Contribution ID: **76** Type: **Poster**

# Performance profiling and design choices of an RDMA implementation using FPGA devices

*Tuesday 3 October 2023 15:00 (20 minutes)*

RDMA communication is an efficient choice for many applications, such as data acquisition systems, data center networking and any other networking application where high bandwidth and low latency are necessary. RDMA can be implemented using a large array of options which need to be tailored to the needed use case in order to get optimal results. Aspects such as the effects of using multiple simultaneous connections, using various transport functions such as RDMA Write and RDMA Send and communication models such as sending individual bursts or continuous streams of data will be investigated for implementing RDMA on FPGA devices.

## Summary (500 words)

Starting with LHC Run3, the FELIX (Front-End Link eXchange) system is used for implementing the data acquisition system of a small number of ATLAS subdetectors, and from LHC Run4 onward, FELIX is planned to become the data acquisition system used by all ATLAS subdetectors. The FELIX system is based on a custom FPGA board which receives data from the front-end detector electronics via optical links and outputs data via a PCIe interface to a host computer which manages processing and relaying the data further to the readout system. The host uses the RDMA (Remote Direct Memory Access) support offered by network interface cards with RoCE (RDMA over Converged Ethernet) support to transmit data further towards the readout systems over Ethernet.

Taking into account the foreseen increase in data rate which will happen as a consequence of the High Luminosity LHC upgrade, a possible improvement of the FELIX system was proposed. This improvement would avoid the potential bottleneck of the PCIe interface and of the FELIX PC CPU by implementing RDMA support in the FELIX FPGA itself, thus removing the FELIX PC itself from the data path of the readout system. The data would now be flowing directly from the FELIX board to the Software ROD PCs in the server farm.

This proposed FPGA implementation of RDMA was demonstrated both outside FELIX (TWEPP2021) and with a modified version of FELIX (TWEPP2022). This showed that the FPGA RDMA implementation can reach the expected performance of an RDMA link.

To make this FPGA RDMA implementation useful in a production setting, a number of issues need to be explored, understood and handled:

First, there is the issue of multiple simultaneous connections between the FPGA board running the RDMA implementation and multiple potential receivers. There is the matter of how many simultaneous connections can be handled, and what will be their effect on the operation of the FPGA device and the RDMA link(s).

Then, there is the issue of the RDMA transport function being used. RDMA Send is similar to how TCP/IP sockets work, offer easy to use streaming of RDMA messages but it involves both endpoints of an RDMA link, which can lead to higher overhead and latency. This is what is currently in use with FELIX software and Software ROD. RDMA Write needs a pre-allocated memory region on the receiver, can, theoretically, lead to higher bandwidth and lower latency, but, on the other hand, it will require some memory management mechanism to make sure that data is not lost when the memory region becomes full. Several approaches for the implementation of this mechanism for Software ROD will be investigated.

Finally, the existing demonstrator was working with bursts of data. But in a production environment data will

need to be streamed continuously from the FPGA board to the receiver(s). This will need to be investigated to see how the throughput varies depending on time, incoming data rate or trigger frequency, especially in the context of the issue described in the previous paragraph.

**Author:** OH, Alexander (University of Manchester (GB))

**Presenter:** VASILE, Matei (IFIN-HH (RO))

**Session Classification:** Tuesday posters session

**Track Classification:** Programmable Logic, Design and Verification Tools and Methods