 (\$)



资料来源: 集邦存储市场、华金证券研究所

## 4、风险提示

国内算力不及预期风险； AI 发展不及预期影响算力基建投资； 国际市场竞争加剧风险； 供应链变动风险， 国际制裁风险。

## 投资评级说明

公司投资评级：

买入 — 未来 6-12 个月内相对同期相关证券市场代表性指数涨幅大于 15%；

增持 — 未来 6-12 个月内相对同期相关证券市场代表性指数涨幅在 5%至 15%之间；

中性 — 未来 6-12 个月内相对同期相关证券市场代表性指数涨幅在 -5%至 5%之间；

减持 — 未来 6-12 个月内相对同期相关证券市场代表性指数跌幅在 5%至 15%之间；

卖出 — 未来 6-12 个月内相对同期相关证券市场代表性指数跌幅大于 15%。

行业投资评级：

领先大市 — 未来 6-12 个月内相对同期相关证券市场代表性指数领先 10%以上；

同步大市 — 未来 6-12 个月内相对同期相关证券市场代表性指数涨跌幅介于 -10%至 10%；

落后大市 — 未来 6-12 个月内相对同期相关证券市场代表性指数落后 10%以上。

基准指数说明：A 股市场以沪深 300 指数为基准；新三板市场以三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）为基准；香港市场以恒生指数为基准，美股市场以标普 500 指数为基准。

## 分析师声明

熊军声明，本人具有中国证券业协会授予的证券投资咨询执业资格，勤勉尽责、诚实守信。本人对本报告的内容和观点负责，保证信息来源合法合规、研究方法专业审慎、研究观点独立公正、分析结论具有合理依据，特此声明。

### 本公司具备证券投资咨询业务资格的说明

华金证券股份有限公司（以下简称“本公司”）经中国证券监督管理委员会核准，取得证券投资咨询业务许可。本公司及其投资咨询人员可以为证券投资人或客户提供证券投资分析、预测或者建议等直接或间接的有偿咨询服务。发布证券研究报告，是证券投资咨询业务的一种基本形式，本公司可以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，并向本公司的客户发布。

### 免责声明：

本报告仅供华金证券股份有限公司（以下简称“本公司”）的客户使用。本公司不会因为任何机构或个人接收到本报告而视其为本公司的当然客户。

本报告基于已公开的资料或信息撰写，但本公司不保证该等信息及资料的完整性、准确性。本报告所载的信息、资料、建议及推测仅反映本公司于本报告发布当日的判断，本报告中的证券或投资标的价格、价值及投资带来的收入可能会波动。在不同时期，本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。本公司不保证本报告所含信息及资料保持在最新状态，本公司将随时补充、更新和修订有关信息及资料，但不保证及时公开发布。同时，本公司有权对本报告所含信息在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。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，一切须以本公司向客户发布的本报告完整版本为准。

在法律许可的情况下，本公司及所属关联机构可能会持有报告中提到的公司所发行的证券或期权并进行证券或期权交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务，提请客户充分注意。客户不应将本报告为作出其投资决策的惟一参考因素，亦不应认为本报告可以取代客户自身的投资判断与决策。在任何情况下，本报告中的信息或所表述的意见均不构成对任何人的投资建议，无论是否已经明示或暗示，本报告不能作为道义的、责任的和法律的依据或者凭证。在任何情况下，本公司亦不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。

本报告版权仅为本公司所有，未经事先书面许可，任何机构和个人不得以任何形式翻版、复制、发表、转发、篡改或引用本报告的任何部分。如征得本公司同意进行引用、刊发的，需在允许的范围内使用，并注明出处为“华金证券股份有限公司研究所”，且不得对本报告进行任何有悖原意的引用、删节和修改。

华金证券股份有限公司对本声明条款具有惟一修改权和最终解释权。

### 风险提示：

报告中的内容和意见仅供参考，并不构成对所述证券买卖的出价或询价。投资者对其投资行为负完全责任，我公司及其雇员对使用本报告及其内容所引发的任何直接或间接损失概不负责。

华金证券股份有限公司

办公地址：

上海市浦东新区杨高南路 759 号陆家嘴世纪金融广场 30 层

北京市朝阳区建国路 108 号横琴人寿大厦 17 层

深圳市福田区益田路 6001 号太平金融大厦 10 楼 05 单元

电话：021-20655588

网址：[www.huajinsec.cn](http://www.huajinsec.cn)