证券研究报告

开源证券

全视角科技投资脉络：从大模型百家争鸣到算力需求井喷

计算机行业2027年度策略

姓名 张越（分析师）
证书编号：S0790524090003
邮箱：zhangyue1@kysec.cn

姓名 叶进新（联系人）
证书编号：S0790126060053
邮箱：yejinxin@kysec.cn

2026 年 11 月 9 日

核心观点

1.大模型能力提升与产业应用同步推进

Transformer架构、数据工程、训练优化与后训练共同决定模型能力；推理效率和部署成本影响模型能否进入真实业务流程。国产模型竞争从单点性能延伸至开放程度、调用成本、产品体验和生态适配。智谱围绕GLM、Coding、多模态与Agent扩展应用。Anthropic案例展示企业客户、用量收入与算力投入的联动。

2.物理AI加速落地，推理需求带动AI服务器与国产算力扩容

物理AI通过数据采集、仿真训练和真实场景部署形成完整链条，搬运等任务已有进展，精细操作与跨场景泛化仍需真机验证；具身大脑有望与机器人本体分开收费，商业化仍取决于应用效果与客户的付费意愿。复杂推理与Agent任务增加持续性算力消耗，服务器需求从训练集群延伸至高并发推理，国产芯片厂商通过软硬件适配、集群交付和客户场景推进规模应用，采购需求正从测试验证转向实际部署。

3.算力租赁、液冷与先进封装承接基础设施需求

算力租赁让客户按需使用算力，供给方涵盖云厂商、运营商与专业服务商；机柜功率密度上升推动散热升级，冷板式液冷适合现有服务器架构，浸没式、喷淋式液冷继续探索更高密度应用场景，CDU、歧管与快接头等部件集中了液冷系统的价值。先进封装涵盖2.5D、3D及面板化路线，玻璃基板以材料特性、应用形态与核心工艺为基础，设备订单、载板验证和产业链配套决定产业化节奏。

4.受益标的

国产算力：寒武纪、海光信息、禾盛新材、沐曦股份、浪潮信息、工业富联等；海外算力：东山精密、胜宏科技、联特科技、德邦科技等；算力租赁：宏景科技、利通电子、协创数据等；大模型：智谱、MiniMax、阿里巴巴、腾讯控股；液冷：盾安环境、同飞股份、恒铭达等；玻璃基板：沃格光电、京东方、彩虹股份、帝尔激光、东威科技、德龙激光等。

5.风险提示：大模型技术迭代不及预期风险； AI技术应用不及预期风险；AI Infra建设不及预期风险；技术研发进度不及预期的风险；玻璃基板技术产业化进展不及预期风险；玻璃基板市场需求不及预期风险；行业竞争加剧的风险；相关政策不及预期风险。

目 录

CONTENTS

1 大模型百花齐放，coding之外谁主沉浮

2 算力需求井喷，GPU、算租、液冷充分受益

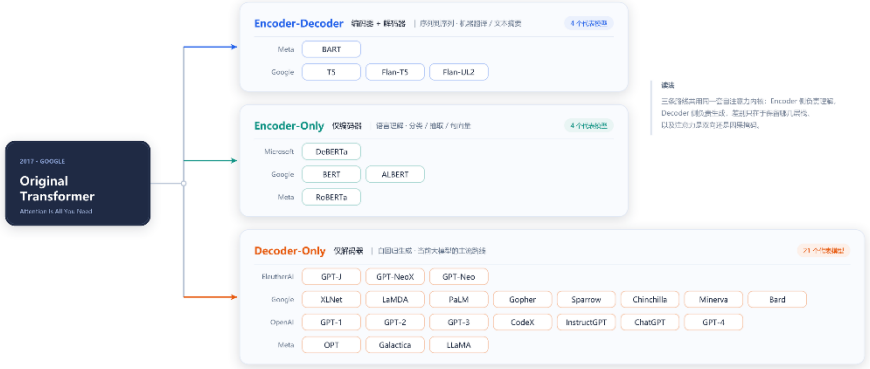
3 功耗问题事关全局，技术迭代趋势明确

4 风险提示

1.1 模型能力与技术底座：架构、训练、推理及行业应用

大模型按处理模态可分为语言大模型、视觉大模型、语音大模型以及多模态大模型。语言大模型以自然语言理解与生成为核心，以GPT、DeepSeek、Kimi为代表，广泛应用于对话、创作、代码及知识检索等场景；视觉大模型聚焦图像与视频的理解、分割与生成，以SAM、Stable Diffusion为代表，在自动驾驶、医疗影像及工业质检领域加速渗透；语音大模型围绕音频识别、合成与语义解析展开，是智能客服与车载交互的关键底座。多模态大模型通过统一架构融合文本、图像、语音等信息，实现跨模态理解与生成，以GPT-4o、Gemini为标杆，被视为通向AGI的核心路径，是当前技术演进的主流方向。

图1：Transformer架构催生了Encoder-Decoder、Encoder-Only与Decoder-Only三条差异化的技术路线



资料来源：Sebastian Raschka《Machine Learning Q and AI》、开源证券研究所

图2：MoE架构中路由网络为每个token挑选少数专家，其余专家不参与计算



资料来源：云雀AI学工台公众号

1.2 智谱沿 GLM 主线构建全栈能力，由技术积累迈向商业化

智谱围绕一体化MaaS平台组织产品供应，平台由全面的模型组合、可扩展的应用和便捷的基础设施适配性三个层级构成。模型层覆盖语言、多模态、智能体及代码四类模型，并提供模型微调、部署及智能体开发工具；应用层面向科技、金融服务、互联网、智能设备、医疗、零售等场景，

智谱ARR高增长的核心支撑在于量价齐升与单位经济模型的改善，token经济模式已初步跑通。截至8月末，平台Token调用量较年初增长超40倍，API平均售价提升约101%，打破行业价格战下“以价换量”的常规路径。开放平台及API业务毛利率由上年同期的-0.4%转正至24.6%，验证了推理成本优化（单位Token推理成本较年初下降80%）对盈利模型的实质贡献。9月11日披露总额约50亿美元再融资后，管理层进一步将年末ARR目标上调至30亿美元。

图3：智谱ARR持续增长，2026年末指引达30亿美元

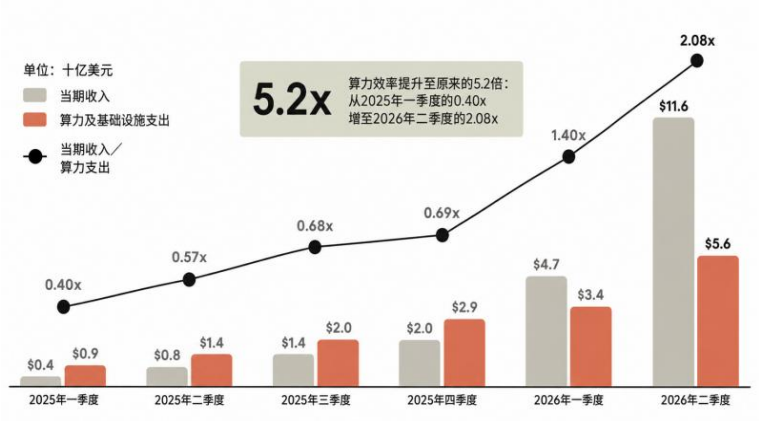


资料来源：中国证券网、网易新闻、同花顺新闻、开源证券研究所

1.3 Anthropic商业化：模型与产品双轮驱动兑现盈利拐点

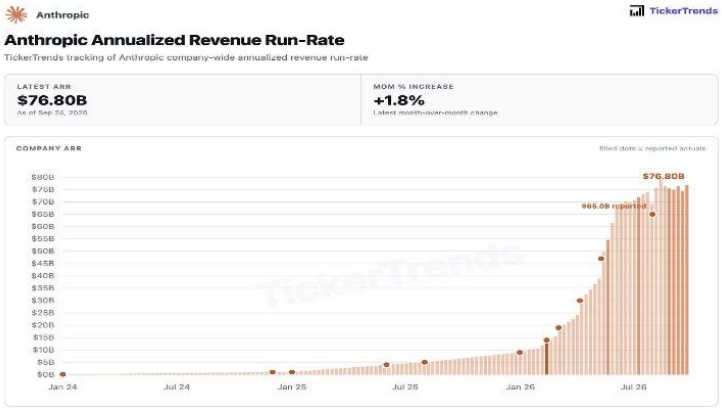
主营业务已跨越盈亏平衡：收入对算力支出的覆盖自2026年一季度转正，单季调整后营业利润率于二季度转正，GAAP净亏损主要源自非经营性公允价值变动。分季度看，当期收入由2025年一季度的4亿美元增至2026年二季度的116亿美元，同期算力及基础设施支出由9亿美元增至56亿美元，收入对算力支出之比由0.40倍升至2.08倍——算力效率提升至原来的5.2倍，并自2026年一季度起收入开始覆盖算力支出；调整后营业利润率由2025年一季度的-221%逐季收窄至-129%、-93%、-84%，2026年一季度为-13%，二季度转正至19%，累计改善240个百分点，单季营业利润率同步转正至8%。年度与半年度口径下，收入由2023年4,100万美元、2024年3.86亿美元增至2025年45.85亿美元，2026年上半年达162.86亿美元、经营亏损收窄至6.93 亿美元，调整后营业利润16.18亿美元、利润率10%（上年同期为亏损18.49 亿美元、利润率-159%）。同期净亏损807.48亿美元，其中可转换票据、衍生负债与SAFE公允价值变动贡献724亿美元——该非现金项目随公司估值上升而扩大。

图4：每美元算力支出带来的收入显著增加



资料来源：Anthropic招股说明书、开源证券研究所

图5：Anthropic的ARR增速放缓

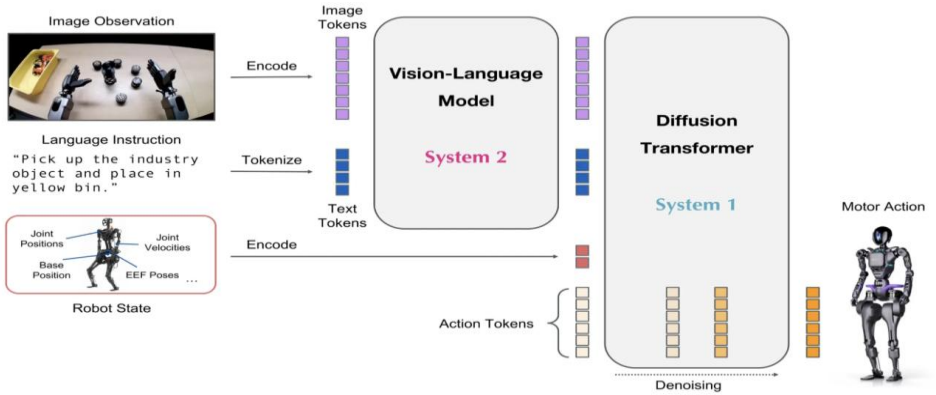


资料来源：TickerTrends Research

1.4 物理AI或成编码外最大落地场景，全球巨头重点发力

物理AI通过感知、推理与行动参与物理世界，形成“感知—决策—行动—反馈”闭环。从机器人应用看，可围绕环境预测、动作生成、仿真验证与实体执行理解其技术体系，分别涉及世界模型、VLA、仿真与数字孪生及机器人本体。世界模型从视频及交互数据中学习环境动态，在观测及动作等条件下预测未来状态或其特征表示，为规划、数据合成和策略评估提供支持；VLA（视觉—语言—动作模型）结合视觉、语言指令及机器人状态生成动作序列，输出末端位姿、关节目标等，与底层控制系统协同执行；物理仿真提供可重复的训练与测试环境，数字孪生构建与真实设备或场景对应、可结合实测数据更新的数字表示；机器人本体承担物理执行。上述环节可按任务组合，预测、推理与动作生成也可在统一模型中融合，共同支撑具身智能应用。

图6：英伟达GR00T双系统VLA架构，由VLM规划任务、扩散模型生成动作



资料来源：英伟达

1.4 物理AI或成编码外最大落地场景，全球巨头重点发力

Astra的核心定位为空间推理与任务规划基座，项目与独立机器人硬件团队并行研发。GPT-6Astra在视觉定位、复杂任务推理和长程交互方面更强，为物理AI的场景理解、任务拆解及动作规划提供了更有力的通用模型基础。在机器人的操作流程上，研究人员以GPT-6Astra为模型基座，依托 Franka FCI 高速控制接口与 ROS 生态，研究团队通过将API接入Astra，由Astra读取相机流与机器人本体状态，接收自然语言指令完成场景理解、任务拆解，输出笛卡尔空间末端目标点，再交由 Franka 底层力控控制器执行动作。

表1：英伟达GR00T双系统VLA架构，由VLM规划任务、扩散模型生成动作

能力	GPT-5.6 Sol	GPT-6 Astra	对比
视觉定位：ScreenSpot-Pro（无工具）	76.90%	92.70%	提升 15.8 个百分点
电脑操作：OSWorld 2.0	65.70%	72.60%	提升 6.9 个百分点
专业工作流：AutomationBench	18.10%	41.40%	提升 23.3 个百分点
软件工程：Terminal-Bench 4.0	37.30%	57.90%	提升 20.6 个百分点
科研任务：Terminal-Bench Science 0.1	22.40%	64.60%	提升 42.2 个百分点
长上下文检索：MRCR v2，8-needle，512K–1M	73.80%	96.30%	提升 22.5 个百分点
任务边界：内部电脑操作安全评测（越低越好）	22.00%	2.40%	降低 19.6 个百分点
幻觉控制：内部幻觉评测（越低越好）	12.20%	4.20%	降低 8.0 个百分点

资料来源：OpenAI、开源证券研究所

2.1 Agent推理浪潮驱动需求井喷，服务器量价齐升延续景气上行

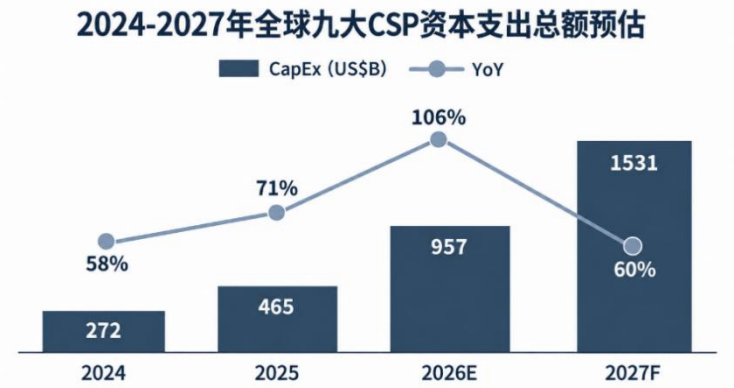
AI产业重心自模型训练转向推理规模化落地，算力需求的性质发生根本改变，由项目制建设转变为持续性消耗

2026年被业界视为Agent规模化部署元年，中国信息通信研究院披露，截至2026年6月我国日均词元调用量已突破500万亿，较2025年12月增长近四倍；IDC预计国内活跃Agent数量将从2026年的495万个增至2030年的2亿个。Agent工作负载以多轮交互、长上下文与多智能体协同为特征，任务轮次与并发规模持续放大Token消耗，用量的增长具备自我强化属性。推理需求的扩张是刚性的：一旦嵌入业务流程便不可撤回，其持续放量构成下游资本开支上修的需求底座。

云厂商资本开支扩张斜率放缓，但算力军备竞赛的投入强度不减，供应获取能力取代预算规模成为竞争核心

TrendForce集邦咨询10月14日发布研究，将Google、Amazon、Meta、Microsoft、Oracle等九大CSP资本支出预期由年增50%上调至60%：2026年总额逾9570亿美元、同比增长106%，2027年有望突破1.5万亿美元、同比增长60%，均优于先前预期，绝对规模逐年抬升。上调由三项因素驱动：服务器整机价格上涨、零部件缺货推高报价、部分供应商要求预付款以确保供货，CSP被迫提早支出现金。资本开支节奏已由建设周期主导转向受供给瓶颈与商务条款共同塑造，军备竞赛的胜负手正从预算规模转向供应获取能力。

图7：云厂商资本开支扩张斜率放缓，但算力军备竞赛的投入强度不减



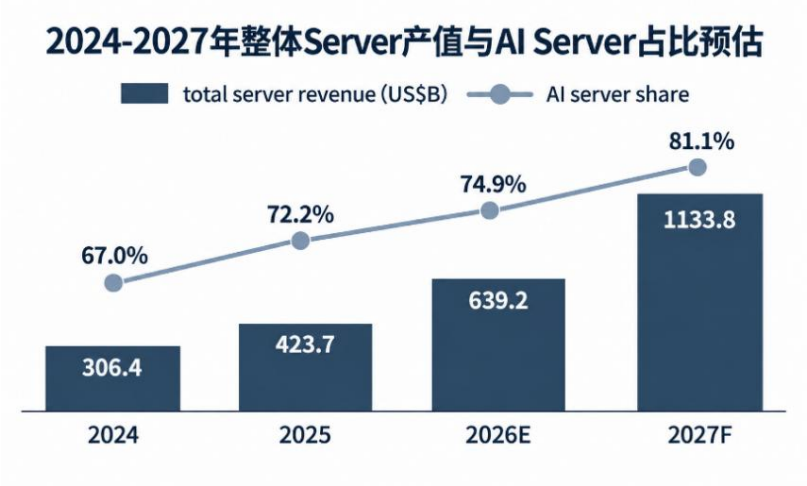
数据来源：TrendForce集邦咨询公众号、开源证券研究所

AI服务器市场扩张由台数驱动转向价值驱动，产值增速显著跑超出货增速，量价齐升成为本轮景气周期的核心特征

云厂商资金充沛带动强劲出货，预计2027年AI服务器出货增长将继续扩大

出货端，TrendForce集邦咨询预计2026年AI服务器出货同比增长32.2%，2027年在超大型CSP、Neocloud与主权AI项目共同拉动下再增近36%，增速平稳抬升。产值端，2026年AI服务器产值约4790亿美元，在整体服务器约6392亿美元产值中占比接近75%；2027年产值预计跃升至近9200亿美元，同比增长超过90%，占整体服务器产值比重升至81%左右，推动整体服务器产值首次跨越1万亿美元。价值量攀升源于新一代整柜方案均价上行与存储器为首的零部件涨价，GB300、VR200及Helios等整柜式方案或成为2027年订单主力。

图8：2024-2027年服务器产值稳步增长，AI服务器占比持续抬升



数据来源：TrendForce集邦咨询公众号、开源证券研究所

服务器需求放量正与供给刚性碰撞，上游供应链或步入较长紧缺周期

高盛将2026年DRAM涨幅预期由约150%上调至250%-280%，NAND由约100%上调至200%-250%，ABF载板、CCL同步上修，其预计DRAM供需缺口达-4.9%，或为十五年来最严重短缺，紧缺延续至2028年。HBM耗片量约为普通DRAM的数倍，持续挤占通用产能，缺口根源主要为供给弹性不足叠加AI服务器需求强劲。先进制程新产能预计2028-2029年方逐步释放，云厂商长约进一步锁定存量供给，供给弹性趋于收缩。涨价信号自存储向载板、被动元件与材料逐级传导，AI服务器价值构成调整或由需求侧逻辑演变为供需共振逻辑。

图9：半导体多环节涨价与供需预期4月全面上修，紧缺态势持续升级



主要变化（4月较1月）	
DRAM	2026年第四季度及2027年上半年供需趋紧；2026年涨幅上调至250%~280%（此前约150%）。
NAND	2026年第四季度及2027年上半年供需趋紧；2026年涨幅上调至200%~250%（此前约100%）。
台积电	2027年上半年由紧张转为非常紧张。
中国大陆晶圆代工	2026年及2027年上半年由紧张 / 平衡转为不紧张；2026年预计涨价10%（此前0%）。
MLCC	2026年第二季度至2027年上半年由紧张转为非常紧张；2026年预计涨价0%~5%（此前略低于0%）。
ABF载板	2026年下半年及2027年第一季度由紧张转为非常紧张；2026年涨幅上调至30%~35%（此前20%~30%）。
CCL	2026年第四季度及2027年第一季度由不紧张转为紧张；2026年涨幅上调至35%~40%（此前5%~10%）。

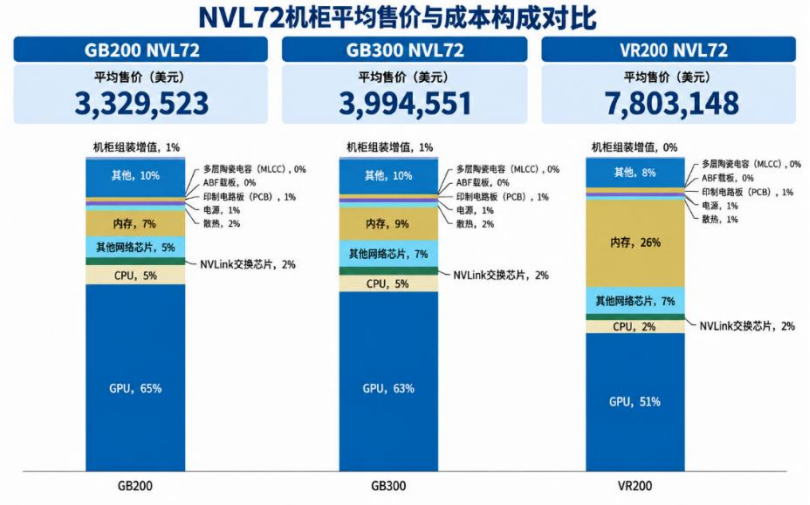
数据来源：电子工程世界、开源证券研究所

2.2 GPU 代际跃升驱动价值扩散，国产算力商业化全面提速

Rubin VR200机柜核心意义并非单柜ASP接近翻倍，而在于AI硬件价值主线已由GPU单点扩张迈入整机柜系统密度扩张阶段

据摩根士丹利测算，GPU仍为最大美元价值池，单柜价值由GB300的252万美元增至396万美元，增幅约57%，仍是ASP提升核心驱动；但占BOM比重由近三分之二回落至约51%，主要为整柜BOM接近翻倍下内存与系统件增速更快所致，并非GPU需求见顶。内存成为第二大价值池，单柜价值由约37.4万美元增至200.2万美元，增幅435%，占比升至26%，HBM、SOCAMM的价格和供应不再只是半导体内部变量，已直接影响整机报价与ODM收入。PCB、MLCC、ABF等二级供应链呈低占比、高弹性特征，增速显著高于整柜；电源与液冷亦由配套部件升级为高密度算力核心约束项。AI硬件价值正从单一算力芯片主导，转向机柜级成分占比与系统复杂度的综合承载。

图10：AI服务器硬件价值主线已由GPU单点扩张迈入整机柜系统密度扩张阶段



数据来源：wallstreetcn、开源证券研究所

表2：VR200服务器硬件成本结构正发生结构性重构

Nvidia NVL72 Bill of Materials	GB300	VR200	Diff.
GPU	$ 2520000	$3960000	57%
CPU	$180000	$180000	0%
NVLink Switch chip	$64800	$144000	122%
Other networking chips	$261000	$576000	121%
Memory	$373939	$2001600	435%
Cooling	$64610	$72080	12%
Power supply	$57600	$76000	32%
PCB	$35100	$116730	233%
ABF Substrate	$11160	$20340	82%
MLCC	$1530	$4320	182%
Others	$402412	$623278	55%
Rack assembly value add	$22400	$28800	29%
Total	$3994551	$7803148	95%

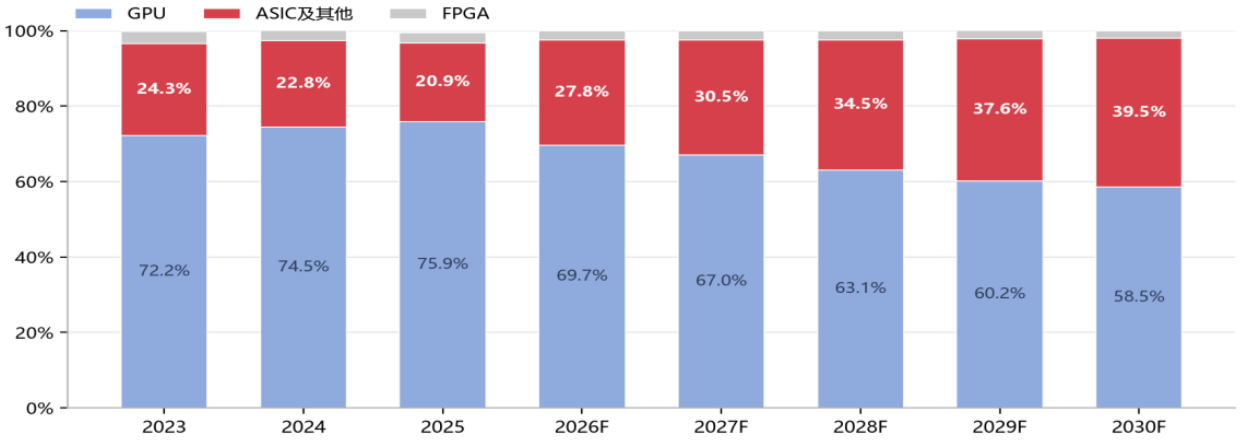
数据来源：电子工程世界、开源证券研究所

2.2 GPU 代际跃升驱动价值扩散，国产算力商业化全面提速

全球AI算力市场正从通用GPU独大走向GPU与定制ASIC并行发展的多元化时代，ASIC出货占比进入陡峭上行通道

全球AI算力市场正从通用GPU独大走向GPU与定制ASIC并行发展的多元化时代，ASIC出货占比进入陡峭上行通道。据TrendForce集邦咨询研究，2026年全球AI服务器出货量预计同比增长28%以上，从所用芯片看，GPU占69.7%仍为最大宗；而在Google、Meta等北美业者积极扩张自研ASIC方案的情况下，预计2026年ASIC AI服务器出货占比将提升至27.8%，为2023年以来最高，出货增速也超越GPU AI服务器。该机构进一步预计，随着Google、Amazon等CSP大力自研ASIC芯片的态势扩大，ASIC AI服务器出货占比将从2026年的27.8%攀升至2030年的近40%。芯片层的供给结构正由单一GPU采购转向GPU与ASIC的组合配置。

图11：AI服务器出货芯片结构:GPU占比缓降,ASIC预计占比升至2030年近40%



数据来源：TrendForce集邦咨询、开源证券研究所

2.2 GPU 代际跃升驱动价值扩散，国产算力商业化全面提速

国产算力规模商用持续深化，系统配套、集群部署与客户应用构成头部厂商的竞争优势

华为昇腾依托完整软硬件体系及广泛行业覆盖，在综合成熟度方面保持领先；海光信息与寒武纪分别凭借服务器平台生态和云端产品积累形成规模商业基础，昆仑芯和平头哥则通过万卡集群交付及云服务部署拓展应用。摩尔线程、沐曦、壁仞、天数智芯和燧原加快交付放量，集群落地与客户拓展同步推进；景嘉微立足成熟图形显控业务，持续拓宽行业计算场景。

算力需求增加、供应链自主与政策牵引构成国产GPU产业链景气上行的三重驱动，国产份额逐步提升

需求侧，大模型训练与推理需求推动智算中心资本投入持续加码，算力需求重心由训练向推理迁移；供给侧，海外高端AI芯片供货受限，倒逼互联网大厂、运营商与政企客户主动导入国产算力。IDC数据显示，2025年中国AI加速卡总出货量约400万张，其中国产厂商合计出货165万张、市场份额达41%，英伟达份额降至55%；2026年上半年英伟达份额进一步降至34%，华为昇腾以28%紧随其后。中商产业研究院预计，2028年前后国产GPU在推理市场占有率将突破40%、训练市场突破25%。

图12：2025年中国AI加速卡份额升至41%，英伟达降至55%



数据来源：AIX财经、开源证券研究所

2.3 算力需求由项目制建设转向持续性消耗，算租承接刚性付费

AI算力需求结构性增长，算力租赁成为灵活供给核心路径

AI大模型从训练向推理加速演进，国内算力需求呈现快速增长态势。随着Agent智能体、多模态大模型的规模化落地，推理侧算力需求快速增长，成为供需缺口的核心来源。中国信通院数据显示，2026年一季度国内AI算力需求同比增长417%，有效供给增速仅128%，供需缺口持续拉大。传统IDC建设周期长达3年以上，难以匹配AI算力快速迭代需求，面临“建成即落后”的扩容窘境。算力租赁以“按需付费、弹性扩容”的轻资产模式应运而生，客户无需大额资金自建GPU集群，可快速补足算力缺口并灵活调配规模与区域，或将成为填补供给缺口、降低企业资本开支的核心路径

训练算力战略地位稳固，高端算力租赁价格攀升维持行业高景气

训练端算力仍是大模型厂商核心竞争力的根本支点，终端用户的核心决策锚点始终围绕模型的实际效果与使用体验，缺乏训练端算力支撑、产品力不足的厂商将快速被主流市场出清。受供给约束与需求增长共振，H100 GPU一年期租赁合约价格从2025年10月的1.70美元/小时低点上涨至2026年3月的2.35美元/小时，累计涨幅近40%；国内B300顶配现货口头报价站上1450万元，H200整机月租同步持续上涨，大模型与智能体项目扩张下训练、微调刚需持续，B系列新品尚不能完全填补高端训练算力缺口，优质算力进入抢货模式，现货上架即被溢价敲定。2026年一季度算力租赁市场规模680亿元，同比增62%，全年预计突破2600亿元，高景气度持续。

算力租赁综合壁垒多维构筑，GPU获取能力成核心护城河

算力租赁行业属重资产驱动赛道，综合竞争力由GPU获取能力、融资能力、订单回款、成本管控及合规风控等多维构筑，行业马太效应持续凸显。GPU获取能力为核心生存壁垒，体现为供给稳定、交付确定、规模充足、品质保障四大维度。英伟达高端GPU在性能、集群及CUDA生态上显著领先国产芯片，高阶认证厂商可获原厂直供配额，配额量相较普通伙伴多30%-50%，交付周期缩短6-12个月，缺乏认证的中小厂商依赖二级市场采购，供需紧张时断供压力加大。头部厂商凭借全维度壁垒优势持续抢占高端市场份额，尾部中小厂商生存空间持续收窄。

2.4 功率密度跃升击穿风冷极限，液冷迈入标配时代

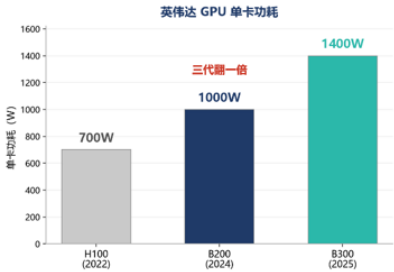
液冷由可选走向标配，高密算力持续打开产业成长空间

AI服务器功耗与单机柜功率密度持续提升，液冷已从传统辅助散热升级为高密算力基础设施的确定性刚需。当前冷板式液冷凭借技术成熟、供应链完善及对现有服务器架构的兼容性占据主流；下一代AI芯片功耗进一步提升，两相冷板式将成为高功率密度场景的重要技术储备，浸没式则持续探索超高密度场景。液冷价值集中于二次侧，CDU、冷板等核心部件壁垒与价值占比较高，产业价值正向核心零部件集中。同时液冷已从服务器配套升级为影响算力释放、能耗与交付效率的核心基础设施。国内液冷产业体系加快完善，多类厂商持续加码，竞争正从单一部件供应向系统集成和全栈散热能力升级。

终端、服务器与光互连协同升级，散热产业成长空间持续打开

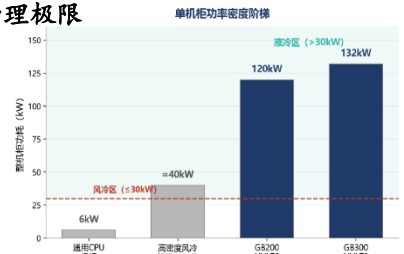
AI算力密度与持续运行负载提升，推动散热由配套环节升级为支撑性能释放的关键基础设施，消费电子、AI服务器与光模块形成多场景需求共振。终端侧，端侧大模型部署带动VC均热板渗透，主动散热与被动均温协同，提升高负载下的持续性能。服务器侧，芯片功耗与机柜功率密度上行，冷板式液冷加速普及，微通道冷板、快速接头与CDU构成核心价值环节；VC与冷板复合、高导热材料及热界面材料升级，进一步拓展单机价值空间。光模块侧，1.6T阶段以超薄VC配合外部风冷为主要方案，高密度场景逐步引入液冷，更高速率演进有望持续提升散热部件价值量。随着散热技术向封装级、芯片级延伸，跨场景工艺复用与核心部件升级有望共同带动产业链成长。

图13：从英伟达H100到B300，单卡功耗翻倍增长



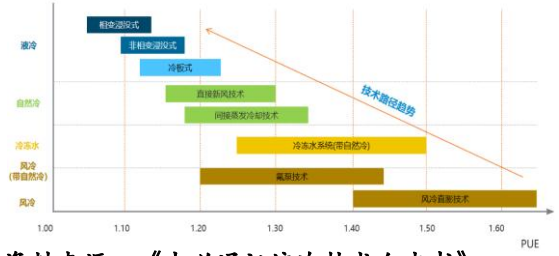
数据来源：Nvidia、Tom's Hardware、开源证券研究所

图14：传统风冷散热极限约在30kW/柜，已触及物理极限



数据来源：ZTE、Vertiv、Supermicro、HPE、开源证券研究所

图15：图3：风冷数据中心已无法满足最新PUE要求，液冷成为唯一可行的技术路径



资料来源：《中兴通讯液冷技术白皮书》

制程微缩的红利正在衰减，性能提升从单点工艺转向架构与封装的系统级协同

① 制程微缩撞上多重约束

寄生延迟、漏电、频率提升困难、布线瓶颈与热密度集中，节点继续推进的技术难度与制造成本同步上升。

关键判断：芯片集成度提高，并不等于系统性能与能效同比例改善。

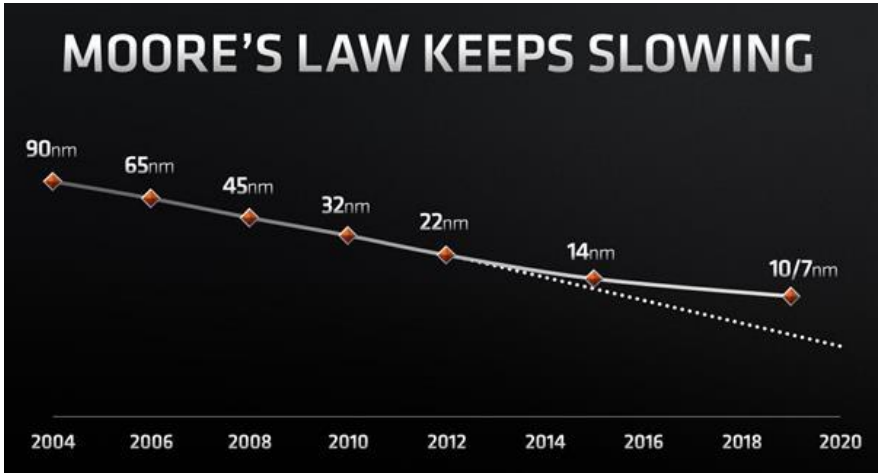
② 瓶颈从算力转向互连

计算单元之间的数据供给与散热，取代晶体管密度成为制约性能释放的关键变量。

③ 出路：Chiplet 拆分 + 先进封装重组

大型 SoC 拆为独立芯粒——计算核心用先进逻辑制程，I/O、电源管理等用适配节点，再在封装层级高密度重组；以更短的信号距离换取性能、能效与成本协同。

图16：制程微缩面临多重约束，芯片性能提升亟需技术创新



资料来源：电子产业世界

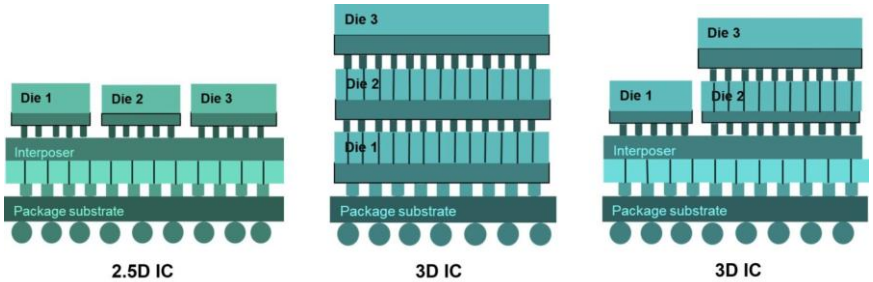
3.2 先进封装突破互连瓶颈，异构集成推动能效提升

封装不再是保护与连接的末端工序，而是决定系统带宽与能效的架构环节

芯粒之间的数据传输效率成为影响系统性能与能效的关键；SiP 把处理器、存储器、传感器与无源件集成于一个封装体，形成完整的系统或子系统，封装由此从物理保护与电气连接变成系统架构的组成部分。

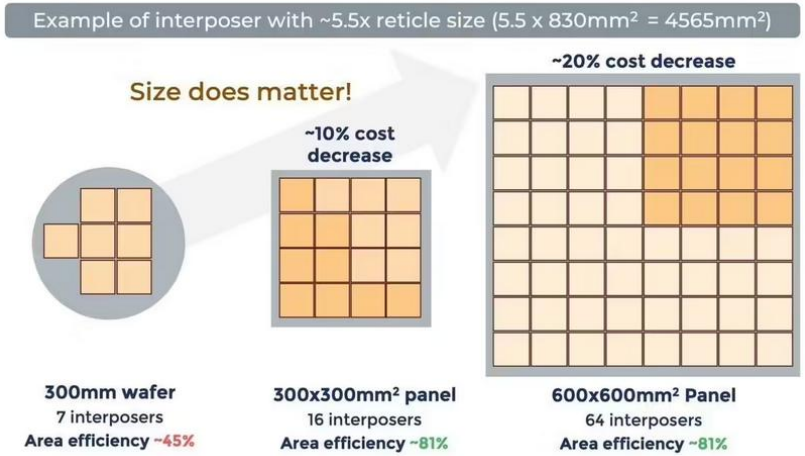
2.5D 通过中介层实现横向高密度互连，CoWoS 把 GPU 与多颗 HBM 装进同一封装；3D 沿垂直方向堆叠，靠 TSV、微凸点与混合键合把间距做细。大尺寸封装又受中介层产能与圆形晶圆边缘损耗制约，CoPoS 转向方形面板摊薄成本。

图17：2.5D封装与3D堆叠协同，实现多芯粒异构集成



资料来源：SIEMENS官网

图18：面板级封装提升面积利用率，为降低封装成本提供空间



资料来源：IT之家

3.3 传统基材性能瓶颈显现，玻璃基板打开材料升级空间

有机载板困于机械稳定性，硅中介层困于制造经济性——玻璃试图同时解决大尺寸、高密度与低损耗

有机（ABF）载板工艺成熟、成本较低，但尺寸稳定性、翘曲控制与热膨胀系数匹配在向大尺寸、高密度演进时面临更大挑战，可能影响细线路加工精度、芯片贴装及焊点可靠性；硅中介层具备高密度布线能力、CTE 与硅芯片匹配，但受晶圆尺寸、加工工艺与制造成本约束，大面积互连结构的扩展空间有限。

玻璃以特种玻璃为基底、通过 TGV 玻璃通孔构建垂直电气连接。机械上刚性与平整度高，CTE 可按材料配方调整（约 3–10ppm/℃），抑制大面积翘曲与图形偏移；电学上绝缘性好、介电损耗低（Dk 约 2.5–3.0，损耗因子比硅低 2–3 个数量级），适配更高带宽与更高互连密度；经济上支持大尺寸方形面板加工，突破圆形晶圆的面积约束，并可复用显示产业在大尺寸玻璃加工、搬运及检测上的积累。

但玻璃并非万能替代：硅中介层胜在布线密度，ABF 胜在成熟度与成本，玻璃承接的是大尺寸、高带宽的增量场景，其成本优势仍需通过专用工艺开发与良率提升逐步兑现。

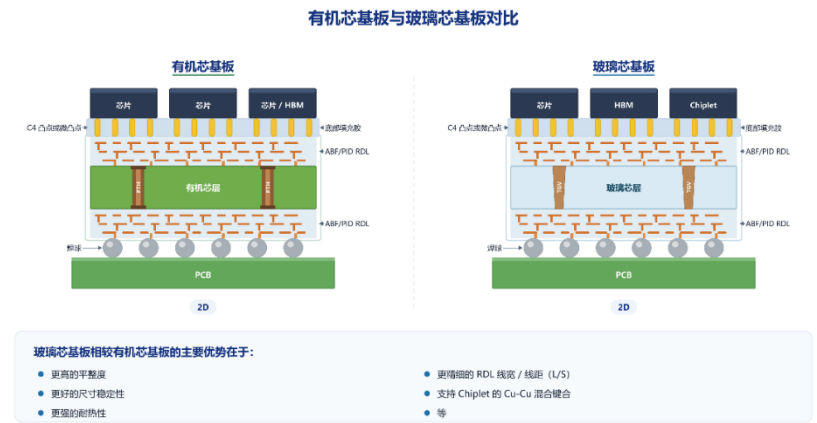
3.3 传统基材性能瓶颈显现，玻璃基板打开材料升级空间

玻璃芯载板改的是底座，玻璃中介层换的是硅桥——同一材料，两种完全不同的导入逻辑

玻璃芯载板用于 AI 加速器等大尺寸芯片封装，主要替代有机载板中的芯层，通过 TGV 贯通玻璃芯层、两面叠加积层与铜线路，充当芯片阵列与 PCB 之间的结构底座，核心是改善尺寸稳定性与翘曲控制。

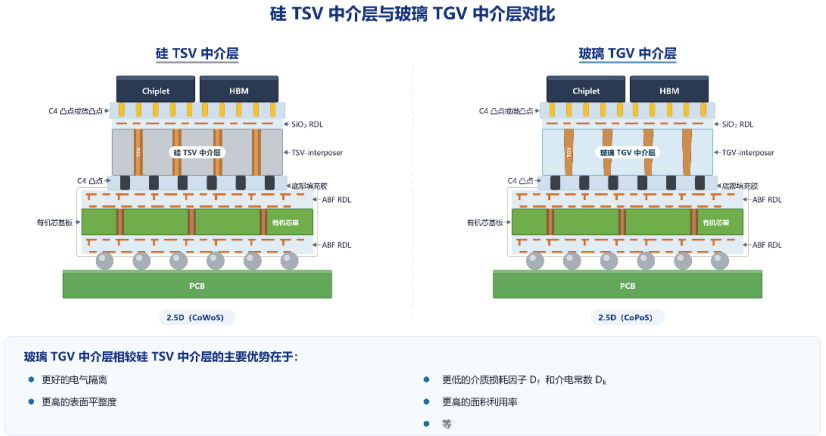
玻璃中介层位于芯片阵列与封装载板之间，承担 GPU 芯粒与多颗 HBM 的高密度互连，是硅中介层的潜在替代方案；此外玻璃还可用于临时载板、玻璃基 IPD 与 FAU 基板，四类应用成熟度差异明显，产业化梯次推进。

图19：玻璃芯载板替代有机芯层，支撑大尺寸芯片封装



资料来源：John H. Lau《Glass Packaging and Its Reliability》、开源证券研究所

图20：玻璃中介层连接GPU与HBM，实现高密度芯片间互连



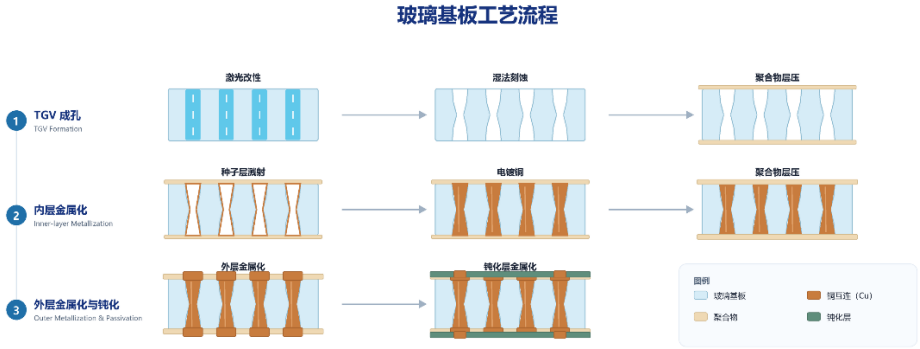
资料来源：John H. Lau《Glass Packaging and Its Reliability》、开源证券研究所

TGV 成孔与金属化决定良率，设备订单先于载板放量——这是跟踪产业化最有效的先行指标

工艺上两个难点：成孔要控制孔位精度、孔壁质量与大面积一致性，避免崩边与微裂纹；填铜要解决种子层附着力难、空洞造成导电失效与应力集中，铜与玻璃的 CTE 差异还会放大界面应力，良率因此成为成本前提。

节奏上，临时载板与玻璃基 IPD 已批量供货，玻璃芯载板加快送样认证（英特尔首款玻璃芯载板处理器已出货，Absolics、三星电机认证中），玻璃中介层仍在研发；设备端激光成孔、电镀及清洗已获订单，产线建设先于载板放量。

图21：玻璃基板制造涉及多道关键工序，全流程协同决定产品良率



资料来源：半导体产业纵横、开源证券研究所

图22：玻璃基板各应用成熟度存在差异，产业化呈梯次推进态势

赛道（成熟度）	玩家	关键节点	当前状态
① 临时载板 已量产 目前已盈利	康宁（美）	全球龙头，市占超60%；12英寸载板月产约20万片；客户覆盖台积电CoWoS及SK海力士HBM产线。	已量产
	AGC（旭硝子）／电气硝子（日）	AGC市占约10%，主攻车载与射频封装；电气硝子聚焦6-8英寸，适配CIS图像传感器封装。	已量产
② 玻璃芯载板 送样—试点 2028-2030年规模量产	英特尔（美）	2026年6月出货首款玻璃芯载板商用处理器Xeon 6+（Clearwater Forest）；里奥兰乔工厂升级35亿美元，计划2030年全面商用。	唯一已出货
	三星电机（韩）	世宗试产线实现TGV深宽比10:1；与东友精细化学合资Glassem（66:34）。量产推迟至2028年以后，现处客户可靠性验证阶段。	原定2025年底量产，已多次推迟
	Absolics（韩资；SKC×应用材料合资）	佐治亚Covington首座玻璃基板厂获CHIPS法案7,500万美元资助；2026年一季度向AMD送样，现于台湾做封装级可靠性评估，目标2026年底至2027年量产。	认证阶段，暂未出货
	新光电气（日）	完成22层玻璃芯基板，玻璃芯两侧各11层铜布线。	工艺攻关
③ 玻璃中介层 原型—研发	Rapidus（日）	2025年发布600×600mm方形玻璃中介层原型；公开规划目标2028年量产。	研发阶段
	三星电子＋三星显示器（韩国）	已组建研发团队，复用面板沉积、光刻及蚀刻能力，主攻玻璃基底RDL图形化；据报道2026年出原型，2028年前导入大规模客户。	研发阶段
④ CPO玻璃基板 刚起步 光电同板	康宁（美）	GlassBridge于2026年6月24日首次发布，与GlobalFoundries合作部署硅光平台。	产品已发布，首批认证待落地

资料来源：HiPi联盟公众号、开源证券研究所

4 风险提示

1. **大模型技术迭代不及预期风险。**多模态感知、复杂推理和智能体执行处于攻坚期，若核心技术突破或模型能力提升慢于预期，可能制约高价值场景适配与功能升级，延缓产业落地。

2. **AI技术应用不及预期风险。**行业数字化基础、客户付费意愿和投入产出验证制约应用落地，若垂直行业适配及企业付费转化慢于预期，将拉长商业价值兑现周期。

3. **AI Infra建设不及预期风险。**算力集群、高速互联和存储配套投入大、周期长、协同要求高，受硬件供给、资金和工程能力约束，若高端算力受限或建设进度偏慢，可能形成供给瓶颈，制约模型训练迭代和应用规模化落地。

4. **技术研发进度不及预期的风险。**技术开发投入较大，若新品研发或核心技术突破慢于预期，投入产出回报可能在较长时期内滞后。

5. **玻璃基板技术产业化进展不及预期风险。**工艺链长、设备与材料协同难度高，若TGV打孔、镀膜、RDL制备的良率提升、产线调试或配套成熟度不及预期，将延缓从样品验证到规模量产。

6. **玻璃基板市场需求不及预期风险。**下游AI芯片及高端封装模组的资本开支和产品迭代影响导入节奏，若算力产品放量放缓、客户改选技术路线或传统方案竞争力较强，需求释放可能延后。

7. **行业竞争加剧的风险。**新兴参与者增加，竞争优势偏弱的企业可能面临出清，中低端品类的毛利率可能承压。

8. **相关政策不及预期风险。**数据安全合规、技术监管、产业扶持及市场准入制度仍在完善，监管趋严、政策落地偏慢或合规成本上升，可能扰动技术迭代、商业化和行业竞争格局。

分析师声明

负责准备本报告以及撰写本报告的所有研究分析师或工作人员在此保证，本研究报告中关于任何发行商或证券所发表的观点均如实反映分析人员的个人观点。负责准备本报告的分析师获取报酬的评判因素包括研究的质量和准确性、客户的反馈、竞争性因素以及开源证券股份有限公司的整体收益。所有研究分析师或工作人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中具体的推荐意见或观点有直接或间接的联系。

特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已于2017年7月1日起正式实施。根据上述规定，开源证券评定此研报的风险等级为R3（中风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C3、C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C3、C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

股票投资评级说明

	评级	说明
证券评级	买入（Buy）	预计相对强于市场表现20%以上；
	增持（Outperform）	预计相对强于市场表现5%～20%；
	中性（Neutral）	预计相对市场表现在－5%～＋5%之间波动；
	减持（Underperform）	预计相对弱于市场表现5%以下。
行业评级	看好（Overweight）	预计行业超越整体市场表现；
	中性（Neutral）	预计行业与整体市场表现基本持平；
	看淡（Underperform）	预计行业弱于整体市场表现。

备注：评级标准为以报告日后的6~12个月内，证券相对于市场基准指数的涨跌幅表现，其中A股基准指数为沪深300指数、港股基准指数为恒生指数、新三板基准指数为三板成指（针对协议转让标的）或三板做市指数（针对做市转让标的）、美股基准指数为标普500或纳斯达克综合指数。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，表示投资的相对比重建议；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。

分析、估值方法的局限性说明

本报告所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。

法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（以下简称“本公司”）的机构或个人客户（以下简称“客户”）使用。本公司不会因接收人收到本报告而视其为客户。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用，如接收人并非开源证券客户，请及时退回并删除。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。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金融工具的邀请或向人做出邀请。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，不应视本报告为做出投资决策的唯一因素。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资目标、财务状况或需要。本公司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。若本报告的接收人非本公司的客户，应在基于本报告做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。

本报告可能附带其它网站的地址或超级链接，对于可能涉及的开源证券网站以外的地址或超级链接，开源证券不对其内容负责。本报告提供这些地址或超级链接的目的纯粹是为了客户使用方便，链接网站的内容不构成本报告的任何部分，客户需自行承担浏览这些网站的费用或风险。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及
的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

本报告的版权归本公司所有。本公司对本报告保留一切权利。除非另有书面显示，否则本报告中的所有材料的版权均属本公司。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。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。

开源证券研究所

上海：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼3层
邮箱：research@kysec.cn
北京：北京市西城区西直门外大街18号金贸大厦C2座9层
邮箱：research@kysec.cn
深圳：深圳市福田区金田路2030号卓越世纪中心1号楼45层
邮箱：research@kysec.cn
西安：西安市高新区锦业路1号都市之门B座5层
邮箱：research@kysec.cn

THANKS

感　　谢　　聆　　听

开源证券