Contribution ID: **145** Type: **Oral**

Improvements for the implementation of RDMA on FPGA devices

Wednesday 2 October 2024 15:50 (20 minutes)

RDMA communication can be a good solution for many communication use cases, such as in data acquisition systems and any other system requiring high bandwidth and low latency. Multiple options for an RDMA-based communication system have already been tested, such as profiling based on message size and message count, using multiple simultaneous clients for FPGA-based RDMA senders, or streaming data over RDMA links for software senders. Now, all of these are being put together in a system supporting streaming of data from a FPGA-based system over an RDMA link to one or more clients and new features are being added.

Summary (500 words)

Starting with LHC Run3, the FELIX (Front-End Link eXchange) system is used for implementing the data acquisition system for ATLAS subdetectors. It uses 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. 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 data rate increase that will come with 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 host PC by implementing RDMA support in the FPGA itself, thus removing the PC from the data path of the readout system.

The work presented until now demonstrated the ability to send data over an RDMA link, from an FPGA-based system using all the available bandwidth of a 100 Gbps link, either with individual data bursts from a single sender to a single receiver, or from a single sender to multiple simultaneous receivers. Subsequently, a streaming system from a software RDMA sender to one receiver or to multiple simultaneous receivers has been implemented. The potential of this technology was recognized also outside the project, in applications requiring high and reliable data streaming form FPGA to a computer infrastructure.

The FPGA sender and the streaming software sender are being combined to implement the functionality of sending a continuous stream of data from a FPGA-based sender to one or more receivers.

Previous testing has shown that sending a continuous stream of data from a single sender to a single receiver , the ability of that single receiver to handle all the received data on its own, on a single thread, is overwhelming. Further investigations revealed that this was caused by the limits of the system's memory bandwidth. When using multiple receiver processes on the same client device, this loss of total receive bandwidth was seen to be decreasing with increasing the number of receiver processes. A multi-threaded implementation for the receiver process may be able to replicate the same performance, that was seen when using multiple receiver processes, fully utilizing the available bandwith., this multi-threaded receiver was designed, implemented, tested and its performance evaluated.

Finally, a new data forwarding controller has been implemented for the FPGA sender. Until now, the data being sent over the RDMA links from the FPGA was simple testing data. Now, we are implementing a data-driven controller that manages buffering, partitioning and sending over RDMA of data coming from the front-end.

Author: VASILE, Matei (IFIN-HH (RO))

Co-authors: ULMAMEI, Andrei-Alexandru (Universitatea Nationala de Stiinta si Tehnologie Politehnica Bucuresti (RO)); Dr BIRA, Calin (Universitatea Nationala de Stiinta si Tehnologie Politehnica Bucuresti (RO)); VASILE, Costin-Emanuel (Universitatea Nationala de Stiinta si Tehnologie Politehnica Bucuresti (RO)); STOICEA, Gabriel (Horia Hulubei National Institute of Physics and Nuclear Engineering (RO)); BOUKADIDA, Nayib (Nikhef National institute for subatomic physics (NL)); Dr HOBINCU, Radu (Universitatea Nationala de Stiinta si Tehnologie Politehnica Bucuresti (RO)); MARTOIU, Sorin (Horia Hulubei National Institute of Physics and Nuclear Engineering (RO)); IANCU, Voichita (University Politehnica of Bucharest (RO))

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

Session Classification: Programmable Logic, Design and Verification Tools and Methods

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