下载日志已记录，仅供内部参考

国海证券 SEALAND SECURITIES

证券研究报告
半导体
2026年10月08日

沐曦股份（688802.SH）深度报告：高性能GPU提供商，互联/集群与软件生态持续完善

评级：增持(维持)

刘熹(证券分析师)
S0350523040001
liux10@ghzq.com.cn

谢婧茹(联系人)
S0350125070015
xiejr01@ghzq.com.cn

下载日志已记录，仅供内部参考

最近一年走势



相对沪深300表现

表现	1M	3M	12M
沐曦股份	-26.3%	-40.0%	-
沪深300	-5.5%	-12.5%	-

市场数据 2026/09/30

当前价格（元）	497.08
52周价格区间（元）	451.10-1,033.00
总市值（百万）	198,881.71
流通市值（百万）	16,152.41
总股本（万股）	40,010.00
流通股本（万股）	3,249.46
日均成交额（百万）	1,418.13
近一月换手（%）	14.26

相关报告

《沐曦股份（688802）科创板公司普通报告：2026Q2归母扣非实现盈利，曦云C600于2026年5月完成量产（增持）*半导体*刘熹》——2026-09-07

《沐曦股份（688802）2025年业绩快报点评：营收倍增带动利润减亏，互联网/运营商客户有望取得突破（增持）*半导体*刘熹》——2026-03-03

下载日志已记录，仅供内部参考

核心提要

- **本篇报告解决了以下几点问题：1、沐曦股份公司概况与业务布局；2、全栈GPU产品矩阵与性能迭代；3、互联/集群与软件生态建设；4、公司优势。**
- **公司概况：高性能GPU提供商，产品迭代持续提速**
  - 沐曦股份是国内领先的高性能GPU设计企业及智算解决方案提供商。公司以自主研发的“曦云C系列、曦思N系列、曦彩G系列、曦索X系列”全栈GPU芯片为核心，配合自研“MXMACA软件栈”，分别面向云侧AI大模型训推、云/端AI推理以及图形渲染、科学智能等场景提供国产算力基座。
  - **截至2025年底，公司GPU产品累计销量超5.5万颗，并批量部署于国家级智算平台、运营商智算平台及商业智算中心等10余个智算集群**，应用领域覆盖金融、能源、教科研、医疗健康、大文娱等多个行业。
- **GPU产品：四大系列覆盖全场景，曦云C600有望成为下一代收入主力**

公司基于自研GPU指令集、SoC及高速互联等技术，推出曦思N系列（智算推理）、曦云C系列（训推一体）、曦彩G系列（图形渲染）及曦索X系列（科学智能）四大系列GPU：

  - 曦云C500综合性能对标英伟达A100，于2024年2月量产；**曦云C600于2026年5月正式量产，搭载HBM3e显存，新增FP8 Tensor及Tensor转置指令，有望成为下一代主力产品。曦云C700预计2027年开启小批量量产；**
  - 曦思N100/N260/N300覆盖传统人工智能到云/端AI推理；曦彩G100已完成GPU IP设计与验证；曦索X206/X302面向科学智能。
- **互联/集群与软件生态：MetaXLink带宽对标H200，MXMACA兼容CUDA**
  - 公司自研MetaXLink卡间高速互联协议，根据公司招股意向书，**其互连带宽已达到与英伟达H200相当的性能水平，**支持2卡至64卡等多种互连拓扑，可支撑千亿参数以上规模AI大模型训练算力集群建设。
  - **于2026年WAIC大会发布新一代AI超节点产品曦景S600，采用OEX架构设计，实现单机柜64张GPU高速全互连，并支持灵活扩展至万卡级集群。**
  - 软件生态方面，**自研MXMACA全栈软件工具链实现对CUDA生态的较好兼容，**已支持超过6000个CUDA应用及超过2200个高性能算子，注册用户及实际使用客户合计超15万人。

下载日志已记录，仅供内部参考

核心提要

- **公司优势：覆盖国家/运营商/互联网客户，较早启动国产供应链布局**
  - 公司客户结构多元，**主要基本盘为国家智算中心、运营商智算中心及互联网厂商智算中心，产品覆盖北京、上海、杭州、长沙、无锡及香港等地智算平台；2026年上半年新增与上海联通签署战略合作协议。**除基本盘外，公司积极布局科研、金融、交通、能源、医疗健康、大文娱六大行业市场。
  - 模型适配方面，**自2025年底至今完成37个主流旗舰模型的Day0适配，覆盖智谱AI、阿里千问、MiniMax、DeepSeek、阶跃星辰、腾讯混元、百度、美团等模型厂商。**供应链方面，公司**较早启动国产化供应链布局，在晶圆生产制造、封装测试、存储颗粒、EDA、IP等核心环节的国产化上均已取得一定突破；**根据公司招股意向书，其仅用3年时间实现两款芯片一次流片并成功量产。
- **投资建议：**公司是国内领先的高性能GPU芯片设计公司，在大模型算力需求增长和国产替代趋势下有望充分受益。我们预计公司2026~2028年营收分别为38.88/87.63/199.51亿元，归母净利润分别为3.23/23.90/73.68亿元，维持“增持”评级。
- **风险提示：**经营风险、管理风险、财务风险、供应链安全风险、产品商业化不及预期风险、性能与生态不及宣传风险、盈利预测假设不及预期风险、可比公司估值局限性提示、国内外公司不具备完全可比性及对标资料局限性风险、非经常性损益不可持续及扣非持续亏损风险、估值过高及股价高波动风险、客户集中度风险、流通盘小/换手率高/解禁波动风险、模型Day 0适配不等于订单及收入风险、2026H1归母盈利主要来自公允价值变动收益且扣非仍亏损风险、技术迭代风险、地缘政治及供应链受限风险。

下载日志已记录，仅供内部参考

核心提要

- **公司优势：覆盖国家/运营商/互联网客户，较早启动国产供应链布局**
  - 公司客户结构多元，**主要基本盘为国家智算中心、运营商智算中心及互联网厂商智算中心，产品覆盖北京、上海、杭州、长沙、无锡及香港等地智算平台；2026年上半年新增与上海联通签署战略合作协议。**除基本盘外，公司积极布局科研、金融、交通、能源、医疗健康、大文娱六大行业市场。
  - 模型适配方面，**自2025年底至今完成37个主流旗舰模型的Day0适配，覆盖智谱AI、阿里千问、MiniMax、DeepSeek、阶跃星辰、腾讯混元、百度、美团等模型厂商。**供应链方面，公司**较早启动国产化供应链布局，在晶圆生产制造、封装测试、存储颗粒、EDA、IP等核心环节的国产化上均已取得一定突破；**根据公司招股意向书，其仅用3年时间实现两款芯片一次流片并成功量产。
- **投资建议：**公司是国内领先的高性能GPU芯片设计公司，在大模型算力需求增长和国产替代趋势下有望充分受益。我们预计公司2026~2028年营收分别为38.88/87.63/199.51亿元，归母净利润分别为3.23/23.90/73.68亿元，维持“增持”评级。
- **风险提示：**经营风险、管理风险、财务风险、供应链安全风险、产品商业化不及预期风险、性能与生态不及宣传风险、盈利预测假设不及预期风险、可比公司估值局限性提示、国内外公司不具备完全可比性及对标资料局限性风险、非经常性损益不可持续及扣非持续亏损风险、估值过高及股价高波动风险、客户集中度风险、流通盘小/换手率高/解禁波动风险、模型Day 0适配不等于订单及收入风险、2026H1归母盈利主要来自公允价值变动收益且扣非仍亏损风险、技术迭代风险、地缘政治及供应链受限风险。

下载日志已记录，仅供内部参考

预测指标

预测指标	2025A	2026E	2027E	2028E
营业收入（百万元）	1644	3888	8763	19951
增长率(%)	121	136	125	128
归母净利润（百万元）	-789	323	2390	7368
增长率(%)	—	—	641	208
摊薄每股收益（元）	-1.97	0.81	5.97	18.41
ROE(%)	-6	2	14	30
P/E	—	616.56	83.20	26.99
P/B	17.63	13.71	11.78	8.20
P/S	141.17	51.16	22.70	9.97
EV/EBITDA	—	363.62	71.78	24.98

资料来源：Wind资讯、国海证券研究所

注：亏损或扭亏口径不适用同比测算

下载日志已记录，仅供内部参考

一、公司概览：

高性能GPU提供商，产品迭代持续提速

下载日志已记录，仅供内部参考

1.1公司概况：高性能GPU/智算方案提供商，面向云/端训推及图形渲染全场景

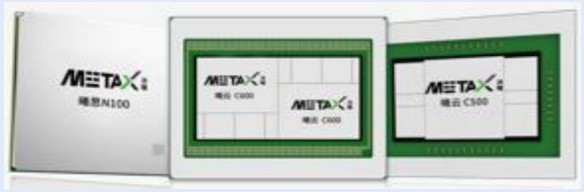
◆**沐曦股份是国内领先的高性能GPU设计企业及智算解决方案提供商。**公司以自主研发的**“曦云C系列、曦思N系列、曦彩G系列、曦索X系列”**全栈GPU芯片为核心，配合自研**“MXMACA软件栈”**，分别面向**云侧AI大模型训推、云/端AI推理以及图形渲染、科学智能**等场景，提供国产算力基座。

◆截至2025年底，公司GPU产品累计销量超5.5万颗，并批量部署于**国家级智算平台、运营商智算平台及商业智算中心**等10余个智算集群，应用领域覆盖金融、能源、教科研、医疗健康、大文娱等多个行业。

图：公司业务体系

核心硬件：高性能GPU

- 曦云 C500/C600：云侧训推一体
- 曦思 N100/N260/N300：云/端推理
- 曦彩 G100：图形渲染
- 曦索X206/X302：科学智能



应用部署

国家级智算平台、运营商智算中心、商业化智算中心等

产品交付：
支持单卡/模组、服务器/超节点、智算集群三类形式







- PCIe/OAM模组
- 服务器/一体机/超节点
- 集群

技术基座

高性能GPU架构+MXMACA软件栈

下载日志已记录，仅供内部参考

1.2 历史沿革：成立后研发/商业化推进顺利，产品迭代持续提速

◆ 2020年9月，沐曦股份由陈维良创立，与联合创始人彭莉、杨建均来自AMD。公司成立后，资本与团队双轮驱动下，**产品研发与商业化快速推进：先后完成曦思N100、曦云C500和C600的流片与量产；并于2025年12月完成科创板上市。**上市后，公司持续推动**下一代产品曦云C700、曦索X206等的研发及产业化落地，产品迭代节奏延续。**

图：公司历史沿革

2020年：创立与起步

公司于2020年9月由陈维良创立，与联合创始人彭莉、杨建均来自AMD。成立仅两个月，公司即完成天使轮融资，由和利资本与泰达科投共同投资。

2021年：融资与研发开启

2021年初，公司完成了Pre-A轮及Pre-A+轮融资，红杉资本、真格基金、经纬创投、光速中国等风险投资机构开始入场。**并于上半年，先后完成推理GPU曦思N100系列和训推一体GPU曦云C500系列的研发立项。**

2022-2023年：流片/量产与商业化

- 2022年1月，首款推理GPU 曦思N100交付流片。
- 2023年4月，推理GPU 曦思N100正式量产；6月首款训推一体GPU 曦云C500回片；公司全年实现营收0.53亿元。

2024-2025年：商业化加速与上市

- 2024年2月，训推一体GPU **曦云C500正式量产；**公司全年营收攀升至7.43亿元。
- 2025年7月，训推一体GPU **曦云C600回片并成功点亮；**12月，公司完成科创板挂牌上市。

2026年：下一代产品研发与落地

2026年，公司布局并推动下一代产品曦云C700 、曦索 X206等的研发及产业化落地。**其中曦云C700预计于年内开启流片测试，2027年开启小批量产。**

2020 — 2021 — 2022-2023 — 2024-2025 — 2026

资料来源：财经网、智东西、沐曦股份招股书、铅笔道公众号、澎湃新闻、同花顺、36Kr、国海证券研究所

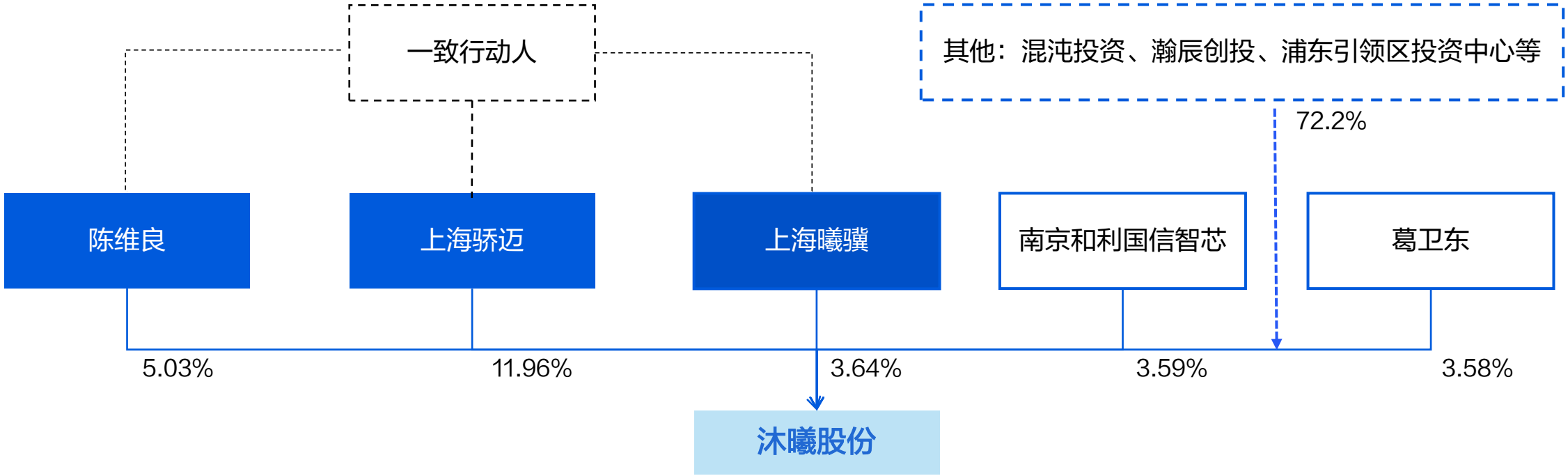
下载日志已记录，仅供内部参考

1.3 股权架构：股权结构清晰，获国家资本支持

◆**股权结构清晰，创始人把握控制权：**截至2026年二季度末，公司创始人/董事长陈维良先生通过直接持股5.03%以及与上海骄迈、上海曦骥的一致行动关系，合计控制公司约20.63%的股权，对公司战略决策具有决定性影响。

◆**获国家资本支持：**上海浦东引领区投资中心持有公司2.41%的股份，中国国有企业结构调整基金股份有限公司持有1.76%的股份，为公司提供重要背书。

图：股权架构（截至2026年二季度末）



下载日志已记录，仅供内部参考

1.4 管理层：创始团队来自AMD，曾负责其多款GPU设计量产

◆ **公司管理层具备深厚技术背景，创始团队曾于AMD负责多款GPU设计量产：**创始人/董事长/总经理**陈维良先生曾长期担任AMD全球GPGPU设计总负责人，主导AMD 15款高性能GPU产品的流片与量产；**联合创始人/CTO**彭莉为芯片架构资深专家，历任AMD首席SoC架构师、系统架构师、GFXIP架构师；**联合创始人/CTO杨建**负责MXMACA软件栈**，为AMD大中华区首位华人科学家，曾任华为海思首席架构师。

表：管理层简介

职务	姓名	简介
创始人/董事长/总经理	陈维良	创始人**陈维良曾在AMD任职17年，担任过其全球GPGPU设计总负责人**，全面负责全球通用计算GPU产品线的整体设计与管理，**并主导了AMD共15款高性能GPU产品的流片与量产**
联合创始人/副总经理/CTO	彭莉	**公司首席硬件架构师，负责芯片架构设计；曾于AMD工作13年**，为AMD全球首位华人女科学家，历任AMD首席SoC架构师、系统架构师、GFXIP架构师，主导多款GPU产品的架构设计与量产
联合创始人/副总经理/CTO	杨建	**公司首席软件架构师**，负责MXMACA软件栈；**曾于AMD工作14年，为AMD大中华区首位华人科学家，主导或参与AMD数十款高性能 GPU 产品量产；**历任华为海思首席架构师
核心技术人员/董事	王爽	负责产品研发管理；先后就职于中兴微电子、华为海思，历任研发工程师、研发经理等

资料来源：沐曦股份招股书、澎湃新闻、二十一世纪经济报道、中国矿业大学官网、上海科技大学信息与技术学院官网、财金职业圈公众号、江苏省睢宁高级中学公众号、南京理工大学公众号、国海证券研究所

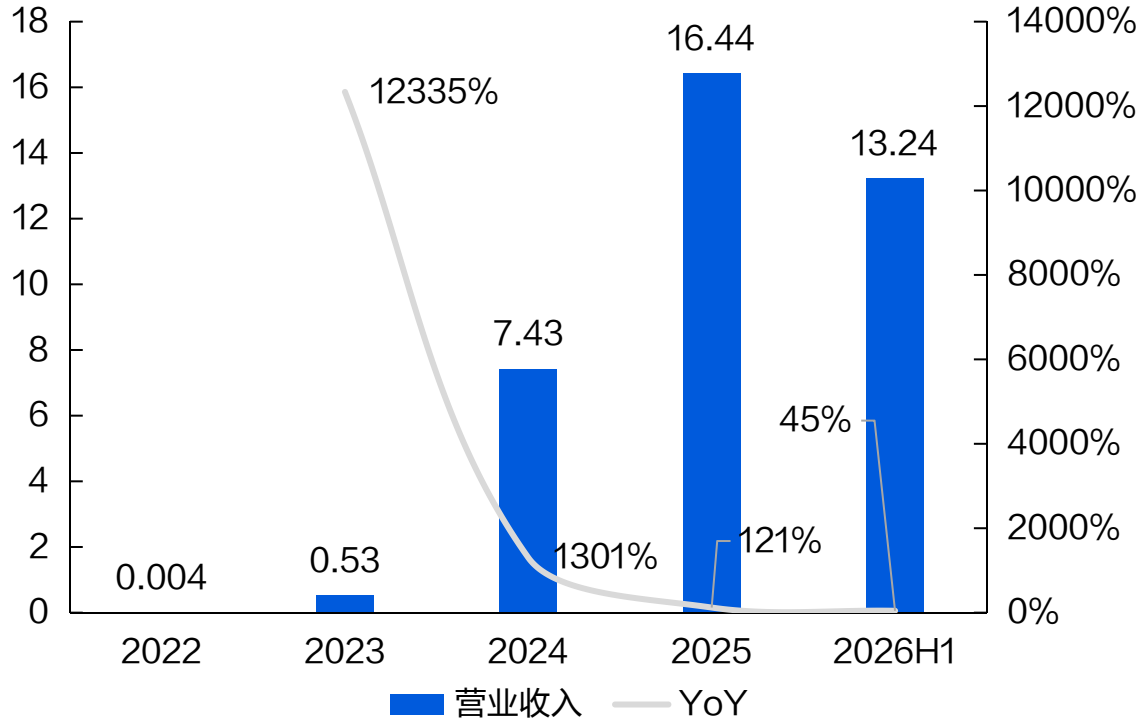
下载日志已记录，仅供内部参考

1.5 公司业绩：营收规模快速增长，曦云C600有望成为下一代收入主力

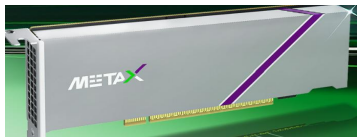
◆**自2023年4月智算推理GPU曦思N100和2024年2月训推一体GPU曦云C500量产以来，公司营收规模快速增长。**2022年至2025年，分别实现营业收入0.004亿元、0.53亿元、7.43亿元和16.44亿元，同比增长12335%、1301.46%和121.26%。

◆**2026年上半年，受益于出货量提升及客户采购规模扩大，**公司实现营收13.24亿元，同比增长44.67%，延续高增态势；**同时曦云C600已于2026年5月成功量产，有望成为下一代收入主力。**

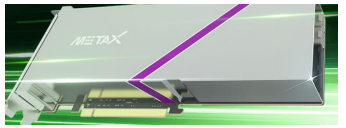
图：公司2022-2026H1营收（单位：亿元，%）



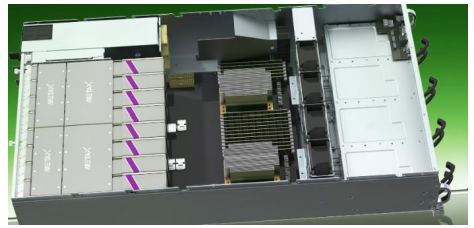
图：公司部分已量产销售产品



• 曦思 N100 PCIe板卡



• 曦云 C500 PCIe板卡



• 曦云 C500 训推一体服务器

➢ 8卡/16卡两类配置规格

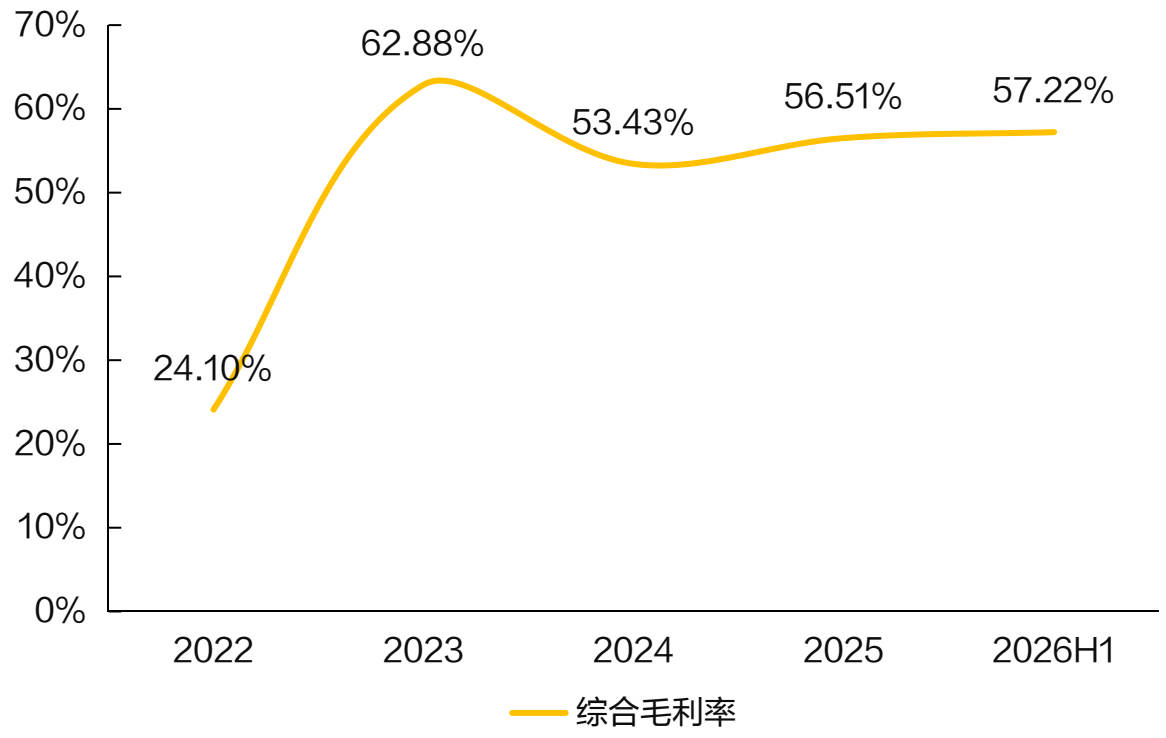
➢ 支持国内外主流CPU

下载日志已记录，仅供内部参考

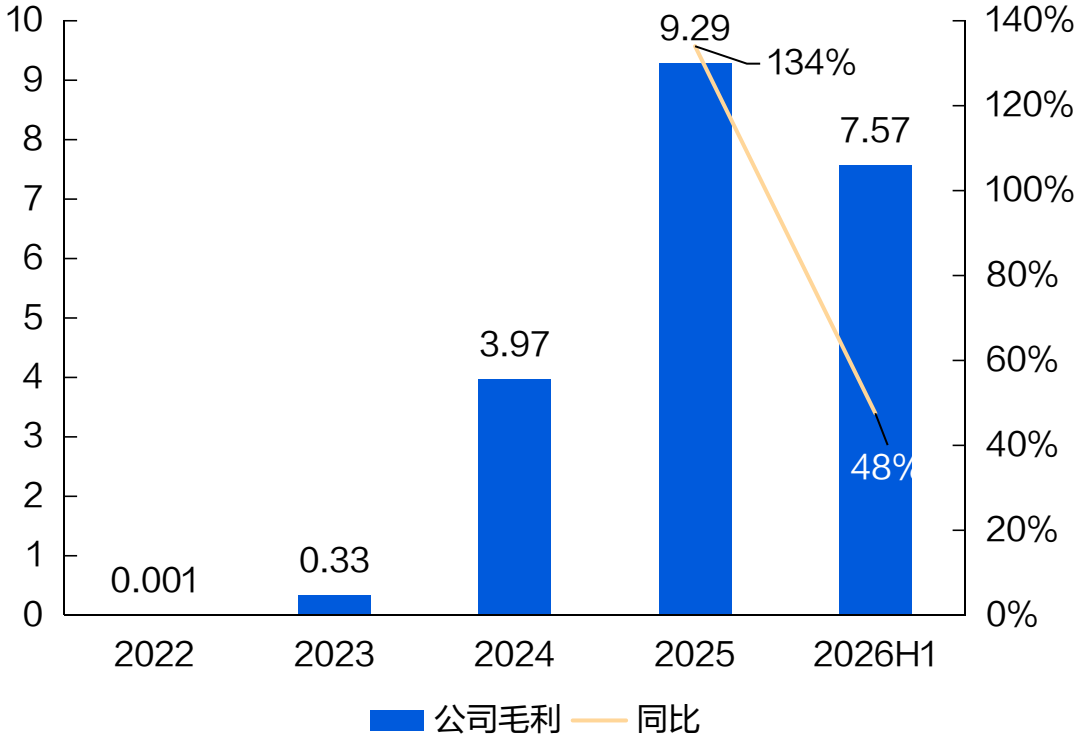
1.6 公司业绩：大客户导入后毛利率仍维持较高水平

◆**大客户导入后毛利率仍维持较高水平：**2023年至2025年，公司综合毛利率分别为62.88%、53.43%、56.51%。2024年后，公司为加速市场导入对大客户给予价格优惠，综合毛利率较2023年高点有所下调，但仍维持较高水平。2026H1综合毛利率57.22%，同比提升1.14pct，毛利率进一步正向改善。

图：公司综合毛利率（%）



图：公司整体毛利及同比（亿元，%）



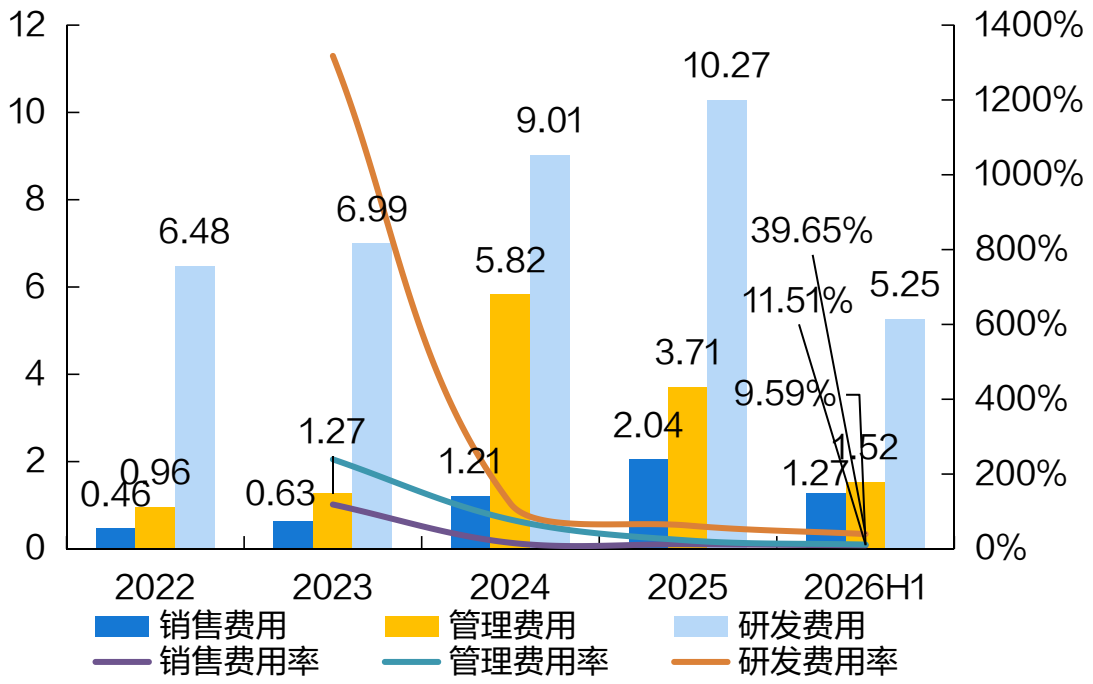
下载日志已记录，仅供内部参考

1.7 公司业绩：费用扩张趋于稳定，收入增长有望驱动利润转化

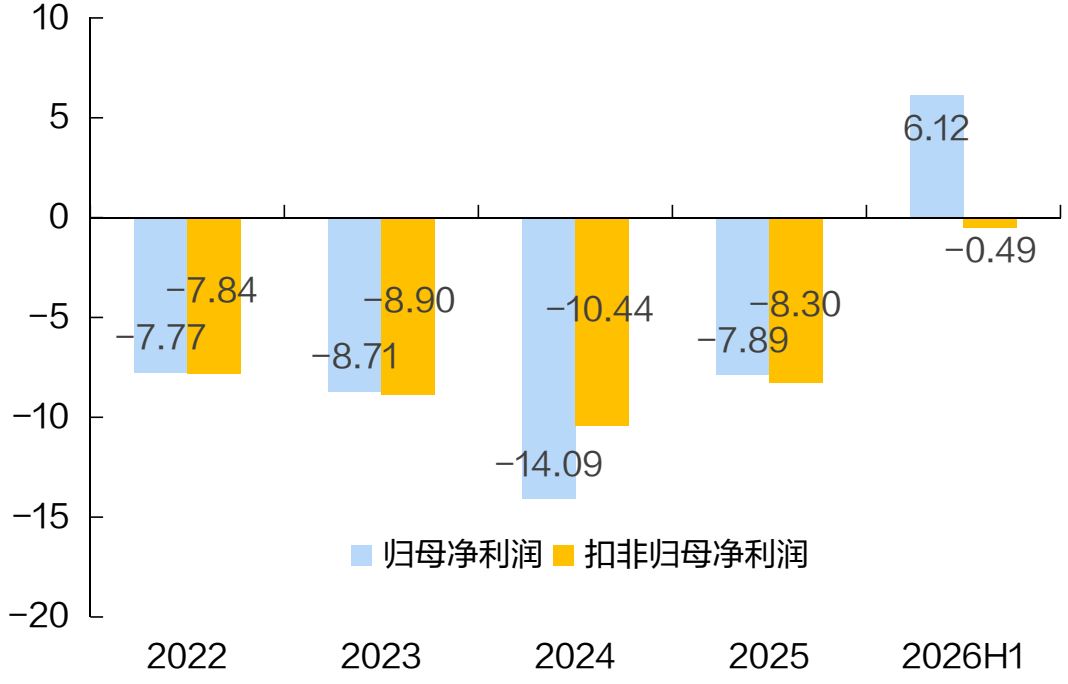
◆ **公司费用扩张趋于稳定，费率随营收高增快速摊薄。**2026年上半年，销售、管理、研发费用率分别降至9.59%、11.51%和39.65%，规模效应逐步显现。同时公司指出未来将持续进行费用控制，为利润释放创造更大弹性空间。

◆ **收入增长有望驱动利润转化。**公司归母净利润亏损幅度于2024年后明显收窄；2026年上半年实现归母净利润6.12亿元，得益于交易性金融资产公允价值变动收益8.87亿元；扣非归母净利润亦大幅收窄至-0.49亿元。**我们认为，随着核心产品放量及费用趋稳，收入增长有望驱动利润转化。**

图：销售、管理、研发费用及费用率（单位：亿元，%）



图：归母净利润/扣非归母净利润（亿元）



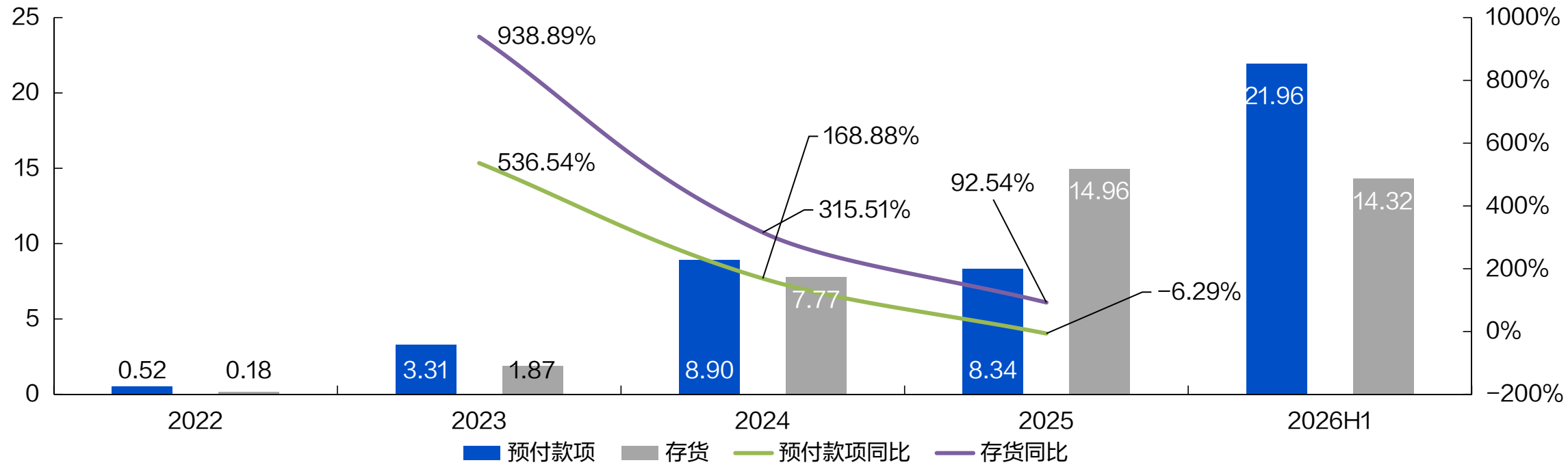
下载日志已记录，仅供内部参考

1.8 公司业绩：预付款项/存货维持较高水平，有望为后续收入转化提供支撑

◆公司预付款项及存货维持高位：

- 截至2026年上半年末，预付款项达21.96亿元，较2025年末的8.34亿元增长163.24%，反映公司为保障产能供应而提前备货，未来业绩有望进一步实现增长。
- 存货方面，截至2026年H1末达14.32亿元，较2025年末下降4.28%，其中主要为原材料从8.02亿元降至3.40亿元，而半成品及发出商品仍保持高位，我们认为有望为后续收入转化提供较强支撑。

图：公司预付款项、存货（单位：亿元；预付款项/存货对应左轴，同比对应右轴）



下载日志已记录，仅供内部参考

二、曦云/曦思/曦彩/曦索 GPU：
基于统一GPU计算架构，覆盖云训推/图形渲染/科学智能

下载日志已记录，仅供内部参考

2.1 GPU产品矩阵：推出曦思N/曦云C/曦彩G/曦索X四大系列GPU

◆ 基于自研GPU指令集、SoC及高速互联等技术，**公司以统一的GPU计算/渲染架构，先后推出了智算推理曦思N系列、训推一体曦云C系列、图形渲染曦彩G系列以及科学计算曦索X四大系列GPU：**

➢ **曦思N系列**搭载**多精度混合算力，配备大容量显存及较高能效比，适配云侧及边缘侧推理场景；曦云C系列**在此基础上**升级运算核心密度，强化并行计算能力，适配密集向量与矩阵运算，用于云端训推一体；**

➢ **曦彩G系列**则在通用GPU架构上**增加图形处理单元，**实现对图形渲染等任务的支持；**曦索X系列**面向科学智能，具备全精度混合算力与大容量高带宽显存，**为材料科学、生命科学等AI4S前沿研究提供算力支撑。**

图：公司自研全栈GPU产品

曦思N系列

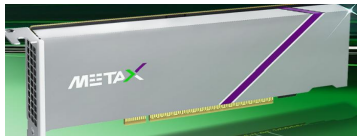
云及边缘推理



• 曦思 N300 PCIe板卡



• 曦思 N260 PCIe板卡



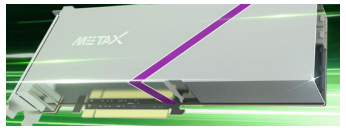
• 曦思 N100 PCIe板卡

曦云C系列

云训推一体



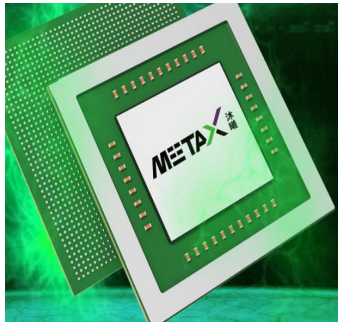
• 曦云 C600 OAM模组



• 曦云 C500 PCIe板卡

曦彩G系列

图形渲染



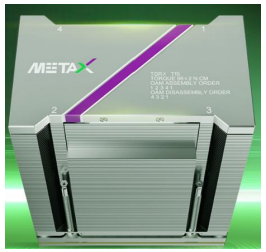
• 曦彩 G100 GPU

曦索X系列

科学智能



• 曦索 X302 OAM模组



• 曦索 X206 OAM模组

资料来源：公司官网、国海证券研究所

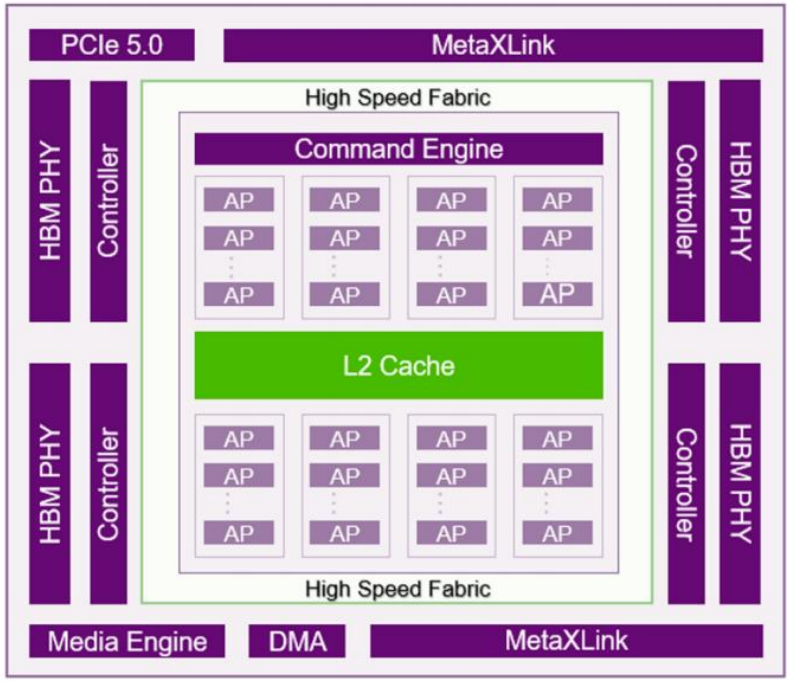
下载日志已记录，仅供内部参考

2.2 曦云系列GPU：高计算密度与存储/互联能力，面向云端训推一体

◆ 为满足云端AI训推场景下大规模数据密集计算与互联通信的需求，**曦云C系列GPU基于自主研发的XCORE架构及指令集，集成运算单元、高速缓存、控制单元等多元组件，具备高计算密度与存储/互联能力：**

- **计算能力方面，**曦云GPU XCORE架构内部包含标量计算单元、矢量计算单元和张量计算单元三类核心计算单元，**单元密度较高，同时可高效支持FP64、FP32、TF32、FP16、FP8等多种混合精度。**
- **存储与互联方面，曦云GPU配备大容量显存与高显存带宽，有效缓解大规模模型训练中的“内存墙”瓶颈；同时支持卡间高速互联，并提供丰富接口以兼容多种通信协议，**有助于多卡扩展场景下的数据传输效率与可扩展性。

表：训推一体GPU——曦云C系列



曦云 C500	曦云 C600	曦云 C700
		
➢ 2023年6月回片 ➢ 2024年2月量产	➢ 2025年7月回片/发布 ➢ 2026年5月量产	➢ 2025年4月立项 ➢ 预计2027年Q3开启小批次量产

下载日志已记录，仅供内部参考

2.2 曦云系列GPU：C500综合性能对标英伟达A100，C600于2026年5月量产

◆ **曦云C500**为上一代训推一体GPU主力产品，于2024年2月正式量产。该产品基于自研XCORE 1.0架构及指令集，**单卡FP16算力达280 TFLOPS，配备64GB HBM2e显存及1.8TB/s带宽，**并可通过MetaXLink实现多卡全互联。根据电子发烧友网Elecfans公众号，**其综合性能对标英伟达A100，且相较H20在大模型训练场景下更具性能优势。**

◆ **曦云C600**于2025年7月回片点亮，**并于2026年5月正式量产，有望成为下一代主力产品。**相较于上一代C500，**C600在算力上有较大提升，同时搭载HBM3e显存，并在精度支持与指令集方面引入新技术，**新增FP8 Tensor及Tensor转置指令，进一步强化了对大规模AI训练任务的支撑能力。

表：训推一体GPU——曦云C系列

规格	曦云C500	曦云C600
工艺架构	自研XCORE1.0 7nm	自研XCORE1.5
FP16算力	280 TFLOPS	/
INT8算力	560 TOPS	/
显存	64GB HBM2e 带宽1.8TB/s	HBM3e
视频编解码	160路解码/12路编码	/
计算精度	支持多精度计算	支持多精度计算，新增支持FP8 Tensor指令
互联	支持卡间高速互联	支持卡间高速互联
软件生态	MXMACA 兼容CUDA/主流AI框架	MXMACA 兼容CUDA/主流AI框架

资料来源：沐曦股份招股书、XG云数智公众号、国海证券研究所

下载日志已记录，仅供内部参考

2.3 曦思系列GPU：推出N100/260/300，覆盖传统人工智能及云/端AI推理

◆ **曦思系列为公司智算推理GPU产品，已推出曦思N100、曦思N260及曦思N300共三个型号，覆盖从传统人工智能场景到云/端AI推理的差异化需求：**

- 其中，**曦思N100面向传统人工智能场景，**内置高性能视频处理器与运算核心，可广泛应用于智慧城市、智慧交通、智慧教育、智能视频处理等领域，**用于人脸识别、图像识别、视频编解码等典型任务；**
- 曦思N260与N300作为后续迭代产品，**拥有多精度混合算力，配备大容量显存及新一代高速I/O接口。**N260可用于搭载DeepSeek等主流大模型的一体机、液冷工作站等，**N300则进一步适用于云端智算中心的AI推理计算任务。**

表：智算推理GPU——曦思N系列

规格	曦思N100	曦思N260	曦思N300
架构	自研XCORE 0.5架构	自研XCORE 1.0架构	自研XCORE 1.5架构
计算精度	支持多种混合精度计算	支持多种混合精度计算	支持多种混合精度计算
显存	HBM2e显存，容量16GB	HBM2e显存，容量64GB	HBM3显存
视频编解码	集成视频图像编解码硬件单元，支持H.264/H.265/AV1/AVS2硬件解码	/	/
目标场景	传统人工智能场景	/	云端智算中心
典型应用	人脸识别、图像识别、视频编解码	搭载DeepSeek等主流大模型的一体机、液冷工作站等	AI推理任务

下载日志已记录，仅供内部参考

2.4 曦彩/曦索系列GPU：G100完成IP设计/验证，X206/302具备全精度算力/大显存

◆ 公司产品矩阵中还包括面向图形渲染的曦彩GPU和面向科学智能的曦索GPU：

- **曦彩GPU主要型号为G100**，其基于自研XCORE2.0架构，内置几何引擎、光栅化单元、像素处理单元、纹理单元及光线追踪单元等，可用于云渲染、游戏、数字孪生、影视动画制作、专业制图等场景。**目前产品已完成GPU IP设计与验证，进入后续开发阶段。**
- 曦索GPU主要型号则为X206和X302等，主要面向科学智能。该系列**具备全精度混合算力，搭载大容量高带宽显存，同时支持高速多卡互连，具有较强的并行计算能力与数据吞吐效率，**可高效支撑材料科学、生命科学等前沿科学计算任务。

表：图形渲染GPU及科学智能GPU——曦彩G系列、曦索X系列

类别	产品型号	产品描述	应用场景
图形渲染GPU	曦彩 G100	曦彩G系列GPU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	云侧及边缘侧图形处理
科学智能GPU	曦索 X206	曦索X206、X302等具备全精度混合算力，搭载超大容量高带宽显存，支持高速多卡互连，具有较强的并行计算能力与数据吞吐效率，可高效支撑经典科学计算任务，广泛赋能AI4S驱动的前沿科研范式	云侧科学计算等
	曦索 X302		

下载日志已记录，仅供内部参考

三、互联/集群与软件生态：MetaXLink带宽对标H200，软件栈较好兼容CUDA

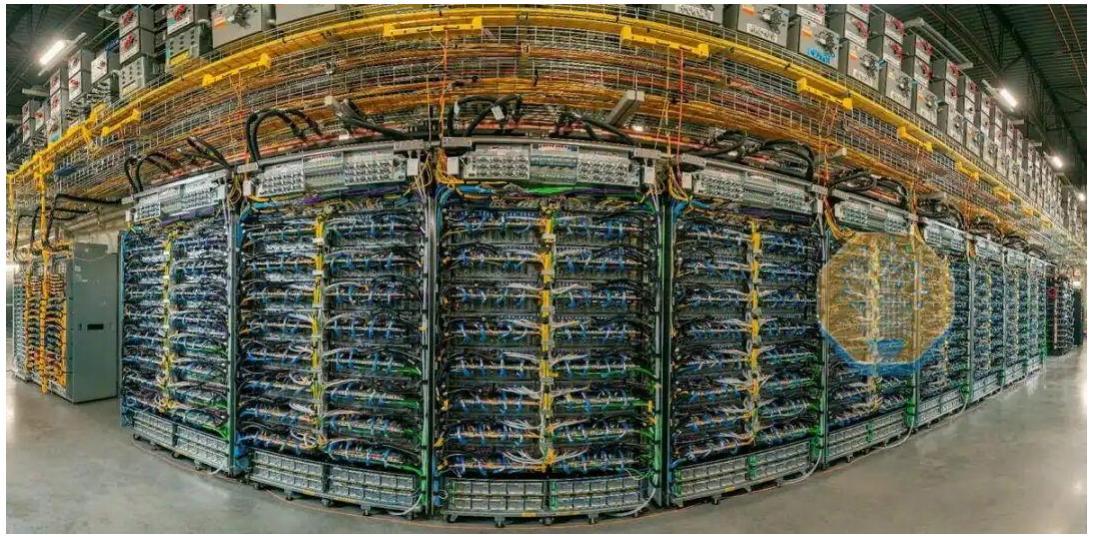
下载日志已记录，仅供内部参考

3.1 互联/集群：集群互联为主流AI训推选择，卡间高速互联/网络通信作用关键

◆**模型规模增长之下，为承载大规模训练与推理任务，包括海外大厂等在内的主流厂商普遍采用分布式并行计算策略，即通过集群互联来满足海量数据与多任务的处理需求。**构建路径为：增加单台服务器内部的GPU数量形成超节点和增加服务器数量来构建大规模计算集群。**在此过程中，卡间高速互联技术与网络通信技术发挥着关键作用，是集群有效算力的核心支撑。**

➢ Google、Microsoft、Meta等海外大厂已纷纷布局大集群超级计算机，如Google基于自研芯片搭建TPUv7集群，微软于威斯康辛州建设十万卡英伟达GB200集群，通过大规模服务器组成集群计算优势。

图：谷歌TPUv7 集群



图：微软十万卡GB200集群



资料来源：热设计公众号、半导体行业观察公众号

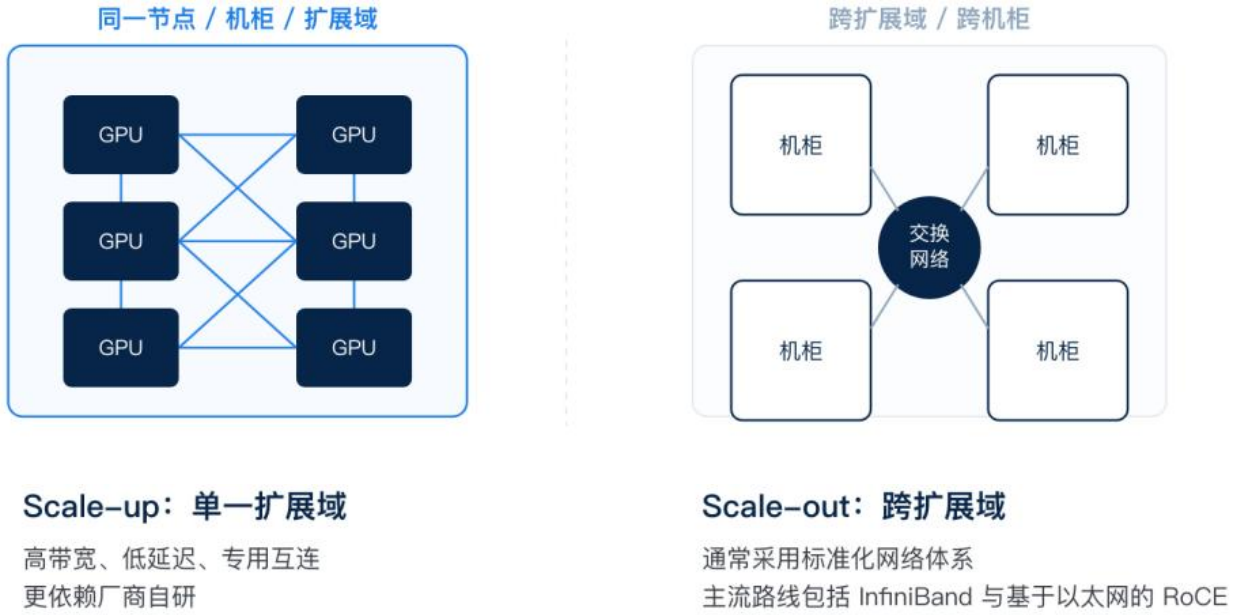
下载日志已记录，仅供内部参考

3.1 互联/集群：公司自研MetaXLink卡间高速互联协议，带宽对标NV H200

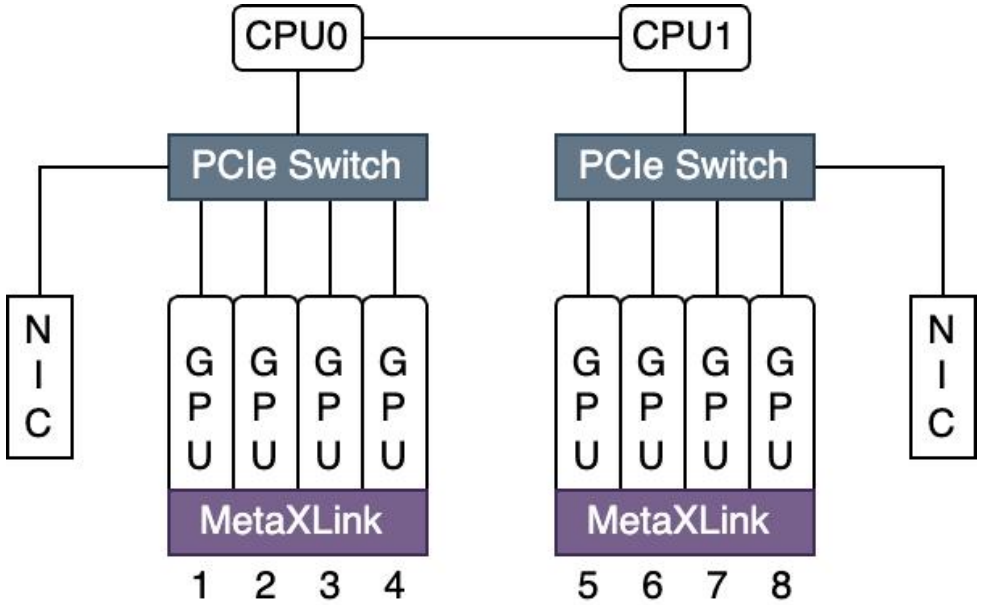
◆**公司自研卡间高速互连技术MetaXLink**，其根据GPU与GPU之间的数据交互需求优化了传输协议，并支持MetaXLink端口之间数据直接转发，可为大规模AI模型训推提供充足通信带宽保障；根据公司招股意向书，**其互连带宽已达到与英伟达H200相当的性能水平。**

◆基于MetaXLink互连，**公司训推一体GPU曦云C500与C600均配置了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**，支持千亿参数以上规模AI大模型训练算力集群的建设。

图：Scale-up单一扩展域内互联依赖厂商自研



图：沐曦MetaXLink示意图



下载日志已记录，仅供内部参考

3.1 互联/集群：可商业化落地千卡集群，进一步发布64卡曦景S600超节点机柜

◆ 集群落地方面，国外头部企业已能够支持万卡乃至十万卡级超大规模集群，**国内目前仍处于以千卡规模为主的阶段。公司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，并正积极推进万卡级集群的研发与落地：**

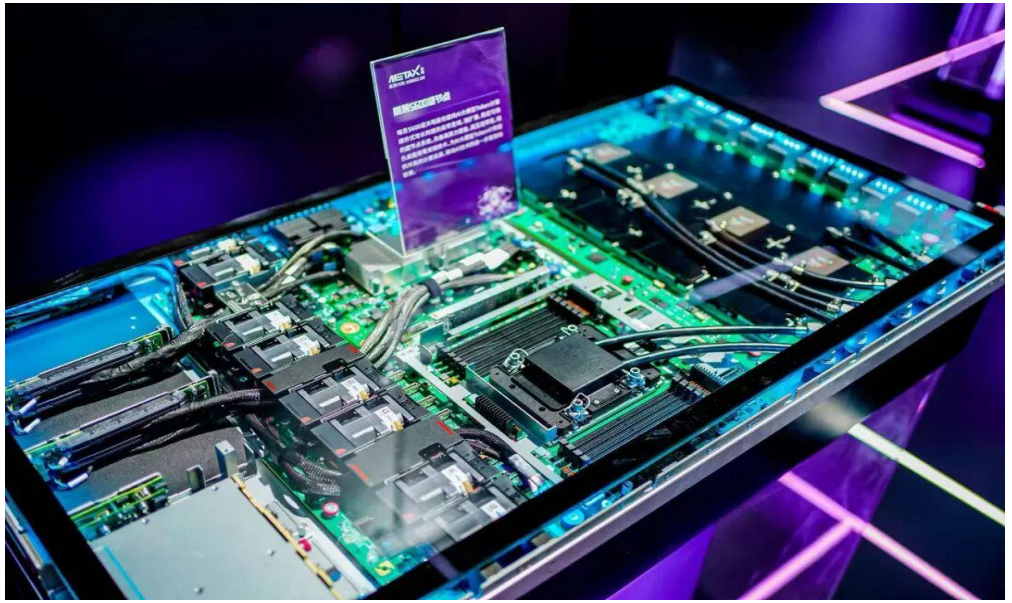
➢ 为进一步提升大规模集群的部署能力，公司于2026年WAIC大会上发布了新一代AI超节点产品曦景S600。该产品采用OEX架构设计，**通过正交连接器与两级交换拓扑，实现单机柜64张GPU高速全互连，**并支持灵活扩展至万卡级集群。

➢ 在物理设计上，计算节点、交换节点以及二者之间均实现**零线缆连接，大幅简化了系统布线复杂度。**同时，产品采用**解耦模块化设计，计算、交换、电源及液冷等关键模块均支持快速部署与维护，**提升了集群运维效率与可扩展性。

图：曦景S600超节点机柜



图：曦景S600超节点机柜内部结构



下载日志已记录，仅供内部参考

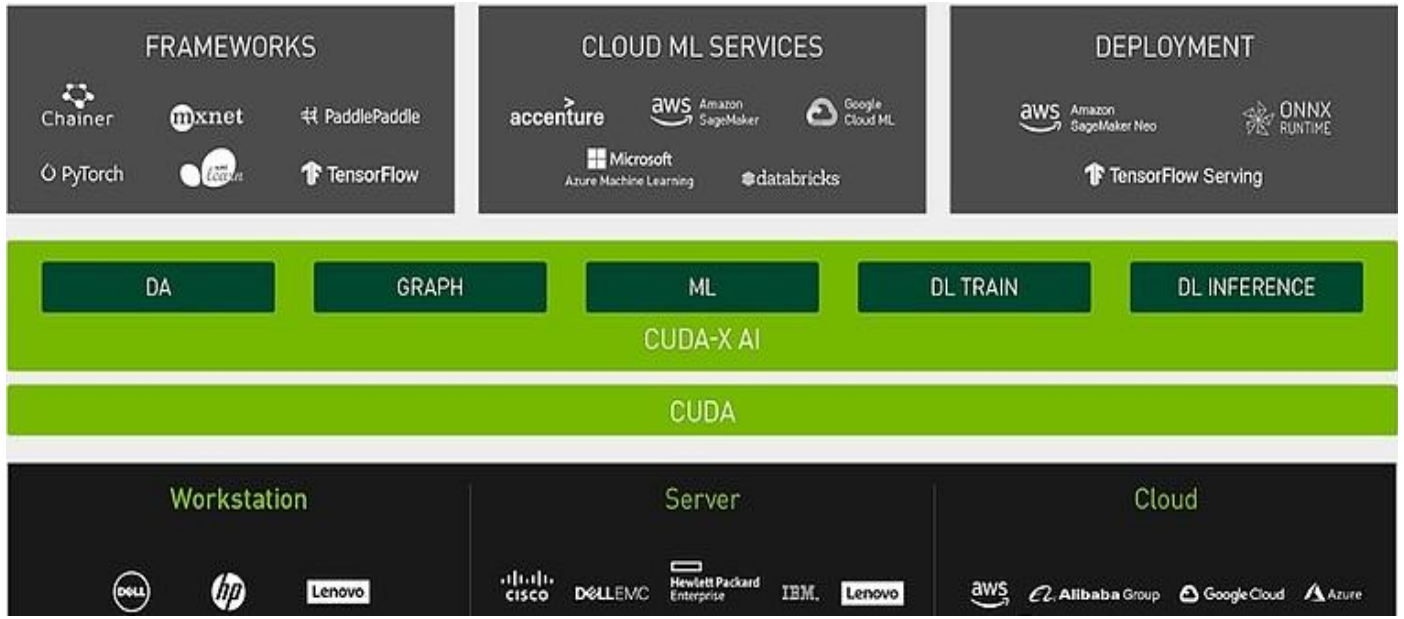
3.2 软件生态：CUDA生态深厚，国产GPU兼容是提高产品可用性的关键路径

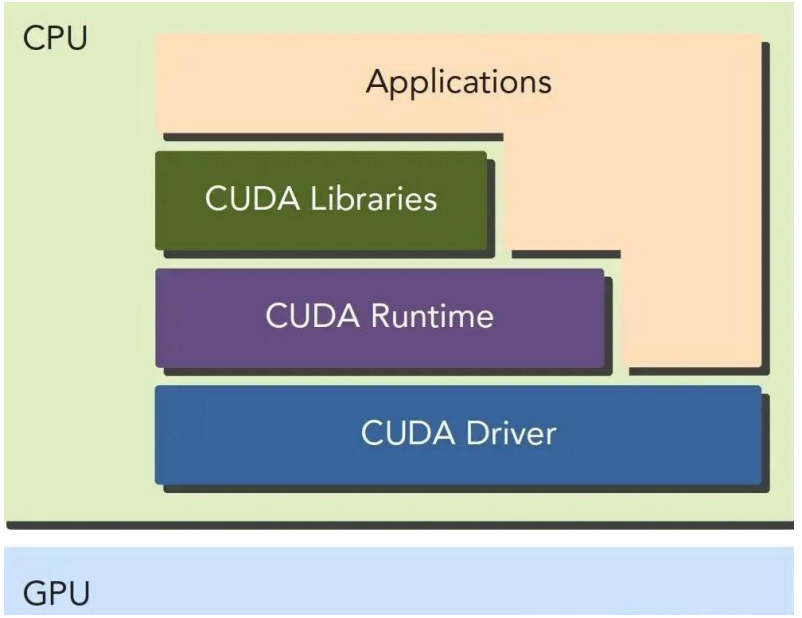
◆ **软件生态是GPU易用性的核心，主要由上层AI框架与算法应用、中层通信库与数学库及各类工具链、底层编译器与驱动程序等构成：**

➢ **英伟达于2006年推出的CUDA计算平台，**凭借完整的工具链与庞大的内部组件，构建了成熟的软件生态。**当前AI计算的应用生态大多建立在CUDA之上，其软件库覆盖率、AI框架及算子库的支持度已较为完善，在AI领域占据绝大部分市场份额。**

➢ 对于国产GPU厂商而言，自建完整生态需要较长时间才能完成工具链完善与开发者迁移。**保持对CUDA生态的较好兼容，**能够有效实现软件层的自主可控与平稳过渡，降低迁移成本、提高迁移效率，我们认为其是**提升自身产品易用性以及客户使用意愿的选择。**

图：英伟达CUDA生态





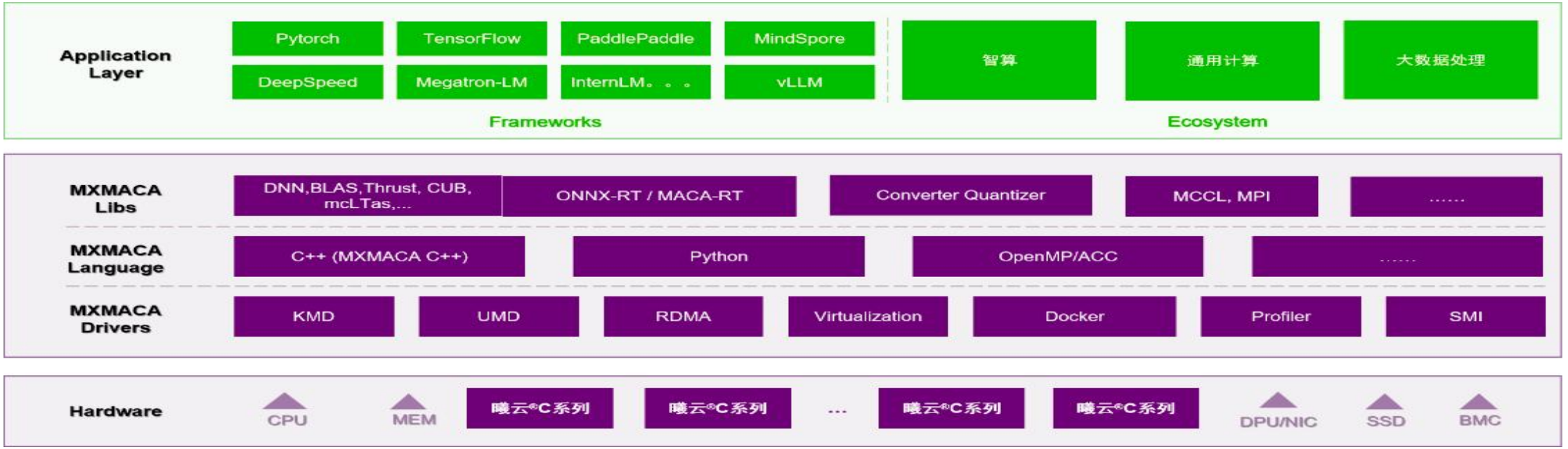
下载日志已记录，仅供内部参考

3.2 软件生态：自研MXMACA软件全栈平台，支持超6000个CUDA应用

◆ MXMACA是公司为自有GPU产品打造的统一、完整的全栈式软件工具链，**分为算法应用层、数学库及通信库层、编程语言层以及软件驱动层四个层级，**涵盖应用开发、功能调试和性能调优等核心环节，为客户提供了从底层驱动到上层应用的完整开发环境。

◆ **在API层面，MXMACA软件栈实现了对CUDA生态的较好兼容。**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GPU产品上运行，大幅降低了客户的应用迁移成本、提升了迁移效率。据公司招股书，截至2025年10月22日，**MXMACA已支持超过6000个CUDA应用及超过2200个高性能算子，MXMACA软件栈注册用户及实际使用客户合计超15万人。**

图：MXMACA软件栈框架结构



下载日志已记录，仅供内部参考

四、公司优势：

覆盖国家/运营商/互联网客户，较早启动国产供应链布局

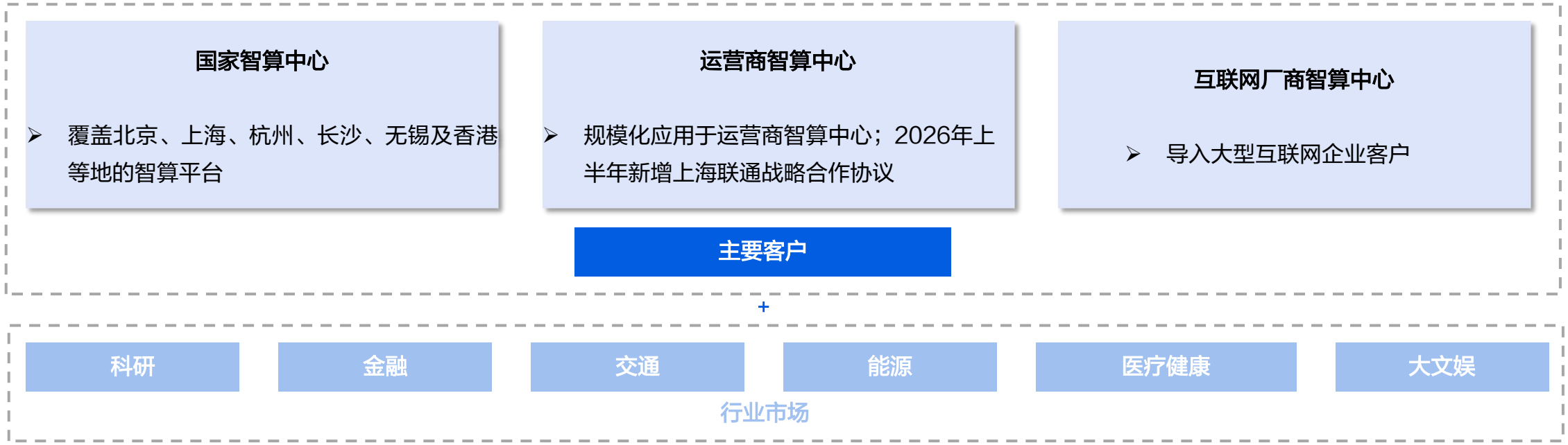
下载日志已记录，仅供内部参考

4.1 客户优势：主要覆盖国家/运营商/互联网智算中心，同步卡位垂类行业市场

◆ **公司客户结构多元，主要基本盘为国家智算中心、运营商智算中心及互联网厂商智算中心：在国家智算中心方面，**产品覆盖北京、上海、杭州、长沙、无锡及香港等地的智算平台。其中**总建设投入25亿元的无锡万卡集群**于2026年上半年实现一期点亮。**在运营商智算中心方面，**公司上半年新增与上海联通签署战略合作协议，客户基础与部署规模持续扩大。

◆ **除基本盘外，公司积极布局科研、金融、交通、能源、医疗健康、文娱六大行业市场，发展垂类场景的销售机会：**2026年上半年，与江南大学开展合作，参与**清华大学千卡GPU集群建设，并与迈富时签署战略合作协议，围绕算力基础设施、**企业级智能体中台及行业场景落地开展深度协作。

图：公司客户结构



下载日志已记录，仅供内部参考

4.2 模型适配：支持万亿参数/长上下文LLM以及多模态生成模型Day0适配

◆ **自2025年底至今，公司完成37个主流旗舰模型的Day 0适配，**覆盖智谱AI、阿里千问、MiniMax、DeepSeek、阶跃星辰、腾讯混元、百度、美团等模型厂商，**支撑包括通用语言大模型、多模态生成、OCR、机器翻译、文生图等领域推理需求，其中：**

➢ **万亿参数/长上下文模型：对GPU算力吞吐、显存容量、长文本推理及多卡协同调度能力要求高。**曦云C系列GPU**可实现对万亿参数模型的超大规模及1M长上下文推理需求支撑**，其Day 0适配序列中已包括智谱及腾讯混元的旗舰通用模型

➢ **多模态模型：凭借多精度混合算力与高带宽显存架构，可高效支撑图像、视频、音频等多模态数据的并行处理与融合推理。**以MiniMax H3为例，公司已完成对其的Day 0适配，可支撑其文本、图片、音频和视频等多种输入形式，以及15秒2K分辨率视频的输出能力。

表：与万亿参数/长上下文LLM以及多模态生成模型Day0适配

模型类型	适配要求	Day0适配序列
万亿参数/ 长上下文模型	要求GPU具备高算力吞吐、显存容量，同时能较好实现多卡调度协同	智谱GLM5.3/5.2/5.1 腾讯混元Hy4 preview/Hy3/Hy3 preview 阿里Qwen3.5系列模型 MiniMax M2.7等
多模态模型	面向图像、视频、音频等不同模态数据的并行处理及融合推理需求，要求具备混合算力及高带宽显存架构	MiniMax H3/M3 阿里Qwen-Image-2.1 百度文心ERNIE-Image文生图等

下载日志已记录，仅供内部参考

4.3 产业优势：较早启动国产化供应链布局，三年内实现两款芯片量产

◆**公司较早启动国产供应链布局，提前规划并设计了基于国产供应链的产品；据公司披露，其在晶圆生产制造、封装测试、存储颗粒、EDA、IP等核心环节的国产化上均已取得一定突破，**有助于产品可持续的供应与交付。同时拥有动态、灵活的库存管理机制，供应链管理经验丰富。

◆得益于稳定的交付能力与高效的研发能力，根据公司招股意向书，其仅用3年时间实现两款芯片一次流片并成功量产，快速完成了从产品验证到商业化、规模化落地的跨越；根据公司披露，**截至2025年末，公司GPU累计销量超过5.5万颗；2026年上半年出货量同比继续显著增长。**

图：半导体供应链



资料来源：沐曦股份招股书、国海证券研究所

下载日志已记录，仅供内部参考

五、投资建议与风险提示

下载日志已记录，仅供内部参考

5.1 盈利预测

- 公司是国内领先的高性能GPU芯片设计公司，在大模型算力需求增长和国产替代趋势下有望充分受益；其中训推一体系列有望成为主要增长动力。假设训推一体C系列GPU上线进度及客户拓展顺利，我们预计2026-2028年公司营业收入为38.88/87.63/199.51亿元，同比+136.46%、+125.41%、+127.68%。

表：沐曦股份分业务盈利预测（百万元）

	2024	2025	2026E	2027E	2028E
营业收入	743.07	1,644.09	3,887.54	8,762.72	19,951.34
增长率%	-	121.26%	136.46%	125.41%	127.68%
毛利率	53.43%	56.51%	55.25%	55.54%	55.50%
分业务					
训推一体系列					
收入	721.74	1,617.05	3,851.81	8,712.79	19,882.58
增长率%	-	124.05%	138.20%	126.20%	128.20%
智算推理系列					
收入	3.49	4.02	5.40	10.06	16.12
增长率%	-	15.20%	34.22%	86.30%	60.26%
IP授权					
收入	1.00	1.10	1.26	1.44	1.65
增长率%	-	10.20%	14.00%	15.00%	14.50%
其他					
收入	16.85	21.92	29.08	38.43	50.99
增长率%	-	-	32.66%	32.15%	32.68%

下载日志已记录，仅供内部参考

5.2 投资建议

- **公司主要面向头部行业客户/运营商/国家智算中心等提供GPU及解决方案,**与寒武纪、摩尔线程的主营业务较为类似，因此我们选取上述公司进行对比。我们预计公司2026~2028年营收分别为38.88/87.63/199.51亿元，归母净利润分别为3.23/23.90/73.68亿元，维持“增持”评级。

沐曦股份业绩与估值指标	2025A	2026E	2027E	2028E
营业收入（百万元）	1,644.09	3,887.54	8,762.72	19,951.34
归母净利润（百万元）	-789.45	322.57	2,390.50	7,367.79
摊薄每股收益（元）	-1.97	0.81	5.97	18.41
P/E	—	616.56	83.20	26.99
P/S	141.17	51.16	22.70	9.97

表：可比公司估值表 (截至2026年9月30日)

	总市值（亿元）	预测净利润（亿元）				PE		
		2025A	2026E	2027E	2028E	TTM	2027E	2028E
沐曦股份-U（688802）	1,988.82	-7.89	3.23	23.90	73.68	—	83.20	26.99
可比公司								
寒武纪（688256）	6,345.20	20.58	55.41	115.81	200.26	190.43	54.79	31.68
摩尔线程-U（688795）	1,851.63	-10.01	-1.71	5.25	16.76	—	352.55	110.46

资料来源：iFind、国海证券研究所（注：沐曦股份盈利预测来自国海证券研究所，其他公司来自iFind一致预期，可比公司所处商业化阶段、产品矩阵、收入结构存在显著差异，PE/P/S 估值横向对比仅供有限参考，不代表完全可比）

下载日志已记录，仅供内部参考

5.3 风险提示

1）经营风险：核心技术及知识产权保护、纠纷风险；技术人才流失或储备不足的风险；产品迭代和技术创新无法满足市场需求的风险；募投项目未能顺利实施或效益不达预期的风险；单一产品收入占比较高的风险；

2）管理风险：规模扩张带来的管理风险；控制权稳定风险，公司控股股东、实际控制人的持股比例可能被进一步稀释，对公司的控制权稳定造成不利影响；

3）财务风险：持续亏损和存在累计未弥补亏损的风险；经营活动现金流持续为负的风险；应收账款回收风险；预付款资金安全风险；存货周转及跌价风险；汇率波动风险；大额股份支付的风险；

4）供应链安全风险：公司在晶圆代工、封装测试、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，涉及主要终端供应商来自于境外的情况；

5）产品商业化不及预期风险：曦云 C600 虽完成量产，但下游智算中心、行业客户实际采购规模不及公司及本报告假设的风险；自2025年底至今，公司完成37个主流旗舰模型的Day 0适配，不代表已经形成商业化采购订单，后续商业化落地仍存在不确定性；

6）性能与生态不及宣传风险：报告引用多项公司披露的性能、软件兼容指标，若大规模集群实际落地后性能、CUDA 迁移兼容效果不及公开宣传，将影响收入兑现；

7）盈利预测假设不及预期风险：报告 2026-2028 年收入高增长基于国产替代、产品放量假设，如果行业竞争加剧、GPU 产品价格下行，将导致收入、毛利率不及预测；

8）可比公司估值局限性提示：各家 GPU 厂商产品所处生命周期、收入结构差异较大，横向 PE/P/S 估值对比仅具备有限参考意义；

9）国内外公司不具备完全可比性及对标资料局限性风险：公司所处GPU/算力芯片行业在竞争格局、产品生命周期、收入结构、客户结构、研发阶段、盈利水平及估值体系等方面，与境外龙头及部分国内厂商存在较大差异，国内外公司并不具备完全可比性；

10）非经常性损益不可持续、扣非持续亏损风险：公司净利润可能受政府补助、公允价值变动收益、投资收益等非经常性损益影响较大，该类损益不具可持续性；

11）估值过高及股价高波动风险：公司估值水平可能显著高于传统行业或部分可比公司；

12）客户集中度风险：公司收入可能主要来自少数智算中心、行业客户或整机厂商。若主要客户资本开支放缓、采购计划调整，将对公司经营业绩产生不利影响；

13）流通盘小、换手率高及解禁波动风险：公司上市初期流通股占总股本比例可能较低，流通盘较小，少量资金即可对股价产生较大影响，容易出现换手率偏高、股价波动加剧的情形。未来限售股解禁后，可流通股份增加，若相关股东减持，可能增加二级市场供给并冲击股价，投资者需关注解禁节奏及股东减持安排；

14）模型Day 0适配不等于订单及收入风险：公司完成多项主流旗舰模型的Day 0适配，主要体现产品在软件生态、兼容性验证、联合调优等方面的进展，不等同于下游客户已形成商业化采购订单或收入。适配后能否转化为实际采购、采购规模、价格、交付节奏及收入确认仍存在不确定性，不构成对公司未来业绩的承诺；

15）2026H1归母盈利主要来自公允价值变动收益、扣非仍亏损风险：公司2026年上半年归母净利润可能主要受公允价值变动收益等非经常性项目影响，扣除非经常性损益后仍可能为亏损。公允价值变动收益受市场环境、估值变化等因素影响，具有较大不确定性和不可持续性，不能代表公司主营业务的真实盈利能力；若相关资产公允价值回落，可能对公司利润造成反向影响；

16）技术迭代风险：GPU及算力芯片行业技术升级快、研发投入高、产品生命周期相对较短。若公司未能准确把握技术路线、产品定义和客户需求，或竞争对手率先推出性能更强、生态更完善、性价比更高的产品，公司现有产品可能面临性能落后、价格下降、毛利率下滑、市场份额流失及研发投入无法收回的风险；

17）地缘政治及供应链受限风险：公司在晶圆代工、封装测试、HBM、EDA工具及部分接口IP等方面涉及境外主要终端供应商。若国际贸易摩擦、出口管制、制裁、关税政策或地缘政治冲突升级，可能导致关键物料、工具、IP或服务供应受限、采购成本上升、交付延迟，进而影响公司产品研发、量产、迭代和商业化进度。

下载日志已记录，仅供内部参考

沐曦股份盈利预测表

证券代码： 688802　　股价： 497.08　　投资评级： 增持(维持)　　日期： 20260930

资产负债表（百万元）	2025A	2026E	2027E	2028E
现金及现金等价物	9362	10615	10417	9509
应收款项	735	594	1095	2771
存货净额	1496	1691	2597	7275
其他流动资产	1054	2167	4377	9909
流动资产合计	**12647**	**15067**	**18486**	**29465**
固定资产	160	194	221	303
在建工程	0	0	0	0
无形资产及其他	866	831	802	851
长期股权投资	1	5	9	13
资产总计	**13675**	**16096**	**19518**	**30631**
短期借款	0	5	30	45
应付款项	122	396	736	1850
合同负债	116	428	613	1796
其他流动负债	188	445	854	2023
流动负债合计	**427**	**1273**	**2233**	**5713**
长期借款及应付债券	0	0	0	0
其他长期负债	82	308	308	308
长期负债合计	**82**	**308**	**308**	**308**
负债合计	**509**	**1581**	**2540**	**6020**
股本	400	400	400	400
股东权益	13165	14515	16978	24611
负债和股东权益总计	**13675**	**16096**	**19518**	**30631**

利润表（百万元）	2025A	2026E	2027E	2028E
营业收入	**1644**	**3888**	**8763**	**19951**
营业成本	715	1740	3896	8878
营业税金及附加	5	15	35	78
销售费用	204	233	351	599
管理费用	371	700	876	1197
财务费用	0	-32	-37	-36
其他费用/（-收入）	1027	1050	1490	1995
营业利润	**-768**	**342**	**2483**	**7640**
营业外净收支	-12	-7	-5	-6
利润总额	**-779**	**334**	**2478**	**7634**
所得税费用	10	0	0	2
净利润	**-789**	**334**	**2477**	**7633**
少数股东损益	0	11	87	265
归属于母公司净利润	**-789**	**323**	**2390**	**7368**
现金流量表（百万元）	**2025A**	**2026E**	**2027E**	**2028E**
经营活动现金流	**-1260**	**396**	**-68**	**-889**
归属于母公司净利润	-789	323	2390	7368
少数股东损益	0	11	87	265
折旧摊销	198	223	223	91
公允价值变动	-1	0	0	0
营运资金变动	-809	-329	-2683	-8421
投资活动现金流	**-3309**	**-170**	**-138**	**-30**
资本支出	-194	-174	-181	-177
长期投资	-3188	-4	-3	-3
其他	74	8	46	150
筹资活动现金流	**10268**	**27**	**8**	**12**
债务融资	-494	5	25	15
权益融资	10812	0	0	0
其它	-49	22	-17	-3
现金净增加额	**5697**	**1253**	**-198**	**-907**

每股指标与估值	2025A	2026E	2027E	2028E
每股指标				
EPS	-2.44	0.81	5.97	18.41
BVPS	32.91	36.25	42.19	60.60
估值				
P/E	—	616.6	83.2	27.0
P/B	17.6	13.7	11.8	8.2
P/S	141.2	51.2	22.7	10.0
财务指标	**2025A**	**2026E**	**2027E**	**2028E**
盈利能力				
ROE	-6%	2%	14%	30%
毛利率	57%	55%	56%	55%
期间费率	35%	23%	14%	9%
销售净利率	-48%	8%	27%	37%
成长能力				
收入增长率	121%	136%	125%	128%
利润增长率	44%	141%	641%	208%
营运能力				
总资产周转率	0.19	0.26	0.49	0.80
应收账款周转率	2.73	5.89	10.37	10.32
存货周转率	0.63	1.09	1.82	1.80
偿债能力				
资产负债率	4%	10%	13%	20%
流动比	29.61	11.83	8.28	5.16
速动比	23.69	8.97	5.32	2.32

下载日志已记录，仅供内部参考

研究小组介绍

计算机小组介绍

刘熹，计算机行业首席分析师，上海交通大学硕士，多年计算机行业研究经验，善于把握由技术、政策驱动的科技产业新趋势，致力于进行前瞻重磅推荐。2024年Wind金牌分析师，2024年同花顺最受欢迎分析师，2025年Wind金牌分析师。
唐锦珂，计算机行业研究助理，中山大学岭南学院硕士，主要覆盖服务器、液冷、算力租赁、IDC等方向。
谢婧茹，计算机行业研究助理，上海财经大学经济学院硕士，主要覆盖半导体、云计算、教育IT等方向。
朱谞晟，计算机行业研究助理，复旦大学金融硕士，主要覆盖存储、网络等方向。

分析师承诺

刘熹, 本报告中的分析师均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，独立，客观的出具本报告。本报告清晰准确的反映了分析师本人的研究观点。分析师本人不曾因，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收取到任何形式的补偿。

国海证券投资评级标准

行业投资评级

推荐：行业基本面向好，行业指数领先沪深300指数；
中性：行业基本面稳定，行业指数跟随沪深300指数；
回避：行业基本面向淡，行业指数落后沪深300指数。

股票投资评级

买入：相对沪深300 指数涨幅20%以上；
增持：相对沪深300 指数涨幅介于10%～20%之间；
中性：相对沪深300 指数涨幅介于-10%～10%之间；
卖出：相对沪深300 指数跌幅10%以上。

下载日志已记录，仅供内部参考

免责声明和风险提示

免责声明

本报告的风险等级定级为R4，仅供符合国海证券股份有限公司（简称“本公司”）投资者适当性管理要求的客户（简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。客户及/或投资者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，需以本公司的完整报告为准，本公司接受客户的后续问询。

本公司具有中国证监会许可的证券投资咨询业务资格。本报告中的信息均来源于公开资料及合法获得的相关内部外部报告资料，本公司对这些信息的准确性及完整性不作任何保证，不保证其中的信息已做最新变更，也不保证相关的建议不会发生任何变更。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。报告中的内容和意见仅供参考，在任何情况下，本报告中所表达的意见并不构成对所述证券买卖的出价和征价。本公司及其本公司员工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。本公司或关联机构可能会持有报告中所提到的公司所发行的证券头寸并进行交易，还可能为这些公司提供或争取提供投资银行、财务顾问或者金融产品等服务。本公司在知晓范围内依法合规地履行披露义务。

风险提示

市场有风险，投资需谨慎。投资者不应将本报告为作出投资决策的唯一参考因素，亦不应认为本报告可以取代自己的判断。在决定投资前，如有需要，投资者务必向本公司或其他专业人士咨询并谨慎决策。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，请公众投资者慎重使用未经授权刊载或者转发的本公司的证券研究报告。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。

若本公司以外的其他机构（以下简称“该机构”）发送本报告，则由该机构独自为此发送行为负责。通过此途径获得本报告的投资者应自行联系该机构以要求获悉更详细信息。本报告不构成本公司向该机构之客户提供的投资建议。

任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。本公司、本公司员工或者关联机构亦不为该机构之客户因使用本报告或报告所载内容引起的任何损失承担任何责任。

郑重声明

本报告版权归国海证券所有。未经本公司的明确书面特别授权或协议约定，除法律规定的情况外，任何人不得对本报告的任何内容进行发布、复制、编辑、改编、转载、播放、展示或以其他任何方式非法使用本报告的部分或者全部内容，否则均构成对本公司版权的侵害，本公司有权依法追究其法律责任。

下载日志已记录，仅供内部参考

国海证券 SEALAND SECURITIES

国海证券·研究所·计算机研究团队

心怀家国，洞悉四海

国海研究上海

上海市黄浦区绿地外滩中心C1栋国海证券大厦

邮编：200023

电话：021-61981300

国海研究深圳

深圳市福田区竹子林四路光大银行大厦28F

邮编：518041

电话：0755-83706353

国海研究北京

北京市海淀区西直门外大街168号腾达大厦25F

邮编：100044

电话：010-88576597