

Marvell：AI 连接与定制算力双轮驱动，CXL 与硅光开启增长新周期

摘要

- AI 业务驱动 Marvell 转型，数据中心季度营收已超 20 亿美元，AI 连接与算力需求推动公司规模增长超 5 倍。
- 连接技术进入爆发期，通过 DSP、硅光及宽带模拟组件解决机架间通信瓶颈，CPO 与近封装光学技术即将大规模落地。
- 内存扩展（CXL）成为核心战略，通过解耦 DRAM 提升推理场景 KV 缓存能力，预计未来市场空间达百亿美元规模。
- 定制 XPU 渗透率持续提升，Marvell 通过 Avera 团队深度绑定超大客户，定制化与商用芯片将长期共存并实现互操作。
- XPU 附加组件业务势头强劲，涵盖定制网卡、安全及内存芯片，单台加速器内容价值达 500 至 2000 美元。
- AI 革命仍处早期，未来增长点在于跨规模扩展（Scale-out）及通过 DCI 技术实现跨数据中心集群的协同连接。

Q&A

从宏观层面来看，当前正处于历史上规模最大的基础设施建设浪潮中，您如何看待这场正在发生的转型？尽管当前规模已十分庞大，市场是否仍在低估其长期影响？

我们正处于部署的极早期阶段。Marvell 在过去十年间进行战略转型，聚焦于我们自定义的数据基础设施市场。当时的核心判断是，数十亿设备产生的海量数据需要被处理、变现、传输、流转和存储，而云计算和数据中心技术正迅速兴起。这项业务在转型初期占公司营收不足 10%，未来将超过 80%，公司规模在此期间增长了五倍以上。AI 已成为数据基础设施领域的杀手级应用。自 2022 年底 ChatGPT 推出以来，市场一直在关注增长的可持续性，但从行业实践来看，我们仍处于部署的初期。更重要的是，整个生态系统在技术上仍处于将 AI 扩展至所需规模的早期阶段。目前市场焦点集中在算力竞争，即 GPU、XPU 和定制 ASIC 芯片的性能比拼，同时内存和存储领域也经历了一轮显著的超级周期。接下来的趋势是释放连接能力，通过互联技术实现内存、算力及所有数据在芯片间、机架

内、跨机柜、网络上层乃至数据中心之间的流动，从而最终惠及消费者。我们目前正处于这一趋势的开端。

Marvell 在五年收购 Inphi，布局光子学领域，这与当前行业热议的“铜墙”等连接瓶颈问题高度相关。请问当时看到了怎样的机遇，如何看待从电连接到光连接的转型趋势，以及共封装光学等技术的发展路径和实现时间？

我们在 2021 年 4 月完成了对 Inphi 的收购，但早在 2016 年 12 月，即我担任 CEO 约五个月后，就开始关注这家公司。2020 年时机成熟，当时数据中心内部光通信正处于一个转折点，基于 PAM 调制技术的数字信号处理器开始大规模应用，这是推动下一代高频光通信所必需的架构。我们的 DSP 业务表现出色，但模块中还需要高性能模拟器件，如基于硅锗技术的跨阻放大器和驱动器。这些宽带模拟组件不仅是光基 DSP 的关键，在近封装光学和共封装光学的线性驱动中也至关重要。通过收购 Inphi，我们获得了这项技术，以及硅光技术。Marvell 与 Inphi 联合设计并量产的硅光技术已在数据中心互联等长距离应用中交付数百万台设备，积累了 150 亿小时的可靠性数据。当前，硅光、宽带模拟组件和 DSP 等技术不仅是构建横向扩展网络的基础，也是纵向扩展乃至向内扩展的核心。为加速发展，我们在 2025 年底收购了拥有定制化 CPO 和光子架构解决方案的 Celestial AI，并整合了双方团队。目前，Marvell 拥有业内最广泛、最多元的硅光和光器件团队。这些技术并非独立产品，它们可以与自研的定制芯片或第三方的 GPU 集成，并通过以太网、UALink 或 NVLink 交换机传输数据，从而驱动数千乃至未来数百万个 GPU 之间的通信。

除了连接性瓶颈，内存墙问题也是一大挑战，尤其是在推理规模不断扩大的背景下。从您的角度看，为什么内存墙是一个如此棘手的问题？

内存墙问题确实是一个重大挑战。

贵公司在内存扩展技术方面有哪些布局，该技术的发展动力和市场前景如何？

公司在内存扩展技术领域拥有完全自主研发的技术。该项业务的启动早于当前的内存短缺问题，最初的尝试是基于大约五年前出现的 CXL 行业标准。CXL 技术最初为标准服务器设计，旨在解决传统 CPU 内存控制器端口数量固定，导致用户若想增加 DRAM 就必须购买更多 CPU 的限制。其核心是通过 CXL 内存扩展器或内存池设备，实现内存解耦和更大容量的内存组合，无需随内存扩容同步增加 CPU。当前，AI 的发展，特别是在推理场景中，XPU 需要配备远超常规比例的内存用于 KV 缓存，这极大地加速了此类技术的应用。同时，目前的内存短缺问题也促使客户加速采纳我们的解决方案，因为通过集成内存扩展能力，可以减少所需采购的内存总量。内存扩展已成为公司的核心战略之一，我们提供包括 CXL 内存扩展器、CXL 交换机和 CXL 重定时器在内的端到端解决方案，涵盖定制化方案和标准线。尽管 x86 应用因 AI 兴起而有所放缓，但内存扩展业务反而进入了加速阶段，预计未来将成长为价值数十亿乃至上百亿美元的业务，公司在此领



备产品和市场的领导地位。

对于超大规模数据中心持续投资定制芯片的趋势，以及贵公司在该领域的定位和业务模式，您如何解读？

定制芯片领域的演变非常显著。公司通过 2019 年收购源自 IBM 定制芯片设计团队的 Avera Semiconductor 进入该业务，并将其接入台积电的技术平台，迅速在数据中心市场赢得了多个定制硅芯片订单。几年前市场还在质疑定制 XPU 能否规模化生产和部署，但如今，我们和其他大型同行已经将非常复杂的定制 XPU 投入生产，并成功应用于训练和推理工作负载。两年前曾预测定制硅芯片的市场渗透率约为 25%，现在普遍认为实际数量占比可能会更高。这背后的原因是客户比任何人都更了解自身的工作负载，他们发现对于部分算力集群，自行设计芯片能在经济、技术及架构上带来显著优势。定制化与商用芯片将长期共存，而非完全取代。今年（2026 年），NVIDIA 向我们投资了 20 亿美元，协议的一部分允许我们在定制产品中使用其丰富的知识产权，以实现与商用产品的互操作性，这表明市场正朝着支持最优技术以实现客户最佳性能和回报的方向发展。

除了加速器本身，贵公司在 XPU 附加组件（XPU add-ons）业务上有何具体布局 and 进展？

我们明确定义并大力发展一类名为“XPU 附加组件”的业务，它涵盖了 XPU 周边的所有关键定制硅组件。这不仅包括内存扩展产品，还包括定制化的网卡、网络接口产品以及安全产品等。该业务领域曾被认为规模尚小，但目前我们的设计中标势头强劲，已在 TPU、Trainium 和 MTIA 等多个生态系统中获得应用。值得注意的是，我们的 XPU 附加组件不仅适用于第三方定制 XPU，也可以与商用 GPU 搭配使用，部分产品已同时部署在两类服务器上。这印证了我们早在 2020 年提出的观点：每个云平台都是独特的市场和架构，最终所有环节都需要一定程度的定制化。在去年（2025 年）的 AI 日上曾提到，我们在四大超大规模厂商及其他客户中，拥有 15 到 18 种不同的 XPU 附加产品设计，每台加速器的内容价值可达 500 至 2000 美元。这些芯片通常由我们依据规格进行设计和制造，绝大部分基于 Marvell 的 IP，客户提供的 RTL 或设计内容较少。这项业务结合我们在 XPU 产品以及与超大规模厂商的广泛合作，能够创造巨大价值。

市场曾认为芯片领域的竞争是“赢者通吃”，但现在似乎进入了“充裕时代”，您如何看待这一转变以及贵公司的协同部署战略？

市场确实曾存在一种“非此即彼”的观点，认为拥有某个关键产品就能胜出，否则就处于劣势。当我们最初提出 XPU 协同部署战略时，外界甚至有些不解。但我们始终保持战略一致性：聚焦自身业务、清晰勾勒市场机遇并坚定执行。事实证明，我们对这类机遇的判断非常精准。我们明确指出，XPU 市场已形成规模，我们在该领域进展顺利，并且市场体量可观。同时，我们已成功赢得数十种适配其他处理器平台的机会，这些项目将贡献可观收入，并且具有高度战略意义，因为

这些芯片解决方案深度嵌入客户的专属系统架构，支撑其核心业务并构筑差异化优势。结合我们在连接和交换领域的实力，我们能提供完善的端到端架构方案，并在不同解决方案间共享大量知识产权。在当前这个不容有失的时代，客户依赖大型、可靠的供应商按时执行。这绝非零和博弈，市场增长空间巨大，主要参与者都能取得良好发展。

从您的视角来看，当前的人工智能革命处于哪个发展阶段？是早期、中期还是后期？

当前仍处于早期阶段。虽然自 2023 年初 ChatGPT 发布以来，行业势头已经非常强劲，但整体技术发展仍处于初期。例如，我们在 2023 年 5 月预测当年 AI 业务收入为 2 亿美元，次年（2027 年）为 4 亿美元，而我们的数据中心业务季度营收已超过 20 亿美元，这反映了 AI 业务的规模和增长速度。展望未来，总可寻址市场的机会非常巨大。一个关键点在于，GPU 和 AI 加速器的大规模部署尚未真正落地。当能够将机架内的多个 TPU、多个机架和计算单元大规模串联起来时，将会释放出巨大的计算和内存能力。跨规模扩展（scale-out）以及通过 DCI 技术实现不同数据中心集群的协同连接，将是满足 AI 算力需求的下一个关键路径，这将催生全新的创新浪潮和应用场景，并有助于降低成本、提升性能。例如，CPO 技术即将到来，但近封装光学可能会先以更大规模落地。从公司年度战略评审来看，无论是我们面对的 TAM，还是可以拓展的解决方案，都表明 AI 市场在技术进步与所需芯片的匹配度方面，仍处于非常早期的阶段。