长江证券 CHANGJIANG SECURITIES

| 证券研究报告 |



联合研究 | 行业点评

国内 AI 算力链：模型进步×CSP 资本开支加速，超节点与上游紧缺共振

请阅读最后评级说明和重要声明

报告要点

我们看好中国国内 AI 算力链的投资机会 1）国内大模型加速进步，智谱 GLM-5.2 跻身全球第一梯队，带动 token 调用与应用收入指数级增长；2）国内 CSP 资本开支加速；3）上游紧缺驱动盈利改善，InP 光芯片产能持续短缺，叠加存储/铜连接紧张，龙头量价齐升、毛利率修复；4）国产超节点密集落地，标配液冷并以 LPO/NPO 全光互联拉动铜连接/光互联/交换/电源。

分析师及联系人



邬博华

SAC：S0490514040001

SFC：BQK482



杨洋

SAC：S0490517070012

SFC：BUW100



赵智勇

SAC：S0490517110001

SFC：BRP550



高超

SAC：S0490516080001

SFC：BUX177



黄天佑

SAC：S0490522050005



郭敬超

SAC：S0490525120002

| 证券研究报告 |

2026-07-07

联合研究 | 行业点评

国内 AI 算力链：模型进步×CSP 资本开支加速，超节点与上游紧缺共振

事件描述

智谱 2026 年 6 月发布的旗舰模型 GLM-5.2 表现亮眼：在前端开发盲测榜单 Code Arena 位列全球可用模型第一。“模型进步→token 用量↑→应用收入↑→算力需求↑”的飞轮已启动。

事件评论

- **国内大模型进步驱动 token 与应用兑现：智谱领跑。**智谱 2026 年 6 月发布的旗舰模型 GLM-5.2 表现亮眼，在前端开发盲测榜单 Code Arena 位列全球可用模型第一。叠加 DeepSeek、通义、混元、文心、可灵（Kling）持续迭代，国产模型能力与性价比快速提升。模型能力兑现为 token 调用与商业化加速。智谱营收呈指数级增长，2026 年 1 月登陆港股（02513.HK）、并推进科创板 A 股（拟募约 150 亿元），是国产大模型商业化提速的缩影。国内大模型加速逼近全球第一梯队。

- **CSP 资本开支加速：国内 BAT+字节投入提速。**阿里/腾讯/百度 capex 三大合计由 2024 年 1537 亿元增至 2025 年 2207 亿元（+44%）。字节强劲现金流为超大规模算力投入提供支撑，或将 AI/数据中心资本开支由 2025 年约 250 亿美元，提升至 2026 年最高约 700 亿美元、2027 年或达约 1000 亿美元；国内算力链需求确定性强。

- **供给与盈利：上游紧缺驱动利润率改善。**光芯片/激光器紧缺：AI 数据中心拉动 InP 系激光器需求爆发。Lumentum 预计 AI 数据中心 InP 光通道（lane）需求 2026-2030 年 CAGR 约 85%，产能扩张持续落后于需求、呈结构性短缺；AI 服务器拉动 HBM/DRAM、NOR 等存储需求与涨价；高速铜连接（112G/224G 线缆、背板连接器）供需偏紧。

- **超节点方案落地，拉动铜连接/光互联/交换/液冷。**2025 年以来国产芯片厂商密集发布"超节点"并全部标配液冷（阿里云 UPN512、华为 CloudMatrix384、曙光 scaleX640 等），将数百卡以 Scale-up 高速互联整合为统一算力域，对标英伟达 NVL72/NVL576 机柜级方案，直接拉动铜连接、光互联、交换、电源与液冷的用量与价值量。NPO/CPO 高密度方案对 PIC、CW 光源、保偏光纤、MPO、FAU 等有源/无源环节用量显著高于传统可插拔。

- **投资建议**：国内 AI 算力链"模型进步→token 用量↑→应用收入↑→算力需求↑"的飞轮已启动。——需求来自国产模型与 CSP 真实资本开支、供给紧缺带来量价齐升、超节点带来结构性新增量，板块β与龙头α共振。沿“算力芯片→服务器/一体机→网络（光模块/铜连接/交换）→IDC/算租/云→液冷/电源”布局，优选高需求确定性与技术壁垒的龙头：**①光互联：**推荐中际旭创、新易盛、天孚通信，关注太辰光、长盈通、源杰科技；**②超节点直接受益的铜连接/交换/液冷：**推荐沃尔核材、华丰科技、英维克、紫光股份；**③国产算力芯片与服务器：**推荐寒武纪、海光信息、工业富联，关注中科曙光；**④电源/IDC：**推荐麦格米特、润泽科技。

风险提示

1、云商/CSP 投入 AI 力度不及预期；2、国产算力供给与国际环境变化

3、行业竞争加剧；4、新技术迭代不确定性

更多研报请访问
长江研究小程序

目录

需求侧：国内模型加速进步，CSP 资本开支进入上行周期.....5
国内大模型进步驱动 token 与应用兑现：智谱领跑.....5
CSP 资本开支加速：国内 BAT+字节投入提速.....5
供给与盈利：上游紧缺驱动利润率改善.....6
超节点方案落地，拉动铜连接/光互联/交换/液冷.....7
重点剖析：阿里云 UPN512 超节点.....8
投资建议.....9
风险提示.....10

图表目录

图 1：GLM-5.2 编程能力对标海外一流模型（截至 2026/6/17）（%）.....5
图 2：智谱营业收入（亿元）.....5
图 3：国内三大云商资本开支（亿元）.....6
图 4：字节跳动 AI/数据中心资本开支（亿美元）.....6
图 5：中国智能算力市场规模（亿美元）.....6
图 6：国产 AI 芯片出货量（万张）.....6
图 7：Lumentum Inp 需求 CAGR 达 85%且产能缺口持续.....7
图 8：Lumentum EML 出货量增长预计达 8 倍.....7
图 9：Coherent Inp 产能 2026/2027 连续翻倍.....7
图 10：Coherent 具备超 20 年激光器相关垂直技术积累.....7
图 11：UPN 与 HPN 指标对比及 UPN512 拓扑.....8
图 12：UPN512 系统布局与全光互联介质.....8
图 13：UPN512 不同规模下 Scale-up 网络成本对比（截至 2025 年 10 月）.....9

表 1：国产芯片厂商超节点方案.....8

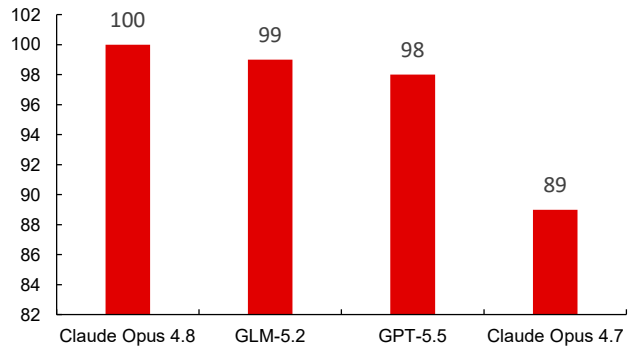
需求侧：国内模型加速进步，CSP 资本开支进入上行周期

国内大模型进步驱动 token 与应用兑现：智谱领跑

国内大模型加速逼近全球第一梯队。智谱 2026 年 6 月发布的旗舰模型 **GLM-5.2** 表现亮眼：在前端开发盲测榜单 Code Arena 位列全球可用模型第一；在 FrontierSWE 等编程基准上仅落后 Claude Opus 4.8 约 1%、并反超 GPT-5.5 约 1%、领先 Claude Opus 4.7 约 11%，综合编程能力介于 Opus 4.7 与 4.8 之间、稳居开源 SOTA（前代 GLM-4.6 在 AIME25/GPQA/LiveCodeBenchv6/SWE-bench Verified 分别达 93.9%/81.0%/82.8%/68.0%，对标 Claude Sonnet 4.6）。叠加 DeepSeek、通义、混元、文心、可灵（Kling）持续迭代，国产模型能力与性价比快速提升。

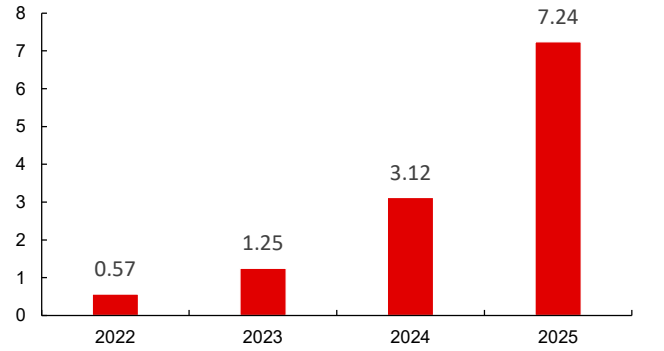
模型能力兑现为 token 调用与商业化加速。智谱营收呈指数级增长：2023/2024/2025 年分别为 1.25/3.12/7.24 亿元；2026 年 1 月登陆港股（02513.HK）、并推进科创板 A 股（拟募约 150 亿元），是国产大模型商业化提速的缩影。"模型进步→token 用量↑→应用收入↑→算力需求↑"的飞轮已启动。

图 1：GLM-5.2 编程能力对标海外一流模型（截至 2026/6/17）（%）



资料来源：智谱，长江证券研究所

图 2：智谱营业收入（亿元）



资料来源：量子位，智谱公司年报，长江证券研究所

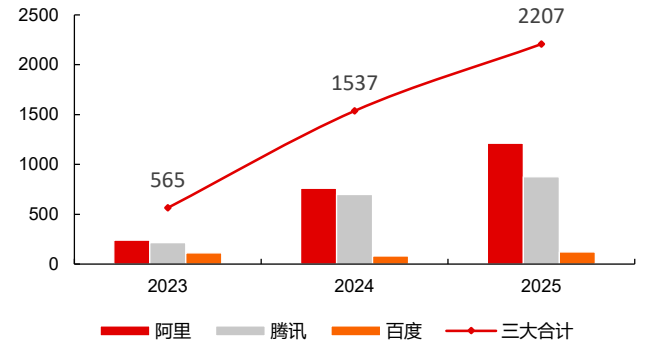
CSP 资本开支加速：国内 BAT+字节投入提速

国内三大云商资本开支进入上行周期。据公司公告+Bloomberg 预测，阿里/腾讯/百度 capex 2024 年为 759/697/81 亿元、2025 年升至 1212/875/121 亿元，三大合计由 2024 年 1537 亿元增至 2025 年 **2207 亿元（+44%）**；2026E/2027E 三大合计进一步升至 2454/2789 亿元。其中阿里以自研（磐久超节点+平头哥 ASIC）与阿里云驱动、capex 强度最高，腾讯 2026E 大幅提速，百度受昆仑芯自给支撑稳步增长。

字节跳动（未上市）是国内最大的增量变量。据 Bloomberg（2026-05-27），在利润驱动下（2025 年净利约 500 亿美元），字节正讨论将 AI/数据中心资本开支由 2025 年约 250 亿美元，提升至 **2026 年最高约 700 亿美元、2027 年或达约 1000 亿美元**（含海外、以数据中心为主）；字节 2024 年全球营收约 1550 亿美元（+29%）、2025 年估约

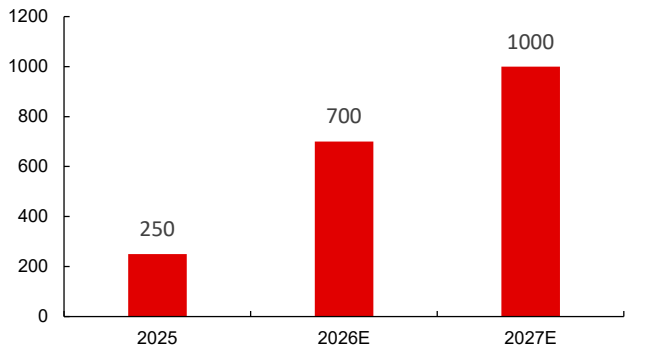
1860 亿美元，强劲现金流为超大规模算力投入提供支撑。BAT+字节共振，国内算力链需求确定性强。

图 3：国内三大云商资本开支（亿元）



资料来源：公司公告，Bloomberg，长江证券研究所

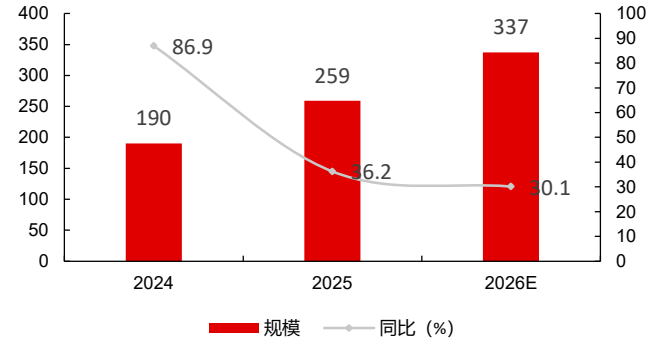
图 4：字节跳动 AI/数据中心资本开支（亿美元）



资料来源：Bloomberg，zol，长江证券研究所

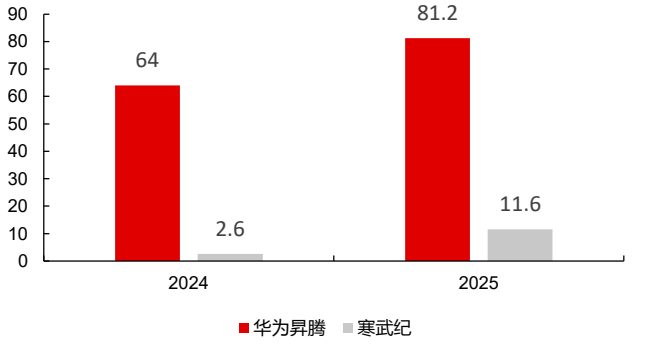
从市场规模与国产替代看，IDC 预计中国智能算力市场 2024/2025/2026E 为 190/259/337 亿美元（2026E 约为 2024 年的 1.77 倍）；H20 受限后国产算力补位，2025 年本土 AI 加速卡份额首破 41%，华为昇腾出货约 81 万张、寒武纪约 11.6 万张，IDC 招标重启、token 需求创新高，服务器/交换/IDC/液冷等配套环节充分受益。

图 5：中国智能算力市场规模（亿美元）



资料来源：IDC，长江证券研究所

图 6：国产 AI 芯片出货量（万张）

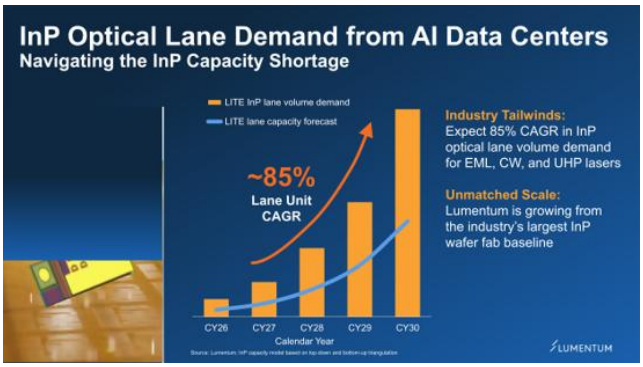


资料来源：新浪财经，长江证券研究所

供给与盈利：上游紧缺驱动利润率改善

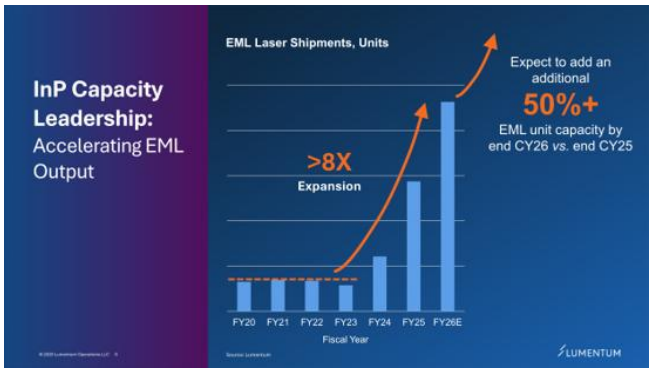
光芯片/激光器紧缺：AI 数据中心拉动 InP 系激光器需求爆发。Lumentum 预计 AI 数据中心 InP 光通道（lane）需求 2026-2030 年 CAGR 约 85%，产能扩张持续落后于需求、呈结构性短缺；其 EML 激光器出货量自 FY20 以来已扩张超 8 倍，并预计 2026 年底 EML 单元产能较 2025 年底再增 50%+。Coherent 规划 InP 产能 2026、2027 连续两年翻倍以应对"前所未有的需求"，并量产业界首条 6 英寸 InP 产线，垂直一体化覆盖 EML/CW/VCSEL。

图 7：Lumentum Inp 需求 CAGR 达 85%且产能缺口持续



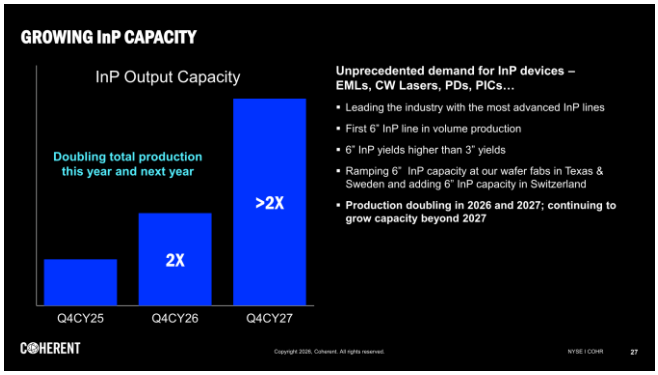
资料来源：Lumentum，长江证券研究所

图 8：Lumentum EML 出货量增长预计达 8 倍



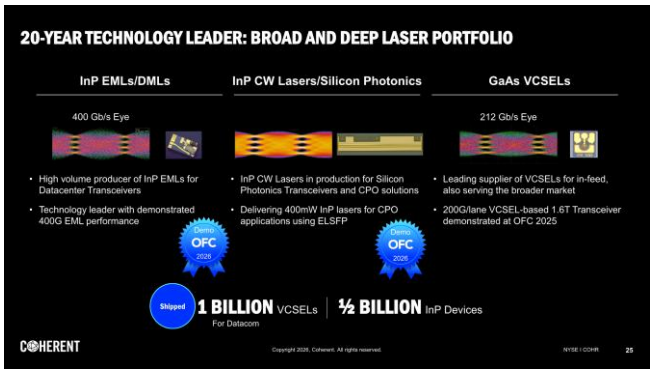
资料来源：Lumentum，长江证券研究所

图 9：Coherent Inp 产能 2026/2027 连续翻倍



资料来源：Coherent，长江证券研究所

图 10：Coherent 具备超 20 年激光器相关垂直技术积累



资料来源：Wind，长江证券研究所

国产光芯片量价齐升：国内源杰科技、仕佳光子分别拟投 12.51/12.65 亿元建设高速光芯片产业化项目，东山精密子公司索尔思拟投 12 亿美元扩产，高端 EML/CW 光源国产供给提升、份额与价格同步上行。

存储与铜连接同步紧张：AI 服务器拉动 HBM/DRAM、NOR 等存储需求与涨价；高速铜连接（112G/224G 线缆、背板连接器）供需偏紧。上游紧缺背景下，具备物料与供应链掌控力的龙头议价权增强，毛利率/净利率自上游向中游修复——这是本轮算力链盈利质量提升的核心。

超节点方案落地，拉动铜连接/光互联/交换/液冷

2025 年以来国产芯片厂商密集发布"超节点"并全部标配液冷（阿里云 UPN512、华为 CloudMatrix384、曙光 scaleX640 等），将数百卡以 Scale-up 高速互联整合为统一算力域，对标英伟达 NVL72/NVL576 机柜级方案，直接拉动铜连接、光互联、交换、电源与液冷的用量与价值量。

表 1：国产芯片厂商超节点方案

方案名称	厂商	算力卡/规模	发布时间
CloudMatrix 384	华为云	384×昇腾 910C NPU+192×鲲鹏 CPU（BF16≈300PFLOPs），可组 16 万卡集群	2025-07（WAIC）
磐久 128（Panjiu）/UPN512 超节点	阿里云	磐久 128：单柜 128－144 颗 AI 芯片，无背板正交架构+单相浸没液冷；UPN512 单层（单跳）即可互联最大 512 张 xPU、并预留扩展至 1024 卡及以上（基于 High Radix 以太交换、复用以太成熟生态）	2025-09（云栖）
scaleX640	中科曙光	单机柜 640 卡（全球首个），16 节点互连成 10240 卡万卡集群（>5EFLOPs）	2025-12

资料来源：阿里云白皮书，量子位，云栖大会，中科曙光官网，财联社，长江证券研究所

重点剖析：阿里云 UPN512 超节点

阿里云 **UPN512 技术架构白皮书**提出的 UPN（Ultra Performance Network）是面向 Scale-up 的超节点网络系统：单层（单跳）即可互联最大 **512 张 xPU**、并预留扩展至 1024 卡及以上（基于 High Radix 以太交换、复用以太成熟生态）；单 xPU 的 scale-up 带宽达 **6.4T**（对比 HPN 的 400G/800G），采用 Load/Store over ETH+ 内存语义互联与低时延 ETH+协议。

图 11：UPN 与 HPN 指标对比及 UPN512 拓扑

	HPN	UPN
GPU带宽	400G / 800G	6.4T
互联协议	基于Message，RDMA / Solar，over ETH+	Load / Store over ETH+
互联规模	万卡 / 十万卡	512 / 1024

图2. HPN 与 UPN

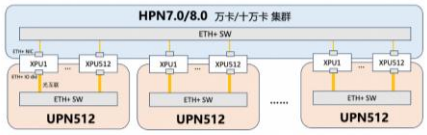


图3. HPN 和 UPN 组成的高性能系统

资料来源：阿里云白皮书，长江证券研究所

图 12：UPN512 系统布局与全光互联介质

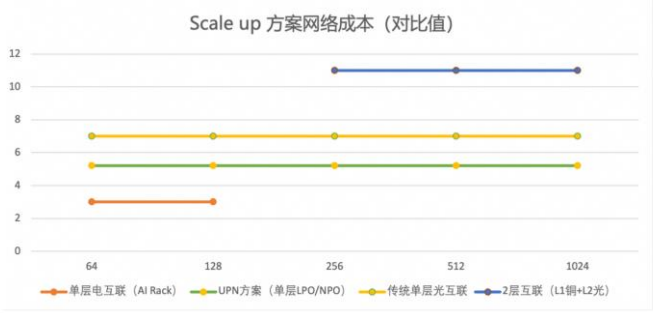


介质/形态	100G/lane 典型覆盖距离	200G/lane 典型覆盖距离	适用
无源铜缆 DAC	2 m	1m	单柜-AI Rack
有源铜缆ACC(均衡/放大，无重定时)	5 m	3m	2～4柜-AI Rack
有源AEC(带重定时的主动电缆)	7 m	5m	
光模块-MMF（VR多模）	50m	50m	多柜-UPN512
光模块-MMF（SR多模）	100m	100m	
光模块-SMF（DR单模）	500m	500m	

资料来源：阿里云白皮书，长江证券研究所

架构上 UPN512 采用盒式（box）化、设备与机柜解耦的单层全光互联：机柜内/近距（2－7m）用无源 DAC/有源 AEC 铜缆，跨柜用多模（50－100m）/单模（约 500m）光模块，并优选 **LPO（线性直驱）/NPO（近封装）**以降低功耗与时延（NPO 当前 3.2T、未来可演进至 6.4T）。相比传统两层（L1 铜+L2 光）方案，UPN 单层光互联在规模放大时成本显著更低（见图 13），直接拉动 LPO/NPO 光模块、单模/多模光器件与以太交换芯片需求；且 LPO/NPO 去 DSP 后链路更稳（月化抖动链路比例 LPO 约 2x、FRO 约 6x，铜缆为基准 1x）。

图 13：UPN512 不同规模下 Scale-up 网络成本对比（截至 2025 年 10 月）



资料来源：阿里云白皮书，长江证券研究所

Scale-up 光互连产业弹性：Lumentum 预测 25-30 年 AI 光通信 TAM 从 180 亿美元增至超 900 亿美元（CAGR 约 40%）；NPO/CPO 高密度方案对 PIC、CW 光源、保偏光纤、MPO、FAU 等有源/无源环节用量显著高于传统可插拔。

投资建议

总览：我们看好中国国内 AI 算力链——国内模型进步（智谱 GLM-5.2 跻身全球第一梯队）驱动 token 与应用兑现；BAT+字节资本开支共振上行；上游光芯片（持续短缺）、存储与连接紧缺推动龙头量价齐升、利润率改善；国产超节点（阿里的 UPN512 等）以 LPO/NPO 全光互联密集落地，拉动铜连接/光互联/交换/液冷价值量。

配置主线：沿“算力芯片→服务器/一体机→网络（光模块/铜连接/交换）→IDC/算租/云→液冷/电源”布局，优选高需求确定性与技术壁垒的龙头：①光互联：推荐中际旭创、新易盛、天孚通信，关注太辰光、长盈通、源杰科技；②超节点直接受益的铜连接/交换/液冷：推荐沃尔核材、华丰科技、英维克、紫光股份；③国产算力芯片与服务器：推荐寒武纪、海光信息、工业富联，关注中科曙光；④电源/IDC：推荐麦格米特、润泽科技。

核心逻辑：本轮区别于以往行情，关键在于"国内自主驱动+盈利兑现"——需求来自国产模型与 CSP 真实资本开支、供给紧缺带来量价齐升、超节点带来结构性新增量，板块β与龙头α共振。

风险提示

1、**云商/CSP 投入 AI 力度不及预期：**若国内外云商对 AI 大模型与算力的资本开支不及预期，算力链各环节增长可能放缓。

2、**国产算力供给与国际环境变化：**先进制程、HBM、设备与 EDA 等受限可能影响国产芯片供给节奏；大国博弈与潜在制裁带来不确定性。

3、**行业竞争加剧：**光模块、服务器、IDC、液冷等环节竞争激烈，若龙头技术迭代或成本控制不及预期，盈利能力可能承压。

4、**新技术迭代不确定性：**CPO/硅光/液冷/HVDC 等新技术良率、成本与客户验证进度存在不确定性，可能影响放量节奏。

投资评级说明

行业评级	报告发布日后的 12 个月内行业股票指数的涨跌幅相对同期相关证券市场代表性指数的涨跌幅为基准，投资建议的评级标准为：	
	看　　好：	相对表现优于同期相关证券市场代表性指数
	中　　性：	相对表现与同期相关证券市场代表性指数持平
	看　　淡：	相对表现弱于同期相关证券市场代表性指数
公司评级	报告发布日后的 12 个月内公司的涨跌幅相对同期相关证券市场代表性指数的涨跌幅为基准，投资建议的评级标准为：	
	买　　入：	相对同期相关证券市场代表性指数涨幅大于 10%
	增　　持：	相对同期相关证券市场代表性指数涨幅在 5%～10%之间
	中　　性：	相对同期相关证券市场代表性指数涨幅在-5%～5%之间
	减　　持：	相对同期相关证券市场代表性指数涨幅小于-5%
	无投资评级：	由于我们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，致使我们无法给出明确的投资评级。

相关证券市场代表性指数说明：沪深两市以沪深 300 指数为基准；北交所市场以北证 50 指数为基准；新三板市场以三板成指为基准；香港市场以恒生指数为基准。

办公地址

上海
Add /虹口区新建路 200 号国华金融中心 B 栋 22、23 层
P.C /（200080）

武汉
Add /武汉市江汉区淮海路 88 号长江证券大厦 37 楼
P.C /（430023）

北京
Add /朝阳区景辉街 16 号院 1 号楼泰康集团大厦 23 层
P.C /（100020）

深圳
Add /深圳市福田区中心四路 1 号嘉里建设广场 3 期 36 楼
P.C /（518048）

分析师声明

本报告署名分析师以勤勉的职业态度，独立、客观地出具本报告。分析逻辑基于作者的职业理解，本报告清晰准确地反映了作者的研究观点。作者所得报酬的任何部分不曾与，不与，也不将与本报告中的具体推荐意见或观点而有直接或间接联系，特此声明。

法律主体声明

本报告由长江证券股份有限公司及/或其附属机构（以下简称「长江证券」或「本公司」）制作，由长江证券股份有限公司在中华人民共和国大陆地区发行。长江证券股份有限公司具有中国证监会许可的投资咨询业务资格，经营证券业务许可证编号为：10060000。本报告署名分析师所持中国证券业协会授予的证券投资咨询执业资格书编号已披露在报告首页的作者姓名旁。

在遵守适用的法律法规情况下，本报告亦可能由长江证券经纪（香港）有限公司在香港地区发行。长江证券经纪（香港）有限公司具有香港证券及期货事务监察委员会核准的“就证券提供意见”业务资格（第四类牌照的受监管活动），中央编号为：AXY608。本报告作者所持香港证监会牌照的中央编号已披露在报告首页的作者姓名旁。

其他声明

本报告并非针对或意图发送、发布给在当地法律或监管规则下不允许该报告发送、发布的人员。本公司不会因接收人收到本报告而视其为客户。本报告的信息均来源于公开资料，本公司对这些信息的准确性和完整性不作任何保证，也不保证所包含信息和建议不发生任何变更。本报告内容的全部或部分均不构成投资建议。本报告所包含的观点、建议并未考虑报告接收人在财务状况、投资目的、风险偏好等方面的具体情况，报告接收者应当独立评估本报告所含信息，基于自身投资目标、需求、市场机会、风险及其他因素自主做出决策并自行承担投资风险。本公司已力求报告内容的客观、公正，但文中的观点、结论和建议仅供参考，不包含作者对证券价格涨跌或市场走势的确定性判断。报告中的信息或意见并不构成所述证券的买卖出价或征价，投资者据此做出的任何投资决策与本公司和作者无关。本研究报告并不构成本公司对购入、购买或认购证券的邀请或要约。本公司有可能会与本报告涉及的公司进行投资银行业务或投资服务等其他业务(例如:配售代理、牵头经办人、保荐人、承销商或自营投资)。

本报告所包含的观点及建议不适用于所有投资者，且并未考虑个别客户的特殊情况、目标或需要，不应被视为对特定客户关于特定证券或金融工具的建议或策略。投资者不应以本报告取代其独立判断或仅依据本报告做出决策，并在需要时咨询专业意见。

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、价值及投资收入可升可跌，过往表现不应作为日后的表现依据；在不同时期，本公司可以发出其他与本报告所载信息不一致及有不同结论的报告；本报告所反映研究人员的不同观点、见解及分析方法，并不代表本公司或其他附属机构的立场；本公司不保证本报告所含信息保持在最新状态。同时，本公司对本报告所含信息可在不发出通知的情形下做出修改，投资者应当自行关注相应的更新或修改。本公司及作者在自身所知情范围内，与本报告中所评价或推荐的证券不存在法律法规要求披露或采取限制、静默措施的利益冲突。

本报告版权仅为本公司所有，本报告仅供意向收件人使用。未经书面许可，任何机构和个人不得以任何形式翻版、复制和发布给其他机构及/或人士（无论整份和部分）。如引用须注明出处为本公司研究所，且不得对本报告进行有悖原意的引用、删节和修改。刊载或者转发本证券研究报告或者摘要的，应当注明本报告的发布人和发布日期，提示使用证券研究报告的风险。本公司不为转发人及/或其客户因使用本报告或报告载明的内容产生的直接或间接损失承担任何责任。未经授权刊载或者转发本报告的，本公司将保留向其追究法律责任的权利。

本公司保留一切权利。