

## 投资者关系活动记录表

编号：2026-01

投资者关系活动类别	<input checked="" type="checkbox"/> 特定对象调研 <input type="checkbox"/> 分析师会议 <input type="checkbox"/> 媒体采访 <input type="checkbox"/> 业绩说明会 <input type="checkbox"/> 新闻发布会 <input checked="" type="checkbox"/> 路演活动 <input checked="" type="checkbox"/> 现场参观 <input type="checkbox"/> 其他（ <u>请文字说明其他活动内容</u> ）
参与单位名称	贝莱德、彬元资本、东方红资产管理、富达国际、高瓴、杭州德亚私募基金、华安基金、汇添富基金、景顺长城基金、摩根大通、摩根基金、南方基金、三星证券、申万菱信基金、兴业基金、兴业证券、永赢基金、中投公司、中信证券、Aspex、Baillie Gifford、Capital International Investors、Causeway Capital Management、Coronation、GIC、Ilmarinen、Invesco、KOO&WONG、Legatum Capital、Mirae Asset Financial Group、Oman State General Reserve Fund、Schroders Plc、Star Capital、Universities Superannuation Scheme
时间	2026年5月6日-5月8日、5月11日
地点	北京、上海
上市公司接待人员	董事长、董办投关人员
投资者关系活动内容介绍	<p><b>1、能否解释一下公司通用全功能 GPU 架构理念，相比于业界将计算与图形分开的做法，我们为什么要把这些能力放在一起？</b></p> <p>摩尔线程定位全功能 GPU，公司基于自主研发的 MUSA 统一系统架构，率先实现了在同一颗芯片上同时支持 AI 计算加速、图形渲染、物理仿真以及超高清视频处理所需计算能力的突破，推动了我国 GPU 产业的自主可控进程，相比采用 GPGPU、ASIC 等技术路线的其他单一 AI 加速卡产品，MUSA 架构技术具备更强的计算通用性、更优的技术演进能力、更高的生态兼容性以及更广泛的市场适应性。</p> <p>在世界模型、具身智能、AI4S、物理 AI 等未来 AI 的发展趋势下，全功能 GPU 拥有更多的竞争优势，可适配未来更多的</p>

应用场景，满足多功能与全精度的需求，具有更大的灵活性和通用性。

## **2、对于 CUDA 的市场地位，以及兼容性的长期策略怎么看？**

目前英伟达 CUDA 生态仍然占据绝对主导，一般而言，开发者或者客户从 CUDA 生态切换至新平台迁移成本高、适配工作量较大。

公司坚持既兼容主流又独立发展的策略，实现 MUSA 架构对 CUDA 生态高度兼容，旨在降低迁移门槛，减少双方人力与资源投入。同时，公司通过自研 MUSA 开发平台，提供完整的开发、调试等工具，原生适配了 PyTorch、vLLM、SGLang 等主流框架，开发者可以基于原生 MUSA 架构开发新项目。目前公司开发者社区已经覆盖超过 45 万人，覆盖 200 多所高校，去年举办了首届 MUSA 开发者大会，今年还在快速扩张。

## **3、公司毛利率维持较高水平的核心原因是什么？**

公司当前聚焦高性能训练市场为主，兼顾高质量的推理市场，该业务定位使得公司具备较高的技术门槛，拥有优质的客户结构与较低的市场竞争，从而支撑较高毛利率水平。

## **4、公司 GPU 产品是否可以适配市场上的主流模型？**

公司产品已与业界主流 SOTA 大模型进行了 Day0 适配。公司依托 MTT S5000 原生 FP8 能力与完善的 MUSA 软件生态，快速完成对 DeepSeek V4、Minimax M2.7、GLM-5.1、中国移动九天大模型等的 Day0 适配。

## **5、公司如何看待训练与推理的差异？**

训练市场更强调系统级平台的创新能力。目前以大模型厂商为代表的企业算法不断更新迭代，GPU 厂商需提供通用、灵活的工具链以满足客户的需求，而摩尔线程全功能芯片具有更

强的适应能力；推理市场则更关注性价比，当企业对于推理使用芯片产品需求量大时，可能会倾向于定制优化芯片性能。

推理市场的需求来源于终端用户交互，是对市场中已存在的模型进行应用，例如每天大量用户与大模型进行交互会消耗大量的词元（Token）。训练则是持续提升大模型智力水平的过程，技术门槛更高，每次训练大模型投入大。当大模型参数越多，数据量也会越多，所需算力水平呈指数增长。

#### **6、公司在 GPU 互联通信方向的布局如何？**

摩尔线程在 GPU 通信领域构建了自主可控的三层技术体系：芯片层面通过 ACE 异步通信引擎实现通信与计算物理级并行，减少 15% 计算资源损耗；卡间互联层面自研 MTLINK2.0 技术，带宽性能高出国内行业平均水平，支持万卡以上规模集群扩展；生态标准层面深度参与中国移动 OISA 开放互联架构制定，并在下一代高密超节点引入对该标准的支持，将跨节点带宽推向 TB/s 级，为大规模智算中心提供了从芯片硬件到集群互联、从私有技术到开放标准的全栈解决方案。

#### **7、公司当前的产品进展及下一代产品规划**

公司旗舰级训推一体全功能智算卡 MTT S5000 产品已量产并实现商业化落地，MTT S5000 是国内为数不多的支持 FP8 精度的训推一体芯片。

公司下一代的产品基于“花港”芯片架构，计划推出两款芯片：“庐山”面向高性能图形渲染，“华山”则专注 AI 训练与推理加速。

#### **8、目前云边端同时进军，战略考量是什么？**

公司的底层技术架构都是以全功能 GPU 为核心，以 MUSA 统一架构为基础。而面对未来 AI 赋能全场景的趋势，无论是云端、边缘端还是终端均会迎来全面爆发。

	<p>公司目前已构建起从云到端的产品矩阵，云端产品以训练为主，持续推进高性能国产 GPU、先进互连网络、集群软件栈和算力调度平台的一体化布局，构建支撑万卡级、十万卡级智能算力集群的自主可控技术底座。在边缘与终端方向，加快推进 SoC 的迭代升级，围绕智能体和端侧推理发展趋势，前瞻布局端侧算力、轻量模型与软硬一体平台能力，推动终端设备从“可连接”向“可感知、可理解、可执行”的智能体载体升级。</p>
<p>关于本次活动是否涉及应当披露重大信息的说明</p>	<p>无</p>