

基于FPGA的RAID控制卡实现及仿真研究*

刘景宁, 刘晓芳, 范俊, 童薇

(华中科技大学 计算机学院 信息存储系统教育部重点实验室, 湖北 武汉 430074)

摘要: 首先介绍了基于FPGA的一种RAID控制卡的原理及系统设计、印刷电路板(PCB)的具体实现。由于板卡运行在66MHz总线时钟之上,必须考虑高频下电路的性能及电路的信号完整性等,因而在PCB设计阶段对电路的仿真显得尤为重要。还将介绍基于IBIS模型的信号完整性仿真分析,并利用信号噪声分析软件(Hyperlynx)对高速电路设计中的PCB布局布线、匹配电阻设计和并行线串扰分析进行深入研究。根据仿真分析结果调整原设计,从而提高了信号质量,减少开发成本。

关键词: 现场可编程逻辑阵列; 磁盘阵列; 仿真; 输入/输出缓冲器信息标准; 信号完整性

中图分类号: TP302.2

文献标识码: A

文章编号: 1001-3695(2007)01-0261-03

Realization and PCB Simulation of FPGA-based RAID Control Card

LIU Jing-ning, LIU Xiao-fang, FAN Jun, TONG Wei

(Key Laboratory of Data Storage System, Ministry of Education, College of Computer, Huazhong University of Science & Technology, Wuhan Hubei 430074, China)

Abstract: This paper first introduces the principle, architecture and PCB design of the RAID expansion card based on FPGA. For the main clock is 66MHz, we must care about the function and Signal Integrity (SI) of the high-speed circuit PCB design. So it focuses on the analysis of SI based on IBIS model with software Hyperlynx. Furthermore, it find a method to improve the PCB design's reliability. This paper is also valuable to electronics design.

Key word: FPGA; RAID; Simulation; IBIS Model; SI

本项目是针对RAID5系列的高性能光纤磁盘阵列系统^[1],采用FPGA技术设计一款RAID系统控制卡,实现磁盘阵列启动、数据缓存(Cache)以及数据XOR校验等功能。

廉价冗余磁盘阵列(Redundant Array of Inexpensive Disks, RAID)通过将数据分布在多个通道的多个磁盘上,利用并行处理机制使多个独立磁盘协调工作,达到提高存取速度和增大存储容量的目的。此外,磁盘阵列通过增加数据的冗余度提高数据的安全性。但计算出冗余信息却需要消耗系统大量的处理资源。以RAID5为例,在数据的写操作和数据的恢复过程中需要执行大量异或运算,过多的校验计算任务占用了大量的CPU资源,这将导致磁盘阵列服务器对用户请求的响应时间增加,系统的整体性能被削弱。因而采用扩展的硬件电路实现XOR运算,能够将CPU从繁重的校验任务中解脱出来,从而提高阵列服务器整体性能。传统的软件实现RAID系统需要操作系统的支持,RAID的配置信息存在系统信息中,而不是存在硬盘上。当系统崩溃需重新安装时,RAID的信息也会丢失。因此目前广泛采用专用的硬件控制卡方式实现RAID系统。RAID控制卡实现系统的启动及自身初始化及配置等工作。RAID控制卡是一个PCI从设备,接收并执行来自系统的命令,主要实现磁盘阵列系统的初始化、配置及系统的启动、阵列管理等工作。另外,在磁盘阵列中采用Cache,其作用主要是加

速读操作和缓存小写,解决RAID5中小写效率低下的问题。因此,RAID控制卡使用NVRAM(Non-Volatile RAM)即非易失性随机存储器来增强阵列Cache的性能。

在板卡设计过程中,由于PCB板上电子元器件密度较大,走线较密,信号频率最高达到66MHz,因此不可避免地要出现EMC(电磁兼容)和EMI(电磁干扰)问题。另外,RAID控制卡在进行PCB设计时由于要满足PCI协议、集成电路芯片及其他元件的布线要求,其PCB设计要注意的问题更多。本文从项目系统设计、印刷电路板(PCB)设计等方面对整个电路作详细介绍,并对基于高速数字电路仿真的信号完整性、并行线串扰以及EMC等问题进行深入研究分析,提出基于PCI总线的FPGA的PCB设计中需注意的问题及解决方法,相信对高速电子电路设计有很好的参考作用。

1 系统原理及硬件电路设计

本设计中采用PCI总线来作为控制卡与RAID之间数据传输的桥梁^[2]。控制卡PCI接口工作在32位数据宽度、33MHz~66MHz的总线时钟之上,数据最大传输率高达266Mbps。系统原理如图1所示。

系统工作原理:系统上电后,FPGA由其专用的配置芯片进行主动配置,10秒钟内控制卡完成自身初始化工作。在磁盘阵列BIOS的Init功能执行过程中,控制卡截获中断Int19H,并将Flash中的阵列控制代码作为新的中断服务程序执行。至此,控制卡即实现对磁盘阵列的主动控制,从而接管数据校验、数据缓存等操作。控制卡启动后控制和协调整个系

收稿日期: 2005-10-06; 修返日期: 2005-11-25

基金项目: 国家自然科学基金资助项目(60273074); 全国优秀博士学位论文专项基金资助项目

统,管理数据的流动。当数据用户请求写入数据时,控制卡缓存数据、计算数据的校验和,并命令阵列存储数据及校验和。读出数据时,控制卡命令阵列提取数据和相应的校验信息,并缓存数据、校验数据有效性。最近写入或读出的信息由控制卡缓存存在本地的非易失的 Cache 中,以便进行数据保护和快速提取。

1.1 数据缓存的控制功能模块设计

本设计采用 NVRAM 实现磁盘阵列 Cache。NVRAM 由小型锂电池和低功耗的 SRAM(Static RAM,静态随机存储器)组成。当系统断电后,由此锂电池对 SRAM 供电,所以 NVRAM 既继承 ROM(Read Only Memory,只读存储器)的优点——具有非易失性,又摒弃 ROM 的缺点,支持快速写操作。在本设计中,磁盘阵列控制器通过 PCI 总线读写 NVRAM(选用 ST 公司 M48Z2M1Y),访问磁盘阵列 Cache。由 FPGA 芯片对 NVRAM 进行读写控制。NVRAM 接口的引脚由信号 Addr、DQ、En、Wn 和 Gn 组成,其中信号 Addr 为单向地址信号;DQ 为双向数据信号。En、Wn 和 Gn 信号分别为芯片使能、输入使能及输出使能等控制信号。当 Wn 为 0 写 NVRAM,为 1 读 NVRAM;当 Gn 信号为 0 时,DQ 信号才输出读 NVRAM 的数据。读 NVRAM 时,在 A 信号输入地址,驱动控制信号($En = 0; Wn = 1; Gn = 0$),一个 NVRAM 读周期(大约 80ns)后,DQ 信号输出数据;写 NVRAM 时,在 A 信号输入地址,在 DQ 信号输入写数据,驱动控制信号($En = 0; Wn = 0; Gn = 1$),一个 NVRAM 写周期(约 80ns)后,数据写入 NVRAM。电路原理图如图 2 所示。



1.2 阵列传输数据的 XOR 校验功能模块电路设计

在工作原理上,当接收到校验计算任务后该控制卡能以 PCI Master 方式从内存读取参加校验计算的数据并通过硬件执行校验计算。当校验结果计算完毕后再以 Master 方式将其写回内存。模块的硬件逻辑设计采用了并行的思想,令数据在 PCI 总线上的传输过程和校验计算过程在时间上重叠,使得整个校验过程耗费的时间等同于数据在总线上传输的时间,从而最大限度地提高了校验速度。该功能模块不需要附加外围电路,仅消耗 FPGA 16KB 的片内 RAM 以实现模块功能。

1.3 阵列启动的控制功能模块电路设计

根据 PCI 协议中利用扩展 ROM 实现系统启动的相关介绍,我们可以在该阵列控制卡上实现扩展 ROM,扩展 ROM 的代码主要包含两部分,即一部分是在 Init 功能中截获 BIOS 的 19H 中断;另一部分是自定义新的 Int 19H 中断服务程序实现从控制卡启动操作系统。具体方法是在板卡上用一块 Flash 放置扩展 ROM 代码和磁盘阵列控制代码并编写相应的控制逻辑以允许程序对其进行访问。本设计选用 AMD 公司闪存系列——AM29DL640D,其存储空间分为四个,两个 8Mb 区和

两个 24Mb 区,以此来实现并发的存取操作。另外,AM29DL640D 支持 8 位或 16 位数据模式,由 BYTE#信号控制。当 BYTE#为高时,DQ₀ ~ DQ₁₅有效;为低时,DQ₀ ~ DQ₇有效。另外,CE#,OE#,WE#分别为片选信号,输出使能信号和写使能信号。Flash 的控制逻辑由 FPGA 实现。

1.4 PCI 总线接口电路设计

目前大多数系统厂家的 PCI 总线接口均是采用国外的专用控制芯片,但用户可能只用到 PCI 接口芯片的部分功能,造成一定的资源浪费,并且在设计上也不能根据系统要求灵活配置。基于以上考虑,本设计采用 FPGA(现场可编程门阵列)实现 PCI 总线协议,将 PCI 接口逻辑与用户其他逻辑集成到一个 FPGA 芯片上,从而实现紧凑的系统设计。当系统升级时,只需对 FPGA 器件重新进行逻辑设计,而无须更新 PCB 版图。PCI 规范定义了一系列的电气兼容要求,如信号环境和 I/O 缓冲等,因而,考虑开发一个 PCI 接口之前必须根据这些电气要求选择合适的 FPGA。当然 FPGA 的选择还要考虑其他一些设计要求,如逻辑资源是否丰富,是否需要实现内部 RAM,内部 RAM 的规模有多大等等。除电气规范外,FPGA 必须同时满足 PCI 规范的严格时序要求。本设计选用 Altera 公司 Cyclone 系列——EP1C12Q240C6^[3,7]。我们复用 Altera 的 pci_mt32 宏核来实现对设备的 PCI 接口的设计。通过 IP 复用技术可以减少开发时间,并且有助于提高设计的稳定性。

PCI 总线接口电路单元是整个设计的核心部分,在进行 FPGA 引脚分配时应充分考虑 PCI 协议和 PCB 布局布线等因素的制约。另外,板卡兼容 3.3V 和 5V(即通用卡)的 PCI 信号,然而 Cyclone 系列 FPGA 引脚只支持 3.3V PCI 信号环境,因此在 PCI 金手指与 FPGA 芯片之间必须进行电平转换。在硬件电路上采用 IDT 公司 QuickSwitch 系列芯片——QS3861。当 QS3861 处于使能状态时,其输出电压随输入电压变化,当输入电压上升时,芯片的内阻和输出端电压值也随之增加,但输出电压的最大值由芯片工作电压值决定,即

$$\text{输出电压最大值} = \text{工作电压 } V_{cc} - \text{阈值电压 } V_t$$

阈值电压 V_t 的典型值为 1V。例如,当 V_{cc} 为 5V 时,输出电压最大值为 4V,即当输入电压值再继续增大时,输出端电压被箝位在 4V。利用这个特性,我们可以实现 5V PCI 信号到 3.3V PCI 信号的转换^[4,7]。电路原理图如图 3 所示。

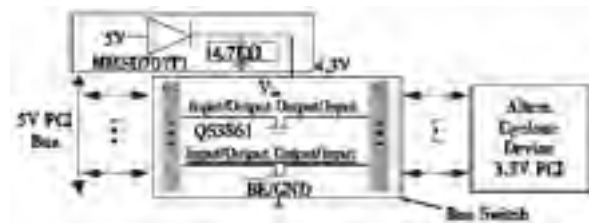


图 3 PCI 接口模块电路原理图

1.5 FPGA 配置电路设计

Cyclone 系列 FPGA 支持三种配置方式,即主动串行配置(AS)、被动串行配置(PS)和基于 JTAG(Joint Test Action Group)配置。控制卡支持 AS 和 JTAG 两种方式。对于 AS 配置方式,需由专用串行配置芯片(本设计选用 EPCS4,容量为 4MB)对 FPGA 进行主动配置,作为目标设备的 FPGA 产生控制信号和同步时序控制信号。当系统上电时,Cyclone 先通过 nCSO 信号拉低 nCS,再通过 ADSO 信号向配置芯片发送配置

命令和地址,作为对配置命令的应答。配置芯片在 DCLK 的下降沿通过 DATA0 信号发送配置数据;在上升沿通过 Cyclone 锁存配置数据。当配置完成时,Cyclone 释放 CONF_DONE 信号,由于外加上拉电阻作用,CONF_DONE 置 1。若 CONF_DONE 信号在配置结束后没有被正常拉高,Cyclone 拉低 nSTATUS 以启动配置芯片重新进行配置。另外,由 ByteBlaster II 下载电缆通过 Quartus 软件将配置文件下载到配置芯片。电路原理图如图 4 所示。对于 JATG 配置方式,主要适用于在线配置,相关参考文章较多,本文不再介绍。

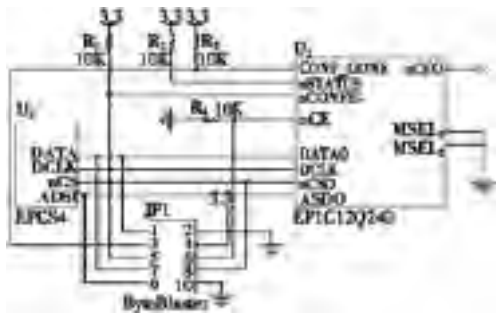


图4 配置电路原理图

1.6 其他电路模块

控制卡除以上主要电路模块以外,还包括了电源模块、USB 接口模块和 LCD 接口模块等,由于均为通用电路模块,本文略。

2 PCB 关键技术

本设计采用四层板结构:信号层-地层-电源层-信号层。普通信号线宽 10mil,地线 $\geq 30\text{mil}$,电源线 $\geq 20\text{mil}$ 。控制卡主要使用的电压为 3.3V 和 1.5V,因此对电源层进行了分割。其他电源(5V)则采取在信号层上走线(遵循电源布线原则)的方式。

由于元件密度较大,信号频率高和 PCI 协议规范复杂等问题,在控制卡关键电路的 PCB 设计中,应注意如下几点:

(1) 关键信号一般要相邻地平面^[5], 如 PCI 中的高频信号、高速信号、时钟信号等要相邻地平面布线, 以减少电磁辐射防止 EMI。对于 PCI 控制信号中的 CLK, CBEN[0-3], FRAME#, TRDY#, IRDY# 等信号, 布线时应该尽量走第一信号层。

(2) PCI 信号线的走线长度的要求。CLK 信号线的长度为 $2\,500\text{mil} \pm 100\text{mil}$ 。其他信号走线长度应小于 $1\,500\text{mil}$ 。对于时钟线可采取蛇形走线的方式。

(3) PCI 总线信号 PRSNT1#的 PRSNT2#中必须有一个接地。

(4)对于不实现 JTAG 边界扫描的板子,必须将引脚 TDI 与 TDO 连接起来,以使扫描链不至于断开。

(5)从 PCI 金手指到去耦电容器焊盘的走线长度不超过 250mil,线宽不小于 20mil。

(6) 注意 FPGA 去耦电容的布局, 使去耦电容与 FPGA 电源和地端连线尽量短。

3 PCB 仿真

由于控制板 PCI 时钟频率较高和布线密度较大,为保证板卡质量,减少开发成本,在设计中我们采用了仿真技术来预测可能引起的信号完整性、串绕以及 EMC 问题,检测已经完成的

布线的传输性能,综合多方面考虑,定出合理的布线约束条件和终端匹配策略等。本设计选用高速仿真工具——Hyperlynx,其提供的 BoardSim 功能用于布线完成后对 PCB 的仿真,它包含快速仿真和详细仿真两种方式。快速仿真用于快速地分析设计中的信号完整性、电磁兼容性和串扰问题,生成串扰强度报告等。详细仿真用于对指定网络进行仿真,可以从示波器观察波形,并可利用软件提供的终端适配向导为电路计算出合适的匹配电阻,从而改善信号质量。

3.1 仿真模型的建立

器件的仿真模型选用了 IBIS 模型^[6]。IBIS 模型是一种基于 V/I 曲线对输入/输出缓冲器快速、准确地提取电气特性的模型,是反映芯片驱动和接收电气特性的一种国际标准。它提供一种标准的文件格式来记录驱动源输出阻抗、上升一下降时间、输入负载等参数,非常适合作振铃、串扰等高频效应的计算与仿真。IBIS 模型一般可从芯片供应商的网站下载。本设计从相关网站上分别获取以下器件的 IBIS 模型: QS3861, Cyclone, AM29DL640D, M48Z2M1Y 等。

3.2 仿真策略及结果分析

由于控制卡电路结构紧凑,PCB 按照电路功能模块进行布局。因此在仿真阶段,主要从以下几个方面对布线后的 PCB 进行板级仿真(BoardSim)。

3.2.1 快速分析整板信号完整性(SI)和 EMC 问题

通过设置 Board Wizard 对整板进行 SI 和 EMC 快速分析。分析结果以报告形式给出。根据报告分析板上所有网络的信号完整性,以下信号串扰较强:LPCI_REQN,LPCI_AD31,LPCI_AD21,LPCI_IDSEL,PCI_FRAMEN 等。因此在关键信号仿真策略中,将对它们进行详细仿真,并寻找其解决方法。

3.2.2 时钟网络的 SI 分析

这部分仿真主要针对 PCI 时钟信号进行仿真。将时钟激励源设置为 66MHz。图 5 为时钟布线及修改前后仿真的波形。



(a) 修改前 (b) 时钟走线 (c) 修改后

图5 时钟走线及修改前后的仿真波形图

从图 5(a) 可以看出,时钟接收端波形正向过冲、负向过冲、振铃现象较为严重。通过分析,在本设计中这些信号的完整性问题主要源于传输线过长、信号线网的阻抗失配,或阻抗不连续。由于 PCI 协议中对时钟走线长度有很严格的规定,因此在保证传输线长度的前提下必须给时钟信号加上适合的终端匹配来改善时钟信号线的传输特性。终端匹配电阻的阻值可以通过 Terminator Wizard 来精确地计算出,匹配方式采取在信号接收端串联一个 50Ω 的电阻。修改后的信号仿真波形得到了很大的改善(图 5(c))。

3.2.3 PCI 关键信号详细仿真

本仿真策略主要是针对 PCI 中的高频信号进行详细仿真,对串扰严重的信号确定其调整策略。从 PCI 金手指到 FPGA 的布线密度较大许多信号线不可避免的需要平行 (下转第 266 页)