



计算机行业研究

买入（维持评级）
行业周报

证券研究报告

计算机组

 分析师：刘高畅（执业 S1130525120005） 分析师：郑元昊（执业 S1130525120004） 分析师：孙恺祈（执业 S1130526060005）
liugaocang@gjzq.com.cn zhengyuanhao@gjzq.com.cn sunkaiqi@gjzq.com.cn

再谈 Q3 科技最大的驱动力：VERA RUBIN 与谷歌 V8

行业观点

- **Q3 科技最大的驱动力，是英伟达 Vera Rubin 与谷歌 TPU 均逐渐进入量产阶段，算力紧缺仍是产业主要矛盾。**
Rubin 侧，英伟达官宣 Vera Rubin 正在全球加速部署：在 300 家全球合作伙伴支持下，Vera Rubin NVL72 生产持续爬坡，机柜已在 CoreWeave、Google Cloud/Microsoft Azure/Oracle 云基础设施运行；CoreWeave 公布的首个实测基准显示，DeepSeek-R1 推理的每兆瓦 token 吞吐量达 Grace Blackwell NVL72 的 10 倍。TPU 侧，Alphabet 确认 TPU 系统销售已于 Q2 开始向客户交付，收入将在 26 年加速爬坡、多数于 27 年确认。供给端的紧缺同样明确：HBM4 三家供应商全部认证量产、全面规模化量产节点落在 Q3，SK 海力士表示未来三年 HBM 客户需求已超供给；Alphabet 将全年 CapEx 指引从 1800-1900 亿美元上调至 1950-2050 亿美元，并明确 Q3 将租用第三方产能作为过桥方案。上周周报我们提出海外算力 Q3 迎来加速期，本周我们进一步继续强调该观点。
- **Vera Rubin 强在五类机柜在 Pod 层面的系统性耦合。**单看芯片，Rubin GPU 搭载 288GB HBM4、带宽 22TB/s（约为 Blackwell 的 2.8 倍），NVFP4 推理算力 50 PFLOPS、训练 35 PFLOPS，分别为上代的 5 倍与 3.5 倍；NVLink 6 将 GPU 间带宽从 1.8TB/s 翻倍至 3.6TB/s。放到机柜，Vera Rubin NVL144 实现 3.6 EFLOPS 推理算力、20.7TB HBM4 容量与 260TB/s 纵向扩展带宽。再放到 Pod，英伟达以 GPU 计算、存储、CPU、低时延推理、网络五类机柜协同构建 AI 工厂：存储侧官方定义 G3.5 层并推出 CMX 平台，单柜管理约 9600TB 闪存承接 KV Cache；网络侧 Spectrum-6 以太网交换机官方确认采用 CPO 共封装光学。芯片-机柜-工厂的每一级放大，都对应 PCB、HBM、存储、光互连、液冷、供电同步放量，这是预计 Q3 上游供应商业绩环比加速的产业基础。
- **谷歌财报确认 TPU 商业化启动，云与资本开支数据全面超上一轮预期。**Alphabet Q2 收入 1198 亿美元、同比增长 24%，经营利润 408 亿美元、同比增长 30%；Google Cloud 收入 248 亿美元、同比增长 82%，云业务积压单达 5140 亿美元、环比+500 亿美元。Q2 CapEx 449 亿美元，其中六成投向服务器；管理层预告 27 年 CapEx 将较 26 年显著增长。TPU 系统销售模式将芯片直接部署到客户数据中心，打开云服务外第二增长；算力紧缺下，谷歌 6 月与 SpaceX 签署 9.2 亿美元/月算力租用协议，优先保障 AGI 前沿研发、其次为搜索与 YouTube、再次为云服务。TPU 出货加速与外销启动，将同步带动上一篇梳理的机柜组装、光互连、液冷、供电链条。
- **机柜供应商、核心 PCB 板厂、光互联是产业链主线。**机柜、PCB、光模块跟随 Rubin 代际切换的逻辑不变：工业富联 Rubin 机柜 Q3 启动量产，PCB 中板替代线缆的增量价值兑现即，Spectrum-6 将 CPO 确立为官方标配、光互连从验证叙事升格为出货叙事。功耗主线上，液冷与 800VDC 供电架构的确定性此前已充分论述；电容侧：三星电机自 2 月起释放涨价信号、4 月计划对 MLCC 实施两位数涨价、5 月再酝酿 5%至 10%的追加调涨，现货市场 MLCC 价格自 2 月底以来普涨 15%至 20%、AI 服务器高容值产品涨幅达 50%至 60%，7 月初公司再与全球大型科技企业签署 4540 亿韩元的 AI 服务器 MLCC 长期供货合同。

相关标的

- 1) 超核心供应商：工业富联、胜宏科技
- 2) 海外算力：中际旭创、新易盛、东山精密、江海股份、中钨高新、蓝思科技、东阳光、光智科技、先导基电、火炬电子、三环集团、欧科亿、天孚通信、鼎泰高科、领益智造、兆易创新、鹏鼎控股、唯科科技、天岳先进、大普微、源杰科技、麦格米特、景旺电子、英维克、京东方等；英特尔、SK 海力士、Lumentum、闪迪、铠侠、美光。
- 3) 国内算力：寒武纪、海光信息、东阳光、中芯国际、华虹半导体、禾盛新材、摩尔线程、天数智芯、沐曦股份、壁仞科技、协创数据、华丰科技、杰华特、京基智农、华达科技、中科曙光、利通电子、浪潮信息、利扬芯片、胜蓝股份、华勤技术、国科微、中国长城、晶科科技、罗曼股份、盈峰环境、芯原股份、亿田智能、豫能控股、星环科技、鸿日达、盛视科技、神州数码、润泽科技、大位科技、润建股份、奥飞数据、瑞晟智能、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子。
- 4) 大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、首都在线、网宿科技、云赛智联、青云科技等。

风险提示

- Rubin 量产爬坡不及预期的风险；ASIC 出货节奏低于预期的风险；数据中心基础设施改造进度的风险；贸易摩擦与出口管制的风险；第三方预测数据时效性的风险。



内容目录

一、Vera Rubin: Q3 步入量产部署阶段，五类机柜协同构建 AI 工厂	3
1.1 从单机柜到 AI 工厂: Rubin 的竞争力强在系统耦合	3
1.2 计算与互联: HBM4、FP4 引擎与 NVLink 6 三箭齐发	4
1.3 存储机柜: G3.5 层与 CMX 平台, KV Cache 成为新的硬件需求	4
1.4 异构与网络: CPU 与 LPX 各司其职, Spectrum-6 将 CPO 确立为官方标配	5
二、谷歌财报更新: TPU 系统销售启动交付, 算力供给依然紧张	6
2.1 Q2 财报全面超预期, 资本开支指引再度上调	6
2.2 TPU 系统销售 Q2 启动交付, 2027 年进入收入放量期	6
三、产业链主线延续: 机柜、PCB 与光模块跟随代际切换	7
四、功耗主线延续: 液冷与供电确定性不变, 电容迎来涨价落地	7
4.1 液冷与 800VDC: 官方口径的确定性方向	7
4.2 电容: 三星电机涨价从信号到落地, 供需紧张有望延续至 2027 年上半年	8
五、相关标的	8
六、风险提示	8

图表目录

图表 1: 英伟达官方将 Vera Rubin 平台定义为六颗协同设计芯片的组合	3
图表 2: NVLink-C2C 一致性内存架构	3
图表 3: NVIDIA Groq 3 LPX 机架级系统	3
图表 4: NVIDIA Vera Rubin NVL72 和 LPX 提供了 10 倍的收入增长机会	4
图表 5: NVIDIA Vera Rubin 集成 6 芯片的平台设计	4
图表 6: SK 海力士 1Q26 业绩数据	4
图表 7: 英伟达 Dynamo 框架下 KV Cache 在 G1 至 G4 层间的卸载序列	5
图表 8: 不同 AI 工作负载的相对计算和交互需求, 包含聊天机器人、语音助手、深度研究、代码助手等, 愈发强调代理群集对高吞吐量和低延迟的需求	6
图表 9: Alphabet 2Q26 收入 1198 亿美元, 同比增长 24%	6
图表 10: Alphabet 2Q26 资本开支 449 亿美元	6
图表 11: TPU 8t ASIC block diagram	7

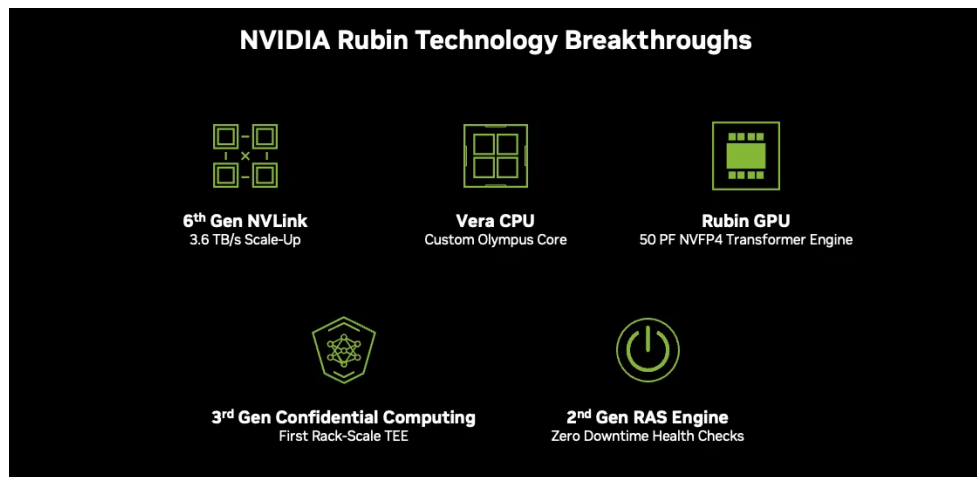


一、Vera Rubin: Q3 步入量产部署阶段，五类机柜协同构建 AI 工厂

1.1 从单机柜到 AI 工厂：Rubin 的竞争力强在系统耦合

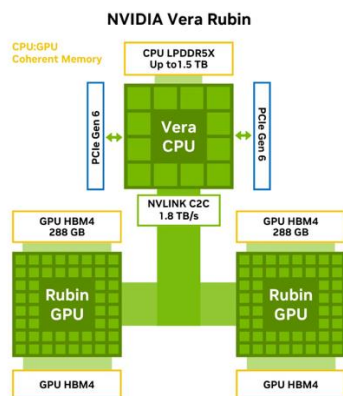
Vera Rubin 是一套 AI 工厂级系统工程方案，而非单一 GPU 或机柜升级。英伟达官方将 Vera Rubin 平台定义为六颗协同设计芯片的组合：Rubin GPU、Vera CPU、NVLink 6 交换芯片、ConnectX-9 SuperNIC、BlueField-4 DPU 与 Spectrum-6 以太网交换芯片，共同构成一台同步运转的 AI 超级计算机——GPU 执行 Transformer 时代的计算负载，CPU 编排数据与控制流，纵向与横向扩展网络高效搬运 token 与状态，专用基础设施处理器负责 AI 工厂自身的运维与安全；3 月 GTC 后平台再纳入第七颗芯片 Groq 3 LPX 低时延推理加速器。

图表1：英伟达官方将 Vera Rubin 平台定义为六颗协同设计芯片的组合

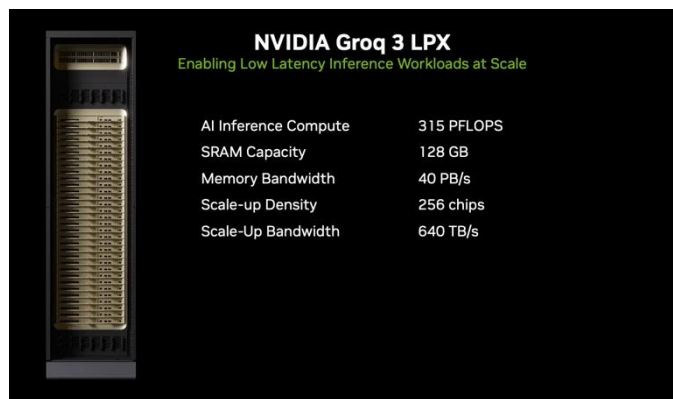


来源：英伟达官方科技博客，国金证券研究所

图表2：NVLink-C2C 一致性内存架构



图表3：NVIDIA Groq 3 LPX 机架级系统



来源：英伟达官方科技博客，国金证券研究所

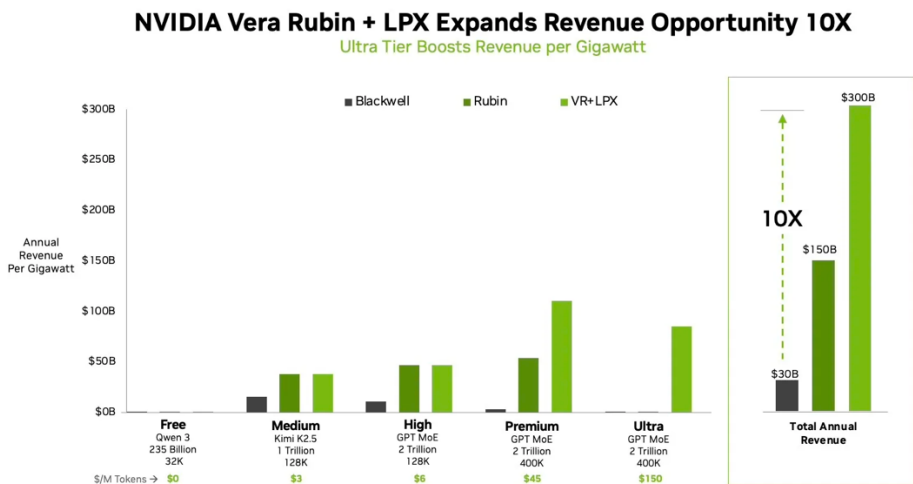
来源：英伟达官方科技博客，国金证券研究所

Vera Rubin 正式步入量产部署阶段。7 月 21 日，英伟达在官网博客官宣 Vera Rubin 正在全球加速部署，并将平台明确概括为跨七颗芯片与五类机架托盘的极致协同设计：Vera Rubin NVL72 计算机柜、Vera CPU 机柜、Groq 3 LPX 机柜、Spectrum-6 SPX 网络机柜与 Vera BlueField-4 STX 存储机柜，作为单一系统进行工程化设计而非独立产品组装——GPU 计算、CPU、低时延推理、网络与存储五类机柜的协同框架由此获得官方确认。英伟达要解决的不是单颗 GPU 性能问题，而是 AI 工厂内部计算、存储、网络、调度与并发之间的系统协同问题，对产业链而言，系统化意味着出货弹性从 GPU 单点扩展到五类机柜的全部构成环节。

实测性能大幅提升，AI Factory 已来。据官宣，Vera Rubin NVL72 机柜已在 CoreWeave、Google Cloud、Microsoft Azure 与 Oracle 云基础设施运行，供应链覆盖 30 个国家的 350 多个工厂站点，是有史以来规模最大、最成熟的机柜级供应链；CoreWeave 在 DeepSeek-R1 上的首个实测基准显示，每兆瓦 token 吞吐量达 Grace Blackwell NVL72 的 10 倍，直接命中电力受限时代 AI 工厂最关键的指标——每瓦性能与最低 token 成本。



图表4: NVIDIA Vera Rubin NVL72 和 LPX 提供了 10 倍的收入增长机会



来源: 英伟达官方科技博客, 国金证券研究所

1.2 计算与互联: HBM4、FP4 引擎与 NVLink 6 三箭齐发

单芯片与单机柜的规格升级为系统协同提供硬件底座。Rubin GPU 采用双芯粒设计,集成 3360 亿晶体管,搭载 288GB HBM4、内存带宽达 22TB/s,较 Blackwell 提升约 2.8 倍——对多 Agent 并发与超长上下文推理,内存带宽已是决定体验与效率的关键变量。第三代 Transformer 引擎通过自适应压缩让更多训练与推理任务使用 NVFP4 精度,单芯片 NVFP4 推理算力 50 PFLOPS、训练 35 PFLOPS,分别为 Blackwell 的 5 倍与 3.5 倍,本质是单位 token 产出效率提升与单位 token 成本下降。NVLink 6 将 GPU 间双向带宽从 1.8TB/s 翻倍至 3.6TB/s,对 MoE 多专家模型中 token 跨 GPU 流转尤为关键。机柜层面, Vera Rubin NVL144 由 72 颗 Rubin GPU 与 36 颗 Vera CPU 构成,合计 3.6 EFLOPS NVFP4 推理与 2.5 EFLOPS 训练算力、20.7TB HBM4 容量、1.6PB/s 聚合内存带宽与 260TB/s 纵向扩展带宽。Vera CPU 配备 88 颗自研 Olympus 核心、176 线程,通过第二代 NVLink-C2C 提供 1.8TB/s 的 CPU-GPU 一致性带宽,使 LPDDR5X 与 HBM4 构成统一内存池,直接支撑 KV Cache 卸载与多模型执行。

供给端看, HBM 是 Rubin 量产爬坡中供需最紧的环节,进入 Q3 全面量产窗口。黄仁勋 6 月 5 日在首尔确认三星电子、SK 海力士、美光三家 HBM4 供应商全部通过认证并处于量产状态;据韩国证券业分析口径, HBM4 的全面规模化量产于 2026 年第三季度启动。SK 海力士 2026 年第一季度营收 52.58 万亿韩元、经营利润 37.61 万亿韩元、经营利润率达 72%, 季度营收首次突破 50 万亿韩元; 公司 HBM 销售负责人金基泰表示, HBM4 与主要客户从早期阶段即联合开发,未来三年客户需求已超过供给能力。Rubin 的量产爬坡与 HBM4 的 Q3 全面量产互为因果,供给端的紧绷恰恰印证本篇的核心判断: 算力仍缺。

图表5: NVIDIA Vera Rubin 集成 6 芯片的平台设计



图表6: SK 海力士 1Q26 业绩数据

	1Q26	4Q25	QoQ	1Q25	YoY
Revenue	52576.3	32826.7	60%	17639.1	198%
Operating Profit	37610.3	19169.6	96%	7440.5	405%
Operating Margin	72%	58%	14%P	42%	30%P
Net Income	40345.9	15246	165%	8108.2	398%

来源: Video Cardez, 国金证券研究所

来源: PR Newswire, SK 海力士公司公告, 国金证券研究所

1.3 存储机柜: G3.5 层与 CMX 平台, KV Cache 成为新的硬件需求

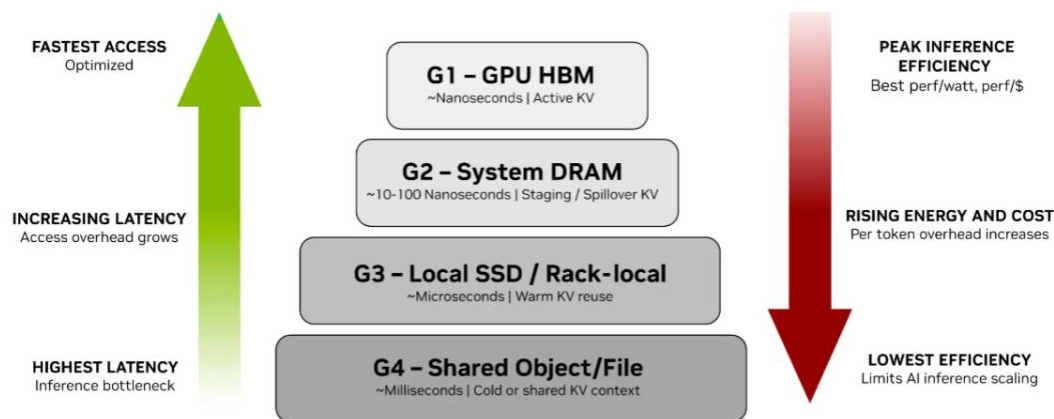
AI 工厂将存储按调用热度分层, G3.5 是本代际新增的关键一层。AI 推理系统的存储层级按热度分为 G1 至 G4: 正在生成 token 时立即使用的数据放在 GPU HBM (G1), 很快使用的数据放在 CPU DRAM (G2), 短期温数据放在本地 SSD (G3), 长期冷数据进入共享存储 (G4)。



Agent 与超长推理带来巨量 KV Cache（上下文缓存），其直接关系模型记忆能力，但 HBM 与 DRAM 容量有限、本地 SSD 难以跨节点共享、共享存储延迟偏高。英伟达官方据此定义 G3.5 层：一个以太网直连、专为 KV Cache 优化的闪存层，作为 AI 基础设施 Pod 的 Agent 长期记忆，容量足以同时承载众多 Agent 的共享上下文，又足够贴近计算侧、可在不阻塞解码的前提下将上下文频繁预取回 GPU 与主机内存。

CMX 平台将 G3.5 从概念落为机柜级产品，存储产业链获得全新增量。英伟达于 2026 年 1 月发布 CMX 上下文内存存储平台（原名 ICMS），由 BlueField-4 DPU 管理、通过 Dynamo 推理框架与 NIXL 传输库编排上下文在各存储层间的流动；据 TrendForce，CMX 围绕 BlueField-4 STX 机柜构建，单柜以 64 颗 BlueField-4 DPU 管理约 9600TB 闪存容量。英伟达官方口径下，相较传统存储方案，该平台通过可靠承接与预取 KV Cache，最高可带来 5 倍的 token 生成速度与 5 倍的能效提升。黄仁勋在 6 月 GTC Taipei 明确表示，AI 的记忆系统将使存储系统被彻底革新。对产业链而言，KV Cache 从软件问题升格为硬件需求：SSD、DPU、以太网交换与存储机柜组装环节获得独立于 GPU 机柜之外的第二类机柜增量。

图表7：英伟达 Dynamo 框架下 KV Cache 在 G1 至 G4 层间的卸载序列



来源：Expon Tech，国金证券研究所

1.4 异构与网络：CPU 与 LPX 各司其职，Spectrum-6 将 CPO 确立为官方标配

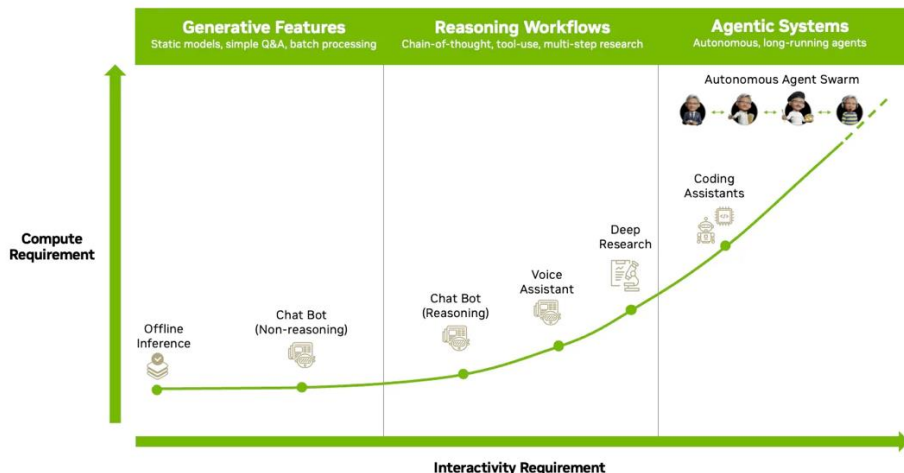
异构协同覆盖 Agent 与低时延推理两类新兴负载。Vera CPU 专为数据搬运与 Agent 推理设计——Agent 需要大量调用工具、执行代码并运行沙箱，这类任务并不适合由 GPU 承担，CPU 机柜的价值在于承接非 GPU 型任务、让 GPU 资源集中于高强度并行计算。Groq 3 LPX 作为平台第七颗芯片，定位低时延推理加速器，服务语音交互、实时翻译、代码助手等对 token 输出时延敏感的场景，与 GPU 形成 prefill 与 decode 阶段的分工协同，并计划于第三季度出货。异构编排同时是英伟达应对 ASIC 竞争的系统性 answer：其护城河已从 GPU 与 CUDA 生态，演进为定义 AI 工厂、组织全球供应链围绕其路线协同投入的系统号召力。

Spectrum-6 官方采用 CPO，光互连从验证叙事升格为出货叙事。英伟达官方明确，Spectrum-6 以太网交换机采用共封装光学（CPO）实现横向扩展连接，以提升能效与可靠性——CPO 将硅光引擎与交换芯片封装在一起，缩短走线、降低电信号损耗与功耗。7 月 21 日官宣进一步确认，采用 CPO 的 NVIDIA Photonics 横向扩展交换机是业界首款进入量产阶段的此类产品，相比可插拔收发器功耗降低 5 倍、平均无故障时间（MTBI）提升 10 倍，CoreWeave、Lambda 与 Oracle 云为首批采用者；Spectrum-6 交换机的首批引入者还包括微软、SpaceX AI 与特斯拉。延续我们光通信系列报告的框架：可插拔 1.6T 光模块仍是 2026 年下半年放量主力，而 CPO 随 Spectrum-6 进入旗舰平台标配清单，其规模化节点较此前预期更为清晰，网络机柜（含交换芯片、CPO 光引擎、FAU 与 MT/MPO 精密连接器件）成为 Pod 五类机柜中弹性最陡的一类。从实质性的 PCB 投产，到存储、网络、计算五类机柜的落地出货，Pod 体系对第三季度产业链业绩的带动是全方位的。



图表8：不同 AI 工作负载的相对计算和交互需求，包含聊天机器人、语音助手、深度研究、代码助手等，愈发强调代理群集对高吞吐量和低延迟的需求

The Era of Agentic Scaling Demands More Multi-Agent Applications Require Low-Latency & High-Throughput



来源：英伟达官方技术博客，国金证券研究所

二、谷歌财报更新：TPU 系统销售启动交付，算力供给依然紧张

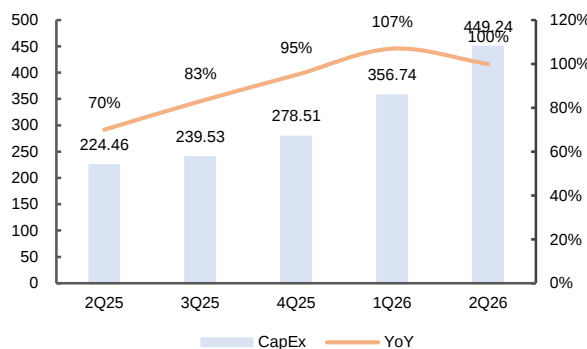
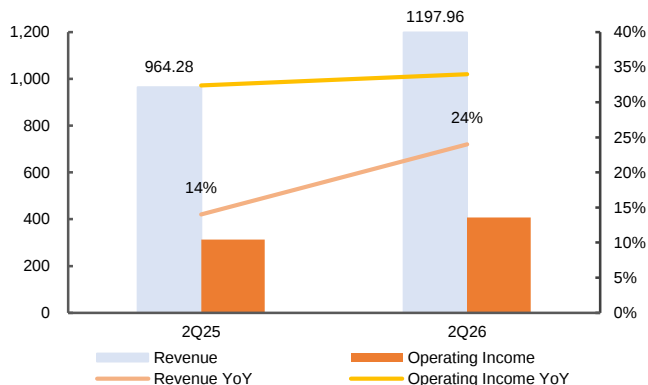
2.1 Q2 财报全面超预期，资本开支指引再度上调

云业务的增速与订单储备印证 AI 需求的强度。Alphabet 于 7 月 22 日盘后发布 2026 年第二季度财报：收入 1198 亿美元、同比增长 24%；经营利润 408 亿美元、同比增长 30%。Google Cloud 收入 248 亿美元、同比增长 82%，云业务积压订单达 5140 亿美元、环比增加超过 500 亿美元。第二季度资本开支 449 亿美元，管理层披露其中约六成投向服务器、四成投向数据中心与网络设备；当季自由现金流为负 59 亿美元，反映投入强度。管理层将 2026 年全年资本开支指引从 1800 亿至 1900 亿美元上调至 1950 亿至 2050 亿美元，上调主因是产能交付加速以满足增长的需求，并预告 2027 年资本开支将较 2026 年显著增长。

供给紧张贯穿管理层表态，第三方产能过桥凸显算力缺口。管理层表示，在供给受限环境下，计划于第三季度扩大第三方产能的使用作为过桥策略，在建设内部产能的同时继续扩大客户群，短期将带来温和的利润率压力。产能分配次序上，基础算力优先保障 AGI 前沿研发，其次是搜索与 YouTube 等核心产品，再次是云服务。此前 6 月，谷歌与 SpaceX 签署每月 9.2 亿美元的 AI 算力租用协议。作为四大云厂商中最早披露季报的一家，Alphabet 的加码直接对冲了市场对 AI 资本开支持续性的担忧——说到底，现在还是缺算力。

图表9：Alphabet 2Q26 收入 1198 亿美元，同比增长 24%

图表10：Alphabet 2Q26 资本开支 449 亿美元



来源：Alphabet 官网，国金证券研究所

来源：Alphabet 官网，国金证券研究所

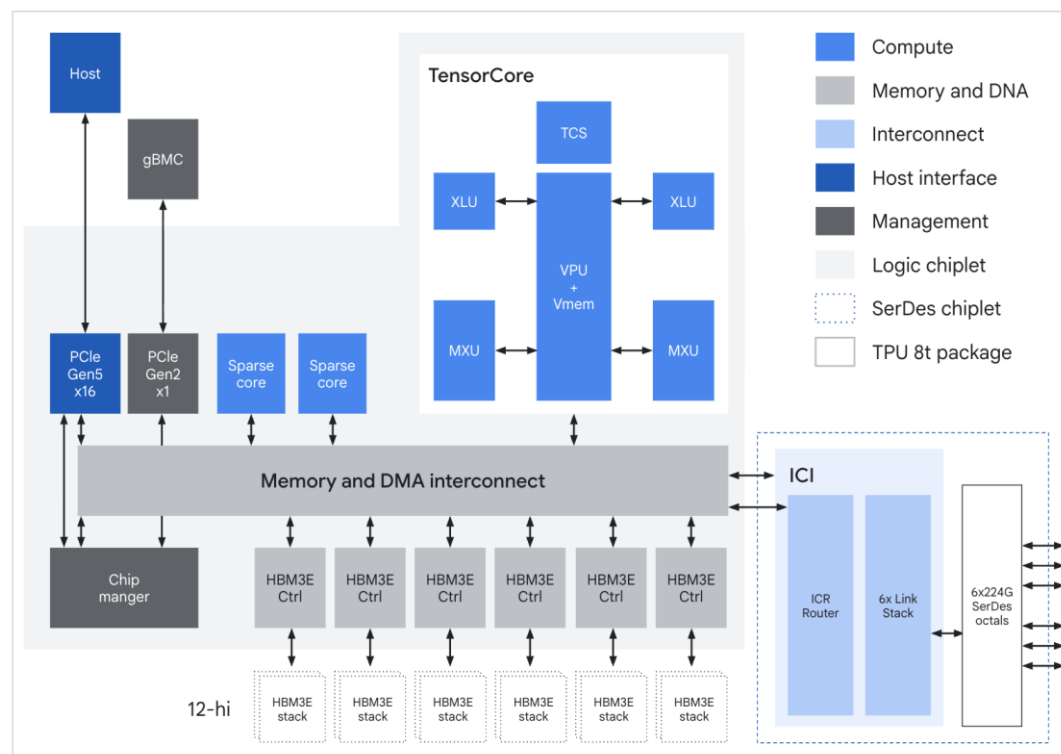
2.2 TPU 系统销售 Q2 启动交付，2027 年进入收入放量期

TPU 商业化从租用走向外销，第二增长曲线正式启动。管理层确认，TPU 系统已于第二季度



开始向客户数据中心交付——这是谷歌首次将自研芯片以硬件形态部署到客户侧，打开云租用之外的可寻址市场。CFO 在电话会上解释收入确认节奏：TPU 系统销售协议签署后计入云业务积压订单，公司先行备货建库存，对当期经营现金流形成影响，交付启动后开始确认收入；2026 年确认的收入占协议总额的比例较小，将在 2026 年退出时持续爬坡，绝大部分收入于 2027 年确认。结合上一篇的梳理，TPU v8 双芯片（8t 训练、8i 推理）已于 4 月发布并将于年内向谷歌云客户开放，Ironwood 持续承担放量任务，DIGITIMES Research 预计 2026 年谷歌 TPU 出货量达 332.5 万颗、稳居 ASIC 阵营第一。TPU 加速出货构成本篇标题所指的第二重加速，机柜组装、光互连（含谷歌自研 OCS 全光交换）、液冷与供电链条同步受益。

图11: TPU 8t ASIC block diagram



来源：Google Cloud，国金证券研究所

三、产业链主线延续：机柜、PCB 与光模块跟随代际切换

主流大板块的代际切换逻辑不变，Q3 进入兑现期。机柜环节，工业富联作为 Rubin 整机柜首发制造商，Rubin 平台自 6 月起小批量出货客户测试、Q3 启动量产，液冷冷板、管路、结构件、电源等核心零部件自供比例较 GB 系列进一步提升。PCB 环节，Vera Rubin NVL144 以中央 PCB 中板取代传统线缆连接，机柜内互连从线缆方案向大尺寸高层数 PCB 方案迁移属于从无到有的增量价值，叠加计算托盘、交换托盘与模组板的高多层、HDI 用量，单机柜 PCB 价值量较 GB 代际系统性抬升；展望 Rubin Ultra（2027 年）整机柜正交背板方案，PCB 价值量曲线有望进一步推陡，延续斜率之王叙事。光模块环节，ConnectX-9 将端口带宽提升至 1.6Tb/s，800G 向 1.6T 的切换斜率决定下半年至 2027 年板块业绩弹性；本篇第一章的增量在于 CPO 随 Spectrum-6 确立为官方标配，光器件与连接器环节（FAU、MT/MPO 插芯、高速连接器）跟随 1.6T 与 CPO 双线演进的确定性进一步增强。

四、功耗主线延续：液冷与供电确定性不变，电容迎来涨价落地

4.1 液冷与 800VDC：官方口径的确定性方向

散热与供电重构是 Rubin 代际的官方确定项。英伟达官方口径下，Vera Rubin NVL144 计算托盘为 100%液冷模块化设计、机柜采用 45 摄氏度液冷并配备液冷母排，机柜储能较上代提升 20 倍；7 月 21 日官宣补充了两个工程口径：托盘内无线缆、无风扇、无软管，计算托盘装配时间从数小时缩短至一分钟；45 摄氏度液冷入口温度设计可实现无冷水机的干冷器运行，新建 AI 工厂每兆瓦每年可节省数百万加仑水资源。800VDC 高压直流已确立为面向 Rubin Ultra Kyber 机柜体系的参考架构。单机柜功耗沿 Hopper 约 40 千瓦、GB300 约 120 至 130 千瓦、VR200 约 190 至 230 千瓦、Kyber 约 600 千瓦的路径逐代跃升，液冷渗透率与单柜价



值量双升、供电架构从 power shelf 向独立 power rack 形态演进的逻辑此前已充分论述，本篇不再展开。

4.2 电容：三星电机涨价从信号到落地，供需紧张有望延续至 2027 年上半年

头部厂商的涨价动作为电容环节提供最直接的景气验证。涨价信号自年初逐级强化：2 月下旬，据韩国媒体 The Elec 报道，三星电机考虑上调 MLCC 价格，公司发言人表示在 AI 与电动车需求带动下、预计第二季度 MLCC 平均售价上行，行业龙头村田亦释放 3 月涨价信号；据 TrendForce 援引工商时报，三星电机计划自 4 月起对 MLCC 实施涨价、幅度预计达两位数百分比；5 月，据 DIGITIMES 报道，公司再酝酿 5% 至 10% 的追加调涨。价格端已实质性反应：据 36 氪援引产业信息，2 月底以来 MLCC 现货价格普涨 15% 至 20%，AI 服务器用高容值产品涨幅达 50% 至 60%、部分紧缺型号价格更高。订单端同步落地：7 月初，据韩媒报道，三星电机与一家全球大型科技企业签署 4540 亿韩元（约 2.9 亿美元）的 AI 服务器 MLCC 长期供货合同。需求端的量级同样清晰：据 TrendForce 转引每日经济新闻的分析师测算，英伟达 VR200 NVL72 服务器单机将使用约 60 万颗 MLCC、较现有 GB300 平台高出逾 30%，且服务器客户高度依赖村田、三星电机等一线供应商；村田社长中岛规巨亦公开表示，云大厂驱动的数据中心投资预计将持续三至五年，公司 MLCC 客户询盘量约为当前供给能力的两倍。

供给刚性决定本轮涨价的持续性。高端高容 MLCC 产线已满载，扩产周期长达 1.5 至 2 年：三星电机天津工厂改造的高容产能预计 2026 年第四季度起逐步释放、菲律宾扩产项目 2027 年完成，太阳诱电韩国服务器产能亦要到 2027 年落地——2027 年上半年之前，高端高容 MLCC 的供需缺口难以根本缓解。需要客观提示的是，据 TrendForce，本轮紧张集中于 AI 服务器等高端产线，消费级产能利用率仍偏低，市场呈现高端紧俏、标准品宽松的分化结构，涨价弹性应聚焦高容高可靠产品线而非全品类线性外推。结合上一篇的分层框架：机柜储能层（大容量 MLCC）与增量最确定、板级去耦层（MLCC）用量随功率密度刚性增长，本轮涨价正落在后者，量价齐升的弹性开始进入报表可见区间。

五、相关标的

- 1) 超核心供应商：工业富联、胜宏科技
- 2) 海外算力：中际旭创、新易盛、东山精密、江海股份、中钨高新、蓝思科技、东阳光、光智科技、先导基电、火炬电子、三环集团、欧科亿、天孚通信、鼎泰高科、领益智造、兆易创新、鹏鼎控股、唯科科技、天岳先进、大普微、源杰科技、麦格米特、景旺电子、英维克、京东方等；英特尔、SK 海力士、Lumentum、闪迪、铠侠、美光。
- 3) 国内算力：寒武纪、海光信息、东阳光、中芯国际、华虹半导体、禾盛新材、摩尔线程、天数智芯、沐曦股份、壁仞科技、协创数据、华丰科技、杰华特、京基智农、华达科技、中科曙光、利通电子、浪潮信息、利扬芯片、胜蓝股份、华勤技术、国科微、中国长城、晶科科技、罗曼股份、盈峰环境、芯原股份、亿田智能、豫能控股、星环科技、鸿日达、盛视科技、神州数码、润泽科技、大位科技、润建股份、奥飞数据、瑞晟智能、科华数据、潍柴重机、欧陆通、杰创智能、奥尼电子。
- 4) 大模型&云：智谱、MiniMax、阿里巴巴、腾讯控股、金山云、百度集团、优刻得、首都在线、网宿科技、云赛智联、青云科技等。

六、风险提示

■ Rubin 量产爬坡不及预期的风险

Pod 体系涉及五类机柜与七颗芯片的协同交付，任一环节的良率或配套瓶颈都可能拖累整体出货节奏。



■ TPU 系统销售收入确认节奏的风险

TPU 外销收入绝大部分于 2027 年确认，若交付或客户建设进度慢于预期，产业链拉货节奏可能相应后移。

■ HBM 与高端元器件供给瓶颈反噬的风险

HBM4、高容 MLCC 等环节供给紧张，若缺口超预期扩大，可能反过来制约整机出货并推高整机成本。



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	北京	深圳
电话：021-80234211	电话：010-85950438	电话：0755-86695353
邮箱：researchsh@gjzq.com.cn	邮箱：researchbj@gjzq.com.cn	邮箱：researchsz@gjzq.com.cn
邮编：201204	邮编：100005	邮编：518000
地址：上海浦东新区芳甸路 1088 号 紫竹国际大厦 5 楼	地址：北京市东城区建国内大街 26 号 新闻大厦 8 层南侧	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