东兴证券 DONGXING SECURITIES

超节点行业：从计算托盘角度拆解英伟达 VR NVL72，通信速率三重升级，超级网卡价值显著提升

2026 年 5 月 22 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：

2026 年英伟达最新发布的超节点 Vera-Rubin NVL72，是全球领先的 Scale up 网络算力平台。Rubin 平台由六款全新芯片组成，包括 Vera CPU、Rubin GPU、NVLink 6 交换机、ConnectX-9 SuperNIC、BlueField-4 DPU 和 Spectrum-6 以太网络交换机。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。

在 VR NVL72 中，AI 计算任务从外部网络进来，数据经过 ConnectX-9、BlueField-4、Vera CPU，再分配到 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 OSFP 之间通过 800G 以太网/ InfiniBand 互联。

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。

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。

采用以太网/ InfiniBand 协议实现超级网卡 CX-9 与 OSFP 光模块笼口互联。CX-9 一项重要升级在于，其在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 OSFP 笼位的数量有两种方案：一种是每颗 GPU 配 1 个 1.6T OSFP 笼口，则每个计算托盘共 4 个 1.6T OSFP 笼口；另一种则是每颗 GPU 配 2 个 800G OSFP 笼口，则每个计算托盘共 8 个 800G OSFP 笼口。

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。

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 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。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。

投资策略：

自 2025 年开始，超节点成为 AI 算力网络重要的技术创新方向。本篇报告从计算托盘角度拆解英伟达 VR NVL72，可以看到，英伟达 VR NVL72 以 1.8TB/s NVLink-C2C+PCIe Gen6+800G SuperNIC 构建三重高速通信壁垒，其核心竞争力源于芯片级互联、高速总线、超级网卡的全栈技术垄断。当前我国 AI 算力网络在超高速互联协议、800G SuperNIC、PCIe Gen6 交换芯片等领域仍存代差，自主可控需求迫切。建议聚焦高速互联芯片、800G/400G SuperNIC、高端光模块、高速 PCB / 覆铜板、超节点整机方案五大国产替代主线。

相关公司：

（1）高速网卡：裕太微（688515）、盛科通信（688702）；

（2）光模块：中际旭创（300308）、新易盛（300502）、天孚通信（300394）、东山精密（002384）、华工科技（000988）、光迅科技（002281）；

（3）高速交换芯片：盛科通信（688702）、紫光股份（000938）、锐捷网络（301165）、中兴通讯（000063）；

（4）高速互联芯片：海光信息（688041）、龙芯中科（688047）、长电科技（600584）；

（5）高速 PCB / 覆铜板：胜宏科技（300476）、生益科技（600183）、深南电路（002916）、东材科技（601208）；

（6）国产超节点：浪潮信息（000977）、中科曙光（603019）、工业富联（601138）、华勤技术（603296）。

风险提示：（1）技术代差；（2）国产超节点生态割裂；（3）地缘政治风险。

目 录

插图目录

表格目录

1. 超节点 VR NVL72，全球领先的 Scale up 网络算力平台

随着大模型技术的发展，越来越多的算力集群需要同时承载训练和推理负载，但 AI 训练与推理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1：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

在最新的 AI 大模型训练中，张量并行与专家并行计算同样要求高带宽与极低时延。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 会被路由到不同专家，专家分布在不同设备上，就会产生大量分发和聚合通信。专家越多，并发越高，通信越重。

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

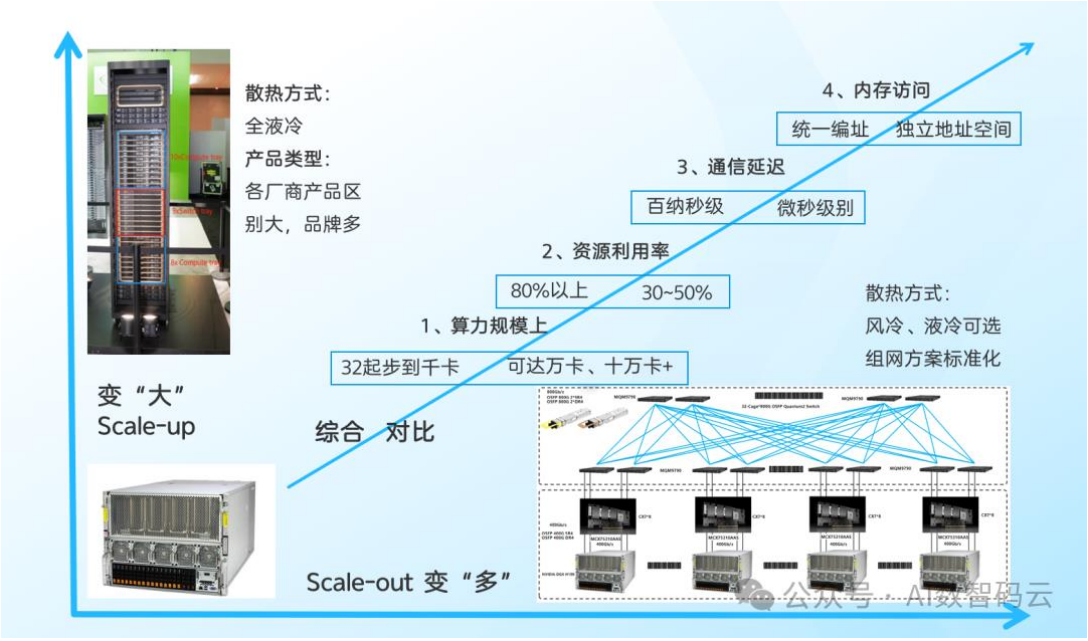
Scale up 与 Scale out 网络均衡发展，构建高带宽互联的集群算力网络。随着模型参数和集群规模继续扩大，大模型训练并不是简单的“卡越多越快”。如果通信跟不上，更多算力卡只会带来更多等待，而等待同步的时间就会吞掉算力收益。因此，构建 Scale up 网络，把数十至数千算力卡组织进一个更紧密的高性能计算单元，实现高带宽通信范围扩大，从而满足需要频繁交换数据、对通信极其敏感的计算单元和应用场景。

Scale up 网络与 Scale out 网络特点对比如下：

Scale up（左）vs Scale out（右）

- 算力规模：数十卡至千卡级 vs 万卡至十万卡级；
- 资源利用率：80%以上 vs 30%-50%；
- 通信延迟：百纳秒级 vs 微秒级；
- 内存访问：统一内存或全局地址空间 vs 独立内存空间；
- 标准化：定制化程度高 vs 基于开放网络标准，相对统一。

图2：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

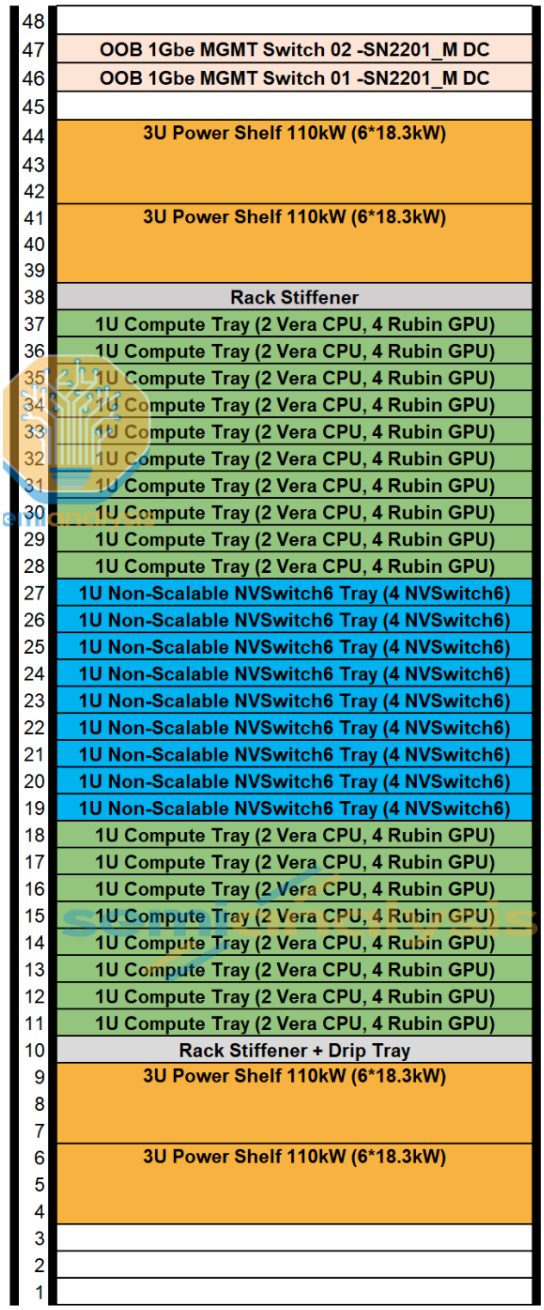
2026 年英伟达最新发布的超节点 Vera-Rubin NVL72，是全球领先的 Scale up 网络算力平台。Rubin 平台由六款全新芯片组成，包括 Vera CPU、Rubin GPU、NVLink 6 交换机、ConnectX-9 SuperNIC、BlueField-4 DPU 和 Spectrum-6 以太网络交换机。据英伟达公布的数据，Rubin 平台的训练性能达到前代 Blackwell 的 3.5 倍，运行 AI 软件的性能提升 5 倍。此外，与前一代相比，英伟达 Rubin 平台在训练 MoE 模型时所需的 GPU 数量减少至原来的四分之一，进一步推动人工智能的普及应用。

图3：GTC2026 大会上 NVIDIA Vera Rubin NVL72 机架

资料来源：GTC2026 大会，东兴证券研究所

VR NVL72 机柜构成包括：顶部的 OOB 管理交换机、电源柜、机架加强件、运算托盘与交换托盘。其中计算托盘 18 个，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；交换托盘 9 个，每个托盘集成 4 颗第六代 NVSwitch 芯片；共计 72 个 Rubin GPU 封装、36 个 Vera CPU 与 36 个 NVLink 6 Switch ASIC。

图4：VR NVL72 机柜组织图



资料来源：SemiAnalysis，东兴证券研究所

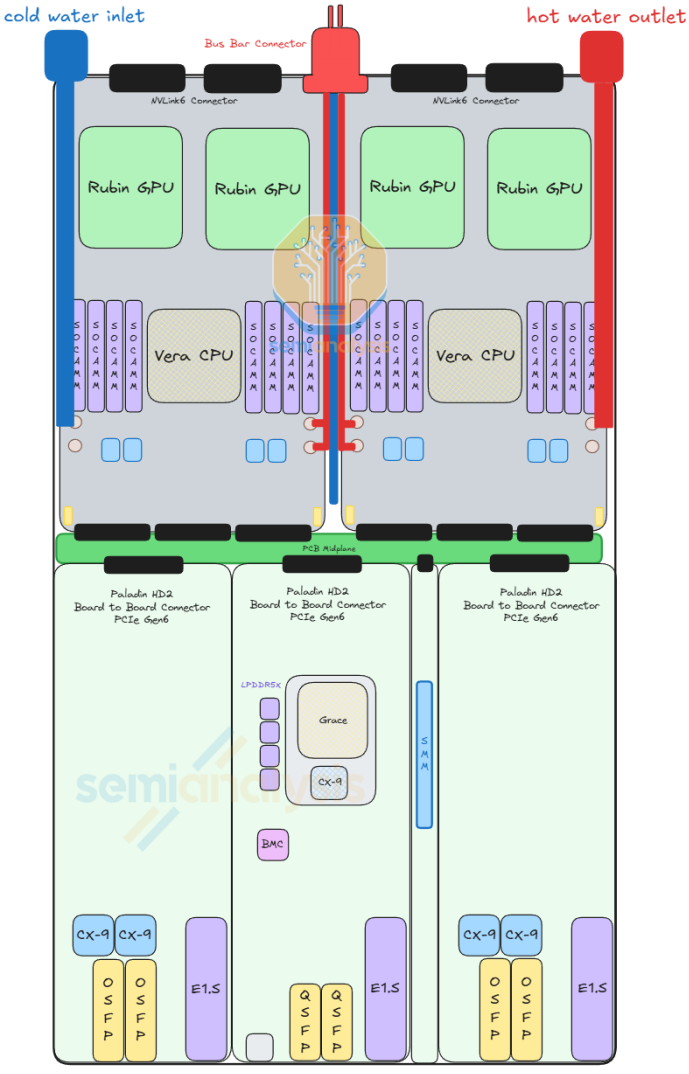
2. 拆解 VR NVL72 计算托盘，通信速率三重升级：1.8TB/s NVLink-C2C+ PCIe Gen6+800G 以太网

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 BlueField-4 模块、1 块 PDB 电源分配板、1 套 SMM 系统管理模块。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5：VR NVL72 机柜计算托盘顶视图



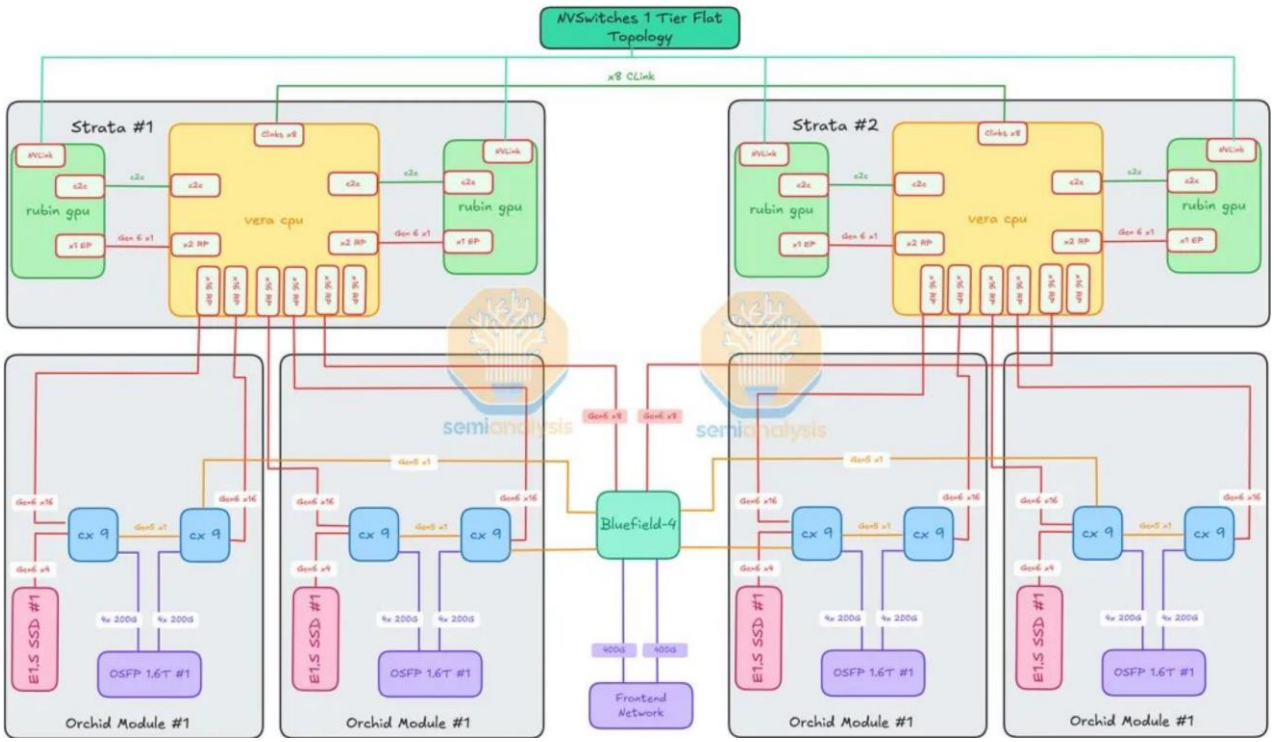
资料来源：SemiAnalysis，东兴证券研究所

在 VR NVL72 中，AI 计算任务从外部网络进来，数据经过 ConnectX-9、BlueField-4、Vera CPU，再分配到 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 OSFP 之间通过 800G 以太网/ InfiniBand 互联。

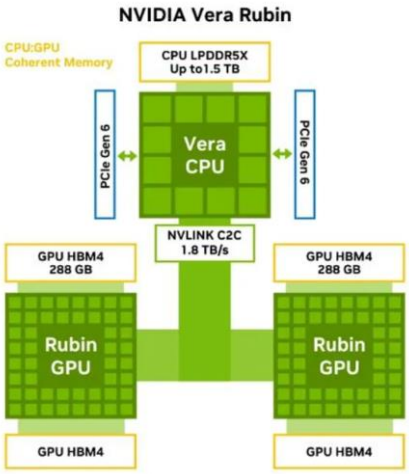
图6：VR NVL72 计算托盘拓扑结构



资料来源：SemiAnalysis，东兴证券研究所

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7：VR NVL72 中 Rubin-Vera 芯片互联方式

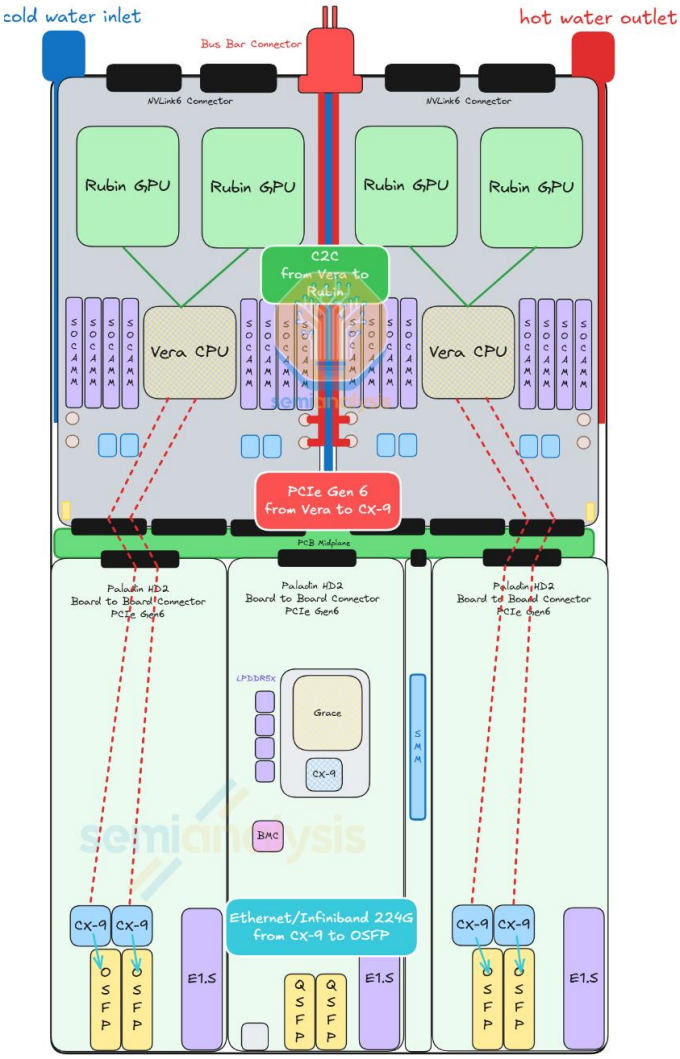


资料来源：英伟达官网，东兴证券研究所

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8：VR NVL72 运算托盘信号路径图



资料来源：SemiAnalysis，东兴证券研究所

采用以太网/ InfiniBand 协议实现超级网卡 CX-9 与 OSFP 光模块笼口互联。CX-9 一项重要升级在于，其在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。

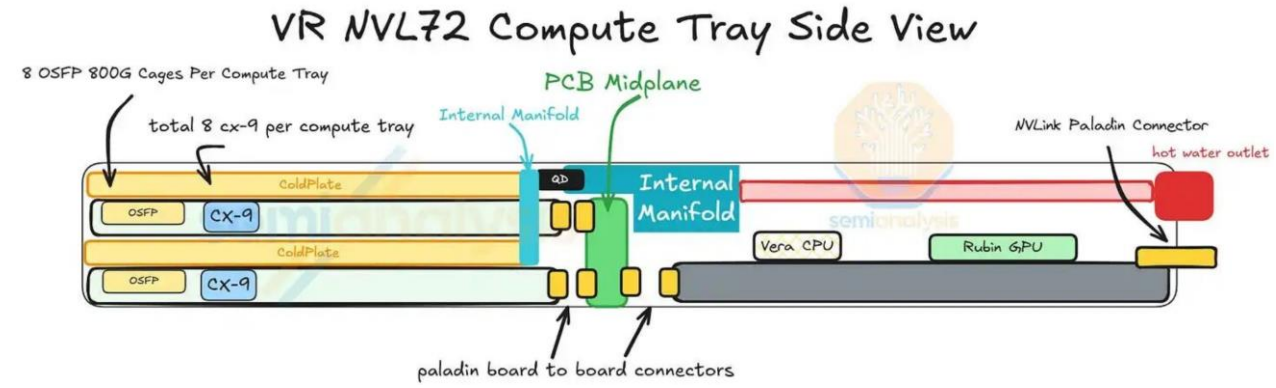
在 VR NVL72 计算托盘中，8 个 800G 的 CX-9 网卡对应 OSFP 笼位的数量有两种方案：一种是每颗 GPU 配 1 个 1.6T OSFP 笼口，则每个计算托盘共 4 个 1.6T OSFP 笼口；另一种则是每颗 GPU 配 2 个 800G OSFP 笼口，则每个计算托盘共 8 个 800G OSFP 笼口。

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9：VR NVL72 计算托盘侧视图



资料来源：SemiAnalysis，东兴证券研究所

3. 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10：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。

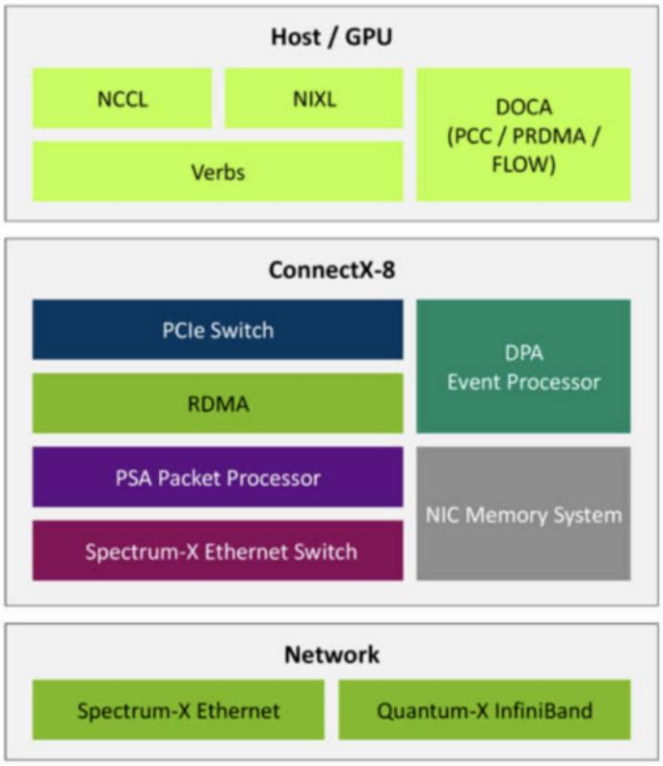
资料来源：英伟达，东兴证券研究所

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）、FLOW（流处理）等功能。
- 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11：ConnectX-8 网卡技术栈

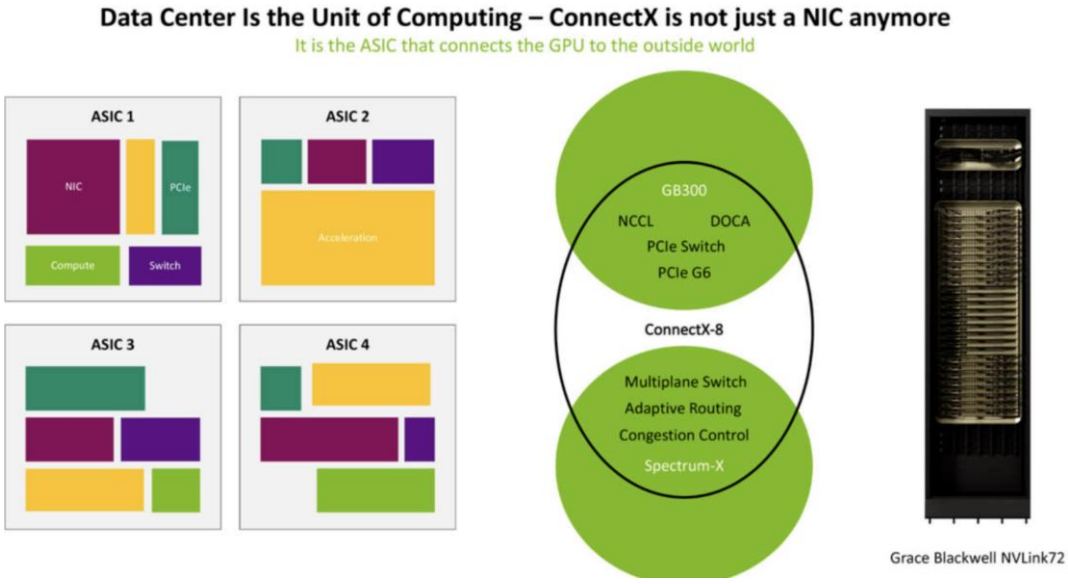


资料来源：英伟达，东兴证券研究所

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12：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

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13：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

4. 投资建议

自 2025 年开始，超节点成为 AI 算力网络重要的技术创新方向。本篇报告从计算托盘角度拆解英伟达 VR NVL72，可以看到，英伟达 VR NVL72 以 1.8TB/s NVLink-C2C+PCIe Gen6+800G SuperNIC 构建三重高速通信壁垒，其核心竞争力源于芯片级互联、高速总线、超级网卡的全栈技术垄断。当前我国 AI 算力网络在超高速互联协议、800G SuperNIC、PCIe Gen6 交换芯片等领域仍存代差，自主可控需求迫切。建议聚焦高速互联芯片、800G/400G SuperNIC、高端光模块、高速 PCB / 覆铜板、超节点整机方案五大国产替代主线。

相关公司：

（1）高速网卡：裕太微（688515）、盛科通信（688702）；

（2）光模块：中际旭创（300308）、新易盛（300502）、天孚通信（300394）、东山精密（002384）、华工科技（000988）、光迅科技（002281）；

（3）高速交换芯片：盛科通信（688702）、紫光股份（000938）、锐捷网络（301165）、中兴通讯（000063）；

（4）高速互联芯片：海光信息（688041）、龙芯中科（688047）、长电科技（600584）；

（5）高速 PCB / 覆铜板：胜宏科技（300476）、生益科技（600183）、深南电路（002916）、东材科技（601208）；

（6）国产超节点：浪潮信息（000977）、中科曙光（603019）、工业富联（601138）、华勤技术（603296）。

5. 风险提示

（1）技术代差；（2）国产超节点生态割裂；（3）地缘政治风险。

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。

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 46F
邮编：518038
电话：0755-83239601
传真：0755-23824526