



通信行业研究

买入（维持评级）

行业周报
证券研究报告

通信组

分析师：张真桢（执业 S1130524060002）

zhangzhenzhen@gjzq.com.cn

字节计划 700 亿美元加码 AI 基建，Marvell 业绩超预期

通信周观点：

1) 华为在 2026 国际电路与系统研讨会上正式提出“ τ 定律”，以“时间缩微”替代“几何缩微”，通过逻辑折叠提升晶体管密度。过去六年已量产 381 款芯片；预计 2031 年高端芯片晶体管密度达 1.4 纳米制程同等水平。看好国产 AI 芯片放量带动国产算力链加速向上。2) 高通与字节跳动达成 AI ASIC 合作，字节将采购数百万颗定制芯片，用于 AI 算力支持。3) 小米宣布 MiMo-V2.5 系列大模型 API 永久降价，最高降幅达 99%，且不再区分上下文窗口长度。其中 MiMo-V2.5 Pro 输入缓存命中价格降至 0.025 元/百万 tokens，输出价格最低降至 6 元/百万 tokens；同步优化 Token Plan 计费体系，用量提升至原来的 5 至 8 倍。4) 台积电 3 纳米芯片下半年报价将调涨 15%，明年可能再涨 5% 至 10%。受 AI 加速器、ASIC、手机芯片等需求同步涌入，3 纳米产能持续满载，客户排队状况未明显缓解，月产能已拉升至 16 万至 17.5 万片，AI 算力需求旺盛。5) 字节跳动据悉考虑今年支出至多 700 亿美元用于 AI 基础设施建设，资金部分来自 2025 年约 500 亿美元利润；若条件有利，明年资本支出或将提高至约 1000 亿美元。6) Snowflake 与 AWS 签订五年 60 亿美元采购协议，涵盖 AWS 自研 Graviton 芯片及 GPU。公司第一财季营收 13.9 亿美元，同比增长 33%，全面超预期。此外，Snowflake 宣布收购 AI 初创公司 Natoma，进一步强化 AI 布局。7) Marvell 发布 2027 财年第一季度财报，营收 24.18 亿美元，同比增长 28%，数据中心业务贡献 18.3 亿美元，占总营收 76%。公司大幅上调 2028 财年营收目标至约 165 亿美元。此外，Marvell 宣布与英伟达扩大战略合作，涵盖光学互联、NVLink Fusion 集成及 AI-RAN 等领域。8) 苹果加码推进端侧 AI，依托自研芯片在 iPhone、Mac 等终端设备本地运行 AI 模型，以降低隐私泄露风险并减少云端算力依赖。同时正探索通过模型蒸馏训练轻量化模型，并考虑收购端侧 AI 初创公司 Liquid AI。9) 戴尔科技 2026 财年全年营收 1135 亿美元，同比增长 19%，创历史新高。第四季度 AI 优化服务器收入 90 亿美元，同比增长 342%；预计 2027 财年 AI 服务器收入将翻倍至约 500 亿美元。10) 和硕股东会表示，今年 AI 伺服器业务将实现 10 倍成长，下半年开始显效；美国工厂已完成测试，预计下半年启动生产。

细分赛道：

服务器：本周服务器指数-4.23%，本月以来，服务器指数-0.31%。戴尔科技 2026 财年第四季度 AI 优化服务器收入 90 亿美元，同比增长 342%。AI 服务器发展略超市场预期，全球 AI 算力需求持续增长。

光模块：本周光模块指数+2.45%，本月以来光模块指数+17.98%。Marvell 数据中心业务营收 18.3 亿美元，占总营收 76%，同时与英伟达扩大战略合作，涵盖光学互联、NVLink Fusion 集成及 AI-RAN 等领域，利好光模块及上游光器件。

IDC：本周 IDC 指数-1.38%，本月以来，IDC 指数+6.50%。华为提出“ τ 定律”，通过逻辑折叠提升晶体管密度，预计 2031 年高端芯片密度达 1.4 纳米制程同等水平；字节跳动考虑今年支出至多 700 亿美元用于 AI 基础设施建设，明年或提高至约 1000 亿美元；高通与字节达成 AI ASIC 合作，字节将采购数百万颗定制芯片。国产算力链加速向上。

核心数据更新：

电信业务量收增速逐步提升。2026 年 1-4 月电信业务收入累计完成 5941 亿元，同比下降 1.7%。按照上年不变价计算的电信业务总量同比增长 8%。2026 年 4 月我国光模块出口数据同比+34.58%；1-4 月累计同比+10.7%。

投资建议与估值

建议关注国内 AI 发展带动的服务器、IDC 等板块，以及海外 AI 发展带动的服务器、光模块等板块。

风险提示

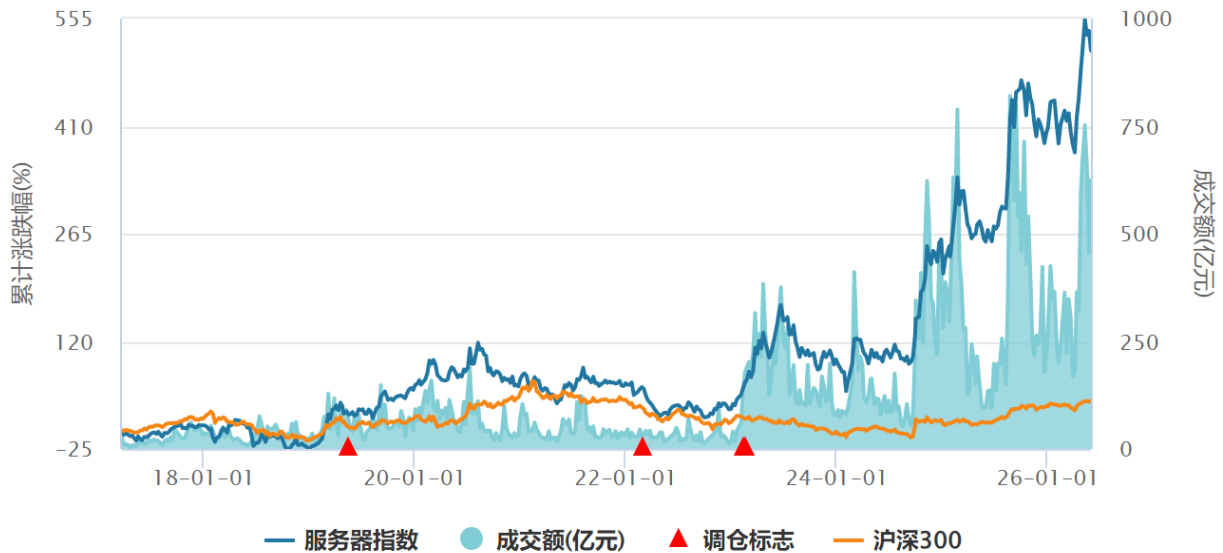
AI 建设不及预期、中美关税波动加剧、原材料供应不足。



一、细分行业观点

服务器：本周服务器指数-4.23%，本月以来，服务器指数-0.31%。戴尔科技 2026 财年第四季度 AI 优化服务器收入 90 亿美元，同比增长 342%。和硕今年 AI 伺服器业务将实现 10 倍成长，美国工厂预计下半年启动生产。AI 服务器发展略超市场预期，全球 AI 算力需求持续增长。

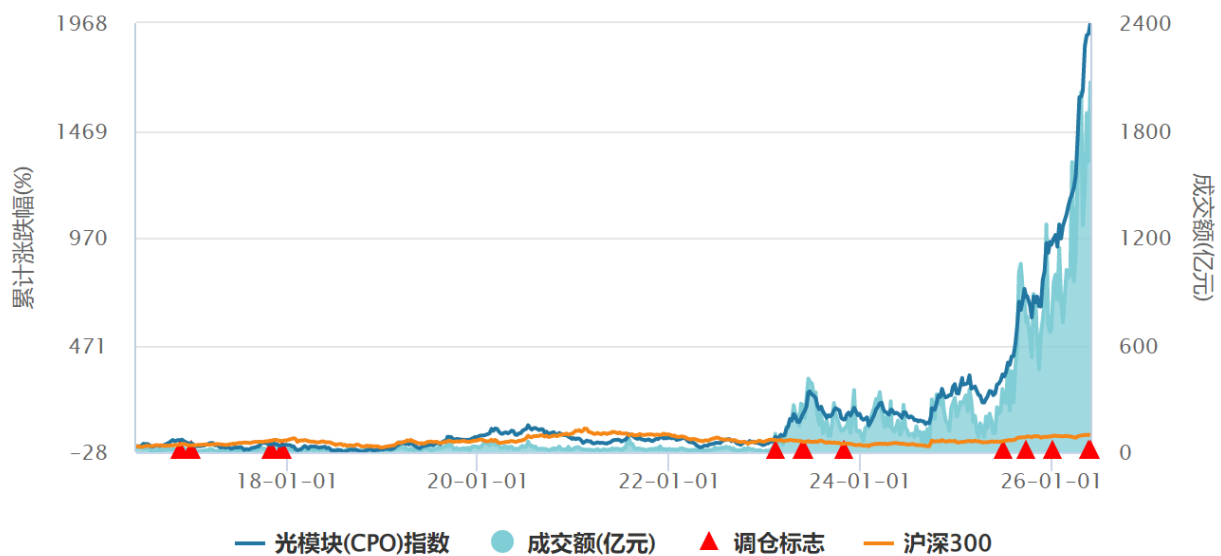
图表1：服务器指数 (8841058.WI) 走势



来源：wind，国金证券研究所

光模块：本周光模块指数+2.45%，本月以来光模块指数+17.98%。Marvell 数据中心业务营收 18.3 亿美元，占总营收 76%，同时与英伟达扩大战略合作，涵盖光学互联、NVLink Fusion 集成及 AI-RAN 等领域，利好光模块及上游光器件。

图表2：光模块(CPO)指数 (8841258.WI) 走势



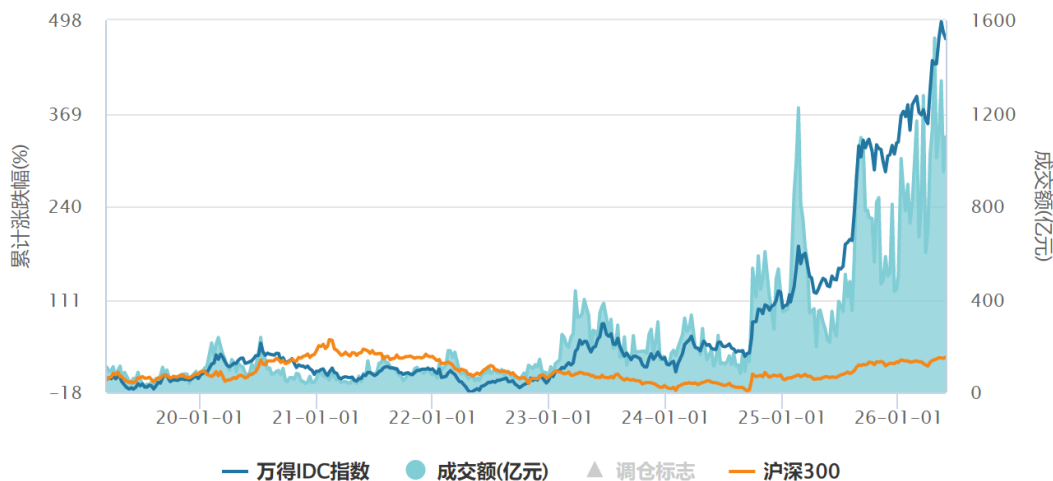
来源：wind，国金证券研究所

IDC：本周 IDC 指数-1.38%，本月以来，IDC 指数+6.50%。华为提出“辐（T）定律”，通过逻辑折叠提升晶体管密度，预计 2031 年高端芯片密度达 1.4 纳米制程同等水平；字节跳动考虑今年支出至多 700 亿美元用于 AI 基础设施建设，



明年或提高至约 1000 亿美元；高通与字节达成 AI ASIC 合作，字节将采购数百万颗定制芯片。国产算力链加速向上。

图表3: IDC 指数 (866052.WI) 走势



来源: wind, 国金证券研究所

图表4: 本周通信板块景气度

板块	景气度指标	本期景气度说明
运营商	稳健向上	5G 投资周期结束，电信行业端承压，但云与 IDC 业务放量接力成长，整体景气度稳健向上。
光模块	稳健向上	Marvell 在业绩会上表示已与多家客户就 6.4T 光引擎展开合作，看好 NPO 在 scaleup 领域的渗透。
服务器	稳健向上	戴尔预计今年 AI 伺服器业务将实现 10 倍成长，下半年开始显效，AI 需求持续加速。
交换机	稳健向上	思科已从超大规模客户获得 53 亿美元订单，并将全年订单预期上调至 90 亿美元。看好交换机作为 AI 重要硬件放量。
连接器	稳健向上	康宁计划将美国光连接产品制造产能提升 10 倍并将美国光纤产能扩大 50% 以上，同时新建三座先进制造工厂。推动光纤、连接器等需求增长。
IDC	加速向上	字节考虑今年支出至多 700 亿美元用于 AI 基础设施建设，看好国产算力链加速向上。
物联网	加速向上	苹果加码推进端侧 AI，依托自研芯片在 iPhone、Mac 等终端设备本地运行 AI 模型，同时正考虑收购端侧 AI 初创公司 Liquid AI。
液冷	高景气维持	谷歌正与中国英维克等企业洽谈采购数据中心液冷系统，液冷板块再受催化。

来源: 华尔街见闻, 新浪财经, 国金证券研究所

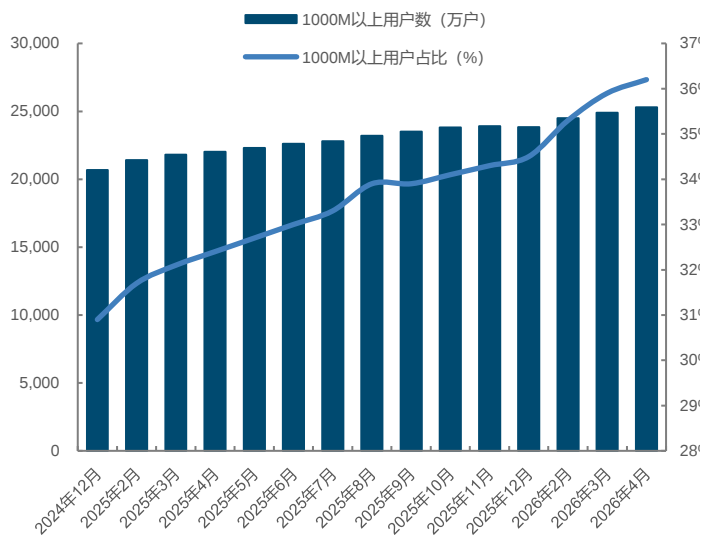
二、核心数据更新

运营商: 运营商数据维持稳健增长

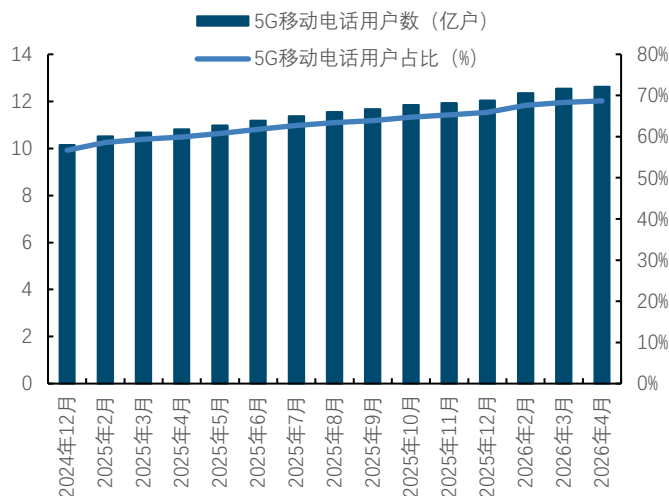
截至 4 月末，三家基础电信企业及中国广电的移动电话用户总数达 18.38 亿户，比上年末净增 1092 万户。其中，5G 移动电话用户达 12.62 亿户，比上年末净增 5758 万户，占移动电话用户的 68.7%。



图表5：千兆用户占比超三成



图表6：截至4月末5G用户占比超六成

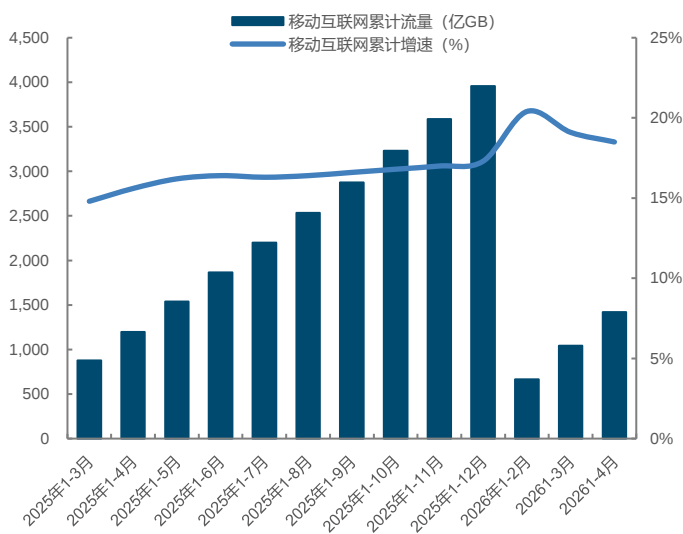


来源：工信部，国金证券研究所

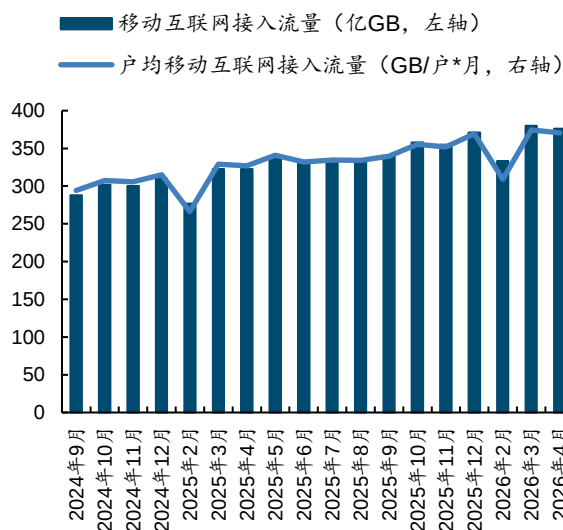
来源：工信部，国金证券研究所

前4个月，移动互联网累计流量达1420亿GB，同比增长18.5%。截至4月末，移动互联网用户数达16.24亿户，比上年末净增1416万户。4月当月户均移动互联网接入流量(DOU)达到23.17GB/户·月，同比增长13.5%，比上年底高0.13GB/户·月。

图表7：前4个月移动互联网累计流量同比增长18.5%



图表8：4月当月DOU达23.17GB/户·月



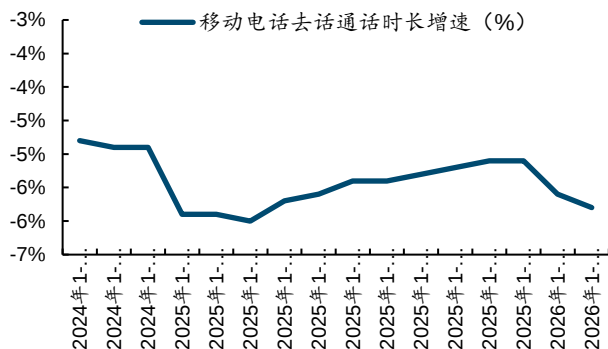
来源：工信部，国金证券研究所

来源：工信部，国金证券研究所

前4个月，移动电话去话通话时长完成6292亿分钟，同比下降5.8%；固定电话主叫通话时长完成182.2亿分钟，同比下降21.5%。前4个月，全国移动短信业务量同比下降5.8%；移动短信业务收入同比下降11.5%。

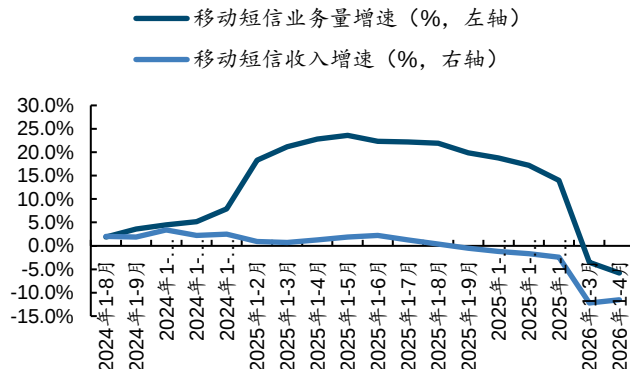


图表9：电话通话量持续下滑



来源：工信部，国金证券研究所

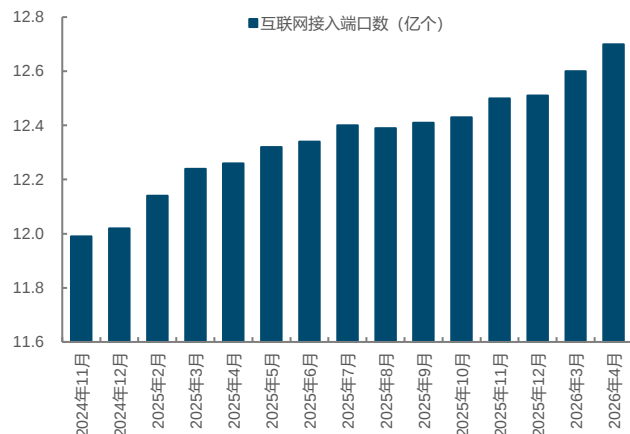
图表10：移动短信业务量大幅下降



来源：工信部，国金证券研究所

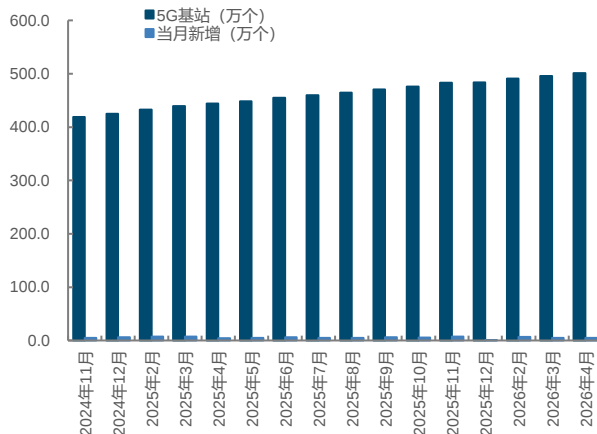
截至4月末，全国互联网宽带接入端口数量达12.7亿个，比上年末净增1894万个。其中，光纤接入（FTTH/O）端口达到12.3亿个，比上年末净增1669万个，占互联网宽带接入端口的96.6%。截至4月末，具备千兆网络服务能力的10G PON端口数达3235万个，比上年末净增72.6万个。

图表11：千兆光纤宽带网络建设稳步推进



来源：工信部，国金证券研究所

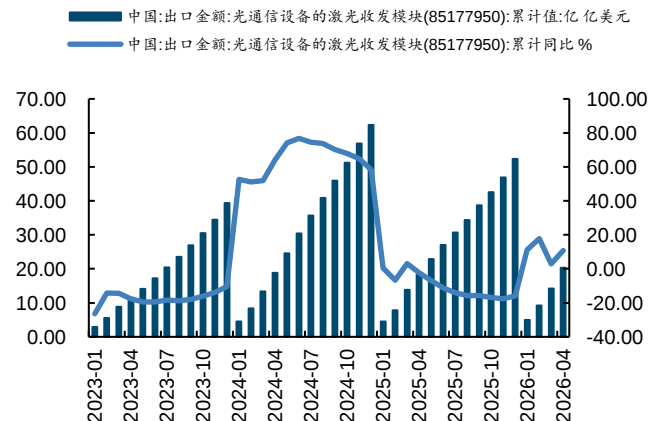
图表12：5G网络建设持续深化



来源：工信部，国金证券研究所

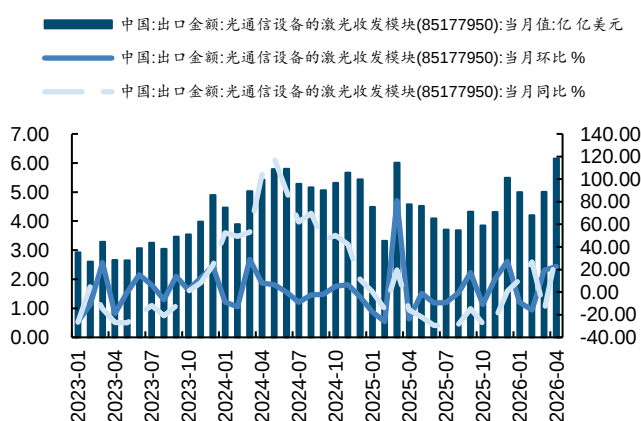
光模块数据：2026年4月我国光模块出口数据同比+34.58%；1-4月累计同比+10.7%。

图表13：4月光模块出口金额累计同比增加10.7%



来源：wind，国金证券研究所

图表14：4月光模块出口金额当月同比+34.58%

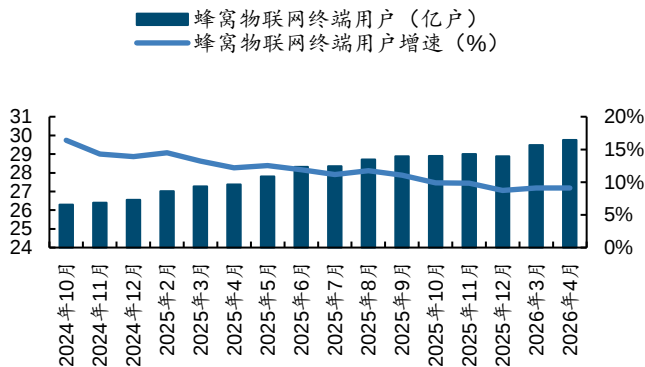


来源：wind，国金证券研究所

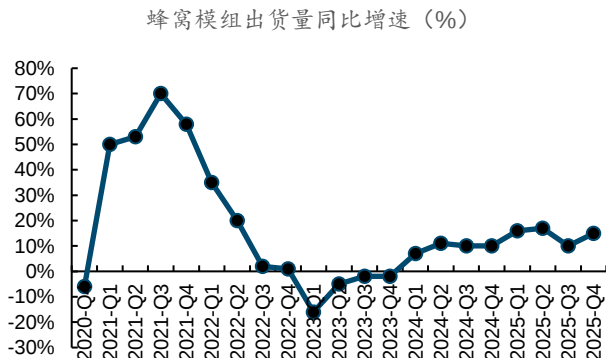


物联网数据：截至4月末，三家基础电信企业发展移动物联网终端用户 29.76 亿户，比上年末净增 8763 万户。互联网电视（IPTV、OTT）用户数达 4.1 亿户，比上年末净增 145.6 万户。

图表15：截至4月末蜂窝物联网终端用户数同比增长8.11%



图表16：2025年全球蜂窝物联网模组出货量同比增长15%



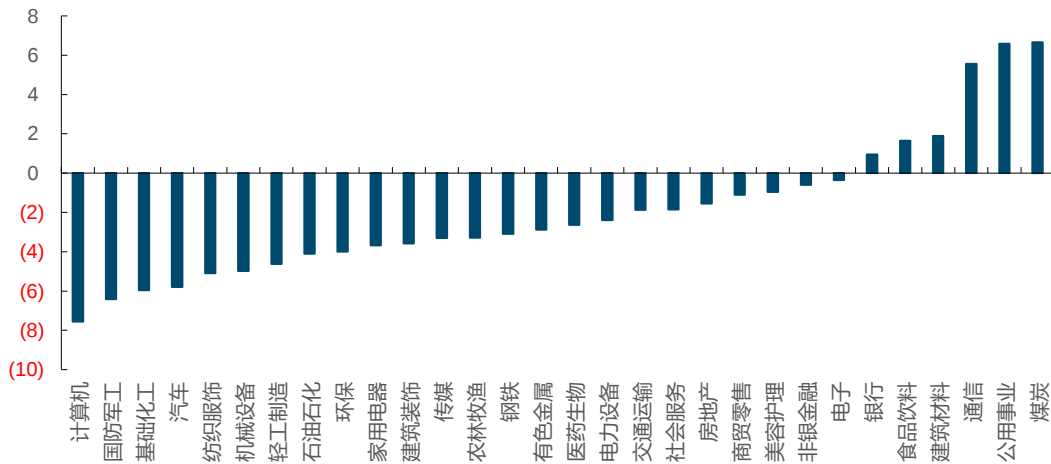
来源：工信部，国金证券研究所

来源：国金证券数字未来实验室，RFID 世界网，C114 通信网，国金证券研究所

三、本周行情

回顾本周行情（5月25日-5月29日），参考申万一级行业划分，通信板块涨跌幅为+5.56%，排名全行业第3。

图表17：板块周涨跌幅排序 (%)

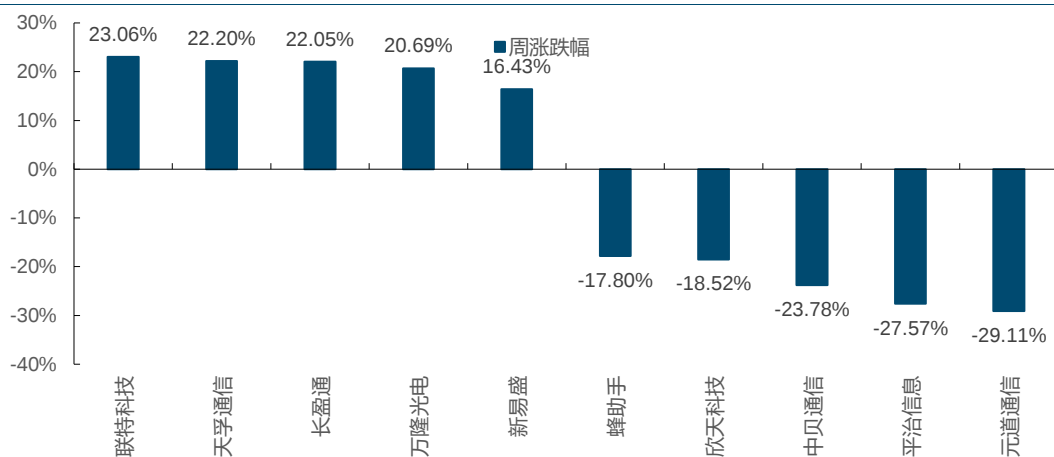


来源：wind，国金证券研究所

从个股情况来看，本周联特科技、天孚通信、长盈通、万隆光电、新易盛为通信（申万）涨幅前五大公司，涨跌幅分别为+23.06%、+22.20%、+22.05%、+20.69%、+16.43%。通信（申万）跌幅前五为蜂助手、欣天科技、中贝通信、平治信息、元道通信，涨跌幅分别为-17.80%、-18.52%、-23.78%、-27.57%、-29.11%。



图表18: 通信板块个股周涨跌幅 (剔除*ST公司)



四、本周重要新闻

4.1 行业新闻

华为：

5月25日财联社讯，在2026国际电路与系统研讨会（上海）上，华为董事、半导体业务部总裁何庭波发表题为《半导体新路径探索与实践》的主旨演讲，正式提出“ t 定律”。这是中国在全球半导体领域首次提出指导产业发展的新原则。该定律以“时间缩微”替代“几何缩微”，通过逻辑折叠等创新技术持续压缩信号时延，构建从器件到系统的多层级协同优化体系。据披露，基于该定律，华为过去六年已成功设计并量产了381款芯片。今年秋季，华为将发布新一代麒麟手机芯片，完整采用逻辑折叠技术，大幅提升性能。预计到2031年，基于该定律的高端芯片晶体管密度将达到1.4纳米制程的同等水平。

高通：

5月27日新浪财经讯，据彭博社报道，高通（Qualcomm）已与字节跳动达成AI ASIC合作协议。字节跳动将向高通采购数百万颗定制芯片，为其AI服务提供算力支持。另一消息人士透露，该交易将帮助字节跳动将已完成的内部芯片设计转变为生产就绪的半导体。此前高通在4月末表示，将在今年向某超大规模云服务商交付首款ASIC，分析师曾猜测字节跳动和亚马逊为其ASIC设计服务客户。

小米：

小米5月27日宣布，旗下自研大模型MiMo-V2.5系列API迎来永久性降价，最高降幅达99%，且不再区分上下文窗口长度。具体来看，MiMo-V2.5标准版输入缓存命中价格降至0.005元/百万tokens，输出价格降至2元/百万tokens，降幅分别为98%和93%；MiMo-V2.5 Pro输入缓存命中价格降至0.025元/百万tokens，输出价格降至6元/百万tokens，降幅分别为99%和86%。除API单价下调外，小米同步优化了Token Plan计费体系，在价格不变的前提下，可用用量提升至原来的5至8倍，并引入“Credits”概念，让计费规则更加清晰易懂。此前于4月28日启动的“百万亿Token创造者激励计划”，已面向全球AI用户在30天内发放合计100万亿免费Token，该计划已于5月26日提前圆满收官。支撑本轮大幅降价的背后，是小米技术团队在推理系统层面的持续突破。小米基于SGLang HiCache完整支持SWA（滑动窗口注意力机制），将KV Cache在GPU显存、CPU内存、SSD等多级存储之间的数据搬运量降低至优化前的近1/7，同时将可缓存token数量提升至近5倍，显著提升了缓存命中率和推理效率。此外，通过优化专家并行方案和输入长度分桶策略，集群输入吞吐能力进一步提升，在保障服务质量的同时持续压低单位token服务成本。

台积电：

5月27日工商时报讯，供应链传出，台积电下半年将再度调涨3纳米报价，涨幅上看15%，2027年亦可能再涨5%至10%。随AI加速器、客制化ASIC、旗舰手机芯片与高效能运算（HPC）需求同步涌入，3纳米产能持续满载，先进制程已成为全球半导体竞争最核心的战略资源。MSCI最新半年度调整于5月29日盘后生效，台积电成为MSCI台湾指数权重调升最多个股，一次调高0.56个百分点，调整后权重达58.33%。外资目前持股比重高达70.35%。摩根士丹利证券同日邀请台积电参加投资人说明会；紧接着6月4日股东会登场，董事长魏哲家预计再度就AI需求、先进制程与海外布局“说清楚、讲明白”。ASIC业者指出，这波3纳米涨价并非单一客户急单带动，而是整体先进制程供需结构出现根本性变化。过去3纳米主要由智能手机SoC支撑，但随着英伟达、AMD、谷歌、AWS等全面加速导入3纳米，投片需求快速升温。供应链透露，台积电3纳米主力厂区Fab18稼动率维持高位，年初月产能约13万片，第二季逐步拉升至16万至17.5万片，但AI需求成长速度仍远超预期。此外，海外设厂成本提高、先进制程折旧压力增加，以及2纳米量产前期良率持续爬坡，调涨价格也有助支撑毛利率表现。目前3纳米已成为AI芯片最成熟且稳定的量产节点，相较仍在初期的2纳米，对AI客户而言更具量产与成本优势。



4.2 公司新闻

Snowflake:

5月27日，云数据平台 Snowflake 宣布与亚马逊 AWS 扩大合作，签署一份为期五年的 60 亿美元采购协议，涵盖亚马逊自研 Arm 架构 Graviton 通用芯片及 AI 用 GPU。受此消息刺激，Snowflake 盘后股价暴涨逾 30%。这是继 Anthropic 承诺十年内在 AWS 投入超 1000 亿美元后，AWS 在 AI 领域赢得的又一重大客户承诺。同日公布的 2026 财年第一季度业绩全面超预期：调整后每股收益 39 美分，营收 13.9 亿美元，同比增长 33%，分别高于分析师预期的 32 美分和 13.2 亿美元。公司预计第二财季产品营收 14.15 亿至 14.20 亿美元，调整后运营利润率 12.5%，均超越市场预期。此外，Snowflake 还宣布以未披露金额收购 AI 初创公司 Natoma，进一步强化 AI 布局。此次 60 亿美元协议意味着 Snowflake 平均每年在 AWS 上支出约 12 亿美元。相较过往，Snowflake 上市时曾披露与某云服务商的五年 12 亿美元合同，该合作方正是亚马逊；2023 年协议规模扩大至 25 亿美元；本次进一步跃升至 60 亿美元。与亚马逊对 Anthropic 及 OpenAI 的协议不同，此次交易不涉及股权投资。扩大采购的核心之一——亚马逊 Graviton 芯片正加速渗透数据中心。随着 AI 应用从问答式聊天机器人向任务导向的智能体 AI 转型，具备通用计算能力的 CPU 需求复苏。亚马逊 CEO Andy Jassy 此前表示，Graviton 能让 Meta 以所需的性能和效率运行智能体 AI 背后的 CPU 密集型工作负载。Meta 此前也表示将调用数十万颗 Graviton 芯片，谷歌和微软随后也推出了各自的自研 Arm 芯片，Arm 架构正加速替代 x86 架构。

Marvell:

5月28日华尔街见闻报道，AI 定制芯片与数据中心互联龙头 Marvell 发布 2027 财年第一季度财报，营收创纪录并再度大幅上调全年指引。CEO Matt Murphy 在电话会上直言“我们的数据中心业务在燃烧”，“订单异常强劲”。财报显示，Q1 营收 24.18 亿美元，同比增长 28%，环比增长 9%，略超分析师预期的 24.1 亿美元；非 GAAP 每股收益 0.80 美元，与预期持平。GAAP 净利润为 3450 万美元，较上年同期下滑，主要因收购 Celestial AI 和 XConn 带来的一次性费用及非现金摊销。数据中心业务贡献营收 18.3 亿美元，占总营收 76%，同比增长 27%，环比增长 11%。公司同步上调 2028 财年营收目标至约 165 亿美元，较上季度指引高出约 15 亿美元，对应同比增长约 45%。Murphy 在财报声明中表示，AI 相关订单异常强劲，数据中心业务增速正在加快——2026 财年数据中心增长 46%，预计 2027 财年约 50%，2028 财年约 55%。互联业务成为最大亮点，增速从 30%→50%→70%，Murphy 称“还有向上偏的空间”。随着推理模型和混合专家模型等复杂架构部署，AI 集群内部数据传输量急剧增加，网络互联重要性大幅提升。Marvell 传统 DSP 业务明年将大幅跃升，1.6T 产品线、DCI 加速、retimer、AEC 及规模化上行光学等新业务共同构成重大增长周期的开端。定制芯片业务方面，现有 XPU 旗舰项目持续增长，超过 10 个 XPU 附属项目进入更高量产阶段，一个新头部 XPU 项目已进入量产。Murphy 确认分析师关于 2028 财年定制芯片营收超 40 亿美元、2029 财年跳升至超 100 亿美元的判断。本季度公司还宣布与英伟达扩大战略合作，涵盖三大方向：共同开发硅光子技术（纵向扩展网络关键使能技术）、NVLink Fusion 集成允许 Marvell 构建与英伟达基础设施无缝对接的自定义芯片和网络半导体、以及 AI-RAN（在相同硬件平台上同时运行 5G/6G 与 AI 负载）。供应链方面，公司计划在 2027 财年进行约 10 亿美元供应商预付款，首批从 Q2 开始。COO 表示，AI 相关一切自始至终供应受限，Marvell 与少数核心供应商建立紧密关系，提供五年需求预测并用预付款支撑。Q1 经营性现金流达创纪录的 6.39 亿美元，公司回购股票 2 亿美元并支付股息 5400 万美元，期末净债务/EBITDA 为 0.32 倍。

戴尔科技:

2月27日经济观察网讯，戴尔科技发布 2026 财年第四季度及全年财报，多项核心数据创历史新高，AI 服务器业务增长显著超出市场预期。财报发布后，戴尔股价当日收盘大涨 21.93%，成交额约 48.3 亿美元，与美股大盘走势形成反差。财报显示，2026 财年第四季度营收达 334 亿美元，同比增长 39%，高于市场预期的 317 亿美元；非 GAAP 每股收益 3.89 美元，同比增长 45%。全年营收 1135 亿美元，同比增长 19%，创历史纪录。第四季度 AI 优化服务器收入为 90 亿美元，同比增长 342%，成为增长核心动力。公司预计 2027 财年 AI 服务器收入将翻倍至约 500 亿美元，同比增长 103%。截至 2027 财年初，AI 相关积攒订单达 430 亿美元，创历史新高。公司同时宣布将现金股利上调 20%，强化资本回报。机构分析指出，AI 数据中心需求持续强劲，戴尔作为英伟达等芯片厂商的核心合作伙伴，在服务器领域的技术优势和订单能见度为短期增长提供支撑。但需注意，存储芯片供应紧张可能对传统 PC 业务造成压力。

和硕联合科技:

5月29日工商时报讯，和硕召开股东会，共同执行长郑国彦表示，看好今年 AI 服务器业务较去年将有 10 倍数增长，上半年处于调整阶段，下半年将展现显著成果。共同执行长郑光志透露，因应客户需求，下半年将投入更多生产人力与空间，人力方面可能需增加 3 至 4 倍，方能在“秋收”季节实现 10 倍增长。此外，美国工厂已完成测试，预计下半年启动生产。在业务模式上，和硕已完成内部组织调整，将服务器业务梳理为四种执行模式：云服务供应商（CSP）模式、新兴云端业者（NeoCloud）模式、ODM/Off-the-Shelf（标准化企业级产品）模式及纯代工模式。郑光志指出，新云端客户下半年将开花结果，并产生外溢效应，吸引介于 CSP 与新兴云端业者之间的企业及 ODM 项目。关于供应链，郑国彦表示，PC 端记忆体大多由客户拉料，虽整体出货量下滑，但高阶产品比重增加，营收未受明显影响。郑光志则指出，CPU 供应紧张程度可能比记忆体更高，但和硕掌握大客户，且自去年底提前布局，目前无缺料隐忧。针对市场传闻 NVIDIA 拟入股和硕 200 亿元，董事长童子贤澄清，入股在法律与资本结构上为严谨名词，目前并无发行新股让对方认购。他表示，当前国际客户常见合作模式有两种：一是直接给予高额预付款；二是客户出资购买专用设备、产权归客户，搬入台厂运作，后续从产品售价中分摊扣除，两者皆为鼓励投资之意。

4.3 海内外大厂重点跟踪



英伟达：

5月21日澎湃新闻讯，AI芯片龙头英伟达发布2027财年第一财季财报，期内营收816.15亿美元，同比增长85%，高于市场预期的789亿美元；GAAP净利润583.21亿美元，同比大增211%；Non-GAAP每股摊薄收益1.87美元，超出市场预期的1.76美元。公司第一财季毛利率为75%。数据中心业务营收752亿美元，同比增长92%，占总营收92%，高于市场预期的731亿美元。其中超大规模数据中心运营商贡献380亿美元，AI云业务收入同比增长超三倍。公司当季通过股票回购和分红向股东返还200亿美元，并新批800亿美元回购授权，季度股息大幅上调至每股0.25美元。展望未来，英伟达预计2027财年第二财季销售额达910亿美元（上下浮动2%），高于市场平均预期的870亿美元。CEO黄仁勋表示，下一代AI计算平台Vera Rubin开局强劲，预计从2027财年下半年开始出货，“将比Grace Blackwell更成功”，且产品生命周期内供应持续紧张。他称，代理式AI已到来，在企业与行业中快速普及，AI工厂建设正以惊人速度加速推进。首席财务官Colette Kress则表示，英伟达目标成为“全球领先CPU供应商”，新款Vera CPU将创造高达2000亿美元的新业务，预计今年CPU总收入达200亿美元。此外，公司第一财季向AI模型制造商和相关基础设施基金投资186亿美元。对于定制AI芯片，黄仁勋表示收购Groq后推出的Groq 3 LPU为小众产品，用例并不广泛。财报发布后，英伟达盘后股价震荡，一度下跌约3%后跌幅收窄。

苹果：

5月29日新浪科技讯，据家电网报道，在下月举行的苹果年度开发者大会上，一系列延期已久的iPhone人工智能功能升级将成为重头戏。外界预计，苹果将重点展示其十五年来为iPhone、Apple Watch、Mac自主研发定制芯片的技术积累，依托全球海量苹果设备直接在终端设备上运行AI模型，发力脱离云端运行模式。知情人士透露，苹果正利用谷歌完整版Gemini大模型，通过模型蒸馏技术训练可在苹果设备本地运行的轻量化模型。同时，苹果也在物色可协助完成AI模型轻量化改造、适配终端设备运行的小型企业，总部位于美国马萨诸塞州剑桥市的端侧AI初创公司Liquid AI已进入收购考察名单。部分Siri指令仍将依托谷歌云平台、调用谷歌Gemini大模型的授权版本运行，苹果近期许可在该场景下使用英伟达的隐私保护技术（机密计算系统），可在数据与模型运算过程中全程加密，处理部分基于Gemini大模型的复杂运算。苹果在终端本地运行AI模型，既能降低用户数据泄露风险，也能杜绝广告公司利用个人信息牟利，对企业客户还可减少令牌使用量、降低成本。对苹果自身而言，将更多AI运算下放至终端设备，可避免像其他科技巨头那样在数据中心领域投入巨额资金——2025年Meta全年资本开支达720亿美元，微软达880亿美元，而同期苹果仅为127.2亿美元。阿雷特研究公司分析师估算，全球苹果终端芯片所汇聚的算力规模相当于一套价值500亿美元的算力资源。苹果前高级工程项目经理马克·苏曼表示，数十亿台苹果设备整合在一起本身就是一股强大的AI算力，苹果有能力搭建全球规模最大的边缘计算AI体系。自2024年推出Apple Intelligence系列AI功能后，该板块进展趋于沉寂，新功能市场反响平平，新版Siri再度延期上线。然而，随着行业开始担忧盲目堆砌云端算力，部分技术从业者对苹果相对审慎的布局思路逐渐改观。奥斯汀AI初创企业webAI首席执行官戴维·斯托特表示，AI技术正朝着轻量化方向发展，绝大多数运算任务终将转移至边缘终端，苹果在这一点上押对了方向。

风险提示

- 1、AI商业价值不及预期的风险：目前AI市场应用仍处于初级阶段，盈利模式仍需探索，市场尚未成熟。若商业模式无法持续发展新客户，需求大幅减弱，或市场接受度偏低，可能对营业收入造成较大负面影响，损害相关公司的盈利能力及产品或服务的商业价值。
- 2、技术发展速度不及预期的风险：目前AI模型的使用仍受限于诸多因素，在特定领域无法达到预期的提高生产力效果。该领域目前仍面临较大的技术挑战，包括模型训练效果不稳定、算法不成熟等问题。若技术落地不及预期，可能影响AI的应用领域和运行效率，造成较大的投资损失。
- 3、供应链集中度过高的风险：AI行业基础设施建设目前高度依赖某几家核心供应商，极易受到相关供应商供应短缺的影响。此外在训练方面，AI技术依赖于大量优质数据的输入。不可靠、低质量的数据来源会一定程度上影响AI模型训练的性能，同时提高训练过程中的不可控成本。
- 4、行业监管加剧的风险：目前生成式AI工具仍存在法律、伦理、安全风险。AI生成内容的产权问题仍存在较大争议。各国可能针对AI的使用及AI生成内容进行更严格的监管及抵制，影响投资预期，并阻碍AI技术在产业上进一步落地。公司面临法律诉讼和声誉受损等负面影响风险。
- 5、市场竞争加剧的风险：在如今巨头科技公司加大AI投入，大量创业公司涌入竞争的大环境下，技术的迅速迭代及新算法的涌现可能使得公司技术迅速落后竞争对手，影响相关公司的市场份额和投资回报的稳定性。



行业投资评级的说明：

买入：预期未来 3—6 个月内该行业上涨幅度超过大盘在 15%以上；
增持：预期未来 3—6 个月内该行业上涨幅度超过大盘在 5%—15%；
中性：预期未来 3—6 个月内该行业变动幅度相对大盘在 -5%—5%；
减持：预期未来 3—6 个月内该行业下跌幅度超过大盘在 5%以上。



特别声明：

国金证券股份有限公司经中国证券监督管理委员会批准，已具备证券投资咨询业务资格。

本报告版权归“国金证券股份有限公司”（以下简称“国金证券”）所有，未经事先书面授权，任何机构和个人均不得以任何方式对本报告的任何部分制作任何形式的复制、转发、转载、引用、修改、仿制、刊发，或以任何侵犯本公司版权的其他方式使用。经过书面授权的引用、刊发，需注明出处为“国金证券股份有限公司”，且不得对本报告进行任何有悖原意的删节和修改。

本报告的产生基于国金证券及其研究人员认为可信的公开资料或实地调研资料，但国金证券及其研究人员对这些信息的准确性和完整性不作任何保证。本报告反映撰写研究人员的不同设想、见解及分析方法，故本报告所载观点可能与其他类似研究报告的观点及市场实际情况不一致，国金证券不对使用本报告所包含的材料产生的任何直接或间接损失或与此有关的其他任何损失承担任何责任。且本报告中的资料、意见、预测均反映报告初次公开发布时的判断，在不作事先通知的情况下，可能会随时调整，亦可因使用不同假设和标准、采用不同观点和分析方法而与国金证券其它业务部门、单位或附属机构在制作类似的其他材料时所给出的意见不同或者相反。

本报告仅为参考之用，在任何地区均不应被视为买卖任何证券、金融工具的要约或要约邀请。本报告提及的任何证券或金融工具均可能含有重大的风险，可能不易变卖以及不适合所有投资者。本报告所提及的证券或金融工具的价格、价值及收益可能会受汇率影响而波动。过往的业绩并不能代表未来的表现。

客户应当考虑到国金证券存在可能影响本报告客观性的利益冲突，而不应视本报告为作出投资决策的唯一因素。证券研究报告是用于服务具备专业知识的投资者和投资顾问的专业产品，使用时必须经专业人士进行解读。国金证券建议获取报告人员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。报告本身、报告中的信息或所表达意见也不构成投资、法律、会计或税务的最终操作建议，国金证券不就报告中的内容对最终操作建议做出任何担保，在任何时候均不构成对任何人的个人推荐。

在法律允许的情况下，国金证券的关联机构可能会持有报告中涉及的公司所发行的证券并进行交易，并可能为这些公司正在提供或争取提供多种金融服务。

本报告并非意图发送、发布给在当地法律或监管规则下不允许向其发送、发布该研究报告的人员。国金证券并不因收件人收到本报告而视其为国金证券的客户。本报告对于收件人而言属高度机密，只有符合条件的收件人才能使用。根据《证券期货投资者适当性管理办法》，本报告仅供国金证券股份有限公司客户中风险评级高于 C3 级（含 C3 级）的投资者使用；本报告所包含的观点及建议并未考虑个别客户的特殊状况、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。对于本报告中提及的任何证券或金融工具，本报告的收件人须保持自身的独立判断。使用国金证券研究报告进行投资，遭受任何损失，国金证券不承担相关法律责任。

若国金证券以外的任何机构或个人发送本报告，则由该机构或个人为此发送行为承担全部责任。本报告不构成国金证券向发送本报告机构或个人的收件人提供投资建议，国金证券不为此承担任何责任。

此报告仅限于中国境内使用。国金证券版权所有，保留一切权利。

上海

电话：021-80234211

邮箱：researchsh@gjzq.com.cn

邮编：201204

地址：上海浦东新区芳甸路 1088 号

紫竹国际大厦 5 楼

北京

电话：010-85950438

邮箱：researchbj@gjzq.com.cn

邮编：100005

地址：北京市东城区建国内大街 26 号

新闻大厦 8 层南侧

深圳

电话：0755-86695353

邮箱：researchsz@gjzq.com.cn

邮编：518000

地址：深圳市福田区金田路 2028 号皇岗商务中心

18 楼 1806



【小程序】
国金证券研究服务



【公众号】
国金证券研究