

基于软件定义的多协议控制器架构设计

夏云飞¹, 张丽², 杨堃¹, 李沛杰², 许立明¹

(1. 天津市滨海新区信息技术创新中心, 天津 300457;

2. 国家数字交换系统工程技术研究中心, 河南 郑州 450002)

摘要: 为了满足高性能计算平台对软件定义互联的通信需求, 设计了一种支持多种通信协议规范, 且支持异构协议互联的软件定义多协议控制器架构。通过嵌入可编程硬件电路实现 Fiber Channel、Ethernet 以及 SRIO 这 3 种协议控制器电路公共算粒和私有算粒的互联。多协议控制器可以通过软件定义切换为不同协议, 从而满足系统应用中的异构需求。所设计的软件定义多协议控制器基于 40 nm 工艺实现测试样片流片, 经测试, 满足 3 种协议的性能与功能指标要求。

关键词: 多协议兼容; 软件定义; 电路可重构; Fiber Channel; Ethernet; SRIO

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2020065

Architecture design of multi-protocol controller based on software definition

XIA Yunfei¹, ZHANG Li², YANG Kun¹, LI Peijie², XU Liming¹

1. Information Technology Innovation Center of Tianjin Binhai New Area, Tianjin 300457, China

2. China National Digital Switching System Engineering & Technological R & D Center, Zhengzhou 450002, China

Abstract: In order to fulfill the requirement of high-performance computing system for software definition interconnected communication, a kind of multi-protocol controller architecture was designed. The architecture supported a variety of protocol specifications and software-defining heterogeneous architecture. The interconnections of Fiber Channel, Ethernet and SRIO were realized by embedding programmable hardware circuits. The work mode of multi-protocol controllers was managed by software definition, which meet the heterogeneous requirements of system applications. The multi-protocol controller of the thesis has been realized in the 40 nm process, that the measurements show that the performance and function satisfy the standard of the three protocols.

Key words: multi-protocol compatibility, software define, reorganized circuit, Fiber Channel, Ethernet, SRIO

收稿日期: 2019-12-06; 修回日期: 2020-03-19

通信作者: 张丽, yezizhang2003@126.com

基金项目: 国家科技重大专项 (2016ZX01012101)

Foundation Item: The National Science and Technology Major Project (2016ZX01012101)

论文引用格式: 夏云飞, 张丽, 杨堃, 等. 基于软件定义的多协议控制器架构设计[J]. 网络与信息安全学报, 2020, 6(5): 139-147.

XIA Y F, ZHANG L, YANG K, et al. Architecture design of multi-protocol controller based on software definition[J]. Chinese Journal of Network and Information Security, 2020, 6(5): 139-147.

1 引言

随着新型网络技术的发展,网络应用已经从早期非实时业务,发展到目前以文本、图像、语音、视频综合内容承载为主,并呈现出终端类型多元化、接入方式多元化、设备形态多元化、业务场景多元化等趋势^[1]。基于开放架构的软件定义网络(SDN, software defined network)^[2-3],更加强调系统的多标准、多协议、高鲁棒、高效能等特性。面对以上需求,现有网络中的交换节点接口标准单一固定,难以满足接入方式与设备形态多元化的趋势^[4]。通过调研,国内外还没有能满足多种高速串行通信协议的控制器的,无论是国内领先的天河二号超算中心,还是国外领先的 Intel 极光数据中心,内部都是通过集成 PCI-E 或光纤等单一协议接口来支撑数据通信网络,因此一个支持多种接口规范的交换器件十分必要。

现有数据中心中被广泛使用的高速协议为:以太网 Ethernet 10GBASE-KR(简称 Ethernet)、Serial RapidIO(简称 SRIO)以及 Fiber Channel(简称 FC)。Ethernet 协议适用于数据中心间的远距离传输与网络拓扑,FC 协议大多作为数据中心内部的高速存储访问接口,而 SRIO 协议则大多作为嵌入式系统中拓展高性能 DSP 或 GPU 所使用的协议接口,3 种协议的应用场景、传输方式以及兼容频点各不相同。而数据中心正以多协议接口设备作为多形态组网的解决方案,但造成系统体积与能耗骤升^[5]。随着高性能计算平台对系统架构可重构需求的增加,单纯固化集成多种单一协议的方案不利于构建高性能可重构互联网络^[6]。

本文对上述 3 种高性能串行协议功能进行了分析与归一化整合,并利用软件定义与可编程硬件技术实现了一种基于软件定义的高性能多协议控制器架构,最终利用 40 nm 工艺实现支持多协议 SerDes 电路样片配合可编程电路构建了验证系统。

2 多协议控制器架构说明

Ethernet、SRIO 以及 FC 这 3 种高速串行协

议控制器的基本结构,均包含 3 部分电路:物理传输子层(PMA, physical media access)电路、物理编码子层(PCS, physical coding sub-layer)电路以及介质访问控制器(MAC, media access control)电路,其结构如图 1 所示。

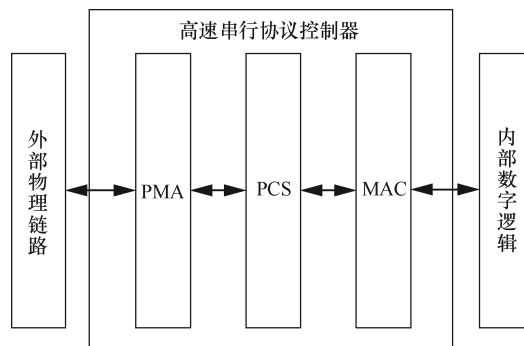


图 1 高速串行协议控制器结构

Figure 1 The architecture of high-speed serial protocol controller

PMA 电路主要完成数据的串/并转换,进而实现高速物理传输^[7],本文所设计的多协议控制器中的 PMA 电路根据 3 种协议的物理规范、通道绑定与频点等需求,采用软件定义互联思想设计多协议高速串行收发器电路。

PCS 电路主要完成数据编解码、通道绑定、位宽转换等功能。本文提取 3 种协议在 PCS 子层功能实现中的公共算粒,并将各个协议私有算粒归一为可配置逻辑,最终实现多协议 PCS 子层电路。

各协议 MAC 电路主要负责数据包流量控制功能,不会对数据包进行帧解析。而 3 种协议在流量控制与校验重传处理上差异较大,私有算粒比重高,无法通过复用重构进行归一化。因此,对 MAC 硬件可重构电路,通过动态加载不同的电路结构,实现多协议异构。

多协议控制器架构如图 2 所示,其基本工作流程如下。

1) 控制器上电,主控根据上层应用,经配置总线接口通过 SPI 接口加载 MAC 硬件可重构电路配置,实现对应协议控制器 MAC 逻辑。

2) MAC 电路配置加载完毕后,内部主控通过配置接口,对 PMA 与 PCS 电路进行相应模式配置,内部主控监测反馈信息以及链路状态。

3) 上层应用通过配置接口轮询确认反馈信

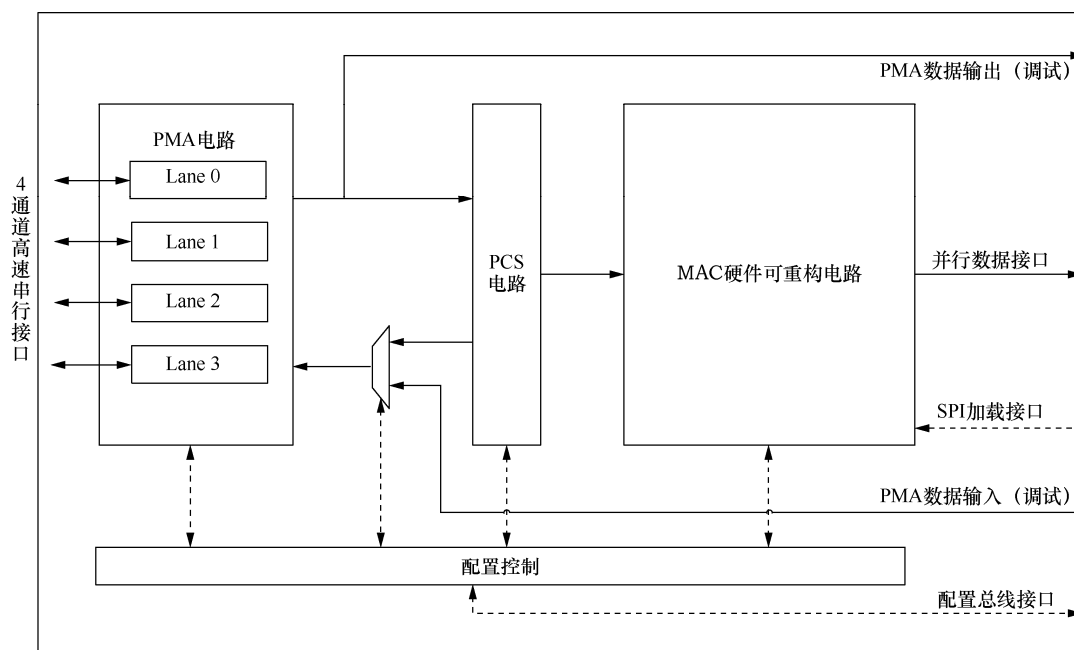


图 2 多协议控制器架构
Figure 2 The architecture of multi-protocol controller

息与链路状态无误后，开始数据传输。

所有工作流程均通过内部主控完成软件定义硬件结构配置，内部主控通过内嵌的一个轻量化 MCU 实现，用于验证软件定义功能。整个协议控制器的工作模式，都可以通过软件定义的形式进行灵活切换，从而满足系统不同应用场景下的异构需求。

3 多协议控制器架构设计

3.1 可重构多协议 MAC

多协议控制器架构主要应用于数据中心软件定义互联的交换设备或芯片中，因此本文多协议控制器 MAC 电路并不解析数据包，而是实现数据链路层功能。而每种协议 MAC 电路具体功能各不相同，对于 FC 协议而言，MAC 层处理包括协议数据链路层和交换适配层^[8]，而其中 FC 协议的交换适配层是 FC 的特化处理算粒（简称 FC-AE-ASM 控制器），无法与其他协议兼容；以太网 MAC 电路除必要的流控功能外，其特殊设计的数据校验与重传机制^[9]都是差异化较大的处理；SRIO 协议 MAC 层采用基于信用的流控处理，而且具备 SRIO 维护包管理层功能^[10]，处理复杂且无法与以上两种协议 MAC 处理兼容。

因此，多协议控制器 MAC 电路很难采用归一公共算粒的设计方法，并且由于存在不同通道绑定形式下的多种协议应用，单纯并行堆叠不同的 MAC 电路，势必带来庞大的冗余设计，而且无法适应软件定义互联的灵活多变的应用形态。综上，多协议控制器的 MAC 电路采用基于软件定义的硬件可重构电路进行实现。

针对 3 种协议控制器使用电路综合工具进行实现，并根据综合结果统计 3 种协议 MAC 部分的逻辑资源，如表 1 所示。

表 1 3 种协议以及多协议 MAC 电路资源统计
Table 1 Resource statistics of 3 protocols and multi-protocol MAC circuit

协议类型	MAC 逻辑资源占用	
	组合逻辑电路门数	时序电路门数
以太网	5 389	6 978
FC	2 379	3 962
SRIO	186 719	70 893
多协议控制器	192 103	74 156

根据以上各电路的资源数据，为了最大限度地复用 PCS 与 PMA 电路，从而确定可编程电路的资源要求，可以同时放下 4 个以太网或 FC 协议的 MAC 电路。

然而,SRIO 协议控制器 MAC 电路规模庞大,如果以 SRIO 资源为依据来确定可编程电路资源,会造成较大的资源冗余。所以,对 SRIO 协议 MAC 电路进行功能拆分,将 SRIO 协议中接口处理部分统一在 FC 与以太网可重构电路结构中实现,其他功能单独使用专用电路实现。这样既满足 FC 或以太网协议 4 个 1x 通道工作,又支持 SRIO 协议的功能需求,可以最大限度地完全利用 PMA 电路的 4 个 Lane 通道资源。

3.2 软件定义多协议 PCS 电路

本文中的协议控制器支持 3 种不同协议,如果根据协议种类设计多个完全独立 PCS 电路,虽解决了多协议的 PCS 电路需求,但由于电路设计复用度低,导致电路本身面积与功耗开销增加,从而直接影响系统的体积、重量、功耗等指标。

通过对比 SRIO 协议、FC 协议以及 Ethernet 协议对 PCS 子层的定义,提取 4 种协议对 PCS 子层的功能需求。多协议 PCS 要求支持包括编解码功能、加扰功能、变速箱功能、极性控制、位序控制、PRBS 生成、Comma 检测、Headers 检测、时钟补偿、通道绑定、PRBS 检测等。

提取 3 种协议在 PCS 子层中的公共算粒作为通用逻辑,并把差异化私有算粒做成可配置逻辑,从而实现多协议通用的 PCS 电路设计。PCS 电路

通过软件配置模式的不同,可以动态支持 SRIO 协议、FC 协议以及 Ethernet 协议规范数据的正确传输。同时支持与 SRIO Controller 动态捆绑模式的特殊需求,软件定义多协议 PCS 电路如图 3 所示。

编码扰码、解码解扰逻辑以及弹性 Buffer 逻辑都能随配置模式的改变而进行动态切换;弹性 Buffer 通过与通道绑定状态机配合能够完成 SRIO 协议需要的通道动态绑定功能。同时,通过配置模式选择是否使用弹性 Buffer。上层软件定义可以通过配置寄存器实现多协议 PCS 子层工作模式动态切换。

3.3 多协议 PMA 电路

多协议 PMA 电路是软件定义互联芯片的关键,是多协议互联的物理层基础,通过分析 3 种协议的特性可知,其各自物理层的功能结构差异不大。可以通过较为简单的归一化设计,实现同时满足 3 种协议的 PMA 电路,表 2 为不同协议需实现的功能列表。

由表 2 分析可知,3 种协议在 PMA 层的设计功能基本一致,其主要功能是将数据由并行转化为高速串行信号并发送至信道上(包括其逆过程),同时可以将高速的串行数据从信道上接收并转化为并行信号供上层使用,其电路系统如图 4 所示。

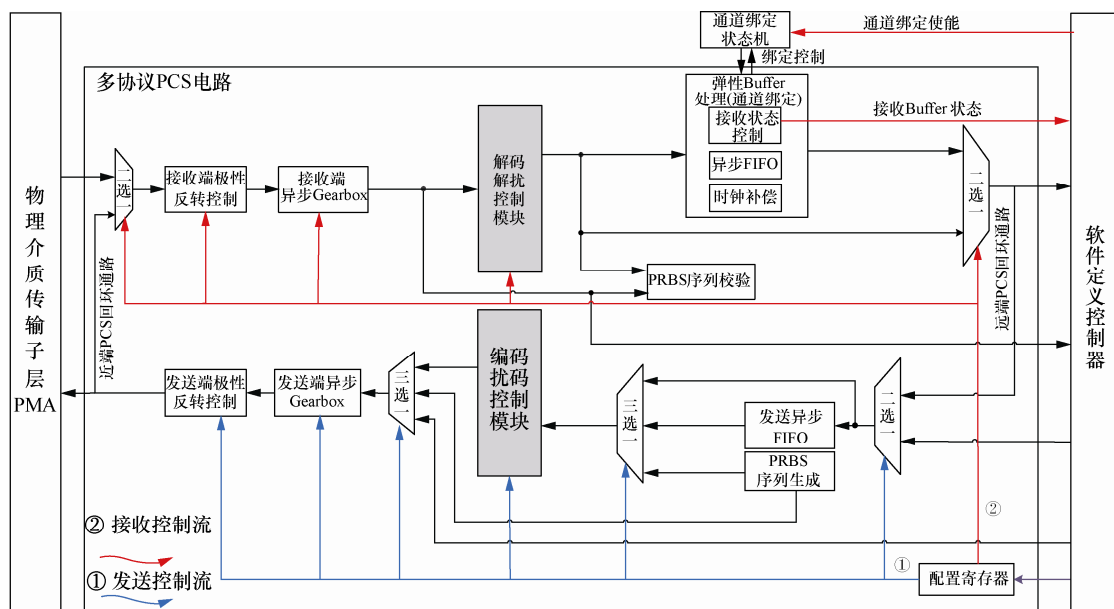


图3 软件定义多协议 PCS 电路
Figure 3 The circuits of software defines multi-protocol PCS

表 2 3 种协议功能列表
Table 2 Feature list of three protocols

功能	SRIO	FC	Ethernet
接收阻抗校正	支持	支持	支持
接收线性均衡 CTLE	支持	支持	支持
接收判决反馈均衡 DFE	支持	支持	支持
CDR	支持	支持	支持
串并转换	支持	支持	支持
内建 BIST 接收	支持	支持	支持
On-die-scope	支持	支持	支持
发送电阻校正	支持	支持	支持
发送摆幅调节	支持	支持	支持
并串转换	支持	支持	支持
可编程预加重	支持	支持	支持
内建 BIST 发送	支持	支持	支持
按序复位	支持	支持	支持
多频点	1.25/2.5/3.125/5/6.25/10.3125 Gbit/s	2.125/4.25/8.5 Gbit/s	1.25/10.3125 Gbit/s
多种环回模式	支持	支持	支持

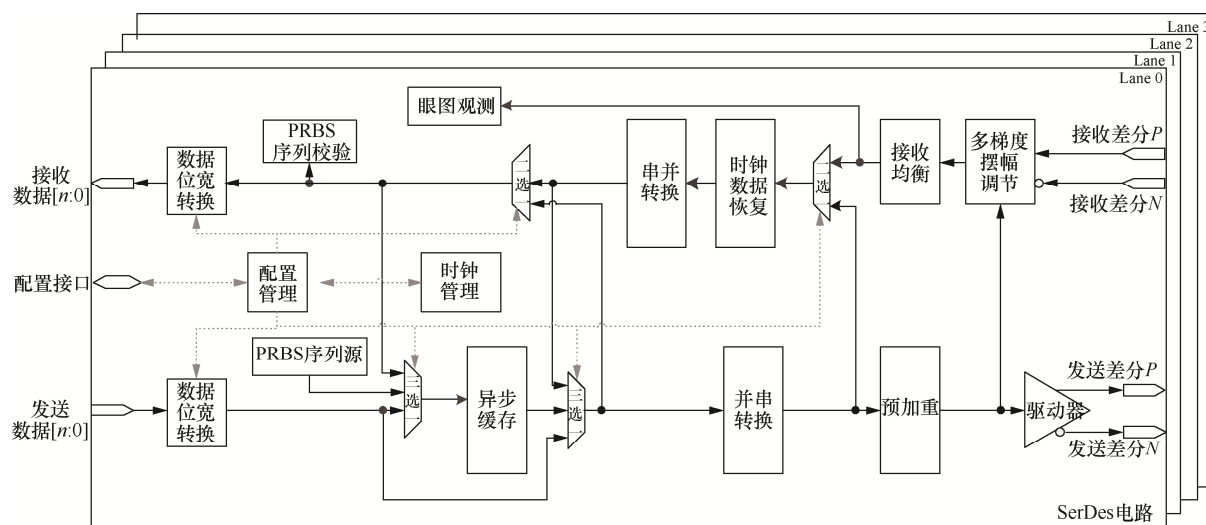


图 4 多协议 SerDes 设计架构
Figure 4 Multi-protocol SerDes design architecture

PMA 电路按照功能子模块的方式进行设计，功能模块之间的数据与控制关系如图 4 所示。由于 3 种协议对 PMA 电路的功能要求基本一致，因此将功能相似的部分设计为通用逻辑，电路按照 4 个 Lane 一组进行设计，主要由 3 部分组成：收发主通道、时钟处理模块、配置管理及复位模块。

接收器：包括阻抗校正电路、均衡器（连续时间线性均衡 CTLE 和判决反馈均衡 DFE）、时

钟数据恢复电路和逻辑（CDR）、相位差值器、并行数据位宽可配置的串并转换电路、PRBS 检测、内建接收通路环回等。主要根据输入数据来调整时钟，保证对数据的正确采样，同时将串行数据转化为并行数据。

发送器：包括并串转换电路、输出驱动器、预加重控制电路、边沿控制电路、阻抗校正电路、PRBS 校验生成、内建发送通路环回以及数据位

宽可配置的并串转换电路等, 完成由并行到串行数据的转化和驱动。

时钟管理: 差分驱动模块, 主要包括 4 个 Lane 共用的锁相环电路 QPLL 以及各 Lane 使用的锁相环电路 CPLL、时钟选择、分频电路等。该模块可以根据软件配置, 为 PMA 中的 SerDes 提供 3 种协议各自频点需求的时钟。

配置管理: 主要包括配置总线管理模块、寄存器管理模块, 负责对数据路径选择、时钟频率、数据位宽以及子模块工作模式实现软件配置控制。

整个 PMA 电路通过软件定义可配置电路, 可以实现针对不同协议、不同频点的结构切换, 从而达到软件定义的协议异构功能。

4 原型实现与测试分析

4.1 系统搭建

基于本文架构设计的控制器电路, 在 40 nm 工艺节点下进行 MPW 流片。测试芯片中包括可编程电路与 ASIC 电路, PMA 电路中的 4 通道 SerDes 电路, 设计可以不经 PCS 电路的测试接口, 直接进行单独数据测试。图 5 为多协议控制器测试芯片布局。

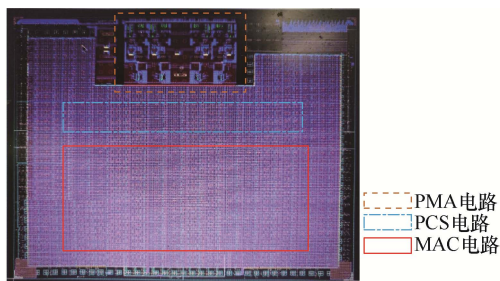


图 5 多协议控制器测试芯片布局
Figure 5 The floorplan of the multi-protocol controller

围绕测试芯片, 本文构造了一个多协议原型验证平台。验证平台主要由测试芯片验证板与端点设备(第三方标准测试仪或 FPGA 端点设备), 原型验证平台架构如图 6 所示。

4.2 电气特性测试

高速协议在进行具体的协议端点设备测试前, 需要先确认 PMA 电路模拟 SerDes 电路的功能与性能。因此, 首先使用高速示波器对 SerDes 进行测试, 确认模拟电路无误后, 再展开协议功能测试。图 7~图 9 为针对 FC、SRIO

以及以太网等协议需要的最高频点下的眼图测试结果, 其中横坐标代表测试时间周期, 纵坐标代表输出信号电压幅值。

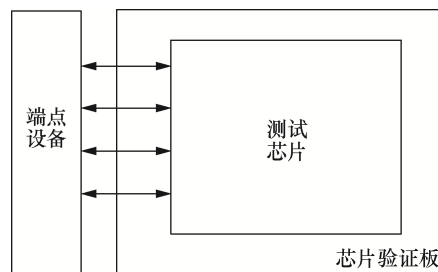


图 6 原型验证平台架构
Figure 6 The diagram of prototype verification platform

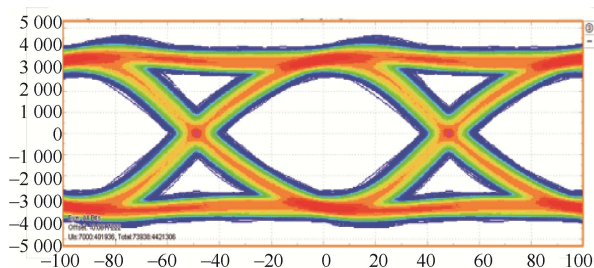


图 7 1x 通道 8.5G 频点 TX 眼图
Figure 7 The eye diagram of 1x 8.5Gbit/s TX Channel

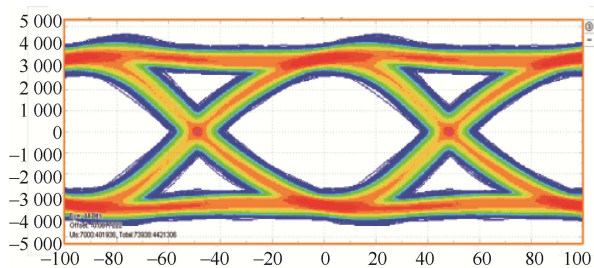


图 8 4x 通道 10.3125G 频点 TX 眼图
Figure 8 The eye diagram of 4x 10.3125 Gbit/s TX Channel

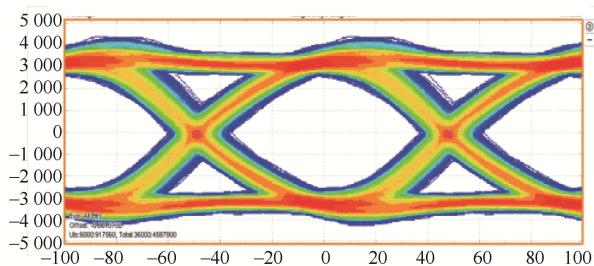


图 9 1x 通道 10.3125G 频点 TX 眼图
Figure 9 The eye diagram of 1x 10.3125 Gbit/s TX Channel

由测试结果可知, 输出发送信号均符合相关协议要求, 信号随机输出抖动 RJRMS 参数在 1.81~1.95 皮秒。同时在以上测试中, 不同协议频

点使用软件配置, 实现动态切换测试。

SerDes 功能性能确认后，分别对 3 种协议进行端点设备通信测试。以太网与 FC 协议均采用同第三方标准协议分析仪进行数据流量测试，并通过测试仪的上位机调试软件确认链路状态与数据包校验结果。SRIO 协议测试使用 Xilinx FPGA 开发板作为测试端点设备，端点设备实现对应频点速率的标准 SRIO 协议 IP 控制器，最终通过读取端点设备数据校验结果确认数据通路完整性。

4.3 协议一致性测试

4.3.1 以太网协议测试结果

软件配置控制器为以太网协议工作模式（包含 MAC 可编程电路的加载配置）。完成协议控制初始化后，对以太网协议分析仪配置数据接收/发送为标准 Ethernet 格式，数据率为 10.312 5 Gbit/s，数据包格式配置如图 10 所示。

完成配置后启动协议分析仪发送数据包，确认发送与接收端流量与速率，测试结果如图 11 所示。

4.3.2 FC 协议测试结果

与以太网测试类似，首先软件配置控制器切换为 FC 协议工作模式，再对 FC 协议分析仪进行配置，设置数据率为 8.5 Gbit/s，测试模式为端点设备与测试设备互发通信，链路和数据测试结果如图 12 所示。

测试同时 4 个 FC 通道在 8.5 Gbit/s 数据率下, 发送与接收端数据传输情况。由结果可知链路建立

稳定, 无失锁与误码现象, 满足 FC 协议通信要求。

Name	Value
Frame	
EthernetII	
Preamble (hex)	ff-55-55-55-55-55-55-55-55
Destination MAC	00:10:9a:00:00:03
Source MAC	00:10:9a:00:00:02
EtherType (hex)	<auto> Internet IP
IPv4 Header	
Version (int)	<auto> 4
Header length (int)	<auto> 5
Tos/Diffserv	tos (Dx00)
Total length (int)	<auto> calculated
Identification (int)	0
Control Flags	
Reserved (bit)	0
DF Bit (bit)	0
MF Bit (bit)	0
Fragment Offset (int)	0
Time to live (int)	255
Protocol (int)	<auto> UDP
Checksum (int)	<auto> 14976
Source	192.85.1.2
Destination	192.0.0.1
Header Options	
Gateway	192.85.1.1
UDP Header	
Source port (int)	1024
Destination port (int)	<auto> 1024
Length (int)	<auto> 0
Checksum (int)	<auto> 30357

图 10 Ethernet 协议分析仪数据包格式配置

Figure 10 The package format configuration of ETH protocol analyzer

4.3.3 SRIO 协议测试结果

SRIO 协议测试中的端点使用另一块 FPGA 平台，利用 FPGA 端点设备中的 Chip scope 配合 FPGA EDA 工具实时抓取数据，以确认测试数据结果。

测试方法是端点设备发送给多协议控制器后,多协议控制器将接收到的数据发回给端点设备确认结果。测试中配置端口设备与测试端频点速率均为 10.312 5 Gbit/s,在确认链路锁定成功后,启动端点设备数据通信并实时抓取数据。

经测试, 确认 SRIO 协议在 10.312 5 Gbit/s 速率下, 在 1 h 连续通信测试中误码率为 10^{-14} , 满足 RapidIO 协议误码率低于 10^{-12} 的要求。测试



(a) 接收端口速率测试结果



(b) 发送端口速率测试结果

Port Traffic and Counters > Basic Traffic Results					Change Result View				
Basic Counters		Errors	Triggers	Protocols	Undersize/Override/Unblock	PFC Counters		User De	
Port Name	Total Tx Count (Frames)	Total Rx Count (Frames)		Total Tx Count (bits)		Total Rx Count (bits)			
Port 1/1/1	2,079,337,882	0		2,129,241,991,168		0			
Port 1/1/2	0	2,079,337,882		0		2,129,241,991,168			

(c) 接收端口流量测试结果

Streams > Detailed Stream Results						
Change Result View 1 of 1 Select Tx Ports: All						
All Ports Change Counter Mode: Basic Mode Reset Example						
Basic Counters		Errors		Basic Sequencing		Histograms
Name/ID	Tx Port Name	Rx Port Names	Tx Count (Frames)	Rx Count (Frames)	Dropped Count (Frames)	Dropped Frame Percent
StreamBloc:	Port //1/1	Port //1/2	2,879,337,882	2,879,337,882	0	0.000

(d) 发送端口流量测试结果

图 11 Ethernet 协议分析仪流量测试结果

Figure 11 Test result of Ethernet protocol analyzer traffic

所述, 考虑上述 5 种应用场景, 可支持的电路数据计算如下。

$$2^1 C_4^1 + 2^2 C_4^2 + 2^3 C_4^3 + 2^4 C_4^4 + 1 = 81$$

对应的 MAC 电路可实现 81 种不同的软件定义可重构控制器形态。经过使用原型环境分别加载以上 81 种软件定义控制器, 各种应用形式均通过 4.2 节与 4.3 节的电气特性与协议一致性测试, 符合支持多协议控制器多形态应用设计预期。

5 结束语

本文设计的基于软件定义的多协议控制器, 支持 FC、以太网以及 SRIO 这 3 种高速串行协议标准, 并具备通过软件定义切换不同协议工作模式的功能。为今后拟态计算平台高异构性需求系统的数据互联, 提供适合的解决方案。同时, 构建的原型验证的测试结果表明, 本文提出的基于软件定义的多协议控制器可以满足 Fiber Channel、Ethernet 以及 SRIO 协议的要求, 满足设计预期。

参考文献:

- [1] 郭江兴. 新型网络技术发展思考[J]. 中国科学:信息科学, 2018, 48(8): 144-153.
WU J X. Thoughts on the development of novel network technology[J]. Science in China: Information Sciences, 2018, 48(8): 144-153.
- [2] MCKEOWNN, ANDERSON T, BALAKRISHNAN H, et al. Open-flow: enabling innovation in campus networks[J]. ACM Sigcomm Computer Communication Review, 2008, 38(2): 69-74.
- [3] HAKIRIA, GOKHALE A, BERTHOUE P, et al. Software-defined networking: challenges and research opportunities for future Internet[J]. Computer Networks, 2014, 75: 453-471.
- [4] LAN L, HAN M O, HU Y X. SDN Architecture: research and practice[J]. ZTE Technology Journal, 2013.
- [5] 田铭, 郭江兴, 兰巨龙, 等. 信息中心网络元模块承载的差异化服务模型[J]. 电子与信息学报, 2016, 38(11): 2940-2947.
TIAN M, WU J X, LAN J L, et al. Differentiated service model based on meta module in information centric networking[J]. Journal of Electronics & Information Technology, 2016, 38(11): 2940-2947.
- [6] 兰巨龙, 程东年, 胡宇翔. 可重构信息通信基础网络体系研究[J]. 通信学报, 2014(1): 128-139.
LAN J L, CHEN D N, HU Y X. Research on reconfigurable information communication basal network architecture[J]. Journal on Communications, 2014(1): 128-139.
- [7] 唐龙飞, 田泽, 邵刚, 等. 一种多协议统一架构 CMOS Serdes 发送器电路设计[J]. 计算机技术与发展, 2015(5).
TANG L F, TIAN Z, SHAO G, et al. Design of a multi-protocol unified architecture CMOS serdes transmitter circuit[J]. Computer Technology and Development, 2015(5).
- [8] Fiber channel avionics environment anonymous subscriber messaging, Rev 1.2[R]. 2006.
- [9] IEEE Standard for Information technology telecommunications and information exchange between systems local and metropolitan area networks specific requirements[S]. IEEE Std. 802.3-2008.
- [10] RapidIO interconnect specification part 1: input/output logical specification, Rev 3.1[R]. 2014.

[作者简介]



夏云飞 (1987-), 男, 黑龙江海伦人, 天津市滨海新区信息技术创新中心工程师, 主要研究方向为软件定义互联、高性能芯片架构设计以及信号处理算法高性能实现。



张丽 (1982-), 女, 河南辉县人, 国家数字交换系统工程技术研究中心副研究员, 主要研究方向为软件定义互联、SoC 芯片设计。



杨堃 (1976-), 男, 黑龙江哈尔滨人, 天津市滨海新区信息技术创新中心工程师, 主要研究方向为软件定义互联、拟态计算以及 SoC 芯片全流程实现。



李沛杰 (1990-), 男, 山西襄汾人, 国家数字交换系统工程技术研究中心助理研究员, 主要研究方向为软件定义互联、拟态计算以及高速 SerDes 电路设计。



许立明 (1990-), 男, 山东潍坊人, 天津市滨海新区信息技术创新中心工程师, 主要研究方向为工作高速数据电路实现、芯片设计前端流程以及 SoC 芯片系统时钟架构设计。