

# PCI Express 专用 DMA 控制器设计与实现

牛少平, 田 泽, 廖寅龙

(中国航空计算技术研究所, 陕西 西安 710119)

**摘 要:** 针对 PCI Express 设备用户端较难充分利用 PCI Express 高数据带宽的问题, 文中使用硬件描述语言设计了一种专用于 PCI Express 设备数据传输的高性能 DMA 控制器。该 DMA 控制器与 PCI Express 事务层直接连接, 可以适用于 Gen1 到 Gen3 不同协议版本; 用户端接口具有较高的通用性和扩展性, 可高效的并发完成多个系统到卡 (S2C) 和卡到系统 (C2S) 的 DMA 操作。该 DMA 控制器的设计经过 RTL 仿真、FPGA 验证, 功能正确, 工作稳定。测试结果表明, 该 DMA 控制器对 PCI Express 带宽的利用率超过 62%, 达到了设计预期目的。

**关键词:** PCI Express; DMA 控制器; S2C; C2S

**中图分类号:** TP39

**文献标识码:** A

**文章编号:** 1673-629X(2012)07-0174-03

## Design and Implementation of PCI Express Specific DMA Controller

NIU Shao-ping, TIAN Ze, LIAO Yin-long

(Aeronautical Computing Technique Research Institute, Xi'an 710119, China)

**Abstract:** Aiming at the problem of PCI Express client hardly utilizing the PCI Express high band, introduce a high-powered DMA controller dedicated to PCI Express based on hardware description language. This DMA controller connects directly with transaction level of PCI Express device, compatible with Gen1 to Gen3 of PCIe protocol. The client interface has high currency and expansibility. It could complete many operations between card and system (S2C and C2S) in high efficiency. With RTL simulation, FPGA verification, the DMA controller functions correctly, works well. The test result indicates the DMA controller could utilize PCI Express band over 62%, reaching design goal.

**Key words:** PCI Express; DMA controller; S2C; C2S

### 0 引 言

PCI Express 是继第一代 ISA、EISA、VESA 总线, 第二代 PCI、AGP、PCI-X 总线之后的第三代高性能 I/O 总线, 是一种能够广泛应用于移动设备、计算系统、嵌入式计算和通信平台等所有周边 I/O 设备互连的总线<sup>[1]</sup>。以 PCI Express Gen1 标准为例, 每个 PCI Express 设备最多支持 32 lane, 每个 lane 支持串行速率高达 2.5Gb/s, 理论数据传输带宽达到了 16GB/s<sup>[2-4]</sup>。如此高效的数据带宽怎样才能体现到集成后的系统性能中, 必然成为 PCI Express 推广应用的重点和难点。

文中在研究 Xilinx FPGA 芯片中 PCI Express 固核<sup>[5-7]</sup>等 PCI Express 设备实现的基础上, 设计实现了

一款适用于多种 PCI Express 设备的高性能 DMA 控制器, 可以适用于各种 FPGA 产品开发与 ASIC 电路设计。设计使用 Verilog HDL 硬件描述语言实现, 在设计完成后与 Xilinx FPGA 中的 PCI Express 固核连接构建了仿真与验证平台, 进行了 RTL 仿真、FPGA 验证。文中着重介绍了 DMA 控制器的电路结构及设计, 仿真验证平台的构建与仿真验证工作的执行, 最后给出了 FPGA 平台性能测试结果。

### 1 DMA 控制器工作原理及体系结构设计

文中介绍的 PCI Express 专用 DMA 控制器与 PCI Express 设备事务层接口直接连接, 可向 PCI Express 设备发送或接收事务层数据包<sup>[8,9]</sup>。用户端接口包括 3 个 BAR 空间读写接口, 两个系统到板卡 (S2C) 方向的 DMA 控制数据接口, 两个板卡到系统 (C2S) 方向的 DMA 控制与数据接口。

DMA 控制器主要完成三个方面的功能:

一是解析 PCI Express 事务层请求包, 并根据需要

收稿日期: 2012-03-02; 修回日期: 2012-06-22

基金项目: 总装“十二五”预研项目 (51308010601); 中国航空工业集团公司创新基金 (2010BD63111)

作者简介: 牛少平 (1979-), 男, 陕西岐山人, 硕士, 工程师, 研究方向为 SoC 设计与验证; 田 泽, 博士, 研究员, 中国航空工业集团首席技术专家, 研究方向为 SoC 设计、嵌入式系统设计与 VLSI 设计。

发送完成包;

二是接收用户端 DMA 请求并将其组装成 DMA 请求包发送到 PCI Express 事务层接口,并根据需要接收 DMA 请求完成包的接收;

三是对前面两个功能共用的发送请求进行优先级仲裁,对接收包进行路由。

DMA 控制器电路结构如图 1 所示。

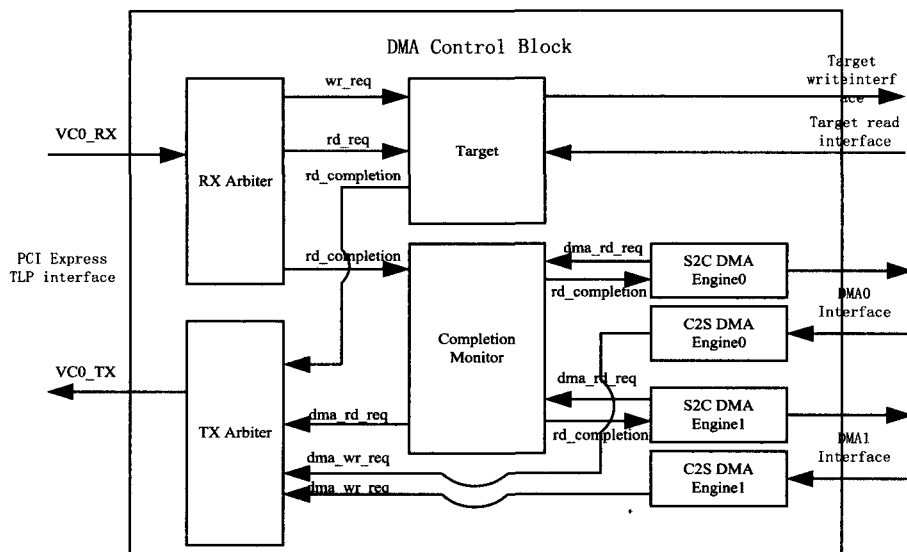


图 1 DMA 控制器体系结构

DMA 控制器主要由八个模块组成:

RX Arbiter 模块用于接收 PCI Express 事务层数据包,包括读请求包、写请求包、读完成包。前两种包对应于主机对 BAR 空间的读写操作,RX Arbiter 模块将其路由到 Target 模块;读完成包对应于 S2C DMA 请求的完成操作,RX Arbiter 模块将其路由到 Completion Monitor 模块。

TX Arbiter 模块接收 Target 模块的读完成包、C2S DMA 请求包、S2C DMA 请求包,对接收到的五类包进行优先级仲裁并发送到 PCI Express 事务层接口<sup>[10]</sup>。本设计中优先级顺序由高到低依次为 C2S DMA0、C2S DMA1、S2C DMA0、S2C DMA1、读完成包。

Target 模块将接收到的读请求或写请求按照不同 BAR 空间的配置属性转换为相应的接口时序,对于写请求包直接转换为写操作时序,对于读请求包先转换为读操作时序,再将读到的数据组成读完成包发送到 TX Arbiter 模块。

Completion Monitor 模块将两个 S2C DMA 请求模块发出的读请求发送到 TX Arbiter 模块,同时将返回的读完成包路由到对应的 S2C DMA 请求模块。

C2S DMA 模块(两个)接收用户端 C2S DMA 请求并将其组装成写请求包,发送到 TX Arbiter 模块。

S2C DMA 模块(两个)接收用户端 S2C DMA 请求并组装成读请求包,发送到 Completion Monitor 模块,

并负责将 Completion Monitor 路由的读完成包按时序送到 S2C DMA 请求接口,完成相应一次 S2C DMA 操作。

## 2 DMA 控制器操作流程

S2C DMA 工作流程<sup>[11]</sup>:在 DMA 请求发出之前,由主机将准备由 S 发到 C 的数据起始地址及大小信息写

入 Bar0 空间规定的寄存器单元,DMA 请求管理逻辑监测到有 S2C DMA 数据准备好后,向 DMA 控制器发出请求。DMA 控制器将该请求组装成 DMA 读请求包,通过 PCI Express 设备发送到主机。根复合体接收到读请求包后,以读完成包的形式将待 DMA 数据发送到端点。DMA 控制器收到 DMA 读完成包后,

通过解析,将读回的数据送到 S2C DMA 请求接口,待传送完成后,DMA 控制器发出完成信号。DMA 请求管理逻辑接收到完成信号后以中断通知主机本次 S2C DMA 操作完成。DMA 控制器一次最多可以接受八个 S2C DMA 请求,缓存在内部的 FIFO 中,每完成一个 DMA 请求可以继续接受一个新的请求。

C2S DMA 工作流程:在请求发出之前,由主机将可以接收数据的存储空间地址写入 Bar0 空间规定的寄存器中,准备好接收板卡 DMA 操作发过来的数据。用户端在有数据需要 C2S DMA 后,DMA 请求管理逻辑向 DMA 核发出请求,DMA 核将该 DMA 请求组装成 DMA 写请求包,经 PCI Express 设备发送到根复合体,根复合体接收到写数据包后负责将数据写到主机准备好的存储空间。待一次 DMA 操作完成后,DMA 控制器发出完成信号,DMA 请求管理逻辑接收到完成信号后以中断通知主机本次 DMA 操作完成。DMA 控制器最多一次可以接受一个 DMA 请求,在完成当前一次 DMA 请求后可以接受下一个的请求。

## 3 仿真与验证平台构建

文中为了验证 DMA 控制器设计的功能和性能,在设计完成之后对其进行 RTL 仿真和 FPGA 验证<sup>[12]</sup>。仿真、验证平台在平台架构和工作原理上一致,区别只是仿真、验证平台中各部分的表现形式不同,所以此处

统一介绍。PCI Express 专用 DMA 控制器的仿真、验证平台原理如图 2 所示。

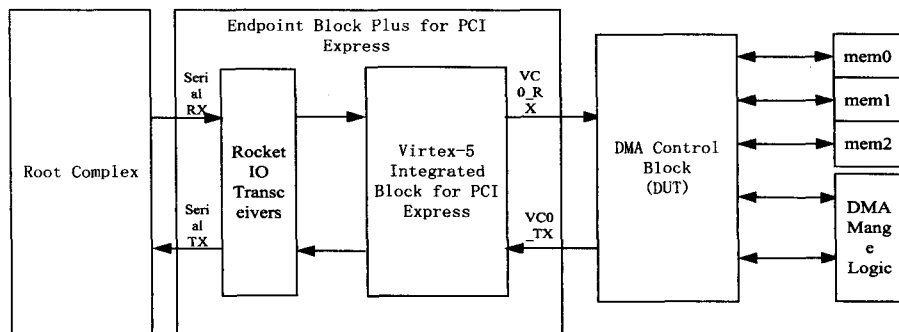


图 2 DMA 控制器仿真、验证平台原理图

PCI Express 根复合体,在系统上电后主动发起 PCI Express 链路训练,对端点设备进行配置,向端点设备发起读写请求、接收端点设备返回的读完成包,或接收端点设备发起的 DMA 请求操作、发送 S2C DMA 请求完成包。

PCI Express 端点设备,在链路建立好并被根复合体正确配置之后,负责接收根复合体发起的读写请求操作、发送读完成包,并负责向根复合体发出 DMA 读写请求包、接收根复合体发送过来的 S2C DMA 请求完成包。

DMA 控制器负责接收 PCI Express 端点设备的读写请求包,经过解析后将读写请求路由到不同的 Bar 空间并转换成正确的读写接口时序,组织读完成包发送到 PCI Express 端点设备发送通道;DMA 控制器同时负责接收 DMA 读写请求,在对五种发送数据包进行优先级仲裁后发送到 PCI Express 端点设备事务层发送通道,同时负责接收 S2C DMA 请求完成包并路由到相应的 DMA 请求接口,转化为正确的读写接口时序。

为了验证 DMA 控制器的功能,在用户端接口还设计了以下电路:BAR0 为 32 位存储器空间、64KB、不支持突发操作;BAR1 为 32 位存储器空间、4KB、支持突发操作;BAR2 为 32 位存储器空间、256MB、不支持突发操作;DMA 请求接口连接了 DMA 请求管理逻辑,S2C DMA 接口分别连接了小包(不超过 4KB)DMA 请求、大包(不超过 16MB)DMA 请求;两个 C2S DMA 接口分别连接了小包 DMA 请求、大包 DMA 请求。

#### 4 仿真验证与性能测试

在 RTL 仿真平台中各部分均采用硬件描述语言描述。根复合体、PCI Express 端点设备使用 Xilinx 提供的仿真模型<sup>[13]</sup>,DMA 控制器使用 RTL 级硬件描述语言编写,mem0、mem1、mem2 使用 Xilinx 提供的存储器模型,DMA 请求管理逻辑使用 RTL 级硬件描述语言编写。

Xilinx 提供的根复合体模型中只提供了简单的读写请求用户接口。为了测试 DMA 功能,在根复合体模型中开发了接收 S2C、C2S

DMA 请求接口、发送 S2C DMA 请求完成包接口,开发了用于 DMA 数据存储的大容量存储器模型。

文中使用上文所述的仿真平台,可以在根复合体端构造不同的测试向量向端点发起读写操作、接收端

点返回的读完成包并检查数据正确性;可以接收用户端发起的 C2S DMA 请求,将数据写入大容量存储器模型并检查正确性;可以接收用户端发起的 S2C DMA 请求并向端点发送 DMA 请求完成包。

在 FPGA 验证平台中,包括主机板、验证板两块电路板,两块电路板通过高速连接器连接。PC 机通过以太网口对主机板进行调试。

主机板主要包括 MPC8641 芯片作为主处理器,该芯片中集成了 PCI Express 根复合体;高性能大容量的 DDR2 存储器,作为处理器程序运行空间以及 DMA 数据存储空间;以太网口作为调试接口。

验证板上选用 Xilinx Virtex5 LX330T FPGA 芯片,PCI Express 端点设备使用的是 FPGA 中的固核,mem0、mem1、mem2 使用的是 FPGA 中的存储块,DMA 控制器、DMA 请求控制逻辑使用 RTL 级描述语言编写,上述模块使用 ISE 工具综合下载到 FPGA 中,共同完成 PCI Express 端点设备和用户端逻辑功能。

文中 PCI Express 设备使用的典型配置为 4lane,2.5Gb/s,最大包负载 128B。经过 FPGA 环境验证得出:根复合体直接读写的数据带宽约为 20MB/s,主要受限于数据包负载大小以及用户端读写操作时序的快慢;在 DMA 请求大小均为 4KB 的情况下,S2C DMA 数据传输带宽约为 502MB/s,C2S DMA 数据传输带宽约为 530MB/s。通过计算可以得出,该带宽利用率超过了 PCI Express 理论带宽的 62%。

#### 5 结束语

文中采用硬件描述语言设计了一种 PCI Express 专用 DMA 控制器,可以充分的利用 PCI Express 设备的高带宽提升整个系统的性能。该设计对不同的 PCI Express 设备及用户端应用需求具有较高的通用性,可以适用于不同的 FPGA 产品开发及 ASIC 设计。经 FPGA 环境测试表明,该 DMA 控制器的设计思想是成功的,PCI Express 根复合体与端点之间的数据传输带宽

(下转第 180 页)

比较图 a 和图 c 点对点的定向天线模型,在波束较小的情况下能耗更小,这是因为 MST 算法和 d-BIP 算法能耗与宽度  $\theta$  线性相关。而 c-d-BIP 算法的能耗几乎和宽度  $\theta$  无关,因为 c-d-BIP 算法宽度可变,最小宽度  $\theta$  几乎很少使用。

MST 算法和 d-BIP 算法每个消息值发送给一个节点,而 c-d-BIP 算法一个消息可以同时发送给多个节点,所以在密度较大的网络中 c-d-BIP 算法的性能优于 MST 和 d-BIP 算法。

### 3 结束语

文中以基于点对点的定向天线模型的 d-BIP 算法为基础,提出了一种改进的 c-d-BIP 算法。实验证明 c-d-BIP 算法使用了波束宽度可以改变的通信模型在节点密度较大的网络结构中,使用 c-d-BIP 算法比定向天线的 MST 算法和 d-BIP 算法性能更优越。接下来的工作可以在 d-RBOP 算法中引入波束宽度可以改变的通信模型以进一步改进 d-RBOP 算法。因为以一个宽度较大的传输光束来传输信号给两个或更多邻节点比以一个较小的光束来一个个的传输信号给这几个节点的效率要高很多。

#### 参考文献:

- [1] 李应娣,单志龙. 无线传感器网络定向扩散路由协议研究[J]. 计算机技术与发展,2010,20(4):40-43.
- [2] 李德英,陈文萍. 无线传感器网络能量高效综述[J]. 计算

机科学,2008,35(11):8-12.

- [3] 崔吉,徐桂云. 无线传感器网络的概述[J]. 山东通信技术,2006,16(4):28-31.
- [4] Wieselthier J E, Nguyen G D, Ephremides A. On the construction of energy-efficient broadcast and multicast trees in wireless networks[DB/OL]. 2000. <http://www.ee.washington.edu/research/nsl/class/565/2002wtr/wieselthier.pdf>.
- [5] Cartigny J, Ingelrest F, Simplot-Ryl D. Localized LMST and RNG based minimum-energy broadcast protocols in ad hoc networks[J]. Ad Hoc Networks, 2005, 3(1):1-16.
- [6] 李政,李德英. 无线自组织网络中能量有效的广播与组播[J]. 软件学报,2010,21(8):2023-2036.
- [7] Ramanathan R. On the performance of ad hoc networks with beam forming antennas[DB/OL]. 2001. <http://www.ir.bbn.com/~ramanath/pdf/mobihoc01-beamform.pdf>.
- [8] 陈帅,钟先信,廖晓伟. 无线传感器网络节点实现模型[J]. 计算机技术与发展,2007,17(3):176-178.
- [9] 唐勇,周明天. 无线传感器网络中最小化能量广播算法[J]. 通信学报,2007,28(4):80-86.
- [10] Guo S, Yang O W W, Leung V C M. Energy-aware multicasting in wireless ad hoc networks: a survey and discussion[J]. Elsevier Computer Communications, 2007, 30(4):2129-2148.
- [11] Argali M, Hubaux J P, Enz C. Minimum-energy broadcast in all-wireless networks: NP-completeness and distribution issues[DB/OL]. 2002. <http://www.sigmobility.org/mobicom/2002/papers/p044-cagalj.pdf>.
- [12] Rodoplu V, Meng T H. Minimum energy mobile wireless networks[J]. Selected Area in Comm, 1999, 17(8):1333-1344.

(上接第 176 页)

得到了很大的提高,对于加速 PCI Express 设备在嵌入式计算和通信平台中的推广应用具有一定的促进作用。

#### 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] PCI-SIG. PCI Express 2.0 Base Specification Revision 0.9[S]. 2006.
- [3] PCI-SIG. PCI Express Card Electromechanical Specification Revision 1.0[S]. 2002.
- [4] Budruk R, Anderson D, Shanley T, et al. PCI Express System Architecture[M]. [s. l.]: Addison-Wesley Developer's Press, 2003.
- [5] Xilinx. Virtex5 Integrated Endpoint Block for PCI Express Designs[EB/OL]. 2009. <http://www.xilinx.com/support/documentation/>.
- [6] Xilinx. Virtex5 FPGA RocketIO GTP Transceiver User Guide

[EB/OL]. 2009. <http://www.xilinx.com/support/documentation/>.

- [7] Xilinx. LogiCORE IP Endpoint Block Plus v1.11 for PCI Express[EB/OL]. 2009. <http://www.xilinx.com/support/documentation/>.
- [8] 廖寅龙,田泽. FC 网络通信中 PCIe 的接口的设计与实现[J]. 航空计算技术,2010(4):127-130.
- [9] 王伟,傅其祥. 基于 PCIe 总线的超高速信号采集卡的设计[J]. 电子设计工程,2010(5):43-45.
- [10] 崔骞,高小鹏,龙翔. 加权轮询策略的 PCI Express DMA 控制器的设计与实现[J]. 微计算机信息,2010(23):156-158.
- [11] 单天昌,陆达. 基于 FPGA 的 PCI 接口 DMA 传输的设计与实现[J]. 计算机技术与发展,2010,20(4):215-218.
- [12] 徐欣,于红旗. 基于 FPGA 的嵌入式系统设计[M]. 北京:机械工业出版社,2005.
- [13] 孙航. Xilinx 可编程逻辑器件应用与系统设计[M]. 北京:电子工业出版社,2008.