仅供内部参考，请勿外传

东兴证券股份有限公司证券研究报告

东兴证券 DONGXING SECURITIES

光通信行业 2026 年中期策略：Scale up 成为 AI 数据中心网络创新方向，光互联供应链紧缺是主线

2026 年 6 月 18 日

看好/维持

通信 | 行业报告

分析师 | 石伟晶 电话：021-25102907 邮箱：shi_wj@dxzq.net.cn | 执业证书编号：S1480518080001

投资摘要：

年初至今英伟达与台积电均实现正向涨幅，其中台积电涨幅显著领先英伟达。在全球 AI 时代，英伟达与台积电是 AI 算力产业链最核心供需同盟，英伟达于 2025 年正式取代苹果，成为台积电第一大客户，双方在技术路线、产能规划以及先进半导体制造良率与效率等维度深度合作与协同。2026 年初至 6 月 10 日，台积电股价累计涨幅 34%，英伟达股价累计涨幅 10%。其中英伟达与台积电在 3 月份均有显著回调，主要受到中东冲突升级影响；而年初至今，台积电股价涨幅超过英伟达，则源于 2026 年 AI 算力供应链处于紧缺状态，尤其受到高带宽内存（HBM）和先进封装产能限制。

年初至今北美光通信板块多点开花，投资机会显著。我们整理 15 家北美重要的光通信企业年初至今股价涨幅，并分析市场对其定价逻辑。可以看到，北美 AI 数据中心大规模建设趋势下，受产能短缺影响，市场对光模块上游核心材料、重要器件、测试厂商给予估值溢价。CPO、DCI、智能体等新兴技术方向也获得市场关注，相关公司股价获得不错涨幅。具体如下：（1）上游衬底新材料供需紧缺，AXTI 生产的磷化铟（InP）衬底是 AI 数据中心高速光模块/CPO 所需激光器的关键材料，其公司累计涨幅 453%；（2）中游环节光模块处于扩产周期，数据中心光模块/光芯片供应商股价涨幅突出，AAOI 涨幅 340%、LITE 涨幅 132%、COHR 涨幅 113%、MACOM 涨幅 91%；（3）测试设备是刚需，订单同步高增，全球光通信测试企业 Viavi Solutions (VIAV)股价涨幅 173%；（4）CPO 产业化加速，Marvell 布局硅光子集成技术，股价涨幅 208%；（5）DCI 互联受益于 AI 跨集群长距传输需求，诺基亚股价涨幅 113%，Ciena 股价涨幅 79%，思科股价涨幅 58%；（6）AI 数据中心建设提速，北美地区光纤需求大幅增长，光纤、光缆产品与一体化连接方案供应商康宁股价涨幅 92%；（7）AI 工作负载从简单的文本生成向复杂的智能体和强化学习演进，CPU 与 GPU 配比提升，英特尔股价涨幅 176%。

受北美市场定价逻辑映射，年初至今 A 股光通信板块上涨显著。全球高速率光模块处于扩产周期以及国内头部企业提前卡位产能，中际旭创股价涨幅 97%、新易盛股价涨幅 94%、天孚通信股价涨幅 107%、源杰科技股价涨幅 179%、仕佳光子股价涨幅 72%、长光华芯股价涨幅 147%。测试设备是光模块扩产刚需，订单同步高增，联讯仪器股价涨幅 171%，罗博特科股价涨幅 113%；CPO、NPO 产业化加速，光迅科技股价涨幅 181%，华工科技股价涨幅 79%，炬光科技股价涨幅 100%，福晶科技股价涨幅 27%；DCI 互联受益于 AI 跨集群长距传输需求，德科立股价涨幅 48%；AI 数据中心建设提速，光纤需求大幅增长，亨通光电股价涨幅 335%，中天科技股价涨幅 172%。

关于 2026 年下半年光通信板块行业主要发展趋势以及投资策略，观点如下：

Scale up 成为 AI 数据中心网络创新方向。超节点是通过高速互联协议与专用交换芯片构建的高带宽域（High-Bandwidth Domain），将数十至数百颗 GPU 芯片在逻辑上整合为统一编址、低延迟、高带宽的协同计算系统。而超节点的实现核心在于构建高带宽、低延迟的 Scale-Up（纵向扩展）通信域。随着模型参数规模的增加，张量并行和专家并行规模随之扩大，对超节点内 HBD 域规模的需求也越来越大。基于以

仅供内部参考，请勿外传

太网的 Scale up 网络开放生态处于快速发展中。以英伟达、Google 为典型代表，二者均采用专有协议：NVLink 向第三方半开放 CPU/Chiplet 接入权限；Google ICI Link 则服务于自研 TPU 集群；基于 Ethernet 的开放技术方向，以各大互联网和云计算公司以及一些 GPU 芯片公司为代表。其中 UALink 基于标准以太网组件打造开放互联协议，华为灵衢协议从 2.0 版本起转向开放标准。

英伟达确立超节点技术范式。英伟达超节点的优势建立在 NVLink 和 NVLink Switch。2026 年 1 月，英伟达发布第六代 NVLink 以及 NVLink 交换机，两者支持最新的 Rubin 架构。从性能指标看，在最新 VR NVL72 超节点中，第六代 NVLink 支持 3.6TB/s 的 GPU-to-GPU 通信带宽，NVLink 交换机提供 260TB/s 聚合带宽。此外，在 VR NVL72 中，英伟达采用 PCIe Gen6 协议实现 Vera CPU 与超级网卡 CX-9 互联，Vera 与 CX-9 之间接口双向总带宽达到 768GB/s；采用以太网/ InfiniBand 协议实现超级网卡 CX-9 与 OS 互联，单个端口即可提供 1x800G 的传输能力；超级网卡 ConnectX-8/9 内置 PCIe Gen6 交换模块，替代传统独立 PCIe 交换机。

CW 激光器成为硅光架构重要配置，市场增速陡峭。数据中心下游 scale up 网络的发展，成为 CPO/NPO 等新兴光互联技术重要的应用场景；而 CPO/NPO 的规模化应用将驱动硅光架构有望成为光互联行业主流技术路径。对于硅光光互连产品，BOM 结构重构，原有的分立调制器与大量无源光器件被集成为一颗硅光芯片（PIC），PCB 与机构件也被大幅简化；由于硅材料本身发光效率低，难以直接实现高效光发射。当前外置 CW（连续波）光源成为硅光光模块的主流方案。在此技术趋势下，数据中心激光器芯片市场增速陡峭，2030 年市场或将超过 200 亿美元。

当下 CPO 已经成为光模块行业确定趋势。传统方案是采用可插拔光模块，部署在交换机前面板。但根据 AI 网络带宽的发展路线图，互连的速度、距离、密度及可靠性要求即将超越传统光模块所能提供的极限。为突破上述局限，CPO（光电共封装）已成为业界公认的 AI 及超算高密度互连终极方案，其通过将光引擎 OE（Optical Engine）和交换芯片 ASIC 共基板封装在一起，实现极致能效、带宽密度与低时延。LightCounting 在 2026 年 4 月对 1.6T CPO 产品出货量进行显著上调。2023-2026 年，1.6T CPO 产品为技术导入期，出货量几乎为零，处于试点部署阶段；2027 年起进入大规模放量阶段，增长曲线陡峭上扬，2029 年 1.6T CPO 产品出货量预测从约 200 万个大幅上调至约 900 万个，2031 年则从约 500 万个上调至约 1300 万个。市场规模方面，CPO 市场 2027 年有望突破 50 亿美元，2030 年增长至 150 亿美元。

受益于数据中心建设浪潮，G.657 单模光纤和多模光纤将迎来规模化应用。根据 CRU 最新报告，2026 年全球数据中心光纤需求预计达到 9160 万芯公里，同比增长 32%。G.657 单模光纤适用于数据中心，具有优异的耐弯曲特性，其弯曲半径可实现常规的 G.652 光纤的弯曲半径的 1/4～1/2。多模光纤产品（OM3、OM4、OM5），适用于数据中心短距离、高带宽的场景。字节跳动 AI 智算中心规模化交付，驱动国内数据中心光纤市场超过 50 亿元。2026 年是字节跳动自建 AI 智算中心规模化交付之年。2024 年至今，我们统计字节跳动在国内自建 6 座大型 AI 智算中心，总投资规模 357 亿元。其中火山云太行算力中心项目一期工程投资规划 28 亿元，规划 12 万台高性能 AI 服务器，2025 年进入运营阶段；其他 4 座大型 AI 智算中心预计 2026 年投入运营；第 6 座火山引擎内蒙古和林格尔算力中心二期项目预计 2027 年投入运营。经测算，2026 年字节跳动自建 1GW 智算中心，预计光纤资本支出约 9.1-10.4 亿元；2026 年国内数据中心光纤市场规模约 55-63 亿元。我们认为，字节跳动庞大的资本开支具有扎实的市场需求，具有可持续性和引领性。从短期看，2026 年是字节跳动自建 AI 智算中心规模化交付之年，有望显著提升国内数据中心光纤市场需求。从中长期看，随着国内 AI 应用市场进一步繁荣发展，国内更多科技企业将加大 AI 智算中心投资。

投资策略：第一，超节点与 Scale up 网络是突破算力通信瓶颈、支撑万亿级大模型与高实时性应用的关键基础设施。2026 年国内和北美超节点陆续进入量产交付期，英伟达确立超节点技术范式。国内 AI 算力网络在高速互联领域仍存在技术代差，自主可控需求迫切。第二，当前光互联供应链持续性紧缺，有望贯穿

仅供内部参考，请勿外传

2026 全年。我们认为，国内光芯片企业有望受益于光互联市场总量增长、硅光架构 CW 激光器结构性机会、国产份额提升三重机遇；Scale up 网络正转向光互联，CPO 生态逐步形成，全球硅光子将从碎片化定制研发转向标准化代工量产模式。国内硅光新材料、无源以及有源器件、以硅光器件设计套件（PDK）、集成测试等厂商有望进入英伟达、台积电供应链体系，加速硅光子国产化验证与量产落地。此外，全球 AI 应用及数据中心建设快速发展，产能缺口驱动光纤价格持续提升。国内光纤企业具有数据中心特种光纤产能优势，有望显著并持续受益。

相关公司：

国产超节点：华勤技术、浪潮信息、中科曙光、工业富联；

国产高速交换芯片：盛科通信、紫光股份、锐捷网络、中兴通讯；

光模块、CPO/NPO：中际旭创、新易盛、天孚通信、华工科技、光迅科技；

光芯片/电芯片：源杰科技、东山精密、仕佳光子、长光华芯、优迅股份；

OCS：腾景科技、福晶科技、矩光科技、德科立；

CPO 设备：联讯仪器、罗博特科；

光纤：亨通光电、中天科技、长飞光纤、烽火通信。

风险提示：（1）英伟达超节点出货量低于预期；（2）CPO 技术路线碎片化；（3）光纤产能过剩与价格下行；（4）AI 应用端增长不及预期；（5）地缘政治风险。



仅供内部参考，请勿外传

目 录

1. AI 产业链 2026 年上半年行情复盘：光互联供应链紧缺是主线 7
2. Scale up 成为 AI 数据中心网络创新方向 12
3. 超节点：AI 数据中心标准化算力单元，英伟达技术生态处于领先地位 17
3.1 英伟达引领超节点技术进步 17
3.2 构建超节点的关键在于 NVLink 和 NVLink Switch 21
3.3 GB200 NVL72 超节点总交换容量 129.6TB/s 24
3.4 VR200 NVL72 超节点总交换容量翻倍至 259.2TB/s 26
3.5 VR NVL72 计算托盘通信速率三重升级：1.8TB/s NVLink-C2C+ PCIe Gen6+800G 以太网 28
3.6 ConnectX 超级网卡价值显著提升：内置 PCIe 交换模块与以太网交换逻辑 33
4. 光芯片：CW 激光器重构全球光芯片产业格局，IDM 模式构建产业链壁垒 37
4.1 技术路径分化趋势清晰，CW 激光器成为硅光架构重要配置 37
4.2 激光器芯片市场增速陡峭，2030 年市场或将超过 200 亿美元 39
4.3 激光芯片制造价值链：外延生长与光栅工艺具有技术壁垒及产能瓶颈 42
4.4 国内外企业同步研发高功率 CW 激光器芯片，不存在代际差距 48
5. CPO：产业化提速，台积电 COUPE 引领硅光集成落地 51
5.1 CPO 成为光模块终极方案，硅光子产业链迎来发展拐点 51
5.2 CPO 架构下光引擎跃升为价值链核心，形成 2D 封装到 3D 单片集成多元技术路线 54
5.3 台积电 COUPE 平台实现技术落地与迭代，绑定标杆客户 57
6. 光纤：受益于数据中心建设浪潮，迎来景气周期 59
6.1 G.657 单模光纤和多模光纤将迎来规模化应用 59
6.2 海外康宁光通信业务高增，映射国内数据中心光纤市场机会 63
6.3 字节跳动 AI 智算中心规模化交付，驱动国内数据中心光纤市场超过 50 亿元 65
7. 通信行业投资策略 69
8. 风险提示 69

插图目录

图 1： 2026 年年初至今英伟达与台积电股价累计涨跌幅 7
图 2： 2026 年年初至今北美五大 CSP 股价累计涨跌幅 8
图 3： 2026 年年初至今北美光通信板块上市企业累计涨幅 10
图 4： 2026 年年初至今 A 股光通信板块上市企业累计涨幅 11
图 5： 人工智能训练与推理对网络需求存在显著差异 12
图 6： Scale up 网络（左）与 Scale out 网络（右）特点对比 14
图 7： 全球主流算力方案对应 Scale Up 协议 15
图 8： 全球主流算力芯片厂商旗下 Scale up 协议特点 16
图 9： GTC2026 大会上 NVIDIA Vera Rubin NVL72 机架 18
图 10： VR NVL72 机柜组织图 19
图 11： VR200 NVL72 机柜计算托盘 20

仅供内部参考，请勿外传

图 12： Vera Rubin NVL72 机柜交换机托盘 20
图 13： NVLink 技术规格参数对比 21
图 14： NVLink 交换机规格参数对比 21
图 15： NVLink 网状拓扑结构提供高速双向带宽 22
图 16： NVLink 交换网络的演进过程（1） 23
图 17： NVLink 交换网络的演进过程（2） 23
图 18： GB200 NVL72 中计算托盘 24
图 19： GB 200 NVL72 中 NVLink 交换机托盘 24
图 20： B200 端口 Port 示意图 25
图 21： NVLINK Switch5 芯片 Port 示意图 25
图 22： GB200/300 NVL72 单层计算托架的互联拓扑 25
图 23： VR200 NVL72 机柜计算托盘 26
图 24： Vera Rubin NVL72 机柜交换机托盘 26
图 25： VR200 NVL72 机柜中 GPU 互联拓扑结构 27
图 26： VR NVL72 机柜计算托盘顶视图 28
图 27： VR NVL72 计算托盘拓扑结构 29
图 28： VR NVL72 中 Rubin-Vera 芯片互联方式 30
图 29： VR NVL72 运算托盘信号路径图 31
图 30： VR NVL72 计算托盘侧视图 32
图 31： ConnectX-9 SuperNIC 与 ConnectX-8 SuperNIC 产品 33
图 32： ConnectX-8 网卡技术栈 34
图 33： 超级网卡是连接 GPU 集群与外部的关键 ASIC 35
图 34： Spectrum-X 交换机到超级网卡 ConnectX 的端到端网络处理，带来高性能以太网表现 36
图 35： 数据中心光互联产业链 37
图 36： 硅光模块内部结构 38
图 37： 全球光互联市场规模预测（亿美元） 39
图 38： 全球数据中心光互联市场规模预测（亿美元） 39
图 39： 全球激光器芯片市场规模预测（亿美元） 40
图 40： CW 激光器芯片合计市场规模预测（亿美元） 40
图 41： 不同功率 CW 激光器芯片市场规模预测（亿美元） 41
图 42： 激光器芯片行业主要采用 IDM 模式 42
图 43： CW 激光器芯片关键制造步骤及相关设备 43
图 44： CW 激光器芯片外延生长步骤 44
图 45： 两种光栅工艺制造的光栅结构对比 45
图 46： 两种光栅工艺对激光器芯片产品特性影响 46
图 47： 光栅结构制作流程具体步骤 46
图 48： 光栅环节委外加工与自产平均成本对比（元/片） 47
图 49： 2025 年全球用于 400G 及以上光互联产品的激光器芯片销售收入排名（亿元，%） 49
图 50： 2025 年全球用于 400G 及以上硅光光互联产品的激光器芯片销售收入排名（亿元，%） 49
图 51： Coherent 公司推出 400 mW CW InP 激光器芯片 50

仅供内部参考，请勿外传

图 52： 源杰科技研发 300mW 高功率 CW 激光器芯片 50
图 53： 英伟达首款 CPO 交换机 Quantum-X 内部结构 51
图 54： 全球硅光子产业生态与供应链 52
图 55： 全球 1.6T CPO 共封装光学出货预测 53
图 56： 硅光子系统级应用架构 54
图 57： CPO 涉及多种光引擎集成封装方法 56
图 58： 台积电发布的 CPO 技术路线图 57
图 59： 全球数据中心光纤应用规模（万芯公里） 59
图 60： 多模光纤产品示意图 61
图 61： 单管主干跳线与多单元主干跳线产品示意图 61
图 62： 单模光纤和多模光纤产品传输距离对比 62
图 63： 康宁光通信业务收入及增速（亿美元，%） 63
图 64： 康宁光通信业务净利润及增速（亿美元，%） 63
图 65： 康宁光通信数据中心营收及增速（亿美元，%） 64
图 66： 康宁光通信运营商营收及增速（亿美元，%） 64
图 67： 康宁光通信净利率情况（%） 64
图 68： 2025 年 12 月国内 AI 原生 App 月活数据（万人） 65
图 69： 豆包日均 Tokens 使用量变化 66
图 70： 2025H1 中国公有云大模型服务调用量分布 66

表格目录

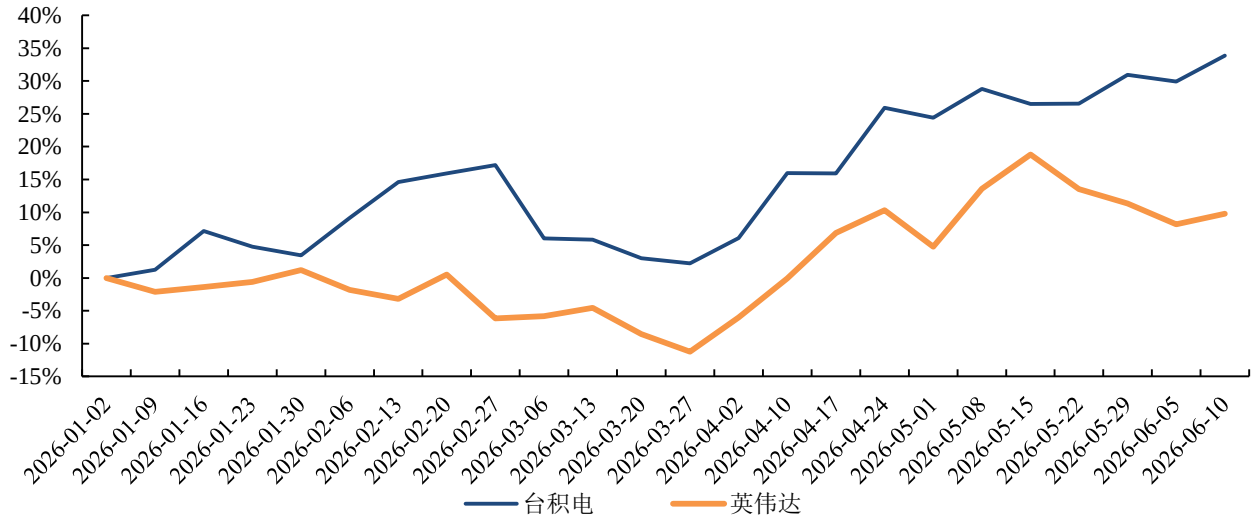
表 1： AI 大语言模型训练中多种并行计算方式对比 13
表 2： 英伟达超节点 Scale up 迭代路线 17
表 3： 激光器芯片价值链分布情况 43
表 4： 2025 年全球激光器芯片市场头部企业市场份额 48
表 5： 英伟达与博通发布的 CPO 交换机产品矩阵 58
表 6： ITU 标准下单模光纤产品 60
表 7： 2024 年至今字节跳动国内数据中心自建项目 67
表 8： 字节跳动 AI 数据中心光纤资本开支测算 68

仅供内部参考，请勿外传

1. AI 产业链 2026 年上半年行情复盘：光互联供应链紧缺是主线

年初至今英伟达与台积电均实现正向涨幅，其中台积电涨幅显著领先英伟达。在全球 AI 时代，英伟达与台积电是 AI 算力产业链最核心供需同盟，英伟达于 2025 年正式取代苹果，成为台积电第一大客户，双方在技术路线、产能规划以及先进半导体制造良率与效率等维度深度合作与协同。2026 年初至 6 月 10 日，台积电股价累计涨幅 34%，英伟达股价累计涨幅 10%。其中英伟达与台积电在 3 月份均有显著回调，主要受到中东冲突升级影响；而年初至今，台积电股价涨幅超过英伟达，则源于 2026 年 AI 算力供应链处于紧缺状态，尤其受到高带宽内存（HBM）和先进封装产能限制。

图1：2026 年年初至今英伟达与台积电股价累计涨跌幅

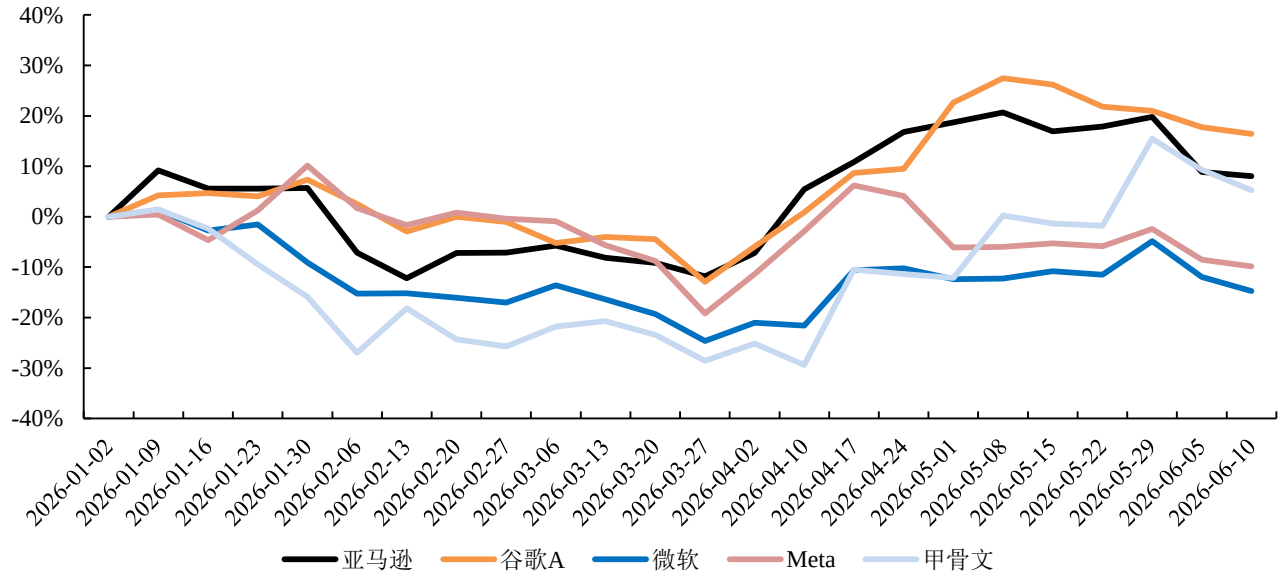


资料来源：i

仅供内部参考，请勿外传

年初至今五大 CSP 股价表现一般，市场对 AI 基建持续性存在分歧。2026 年初至 6 月 10 日，谷歌股价累计涨幅 16%，亚马逊股价累计涨幅 8%，甲骨文股价累计涨幅 5%，Meta 股价累计涨幅-10%，微软股价累计涨幅-15%。2026 年一季度，北美五大 CSP 合计资本开支 1391 亿美元，全年预计资本开支计划 7300 亿美元-7600 亿美元，一季度资本开支占比全年约 18%-19%。可以看到，北美五大 CSP 保持对 AI 基础设施的巨额投入，但由于市场担心现阶段 AI 投资－需求节奏存在错配，投资或不具备可持续性。

图2：2026 年年初至今北美五大 CSP 股价累计涨跌幅



资料来源：i

仅供内部参考，请勿外传

年初至今北美光通信板块多点开花，投资机会显著。我们整理 15 家北美重要的光通信企业年初至今股价涨幅，并分析市场对其定价逻辑。可以看到，北美 AI 数据中心大规模建设趋势下，受产能短缺影响，市场对光模块上游核心材料、重要器件、测试厂商给予估值溢价。CPO、DCI、智能体等新兴技术方向也获得市场关注，相关公司股价获得不错涨幅。

其中上游衬底新材料供需紧缺，AXTI 生产的磷化铟（InP）衬底是 AI 数据中心高速光模块/CPO 所需激光器的关键材料，其公司累计涨幅 453%；

中游环节光模块处于扩产周期，数据中心光模块/光芯片供应商股价涨幅突出，AAOI 涨幅 340%、LITE 涨幅 132%、COHR 涨幅 113%、MACOM 涨幅 91%；

测试设备是刚需，订单同步高增，全球光通信测试企业 Viavi Solutions (VIAV)股价涨幅 173%；

CPO 产业化加速，Marvell 布局硅光子集成技术，股价涨幅 208%；

DCI 互联受益于 AI 跨集群长距传输需求，诺基亚股价涨幅 113%，Ciena 股价涨幅 79%，思科股价涨幅 58%；

AI 数据中心建设提速，北美地区光纤需求大幅增长，光纤、光缆产品与一体化连接方案供应商康宁股价涨幅 92%；

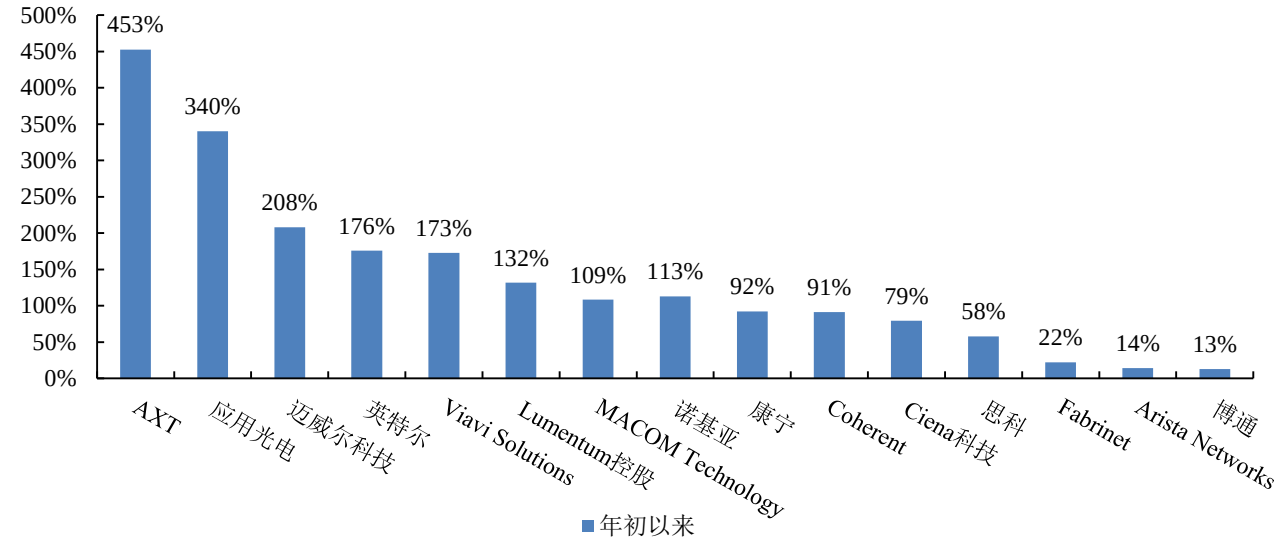
AI 工作负载从简单的文本生成向复杂的智能体和强化学习演进，CPU 与 GPU 配比提升，英特尔股价涨幅 176%。

年初至今，
方开放图纸标准，同时给多家竞争设计方代工，无排他捆绑，承接博通 CPO 光子封装、Arista 整机配套光引擎、Lumentum/Nvidia 光模块的制造封装；博通专注芯片层
DSP、光子芯片，把硬件组装、系统软件、光电封装全部外包给生态伙伴；Arista 专注整机软件层，放弃芯片 - 整机 - 制造一体化重资产模式，只做整机方案与软件价值。标准化采购博通商用 ASIC，自研 EOS 开放操作系统。

我们认为，三家公司商业模式具有相似性，均扎根定位数据中心与光通信开放解耦生态。但 2026 年 AI 供应链处于紧缺周期，开放解耦生态则受到显著影响。相反，英伟达为保证供应链稳定，不再只靠采购下单，而是用金融深度绑定，锁定扩产产能。例如英伟达战略入股康宁、Marvell、Lumentum、Coherent 等关键上游供应链，与台积电建立深度合作关系。

仅供内部参考，请勿外传

图3：2026 年年初至今北美光通信板块上市企业累计涨幅



资料来源：i

仅供内部参考，请勿外传

受北美市场定价逻辑映射，年初至今 A 股光通信板块上涨显著。

全球高速率光模块处于扩产周期以及国内头部企业提前卡位产能，中际旭创股价涨幅 97%、新易盛股价涨幅 94%、天孚通信股价涨幅 107%、源杰科技股价涨幅 179%、仕佳光子股价涨幅 72%、长光华芯股价涨幅 147%。

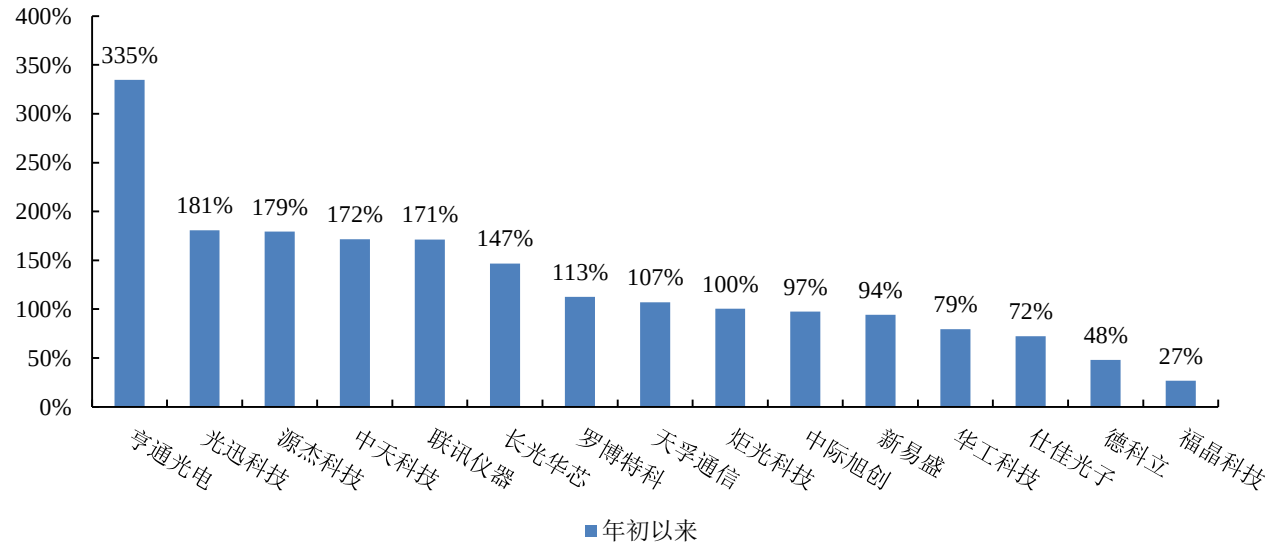
测试设备是光模块扩产刚需，订单同步高增，联讯仪器股价涨幅 171%，罗博特科股价涨幅 113%；

CPO、NPO 产业化加速，光迅科技股价涨幅 181%，华工科技股价涨幅 79%，炬光科技股价涨幅 100%，福晶科技股价涨幅 27%；

DCI 互联受益于 AI 跨集群长距传输需求，德科立股价涨幅 48%；

AI 数据中心建设提速，光纤需求大幅增长，亨通光电股价涨幅 335%，中天科技股价涨幅 172%。

图4：2026 年年初至今 A 股光通信板块上市企业累计涨幅



资料来源：i

仅供内部参考，请勿外传

以下章节是我们对 2026 年下半年光通信行业主要发展趋势总结。

2. Scale up 成为 AI 数据中心网络创新方向

AI 训练和推理是大语言模型两个核心工作负载，但两者对网络需求存在显著差异。

AI 推理网络具有分布式、低延迟、按需任务调度、高效率、强外部交互特点。AI 推理场景下，推理请求往往是零散、独立的（比如用户单次提问、单次图像识别），任务被拆分成多个轻量、并行的小任务，分布在不同的服务器/计算单元上处理；用户对推理结果的等待时间有严格要求，毫秒级延迟都会直接影响体验；推理流量是突发的、按需的，需要网络能灵活调度任务，同时依赖大规模集群来分摊负载、保证整体效率；推理系统需要频繁与外部数据交互。

AI 训练网络具有长周期、大规模、高同步性与稳定性、弱外部交互特点。AI 训练场景下，训练是一个持续数天甚至数周的过程，需要大量 GPU/TPU 之间频繁同步模型参数、梯度，对网络的同步性和稳定性要求极高；训练通常是单个超大规模任务，需要跨机房、跨园区的集群协同，因此网络需要支持长距离、高带宽的稳定传输；训练中如果出现个别节点/链路的延迟过高（尾部延迟），会拖慢整个集群的同步节奏，导致整体训练效率下降，因此网络需要严格控制延迟抖动；训练过程主要是集群内部的数据传输，和外部用户/系统的交互很少，因此对外部接口的需求低，更关注内部网络的性能。

图5：人工智能训练与推理对网络需求存在显著差异



资料来源：英伟达，东兴证券研究所

仅供内部参考，请勿外传

张量并行与专家并行计算要求 AI 网络高带宽与极低时延。大模型参数规模从千亿级向万亿级乃至十万亿级演进，跨服务器张量并行计算成为必然选择。张量并行要求多张卡一起完成一个层内计算，因此算力网络会在模型前向、反向过程中反复通信；此外，混合专家（MoE）模型在 Transformer 架构大模型中规模化应用，每个 token 会被路由到不同 GPU，产生大量分发和聚合通信。

表1：AI 大语言模型训练中多种并行计算方式对比

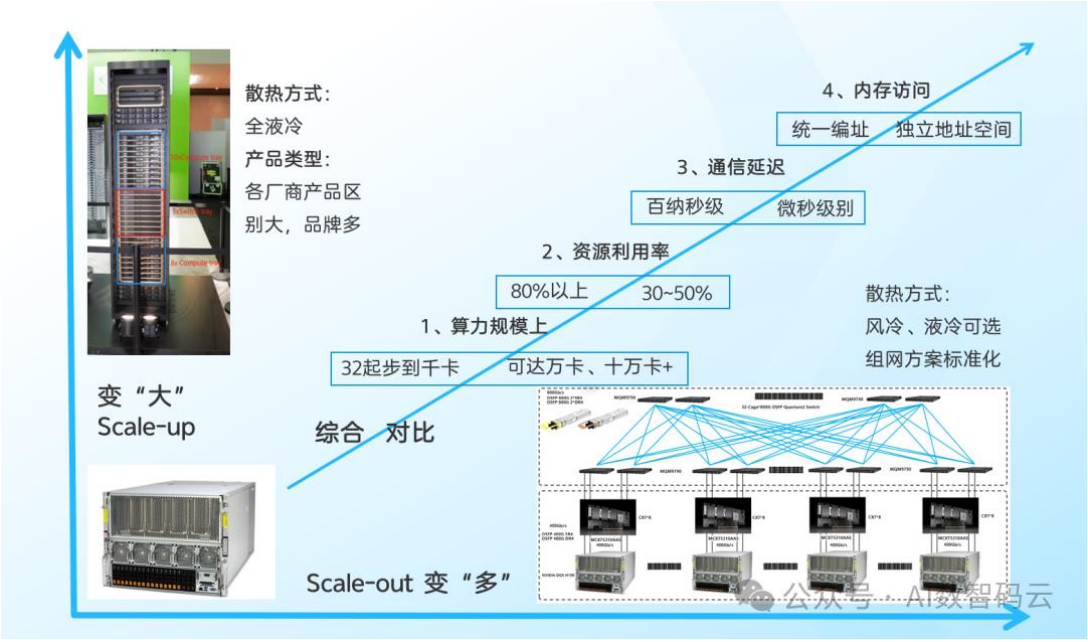
并行方式	带宽要求	延迟要求	说明
张量并行(TP)	数百至数千 GB/s 级	延迟要求极高	将单个运算（如矩阵乘法）拆分到不同 GPU 上运行，通常在机内完成
专家并行(EP)	数百至数千 GB/s 级	延迟要求极高	基于不同的任务选择不同专家进行训练，引入 All to All 流量，适合机内完成
流水线并行（PP）	MB/s 至 GB/s 级	延迟要求较高	将模型的不同层划分为若干个阶段，每个阶段可以在不同的 GPU 上执行，通常在机间完成
数据并行（DP）	GB/s 级	延迟要求较高	将同一批数据分割成多个子集，并将每个子集分配给不同 GPU 上（模型实例相同）运行，通常在机间完成

资料来源：网络技术趋势洞察公众号，东兴证券研究所

仅供内部参考，请勿外传

超节点是通过高速互联协议与专用交换芯片构建的高带宽域（High-Bandwidth Domain），将数十至数百颗 GPU 芯片在逻辑上整合为统一编址、低延迟、高带宽的协同计算系统。而超节点的实现核心在于构建高带宽、低延迟的 Scale-Up（纵向扩展）通信域。随着模型参数规模的增加，张量并行和专家并行规模随之扩大，对超节点内 HBD 域规模的需求也越来越大，Scale up 成为 AI 数据中心网络创新方向。

图6：Scale up 网络（左）与 Scale out 网络（右）特点对比

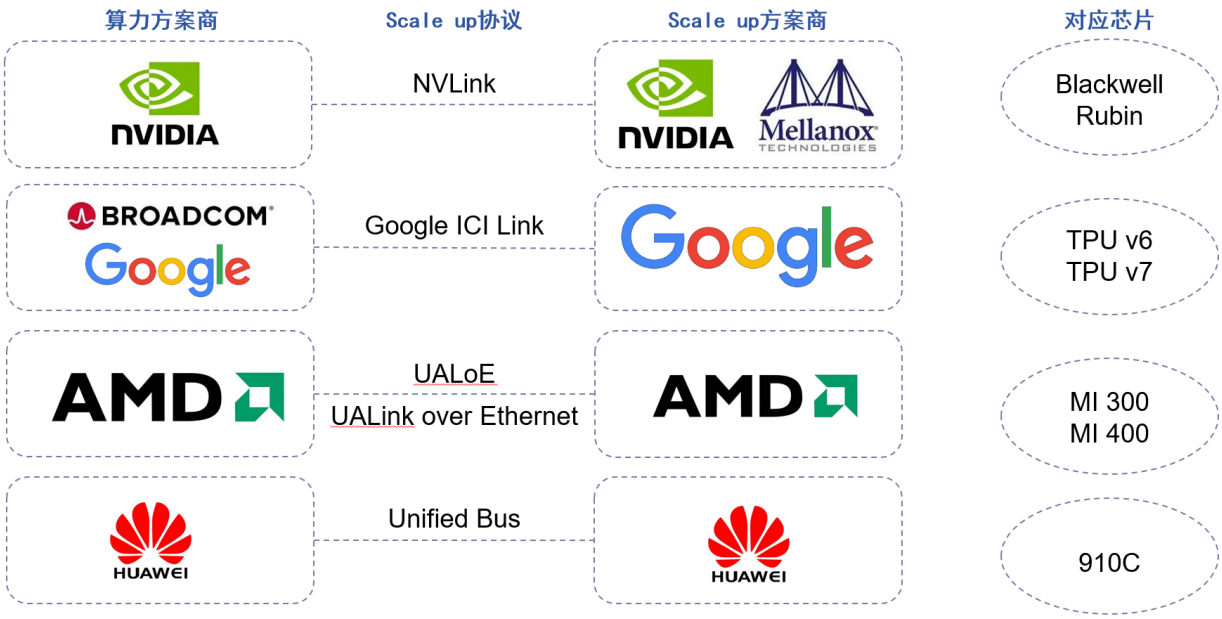


资料来源：AI 数智码云公众号，东兴证券研究所

仅供内部参考，请勿外传

全球主流算力芯片厂商推出各自 Scale up 网络标准。英伟达在 Scale up 网络中采用自研的 NVLink 高速互连技术；AMD 与 AWS、思科、谷歌等公司组成超以太网联盟（UALink）；Google 采用私有 ICI 协议，机柜之间运用 OCS 光交换技术；华为推出自研的灵衢协议技术（UB）。

图7：全球主流算力方案对应 Scale Up 协议



资料来源：傅里叶的猫公众号，东兴证券研究所

仅供内部参考，请勿外传

基于以太网的 Scale up 网络开放生态处于快速发展中。以英伟达、Google 为典型代表，二者均采用专有协议：NVLink 向第三方半开放 CPU/Chiplet 接入权限；Google ICI Link 则服务于自研 TPU 集群；基于 Ethernet 的开放技术方向，以各大互联网和云计算公司以及一些 GPU 芯片公司为代表。其中 UALink 基于标准以太网组件打造开放互联协议，华为灵衢协议从 2.0 版本起转向开放标准。

图8：全球主流算力芯片厂商旗下 Scale up 协议特点

	NVIDIA NVLink™	Google ICI Link	ULTRA ACCELERATOR LINK™	灵衢 UnifiedBus
主导方	英伟达	谷歌	超以太网联盟（Meta、微软、英特尔、AMD 等）	华为
协议性质	专有协议	专有协议	开放标准	自灵衢 2.0 起开放标准
技术标准	基于 SerDes 的私有协议	ICI 协议（支持 3D 环面拓扑）	使用标准以太网组件	基于高速 SerDes 的物理层
连接形式	机柜内通过铜缆实现 GPU 互联	结合铜缆与光缆连接，立方体内部使用铜缆连接，立方体外部使用光缆连接	同时支持铜缆和光缆	机柜内使用电缆互联，不采用光模块
生态支持	NVLink Fusion，通过半定制 CPU 或 Chiplet 向第三方 GPU 开放生态系统	需通过 XLA 编译器将计算图转为 TPU 机器码，深度绑定 Google 生态	兼容多厂商 GPU（如 AMD Instinct MI 系列、英特尔 Gaudi 等）	全栈开放协议，处于建设初期，第三方商用 GPU 产品尚未大规模上市

资料来源：Semi Analysis，CSDN，东兴证券研究所

仅供内部参考，请勿外传

3. 超节点：AI 数据中心标准化算力单元，英伟达技术生态处于领先地位

3.1 英伟达引领超节点技术进步

自 2025 年开始，Scale up 网络与超节点成为 AI 算力网络重要的技术创新方向。从 AI 基建竞争维度，AI 芯片厂商从芯片算力性能竞争延续至芯片+Scale up 网络的双战场。因此，除了原先英伟达、华为、AMD 以及谷歌等芯片公司，全球更多厂商加入超节点赛道的竞争，包括微软、Meta、Amazon、中国移动、阿里巴巴、腾讯、百度、中科曙光、中兴通讯、浪潮信息、紫光股份等。

英伟达引领超节点技术进步。2024-2025 年，英伟达陆续推出 GH200 NVL72、GB200/ GB300 NVL72 等成熟超节点解决方案。展望 2026-2027 年，英伟达计划推出 Vera Rubin NVL144 和 Rubin Ultra NVL576。互联 GPU 数将从 72 颗进一步向 576 颗发展。

表2：英伟达超节点 Scale up 迭代路线

架构	Blackwell Ultra	Vera Rubin NVL72	Vera Rubin NVL144	Rubin Ultra NVL576	
首发时间	2025-03	2026-01	预计 2026 下半年	预计 2027 年	预计 2028 年
核心平台	GB300 NVL72	VR200 NVL72	VR200 NVL144	Rubin Ultra NVL576	
计算托盘	18 个（单盘 4GPU+2CPU）	18 个（单盘 4GPU+2CPU）	36 个（单盘 4GPU+2CPU）	72 个（单盘 8GPU+4CPU）	144 个（单盘 8GPU+4CPU）
CPU	36 Grace CPUs（72 核）	36 Vera CPUs（72 核）	72 Vera CPUs（88 核）	288 Vera Ultra CPUs（176 核）	576（256 核）
	单颗内存带宽 3.6Tpbs	单颗内存带宽 4.8Tpbs	单颗内存带宽 4.8Tpbs	单颗内存带宽 9.6Tpbs	单颗内存带宽 19.2Tpbs
	NVLink C2C 0.9Tpbs	NVLink C2C 1.8Tpbs	NVLink C2C 1.8Tpbs	NVLink C2C 3.6Tpbs	NVLink C2C 7.2Tpbs
GPU	72 GB300 GPUs	72 VR200 GPUs	144 VR200 GPUs	576 VR300 GPUs	1152
	单颗 288G HBM3E	单颗 512G HBM3E	单颗 512G HBM3E	单颗 1TB HBM4E	单颗 2TB HBM5E
	单颗 MV	单颗 MV 50P	单颗 MV 50P	单颗 MV 100P	单颗 MV 200P
Scale up	铜缆背板	铜缆背板	铜缆背板+板载无源光引擎（非 CPO）	3.2T CPO 硅光（规划）	6.4T CPO 硅光（规划）
	18 个 NVLink5	36 个 NVLink6	72 个 NVLink6	144 个 NVLink7	72 个 NVLink8 硅光交换机
	单个 144*200G 28.8T	单个 72*400G 28.8T	单个 72*400G 28.8T	单个 144*800G 115.2T	单个 288*1.6T 460.8T
	GPU 侧 NVLink 带宽 18*1.8TBps	GPU 侧 NVLink 带宽 18*3.6TBps	GPU 侧 NVLink 带宽 18*3.6TBps	GPU 侧 NVLink 带宽 36*7.2TBps	GPU 侧 NVLink 带宽 72*14.4TBps
Scale out	Spectrum-5 800G OS 可插拔	ɔectrum-6 CPO 硅光	Spectrum-6 CPO 硅光	Spectrum-7 CPO 硅光	Spectrum-8 CPO 硅光
	64*800G 51.2T	128*800G 102.4T	128*800G 102.4T	256*1.6T 409.6T	512*3.2T 1638.4T

注：28.8Tbps=3.6TBps

信息来源：光芯之路公众号，东兴证券研究所

仅供内部参考，请勿外传

2026 年 CES 展会上，英伟达发布新一代超节点 VR200 NVL72，也是全球领先的 Scale up 网络算力平台。据英伟达公布的数据，Rubin 平台的训练性能达到前代 Blackwell 的 3.5 倍，运行 AI 软件的性能提升 5 倍。此外，与前一代相比，英伟达 Rubin 平台在训练 MoE 模型时所需的 GPU 数量减少至原来的四分之一，进一步推动人工智能的普及应用。

图9：GTC2026 大会上 NVIDIA Vera Rubin NVL72 机架

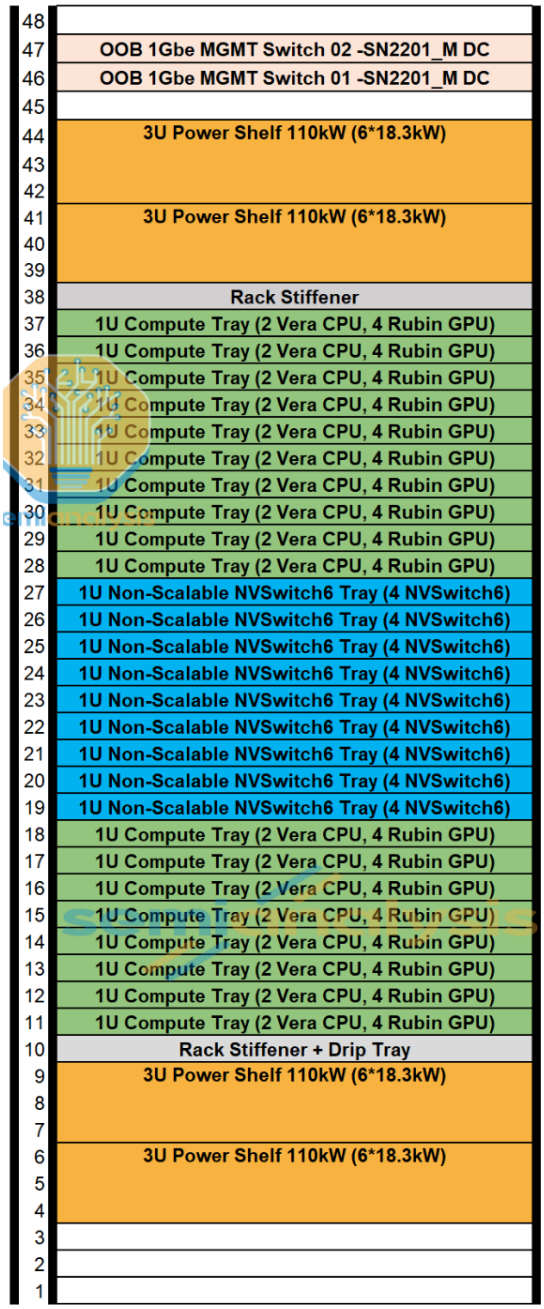


资料来源：GTC2026 大会，东兴证券研究所

仅供内部参考，请勿外传

超节点产品主要由计算节点、交换节点和 Scale up 网络互联构成。以 VR NVL72 机柜为例，其硬件构成包括：运算托盘与交换托盘、顶部的 OOB 管理交换机、电源柜、机架加强件。

图10：VR NVL72 机柜组织图

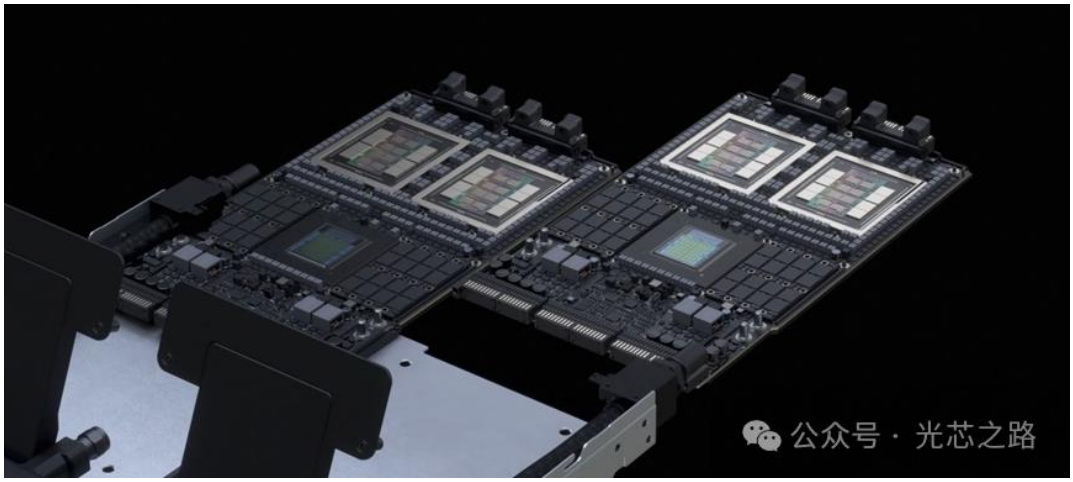


资料来源：SemiAnalysis，东兴证券研究所

仅供内部参考，请勿外传

计算节点：Rubin NVL72 机架通过无缆互联架构整合 18 个计算托盘，每个托盘 2 颗超级芯片，每颗超级芯片集成 1 个 Vera CPU 与 2 块 Rubin GPU，共 72 GPU 与 36 CPU。

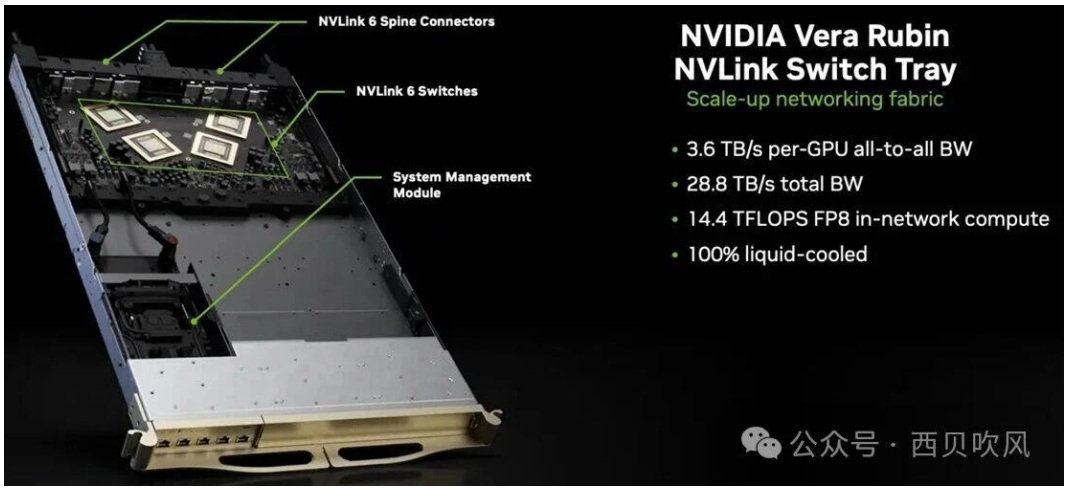
图11：VR200 NVL72 机柜计算托盘



资料来源：光芯之路公众号，东兴证券研究所

交换节点：VR200 NVL72 配置 9 个交换托盘，每个托盘集成 4 颗第六代 NVSwitch 芯片，全机柜部署 36 颗 NVSwitch。相比 GB200 NVL72，NVSwitch 芯片数量实现翻倍。

图12：Vera Rubin NVL72 机柜交换机托盘



资料来源：西北吹雪公众号，东兴证券研究所

仅供内部参考，请勿外传

3.2 构建超节点的关键在于 NVLink 和 NVLink Switch

NVLink 协议是英伟达构建超节点的关键。NVLink 协议从早期点对点互联发展到如今全互联通信，并支持多代 GPU 架构算力芯片。2026 年 1 月，英伟达发布第六代 NVLink 以及 NVLink 交换机，两者支持最新的 Rubin 架构。从性能指标看，在最新 VR NVL72 超节点中，第六代 NVLink 支持 3.6TB/s 的 GPU-to-GPU 通信带宽，NVLink 交换机提供 260TB/s 聚合带宽。

图13：NVLink 技术规格参数对比

NVLink　NVLink 交换机

	第四代	第五代	第六代
每 GPU 的 NVLink 带宽	900GB/s	1,800GB/s	3,600 GB/s
每 GPU 的最大链路数	18	18	36
支持的 NVIDIA 架构	NVIDIA Hopper™ 架构	NVIDIA Blackwell 架构	NVIDIA Rubin 平台

资料来源：英伟达官网，东兴证券研究所

图14：NVLink 交换机规格参数对比

NVLink　NVLink 交换机

	NVLink 4 交换机	NVLink 5 交换机	NVLink 6 交换机
NVLink GPU 域	8	8 | 72	8 | 72
NVLink 交换机 GPU 到 GPU 带宽	900 GB/s	1,800 GB/s	3,600 GB/s
总聚合带宽	7.2 TB/s	130 TB/s (NVL72)	260 TB/s (NVL72)
支持的 NVIDIA 架构	NVIDIA Hopper™ 架构	NVIDIA Blackwell 架构	NVIDIA Rubin 平台

资料来源：英伟达官网，东兴证券研究所

仅供内部参考，请勿外传

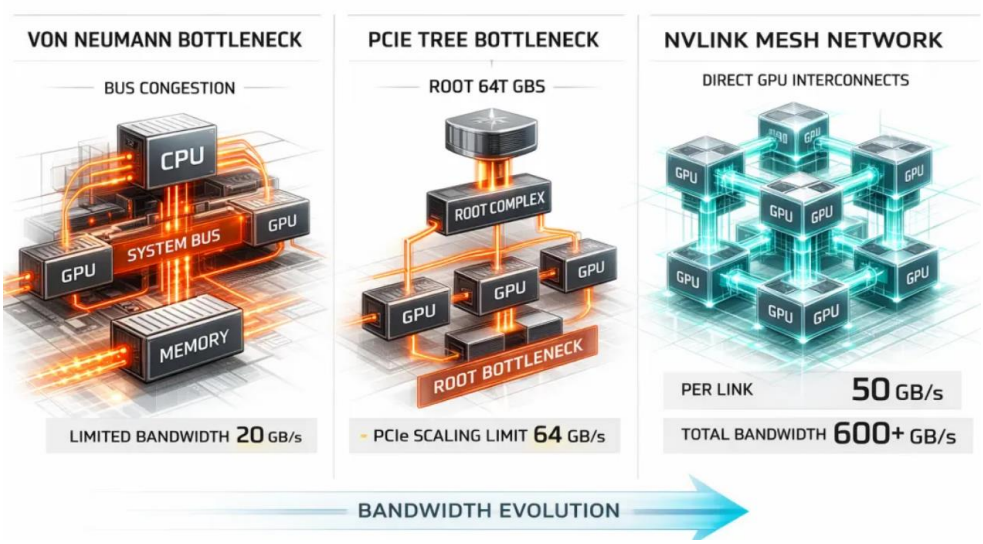
NVLink 核心优势在于创新设计通信架构，推出网状拓扑理念。为实现 AI 训练集群高带宽与低延迟数据传输，NVLink 允许 GPU 之间形成多对多的直接通信网络，每个 GPU 都可以同时与多个其他 GPU 建立高速通信链路。NVLink 协议创新如下：

在物理层面，NVLink 采用差分信号传输技术，具有高带宽和高抗干扰性能。每个链路由多对差分信号线组成，每对信号线负责传输一个方向的数据。SerDes 模块是 NVLink 物理层的核心组件，负责将并行数据转换为高速串行流，并在接收端进行反向转换。NVLink 的 SerDes 设计采用时钟数据恢复技术，以及集成复杂的自适应均衡电路。

在链路层，NVLink 定义多种类型的符号，包括数据符号、控制符号和填充符号，实现复杂的通信协议功能；设计精细的信用机制，实现不同优先级的流量调度。

除此之外，NVLink 其他创新之处包括多 Lane 绑定技术、统一内存空间等。

图15：NVLink 网状拓扑结构提供高速双向带宽



资料来源：仰望 7866 公众号，东兴证券研究所

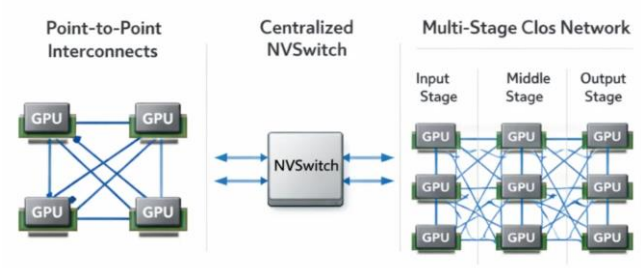
仅供内部参考，请勿外传

NVSwitch 是实现 Scale up 网络复杂交换的关键设备。

早期的 NVLink 实现主要采用点对点连接模式，GPU 之间通过直接的串行链路进行通信。当系统包含多个 GPU 时，点对点模式的连接复杂度呈平方级增长。

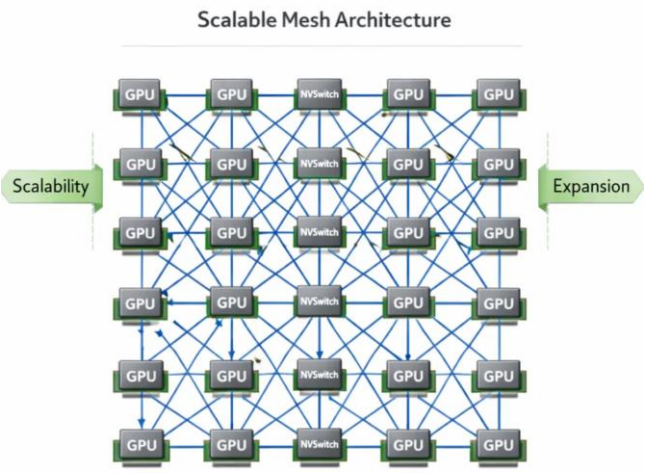
作为专门的交换芯片，NVSwitch 可以提供多端口的高速交换能力。NVLink 的交换网络采用多阶 Clos 网络架构，Clos 网络通过多级交换结构实现输入端口到输出端口的任意连接。

图16：NVLink 交换网络的演进过程（1）



资料来源：仰望 7866 公众号，东兴证券研究所

图17：NVLink 交换网络的演进过程（2）



资料来源：仰望 7866 公众号，东兴证券研究所

仅供内部参考，请勿外传

3.3 GB200 NVL72 超节点总交换容量 129.6TB/s

Blackwell 架构推动英伟达 Scale up 标准化。英伟达 GB200 NVL72 将 Scale-up 规模稳定在 72 个 GPU/机柜，形成可复制标准化方案。

单台 GB 200 NVL72 机柜有 18 个计算节点。GB200 NVL72 超节点主要由 18 个 Compute Tray（计算托盘）和 9 个 Switch Tray（网络交换托盘）构成。每个计算托盘容纳 4 颗 B200 GPU 和 2 颗 Grace CPU，构成两个 GB200 超级芯片。

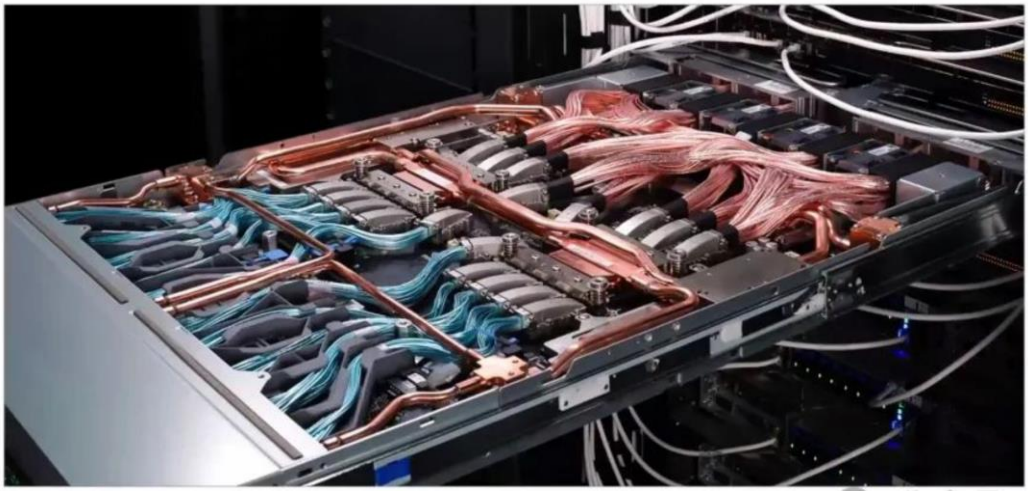
图18：GB200 NVL72 中计算托盘



资料来源：光芯公众号，东兴证券研究所

GB 200 NVL72 机柜有 9 个网络交换托盘。每个网络交换托盘中包含两颗 NVLINK Switch5 芯片，合计 18 颗 NVSwitch5 芯片。单颗 NVSwitch5 芯片交换容量为 7.2TB/s，总交换容量 129.6TB/s。

图19：GB 200 NVL72 中 NVLink 交换机托盘



资料来源：光芯公众号，东兴证券研究所

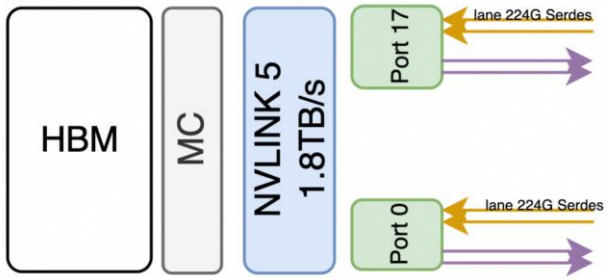
仅供内部参考，请勿外传

GB200 NVL72 实现 72 颗 B200 完全互联，总交换带宽 129.6TB/s。

计算节点访存带宽为 7.2TB/s：B200 设置 18 个端口（Port）。每个端口采用 224G Serdes，由四对差分线构成。每个端口的传输速率为 200Gbps * 4（4 对差分线）/8 =100GB/s（双向）。每个计算托盘容纳 4 颗 B200 GPU，则每个计算节点 72 个 NVLink5 Port，总访存带宽为 7.2TB/s。

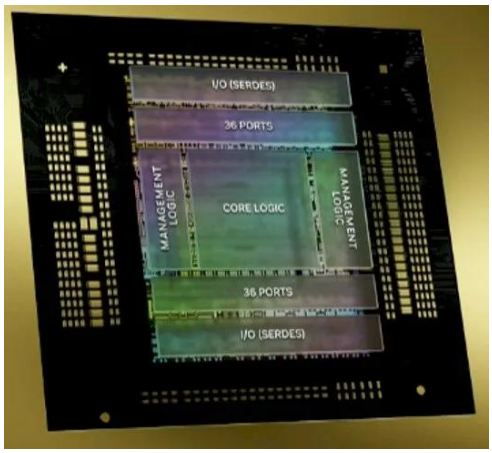
交换节点访存带宽为 14.4TB/s：NVSwitch5 芯片由 72 个 NVLINK Port（上下各 36 个 Port）。同样，每个 Port 采用双路 200Gbps 速率的 SerDes 高速串行接口，则每个 Port 带宽为 100GB/s。每个交换托盘两颗 NVLINK Switch5 芯片。每个交换节点 144 个 NVLINK Port，总访存带宽为 14.4TB/s。

图20：B200 端口 Port 示意图



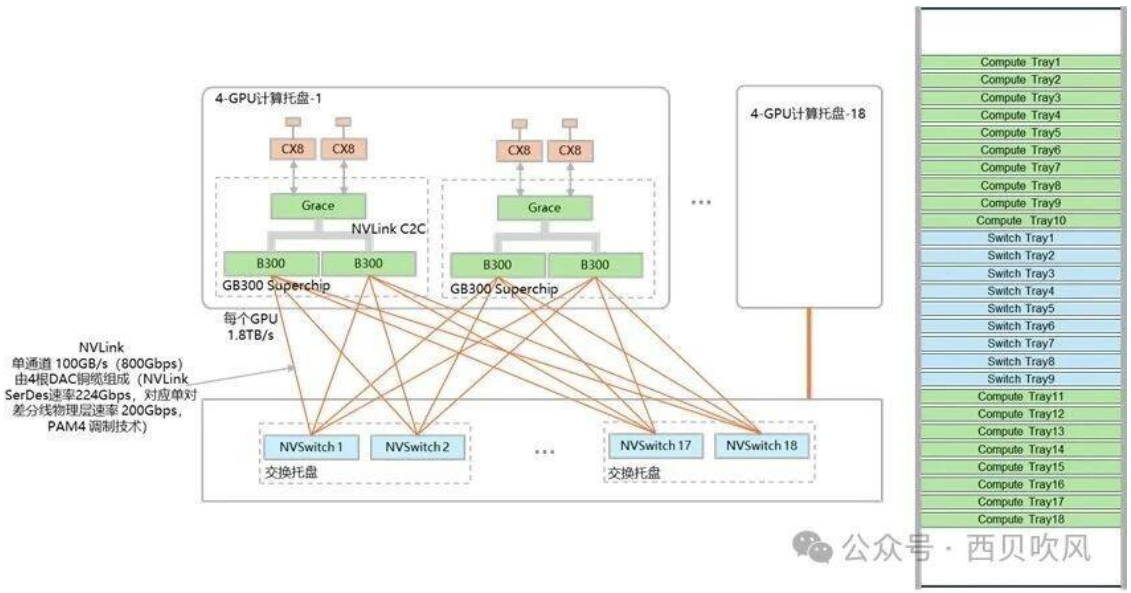
资料来源：zartbot 公众号，东兴证券研究所

图21：NVLINK Switch5 芯片 Port 示意图



资料来源：zartbot 公众号，东兴证券研究所

图22：GB200/300 NVL72 单层计算托架的互联拓扑



资料来源：西贝吹风公众号，东兴证券研究所

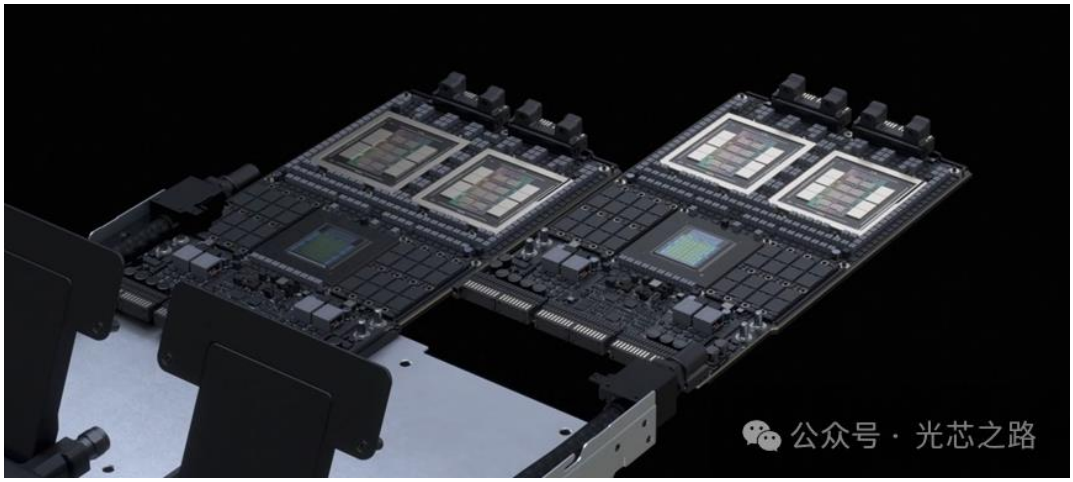
仅供内部参考，请勿外传

3.4 VR200 NVL72 超节点总交换容量翻倍至 259.2TB/s

Rubin 架构推动 Scale up 方案带宽倍增。2026 年 1 月 CES 展会，英伟达发布 Rubin 架构 VR200 NVL72。其中 NVLink 6 Switch 实现单 GPU 的互连带宽提升至 3.6 TB/s，上代为 1.8TB/s。

计算节点：Rubin NVL72 机架通过无缆互联架构整合 18 个计算托盘，每个托盘 2 颗超级芯片，每颗超级芯片集成 1 个 Vera CPU 与 2 块 Rubin GPU，共 72 GPU 与 36 CPU。

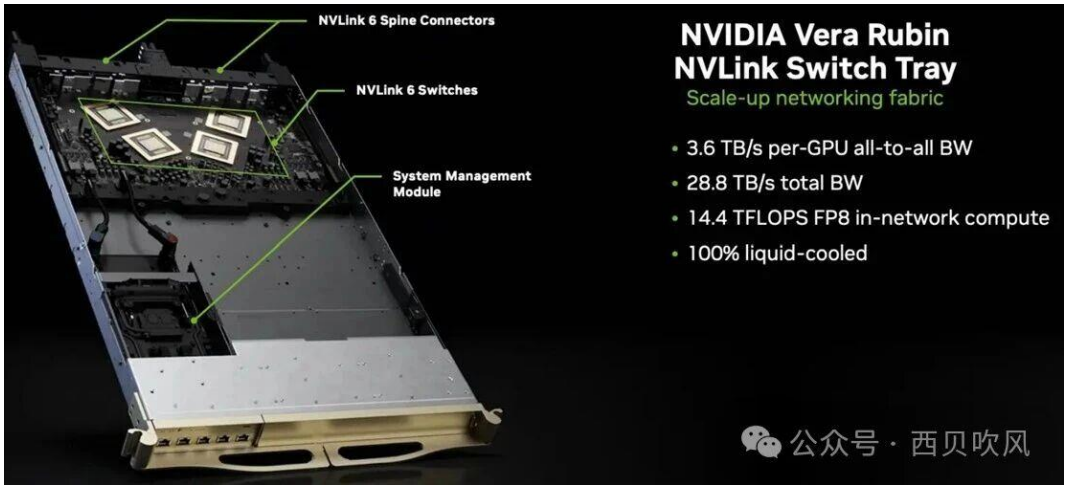
图23：VR200 NVL72 机柜计算托盘



资料来源：光芯之路公众号，东兴证券研究所

交换节点：VR200 NVL72 配置 9 个交换托盘，每个托盘集成 4 颗第六代 NVSwitch 芯片，全机柜部署 36 颗 NVSwitch。相比 GB200 NVL72，NVSwitch 芯片数量实现翻倍。单颗第六代 NVSwitch 交换容量为 7.2TB/s，相比 NVSwitch5 芯片，保持不变。

图24：Vera Rubin NVL72 机柜交换机托盘



资料来源：西北吹雪公众号，东兴证券研究所

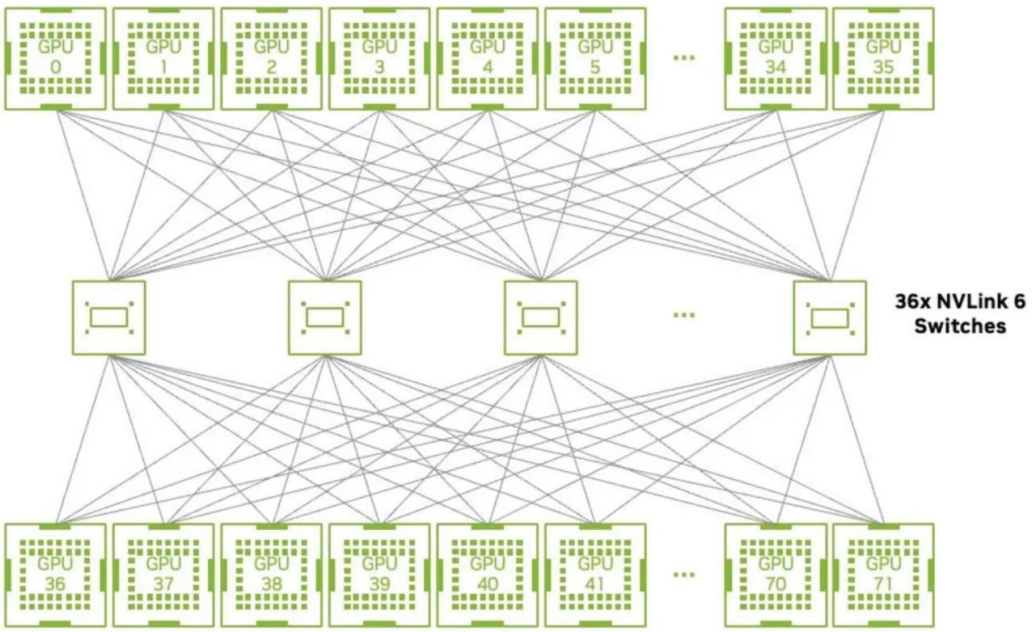
仅供内部参考，请勿外传

VR200 NVL72 Scale up 方案实现总交换容量 259.2TB/s，对比 GB200 NVL72，提升一倍。

计算节点：VR200 设置 72 个端口。每个端口带宽 100GB/s。每个计算托盘 2 颗 VR200 GPU，则每个计算节点 144 个 NVLink 6.0 端口，总访存带宽为 14.4TB//s。

交换节点：NVSwitch6 芯片 72 个 NVLink 6.0 端口。每个 NVLinkPort 交换容量 100GB/s。每个交换托盘 4 个 NVLink 6 Switch 芯片，每个交换节点 288 个 NVLink Port，总访存带宽为 28.8TB/s。

图25：VR200 NVL72 机柜中 GPU 互联拓扑结构



资料来源：英伟达官网，东兴证券研究所

仅供内部参考，请勿外传

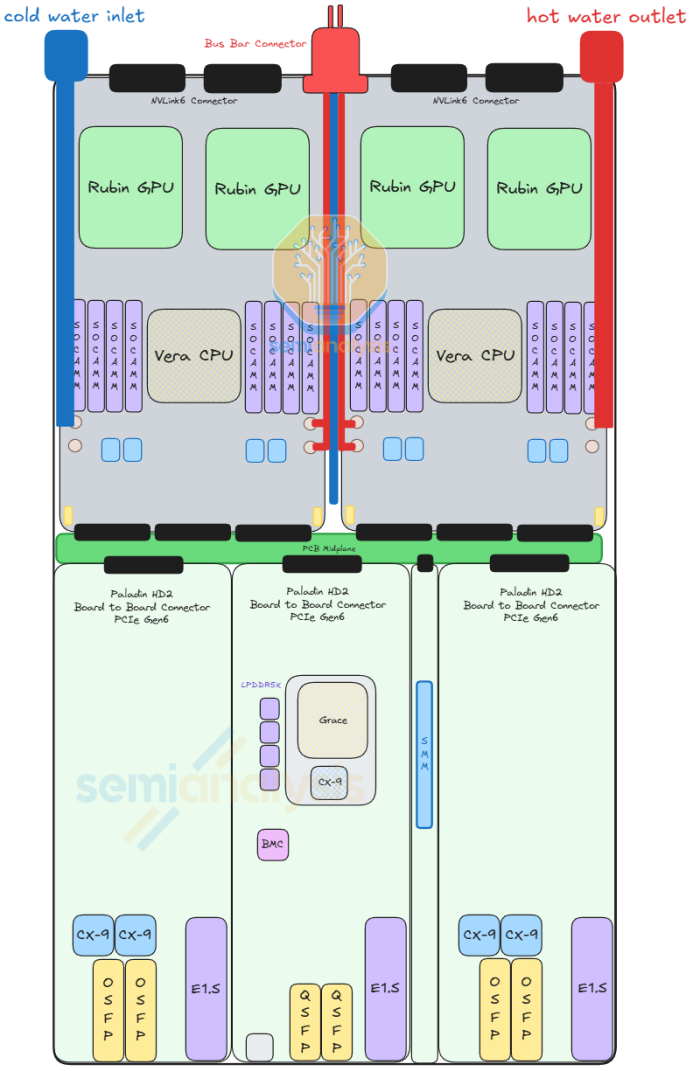
3.5 VR NVL72 计算托盘通信速率三重升级：1.8TB/s NVLink-C2C+ PCIe Gen6+800G 以太网

VR NVL72 计算托盘由六类模块拼合而成：后半部是 2 块 Strata 模块；前半部是 4 块 Orchid 模块；托盘中央垂直插着 1 块 PCB Midplane；前部中央还有 1 块 Blue
统管理模块。各模块之间通过板对板连接器相互连接。

Strata 模组容纳两个 Rubin GPU 与一个 Vera CPU，并引入 SOCAMM 插槽供 Vera CPU 的 LPDDR 内存使用，位于 Vera CPU 左右两侧的八个 SOCAMM 插槽支持 192GB 或 128GB 模组，使每个 Vera CPU 能实现 1024-1534GB 的弹性内存配置。

每个 Orchid 模组容纳两个 ConnectX-9 NIC、两个 800G 收发器笼与一个 E1.S 模组插槽供本地存储。

图26：VR NVL72 机柜计算托盘顶视图



资料来源：SemiAnalysis，东兴证券研究所

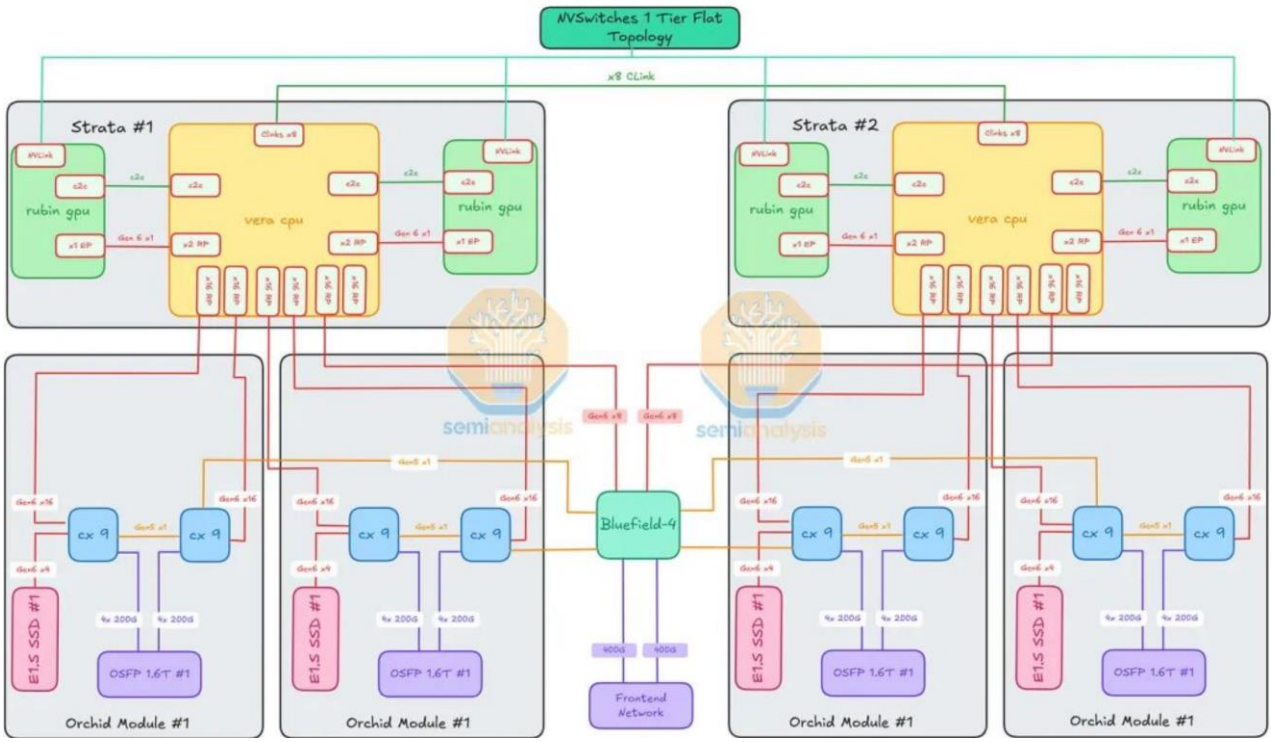
仅供内部参考，请勿外传

在 VR NVL72 中，AI 计算任务从外部网络进来，数据经过 ConnectX-9、Blue GPU 和机架内其他 71 颗 GPU 协同完成计算，最后计算结果通过网络传送出去。

在计算托盘中，数据传输路径可以分为三段：

- Vera CPU 至 Rubin GPU 之间通过 NVLink C2C 高速链路互联；
- Vera CPU 至 CX-9 之间通过两条 PCIe Gen6 链路分别连接到两个 CX9 的 PCIe Switch 模块；
- 以及 CX-9 至 OS

图27：VR NVL72 计算托盘拓扑结构

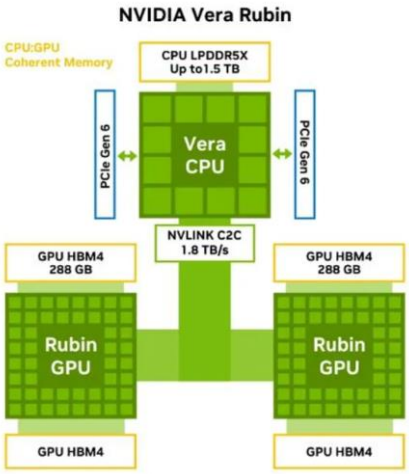


资料来源：SemiAnalysis，东兴证券研究所

仅供内部参考，请勿外传

NVLink-C2C 技术重构异构计算的互联范式，在裸片/芯片间互联领域建立巨大的领先优势。在 VR200 NVL72 中，Rubin-Vera 之间依托 NVLink-C2C（Chip-to-Chip，芯片到芯片互联）实现双向带宽 1.8TB/s CPU-GPU 互联，延迟纳秒级，相比 GB200 NVL72 的 NVLink-C2C 的 900GB/s，提升一倍。而主流 PCIe Gen5 架构双向带宽为 128GB/s 带宽，非一致性内存访问增加编程复杂性以及计算资源闲置等待。NVLink-C2C 的核心技术原理在于：通过 AMBACHI 协议实现硬件级缓存一致性，CPU 和 GPU 缓存自动同步；CPU 内存与 GPU 显存在软件视角呈现为单一内存池；对系统范围跨处理器的原子读写无需额外同步原语。

图28：VR NVL72 中 Rubin-Vera 芯片互联方式



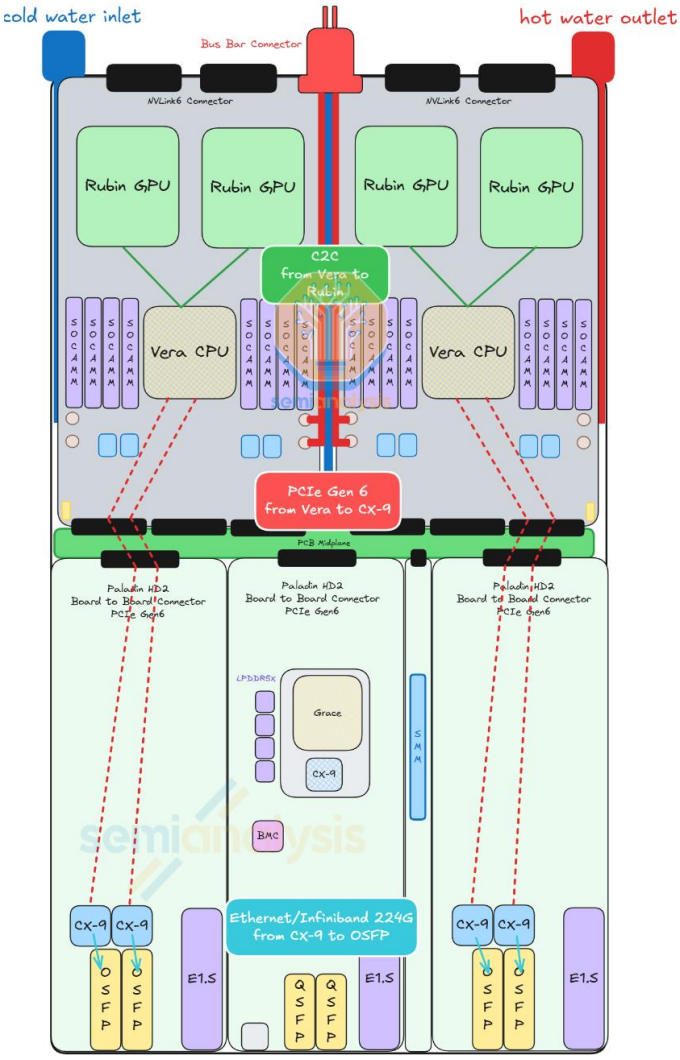
资料来源：英伟达官网，东兴证券研究所

仅供内部参考，请勿外传

采用 PCIe Gen6 协议实现 Vera CPU 与超级网卡 CX-9 互联。PCIe Gen6 是第六代高速外设互联标准，CPU 与网卡、存储等外设的通用接口。PCIe 6 接口支持 48 条 Lane，每条 Lane 单向速度 64 Gbps。因此，Vera 与 CX-9 之间接口双向总带宽达到 768GB/s。

PCIe Gen6 信号需要使用高端 PCB 与玻纤布传输。在 VR200 NVL72 计算托盘中，PCIe Gen6 信号从 Strata 模块传输到 Orchid 模块前端，PCB 距离长达约 500mm。为实现信号完整性，VR200 NVL72 除了升级双向 SerDes 技术外，还需要升级 PCB 材料。在材料层面，CCL（覆铜板）从 M7 升级到 M8/M9,，主计算板和网络板的铜箔升级到 HVLP4，材料价值显著上升；为了降低介质损耗，玻璃纤维布或价值更高的石英材料被用于 Orchid 板和中置板。

图29：VR NVL72 运算托盘信号路径图



资料来源：SemiAnalysis，东兴证券研究所

仅供内部参考，请勿外传

采用以太网/ InfiniBand 协议实现超级网卡 CX-9 与 OS

以太网模式下通过单个端口即可提供 1x800G 的传输能力，无需依赖多链路聚合实现总吞吐量。相比之下，CX-8 仅在 InfiniBand 架构下支持 800G 速率，但在以太网模式下通常以 2x400G 的配置呈现。

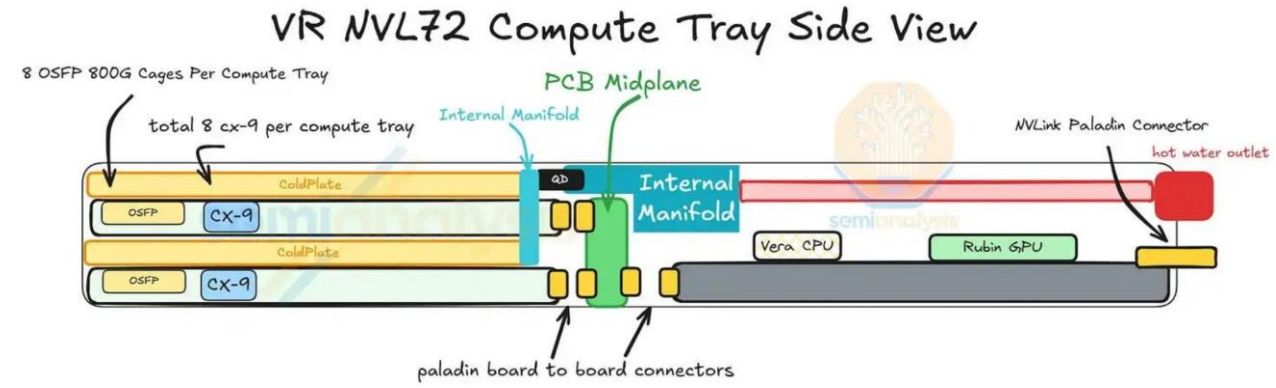
在 VR NVL72 计算托盘中，8 个 800G 的 CX-9 网卡对应 OS
配 1 个 1.6T OS
笼口，则每个计算托盘共 8 个 800G OS

800G 以太网的核心技术原理是通过 4×200G PAM4 串行链路实现。PAM4（Pulse Amplitude Modulation 4-level，四电平脉冲幅度调制）是一种在相同时间内传递更多比特的调制技术——普通信号只有高/低两种电压，代表 0 和 1；PAM4 使用四种电压等级（比如 0V/0.33V/0.67V/1V），分别代表 00/01/10/11，每次传输 2 个比特，实现单位时间信息密度翻倍。

InfiniBand（无限带宽）：起源于 1999 年，最初为 HPC 超算集群设计，特点：低延迟（约 1 微秒端到端）、高带宽、支持 RDMA。英伟达 2019 年以 69 亿美元收购 Mellanox 后获得 InfiniBand 全栈。Quantum-3 是当前最新的 InfiniBand 交换 ASIC。

Ethernet（以太网，IEEE 802.3 标准）：1973 年由 Xerox PARC 发明，全球最通用的有线网络标准。AI 数据中心以太网已从 100GbE 升级到 400GbE、800GbE，下一代 1.6TbE 正在标准化。英伟达 Spectrum-X 和博通 Tomahawk 系列是 AI 以太网主要竞争方案。

图30：VR NVL72 计算托盘侧视图



资料来源：SemiAnalysis，东兴证券研究所

仅供内部参考，请勿外传

3.6 ConnectX 超级网卡价值显著提升：内置 PCIe 交换模块与以太网交换逻辑

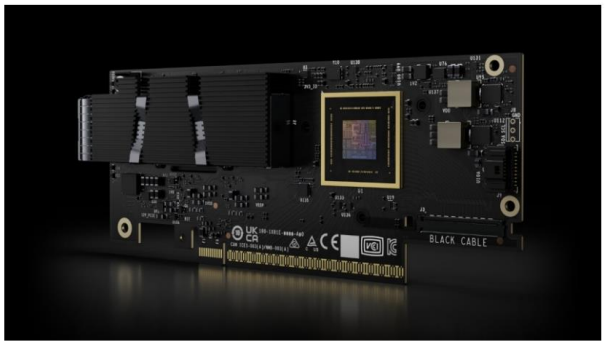
ConnectX-8/9 定位超级网卡（SuperNIC），性能远超传统网卡。2025 年 8 月，英伟达正式发布专为 Blackwell 架构和加速超大规模 AI 工作负载而设计的 ConnectX-8 SuperNIC。ConnectX-8 SuperNIC 单端口 800Gb/s InfiniBand（XDR）或双端口 400Gb/s Ethernet（Spectrum-X），为上一代 ConnectX-7（200Gb/s）的 4 倍，是当前业界最高带宽网卡。2026 年 1 月，英伟达推出高性能智能网络接口卡 ConnectX-9，核心变革在于实现单端口 800Gb/s 的以太网传输能力。

图31：ConnectX-9 SuperNIC 与 ConnectX-8 SuperNIC 产品



ConnectX-9

NVIDIA ConnectX-9 SuperNIC 借助突破性的网络技术、优化的网络连接和加速的性能为每个 GPU 提供高达 1.6 Tb/s 的吞吐量，为十亿级 AI 工厂赋能。



ConnectX-8

ConnectX-8 InfiniBand SuperNIC 可提供高达 800 Gb/s 的数据吞吐量，并支持 NVIDIA 网络计算加速引擎，可提供支持万亿参数级 AI 工厂和科学计算工作负载所需的性能和各种强大功能。

资料来源：英伟达，东兴证券研究所

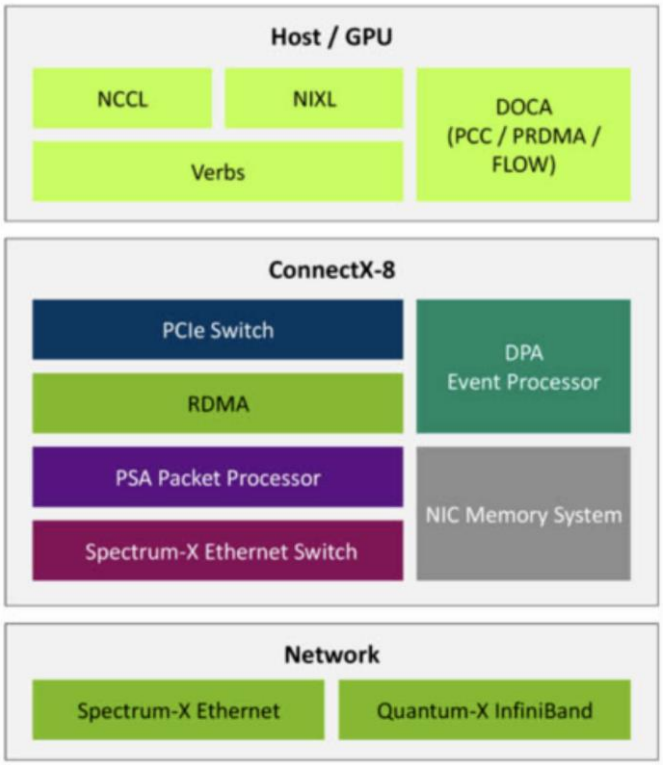
仅供内部参考，请勿外传

超级网卡内置 PCIe Gen6 交换模块，替代传统独立 PCIe 交换机。ConnectX-8 内置 48 通道 PCIe Gen6 交换机，单芯片实现“网络接口 + GPU 间交换”二合一，有助于消除 IO 瓶颈，并加快 GPU、NIC 和存储之间的数据移动速度。基于 ConnectX-8 的优化设计可为集群内的所有 GPU 间通信提供高达每个 GPU 50 GB/s 的 IO 带宽，因为 NCCL 直接通过网络转发所有流量。

ConnectX-8 技术栈分为三层。

- 上层：Host/GPU 侧的 AI 通信专用 API。NCLL 是 GPU 间集合通信的核心库，是大模型训练中多卡同步的基础。NIXL 支持网络内集合通信，进一步降低主机侧的通信开销。Verbs 是 RDMA 的通用编程接口，是高性能通信的底层标准。DOCA 是 NVIDIA 的芯片级数据中心编程框架，包含 PCC（拥塞控制）、PRDMA（可编程 RDMA）、
- 中层：ConnectX-8 硬件架构。PCIe Switch 是内置 PCIe 交换模块，优化主机/GPU 与网卡之间的数据传输路径；RDMA 是硬件级 RDMA 引擎，实现低延迟、高吞吐的直接内存访问，是 AI 通信的核心；PSA Packet Processor 是可编程数据包处理器，支持灵活的数据包处理逻辑，适配不同的网络协议和工作负载；Spectrum-X Ethernet Switch 是内置以太网交换逻辑，与 NVIDIA Spectrum-X 交换机深度协同，实现端到端的网络优化；DPA Event Processor 是专用事件处理器，处理网络事件和任务调度，减轻主机 CPU 负担；NIC Memory System 是网卡本地存储系统，用于缓存数据、优化流控和拥塞处理。
- 下层：网络层协议支持。ConnectX-8 同时支持两种主流高性能网络协议：Spectrum-X Ethernet 基于以太网的高性能网络，兼容标准以太网架构，适合大规模数据中心部署；Quantum-X InfiniBand：InfiniBand 协议的高性能网络，专为超算和 AI 集群设计，提供极致的低延迟和高吞吐。

图32：ConnectX-8 网卡技术栈



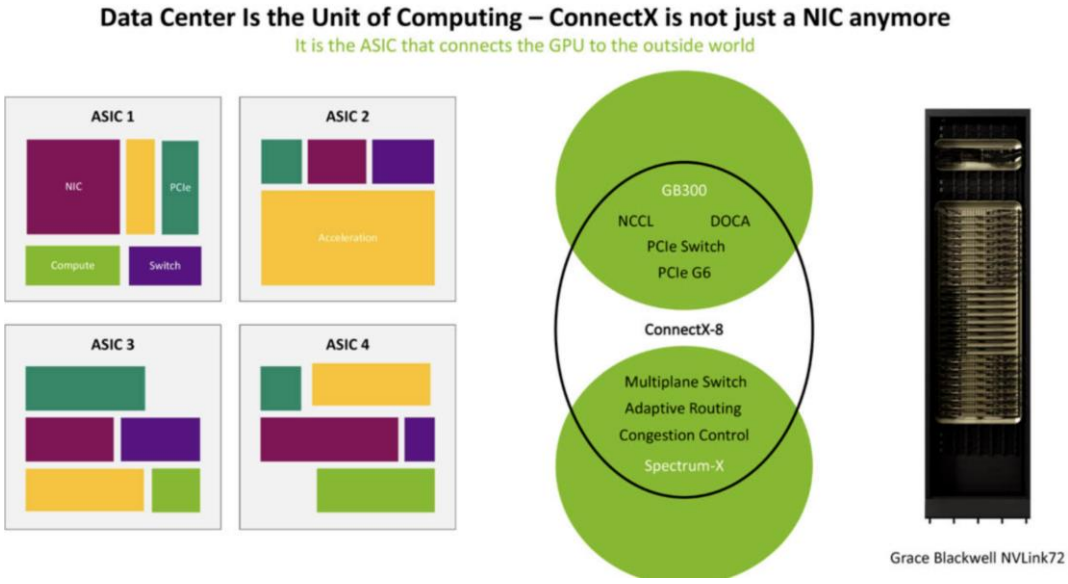
资料来源：英伟达，东兴证券研究所

仅供内部参考，请勿外传

ConnectX 集成 Spectrum-X 交换逻辑，构成端到端 800G AI 网络。SuperNIC 内部集成 Spectrum-X 风格的交换与加速逻辑”并作为 Spectrum-X 以太网平台的终端侧关键组件，与外部 Spectrum-X 交换机（如 SN5600 列）端到端协同。

- ConnectX-8 与 GB300 平台的 GPU / 主机架构深度协同，关键能力包括 NCCL、DOCA、PCIe Switch、PCIe Gen6；
- 与 Spectrum-X 交换机形成端到端的高性能网络，关键能力包括：Multiplane Switch（多平面交换技术）、Adaptive Routing（自适应路由）、Congestion Control（硬件级拥塞控制）、Spectrum-X（端到端的 AI 通信优化）。

图33：超级网卡是连接 GPU 集群与外部的关键 ASIC



资料来源：英伟达，东兴证券研究所

仅供内部参考，请勿外传

交换机（Spectrum-X Switch）与终端 SuperNIC（ConnectX-8）协同优化，为 AI / 超算以太网带来的五大关键性能提升。负载均衡方面，实现 1.6X 更高有效带宽；尾延迟优化方面，实现 1.3X 更高集合通信带宽；噪声隔离方面，实现 2.2X 更高 All-reduce 带宽；弹性性能方面，实现 1.3X 更高 All-to-all 带宽；高频遥测方面，实现 1000X 更快遥测采集。

图34：Spectrum-X 交换机到超级网卡 ConnectX 的端到端网络处理，带来高性能以太网表现



资料来源：英伟达，东兴证券研究所

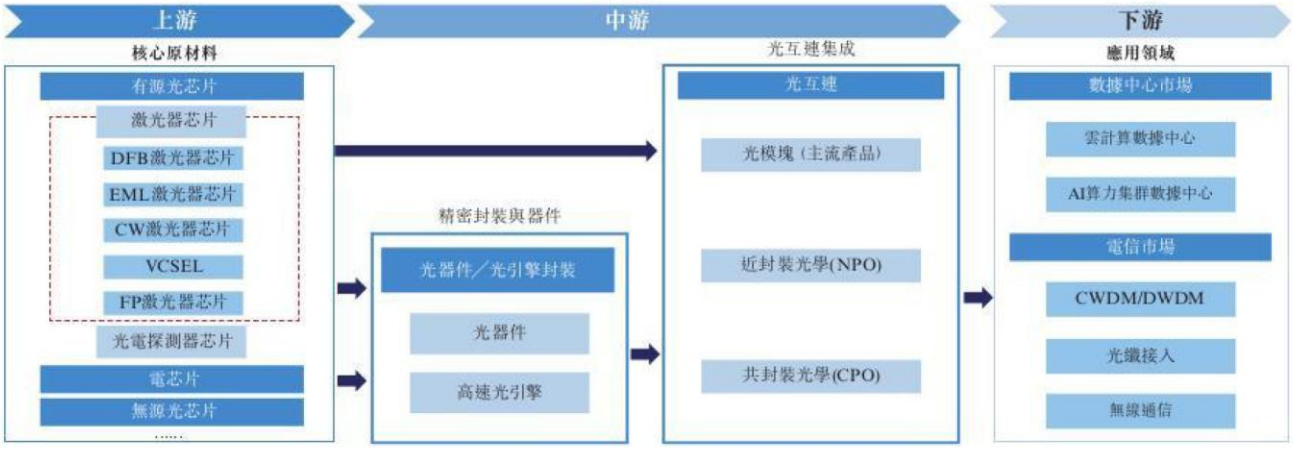
仅供内部参考，请勿外传

4. 光芯片：CW 激光器重构全球光芯片产业格局，IDM 模式构建产业链壁垒

4.1 技术路径分化趋势清晰，CW 激光器成为硅光架构重要配置

数据中心光互联技术路径分化趋势清晰。一种是早期 EML 设计，依靠高性能 EML 激光器芯片实现高速率数据传输（如用于 400G 和 800G 光模块的 EML 激光器芯片）；第二是集成度更高的硅光架构，该架构需要 CW 激光器芯片作为外部激光源，支持其高集成度。我们认为，数据中心下游 scale up 网络的发展，成为 CPO/NPO 等新兴光互联技术重要的应用场景；而 CPO/NPO 的规模化应用将驱动硅光架构有望成为光互联行业主流技术路径。

图35：数据中心光互联产业链

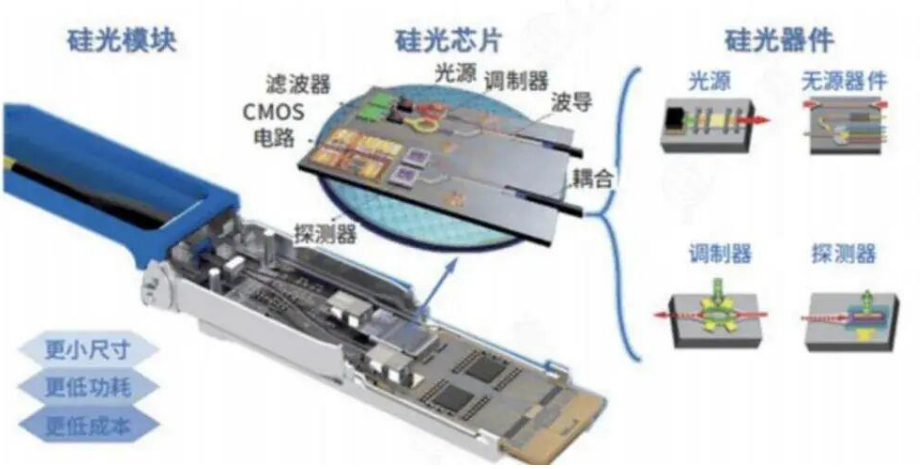


资料来源：灼识咨询，东兴证券研究所

仅供内部参考，请勿外传

硅光架构的技术路径下，激光器芯片占据重要价值链地位。以非硅光光模块为例，其硬件成本结构主要包括光芯片、电芯片、无源光器件、PCB 及机构件四大板块。对于硅光光互连产品，BOM 结构重构，原有的分立调制器与大量无源光器件被集成为一颗硅光芯片（PIC），PCB 与机构件也被大幅简化；由于硅材料本身发光效率低，难以直接实现高效光发射。当前外置 CW（连续波）光源成为硅光光模块的主流方案，且进一步应用于 CPO/NPO 等场景。因此，硅光光模块 BOM 聚焦于‘硅光芯片’与‘激光器芯片’两大核心。其中激光器芯片性能直接决定了下游光器件、光模块乃至整个光通信系统的传输速率与能效表现。

图36：硅光模块内部结构



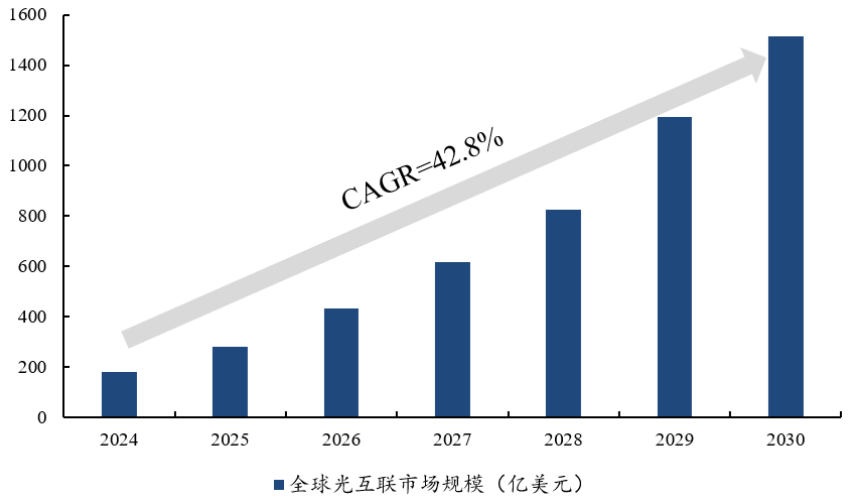
资料来源：Intel，东兴证券研究所

仅供内部参考，请勿外传

4.2 激光器芯片市场增速陡峭，2030 年市场或将超过 200 亿美元

Scale up 网络驱动光互连行业高景气发展。Scale up 网络要求实现超高带宽、超低延迟。光互联成为支撑万亿级大模型与高实时性应用的关键解决方案。在此技术趋势下，全球光互连市场规模有望从 2024 年 179 亿美元，攀升至 2030 年的 1514 亿美元，年复合增长率高达 42.8%。

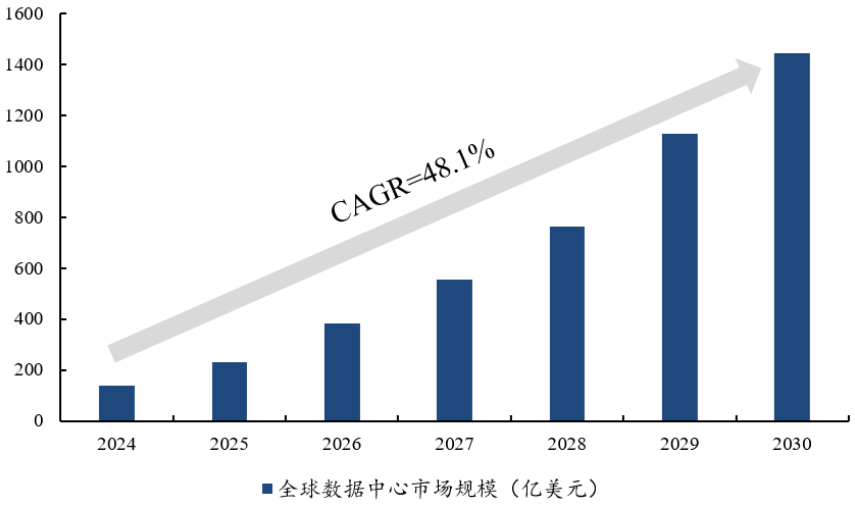
图37：全球光互联市场规模预测（亿美元）



资料来源：LightCounting，灼识咨询，东兴证券研究所

其中，2024 年全球数据中心光互连市场规模约 137 亿美元，预计 2030 年将增长至 1444 亿美元，对应年复合增长率 48.1%。

图38：全球数据中心光互联市场规模预测（亿美元）



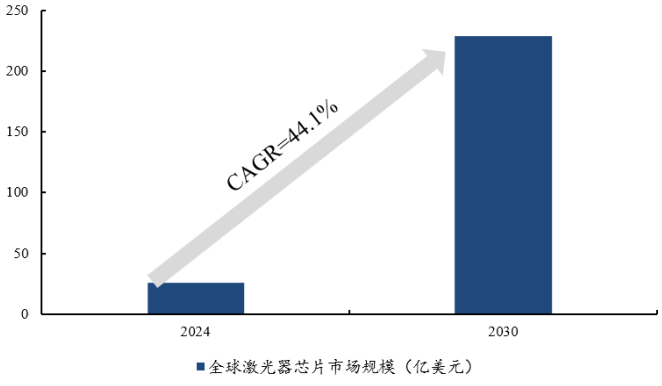
资料来源：LightCounting，灼识咨询，东兴证券研究所

仅供内部参考，请勿外传

受益于光互联高速发展，2030 年激光器芯片市场或将超过 200 亿美元。全球激光器芯片市场规模或将由 2024 年的 26 亿美元增长至 2030 年 229 亿美元，年复合增长率为 44.1%。其中数据中心领域贡献主要增量，2024 年市场规模 16 亿美元，预计 2030 年达到 211 亿美元，年复合增长率高达 53.4%。

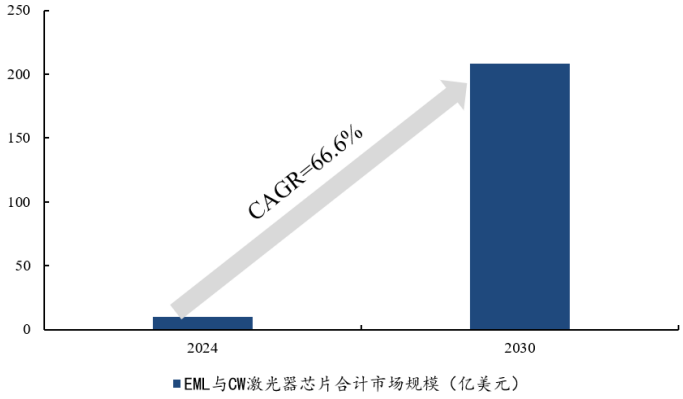
其中，EML 与 CW 激光器芯片占据主要市场份额。EML 激光器芯片作为早期开发的解决方案，在 400G 及以上光互连产品中广泛应用。而近年来，具备高集成、低成本优势的硅光解决方案成为演进方向，需配套大功率 CW 激光器芯片。2024 年两类高端芯片合计市场规模达 9.7 亿美元，市场占比约 38.1%。随着 800G、1.6T 及以上高速方案快速渗透，到 2030 年预计合计收入将达 208.0 亿美元，年复合增长率达 66.6%，市场占比将达 90.9%。

图39：全球激光器芯片市场规模预测（亿美元）



资料来源：LightCounting，灼识咨询，东兴证券研究所

图40：EML 与 CW 激光器芯片合计市场规模预测（亿美元）

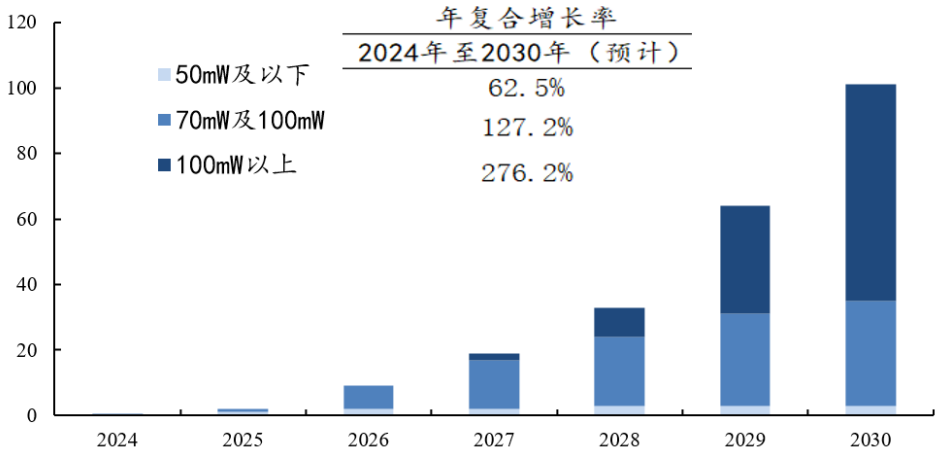


资料来源：LightCounting，灼识咨询，东兴证券研究所

仅供内部参考，请勿外传

NPO 及 CPO 驱动激光器芯片向大功率 CW 激光器芯片方向升级。在当前 400G、800G 乃至 1.6T 的主流硅光高速光互连产品中，所采用的主流 CW 激光器芯片涵盖 50mW、70mW、100mW 等成熟功率型号。此外，在 NPO 及 CPO 等下一代光电集成新技术推动下，150mW、300mW、400mW 更高功率规格 CW 光源，正逐步落地新一代光互连产品商业化研发进程。从细分增速结构来看：50mW 及以下产品 2024-2030 年 CAGR 为 62.5%，70mW 及 100mW 主力产品 CAGR 高达 127.2%，100mW 以上大功率高端 CW 芯片 CAGR 更是达到 276.2%。

图41：不同功率 CW 激光器芯片市场规模预测（亿美元）



资料来源：LightCounting，灼识咨询，东兴证券研究所

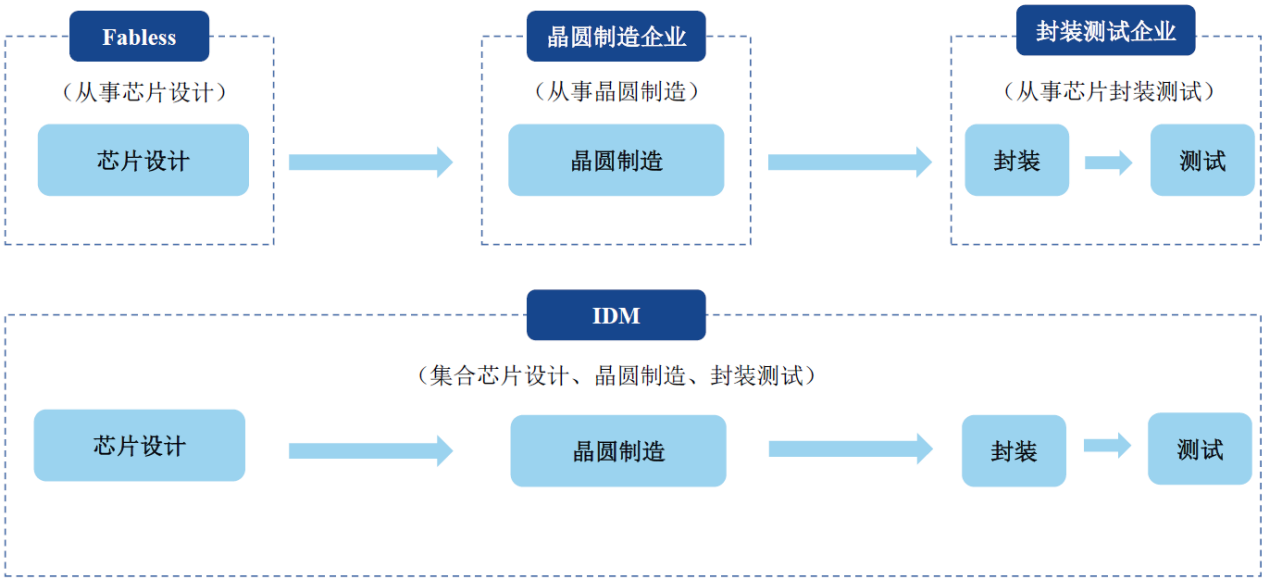
仅供内部参考，请勿外传

4.3 激光芯片制造价值链：外延生长与光栅工艺具有技术壁垒及产能瓶颈

现阶段，激光器芯片企业主要采用 IDM 垂直整合制造模式，价值链覆盖研发、晶圆制造、芯片加工、封装及测试等环节。根据是否从事晶圆制造、封装测试等生产环节，半导体行业的经营模式主要分为 IDM（Integrated Device Manufacture，垂直整合制造）模式与

稳定的产品质量、可靠的供应以及可扩展的制造产能，有助于快速研发新产品并加速产品的迭代，但另一方面，前期资本开支与技术积累要求较高。

图42：激光器芯片行业主要采用 IDM 模式



资料来源：长光华芯招股书，东兴证券研究所

仅供内部参考，请勿外传

IDM 模式下，高端激光器芯片附加值较高，毛利率超过 50%。2023-2025 年，源杰科技光芯片业务逐步向数据中心高端产品升级，产品毛利率稳中有升，2025 年达到 58%，产品附加值可以对标美国马科姆公司。此外 IDM 模式下，晶圆制造环节附加值稳定。2023-2025 年，台湾全新光电公司，主要产品为以 MOCVD 外延成长法为核心技术的晶圆片，毛利率维持在 36%-41%。

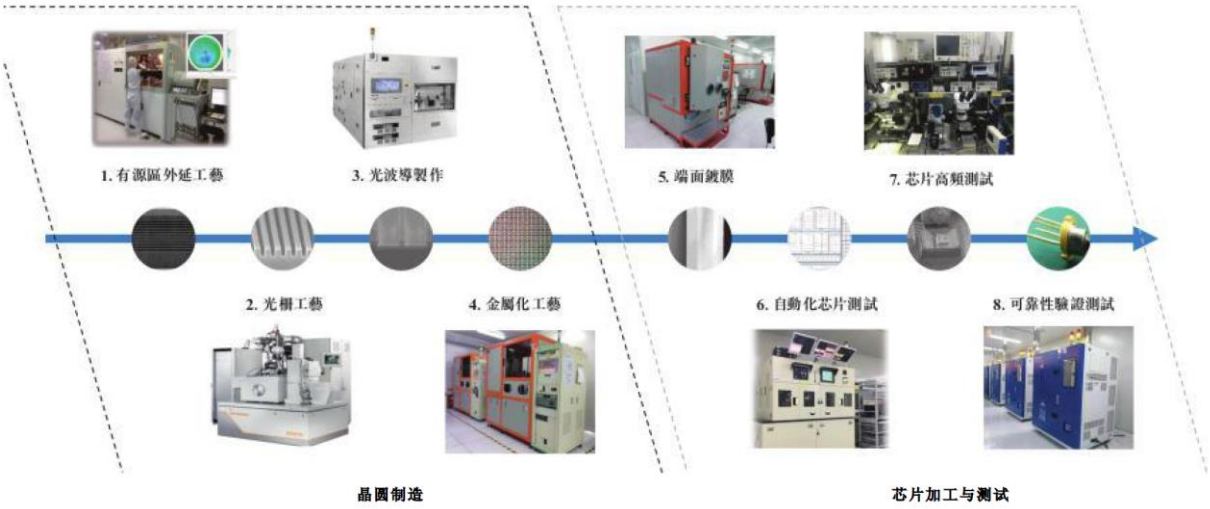
表3：激光器芯片价值链分布情况

公司名称	主营业务	2023 年度	2024 年度	2025 年度
马科姆（MACOM）	设计并制造应用于电信、工业和国防及数据中心领域的半导体产品	60%	54%	55%
全新光电	以 MOCVD 外延成长法为核心技术的 III-V 族化合物半导体	41%	39%	36%
源杰科技	光芯片的研发、设计、生产与销售	42%	33%	58%
仕佳光子	覆盖光芯片及器件、室内光缆、线缆材料等	19%	27%	35%
长光华芯	半导体激光芯片的研发、设计及制造	34%	24%	36%

资料来源：同花顺，源杰科技公告，东兴证券研究所

进一步对激光器芯片价值链进行划分，可以分为两个核心价值点：晶圆制造、芯片加工及测试。其中晶圆制造环节主要技术门槛有：1）有源区外延工艺；2）光栅工艺；3）光波导制作；4）金属化工艺。芯片加工及测试环节主要技术门槛有：5）端面镀膜；6）自动化芯片测试；7）芯片高频测试；8）可靠性验证测试。

图43：CW 激光器芯片关键制造步骤及相关设备



资料来源：源杰科技招股书，东兴证券研究所

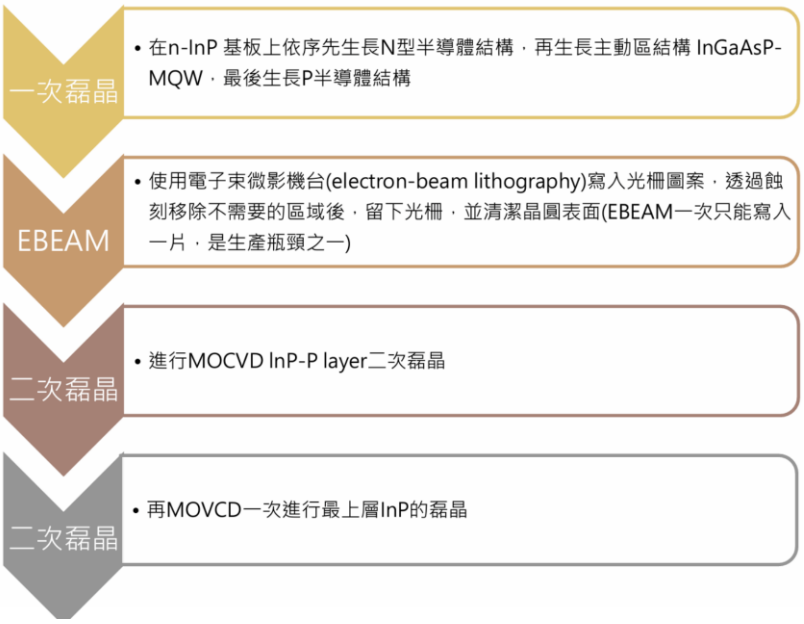
仅供内部参考，请勿外传

晶圆制造环节，有源区外延工艺与光栅工艺不仅技术门槛高，而且具有制造产能扩展瓶颈的特点。

外延生长是激光器芯片设计及制造的核心工艺。有源区外延生长是指通过 MOCVD 技术，生长可精确控制厚度及组分的多层量子阱外延晶片。MOCVD 外延生长技术包括外延工艺、MOCVD 外延设备改进工艺，如针对温度场、气场分布与 III/V 比等进行调整，建立高铟组分应变量子阱外延生长动力学模型，得到高质量的外延晶体材料。

在生产制造中，每台 MOCVD 机器的每批产品都必须对每个客户和应用进行单独认证，这在扩大生产规模时形成显著的扩产瓶颈。

图44：CW 激光器芯片外延生长步骤



资料来源：逍遥设计自动化公众号，东兴证券研究所

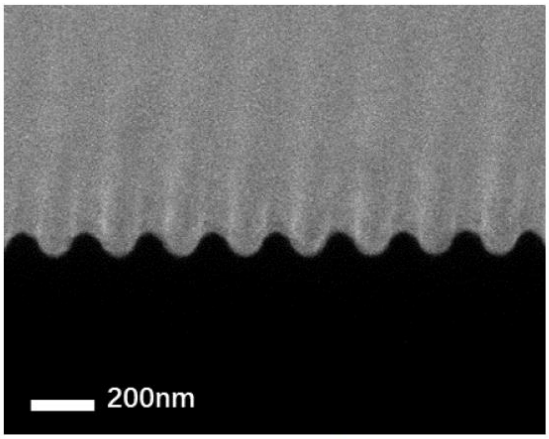
仅供内部参考，请勿外传

光栅工艺同样是激光器芯片晶圆制造重要环节。光栅工艺主要在涂有光刻胶的基板上定义出光栅结构对应的掩膜图形，再利用刻蚀技术将掩膜上的图形转移至衬底上形成最终的光栅结构。

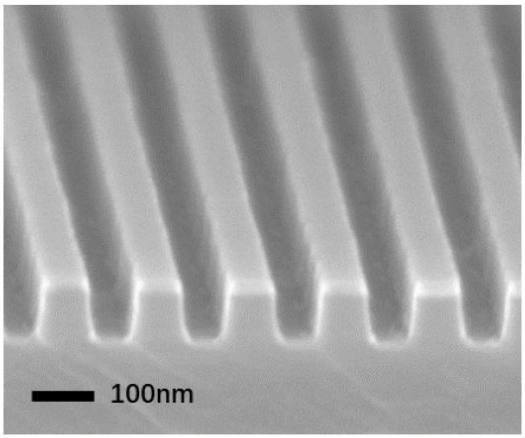
光栅结构影响激光器芯片产品指标包括出光功率、单模良率、芯片波长、极限工作温度特性（商业级激光器芯片工作温度范围为0～70℃，而工业级激光器芯片工作温度范围为-40～85℃）、芯片模态稳定性（影响光信号在传输网络中的传输距离和出错概率）、高频特性等。

光栅工艺主要分为两种，一种是全息光栅工艺（Holographic Grating），即利用两束激光的干涉条纹定义周期性掩膜图形，全息光栅工艺在 2.5G 激光器芯片生产中广泛使用；另外一种是电子束光栅工艺（Electron-Beam Technology），即利用电磁场控制电子形成电子束，利用电子束定义掩膜图形，该工艺技术较全息光栅工艺更为先进，能大幅提高光栅的控制精度，且实现非等周期光栅结构，国内掌握的厂家较少。

图45：两种光栅工艺制造的光栅结构对比



全息光栅工艺制造的光栅



电子束光栅工艺制造的光栅

资料来源：源杰科技公告，东兴证券研究所

仅供内部参考，请勿外传

相比于全息光栅，电子束光栅的工艺更为复杂，产品的光功率、单模良率、芯片波长、极限工作温度、高频特性等指标更好，有利于提高芯片产品的质量和可靠性。

图46：两种光栅工艺对激光器芯片产品特性影响

项目		全息光栅工艺	电子束光栅工艺
工艺复杂度		适中	复杂
产品特性	光功率	功率离散	功率一致性好
	单模良率	30%-50%	50%-90%
	芯片波长	<±3nm	<±2nm
	极限工作温度	<30%	>50%
	高频特性	差	好

资料来源：源杰科技公告，东兴证券研究所

电子束写入光栅图案一次只能写入同样在产能扩展瓶颈。图样掩膜定义步骤主要利用电子束光栅设备，将光栅图样编程步骤所涉及的光栅图样直写于晶圆上方的光刻胶掩膜上。光栅结构制作具体步骤包括：1）光栅图样编程；2）图样掩膜定义；3）掩膜转移刻蚀；4）光栅形貌观测与判定；5）掩膜清除；6）二次外延披覆光栅层。

图47：光栅结构制作流程具体步骤

环节	具体流程/步骤	
晶圆制造	（A）晶圆外延结构生长	
	（B）光栅结构制作	（B1）光栅图样编程
		（B2）图样掩膜定义（发行人委托加工内容）
		（B3）掩膜转移刻蚀
		（B4）光栅形貌观测与判定
		（B5）掩膜清除
		（B6）二次外延披覆光栅层
	（C）波导光刻工艺	
	（D）金属化制程	
	（E）减薄退火工艺	
芯片制造	（F）解理镀膜工艺	
	（G）封测分选	
	（H）可靠性验证	

资料来源：源杰科技公告，东兴证券研究所

仅供内部参考，请勿外传

光栅环节是激光器芯片 IDM 模式核心竞争力重要体现。电子束光栅工艺使用的专用设备价值较高，另一方面设备的操作和调试需要较高的经验和技术储备。2019-2021 年，根据源杰科技公告，其激光芯片晶圆光栅环节委外加工平均价格 2000-5300 元/片，而自产平均成本只需要 700 元/片左右。

图48：光栅环节委外加工与自产平均成本对比（元/片）

项目	2021 年	2020 年	2019 年
委外加工平均单价	2,097.69	3,985.56	5,317.85
自产平均成本	778.27	747.85	694.95
差异率	62.90%	81.24%	86.93%

注：差异率=1-自产平均成本/委外加工平均单价。

资料来源：源杰科技公告，东兴证券研究所

仅供内部参考，请勿外传

4.4 国内外企业同步研发高功率 CW 激光器芯片，不存在代际差距

中国光芯片产业进入加速发展阶段。全球激光器芯片呈现高增长蓝海市场态势，市场规模将由 2024 年的 26 亿美元增长至 2030 年 229 亿美元，年复合增长率为 44.1%。现阶段激光器芯片市场头部企业主要为美日企业，包括 Lumentum、Coherent、住友电工、三菱电机、博通五家公司。相比海外头部企业，源杰科技等中国头部企业，收入规模仍有较大差距。但受益于中国政府及头部企业积极投资光芯片领域以及全球市场的高增长红利，中国光芯片产业实现加速发展。

国家层面政策持续加码。2025 年 9 月，工业和信息化部、市场监督管理总局联合印发《电子信息制造业 2025-2026 年稳增长行动方案》，明确提出“面向光子领域重点环节开展技术攻关，加大对高速光芯片、光电共封等领域的研发投入力度，推动光架构与现有电架构体系生态融合”。

地方产业基金精准布局。2025 年 10 月，首期规模 50 亿元的深圳市半导体与集成电路产业投资基金正式揭牌，该基金重点投向光电子及传感器等芯片，以及关键制造设备、零部件及材料与先进封测等核心领域和薄弱环节，投资阶段以初创期、成长期为主。苏州则设立总规模超百亿元的光子及集成电路产业母基金，汇聚上下游企业超 350 家，形成全国领先的光子产业创新集群。

龙头企业密集扩产投资。源杰科技 2026 年 2 月宣布拟投资 12.51 亿元建设光电通讯半导体芯片和器件研发生产基地二期项目，同时将 50G 光芯片产业化建设项目投资额由 4.87 亿元调增至 7.57 亿元，以应对下游需求的快速增长。仕佳光子拟斥资 12.65 亿元建设高速光芯片与器件开发及产业化项目，投资额接近 2025 年全年营收的 60%，彰显公司在 AI 算力高景气周期对高端光芯片赛道下重注的决心。

表4：2025 年全球激光器芯片市场头部企业市场份额

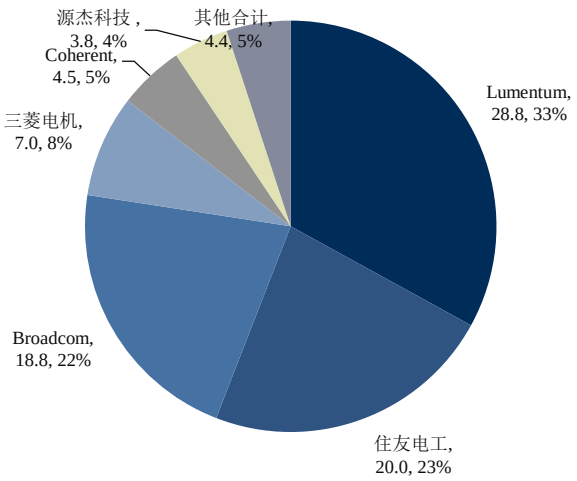
排名	公司名称	国家	销售收入（亿元人民币）	市占率
1	Lumentum	美国	32.4	16.7%
2	Broadcom	美国	28.1	14.5%
3	三菱电机	日本	25.9	13.4%
4	住友电工	日本	25.7	13.3%
5	Coherent	美国	7.2	3.7%
6	源杰科技	中国	6.0	3.1%
前六大公司合计			125.3	64.8%

资料来源：公司年报，灼识咨询，东兴证券研究所

仅供内部参考，请勿外传

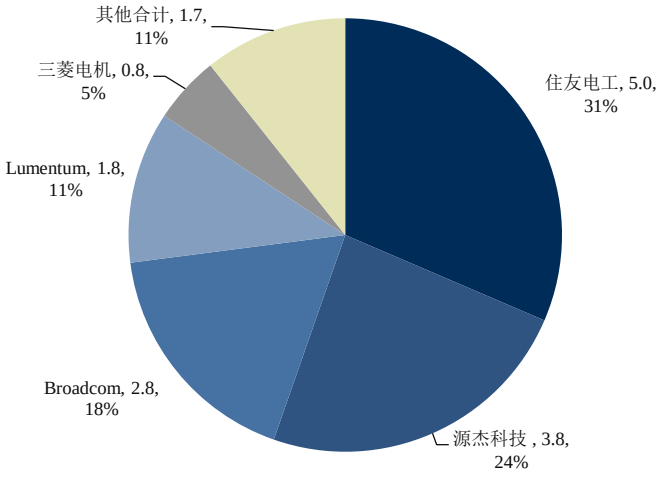
国内企业有望在硅光高速率光互连产品 CW 激光器实现追赶。2025 年全球用于 400G 及以上数据速率的高速率光互连产品激光器芯片市场规模约 87.3 亿元，包含 100G 及以上的 EML 激光器芯片，以及输出功率为 50mW 或以上的 CW 激光器芯片。其中应用于 400G 及以上硅光高速率光互连产品的 CW 激光器芯片市场 15.9 亿元，国外头部企业住友电工相关产品收入规模 5 亿元，国内头部企业源杰科技相关产品收入规模 3.8 亿元，双方差距不大。

图49：2025 年全球用于 400G 及以上光互联产品的激光器芯片销售收入排名（亿元，%）



资料来源：公司年报，灼识咨询，东兴证券研究所

图50：2025 年全球用于 400G 及以上硅光光互联产品的激光器芯片销售收入排名（亿元，%）



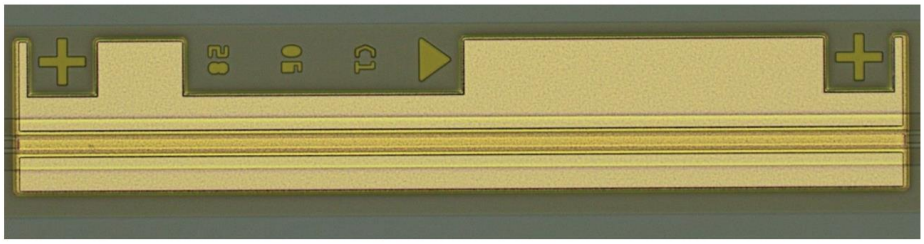
资料来源：公司年报，灼识咨询，东兴证券研究所

仅供内部参考，请勿外传

激光器芯片国产替代可期，国内外光芯片厂商处于同一起跑线。为满足 NPO 及 CPO 等新型封装技术，国内外企业同步研发 150mW、300mW 及 400mW 型号在内的高功率 CW 激光器芯片，双方产品不存在代际差距。

2025 年海外厂商 Coherent 推出 400 mW CW 激光器。从产品性能及指标看，该器件工作波长为 1311 nm，在 55℃条件下输出功率超过 400 mW，光谱线宽低于 200 kHz，相对强度噪声（RIN）低于-145 dB/Hz，并采用载体贴装封装形式封装，对应共封装光学与硅光应用场景。光波导结构设计采用 BH D
分布反馈）激光器平台，BH D
合连续波、高功率和低噪声要求较高的场景。从商业化节奏看，Coherent 表示该产品在 2025 年进入工程样品阶段，并预计于 2026 年第三季度进入批量生产和全面供货。

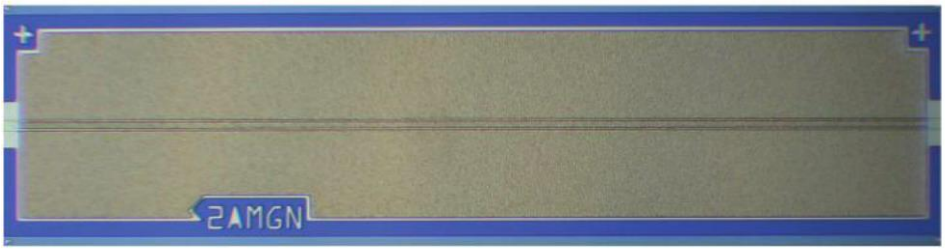
图51：Coherent 公司推出 400 mW CW InP 激光器芯片



资料来源：公司官网，东兴证券研究所

2025 年国内头部企业源杰科技同样前瞻研发 300mW 等高功率的 CW 光源，产品性能处于良率优化阶段，达到国际先进水平。商业化进展方面，2026 年该产品处于研发及客户验证阶段。

图52：源杰科技研发 300mW 高功率 CW 激光器芯片



资料来源：源杰科技公告，东兴证券研究所

仅供内部参考，请勿外传

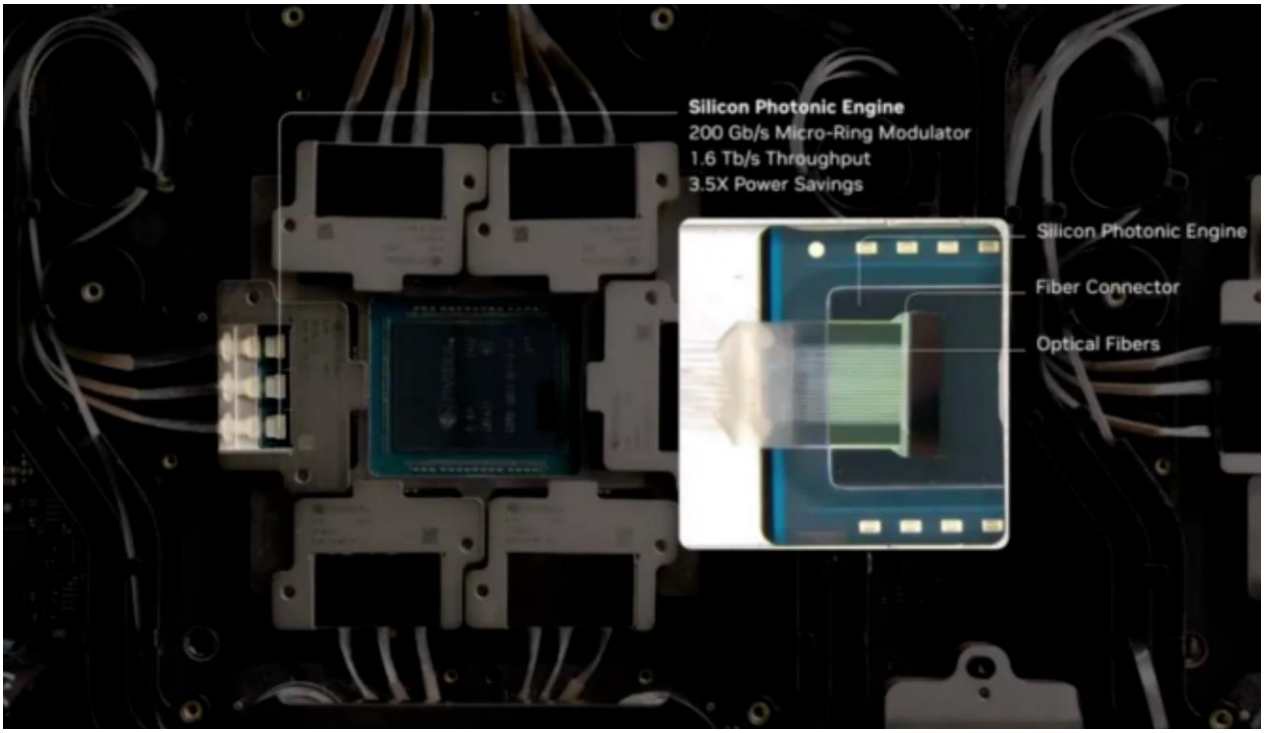
5. CPO：产业化提速，台积电 COUPE 引领硅光集成落地

5.1 CPO 成为光模块终极方案，硅光子产业链迎来发展拐点

当下 CPO 已经成为光模块行业确定趋势。传统方案是采用可插拔光模块，部署在交换机前面板。但根据 AI 网络带宽的发展路线图，互连的速度、距离、密度及可靠性要求即将超越传统光模块所能提供的极限。为突破上述局限，CPO（光电共封装）已成为业界公认的 AI 及超算高密度互连终极方案，其通过将光引擎 OE（Optical Engine）和交换芯片 ASIC 共基板封装在一起，实现极致能效、带宽密度与低时延。

从硬件架构来看，典型的 CPO 交换机芯片模块由五大核心单元协同组成：交换 / 计算 ASIC 封装体、光引擎模块、高性能互连基板、先进散热系统以及高密度光纤管理与连接系统。以英伟达在 GTC 2025 公布，2025 年下半年上市的首款 CPO 交换机 Quantum-X 为例。该款 CPO 交换机提供 115.2Tb/s 的硅光交换机，能够提供 144 个 800G 端口；交换机内部包含 4 个 switch 芯片模组，单片数据吞吐量为 28.8Tb/s；每个 switch 周围有 6 个光学组件即 optical sub-assemblies，每个光学组件内包含 3 个 1.6T 硅光引擎，共 18 个（18*1.6=28.8T）。每个硅光引擎上有两个外置激光器的输入口，连接方式是光纤，输出是 16 路光纤，共 324 路光纤（输出光纤路数 16*3*6=288，输入光纤路数 2*3*6=36）。

图53：英伟达首款 CPO 交换机 Quantum-X 内部结构



资料来源：英伟达，东兴证券研究所

仅供内部参考，请勿外传

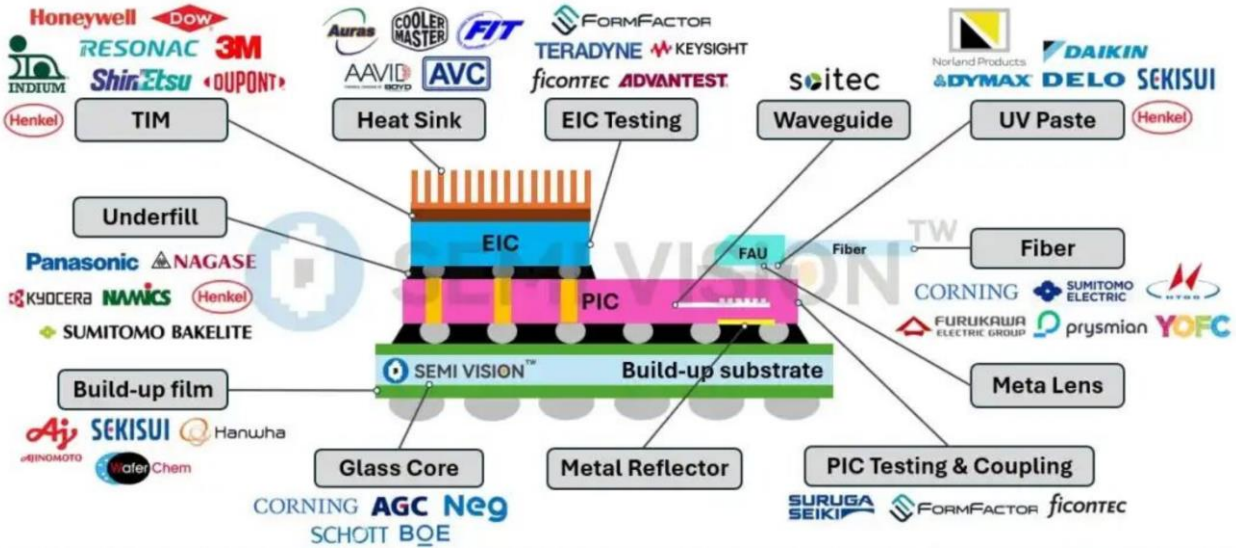
CPO 方案将带动全球硅光子（SiPh）产业链迎来发展拐点。过往硅光子依托 CMOS 工艺实现光电器件集成，处于小批量定制化阶段，始终缺乏大规模落地场景。而 CPO 则是硅光子规模化落地的刚需载体。受益于英伟达 Spectrum-X 以太网硅光技术 2026 年 6 月全面量产，CPO 技术有望在 2027 年迎来规模化部署。

硅光子 SiPh 产业链可以拆解为“材料-器件-封测-系统”四个层级。其中底层是材料层，主要包括热界面材料、底部填充材料、积层膜、玻璃芯基板、紫外胶等；第二层是核心有源及无源器件，主要包括散热部件、光学部件（波导、光纤、超构透镜）、激光源、光纤阵列单元等；第三层是代工制造与封测层，主要包括光学封装平台、电芯片测试、光芯片测试与耦合等；最后一层是系统模块，主要包括 CPO 光引擎以及 CPO 交换机芯片模组。

以英伟达第二代 CPO 交换机 Spectrum-X 为例，Spectrum-X Photonics 计划于 2026 年下半年上市，2026 年 6 月实现量产。Spectrum-X 交换机通过与中国台湾地区半导体和系统生态合作伙伴的深度协同工程实现量产，台积电、日月光（SPIL）、天孚通信（T
层提供了突出贡献。

- 台积电先进的硅光制造技术，将英伟达突破性设计转化为可投入生产的芯片。
- SPIL 的芯片级封装、组装和测试技术，将电气和光学组件以微米级精度结合在一起。
- T
-
- NVIDIA AI 工厂系统在 NVIDIA 自有和运营的 AI 工厂内进行拆箱、安装和通电，在客户发货前验证整体工作流。

图54：全球硅光子产业生态与供应链

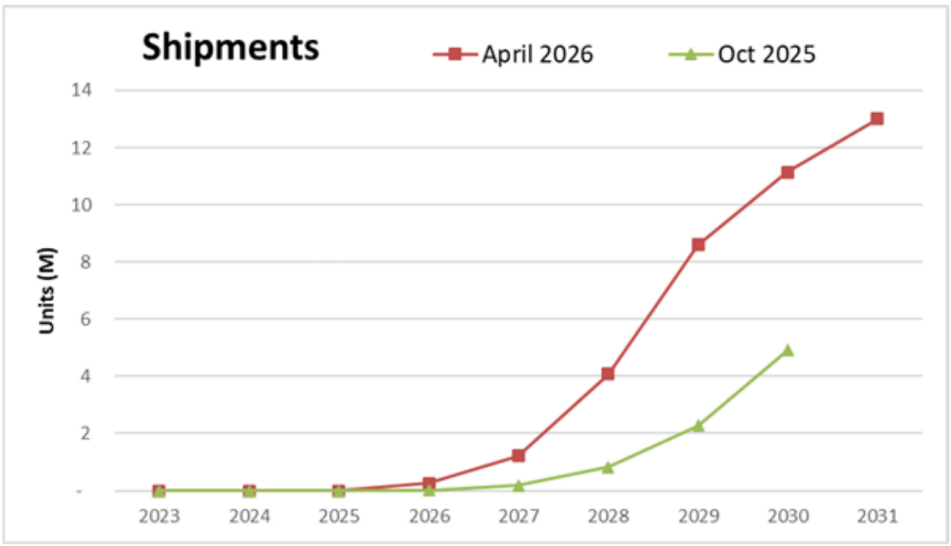


资料来源：SemiVision，东兴证券研究所

仅供内部参考，请勿外传

CPO 市场 2027 年有望突破 50 亿美元。LightCounting 在 2026 年 4 月对 1.6T CPO 产品出货量进行显著上调。2023-2026 年，1.6T CPO 产品为技术导入期，出货量几乎为零，处于试点部署阶段；2027 年起进入大规模放量阶段，增长曲线陡峭上扬，2029 年 1.6T CPO 产品出货量预测从约 200 万个大幅上调至约 900 万个，2031 年则从约 500 万个上调至约 1300 万个。市场规模方面，CPO 市场 2027 年有望突破 50 亿美元，2030 年增长至 150 亿美元。

图55：全球 1.6T CPO 共封装光学出货预测



资料来源：LightCounting，东兴证券研究所

仅供内部参考，请勿外传

5.2 CPO 架构下光引擎跃升为价值链核心，形成 2D 封装到 3D 单片集成多元技术路线

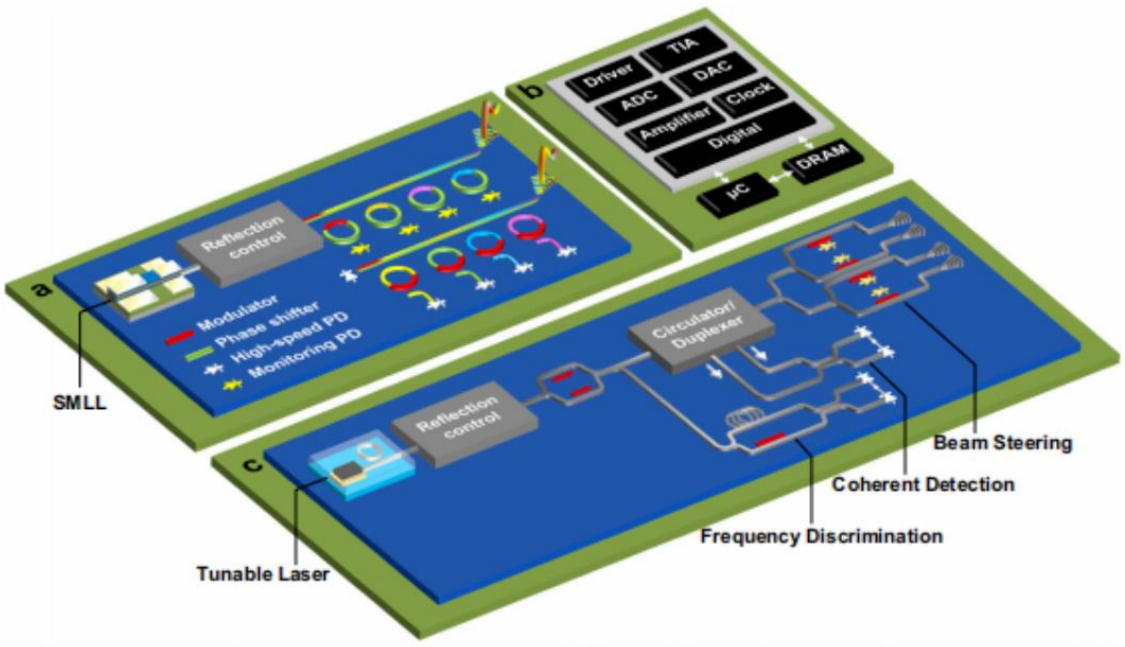
CPO 架构下，光引擎占据重要价值链地位。可插拔光模块价值量主要集中在 DSP、光芯片、封装与测试等环节；进入 CPO 时代，光引擎（PIC 与 EIC 芯片设计与集成）成为价值占比最高环节，外置光源方案（ELS）是当前 CPO 商用化主流路径，价值占比稳定；无源器件成为大规模商用制约瓶颈。

从硅光子系统级应用架构看，光引擎是光电集成的核心载体，主要由 PIC 光子集成电路与 EIC 电子集成电路两部分集成构成。

其中 PIC 负责光信号处理。行业主流方案是以硅光 SiPh 技术为核心，依托 CMOS 工艺，实现波导、调制器、探测器等器件的高密度集成；此外也可采用磷化铟 InP、薄膜铌酸锂 T
光子集成电路（PIC）核心模块包括三类：（1）SMLL（单频微腔激光器）/Tunable Laser（可调激光器）：提供光源；（2）Modulator（调制器）+ Phase shifter（移相器）：实现电光调制，加载信号；（3）High-speed PD（高速光电探测器）+ Monitoring PD（监测探测器）：实现光信号接收与功率监控；

EIC 负责电信号驱动与接收，实现电光信号的转换与处理。电子集成电路（EIC）核心模块主要包括：Driver（驱动）、TIA（跨阻放大器）、DAC/ADC（数模 / 模数转换器）、时钟、数字处理单元等。

图56：硅光子系统级应用架构



资料来源：《Roadmapping the next generation of silicon photonics》（John E. Bowers 等），东兴证券研究所

仅供内部参考，请勿外传

从集成技术看，光引擎（PIC-EIC）形成传统 2D 封装到 2.5D 中介层，再到 3D 单片集成的完整技术谱系，集成工艺越先进，带宽、功耗优势越显著。其中台积电 COUPE 通用光引擎代工平台使用 SoIC 键合工艺进行接合，是仅次于 3D 单片集成下的最短走线长度方案，技术路线处于行业领先地位。

（1）2D 封装：左侧方案是引线键合，将 PIC 和 EIC 并排贴装在共封装基板上，用金属引线键合实现互连；该方案工艺最简单、成本最低，但引线寄生参数大，电串扰严重，仅适合低速信号，已无法支撑高速收发器需求。右侧方案是倒装芯片，通过微凸点（μ-bump）将 PIC 和 EIC 倒装键合在基板上，缩短互连距离；该方案比引线键合的寄生参数更小，适合中高速场景，是早期 400G / 部分 800G 模块的方案之一。

（2）2.5D 封装：PIC 和 EIC 通过硅 / 有机中介层（Interposer）倒装键合，中介层上的高密度布线实现两者的高速互连。该方案互连距离短，寄生参数和电串扰大幅降低，支持 50GHz + 高速信号；工艺成熟度高，可复用成熟的倒装芯片和中介层技术，良率和成本可控。

（3）3D-WB/
时通过引线键合补充互连；右侧方案是 PIC 倒装在 EIC 上方（
微凸点实现垂直互连；3D-WB/
但工艺复杂度提升，目前多用于特定高性能场景。

（4）3D Hybrid TSV/
直互连。该方案信号路径极短，寄生参数和串扰降至最低，支持 100GHz + 超高速信号；集成密度大幅提升，适合 CPO（共封装光学）等高集成度场景；TSV 工艺复杂度高，成本昂贵，目前处于商用早期阶段。

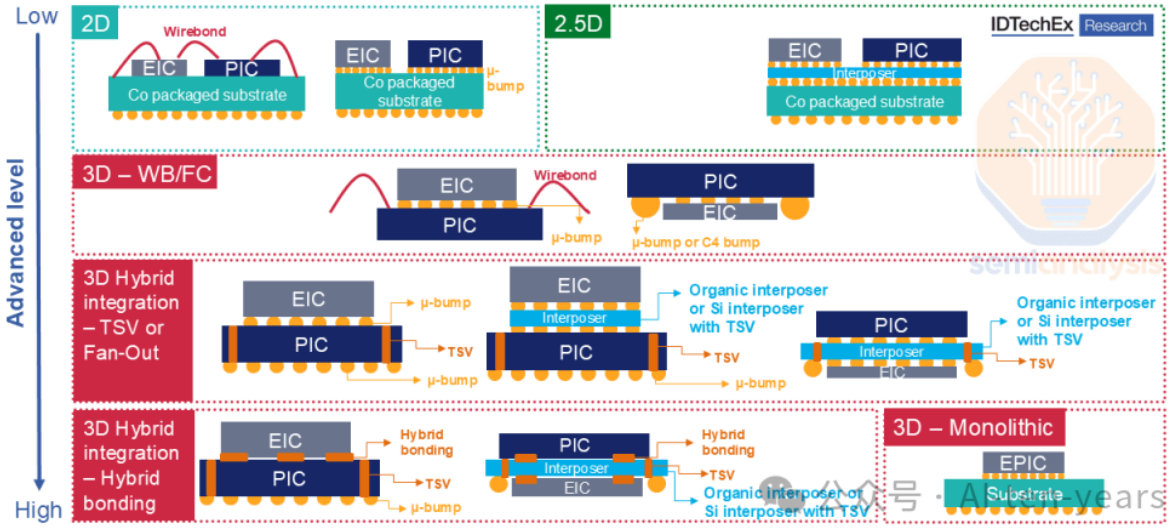
- 左侧方案：EIC 倒装 + TSV PIC：EIC 通过微凸点倒装在 PIC 上，PIC 内部的 TSV 实现垂直信号传输；
- 中间方案：中介层 + TSV PIC：PIC 通过带 TSV 的中介层与 EIC 互连，进一步提升布线密度；
- 右侧方案:PIC 倒装 + 带 TSV 的中介层:PIC 倒装在中介层上,通过 TSV 实现与下方 EIC 的互连；

（5）3D Hybrid Bonding：更先进的异质集成方案，采用混合键合（Hybrid Bonding）技术实现 PIC 与 EIC 的直接键合。通过金属介质混合键合，将 PIC 和 EIC 直接键合，无需微凸点，实现晶圆级的高密度互连；互连密度比 TSV 方案更高，键合界面平整，寄生参数极低，是未来高性能硅光子的核心技术之一；超大规模数据中心 CPO、相干通信等对带宽和密度要求极高的场景。

（6）3D Monolithic（单片 EPIC 集成）：硅光子封装的 终极形态，在同一晶圆上同时制作光子和电子器件（EPIC）。在 CMOS 或 BiCMOS 工艺中直接集成光子器件，实现 PIC 与 EIC 的单片集成；无封装寄生，性能理论最优，成本在大批量时可大幅降低；硅光子与 CMOS 工艺兼容性差，光子器件面积远大于电子器件，良率和成本控制难度极大，目前仍处于研发阶段。

仅供内部参考，请勿外传

图57：CPO 涉及多种光引擎集成封装方法



资料来源：IDTechEx，东兴证券研究所

仅供内部参考，请勿外传

5.3 台积电 COUPE 平台实现技术落地与迭代，绑定标杆客户

2026 年 4 月，台积电 COUPE 平台经过反复仿真与制程优化，已转化为完整的半导体制程技术，2026 年 6 月进入量产。COUPE 平台技术路线分三步走，产品性能与集成度逐级跃升。

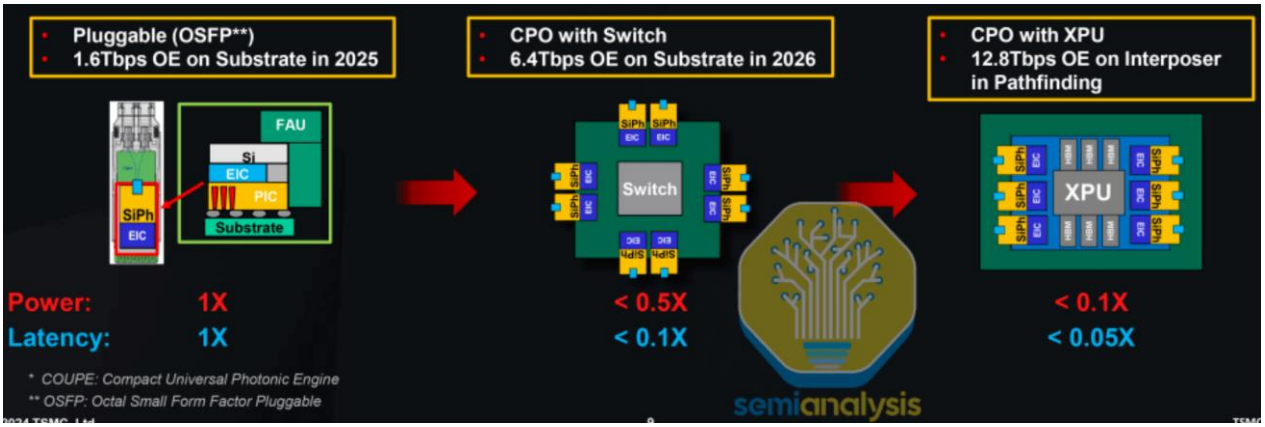
阶段 1：2025 年实现基于 OS（1.6Tb/s）光学引擎。

阶段 2：2026 年实现基于台积电 CoWoS（Chip-on-Wafer-on-Substrate）中介层技术的共封装光学（CPO），光学引擎直接集成在交换机基板上；规格为 6.4Tbps 光学引擎。

阶段 3：2026 年之后探索光学引擎直接集成在 XPU（AI/CPU/GPU）的 CoWoS 中介层上，实现“芯片级光互联”；规格为 12.8Tbps 光学引擎。

以 2025 年 OS 10 倍；基于 XPU 的 CoWoS CPO，以进一步实现 5 倍的功耗降低和 2 倍的延迟降低。

图58：台积电发布的 CPO 技术路线图



资料来源：台积电，东兴证券研究所

仅供内部参考，请勿外传

英伟达与博通成为台积电光引擎代工标杆客户。英伟达的 Quantum-X Photonics 交换机架构采用基于 COUPE 光引擎的光子组件，每个组件在 8 条 200 Gbps PAM4 发送通道和 8 条 200 Gbps PAM4 接收通道上提供 1.6 Tbps 的发送吞吐量和 1.6 Tbps 的接收吞吐量； Spectrum-X Ethernet Photonics 封装密度更高，在一个多芯片模块中集成了 32 个硅光子引擎，每个引擎的吞吐量为 3.2 Tbps。2026 年 6 月，基于台积电 COUPE 平台，英伟达新一代 CPO 交换机 Spectrum-X 实现量产。此外英伟达 COUPE 平台并不局限于单一的垂直整合生态系统。博通 Tomahawk 6 Davisson 交换机也将采用基于台积电 COUPE 的光引擎。

表5：英伟达与博通发布的 CPO 交换机产品矩阵

发布时间	代表产品	核心参数
GTC 2025 发布	英伟达 Quantum-X800-Q3450	1) 总带宽：115.2 Tb/s（4×28.8 Tb/s Quantum-X800 ASIC） 2) 端口：144×MPO；144×800G / 72×1.6T 逻辑端口 3) 光引擎：每 ASIC 18× 光引擎，单引擎 1.6 Tb/s 4) 散热：液冷；单端口功耗 9 W
GTC 2025 发布	英伟达 Spectrum-X SN6800/SN6810	1) 交换芯片：102.4 Tb/s，TSMC 3 nm，COUPE 硅光共封装 2) 端口：128×800G / 512×200G（PAM4） 3) 封装：EIC+PIC 3D 键合，无独立 DSP 4) 功耗：单端口 9 W；整机较传统 −50%+
2024 发布	博通 BCM56990（Tomahawk 5-Bailly，第二代 CPO）	1) 总带宽：51.2 Tb/s；8×6.4 Tb/s 光引擎 2) 端口：512×100G 3) 工艺：7 nm CMOS EIC； 4) 架构：硅光 PIC + 锗硅 EIC，无独立 DSP
2025-10 正式发布	博通 BCM78919（Tomahawk 6-Davisson，第三代 CPO）	1) 总带宽：102.4 Tb/s；16×6.4 Tb/s Davisson DR 光引擎 2) 端口：512×200G（PAM4） 3) 工艺：TSMC 3 nm； COUPE 紧凑型通用光子引擎 4) 激光模块：ELS

资料来源：GTC，英伟达官网，博通官网，Semianalysis，东兴证券研究所

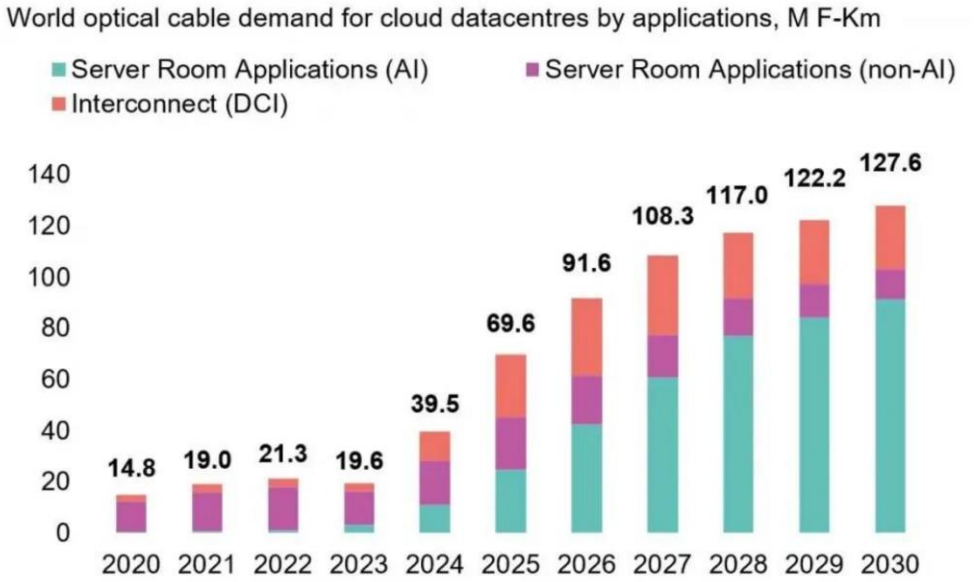
仅供内部参考，请勿外传

6. 光纤：受益于数据中心建设浪潮，迎来景气周期

6.1 G.657 单模光纤和多模光纤将迎来规模化应用

AI 应用驱动数据中心光纤需求快速增长。根据 CRU 最新报告，2026 年全球数据中心光纤需求预计达到 9160 万芯公里，同比增长 32%。根据具体应用场景，可以看到，2026 年 AI 应用、非 AI 应用和数据中心互连（DCI）应用的光纤需求均实现较快增长。展望 2030 年，全球数据中心光纤需求预计达到 1.28 亿芯公里，其中 AI 应用的光纤需求超过 8000 万芯公里，成为重要增长驱动力。

图59：全球数据中心光纤应用规模（万芯公里）



资料来源：CRU，东兴证券研究所

仅供内部参考，请勿外传

G.657 单模光纤适用于数据中心。根据 ITU 标准下传输等级分类，单模光纤产品可以分为 G.652/G.653/G.654/G.655/G.656/G.657 等类型。其中 G.652D 是所有 G.652 级别中指标最严格的，兼容性好，是光纤到户（
G.652 光纤的弯曲半径的 1/4～1/2,适用于数据中心机柜内、楼道配线盒等空间受限或可能频繁弯折的场景。

表6：ITU 标准下单模光纤产品

等级分类	核心特性	典型应用场景
G.652（常规单模光纤）	• 零色散点在 1310nm 波长附近，在此波长传输性能最佳；1550nm 波长处损耗最低（约 0.2dB/km） • 分为 G.652.A、B、C、D 四个子类，其中 G.652.D（低水峰光纤）是当前应用最广泛的型号，消除了 1383nm 附近的损耗峰，支持更宽的传输波段	• 适用于城域网、接入网等绝大多数通信场景 • G.652D 成本较低，兼容性好，是光纤到户（和接入网的常见选择
G.653（色散位移光纤）	• 将零色散点从 1310nm 移至 1550nm，使最低损耗和零色散在同一窗口	• 目前已基本被淘汰，不再在新工程中部署
G.654（低损耗光纤）	• 在 1550nm 波长处具有极低的衰减系数（部分型号可低至 0.15dB/km 以下）	•长距离、大容量通信的首选，如海底光缆、跨洋通信 • 适用于 100G、400G 及更高速率的骨干网传输
G.655（非零色散位移光纤）	• 在 1550nm 窗口保留少量且可控的色散	• 适合长距离骨干网和高速 DWDM 系统（如 40G/100G），在保持低损耗的同时保证了多波长的稳定传输
G.656（宽波长范围光纤）	• 在更宽的波段（如 1460-1625nm 的 S+C+L 波段）保持一定的色散系数	• 面向未来超宽带 DWDM 系统设计，可支持更多的通信信道 • 适用于未来 5G/6G 的承载网
G.657（抗弯曲光纤）	• 具备优异的抗弯曲性能，最小弯曲半径可小至 5-7.5mm（传统 G.652 需 30mm） • 性能上通常兼容 G.652 光纤	• 室内布线（如光纤到户 • 适用于数据中心机柜内、楼道配线盒等空间受限或可能频繁弯折的场景

资料来源：电子发烧友网，东兴证券研究所

仅供内部参考，请勿外传

多模光纤产品可以分为 OM3、OM4、OM5，适用于数据中心短距离、高带宽的场景。其中，OM3 支持多种光模式并实现快速数据传输；OM4 是当前新建高性能数据中心的主流选择，在相同速率下提供比 OM3 更长的传输距离；OM5 光纤完美支持短波波分复用（SWDM）技术，可在同一根光纤上同时传输多个波长，相当于将单车道公路拓宽为四车道，传输容量提升至原来的 4 倍。在北美数据中心，多模 OM4 和 OM5 光纤总体用量较少。

图60：多模光纤产品示意图



资料来源：科兰布线公众号，东兴证券研究所

根据结构划分，光缆可分为单管主干跳线或者多单元主干跳线。单管主干跳线内部设计采用单一的光纤管来容纳多根光纤；多单元主干跳线采用多个独立的光纤单元来组成整个光缆，每个光纤单元内部都包含一定数量的光纤。常见的单管有 8 芯、12 芯、24 芯，而多单元线缆也根据需求定制单元数及每单元的芯数，如 48 芯、96 芯、144 芯等。

图61：单管主干跳线与多单元主干跳线产品示意图



资料来源：科兰布线公众号，东兴证券研究所

仅供内部参考，请勿外传

单模光纤和多模光纤均可以支持数据中心内高传输速率。在相同应用、速率下，单模光纤的传输距离比多模光纤要远得多。例如在 400Gbps 传输应用下，多模光纤的传输距离约 100 米，而单模光纤的传输距离最多可以做到 10000 米。在光纤造价上，多模光纤比单模光纤的纤芯尺寸大得多，所以多模光纤成本比较高；发光器造价上，在同样的传输速率下，单模激光器造价往往会比多模激光器要高。因此，尽管多模光纤本身的成本较高，但由于连接、安装、低功耗以及持续维护和可用性的成本较低。

图62：单模光纤和多模光纤产品传输距离对比

光纤类型	应用类型	波长	连接器类型	调制技术	协议标准	最大传输距离
单模光纤	400G OSFP DR4	1310nm	MPO/MTP-12	100G PAM4	IEEE 802.3bs	500米
	400G OSFP DR4+	1310nm	MPO/MTP-12	100G PAM4	/	2000米
	400G OSFP FR4	CWDM4波长	LC	100G PAM4	100G lambda多源协议（MSA）	2000米
	400G OSFP FR8	CWDM4波长	CS	50G PAM4	IEEE 802.3bs	2000米
	400G OSFP LR4	CWDM4波长	LC	100G PAM4	100G lambda多源协议（MSA）	10000米
多模光纤	400G OSFP SR8	850nm	MPO/MTP-16	50G PAM4	IEEE P802.3cm/IEEE 802.3bs	100米

资料来源：科兰布线公众号，东兴证券研究所

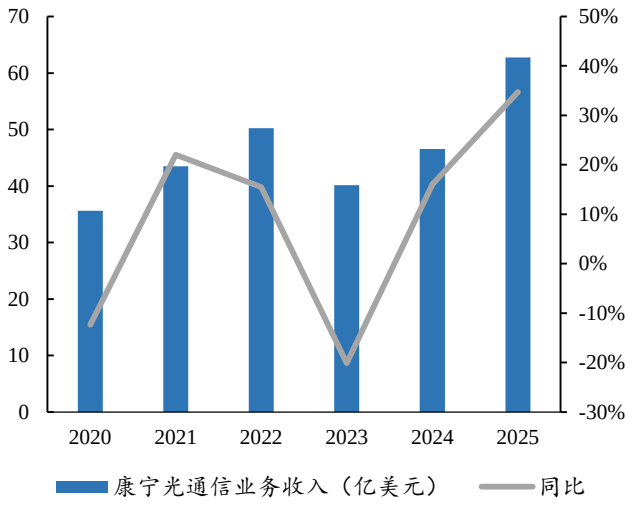
仅供内部参考，请勿外传

6.2 海外康宁光通信业务高增，映射国内数据中心光纤市场机会

康宁光通信业务高增，是由生成式 AI 引发的数据中心架构革命所驱动的。这种需求并非北美独有，而是全球算力基础设施建设的共性趋势。目前北美 AI 数据中心建设进展领先国内，因此我们通过总结康宁近几年财务数据表现，从而分析国内数据中心光纤市场空间。

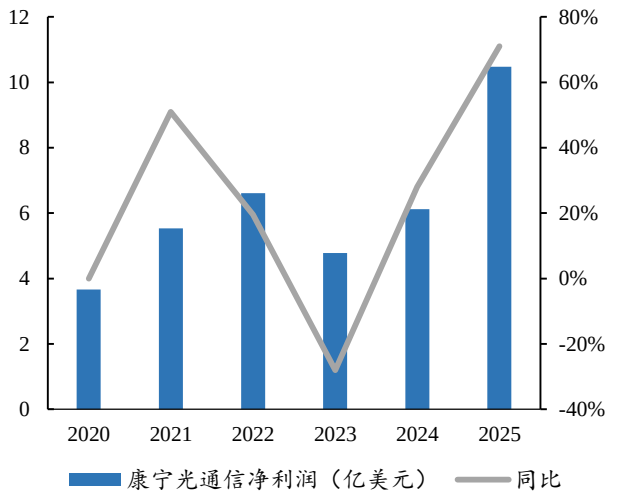
受益于市场对康宁新型生成式人工智能（Gen AI）产品的持续强劲采用，2024-2025 年康宁光通信业务实现强劲增长。2024 年，康宁光通信业务实现销售收入 46.57 亿美元，同比增长 16%；净利润 6.12 亿美元，同比增涨 28%。2025 年，康宁光通信业务收入进一步增长至 62.7 亿美元，同比增长 34.7%；净利润进一步增长至 10.5 亿美元，同比增长 71%。展望未来，公司与大型科技企业的合作，预计将使得公司收入与利润高增长具有持续性。近期康宁宣布与 Meta 达成一项长期供应协议，合同总金额最高可达 60 亿美元，合作期限持续至 2030 年，主要提供光纤、光缆及光连接解决方案。

图63：康宁光通信业务收入及增速（亿美元，%）



资料来源：公司年报，东兴证券研究所

图64：康宁光通信业务净利润及增速（亿美元，%）



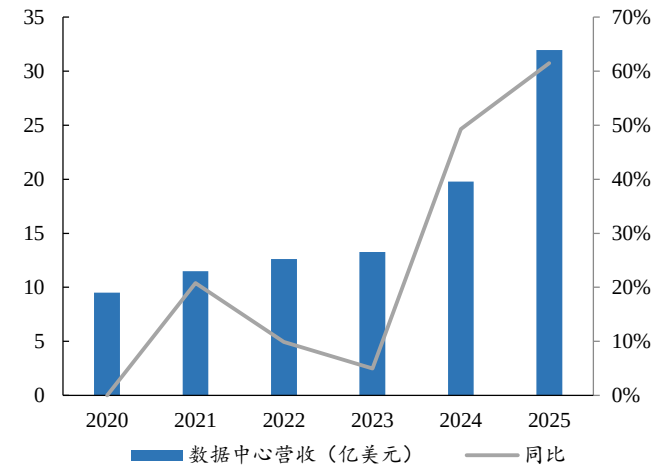
资料来源：公司年报，东兴证券研究所

康宁光通信业务主要分为两大应用领域：运营商业务和数据中心业务。其中，运营商业务主要为电信运营商提供用于视频、数据及语音通信的光通信基础设施，而数据中心业务则主要为企业客户、政府机构以及超大规模云计算数据中心提供光连接解决方案。

仅供内部参考，请勿外传

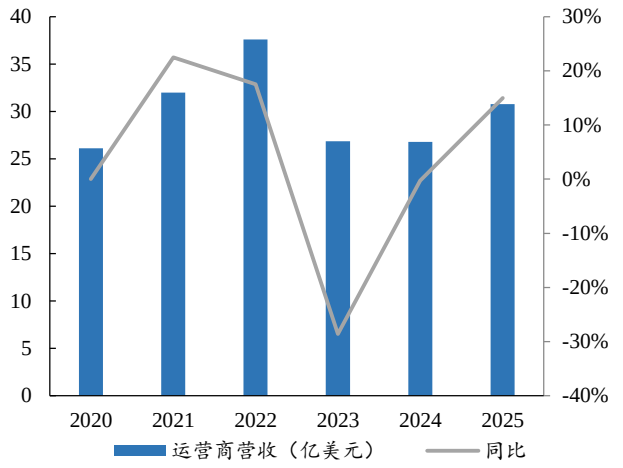
海外 AI 数据中心扩张，为康宁光通信业务带来每年 19 亿美元的增量市场。2023 年，康宁光通信业务数据中心板块实现收入 13.3 亿美元，2025 年该板块增长至 32.0 亿美元。康宁光通信业务运营商板块整体收入稳健，2025 年运营商板块收入 30.8 亿美元，2025 年康宁光通信业务数据中心板块收入规模超过运营商板块。

图65：康宁光通信数据中心营收及增速（亿美元，%）



资料来源：公司年报，东兴证券研究所

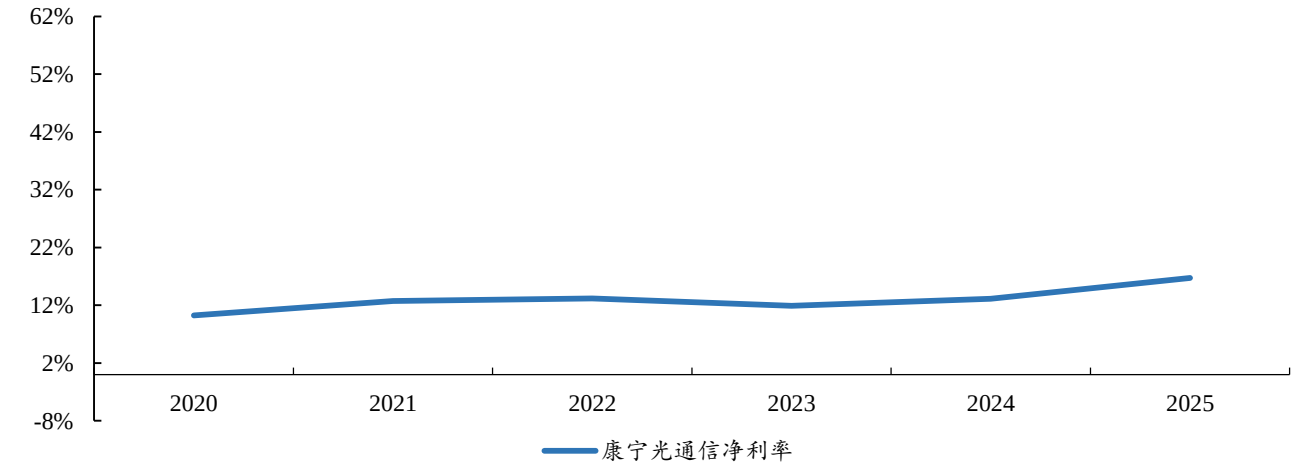
图66：康宁光通信运营商营收及增速（亿美元，%）



资料来源：公司年报，东兴证券研究所

数据中心光纤光缆产品及解决方案附加值高，康宁光通信业务净利润率显著提升。2023 年公司光通信业务净利率 12%，2024 年净利率回升至 13%。2025 年公司净利率进一步提升至 17%，达到近年来最高水平。整体来看，随着数据中心等高附加值业务占比持续提升，康宁光通信业务的盈利能力有望进一步改善。

图67：康宁光通信净利率情况（%）



资料来源：公司年报，东兴证券研究所

仅供内部参考，请勿外传

6.3 字节跳动 AI 智算中心规模化交付，驱动国内数据中心光纤市场超过 50 亿元

字节跳动作为国内科技头部企业，深入布局 AIGC 产业链，旗下 AI 产品（如豆包）月活用户已稳居国内榜首，且使用时长呈指数级增长；同时，火山引擎 MaaS（模型即服务）业务快速增长。**我们认为，字节跳动庞大的资本开支具有扎实的市场需求，具有可持续性和引领性。**从短期看，2026 年是字节跳动自建 AI 智算中心规模化交付之年，有望显著提升国内数据中心光纤市场需求。从中长期看，随着国内市场进一步繁荣发展，国内更多科技企业将加大 AI 智算中心投资。

AI 应用规模化落地，豆包与 Deepseek 建立头部优势。根据 QuestMobile 2025 年 12 月 AI 原生 App 月活数据，字节跳动旗下豆包以约 2.27 亿月活跃用户规模稳居行业第一，第二名 DeepSeek 月活约 1.36 亿，元宝、蚂蚁阿福、千问分别以 4071 万、2689 万、2517 万位列第三至第五。

图68：2025 年 12 月国内 AI 原生 App 月活数据（万人）



资料来源：Questmobile，东兴证券研究所

仅供内部参考，请勿外传

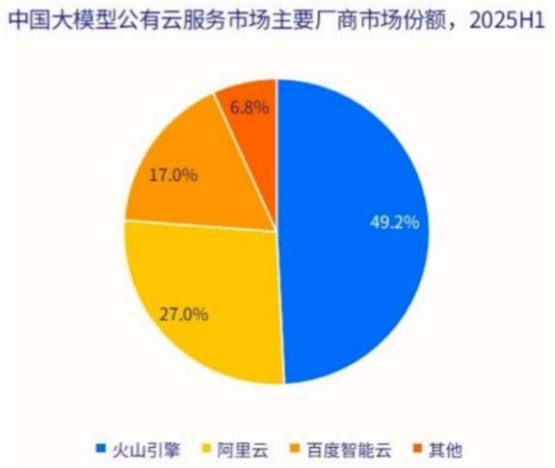
AI 应用驱动大模型公有云服务调用市场快速发展。截至 2025 年 12 月，豆包大模型日均 token 使用量已突破 50 万亿，较去年同期增长超 10 倍。根据 IDC 报告，2025 年上半年，中国公有云大模型调用量达 536.7 万亿 tokens。其中，火山引擎以 49.2%的市场份额位居中国市场第一，阿里云、百度智能云分别占比 27%、17%，位列第二、第三位。

图69：豆包日均 Tokens 使用量变化



资料来源：字节跳动，东兴证券研究所

图70：2025H1 中国公有云大模型服务调用量分布



资料来源：IDC，东兴证券研究所

仅供内部参考，请勿外传

2026 年是字节跳动自建 AI 智算中心规模化交付之年。2024 年至今，我们统计字节跳动在国内自建 6 座大型 AI 智算中心，总投资规模 357 亿元。其中火山云太行算力中心项目一期工程投资规划 28 亿元，规划 12 万台高性能 AI 服务器，2025 年进入运营阶段；其他 4 座大型 AI 智算中心预计 2026 年投入运营；第 6 座火山引擎内蒙古和林格尔算力中心二期项目预计 2027 年投入运营。

表7：2024 年至今字节跳动国内数据中心自建项目

序号	项目名称	投资金额（亿元）	数据中心规模	建设进展
1	火山引擎长三角算力中心项目	80	规划服务器机柜数 2.2 万台	2024.07 获得审批； 2024.09 环境影响报告表编制； 2025.03 获得批复； 2025.03 开工建设； 2025.11 投入调试
2	火山引擎长三角无为算力中心项目	80	规划服务器机柜数 2.1 万台； 2 万 P	2024.11 签约； 2025.02 纳统； 2025.06 环境影响报告表受理公示； 2025.08 主体土建工程已完成 85% 预计 2026 年机房将陆续投入使用
3	火山引擎内蒙古和林格尔算力中心一期项目	56	规划 IT 容量 220MW； 规划标准机架 8 万个	2024.11 开工建设； 2025.04 一期 A 项目 110kV 变电站工程核准获批复； 2025.05110kV 变电站主体结构封顶； 2025.09 审议通过一期 B 项目数据中心机房楼规划设计方案； 2025.12 成功送电
4	火山引擎内蒙古和林格尔算力中心二期项目	68	-	2025.04 备案； 2026.02 项目配套 110 千伏供电线路工程施工招标
5	火山云太行算力中心项目一期工程	28	规划 12 万台高性能 AI 服务器	2023.05 开工建设； 2024.07 110kV 变电站完成送电； 2024 年 11 月投入运营
6	火山云太行算力中心项目二期工程	45	规划 12KW 服务器机柜 1.6 万台	2024.07 前期准备工作就绪； 项目于 2025 年 1 月获批； 计划于 2026 年 5 月完工

资料来源：芜湖市人民政府，全国建设项目环境信息公示平台，无为市人民政府，呼和浩特市人民政府，大同市行政审批服务管理局等，东兴证券研究所

仅供内部参考，请勿外传

2026 年字节跳动投入运营的自建数据中心有火山引擎长三角算力中心项目、火山引擎长三角无为算力中心项目、火山引擎内蒙古和林格尔算力中心一期项目、火山云太行算力中心项目二期工程，合计投资规模 261 亿元。

其中火山引擎内蒙古和林格尔算力中心一期项目投资规模 56 亿元，规划 IT 容量 220MW，规划标准机架 8 万个。参考该项目建设 220MW 数据中心需要投资 56 亿元。经测算，字节跳动 2026 年投资 261 亿元，对应约 1GW 智算中心。

根据产业调研和券商公开测算，我们假设：（1）常规情况下，光纤资本开支占比数据中心总投资规模 3.5%；（2）涨价情况下，光纤资本开支占比数据中心总投资规模 4.0%。

经测算，2026 年字节跳动自建 1GW 智算中心，预计光纤资本支出约 9.1-10.4 亿元。

根据产业调研以及国内互联网企业年度资本开支，我们假设：（1）字节跳动智算中心投资规模占比国内互联网行业的 1/3；（2）互联网行业在数据中心领域投资规模占国内数据中心投资规模的 50%。

则 2026 年国内数据中心光纤市场规模约 55-63 亿元。

表8：字节跳动 AI 数据中心光纤资本开支测算

项目	单位	规模
投资规模	亿元	261
AI 智算中心规模	MW	1025
光纤资本开支占比（常规）		3.5%
光纤资本开支（常规）	亿元	9.1
光纤资本开支占比（涨价）		4.0%
光纤资本开支（涨价）	亿元	10.4

资料来源：东兴证券研究所

仅供内部参考，请勿外传

7. 通信行业投资策略

投资策略：第一，超节点与 Scale up 网络是突破算力通信瓶颈、支撑万亿级大模型与高实时性应用的关键基础设施。2026 年国内和北美超节点陆续进入量产交付期，英伟达确立超节点技术范式。国内 AI 算力网络在高速互联领域仍存在技术代差，自主可控需求迫切。第二，当前光互联供应链持续性紧缺，有望贯穿 2026 全年。我们认为，国内光芯片企业有望受益于光互联市场总量增长、硅光架构 CW 激光器结构性机会、国产份额提升三重机遇；Scale up 网络正转向光互联，CPO 生态逐步形成，全球硅光子将从碎片化定制研发转向标准化代工量产模式。国内硅光新材料、无源以及有源器件、以硅光器件设计套件（PDK）、集成测试等厂商有望进入英伟达、台积电供应链体系，加速硅光子国产化验证与量产落地。此外，全球 AI 应用及数据中心建设快速发展，产能缺口驱动光纤价格持续提升。国内光纤企业具有数据中心特种光纤产能优势，有望显著并持续受益。

相关公司：

国产超节点：华勤技术、浪潮信息、中科曙光、工业富联；

国产高速交换芯片：盛科通信、紫光股份、锐捷网络、中兴通讯；

光模块、CPO/NPO：中际旭创、新易盛、天孚通信、华工科技、光迅科技；

光芯片/电芯片：源杰科技、东山精密、仕佳光子、长光华芯、优迅股份；

OCS：腾景科技、福晶科技、矩光科技、德科立；

CPO 设备：联讯仪器、罗博特科；

光纤：亨通光电、中天科技、长飞光纤、烽火通信。

8. 风险提示

（1）英伟达超节点出货量低于预期；（2）CPO 技术路线碎片化；（3）光纤产能过剩与价格下行；（4）AI 应用端增长不及预期；（5）地缘政治风险。

仅供内部参考，请勿外传

分析师简介

石伟晶

首席分析师，覆盖传媒、互联网、云计算、通信等行业。上海交通大学工学硕士。10 年证券从业经验，曾供职于华创证券、安信证券，2018 年加入东兴证券研究所。

分析师承诺

负责本研究报告全部或部分内容的每一位证券分析师，在此申明，本报告的观点、逻辑和论据均为分析师本人研究成果，引用的相关信息和文字均已注明出处。本报告依据公开的信息来源，力求清晰、准确地反映分析师本人的研究观点。本人薪酬的任何部分过去不曾与、现在不与，未来也将不会与本报告中的具体推荐或观点直接或间接相关。

风险提示

本证券研究报告所载的信息、观点、结论等内容仅供投资者决策参考。在任何情况下，本公司证券研究报告均不构成对任何机构和个人的投资建议，市场有风险，投资者在决定投资前，务必要审慎。投资者应自主作出投资决策，自行承担投资风险。

仅供内部参考，请勿外传

免责声明

本研究报告由东兴证券股份有限公司研究所撰写，东兴证券股份有限公司是具有合法证券投资咨询业务资格的机构。本研究报告中所引用信息均来源于公开资料，我公司对这些信息的准确性和完整性不作任何保证，也不保证所包含的信息和建议不会发生任何变更。我们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，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。

我公司及报告作者在自身所知情的范围内，与本报告所评价或推荐的证券或投资标的不存在法律禁止的利害关系。在法律许可的情况下，我公司及其所属关联机构可能会持有报告中提到的公司所发行的证券头寸并进行交易，也可能为这些公司提供或者争取提供投资银行、财务顾问或者金融产品等相关服务。本报告版权仅为我公司所有，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。如引用、刊发，需注明出处为东兴证券研究所，且不得对本报告进行有悖原意的引用、删节和修改。

本研究报告仅供东兴证券股份有限公司客户和经本公司授权刊载机构的客户使用，未经授权私自刊载研究报告的机构以及其阅读和使用者应慎重使用报告、防止被误导，本公司不承担由于非授权机构私自刊发和非授权客户使用该报告所产生的相关风险和责任。

行业评级体系

公司投资评级（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数）：

以报告日后的 6 个月内，公司股价相对于同期市场基准指数的表现为标准定义：

强烈推荐：相对强于市场基准指数收益率 15%以上；

推荐：相对强于市场基准指数收益率 5%～15%之间；

中性：相对于市场基准指数收益率介于-5%～+5%之间；

回避：相对弱于市场基准指数收益率 5%以上。

行业投资评级（A 股市场基准为沪深 300 指数，香港市场基准为恒生指数，美国市场基准为标普 500 指数）：

以报告日后的 6 个月内，行业指数相对于同期市场基准指数的表现为标准定义：

看好：相对强于市场基准指数收益率 5%以上；

中性：相对于市场基准指数收益率介于-5%～+5%之间；

看淡：相对弱于市场基准指数收益率 5%以上。

东兴证券研究所

北京
西城区金融大街 5 号新盛大厦 B 座 16 层
邮编：100033
电话：010-66554070
传真：010-66554008

上海
虹口区杨树浦路 248 号瑞丰国际大厦 23 层
邮编：200082
电话：021-25102800
传真：021-25102881

深圳
福田区益田路 6009 号新世界中心 46
邮编：518038
电话：0755-83239601
传真：0755-23824526