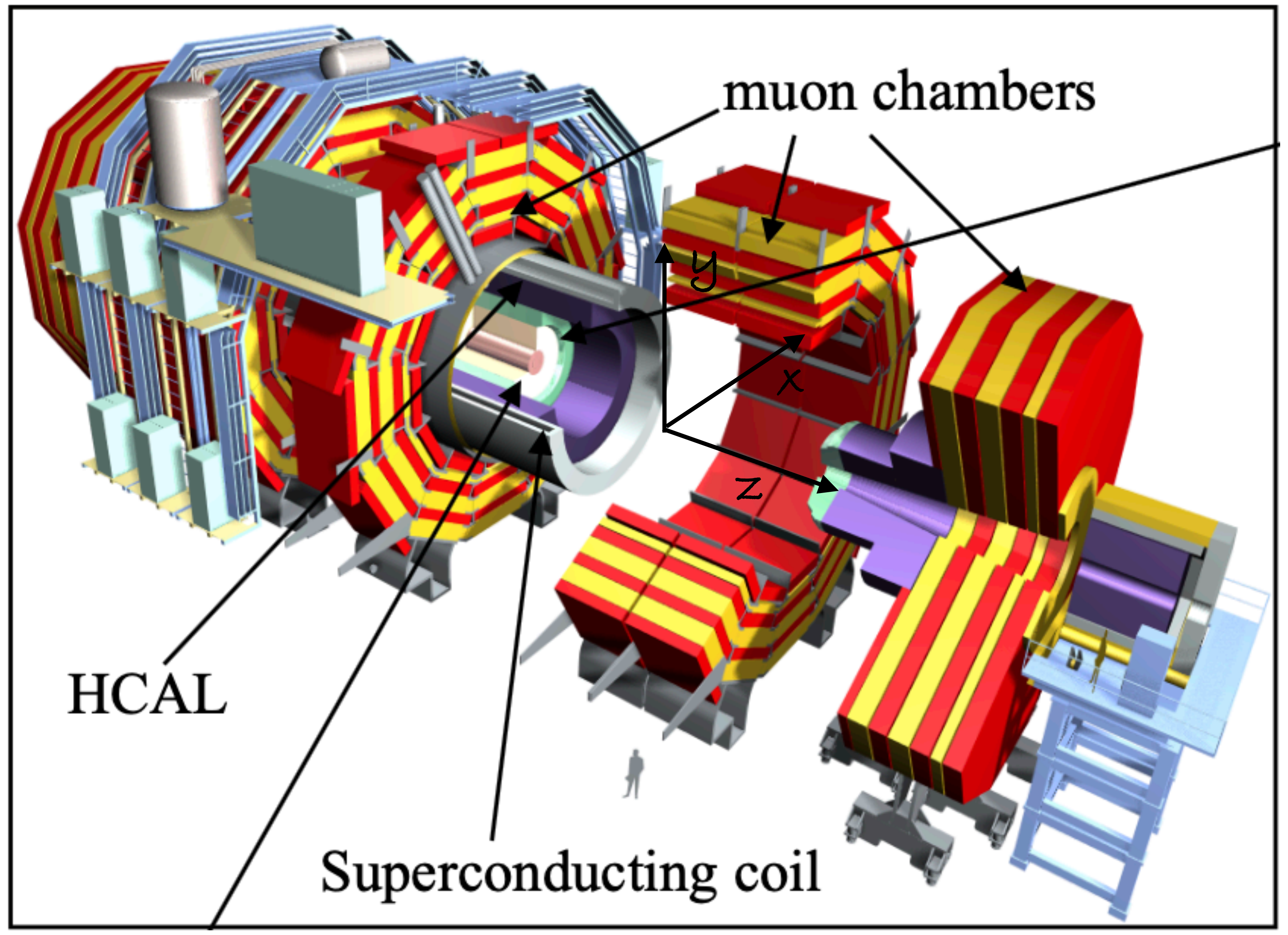
e/γ long exercise

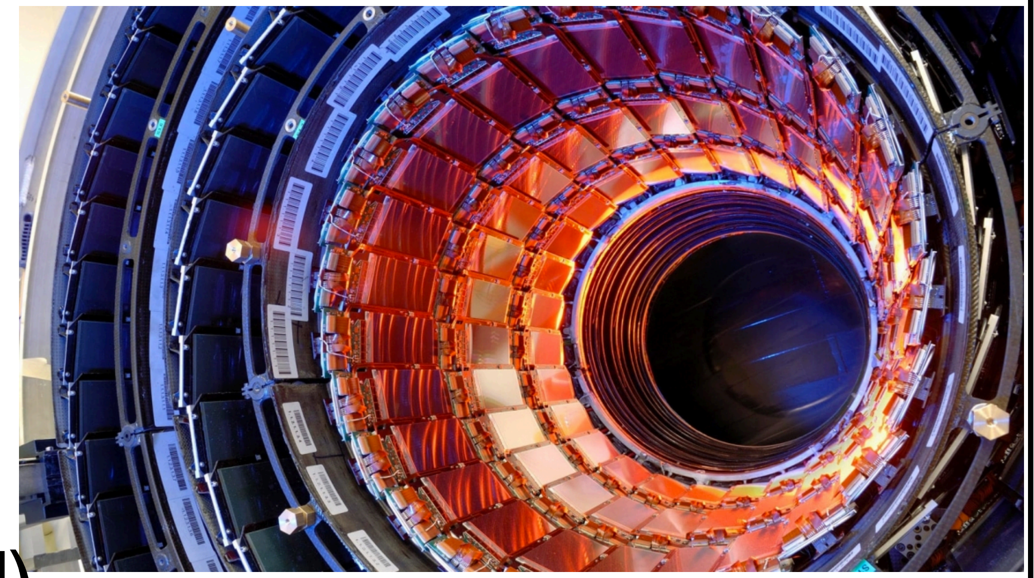
Ying AN, Afiq Aizuddin Anuar, Anshul Kappor

CMSPO&DAS 2023 Hamburg DESY
11 Oct. 2023

CMS detector



Tracker



- **Pixels**
  4 layers (barrel)
  3X2 disks (forward)

- **Stripes**
  Inner Tracker
  Outter Tracker

|η| < 2.5

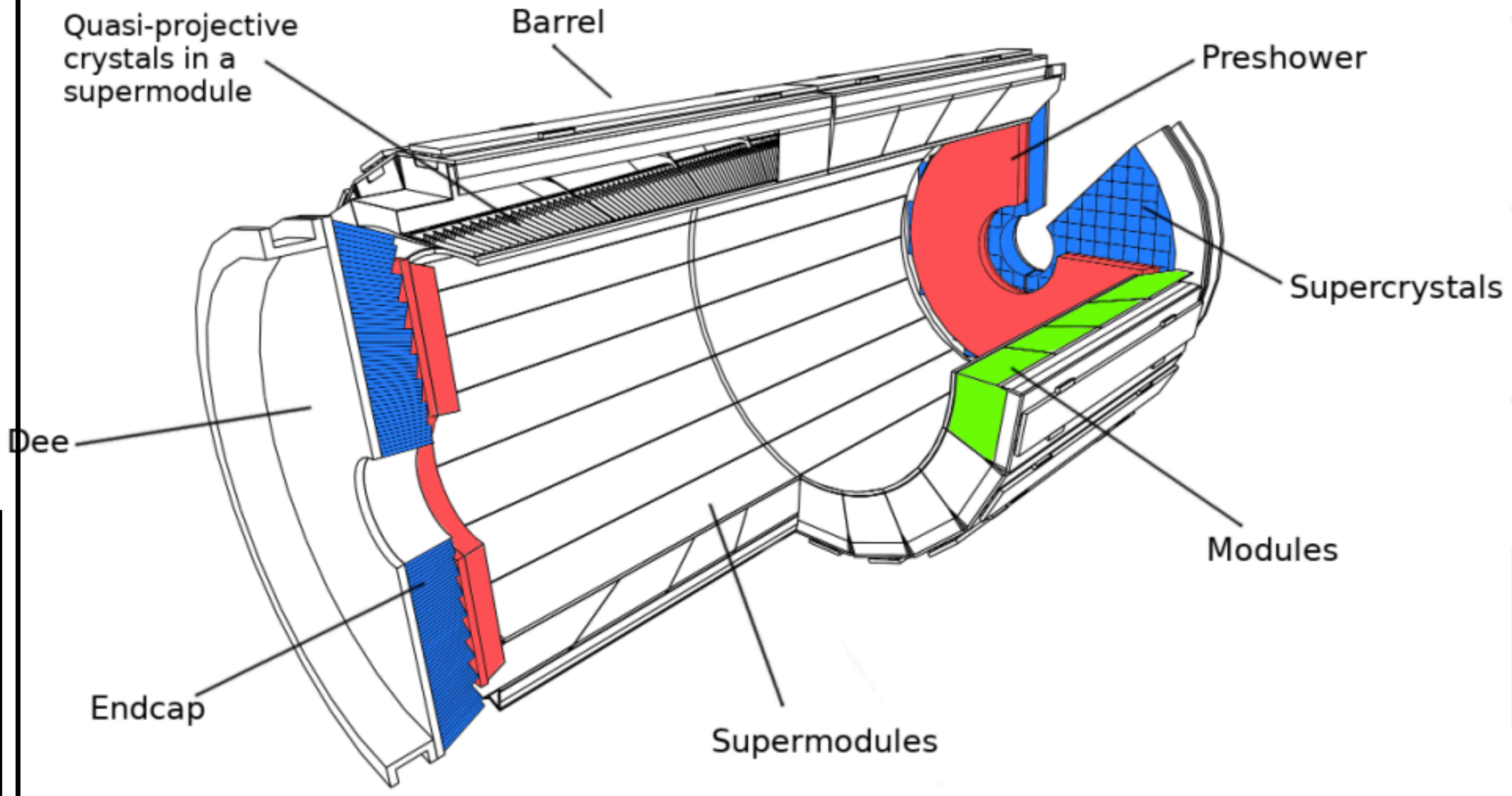
4T solenoidal field

ECAL

≈76000 **scintillating PbWO4 crystals**

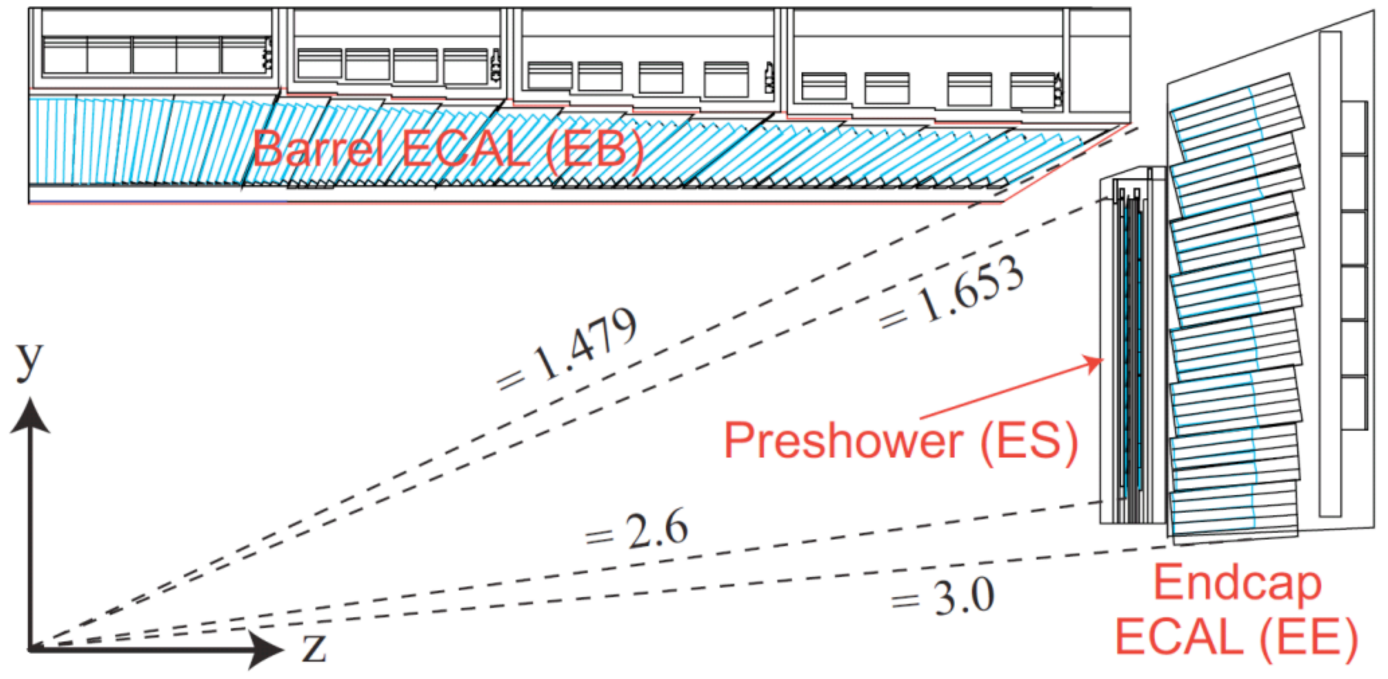
X_0 = 0.89 cm and M_R = 2.2 cm

|η| < 3



Barrel (EB) |η| < 1.48

- 36 supermodules (1700 channels)
- Total of 61200 PbWO4 crystals
- Avalanche Photodiode readout (APD)



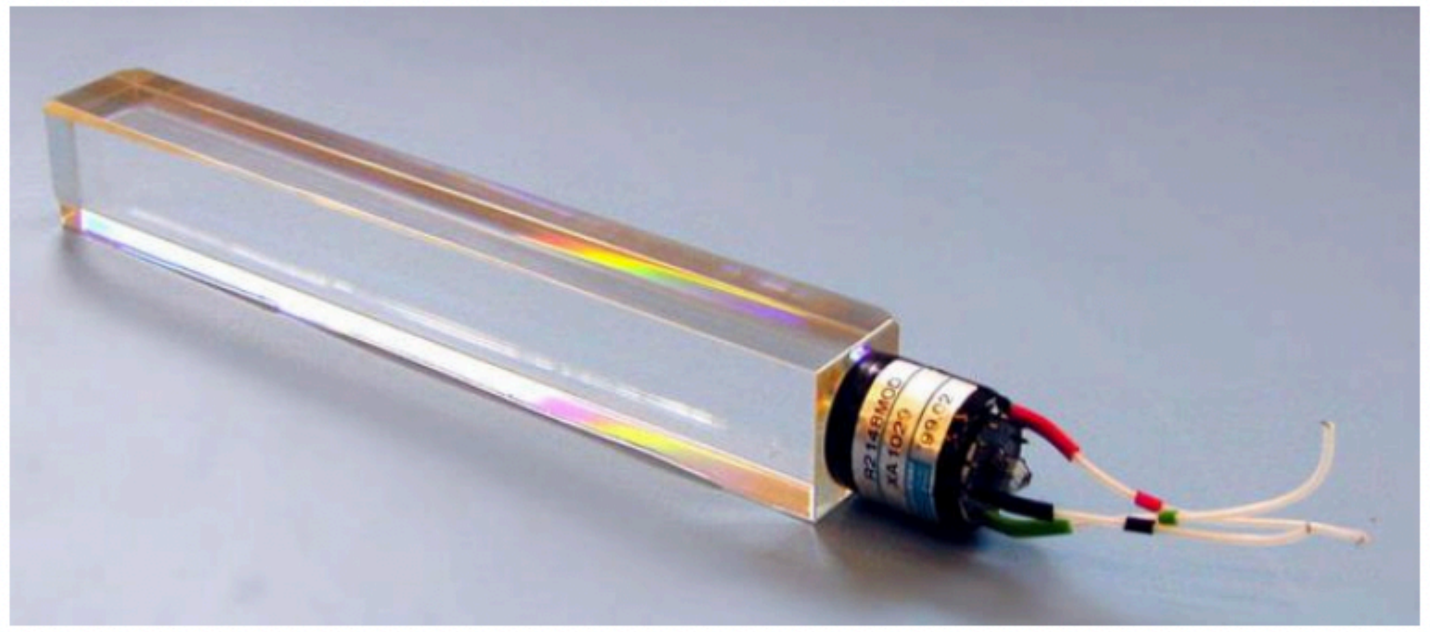
Endcaps (EE) 1.48<|η|<3.0

- Four half-disk Dees (3662 channels)
- Total of 14648 PbWO4 crystals
- Vacuum Photo Triode readout

Preshower 1.65 <|η|< 2.6

- Two Lead/Si planes
- 137216 Si strips ($1.8 \times 61 mm^2$)

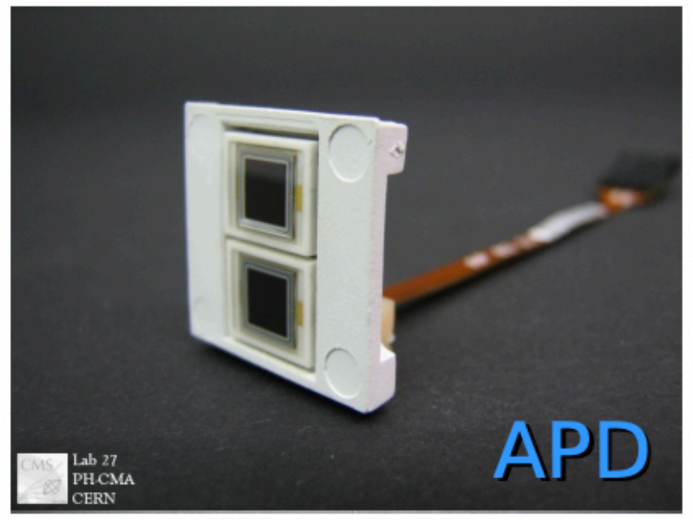
e/γ showers

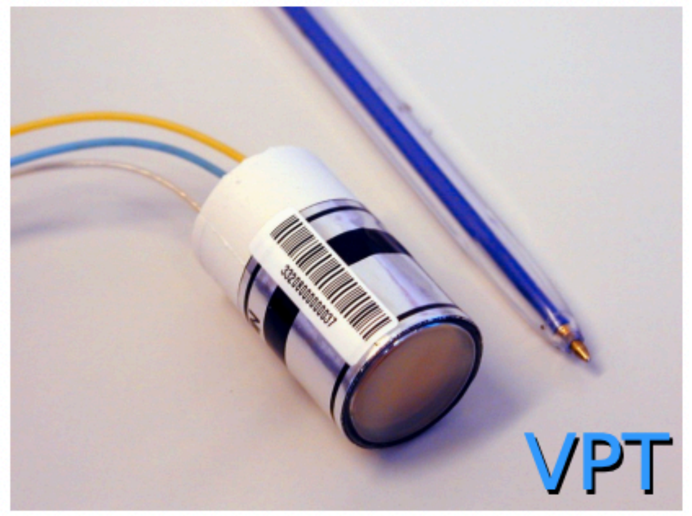


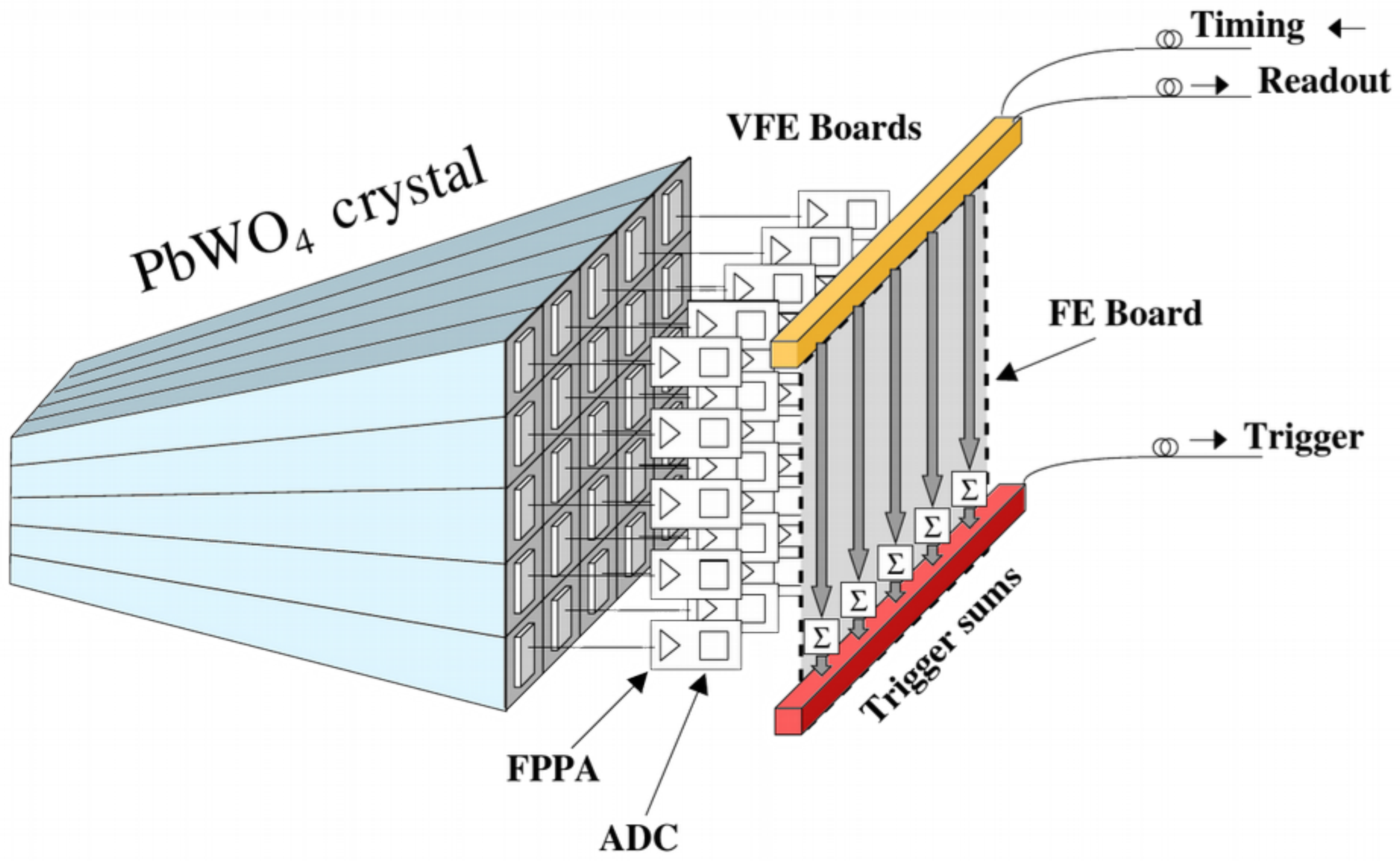
- ECAL Made of 75848 **PbWO4** crystals
- Each crystal has an APD (barrel) or a VPT (endcaps)

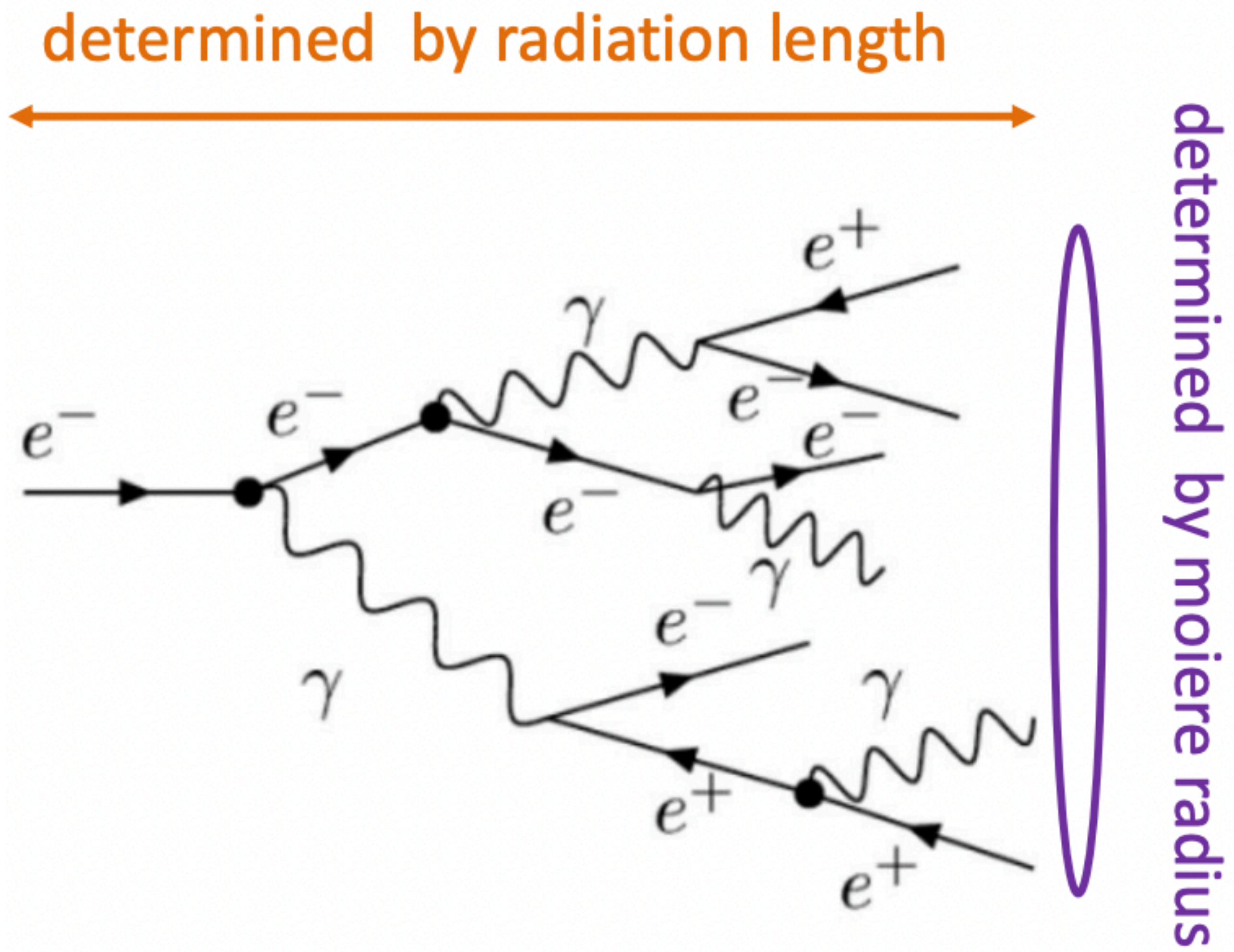
PbWO4 properties

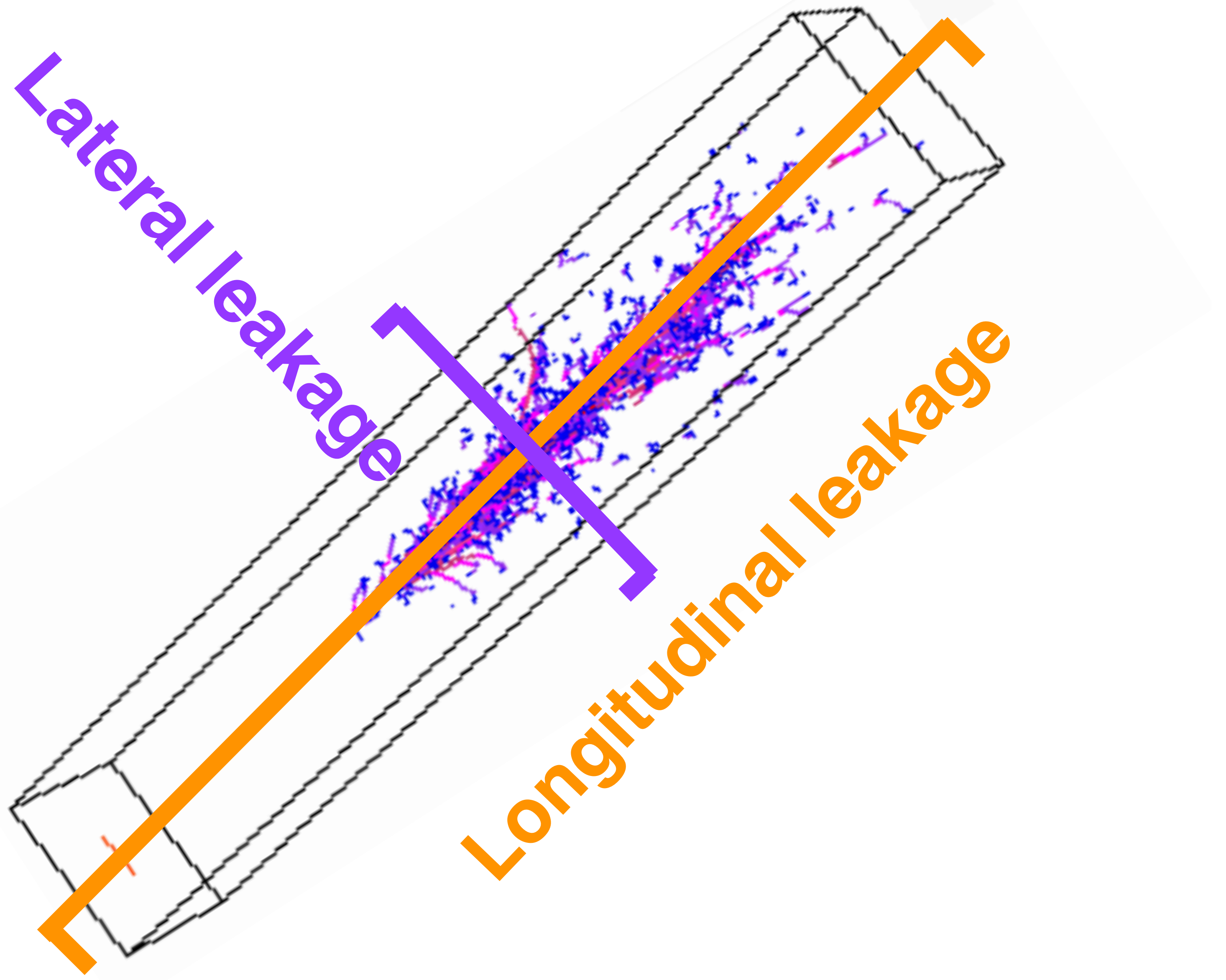
- 80% of light in 25 ns
- Density of 8.28 g/cm^3
- X_0 = 0.89 cm and M_R = 2.2 cm
- Front face of 22x22mm^2; rear face of 26x26mm^2
- Low light yield (100 γ/MeV); Need photodetectors with gain in magnetic field





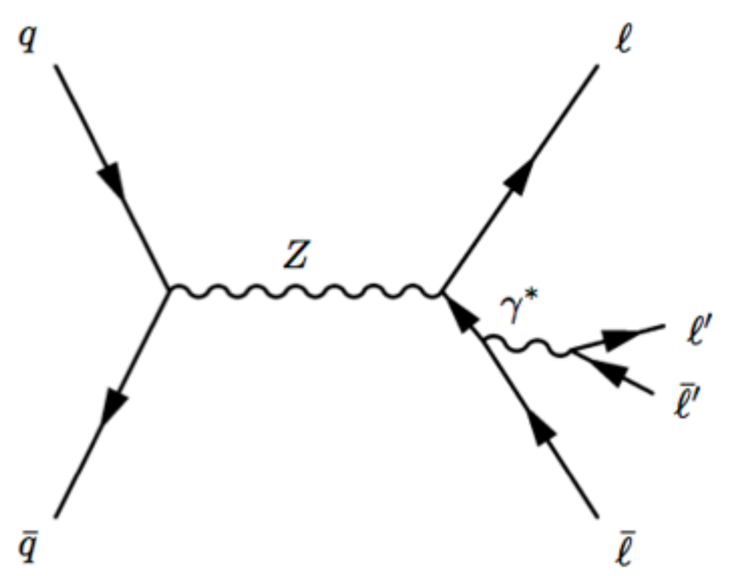




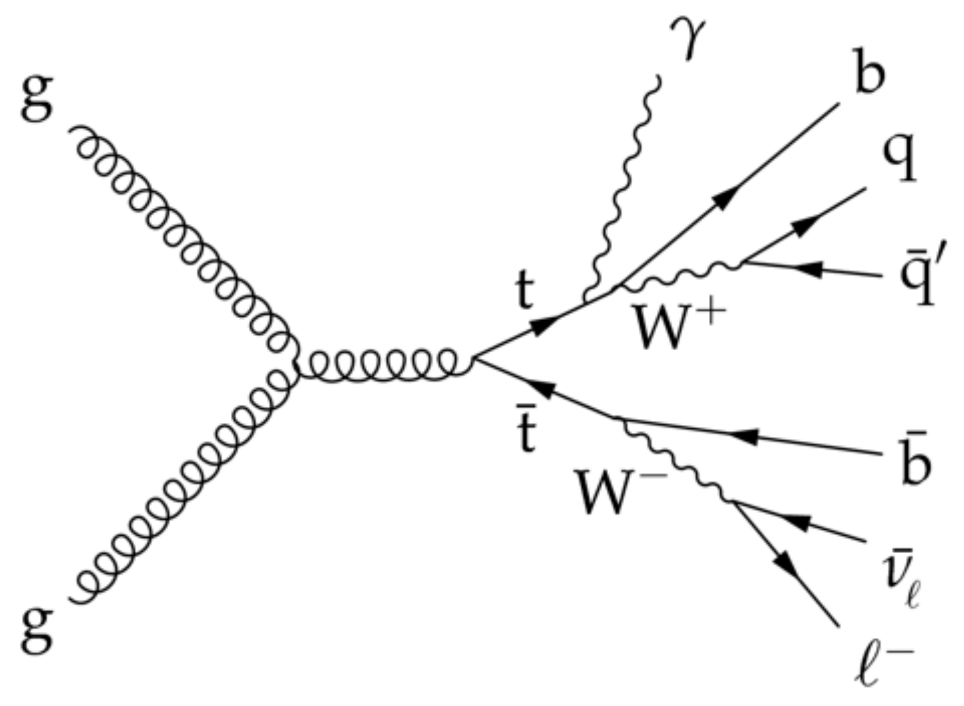


Produce prompt/nonprompt e/γ

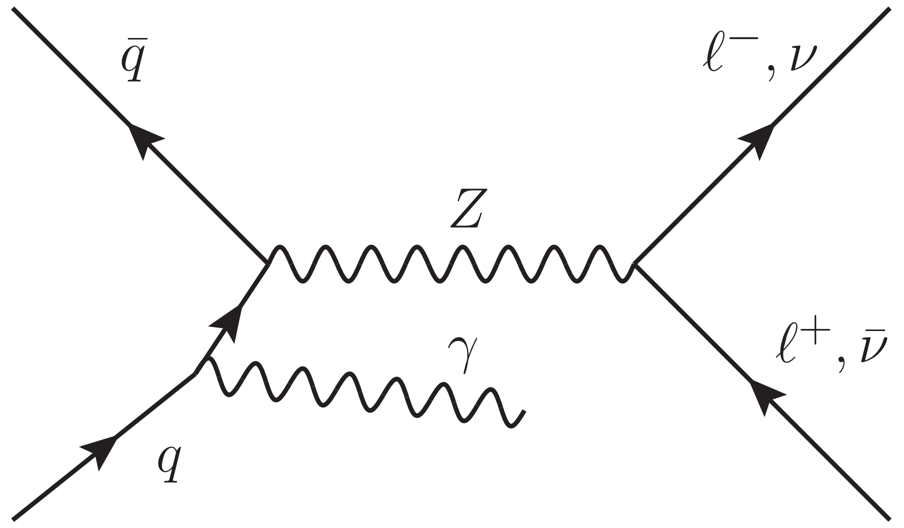
Z production



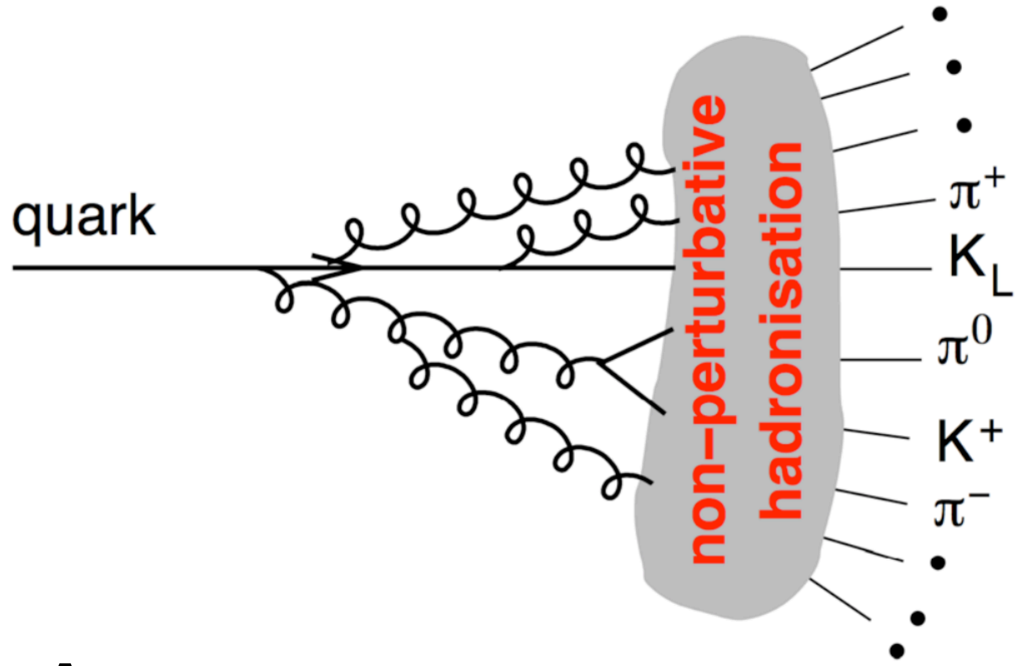
$t\bar{t}$ production



Zγ production



Hadronic jets



Appear as nonprompt e/γ

- Let's build samples containing prompt and nonprompt electrons/photons
- It's performed by
  1. finding final-state electron and photon particles in the generator-level table that are directionally close ($\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)
  2. Match the PDG ID or parent ID
  3. The flavour of the particle in the generator-level

Exercise

1. Open the NanoAOD files and draw the distribution of e/γ p_T and η
2. Prepare Ntuples
   - Save prompt and nonprompt electrons/photons
   - Save Z events

e/γ reconstruction and energy measurement

e/γ reconstruction

Goal:

- Reconstruction the e/γ energy
- Reconstruct the pT, η, ϕ
- Reconstruct the electron track

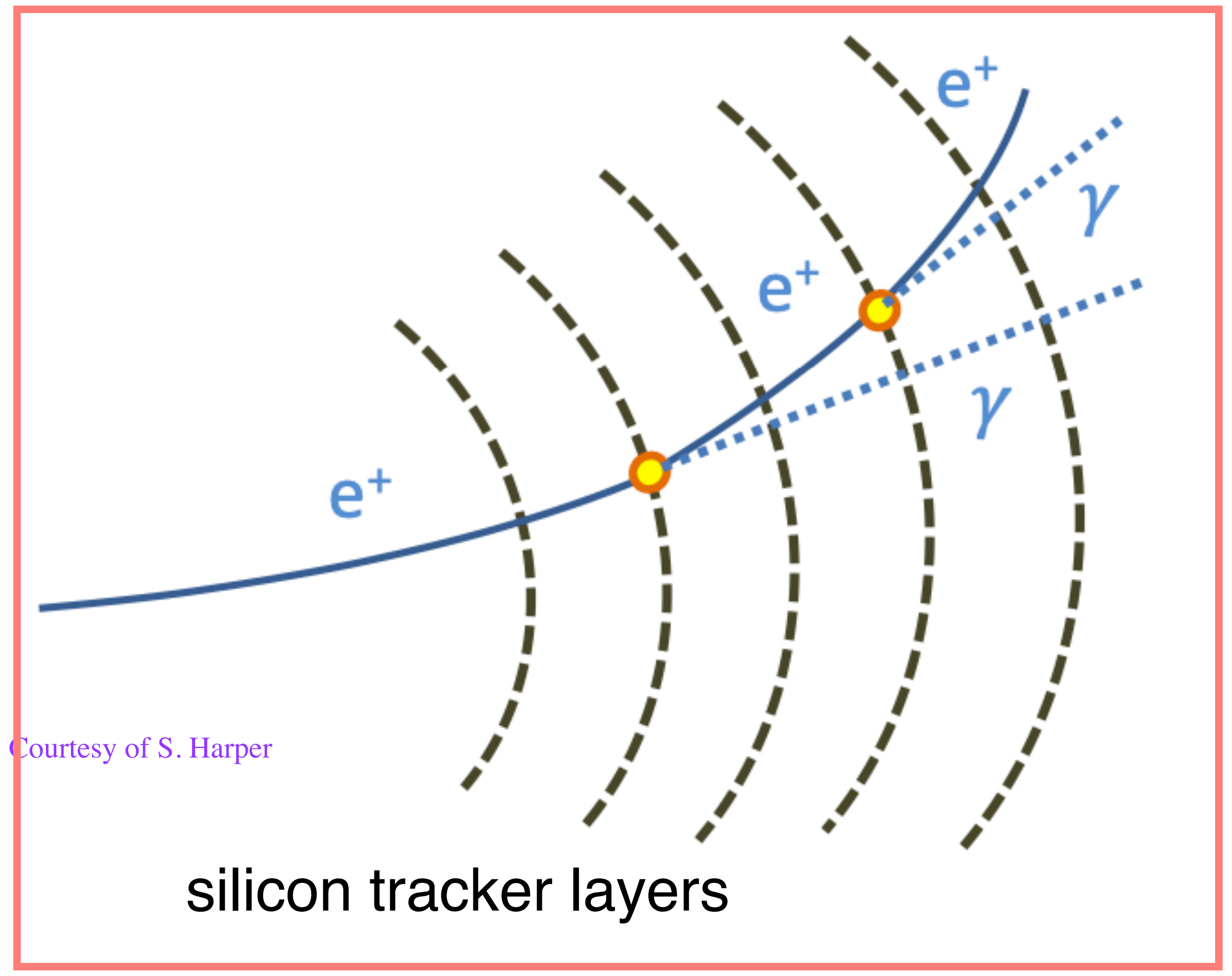
e/γ reconstruction

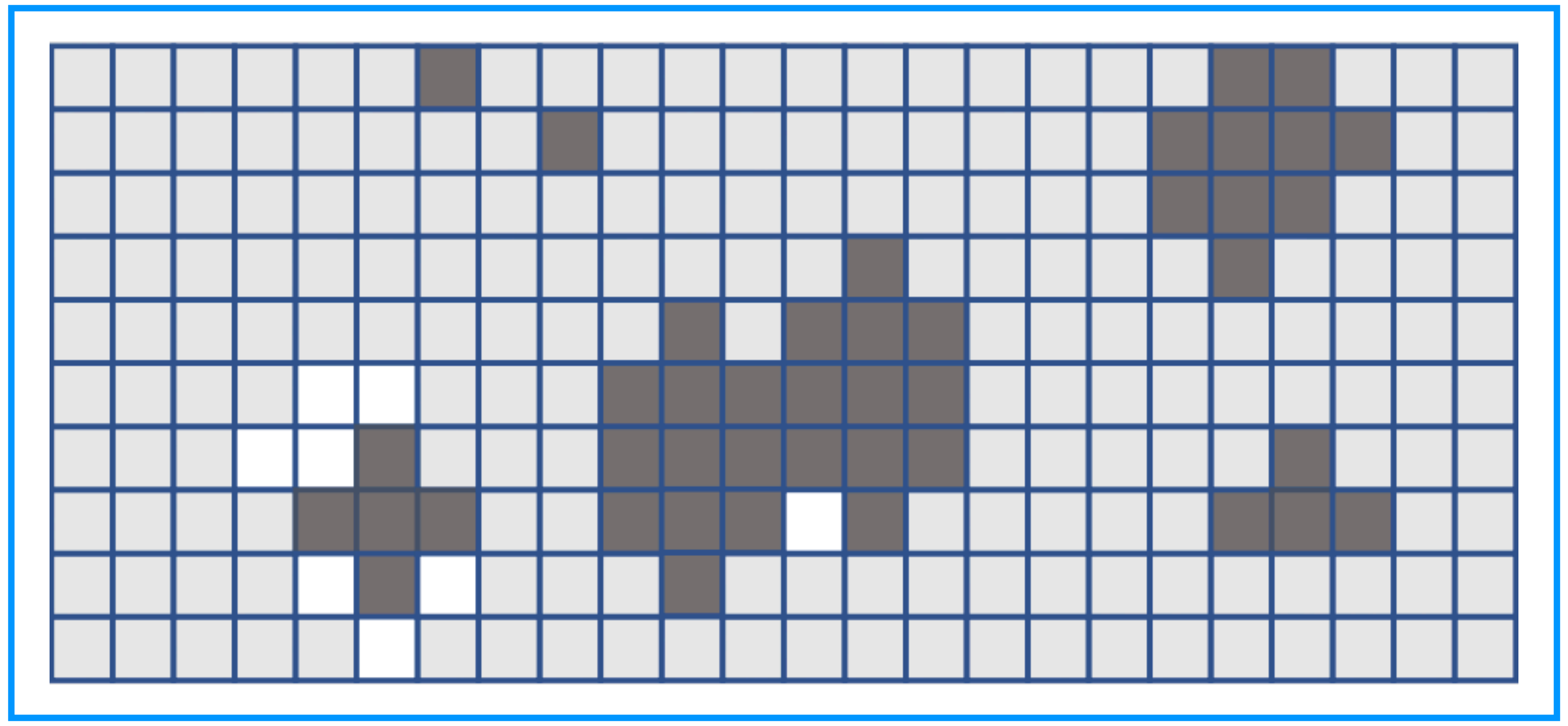
Goal:

- Reconstruction the e/γ energy
- Reconstruct the pT, η, ϕ
- Reconstruct the electron track

What we have:

- Information in tracker
- Information in ECAL





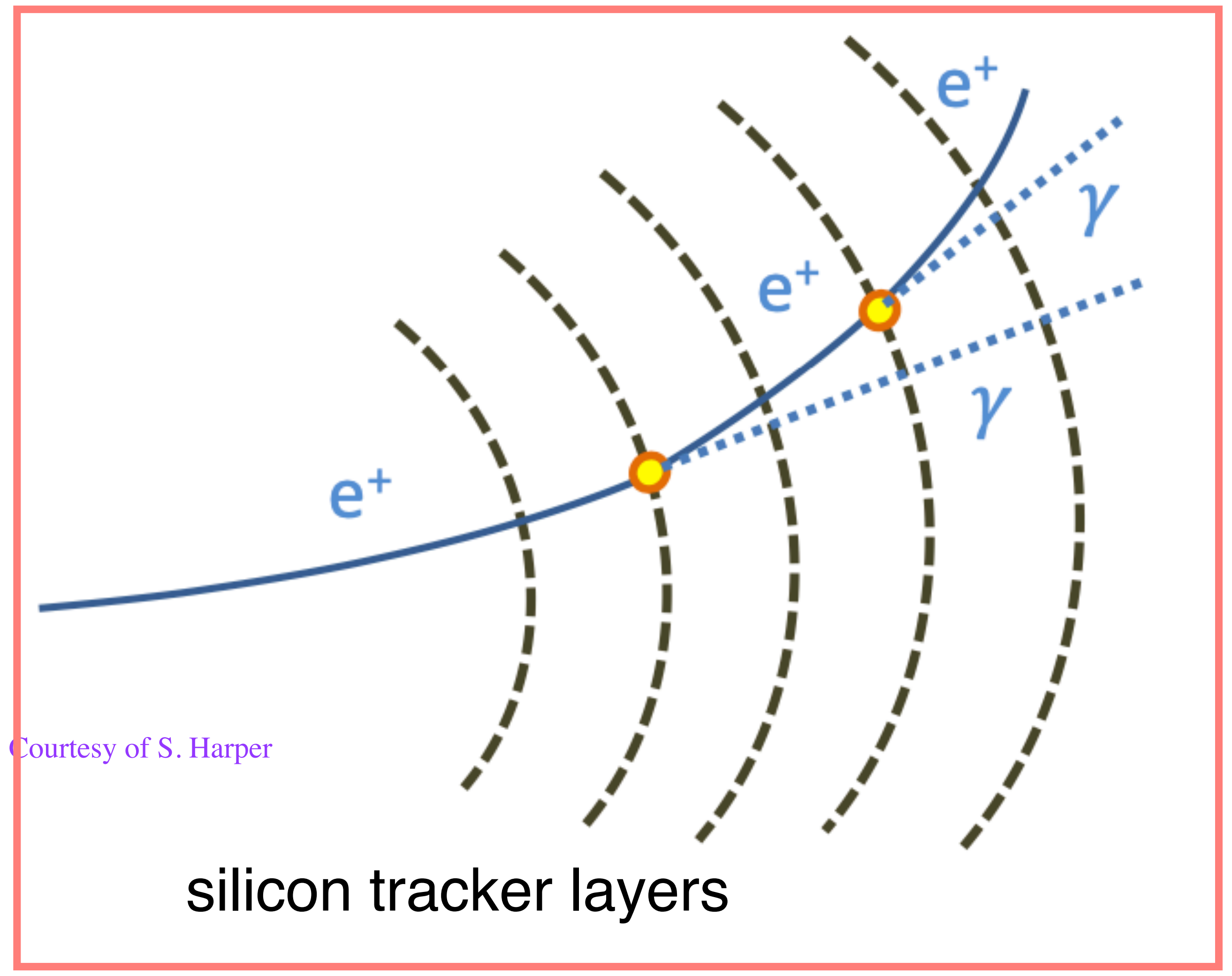
e/γ reconstruction

Goal:

- Reconstruction the e/γ energy
- Reconstruct the pT, η, ϕ
- Reconstruct the electron track

What we have:

- Information in tracker
- Information in ECAL



- Electrons loss energy via bremsstrahlung and bend in the magnetic field
- Photons may instantly convert to e^+e^- pairs

There is a cluster of electrons and photons before reach the ECAL

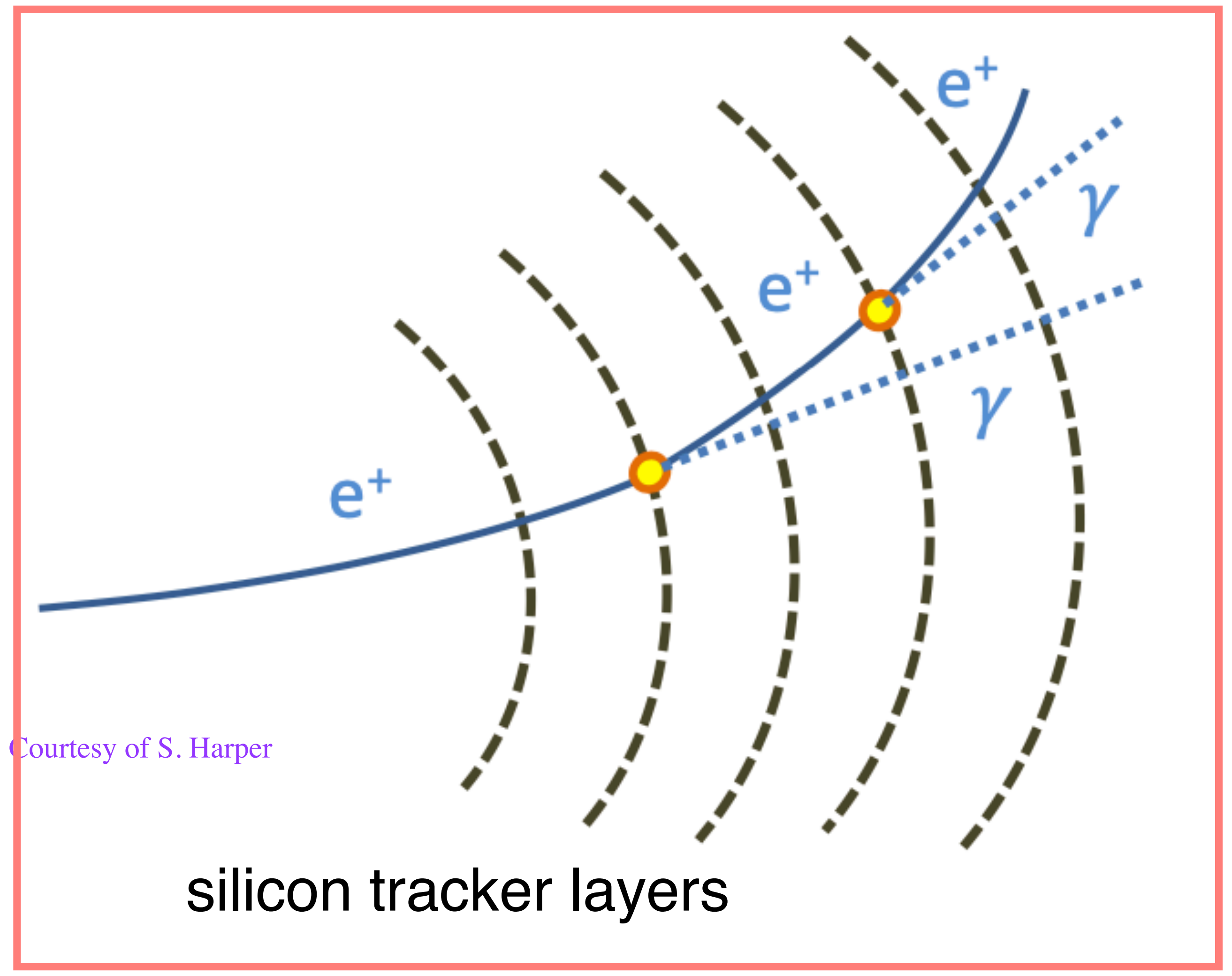
e/γ reconstruction

Goal:

- Reconstruction the e/γ energy
- Reconstruct the pT, η, ϕ
- Reconstruct the electron track

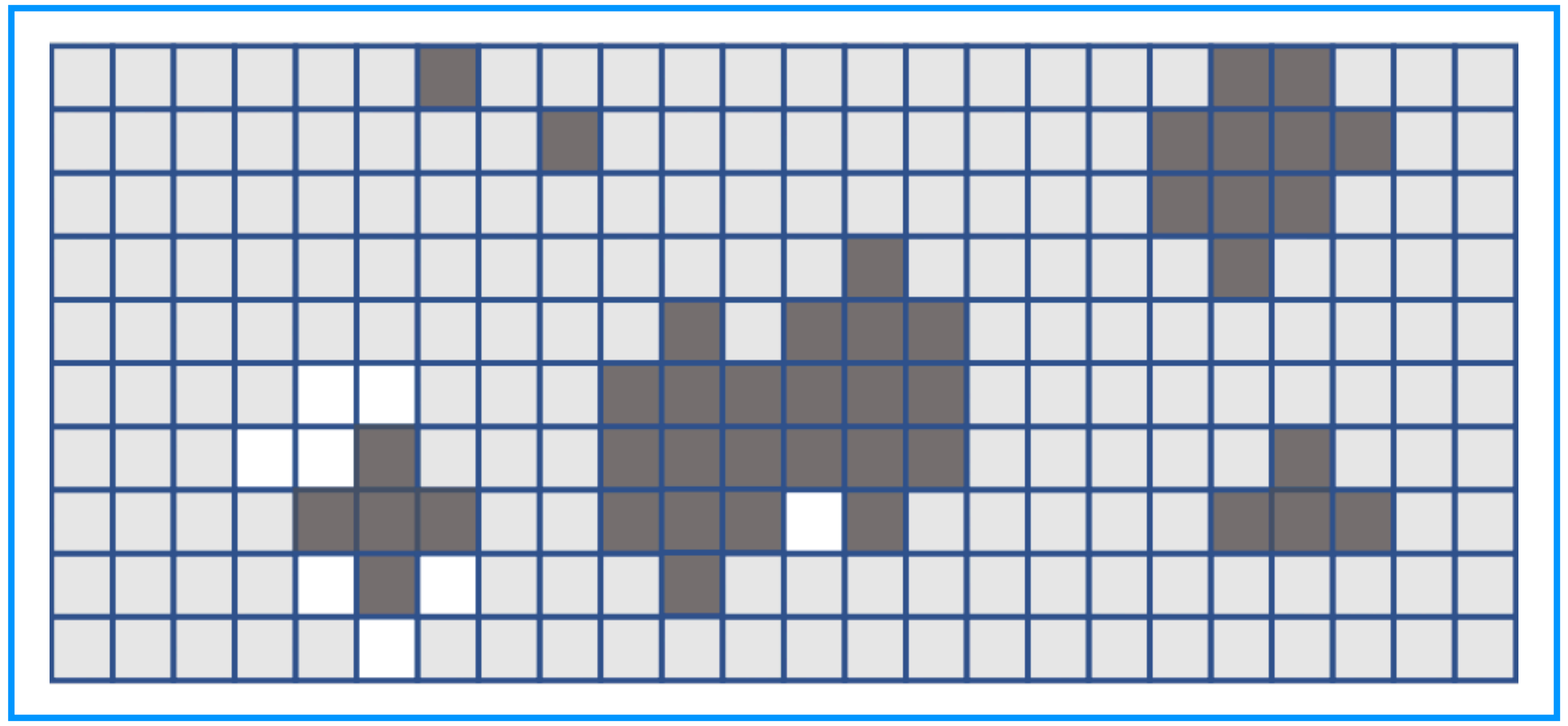
What we have:

- Information in tracker
- Information in ECAL



- Electrons loss energy via bremsstrahlung and bend in the magnetic field
- Photons may instantly convert to e^+e^- pairs

There is a cluster of electrons and photons before reach the ECAL



- Electrons with high energy (>100 GeV) deposit 97% energy in a 5x5 crystals of the ECAL
- More average electrons lose a amount of the energy as bremsstrahlung photons before they reach the ECAL

e/γ energy is inaccurate from ECAL only

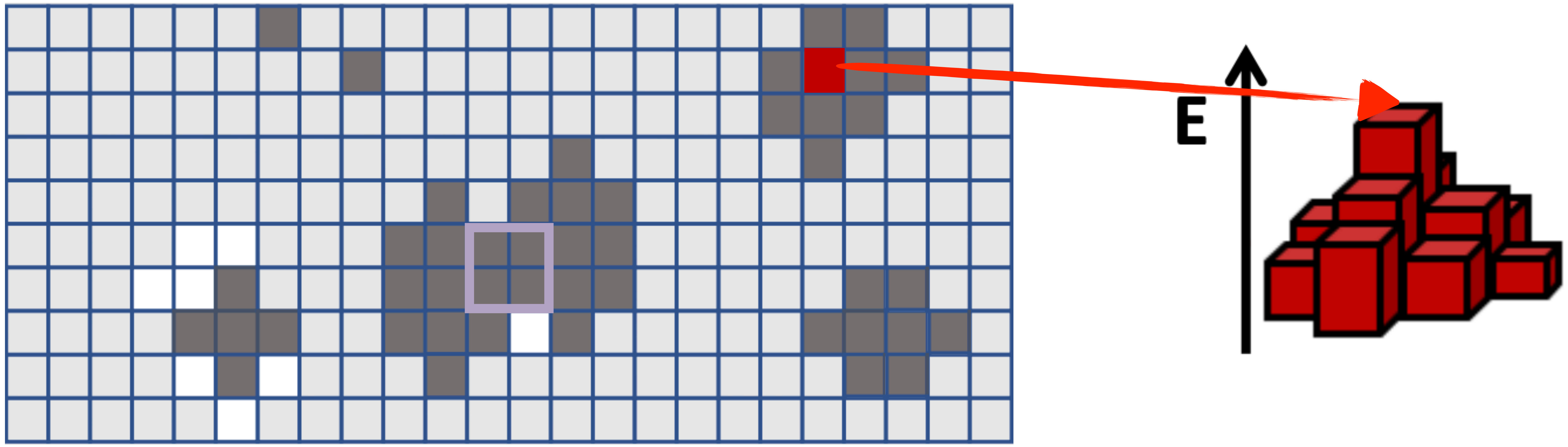
e/γ reconstruction — ECAL calibration

- The ECAL of CMS should be regularly calibrated in situ with physics events.
- The response of the crystals is calibrated such that the position of the peak of the di-photon invariant mass is at the PDG value.
- The absolute value of the calibration is not used, but rather the relative response of each crystal respect to all the others is derived (inter-calibration)

ECAL calibration exercise 1

Open the exercise-1.ipynb

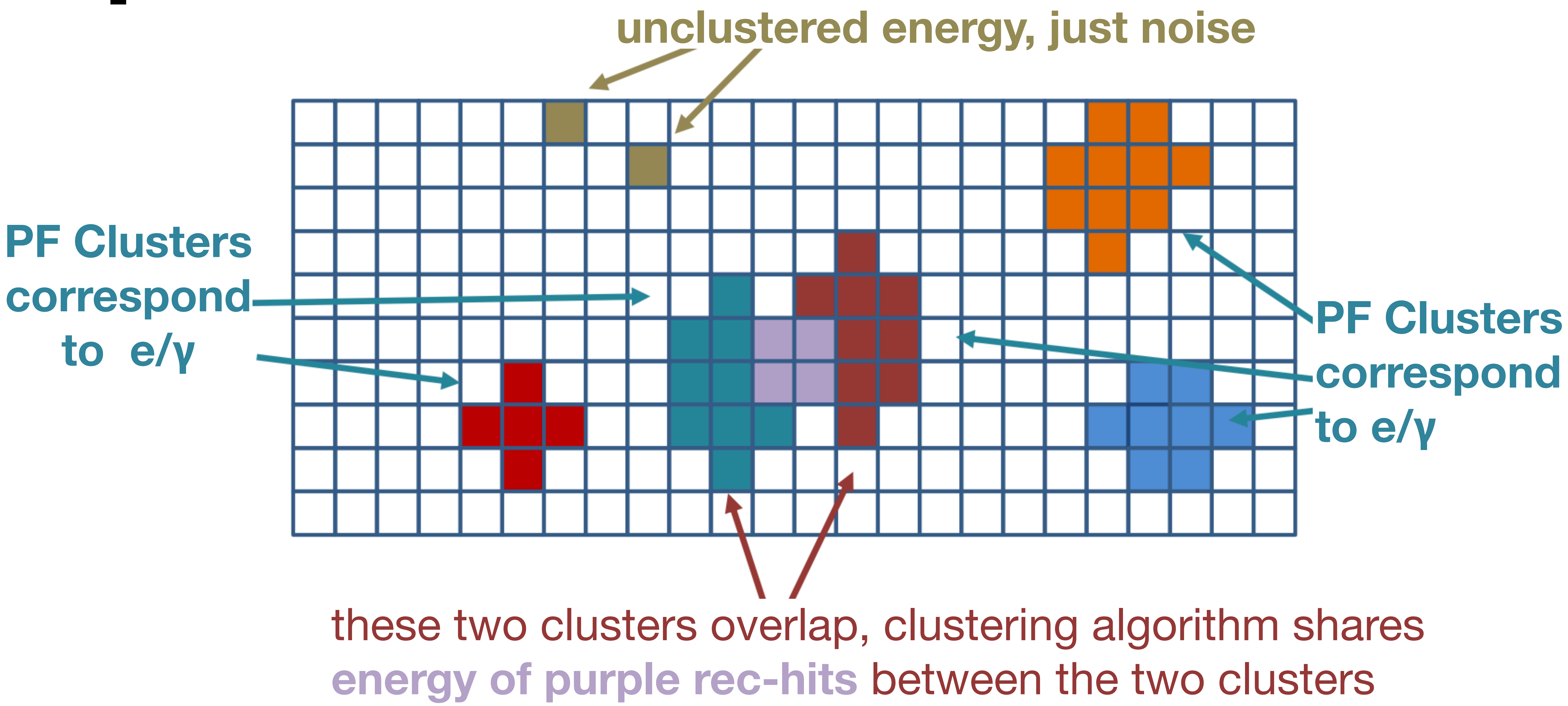
e/γ reconstruction — PF clustering



Each cluster represents the energy deposits *(after ECAL calibration added)* of a particle, such as a photon or electron, and is reconstructed from Particle Flow (PF) Clustering algorithm, usually called PF cluster

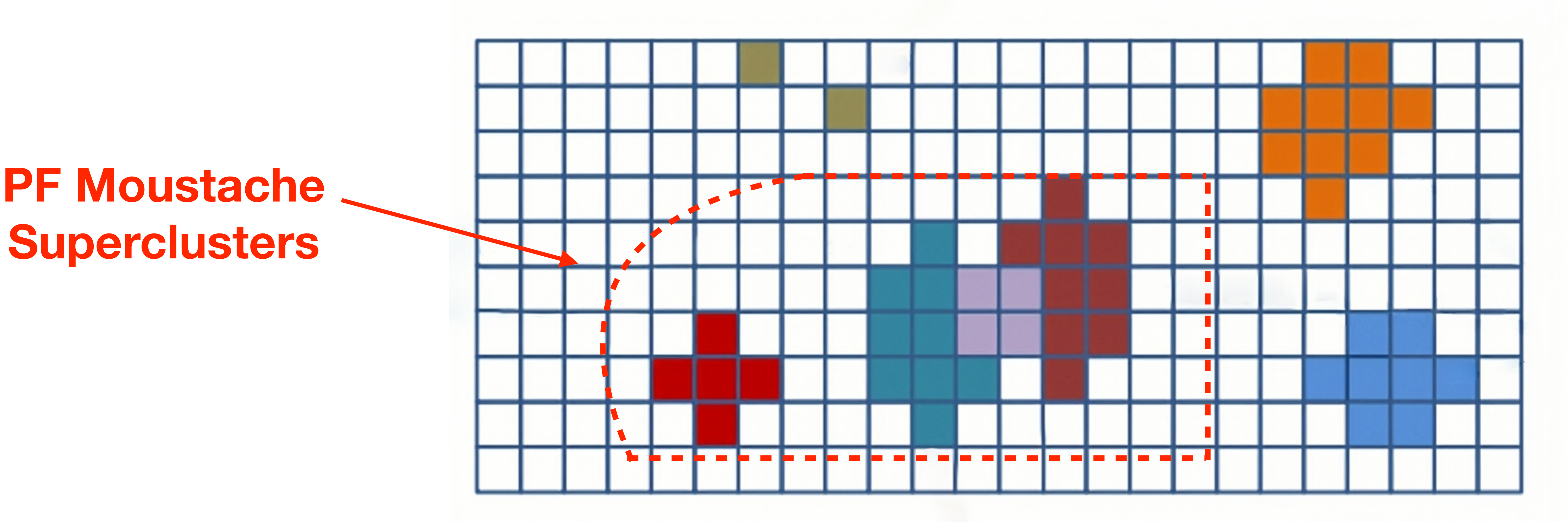
- It has a peaked profile (local maxima)
- The local maxima (seed crystal) in the cluster is above a given threshold (X GeV)
- Energies in crystals between overlapping adjacent clusters are shared

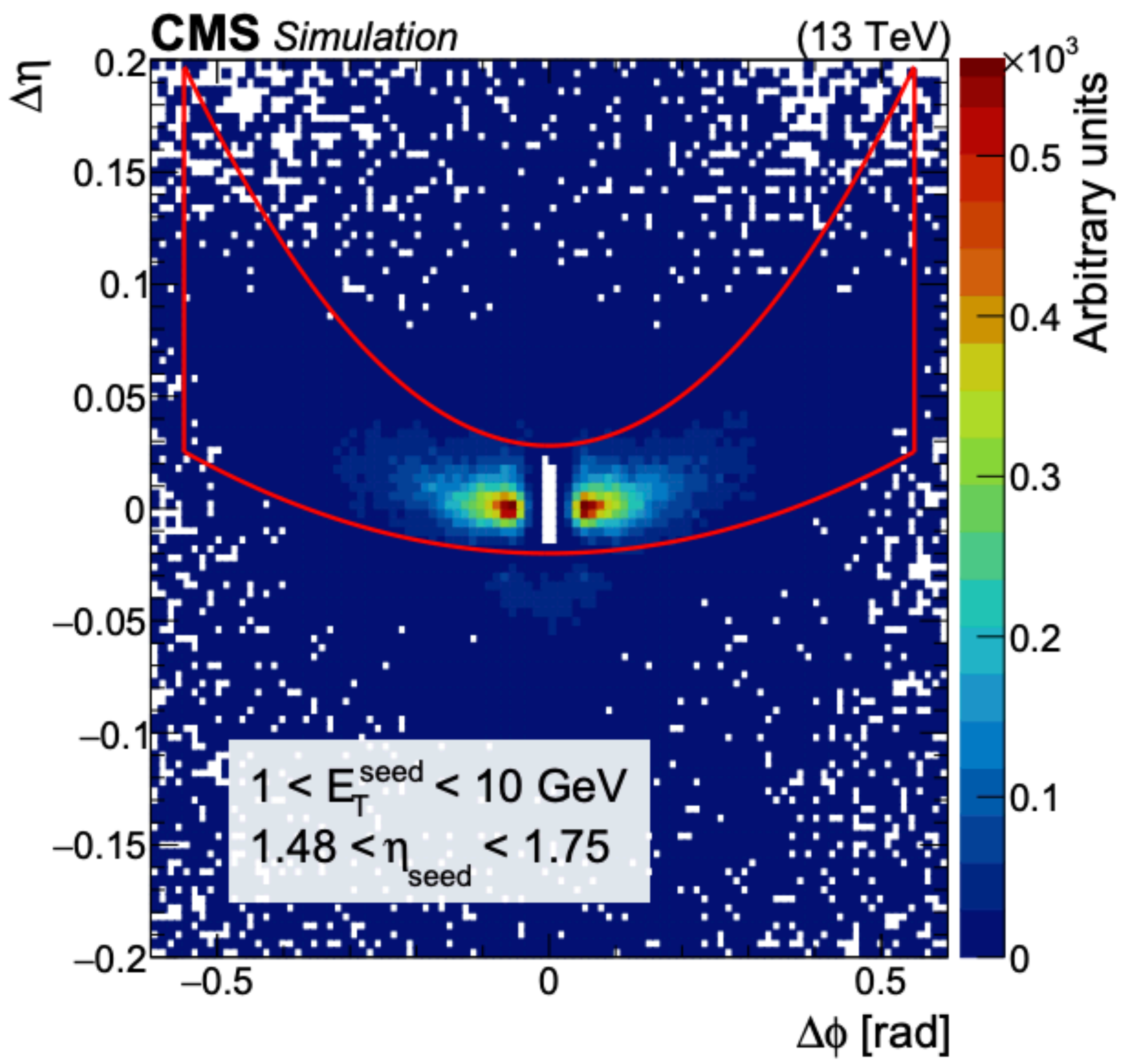
e/γ reconstruction — Superclustering



- As we mentioned before, there is a shower of electrons and photons before reach the ECAL
- The different PF clusters could be from the same origin electron or photon
- Our **goal** first is to **merge these possible PF clusters**, usually call **superclustering**

e/γ reconstruction — Superclustering

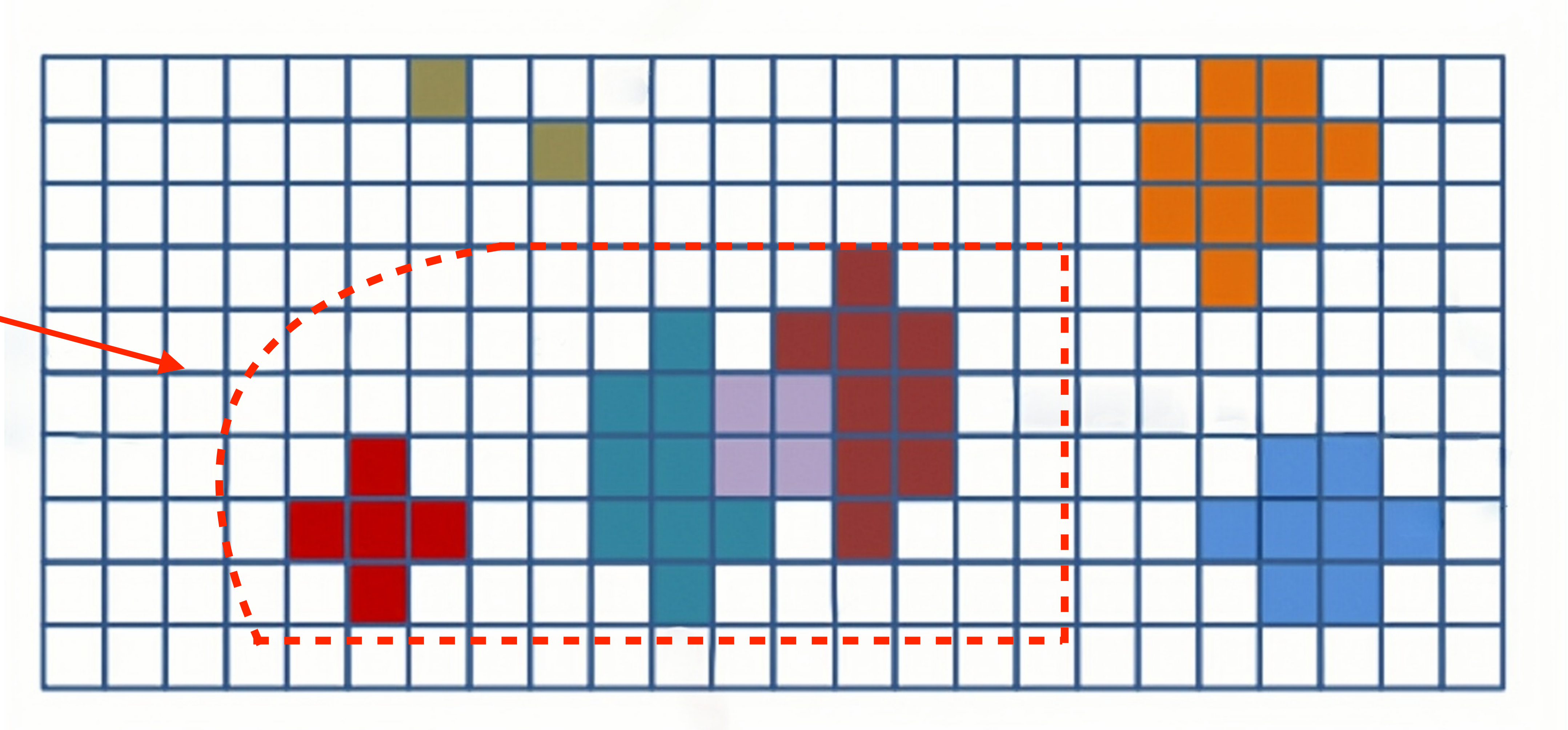


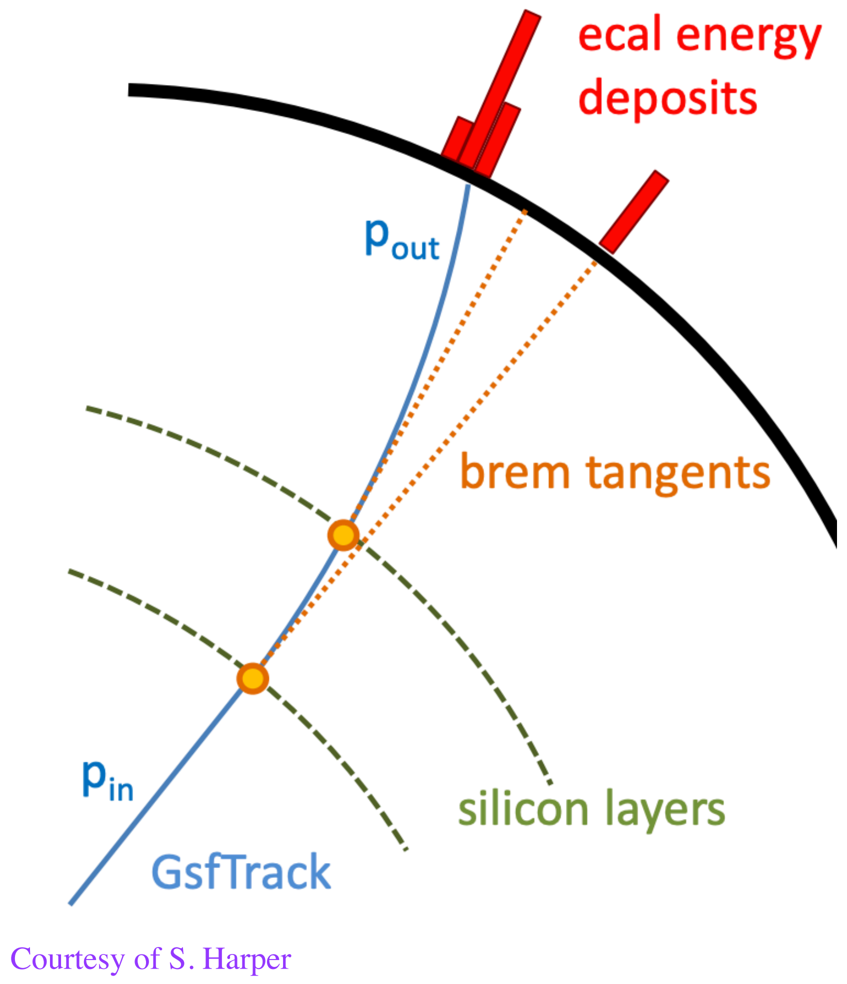


- SuperCluseters start by taking the highest energy cluster (seed cluster, in this case the **teal** one)
- Cluster is tested if it is geometrically compatible with the ***seed cluster*** in ϕ, and η
  1. Allowed η/ϕ distance to seed cluster is dependent on the energy/E_T and η of the cluster
  2. Due to B-field, bremsstrahlung can have large difference in ϕ, much smaller difference in η

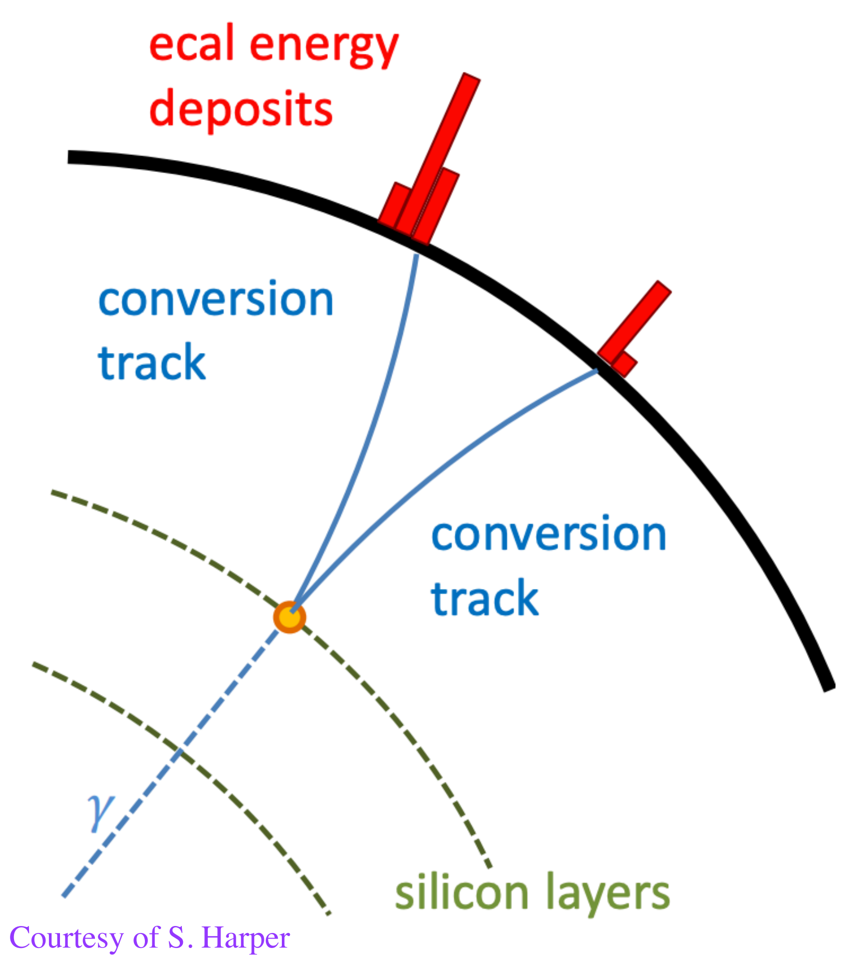
e/γ reconstruction — electron track

PF Moustache Superclusters





Brem

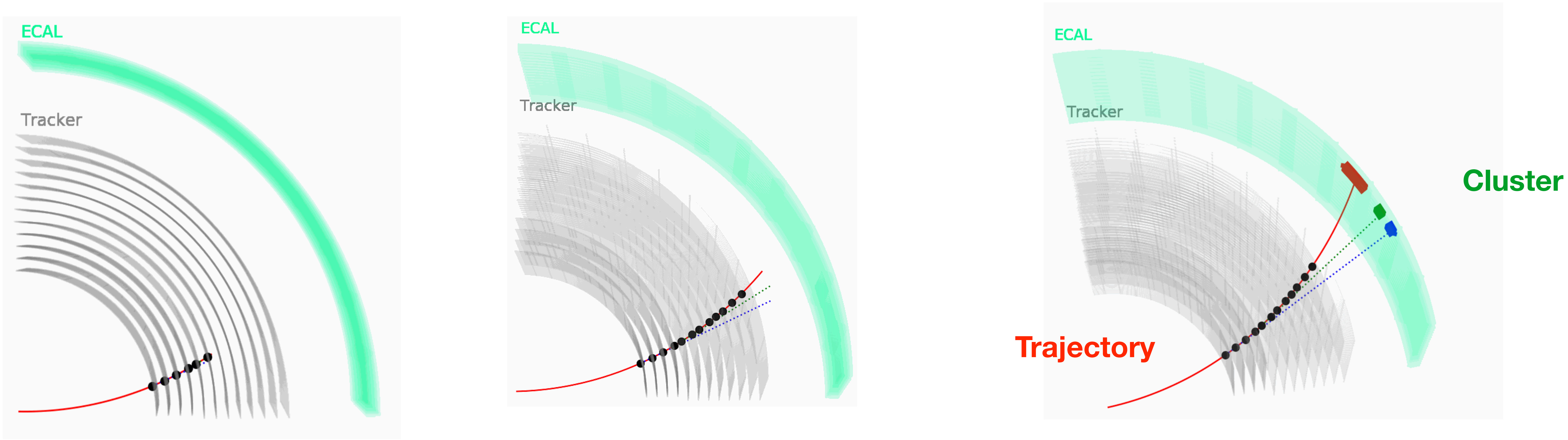


Conversions

- Several clusters could be from the same e/γ, but are far from the seed cluster in η/ϕ
- Fail to be added into the PF Moustache superclusters

Need track information !!!!

e/γ reconstruction — track seeding



Track reconstruction starts from the track **seed generation** *(identification of a hit pattern that might lie on an electron trajectory)*, such as the black point in every pixel layer as above.

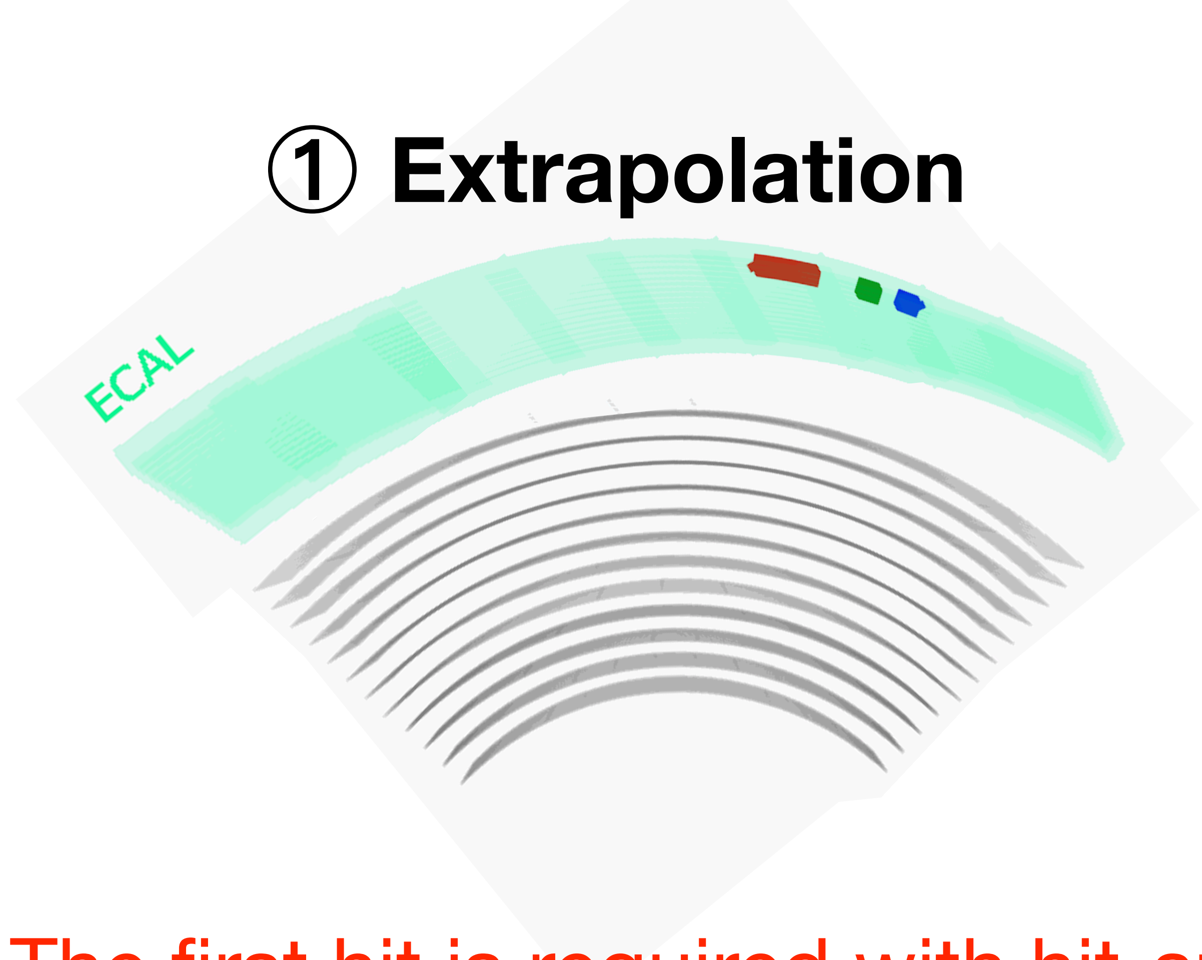
- **ECAL-based seeding** 95% for $p_T > 10$ GeV for electrons from Z boson decay
- **Tracker-based seeding** ≈50% with $p_T \approx 3$ GeV and drops to less than 5% for $p_T > 10$ GeV for electrons from Z boson decay

e/γ reconstruction — track seeding

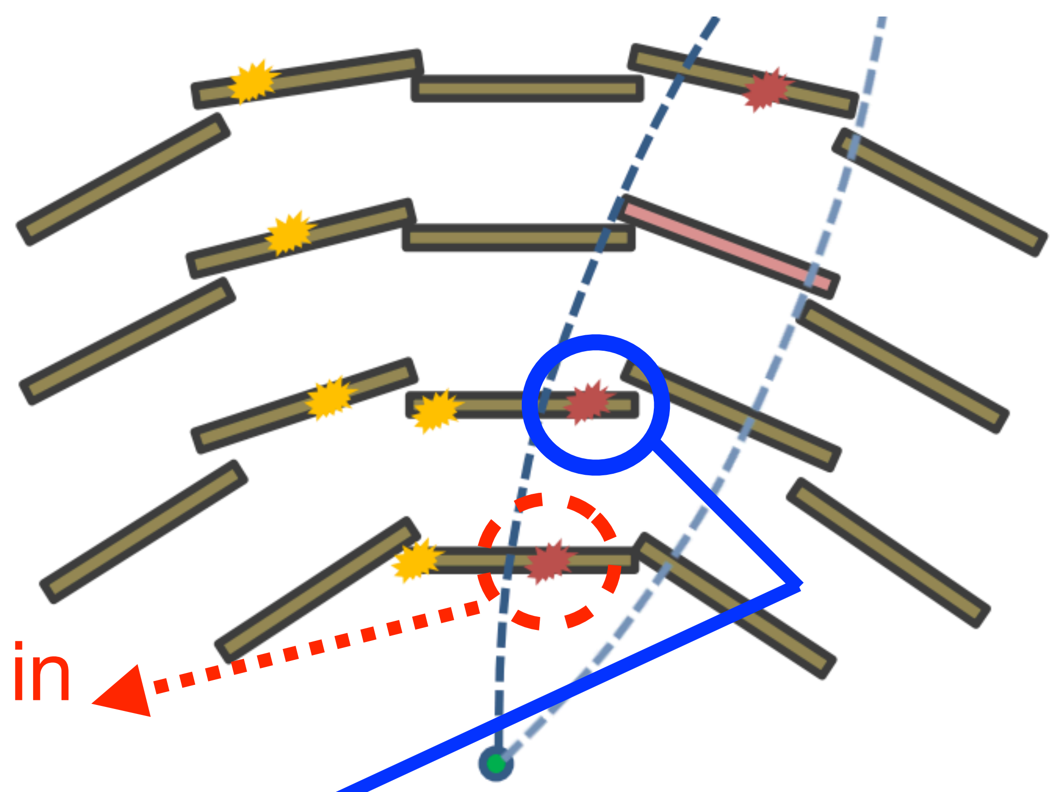
ECAL-Driven (aka outside-in)

1. Mustache SCs are used to extrapolate the collision vertex where there is no bremsstrahlung
2. The first matched hit is used to form a new trajectory starting from the beamspot

① **Extrapolation**



② **Hits matching from the most inner pixel layer**



The first hit is required with hit-supercluster matching in loose ϕ and z requirements

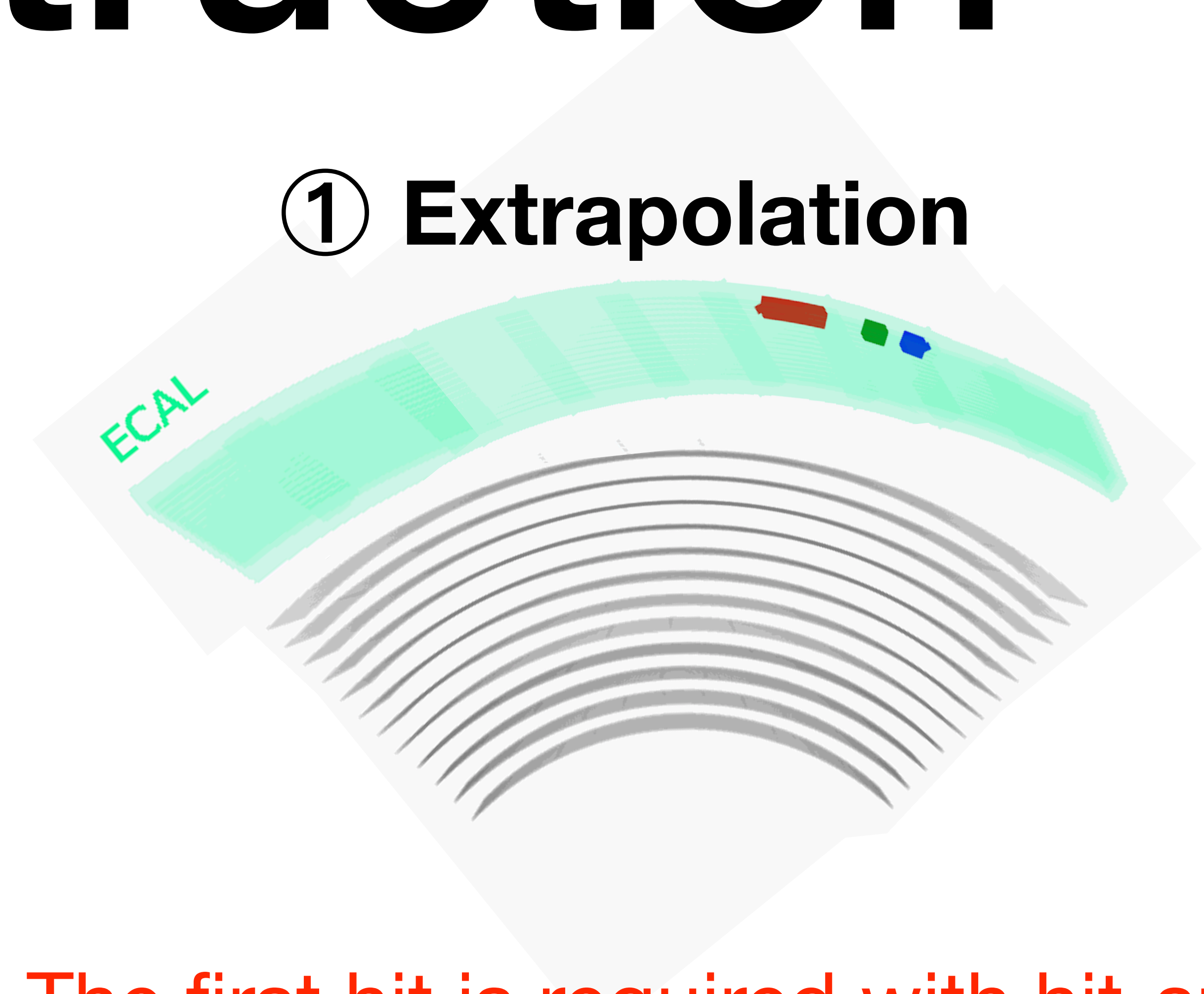
The second hit has the rather tight ϕ and z requirement

e/γ reconstruction — track seeding

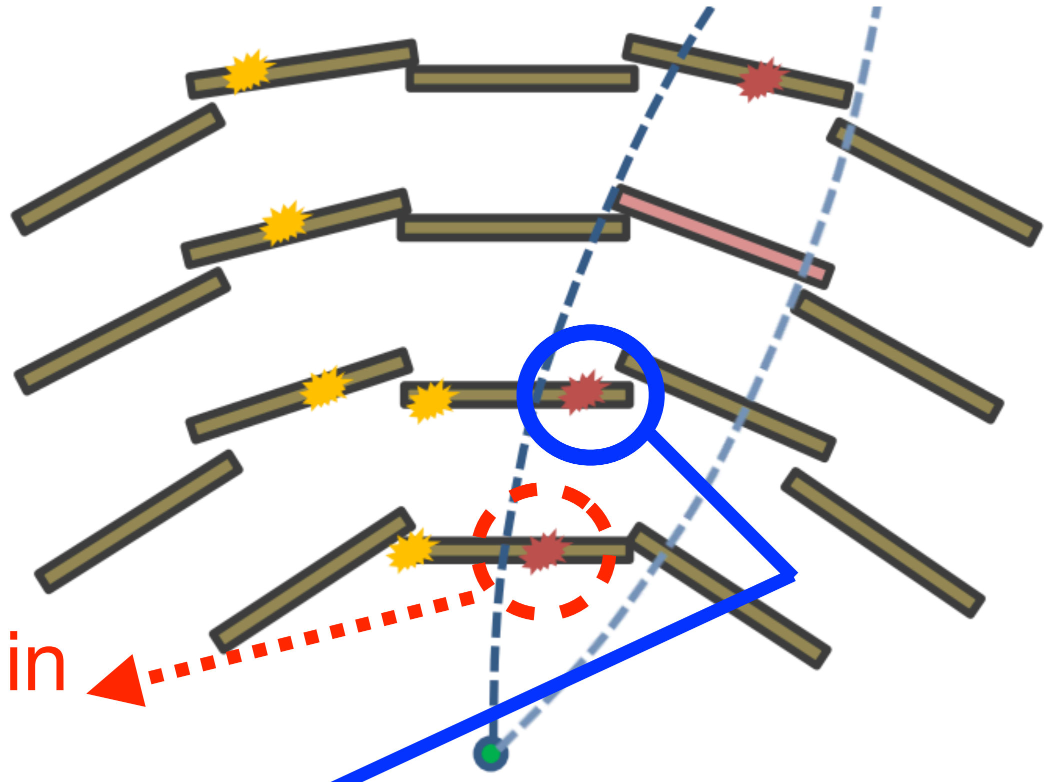
ECAL-Driven (aka outside-in)

1. Mustache SCs are used to extrapolate the collision vertex where there is no bremsstrahlung
2. The first matched hit is used to form a new trajectory starting from the beamspot

① **Extrapolation**



② **Hits matching from the most inner pixel layer**



The first hit is required with hit-supercluster matching in loose ϕ and z requirements

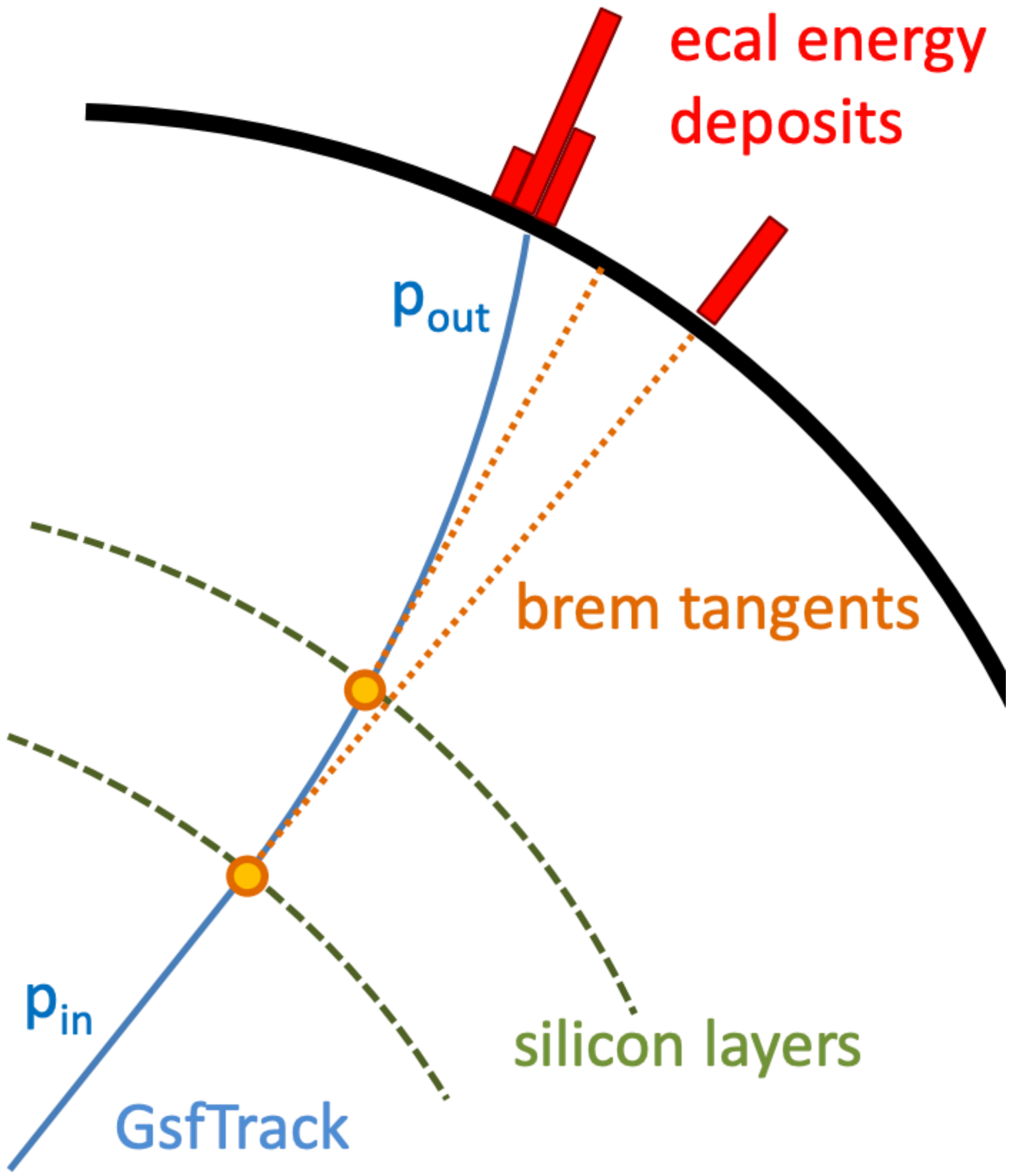
The second hit has the rather tight ϕ and z requirement

Tracker-Driven (aka inside-out)

Find a track and then matches with PFClusters using particle flow techniques

- An electron is typically as both tracker driven and ECAL driven
- **We get the hits information then can start the track reconstruction**

e/γ reconstruction — track reconstruction

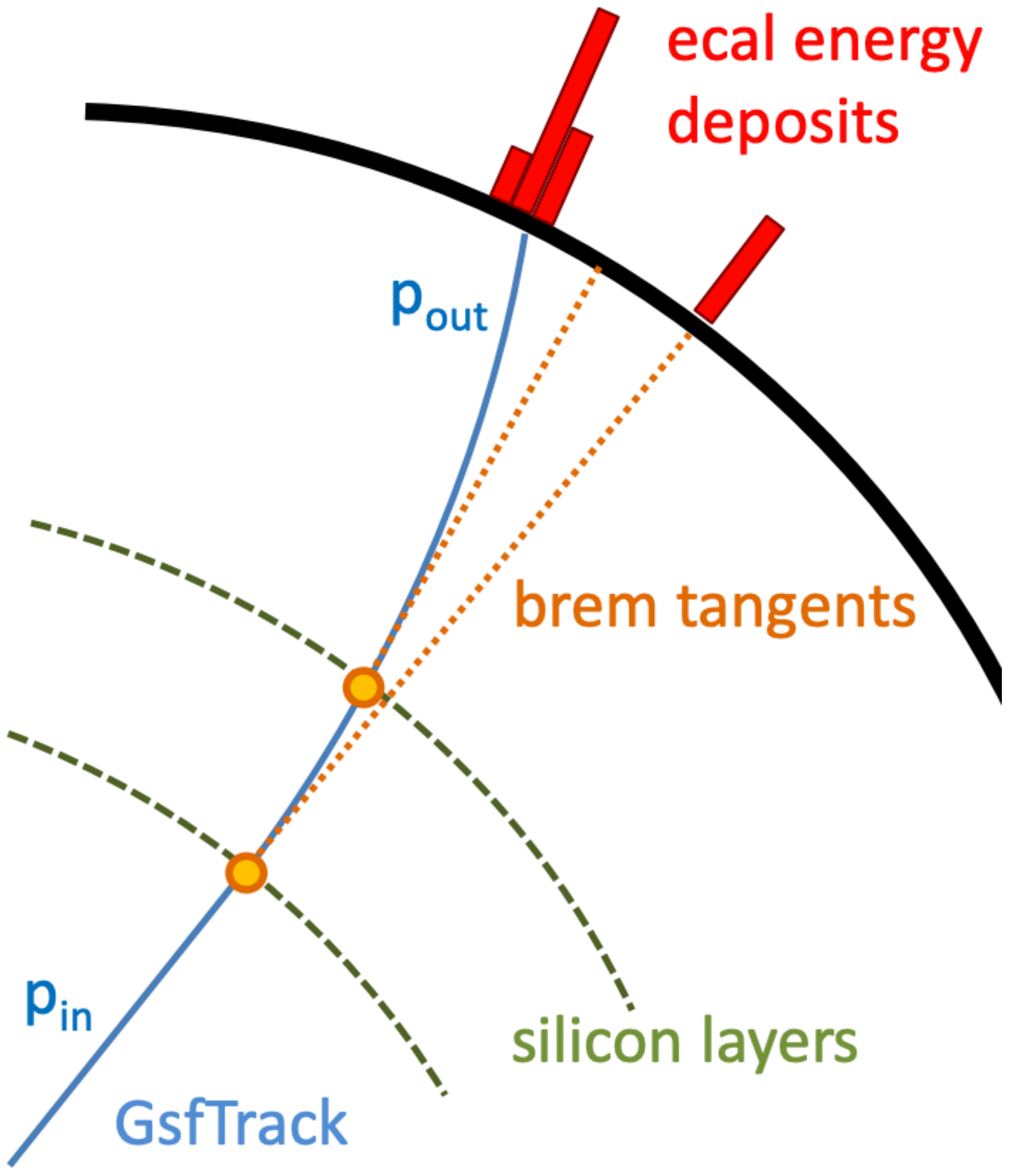


Courtesy of S. Harper

Goal is to take into account radiative losses due to brem so we can measure $\mathbf{p_{in}}$ and $\mathbf{p_{out}}$ (and p at intermediate layer)

- p_{in} = initial (or inner) momentum of the electron before it traverses the tracker, so original momentum of electron
- p_{out} = final (or outer) momentum of the electron after it has gone through the tracker and radiated photons

e/γ reconstruction — track reconstruction



Courtesy of S. Harper

Goal is to take into account radiative losses due to brem so we can measure $\mathbf{p_{in}}$ and $\mathbf{p_{out}}$ (and p at intermediate layer)

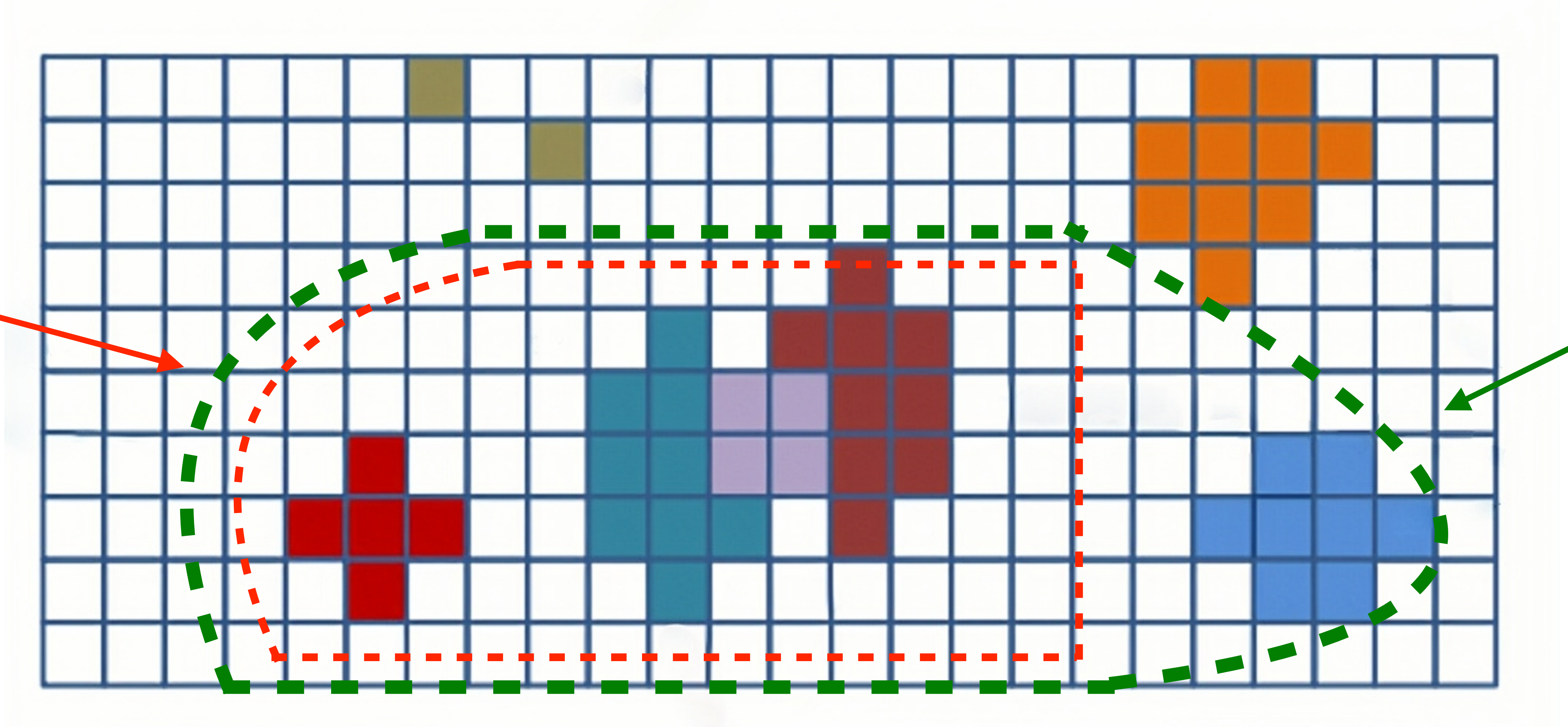
- p_{in} = initial (or inner) momentum of the electron before it traverses the tracker, so original momentum of electron
- p_{out} = final (or outer) momentum of the electron after it has gone through the tracker and radiated photons

CMS use a dedicated tracking algorithm known as **GSF** tracking and the track is known as GSF track

Used to product brem tangents for later **matching to missed ECAL deposits**

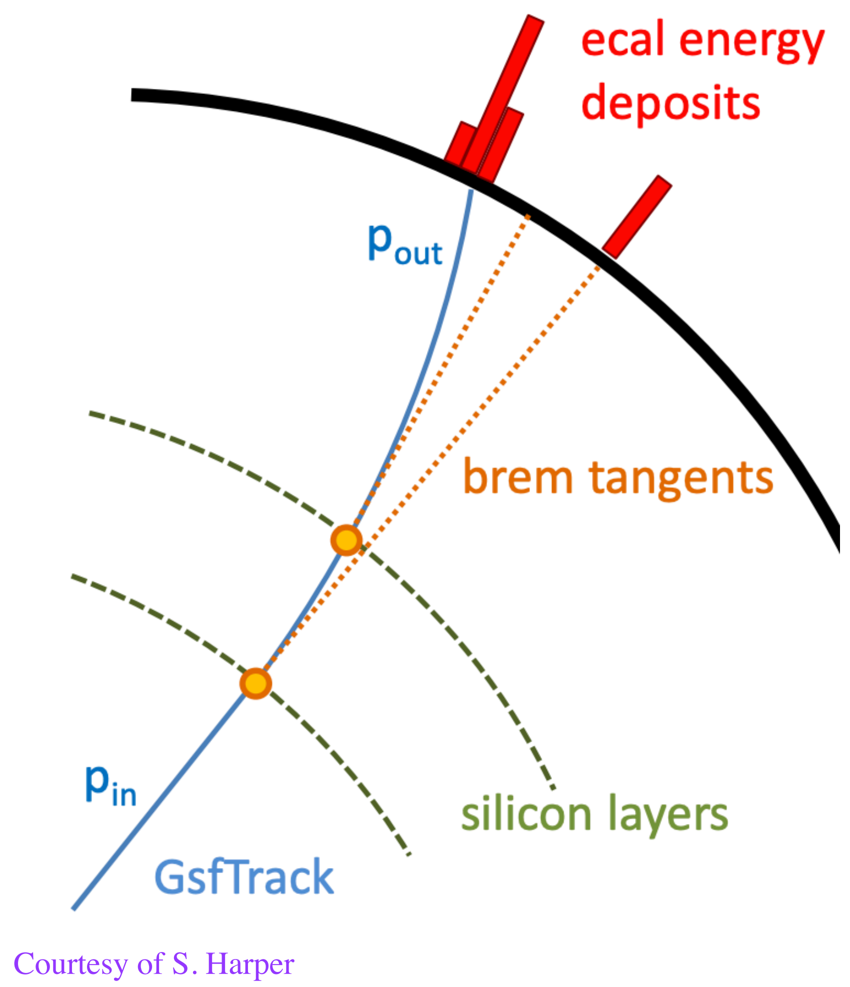
e/γ reconstruction — Refined SuperClustering

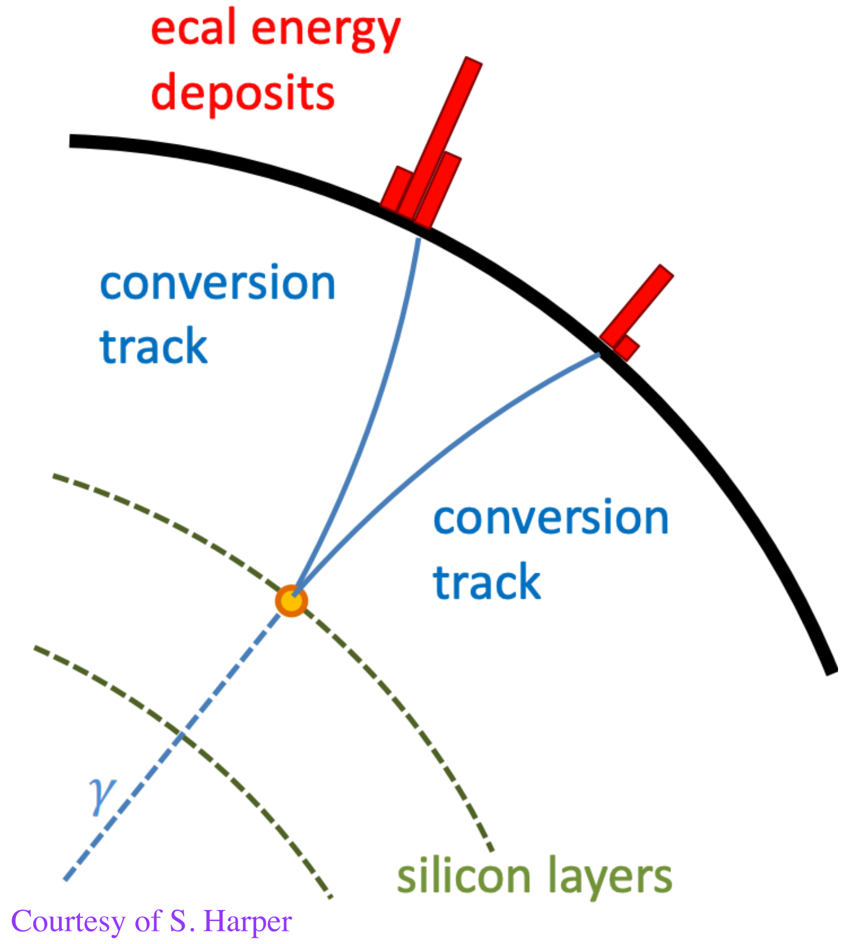
PF Moustache Superclusters



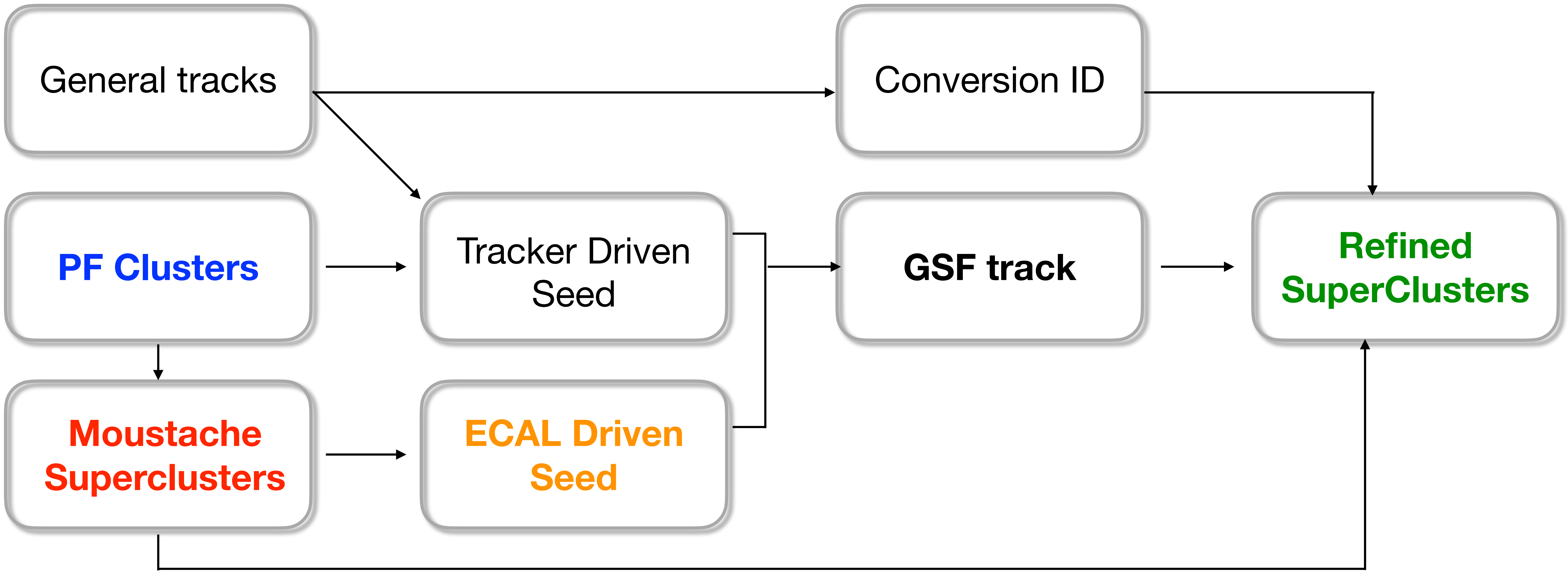
Refined Supercluster recovers bremsstrahlung and adds it to the PF moustache SC

- Associate additional bremsstrahlung to the moustache supercluster by looking at GSF track tangents
- Associate additional clusters matched conversion tracks
- Reject some clusters which are highly incompatible with its matched tracks
- This is a **refined supercluster**





e/γ reconstruction

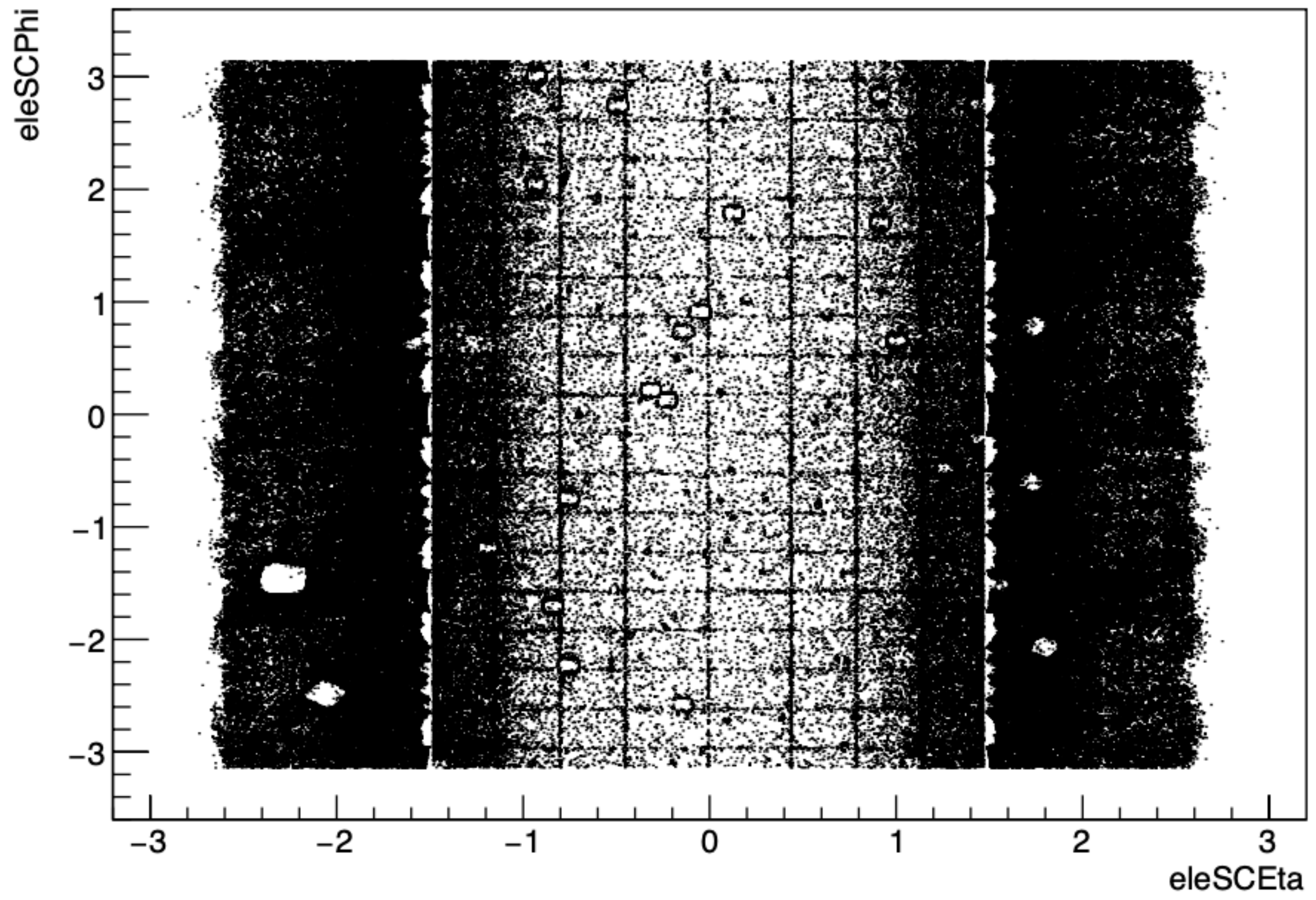


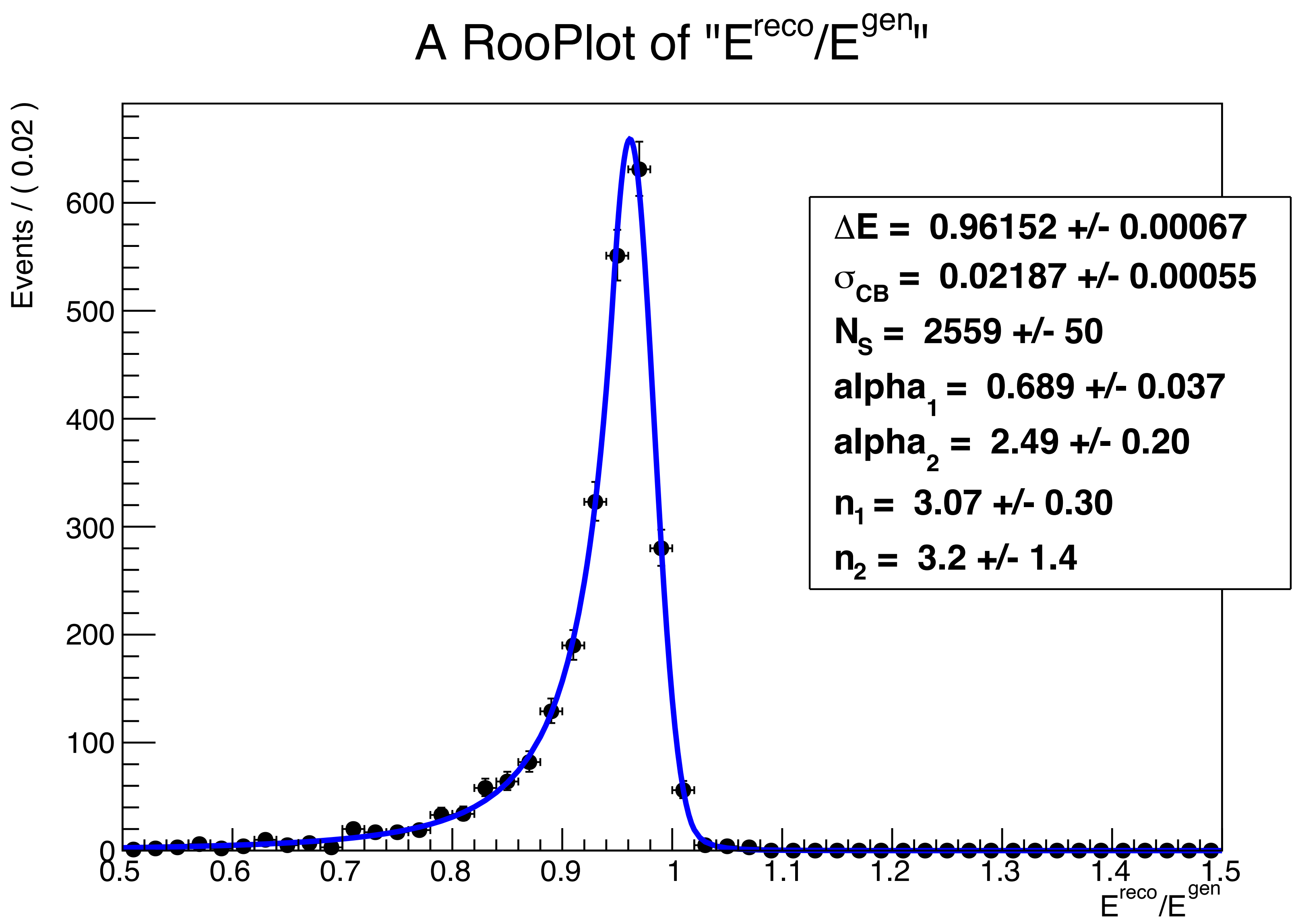
- PF clusters are energy-deposits of electron and photon in the ECAL
- Moustache Superclusters are from the combination of the individual electrons + photons
- ECAL Diven seed is the method to find hits in pixels as seeds for track reconstruction
- GSF track is reconstructed by GSF algorithm in the step of electron trajectory building
- Refined Superclusters are the superclusters used for all supercluster related with e/γ

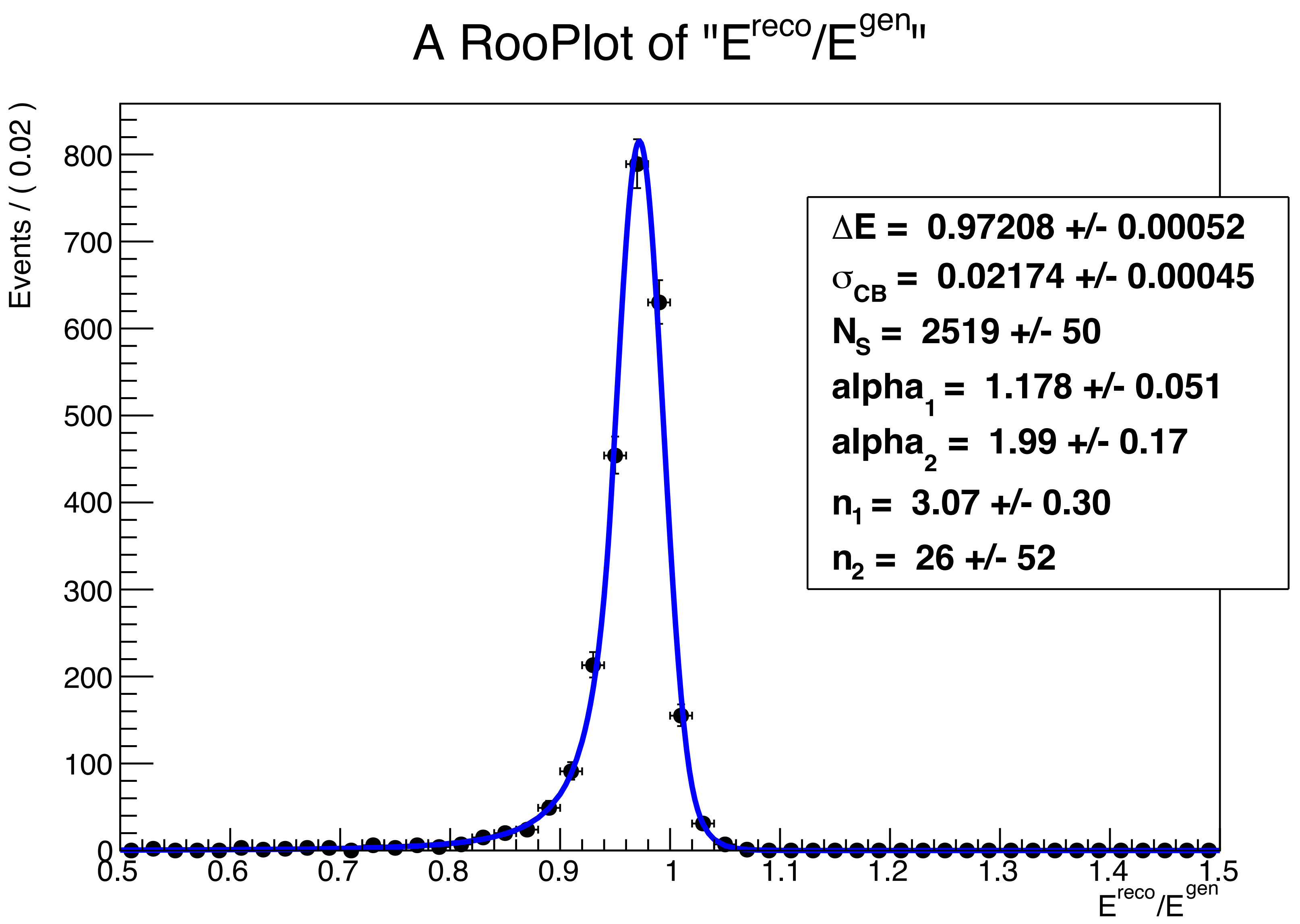
Exercise 2.1

1. Draw the distributions of the supercluster energy, seed cluster energy, and p_{in} from GSF track
   - Branch name: `ele_SCRawEn` (Refined SC), `ele_SeedRawEn`, and `ele_GsdTrkPInn`
2. Use the variable `ele_SCRawEn` to see the resolution distribution (divided by `eleGen_pt`) in different η regions and fit it with different functions (functions `makeDCBFit` and `makeCruijffFit`)
   - What do you observe? Why the mean value is not very close to 1?
3. Draw the occupancy plot of η and ϕ for electrons in ECAL
   - What does this occupancy plot means?

e/γ energy regression

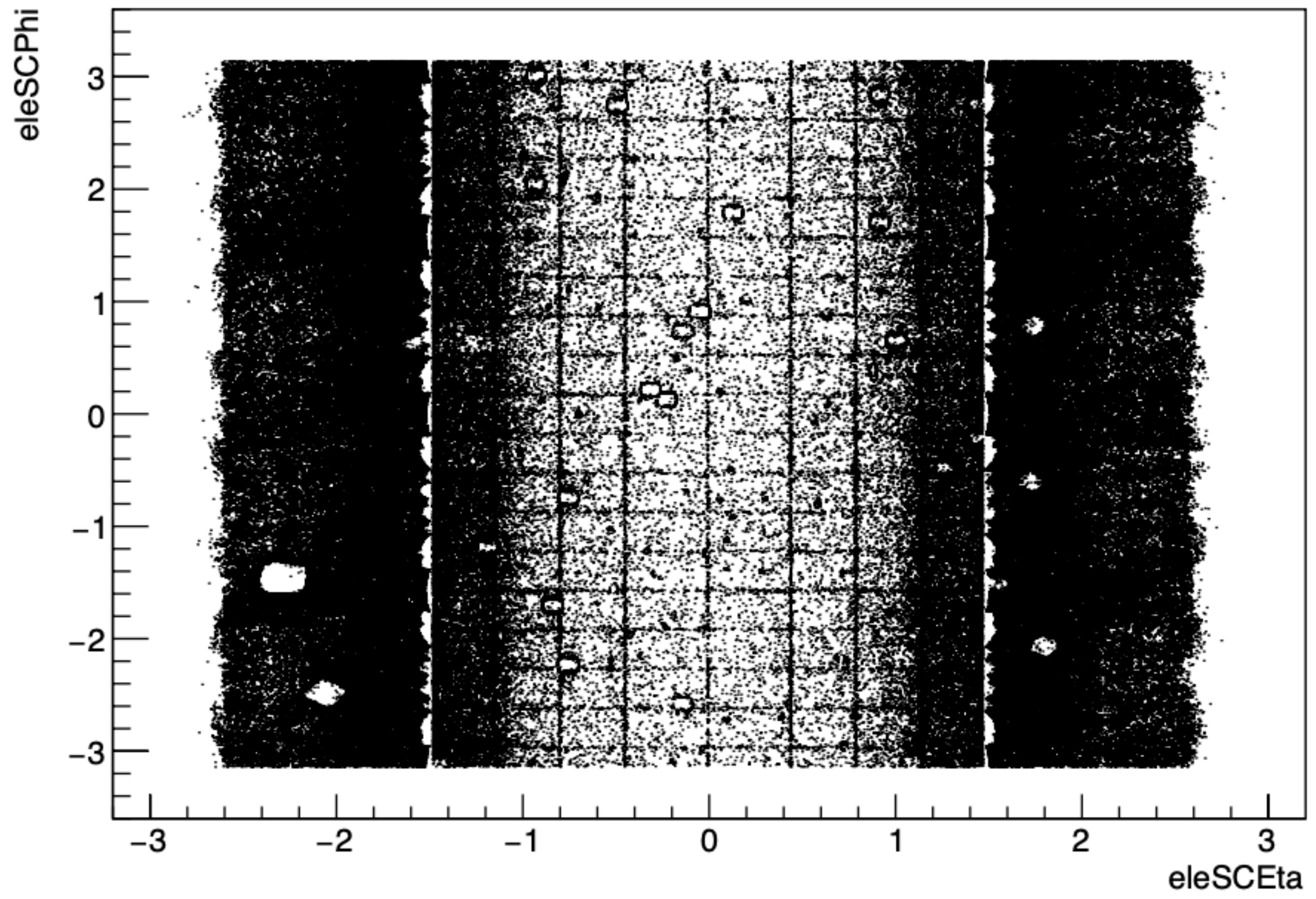


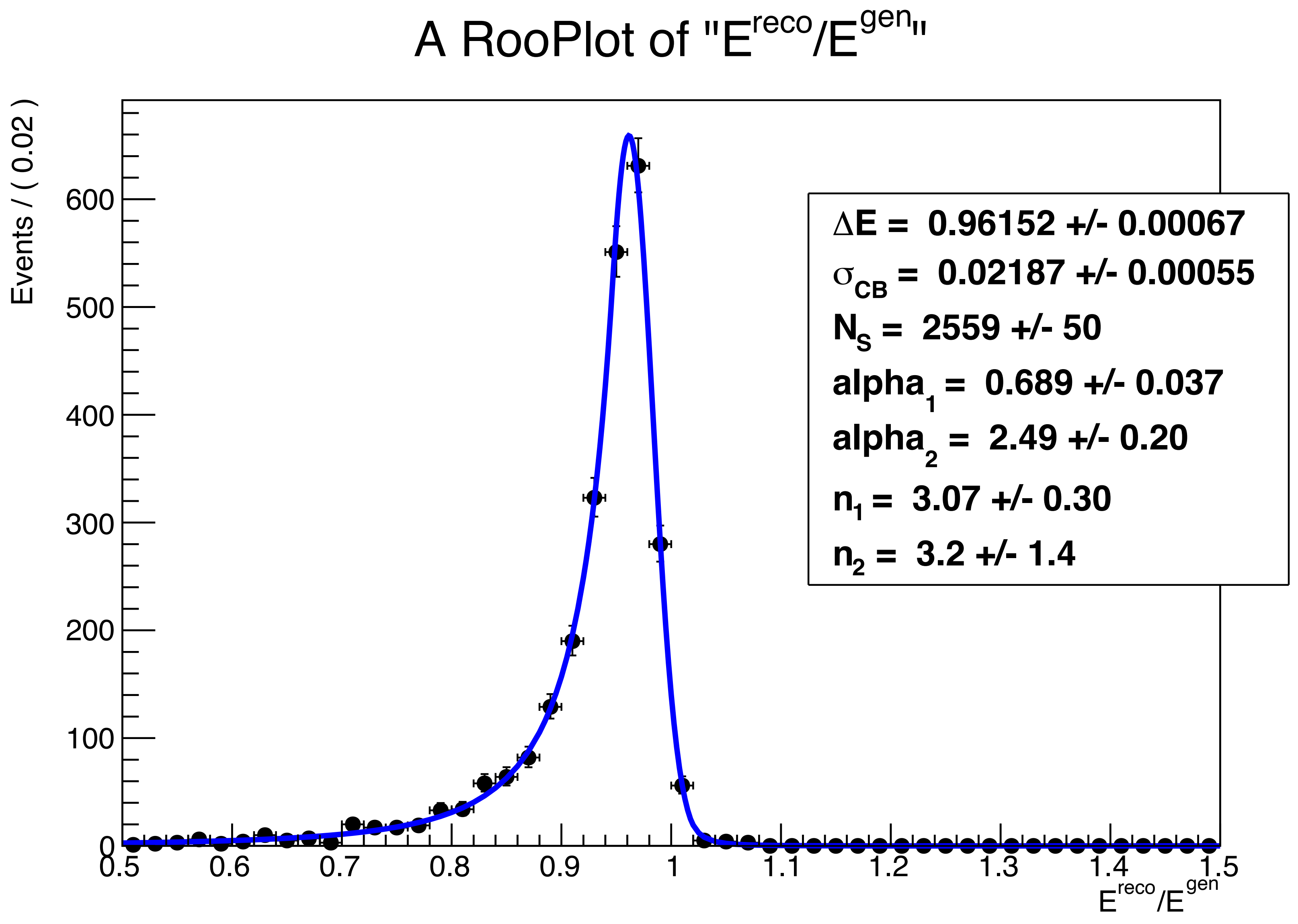


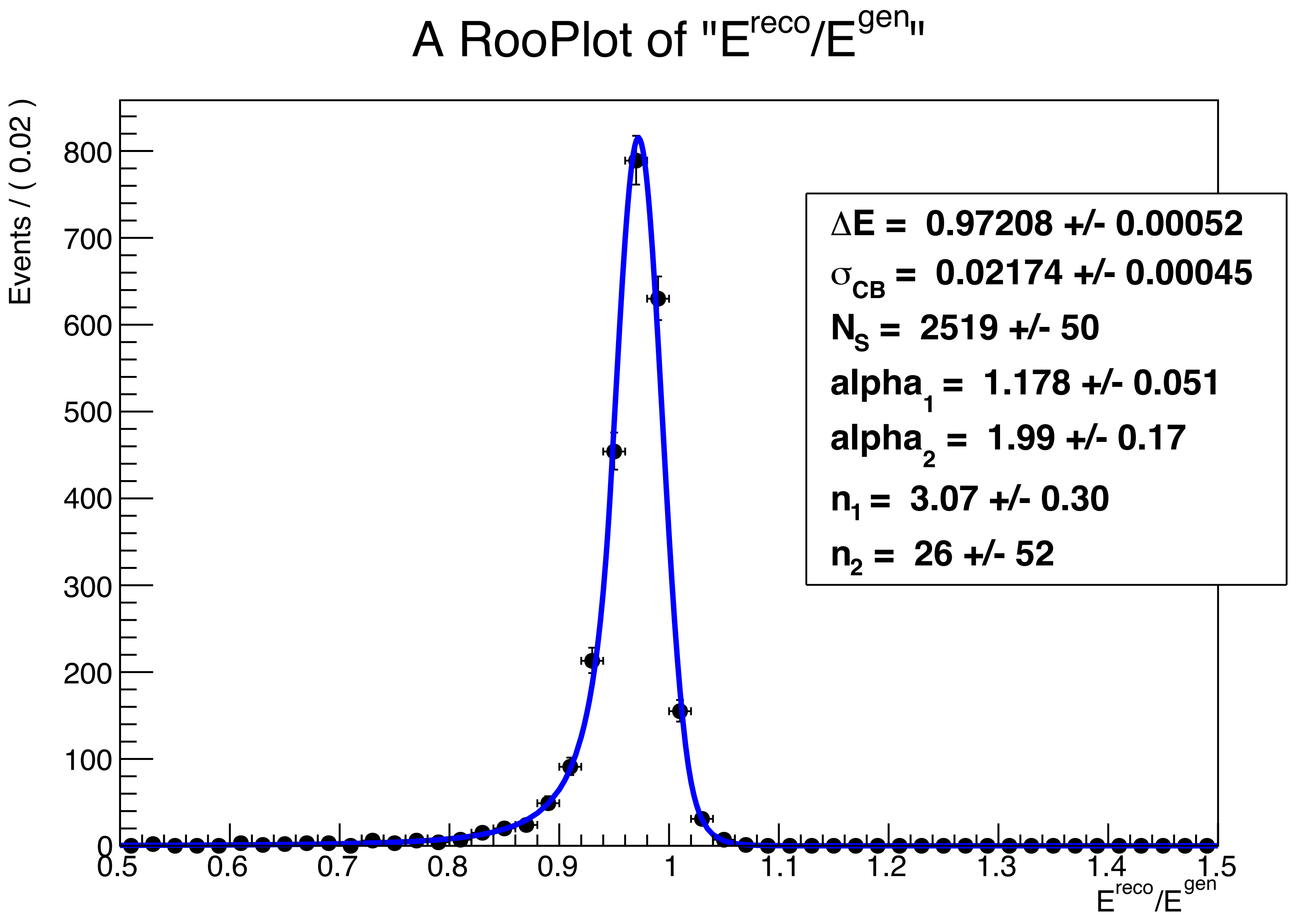


Mean value of the $E^{reco}/E^{gen} \neq 1$, means that there are energy loss and the effect is different in different detector regions

e/γ energy regression





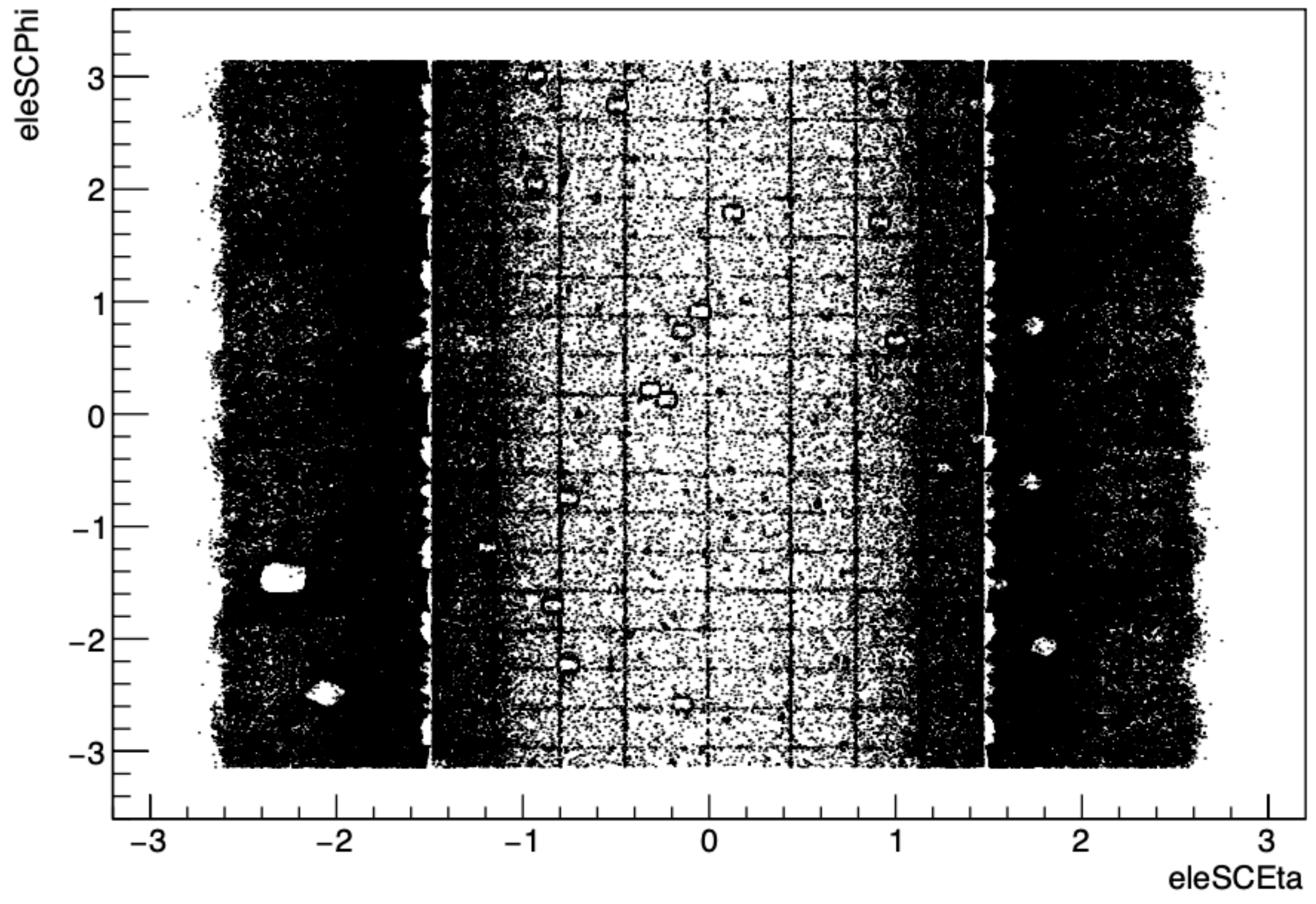


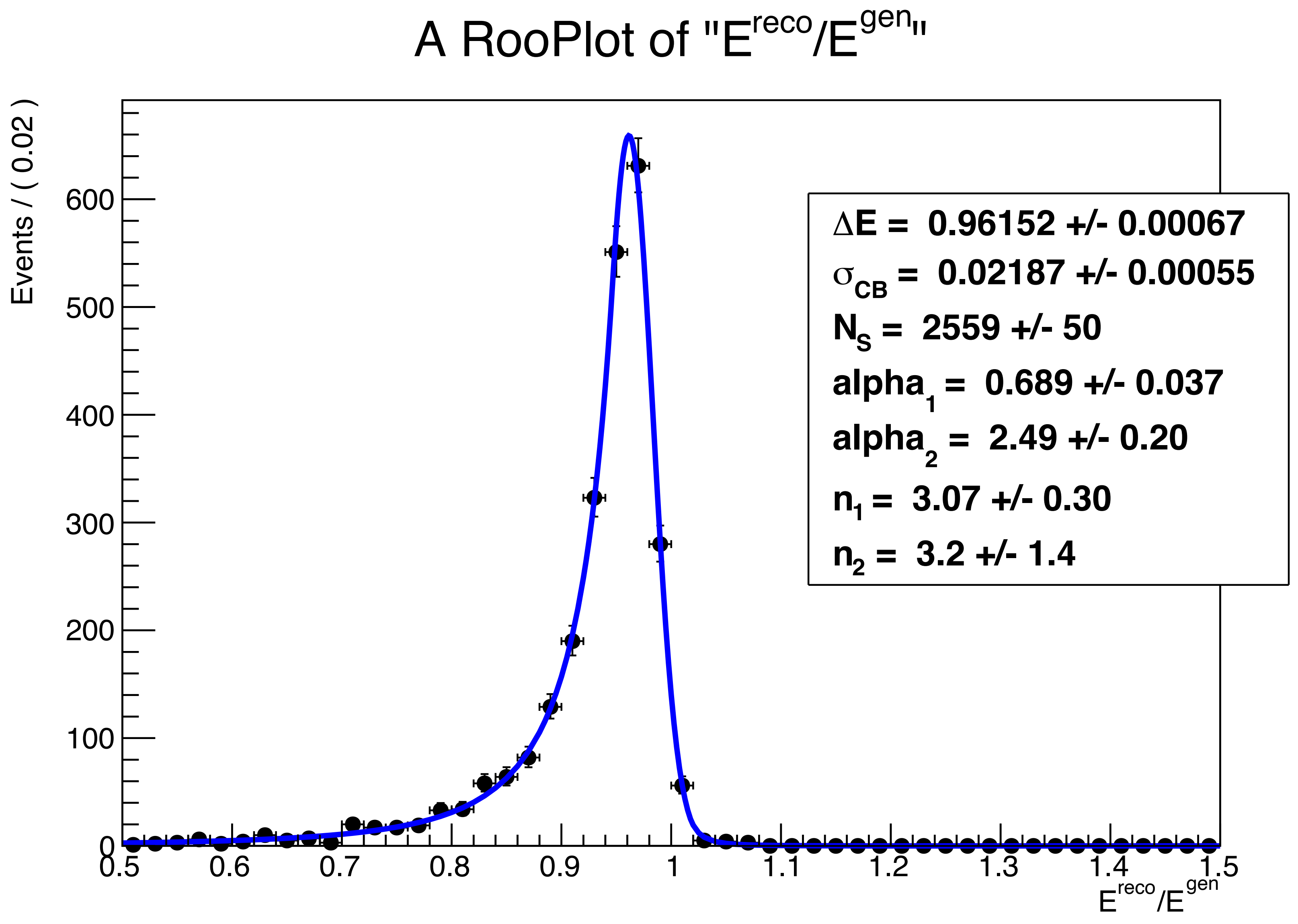
Mean value of the $E^{reco}/E^{gen} \neq 1$, means that there are energy loss and the effect is different in different detector regions

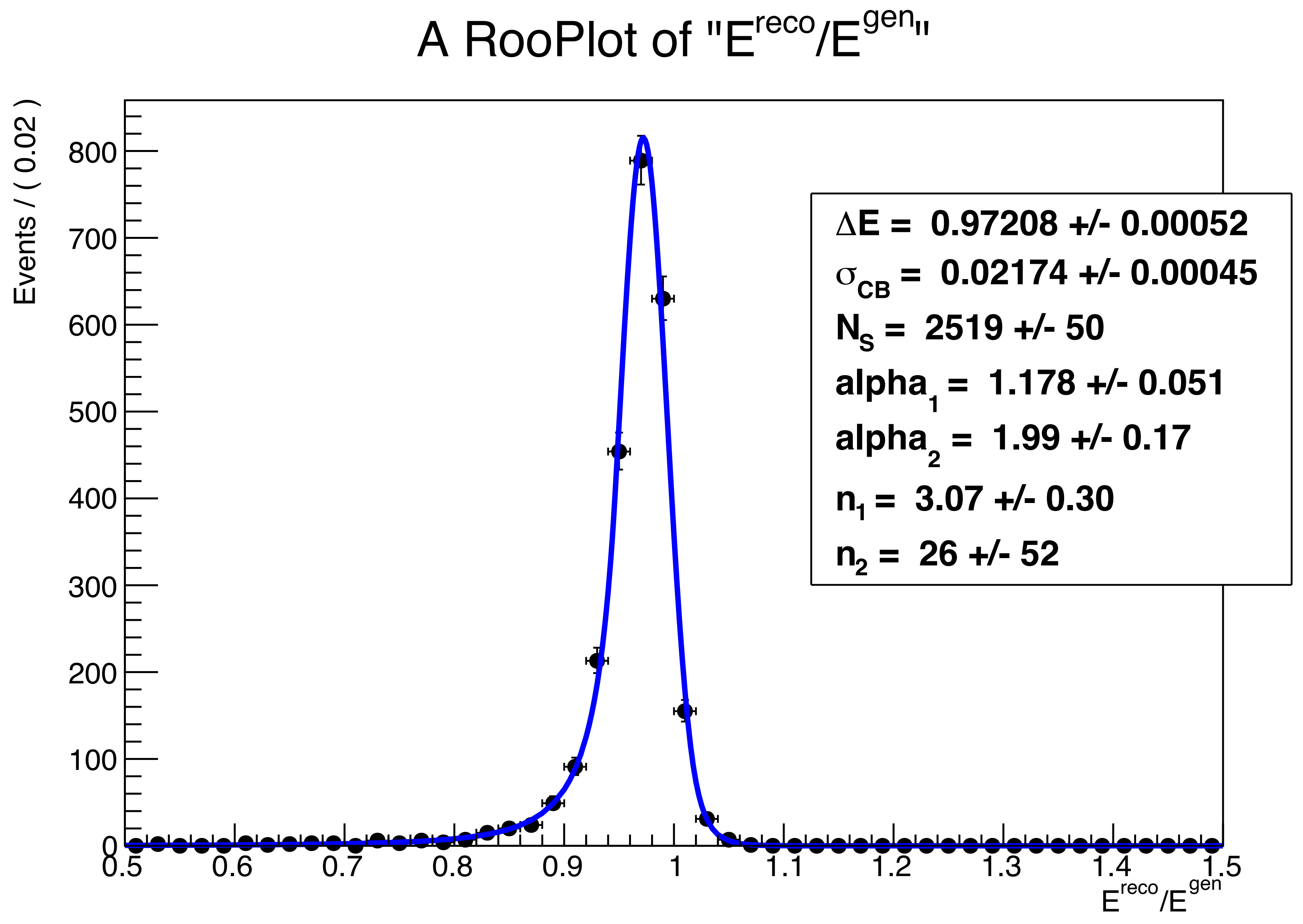
Energy loss for several reasons

- Energy lost in gaps, lateral and longitudinal shower leakage, and dead crystals
- Large amount of material upstream of the calorimeter
- The presence of pileup, energy is mis-reconstructed

e/γ energy regression







Mean value of the $E^{reco}/E^{gen} \neq 1$, means that there are energy loss and the effect is different in different detector regions

The **regression technique** is used to derive corrections so that the reconstructed energy is calibrated back to the generator level energy

- **Based on simulation**
- Uses machine learning techniques
- Applied on data and MC

The ECAL-Track (E-p) combination is only performed for electrons with energies less than 200 GeV

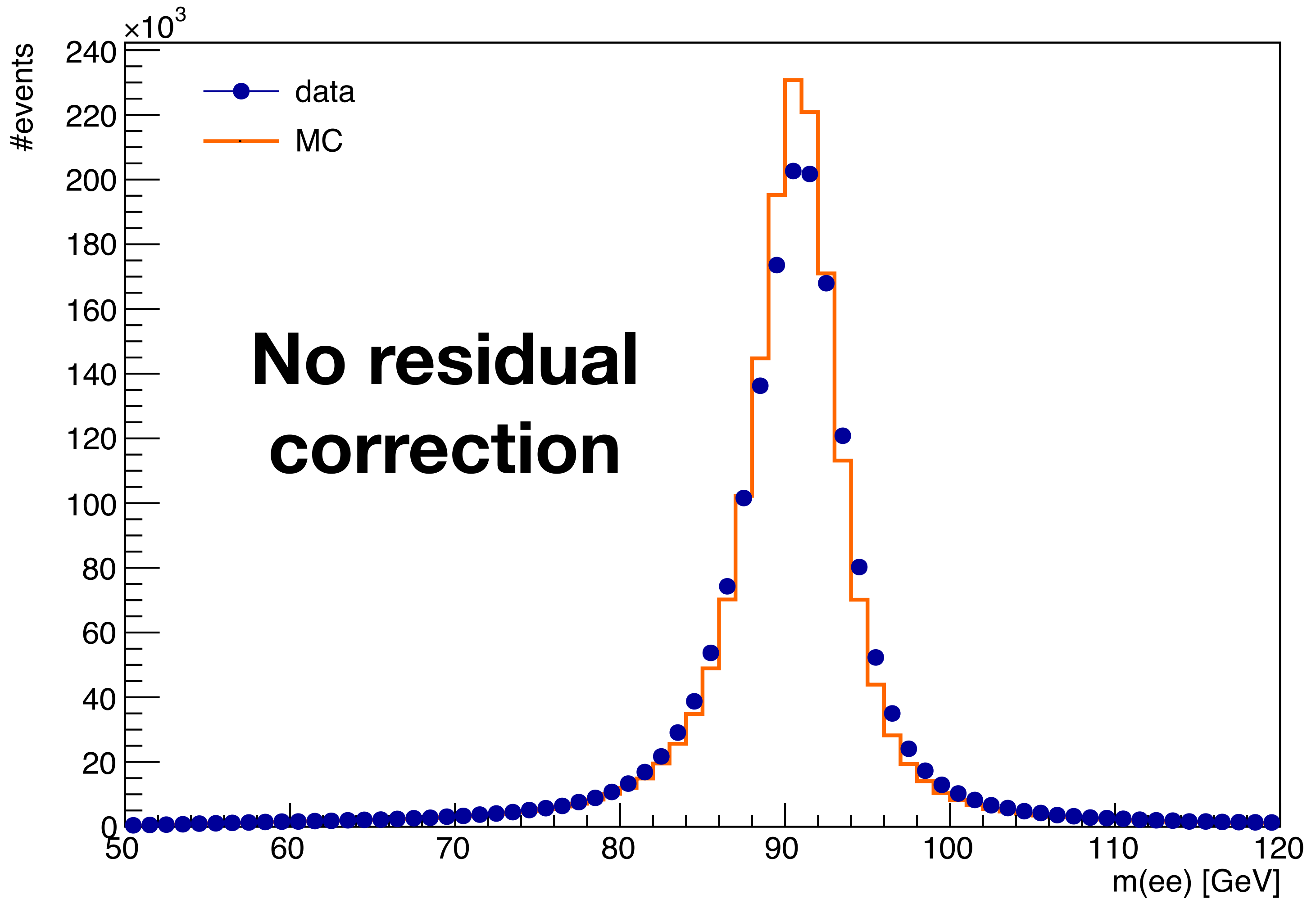
Energy loss for several reasons

- Energy lost in gaps, lateral and longitudinal shower leakage, and dead crystals
- Large amount of material upstream of the calorimeter
- The presence of pileup, energy is mis-reconstructed

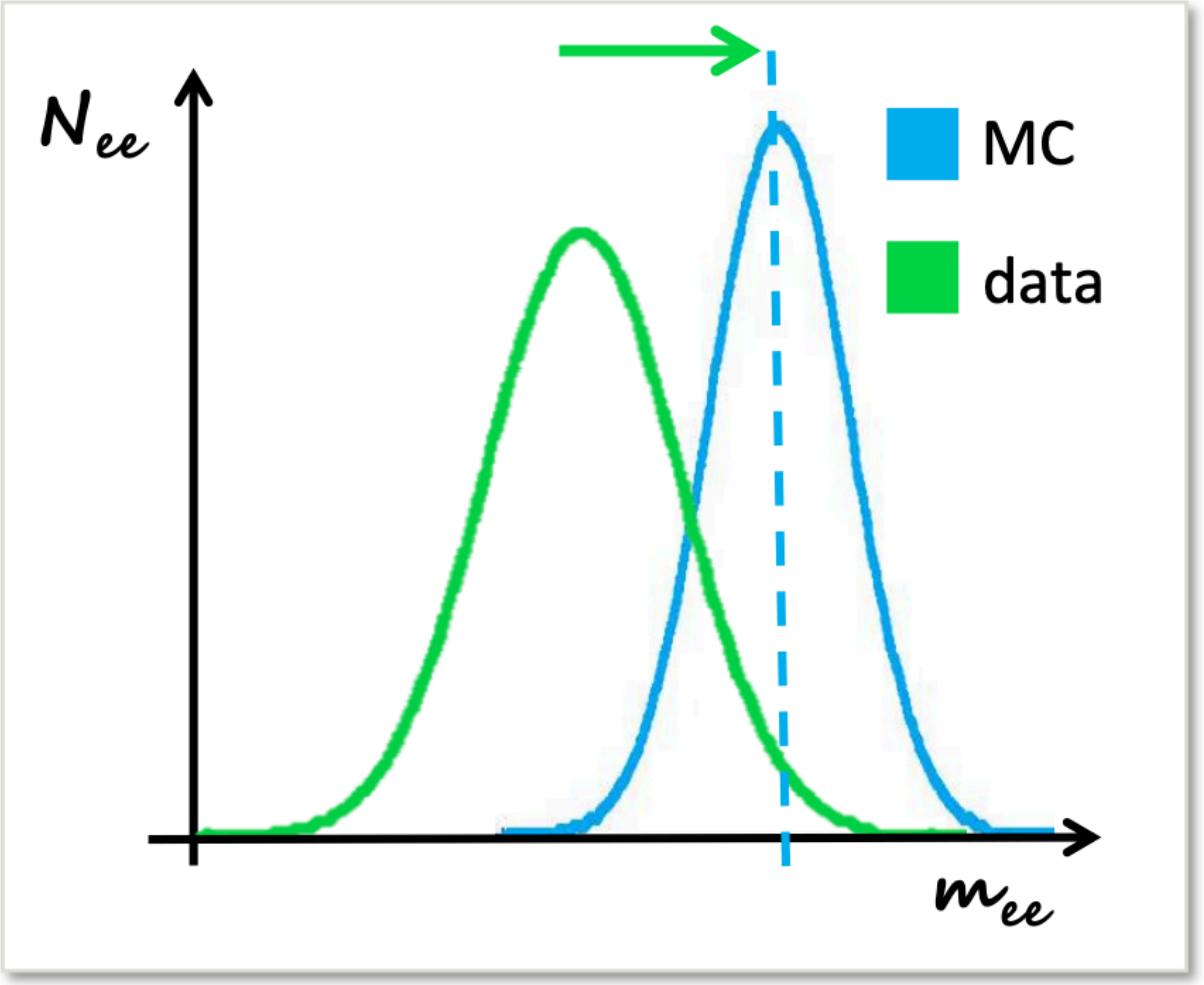
Exercise 2.2

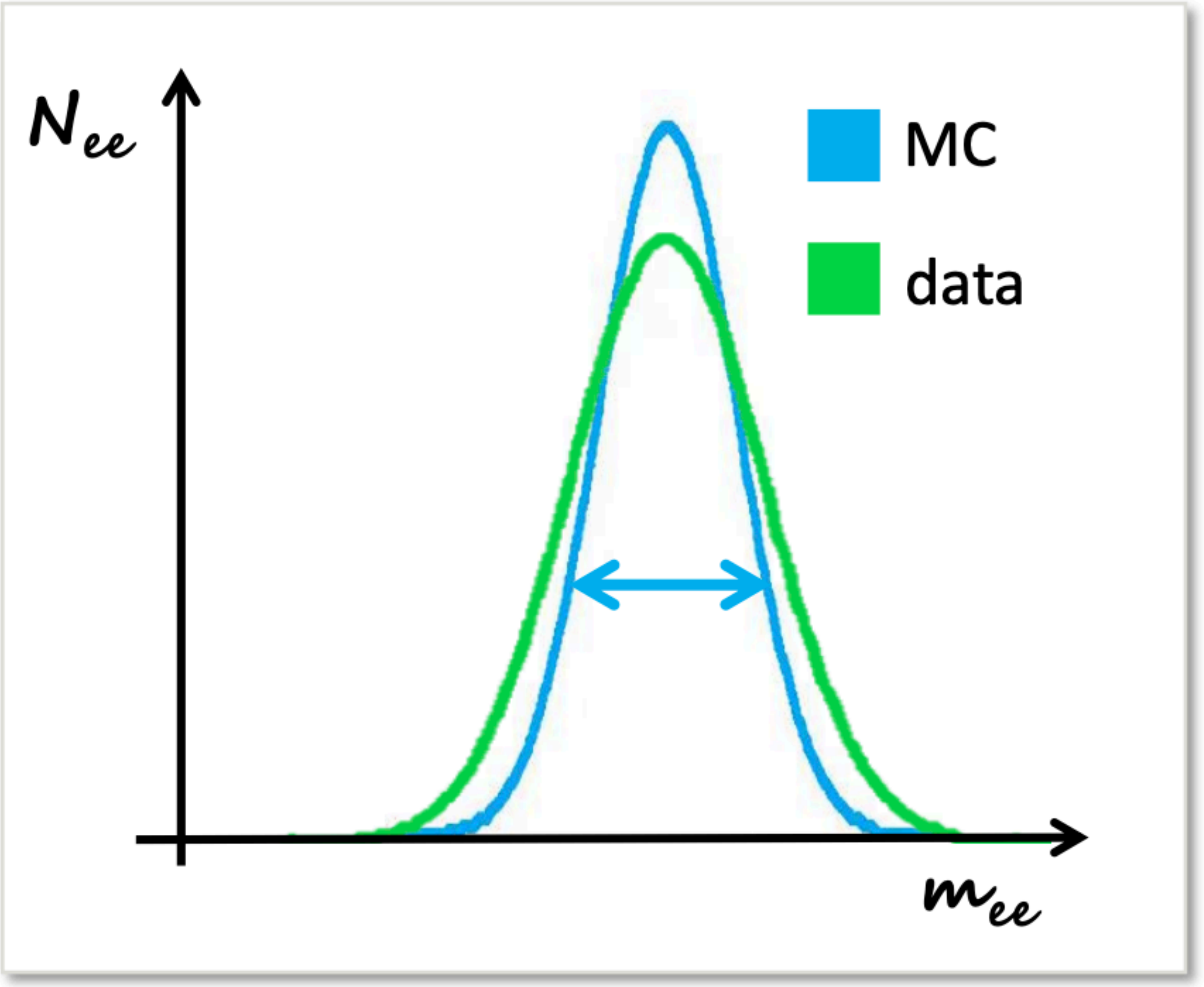
1. Correct the `ele_SCRawEn/ele_GenEn` with the mean value got in the fit of the Exercise 2.1 and repeat the fit
2. Draw distributions of `ele_SCRawEn, ele_SCEn, ele_EcalEn`, and `ele_En`
   - Give explanations
3. When `eleGen_pt` ranges from (10, 20), (20, 30), (30, 40), and (40, 50), fit the resolution distributions of `ele_SCEn/eleGen_pt`, `ele_GsdTrkPInn/eleGen_pt`, and `ele_En/eleGen_pt`
   - What do you observe from the variations of the resolution with ECAL or tracker?
4. Compare the Z mass distribution after the energy regression

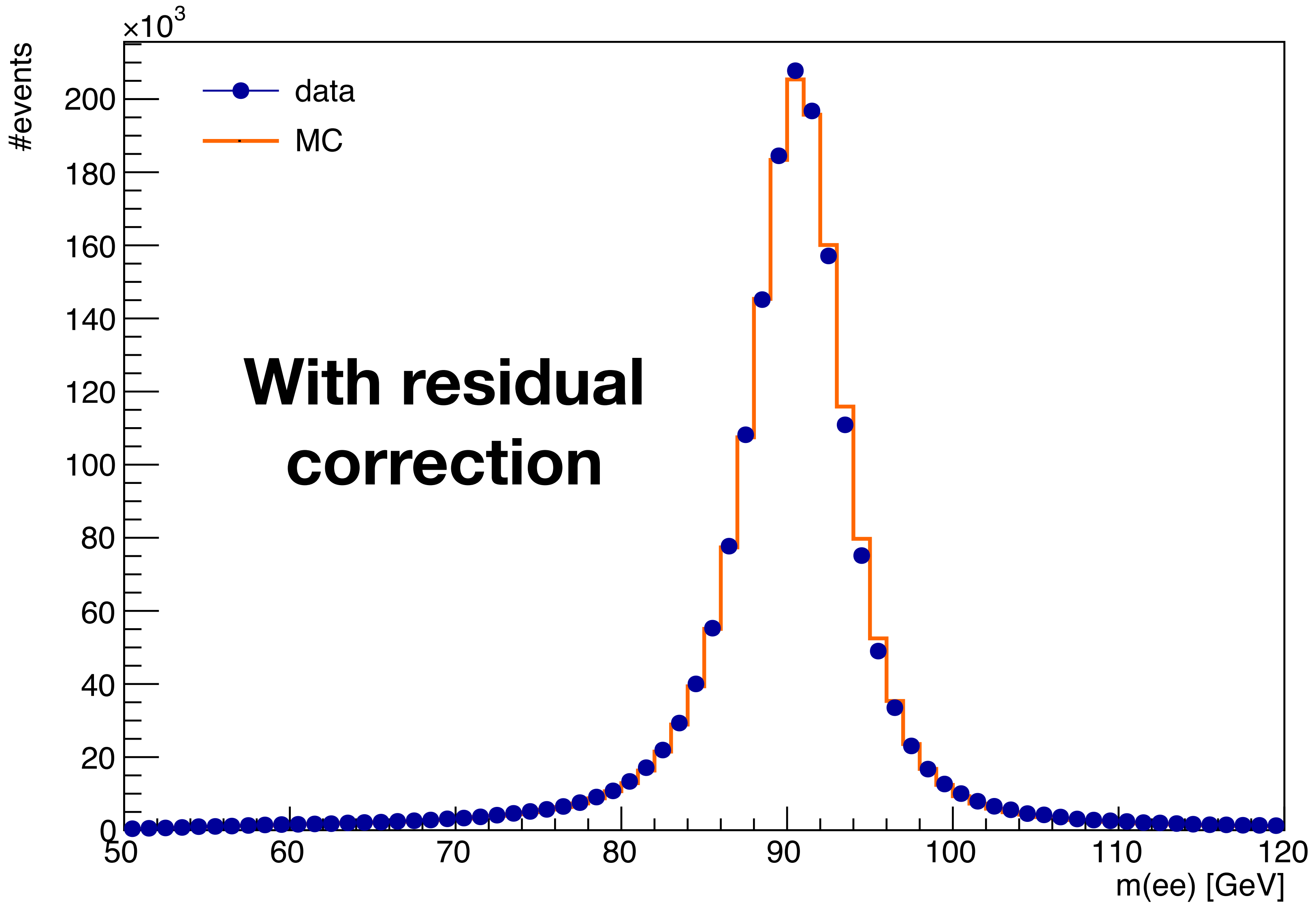
e/γ energy residual corrections



- Regression correction in Refined supercluster is MC-based
- **Residual data/MC discrepancies** corrected using the Z mass and width, by comparing Z → ee events in data and MC
- Simultaneously adjust **energy scale** (data) and **resolution** (MC) → e/γ energy scale& smearing correction







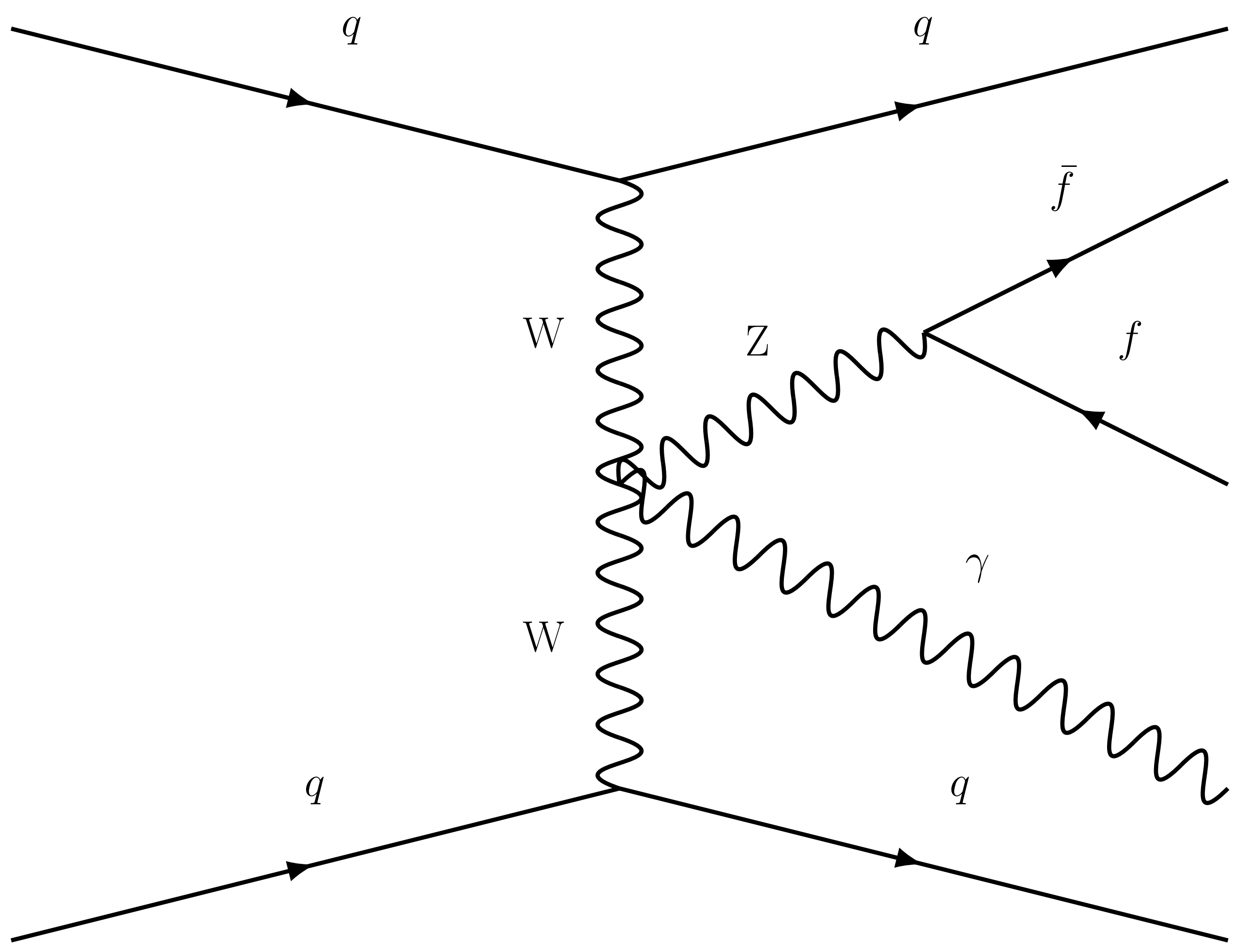
Exercise 2.3

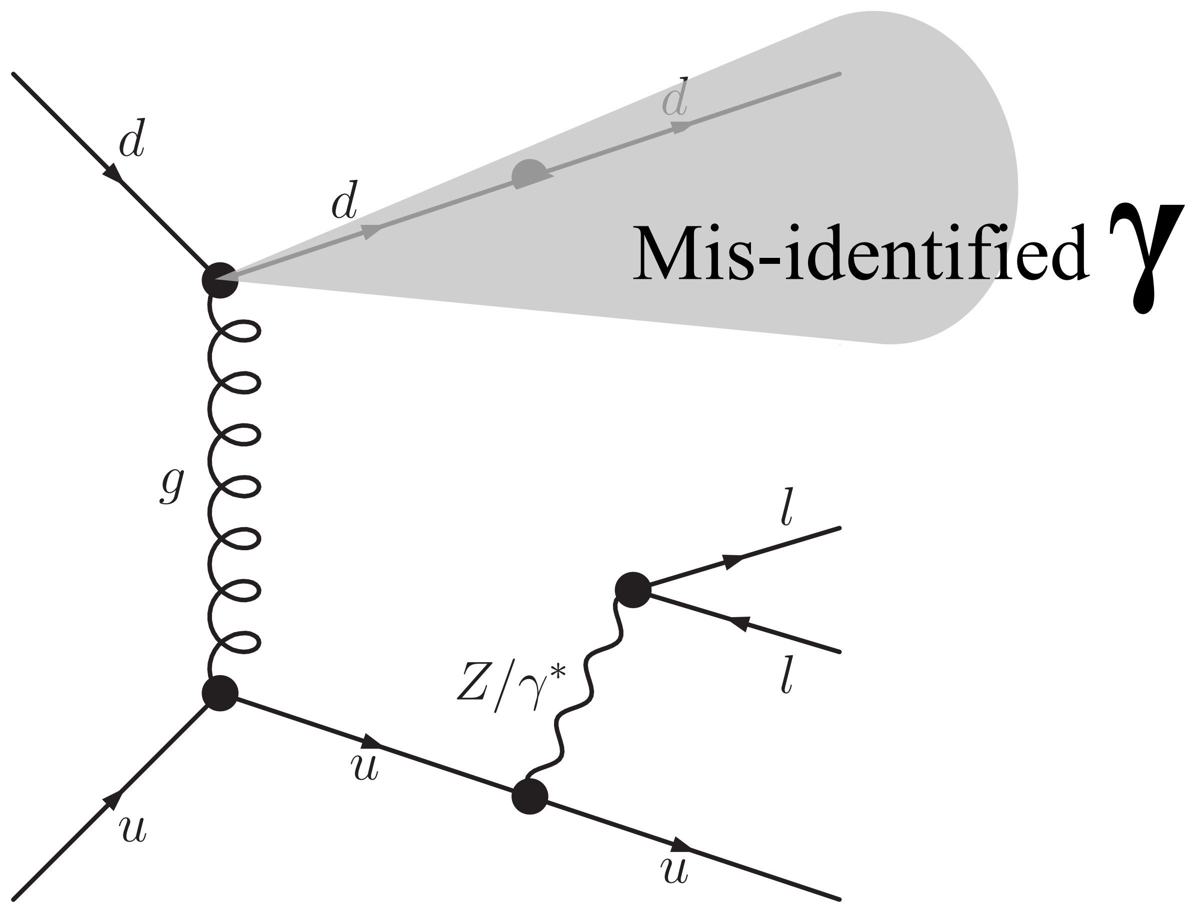
1. Draw the Z mass distribution after the residual energy correction
2. Draw a workflow of the full electron reconstruction including the energy correction

e/γ identification

e/γ identification

Identify an/a electron/photon → Select a real/prompt electron/photon





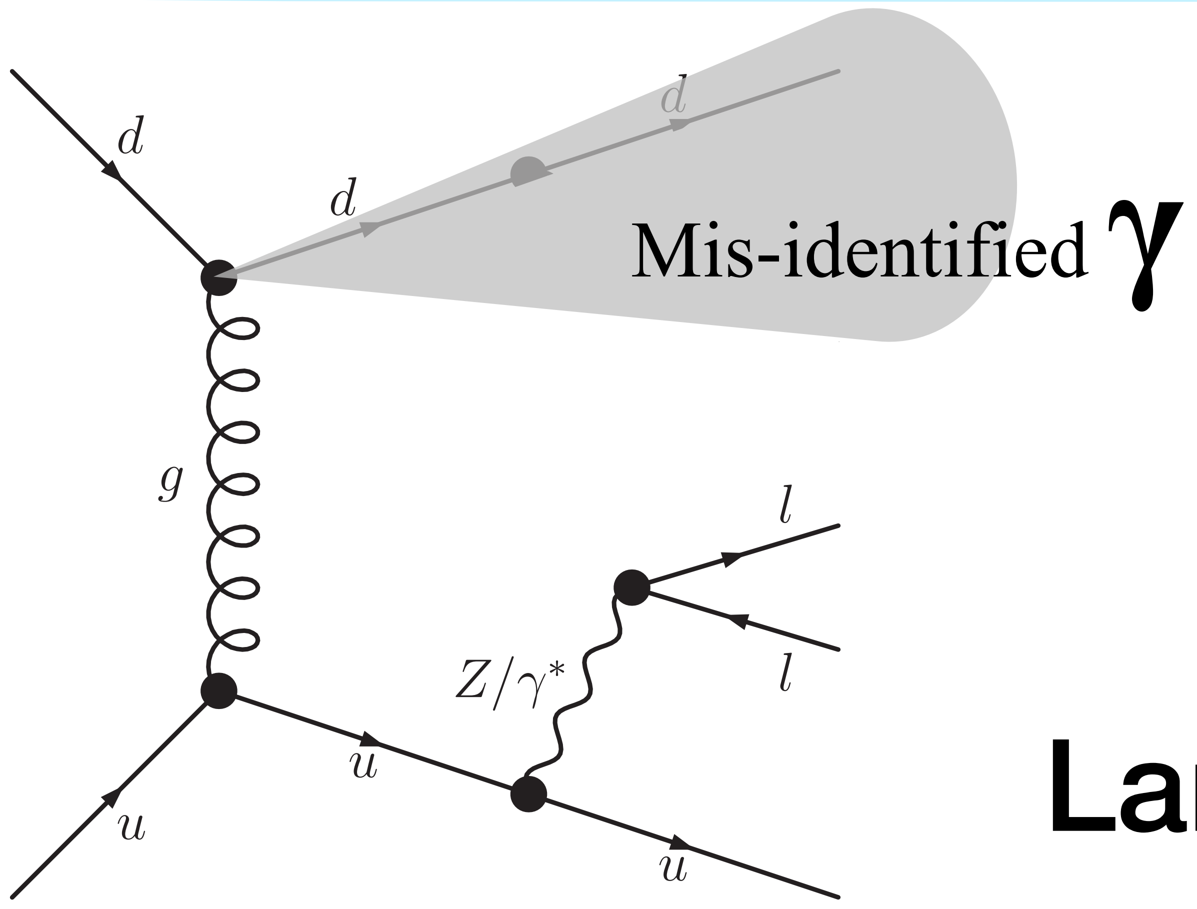
e/γ identification

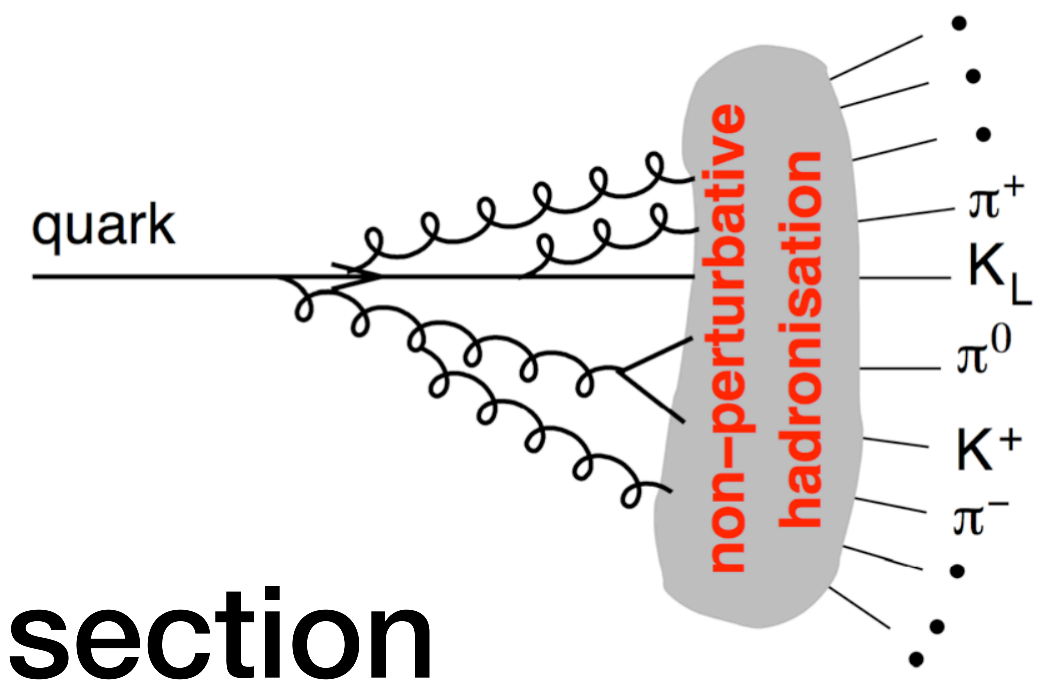
Identify an/a electron/photon → Select a real/prompt electron/photon

The fact is that **majority** of "electrons" and "photons" reconstructed in the data **are actually jets**

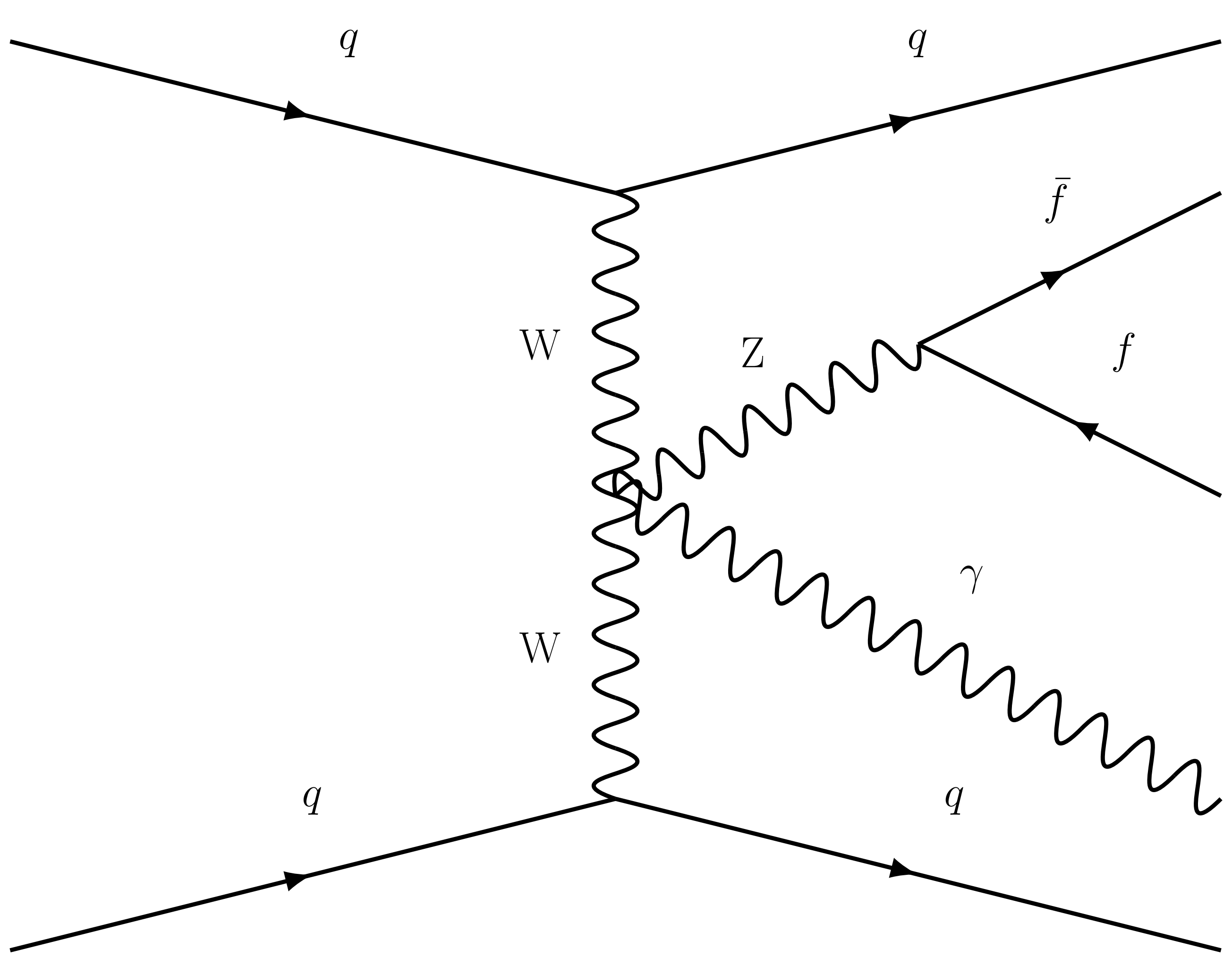
Background →

They are really from the Z/W/τ decay
From initial state radiation
....





Large cross section



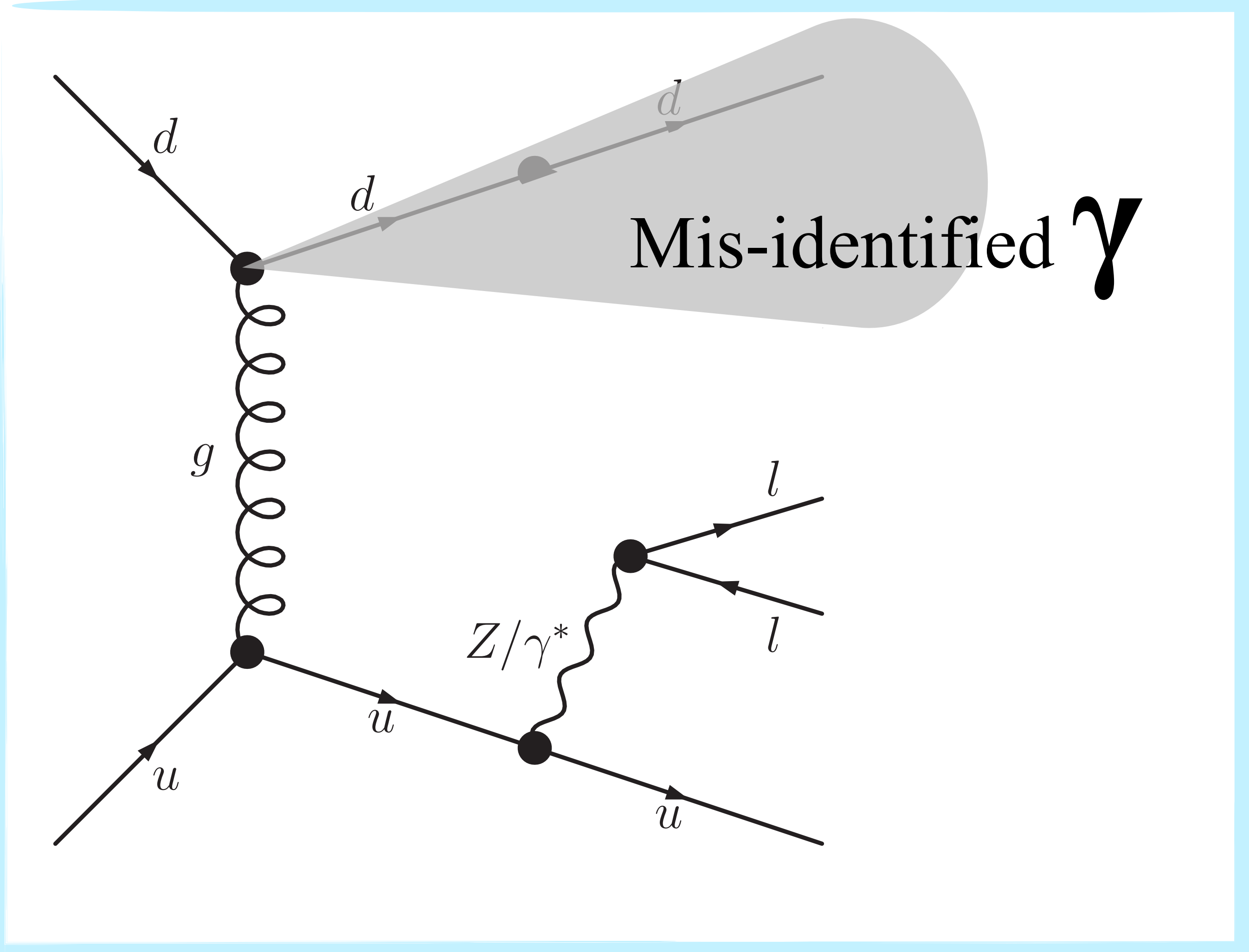
e/γ identification

Identify an/a electron/photon → Select a real/prompt electron/photon

The fact is that **majority** of "electrons" and "photons" reconstructed in the data **are actually jets**

Background →

They are really from the Z/W/τ decay
From initial state radiation
....

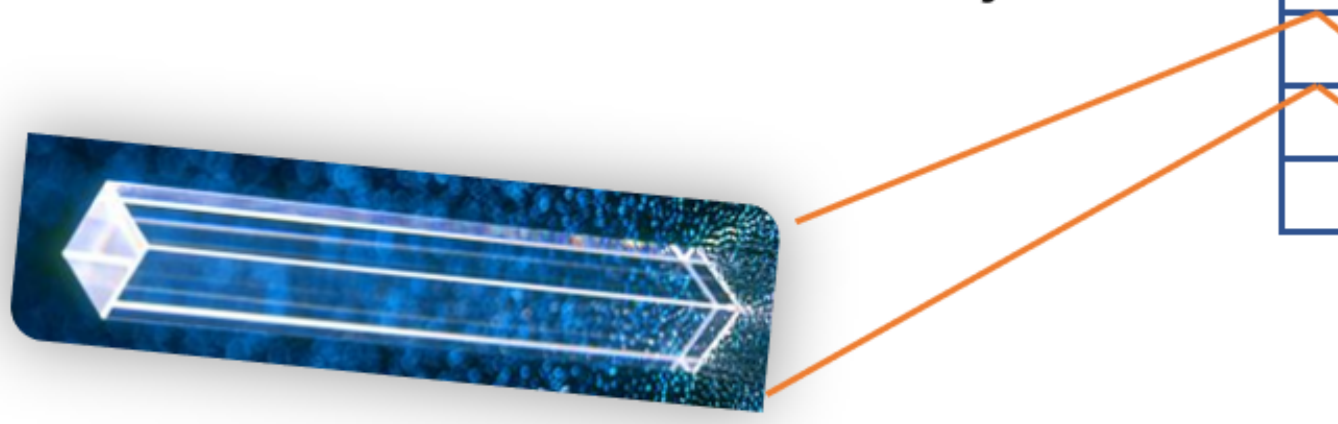


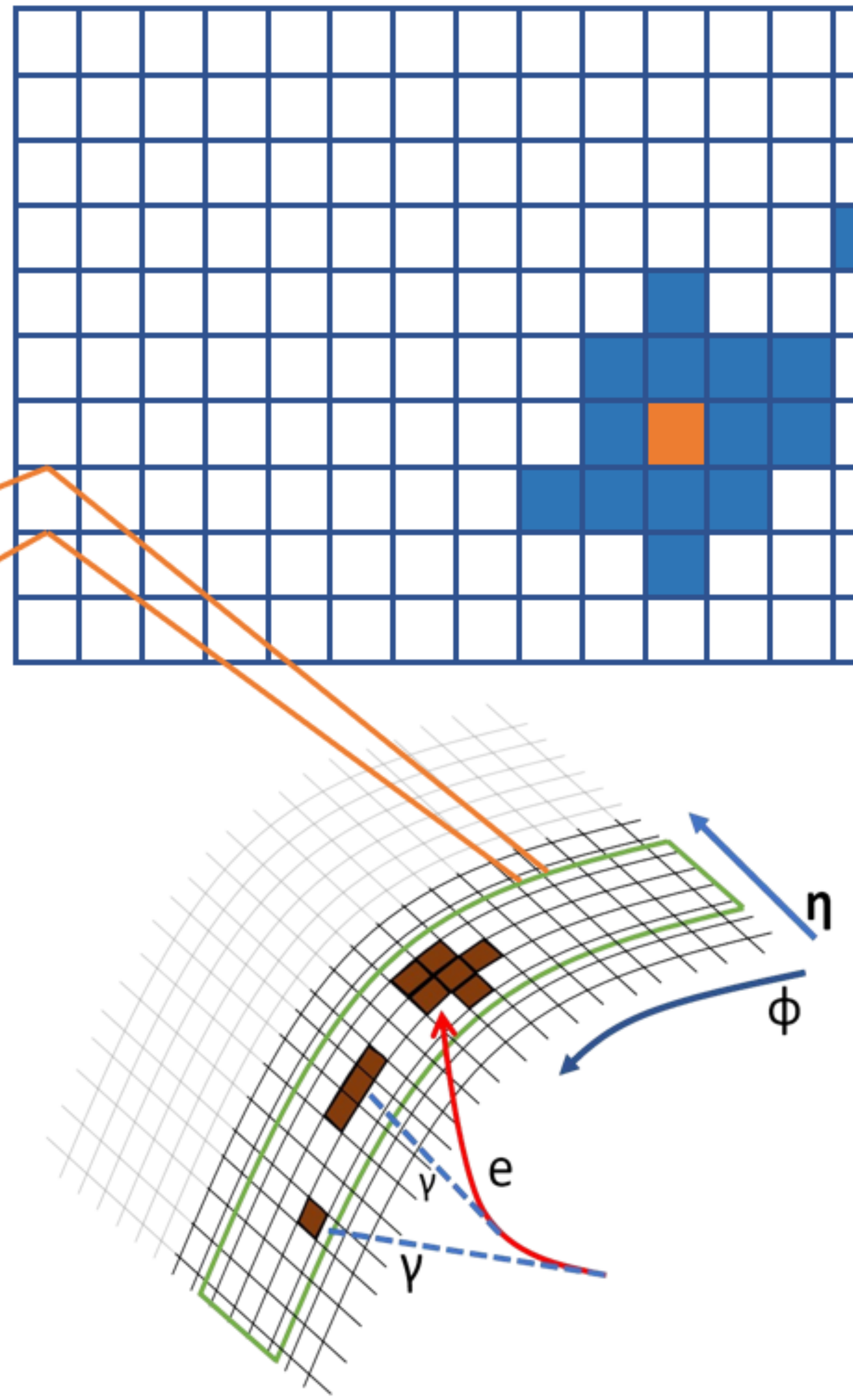
To remove this jet fakes background we apply **a selection of key quantities** that **distinguish electrons & photons from hadronic jets**

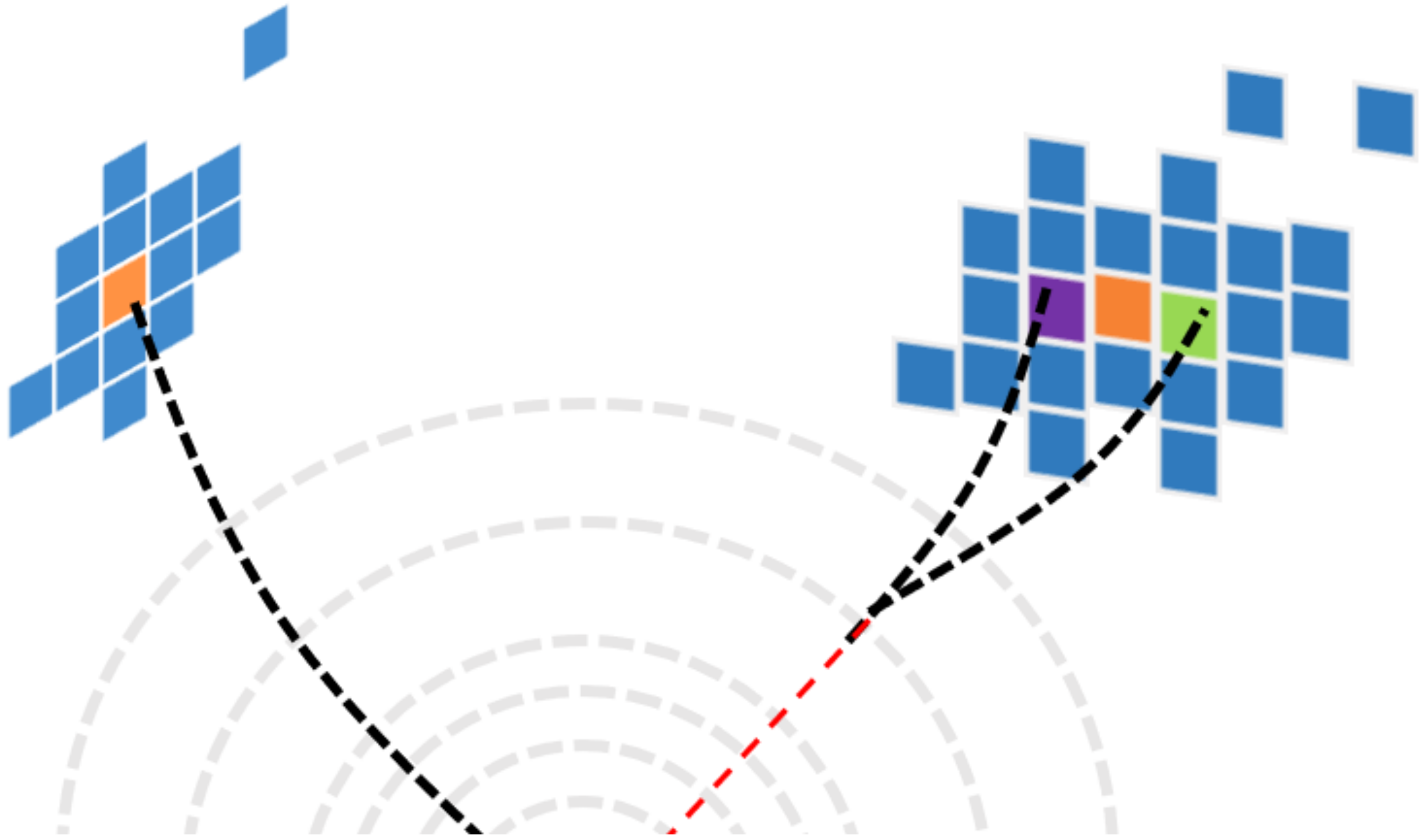
e/γ identification

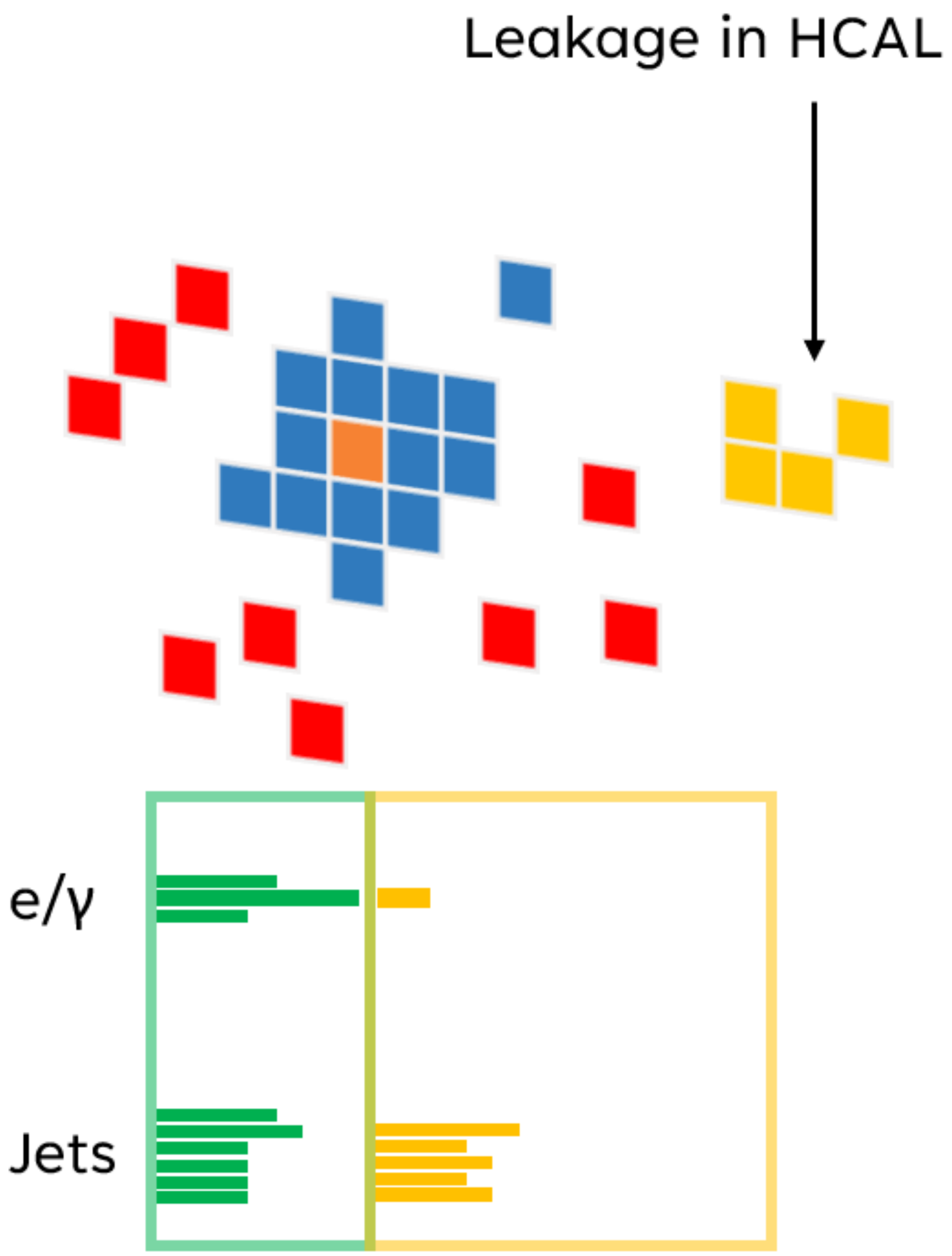
What can we use to identify electrons and photons?

- Description of the EM shower shape
- Tracking and clustering matching parameters
- Quantification of isolation of these objects









Exercise 3

Open the exercise-3.ipynb

Tag & Probe method

Tag & Probe

A final prompt selected e/γ used in the analysis needs to pass:

- reconstruction algorithm → *for data and MC independently*
- energy regression → based on MC only, needs residual correction
- residual correction → simultaneous adjustment for data and MC
- identification → *for data and MC independently*
- HLT → *for data and MC independently*

Tag & Probe

A final prompt selected e/γ used in the analysis needs to pass the

- reconstruction algorithm → *for data and MC independently*
- energy regression → based on MC only, needs residual correction
- residual correction → simultaneous adjustment for data and MC
- identification → *for data and MC independently*
- HLT → *for data and MC independently*

Tag & Probe

A final prompt selected e/γ used in the analysis needs to pass the

- reconstruction algorithm → *for data and MC independently*
- energy regression → based on MC only, needs residual correction
- residual correction → simultaneous adjustment for data and MC
- identification → *for data and MC independently*
- HLT → *for data and MC independently*

Might cause different performance (efficiency) in data and MC. Needs to correct MC to match data

Tag & Probe

- The efficiencies of data and MC are

$$\epsilon_{\mathrm{MC/data}} = \frac{N_{\mathrm{pass}}}{N_{\mathrm{pass}} + N_{\mathrm{fail}}}$$

- The scale factors are then derived from the ratio

$$\mathrm{SF} = \epsilon_{\mathrm{data}}/\epsilon_{\mathrm{MC}}$$

- The scale factors should be applied to MC

Tag & Probe

$$\epsilon_{\mathrm{MC/data}} = \frac{N_{\mathrm{pass}}}{N_{\mathrm{pass}} + N_{\mathrm{fail}}}$$

- The first step is to have an electron collection
- In order to make it sure that the electron is prompt, we usually select it from a Z candidate
  - One electron is with such tight selection called “Tag” electron
  - One electron is with loose selection called **“Probe” electron**
  - The invariant mass of the tag and probe electrons are required

Tag & Probe

$$\epsilon_{\text{MC/data}} = \frac{N_{\text{pass}}}{N_{\text{pass}} + N_{\text{fail}}}$$

- The first step is to have an electron collection
- In order to make it sure that the electron is prompt, we usually select it from a Z candidate
  - One electron is with such tight requirements called “Tag” electron
  - One electron is with loose requirements called **“Probe” electron**
  - The invariant mass of the tag and probe electrons are required

The probe electron collection is prompt and unbiased, as it’s from the Z candidates and with loose requirement.

Tag & Probe

For example, we want to measure the efficiency of 'cut-based-medium' electron ID or HLT_Ele32_WPTight_Gsf

Tag cuts: $p_T > 30$ GeV, $|\eta|<2.5$
Pass HLT
cutBasedElectronID-Fall17-94X-V2-medium

Mass cut: 50<pair mass<130, opposite charge

$$\epsilon_{\text{MC/data}} = \frac{N_{\text{passing probes}}}{N_{\text{probes}}}$$

Probe cuts: $p_T > 25$ GeV, $|\eta|<2.5$

Passing Probes: cutBasedElectronID-Fall17-94X-V2-medium

Probe cuts: $p_T > 25$ GeV, $|\eta|<2.5$
cutBasedElectronID-Fall17-94X-V2-medium

Passing Probes: HLT_Ele23_WPTight_Gsf

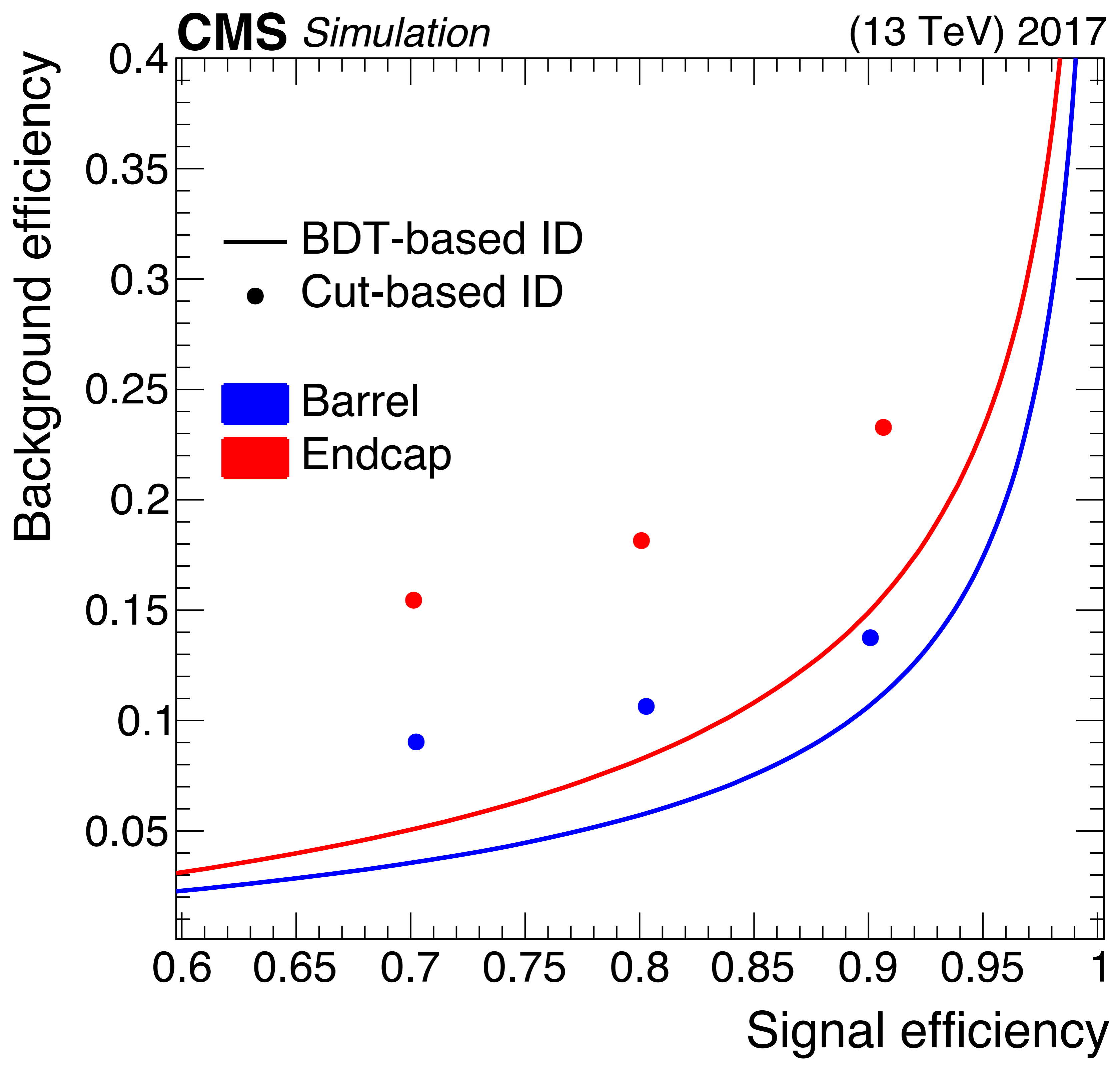
e/γ MVA identification

e/γ MVA identification

- Identify prompt electron/photon by machine learning technique
- Better signal efficiency

Analyst could train their own MVA Identification according to the demand and feature

More in exercise-5.ipynb



Backup

e/γ reconstruction — track seeding

Tracker-Driven (aka inside-out)

Find a track and then matches with PFClusters using particle flow techniques

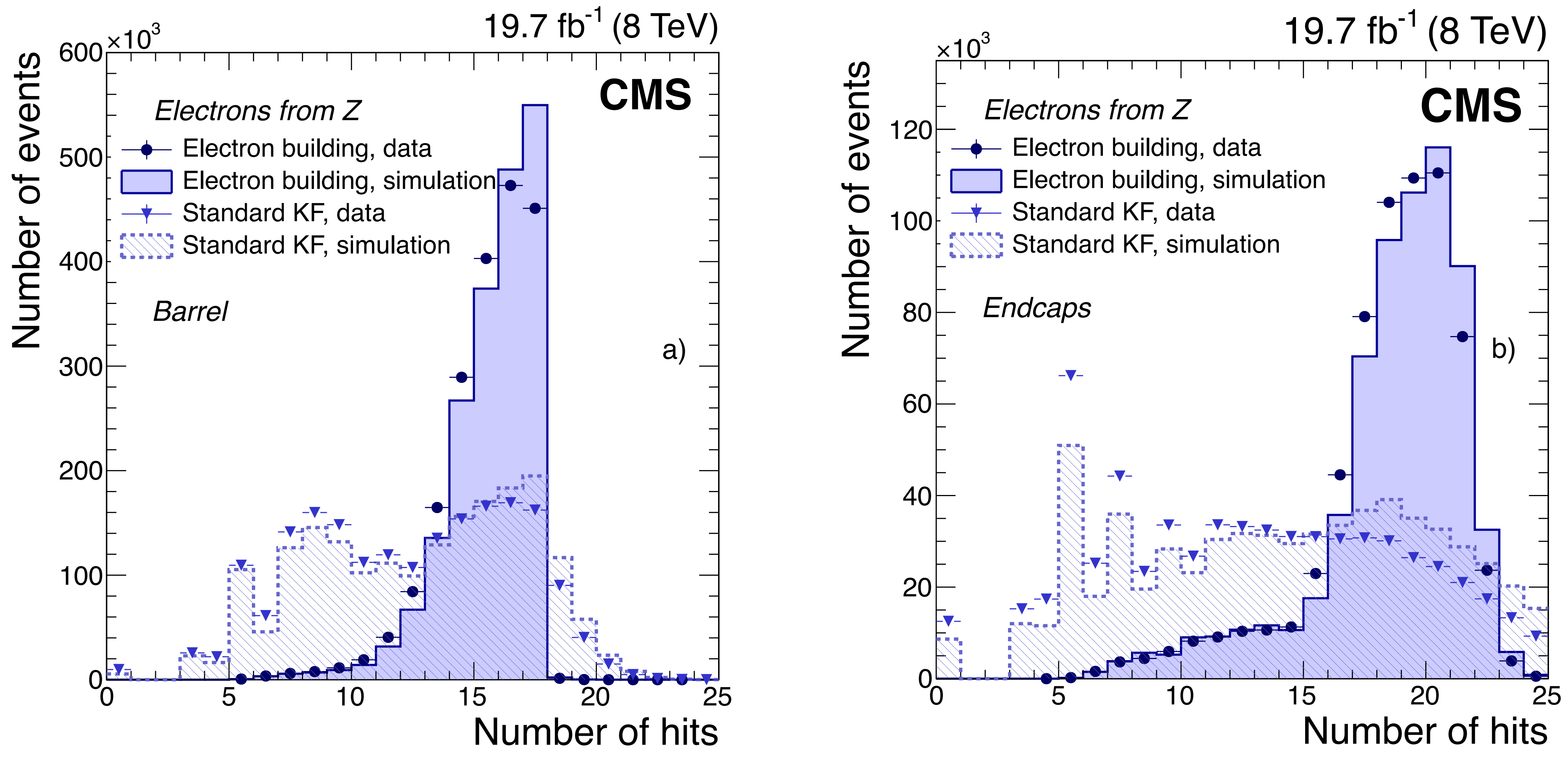
- KF algorithm collects hits if $r_{th} < E/p < 1$ ($r_{th} = 0.65$ or 0.75 for $2 < p_T < 6$ GeV or $p_T \geq 6$ GeV)
- Refit the KF tracks with a small number of hits or a large χ^2_{KF} by GSF
- N_{hits}, the χ^2_{KF}, the χ^2_{GSF}, and the matching of the ECAL and tracker in geometrical and energy are used in a multivariate (MVA) method

Complement at low p_T

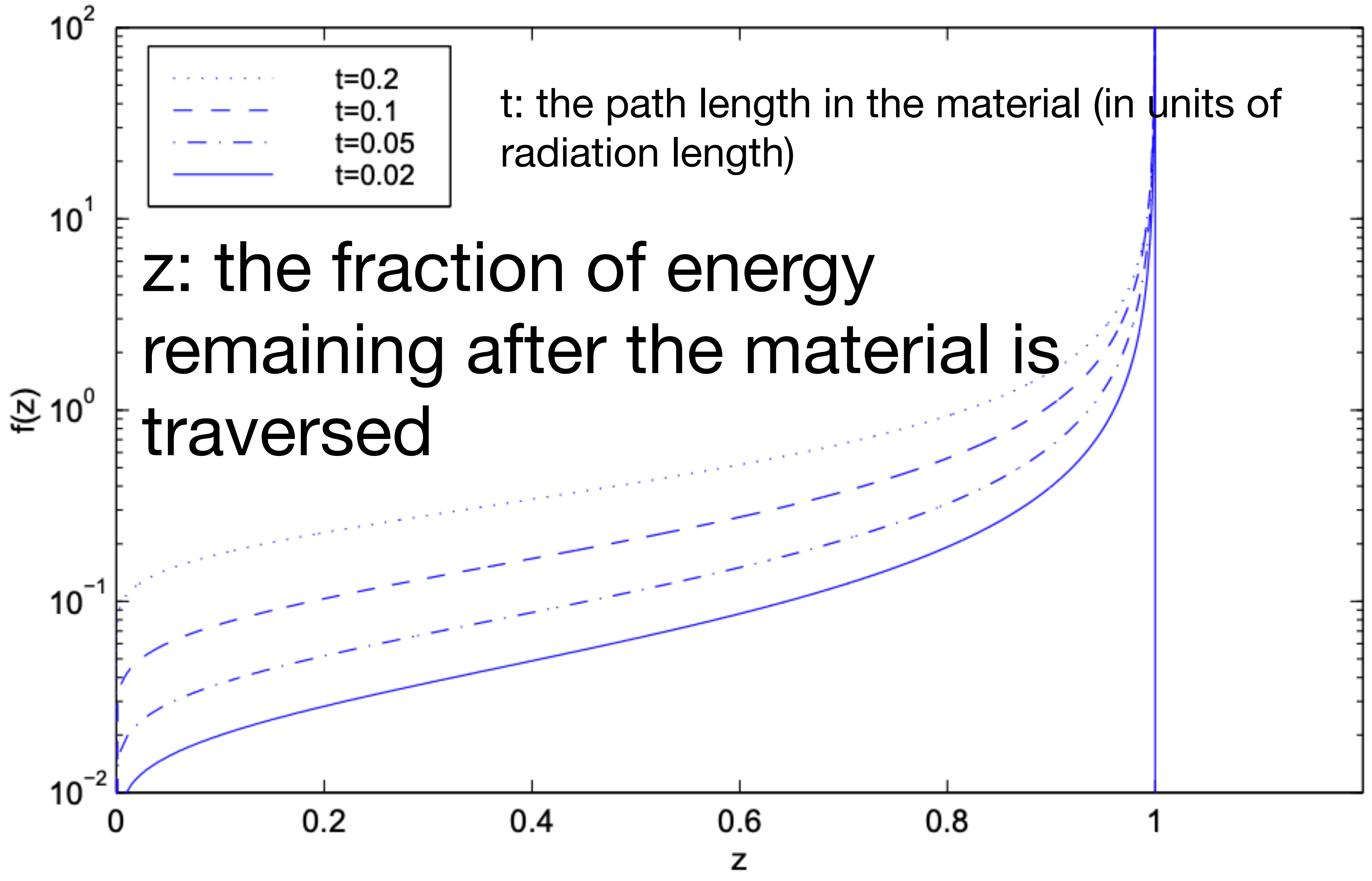
The overall efficiency by these two methods are more than 95%

e/γ reconstruction — track reconstruction

- A minimum of five hits is finally required to create a track
- If many compatible hits are found on a layer, with a limit of five candidate trajectories are grown in parallel
- At most, one missing hit is allowed for an accepted trajectory candidate with increased χ^2 penalty applied



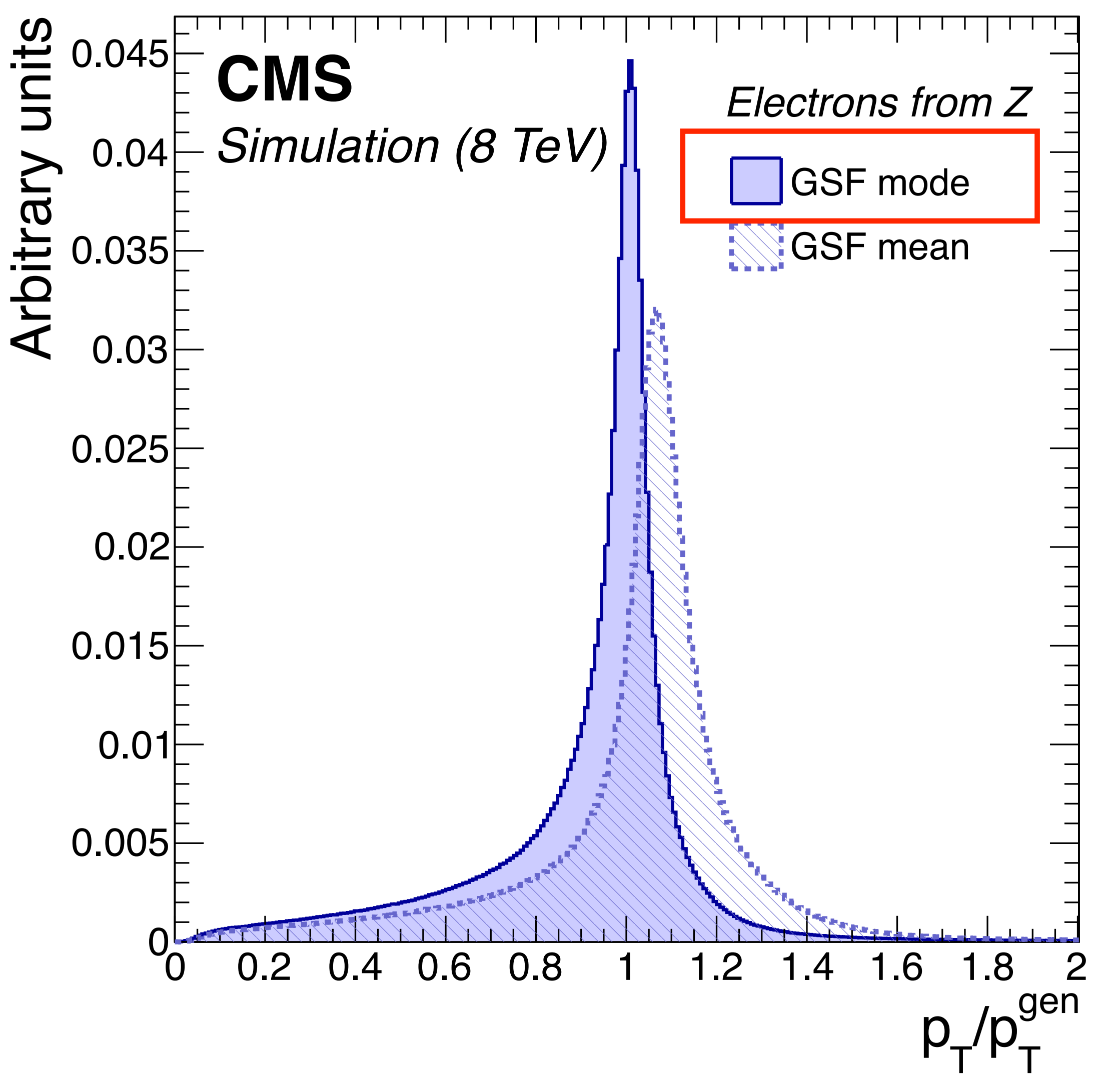
arxiv:1502.02701 (EGM-13-001)

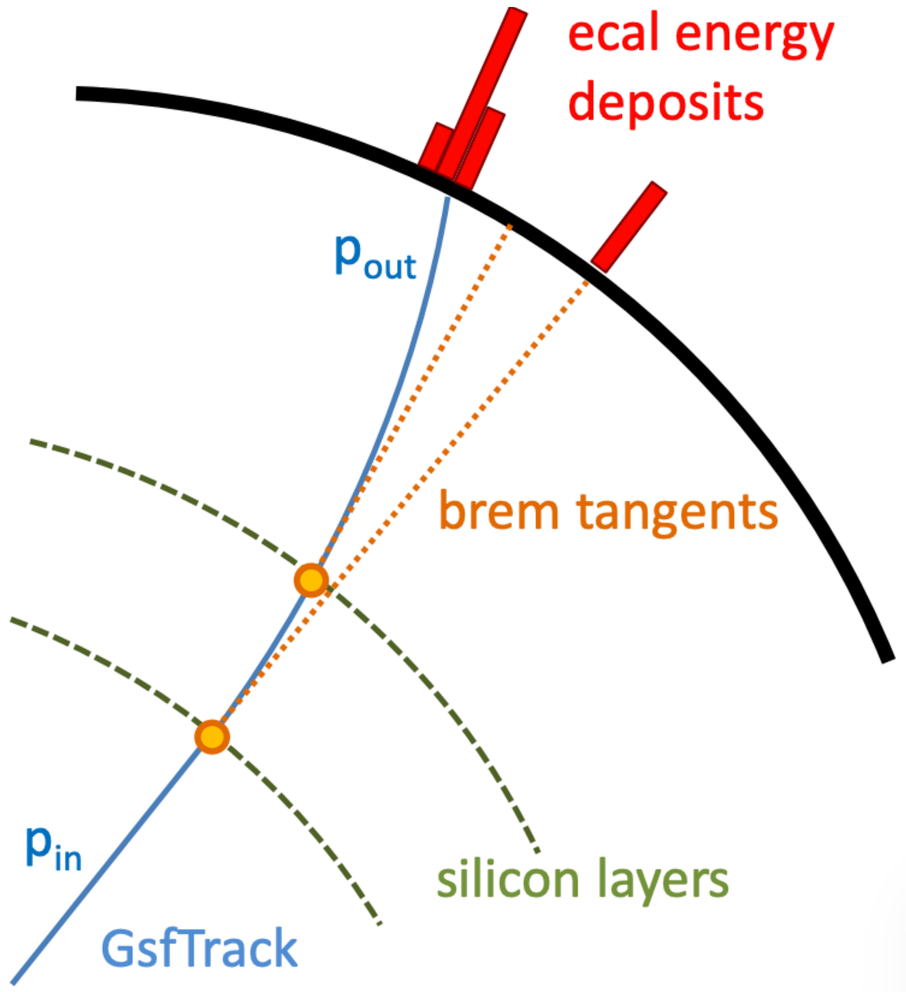


- Search for the compatible hits on the next silicon layers and apply the GSF fit
- The electron energy loss is modelled through a **Bethe–Heitler function**
- The energy loss in each layer is approximated by a **mixture of Gaussian distributions (GSF)**
- A weight is attributed to each Gaussian distribution that describes the associated probability

e/γ reconstruction — track reconstruction

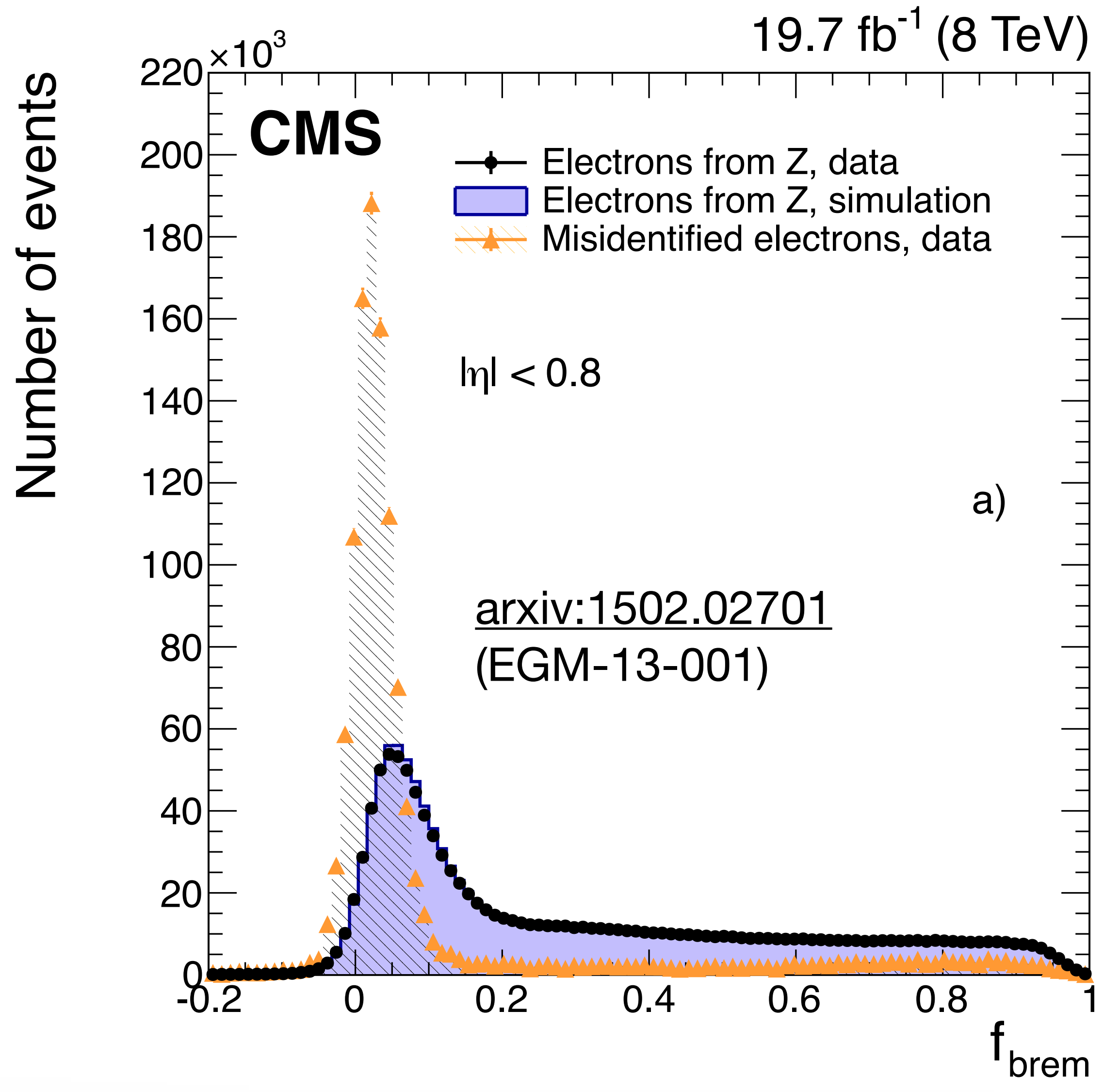
arxiv:1502.02701 (EGM-13-001)

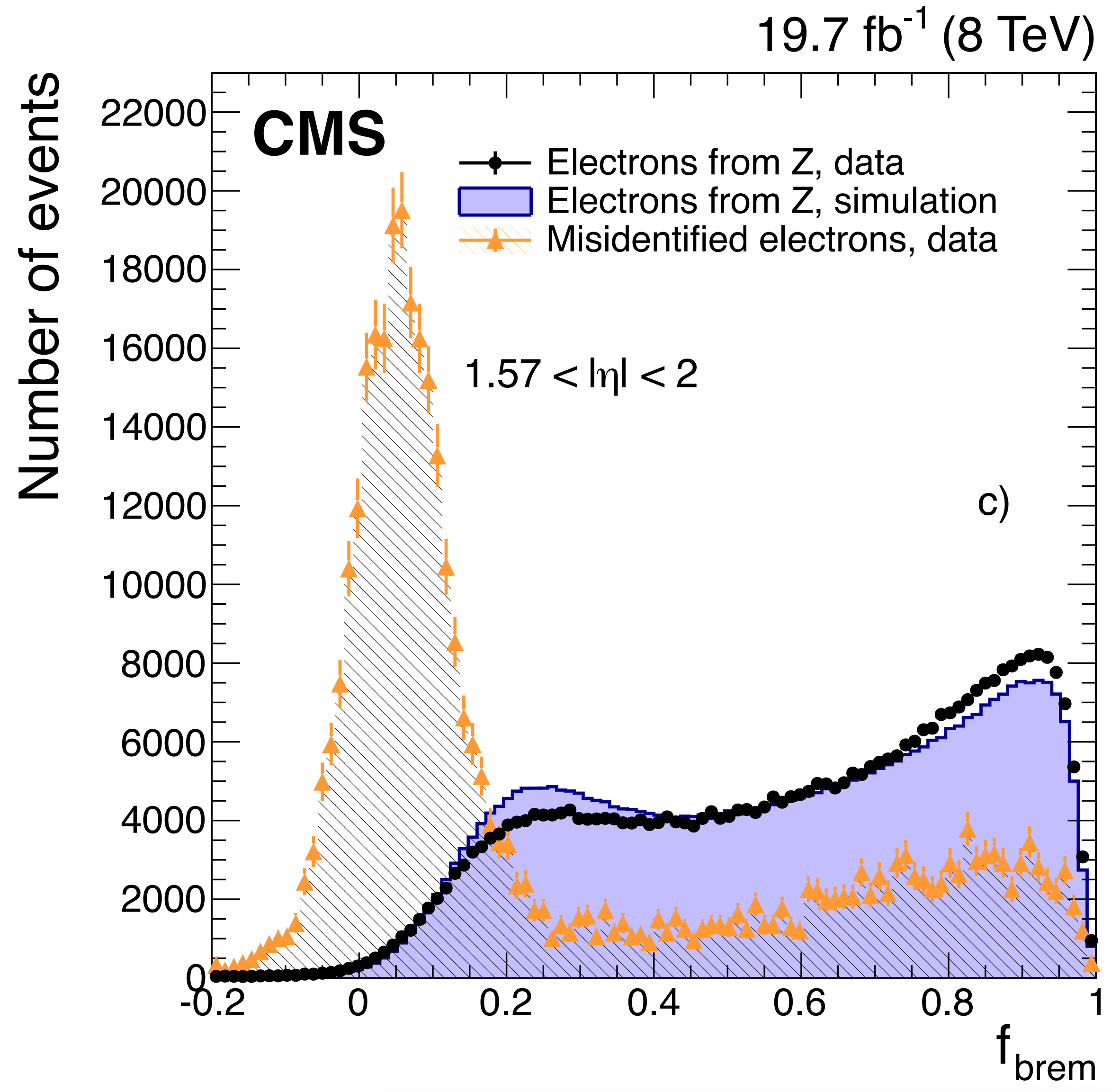




$$f_{brem} = (p_{in} - p_{out})/p_{in}$$

- Extract track parameters from the GSF track
  1. Take the weighted mean of all the components, given the track state on each layer
  2. Take only **the most probable value (mode) of the probability distribution function (PDF)**, giving more importance to the highest weight component





Electron candidates finally defined by loose track–superclusters matching criteria

- $|\Delta\eta|<0.02$, $|\Delta\phi|<0.15$