

芯粒集成系统封装 I/O 高速总线架构设计及实现*

张转转^{1,2,3} 缪旻^{1,2,3**} 朱仕梁^{1,2,3} 段晓龙^{1,2,3}

(¹ 北京信息科技大学 信息与通信系统信息产业部重点实验室, 北京 100101) (² 北京信息科技大学 光电测试技术及仪器教育部重点实验室, 北京 100192) (³ 北京信息科技大学 智能芯片与网络研究中心, 北京 100101)

2023-07-24 收稿, 2023-11-24 收改稿

摘要:随着集成密度和单片处理速度的不断提升,芯粒集成系统封装(Chiplet SiP)中互连网络日趋复杂且信号与电源完整性、传输能耗问题日趋严重,芯粒与 SiP 外部的数据交换 I/O(Input/Output)容量的提升遭遇瓶颈。为提升芯粒集成度、提高数据传输速率与准确率、降低系统功耗,根据芯粒间通信的最新标准通用芯粒互连技术(Universal chiplet interconnect express, UCIe),利用高速串行计算机扩展总线标准(Peripheral component interconnect express, PCIe)在高速数据存储及传输方面的技术优势,设计出一种芯粒高速 I/O 通信的架构,并用 FPGA 验证了此架构的可行性,为 UCIe 标准的落地提供了一种实现途径。

关键词:PCIe; 系统芯片; 芯粒; 系统封装; 高速总线

中图分类号:TN43 文献标识码:A 文章编号:1000-3819(2024)01-0045-05

Design and Implementation of a Chiplet Integrated System-in-package Based on I/O High-speed Bus Architecture

ZHANG Zhuanzhuan^{1,2,3} MIAO Min^{1,2,3} ZHU Shiliang^{1,2,3} Duan Xiaolong^{1,2,3}

(¹ Key Laboratory of Information and Communication Systems, Ministry of Information Industry, Beijing Information Science and Technology University, Beijing, 100101, CHN) (² Key Laboratory of the Ministry of Education for Optoelectronic Measurement Technology and Instrument, Beijing Information Science and Technology University, Beijing, 100192, CHN) (³ Academy of Smart IC and Network (ASICNet), Beijing Information Science and Technology University, Beijing, 100101, CHN)

Abstract: With increasing integration density and chip processing speeds, the interconnect network among chiplet SiP has become increasingly complex, and the difficulty of wiring has sharply risen, resulting in a growing concern over signal and power integrity as well as transmission energy consumption. As a consequence, the data exchange I/O (Input/Output) capacity between the outside of the SiP and chiplet is constrained. In order to improve chiplet integration, data transmission rate and accuracy, and reduce system power consumption, a chiplet high-speed I/O communication architecture is designed. The architecture is based on the latest standard UCIe (Universal chip interconnect express) for inter chip communication. The design takes advantage of PCIe (Peripheral component interconnect express) bus in high-speed data storage and transmission, an architecture for implementing high-speed I/O communication between Chips is designed. Its feasibility is verified by using FPGA resources, which provide an approach for the physical implementation of the UCIe standard.

Key words: PCIe; SoC; chiplet; SiP; high-speed bus

* 基金项目:国家自然科学基金资助项目(62074017)

** 联系作者: E-mail:miaomin@bistu.edu.cn

引言

21 世纪第二个十年兴起的多芯片异构集成被广泛视为实现拓展摩尔定律 (More than Moore) 的主要载体之一^[1], 系统封装技术 (System in Package, SiP) 则被视为多芯片异构集成的一种主要实现形式^[2], 其 2.5D 的实现方案如图 1 所示^[3-5]。作为最新一代的 SiP 技术形态, 芯粒集成 SiP 是以功能、接口设计标准化的小芯片 (即芯粒) 为功能单元, 借助立体微组装技术进行系统集成的多芯片组件技术^[6], 其核心技术优势在于 IC 芯片设计的高度复用性和组合的灵活性^[7]。但随着集成密度和单片处理速度的不断提升, 当前芯粒集成系统封装 (Chiplet SiP) 中互连网络日趋复杂且信号与电源完整性、传输能耗问题日趋严重。此外, 目前高速数据交换方面尚缺乏统一标准和广泛接受的标准落地方式, 也限制了芯粒集成 SiP 技术的发展与应用。于是, 芯粒 SiP 对外的数据通信架构成为业界和学术界近几年的研究重点。

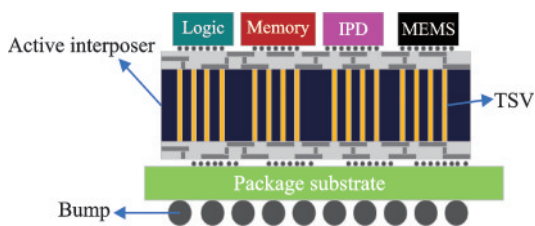


图 1 SiP 结构图

Fig.1 SiP structure

2022 年由 Intel、AMD、ARM 等公司联合推出了 Die-to-die 互连标准——通用芯粒技术 (Universal chiplet interconnect, UCIe), 这是一种开放的行业标准互连。由于其可以根据特定产品集成不同数量的计算单元、内存、I/O 以及加速器等小芯片, 无需根据不同产品对芯片进行全新设计, 从而达到了降低设计、验证和产品成本的目的, 并有望为异构芯片提供高带宽、低延时、高能效的封装互连^[8]。

UCIe 协议可以利用现有的高速串行计算机扩展总线标准 (Peripheral component interconnect express, PCIe) 和 CXL (Compute express link) 协议作为底层通信协议^[9]。基于此, 本文提出了一种基于 PCIe 高速串行总线的芯粒 I/O 设计, 可以作为 UCIe 标准的其中一种跨 SiP 通信方式。这种方法可以提高多芯片架构的可扩展性以及设计的可重用性。借助 Xilinx 公司的 PCIe、AXI (Advanced xtensible

interface)、直接存储器存取 (Direct memory access, DMA)、双倍速率同步动态随机存储器 (Double data rate, DDR) IP 核和其他的 FPGA 片上资源实现了方案并验证了其可行性。

1 芯粒集成 SiP 间数据通信整体架构

本文重点介绍一种以高速总线 PCIe 实现芯片间通信的架构方案, 图 2 是实现芯粒集成 SiP 间通信的整体架构图, 其中芯粒都组装在硅转接板上。

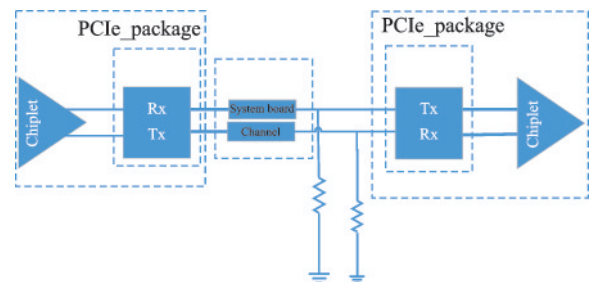


图 2 系统整体架构

Fig.2 Overall system architecture

目前主流的 PCIe SerDes 结构如图 3 所示。主要包括发送端 (Tx)、接收端 (Rx)、锁相环 (Phase locked loop, PLL)、时钟数据恢复 (Clock and data recovery, CDR) 四部分^[10]。其中发送端和接收端进行信号的传输, 锁相环用于产生发送端的时钟, 时钟数据恢复用于接收接收端的时钟。发送端使用前向反馈均衡 (Feed forward equalization, FFE) 校正发送端电压电平, 接收端通过连续时间线性均衡器 (Continuous time linear equalizer, CTLE) 进行一定程度的信道补偿, 最后通过判决均衡反馈器 (Decision feedback equalizer, DFE) 消除由有损信道引起的符号间干扰^[11] (Inter-symbol interference, ISI)。数据发送时, 首先通过并转串 (Serializer) 模块将发送端数据由并行转为串行, 其编码方式为 128b/130b 编码格式, 数据发送至传输信道前采用 FFE 均衡方式; 接收端接收到信号后通过 CTLE 和 DFE 对接收到的信号进行均衡, 最后通过串转并 (Deserializer) 模块将信号转换为并行信号^[12]。

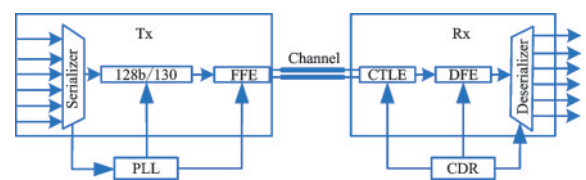


图 3 PCIe 物理层

Fig.3 PCIe physical layer

2 系统架构设计

2.1 基于 MATLAB 的系统架构设计

针对基于 MATLAB™ 开发的一种符合 PCIe3.0 规范的测试仿真模型开展研究。其中传输特性的分析基于 IBIS-AMI 仿真模块,发送端、接收端由对应的 IBIS-AMI 模型等效,最终搭建的模型架构如图 4 所示。将 PCIe 的 IBIS-AMI 模型导入 Simulink 后,所得模型架构如图 5 所示。

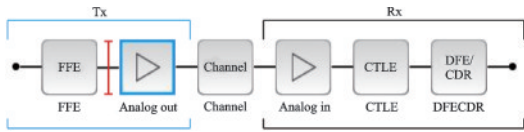


图 4 基于 IBIS 模型的 PCIe 系统架构

Fig.4 Design of PCIe system architecture based on IBIS model

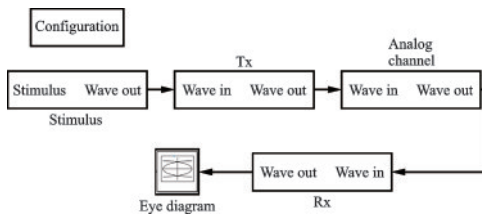


图 5 基于 Simulink 的 PCIe 模型

Fig.5 PCIe model based on Simulink

发射器使用 3 抽头 (Tap) FFE 均衡方式,接收器模型使用具有 7 个预定义设置的 CTLE 和 2 抽头 DFE。为支持此配置, SerDes 系统设置如表 1 所示。

表 1 发射端参数设置

Tab.1 Parameter settings at the transmitter end

Parameter	Value
Symbol time	62.5 ps
Target BER	1×10^{-12}
Anolog out voltage	1.05 V
Rise time	12 ps
Output resistance	50 Ω
Capacitance	0.5 pF
Channel loss	15 dB
Target frequency	8 MHz
Defferential impedance	100 Ω

2.2 基于 Xilinx FPGA 的硬件实现

为构建原型系统,以个人计算机(Personal com-

puter, PC)端作为上位机模拟其中一个芯粒组件,以 FPGA 模拟另一个芯粒组件。根据 PCIe 架构 PIPE 规范可以得出 PC 端的 PCIe 物理层架构,如图 6 所示^[13]。PCIe 物理层架构由物理媒体附着层(Physical media attachment layer, PMA)、物理编码子层(Physical coding sublayer layer, PCS)和媒体访问层(Media access layer, MAC)三部分组成。MAC 层由链路训练状态机(Link training status state machine, LTSSM)、发射机和接收机组成。作为 PCIe 物理层的主控制器的链路训练状态机生成控制信号,对接收机和发射机进行链路训练。发射机具有多路选择器,可以通过它判断下一步传输数据将发送至事务层(Transaction layer package, TLP)、数据链路层(Data link layer package, DLLP)还是物理层(Physical layer package, PLP)。PCS 和 PMA 对从 MAC 层接收到的数据进行编码并通过 LINK 发送至下一 PCIe 接口^[14]。

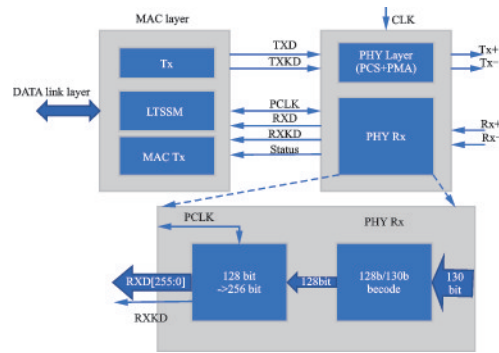


图 6 PC-PCIe 物理层接口

Fig.6 Physical Layer of PC-PCIe

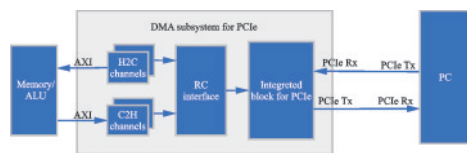


图 7 硬件平台

Fig.7 Hardware Platform

采用 Xilinx 的知识产权 (Intellectual property, IP) 构建如图 7 所示的系统硬件平台。存储器 (Memory) 是采用 FPGA 中的 DDR IP 核, DDR 控制器通过 AXI 接口与 DMA 连接, 实现 DDR 数据的读写。PC 通过设置 PCIe 硬核和与 DMA 连接, 最终可以实现存储器和 PC 进行通信。本文使用的 DMA 是 Xilinx 官方提供的, 在分散收集 (Scatter gather, SG) 模型下, 主机将传输的数据组成链表形式, 然后依次将链表的首地址发送给 XDMA, XDMA 会根据链表结构完成数据传输任务。

系统采用 100 MHz 时钟, FPGA 选用 UltraScale 系列芯片, CPU 为 XCKU040-FFVA1156-2-I。PL 端搭载 64 bit 2GB DDR4 内存。PCIe 接口链路设置为 8, AXI 数据总线的位宽为 256 bit, AXI 参考时钟为 250 MHz, 最快的链路传输速度为 8.0 Gbps。板卡 DDR4 采用 MT40A256 芯片, 由于其时钟频率为 1200 MHz, 所以 DDR 控制器 MIG 接口速度设置为 833 ps, PS 参考时钟为 100 MHz。

基于 DMA IP 核的高速数据传输系统, 完成了 DMA 硬核用户时序接口和传输逻辑的设计, 通过驱动程序和板卡的联调, 在硬件开发平台上完成对基于 XDMA 的 PCIe 高速传输设计的测试验证。

3 基于 MATLAB 仿真结果

图 8 为基于 IBIS-AMI 搭建无均衡方式的仿真眼图。图 9 为将 Rx FFE(接收端前向反馈均衡)配置选择参数值(ConfigSelect parameter, 控制发送均衡参数调节)从 0 更改为 6 的 Matlab 仿真结果图; 图 10 为将 Tx CTLE tap weights(发送端连续时间线性均衡器中滤波器抽头权重)的值从 $[0 \ 1 \ 0]$ 更改

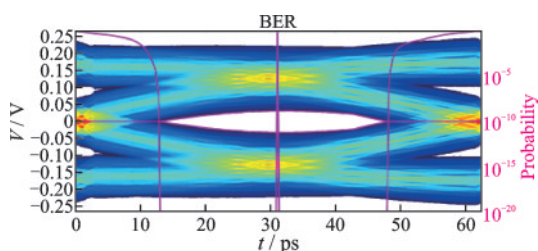


图 8 均衡前眼图

Fig.8 Eye map before equalize

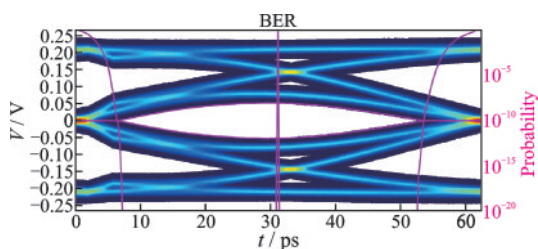


图 9 FFE 均衡后的眼图

Fig.9 Eye map after FFE equalize

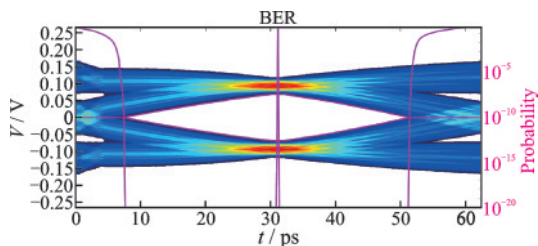


图 10 CTLE 均衡后的眼图

Fig.10 Eye map after CTLE equalize

为 $[-0.125 \ 0.750 \ -0.125]$ 后的眼图仿真结果; 图 11 为将 Rx DFE 模式更改为 Adapt(自适应模式, 使 DFE 达到最优值) 时眼图仿真结果。

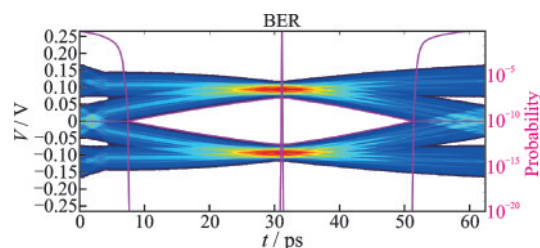


图 11 DFE 均衡后的眼图

Fig.11 Eye map after DFE equalize

通过上述仿真结果可以得出均衡方式对信号传输影响较大, 对信号质量的传输均有正向的帮助。

4 基于 FPGA 测试结果

为了测试和验证 PCIe 传输架构的传输功能, 需要软硬件协同处理。使用 FPGA 作为硬件板卡和 PC 主机, 上机位为 Visual Studio 配合 QT 软件, Visual Studio 用来开发 PCIe 驱动程序。QT 软件用来对传输的数据进行处理并控制数据传输位置。

首先采用回环测试分析 DMA 读写数据的完整性, 通过对 IP 核的参数配置, 让 AXI 4 总线突发一次数据传输, 数据的长度为 16, 数据总线宽度为 256 bit, 所以每次传输数据为 512 byte。

使用 FPGA 开发板的 PCIe 金手指与电脑主机 PCIe 口相插, 如图 12 所示, 图中用红框标注的为 FPGA 板卡。先将比特流文件烧写进 FPGA 板卡然后重启电脑, 主机检测到 PCIe 设备然后进行下一步的工作。集成逻辑分析仪采用 Vivado 的在线逻辑分析仪。基于集成逻辑分析仪(Integrated logic analyzer, ILA) 捕获的信息来验证系统功能。

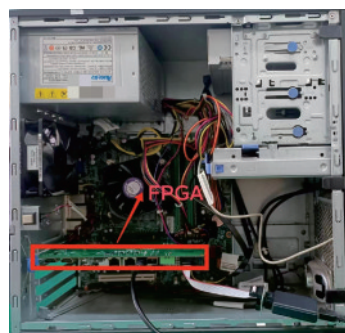


图 12 测试平台实物图

Fig.12 Physical diagram of testing platform

从 PC 端将 10 240 个数据写入 FPGA DDR 存储器中, 通过统计传输正确的数据个数与传输数据

总数,可测试数据传输准确率。

图 13 是 ILA IP 核通过 FPGA 捕获到的传输数据信号。经验证,捕获到的数据与上机位写入的数据一致,此系统可以实现对数据的处理及数据的交换。为了测试该系统数据传输准确率,本实验通过测试五组数据,每次传输的数据个数为 10 240 个,通过 QT 控件,可获得每次传输正确的数据与错误的的数据个数,从而获取数据传输准确率。图 14 为一次数据传输准确率的测试结果,其中 error_cnt 表示接收到的错误数据个数,buf1 表示传输的数据个数,可得数据传输准确率为 100%。

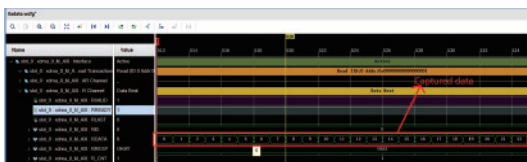


图 13 ILA 捕获的波形

Fig.13 Waveform captured by ILA

名称	值	类型
> buf1	unsigned char [10240]*	unsigned char [10240]
▼ buf2	"007010"	unsigned char *
"buf2"	7	0x07 unsigned char
error_cnt	0	unsigned int
i	7	unsigned int
▼ this	0xf9f830	MainWindow *
> QMainWindow	class QMainWindow	QMainWindow
> staticMetaObject	struct QMetaObject	QMetaObject
> ui	0x1484080	Ui::MainWindow *

图 14 数据传输结果

Fig.14 Transfer results of data

通过 QT 软件开启两个定时器,这两个定时器分别运用 H2C (Host to card) 和 C2H (Card to host) 通道进行设置,每 100 ms 进行一次读写操作。

图 15 是通过 QT 软件获取的数据读写速率,测试结果表明,在 4 kB 的数据包传输中,PCIe 写带宽可达 1 336 MBps,读带宽可达 1 250 MBps。文献 [15] 基于 ZYNQ 平台的 XDMA IP 核性能研究中有关于 PCIe 和 XDMA 测试速率最大为 670 MBps,与相比本文的最大传输速率大致提升了 2 倍。



图 15 FPGA 读写数据速率(F 指 FPGA, P 指 PC)

Fig.15 Read/write data rate of FPGA (F is FPGA, P is PC)

图 16 为 Vivado 综合后的功耗分布图,系统总功耗约为 5.353 W。具体来说,硬核 (Hard IP) 占

6% (0.298 W), 高速收发器 (Gigabyte transceiver H, GTH) 占 41% (2.215 W), 设备静态功耗 (Device static) 占总功耗的 11% (0.572 W), 动态功耗 (Dynamic) 模块占总功耗的 42% (2.268 W)。动态功耗由 Clocks、Signal、Logic、BRAM、I/O 等构成,其中, BRAM 消耗了动态功耗的将近 17% (0.378 W), Clocks 约占动态功耗的 18% (0.398 W), Single 约是动态功耗的 6% (0.137 W)。此外,实际 I/O 消耗的大约为 0.843 W, 占动态功耗的 36%, 占总功耗约为 16%, 只占据了整体功耗的一小部分。

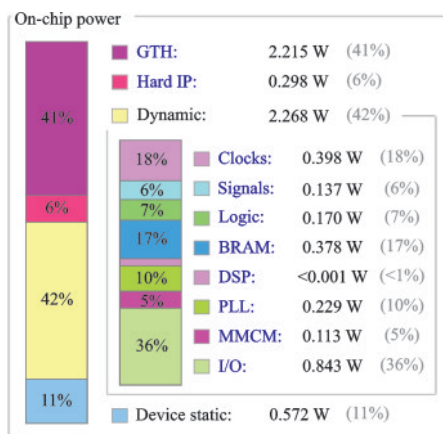


图 16 功耗分布图

Fig.16 Power consumption distribution diagram

5 结 论

本文提出了一种基于 PCIe 接口的芯粒集成 SiP 的高速数据 I/O 传输系统,使用 PCIe 高速串行总线来传输数据,并且该方法适用于传统封装,开发成本较低。该方案实现了高采样率、高传输率、实时存储、高可靠性等性能和低功耗,并具有良好的可拓展性。此外,该方案无需大的设计修改即可灵活地移植到不同系列的 FPGA,同时为高性能计算板卡上的数据存储和信号传输等应用提供参考。

参 考 文 献

- [1] 缪旻,金玉丰.微系统集成全新阶段——IC 芯片与电子集成封装的融合发展[J].微电子学与计算机,2021,38(1):1-6.
- [2] Green D S, Dohrman C L, Demmin J, et al. Path to 3D heterogeneous integration[C].2015 International 3D Systems Integration Conference (3DIC). Sendai, Japan: IEEE, 2015, (7):1-3.
- [3] Tai K L. System-in-package (SIP) challenges and opportunities[C].Proceedings of the 2000 Asia and South Pacific Design Automation Conference. Yokohama, Japan: IEEE, 2000: 191-196.

(下转第 58 页)