

基于 PCI Express 总线的高速光纤接口卡的设计与实现

陈雪^{1,2}, 李佩玥^{1,2}, 章明朝¹, 隋永新¹, 杨怀江¹

(1.中国科学院长春光学精密机械与物理研究所, 吉林 长春 130033;

2.中国科学院研究生院, 北京 100039)

摘要: 针对高速数据采集和远程传输的需求, 提出了一种基于 PCI Express 总线的高速光纤接口卡的设计方案和功能实现。该接口卡基于单片 Virtex-5 FPGA, 采用 PCI Express 总线架构, 以 LocalLink 做为内部链路, 并利用 Aurora 协议实现远距离光纤传输。通过采用 DMA 的数据传输方式, 接口卡的传输速率达到了 700 Mb/s。

关键词: PCI Express; 直接存储器访问; LocalLink; Aurora 协议

中图分类号: TN929.11

文献标识码: B

文章编号: 0258-7998(2011)05-0064-04

Design and implementation of a high-speed fiber card based on PCI Express bus

Chen Xue^{1,2}, Li Peiyue^{1,2}, Zhang Mingchao¹, Sui Yongxin¹, Yang Huaijiang¹

(1.Changchun Institute of Optics, Fine Mechanics and Physics, the Chinese Academy of Science, Changchun 130033, China;

2.Graduate School of the Chinese Academy of Science, Beijing 100039, China)

Abstract: To fit the increasing demand for high-speed data acquisition and remote transfer, the paper introduces a high-speed fiber card based on PCI Express bus. The card is realized by a Virtex-5 FPGA chip. It uses PCI Express bus at endpoint, takes LocalLink as internal link interface, and realizes remote fiber transfer on Aurora protocol. By the DMA control of PCI Express, the card can reach a speed of 700 Mb/s.

Key words: PCI Express; DMA; LocalLink; Aurora protocol

在远程实时控制、高速数据交换、高速光纤组网等领域, 如何实现数据的高速实时采集和远程传输已经成为关键问题。目前的解决方案大都基于 PCI 总线, 并利用电缆进行远程传输。这种方式的共享总线, 可达到的理论极限速率为 133 Mb/s, 远不能满足需求。而且在远距离传输时损耗高, 抗电磁干扰性低, 信息容易被窃取。

与 PCI 总线相比, 第三代高性能 I/O 总线——PCI Express (以下简称 PCI-E) 总线在总线结构上采取了根本性的变革。一是由并行总线变为串行总线, 二是采用点到点的互连。PCI-E 总线×1 通路数据传输率达 0.5 Gb/s, 32 通路时可高达可达到 16 Gb/s, 大大高于目前任何一种总线, 可以满足当前及将来一段时期高速设备的带宽

需求。同时, 由于总线变成链路, PCI-E 引脚数大大减少, 平均带宽大幅提升^[1]。

与电缆等其他通信介质相比, 光纤通信具有传输速度快、损耗低、容量大、占用带宽低、抗电磁干扰的优点, 非常适合于远程高速数据传输。

由此, 本文提出了基于 PCI Express 总线的高速光纤通信方式。在分析 PCI Express 协议、LocalLink 协议和 Aurora 协议的基础上, 使用 Virtex 5 XC5VLX50T FPGA 进行接口卡的硬件设计, 实现了 PCI-E 的 DMA 控制和光纤远程传输, 满足了大容量数据的高速实时传输需求。

1 系统框图

高速光纤接口卡主要分为 PCI-E 金手指、FPGA、《电子技术应用》2011年 第37卷 第5期

光模块、配置 PROM、时钟管理、电源管理 6 个模块,如图 1。

当主机通过板卡向外发送数据时,数据经 PCI-E 插槽传给 FPGA 的 PCI-E 差分接收端,并经 DMA 控制后,被大量取出,放入发送 FIFO 中进行缓存。在发送 FIFO 控制逻辑下,FIFO 数据被转化为 16 bit 的 LocalLink 格式传给 Aurora IP(Intelligence Property)核。Aurora IP 核通过 GTP 差分发送端将数据传送给光模块。数据经过光模块的电光转化后,由光纤传出板卡。主机通过板卡接收外界数据的过程与上述过程基本相反,如图 1。

FPGA 是板卡的核心器件。在此选用 Xilinx 公司 Virtex-5 系列 XC5VLX50T 芯片。FPGA 负责实现板卡的主要逻辑,包括:(1)PCI Express 的物理层、数据链路层、事务层的实现;(2)PCI Express 的 DMA 控制;(3)控制发送、接收缓冲;(4)LocalLink 链路接口;(5)利用 Aurora 协议实现光纤接口。

LocalLink 接口是一种点对点的高效同步连接协议,可以连接用户设计与系统级 Xilinx IP 核。在 PCI-E 高速板卡的设计中,Xilinx 提供的 Aurora 核和 PCI-E 核均采用 LocalLink 接口与用户逻辑进行通信^[2]。

2 系统实现

2.1 PCI-E 接口实现

以往 PCI-E 的实现主要使用专用接口芯片(如 PEX-8311)或使用 FPGA 编程实现 PCI-E 各层。第一种方法利用接口芯片实现物理层、数据链路层和事务层的逻辑控制,但使用芯片多、成本高。第二种方法具有灵活的可编程性,但是开发难度比较大,开发周期长。因此,Xilinx 为 PCI-E 设计提供了第三种快速高效的解决方案,即利用 Virtex-5 芯片内嵌的 Rocket I/O 模块、硬核端点 PCI-E 模块及可选的软核 IP,大大降低了 PCI-E 的开发难度。

板卡设计中,使用 IP 核 Endpoint Block Plus for PCI Express,并配置其为×4 PCI-E 通路,传统 INTx 中断,仅

使用 BAR0 区用作程序中的控制寄存器区。该 IP 核与 PCI Express V1.1 规范完全兼容,达到 2.5 Gb/s 的线速率,支持×1、×4、×8 多种通路,支持 32vector 的 MSI 中断,且具有时钟补偿、自动时钟恢复、8B/10B 编码、流控等多种功能^[3]。

2.2 PCI-E 高速 DMA 控制

PCI Express 端点有 PIO 和 DMA 两种数据传输方式。PIO 方式是由 CPU 通过根复合器发起的读写操作,由于需要 CPU 的发起,所以速度和效率不是很高。DMA 方式直接对内存进行读写,由本地 PCI Express 端点发出,其效率比 PIO 模式高出很多倍。DMA 传输的基本流程如图 2。

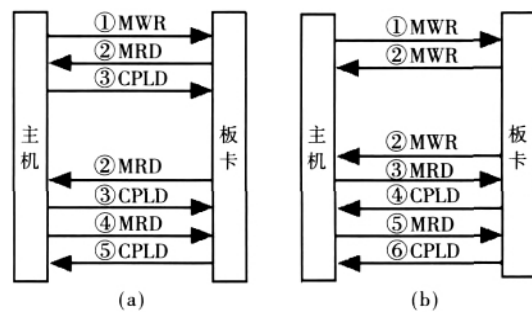


图 2 PCI-E DMA 发送、接收数据过程图

当主机通过板卡向外发送数据时,主机的 PCI Express 端口向板卡写大量数据。如图 2(a)所示,可分为以下几步:

- ①主机向板卡发送 MWR(存储器写)TLP(事务层包),依次写入要发送数据的地址、总长度 len_sum,及控制寄存器的发送开始位,从而启动 DMA 发送。
- ②板卡根据①中的地址和总长度,发送 MRD(存储器读)请求,申请读取长度为 len_temp 的有效数据。注意读请求申请的长度 len_temp 不得大于主机 PCI-E 设备控制寄存器中 Max Read Request Size 所设定的长度,否则将不会返回正确长度的 CPLD(带有数据的完成)。

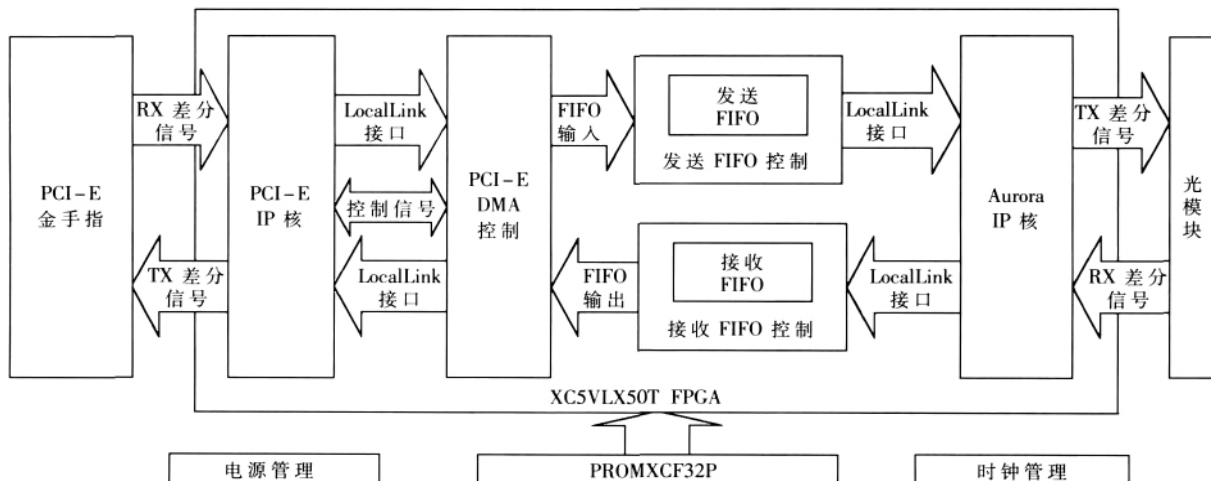


图 1 板卡系统框图

③ 主机返回 len_temp 长度的 CPLD。重复过程②和③，直至板卡总共收到 len_sum 长度的数据，DMA 过程完成，置位中断标志寄存器的 XMIT_COMP 标志位。

④ 主机发送 MRD，读取中断标志寄存器的值。

⑤ 板卡收到读取中断标志寄存器的请求后，返回中断标志寄存器的值，并立即将中断标志寄存器清 0。

当主机通过板卡接收外界数据时，主机的 PCI-E 端口从板卡读大量数据。如图 2(b)所示，可分为以下几步：

① 主机向板卡发送 MWR，依次写入主机中准备存放数据的地址、控制寄存器的接收开始位，从而启动 DMA 传输。

② 当接收 FIFO 中收到不少于某一数量的数据后，板卡通过多个 MRD 将这些数据写入主机，完成一个包的 DMA 传输。注意每次写请求的长度不得大于主机 PCI-E 设备控制寄存器中 Max Payload 所设定的长度，否则将不能完整写入数据。

重复②直至写入 MAX_PACK_COUNT 个包或超时时间到达，置位中断标志寄存器的 RECV_COMP 位，并将收到包的总个数保存到接收包个数寄存器。

③ 主机发送存储器读请求，读取中断标志寄存器。

④ 板卡收到读取中断标志寄存器的请求后，返回中断标志寄存器的值，并立即将中断标志寄存器清 0。

⑤ 主机发送存储器读请求，读取接收包个数寄存器的值。

⑥ 板卡返回 CPLD，携带数据为接收包的个数。

DMA 控制逻辑的组成如图 3 所示。

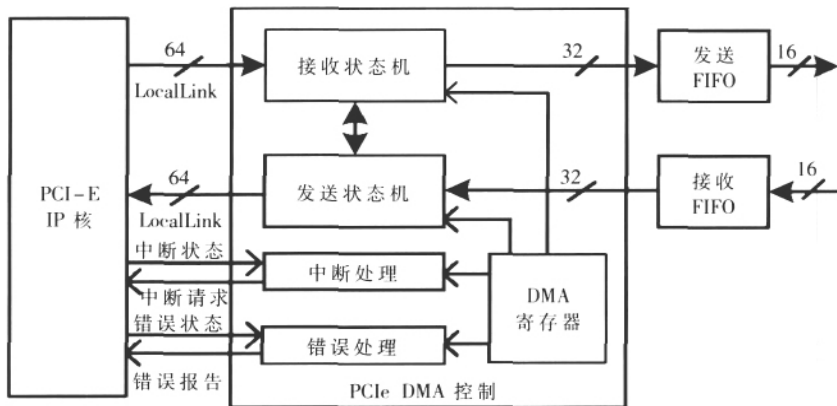


图 3 DMA 逻辑控制框图

接收状态机负责根据板卡从主机接收到的 TLP 包类型进行相应的处理，并将欲发送给光纤的数据存入发送 FIFO。根据图 2 可见，接收状态机需要处理的事务包括 MWR、MRD、CPLD。

发送状态机负责从接收 FIFO 中取数据，并根据 DMA 控制的状态，发送相应的 TLP 包给主机。根据图 2 可见，接收状态机需要处理的事务包括 MWR、MRD、CPLD。

中断处理可以产生 MSI 或 INTx 两种模式的中断^[1]。

具体模式可以通过配置 MSI 控制寄存器的值来选择。板卡中定义的中断有 XMIT_COMP、RECV_COMP。板卡收到主机的读中断请求后，要及时清除中断标志，以防止中断再次误触发。主机根据读取到的中断标志寄存器的值进入相应的中断处理程序进行处理。

错误处理可以将 DMA 控制模块中产生的错误汇报给 PCI-E IP 核，包括非转发 UR (Unexpected Request) 错误、转发 UR 错误、超时错误等。

发送 FIFO 及接收 FIFO 均为异步 FIFO，除进行数据缓存外，还完成了数据宽度 32 bit 和 16 bit 之间的相互转化及时钟域的变化。

2.3 光纤通道实现

光纤通道通过 Aurora 协议实现。Aurora 协议是一种可升级、小型链路层协议，可以用来进行点到点的串行链路数据传输，同时消除了其他串行协议的资源不足、效率低等问题。Aurora 是一个开放的协议，可以在包括 FPGA、ASIC、ASSP 在内的任何硅器件技术中实现。Aurora 支持光纤、TCP/IP、以太网等高层协议^[4]。

在 Xilinx 的 FPGA 器件中，可以通过 IP 核 GTP Aurora 实现。其不仅定义了 Aurora 协议的物理接口，而且定义了包结构、数据提取和流量控制等。IP 核生成文件中包含有效链路的初始化程序、发生过量错误的链路相关程序、嵌入其他协议包的推荐程序等。

在本文的高速光纤网卡设计中，采用 125 MHz 参考时钟，进行 1 通道带流量控制的全双工帧接口通信，可达到 1.25 Gb/s 线速率。

3 软件仿真和性能测试

在板卡的开发设计过程中，对其进行了软件仿真及实际性能测试。

软件仿真中，利用 Modelsim 前仿真功能，对设计的正确性进行验证；利用其后仿真功能，分析布局布线对设计的影响。仿真时，将 Xilinx 提供的下行端口 (Downstream port model) 测试平台与自己的设计相连。下行端口模型的用途是模拟主机 PCI-E 端口，为生成下行 PCI Express TLP 流量提供源机制，以驱动客户设计，同时提供目标机制，以在仿真环境中接收来自客户设计的上行 PCI Express TLP^[5]。后仿真中，发现布局布线会对功能有一定的影响。对关键路径进行修改后得到的最终发送和接收的 DMA 仿真波形如图 4 和图 5 所示。由图可见，设计实现了预期的功能，且通过 DMA，有效提高了 PCI-E 的数据速率。

实际性能测试中，将插有 Agilent 1310 nm 双向单模 SFP 光模块 (极限速率 1.25 Gb/s) 的两块高速光纤网卡分别插入两台服务器 Dell Power Edge R310 的 PCI-E 插槽中，并使用 chipscope 对关键信号进行逻辑分析。通

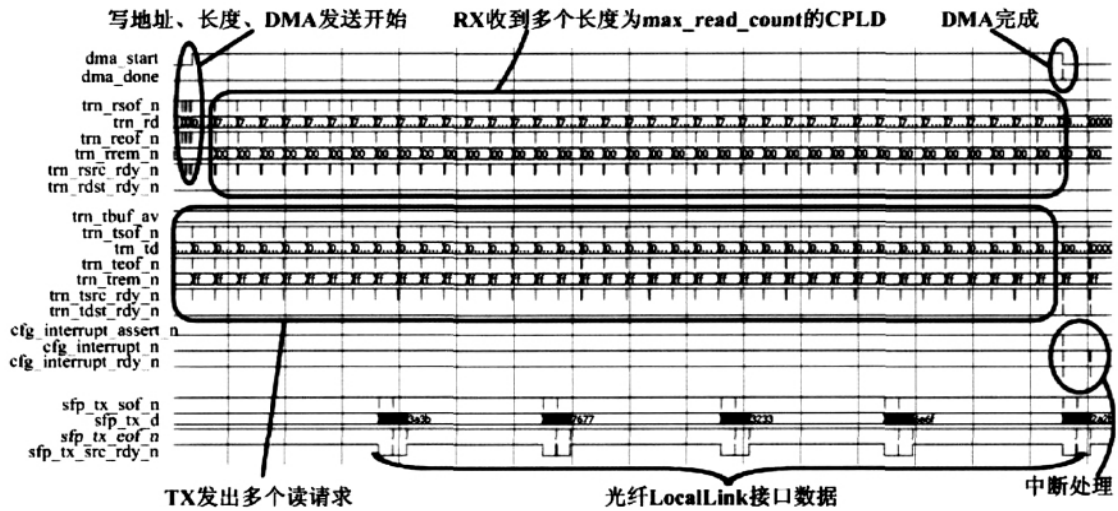


图4 发送DMA仿真波形

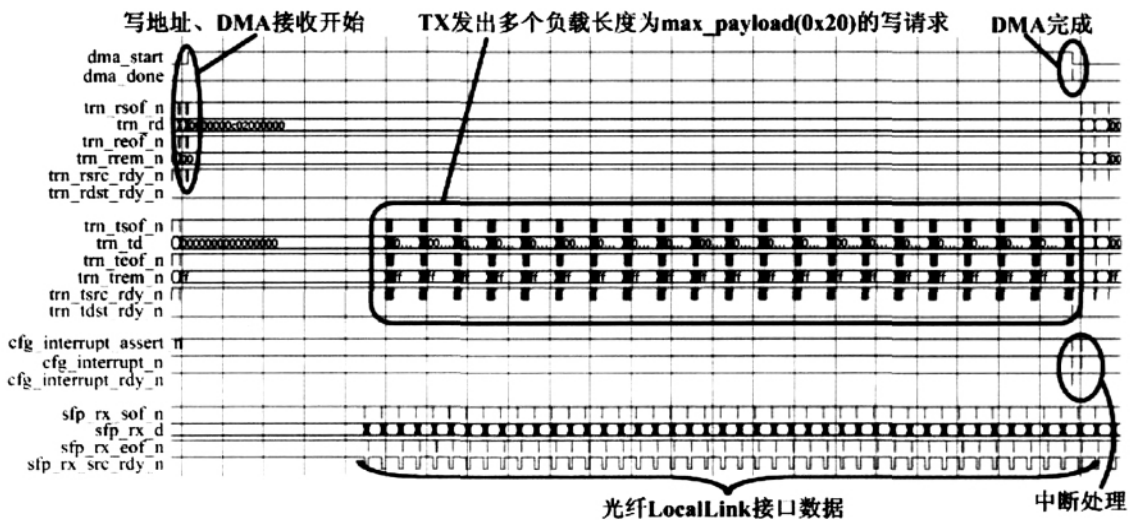


图5 接收DMA仿真波形

过连续发送大批量数据,对板卡进行性能测试。测试结果表明,板卡可以稳定地达到700 Mb/s的通信速率,满足设计要求。

板卡的速度瓶颈主要在于光纤接口。若要提高板卡速率,可以使用更高速的光模块或使用多个光纤链路。

经过测试,该接口卡可以实现数据的高速、实时、稳定传输。该光纤接口卡在高速数据远程采集、远程实时控制、高速数据交换、高速光纤网络通信等领域,将有广阔的应用前景。

参考文献

[1] 马鸣锦,朱剑冰,何红旗,等.PCI、PCI-X和PCI Express的原理及体系结构[M].北京:清华大学出版社,2007.
 [2] Xilinx.LocalLink Interface Specification SP006.July 25, 2005.

[3] Xilinx.Endpoint Block Plus v1.9 for PCI Express DS551. 2008(9), 19.
 [4] 田耘,胡斌,徐文波.Xilinx ISE Design Suite 10.x FPGA开发指南——DSP、嵌入式与高速设计篇[M].北京:人民邮电出版社,2008.
 [5] Xilinx.LogiCORE IP Endpoint Block Plus v1.9 for PCI Express User Guide UG341.2008(9), 19.

(收稿日期:2010-12-25)

作者简介:

陈雪,女,1987年生,硕士研究生,主要研究方向:信息安全、嵌入式设计。