Contribution ID: **114**   Type: **Oral**

APEnet+ 34 Gbps Data Transmission System and Custom Transmission Logic.

Thursday 26 September 2013 15:40 (25 minutes)

APEnet+ is a point-to-point, low-latency, 3D-torus network controller integrated in a PCIe Gen2 board based on Altera Stratix IV FPGA.
We characterize the transmission system (embedded transceivers driving external QSFP+ modules), analyzing signal integrity, throughput, latency, BER and jitter at different data-rate up to 34Gbps.
We estimate the efficiency of custom logic able to sustain 2.6 GB/s per link with a memory consumption of 40KB, guaranteeing deadlock free routing and systemic awareness of faults.
Finally, we show the preliminary results obtained with next-generation FPGA embedded transceivers and propose a new protocol to increase the performance with the same memory consumption.

Summary

In future particle and astroparticle physics experiments an increasing importance has been recently achieved by high speed data transfer for trigger and data acquisition systems. We present final results of characterization of our data transmission system based on FPGA embedded transceivers driving external QSFP+ modules and the custom control logic implementing fault-awareness capabilities free of any detrimental effect on the data transmission performance.

The APEnet+ project delivered a point-to-point, high performance, low latency, 3D torus network controller integrated in a PCIe Gen2 based board. The APEnet+ board exploits 32 8.5Gbps embedded transceivers of Altera Stratix IV devices, obtaining the impressive aggregated bandwidth of 400 Gbps per single device.

The QSFP+ (Quad Small Form Pluggable) standard (SFF-8436) is a technology intended for high-density and low-power applications and specifies a hot-pluggable transceiver with a bandwidth of 40 Gbps per direction. The APEnet+ card hosts 6 channels using the QSFP+ electrical and mechanical standard.

In order to produce the clearest signal and thus being able to increase signal clock frequency over the cable, Altera provides a Physical Medium Attachment (PMA). A fine tuning of the Equalization, Pre-emphasis, DC-Gain and Voltage Output Differential (VOD) is required.

Each transceiver implements an 8b/10b encoding to maintain the DC balance in the serial data transmitted and a byte ordering system at receiver side. Deskew logic and 128-bit word-level alignment is implemented to preserve data integrity along the four bonded lanes. Indeed, the 6 channels are bi-directional and can work simultaneously, reaching a data rate of 34 Gbps per direction.

We characterize all parts of our transmission system (transmission lines, connectors, cables) from the signal integrity point of view, then we characterize throughput, latency, bit error rate and jitter of the link at different data rates up to the maximum achievable data rate, with optical and electrical cables of different lengths.

The implemented control logic manages the data flow by encapsulating packets into a light, low-level, word-stuffing protocol able to detect transmission errors via CRC.

We develop a model of the transmission mechanism to estimate the efficiency of the control logic. The data transmission model is validated by comparison with actual performance achieved at different transceiver clock frequencies. We show the relation between the performance achieved with the adopted solution and the used memory resource. The current implementation of the data transmission system is able to sustain the link bandwidth of about 2.6 GB/s per link with a memory consumption limited to 40KB per link and guaranteeing a deadlock free routing with the adoption of virtual channels.

HPC systems in the peta/hexa-scale require techniques that aim at maintaining an acceptapble Failure-In-Time ratio. The diagnostic messages necessary to create a systemic awareness of fault and critical events are embedded in the APEnet+ transmission protocol to avoid performance degradation.

As conclusion, we show the preliminary results obtained with next-generation FPGA embedded transceivers of Altera Stratix V and propose a new data transmission protocol to increase the performance with the same memory consumption.

Primary authors: Dr LONARDO, Alessandro (INFN); BIAGIONI, Andrea (INFN); ROSSETTI, Davide (INFN); LO CICERO, Francesca (INFN); SIMULA, Francesco (INFN); TOSORATTO, Laura (INFN); FREZZA, Ottorino (INFN); PAOLUCCI, Pier Stanislao (INFN); Dr VICINI, Piero (INFN); AMMENDOLA, Roberto (INFN)

Presenter: BIAGIONI, Andrea (INFN)

Session Classification: Programmable logic, design tools and methods