



2025年9-12月考核报告

导 师：朱科军，章红宇

专 业：核技术及应用

2025年12月25日



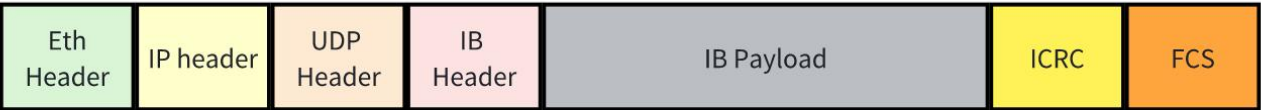
- 背景介绍
- 固件开发与调试
- 软件开发与调试
- 数据传输测试
- 总结与计划



背景介绍

● 研究目标

- 使用RDMA技术代替传统数据传输协议进行FPGA到DAQ系统的数据传输，提高数据传输效率，节约计算资源
- 建立连接、RoCE v2 功能支持（Write、Send、Receive、Read）



RoCE v2 数据包结构

● 参考项目

- RoCE core+Microblaze+CMAC+DDR4 MIG
 - 点对点数据传输，实现 RDMA Write 等操作，基于 vcu128，需要移植

● 项目进展

- 基本完成项目的移植，解决移植过程中的各种问题
- 搭建点对点传输测试平台，物理层、链路层连接成功
- RDMA 连接成功建立，不能进行数据传输

难点

- 协议复杂性高，状态机设计复杂
- 数据包处理需在极小时钟周期内完成
- 缓存报文、维护连接状态表需大量存储资源
- 拥塞控制与流量管理

兼顾高性能、低延迟、协议复杂性和资源效率

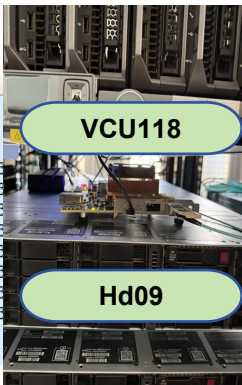
需要完成固件调试

Source	Destination	Protocol	Length	Info
192.168.160.10	192.168.160.32	RRoCE	322	CM: ConnectRequest
192.168.160.32	192.168.160.10	RRoCE	322	CM: ConnectReply
192.168.160.10	192.168.160.32	RRoCE	322	CM: ReadyToUse
192.168.160.10	192.168.160.32	RRoCE	74	RC Send Only QP=0x000003
192.168.160.10	192.168.160.32	RRoCE	74	RC Send Only QP=0x000003
192.168.160.10	192.168.160.32	RRoCE	74	RC Send Only QP=0x000003
192.168.160.10	192.168.160.32	RRoCE	74	RC Send Only QP=0x000003
192.168.160.10	192.168.160.32	RRoCE	74	RC Send Only QP=0x000003
192.168.160.10	192.168.160.32	RRoCE	74	RC Send Only QP=0x000003
192.168.160.10	192.168.160.32	RRoCE	74	RC Send Only QP=0x000003
192.168.160.10	192.168.160.32	RRoCE	322	CM: DisconnectRequest
192.168.160.32	192.168.160.10	RRoCE	322	CM: DisconnectReply

连接建立

数据传输过程，
FPGA 不向 PC 发
送回包

连接断开



wireshark 抓包结果

Hd09 与 VCU118 开发板

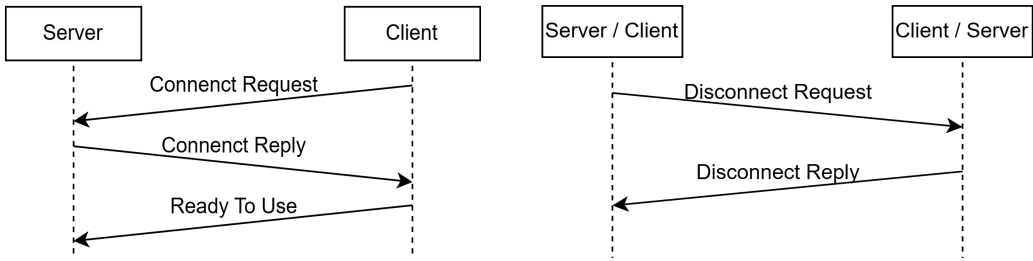


固件开发与调试

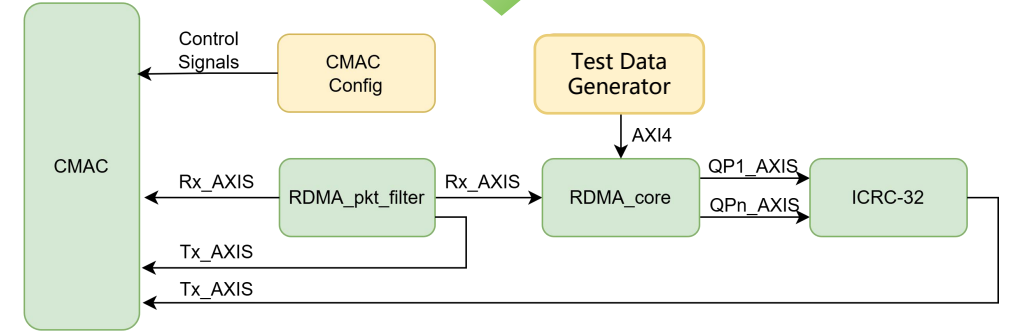
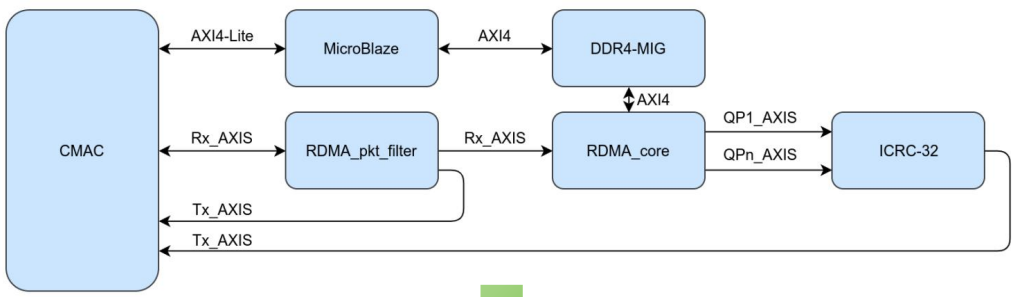
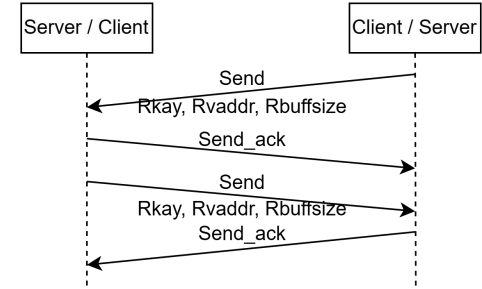
● 固件调试

- 修改 CMAC IP 核的配置方法
- 时序问题优化
- 调试过程中完成 ICRC 校验脚本以及数据包解码脚本
- 添加测试数据生成模块，用于数据传输测试

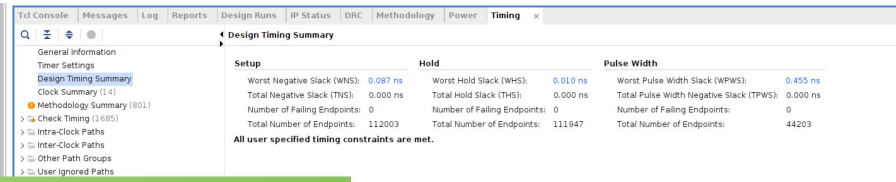
● 连接建立与断开



● Send 功能的实现



固件结构调整



时序结果

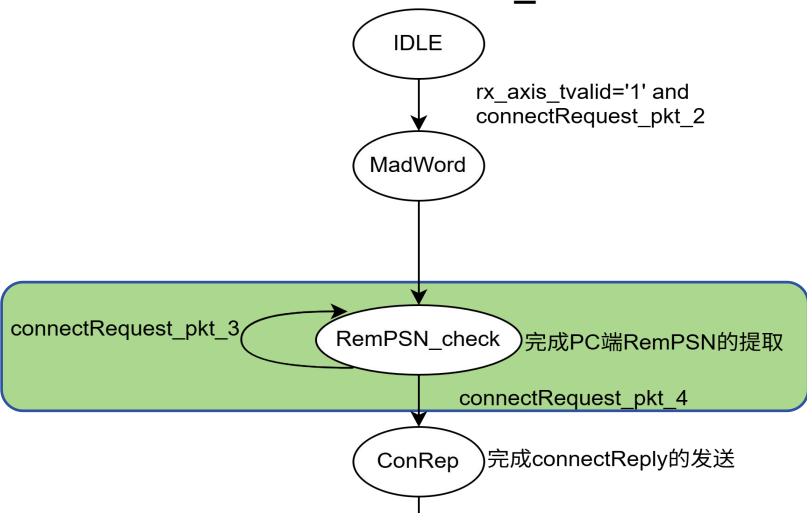
- 解决时序问题
- ILA 正常工作
- 完成连接建立数据包、Send 数据包结构检查
- 需要进一步完成数据传输测试



固件开发与调试

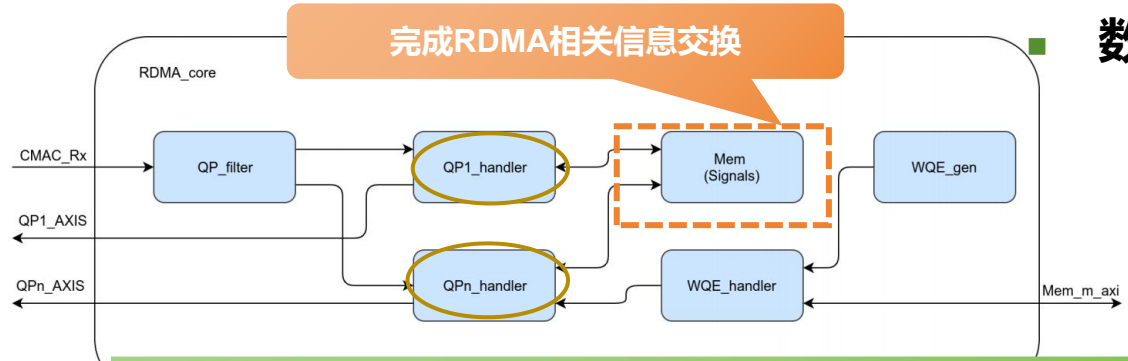
● 状态机展示

■ 连接建立 (QP1_handler)



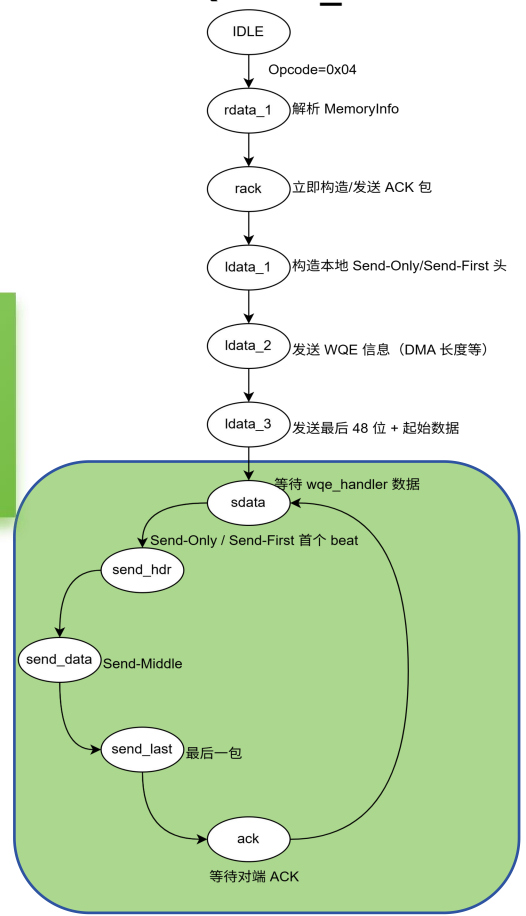
- **QPN (Queue Pair Number):**
 - RDMA 数据连接过程中每个连接需要维护一个对应队列对，每个队列对有相应编号
 - 连接建立过程中 QPN=0x01，数据传输过程中 QPN 由 Client (PC) 决定
 - 在 ConnectRequest 数据包中提取 RemQPN
- **PSN (Packet Sequence Number):**
 - PSN 用于检测有无丢包
 - 在 ConnectRequest 数据包中提取 RemPSN

QP1_handler



- 解决连接建立过程以及数据发送过程中的问题，实现连续数据发送
 - PSN，QPN 以及数据发送逻辑

■ 数据包生成(QPn_handler)



- **数据包组装**
 - 无 ACK 数据包: QPN 不对应
 - PC 端应用层接收数据失败: PSN 不对应
 - 在打包过程中，添加相应的 QPN，PSN 字段
 - 分别维护 ACK 数据包与 Send 数据包的 PSN 字段
- **修改数据发送逻辑**
 - 修改状态机逻辑
 - 使用 Send 功能完成测试数据的发送

QPn_handler



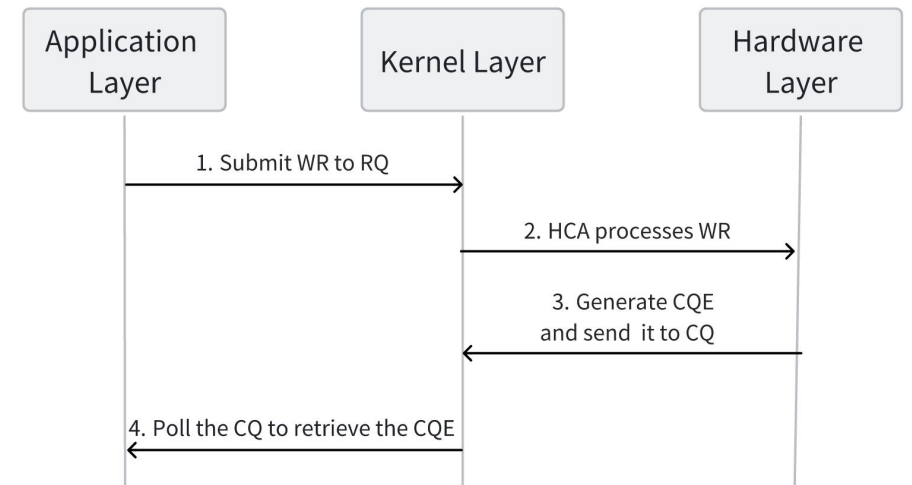
软件开发与调试

● 连接建立

- 使用 RDMA CM 提供的 API 完成连接建立
- 不需要提前指定对应的 QPN, PSN
- MEM Base ADDR, MEM Key 通过 Send 功能发送

● 数据传输过程

- 提交工作请求 (Work Request - WR)
- 硬件处理 WR
- 生成完成队列项 (Completion Queue Entry - CQE)
- CQE 进入完成队列 (Completion Queue - CQ)
- 轮询 CQ (Polling CQ)



RDMA Workflow on PC

■ 忙轮询 (Busy-Polling) - 速度最快，但最耗费资源

- 应用程序在一个紧凑的循环中，不停地、连续地调用轮询函数（如 `ibv_poll_cq`）来检查完成队列（CQ）中是否有新的完成队列项（CQE）

■ 事件驱动/阻塞式轮询 (Event-Driven / Blocking) - 效率最高，但延迟较高

- 应用程序不主动轮询，而是请求硬件在有新的 CQE 到达时通过一个“事件”来通知它。应用程序的线程会“睡眠”或阻塞，直到事件发生

■ 混合轮询 (Hybrid Polling) - 最佳实践的折中方案

- 结合以上两种方式的优点。先进行一小段时间的忙轮询。如果在这段时间内没有等到 CQE，就转为事件驱动的阻塞模式，等待下一次通知

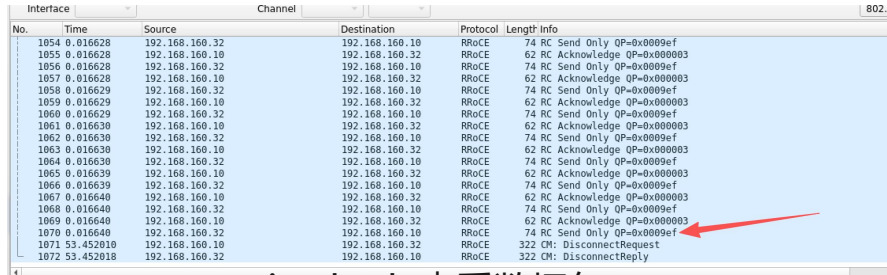
数据传输测试

FPGA —> PC (Send)

- 一段时间后，FPGA停止发送数据，PC端持续等待
- PC端：接收到532个来回数据包，并完成数据包内容的打印
- TCPdump: 1072个数据包，最后一个 send 数据包缺少 Send_ack 回包

PC —> PC (Send)

- 最高稳定数据传输速率：1 Gbps



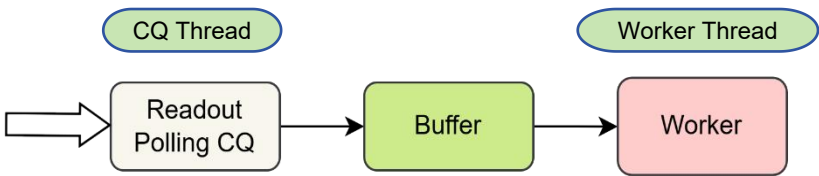
wireshark 查看数据包

```
[Status] Total received: 531, Queue size: 1
+++ [Client] RECEIVED PACKET #531 +++
WR ID: 630
Buffer Used: buffer[18]
Received Length: 16 bytes

Data Content (16 bytes):
0000: 00 00 00 00 23 ed 53 00 00 00 1d ce 00 00 00 04 |...#.S.....|

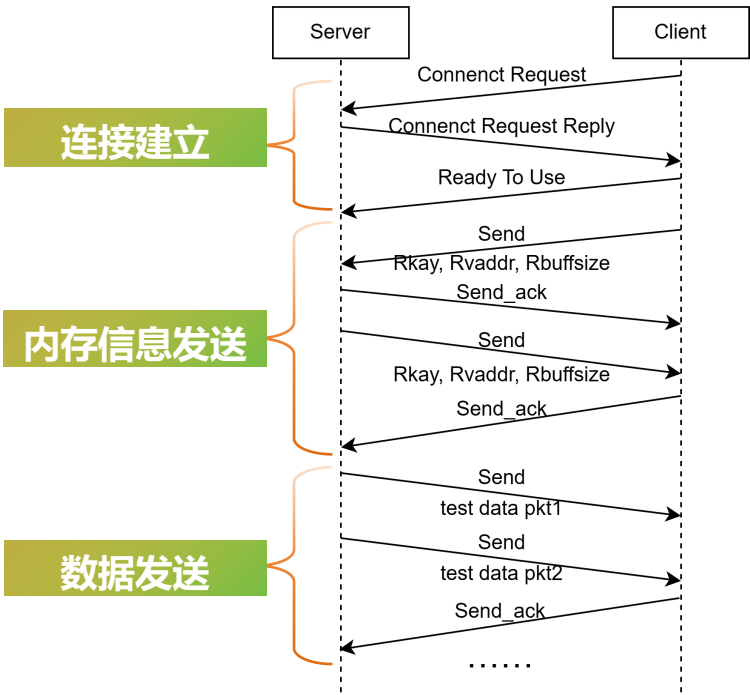
[Status] Total received: 531, Queue size: 0
[Client] Waiting for packets... (total received: 531, printed: 531)
[Client] QP State: RTS
[Client] Waiting for packets... (total received: 531, printed: 531)
[Client] QP State: RTS
[Client] Waiting for packets... (total received: 531, printed: 531)
[Client] QP State: RTS
[Client] Waiting for packets... (total received: 531, printed: 531)
[Client] QP State: RTS
```

PC 端打印内容



Software Structure

PC 端软件性能不够



固件与软件交互情况

PC	
CPU	Intel(R) Xeon(R) Gold 6226 CPU @ 2.70GHz
NIC	ConnectX-5
Server/Client	Client
FPGA board	
Model	VCU118 Evaluation Kit
QSFP28	Capable of 100 Gbps Ethernet
Server/Client	Server

测试条件

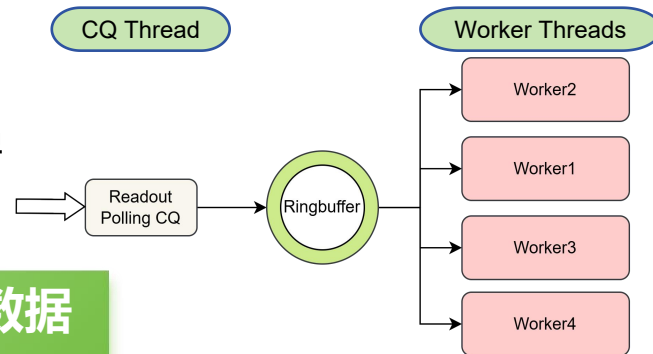


数据传输测试

● 软件优化 (PC->PC)

- 使用Ringbuffer做线程之间的数据传输，4个 Worker 线程共同消费 CQ 中数据
- 优化软件工作逻辑，指针传递代替数据拷贝
- 最高稳定数据传输速率：10 Gbps

软件端性能不够：调试用打印过多，单机数据处理，WR 分配数量不足



Software Structure

● 吞吐测试 (PC->PC)

- 去掉多余打印，暂时放弃数据处理
- 最高稳定数据传输速率：90 Gbps
- 可用于与FPGA的数据传输性能测试

```
=====
[Server] TRANSMISSION COMPLETE
Total packets: 82130944
Total bytes: 1345633386496
Total time: 120.00 sec
Average rate: 89.71 Gbps
=====
[Server] Done
```

数据传输情况

Server hd01	
CPU	Intel(R) Xeon(R) Gold 6226 CPU @ 2.70GHz
NIC	ConnectX-5
Server/Client	Client

Client hd04	
CPU	Intel(R) Xeon(R) Gold 6226 CPU @ 2.70GHz
NIC	ConnectX-5
Server/Client	Client

测试条件

- 通信链路基本搭建完成，Send 功能实现
- 需要实现 RoCE v2 的其他功能，并完成进一步测试

总结与计划



● 总结

■ 完成开源固件的移植与数据传输功能的调试

- Send_ack 数据包的产生与发送
- RDMA 建立连接过程中关键字段的提取
- Send 数据包的组装
- 连续 Send 数据包的发送

■ 初步完成 PC 端软件开发

- 进行FPGA与PC端数据传输测试

■ 报告

- 参加 2025 CEPC Workshop 会议
- 报告题目: FPGA based RDMA for BEE Readout

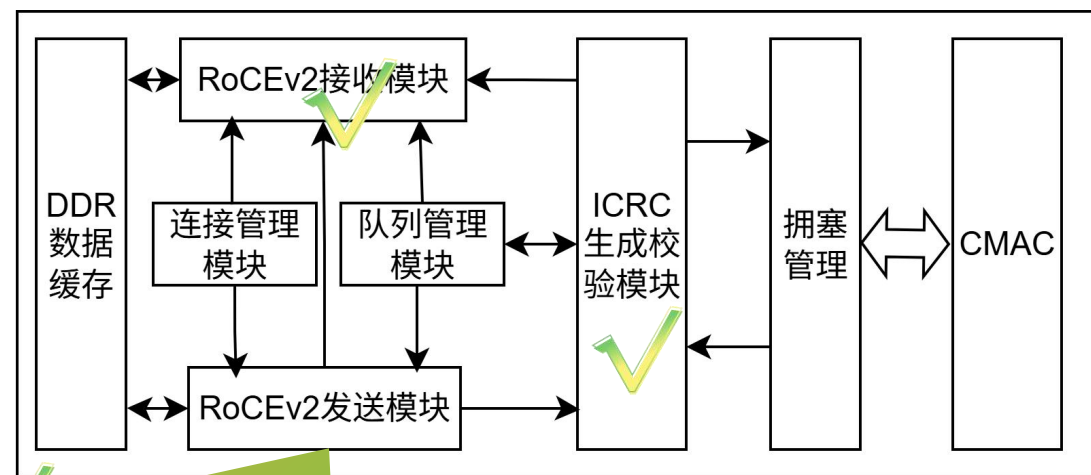
● 工作计划

■ 基于RDMA的高性能数据传输研究

- 进行数据传输速率测试
- 继续完善固件中 RoCE v2 协议相关功能 (Write)
- 固件中数据发送速率可调节
- 实现 RDMA QP 管理



2025 CEPC Workshop



- Send ✓
- Write(To be Continued...)

固件框架调整



感谢！