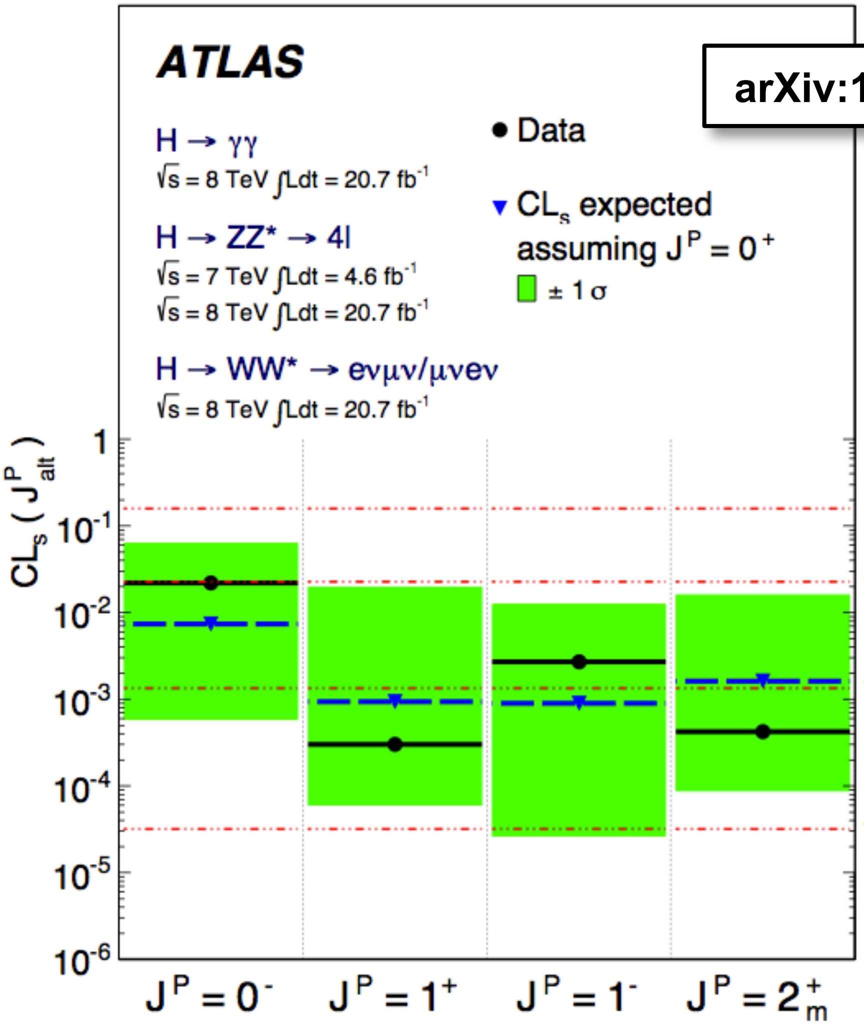
Challenges and opportunities for a First Level Trigger based on Silicon Trackers for the High Luminosity LHC

Fabrizio Palla
INFN Pisa and CERN

Run1 legacy

Higgs-like particle has now become a (the) Higgs boson

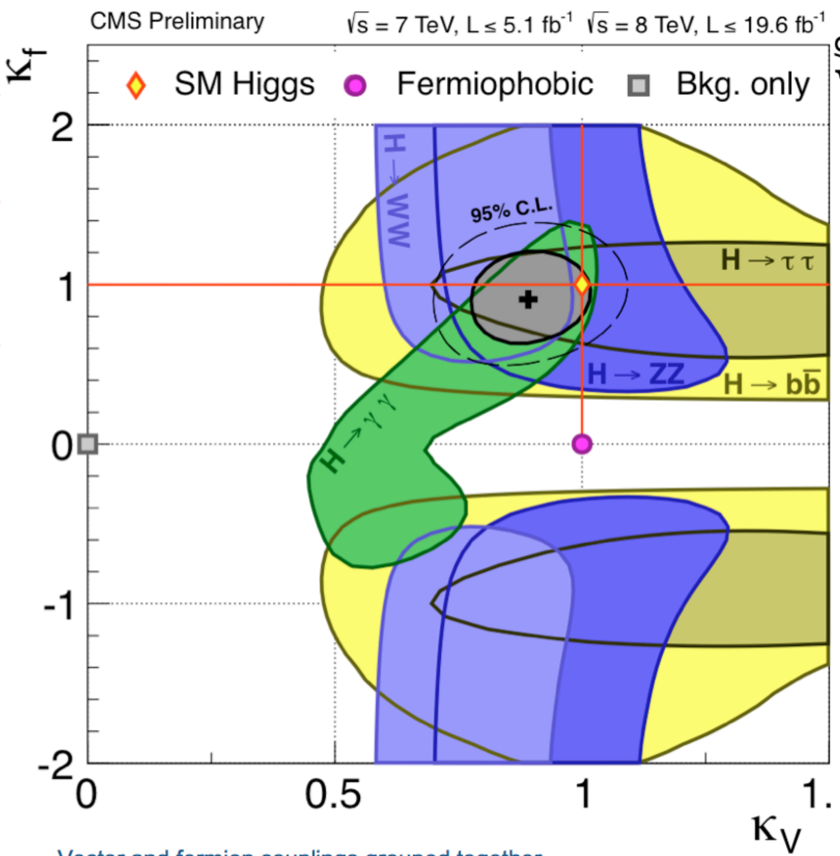


arXiv:1307.1432

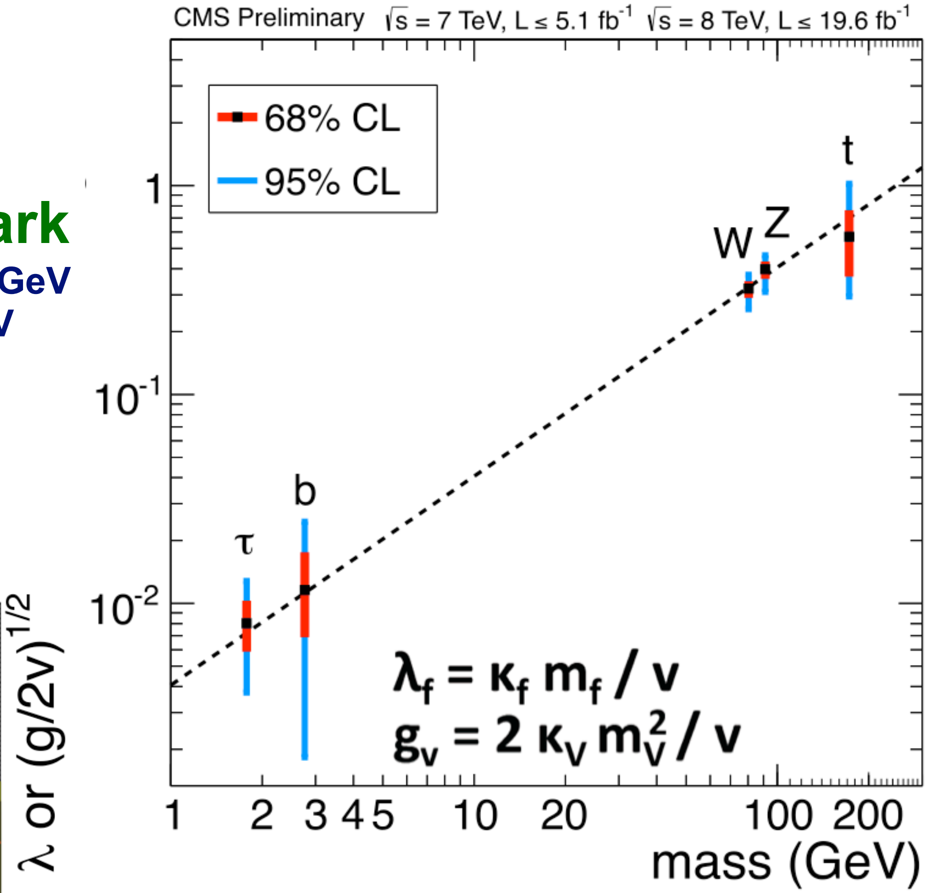
$J^{PC} = 0^{++}$ strongly favored over alternative hypotheses

Higgs mass already measured more precisely than top quark

ATLAS: m(H) = 125.5 ± 0.2 $^{+0.5}_{-0.6}$ GeV
CMS: m(H) = 125.5 ± 0.3 ± 0.3 GeV



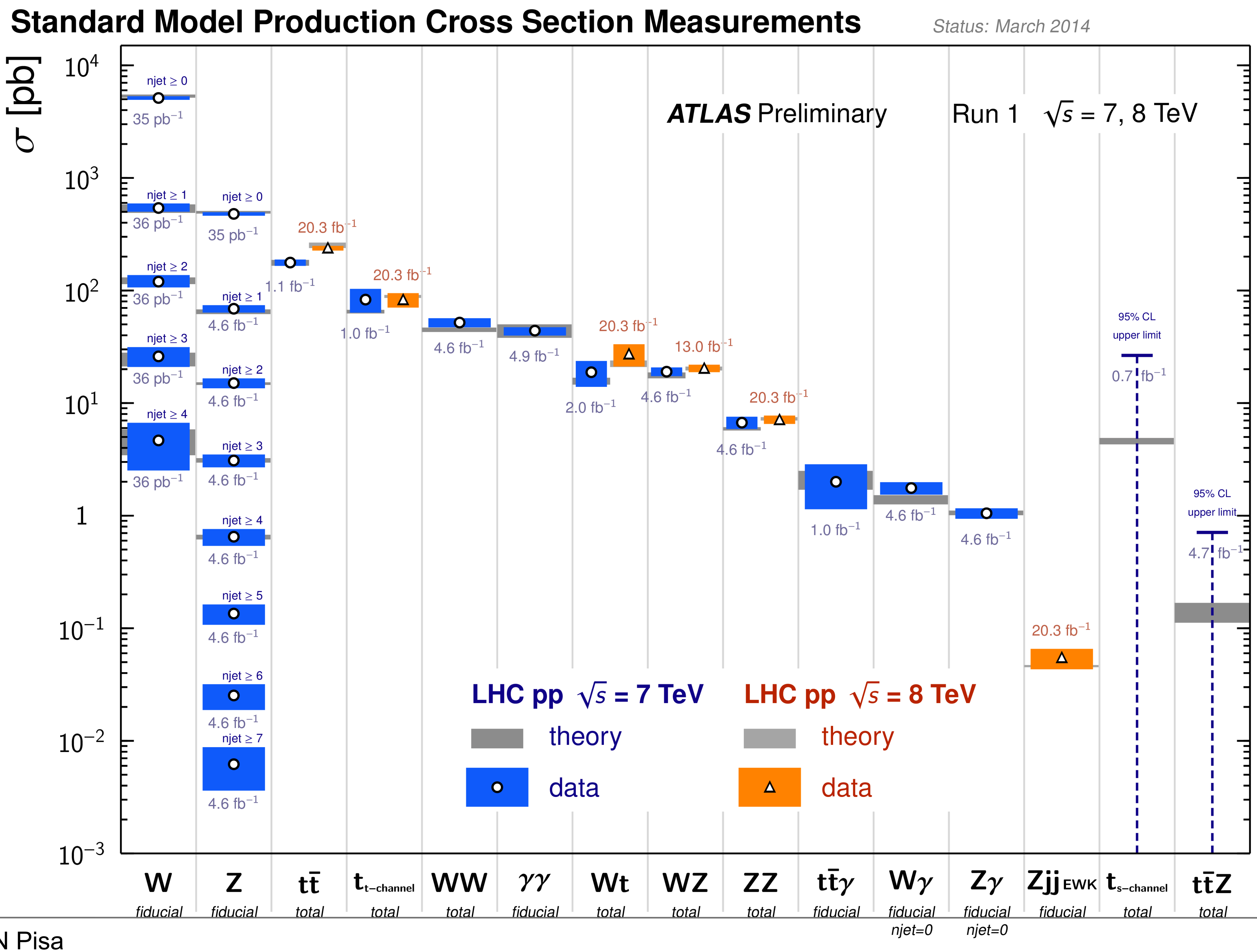
Vector and fermion couplings grouped together



$$\lambda_f = \kappa_f \, m_f / v$$
$$g_V = 2\,\kappa_V \, m_V^2 / v$$

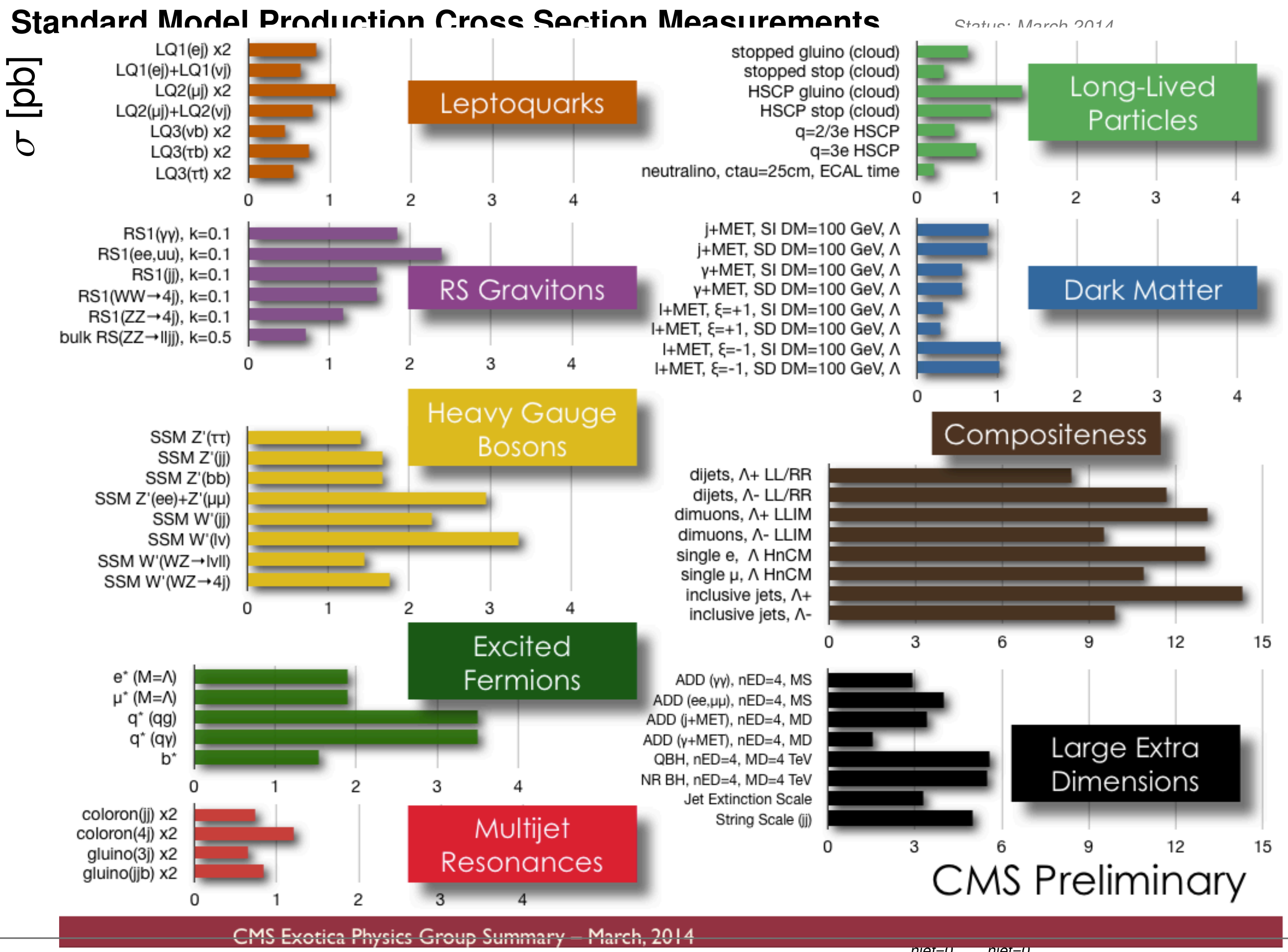
No evidence of New Physics (yet!)

No evidence of New Physics (yet!)



F. Palla INFN Pisa

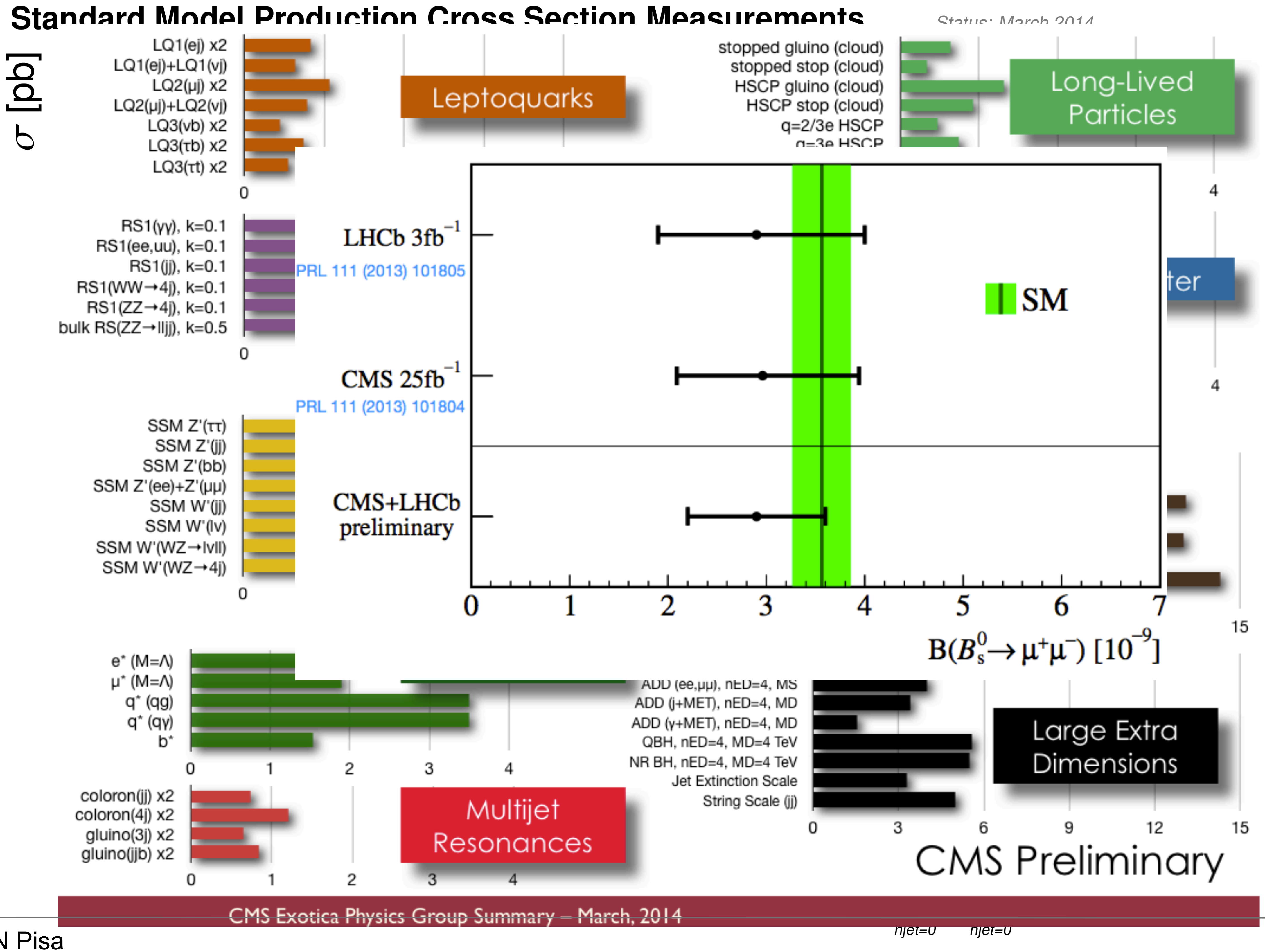
No evidence of New Physics (yet!)



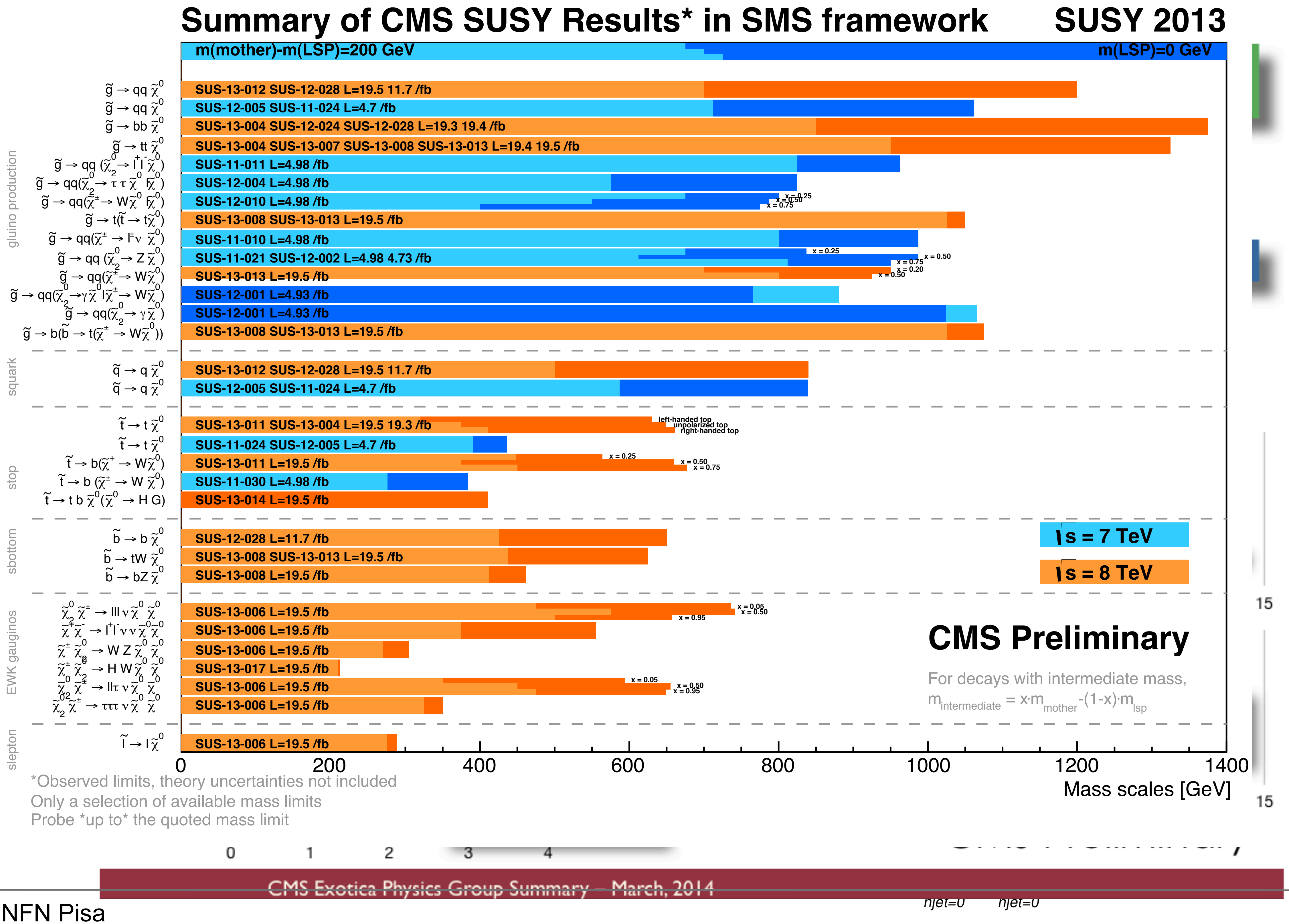
njet=0 njet=0

F. Palla INFN Pisa

No evidence of New Physics (yet!)

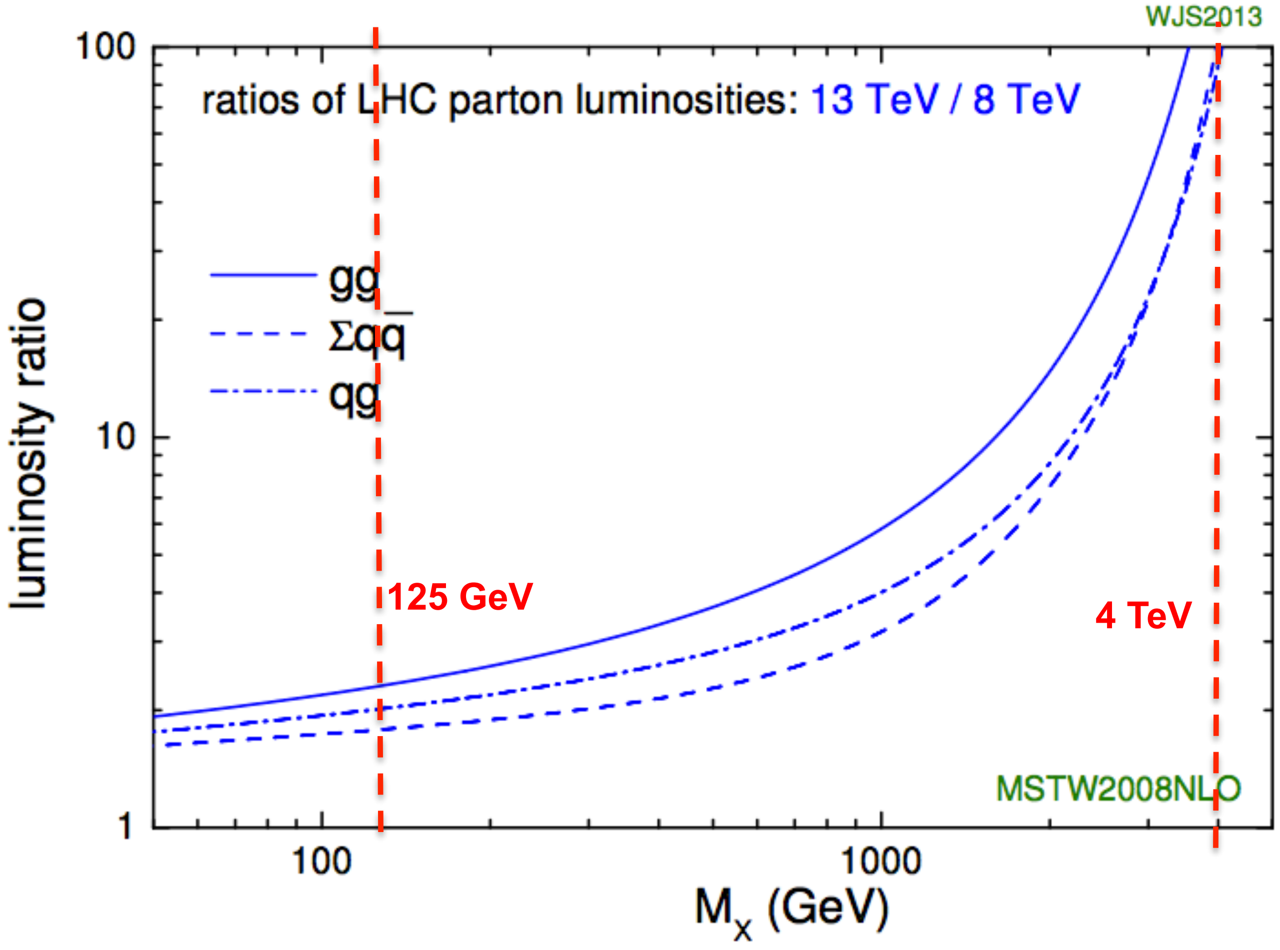


No evidence of New Physics (yet!)



What could be done next?

- Run 2 (2015-18):
  - ~100 fb^{-1} at 13 TeV, 10^{34} cm^{-2} s^{-1}
    - Gain depends on coupling (qq, gg, qg)
      - 1 year worth ~250 years of 8 TeV data for e.g. 4 TeV Z’ or 2 TeV squarks big gains in sensitivity



The HL-LHC scenario

13 TeV, 10^{34} /cm²/s

~100 fb^{-1}

13-14 TeV, ~2x10^{34} /cm²/s

LHCb Upgrade
ATLAS/CMS Upgrades phase 1

~300 fb^{-1}

14 TeV, ~5x10^{34} /cm²/s

HL-LHC
14 TeV
>250 fb-1/year

~3000 fb^{-1}

ATLAS/CMS
Upgrades phase 2

At HL-LHC ~140 events/bx spread over ±15 cm (3 σ)

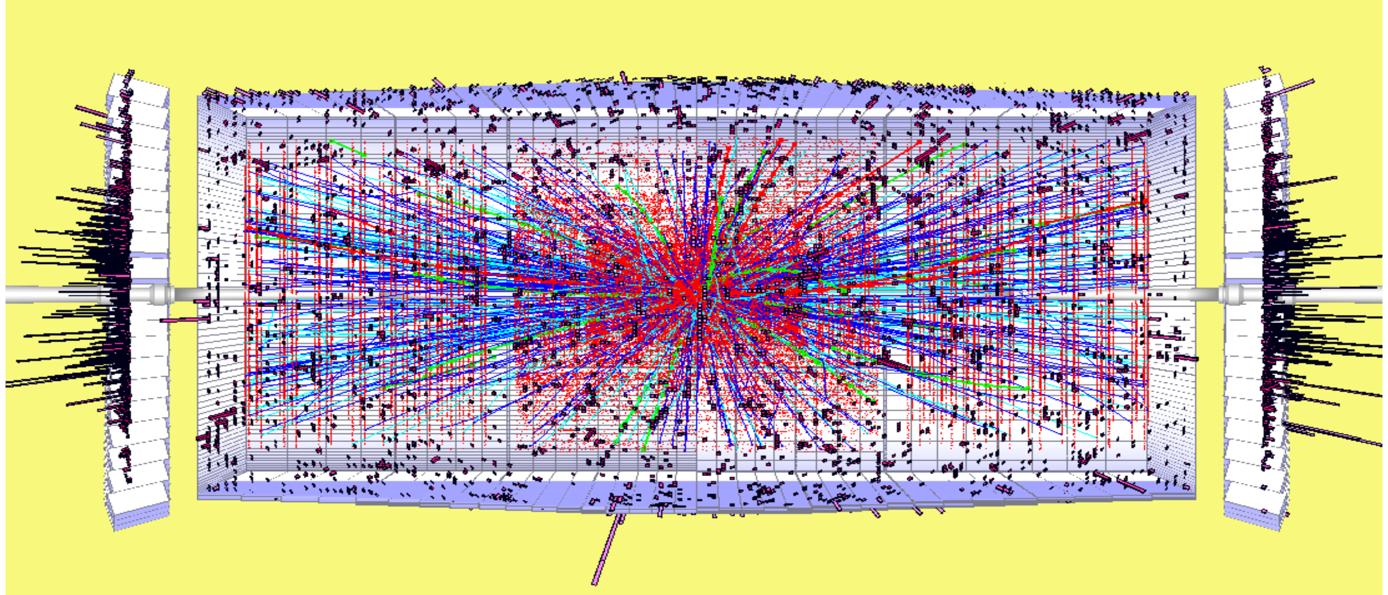
~3000 fb^{-1} expected in 10 years running

Excellent opportunity to search for (rare) and new phenomena

Need anyway still to trigger on “SM” objects (leptons, b, jets, MET)

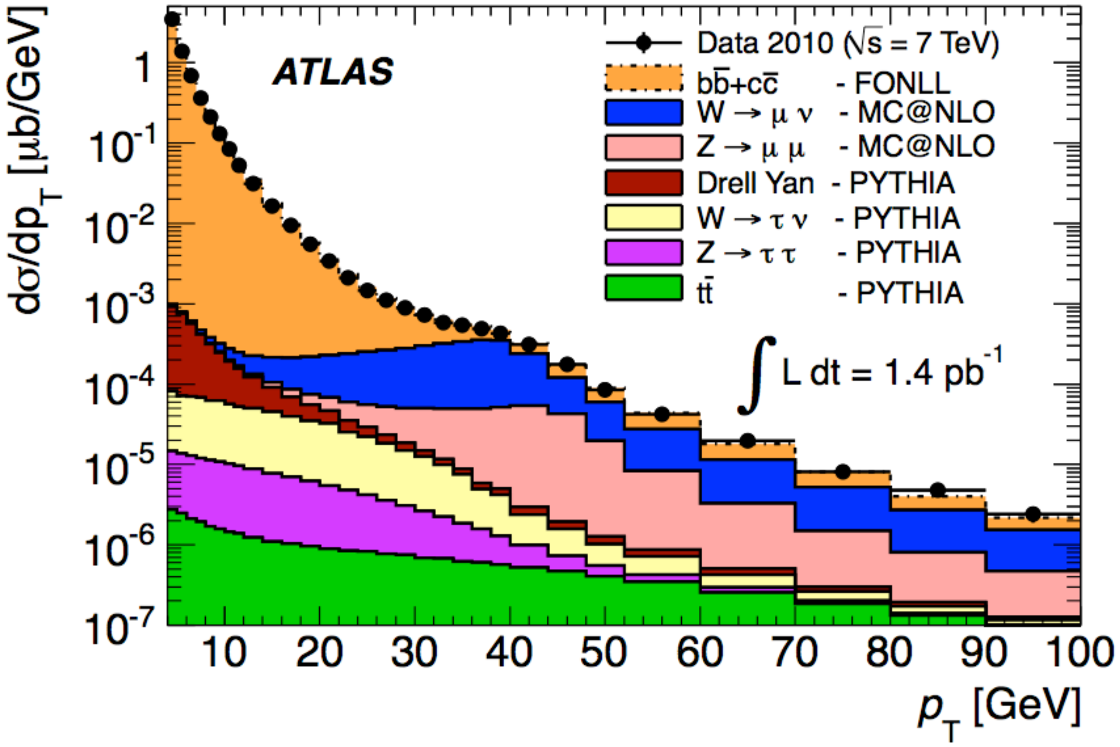
Implications for the Trigger

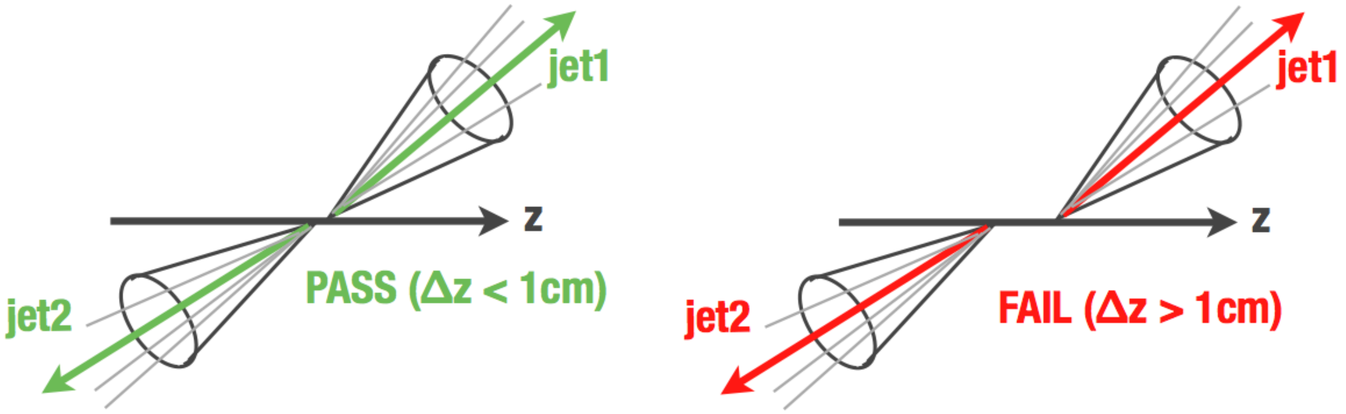
- At 5×10^{34} cm^{-2} s^{-1} up to ~140 interactions per bunch crossing
  - About 6k primary tracks per bunch crossing in the Tracker volume $|\eta|<2.5$ …
    - …plus any other coming from γ conversions and nuclear interactions
  - ~ one order of magnitude larger wrt LHC
  - Severe Triggering conditions
    - Too many primary vertices, need to have smarter triggers combining information from several subdetectors
    - Need to maintain low thresholds for basic objects, even with an increase in the L1-Accept bandwidth (currently at 100 kHz)
  - Both ATLAS and CMS will replace their “inner trackers” to cope with the nasty environmental conditions
    - The usage of the Tracker would help to disentangle among those 140 pileup events



Why a Track Trigger at L1

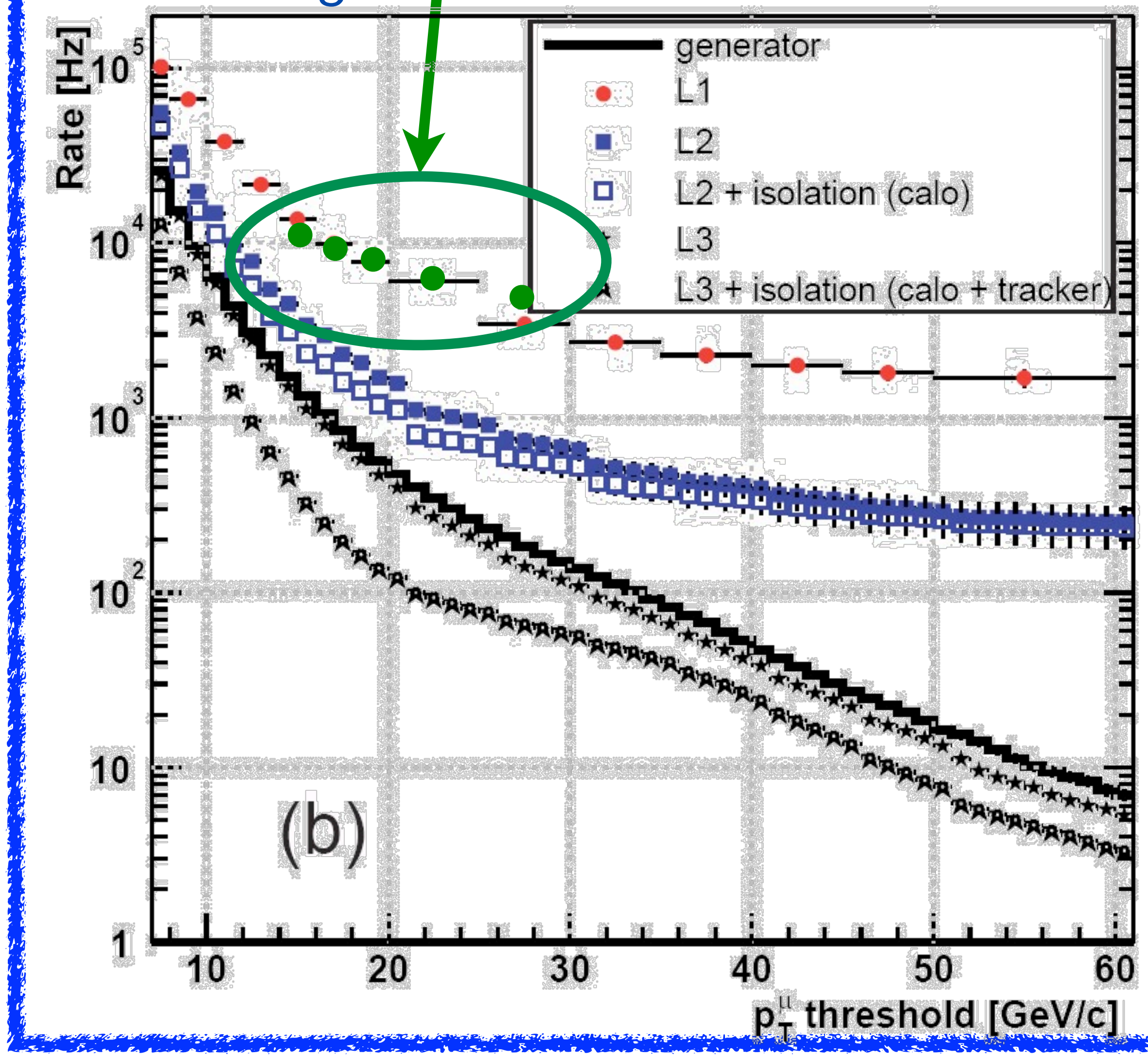
- HL-LHC physics goals require excellent Trigger selectivity on basic objects (leptons, jets, taus, b-jets, MET)
  - This might be jeopardized by the increased level of pileup events (140 on average)
    - Huge rate of *μ* from heavy flavors ⇨ use better p_T resolution from tracker
    - Prompt electrons 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





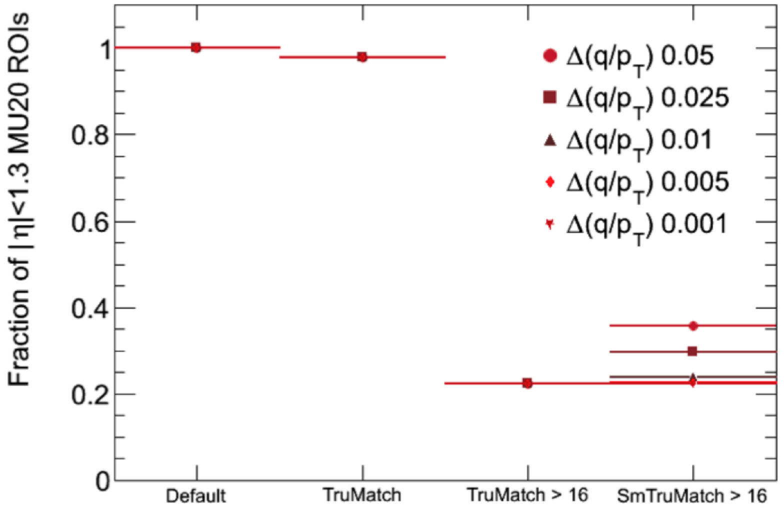
Examples from ATLAS and CMS

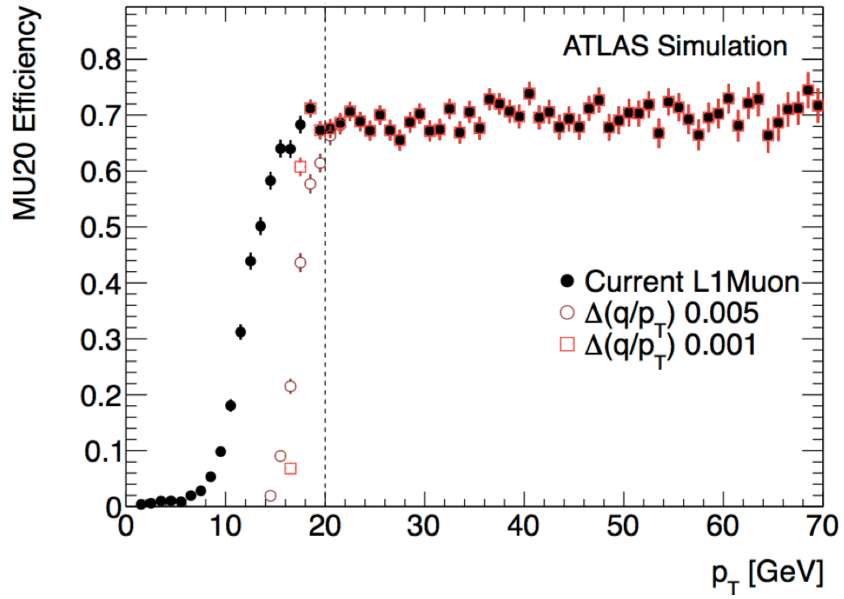
- CMS simulation for L=10^{34} cm^{-2} s^{-1}
- Add measured data rates at 8 TeV, extrapolated to 10^{34} cm^{-2} s^{-1}
  - No p_T threshold may reduce the rate enough!



ATLAS simulation

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





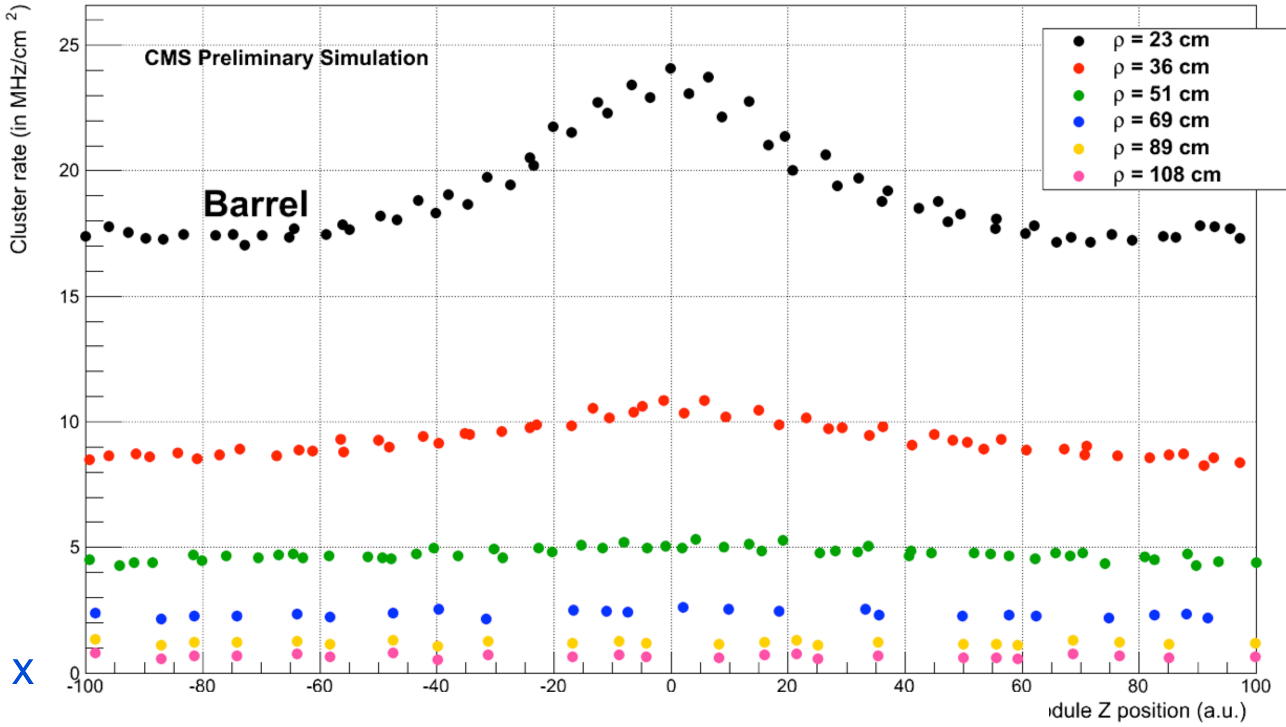
The challenge and the way out

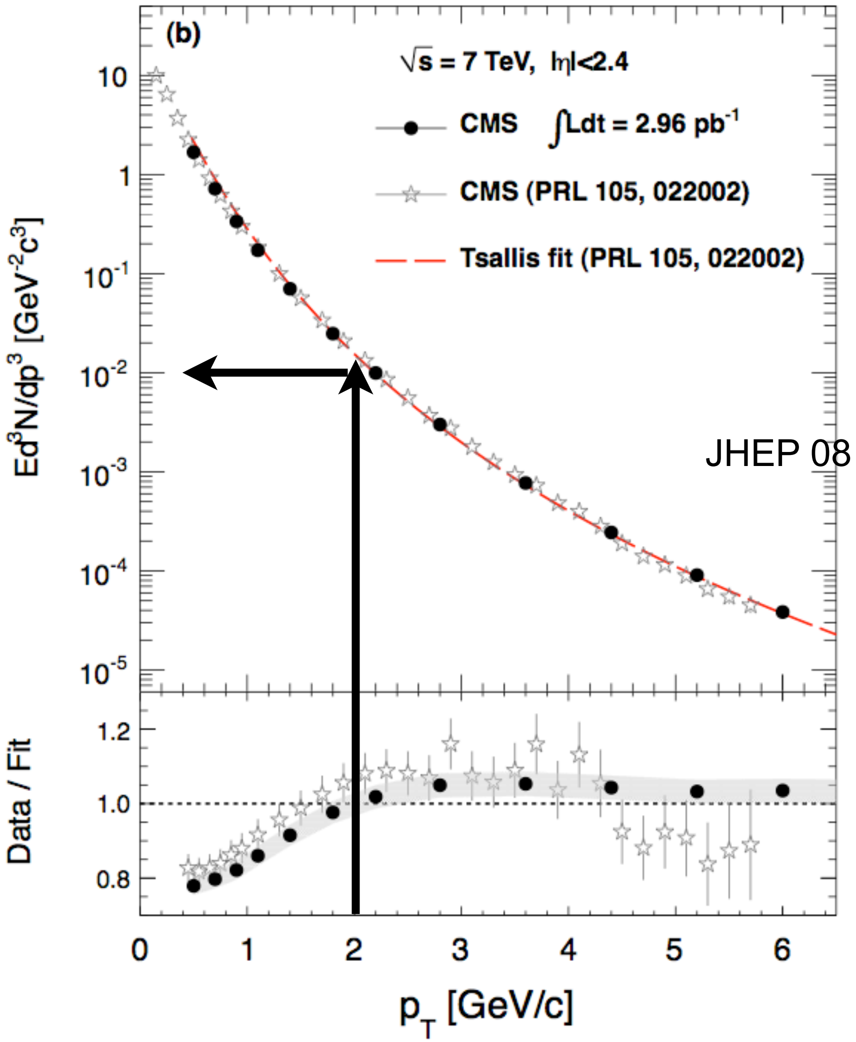
- Take data off the tracker
  - ~ 4k primary tracks within $|\eta|<2.5$
    - Large data rates (up to 25 MHz/cm^2)
      - huge contribution from nuclear interactions and photon conversions
    - ~1.3 events/mm × Gauss(σ=4 cm)
    - Short L1A trigger latencies (10-20 μs)
    - Cannot read all (~60 M 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
- Once data are off-detector, find tracks and
  - formidable pattern recognition problem
    - need latencies of ~5 μs

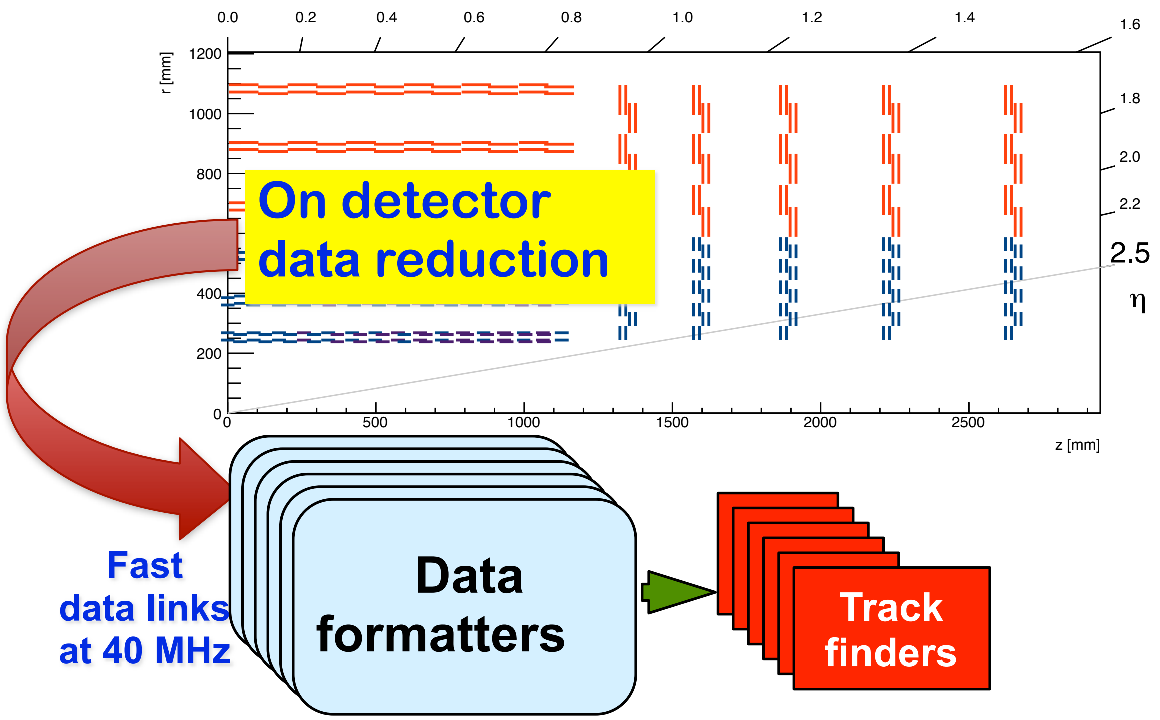


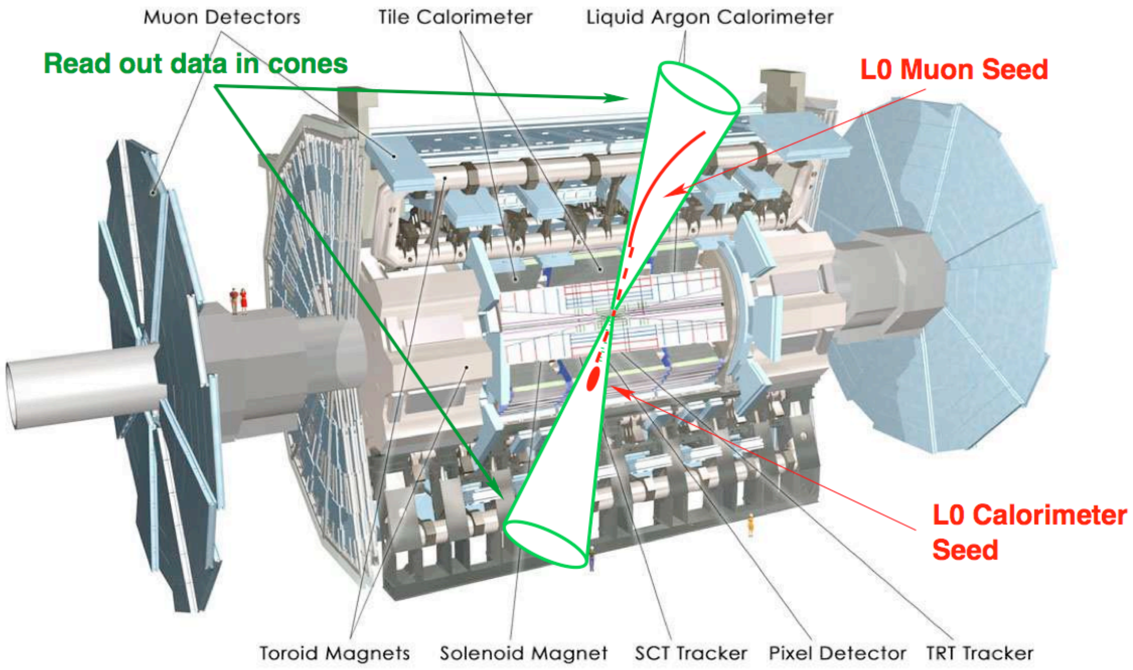


JHEP 08 (2011) 086

Trigger architectures

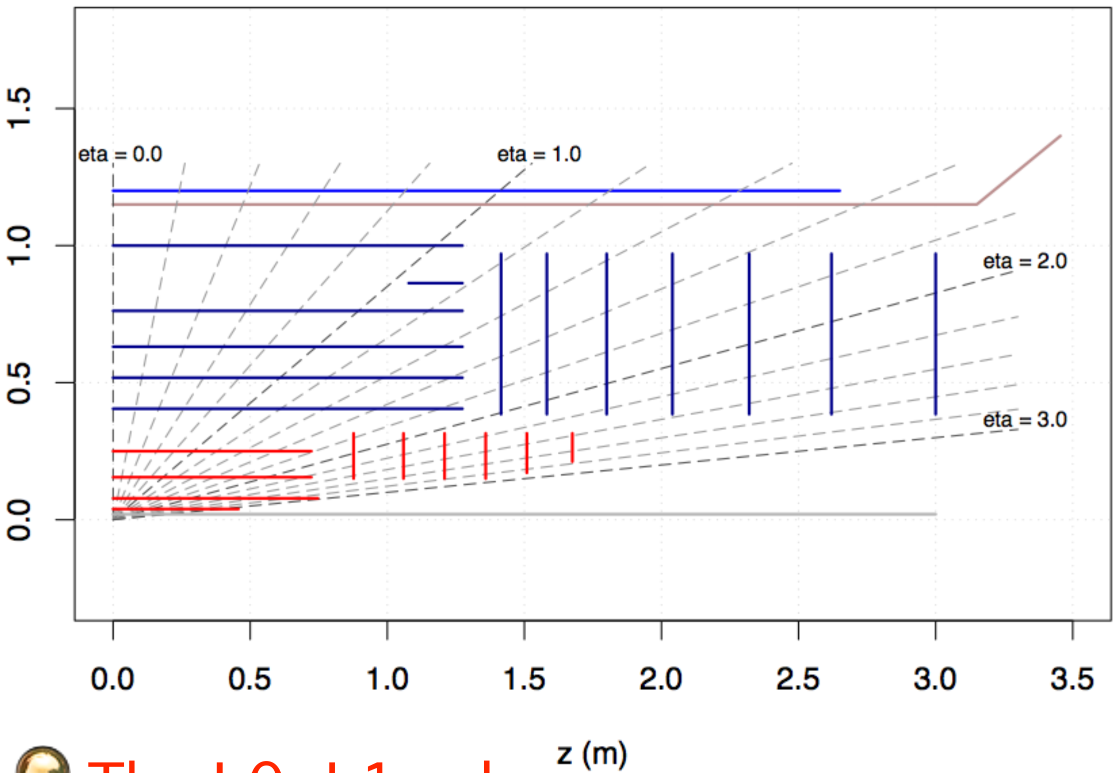
- PUSH path (CMS)
  - Reduced Tracker information readout at 40 MHz and then combined with calorimeter & muon at L1
    - Trigger objects made from tracking, calorimeter & muon inside a Global Trigger module
- PULL path (ATLAS)
  - Use calorimeter & muon detectors to produce a “Level-0” to request tracking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

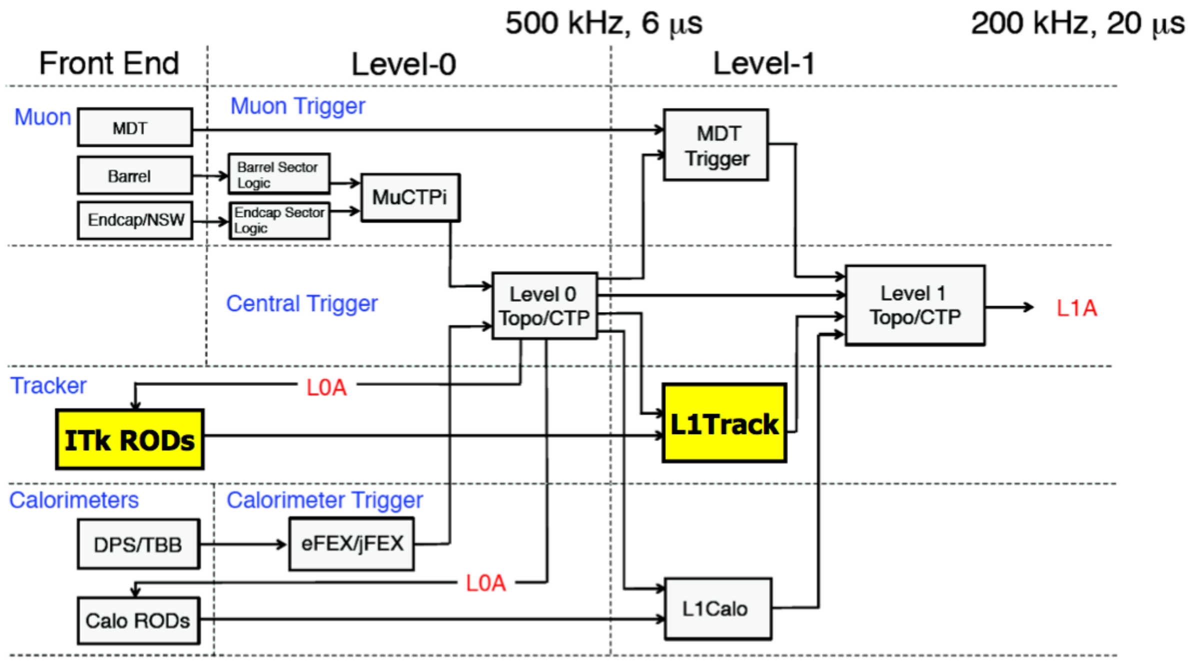




Data reduction

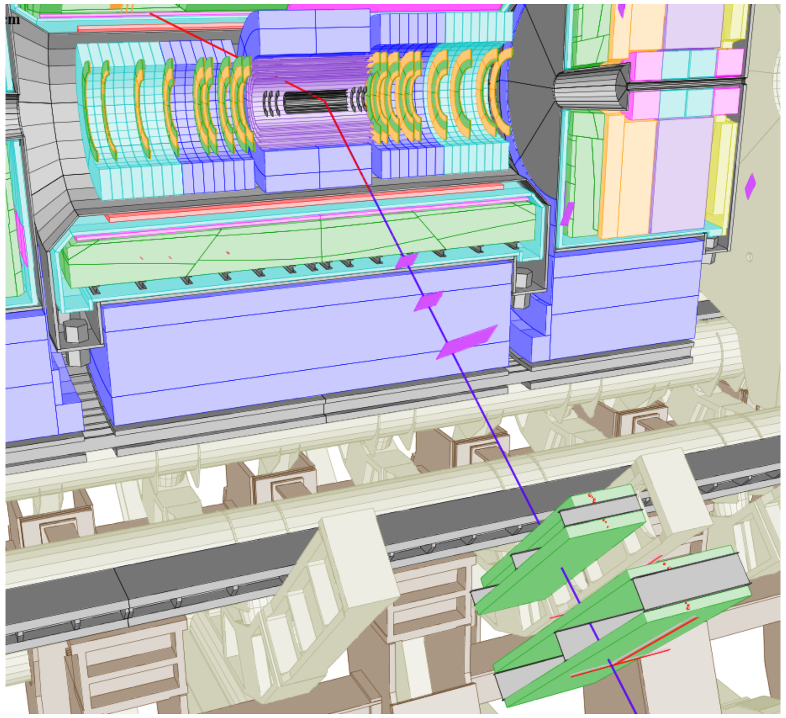
Track Trigger with pull architecture (ATLAS)





The L0+L1 scheme

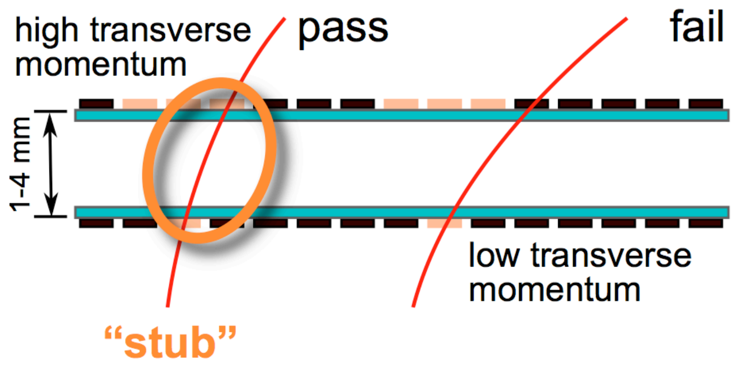
- Level-0:
  - Coarse calo and muon data
  - Rate 40 MHz →500 kHz
  - Latency < 6.4 μs
  - Defines Region of Interest (ROIs) for L1
- Level-1:
  - Tracker data only from ROIs
  - Refined information from calo and muons
  - Rate 500 kHz →200 kHz
  - Latency < 20 μs

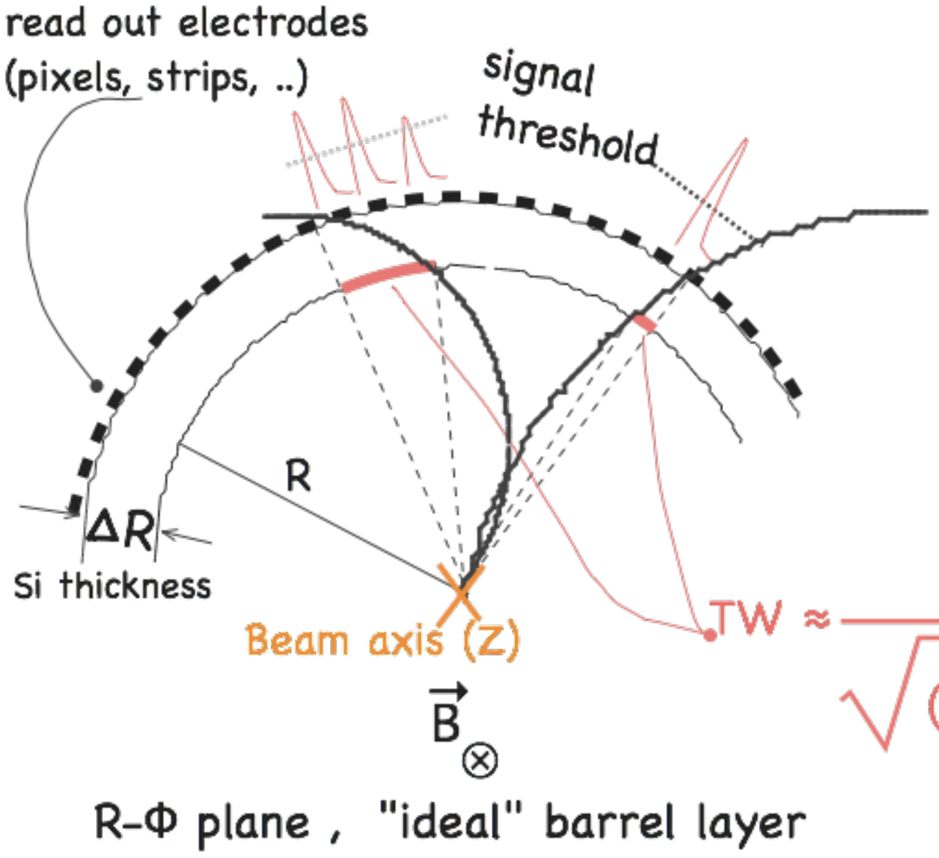


- Need to provide data from the FE in less than ~5μs.

Select only hits from “high-p_T” tracks

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





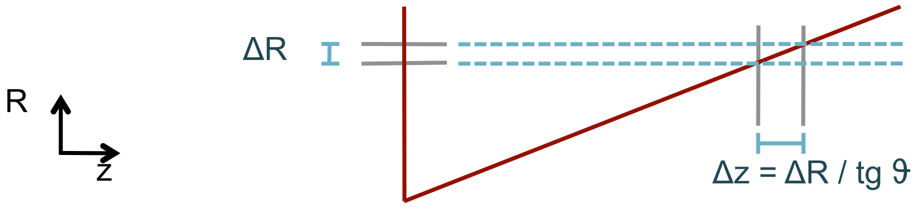
F. Palla, G. Parrini, PoS VERTEX2007 (2007) 034, http://pos.sissa.it/archive/conferences/057/034/Vertex%202007_034.pdf

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

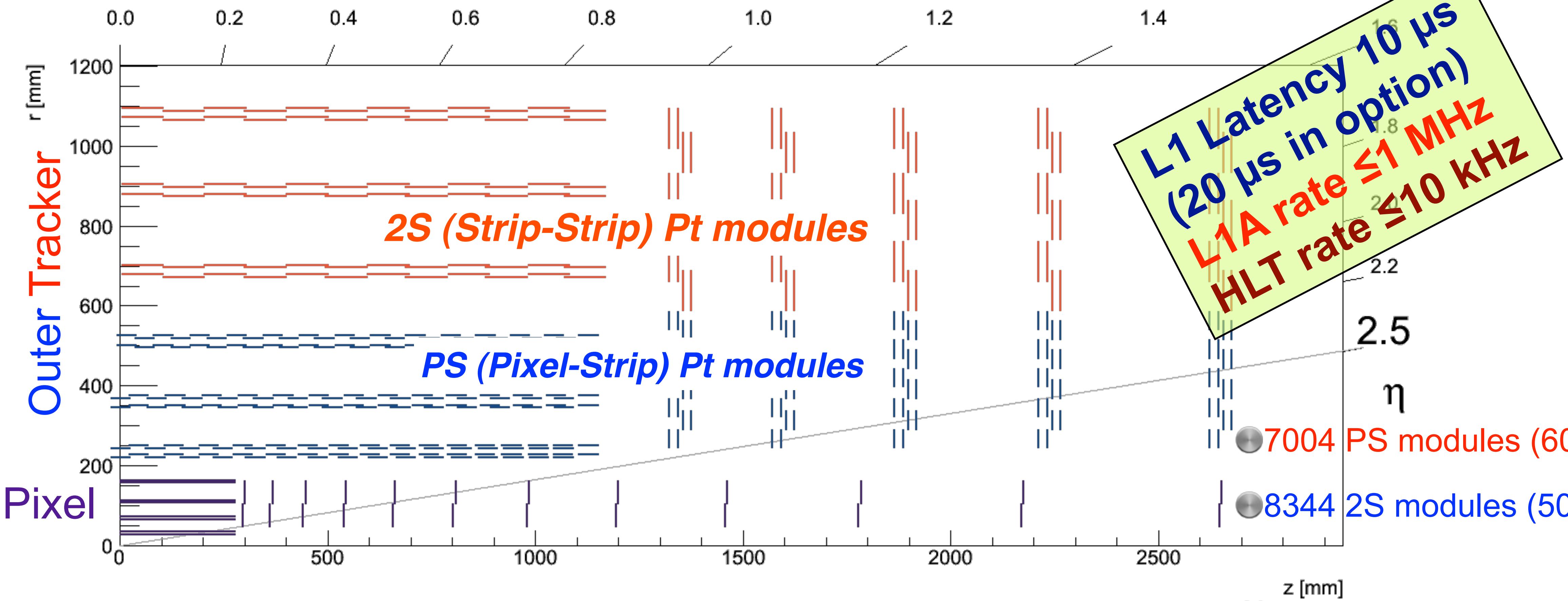
$$TW \approx \frac{\Delta R}{\sqrt{\left(\frac{pT}{pT_{min}}\right)^2 - 1}} \approx \Delta R \frac{pT_{min}}{pT} = 0.15\,B\,\Delta R\,\frac{R}{pT}$$

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

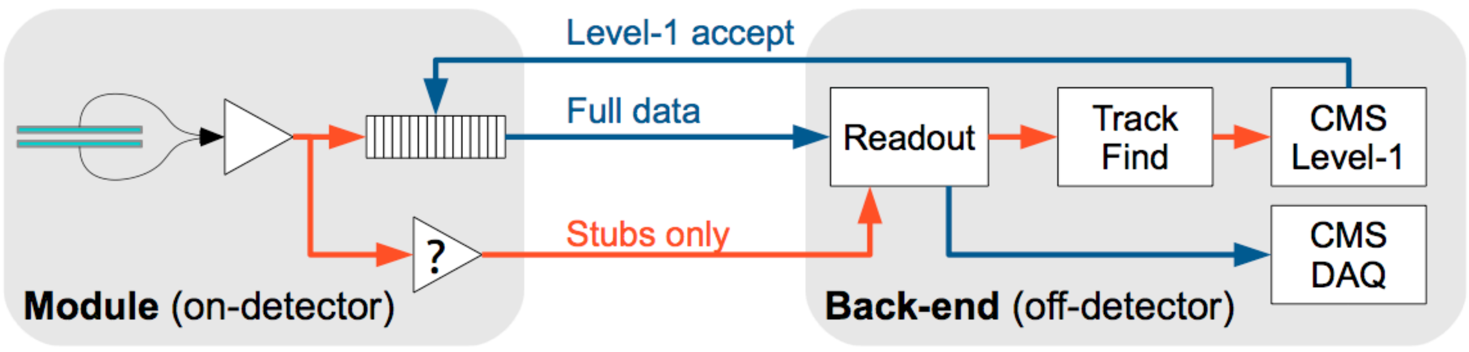


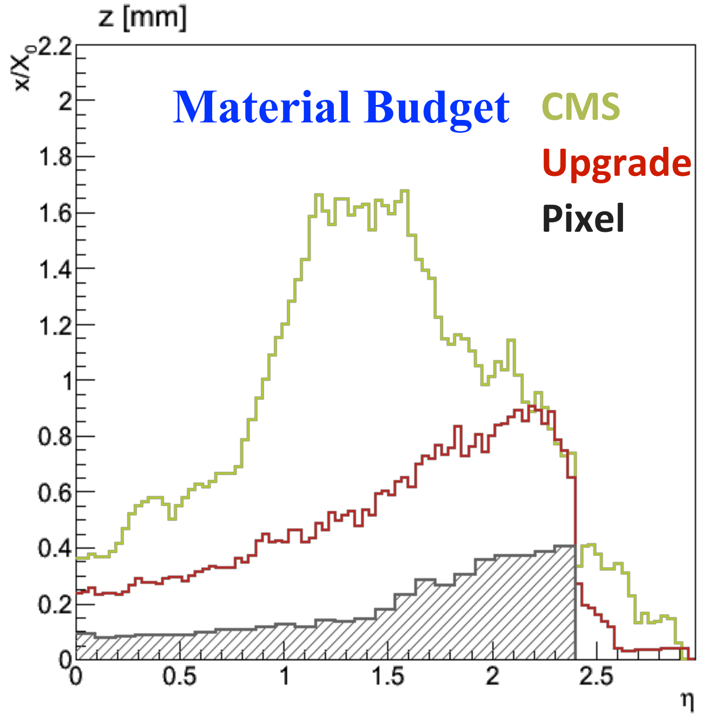
L1 Latency 10 μs
(20 μs in option)
L1A rate ≤1 MHz
HLT rate ≤10 kHz

- 7004 PS modules (60% in the barrel)
- 8344 2S modules (50% in the barrel)

Better p_T resolution and lighter than current tracker

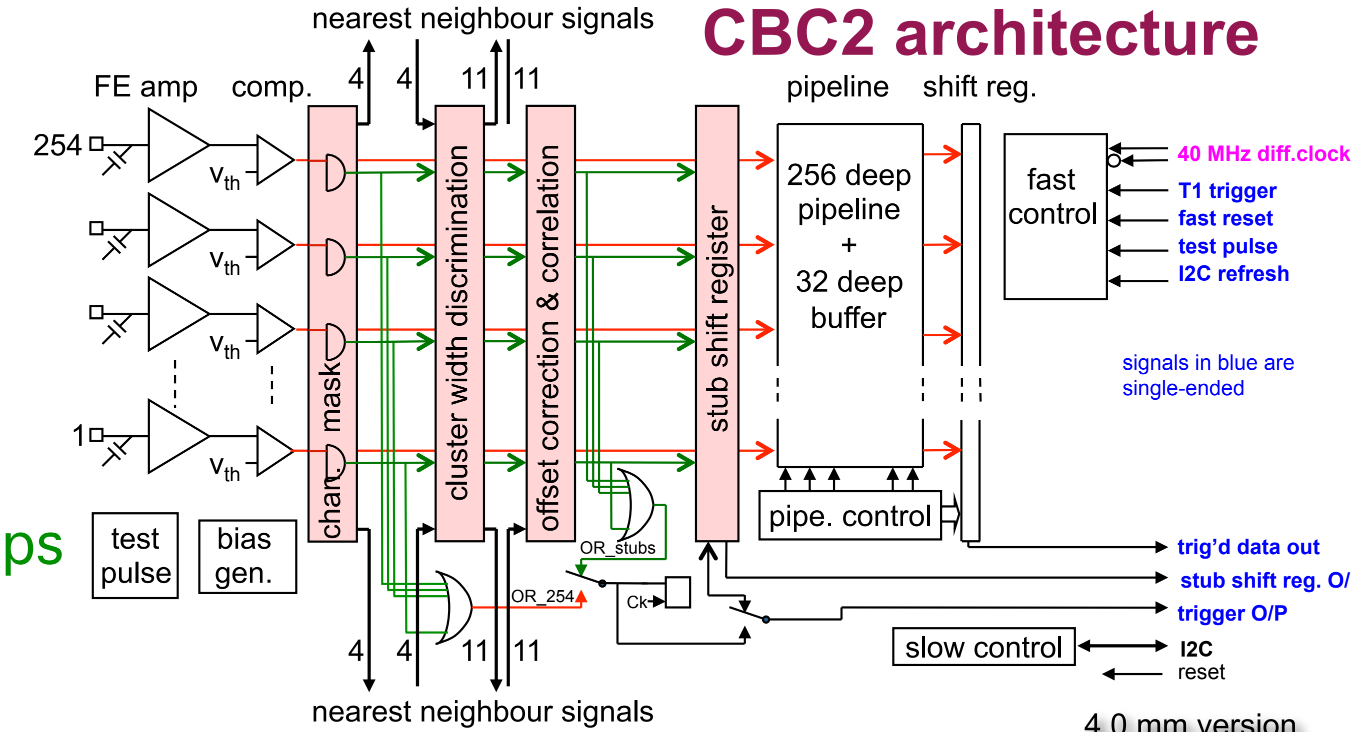
Readout and Trigger schematics

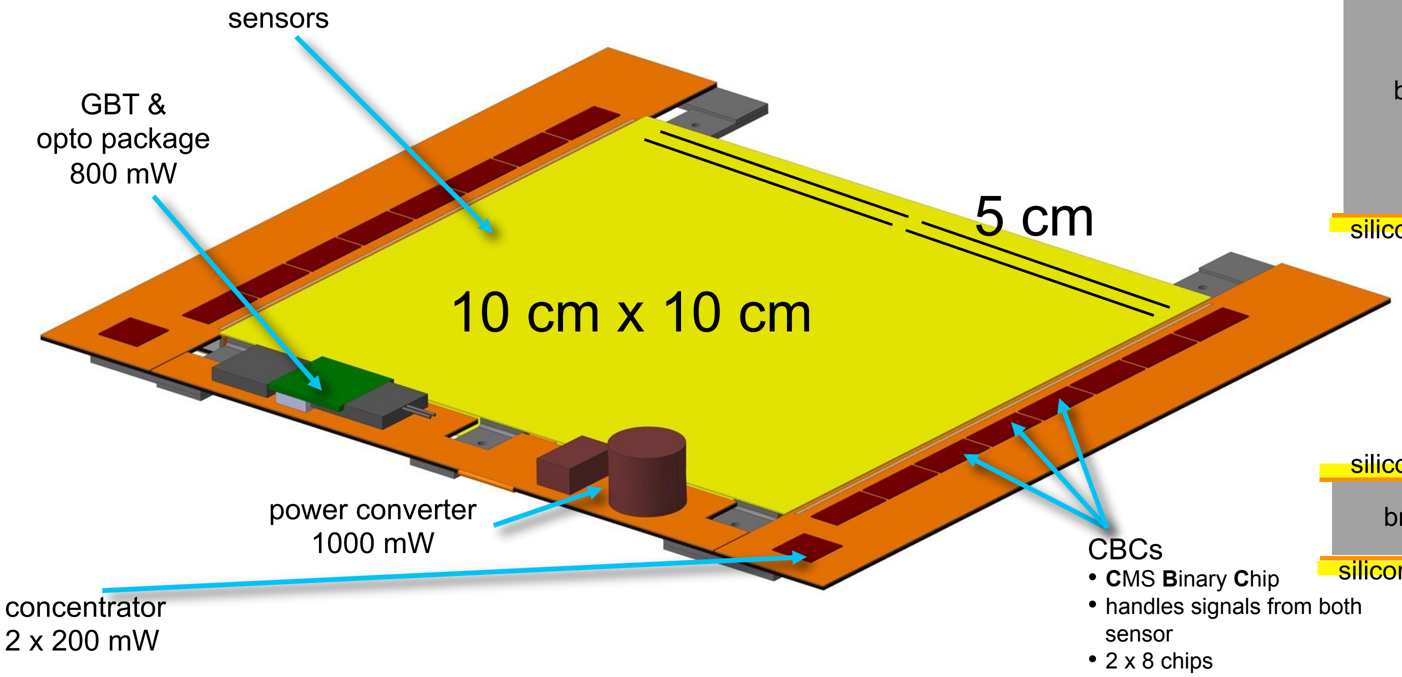




CMS 2S modules

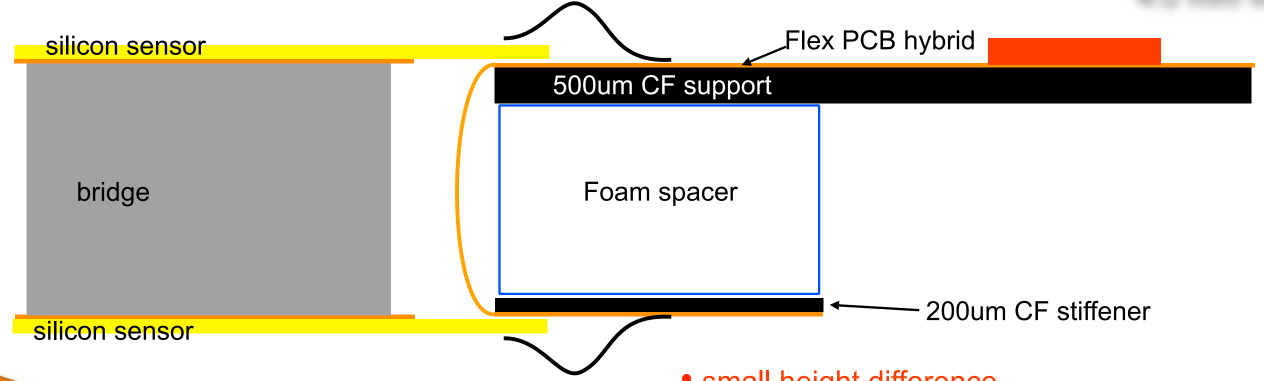
- 2S(trip) sensors modules
  - 100 μm x 5 cm long strips on both sensors
  - readout by 8 CBC on either sides
    - First discriminates signals by rejecting large clusters; then form a coincidence between the two sensor planes
    - Concentrator chip sends data from 8 chips to GBT


CBC2 architecture



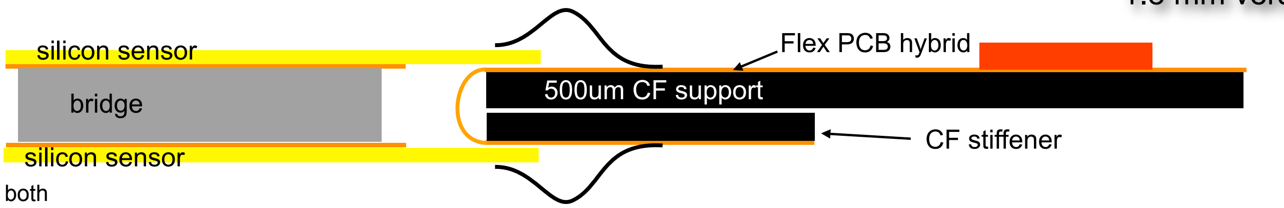
- **C**MS **B**inary **C**hip
- handles signals from both sensor
- 2 x 8 chips
- 1200 mW

4.0 mm version

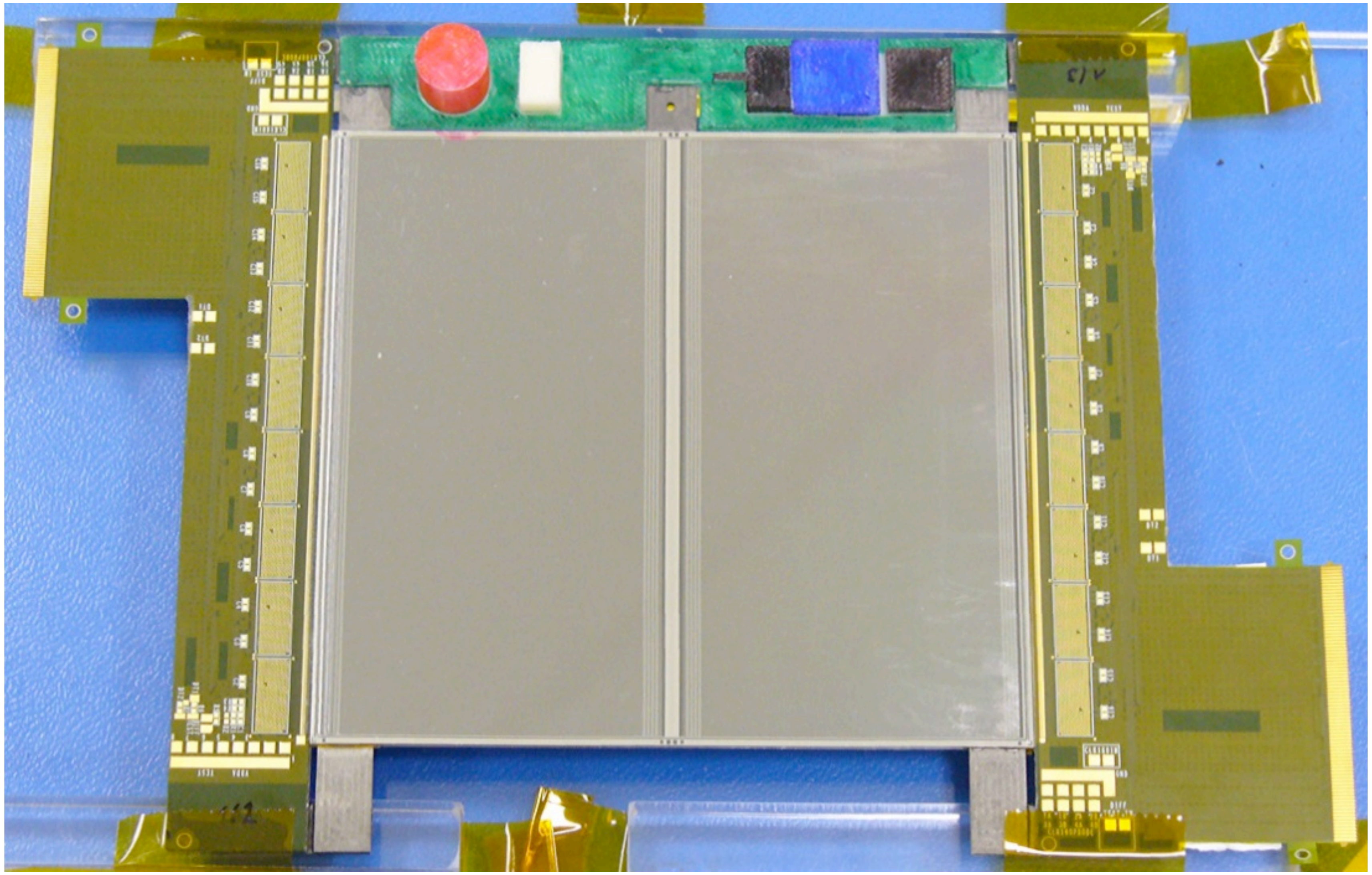


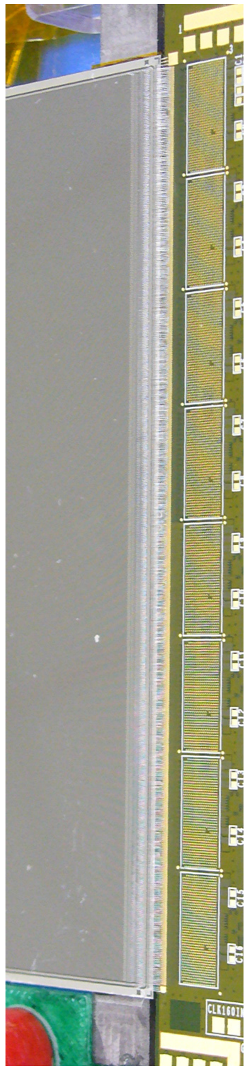
- small height difference
- short wires (<3 mm)
- encapsulation of bond wires possible

1.8 mm version



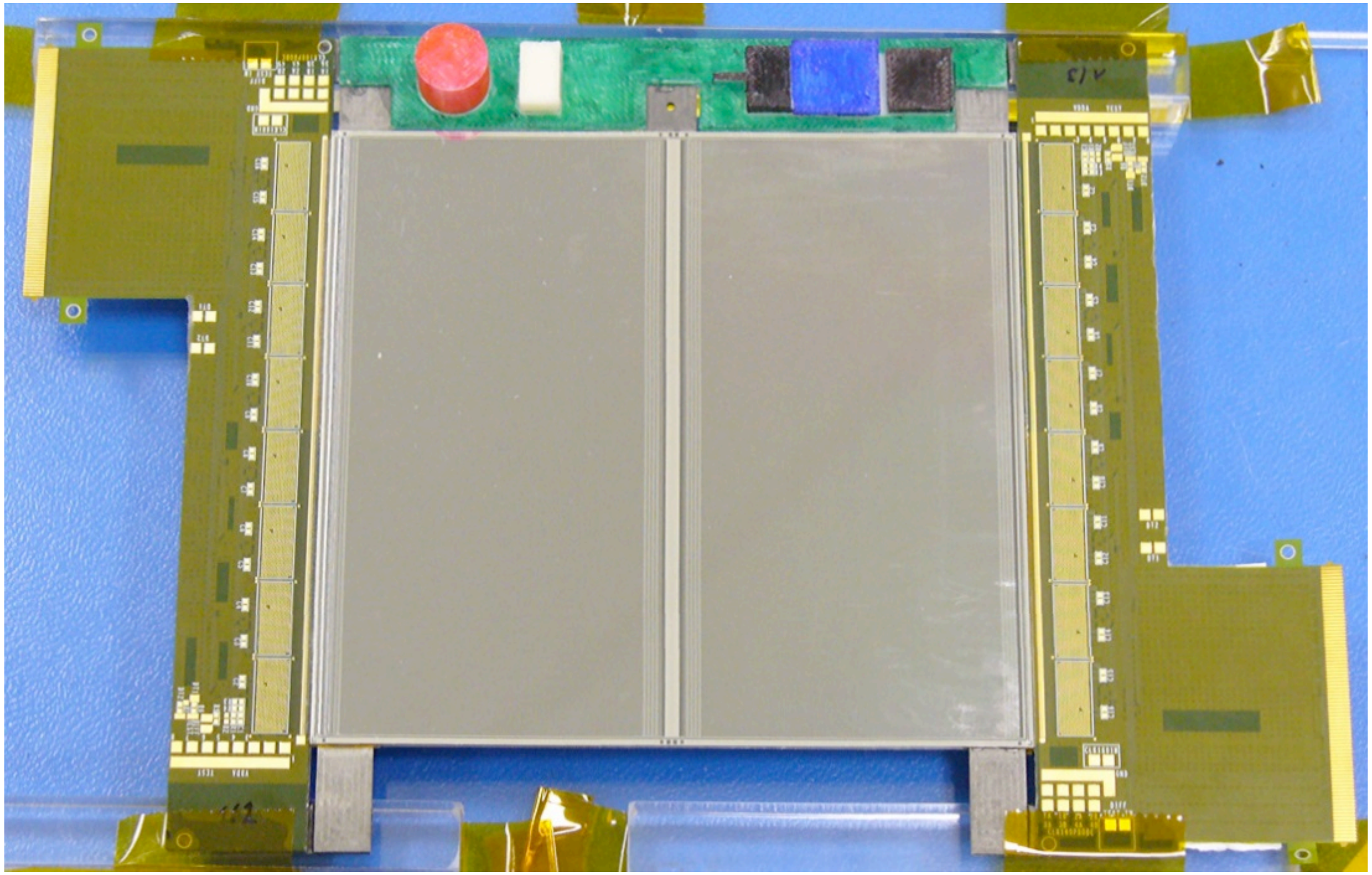
2S module prototype

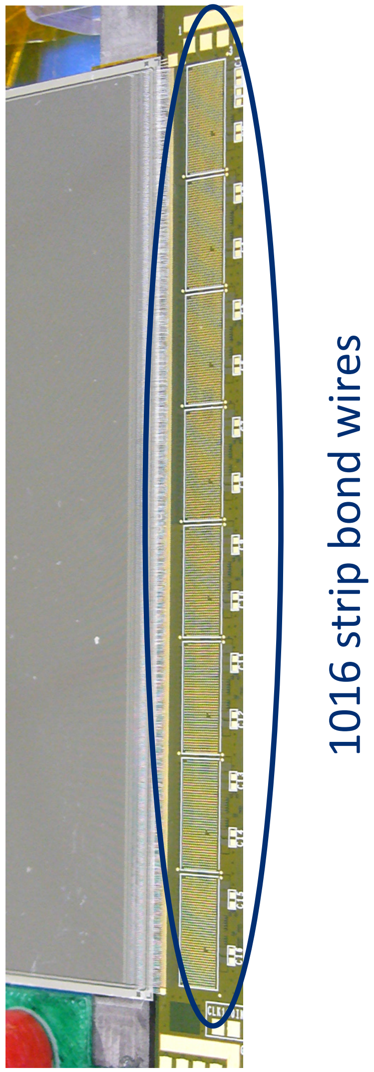




F. Palla INFN Pisa

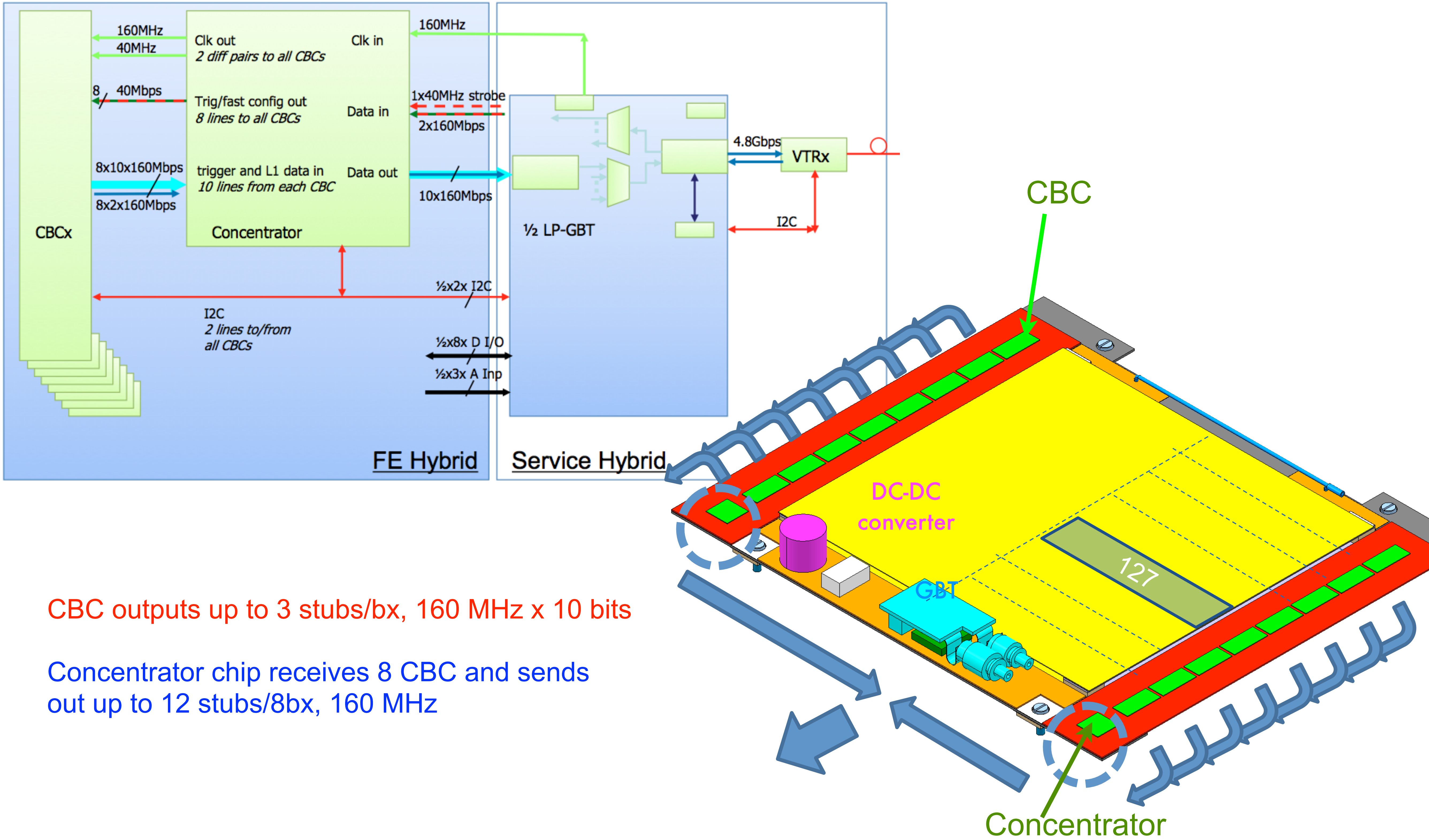
2S module prototype




1016 strip bond wires

F. Palla INFN Pisa

Electronics



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

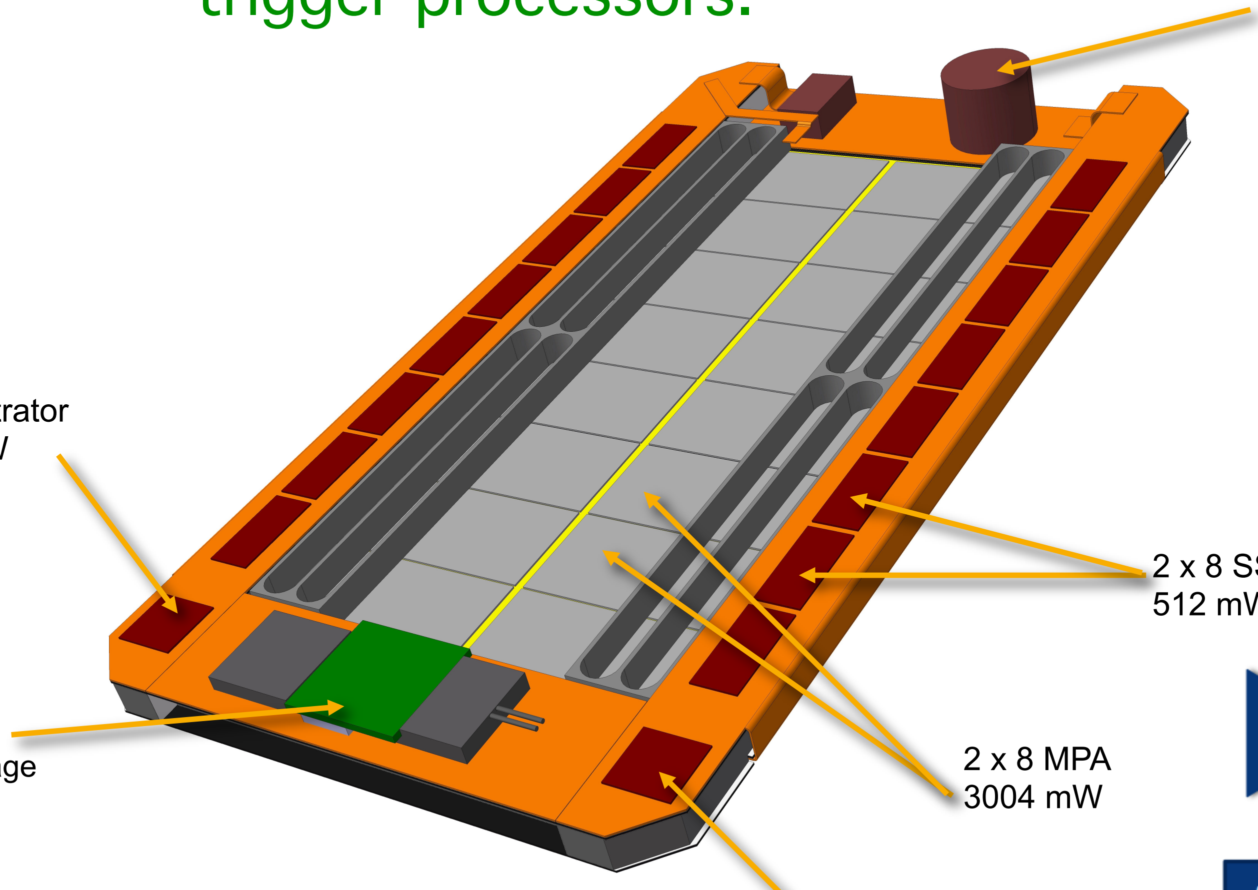
Concentrator chip receives 8 CBC and sends out up to 12 stubs/8bx, 160 MHz

F. Palla INFN Pisa

CMS PS modules

P(ixel)S(strip) 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.



power converter
2 W

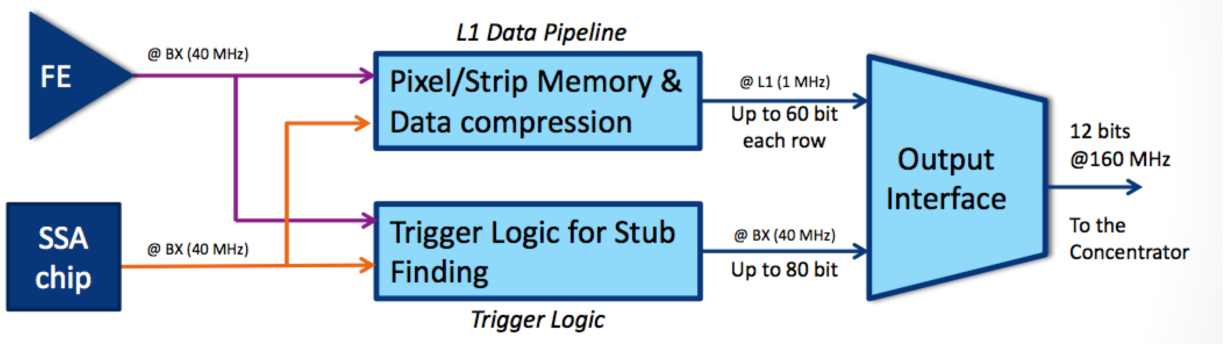
concentrator
200 mW

2 x 8 SSA chips
512 mW

GBT &
opto package
800 mW

2 x 8 MPA
3004 mW

concentrator
200 mW



L1 Data Pipeline

FE

@ BX (40 MHz)

Pixel/Strip Memory & Data compression

@ L1 (1 MHz)
Up to 60 bit each row

Output Interface

12 bits
@160 MHz

To the Concentrator

SSA chip

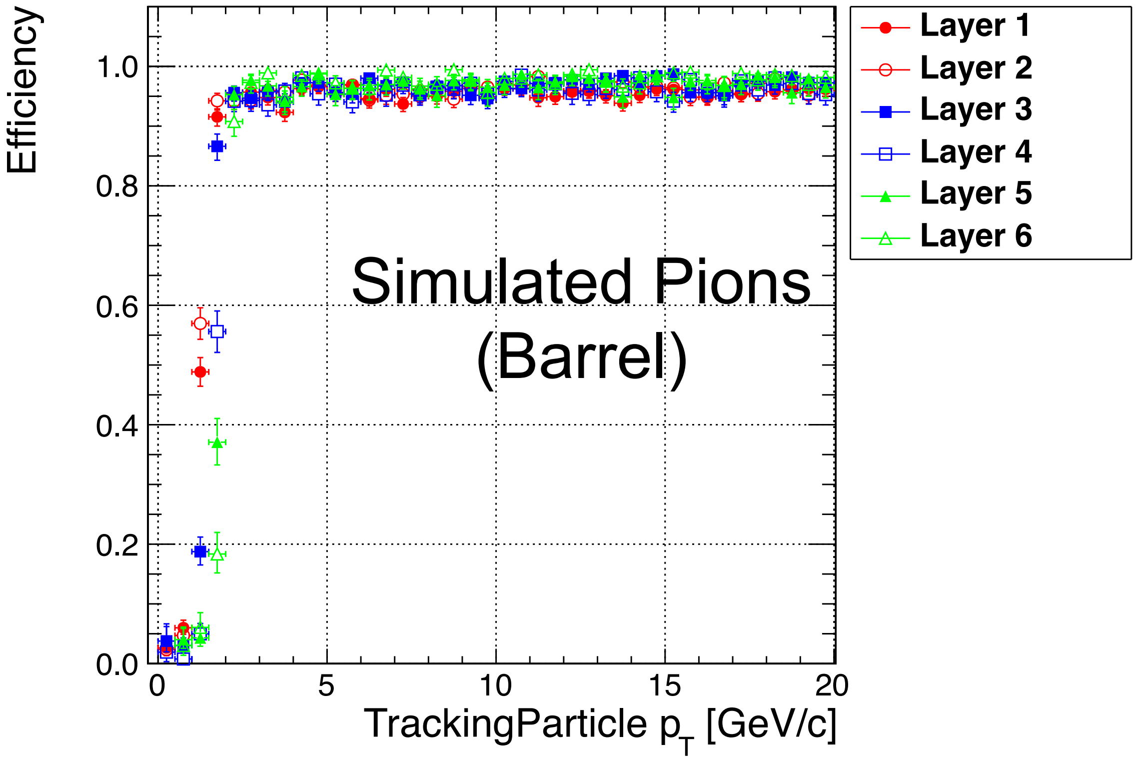
@ BX (40 MHz)

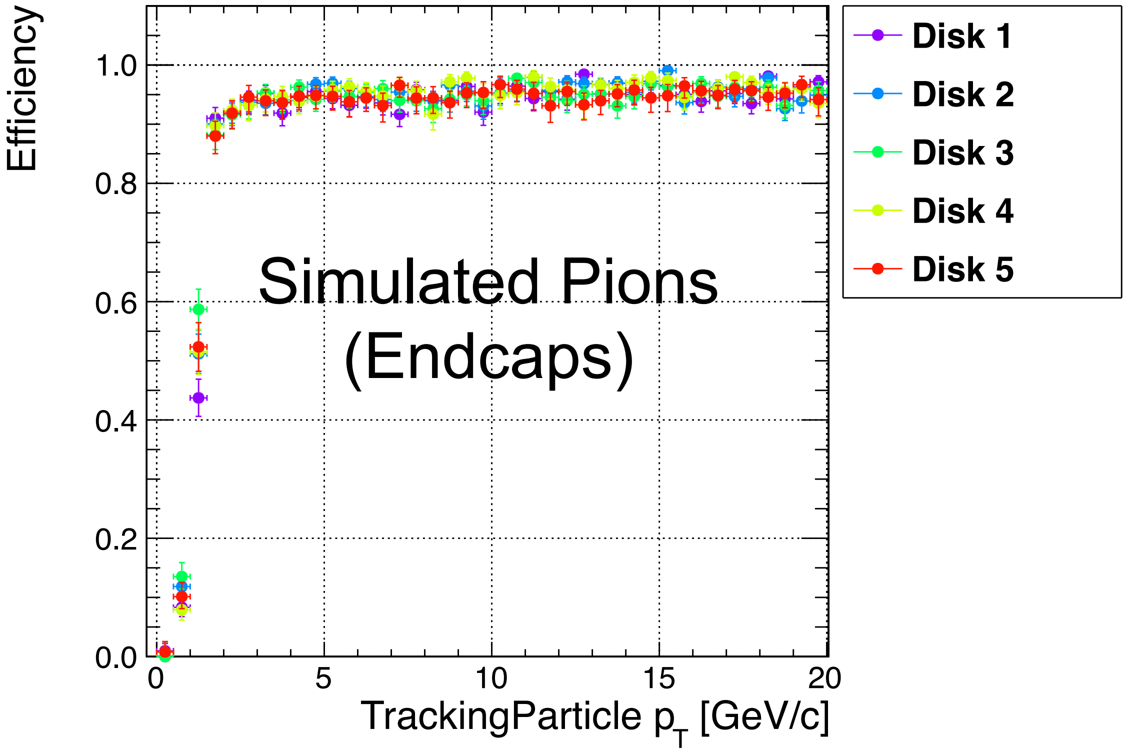
Trigger Logic for Stub Finding

@ BX (40 MHz)
Up to 80 bit

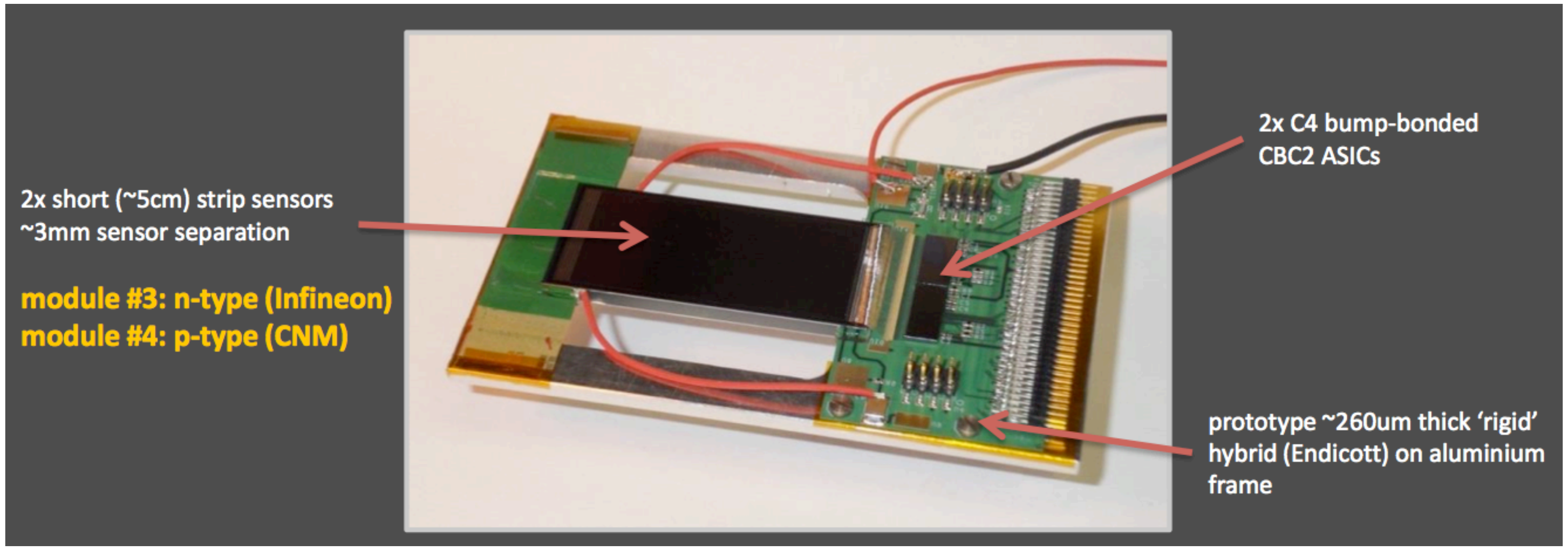
Trigger Logic

p_T modules performances





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

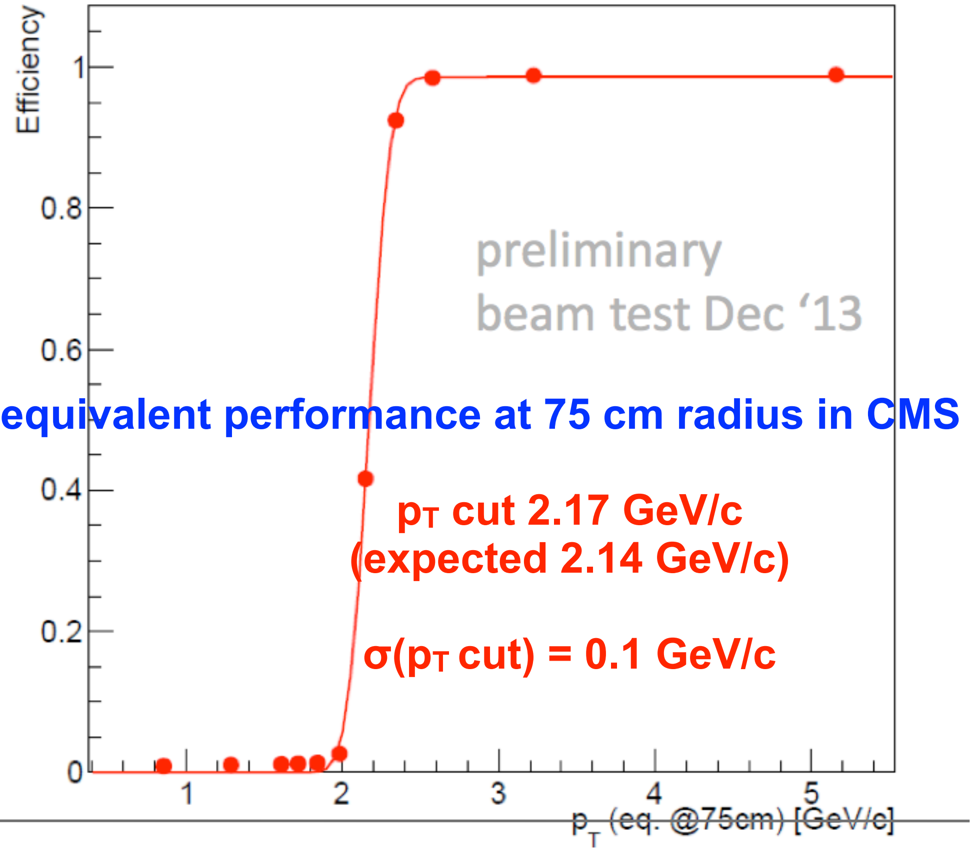


2x short (~5cm) strip sensors
~3mm sensor separation

module #3: n-type (Infineon)
module #4: p-type (CNM)

2x C4 bump-bonded
CBC2 ASICs

prototype ~260um thick 'rigid' hybrid (Endicott) on aluminium frame



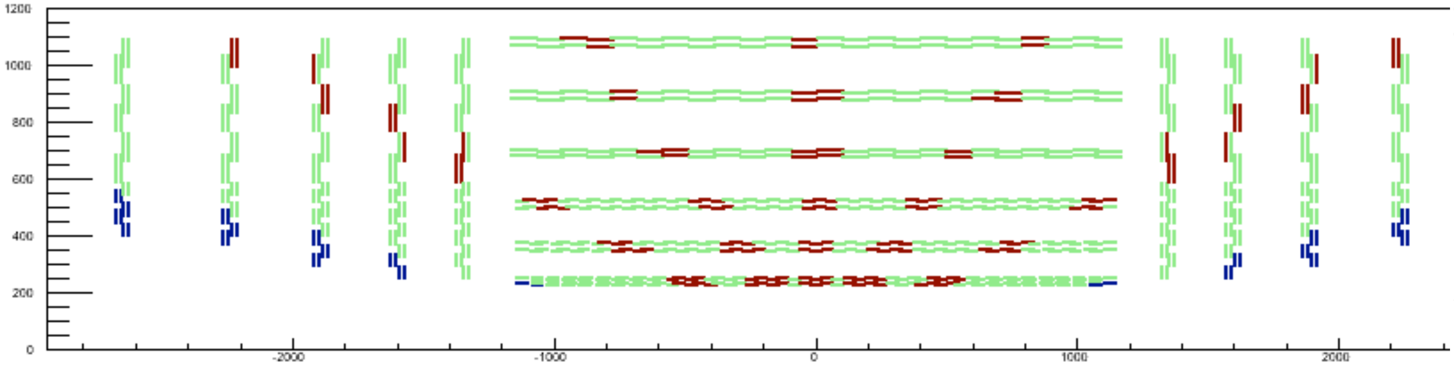
equivalent performance at 75 cm radius in CMS

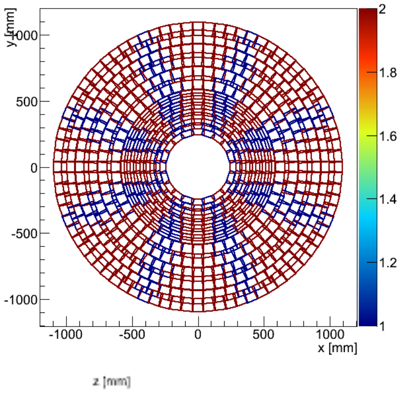
p_T cut 2.17 GeV/c
(expected 2.14 GeV/c)

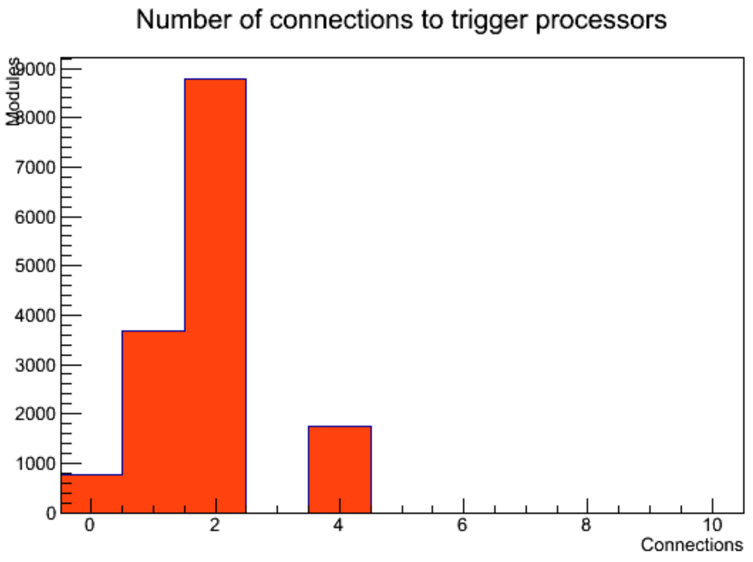
$\sigma(p_T$ cut$)$ = 0.1 GeV/c

Data organization and dispatch

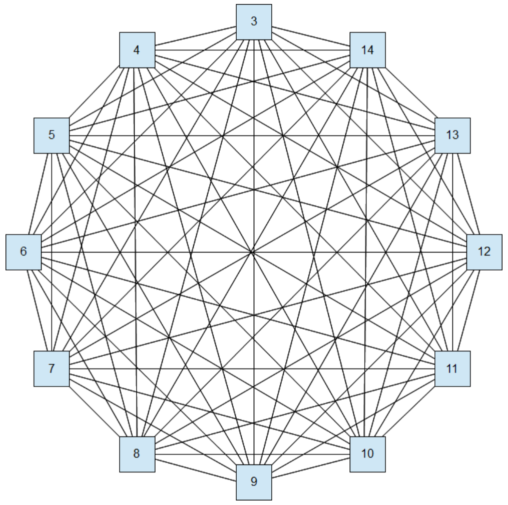
- Subdivide tracker into trigger towers
  - Example CMS: 8(r-ϕ)x6(r-z) trigger sectors (some 10% overlapping)
    - Each sector ~200 stubs on average; tails up to ~500 stubs/event in 140 evts pileup+ttbar (to be compared with ATLAS-Phase 1 ~2000)
      - About 600 Gb/s per one trigger tower





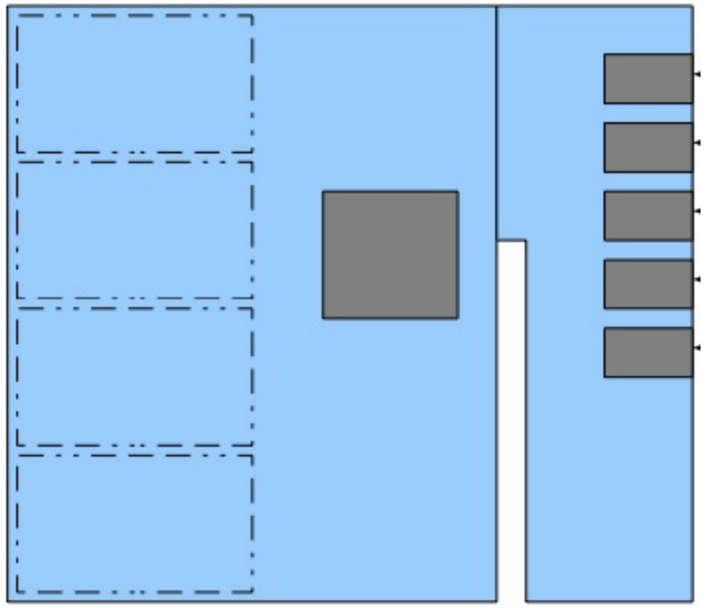


- Send data to Track-finding processors
  - Full mesh ATCA shelfs
    - Capable of "40G" full-mesh backplane on 14 slots = 7.2 Tb/s
    - Several options being investigated, all include time multiplexing data transfer from a set of receiving processors boards to pattern recognition and track finding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



CMS: toward a demonstrator

Input Data Board x 8

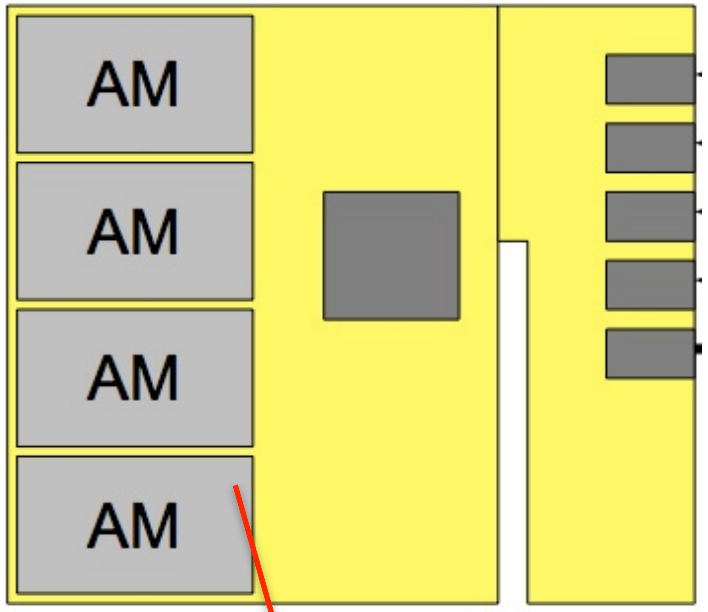


Up to 48 pairs of connection, Bidirectional.

Up to 48 inputs with 3.25Gbps each

With 40G full-mesh backplane

Pattern Recognition Board x 4



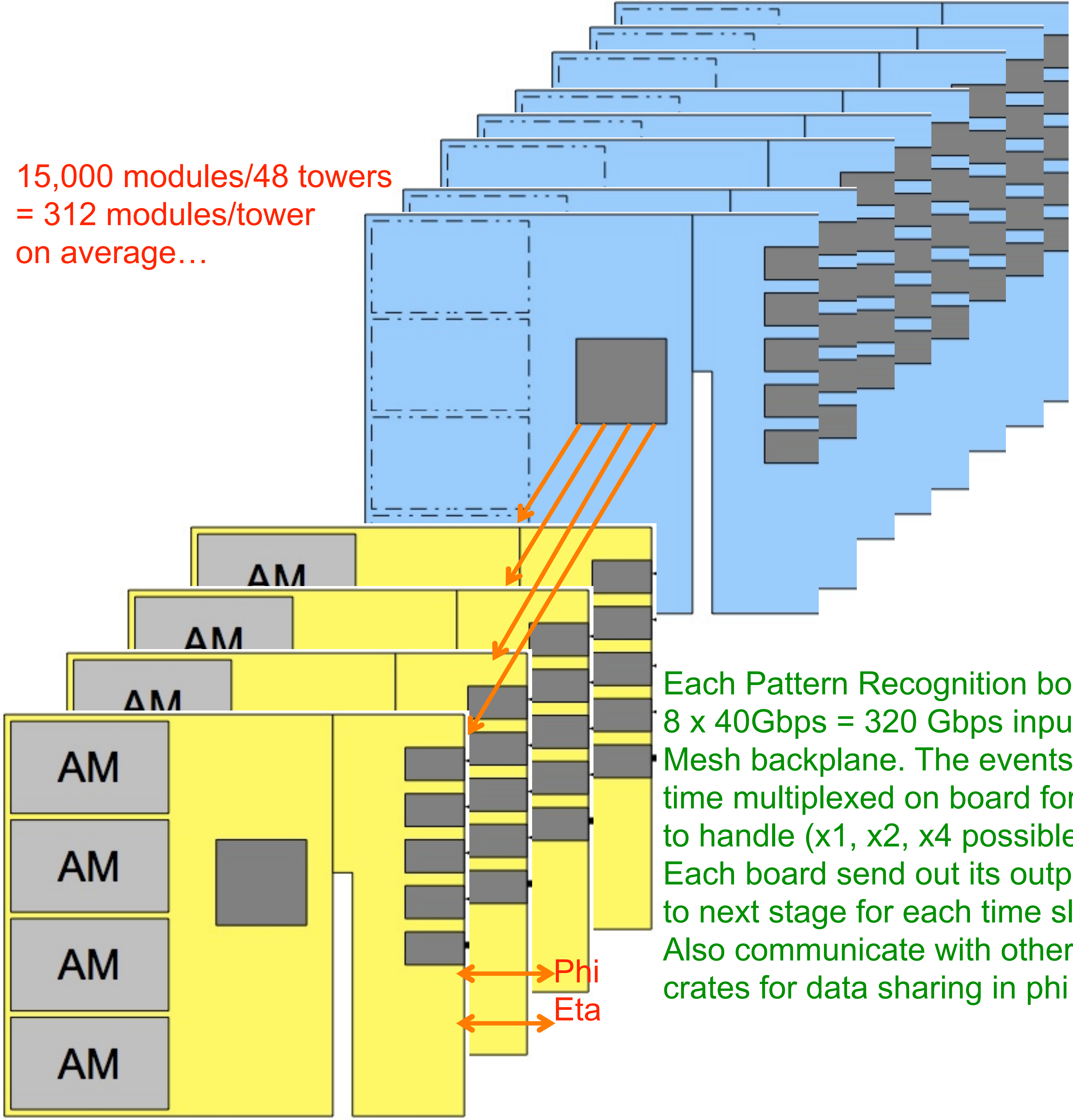
Up to 48 pairs of connection, Bidirectional.

10Gbps each

With 40G full-mesh backplane

Could also be pure FPGA based approach

One Trigger Tower



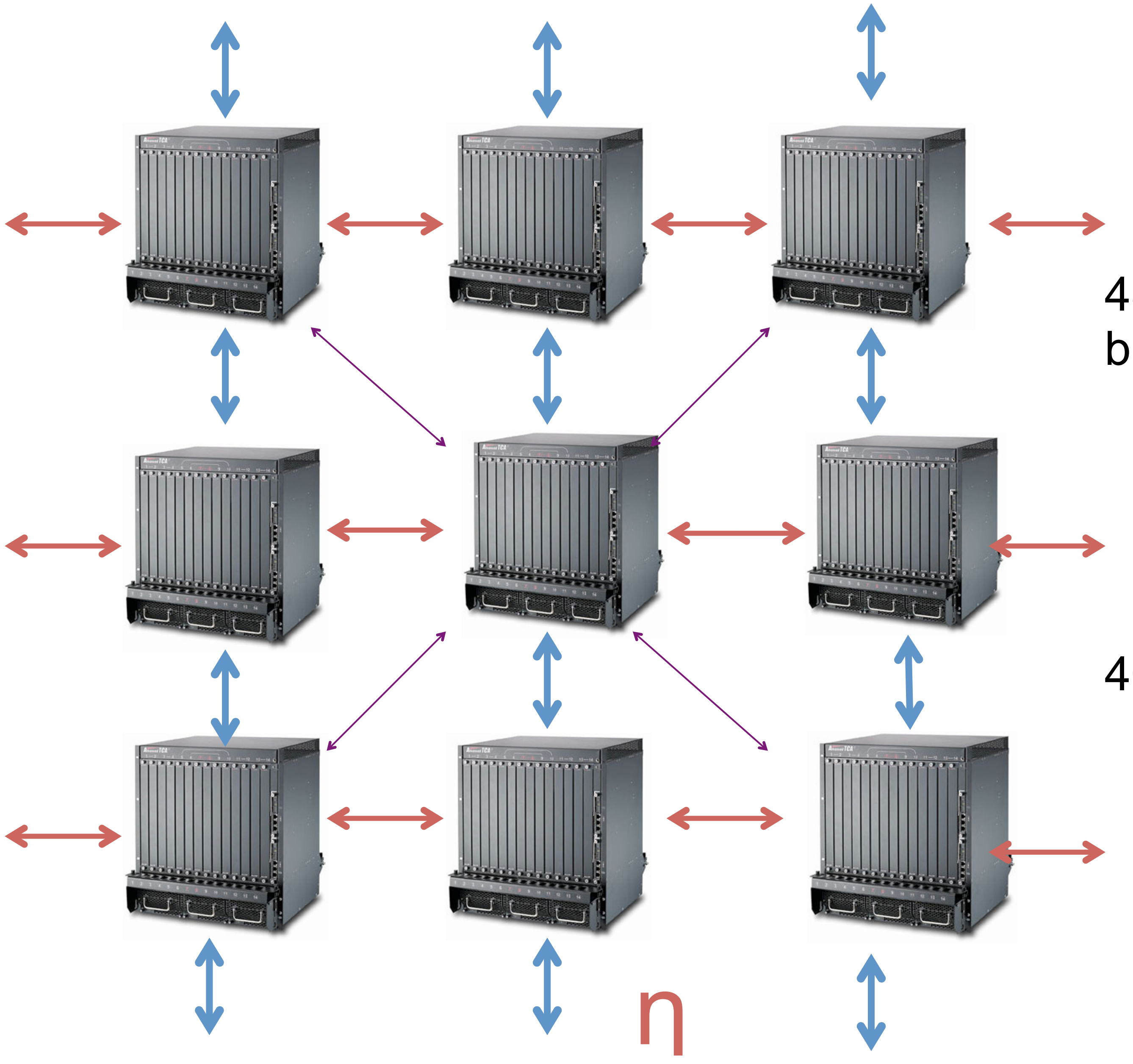
15,000 modules/48 towers
= 312 modules/tower
on average…

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

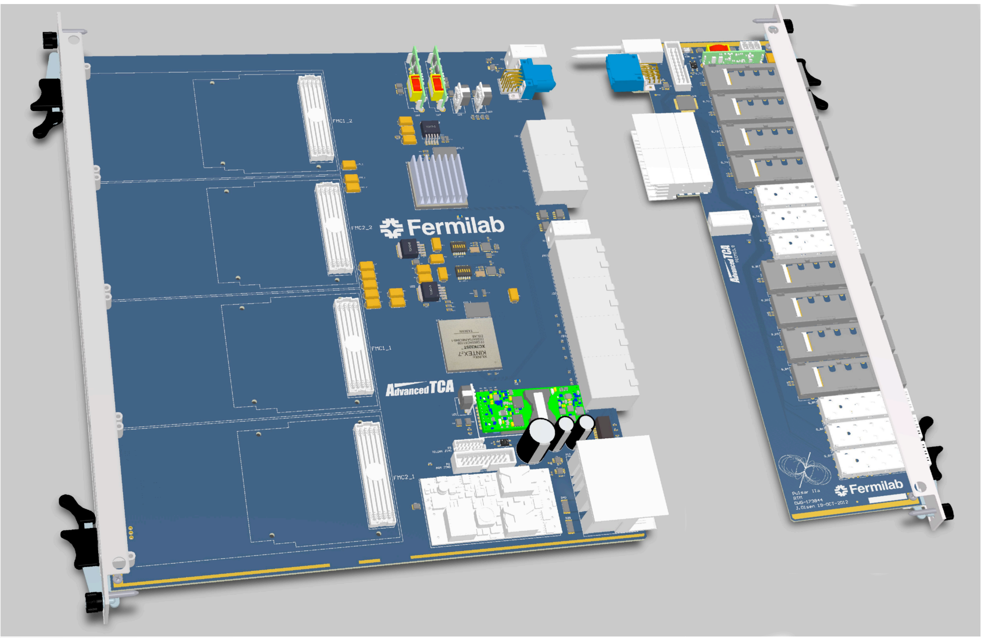
Tower interconnections



48 x 10 Gbps
bidirectional

φ

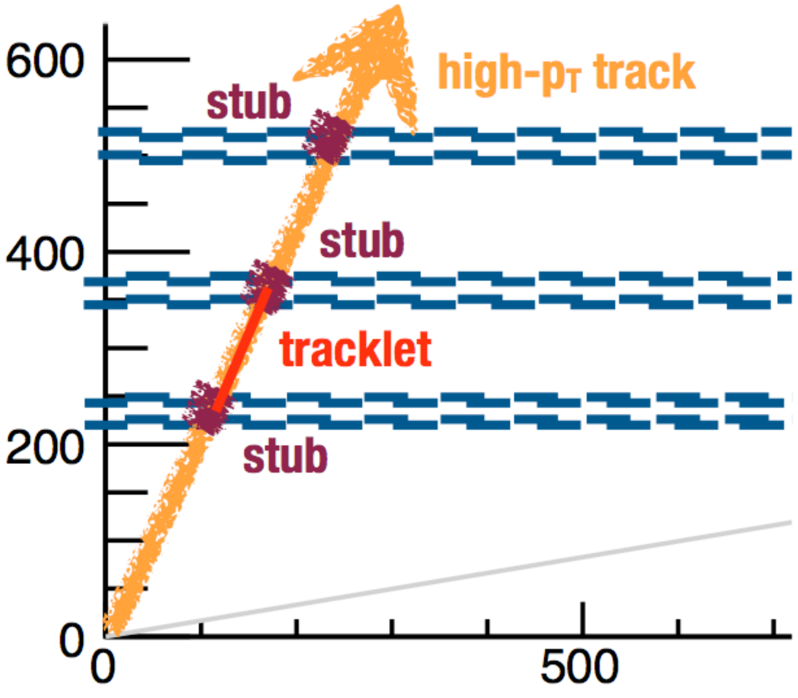
40G full-mesh backplane

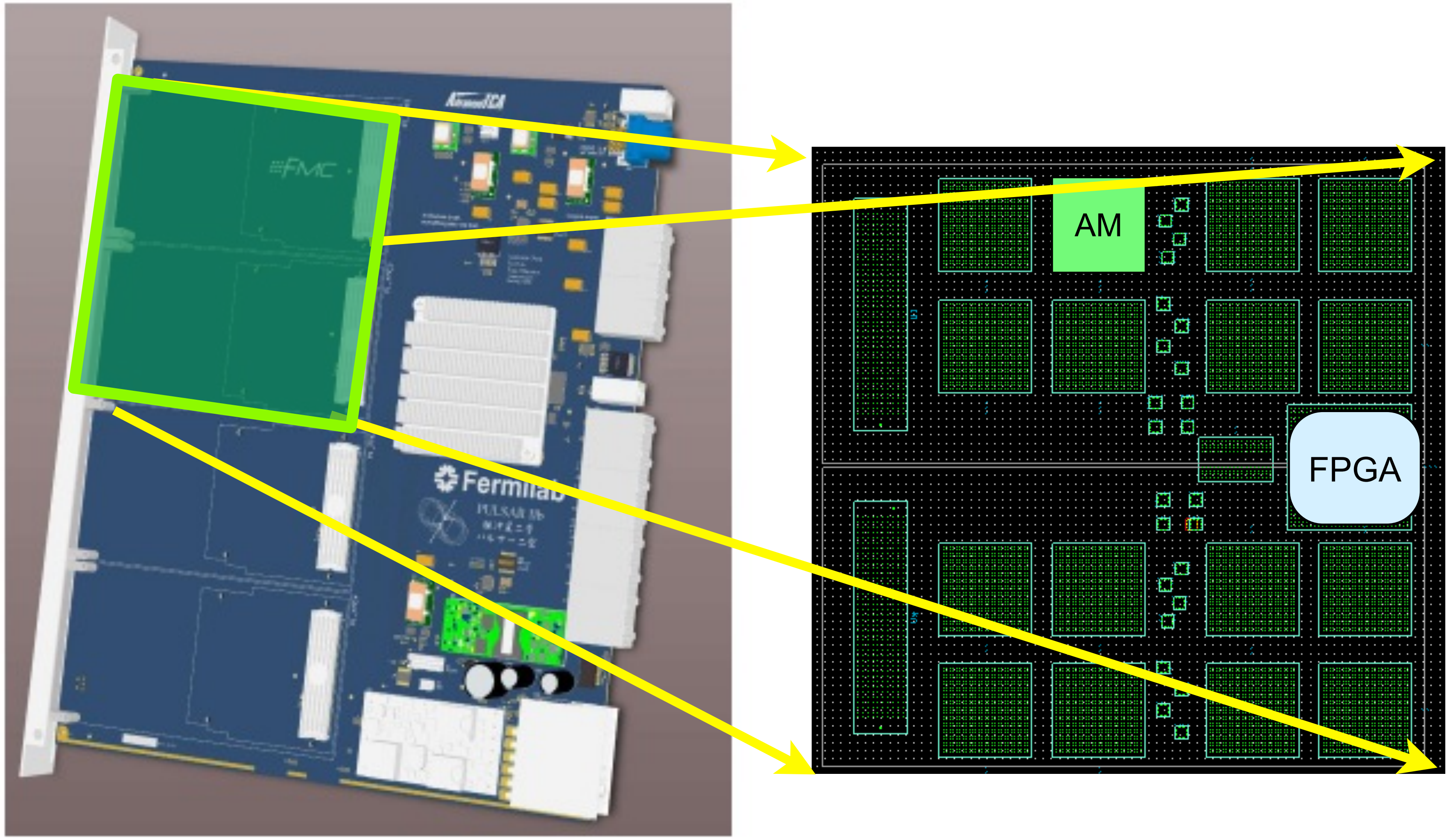


Neighbors
data sharing

Pattern recognition and Track fit

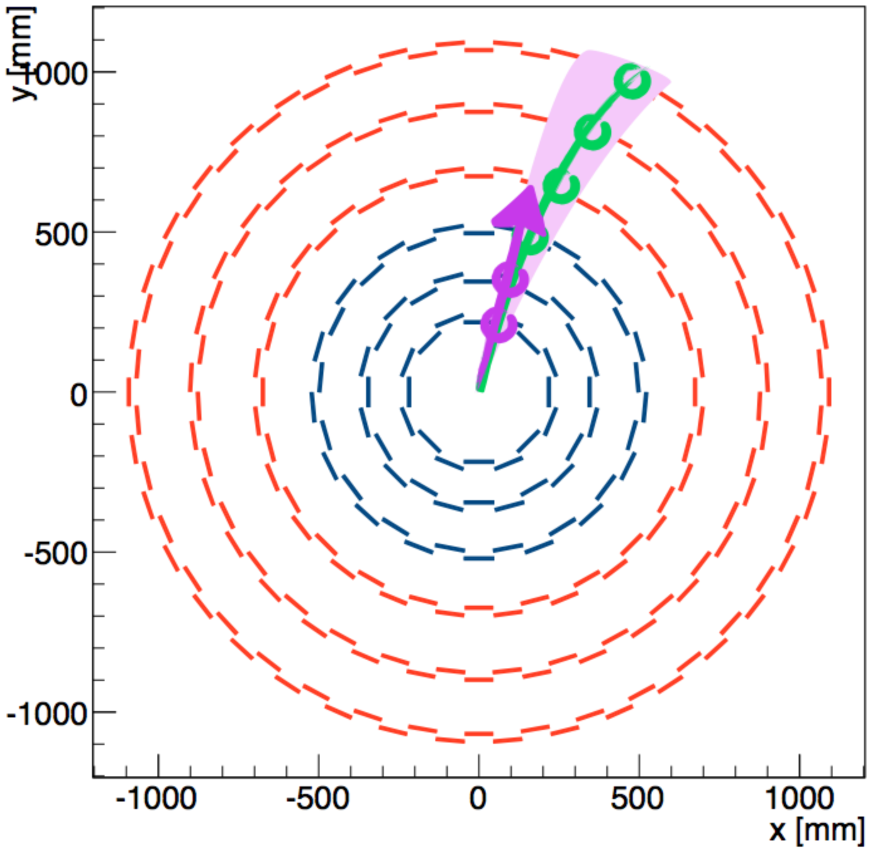
- Associative Memories (pattern recognition) + FPGA (track fit)
  - CMS trigger sectors need ~1M patterns: only 8 state of the art AM06-chip
    - Higher I/O speed (currently 2Gb/s/layer) to reduce time multiplexing
  - ~3000 Track fitting engines using Principal Component Analysis
    - Alternative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





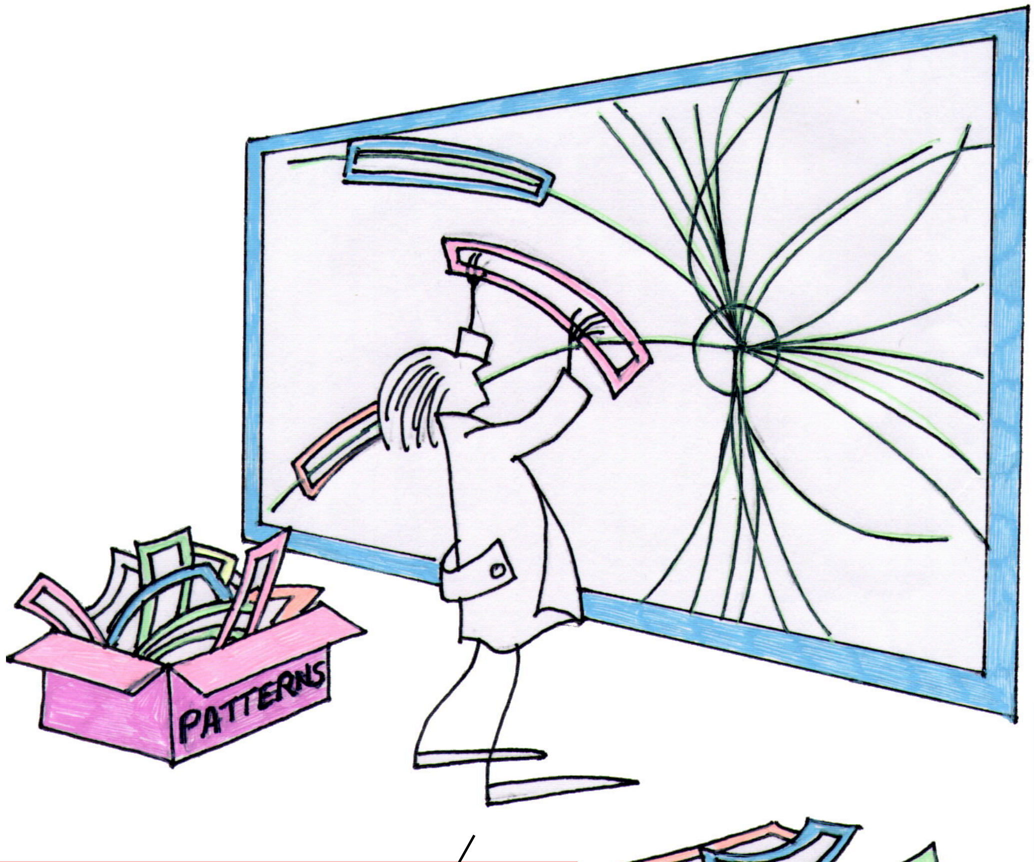
FPGA -only based L1 Track finding

- Find track seeds by correlating stubs in subsequent layers
  - Very much the same approach of “Kalman Filter”
    - Propagate stubs from one layer to another and concatenate stubs within matching windows
    - Linearized Track Fitter
    - Ghost removal



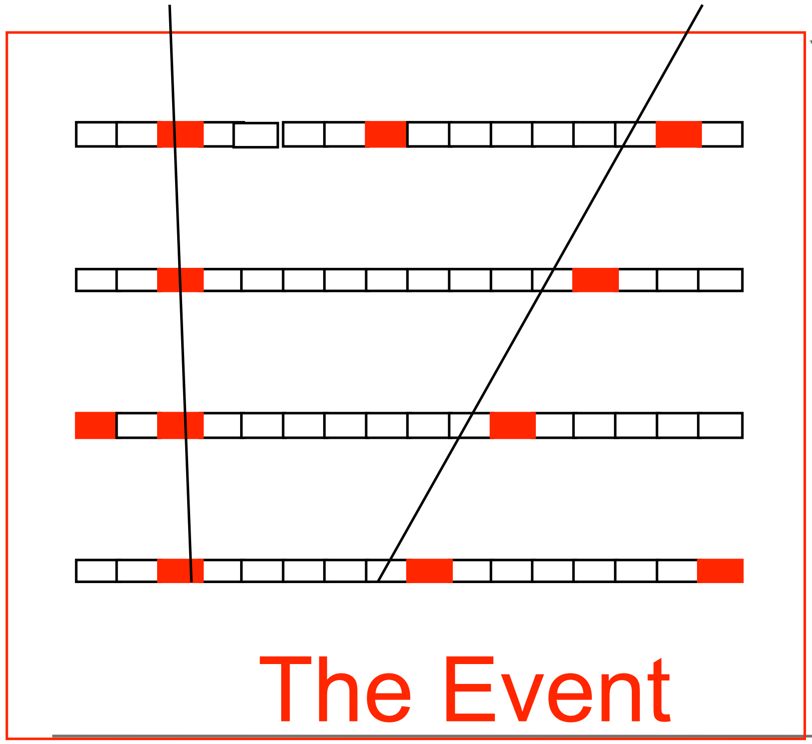
F. Palla INFN Pisa

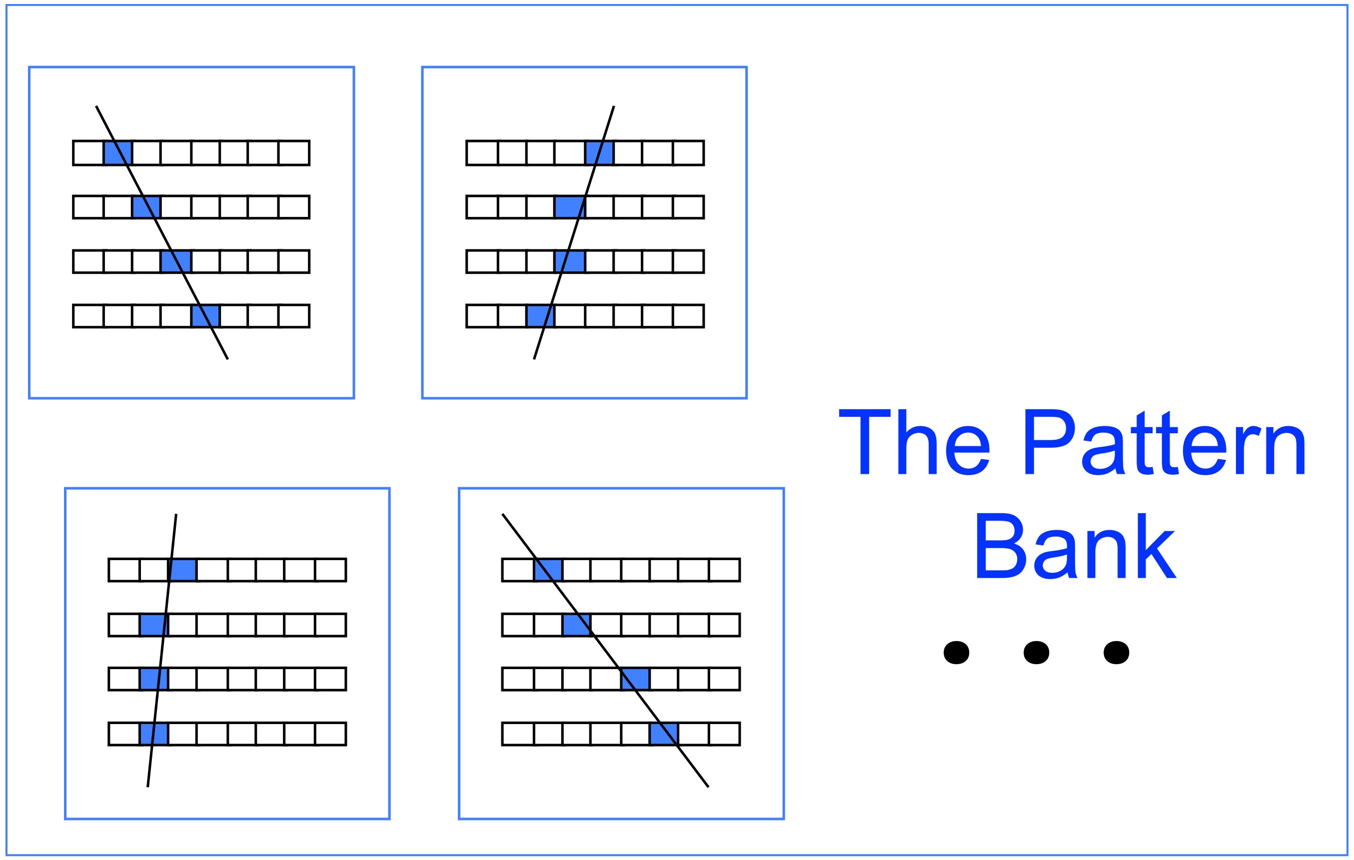
Pattern matching in CDF



The pattern bank is flexible set of pre-calculated patterns:

- can account for misalignment
- changing detector conditions
- beam movement
- …

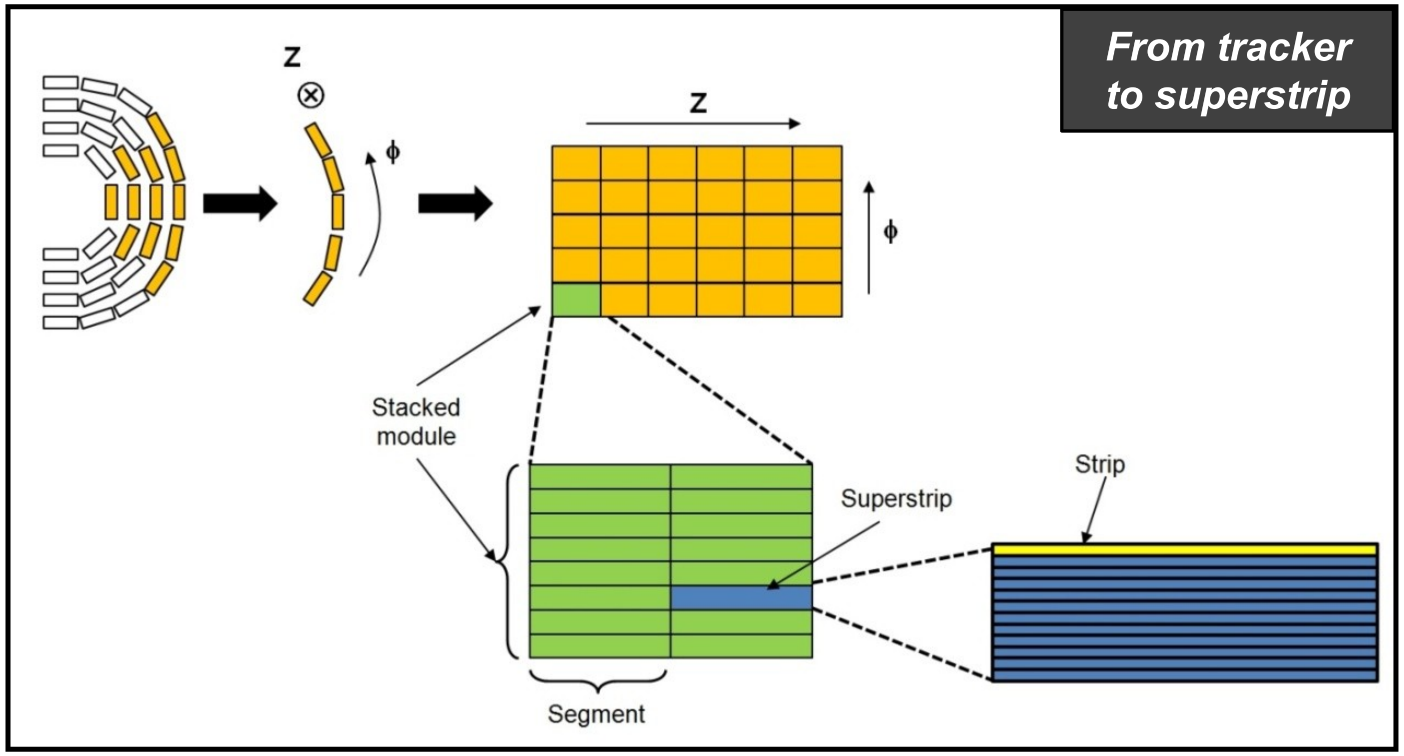

The Event


The Pattern Bank

• • •

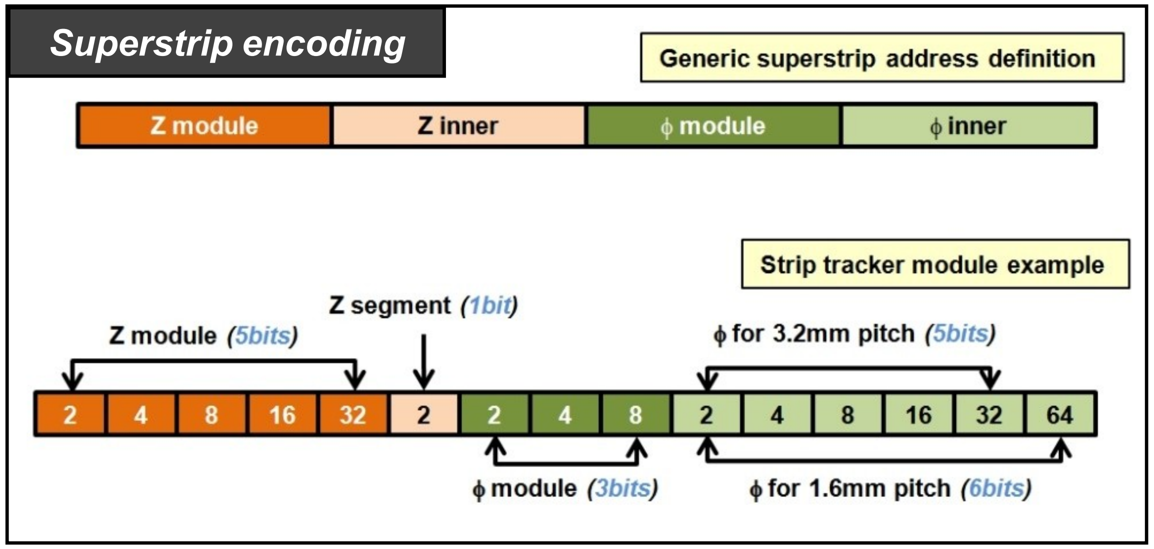
A pattern

→ **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.

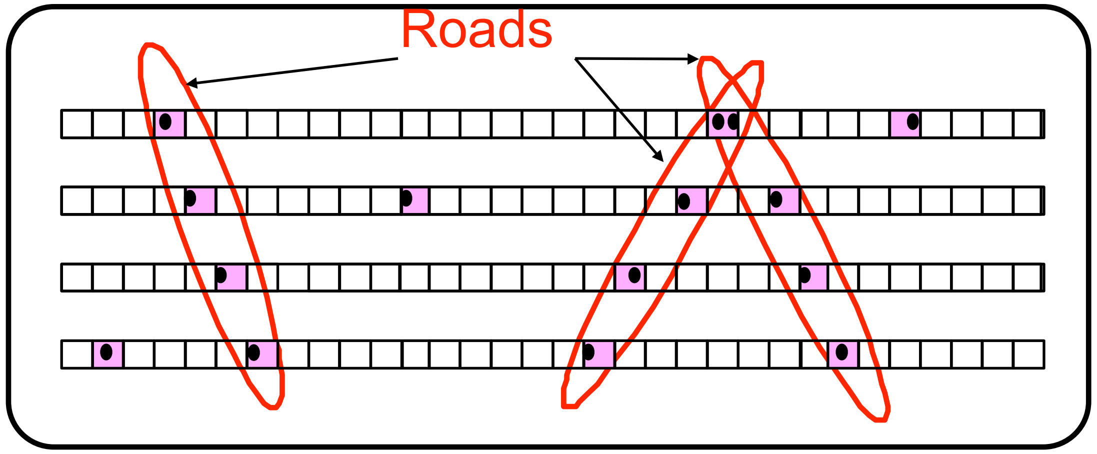


→ The encoding is divided into 4 parts, giving module and intra-module SS position in Z and ϕ direction (*R is not necessary*)

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

Two step approach

1. Find low resolution track candidates called “roads”. Solve most of the pattern recognition

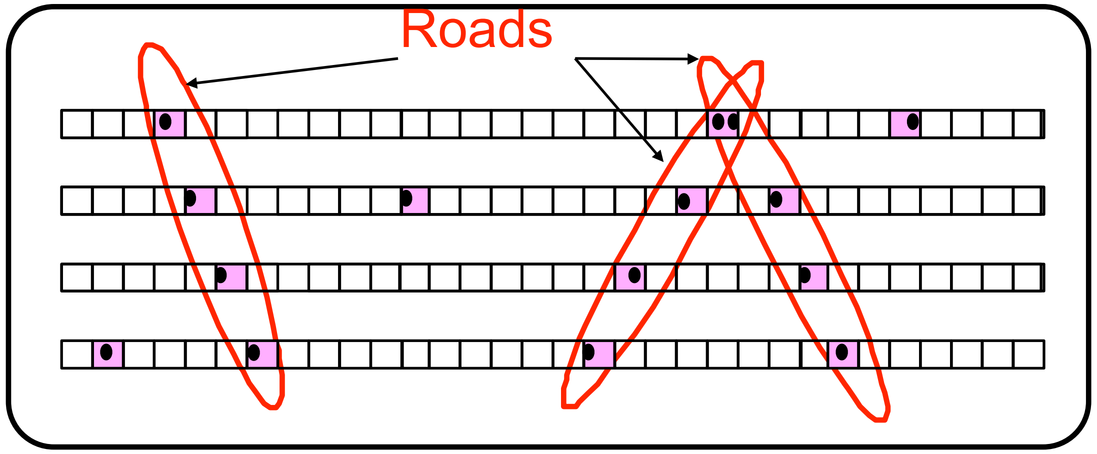


AM chip

FPGA

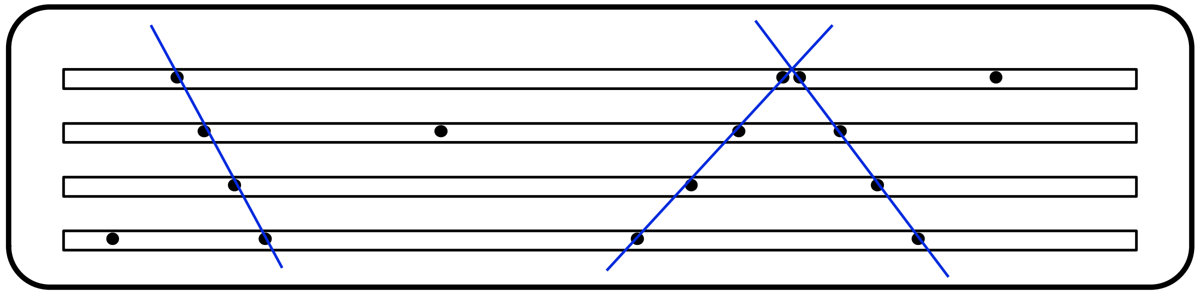
Two step approach

1. Find low resolution track candidates called “roads”. Solve most of the pattern recognition



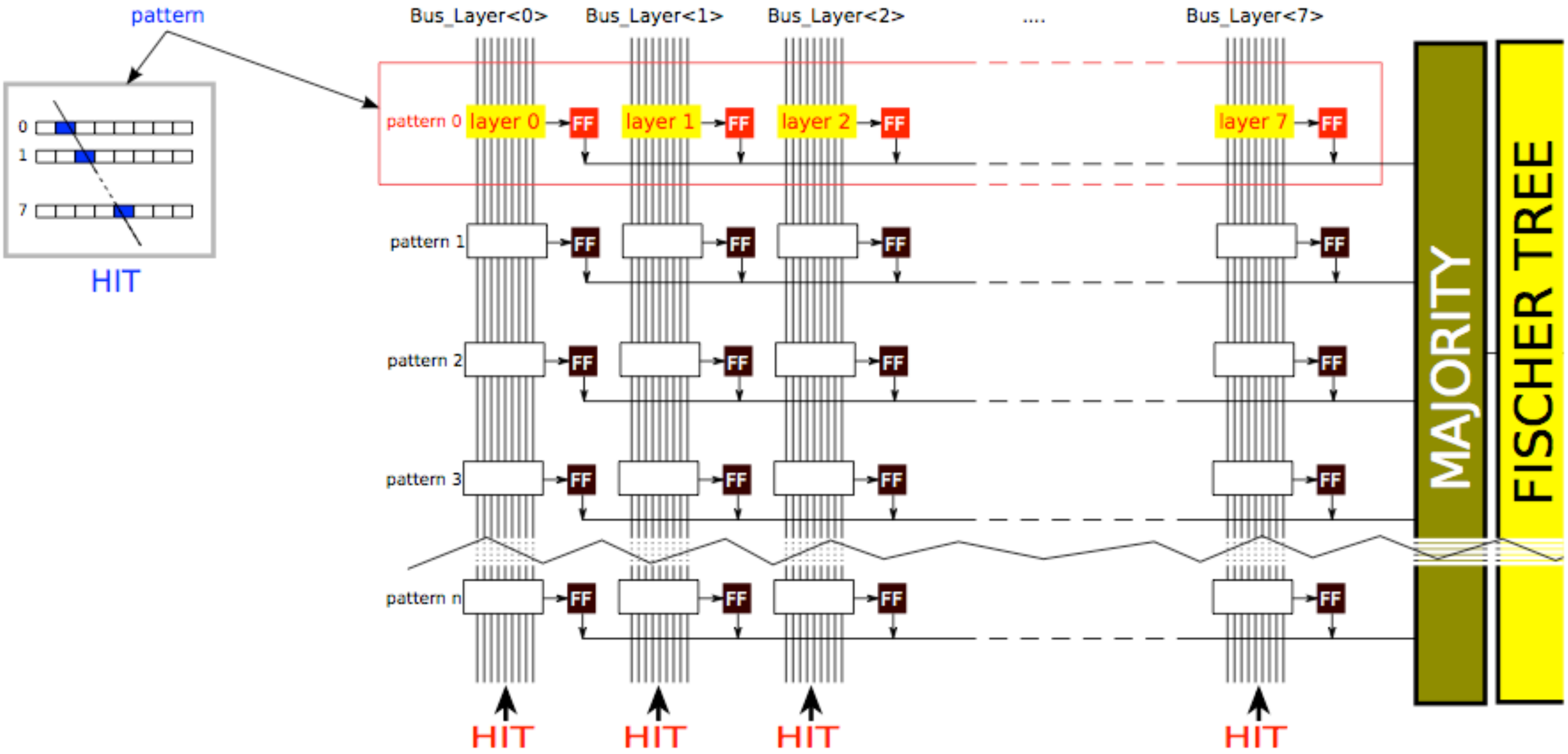
AM chip

2. Then fit tracks inside roads.
   Thanks to 1st step it is much easier



FPGA

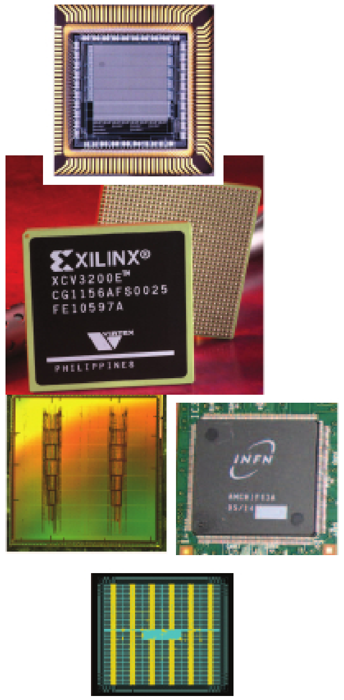
The AM chip at work



The event hit positions are received over 8 input buses of 15 bits each.
All the hits are then compared with the data stored inside each layer block, as soon as they are loaded into the chip, each one in the corresponding bus. If a layer block is matched, the corresponding Flip-Flop (FF) is set. It should be noted that each hit is fed into the memory only once. In fact the bus line transmits the information to all the layer blocks, and, if matched, all the corresponding FF are set simultaneously. **Finally, a given pattern is matched with a logic that counts the number of FF set to 1 within a row, using a majority logic: that means that one could ask a minimum number of FF set**

AM chips history

Several versions of the AMChip:



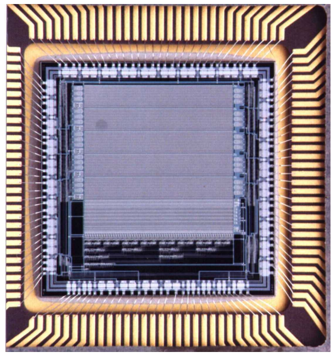
Vers.	*Design*	*Tech.*	*Area*	*Patterns*	*Package*	**Year**
1	Full custom	700 nm		128	QFP	**1992**
2	FPGA	350 nm		128	QFP	
3	Std cells	180 nm	100 mm^2	5 k	QFP	**2000**
4	Std cells + Full custom	65 nm	14 mm^2	8 k	QFP	
mini-5 **5**	**Std cells + Full custom + SERDES IP blocks**	**65 nm**	**4 mm^2** **12 mm^2**	**0,5 k** **3 k**	**QFP** **BGA**	**2014**
6	**Std cells + Full custom + SERDES IP blocks**	**65 nm**	**150 mm^2**	**128 k**	**BGA**	**2015**

red = under fabrication

blue = under design (figures are estimated)

AM technological evolution

SVT
AM chip



- (90's) **Full custom VLSI chip -** 0.7µm (INFN-Pisa)
- **128 patterns, 6x12bit words each, 30MHz**

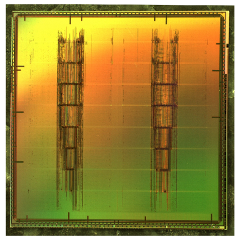
F. Morsani et al., IEEE Trans. on Nucl. Sci., vol. 39 **(1992)**

Alternative **FPGA** implementation of SVT AM chip

P. Giannetti et al., Nucl. Intsr. and Meth., vol. A413/2-3, **(1998)**

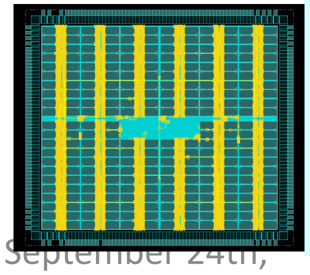
G Magazzù, 1st std cell project presented @ LHCC (1999)

SVT upgrade



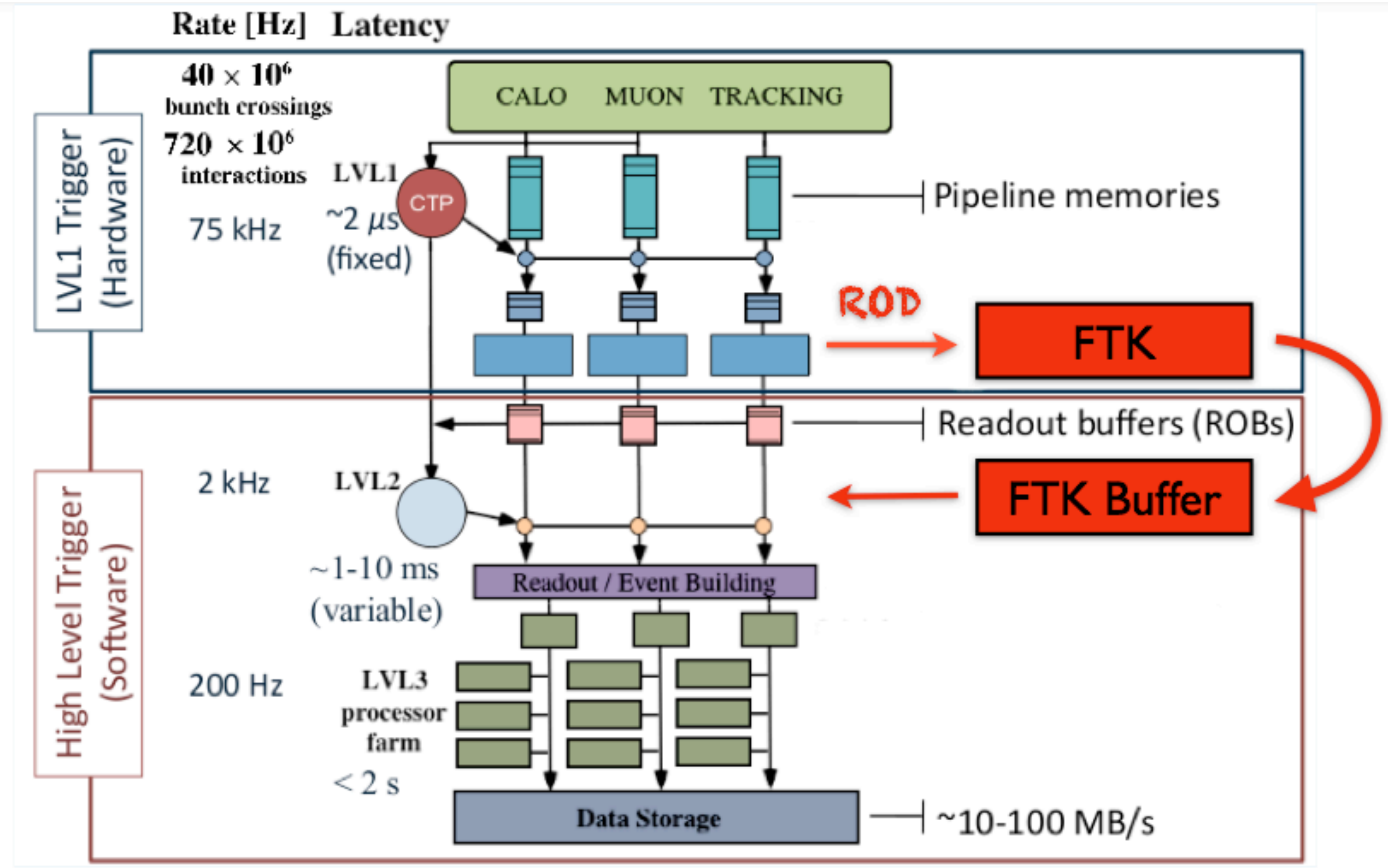
Standard Cell 0.18 µm → 5000 pattern/AM chip
SVT upgrade total: 6M pattern, 40MHz
A. Annovi et al., **IEEE TNS**, Vol 53, Issue 4, Part 2, **2006**

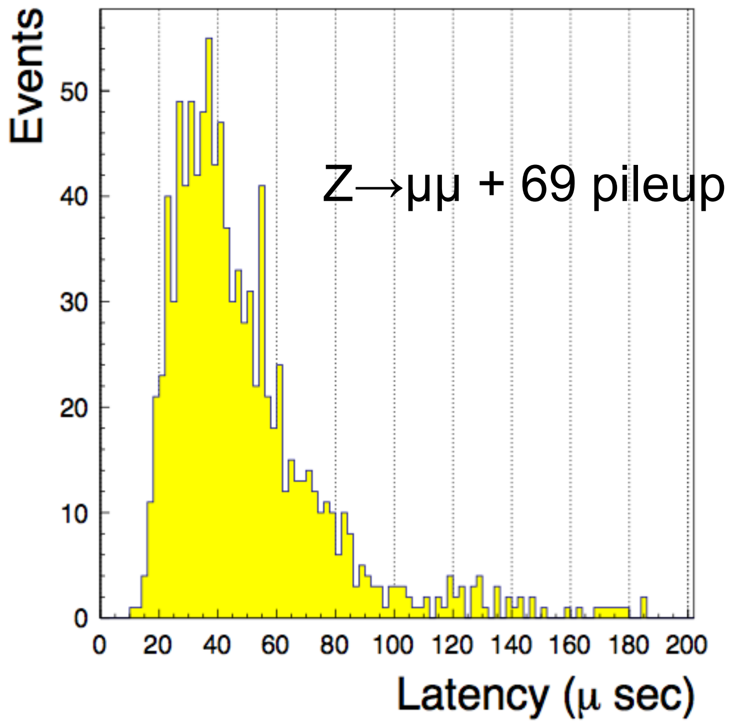
FTK R&D



AMchip04 –65nm technology, std cell & full custom, 100MHz
Power/pattern/MHz ~30 times less. Pattern density x12.
First variable resolution implementation!
F. Alberti *et al 2013 JINST* ***8* C01040,** ***doi:***10.1088/1748-0221/8/01/C01040

Usage in ATLAS @L1.5





Z→μμ + 69 pileup

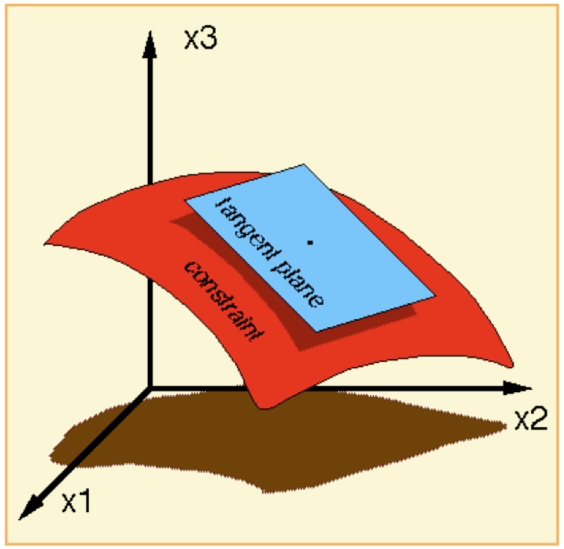
Track fitting - high quality helix parameters and χ^2

Principal component analysis

- 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.

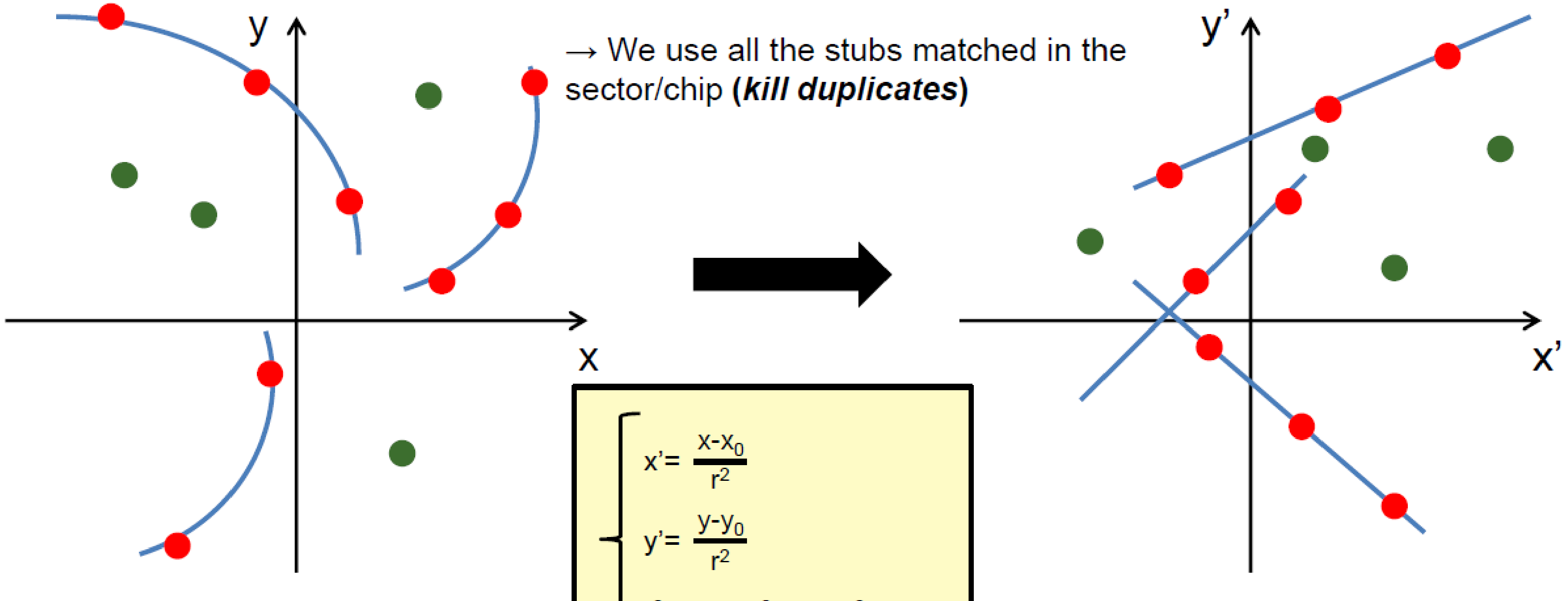
- ~3000 fitting engines/trigger sector for CMS

Hough Transform (1)

→ **Fast hough transform step 1:** ***conformal space mapping***

→ Transform circles into lines. Need one hit on the circle **(x_0,y_0)**.

→ We use all the stubs matched in the sector/chip **(*kill duplicates*)**



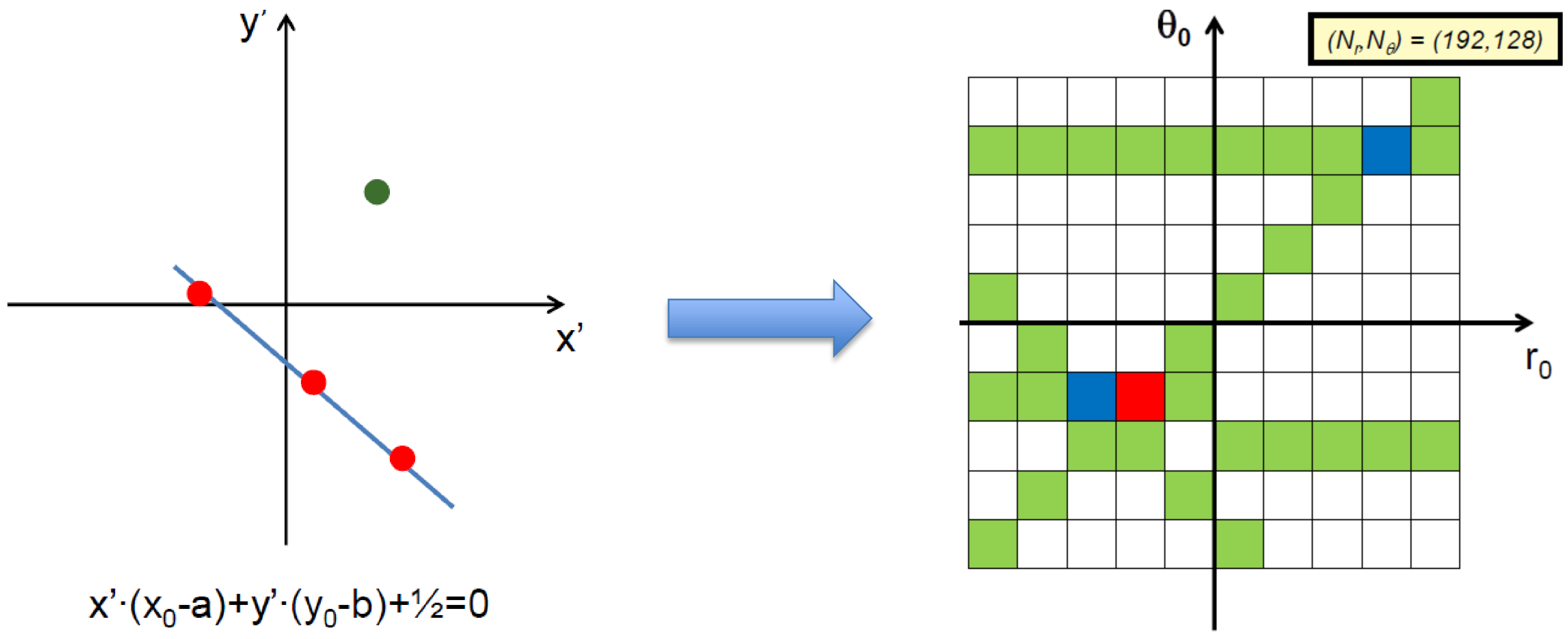
$$\begin{cases} x' = \dfrac{x - x_0}{r^2} \\ y' = \dfrac{y - y_0}{r^2} \\ r^2 = (x-x_0)^2 + (y-y_0)^2 \end{cases}$$

Conformal space

$(x-a)^2+(y-b)^2=R^2$

$x' \cdot (x_0-a) + y' \cdot (y_0-b) + \frac{1}{2} = 0$

Hough Transform (2)



$x'\cdot(x_0-a)+y'\cdot(y_0-b)+\frac{1}{2}=0$

$x'\cdot\cos(\theta_0)+y'\cdot\sin(\theta_0)=r_0$

$x'\cdot\cos(\theta_0^i)+y'\cdot\sin(\theta_0^i)=r_0^i$

First Hough transform made with a coarse binning (simple and fast process to identify accretion)

Retina

→ Ristori, An artificial retina for fast track finding, Nucl. Instrum. Meth. A 453 (2000) 425

1. A track is described by $n(\leq 5)$ parameters (p_T, ϕ, ...)
2. Build a n-dimensional grid with cell indexes i, j, ...
3. Each cell of this space represents 1 single track k with given parameters $p_{T,i}$, ϕ_j, ...
4. For each event, **fill** each cell with this *weight*:

$$R_{ij\ldots} = \sum_m e^{-\frac{s^2_{ij\ldots,m}}{2\sigma^2}} \tag{1}$$

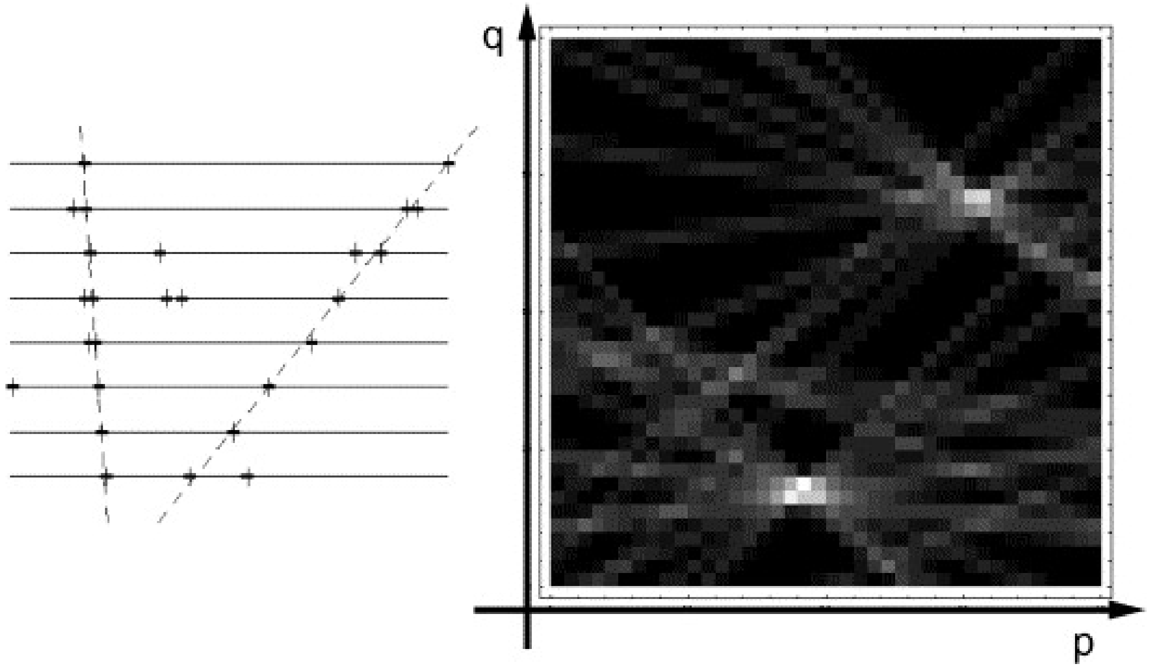
   where

$$s_{ij\ldots,m} = x_m - x_k(p_{T,i}, \phi_j, \ldots) \tag{2}$$

   is the distance between the hit m and the track k (on the hit layer)

5. The sum is on every hit of the event
6. σ is an adjustable parameter
7. The weight is higher if the parametrized track is closer to the hits

Retina (2)



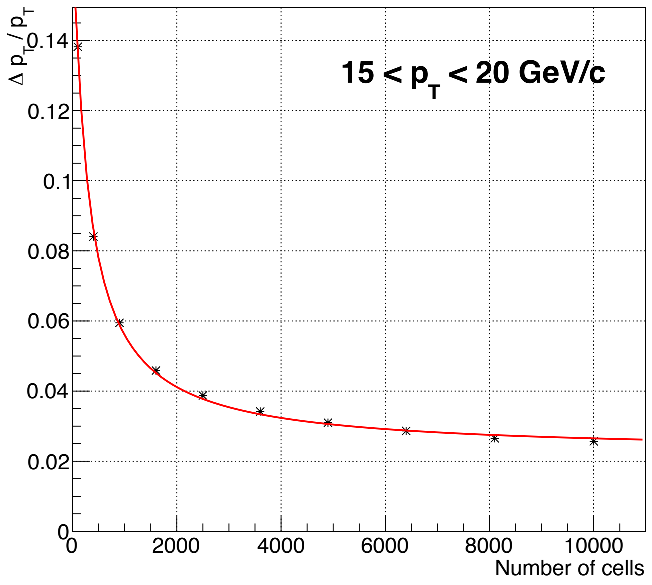
❶ If a weight is high for a given cell, a track with the parameters corresponding to that cell has produced the measured hits

❷ Find maxima on the grid

❸ Each maximum is a track

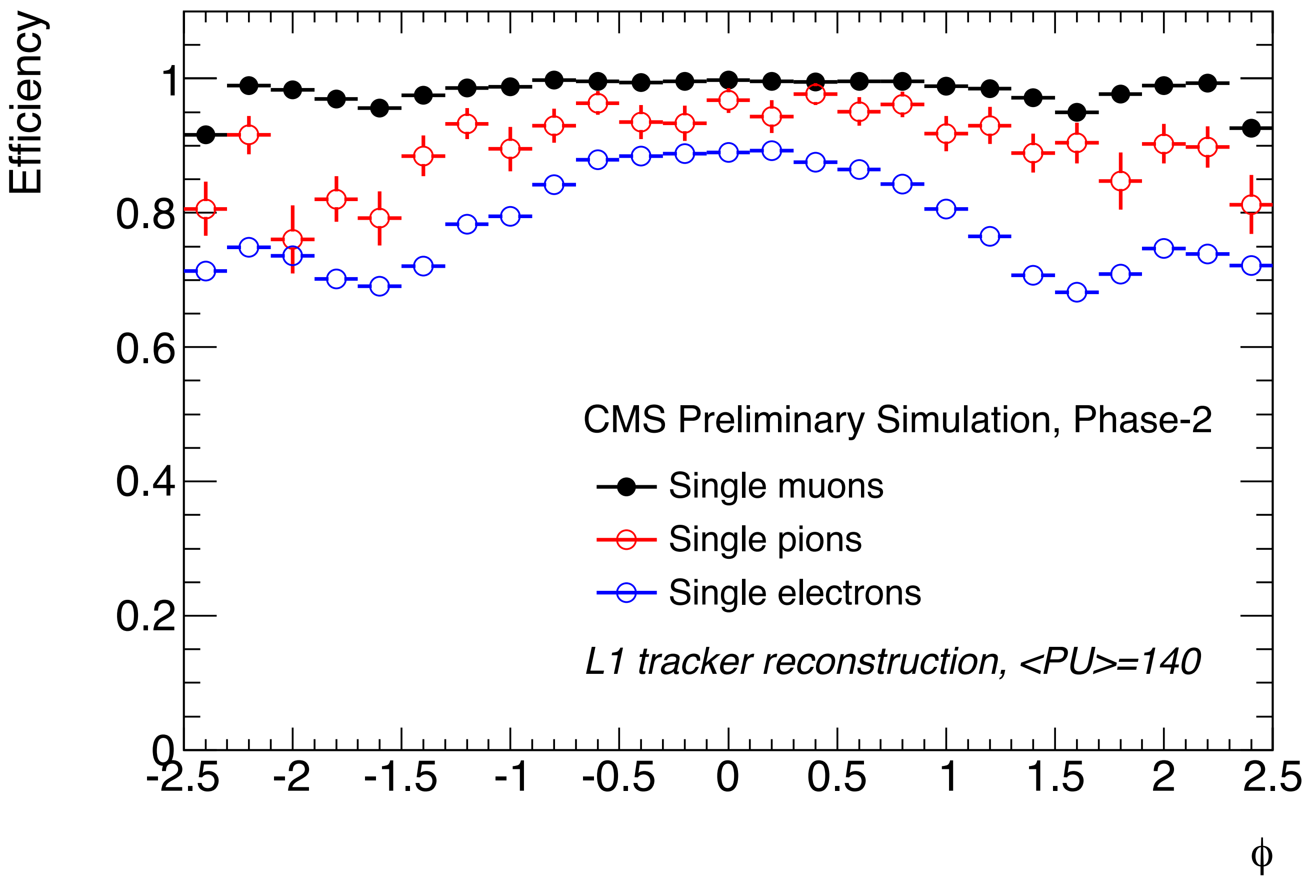
❹ Interpolate between near cells to improve resolution

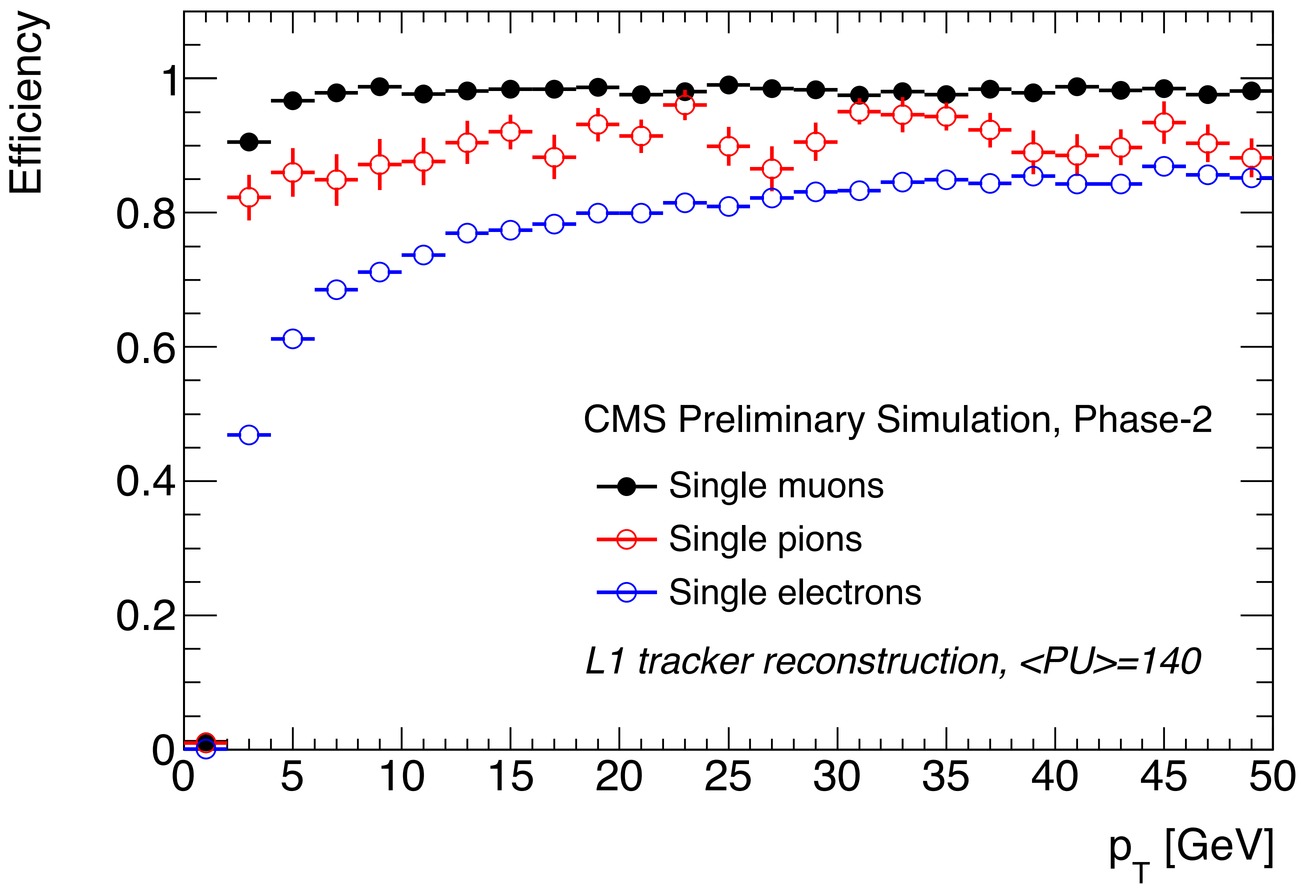
- First implementation in CMS on-going
    - some encouraging initial studies



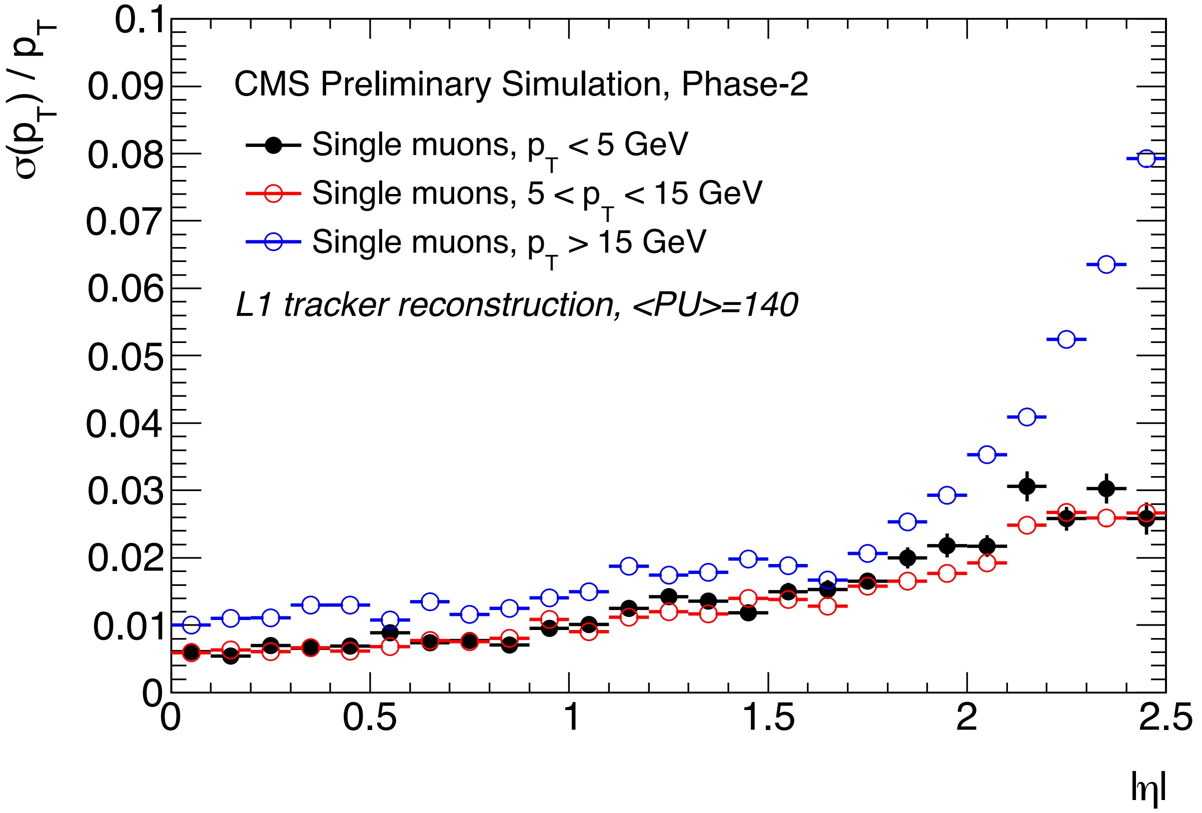
Expected performances (tracklet)

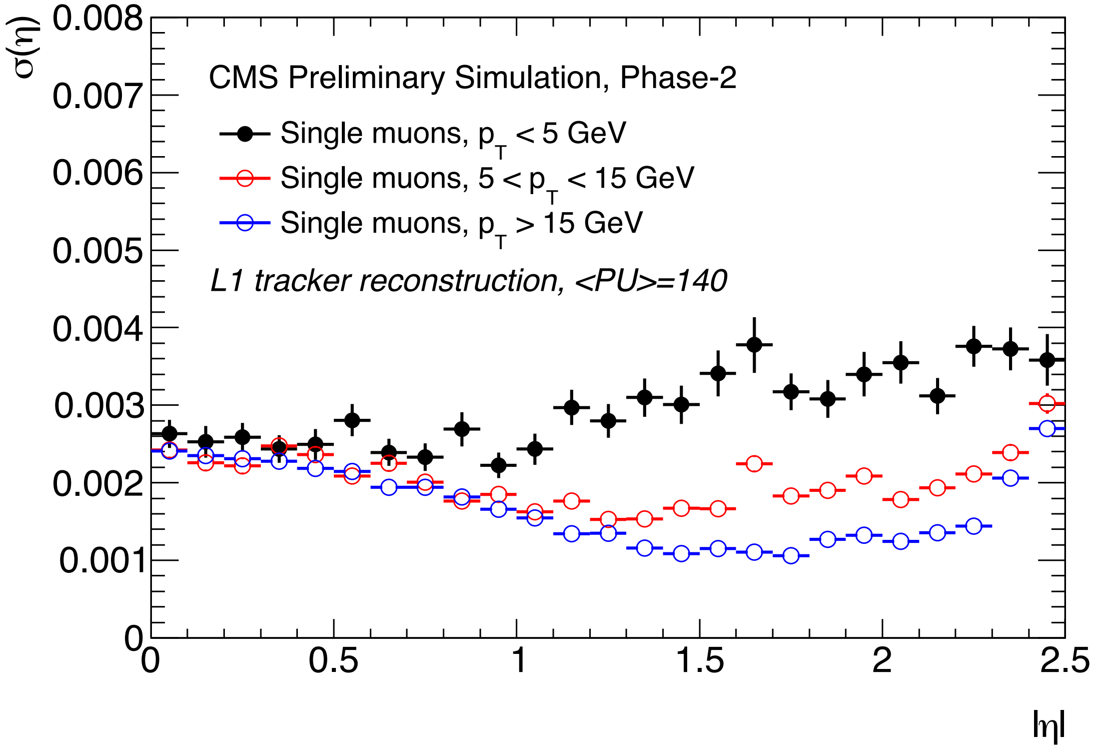
- L1 tracking efficiency as function of η & p_T for single μ, π, e with <PU>=140
  - **Muons** Sharp turn-on at 2 GeV & high efficiency across all η
  - **Pions** Somewhat lower efficiency due to higher interaction rate
  - **Electrons** Slower turn-on curve, efficiency reduced from bremsstrahlung
- For |η| < 1.0 & p_T > 2 GeV, efficiency for **μ**, **π**, **e** is **>99%**, **95%**, **87%**

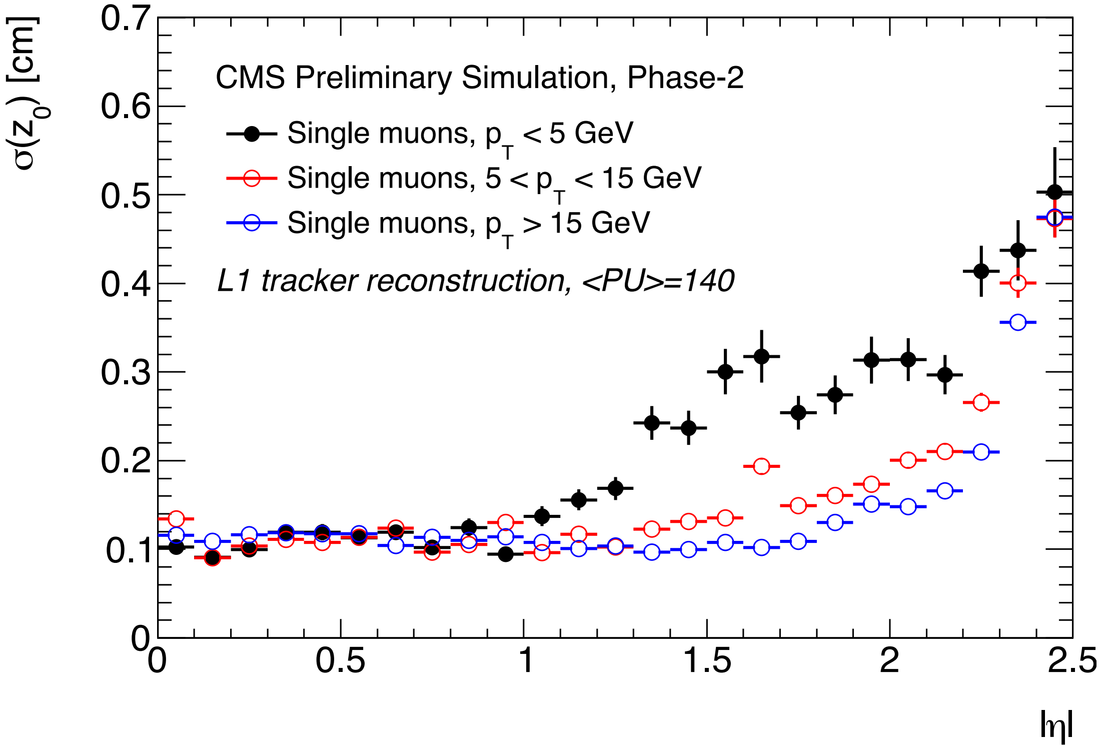




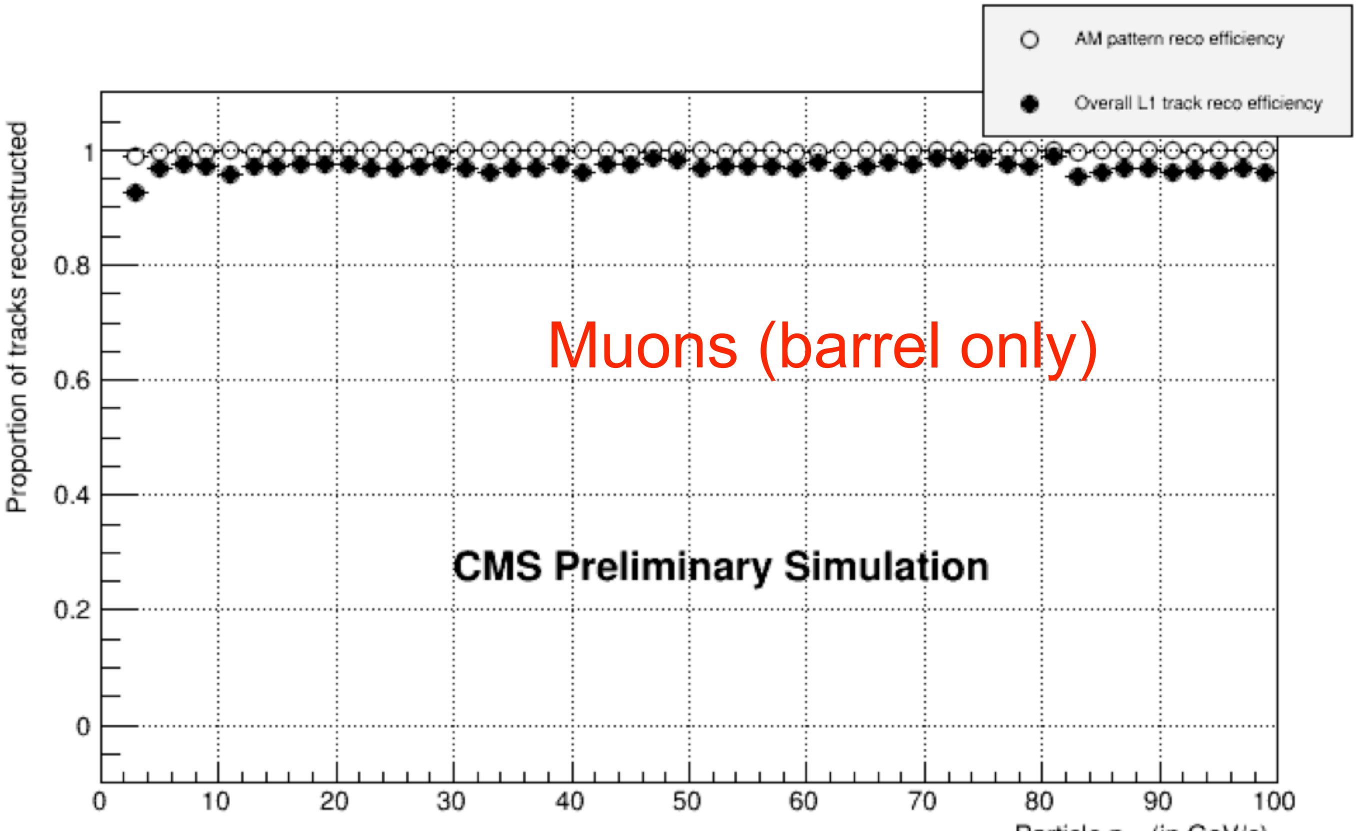
Preliminary resolutions



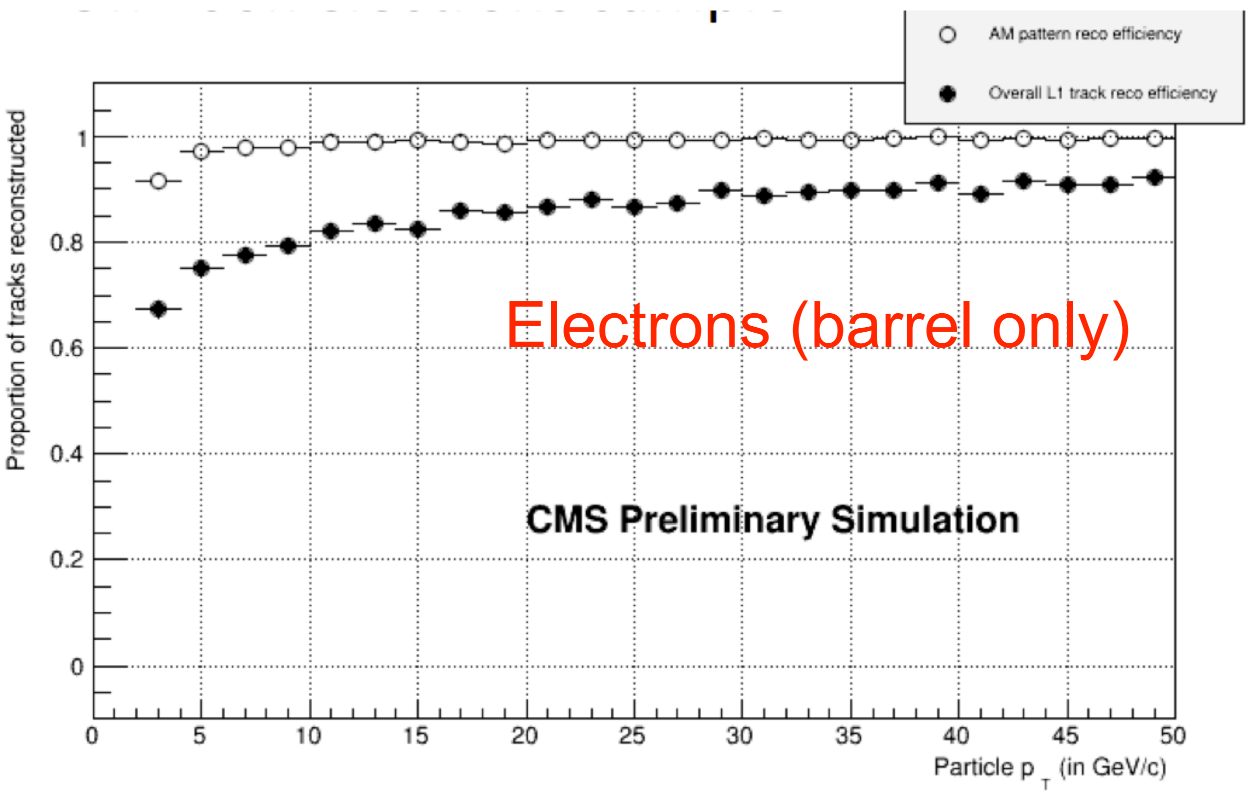




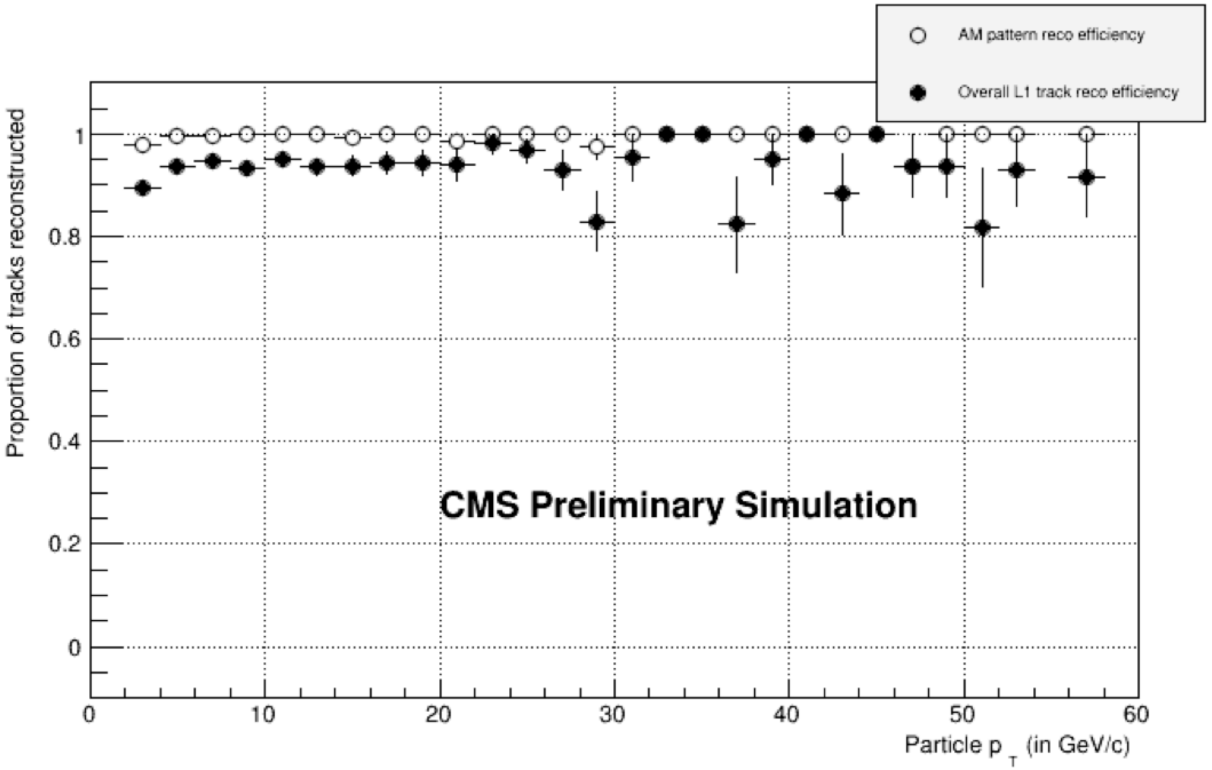
Expected performances (Hough)



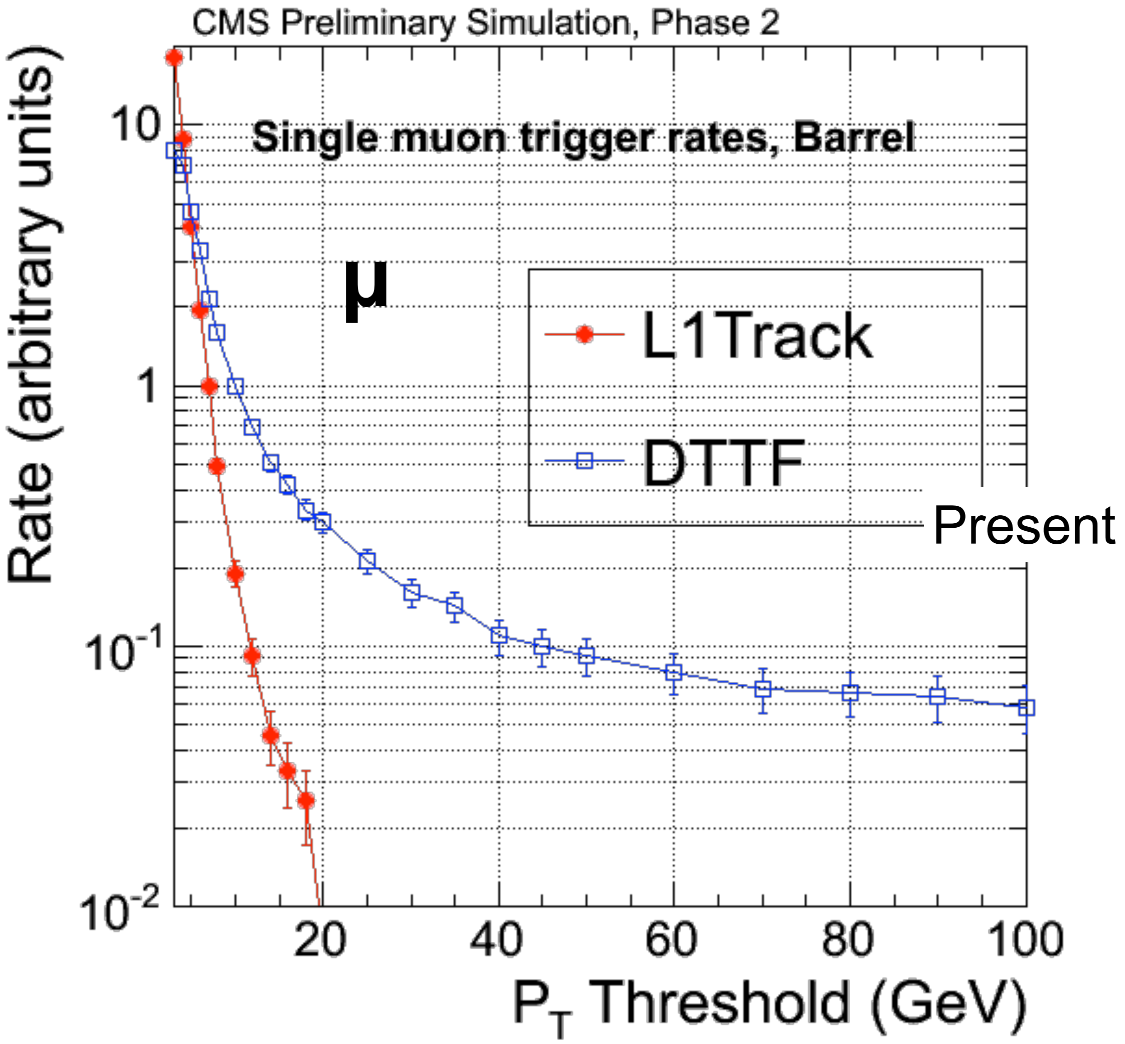
Muons (barrel only)

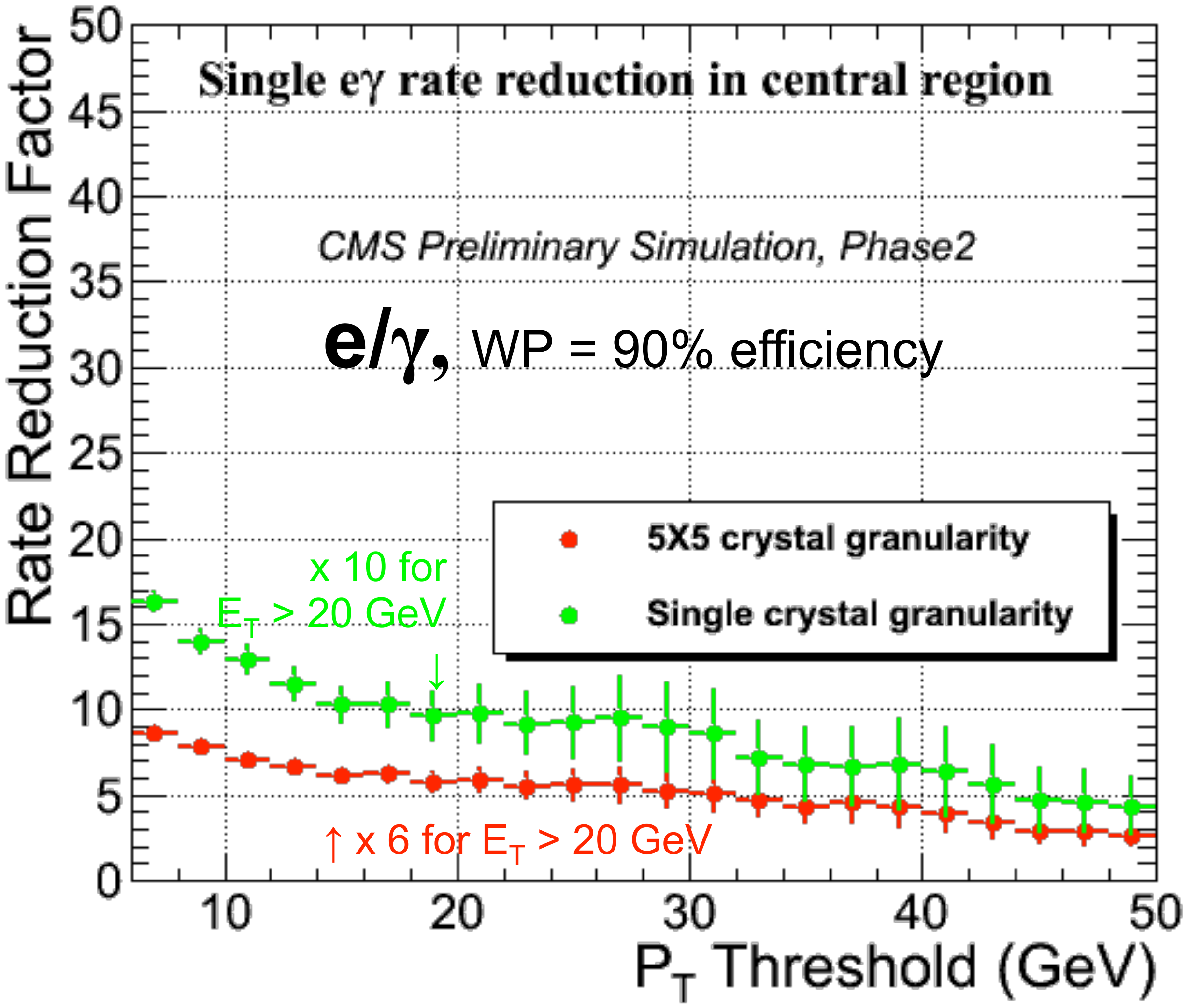


Electrons (barrel only)



Usage of L1 Tracks - CMS



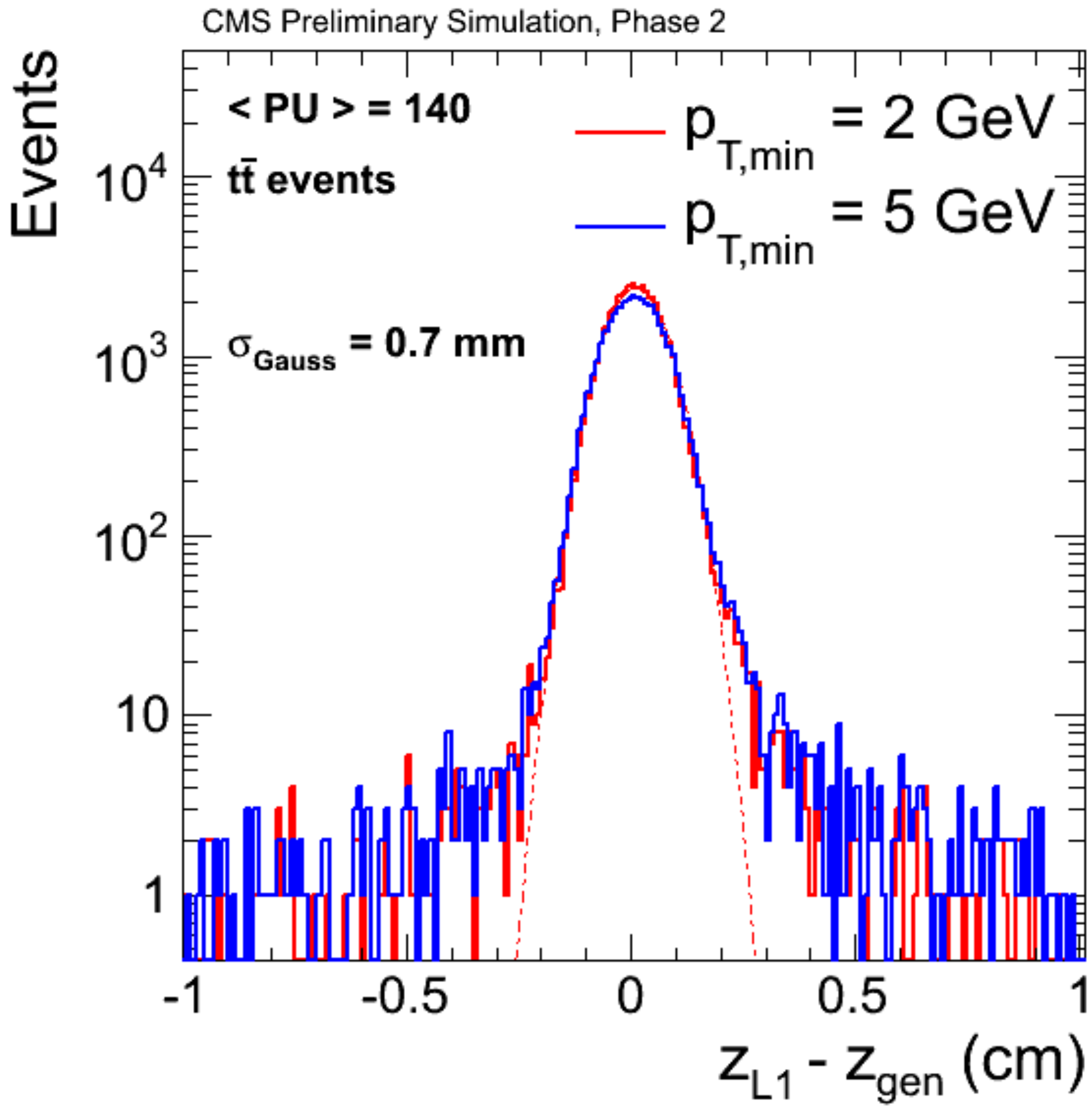


Matching Drift Tube trigger primitives with L1Tracks: **large rate reduction:**
> 10 at threshold > ~ 14 GeV. Normalized to present trigger at 10 GeV. Removes flattening at high P_t

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

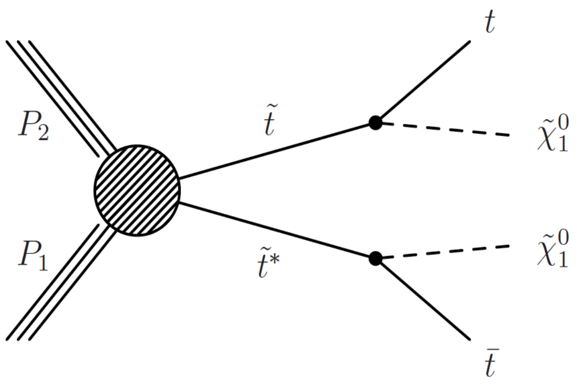
Primary Vertex and Track MET

- L1 tracks can also be used to reconstruct primary vertex of event
- Resolution of primary vertex using L1 tracks with p_T > **2 GeV** or **5 GeV**
  - <1mm for events with large track multiplicity
    - *Here: ttbar <PU>=140*
  - Similar performance with the higher track p_T threshold
- **Track “MET”**
  - Define L1 track-based missing transverse momentum from L1 tracks coming from primary vertex

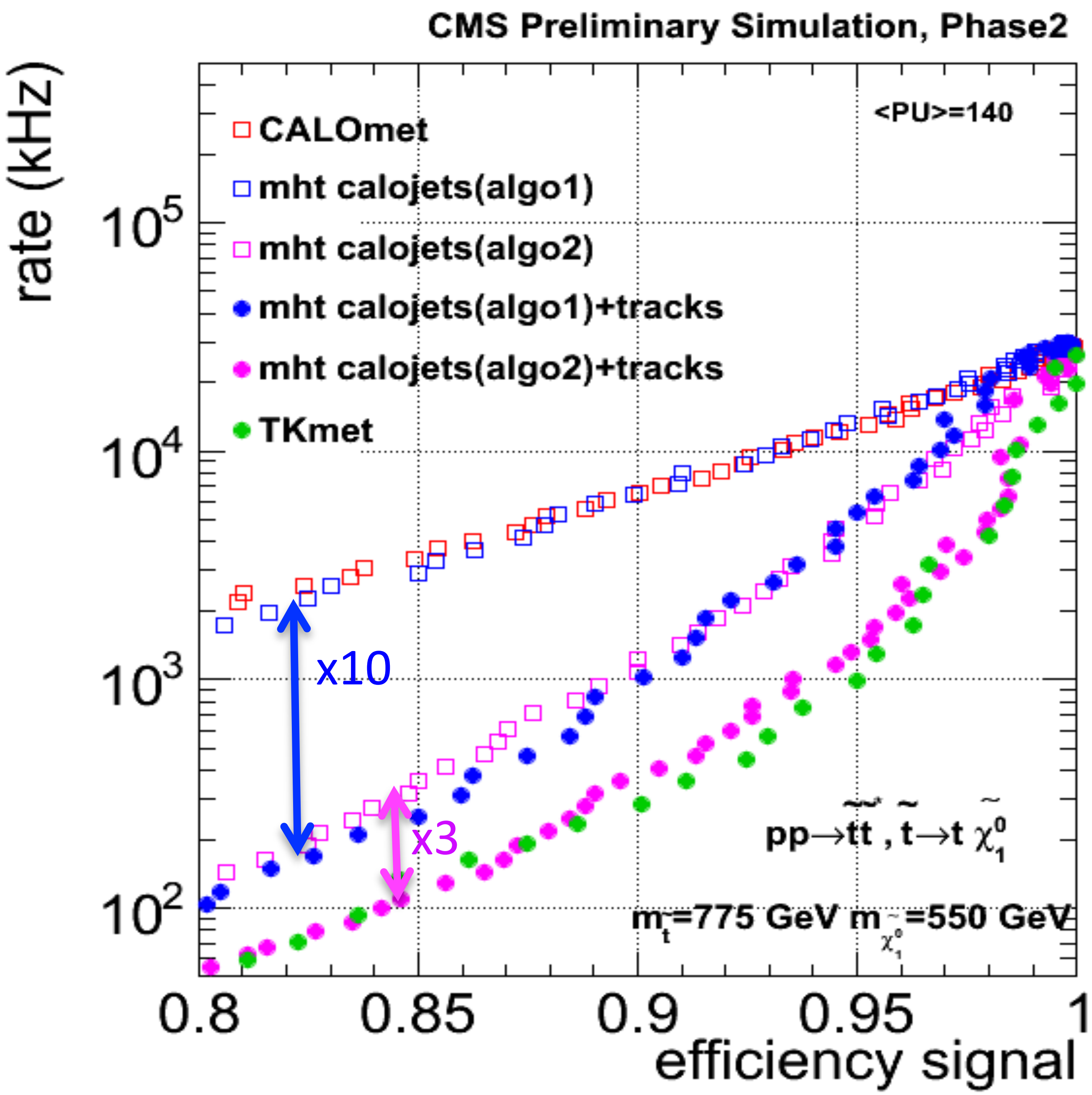


SUSY Signal point

- Rate reductions using L1 tracks for SUSY signal
  - Stop pair production with hadronic top decays (stop=775 GeV, LSP=550 GeV)
  - Signal defined by genMET > 100 GeV



- Missing H_T determined with/without vertex association
  - Algo1 & Algo2: Calorimeter-based L1 jet algorithms with different PU subtraction methods
- Sizable rate reductions achieved with tracking information!



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 ($\|\eta\| < 1$)	~ 90 %	Tracker isolation may help further.
Single Electron, 20 GeV	Match with cluster	> 6 (current granularity) >10 (crystal granularity) ($\| \eta \| < 1$)	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.

Conclusions

- Tracker information helps reducing drastically the rate of uninterested events
  - This will become a new “must” for all future detectors
- HL-LHC detectors will make use of tracking information in the Level-1 Triggers
  - Several trigger architectures exploited
    - Full readout @40 MHz, on-detector data reduction using p_T-modules
    - 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
  - High statistics of useful events for precision physics available
  - Stay tuned!