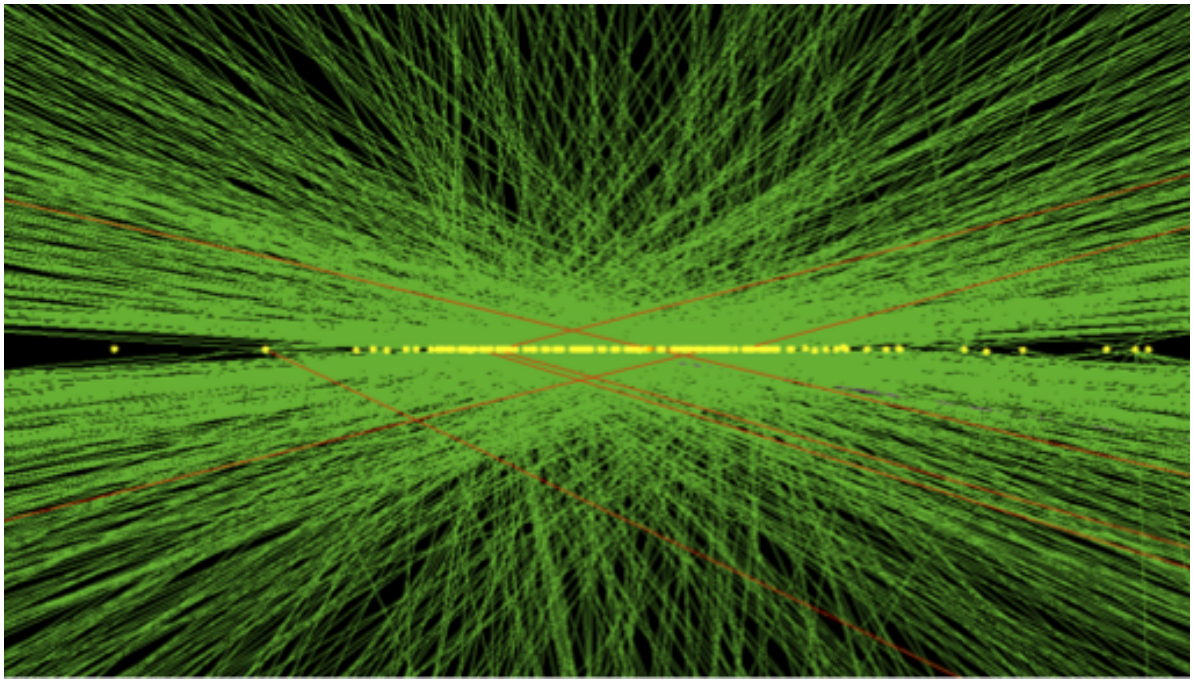
Tracking in the Trigger using FPGAs

Brian L. Winer
Ohio State University

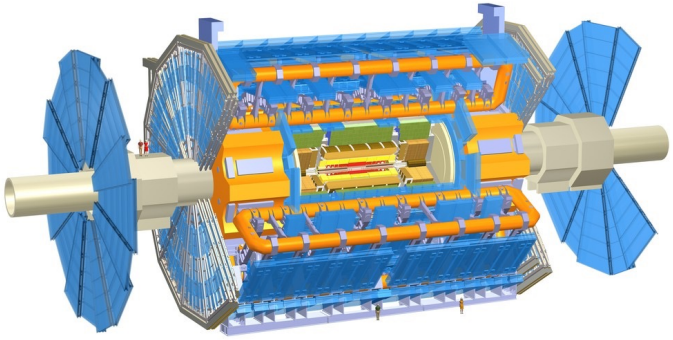
October 14, 2016



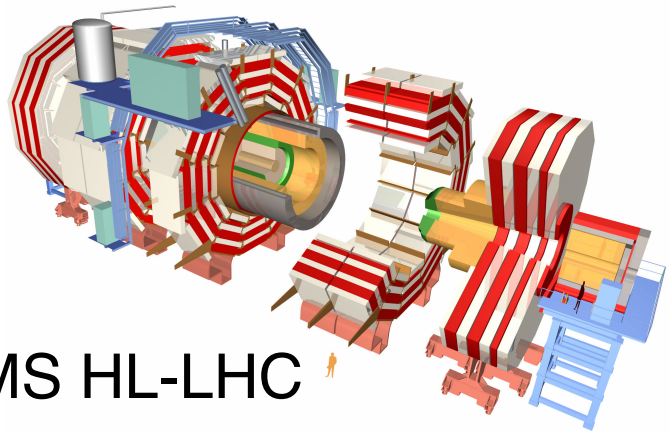
THE OHIO STATE UNIVERSITY

Introduction

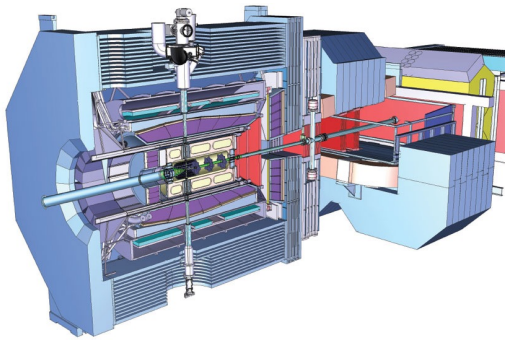
- History:
  - ➡ Using FPGAs for Track Triggers
  - ➡ State of the Art ~2000
- Modern Challenges
  - ➡ Beams and Detectors
  - ➡ HL-LHC Triggering
    - ▸ Usage of Tracking Info
  - ➡ Strategies:
    - ▸ Pattern recognition
    - ▸ Extracting Track parameters
- Future Approach
  - ➡ CMS “Tracklet" Approach
  - ➡ Current R&D
- Summary



ATLAS



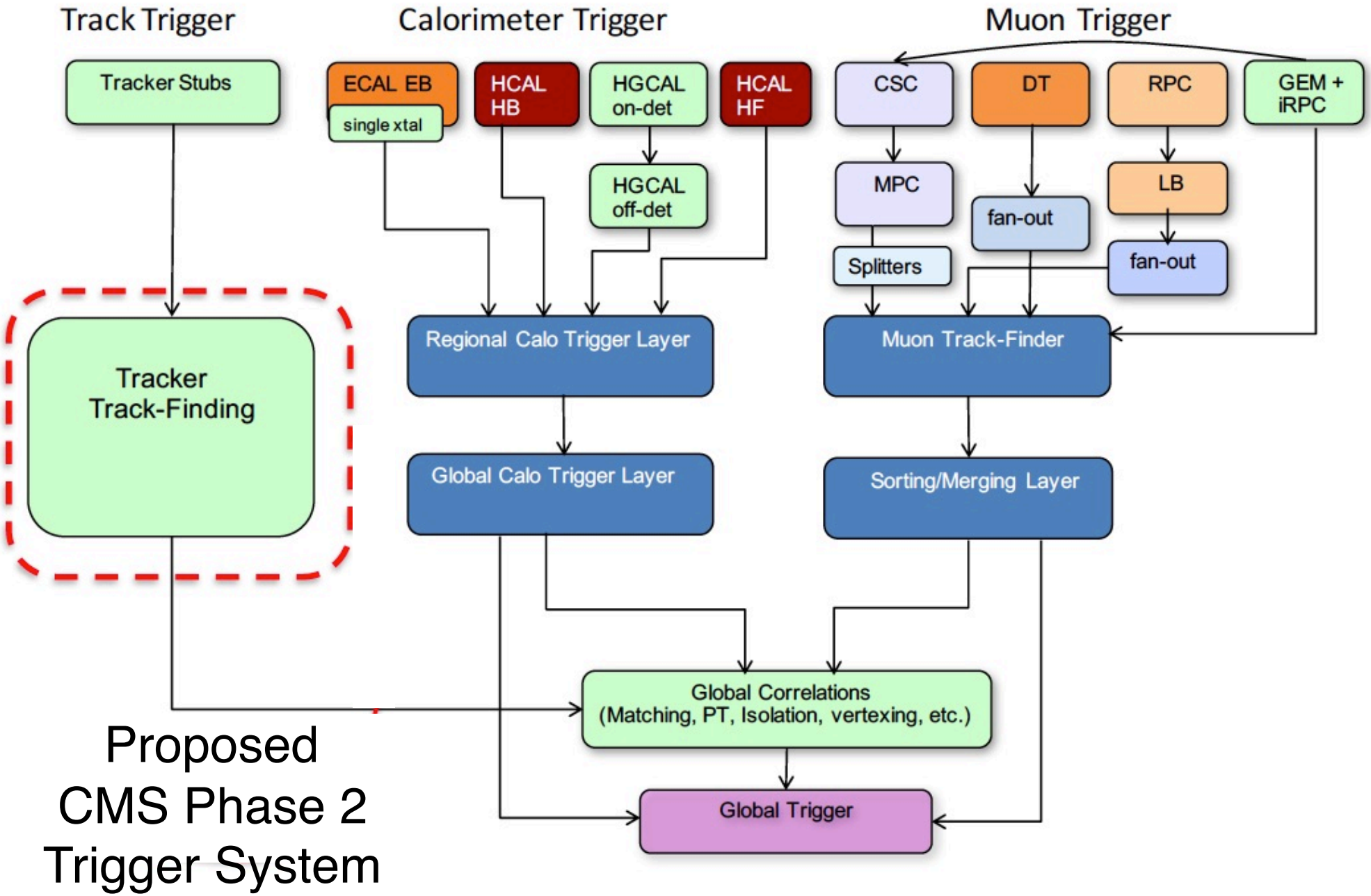
CMS HL-LHC



PANDA @ FAIR

THE OHIO STATE UNIVERSITY

Tracking — One piece

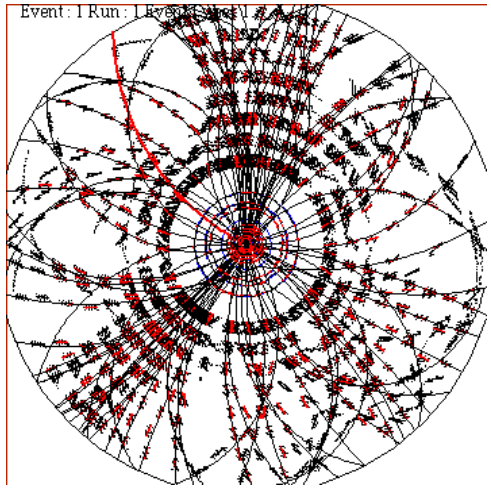


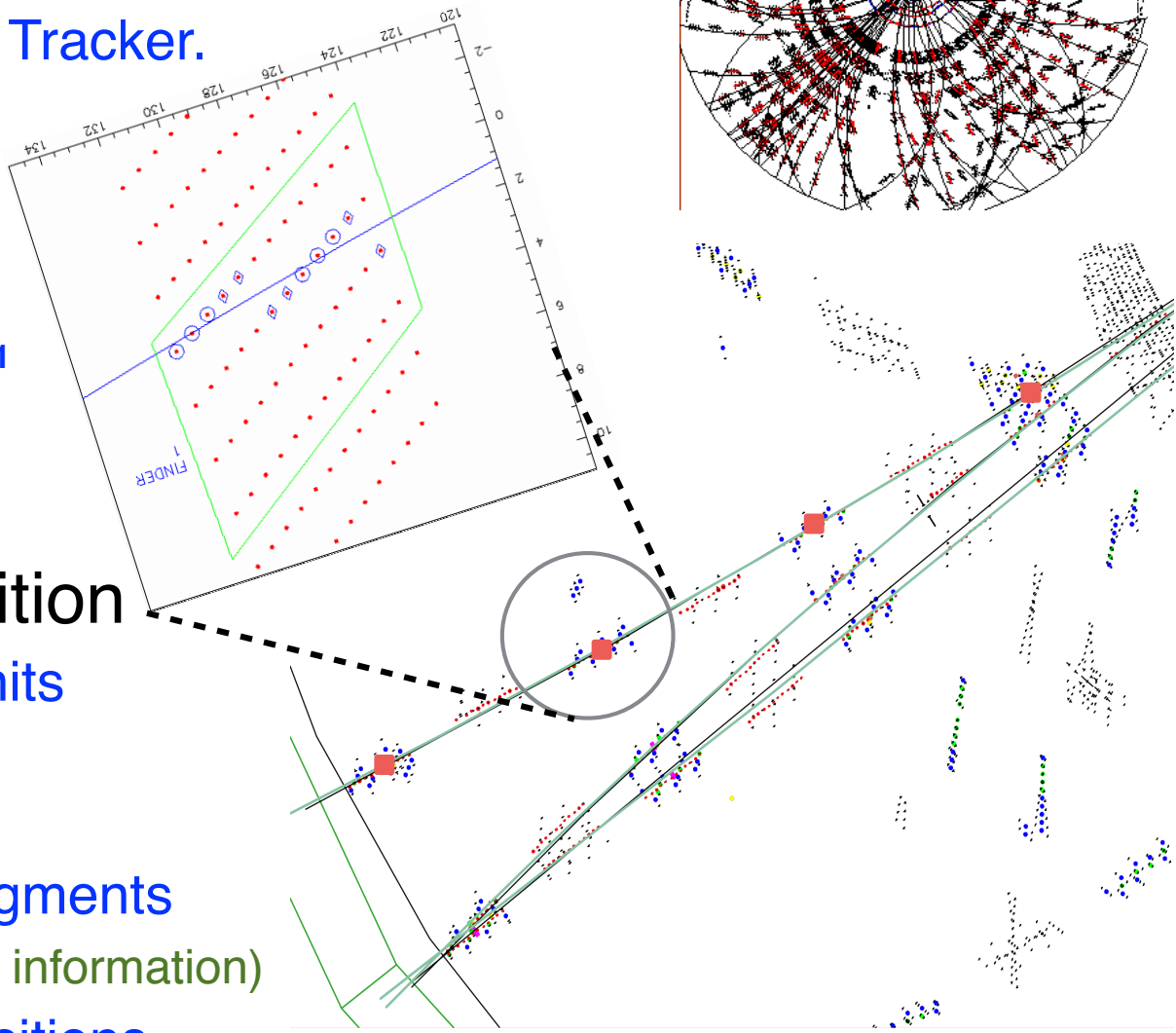
Proposed CMS Phase 2 Trigger System

THE OHIO STATE UNIVERSITY

Previous Generation

- CDF Level 1 e**X**tremely **F**ast **T**racker (XFT)
  - ➡ Used the CDF Central Outer Tracker.
  - ➡ Designed in late 1990's
  - ➡ Operated from ~2001 - 2011
- Beam Conditions
  - ➡ Luminosity: **1-3 x 10^{32} cm^{-2} s^{-1}**
  - ➡ Interactions/crossing: **3 - 9**
  - ➡ Bunch Spacing: **396 ns**
- Pipelined pattern recognition
  - ➡ 1st Stage: Coarse timing of hits
    - ▸ (2 bits "prompt" and "delayed")
    - ▸ On Front-end TDC
  - ➡ 2nd Stage: "Finding" Line segments
    - ▸ Output: position and slope (P_T information)
  - ➡ 3rd Stage: "Linking" stub positions
    - ▸ Stored patterns corresponding to possible tracks (P_T, ϕ)

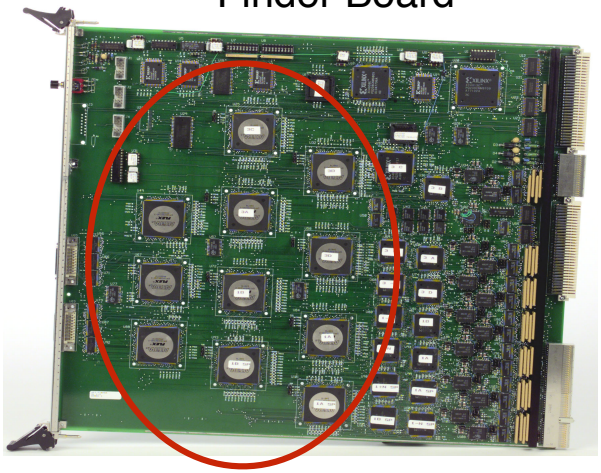




Using FPGAs

- Hardware Configuration
  - ➡ COT divided into sections for parallel processing
  - ➡ Segment Finding (48 9U VME boards)
  - ➡ Segment Linking (24 9U VME boards)

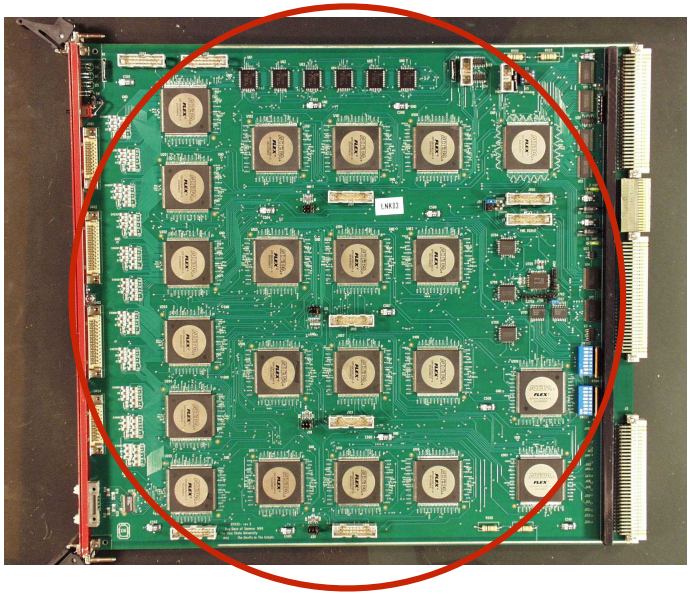
Finder Board



11 ALTERA Flex 10K FPGAs

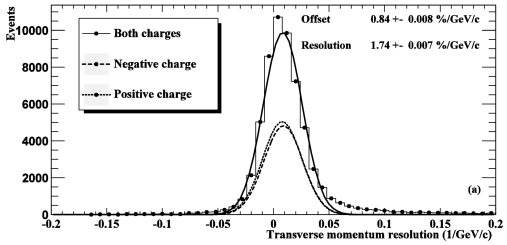
For both stages valid patterns were stored in logic and evaluated in parallel. Patterns could be adjusted for varying beam conditions (e.g. beam offset)

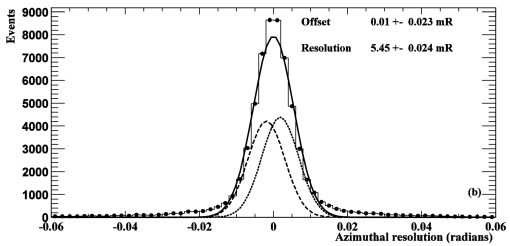
Linker Board

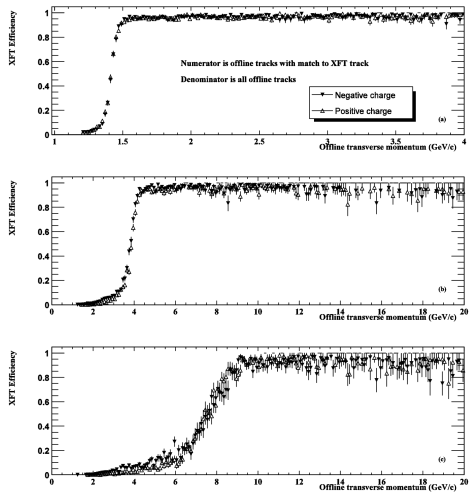


21 ALTERA Flex 10K FPGAs

- Performance:
  - ➡ P_T resolution: 1.7%
  - ➡ ϕ_0 resolution: 6 mRad
  - ➡ Efficiency > 95%



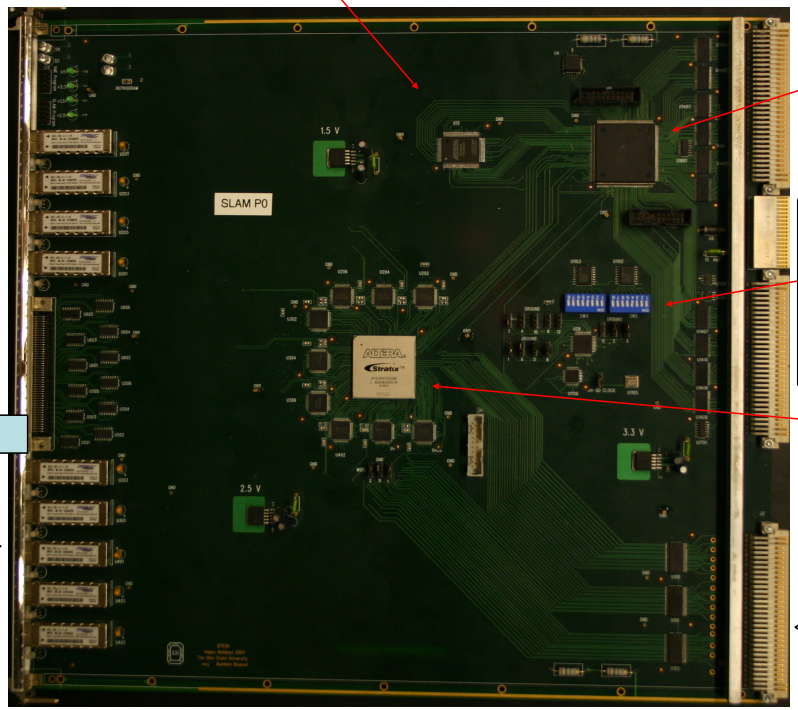




Upgrade circa 2005

XFT was upgraded from 2-D pattern recognition to “3-D”.
New FPGAs provide power.

Trending to the future:
Optical input rather than copper.

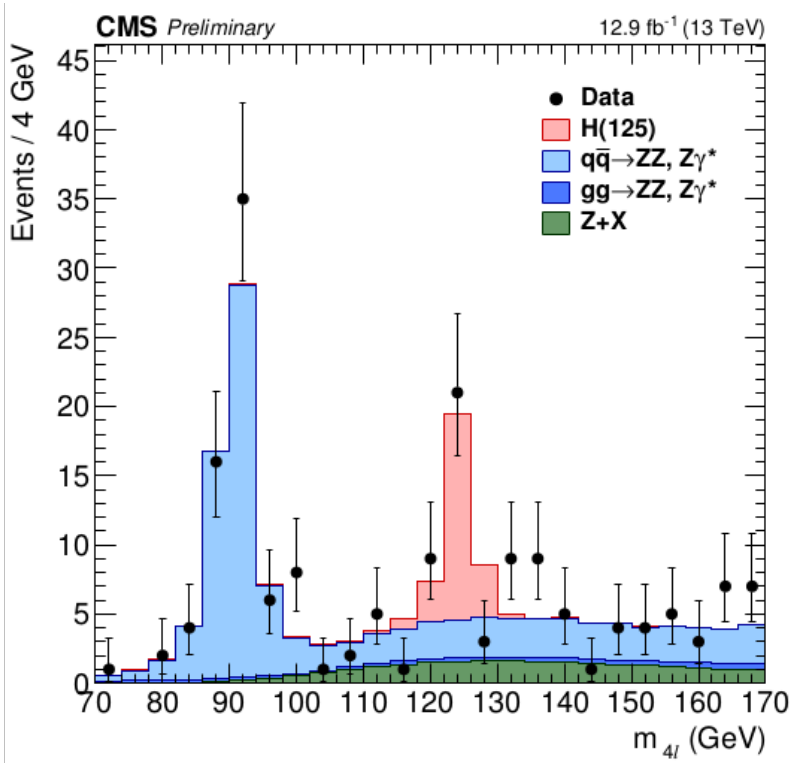


XFT was a success and provided critical input for triggering on physics signatures.

THE OHIO STATE UNIVERSITY

Modern Challenges

- Beam conditions have become more challenging
  - ➡ Luminosity: **1 x 10^{32} cm^{-2} s^{-1} —> 2 x 10^{34} cm^{-2} s^{-1} (LHC)—> 8 x 10^{34} cm^{-2} s^{-1} (HL - LHC)**
  - ➡ Pileup: 2-3 (Run 2 Tevatron) —> 50 (Run 2 LHC) —> 200 (HL-LHC)
  - ➡ Bunch spacing (396 ns —> 25 ns)
- Bandwidth and latencies
  - ➡ Bigger bandwidths but data volumes are much bigger
    - ▸ CDF COT: (wires in 2-D XFT — 16K channels)
    - ▸ HL-LHC CMS Silicon Tracker (~40 Tb/s)
      - ◉ Sparsification on front-end.
  - ➡ Latency budget for trigger
    - ▸ CDF/Tevatron 5.5/1.9 µsec Total/Tracking
    - ▸ CMS/HL-LHC 12.5/4.0 µsec Total/Tracking
- Physics analyses dictate thresholds
  - ➡ W, Z, Higgs decays define a scale
  - ➡ Want to be efficient for leptons below 40 GeV
  - ➡ Current thresholds @ CMS
    - ▸ Single Muon $P_T > 20$ GeV (~6 kHz @ 1e34)
    - ▸ Single EG (electron) $P_T > 30$ GeV (~20 kHz @ 1e34)



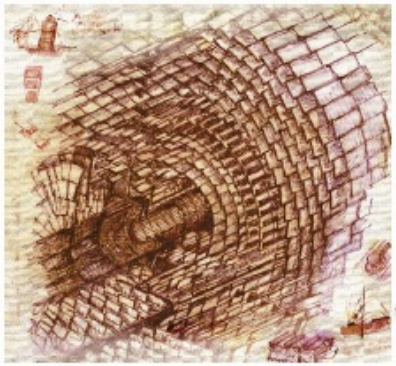
Trigger Rates at HL-LHC

- Current CMS Level-1 Trigger
  - ➡ No tracking information
  - ➡ Electrons/Gamma (EG), Taus, Jets based solely on calorimeter deposits
  - ➡ Muons reconstructed from tracks in the muon chambers
  - ➡ **Maximum Bandwidth: 100 kHz**
- HL-LHC
  - ➡ Current Trigger System
    - ▸ EG rate @ 25 GeV > 100 kHz
    - ▸ Muon rate plateaus
    - ▸ Overall Trigger rate > 1000 kHz
  - ➡ Upgraded System
    - ▸ Must increase total bandwidth
    - ▸ Must increase trigger capabilities
    - ▸ **Level-1 Tracking is a completely NEW handle.**

$L = 5.6 \times 10^{34}$ cm^{-2}s^{-1} $\langle PU \rangle = 140$	Level-1 Trigger with Level-1 Tracks	
Trigger Algorithm	Rate [kHz]	Offline Threshold(s) [GeV]
Single Mu (tk)	14	18
Double Mu (tk)	1.1	14 10
ele (iso tk) + Mu (tk)	0.7	19 10.5
Single Ele (tk)	16	31
Single iso Ele (tk)	13	27
Single γ (tk-iso)	31	31
ele (iso tk) + e/γ	11	22 16
Double γ (tk-iso)	17	22 16
Single Tau (tk)	13	88
Tau (tk) + Tau	32	56 56
ele (iso tk) + Tau	7.4	19 50
Tau (tk) + Mu (tk)	5.4	45 14
Single Jet	42	173
Double Jet (tk)	26	2@136
Quad Jet (tk)	12	4@72
Single ele (tk) + Jet	15	23 66
Single Mu (tk) + Jet	8.8	16 66
Single ele (tk) + H_T^{miss} (tk)	10	23 95
Single Mu (tk) + H_T^{miss} (tk)	2.7	16 95
H_T (tk)	13	350
Rate for above Triggers	180	
Est. Total Level-1 Menu Rate	**260**	

THE OHIO STATE UNIVERSITY

Three distinct challenges



Data transfer(50-100 Tbs)

Partition detector into trigger regions
Bring data from each region to the corresponding processing engine

- ***Associative Memory (AM), Tracklets, or TMT (UK)***

Fit hits in FPGA to determine track parameters

Data formatting

Pattern Recognition

Track fitting and duplicate removal

Δt1

Δt2

Δt3

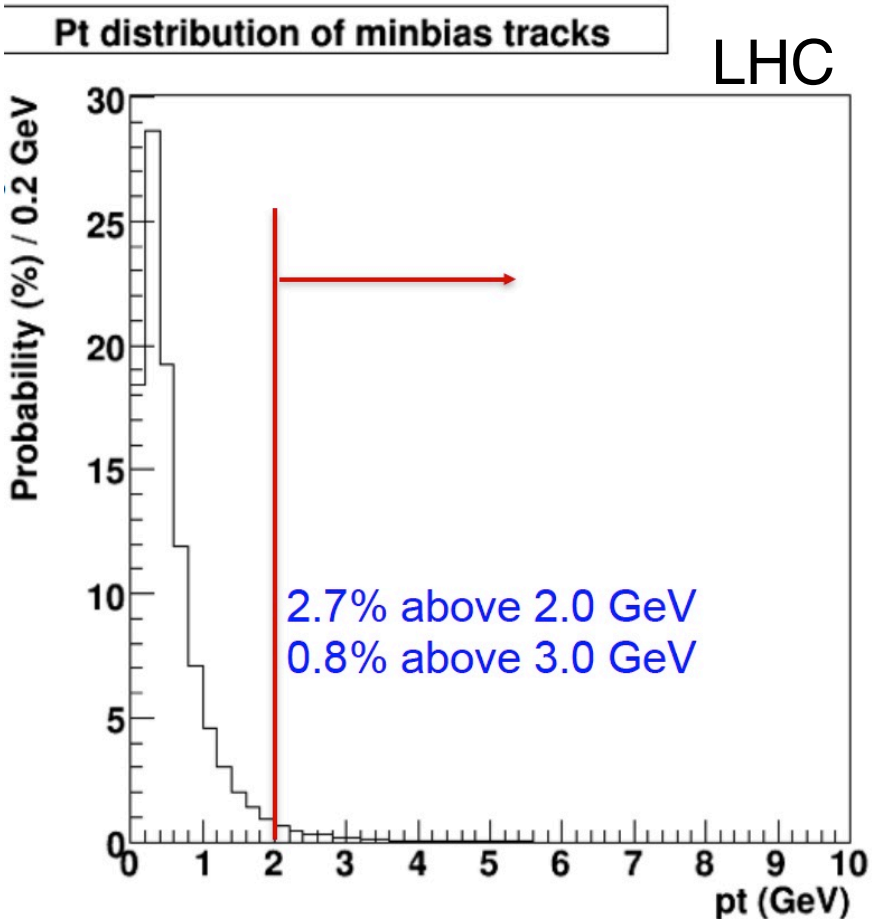
Total processing latency Δt ?

Tracks out

Focus on these two steps

Pattern Recognition Challenges

- Timing constraints
  - ➡ Fixed latency
    - ▸ Worry about tails of distributions
  - ➡ Must be able to process **every** event
    - ▸ beam crossing timescales
- Data Volume
  - ➡ High speed optical links
  - ➡ High volume of data to single point (Chip
- Combinatorics
  - ➡ Initial combinatorics are very high.
  - ➡ Must have a fast and efficient means to focus on the relevant combinations that lead to particle tracks.
  - ➡ Many different options being explored
    - ▸ A few mentioned below

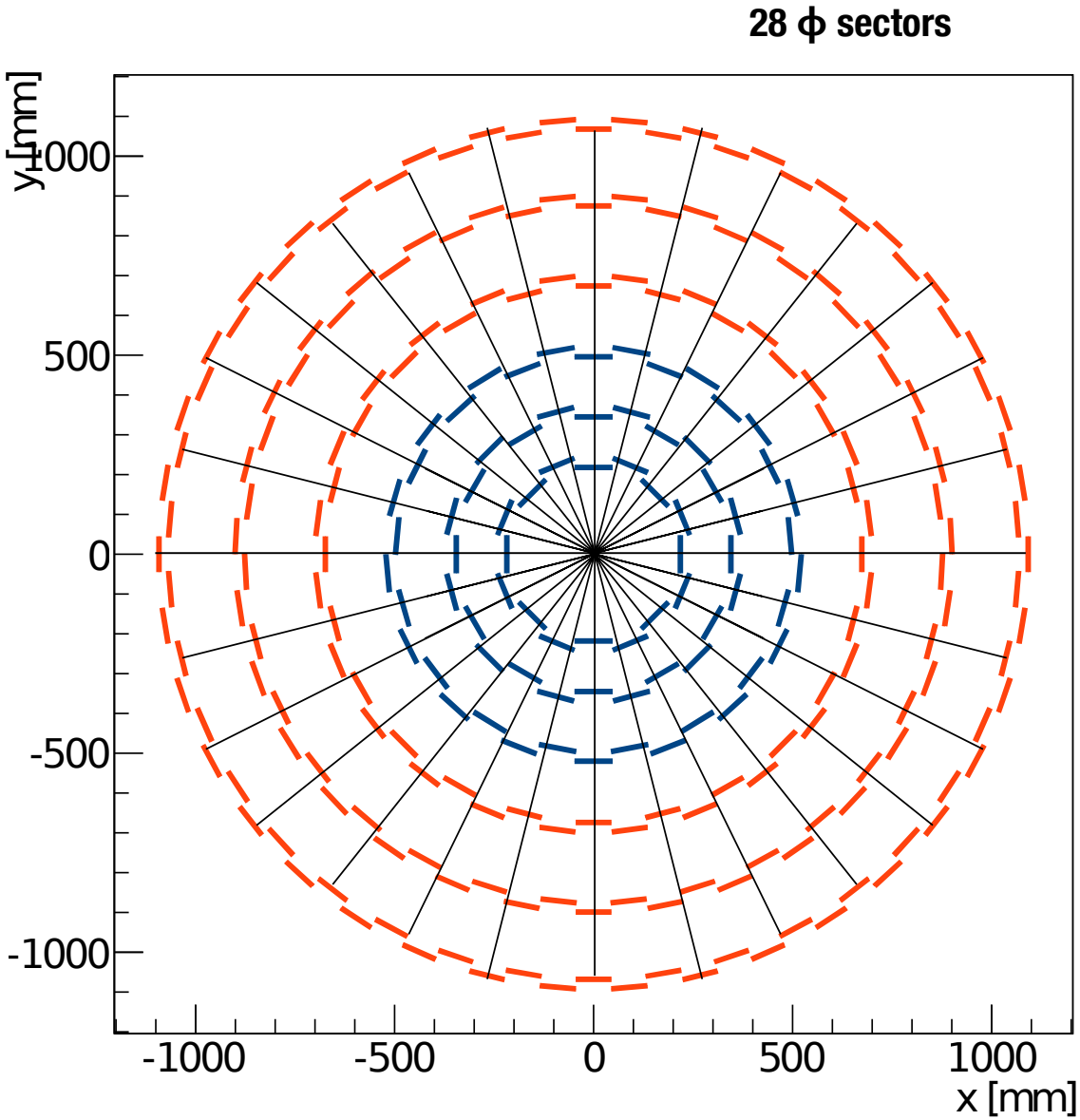


Only attempting to reconstruct a small fraction of physical tracks. However, detector hits from low momentum particles cause confusion in pattern recognition.

THE OHIO STATE UNIVERSITY

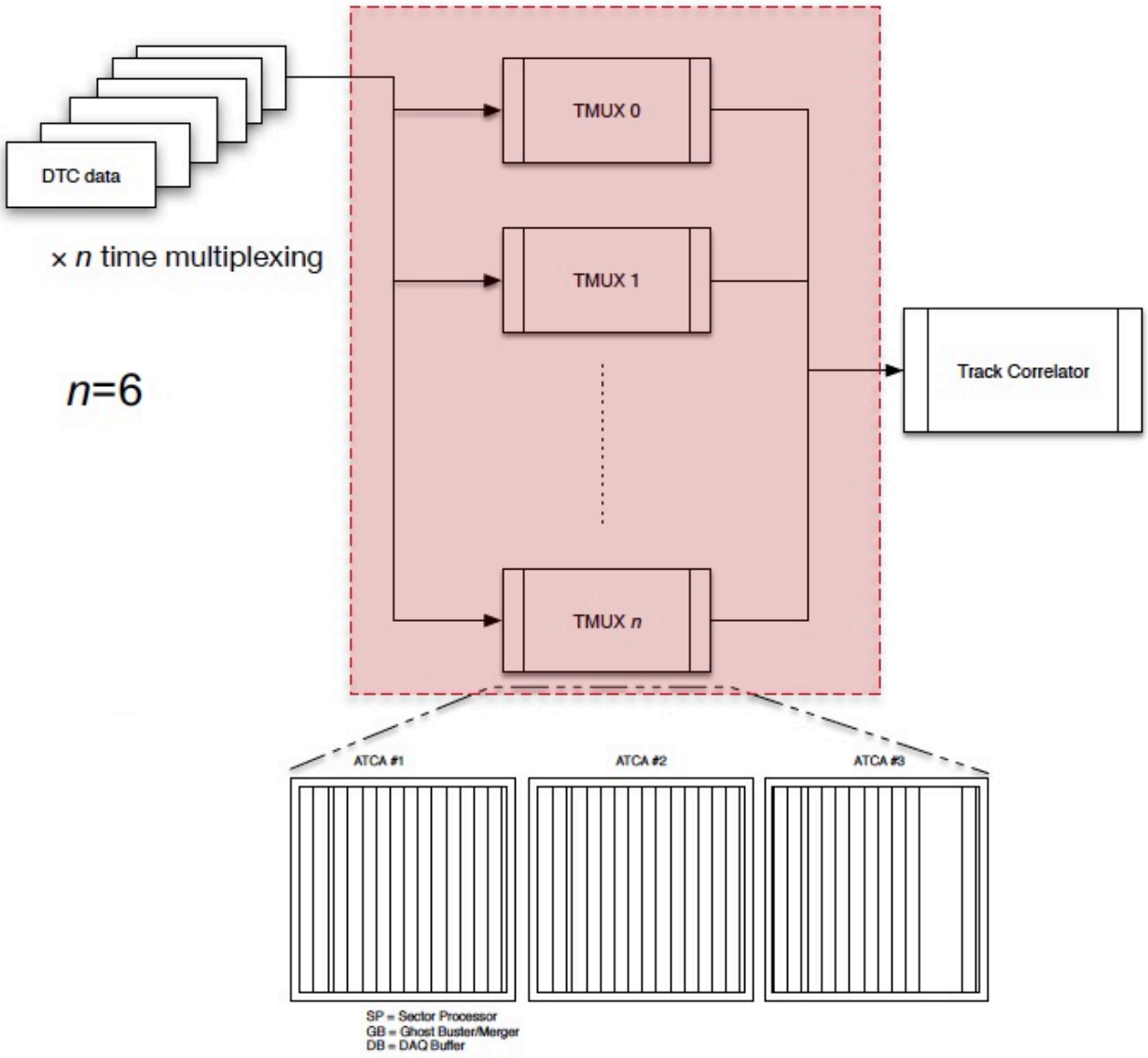
Fighting latency…

- Parallel Process
  - ➡ Different region of the detector are independent.
  - ➡ Divide detector
- Regions naturally map to boards w/ FPGAs
- Sharing between regions
  - ➡ Low P_T tracks can cross boundaries.
  - ➡ Must share information to neighbor boards.
  - ➡ Example: CMS “Tracklet”
    - 28 sectors
    - Smallest number that still restricts 2 GeV tracks to only two sectors.



Buying time…at a co$t

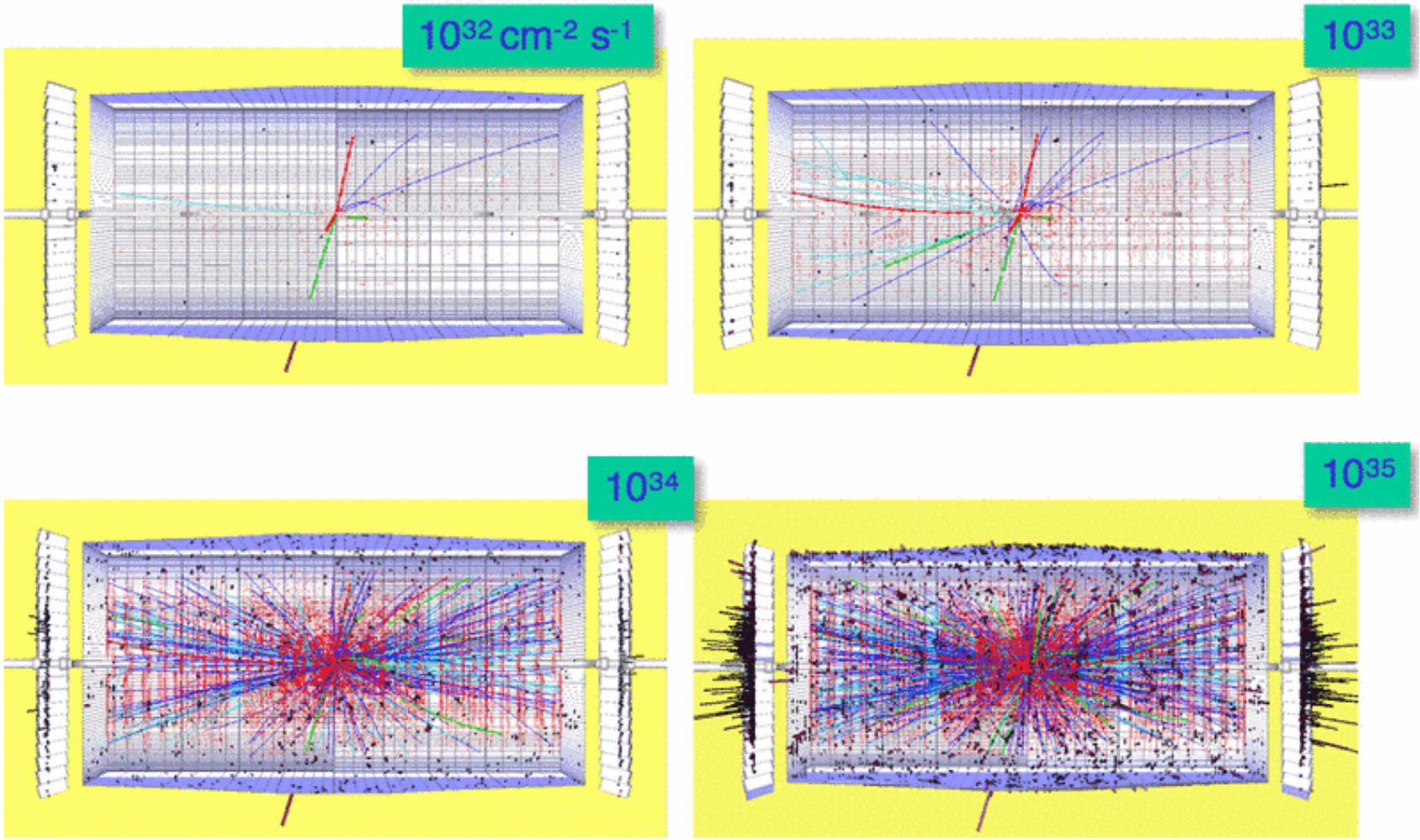
- Process each collision
  - ➡ Must be ready for new input at beam crossing time. (e.g. 25 ns)
  - ➡ Impacts time for each pipelining step.
- Extend time for each pipeline step by using **T**ime **Mu**ltiple**x**ing.
- TMux:
  - ➡ Replicate system N times
  - ➡ Route input data to each copy 1/N times
  - ➡ Processing Step: N*25 ns
  - ➡ **Hardware Costs: x N**



Other TMux approaches exist which use a large TMux Factor to reduce number of regions in the detector.

THE OHIO STATE UNIVERSITY

Reducing Combinatorics



Pattern Recognition must quickly reduce combinations of hits as a first step.

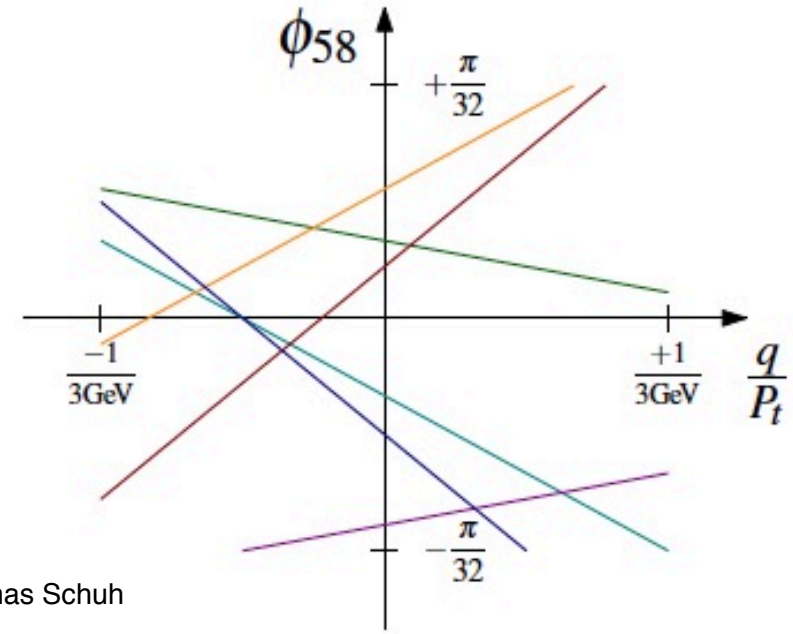
* Associative Memory (AM): Normally dedicated ASIC
* Hough Transform: Good for FPGA —"binning approach"
* Finding small Track Segments ("Tracklet" — more later)

See Talk by Marco Trovato (Tues)

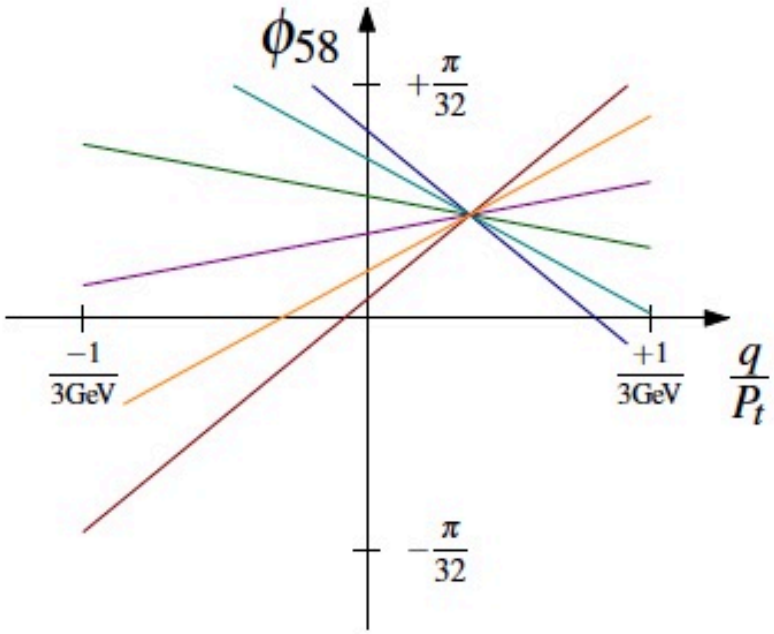
THE OHIO STATE UNIVERSITY

Hough Transform Approach

- Considered by several different experiments
  - ➡ PANDA, Belle II, option for CMS Phase 2 Track Trigger.
  - ➡ Suited for FPGA (Histogram approach)
- Transform each detector hit from x,y to ϕ, q/P_T
  - ➡ Each hit transforms to a line in ϕ, q/P_T space.
  - ➡ Hits from the same physical track will form intersecting lines.



6 random stubs

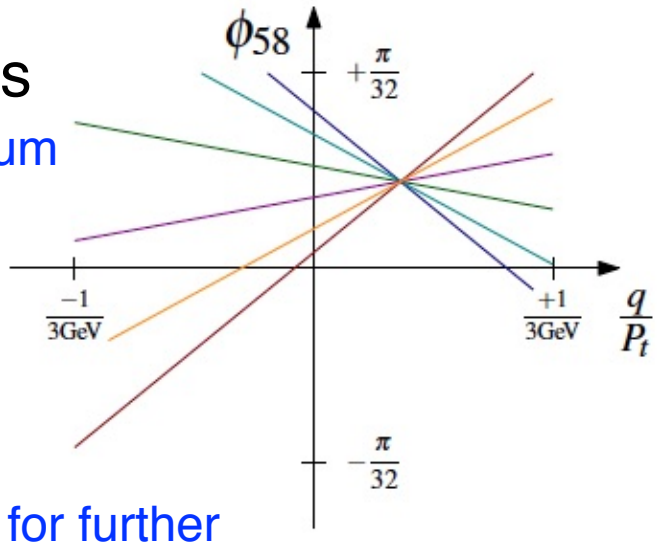


6 stubs from same particle

Fig: Thomas Schuh

Hough Transform

- Divide the space into bins
  - ➡ In each bin, require a minimum number of unique layers contributing to bin.
- Advantages
  - ➡ Fast: once bins are filled all tracks are found.
  - ➡ Easy to P_T order candidates for further processing.
  - ➡ Can follow this stage with more precise fitting of hits
- Some limitations from bin size
  - ➡ Mitigate with clever binning approaches.



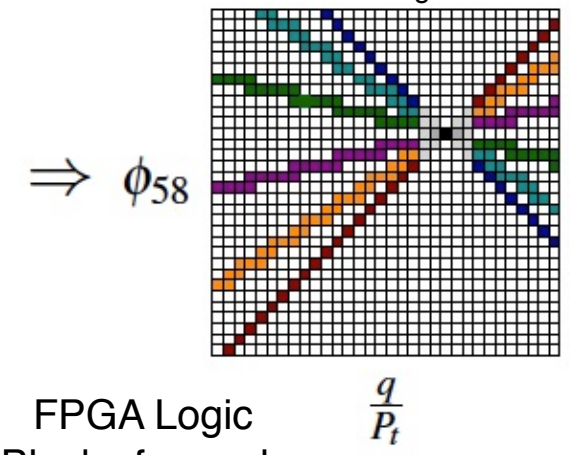
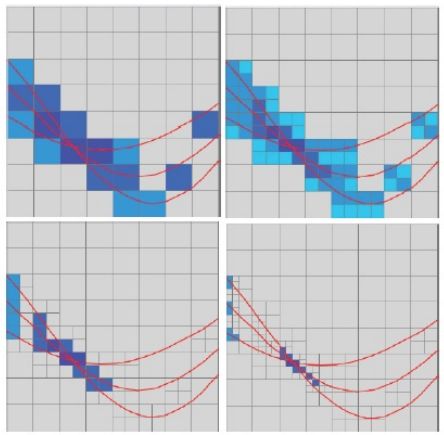


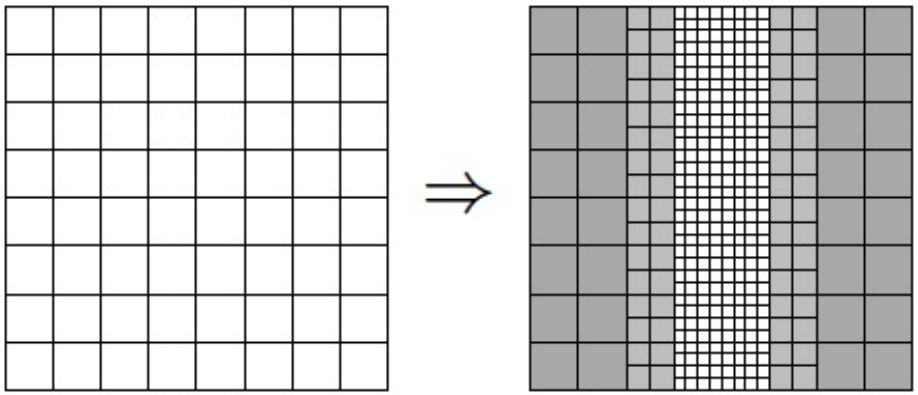
Fig: Thomas Schuh

FPGA Logic Blocks for each bin in histogram

PANDA/Belle II: adaptive binning approach. Iterations with smaller bin sizes.



CMS Approach investigating inhomogeneous binning

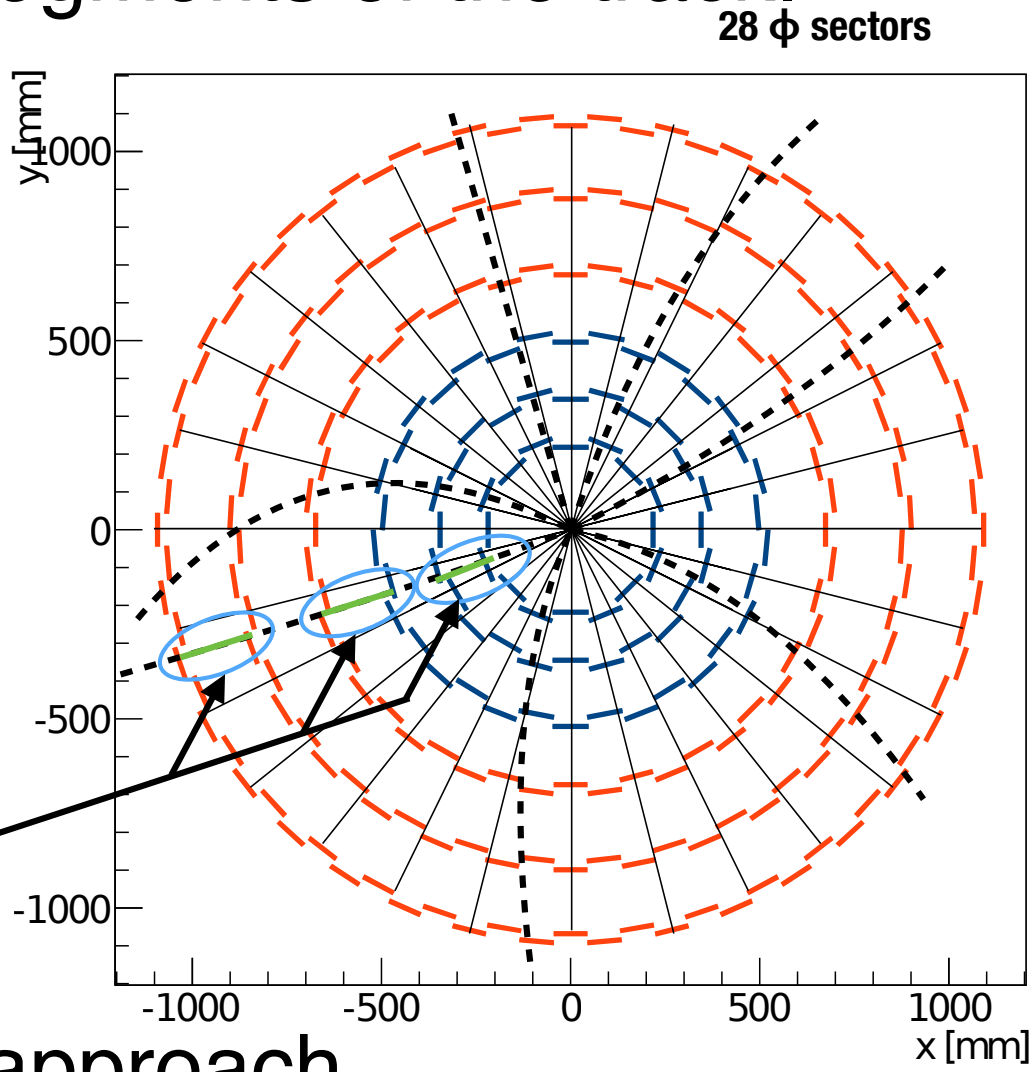


THE OHIO STATE UNIVERSITY

"Tracklet Approach"

Combinatorics can also be reduced by searching for small segments of the track.

CMS is considering this approach for the Phase 2 HL-LHC upgrade.

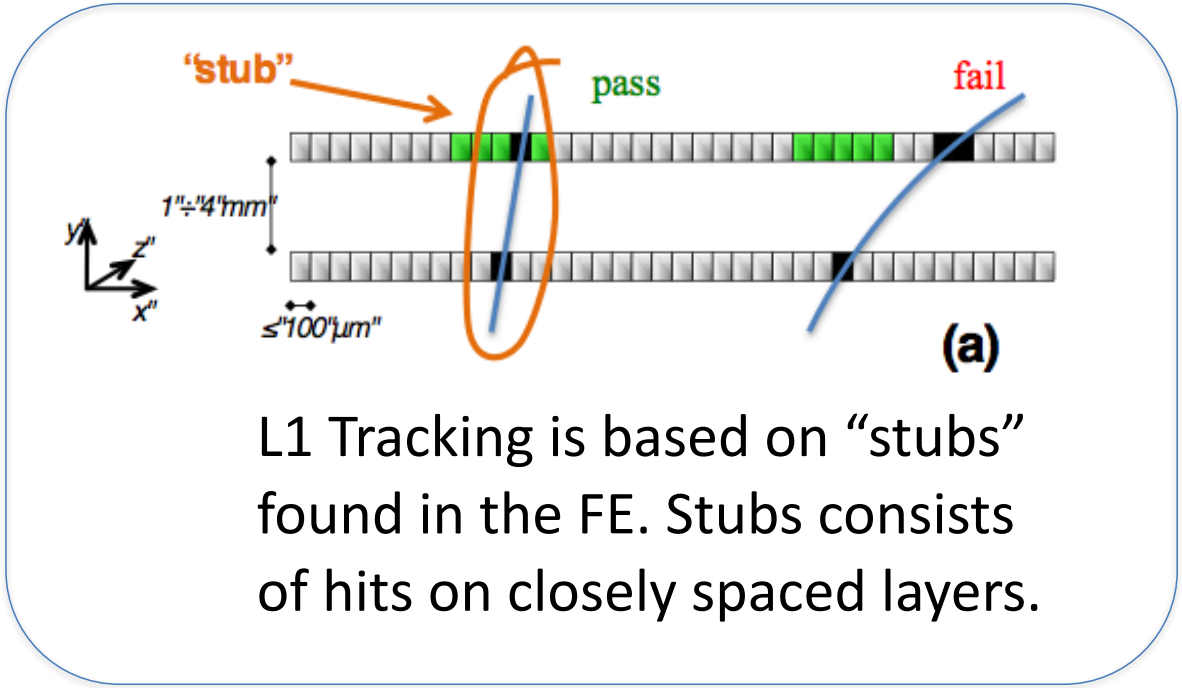


* Search for small segments of tracks on adjacent layers.

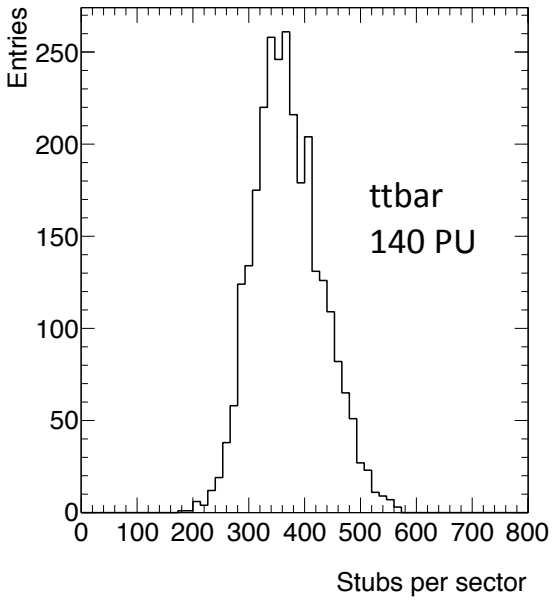
Go deeper into this approach…

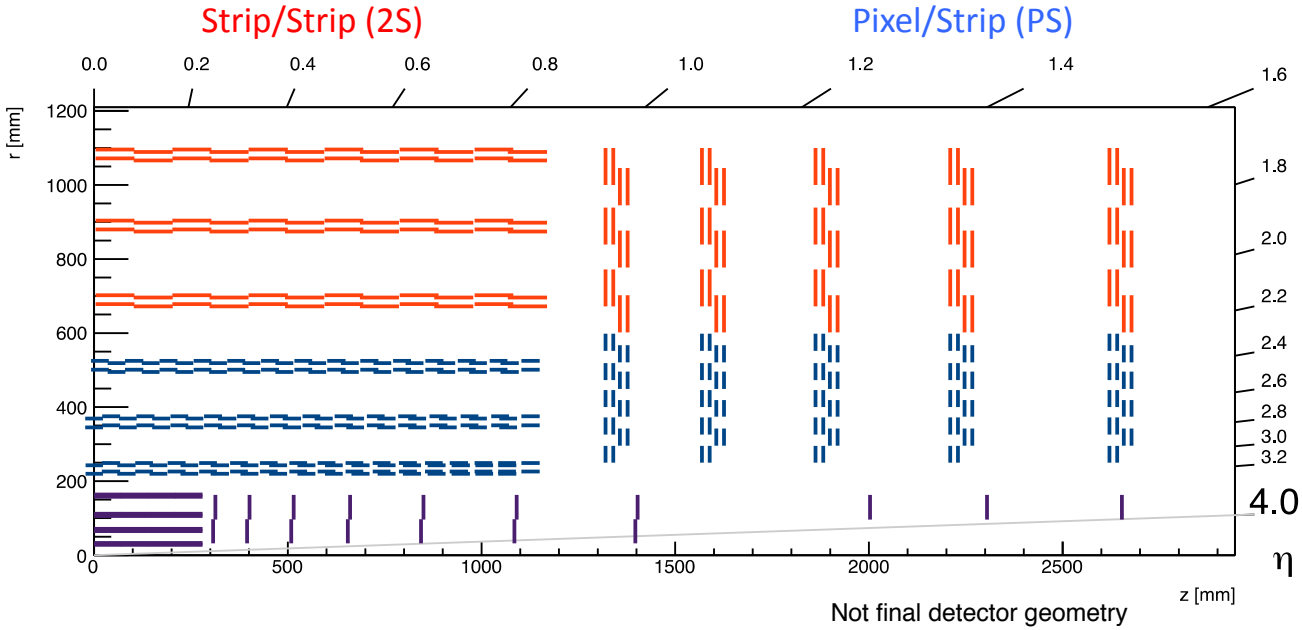
CMS Tracking Input

- Silicon pixel/strip Tracker
  - ➡ Inner pixel detector
  - ➡ Fwd Pixel Detector
  - ➡ Outer Pixel/Strip Detector
- Outer Tracker Geometry
  - ➡ Barrel: 6 layers
  - ➡ Endcaps: 5 disks.
- Each beam crossing
  - ➡ ~125 charged particles $P_T > 2.0$ GeV
  - ➡ ~10,000 stubs (~10% assoc. with $P_T > 2$ GeV)
  - ➡ Input bandwidth 20 - 40 Tbits/sec



L1 Tracking is based on "stubs" found in the FE. Stubs consists of hits on closely spaced layers.

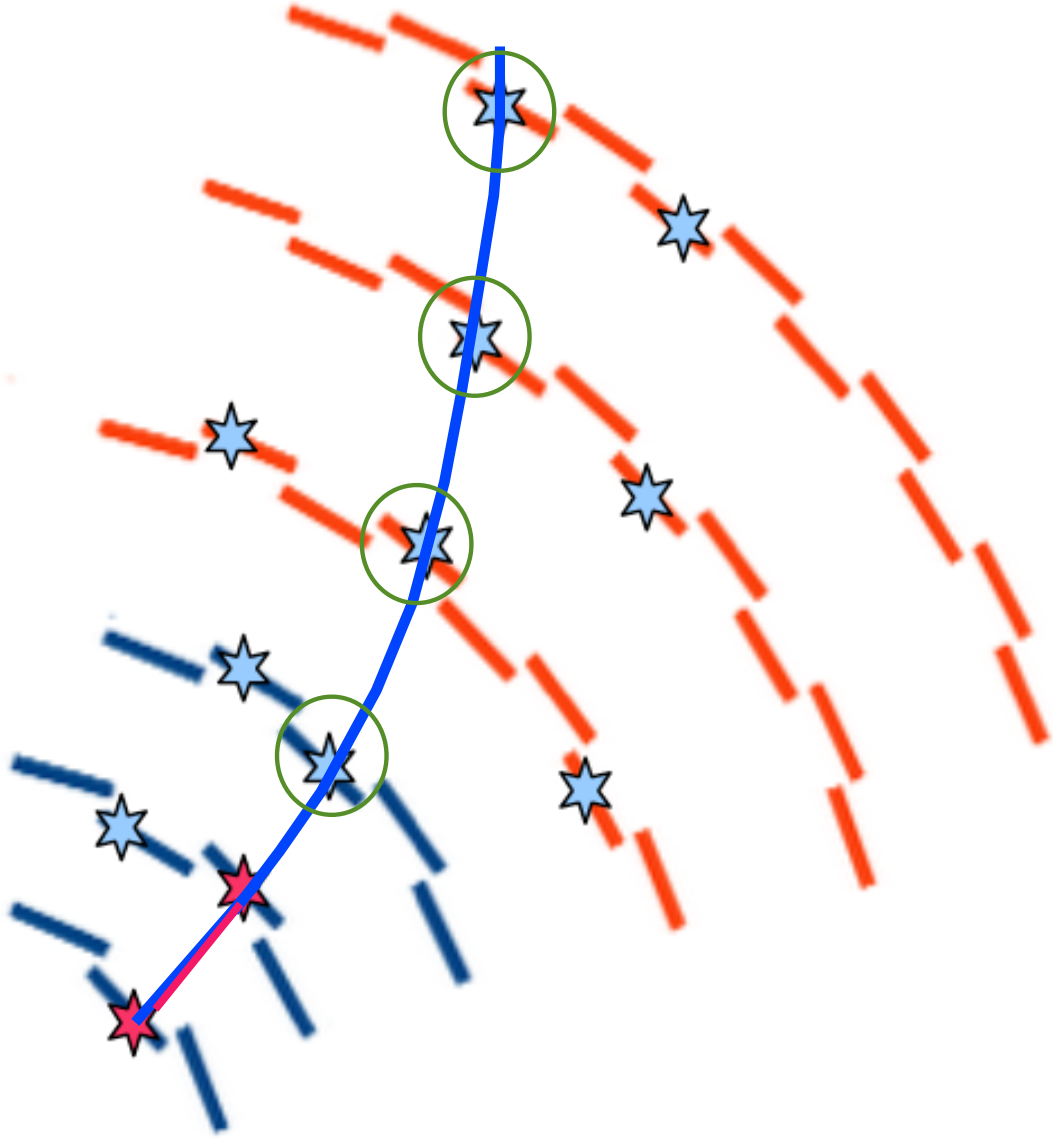




THE OHIO STATE UNIVERSITY

Basic Approach

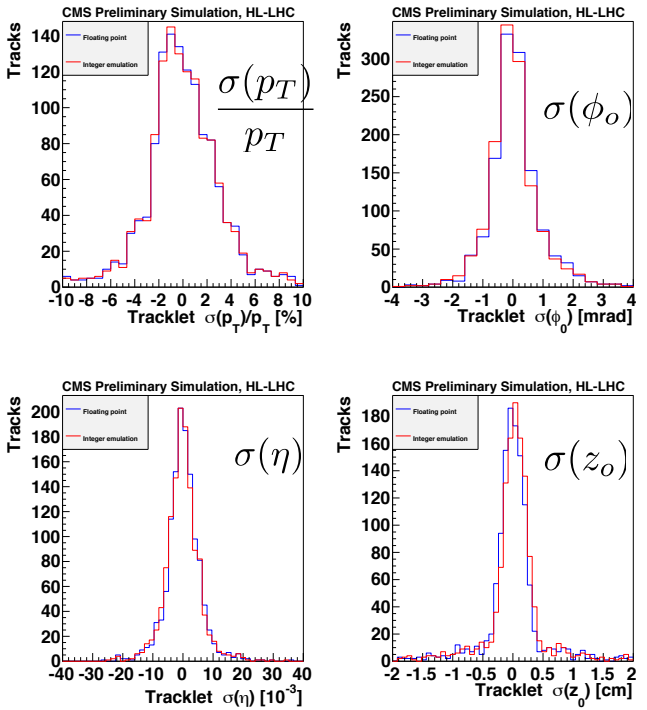
- (1) Form "Tracklets"
  - ➡ Stubs from adjacent layers
  - ➡ Find initial track parameters using beam constraint
- (2) Project into other layers
  - ➡ Locate additional stubs
- (3) Perform a fit to all stubs
  - ➡ Linearized χ^2 fit
  - ➡ Extract final track parameters
- (4) Remove duplicate tracks
  - ➡ Seeding tracks on multiple layers leads to the same particle being found multiple times.

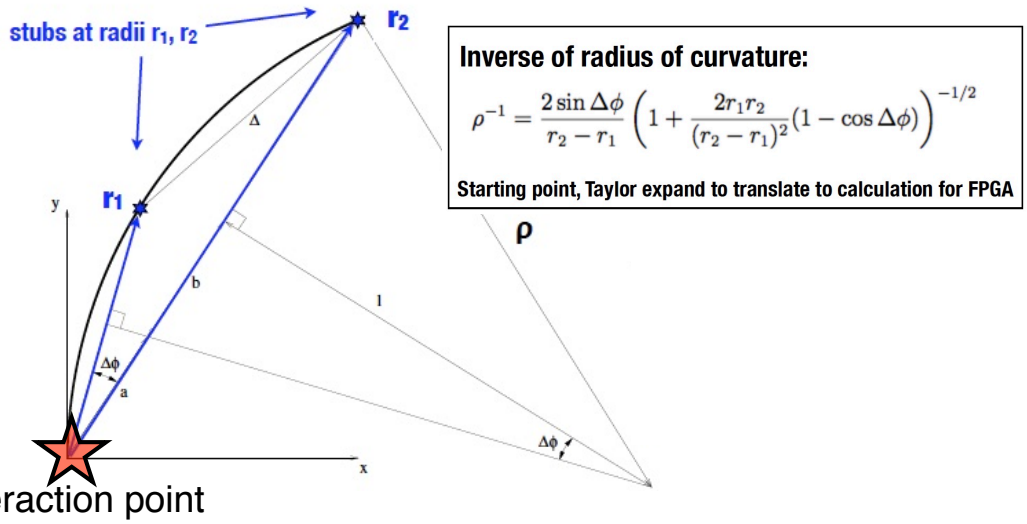


Stage 1: Tracklet Formation

- Tracklets are combinations of stubs in adjacent layers
  - ➡ Barrel Layers:
    - L1+L2, L3+L4, L5+L6
  - ➡ Disk Layers:
    - D1+D2, D3+D4
  - ➡ Overlap:
    - L1+D1, L2+D1
  - ➡ Tracklets must be consistent with $P_T > 2$ and $|z_o| < 15$ cm
- Initial Track parameters
  - ➡ Two stub positions
  - ➡ Beam Constraint
- Even these simple tracks have good resolution on track parameters.
  - ➡ Important for projection phase.

Tracklet Parameter Resolutions

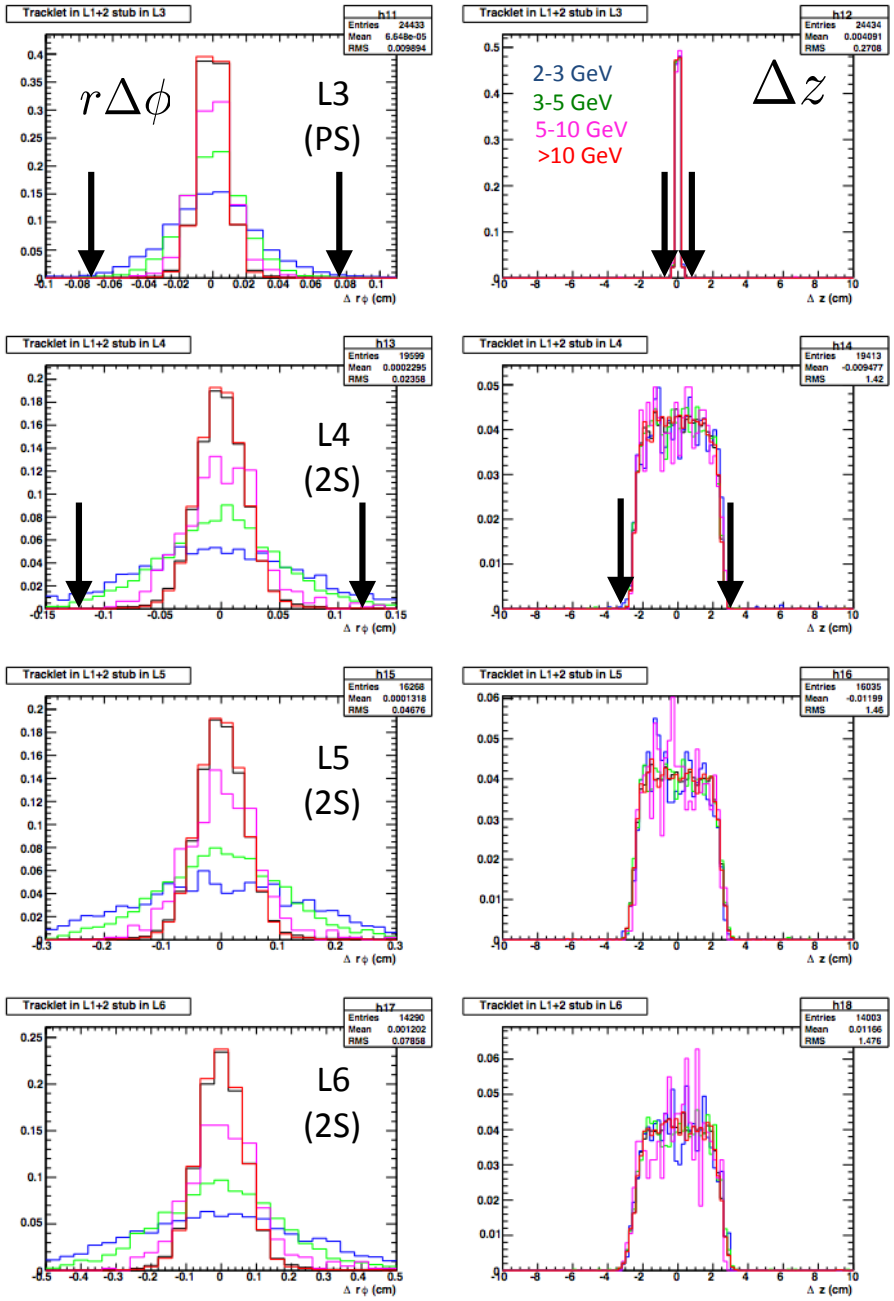




THE OHIO STATE UNIVERSITY

Stage 2: Tracklet Projections

- Tracklets are projected to other layers
  - ➡ Search for hits in a window around extrapolated P_T.
  - ➡ Window widths based on resolution.
  - ➡ Example plot: L12 tracklet projected to other layers.
- Find hit with smallest residual in each layer
  - ➡ Hits kept for full track fitting stage.
  - ➡ Some extra handles available such as sign/slope of stub.
    - ▸ Can ensure charge/P_T consistency
    - ▸ Not used in the current design
- Projection may be in neighboring sector (board)
  - ➡ Hits/Residuals must be communicated between boards.

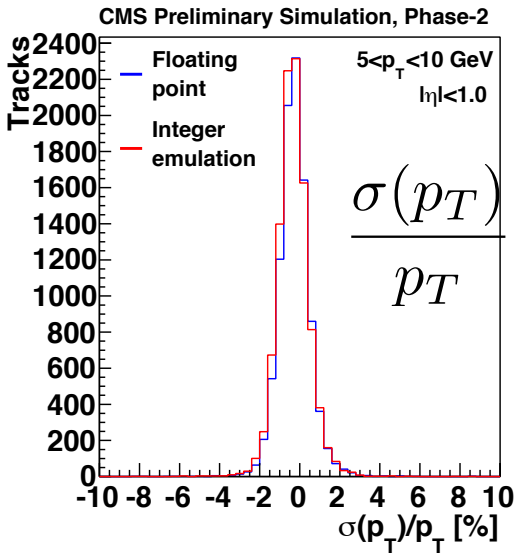


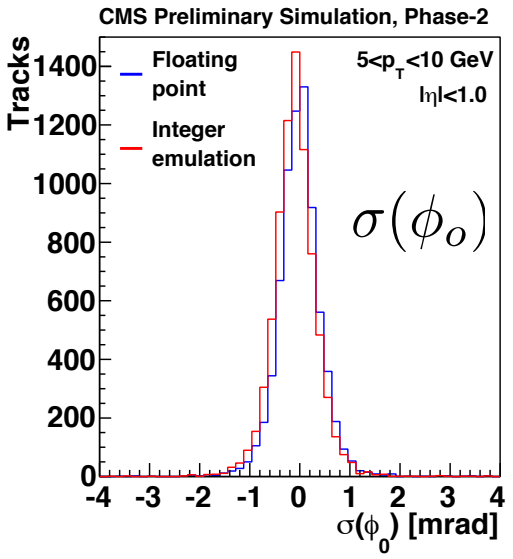
THE OHIO STATE UNIVERSITY

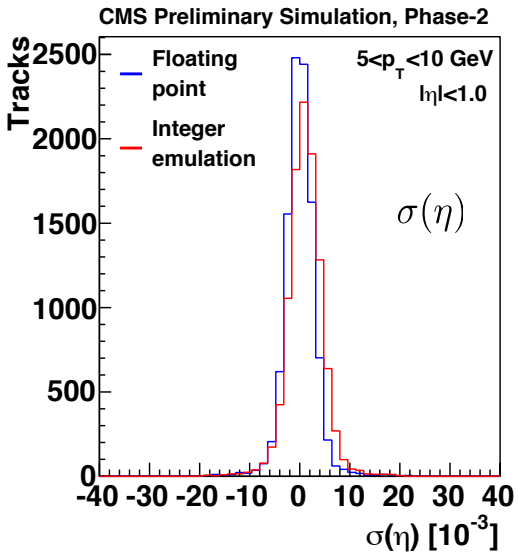
Stage 3: Track Fitting

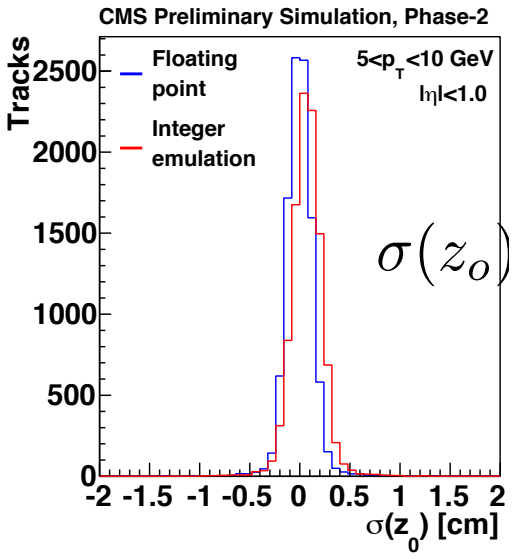
- Track Fitting
  - ➡ Use all hits, both from tracklets and projections.
  - ➡ Use linearized χ^2
    - Series of simple computational steps
    - More complex computations are tabulated in LUTs.
    - Good agreement between integer and floating pt calculations.
  - ➡ Heavy use of FPGA DSPs
- Extract Track parameters
  - ➡ P_T , ϕ_o , z_o , η
  - ➡ d_o (optionally)
- Use the χ^2 as a measure of track quality.
  - ➡ Can be used downstream in the trigger.
  - ➡ e.g. track entering a MET calc.

Tracks in Barrel: Floating Pt vs Integer





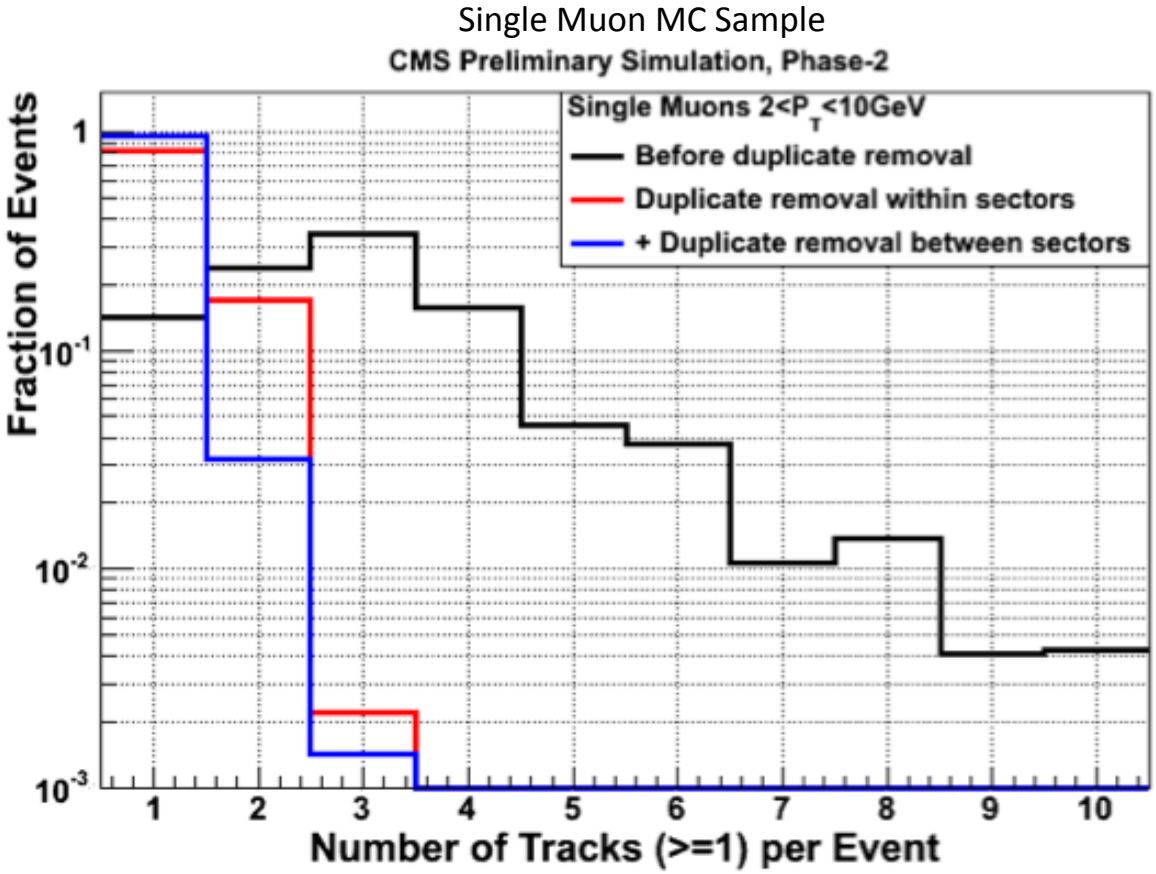




Stage 4: Duplicate Removal

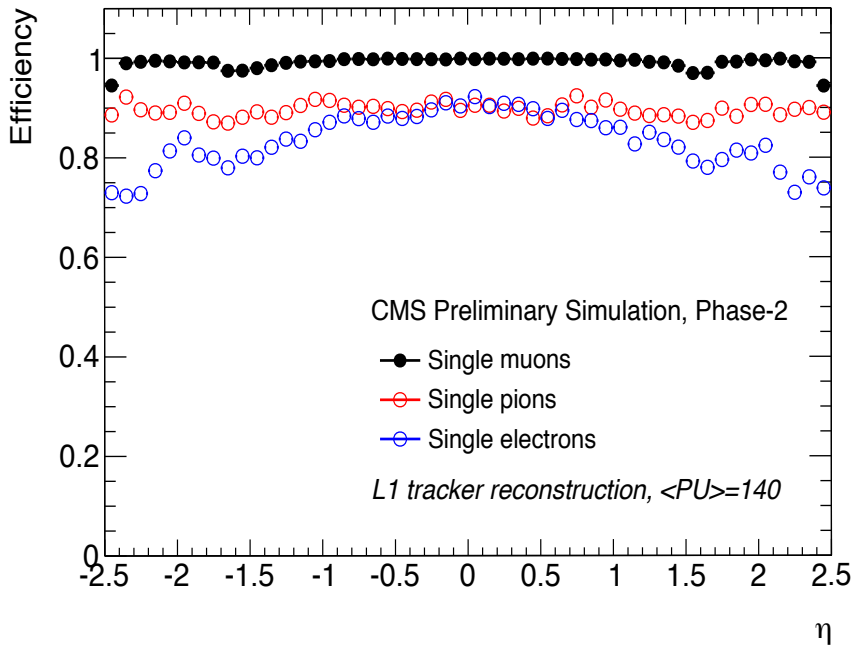
- Each physical particle can generate multiple tracks during the pattern recognition process
  - ➡ Good for efficiency and robustness against detector failures
  - ➡ Must pare out these duplicates to avoid flooding downstream trigger system.

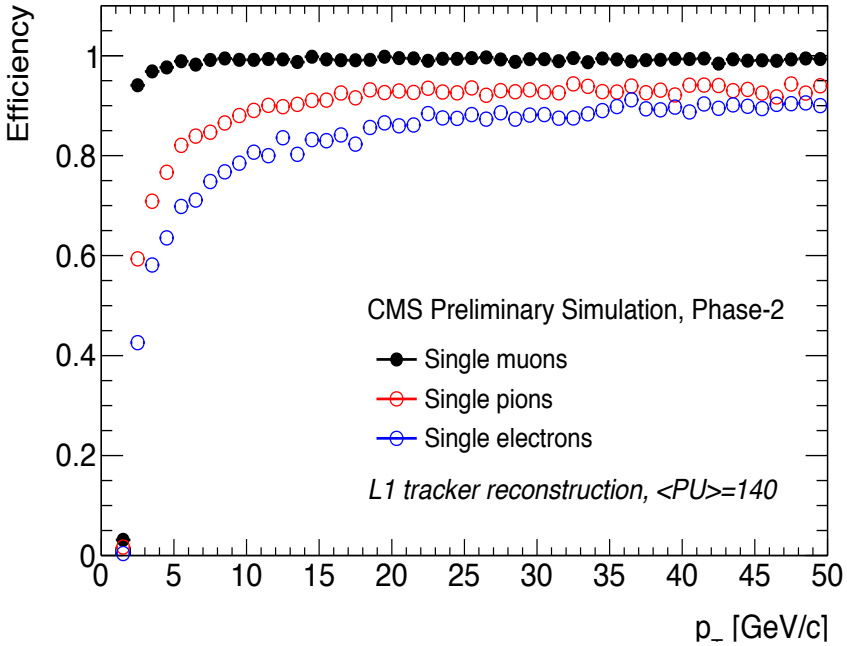
- Removal Algorithm:
  - ➡ Compare pairs of tracks and count the number of shared stubs.
  - ➡ Select on the #indep. stubs
  - ➡ If many shared stubs, remove one of the tracks as a duplicate.
  - ➡ Must consider tracks found on adjacent boards.

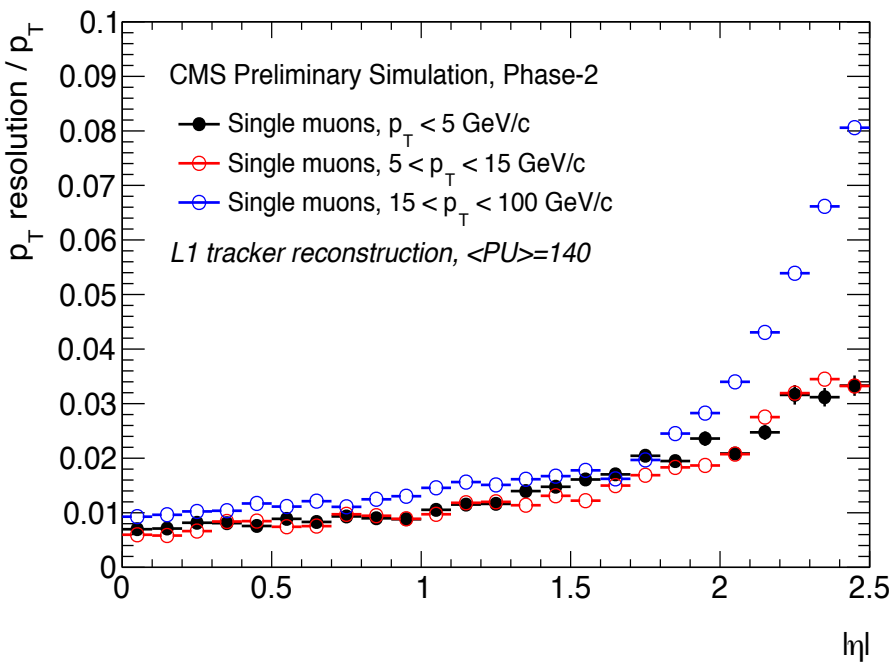


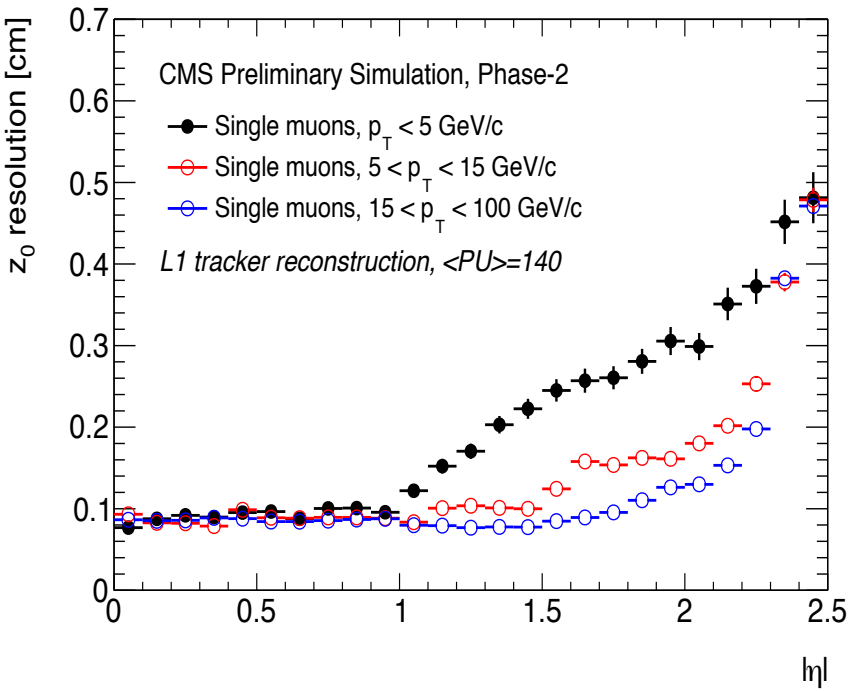
THE OHIO STATE UNIVERSITY

Performance







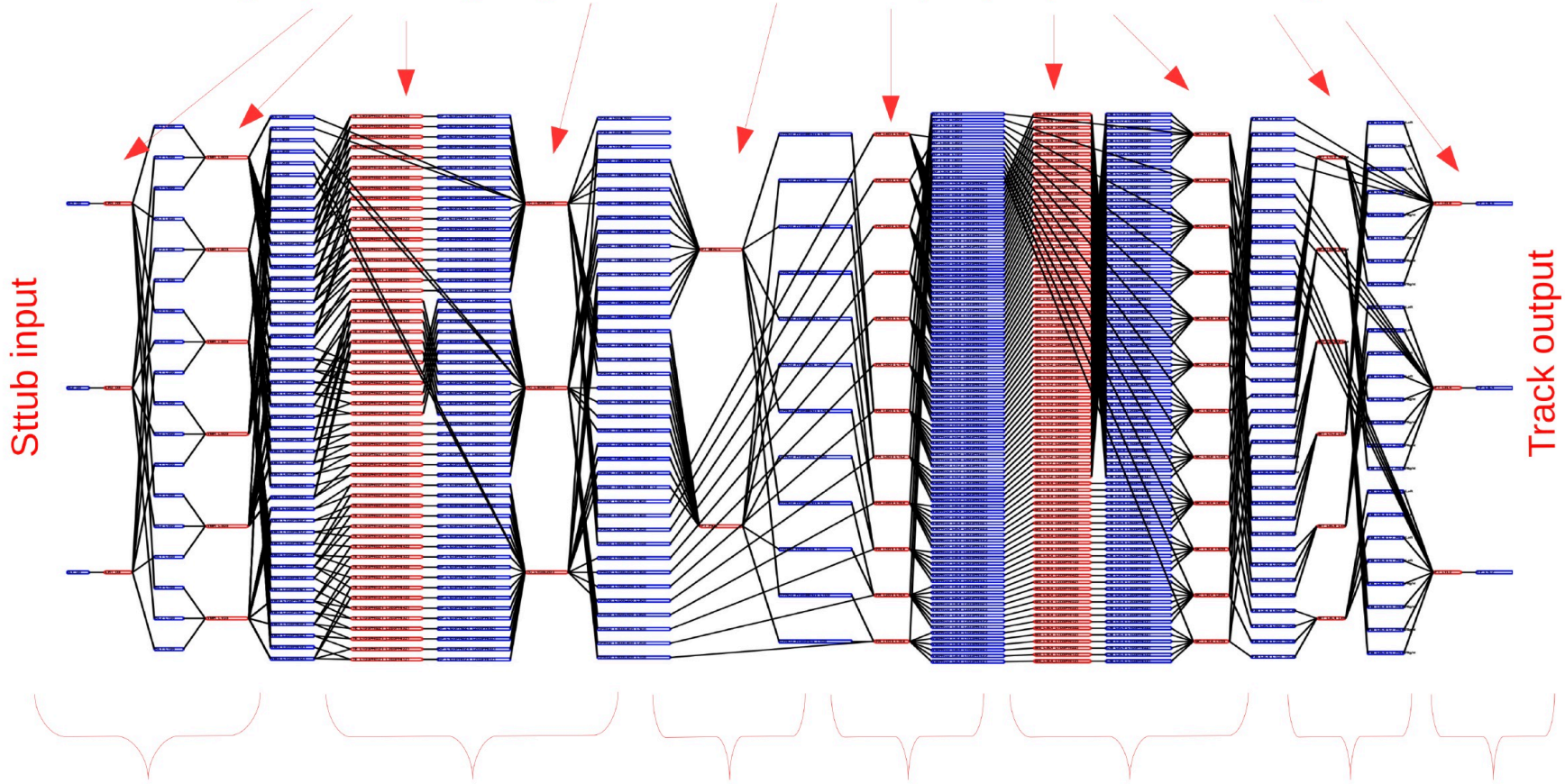


THE OHIO STATE UNIVERSITY

Firmware Implementation

Python code generates Verilog code, emulator configuration, and this picture

Eight processing steps + two transmission (red) implements the algorithm



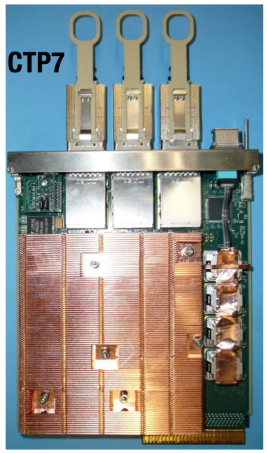
Vertex-7 is not large enough to hold a design for the full fiducial volume. We have separate designs that implement the different regions of the detector.

THE OHIO STATE UNIVERSITY

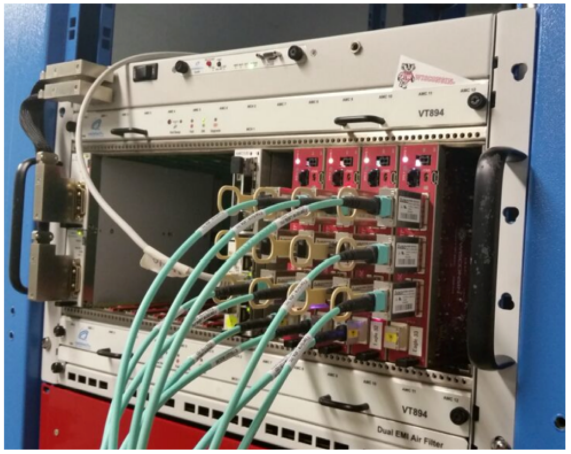
R&D Testing

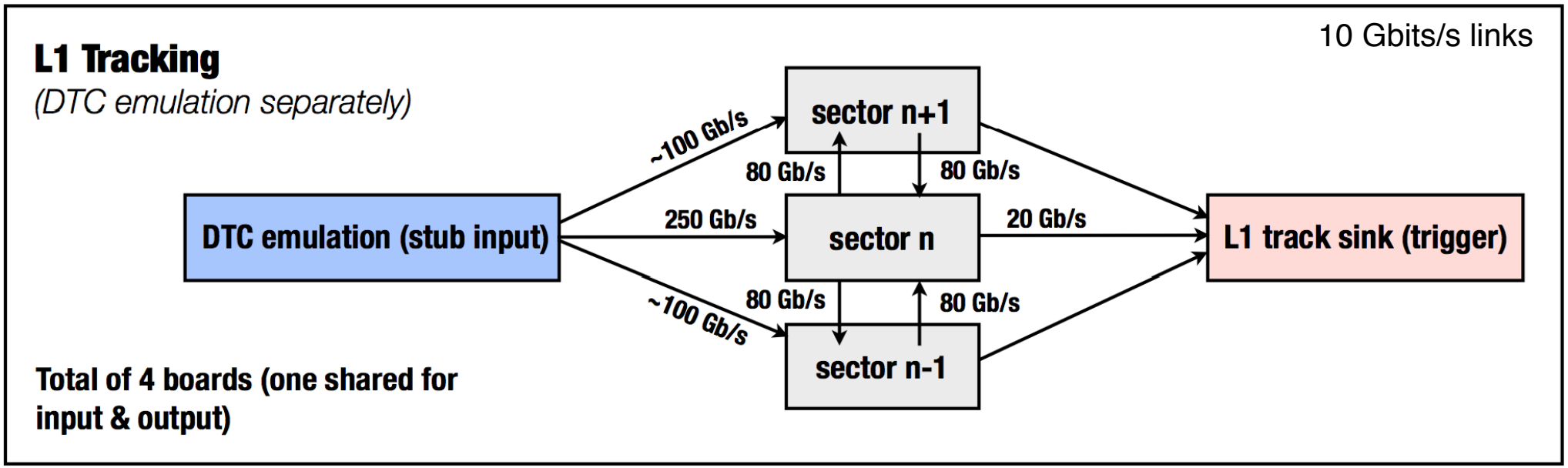
- Hardware: Existing Trigger Boards
  - ➡ 4 CTP7 boards in uTCA
    - 3 Sector Boards
    - 1 Data Source/Sink
  - ➡ Represents 1 of N TMux Slices
  - ➡ Neighbor communication
- Firmware: Portion of acceptance

Developed at Univ. Wisconsin



Test Stand at CERN (4 CTP7)





THE OHIO STATE UNIVERSITY

Summary

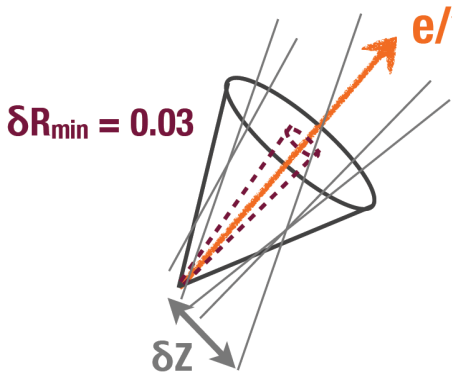
- Tracking information is very useful for trigger system
  - ➡ Helps control trigger rates.
  - ➡ Keeps thresholds lower for more physics acceptance
- Tracking is a pattern recognition problem
  - ➡ Conducive to parallel processing
  - ➡ Conducive to pipelining
  - ➡ Well suited for FPGAs
- FPGAs are an excellent approach
  - ➡ “Cheap” hardware costs
  - ➡ Flexibility: Adapt to changing conditions/requirements
  - ➡ Modern FPGAs have many built in capabilities that allow for complex algorithms
- Track Triggers with FPGAs have been used for several decades and will be part of the next generation as well.

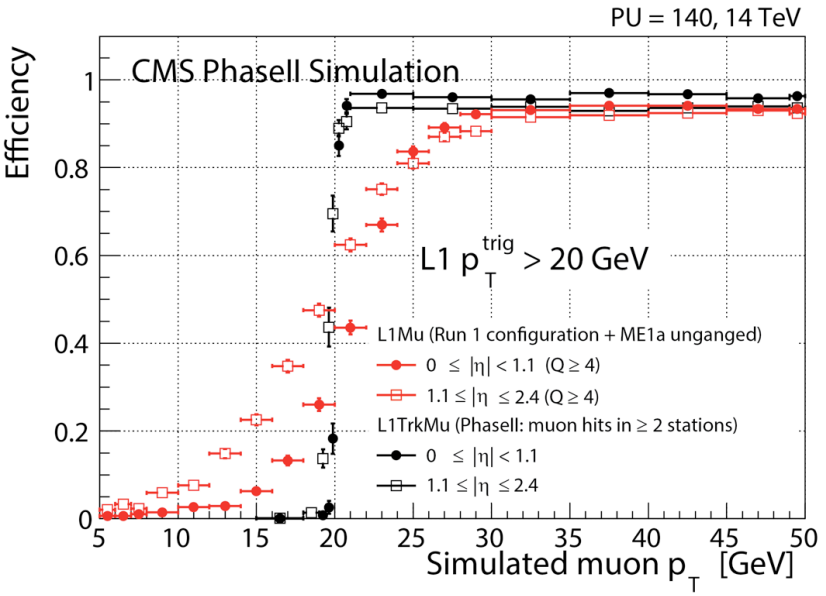
THE OHIO STATE UNIVERSITY

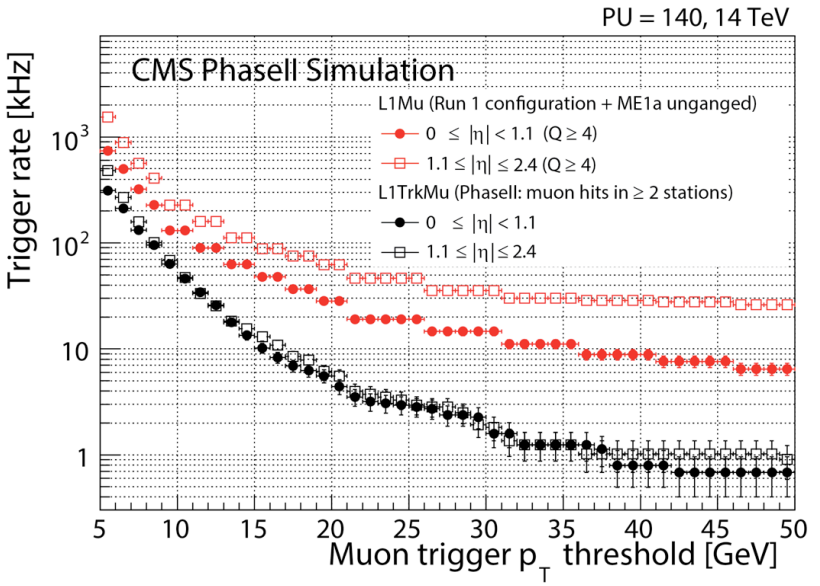
Backup

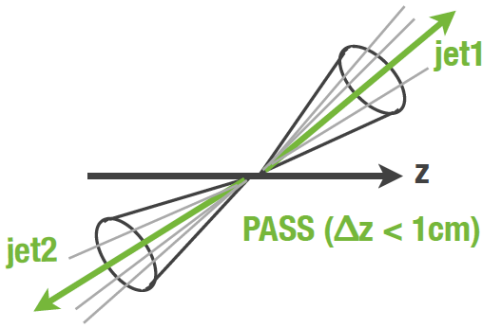
Impact of Tracking @ Trigger

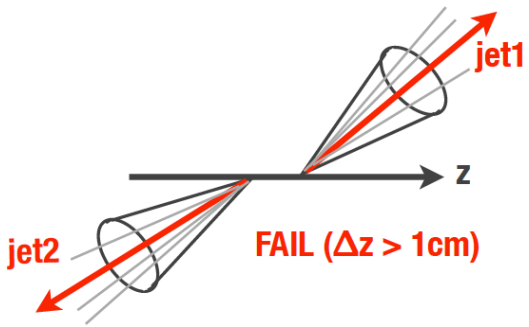
- Electron/Photons
  - ➡ Extra measurement - rate reduction
  - ➡ Isolation
- Muons
  - ➡ Excellent P_T resolution
  - ➡ Isolation
- Tau Triggers
  - ➡ Search for multi-prong
- Pileup mitigation
  - ➡ Separation of multiple interactions for multi-object triggers
  - ➡ Track-based missing energy











R&D Testing

- Validation of Algorithm/Firmware
  - ➡ Comparisons between simulations and hardware
    - ▸ Known MC input data
    - ▸ Compare parameters of found tracks
  - ➡ Comparisons between chip simulation and algorithm emulation (C++) allows debugging of intermediate stages.
  - ➡ Allows testing of communications between boards.
- Latency Measurements
  - ➡ Latency is an important constraint that must be met.
  - ➡ With test stand we can **measure** the latency and compare against our models.
    - ▸ Current ~3.2 µsec