Particle flow in CMS

Matthew Nguyen
ePIC Jets & HF meeting
July 13th, 2023

My CMS PFA credentials

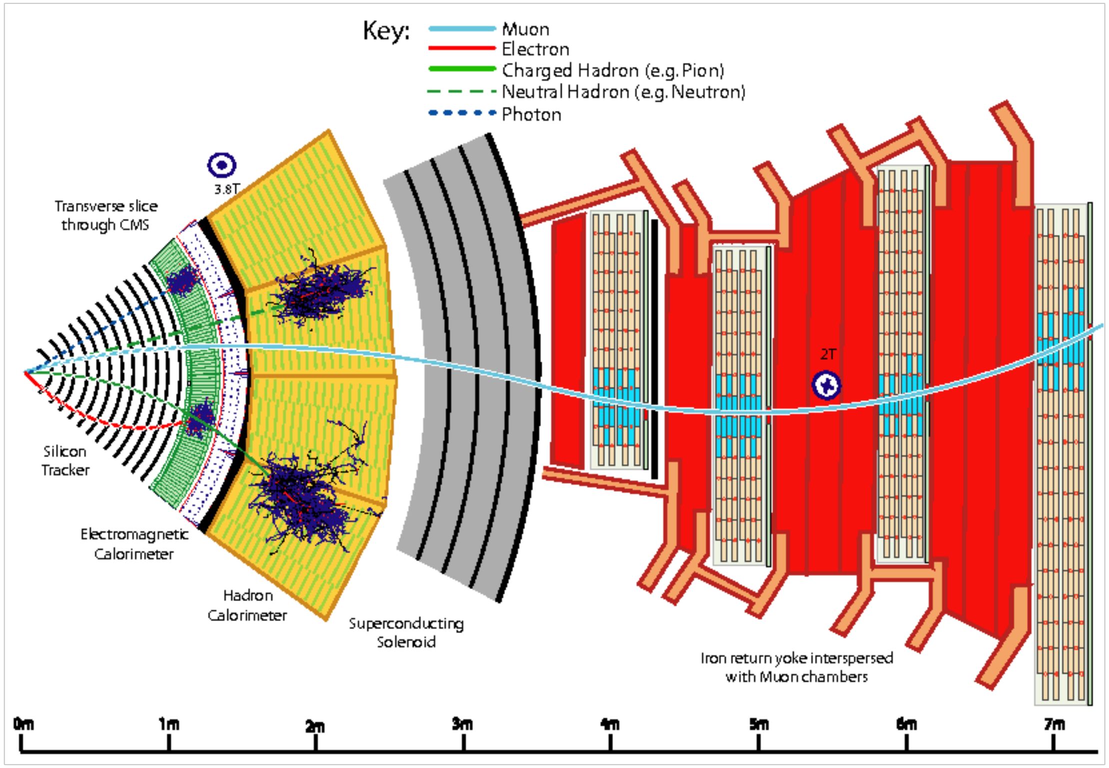
- CMS member since LHC start-up in 2009
- During ramp-up period (2009 – 2011), I worked on commissioning of the CMS particle flow (PF) algorithm
- I adapted the CMS PF algo for heavy ion collisions
- I currently serve as reconstruction co-convener of CMS (2022 – 2024)

All material drawn from
- CMS PF JINST
- Talks from P. Janot & C. Bernet [1,2]

Code lives here:
https://github.com/cms-sw/cmssw/tree/master/RecoParticleFlow

What is particle flow and why do it?

- Reconstruction based on *physics objects* (vs. detector)
- First developed for e^+e^- collisions w/ ALEPH detector
- For CMS: charged hadron, neutral hadron, photon, e, μ



- PF = optimized combination of information from sub-detectors
- Simplifies analysis, at cost of more complex reconstruction

Detectors for particle flow

The ideal PF detector

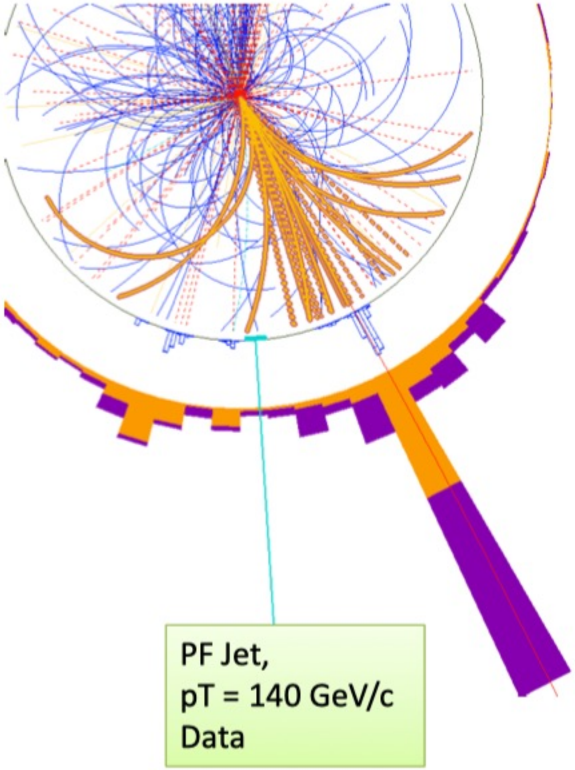
- Tracker
    - High granularity & B field
    - High efficiency / low fake
- Calorimeters
    - Segmentation above all else
    - Long. segmentation a plus
    - Energy resolution secondary
- Material:
    - As little inactive as possible

CMS: NOT made for PF

- Tracker
    - ✓ Pixel + strips, $\sigma_p/p \approx 1\%$
    - x Eff. limited by material
- Calorimeters
    - ✓ ECAL: Excellent spatial and energy resolution
    - x HCAL: Modest spatial and energy resolution
- Material:
    - ✓ Calorimeters inside magnet
    - x Too much material in tracker

pp collision environment more challenging than e^+e^-, particularly at high PU (not to mention heavy ions)

Separating particles



CMS magnetic field

- B = 3.8 T
- ECAL surface at 1.29 m
- B*r = 4.9 T*m

For comparison

- ALEPH: 1.5*1.8 = 2.7 T*m
- ATLAS: 2.0*1.2 = 2.4 T*m
- CDF: 1.5*1.5 = 2.3 T*m
- D0: 2.0*0.8 = 1.6 T*m

Limited segmentation can be compensated by strong B field

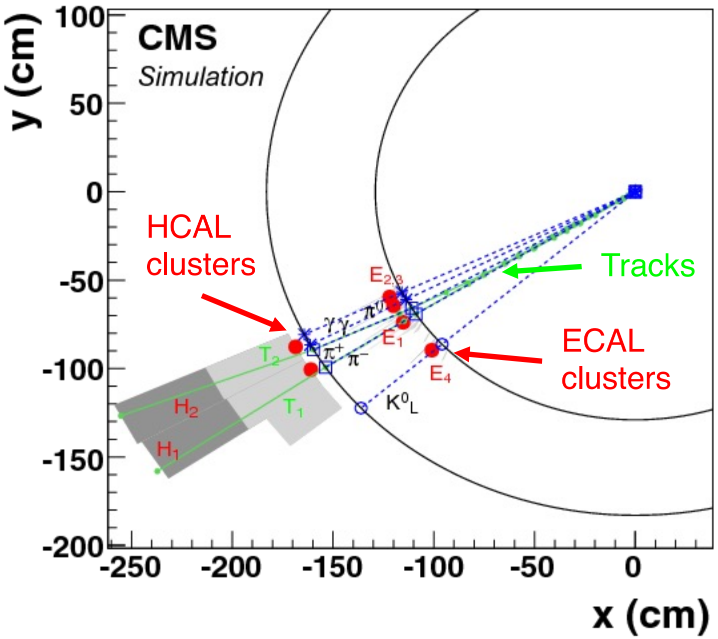
NB: For calo jets, low p_T charged hadrons are pushed out of cone

PF recipe

1) Local reco: (super) clustering, track finding, lepton ID*,
2) Link between elements from different detector subsystems to form blocks
3) Resolve blocks into particles w/ appropriate calibrations (calibrations discussed before linking)
4) Post-processing (cleaning)

Illustrative example jet

50 GeV jet containing:
- 2 charged hadrons: π^+, π^-
- 2 photons from a π^0 decay
- 1 neutral hadron: K^0_L



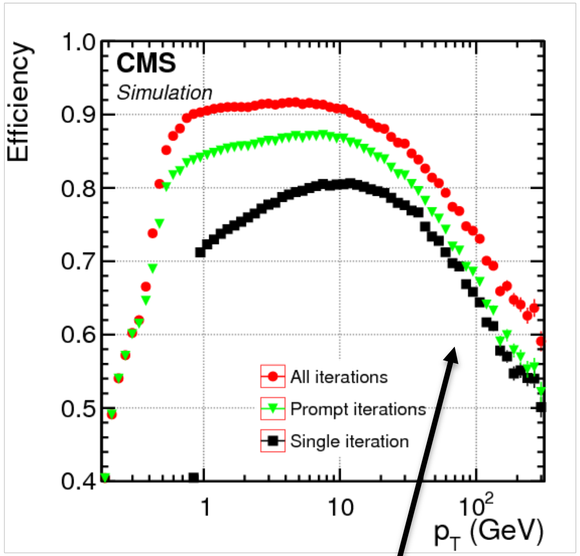
---

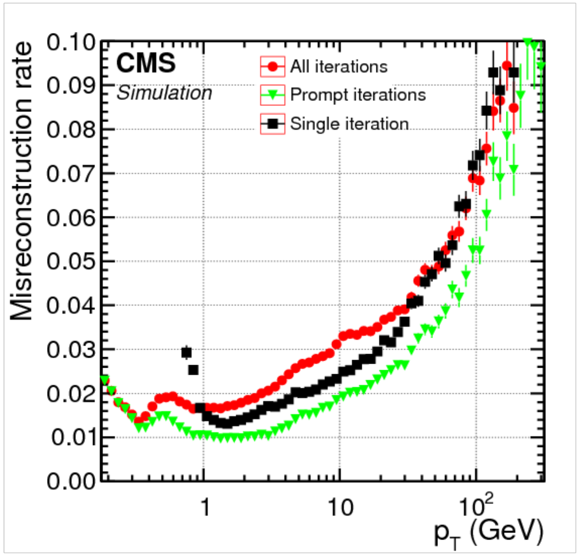
*Leptons most neglected in this talk, will come back to muons in discussion of post-processing

Local reco: tracking

“Iterative tracking” for higher efficiency, improved CPU timing

Description of tracking in CMS
[JINST 9 (2014) 10, P10009]



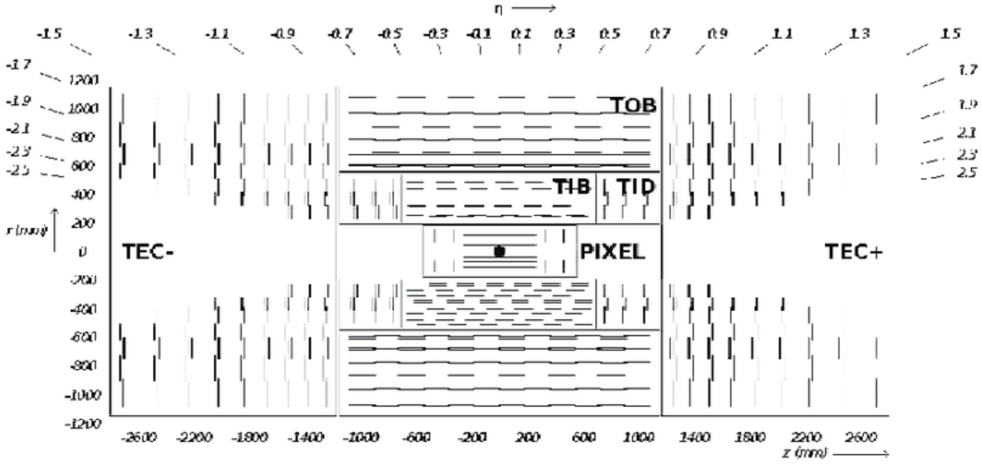


- Performance degradation at large p_T due to track merging / hit confusion
- Much improved now after years of development (cluster splitting, DNN, etc.)
- Large E charged hadrons are anyway well-measured in calorimeters

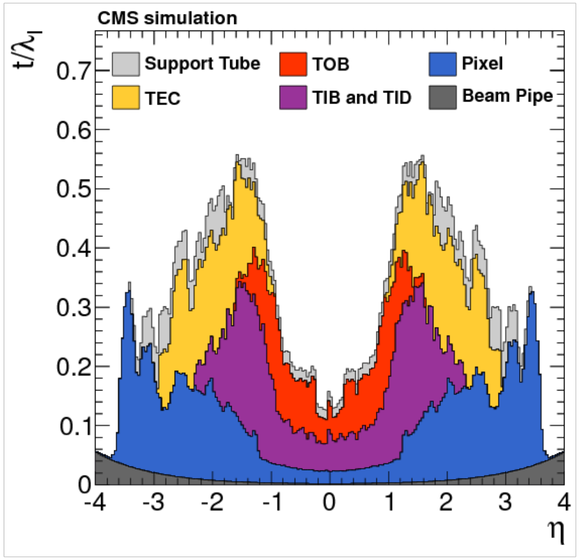
CMS tracker

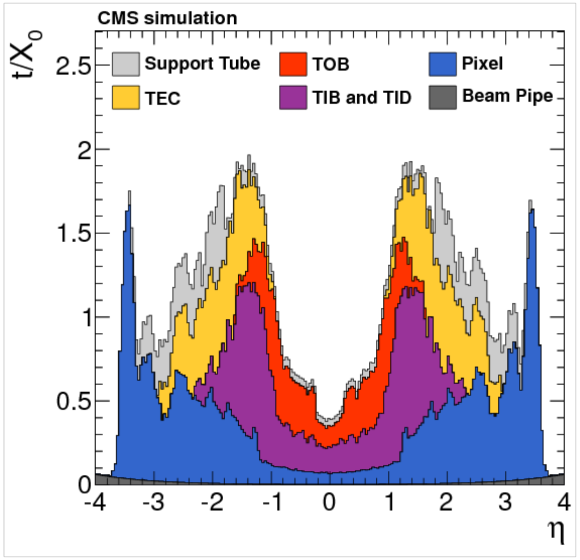
Required to be fast and rad-hard

→ large material budget



- At worst (|η| ≈ 1.5)
  - 85% γ conversion / e brem
  - 20% $h^{+/-}$ have nuclear interaction before ECAL
- Secondaries are a major complication of CMS PFA





Local reco: clustering

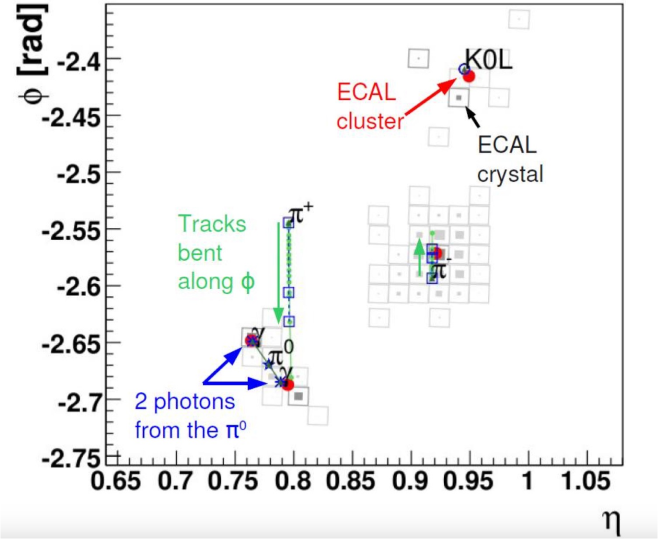
Dedicated PF clustering algorithm is designed to be outlier resistant

- Seeds: cells above a given threshold & higher than neighboring cells
- Topo clusters: seed + cells sharing a side (ECAL & HCAL) or a corner (ECAL only)
- Final clusters: obtained with Gaussian mixing model for energy sharing

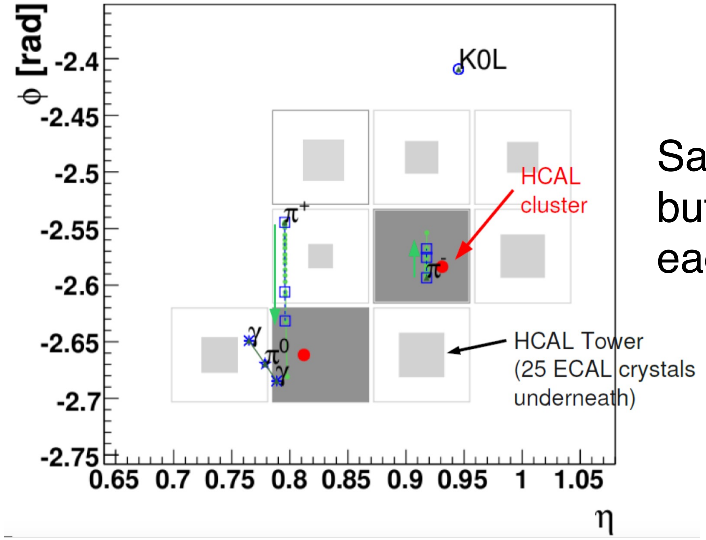
	ECAL		HCAL	
	barrel	endcaps	barrel	endcaps
Cell E threshold (MeV)	80	300	800	800
Seed # closest cells	8	8	4	4
Seed E threshold (MeV)	230	600	800	1100
Seed E_T threshold (MeV)	0	150	0	0
Gaussian width (cm)	1.5	1.5	10.0	10.0

* Omitting pre-shower from this presentation for simplicity

ECAL view



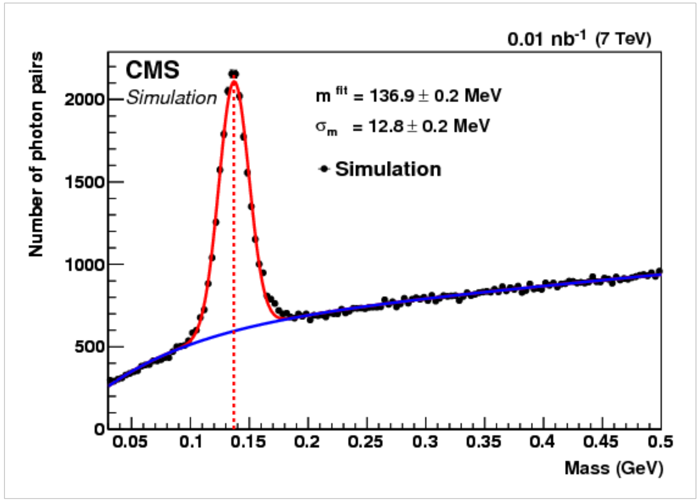
HCAL view

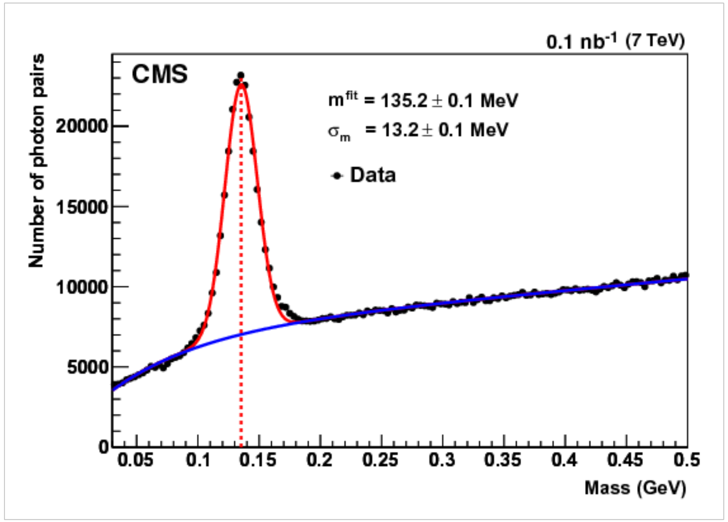


Same example event, but in the η-ϕ plane of each calorimeter

Cluster calibrations for photons (ECAL)

- Thresholds in ECAL clustering require energy scale correction
- Derived from (un-converted) photon gun GEANT simulation vs E and η
- Correction factor can be as large as 20% at low E





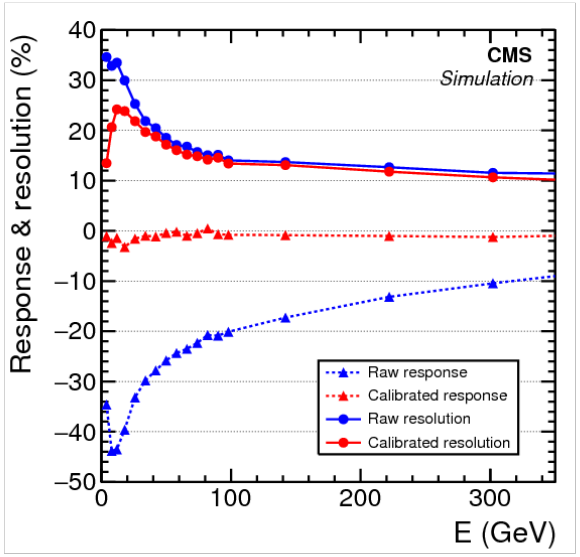
- Resulting π^0 peak in data within 1% of PDG for all E & η validating simulation
- Note that π^0 are mostly merged in jets, but not relevant for PF

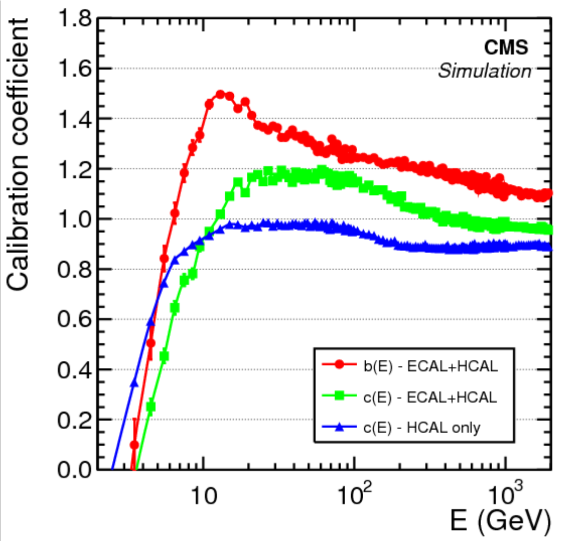
Cluster calibrations for hadrons

- Initial HCAL calibrations derived from test beam w/ 50 GeV pions w/o ECAL interaction
- But HCAL response to charged hadrons:
  - depends on energy deposited in ECAL (≈ 1λ)
  - Is non-linear
- Response derived from K^0_L gun MC, then corrected w/ isolated hadron data

$$E_{\text{calib}} = a + b(E) f(\eta) E_{\text{ECAL}} + c(E) g(\eta) E_{\text{HCAL}}.$$

- *b* & *c* determined by iterative χ^2 minimization
- *a* represents energy lost to thresholds
  - obtained by minimizing dependence of *b* & *c* on *E*, for $E > 10$ GeV
  - $a = 3.5$ (2.5) GeV for hadrons showering in ECAL & HCAL (HCAL only)

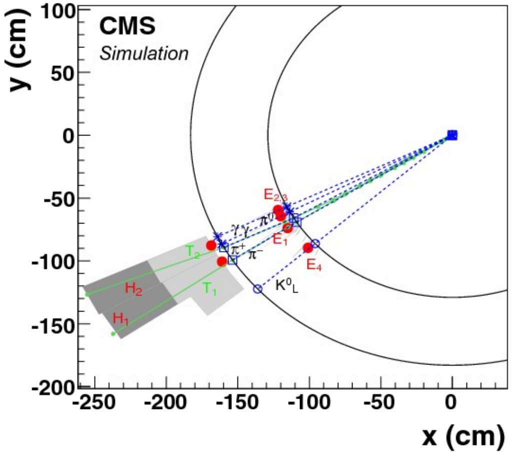




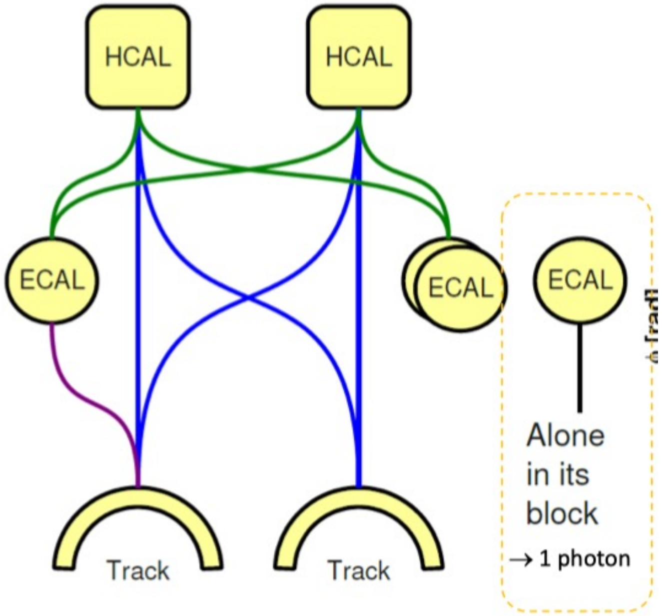
Calibration procedure applied directly to isolated hadrons, for non-isolated hadrons, first have to discuss *linking* algo

Linking

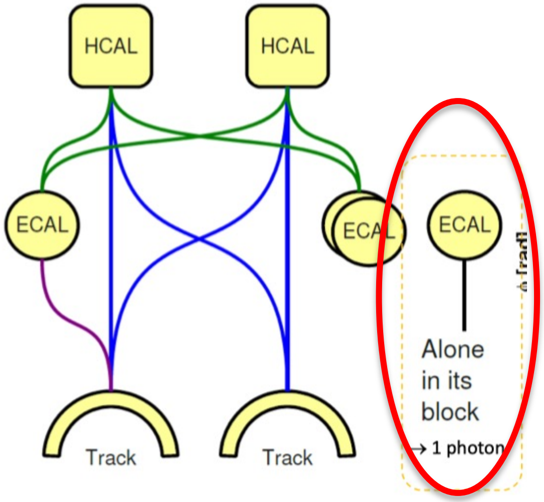
- *Elements* are linked into *blocks*
  - Purity driven by granularity & particle density
  - Efficiency driven by material (kinks, secondaries)
- Track-calo matching:
  - Track extrapolated from outermost hit to *e* shower max (one interaction length) for ECAL (HCAL)
  - Extrapolation must intersect cluster boundary + an envelope that accounts for cracks, uncertainty in shower max position, and for multiple scattering
- Calo-calo matching:
  - Match if ECAL cluster lies with in HCAL cluster
  - If ECAL matches multiple HCAL, choose closest
- Linking time quadratic w/ multiplicity → pairs of elements restricted to nearest neighbors using a k-dimensional tree

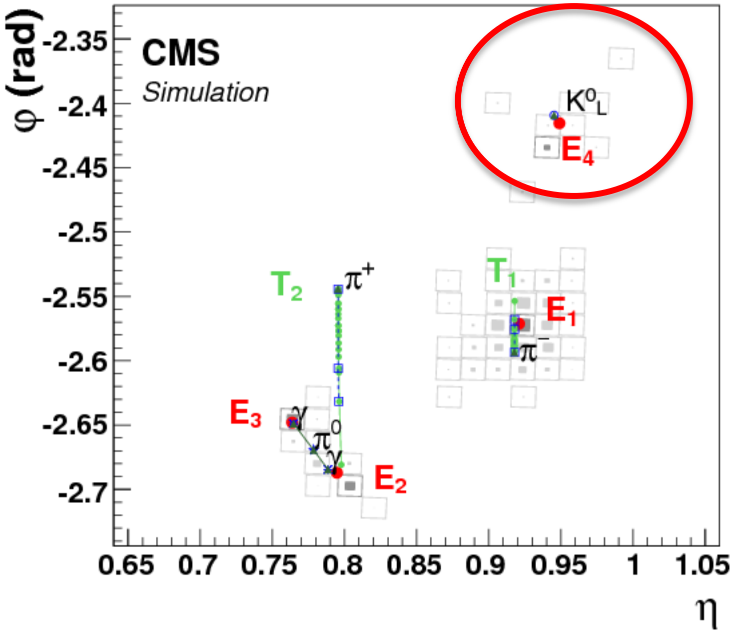


This event gives 2 *blocks*



“Photon precedence”



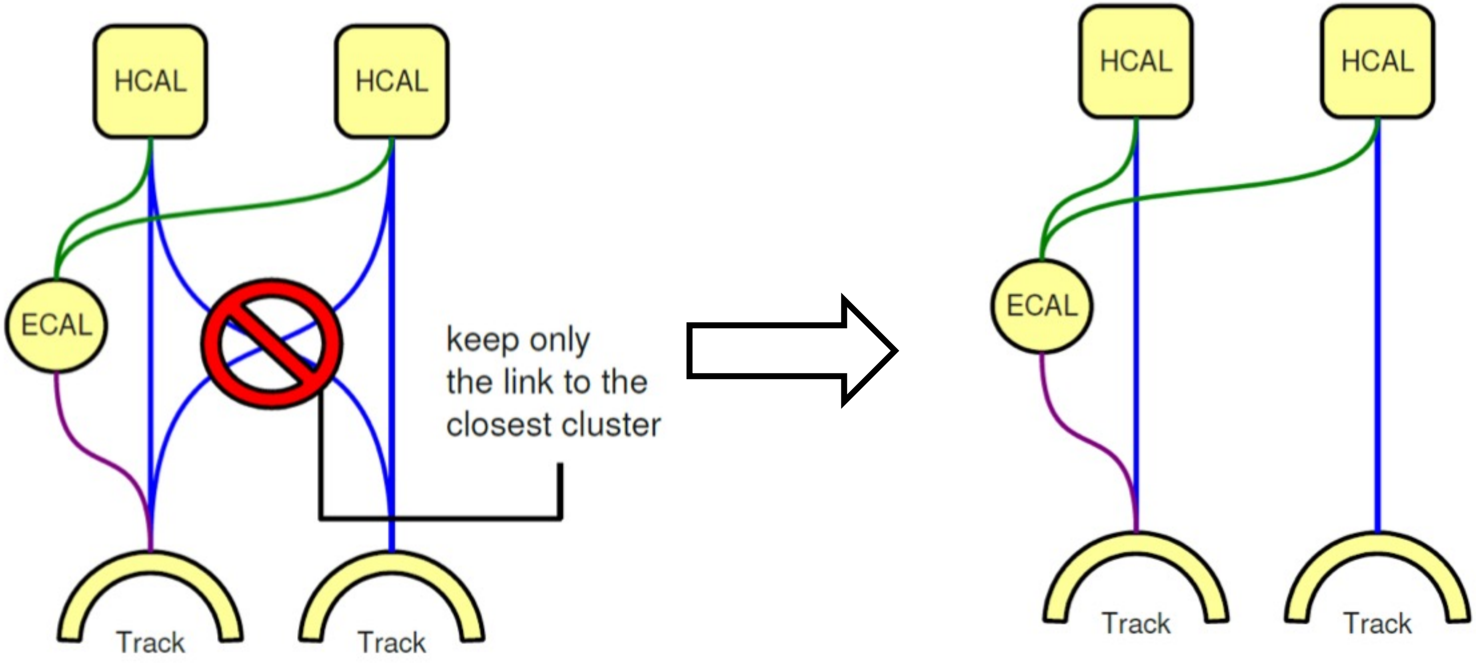


- K^0_L deposits all its energy in ECAL
  → labeled a photon
- Choice justified by jet composition
  - 25% of jet energy in photons
  - 3% of jet energy from neutral hadrons in ECAL
- Calibration based on EM hypothesis
  → Response for these $h^0 \approx 30\%$ low, w/ JES ≈ 0.5% low (left for JECs)

Link disambiguation

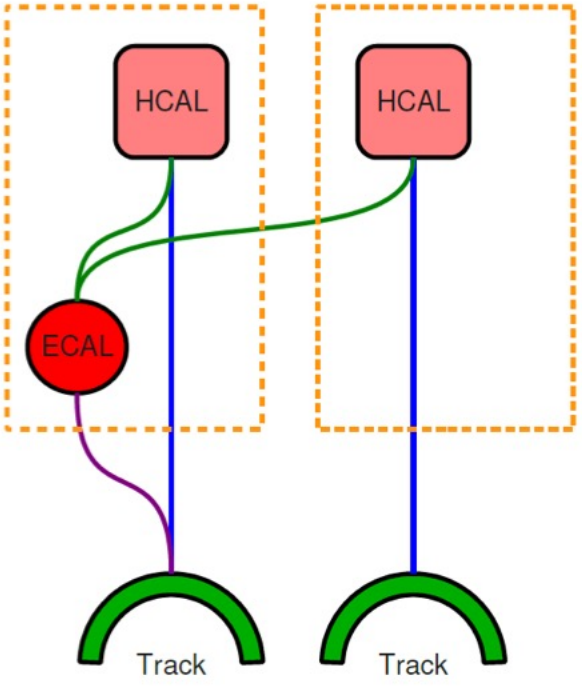
Keep only closest link if …

- A track matches multiple HCAL clusters
- ECAL cluster matches multiple HCAL clusters
- An ECAL cluster matches multiple tracks



keep only the link to the closest cluster

Resolving blocks



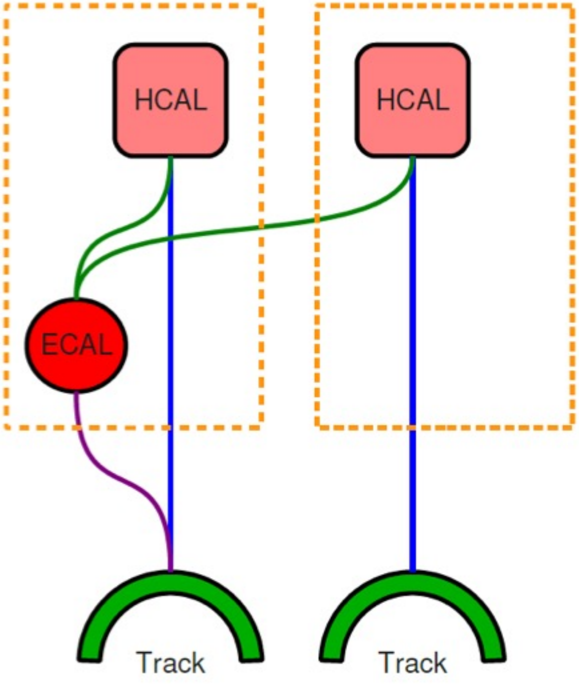
For each HCAL cluster
compare sum track p vs calo E*

- If p and E compatible:
  - ➢ Create $h^{+/-}$ (1 per track)
- if E >> p + 120%*√p:
  - ➢ Create $h^{+/-}$ + neutrals
- If E << p: something is fishy
  - ➢ Re-check for muons or fake tracks. If not, create $h^{+/-}$

* Reference E based on hadron hypothesis

$$E = a + bE_{ECAL} + cE_{HCAL}$$

Energy/momentum assignment

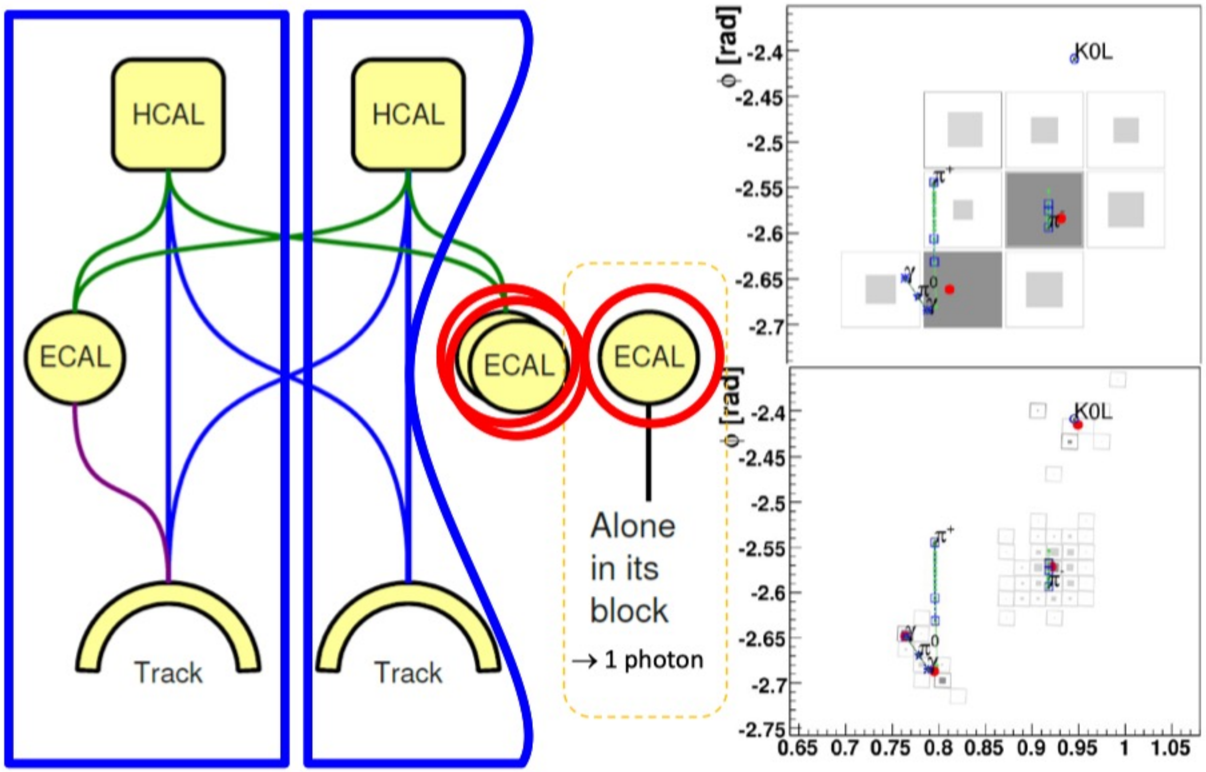


$$E = a + bE_{ECAL} + cE_{HCAL}$$

- If p & E compatible, h^{+-} from:
  - fit to p_i & E according to $\sigma_{p,E}$
    - → p_i for small $p_{T,i}$,
    - → calo measurement at large E
- If E significantly larger than p:
  - $h^{+/-}$ with p_i + 1 or more neutrals:
    - If E from HCAL or ECAL only:
      - HCAL → h^0 with E – p
      - ECAL → γ with E - p/b
    - If *E* from both HCAL & ECAL:
      - If $E\text{-}p > E_{ECAL}$: γ w/ E_{ECAL} + h^0 w/ rest
      - If $E\text{-}p < E_{ECAL}$: γ with $(E – p)/b$ (“photon precedence”)

Wrapping up our event of interest:

2 charged hadrons, 3 photons

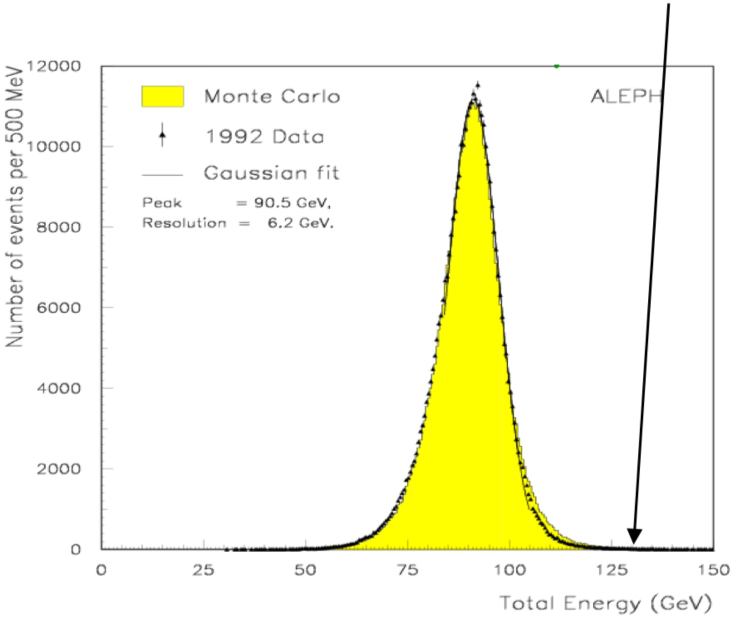


Alone in its block → 1 photon

Post-processing: muon cleaning

- Have ignored leptons so far in this presentation
  - Muon ID is done before other particles: High quality tracks with matching high quality muon road are removed from blocks
  - e ID is another story, but specific to CMS (thick tracker, brem collection)
- Post-processing: Revisit particle assignment using high-level quantities

Example from ALEPH:
Look at tail of hadronic Z decays



CMS: Scan large missing E_T (MET) events
→ Post-processing largely concerns muons

Several sources of large MET identified:

1) Cosmics → large impact parameter muons
2) Mis-reco → Poor agreement between momentum in tracker and muon system
3) Punch-through → High E muon w/ fake h^0
4) Missed muons → Fake h^{+-} "eats" nearby h^0

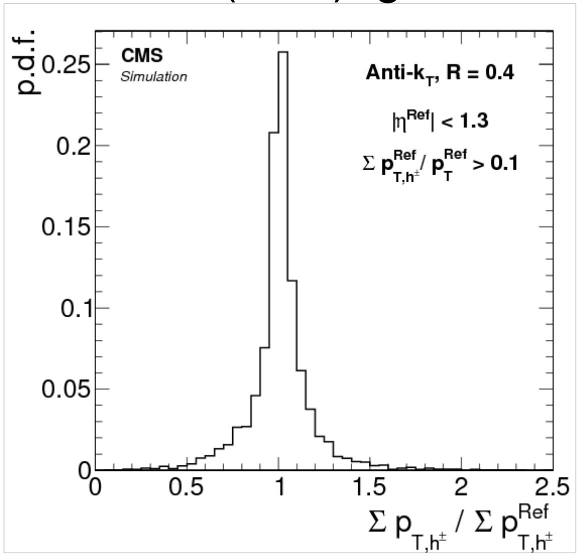
If it decreases MET, action taken:

1) Remove muon
2) Choose different muon momentum estimate
3) Change muon to charged hadron
4) Change charged hadron to muon + h^0

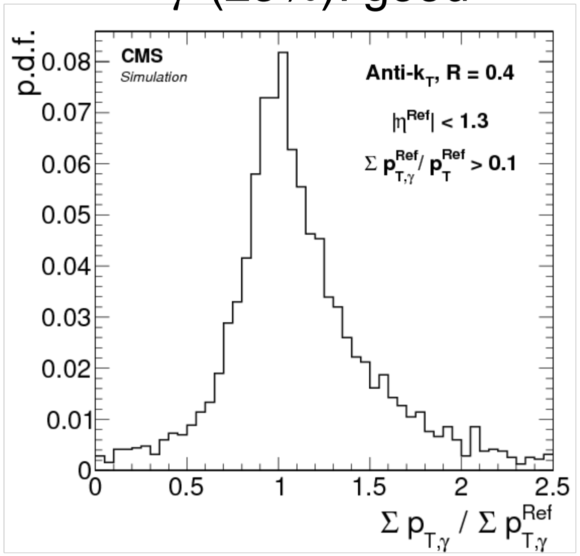
How well does it work?

Particle resolution

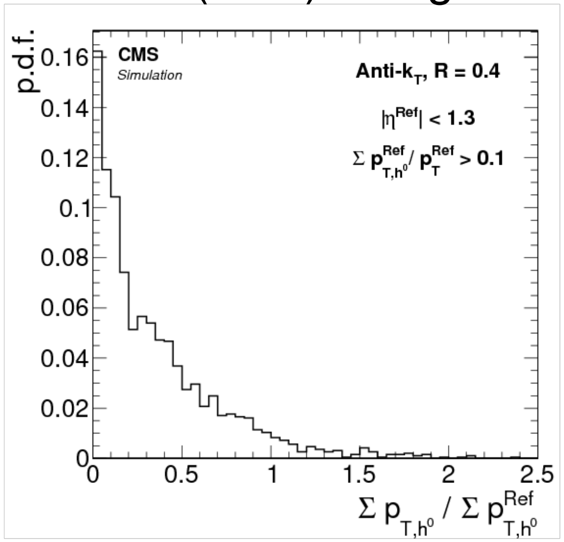
$h^{+/-}$ (65%): great



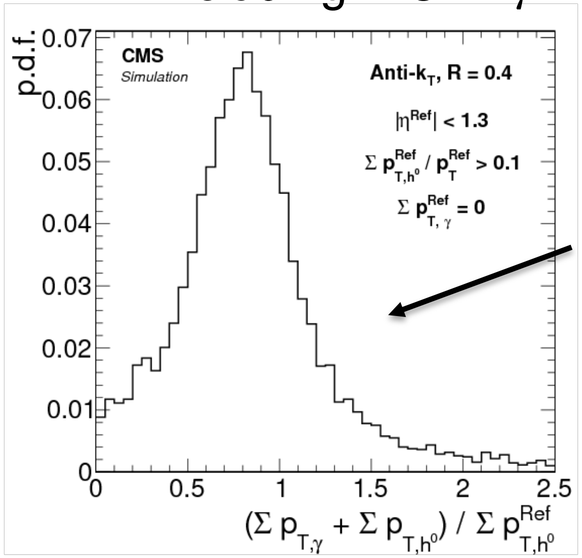
γ (25%): good



h^0 (10%): not great

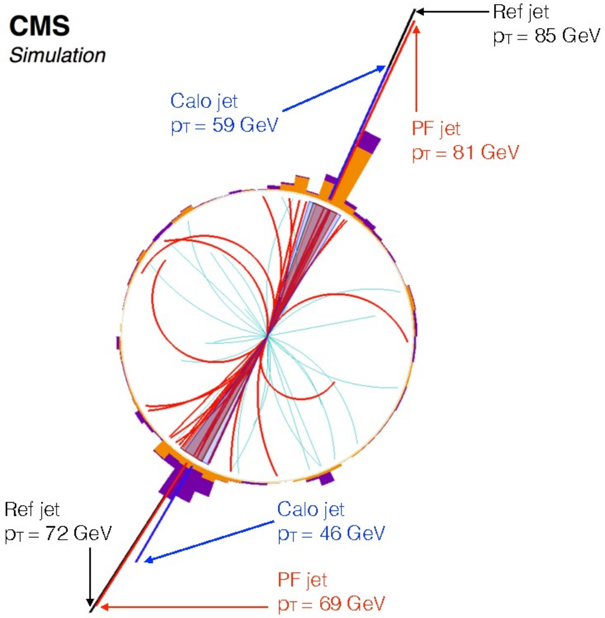


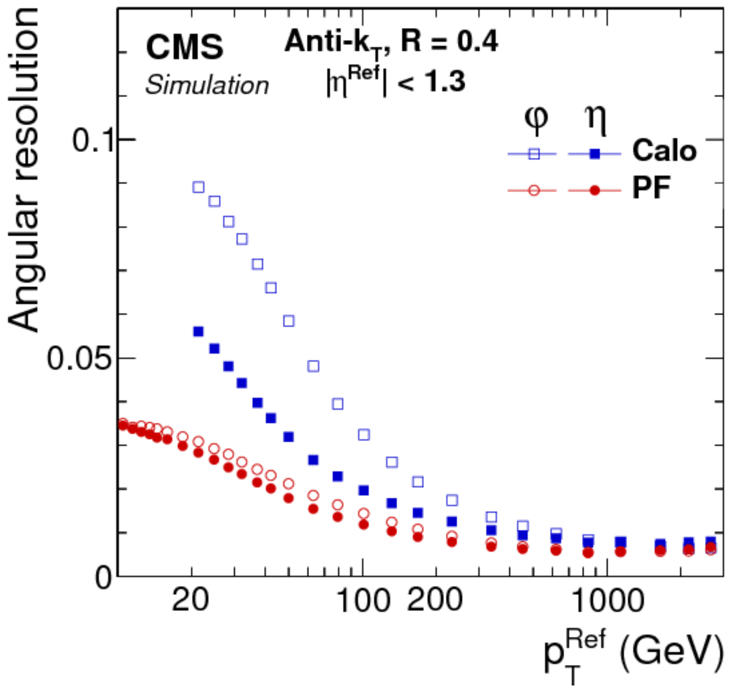
h^0 including mis-ID γ: less bad



* Evaluated w/ jets containing no real γ

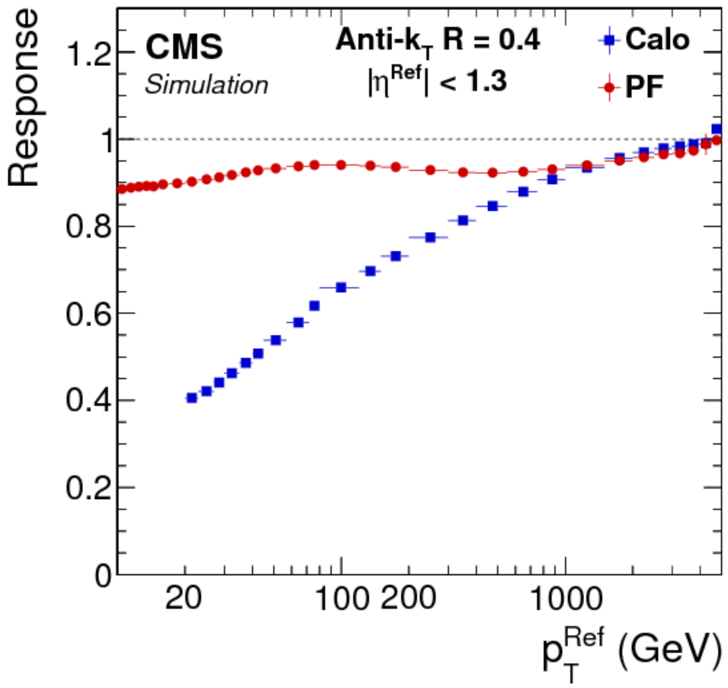
Jet angular resolution

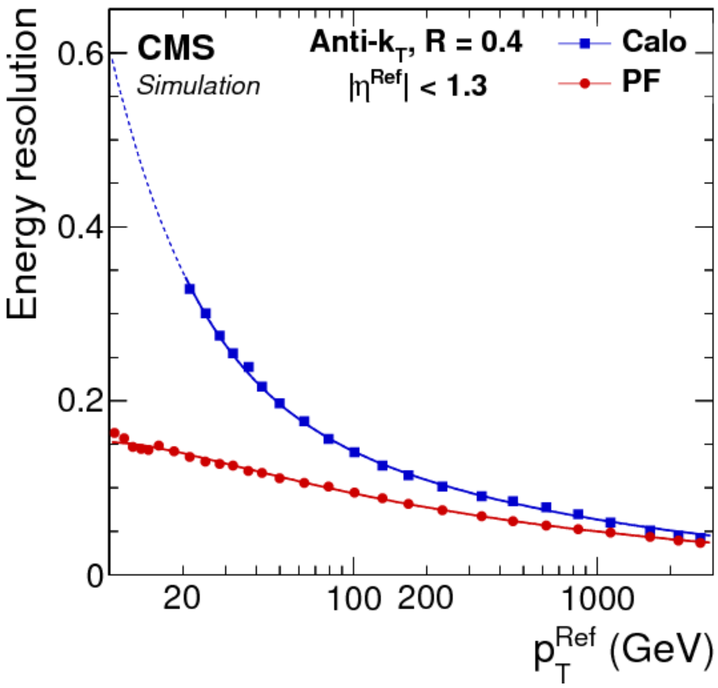




- Mitigates effect of coarse HCAL segmentation
- Recovers h+/- that are bent by B field

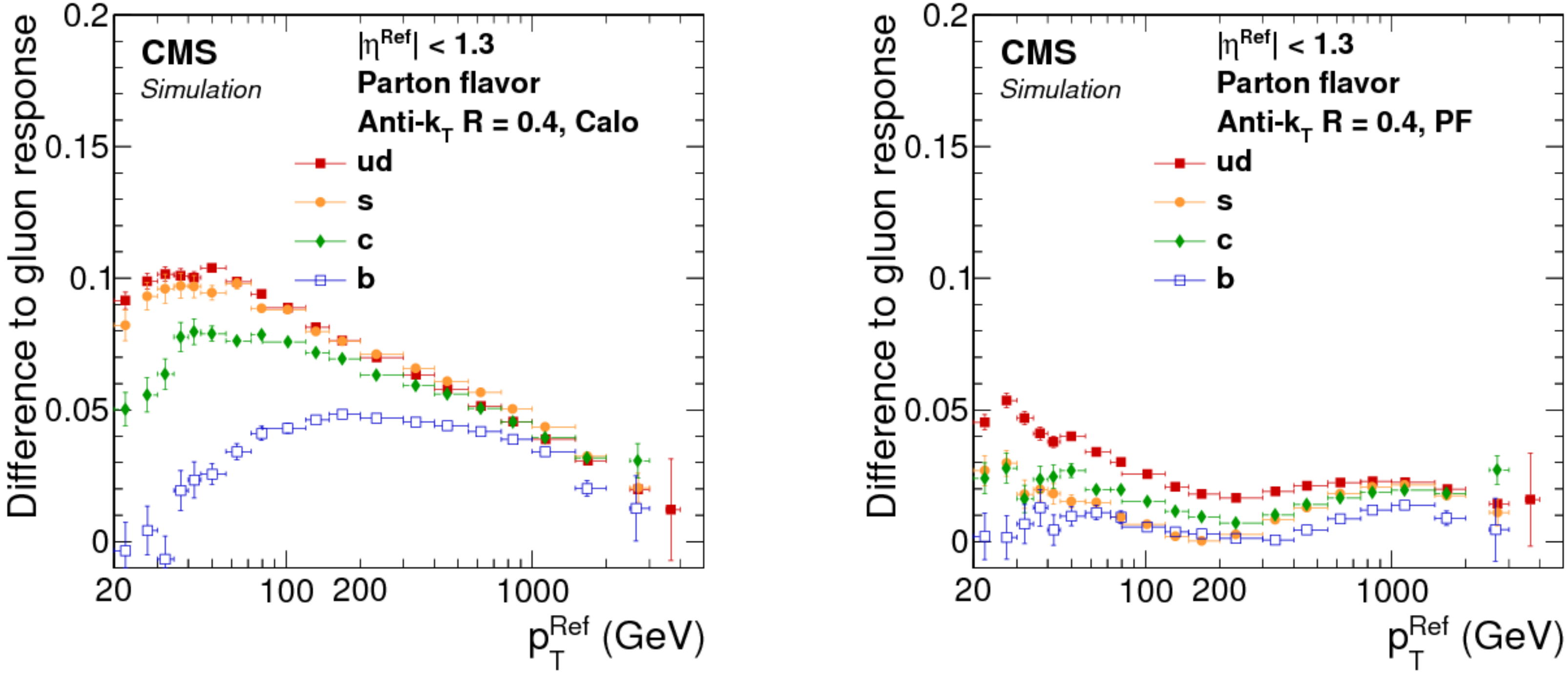
Jet energy resolution





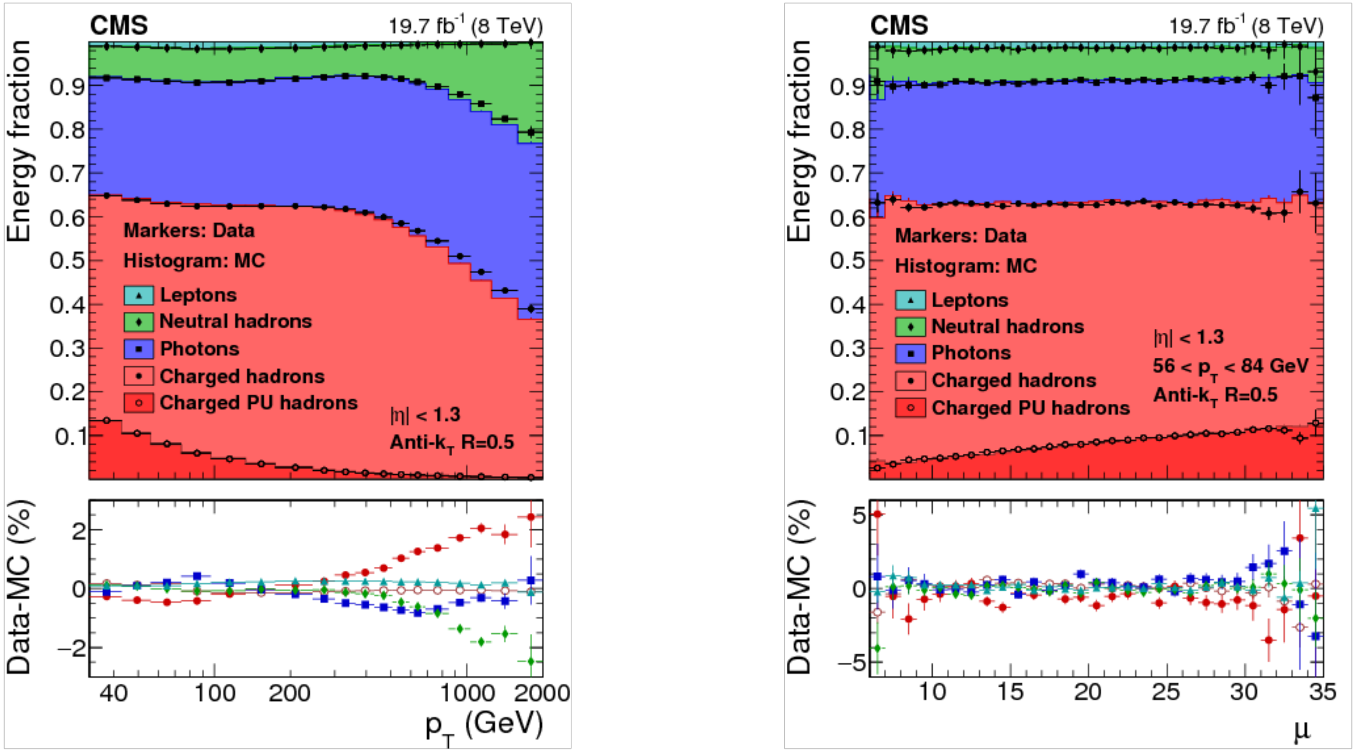
- Raw response is already close to unity, reducing size of jet energy correction
- Jet energy resolution improved, especially at low p_T where tracker dominates

Jet flavor sensitivity



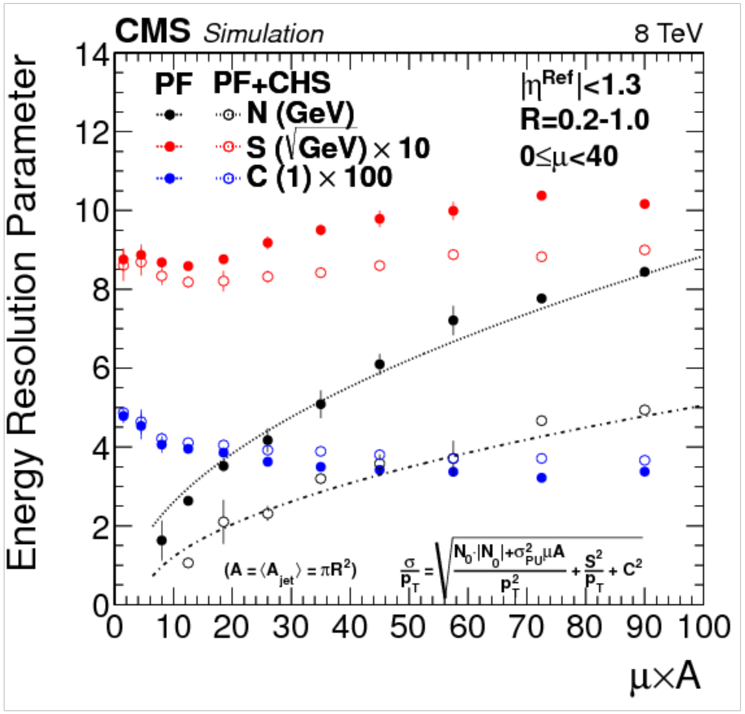
- Flavor dependence is one of the leading contributors to JES uncertainty
- Reduced by ≈ 2x at low-to-mid p_T

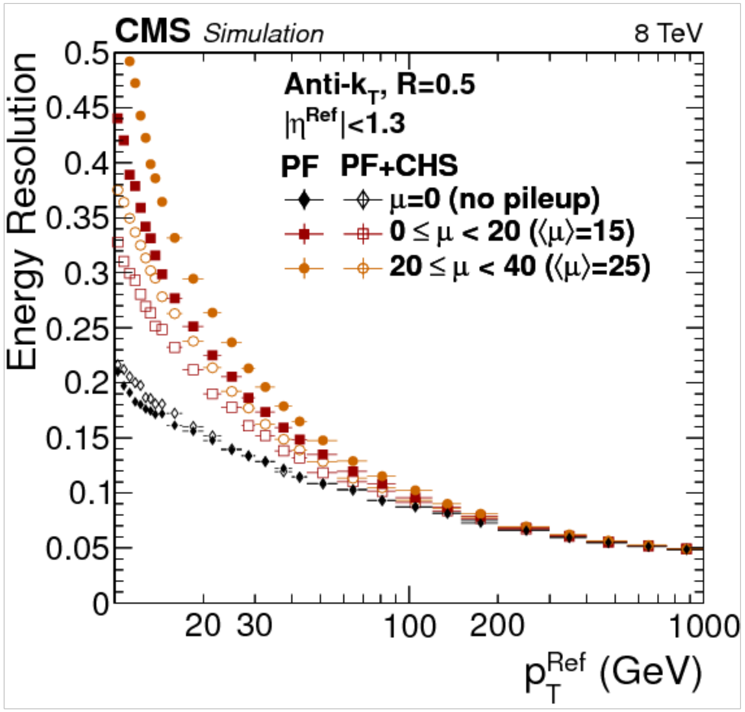
Particle composition



- Important to test accuracy of simulation by checking particle composition
- Within 1% until very large jet p_T
- NB: Baseline JECs derived from MC, but residual data/MC scale factors obtained from dijet & boson+jet balancing (described here)

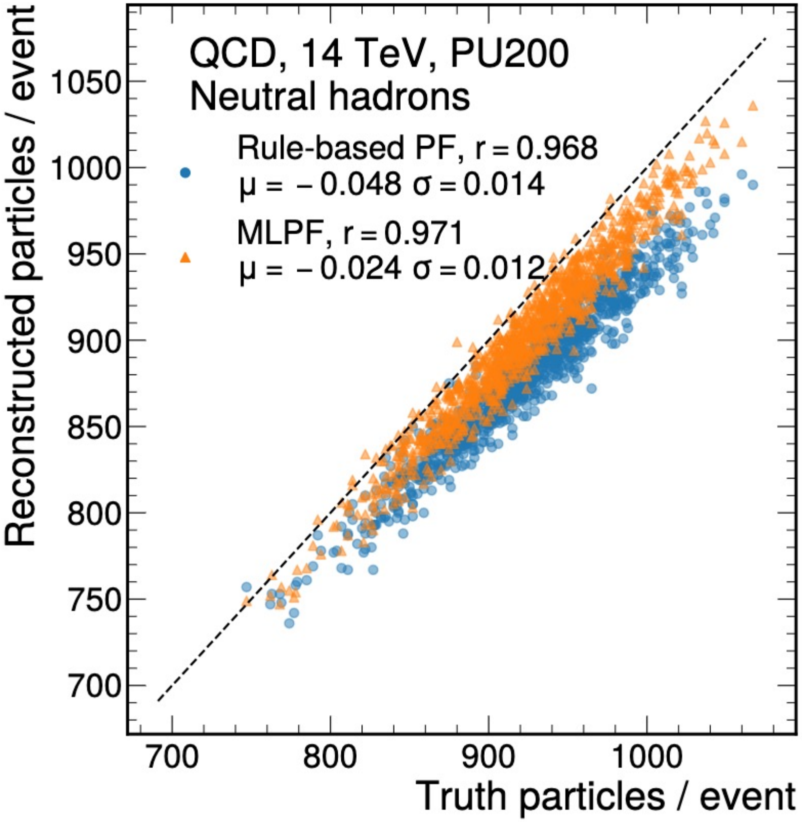
Pile-up mitigation



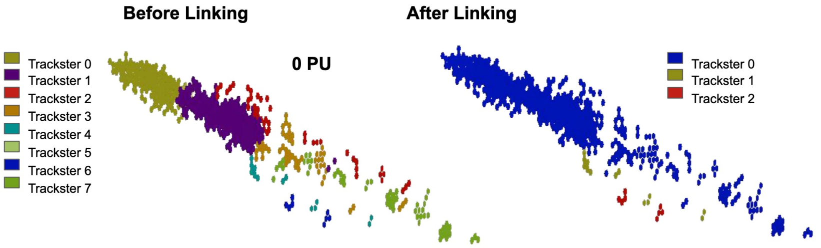


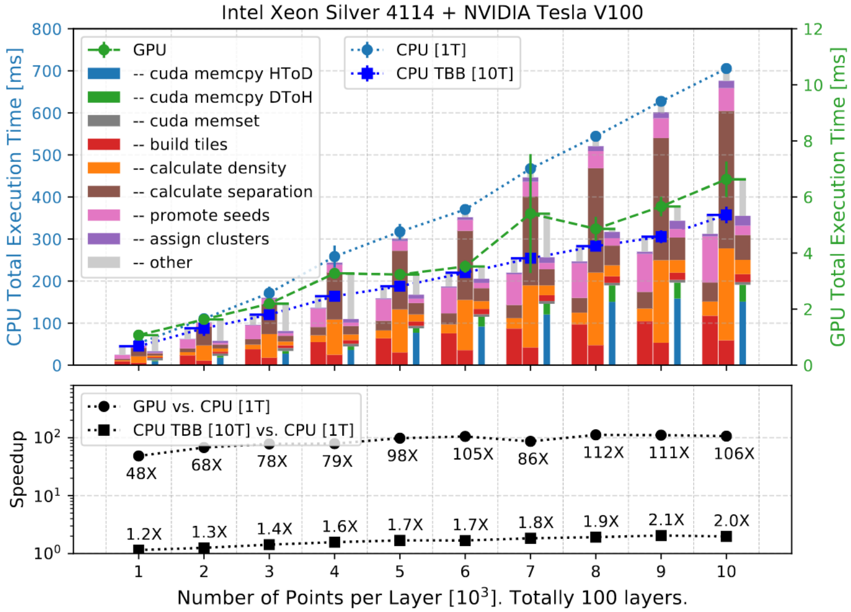
- Charged hadron subtraction (CHS): $h^{+/-}$ from pile-up removed
- More advanced method uses proximity to $h^{+/-}$ to also mitigate effector of PU on neutrals ("PUPPI")

Advanced topics



[ML-based PF](#)





Clustering (CLUE) and linking (TICL) for High Granularity Calorimeter, written for heterogenous architectures

Conclusions & Outlook

- Particle flow reconstruction provides the default event interpretation of CMS
  - A first for hadronic collisions
  - Despite a detector not designed for PF with several short-comings (thick tracker, modest HCAL segmentation, etc.)
- Particle flow improved performance
  - Of physics objects: jets, MET, tau, etc.
  - And mitigating effects of pile-up
- Expect better performance w/ detector designed for PF, e.g., Phase-2 CMS w/ HGCAL
- Elements from the CMS PFA may be useful ePIC