Fermilab FERMILAB-Conf-03/226-E August 2003

The BTeV Trigger Architecture

Michael H.L.S. Wang

For the BTeV Collaboration

Fermi National Accelerator Laboratory, Batavia, IL 60510, USA

Abstract

BTeV is a high-statistics B-physics experiment that will achieve new levels of sensitivity in testing the Standard Model explanation of CP violation, mixing, and rare decays in the b and c quark systems by operating in the unique environment of a hadron collider. In order to achieve its goals, it will make use of a state-of-the-art Si-pixel vertex detector and a novel 3-level hierarchical trigger that will look at every single beam crossing to detect the presence of heavy quark decays. This talk will describe the trigger architecture focusing on key design aspects that allow the use of commercially available technology in a highly feasible and practical solution that meets the demanding physics requirements of the BTeV experiment.

1 Introduction

BTeV is a collider B-experiment that will begin taking data in 2009 at the C0 interaction region of the Fermilab Tevatron. In addition to the large $b\overline{b}$ cross sections at hadron colliders ($\sigma(b\overline{b}) \sim 100$ μb), it will exploit the unique characteristics of hadronic b production in the forward region which includes the large production cross sections for highly correlated $B\overline{B}$ pairs and the large boost received by B mesons that result in longer observed decay lengths and reduced effects due to multiple scattering [1,2].

To take advantage of these features, it will employ a unique forward spectrometer consisting of a muon detector, a Ring Imaging Cherenkov detector, and an electromagnetic calorimeter for particle ID, and a combination of straw tubes and Si-microstrips for charged particle tracking [2]. The real strength of the BTeV detector, however, is the 30 station Si-pixel vertex detector spanning $\sim$120 cm, centered at the C0 collision point, and immersed in the 1.5 Tesla dipole field of the SM3 magnet [3]. Each pixel station has over 7.6×10^5 rectangular pixels measuring 50×400 μm^2 distributed over bend and non-bend

views for a total of over 22×10^6 pixels in the full detector. The long dimension of pixels in the bend view are oriented parallel to the dipole field while those in the non-bend view are perpendicular to the field.

In order to deal with the rare occurence of a b event (about 1/1000 $p\overline{p}$ collisions at the Tevatron) and the high track multiplicities in these events, BTeV will couple the superior pattern recognition capability of the Si-pixel detector with a unique online trigger system to acquire data at the full crossing [1] rate of 7.6 MHz. With an estimated event size of 100 KB, data will come out of the detector front-end at an extremely high rate of 800 GB/sec. To handle this, BTeV will employ a three-level hierarchical trigger architecture which will be described in some detail below.

2 Level-1 Trigger Algorithm

We begin our discussion of the trigger architecture by briefly describing the two stage algorithm employed in the first level trigger (L1) [4–6]. The first stage, referred to as the segment finding stage, searches for track segments using hit clusters spanning three adjacent pixel stations. Its search is confined to two separate regions of the pixel detector, an inner region close to the beam axis and an outer region close to the edge of the pixel planes. Segments found in the inner region are called inner triplets which represent the beginning of a track. Segments found in the outer region are called outer triplets which represent the end of a track just before exiting the pixel detector volume.

The second stage consists of two phases, a track finding phase followed by a vertex finding phase. The track finding phase projects the inner triplets downstream based on their slopes in the non-bend plane and their "curvature" (slope change) in the bend plane. Outer triplets within a certain distance of the projected trajectories are matched to the inner triplets to form complete tracks. The vertex finding phase uses these found tracks to locate the primary interaction vertices. Remaining tracks not belonging to these primary interaction vertices are then tested for their detachment from these vertices to see if they belong to candidate B-decay products. A L1 trigger accept is generated if there are at least 2 tracks in the instrumented arm of the BTeV detector meeting the following criteria: $p_T^2 \geq 0.25$ (GeV/c)^2, $b/\sigma \geq n$, and $b \leq 2$ mm where b is the impact parameter, σ is the error in b, and n is a preset value ranging from 4-7.

[1] In this paper, the terms crossing and event will be used interchangeably

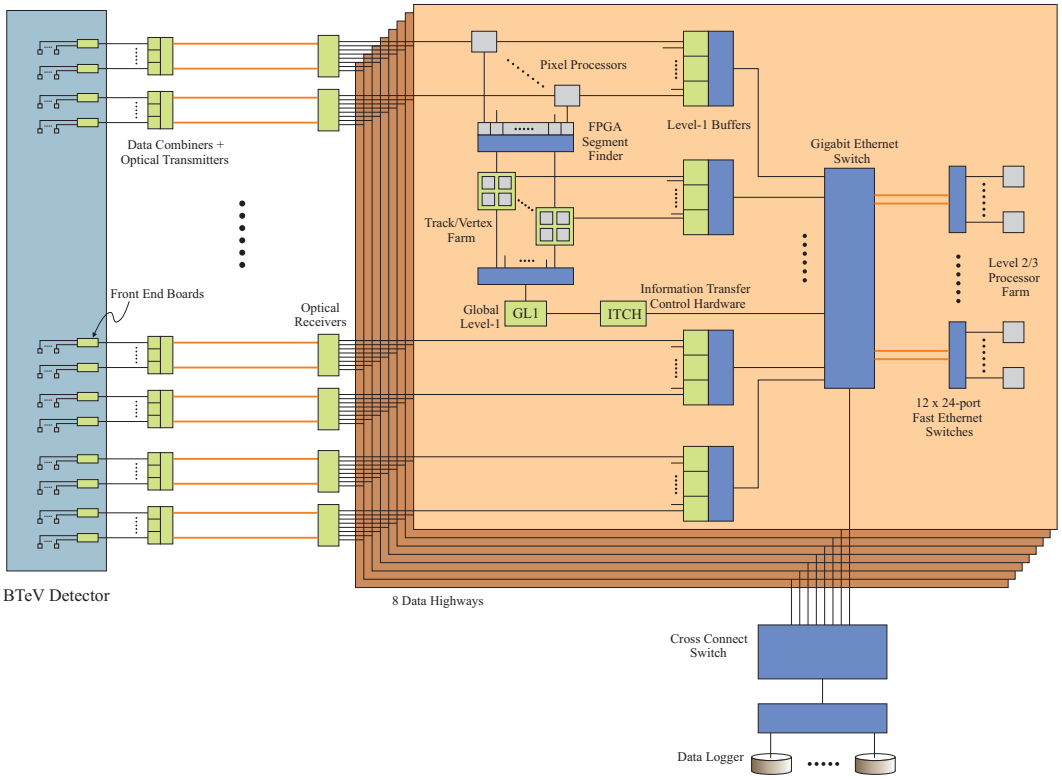

Fig. 1. BTeV Three-Level Eightfold-way Trigger Architecture

3 Trigger Architecture

The three-level trigger system shown in Fig. 1 consists of eight parallel data pathways called "highways". Data from each beam crossing is distributed uniformly to one of these eight highways, each of which forms a complete and independent three-level trigger system. This reduces the full data rate of 800 GB/s from the detector to 100 GB/s into each highway allowing the use of cheap components such as commercially available ethernet switches.

Starting from the left hand side of Fig. 1, digitized data packets from the front-end boards are sent via high-speed copper links to data-combiner boards (DCB) located in the collision hall [7]. DCB's merge multiple input streams to fewer output streams by combining smaller packets from many input channels into larger packets. Each DCB interfaces with eight optical transmitters, each connected to one of the eight highways. Data packets from the output stream for one crossing are sent by one of these transmitters via optical fibers over a distance of about 30 m from the collision hall to the counting room. Optical receivers in the counting room then route the data packets to the appropriate highway. Three data-combiner boards are grouped together to form a DCB subsystem.

For each highway, data from all detector components is sent from the optical

receivers to L1 buffers for temporary storage as trigger decisions are made. An exception is the vertex detector whose data goes through the pixel-processors before being stored in L1 buffers. Data from the pixel processors are also sent to an FPGA based segment tracker that executes the segment finding stage of the algorithm. Inner and outer triplets belonging to the same crossing are then routed by a network switch from the segment tracker to one node in a farm of over 300 programmable embedded processors that execute a C version of the track and vertex finding stage of the algorithm. For each node, complete processed results are routed to the L1 buffers while summarized trigger results are sent to a Global Level-1 (GL1) processor responsible for the ultimate trigger decision. Crossings accepted by GL1 are then maintained as a list in the Information Transfer Control Hardware (ITCH) which also broadcasts accept messages to all L1 buffers indicating which crossings are to be saved.

The basic building blocks of the L1 buffers are modules consisting of commercial DRAM configured as circular buffers [7]. Three modules are grouped into an L1 buffer subsystem for a total of 32 subsystems in each highway. Each of these subsystems has 24 input buffers serving 8 DCB subsystems in the collision hall. A Level-1 accept concatenates the data from all 24 input buffers copying them to an output buffer where it remains until transferred to a Level-2/3 (L2/3) node. The use of low-cost DRAM allows enough memory to buffer $\sim$300 thousand crossings in each highway corresponding to $\sim$300 ms of L1 trigger decision time (since the crossing time of 132 ns is increased $8\times$ to 1 μs per highway). This is nearly three orders of magnitude more than the average L1 processing time of 330 μs. The use of large buffers with circular access is far more cost effective than smaller ones employing sophisticated memory management. It also allows the system to handle the long processing times of events in the tail of the L1 time distribution. Each of the 32 L1 buffer subsystems feeds a port in a highway switch consisting of a commercial 64-port gigabit ethernet switch. 12 pairs of ports on this switch feed the paired gigabit uplink ports of a dozen 24-port fast-ethernet fanout switches. These fanouts, in turn, feed data to nearly 300 commodity Linux-PC's that make up the L2/3 processor farm in each highway where the DRAM in each of the nodes functions as a Level-2/3 buffer.

After receiving a request from an idle node in the L2/3 processor farm for an event, the ITCH responds by assigning an accepted crossing number to that node. Once it receives its assignment, the L2/3 node sends a request to a subset of the L1 buffers which respond by sending their data to that node. All requests and data transfers between the L2/3 farm and the L1 buffers and the ITCH are routed through the highway and fanout switches. Upon receiving the data, the L2/3 node executes the L2 trigger algorithm which now has the option of using additional information from the first few stations of the straw and Si-microstrip trackers in doing a more refined analysis of the

event. If the event passes L2, data from the rest of the L1 buffers is transferred to the same node to execute L3. At this stage, the processing node has, at its disposal, particle ID information in addition to that from the rest of the forward tracking stations to further improve upon the L2 results.

In practice, each L2/3 node issues multiple event requests, storing data for ∼16-32 events in the L2/3 buffers at any given instant. When a Level-2/3 processor completes executing the L2 trigger algorithm on an event, it performs a context switch to process one of the other events in the buffer while new data is requested to fill the buffers. Processing of events that pass L2 is temporarily suspended while the L2/3 processor switches to another event, resuming L3 processing on the L2-accepted event after additional data has been transferred. This way, no dead-time is incurred between events due to data transfers between the L1 and L2/3 buffers.

If the event passes L3, the processed results are propagated back up the fanout and highway switches to an external cross-connect switch that routes accepted events from all 8 highways to a small cluster of data-logging nodes for archival.

4 Data Rates

Each of the 256 DCB subsystems serving the detector front-end interfaces with two dozen 2.5 GB/s optical links to bring the data from the collision hall to the optical receivers in the counting room [7]. This makes it possible to achieve a peak design rate of 1.5 TB/s providing sufficient headroom for the estimated data rate of 800 GB/s. The data rate going into the L1 buffers in each highway is reduced 8× to 100 GB/s through the use of the parallel highway architecture described above. The data coming out of the pixel detector front-ends is on the order of 40 GB/s which is increased to about 70 GB/s after additional information, such as the ID's of the pixel read-out chips, is inserted into the data stream by the pixel DCB's. This means the data rate going into the pixel processors in each highway is ∼9 GB/s. Assuming an average size of 4 KB for the total number of triplets found by the L1 segment tracker for each crossing, the data rate going into the L1 track/vertex processor farm is ∼4 GB/s. The resulting data rate into each node in a farm of over 300 processors is an easily manageable ∼13 MB/s. If each node produces 20 bytes of summarized trigger results, the total data rate going from the track/vertex farm to the GL1 processor in each highway is ∼20 MB/s.

The L1 trigger rejects 99% of the input to the L1 buffers reducing the output from the buffers by 100× to 1 GB/s (for simplification, we will treat L2 and L3 as a single trigger stage in this discussion). This means the data rate coming out of each of the 32 L1 buffer subsystems is ∼31 MB/s which can readily be

handled by a single gigabit ethernet port on the highway switch. Since there are 12 fanout switches, the data rate into each fanout is ∼83MB/s which can easily be accomodated by the pair of gigabit uplink ports available on the fanouts. Data is distributed to each node in the L2/3 processor farm at ∼3.5 MB/s using one of the 24 fast ethernet ports on the fanout switches. The L2/3 trigger rejects 95% of the incoming data reducing this by 20× to 50 MB/s. A 2× data compression further reduces this to 25 MB/s. The resulting data rate from all 8 highways into the cross-connect switch and the data-logging cluster is a mere 200MB/s which can easily be handled by commercially available storage technology.

5 Conclusion

We have presented a detailed description of BTeV's three-level, eightfold-way trigger architecture. We have shown how key design aspects of this architecture allow us to keep the data rates at very manageable levels. This, in turn, permits us to build a sophisticated trigger system based on low-cost commercially available components that meets the demanding physics requirements of the BTeV experiment.

References

[1] A. Kulyavtsev et al., BTeV proposal, Fermilab, May 2000, BTeV-doc-66.

[2] G. Y. Drobychev et al., Update to BTeV proposal, Fermilab, March 2002, BTeV-doc-316.

[3] S. Kwan, Pixel Detector and Silicon Forward Tracker, Vertex 2002: 11th International Workshop on Vertex Detectors, Nov. 3-8, 2002, Kailua-Kona, Hawaii.

[4] M. Wang, BTeV Level 1 Vertex Trigger Algorithm, BTeV-doc-1179.

[5] M.H.L.S. Wang for the BTeV Collaboration, BTeV Level 1 Vertex Trigger, Nucl. Instrum. Meth. A 501 (2003) 214.

[6] E.E. Gottschalk, BTeV detached vertex trigger, Nucl. Instrum. Meth. A 473 (2001) 167.

[7] The BTeV DAQ Group, Overview of the BTeV Readout and Controls System, BTeV-doc-1091.