下载日志已记录，仅供内部参考

半导体

2026 年 10 月 15 日

投资评级：看好（维持）

AI 互连系列报告（二）：围绕 CPU-外设及 GPU-GPU 两条主线探索当前超节点的柜内互连

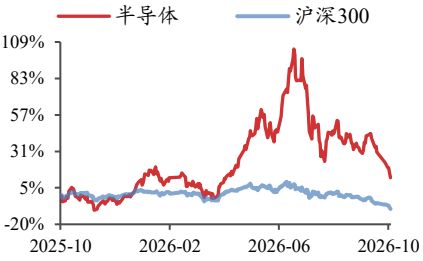
——行业深度报告

陈蓉芳（分析师）
chenrongfang@kysec.cn
证书编号：S0790524120002

陈瑜熙（分析师）
chenyuxi@kysec.cn
证书编号：S0790525020003

陈凯（分析师）
chenkai1@kysec.cn
证书编号：S0790526030004

行业走势图



数据来源：聚源

相关研究报告

《以台积电发展史为镜，看本土晶圆代工行业的战略机遇—行业深度报告》-2026.3.23

《先进封装龙头积极抢滩布局，产业进入“扩产+提价”新阶段—行业点评报告》-2026.1.18

《AI 算力自主可控的全景蓝图与投资机遇—行业投资策略》-2026.1.8

● **超节点是大规模算力集群构建的基础单元，柜内互连为超节点提供基础支撑**

超节点的本质是柜内资源高密度整合，通过增加 GPU 等计算单元，并依赖高速低延迟内部网络实现高效通信。当前构建大规模算力集群离不开 scale-up 与 scale-out 两个层级的互连，其中 scale-up 是核心环节，它指的是在单个计算节点内通过增加计算单元（GPU、CPU）、内存等资源，并采用极高带宽、极低延迟的内部互连技术从而提升该节点的整体计算能力和通信效率。由此当我们把视角**聚焦在超节点柜内，围绕 CPU-外设和 GPU-GPU 的互连成为当下决定超节点形态的核心议题，它关系着超节点的带宽、延迟、组网模式和架构等关键指标。**

● **CPU-外设：PCIe 承担着最广泛的互连，私有协议部分渗透 CPU-GPU**

PCIe 发展至今已拥有成熟的生态和最广泛的兼容性，成为 AI 互连领域的通用标准。PCIe 及其配套的 PCIe Switch 在 CPU-外设的互连层级提供了关键的支撑，不仅所有 GPU 在与主机 CPU 通信时都支持 PCIe，CPU/GPU 同其他外设的互连，再到外设拓展同样离不开 PCIe 的连接及交换芯片的辅助。**由此可见，PCIe 在 CPU-外设层级的地位是无可动摇的，直接决定超节点构建时计算托盘的基础布局。**英伟达针对 CPU-GPU 的互连采用了自有生态的 NVLink C2C 以实现 CPU-GPU 的高带宽低延迟互连，而 CPU/GPU 同其他外设的互连仍然依托 PCIe 实现，这种 CPU-GPU 的私有互连为超节点的深入发展提供了全新思路。

● **GPU-GPU：私有协议群雄逐鹿，开放生态逐步渗透**

实现柜内海量 GPU-GPU 间高带宽、低延迟的互连是构建超节点服务器的关键，也是当前 AI 大模型加速迭代发展的题眼，GPU-GPU 层级下的各类协议及其适配的交换芯片直接决定了超节点的搭建和整体拓扑。在 GPU-GPU 层级下，具有完整生态及 GPU 定义权的大厂如英伟达、华为、AMD 等为其自身加速卡搭建了完整的私有 GPU-GPU 互连体系，如 NVLink、UB、Infinity Fabric 等，私有协议也凭借自身 GPU 的出货量在整个 GPU-GPU 层级下占据了相当的覆盖度，呈现出群雄逐鹿的态势。在私有协议之外，以 PCIe 和以太网为核心的技术凭借成熟的生态和硬件体系，正逐步向 GPU-GPU 层级渗透，如 SUE、UALink 正逐步开始各大服务器厂商及 CSP 厂商侧得到认可，助力其专属超节点的搭建。

● **AI 互连逐步成为当前超节点建设的核心通胀环节，看好国产核心替代者**

综上我们认为，随着全球超节点形态的确认及 AI 算力集群进入加速建设阶段，以各类 AI 互连协议配套交换芯片为核心的 AI 互连板块将逐步成为继算力和存力的 AI 产业第三增长极。看好卡位关键节点且占据先发优势的国产核心替代者。**核心推荐标的：万通发展、盛科通信、澜起科技**

● **风险提示：**AI 算力建设不及预期、技术进步/标准落地不及预期、生态适配不及预期。

下载日志已记录，仅供内部参考

目 录

1、 “CPU-外设”+“GPU-GPU”构成了超节点内主要互连 4
2、 CPU-外设：决定超节点层级的基础组网 5
2.1、 PCIe：应用最广泛，生态最成熟的 CPU-外设的互连总线 5
2.2、 NVLink-C2C：针对 CPU-GPU 的高速 NVLink 6
2.3、 CXL：建立在 PCIe 之上的跨设备缓存一致性高速互连标准 7
3、 GPU-GPU：利用多种私有/开放协议实现计算资源高效互连 8
3.1、 PCIe：在超节点层级实现 GPU-GPU 间大规模互连 8
3.2、 NVLink：专属 NVIDIA 的服务器内部紧耦合 GPU 高速互连方案 10
3.3、 Infinity Fabric：缓存一致与统一显存提升多 GPU 协同效率 11
3.4、 华为 Ascend UB 互连：面向昇腾平台形成自主可控高速互连 12
3.5、 XLink 及其他方案：定制化提升局部性能但生态碎片化明显 15
3.6、 SUE/UALink/ETH+/…：依托于成熟的以太网生态的多种开放 scale-up 互连方案 15
3.6.1、 SUE：利用优化以太网协议栈，为 XPU 集群提供高效的互连解决方案 16
3.6.2、 UALink：全新开放的超节点内部紧耦合高速互连方案 17
4、 投资建议 19
5、 风险提示 19

图表目录

图 1： Scale-up& scale-out 在大规模 AI 集群构建中起到了核心互连的作用 4
图 2： （超）节点（柜）内互连模式可分为“CPU-外设”与“GPU-GPU” 5
图 3： PCIe 是 AI 服务器柜内 CPU-外设最核心的互连总线 5
图 4： NVLink C2C 在 NVL72 机柜中起到 CPU-GPU 互连的作用 6
图 5： PCIe 同 NVLink C2C 一起承担 NVL72 中的 CPU-外设互连 7
图 6： CXL 定义的三类设备及其 PHY、控制器与应用层 7
图 7： 利用 Scorpio X-Series 320 Lane 实现的 High Radix, Multi-level Switch Scale-up 组网 9
图 8： Astera Labs 公司 Scorpio P-/X-Series PCIe Switch 概览 9
图 9： NVLink Scale-up 技术成熟，创新步伐加快 10
图 10： NVIDIA 核心 GPUxNVLink 架构演变史 11
图 11： AMD Instinct MI300X&MI325X 平台利用 Infinity Fabric 实现 8GPU 全互连 12
图 12： UB 协议架构分层技术说明 13
图 13： CloudMatrix384 超节点的对等直连硬件架构及 UB 交换系统示意图 14
图 14： 华为昇腾新一代 Atlas 系列产品 14
图 15： 谷歌 TPUv4 中 TPU 通过 ICI 沿着 x/y/z 三个维度完成互连形成 4x4x4 结构 15
图 16： 标准 SUE 协议栈及工作流程示意图 16
图 17： XPU 通过部署 12 个 SUE 并使用 12 台 64 个端口的以太网交换机实现全交换连接（示例） 17
图 18： XPU 利用 SUE 直接互连网状拓扑结构（示例） 17
图 19： 基于 UALink 的 Scale-Up 机柜架构中的多节点与多平面交换 18
图 20： 基于 UALink 的多加速器系统 18

表 1： CXL 各代际说明及核心变化对比 8
表 2： NVLink 各代际速率及适配情况 10

下载日志已记录，仅供内部参考

表 3： 核心推荐标的汇总表（截至 2026 年 10 月 15 日）.....19

下载日志已记录，仅供内部参考

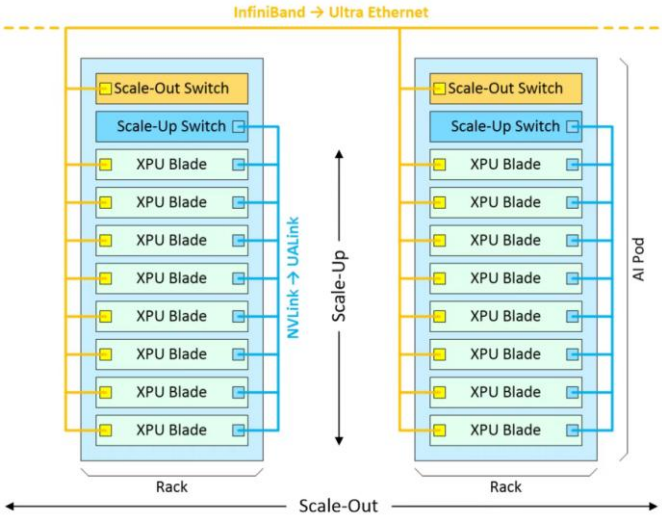
1、 “CPU-外设”+“GPU-GPU”构成了超节点内主要互连

超节点在算力需求量级增长背景下诞生，已成为支撑万亿参数大模型训练与推理的核心基石。随着大模型参数规模的不断增长，对 GPU 集群的规模需求也在不断增长。而当 AI 模型参数从亿级跃升至万亿级，传统服务器已难以满足日益增长的算力需求与效率要求。为了打破这个瓶颈，超节点（Super Node）这一全新概念应运而生。**超节点服务器是大量高速计算芯片的高效统一结构，它并非计算硬件的简单堆砌，而是将海量的计算单元（CPU/GPU/TPU 等）以前所未有的密度和效率整合在一起，专门为应对人工智能任务带来的庞大算力需求而生。**超节点的澎湃算力离不开其内部高速、低延迟的通信网络，这张网络即是超节点的“神经中枢”，其性能直接决定了整个系统的效率与上限。超节点是大规模 AI 集群的组成单元，而构建大规模 AI 集群，离不开两种基本的扩展策略：Scale-Up（纵向扩展）与 Scale-Out（横向扩展）：

Scale-Up：在单个计算节点或紧密耦合的计算单元内部，**通过增加计算单元（GPU、CPU）、内存等资源**，并采用极高带宽、极低延迟的内部互联技术，来提升该单元的整体计算能力和通信效率。**其目标是形成一个更为强大的“超级计算核心”，也就是所谓的“超节点”。**

Scale-Out：当单个计算单元的扩展达到物理或经济上的限制时，通过高速的外部网络将多个这样的计算单元连接起来，以构建更大规模的分布式计算集群。

图1：Scale-up& scale-out 在大规模 AI 集群构建中起到了核心互连的作用



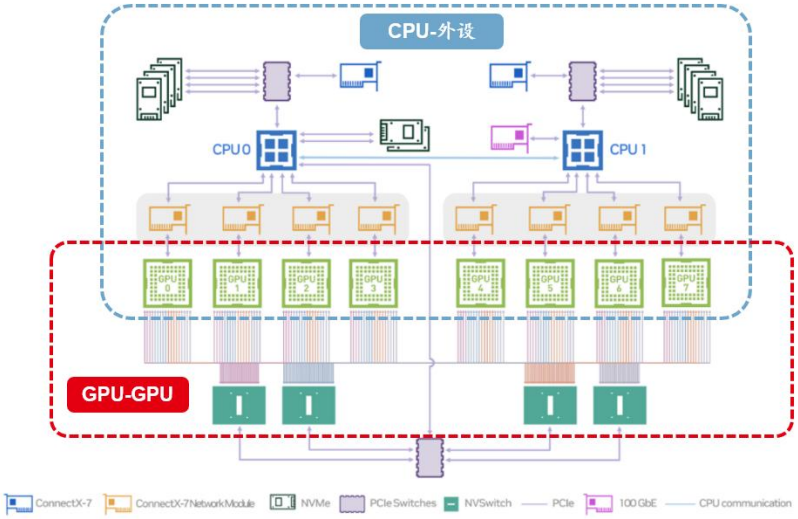
资料来源：半导体行业观察公众号

直击要点，从“柜内”角度解析超节点的互连方式。由上文总结可知，随着 AI 模型参数从亿级向万亿级跃升，AI 算力需求持续增长催生超节点，而超节点服务器就是大量高速计算芯片的高效统一结构，面对 AI 算力大潮的最优解。它并非计算硬件的简单堆砌，而是将海量的计算单元（CPU/GPU/TPU 等）以前所未有的密度和效率整合在一起，这需要 NVLink 等高速传输技术在 GPU 间搭建起超宽车道，让节点内的 GPU 之间能够“融为一体”，以最快的速度互相通信，同时也需要 CPU 在中心做全局互连和调度，以及网卡和固态硬盘等外设提供基础支持。由此，超节点内便以 CPU 和 GPU 为中心，构成了 CPU-外设以及 GPU-GPU 两种基础互连，前者承担了全局调度和指令下发，而后者即 sacle-up，担起了超节点 GPU 间高效互连。例如在下方英伟达的 DGX H100 系统中，将多颗 GPU、CPU、高速内存、NVLink/NVSwitch

下载日志已记录，仅供内部参考

等高度集成在一个优化设计的机箱内，这便可以看作一个典型的 8 卡“超节点”单元。接下来我们将围绕着这两个互连层级，分析各自使用的一般互连方式。

图2：（超）节点（柜）内互连模式可分为“CPU-外设”与“GPU-GPU”



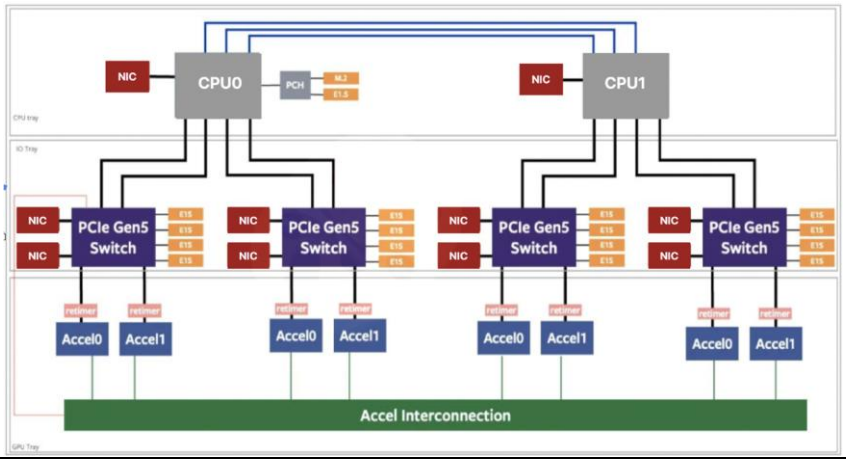
资料来源：英伟达、魔形智能 Magik Compute 公众号、开源证券研究所

2、 CPU-外设：决定超节点层级的基础组网

2.1、 PCIe：应用最广泛，生态最成熟的 CPU-外设的互连总线

PCIe 是通用高速串行总线，在 AI 服务器中具有最基础的互连作用。此前我们在专题(一)中详细探讨了以 PCIe 总线为核心的 AI 互连基础架构，且经由 PCIe Switch 的辅助，AI 服务器能够挂载更多的外设资源，由此可见 PCIe 作为通用标准，具有广泛兼容性和成熟的生态。所有 GPU（无论 NVIDIA、AMD 或国内 GPU）在与主机 CPU 通信时都支持 PCIe，是 GPU 与主板/CPU 连接的基础接口；PCIe 还承担着 CPU 同其他外设如网卡及固态硬盘的连接。PCIe 还具有丰富的软件支持，几乎所有操作系统和驱动默认支持 PCIe 设备，使得基于 PCIe 的系统集成最为便利。

图3：PCIe 是 AI 服务器柜内 CPU-外设最核心的互连总线



资料来源：《Hardware Evaluation with Telemetry Enablement for Grand Teton》

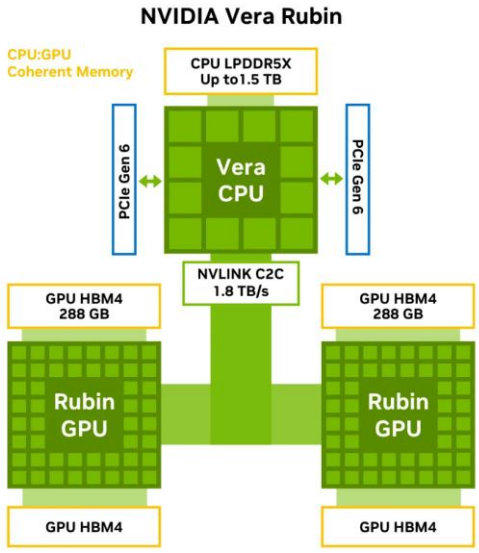
下载日志已记录，仅供内部参考

从当前部署来看，PCIe 依然是基础通信框架：即使使用 NVLink 的系统，GPU 也通过 PCIe 与 CPU 通信。因此从占有率看，PCIe 的部署率几乎是 100%，是所有 GPU 服务器都具备的底层总线。综合来看，PCIe 具有极强的普适性：标准开放，支持热插拔和灵活拓扑（可通过 PCIe Swich 扩展设备数），但受制于延迟且缺少跨设备的内存一致性（CXL 出现前），因此在当前的高端 AI 服务器中，PCIe 在柜内互连中更多的承担 CPU-外设的作用，关于 GPU-GPU 的互连我们会在下一章节详细说明。

2.2、 NVLink-C2C：针对 CPU-GPU 的高速 NVLink

英伟达采用 NVLink-C2C 来强化 CPU-GPU 间的互连。英伟达在 GPU 间 NVLink 基础上，推出了 NVLink-C2C 技术，能在英伟达的 CPU 和 GPU 之间实现高速低延迟传输。例如在英伟达当前的 NVL72 机柜中，NVIDIA Vera Rubin/GB200 超级芯片分别通过 NVLink-C2C 将两个 Rubin GPU 和一个 Vera CPU/两个 Blackwell GPU 和一个 Grace CPU 进行连接。如图 4 中第二代 NVLink-C2C 在 Vera CPU 与 Rubin GPU 之间提供 1.8 TB/s 的一致性带宽，实现 CPU 内存与 GPU 显存的统一地址空间。

图4：NVLink C2C 在 NVL72 机柜中起到 CPU-GPU 互连的作用



资料来源：英伟达官网

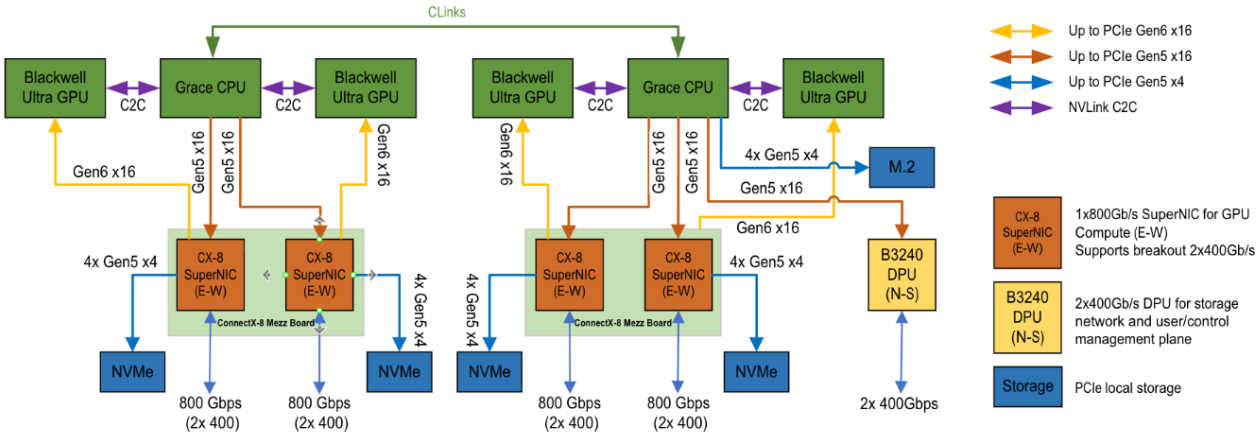
在英伟达体系内的 CPU-外设层级，PCIe 仍承担 CPU-GPU 之外的互连任务。根据英伟达官网展示的 GB300 NVL72 基础互连示意图可知，其计算托盘采用全集成 NVLink 机柜域扩展架构，在整机柜内每一块 GPU 都能够通过 NVLink 与其余所有 GPU 直连基础上，CPU 与 GPU 的连接也采用了 NVLink C2C 实现高速直连。然而 CPU-其他外设，以及 GPU 到其他外设则全部采用 PCIe 链路，如以 CPU 为中心的互连 CPU-网卡、CPU-固态硬盘以及 CPU-DPU，这一系列的 PCIe 互连链路通过内置 PCIe 6.0 Switch 的 CX-8 超级网卡承担，除了 CPU-外设的互连，CX-8 同样通过 PCIe 链路将 GPU 纳入互连体系。由此我们可得到两个结论：

（1）NVLink-C2C 应对的是 CPU 与 GPU 之间的高带宽互连，并将 CPU 与 GPU 紧密集成为一个“超级芯片”；

（2）PCIe 仍然承担了 CPU-其他外设以及 GPU-网卡等其他外设的核心互连作用且 PCIe Switch 集成于 CX-8 超级网卡中。

下载日志已记录，仅供内部参考

图5：PCIe 同 NVLink C2C 一起承担 NVL72 中的 CPU-外设互连

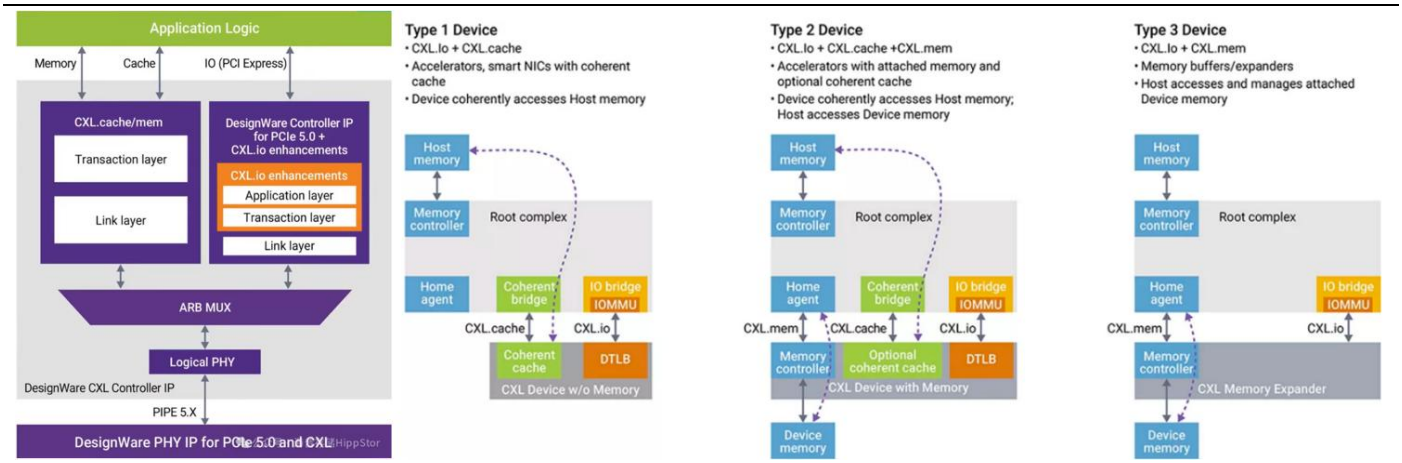


资料来源：英伟达官网

2.3、 CXL：建立在 PCIe 之上的跨设备缓存一致性高速互连标准

CXL 建立在 PCIe 物理层之上，提供了缓存一致（Cache-Coherent）和内存共享的协议，旨在高效连接 CPU 与加速器、内存设备等。CXL 支持三种协议，分别是 CXL.io、CXL.cache 和 CXL.memory。其中 CXL.io 主要用于传统的 I/O 操作，与 PCIe 类似；CXL.cache 和 CXL.memory 则提供了缓存一致性和内存访问能力，使得 CPU 与加速器或内存扩展设备之间能够共享和一致性地访问内存。这对于加速器（如 GPU、FPGA）而言尤为重要，因为它们可以更有效地访问系统内存，而不必通过缓慢的 I/O 通道。在现代计算机处理器和加速器之间的高带宽、低延迟通信需求下，CXL 技术的匹配性和优势要优于 PCIe。**虽然 CXL 最初聚焦于 CPU–设备互连，但其最新规范（CXL 3.0）已经支持多头设备（Multi-Head）和互联拓扑，意味着多个处理器/加速器可以通过 CXL 构成复杂的交换网络。**对于 GPU 而言，CXL 有望成为跨 CPU/GPU 之间共享内存、横向扩展加速器的一种通用方式。

图6：CXL 定义的三类设备及其 PHY、控制器与应用层



资料来源：Synopsys、开源证券研究所

CXL 依托 PCIe 底层带宽迭代完成版本升级，从 CXL 1.1 到 CXL 4.0 带宽持续翻倍，组网与内存共享能力逐层增强。当前 CXL 4.0 基 PCIe 7.0 再度翻倍带宽，强化内存一致性与共享能力。CXL 凭借开放标准和兼容 PCIe 生态的特性，具有广泛的

下载日志已记录，仅供内部参考

支持面，其缓存一致性支持使异构计算编程简化，对于多 GPU 系统，未来若 GPU 具备 CXL.cache 功能，则彼此间也能通过共享内存进行通信；CXL 还支持内存池化，这对 GPU 利用主存扩展容量、多个 GPU 共享高速存储器都有潜在好处。但当下 CXL 的带宽还受限于 PCIe 标准，远低于 NVLink 等专用方案，而且由于协议更复杂（涉及一致性维护），延迟开销可能偏高，不一定适合需要极致低延迟的 GPU 梯度交换。未来随着 CXL 3.0 硬件逐步推出（如支持 CXL 3.0 的 PCIe 6.0 交换芯片），有望看到 GPU 开始兼容 CXL，用于内存协同和跨主机互连的新模式。

表1：CXL 各代际说明及核心变化对比

CXL 版本	底层 PCIe 标准&x16 链路带宽	核心特点&功能概述
CXL 1.1	PCIe 5.0（32GT/s） 单向 64GB/s，双向 128GB/s	为直接连接到主机的设备引入了一致性和内存语义，单节点一致性互连，低延迟
CXL 2.0	PCIe 5.0（32GT/s） 单向 64GB/s，双向 128GB/s	实现了资源池功能，可在一段时间内将相同的资源分配给不同的主机；增加了热插拔、单级交换、内存 QoS、内存池、设备池和全局持续刷新（GPF）等功能
CXL 3.0	PCIe 6.0（64GT/s） 单向 121GB/s，双向约 242GB/s	实现跨主机共享一致的内存和信息传递；支持 Fabric 拓扑；近内存处理，允许在内存附近执行计算，以提高性能和能效
CXL 4.0	PCIe 7.0（128GT/s） 单向 242GB/s，双向近 500GB/s	CXL 捆绑端口、内存 RAS 增强；支持 CXL 设备间的高速数据传输，直接提升系统性能；提供内存可靠性、可用性和可维护性方面的增强功能

资料来源：Saniffer 公众号、半导体行业观察公众号、开源证券研究所

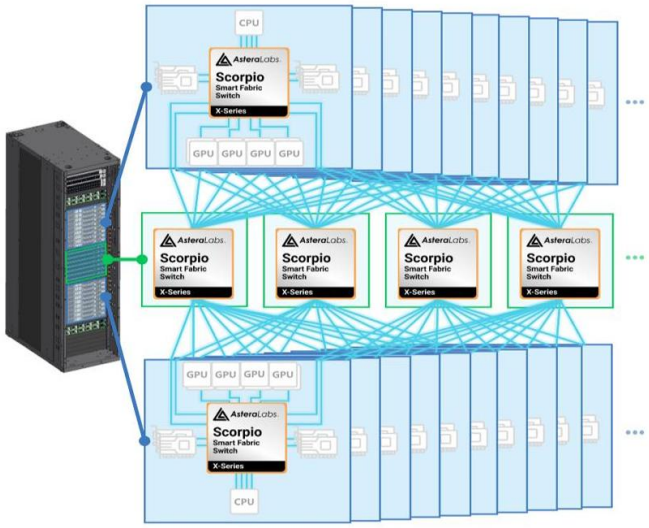
3、 GPU-GPU：利用多种私有/开放协议实现计算资源高效互连

3.1、 PCIe：在超节点层级实现 GPU-GPU 间大规模互连

PCIe&PCIe Switch 在超节点大规模 GPU 组网中依然是重要的选项。在多 GPU 服务器中，PCIe 链路依然可以经由 PCIe Switch 助力大规模 GPU 间的互连组网，实现超节点服务器的搭建。根据 Astera Labs 的 Scorpio Smart Fabric Switch（X-Series）应用示意图可推断，超节点内众多 GPU 通过强大的高通道 Scorpio Smart Fabric Switch（X-Series）组网，该芯片可提供高达 320x PCIe 通道，每个计算托盘内的 GPU 通过本托盘的 PCIe Switch 连接到位于机柜内中部交换托盘的 PCIe Switch，形成叶脊结构，实现超节点柜内所有 GPU 的 all-to-all 互连。

下载日志已记录，仅供内部参考

图7：利用 Scorpio X-Series 320 Lane 实现的 High Radix, Multi-level Switch Scale-up 组网



资料来源：Astera Labs 官网

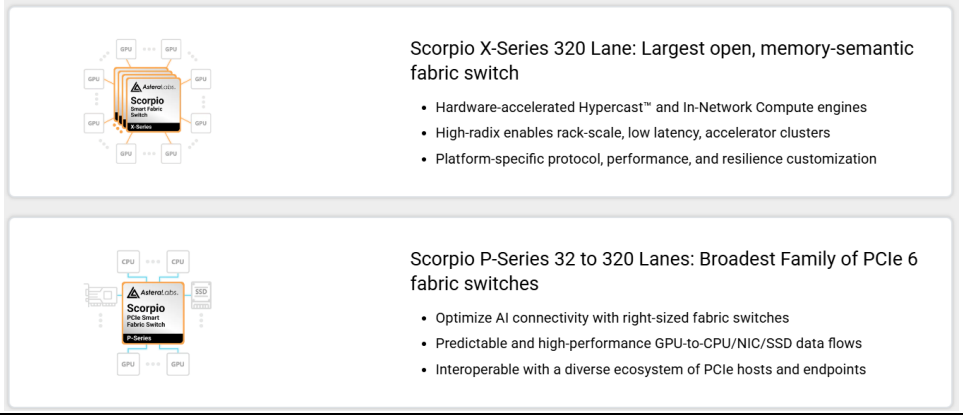
结合前文表述，我们认为 PCIe 在 AI 服务器逐步迈向超节点的时代，仍然承担着最核心的互连作用：

（1）CPU-外设（网卡/固态硬盘/GPU）：该层级下的互连需要 PCIe Switch 提供充足的 PCIe 链路实现，即使出现 NVLink C2C 及华为 UB 这样的私有互连，在一定程度取代了其体系内 CPU-GPU 的互连，但 CPU-网卡/固态硬盘以及 GPU 到其他外设依然需要依托 PCIe 链路。**综合来看，PCIe 依然是该层级下最广泛的互连方式，支撑着超节点 CPU 侧的基础组网；**

（2）GPU-GPU：PCIe 作为行业标准，具有极强的适用性，任何厂商的 GPU/加速卡都支持 PCIe 接口，PCIe 还具有丰富的软件支持（几乎所有操作系统和驱动默认支持 PCIe 设备），通过集中式交换架构和成熟的高通道 PCIe Switch 扩展，可以实现超节点内大规模 GPU 的组网。

正如 Astera Labs 公司 Scorpio PCIe Switch 两大系列典型应用所述：P-Series Scorpio PCIe Switch 实现 CPU-外设的超节点基础搭建，而 X-Series Scorpio PCIe Switch 则承担着超节点大规模 GPU 的组网，二者共同构成了 Astera Labs 在 AI 互连领域的产品矩阵，支撑超节点系统搭建。

图8：Astera Labs 公司 Scorpio P-/X-Series PCIe Switch 概览



资料来源：Astera Labs 官网

下载日志已记录，仅供内部参考

3.2、 NVLink：专属 NVIDIA 的服务器内部紧耦合 GPU 高速互连方案

应对 GPU-GPU 互连的专属高带宽互连协议，NVLink 已具有成熟且久经验证的体系。NVLink 是 NVIDIA 开发的高带宽 GPU 互连总线和通信协议，支持 GPU-GPU 以及 GPU-CPU 直接通信，自 2016 年推出以来已迭代多代，广泛应用于 NVIDIA 数据中心 GPU 产品中，自 Pascal 至 Blackwell 持续提高链路数量与速率。**NVLink 采用高速度差分 SerDes 点对点链路连接 GPU，最初设计用以超越 PCIe 带宽，满足 GPU 之间大量张量数据交换的需求。**NVLink 现已推出第六代，借助 Vera Rubin NVL72，第六代 NVLink 可在 72 GPU 域中为每个 GPU 提供 3.6TB/s 的 GPU 到 GPU 双向带宽和 260TB/s 的机架级 GPU 带宽。与基于现有以太网的替代解决方案相比，GPU 到 GPU 传输的端到端延迟低 3 倍，数据包速率高 10 倍。第六代 NVLink 可提供更大限度提高工厂产出所需的性能和弹性，并经过十多年的技术投资和成熟部署，包括世界领先的超大规模企业、云服务提供商（CSP）和超级计算中心。

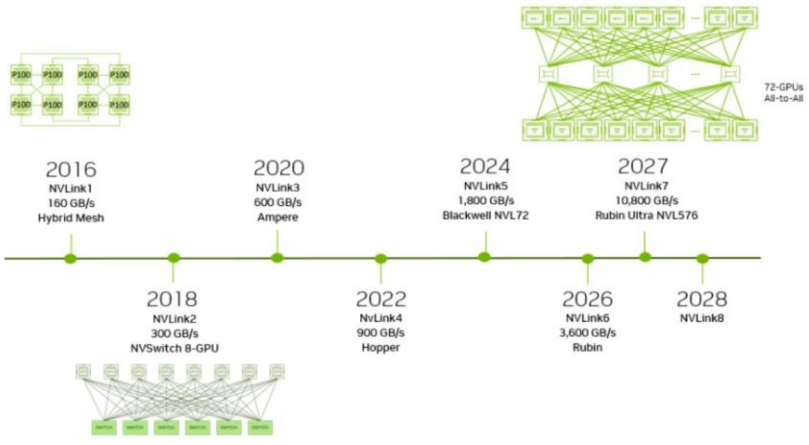
表2：NVLink 各代际速率及适配情况

型号	速率	适配情况
NVLink 1.0	每链路约 20GB/s；典型 4 链路总带宽约 80GB/s	Pascal P100 时代
NVLink 2.0	每链路约 25GB/s（双向）；最高总带宽约 300GB/s	Volta V100，6 条 NVLink 2 链路
NVLink 3.0	12 条链路；GPU 对 GPU 总带宽约 600GB/s	Ampere A100（80GB SXM4 版）
NVLink 4.0	100Gbps/通道；18 条链路总双向带宽约 900GB/s	Hopper H100；8 卡 HGX 系统
NVLink 5.0	18 条链路；每链路 100GB/s；单 GPU 总带宽约 1.8TB/s	Blackwell 架构 GPU
NVLink 6.0	单 GPU 总带宽约 3.6TB/s 和 260TB/s 的机架级 GPU 带宽	Rubin 架构 GPU

资料来源：Saniffer 公众号、开源证券研究所

NVSwitch 提供 NVLink 端口集中交换，实现每 GPU 对等互连。在 GPU 数量更多时，引入专用交换芯片 NVSwitch 构成星型/集群拓扑。2018 年，NVIDIA 推出了 NVLink Switch 技术，实现了在 8 个 GPU 的网络拓扑中每对 GPU 之间高达 300 GB/s 的 all-to-all 带宽，为多 GPU 计算时代的 scale-up 网络奠定了基础。随后，在第三代 NVLink Switch 中引入了 NVIDIA 可扩展分层聚合与归约协议（SHARP）技术，进一步提升了性能，有效优化了带宽性能并降低了集合操作的延迟。随着 2024 年第五代 NVLink 的发布，进一步增强的 NVLink Switch 支持 72 个 GPU 实现全互联通信，通信速率达 1800 GB/s，聚合总带宽高达 130 TB/s，较第一代产品提升了 800 倍。

图9：NVLink Scale-up 技术成熟，创新步伐加快

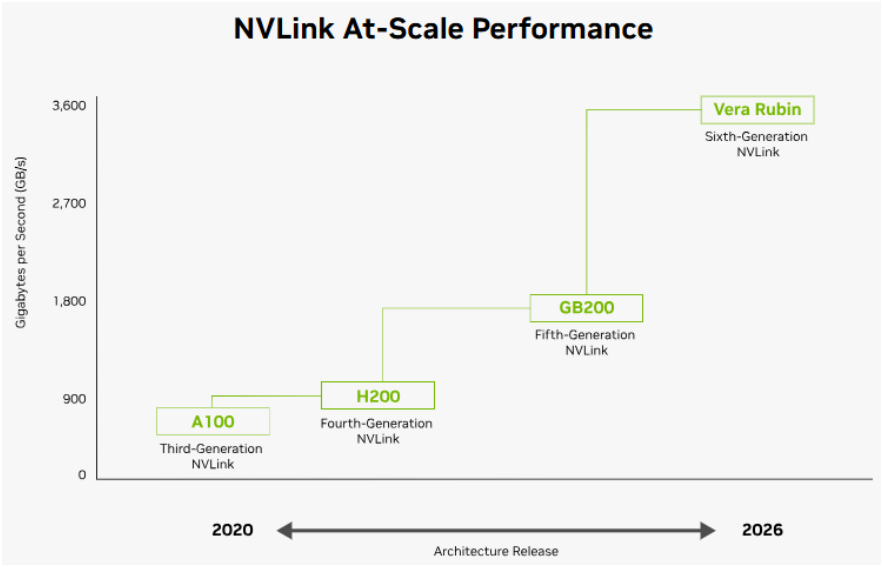


资料来源：NVIDIA DEVELOPER

下载日志已记录，仅供内部参考

NVLink/NVSwitch 是目前带宽最高、延迟最低的 GPU 互联方案之一，使 NVIDIA GPU 可以组成大规模共享内存域（如 DGX 系列系统）。但 NVLink 仅适用于 NVIDIA 自家 GPU/CPU，不兼容第三方硬件。这种锁定生态使系统灵活性受限，且 NVLink 需要特殊物理接口（如 NVIDIA SXM 模块）和交换芯片支持，增加成本。尽管如此，依托 NVIDIA 在 AI 市场的主导地位，NVLink 的部署占有率在专用 GPU 互联中位居第一。

图10：NVIDIA 核心 GPUxNVLink 架构演变史



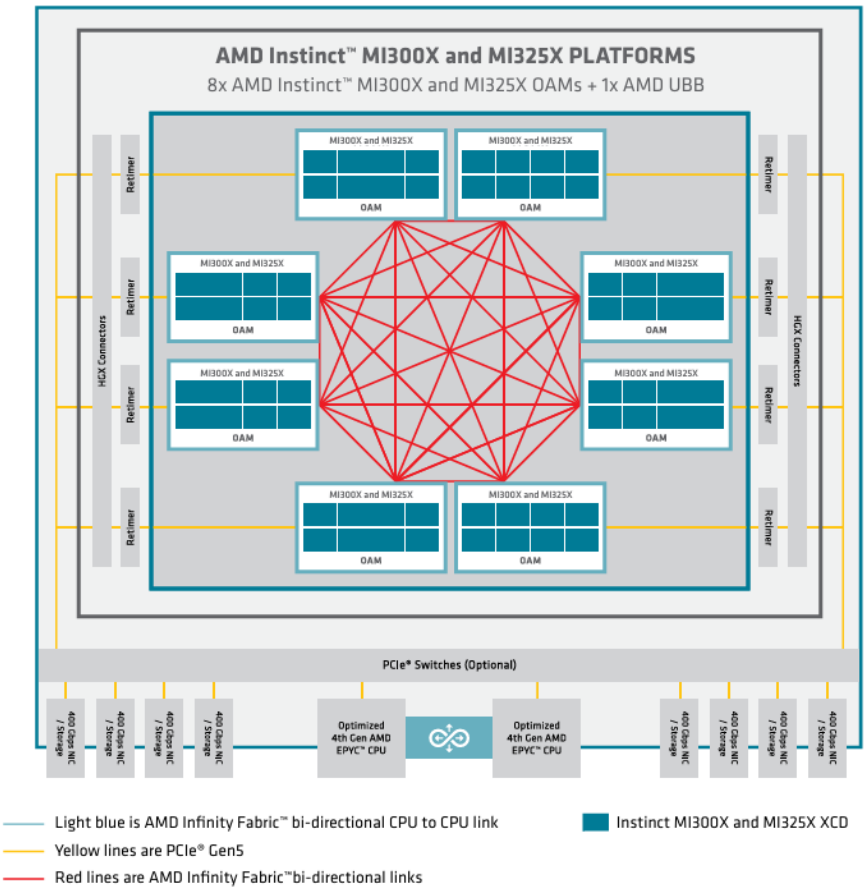
资料来源：NVIDIA 官网

3.3、 Infinity Fabric：缓存一致与统一显存提升多 GPU 协同效率

Infinity Fabric 是 AMD 开发的片上/片间高速互连架构，广泛用于 AMD CPU 和 GPU 内部。针对 GPU 对 GPU 互联，AMD 提供了 XGMI（External Global Memory Interconnect）协议，构建在 Infinity Fabric 之上，XGMI 能够提供与 NVLink 类似的高速直连，但更强调缓存一致性和内存共享：AMD 的多 GPU 方案构建“hive”（蜂巢）架构，使多 GPU 共享统一的分布式 HBM 显存池，实现单一内存空间下的协同计算。这类似于把多 GPU 当作一个统一内存的计算单元，简化并行编程模型。以 MI300 系列为例，每块 MI300X GPU 具有最多 7 条 Infinity 链路，每条链路为 16 通道，第四代 Infinity Fabric 速率约 32 Gbps/通道，折合每条 XGMI 链路双向带宽约 128 GB/s（单向约 64 GB/s）。这比 PCIe Gen5 x16 的 64 GB/s 单向带宽高出一倍。多个 MI300 GPU 可以通过 XGMI 组成所谓“hive”架构，所有 GPU 共享各自 HBM 组成的统一内存池。

下载日志已记录，仅供内部参考

图11：AMD Instinct MI300X&MI325X 平台利用 Infinity Fabric 实现 8GPU 全互连



资料来源：AMD CDNA 3 白皮书

Infinity Fabric/XGMI 最大的优势是原生缓存一致性和异构内存共享。XGMI 不仅传输数据，还携带缓存协议，使多 GPU 可以像多 CPU 那样保持内存一致，实现更灵活的编程模型（例如直接执行 Load/Store 访问远端 GPU 内存）。在 AMD 的 ROCm 软件栈中，XGMI 结合 RCCL 通信库，可高效实现 GPU 的集合通信和一致内存操作。此外，AMD 正推动 Infinity Fabric 的开放化。但受限于 AMD 在 AI 加速器市场份额相对有限，XGMI 的生态普及率较低，仅 AMD Instinct 加速卡和配套平台支持，因此 XGMI 部署量远不及 NVLink。不过随着 MI300 等产品在超算和云服务中的应用增长，以及开放标准 UALink 的推进，Infinity Fabric 有望被更多厂商和系统采用，成为对抗 NVLink 的一支重要力量。

3.4、华为 Ascend UB 互连：面向昇腾平台形成自主可控高速互连

Ascend UB（UnifiedBus，灵衢）是华为昇腾面向超节点 AI 加速器采用的专有高速互连，将 IO、内存访问和各类处理单元间的通信统一在同一互联技术体系，实现高性能数据搬移、资源统一管理、资源灵活组合、处理单元间高效协同和高效编程，用于连接多颗 Ascend 处理器并构建大规模训练集群。UB 协议形成了六大核心特征，这些特征通过以下设计得以实现：

（1）总线级互联：满足带宽和时延的总线级要求，支持 112G/224G 速率并探索工程极限，通过多端口聚合提供更高带宽。

（2）平等协同：采用 EID/CNA 两层地址体系，实现互联域内直接点对点通信，

下载日志已记录，仅供内部参考

免协议转换，极大降低时延。

（3）全量池化：将 UBPU 中的能力以 Entity 粒度分配给其他 UBPU 使用，实现设备级虚拟化共享，资源利用率更高。

（4）协议归一：整合多套协议为一套，消除数据路径转换开销。

（5）大规模组网：支持 1D/2D/CLOS 等多种拓扑，工程上可达 8192 节点以上；同时支持以太、UBoE 等组网方式。

（6）高可用性：针对光模块多发故障设计换路重传、链路重传等系统级机制，提升超节点系统可靠性。

图12：UB 协议架构分层技术说明



资料来源：灵衢 UnifiedBus 公众号

在 CloudMatrix384 超节点中，总计部署 16 个机柜：其中 12 个为计算机柜，合计部署 48 台昇腾 910C 节点（总计 384 颗 NPU），每颗昇腾 910C 裸片对接两套相互独立的网络平面：

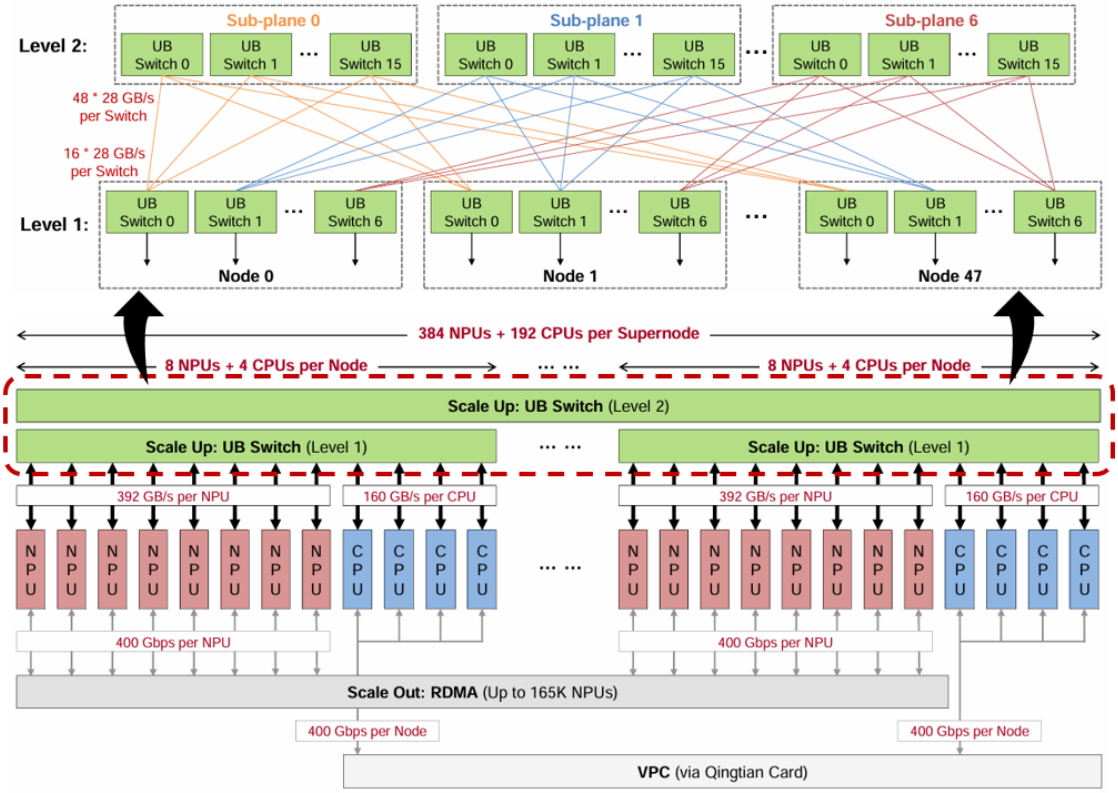
（1）UB 平面：单节点 7 颗板载一级（L1）UB Switch（部署于每台昇腾 910C 节点内部）为纵向扩展 UB 平面提供合计 196 GB/s 单向带宽（双向带宽为 392 GB/s）；

（2）RDMA 平面：每颗裸片配备专用接口，为横向扩展 RDMA 平面提供最高 200Gbps 的单向带宽；

另外 4 个为通信机柜，通信机柜部署二级（L2）UB Switch，用于完成超节点内部全部节点之间的互连。L2 UB Switch 被划分为 7 个独立子平面。每个子平面包含 16 颗 L2 UB Switch，单颗 L2 UB Switch 提供 48 个 28 GB/s 端口。每个节点内部的 7 颗板载 L1 UB Switch 与这 7 个 L2 子平面一一对应。每颗 L1 UB Switch 向外扩展出 16 条链路，分别连接对应子平面内的每一颗 L2 UB Switch。该配置保证节点通往 L2 交换架构的总上行带宽与自身内部 UB 带宽能力完全匹配，使整个超节点始终维持无阻塞特性。

下载日志已记录，仅供内部参考

图13：CloudMatrix384 超节点的对等直连硬件架构及 UB 交换系统示意图



资料来源：《Serving Large Language Models on Huawei CloudMatrix384》、开源证券研究所

灵衢 2.0 高速互联架构作为昇腾 950 系列算力集群核心网络底座，通过高效的卡间通信全面助力昇腾 950 系列芯片性能的释放。目前，昇腾 950 系列芯片及 Atlas 950 SuperPoD 的发布，灵衢 2.0 高速互联架构作为集群核心网络底座，全面升级 UB 专属总线，带宽大幅提升，原生支持 1024 卡超大超节点组网。该架构依托超大卡间带宽、超低跨卡时延及全域统一池化，解决 MoE 推理、长序列训练、强化学习交互场景的通信互联瓶颈难题。此外，灵衢 2.0 采用 UB-Mesh+Clos 混合拓扑与统一 UBoE 协议，针对 LLM 预训练、多智能体等互联网主流业务可显著提升性能。

图14：华为昇腾新一代 Atlas 系列产品



资料来源：华为中国政企业务公众号

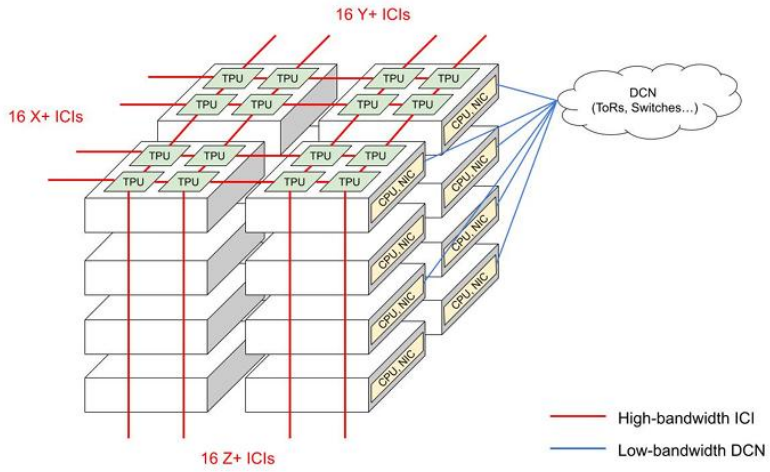
下载日志已记录，仅供内部参考

3.5、 XLink 及其他方案：定制化提升局部性能但生态碎片化明显

XLink 并非统一正式标准，而是对多种自定义高速芯片直连技术的概括。目前已公布的 XLink 包括阿里 xLink C2C、Graphcore IPU-Link、谷歌 TPU ICI Link，以及 Enflame、寒武纪等 AI 芯片的多芯片互连。以上这些由公司和研究机构推出的各自 GPU/AI 加速器互连技术，虽然名称不同，但核心思想都是通过定制协议充分利用硬件特性，实现直连高带宽和低延迟，绕开通用总线的性能瓶颈。

XLink 的最大特点是设计灵活，但通用性差的痛点使其难以广泛推广。这类专有互连可以实现特定产品优化，性能发展潜力大，同时设计灵活，可以根据需要定义链路宽度、拓扑（环形、网格等）和协议层。例如在谷歌 TPUv4 中，每一个独立 TPU 计算单元包含一个 CPU 托盘和一个 TPU 托盘，二者通过 PCIe 互连。**而每个 TPU 托盘搭载 4 颗 TPUv4 芯片，构成 2x2x1 的 ICI 网格；16 个 TPU 计算单元共同组成一个数据中心机柜；机柜内部的 ICI 链路通过 ICI 互连形成 4x4x4 网格，该整套硬件集合即为一个 cube。**

图15：谷歌 TPUv4 中 TPU 通过 ICI 沿着 x/y/z 三个维度完成互连形成 4x4x4 结构



资料来源：《Resiliency at Scale: Managing Google's TPUv4 Machine Learning Supercomputer》

然而站在互连协议的通用性上看，~~各厂~~商的“XLink”彼此不兼容，难以形成统一生态，且由于缺乏行业标准约束，有些私有互连在可靠性、容错等方面成熟度需要长期验证。

3.6、 SUE/UALink/ETH+/…：依托于成熟的以太网生态的多种开放 scale-up 互连方案

基于以太网构建开放的 Scale-up 互连方案，正逐步成为业界的共识和主流选择。随着模型参数量的持续增长叠加 GPU 卡间高速互联带宽需求持续提升，Scale-up 域的规模也在不断扩大。Scale-up 层级的互连从早期的以 NVIDIA NVLink、AMD Infinity Fabric 等私有协议主导市场，到如今基于以太网的开放方案快速崛起，这一转变反映了业界对开放生态、成本效益和供应链多元化的强烈诉求。而以太网则凭借其成熟的产业链、丰富的供应商选择和持续演进的带宽能力（单端口吞吐从 100G 向 800G 乃至 1.6T 迈进），正在从传统的数据中心网络向服务器内部高速互联深度延伸。基于以太网构建开放的 Scale-up 互连方案，正逐步成为业界的共识和主流选择。

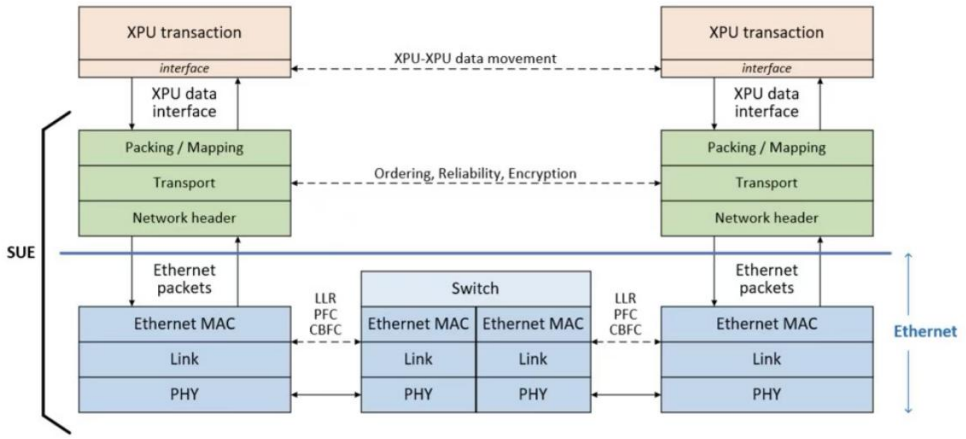
下载日志已记录，仅供内部参考

在此背景下，业界涌现出了多种开放 scale-up 互连协议，它们有的基于以太网增强如 ETH+/SUE/ESUN，有的则采用全新设计如 UALINK；本小节将选取 SUE 和 UALink 进行阐述。

3.6.1、 SUE：利用优化以太网协议栈，为 XPU 集群提供高效的互连解决方案

SUE（Scale-Up Ethernet）是由博通主导的旨在利用以太网生态构建 AI 系统内部 Scale-Up 互联的开放框架。SUE 利用以太网的高速 SerDes、交换芯片和生态基础，针对 Scale-Up 场景重新简化协议与数据路径。SUE 的目标是为 XPU 集群提供超高带宽（9.6Tbps）、超低延迟（RTT＜2μs）及无损传输的高性能通信能力，以支持千卡级规模的扩展。SUE 支持“双版本”交付模式（即完整版 SUE 与 SUE Lite）。标准 SUE 具备对 E2E 可靠传输的支持，可以支持多层网络互联拓扑和超大规模组网。而 SUE Lite 则是去除了可靠传输层，将可靠性保障下放给交换机和网卡的 LLR（链路层重传）机制。综合来看，SUE 是在保持与现有以太网基础设施兼容的前提下，打破私有协议壁垒，为大规模 AI 智算提供一种高效、开放且低成本的互连标准。

图16：标准 SUE 协议栈及工作流程示意图

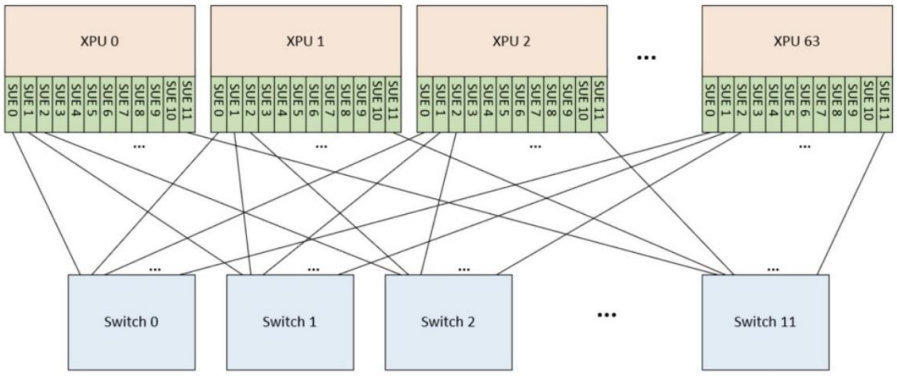


资料来源：SDNLAB 公众号

SUE 在多 XPU 系统中提供 XPU 间的通信，将 XPU 发出的内存类事务高效送到另一颗 XPU。SUE 提供通用的命令与响应通道，具体命令由 XPU 自己定义，SUE 并不理解上层业务语义。每个 SUE 实例可配置为 1、2 或 4 个端口。例如，一个 800G SUE 实例支持 1 个 800G 端口、2 个 400G 端口或 4 个 200G 端口。端口配置的选择基于所需的交换机基数、SerDes 数量以及冗余和故障转移考量。SUE 的设计支持每个 XPU 部署多个 SUE 实例。例如，单个 XPU 可能包含 8 个或 16 个 SUE 实例。图 18 展示了一个 64 个 XPU 的示例配置，每个 XPU 配备 12 个 800G SUE 实例，并使用 12 台各有 64 个端口的以太网交换机。在该示例中，通过全交换连接，任意一对 XPU 之间的带宽可达 9.6Tbps（12×800G）。此外，SUE 还可用于网状拓扑结构，如图 18，8 个 XPU 直接互连形成 Mesh 部署模式。

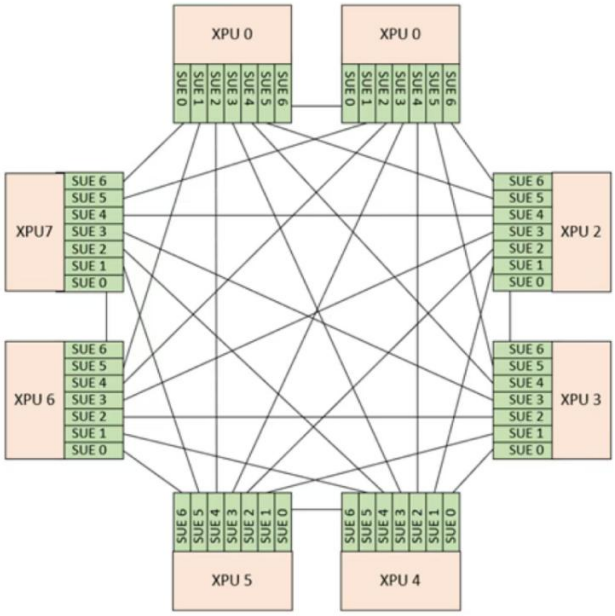
下载日志已记录，仅供内部参考

图17：XPU 通过部署 12 个 SUE 并使用 12 台 64 个端口的以太网交换机实现全交换连接（示例）



资料来源：SDNLAB 公众号

图18：XPU 利用 SUE 直接互连网状拓扑结构（示例）



资料来源：SDNLAB 公众号

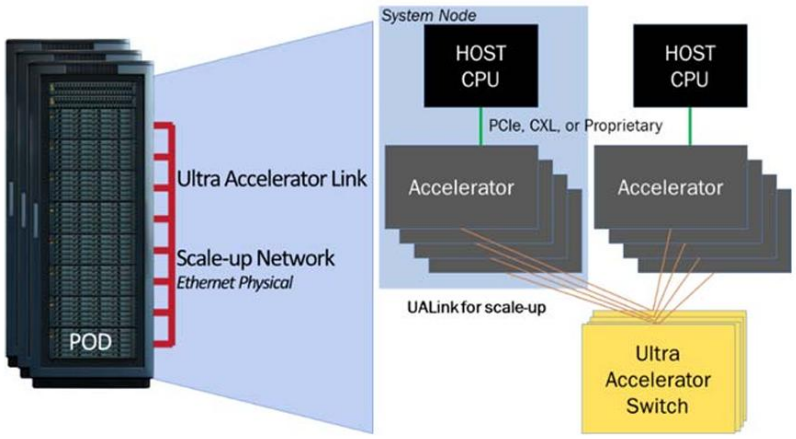
3.6.2、 UALink：全新开放的超节点内部紧耦合高速互连方案

依托于以太网基础设施之上的开放互连规范。UALink（Ultra Accelerator Link）是 2024 年由 AMD、Broadcom 牵头，与 Cisco、谷歌、Intel、Meta、微软等共同成立联盟制定的全新开放互连规范，目标是创建一个多厂商支持的生态系统，以应对日益增长的 AI 模型和复杂计算任务对加速器间通信的极致需求。UALink 直接对标 NVIDIA 的 NVLink/NVSwitch，能够提升双向内存访问的链路效率，以实现最大化数据带宽，同时复用现有以太网基础设施（含线缆、连接器、信号重定时器、管理软件），降低整体拥有成本（TCO）；UALink 还可借助内存语义（支持直接读、写、原子事务），并对主机挂载内存、本地加速器内存、远端加速器内存维持统一的访存序模型，以此降低软件层面的复杂度。

下载日志已记录，仅供内部参考

UALink 强调开放性和避免厂商锁定。如前所述 UALink 的成员覆盖云计算公司、CPU/GPU 厂商、交换芯片厂商等，旨在让不同厂商的加速器都可以通过 UALink 互连。这意味着，将来可能出现第三方 UALink 交换机、适配不同厂商 GPU 的 UALink 接口，从而建立一个统一的加速器互连生态。

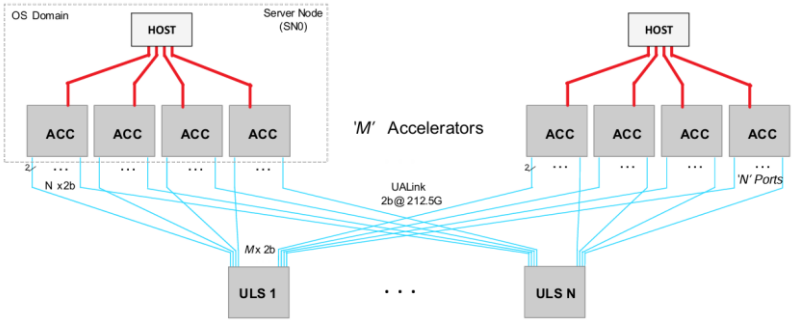
图19：基于 UALink 的 Scale-Up 机柜架构中的多节点与多平面交换



资料来源：Nathan Kalyanasundharam:《Introducing UALink 200G 1.0 Specification》

UALink 可提供可扩展的互连性能，匹配算力、内存带宽与内存容量的增长需求。通过支持多节点系统以及多平面交换生态体系，实现互连性能的可扩展能力。2025 年 4 月 UALink 1.0 规范正式发布，其采用 200Gbps/lane 的以太网物理层，每端口可由多条 lane 组成，并设计了专用 UALink Switch 用于互连，最多可连接 1024 个加速器或端点设备。每个加速器分配唯一的 10 位路由标识符。这些互连的加速器共同构成一个 Scale-Up Pod。UALink Switch 的每个端口分别对接一个独立加速器，其端口数量与所接入加速器的数量保持一致。协议不直接沿用 TCP/IP 或 RDMA，而是以精简的数据格式支持读、写、原子等内存语义，追求亚微秒级延迟和跨厂商加速器、交换机互操作。

图20：基于 UALink 的多加速器系统



资料来源：Nathan Kalyanasundharam:《Introducing UALink 200G 1.0 Specification》

下载日志已记录，仅供内部参考

4、 投资建议

综上我们认为，随着全球超节点形态的确认及 AI 算力集群进入加速建设阶段，以各类 AI 互连协议配套交换芯片为核心的 AI 互连板块将逐步成为继算力和存力的 AI 产业第三增长极。看好卡位关键节点且占据先发优势的国产核心替代者。推荐标的：万通发展、盛科通信、澜起科技。

表3：核心推荐标的汇总表（截至 2026 年 10 月 15 日）

公司代码	公司简称	评级	股价	市值	总营收（亿元）			PS（倍）		
			元	亿元	2026E	2027E	2028E	2026E	2027E	2028E
600246.SH	万通发展	买入	11.39	222.79	10.98	16.65	21.28	20.3	13.4	10.5
688702.SH	盛科通信-U	买入	236.66	970.31	12.14	12.58	13.06	79.9	77.1	74.3
公司代码	公司简称	评级	股价	市值	归母净利润（亿元）			PE（倍）		
			元	亿元	2026E	2027E	2028E	2026E	2027E	2028E
688008.SH	澜起科技	买入	163.00	1,997.14	38.76	49.09	63.72	51.5	40.7	31.3

数据来源：Wind、开源证券研究所（万通发展、盛科通信总营收、澜起科技归母净利润数据采用开源研究所预测）

5、 风险提示

（1）AI 算力建设不及预期：AI 互连协议的发展依托 CSP 厂商及互联网厂商持续的资本开支以及 AI 算力集群的扩张，若终端市场发展缓慢则不利于 AI 互连协议的持续进步；

（2）技术进步/标准落地不及预期：无论是厂商私有生态抑或是 PCIe、UALink、CXL 等开放互连标准仍处于持续迭代推广阶段，若协议完善、厂商兼容适配进度滞后，相关互连产品落地与客户规模化采购也将迟滞；

（3）生态适配不及预期：当前主流 AI GPU 厂商对开放互连方案适配意愿有限，若头部厂商持续锁定自有私有互连技术，开放互连产品市场需求空间将持续受限。

下载日志已记录，仅供内部参考

特别声明

《证券期货投资者适当性管理办法》、《证券经营机构投资者适当性管理实施指引（试行）》已正式实施。根据上述规定，开源证券评定此研报的风险等级为R4（中高风险），因此通过公共平台推送的研报其适用的投资者类别仅限定为专业投资者及风险承受能力为C4、C5的普通投资者。若您并非专业投资者及风险承受能力为C4、C5的普通投资者，请取消阅读，请勿收藏、接收或使用本研报中的任何信息。

因此受限于访问权限的设置，若给您造成不便，烦请见谅！感谢您给予的理解与配合。

分析师声明

本研究报告的署名人员具有中国证券业协会授予的证券投资咨询执业资格或相当的专业胜任能力，以勤勉的职业态度，独立、客观地出具本报告，并对内容和观点负责。本报告清晰准确地反映了署名人员的研究观点，所包含的分析基于各种假设，不同假设可能导致分析结果出现重大不同。本报告采用的各种估值方法及模型均有其局限性，估值结果不保证所涉及证券能够在该价格交易。本报告署名人员保证他们报酬的任何一部分不曾与，不与，也将不会与本报告中具体的推荐意见或观点有直接或间接的联系。

股票投资评级说明

	评级	说明
证券评级	买入（Buy）	预计相对强于市场表现 20%以上；
	增持（Outperform）	预计相对强于市场表现 5%～20%；
	中性（Neutral）	对市场表现在－5%～＋5%之间波动；
	减持（Underperform）	预计相对弱于市场表现 5%以下。
行业评级	看好（Overweight）	预计行业超越整体市场表现；
	中性（Neutral）	预计行业与整体市场表现基本持平；
	看淡（Underperform）	预计行业弱于整体市场表现。
备注：评级标准为以报告日后的 6~12 个月内，证券相对于市场基准指数的涨跌幅表现，其中 A 股基准指数为沪深 300 指数（北交所基准指数为北证 50 指数）、港股基准指数为恒生指数、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）、美股基准指数为标普 500 或纳斯达克综合指数。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取决于个人的实际情况，比如当前的持仓结构以及其他需要考虑的因素。投资者应阅读整篇报告，以获取比较完整的观点与信息，不应仅仅依靠投资评级来推断结论。		

下载日志已记录，仅供内部参考

法律声明

开源证券股份有限公司是经中国证监会批准设立的证券经营机构，已具备证券投资咨询业务资格。

本报告仅供开源证券股份有限公司（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。本报告是发送给开源证券客户的，属于商业秘密材料，只有开源证券客户才能参考或使用。

本报告是基于本公司认为可靠的已公开信息，但本公司不保证该等信息的准确性或完整性。本报告所载的资料、工具、意见及推测只提供给客户作参考之用，并非作为或被视为出售或购买证券或其他金融工具的邀请或向人做出邀请。本报告所载的资料、意见及推测仅反映本公司于发布本报告当日的判断，本报告所指的证券或投资标的的价格、价值及投资收入可能会波动，过往的业绩表现不应作为其日后表现的预示。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。客户应当考虑到本公司可能存在可能影响本报告客观性的利益冲突，不应视本报告为做出投资决策的唯一因素。本报告中所指的投资及服务可能不适合个别客户，不构成客户私人咨询建议。本公司未确保本报告充分考虑到个别客户特殊的投资目标、财务状况或需要。本公司建议客户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。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。在任何情况下，本公司不对任何人因使用本报告中的任何内容所引致的任何损失负任何责任。若本报告的接收人非本公司的客户，应在基于本报告做出任何投资决定或就本报告要求任何解释前咨询独立投资顾问。投资者应自主作出投资决策并自行承担投资风险，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。

开源证券在法律允许的情况下可参与、投资或持有本报告涉及的证券或进行证券交易，或向本报告涉及的公司提供或争取提供包括投资银行业务在内的服务或业务支持。开源证券可能与本报告涉及的公司之间存在业务关系，并无需事先或在获得业务关系后通知客户。

本报告的版权归本公司所有。本公司对本报告保留一切权利。未经本公司事先书面授权，本报告的任何部分均不得以任何方式制作任何形式的拷贝、复印件或复制品，或再次分发给任何其他人，或以任何侵犯本公司版权的其他方式使用。所有本报告中使用的商标、服务标记及标记均为本公司的商标、服务标记及标记。

开源证券研究所

上海
地址：上海市浦东新区世纪大道1788号陆家嘴金控广场1号楼3层
邮编：200120
邮箱：research@kysec.cn

深圳
地址：深圳市福田区金田路2030号卓越世纪中心1号楼45层
邮编：518000
邮箱：research@kysec.cn

北京
地址：北京市西城区西直门外大街18号金贸大厦C2座9层
邮编：100044
邮箱：research@kysec.cn

西安
地址：西安市高新区锦业路1号都市之门B座5层
邮编：710065
邮箱：research@kysec.cn