BTeV Trigger

Erik Gottschalk, Fermilab

(for the BTeV Trigger Group)

- Introduction
- Brief overview of the BTeV detector
- Level-1 trigger algorithm
- Trigger architecture
- Level-1 trigger hardware
  - Original baseline design (digital signal processors – DSPs)
  - New baseline design (commodity hardware)
  - Proposed change to the new baseline design (upstream event builder & blade servers)

Not covered:

- Level-1 muon trigger
- Level-2/3 trigger (high-level trigger – "HLT")
- Real Time Embedded Systems (RTES) Project

- The challenge for the BTeV trigger and data acquisition system was to reconstruct particle tracks and interaction vertices for EVERY proton-antiproton interaction in the BTeV detector, and to select interactions with *B* decays.

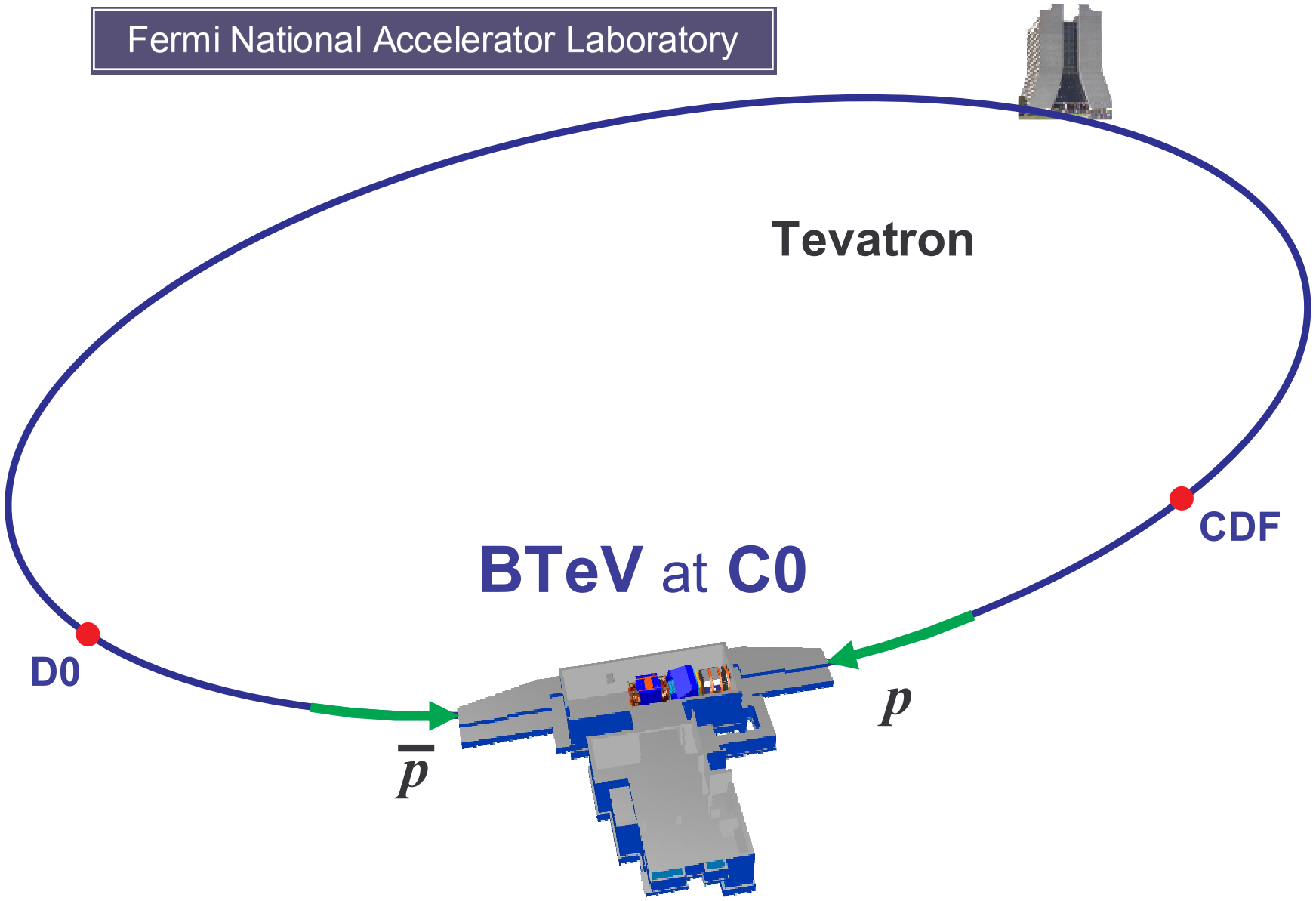
- The trigger system was designed with 3 levels, referred to as Levels 1, 2, and 3:
  "L1" – look at every interaction and reject at least 98% of minimum bias background
  "L2" – use L1 computed results & perform more refined analyses for data selection
  "L3" – reject additional background and perform data-quality monitoring

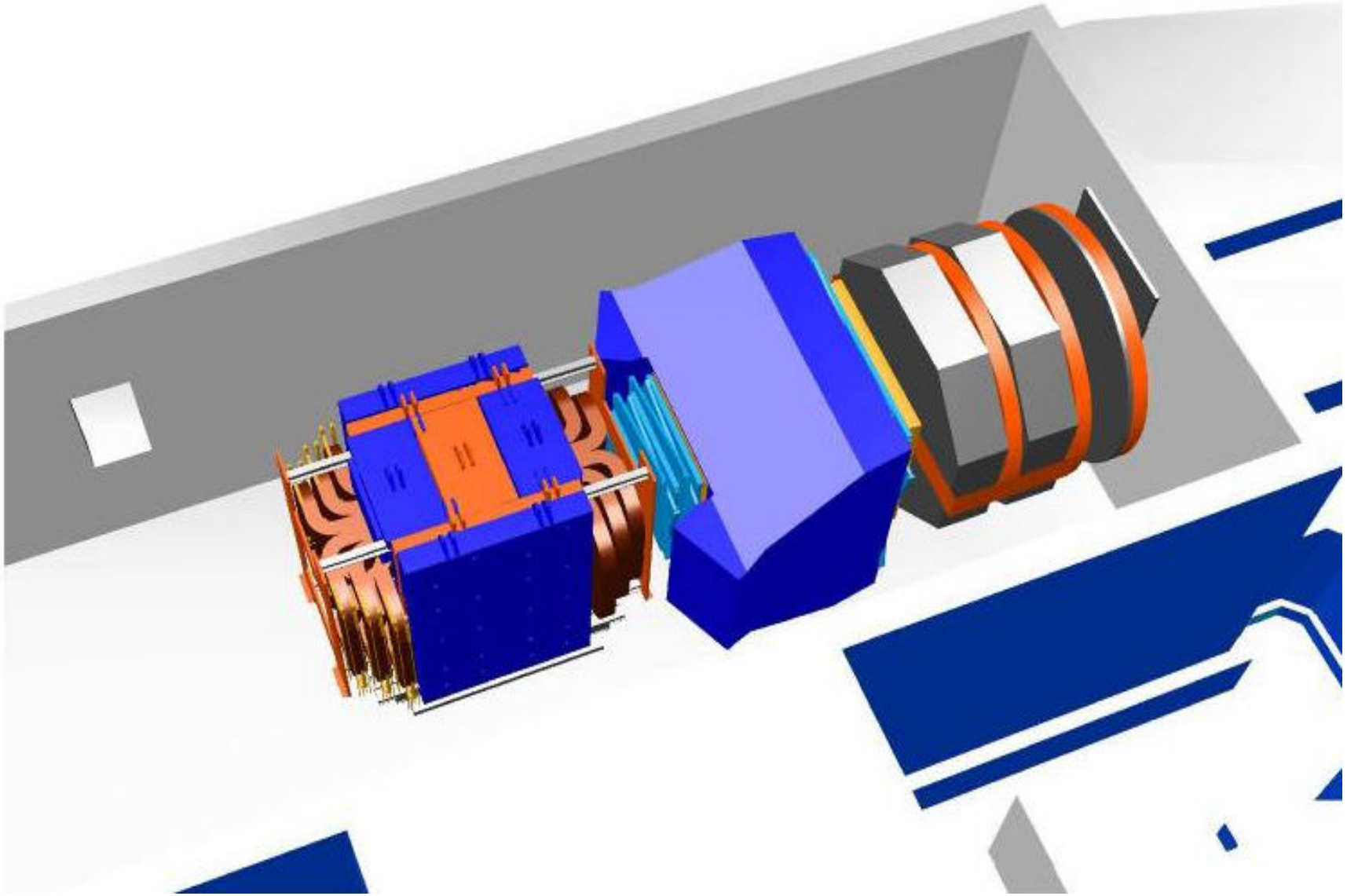
Reject > 99.9% of background. Keep > 50% of *B* events.

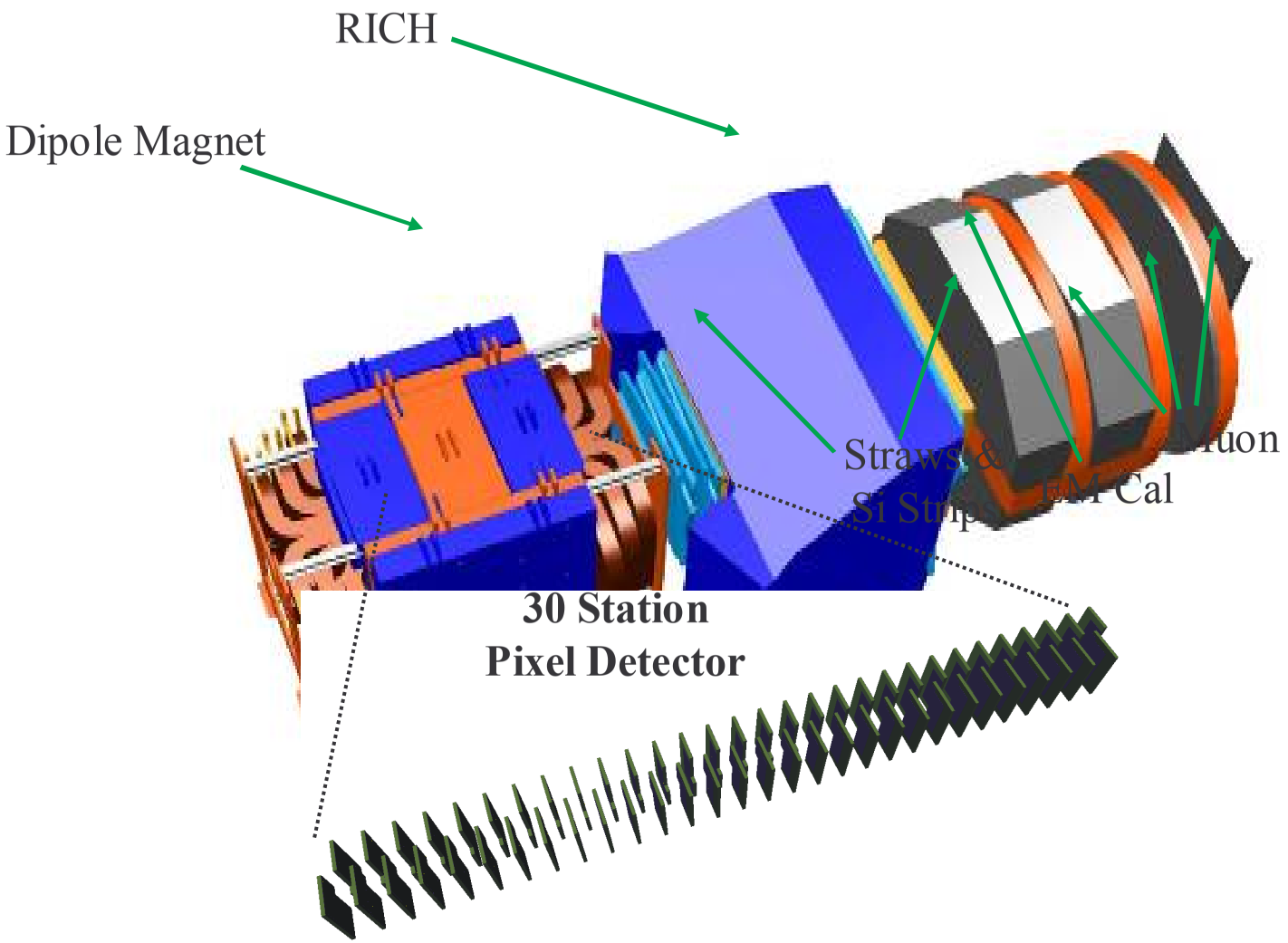
- The data acquisition system was designed to save all of the data in memory for as long as was necessary to analyze each interaction, and move data to L2/3 processors and archival data storage.

- The key ingredients that made it possible to meet this challenge:
  - BTeV pixel detector with its exceptional pattern recognition capabilities
  - Rapid development in technology – FPGAs, processors, networking

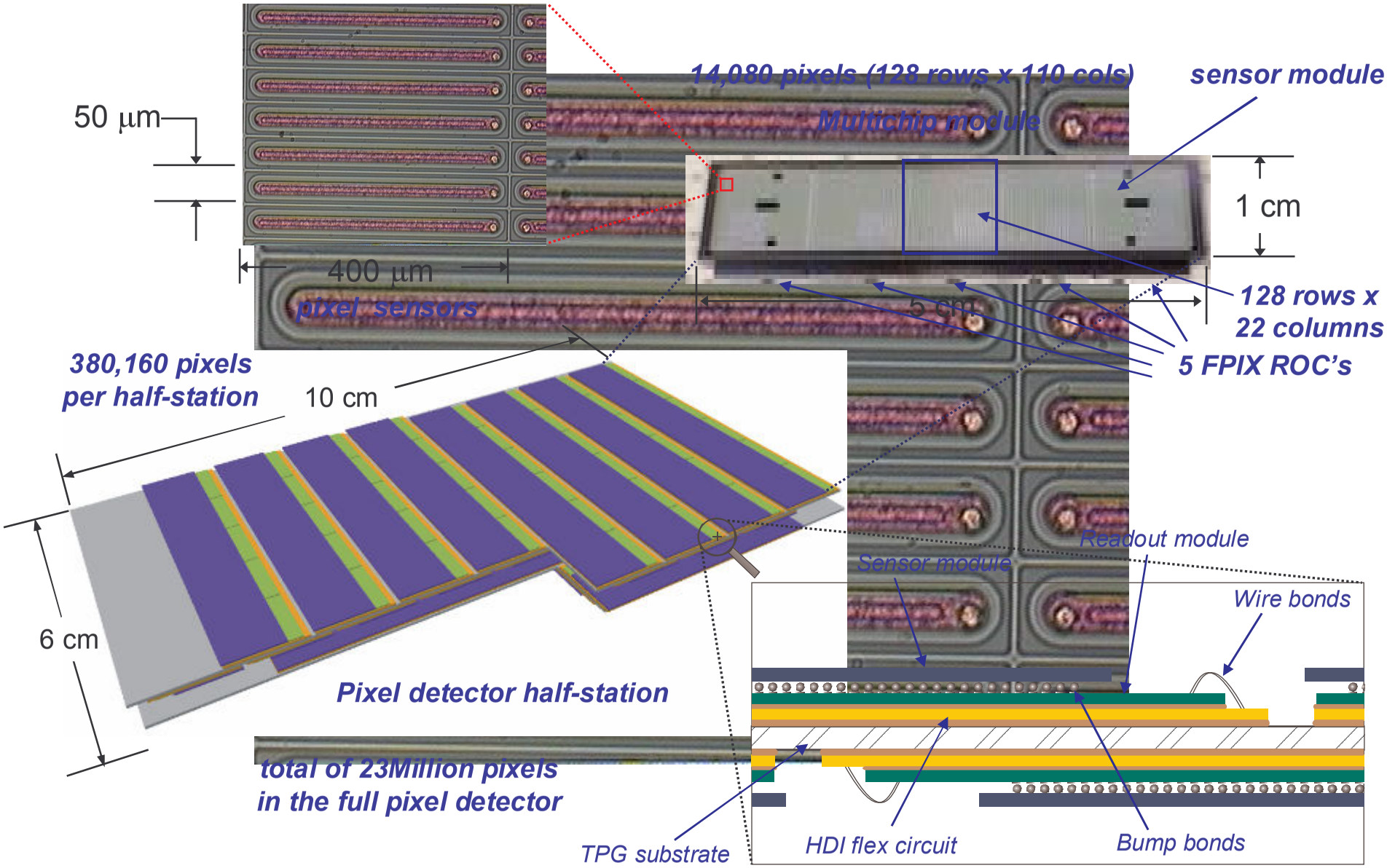
Fermi National Accelerator Laboratory



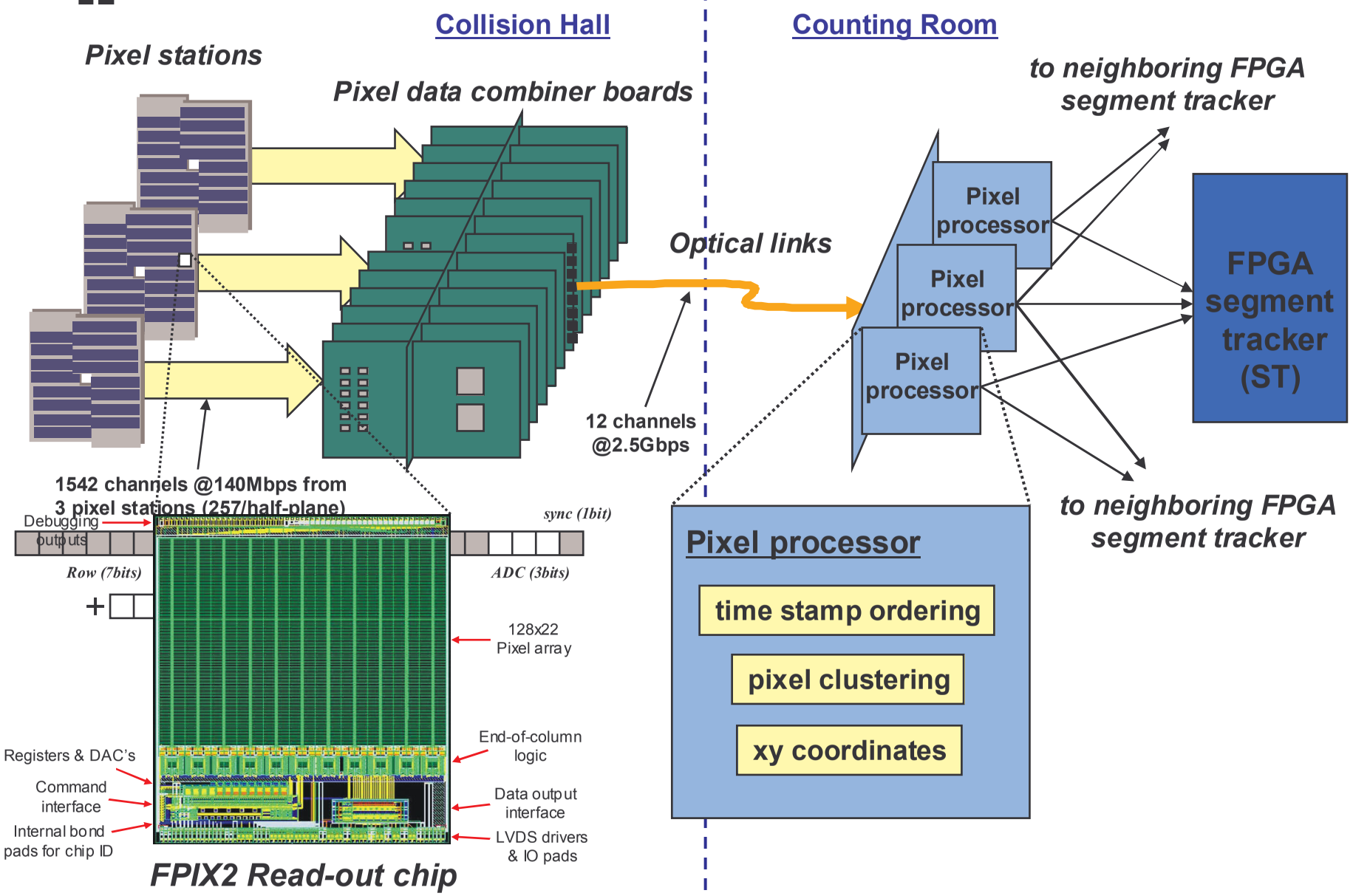


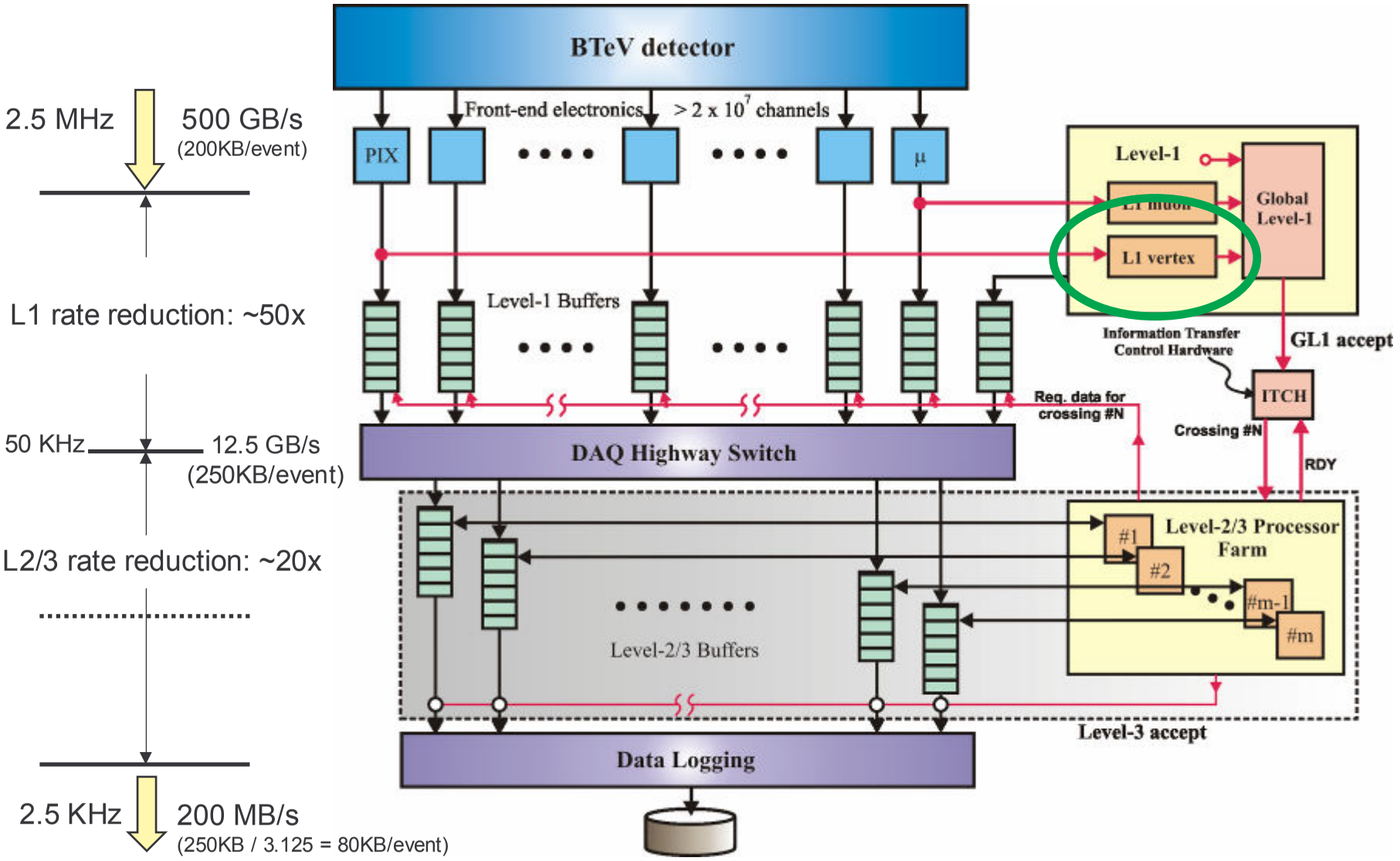

RICH
Dipole Magnet
Straws &
Si Strip
EM Cal
uon
30 Station
Pixel Detector

Silicon Pixel Detector

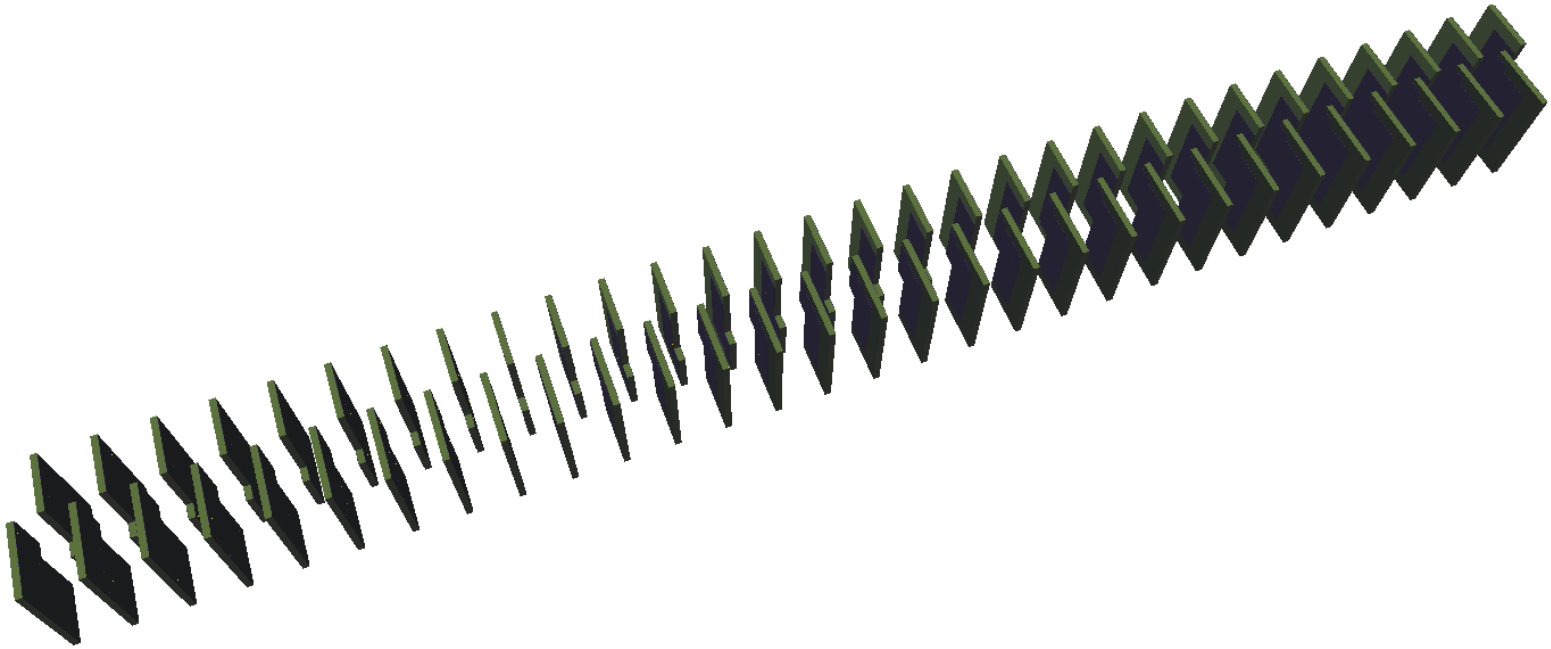


Pixel Data Readout & 1st Part of L1 Trigger

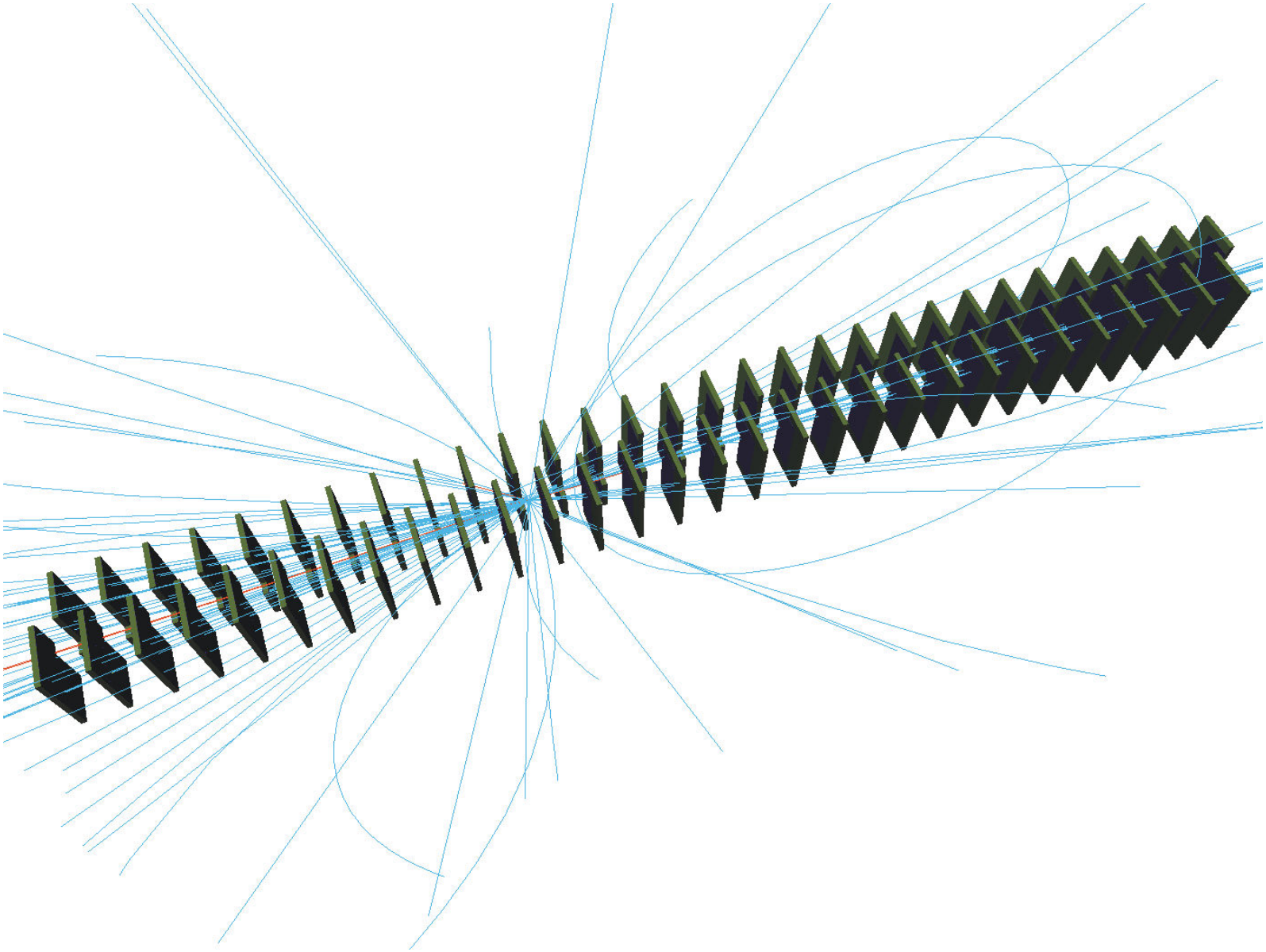



2.5 MHz
500 GB/s
(200KB/event)
L1 rate reduction: ~50x
50 KHz
12.5 GB/s
(250KB/event)
L2/3 rate reduction: ~20x
2.5 KHz
200 MB/s
(250KB / 3.125 = 80KB/event)
BTeV detector
Front-end electronics
> 2 x 10^7 channels
PIX
μ
Level-1
Global Level-1
L1 muon
L1 vertex
Level-1 Buffers
Information Transfer Control Hardware
GL1 accept
ITCH
Req. data for crossing #N
Crossing #N
RDY
DAQ Highway Switch
Level-2/3 Processor Farm
#1
#2
#m-1
#m
Level-2/3 Buffers
Level-3 accept
Data Logging

Simulated B Event

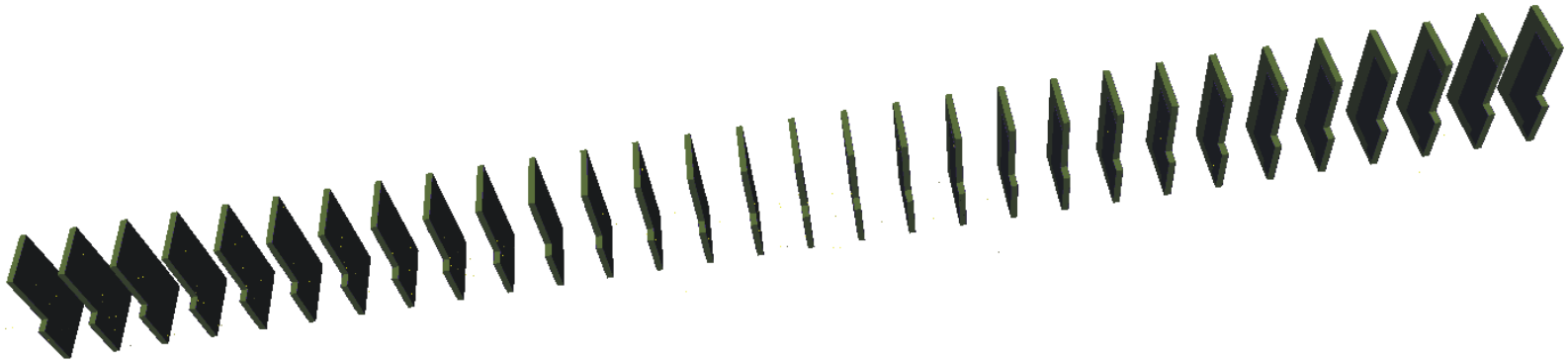


Simulated B Event

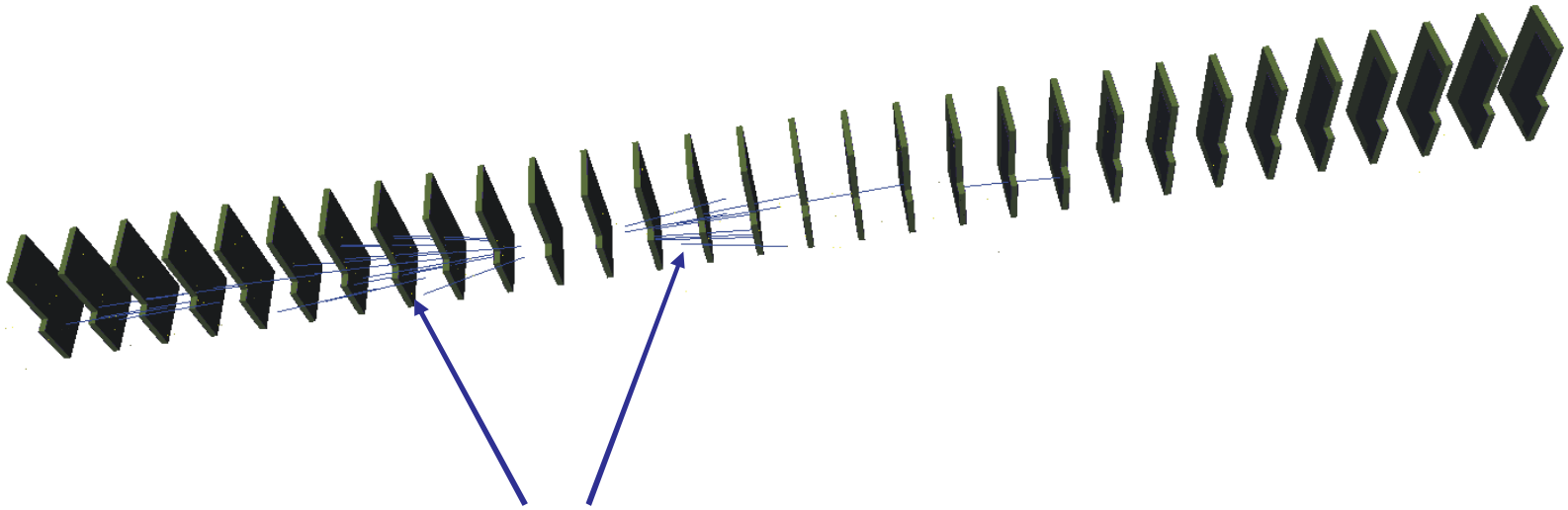


Two stage trigger algorithm:

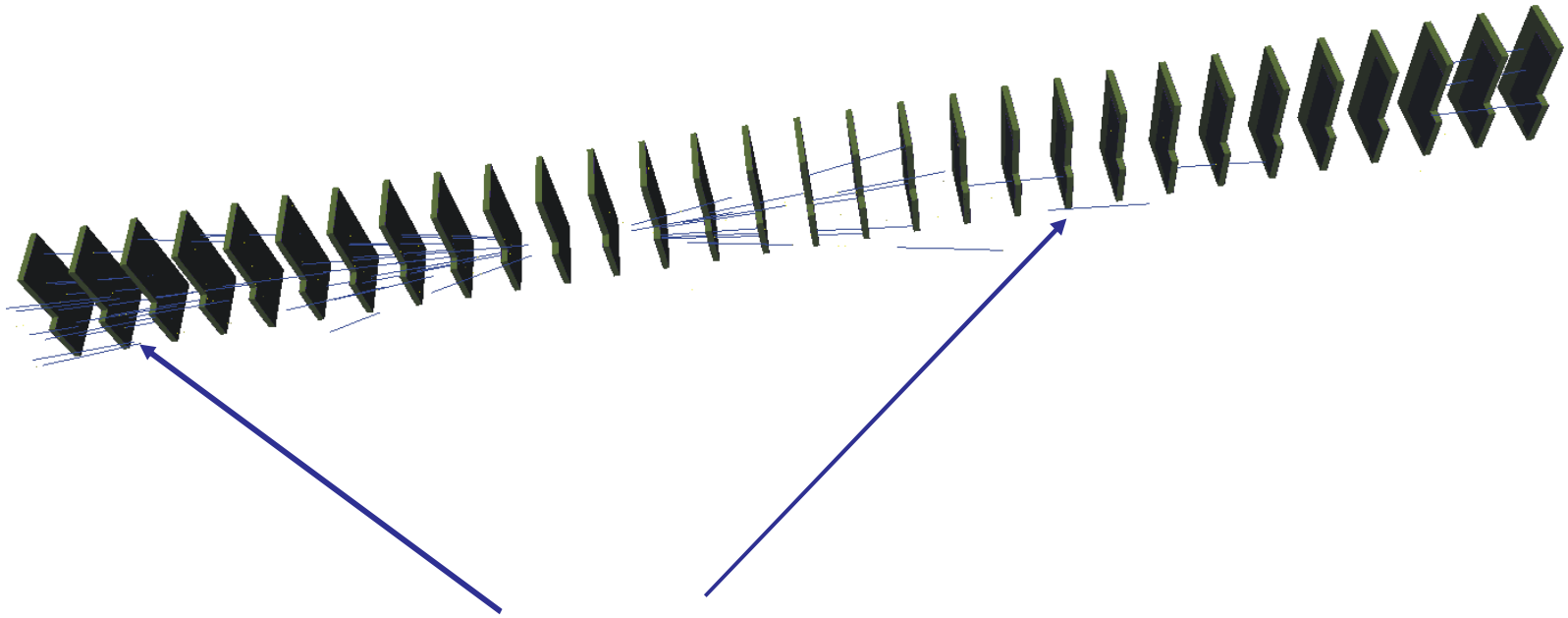
1. ***Segment tracking***
2. ***Track/vertex finding***



1) <u>Segment tracking stage:</u>
Use pixel hits from 3 neighboring stations to find the beginning and ending segments of tracks. These segments are referred to as <u>triplets</u>.



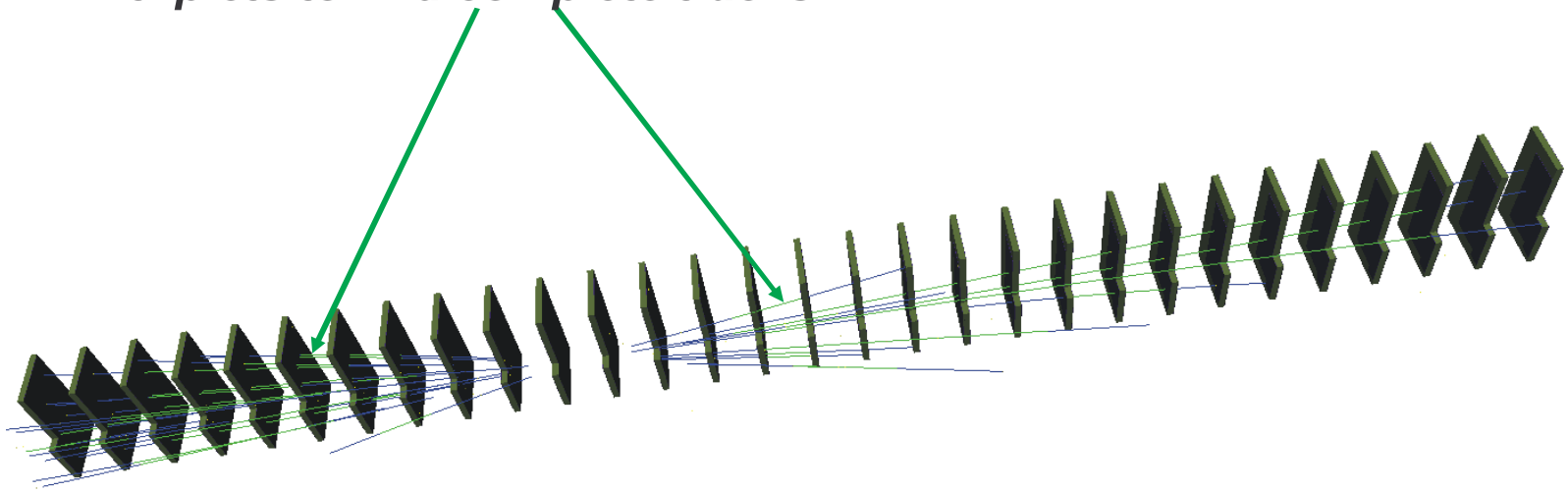
1a) <u>Segment tracking stage: phase 1</u>
Start with inner triplets close to the interaction region.
An inner triplet represents the start of a track.



1b) <u>Segment finding stage: phase 2</u>

Next, find the outer triplets close to the boundaries of the pixel detector volume. An outer triplet represents the end of a track.

2a) <u>Track finding phase:</u>
Match inner triplets with outer triplets to find complete tracks.



2b) <u>Vertex finding phase:</u>

- ***Use reconstructed tracks to locate interaction vertices***
- ***Search for tracks detached from interaction vertices***

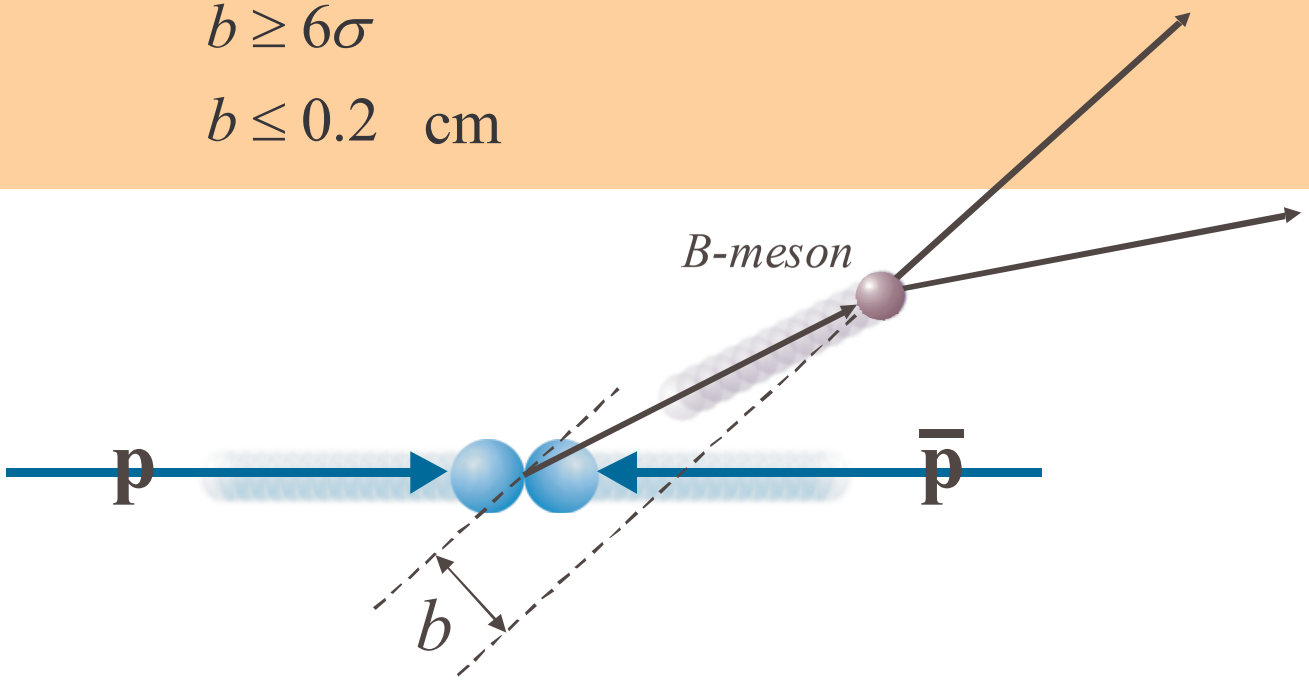
Execute Trigger

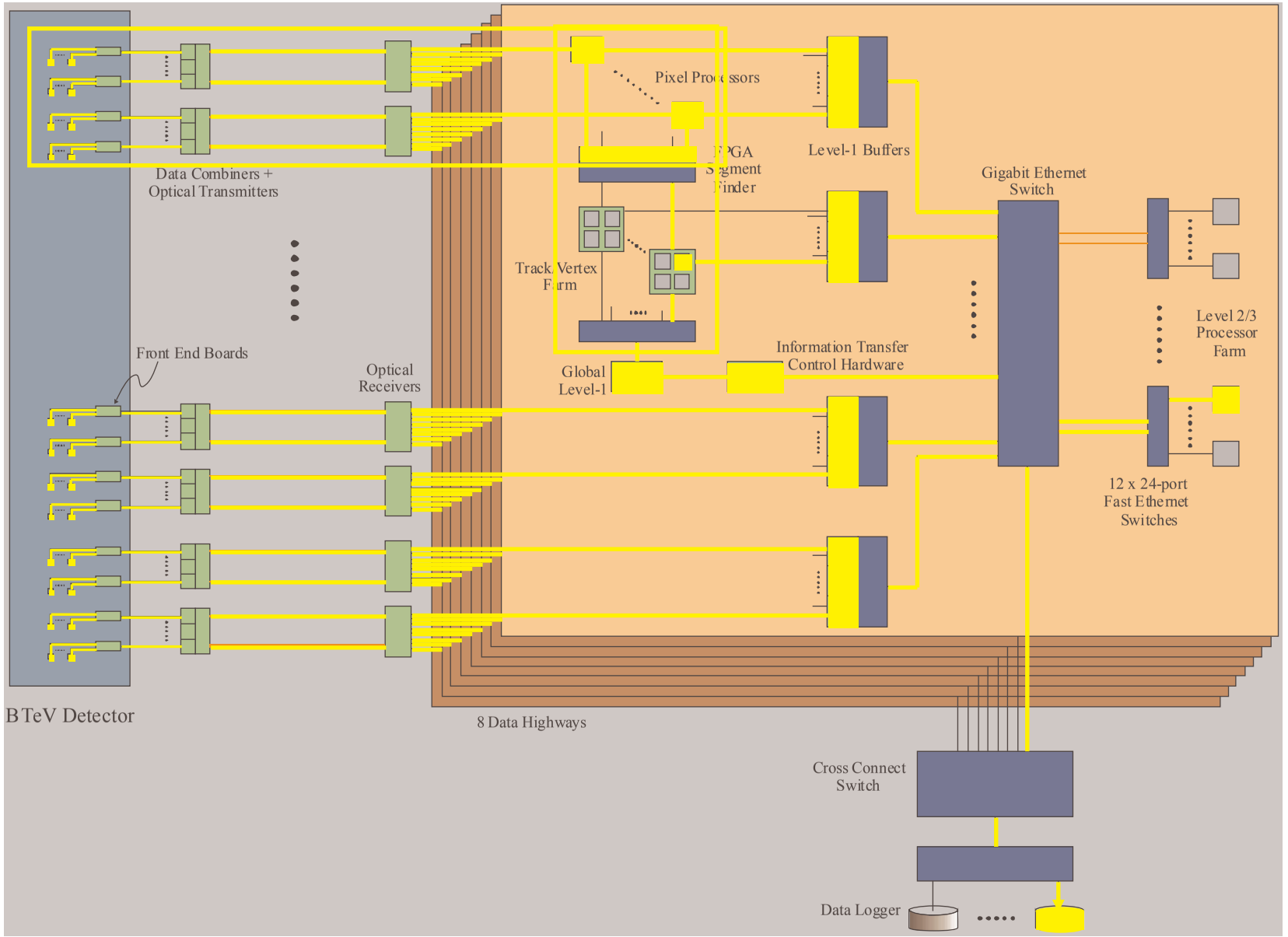
- **Generate Level-1 accept if ≥ 2 "detached" tracks going into the instrumented arm of the BTeV detector with:**

$$p_T^2 \geq 0.25 \quad (\mathrm{GeV/c})^2$$

$$b \geq 6\sigma$$

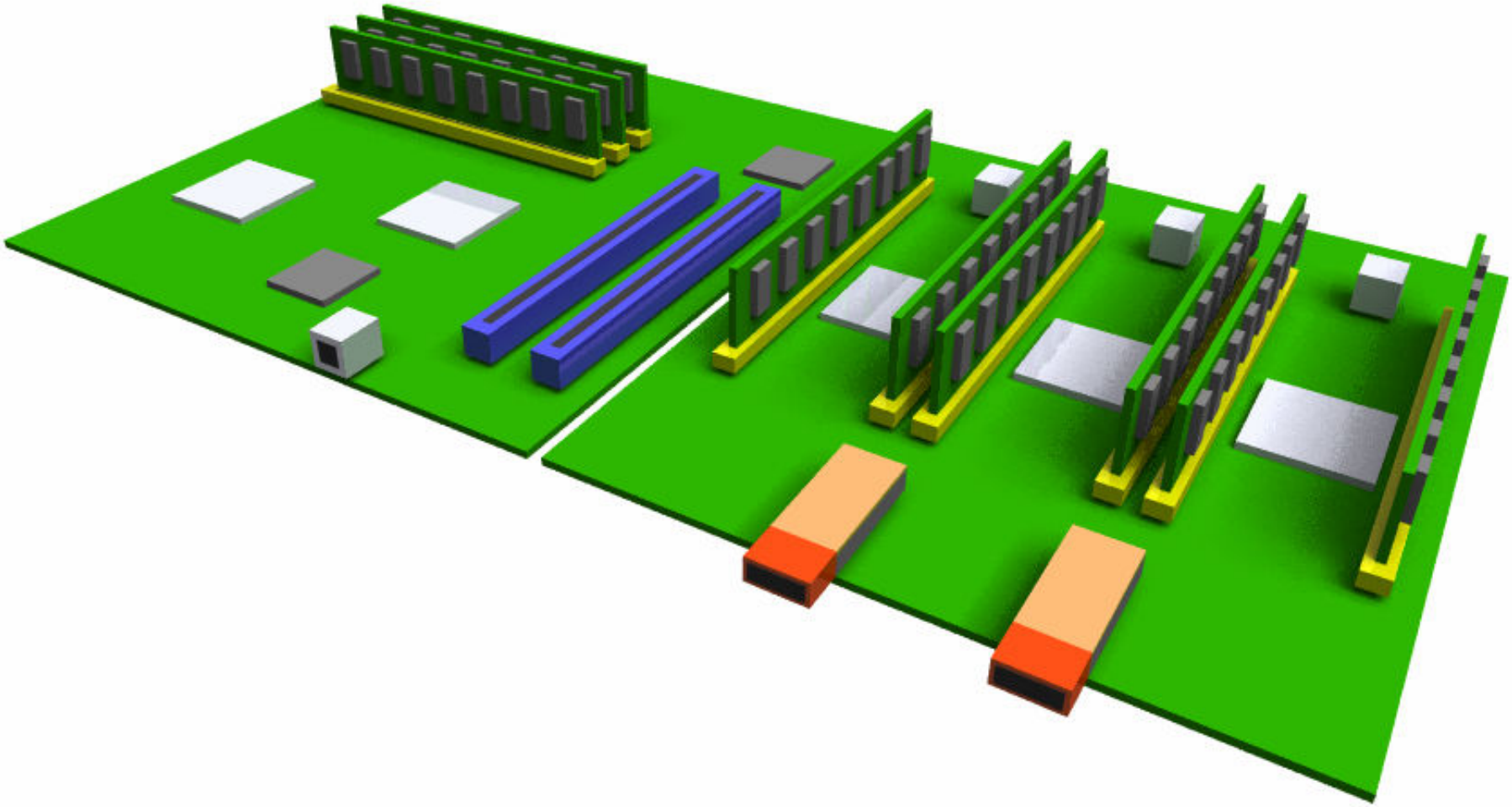
$$b \leq 0.2 \quad \mathrm{cm}$$




Pixel Processors
Level-1 Buffers
Gigabit Ethernet Switch
Data Combiners + Optical Transmitters
Track Vertex Farm
Level 2/3 Processor Farm
Front End Boards
Optical Receivers
Global Level-1
Information Transfer Control Hardware
12 x 24-port Fast Ethernet Switches
BTeV Detector
8 Data Highways
Cross Connect Switch
Data Logger

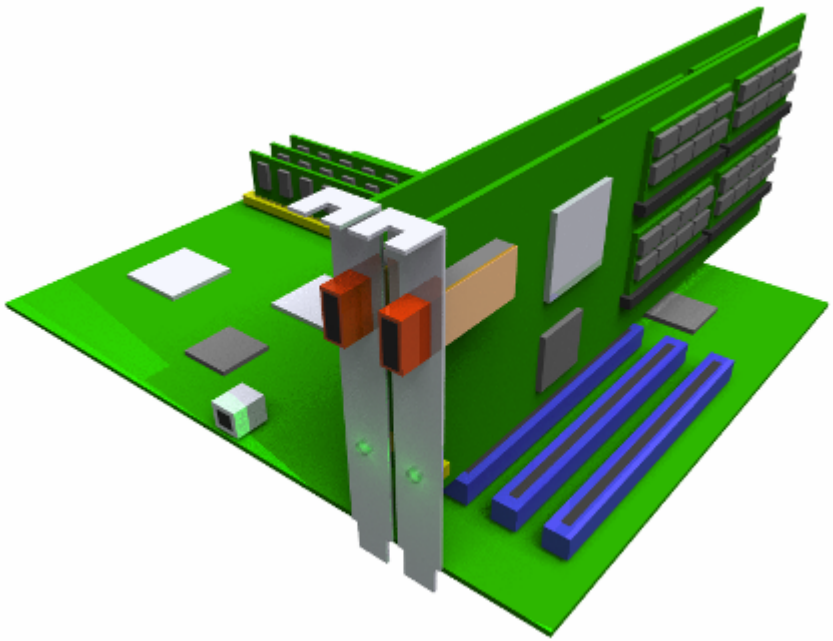
~0.5TB L1 Buffers Based on Commodity DRAM

Output buffer



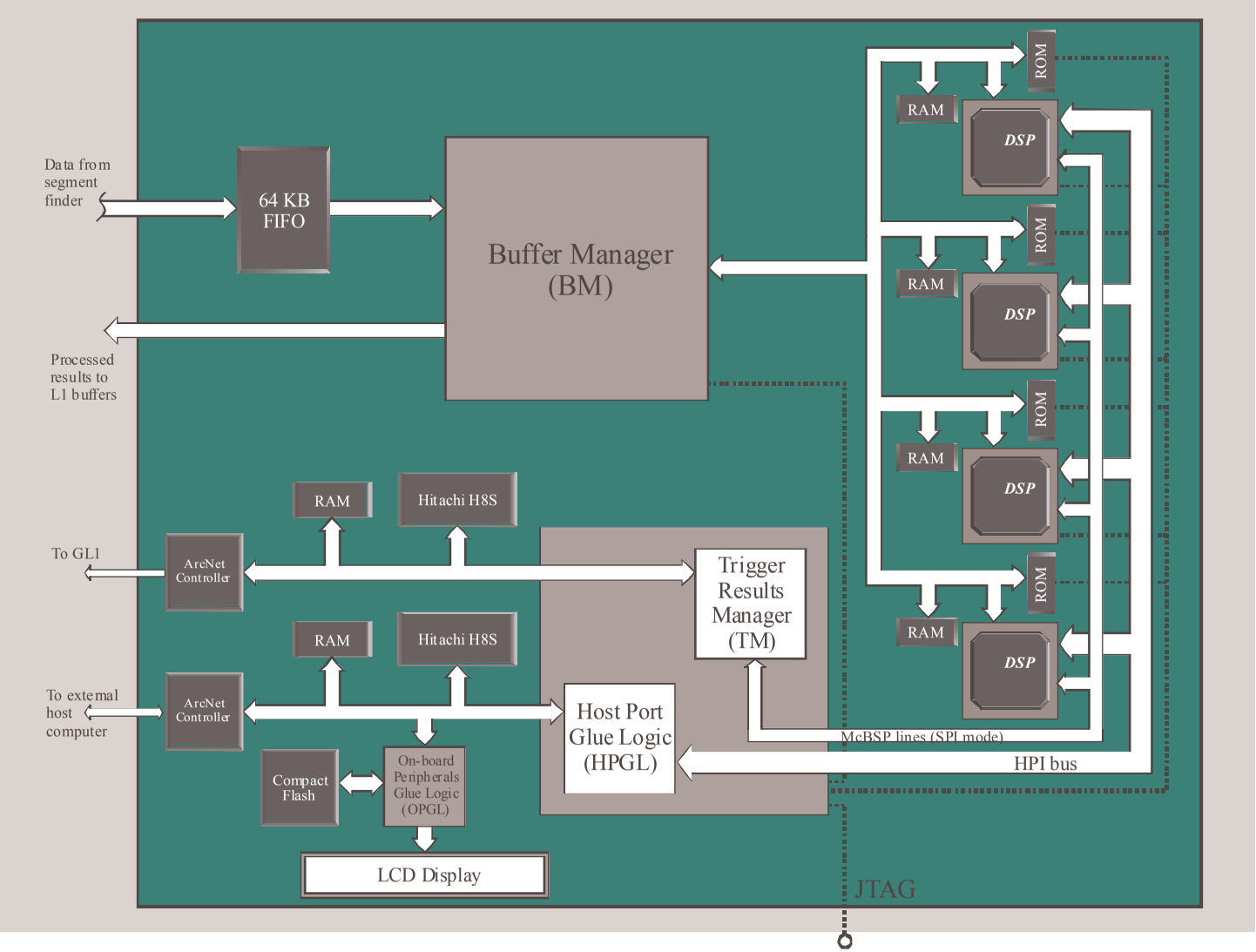
(DDR DRAM)

Prototype L1 Buffer



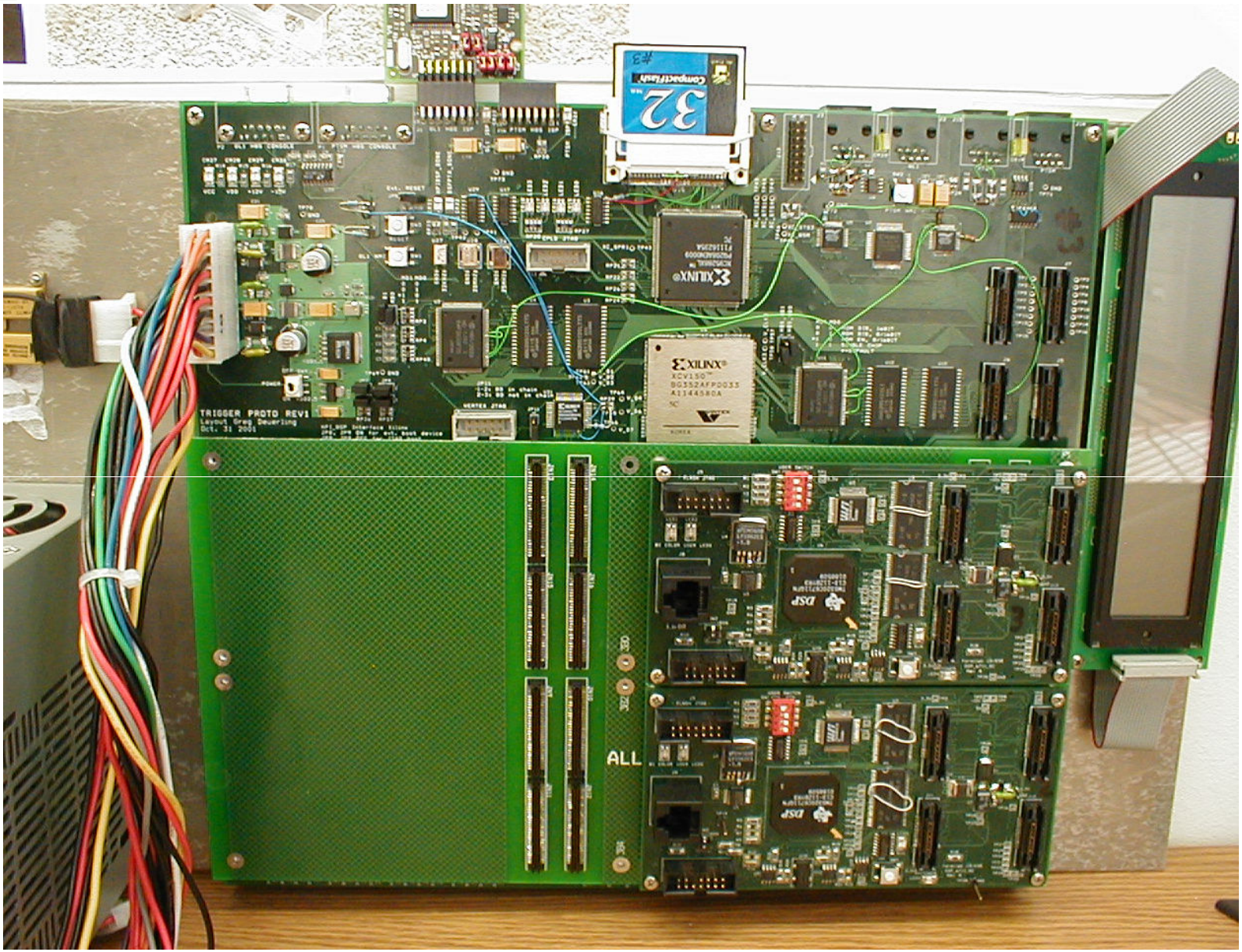
PCI card based L1 buffer
board with commodity DRAM

PC
with Gigabit
ethernet acting
as L1 buffer server

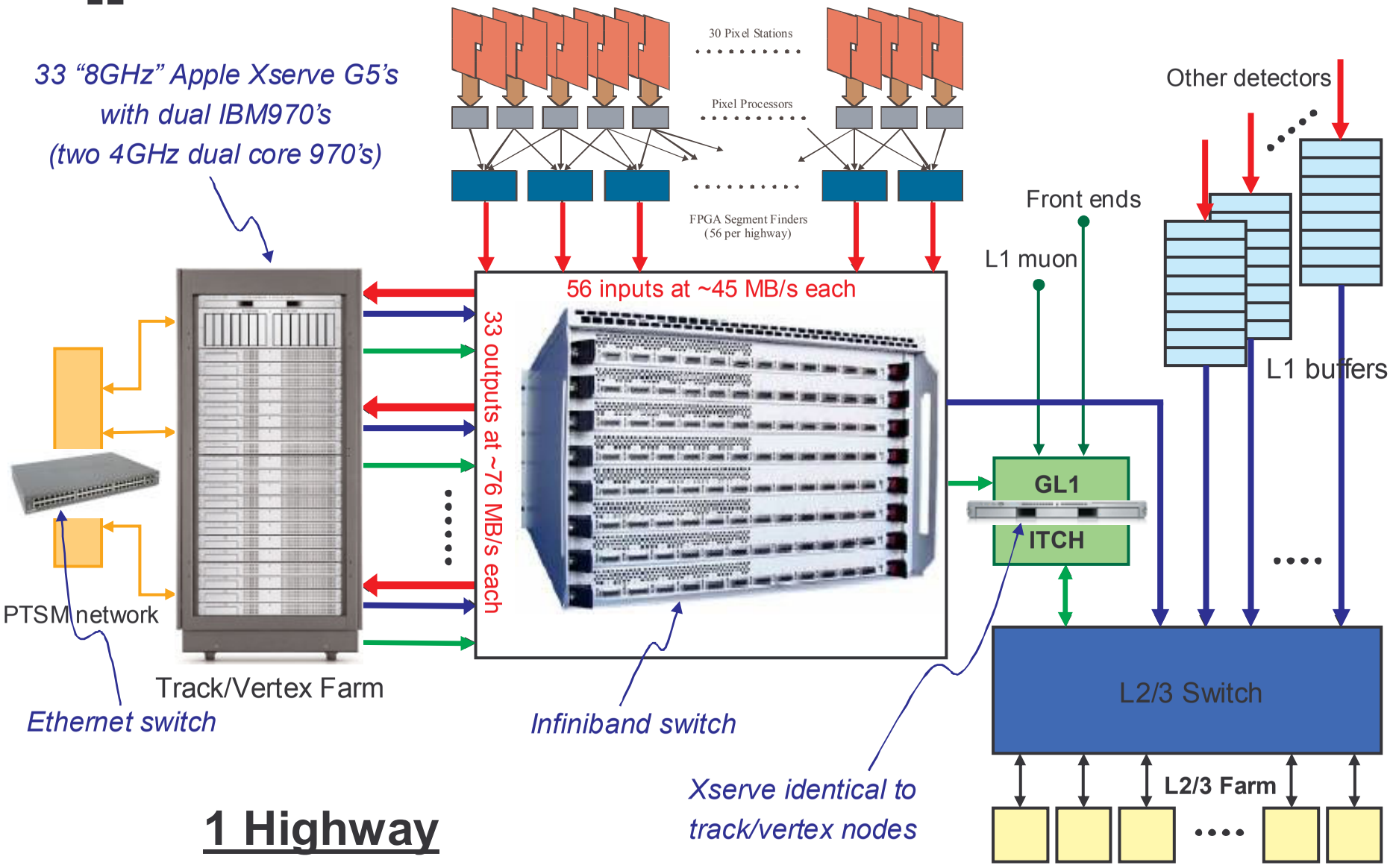


Block diagram of prototype L1 track/vertex farm hardware

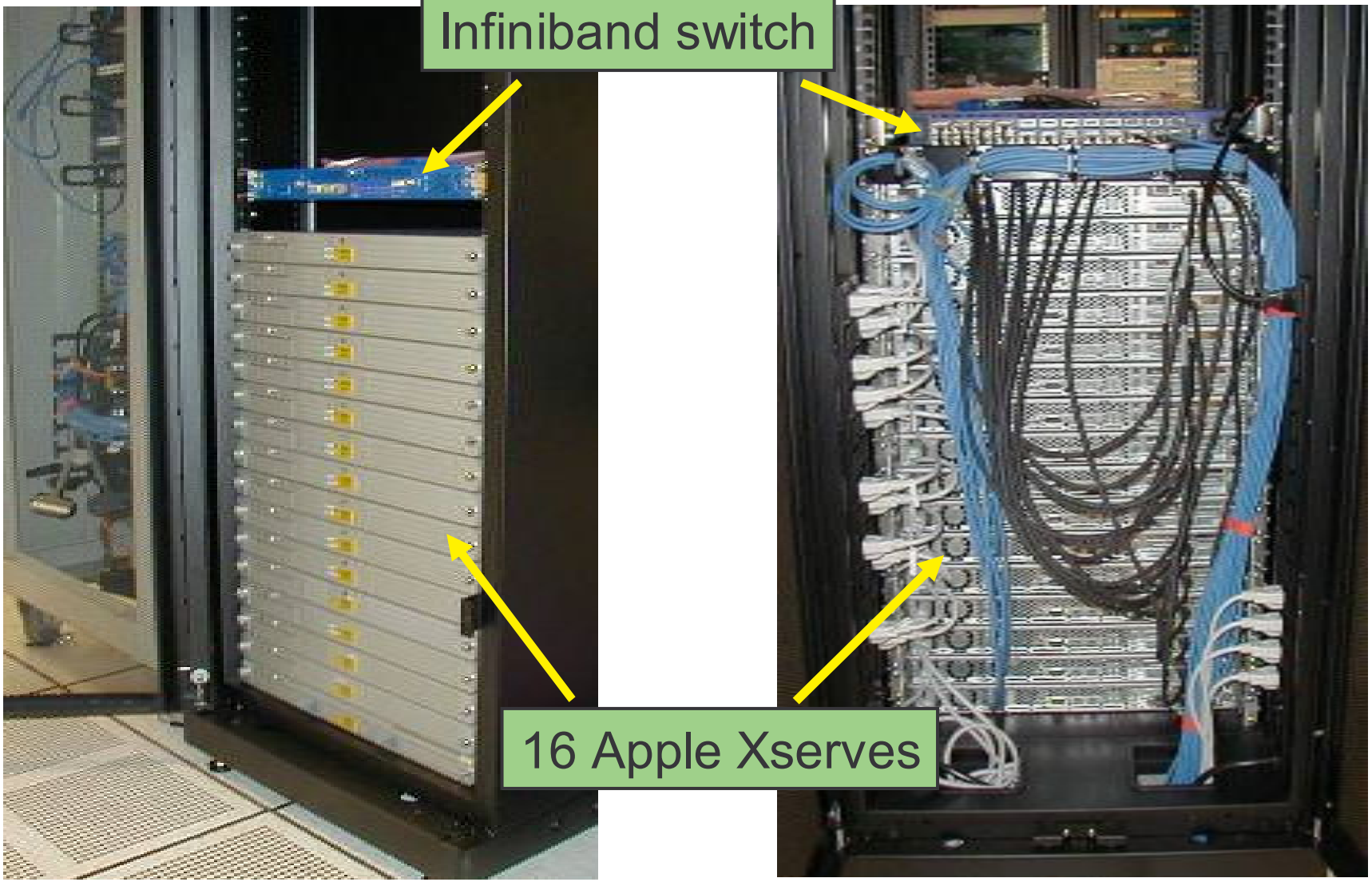
DSP-Based Track/Vertex Hardware Prototype



New L1 Baseline using Commodity Hardware



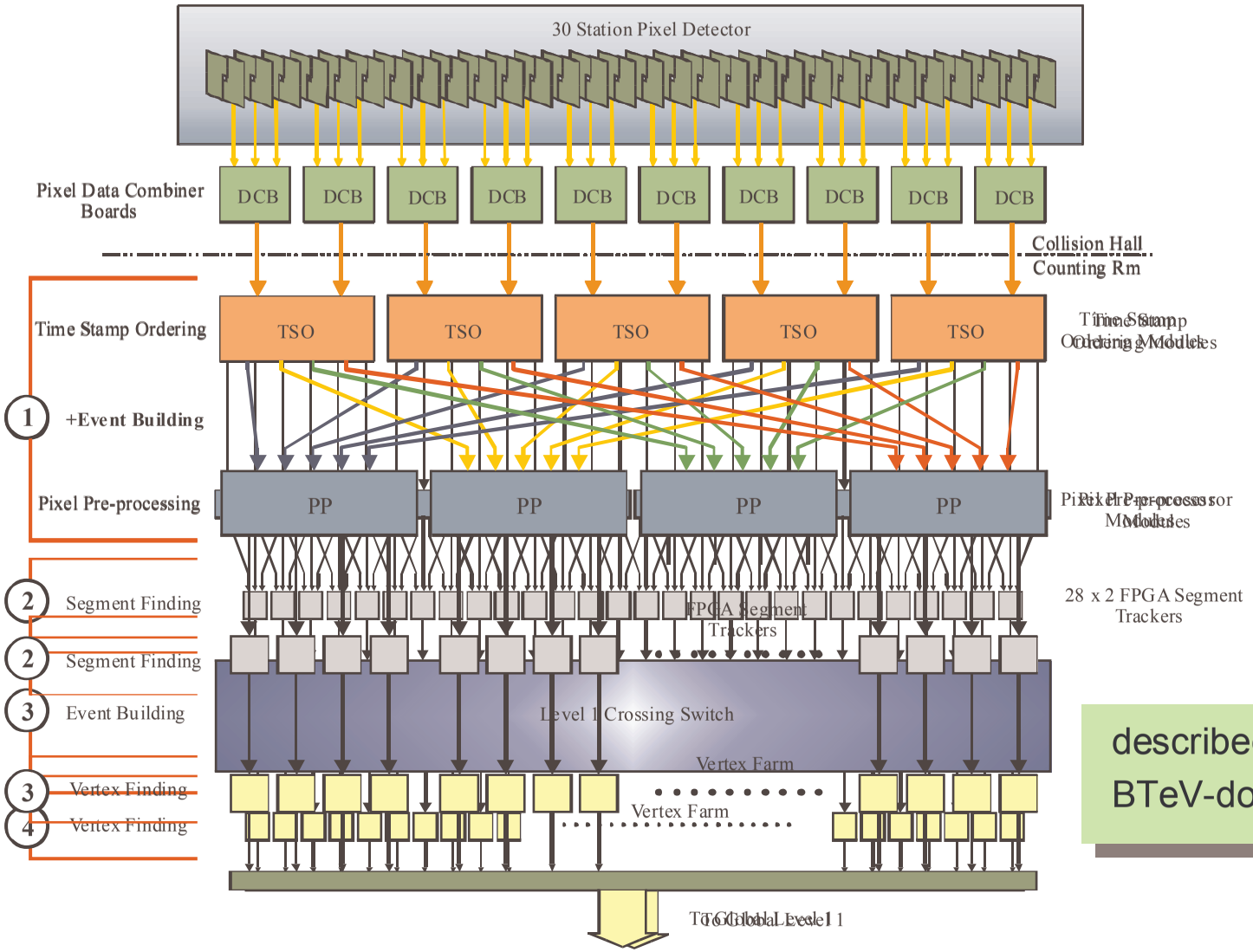
1 Highway



Front

Rear

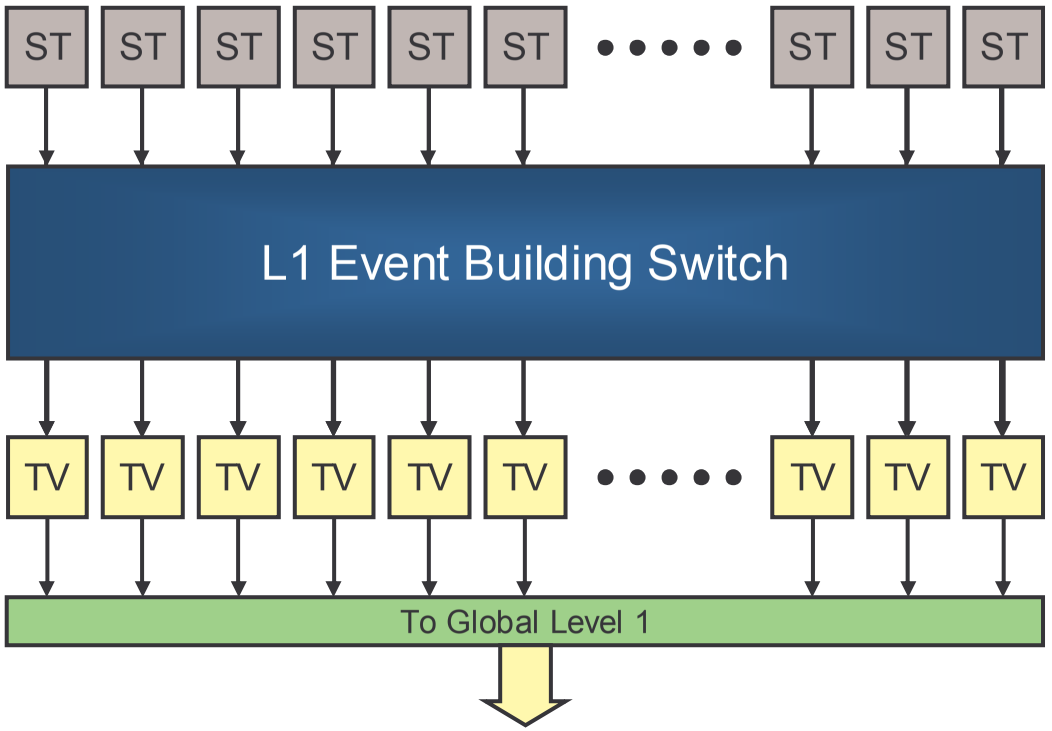
New integrated Current baseline event builder architecture



described in:
BTeV-doc-3342

L1 Farm Transformation

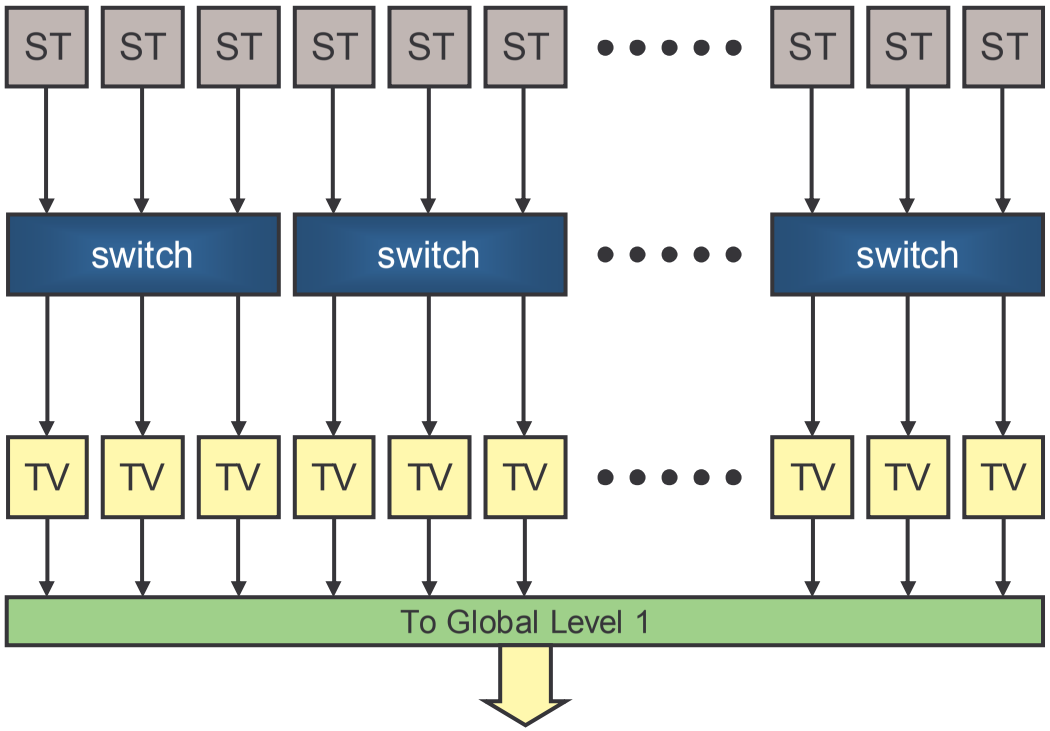
Segment Trackers have complete events. No need for a single large switch for event building.



ST: Segment Tracker

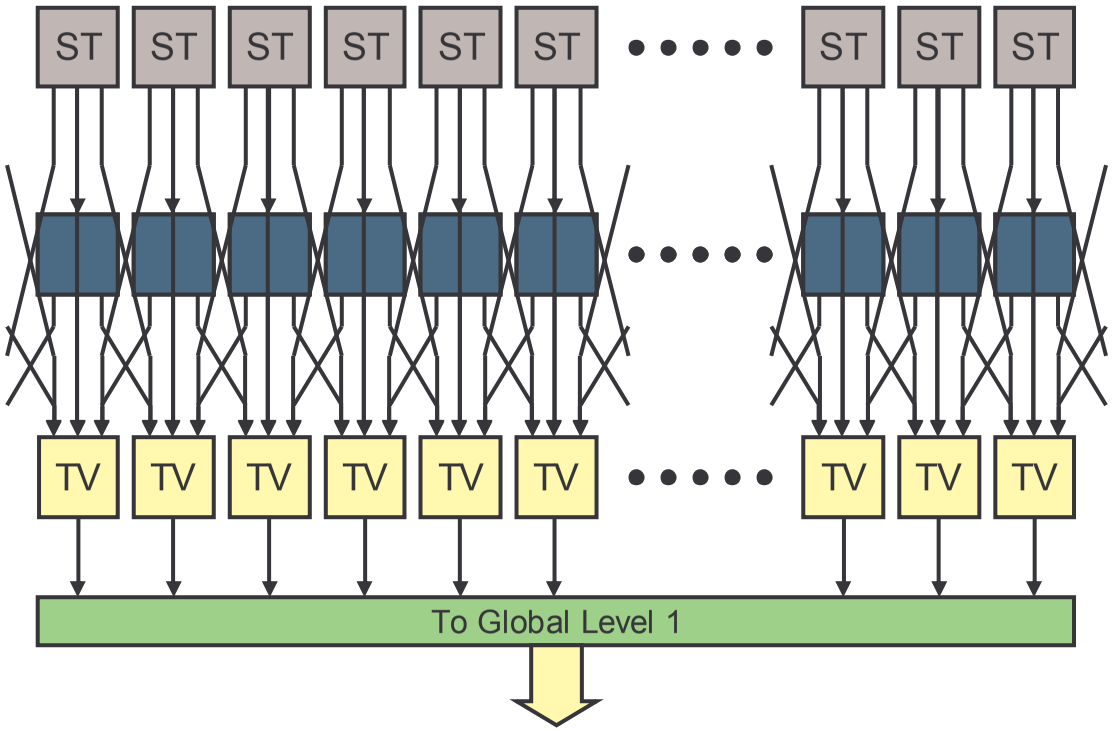
TV: Track/Vertex Finder

L1 Farm Transformation



However, we may still need a switching function to deal with TV node failures. This can be handled by smaller switches.

L1 Farm Transformation

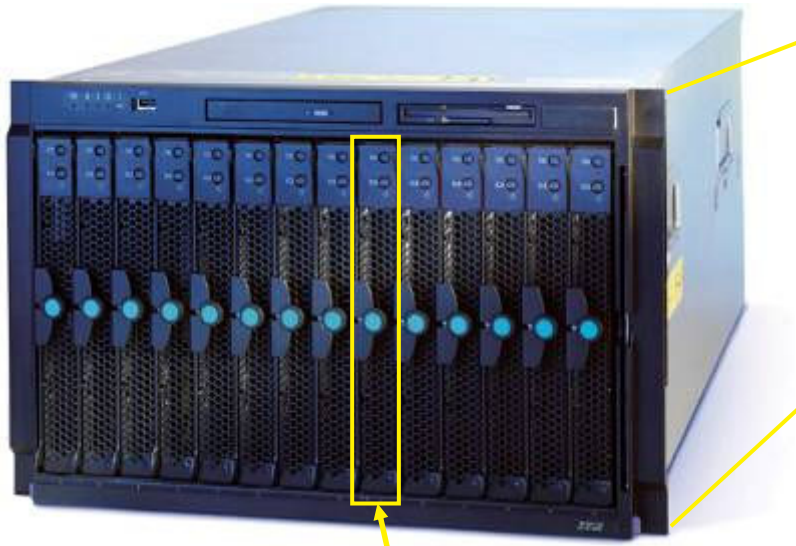


Or, the switching function can be handled by a "Buffer Manager" as it was done in our DSP prototype.

Which could possibly be integrated into the Segment Trackers (ST).

Blade Server Platform

In the proposed architecture we replace Apple Xserves with blade servers.



Intel/IBM blade server chassis

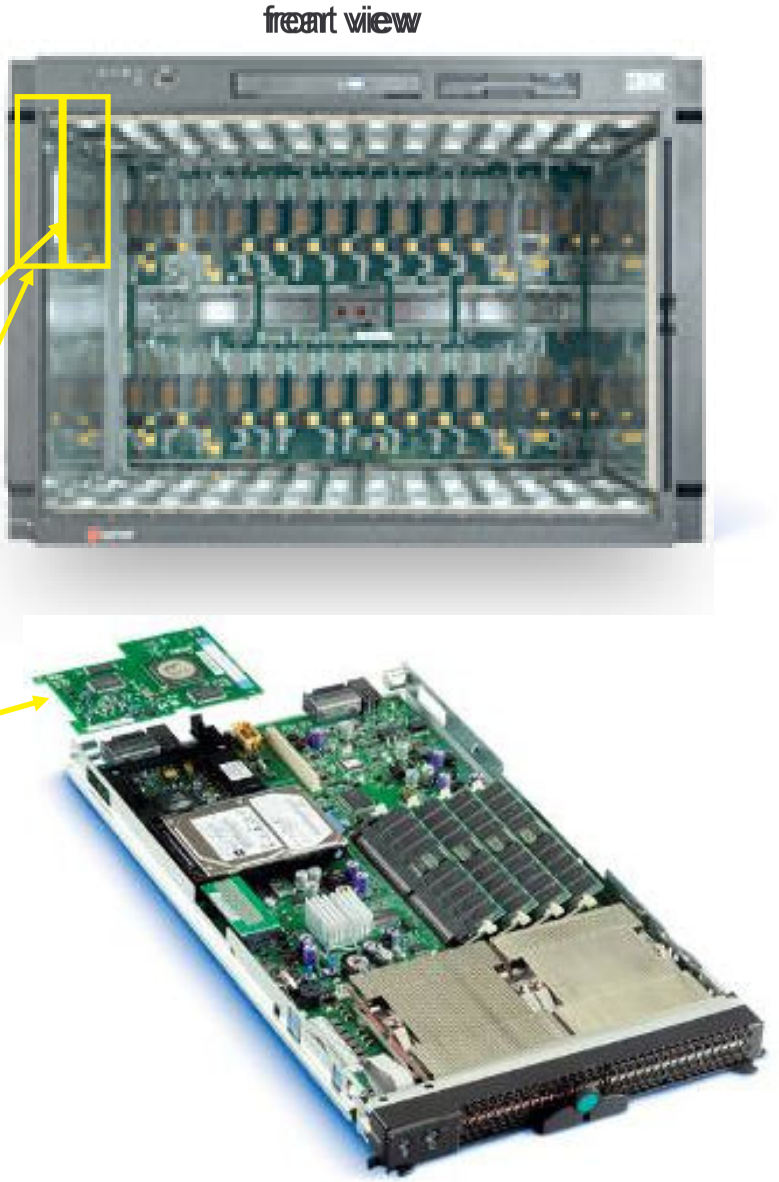
Each 7U crate can hold up to 14 dual-cpu blades which are available with Intel Xeon's (>3.0GHz) or IBM970PPC's.

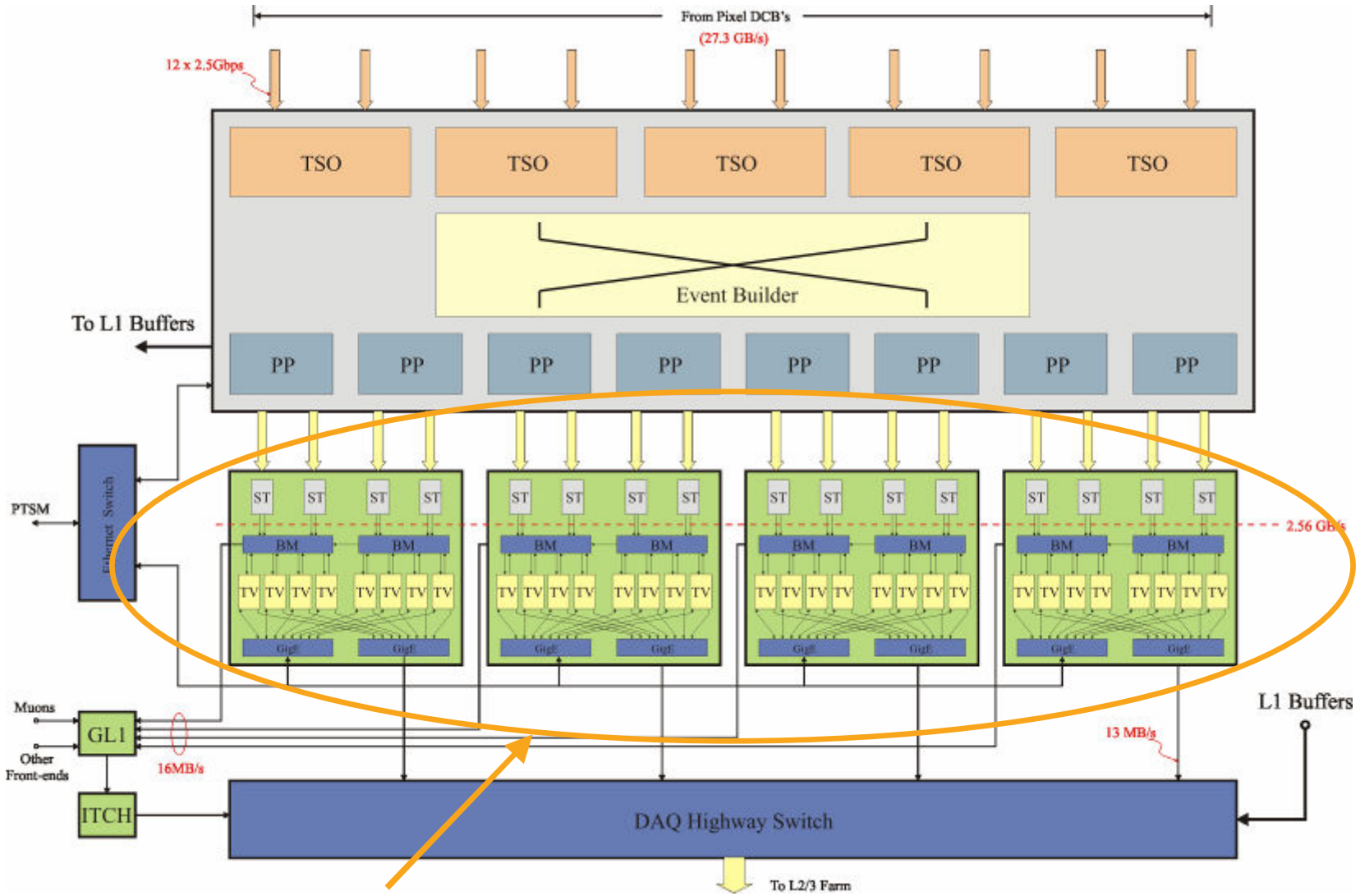
6 of these crates will fit into a standard 42U rack for a total of 168 CPU's (on dual cpu blades) per rack.

Blade Server Platform Features

2 network interfaces are provided on the mid-plane of the chassis:

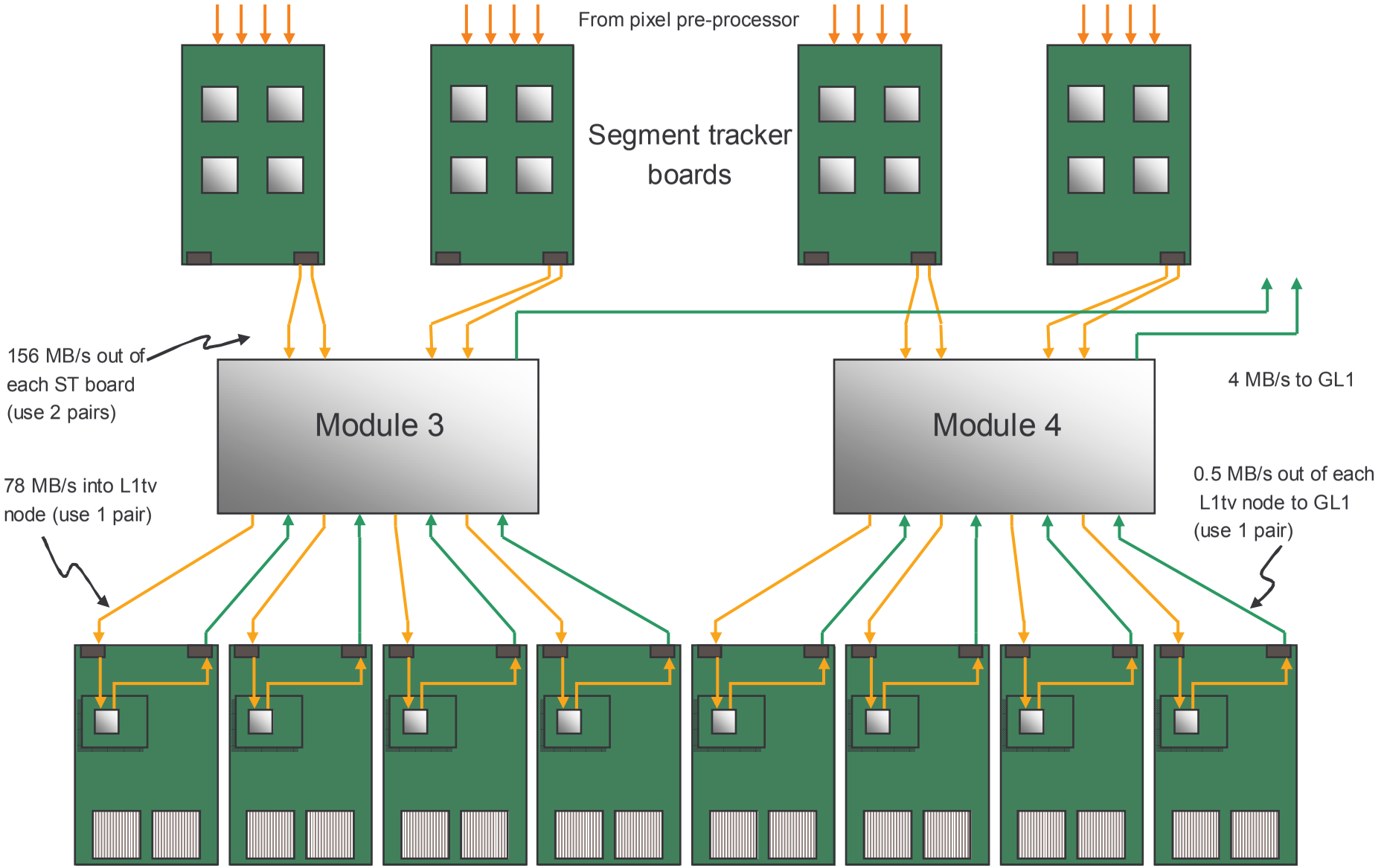
- primary/base network: each slot connects blade's on-board gigabit-ethernet ports to a switch module in the rear
- secondary hi-speed network: each slot also connects ports on the blade's optional I/O expansion card to an optional hi-speed switch module (myrinet, infiniband, etc.)
- the I/O expansion card might even be a custom card with an FPGA to implement a custom protocol.



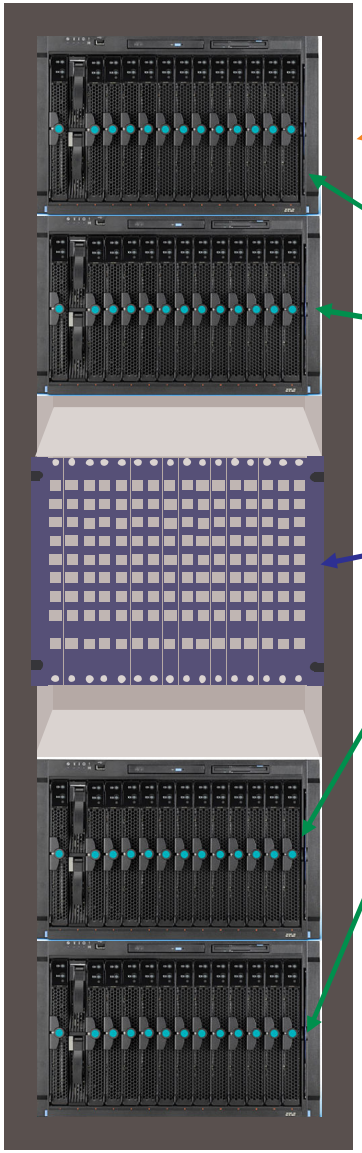


Complete segment tracking and track/vertex hardware for 1 highway housed in 4 blade server crates

L1 Trigger Hardware in a Blade Server Crate



Level 1 Hardware Rack Count



Complete L1 trigger hardware for
one highway can fit in one 42U rack

Segment finder & track/vertex hardware
in 4 blade server crates

Upstream event builder/pixel pre-processor

Entire L1 trigger for all 8 highways
in 8 42U racks

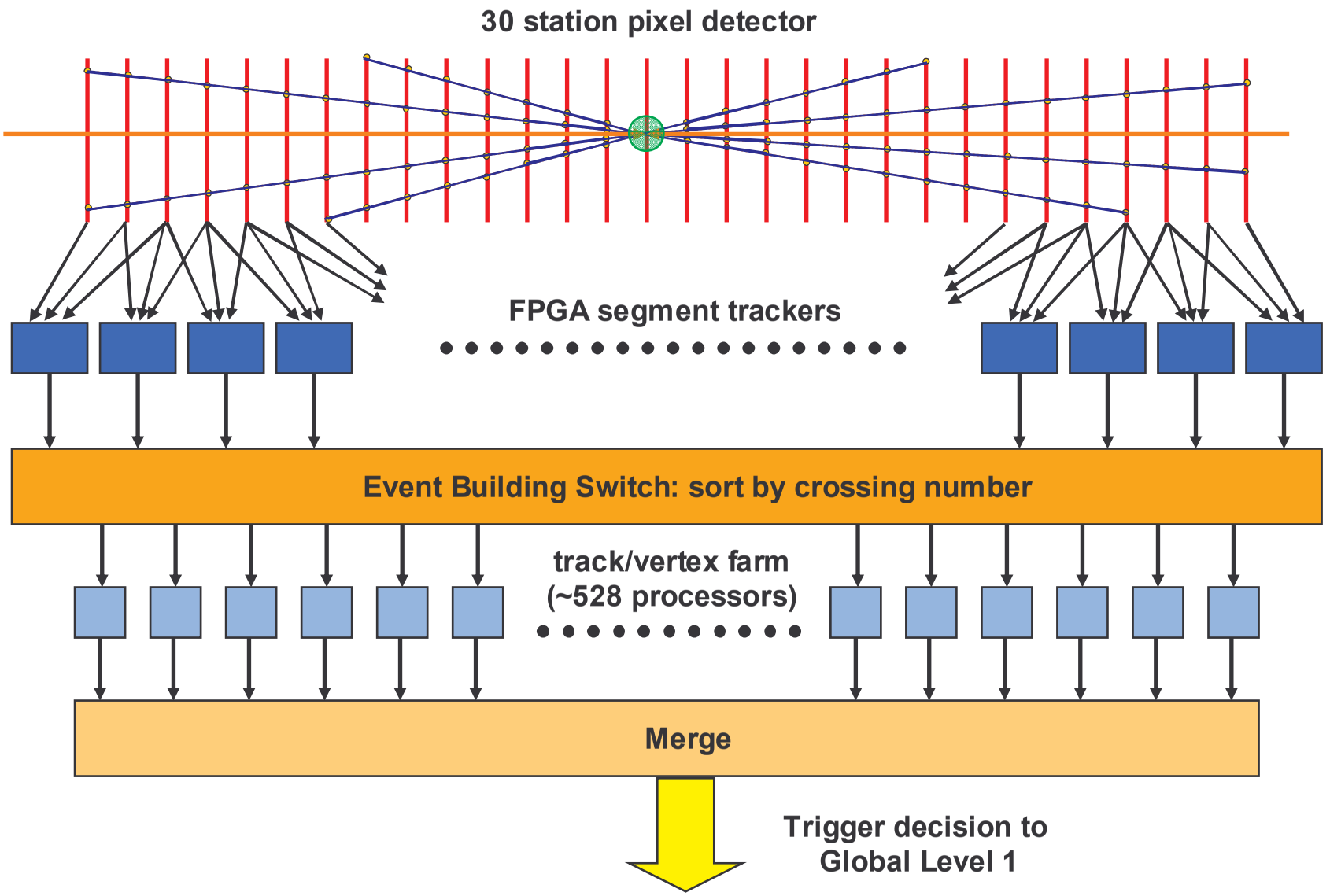
The BTeV trigger evolved from an original baseline design that satisfied all BTeV trigger requirements to a new design that was easier to build, required less labor, had lower cost, and lower risk.

When BTeV was canceled we were on the verge of proposing a new design to the Collaboration. This design was expected to:

- reduce the cost of the trigger system
- reduce the amount of space needed for the hardware
- improve the design of the system architecture
- improve system reliability (redundancy & fault tolerance)
- improve integration of L1 & L2/3 trigger systems

Ideas that we developed for the BTeV trigger are likely to be used in future high-energy physics and nuclear physics projects.

End


30 station pixel detector
FPGA segment trackers
Event Building Switch: sort by crossing number
track/vertex farm
(~528 processors)
Merge
Trigger decision to
Global Level 1

Backup slides

Data Rate into L1 Farm

Using the triplet format proposed in BTeV-doc-1555-v2:

internal triplets: 128 bits = 16 Bytes
external triplets: 74 bits = 10 Bytes

Then assumed 35 internal and 35 external triplets for events with <2> interactions/crossing:

35 x 26 Bytes = 910 Bytes

Extrapolating linearly to 9 interactions/crossing and applying a safety factor of 2:

2 x (4.5 x 910 Bytes = 4095 Bytes) = 8190 Bytes

Total rate going into L1 track/vertex farm in all 8 highways :

2.5 MHz x 8190 Bytes ~ 20 GBytes/s

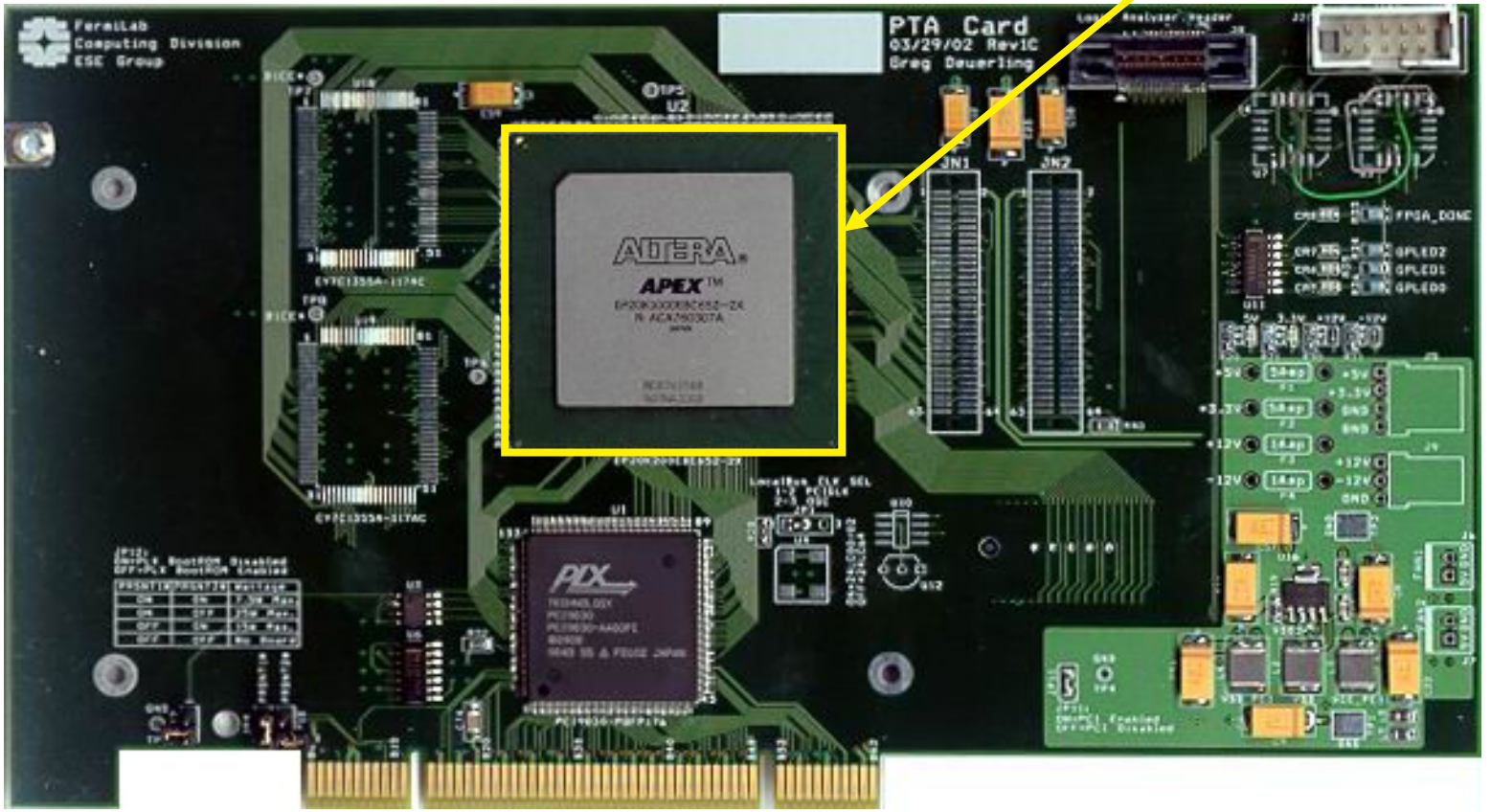
Required L1 Farm Computing Power

- **Assume an "8.0 GHz" IBM 970:**
  - L1 trk/vtx code (straight C code without any hardware enhancements like hash-sorter or FPGA segment-matcher) takes 379 μs/crossing on a 2.0 GHz IBM 970 (Apple G5) for minimum bias events with <6> interactions/crossing
  - assume following:
    - L1 code: 50%, RTES: 10%, spare capacity: 40%
    - 379 μs + 76 μs + 303 μs = 758 μs
  - 758 μs ÷ 4 = 190 μs on a 8.0 GHz IBM 970
  - Include additional 10% processing for L1-buffer operations in each node
    - 190 μs + 19 μs = 209 μs

- **Number of "8.0 GHz" IBM 970's needed for L1 track/vertex farm:**
  - 209 μs/396 ns = 528 cpu's for all 8 highways, 66 cpu's per highway
  - 33 dual cpu Apple Xserve G5's per highway

L1 Segment Tracker on a PTA Card

Uses Altera APEX EPC20K1000 instead of EP20K200 on regular PTA



Modified version of PCI Test Adapter card developed at Fermilab for testing hardware implementation of 3-station segment tracker (a.k.a. “Super PTA”)

Prototype L2/3 farm using nodes from retired FNAL farms

Real Time Embedded Systems (RTES)

- RTES: NSF ITR (Information Technology Research) funded project
- Collaboration of computer scientists, physicists & engineers from: Univ. of Illinois, Pittsburgh, Syracuse, Vanderbilt & Fermilab
- Working to address problem of reliability in large-scale clusters with real time constraints
- BTeV trigger provides concrete problem for RTES on which to conduct their research and apply their solutions