一个体全息存储高速数据通道的设计*)

刘 洋 谢长生 吴 非 胡迪青 罗东健
(华中科技大学国家专业存储实验室 武汉430074)

摘 要 为了实现新一代存储技术---体全息存储技术在千兆网中的具体运用,本文介绍了一个体全息存储高速数据通道的硬件与软件的体系结构,阐述了在嵌入式操作系统 VxWorks 上的编写 PCI 设备驱动程序的原理和方法,并给出了我们在 Intel IXP1200评估系统上关于体全息存储高速数据通道的设计与实现。

关键词 体全息存储,数据通道,VxWorks,PCI 设备驱动程序,IXP1200

A Realization for High Speed Holographic Data Storage Channel

LIU Yang XIE Chang-Sheng WU Fei HU Di-Qing LUO Dong-Jian
(Huazhong University of Science&Technology National Storage Laboratory, WuHan 430074)

Abstract To bring a newly storage technology--- holographic storage to real using in the Gigabit Ethernet, this paper introduces the hardware architecture and software architecture of a High Speed Holographic Data Storage Channel. Also i gives the principles and methods to design a PCI device driver based on VxWorks. In the end it presents the design and realization to the high speed holographic data storage channel we builded on the Intel IXP1200 evaluation system.

Keywords Holographic storage, Data storage channel, VxWorks, PCI device driver, IXP1200

1 引言

与传统存储相比,体全息存储具有存储容量大、数据传输率高、存取时间短,数据保存时间长(不怕灰尘、擦伤和电磁场干扰)、数据保密性好、应用灵活以及可快速进行图像匹配和内容相关寻址操作等特点[1,2],成为最有吸引力的技术之一。近十余年来,国内和国际上在体全息存储技术的研究十分活跃,发展也相当迅速。但是很少有人关注与体全息存储相适应的数据通道的设计[3],而随着体全息存储材料等各项技术的成熟,在实现体全息存储实用化的过程中,从计算机用户到体全息存储数据的通道问题日益突现出来。体全息数据存储系统如图1所示[1]。

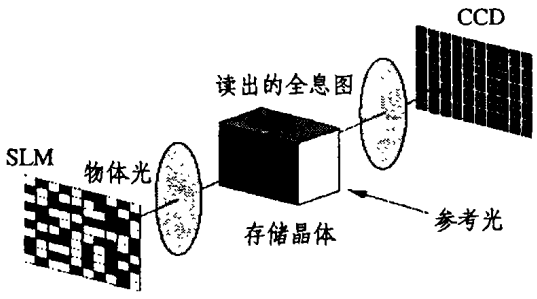

图1 体全息存储系统示意图

全息存储记录数据时,由一束激光照射空间光调制器(spatial light modulator, SLM),包含二元数字0与1的电子数据经过 SLM 后被调制到物体光束上,物体光束与同一束激光产生的参考光束在记录介质中发生干涉形成体全息图并被记录下来。采用各种多重复用方法(multiplexing scheme),可在一个公共体积内记录很多全息图,使之具有非常高的数据存储密度。当要读出某一幅存储信息页时,用适当的参考光照射全息记录介质,根据布喇格衍射定理,衍射光将以物光的强度与方向成像于探测器阵列表面,根据探测器阵列读出的光强大小,光信号又被转化为电子信号。而光电转换的速度很快,所以体全息数据读出也很快。

为了充分发挥体全息存储的优势,使其实用化,我们进行了高速体全息存储通道的设计及实现。经过分析与比较,我们选择 Intel IXP1200评估板为平台,自主开发了与 PCI 2.2规范兼容的编码、解码 PCI 板,基于实时嵌入式操作系统 VxWorks 编写了相应 PCI 驱动程序,搭建起适用于千兆以太网的体全息高速数据通道。

2 体全息存储通道的设计

2.1 体全息存储通道的硬件体系结构

体全息存储系统作为一种新型的存储系统,为了适应千兆以太网的应用,其通道方面需要解决以下几个方面问题:(1)通道要保证体全息存储系统数据的原始误码率通过通道后达到10^{-12};(2)通道的传输速度要求达到100MB/s;(3)提供合适的访问接口标准。

研究表明,从存储系统中读出的数据含有突发噪声和随机噪声[4],体全息数据存储系统也不例外。因此,体全息数据存储系统可以看作一个有噪声的信道,称为全息数据通道。在体全息数据存储系统中,通常采用纠错编码、交错和调制编码相结合的方式来对数据进行编码。

*)本文研究受国家"973"高技术项目资助,课题编号 G1999030106.刘 洋 硕士研究生,主要研究方向为网络存储技术;谢长生 教授,博士生导师,主要研究方向为新型计算机外存储体系结构,网络存储技术,网络多媒体技术;吴 非 博士,讲师,主要研究方向为计算机系统结构,体全息存储;胡迪青 博士后,主要研究方向为体全息存储,精密自动化控制;罗东健 博士研究生,主要研究方向为磁盘阵列,网络存储技术。

为此，我们设计了两块 PCI 板来实现，一块负责编码和 SLM 接口（相当于完成写数据的功能），一块负责解码和 CCD 接口（相当于完成读数据功能），并将其插于 IXP1200评估系统的 PCI 槽上，如图2所示。

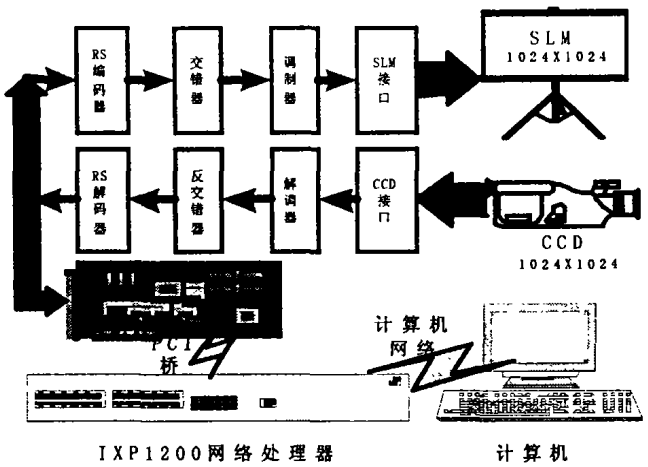

图2 体全息存储高速通道

系统中的 IXP1200处理器由一个高性能的32bit RISC Intel StrongARM 核和六个可编程的微引擎等构成。其中，每个微引擎具有2k 字节的控制存储器以用于开发者存放控制代码。这些微引擎在强大 StrongARM 核的协调管理下，有机地结合在一起，以同样的速度运行，具有强大的网络协议处理能力和很高的网络数据吞吐量。在 IXP1200评估系统中，IX 总线为 IXP1200提供了非复用的32bit 收、发外部数据信道。并且，进入 IX 总线的数据可以直接传送到微引擎或外部的内存中。我们设计的高速通道中，数据及读写请求就是通过连在 IX 总线上千兆以太网口进出 IXP1200内核。另外，在 IXP1200评估系统中还有一条通用 PCI 总线，独立的 SDRAM 和 SRAM 控制器，串口控制器等各种各样的功能部件[5]。我们的编码、解码 PCI 板卡及系统自带的一块 PCI 以太网卡正是通过这条 PCI 总线与 IXP1200建立连接和通信。

在所设计的 PCI 板中使用 PCI9054 I/O 加速芯片和 Serial EEPROM 实现图2通道方案中 PCI 桥的功能，而使用 Altera Stratix 系列 FPGA 则完成体全息存储数据的编码、解码处理和实现与 SLM 与 CCD 的接口。

2.2 体全息存储通道的软件体系结构

为了使所设计成为真正适用于千兆网的体全息存储高速数据通道，我们在 IXP1200网络处理器上还需要三个程序模块，如图3所示。其中网络模块主要由运行于 IXP1200微引擎上的微码和 IXP1200 StrongARM 上伪以太网络驱动程序构成；微引擎接受从千兆网口发来的网络数据包，交付微码完成 IP 等网络协议的解析，如果涉及到更上层的应用则通过伪以太网络驱动接口交给 StrongARM 核处理；伪以太网络驱动除了负责 StrongARM 核与微引擎的通信，还要解析上层的网络协议，如是对体全息存储数据的访问请求则交给体全息存储文件系统处理。而文件系统则针对体全息存储体以二维页面为访问单位进行专门的设计，透明化体全息存储的读写操作。最后 PCI 设备驱动程序对两块编码、解码 PCI 板驱动，真正实现对体全息存储数据的访问。网络模块部分已由 IXP1200评估系统提供。一般的计算机系统中，文件系统与相应的磁盘驱动常是合作一体的。这里把它们分开是为了容易开发，更因为 PCI 驱动与体全息数据直接相关，并是整个软件开发中的难点，所以开发 PCI 驱动是我们设计的一个关键。在后面我们将更详细地介绍 PCI 驱动程序的设计。

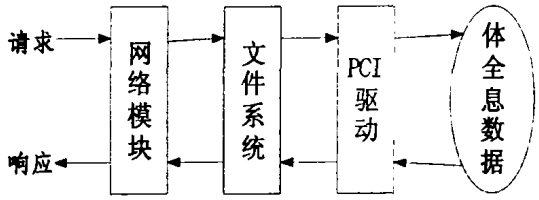

图3 体全息数据通道数据流图

由于 IXP1200评估系统自带了一个可引导的 VxWorks 映像，更因为 VxWorks 是一种类似于 UNIX 的高性能嵌入式实时操作系统，具有高度可裁剪的微内核结构，高效的多任务调度，灵活的任务间通信手段，确定的 us 级的中断延迟时间等诸多优点，所以我们的软件开发基于 VxWorks。VxWorks 的系统结构如图4所示。

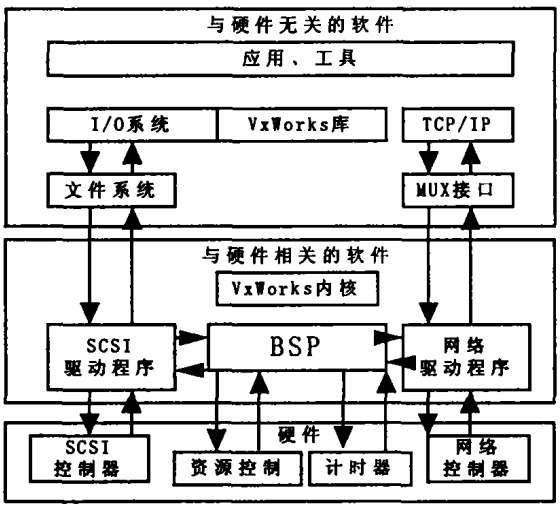

图4 VxWorks 系统结构

其中，BSP（Board Support Package）主要用来完成 VxWorks 对专用目标板的支持。一个 BSP 包括硬件初始化、中断处理和产生、硬件时钟管理、本地和总线内存空间映射，同时也包含定制 VxWorks 映像。它贯穿着硬件级、操作系统级和应用程序级3层。

按照操作系统相应规范编写对硬件的驱动，然后挂接于 VxWorks 的内核中与 VxWorks 一起为应用程序级提供服务。在 VxWorks 中，硬件驱动程序被分为两类：通用的和 BSQ 专用的。通用的驱动程序管理可以在不同的目标环境之间移动的设备，例如网卡；而 BSP 驱动程序管理专用于某种目标环境之间移动的设备，例如中断控制器。因此，在编写设备驱动程序时，可以根据具体情况将其放合适的位置。而我们开发 PCI 驱动程序参照 IXP1200评估系统的 PCI 网卡驱动，并将其挂接在例程 usrRoot()中（usrRoot()是系统加电后启动 VxWorks 所要产生的根任务）。

3 PCI 设备驱动程序的设计

在 VxWorks 中，设备驱动程序又分查询方式和中断方式两种。无论采用哪一种方式，设备驱动程序的基本流程都是相同的：首先是获取接口参数，接着设置硬件寄存器和实现接口函数，最后启动设备。

在第一步中获取的硬件接口参数包括内存映射地址，

（下转第227页）

例如：$f(a,b,c,d)=\sum_m(0,1,2,5,7,8,10,13,15)$

ab \ cd	00	01	11	10
00	1	1		1
01		1	1	
11		1	1	
10	1			1

图2　$n=4$卡诺图

按上述规则：

(1)S^0中有：$S_0^0=m_0=(0000)$；$S_1^0=(m_1,m_2,m_8)=(0001,0010,1000)$；$S_2^0=(m_5,m_{10})=(0101,1010)$；$S_3^0=(m_7,m_{13})=(0111,1110)$；$S_4^0=(m_{15})=(1111)$

(2)求S^1。由S_0^0，S_1^0可得：

$S_0^1=(m_0,m_1)(m_0,m_2)(m_0,m_8)=(000X,00X0,X000)$

由S_1^0，S_2^0可得：

$S_1^1=(m_1,m_5)(m_{10},m_2)(m_{10},m_8)$

$=(0X01,X010,10X0)$

$S_2^1=(m_5,m_7)(m_5,m_{13})$

$=(01X1,00X0,X101)$

$S_3^1=(m_7,m_{15})(m_{13},m_{15})=(X111,11X1)$

(3)求S^2。由S_0^1，S_1^1可得：

$S_0^2=(X000,X010)(00X0,10X0)$

$=(X0X0,X0X0)$

$S_1^2=(X111,X101)(11X1,01X1)$

$=(X1X1,X1X1)$

(4)求S^3不存在。

2.2　找出其中不能再合并的项

即：$000X$，$0X01$，$X0X0$，$X1X1$是素隐含项，从卡诺言图化简可得到$0X01$是冗余项，在Q-M算法里还得用最小覆盖去掉冗余项。

2.3　最小覆盖

一个函数的覆盖由包含此函数的所有"0"维立方体的素含项组成，当覆盖不包含另一个覆盖时，此覆盖为最小的。为了确定最小覆盖在Q-M算法中需要进行：识别EPI(素隐含项)，更新集合和除去冗作项。

为简化说明此方法，依照卡诺图中在相邻最小项合并乘积项圈画的原则：在圈的最小项中至少可以找到一项未被圈过一次的最小项，这样可以将函数所包含的最小项和已确定素隐含项组成最小项——素隐含项表，在表中将素隐含项所包括的最小项的位置上画"○"，然后把纵列上只有一个"○"换成"◎"，则"◎"所在行对应的素隐含项不是冗余项。在余下素隐含项中若除去此行，余下的素隐含项存在某纵列只有一个"○"则除去的素隐含项是冗余项。

按照此方法，上述例子中，显然$0X01$或者$000X$是冗余项。

由图3最小覆盖的素隐含项仅为$000X$，$X0X0$，$X1X1$或者$0X01$，$X0X0$，$X1X1$，化简结果为：$y=f(a,b,c,d)=\overline{A}\overline{B}\overline{C}+\overline{B}\,\overline{D}+BD$，与卡诺图化简的结果完全相同。

m	m_0	m_1	m_2	m_5	m_7	m_8	m_{10}	m_{13}	m_{15}
000x	○	○							
0x01		○		○					
x0x0	○		◎			○	○		
x1x1				○	○			◎	○

图3　最小覆盖化简

小结　在决定函数的素隐含项时，找出所有最小项，采用逐项逐位比较的办法即可找出素隐含项，这是计算机程序设计善长之处。

在找出函数的最小覆盖时，用矩阵的办法并不难找出冗余项。

Cube运算的实质仍然是卡诺图化简，但卡诺图是一个二维的平面图，Cube运算将其扩展为多维的空间坐标体系；卡诺图化简时简单直观，但变量不能太多，Cube运算繁锁抽象，适合计算机编程实现。可见Cube运算是EDA中对电路综合的行之有效的化简方法。

参 考 文 献

1　李亚民. 计算机组成与系统结构. 北京：清华大学出版社，2000
2　孟宪元，李广军. 可编程ASIC设计与应用. 成都：电子科技大学出版社，2000
3　江国强. 现代数字逻辑电路. 北京：电子工业出版社，2002
4　薛宏熙，边计年. 数字系统设计自动化. 北京：清华大学出版社，2002

(上接第206页)

I/O端口和系统中断控制器的输入(IRQ)。获取这些参数的方法由硬件的接口方式决定。PCI总线作为一种即插即用的总线结构，在BOOTROM和操作系统的支持下，能够自动为设备分配合适的硬件接口参数。

硬件的行为和特性是由内部的寄存器控制的。基于PCI总线的系统采用内存映射来访问寄存器。

对于采用中断方式的硬件设备，在接口函数中必须实现中断服务程序。中断程序的编写必须遵循一条规则：不能有运行时间过长的代码；不能独占共享资源以避免死锁；程序结束后尽可能地快速返回。

上述步骤完成后，即可启动硬件设备。

结束语　考虑到作为下一代存储技术的体全息存储的存储容量大，并行速度高，又兼顾刚开始价格较高，我们将其定位于高速网络数据存储服务器应用，所以面向千兆以太网来设计这个的体全息存储数据通道。

我们在利用网络处理器IXP1200的传统网络处理的同时，基于嵌入式实时系统VxWorks充分挖掘了其StrongARM核的强大潜能，使得体全息存储与千兆以太网融为一体，开辟了一条通向新型体全息存储体的高速数据通道。经过测试，我们所设计的体全息高速数据通道的有效数据传输速率达100MB/s，完全满足当初的设计要求。

参 考 文 献

1　Psaltis D, Burr GW. Holographic Data Storage. Computer, 1998, 31(2): 52～60
2　Kevin C, William W, Lisa D. Commercialization of Holographic Storage at InPhase Technologies. IEEE Invited Paper, 2002
3　Wu Fei, XIE ChangSheng, Hu DiQing, Wu Ming. The Design and Research of High-Speed Channel for Volume Holographic Data Storage. APOC 04-157, 2003
4　Ju C, Dai F, Hong J. Electron. Lett., 1996, 32(15): 1400
5　Intel Corp. IXP1200 Evaluation System User's Manual, 2001