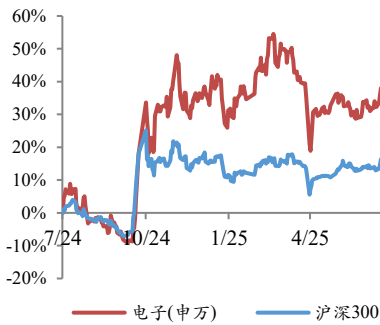


电子行业周报：科创招股书梳理之沐曦篇

行业评级：增持

报告日期：2025-07-07

行业指数与沪深300走势比较



分析师：陈耀波

执业证书号：S0010523060001

邮箱：chenyaobo@hazq.com

分析师：李元晨

执业证书号：S0010524070001

邮箱：liyyc@hazq.com

相关报告

- 关税影响下的自主可控持续关注 20250420
- 构建自主可控产业链,关税战的本质是科技战 20250413
- 地缘影响下科技自主可控重要性凸显,成熟制程工艺反倾销调查国产替代有望加速 20250405
- 国产存储双雄崛起,存储芯片国产化持续进行 20250621

主要观点：

● 本周行情回顾

从指数表现来看,本周(2025-06-30至2025-07-04),上证指数周涨跌+1.40%,深证成指涨跌幅为+1.25%,创业板指数涨跌幅+1.50%,科创50涨跌幅为-0.35%,申万电子指数涨跌幅+0.74%。板块行业指数来看,表现最好的是集成电路封测,涨幅为8.32%,面板表现较弱,涨幅为3.11%;板块概念指数来看,表现最好的是印制电路板指数,涨幅为8.31%,表现最弱的是数字芯片设计,跌幅为2.38%。

● 沐曦集成电路：历经五载发展风雨兼程，公司成为国内高性能 GPU 产品的主要领军企业

2020-2022年原始积累阶段，以陈维良先生、彭莉女士、杨建先生为核心的创始团队于2020年9月创立沐曦，经过“从零到一”全自研的技术沉淀，沐曦首款智算推理芯片曦思 N100 于2022年1月交付流片，2022年9月流片成功。并于2023年4月正式量产。

2023-2024年产品矩阵逐渐丰富，曦思 N100 芯片的设计、流片到量产，为公司打通了从研发到生产再到销售的全业务流，完善了产品、生产供应链和销售渠道的资源储备，奠定了公司的经营模式和商业模式。在此基础上，2024年2月，沐曦训推一体芯片曦云 C500 正式量产，同时沐曦提前布局国产供应链，并基于国产供应链研发设计了新一代训推一体芯片曦云 C600，于2024年10月交付流片，为进一步提升国产化奠定基础。

2025年至今，截至2025年第一季度，沐曦 GPU 产品累计销量超过 25,000 颗。沐曦深度构建“1+6+X”生态与商业布局，基础算力底座方面相继交付 9 大智算集群，算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心，区域横跨北京、上海、杭州、长沙、中国香港等地区，并逐渐向更多区域延伸。

● 沐曦主要产品收入来自沐曦曦云 C 系列训推一体 GPU

2022年，沐曦实现首款产品曦思 N 系列小批量销售实现营业收入 42 万元；2023 年曦思 N 系列产品实现正式量产，带动公司 2023 年当年智算推理 GPU 产品销售收入及 IP 授权收入大幅增长。

2024 年 2 月，沐曦核心产品训推一体 GPU 产品曦云 C 系列正式量产，公司加大力度开拓下游客户，叠加人工智能与大模型、智算中心建设的蓬勃发展以及 GPU 国产替代加速，公司 2024 年训推一体系列产品销量、销售收入均实现大幅增长。核心产品训推一体 GPU 板卡是公司收入增长的核心来源，分别占比 2023 年、2024 年和 2025 年 Q1 季度的收入的 30.09%、68.99%和 97.55%

2025 年 1-3 月，沐曦主营业务收入延续了良好的增长趋势，核心产品训推一体 GPU 销量、销售收入均大幅增长。随着公司产品矩阵逐渐丰富以及销量的增长，公司销售收入的季节性特征将进一步减弱。

● **沐曦与国产供应链加强合作，未来或进一步提升国产供应链使用比例**

训推一体 GPU 曦云 C 系列最新一代在研产品为曦云 C600。该产品基于国产供应链，采用沐曦自研的 XCORE1.5 架构及指令集，增加了 FP8 Tensor 及 Tensor 转置指令，面向云端人工智能训练与推理、通用计算、AI for Science 等计算任务。曦云 C600 采用 HBM3e 显存，支持卡间高速互连。

曦思 N 系列最新一代在研产品为曦思 N300。该产品基于国产供应链，采用沐曦自研的 XCORE1.5 架构及指令集，面向云端推理计算任务。曦思 N300 采用 HBM3 显存，典型应用场景为云端智算中心以及搭载 DeepSeek 等主流大模型的一体机、液冷工作站等。随着沐曦最新一代的产品导入国产供应链设计，未来或将进一步提升国产供应链在国产 GPU 中的使用占比。

● **创始团队行业经验丰富，推动公司研发行稳致远**

沐曦创始人、CEO 陈维良曾在全球顶尖的 GPU 厂商担任全球 GPU SoC 设计总负责人、通用 GPUMI 产品线设计总负责人，拥有出色的团队管理能力，顶尖的技术研发和量产经验。硬件首席架构师彭莉女士是 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。软件首席架构师杨建博士是 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。

● **沐曦汇聚诸多知名投资人和国资及市场化机构**

在申请上市前，沐曦股份共进行了八轮融资，诸多知名机构参与投资。国资方面，上海科创基金、引领区基金、国调基金等纷纷入局；市场化机构方面，红杉中国、经纬中国、葛卫东的混沌投资、和利资本、东方富海、光速创投、复星锐正、真格基金等知名机构也参与其中。

● **募资投入加速核心产品迭代和商业化落地**

沐曦本次募资主要用于“新型高性能通用 GPU 研发及产业化项目”，目标是研发公司曦云 C 系列训推一体芯片的第二代和第三代产品，包括 C600 和 C700 两个子项目。

曦云 C600 是基于国产先进工艺开发的通用 GPU 芯片，于 2024 年 2 月立项，计划建设期为 3 年。预计曦云 C 系列训推一体 C600 芯片将于 2025 年 Q4 实现小批次量产，并在 2026 年实现大规模量产和持续的算法优化。

曦云 C700 是基于国产先进工艺开发的下一代通用 GPU 芯片，拥有更高的性能，预计曦云 C 系列训推一体 C700 芯片将于 2026 年 Q2 实现流片和测试任务，并于 2027 年 Q3 开始实现小批次量产，并同步在 2027 年下半年实现大规模量产和持续的算法优化。

● **风险提示**

1) 下游需求不及预期；2) 资本开支不及预期；3) 技术迭代不及预期。

正文目录

1 沐曦集成电路：国内高性能 GPU 产品的主要领军企业，把握时代机遇助力国产替代	5
1.1 历经五载发展风雨兼程，公司成为国内领先的 GPU 公司.....	5
1.2 沐曦公司核心产品包括智算推理的曦思 N 系列和用于训推一体和通用计算的曦云 C 系列.....	6
2 沐曦智算推理曦思 N 系列奠定公司发展基础，训推一体曦云 C 系列接棒助力国产算力发展	9
2.1 沐曦公司产品线日趋丰富，营收快速增长，季节性特征逐渐减弱.....	9
2.2 沐曦曦云 C 系列训推一体 GPU 板卡是公司核心营收来源，分别占比 2023 年、2024 年和 2025 年 Q1 季度总收入的 30.09%、68.99%和 97.55%.....	11
2.3 沐曦训推一体 GPU 服务器占比公司 2024 年营业收入的 28.29%.....	12
2.4 沐曦曦思 N 系列面向传统人工智能是公司 2022 年推出的首款产品.....	12
2.5 沐曦公司借助经销模式扩大终端客户覆盖范围提升产品服务能力.....	13
2.6 沐曦汇聚诸多知名投资人和国资及市场化机构.....	14
3 沐曦加强国产供应链合作，持续研发投入夯实竞争优势	15
3.1 沐曦公司与供应链伙伴持续合作，加强国产供应链配置.....	15
3.2 沐曦公司重视研发投入，夯实竞争优势，核心创始团队产业经验积累丰富.....	17
3.3 沐曦公司持续积累核心技术并承担国家重大专项.....	18
3.4 沐曦募资项目重点投入高性能通用 GPU，核心产品将于 2026 年(第二代高性能通用 GPU 芯片(C600))和 2027 年下半年(第三代高性能通用 GPU 芯片(C700))加速商业化落地量产.....	23
3.5 沐曦具备多重竞争优势包括技术优势、通用性优势、高效能优势、易用性优势、产业化优势、团队优势.....	25
4 市场行情回顾	28
4.1 行业板块表现.....	28
4.2 电子个股表现.....	31
风险提示：	31

图表目录

图表 1 沐曦集成电路历史沿革	6
图表 2 沐曦集成电路公司核心产品	7
图表 3 沐曦集成电路公司核心产品（训推一体 GPU——曦云 C 系列）	8
图表 4 沐曦集成电路公司核心产品（智算推理 GPU——曦思 N 系列）	8
图表 5 沐曦集成电路公司核心产品产量和销量	9
图表 6 沐曦集成公司 2022-2025Q1 营业收入主要拆分	10
图表 7 沐曦集成公司 2022-2025Q1 营业收入按照季度拆分	10
图表 8 沐曦集成公司 2022-2025Q1 销售收入和销售数量及单价（训推一体 GPU 板卡）曦云 C	11
图表 9 沐曦集成公司 2022-2025Q1 销售收入和销售数量及单价（训推一体 GPU 服务器）	12
图表 10 沐曦集成公司 2022-2025Q1 销售收入和销售数量及单价（智算推理 GPU 板卡）曦思 N	13
图表 11 沐曦集成公司 2022-2025Q1 销售收入及收入占比（直销 vs 经销）	13
图表 12 沐曦集成公司 2022-2025Q1 前五大客户销售金额和占比	14
图表 13 沐曦集成公司截至 2025 年 3 月 31 日履行完毕和正在履行的合同金额在 7,500 万元以上的销售合同	14
图表 14 沐曦集成公司股权结构	15
图表 15 沐曦集成公司主要采购物料金额和占比（晶圆和 HBM 内存在成本占比中较高）	16
图表 16 沐曦集成公司前五大供应商采购情况和占比	16
图表 17 沐曦集成公司主要创始人和核心技术人员	17
图表 18 沐曦集成公司核心技术（高性能硬件设计技术）（公司核心技术均自主研发并实现国内领先）	19
图表 19 沐曦集成公司核心技术（高性能软件设计技术）（基于 MXMACA 软件栈，应用于全系列 GPU 产品）	21
图表 20 沐曦集成公司承担的重大科研项目	22
图表 21 沐曦集成公司在研发项目和进展	23
图表 22 沐曦集成公司募集资金运用领域	23
图表 23 沐曦曦云 C600 训推一体项目投资具体拆分	24
图表 24 沐曦曦云 C600 训推一体项目实施进度安排	24
图表 25 沐曦曦云 C700 训推一体项目投资具体拆分	24
图表 26 沐曦曦云 C700 训推一体项目实施进度安排	24
图表 27 沐曦新一代人工智能推理 GPU 研发及产业化项目投入拆分	25
图表 28 沐曦新一代人工智能推理 GPU 研发及产业化项目进展安排	25
图表 29 沐曦面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目投入拆分	25
图表 30 沐曦面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目进展安排	25
图表 31 沐曦集成公司具备六大优势（技术、通用性、高效能、易用性、产业化、世界级技术团队）	26
图表 32 沐曦（发行人）与竞争对手对比	27
图表 33 沐曦（发行人）与竞争对手对比（主要财务数据）	27
图表 34 板块指数	28
图表 35 行业板块涨跌幅和换手率（本周电子在申万一级行业指数中 16/26 位）	28
图表 36 电子行业细分板块涨跌幅和换手率	29
图表 37 电子行业热门细分概念指数涨跌幅和换手率	29
图表 38 电子行业热门细分概念指数涨跌幅和换手率	30
图表 39 电子行业行情图	30
图表 40 个股涨跌幅（%）	31

1 沐曦集成电路：国内高性能 GPU 产品的主要领军企业，把握时代机遇助力国产替代

1.1 历经五载发展风雨兼程，公司成为国内领先的 GPU 公司

(1) 沐曦公司历经三个发展阶段，核心技术不断创新积累，GPU 产品逐渐丰富，产品影响力持续增强

1、研发优势:自设立即专注 GPU 业务，打造全建制团队并形成相关技术的原始积累(2020-2022 年)

2020 年，人工智能技术迈向新阶段，从“小模型+判别式”转向“大模型+生成式”，从传统的人脸识别、目标检测、文本分类，升级到文本生成、文生图、语音生成等，人工智能产业的发展深刻重塑全球秩序和发展格局，成为国家竞争的主战场，是彰显国家科技硬实力的关键环节。在此历史性时刻，以陈维良先生、彭莉女士、杨建先生为核心的创始团队凭借在 GPU 及集成电路设计行业近二十年的从业经验，精准把握行业发展趋势，于 2020 年 9 月创立沐曦有限，致力于打造世界一流的 GPU 芯片公司。

经过“从零到一”全自研的技术沉淀，沐曦逐渐在硬件、软件层分别积累了大量核心技术，形成了高性能 GPU 处理器架构设计技术、高性能 GPU 流处理器设计技术、高性能 GPU 多级缓存结构及内存管理设计技术、高性能 GPU 系统控制及虚拟化技术、MetaXLink 高速互连技术、芯粒架构与先进封装设计技术、高性能 GPU 动态功耗管理技术、高性能 GPU 验证方法学及智能验证技术、GPU 异构计算编程语言与开发环境、人工智能编译器与编译技术、GPU 高性能通信库设计与优化技术、GPU 高性能数学库设计与优化技术、GPU 核心驱动程序、GPU 软硬件协同优化技术等核心技术体系，打造了第一代 IP 核、超过 600 条指令集，并全面应用到公司产品中。

2、产品优势:产品矩阵逐渐丰富，高性能 GPU 产品进入量产阶段(2023-2024 年)

自 2023 年以来，沐曦逐渐将研发优势转换为产品优势，陆续量产了智算推理曦思 N 系列 GPU 产品、训推一体曦云 C 系列 GPU 产品，产品性能达到了行业先进水平，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具备较强的核心竞争力。

沐曦首款智算推理芯片曦思 N100 于 2022 年 1 月交付流片，并于 2023 年 4 月正式量产。曦思 N100 芯片的设计、流片到量产，为公司打通了从研发到生产再到销售的全业务流，完善了产品、生产供应链和销售渠道的资源储备，奠定了公司的经营模式和商业模式。2024 年 2 月，沐曦训推一体芯片曦云 C500 正式量产，陆续在多个国家人工智能公共算力平台和商业化智算中心批量应用。同时，基于对国际政策形势的研判，沐曦提前布局国产供应链，并基于国产供应链研发设计了新一代训推一体芯片曦云 C600，于 2024 年 10 月交付流片，能够确保生产经营的可持续性。至此，沐曦已量产的产品线覆盖了人工智能训练及推理、通用计算等计算场景。

这一阶段，沐曦在对技术进行产业化输出的同时持续强化研发投入，在图形渲染领域也取得了进一步突破，构建了涵盖推理、训练、渲染全套核心 GPU IP 框架，在全自研的统一计算和渲染的 GPU 架构下推出了第二代 IP 核、超过 800 条指令集，进而形成了完善的 GPU 产品矩阵。

3、市场领先优势:业绩实现跨越式增长，行业地位日益稳固(2025 年至今)

自 ChatGPT 发布以来，全球人工智能算力建设进入加速通道，下游爆发的算力需求叠加持续升级的地缘政治风险，推动沐曦 GPU 产品快速实现市场拓展，从而将产品优势转换为市场领先优势。凭借突出的产品性能和稳定的供应能力，截至 2025 年第一季度，沐曦 GPU 产品累计销量超过 25,000 颗。沐曦深度构建“1+6+X”生态与商业布局，基础算

力底座方面相继交付 9 大智算集群，算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心，区域横跨北京、上海、杭州、长沙、中国香港等地区，并逐渐向更多区域延伸。沐曦 GPU 产品深度赋能众多行业应用场景，已率先布局教科研、金融、交通、能源、医疗健康、大文娱等行业，产品赋能真实应用场景的竞争力和交付能力得到充分验证。2025 年，沐曦作为创始成员单位加入由中国计算机协会开源发展委员会(CCF)的 Allinfra 开源工作组，依托 CCF 链接科教资源、产业资源和社会资源等，形成产、学、研、用联动的开源创新模式，真正推动我国智能算力自主可控和良性发展，市场认可度和品牌影响力大幅提升。

图表 1 沐曦集成电路历史沿革

年份	公司该阶段主要进展	进展细节	产品技术/销售核心进展
2020-2022 年	自设立即专注 GPU 业务，打造全建制团队并形成相关技术的原始积累	2020 年，以陈维良先生、彭莉女士、杨建先生为核心的创始团队凭借在 GPU 及集成电路设计行业近二十年的从业经验，精准把握行业发展趋势，于 2020 年 9 月创立沐曦有限，致力于打造世界一流的 GPU 芯片公司。	经过“从零到一”全自研的技术沉淀，沐曦逐渐在硬件、软件层分别积累了大量核心技术。
2023-2024 年	产品矩阵逐渐丰富，高性能 GPU 产品进入量产阶段	沐曦首款智算推理芯片曦思 N100 于 2022 年 1 月交付流片，并于 2023 年 4 月正式量产。2024 年 2 月，沐曦训推一体芯片曦云 C500 正式量产，陆续在多个国家人工智能公共算力平台和商业化智算中心批量应用。同时，基于对国际政策形势的研判，沐曦提前布局国产供应链，并基于国产供应链研发设计了新一代训推一体芯片曦云 C600，于 2024 年 10 月交付流片，能够确保生产经营的可持续性。	沐曦逐渐将研发优势转换为产品优势，陆续量产了智算推理曦思 N 系列 GPU 产品、训推一体曦云 C 系列 GPU 产品，构建了涵盖推理、训练、渲染全套核心 GPU IP 框架，在全自研的统一计算和渲染的 GPU 架构下推出了第二代 IP 核、超过 800 条指令集，进而形成了完善的 GPU 产品矩阵。
2025 年至今	业绩实现跨越式增长，行业地位日益稳固	沐曦深度构建“1+6+X”生态与商业布局，基础算力底座方面相继交付 9 大智算集群，算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心，区域横跨北京、上海、杭州、长沙、中国香港等地区，并逐渐向更多区域延伸。沐曦 GPU 产品深度赋能众多行业应用场景，已率先布局教科研、金融、交通、能源、医疗健康、大文娱等行业，产品赋能真实应用场景的竞争力和交付能力得到充分验证。	凭借突出的产品性能和稳定的供应能力，截至 2025 年第一季度，沐曦 GPU 产品累计销量超过 25,000 颗。

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

1.2 沐曦公司核心产品包括智算推理的曦思 N 系列和用于训推一体和通用计算的曦云 C 系列

(1) 公司主要产品全面覆盖人工智能计算、通用计算和图形渲染三大领域。先后推出用于智算推理的曦思 N 系列 GPU、用于训推一体和通用计算的曦云 C 系列 GPU，以及正在研发用于图形渲染的曦彩 G 系列 GPU

公司现有产品采用自主研发的 GPU IP 和指令集、统一的 GPU 计算和渲染架构，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，并已实现多款高性能 GPU 产品的量产销售。

图表 2 沐曦集成电路公司核心产品

产品类型	型号	产品特点	应用场景
训推一体 GPU	曦云 C500 系列	公司曦云 C 系列的产品拥有多精度混合算力，内置大量运算核心，具有较强的并行计算能力和较高的能效比，适用于向量计算和矩阵计算等计算密集型应用，可广泛应用于智算训练和推理、通用计算、AI for science 等场景	云端智算(训推一体)、通用计算、AI for Science 等
	曦云 C600 系列		
智算推理 GPU	曦思 N100 系列	公司曦思 N100 产品系面向传统人工智能场景，内置性能强劲的视频处理器和运算核心，可广泛应用于智慧城市、智慧交通、智慧教育、智能视频处理等场景	云端及边端推理、视频转码
	曦思 N260 系列	公司曦思 N 系列后续迭代产品系面向生成式人工智能场景，拥有多精度混合算力、大容量显存和较高的能效比，可广泛应用于大模型推理、生成式应用等场景	
	曦思 N300 系列		
图形渲染 GPU	曦彩 G100 系列	公司曦彩 G 系列产品系面向图形处理场景，内置性能强大的图形处理器，可广泛应用于云游戏、数字孪生、云渲染、影视动画和专业制图等场景	云端及边端图形处理

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

(2) 沐曦公司于 2023 年推出了首款训推一体 GPU 芯片曦云 C500，并在此基础上推出了曦云 C550;该系列基于国产供应链的产品曦云 C600 已完成流片

沐曦公司曦云 C 系列产品针对云端计算场景，为人工智能训练和推理、通用计算提供算力底座，具备高性能、高自主可控、高扩展性。曦云 C 系列产品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。

曦云 C 系列 GPU 芯片的结构逻辑高度复杂、计算单元较多，内置运算单元、高速缓存、控制单元等，支持 PCIe、CXL 等协议和自主研发的 MetaXLink 协议，拥有丰富的接口，能够满足不同应用场景的需求。

沐曦已量产的训推一体 GPU 芯片具有强大的大规模并行计算能力。通过与新华三等服务器整机厂通力合作，沐曦的 GPU 产品已在多个智算中心中获得规模化应用。同时，随着大模型不断更新迭代，大规模集群算力需求快速增长，基于领先的 MetaXLink 互连技术，沐曦可在大规模集群实现优异的线性扩展，将在智算集群实际运营中发挥强大的性能。

图表 3 沐曦集成电路公司核心产品（训推一体 GPU——曦云 C 系列）

产品型号	产品概况	产品特性
曦云 C500	<ol style="list-style-type: none"> 1) 采用公司自研的 XCORE1.0 架构及指令集，面向云端人工智能训练与推理、通用计算、AlforScience 等计算任务 2) 拥有丰富的标量、矢量和张量计算单元，支持多种混合精度计算 3) 采用 HBM2e 显存，显存容量 64GB 4) 支持卡间高速互连 	<ol style="list-style-type: none"> 1) 基于 MetaXLink 互连技术，公司的单颗芯片产品拥有 7 个高速 MetaXLink 互连接口，可实现 2 卡、4 卡、8 卡、16 卡、32 卡、64 卡等多种互连拓扑，具备国内稀缺的高带宽、超多卡互连能力，高度支持千亿参数以上规模 AI 大模型训练算力集群建设
曦云 C550	<ol style="list-style-type: none"> 1) 采用公司自研的 XCORE1.0 架构及指令集，面向云端人工智能训练与推理、通用计算、AlforScience 等计算任务 2) 拥有丰富的标量、矢量和张量计算单元，算力密度更高，支持多种混合精度计算 3) 采用 HBM2e 显存，显存容量 64GB 4) 支持卡间高速互连 	<ol style="list-style-type: none"> 2) 基于自研的 MXMACA 软件栈，适配主流算法框架、运算库、通信库、操作系统编程语言、调试和运维管理工具等，其编程接口在 API 层面实现了对 GPU 行业国际主流 CUDA 生态的高度兼容
曦云 C600 (在研)	<ol style="list-style-type: none"> 1) 基于国产供应链，采用公司自研的 XCORE1.5 架构及指令集，增加了 FP8Tensor 及 Tensor 转置指令，面向云端人工智能训练与推理、通用计算、AlforScience 等计算任务 2) 拥有丰富的标量、矢量和张量计算单元，支持多种混合精度计算 3) 采用 HBM3e 显存 4) 支持卡间高速互连 	

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

沐曦先后实现智算推理 GPU 产品在传统 AI 任务和生成式 AI 任务中的落地。其中曦思 N100 产品是沐曦 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。曦思 N 系列的后续产品如曦思 N260、曦思 N300(在研)均主要面向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 I/O 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。

图表 4 沐曦集成电路公司核心产品（智算推理 GPU——曦思 N 系列）

产品型号	产品概况
曦思 N100	<ol style="list-style-type: none"> 1) 采用公司自研的 XCORE0.5 架构及指令集，面向云端和边端推理、视频转码等计算任务 2) 支持多种混合精度计算 3) 采用 HBM2e 显存，显存容量 16GB 4) 集成了丰富的视频图像编解码硬件单元，支持 H. 264/H. 265/AV1/AVS2 硬件解码 5) 典型应用场景为人脸识别、图像识别、视频编解码等传统人工智能场景
曦思 N260	<ol style="list-style-type: none"> 1) 采用公司自研的 XCORE1.0 架构及指令集，面向云端推理计算任务 2) 支持多种混合精度计算 3) 采用 HBM2e 显存，显存容量 64GB 4) 典型应用场景为搭载 DeepSeek 等主流大模型的一体机、液冷工作站等，支持一机二卡、四卡等不同配置
曦思 N300 (在研)	<ol style="list-style-type: none"> 1) 基于国产供应链，采用公司自研的 XCORE1.5 架构及指令集，面向云端推理计算任务 2) 支持多种混合精度计算 3) 采用 HBM3 显存 4) 典型应用场景为云端智算中心以及搭载 DeepSeek 等主流大模型的一体机、液冷工作站等

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

2022 年到 2025 年 Q1, 沐曦 GPU 板卡产品的产销率分别为 21.67%、39.00%、39.82% 和 132.02%, 主要原因系:(1)沐曦 N100 系列智算推理 GPU 芯片、C500 系列训推一体 GPU 芯片分别于 2023 年、2024 年实现量产, 处于前期业务快速拓展阶段, 面对较为旺盛的行业下游需求, 沐曦需提前进行下单生产;(2)受国际政治局势变动影响, 为保障未来销售需求, 沐曦进行战略性生产备货。2025 年 1-3 月, 沐曦 GPU 板卡产品产销率为 132.02%, 销售情况良好, 产销率较高。

图表 5 沐曦集成电路公司核心产品产量和销量

单位: 张

年份	产品类型	产量	销量	产销率
2025 年 1-3 月	训推一体 GPU 板卡	6,058	8,055	132.96%
	智算推理 GPU 板卡	1,066	1,350	126.64%
	合计	7,124	9,405	132.02%
2024 年	训推一体 GPU 板卡	32,563	13,606	41.78%
	智算推理 GPU 板卡	2,895	512	17.69%
	合计	35,458	14,118	39.82%
2023 年	训推一体 GPU 板卡	1,444	272	18.84%
	智算推理 GPU 板卡	4,489	2,042	45.49%
	合计	5,933	2,314	39.00%
2022 年	训推一体 GPU 板卡	-	-	-
	智算推理 GPU 板卡	240	52	21.67%
	合计	240	52	21.67%

注: 上述销量包含公司直接以板卡形态销售及集成为服务器销售的板卡产品数量。

资料来源: 沐曦集成电路(上海)股份有限公司招股说明书, 华安证券研究所

2 沐曦智算推理曦思 N 系列奠定公司发展基础, 训推一体曦云 C 系列接棒助力国产算力发展

2.1 沐曦公司产品线日趋丰富, 营收快速增长, 季节性特征逐渐减弱

(1) 2023 年 4 月曦思 N 系列量产和 2024 年 2 月训推一体曦云 C 系列量产带动公司营收大幅增长

2022 年营收构成方面, 主要来自智算推理曦思 N 系列首代产品: 智算推理 GPU 产品曦思 N 系列于 2022 年 9 月流片成功, 2022 年当年实现小批量销售, 2022 年销售收入金额较小, 营业收入为 42.64 万元。

2023 年营收构成方面, 主要增量来自智算推理曦思 N 系列量产产品推向市场: 2023 年 4 月, 曦思 N 系列产品实现正式量产, 同时公司逐渐开拓下游市场, 2023 年当年智算推理 GPU 产品销售收入及 IP 授权收入大幅增长。

2024 年营收构成方面, 主要增量来自训推一体 GPU 产品曦云 C 系列正式量产: 2024 年 2 月, 公司核心产品训推一体 GPU 产品曦云 C 系列正式量产, 公司加大力度开拓下游客户, 叠加人工智能与大模型、智算中心建设的蓬勃发展以及 GPU 国产替代加速, 公司 2024 年训推一体系列产品销量、销售收入均实现大幅增长。

2025 年 1-3 月, 公司主营业务收入延续了良好的增长趋势, 核心产品训推一体 GPU 销量、销售收入均大幅增长。

图表 6 沐曦集成公司 2022-2025Q1 营业收入主要拆分

单位：万元

产品类型		2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
		金额	占比	金额	占比	金额	占比	金额	占比
训推一体系列	GPU 板卡	31,255.24	97.55%	51,181.39	68.99%	1,546.81	30.09%	-	-
	GPU 服务器	104.03	0.32%	20,992.13	28.29%	-	-	-	-
智算推理系列	GPU 板卡	400.75	1.25%	349.08	0.47%	1,376.32	26.77%	42.64	100.00%
IP 授权		25.00	0.08%	100.00	0.13%	2,216.67	43.12%	-	-
其他		256.50	0.80%	1,568.28	2.11%	1.29	0.03%	-	-
合计		32,041.53	100.00%	74,190.87	100.00%	5,141.09	100.00%	42.64	100.00%

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

2022-2024 年度，公司第四季度收入分别为 42.64 万元、4,576.32 万元和 51,975.87 万元，占全年主营业务收入的比例分别为 100.00%、89.01%、70.06%，公司 2023 年和 2024 年第四季度收入占比较高，主要原因系公司核心产品研发成功、实现量产并大规模出货的时间点对销售收入的季节性影响较大。具体而言，沐曦智算推理系列产品于 2023 年 4 月实现正式量产、训推一体系列产品于 2024 年 2 月实现正式量产，产品上市后下游客户通常先小规模采购进行试用，再进行大规模采购。沐曦核心产品均于当年上半年量产，至下半年或第四季度与重要客户签署大额订单并实现销售，符合新产品上市后销量爬坡需要一定时间的客观规律。

2025 年一季度，公司实现收入 32,041.53 万元，相较 2024 年一季度大幅增加，未来随着公司产品矩阵逐渐丰富以及销量的增长，公司销售收入的季节性特征预计将有所减弱。

图表 7 沐曦集成公司 2022-2025Q1 营业收入按照季度拆分

单位：万元

项目	2025 年 1-3 月		2024 年度		2023 年度		2022 年度	
	金额	占比	金额	占比	金额	占比	金额	占比
第一季度	32,041.53	100.00%	458.16	0.62%	11.58	0.23%	-	-
第二季度	-	-	17,560.50	23.67%	407.48	7.93%	-	-
第三季度	-	-	4,196.35	5.66%	145.71	2.83%	-	-
第四季度	-	-	51,975.87	70.06%	4,576.32	89.01%	42.64	100.00%
合计	32,041.53	100.00%	74,190.87	100.00%	5,141.09	100.00%	42.64	100.00%

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

2022年、2023年、2024年和2025年Q1，归属于母公司所有者的净利润分别为-77,696.52万元、-87,115.82万元、-140,887.94万元和-23,251.22万元，尚未实现盈利。

沐曦公司截至2025年Q1尚未实现盈利，主要系：

(1)中国GPU芯片市场曾长期被国外巨头垄断，国产芯片渗透率低，面临技术标准适配及用户习惯迁移障碍，生态建设需突破，市场拓展呈渐进式发展；

(2)沐曦智算推理GPU芯片曦思N100系列、训推一体GPU芯片曦云C500系列分别于2023年4月和2024年2月正式量产，因公司产品进入重点行业客户需经历严苛的技术验证及生态适配周期，公司销售规模处于快速爬坡阶段，目前收入规模仍然难以覆盖成本费用支出；

(3)沐曦所处GPU芯片行业属资本与技术双密集型，有极高技术壁垒与密集研发投入特点，公司为夯实产品竞争优势持续加大研发投入；

(4)沐曦因实施股权激励而确认了较大金额的股份支付费用。

2.2 沐曦曦云C系列训推一体GPU板卡是公司核心营收来源，分别占比2023年、2024年和2025年Q1季度总收入的30.09%、68.99%和97.55%

2023年，沐曦推出曦云C系列训推一体GPU产品，该系列产品拥有多精度混合算力，内置大量运算核心，具有较强的并行计算能力和较高的能效比，适用于向量计算和矩阵计算等计算密集型应用，可广泛应用于智算训练与推理、通用计算、AlforScience等场景。2023年、2024年和2025年1-3月，公司训推一体GPU板卡分别实现销售收入1,546.81万元、51,181.39万元和31,255.24万元，占公司同期主营业务收入的比例分别为30.09%、68.99%和97.55%，销售收入呈大幅增长趋势，是公司收入增长的主要来源。

2023年、2024年和2025年1-3月，沐曦训推一体GPU板卡销量分别为272张、10,905张和8,041张，2023年销量较少，主要系训推一体系列产品于2023年6月首批芯片回片，当年尚未实现正式量产。2024年2月，训推一体GPU板卡正式量产后产品影响力逐渐提升，同时受益于人工智能、大模型以及智算中心建设需求的增长，销量及销售收入大幅增长。

2023年、2024年和2025年1-3月，沐曦训推一体GPU板卡平均单价(不含税)分别为5.69万元/张、4.69万元/张和3.89万元/张，价格内有所下降。主要原因为2023年训推一体GPU板卡客户单笔采购量较小，单价较高；随着2024年2月训推一体GPU板卡正式量产，2024年以及2025年1-3月销量均同比大幅上升，特定客户单笔采购量大规模增加，公司给予了相应的优惠措施。

图表8 沐曦集成电路(上海)股份有限公司2022-2025Q1销售数量和单价(训推一体GPU板卡)曦云C

项目	2025年1-3月	2024年度	2023年度	2022年度
销售收入(万元)	31,255.24	51,181.39	1,546.81	-
销售数量(张)	8,041	10,905	272	-
平均单价(万元/张,不含税)	3.89	4.69	5.69	-

注：上表中销售数量不包含装配在训推一体GPU服务器中实现销售的板卡数量。

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

2.3 沐曦训推一体 GPU 服务器占比公司 2024 年营业收入的 28.29%

(1) 训推一体 GPU 服务器是人工智能和高性能计算的核心硬件，是沐曦公司 2024 年重要营收，占比公司 2024 年营业收入的 28.29%

2024 年，公司与市场知名服务器厂商合作，委托服务器厂商将公司训推一体 GPU 板卡，以及其他品牌的 CPU、内存等硬件集成为 GPU 服务器整机，公司将 GPU 服务器整机交付给部分客户后最终应用于智算中心集群的建设。

公司训推一体 GPU 服务器主要于 2024 年形成销售并产生收入，2024 年和 2025 年 1-3 月，分别实现收入 20,992.13 万元和 104.03 万元，实现销量分别为 343 台、2 台，销售平均单价分别为 61.20 万元/台、52.01 万元/台。公司销售的训推一体 GPU 服务器通常每台包含数张训推一体 GPU 板卡，同时还包含 CPU、内存、硬盘等其他硬件，其销售价格受单台服务器包含的训推一体 GPU 板卡数量以及其他硬件配置的不同而有较大差异。2024 年公司交付的训推一体 GPU 服务器主要应用于智算中心集群，通常每台服务器包含 8 张训推一体 GPU 板卡。2025 年 1-3 月训推一体 GPU 服务器销售单价有所降低，主要因为当期训推一体 GPU 服务器销量较少，且均为根据客户需求定制化的产品，包含的板卡数量总体较少。

图表 9 沐曦集成公司 2022-2025Q1 销售收入和销售数量及单价（训推一体 GPU 服务器）

项目	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
销售收入（万元）	104.03	20,992.13	-	-
销售数量（台）	2	343	-	-
平均单价（万元/台，不含税）	52.01	61.20	-	-

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

2.4 沐曦曦思 N 系列面向传统人工智能是公司 2022 年推出的首款产品

(1) 智算推理 GPU 板卡主要面向传统人工智能场景，智算推理 GPU 曦思 N 系列产品是公司于 2022 年推出的首款产品

智算推理 GPU 板卡主要面向传统人工智能场景，可广泛应用于智慧城市、智慧交通、智慧教育、智能视频处理等场景。2022 年、2023 年、2024 年和 2025 年 Q1 各期，公司智算推理 GPU 板卡实现收入分别为 42.64 万元、1,376.32 万元、349.08 万元和 400.75 万元，占同期主营业务收入的比例分别为 100%、26.77%、0.47%和 1.25%，实现销量分别为 52 张、2,042 张、512 张和 1,350 张，均价分别为 0.82 万元/张、0.67 万元/张、0.68 万元/张和 0.30 万元/张。

智算推理 GPU 曦思 N 系列产品是公司于 2022 年推出的首款产品，定位为应用于人脸识别、图像识别、文本分类等传统人工智能推理场景的算力芯片。2023 年末至 2024 年，生成式 AI 大模型领域的井喷式增长促使市场对于传统人工智能推理芯片的需求下降，导致 2024 年公司曦思 N 系列产品的销售收入、销量有较大幅度下降。2025 年 1-3 月，公司根据市场需求调整其定价策略，该系列产品销量有所上升。

图表 10 沐曦集成公司 2022-2025Q1 销售收入和销售数量及单价（智算推理 GPU 板卡）曦思 N

项目	2025 年 1-3 月	2024 年度	2023 年度	2022 年度
销售收入（万元）	400.75	349.08	1,376.32	42.64
销售数量（张）	1,350	512	2,042	52
平均单价（万元/张，不含税）	0.30	0.68	0.67	0.82

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

2.5 沐曦公司借助经销模式扩大终端客户覆盖范围提升产品服务能力

2022 年、2023 年、2024 年和 2025 年 Q1，沐曦经销收入分别为 0、2,252.23 万元、27,758.94 万元和 17,849.70 万元，占公司各期主营业务收入的比例分别为 0、43.81%、37.42%和 55.71%。公司 GPU 产品具有功能通用性高、下游应用领域广泛等特点，现阶段综合考虑到自身品牌影响力、经营规模、销售人员数量及市场服务能力有限等因素，公司通过采用经销模式，可以充分发挥经销商区位优势、行业资源的优势，扩大终端客户覆盖范围并提高服务能力，是公司结合自身实际情况自主选择的销售策略，符合公司实际情况，具有必要性、商业合理性。

图表 11 沐曦集成公司 2022-2025Q1 销售收入及收入占比（直销 vs 经销）

单位：万元

销售模式	2025 年 1-3 月		2024 年		2023 年		2022 年	
	金额	占比	金额	占比	金额	占比	金额	占比
直销	14,191.82	44.29%	46,431.94	62.58%	2,888.86	56.19%	42.64	100.00%
经销	17,849.70	55.71%	27,758.94	37.42%	2,252.23	43.81%	-	-
合计	32,041.53	100.00%	74,190.87	100.00%	5,141.09	100.00%	42.64	100.00%

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

在经销商方面。根据超讯通信公司 2024 年年报披露显示，2023 年，超讯通信与国产 GPU 行业领先的设备厂商“沐曦股份”达成合作，成为其“沐曦”品牌 GPU 产品特定行业总代理。2024 年，超讯通信自有算力设备品牌“元醒”成功推出市场，目前该系列服务器能覆盖模型训练、推理、定制计算以及储存等应用场景。2024 年，超讯通信围绕算力板块持续布局，与上海国和投资就人工智能与硅光技术投资建立战略合作关系，双方将整合各自优势资源，共同打造科技赋能与产融结合的投资新模式。

2025 年初，超讯通信携手“沐曦股份”等合作伙伴共同成立合资公司“讯曦智能”，向算力设备产业链上游底层基础技术迈进，承接芯片封测、服务器整机生产、销售和维修服务，以围绕产业链上下游获取更多技术和市场渠道，进一步巩固超讯通信的核心竞争力和品牌影响力。

在销售合同方面，截至 2025 年 3 月 31 日，沐曦及其子公司履行完毕和正在履行的合同金额在 7,500 万元以上的销售合同以及无具体合同金额但对公司经营活动、财务状况、或未来发展等具有重要影响的框架销售合同中包括与超讯通信最早于 2023 年签订的板卡销售订单。2025 年 Q1 超讯通信是公司第一大客户销售金额为 1.28 亿元占比公司营业收入的 39.95%。

图表 12 沐曦集成公司 2022-2025Q1 前五大客户销售金额和占比

图表 13 沐曦集成公司截至 2025 年 3 月 31 日履行完毕和正在履行的合同金额在 7,500 万元以上的销售合同

单位：万元

年份	序号	客户名称	销售金额	占主营业务收入比例
2025 年 1-3 月	1	超讯通信股份有限公司	12,800.17	39.95%
	2	上海源庐加佳信息科技有限公司	9,095.93	28.39%
	3	新华三信息技术有限公司	2,918.69	9.11%
	4	北京创世云科技股份有限公司	2,359.28	7.36%
	5	华科智谷(上海)科技发展有限公司	1,136.14	3.55%
		合计	28,310.22	88.35%
2024 年	1	新华三信息技术有限公司	15,095.15	20.35%
	2	苏州瑞芯智能科技有限公司	13,046.02	17.58%
	3	香港算丰信息有限公司	10,733.34	14.47%
	4	汇天网络科技有限公司	7,332.18	9.88%
	5	浙江武路智慧城市技术有限公司	6,532.66	8.81%
		合计	52,739.35	71.09%
2023 年	1	公司 1	2,216.67	43.12%
	2	北京创世云科技股份有限公司	1,635.26	31.81%
	3	上海闪马智能科技有限公司	600.00	11.67%
	4	中科远见(重庆)科技有限公司	139.51	2.71%
	5	湖南沐仕杰半导体科技有限公司	116.58	2.27%
		合计	4,708.02	91.58%
2022 年	1	上海东焱信息技术有限公司	42.64	100.00%
		合计	42.64	100.00%

序号	客户名称	合同金额(万元)	销售内容	签订年度	履行情况
1	芯原微电子(上海)股份有限公司	10,000.00	IP 授权	2023	正在履行
2	超讯通信股份有限公司	根据订单	板卡	2023	正在履行
3	新华三信息技术有限公司	根据订单	根据订单	2023	正在履行
4	上海算丰信息有限公司	10,592.00	服务器	2023	履行完毕
5	浙江武路智慧城市技术有限公司	根据订单	板卡	2024	正在履行
6	公司 3	10,656.00	服务器	2024	正在履行
7	汇天网络科技有限公司	10,870.00	服务器	2024	正在履行
8	苏州瑞芯智能科技有限公司	根据订单	板卡	2024	正在履行
9	华科智谷(上海)科技发展有限公司	根据订单	板卡	2025	正在履行
10	汇天网络科技有限公司	21,165.16	服务器	2025	正在履行

注：上述客户均为合并口径。北京创世云科技股份有限公司包括该公司及其子公司上海燃云科技有限公司；香港算丰信息有限公司包括该公司及其子公司上海算丰信息有限公司。

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

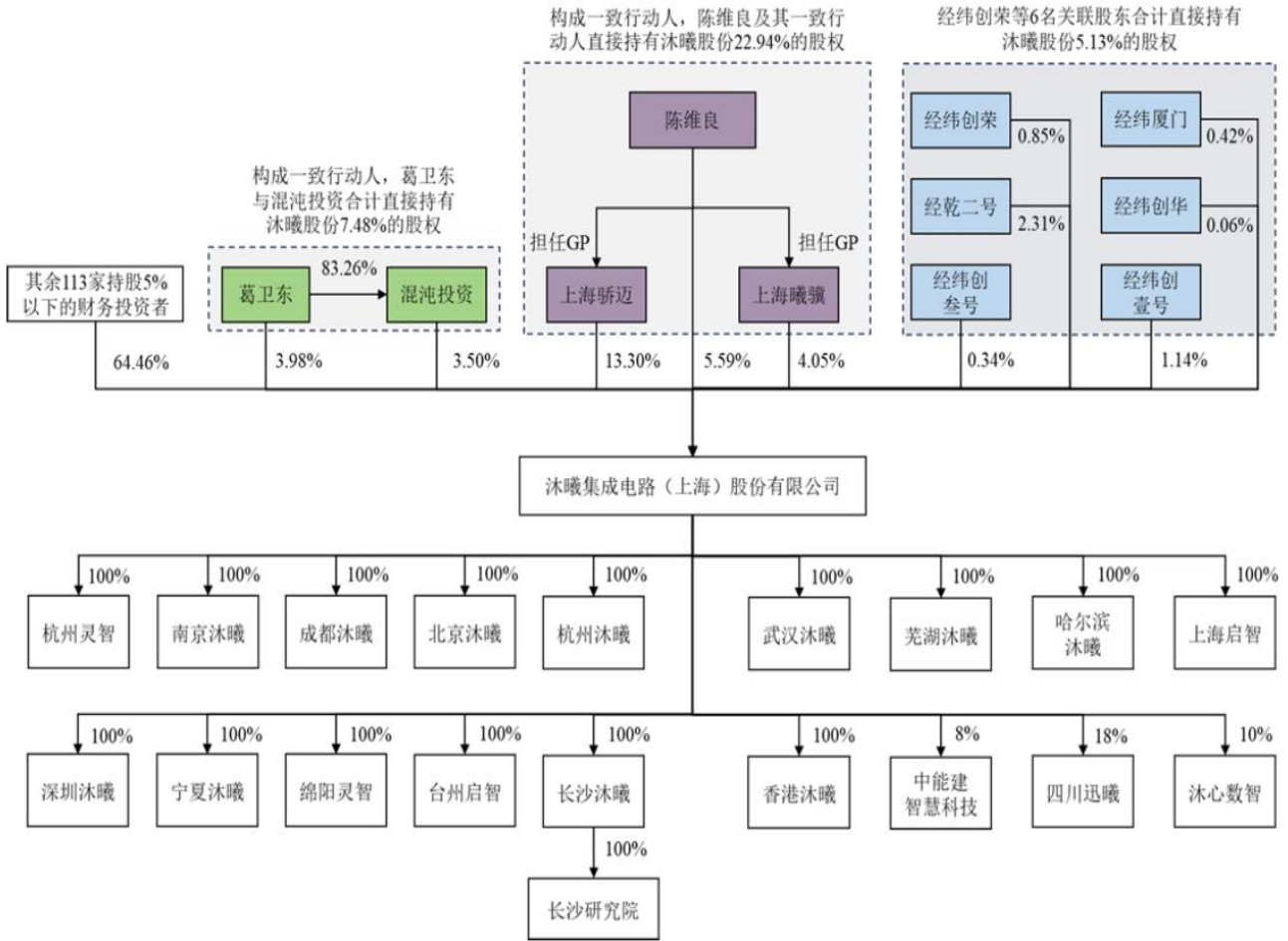
资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

2.6 沐曦汇聚诸多知名投资人和国资及市场化机构

根据中国基金报的披露。在申请上市前，沐曦股份共进行了八轮融资，诸多知名机构参与投资。国资方面，上海科创基金、引领区基金、国调基金等纷纷入局；市场化机构方面，红杉中国、经纬中国、葛卫东的混沌投资、和利资本、东方富海、光速创投、复星锐正、真格基金等知名机构也参与其中。

从公司股权结构来看，陈维良是公司实际控制人，陈维良和陈维良实控的上海骄迈、上海曦骥，合计持股 22.94%。葛卫东的混沌投资在 2022 年 7 月的 Pre-B 轮首次投资，目前葛卫东与其控制的混沌投资合计持股 7.48%。经纬中国通过几个实体合计持股 5.13%，红杉中国旗下基金持有 4.19%，东方富海旗下基金持有 2.82%，国调基金和相关合伙企业持有 2.52%，中金旗下基金持有 1.85%。

图表 14 沐曦集成电路公司股权结构



资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

3 沐曦加强国产供应链合作，持续研发投入夯实竞争优势

3.1 沐曦公司与供应链伙伴持续合作，加强国产供应链配置

根据沐曦招股说明书披露，公司核心训推一体 GPU 曦云 C 系列最新一代在研产品为曦云 C600。该产品基于国产供应链，采用公司自研的 XCORE1.5 架构及指令集，增加了 FP8Tensor 及 Tensor 转置指令，面向云端人工智能训练与推理、通用计算、AlforScience 等计算任务。曦云 C600 采用 HBM3e 显存，支持卡间高速互连。

智算推理 GPU 曦思 N 系列最新一代在研产品为曦思 N300。该产品基于国产供应链，采用公司自研的 XCORE1.5 架构及指令集，面向云端推理计算任务。曦思 N300 采用 HBM3 显存，典型应用场景为云端智算中心以及搭载 DeepSeek 等主流大模型的一体机、液冷工作站等。

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 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 板卡的组装由加工厂商完成，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。

图表 15 沐曦集成公司主要采购物料金额和占比（晶圆和 HBM 内存在成本占比中较高）

单位：万元

类别	2025 年 1-3 月		2024 年		2023 年		2022 年度	
	金额	占比	金额	占比	金额	占比	金额	占比
晶圆	1,476.65	7.35%	40,637.47	37.15%	8,321.68	44.89%	260.33	14.40%
HBM 内存	10,088.98	50.21%	33,750.64	30.85%	6,361.94	34.32%	800.67	44.28%
外购硬件及配件	5,004.25	24.91%	19,908.20	18.20%	1,351.50	7.29%	-	-
委托加工	3,033.08	15.10%	12,848.13	11.75%	1,726.55	9.31%	496.59	27.46%
电子元器件	489.34	2.44%	2,241.28	2.05%	777.47	4.19%	250.65	13.86%
合计	20,092.30	100.00%	109,385.72	100.00%	18,539.14	100.00%	1,808.23	100.00%

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

2023 年、2024 年 2025Q1，沐曦集成公司前五大供应商采购金额占当期采购总额的比例为 95.01%、91.30%和 93.87%，前五大供应商采购占比相对稳定。其中，公司 D、公司 E、公司 C 为芯片供应商，公司向其采购晶圆及委托加工服务；公司 A 为 HBM 内存等产品知名供应商，公司向其采购 HBM 内存，因此沐曦对上述供应商采购金额相对较高，不属于对单一供应商的依赖情况。

图表 16 沐曦集成公司前五大供应商采购情况和占比

单位：万元

年份	序号	供应商名称	采购金额	占采购总额比例	采购内容
2025 年 1-3 月	1	公司 A	9,207.61	45.83%	HBM 内存
	2	万司信息技术(上海)股份有限公司	4,998.67	24.88%	硬件及配件
	3	公司 B	1,760.55	8.76%	委托加工
	4	公司 C	1,531.85	7.62%	晶圆、委托加工
	5	环鸿电子(昆山)有限公司	1,362.29	6.78%	委托加工、电子元器件
			合计	18,860.97	93.87%
2024 年	1	公司 D	39,804.75	36.39%	晶圆、委托加工
	2	公司 A	32,680.18	29.88%	HBM 内存
	3	万司信息技术(上海)股份有限公司	14,255.76	13.03%	硬件及配件
	4	环鸿电子(昆山)有限公司	7,205.43	6.59%	委托加工、电子元器件
	5	公司 B	5,922.59	5.41%	委托加工
			合计	99,868.70	91.30%
2023 年	1	公司 A	6,361.94	34.32%	HBM 内存
	2	公司 D	5,668.86	30.58%	晶圆、委托加工
	3	公司 E	3,650.43	19.69%	晶圆、委托加工
	4	万司信息技术(上海)股份有限公司	1,351.50	7.29%	硬件及配件
	5	环鸿电子(昆山)有限公司	581.49	3.14%	委托加工、电子元器件

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

沐曦公司采用 Fabless 经营模式，主要负责芯片的研发、设计与销售，生产环节由专业的外协厂商完成，主要包括晶圆代工厂和封装测试厂，沐曦公司与部分境外供应商进行合作；此外，公司在重要物料 HBM 以及芯片研发设计所需的 EDA 工具和部分接口 IP 等方面的采购，也涉及主要终端供应商来自于境外的情况。在中美科技博弈持续升级的背景下，因美国政府相关政策影响，沐曦公司目前在先进制程晶圆代工和 HBM 供应等方面受到不利限制。随着沐曦最新一代的产品导入国产供应链设计，未来或将进一步提升国产供应链在国产 GPU 中的使用占比。

3.2 沐曦公司重视研发投入，夯实竞争优势，核心创始团队产业经验积累丰富

核心技术人员共计 4 名，均系对集成电路行业拥有深刻理解并具备多年技术专业经验，**彭莉**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；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，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，包含操作系统内核态驱动、用户态驱动、编译器、数学库、AI 应用框架以及大模型训练和推理的框架；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，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，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。

公司核心创始人实际控制人陈维良，1999 年毕业于电子科技大学微电子科学与工程专业，2002 年考取清华大学微电子学研究所硕士学位。2002 年 7 月至 2003 年 9 月，任泰鼎多媒体技术(上海)有限公司高级工程师；2003 年 10 月至 2006 年 3 月，任远弘科技(上海)有限公司 GPU 设计经理；2006 年 3 月至 2006 年 12 月，任亚鼎视频科技(上海)有限公司 GPU 设计经理；2007 年 1 月至 2020 年 8 月，任超威半导体(上海)有限公司高级总监；2020 年 9 月至今于公司任职，目前担任沐曦董事长、总经理。

图表 17 沐曦集成公司主要创始人和核心技术人员

核心技术/公司创始团队	学历背景	取得的专业资质、重要科研成果/主要经历	对公司研发活动的贡献/主要经历	目前职务
陈维良	本科电子科技大学微电子科学与工程专业，硕士清华大学微电子学研究所学位。	2002 年 7 月至 2003 年 9 月，任泰鼎多媒体技术(上海)有限公司高级工程师；2003 年 10 月至 2006 年 3 月，任远弘科技(上海)有限公司 GPU 设计经理；2006 年 3 月至 2006 年 12 月，任亚鼎视频科技(上海)有限公司 GPU 设计经理；2007 年 1 月至 2020 年 8 月，任超威半导体(上海)有限公司高级总监		2020 年 9 月至今于公司任职，目前担任沐曦董事长、总经理。
彭莉	硕士，上海交通大学电路与系统专业。	从业经历与科研成果：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。 2007 年 7 月至 2020 年 9 月，任超威半导体(上海)有限公司企业院士。	(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。	2020 年 9 月至今于公司任职，目前担任沐曦董事、副总经理、首席技术官。
杨建	博士，浙江大学计算机科学与	从业经历与科研成果：拥有 20 年以上高性能 GPU 软硬	(1) 作为软件技术负责人，领导团队进行软件研发工作，推	2020 年 9 月至今于公司任

技术专业

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。

2007 年 1 月至 2019 年 11 月，任超威半导体(上海)有限公司企业院士;2019 年 11 月至 2020 年 3 月，任上海华为技术有限公司架构师;2020 年 4 月至 2020 年 9 月，任海思技术有限公司架构师。

从业经历与科研成果: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。

2009 年 7 月至 2013 年 5 月，任深圳市中兴微电子技术有限公司工程师;2013 年 5 月至 2021 年 1 月，任深圳市海思半导体有限公司研发经理。

动研发与设计 MXMACA 软件栈，包含操作系统内核态驱动、用户态驱动、编译器、数学库、AI 应用框架以及大模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。

职，目前担任沐曦董事、副总经理、首席技术官。

王爽

硕士，南京理工大学信号与信息处理专业

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。

2021 年 1 月至今于公司任职，目前担任沐曦董事、资深研发总监。

吴志华

硕士，浙江大学数学专业

从业经历与科研成果: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。

2020 年 11 月至今于公司任职，目前担任沐曦软件研发部副总裁。

资料来源:沐曦集成电路(上海)股份有限公司招股说明书,华安证券研究所

3.3 沐曦公司持续积累核心技术并承担国家重大专项

沐曦致力于推动“芯片—系统—算法—应用”的全栈技术闭环，通过研发 GPU 硬件技术从而为人工智能产业发展提供算力基座;系统软件介于芯片与应用程序之间，沐曦通过研发系统软件，向下实现对 GPU 芯片的管理、调用、优化及控制，向上为应用程序的开发与运行提供全面支撑。沐曦的硬件和软件技术有效支撑开发者构建多元化的人工智能算法和人工智能应用程序，推动人工智能在各场景快速落地。

GPU 产品的研发需要全面掌握集成电路设计与系统软件开发的大量关键技术，同时需要对各类人工智能算法和应用场景有着深入的研究和理解，是一项技术难度大、涉及方向广，且极为复杂的系统性工程。沐曦长期聚焦关键核心技术的攻关，不断强化科技创新能力，坚持核心技术自主可控的发展路线，通过自主研发形成了具备较强行业竞争力的核心技术体系。在集成电路设计领域，沐曦已全面掌握多项关键技术，涵盖了高性能 GPU 处理器架构设计技术、高性能 GPU 流处理器设计技术、高性能 GPU 多级缓存结构及内存管理设计技术、高性能 GPU 系统控制及虚拟化技术、MetaXLink 高速互连技术、芯粒架构与先进封装设计技术、高性能 GPU 动态功耗管理技术、高性能 GPU 验证方法学及智能验证技术等;系统软件方面，沐曦已系统掌握 GPU 异构计算编程语言与开发环境、人工智能编译器与编译技术、GPU 高性能通信库设计与优化技术、GPU 高性能数学库设计与优化技术、GPU 核心驱动程序、GPU 软硬件协同优化技术等核心技术。相关技术在沐曦各系列产品

的研发与应用中发挥了关键作用，显著提升了产品的性能、效率和竞争力。

图表 18 沐曦集成公司核心技术（高性能硬件设计技术）（公司核心技术均自主研发并实现国内领先）

核心技术名称	核心技术概况	核心技术先进性能表征	应用相关核心技术的产品或服务
高性能 GPU 处理器架构设计技术	公司 GPU 芯片采用高性能 SoC 架构设计，代表了行业内复杂异构系统设计的先进技术，主要包括接口处理模块、大容量高带宽显存控制器、多媒体处理单元等。基于一体化的 SoC 集成环境和流程，公司 GPU 产品实现从 IP 到 SoC，最后到芯片组的无缝集成。	在接口处理方面，公司接口处理模块可处理各种高速接口，为公司实现丰富的产品形态、特定场景及不同规模的大模型互连应用奠定了坚实的基础。在显存控制方面，公司显存控制器支持各种大容量高带宽 DRAM，第二代曦云系列 GPU 已引入了 HBM3/HBM3e 存储技术。在多媒体处理方面，公司多媒体处理单元支持各种音频编解码、视频编解码格式。	曦云 C 系列、曦思 N 系列、曦彩 G 系列
高性能 GPU 流处理器设计技术	公司迄今已自主研发了两个大代际、五个小代际流处理器架构，基于 SIMT(单线指令多线程结构)以及 MXMACA 指令集，定义了 GPU 可以执行的基本操作，如算术运算、逻辑运算、内存访问和分支跳转等，具有丰厚的技术积累。基于流处理器的自研 GPU 架构，公司设计了全栈 GPU 芯片产品，全面覆盖大规模人工智能训练、高并发实时人工智能推理、数据挖掘和数据分析等通用计算，以及图形渲染等领域，兼具高性能和通用性。公司的 MXMACA 指令集不仅满足了当前复杂多样的计算需求，还为未来的技术发展和应用场景提供了坚实的基础。	在指令集的通用性方面，公司自主研发了超过 600 条 (XCORE1.0 计算 GPU)、800 条 (XCORE2.0 渲染 GPU) 指令集，具有强大的通用性和灵活性。在计算单元的精度和丰富度方面，公司矩阵运算单元支持 FP64、FP32、TF32、FP16、BF16、BF8、FP8、FP6、FP4、INT8 等多种精度的加速，适用于各种人工智能及高性能计算场景。在访存优化方面，通过动态调整存储大小来灵活适配计算需求，降低了 GPU 访问带宽需求、延迟和功耗。在指令执行方面，采用标量、矢量和张量混合的超标量指令发射流水线技术、内部低延迟的取指操作，以及零开销的线程束切换机制，确保了指令执行的高效性和灵活性。在软硬件融合设计方面，创新地设计了软件双发射机制，在指令编译阶段识别出可以双发射的指令，设置提醒位以指导硬件进行双发射，大幅简化了硬件设计。	曦云 C 系列、曦思 N 系列、曦彩 G 系列
高性能 GPU 多级缓存结构及内存管理设计技术	公司高速缓存系统凭借独特的缓存组网技术，构建起多维度的缓存网络架构，基于多级高速缓存和高效率、高吞吐的数据访问网络，可以针对不同场景采用灵活的多级缓存控制策略，实现对高速缓存的精细管控，达成数据访问效率的最大化。公司设计的内存管理单元支持虚拟地址到物理地址的转换、内存保护、缓存控制、设备虚拟化等功能并支持 GPU 虚拟地址空间和 CPU 虚拟地址空间相融合或隔离的不同工作方式。	在缓存技术方面，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，有效提升缓存命中率。公司开发了针对 GPU 场景优化的缓存一致性协议，以及配套的软件和硬件缓存一致性机制，能够在多卡协同工作中提供高性能。同时，公司基于定制的交流网络来实现高效率高带宽的缓存访问，促使海量数据均衡、高效流通，大幅提高 GPU 整体算力利用率。在内存管理技术方面，通过多级大页表机制、虚拟化技术、共享虚拟内存技术等，为 GPU 计算核心并发访存提供了低延迟、低功耗的地址转换服务，能够满足各种多卡互连场景下的内存空间需求。	曦云 C 系列、曦思 N 系列、曦彩 G 系列
高性能 GPU 系统控制及虚拟化技术	公司在系统控制中的核心技术涵盖任务调度与资源管理、可靠性/可用性/可维护性管理，为整个系统能够长期高效、稳定、安全运行提	在任务调度及资源管理效率方面，公司运用创新性资源调度算法，实现了任务派发和任务并行处理的高效性，提升了 GPU 整体运行效率。在稳定性方面，公司通过可靠性、可	曦云 C 系列、曦思 N 系列、曦彩 G 系列

供了保障。

虚拟化技术可将单 GPU 分割为多个完全独立的虚拟 GPU，使得 GPU 在多租户的场景中既能满足用户的性能要求又能实现租户间的安全隔离。

用性及可维护性方面的系统协同设计，有效提高了 GPU 在集群训练中的稳定性，可以长时间无故障支持训练任务。在虚拟化方面，可将单个大算力 GPU 分割为 1/2/4/8 个独立用户同时使用，实现用户间的安全隔离，同时大幅降低虚拟化调度带来的性能损失。

**MetaXLink
高速互连
技术**

公司自主研发的 MetaXLink 高速互连技术采用创新的架构设计，支持多种 GPU 互连拓扑，供开发者自由选择以适应自身场景需求。

MetaXLink 高速互连技术突破了传统 PCIe 总线在带宽和延迟方面的限制，为大规模 GPU 集群的高效互连提供了关键支撑，在性能、灵活性和可扩展性方面均处于行业领先地位。

在互连带宽方面，公司单 GPU 芯片创新性集成 7 个 MetaXLink 接口，达到了英伟达 4nm 制程工艺下旗舰产品(H200)相当的互连带宽性能，处于国内领先水平，为大规模 AI 模型训练提供了充足的通信带宽保证。MetaXLink 可以灵活适配大模型发展及其不同切分方式下对于互连延时和带宽的特定需求，满足训练和推理等不同应用场景。在互连拓扑结构方面，MetaXLink 支持 Full-MeshHybridCubeMesh 等多种复杂互连拓扑结构，并且通过协议层设计创新，MetaXLink 支持互连拓扑重构，能够根据应用需求灵活调整互连拓扑，最大化接口带宽的利用效率。目前公司可实现 2-64 卡等多种互连拓扑及超节点架构，满足从中小型训练到超大规模集群的差异化需求，为构建高密度算力和大规模集群、处理更为复杂的人工智能任务提供了关键技术支撑。

曦云 C 系列、曦思 N 系列、曦彩 G 系列

芯粒架构与先进封装设计技术

公司掌握了行业内先进的 MCM、2.5DCoWoS-S 封装设计能力。公司芯粒架构有较强的通用性和可重构性，根据市场需求，公司的产品可以采用单芯粒、多芯粒同构或异构封装等多种封装形式实现。

公司可以根据需求灵活地将单颗或者多颗 SoC 芯片与高带宽存储芯片进行封装，实现不同尺寸和多种功耗规模的封装集成。公司已完成或正在进行国内外多家知名封装厂先进封装的导入和量产验证，并引入对 2.5DCoWoS-L 和 SiliconBridge 等先进封装形态的应用扩展。

曦云 C 系列、曦思 N 系列、曦彩 G 系列

高性能 GPU 动态功耗管理技术

公司在 GPU 动态功耗管理方面创造性地采取 GPU 微架构的高主频实现、工艺偏差自适应机制、供电网络抖动自适应技术以及运行时电压与频率的实时调节技术等，有效提高了 GPU 在复杂多变的应用场景下的能效比。

基于微架构设计与物理实现协同优化的方法，实现了 GPU 主频的稳步提升、微架构性能与功耗的平衡。基于工艺偏差自适应机制，公司 GPU 可根据自身工艺特性和运行时的温度信息，用更低的电压实现目标频率，进而达成更高性能功耗比。基于供电网络抖动自适应技术及运行时频率响应电压瞬态变化的主动调节技术，公司 GPU 产品能够在负载突变的瞬间主动调节 GPU 运行频率。

曦云 C 系列、曦思 N 系列、曦彩 G 系列

高性能 GPU 验证方法学及智能验证技术

公司拥有成熟先进、多层次的 GPU 验证方法学及智能验证技术，有效保障芯片产品的流片成功率和高质量交付，涵盖模拟仿真验证、形式验证、覆盖率驱动验证、脚本语言等。

公司自主研发了自主可控、多线程、高并发、高性能设计验证核心平台，可以实现高度的设计验证自动化、关键设计验证信息的归一化。公司验证 IP(VIP)及测试平台可以自动化产生 70-85%的代码，大大提高了效率，降低可开发周期和调试复杂度。公司自研并部署大型数据库，可以实现前后端全栈数据的存储、可视化分析和决策，进一步提高研发效率、降低成本。

曦云 C 系列、曦思 N 系列、曦彩 G 系列

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

图表 19 沐曦集成公司核心技术（高性能软件设计技术）（基于 MXMACA 软件栈，应用于全系列 GPU 产品）

核心技术名称	核心技术概况	核心技术先进性表征
GPU 异构计算编程语言与开发环境	MXMACA 是由公司推出的一种采用通用并行计算架构解决复杂计算问题的运算平台。它基于自研指令集以及 GPU 并行计算引擎，集成了通用计算和机器学习框架，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。此外，MXMACA 架构的编程接口在 API 层面实现了对 GPU 行业国际主流 CUDA 生态的高度兼容，为人工智能、科学计算和大数据处理等领域的现有应用提供了高效迁移至 MXMACA 运算平台的能力。	在生态兼容性方面，提供了一种简单易用的编程语言，高度兼容 CUDA 生态及业内 6,000 多个 GPU 计算的生态应用，改善了高性能异构计算场景中迁移困难、迁移成本高的痛点，兼容性、易用性、迁移效率、性能领先同行业。在计算资源的管理方面，MXMACA 拥有完善、丰富的工具链，平台内置的智能资源管理引擎和协同计算框架实现了对云边缘异构硬件资源的统一管理、动态调度和高效协同，支持大模型训练任务的自动扩缩容，大幅提高计算资源利用率，降低开发复杂度。在国产平台的支持度方面，兼容大多数国产操作系统、交换机、CPU 架构和服务器等，并支持在此基础上进行性能优化。
人工智能编译器与编译技术	公司研发了可以将深度学习框架(包括 PyTorch/TensorFlow/JAX/PaddlePaddle 等)模型代码转换为 GPU 底层硬件指令的 AI 编译器。更进一步，公司研发了基于开源大模型融合内部 GPU 体系结构知识库的编译优化技术,使用开源大模型优化 CUDA 算子的生成及编译过程。	通过算子自动调优技术、算子融合、动态内存分配、指令调度、代码结构优化等编译优化技术，公司自研的 AI 编译器可以促使各类 AI 框架下的模型都能够高效地部署运行在公司 GPU 上，充分释放硬件性能。公司的编译优化技术支持前沿的体系结构设计，能够优化计算和内存访问的调度，充分发挥硬件各流水线峰值性能，实测算子性能的提升可超过 10 倍。
GPU 高性能通信库设计与优化技术	GPU 高性能通信库设计与优化技术是构建计算芯片多机多卡互连通信的核心技术，是保障智算集群卡间、机间协作运行的基本组件。公司的 GPU 高性能通信库设计与优化技术基于行业内先进的通信技术基础，结合自身硬件特征，设计了多种高性能通信机制，形成了通用的高性能通信库。	该技术实现了公司产品对各类 Scaleup 和 Scaleout 拓扑架构的兼容与自适应性，减少了集合运算过程中的卡间同步次数和通信时延，提高了混用高低带宽链路拓扑的通信效率，降低了网卡负载不均现象、提升了集群容错能力。同时，该技术实现了异构集群互连互通，可以增加客户现有算力资源的利用率。基于公司 GPU 产品的千卡集群执行 128B 的 MoE 大模型全量预训练的性能线性度在 95%左右，无故障时间超 14 天(训练任务按需结束)，处于行业领先水平。
GPU 高性能数学库设计与优化技术	开发者在编写程序时能够以调用高性能数学库的形式实现常用的数学运算，从而快速实现预期的功能并获得较好的性能。公司针对开发者在 AI 大模型、高性能计算、图形图像处理和向量数据库领域，分析和设计了一系列高性能数学库，使开发者快速、灵活、友好利用硬件能力。	公司提供高效的标量、向量和矩阵计算接口、卷积计算接口，访存高效的 attention 类算子库、pytorch 优化算子库以及与业务场景相关的算子库，充分释放各产品线硬件潜力。公司开发了离线优化算子库的工具链，针对客户特定模型对算子库的需求，工具链可自动生成算子优化问题集，离线优化工具链自动化、全流程搜索对特定问题集的更优 GPUkernel 选择和配置。
GPU 核心驱动程序	公司自主研发的核心 GPU 驱动程序(包括 OS/操作系统层驱动、用户态驱动等)，作为全系列产品的关键底层支撑,集成了先进的虚拟化管理机制、内存管理机制、基于计算图描述的智能任务调度系统、实时状态监控与性能分析工具、高效数据通信接口以及多芯片协同管理、GPU 芯片的高可靠性服务等功能模块。	在硬件兼容性方面，公司的驱动程序采用创新的异构计算架构，可实现 GPU 资源的高效分配与动态优化，确保芯片在各类操作系统环境中稳定运行。在硬件资源管理方面，该技术通过精细化的内存管理和智能任务调度算法，显著提升了 GPU 的资源利用率和任务处理效率；同时，公司自主研发的虚拟化软件能够将 GPU 虚拟化为可配置数量、可选规模且方法灵活的虚拟 GPU 设备，确保了资源分配的高效性和灵活性，能够简化智算中心 IT 资产管理、提升智算中心场景下 GPU 资源利用率。
GPU 软硬件协同优化技术	公司自主研发的软硬件协同优化技术,通过创新的架构设计和算法优化,有效提升 GPU 资源利用效率,为全系列产品在人工智能深度学习、高性能科学计算和大规模数据处理等关键应用领域提供了强有力的支持。GPU 软硬件协同优化技术包括不同硬件特性抽象层的编程接口和性能评估工具链等。	在软件抽象层与硬件架构特性的匹配性方面，公司使用 C/C++或其他高级语言对硬件资源进行抽象，使得开发者能够直接调用硬件特性，根据不同的工作负载选择更优的优化算法，并且可使用高级语言并发调度多种类型的指令，从而获得 GPUkernel 的更优性能。在性能评估工具链方面，公司性能评估工具功能丰富，使得开发者能够详细分析 GPU 目标计算任务的性能瓶颈，实现最优化的 GPUkernel 实现。

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

公司重视研发并承担多项重大科研项目，在部分课题中作为牵头单位助力国产自主可控。项目 A 课题中，由公司进行牵头，工信部主管，旨在加快推动服务器产业链高质量发展，完成重点任务攻关工作，助力智能服务器标志性产品自主可控；《基于 MetaXLink 的 GPU 互联核心技术研发及产业化项目》课题方面，公司独立承担，是 2024 年上海促进产业高质量发展专项先导产业创新发展(人工智能专题)项目，项目建设内容主要包括 GPU 互连拓扑的 MetaXLink 架构设计核心技术攻关、芯片制程和封装工艺优化以及算力互连，实现芯片架构、制程、封装、显存、计算单元设计和互连、生态建设等功能。

图表 20 沐曦集成公司承担的重大科研项目

课题名称	主要研究内容	主管单位	承担角色	项目进展
项目 A	本课题旨在加快推动服务器产业链高质量发展，完成重点任务攻关工作，助力智能服务器标志性产品自主可控	工信部	牵头单位	研发阶段
《多模态边缘转发与算力增强》	本课题聚焦于研发支持算力增强的多模态边缘网络路由设备与控制系统，包括三项研究内容，即多模态智慧协同控制系统的研制、多模态边缘网络路由转发设备的研制和应用于多模态边缘网络路由转发设备的多元异构算力增强模块的研制。	科技部	参与单位	验收阶段
《分布式加速服务器》	研制基于以太网的全国产分布式 GPU 芯片、加速卡和单机 32 卡的高密度 GPU 池化服务器软硬件平台，在算力中心完成验证推广。	工信部	参与单位	验收阶段
《自主高性能 GPU 及高密度视觉智能服务器研发》	本课题紧扣“自主高性能 GPU 及高密度视觉智能服务器研发”重点专项指南，研发自主高算力高效比 GPU 芯片及配套生态软件栈。研究内容涵盖“基于异构多核的高性能 GPU 架构”“低功耗高算力芯片电路设计技术”和“基于层次化架构的高通用生态软件栈”等三个课题。	科技部	参与单位	验收阶段
项目 B	本课题研发形成行业大数据分析核心模型及基于大数据分析预测中间件。	科技部	参与单位	研发阶段
《基于 MetaXLink 的 GPU 互联核心技术研发及产业化项目》	本课题是 2024 年上海促进产业高质量发展专项先导产业创新发展(人工智能专题)项目，项目建设内容主要包括 GPU 互连拓扑的 MetaXLink 架构设计核心技术攻关、芯片制程和封装工艺优化以及算力互连，实现芯片架构、制程、封装、显存、计算单元设计和互连、生态建设等功能。	上海市经信委	独立承担	验收阶段
《基于存内计算架构的推理 GPU 研发与产业化》	本课题针对神经网络、并行计算领域的加速需求，开展基于电阻型存储器异质集成工艺下推理 GPU 芯片设计及关键工艺研究，并形成应用示范。	四川省科技厅	牵头单位	验收阶段
《面向芯片安全的 EDA 工具研发》	本课题围绕 RTL 硬件安全漏洞检测中的基础理论与关键技术展开研究，主要包括关键链路提取技术、硬件安全形式化验证技术和低活性节点定位技术研究等。	湖南省科技厅	参与单位	研发阶段
《动态混合精度 AI 计算芯片研发及异构融合系统搭建》	本课题面向新一代人工智能大规模语言模型 (LLM) 和自动驾驶等应用，从计算架构、核心电路、软件工具链和原型验证系统四个方面开展动态混合精度 AI 芯片及其与 GPU 异构融合架构的研究。	北京市科学技术委员会	参与单位	研发阶段

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

3.4 沐曦募资项目重点投入高性能通用 GPU，核心产品将于 2026 年(第二代高性能通用 GPU 芯片(C600))和 2027 年下半年(第三代高性能通用 GPU 芯片(C700))加速商业化落地量产

沐曦公司本次募集资金用于投资“新型高性能通用 GPU 研发及产业化项目”、“新一代人工智能推理 GPU 研发及产业化项目”和“面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目”，均系围绕公司主营业务投向科技创新领域，符合国家产业政策和公司发展战略。

“新型高性能通用 GPU 研发及产业化项目”包括公司第二代高性能通用 GPU 芯片(代号 C600)和第三代高性能通用 GPU 芯片(代号 C700)两个研发子项目，基于国产先进工艺开发具备较高性能和更高性能的两款通用 GPU 计算芯片，应用于 AI 训练及推理、通用计算等场景，是公司曦云 C 系列训推一体芯片的后续主力产品。“新一代人工智能推理 GPU 研发及产业化项目”系公司研发下一代基于先进封装技术的云端大模型推理芯片(代号 Nx)，主要用于生成式 AI 推理。前述募投项目的顺利实施，将帮助公司完成现有核心产品线的持续迭代升级，巩固行业领先地位，进一步扩大市场份额，是公司未来几年主营业务发展及收入增长的重要基础和核心驱动。

“面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目”围绕前沿芯片技术研发和前沿 GPU 系统研发两大方向，具体包括但不限于超高带宽显存、芯粒架构、设计工艺联合优化、GPU 光互连技术、大功率 POL 电源设计优化、GPU 系统散热技术研究、scaleup 互连方式优化改进、超节点服务器系统交换架构优化等研发内容。通过实施本项目，有助于公司增强整体研发实力，丰富前沿技术储备，构筑更高的竞争壁垒，为公司未来长期稳定发展、实现经营战略目标奠定坚实的基础。

图表 21 沐曦集成公司在研发项目和进展

图表 22 沐曦集成公司募集资金运用领域

序号	项目名称	研发内容和目标	所处阶段	项目预算(万元)
1	第二代高性能通用 GPU 研发项目	本项目面向云端人工智能训练、推理及通用计算场景，研发基于国产供应链的高性能计算 GPU 芯片，将升级现有芯片架构以获得更高计算效率、更大存储，内置多精度混合算力、新一代 MetaXLink 高效互连系统并搭载新一代 MXMACA 软件栈。	试产阶段	136,977.30
2	第三代高性能通用 GPU 研发项目	本项目面向云端人工智能训练、通用计算、高性能科学计算等场景，研发基于国产供应链的下一代超高性能计算 GPU 芯片，单芯片峰值计算能力、存储、通信带宽、多芯片互连等能力得到显著提升。	研发阶段	204,015.07
3	高性能通用 GPU 优化升级项目	本项目系在第一代训练产品上的持续升级与适配，支撑在大模型训推一体、超智融合、高性能计算等各种应用场景发挥高性能和高应用适配迁移能力。	产品持续迭代阶段	不适用
4	人工智能推理 GPU 优化升级项目	本项目系在第一代推理产品上的持续升级与适配，支持在人脸识别、图像识别、音视频 AI 处理等各种应用场景发挥高性能和高应用适配迁移能力。	产品持续迭代阶段	不适用
5	图形渲染 GPU 研发项目	本项目旨在推出公司图形渲染产品线第一代产品，在人工智能推理基础上，进一步满足云游戏、图形渲染等多种应用场景对 AI 算力的需求。	研发阶段	110,000.00

单位：万元

序号	项目名称	投资总额	拟使用募集资金投入金额	备案情况
1	新型高性能通用 GPU 研发及产业化项目	340,992.38	245,919.76	已备案
1.1	其中：第二代高性能通用 GPU	136,977.30	84,101.70	
1.2	第三代高性能通用 GPU	204,015.07	161,818.06	
2	新一代人工智能推理 GPU 研发及产业化项目	57,813.35	45,305.64	已备案
3	面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目	102,115.42	99,141.19	已备案
合计		500,921.15	390,366.59	

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

新型高性能通用 GPU 研发及产业化项目建设目标是研发公司曦云 C 系列训推一体芯片的第二代和第三代产品，包括 C600 和 C700 两个子项目。其中，C600 是基于国产先进工艺开发的通用 GPU 芯片，拥有较高性能；C700 是基于国产先进工艺开发的下一代通用 GPU 芯片，拥有更高的性能。两款产品均依托于公司自主创新的核心 GPU IP 及 MXMACA 软件栈，可提供混合精度算力支持，并搭载超高带宽显存，深度融合计算密度与互连系统优化技术，支持从单卡到超大规模集群的全场景加速需求，是针对 AI 训练、推理及通用计

算的高效解决方案。

第二代高性能通用 GPU 芯片(C600)研发项目,投资总额为 136,977.30 万元,其中拟使用募集资金投入金额为 84,101.70 万元。第二代高性能通用 GPU 芯片(C600)研发项目为公司目前在研项目,于 2024 年 2 月立项,计划建设期为 3 年。预计曦云 C 系列训推一体 C600 芯片将于 2025 年 Q4 实现小批次量产,并在 2026 年实现大规模量产和持续的算法优化。

图表 23 沐曦曦云 C600 训推一体项目投资具体拆分

单位: 万元

序号	项目	投资金额	占比	使用募集资金投入
1	设备购置	21,157.29	15.45%	20,483.00
2	软件购置	2,580.52	1.88%	2,010.00
3	IP 授权费用	5,670.00	4.14%	-
4	人员工资	42,080.00	30.72%	34,643.00
5	委外研发费用	27,374.39	19.98%	18,914.10
6	流片费用	7,548.70	5.51%	5,196.60
7	租赁费用	3,023.02	2.21%	2,855.00
8	基本预备费	3,283.02	2.40%	-
9	铺底流动资金	24,260.36	17.71%	-
项目总投资		136,977.30	100.00%	84,101.70

图表 24 沐曦曦云 C600 训推一体项目实施进度安排

项目阶段	2024 年				2025 年				2026 年			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
项目立项												
软硬件购置												
产品设计开发												
流片及测试												
小批次量产												
大规模量产												
算法优化												

注: C600 已于 2024 年 10 月交付流片,并于 2025 年 5 月回片。

资料来源:沐曦集成电路(上海)股份有限公司招股说明书,华安证券研究所

资料来源:沐曦集成电路(上海)股份有限公司招股说明书,华安证券研究所

第三代高性能通用 GPU 芯片(C700)研发项目,投资总额为 204,015.07 万元,其中拟使用募集资金投入金额为 161,818.06 万元。第三代高性能通用 GPU 芯片(C700)研发项目于 2025 年 4 月立项,计划建设期为 3 年。预计曦云 C 系列训推一体 C700 芯片将于 2026 年 Q2 实现流片和测试任务,并于 2027 年 Q3 开始实现小批次量产,并同步在 2027 年下半年实现大规模量产和持续的算法优化。

图表 25 沐曦曦云 C700 训推一体项目投资具体拆分

单位: 万元

序号	项目	投资金额	占比	使用募集资金投入
1	设备购置	23,452.50	11.50%	23,452.50
2	软件购置	7,671.56	3.76%	7,671.56
3	IP 授权费用	8,750.00	4.29%	8,750.00
4	人员工资	60,480.00	29.64%	60,017.00
5	委外研发费用	40,700.00	19.95%	40,700.00
6	流片费用	17,927.00	8.79%	17,927.00
7	租赁费用	3,300.00	1.62%	3,300.00
8	基本预备费	4,868.43	2.39%	-
9	铺底流动资金	36,865.58	18.07%	-
项目总投资		204,015.07	100.00%	161,818.06

图表 26 沐曦曦云 C700 训推一体项目实施进度安排

项目阶段	2025 年				2026 年				2027 年			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
项目立项												
软硬件购置												
产品设计开发												
流片及测试												
小批次量产												
大规模量产												
算法优化												

资料来源:沐曦集成电路(上海)股份有限公司招股说明书,华安证券研究所

资料来源:沐曦集成电路(上海)股份有限公司招股说明书,华安证券研究所

新一代人工智能推理 GPU 研发及产业化项目建设目标是研发公司下一代基于先进封装技术的云端大模型推理芯片(代号 Nx),专为生成式 AI 推理市场设计。该芯片计划采用公司最新的第三代指令集和先进封装工艺,通过超高显存带宽和大容量显存提高 token 吞吐量,同时保持低功耗和高能效,大幅降低 AI 推理成本。

项目完成后,相关产品可以广泛应用于国内商业化智算中心,为其提供高性价比的大模型推理解决方案;或者用于 AIPC 等端侧低延迟、独立用户推理场景,支持从百亿到万亿各种参数规模的模型推理。同时,也可以与公司的训练芯片进行互连,提升协同效率。

项目投资总额为 57,813.35 万元,其中拟使用募集资金投入金额为 45,305.64 万元,项目计划建设期为 3 年。

图表 27 沐曦新一代人工智能推理 GPU 研发及产业化项目投入拆分

单位：万元				
序号	项目	投资金额	占比	使用募集资金投入
1	设备购置	8,074.00	13.97%	8,074.00
2	软件购置	1,448.64	2.51%	1,448.64
3	IP 授权费用	6,550.00	11.33%	6,550.00
4	人员工资	14,400.00	24.91%	14,400.00
5	委外研发费用	5,740.00	9.93%	5,740.00
6	流片费用	9,093.00	15.73%	9,093.00
7	基本预备费	1,359.17	2.35%	-
8	铺底流动资金	11,148.54	19.28%	-
项目总投资		57,813.35	100.00%	45,305.64

图表 28 沐曦新一代人工智能推理 GPU 研发及产业化项目进展安排

项目阶段	T+1 年				T+2 年				T+3 年			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
项目立项												
软硬件购置												
产品设计开发												
流片及测试												
小批量量产												
大规模量产												

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目定位于行业前沿技术及基础研究，公司将主要围绕前沿芯片技术研发和前沿 GPU 系统研发两大方向开展研发工作。其中，前沿芯片技术研发具体包括但不限于超高带宽显存、先进工艺芯粒架构、设计工艺联合优化(DTCO)、GPU 光互连技术等研发内容；前沿 GPU 系统研发具体包括但不限于大功率 POL 电源设计与优化、GPU 系统散热技术研究、scaleup 互连方式优化改进、超节点服务器系统交换架构的演进与优化等研发内容。

面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目不直接产生经济效益，主要通过扩充公司现有研发团队，引进一批跨学科、多领域的高精尖专业人才，同步引入各类高端研发设备，以提升公司整体研发实力，实现关键技术突破，丰富前沿技术储备。项目成功实施后，不仅能够增厚公司技术积淀，构筑更高的竞争壁垒，为公司未来长期发展奠定坚实基础，亦有助于推动国内 GPU 整体行业的技术创新发展。

图表 29 沐曦面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目投入拆分

单位：万元				
序号	项目	投资金额	占比	使用募集资金投入
1	场地费用	1,076.49	1.05%	1,076.49
2	设备购置	34,842.20	34.12%	34,842.20
3	软件购置	10,030.00	9.82%	10,030.00
4	IP 授权费用	4,500.00	4.41%	4,500.00
5	人员工资	24,000.00	23.50%	24,000.00
6	委外研发费用	6,000.00	5.88%	6,000.00
7	流片费用	14,420.00	14.12%	14,420.00
8	租赁费用	4,272.49	4.18%	4,272.49
9	基本预备费	2,974.24	2.91%	-
项目总投资		102,115.42	100.00%	99,141.19

图表 30 沐曦面向前沿领域及新兴应用场景的高性能 GPU 技术研发项目进展安排

项目阶段	T+1 年		T+2 年		T+3 年		T+4 年	
	H1	H2	H1	H2	H1	H2	H1	H2
场地租赁								
人员招聘								
软硬件购置								
技术与开发								
测试验证及流片								

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

3.5 沐曦具备多重竞争优势包括技术优势、通用性优势、高效优势、易用性优势、产业化优势、团队优势

在实际人工智能计算场景中，芯片的效率输出由芯片性能、集群性能和软件生态的易用性共同决定。沐曦坚持以自主技术创新为导向，持续深耕 GPU 行业，在专业知识储备深厚、行业经验丰富的研发团队带领下，沐曦通过技术创新和设计优化，持续提升 GPU 架构的通用性、GPU 产品的高能效(单卡性能和集群性能)与 GPU 软件生态的易用性，率先实现了“技术-产品-市场”的产业化应用。

图表 31 沐曦集成公司具备六大优势（技术、通用性、高能效、易用性、产业化、世界级技术团队）

主要优势	优势说明
<p>技术优势:技术积累深厚全面,核心技术自主可控</p>	<p>是国内少数几家全面系统掌握了 GPU 架构、GPU IP、先进制程 GPU 芯片及其基础系统软件研发、设计和量产等全流程核心技术的企业之一,拥有自主研发的 GPU IP、GPU 指令集和架构等,充分掌握并构建了高性能 GPU 处理器架构设计技术、高性能 GPU 流处理器设计技术、高性能 GPU 多级缓存结构及内存管理设计技术、高性能 GPU 系统控制及虚拟化技术、MetaXLink 高速互连技术、芯粒架构与先进封装设计技术、高性能 GPU 动态功耗管理技术、高性能 GPU 验证方法学及智能验证技术、GPU 异构计算编程语言与开发环境、人工智能编译器与编译技术、GPU 高性能通信库设计与优化技术、GPU 高性能数学库设计与优化技术、GPU 核心驱动程序、GPU 软硬件协同优化技术等一整套完善且自主可控的底层核心技术体系。</p>
<p>通用性优势:以 GPU 架构和 GPU IP 打造全系列 GPU 产品,具备通用性、灵活性和自主性</p>	<p>公司通用、灵活的 GPU 架构能够高效适应云端快速迭代的复杂算法和大量非传统模型结构,从而支持更为广泛的人工智能应用场景。</p> <p>GPU IP 与指令集方面,公司是国内少数坚持自主研发核心 GPU IP 的集成电路设计企业,经过多年产品设计及落地经验的积累,已形成了全面覆盖计算(人工智能训练、推理、通用计算)和渲染场景的核心 GPU IP,打造了由超过 600 条(XCORE1.0 计算 GPU)、800 条(XCORE2.0 渲染 GPU)指令组成自主安全可控的 MXMACA 指令集。自研核心 GPU IP 具有灵活性和自主性优势,能够通过 GPU IP 的设计优化摆脱对最先进工艺的依赖、弥补工艺代差带来的性能损失。公司自主研发的核心 GPU IP 不仅满足了当前复杂多样的计算需求,还为未来的技术创新和产品迭代奠定了坚实的基础。</p>
<p>高能效优势:GPU 硬件设计在多个维度突破国外垄断</p>	<p>公司训推一体 GPU 具备国内领先的计算能力,拥有丰富的标量、矢量和张量计算单元,支持 FP64、FP32、TF32、FP16、BF16、FP8、INT8、FP6、FP4 等混合精度计算,能够满足人工智能训练和推理、通用计算、科学计算等多元化的算力需求,大幅提升了智算集群的多场景融合应用能力。同时,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,能够针对不同场景灵活自适应多级缓存控制策略,实现对高速缓存的精细管控、有效提升缓存命中率,满足 GPU 高算力下对于高带宽的需求。</p> <p>公司具备国内稀缺的高带宽、超多卡互连能力,在集群性能、灵活性和可扩展性方面均处于行业领先地位。随着大模型参数规模指数级增长,单卡有效算力难以应对超大规模的算力需求,高带宽、低延迟的卡间互连技术成为关键竞争要素。公司自主研发的 MetaXLink 高速互连技术突破了传统 PCIe 总线在带宽和延迟方面的限制,达到了与英伟达 4nm 制程工艺下旗舰产品(H200)相当的互连带宽性能。同时,公司的 GPU 产品可实现 2-64 卡等多种互连拓扑及超节点架构,满足从中小型训练到超大规模集群的差异化需求,为构建高密度算力和大规模集群、处理更为复杂的人工智能任务提供了关键技术支撑。</p>
<p>易用性优势:“自主创新与开放兼容”双轨并行,打造了具备易用性和可扩展性的软件生态</p>	<p>易用性方面,沐曦自主研发的 MXMACA 软件栈为开发者提供了完整的 AI 开发工具链,利用自有技术完成对相关开源框架和工具的适配、编译、验证、优化等,不断降低开发者使用门槛。高性能方面,沐曦基于 GPU 软硬件协同优化技术,从软件实际需求出发,精准定义硬件的功能与性能,实现软硬件之间的深度协同与优化,充分释放 GPU 硬件性能潜力。</p> <p>通用性方面,沐曦 MXMACA 软件栈实现了对 CUDA 生态的高度兼容,为人工智能、通用计算和大数据处理等领域的现有应用提供了高效迁移至 MXMACA 软件栈的能力,开发者无需过多修改现有代码即可将应用快速部署到公司的 GPU 产品上运行,大幅降低客户应用迁移本、提升迁移效率。MXMACA 软件栈在 AI 框架、高性能算子库的支持度和模型的适配度方面具备深厚的积累,支持超过 6,000 个 CUDA 应用,与超过 1,000 个模型实现了原生适配。基于较强的生态兼容性,沐曦 MXMACA 软件栈能够开放拥抱全球开发者丰富的开源成果,拥有丰富的人工智能落地场景。坚持推动软件生态开源共享,通过产学研合作不断助力大模型应用在千行百业加速落地。自 2025 年 2 月 14 日沐曦开源社区开放以来,截至 2025 年 6 月 1 日,公司 MXMACA 软件栈注册用户超过 11,000 人,网络 API 调用次数超过 253 万次,文件下载量超过 4 万次。公司是国内少数提供核心软件生态开源、提供版本 SDK 公开下载渠道、提供英文开发文档资料的 GPU 公司, MXMACA 软件栈已公开 200 余份开发文档和手册,并已成功引入中国高校教育体系,覆盖超过半数的中国 C9 高校,有望打破国外软件生态垄断、支持中国人工智能产业转型升级。</p>
<p>产业化优势:优质的产业链上下游资源,快速实现商业化落地</p>	<p>公司仅用 3 年时间即实现两款芯片一次流片并成功量产,快速实现了商业化、规模化落地,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。凭借突出的产品性能和稳定的供应能力,截至 2025 年 Q1 末,沐曦 GPU 产品累计销量超过 25,000 颗。沐曦深度构建“1+6+X”生态与商业布局,基础算力底座方面相继交付 9 大智算集群,算力网络覆盖国家人工智能公共算力平台、运营商智算平台和商业化智算中心,区域横跨北京、上海、杭州、长沙、中国香港等地区,并逐渐向更多区域延伸。沐曦 GPU 产品深度赋能众多行业应用场景,已率先布局教研、金融、交通、能源、医疗健康、大文娱等行业,产品赋能真实应用场景的竞争力和交付能</p>

力得到充分验证。

公司构建了稳定的供应链体系，积累了丰富的供应链管理经验和拥有动态、灵活的库存管理机制，并较早启动国产供应链布局、提早规划并设计了基于国产供应链的产品，能够确保产品高质量及可持续的供应与交付。截至本招股说明书签署日，公司在晶圆生产制造、封测、存储颗粒、EDA、IP 等核心环节的国产化上均已取得一定突破。随着市场影响力和产业链资源逐渐提升，公司将持续培育国产供应商，进一步推动关键供应环节的国产配套，全面提升国内集成电路产业链的竞争力。

团队优势:拥有稀缺、稳定的全建制世界级技术团队

公司创始团队拥有深厚的 GPU 技术积累及全流程量产经验，在国际领先 GPU 公司负责设计并量产了多款高性能计算 GPU 和高性能渲染 GPU，具备敏锐的市场洞察能力、前瞻性视野和国内稀缺的万卡大集群落地经验。创始人陈维良先生拥有 20 年以上 GPU 设计及量产经验，带领团队主导并完成多款高性能 GPU 产品的流片和量产，曾获“上海产业菁英高层次人才-产业领军人才”、“中国信息协会常务理事”等称号或荣誉。联合创始人彭莉女士是芯片架构资深专家，曾作为主架构师完成了多款高端复杂的 GPU 芯片设计。联合创始人杨建博士是三维图形与高性能计算生态领域资深专家，拥有丰富的 GPU 芯片设计及软件生态开发经验。彭莉女士和杨建博士曾在国际领先 GPU 公司获得企业院士称号。

公司创始团队以近二十年 GPU 行业的工作历程，聚合起一批扎根芯片设计、GPU 和人工智能行业的资深工程师，组成了核心技术研发团队和经营管理团队，形成了一支富有远见、深刻洞察全球 GPU、人工智能行业技术发展趋势，具有持续自主创新能力的技术研发团队。截至 2025 年 3 月 31 日，公司已在北京、南京、成都、武汉、杭州、中国香港等地建立全资子公司及研发中心，共有 652 名研发人员，占员工总人数的比例为 74.94%;公司员工中有 520 人拥有硕士及以上学位，为公司持续的技术创新和产品研发提供了重要的人才基础。此外，公司的市场、运营等部门的核心团队均拥有行业内知名公司多年的工作经历，具有丰富的产业经验和专业的管理能力。

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

竞争格局方面，国内市场来看，国内人工智能芯片市场呈现多元化和快速发展的特征。GPU 作为人工智能芯片的重要组成部分，在人工智能计算市场占据核心地位。近年来，受到美国高性能计算芯片出口管制与国内政策推动等因素影响，海外厂商在中国市场的份额呈现明显下降趋势。国产人工智能芯片公司迎来黄金发展期，以不同技术路径切入市场，各具特色。因此，在人工智能计算领域，公司国内主要竞争对手包括两类:1)设计、研发及销售 GPU 芯片的公司，除国际巨头英伟达、AMD 外，还包括国内上市公司海光信息，国内未上市公司天数智芯、壁仞科技、摩尔线程等;2)采用其他技术路线(例如 ASIC/DSA 等)的 AI 芯片公司，包括国内上市公司寒武纪，国内未上市公司华为海思、昆仑芯、燧原科技等。此外，在图形渲染领域，以景嘉微、摩尔线程为代表的国内 GPU 厂商也已在各自的细分领域取得了阶段性产业成果。

图表 32 沐曦(发行人)与竞争对手对比

项目	英伟达	AMD	寒武纪	海光信息	景嘉微	龙芯中科	发行人
股票代码	NVDA.O	AMD.O	688256.SH	688041.SH	300474.SZ	688047.SH	不适用
主要产品及应用领域	GPU 显卡、图形显卡及配套产品，主要应用于数据中心、桌面游戏、专业视觉设计、智能驾驶、机器人等场景	CPU 显卡、GPU 显卡、图形显卡、SoC 等，主要应用于数据中心、个人电脑、工作站等场景	智能芯片及加速卡、智能处理器 IP，主要应用于数据中心、云计算、消费电子等场景	GPU、DCU (GPU) 及配套产品、技术，其 DCU 产品主要应用于数据中心	图形显控(图形 GPU)、小型专用化雷控类处理器等	工控类处理器、信息类处理器	GPU 产品及配套软件与计算平台，广泛应用于人工智能、通用计算、智慧城市、自动驾驶、云渲染等前沿领域
主要产品型号	B200、H100、A100	M1325X、MI250X	MLU590、MLU370	BW100、K100	JM9 系列	龙芯 1 号、龙芯 2 号、龙芯 3 号	曦云 C500
2024 年营业收入(万元)	93,617,242.83	18,497,801.15	117,446.44	916,214.81	46,634.24	50,425.72	74,307.16
2022-2024 年营业收入复合增长率	119.95%	4.52%	26.92%	33.70%	-36.43%	-17.38%	4074.52%
2024 年收入构成	数据中心产品: 88.27%; 游戏 GPU 和相关产品: 8.70%; 专业视觉设计 GPU: 1.44%; 智能驾驶 GPU 及相关产品: 1.30%; 其他业务: 0.30%	数据中心产品: 48.78%; 游戏: 10.06%; 企业、嵌入式和半定制: 13.79%; 客户端: 27.36%	云端智能芯片及加速卡: 99.30%; 边缘端智能芯片及加速卡: 0.56%; 其他: 0.15%	高端处理器: 99.70%; 技术服务: 0.30%	图形显控领域产品: 52.53%; 小型专用雷控类领域产品: 13.01%; 其他: 34.46%	信息化类芯片: 53.30%; 解决方案: 28.88%; 工业控制及仪表芯片: 0.47%; 其他业务: 0.01%	训练一体 GPU 显卡: 68.99%; 训练一体 GPU 服务器: 28.29%; 推理推理 GPU 显卡: 0.47%; IP 授权: 0.13%; 其他: 2.11%

图表 33 沐曦(发行人)与竞争对手对比(主要财务数据)

项目	英伟达	AMD	寒武纪	海光信息	景嘉微	龙芯中科	发行人
2024 年归母净利润(万元)	52,283,383.20	1,177,236.99	-45,233.88	193,099.05	-16,511.95	-62,534.71	-140,887.94
2024 年毛利率	74.99%	49.35%	56.71%	63.72%	43.70%	31.04%	53.43%
2024 年研发投入(万元)	9,264,374.46	4,631,469.84	121,587.36	344,575.56	28,064.86	43,019.36	90,089.04
2024 年研发投入率	9.90%	25.04%	103.53%	37.61%	60.18%	85.31%	121.24%
2024 年末总资产(万元)	80,061,441.39	49,662,040.14	671,781.25	2,855,949.20	756,481.43	349,748.34	388,958.25
2024 年末归母净利润(万元)	56,908,396.53	41,298,707.52	542,265.87	2,025,095.92	707,031.99	293,639.71	117,830.41

注 1: 数据来源 Wind 及公开资料整理。
 注 2: 英伟达财务报告期间为 2024 年 1 月末至 2025 年 1 月末。
 注 3: 英伟达数据中心产品为通用 GPU、游戏 GPU 和专业视觉设计 GPU 为图形 GPU; AMD 数据中心产品主要为 GPU。
 注 4: 海光信息高端处理器收入包括 CPU 及 DCU 收入，未披露其中 DCU 收入占比。根据海光信息《首次公开发行股票并在科创板上市招股说明书(注册稿)》，海光信息 2021 年度 DCU 收入占比为 10.34%。
 注 5: 景嘉微主要产品为图形 GPU，主要应用于军工领域。
 注 6: 龙芯中科主要产品为 CPU，GPU 产品占比相对较小且其 GPU 产品主要应用于 PC 端。

资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

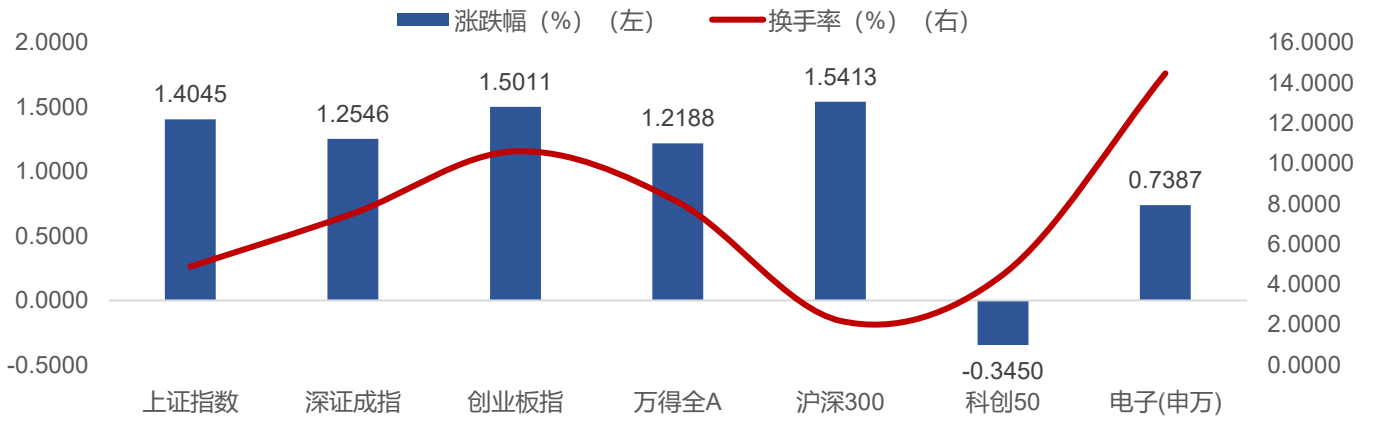
资料来源：沐曦集成电路(上海)股份有限公司招股说明书，华安证券研究所

4 市场行情回顾

4.1 行业板块表现

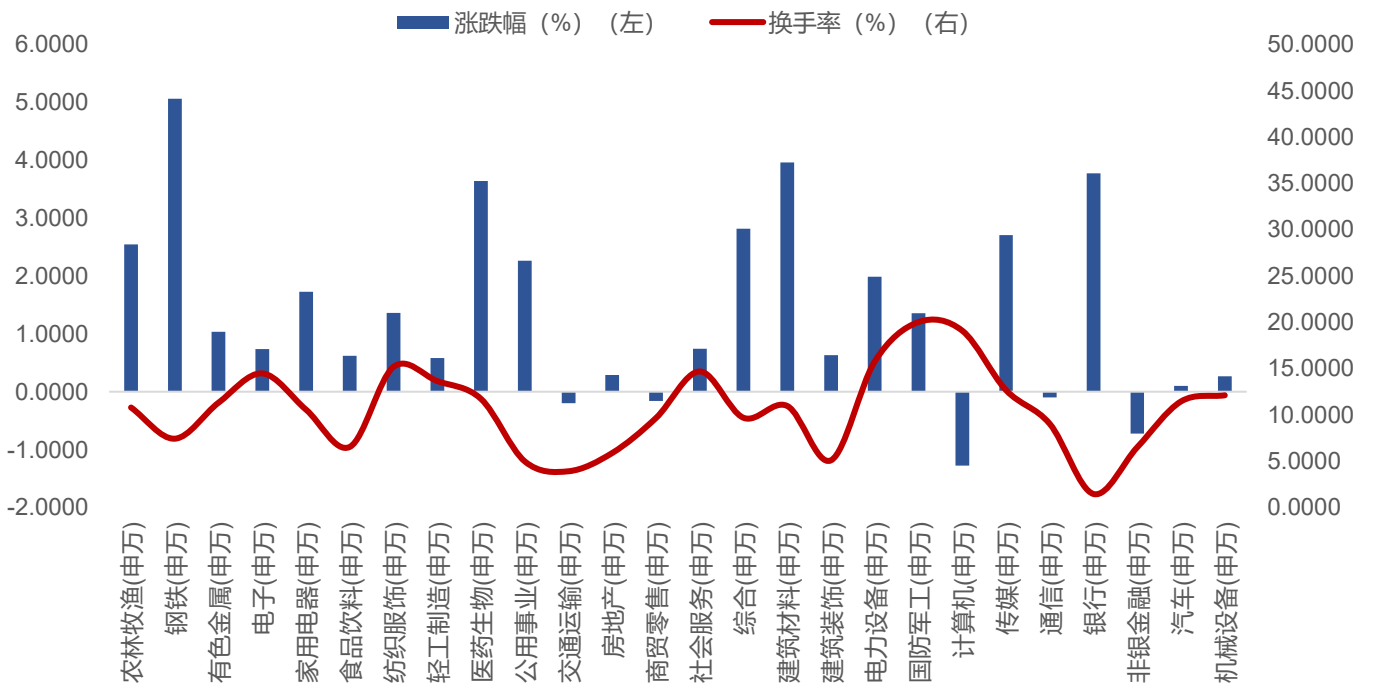
从指数表现来看，本周（2025-06-30至2025-07-04），上证指数周涨跌+1.40%，深证成指涨跌幅为+1.25%，创业板指数涨跌幅+1.50%，科创50涨跌幅为-0.35%，申万电子指数涨跌幅+0.74%。板块行业指数来看，表现最好的是集成电路封测，涨幅为8.32%，面板表现较弱，涨幅为3.11%；板块概念指数来看，表现最好的是印制电路板指数，涨幅为8.31%，表现最弱的是数字芯片设计，跌幅为2.38%。

图表 34 板块指数



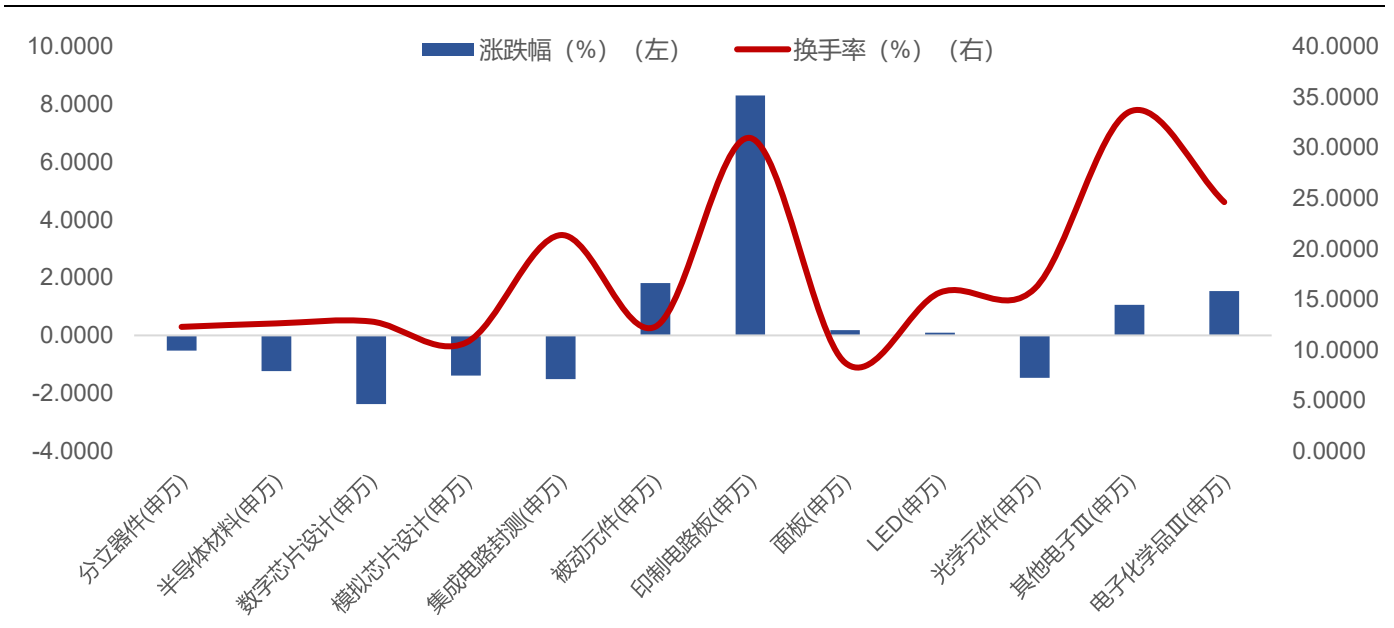
资料来源：Wind，华安证券研究所

图表 35 行业板块涨跌幅和换手率（本周电子在申万一级行业指数中 16/26 位）



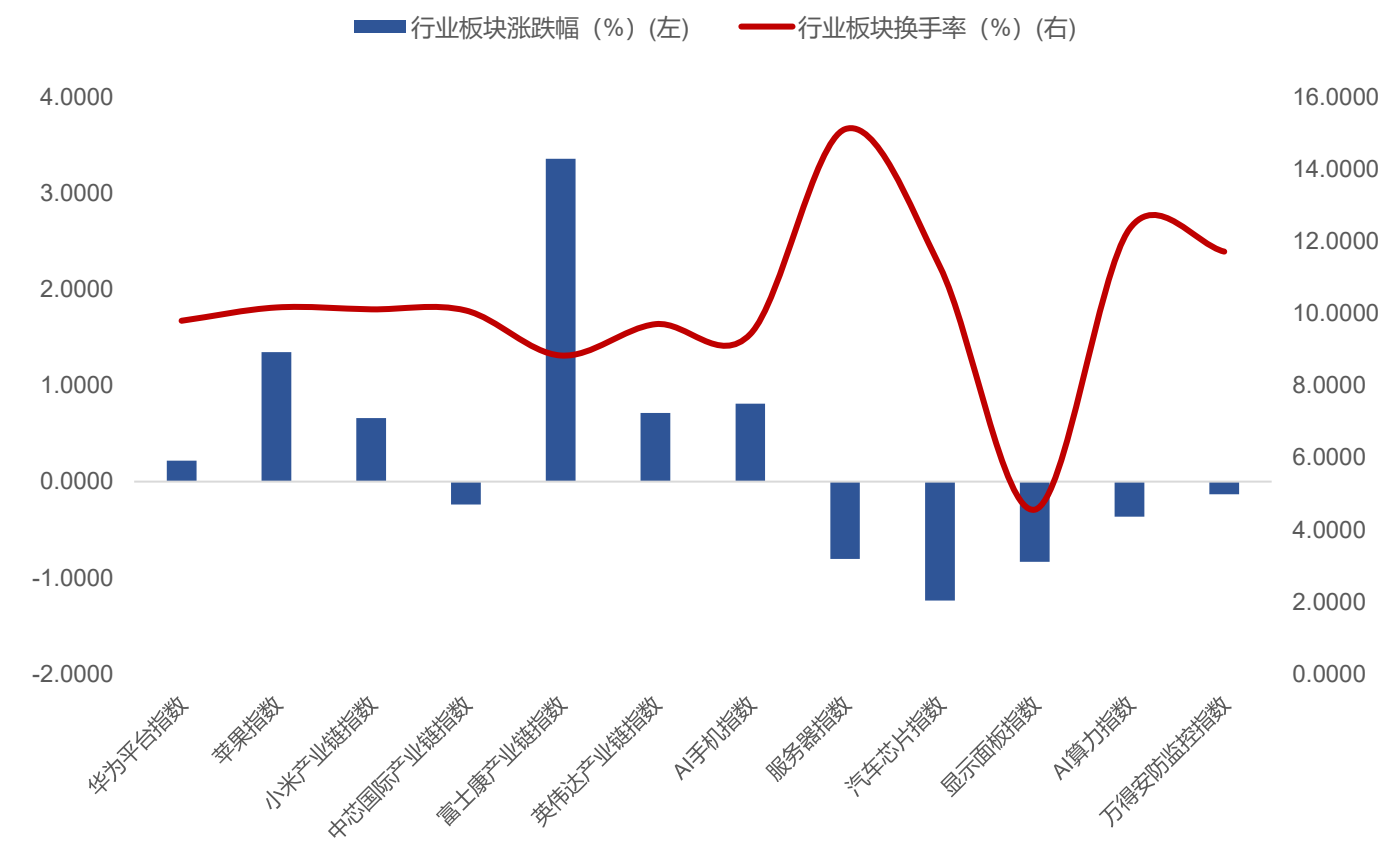
资料来源：Wind，华安证券研究所

图表 36 电子行业细分板块涨跌幅和换手率



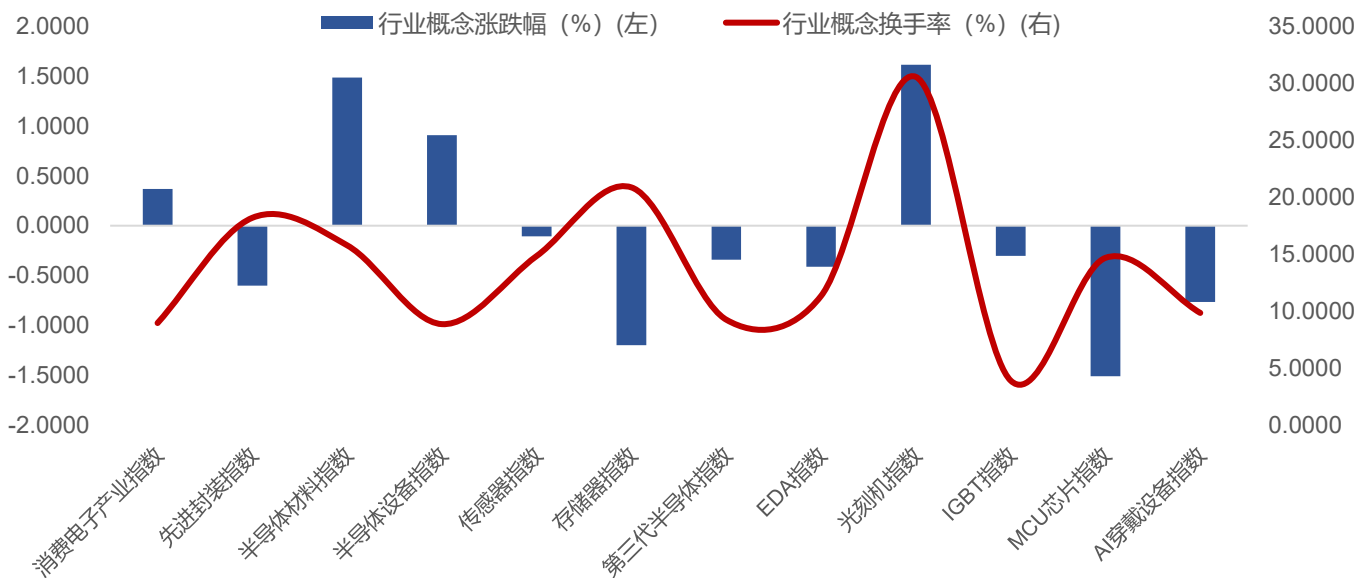
资料来源: Wind, 华安证券研究所

图表 37 电子行业热门细分概念指数涨跌幅和换手率



资料来源: Wind, 华安证券研究所

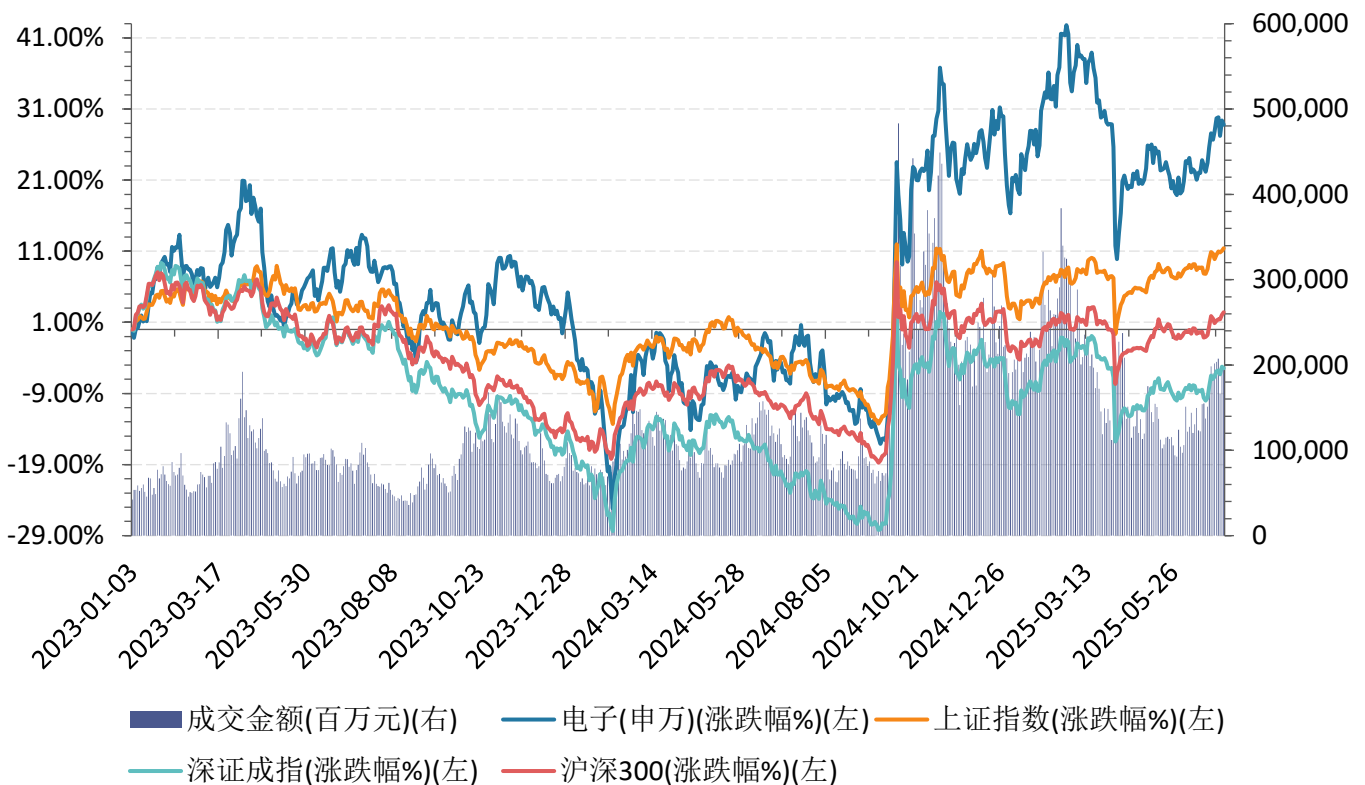
图表 38 电子行业热门细分概念指数涨跌幅和换手率



资料来源: Wind, 华安证券研究所

图表 39 电子行业行情图

电子(申万)-行情图



资料来源: Wind, 华安证券研究所

4.2 电子个股表现

从个股表现看，本周（2025-06-30 至 2025-07-04）表现最好的前五名分别是逸豪新材、隆扬电子、景旺电子、金安国纪、奥士康；光智科技、雅葆轩、合力泰、太龙股份、显盈科技表现较弱。

从今年表现来看，表现最好的前五名分别是胜宏科技、南亚新材、逸豪新材、迅捷兴、好上好；盛科通信-U、国星光电、星宸科技、国光电器、润欣科技表现较弱。

图表 40 个股涨跌幅 (%)

周表现最好前十		周表现最差前十		今年以来表现最好前十		今年以来表现最差前十	
逸豪新材	38.60	光智科技	-15.82	胜宏科技	249.02	盛科通信-U	-30.42
隆扬电子	29.81	雅葆轩	-10.50	南亚新材	119.20	国星光电	-29.23
景旺电子	22.45	合力泰	-9.49	逸豪新材	109.07	星宸科技	-28.86
金安国纪	21.54	太龙股份	-9.34	迅捷兴	103.62	国光电器	-27.57
奥士康	19.92	显盈科技	-9.05	好上好	101.92	润欣科技	-27.52
好上好	18.04	智新电子	-8.08	中京电子	96.58	光智科技	-27.21
盈方微	17.99	长阳科技	-7.70	慧为智能	95.59	电连技术	-26.03
广信材料	15.71	龙迅股份	-7.67	隆扬电子	93.59	天键股份	-25.73
宝明科技	15.23	威贸电子	-7.61	天山电子	86.87	龙图光罩	-25.00
鹏鼎控股	14.57	科翔股份	-7.50	旭光电子	85.82	灿芯股份	-23.29

资料来源：Wind，华安证券研究所

风险提示：

1) 下游需求不及预期；2) 资本开支不及预期；3) 技术迭代不及预期。

分析师与研究助理简介

陈耀波 (执业证书号: S0010523060001): 北京大学管理学硕士, 香港大学金融学硕士, 华中科技大学电信系学士。8年买方投研经验, 历任广发资管电子研究员, TMT组组长, 投资经理助理; 博时基金投资经理助理。行业研究框架和财务分析体系成熟, 擅长买方视角投资机遇分析对比, 全面负责团队电子行业研究工作。

李元晨 (执业证书号: S0010524070001): 墨尔本大学会计和金融学本科, 悉尼大学数据分析和金融学硕士。2022年加入华安证券研究所, 目前重点覆盖 AI 芯片、半导体材料设备、科创新股、MEMS 和传感器等。

重要声明

分析师声明

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, 以勤勉的执业态度、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, 本报告所采用的数据和信息均来自市场公开信息, 本人对这些信息的准确性或完整性不做任何保证, 也不保证所包含的信息和建议不会发生任何变更。报告中的信息和意见仅供参考。本人过去不曾与、现在不与、未来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接收任何形式的补偿, 分析结论不受任何第三方的授意或影响, 特此声明。

免责声明

华安证券股份有限公司经中国证券监督管理委员会批准, 已具备证券投资咨询业务资格。本报告由华安证券股份有限公司在中华人民共和国(不包括香港、澳门、台湾)提供。本报告中的信息均来源于合规渠道, 华安证券研究所力求准确、可靠, 但对这些信息的准确性及完整性均不做任何保证。在任何情况下, 本报告中的信息或表述的意见均不构成对任何人的投资建议。在任何情况下, 本公司、本公司员工或者关联机构不承诺投资者一定获利, 不与投资者分享投资收益, 也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。投资者务必注意, 其据此做出的任何投资决策与本公司、本公司员工或者关联机构无关。华安证券及其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并进行交易, 还可能为这些公司提供投资银行服务或其他服务。

本报告仅向特定客户传送, 未经华安证券研究所书面授权, 本研究报告的任何部分均不得以任何方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。如欲引用或转载本文内容, 务必联络华安证券研究所并获得许可, 并需注明出处为华安证券研究所, 且不得对本文进行有悖原意的引用和删改。如未经本公司授权, 私自转载或者转发本报告, 所引起的一切后果及法律责任由私自转载或转发者承担。本公司并保留追究其法律责任的权利。

投资评级说明

以本报告发布之日起 6 个月内, 证券(或行业指数)相对于同期相关证券市场代表性指数的涨跌幅作为基准, A 股以沪深 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; 美国市场以纳斯达克指数或标普 500 指数为基准。定义如下:

行业评级体系

- 增持—未来 6 个月的投资收益率领先市场基准指数 5% 以上;
- 中性—未来 6 个月的投资收益率与市场基准指数的变动幅度相差 -5% 至 5%;
- 减持—未来 6 个月的投资收益率落后市场基准指数 5% 以上;

公司评级体系

- 买入—未来 6-12 个月的投资收益率领先市场基准指数 15% 以上;
- 增持—未来 6-12 个月的投资收益率领先市场基准指数 5% 至 15%;
- 中性—未来 6-12 个月的投资收益率与市场基准指数的变动幅度相差 -5% 至 5%;
- 减持—未来 6-12 个月的投资收益率落后市场基准指数 5% 至 15%;
- 卖出—未来 6-12 个月的投资收益率落后市场基准指数 15% 以上;
- 无评级—因无法获取必要的资料, 或者公司面临无法预见结果的重大不确定性事件, 或者其他原因, 致使无法给出明确的投资评级。