

# 电子行业深度报告

## AI 基建，光板铜电—光&铜篇 主流算力芯片 Scale up&out 方案全解析

增持（维持）

2025 年 12 月 27 日

证券分析师 陈海进

执业证书：S0600525020001  
chenhj@dwzq.com.cn

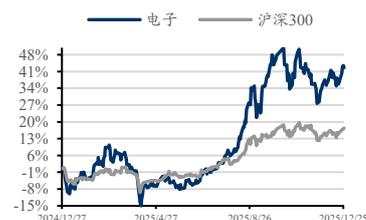
研究助理 解承堯

执业证书：S0600125020001  
xiechy@dwzq.com.cn

### 投资要点

- **英伟达 Rubin，集群互联高带宽低时延方向持续升级。** Scale up 端，依托 NVLink 6.0 与六代 NVSwitch，实现计算与交换侧 129.6TB/s 对称带宽，有效保障机柜级算力密集型场景下的参数同步效率；Scale out 端，基于胖树无阻塞拓扑，通过多层组网支撑超大规模 GPU 集群无阻塞互联，缩短传输路径、降低时延。根据我们测算，Rubin 若满配 CPX 芯片，三层组网下 GPU 光模块比例将达到 1:12。
- **谷歌 TPU 集群以高带宽互联与动态扩展为核心升级方向，构建 Scale up&out 组网能力。** Scale up 端，64 卡机柜内 TPU（Ironwood V7）采用 3D Torus 拓扑，通过 PCB 走线、铜缆/AOC&OCS 多链路互联；Scale out 端依托 DCN 分层架构，以 9216 ICIPOD 为基础模块，经 Tor/Leaf/Spine 交换机实现流量汇聚，最终通过 64 台 300\*300 端口 OCS 设备达成全局动态非阻塞互联，完成 147456 颗 TPU 集群组网。
- **亚马逊 Trainium3 组网突出“高密度互联+灵活扩展”。** Scale up 端依托 Scropio X 交换芯片与 PCIe6.0 协议，通过 AEC 铜缆实现三层互联（PCB/背板/跨机架），144 颗 Trainium3 经 NeuronLink4 端口达成高密度算力聚合；Scale out 端采用 ENA/EFA 双网分工，搭配高基数低速率交换机，以 Clos 拓扑构建无阻塞网络，交换机带宽升级可实现集群规模线性扩展，全方位适配大规模 AI 训练的算力与通信需求。
- **Meta 组网聚焦超大规模 AI 训练需求。** Scale up 端依托 Tomahawk5 交换芯片与 cable backplane（112G PAM4 铜缆背板），实现 16 颗 MTIA 与交换侧 204.8Tbps 的对称带宽互联；Scale out 端基于 DSF 解耦架构，以 Jericho3 交换芯片构建机柜级 RDSW、Ramon3 交换芯片组成 FDSW 分层组网，1:1 收敛比保障非阻塞传输，芯片与光模块比例可达 1:10。
- **光&铜观点：**2026 年商用 GPU 持续放量，也是 CSP ASIC 进入大规模部署的关键一年，数据中心 Scale up 催生超节点爆发，铜缆凭短距低耗低成本，成为柜内互联最优解；Scale out 带动集群持续扩容，光模块与 GPU 配比飙升，产品放量使得光芯片缺口凸显。一铜一光双线共振，互联需求迎来量价齐升，建议重点关注光铜两大核心赛道。
- **产业链相关公司：**  
光芯片：长光华芯、源杰科技、仕佳光子等  
铜缆：华丰科技、兆龙互连、沃尔核材等
- **风险提示：**算力互联需求不及预期，客户拓展及份额提升不及预期，产品研发及量产落地不及预期，行业竞争加剧

### 行业走势



### 相关研究

《从云端算力国产化到端侧 AI 爆发，电子行业的戴维斯双击时刻——电子行业 2026 年投资策略》

2025-12-10

《Credo 营收超预期&Marvell 收购 Celestial AI，催化光铜走强》

2025-12-08

## 内容目录

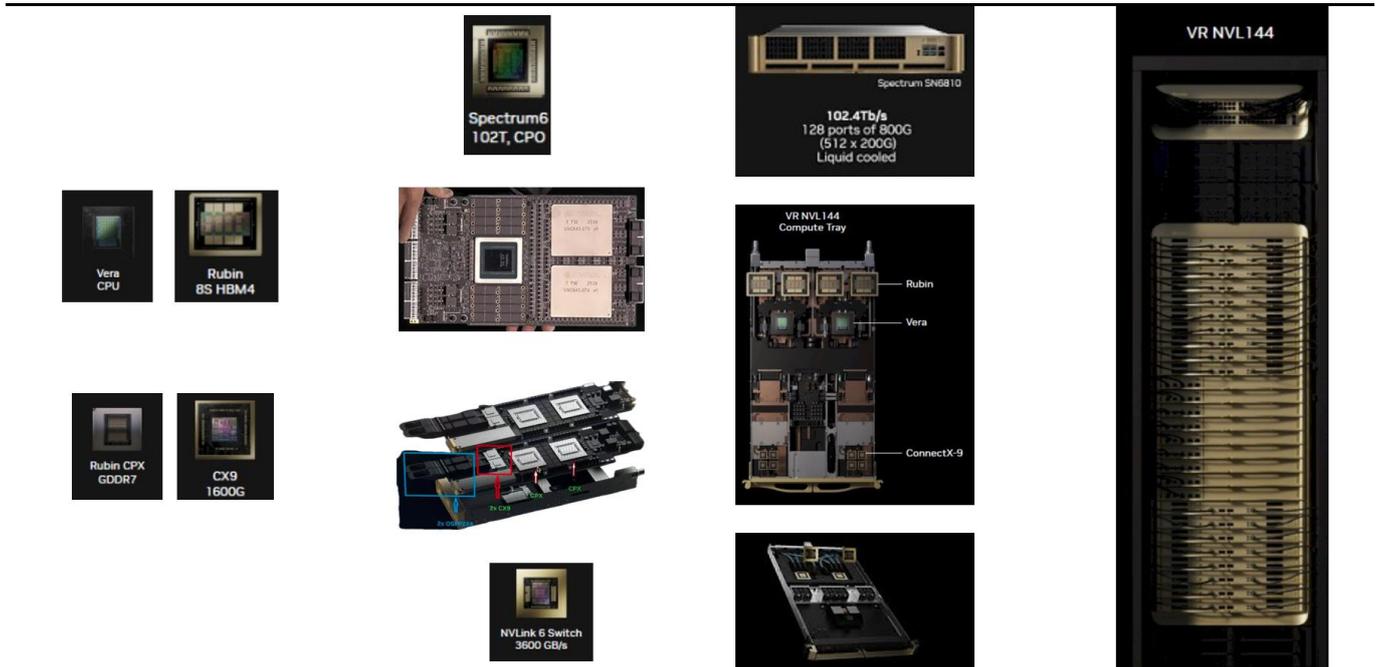
1. 英伟达 Rubin，集群互联高带宽低时延方向持续升级 .....	4
1.1. 第六代 NVLink 赋能，Rubin Scale up 持续突破.....	4
1.2. 满配 CPX，Rubin NVL 144 Scale out 芯光比最高可达 1:12.....	5
2. 谷歌 TPU Scale up&out，3D Torus+DCN 低时延互联升级 .....	7
2.1. 基于 3D Torus 拓扑，谷歌 TPU 多链路 Scale up 扩展 .....	7
2.2. TPU Scale out: DCN 多层组网+ OCS 支撑超大规模低时延互联.....	9
3. 亚马逊 Trainium3 组网突出“高密度互联 + 灵活扩展” .....	10
3.1. Scropio X 赋能，Tr3 基于 PCIe6+AEC 完成 Scale up 扩展.....	10
3.2. ENA/EFA 双网分工，高基数低速率交换机支撑 Tr3 Scale out.....	11
4. Meta 组网聚焦超大规模 AI 训练需求 .....	13
4.1. Meta Minarva Scale up: TH5+112G 铜缆 204.8T 对称互联 .....	13
4.2. Meta Scale out: DSF 网络架构，专为 AI 训练设计 .....	14
5. 投资建议 .....	16
6. 风险提示 .....	16

## 图表目录

图 1:	Rubin NVL144 CPX 拆解.....	4
图 2:	GB200 NVL 72 网络端口拆分.....	5
图 3:	TPU 计算托盘板卡.....	7
图 4:	TPU 服务器机柜.....	7
图 5:	TPU V7 3D Torus 拓扑图.....	8
图 6:	谷歌 TPU 147456 DCN 网络 Scale out 网络互联.....	10
图 7:	Trainium Teton3 NVL72*2 交换机架构.....	11
图 8:	Tranium3 131072 万卡集群 Scale out 组网.....	13
图 9:	Minerva 机柜网络拓.....	14
图 10:	MTIA 托盘图.....	14
图 11:	Jericho 托盘图.....	14
图 12:	Tomahawk 托盘图.....	14
图 13:	Meta DSF Fabric 网络拓扑.....	15
图 14:	Meta DSF Dual-stage Fabric 网络拓扑.....	15
表 1:	Rubin NVL 144 光模块需求比例测算.....	6
表 2:	谷歌 Ironwood TPU 3D Torus 连接端口配比关系测算.....	8

## 1. 英伟达 Rubin, 集群互联高带宽低时延方向持续升级

图1: Rubin NVL144 CPX 拆解



数据来源: 英伟达, 东吴证券研究所

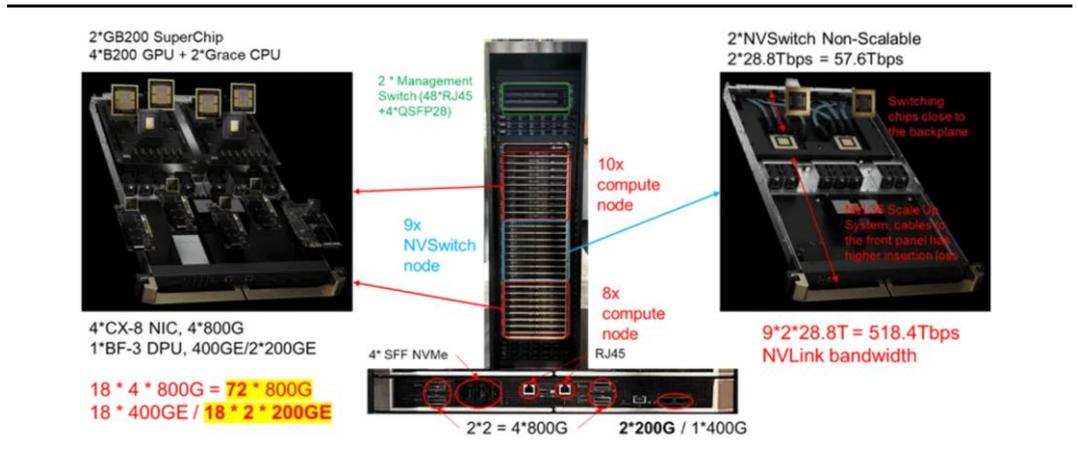
### 1.1. 第六代 NVLink 赋能, Rubin Scale up 持续突破

- Rubin NVL 144 Scale up 节点配置
  - 计算节点:
    - ✓ 18 个计算托盘: 每个托盘集成 4 个 Rubin GPU 模组(双 GPU die)。
    - ✓ 单颗 Rubin GPU 模组单向互联带宽 1.8TB/s, 整机柜计算侧单向带宽为  $72 * 1.8TB/s = 129.6TB/s$ 。
  - 交换节点:
    - ✓ 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。
    - ✓ 第六代 NVSwitch 单向带宽 3.6TB/s, 整机柜交换侧单向带宽 129.6TB/s。
- 基于 GB200 NVL 72 的 Scale up 方案配置推演
  - GB200 中, 72 颗 B200 芯片通过 5184 对 112G 差分对(对应 AEC/ACC

铜缆), 与 18 颗第五代 NVSwitch 完成互联, 形成约 64.8TB/s 单向带宽的机柜级 NVLink 网络。

- 类比至 Rubin NVL144, 在 GPU 数量与总带宽翻倍的情况下, 预计将沿用相近的物理端口规模, 通过将单通道速率提升至 224G, 采用 1.6TAEC 铜缆 (5184 对 224G 差分对), 完成 GPU 与 NVSwitch 之间的互联, 从而构建更高带宽密度的 scale-up NVLink 节点。

图2: GB200 NVL 72 网络端口拆分



数据来源: 陆玉春《Nvidia AI 芯片演进解读与推演 (二)》, 东吴证券研究所

## 1.2. 满配 CPX, Rubin NVL 144 Scale out 芯光比最高可达 1:12

AI 集群在模型训练过程中, 需实现大规模服务器间海量参数的高频同步。在有损网络环境下, 随着服务器集群规模扩容, 网络拥塞问题会显著加剧, 进而引发数据包丢失, 严重影响训练效率。胖树 (Fat Tree) 架构作为解决 AI 训练网络核心挑战的关键方案, 其技术优势通过以下三点实现精准突破:

- **无收敛拓扑设计:** 采用 1:1 上下行带宽配比的无收敛拓扑配置, 确保网络传输过程中无带宽瓶颈, 实现海量数据的无阻塞传输, 为参数同步提供高吞吐保障。
- **分层通信优化:** 单集群内所有节点整合为统一资源池, 池内节点间通信实现单跳可达; 跨集群通信通过汇聚层与核心层交换机协同转发, 可在三跳内完成数据传输, 大幅缩短传输路径, 提升数据交互效率。
- **RDMA 技术集成:** 架构内置远程直接内存访问 (Remote Direct Memory Access, RDMA) 技术, 通过绕开操作系统内核态, 实现主机间内存的直接数据读写, 在单集群单跳可达场景下, 显著降低端到端通信时延, 进一步适配 AI 训练的低时延需求。

无阻塞网络指系统内所有算力节点可相互连接, 无需中断现有连接。以 Rubin NVL144 机柜 (无 CPX) 与 Spectrum 6 交换机组网为例: Rubin NVL144 每个计算托盘后端扩展网络配置 4 个 CX-9 网卡 (满配 8 颗 CPX 时则配置 8 个), 分别对应托盘内 4

个 GPU 模组，通过 CX-9 的 4 个 OSFP 端口实现外部连接；Spectrum 6 交换机具备 102.4Tbps 交换带宽与 64 个 1.6T 端口，适配光模块连接。若需实现 9216 张 Rubin 的算力集群无阻塞互联，其组网逻辑如下：

- 第一层网络的服务器上行端口与 Tor 交换机下行端口等价。9216 个 1600Gbps 网卡流量对应 288 台 64\*1600Gbps 的 Tor 交换机 50%端口单向下行,Tor 交换机剩余 50%端口上行连接叶交换机。
- 第二层网络中，单台 Tor 交换机上行的 32\*1600G 带宽将占据 64\*1600G 叶交换机的 1/2, 叶交换机之间互不相连。考虑网络全互联端口充分利用，64\*1600G 端口的 Spectrum 6 交换机在双层胖树架构下最多仅可容纳  $64*64/2=2048$  张 GPU 互联，GPU 数量超过 2048 则需要引入第三层脊交换机。因此接纳 288 台 Tor 交换机一半端口上行，再预留同等数量端口向第三层脊交换机上行互联，第二层网络共需要 288 台 64\*1600G 的叶交换机。
- 第三层网络上脊交换机互不相连，完成第二层叶交换机的无阻塞互联需要一半叶交换机的数量，即 144 台脊交换机。

**表1: Rubin NVL 144 光模块需求比例测算**

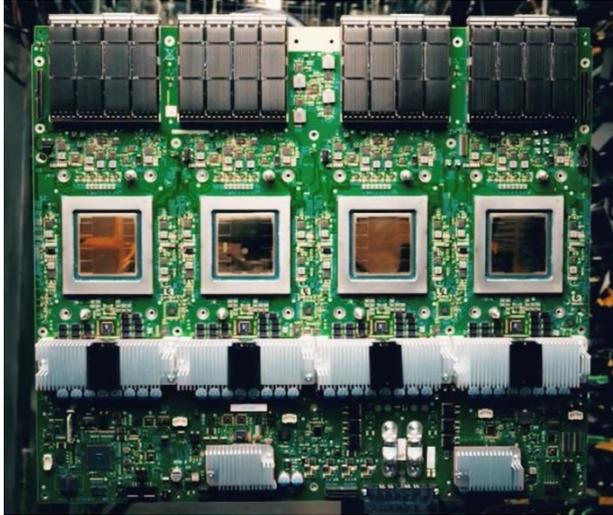
集群芯片量 (颗)	1152 (无 CPX)	9216(无 CPX)	1152 (满配 CPX)	9216 (满配 CPX)
网络层数 (层)	2	3	3	3
Tor 交换机数量 (台)	36	288	72	576
叶交换机数量 (台)	18	288	72	576
脊交换机数量 (台)		144	36	288
服务器 1.6T 端口 (个)	1152	9216	2304	18432
Tor 交换机 1.6T 端口 (个)	2304	18432	4608	36864
叶交换机 1.6T 端口 (个)	1152	18432	4608	36864
脊交换机 1.6T 端口 (个)		9216	2304	18432
总光模块需求数量 (个)	4608	55296	13824	110592
芯片/光模块比例	1:4	1:6	1:12	1:12

数据来源：英伟达，东吴证券研究所测算

## 2. 谷歌 TPU Scale up&out, 3D Torus+DCN 低时延互联升级

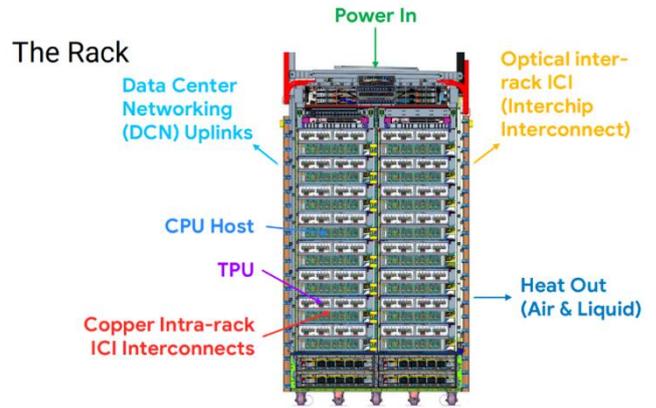
### 2.1. 基于 3D Torus 拓扑, 谷歌 TPU 多链路 Scale up 扩展

图3: TPU 计算托盘板卡



数据来源: Hot chips, 东吴证券研究所

图4: TPU 服务器机柜



数据来源: Hot chips, 东吴证券研究所

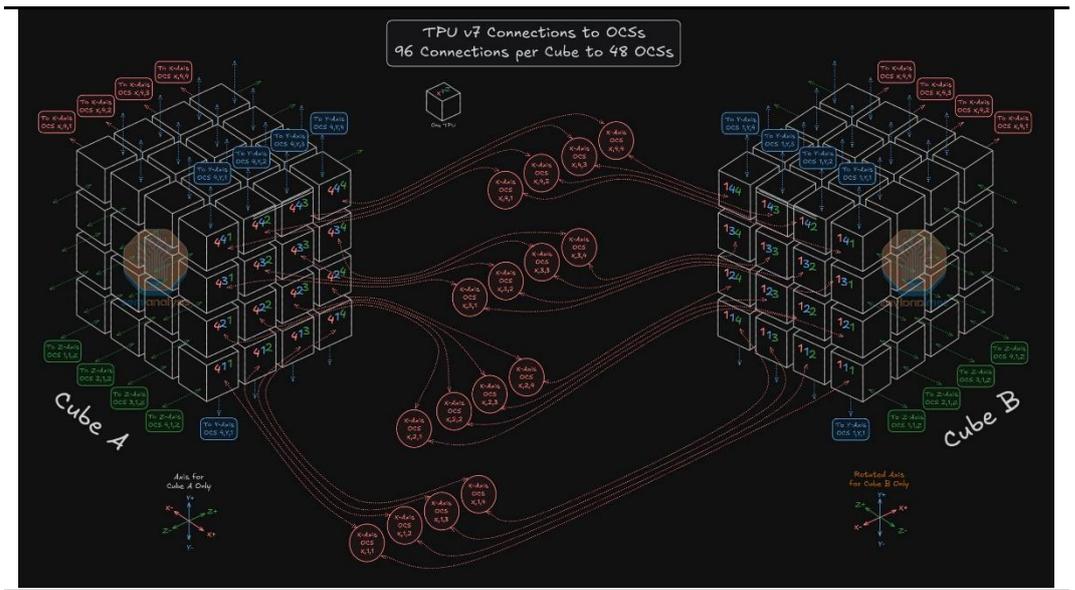
- 基础构建单元:

- POD 集群: 集群由 144 个 64 卡机柜组成, 总计 144 机柜\*64 卡=9216 颗 TPU 芯片, 整体集群即为谷歌的 ICI POD。
- 64 卡机柜: 单机柜包含 16 个 TPU 托盘, 每个托盘包含 4 颗 TPU。
- TPU 芯片 (Ironwood V7):
  - ✓ Scale up I/O 端口: 每颗 TPU 配置 6 个 I/O 端口。
  - ✓ 单向带宽: 每个 I/O 端口提供 800Gbps 带宽。
  - ✓ 芯片双向互联速率:  $800\text{Gbps} * 2 * 6/8 = 4.8\text{TB/s}$ 。

- POD 内部互联: 3D Torus 拓扑与 I/O 端口分配:

- 在每个 64 卡机柜内部, TPU 芯片采用 3D Torus 拓扑结构实现超高带宽芯片间互联。
  - ✓ 64 卡总 I/O 端口数=64 卡\*6 端口=384 端口。
  - ✓ 96 个 I/O 端口通过光模块+OCS 连接 Rack 外 TPU。
  - ✓ 128 个 I/O 端口通过 PCB 内部走线连接板上相邻 TPU。
  - ✓ 160 个 I/O 端口通过铜缆/AOC 与柜内不同板 TPU 连接。

图5: TPU V7 3D Torus 拓扑图



数据来源: Semi analysis, 东吴证券研究所

表2: 谷歌 Ironwood TPU 3D Torus 连接端口配比关系测算

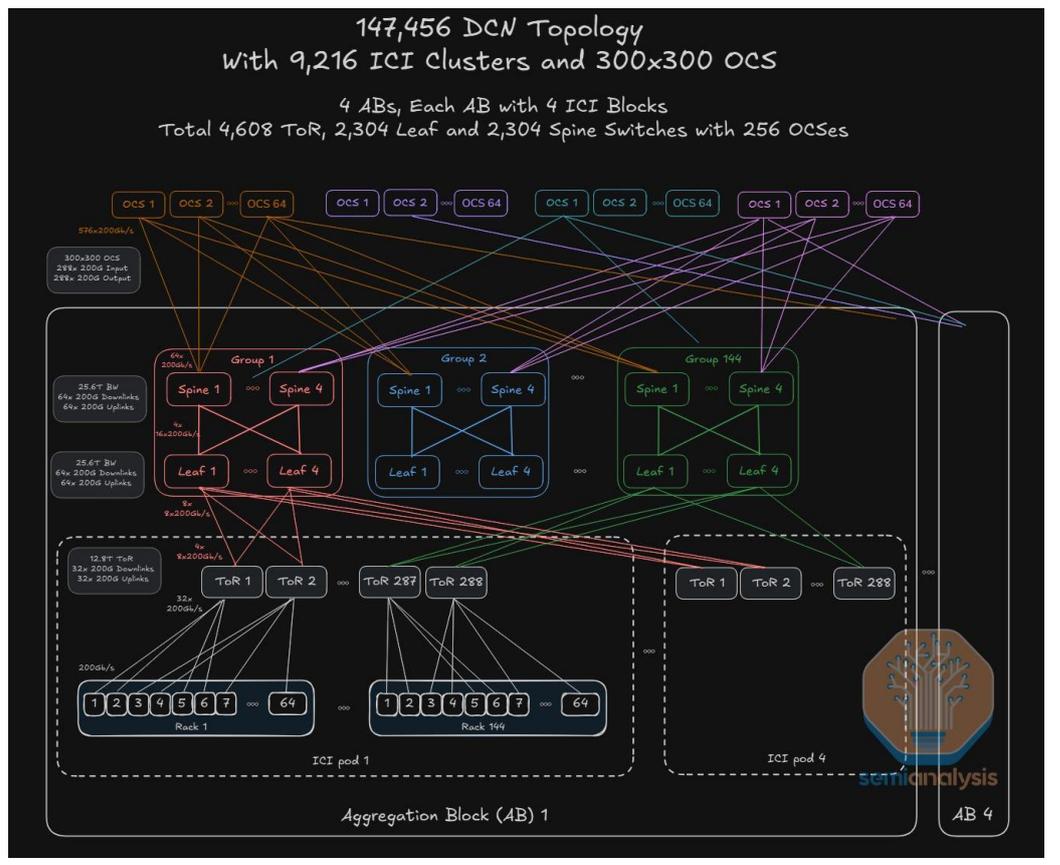
	每张 TPU 对应数量	整机柜总用量
8 颗内部 TPU		
铜缆端口 (个)	4	32
PCB 端口 (个)	2	16
光模块端口 (个)	0	0
8 颗角 TPU		
铜缆端口 (个)	1	8
PCB 端口 (个)	2	16
光模块端口 (个)	3	24
24 颗边 TPU		
铜缆端口 (个)	2	48
PCB 端口 (个)	2	48
光模块端口 (个)	2	48
24 颗面 TPU		
铜缆端口 (个)	3	72
PCB 端口 (个)	2	48
光模块端口 (个)	1	24
整机柜 64 TPU 与连接介质比例及用量		
铜缆 (根)	1.25	80
PCB (块)	1	64
光模块 (个)	1.5	96

数据来源: Semi analysis, 东吴证券研究所

## 2.2. TPU Scale out: DCN 多层组网+ OCS 支撑超大规模低时延互联

- 接入层: 64 卡机柜与计算托盘:
  - 计算托盘: 每个计算托盘集成 4 颗 TPU。
  - CDFP 端口: 每个计算托配置 4 个 CDFP PCIe 端口。
  - 机柜网络接入: TPU 通过 CDFP PCIe 与 CPU 托盘连接, 再通过 CPU 托盘中网卡上行与机柜内 Tor 交换机互联。
- 汇聚层: 9216 ICI POD 构成网络基本模块, 处理 POD 流量汇聚:
  - POD 规模: 144 个 64 卡机柜组成一个 9216 ICI POD。
  - TOR 交换机: 该 POD 使用 288 台 12.8T TOR 交换机提供接入服务。288 台 TOR 交换机采取水平分拆策略, 50%端口下行连接到 144 个机柜, 50%的端口上行连接到上层 Leaf 交换机。
- 核心层: 4 个 ICI POD 组成一个聚合模块, 是核心流量转发的主体:
  - Leaf/Spine 组: 每个聚合模块 (AB) 包含 4 组 Spine/Leaf 交换机, 用于连接 AB 内的 4 个 ICI POD:
    - ✓ 每组含 4 台 25.6T Spine 交换机和 4 台 25.6T Leaf 交换机。
  - Leaf 交换机: 576 台 25.6T Leaf 负责聚合模块内的流量汇聚:
    - ✓ 50%端口下行连接到 4 个 POD 内 1152 台 TOR 交换机。
    - ✓ 50%端口上行连接到 Spine 交换机将流量转发给核心层。
  - Spine 交换机: 负责提供 Leaf 交换机所需的上行带宽:
    - ✓ 50%端口用于聚合模块内的 Leaf 交换机连接。
    - ✓ 50%端口预留给跨聚合模块的互联, 即到顶层 OCS 系统。
- 集群互联层: 4 个聚合模块 (AB) 共 147456 颗 TPU 构成了整个大规模集群, 通过 OCS 实现全局, 动态互联:
  - ✓ OCS 数量和端口: 谷歌使用 64 台 300\*300 端口 OCS 将 2304 台 Spine 交换机 (4 个 AB 模块总 Spine 数量) 连接起来。每台 OCS 使用 288 个输入端口和 288 个输出端口。

图6：谷歌 TPU 147456 DCN 网络 Scale out 网络互联



数据来源：Semi analysis，东吴证券研究所

### 3. 亚马逊 Trainium3 组网突出“高密度互联 + 灵活扩展”

#### 3.1. Scropio X 赋能，Tr3 基于 PCIe6+AEC 完成 Scale up 扩展

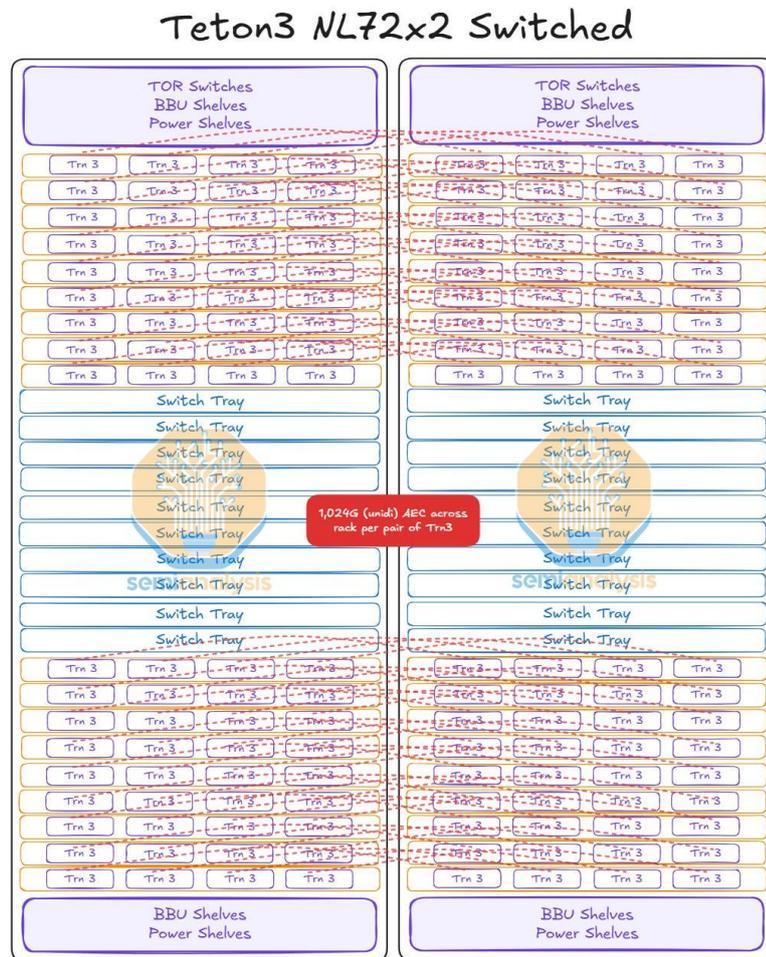
- 基础构建单元：Trainium3 NL72\*2:
  - 计算托盘\*36 个:
    - ✓ Trainium3 XPU\*144 颗 (72 颗 / 机架 \* 2)
    - ✓ Graviton CPU (1 颗 / 计算托盘)
    - ✓ Scropio X 32/64/128 通道 PCIe 6.0 交换芯片，用于同一板上 4 颗 Tr3 互联，按通道数分别需要配置 8/4/2 颗。
  - 交换托盘\*20 个:
    - ✓ Scropio X 320 通道 PCIe 6.0 交换芯片\*40 颗。
- Scale up 组网逻辑与 AEC 线缆需求规模:
  - 端口资源:单颗 Trainium3 芯片配置 160 个 NeuronLink4 端口连接 PCIe 6.0

通道，其中包括 144 个活跃端口，以及 16 个冗余端口

➤ 组网划分:

- ✓ PCB: 每颗 Tr3 分配 64 个 NeuronLink4 端口，通过板上 Scropio X 交换芯片和同托盘 4\*Tr3 互联，无外部线缆。
- ✓ 背板: 每颗 Tr3 分配 80 个端口，经由背板连接器及 AEC 铜缆接入交换托盘。整机柜 144 颗 Tr3\*80 通道=11520 PCIe6 通道，对应 180 根 64 端口 PCIe AEC 铜缆。
- ✓ 跨机架: 每颗 Tr3 分配 16 个端口，经由板上配置的 OSFP-XD 端口及 AEC 铜缆接入邻机架交换托盘。整机柜 144 颗 Tr3\*16 通道=2304 PCIe6 通道，对应 36 根 64 端口 PCIe AEC 铜缆。

图7: Trainium Teton3 NVL72\*2 交换机架构



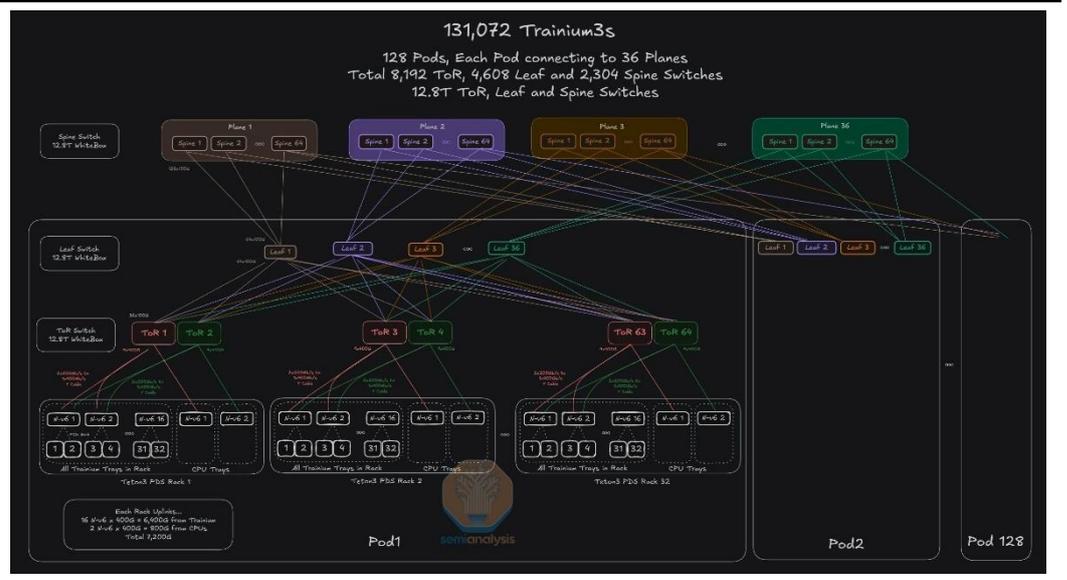
数据来源: Semi analysis, 东吴证券研究所

### 3.2. ENA/EFA 双网分工，高基数低速率交换机支撑 Tr3 Scale out

- AWS Scale out 组网核心特点:

- 双网络分工: ENA+EFA:
  - ✓ ENA (前端/南北向): 依托 400G Nitro-v6 网卡, 承载集群与外部的流量, 包括虚拟机通信、S3/EFS/EBS 存储访问、WAN/互联网接入等。
  - ✓ EFA (后端/东西向): 采用 SRD 定制协议, 提供每 2 颗 Trainium3 共享 400G 的低延迟带宽, 支撑 AI 集群通信的规模化扩展。
- Scale out 网络组网使用高基数低速率交换机选型:
  - ✓ 低速率设备成本更低、部署更灵活, 冗余性更强。
  - ✓ 扩展性线性: 交换机带宽从 12.8T 升级为 25.6T/51.2T 时, 集群规模可直接扩展 2/4 倍。
- AWS Scale out 组网的硬件配置与层级互联:
  - Nitro-v6 网卡配置
    - ✓ 东西向 EFA 后端网络: 2 颗 Trainium3 共享 1 个 400G Nitro-v6 网卡。
    - ✓ 南北向 ENA 前端网络: 1 颗 Graviton CPU 对应 1 个 400G Nitro-v6 网卡。
  - Tr3 及 Graviton 与 Tor 的互联:
    - ✓ Trainium3: 通过带 Gearbox 的 400G Y 型 AEC (56G SerDes 转 112G), 单网卡以 2 条 200G 链路连 NL72\*2 配置的双 ToR。
    - ✓ CPU: 通过直连 AEC/DAC, 连 NL72\*2 配置的双 ToR。
  - Leaf/Spine 组网:
    - ✓ 采用 Clos 拓扑: 8192 台 TOR (12.8T) 上行连接 4608 台 Leaf (12.8T), 进一步上行连接至 2304 台 Spine (12.8T), 层级间接无阻塞规则互联, 分 36 个平面优化传输。

图8: Trainium3 131072 万卡集群 Scale out 组网



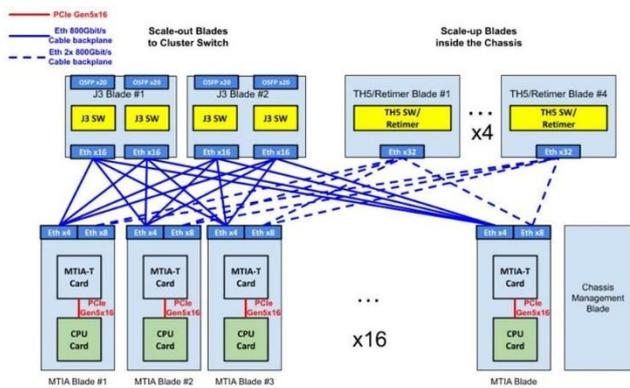
数据来源: Semi analysis, 东吴证券研究所

## 4. Meta 组网聚焦超大规模 AI 训练需求

### 4.1. Meta Minarva Scale up: TH5+112G 铜缆 204.8T 对称互联

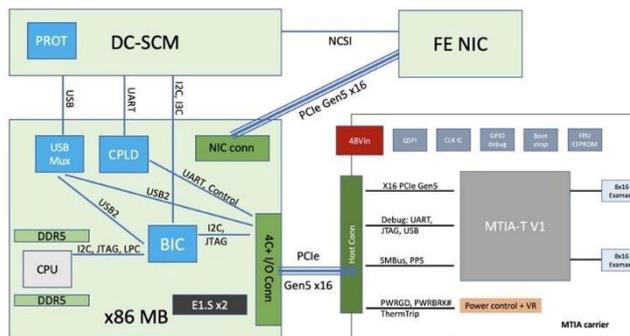
- Minarva 机柜 Scale up 基础构建单元
  - MTIA-T V1 计算托盘: 机柜内共 16 个, 每个托盘集成 1 张 MTIA-T V1 加速卡。
  - Tomahawk5 交换托盘: 机柜内共 4 个, 每个托盘集成 1 颗 TH5 交换芯片, 每个托盘背板接口包含 32 个 800G 以太网端口, 通过 4 组 cable cartridges 连接到 16 颗 MTIA。
  - 铜缆背板: 机柜内的高速互联介质, 采用 4 组 112G PAM4 的差分铜缆。每组包含 384 对不同的电缆线, 其中 256 对用于 Scale up 互联。整个铜缆背板上共有 1024 对电缆线用于 Scale up。
- Minarva 机柜 Scale up 互联拓扑
  - MTIA 单卡对外通过 8 个 2\*800Gbps 端口进行 Scale up 互联。
  - 16 颗 MTIA 芯片总带宽为  $16 * 8 * 2 * 800 = 204.8 \text{Tbps}$ 。
  - 4 颗 Tomahawk5 交换芯片总带宽为  $4 * 51.2 \text{T} = 204.8 \text{Tbps}$

图9: Minerva 机柜网络拓



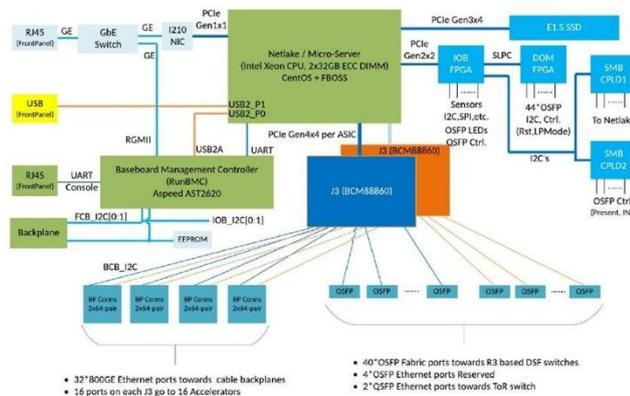
数据来源: Substrack, 东吴证券研究所

图10: MTIA 托盘图



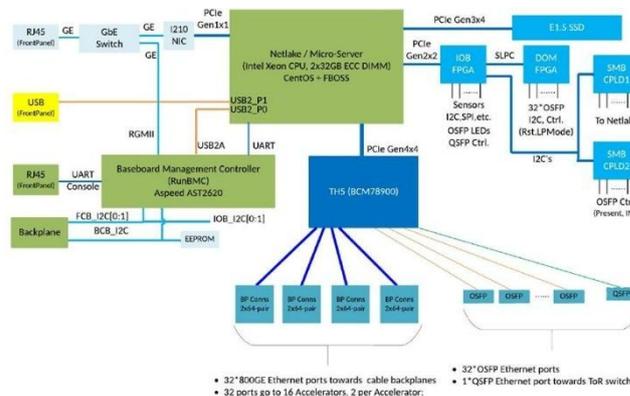
数据来源: Substrack, 东吴证券研究所

图11: Jericho 托盘图



数据来源: Substrack, 东吴证券研究所

图12: Tomahawk 托盘图



数据来源: Substrack, 东吴证券研究所

## 4.2. Meta Scale out: DSF 网络架构, 专为 AI 训练设计

Meta 基于自身大规模 AI 训练需求, 构建了 DSF (Disaggregated Scheduled Fabric) 网络架构, 以解决超大规模算力集群在 Scale-out 场景下面临的带宽瓶颈、拥塞与调度不可控问题。DSF 通过将交换结构与调度能力解耦, 引入 RDSW 与 FDSW 的分层设计, 实现确定性、高利用率的算力互联。

在 Minerva 机柜内部, DSF 的最小实现单元已经完整落地。单机柜内共部署 16 个 MTIA-T V1 计算托盘, 每个托盘集成 1 张 MTIA-T V1 加速卡; 同时配置 2 个 Jericho3 交换托盘, 每个托盘集成 2 颗 Jericho3 交换芯片, 合计 4 颗 Jericho3, 构成机柜级 RDSW (Rack Disaggregated Switch)。

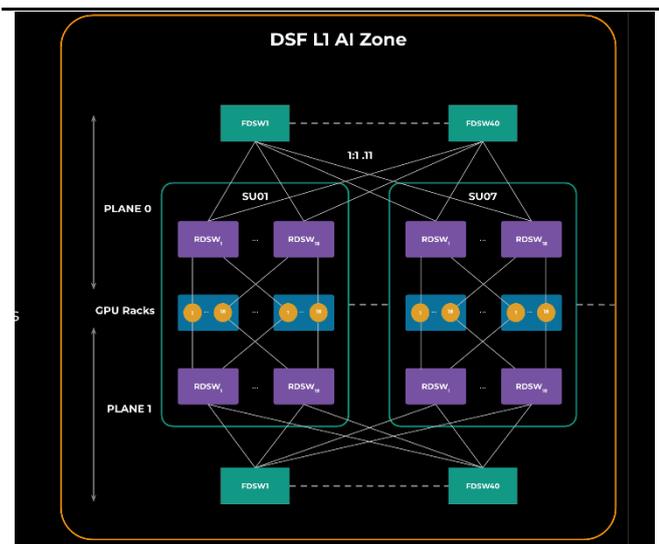
MTIA 与 Jericho3 之间通过机柜高速铜缆背板直连, 基于 112G PAM4 SerDes, 采用 Examax2 高密度连接器实现。每颗 MTIA 通过 4 × 800Gbps 以太网端口接入 RDSW, 16 颗 MTIA 的总上行带宽为 51.2Tbps; 对应的 4 颗 Jericho3 总交换能力约 57.6 Tbps, 形成

略有冗余的无阻塞机柜级 Scale-out 互联。

在机柜之上，RDSW 通过光互连接入 Fabric 层。FDSW(Fabric Disaggregated Switch) 采用 Ramon3 交换芯片，负责跨机柜、跨 Pod 的大规模 Scale-out 互联。RDSW 与 FDSW 之间使用  $2 \times 400G$  FR4 (800G) 光模块连接；两个 Jericho3 交换托盘共提供 40 个 OSFP 端口，用于上联至 FDSW，构成 DSF Fabric 的基础网络单元。

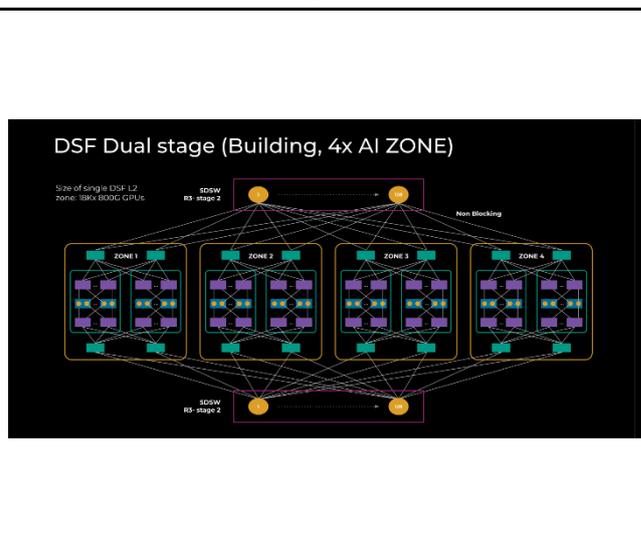
在 DSF Fabric 的基础网络单元之上，进一步可扩展至 DSF Dual-stage Fabric，核心逻辑是引入第二层 Fabric 交换机 (SDSW, Spine Disaggregated Switch)，从而形成一个真正意义上的非阻塞交换网络。FDSW 与 SDSW 之间同样采用 800G 光模块(如  $2 \times 400G$  FR4 或 DR8) 进行长距离互联。通常 SDSW 层的总交换能力会与下层 FDSW 的总上行带宽保持 1:1 的收敛比。由于 Ramon3 芯片专为信元交换设计，其单芯片容量与 Jericho3 匹配，通过多级堆叠，可以构建出一个逻辑上等效于“巨型模块化交换机”的扁平网络。

图13: Meta DSF Fabric 网络拓扑



数据来源: Meta, 东吴证券研究所

图14: Meta DSF Dual-stage Fabric 网络拓扑



数据来源: Meta, 东吴证券研究所

● Meta 18432 MTIA-DSF Dual-stage Fabric Scale out 组网

- ✓ 机柜总数:  $18432/16=1152$  个 Minerva 机柜
- ✓ RDSW 数量: 每个机柜配 2 个 Jericho3 交换托盘 (每个托盘含 2 颗 Jr3 芯片, 共 4 颗芯片) 总计  $1152*2=2304$  台 RDSW 交换托盘
- 第一层 Fabric (FDSW-Leaf)
  - ✓ 下行连接: 需要接入 1152 个机柜产生的  $1152*40=46080$  个 OSFP 端口
  - ✓ FDSW 规模: FDSW 采用 Ramon3 芯片, 规格为  $64*800G=51.2T$ , 1:1 无损转发, 32 口下联 RDSW, 32 口上量 SDSW。共需要  $46080/32=1440$

### 台 FDSW

- 第二层 Fabric (SDSW-Spine)
  - ✓ 上行连接: FDSW 的上行总端口数量同样为  $1440 \times 32 = 46080$  个。
  - ✓ SDSW 规模: 每台 SDSW 提供 64 个端口全速转发, 需要 SDSW 数量  $= 46080 / 64 = 720$  台 SDSW。
- 光模块需求总量
  - ✓ RSDW-FDSW 层:  $46080 \times 2 = 92160$  个
  - ✓ FDSW-SDSW 层:  $46080 \times 2 = 92160$  个
  - ✓ 合计 184320 个 800G OSFP 光模块

## 5. 投资建议

2026 年商用 GPU 持续放量, CSPASIC 同步进入大规模部署的关键一年, 数据中心 Scale up 催生超节点爆发, 铜缆凭短距低耗低成本, 成为柜内互联最优解; Scale out 带动集群持续扩容, 光模块与 GPU 配比飙升, 产品放量让光芯片缺口凸显。一铜一光双线共振, 互联需求迎来量价齐升, 建议重点关注光铜两大核心赛道。产业链相关公司:

光芯片: 长光华芯、源杰科技、仕佳光子等

铜缆: 华丰科技、兆龙互联、沃尔核材等

## 6. 风险提示

算力互联需求不及预期: 下游 AI 算力建设投入力度、算力网络带宽扩容规模若未达市场预期, 将直接导致客户对网络互联相关产品的采购需求放缓, 进而对相关公司的营收增长及盈利表现产生不利影响。

客户拓展及份额提升不及预期: 相关企业若未能按预期开拓潜在客户资源, 或在现有核心客户供应链中的份额提升进度低于预期, 将制约其市场渗透率提升, 进而影响业绩增长的持续性。

产品研发及量产落地不及预期: 针对高潜力场景的新兴网络互联产品, 若相关公司在技术研发突破、产品迭代升级或规模化量产落地环节未达预期, 将错失市场发展机遇, 对公司长期业绩增长构成制约。

行业竞争加剧: 随着网络互联赛道商业化进程加快, 行业参与者增多可能导致竞争

持续加剧，若相关公司未能维持技术壁垒、成本控制或客户服务优势，其产品市场份额存在下滑风险，进而影响盈利能力。。

## 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>