



2026年5-8月考核报告

导 师：朱科军，章红宇

专 业：核技术及应用

2026年8月27日



- 基于 RDMA 的高性能数据传输研究
- 其他工作
- 总结及计划



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

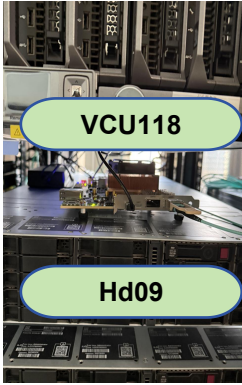
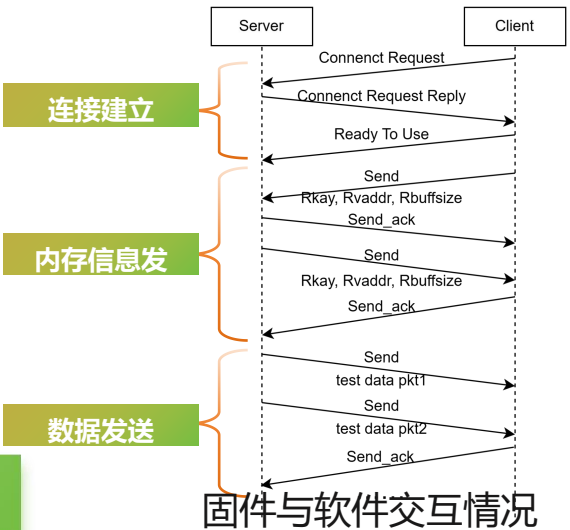
● 项目进展

- **Send 功能已实现，最高数据传输速率 21Gbps**
- **PC 端软件开发**
 - 支持数据传输测试
 - 支持极限吞吐率测试
 - 数据传输过程中错误检查
 - 离线数据包检查
- **PC 与 FPGA 进行数据传输测试**
 - 多次数据传输测试，查找问题出现原因

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



Hd09 与 VCU118 开发板

● **Send 数据包发送成功，最高数据传输速率 21.0 Gbps**



固件开发——提高数据传输速率

● 瓶颈：逐包等待 ACK

- 初始发送端采用Stop-and-Wait机制，每发送一个报文都需要等待ACK
- ACK往返时间导致发送流水频繁中断，FPGA-to-PC吞吐率仅为**21.0 Gbps**

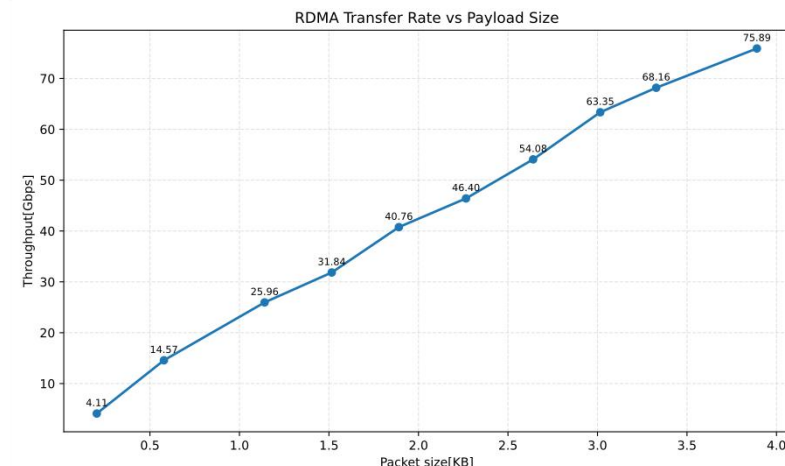
● 固件优化：滑动窗口与端到端反压

- 实现滑动窗口，允许多个未确认报文同时在途
- 将 AXI4-Stream TREADY 从 CMAC 逐级传回数据源
- 下游暂停接收时，上游保持 TDATA/TKEEP/TLAST 稳定，避免 FIFO 溢出和数据丢失
- 实现全流程反压：

Data Generator ← RoCEv2 TX ← ICRC ← TX FIFO ← CMAC

● 测试结果：在 payload 为 4096B 时达到最高，为 75.9Gbps

- 吞吐率随payload增大而提升
- payload=4096 B时，FPGA-to-PC 传输速率达到 **75.9 Gbps**



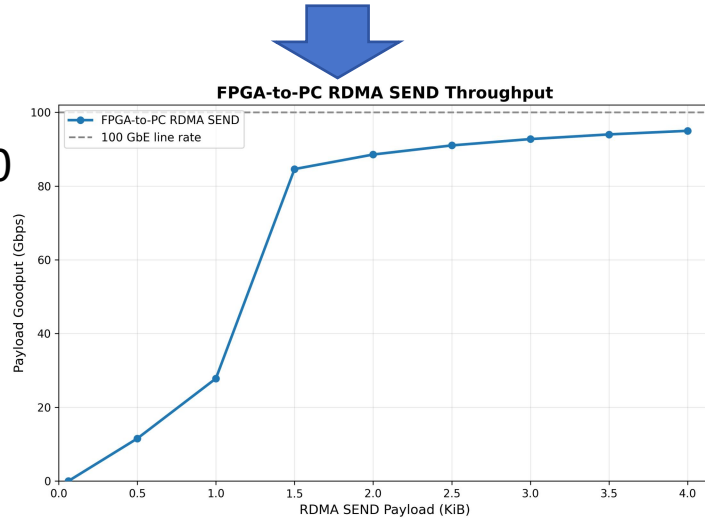
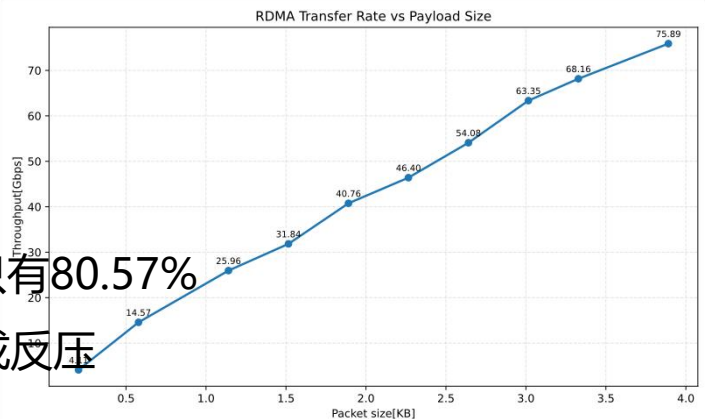
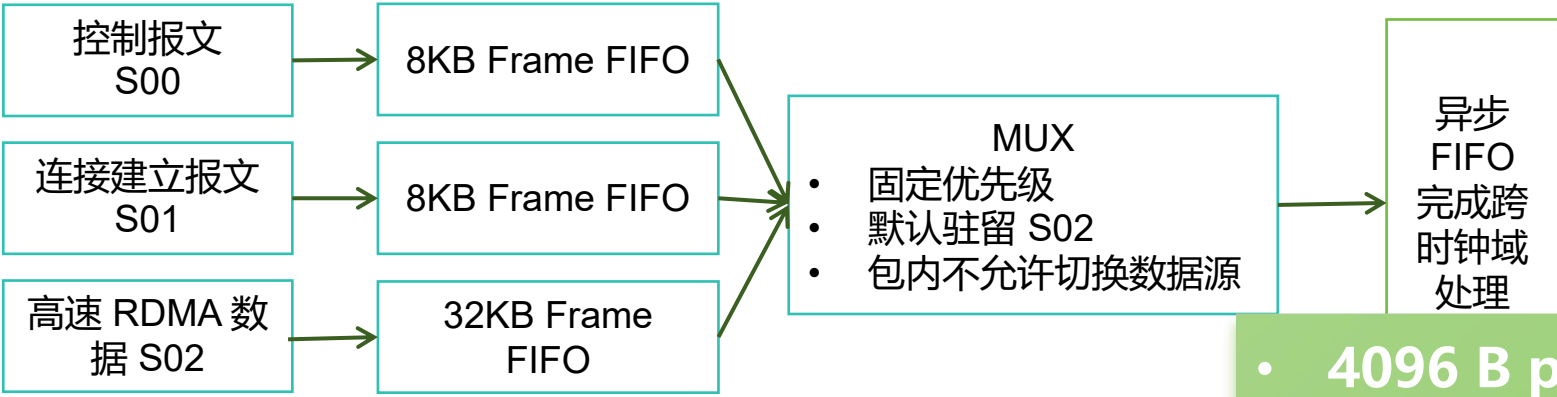
- 滑动窗口消除逐包等待时产生的时间
- 端到端 TREADY 反压保证下游拥塞时数据不丢失
- 4096 B payload 下，数据传输速率由 21.0 提升至 **75.9 Gbps**



固件开发——提高数据传输速率

● 问题排查

- 链路中存在慢速模块，影响数据传输速率
 - 在 QPn、ICRC 输出、Interconnect S02 输入和 CMAC 端分别加入 ILA
 - ICRC 输出 TVALID 有效的比例为100%，CRC 结果持续有效，但下游 TREADY 只有80.57%
 - S00/S01在整个采样窗口中TVALID=0，但 S02 仍周期性被反压，非三路竞争造成反压
 - CMAC侧 TREADY=96.09%、TX enable=100%，排除 CMAC 持续阻塞
- axis_interconnect_0 ——> axis_packet_arb_mux_3 + tx_axis_async_fifo
 - 三路Frame FIFO + 包级固定优先级仲裁 + 异步包FIFO 替换 axis_interconnect_0



Send数据传输速率测试结果

● 4096 B payload 下，数据传输速率由 75.9 Gbps 提升至 94.0 Gbps



固件开发——Write 功能开发

● RDMA WRITE发送功能

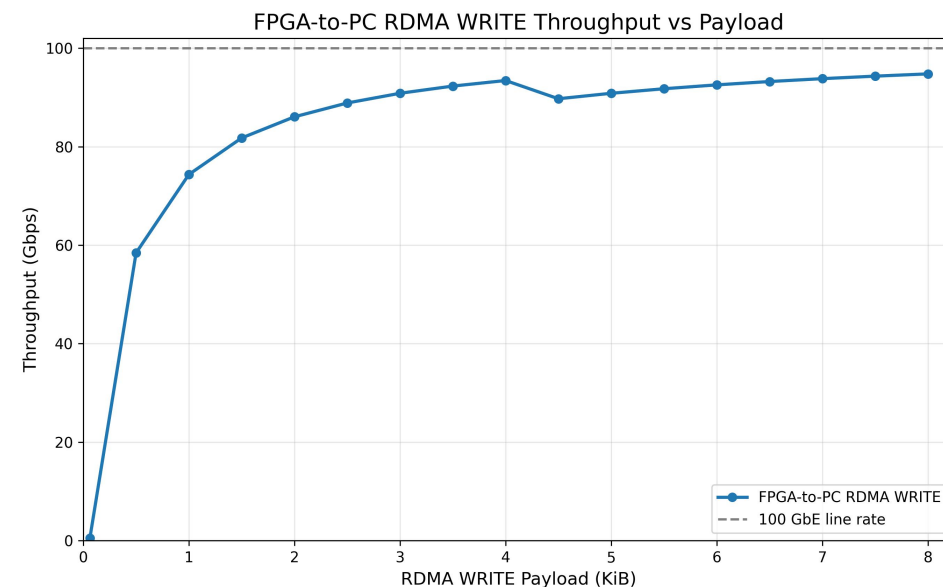
- 实现不带 Immediate Data 的 RC RDMA WRITE
- 根据 Path MTU 自动完成报文分片
 - 生成 WRITE First / Middle / Last / Only 操作码
 - 逐包递增 PSN
 - 逐消息递增远端虚拟地址
- 通过 VIO 独立选择 SEND / WRITE 工作模式

● 软件接收端与验证方法

- 注册并向 FPGA 交换一定大小的 MR 的地址和 RKey
- 通过 MR 尾字节变化检测整段内存是否写满
- 联合统计 QP 状态、PSN 增量及网卡 RDMA 错误计数

● 测试结果

- 完成 64 B ~ 8192 B Payload扫描测试
- 8192 B Payload 最高吞吐率达到 94.77 Gbps
- 测试过程中未出现 PSN 乱序和 ICRC 错误, MR 能够完整写满



Write 数据传输速率测试结果

- 完成 FPGA 到 PC 的 RC RDMA WRITE 闭环, 最高有效吞吐率达到 **94.77 Gbps**



固件开发——累计 ACK 滑动窗口与重传控制

● 累计 ACK 滑动窗口

- 累计 ACK : $\text{ACK_psn} = X$ 表示所有序号小于 X 的数据包均已确认接收
- 发送约束: $\text{psn_unACK} < \text{curr_PSN} < \text{psn_unACK} + \text{window_size}$
 - psn_unACK : 最早未被确认的数据包序号
- 窗口更新: 每次收到 ACK 数据包后, 更新 psn_unACK , 窗口右移

● 重传条件

- 收到 Sequence Error NAK
- 最早未确认数据包等待 ACK 超过 RTO
- 当前配置: RTO 约 10 ms, 最大重试 7 次
- 重放期间暂停 RTO 计时, 连接断开时清除重传状态

● 重传范围

- 采用 Go-Back-N 策略
- 从 NAK 或 RTO 指定的 PSN 开始
- 重传至触发时已经完整缓存的最后一个 PSN
- 所有重传数据重新经过 ICRC 计算后发送

- 当前完成 Sequence NAK 与 RTO 处理
- RNR NAK、Implied NAK及错误分类处理仍待补充



固件开发——BRAM/DDR两级重传实现

● 两级重传缓存

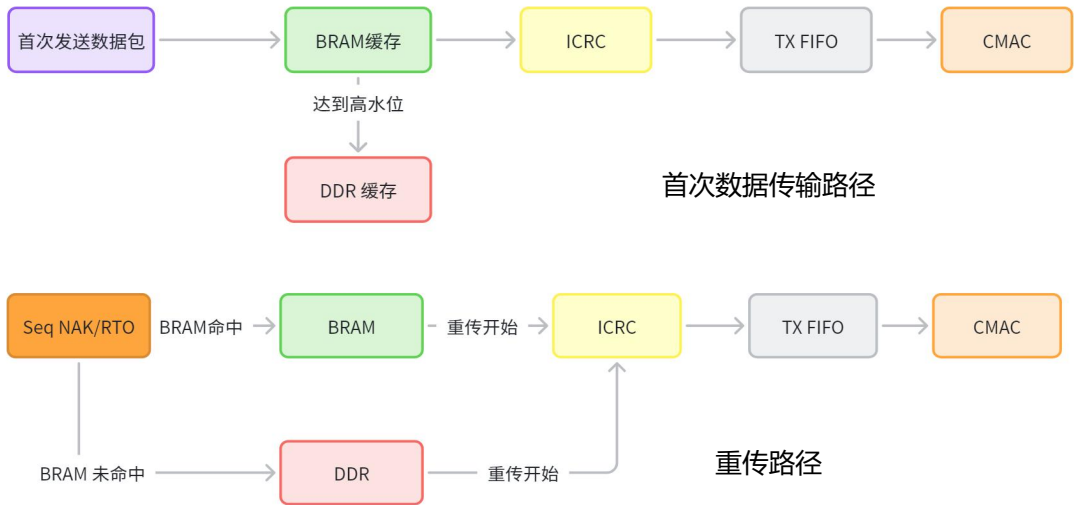
- 完整数据包同时进入网络通路和BRAM热缓存
- BRAM 配置 64 个包槽，保存完整的 pre-ICRC AXIS 报文
- BRAM 占用达到 48 后，将较旧且未确认的数据异步下沉至 DDR
- BRAM 占用回落至 32 后停止下沉
- DDR 仅作为冷数据缓存，不进入正常首发串行路径

● 重传路径

- BRAM 命中：直接读取 BRAM 并低延迟重放
- BRAM 未命中：查询已提交的 DDR 缓存
- DDR 命中：读取 DDR 数据并送入重放通路

- Sequence NAK、RTO及DDR冷重传均完成闭环验证
- 关闭故障注入后，正常通路仍保持约94 Gbps有效吞吐率

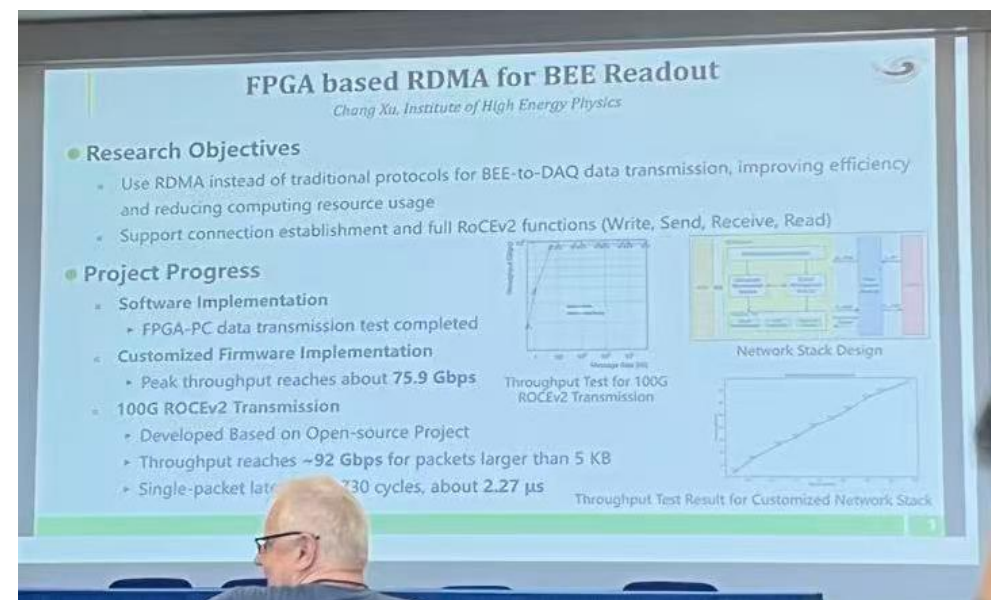
测试项目	故障注入方式	恢复结果
BRAM热重传	缓存后丢弃第64个首发WRITE包	OOS +1, Sequence NAK触发BRAM重传21包, 16 GiB MR完整
RTO恢复	80包突发中接收端丢弃尾部14包	RTO触发1次, BRAM重传14包, 最终80/80包接收完成
DDR冷重传	暂停ACK释放并对DDR内PSN注入Sequence NAK	BRAM按预期miss, DDR重传48包后衔接BRAM重传32包
高速回归	关闭全部故障注入	SEND 94.98 Gbps, WRITE 94.93 Gbps
协议检查	检查QP、网卡及缓存计数器	QP保持RTS, ICRC/AXI错误均为0





● IEEE Real Time Conference 2026

- 参加会议
- 提交论文初稿



Real Time Conference 2026



● 总结

■ 100 Gb/s数据通路优化

- 使用三路 Frame FIFO 与包级仲裁器替换 AXI Interconnect
- FPGA-to-PC SEND、RDMA WRITE 最高达到 **94 Gbps** 左右

■ RoCEv2协议功能实现

- 完成 RDMA WRITE 组包及 MTU 分片
- 实现 Sequence NAK 与 RTO 触发的 Go-Back-N 重传
- 构建 BRAM 热缓存与 DDR 冷缓存两级重传架构

■ 可靠性验证

- 完成真实丢包、尾部丢包及 DDR 冷重传故障注入
- 测试过程中 QP 保持 RTS, PSN、OOS 和 ICRC 错误均无新增

■ 其他

- IEEE Real Time Conference: 参会以及提交论文

● 工作计划

- 完善重传机制
- 实现多 QP 独立状态管理与公平调度
- 开展小时级可靠性及自动化回归测试
- 完成 RoCE v2 协议栈 IP 核封装



感谢！