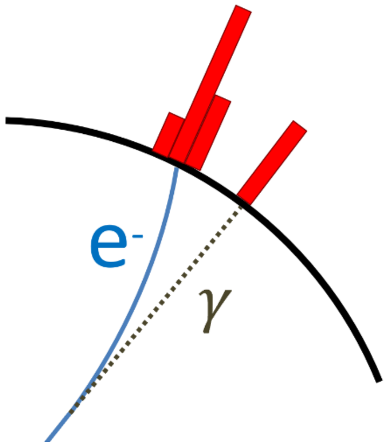
A brief overview of CMS E/gamma Reconstruction



Sam Harper (STFC-RAL)

08/12/22

Dressed Leptons Workshop, IPPP, Durham

Introduction

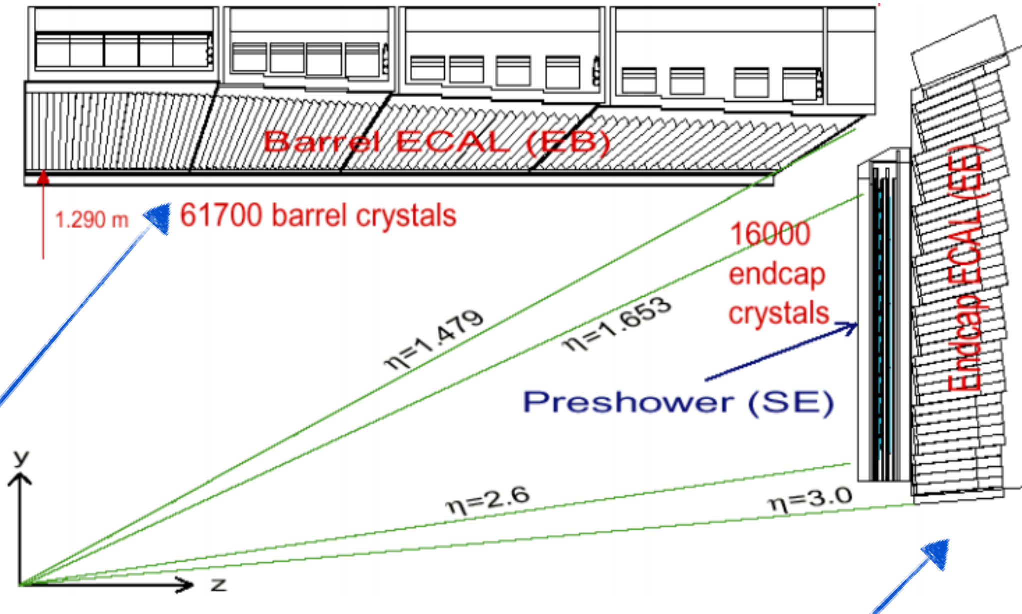
- goal: give a basic understanding of CMS e/gamma reconstruction, with special emphasis on what is being clustered
- this talk will focus on electrons, photons reconstruction wise are mostly identical except there is no associated track
  - a few differences will be pointed out
- CMS has two papers out which are relevant here:
  - Particle Flow: https://arxiv.org/abs/1706.04965
  - E/gamma: https://arxiv.org/abs/2012.06888

A quick primer on GED / PF

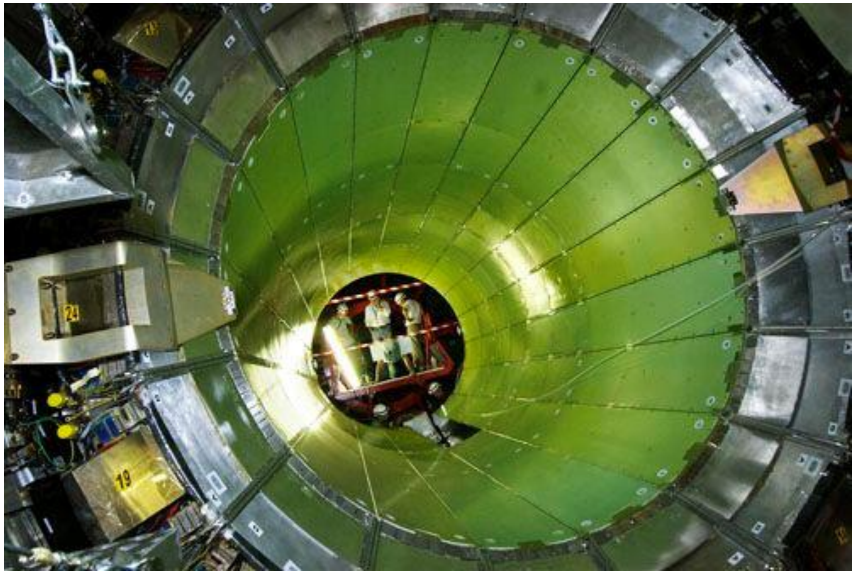
- General Event Description / Particle Flow is the CMS way of reconstructing particles
- CMS first reconstructs tracks, ECAL clusters, HCAL clusters and muons
- It then then feeds the tracks, clusters to the E/gamma algos which try to make electrons/photons out of them with full bremstralung recovery
  - these are saved in the E/gamma collections
- these are then passed to PFAlgo, which will accept or reject the electron/photon as valid
  - if rejected, the ele/phos tracks & clusters are released to PFAlgo to do as it sees fit, if accepted, the associated tracks/clusters are not available to PFAlgo
- PF Algo then creates charged hadrons, non isolated photons and neutral hadrons from the available tracks & clusters in the event
  - the clusters/ tracks originally belonging to the ele/pho found and kept by e/gamma is recorded and tracked to avoid double counting, however this is imperfect
  - photons here are just a single EM cluster and are only the excess of any associated track

The CMS ECAL

- homogeneous calorimeter
- Measures the energy of electrons and photons and the electromagnetic fraction of jets
- Total weight: 88.7 tonnes



BARREL

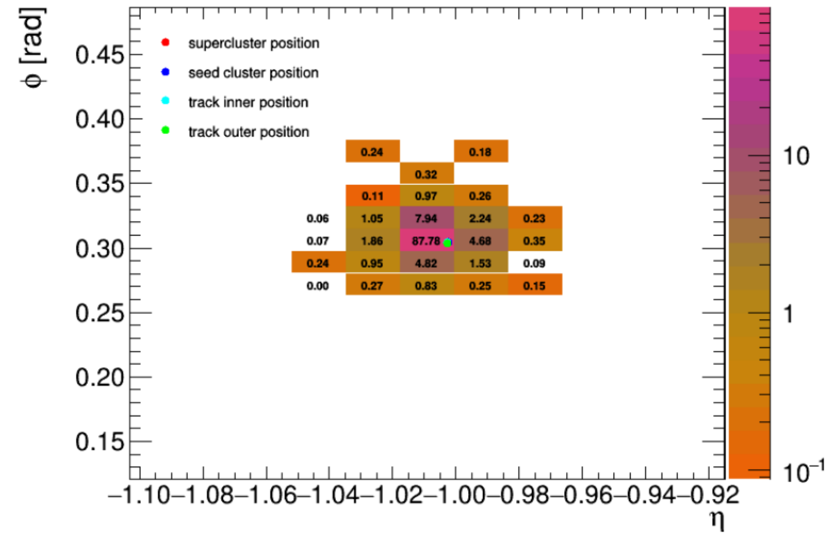


ENDCAP

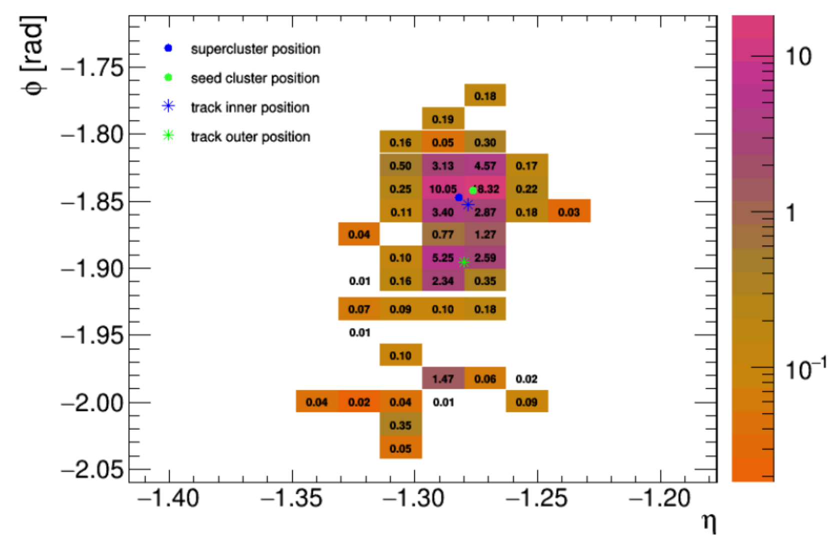
CMS EM Particle Reconstruction

- CMS ECAL is made out of 22x22mm lead crystals (barrel), 28.6x28.6mm (endcap) for a total X_o of 25.8 (24.7)
  - in barrel corresponds to 0.0174 eta x 0.0174 phi
  - endcap it varies
- Moliere radius ~ 22mm, therefore ~90% of the shower's energy contained within a two crystals if incident between them
- a 5x5 array of crystals contains almost all of an unconverted photons energy
  - base unit for how CMS thinks about individual em particles

MC electron, no brem example



MC electron, brem example

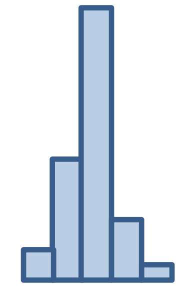


CMS EM Reconstruction

- RunI CMS reco algos simply formed a 5x5 grid on top of each local maxima above a given threshold
- from RunII, we use a clustering algorithm which can resolve overlapping particles: Particle Flow Clustering
  - it has nothing to do with particle flow, its merely the algo the particle flow algorithm used
  - it has marginal impact in CMS E/gamma reconstruction, the overlapping particles are summed back together
- each local maxima with crystal energy > 1 GeV is clustered under the hypothesis that it is a result of a incident EM particle
  - energy is shared between nearby local maxima under the hypothesis each represents a in individual incident em particle

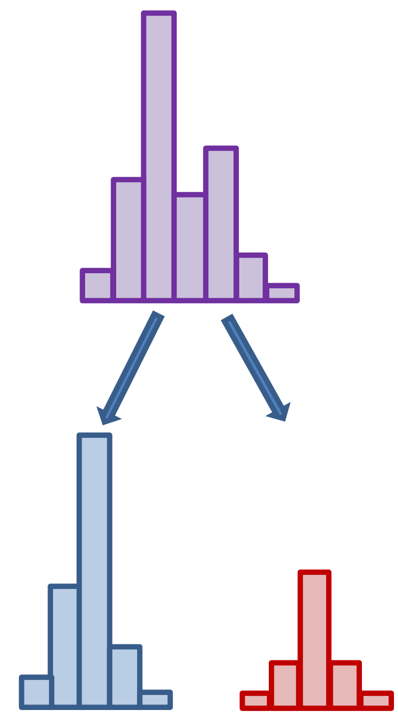
each block represents a CMS ECAL crystal

single EM particle



two close by EM particles

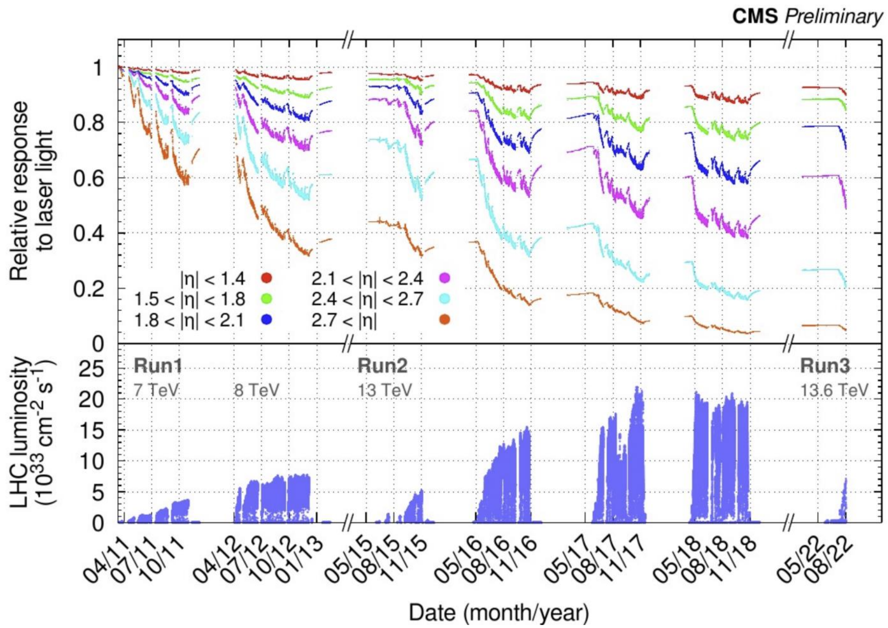
cluster algo sees two local maxima, shares the energy in the crystals between the two clusters



Clustering Thresholds

To combat detector noise, there are minimum energy cuts for:

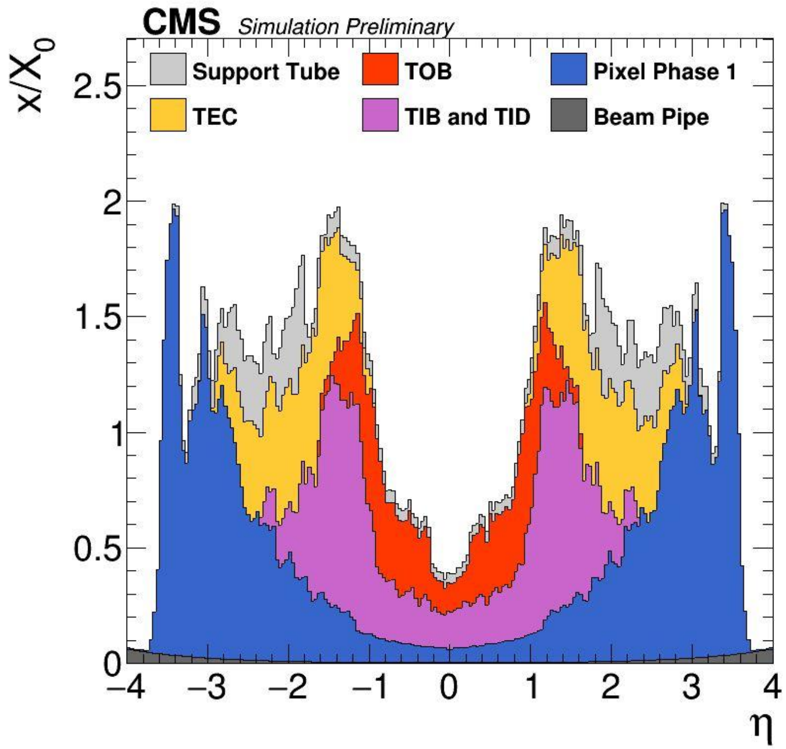
- a crystal to included for clustering
  - this varies across the detector, becoming larger at higher eta
  - of order ~0.35->0.55 GeV barrel and mostly 0.7 GeV in endcap but varies strongly across the detector (0.2 -> 50 GeV)
    - represents about 2xnoise threshold
- a crystal to be considered a local maxima:
  - barrel: E> 0.23 GeV
  - endcap: E > 0.6 GeV && $E_T > 0.15$



- response of ecal to laser light
- inverse of this is applied to measured energy -> larger correction more noise

Brem Recovery

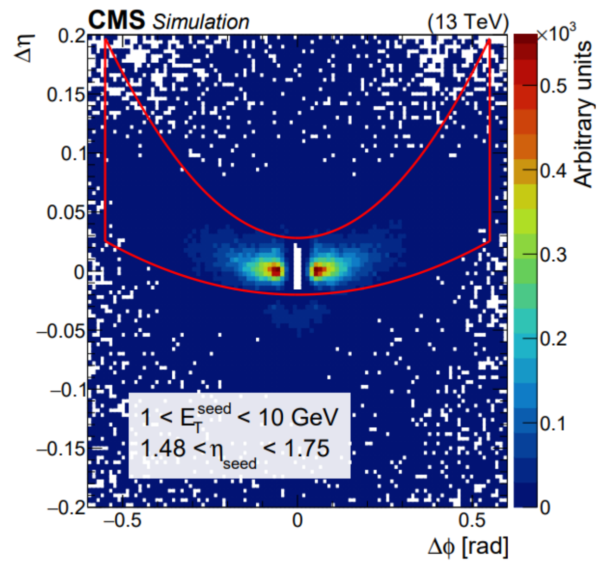
- significant material in front of the ECAL, most electron/photons are showering on arrival at high eta
- a single e/g appears as multiple clusters separated in phi
- brem recovery: **primarily geometric windows** in the calorimeter with **refinement from track extrapolations**
  - windows are narrow eta, wide phi
- any FSR recovery is an accidental side effect of this process

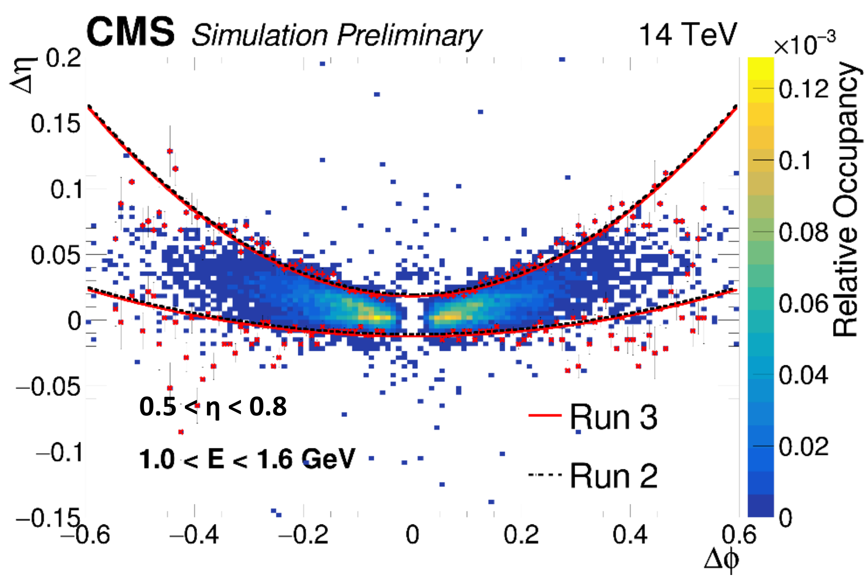


https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackerMaterialBudgetplots

Brem footprint in the calorimeter

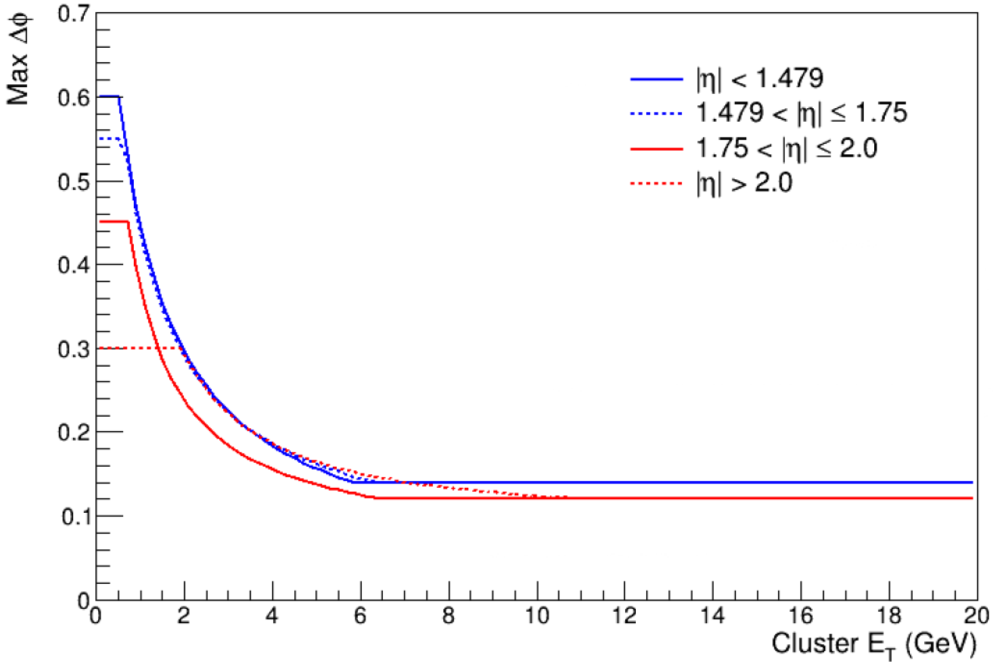
- as well as bending in ϕ, there is a small bending in η
- we use the **“mustache” superclustering** algo which accounts for this geometric spread
- effectively takes all the clustered energy in a given energy dependent η / ϕ window
- requires a seed PF cluster $E_T > 1$ GeV





Moustache Window: $\Delta\phi$

- max $\Delta\phi$ w.r.t to the seed cluster for a cluster to be included in the supercluster as function of cluster E_T



maxDPhi = Y+scale/(1 + exp(logEt – X) / width)

CMS Photon Reconstruction

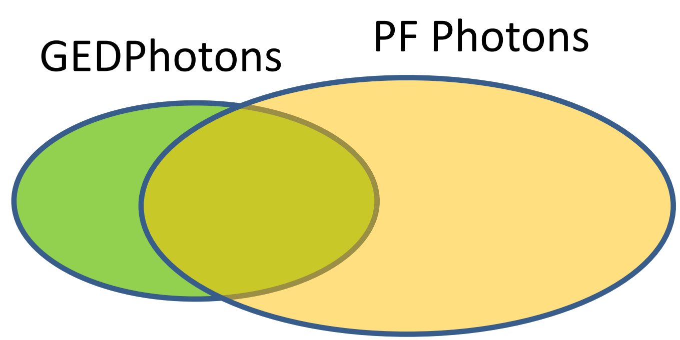
CMS Photon reconstruction is highly confusing, there are two separate algorithms and two separate collections which overlap

GEDPhotonAlgo:

- makes traditional isolated photons, $E_T > 10$ GeV with H/E requirements
- effectively consist of the refined moustache superclusters
- handled by the E/gamma group who assume everything is a photon / electron
- if passes PFEGamma ID (isol/showershape), it will enter PF as a single photon, if it fails, its broken into its compontent parts which are then available to the PFAlgo to do as it sees fit

PFPhotonsAlgo:

- basically makes the EM component of jets
- consist of a single PF ECAL cluster (no showering recovery)
  - corrected with a very simple BDT to take in account detector response mostly due to energy thresholds and gaps
- input: all ECAL clusters not associated to GEDPhotons/Electrons
- handled by PF Group who assume everything is a charged hadron until proven otherwise
  - if there is a track associated to the cluster, it will be IDed as charged hadron, a PF photon will only be produced from any excess in the ECAL over the track p_T



any photon < 10 GeV is at the mercy of the PFAlgo which will try to make it a charged hadron if possible

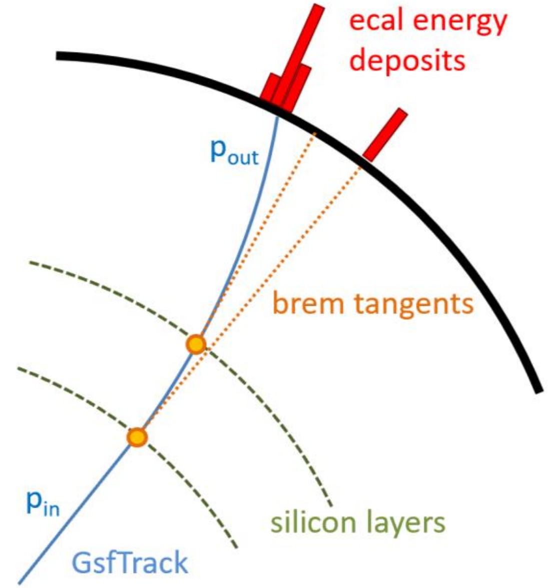
CMS Electron Reconstruction

- "standard" electrons are reconstructed in one of two ways at CMS
  - tracker driven : starts from a track, looks for a calorimeter deposit
  - ecal driven : starts from calorimeter deposits, looks back for a track, starting in the pixels
- ECAL driven is primary algo:
  - tracker driven mostly recovers soft electrons, areas where pixel modules are inoperative
  - most electrons are both ecal and tracker driven
- will focus mostly on the ECAL clustering steps of ECAL driven process and then the common "refining" procedure
- note: there is a class of "low pt" electrons used for b-parking analyses, mostly targeting below 5 GeV
  - effectively a track and some cuts, very little brem recovery attempts
  - unlikely to be relevant to todays topic

Electron Track Reconstruction

not to scale and only showing two tracker layers

- electrons at CMS use a dedicated tracking algorithm known as GSF tracking
  - GSF = Gaussian Sum Filtered
  - electron specific hypothesis
- GSF tracking explicitly takes into account radiative losses due to brem so we can measure p_{in} and p_{out} (and p at intermediate layer)
- use these measurements to be able to product brem tangents for later matching to missed ECAL deposits



p_{in} = initial (or inner) momentum of the electron before it traverses the tracker

- ie before brem, so original momentum of electron

p_{out} = final (or outer) momentum of the electron after it has gone through the tracker and radiated photons

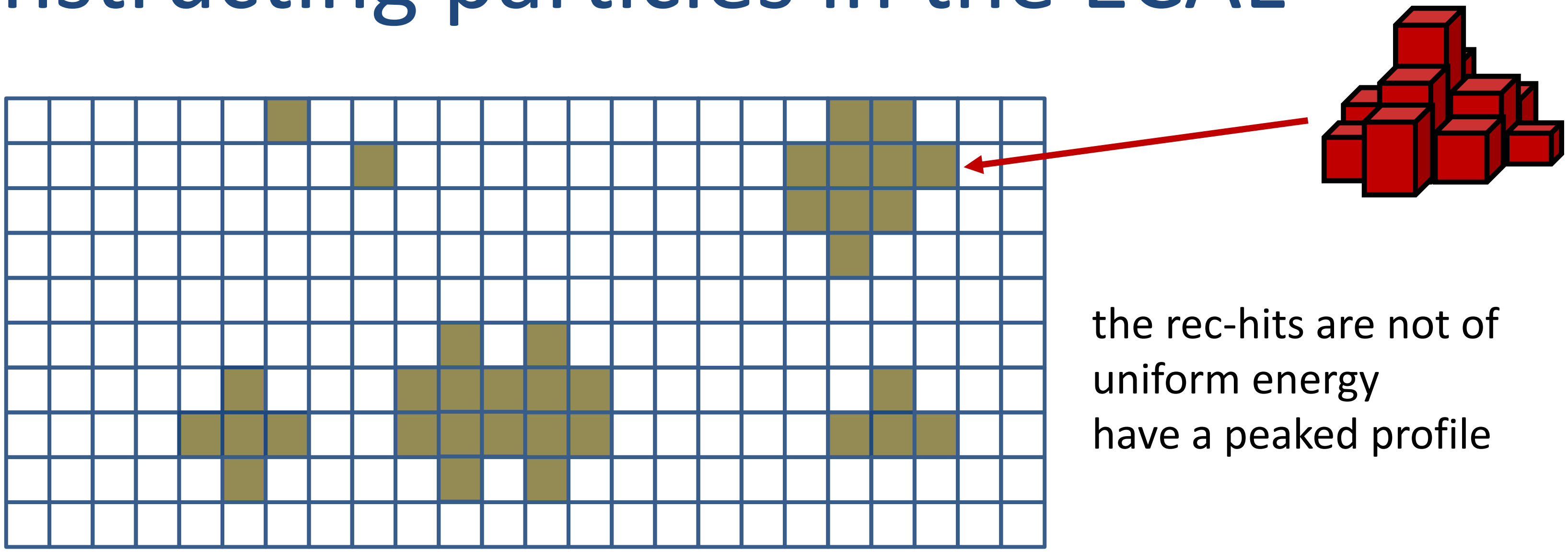
SuperCluster Refining

- these are the final superclusters used by electrons/photons
- use particle flow techniques to further refine which clusters belong to the supercluster
- uses moustache superclusters as the starting point (all clusters of a moustache SC are candidate clusters for the refined)
- in the case of a tracker driven electron with no moustache, it will make a supercluster from scratch
  - fairly rare case, means very little energy found in the ecal
- note: electrons and photons share the same refined supercluster but if a GsfTrack is found, this is exploited to better associate brems with the supercluster

SuperCluster Refining

- links additional clusters associated to:
  - the GsfTrack
  - associated primary tracks
  - associated secondary tracks (conversions)
  - brem tangents from the GsfTrack
- then it unlinks ECAL clusters that mean there is a bad E/p
  - note: clusters within $\eta<0.05$ of the GsfTrack and $\eta < 0.015$ with a respect to a brem tangent are never unlinked

Reconstructing particles in the ECAL



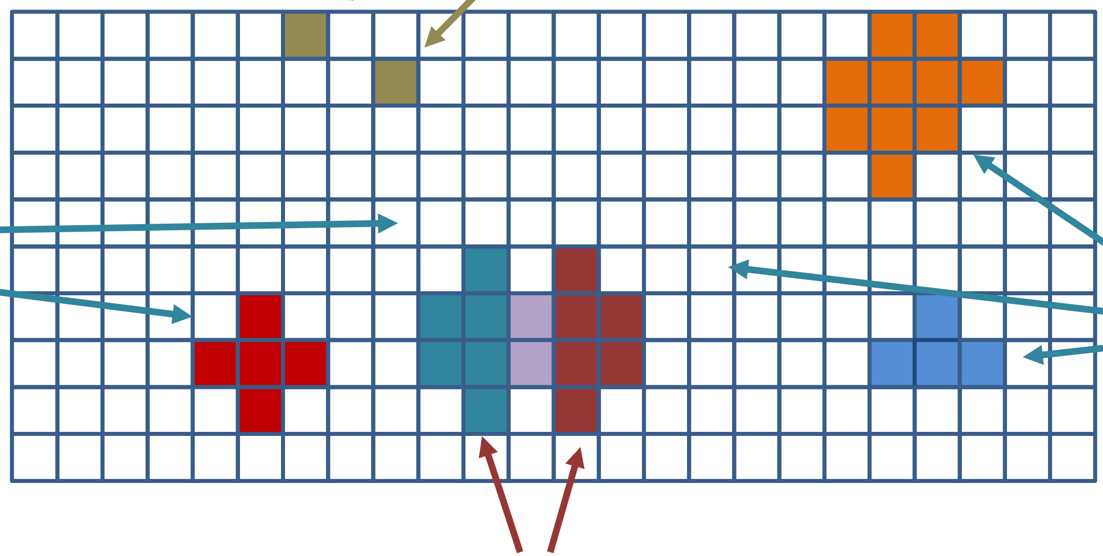
the rec-hits are not of uniform energy
have a peaked profile

- once we have the crystal energies reconstructed, need to construct individual particles from them
  - this is the PFClustering step
- PFClustering looks for a local maxima above a given threshold (X GeV)
  - each local maxima in energy is assumed to represent a cluster
    - referred to as the seed crystal
  - algo is smart enough to share hit energies between overlapping adjacent clusters

After PF Clustering

unclustered energy, just noise

PF Clusters corresponding to electrons / photons



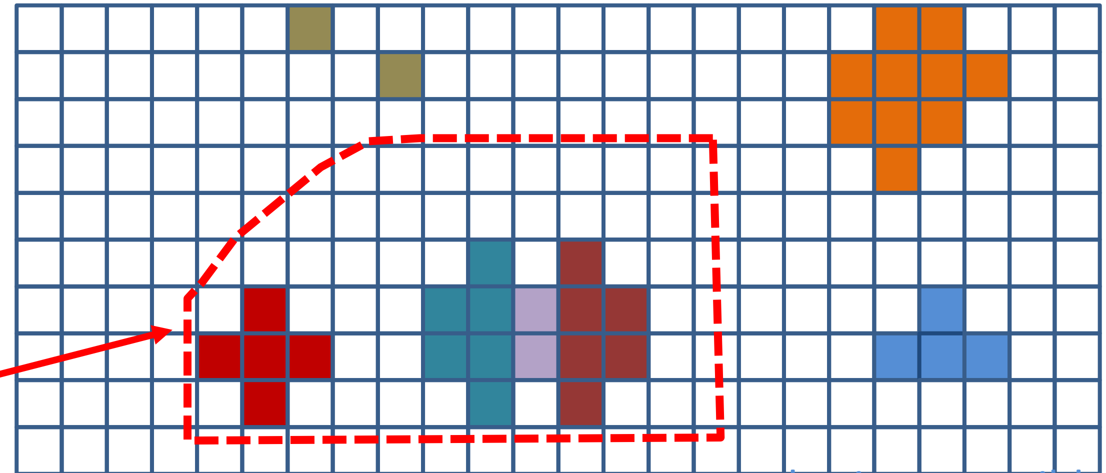
PF Clusters corresponding to electrons / photons

these two clusters overlap, clust algo shares energy of purple rec-hits between the two clusters according to a Gaussian energy profile, each gets a fraction of the rec-hit energy

- found 5 PF Clusters in this region of the ECAL
- each one represents a the energy deposits of a particle, such as a pho or ele
- in practice not this clean, energy sharing mean clusters often have very large number of this with very small energy fractions
  - the seed will always have fraction =1

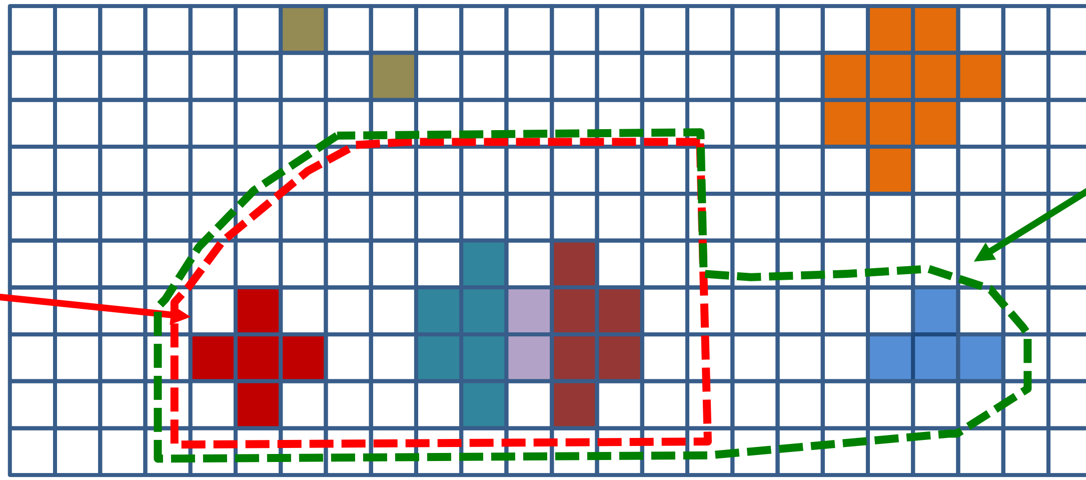
SuperClustering

the PF Moustache superclustering algo looks in ϕ and a little in η to combine the 3 PF Clusters into a single supercluster representing the electron



- an electron or photon can actually consider of many electrons and photons due to pair conversion / bremstrulung inside the detector
- superclustering (clustering on clusters) aims to combine the individual electrons + photons into a single object
- starts by taking our highest energy cluster (in this case the maroon one) and looks for compatible clusters in ϕ, η
  - due to B-field, bremsstrahlung can have large difference in ϕ, much smaller difference in η
  - moustache supercluster algo accounts for the bend in η
    - box supercluster algo (no longer used) does not
  - in the endcap preshower (aka ES) clusters will be associated here

Refined Superclusters



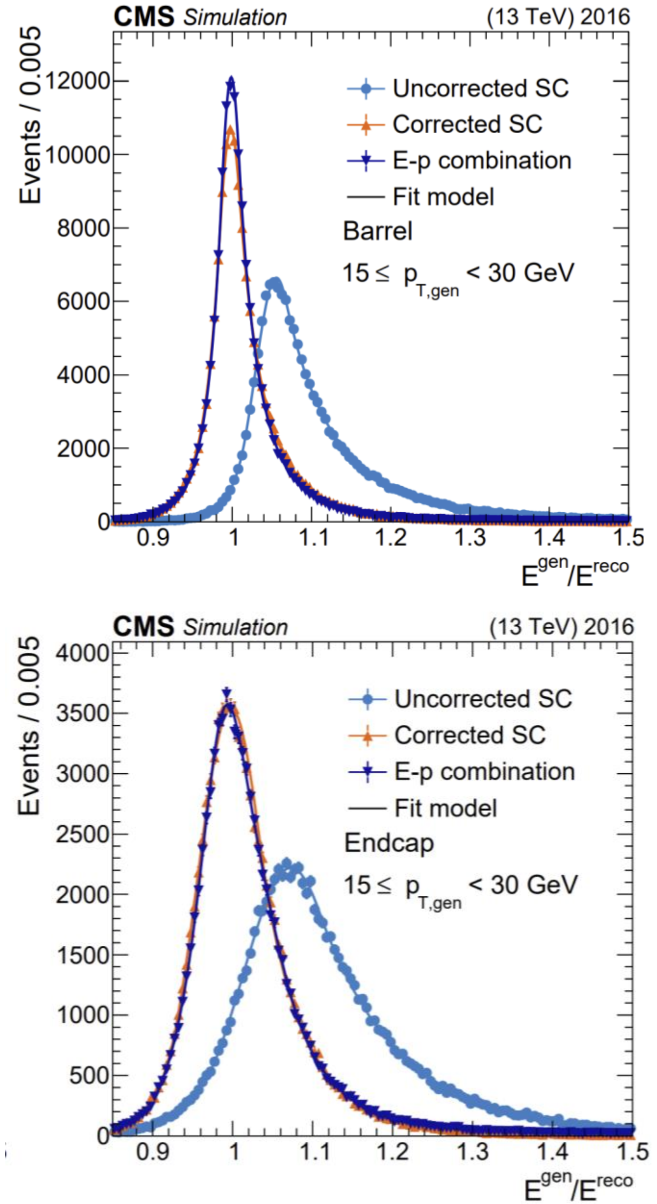
- so far we have only considered the ECAL (+ES) for brem/conversion recovery
  - but we have a tracker! surely that can help tracking the
- use particle flow to further associate soft brems to the moustache supercluster
  - sees if there are compatible tracks, looks at the flow of particles through the event
- this is a **refined supercluster**
  - these are the superclusters which are used for all "supercluster" related things in e/gamma
  - GsfElectron/Photon::supercluster() : refined supercluster
  - GsfElectron/Photon::parentSuperCluster() : original moustache supercluster

ID requirements

- ID requirements will actively reject electrons / photons which it thinks are spread in η
  - about ~95% of energy within two strips of η or ~85% in one strip in the 5x5 area centred on the highest energy crystal
- isolation requirements will kill anything significant outside clustering region (typical cone size is 0.3 but this can vary)

A Word on Energy Corrections

- CMS uses ML to correct the energies back to the gen level energy
  - currently using a BDT but a GNN is planned to be deployed
- training is on a uniform (p_T ,η) sample of “particle gun” electrons, ie no FSR as not from a physics source
  - so w.r.t to bare lepton
- accounts for all possible sources of miss measurement
  - intermodule gaps, PU contamination, dead crystals, min crystal energy thresholds etc etc
- there is a residual correction to scale the data and MC to each other at the Z peak (1-2% effect)



Other considerations

- a team in CMS is developing a DNN based superclustering algorithm to recover beam
  - https://arxiv.org/pdf/2204.10277.pdf
- so far not deployed and I personally have yet to see any practical advantage beyond a final correction
- it is unclear how this will interact with FSR
  - trained on a particle gun sample without FSR

Summary:

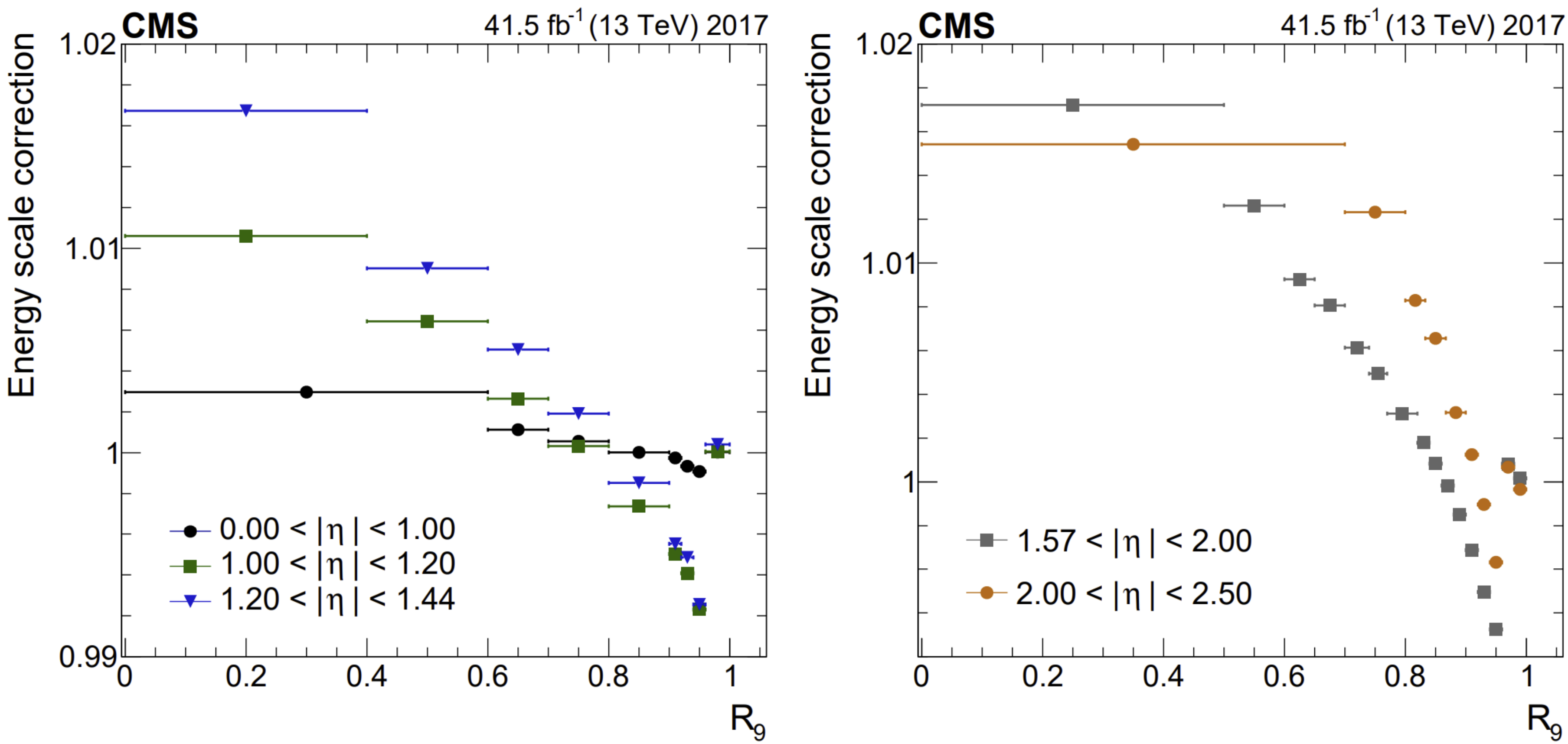
- CMS electrons and photons are narrow objects in η, wide in ϕ
  - some bending in η at high ϕ due to B-field
  - otherwise actively reject electrons/photons spread in η
- thresholds:
  - to enter PF cluster: crystal E > 0.33 to 0.55 GeV (barrel), E>0.2 to 50 GeV with most at 0.7 GeV (endcap) GeV
  - to seed PF cluster: crystal E> 0.23 (barrel) E > 0.6 GeV && $E_T > 0.15$
  - to seed supercluster: cluster $E_T > 1$ GeV
  - to become a SC: SC $E_T > 4$ GeV
  - to become an isolated Photon: $E_T > 10$ GeV

Science and Technology Facilities Council

Any Questions

backups

Residual (aka Z) corrections



Bremstalung

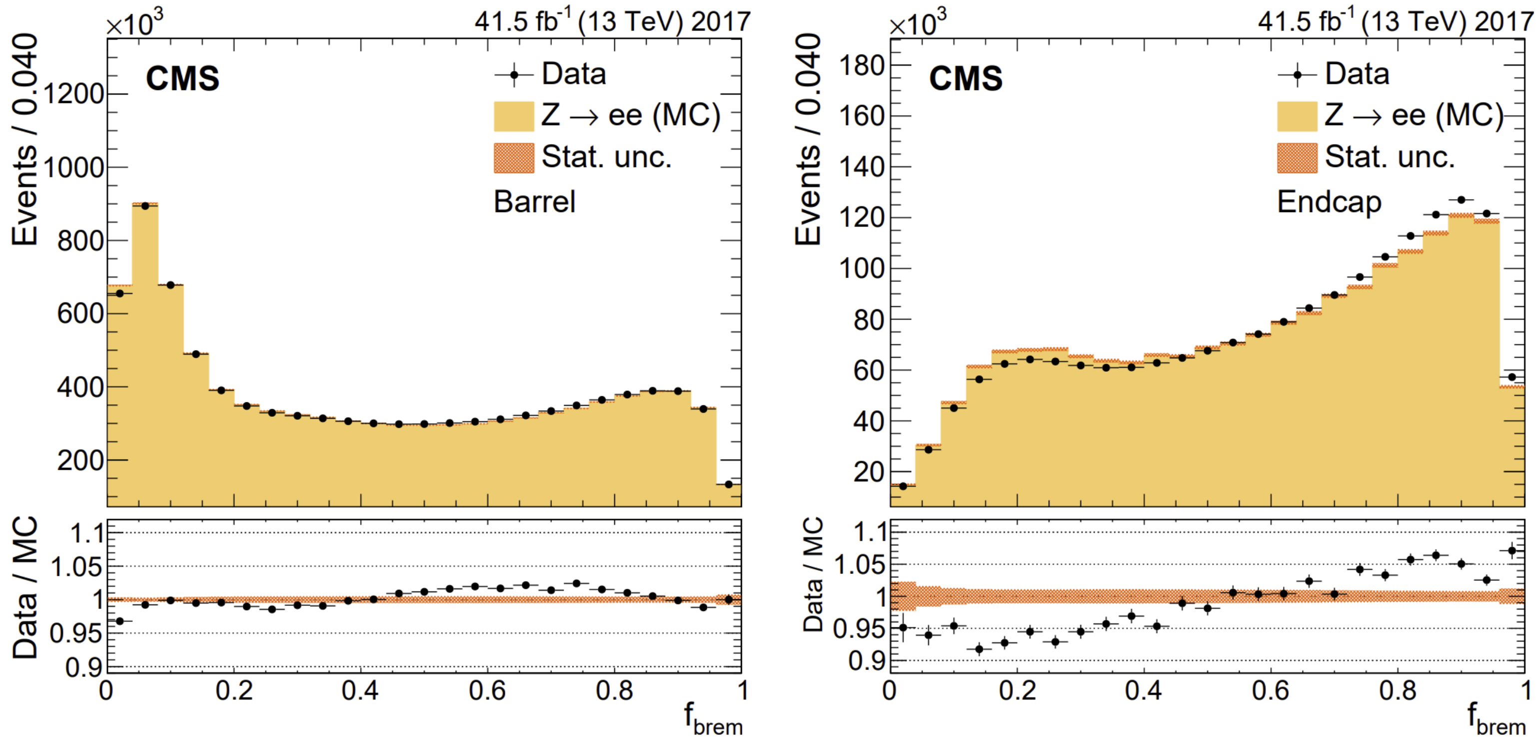
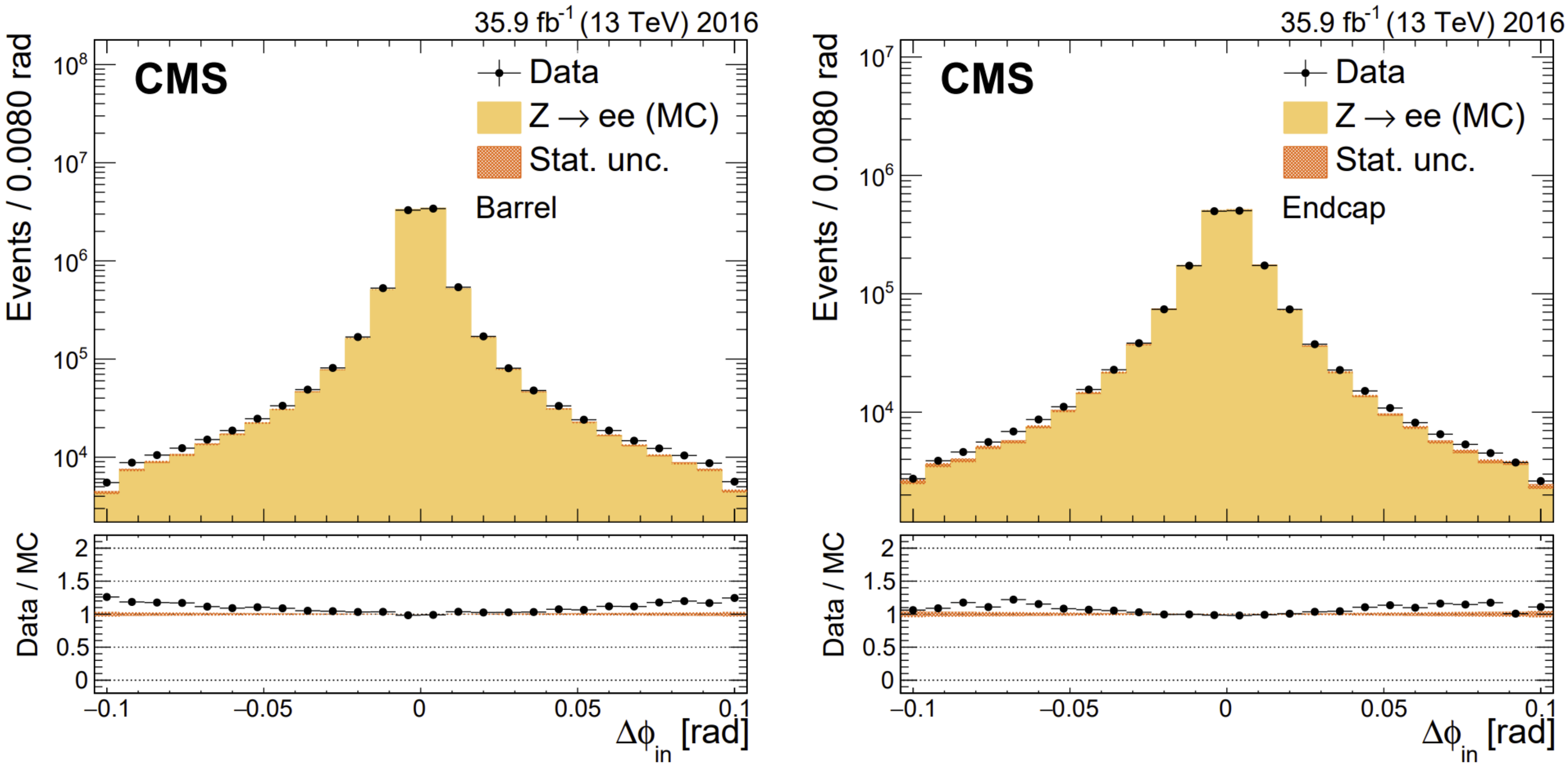


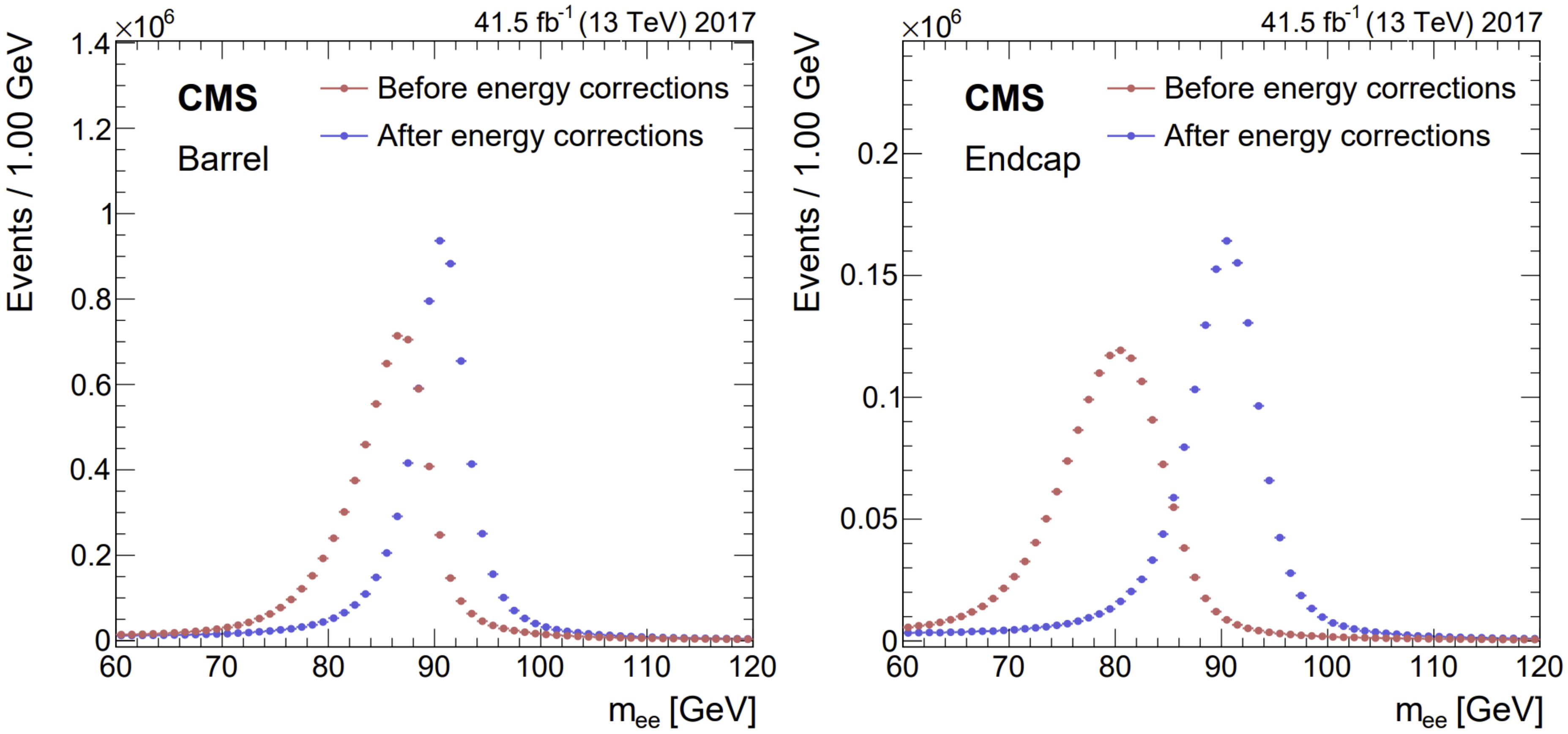
Figure 2: Fraction of the momentum lost by bremsstrahlung between the inner and outer parts of the tracker for electrons from Z boson decays in the barrel (left) and in the endcaps (right). The upper panels show the comparison between data and simulation. The simulation is shown with the filled histograms and data are represented by the markers. The vertical bars on the markers represent the statistical uncertainties in data. The hatched regions show the statistical uncertainty in the simulation. The lower panels show the data-to-simulation ratio.

Track SuperCluster Matching

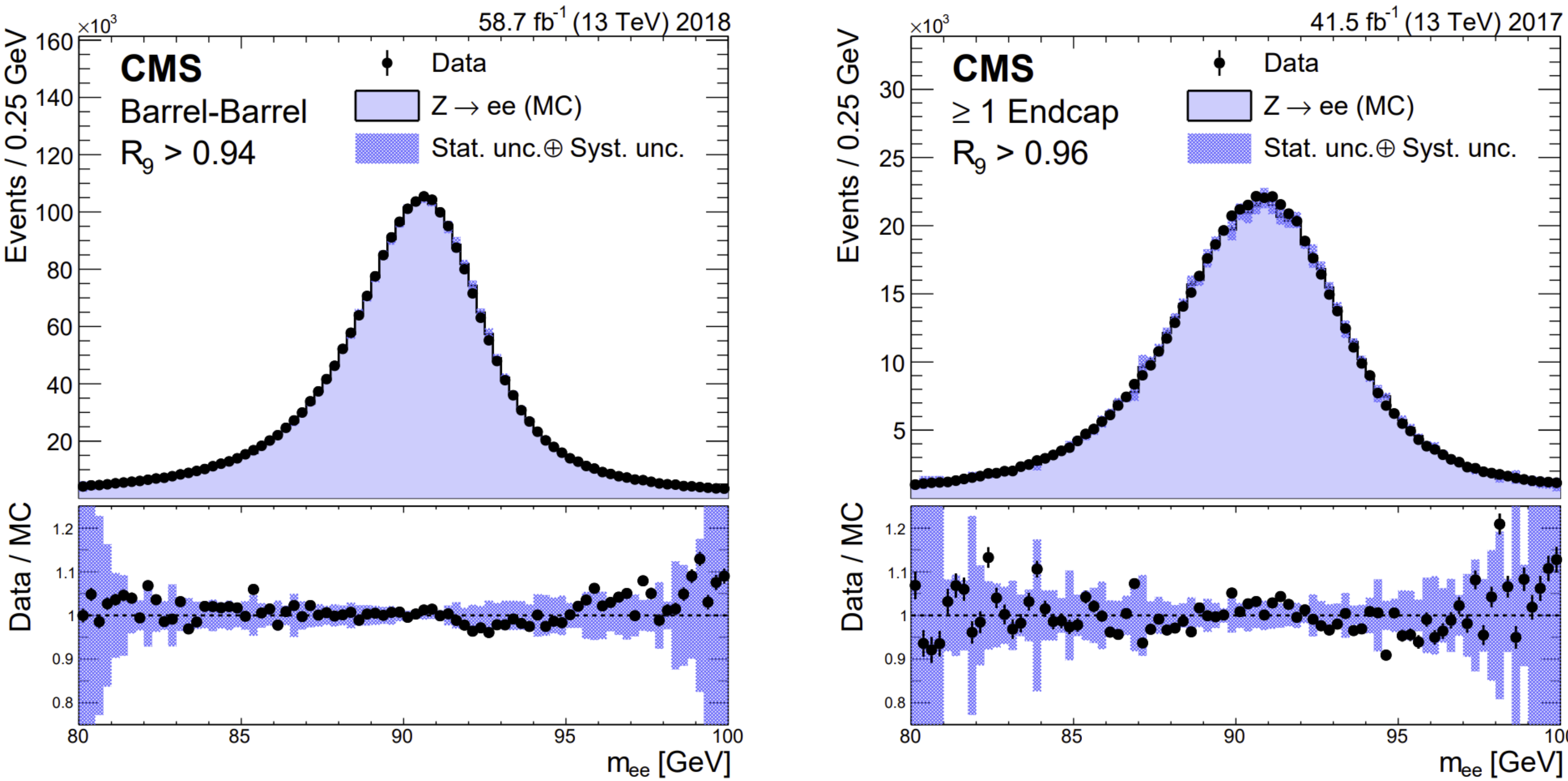


- track position at inner state extrapolated to the calorimeter

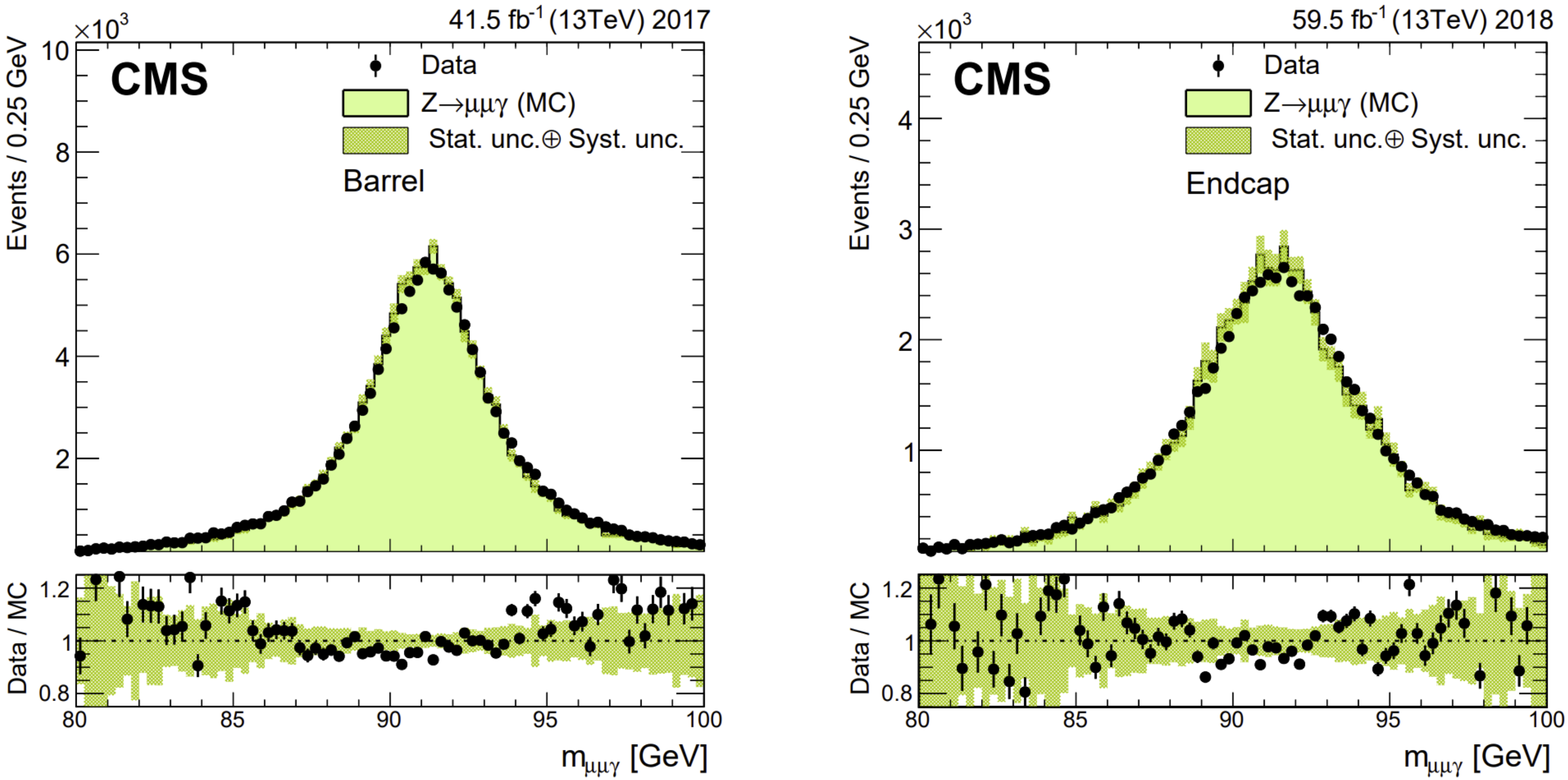
Energy Resolution Before / After Corrections



Z peak comparisons



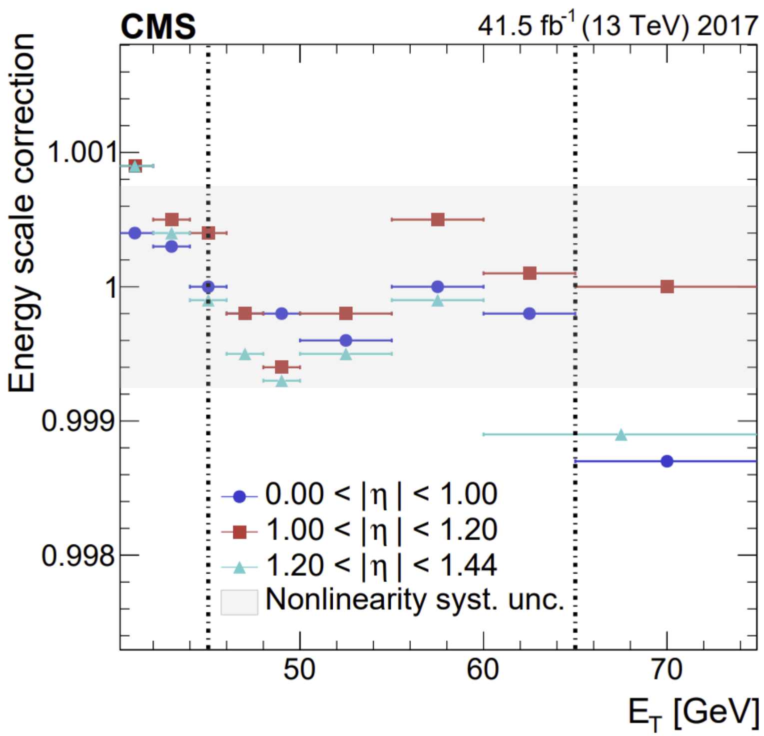
Zmmu gamma



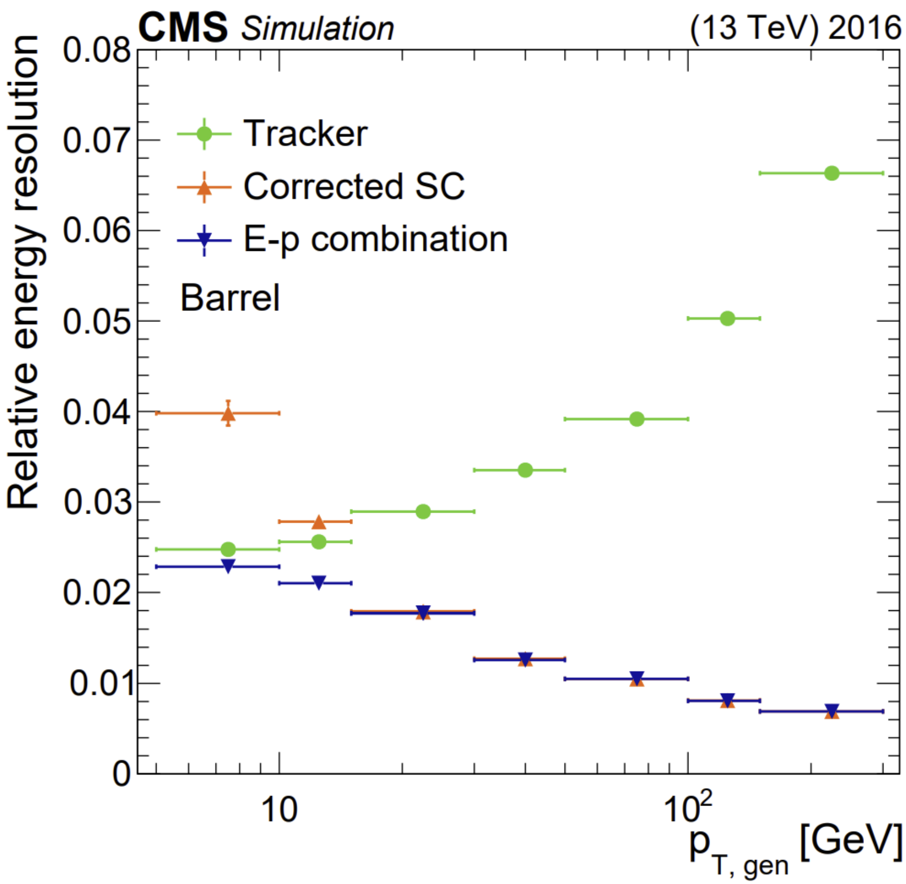
- we use this to validate the photon energy scale

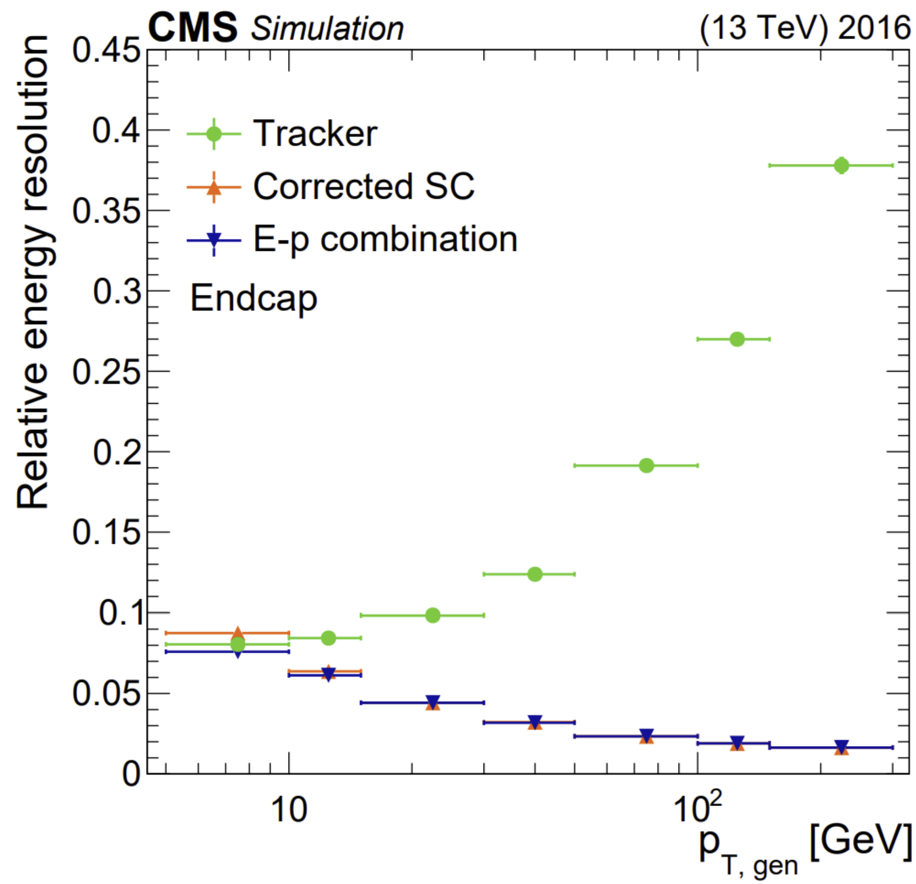
Photon Energy Corrections

- Use Zmumu gamma to validate the photon energy scale corrections obtained with electrons

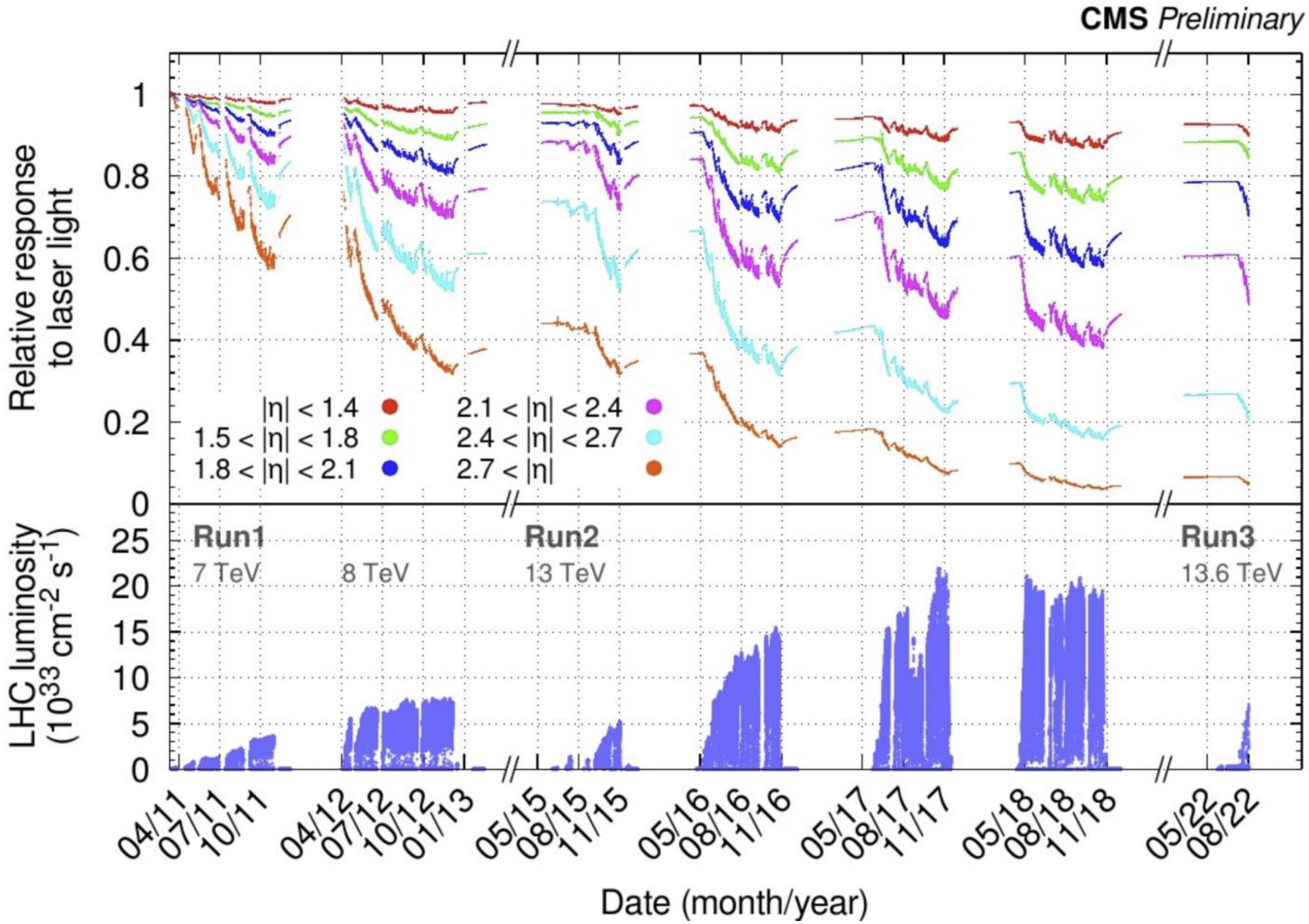


Electron Energy Resolution





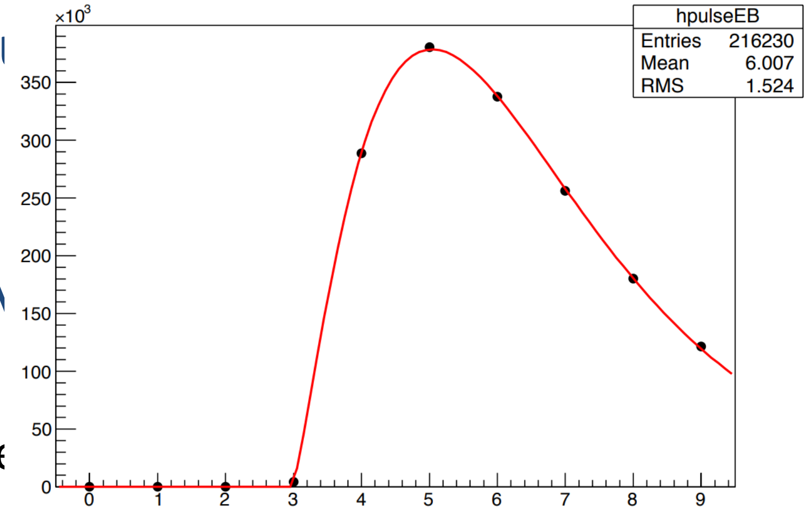
ECAL Response

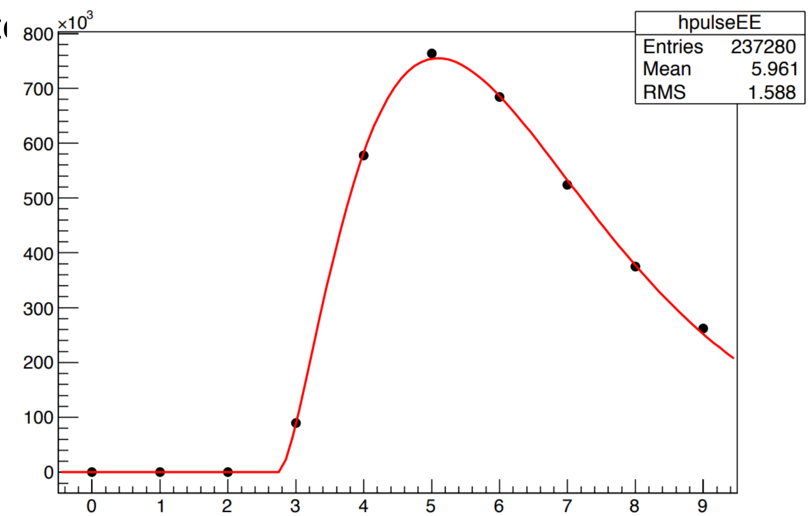


- the inverse of this is applied to the ecal energy to correct to the true value
- this amplifies the noise, leading to threshold increases

ECAL energy reconstruction

E. Di Marco, J. Bendavid
https://indico.cern.ch/event/292930/contributions/671061/attachments/547860/755142/edm-ecalreco-pu-21Aug2014.pdf



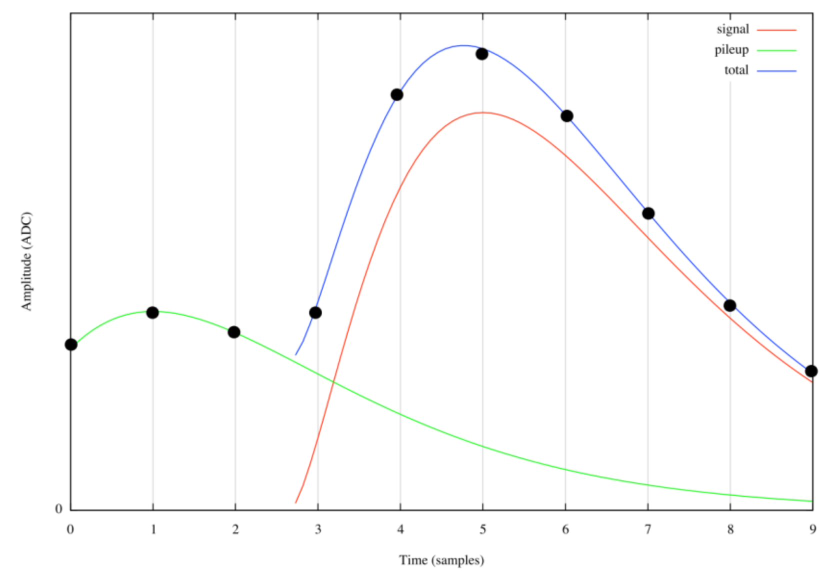


- energy in the ECAL observed as a pu the shower arrive over time
  - need to convert this pulse shape to an
- show example pulse shape of an EM crystal
  - this case is easy to reconstruct, use the
  - $A = \Sigma w_i S_i$ where
    - A is the amplitude of the pulse (ie related t
    - w_i = weights, computed for a known pulse
    - S_i = sample values

ECAL energy reconstruction

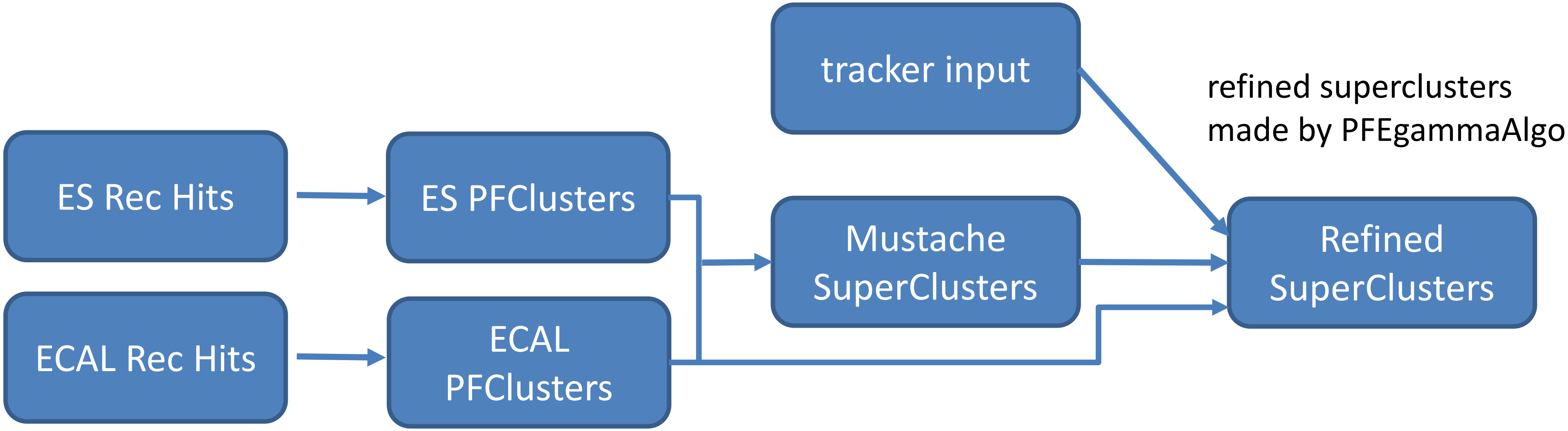
E. Di Marco, J. Bendavid
https://indico.cern.ch/event/292930/contributions/671061/attachments/547860/755142/edm-ecalreco-pu-21Aug2014.pdf

- now lets add in some out of time pile up (OOT)
- therefore multiple pulses from dif
  - the tails of the OOT pile pulses lead t amplitude measurement
- solution: multifit reconstruction
- multifit tries to resolve the amplit by fitting multiple pulse shapes si
  - this means the amplitude of the in ti reconstructed correctly



multifit now fits the black observed distribution with two separate pulse shapes allowing the signal and the pile up contributions to be seperated

CMS Superclustering Chain



this is greatly simplified but is fine for the general picture

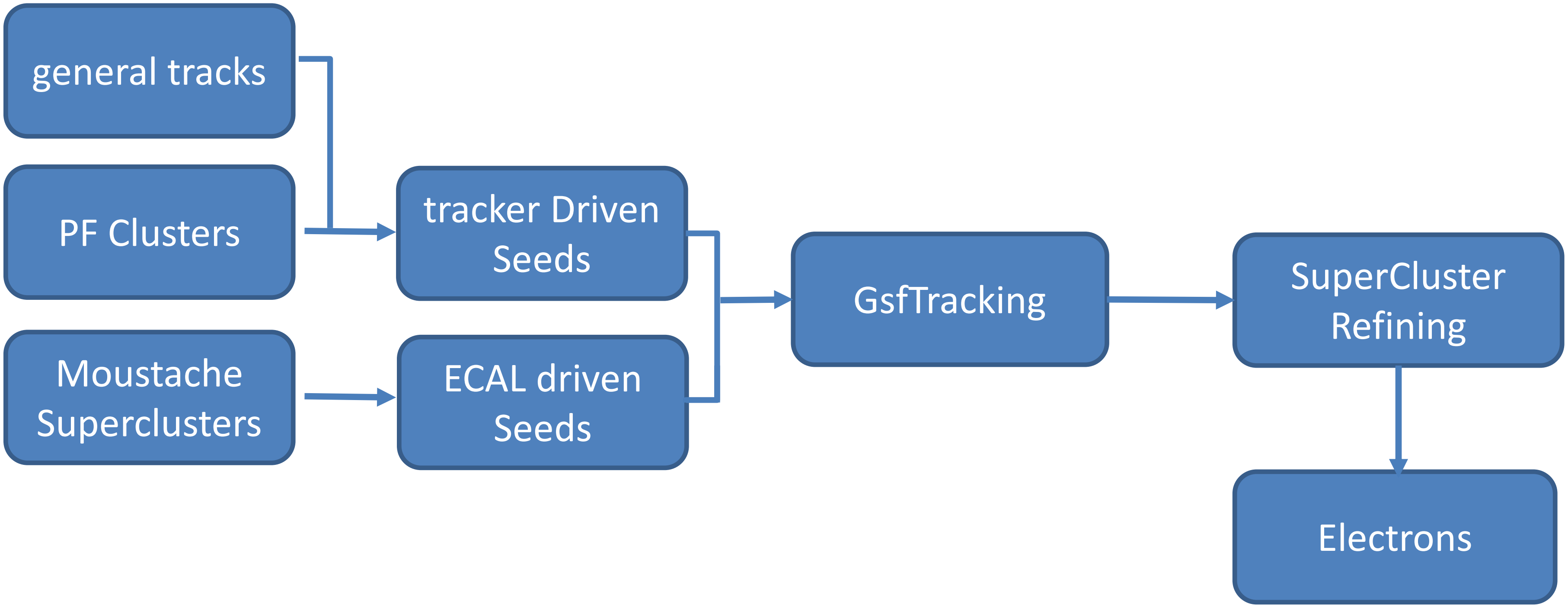
the refined superclusters are then used to make electrons and photons (if they pass preselection)

- electrons additionally require a track (which is also used as input in making the refined superclusters)

ELectron Seeding

- ECAL driven:
  - supercluster + pixel match
  - highly efficient (with working pixels) for high pt (>30 GeV) isolated electrons
  - at HLT, only have ECAL driven seeded electrons
- tracker driven:
  - matches reconstructed tracks with PFClusters using particle flow techiques
    - uses "generalTracks", the general track collection in CMS
    - no explicit requirement on pixel hits
  - most useful for low pt / non-isolated electrons
- an electron will be typically as both tracker driven and ECAL driven
  - each electron type has their own preselection to become a GsfElectron
  - GsfElectron::ecalDrivenSeed()
    - GsfElectron::ecalDriven() as requires that electron passes cut based preselection GsfElectron::passingCutBasedPreselection()
  - GsfElectron::trackerDrivenSeed()

Electron Reconstruction

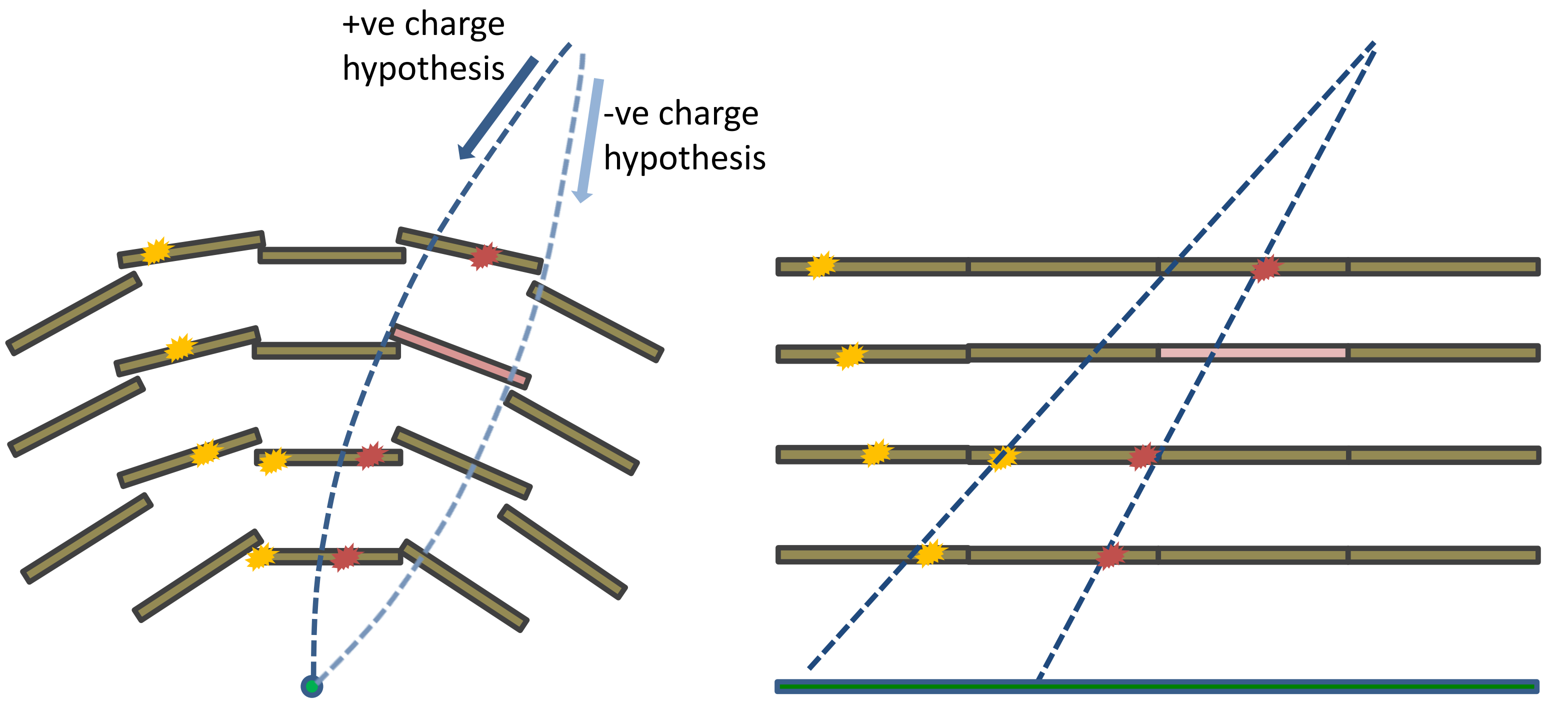


- again slightly simplified as it glosses over the PF part
  - making the tracker driven seeds is rather complex

Technical Details

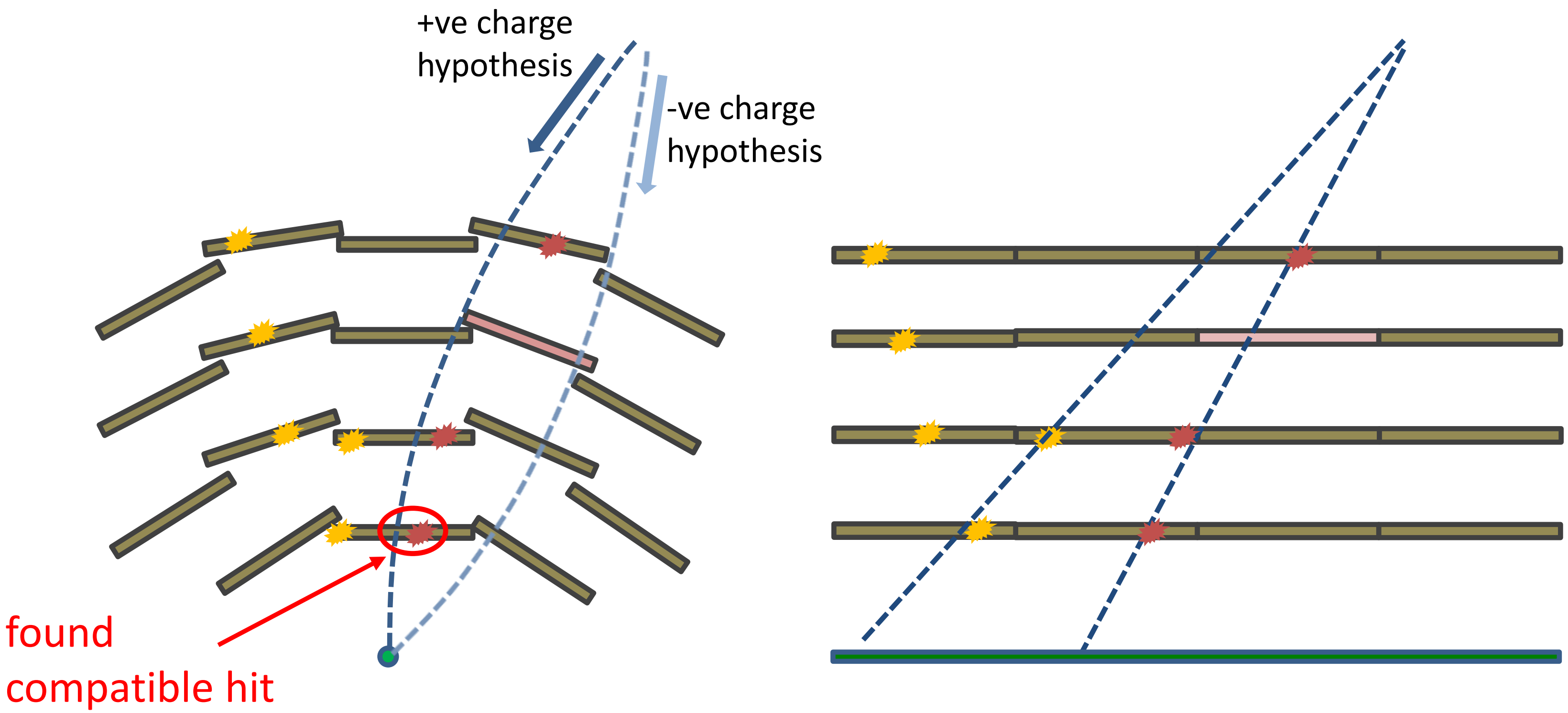
- https://github.com/cms-sw/cmssw/blob/master/RecoEcal/EgammaCoreTools/interface/Mustache.h
- https://github.com/cms-sw/cmssw/blob/master/RecoEcal/EgammaCoreTools/src/Mustache.cc

Pixel Matching: Step 1



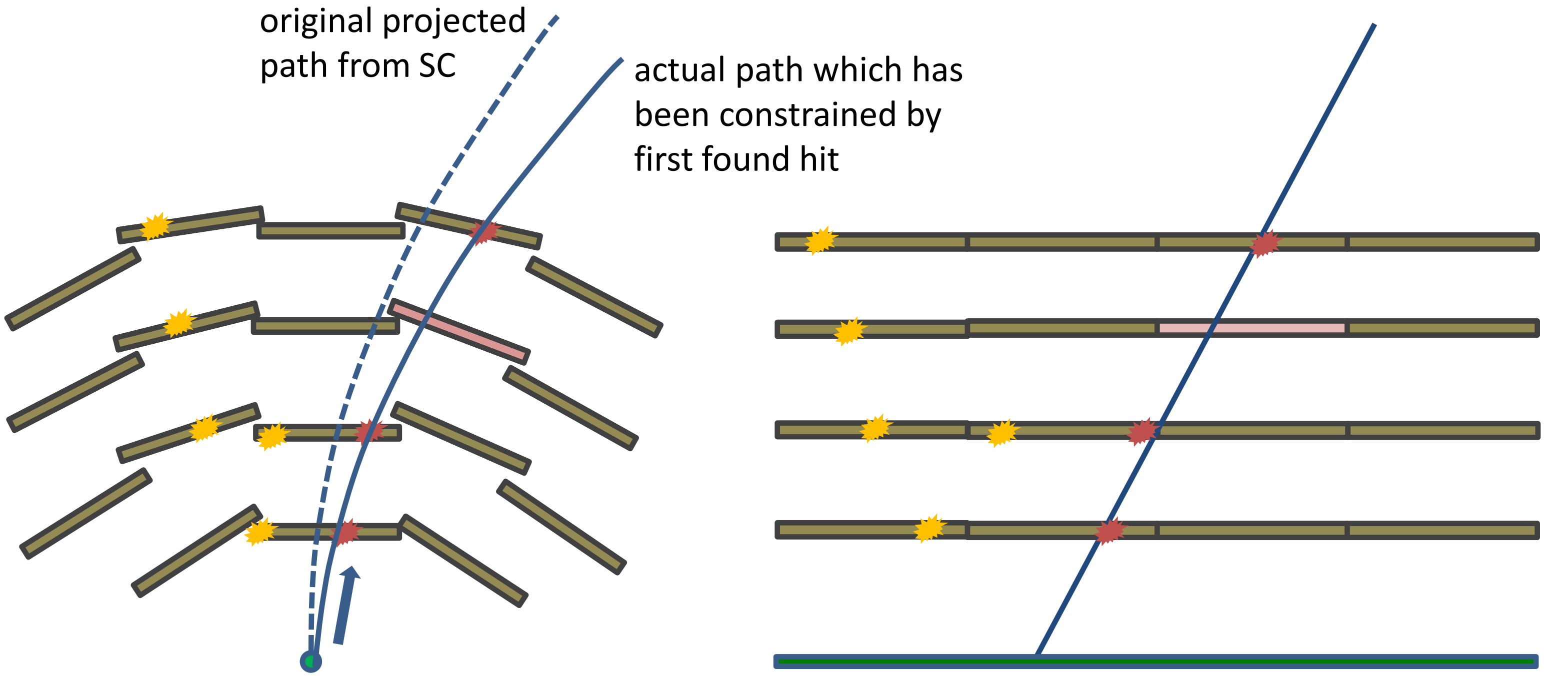
- first extrapolate the trajectory of the supercluster assuming it comes from the beamspot x,y
  - z position is not yet known
  - do it for both charge hypotheses
- look for a hit matching in ϕ , can have any z

Pixel Matching: Step 1



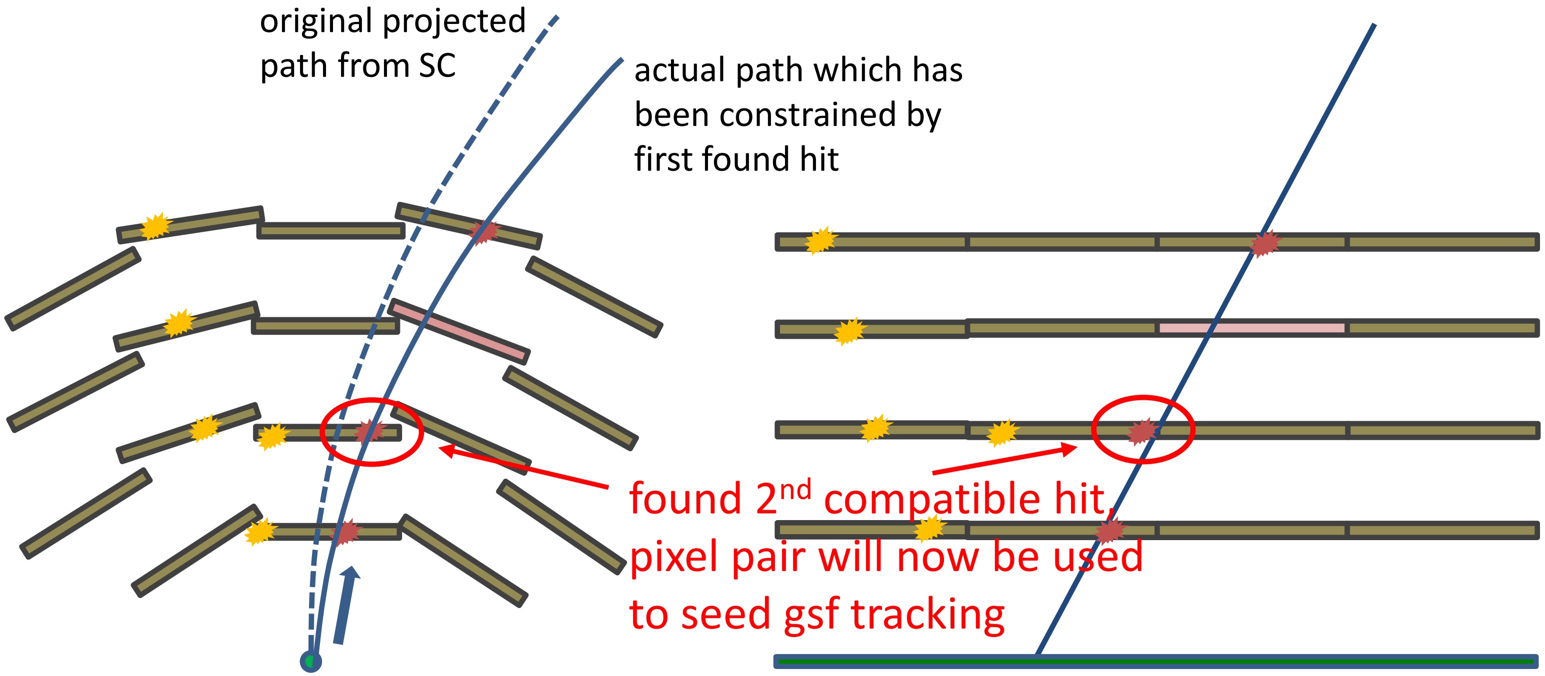
- first extrapolate the trajectory of the supercluster assuming it comes from the beamspot x,y
  - z position is not yet known
  - do it for both charge hypotheses
- look for a hit matching in ϕ , can have any z

Pixel Matching: Step 2



- now use the found hit and use to form a new trajectory starting from the beamspot and passing thought that hit
  - momentum of trajectory is still supercluster energy
  - z of starting point is linearly extrapolated from SC position and hit position
- look for a second hit and matching it in ϕ, z
  - second match has order of magnitude tighter ϕ, z windows
  - in future, require a 3rd hit (already done at the HLT)

Pixel Matching: Step 2



- now use the found hit and use to form a new trajectory starting from the beamspot and passing thought that hit
  - momentum of trajectory is still supercluster energy
  - z of starting point is linearly extrapolated from SC position and hit position
- look for a second hit and matching it in ϕ, z
  - second match has order of magnitude tighter ϕ, z windows
  - in future, require a 3rd hit (already done at the HLT)

Mustache Parameters

```
SC dynamic dPhi parameters:
 Parameters are binned in 4 (E, |eta|) regions.
 Parameters for E_min=0 and |eta_min|=0:
  yoffset:    0.0280506
  scale:      0.946048
  xoffset:    -0.101172
  width:      0.432767
  saturation: 0.14
  cutoff:     0.6
 Parameters for E_min=0 and |eta_min|=1.479:
  yoffset:    0.0497038
  scale:      0.975707
  xoffset:    -0.18149
  width:      0.431729
  saturation: 0.14
  cutoff:     0.55
 Parameters for E_min=0 and |eta_min|=1.75:
  yoffset:    0.05643
  scale:      1.60429
  xoffset:    -0.642352
  width:      0.458106
  saturation: 0.12
  cutoff:     0.45
 Parameters for E_min=0 and |eta_min|=2:
  yoffset:    0.0928887
  scale:      1.22321
  xoffset:    -0.260256
  width:      0.345852
  saturation: 0.12
  cutoff:     0.3
```

```
Parameters for log10(E_min)=-3 and |eta_min|=0:
 pUp:
  [0]: -0.107537
  [1]: 0.590969
  [2]: -0.076494
 pLow:
  [0]: -0.0268843
  [1]: 0.147742
  [2]: -0.0191235
 w0Up:
  [0]: -0.00681785
  [1]: -0.00239516
 w1Up:
  [0]: 0.000699995
  [1]: -0.00554331
 w0Low:
  [0]: -0.00681785
  [1]: -0.00239516
 w1Low:
  [0]: 0.000699995
  [1]: -0.00554331
```

Key Showershape Variables

- sigmaIEtaIEta is the log energy weighted RMS of the shower in units of crystals
  - $\sigma_{i\eta i\eta} = \sqrt{\left(\frac{\Sigma_i^{5\times5} w_i\left(\eta_i - \overline{\eta}_{5\times5}\right)^2}{\Sigma_i^{5\times5} w_i}\right)}$
  - $w_i = 4.7 + \ln\frac{E_i}{E_{5\times5}}$
    - this is effectively a noise cut, each crystal needs to have > 0.9% of 5x5 energy
    - means that very low energy electrons are sensitive to noise as 0.9% of a small number brings it below noise threshold
  - E_i = energy of crystal, $E_{5\times5}$ energy of 5x5
    - likewise for η
  - η is in units of crystals, not absolute η
    - endcap uses $(ix^2 + iy^2)^{1/2}$ to get η in terms of crystals
  - normalised to 0.01745 in barrel and 0.0447 in endcap
  - cut effectively means that all the energy is within two crystals
- one of the most important ele/photon ID variables in CMS
  - calculated by https://github.com/cms-sw/cmssw/blob/CMSSW_9_2_14/RecoEcal/EgammaCoreTools/interface/EcalClusterTools.h#L889