

阿里 Q3 验证 AI 需求高景气，DeepSeek 与谷歌 TPU 引领软硬件进阶

增持（维持）

投资要点

- 本周 AI 产业链股价普遍反弹，在经历“AI 泡沫论”等利空因素回调后，阿里三季度财报表现强劲再次验证 AI 需求仍将持续。本周股价涨幅情况来看，数通 PCB/CCL：生益电子+14.16%，胜宏科技+12.93%，生益科技+8.6%；铜连接：博创科技+20.84%，兆龙互连+7.23%；光芯片/光器件：长光华芯+59.33%，太辰光+22.3%；液冷：思泉新材+12.72%，申菱环境+5.92%；服务器代工：华勤技术+3.49%。**
- 国产算力呈现出“商业兑现加速”与“技术范式突破”共振的强劲势头。**一方面，基础设施端迎来需求井喷，阿里巴巴 Q3 财报中阿里云收入同比高增 34%，充分验证了国内 AI 算力需求的爆发性与持续性，且公司表示服务器交付速度滞后于订单增长的现状，预示着国内云厂商资本开支有望进入新一轮加速期，进一步夯实国产算力底座。另一方面，算法端实现了从“结果拟合”向“逻辑自洽”的关键跃迁，DeepSeekMath-V2 通过引入基于形式化证明的自我验证机制，将大模型竞赛从单一的参数规模堆砌引向了“过程监督”的高阶维度，成功验证了推理端算力 (Test-Time Compute) 带来的巨大边际收益。在这一背景下，我们坚定看好国产算力产业链，认为其已跨越概念导入期，正式进入由技术迭代与产能扩张双轮驱动的业绩兑现黄金周期，具备核心交付能力的算力底座厂商将持续受益于这场从底层架构到顶层应用的全面繁荣。
- 谷歌硬件生态正迎来“技术闭环”向“商业外溢”的关键转折点，重塑 AI 算力供应链格局。**技术层面，新一代 Ironwood TPU 凭借“9216 芯片 +1.77PB 内存”的超大规模集群架构，配合自研 MEMS OCS 光交换技术，彻底打破了传统电交换在能耗与延迟上的物理瓶颈，以极低的网络成本（占比<5%）实现了万卡级集群的高效互连。这一基于“光电协同”的 TCO 壁垒已转化为强大的商业虹吸效应，接连斩获 Anthropic 百万级算力订单及 Meta 潜在部署意向，有力证明了 TPU 已具备比肩英伟达的竞争力。我们认为，随着谷歌 TPU 从自用走向外售，“谷歌链”有望成为独立于英伟达之外的第二增长极，建议重点关注深度受益于 OCS 架构的 MEMS 光器件、光模块、ASIC 设计服务及液冷散热等核心供应链环节的弹性机会。
- 产业链相关公司：**
 - PCB/CCL：**胜宏科技、沪电股份、深南电路、景旺电子、方正科技、生益科技、南亚新材、生益电子、东山精密、威尔高等；
 - 铜缆/铜连接：**沃尔核材、兆龙互连、华丰科技、立讯精密、鸿腾精密等；
 - 光芯片/光器件：**长芯博创、源杰科技、仕佳光子、太辰光、长光华芯等
 - 液冷：**英维克、思泉新材、申菱环境、高澜股份等。
 - 服务器代工：**工业富联、华勤技术。
- 风险提示：**供应链波动风险，下游需求不及预期，行业竞争加剧。

2025 年 11 月 30 日

证券分析师 陈海进

执业证书：S0600525020001

chenhj@dwzq.com.cn

研究助理 解承堯

执业证书：S0600125020001

xiechy@dwzq.com.cn

行业走势



相关研究

《三季度 AI 业绩持续兑现，Mid-Training 开启结构化智能新阶段-算力周报》

2025-10-26

《国产算力认知强化！Tokens 消耗——AI 需求侧核心逻辑正式向多模态大模型延展》

2025-10-08

内容目录

1. 阿里资本开支有望提速，模型迈向“逻辑自洽”新范式	4
1.1. 阿里巴巴 Q3 业绩：阿里云收入同增 34%，资本开支有望加速	4
1.2. DeepSeekMath-V2：从“结果拟合”到“逻辑自洽”的范式跃迁	4
2. Ironwood 实现万卡级集群扩展，自研 OCS 引领互连革命	5
2.1. 谷歌 Ironwood：9216 颗芯片全互连共享 1.77PB 内存	5
2.2. 谷歌 OCS 交换机：MEMS 技术具备低功耗、低成本优势，重塑数据中心互联架构	7
3. 风险提示	10

图表目录

图 1: IMO-ProofBench 基准测试基础子集与进阶子集的专家评估结果	5
图 2: TPUv4: 4096 颗芯片通过 OCS 共享内存	5
图 3: Ironwood: 9216 颗芯片通过 OCS 共享内存	6
图 4: Ironwood 芯片架构	6
图 5: Ironwood 机架	7
图 6: 网络架构变化: a) 传统网络架构, b) 谷歌 Apollo 基于 OCS 网络架构	7
图 7: 不同 OCS 技术路径对比	8
图 8: 谷歌 Palomar OCS 交换机架构	8
图 9: 每个维度通过 16 台 OCS 交换机连接	9

1. 阿里资本开支有望提速，模型迈向“逻辑自洽”新范式

1.1. 阿里巴巴 Q3 业绩：阿里云收入同增 34%，资本开支有望加速

阿里巴巴与 11 月 25 日发布三季报，本季度阿里云收入同比实现 34% 的强劲增长，外部客户收入增速加快至 29%，主要驱动力来自于 AI 的持续爆发性需求和公共云使用量的激增。公司的旗舰模型 Qwen2.5-Max 在真实世界编码任务、智能体工具使用能力等专项评估基准中位列全球前茅。同时，阿里巴巴在“AI+云”战略上的持续投入，进一步巩固了其作为市场领导者的地位，并为加速期奠定了技术基础。公司持续加强从高性能 AI 基础设施到基础模型的全栈 AI 能力，将其视为决定性的竞争优势。阿里云不仅是中国 AI 云市场的明确领导者，市场份额大于第二至第四大供应商的总和，还在混合云和金融云等细分市场保持超越行业的增速。这种市场地位和技术领先性，通过与 NBA、万豪中国等企业在 AI 计划上的合作加速落地，并延伸至消费者侧-Qwen App 在公开测试首周下载量即突破千万，标志着企业 AI 和消费者 AI 双轮驱动的全面启动。

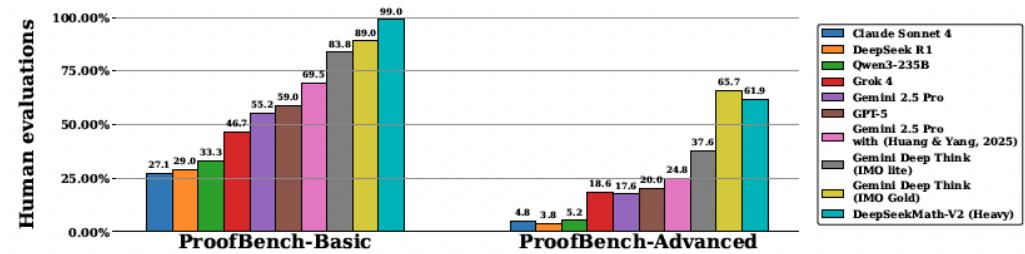
资本开支方面，公司提到的 3800 亿资本开支实际上是一个为期三年的规划数字，但公司服务器上架速度仍然远远跟不上客户订单的增长速度，不排除进一步增加投资以满足客户需求。

1.2. DeepSeekMath-V2：从“结果拟合”到“逻辑自洽”的范式跃迁

从“结果拟合”向“过程严谨性”的范式跃迁，DeepSeekMath-V2 标志着人工智能在复杂逻辑推理领域实现了关键的技术跨越。与传统大模型仅依赖“结果监督”（即只要最终答案正确即可，忽略过程中的逻辑漏洞）不同，该模型引入了基于 Lean 4 形式化证明语言的自我验证机制（Self-Verification），将训练重心升级为全链路的“过程监督”。通过构建高精度的验证器，模型能够在推理过程中自主且严谨地核验每一个逻辑步骤的有效性。这一机制从根本上抑制了现有大模型普遍存在的“逻辑幻觉”问题，确保了推理路径的数学严密性，使其在处理高阶定理证明等容错率极低的复杂任务时展现出显著的性能优势。

DeepSeekMath-V2 成功验证了测试时算力（Test-Time Compute）的高效扩展路径。这意味着通过在推理端增加计算投入（即允许模型进行多路径探索与自主纠错），可以大幅突破模型参数规模带来的性能天花板，获得比单纯堆砌训练算力更高的边际收益。这种具备“自主逻辑闭环”能力的架构，不仅巩固了 DeepSeek 在通用人工智能核心推理赛道上的技术护城河，也为 AI 从简单的辅助工具迈向独立解决未知科学难题奠定了坚实的技术基石。

图1: IMO-ProofBench 基准测试基础子集与进阶子集的专家评估结果



数据来源：《DeepSeekMath-V2: Towards Self-Verifiable Mathematical Reasoning》，东吴证券研究所

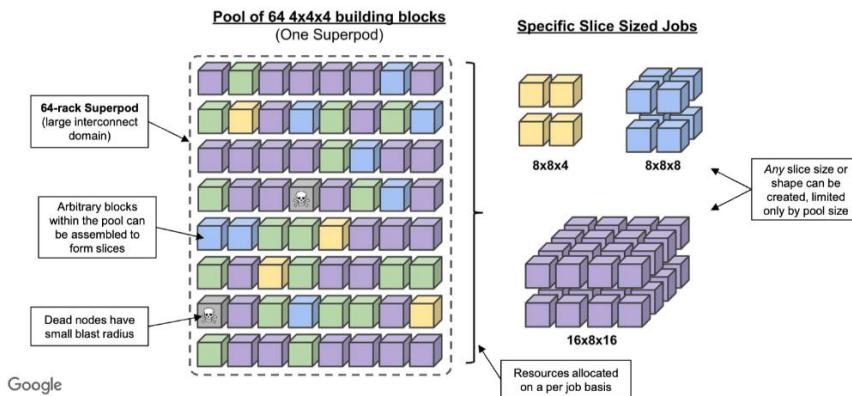
2. Ironwood 实现万卡级集群扩展，自研 OCS 引领互连革命

2.1. 谷歌 Ironwood: 9216 颗芯片全互连共享 1.77PB 内存

在此前 Hot Chips 2025 大会中，谷歌介绍了其新一代 TPU，代号 Ironwood，专为大模型 AI 推理设计。其突破性的创新包括：1) 单 SuperPod 节点最多容纳 9216 颗芯片，使用 OCS 交换机共享内存，相比于 TPUv4 的 4096 颗芯片，Ironwood 每 Pod 芯片数量翻倍；2) 9216 颗芯片总共共享 1.77PB 内存，在 FP8 精度下，单 SuperPod 性能算力可达 42.5EFLOPS，其每瓦性能是上一代 TPU Trillium 的 2 倍；3) 具备可靠性、可用性以及可服务性，采用第三代液冷基础设施。

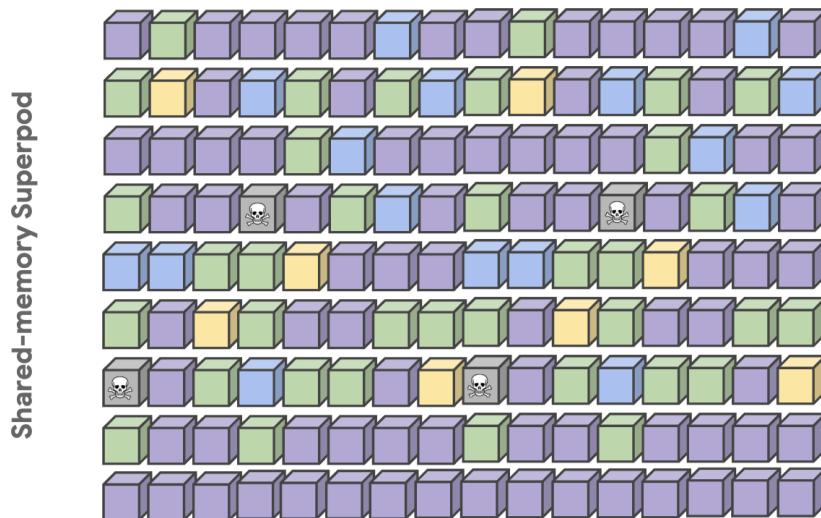
谷歌 Ironwood 互连架构与 TPUv4 类似，其中每 Cube 代表单个机架，单机架由 64 颗 TPU 组成，按 $4 \times 4 \times 4$ 网格排列。在 TPUv4 中搭配 64 个 Cube 形成单台超级计算机 (SuperPod)，总计 4096 颗 TPU 芯片；在 Ironwood 中单 SuperPod 共有 144 个 Cube，总计 9216 颗 TPU 芯片。每 Cube 内芯片使用 ICI 直连，不同 Cube 中芯片通过 OCS 交换机进行 ICI 直连。

图2: TPUv4: 4096 颗芯片通过 OCS 共享内存



数据来源：Hot Chips 2025，东吴证券研究所

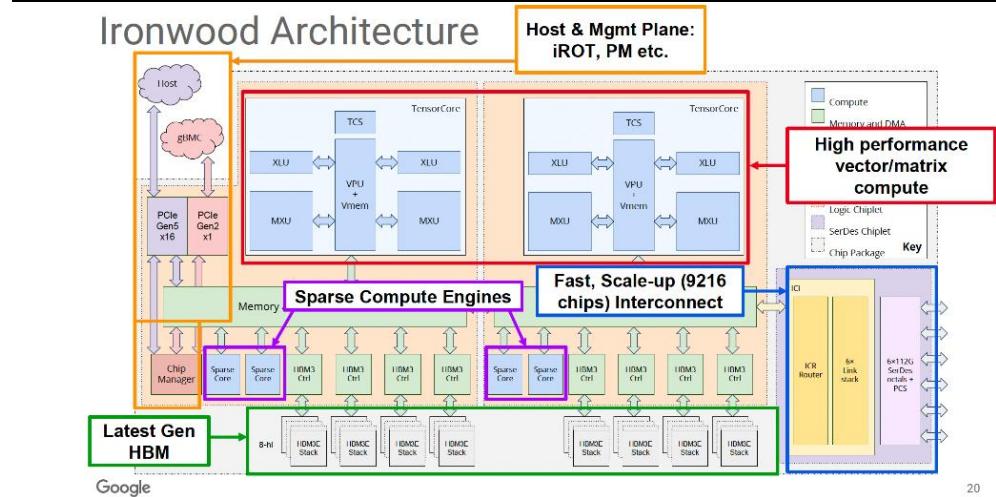
图3: Ironwood: 9216 颗芯片通过 OCS 共享内存



数据来源: Hot Chips 2025, 东吴证券研究所

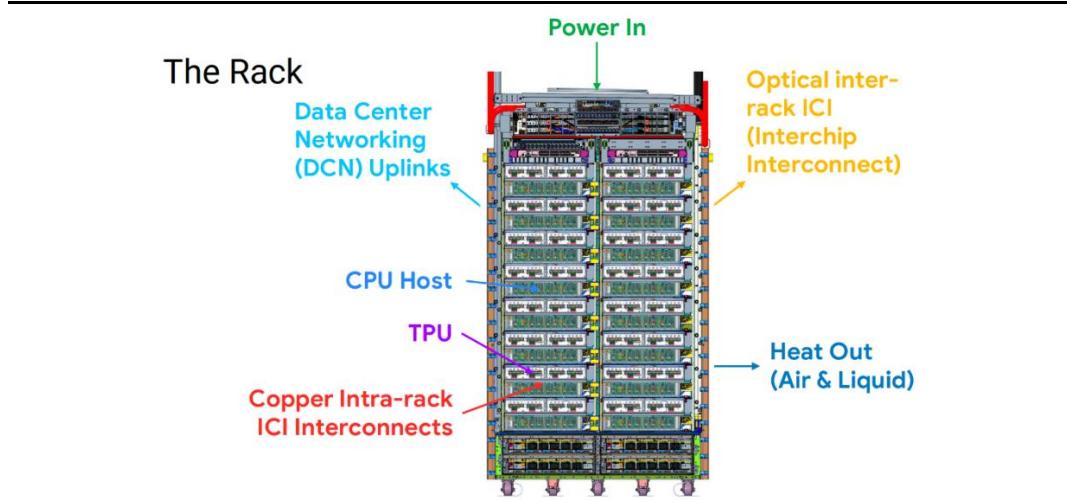
Ironwood 采用双 die TPU 设计, 使用 8 层 HBM3e 内存, 提供 192GB 容量和 7.3TB/s 的带宽。每个 Ironwood Tray 包含 4 颗 TPU, 采用液冷设计, 每机架包含 16 个计算托盘, 总共 64 颗 TPU, 同时与 16 个 CPU 主机机架连接。机架内所有互连均采用铜缆, OCS 提供与其他机架的连接。

图4: Ironwood 芯片架构



数据来源: Hot Chips 2025, 东吴证券研究所

图5: Ironwood 机架

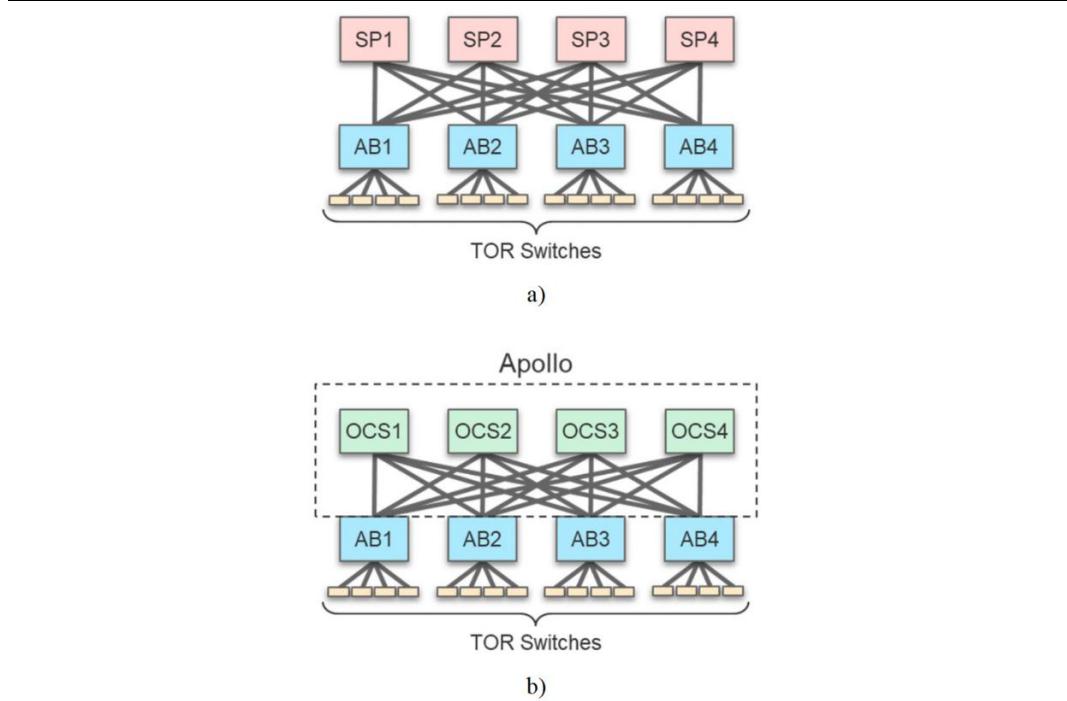


数据来源: Hot Chips 2025, 东吴证券研究所

2.2. 谷歌 OCS 交换机: MEMS 技术具备低功耗、低成本优势, 重塑数据中心互联架构

谷歌通过引入自研的光电路交换系统 (Apollo) 实现了数据中心网络架构的重要演进, 以光交换层替代传统电分组交换核心, 显著降低了网络成本、功耗与延迟, 同时凭借光技术的协议透明性和速率无关性, 支持灵活拓扑重构与双向全双工通信, 为机器学习训练等带宽密集型应用提供了可扩展、高能效的动态互联基础。

图6: 网络架构变化: a) 传统网络架构, b) 谷歌 Apollo 基于 OCS 网络架构



数据来源: 《Mission Apollo: Landing Optical Circuit Switching at Datacenter Scale》, 东吴证券研究所

OCS 技术路径主要包括 MEMS、压电驱动、机器人机械重配置、导波以及波长切换等多种方案，其中基于 MEMS 的方案，是实现高能效、大规模数据中心互联的关键技术。相较于其他光交换技术，MEMS 光交换机凭借其可扩展性（端口数超 1000x1000）和成本效益展现出巨大潜力。其核心优势在于：1) 与数据速率和波长无关，具备宽带无源特性，可支持多代光收发器技术而无需更换硬件；2) 功耗极低，因其无需逐包处理且维持反射镜位置能耗甚微，每比特能耗比电交换低数个量级；3) 延迟极低，仅由光速传播决定，避免了电交换带来的数十至数百纳秒的处理延迟，这对于机器学习等延迟敏感型应用至关重要。

图7：不同 OCS 技术路径对比

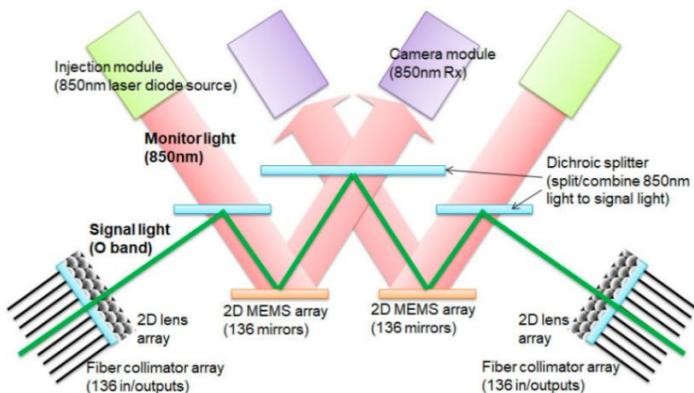
Technology	Relative Cost*	Port Count	Switching Time	Insertion Loss**	Driving Voltage (volts)	Latching
MEMS [25, 26]	Medium	320x320	ms	<3dB	100s	No
Robotic [27]	Medium	1008x1008	mins***	<1dB		Yes
Piezo [28]	High	384x384	ms	<2.5dB	10s	No
Guided Wave [29]	Low	16x16	ms	<6dB	1s	No
Wavelength Switching [30]	TBD	100x100	ns	<6dB	0	Yes

*Based on scale indicated. **Includes connector losses. ***Per connection

数据来源：《Mission Apollo: Landing Optical Circuit Switching at Datacenter Scale》，东吴证券研究所

以谷歌 TPUv4 的 Palomar OCS 交换机为例，通过 136x136 端口实现光路交换：1) 其光学核心采用两组二维 MEMS 反射镜阵列，通过精确驱动反射镜倾斜完成光束转向，实现任意端口间的无阻塞连接；2) 创新性地采用 850nm 监控光束与相机成像构成闭环控制系统，通过图像处理实时优化光路耦合；3) 采用“176 选 136”的冗余设计（每个 MEMS 芯片含 176 个反射镜，筛选最优 136 个使用），显著提升良率和可靠性；4) 整机最大功耗仅 108W，插入损耗<2dB，回波损耗<-38dB。

图8：谷歌 Palomar OCS 交换机架构



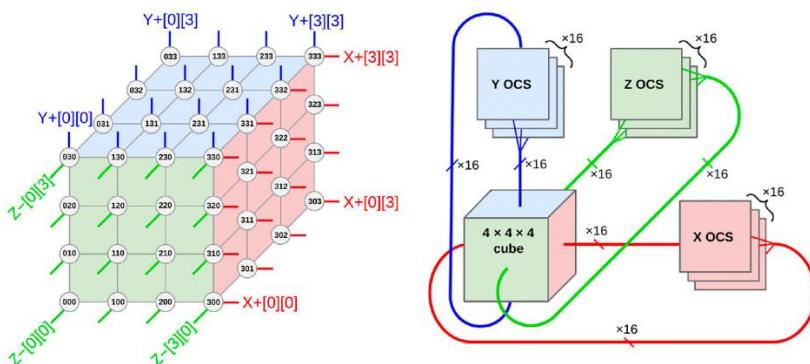
数据来源：《Mission Apollo: Landing Optical Circuit Switching at Datacenter Scale》，东吴证券研究所

谷歌TPUv4超级计算机通过OCS构建其芯片互连(ICI)网络。基本构建块是3D立方体单元，每个单元在X/Y/Z三个维度的每个面上暴露16条ICI链路，每链路单向带宽50GBps，每个维度由16台交换机连接，一个单元总计96条光学链路。一台完整的超级计算机由64个这样的立方体单元组成，通过6144条光学ICI链路连接到48台不同的OCS交换机，形成了一个大规模、高维度的可重构互连网络。

OCS的可重构性由配套软件基础设施管理，能够在每次任务启动时按需建立独特的交叉连接，从而实现极高的灵活性与资源利用率。其关键优势在于：1) 缓解资源碎片化：训练任务可以使用任何物理上不连续的立方体单元；2) 提升系统可用性与容错：硬件故障仅导致受影响的立方体单元被移除，而不中断整个集群的运行，实现了故障隔离与优雅降级。

成本方面，采用OCS方案实现了显著的成本优势，其OCS和光纤成本不到TPUv4 Pod总资本成本的5%，运行功耗小于总功耗的3%，远低于基于Infiniband等电交换机的扩展方案。在协议层面，ICI设计为可编程形式，支持按任务划分网络，为每个任务提供专属的、安全的网络切片，消除了网络共享带来的拥塞和复杂性，完美适配大规模机器学习训练的需求。

图9：每个维度通过16台OCS交换机连接



数据来源：Hot Chips 2025，东吴证券研究所

近期谷歌TPU商业化进程显著加速，先是于10月27日与AI独角兽Anthropic达成“里程碑式”协议，将向其提供多达100万个TPU算力支持；近日据The Information消息称Meta正洽谈在自家数据中心部署谷歌TPU，或成为其又一重量级外部客户。接连获得行业顶尖客户的青睐，有力证明了谷歌TPU正受到市场前所未有的重视。这也标志着经过多代迭代，谷歌TPU在性能表现及总拥有成本(TCO)控制上已具备了与英伟达芯片相媲美的核心竞争力，正成为科技巨头构建差异化算力优势、降低对单一供应商依赖的关键选择。我们认为，随着谷歌TPU从自用走向外售，“谷歌链”有望成为独立于英伟达之外的第二增长极，建议重点关注深度受益于OCS架构的MEMS光器件、光模块、ASIC设计服务及液冷散热等核心供应链环节的弹性机会。

3. 风险提示

- 1、供应链波动风险。**原材料供应（如核心零部件、关键材料）的价格上涨、短缺或交货周期延长，可能直接导致公司的生产成本上升、生产计划受阻，进而影响产品的交付能力和盈利水平。
- 2、下游需求不及预期。**数据中心行业受到宏观经济环境变化、产业政策调整或消费者偏好转移等因素影响，导致市场需求增长放缓甚至下滑，可能造成相关公司订单量减少、库存积压和营收增长承压。
- 3、行业竞争加剧。**现有竞争对手或潜在进入者通过降价促销、加大研发投入或扩展产能等方式，争夺市场份额，可能导致行业内各公司面临毛利率下降、品牌影响力被削弱、以及技术更新换代加速的压力，影响其长期竞争地位。

免责声明

东吴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。

本研究报告仅供东吴证券股份有限公司(以下简称“本公司”)的客户使用。本公司不会因接收人收到本报告而视其为客户。在任何情况下,本报告中的信息或所表述的意见并不构成对任何人的投资建议,本公司及作者不对任何人因使用本报告中的内容所导致的任何后果负任何责任。任何形式的分享证券投资收益或者分担证券投资损失的书面或口头承诺均为无效。

在法律许可的情况下,东吴证券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并进行交易,还可能为这些公司提供投资银行服务或其他服务。

市场有风险,投资需谨慎。本报告是基于本公司分析师认为可靠且已公开的信息,本公司力求但不保证这些信息的准确性和完整性,也不保证文中观点或陈述不会发生任何变更,在不同时期,本公司可发出与本报告所载资料、意见及推测不一致的报告。

本报告的版权归本公司所有,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。经授权刊载、转发本报告或者摘要的,应当注明出处为东吴证券研究所,并注明本报告发布人和发布日期,提示使用本报告的风险,且不得对本报告进行有悖原意的引用、删节和修改。未经授权或未按要求刊载、转发本报告的,应当承担相应的法律责任。本公司将保留向其追究法律责任的权利。

东吴证券投资评级标准

投资评级基于分析师对报告发布日后6至12个月内行业或公司回报潜力相对基准表现的预期(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,新三板基准指数为三板成指(针对协议转让标的)或三板做市指数(针对做市转让标的),北交所基准指数为北证50指数),具体如下:

公司投资评级:

- 买入: 预期未来6个月个股涨跌幅相对基准在15%以上;
- 增持: 预期未来6个月个股涨跌幅相对基准介于5%与15%之间;
- 中性: 预期未来6个月个股涨跌幅相对基准介于-5%与5%之间;
- 减持: 预期未来6个月个股涨跌幅相对基准介于-15%与-5%之间;
- 卖出: 预期未来6个月个股涨跌幅相对基准在-15%以下。

行业投资评级:

- 增持: 预期未来6个月内,行业指数相对强于基准5%以上;
- 中性: 预期未来6个月内,行业指数相对基准-5%与5%;
- 减持: 预期未来6个月内,行业指数相对弱于基准5%以上。

我们在此提醒您,不同证券研究机构采用不同的评级术语及评级标准。我们采用的是相对评级体系,表示投资的相对比重建议。投资者买入或者卖出证券的决定应当充分考虑自身特定状况,如具体投资目的、财务状况以及特定需求等,并完整理解和使用本报告内容,不应视本报告为做出投资决策的唯一因素。

 东吴证券研究所

苏州工业园区星阳街5号

邮政编码: 215021

传真: (0512) 62938527

公司网址: <http://www.dwzq.com.cn>