

引用格式:许晋彰,景乃锋,蒋剑飞. 一种面向多处理器互连的高速串行传输系统设计[J]. 微电子学与计算机,2020,37(8):16-20. [XU J Z,JING N F,JIANG J F. Design of high speed serial transmission system for multiprocessor interconnects[J]. Microelectronics & Computer,2020,37(8):16-20.]

一种面向多处理器互连的高速串行传输系统设计

许晋彰,景乃锋,蒋剑飞

(上海交通大学 电子信息与电气工程学院,上海 200240)

摘 要: 为了应对大数据量实时处理,解决处理器间通信带宽瓶颈. 采用自主设计的混合多 FPGA 平台搭建了面向新型 RISC-V 处理器互连的高速串行传输系统,通过对互连接口的改进,实现了单通道 4 倍传输速率的提升. 实验表明,该系统总吞吐率可达 300 Gbps,单通道速率最高可达 25 Gbps,支持的互连链路大于 40 路,传输误码率低于 $1\text{E}-11$,能够满足高带宽、高可靠性传输需求,同时极大地减少了 IO 管脚数量. 最后,基于该系统,在相同速率下对比和评估了 3 种高速串行协议的实现情况,为传输效率提升和系统优化提供帮助.

关键词: 高速串行;多处理器互连;FPGA 设计;片间通信

中图分类号: TP336

文献标识码: A

文章编号: 1000-7180(2020)08-0016-05

Design of high speed serial transmission system for multiprocessor interconnects

XU Jin-zhang,JING Nai-feng,JIANG Jian-fei

(School of Electronic Information and Electrical Engineering,Shanghai Jiao Tong University,Shanghai 200240,China)

Abstract: In order to deal with the real-time processing of large data volume, and to solve the communication bandwidth bottleneck between processors, a high-speed serial transmission system for the interconnection of new RISC-V processors was built by self-developed hybrid multi-FPGA platform. It has achieved a single-channel 4x transmission rate improvement. Experiments show that the total throughput rate of the system can reach 300 Gbps, the single channel rate can reach 25 Gbps, the supported interconnect links are greater than 40, and the transmission error rate is less than $1\text{E}-11$, which can meet most high-bandwidth and high-reliability transmission requirements, greatly reducing the number of IO pins. Finally, based on this system, at the same rate, compare and evaluate the implementation of three high-speed serial protocols, To help improve transmission efficiency and system optimization.

Key words: high speed serial; multiprocessor interconnect; FPGA;inter-chip communication

1 引言

随着雷达侦测、电子对抗等数字通信类先进电子系统的发展,高速、高带宽、抗干扰、大数据高速通信技术将成为未来主流^[1]. 针对大数据量实时处理,系统性能不仅与处理器的性能和数目有关,更受制

于各个处理器之间的通信传输效率^[2]. 传统方式上,多芯片间互连采取并行总线,但这种互连容易受到线间同步和干扰的影响,以及互连线长度、信号摆幅、工作频率和互连端口的限制,很难实现高速的数据传输. 因此,多芯片间目前大多采用高速 SerDes 技术^[3] 加光传输来实现,通常来说,采用 16 nm~40

nm 工艺,单通道速率可达 6.25 Gbps~32.75 Gbps.本文基于自主设计的混合多 FPGA 平台,搭建并实现了一种面向多处理器片间互连的高速串行传输系统,通过对互连接口的改进,完成了单通道 4 倍传输性能的提升.该系统可适用于多种高速串行协议,具备互连结构可配置,传输效率高,信号质量好等特点.

2 系统结构设计

本文所搭建的系统整体结构如图 1 所示,系统硬件平台由 5 块 FPGA 构成,其中 FPGA1~4 为 Titan 系列的 PGT180H_7CFFBG1152,采用 40 nm 工艺,芯片可编程规模为 174 K 等效 4 输入 LUT ,内部集成了 Data Rate 大于 6.25 Gbps 的高速串行接口模块, (High Speed Serial Transceiver, HSST),可支持 12 个全双工收发 Channel. FPGA5 为 XilinxKintexUltraScale + 系列的 XCKU5P-2FFVB676E,采用 FinFET 工艺,具有 475 K 系统逻辑单元,内部集成了 Data Rate 不低于 25 Gbps 的高速串行接口模块 (Gtwizard Transceiver Y, GTY),可支持 16 个全双工收发的 Channel. FPGA1~5 之间和自身通过可拔插光模块即 QSFP 或 SFP,以及外加光纤的方式实现芯片间的互连^[4].

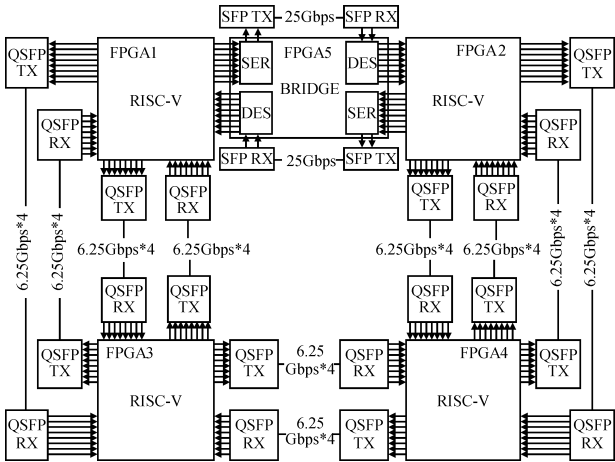


图 1 系统整体结构图

在上述硬件平台上,将由工具链编译生成得到的定制化的新型 RISC-V 处理器进行 RTL 设计和综合,最后下载到 FPGA1~4 中.在完成 RISC-V 处理器与 FPGA 内部高速收发器的互连与集成后,通过运行指令实现数据的加载、高速收发和存储的功能.在此基础上,为实现更高速率的数据传输,本文对系统进一步优化.通过对现有的高速串行接口进行设计和改进,在 FPGA5 上搭建了 2 组四通道合

并成单通道的互连转接电路,大幅度提高了系统的性能.最终整个系统完成 4 个 RISC-V 处理器之间的多种协议可配置、高吞吐率无阻塞通信.

2.1 处理器与高速收发器互连设计

在 FPGA1~4 上进行 RISC-V 与高速收发器互连电路的设计. RISC-V 处理器的设计可通过参数化的配置方法,编译生成可综合的 RTL. 配置生成的 RISC-V 处理器预留了 AXI4 总线接口,通过 AXI4 总线完成 RISC-V 处理器与 HSST 高速收发器的互连. HSST 高速收发器的设计需要考虑不同传输速率,编码方式和高速串行协议的设定.下面对 RISC-V 处理器和 HSST 高速收发器分别展开介绍.

2.1.1 RISC-V 处理器接口设计

为充分发挥高速串行互连所具备的高带宽优势^[5],实现应用性能的提升,本文对处理器的设计做了相应的更改,包括处理核心、高速互连接口等优化.最终完成 RISC-V 处理器与高速收发器的互连接口设计.

此处所采用的 RISC-V 处理器是一款 32 位、5 级流水线、顺序执行的处理器.通过具备强大硬件开发能力的 Chisel 语言^[6]编写,最终配置生成得到.在此基础上进行设计和仿真验证,进而大幅度缩短设计的迭代周期. RISC-V 处理器结构如图 2 所示.

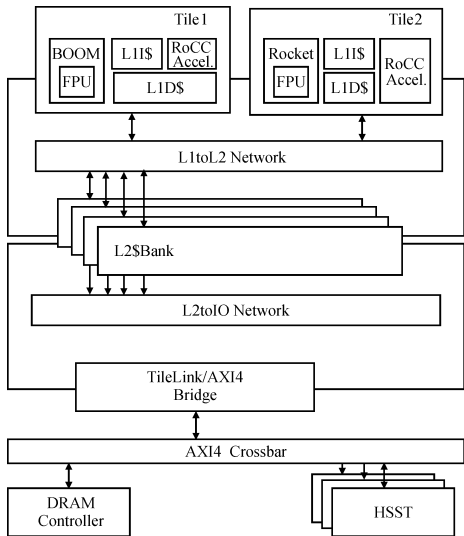


图 2 RISC-V 处理器结构图

该系统可分为三大部分,分别是 Tile,互连网络以及外设. Tile 共有两种,分别包含 BOOM 和 Rocket Core,在每个 Tile 中,除了 L1 Cache 外,还存在一个 RoCC 协处理器. 互连网络主要包含两种总线:TileLink 和 AXI4. 在 Tile 内部,以及 Tile 和

L2 Cache 的互连采用 TileLink 总线;在系统与各外设如 DRAM Controller、HSST 之间的互连采用 AXI4 总线. 在整个设计中,由 3 个 Quad 组成的 HSST 为系统中的一个设备,挂载在 RISC-V 处理器的 AXI4 总线上. 通过对设备进行地址分配^[7],利用 RISC-V 处理器中 Load 和 Store 指令对 HSST 进行访问,将本地 RISC-V 处理器的数据经过高速串行传输发送至相邻 RISC-V 处理器,或者从相邻 RISC-V 处理器获取数据,实现多个 RISC-V 处理器之间高速互连通信.

2.1.2 HSST 高速收发器设计

通常来说,高速收发器包含 PMA 和 PCS 两层^[8],PCS 层主要包括 8B/10B 编码,PRBS 序列产生和校验电路,字节对齐,通道绑定以及时钟补偿电路等. PMA 层主要包含串并和并串转换电路,驱动强度控制电路,均衡器,以及时钟数据恢复电路等,每一层又可再细分为发送部分和接收部分.

在 Titan 系列 FPGA 中,HSST 按照 Quad 分布,每块 FPGA 包含 3 个 HSST Quad,每个 HSST Quad 支持 4 个全双工的收发 Channel,一共可支持 12 个 Channel 高速收发. HSST 高速收发器的结构如图 3 所示.

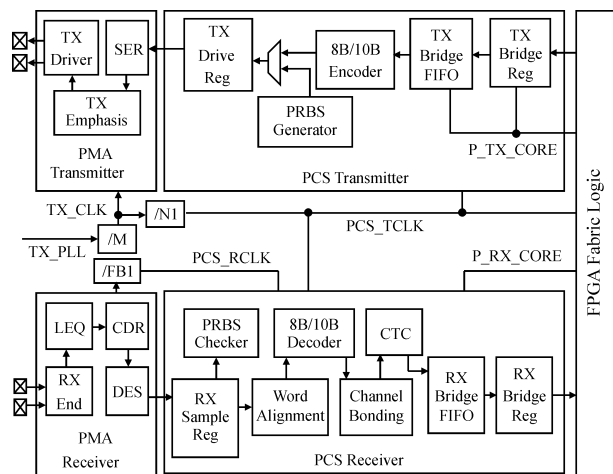


图 3 HSST 高速收发器结构图

整个 HSST 高速收发器由发送通路和接收通路组成. 其中,PCS Transmitter 和 PMA Transmitter 组成发送通路,PCS Receiver 和 PMA Receiver 组成接收通路. 当 HSST 进行数据收发时,所需传输的并行数据从 FPGA Fabric Logic 发出,经过与 PCS Transmitter 的数据桥接,到达 8B/10B Encoder 模块,进行数据的编码,之后通过 TX Drive Reg 到达 PMA Transmitter. 在 PMA Transmitter

经过 Serializer 串化、TX Emphasis 处理后,到达差分数据的发送口,完成数据的发送过程;在接收时,串行数据以差分的形式到达 PMA Receiver,通过内部的 LEQ 均衡器改善信号的信噪比^[9],CDR 完成时钟数据恢复,以及 Deserializer 解串处理后抵达 PCS Receiver. 在 PCS Receiver 主要进行 Word Alignment、8B/10B 解码、Channel Bonding 以及 CTC 时钟补偿等操作,最后数据抵达 FPGA Fabric Logic 完成整个传输过程. 此外,设计中内嵌的 PRBS 测试序列产生和校验电路用于测试传输通道的好坏.

2.2 多芯片间互连接口改进设计

以 Xilinx Kintex UltraScale+ 系列 FPGA 上集成的 GTY 高速收发器为基础,通过 RTL 详细设计,进行多芯片间互连接口改进,实现四通道合并成单通道双路高速收发. 多芯片间互连改进接口结构如图 4 所示. 以用户数据从 FPGA1 发送至 FPGA2 为例,整个设计工作时,从 FPGA1 发出的四路差分串行数据在抵达外部 CHK 端后迅速转化 128 bit 并行数据,并以当前内部高速采样时钟按照 RX Valid 要求写入 RX FIFO 中. RX FIFO 为读写位宽不一致的异步 FIFO,当数据进行改进后高速发送时,通过发送端 TX Ready 信号控制 RX FIFO 的读使能. 数据读出为 64 bit 并行数据,在 TX Support 中将 64 bit 并行数据转化为单通道高速串行数据用于发送. 完成四通道转单通道的发送过程. 发出去的数据经过双 LC 多模 OM3 光纤以差分的形式进行传输^[10]. 接着,数据抵达接收端 RX Support,在此处恢复成 64 bit 并行数据,并以 378.78 MHz 的时钟频率写入 TX FIFO 中,根据 FPGA2 接收端的时序要求,128 bit 有效数据以 156.25 MHz 的时钟频率从 TX FIFO 读出,再拆分成 4 组进行并串转换,最终发送至 FPGA2 完成单通道重新恢复成四通道的高速接收过程. 数据从 FPGA2 发送至 FPGA1 过程同理.

值得注意的是,由于互连接口改进前后实际数据率以及编码效率的不同,需要额外增加 IDLE Data 以及 Data Filter 电路,从而保证整个传输过程中数据的正确性.

3 实验结果分析

使用 Pango Design Suit2018.1B2 对 FPGA1~4 进行设计,使用 Vivado2017.4 对 FPGA5 进行设计. 基于整个混合多 FPGA 硬件平台,进行实验结果分析,包括高速串行系统各部分逻辑资源消耗分

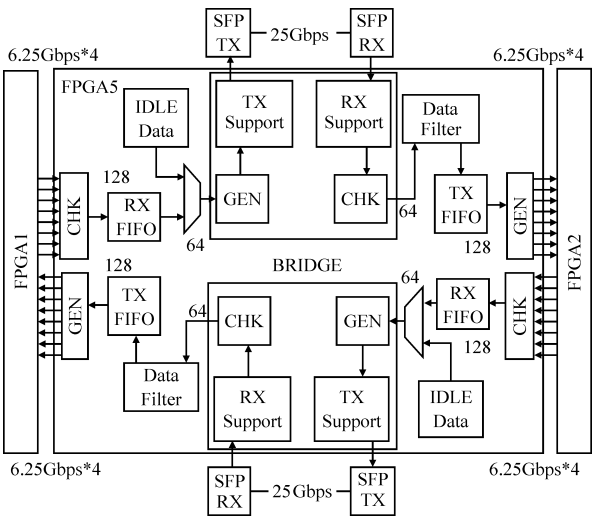


图 4 多芯片间互连接口改进结构图

析,系统测试及性能对比,以及不同高速串行协议的实现评估。

3.1 系统逻辑资源消耗情况

在 FPGA1~4 中,将 RISC-V 处理器与 HSST 通过 AXI4 总线互连,生成 Bit 流,下载至 FPGA 芯片中.运行时,RISC-V 处理器通过系统内部的带宽映射与分配策略,实现与高速串行收发器的互连通信. RISC-V 与 HSST 互连资源占用情况如表 1 所示.

表 1 RISC-V 与 HSST 互连资源占用情况

资源	已使用	可用	占比%
CLMA	12 616	27 131	47
CLMS	4 399	9 123	48
HSST	3	3	100
IO	21	611	3
DRM	87	526	17

其中, CLMA 和 CLMS 均为 Titan 系列 PGT180H_7CFFBG1152 的基本逻辑单元,主要由多功能 LUT5,寄存器等组成. 本文的设计中,基本逻辑单元的占比可等效为 CLMA 和 CLMS 二者的均值,为 47.5%,接近系统总逻辑资源的一半. HSST 以 Quad 为单位,3 个 Quad 共计 12 个 HSST,本文已将 FPGA1~4 的 HSST Quad 全部用于数据的收发,故 HSST 的占用率为 100%. 对于 IO 管脚资源占用情况,相比于并行方式传输下 FPGA 的管脚数为 $(32 \times 4 \times 3 \times 2 =)768$,已超出 FPGA 的 IO 管脚总数,本文所采用的高速串行传输技术可以节省大量的 IO 管脚资源.

对 FPGA5 进行开发,完成对多芯片间高速通信接口的改进设计,实现更高效率的数据传输. 多芯

片接口改进资源占用情况如表 2 所示.

表 2 多芯片接口改进资源占用情况

资源	已使用	可用	占比%
LUT	11 762	216 960	5.42
FF	22 798	433 920	5.25
IO	16	280	5.71
GTY	10	16	62.5

由表 2 可知,系统逻辑单元占比为 5.42%,即使用了少量的逻辑资源,完成了改进的双通道高速接口设计,实现了传输性能的提升. FPGA5 一共有 16 个 GTY,所用到的 10 个 GTY 中,4 个用于与 FPGA1 的四通道互连,4 个用于与 FPGA2 的四通道互连,还有 2 个用于双通道改进后的高速收发互连,GTY 占比 62.5%.

3.2 系统测试及性能对比情况

针对系统高速收发延迟测试,采用基于序列数据传输的动态测算方法,统计发送端和接收端二者之间的序列差值,将该差值带入公式(1),即可得到实际的延迟.

$$Latency = \frac{x_2 - x_1}{f}$$

(1)

式中, x_2 表示当前时刻发送的数; x_1 表示当前时刻接收的数;二者差值 $x_2 - x_1$ 即表示间隔的 Cycle; f 表示当前传输的时钟频率.

针对单通道传输速率的测试,计算方法如公式(2)所示. 其中, f 为示波器所采集的数据收发的时钟频率, $width$ 为数据位宽, eff 为编码传输效率,由此可得单通道的传输速率.

$$v = \frac{f \times width}{eff}$$

(2)

根据上述方法,对改进前后的两种互连通道的传输性能进行评估,得到的单通道性能提升对比情况如表 3 所示.

表 3 单通道性能提升对比情况

指标	改进前	改进后	提升比例
时钟频率	156.25 MHz	378.78 MHz	2.424
位宽	32 bit	64 bit	2
编码方式	8B/10B	64B/66B	—
开销	20%	3.03%	0.15
数据包	1 KB	0.5 KB	0.5
延迟	77 ns	58 ns	0.82
传输效率	80%	96.97%	1.212
接口速率	6.25 Gbps	25 Gbps	4

由表 3 可知,改进后相比于改进前,单通道传输速率由原来的 6.25 Gbps 提升到 25 Gbps,通过消

耗少量的逻辑资源,实现了 4 倍传输速率的提升.其中,时钟频率为原来的 2.424 倍,数据位宽为原来的 2 倍.改进前所采用的编码为 8B/10B,改进后所采用的编码为 64B/66B,二者的传输效率分别为 80%和 99.47%,改进后电路具备更高的传输效率.同时,改进后的单通道 25 Gbps 传输具备更低的传输延迟,相比于改进前的 6.25 Gbps 传输,改进后的电路采样时钟频率更快,传输性能更高.

系统整体性能测试结果如表 4 所示.其中,采用基于 PRBS 原理的内置误码仪对各个通道进行误码率测试和统计,结果表明:6.25 Gbps 通道和 25 Gbps 通道的误码率处于同一数量级,均低于 1E-11,属于理想传输状况.对于链路数,整个系统互连链路总数达 42 路,其中包括 40 路 6.25 Gbps 和 2 路 25 Gbps 的高速传输,系统总体吞吐量为原始部分和改进部分之和,达 300 Gbps,符合高带宽指标要求.

表 4 系统整体性能测试结果

指标	原始部分	改进部分
单路速率(Gbps)	6.25	25
链路数	40	2
误码率	9.936E-12	2.984E-12
吞吐量(Gbps)	250	50

3.3 不同高速串行协议对比情况

基于本文搭建的高速串行传输系统,评估了在同一传输速率 6.25 Gbps 下三种典型的串行协议的具体实现,包括 Aurora、Serial RapidIO 以及 Rawdata 自定义协议.三种高速串行协议的实现对比情况如表 5 所示.

表 5 三种高速串行协议实现对比情况

类别	Aurora	Serial RapidIO	Rawdata
LUT	11 762	11 974	10 881
FF	22 798	23 739	22 586
GTY	10	10	10
延迟(ns)	77	79	39
功耗(W)	4.56	4.61	4.50

实验表明,在资源消耗方面,Serial RapidIO 的基本逻辑资源占用量高于其他两种,原因在于 Serial RapidIO 相比于 Aurora 和 Rawdata 在层级定义、数据包格式和传输机制上更加完善,因而开销更大.在延迟方面,Aurora 和 Serial RapidIO 的延迟相近,且均大于 Rawdata.由于 Aurora 涉及到内部通道绑定及同步等操作,而 Serial RapidIO 为基于请求响应的包交换协议,为保证传输可靠性,延迟会有所增加.在功耗方面,三种协议大致相同,主要原因在于三种协议所用到的高速收发器 GTY 数目一致,而高速收发器所产生的功耗占了总体功耗的很

大一部分,因此三者的功耗差别不大.

4 结束语

本文基于自主设计的混合多 FPGA 硬件平台,开发了一种面向多处理器互连的高速串行传输系统.同时对互连接口进行改进,实现了单通道 4 倍传输速率的提升.整个系统支持 40 路 6.25 Gbps 和 2 路 25 Gbps 同时收发,最终实现了 4 个处理器之间 300 Gbps 的高速互连传输.此外,基于该系统,进行 3 种典型高速串行协议的实现,评估了实际应用的性能.该系统规模大,设计方法新,可靠性好,可为高性能计算和高带宽传输提供支持.

参考文献:

[1] JADHAV A K, KUMAR V S, HARLALKA K. Optimization of the performance of high speed serial communication[C]// IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI). Chennai, India, 2017:1693-1697.

[2] HUANG X, SU H, WU Q, et al. Multi-Processor parallel system based on high-speed serial transceiver [C]// 2010 Second International Workshop on Education Technology and Computer Science. Wuhan, China, IEEE, 2010:178-181.

[3] BAUMER H A. Multi-rate, multi-port, gigabit SERDES transceiver; U. S. Patent 9,330,043[P]. 2016-5-3.

[4] XIAO C H, HUANG Z Q, LI D. Multi-FPGA interconnection structure and partition method for high-performance computing[J]. Computer Application Research, 2015,32(1):150-155.

[5] JIN Y Y. Design and application of high-speed data interconnection module based on FPGA[D]. Chengdu: University of Electronic Science and Technology, 2018.

[6] LI D Z, GONG H R, CHANG Y C. Implementing RISC-V System-on-Chip for Acceleration of Convolution Operation and Activation Function Based on FPGA[C]// 14th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT). Qingdao, China, 2018:1-3.

[7] OLIVIERI M, CHEIKH A, CERUTTI G, et al. Investigation on the optimal pipeline organization in RISC-V multi-threaded soft processor cores[C]// New Generation of CAS (NGCAS). Genova, Italy, IEEE, 2017:45-48.