



2025年07月06日

看好

相关研究

- 《AI 推动国产算力，先进制程版图重塑（AI 硬件深度之一暨 GenAI 系列深度之五十）》 2025/03/09
- 《AI+车载:2025 算力与算法新星熠熠!——2025 年电子&计算机投资策略》 2024/12/17

证券分析师

- 杨海晏 A0230518070003
yanghy@swsresearch.com
- 袁航 A0230521100002
yuanhang@swsresearch.com
- 刘洋 A0230513050006
liuyang2@swsresearch.com

研究支持

- 陈俊兆 A0230124100001
chenjz@swsresearch.com

联系人

- 陈俊兆
(8621)23297818×
chenjz@swsresearch.com

沐曦股份及摩尔线程获科创板 IPO 受理，算力自主化趋势再明确

——AI 算力行业跟踪点评

本期投资提示：

- 事件：**6 月 30 日，国产 GPU 芯片公司沐曦股份和摩尔线程的科创板 IPO 申请，正式获上海证券交易所受理，分别拟募资 39.04 亿元和 80 亿元。
- 沐曦：自研架构+兼容 CUDA，落地千卡集群。**1) **产品矩阵：**包括用于智算推理的曦思 N 系列，用于训推一体和通用计算的曦云 C 系列，以及用于图形渲染的曦彩 G 系列。截至 25Q1，GPU 累计销量超过 25,000 颗。2) **路线图：**曦云 C500 (XCORE1.0; HBM2e) 于 2024 年 2 月正式量产。C600 (XCORE1.5; HBM3e)、C700 预计 26Q1、27Q3 大规模量产。3) **完成 AI GPGPU 软硬件全栈技术布局。**公司管理层具有 AMD GPGPU 资深研发背景，搭建软硬全建制团队。全系列产品采用统一的 XCORE 架构，利于快速迭代。GPU 硬件在计算能力 (C500FP16/BF16 算力处于国内第一梯队)、存储 (C500 使用 HBM2e, HBM3/3e 产品研发中)、互联能力 (MetaXLink 支持 2-64 卡 Scale-up) 等维度具有领先产品力。自研的 MXMACA 软件框架兼容 CUDA。4) **“1+6+X”稳健市场策略。**“1”代表数字算力底座，“6”代表教科研、金融、交通、能源、医疗健康、大文娱六大关键行业，同时布局“X”个长尾行业市场。5) **实现千卡集群落地。**工信部评测结果显示，基于沐曦芯片的曦源一号 SADA 一期 1024 卡集群的生态兼容性、系统稳定性、线性度、模型支持度等多个维度均表现优异。
- 摩尔线程：布局图形+AI 智算，推出万卡集群。**1) **产品矩阵：**推出四代 GPU 架构，形成 AI 智算、专业图形加速、桌面级图形加速、智能 SoC 系列产品，交付模式有芯片、板卡、一体机、集群设备、SoC。2) **路线图：**AI 智算领域产品 MTT S4000 (第三代“曲院”) 于 2023 年底推出；MTT S5000 (第四代“平湖”) 通过 FP8 精度支持等创新提升性能。集群 KUAE1 (基于 S4000) 支持千卡互联；KUAE2 (基于 S5000) 于 2024 年底推出，支持万卡互联。3) **从图形切入 AI 计算，MUSA 兼容 CUDA。**摩尔线程管理层具有鲜明的英伟达技术或销售背景。自主研发的 MUSA GPU 架构同时支持 AI 计算加速、图形渲染、物理仿真和科学计算、超高清视频编解码，核心优势在于其统一编程模型和高度可扩展性。S4000 单卡支持 48GB 显存和 768GB/s 的显存带宽；MTLink 支持 2-8 卡互联，带宽达到 240GB/s。基于夸娥集群管理平台以及夸娥大模型平台，KUAE2 支持万卡级规模扩展能力，单集群可部署超 1,000 个计算节点 (每节点 8 个 OAM)，卡间互联带宽提升至 ~800GB/s。MUSA 软件栈包括指令集架构、MUSA 编程模型、驱动、运行时库、算子库、通讯库、数学库等。通过自研的 MUSIFY 工具，可以实现 CUDA 程序平滑迁移至 MUSA。5) **市场策略为消费和商业及政企市场双轮驱动。**消费端聚焦 AI 工具链开源与国际 GPU 生态兼容计算设备平台推广，培育开发者生态；企业级市场强化金融、医疗等关键基础行业垂直领域的场景化方案输出。
- 看好 AI 算力带动核心环节国产化，除待上市的 AI 芯片标的之外：**1) **基础设施：**澜起科技、德明利、联芸科技、杰华特、泰嘉股份；2) **先进制造：**中芯国际、华虹公司、长电科技、通富微电、甬矽电子；3) **服务器：**浪潮信息、神州数码、中科曙光。
- 风险提示：**技术路线不确定性、技术迭代不及预期、供应链稳定风险、竞争加剧风险



申万宏源研究微信服务号

目录

1. 沐曦：自研架构+兼容 CUDA，落地千卡集群.....	4
1.1 累计出货 2.5 万颗 GPU，C600 预计 26Q1 量产.....	4
1.2 AMD GPGPU 资深研发背景，软硬全建制团队	5
1.3 单卡性能、集群性能、生态兼容性领先	6
1.4 “1+6+X” 稳健市场策略，推动核心供应环节国产化.....	9
2. 摩尔线程：布局图形+AI 智算，推出万卡集群	11
2.1 主打全功能 GPU，AI 智算产品累计实现收入 3.36 亿元	11
2.2 英伟达系技术班底，本土化生态布局	12
2.3 MUSA 兼容 CUDA，推出万卡集群.....	13
2.4 消费和商业及政企市场双轮驱动	15
3. 风险提示	16

图表目录

图 1：沐曦股份营收结构（百万元）	4
图 2：沐曦股份产品销量（万）	4
图 3：沐曦股份财务简况（百万元）	4
图 4：沐曦股份运营费用（百万元）	4
图 5：沐曦股份股权结构（截至 2025 年 6 月 13 日）	5
图 6：沐曦芯片架构	7
图 7：MXMACA 软件栈框架结构	9
图 8：沐曦股份成本结构（百万元）	11
图 9：沐曦股份采购内容（百万元）	11
图 10：摩尔线程产品销量	12
图 11：摩尔线程产品价格	12
图 12：摩尔线程财务简况（百万元）	12
图 13：摩尔线程运营费用（百万元）	12
图 14：摩尔线程股权结构（截至 2025 年 6 月 25 日）	13
图 15：第三代 MUSA 软件栈架构	14
表 1：沐曦股份核心技术人员情况	5
表 2：AI 芯片产品技术实力对比	6
表 3：使用沐曦 GPU 硬件的智算中心/集群（部分）	8
表 4：沐曦股份主要客户情况	10
表 5：摩尔线程主要产品	11
表 6：摩尔线程核心技术人员情况	12
表 7：摩尔线程智算集群规格与可比公司的比较情况	14
表 8：摩尔线程商业拓展情况	15

1. 沐曦：自研架构+兼容 CUDA，落地千卡集群

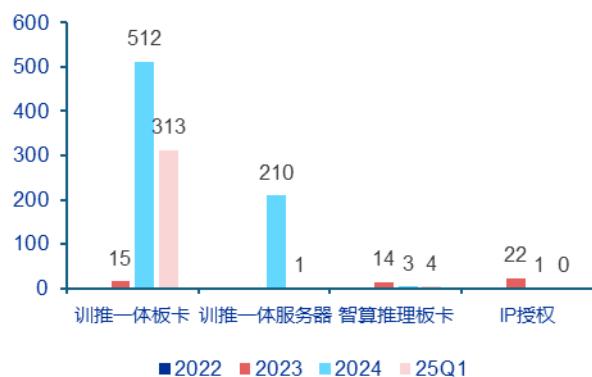
1.1 累计出货 2.5 万颗 GPU，C600 预计 26Q1 量产

沐曦集成电路于 2020 年成立于上海，主营业务是研发、设计和销售应用于人工智能训练和推理、通用计算与图形渲染领域的全栈 GPU 产品，并围绕 GPU 芯片提供配套的软件栈与计算平台。公司深度构建“1+6+X”生态与商业布局，是国内少数真正实现千卡集群大规模商业化应用的 GPU 供应商，并正在研发和推动万卡集群的落地。

公司主要交付模式有 GPU 产品、GPU 服务器产品与 IP 授权。公司产品矩阵包括用于智算推理的曦思 N 系列 GPU 产品，用于训推一体和通用计算的曦云 C 系列 GPU 产品，以及用于图形渲染的曦彩 G 系列 GPU 产品，涵盖了计算（包括训练、推理、通用计算）和渲染的全场景。2022-24 营收复合增长率为 4075%。**截至 25Q1，GPU 累计销量超过 25,000 颗。**

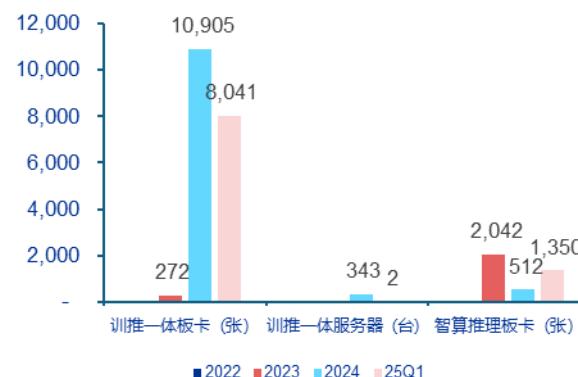
路线图：旗舰产品曦云 C500 (XCORE1.0; HBM2e) 于 2024 年 2 月正式量产。第二代 C600 (XCORE1.5; HBM3e) 已于 2024 年 10 月交付流片，并于 2025 年 5 月回片；预计 26Q1 大规模量产。第三代 C700 预计 26Q2 流片，27Q3 大规模量产。

图 1：沐曦股份营收结构（百万元）



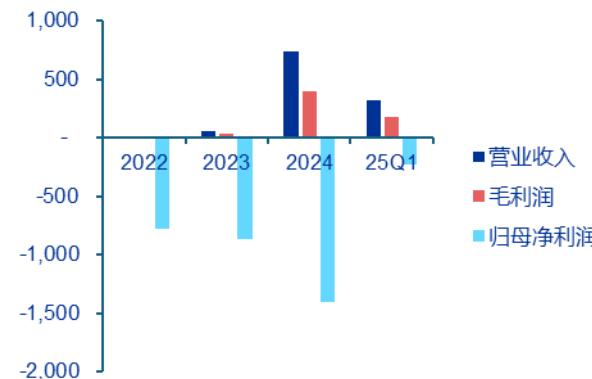
资料来源：沐曦股份招股书，申万宏源研究

图 2：沐曦股份产品销量（万）



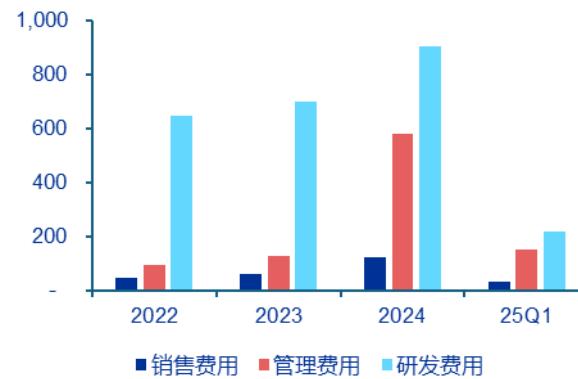
资料来源：沐曦股份招股书，申万宏源研究

图 3：沐曦股份财务简况（百万元）



资料来源：沐曦股份招股书，申万宏源研究

图 4：沐曦股份运营费用（百万元）



资料来源：沐曦股份招股书，申万宏源研究

1.2 AMD GPGPU 资深研发背景，软硬全建制团队

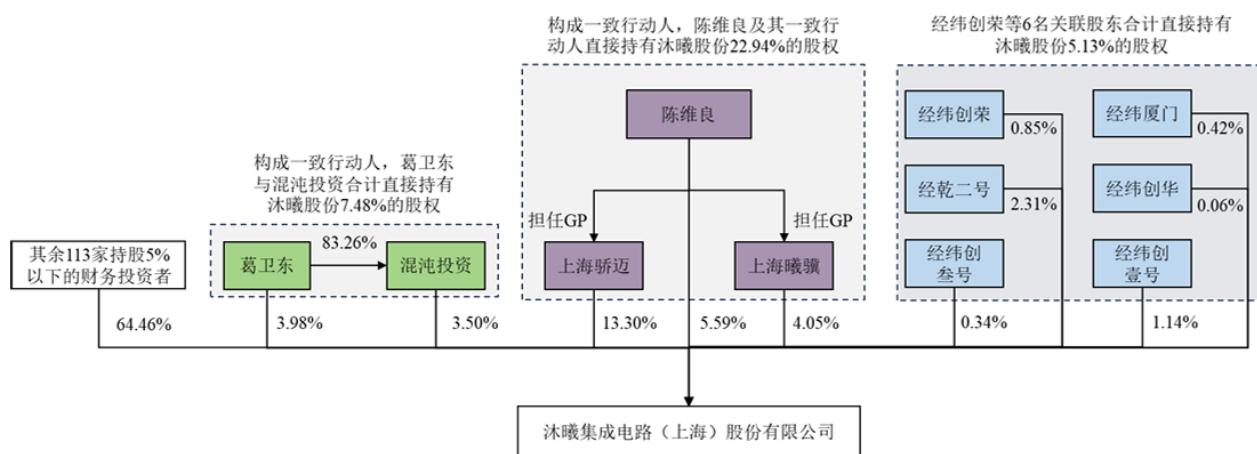
沐曦管理层有顶尖技术背景+全流程实战经验+国际化视野的鲜明特色。创始人兼CEO陈维良曾担任AMD全球GPU SoC设计总负责人、通用GPU MI产品线设计总负责人，负责全球通用计算GPU产品线的整体设计与管理，主导并完成15款高性能GPU产品的流片与量产。陈维良通过直接持股5.59%及控制上海骄迈（持股13.30%）、上海曦骥（持股3.64%），合计控制公司22.94%的股份表决权，为公司实际控制人。硬件负责人彭莉女士是AMD全球首位华人女科学家（Fellow），曾任AMD首席架构师，拥有15年高性能GPU设计经验。软件负责人杨建博士是AMD大中华地区第一位科学家（Fellow），历任AMD、海思等首席架构师，拥有20年大规模芯片及GPU软硬件设计经验。其余高管团队均拥有多年AMD、Trident、ARM、海思、英伟达、英特尔、展讯、Cadence、Synopsys、字节、澜起等顶尖芯片厂商任职经验，在芯片设计、算法、市场和运营方面拥有丰富经验。

表 1：沐曦股份核心技术人员情况

姓名	职务	学历背景	从业经历与科研成果	对公司研发活动的贡献
彭莉	董事、副总经理、首席技术官	硕士，上海交通大学电气与系统专业	拥有20年以上的高性能GPU芯片架构研发与设计经验、近40项授权发明专利，曾主导研发40nm至6nm十余款高性能芯片	(1) 作为硬件技术负责人，领导团队进行GPU产品规划、各层级硬件相关工作 (2) 把控GPU产品架构研发与设计，负责了曦思N系列、曦云C系列以及曦彩G系列等全系列芯片产品研发
杨建	董事、副总经理、首席技术官	博士，浙江大学计算机科学与技术专业	拥有20年以上高性能GPU软硬件架构研发与设计经验、近10项授权发明专利、专著1本，历任多家国际头部芯片设计公司的首席架构师，曾主导研发40nm至6nm十余款高性能芯片	(1) 作为软件技术负责人，领导团队进行软件研发工作，推动研发与设计MXMACA软件栈，包含操作系统内核驱动、用户态驱动、编译器、数学库、AI应用框架以及大型训练和推理的框架 (2) 作为沐曦光启智能研究院负责人，领导团队进行前沿工艺的技术储备相关研究
王爽	董事、资深研发总监	硕士，南京理工大学信号与信息处理专业	拥有16年以上集成电路研发与设计经验、近20项授权发明专利，成功研发从90nm到6nm多款高性能芯片，在国际集成电路设计顶级会议IEEE-ASSCC/DAC/EDTM等发表论文6篇	担任沐曦股份资深研发总监，负责或参与了公司多款芯片产品的研发与设计，主要承担芯片中系统IP开发等，担任项目SoC设计总负责人
吴志华	软件研发部副总裁	硕士，浙江大学数学专业	拥有近20年的集成电路与软件研发与设计从业经验，4项授权发明专利，对GPU产业链上下游有着资深理解	担任沐曦股份软件研发部副总裁，负责或参与了整个软件栈的设计和核心技术研发，包括从固件到内核驱动、用户态驱动，及数学与通讯库的搭建；负责公司全系列产品在应用场景的底层支持；同时也负责技术团队搭建、跨软件层次方案决策等

资料来源：沐曦股份招股书，申万宏源研究

图 5：沐曦股份股权结构（截至2025年6月13日）



资料来源：沐曦股份招股书，申万宏源研究

1.3 单卡性能、集群性能、生态兼容性领先

沐曦完成 AI GPGPU 软硬件全栈技术布局。 公司全系列产品采用统一的 XCORE 架构，兼具通用性和高性能，利于快速迭代。GPU 硬件在计算能力（C500 FP16/BF16 算力处于国内第一梯队）、存储（C500 使用 HBM2e, HBM3/3e 产品研发中）、互联能力（MetaXLink 支持 2-64 卡 Scale-up）等维度具有领先产品力。自研的 MXMACA 软件框架兼容 CUDA，兼具易用性和通用性。旗舰产品曦云 C 系列训推一体 GPU 芯片基于全自研的 GPU IP、指令集和架构，在通用性、单卡性能、集群性能及稳定性、生态兼容与迁移效率等方面均达到国内领先水平，具备较强的综合竞争力。

表 2：AI 芯片产品技术实力对比

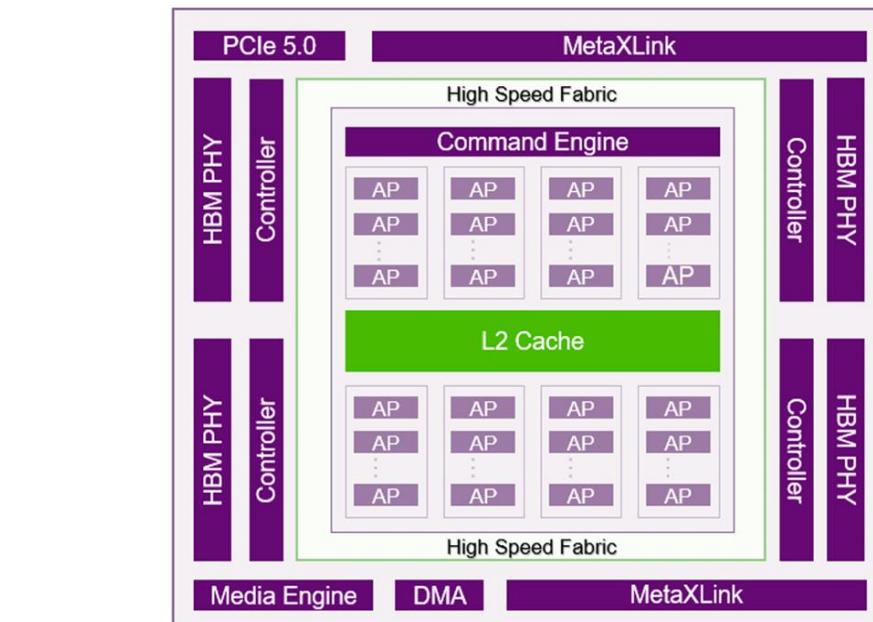
评价指标	国外企业	国内其他企业	沐曦股份
架构/通用性	英伟达、AMD 均采用通用型架构 (GPU)，并且两家公司都已对 GPU 架构进行数次迭代更新。如英伟达常见的 GPU 架构包含 Ampere 架构、Hopper 架构、Blackwell 架构。AMD 常见的有 RDNA 架构、CDNA 架构等。英伟达及 AMD 先进的计算架构决定了其领先的 GPU 计算性能。	国内企业根据自身技术特色沿用了不同的技术路径，包括通用型架构 (GPU) 和专用型架构 (如 ASIC)，不同公司的架构设计各有差异。	公司打造了自主安全可控的 GPU 架构，兼具通用性和高性能；指令集由超过 600 条 (XCORE1.0 计算 GPU)/800 条 (XCORE2.0 渲染 GPU) 指令组成，可广泛支撑多种应用。
计算能力	(1) 英伟达和 AMD 支持多种混合精度 (FP64、FP32、FP16、BF16、FP8、FP4、INT8) 等。 (2) 算力指标上，英伟达主要产品的 FP16/BF16 在 300-2,000TFLOPS 左右，AMD 主要产品的 FP16/BF16 在 300-1,300 TFLOPS 左右。	(1) 国内企业支持部分混合精度，大多不支持 FP64 且对矩阵计算的支持程度较低；采用专用型架构的企业通常以支持 FP16、BF16、INT8 为主。 (2) 算力指标上，国内多数头部企业主流在售产品的 FP16/BF16 在 100-300 TFLOPS 左右。	(1) 公司支持多种混合精度 (FP64、FP32、TF32、FP16、BF16、FP8、INT8、FP6、FP4) 等向量和矩阵计算。 (2) 算力指标上，公司曦云 C500 系列产品的 FP16/BF16 处于国内第一梯队。

		(1) 缓存方面，公司是全球少数几家掌握了复杂多级缓存结构的 GPU 企业，凭借独特的缓存组网技术，构建起多维度的缓存网络架构，有效提升缓存命中率。(2) 显存方面，公司量产产品使用 HBM2e 显存，显存带宽处于国内第一梯队，并已将 HBM3/3e 应用在迭代产品中。
存储能力	(1) 缓存方面，英伟达和 AMD 均使用了复杂的多级高速缓存架构，以提升数据访问效率，减少延迟。(2) 显存方面，英伟达和 AMD 采用高带宽显存 HBM2e、HBM3 及 HBM3e，显存带宽在 2-7TB/s 左右。	
集群能力	(1) 互连通信方面，国外的多卡互连技术主要包括 NVLink（英伟达）及 Infinity Fabric（AMD）。NVLink 现有单卡互连带宽最高约 1.8TB/s，支持 72 卡的 scale up 网络。Infinity Fabric 单卡互连带宽最高约 896GB/s。(2) 千卡集群线性度和稳定性方面，国外企业千卡集群线性度和稳定性整体较高。(3) 集群规模方面，国外企业已能够支持万卡、十万卡级超大集群。	(1) 互连通信方面，国内互连技术与国外存在一定差距，大部分企业单卡互连带宽在 200-400 GB/s。(2) 千卡集群线性度和稳定性方面，国内企业线性度和稳定性与国外存在一定差距。(3) 集群规模方面，只有少数国内企业实现了千卡集群落地运营。
软件生态	英伟达自研 CUDA，拥有成熟完善且全球领先的生态，已建立显著生态优势；AMD 自研 ROCm 并兼容 CUDA，软件生态较为丰富。	国内企业通常采取自建软件生态或兼容国外软件生态两种方式，多数国内企业无显著生态优势，适配和迁移成本较高。

资料来源：沐曦股份招股书，申万宏源研究

单卡性能方面：公司是国内首批具备先进制程 GPU 芯片设计能力和商业化落地能力的企业之一，全栈 GPU 产品基于自主研发和自主知识产权的 GPU IP、GPU 指令集和架构，单卡性能处于国内第一梯队。

图 6：沐曦芯片架构



资料来源：沐曦股份招股书，申万宏源研究

集群性能方面：公司自研的 MetaXLink 具备国内稀缺的高带宽卡间互连能力，可实现 2-64 卡多种互连拓扑，并且在智算集群的线性度和稳定性方面具有较强的产品表现。2024 年 11 月 29 日，沐曦与加佳科技共同启动曦源一号 SADA 万卡集群算力项目，其第一期千卡集群在上海正式落地。随后，一期 1024 卡集群通过工信部中国软件评测中心权威评测认证，评测结果显示，该集群在生态兼容性、系统稳定性、线性度、模型支持度等多个维度均表现优异。算丰信息在上海和中国香港运营的“丰收一号”、“丰收二号”和“丰收三号”算力集群，累计运营规模超过 3,000 卡。

表 3：使用沐曦 GPU 硬件的智算中心/集群（部分）

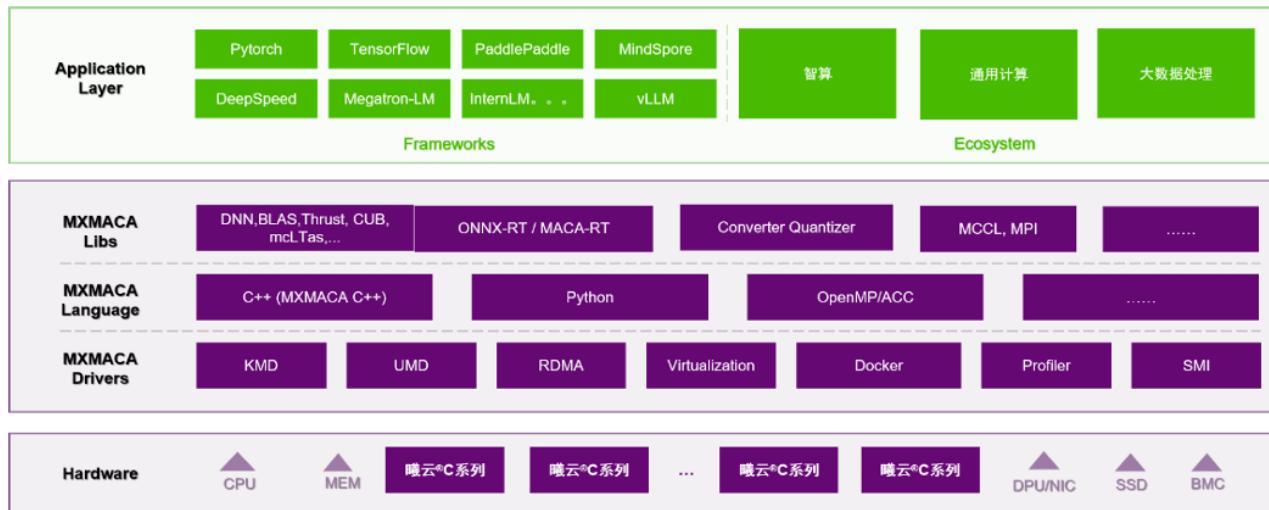
项目名称	合作建设方	地址	算力卡	上线时间	规模及建设进度	应用场景
曦源一号 SADA 万卡集群	加佳科技（运营） 上海联通（设施+调度 管理）	上海/湖 南/江苏 等	曦云 C 系列	2024 年 11 月	目标万卡集群；上海 一期 1024 卡 24M11 上线	新闻传媒、自动驾 驶、生物医药、智 能制造等
北京智源研究院 千卡集群	北京智谱	北京	曦云 C 系列	2024 年 6 月	2024 年 6 月，沐曦曦云 C 系列千卡 集群支撑北京智源研究院完成千亿参 数 MoE 大模型预训练	教科研
丰收一号/二号/ 三号	算丰信息（IDC 运营） 香港宽频（算力服务经 销）	上海、 香港	曦云 C 系列	2024- 2025 年	一号位于 Ucloud 青浦数据中心、二 号位于香港九龙湾、三号位于香港将 军澳。累计运营规模超过 3,000 卡	企业、金融、零 售、物流、工业
国家超算长沙中 心集群	加佳科技（运营） 国家超算长沙中心（设 施）	长沙	曦云 C 系列	2024 年	2025 年 2 月上线上新基于沐曦国产 GPU 的 DeepSeek-R1 大模型	智慧城市治理、医 疗

资料来源：上海联通，加佳科技，中国日报网，芯智讯，算丰信息，上海经信委，湖南大学新闻网，长沙晚报，申万宏源研究

软件生态方面：公司自主构建的 MXMACA 软件栈不仅拥有统一、完整且高效的全栈式工具链，涵盖应用开发、功能调试和性能调优等核心环节，同时高度兼容 GPU 行业国

际主流 CUDA 生态，能够开放拥抱全球开发者丰富的开源成果，具有较高的易用性和迁移效率，在通用性和灵活性上具备独特的竞争力。

图 7：MXMACA 软件栈框架结构



资料来源：沐曦股份招股书，申万宏源研究

沐曦指引 AI GPGPU 技术迭代方向。在架构方面，持续迭代 GPU 核心架构，提高 AI 算力密度；在计算方面，扩展对 FP8、FP4 等数据格式的支持度，增加多种数据转换指令；在存储方面，研发新一代显存控制器接口 HBM3/4，顺应更大容量显存需求；在互连通信方面，研发新一代 MetaXLink 互连技术，结合光互连技术和 GPU 超节点系统设计，打破传统服务器架构的通信瓶颈、丰富公司 GPU 产品的适用范围；在先进封装层面，在 GPU 产品中引入新一代芯粒架构和三维堆叠芯片技术；在软件生态方面，结合产业伙伴的需求优化软件栈的丰富度，持续提升其易用性和应用性能，进一步解决跨平台的应用兼容难题。

1.4 “1+6+X” 稳健市场策略，推动核心供应环节国产化

沐曦的市场策略为“1+6+X”生态与商业布局。“1”代表数字算力底座，包括国家人工智能公共算力平台、商用 AIDC 以及由运营商、互联网厂商等承建的智算中心等，这类客户群体将是公司坚实的基本盘。“6”代表教科研、金融、交通、能源、医疗健康、大文娱六大关键行业。同时布局“X”个长尾行业市场。

销售分为直销和经销两种模式。过往期间主营收入中经销占比逐步提升，2022-2024、25Q1 分别为 0、43.81%、37.42% 和 55.71%。经销商客户包括超讯通信（特定行业总代）、迈信林（特定行业总代）、华科智谷（全系列产品国代）、算丰信息（中国香港总代）。其中，超讯通信与公司合作较为深入，双方合资成立讯曦智能、中能建智慧科技；推出基于沐曦芯片的自有服务器品牌“元醒”。直销客户包括新华三等服务器 OEM，以

及集成+运营商包括加佳科技（共建曦源一号 SADA 万卡集群）、算丰信息（推出曦云 C550 Shanghai Cube 128 卡机柜；基于沐曦芯片运营“丰收一号/二号/三号”算力集群）。

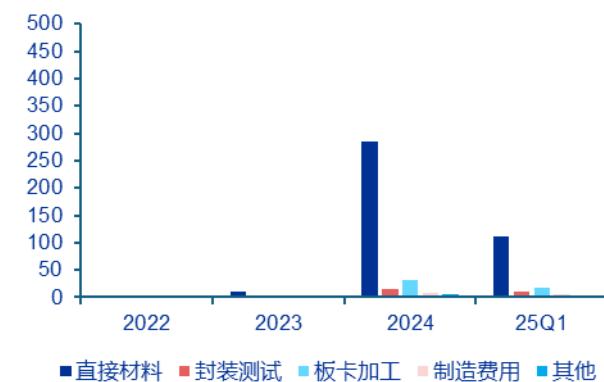
表 4：沐曦股份主要客户情况

客户	主要交付形式	客户类型	销售模式	业务规模	合作详情
新华三，紫光股份 (000938.SZ)子公司	根据订单	服务器 OEM	直销	24 年收入 1.5 亿元 (主营占比 20%) 25Q1 收入 0.3 亿元 (主营占比 9%)	通过与新华三等服务器整机厂通力合作，公司的 GPU 产品已在多个智算中心中获得规模化应用
超讯通信 (603322.SH)	板卡	集成方案商	经销	25Q1 收入 1.3 亿元 (主营占比 40%)	为沐曦特定行业总代；合资成立讯曦智能，主营芯片技术服务和服务器整机生产；推出基于沐曦芯片的自有服务器品牌“元醒”，并于 2025 年 2 月获得中特新联 8.5 亿元、星航智算 6.38 亿元智算集成服务订单；合资成立中能建智慧科技，将主要承担中国能建中电工程“东数西算”各节点大数据项目算力资源的销售和运营
迈信林(688685.SH)	板卡	集成方案商	经销	子公司瑞芯智能于 24 年贡献收入 1.3 亿元 (主营占比 18%)	为沐曦特定行业总代
华科智谷	板卡	集成方案商	经销	25Q1 收入 0.11 亿元 (主营占比 4%)	获得沐曦全系列产品的国代资格
加佳科技	板卡	集成+运营商	直销	25Q1 收入 0.91 亿元 (主营占比 28%)	合作建设曦源一号 SADA 万卡集群；上海一期 1024 卡集群 2024 年 11 月上线；经工信部认证，一期集群在生态兼容性、系统稳定性、线性度、模型支持度等多个维度均表现优异
算丰信息，亿都国际 (259.HK)子公司	服务器	集成+运营商	直销 / 经销	24 年收入 1.1 亿元 (主营占比 14%)	为中国香港总代；合作推出曦云 C550 Shanghai Cube 128 卡机柜；基于沐曦芯片，在上海和中国香港运营“丰收一号/二号/三号”算力集群，累计运营规模超过 3,000 卡。其中，丰收一号千卡集群位于优刻得青浦数据中心。
芯原股份	IP 授权	IP 供应商/芯片设计服务	直销	#N/A	2023 年签订金额 1 亿元的 IP 授权销售合同，正在履行

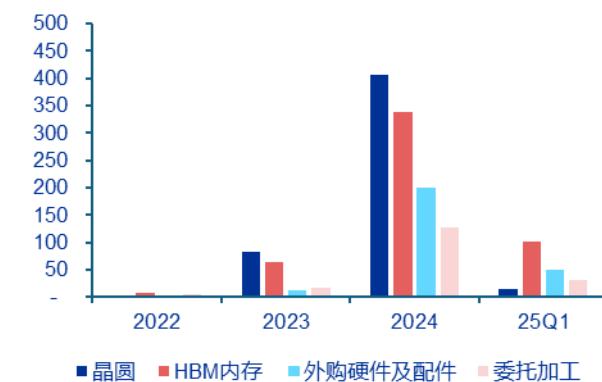
资料来源：沐曦股份招股书，超讯通信公告，国际能源网，迈信林公告，华科智谷官网，加佳科技，中国日报网，算丰信息官网，上海经信委，优刻得，申万宏源研究整理

沐曦正努力推动核心环节实现国产化与技术进步。基于国产供应链的下一代旗舰产品曦云 C600 已完成流片。公司一方面将加速公司在晶圆生产制造、封测、存储颗粒、EDA、IP 等核心环节实现国产配套，保障供应链稳定，另一方面将反哺产业链上下游，带动国内先进工艺良率爬坡以及 IP、EDA、存储颗粒等核心环节技术进步。

目前，先进制程代工晶圆、HBM 采购占比较高。公司采购的主要内容中，晶圆指由晶圆代工厂生产完成的硅片晶圆，HBM 公司内存指高带宽存储器，外购硬件及配件指对外采购的服务器、一体机等硬件设备公司及相关配件，委托加工服务中芯片的封装测试由封测厂商完成，GPU 公司板卡的组公司装由加工厂商完成，电子元器件包括 PCB 板、电容、电阻等各类型用于产品生产的电子原材料等。制造费用主要为光罩、IP 公司使用权的折旧和摊销。

图 8：沐曦股份成本结构（百万元）

资料来源：沐曦股份招股书，申万宏源研究

图 9：沐曦股份采购内容（百万元）

资料来源：沐曦股份招股书，申万宏源研究

2. 摩尔线程：布局图形+AI 智算，推出万卡集群

2.1 主打全功能 GPU，AI 智算产品累计实现收入 3.36 亿元

摩尔线程于 2020 年成立于北京，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，并形成了覆盖 AI 智算、高性能计算、图形渲染、计算虚拟化、智能媒体和面向个人娱乐与生产力工具等应用领域的多元计算加速产品矩阵，产品线涵盖政务与企业级智能计算、数据中心及消费级终端市场。

公司交付模式有芯片、板卡、一体机、集群设备、SoC。公司成功自主研发并量产四代高性能 GPU 架构，并基于此推出了十余款高性能 GPU 加速卡与模组，以及 D800 计算服务器和 KUAE 智能计算集群等产品，构建了包含芯片设计、硬件产品及软硬一体化解决方案的全方位自主可控产品体系。截至 2024 年底，AI 智算产品合计实现收入 3.36 亿元，

路线图：AI 智算领域产品 MTT S4000（第三代“曲院”）于 2023 年底推出；MTT S5000（第四代“平湖”）通过 FP8 精度支持等创新提升性能。集群 KUAE1（基于 S4000）支持千卡互联；KUAE2（基于 S5000）于 2024 年底推出，支持万卡互联。

表 5：摩尔线程主要产品

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” 第三代 GPU “曲院”	S5000 S4000	D800 X1/X	KUAE2 KUAE1
	专业图形 加速 企业级	第二代 GPU “春晓” 第一代 GPU “苏堤”	S3000 S1000/S2000	D200/D400/D800	MCCX
	桌面级图形 加速	消费级 第二代 GPU “春晓”	S70/S80		
		企业级 第二代 GPU “春晓” 第一代 GPU “苏堤”	X300 S10/S30/S50/X100		
智能 SoC 类	企业级 消费级	第一代 SoC “长江”	AI 模组-E300	AI 算力本-A140	

资料来源：摩尔线程招股书，申万宏源研究



图 10：摩尔线程产品销量

销量 (个)	2022	2023	2024
一、AI智算			
集群			3
板卡			2,058
一体机			52
二、专业图形加速			
集群		2	
板卡	4,431	3,074	5,699
一体机	2	13	8
三、桌面级图形加速			
板卡	438	11,672	4,659
一体机		19	20
芯片	173,884	191,473	28,000

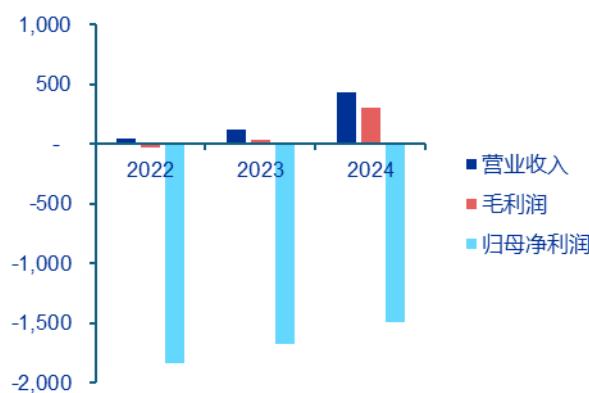
资料来源：摩尔线程招股书，申万宏源研究

图 11：摩尔线程产品价格

价格	2022	2023	2024
一、AI智算 (万元/个)			
集群			6,113
板卡			6.02
一体机			54
二、专业图形加速 (万元/个)			
集群			3,235
板卡		0.29	0.75
一体机		13	17
三、桌面级图形加速 (元/个)			
板卡		1,624	538
一体机	-		6,450
芯片		184	117
			89

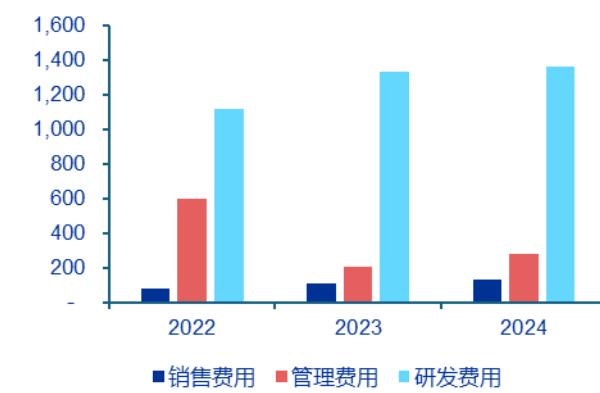
资料来源：摩尔线程招股书，申万宏源研究

图 12：摩尔线程财务简况 (百万元)



资料来源：摩尔线程招股书，申万宏源研究

图 13：摩尔线程运营费用 (百万元)



资料来源：摩尔线程招股书，申万宏源研究

2.2 英伟达系技术班底，本土化生态布局

摩尔线程管理层具有鲜明的英伟达技术或销售背景。创始人兼任总经理张建中曾任英伟达全球副总裁兼大中华区总经理，主导构建了英伟达在中国的完整生态体系。张建中先生通过直接持有或间接控制一致行动人，合计控制公司 36.36% 的股份。联合创始人、职工董事周苑曾任英伟达市场生态高级总监，主导市场生态搭建。联合创始人、董事、副总经理、CTO 张钰勃曾任英伟达 GPU 架构师，在摩尔线程负责芯片架构设计与技术研发。副总经理、战略合作总经理宋学军曾任英伟达高级销售经理，在摩尔线程负责 GPU 产品市场推广，推动公司 AI 智算、图形加速产品商业化落地。副总经理、软件研发部总经理杨上山曾任英伟达 GPU 架构师，在摩尔线程主导软件研发与硬件协同。

表 6：摩尔线程核心技术人员情况

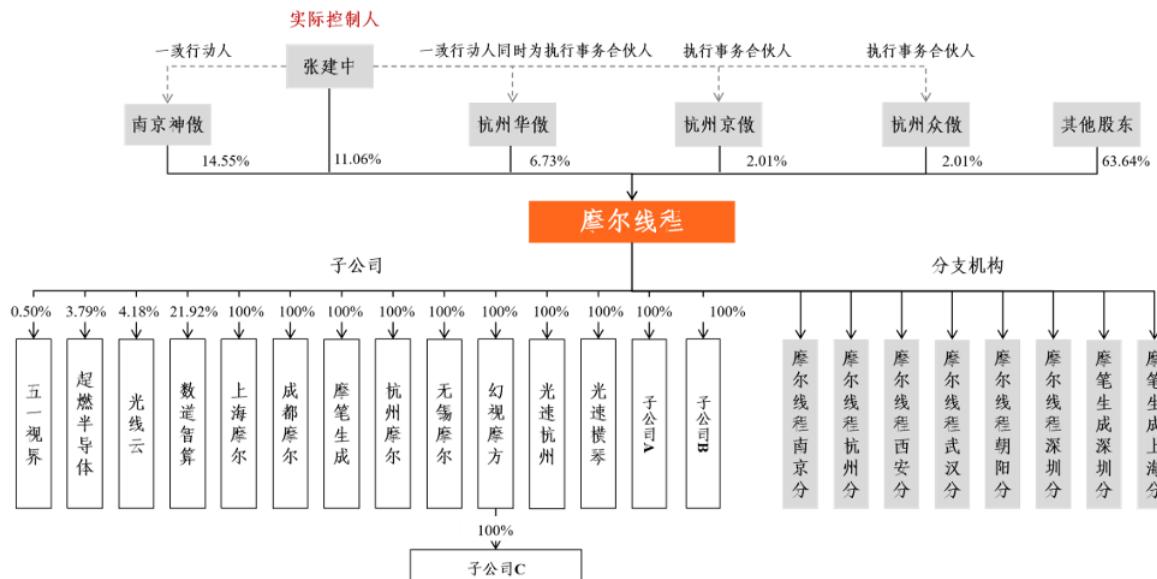
姓名	职位	履历	贡献



张钰勃	董事、副总经理、CTO	英伟达 GPU 架构师、Pony AI Inc. 基础架构部门主任工程师	负责产品研发战略、技术路线规划、核心专利布局，包括 MUSA 系统架构定义、GPU 指令集定义、芯片和软件架构设计
马凤翔	芯片研发部总经理	北京中星微电子有限公司资深芯片设计经理、地平线公司芯片研发总监	负责芯片项目研发管理，包括数字设计、物理设计、设计验证、DFT 和封装设计
杨上山	副总经理、软件研部总经理	上海贝尔阿尔卡特有限公司软件工程师、爱立信（中国）通信有限公司软件工程师、英伟达 GPU 架构师	负责 MUSA 软件研发管理，包括芯片固件、驱动软件、编译器、计算加速库、开发者工具、应用软件和 GPU 研发基础设施平台
王华	云计算与人工智能事业部总经理	威睿信息技术（中国）有限公司网络与安全部门高级研发经理、华为技术有限公司云网络高级专家、深信服科技股份有限公司 CTO	负责研发云计算平台、虚拟化技术、网络技术、AI 智算中心集群系统、基础设施及应用

资料来源：摩尔线程，申万宏源研究

图 14：摩尔线程股权结构（截至 2025 年 6 月 25 日）



资料来源：摩尔线程招股书，申万宏源研究

2.3MUSA 兼容 CUDA，推出万卡集群

架构：自主研发的 MUSA GPU 架构已迭代四代（苏堤、春晓、曲院和平湖），同时支持 AI 计算加速、图形渲染、物理仿真和科学计算、超高清视频编解码，核心优势在于其统一编程模型和高度可扩展性。

单卡：根据摩尔线程官网，S4000 单卡支持 48GB 显存和 768GB/s 的显存带宽；MTLLink 支持 2-8 卡互联，带宽达到 240GB/s；可加速集群从 64 卡到 1024 卡的训练速度以及多卡互连的线性度。

集群：基于夸娥集群管理平台以及夸娥大模型平台，KUAE2 支持万卡级规模扩展能力，单集群可部署超 1,000 个计算节点（每节点 8 个 OAM），卡间互联带宽提升至 ~800GB/s。

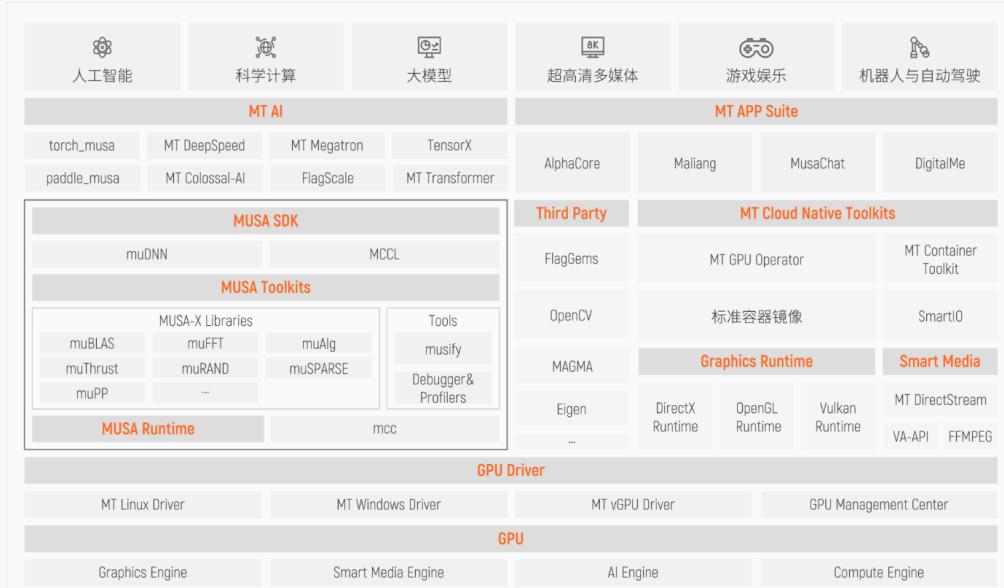
软件：MUSA 软件栈包括指令集架构、MUSA 编程模型、驱动、运行时库、算子库、通讯库、数学库等。通过自研的 MUSIFY 工具，可以实现 CUDA 程序平滑迁移至 MUSA。

表 7：摩尔线程智算集群规格与可比公司的比较情况

公司名称	代表产品	集群规模	FP16 Tensor (EFLOPs)	GPU 数目	网络方案	卡间互连 (GB/s)
摩尔线程	KUAE2	万卡级	不披露	10,240 张	IB/RoCE	~800
微软/OpenAI	Azure AI SuperCluster	十万卡级	100 (预估)	10 万张 H100	IB	900
Meta	AI Research SuperCluster (RSC)	十万卡级	100 (预估)	10 万张 H100	IB/RoCE	900
xAI	Colossus 超级计算集群	十万卡级	200	20 万张 H100	IB	900

资料来源：摩尔线程招股书，申万宏源研究

图 15：第三代 MUSA 软件栈架构



资料来源：摩尔线程官网，申万宏源研究

2.4 消费和商业及政企市场双轮驱动

公司采用直销与经销并存的销售模式。初期以经销为主，过往期间主营收入中直销占比逐步提升，2022-2024 分别为 1%、61% 和 41%。核心经销商为极致电子技术，代理产品包括桌面级图形加速、专业图形加速、AI 智算系列的板卡/芯片/一体机产品，2022-2024 年贡献收入主营占比 67%、39%、2%；客户 J 为核心为业内知名经销商，主要代理专业图形加速板卡、AI 智算板卡与集群，2024 年销售收入 1.54 亿元（主营占比 36%）；北京质能芯科技主要代理 AI 智算一体机，2024 年销售收入 0.28 亿元（主营占比 6%）。直销客户包括客户 C（AI 智算集群设备），2024 年对其销售收入为 1.65 亿元（主营占比 38%）。公司于 2023 年还对百度、中国邮电器材、京东平台销售专业图形加速集群设备或桌面级加速板卡。

摩尔线程的市场策略为消费和商业及政企市场双轮驱动。消费端聚焦 AI 工具链开源与国际 GPU 生态兼容计算设备平台推广，培育开发者生态；企业级市场强化金融、医疗等关键基础行业垂直领域的场景化方案输出。

表 8：摩尔线程商业拓展情况

客户类别	合作模式
运营商/国央企	2024.10 摩尔线程全功能 GPU 和 MUSA 自研软件栈与中国移动原生算力平台深度适配，打造智算基础算力底座； 2024.7 摩尔线程联合中国移动通信集团青海有限公司、中国联通青海分公司、中国能源建设股份有限公司总承包公司等就青海零碳产业园万卡集群项目、青海高原夸娥万卡集群项目、广西东盟万卡集群项目战略签约。
地方政府	2024.5 摩尔线程与盐城市盐南高新区、东华软件达成战略合作。三方将在盐城共同建设华东（盐城）国产算力产业基地。
银行金融机构	2024.10 摩尔线程与浦发银行、邮储银行和北京银行就“摩尔线程国产全功能 GPU 万卡通用智算集群科研项目”签署了重大科技项目合作协议； 2024.9 摩尔线程与建设银行北京市分行签约合作； 2023.6 摩尔线程与北京银行达成战略合作； 2022.6 摆尔线程与招商银行北京分行签署战略合作协议
大模型创业公司	已拓展了智谱 AI、智源研究院、无问芯穹、滴普科技、瑞莱智慧、实在智能、因识智能等国内众多大模型企业
软件服务/企业	2024.8 摆尔线程与东华软件合作，完成了政企客户多种业务系统的 AI 升级 2023.6 摆尔线程与弘信电子签署《战略合作框架协议》，将合作构建面向元宇宙场景的智能计算中心，为客户提供定制化的工厂管理和智慧办公解决方案等。此外，弘信电子还将使用摩尔线程的 GPU 算力服务，开发高效智能的工业质检应用系统。

资料来源：芯智讯，中国日报网，腾讯新闻，搜狐网，摩尔线程官网，弘信电子公告，申万宏源研究

公司正在推动国产化供应替代。公司正与国内 Foundry 联合开发 FinFET 工艺设计套件（PDK），开展 GPU 关键模块的技术研发和流片验证。联合国内封装测试厂商，完成 Chiplet 与 2.5D 封装（国产硅中介层）量产和测试，提升互连密度、性能，降低功耗，实现了先进封装测试国产化。

3. 风险提示

技术路线不确定性：AI 芯片技术仍处于发展阶段，技术迭代速度较快，技术发展路径尚在探索中，尚未形成具有绝对优势的架构和系统生态。

技术迭代不及预期：AI 芯片属于前沿核心科技领域，现有产品升级更新和新产品开发需要持续投入大量的资金和人员，但研发项目的进程及结果具有不确定性，如果厂商在研发方向上未能做出正确判断，在研发过程中未能持续突破关键技术或性能指标未达预期，厂商将面临前期研发投入难以收回、预计效益难以达到的风险，将对厂商业绩产生不利影响。

供应链稳定风险：AI 芯片公司通常采用 Fabless 模式经营，供应商包括 EDA 工具厂商、IP 授权厂商、服务器厂商、晶圆制造厂和封装测试厂等。由于集成电路领域专业化分工程度及技术门槛较高，部分供应商的产品具有稀缺性和专有性，如不能与其保持稳定的合作关系，或由于地缘政治、厂商处于实体清单等其他外部环境因素导致供应商中止与厂商的业务合作，厂商更换新供应商的代价较高，将对厂商生产经营、研发造成不利影响。

竞争加剧风险：随着 AI 应用及算法的逐步普及，AI 芯片受到了多家集成电路企业的重视，该领域也成为多家初创集成电路设计公司发力的重点。随着越来越多的厂商推出 AI 芯片产品，该领域市场竞争日趋激烈。而目前，英伟达在全球 AI 芯片领域中仍占有绝对优势。



信息披露

证券分析师承诺

本报告署名分析师具有中国证券业协会授予的证券投资咨询执业资格并注册为证券分析师，以勤勉的职业态度、专业审慎的研究方法，使用合法合规的信息，独立、客观地出具本报告，并对本报告的内容和观点负责。本人不曾因，不因，也将不会因本报告中的具体推荐意见或观点而直接或间接受到任何形式的补偿。

与公司有关的信息披露

本公司隶属于申万宏源证券有限公司。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。本公司关联机构在法律许可情况下可能持有或交易本报告提到的投资标的，还可能为或争取为这些标的提供投资银行服务。本公司在知晓范围内依法合规地履行披露义务。客户可通过 compliance@swsresearch.com 索取有关披露资料或登录 www.swsresearch.com 信息披露栏目查询从业人员资质情况、静默期安排及其他有关的信息披露。

机构销售团队联系人

华东组	茅炯	021-33388488	maojiong@swhysc.com
银行团队	李庆	021-33388245	liqing3@swhysc.com
华北组	肖霞	010-66500628	xiaoxia@swhysc.com
华南组	张晓卓	13724383669	zhangxiaoze@swhysc.com
华东创新团队	朱晓艺	021-33388860	zhuxiaoyi@swhysc.com
华北创新团队	潘烨明	15201910123	panyeming@swhysc.com

股票投资评级说明

证券的投资评级：

以报告日后的 6 个月内，证券相对于市场基准指数的涨跌幅为标准，定义如下：

买入 (Buy)	: 相对强于市场表现 20%以上；
增持 (Outperform)	: 相对强于市场表现 5% ~ 20%；
中性 (Neutral)	: 相对市场表现在 - 5% ~ + 5% 之间波动；
减持 (Underperform)	: 相对弱于市场表现 5%以下。

行业的投资评级：

以报告日后的 6 个月内，行业相对于市场基准指数的涨跌幅为标准，定义如下：

看好 (Overweight)	: 行业超越整体市场表现；
中性 (Neutral)	: 行业与整体市场表现基本持平；
看淡 (Underweight)	: 行业弱于整体市场表现。

我们在此提醒您，不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系，表示投资的相对比建议；投资者买入或者卖出证券的决定取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。申银万国使用自己的行业分类体系，如果您对我们的行业分类有兴趣，可以向我们的销售员索取。

本报告采用的基准指数 : 沪深 300 指数

法律声明

本报告由上海申银万国证券研究所有限公司（隶属于申万宏源证券有限公司，以下简称“本公司”）在中华人民共和国内地（香港、澳门、台湾除外）发布，仅供本公司的客户（包括合格的境外机构投资者等合法合规的客户）使用。本公司不会因接收人收到本报告而视其为客户。客户应当认识到有关本报告的短信提示、电话推荐等只是研究观点的简要沟通，需以本公司 <http://www.swsresearch.com> 网站刊载的完整报告为准，本公司接受客户的后续问询。

本报告是基于已公开信息撰写，但本公司不保证该等信息的真实性、准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他投资标的的邀请。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动。在不同时期，本公司可发出与本报告所载资料、意见及推测不一致的报告。

客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为作出投资决策的惟一因素。客户应自主作出投资决策并自行承担投资风险。本公司特别提示，本公司不会与任何客户以任何形式分享证券投资收益或分担证券投资损失，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。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投資目标、财务状况或需要。本公司强烈建议客户应考虑本报告的任何意见或建议是否符合其特定状况，以及（若有必要）咨询独立投资顾问。在任何情况下，本报告中的信息或所表述的意见并不构成对任何人的投资建议。在任何情况下，本公司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。市场有风险，投资需谨慎。若本报告的接收人非本公司的客户，应在基于本报告作出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。

版权均属本公司。未经本公司事先书面授权，本报告的任何部分均不得以任何形式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记，未获本公司同意，任何人均无权在任何情况下使用他们。