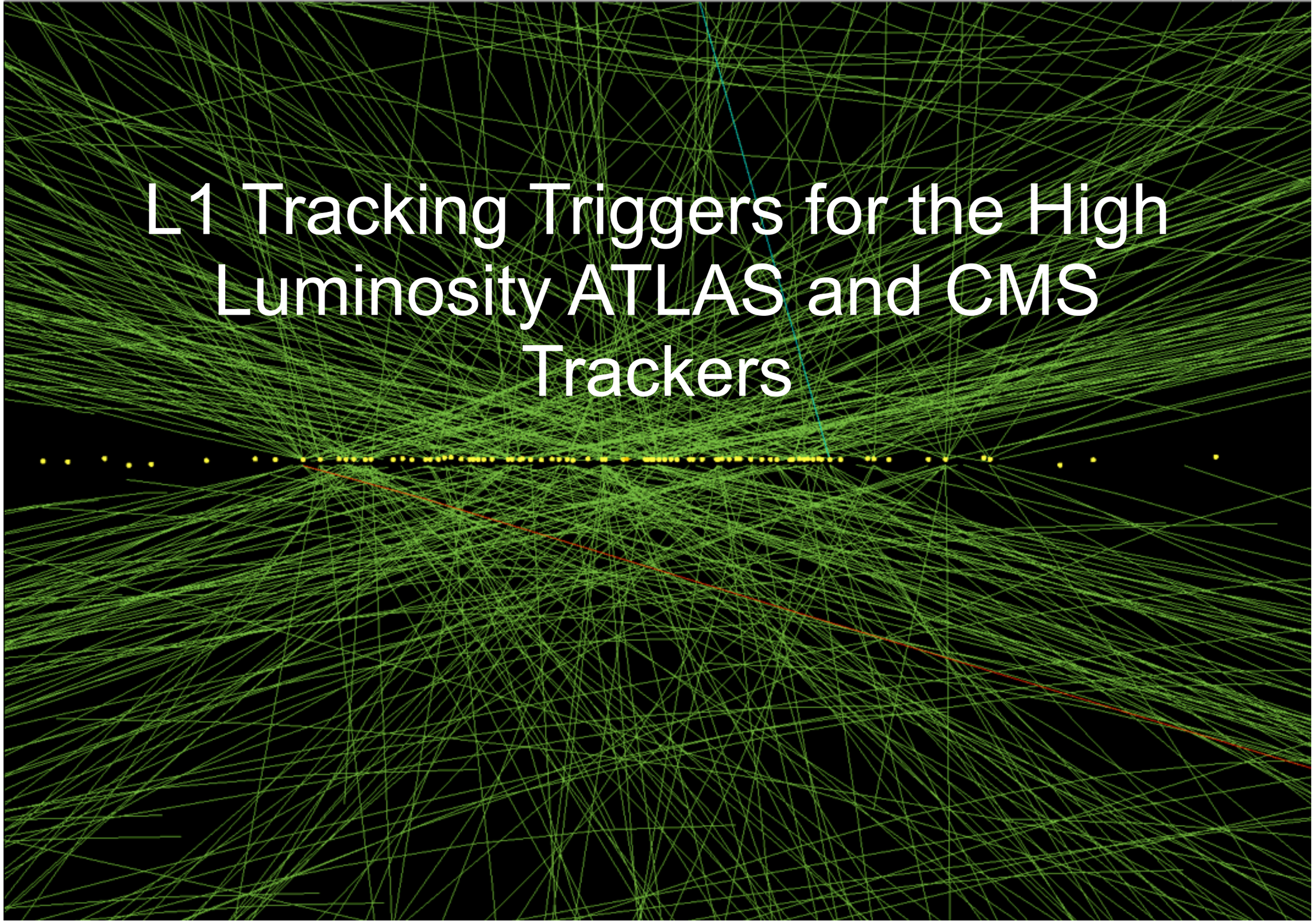
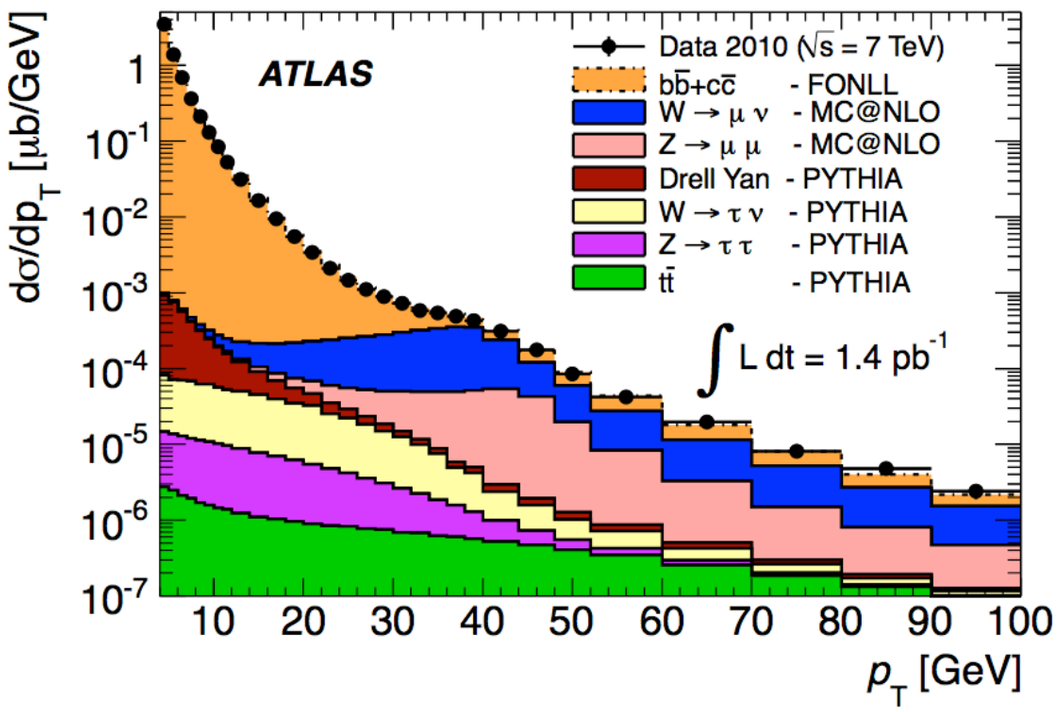
L1 Tracking Triggers for the High Luminosity ATLAS and CMS Trackers

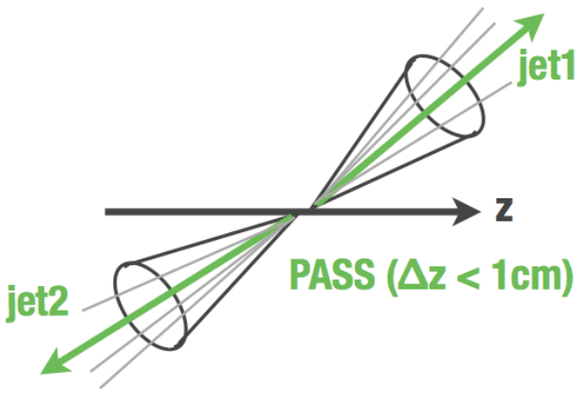


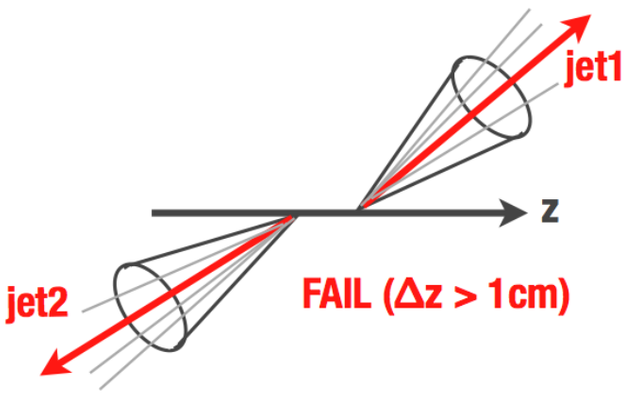
Fabrizio Palla
INFN Pisa
on behalf of ATLAS and CMS

Why a Track Trigger at L1

- HL-LHC physics goals require excellent Trigger selectivity on basic objects (leptons, jets, taus, b-jets, MET)
  - This might be jeopardized by the increased level of pileup events (140 on average)
    - Huge rate of μ from heavy flavors ⇨ use better p_T resolution from tracker
    - Prompt electrons at L1 need to be separated from huge γ ⇨ Tracker tracks
    - High E_T jets from (many) different primary vertices ⇨ jet-vertex association
    - Photon isolation in Calorimeters compromised by large pileup ⇨ use tracks

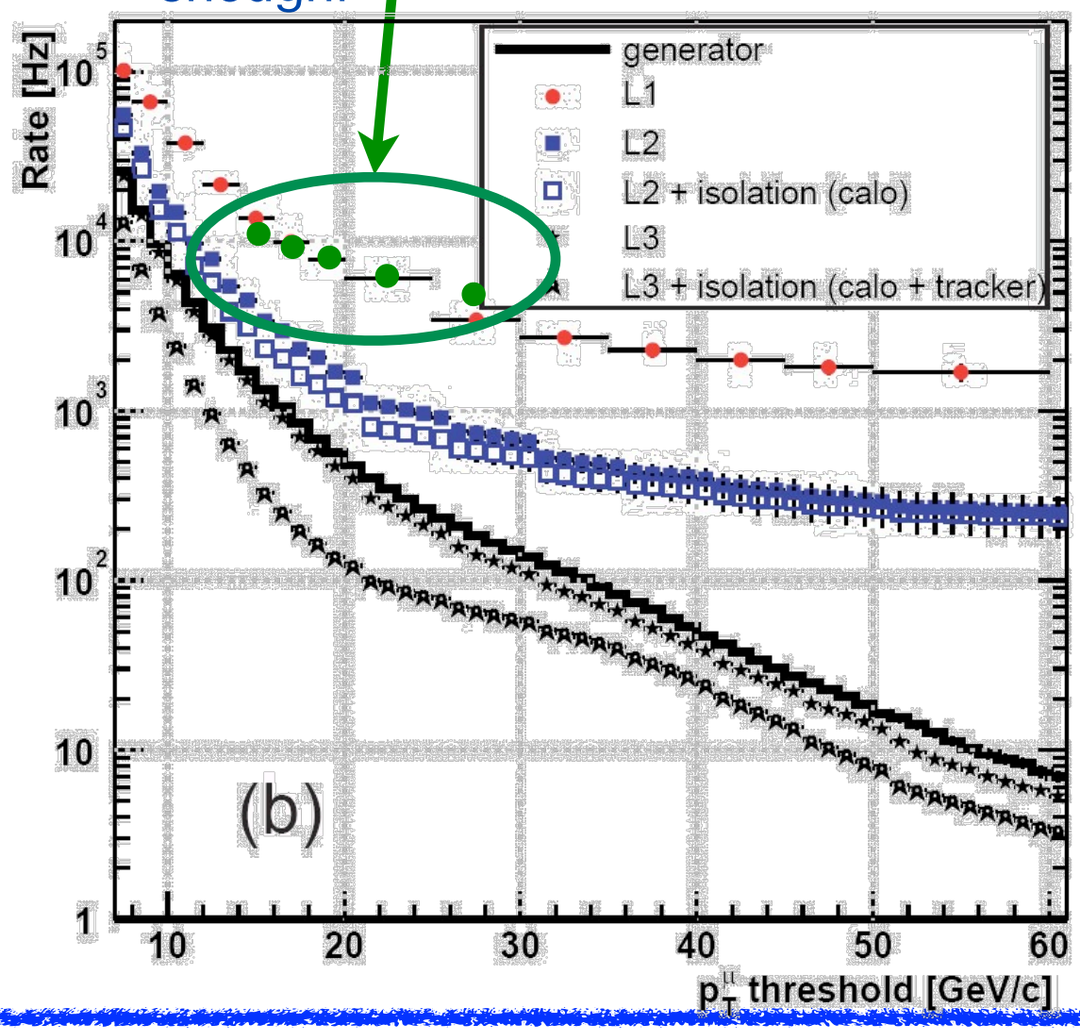




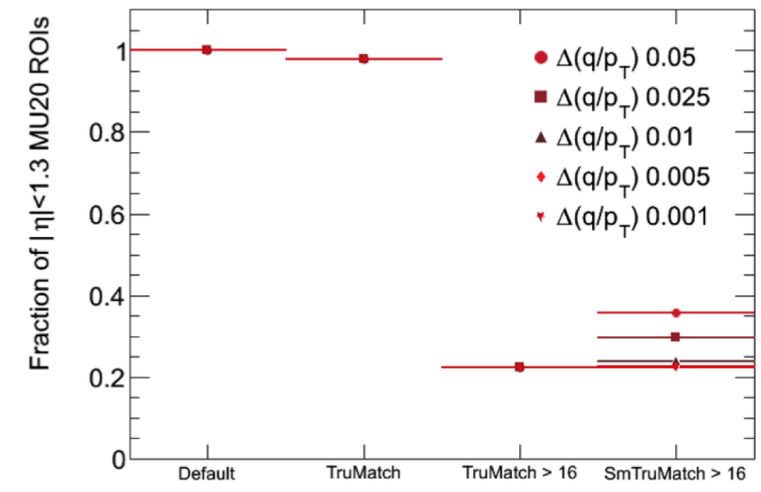


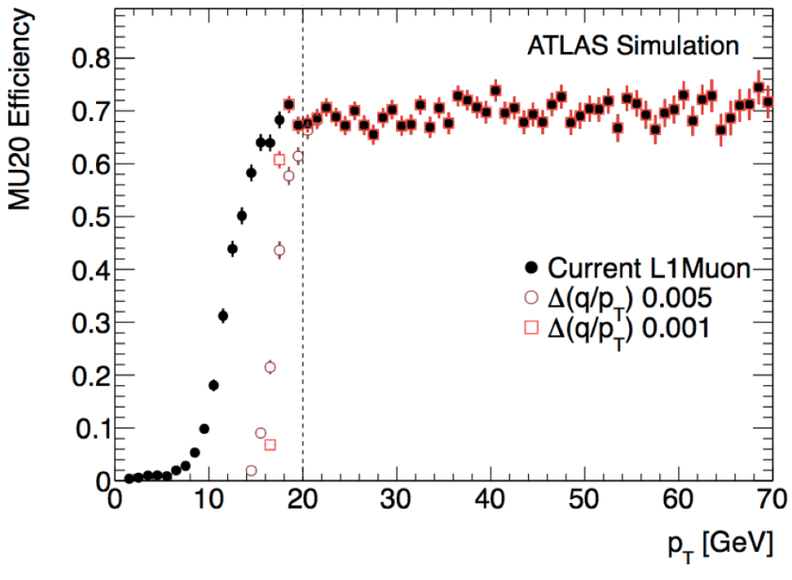
Examples from ATLAS and CMS

- CMS simulation for L=10^{34} cm^{-2} s^{-1}
- Add measured data rates at 8 TeV, extrapolated to 10^{34} cm^{-2} s^{-1}
  - No p_T threshold may reduce the rate enough!



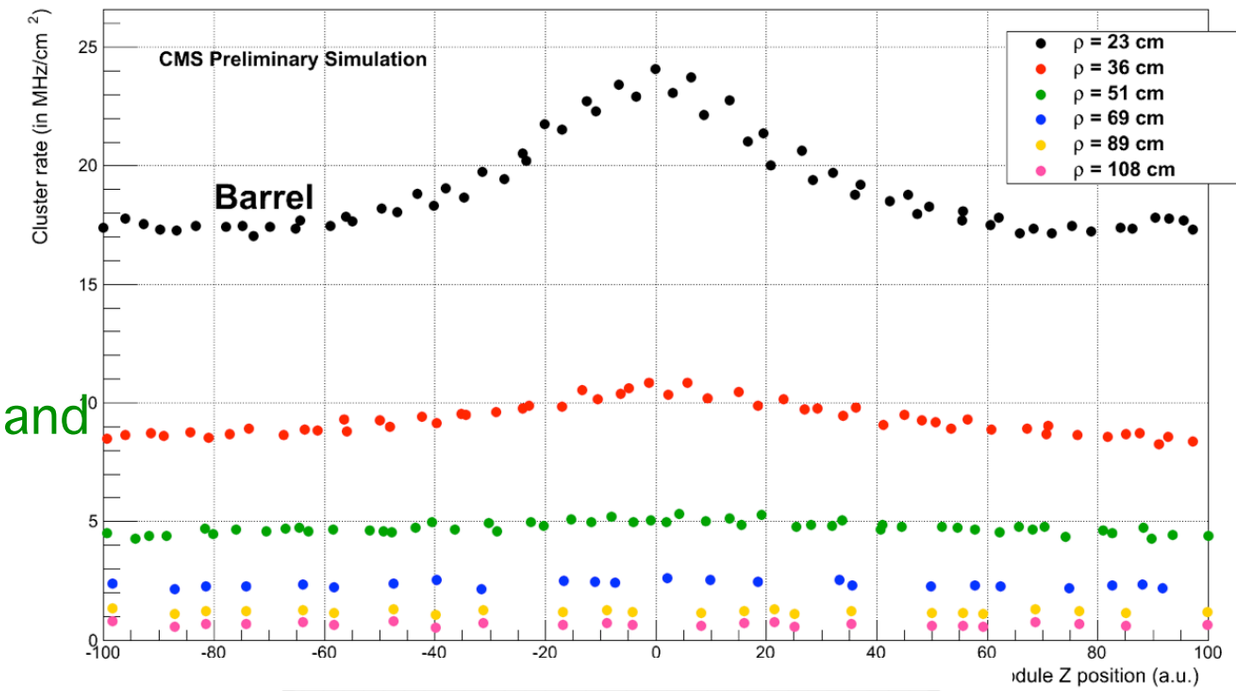
- **ATLAS simulation**
  - **~80% of μ originate from lower p_T**
  - **Sharpening the p_T to reduce the rate at constant efficiency**

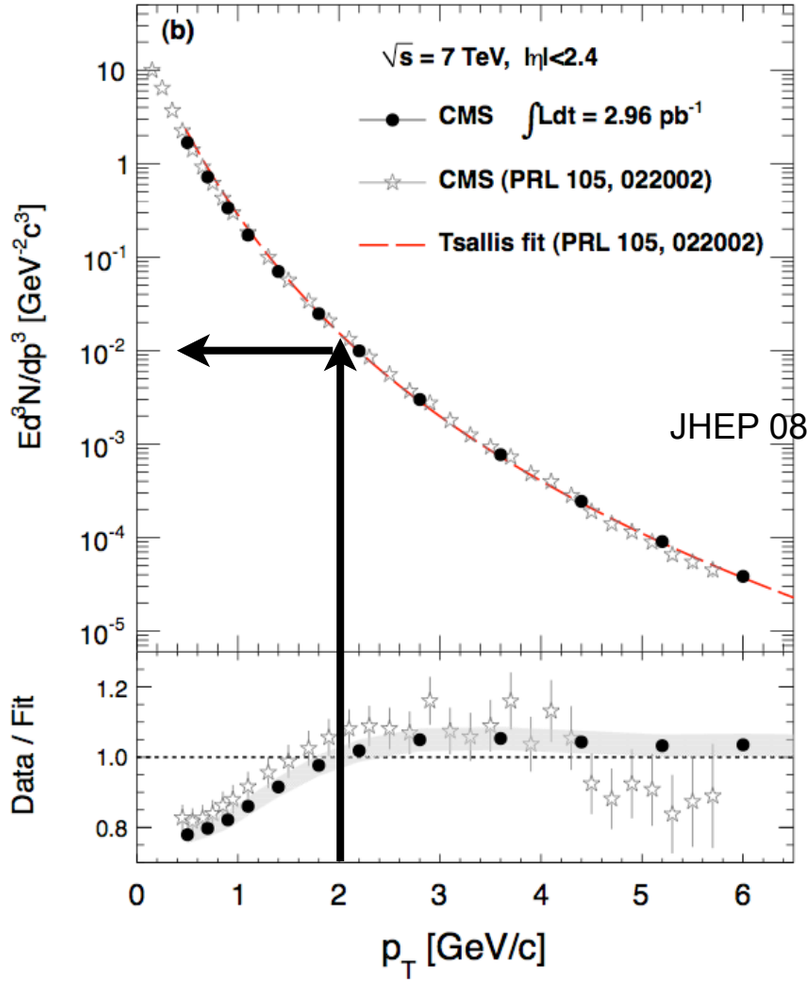




The challenge and the way out

- Take data off the tracker and combine with other primitives
  - ~ 4k primary tracks within |η|<2.5
  - Large data rates (up to 25 MHz/cm^2)
    - huge contribution from nuclear interactions and photon conversions
  - ~1.3 events/mm × Gauss(σ=4 cm)
  - Short L1A trigger latencies (10-20 μs)
  - Cannot read all (~60 M strips) channels at 40 MHz
    - Even a 1% occupancy: 0.5 M channels x 40 MHz x 20 bit = 400 Tb/s
      - ~120k links at 3.25 Gb/s (GBT) - Current CMS Tracker has 40k links (320 Mb/s)
    - Need to
      - suppress hits from low p_T tracks
      - read at smaller (affordable) rate

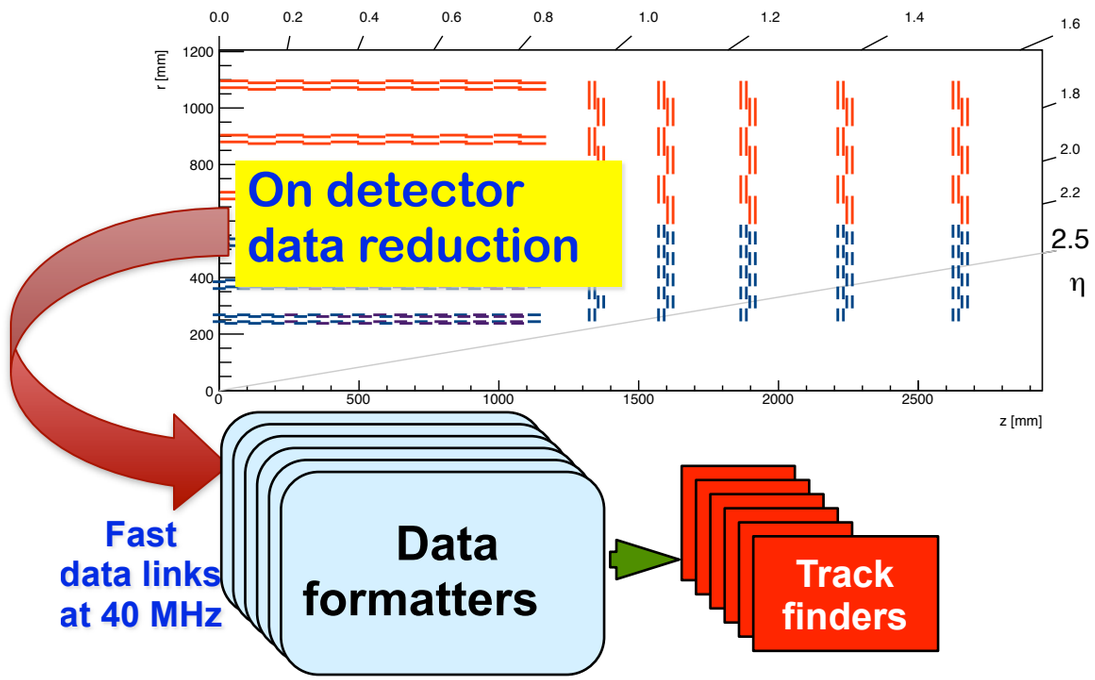


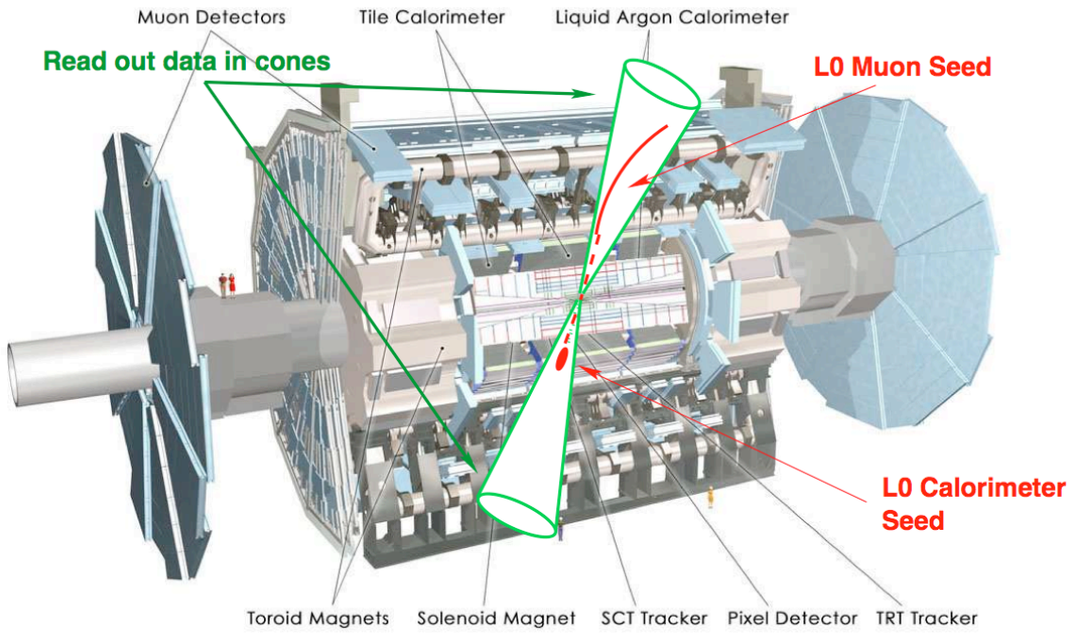


JHEP 08 (2011) 086

Trigger architectures

- PUSH path (CMS)
  - Reduced Tracker information readout at 40 MHz and then combined with calorimeter & muon at L1
    - Trigger objects made from tracking, calorimeter & muon inside a Global Trigger module
- PULL path (ATLAS)
  - Use calorimeter & muon detectors to produce a “Level-0” to request tracking information in specific regions
    - Tracker sends out information from regions of interest to form a new combined L1 trigger

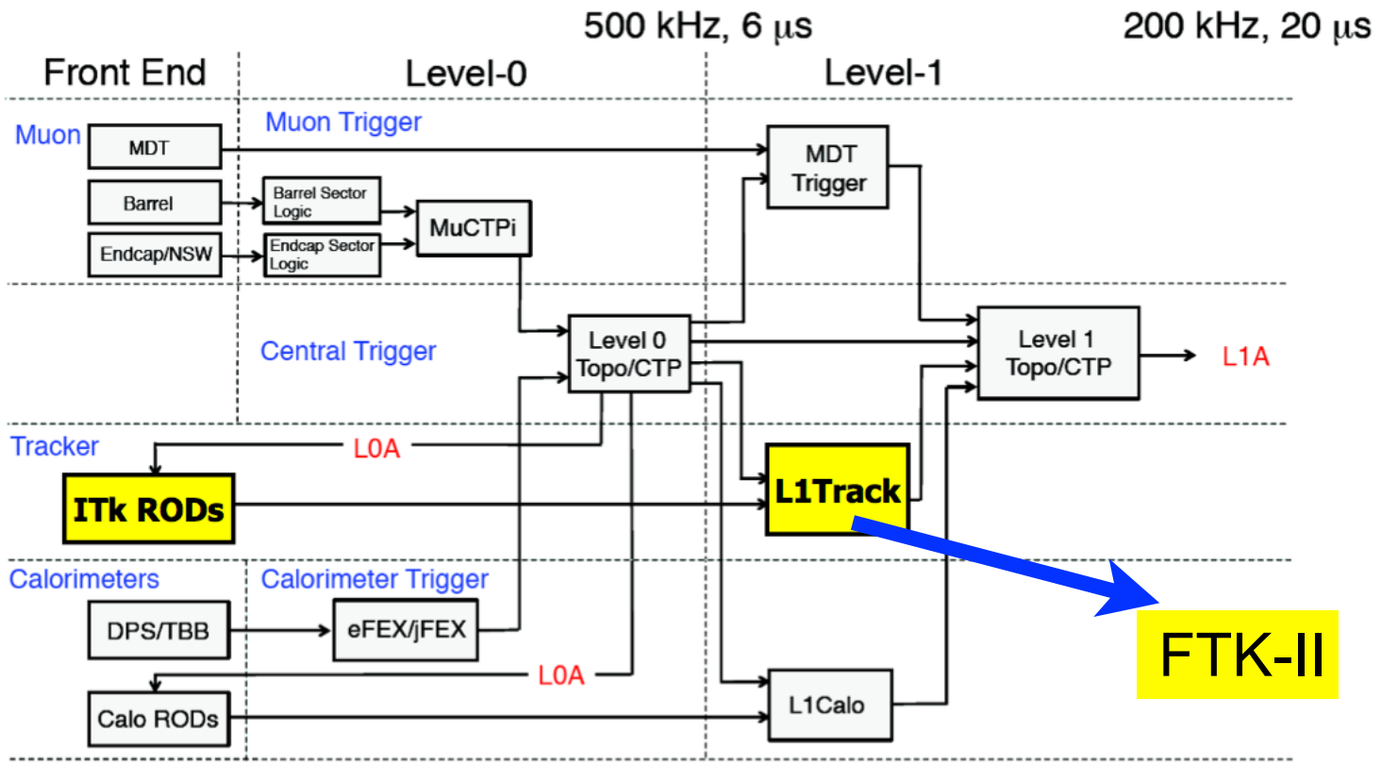




Track Trigger with pull architecture (ATLAS)

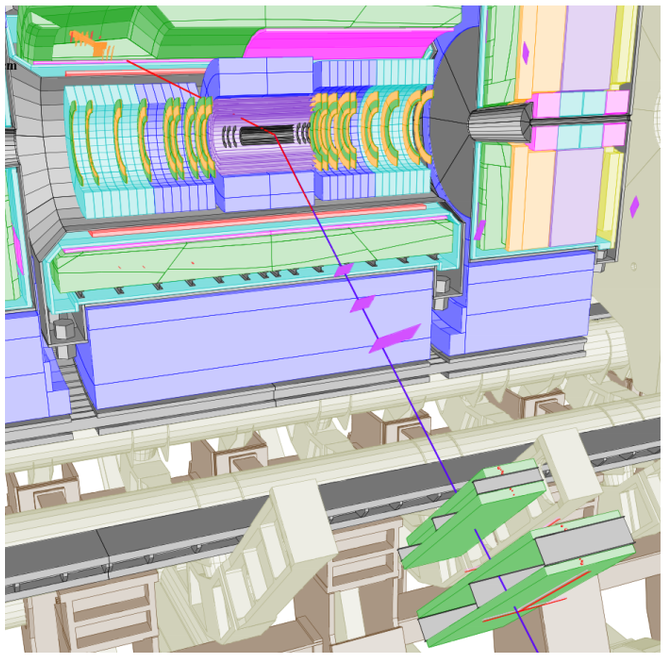
The L0+L1 scheme

- Level-0:
  - Coarse calo and muon data
  - Rate 40 MHz →500 kHz
  - Latency < 6.4 *μs*
  - Defines Region of Interest (ROIs) for L1
- Level-1:
  - Tracker data only from ROIs
  - Refined information from calo and muons
  - Rate 500 kHz →200 kHz
  - Latency < 20 *μs*



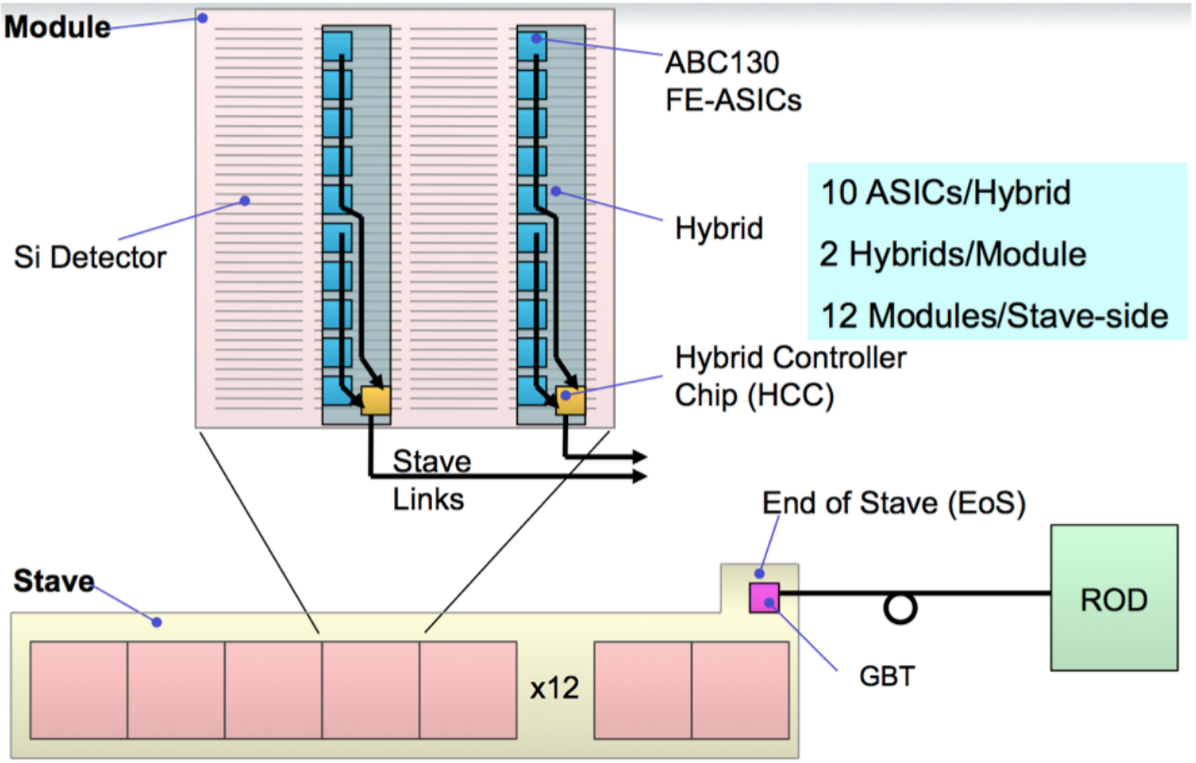
Issues for FTK to be used in Phase 2

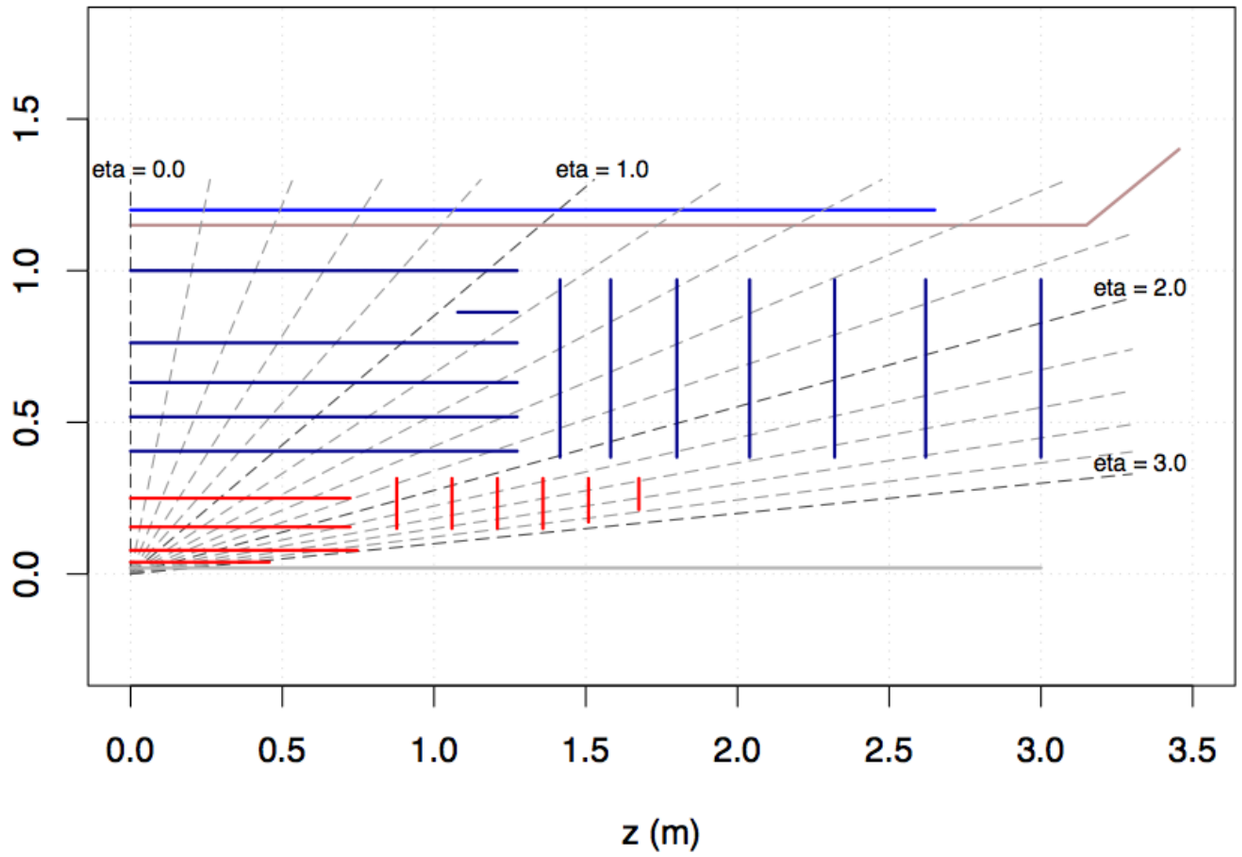
- the larger pileup (x2.5), rate (x5) and granularity
  - increase in the number of patterns by ~one order of magnitude
    - no p_T filtering - rise p_T threshold
- need to cope with shorter latency (20 *μs* instead of 200 *μs*)

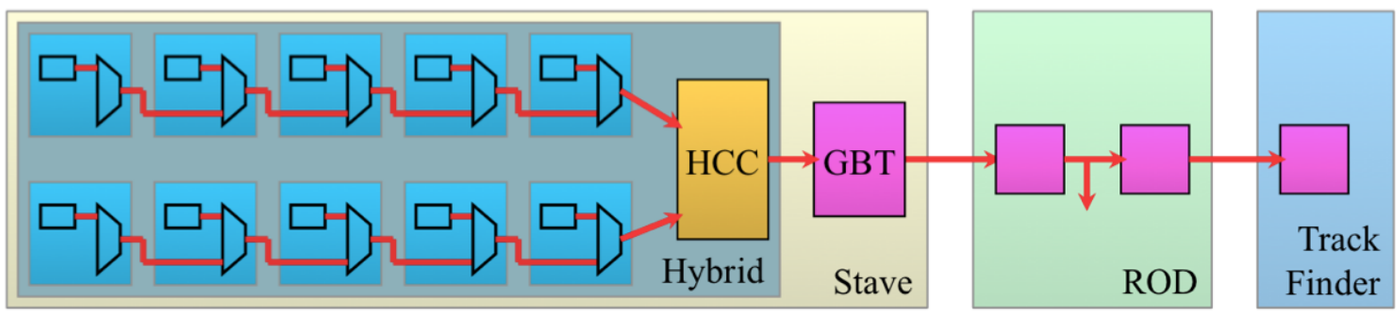


ATLAS readout

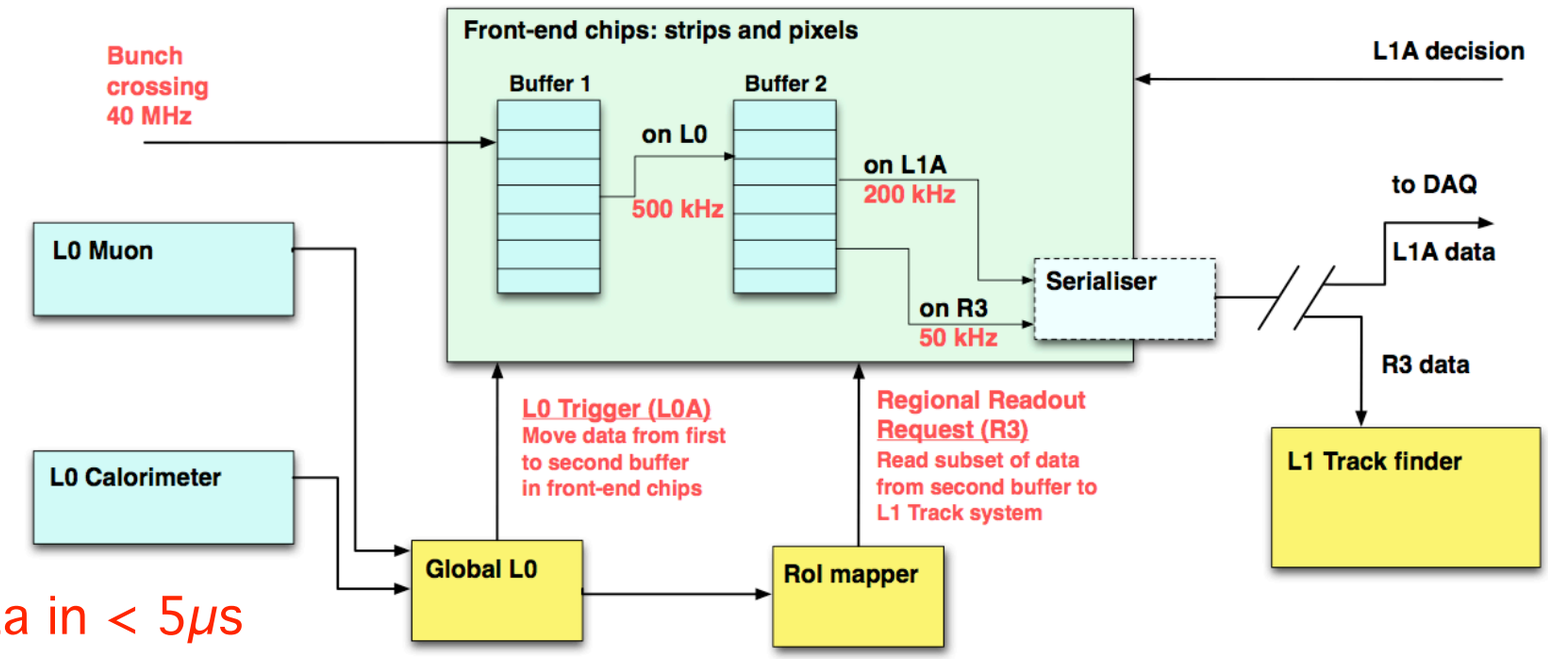
ATLAS Tracker for HL-LHC



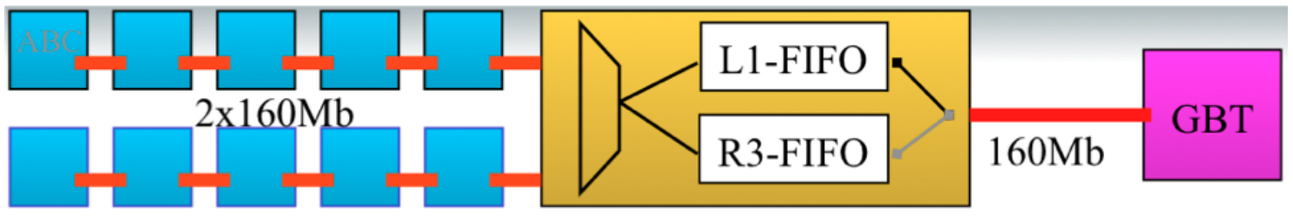


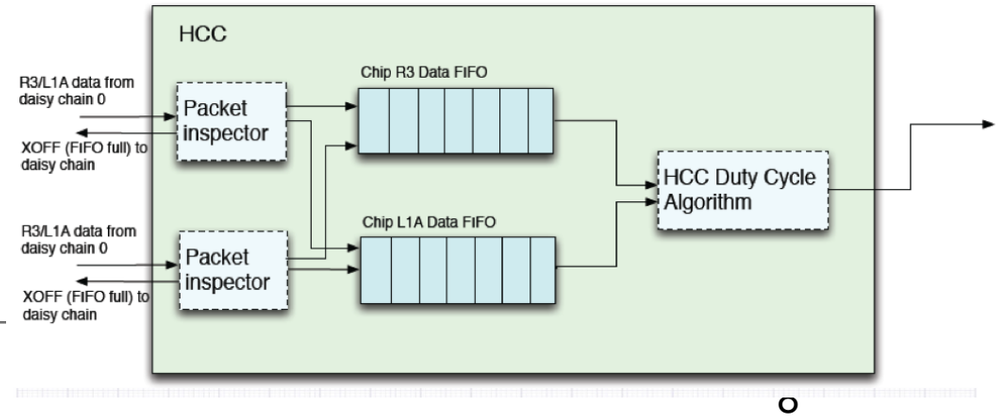


ATLAS L0 and Regional Readout Requests (R3) implementations

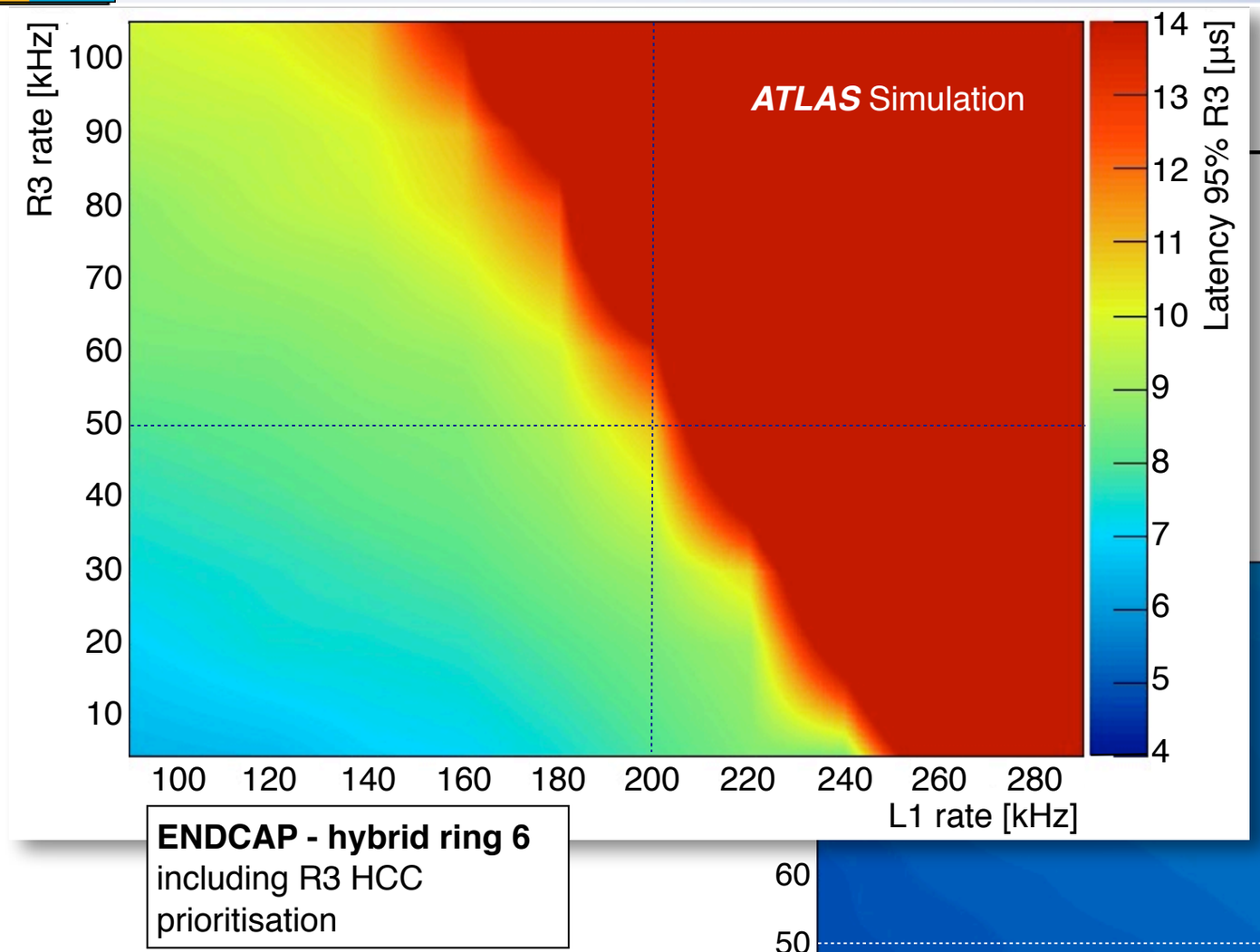


- Trigger data in < 5μs
  - L0 Trigger accept rate 500 kHz
    - On a L0 accept, copy data from primary to secondary buffer
    - Identify “region of interest” (1-10% of the detector on each L0 accept)
    - Generate a “Regional Regional Request” (R3)
      - Reading only ~10% of the Tracker data, the total bandwidth is only 50% more with the Track Trigger than without.
    - To reduce the latency, a prioritization scheme is envisaged, by using a dedicated R3 buffer



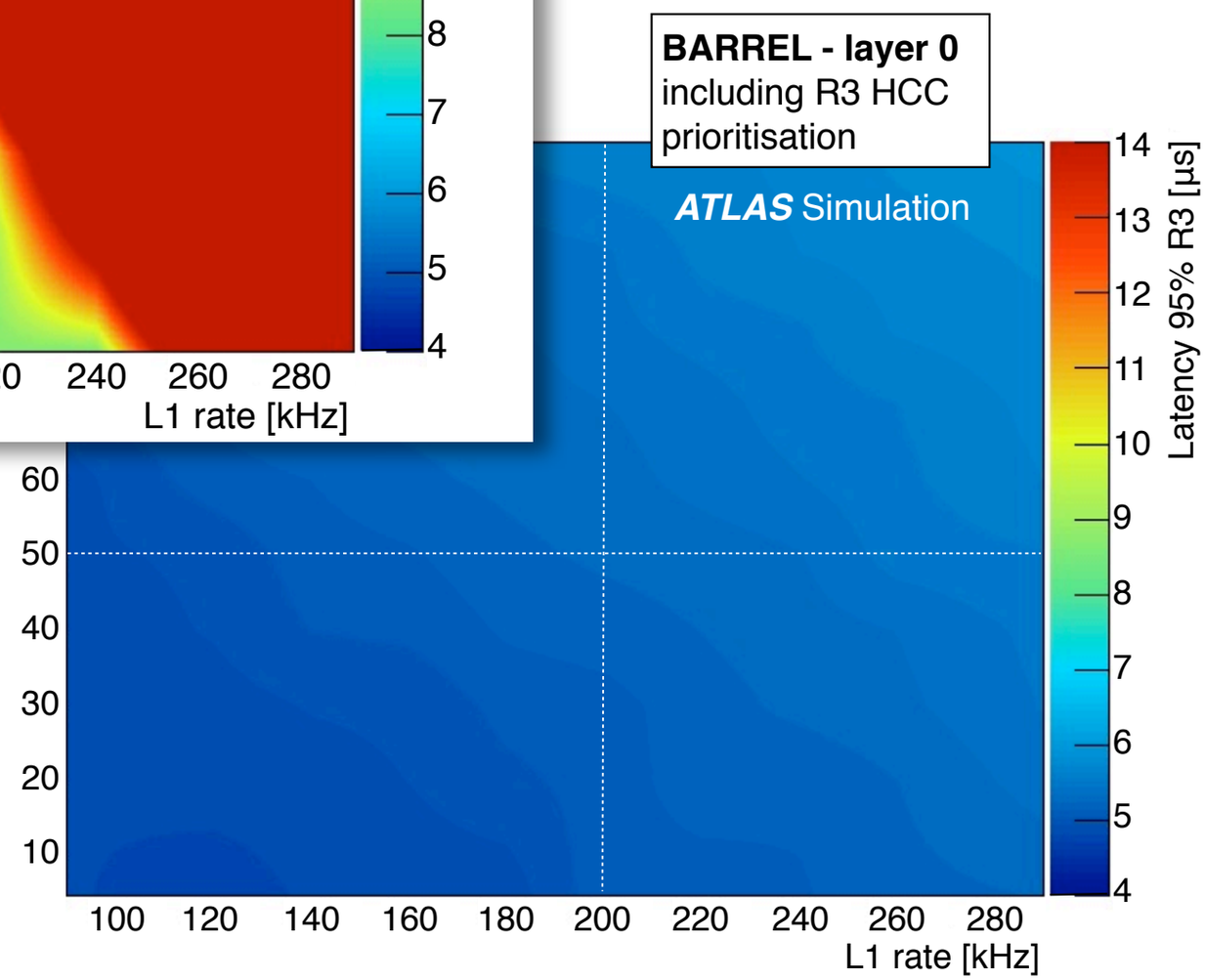


Simulation results - ATLAS



ENDCAP - hybrid ring 6
including R3 HCC prioritisation

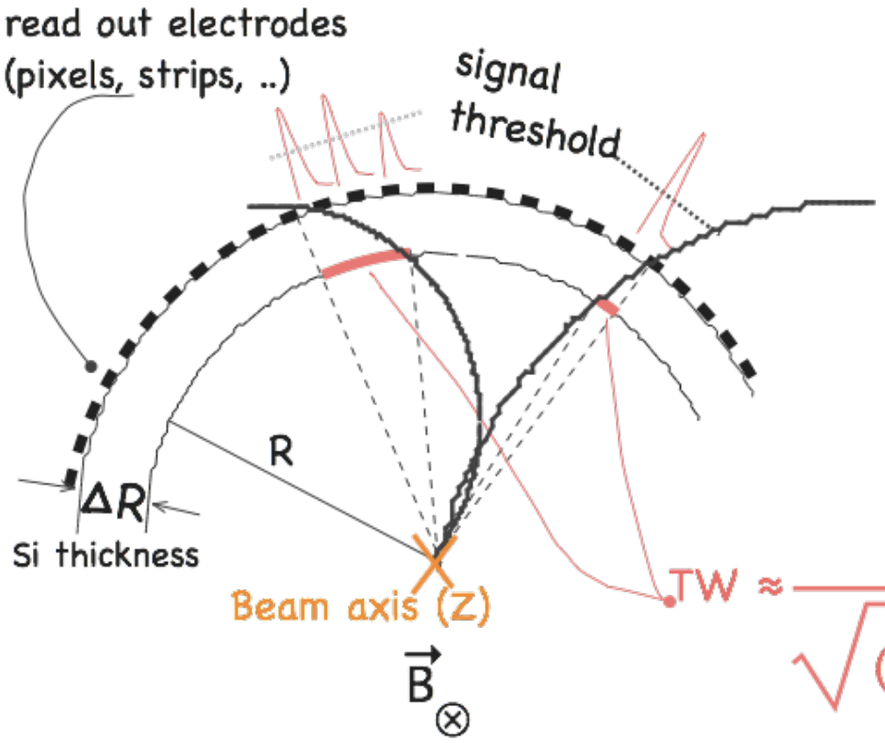
Latency maps

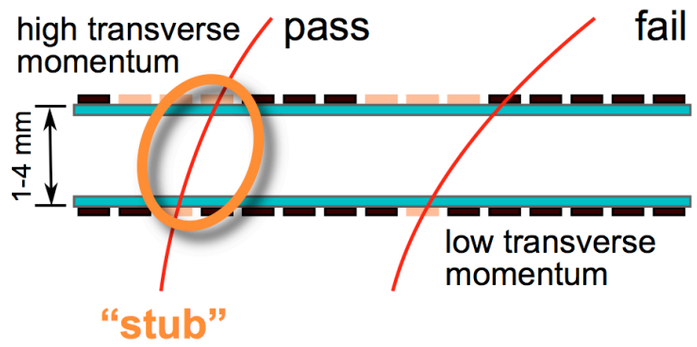


- Barrel acceptable latency over the full phase space when using prioritisation on the HCC
- Endcap hybrids run 4, 5 and 6 have an issue due to limited HCC bandwidth and chip multiplicity

Select only hits from "high-p_T" tracks

- Select "high-p_T" tracks (>2 GeV) by correlating hits in 2 nearby sensors (stub)





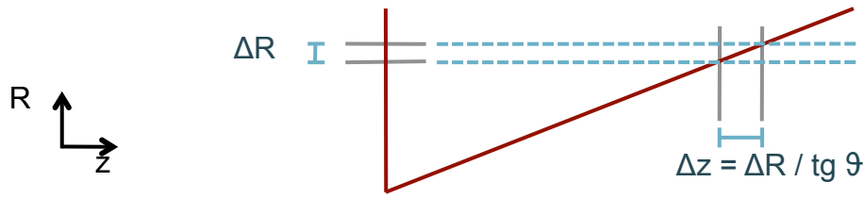
F. Palla, G. Parrini, PoS VERTEX2007 (2007) 034, http://pos.sissa.it/archive/conferences/057/034/Vertex%202007_034.pdf

J. Jones, A. Rose, C. Foudas, G. Hall, http://arxiv.org/pdf/physics/0510228v1.pdf

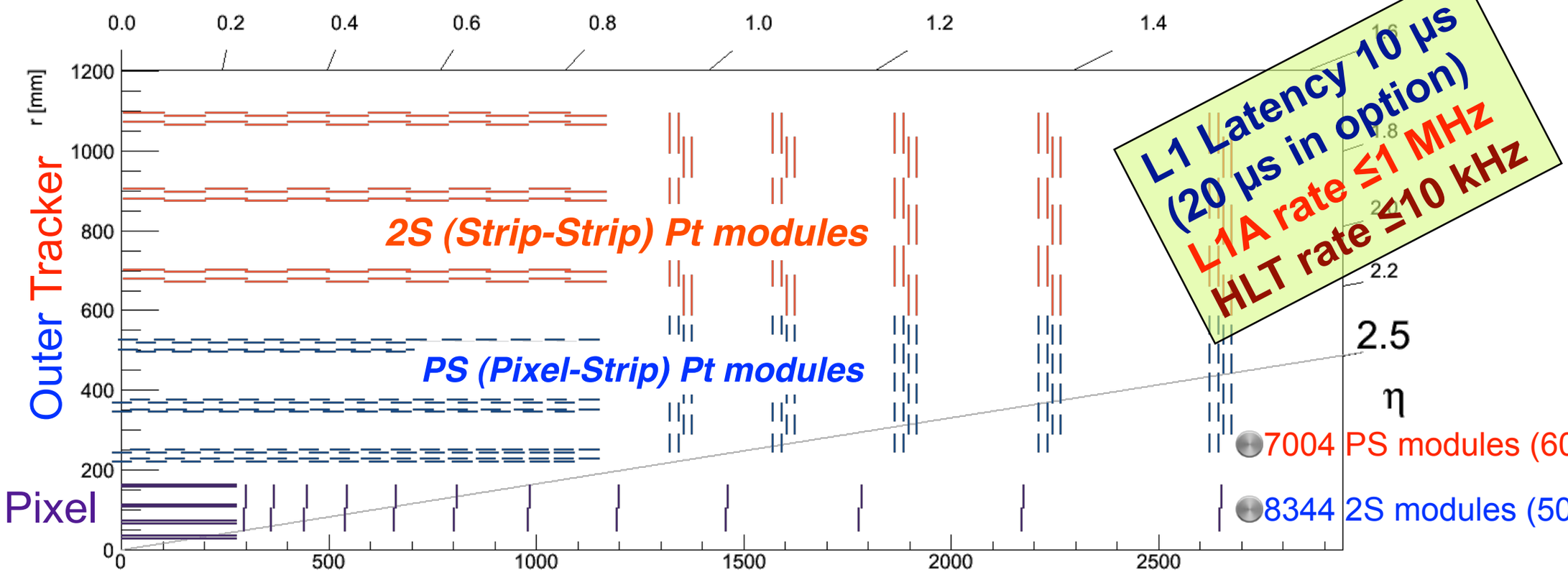
$$TW \approx \frac{\Delta R}{\sqrt{\left(\frac{pT}{pT_{min}}\right)^2 - 1}} \approx \Delta R \frac{pT_{min}}{pT} = 0.15\,B\,\Delta R \frac{R}{pT}$$

Large B field of CMS beneficial!

- In the barrel, ΔR is given directly by the sensors spacing
- In the end-cap, it depends on the location of the detector
  - ➡ End-cap configuration typically requires wider spacing (up to ~ 4 mm)

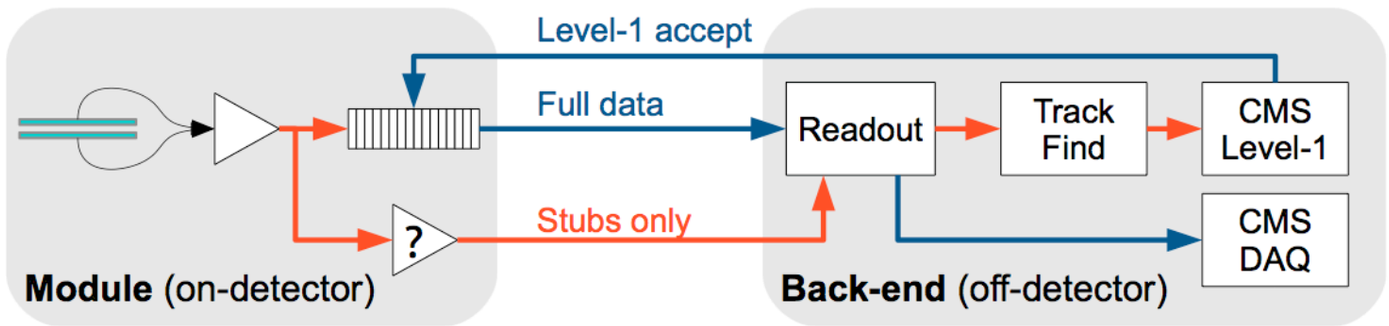


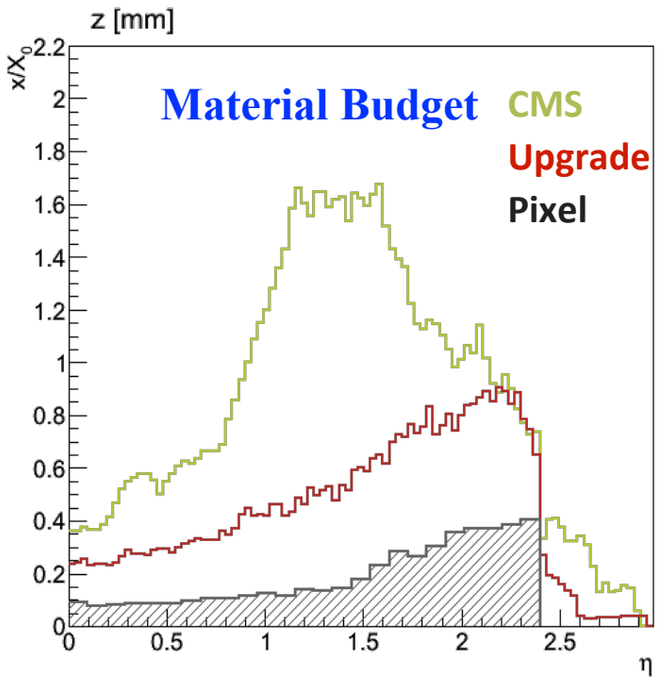
CMS Upgraded Tracker



Better p_T resolution and lighter than current tracker

Readout and Trigger schematics

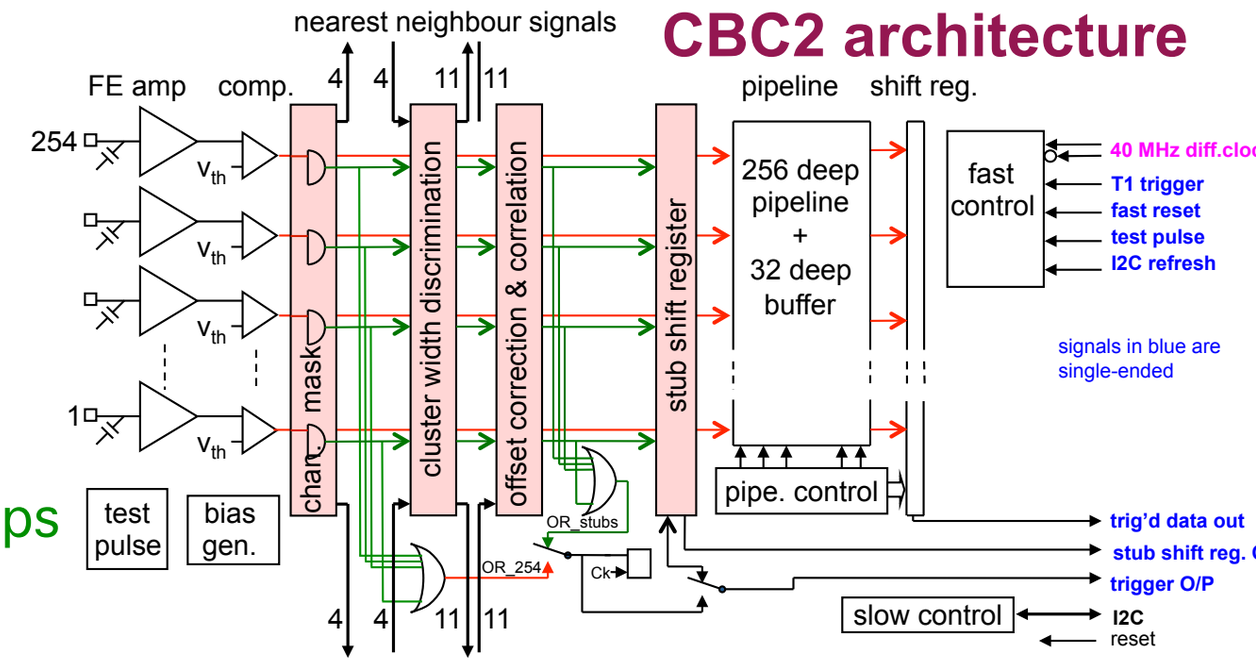


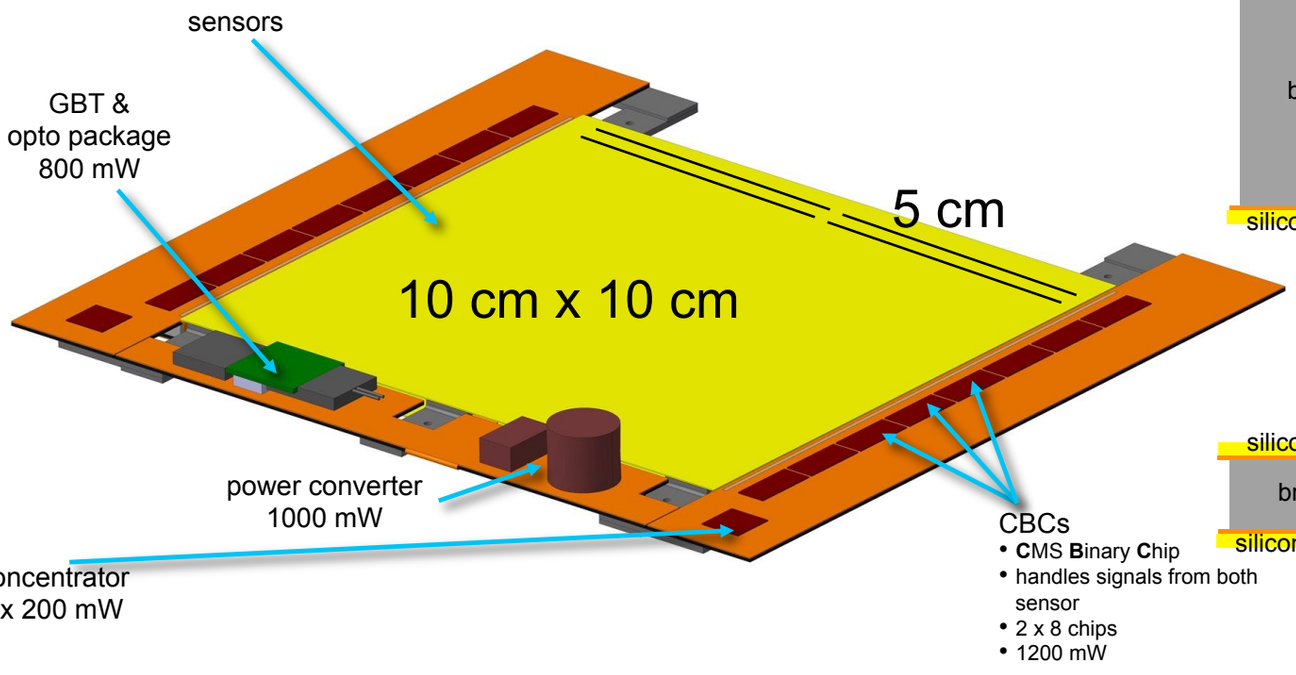


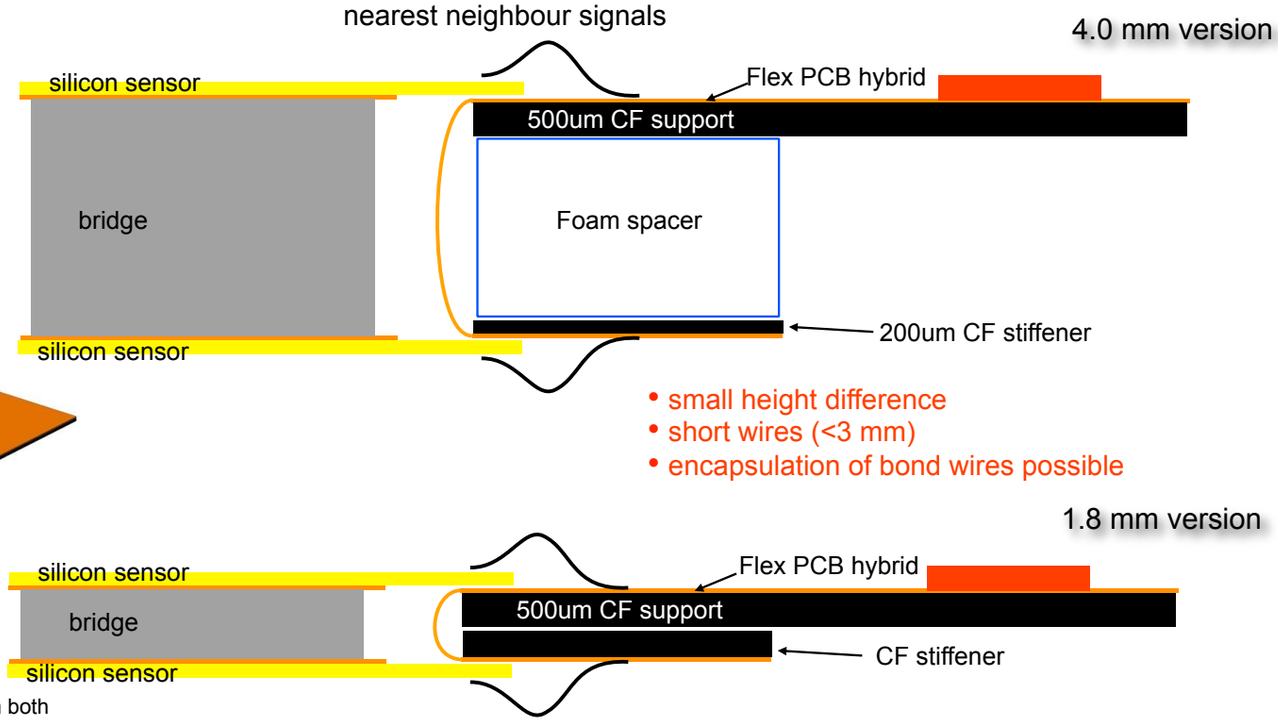
CMS 2S modules

2S(trip) sensors modules

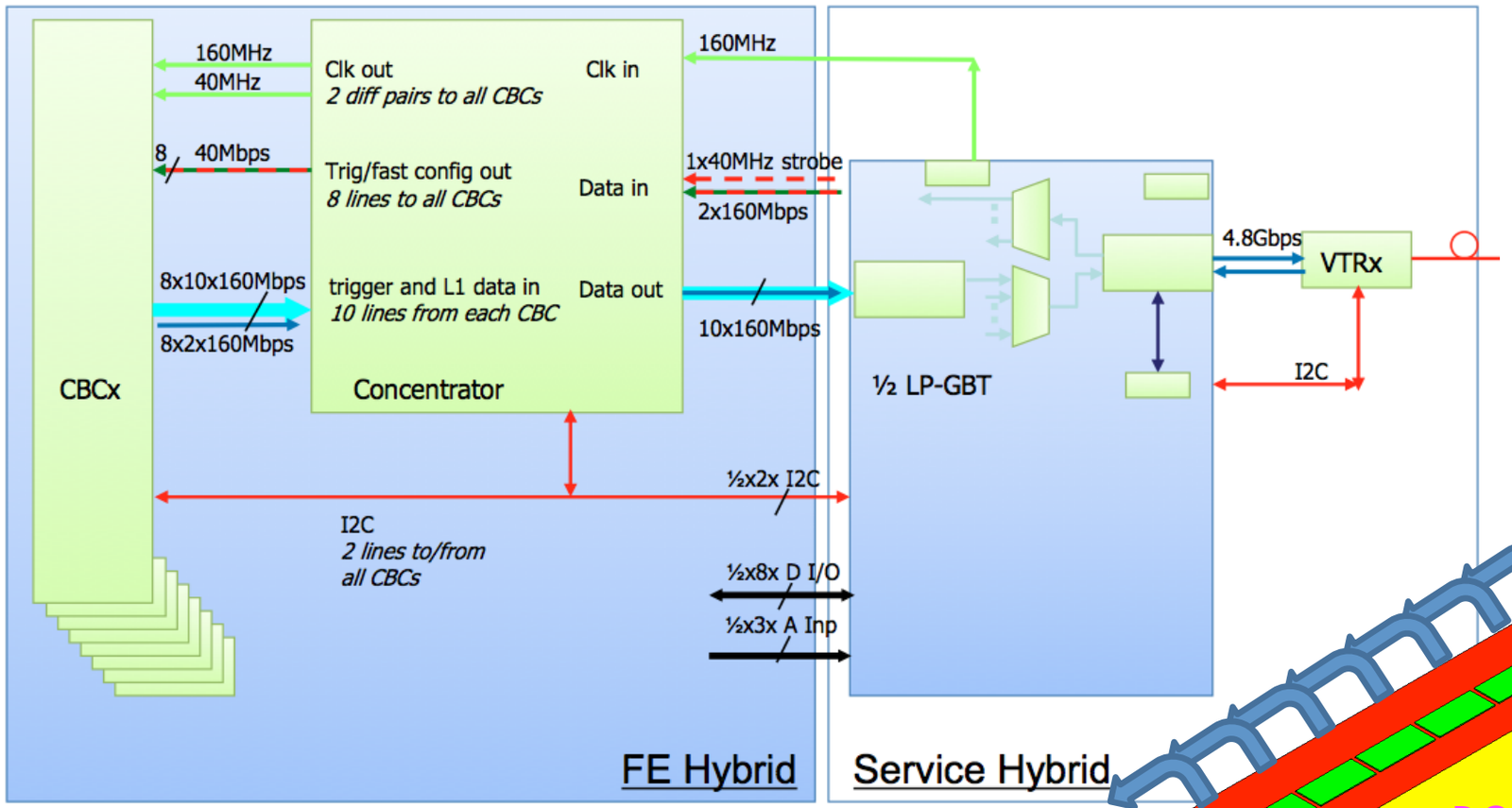
- 100 μm x 5 cm long strips on both sensors
- readout by 8 CBC on either sides
  - First discriminates signals by rejecting large clusters; then form a coincidence between the two sensor planes
  - Concentrator chip sends data from 8 chips to GBT

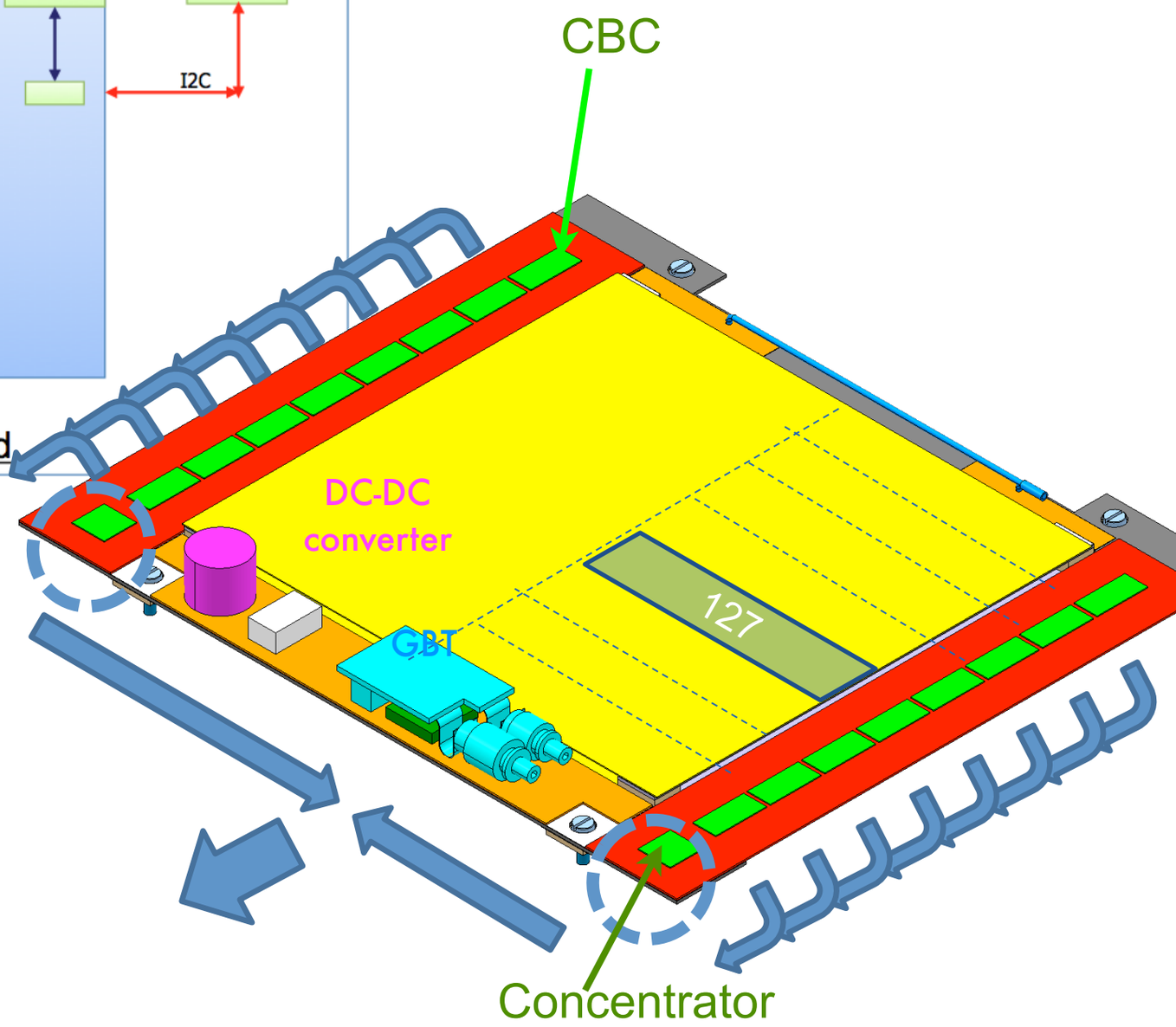






Electronics



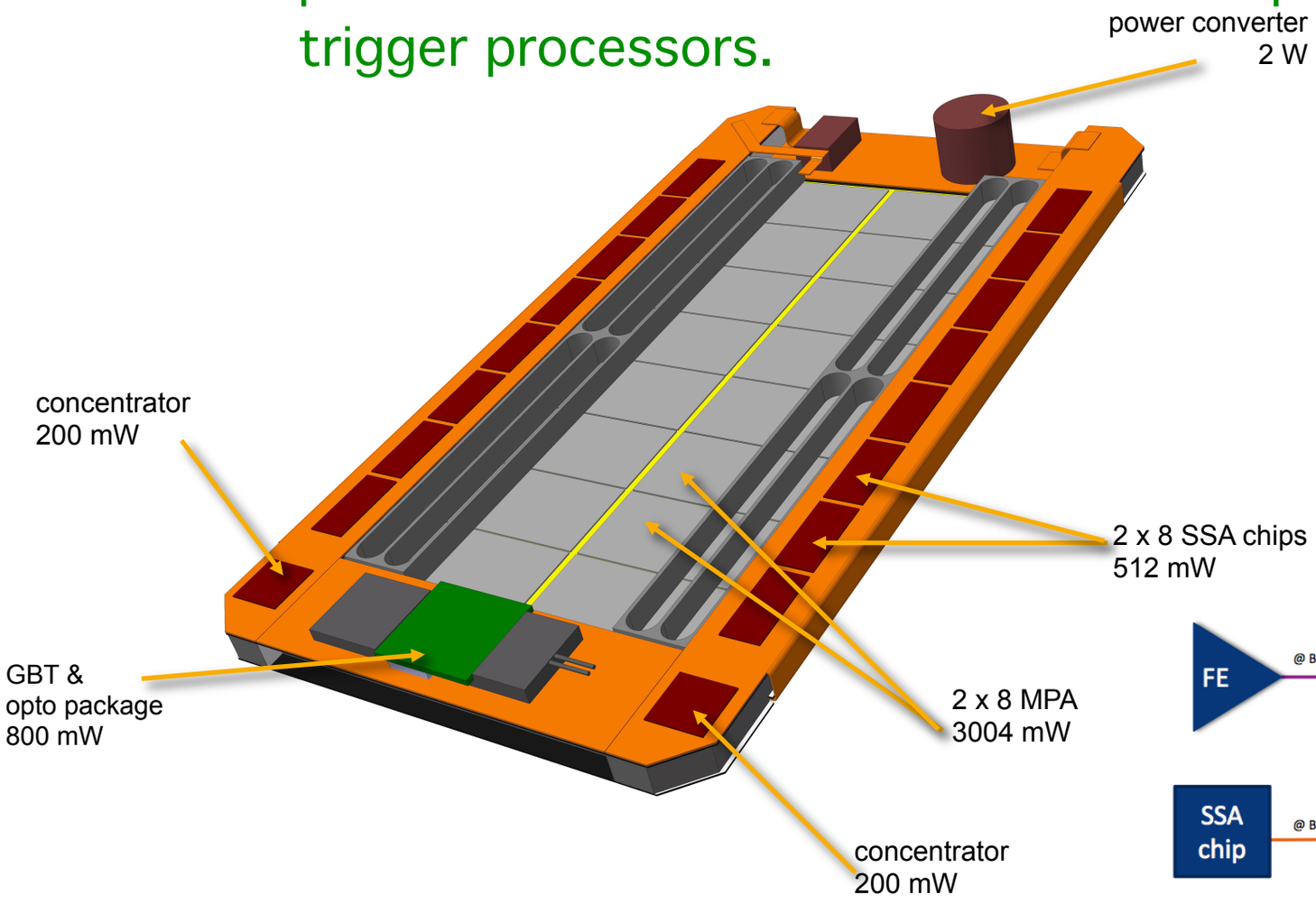


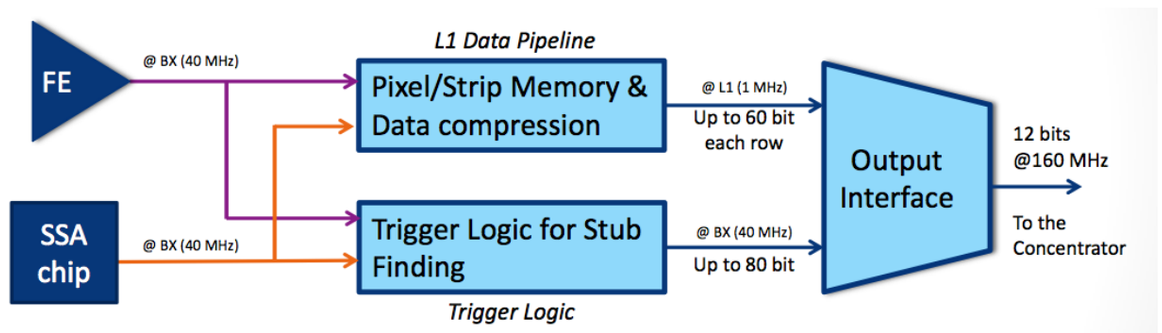
CBC outputs up to 3 stubs/bx, 160 MHz x 10 bits

Concentrator chip receives 8 CBC and sends out up to 12 stubs/8bx, 160 MHz

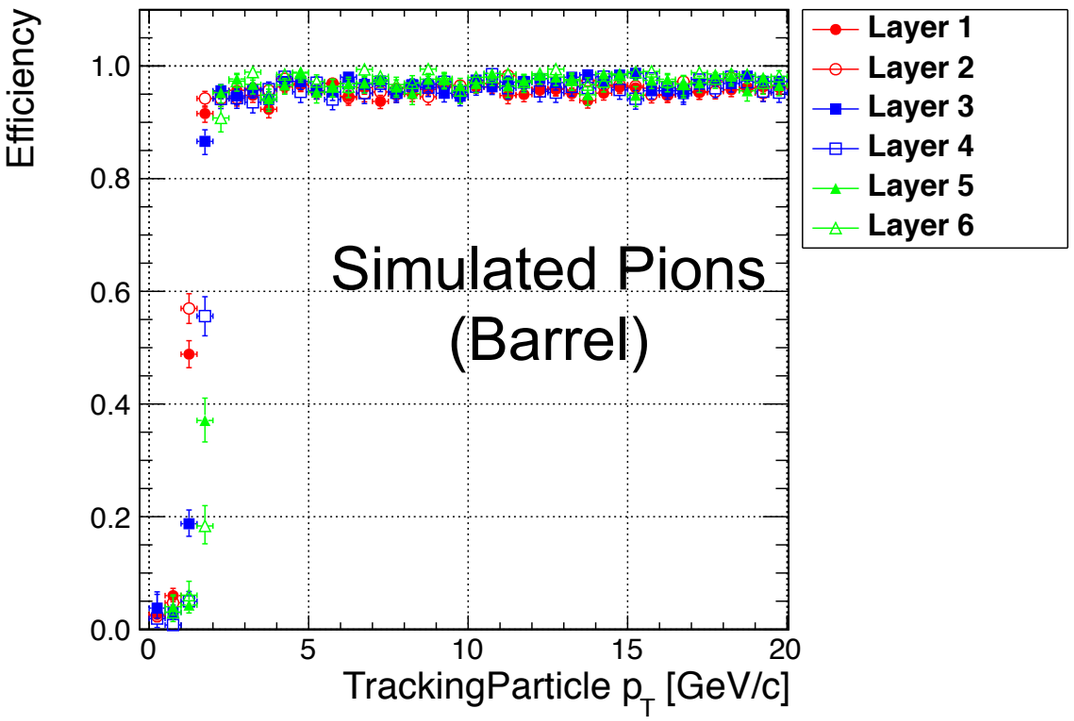
CMS PS modules

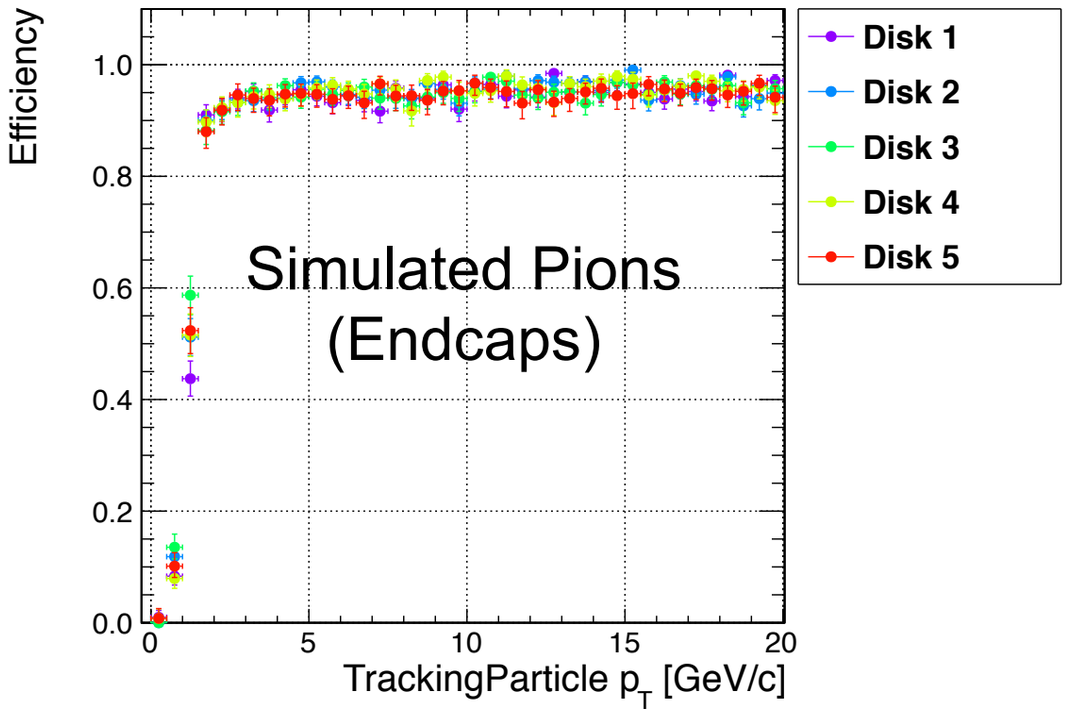
- P(ixel)S(trip) module
  - strips = 100 μm x 2.4 cm
  - pixels = 100 μm x 1.5 mm
    - Pixels are logically OR-ed for finding coincidence in the r-φ plane, and the precise z-coordinate is retained in the pixel storage and provided to the trigger processors.



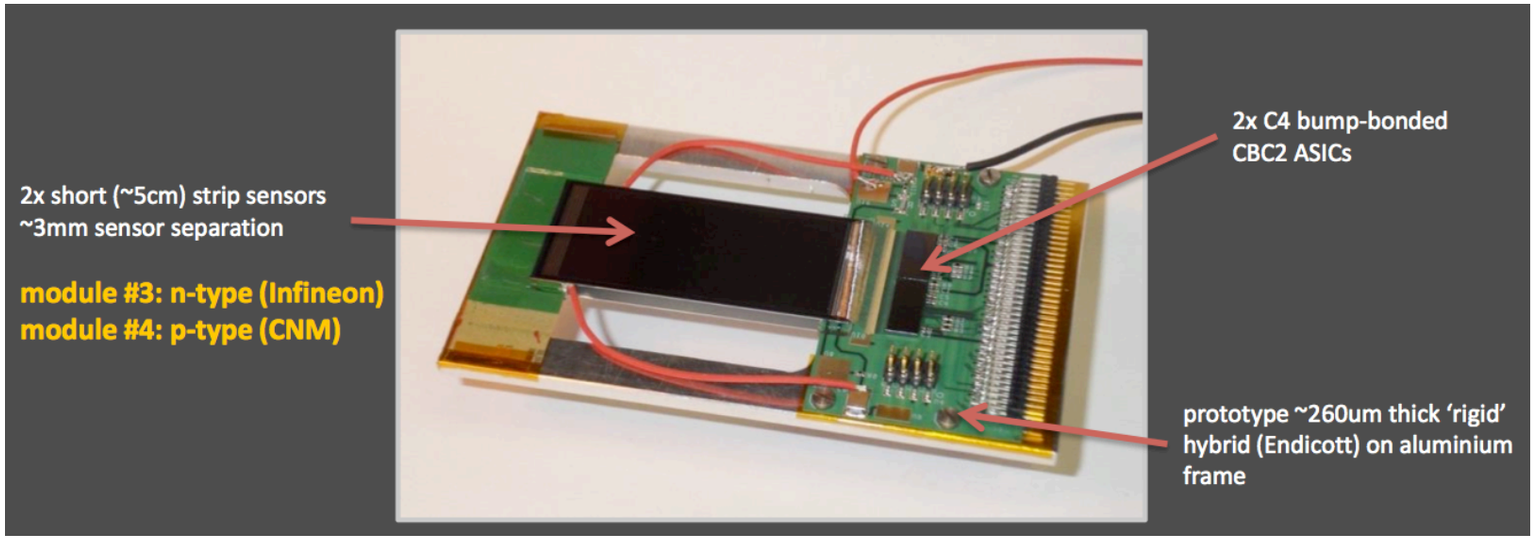


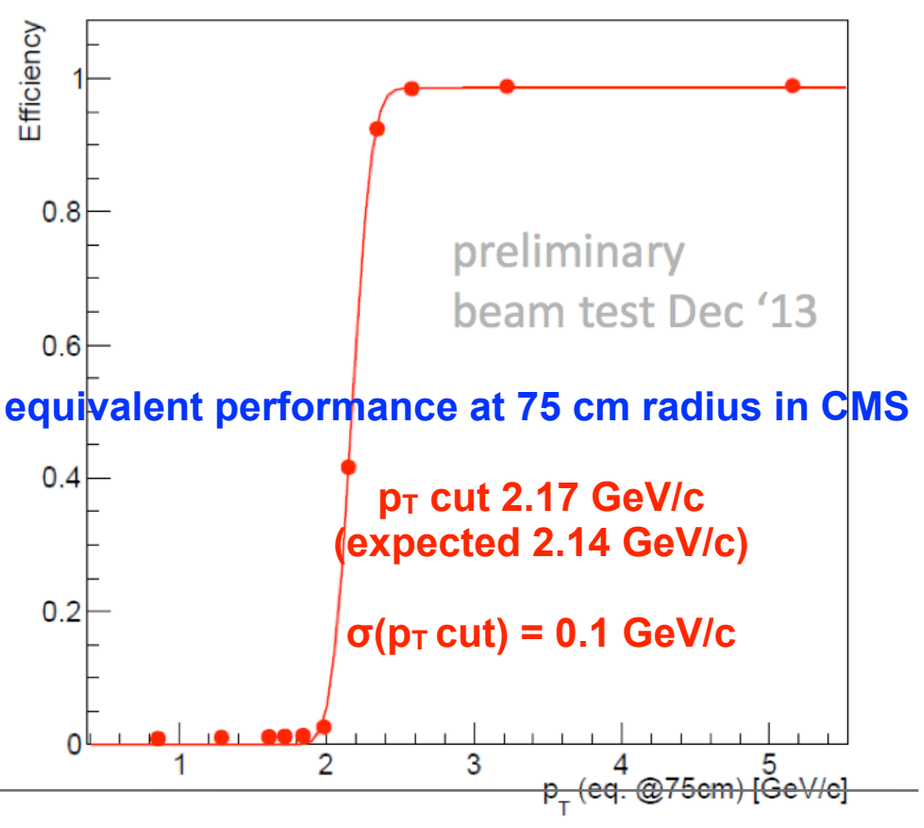
p_T modules performances





Prototype module test beam at DESY (2-4 GeV e^+)

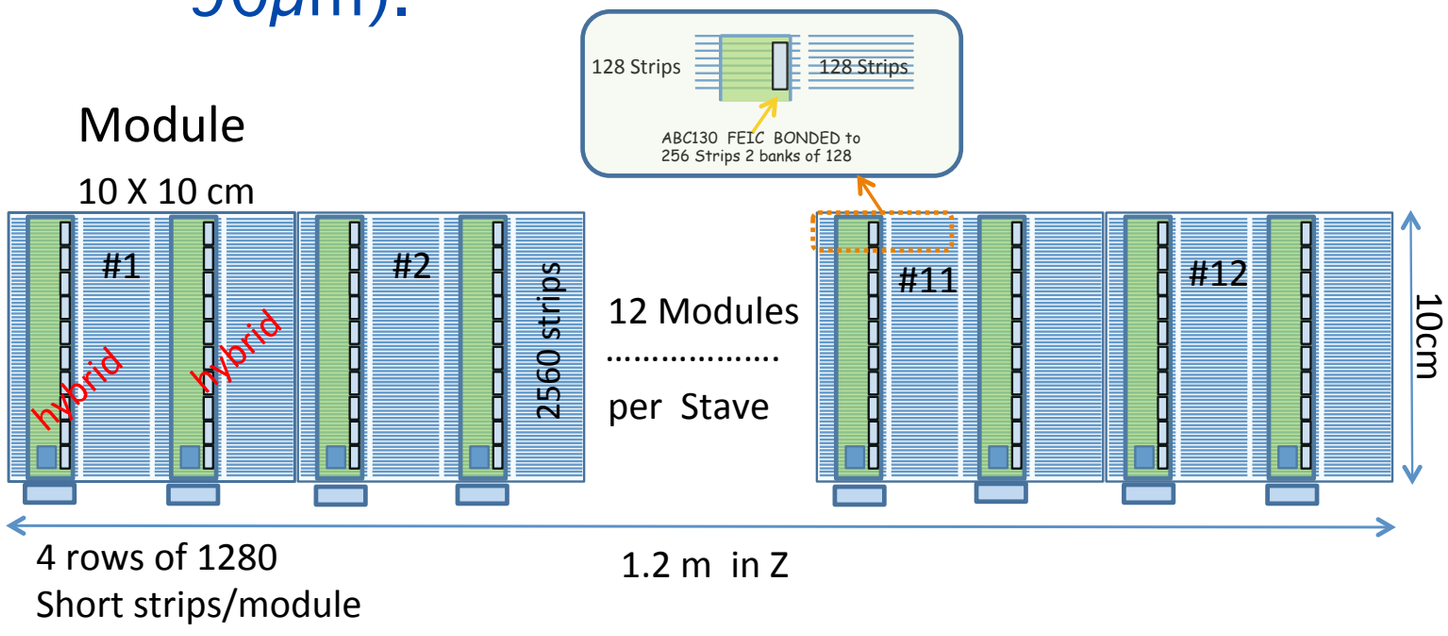


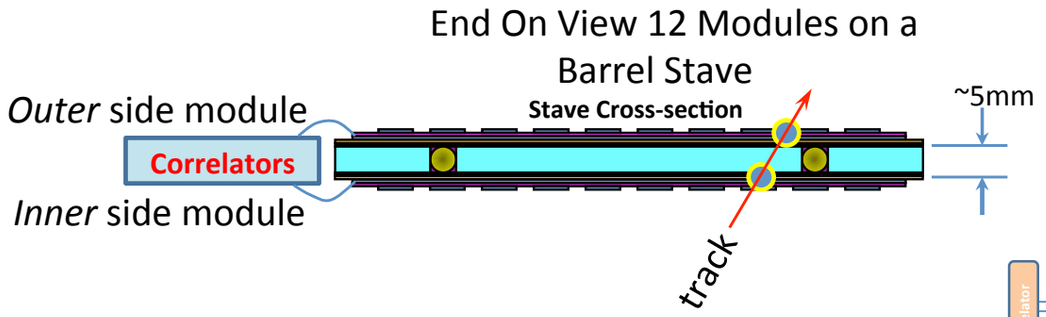


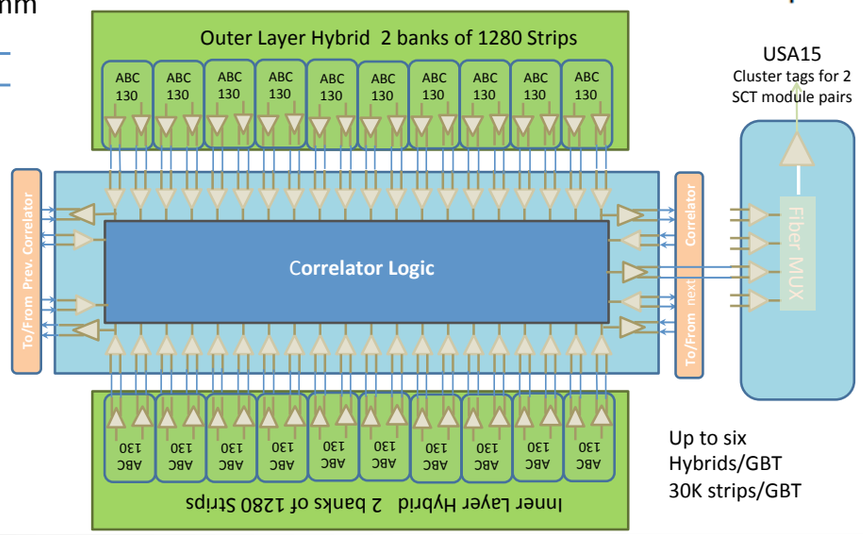
Similar method in ATLAS

ATLAS considering same approach

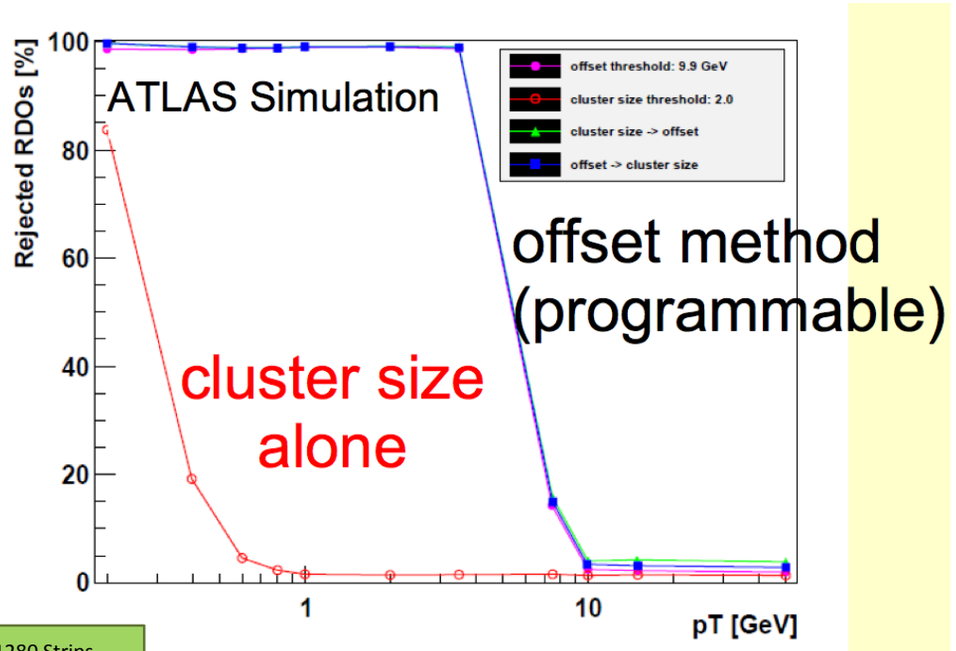
- Needs higher sensors' separation, though smaller pitch (75μm instead of 90μm).





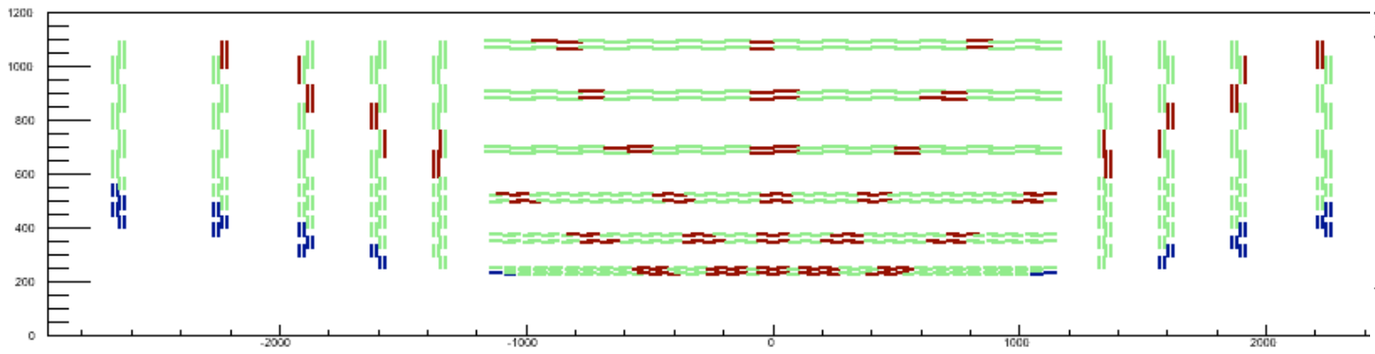


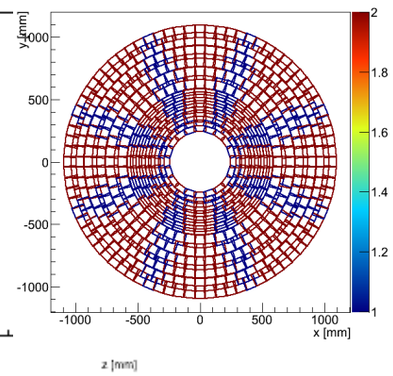
Thresholds@ p_T>10 GeV

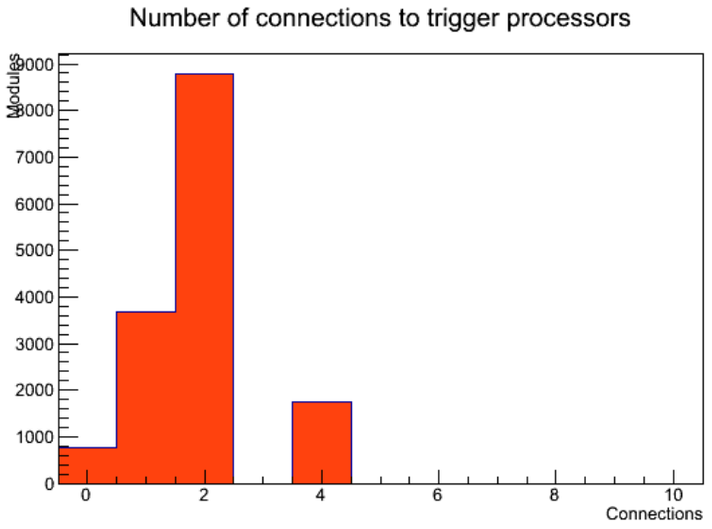


Data organization and dispatch

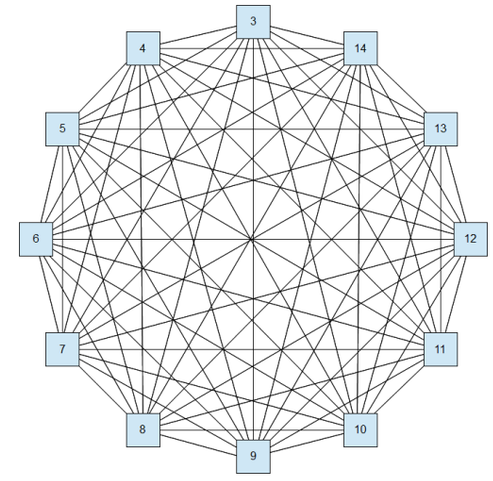
- Subdivide tracker into trigger towers
  - Example CMS: 8(r-ϕ)x6(r-z) trigger sectors (some 10% overlapping)
    - Each sector ~200 stubs on average; tails up to ~500 stubs/event in 140 evts pileup+ttbar (to be compared with ATLAS-Phase 1 ~2000)
      - About 600 Gb/s per one trigger tower



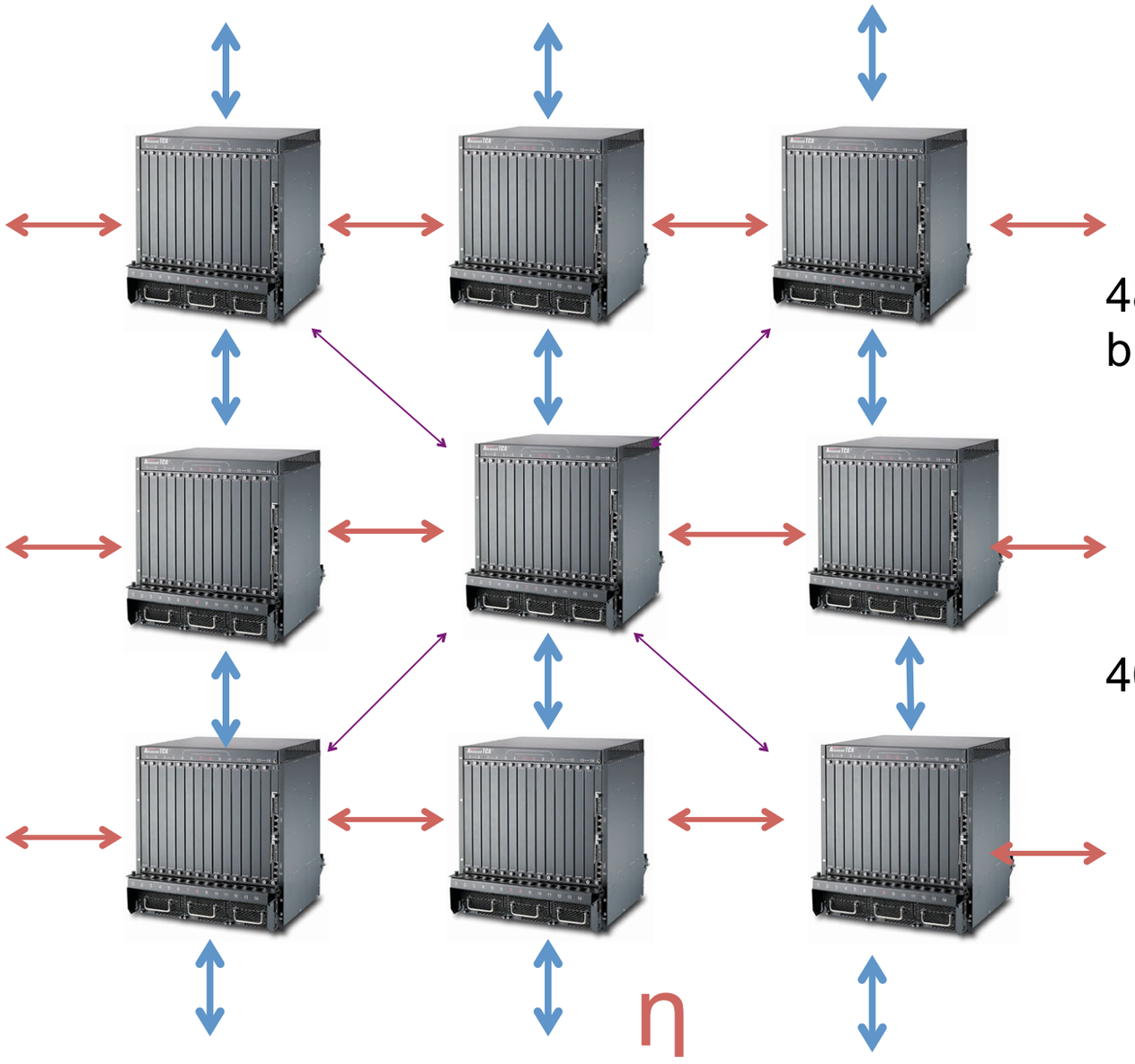




- Send data to Track-finding processors
  - Full mesh ATCA shelfs
    - Capable of “40G” full-mesh backplane on 14 slots = 7.2 Tb/s
    - Several options being investigated, all include time multiplexing data transfer from a set of receiving processors boards to pattern recognition and track finding engines
    - O(10) time multiplexed at the shelf level
    - keep latency < 5 μs, including pattern recognition and track fitting



Tower interconnections



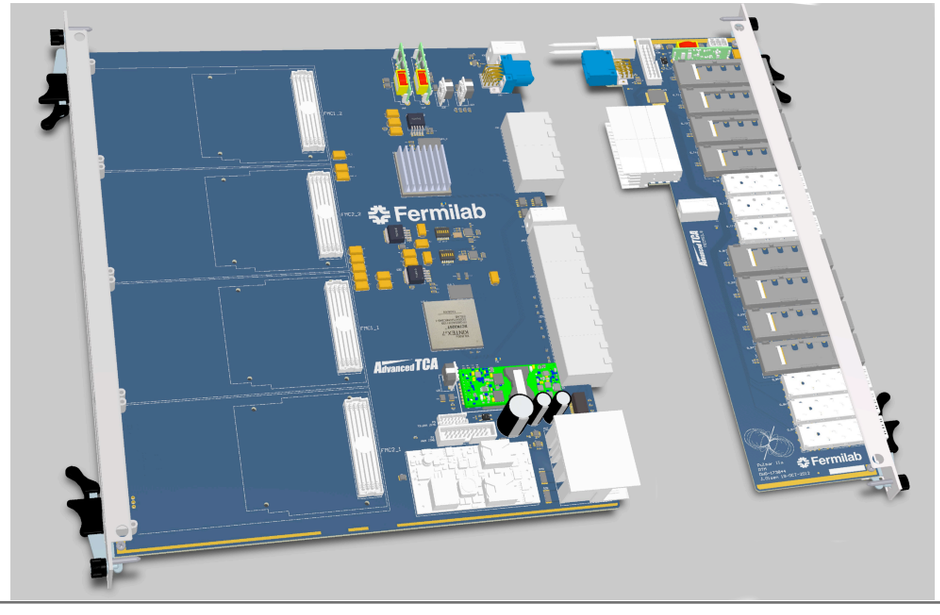
48 x 10 Gbps
bidirectional

φ

40G full-mesh backplane

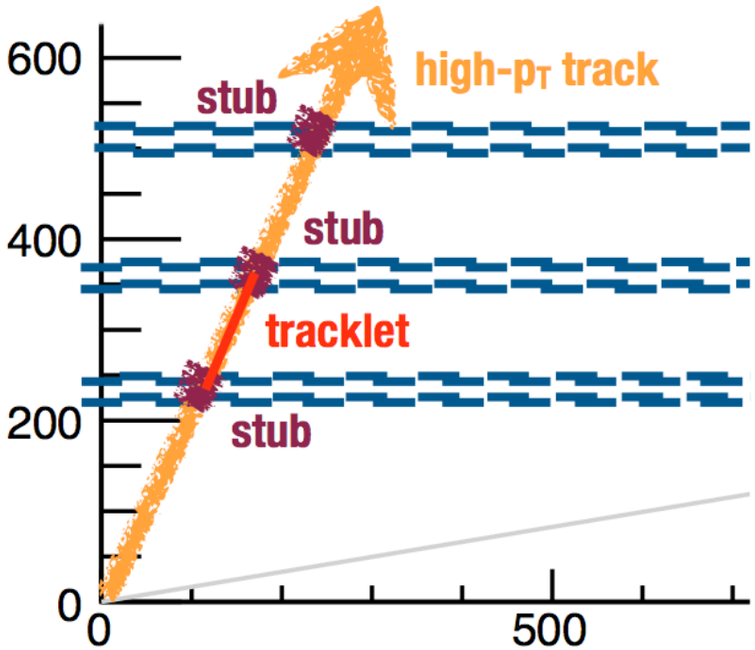
η

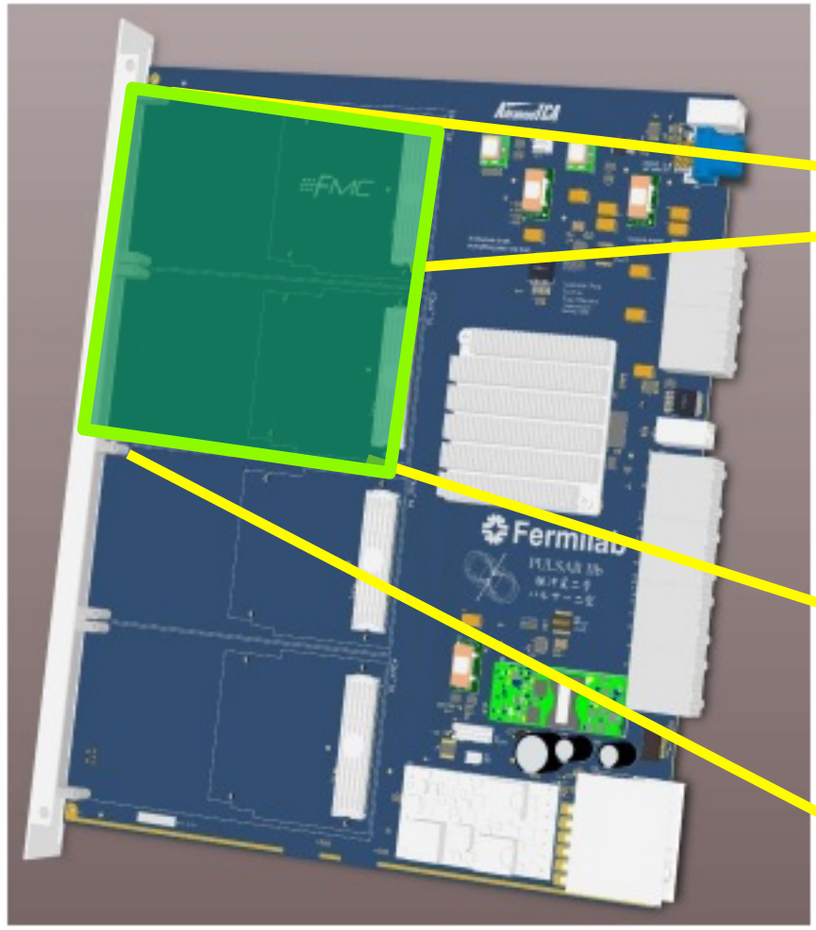
Neighbors
data sharing



Pattern recognition and Track fit

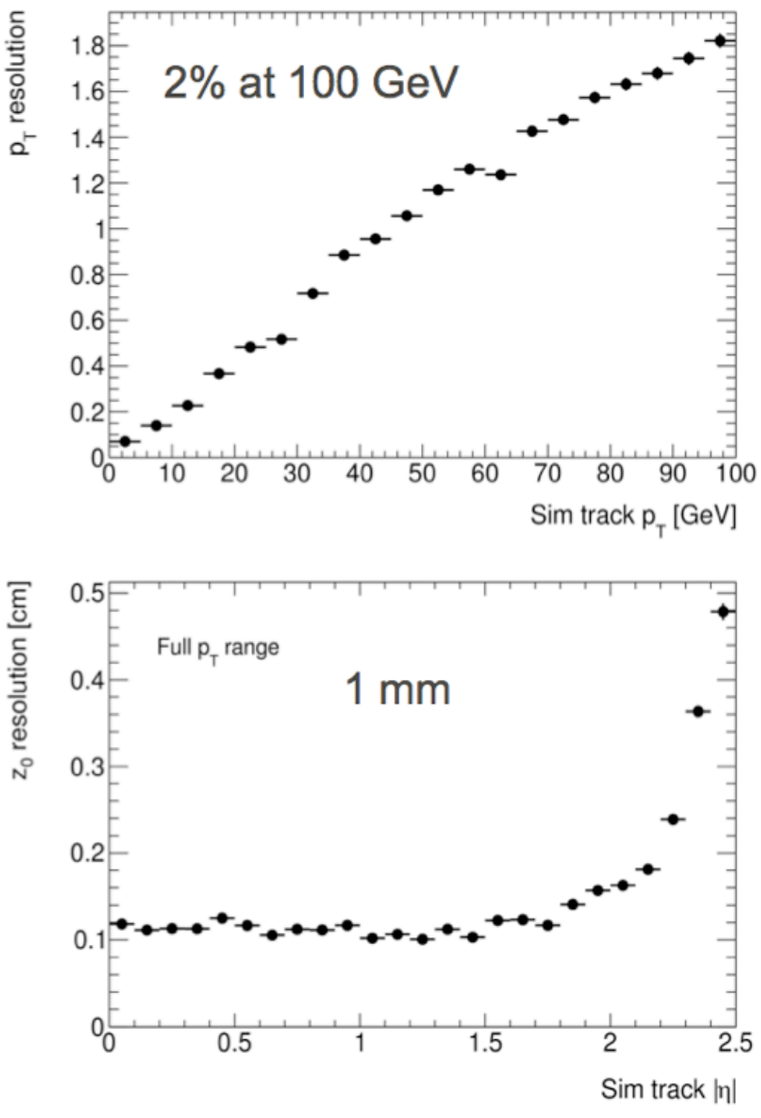
- Associative Memories (pattern recognition) + FPGA (track fit)
  - CMS trigger sectors need ~1M patterns: only 8 state of the art AM06-chip
    - Higher I/O speed (currently 2Gb/s/layer) to reduce time multiplexing
  - ~3000 Track fitting engines using Principal Component Analysis
    - Alternative methods under study (Hough Transform, Retina)
- Alternative approaches under study
  - Purely FPGA based

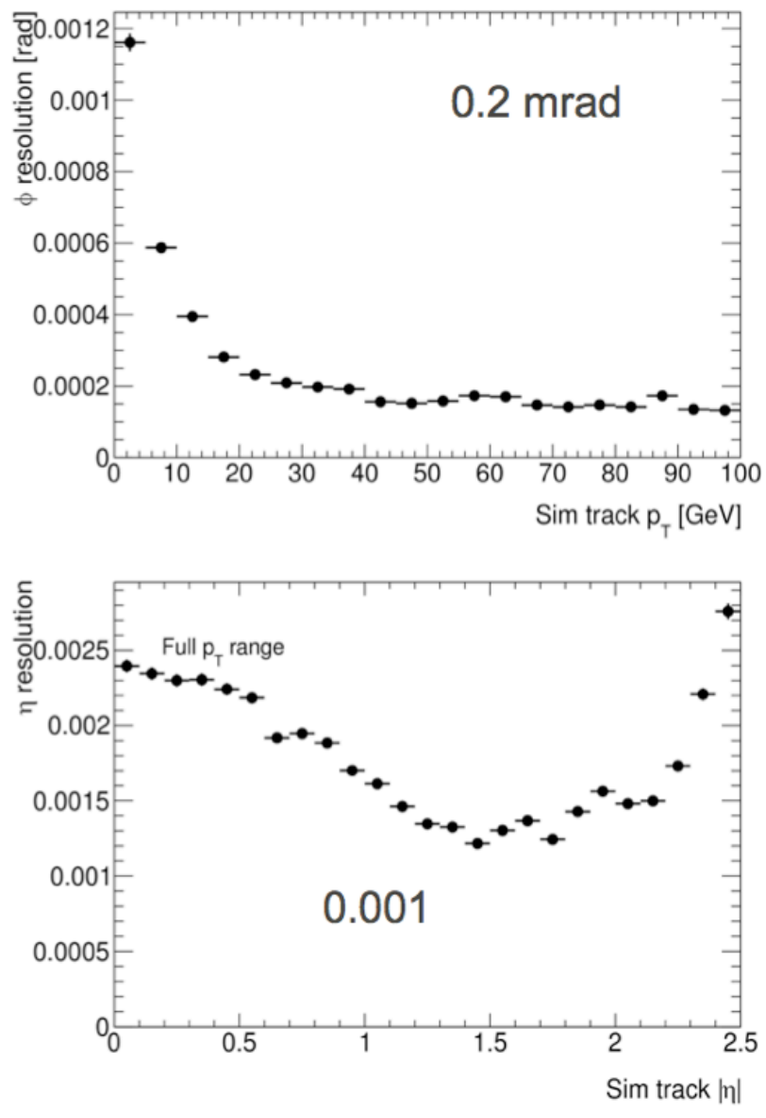




Expected performances

- Pattern recognition efficiency ~99%
- Excellent track parameters resolutions



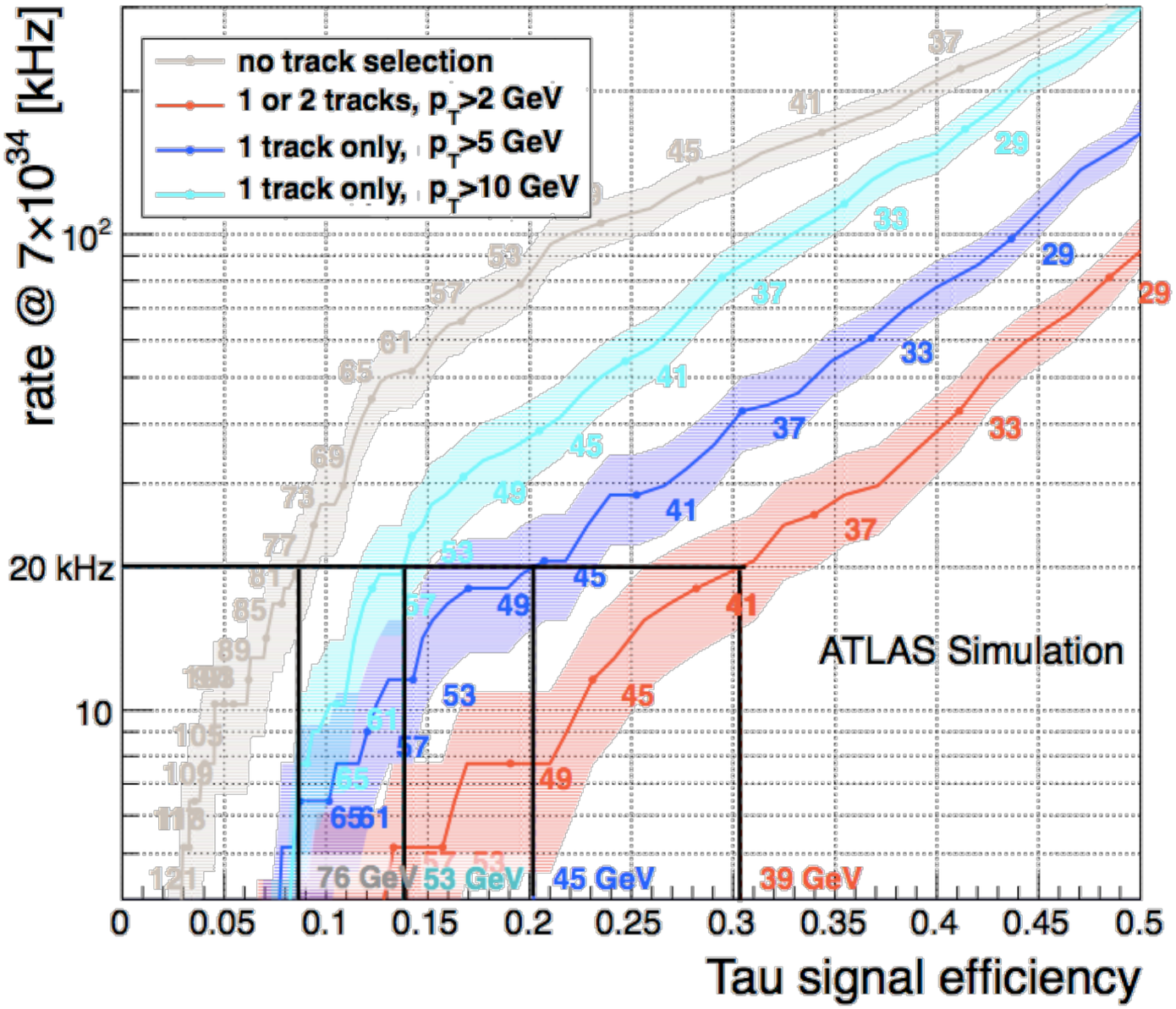


CMS simulation

Usage of L1 Tracks - ATLAS

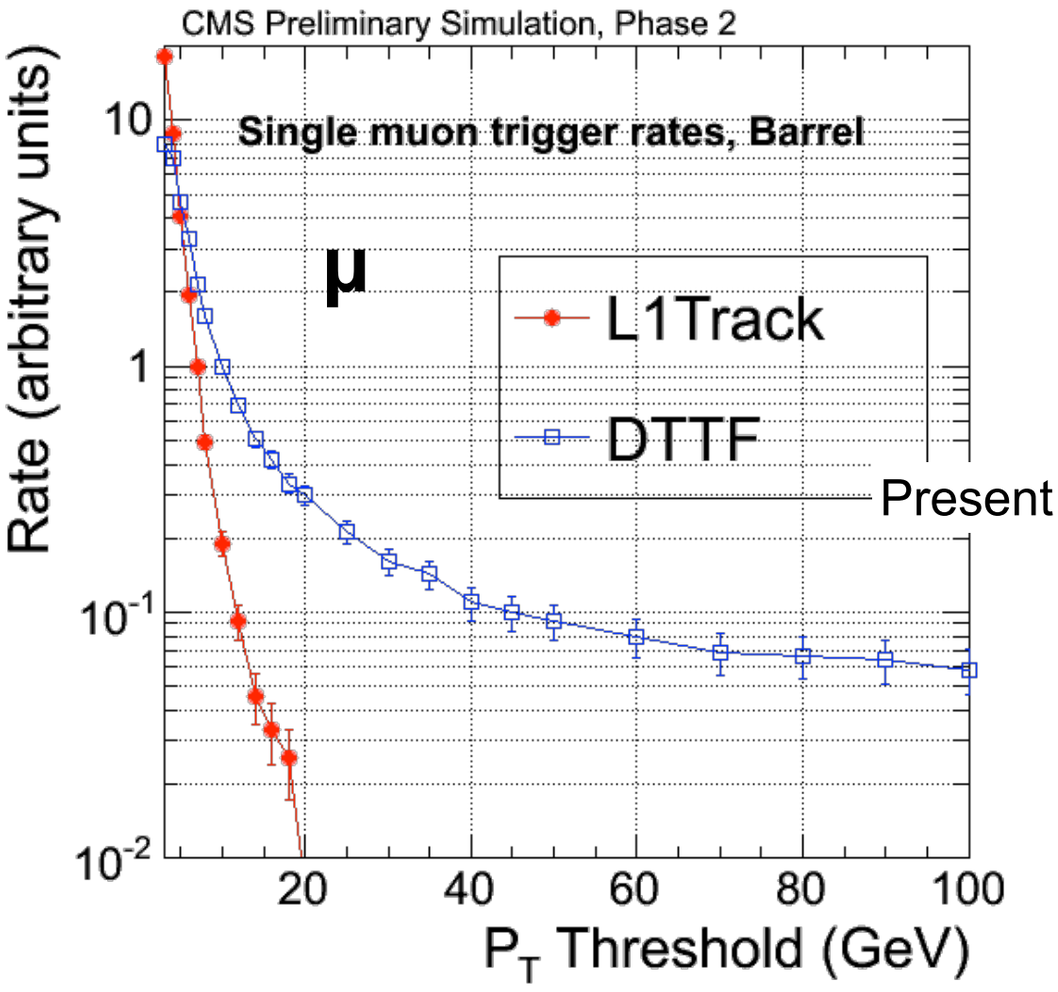
Rate vs. tau finding efficiency curves for taus from the decay of a 120 GeV Higgs boson for the inclusive tau trigger at $7x10^{34}$ cm^{-2} s^{-1} for different track multiplicity and minimum track p_T requirements.

The bands show the rate vs. efficiency parameterized for different L1 cluster E_T thresholds, shown as the small numbers next to the corresponding points on each band.

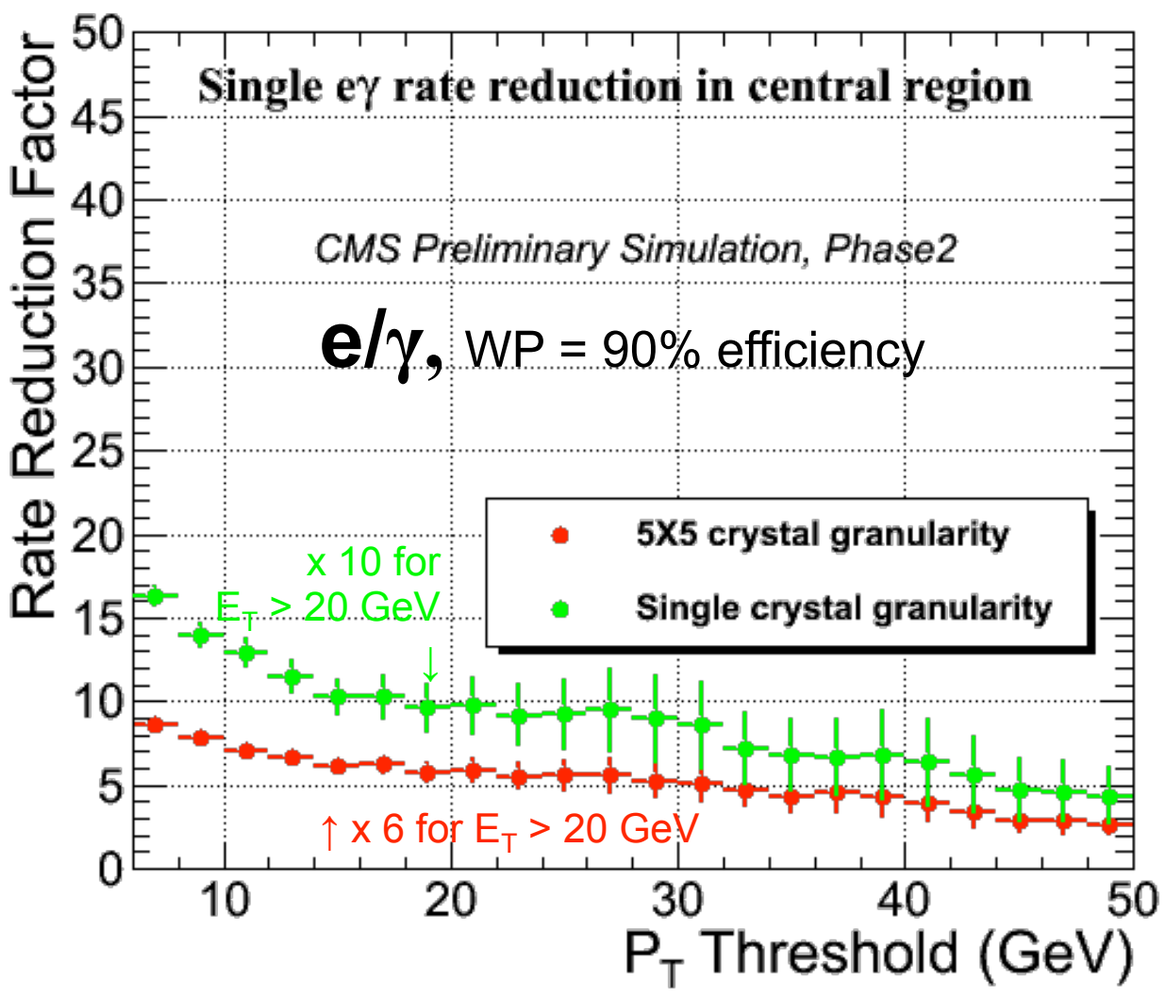


No Trk: 20 kHz, > 76 GeV ⇒ 10% effcy.

Usage of L1 Tracks - CMS



Matching Drift Tube trigger primitives with L1Tracks: **large rate reduction:**
> 10 at threshold > ~ 14 GeV. Normalized to present trigger at 10 GeV. Removes flattening at high P_t



Rate reduction brought by matching L1 e/γ to L1Track stubs for | η | < 1.
Red: with current (5x5 xtal) L1Cal granularity.
Green : using single crystal-level position resolution improves matching

CMS Gains from Track Trigger

- Preliminary simulation studies demonstrate addition of L1 tracking trigger provides significant gains in rate reduction with good efficiency for physics objects. Note these results are “work in progress”.

Trigger, Threshold	Algorithm	Rate reduction	Full eff. at the plateau	Comments
Single Muon, 20 GeV	Improved Pt, via track matching	~ 13 (|η| < 1)	~ 90 %	Tracker isolation may help further.
Single Electron, 20 GeV	Match with cluster	> 6 (current granularity) >10 (crystal granularity) (| η | < 1)	90 %	Tracker isolation can bring an additional factor of up to 2.
Single Tau, 40 GeV	CaloTau – track matching + tracker isolation	O(5)	O(50 %) (for 3-prong decays)	
Single Photon, 20 GeV	Tracker isolation	40 %	90 %	Probably hard to do much better.
Multi-jets, HT	Require that jets come from the same vertex			Performances depend a lot on the trigger & threshold.

Pros and cons

- CMS (based on push architecture)
  - Pros:
    - only a fraction of the tracker data readout - p_T filtering
      - ~200 stubs/sector - 48 sectors
    - large flexibility to use tracks in Global Trigger (including MET)
  - Cons:
    - readout of tracker (trigger) data at 40 MHz
- ATLAS (based on pull architecture)
  - Pros:
    - only portions of tracker data readout at 500 kHz
  - Cons:
    - hits from low p_T tracks not filtered: a problem for pattern recognition
      - Phase 1 ~200-500 hits/sector in 64 sectors
      - Phase 2: increase by a factor ~10 ⇒ larger data rate
    - improvements limited only to few “objects” (μ, E/γ, τ, jets) with L0A compliant rates

Current INFN R&D interests

- INFN interests:
  - CMS
    - Electronics of the PS modules (Pv)
    - L1 track finding (Fi, Pd, Pi, Pg, Ts)
  - ATLAS
    - Electronics (Bo, Ge, Mi)
    - Mechanics (Ge, Mi)
    - L1 track finding (LNF, Mi, Pi, Pv)
- L1 track finding INFN (see A. Annovi's talk)
  - CMS + ATLAS
    - Tracking algorithms (PCA and retina) simulation and firmware development
    - FMC fabrication (includes AM procurements)
    - ATCA and DAQ development
    - New AM chip developments

Collaborations and related projects

- Collaborations and related projects on L1 Track finder
  - On-going collaborations
    - AM chip:LPNHE, IMEC
    - ATCA: FNAL , Kalrsruhe , Lyon, Nortwestern
    - Simulation: LPNHE, Lyon, FNAL, Kalrsruhe , Lyon, Nortuestern, UCL, Uppsala, Purdue, Cornell, CERN, India
  - On-going projects
    - FP7-PEOPLE-2012-IAPP: P. Giannetti
    - PRIN 2012: H-TEAM: G. Tonelli
    - ANR:(LPNHE, IPNL, Lyon)
    - FP7-PEOPLE-ITN INFIERI: F. Palla
  - Future applications for funds
    - SIR, ERC, Pillar II (ICT-4)

Conclusions

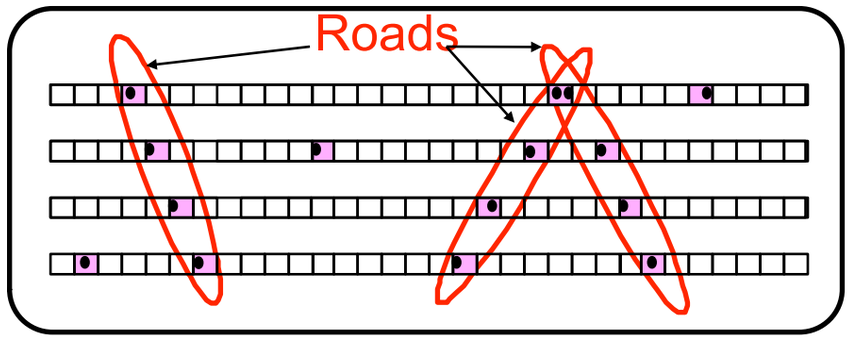
- Tracker detector helps drastically to reduce the rate of uninteresting events at L1
  - Several trigger architectures exploited
    - On-detector data reduction using pT-modules (CMS), L0 pre-trigger (ATLAS)
    - Implications on Tracker detector layouts ongoing
  - Some demonstrators being built to validate the full chain
  - Large gains in combining tracking with other subdetectors
    - Electrons, Muons, Jets and MET

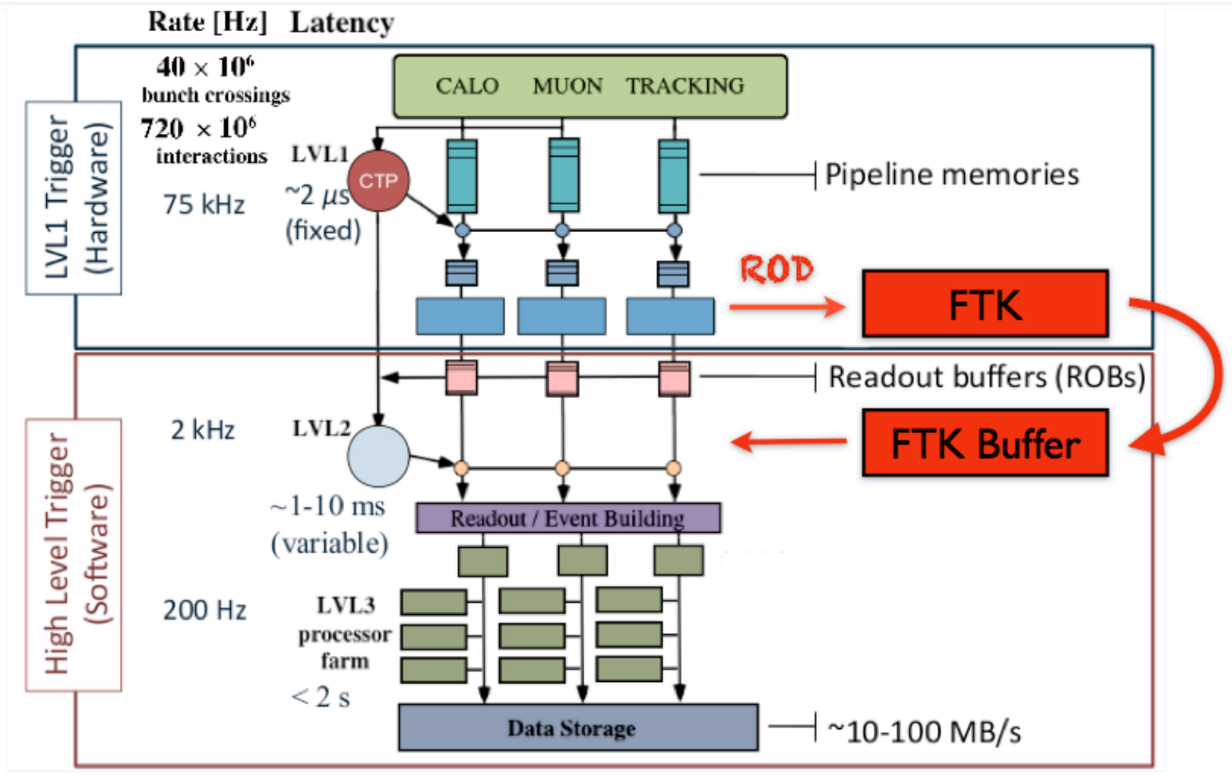
Backup

ATLAS Fast Track - AM based

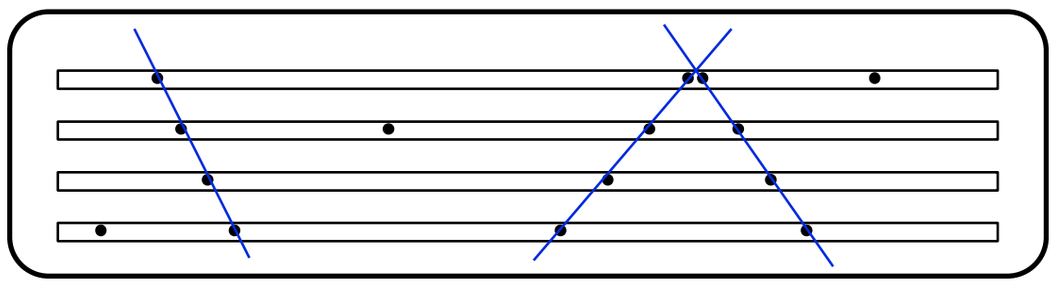
FTK: Fast TracK processor - 20

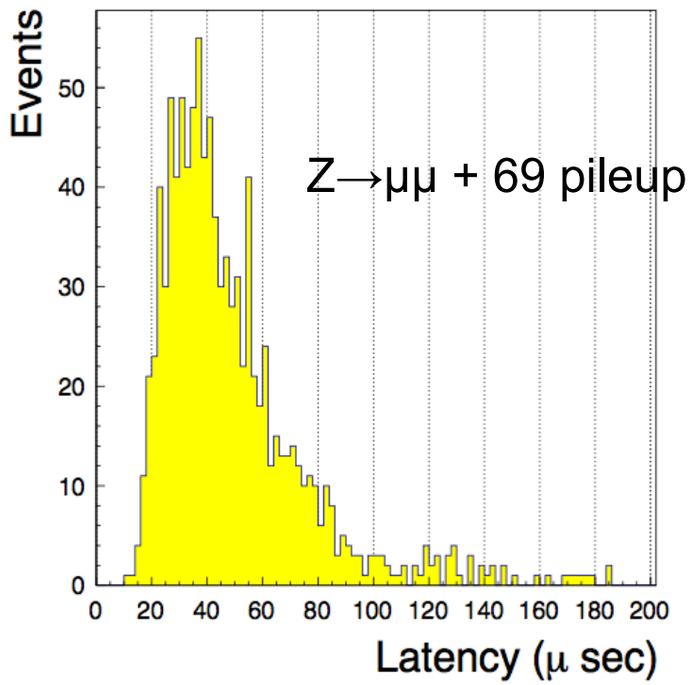
- Uses Associative Memory (AM) approach
  1. Pattern recognition: using coarse resolution with AMs





3. Fit tracks using high-resolution hits with FPGAs. Linear approximation, (instead of full helix) with pre-computed constants





R&D Topics: Trigger

- Increase of rate from Level-0 to HLT to read out
  - Absolute rate & balance between levels
  - L1 complexity vs. HLT input rates
- L1 Trigger Latency
  - How much is needed & consequences on electronics
- L1 Track Triggers
  - Associative Memories
  - Study techniques: sharpen p_T threshold, e- & μ- ID, Isolation, primary vertex for jets, multi-object triggers, possibility of pixel b-tag.
  - Interplay with tracker design
- Impact of higher bandwidth links & denser optical interconnects
- New packaging & interconnect technologies
  - ATCA, μTCA
- Use of FPGAs in L1 Trigger

Issues for FTK for ATLAS Phase 2

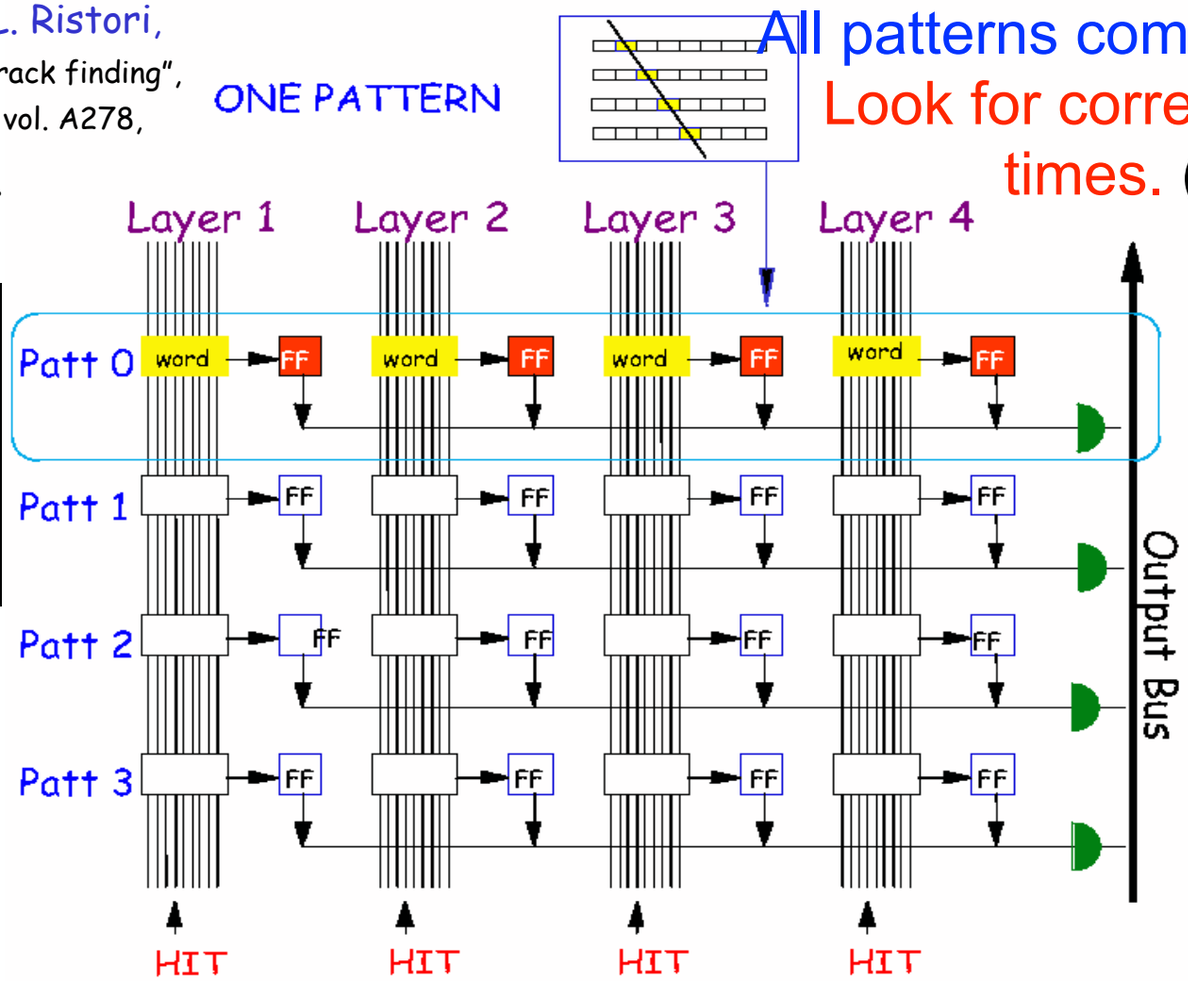
- Increased bandwidth ~12 due to
  - the larger pileup (x2.5)
  - x5 increase in the rate (500 kHz instead of 200 kHz)
- Implications
  - replace DF using an ATCA based system
  - increase the lower pT threshold from 1 to 2 GeV
  - increase in the number of patterns by ~one order of magnitude
  - increase the speed of the processing to cope with the shorter latency

AM working principle

M. Dell'Orso and L. Ristori,
"VLSI structures for track finding",
Nucl. Instr. and Meth., vol. A278,
pp. 436-440, (1989).

All patterns compared in parallel with incoming data.
Look for correlation of data received at different times. (feature unique to AM chip)

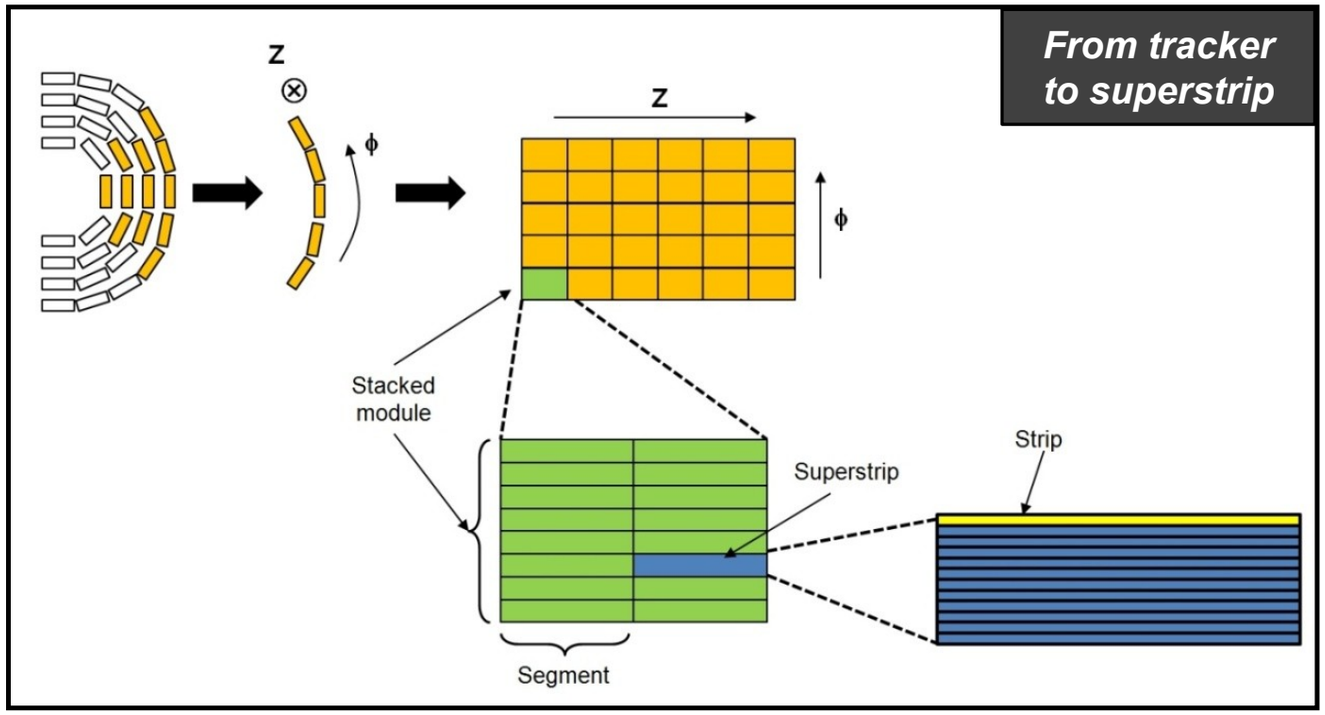
1 register
1 comparator
1 match FF
/ layer
/ pattern



Final chip (AM06) to be submitted by mid-2014
128k patterns/8 layers
100 MHz clock frequency
Serial bus

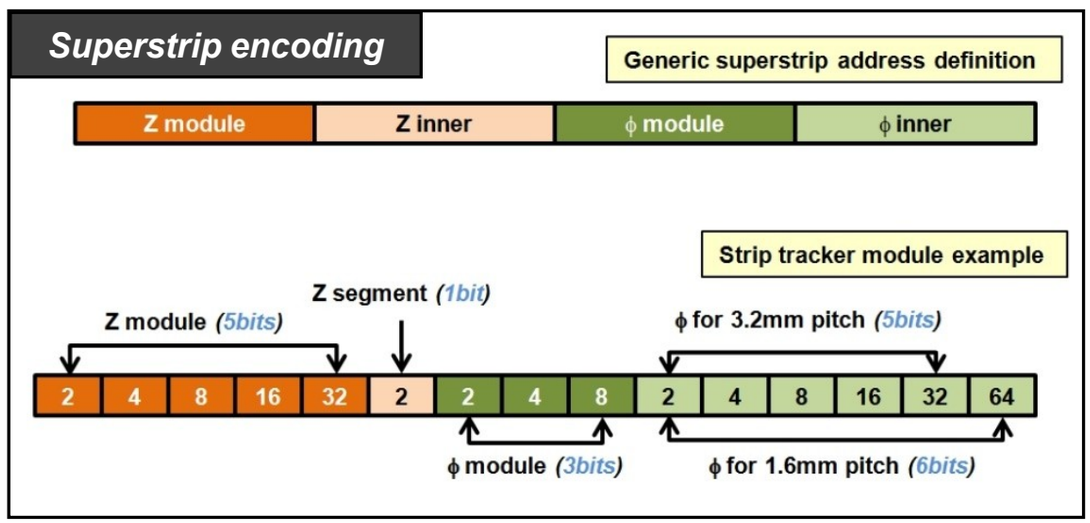
The pattern

→ **Superstrip definition:**



→ **A superstrip is simply a bunch of strips in one module of the tracking detector.**

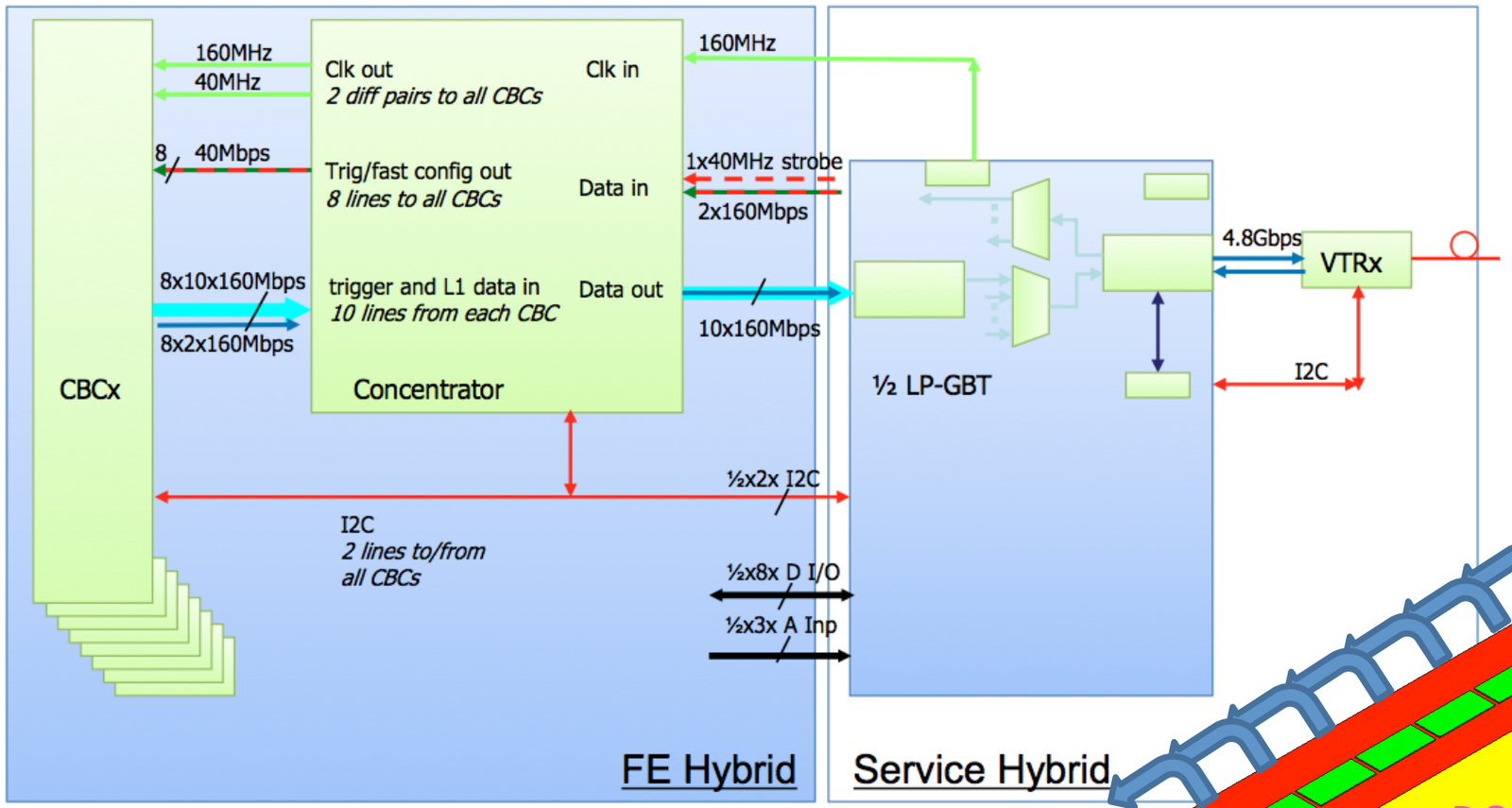
→ **The superstrip address is the info sent to the AM board. Is is coded on a certain number of bits, depending on the superstrip resolution.**

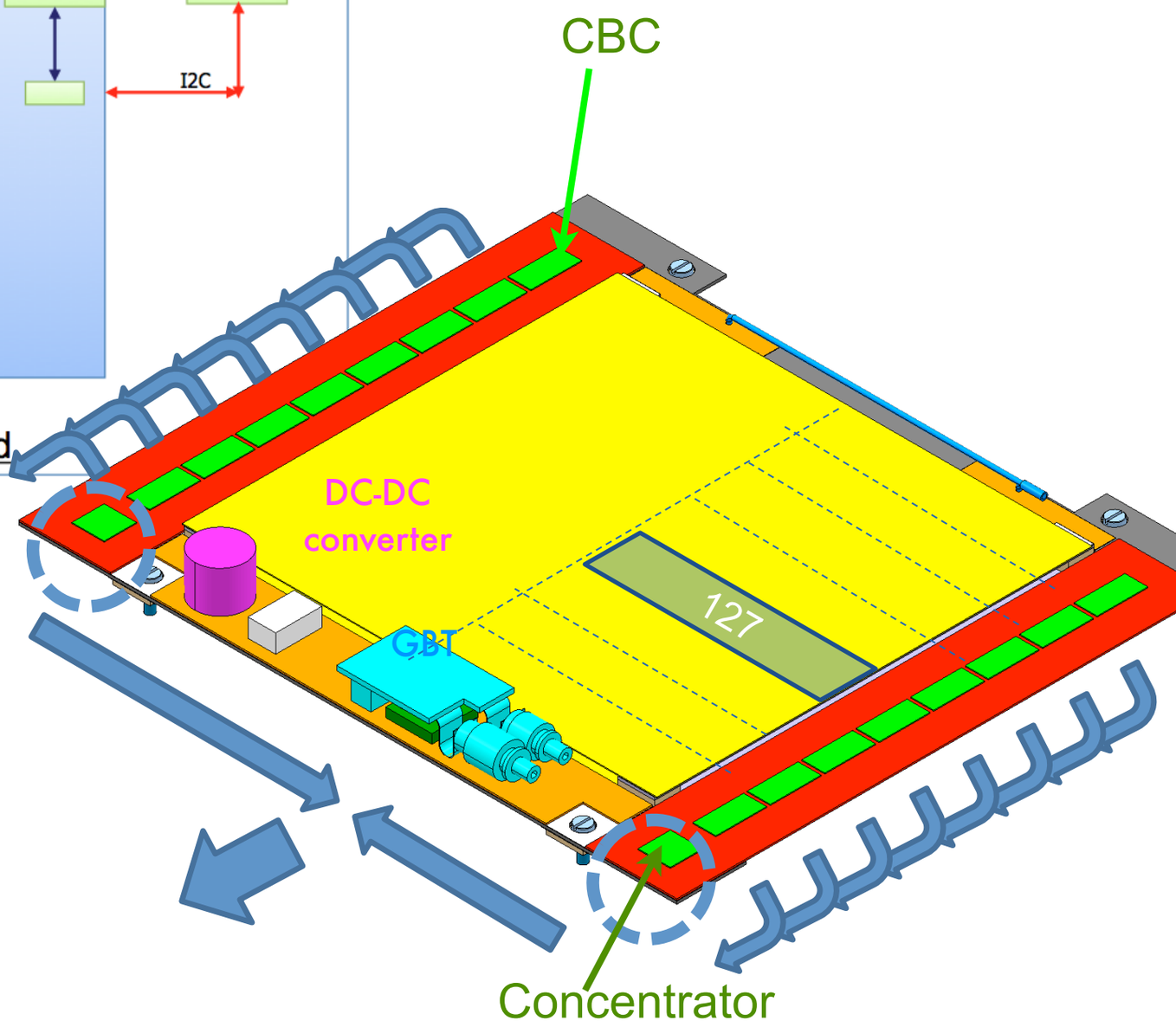


→ **The encoding is divided into 4 parts, giving module and intra-module SS position in Z and ϕ direction (*R is not necessary*)**

→ **We are not using pixel info yet, so our Z intra-module encoding is very basic for the moment.**

Electronics





CBC outputs up to 3 stubs/bx, 160 MHz x 10 bits

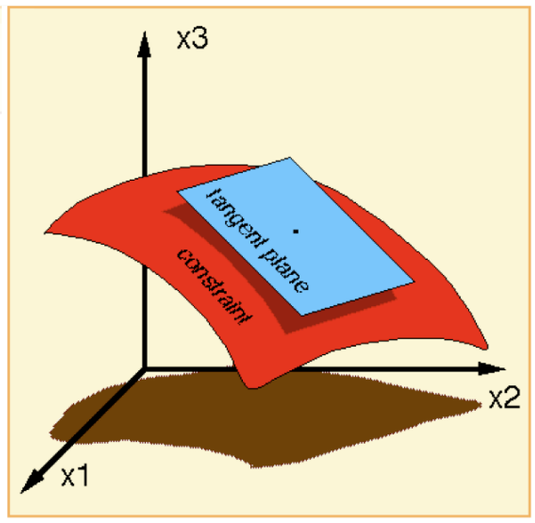
Concentrator chip receives 8 CBC and sends out up to 12 stubs/8bx, 160 MHz

Track fitting - high quality helix parameters and χ^2

- Principal component analysis (Other techniques under consideration)
  - Over a narrow region in the detector, equations linear in the local silicon hit coordinates give resolution nearly as good as a time-consuming helical fit.

Nucl.Instrum.Meth.A623:540-542,2010
doi:10.1016/j.nima.2010.03.063

$$p_i = \sum_{j=1}^{14} a_{ij} x_j + b_i$$

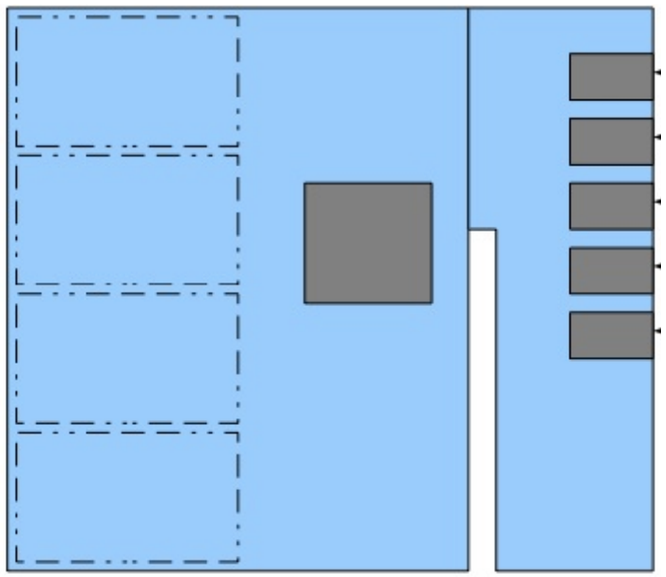


- p_i's are the helix parameters and 2 components.
- x_j's are the hit coordinates in the silicon layers.
- a_{ij} & b_i are pre-stored constants determined from full simulation or real data tracks.
- The range of the linear fit is a "sector" which consists of a single silicon module in each detector layer.
- This is VERY fast in FPGA DSPs.

- ~3000 fitting engines/trigger sector for CMS

CMS: toward a demonstrator

Input Data Board x 8

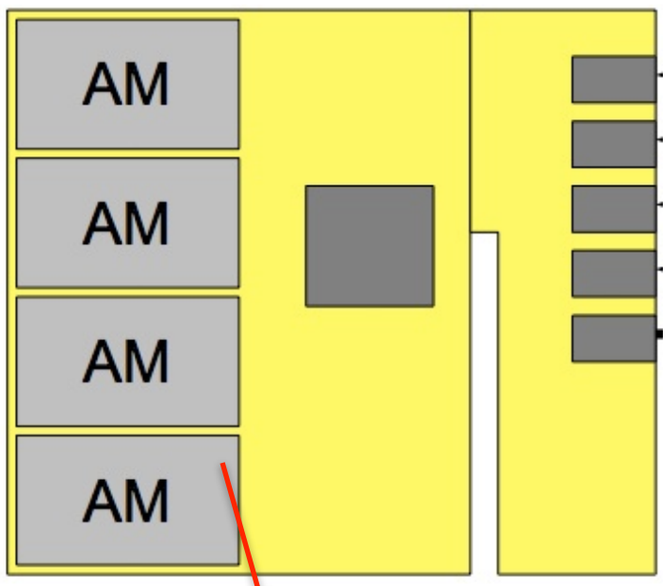


Up to 48 pairs of connection, Bidirectional.

Up to 48 inputs with 3.25Gbps each

With 40G full-mesh backplane

Pattern Recognition Board x 4



Up to 48 pairs of connection, Bidirectional.

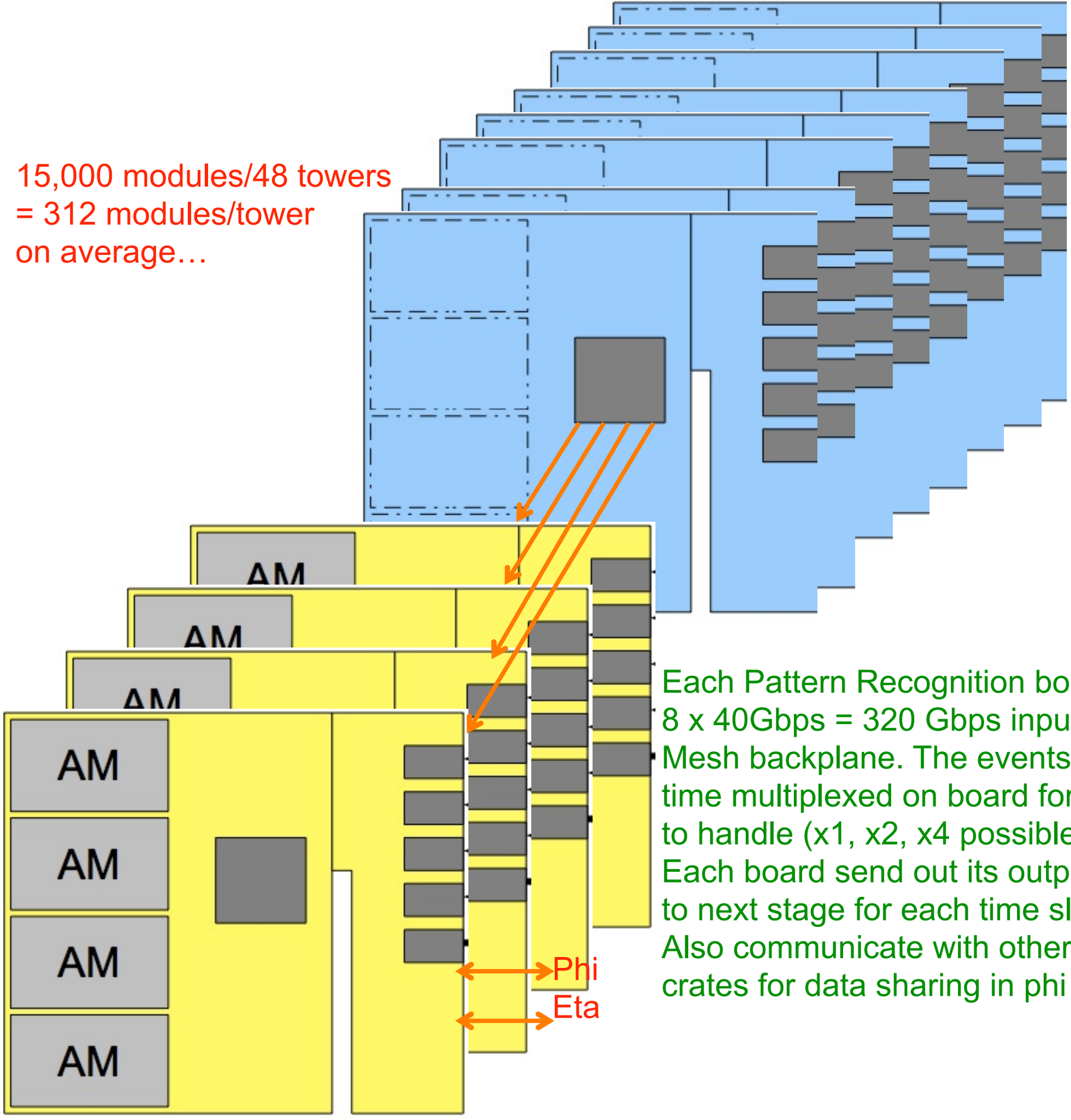
10Gbps each

With 40G full-mesh backplane

One Trigger Tower

Could also be pure FPGA based approach

15,000 modules/48 towers
= 312 modules/tower
on average…



Each board is capable to receive data from up to 48 modules at 3.25Gbps, with total 156 Gbps per Board/RTM. 8 boards can receive up to 384 modules (one trigger tower worth)

The input data is then divided into 4 time slices, each slice is sent to 1 of 4 Pattern Recognition board, with 40Gbps full-mesh (4x40=160Gbps > 156Gbps).

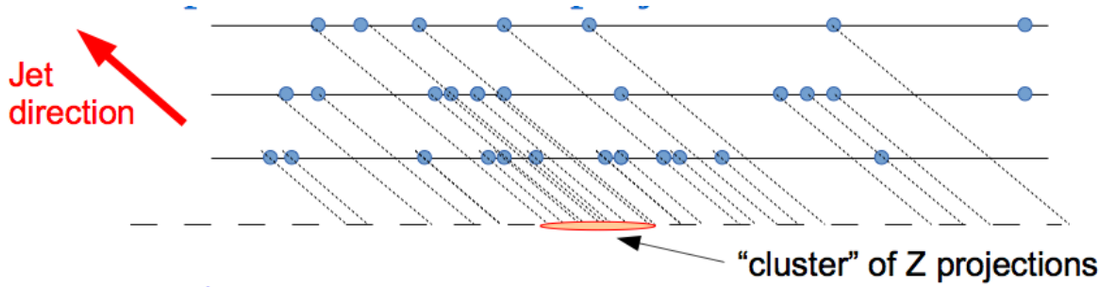
Each Pattern Recognition board receives *up to* 8 x 40Gbps = 320 Gbps input data over full Mesh backplane. The events can then be time multiplexed on board for each mezzanine to handle (x1, x2, x4 possible, flexible).
Each board send out its output from RTM to next stage for each time slice.
Also communicate with other boards in other crates for data sharing in phi & eta for each time slice

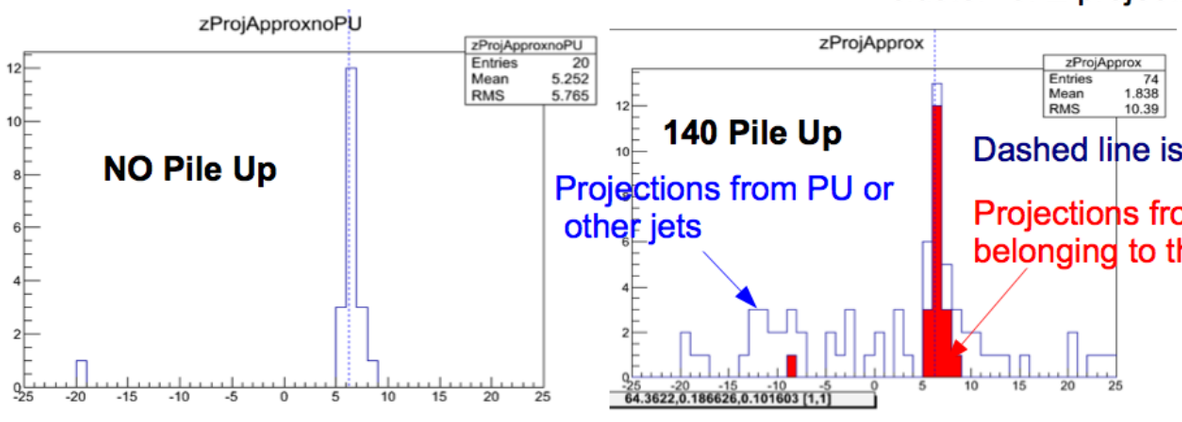
First ideas on Pixel@L1at CMS

Considering to use 3rd generation pixels detectors to trigger

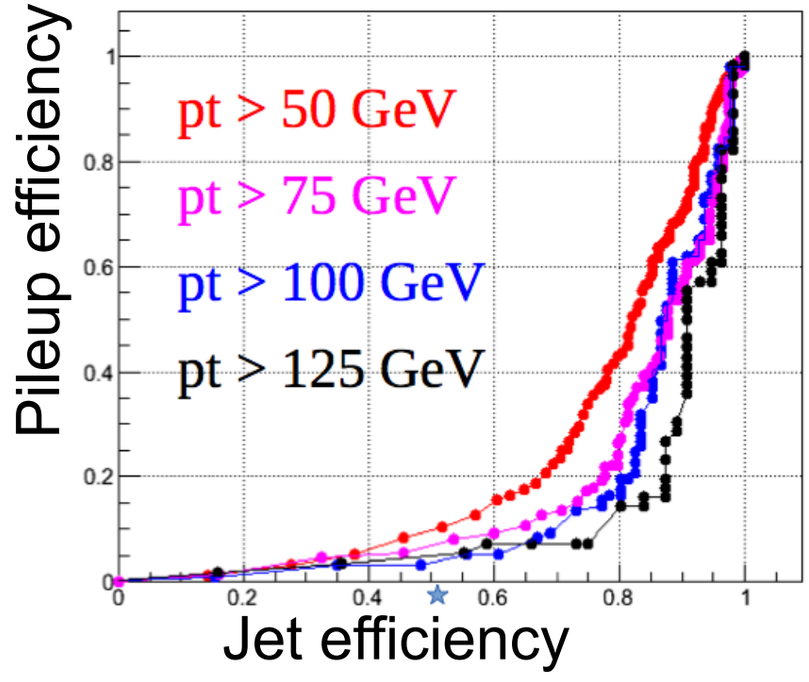
- 65 nm chip, 2x2 cm^2: output bandwidth ~3 Gbps
- Main problem is the bandwidth (~1 GHz/cm^2 hits)
  - Need a L0 trigger and a clusterization algorithm on chip.

Need 20 μs latency

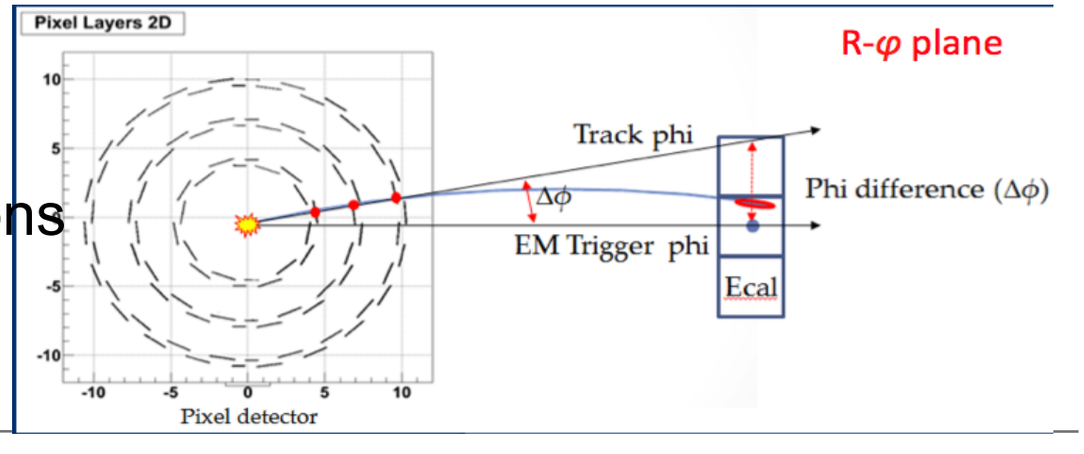




Preliminary!



4 Electrons



Current involvement CMS

- Electronics R&D
  - FE Asics: UK, France, Italy, CERN
  - MPA assembly, including TSV: CERN and US
  - Module design and assembly: CERN, Germany, UK, US
  - Mechanics: CERN, Germany, France, US
  - DAQ: CERN, Germany, France, UK, India
  - L1 Track finding: CERN, Italy, France, UK, US
- L1 track finding (conveners: F. Palla, T. Liu)
  - Simulations (both at High and Low Level): INFN, FNAL, Lyon, Cornell, RAL, KIT
  - Hardware demonstrator (see Annovi's talk): INFN, FNAL, Lyon, Cornell, Northwestern
    - INFN Interests
      - Tracking algorithms (PCA and retina) simulation and firmware development
      - FMC fabrication (includes AM procurements)
      - ATCA development

ATLAS & CMS Triggered vs. Triggerless Architectures

- 1 MHz (Triggered):
  - Network:
    - 1 MHz with ~5 MB: aggregate ~40 Tbps
    - Links: Event Builder-cDAQ: ~ 500 links of 100 Gbps
    - Switch: almost possible today, for 2022 no problem
  - HLT computing:
    - General purpose computing: 10(rate)x3(PU)x1.5(energy)x200kHS6 (CMS)
      - Factor ~50 wrt today maybe for ~same costs
    - Specialized computing (GPU or else): Possible
- 40 MHz (Triggerless):
  - Network:
    - 40 MHz with ~5 MB: aggregate ~2000 Tbps
    - Event Builder Links: ~2,500 links of 400 Gbps
    - Switch: has to grow by factor ~25 in 10 years, difficult
  - Front End Electronics
    - Readout Cables: Copper Tracker! – Show Stopper
  - HLT computing:
    - General purpose computing: 400(rate) x3(PU)x1.5(energy)x200kHS6 (CMS)
      - Factor ~2000 wrt today, but too pessimistic since events easier to reject w/o L1
      - This factor looks impossible with realistic budget
    - Specialized computing (GPU or ...)
      - Could possibly provide this ...