

狂堆GPU路渐窄 算力调优接棒AI竞赛

见习记者 尹靖霏

AI算力的竞争逻辑，正在松动与重构。

过去两年，行业深陷GPU军备竞赛，谁囤卡多、集群大，谁就占先机。当下，这套“以规模论输赢”的法则正被瓦解——在词元(Token)计费模式下，规则从“卖硬件”变为“卖Token”，堆硬件的边际收益递减，效率由此成为新战场。

7月初，京算Token工厂落地，同时，趋境科技半年累计融资超10亿元。前者标志算力供给走向标准化，后者折射调优技术受资本追捧。两条线索指向同一方向：算力调优正从幕后走向前台，成为决定企业盈利、重塑国产算力格局的关键变量。

然而，芯片架构碎片化、全栈人才匮乏、集群稳定性不足等问题依然严峻，行业尚未形成标准化体系，算力软实力的长期竞赛才刚刚开始。

规则重构：调优碾压堆卡

国内算力市场长期由海外主导，按卡时长计费，客户选型看重单卡峰值性能、CUDA生态成熟度，英伟达凭此形成垄断。作为企业级AI平台公司，范式智能首席战略官兼副总裁曾慈雯向证券时报记者表示，国产芯片虽在成本、供给、异构部署灵活性上有优势，但因生态短板沦为备选，难入核心场景。

Token计费模式的普及，正在改变这套固化多年的评价体系。

“Token模式的本质，是算力从卖硬件变成卖结果。”曾慈雯直言，全新规则下，客户不再为硬件参数买单，核心考核指标变成单Token交付成本、有效产出效率、服务稳定性，而底层硬件的品牌、性能差异重要性降级。

这套规则切换，直接盘活国产算力的差异化优势。

曾慈雯表示，国产芯片无需再硬拼单卡峰值短板，依托硬件成本优势，搭配全栈算力调优技术拉升集群整体吞吐能力，就能跑出优于海外产品的性价比，顺利进入此前无法触及的核心使用场景。

曾慈雯告诉证券时报记者，这场变革为国产算力解锁机遇：跳出海外定义的硬件“内卷”赛道、倒逼国产软硬件生态体系化成熟；降低中小企业算力使用门槛；解锁海量下沉市场增量。而AI产业发展的下一阶段，在于推动AI能力走向普惠化。算力调优和基础设施能力的成熟，将成为推动AI普惠化的重要支撑。

国产芯片的成本优势向Token生产优势转化，关键在算力调优而非硬件参数，市场给出了最直接的回应。

2026年，AI行业迎来商业拐点：Token生意能赚钱了。

是石科技正打造全国最大的“Token优化工厂”，公司联合创始人毛运航告诉证券时报记者：去年算力调优只是小众技术探索，哪怕优化10%的效率，也无法扭转整体亏损的局面。但是今年，C端、B端AI推理应用集中爆发，API付费模式跑通，大模型推理赛道整体跨过只见投入不见产出的阶段。

固态电池产业链公司业绩报喜 量产建设持续推进

证券时报记者 康殷

固态电池产业化按下“加速键”，与之相伴的是产业链公司上半年业绩集体飘红。截至7月16日，国轩高科、多氟多、杉杉股份等多家固态电池概念公司披露2026年半年度业绩预告，归母净利润同比增幅普遍在一倍以上，部分企业甚至实现数十倍增长。

业内人士指出，本轮业绩改善既受益于新能源汽车、储能市场需求持续扩张带来的量价齐升，也叠加了固态电池技术验证与中试产线密集落地所释放的产业预期，锂电材料板块景气度已从“底部修复”步入“加速兑现”阶段。

材料端量价齐升

在已披露业绩预告的固态电池产业链公司中，业绩增幅最为亮眼的当数锂盐环节。天齐锂业预告上半年归母净利润为28.5亿元至42.5亿元，同比增幅达3276.35%至4934.91%。



图片来源:AI生成

在毛运航看来，盈利逻辑逐渐成立，让算力调优从“锦上添花”变成“雪中送炭”。每一次效率提升、每一分成本压降，都会直接体现在企业利润表上，这也是当下全行业愿为优化技术重金买单的主因。

毛运航分析，国产芯片存在工艺、产能、单卡代差、技术路线分散及生态碎片化等短板，适配难度高，要靠技术盘活算力，在商用场景中寻找成本效率最优解，实现国产算力商业化突围。

格局重塑：第三方AI Infra崛起

头部芯片厂商软件架构师告诉证券时报记者，市场普遍存在误区，认为算力调优仅针对显卡算子，而实际产业中的专业调优是端到端、全层级的系统工程，覆盖NPU/GPU、CPU、存储、网络、操作系统、算子算法等全链路。

随着算力调优赛道价值爆发，产业入局者已呈全面铺开态势，芯片厂商、云厂商、运营商、大模型企业、第三方AI Infra（人工智能基础设施）同台竞技，形成分层互补的产业格局。

证券时报记者了解到，某头部国产芯片厂商组建千人级调优团队，聚焦自有硬件的底层适配与基础生态搭建。各大模型企业也标配专属调优团队，服务自有模型迭代，持续降低训练、推理成本。手握大规模算力集群的云厂商、运营商，将算力利用率、Token产出效率作为核心经营指标，通过调优盘活重资产。

传统算力调优聚焦硬件集群层面，是超算时代延续至今的基础技术能力。而Token调优的兴起，是2026年AI产业的全新变量，核心源于行业商业模式彻底反转。独立第三方AI Infra公司的快速崛起，在其中扮演关键角色。2023年至2024年上半年，清程极智、硅基流动、无问芯穹、趋境科技等公司纷纷成立，并在此后完成多轮融资；是石科技2026年完成数亿元融资，硅基流动已向港交所递交上市申请，当前估值约77亿元。

从操作路径看，第三方Infra公司普

遍遵循相似的技术架构：基础层做异构算力统一纳管；核心优化层自研推理引擎，覆盖算子开发、键值缓存等专项优化；调度层实现动态负载分配和集群混部；顶层提供标准化API和多模型统一接入。

2023年是AI Infra赛道的产业化拐点。星连资本押中智谱、硅基流动等多家AI Infra独角兽公司，其合伙人王璞告诉证券时报记者，这一创业潮发生在2023年10月美国对华芯片出口管制升级后，国产芯片供给成为常态化约束，如何用更少的芯片训练出更强的模型成为行业最紧迫命题。

第三方Infra公司核心优势在于“中立性”。王璞表示，大厂自研团队无法覆盖市场碎片化需求，大量中小算力服务商涌入Token市场，跨芯片、跨模型的通用适配、调优和调度需求集中爆发，第三方公司凭借专业化服务填补市场空白。

毛运航分析，不同于大厂（生态绑定）、模型公司（重算法轻基建），第三方Infra公司搭建起“硬件适配—全栈调优—智能调度—标准化服务”的完整链路，使算力利用率跃升，Token生产成本骤降。

曾慈雯认为：“现在市场的注意力集中在芯片、大模型、云厂商这些显性环节，但真正决定Token经济运行效率的是中间层——谁能把分散、异构、碎片化的算力转化为稳定、标准化的Token产能，谁就掌握核心生产力。”

王璞预判未来一两年将出现三个变化：上游纯硬件溢价逐步回落；中间转化层的价值占比快速提升；垂直场景的应用与模型方拿到更多增量价值。毛运航判断，第三方AI Infra头部壁垒成型，新玩家窗口收窄，未来细分赛道将呈现寡头垄断格局。

瓶颈与破局：调优的硬仗才刚开始

赛道火热，挑战同样严峻。这些挑战，既有行业层面的共性难题，也有新玩家特有的生存风险。

数企业当前主营收入仍以传统液态锂电材料为主，固态电池相关业务处于中试或小批量验证阶段，尚未形成规模化收入贡献。

头部企业加速布局

从行业整体进展看，固态电池产业化时间表正持续清晰化。国轩高科预告上半年净利润12亿元至15.5亿元，同比增长227.31%至322.77%。公司表示，正加快研发产业化落地，推进产品迭代升级。此前公司已披露固态电池量产线设计工作已基本完成，相关技术研发及产业化工作在稳步推进中。

公开资料显示，国内多条固态电池产线建设取得实质性进展：四川赛科动力宜宾0.1GWh全固态电池智能制造中试线预计2026年底建成，聚焦硫化物路线；恩力动力安徽滁州2GWh半固态电池基地于2026年7月完成产能爬坡达产；国晟环球铁岭10GWh固态电池产业链项目首线设备于7月搬入车间；广州卫蓝能源20亿元固态电池产业园正式动工，计划2027年

对于新兴的第三方AI Infra公司而言，面临的另一重挑战是“技术自嗨”，毛运航直言，过度专注于技术攻坚，却忽略了商业化节奏，“先驱”容易变成“先烈”。

曾慈雯将行业普遍存在的挑战归纳为四大难题：国产芯片架构碎片化，每款硬件都需要定制化底层适配，踩坑成本极高；全栈优化门槛高，需要打通软硬件全链路，复合型团队稀缺；场景化调优难以复制，不同业务的优化逻辑完全不同，需要长期沉淀方法论；大规模异构集群的动态调度复杂度呈指数级上升，对工程化能力要求极高。

就人才方面而言，清程极智联合创始人师天麓告诉证券时报记者，全栈算力调优属于高门槛交叉学科，需要同时精通硬件架构、底层系统、网络通信、大模型算法，培养周期极长。此前赛道小众、商业价值未显现，人才储备长期匮乏。不同于算法赛道的高薪热门，高性能计算方向资深工程人才“一将难求”。

此外，底层产业生态仍有明显短板。前述芯片架构师表示，英伟达凭借全球用户生态实现快速迭代与缺陷修复，国产卡用户基数有限，迭代节奏慢，驱动层bug（漏洞）在规模化部署中仍属常见现象。

师天麓表示，各芯片厂商底层接口开放程度不一，行业暂无统一标准，调优深度高度依赖底层编程接口的开放层级，部分厂商开放有限，压缩了优化空间。不过，华为昇腾近年在开放程度上已有明显改善，反映产业共识正在形成。在硬件层面，国产大规模集群的运行稳定性不足，单卡故障易引发连锁问题，万卡级别集群落地难度较高，难以通过规模效应进一步摊薄Token生产成本。

挑战即机遇，算力调优与AI Infra正以Token产能改写格局，推动国产算力从硬件替代迈向生态自主，开启软科技长牛。



长鑫科技网上发行中签率刷新科创板历史纪录

证券时报记者 胡华雄

7月16日晚间，国内DRAM（动态随机存取存储器）龙头企业长鑫科技披露新股发行中签率。

根据上交所提供的数据，长鑫科技本次网上发行有效申购户数约为942.88万户，有效申购股数约为8169.2亿股，网上发行初步中签率为0.40995452%。

值得注意的是，长鑫科技联席主承销商已按本次发行价格向网上投资者超额配售约10.03亿股，约占初始发行股份数量的15%。另外，发行人和联席主承销商还决定启动回拨机制，对网下、网上发行的规模进行调节，将扣除最终战略配售部分后本次公开发行业股票数量的10%（即约5.02亿股）从网下回拨至网上。

在回拨机制启动后，长鑫科技网上发行最终中签率为0.47141739%。

据Wind数据，长鑫科技上述网上发行最终中签率刷新科创板历史纪录。在长鑫科技之前，科创板网上发行最终中签率最高的新股为2021年发行的和辉光电，彼时和辉光电网上发行最终中签率为0.30314621%。

根据公告，发行人与联席主承销商将于2026年7月17日（T+1日）上午进行本次发行网上申购摇号抽签仪式，并将于2026年7月20日（T+2日）在上交所网站公布网上摇号中签结果。

资料显示，长鑫科技专注于DRAM产品的研发、设计、生产及销售，公司产能规模已位居中国第一、全球第四。若以预计募集资金规模计算，长鑫科技亦是科创板有史以来募资规模最大的IPO。

成本攀升与需求增长并存 磷酸铁锂龙头宣布涨价

证券时报记者 赵黎昀

上游硫磺、硫酸亚铁等原材料价格普涨，叠加下游需求显著增长，国内磷酸铁锂市场呈现出供不应求局面。

近日，国内主要的锂离子电池正极材料供应商湖南裕能对外官宣调价方案，对全部磷酸铁锂产品在现有执行价基础上每吨上调2000元，新规将于今年8月1日正式生效。

分析人士指出，本次调价折射出磷酸铁锂行业当前的深层矛盾，即上游成本全面攀升与下游需求刚性增长并存，行业正处于从产能过剩向供需紧平衡切换的关键周期。

成本上行推涨

“本次调价函直指两大核心驱动因素，一是上游核心原材料价格持续上行，二是公司持续满产，新增产能无法满足全部客户订单增长需求。”对于此番湖南裕能官宣调价，富宝资讯磷酸铁锂分析师苗敏称，2026年上半年，磷酸铁锂价格走出一轮强劲的周期性反弹行情。2025年三季度，磷酸铁锂走出周期底部，从动力型均价35771元/吨、储能型29000元/吨起步，到2026年二季度升至动力型58178元/吨、储能型57016元/吨，涨幅分别达63%和97%。7月高位震荡于5.5万元/吨—6万元/吨区间，龙头公司上调加工费再度推升价格中枢预期。

涨价背后，磷酸铁锂成本端压力不断提升。磷酸铁锂成本主要由碳酸锂（占60%—65%）和磷酸铁加工费（占30%—35%）两部分构成。2026年上半年，两部分成本同步大幅上行。

苗敏介绍，作为磷酸铁锂的上游，磷酸铁价格从2025年6月的10450元/吨涨至如今15400元/吨，涨幅47.4%，占加工费比重已达70%以上。而自2026年初至6月中旬，仅硫磺一项均价就上涨7490元/吨，传导至磷酸铁锂端单吨成本上涨约6000元。叠加硫酸亚铁、双氧水等辅料涨价，非锂成本实际增量超过5000元/吨，但同期加工费上调幅度远未覆盖这一成本增量，正极厂利润被严重侵蚀。

高端供给呈现紧缺态势

对于当前磷酸铁锂市场供需局面，隆众资讯磷酸铁锂分析师翟林林称：“当前磷酸铁锂供给端头部满产，高端高压实产品因技术认证壁垒偏紧溢价，而需求端动力缓修复、储能高景气延续，海外大储及出口刚性采购托底。”

富宝资讯调研报告也显示，当前磷酸铁锂行业名义产能看似充足，但符合下游头部客户认证标准的高端有效产能严重不足。

苗敏表示，当前磷酸铁锂供应端呈现结构性分化，行业最显著的特征是“名义过剩、有效紧缺”。

与此同时，在需求端，动力+储能双轮驱动叠加出口爆发，为磷酸铁锂价格走强提供支撑。据统计，磷酸铁锂下游需求呈现“动力稳健、储能爆发、出口激增”三大特征。2026年全球需求预计556万吨，同比增长约57%。

短期市场紧平衡

虽然磷酸铁锂价格已显著走强，但市场对后市依然持看多预期。

翟林林预计，2026年下半年，我国磷酸铁锂市场将在产能释放与需求回暖的双重驱动下，迎来新一轮的产销两旺周期，我国全年磷酸铁锂总产量预计为638万吨左右。随着前期规划的新增产能在下半年集中兑现落地，全年的设计产能规模将跃升到一个新的量级。在下游需求平稳发展的支撑下，行业整体产能利用率将呈现稳定且小幅优化的态势，相关企业的生产积极性得到有效提振。

“2026年下半年，磷酸铁锂价格或呈现高位震荡、紧平衡持续格局。预计三季度旺季备货支撑价格小幅反弹，四季度随少量新产能爬坡温和回落，但不会跌破5.5万元/吨。”苗敏认为，磷酸铁锂高端产能建设周期长，2026年内无大规模高端增量释放。年底预计新增约25万吨高端产能，缺口收窄但仍无法完全抹平。

不过，苗敏表示，2027年下半年磷酸铁锂或迎来供需转折的关键节点。行业名义产能预计将达1200万吨，若需求增速放缓，产能利用率可能跌破60%，重演2023—2024年的过剩竞争格局。