白皮书

下一代 AI 集群扩展中 448G 互连的性能评估与调制策略

molex

AI 正在推动对数据中心基础设施的快速投资，尤其是在超大规模运营商不断扩展 AI 集群规模与性能的背景下。训练工作负载和推理应用的爆炸式增长，对高带宽互连系统提出了全新的要求，以便支撑成千上万个相互连接的计算节点。要满足这些要求，关键在于将串行链路技术提升至每通道 448G，从而实现未来每端口 3.2Tbps 的以太网接口。

将 AI 集群扩展至此规模对物理层提出了严苛要求，其中互连架构将定义处理器、加速器和网络接口之间可实现的带宽密度。目前光学与铜解决方案均在考虑之列，但如果能克服封装、PCB 布局及连接器转换等方面的设计挑战，铜互连系统仍具吸引力。由于支持候选的 448G 调制格式所需的带宽极高，传统的 PCB 互连系统、Flyover 互连系统及传统连接器外形能否适用于此数据传输率，目前仍是悬而未决的问题。

本白皮书提出一项研究，通过三种拟议调制方案考察 448G 铜互连系统性能：PAM-4、PAM-6 和 PAM-8。每种调制方案对损耗容限、线性度及均衡要求的影响各不相同，进而影响连接器设计与系统架构。

该研究考察元件封装与 QSFP 模块间的各种互连系统选项，并概述要在 AI 数据中心实现可靠的 448G 部署所必须满足的信号完整性要求。

芯片到模块架构中 448G 连接下的信号完整性挑战

扩展 AI 集群对物理层提出了严格要求，即每个互连系统必须实现更高带宽，同时将信号衰减降至最低。Molex 先前的一项模拟研究对比了直接 PCB 布线与 Flyover Twinax 信道，证实了采用走线和通孔方法的传统 PCB 布线局限性。尽管直接 PCB 布线更为简单，但在长距离传输中插入损耗显著增加，且 PCB 通孔处的回波损耗也更为严重，导致在 80GHz 至 90GHz 频段间出现明显的插入损耗滚降。Flyover Twinax 连接器设计虽表现出类似的插入损耗滚降，但总插入损耗大幅降低；因此，Flyover 电缆能提供更长的信道传输距离。

共封装铜缆 (CPC) 作为一种解决方案，因能规避信号在 PCB 和连接器通孔中传输时产生的高损耗，而备受业界关注。这些模块直接连接到封装基板内部的走线，并直接与 Twinax 电缆连接，绕过了限制带宽的 PCB 通孔。这种架构称为芯片到模块 (C2M)，如图 1 所示。

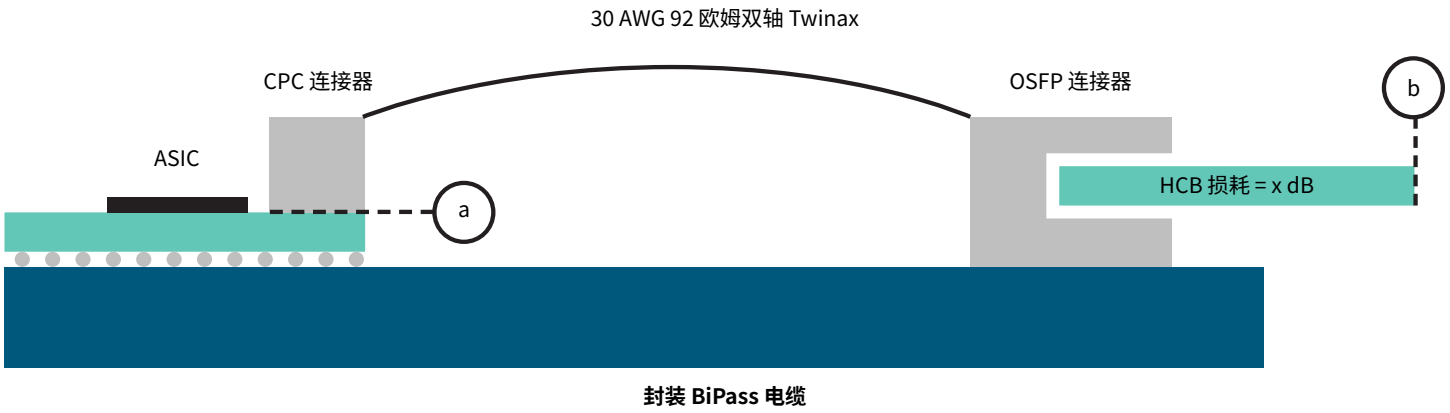


图 1：C2M 架构采用共封装铜缆连接器，并通过 Flyover 电缆连接到 MSA 标准化的连接器与模块

共封装铜缆所用的连接器也带来了信号完整性方面的挑战，这涉及 SMT 设备 (SMD) 焊盘图案、接触短截线长度，以及从封装到连接器的通孔布线。但是，一旦信号穿过连接器并经电缆传输，其插入损耗将大幅降低，并且同轴连接器设计（例如 Twinax）的回波损耗通常远低于 PCB 互连系统。这一特性同样适用于连接至 QSFP/OSFP 连接器以及铜背板连接器的 Flyover 电缆。

信号完整性的性能指标取决于调制格式，因为不同的格式对串行比特流中信号电平间的功率裕度有特定要求。表 1 和表 2 对比了每种拟议调制格式所需的信道带宽与信道传输距离、误码率 (BER) 及信噪比 (SNR)。鉴于当前 AI 数据中心部署的技术是采用 56GHz 信道带宽的 224Gbps-PAM-4，因此各项相对值均与需要 112 GHz 信道带宽的 448Gbps-PAM-4 进行比较。

调制	尺寸	每维度比特数	信令速率, GBd	带宽, GHz	距离缩减, dB
PAM-4	1	2	225 (212.5)	112.5 (106.25)	—
PAM-6	2	2.5	180 (170)	90 (85)	-4.44
PAM-8	1	3	150 (142.5)	75 (71.25)	-7.36

带宽需求降低

检测信号难度增加

表 1：每通道 448G 互连系统中候选调制格式的信号特性与信道传输距离

调制	SNR = 19 dB 时的 BER	Δ	BER = 2.4e−5 时的 SNR	Δ
PAM-4	2.4e−5	—	19	—
PAM-6	3e−3	125x	22.6	+3.6
PAM-8	1.5e−2	625x	25.1	+6.1

表 2：表 1 展示的各调制格式的 BER 和 SNR 要求

表 2 给出了在 PAM-4 调制中，给定目标 SNR 值时所需 BER 的比较，反之亦然。数据显示，互连系统接收端观测到的 SNR 受诸多因素影响：

- 插入损耗
- 回波损耗（反射）
- 串扰
- 偏离/抖动

更高带宽信道中每种调制格式的 BER 与 SNR 要求表明：对于 N 元 PAM，由于星座图密度更高，即便在给定的 BER 目标下，较低带宽的信道也要求更低的反射和串扰。可使用改进的检测技术、链路均衡或纠错机制来放宽 SNR 要求。

在此类数据传输率下，连接器中的机械特性可能限制整体信道性能。例如短截线长度、焊点几何形状、封装到连接器的转换以及连接器到 Twinax 电缆转换等因素，会在低频产生反射，并在高频产生共振。这些因素共同作用，导致高频插入损耗出现滚降，从而在实质上定义了信道带宽的上限。连接器模块的高引脚密度也会产生封装到连接器转换处发生串扰的风险。除了插入损耗，串扰也将决定互连系统接收端的 SNR 与 BER 数值。

使用共封装铜缆打造 448G 互连系统的研究

在之前的模拟研究中观测到的信号完整性特性，驱动了对不同调制频率下共封装铜缆 448G 互连系统的研究。调制格式、各类损耗与 SNR 和 BER 极限之间的明确关系，需要在典型部署环境中通过实际信道进行实验研究。

白皮书本文的后续部分将对采用标准链路架构的 448G 信道中的信号完整性指标进行分析。该研究旨在实现以下目标：

- 评估 PAM-6 与 PAM-8 信号的 SNR、BER 与插入损耗
- 根据插入损耗的滚降确定带宽限制
- 检查 C2M 架构中互连系统在已识别带宽范围内的串扰
- 比较 C2M 架构中主机到模块和模块到主机方向上的信号传播

在 C2M 架构中使用了真正的共封装铜缆和 OSFP 连接器。测量并比较了 PAM-6 和 PAM-8 信号的插入损耗、BER 和 SNR 裕度。

该研究所考察的链路架构及其各部分的预估 BER 极限如图 2 所示。在此 C2M 架构中，研究团队测试了 300mm 与 500mm 的超短距离 (VSR) 信道，涵盖主机到模块与模块到主机双向传输，获得了互连系统的串扰与插入损耗结果。然后，在串扰存在的条件下，改变信道距离与调制格式，同时测量 SNR 与 BER。这些结果为比较支持 PAM-6 与 PAM-8 调制的信道性能提供了充分的数据。

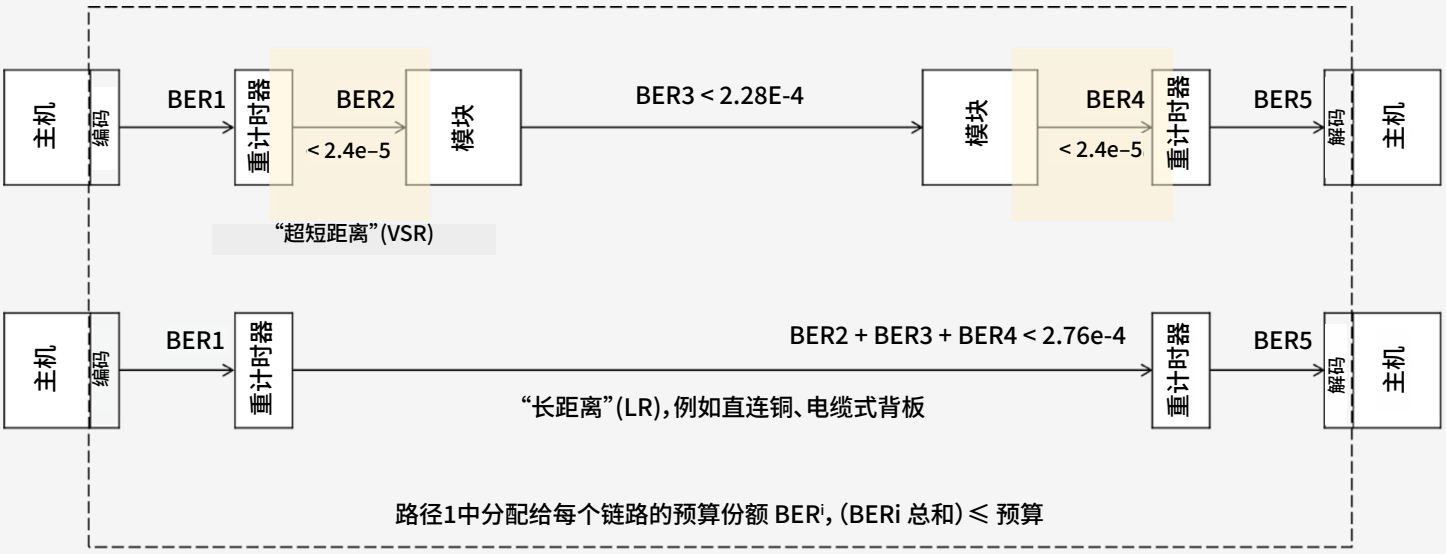


图 2：该研究调查的链路架构

[1] 如果错误发生的方式会损害解码器的性能，链路 BER 可能需要低于目标 BERi。

研究发现与结论

插入损耗和串扰

图 3 展示了主机到模块与模块到主机两个信号传输方向的插入损耗及串扰曲线。两条曲线均显示显著的插入损耗滚降，将信道带宽上限限制在 90GHz 左右。这证实了这些信道能够在 500mm 的传输距离下支持 PAM-6 与 PAM-8 信号传输。

该结果涉及标准 OSFP 收发器模块上的 TX6（主机到模块）与 RX6（模块到主机）信道，该模块是专门为呈现最坏情况下的串扰条件而选定的。串扰在整个信道带宽内保持较低水平，主机到模块方向的 PSNEXT 仅在频率接近 80GHz 时才升至 -60 dB 以上。而主机到模块方向的 PSFEXT 则从大约 75GHz 开始，仅偶尔会超过 -50 dB。

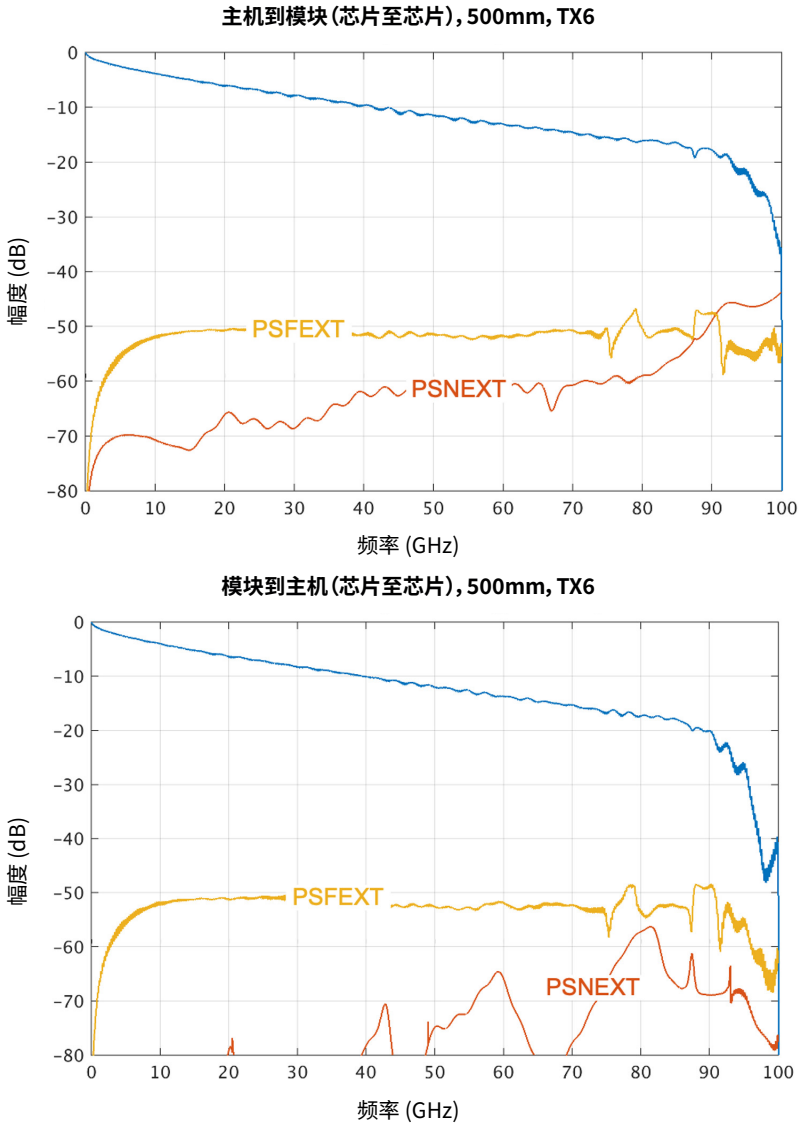


图 3：主机到模块（上图）与模块到主机（下图）方向的芯片至芯片插入损耗及 PSFEXT/PSNEXT 结果；信道长度 = 500mm

BER 和 SNR 裕度

鉴于这些信道中潜在的串扰水平（如图 3 所示），考察预期的 BER 与 SNR 裕度至关重要。表 3 与表 4 分别展示了 PAM-6 与 PAM-8 格式在奈奎斯特频率处的插入损耗、所观察到的 BER 及可用 SNR 裕度。本次评估选择了 TX6/TX8 与 RX6/RX7 信道，因为它们可呈现最坏情况下的串扰条件。

主机到模块		PAM-6				PAM-8			
配对	电缆长度，mm	信令速率，GBd	插入损耗，dB	BER	SNR 裕度，dB	信令速率，GBd	插入损耗，dB	BER	SNR 裕度，dB
TX6*	300	170	15.1	4.8e−7	1.7	145	13.7	1.8e−5	0.2
TX8**	300	170	15.1	4.4e−7	1.7	145	13.7	1.7e−5	0.2
TX6	500	170	16.5	5.4e−7	1.6	145	14.9	2.4e−5	0
TX8	500	170	16.5	4.9e−7	1.7	145	14.9	2.1e−5	0.1

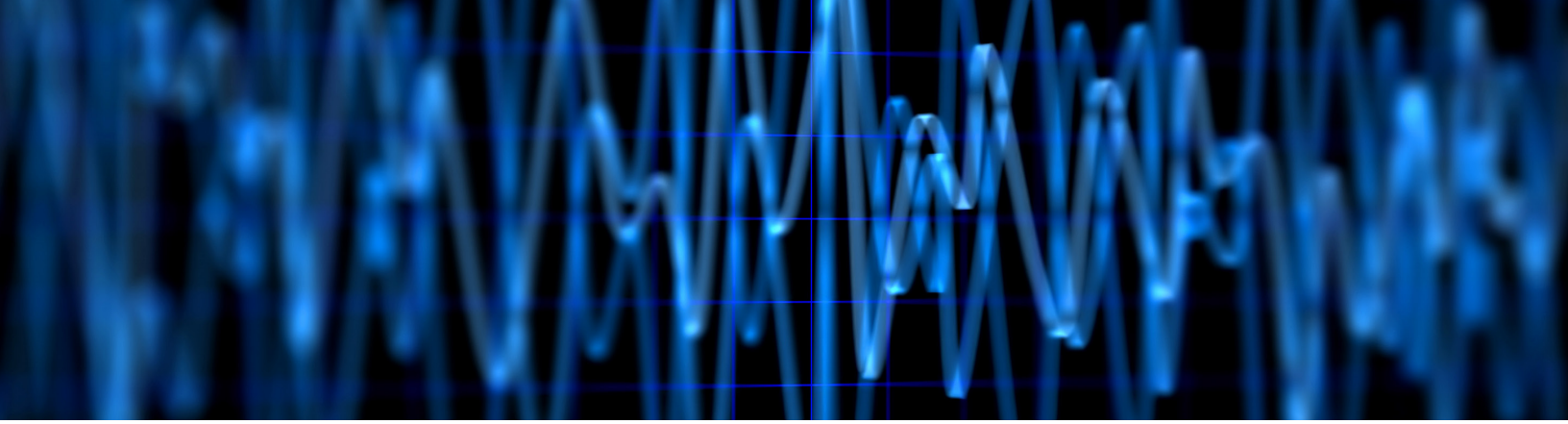
*FEXT 处于最坏情况下的信道
**NEXT 处于最坏情况下的信道

表 3：主机到模块的插入损耗、BER 和 SNR 结果

模块到主机		PAM-6				PAM-8			
配对	电缆长度，mm	信令速率，GBd	插入损耗，dB	BER	SNR 裕度，dB	信令速率，GBd	插入损耗，dB	BER	SNR 裕度，dB
RX6*	300	170	16.5	3.9e−7	1.8	145	14.8	1.8e−5	0.1
RX7**	300	170	16.9	4.2e−7	1.7	145	14.9	1.7e−5	0.2
RX6	500	170	17.9	5.4e−7	1.6	145	15.9	2.3e−5	0
RX7	500	170	18.3	5.3e−7	1.6	145	16	2.1e−5	0.1

*FEXT 处于最坏情况下的信道
**NEXT 处于最坏情况下的信道

表 4：模块到主机的插入损耗、BER 和 SNR 结果



结果表明，在采用 OSFP 连接器的 300mm 和 500mm 信道长度中，PAM-6 与 PAM-8 在技术上均具备可行性。由于 PAM-8 表现出预期中较高的 BER，因此在重定时器/模块接口处相对于 2.4e−5 的 BER 上限几乎没有任何 SNR 裕度（参见图 2）。这使得链路极易受到其他噪声源的影响，例如串扰、外部电磁干扰 (EMI) 或电源轨波动。

而由于插入损耗滚降发生在所需的 112GHz 信道带宽以下，使用传统的 OSFP 桨卡接口无法实现 448Gbps-PAM-4 调制。若要在该数据传输率下使用 PAM-4，则需对多源协议 (MSA) 标准进行修订，以提高滚降阈值。

基于这些结果，PAM-6 成为最具可行性的方案，它能在 BER、SNR 和带宽裕度之间提供更好的平衡。所需信道带宽与 C2M 架构中观察到的插入损耗滚降相吻合，这促使对 OSFP 连接器及桨卡模块的 MSA 标准进行若干修改。

对 MSA 标准的拟议修改

基于上述结果，对桨卡接口进行若干更新有助于进一步支持单信道 448G 数据传输率的 PAM-6 技术方案。这些修改涉及光纤模块 PCB 上的连接器与焊盘布局调整，以便将 C2M 架构中的信道带宽扩展至 90GHz 以上。拟议的修改于表 5 中总结。

拟议	详情
在模块卡侧边添加倒角，以便缩短信号光束尖端	将倒角从 0.25mm 增加至 0.30mm
减小模块卡信号焊盘长度	将从卡侧边到信号焊盘边缘的标称距离从 1.70mm 增加至 1.90mm
减小擦拭长度公差，以便缩短模块卡信号焊盘	从 +/-0.395mm 紧缩至 +/-0.200mm
模块卡上接地焊盘所需的额外暴露长度	将接地焊盘最小长度从 1.40mm 增加至 2.50mm
修改主 PCB 上的 OSFP 连接器焊盘布局	使用 Ø0.36mm，采用焊盘内通孔 (via-in-pad) 技术实现差动信号
消除光纤模块 PCB 顶部和底部焊盘阵列的偏移	将模块卡顶部和底部的焊盘布局对齐

表 5：针对 MSA 标准下桨卡接口的拟议变更，以便支持 448Gbps-PAM-4 及更高阶 N 元 PAM 技术

为实现 448G 铺平道路

本研究证明，采用 OSFP 连接器的传统桨卡接口与 C2M 架构中的共封装铜缆连接器，能够支持使用 PAM-6 或 PAM-8 调制的 448G 信号传输。BER 和 SNR 结果表明，在桨卡和连接器设计符合 MSA 更新的前提下，PAM-6 是首选的调制格式。若无这些更新，PAM-8 或仍可行；但是，必须采用内部编码或高级均衡等增强技术来提升 SNR 裕度。

Molex 依托自身在 112G 及 224G 传输率领域久经验证的领先地位，通过广泛研究与深厚工程技术积淀，为 448G 互连技术铺平道路。通过推动连接器架构与信号完整性技术的发展，Molex 助力数据中心实现更快的数据传输、更出色的信号清晰度，满足 AI 驱动的新型数据环境的性能需求。

如需了解有关实现这一转型的基础技术的更多信息，包括下一代数据中心的设计策略，请访问我们的 224Gbps-PAM-4 高速数据中心技术页面。