2026 年 11 月 18 日

电子行业研究

买入（维持评级）

行业深度研究

证券研究报告

电子组
分析师：樊志远（执业 S1130518070003）
fanzhiyuan@gjzq.com.cn

ASIC 加速放量，ASIC 产业链成长可期

投资逻辑

推理需求持续放量，资本开支效率诉求驱动 ASIC 加速渗透。OpenAI 推理算力支出由 2024 年的 18 亿美元增至 2025 年的 80 亿美元，Agent 与长上下文进一步推高计算和访存需求。截至 2026 年三季度的过去四个季度，北美四大云厂商资本开支合计 6154 亿美元，同比增长约 87%，自由现金流同比下降约 70%。ASIC 围绕目标负载优化成本与能效，成为提升算力投入回报的关键路径。Google TPU、AWS Trainium 已规模部署，Omdia 预计 2028 年数据中心 AI-ASIC 出货量将超过 GPU。

黄仁勋辞任 CEO 强化供应多元化诉求，ASIC 迎来新增催化。11 月 12 日英伟达公告黄仁勋辞任 CEO。我们认为，创始人退出经营管理层将提升客户分散单一供应商依赖的意愿，自研 ASIC 的战略优先级进一步抬升。云厂商已有产品与软件基础，新增推理负载有望加快向已验证的自研平台导入，供给可控与降本增效共同推动 ASIC 部署扩大。

ASIC 市场扩容叠加整柜升级，设计服务及系统制造价值量提升。Marvell 预计 2030 年定制芯片 TAM 约 2350 亿美元，2025—2030 年复合增速约 55%。HBM、高速 SerDes 及先进封装抬高设计壁垒，云厂商多代迭代带来持续研发需求。交付形态向整柜演进，带动高速交换、高阶 PCB 及系统制造升级；224Gbps 与更高层数推动 PCB 产品结构提升，液冷、供电及高速互联的整柜联调进一步凸显制造商价值。

OpenAI 定制芯片 2027—2028 年加速放量，天弘科技盈利弹性突出。博通预计 OpenAI 于 2027 年部署 1.3 GW，2028 年首代及后继 XPU 合计部署超过 5 GW，扩张至 2027 年计划量的 3.8 倍以上。天弘科技已参与 Jalapeño 整柜设计，预计 2027 年启动量产爬坡。公司 2027 年集团收入指引 400 亿美元、调整后经营利润率 9%，据此测算调整后经营利润同比增长约 104%，高于收入约 90%的增幅，规模效应与利润率改善共同放大盈利弹性。

Marvell 深化 Google 合作，定制业务增长与份额提升空间打开。双方已签署定制芯片协议，合作由互联拓展至推理加速器、存储及网络接口控制器，采购挂钩的认股权安排绑定长期合作。Marvell 将 FY2029 定制业务收入目标由超过 100 亿美元上调至超过 120 亿美元，FY2031 中点情景对应超过 300 亿美元。XPU 项目爬坡与配套芯片定点增加，有望推动公司持续获得新增设计份额。

投资建议

我们认为 ASIC 设计服务、高速网络、PCB 与整机柜制造有望受益。博通受益于 Google 与 OpenAI 定制芯片需求，FY26Q3 AI 半导体收入同比增长 221%；天弘科技受益于 ASIC 整柜量产及机柜级网络放量；沪电股份 2026 年上半年数据通讯 PCB 收入同比增长 73.46%，32 层及以上 PCB 收入同比增长 190.83%，高阶产品升级持续兑现。我们看好：天弘科技、博通，以及 ASIC 服务器与交换机 PCB 厂商如鹏鼎控股等。

风险提示

AI 进展及下游需求不及预期；竞争加剧；技术与量产进度不及预期；客户及供应链集中；数据中心配套交付延迟；盈利兑现不及预期；贸易政策变化。

内容目录

一、AI 应用扩张带动算力需求，云厂商持续加码 ASIC 4
1.1 推理需求持续扩张 ASIC 的系统性价比优势凸显 4
1.2 云厂商持续迭代自研芯片 ASIC 应用规模有望扩大 6
二、ASIC 市场空间扩大，带动产业链高速增长 8
2.1 ASIC 市场空间扩大，设计服务需求与技术壁垒同步提升 8
2.2 ASIC 向整机柜交付演进，带动高速网络、PCB 与系统制造需求 11
三、OpenAI 定制芯片 27-28 年预期高速增长，天弘科技有望充分受益 14
3.1 博通客户部署计划逐步清晰，OpenAI 从首批出货走向吉瓦级部署 14
3.2 Jalapeño 兼顾响应速度与能效，双机柜架构提升制造难度 15
3.3 天弘科技参与整柜设计与量产，经营杠杆有望放大利润增长 17
四、设计服务公司中 Marvell 边际变化较明显，长期展望积极 18
4.1 Google 扩大定制开发合作，Marvell 争取新增设计份额的基础增强 18
4.2 Marvell 深化 Google TPU 合作，有望获得更多 ASIC 定点 19
五、投资建议 20
六、风险提示 20

图表目录

图表 1： OpenAI 推理算力支出由 18 亿美元增至 25 年的 80 亿美元（单位：十亿美元） 4
图表 2： TPU 8i 将计算 存储和互联纳入协同设计 5
图表 3： TPU 8i 资源配置更偏向存储容量与带宽 6
图表 4： Google 第八代 TPU 的训练与推理产品 6
图表 5： 亚马逊已经推出 Trainium3 ASIC 芯片 7
图表 6： Meta 与 Microsoft 持续迭代自研加速器 7
图表 7： OpenAI 推出第一代 ASIC Jalapeño 芯片 8
图表 8： Omidia 预计 28 年 ASIC 出货量有望反超 GPU（单位：百万颗） 8
图表 9： Gartner 预计 ASIC 市场 30 年超 2000 亿美元（单位：十亿美元） 9
图表 10： Marvell 预计在 AI 半导体细分市场中，定制芯片将保持高增速 9
图表 11： 高端 XPU 的技术集成具备较高复杂度，核心技术包括 SerDes、Die to die 键合、先进封装等 10
图表 12： 定制化机会由计算芯片延伸至更多系统功能 10
图表 13： 博通 AI 半导体收入快速增长（单位：亿美元） 11
图表 14： Marvell 数据中心收入快速增长（单位：亿美元） 11
图表 15： 天弘科技云与连接解决方案收入快速增长（单位：亿美元） 11
图表 16： 世芯月度营收同比增长 26H2 转正并加速 11

图表 17： Jalapeño 整机柜集成 128 颗 ASIC 与交换芯片 12
图表 18： 柜内互联与跨柜集群带动 scale up（纵向）、scale out（横向）网络需求 12
图表 19： AI 发展带动风冷机柜转向液冷整机柜 13
图表 20： 26H1 沪电股份数据通讯 PCB 收入主要来自交换机与路由（单位：亿元人民币） 14
图表 21： 天弘科技将自动化投入延伸至托盘生产与整机柜装配 14
图表 22： 博通 XPU 客户的年度部署与出货计划，Anthropic 与 OpenAI 28 年预计增长显著 15
图表 23： DeepSeek R1 测试中吞吐量与用户生成速度的关系，OpenAI ASIC Jalapeño 具有较为明显又是... 16
图表 24： Jalapeño 双机柜的计算、交换与主机托盘布局 17
图表 25： 天弘科技投资者日对 2026 及 2027 年的集团指引 17
图表 26： 天弘科技指引 27 年业绩有望实现同比快速增长 18
图表 27： Marvell 与 Google 定制芯片合作要点，覆盖 TPU、存储、网络、内存控制器等..... 18
图表 28： Marvell 对市场空间与分业务收入增长的预测，Marvell 预计公司定制芯片收入增速将超过行业市场增速..... 19
图表 29： Marvell 集团收入从 FY2026 实绩到 FY2031 目标，FY31 有望实现 700~900 亿美金收入..... 19

一、AI 应用扩张带动算力需求，云厂商持续加码 ASIC

1.1 推理需求持续扩张 ASIC 的系统性价比优势凸显

算力投入已压缩云厂商的自由现金流，提升资本开支效率成为必然选择。根据各公司财报，截至 2026 年三季度的过去四个季度，Microsoft、Alphabet、Amazon 和 Meta 合计资本开支 6154 亿美元，同比增长约 87%，相当于同期经营现金流 6768 亿美元的 91%；合计自由现金流降至 615 亿美元，较上年同期的 2070 亿美元下降约 70%。2026 年三季度单季，四家合计资本开支 2019 亿美元，超过经营现金流 1755 亿美元，自由现金流转为－264 亿美元，其中 Alphabet、Amazon 和 Meta 单季均为负。Oracle 截至 2026 年 8 月的过去四个季度自由现金流为－287 亿美元。

算力需求仍在快速增长，云厂商需要让每一笔资本开支产出更多有效算力。我们认为，围绕自身负载定制、单位算力成本与能效更优的 ASIC，是云厂商提升资本开支回报的关键路径，ASIC 成为大势所趋。

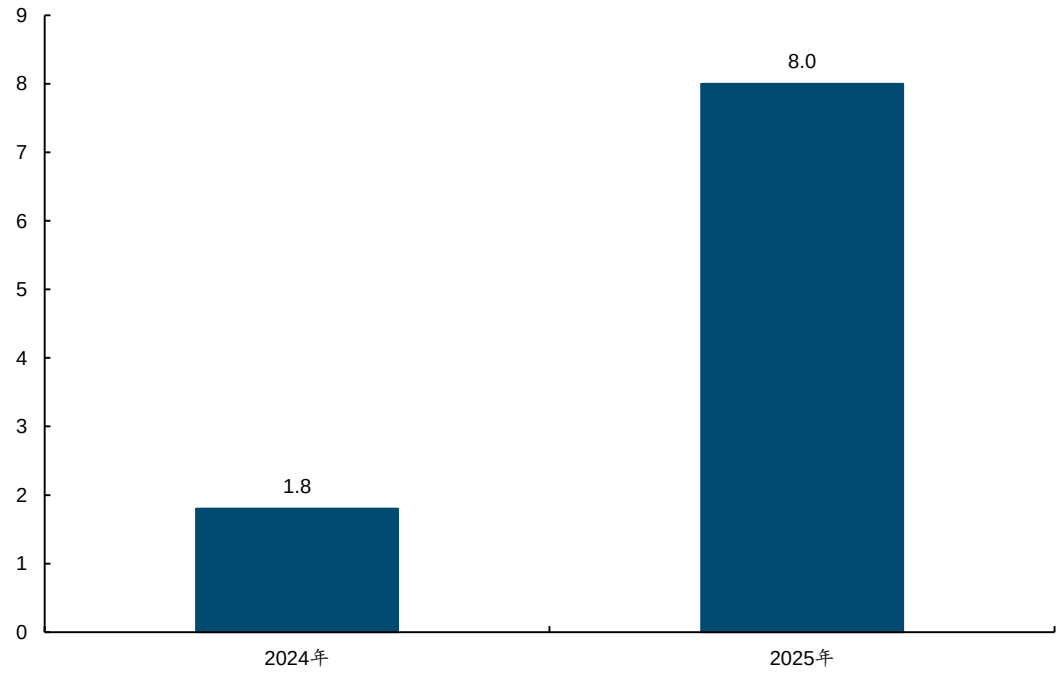
黄仁勋辞任 CEO，ASIC 迎来算力供应多元化的重要催化。英伟达于 2026 年 11 月 12 日公告，黄仁勋辞去总裁、CEO 及董事职务，CFO Colette Kress 出任临时 CEO；据公司 8-K 文件，辞任于 11 月 11 日生效。我们认为，创始人退出经营管理层，将推动大型算力客户重新评估对单一供应商的依赖，自研 ASIC 的战略优先级有望进一步提升。

供给自主可控与降本增效形成合力，有望加快云厂商导入 ASIC。云厂商通过自研芯片掌握产品节奏，并围绕自身模型优化计算、存储和互联配置，可在降低单位推理成本的同时增强算力规划主动权。Amazon 已披露，Bedrock 大部分推理运行在 Trainium 上，说明 ASIC 已具备承接大规模生产负载的基础。我们预计，本轮管理层变动将强化客户对多元算力架构的投入意愿，推动已验证的自研方案扩大部署，新增推理负载是重要切入点。

ASIC 渗透加速有望带动设计服务与整柜产业链共同增长。客户增加自研项目，将扩张定制设计与先进封装需求；规模部署进一步带动高速交换、高阶 PCB 及整机柜制造放量。我们看好已进入头部云厂商供应体系、具备高端设计或量产交付能力的公司，产业链有望同时受益于项目数量增加和单项目规模扩大。

单次推理价格下降，规模化部署仍推高算力支出。Stanford HAI《2025 AI Index》显示，在达到 GPT-3.5 相当 MMLU 水平的模型中，每百万 token 查询价格从 2022 年 11 月的 20 美元降至 2024 年 10 月的 0.07 美元。《2026 AI Index》显示，OpenAI 用于推理的算力支出从 2024 年的 18 亿美元增至 2025 年的 80 亿美元，增至上年的 4.44 倍。我们认为，当推理成为持续性支出，拥有稳定负载的客户更有动力通过专用芯片降低每次任务的成本。

图表 1：OpenAI 推理算力支出由 18 亿美元增至 25 年的 80 亿美元（单位：十亿美元）

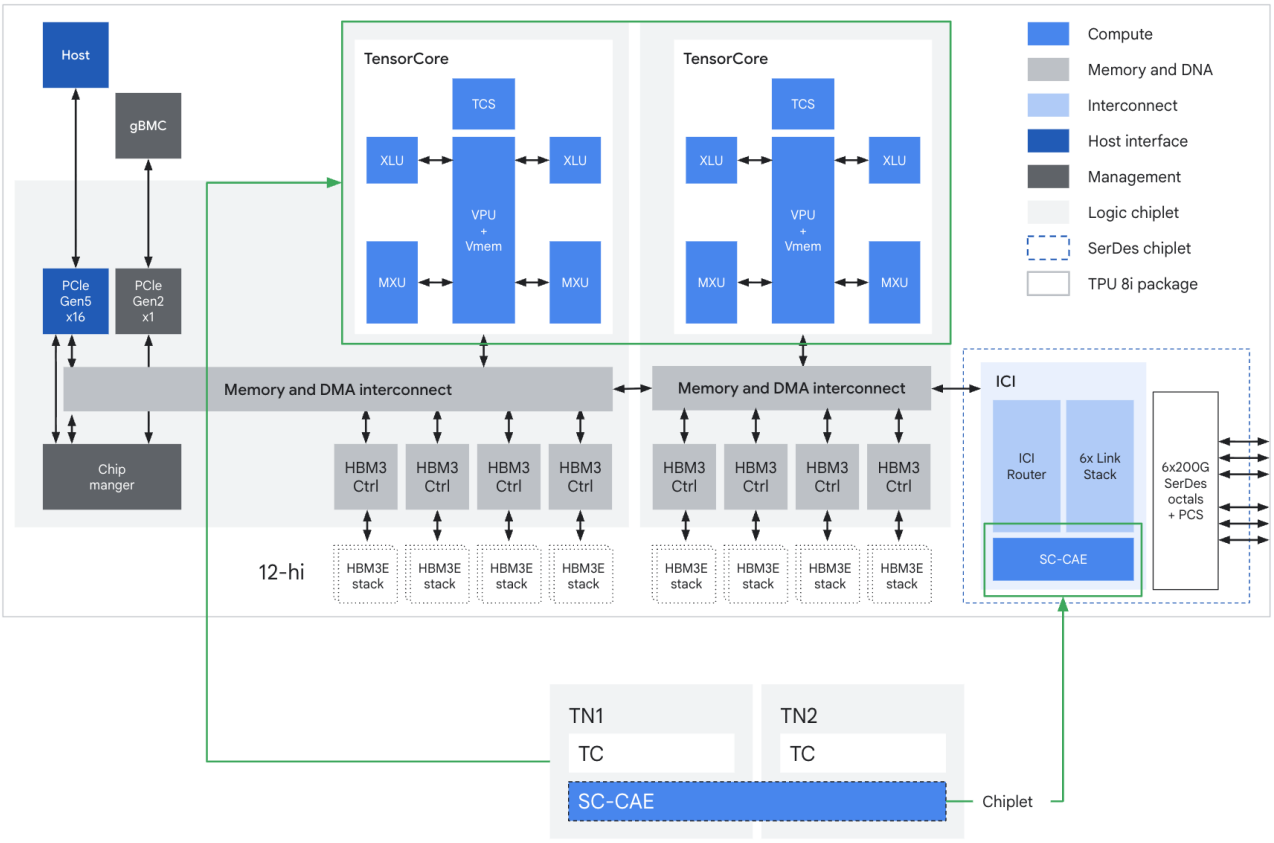


来源：Stanford HAI，国金证券研究所

应用普及与单任务算力增加，共同构成推理需求的增长来源。Agent 需要执行多轮推理、调用工具并保留上下文，部分任务的 token 消耗和访存需求随之上升。Google 在第八代 TPU 的架构说明中，将长上下文、复杂顺序逻辑和大规模 MoE 作为重要设计背景。

ASIC 的优势在于围绕目标负载优化整个系统的成本与能效。Stanford HAI《2026 AI Index》引用的测算显示，在输入和输出各约 1000 个 token 的条件下，GPT-5（high）每次查询平均耗电 21.85 Wh，Claude 4 Opus 为 5.32 Wh。

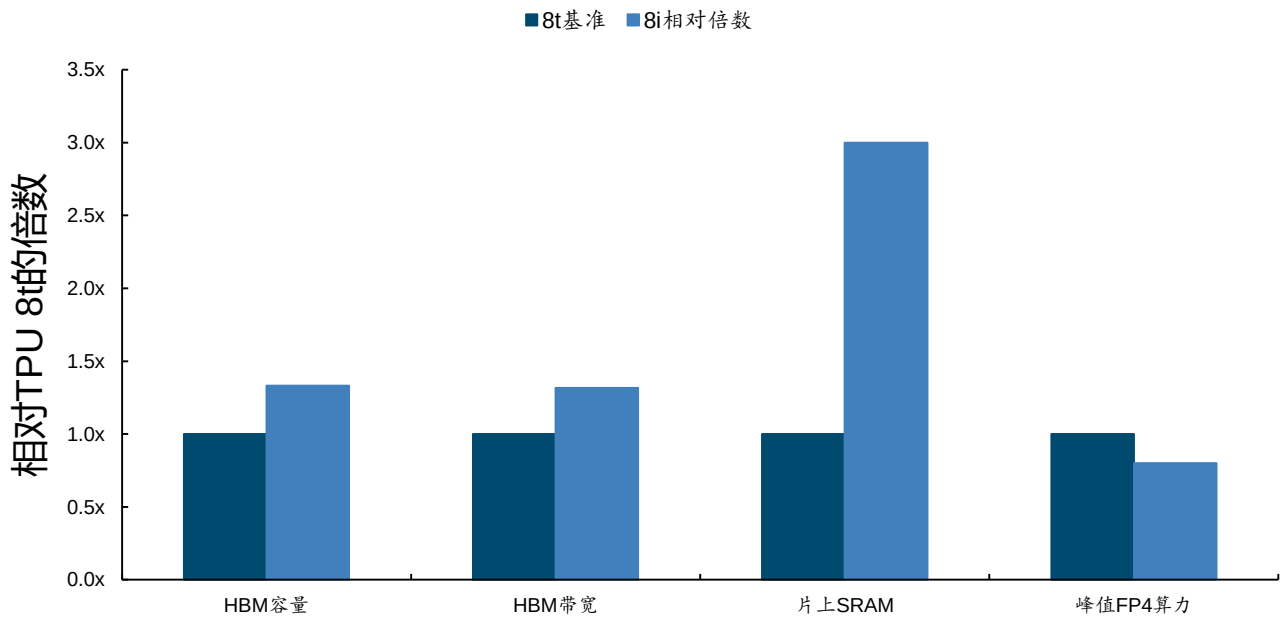
图表2：TPU 8i 将计算 存储和互联纳入协同设计



来源：Google Cloud，国金证券研究所。

推理优化正推动硬件资源向存储和通信重新分配。以 Google 公布的 TPU 8t 和 8i 为例，8i 面向服务与推理，其 HBM 容量、带宽和片上 SRAM 分别为 8t 的约 1.33 倍、1.32 倍和 3 倍，而峰值 FP4 算力约为 0.80 倍。我们认为，这一配置差异说明，推理专用设计的价值主要在于减少数据等待、提升实际利用率，而非直接增加运算单元。

图表3：TPU 8i 资源配置更偏向存储容量与带宽



来源：Google Cloud，国金证券研究所，各指标以 8t=1 标准化。

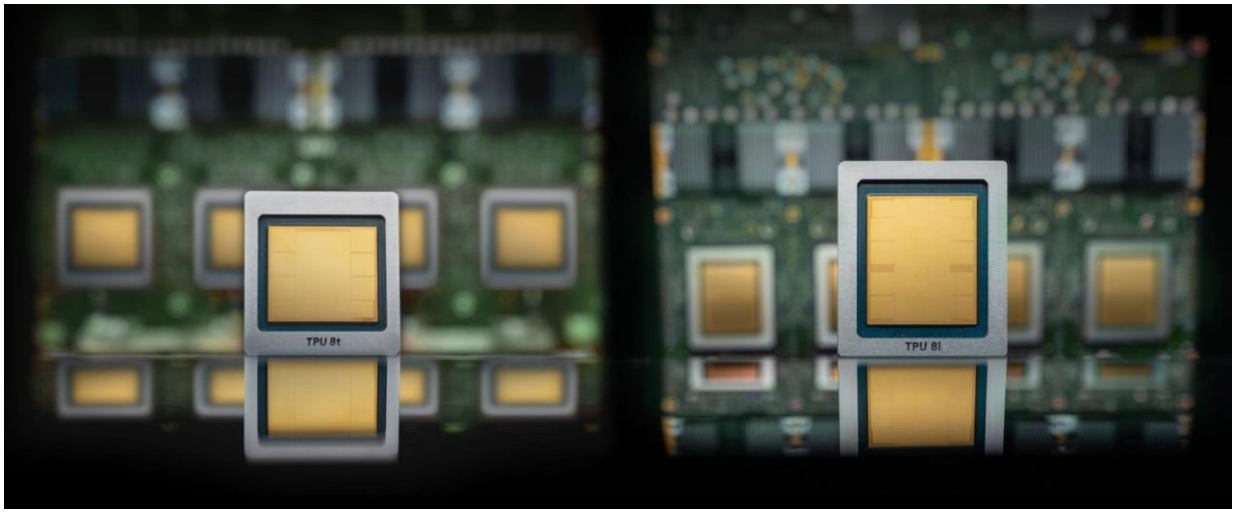
规模效应和软件成熟度决定 ASIC 优势能否兑现。Meta 明确提出复用芯粒、机箱、机架和网络，并围绕 PyTorch、vLLM 及 Triton 建设软件栈。我们认为，硬件平台复用有助于分摊研发投入，软件兼容则有助于降低模型迁移成本。硬件效率、部署规模与软件生态共同决定 ASIC 渗透率的提升速度。

1.2 云厂商持续迭代自研芯片 ASIC 应用规模有望扩大

云厂商的定制芯片已形成多代产品路线，行业参与范围继续扩大。Google、AWS、Meta 和 Microsoft 均已公布持续迭代的自研 AI 加速器，模型公司 OpenAI 也已发布首款自研推理芯片。各家布局的共同出发点，是把内部模型与应用需求转化为硬件优化，并提高大规模算力供给的可控性。

Google 进一步区分训练和推理的硬件需求。2026 年 4 月公布的第八代 TPU 包含 8t 与 8i 两类架构，分别侧重大规模预训练以及采样、服务和推理。8t 采用 3D torus 拓扑，8i 采用 Boardfly 拓扑，并强化存储与集合通信能力。我们认为，产品分化反映出训练与在线推理的优化目标逐渐拉开，有利于围绕不同负载持续开发定制芯片。

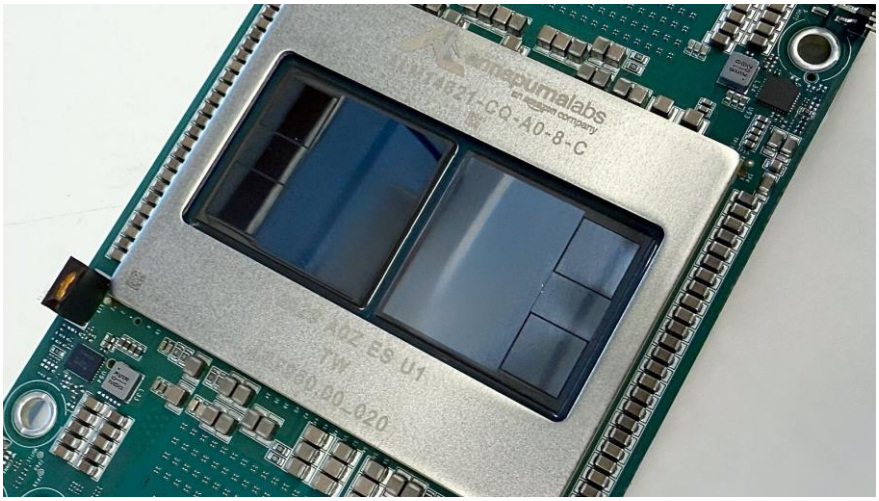
图表4：Google 第八代 TPU 的训练与推理产品



来源：Google Cloud，国金证券研究所。

AWS 的 Trainium 已经从产品迭代进入实际业务承载。Amazon 在 2026 年 4 月发布的股东信中披露，Trainium3 于 2026 年初开始出货，并称其较 Trainium2 价格性能提升 30%至 40%，Amazon Bedrock 大部分推理已运行在 Trainium 上。

图表5：亚马逊已经推出 Trainium3 ASIC 芯片



来源：亚马逊，国金证券研究所

Meta 与 Microsoft 继续扩大 ASIC 的工作负载覆盖。Meta 在 2026 年 3 月披露 MTIA 300 已用于推荐与排序训练，MTIA 400 完成实验室测试，450 与 500 计划于 2027 年扩大部署。Microsoft 于 2026 年 10 月披露 Maia 200 已服务生产工作负载，Maia 300 计划 2027 年开始生产部署。我们认为，ASIC 正在由单一推荐任务扩展至更广泛的生成式 AI 服务。

图表6：Meta 与 Microsoft 持续迭代自研加速器

Meta MTIA 官方配图



Microsoft Maia 300



来源：Meta、Microsoft，国金证券研究所，左为 Meta MTIA，右为 Microsoft Maia 300。

模型厂商进入自研芯片领域，为 ASIC 增加新的需求主体。OpenAI 与博通于 2026 年 6 月公布面向大语言模型推理的 Jalapeño，并于 8 月披露首批厂商测试结果，已从样片演示进入模型、软件与硬件联合验证阶段，初始部署仍计划在 2026 年底。模型厂商向底层基础设施延伸，有助于围绕真实业务负载优化整套系统的效率。

图表7：OpenAI 推出第一代 ASIC Jalapeño 芯片

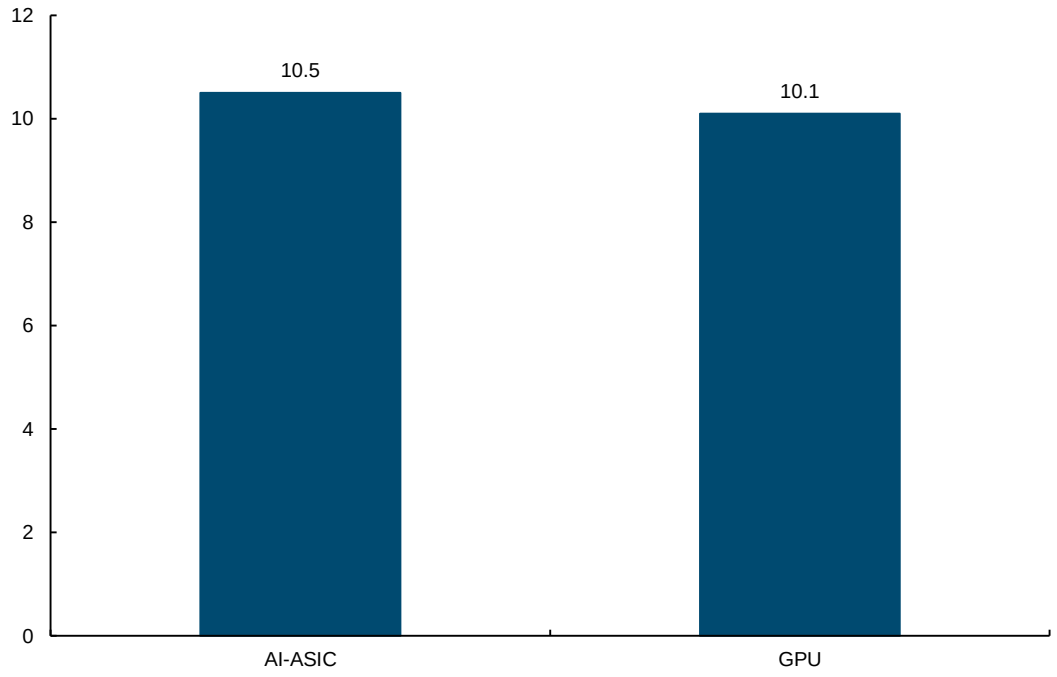


来源：OpenAI，Tom's hardware，国金证券研究所。

基础设施投资与 AI 服务器需求扩张，为 ASIC 提供更大的部署空间。TrendForce 于 2026 年 10 月预计，全球九大 CSP 当年资本开支为 9570 亿美元，2027 年超过 1.5 万亿美元；AI 服务器出货量在 2026 年增长 32.2%，2027 年增速接近 36%。

2028 年 ASIC 出货量有望超过 GPU，收入端 GPU 仍保持领先。Omdia 于 2026 年 8 月预计，2028 年数据中心 AI-ASIC 出货量将达 1050 万颗，略高于 GPU 的 1010 万颗；同时预计 GPU 收入仍将领先至 2032 年。

图表8：Omidia 预计 28 年 ASIC 出货量有望反超 GPU（单位：百万颗）



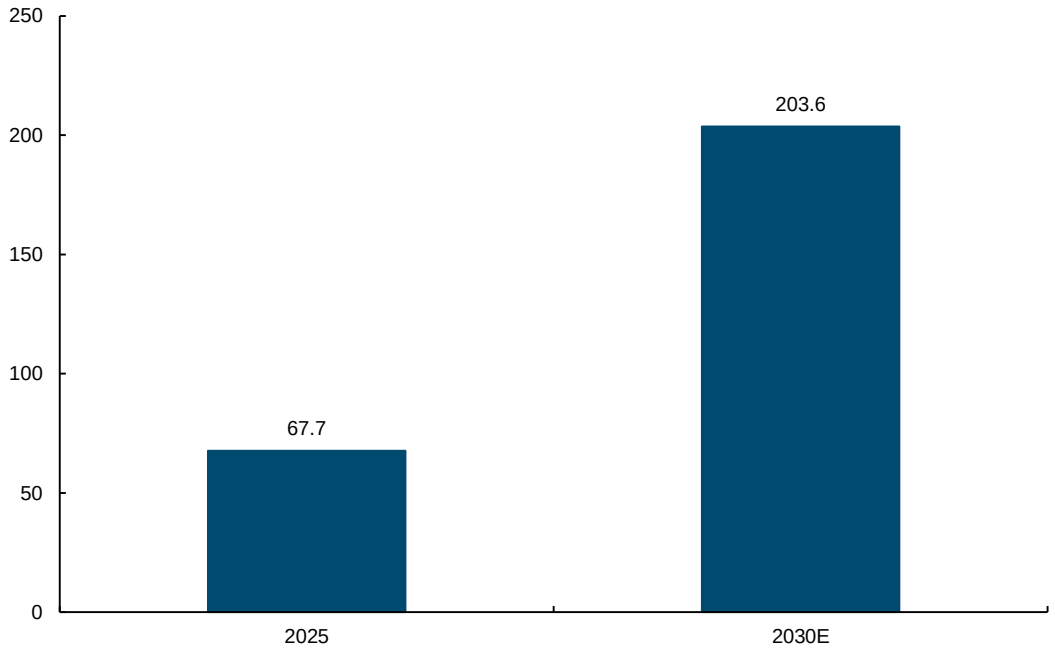
来源：Omidia，国金证券研究所

二、ASIC 市场空间扩大，带动产业链高速增长

2.1 ASIC 市场空间扩大，设计服务需求与技术壁垒同步提升

ASIC 市场扩容为设计服务带来持续项目需求。Gartner 于 2026 年 8 月发布的公开摘要预计，定制半导体 ASIC 市场将由 2025 年的 677 亿美元增至 2030 年的 2036 亿美元，五年复合增速约 24.6%。

图表9：Gartner 预计 ASIC 市场 30 年超 2000 亿美元（单位：十亿美元）



来源：Gartner，国金证券研究所

产业链公司对定制芯片市场的预期，进一步印证 ASIC 的长期扩容空间。Marvell 在 2026 年 10 月 Investor Day 预计，2030 年 Custom 市场空间约 2350 亿美元，2025—2030 年复合增速约 55%，增长来源包括 XPU 及 XPU attach。天弘科技预计，2026—2030 年 Custom AI ASIC 市场扩大至 4 倍以上。

图表10：Marvell 预计在 AI 半导体细分市场中，定制芯片将保持高增速

业务	市场规模复合增速 2025—2030 年	行业增长驱动因素	2030 年市场规模
互联	约 65%	横向扩展增长；新增纵向扩展与跨集群扩展需求	约 650 亿美元
交换与存储	约 40%	横向扩展增长；新增纵向扩展需求	约 850 亿美元
定制芯片	约 55%	XPU 数量增加；新增 XPU 配套芯片需求	约 2,350 亿美元
通信及其他	约 3%	温和增长	约 150 亿美元
合计	约 45%	AI 基础设施扩张带动各层需求	约 4,000 亿美元

来源：Marvell，国金证券研究所

客户自研架构与外部设计服务具有互补关系。自研通常意味着客户掌握负载需求、系统架构与产品定义。ASIC 服务模式涵盖从规格、网表或 GDS 交接，到物理实现、封装、流片、测试和量产管理的不同参与深度。我们认为，客户可按自身团队能力决定外包边界，设计服务行业的机会既来自新增客户，也来自同一客户产品线和代际项目的扩张。

高端 ASIC 的壁垒正从单一设计环节延伸至系统协同能力。HBM 接口、高速 SerDes、芯粒互联等 IP，需要与先进制程及 2.5D/3D 封装结合，同时处理信号完整性、供电和散热问题。Marvell 在 Investor Day 展示了 XPU 向更多计算芯粒、更多 HBM 及光 I/O 演进的路线，并将连接 IP、定制存储和先进封装列为关键能力。我们认为，已验证 IP、封装协同及量产经验共同构成执行壁垒，客户购买的不仅是设计服务公司所提供的设计人力，也包括降低复杂项目失败风险的能力。

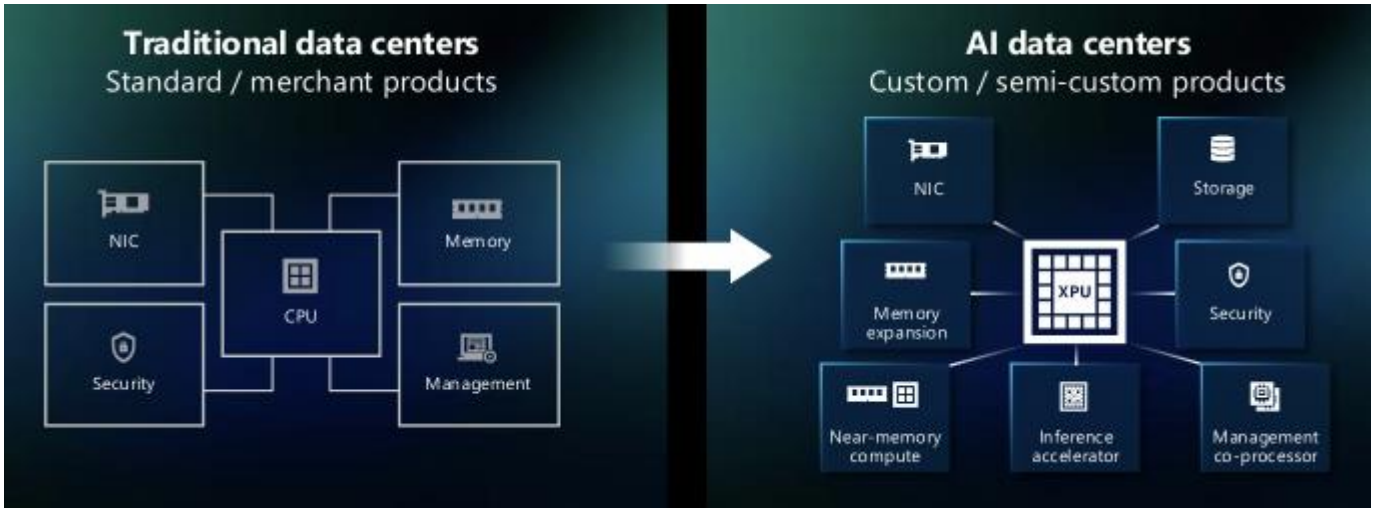
图表11：高端 XPU 的技术集成具备较高复杂度，核心技术包括 SerDes、Die to die 键合、先进封装等



来源：Marvell，国金证券研究所。

设计定点转化为收入，需要经过验证与量产的完整周期。设计服务通常经历方案开发、流片、样片验证、可靠性认证及量产爬坡等阶段。先进工艺、更多存储和复杂封装可能提高单颗芯片价值量，也会提高研发及制造成本。未来随着 ASIC 所采用的制程持续提升，涉前道设计与封装设计复杂度持续增长，以及 ASIC 所需要的系统化工程能力要求增长，配套芯片复杂度提升，行业市场空间有望持续扩容。

图表12：定制化机会由计算芯片延伸至更多系统功能



来源：Marvell，国金证券研究所。

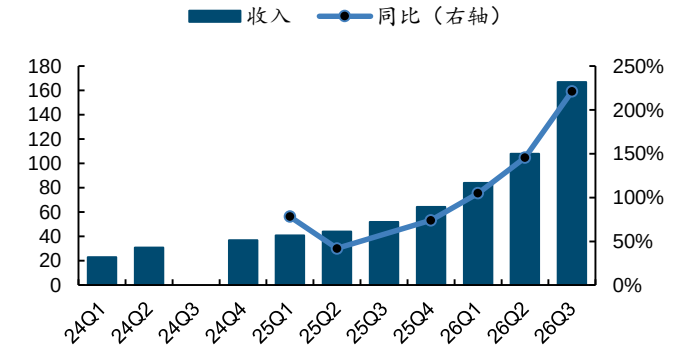
平台复用有望降低开发门槛，工程复杂度仍支撑专业分工。云厂商可以通过复用芯粒、接口与机架平台加快迭代，设计服务平台也可以复用 IP 和验证流程，缩短不同项目的开发路径。但模型演化、先进节点迁移和多芯粒封装同时带来新的工程工作。随着 ASIC 参与者和产品代数增加，行业竞争将围绕交付效率、系统设计与量产能力展开。

相关业务季度收入持续增长，验证 AI 基础设施需求向产业链传导。Broadcom AI 业务、Marvell 数据中心业务及天弘科技 CCS 业务季度收入保持高增长。

相关业务分部的增长，更直接体现 AI 投入向产品需求的传导。Broadcom 披露 FY26Q3 AI 半导体收入 167 亿美元、同比增长 221%；Marvell FY27Q2 数据中心收入 21.715 亿美元、同比增长约 46%；天弘科技 2026Q3 CCS 收入 48.72 亿美元、同比增长约 102%，体现 ASIC 高景气度。

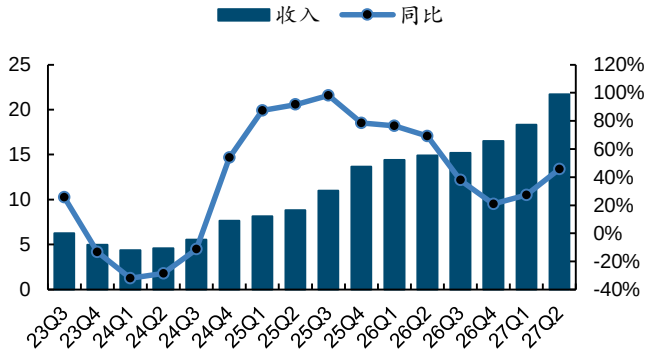
ASIC 景气度逐渐向北美以外厂商扩散，世芯等厂商增速明显。世芯 2026 年 1—5 月单月营收同比仍为负，6 月转为增长 15.4%，7~9 月进一步增至 181.8%、273.6%和 496.0%。公司对 7~9 月增长的公告说明均指向量产产品增加。

图表13：博通 AI 半导体收入快速增长（单位：亿美元）



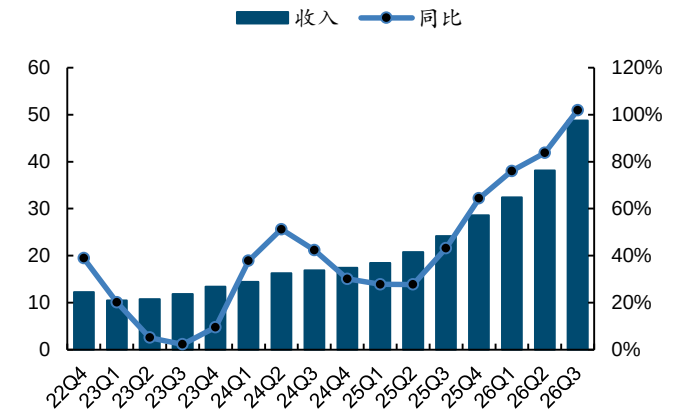
来源：博通公告，国金证券研究所，3Q24 AI 半导体收入未单独披露

图表14：Marvell 数据中心收入快速增长（单位：亿美元）



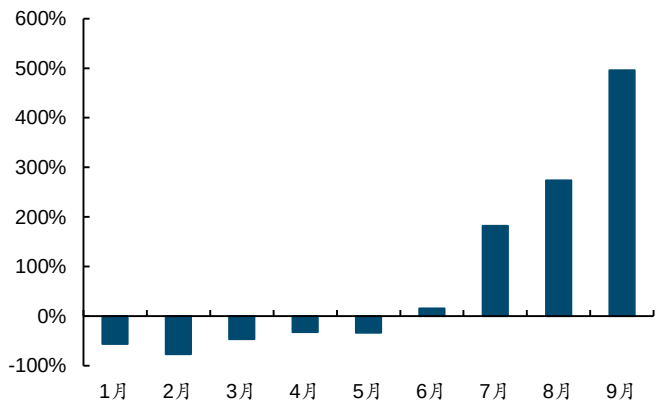
来源：Marvell 公告，国金证券研究所

图表15：天弘科技云与连接解决方案收入快速增长（单位：亿美元）



来源：天弘科技公告，国金证券研究所

图表16：世芯月度营收同比增长 26H2 转正并加速



来源：世芯公告，国金证券研究所

我们认为，考虑到 ASIC 从设计到量产周期较长、客户项目具有多代延续性，多家云厂商投入与应用扩张有望共同推动高端 ASIC 设计服务的长期需求，后续重点跟踪项目验证、客户实际采用与量产交付进度。

2.2 ASIC 向整机柜交付演进，带动高速网络、PCB 与系统制造需求

高密度 ASIC 平台开始以整机柜为设计和交付单位。天弘科技在 2026 年投资者日展示了面向 OpenAI Jalapeño 的方案：128 颗定制加速芯片与 Broadcom Tomahawk 6 交换芯片组成机柜平台，并配套 SONiC 网络操作系统。公司预计该项目于 2027 年开始量产爬坡。

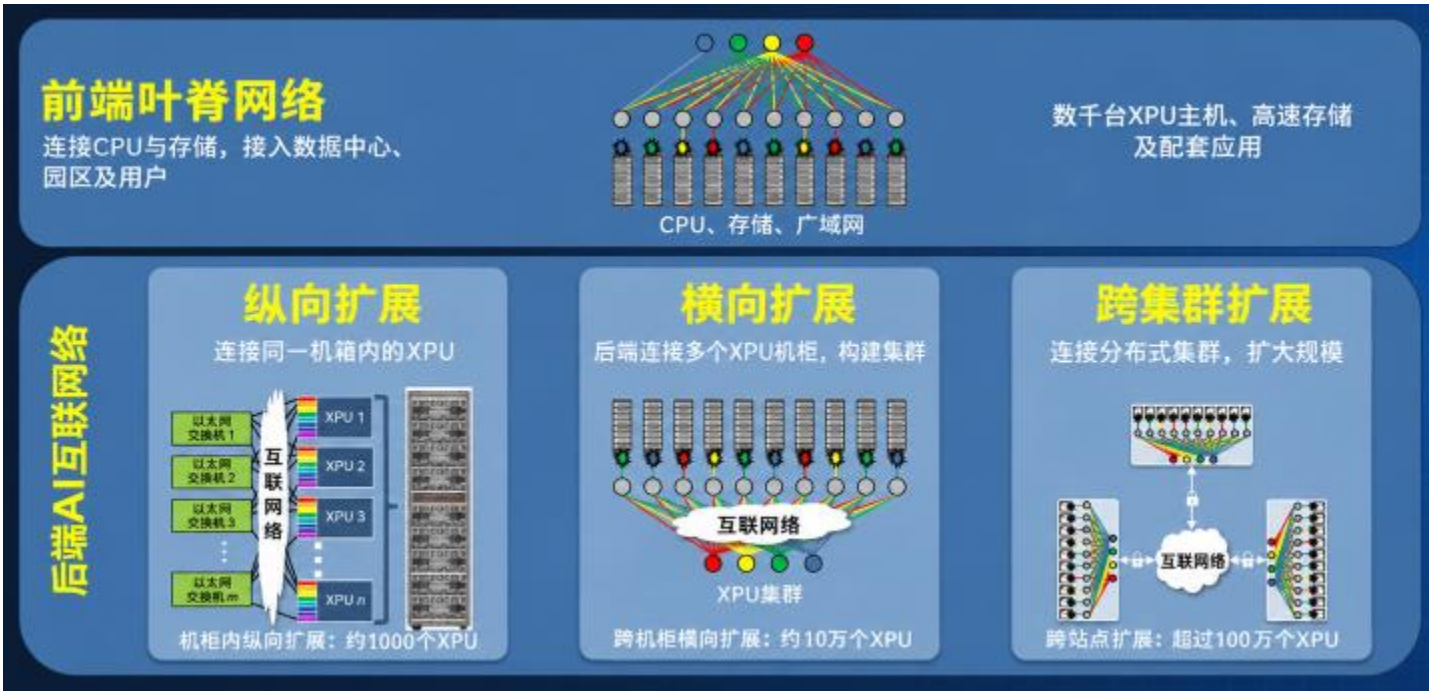
图表17：Jalapeño 整机柜集成 128 颗 ASIC 与交换芯片



来源：天弘科技，国金证券研究所

柜内协同计算与柜间扩容，带来 Scale-up 和 Scale-out 网络需求，前者连接紧密协作的加速器，芯片间要频繁交换中间结果，互联时延会影响计算单元的等待时间。后者把多柜加速器连成集群，跨柜通信更关注可用带宽与拥塞。前端网络另行连接 CPU、存储和用户。

图表18：柜内互联与跨柜集群带动 scale up（纵向）、scale out（横向）网络需求



来源：Arista，国金证券研究所。

更高功率密度改变了网络设备的安装和冷却方式。Arista 列举的分立风冷系统最高功率密度约为每柜 30 至 40 千瓦，液冷整柜平台可达 100 至 150 千瓦或更高。其对照图显示，前者以外部线缆连接独立设备，后者通过内部背板和盲插接口将计算与网络紧密连接。天弘科技预计，其含光学及其他产品的网络业务组合收入将由 2025 年的 71 亿美元增至 2026 年的 118 亿美元，同比增长约 66%。

图表19：AI 发展带动风冷机柜转向液冷整机柜

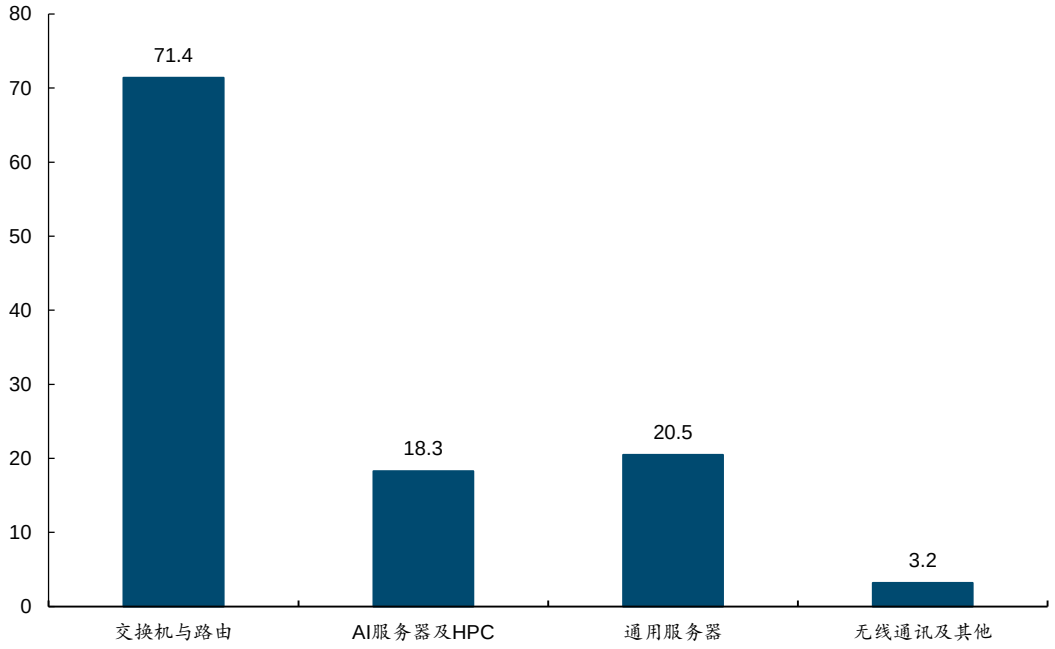
比较维度	传统风冷分立系统	液冷整机柜集成平台
最大功率密度	每柜最高 30－40kW	每柜 100－150kW 以上
冷却介质	高速气流	冷却液 / 水-乙二醇
组件布局	机箱相互独立，预留通风间距	计算与网络背板紧密耦合
风扇附加功耗	较高（系统内部风扇负荷较大）	极低（由高效液冷泵驱动）
互联布线	外部光缆 / 铜缆通路，阻碍气流	内部背板与盲插连接
光模块热环境	标准光模块承受较高热应力	冷却、稳定的运行环境，针对 LPO 优化
典型设施电能使用效率（PUE）	1.3－1.5	1.05－1.15

来源：Arista，国金证券研究所

信号速率提升，使高阶 PCB 的难点落到材料、布线与加工精度上。高速信号经过板材和过孔会产生损耗与反射。沪电股份在 2025 年年报中披露，公司针对 224Gbps 信号推进低损耗材料、背钻残桩和阻抗控制研究。当时 ASIC 平台相关 224Gbps 产品仍在认证，1.6T 交换机相关 PCB 开始小批交付。到 2026 年半年报，适配 224Gbps 的核心产品已小批供货，1.6T 交换机用板已批量出货，448Gbps 和 3.2T 相关产品进入预研。

沪电的分项收入显示，交换与路由用板已是数据通讯业务的主要收入来源。2026 年上半年，沪电数据通讯 PCB 收入约 113.30 亿元，同比增长 73.46%。其中，高速网络交换机及配套路由为 71.39 亿元，AI 服务器及 HPC 为 [illegible]28合计 89.67 亿元，占数据通讯收入约 79.1%。同期公司 32 层及以上 PCB 产品收入同比增长约 190.83%。

图表20：26H1 沪电股份数据通讯 PCB 收入主要来自交换机与路由（单位：亿元人民币）



来源：沪电股份公告，国金证券研究所

整柜制造的价值集中在装配、联调和整机验收。天弘科技在投资者日披露，自动化产线的投入同时覆盖网络托盘、计算托盘和整柜装配，并把机柜级液冷、电源分配以及高速信号完整性列为研发重点。托盘装入机柜后，需完成背板、供电与液冷回路的整体联调，这正是整柜制造商的价值所在。

图表21：天弘科技将自动化投入延伸至托盘生产与整机柜装配



来源：天弘科技，国金证券研究所

三、OpenAI 定制芯片 27-28 年预期高速增长，天弘科技有望充分受益

3.1 博通客户部署计划逐步清晰，OpenAI 从首批出货走向吉瓦级部署

六家 XPU 客户中，Anthropic 和 OpenAI 的年度部署量最清晰。博通在 FY26Q2 电话会预计 2027 年 XPU 出货对应约 10 GW 算力容量，进度偏向下半年。到 Q3，公司披露 Anthropic 在 2026 年部署约 1 GW Ironwood，2027 年预计再部署 5 GW TPU v8i。Meta 首笔 1 GW 订单自 2027 年下半年开始交付。

图表22：博通 XPU 客户的年度部署与出货计划，Anthropic 与 OpenAI 28 年预计增长显著

客户	2026 年进展	2027 年计划	2028 年展望
Google 自用	Ironwood 出货，v8i 开始量产；GW 未披露	后续 TPU 需求增长；自用 GW 未披露	后续 TPU 需求预计增长；自用 GW 未披露
Anthropic	约 1 GW Ironwood	新增约 5 GW TPU v8i	当年新增约 10 GW，专属 Anthropic
OpenAI	首批出货，年底初步部署；GW 未披露	1.3 GW 部署承诺	预计部署超过 5 GW，含 Jalapeño 及后继 XPU
Meta	MTIA 持续迭代；GW 未披露	首笔 1 GW 订单于下半年起交	截至 2028 年底三代 MTIA 累计约 3 GW；当年量未披露
其余两家	已有出货；各自 GW 未披露	Q1 预计出货量超翻倍；各自 GW 未披露	客户未具名，各自 GW 未披露
六家合计	全年 GW 未披露	Q2 规划约 10 GW 出货	未披露

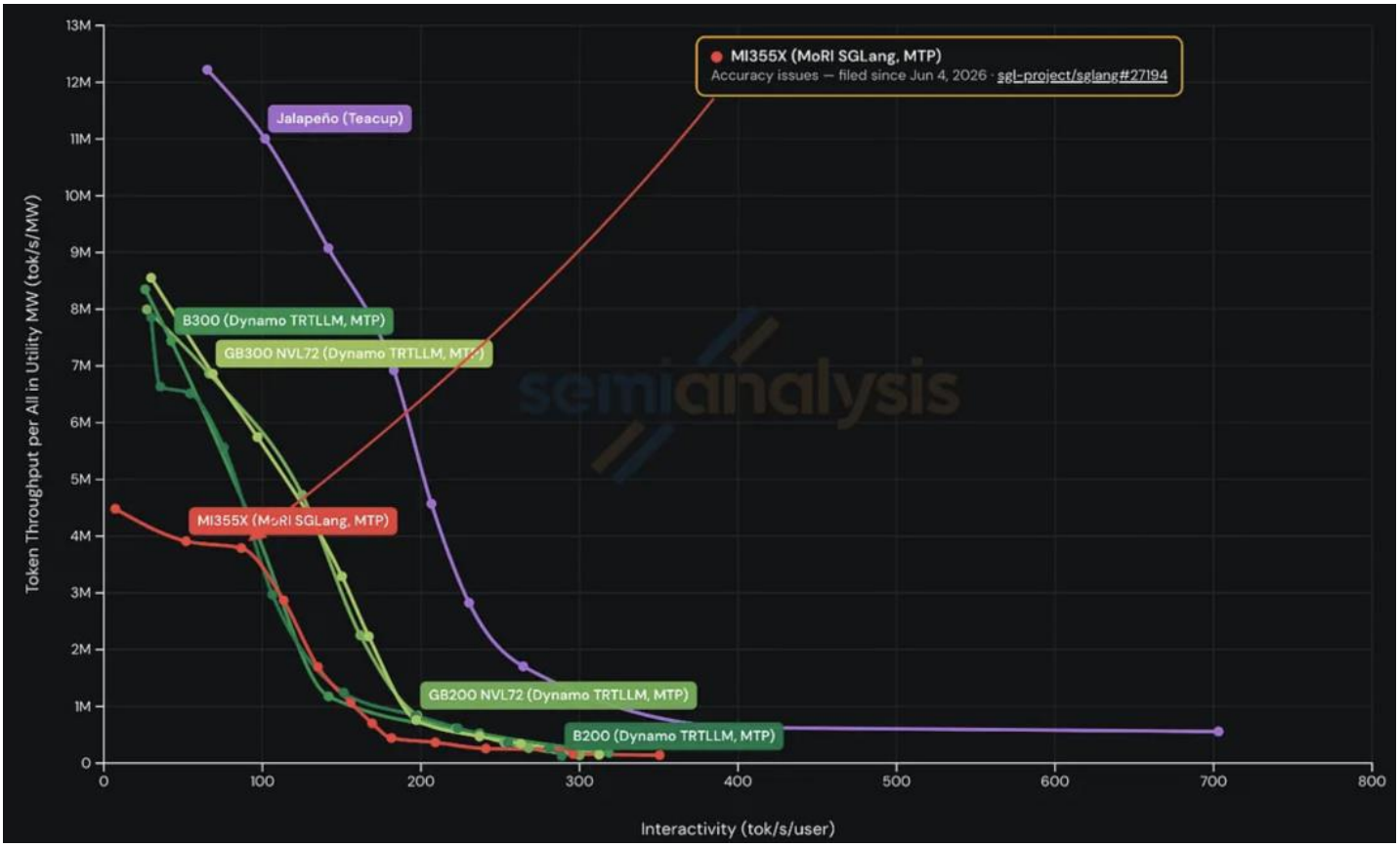
来源：博通，国金证券研究所

OpenAI 是新增量最明确的客户之一。FY26Q3 博通已交付首代 Jalapeño，并预计 Q4 继续出货。OpenAI 计划于 2026 年底开始初步部署，2027 年部署 1.3 GW 已有合同承诺。放量速度体现在后续规划：管理层预计 2028 年首代及后继 XPU 合计部署超过 5 GW，相当于 2027 年计划量的 3.8 倍以上。

3.2 Jalapeño 兼顾响应速度与能效，双机柜架构提升制造难度

对交互式推理而言，芯片价值体现在用户等多久，以及同样的电力能服务多少请求。8 月 25 日，Sam Altman 在 X 上写道："we made a chip and it is fast"。同日 OpenAI 披露，在 GPT-OSS 120B、DeepSeek R1 和 Kimi K2.5 三种公开模型上，按公布的芯片额定功耗归一化，Jalapeño 的峰值单位功耗吞吐量为对比系统的 1.5 至 1.9 倍，高交互负载下的性能为 2.1 至 4.1 倍。对需要连续调用模型的编码助手，单步等待时间会逐轮累积，降低生成时延因而能改善整项任务的体验。

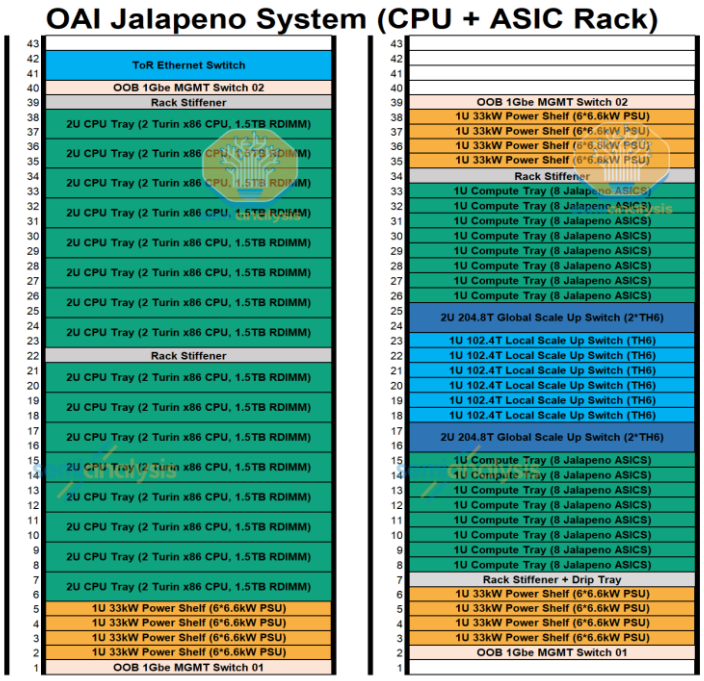
图表23：DeepSeek R1 测试中吞吐量与用户生成速度的关系，OpenAI ASIC Jalapeño 具有较为明显又是



来源：SemiAnalysis，国金证券研究所，横轴为 token/s/用户，竖轴为 token/s/MW

一套 Jalapeño 系统需要同时交付主机柜与加速器柜。SemiAnalysis 拆解，主机柜装有 16 个 Katsu 托盘，加速器柜装有 16 个 Windaloo 计算托盘，每托盘 8 颗 ASIC，单柜合计 128 颗。另有 8 个 Chana 交换托盘，其中 6 个承担本柜互联，2 个连接更大的计算域。每对主机和计算托盘通过 8 根外部 PCIe 线缆互连，按 16 对计算，一套双柜仅这一路径便有 128 根线缆。柜间、柜内布线与背板需要在相同的机械公差、信号损耗和维护空间约束下完成。

图表24：Jalapeño 双机柜的计算、交换与主机托盘布局



来源：SemiAnalysis，国金证券研究所

托盘越多，整柜验收越依赖制造商的系统能力。SemiAnalysis 估算，加速器柜约 130 kW，主机柜生产运行约 31 kW，双柜合计约 160 kW。我们认为，在这个功率密度下，装配完成后的互联误码、散热和供电都可能限制整柜吞吐。能参与早期设计、在产线上定位故障并完成双柜联调的制造商，有机会承接比单纯组装板卡更多的工程和测试工作。

3.3 天弘科技参与整柜设计与量产，经营杠杆有望放大利润增长

我们看好天弘科技在 OpenAI 整柜放量中的制造机会。天弘科技在投资者日确认，已为 Jalapeño 提供 HPS 整柜方案及定制 SONiC 软件，生产爬坡计划从 2027 年开始。公司也在投资计算及网络托盘产线和整柜自动化装配。我们认为，前期设计验证把制造商带入客户的产品迭代过程；进入量产后，工装、测试程序与工程团队有望复用。我们以集团指引评估利润弹性，后续关注客户验收与交付进度。

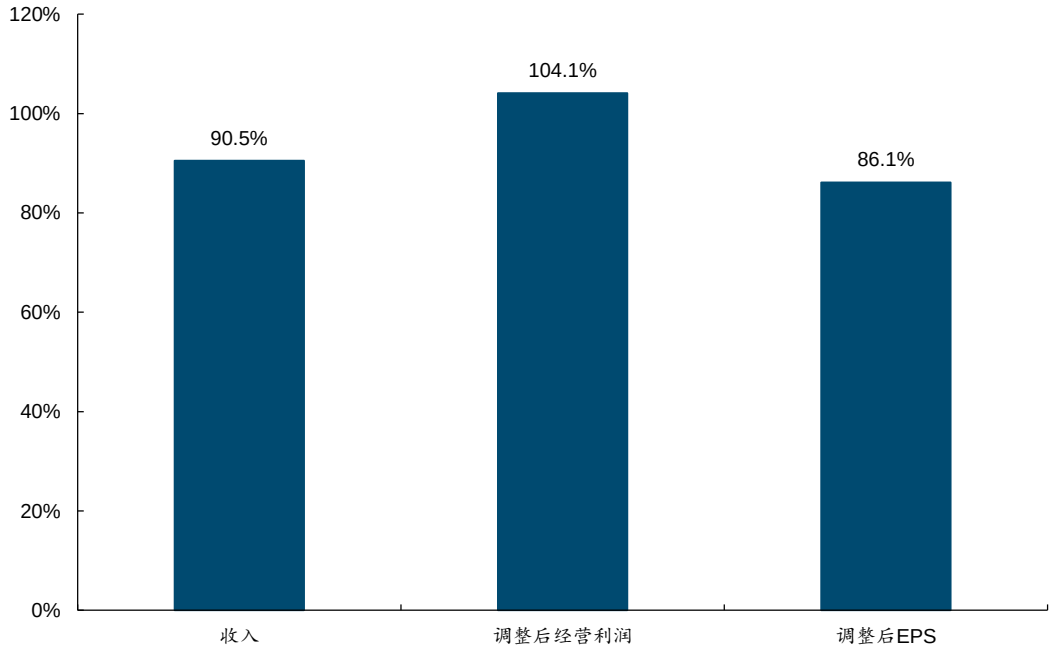
图表25：天弘科技投资者日对 2026 及 2027 年的集团指引

集团收入（亿美元）	210	400	+90%
调整后经营利润率	8.4%	9.0%	+60 个基点
调整后每股收益（美元）	11.55	21.50	+86%
资本开支（亿美元）	约 10	15–20	+75%（按中点）

来源：天弘科技，国金证券研究所

投资者日指引显示，利润弹性首先体现在调整后经营利润。天弘科技预计集团收入从 2026 年的 210 亿美元升至 2027 年的 400 亿美元，调整后经营利润率从 8.4%升至 9.0%。按两项指引相乘，调整后经营利润由 17.64 亿美元升至 36.00 亿美元，增长约 104%，快于收入约 90%的增幅。按先放量、再提利润率拆分，18.36 亿美元经营利润增量中，15.96 亿美元来自收入扩大，2.40 亿美元来自利润率提高 60 个基点。公司指出，调整后毛利润金额的增长快于固定成本，这为规模效应提供了经营依据。

图表26：天弘科技指引27年业绩有望实现同比快速增长



来源：天弘科技，国金证券研究所

利润率提升将进一步放大盈利弹性。以2027年400亿美元收入为基准，调整后经营利润率每提升0.5个百分点，经营利润约增加2亿美元。投资者日预计2027年资本开支为15亿至20亿美元，调整后EPS由11.55美元增至21.50美元，增幅约86%；摊薄加权平均股数假设由约1.20亿股升至1.27亿股。

四、设计服务公司中Marvell边际变化较明显，长期展望积极

4.1 Google扩大定制开发合作，Marvell争取新增设计份额的基础增强

我们看好Marvell在Google新一轮定制芯片项目中的竞争机会。Marvell在8月19日提交的8-K中披露，双方已于7月29日签署商业协议，合作范围包括面向TPU生态的推理加速器、存储控制器、网络接口控制器、内存接口控制器及近内存计算。

认股权与采购挂钩，绑定双方长期合作。Google取得的认股权中，一部分按时间归属，其余部分与Google及关联方的自主采购挂钩。自Marvell FY2027度至FY2033末，每实现5亿美元符合条件的定制产品收入，便可触发一档归属，共240档。全部触发对应1200亿美元的累计收入门槛。

图表27：Marvell与Google定制芯片合作要点，覆盖TPU、存储、网络、内存控制器等

观察项目	已披露内容
合作范围	TPU生态推理加速器及存储、网络、内存等控制器
采购激励	每5亿美元符合条件的定制产品收入触发一档，共240档
公司定制收入目标	FY2029由超过100亿上修至超过120亿美元
远期情景	FY2031中点情景超过300亿美元

来源：Marvell，国金证券研究所。

Marvell自身的定制业务展望也在上修。在10月6日投资者日，公司把FY2029定制业务收入目标从超过100亿美元提高到超过120亿美元，预计较FY2028增长至三倍以上；FY2031中点情景对应超过300亿美元。管理层将增长来源归于已有XPU项目爬坡，以及配套芯片（XPU attach）设计定点增加。我们认为，目标上修反映设计定点正加速转化为收入。

Marvell 预计，定制芯片将成为其可服务市场中规模最大的板块。按公司 2026 年 10 月投资者日测算，到 2030 自然年，其可服务市场（TAM）约 4000 亿美元，其中定制芯片 2350 亿美元，占约 59%；交换与存储 850 亿美元，互联 650 亿美元，通信及其他 150 亿美元。管理层预计 2025 至 2030 自然年定制芯片 TAM 年均增长约 55%，同期整体 TAM 年均增长约 45%。

图表28：Marvell 对市场空间与分业务收入增长的预测，Marvell 预计公司定制芯片收入增速将超过行业市场增速

业务	CY2030 TAM / 市场 CAGR[1]	FY2031 收入中点 / 公司 CAGR[2]
定制芯片	约 2350 亿美元 / 约 55%	约 300 亿美元 / 75%–85%
互联	约 650 亿美元 / 约 65%	约 375 亿美元 / 60%–70%
交换与存储	约 850 亿美元 / 约 40%	约 100 亿美元 / 40%–50%
通信及其他	约 150 亿美元 / 约 3%	约 25 亿美元 / 超过 3%
合计	约 4000 亿美元 / 约 45%	中点 800 亿美元 / 55%–60%
统计期间	CY2025–CY2030，自然年	FY2026–FY2031，公司财年（公司财年截至自然年 1 月底）

来源：Marvell，国金证券研究所

公司对份额提升的判断，体现在定制收入目标跑赢市场的幅度上。Marvell 预计 FY2026 至 FY2031 定制收入复合增速为 75%至 85%，较公司预测的定制市场增速高出约 20 至 30 个百分点。管理层认为，已获定点的 XPU 陆续量产，配套芯片则扩大了可承接的设计项目。按 FY2031 集团收入中点 800 亿美元测算，定制业务约 300 亿美元、互联约 375 亿美元，两者合计约占 84%。

数据中心目标提前一年实现，长期模型指向收入与盈利同步扩张。根据 Marvell 投资者日，公司预计 FY2028 收入约 200 亿美元，较 FY2026 的 82 亿美元增长约 1.4 倍，其中数据中心收入约 180 亿美元，较原计划提前一年实现。公司 FY2031 收入目标为 700 亿至 900 亿美元，经营利润率 44%—46%，隐含每股收益超过 30 美元。

图表29：Marvell 集团收入从 FY2026 实绩到 FY2031 目标，FY31 有望实现 700~900 亿美金收入

公司财年	集团收入（亿美元）	增速与配套指引
FY2026 实绩	约 82	
FY2027E	约 120	同比约 47%
FY2028E	约 200	同比约 67%；数据中心约 180 亿美元
FY2031E	700–900	FY2026–31 复合增速 55%–60%；非 GAAP 经营利润率 44%–46%，隐含 EPS 超 30 美元

来源：Marvell，国金证券研究所

4.2 Marvell 深化 Google TPU 合作，有望获得更多 ASIC 定点

继微软 Maia 后，Marvell 有望借助与 Google 的合作争取更多定制 ASIC 定点。Marvell 此前参与了微软 Maia Braga-R 项目的部分芯粒设计，随后与 Google 的定制芯片合作也推进迅速。据 The Information 于 2026 年 4 月报道，Marvell 与 Google 洽谈开发推理 TPU 及配套内存处理单元；8 月，Marvell 披露已与 Google 签署定制芯片开发协议，双方合作由既有互联业务拓展至推理加速器、存储控制器和网络接口控制器等产品。公司随后在投资者日表示，合作已进入执行阶段。

我们认为，这次合作有助于 Marvell 争取更多 ASIC 定点机会，获得网络接口或存储控制器定点，与博通在配套芯片定制环节直接竞争，给予 Google 更多运算芯片与配套环节的供应商选择。后续可重点关注 Marvell 能否获得推理 TPU 设计定点，以及已签约项目的量产进度。

五、投资建议

我们认为推理算力的需求高速增长，云厂商资本开支持续高企，提升算力投入效率的诉求愈发迫切。ASIC 作为具备较高性价比的产品，有望在推理端得到广泛应用。目前北美 CSP 已经积极布局 ASIC，谷歌 TPU、亚马逊 Trainium 已经有较大规模出货，另外模型厂商如 OpenAI 也在积极布局 ASIC，首代定制芯片已开始流片，有望迎来高速增长。

我们认为 ASIC 设计服务、高速网络、PCB 与整机柜制造有望充分受益。我们看好：天弘科技、博通、鹏鼎控股等，建议关注：沪电股份、Marvell、生益电子。

在 ASIC 设计服务领域，关注 Marvell 在 Google TPU 生态中的新增项目。双方已签署定制芯片商业协议，公司上调了 FY2029 定制业务收入展望。后续关注配套芯片量产与下一代设计定点。在网络环节，我们关注加速器之间的低时延互联和跨柜集群扩容。两类需求同步带动机柜级交换设备升级，我们看好天弘科技机柜级网络产品的放量。

在整机柜制造环节，我们看好天弘科技。公司已参与 OpenAI 首代 ASIC 整柜设计，预计 2027 年开始量产爬坡。

六、风险提示

AI 进展不及预期：如果 AI 进展不及预期，可能导致模型的应用需求不及预期、使用客户数量不及预期，影响 ASIC 的需求，导致相关公司业绩不及预期。

行业竞争加剧：AI 行业趋势明显，布局 AI 的厂商数量持续增多，将导致行业竞争加剧，影响相关公司业绩。

下游需求不及预期：ASIC 终端客户为云厂商，如果云厂商资本开支收缩、需求不及预期，将造成相关公司出货不及预期，导致相关公司业绩不及预期。

技术路线与软件适配风险：大模型架构、算子及推理负载快速变化，若 ASIC 设计与目标负载匹配度下降，或编译器及软件迁移进度滞后，其性能和成本优势可能弱于预期；GPU 平台迭代亦可能压缩 ASIC 的渗透空间。

客户集中与项目调整风险：定制芯片收入与少数云厂商、模型厂商的核心项目密切相关。若大客户推迟项目、调整供应商或扩大内部设计范围，设计服务及整柜供应商的份额与订单可能波动，影响收入增长。

研发验证与量产爬坡风险：先进制程、多芯粒封装及高速互联提高工程复杂度，若流片、良率提升或整柜联调未达预期，可能推迟客户验收和收入确认，并增加研发、返工及售后成本。

关键产能与供应链风险：高端 ASIC 依赖先进制程、先进封装、HBM 及高阶 PCB 等配套。若关键产能扩张不及预期、供应商交付受阻或核心器件涨价，可能限制出货规模并压缩产业链利润空间。

电力与数据中心配套交付风险：吉瓦级算力部署需要机房建设、电网接入及供电、液冷系统同步到位。若电力获取、建设审批或配套工程延迟，可能导致芯片和整柜采购后移，拖慢项目放量。

盈利兑现与营运资金风险：产业链扩产、备货及研发投入通常先于收入兑现。若产能利用率、产品结构或降本进度不及预期，固定成本及存货压力可能上升，经营利润率和现金回报可能低于预期。

贸易政策与地缘政治风险：ASIC 产业链涉及跨境设计、制造与交付，出口管制、关税及技术授权政策变化，可能限制客户覆盖和供应商选择、抬升采购成本或延长交付周期，影响相关公司经营。

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