

面向嵌入式计算机的高速全光互连架构研究

文敏华¹ 田 径¹ 石添介¹

WEN Minhua TIAN Jing SHI Tianjie

摘 要

在嵌入式计算机领域中,随着数据传输速率的快速提升,已有电信号传输架构无法满足实际使用需求。针对嵌入式计算机高处理性能、高传输带宽的应用需求,文章设计了一种高速全光互连架构,构建基于光背板的光交换设备,通过设计扁平化的处理器互连架构,将多个高性能处理器经由高速网卡与光交换设备相连,可支持 100 Gbit/s 传输,实现高可靠、低时延、大带宽的全光互连网络,并搭建原型系统,通过数据收发实验验证了高速全光互连架构的有效性。

关键词

嵌入式计算机; 数据传输; 光交换设备; 全光互连网络; 高可靠; 低时延; 大带宽

doi: 10.3969/j.issn.1672-9528.2025.04.008

0 引言

在嵌入式计算机领域,随着待处理数据的激增,对于处理器性能、存储设备容量以及交换性能提出了更高要求。由于单个模块的性能有限,当前可基于背板连接来构建一个包含多模块的高性能数据交换和信息处理系统,各个模块通过背板实现高密度的互联。为在有限的空间提高交换能力,每对差分线上的数据传输速率逐渐提升,IEEE 组织在 2014 年发布了 802.3bj 标准,其中定义了 100GBase-KR4 的背板以太网接口规范。这一接口允许在单对差分线上传输 25.78 Gbit/s 的数据速率,4 对差分线可提供 100 Gbit/s 的以太网传输能力^[1]。由于单对差分线上的信号速率已经达到了 25 Gbit/s 以上,对于电信号走线的信号质量有更高的要求,严格控制传输通道的损耗和阻抗连续性仍然是很大的挑战,且对于走线距离也存在更为严格的限制^[2],因此通过电信号进行长距离高速信号传输不再是一种理想的架构,需要引入光互连架构。

高速光互连技术具备巨大的潜力,在云计算数据中心中已被广泛运用。随着绿色计算^[3]、Green Cloud^[4]等概念的提出,光互连网络成为网络节能的重要方式,面向数据中心的高带宽、高交换容量、低开销、低能耗要求,光互连技术有效降低数据中心能耗,提升带宽。针对云计算数据中心的典型运用模式,研究人员提出了多种高速光互连网络架构,并在实验中经过验证,具备高吞吐、低时延、高灵活性、低能耗的优良特性^[5]。而在嵌入式计算机领域,也存在通过集中式结构实现的光互连通信,实现对外通信的灵活配置,但通信速率较低,无法满足不断更新的国际标准及实际使用需求。

因此,针对上述问题,将高速光互连架构引入嵌入式计

算领域,构建基于光背板的光交换设备,通过设计扁平化的处理器互连架构,将多个高性能处理器经由高速网卡与光交换设备相连,高速网卡具备高带宽数据通信功能,满足低时延数据处理要求,实现高可靠、低时延、高带宽高速全光互连架构。

1 高速全光互连架构设计

1.1 总体架构

针对嵌入式计算机高处理性能、高传输带宽的应用需求,设计了高速全光互连架构,通过光互连技术将多个高性能处理器通过光交换设备相连,实现高性能处理器对外的高速数据通信,如图 1 所示,高速全光互连架构由两部分构成。

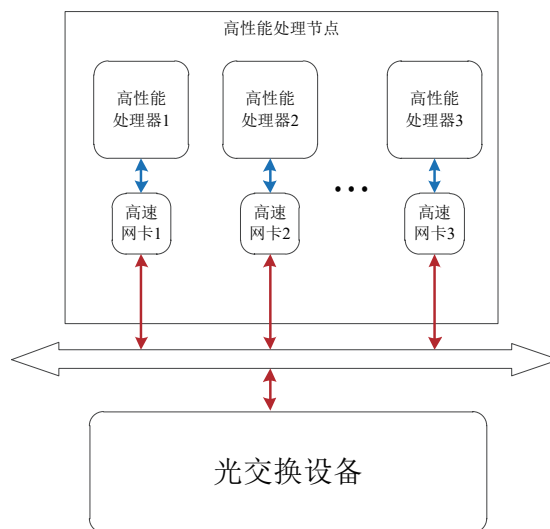


图 1 高速全光互连总体架构

(1) 基于高速网卡的高性能处理架构:通过高速网卡实现高带宽数据的光电转化以及传输,突破高性能处理器数

据处理过程中的带宽瓶颈。

(2) 基于光背板的光交换设备：通过光背板，实现高性能节点与光交换单元的光路映射，完成高性能处理节点的内部信息交换以及对外传输。

1.2 高性能处理架构设计

为应对嵌入式计算机领域日益增长的数据处理需求，各芯片公司陆续推出了针对嵌入式系统的高性能处理芯片，然而，仅更换更高处理性能的计算芯片将会使数据传输成为新的瓶颈。针对此问题，采用了以高速网卡为中心的高性能处理架构，高速网卡承担高速数据传输以及高速信号光电转换等功能。

高速网卡在数据中心已得到广泛运用，当前数据中心网络向高带宽和新型传输体系发展，其网络传输速率迈向 25 Gbit/s，且快速向 100 Gbit/s 发展。而数据中心算力提升遭遇瓶颈，难以匹配快速增长的网络传输速率，此外 CPU 相对更适合处理串行的复杂指令操作，对大量并行的固定模式计算并不适用，如网络传输的协议栈（TCP/IP）等，因此激发了对于高速网卡的需求。基于高速网卡的高性能处理架构工作原理如图 2 所示，高速网卡本质作用在于承载网络侧专用性的网络堆栈算法和传输协议运算迁移，核心效用在于释放 CPU 算力资源，助力其他计算模块高效处理业务数据。在传统处理架构中，由于 CPU 承担存储、数据、网络加密等复杂任务，占用大量业务计算资源，极大影响了可为系统应用提供的处理资源^[6]。

NVIDIA 公司是全球可编程图像处理技术的领军企业，通过收购 Mellanox Technologies，其高速网卡产品在时延与功能均位于市场第一。Mellanox 在 InfiniBand 领域技术专利排名第一，相比于以太网技术，InfiniBand 具备高带宽、低时延的优势。其高速网卡的产品主要有 ConnectX 系列与 BlueField 系列，以 ConnectX-6 Dx 为例，该产品主要特点如下：

(1) 采用虚拟协议互联，支持双端口 200 Gbit/s InfiniBand 和以太网连接，并提供小于 600 ns 的延迟和每秒 2 亿条消息。

(2) 通过提供块级加密，为网络安全提供了重大创新。传输中的数据在存储或检索时进行加密和解密。加密/解密通过 ConnectX-6 硬件进行卸载，从而节省延迟并为 CPU 卸载。由于可以使用不同的加密密钥，ConnectX-6 块级加密卸载可保护用户间共享相同的资源。

(3) 通过 NVMe-of 目标端和发起端卸载，为 NVMe-of 带来进一步优化，提高了 CPU 在应用上的利用率和可扩展性。

(4) 服务器的每个 CPU 通过其专有的 PCIe 接口直接连接到网络，使数据能够绕过总线和其他 CPU，由此改善延迟、

性能和 CPU 利用率。ConnectX-6 Dx 高速网卡基于以上功能，实现了机器学习、Web2.0、云数据分析和存储平台的性能最大化。

在嵌入式计算机领域内，以高速网卡为中心的高性能处理架构能够满足嵌入式计算机对数据处理速度和效率日益提高的要求。这种架构通过利用高速网络技术和智能网卡，优化了数据流的传输和处理，从而提高了整体系统的性能，该种架构具备如下优势：

(1) 该架构采用高速网络技术，如 10、25、50 Gbit/s 甚至 100 Gbit/s 的以太网，以及 InfiniBand 等，确保数据在不同节点间的传输速度和效率。

(2) 高速网卡通常具备可编程性，允许用户根据具体应用需求定制数据处理逻辑。可以根据不同的应用需求进行定制，从而提供了更高的灵活性和可扩展性。

(3) 通过在高速网卡上直接处理数据包，可以减少数据在系统中的传输距离和处理时间，实现更低的延迟，显著提高数据处理能力和响应速度。

(4) 高速网卡的使用可以释放 CPU 资源，使得更多的计算资源用于执行应用程序逻辑，提高计算机的总体处理能力。以高速网卡为中心的高性能处理架构是嵌入式计算机领域内的关键技术，通过提供高速、高效和可定制的数据处理能力，满足了对性能要求极高的应用场景。随着技术的不断进步，这种架构将继续发展，以满足日益增长的数据处理需求。

综上，采用数据中心领域中成熟高速网卡方案，同时添加必要的裁剪适配，可以作为嵌入式计算机中高性能处理架构中的网络传输节点，实现高速网络传输。

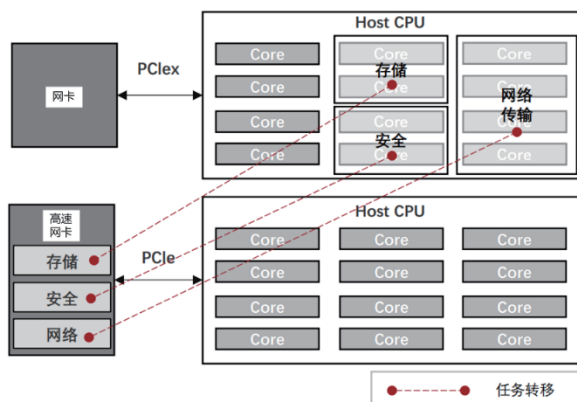


图 2 基于高速网卡的高性能处理架构

1.3 光交换设备设计

光交换设备由光交换单元及光背板组成，光交换单元可提供多路光交换接口，可支持多个处理节点数据交换，便于扩展。光背板，即光纤光互连背板^[7]，将多组带状光纤在背板内部进行分解和组合，构建符合系统内部光路连接的映射

系统^[8]。光背板是将光互连技术与背板制造技术相结合,将光传输介质代替传统的铜导线作为高速信号传输通道,使其具有高带宽、高速率、低功耗、低损耗以及抗电磁干扰等优势,因而也受到了广泛的研究。

光背板具体拓扑关系如图3所示,各处理节点的高速网卡分别引出多路高速光纤与光背板相接,支持扩展100 Gbit/s传输,光背板根据使用需求实现光路映射,系统内光路使用需求为:

- (1) 高性能处理节点间的高速数据交互。
- (2) 高性能处理节点与外界的高速数据交互。针对该使用需求,光背板内部共设计两类映射。
- (3) 高性能节点输出的光路全部接入光交换机提供的光交换接口,借由光交换机的交换功能实现高性能节点间的数据交互。
- (4) 光交换机部分端口通过光背板接到对外输出端口上,以支持处理节点与外部设备的数据交互。为提高光交换设备的稳定性,将光背板内部已经完成布线关系的光纤使用柔性物质固定^[9],同时使用外壳对其进行包装,对光路映射结构进行保护。

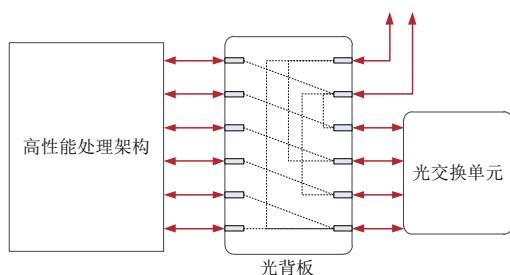


图3 基于光背板的光交换设备

基于光背板的光交换设备实现了处理单元与交换单元的全光互连,因使用光信号的损耗远小于电信号,极大提升了传输距离,进而提升系统的可扩展性^[10]。此外,为进一步提升信号质量,降低链路中的信号损耗,光纤连接处采用斜面物理接触(angled physical contact, APC)光纤,APC光纤端面通常研磨为8°斜面,通过8°斜面连接可以使得光纤端面连接更紧密,降低传输链路的回损,因而提供了更好的连接性能。

光在传输链路中均存在损耗,包括插入损耗和回波损耗,其中回波损耗是因为传输链路中的不连续性,部分信号在传输时反射回信号源所产生的功率损耗,影响回波损耗的因素主要有光纤的精密对准度以及光纤端面的脏污度。为使光纤的端面更好接触,减少光的反射,光纤连接处端面通常被研磨成8°斜面角。如图4所示,光纤连接处端面为平面时,存在部分反射光按原路发射回光源处,从而增大回波损耗,而当光纤连接处端面为8°斜面时,可使部分反射光以一定角度

反射到包层,从而大幅减少反射光在纤芯中反射光返回到光源处,进而降低链路中的回波损耗,因此为进一步提升光信号传输质量,光交换设备中的光纤连接处端面均研磨为8°斜面角。

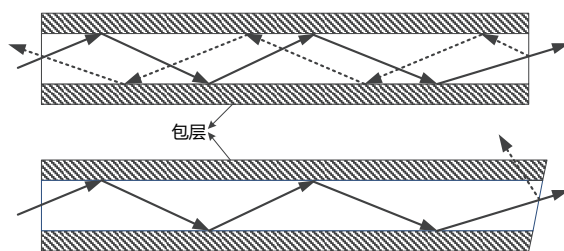


图4 光纤连接处平/斜面对比

2 实验结果与分析

2.1 实验设置

基于高速全光互连架构搭建原型系统,通过测试误码率的方式对该架构的可行性进行验证,具体拓扑互连关系如图5所示,其中误码仪采用联讯仪器mBT5210,以误码仪作为信号发射源,用来模拟高速全光互连架构中的高性能处理器;光模块采用4路收发一体光模块,支持25 Gbit/s传输速率,光模块与光模块测试工装板组成模拟高速网卡,承担高速信号的光电转换以及传输功能;由光模块输出的光链路经由光背板,在外部回环,模拟高性能处理节点间的数据通信。误码仪输出标准码型,通过信号回环后测试对应误码率,逐一对光模块4路光纤进行测试,由于实际使用中可通过前向纠错功能降低误码率,测试过程中当误码率小于5E-5即满足要求。

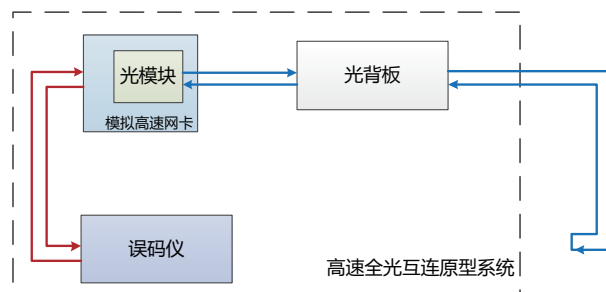


图5 高速全光互连原型系统测试拓扑

为模拟实际使用条件,覆盖嵌入式计算机在多种恶劣环境下使用的场景,在高温、常温、低温3种使用条件下开展测试,以覆盖所有使用温区,此外,分别针对光纤连接面平面/斜面进行误码率测试,以验证设计中采用8°斜面的有效性。

2.2 实验结果

针对原型系统的误码率测试结果如表1所示,由表中数据可得,常温、高温、低温环境下测得的误码率均小于5E-5,满足系统通信要求。此外,表中4个光通道在低温与

高温情况下误码率存在差距, 极限环境条件下光模块因不同通道信号传输质量有偏差, 故测得的误码率存在偏差, 但 4 个光通道中最差情况仍满足小于 $5E-5$ 的要求。

对比光纤连接处为斜面与光纤连接处为平面的误码率测试结果, 低温、高温环境下 4 个光通道, 当光纤连接处为 8° 斜面时, 误码率有明显降低, 从而验证了光纤连接处采用 8° 斜面的有效性, 常温下因自身信号传输质量较好, 光纤连接处为平面或斜面对误码率影响不大。

表 1 原型系统误码率测试结果

	通道	低温	常温	高温
光纤连接处为平面	1	1.799E-7	6.449E-13	2.947E-6
	2	2.106E-8	1.29E-13 (零误码)	1.501E-9
	3	1.137E-9	1.29E-13 (零误码)	1.398E-7
	4	1.466E-6	1.29E-13 (零误码)	1.108E-13
光纤连接处为斜 8°	通道	低温	常温	高温
	1	7.39E-7	1.175E-13 (零误码)	9.046E-11
	2	4.197E-9	1.238E-13 (零误码)	1.236E-13 (零误码)
	3	7.589E-11	1.238E-13 (零误码)	1.236E-13 (零误码)
	4	1.844E-8	1.238E-13 (零误码)	1.618E-11

综上, 本文基于提出的高速全光互连机构构建原型系统, 通过误码仪对整个光链路误码率进行测量, 结合光模块自身工艺影响, 对光模块的多路光通道进行测试, 避免实验偶然性。由实验结果可得, 本文提出的高速全光互连架构, 可以有效满足实际使用中对功能链路的需求, 同时通过在光链路连接节点处的工艺进行改善 (将光纤连接处改为 8° 斜面), 可以在满足使用需求的基础上, 显著改善信号传输质量, 对嵌入式系统高速数据通信具有较大意义。

3 结论与展望

面向嵌入式计算机提出了一种高速全光互连架构, 通过设计扁平化的处理器互连架构, 将多个高性能处理器经由高速网卡与光交换设备相连实现高可靠、低时延、高带宽的全光互连网络, 并通过搭建原型系统, 测试误码率的方式完成了初步的性能验证, 该种高速全光互连架构优点如下:

(1) 高速光信号的损耗及串扰远小于电信号, 同时可提供更长的有效传输距离, 因此光链路可以使用更低的发射光功率, 可有效降低能耗, 同时使用光纤传输高速信号可使系统具备更高的通信带宽及更轻的重量。

(2) 光交换设备中光交换单元对于信号的速率、调制模

式、协议等存在透明传输的特性, 使得网络能够在不更换光交换设备的情况下进行链路带宽升级, 增加系统的可扩展性。

(3) 光交换设备中光交换单元往往具备更大的交换容量以及更高的端口密度, 因此能够实现扁平化的架构设计, 能够很大程度降低网络设备管理开销, 提升系统性能。

面向嵌入式计算机的全光互连架构, 处理器通过电信号与高速网卡进行互联, 在高速网卡上完成电信号到光信号的转换。随着高集成芯片技术的进一步发展, 嵌入式计算机内部的完全光交换已成为可能, 即高性能处理器芯片直接集成光接口, 无需光电转换, 处理器直接通过光信号与其他处理器及外部设备通信, 目前工业界已有集成了光接口的处理芯片原型, 其在嵌入式计算机上应用仍需未来进一步论证。

参考文献:

[1] 羿昌宇, 祖翔宇. 大规模高速背板的信号完整性设计与仿真 [J]. 航空电子技术, 2013,44(4): 1-6.

[2] 杨安毅. 基于高速混合背板的设计及评估方法 [D]. 长沙: 国防科技大学, 2017.

[3] 郭兵, 沈艳, 邵子立. 绿色计算的重定义与若干探讨 [J]. 计算机学报, 2009, 32(12): 2311-2319.

[4] BALIGA J, ARYE R W A, HINTON K, et al. Green cloud computing: balancing energy in processing, storage, and transport[J]. Proceedings of the IEEE, 2011, 99(1): 149-467.

[5] 宋梦洋, 朱虎, 江毅. 数据中心光互联模块发展趋势及新技术研究 [J]. 邮电设计技术, 2024(2):36-40.

[6] 李殿来, 龚欣, 王智君. 预加重在高速电路板中的应用 [J]. 电子科技, 2014, 27(8): 112-114.

[7] 毛久兵, 郭元兴, 刘强等. 光互联背板研究现状 [J]. 激光与红外, 2022,52(9):1280-1287.

[8] 卓良明, 阎德劲. 光电互联基板信号传输检测技术 [J]. 中国科技信息, 2019(11): 103-105.

[9] BENNER A. Optical interconnect opportunities in super-computers and high end computing[C/OL]//2012 Optical Fiber Communication Conference and Exposition and the National Fiber Optic Engineers Conference. Piscataway: IEEE, 2012[2024-09-11].https://doi.org/10.1364/OFC.2012.OTu2B.4.

[10] YOON K B, CHO I K, AHN S H, et al. Optical backplane system using waveguide-embedded PCBs and optical slots[J]. Journal of lightwave technology, 2004, 22(9): 2119-2127.

【作者简介】

文敏华 (1985—), 男, 陕西咸阳人, 硕士, 高级工程师, 研究方向: 机载计算机系统技术。

(收稿日期: 2024-12-02)