FERMILAB-CONF-13-527-CMS-PPD

A Full Mesh ATCA-based General Purpose Data Processing Board

J. Olsen[a]*, T. Liu[a], and Y. Okumura[b]

[a] *Fermi National Accelerator Laboratory, Batavia, Illinois, USA*
[b] *University of Chicago, Chicago, Illinois, USA*
E-mail: jamieson@fnal.gov

ABSTRACT: High luminosity conditions at the LHC pose many unique challenges for potential silicon based track trigger systems. Among those challenges is data formatting, where hits from thousands of silicon modules must first be shared and organized into overlapping trigger towers. Other challenges exist for Level-1 track triggers, where many parallel data paths may be used for high speed time multiplexed data transfers. Communication between processing nodes requires high bandwidth, low latency, and flexible real time data sharing, for which a full mesh backplane is a natural fit. A custom full mesh enabled ATCA board called the Pulsar II has been designed with the goal of creating a scalable architecture abundant in flexible, non-blocking, high bandwidth board- to-board communication channels while keeping the design as simple as possible.

KEYWORDS: Trigger concepts and systems (hardware and software); Modular electronics; Data acquisition concepts.

*Corresponding author.

Operated by Fermi Research Alliance, LLC under Contract No. De-AC02-07CH11359 with the United States Department of Energy.

Contents

1. Introduction

The Pulsar II hardware design process started with the task of implementing Data Formatter system for the ATLAS Fast Tracker (FTK). This design process followed a bottom-up approach whereby we studied the input and output requirements and analyzed the data sharing between processing nodes. Various track trigger architectures and platforms were considered before settling on a hardware design which is a good fit for the Data Formatter application. Our baseline design also works well as a general purpose processor board in scalable systems where highly flexible, non-blocking, high bandwidth board to board communication is required.

1.1 ATLAS Fast Tracker Data Formatter

The ATLAS Fast Tracker [1] is organized as a set of parallel processor units within an array of 64 η-ϕ trigger towers. Due to the fact that the existing silicon tracker and front end readout electronics were not designed for triggering, the data sharing among trigger towers is quite complex. Our initial analysis showed that the data sharing between trigger towers is highly dependent upon upstream cabling and detector geometry. The ideal Data Formatter hardware platform must be flexible enough to accommodate future expansion and allow for changes in input cabling and module assignments.

Many different architectures were considered, including those based around full custom backplanes and discrete cables. In the end we determined that the full mesh Advanced Telecommunication Computing Architecture (ATCA) backplane was found to be a natural fit for the Data Formatter

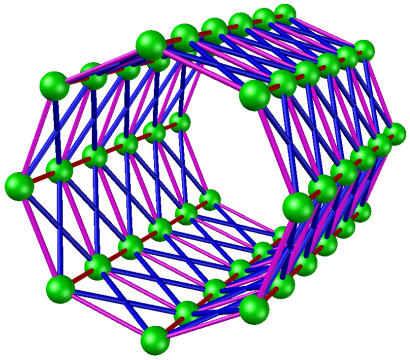

Figure 1. Conceptual view of a proposed CMS phase 2 Level-1 tracking trigger which consists of 48 towers ($6\eta \times 8\phi$). Trigger tower processor crates (shown in green) share data with immediate neighbors only.

design. The ATCA full mesh *Fabric Interface* enables high speed point-to-point communication between every slot, with no switching or blocking. Field Programmable Gate Array (FPGA) devices, which are abundant in local cells, memory, and high speed serial transceivers, were selected for the core processing element on each Data Formatter board [2] [3].

Unlike commercial CPU-based ATCA processors, the Pulsar II design avoids using a network switch and directly couples the FPGA serial transceivers to the backplane Fabric Interface. The direct connection between FPGA and fabric allows firmware designers to utilize low-overhead data transmission protocols which offer high bandwidth and deterministic transmission latency.

1.2 Applications Beyond the Data Formatter

The Data Formatter system is an application where the full mesh architecture is used to share data between directly processing nodes, thereby solving a physical or spacial problem of data duplication and sharing at trigger tower boundaries.

When one considers the many high bandwidth parallel data channels available in the full mesh it also becomes apparent that this architecture is uniquely positioned to support sophisticated and complex time multiplexed data transfer schemes.

An example of one such application is a proposed CMS phase 2 Level-1 track trigger, which consists of 48 tower processors as shown in Figure 1. Each tower processor crate hosts an array of independent track finder engines which are based on a pattern recognition associative memory devices. In this application the full mesh backplane is used to transfer time multiplexed event data from input boards to multiple track processing engines. Here the full mesh backplane is effectively used to blur the distinction between FPGAs and thus is used to support many different crate configurations. Currently we are investigating the performance and backplane channel bandwidth requirements for various track finder processor configurations [5].

The Pulsar II design forms the basic building block of a high performance scalable architecture, which may find applications beyond tracking triggers, and may serve as a starting point for future Level-1 silicon-based tracking trigger research and development.

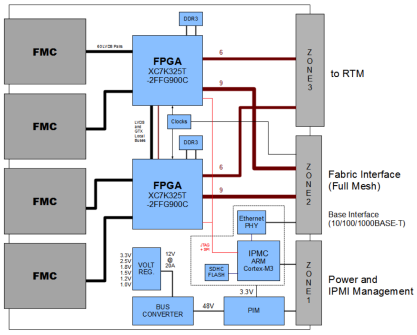

Figure 2. The Pulsar IIa block diagram.

Figure 3. The Pulsar IIa front board and RTM.

2. The Pulsar IIa Prototype

The Pulsar IIa consists of a front board and rear transition module, shown in Figure 3.

2.1 Front Board

Our first prototype board, called the Pulsar IIa, is designed around a pair of FPGAs, as shown in the block diagram in Figure 2. These FPGAs feature multiple high speed serial transceivers which are directly connected to the ATCA full mesh Fabric Interface and to pluggable transceivers on a rear transition module (RTM). The Xilinx Kintex-7 FPGAs we have selected for Pulsar IIa each have 16 10Gbps serial transceivers (GTX) and thus offer a subset of the full mesh backplane and RTM connectivity.

A Cortex-M3 microcontroller is used as an Intelligent Platform Management Controller (IPMC), which is required on all ATCA boards. This microcontroller is responsible for communicating with the ATCA shelf manager boards using the Intelligent Platform Management Interface (IPMI). Through this interface the dual redundant shelf manager boards monitor temperature and other various board sensors, and coordinate hot swap operations, and configure various board functions. In addition to the required IPMI functions, this microcontroller communicates over a secondary Ethernet network called the Base Interface. This network is primarily used for slow control functions such as downloading FPGA configuration images via FTP and providing a command line user interface through a Telnet server.

The ATCA specification was designed by the telecommunications industry and thus strong emphasis has been placed on reliability and high availability; the Pulsar II design embraces these ideas wholeheartedly by supporting hot swap capabilities and advanced telemetry and instrumentation designed into the power regulator subsystems.

2.2 Rear Transition Module

Eight four channel QSFP+ and six single channel SFP+ pluggable transceivers are located on the RTM. When fully loaded with SFP+ and QSFP+ modules the RTM will support an aggregate bandwidth of 380 Gbps. The Pulsar II RTM conforms to the PICMG3.8 standard and is considered an intelligent “field replaceable unit” (FRU) device. A small ARM microcontroller on the RTM

continuously monitors the status of the pluggable transceivers. This microcontroller also communicates with the front board IPMC and coordinates hot swap sequencing, sensor monitoring, and other hardware platform management functions.

Each of the Pulsar IIa FPGAs connects to one QSFP+ transceiver and two SFP+ transceivers on the RTM.

2.3 FMC Mezzanine Card

The Pulsar IIa supports up to four FMC mezzanine cards with the high pin count (HPC) LVDS interface. Mezzanine cards may contain FPGAs, pattern recognition ASICs, fiber optic transceivers, or any other custom hardware. We developed our FMC test mezzanine card in order to become familiar with the FMC form factor and to study high speed LVDS communication between FPGAs.

A test mezzanine card has been designed which features a Xilinx Kintex-7 XC7K160T FPGA, four SFP+ pluggable transceivers, 128MB DDR3 memory, and a 144 pin socket used for testing custom ASIC chips, primarily aimed at testing pattern recognition associative memory devices [4]. Per the VITA 57.1 specification the FMC mezzanines support loads up to 35W, which is supplied in on 12V and 3.3V power rails. An I2C bus and JTAG interface are also provided for slow controls and in-system programming.

3. Pulsar IIa Testing

3.1 Bench Top Testing

The first Pulsar IIa tests were performed on the bench top using a custom single slot "mini backplane" to provide 48VDC power to the front board and RTM. We then verified that the many voltage regulators on the board were quiet and within their allowable voltage range. Using the RJ45 Ethernet connection on the mini backplane we then connected successfully to the IPMC microcontroller and downloaded configuration images to the FPGA and read back various sensors through the Telnet interface.

Once the FPGA was configured we successfully completed various high speed tests involving the GTX transceivers. The mini backplane loops back all Fabric Interface channels so that the FPGA-PCB-connector signal path can be tested. RTM channels were also configured for loop back mode using passive copper SFP and QSFP cables and loopback adapters.

The Kintex-7 GTX transceivers have built-in diagnostic features which provide a mechanism to measure and visualize the receiver performance in real time using the ChipScope IBERT tool. The IBERT GUI allows designers to adjust various transceiver parameters such as pre- and post-emphasis, TX voltage swing, receiver equalization, sample point, and RX voltage offset. As the IBERT tool sweeps these various parameters it creates a 2D graphical depiction of the bit error rate as standard PRBS test patterns are sent over the link.

All GTX transceiver channels have been tested and characterized using the IBERT tool, and the results are shown in Table 1. Furthermore, the IBERT statistical "eye diagram" testing been performed on our Kintex-7 KC705 development board, which provides a "golden reference" for comparison studies. Comparing the Pulsar IIa eye diagrams against the reference design helps us learn more about high speed layout techniques, which will be used in the next iteration of the board.

Table 1. Pulsar IIa GTX Performance (PRBS-31).

	Line Rate	Bit Error Rate
Fabric Interface channels	6.25 Gbps	4.2×10^{-17}
RTM channels	6.25 Gbps	8.3×10^{-17}
Local Bus	10.0 Gbps	1.4×10^{-15}

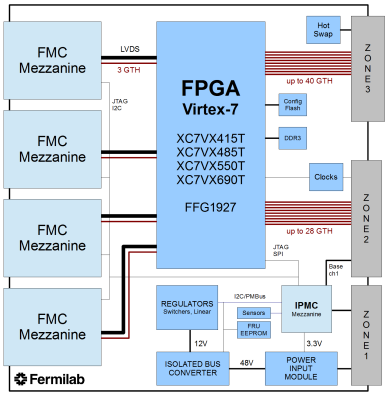

Figure 4. The Pulsar IIb block diagram.

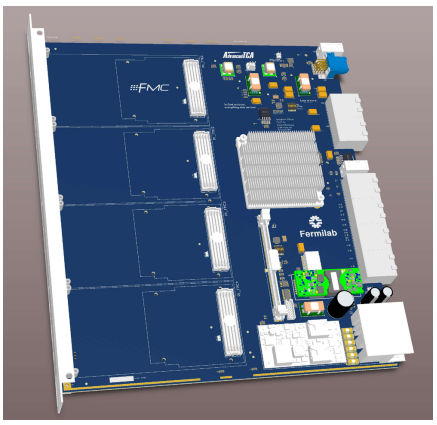

Figure 5. The Pulsar IIb board in layout.

Communication over the LVDS signals between the FMC mezzanine and the main FPGAs has been tested successfully at 400MHz single data rate (SDR) and 200MHz double data rate (DDR). Thirty-four LVDS pairs running at this speed yield a bandwidth of 13Gbps.

3.2 In-System Testing

Upon successful completion of our bench top tests we proceeded to install the Pulsar IIa boards and RTMs into our 14 slot full mesh ATCA shelf. The Pulsar IIa boards were installed in node slots (logical slots 3-10) and a commercial Ethernet switch was installed in slot 1. After logging into the Ethernet switch processor we were then able to Telnet into each Pulsar IIa board and initialize the FPGAs with "test sender" firmware. This firmware image is designed to transmit, receive and check data on the fabric, RTM and local bus GTX transceivers.

The Xilinx IBERT tool has also been used in the shelf to test GTX performance over the Fabric Interface. Technically our "10G" ATCA backplane is rated for only 3Gbps per lane. Despite this apparent speed limitation the Pulsar IIa has performed extremely well and no bit errors have been observed at rates at up to 6.25Gbps. Furthermore, there has been no significant signal degradation observed across the width of the backplane.

4. The Pulsar IIb

Leveraging the experience we gained through designing, building and testing the Pulsar IIa system we are in the final stages of laying out the next generation board, the Pulsar IIb (Figure 4 and Figure 5). The new board design replaces the two Kintex XC7K325T devices with a single Virtex-7 FPGA. The high speed serial transceiver (GTX/GTH) count has increased up to 80 channels,

providing a significant bandwidth increase to the RTM, Fabric and FMC mezzanine cards. The power regulator sections of the board have been redesigned to handle the increased power required by the Virtex-7 FPGA.

The ARM microcontroller, Ethernet PHY chip and other associated circuitry has been moved off the front board and into a small IPMC mezzanine module. The IPMC mezzanine is being developed at LAPP [6] with the goal of providing a modular, standard IPMI interface for ATCA boards in use at LHC experiments. Just as in the the Pulsar IIa, this IPMC will connect to the Ethernet Base Interface port and support FPGA firmware downloads and other non- IPMI user functions. Instrumentation on the Pulsar IIb has been significantly augmented; now more than 40 sensor channels, which include temperature, voltage, and regulator output current, are available to the shelf manager.

The Pulsar IIb boards will be used for the ATLAS FTK Data Formatter system. We anticipate that the boards will also be used for CMS L1 tracking trigger early technical demonstrations.

5. Conclusion

The Pulsar IIa is our first ATCA prototype board and works as designed, as demonstrated by our successful stand-alone and crate-level tests. Through this prototype development process we have gained experience using the latest Xilinx FPGAs and high speed serial transceivers to communicate over the ATCA full mesh backplane. Furthermore, the Pulsar IIa boards have successfully interfaced with other ATCA system components such as Ethernet switch blades and shelf manager cards.

The Pulsar IIb boards will be used in the ATLAS FTK Data Formatter system starting in 2015. The Pulsar IIb design forms the basic building block of a high performance scalable architecture, which may find applications beyond tracking triggers, and may serve as a starting point for future Level-1 silicon- based tracking trigger research and development for ATLAS and CMS.

Acknowledgments

The authors wish to thank Nicolas Letendre and Guy Perrot from LAPP for their work designing and documenting the IPMC mezzanine module. We are also grateful for the assistance provided by Fermilab Post-Docs Hang Yin, Matteo Cremonesi, and Zijun Xu for their work testing Pulsar IIa boards. Thanks to Andrew Rose for alerting us to suspiciously optimistic Virtex-7 FPGA power estimates.

References

[1] The ATLAS Collaboration, *Fast TracKer (FTK) Technical Design Report*, *CERN-LHCC-2013-007* **ATLAS-TDR-021-2013** (2013).

[2] Jamieson Olsen, Tiehui Ted Liu, Yasuyuki Okumura, *The Data Formatter Design Specification*, *Fermilab Preprint* **FERMILAB-TM-2553-E-PPD** (2013).

[3] Jamieson Olsen, et al., *A Data Formatter for the ATLAS Fast Tracker IEEE Real Time Conference Proceedings* **10.1109-RTC.2012.6418210**, (2012).

[4] Ted Liu et al., *A New Concept of Vertically Integrated Pattern Recognition Associative Memory Fermilab Preprint* **Fermilab-CONF-11-709-E** (2011).

[5] Ted Liu et al., *CMS Phase II Level-1 track trigger proposal Fermilab Preprint* **Fermilab-CONF-XXX** (2013).

[6] Letendre, N., *Development of an ATCA IPMI controller mezzanine board to be used in the ATCA developments for the ATLAS Liquid Argon upgrade*, *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)* **10.1109-NSSMIC.2011.6154412** (2011).