

支持动态互连的多处理器系统设计

张 骋, 侯卫民, 蔡惠智

(中国科学院声学研究所, 北京 100080)

摘 要: 针对多处理器系统通信带宽的高需求, 结合 LocalLink 协议和 Batcher-Banyan 交换网络, 提出一种适合在 FPGA 中实现的动态互连网络。与传统总线互连方式相比, 该系统的通信性能有了较大的提高, 仅占用少量硬件资源, 便于集成新的通信接口, 为多处理器系统设计提供了一种高性能且灵活的互连方式。

关键词: LocalLink 协议; Batcher-Banyan 交换网络; 动态互连

Design of Multi-processor System Supporting Dynamic Interconnection

ZHANG Cheng, HOU Wei-min, CAI Hui-zhi

(Institute of Acoustics, Chinese Academy of Sciences, Beijing 100080)

【Abstract】 With the high demand of communication bandwidth for multi-processor system, combining LocalLink protocol and Batcher-Banyan switching network, this paper proposes a dynamic interconnection network which is fit to be realized in FPGA. Compared with traditional interconnection, communication performance is improved great, while only occupying a small quantity of hardware resource. The network is also convenient to integrate with new communication interface, brings forward a high performance and flexible interconnection method for multi-processor system design.

【Key words】 LocalLink protocol; Batcher-Banyan switching network; dynamic interconnection

随着半导体技术的发展, DSP 处理器得到了飞速的发展。但是在许多应用领域, 应用对象日益复杂, 采样速率和处理精度不断提高, 只有多片 DSP 处理器并行处理系统才能满足需求。传统的总线互连方式仅支持独占方式, 总线上的设备越多, 每个设备分享到的带宽越少。而一个高性能系统势必又需要高带宽的交换和路由。本文基于 LocalLink 协议和 Batcher-Banyan 交换网络, 在多处理器系统中引入了一种支持动态互连、无阻塞路由的交换网络, 使系统中的所有处理单元和通信节点间可以真正实现并行通信。

1 多处理器系统硬件结构设计

多处理器系统中的处理单元采用 ADI 公司的 ADSP-TS201S 处理器(以下简称 TS201), 系统互连和交换功能基于一块高性能的 FPGA 芯片, 即 Xilinx 公司的 Virtex-5 LX50T(以下简称 V5) 系统结构如图 1 所示。4 片 TS201 加上一片 V5 构成了一个基本的计算单元, 可以提供 14.4 GFLOPS 的浮点运算能力和大于 7.5 Gb/s 的对外通信带宽。

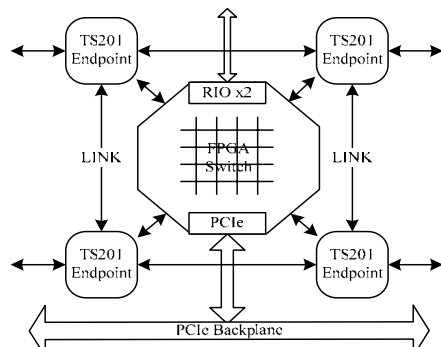


图 1 多处理器系统结构

TS201 是一款 128 位的 DSP 处理器, 最高可以工作在 600 MHz, 可提供 3.6 GFLOPS 的 32 位浮点运算能力。基于它的超标量结构, TS201 在定点运算中也很高效, 每片可以提供 14.4 BOPS 的处理能力。1 024 点的基 2-FFT 运算只需要 15.7 μ s。TS201 在提供领先的系统性能的同时还保持软硬件开发“System-on-Chip”的灵活性: 片内 24 Mb 的 SRAM; 支持 14 个 DMA 通道的 DMA 控制器; 4 个 LVDS 电平的链路口, 每个可提供 2.4 GB/s 的带宽(各 1.2 GB/s 的发送和接收带宽); 还有可以在 TS201 之间实现无缝连接的多处理器接口。TS201 处理器的平衡架构利用 RISC, VLIW 和 DSP 的特性, 提供了一种灵活的“全软件”方法, 能够在降低成本、减少元件用量的同时提升性能。

V5 这款 FPGA 基于创新的 ExpressFabric 架构采用 1.0 V 和 65 nm 三栅极氧化层工艺提升了性能, 降低了功耗。ExpressFabric 技术提供了业界首个真正的 6 输入 LUT 结构, 可以明显提升逻辑性能。新的对角对称布线使逻辑单元 CLB 连线经过的开关点更少, 从而降低了布线延迟。除了提供高性能逻辑功能外, 为满足系统定制的灵活性, V5 集成的诸多硬 IP 方便用户实现高速串行连接(如 PCIe)和信号处理功能。

在互连结构上, 有 3 种不同的通信方式可以满足不同层次的通信需求: (1) 计算单元内部 4 个 TS201 之间既有固定连接的 LINK 通道, 又可以通过 FPGA 提供的动态互连网络通

作者简介: 张 骋(1980 -), 男, 博士研究生, 主研方向: 高速实时信号处理, 嵌入式系统设计; 侯卫民, 博士; 蔡惠智, 研究员、博士生导师

收稿日期: 2008-08-14 **E-mail:** lucas.zhang@yahoo.com.cn

信 ;(2)计算单元之间利用 V5 内置的 PCIe 端点集成了高速的 PCIe 总线,计算单元可以通过背板上的 PCIe 协议进行通信,多个计算单元构成一个机箱;(3)机箱间的通信利用 V5 上的 RocketIO 通道,配合搭载的 Aurora 应用层协议,可以实现机箱间的远距离通信。

2 交换网络设计

影响系统性能的关键部分是 FPGA 交换互连结构的设计,在互连上它必须兼容 TS201 的 LINK 口、PCIe 接口和采用 Aurora 应用层协议的 RIO 收发器;在交换上它要支持 4 个 LINK 口,1 个 PCIe 接口和 2 个 RIO 收发器之间的动态互连特性。因此,需要一种通信协议,它必须满足如下条件:一个基于点对点通信的协议;协议比较简单,以便将其他不同的接口融合进这个交换网络;协议接口信号要少,这样在搭建交换网络时才不会占用很多 FPGA 资源。交换网络的结构如图 2 所示。

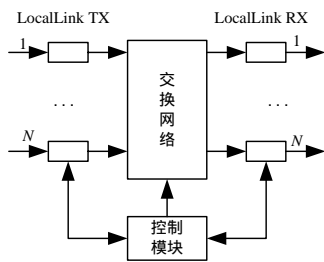


图 2 交换网络结构

基于以上考虑,采用了 Xilinx 公司提出的 LocalLink 接口协议^[1]。LocalLink 是一个同步的点对点互连协议,它支持任意宽度的数据端,可以采用任意长度的帧结构传输数据,同时还提供了上下行的流量控制、奇偶检验错误报告等可以供用户增加载减的功能。

图 3 是 LocalLink 发送端的信号波形图,逻辑关系比较简单,同步信号 TX_SOF_N 和 TX_EOF_N 标记发送帧的帧头和帧尾,握手信号 TX_SRC_RDY_N 和 TX_DST_RDY_N 协调发送和接收方的动作。这种简单的总线协议很适合作为归一化的交换网络接入协议。

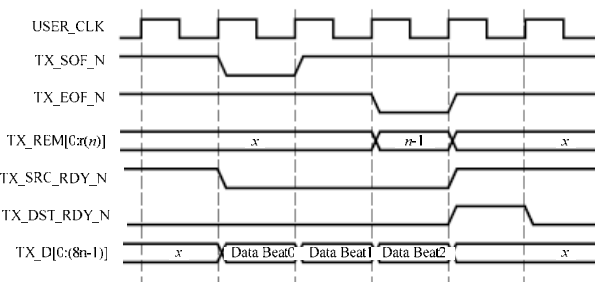


图 3 LocalLink 发送波形

典型的空分交换结构 Crossbar 很高效,而且内部没有阻塞,是一种理想的交换网络选择。但受限于 FPGA 的逻辑结构,只能通过多路选择器来实现,因此,交换延时在端口数增多时也会急剧增加,仅仅适合在 ASIC 中用特殊结构实现。而且考虑到交换网络将来的扩容性,Crossbar 也不是一个好的选择。

Banyan 交换网络由 2×2 无阻塞的基本交换单元(图 4)组成,这种基本交换单元只有直通和交叉 2 种交换方式,易于在 FPGA 中构建。而且 Banyan 交换网络是自路由的,因此,方便逻辑实现控制^[2]。 8×8 的 Banyan 交换网络由 12 个基本

交换单元分 3 级构成(图 5),每个入口都可以将连接导至 8 个出口中的任一个。每个入口提供要建立连接的出口号,用 3 位二进制数表示,Banyan 交换网络从左向右逐位分析 3 位数字,并据此寻径。第 1 级按左边最高位寻径,第 2 级按中间位寻径,第 3 级按右边最低位寻径,从而将入口对应导至出口。例如在图 5 中,入口 0 上收到一个出口标号为 6(110)的连接请求,交换网络即根据出口标号逐位检查。第 1 位为 1,指示第 1 级输出到 1,第 2 位也是 1,指示第 2 级也输出到 1,第 3 位为 0,指示第 3 级输出到 0。于是就建立了到输出口 6 的连接。不同入口可同时建立与不同出口的连接。设在入口 6 上同时收到出口标号为 1(001)的连接请求。交换网络可同时处理,逐位检查逐级处理,将连接导向出口 1。

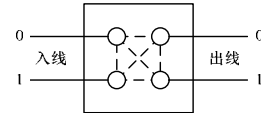


图 4 基本交换单元

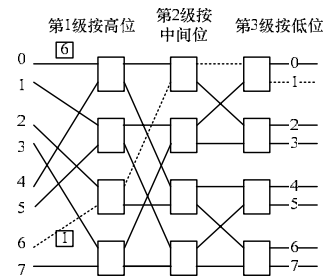


图 5 8×8 的 Banyan 交换网络

但是 Banyan 交换网络存在内部阻塞的缺点:由于 Banyan 交换机可同时接收 8 个输入,几乎肯定会在某一级的基本交换单元中发生冲突,因此会使交换网络陷于瘫痪。

可以通过在交换网络前端加上一个 Batcher 排序网络的方法来解决内部阻塞问题。Batcher-Banyan 交换网络如图 6 所示,Narasimha 证明了 Batcher 网络可以将任意的输入模式排序成二重单调序列,并且该序列通过 Banyan 交换网络后总是以升序或降序输出^[3]。

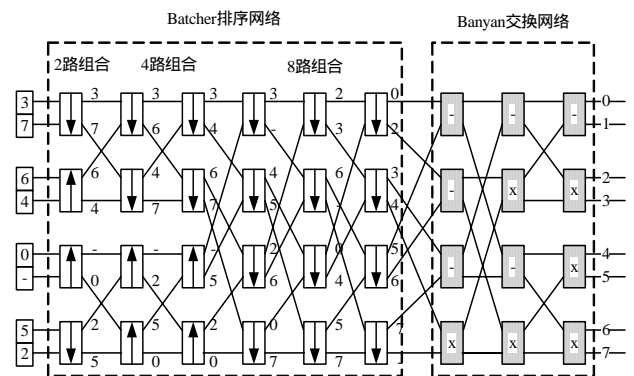


图 6 Batcher-Banyan 排序交换网络

Batcher 排序网络也由多个基本交换单元组成,但其连接方式和工作原理与 Banyan 交换网络不同。当每个基本交换单元收到 2 个请求时,就按其出线标号进行比较然后进行规整。其原则是:较高标号按箭头方向寻径,较低标号按箭头反方向寻径。若只收到 1 个请求,则在 Batcher 交换机中不存在比较操作,只需按箭头反方向寻径^[4]。

(下转第 251 页)