●新特器件应用

千兆以太网交换芯片 BCM5690 及其在交换整机中的应用

王亚平[1]，　蔡勖[1]，　罗小巧[2]

（1. 华中师范大学 物理科学与技术学院，湖北 武汉 430079；2. 华中师范大学 信息技术学院，湖北 武汉 430079）

摘要：BCM5690 是 BroadCOM 公司推出的集成有 12 个千兆端口和 1 个万兆端口的多层交换芯片。文章比较全面地介绍了该芯片的结构和功能特性，给出了它的访问控制方式和数据流程，同时给出了用 BCM5690 设计交换整机的硬件结构和软件实现方法。

关键词：千兆以太网；　BCM5690；　堆叠；　数据流程

分类号：TP393　　**文献标识码**：A　　**文章编号**：1006 – 6977(2005)03 – 0046 – 04

Scalable 12 – port Gigabit Ethernet Multilayer Switch BCM5690 and Its Application in Switchboard

WANG Ya-ping，　CAI Xu，　LUO Xiao-qiao

(Central China Normal University，Wuhan 430079，China)

Abstract: The multilayer switch chip BCM5690 is introduced，which is new product of the Broadcom cooperation and integrated with 12 – port Gigabit and single 10 Gbps XAUI HiGig expansion port. And the chip's characteristics and function are given completely，as well as it's realization in switchboard.

Keywords: Gigabit Ethernet；　BCM5690；　stacking；　data flow

目前，万兆芯片技术不断取得新的发展，尤其在“真”万兆的问题上，只有拥有更先进的芯片，设备厂商才能够在芯片功能与特性的基础上开发自身的交换机体系架构。虽然 Broadcom、Intel、Marvell、美国国

4　典型应用

图 3 是 SAA7185 的一个应用电路。其中电路的模拟和数字电源均为 5V。该电路为 VCD 小影碟机和多媒体电影解压卡中常用的数字编码电路，系统中的输入数据可以是 16 比特 YUV 方式，也可是 CCIR656 格式。图中输入的是 8 比特 MPEG – 1 数据，采用与 CCIR656 兼容的方式。在该系统中，SAA7185 编码器工作于主模式。

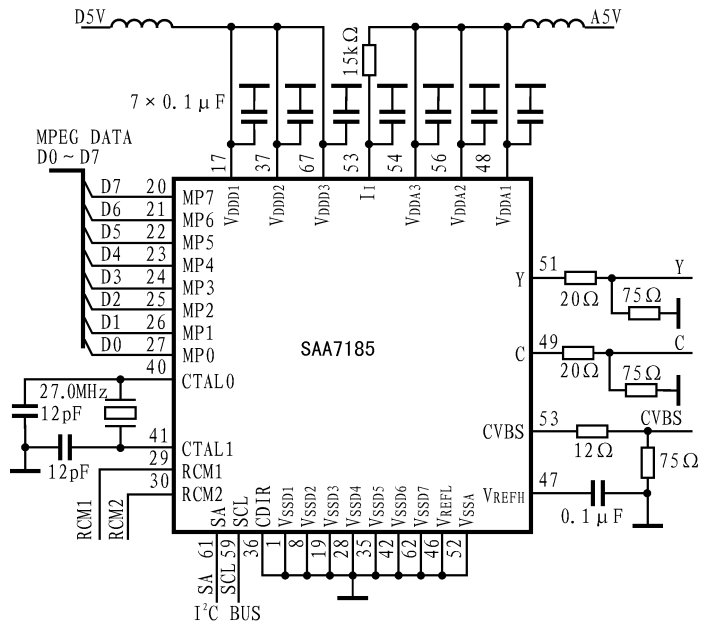

图 3　SAA7185 的典型应用电路图

参考文献

[1] Philips Semiconductors. SAA7185 Digital Video Encoder (DENC2). 1996，7.

[2] 王新成. PAL 全电视信号的特征与处理技术[J]. 电视技术，1997，(5).

收稿日期：2004 – 08 – 27

咨询编号：050315

家半导体（NS）、英飞凌（Infineon）和意法半导体（STMicro electronics）都推出了最新的千兆以太网芯片产品。但万兆芯片的发展无疑会从硬件与架构层面来加快万兆产品的发展速度。

为此，BroadCOM 公司开发了 BCM5690(12 + 1) 单芯片交换方案。该集成电路芯片集成了 12 个千兆端口和 1 个万兆端口，是一款功能比较强大和全面的三层千兆以太网交换芯片。文中将详细介绍 BCM5690 芯片的功能特性以及基于该芯片的交换机实现方法。

1 BCM5690 芯片简介

1.1 BCM5690 芯片结构

BCM5690 是芯片提供有 12 个 GE 接口（千兆端口）和 1 个 HiGig 接口（内联端口），并具有堆叠功能。器件的端口采用 PCI 接口进行管理。其结构框图如图 1 所示。

由图 1 可以看出：BCM5690 芯片由以下一些主要功能模块组成：

（1）GIGA 接口控制器 GPIC：用于提供 GE 口与交换逻辑之间的接口。

（2）内联端口（HiGig）控制器 IPIC：主要提供 HiGig 口与内部交换逻辑之间的接口，有时也被用于多片 BCM5690 之间的堆叠操作。

（3）CPU 管理接口 CMIC：主要提供 CPU 与 BCM5690 设备不同功能块之间的接口，同时也用于诸如 MIIM、I^2C 和灯的处理等功能。该模块通过 PCI 接口与 CPU 相联，可使 CPU 访问和控制 BCM5690，而 DMA 引擎则支持数据从 CPU 传向 BCM5690 或从 BCM5690 传向 CPU。

（4）地址解析逻辑 ARL：该逻辑功能模块可在数据包的基础上确定该数据包的转发策略。它利用二层表（L2-TABLE）、二层组播表（L2MC TABLE）、三层表（L3-TABLE）、三层最长前缀匹配表（DEF-IP-HI 和 DEF-IP-LO）、三层接口表（L3 – INTF）、IP 组播表（L3-IPMC）、VLAN 表（VLAN）以及 spanning tree Group 表（VLAN-STAG）来决定如何转发数据包。

（5）公共缓冲池 CBP：CBP 实际上是 1MB 共享的包缓冲区。CBP 由 8192(8K) 个单元组成，每个单元 128 字节。设备里的每个数据包消耗一至多个单元。

（6）内存管理单元 MMU：BCM5690 有一个单独的内存管理单元。每个 MMU 与设备的功能块（GPIC、IPIC 等）相关联。MMU 负责数据包的缓冲和调度。它首先接收数据包，然后再将数据包缓冲，并在发送时加以调度，同时它还管理交换单元的流控特性。概括来说，就是缓冲逻辑、调度逻辑、流控逻辑。缓冲逻辑从 CP – BUS 接收包并存放在 CBP，同样也从 CBP 获取包并将它们发送到 CP – BUS 上去。包的发送顺序由调度逻辑根据包的优先级别确定。流控逻辑包括 Head – of – Line(HOL) 阻塞预防和 Backpressure 两种方式。

这些功能模块之间可通过两条内部总线联系起来：CP – Bus（图 1 粗黑线所示）、S – Channel Bus（图 1 细黑线所示）。其中 CP – Bus 用于芯片内数据包的高速传输，它支持所有端口的同时线速转发。而 S – Channel Bus 则有两个作用：第一是用于 MMU 到其它功能块的流控；第二是通过 CMIC 利用软件控制来访问内部寄存器和表。

1.2 BCM5690 芯片特性及功能介绍

Broadcom 公司 XGS 系列芯片的重要特性是具有堆叠功能，该功能可以将多个交换芯片组合在一起，以形成一个更大规模的系统，或者将多个带交换芯片的系统组合在一起形成一个完整的系统设备。这种功能最多可以实现 30 个设备的堆叠。

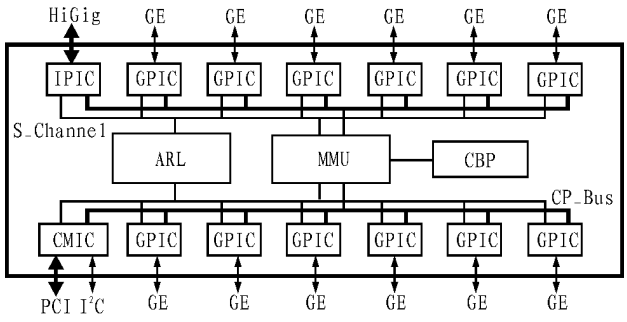

图 1 BCM5690 结构框图

BCM5690 能够通过 Hi – Gig 和 GIGA 来扩展系统容量。它有四种模式，其中 cascade 模式通过 Hi – Gig 口单向互联来形成（环行组网）；而全双工堆叠模式则通过 BCM5690 的 Hi – Gig 口双向地与 BCM567X 相连来扩展容量（环形组网）；第三是 chassis 模式，该模式是将 Hi – Gig 通过背板

互连形成(星形组网)。可实现冗余备份和逐级交换，不需中转，且效率比环行组网高；最后是 SL 形式的堆叠，它通过GIGA 口来互连 BCM5690。图 2 所示为 BCM5690 的逻辑框图。

BCM5690 是一款千兆以太网交换芯片，它支持二层交换、三层路由以及第 2～7 层数据包的分类和过滤等。

地址解析逻辑是 BCM5690 集成电路芯片的中心部件，GPIC 的入口逻辑用它来决定单个包的转发方向。

BCM5690 集成电路芯片中的快速过滤处理器(FFP)是一个通过第 2～7 层数据包进行分类和过滤的引擎。每个 GE 口各有一个 FFP 来负责包的分类和更新。FFP 可以通过配置寄存器 GIMASK 和 GIRULE 来改变符合条件的数据包特性，其中 GIMASK 用于配置匹配选项，GIRULE 用于产生操作命令。对于每个数据包来说，最多可以改变 16 个匹配特性。如果同时有多个特性符合匹配条件，则在 GIMASK 里处于高位的优先配置。

对于包的缓冲和流控，BCM5690 还集成了 1MB 的数据包缓冲区 CBP，这个缓冲区可为所有端口共用。

BCM5690 中的寄存器 MIRROR-CONTROL 和 IMIRROR-CONTROL 用来设置被镜像端口与镜像端口，两个寄存器的内容应保持一致。它支持本芯片内的镜像。

BCM5690 集成电路芯片中的链路聚合(trunk)最大可支持 8 端口的 Trunking，共 32 组 Trunk，并可进行跨芯片的端口 Trunk。另外，BCM5690 还支持速度高达 66MHz 的 PCI 接口，并可对所有数据包的线速交换以及 RMON、SNMP、STP 和 Rapid STP 提供支持。

2　访问控制方式及数据流程

2.1 访问控制方式

BCM5690 支持一系列符合 PCI 标准的寄存器，这些寄存器允许对设备再分配(MODID)、地址空间的自动配置和再映射。CPU 对 BCM5690 的控制都是通过访问 PCI 寄存器来实现的。BCM5690 的寄存器分为直接访问和非直接访问两种。可直接访问的寄存器映射到 PCI 的内存空间，这些寄存器相对于 PCI 控制寄存器有一个固定的地址偏移。上电初始化期间，系统自动配置每个 PCI 设备的基地址及地址范围，以便能够唯一地访问每一个 PCI 设备中的寄存器。BCM5690 的访问机制分为三个类型：一是 PCI 配置空间；二是 PCI memory 映射的 I/O，比如通过 PCI 设备对 DMA、MIIM 和 I^2C 的控制；第三是消息机制。

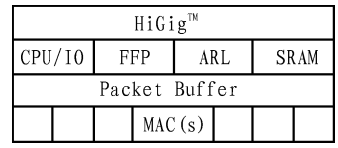

图 2　BCM5690 的逻辑框图

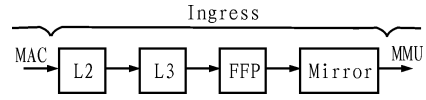

图 3　数据流程

2.2 数据流程

所有的数据流通过交换芯片都要经过输入部分(Ingress)、内存管理单元(MMU)和输出部分(Egress)这三个流程。其数据流程如图 3 所示。

Ingress(输入逻辑)是数据包在每端口上的逻辑流程。每端口都有自己的输入逻辑，输入逻辑负责所有包的转发(交换)策略，决定将包送给哪个端口，根据转发信息将数据包发送给 MMU，进行缓冲和调度，并以线速对包进行处理。输入逻辑与大部分交换功能关联。

MMU(内存管理单元)主要负责数据包的缓冲与调度，它接收从输入逻辑过来的数据包并缓冲这些包，同时对这些包进行调度并将它们送给输出逻辑。数据包进入 MMU 时将存储在 CBP 里。CBP 有 1MB 的大小供所有端口共用。MMU 主要有四部分功能：资源计数、背压、HOL 预防机制、调度。其中资源计数主要是统计当前消耗的 CBP 单元数或 CBP 数据包个数，决定数据包什么时候进入背压或 HOL 预防；调度则是根据优先级和 COS 的四种调度准则来确定包发送的先后顺序和权重。

Egress(输出逻辑)主要从 MMU 获取数据包并将其送入各个端口，这是整个流程中最简单的一环。它先将包发给输出端口，然后确定是否在发送的数据包上添加 tag。具体流程如下：首先从 MMU 请求数据包，如果发送的包要求不带 tag，则负责将 tag 去

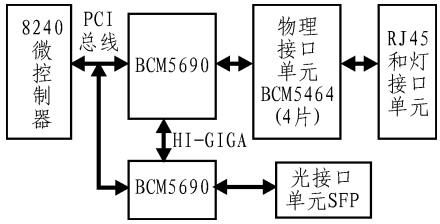

图 4 硬件整体结构图

掉；然后计算数据包的 CRC；最后将包交给 MAC 发送(特殊情况下，CMIC 将直接把数据包以 DMA 方式发送给 CPU)。

3 基于 BCM5690 的交换整机设计

3.1 硬件实现

笔者在设计中采用了两片 BCM5690 和四片 BCM5464，外加四个 SFP 接口的千兆光接口来实现 16 端口的 10/100/1000M 的电口和 4 个千兆光口的交换整机。BCM5464 是 BroadCom 的四端口的千兆 PHY。

图 4 是该交换整机的主控板硬件结构，其硬件电路由交换单元、物理接口单元、RJ45 和灯接口单元、光接口单元、控制单元、CPU 连接器单元、时钟单元、电源单元组成。其中交换单元选用了两片 BCM5690，它们之间通过 Hi – Gig 口背靠背连接实现通信，带有与其它大部分单元的接口。每片 BCM5690 通过 PCI 接口与 CPU 连接器相连，主要用来与 CPU 通信以实现对芯片的管理。

图 4 所示电路中的 CPU 采用 Motorola 公司的 PPC8240，主要负责整个系统的运转调度。12 个电口通过背板总线与各个线卡相连，以实现各线卡的上联功能。当主控板单独作为一个独立的三层交换机时，它将同时作为与其它三层设备互联的接口。4 个千兆光接口与本设备上联可扩展以太网的传输距离。

本系统使用了两个背靠背连接的 BCM5690 来进行设计，这样可将该系统归为 cascade 模式。在堆叠完成之后，通过在堆叠口的以太网包首部加上头信息，可使芯片与芯片之间、系统与系统之间通过 Broadcom 专有的通信协议来实现相互之间的信息传输，从而实现数据包在不同芯片、不同系统之间的转发。

另外，还可以通过设置 IPIC CONFIG 寄存器(IPIC-CASCADE，MY-MODID)、MODPORT 表和 Gigabit 端口 CONFIG 寄存器来对堆叠进行配置。

3.2 软件实现

a. 初始化流程

在设计软件模块的初始化流程时，首先是头模式的设置，由于 BCM5690 固定工作在小头(little-endian)模式，而 PPC8240 工作在大头模式，因此需要对头模式进行设置；接下来是查找 PCI 设备，以获取各个 PCI 设备的设备号以及各自的基地址；之后应对堆叠模式进行设置，以便两片 BCM5690 之间的二层表内容能够互通；最后是基本交换功能和 DMA 通道的配置。

b. 软件结构

图 5 是笔者设计的软件结构简图。其中涉及驱动程序和驱动程序封装的是 SAL 层、Driver 层和 BCM 层。SAL 层可对操作系统与驱动层进行隔离，可提供 PCI 中断以及 PCI 设备的查找、线程、中断、同步和内存管理。Driver 层包括 BCM5690 寄存器的访问方式实现、表的初始化、内存初始化、芯片堆叠模式的设置、L2 和 L3 地址的操作与查找、数据包的发送/接收以及端口的管理等。

参考文献

[1]StrataXGS BCM5690/BCM5691/BCM5692/BCM5693 Programmer's Reference Guide.

[2]StrataXGS BCM5690 Theory of Operations.

[3]Broadcom StrataSwitch Switching API Software Overview.

[4]Broadcom BCM5690 Producu Brief.

收稿日期：2004 – 08 – 02

咨询编号：050316

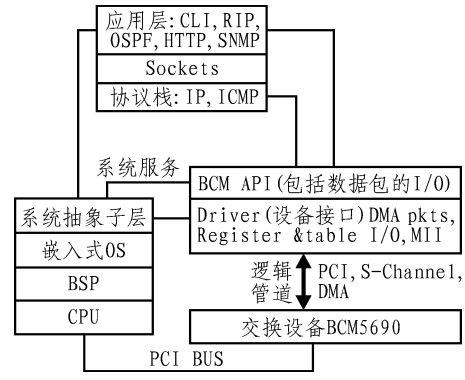

图 5 软件结构框图