文章编号: 1001-3806(2004)02-0121-03

# 一种宽带光总线互连网络机群系统设计

黄 平, 罗风光, 曹明翠, 王江义
(华中科技大学 激光技术国家重点实验室, 武汉 430074)

摘要: 介绍了一种宽带光总线互连网络机群系统。该系统利用宽带多层自由空间光学数据总线和 VCSEL/MSM-CMOS 光电子集成器件实现 4 台计算机之间互连。专用通讯适配器实现高性能并行接口(high performance parallel interface, HIPPI) 到外设部件互连 (peripheral component interconnection, PCI) 协议转换。该系统光互连网络传输容量可达 60Gbit/s, 可用于计算机之间的高速通讯。

关键词: 垂直腔表面发射激光器; 光互连网络; 机群; HIPPI; PCI
中图分类号: TP393.03 文献标识码: A

# Design of PCs cluster system of broadband optical bus interconnection network

*HUANG Ping, LUO Feng-guang, CAO Ming-cui, WANG Jiang-yi*
(National Laboratory of Laser Technology, HUST, Wuhan 430074, China)

**Abstract:** A PCs cluster of broadband optical bus interconnect network is introduced in this paper. The broadband multiplayer free-space optical data bus and VCSEL/MSM-CMOS integrate device are used to realize the interconnection of four computers in this system. The special communication adapter is used to realize the conversion between HIPPI (high performance parallel interface) protocol and PCI (peripheral component interface) protocol. The capacity of the optical interconnection network is 60Gbit/s which is fit for high speed communication between the computers.

**Key words:** VCSEL; optical interconnection network; PCs cluster; HIPPI; PCI

版权所有 © 《激光技术》编辑部

## 引 言

随着计算机和通讯技术, 特别是高速信息通讯网的飞速发展, 要求计算机处理速度向着每秒 1 千亿次、1 万亿次甚至更快的高性能计算机方向发展。为满足以上要求, 发展高性能并行多处理计算机系统一直是计算机领域研究的热点课题[1,2]。传统的电子计算机采用"线互连"方式, 但电子"线互连"方式存在严重的串话、RC 时延、时钟歪斜、瓶颈阻塞和随着电路规模的增大功耗急剧增加等缺点[2,3]。而"光互连"具有极高时间空间带宽积、高速、高密度、无干扰和传输功耗极低等优点。所以用"光互连"替代"电互连"是必然趋势。笔者设计利用光互连的优势, 采用多层自由空间光学数据总线和VCSEL/MSM-CMOS 光电子集成器件实现 4 台计算机之间的互连[4]。

## 1 系统结构及工作原理

该系统主要由光互连交换网络系统和计算机机群组成。

基于 VCSEL-CMOS 光电集成器件的光互连交换网络系统, 主要由二维光纤列阵光电收发接口组件、宽带多层自由空间光学数据总线模块和VCSEL/MSM-CMOS 光交换光电集成数据收发模块 3 大部分组成, 如图 1 所示。该系统采用分布式的 4 组

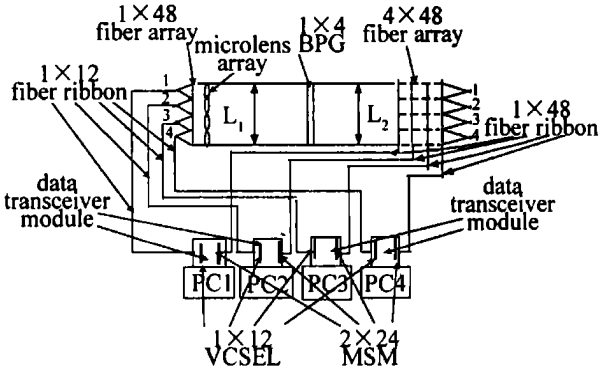


Fig. 1 The structure of parallel optical interconnection network

基金项目: 国家八六三计划资助项目(2001AA111030)

作者简介: 黄 平(1979-), 男, 博士研究生, 现主要从事光通信、光互连网络方面的研究。
E-mail: huangi_0@etang.com

收稿日期: 2003-05-12; 收到修改稿日期: 2003-06-18

1× 12 列阵的 VCSEL 垂直腔表面光发射器件和 2× 24 列阵的 MSM 探测器列阵, 并针对 VCSEL/ MSM-CMOS 光电集成模块光收发像元列阵窗口规则分布的特点, 研制了具有高密度规则排列的二维光纤列阵光电收发接口组件( 专利号: ZL 98 1 13444. 0) , 利用由微透镜阵列( microlens array, MLA) 、棱镜 $L_1$、棱镜 $L_2$ 以及二元位相光栅( binary phase grid, BPG) 组成的 4*f* 系统将数据进行空间广播, 从而构成了能互连 4 台计算机, 每台计算机的数据信号可通过 12 条光纤并行传送的自由空间光互连网络系统。

该系统可实现 4 台计算机之间的双向光互连。每台计算机接口位宽为 12bit。每台计算机都有 1× 12VCSEL 垂直腔表面发射激光器列阵和 2× 24 MSM 探测器列阵收发模块。计算机中的数据信号通过并/串转换进入 12 条 VCSEL 发送通道, 由 1× 12 VCSEL 发射窗口发出的数据光信号通过光纤列阵接口器件耦合到光纤列阵中进行并行传输。4× 12 并行传输光纤列阵在宽带多层自由空间光学数据总线的入口处形成 1× 48 的输入信号光束列阵, 输入到自由空间光互连光学数据总线模块。光互连数据总线由 4*f* 成像系统和位相计算全息光栅构成。经过 1× 4 位相计算全息光栅分束和 4*f* 系统成像, 在光学数据总线输出口处形成 4× 48 的输出信号光束列阵, 由 4× 48 高密度规则排列的多层二维光纤列阵光电收发接口器件耦合后, 经重新排列, 分别输入到每台计算机的 2× 24MSM 探测器的窗口列阵, 经光电转换后的电信号, 由 CMOS 选通电路选择其中的一组 1× 12 并行输入数据信号输入到该计算机, 再由专用通讯适配器进行数据转换, 通过 PCI 总线送入计算机, 从而实现任意 4 台计算机之间的双向互连通信。

该并行计算平台光互连系统采用微光学器件构成自由空间光学广播网络, 采用了 VCSEL 列阵集成光发射模块, MSM 列阵集成光接收模块, 高精度光纤列阵 I/ O 接口, 使得整个互连网络具有集成度高、体积小的优点。由于采用 VCSEL 列阵作为激光源, 每个节点拥有 12 个光通道, 每条通道速率达 1. 25Gbit/ s, 整个光互连网络总的传输容量可达 60Gbit/ s。

## 2 专用通讯适配器的设计

每台计算机都有一个通讯适配器。适配器主要由 O/ E, E/ O 转换装置、P/ S, S/ P 模块、选择模块和 HIPPI-PCI 协议转换模块组成。如图 2 所示当主机接收数据时, 该适配器负责将输入的 4 路 1× 12 串行光信号转换为 4 组 1× 48 的并行电信号, 并且通过仲裁选通其中一组信号接收, 通过 HIPPI-PCI 协议转换模块进行转换, 通过计算机的 PCI 接口将数据送入主机处理。当主机要发送数据时, 数据通过 PCI-HIPPI 协议转换模块输出, 信号并/ 串转换后, 通过发射装置发送出去。适配器原理图如图 2 所示。

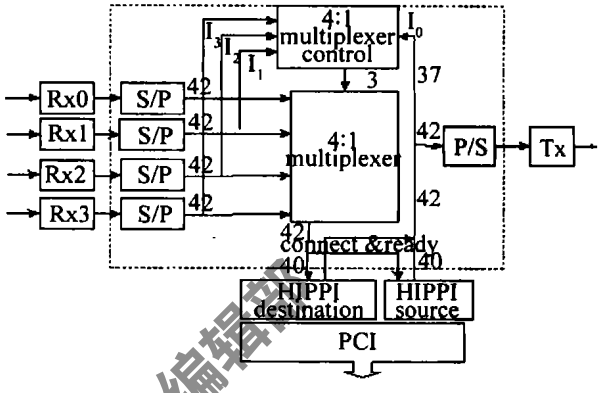


Fig. 2 Schematic diagram of network adapter

## 3 PCI 与 HIPPI

PCI 局部总线[5] 是一种高性能、32 位或 64 位地址/ 数据线复用的总线。其用途是在高度集成的外设控制器器件、扩展板和处理器之间提供一种连接机制。

如图 3 所示, CPU/ cache/ DRAM 通过一个 PCI 桥联接。外设板卡, 如 SCSI 卡、网络卡( LAN card) 、声卡( audio card) 、视频卡( graphics card) 、图像处理卡( motion video card) 等高速外设, 挂接在 PCI 总线上。基本 I/ O 设备, 或一些兼容 ISA 总线的外设, 挂接在 ISA 总线上。ISA 总线与 PCI 总线之间由扩展总线桥连接。典型的 PCI 总线一般仅支持 3 个 PCI 总线负载, 由于特殊环境需要, 专门的 PCI 总线变种-工业 PCI 总线, 可以支持多于 3 个 PCI 总线负载。

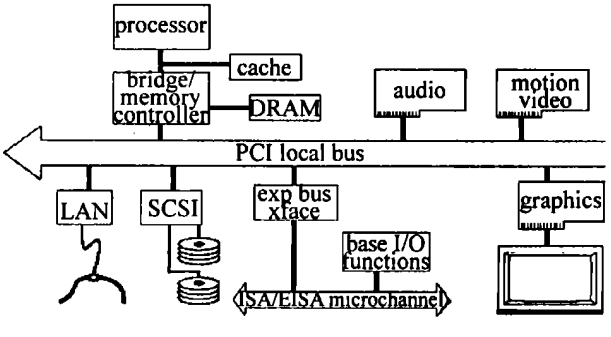


Fig. 3 Schematic diagram of PCI bus system

HIPPI 协议全称为高性能并行接口协议。该协议 1987 年由美国 Los Alamos 国家实验室提出, 最早

版权所有 ©《激光技术》编辑部

用于实现超级计算机和数据存储器之间的高速数据通道, 3 年后成为 ANSI 标准。HIPPI 协议包括 800Mbit/s 和 1.6Gbit/s 两种规范。尽管 HIPPI 协议最早用于超级计算机应用系统中, 但后来的研究表明, 该协议还可以用在其它的数据设备中, 例如工作站, 监视器, 帧缓冲设备等。

HIPPI 协议[6~8]包括机械、电气和信号定义规范 HIPPI-PH; 帧协议规范 HIPPI-FP; 交换协议规范 HIPPI-SC; 以及 HIPPI-LE, HIPPI-FC, HIPPI-IPI 等规范。

HIPPI 协议为单向点到点传输, 数据发送一端称为源端(source), 数据接收一端称为目的端(destination), 见图 4。

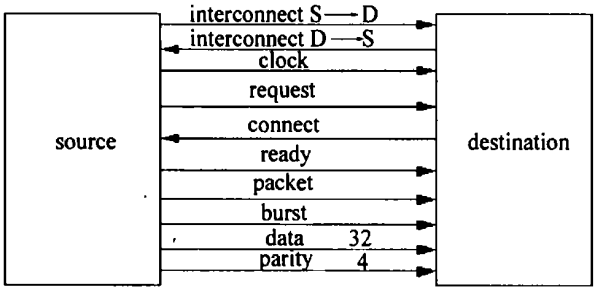


Fig. 4　Description of HIPPI protocol signal

## 4　HIPPI-PCI 协议转换模块设计

HIPPI-PCI 模块完成接口卡和计算机总线之间的接口逻辑, 见图 5。该模块采用 AMCC 公司的专用芯片 S5933, S2020/S2021。外围控制逻辑采用 Altera 公司的 EPLD 逻辑器件。

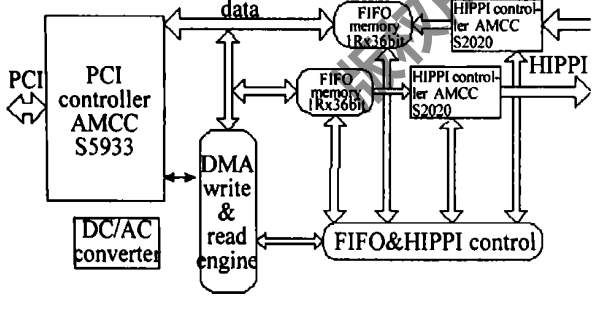


Fig. 5　Schematic diagram of HIPPI-PCI protocol conversion

其中 S5933 实现与 PCI 接口之间的联系。为了数据的高速传输, 决定采用 DMA 模式。外部 DMA write & read engine 用来控制 S5933 内置 FIFO 与外置 FIFO 之间的 DMA 方式的数据传输。S2020 负责接收外部输入的 HIPPI 协议信号, S2021 负责以 HIPPI 协议方式发送主机给外设的数据。FIFO&HIPPI control 负责控制外置 FIFO 与 S2020, S2021 芯片。DMA write & read engine 与 FIFO&HIPPI control 采用 Altera 公司的 EPLD 逻辑器件。在电路设计过程中要注意 S5933 芯片要尽量接近金手指, 这样有利于数据的高速传输。

## 5　总　结

设计了一种宽带光总线互联网络机群系统。该系统上层采用光互连交换网络, 利用光互连的高带宽、大容量的优点。下层采用自行设计的网络适配器实现计算机与光互连网络之间的数据传递。该系统可用于计算机之间的高速通讯。

参　考　文　献

[1] MICHAEL R R, CHRISTENSEN M P, HANEY M W. J Lightwave Technol, 1996, 14(9): 1970~1978.

[2] HANEY M W, CHRISTENSEN M P. Appl Opt, 1998, 37(14): 2886~2894.

[3] NAKAHARA T, MATSUO S, FUKUSHIMA S. Appl Opt, 1996, 35(5): 860~871.

[4] 罗风光, 曹明翠, 徐　军 *et al*. 光纤通信, 2001, 23(5): 28~29.

[5] 曾繁泰, 冯保初. PCI 总线与多媒体计算机. 北京: 电子工业出版社, 1998. 12~15.

[6] RONALD R. High-performance parallel interface-physical switch control(HIPPI-SC). http://www.hippi.org/dsc32m.pdf, 1999-02-04/2003-02-04.

[7] RONALD R. Information system-high-performance parallel interface-mechanical, electrical, and signalling protocol specification(HIPPI-PH). http://www.hippi.org/dPH82.pdf, 1999-02-04/2003-02-04.

[8] RONALD R. Information technology-high-performance parallel interface framing protocol(HIPPI-FP). http://www.hippi.org/dfP48m.pdf, 1999-02-04/2003-02-04.

版权所有©《激光技术》编辑部