Convergence, 1952 by Jackson Pollock

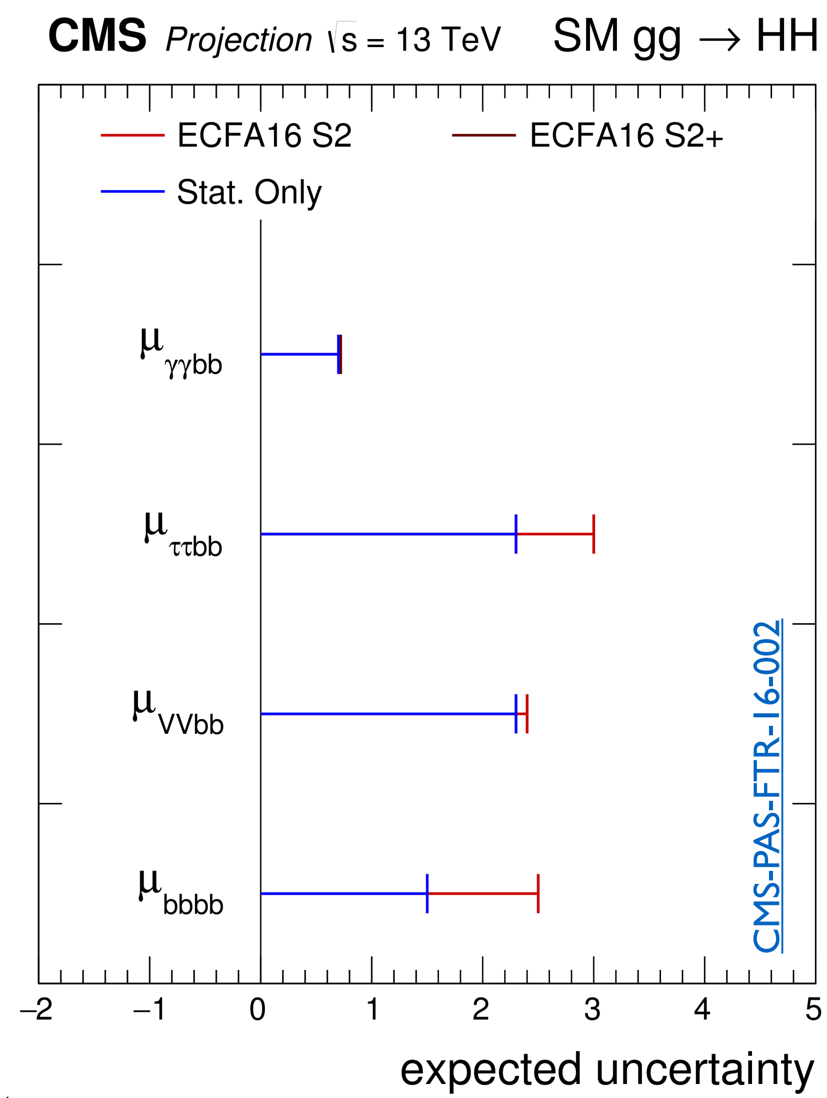
Reconstruction and clustering for the CMS High Granularity Calorimeter

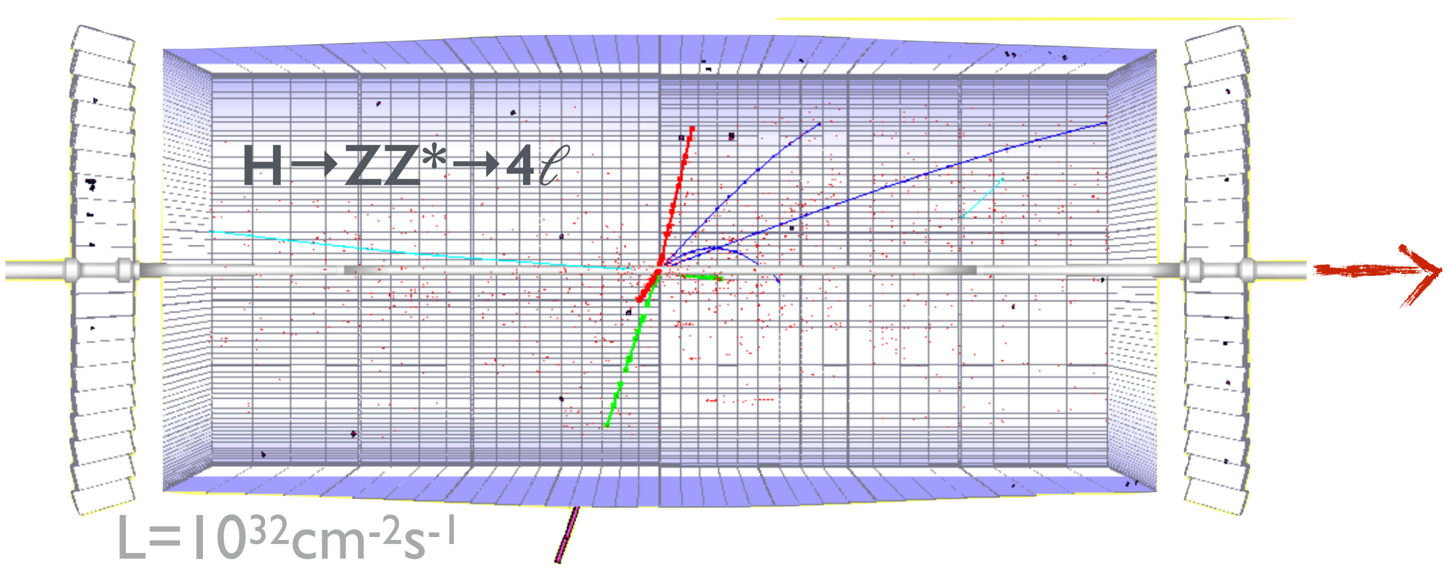
P. Ferreira da Silva (CERN) *for the CMS Collaboration*

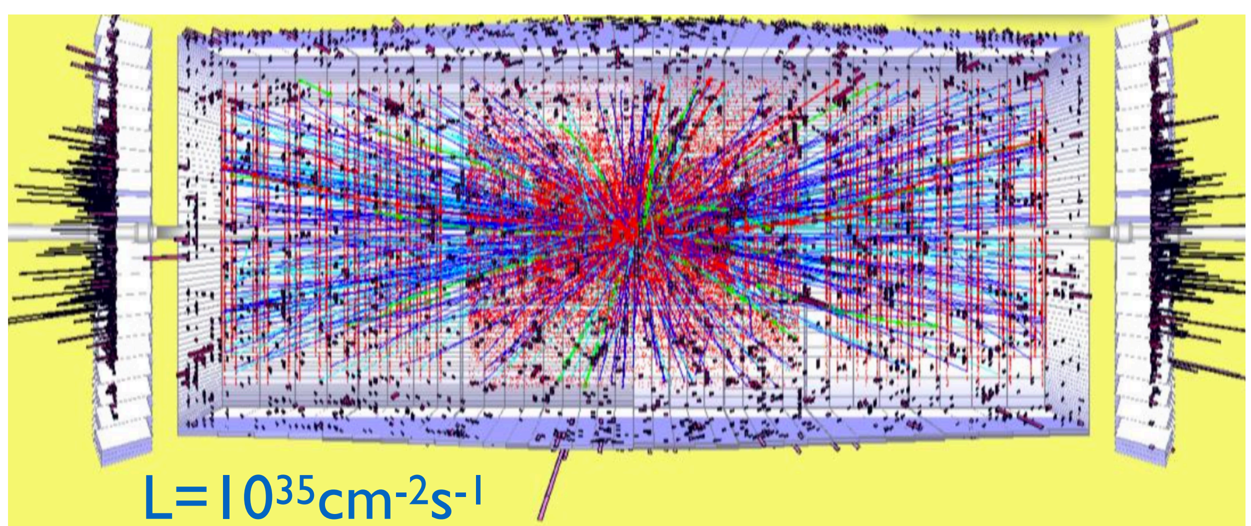
CALOR 2018 May 21-25, 2018, Eugene, USA

Towards new discoveries in the high luminosity regime

- **3ab^{-1} @ 14 TeV are needed = a long road ahead to**
  - pin down Higgs couplings to <5-10%:
  - VBF Higgs measurements at the level of 10%
  - claim indication (3σ) for HH after combination →
  - … but also looking for rare processes (tails or weakly coupled)
- **All cases need good control of b-jets, τ-leptons, photons,…**
  - **particle flow** is the reconstruction backbone for CMS
- **Processes tends to go forward with the increase in $\sqrt{s}$**
  - where most of the 140-200 soft min. bias pileup events will accumulate…







140 pileup events super-imposed

Fine granularity in CMS Phase 2

Main goals: mitigate pileup contamination to signals, maintain or improve Phase I performance

longitudinal

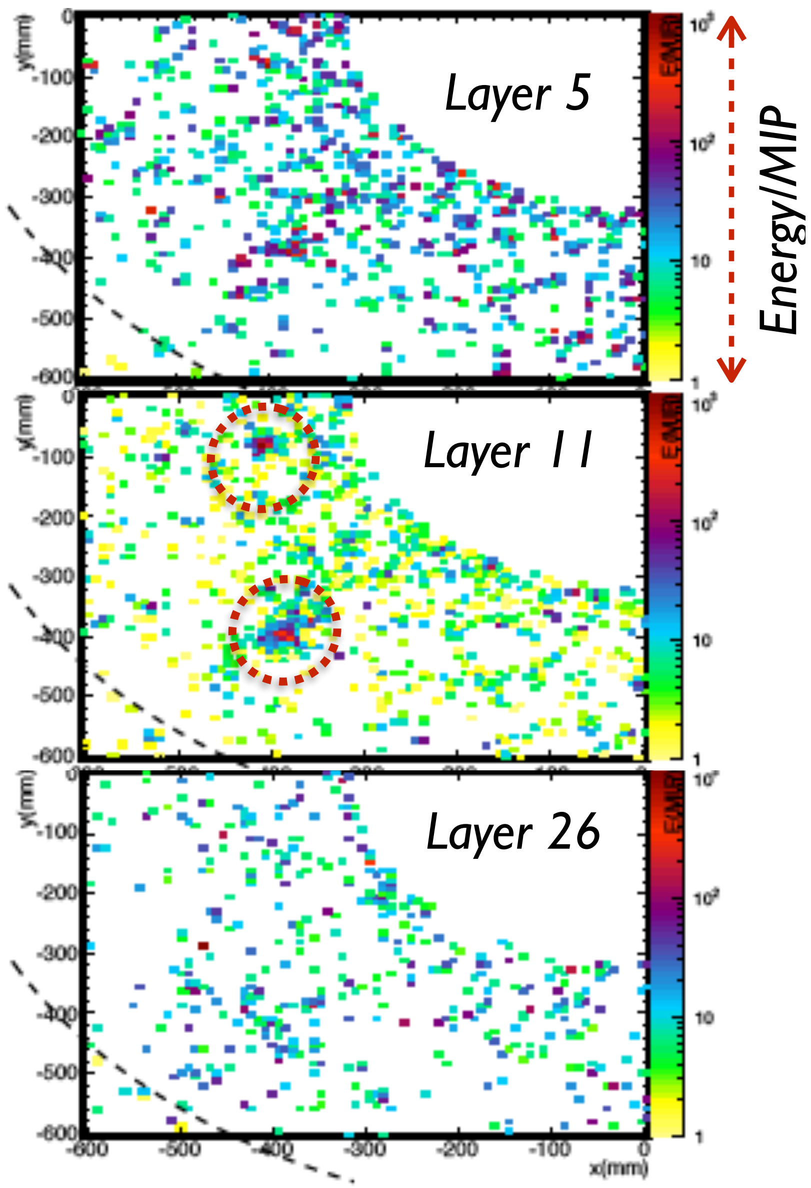
enable feature extraction

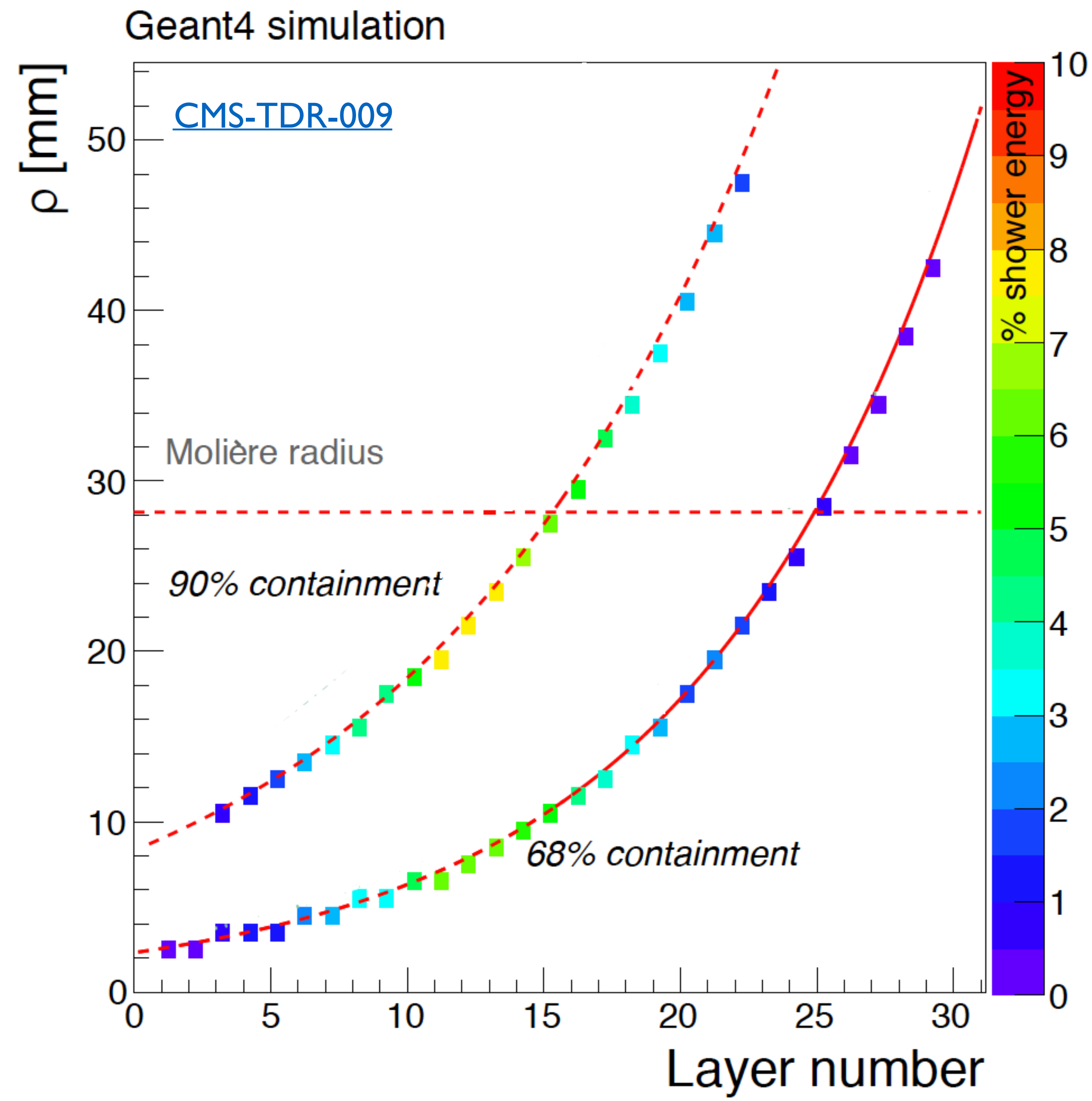
absorb low energy in the first layers

transverse granularity

reduce flux/cell

resolve fine structures



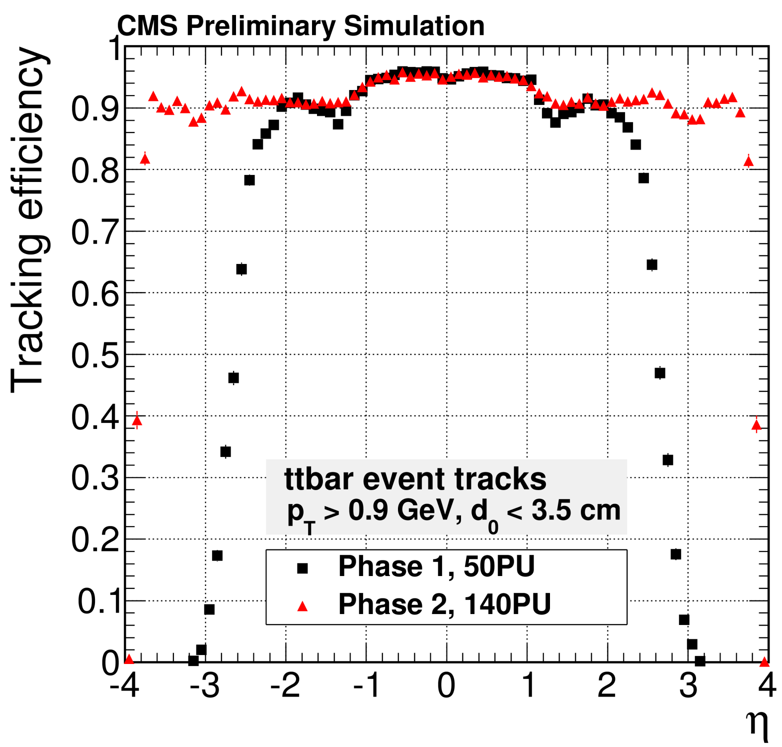


Fine granularity in CMS Phase 2 - cont.

Main goals: mitigate pileup contamination to signals, maintain or improve Phase I performance

extended tracking coverage

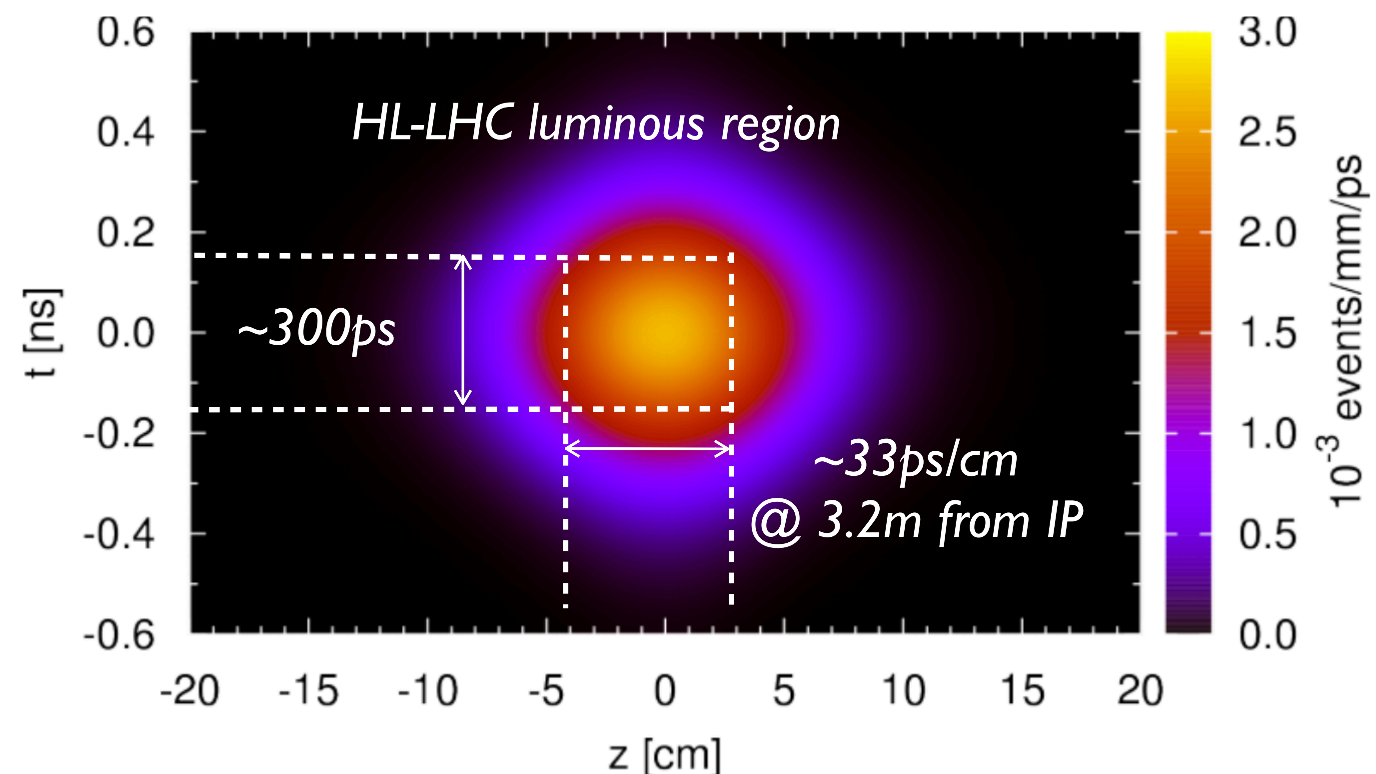
up to η=4 for better particle flow



90% efficient
for 2-8% fakes

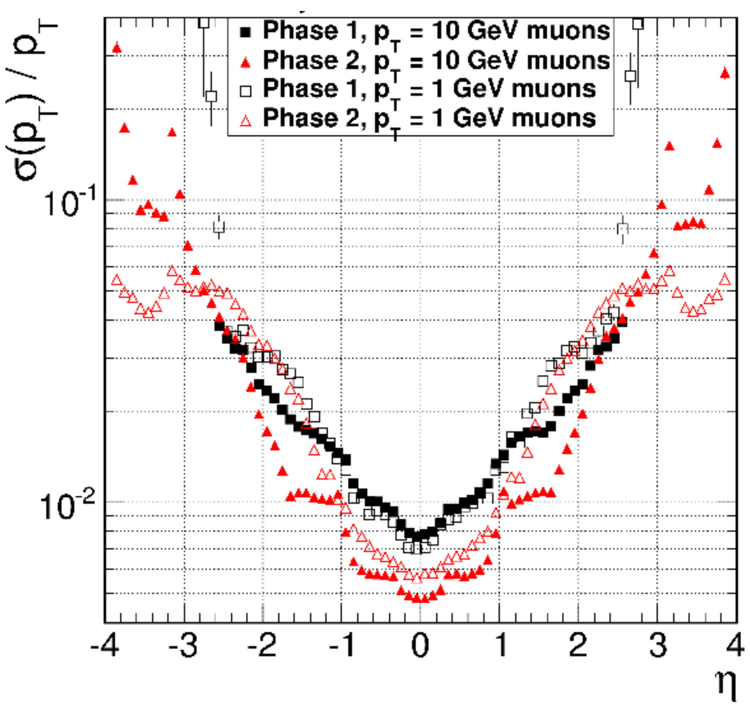
3D reconstruction+timing (calorimeter+timing layers)

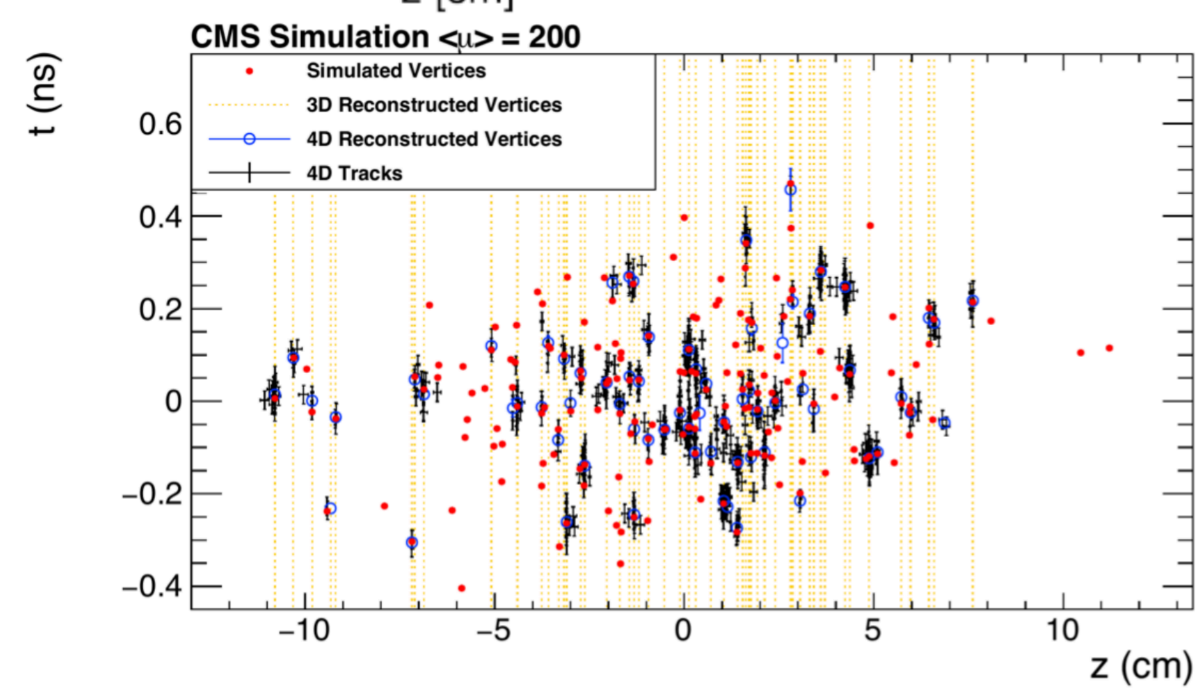
associate energy deposits to vertices



TDR-15-002

<5% resolution
up to η~3





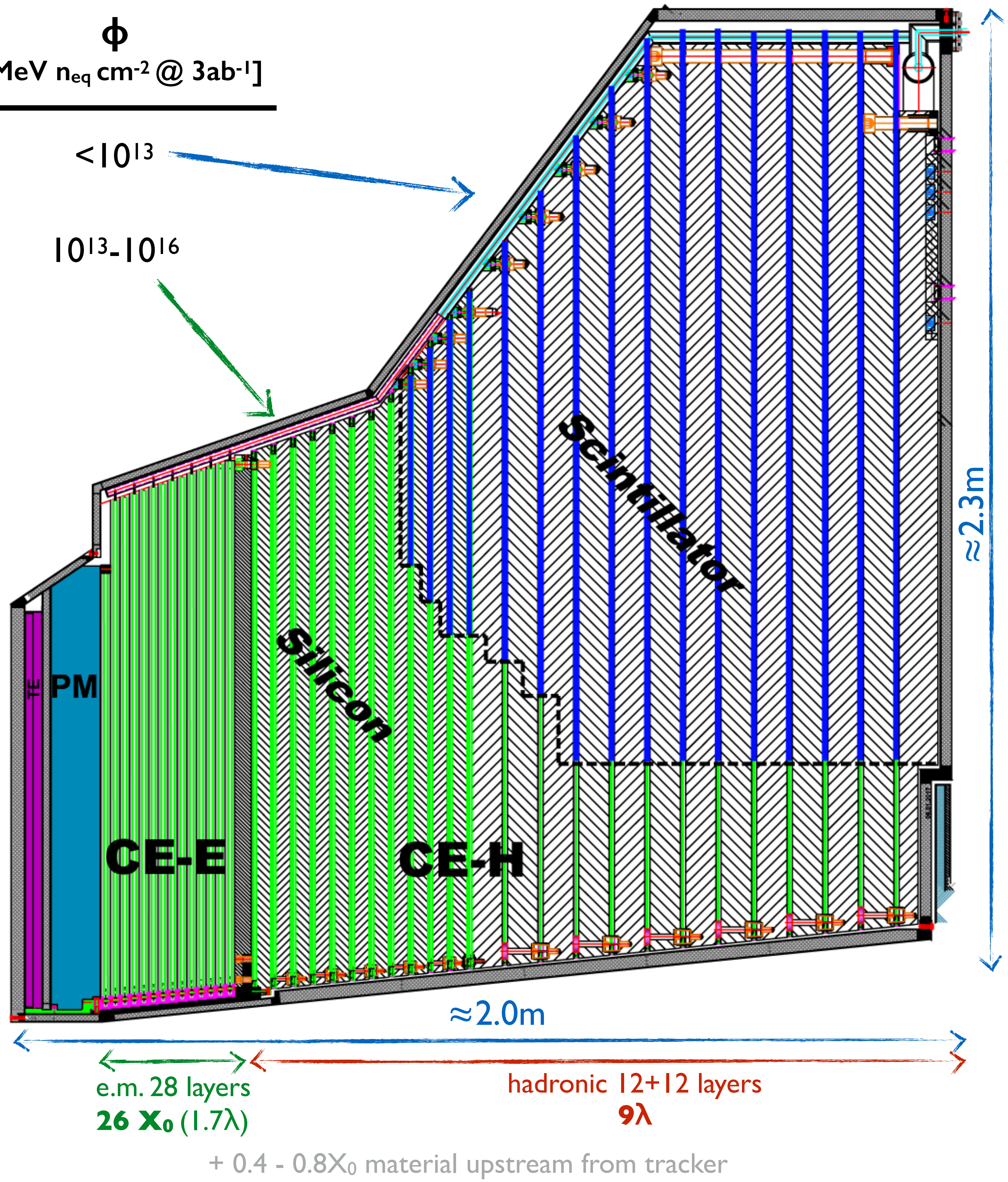
High Granularity CALorimeter Overview

Active elements	Absorbers	ϕ [1MeV n_{eq} cm^{-2} @ $3ab^{-1}$]
Scintillating tiles (SiPM readout)	Stainless steel	$<10^{13}$
Si (hexagonal modules)	Pb/CuW/Cu (CE-E) Stainless steel (CE-H)	10^{13}-10^{16}

Key parameters

- provide $1.5<\eta<3.0$ coverage
- operated at -243°K (-30°C)
- ~600 (500) m^2 of Si (scintillator)
- 6M Si channels, 0.5 or 1.1 cm^2 cell size arranged in 27k Si modules
- readout data from all layers
- trigger from alternate layers in CE-E and all layers in CE-H

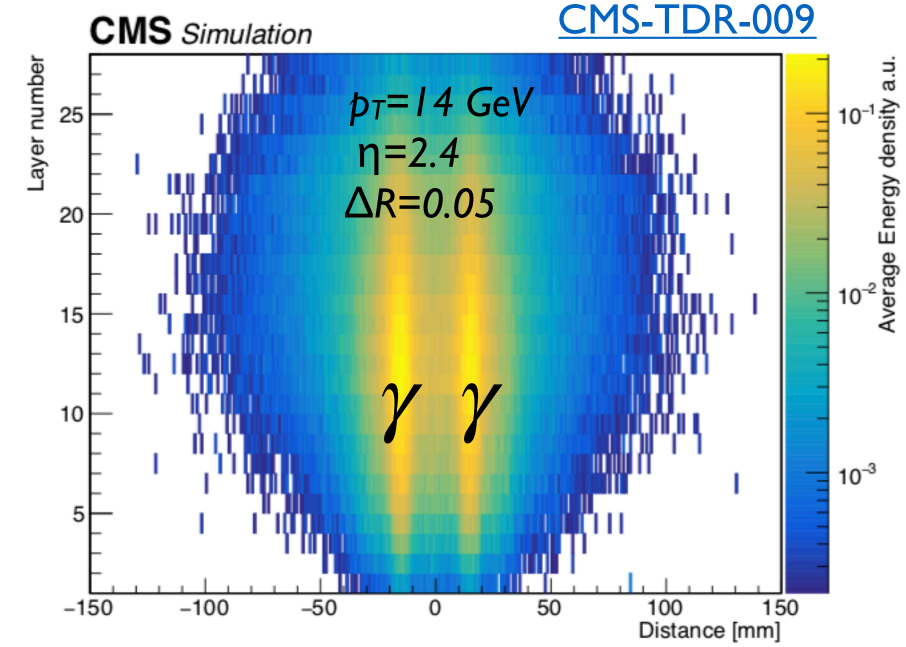
☞ ***status and details in N. Akchurin's talk***



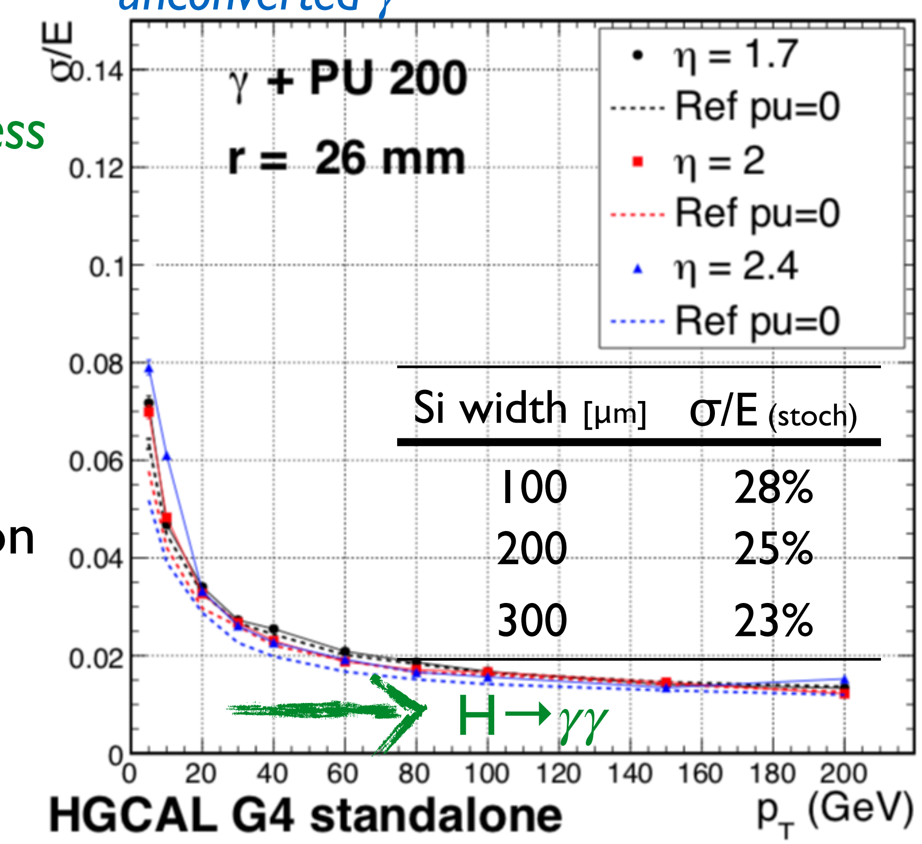
HGCAL and particle flow

- **Benefit from high granularity for feature extraction**
  - *lateral granularity ~1/3-1/4R_M in Si sections*
    good two-shower separation →
    non-pointing but can be used as coarse tracker (δθ~7mrad for γ)
    excellent tracker-calorimeter linking
  - *timing capabilities* (<30ps per cell for S/N>60)
    adds extra dimension for cleaning and clustering algorithms
  - *resolution only partially spoiled by longitudinal sampling/sensor thickness*
    expected maintain competitiveness for physics of interest →
    granularity enables fine-grained pileup subtraction
- **Dense absorbers and thin sensors**
  - MIP sensitivity throughout lifetime allows individual pad calibration
    use isolated track approach with all the pileup π^-
- **Sophisticated software needed!**
  - the results that follow rely on initial studies
  - we are still at the infancy with respect to reconstruction (and some simulation aspects)

CMS-TDR-009



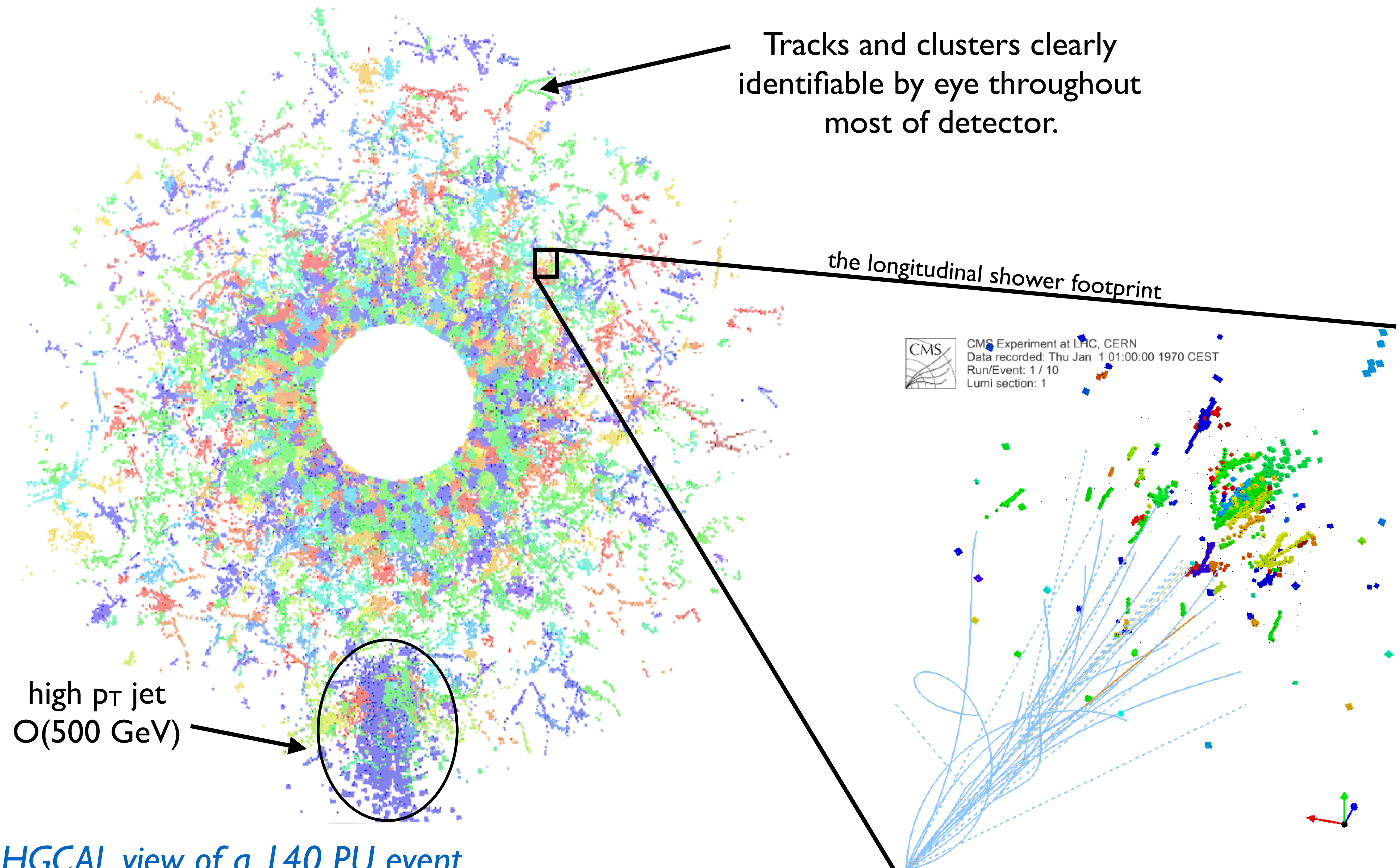
unconverted γ



Si width [μm]	σ/E (stoch)
100	28%
200	25%
300	23%

From visualisation to reconstruction

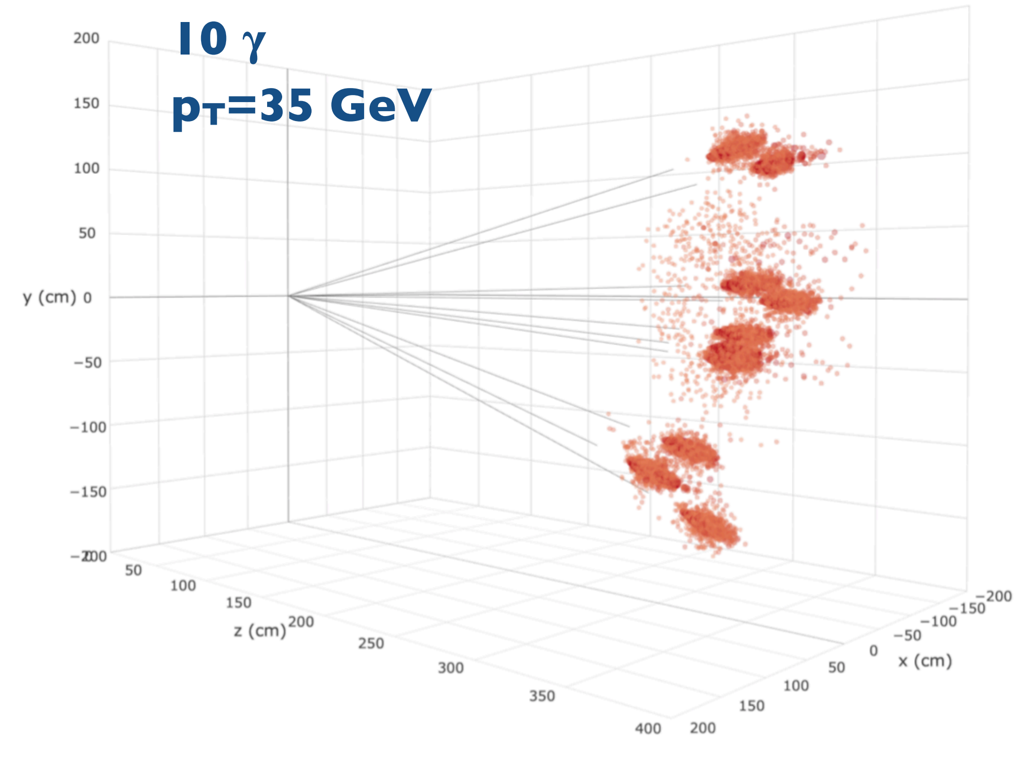
- **Intrinsic potential to visualise individual shower components at high pile-up**
  - how to reconstruct and identify showers offline? ☞ for trigger aspects see T. Strebler



HGCAL view of a 140 PU event

Current approach to clustering in HGCAL

- **Partition reconstruction space k-d tree approach**
  doi:10.1145/361002.361007
  - k-d tree creation per event has O(N logN) complexity
  - but nearest-neighbour search complexity O(N) → O(log N)

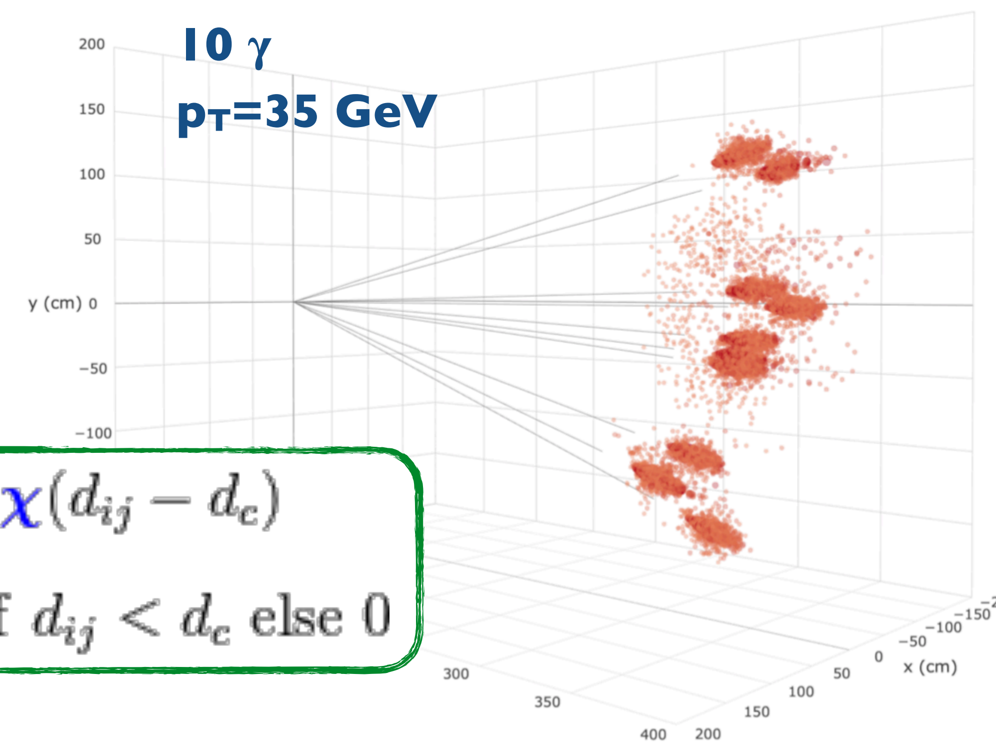


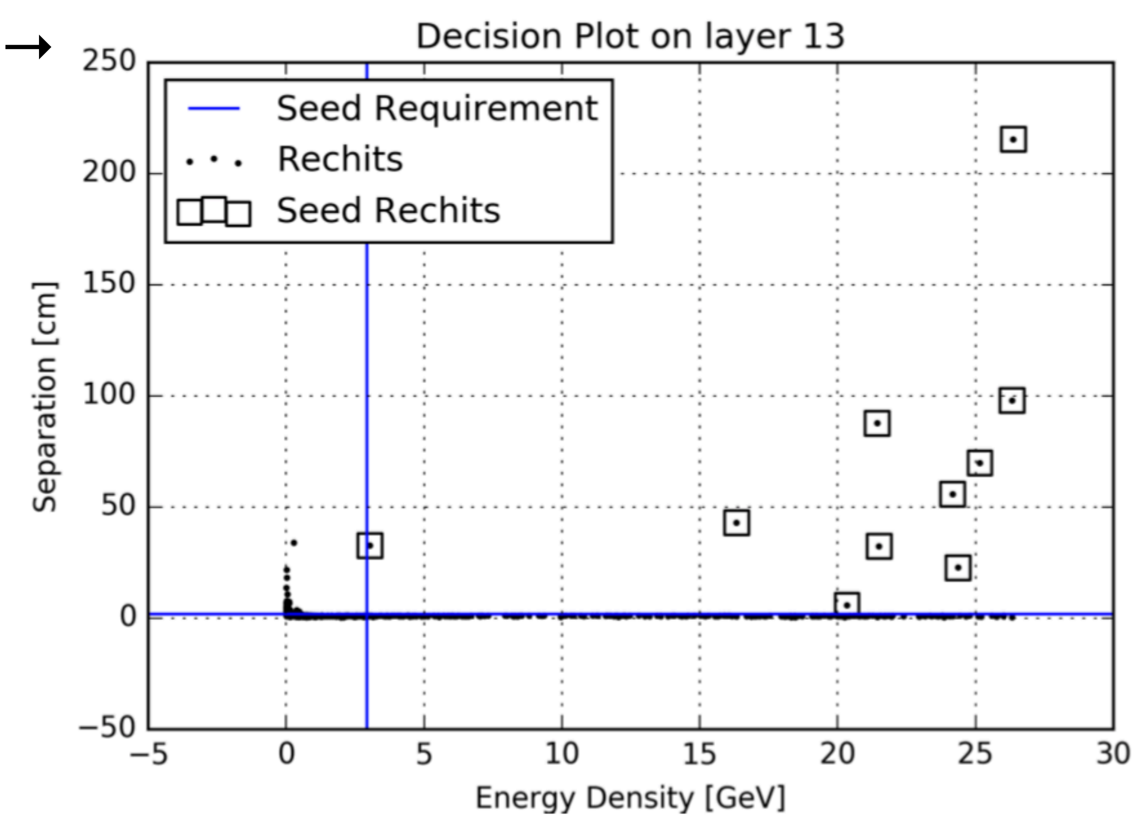
Current approach to clustering in HGCAL

- **Partition reconstruction space k-d tree approach**
  doi:10.1145/361002.361007
  - k-d tree creation per event has O(N logN) complexity
  - but nearest-neighbour search complexity O(N) → O(log N)
- **2D clustering proceeds on a layer-by-layer basis**
  - compute local energy density per point based on

$$\rho_i = \sum_j \chi(d_{ij} - d_c)$$
$$\chi = E_j \text{ if } d_{ij} < d_c \text{ else } 0$$

  - evaluate distance to closest cell with higher density
  - **seed cells** trivially found in the separation-density plane →
  - cluster neighbouring cells within d_c
    (2cm for EE/FH, 5cm for BH)



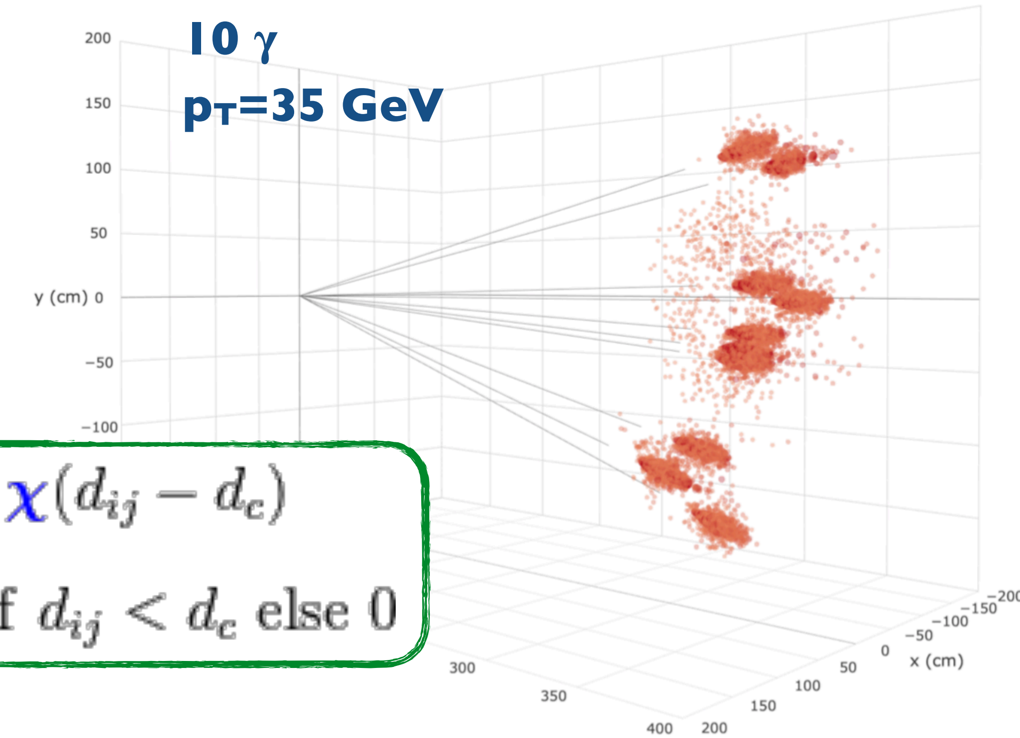


CMS-CR-2017-048

DOI: 10.1126/science.1242072

Current approach to clustering in HGCAL

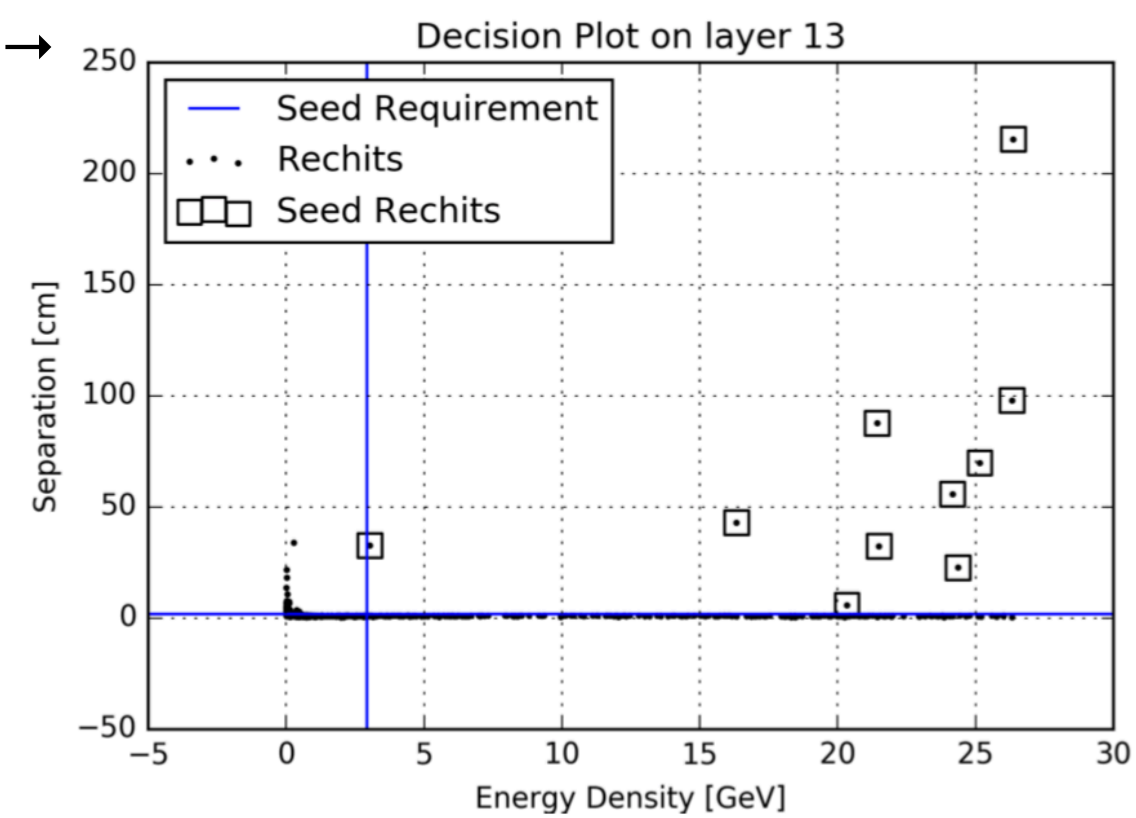
- **Partition reconstruction space k-d tree approach**
  doi:10.1145/361002.361007
  - k-d tree creation per event has O(N logN) complexity
  - but nearest-neighbour search complexity O(N) → O(log N)



- **2D clustering proceeds on a layer-by-layer basis**
  - compute local energy density per point based on

$$\rho_i = \sum_j \chi(d_{ij} - d_c)$$
$$\chi = E_j \text{ if } d_{ij} < d_c \text{ else } 0$$

  - evaluate distance to closest cell with higher density
  - **seed cells** trivially found in the separation-density plane →
  - cluster neighbouring cells within d_c
    (2cm for EE/FH, 5cm for BH)

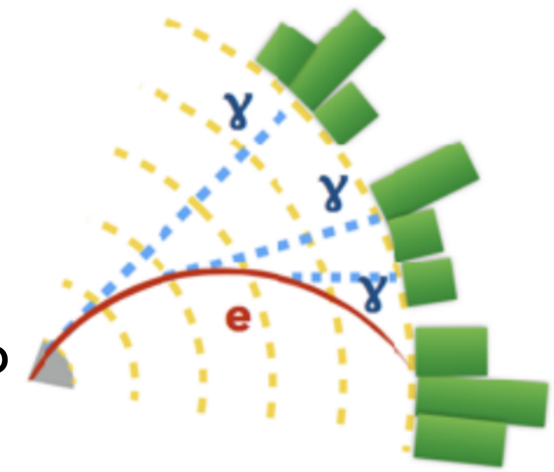


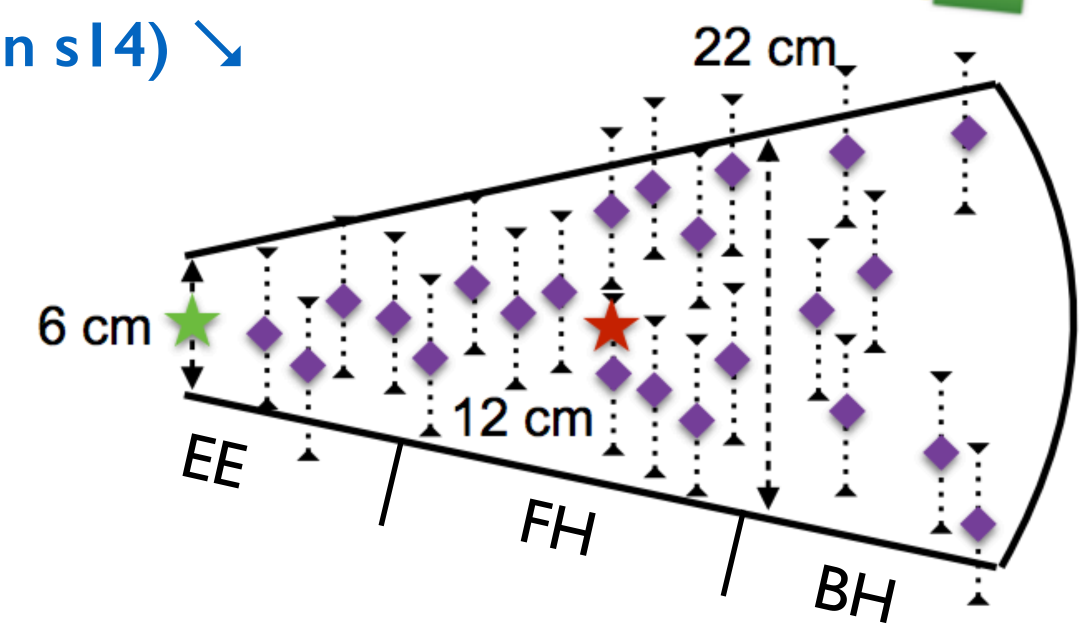
CMS-CR-2017-048
DOI: 10.1126/science.1242072

- **Simple, fast and robust algorithm**
  - energy densities @ 140PU: 0.4 s/event
    (73% CPU time reduction wrt to brute force search)
  - O(2.5k) clusters at e.m. shower max for PU=200
  - suitable for future implementation in GPUs ⇒ HLT?

Linking consecutive layers: multiclusters

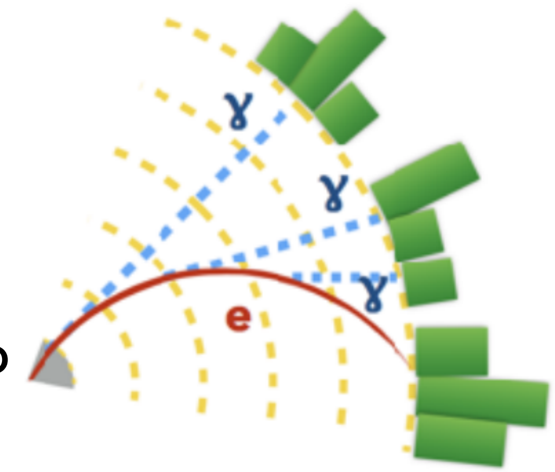
- **Geometric approach based on straight line propagation from centre of detector**
- **e.m. showers**
  - linked within the first 28 layers with a 2cm alignment requirement
  - gather additional bremsstrahlung emitted upfront to HGCAL with a road in ϕ
- **Hadronic showers: mega-cluster approach (more on s14) ↘**



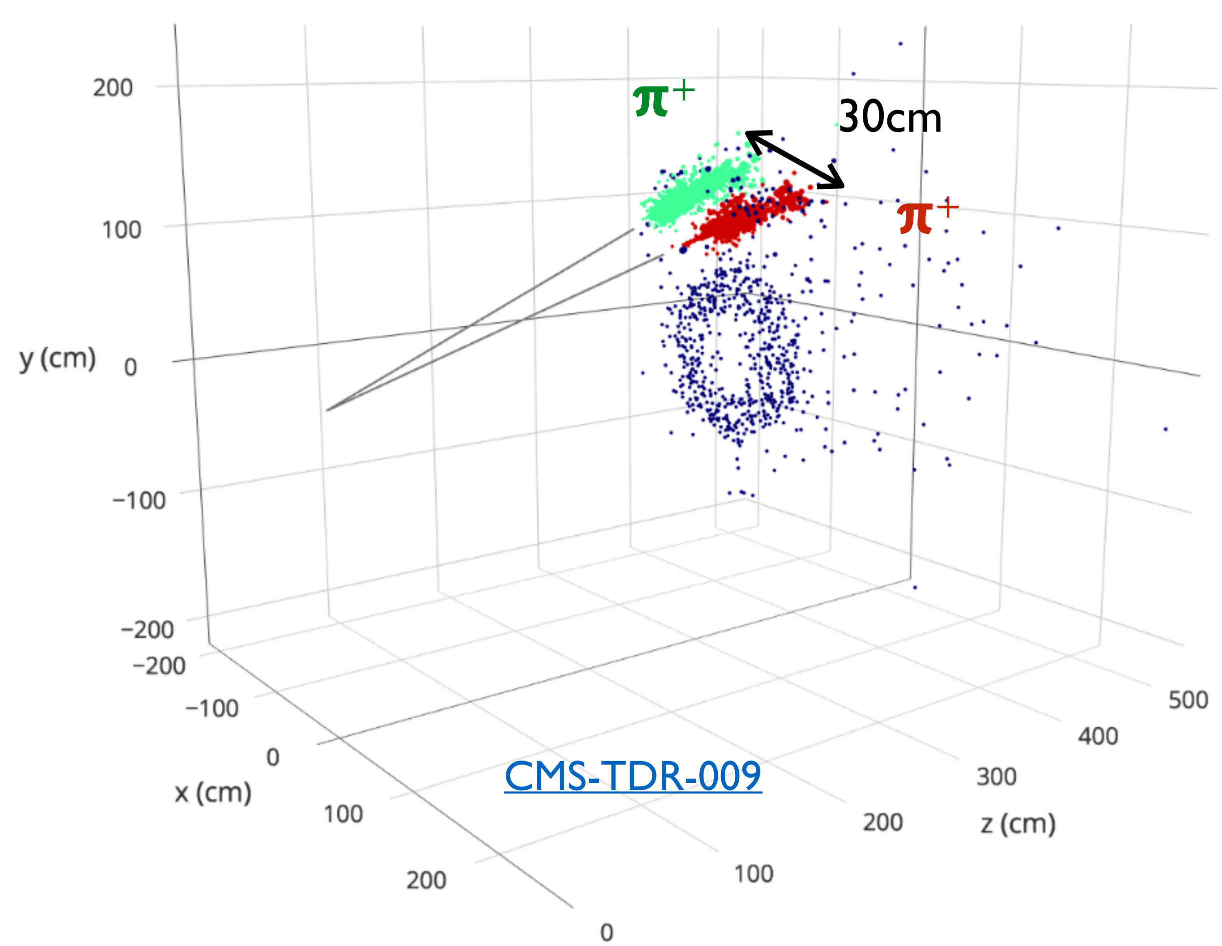


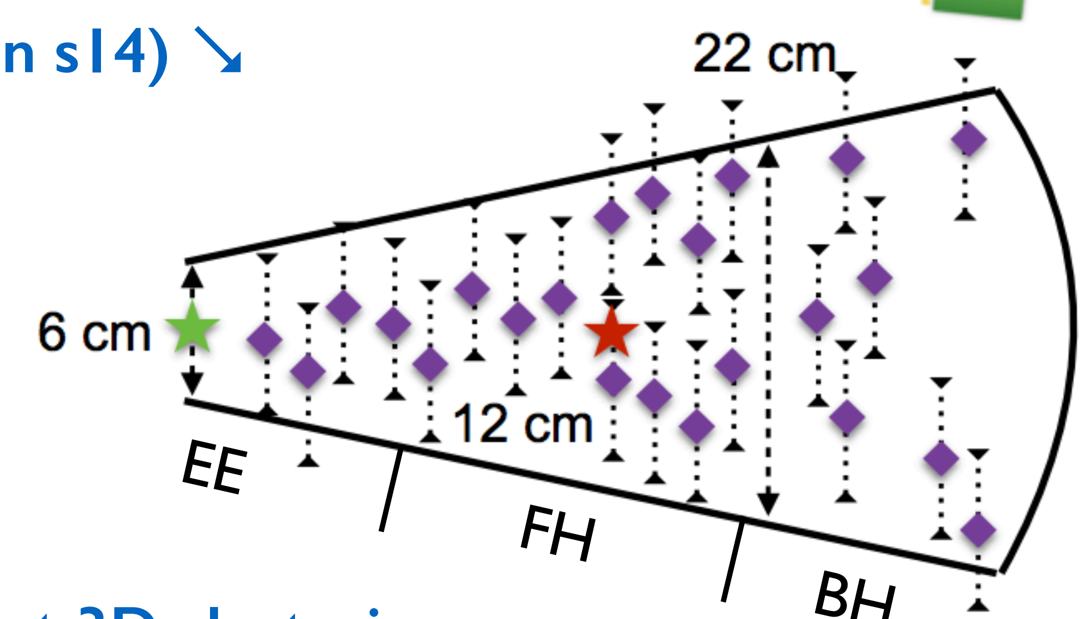
Linking consecutive layers: multiclusters

- Geometric approach based on straight line propagation from centre of detector
- e.m. showers
  - linked within the first 28 layers with a 2cm alignment requirement
  - gather additional bremsstrahlung emitted upfront to HGCAL with a road in ϕ



- Hadronic showers: mega-cluster approach (more on s14) ↘

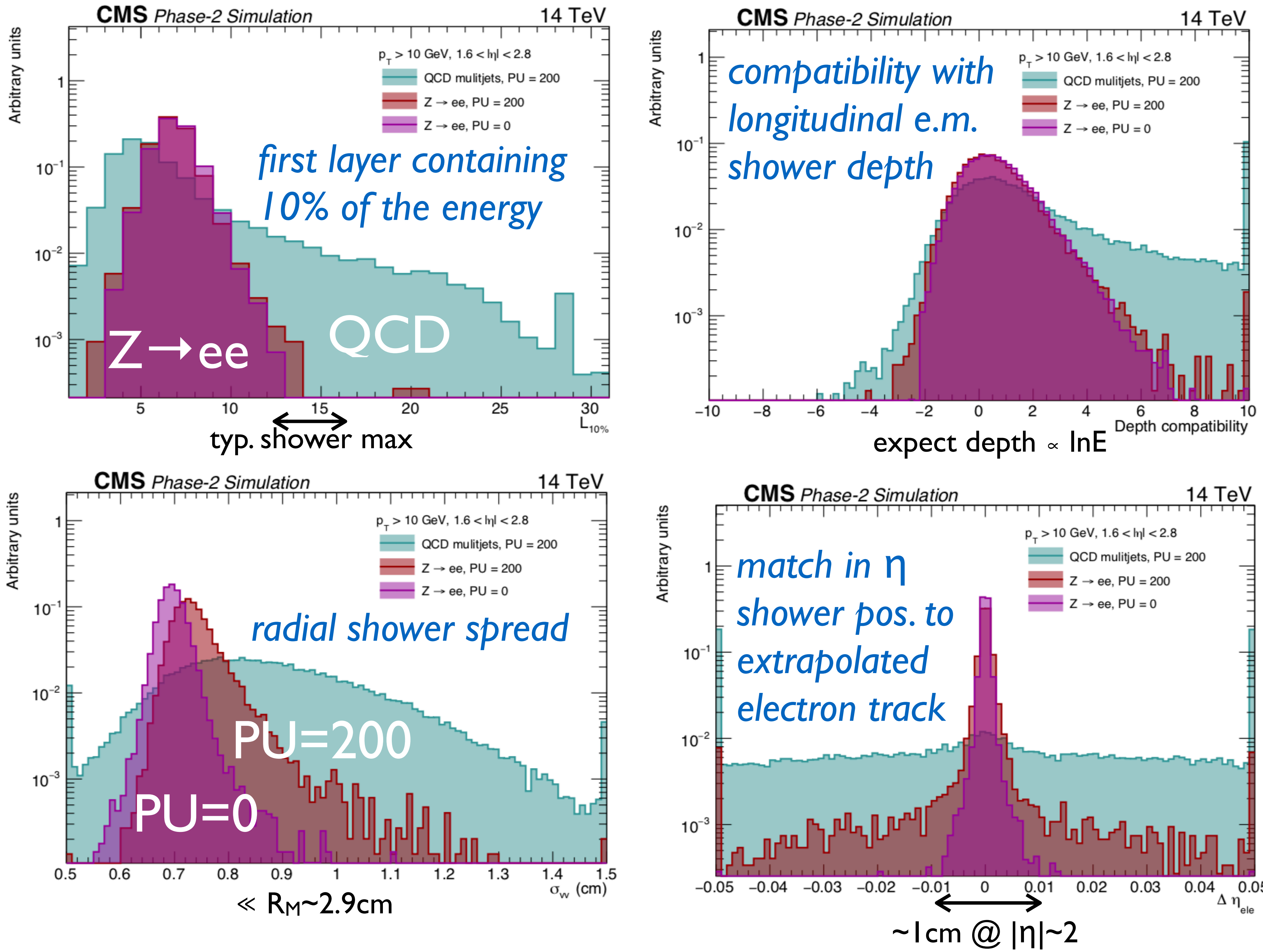




← Direct 3D clustering

- simple extension to a 3D kernel yield already meaningful results
- but further work is needed - e.g. accommodate for different sampling dE/dx, optimise metric in use, etc.

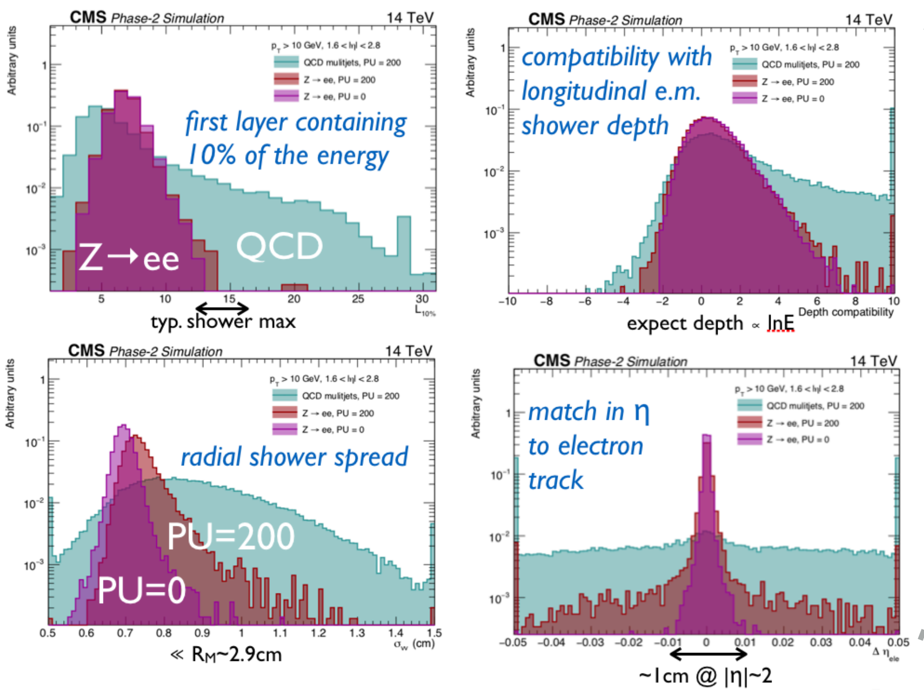
e/γ performance: identification

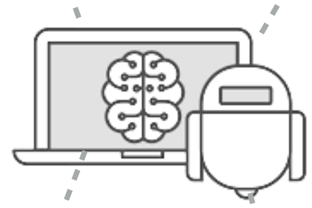


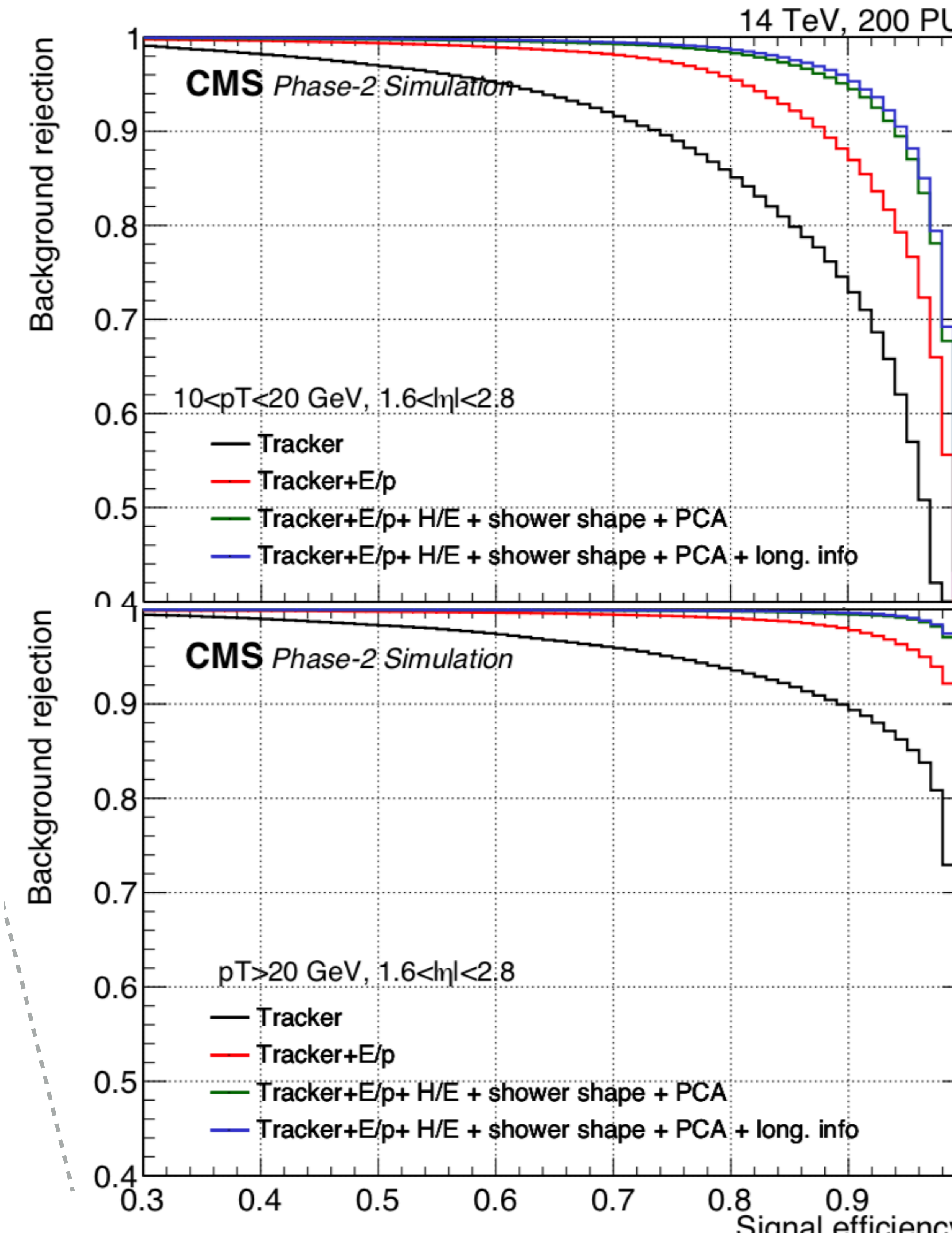
CMS-TDR-009

Longitudinal+lateral granularity provide unique handles for identification

e/γ performance: identification





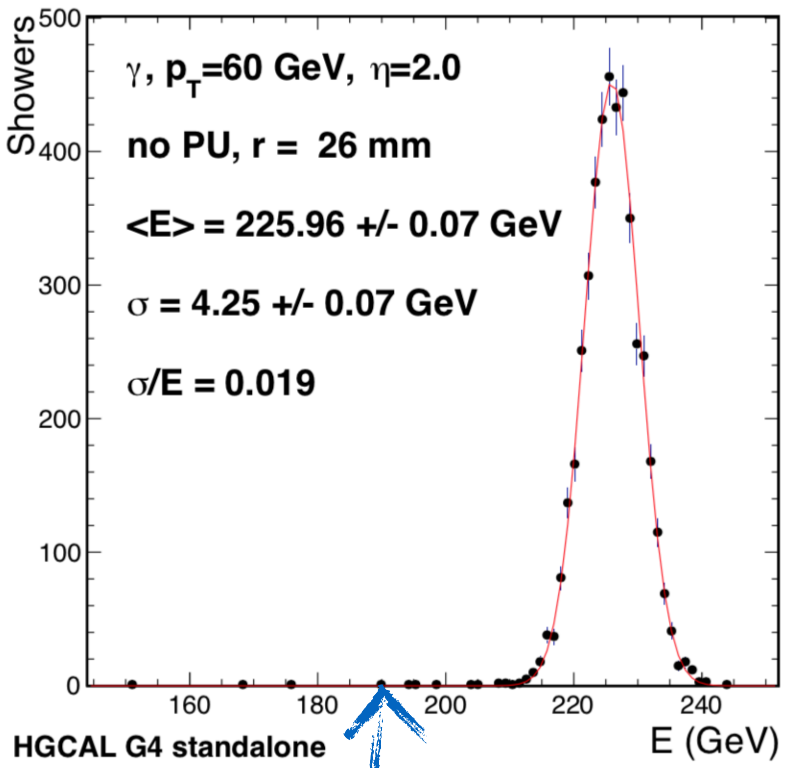


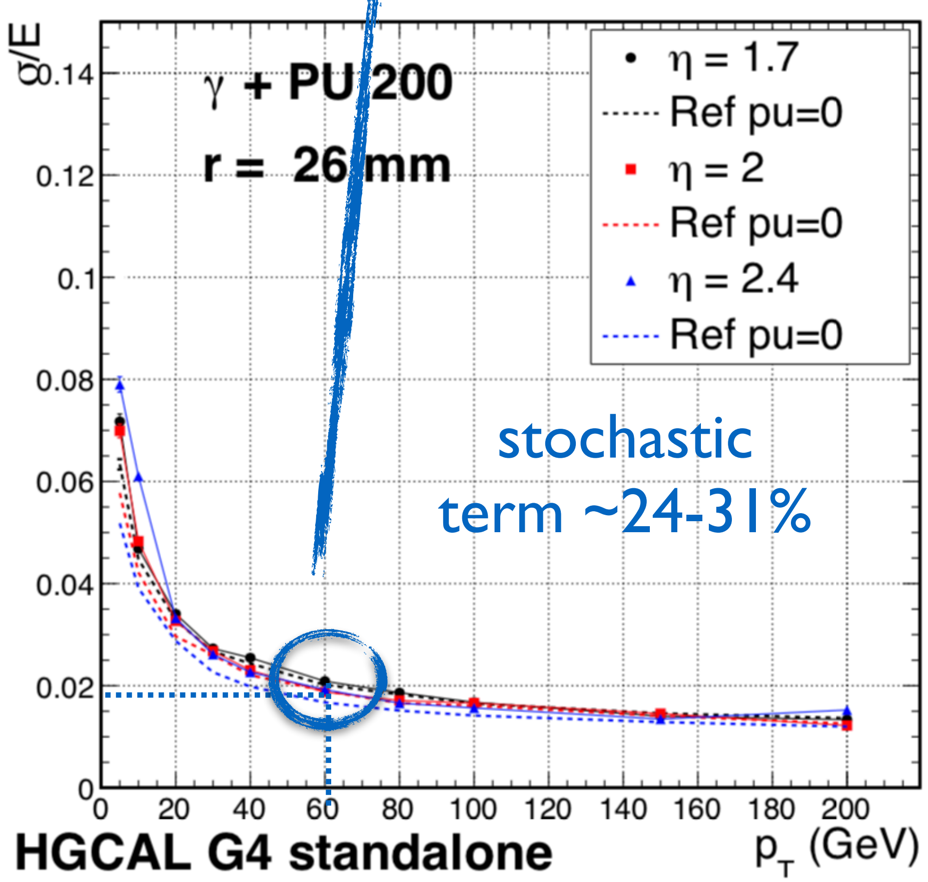
CMS-TDR-009

- **Multivariate analysis for ultimate performance**
- **Initial studies indicate**
  - large gain from shower shapes + principal component analysis (PCA)
  - longitudinal information improving background rejection for p_T<20 GeV by 2-5%

e/γ performance: energy resolution

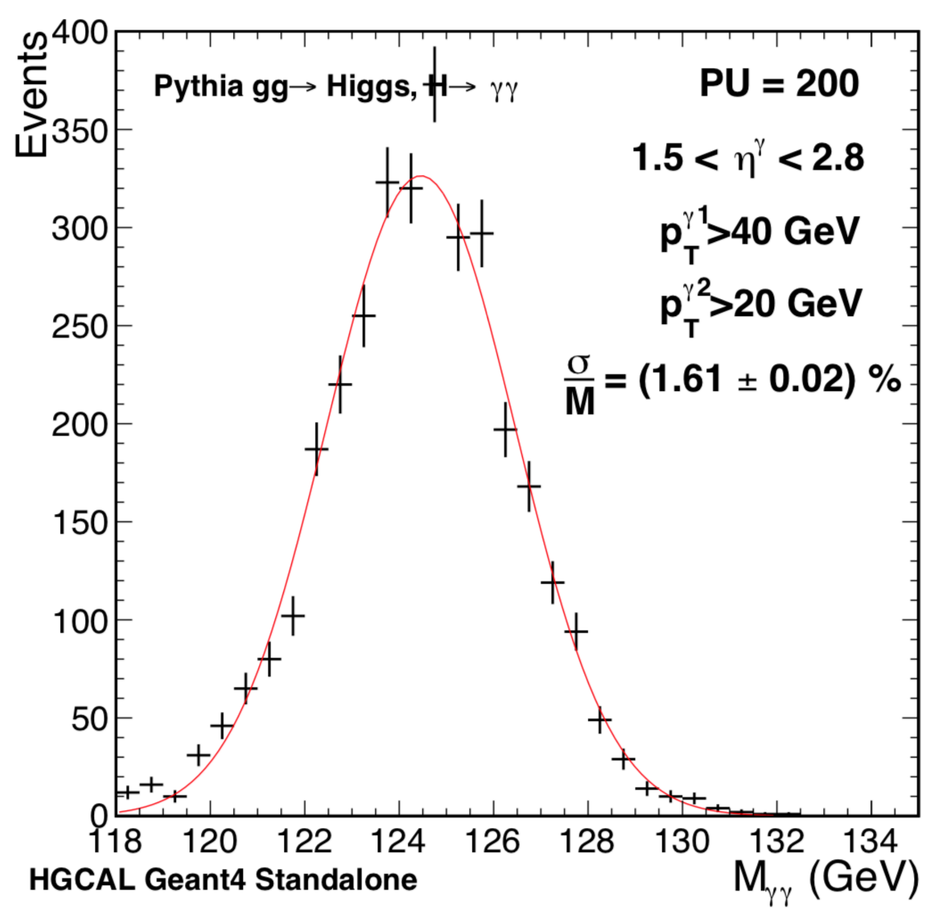
CMS-TDR-009





- **For energy reconstruction may not use all hits**
  - simplified photon reconstruction using fixed radius (26mm around most energetic cell)
  - pileup subtraction based on random cone in detector
  - leakage parametrised as function of energy in last 4 layers (F_4)

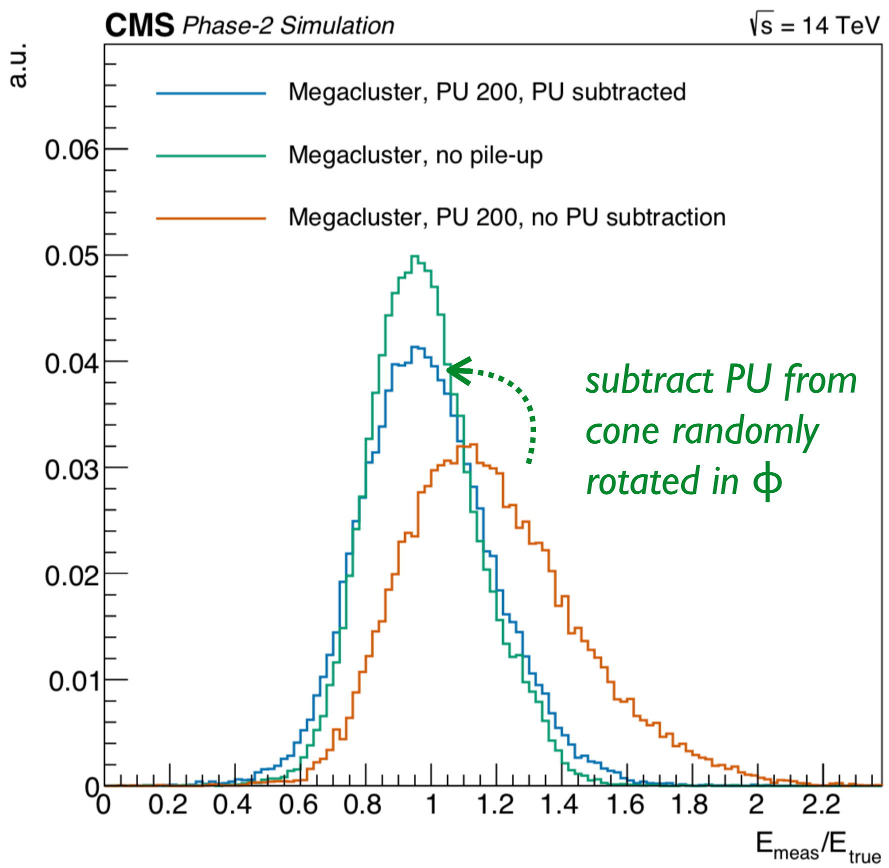
(typically F_4<5%, $\Delta E/E$~7% for F_4~15%)



Pileup-resilient energy estimation!

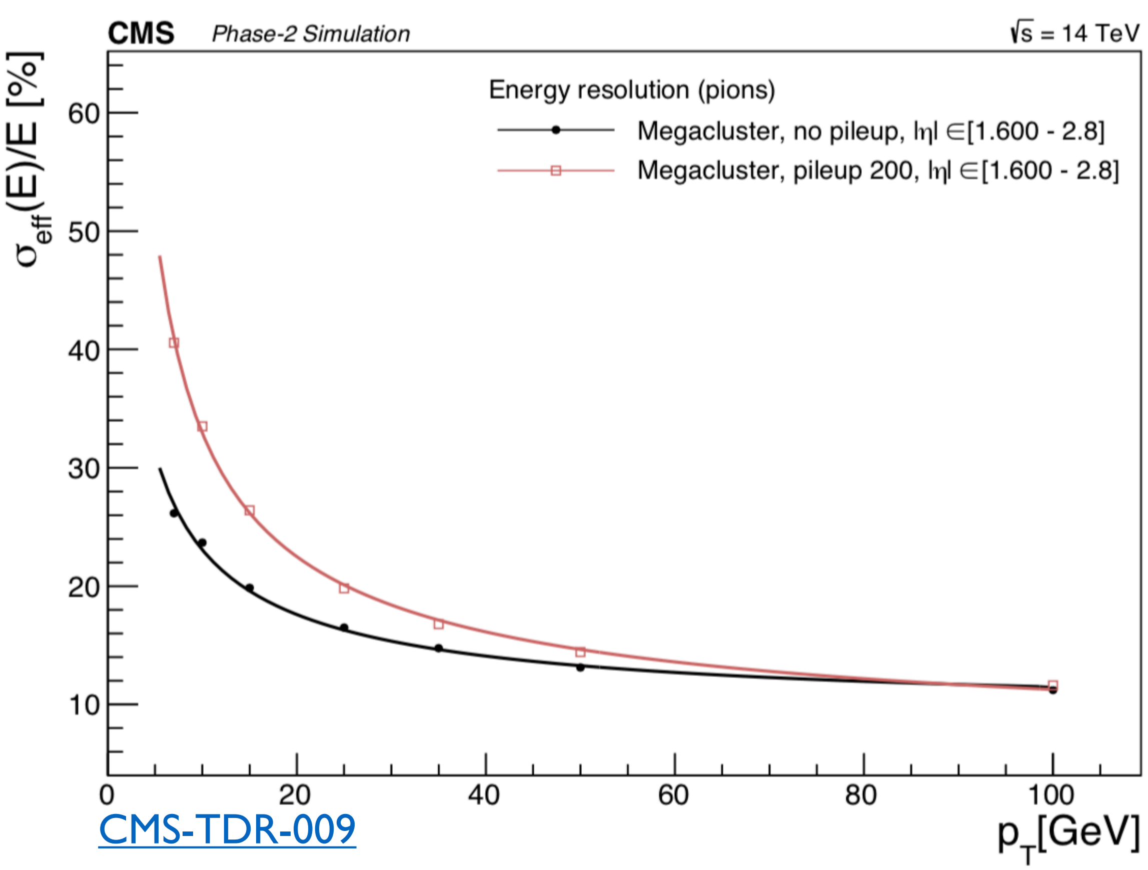
Resolution assuming vertex position comparable to Run I

Performance for single hadrons

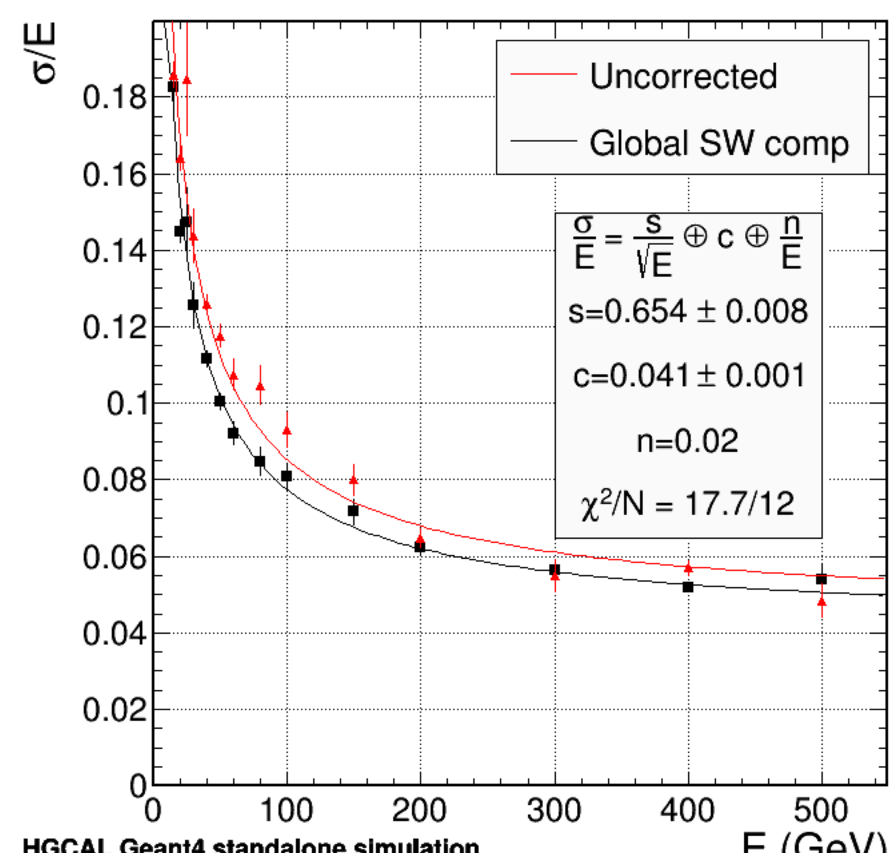


- **"Megaclustering" approach as a robust first approach**
  - pileup-robust estimation of energy for p_T>35 GeV but saturated at σ_E/E~12% constant term
  - NB: HGCAL is non-compensating with non-linear π/e ⇒ corrections of this type not applied in this study …

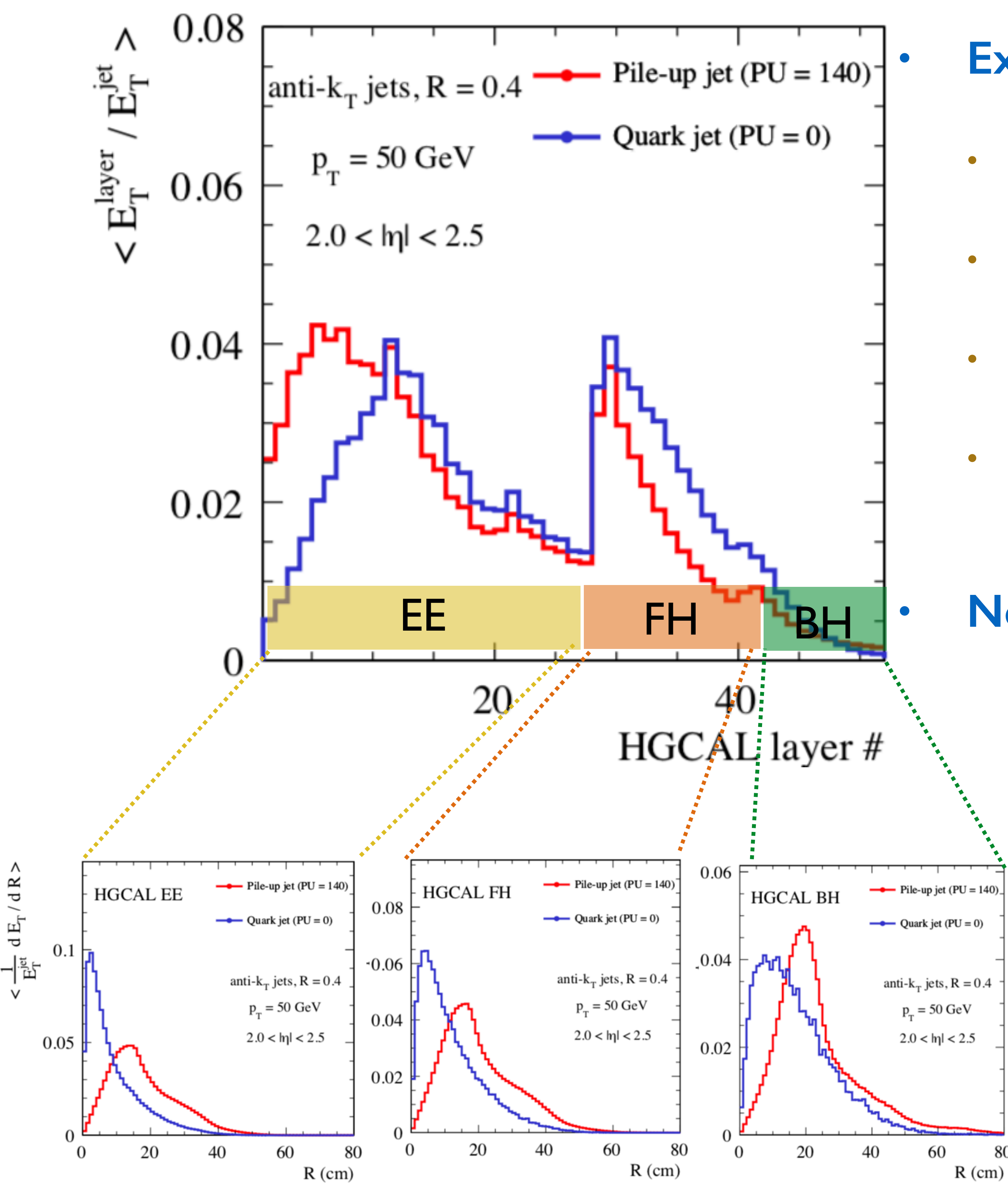
    …but potential is there from early studies ↓



CMS-TDR-009

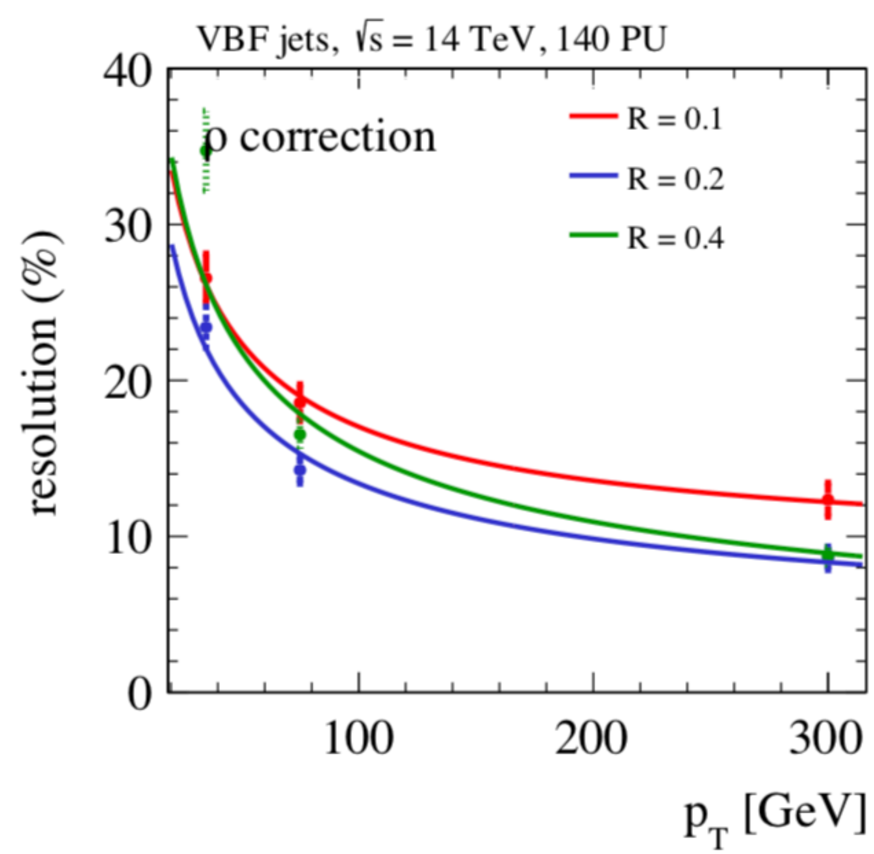


Performance for jets



radial profiles in the different sections

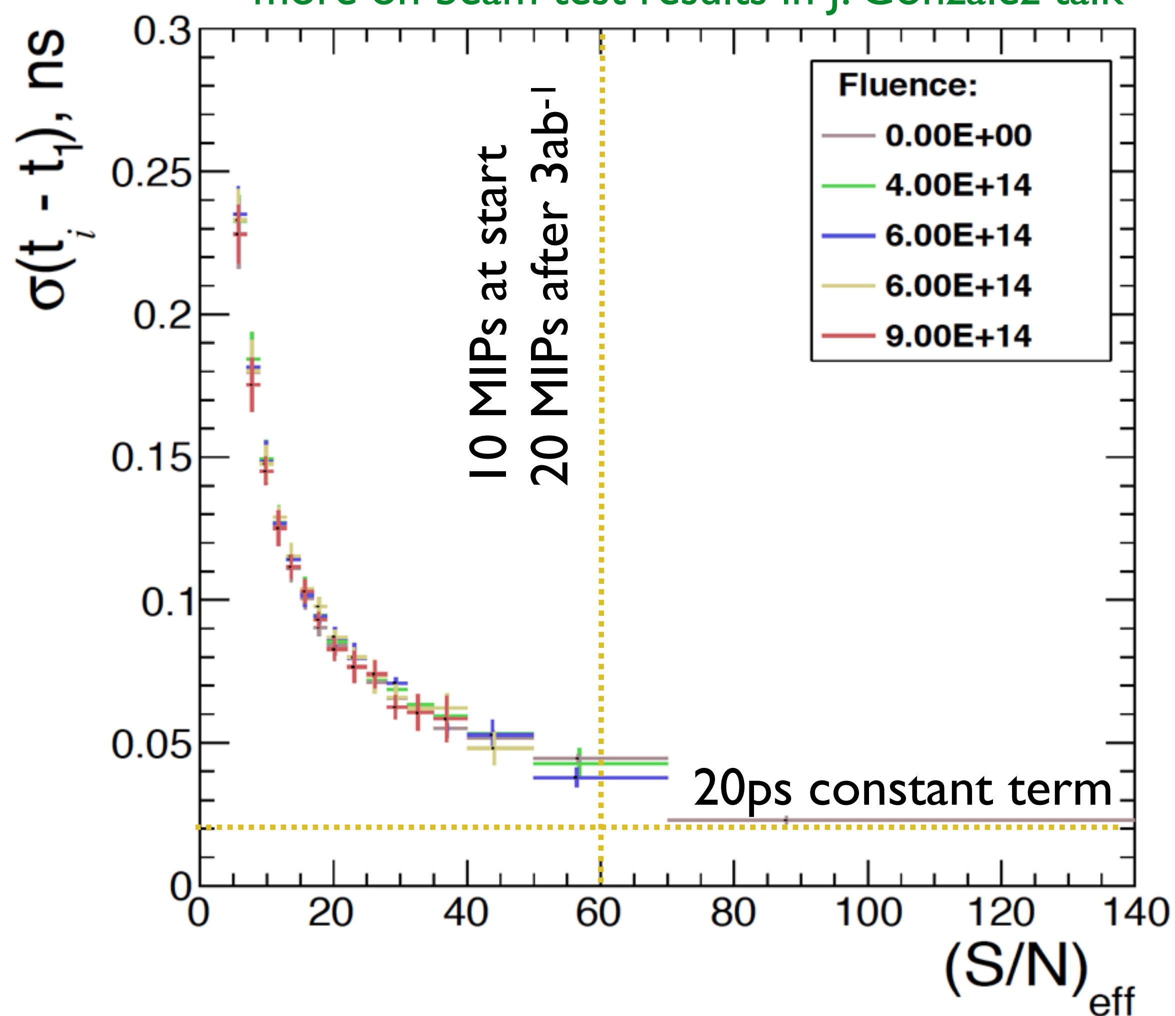
- **Exercise clustering RecHits with anti-k_T algorithm**
  - distinct longitudinal and transverse profiles wrt to PU
  - dynamical definition of jet axis is possible
  - ingredients for future PU jet id and q/g discrimination
  - allow to keep VBF L1 rate below 10 kHZ @ 80% efficiency using jets with p_T>25 GeV
- **No full reconstruction ⇒ non-optimal resolution**



CMS-TDR-009

Precision timing: towards 5D particle flow…

☞more on beam test results in J. Gonzalez talk

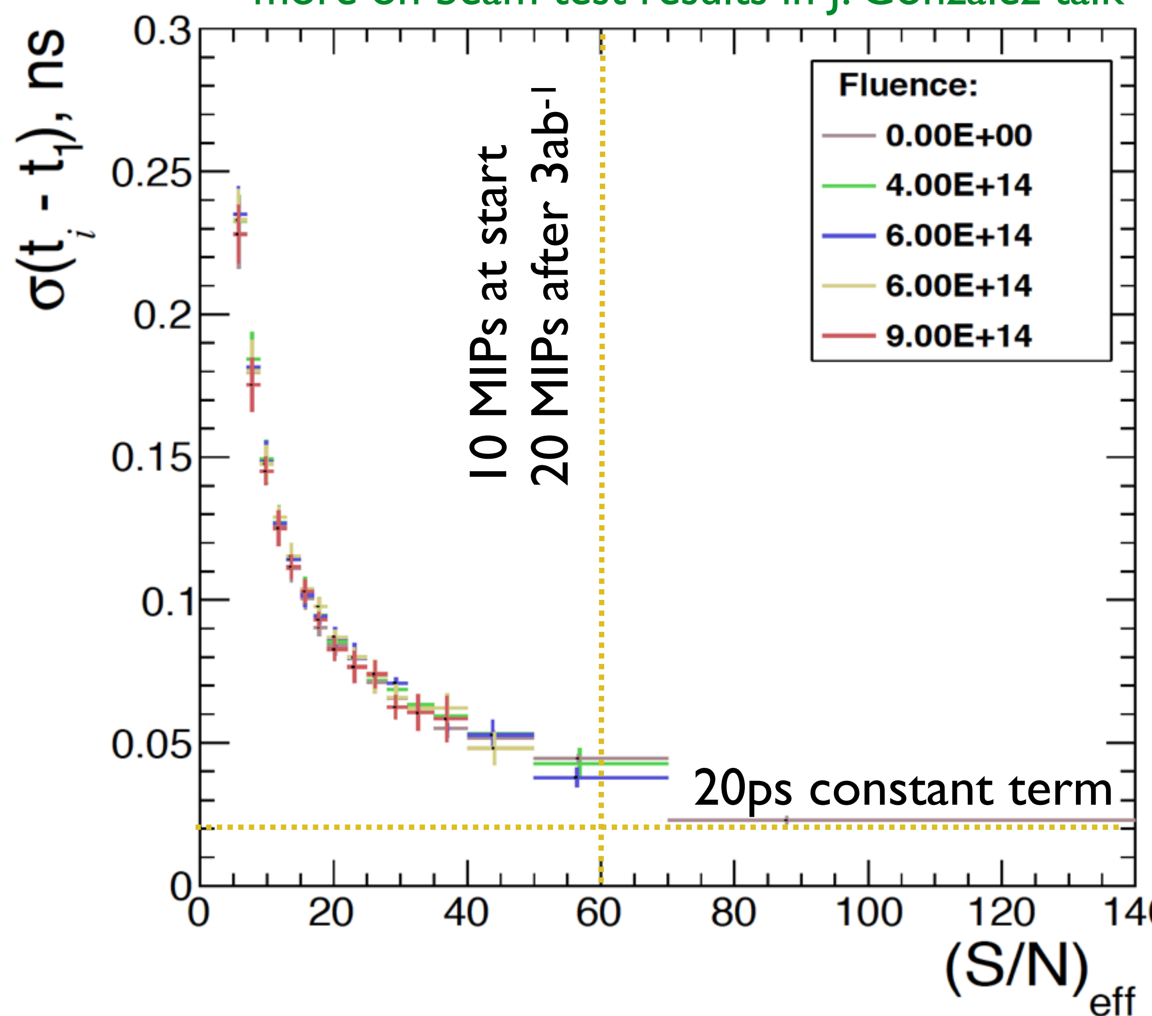


- **Good timing resolution for Si**
  - maintained by electronics, ~20ps floor assumed
  - measurement for several cells in a shower

p_T=5 GeV	ε [%]	δt [ps]
γ	100%	10-15
K_L^0	95%	20-30

Precision timing: towards 5D particle flow…

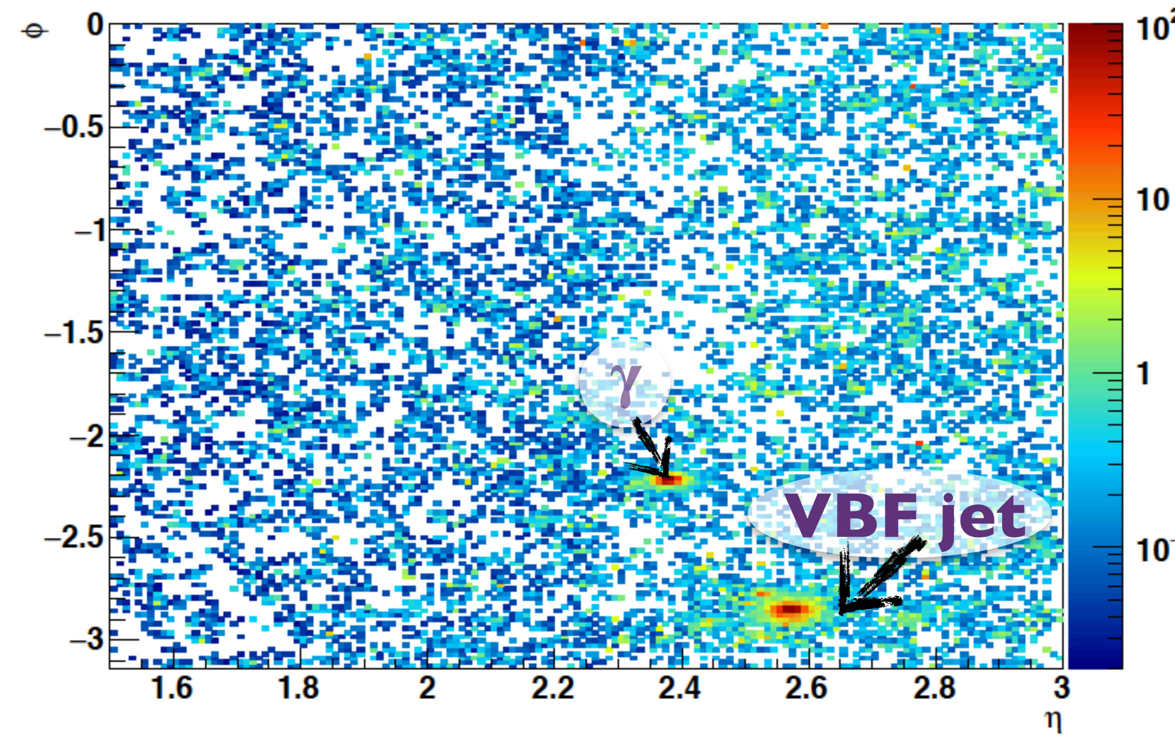
☞more on beam test results in J. Gonzalez talk



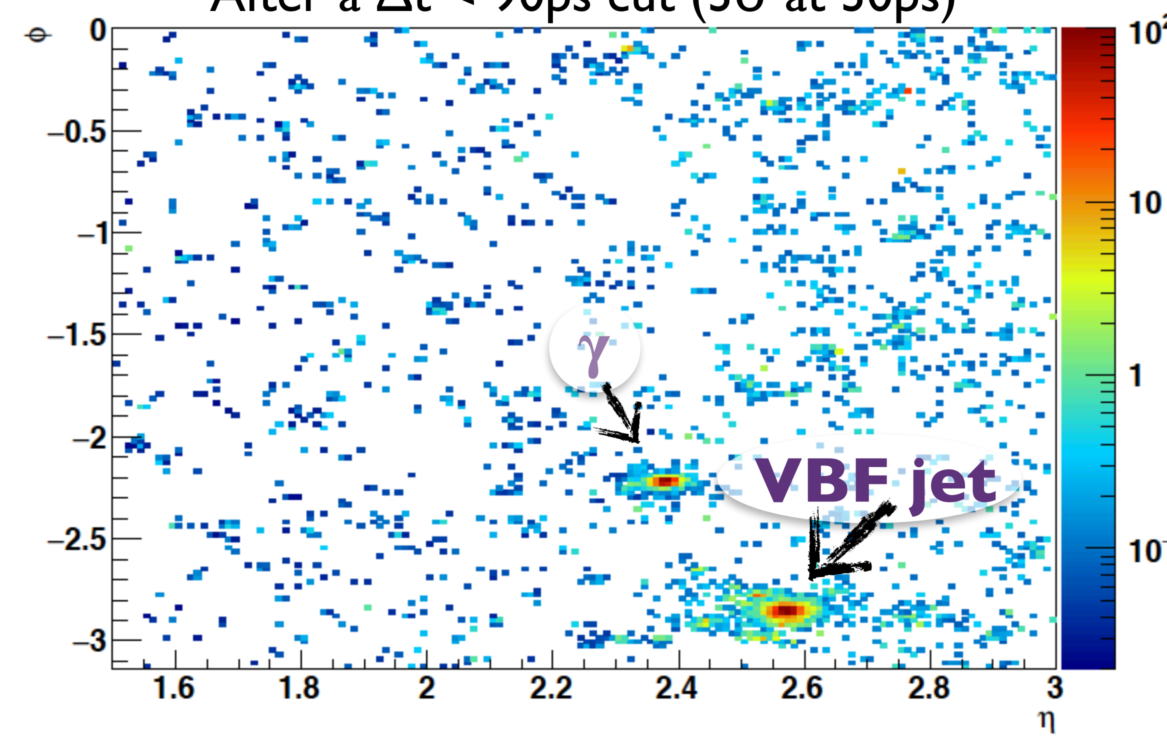
- **Good timing resolution for Si**
  - maintained by electronics, ~20ps floor assumed
  - measurement for several cells in a shower

p_T=5 GeV	ε [%]	δt [ps]
γ	100%	10-15
K_L^0	95%	20-30

All cells with q>12fC (~3.5MIP @ 300um) threshold for time measurements calorimeter front-face projection



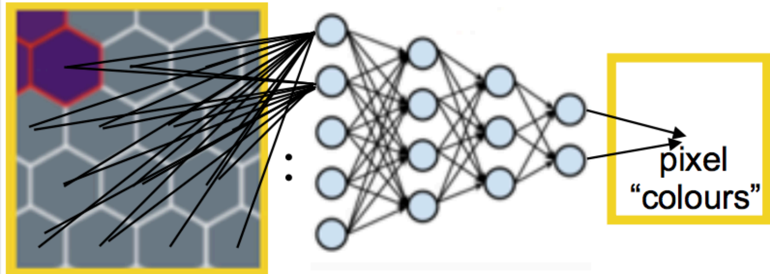
After a Δt < 90ps cut (3σ at 30ps)

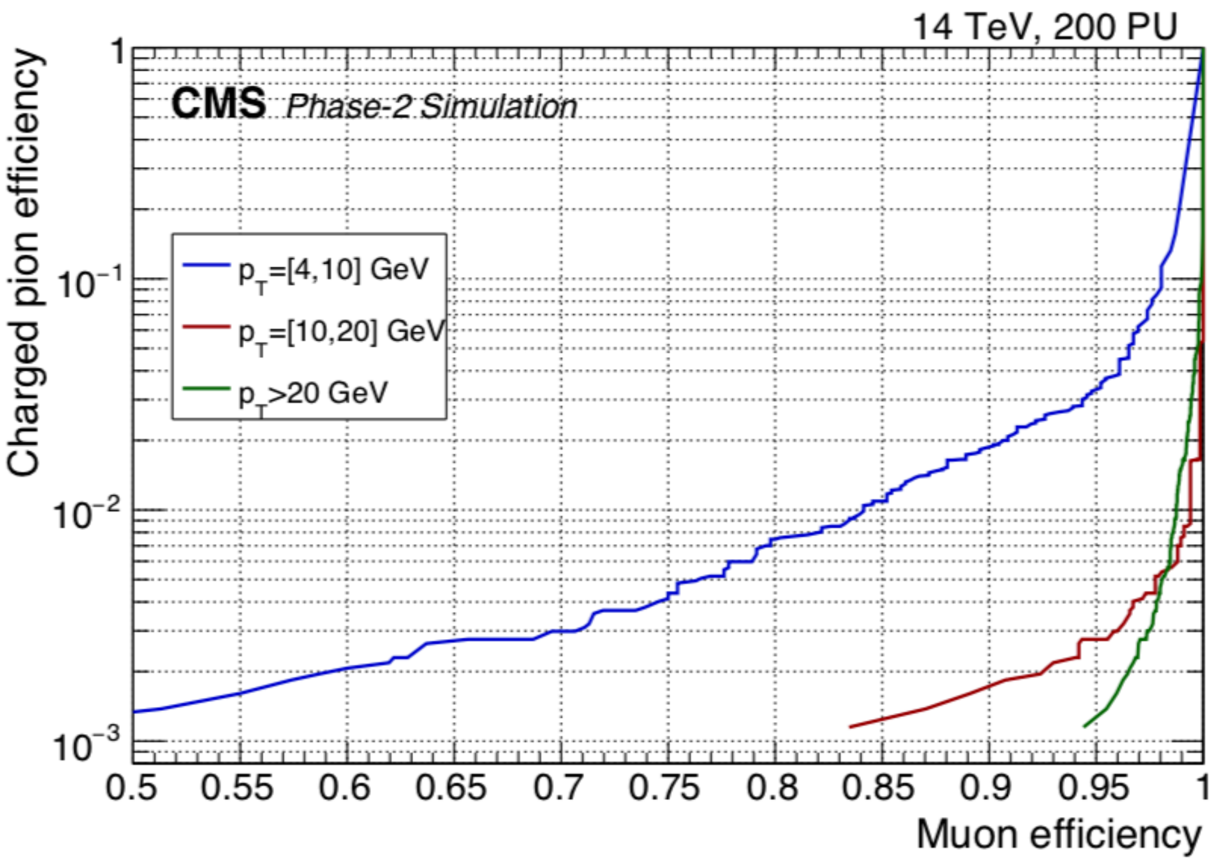


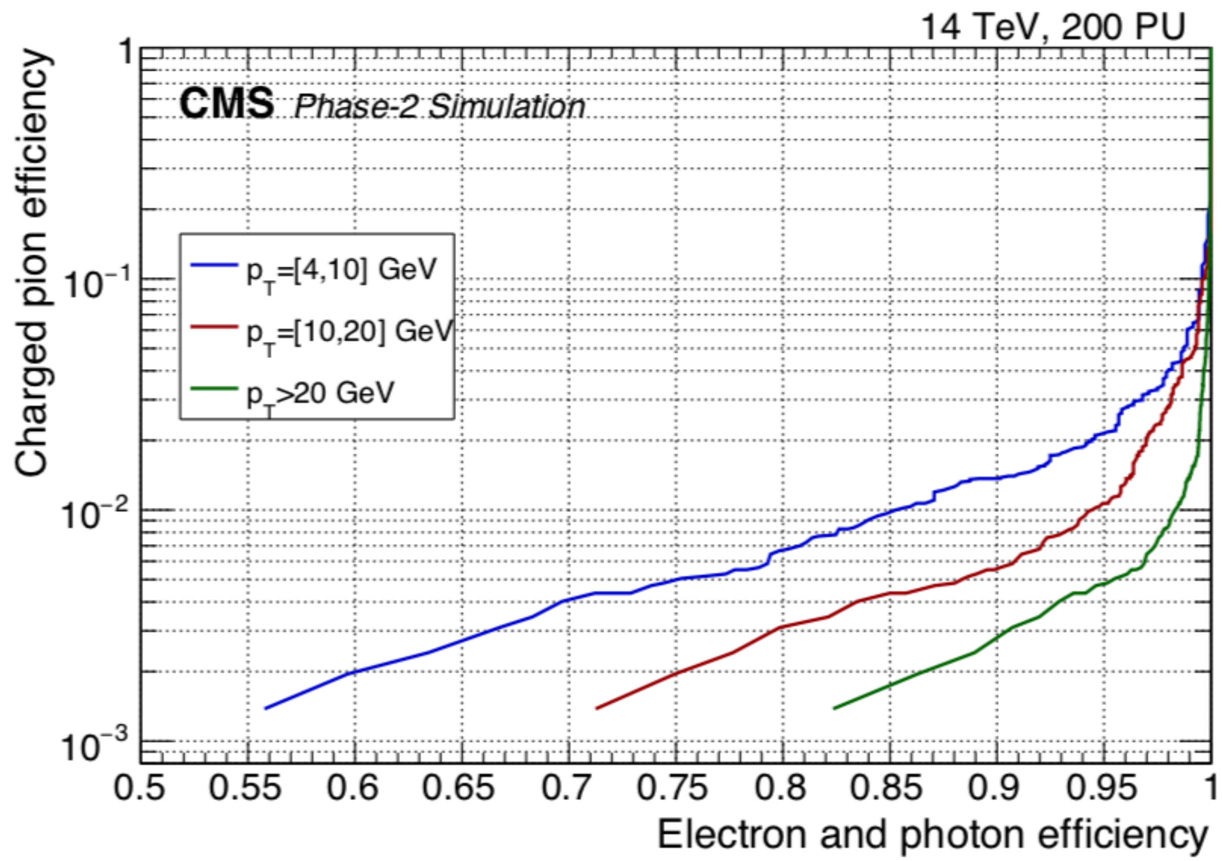
…to be integrated in clustering/PF algorithm

…and beyond (the machine learning era)

- **Imaging calorimetry goes hand-in-hand with pattern recognition**
  - playground to deploy machine learning algorithms
- **First promising results using convolutional neural networks**
  - coarse pixel-isation (up to 6 cells per pixel)
  - energy, time processed by shallow 32x12 DNN (relative positions accounted for) ↑
  - 3D image fed to two parallel 3D CNNs (one for particles that shower, the other for muons)







CMS-TDR-009

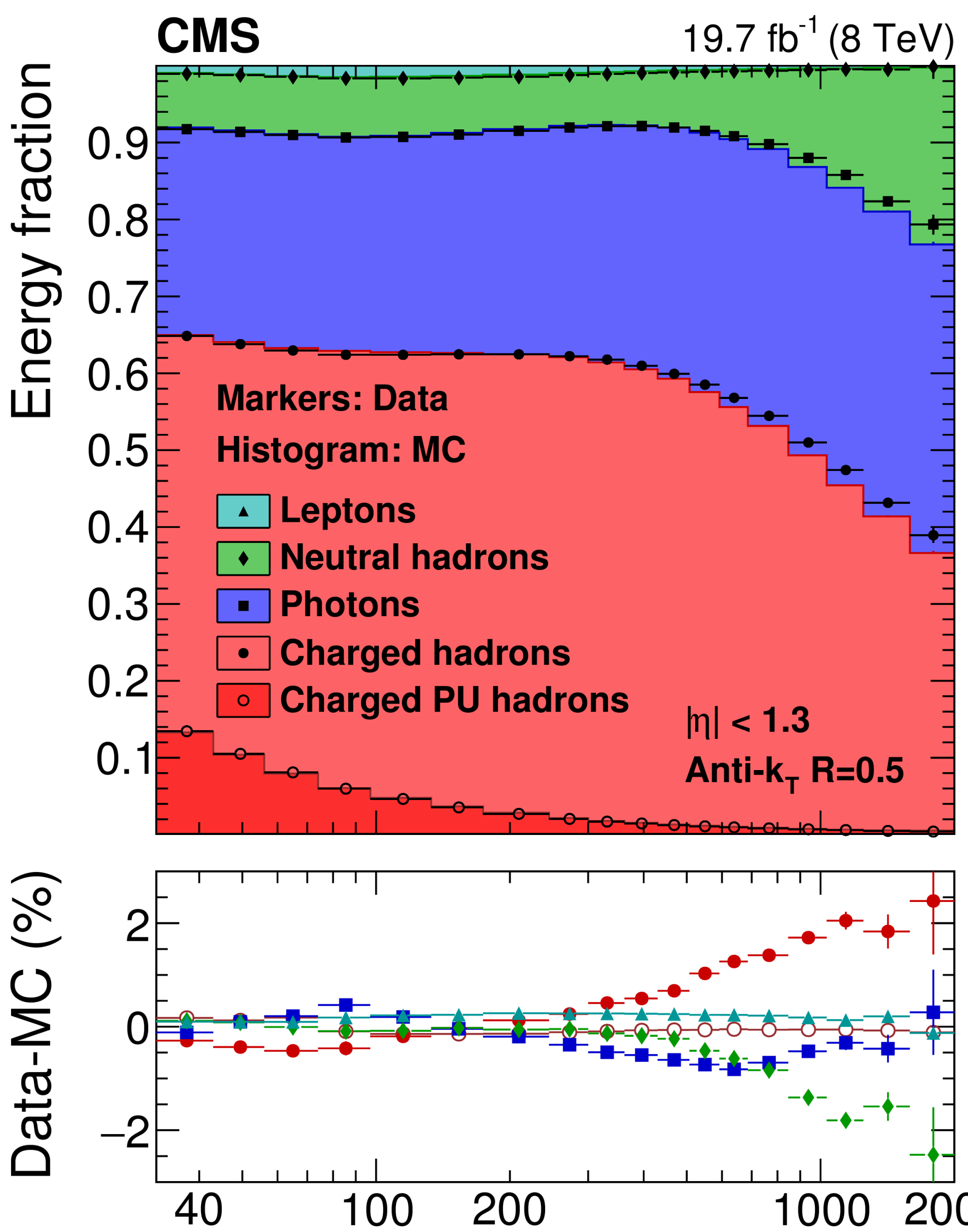
Conclusions

- HGCAL will provide much more information than any previous calorimeter.
  - deploying needed handles to cope with dramatic increase of pileup at the HL-LHC
  - designed to boost particle flow, the reconstruction paradigm of CMS
- Building and exploiting the HGCAL brings major technological challenges
  - software development is one of the development cores
  - fast, robust clustering algorithms bringing offline and online worlds as close as possible
  - still a long way to deploy the final particle flow algorithm for Phase 2
- Exciting times ahead

Backup

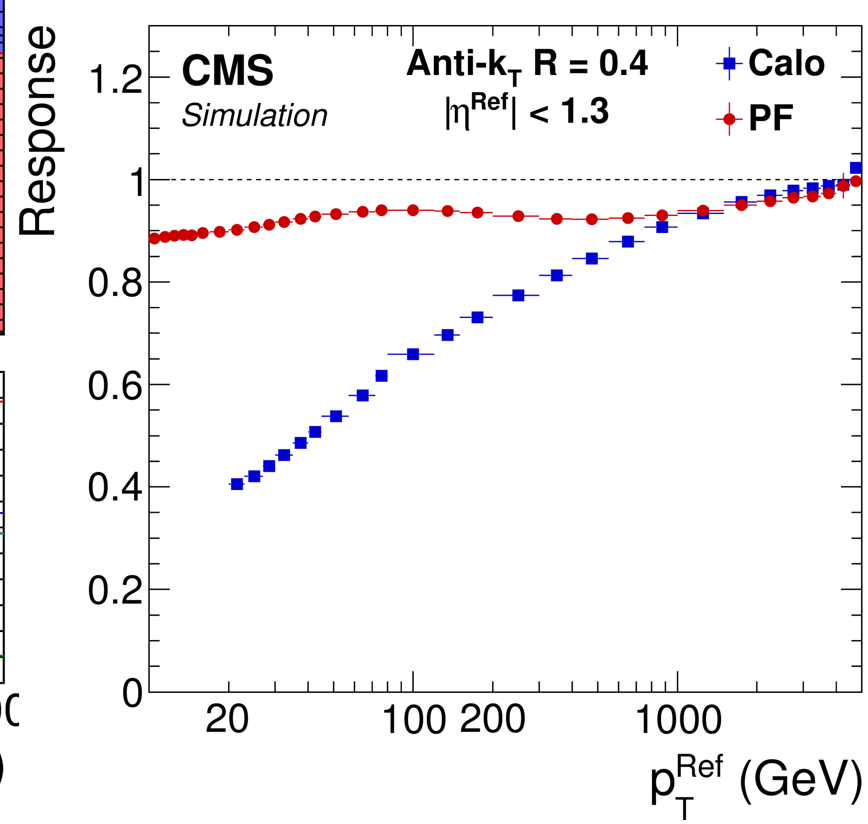
Particle flow performance

- **>80% of the jet components are reconstructed using high resolution detectors**
    - *tracking*: π^+,K^+ and other charged hadrons are approximately O(60%)
    - *ECAL*: by isospin symmetry $\pi \rightarrow \gamma\gamma$ contribute in second place with O(20%)

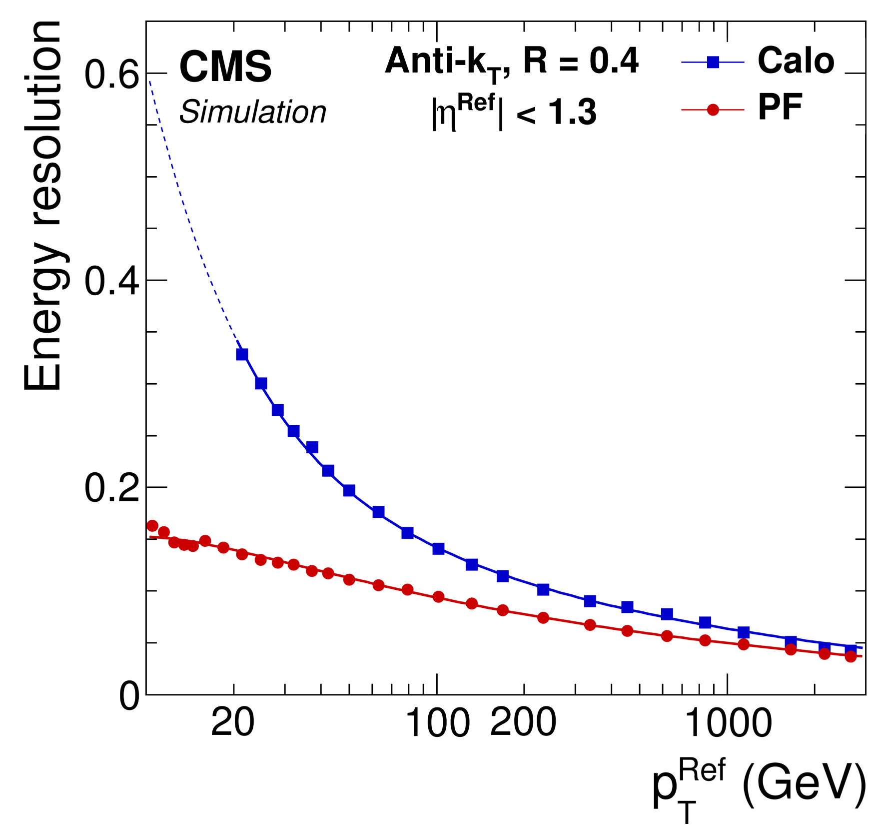


arXiv:1706.04965

Excellent out-of-the-box response (p_T^{reco}/p_T^{ref})



Ability to reconstruct jets at lower pT, gain>60% in resolution ($\Delta p_T/p_T^{ref}$)

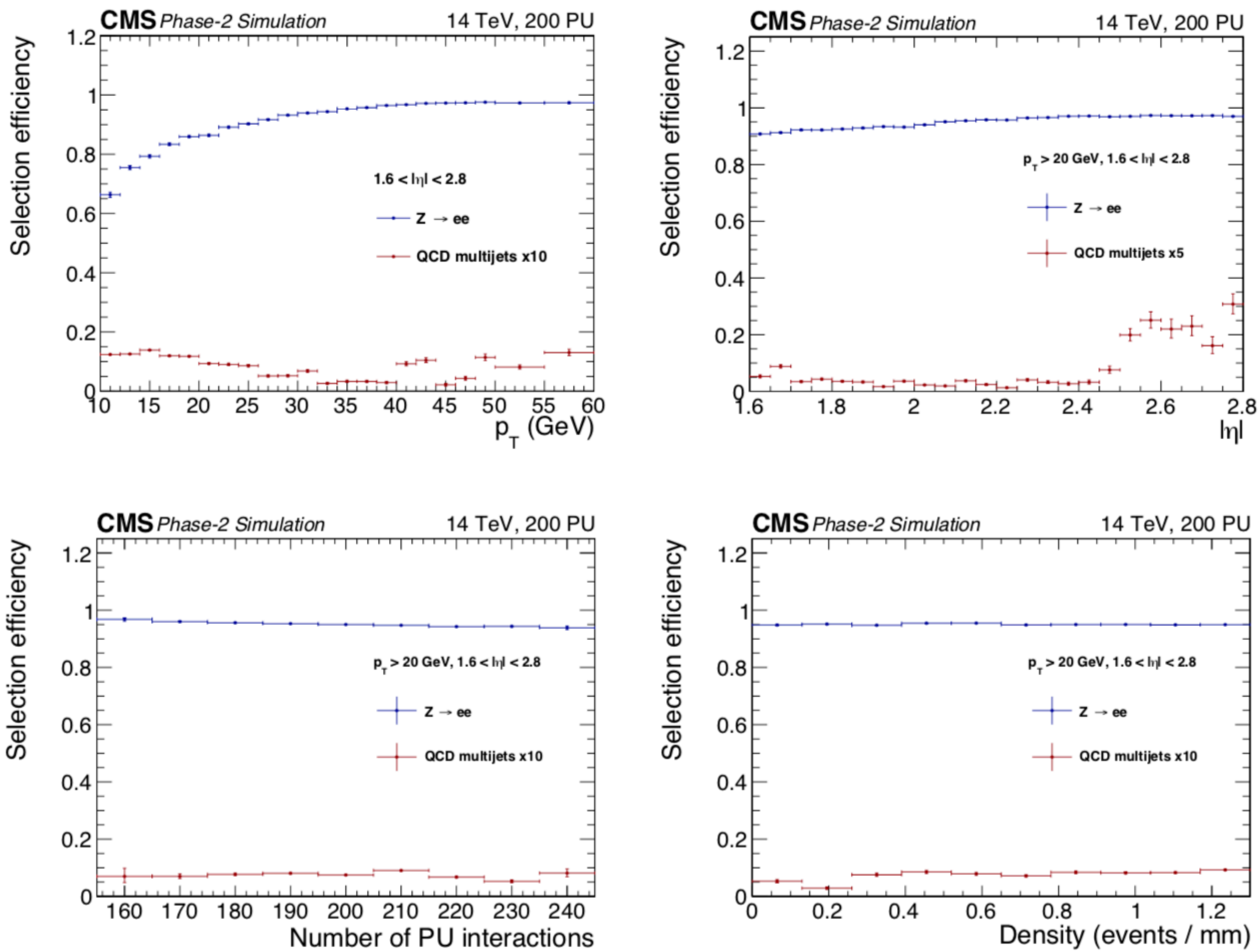


HGCAL geometry: TDR versus CMS

- Implementation of the full geometry as described in the TDR is work in progress
    - closest implementation available in a standalone Geant4 setup with fixed Si width
    - standard CMS simulation has full Scintillator BH and differs in sampling fractions for FH/BH
    - both setups implement hexagonal-like pads for the Si sections

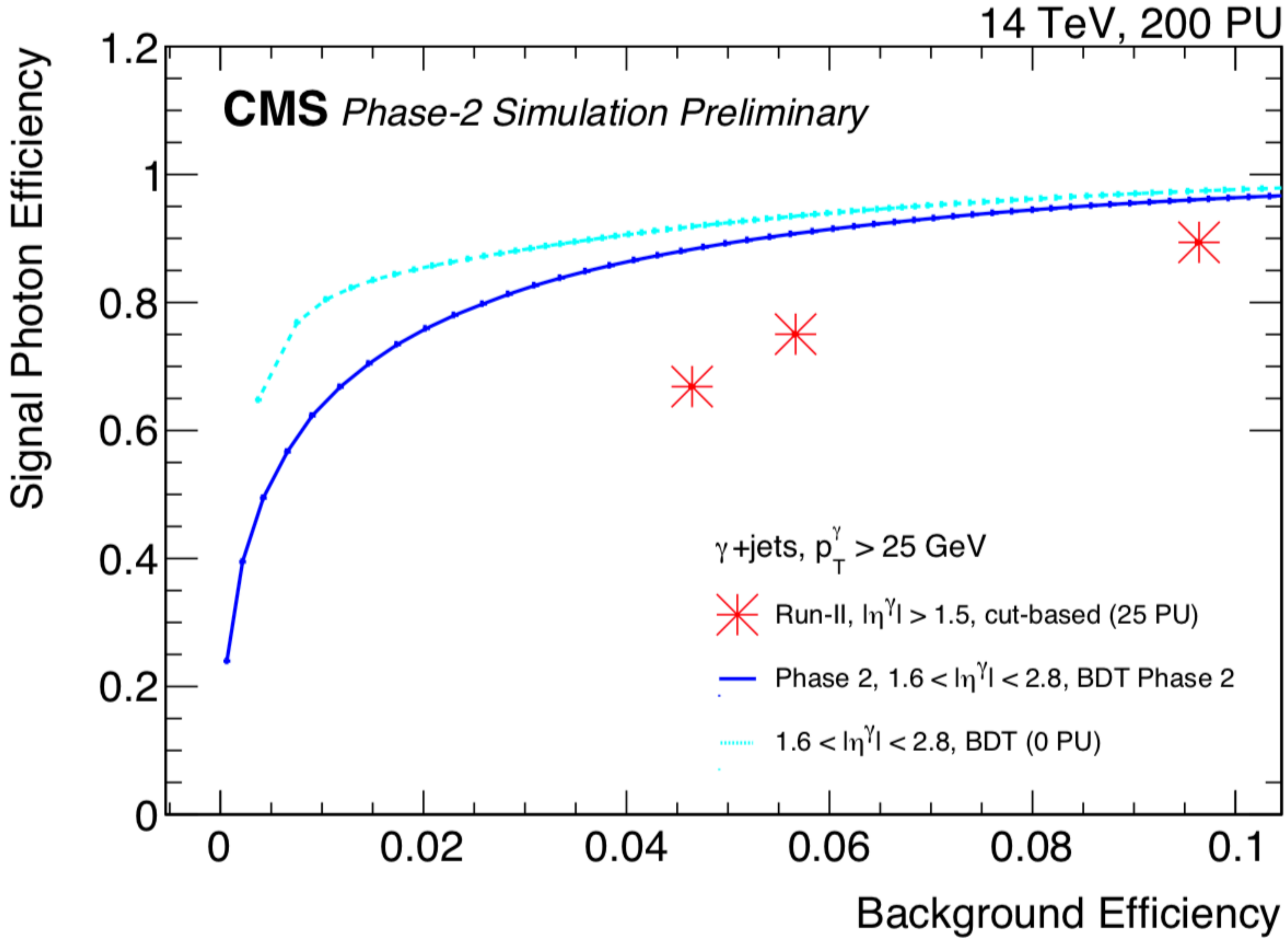
	CMSSW			TDR design		
	cm	X_0	λ	cm	X_0	λ
Neutron moderator	18.0	0.4	0.2	15.7	0.3	0.2
Electromagnetic section	32.1	26.7	1.6	33.9	25.4	1.6
1^{st} hadronic section	59.7	33.8	3.1	60.2	35.2	3.7
2^{nd} hadronic section	110.6	62.8	5.8	100.6	49.9	5.2
Total	220.4	123.6	10.7	210.4	110.9	10.7

Further details on electron selection efficiencies

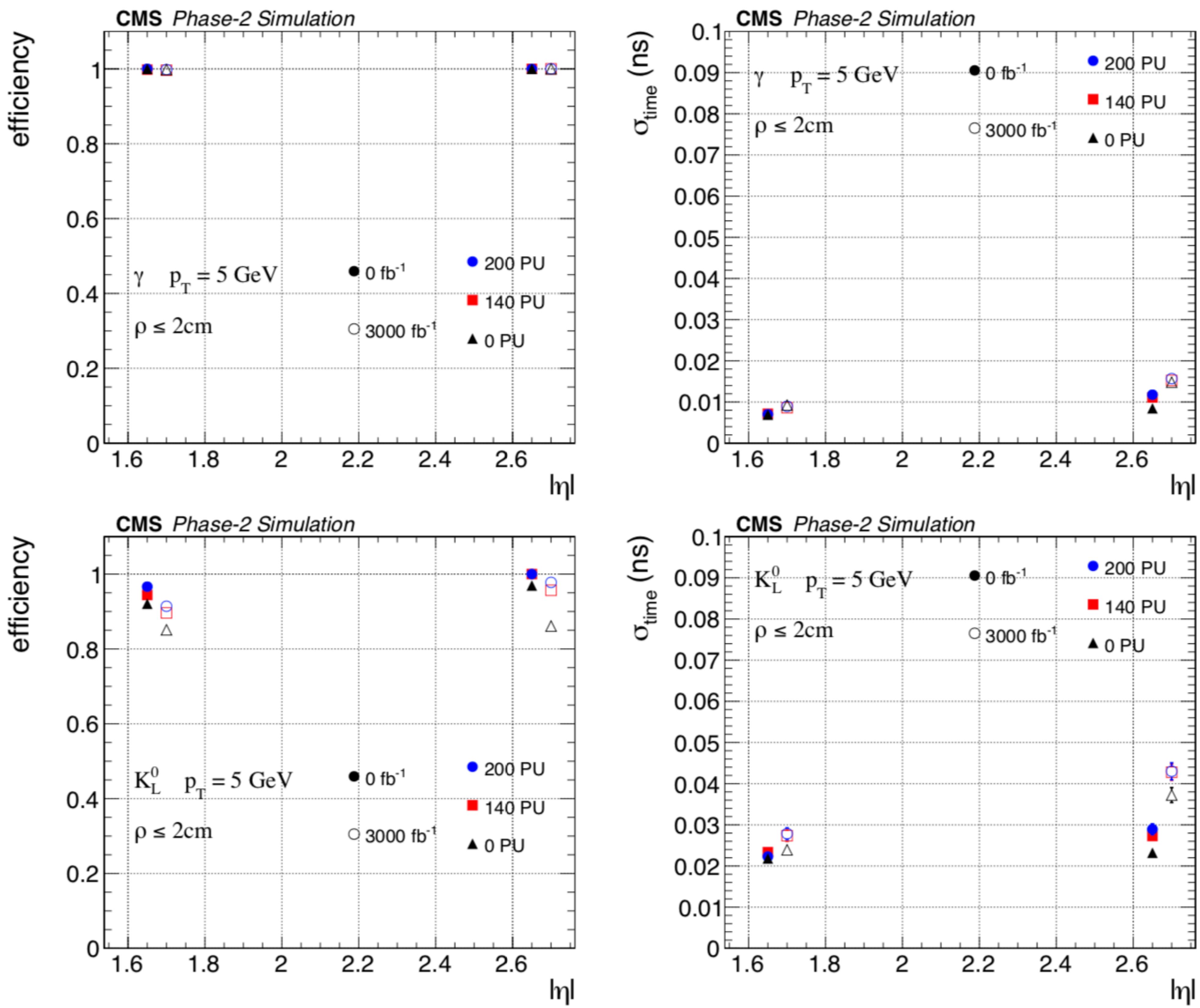


Photon identification efficiency

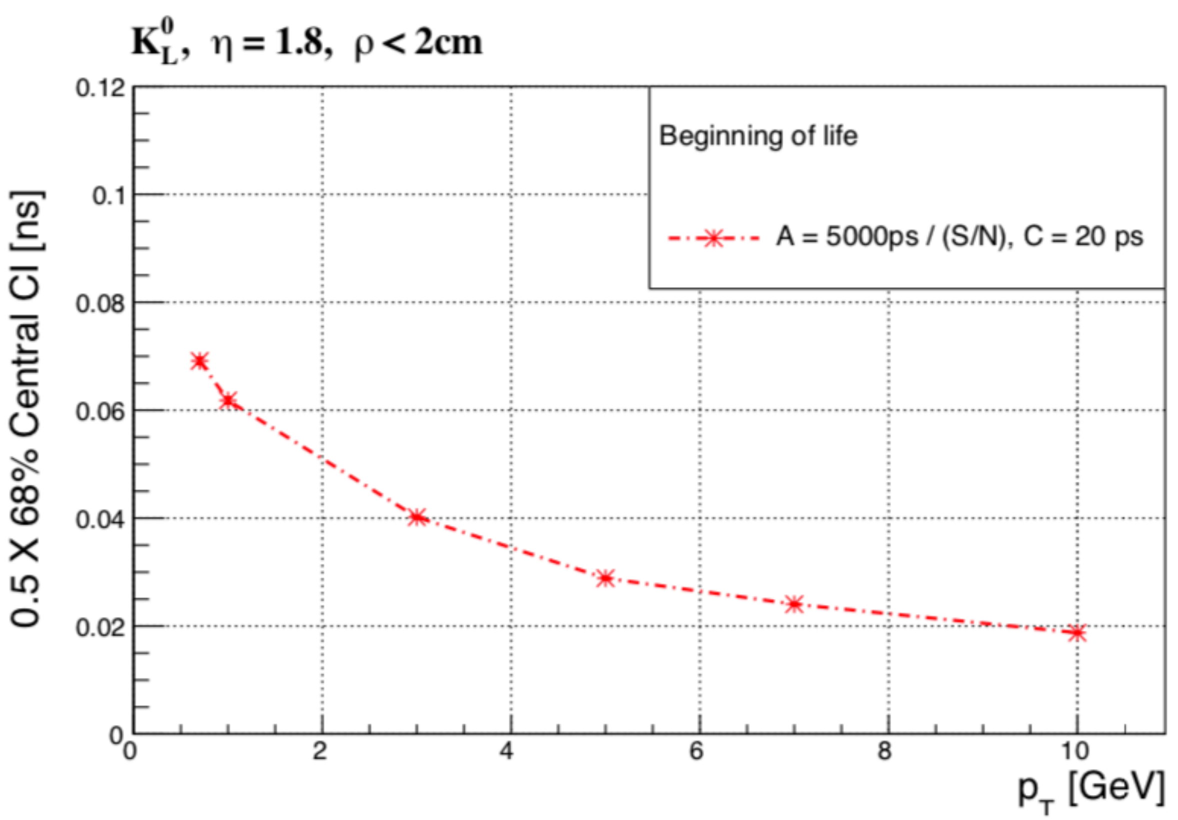
- Based on analogous variables as the ones used for electrons (but no track)

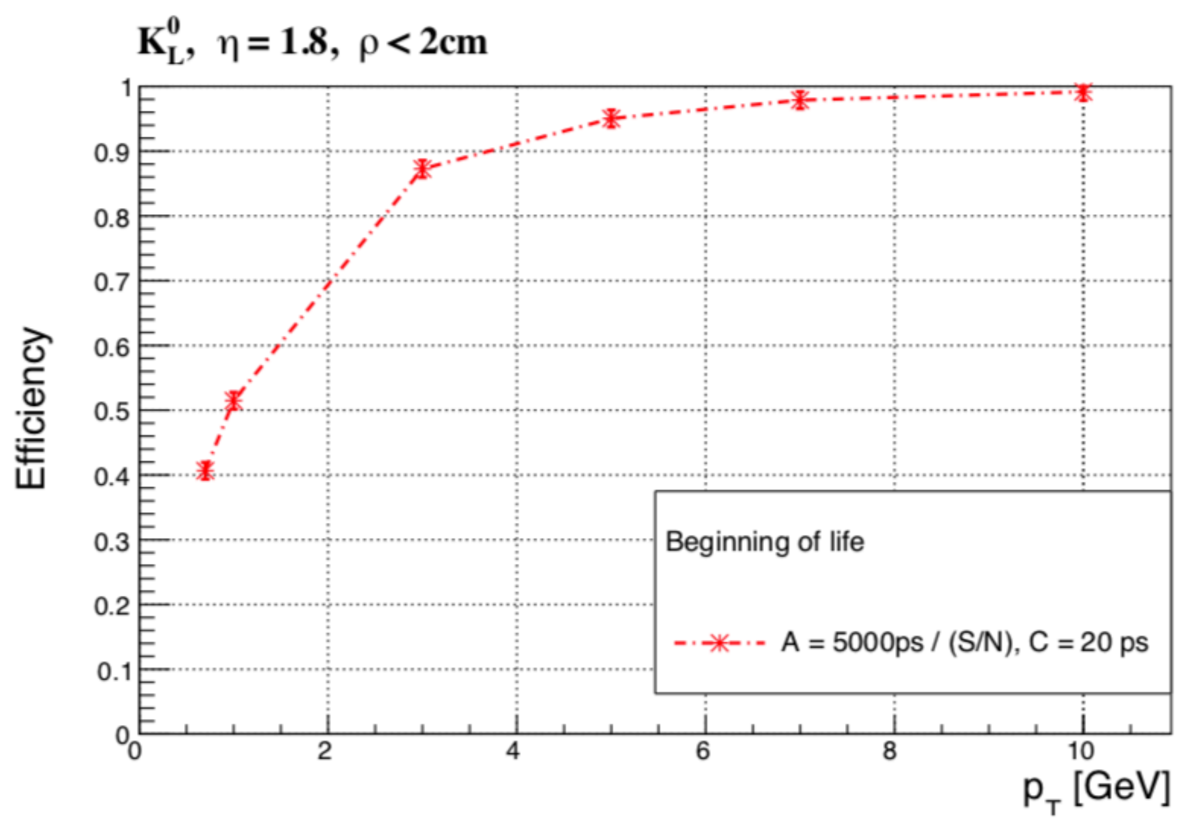


Vertex association efficiency and time resolution I



Vertex association efficiency and time resolution II





Hadron energy resolution with CNN