

CDF Level 2 Trigger Upgrade

K. Anikeev, M. Bogdan, R. Demaat, W. Fedorko, H. Frisch, K. Hahn, M. Hakala, P. Keener, Y. Kim, J. Kroll, S. Kwang, J. Lewis, C. Lin, T. Liu, F. Marjamaa, T. Mansikkala, C. Neu, S. Pitkanen, B. Reisert, V. Rusu, H. Sanders, M. Shochet, H. Stabenau, R. Van Berg, P. Wilson, D. Whiteson, P. Wittich

Abstract— We describe the new CDF Level 2 Trigger, which was commissioned during Spring 2005. The upgrade was necessitated by several factors that included increased bandwidth requirements, in view of the growing instantaneous luminosity of the Tevatron, and the need for a more robust system, since the older system was reaching the limits of maintainability. The challenges in designing the new system were interfacing with many different upstream detector subsystems, processing larger volumes of data at higher speed, and minimizing the impact on running the CDF experiment during the system commissioning phase. To meet these challenges, the new system was designed around a general purpose motherboard, the PULSAR, which is instrumented with powerful FPGAs and modern SRAMs, and which uses mezzanine cards to interface with upstream detector components and an industry standard data link (S-LINK) within the system.

Index Terms—FPGA based hardware, trigger, real-time systems.

I. OVERVIEW OF CDF TRIGGER SYSTEM

The CDF trigger [1] is a three level system. The Level 1 and Level 2 triggers, shown in Fig. 1, use custom-designed hardware to find physics objects in a subset of the event information. The Level 3 trigger uses the full set of information and the latest calibrations for complete event reconstruction in a farm of x86 PCs. The goal of the trigger is to retain the most interesting events for physics analysis while respecting the bandwidth limitations of the CDF data acquisition system. Each stage in the trigger must reject a sufficient fraction of the events to allow the next stage to process the accepted events with minimal dead time.

The Level 1 system has a synchronous 40 stage pipeline. When an event is accepted by the Level 1 trigger, all data are moved to one of the four Level 2 data buffers (in the front end electronics), while a predefined subset of these data is sent to the asynchronous Level 2 trigger. Here, some additional, but still limited, event reconstruction is performed and a Level 2 decision is evaluated. Level 2 has at its disposal all trigger objects used in Level 1, such as tracks from the eXtremely Fast Track trigger (XFT/XTRP), muon primitives, and global

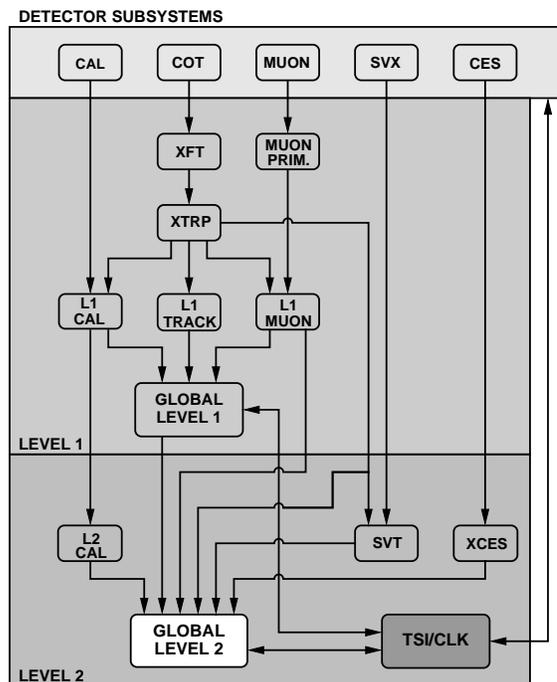


Fig. 1. The CDF Trigger architecture. Only the first two levels are shown. “Global Level 2” box represents the system described in this paper.

energy information, as well as the complete Level 1 trigger decision information. In addition, the ShowerMax (CES/XCES) information for electron/photon identification and objects found in two other dedicated Level 2 sub-systems, the Secondary Vertex Tracker (SVT) and the Level 2 Calorimeter (L2CAL), are available.

Dead time arises when an event is accepted by the Level 1 trigger while all four Level 2 data buffers are occupied. Thus minimizing both loading and processing times is critical. Loading time refers to how long it takes to deliver data to the Level 2 decision CPU counting from the Level 1 Accept. Processing time refers to the time it takes to unpack the data, to form objects and to make a Level 2 decision based on the results of executing an algorithm that evaluates these physics objects.

The original Level 2 trigger was designed and built in the mid to late 1990’s based on technology available at that time. The design relied on a custom bus (Magic Bus), a now obsolete processor (DEC Alpha) on a custom board, and a set of different custom interface boards. The system was able to handle a

Manuscript received on November 17, 2004; revised July 22, 2005

Project contact person: T. Liu, thliu@fnal.gov,

<http://hep.uchicago.edu/~thliu/projects/Pulsar/>

M. Bogdan, Y. Kim, W. Fedorko, H. Frisch, S. Kwang, V. Rusu, H. Sanders, M. Shochet, University of Chicago, Chicago, IL 60637, USA

K. Anikeev, R. Demaat, M. Hakala, J. Lewis, C. Lin, T. Liu, T. Mansikkala, F. Marjamaa, S. Pitkanen, B. Reisert, P. Wilson, FNAL, P.O. Box 500, Batavia, IL 60510, USA

K. Hahn, P. Keener, J. Kroll, C. Neu, H. Stabenau, R. Van Berg, D. Whiteson, P. Wittich, University of Pennsylvania, Philadelphia, PA 19104, USA

Level 1 accept rate up to 25 kHz with a Level 2 accept rate around 300 Hz.

The strategy we chose for the new system was to convert and pre-process all trigger data fragments from upstream into a self-describing format using a universal interface board. Multiple copies of the board running different types of firmware are employed to process data from all of the upstream systems. After initial processing, the data are merged (again, using the same type of board) and transferred via an industry standard data link (*e.g.*, S-LINK or Gigabit Ethernet) into a CPU for decision making. The new system was designed to be able to handle a Level 1 Accept rate above 30 kHz and to be able to produce Level 2 Accept rate near 750 Hz.

A detailed description of the new Level 2 trigger is presented below. Some additional information can be found in [2] and references therein.

II. CHALLENGES FOR THE LEVEL 2 TRIGGER UPGRADE

The second portion of Run II of the Tevatron, known as Run IIb, is going to be marked by a multi-fold increase in the instantaneous luminosity over what used to be the norm during Run IIa. The increase will be achieved at the 396 ns bunch spacing rather than previously planned 132 ns. At the expected peak Run IIb luminosity of $3 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$, we shall see 10 interactions per proton anti-proton bunch crossing. This large number of interactions will lead to much larger data size and combinatorics (for multi-object triggers) per event. The increase in latency and data size upstream of the Level 2 system must be compensated for by an increase in effective bandwidth to transfer the trigger data from the Level 2 input into the decision CPU memory. With the increase in data size, the amount of processing increases, especially for multi-object triggers. A higher demand in terms of processing must be offset by increased CPU power. Given the situation above, a seemingly modest increase of 5 kHz in the allowed input rate of the new system over the old one is in fact a serious achievement if one takes into account that the event size is substantially larger at the instantaneous luminosities the new Level 2 trigger is designed to operate.

Another challenge of this upgrade lies in the diversity of the interfaces involved (hardware as well as protocol) and the desire to handle them all using a single board type. CDF uses different types of ribbon cables and optical fiber links for data interfaces at Level 2. For example, different types of cables and data protocols are used to transmit the Level 1 trigger decision, global energy information, and tracks from the XFT/XTRP and the SVT. Interfaces with the Muon, the Level 2 Calorimeter and the ShowerMax sub-systems are implemented using different types of optical fiber links. In addition, the board has to be able to interface with the Trigger Supervisor (TSI) as well as Level 2 decision CPU(s).

Yet another challenge was to commission the new system with minimal impact on the running experiment.

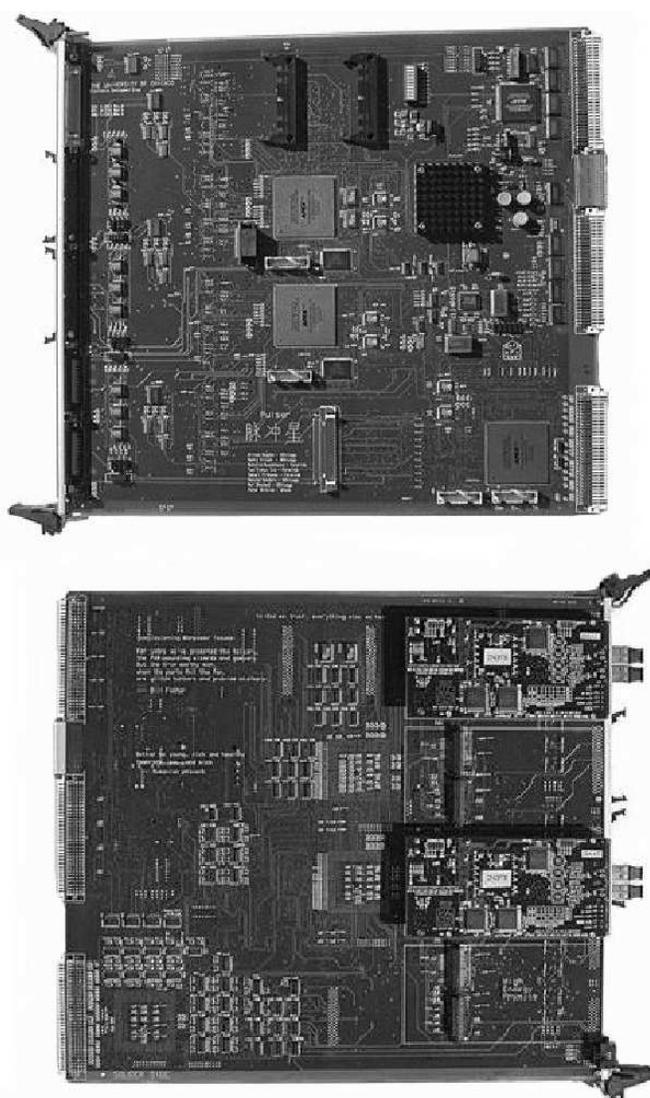


Fig. 2. Top and bottom view of the PULSAR board.

III. PULSAR BOARD DESIGN

In order to meet the challenges, the design philosophy of the upgrade was to use one kind of general purpose motherboard, with powerful FPGAs and SRAMs, and to interface any custom data link with an industry standard link through the use of mezzanine cards. The design asserted that the motherboard could be configured either as a data sink or data source, thus the name PULSAR (PULSer And Recorder). This feature makes testing an individual board or the entire system easier as one does not need boards of any other kind.

Fig. 2 shows an actual PULSAR board, while Fig. 3 shows its design. The PULSAR is a general purpose 9U VME board. It implements all the interfaces to the Level 2 trigger system. There are three different types of dedicated LVDS connections at the front of the board, which are specific to the application at CDF: two 34-bit wide cable connectors for Level 1 trigger information and global energy sums, a 25-bit wide cable

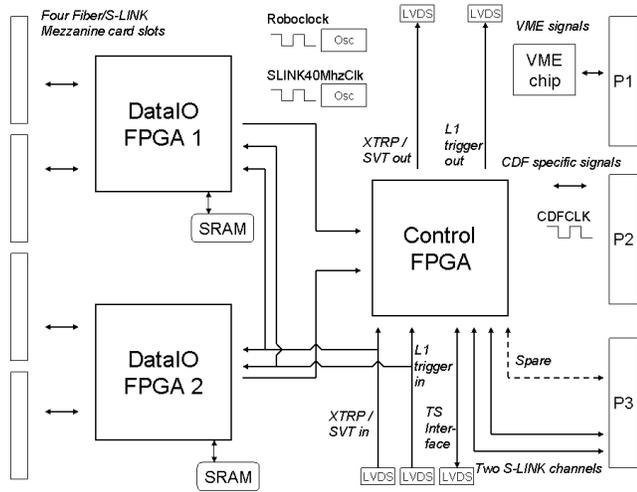


Fig. 3. PULSAR board design.

connector for track information from the XFT/XTRP as well as from the SVT, and a 11-bit wide connection interfacing with the TSI. The data transfers of the remaining CDF sub-systems providing data to the Level 2 trigger are implemented via optical fiber links. The interfaces to the optical fibers are absorbed in two types of custom mezzanine cards (Hotlink [3] and Taxi [4]).

The key devices on the PULSAR board are three FPGAs (Altera APEX 20K400BC-652-1XV [5]): two DataIO FPGAs and one Control FPGA. Each DataIO FPGA is coupled to a 128K×36 synchronous-pipelined Burst SRAM equipped with advanced No Bus Latency™ (NoBL™) logic. The maximum access delay from the clock rise is 4.0 ns [6]. Both DataIO FPGAs provide interfaces to two mezzanine cards each. The mezzanine card connections are bidirectional, *i.e.*, one can plug in either transmitter or receiver cards. The implementation is similar to the CMC (Common Mezzanine Card) standard [7] and the actual design followed S-LINK64 specifications [8]; thus custom designed Hotlink and Taxi mezzanine cards as well as commercially available S-LINK mezzanine cards can be mounted on the motherboard. Each mezzanine card slot at the front of the board (on the bottom side) has up to 83 user defined signals directly accessible by the motherboard FPGAs. The PULSAR has a user defined interface to the P3 connector with up to 117 signals directly interfacing with the Control FPGA. This allows users to define which custom or standard link to interface with on the transition module on the back of the VME crate. The board also has a user defined interface to the P2 connector with up to 50 signals visible to all three main FPGAs via buffer chips. The interfaces to both P3 and P2 are all bidirectional. The board was designed to be programmable via JTAG as well as through the VME bus.

In the CDF Level 2 trigger application, the PULSAR is used as a universal interface board to convert (*e.g.*, perform data compression via zero suppression or select information relevant for Level 2 decision making) and merge many different

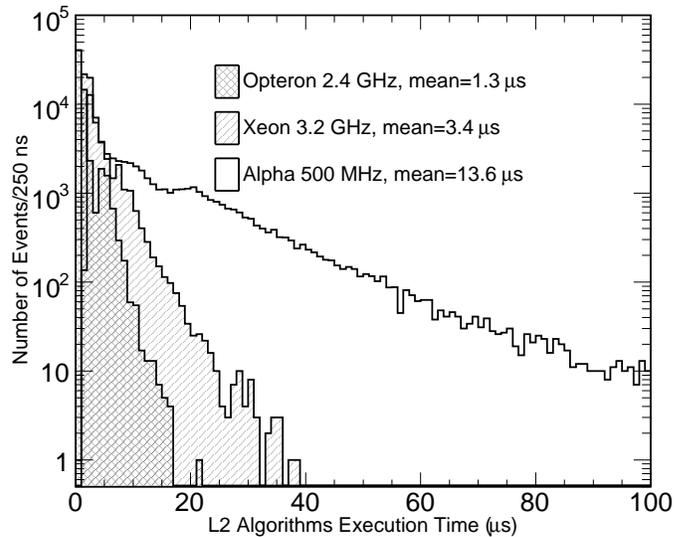


Fig. 4. Comparison of the event processing time of the Run IIa processor, DEC Alpha, and two choices for the Run IIb processor, Intel Xeon and AMD Opteron.

trigger data into an S-LINK standard packet. Although the PULSAR has been specifically designed for its application within CDF, the use of standardized mezzanine card connectors should provide sufficient flexibility for applications of PULSAR outside CDF as well.

A significant fraction of the design effort was dedicated to extensive verifications by using state-of-the-art CAD tools. The tools used for FPGA firmware development and gate level simulation were Leonardo Spectrum for VHDL synthesis and Quartus II for placing and routing of logic arrays. Mentor Graphics QuickSim II using Smart Models together with netlist files created by Quartus II was used for board and multi-board level simulation. In addition, the Interconnect Synthesis tool was used for trace and cross talk analysis to check signal integrity. The IS MultiBoard tool was used for signal integrity checks between the motherboard and mezzanine cards. These sophisticated tools helped to streamline the design process significantly.

The prototype boards were tested with on-board clock speeds up to 100 MHz. No problems were found. No layout or fabrication errors were found on the prototype boards, allowing them to serve as production boards. Furthermore, the firmware design tools made it possible to gain confidence that the chosen FPGA provides sufficient logic and storage resources, and meets the speed requirements for its application as data pre-processor and merger, before starting the production of the boards. The throughput of the PULSAR board with two S-LINK mezzanine cards can reach 2×160 MB/s, thus the performance of the Level 2 trigger is limited by the arrival time of the data from the Level 2 sub-systems.

IV. DECISION NODE AND COMMUNICATION WITHIN THE LEVEL 2 TRIGGER

To reconstruct events and make trigger decisions the Run IIa system utilized a 500MHz DEC Alpha processor on a custom designed processor board, which was physically located in the same crate as the other Level 2 trigger boards. Our choice for the decision node was a commodity dual-processor PC, external to the PULSAR crate. Historically, to guarantee performance, real-time operating systems have been required. The standard 2.6 Linux kernel, however, provides a mechanism to schedule processes with real time priority. In addition, it provides the means to bind peripheral interrupts to a specific CPU in a multiprocessor environment; we use this feature to leave the second CPU free to process data and make trigger decisions. Together, these features allow operational performance which approaches that of a real time system.

In order to investigate what kind of a PC best fits the purpose (memory and PCI bus architectures are of importance), we compared the performance of the DEC Alpha processor to an Intel XEON 3.2GHz processor and an AMD Opteron 2.4GHz processor running the trigger algorithms on real events. To isolate the CPU requirements, we removed data transmission delays from the timing measurements. Fig. 4 shows a marked decrease in the mean processing time and a sharp reduction in the long tail. The differences between XEON and Opteron are to a large extent due to differences in the memory bus architecture of the two processors. For the Level 2 application the Opteron 2.4 GHz running in 64-bit applications was chosen.

The communication between PULSAR boards and the decision CPU is done via commercially available, high bandwidth and low latency S-LINK-to-PCI interface cards S32PCI64 [9]. The S32PCI64 is designed to have low PCI bus utilization and needs minimal host processor control. The S32PCI64 cards are used both in receiver mode (to send data into the CPU) and in transmitter mode (to send the Level 2 decision back to the PULSAR crate).

The transmission latency, which is comprised of the times needed to transport the data from the upstream systems to the PULSAR crate, from the PULSAR crate to the decision node and communicate decision back to the PULSAR crate, was evaluated in a controlled environment. When no data processing or event reconstruction were performed, tests showed a total time of about $8 \mu s$. When in addition the data were unpacked and the trigger algorithms run, the total time was well behaved, in agreement with the expectations from the studies of the processing time illustrated by Fig. 4.

V. SYSTEM CONFIGURATION FOR CDF LEVEL 2 UPGRADE

The PULSAR system configuration for the initial phase of the CDF Level 2 trigger upgrade is shown in Fig. 5. Six PULSAR boards (those labeled “Rx”) act as sinks for all data paths upstream:

- “Muon Rx” receives muon information, tracks from XFT/XTRP and Level 1 trigger bits;

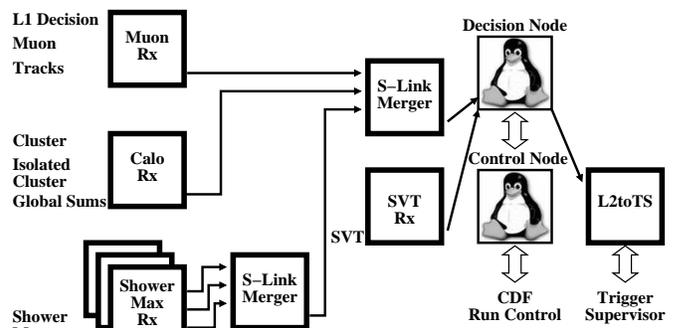


Fig. 5. CDF Level 2 Trigger upgrade system configuration.

- “Calo Rx” receives global energy sums from L1CAL and information on energy clusters as well as isolated clusters from L2CAL;
- three “ShowerMax Rx” boards receive information from ShowerMax (CES/XCES);
- “SVT Rx” receives track information from the SVT sub-system.

The S-LINK formatted output of the triplet of the ShowerMax receivers is merged using one S-LINK merger PULSAR. Its output is then merged with the outputs of the Muon and Calorimeter receivers in the final S-LINK merger. The output of this merger PULSAR is fed to the decision node via one S32PCI64 card. The SVT data arrives via a separate path because it has the longest latency; after being converted into S-LINK format it goes directly to the decision node via another S32PCI64 card. The third S32PCI64 card in the decision node is used to send the trigger information (including the decision) to yet another PULSAR (labeled “L2toTS”), which communicates the decision to the TSI.

Current system configuration employs nine PULSAR boards, many of which carry different mezzanine cards and FPGA firmware, for a total of six different types. In addition to the decision node, the system includes another PC (labeled “Control Node” in Fig. 5) which handles the task of communicating between the decision node and the rest of the CDF DAQ. A copy of the system (including both PCs) is maintained as a spare as well as for development purposes.

VI. COMMISSIONING STRATEGY AND INITIAL EXPERIENCE

The universality of the PULSAR board design allows us to test each data path, hardware as well as firmware, in a test stand using additional PULSARs configured in transmitter mode. As described in the previous section, there are six different types of PULSAR boards in the CDF Level 2 trigger. Correspondingly, there are six different types of transmitter PULSAR boards used in the test stand configuration. All hardware and firmware were tested extensively in this controlled environment before integrating the new system into the data taking.

In order to minimize the impact on the operation of the CDF experiment during the system commissioning phase of the Level 2 upgrade, all input data paths have been split so that

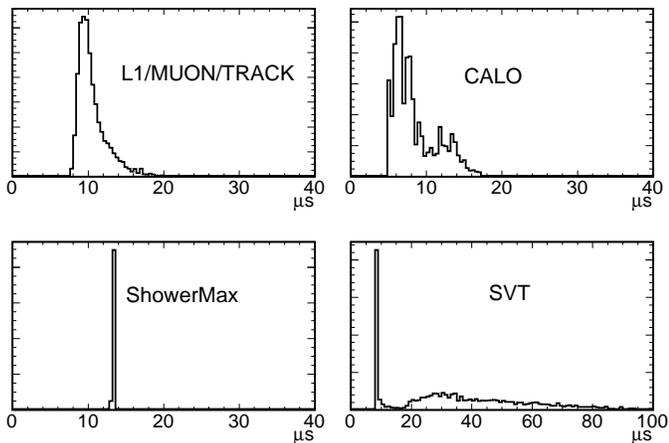


Fig. 6. The arrival time with respect to the Level 1 Accept for all the Level 2 input paths.

a copy of the input data is made available to the new system, while the original system continues to drive the data acquisition.

The initial system commissioning work has been done using cosmic rays and other non-beam trigger configurations. Subsequently, the new system was tested with beam in pure parasitic mode, *i.e.*, the TSI would not wait for, and would not listen to, decisions made by the new system. The system has been tested extensively using this methodology before we requested dedicated beam time to allow the new Level 2 system to drive the data acquisition. The PULSAR based Level 2 trigger worked on the first attempt in the initial commissioning test run with dedicated beam time. For all Level 2 trigger algorithms the trigger decisions from the upgrade system matched perfectly to those expected from the original system.

For all PULSAR boards used in the system, diagnostic DAQ buffers have been implemented, allowing us to readout the intermediate information (data as well as timing information) into the data stream. The information present in the DAQ buffers can be saved into the event data structure for offline analysis. This design feature is essential for commissioning, optimizing, as well as long term maintenance of the system.

Using this functionality, an important measurement, that of the inherent latency of all Level 2 sub-systems, has been done (see Fig. 6) early on in the commissioning effort. The ShowerMax path dictates the minimum latency of the system, with data arriving at fixed time after the Level 1 Accept. The tails of the Level 2 latency are driven by the SVT path. These and similar measurements at different instantaneous luminosities have helped to choose the optimal design for the new Level 2 trigger and identified areas that need improvement. An example of the latter is the SVT sub-system, which is currently being upgraded for faster processing. The SVT upgrade team chose to implement a large fraction of the SVT functionality using PULSAR boards.

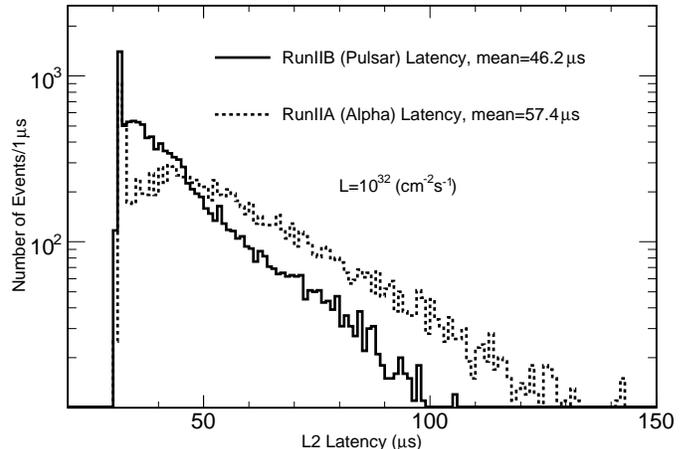


Fig. 7. Global Level 2 latency, from Level 1 Accept to broadcast of Level 2 decision, for the legacy system and new PULSAR-based system.

VII. PERFORMANCE AND RELIABILITY

The overall performance of the upgraded Level 2 trigger has been measured using data collected with the CDF detector. During the short commissioning phase we have been able to record the latency for the old and the new system at the same time and compare them on an event by event basis. The latency was defined as the time between the Level 1 Accept and the broadcast of the Level 2 decision. The same measurement has been repeated for different luminosities in order to map out the dynamical behavior of the system. These measurements have had almost no impact on the routine operation of the CDF detector. Fig. 7 shows a comparison between the two systems at one typical luminosity of $1 \times 10^{32} \text{ cm}^{-2} \text{ s}^{-1}$. The new system improves the mean latency by about $11 \mu\text{s}$, which translates into a more than 20% increase in the total bandwidth allowed for the Level 1 trigger. This represents a significant increase in the physics capabilities of the CDF detector, especially at higher instantaneous luminosities.

The new Level 2 trigger for the CDF experiment has been running without any problem since late March 2005. After a month of serving as a hot spare, the old system was decommissioned.

VIII. FUTURE IMPROVEMENTS

The Level 2 trigger performance can be improved in various ways. For example, at board level, the data volume can be suppressed further for some data paths, and the timing of the firmware can be improved.

At the system level, one promising upgrade option is to use a CERN Four Input Links for Atlas Readout (FILAR) card [10] instead of the S32PCI64. The FILAR PCI interface is based on the design of the S32PCI64, but can move data from up to four S-LINK channels. Therefore, we eliminate the need for one PULSAR S-LINK merger, thus allowing all data fragments to be sent directly to the CPU memory via PCI bus. Using FILAR also makes it possible to run the S-LINK mezzanine cards at

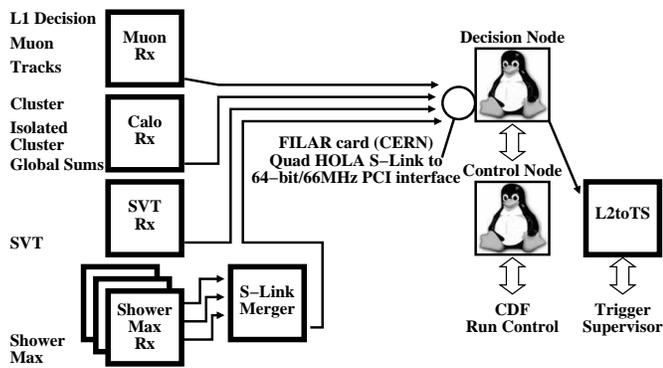


Fig. 8. Possible future system configuration of CDF Level 2 trigger upgrade using S-LINK FILAR cards.

higher speed. In addition, FILAR has less PCI overhead relative to S32PCI64. Fig. 8 shows one possible system configuration using FILAR.

Further improvements may be achieved by using four Level 2 decision nodes, each dedicated to a given Level 2 buffer event. This is possible since PULSAR has two S-LINK channels over P3, the potential gain here is that the processing of a given event does not have to wait for the previous event processing to be finished.

IX. SUMMARY

We have presented the upgrade of the CDF Level 2 trigger. The design of the new system departs significantly from the previous implementation. It makes use of PULSAR, a general purpose 9U VME interface board developed for HEP applications, and an easily upgradeable commodity CPU to run decision algorithms. The new system is designed to have a safety margin in performance and flexibility to meet the Run IIb trigger challenges, and to use built-in test capabilities to speed up the commissioning process and to ease the long term maintenance effort. The upgrade of the CDF Level 2 trigger is a project where the S-LINK technology developed at CERN for the LHC experiments is used for the first time in a high rate hadron collider environment. Knowledge gained by using S-LINK at CDF is transferable back to the LHC community.

REFERENCES

- [1] R. Blair *et al.* (CDF II Collaboration), The CDF Run II Detector Technical Design Report, 1996. FERMILAB-Pub-96/390-E.
- [2] M. Bogdan *et al.*, "CDF Level 2 Trigger Upgrade: the Pulsar Project", Proceedings of the 10th Workshop on Electronics for LHC and Future Experiments, Boston, 2004.
- [3] Cypress Semiconductor Corporation, "HOTLink™ Transmitter/Receiver (CY7B923/CY7B933)," Data Sheet, 1999.
- [4] TAXIchip™ Integrated Circuits, "Transparent Asynchronous Transmitter/Receiver Interface Am7968/Am7969-125 Am7968/Am7969-175," Data Sheet and Technical Manual, 1994.
- [5] ALTERA Pub., "APEX 20K Programmable Logic Device Family," Data Sheet v.5.1, 2004.
- [6] Cypress Semiconductor Corporation, "CY7C1350, 128K×36 Pipelined SRAM with NoBL™ Architecture," Data Sheet, 1999.
- [7] "IEEE Standard for a Common Mezzanine Card Family: CMC," Document 1386-2001, ANSI Pub., 2001.

- [8] E. van der Bij *et al.* (1997, Sep.). "S-Link, a Data Link Interface Specification for the LHC Era." Presented at 10th IEEE Real Time Conference. [Online]. Available: <http://hsi.web.cern.ch/HSI/s-link/>
- [9] W. Iwanski *et al.* (2001, Aug.). "Designing an S-LINK to PCI Interface using an IP core." Presented at 12th IEEE-NPSS Real Time Conference. [Online]. Available: <http://hsi.web.cern.ch/HSI/s-link/devices/s32pci64/>
- [10] S. Haas. (2004, Sep.). "Design and Performance of a PCI Interface with four 2 Gbit/s Serial Optical Links." Presented at 10th Workshop on Electronics for LHC and Future Experiments. [Online]. Available: <http://agenda.cern.ch/fullAgenda.php?ida=a043274>