GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

III. Physikalisches Institut A

RWTH AACHEN UNIVERSITY

Electron and photon performance in CMS in Run2 and prospects for Run3

Swagata Mukherjee (RWTH Aachen University, Germany)

on behalf of the CMS Collaboration

PLAN OF THIS TALK

Main steps

- Reconstruction
- Identification
- Energy correction

To be performed at

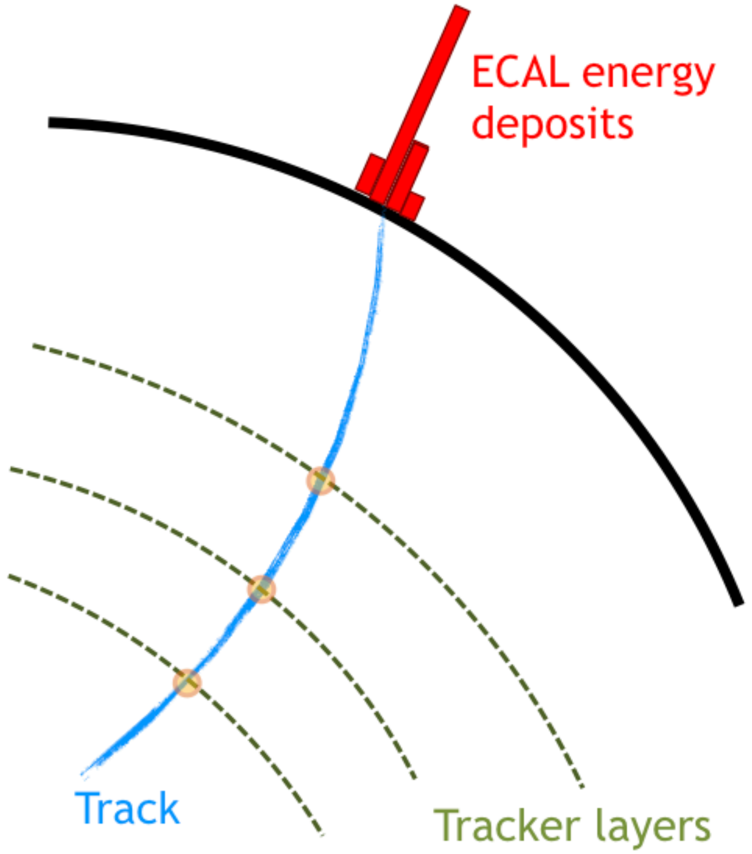
- Online (trigger)
- Offline

- ✓ This talk will discuss reconstruction, identification, energy corrections in
  - ✓ **Offline,** which is used in physics analyses
  - ✓ Also briefly touch upon **Online**, i.e. in software-based high-level trigger(HLT). Hardware-based L1 trigger will not be discussed in this talk
- ✓ Reconstruction algorithms at HLT and offline are similar
  - ✓ HLT is optimised for speed, thus **simpler algorithms** wherever possible
- ✓ Identification variables used to reject background(fake e/ɣ objects) are similar at HLT and offline
- ✓ Energy correction at HLT preliminary, at offline it is much more precise

I'll mostly focus on electrons, but photons are, in most cases, very similar

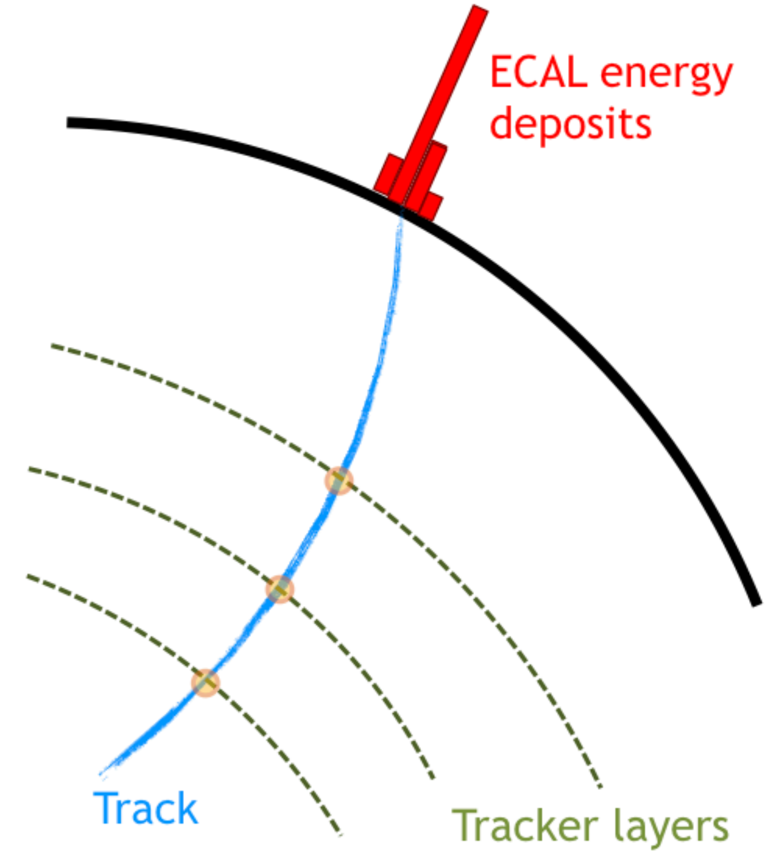
SIGNATURE OF ELECTRON IN CMS

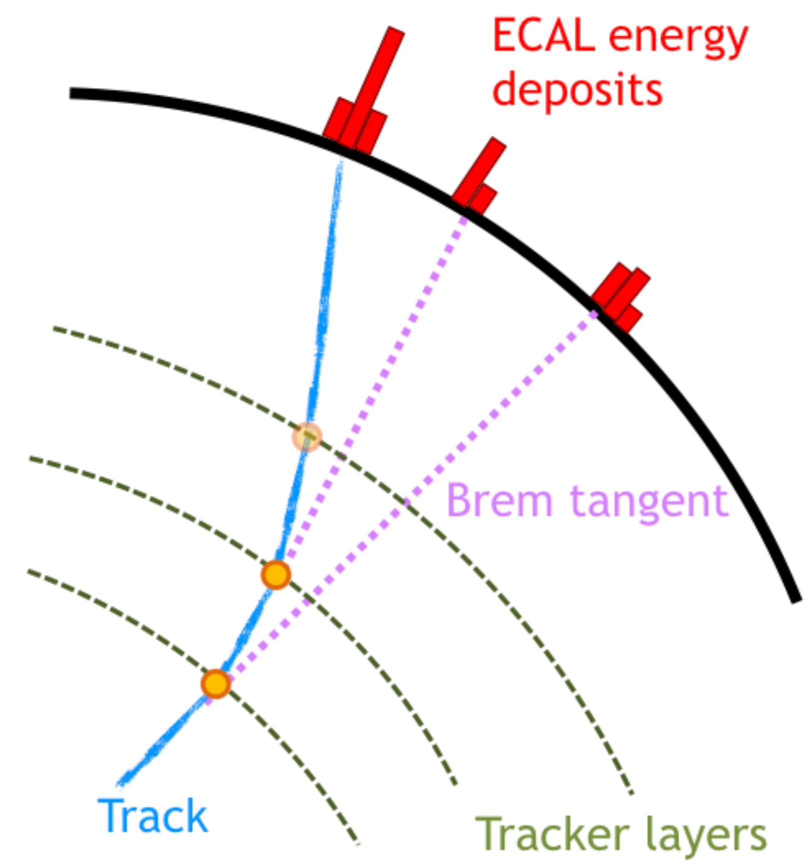
✓ A track in tracker + energy deposit in electromagnetic calorimeter(ECAL)



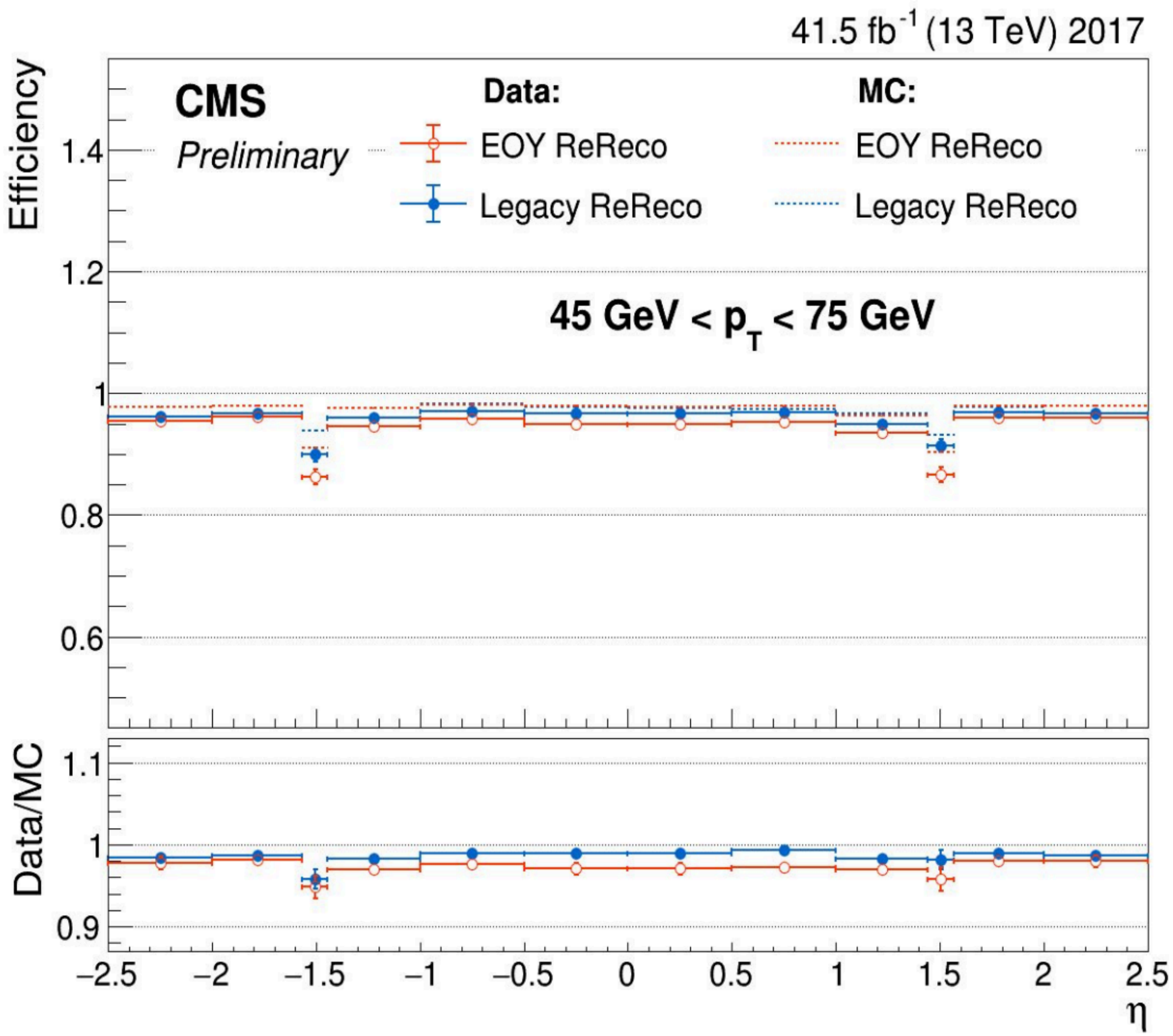
SIGNATURE OF ELECTRON IN CMS

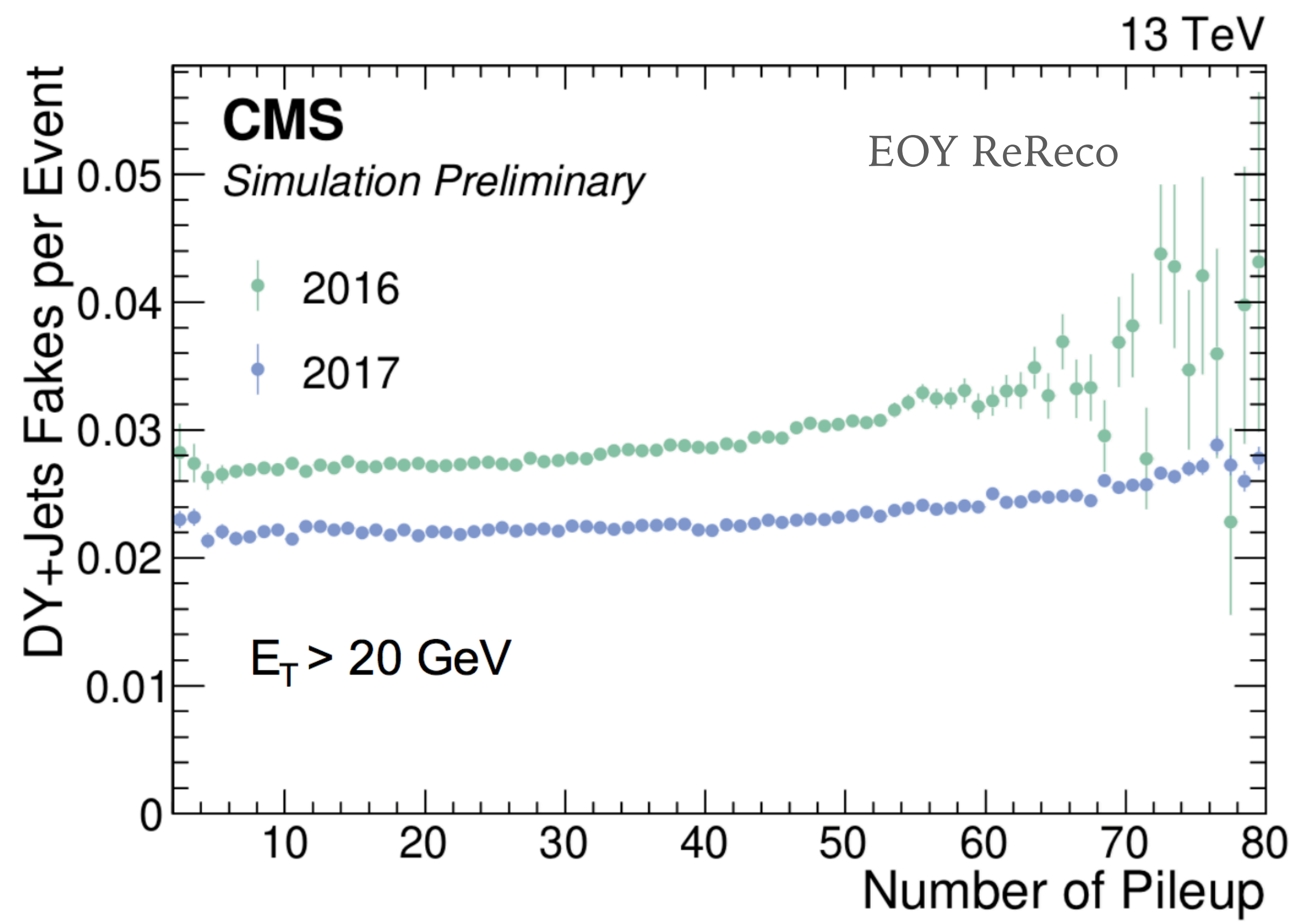
- A track in tracker + energy deposit in electromagnetic calorimeter(ECAL)
- Not so simple always
- Electron can brem in tracker
  - Multiple energy deposits in ECAL
  - Draw brem tangents to find them
  - Add their energy to reconstruct full energy of electron
- Special tracking for electrons to take into account brems
  - Gaussian Sum Filter (**GSF**), while other objects use Kalman Filter (KF)
- Electron reconstruction algorithm takes into account all these complexities





ELECTRON RECONSTRUCTION EFFICIENCY AND FAKE RATE (OFFLINE)





Highly efficient (~95%) electron track reconstruction algorithm.

Improvement in **Legacy reprocessing**(improved calibration, better description of data conditions in simulation) w.r.t **End of the Year (EOY)** processing.

Lower fake rate in 2017 due to additional layer of inner tracker

Fake rate <3% even at high pile up(70-80)

https://cds.cern.ch/record/2718815/files/DP2020_024.pdf & https://cds.cern.ch/record/2320638/files/DP2018_017.pdf

VARIABLES USED FOR ELECTRON IDENTIFICATION

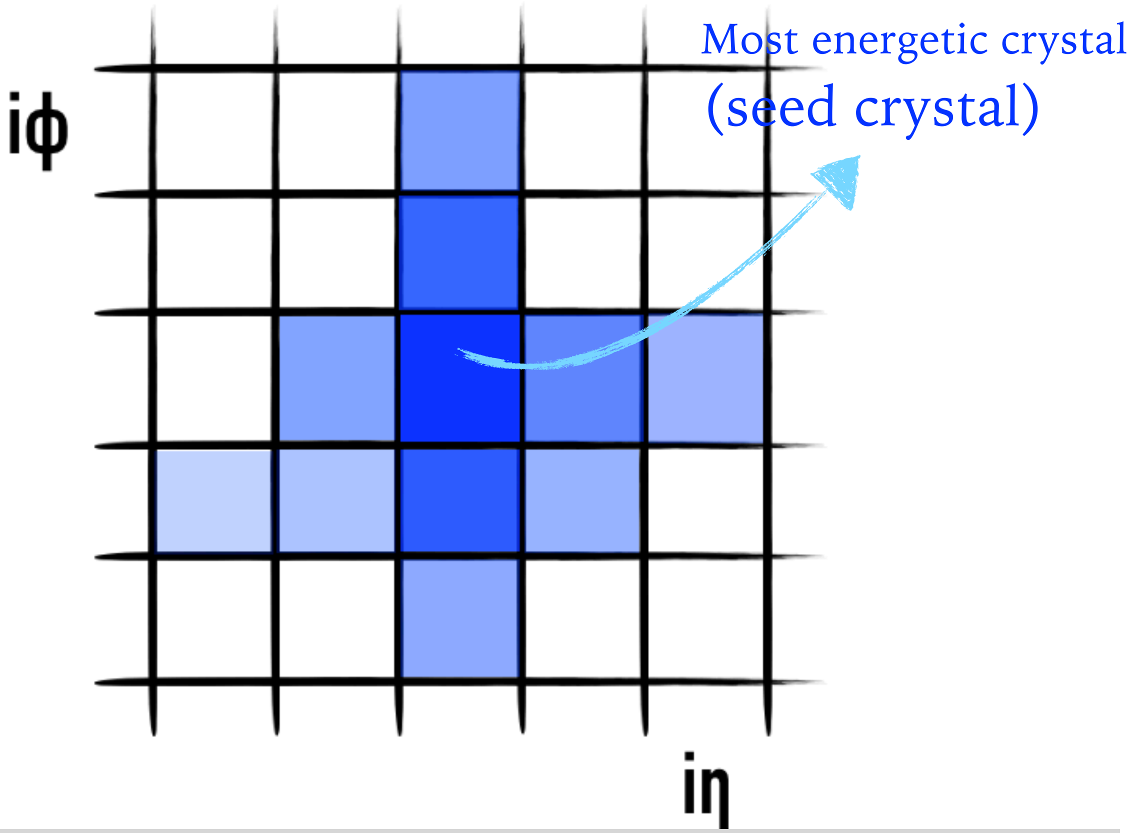
An ID variables should have good discrimination power between signal(real electron) and background(fake electron). Broadly, 3 types of ID variables:

Shower shape	How is the shape of the shower in ECAL? Is there any energy deposit in HCAL?
Isolation	Is there any activity around the electron object? In ECAL / HCAL / Tracker
Compatibility between track and ECAL energy deposit	Two independent position and energy measurements One from **tracker**, another from **ECAL** How compatible are they?

Similar identification variables used at HLT and offline

LATERAL SHOWER SHAPE σiηiη

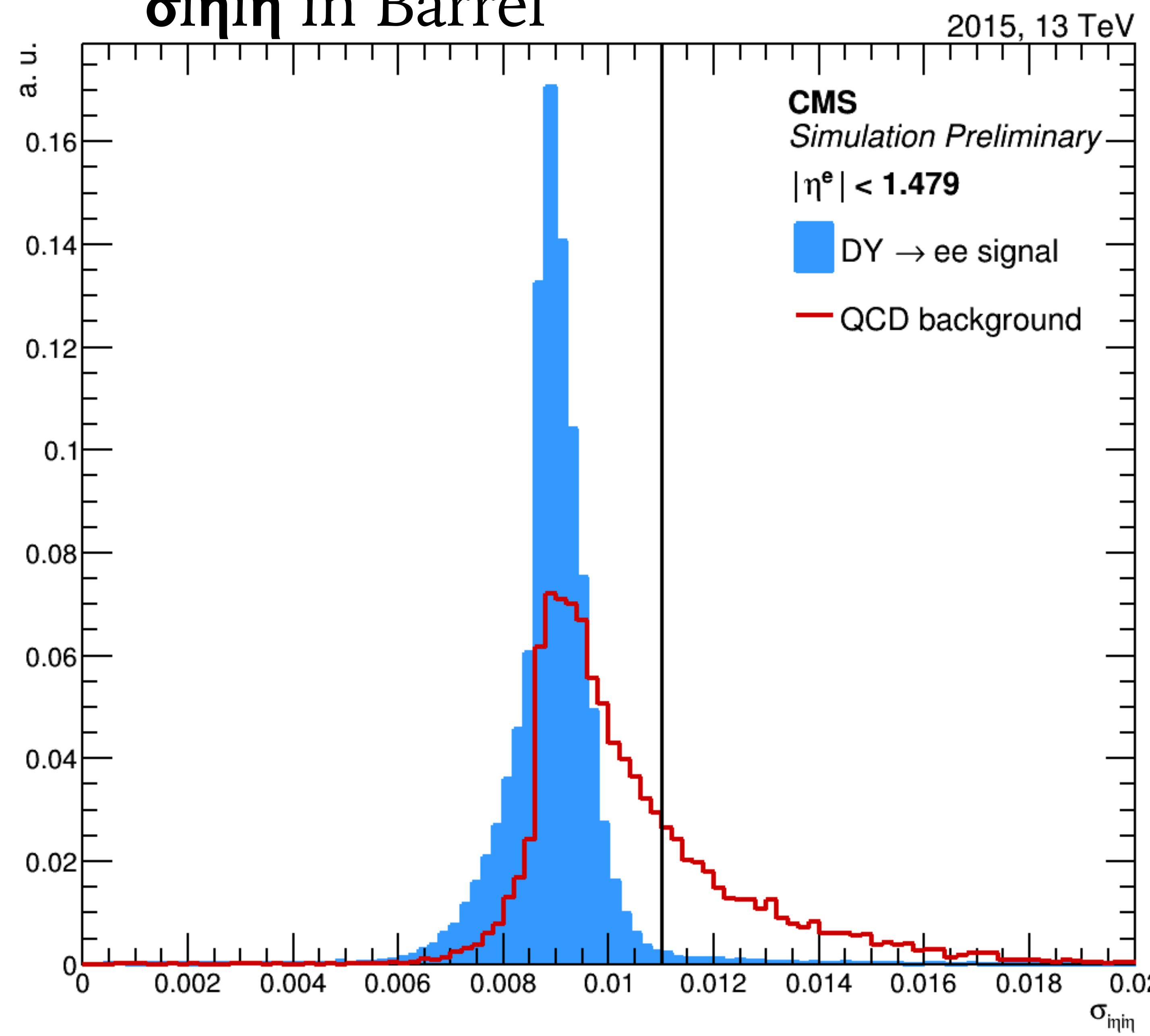
5x5 array of ECAL crystals around the seed



σiηiη measures spread of a EM shower along **η** direction

$$\sigma_{i\eta i\eta} = \sqrt{\left(\frac{\Sigma_i^{5\times5} w_i(\eta_i - \bar{\eta}_{5\times5})^2}{\Sigma_i^{5\times5} w_i}\right)}$$

σiηiη in Barrel



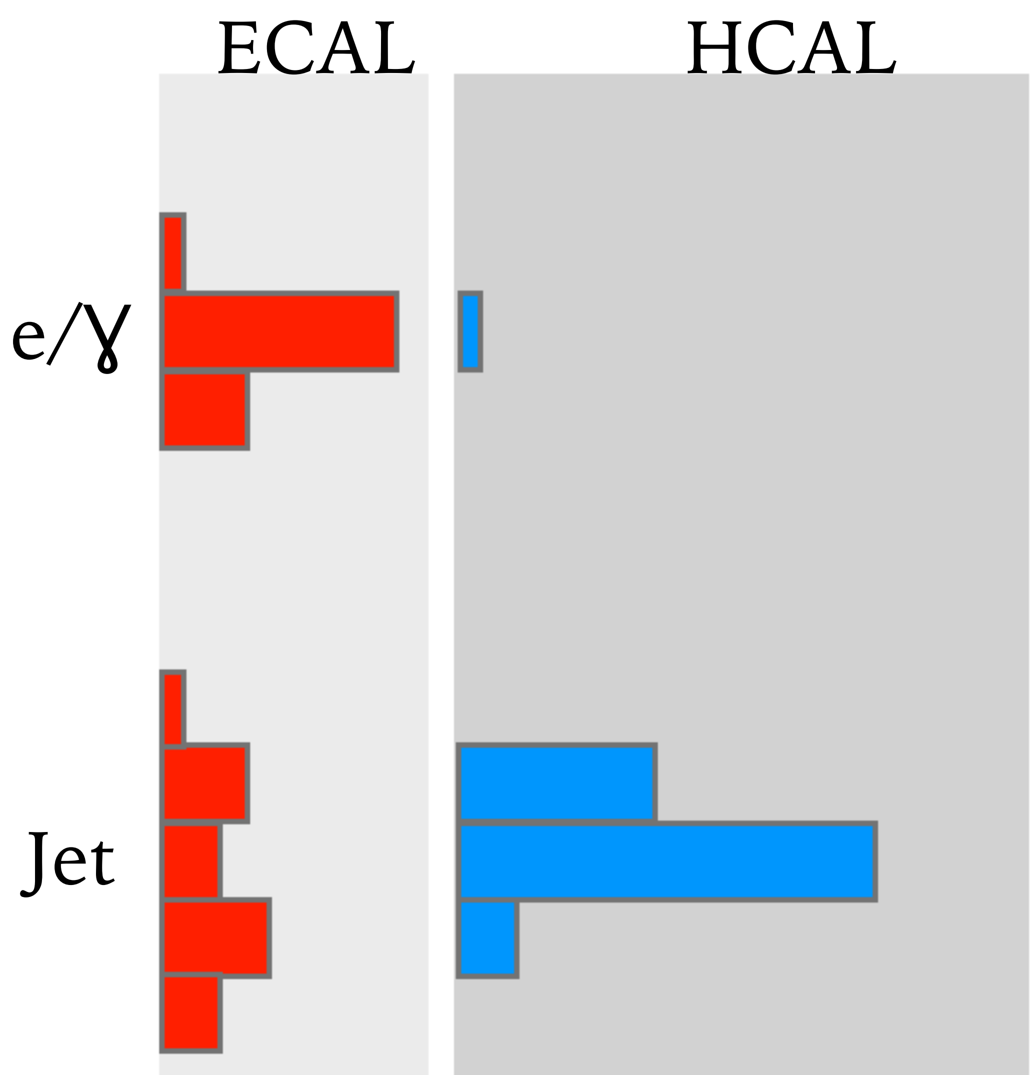
$w_i = \max(0, 4.7 + \ln(E_i/E_{5\times5}))$

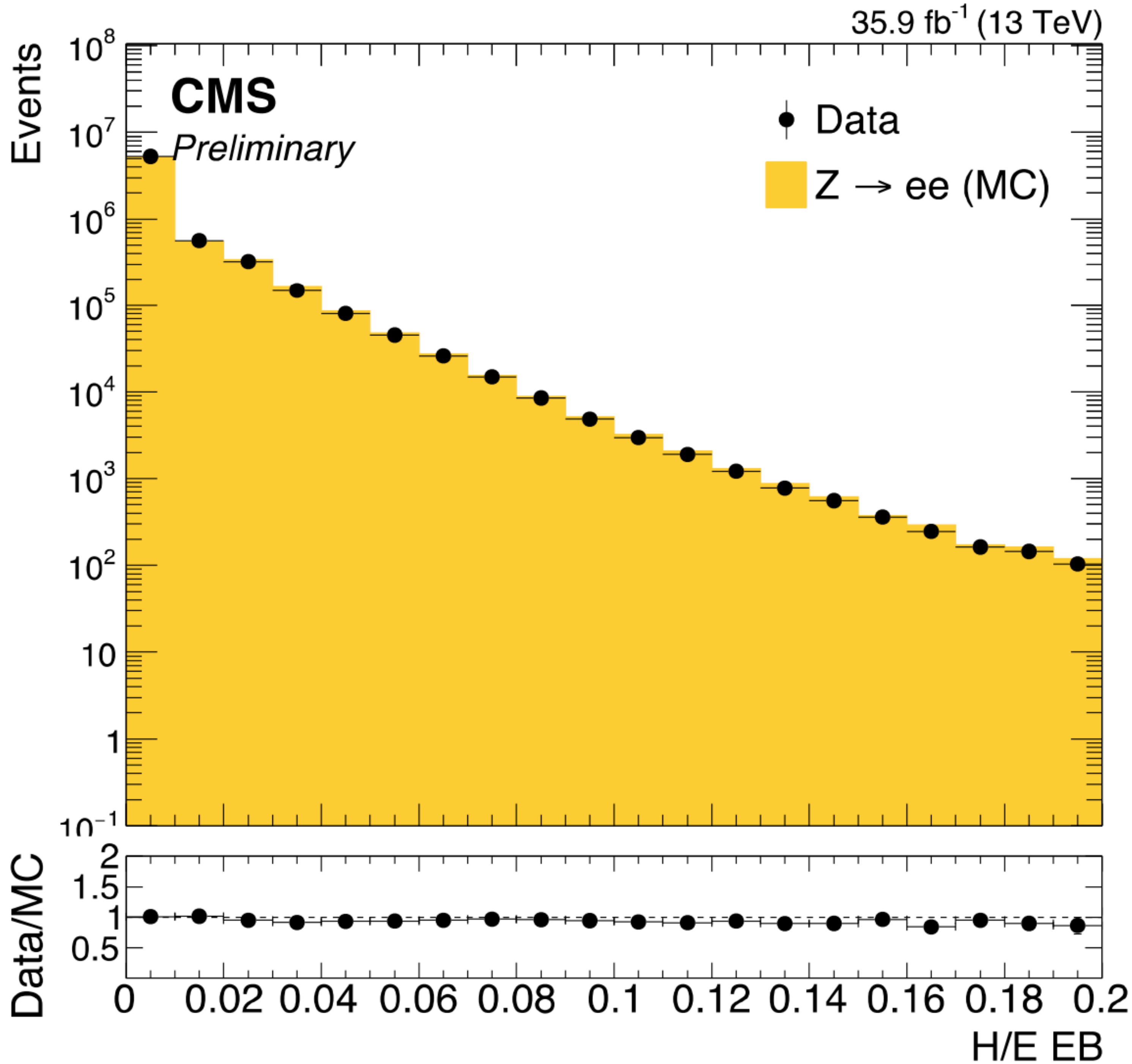
Which translates to a energy cut

$E_i > 0.9\%$ of $E_{5\times5}$

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

LONGITUDINAL SHOWER SHAPE H/E





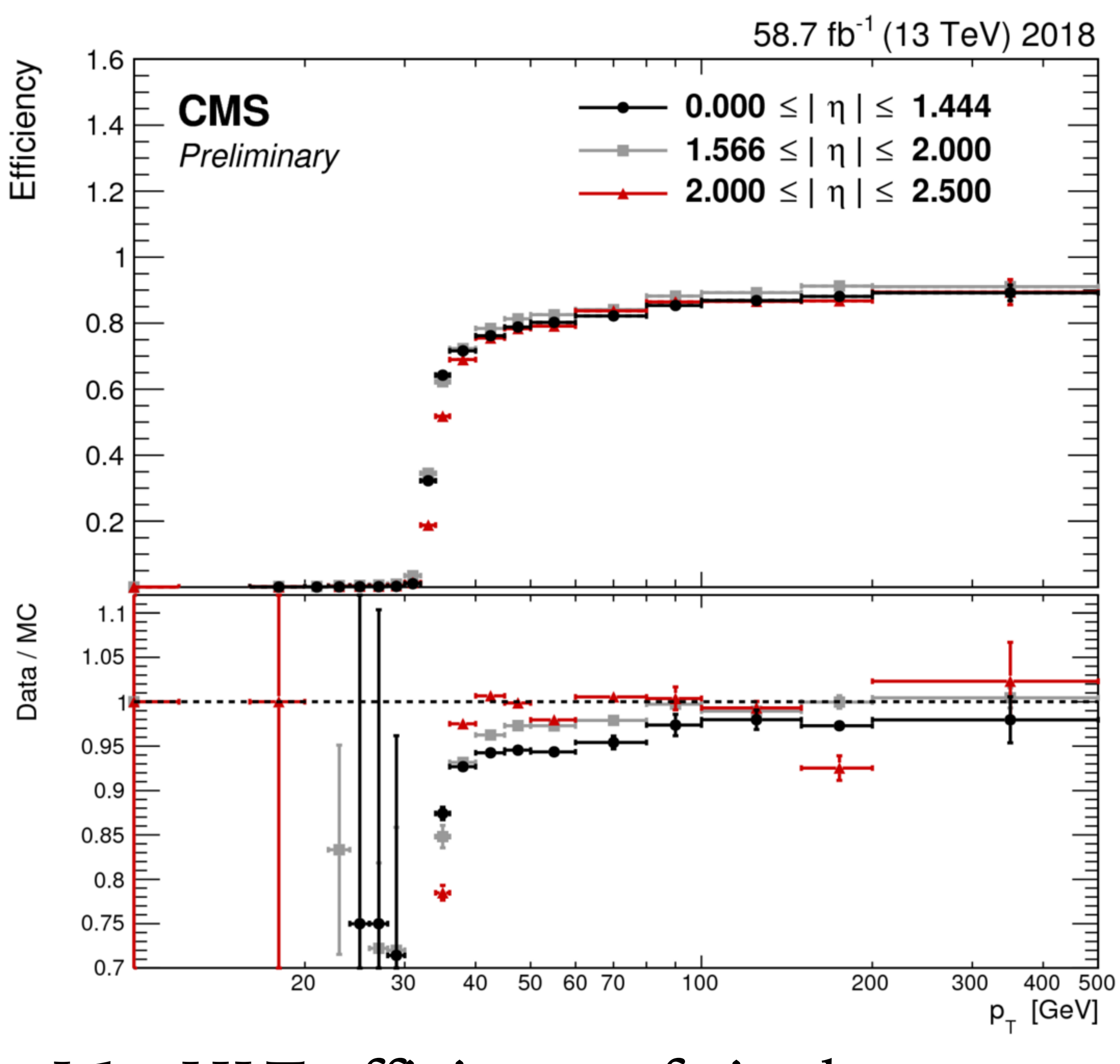
H/E

Excellent ID variable used in electron and photon ID

H/E is very well modelled in simulation

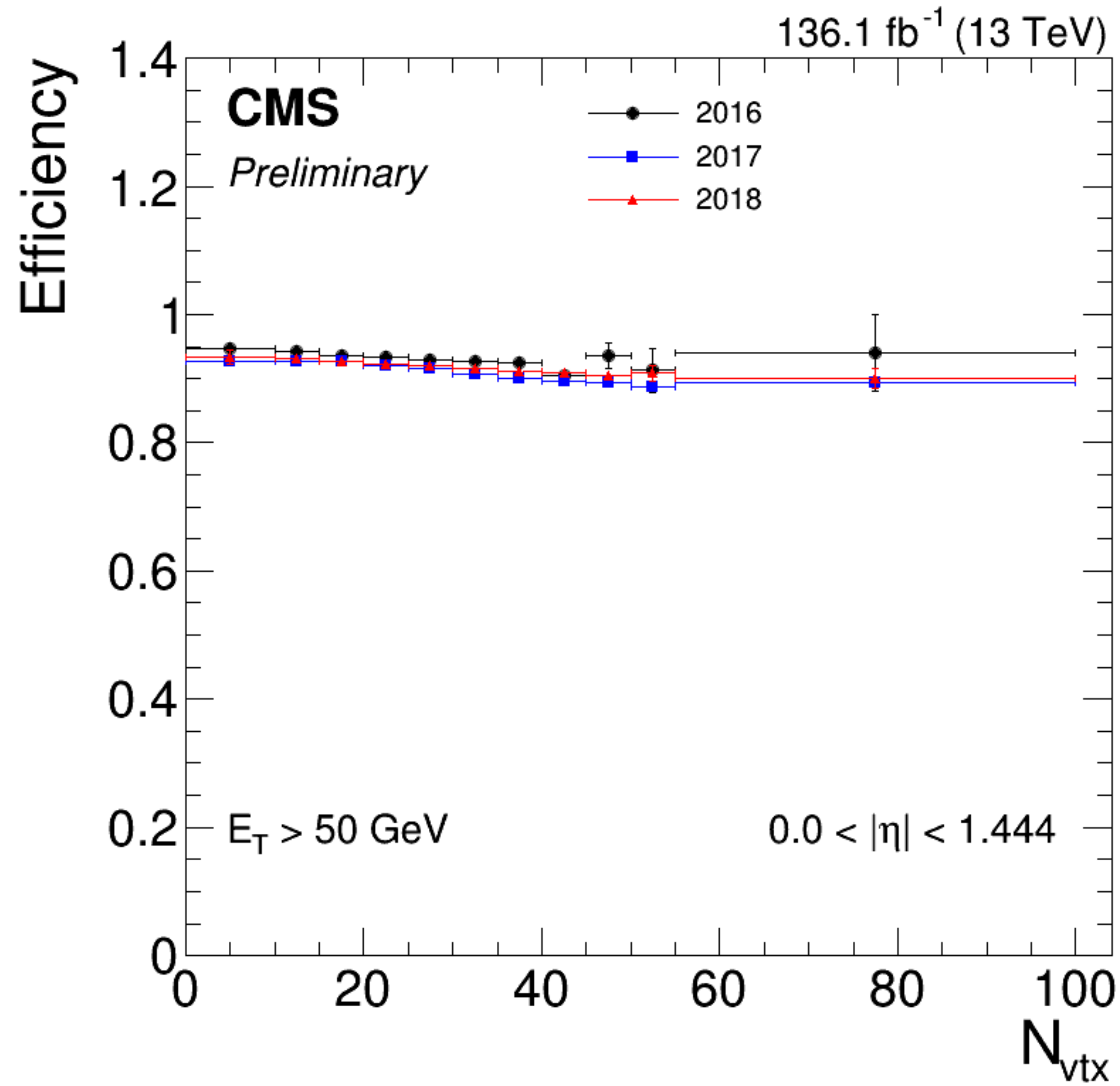
https://cds.cern.ch/record/2255497/files/DP2017_004.pdf

ELECTRON TRIGGER EFFICIENCY



L1+HLT efficiency of single electron trigger as a function of pT of offline electron

80-90% efficient depending on pT



Double electron trigger efficiency, stable within 5% as a function of pile up

https://twiki.cern.ch/twiki/bin/view/CMSPublic/EgammaFullRun2Data

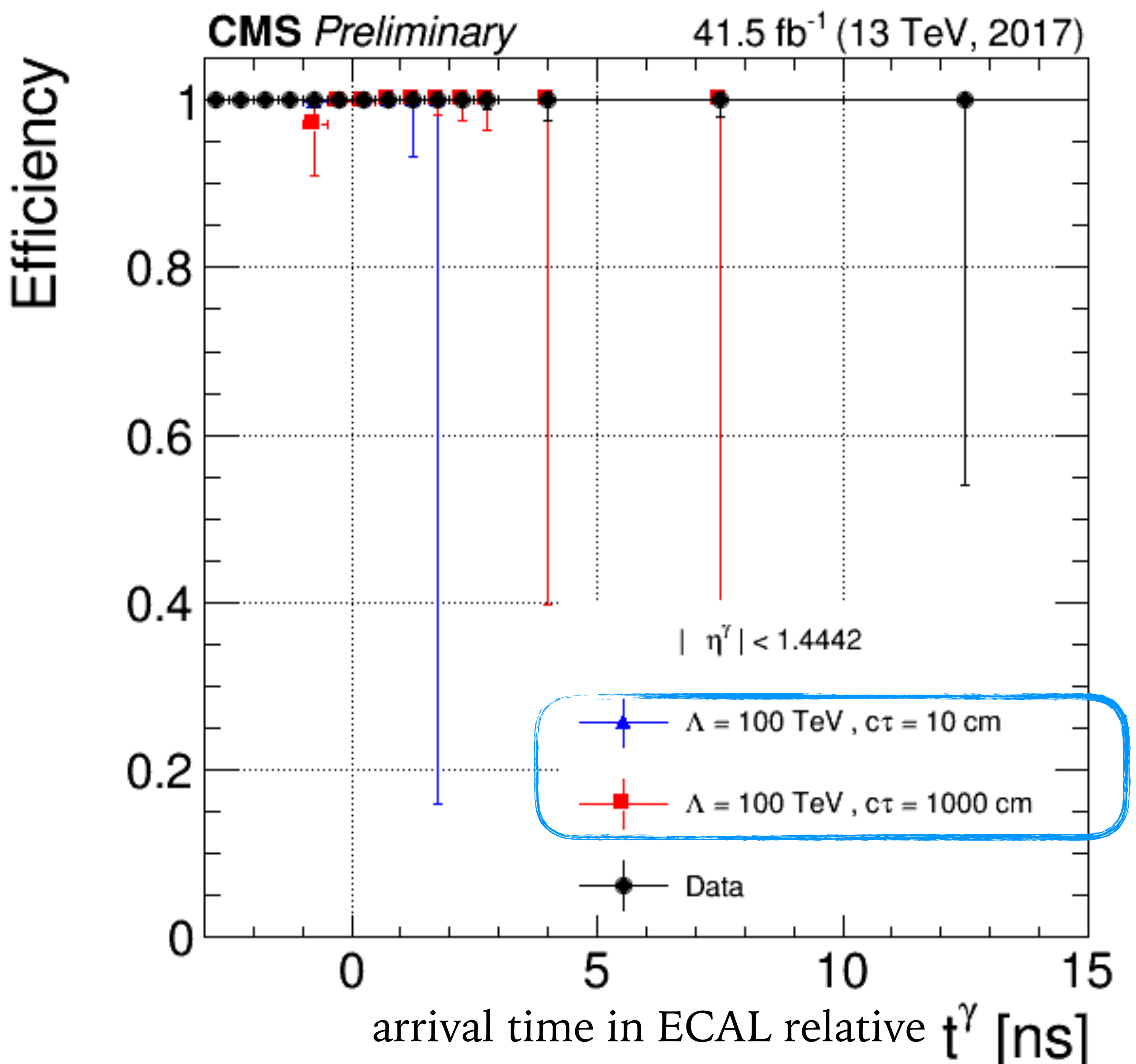
DISPLACED PHOTON TRIGGER

Dedicated displaced photon trigger to aid searches for long-lived particles

Designed to be efficient for both prompt and delayed photons

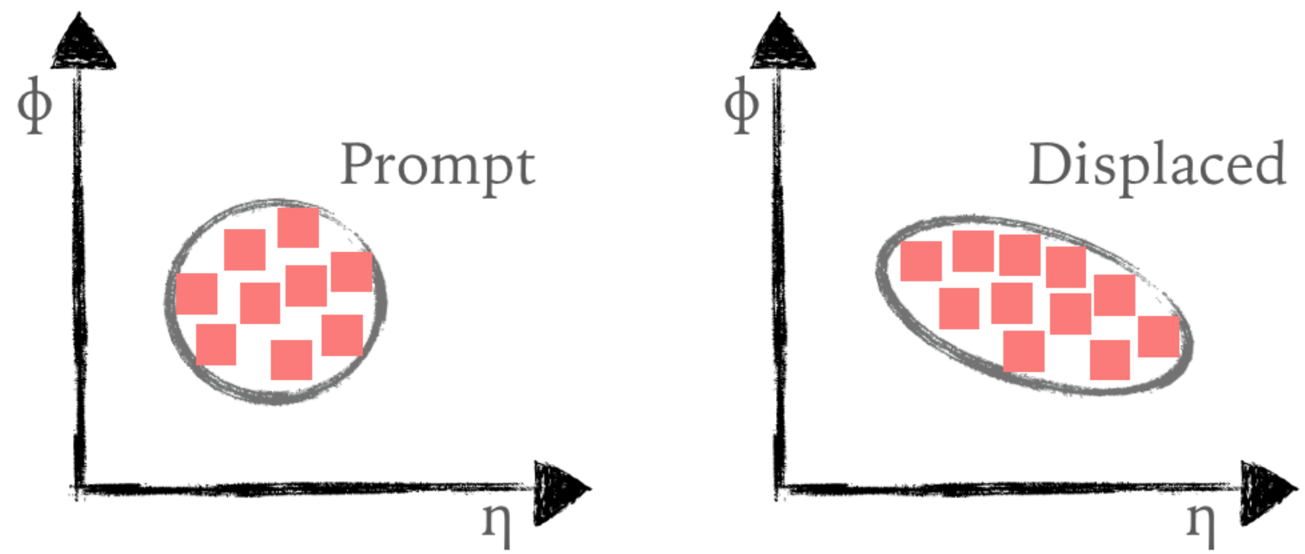
To facilitate background estimation techniques

Exploit elliptical energy deposition pattern in ECAL



arrival time in ECAL relative to time of bunch crossing

GMSB, SUSY breaking scale 100 TeV



Efficiency of displaced photon identification sequence vs photon seed time

https://cds.cern.ch/record/2714882/files/DP2020_015.pdf

ELECTRON IDENTIFICATION (OFFLINE)

Generic MVA ID

2 flavors: Trained with and w/o isolation variables

Excellent background rejection at a given signal efficiency: **Powerful**.

Generic cut based ID

Flip any cut to perform side band study: **Flexible**.

Dedicated high pT cut based ID

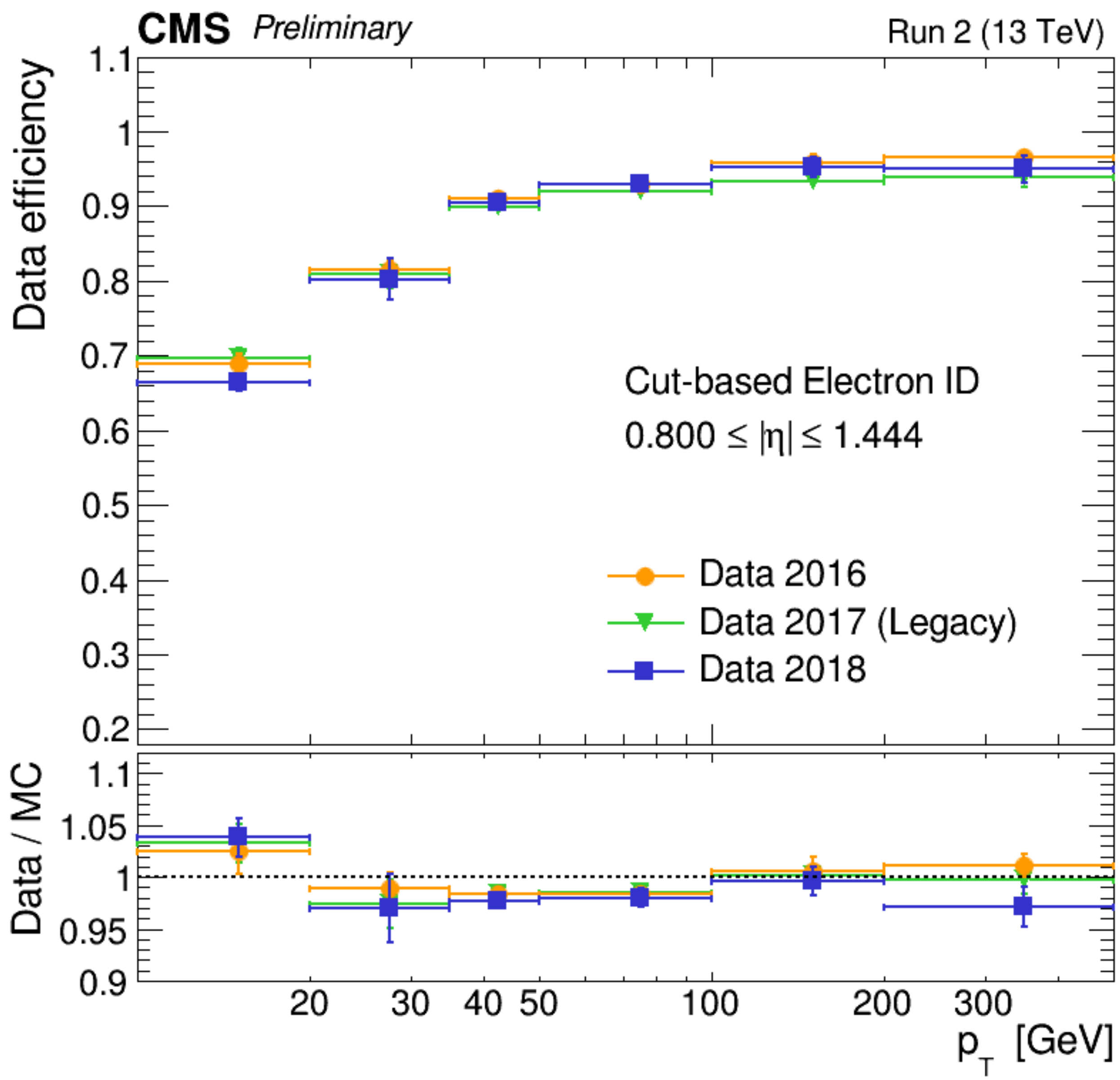
Stable and robust efficiency vs pT

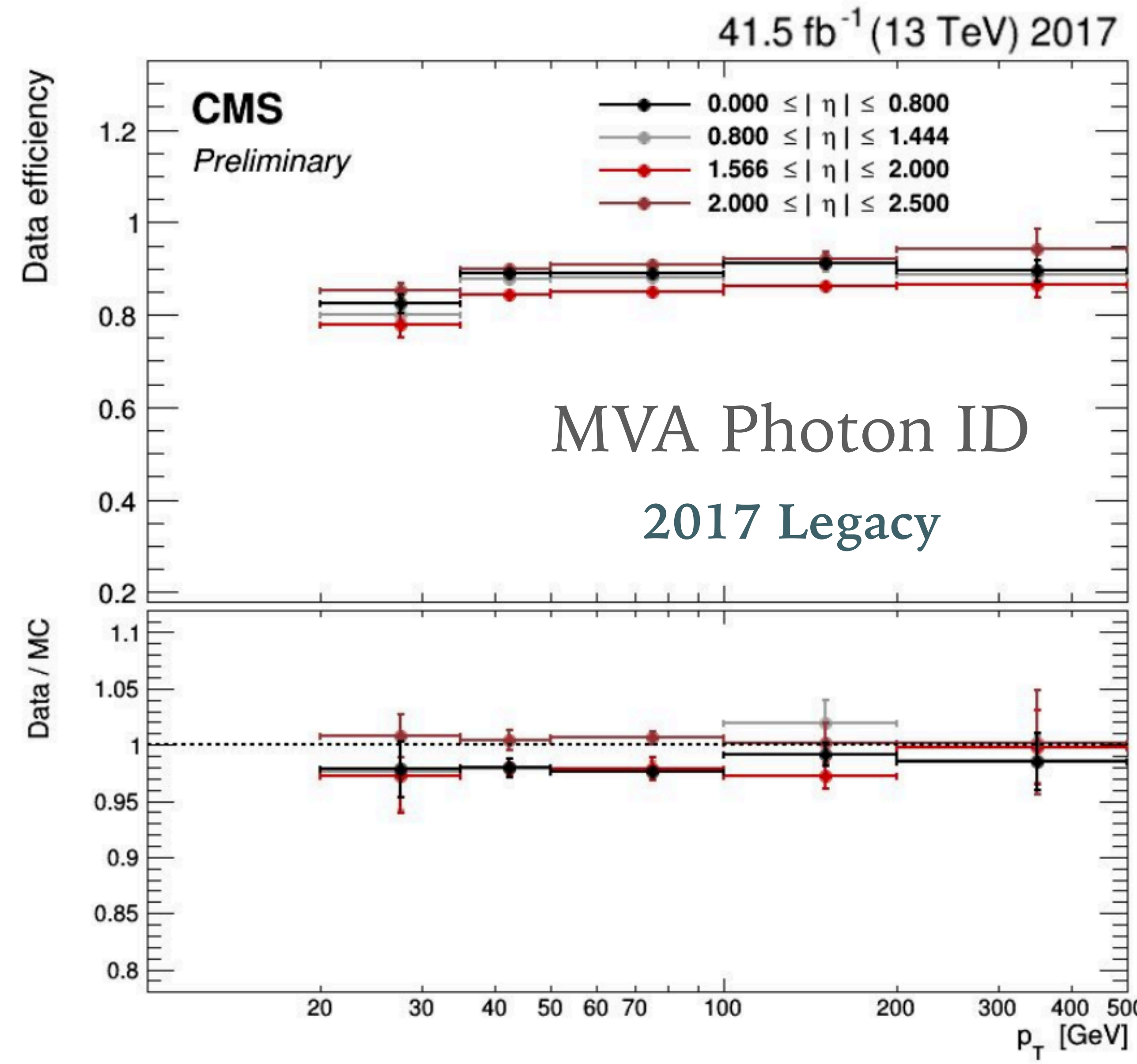
Generic IDs have different working points, corresponding to different signal efficiency & background rejection.

Looser working points: Good for high mass searches where background is low. When efficiency more important than purity.

Tighter working points: Good for precision measurements / searches that deal with significant amount of background. When purity more important than efficiency.

IDENTIFICATION EFFICIENCY (OFFLINE)





Cut based electron ID

Similar efficiency and data/MC correction factors across 3 years

MVA photon ID

Data/MC correction factors always within 3%

https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSEgammaFullRun2PerformancePlotsICHEP2020
https://cds.cern.ch/record/2718815/files/DP2020_024.pdf

ENERGY SCALE AND RESOLUTION

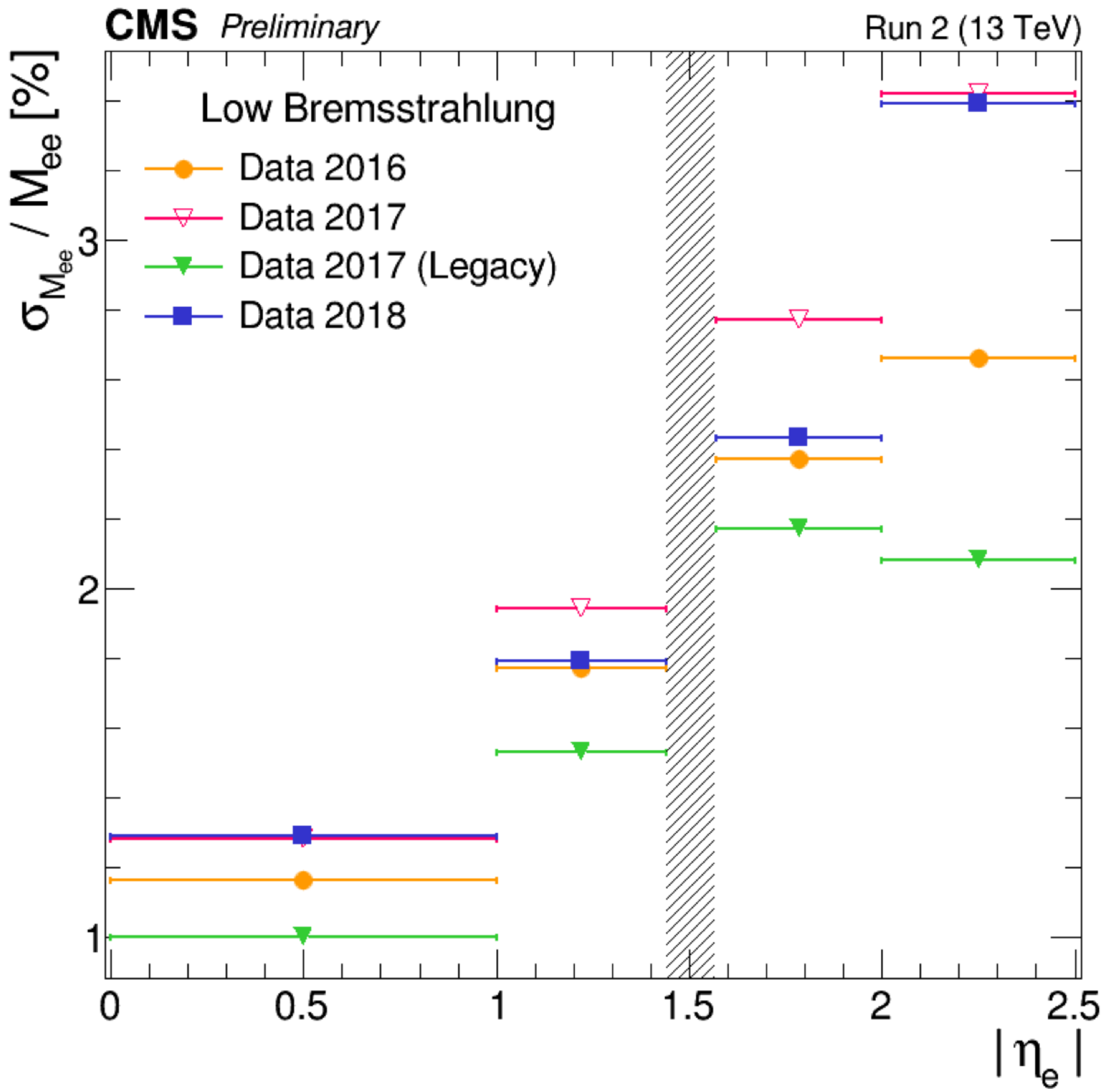
Multi-step BDT-based **energy regression**. Energy correction factors obtained from MC, applied on data and MC.

After applying corrections from regression, data have slightly different energy scale and slightly worse resolution compared to MC

Estimate **residual scale and smearing** corrections from Z(ee) events

Calibrate data so that scale in data matches scale in MC

Smear MC resolution to match that of data

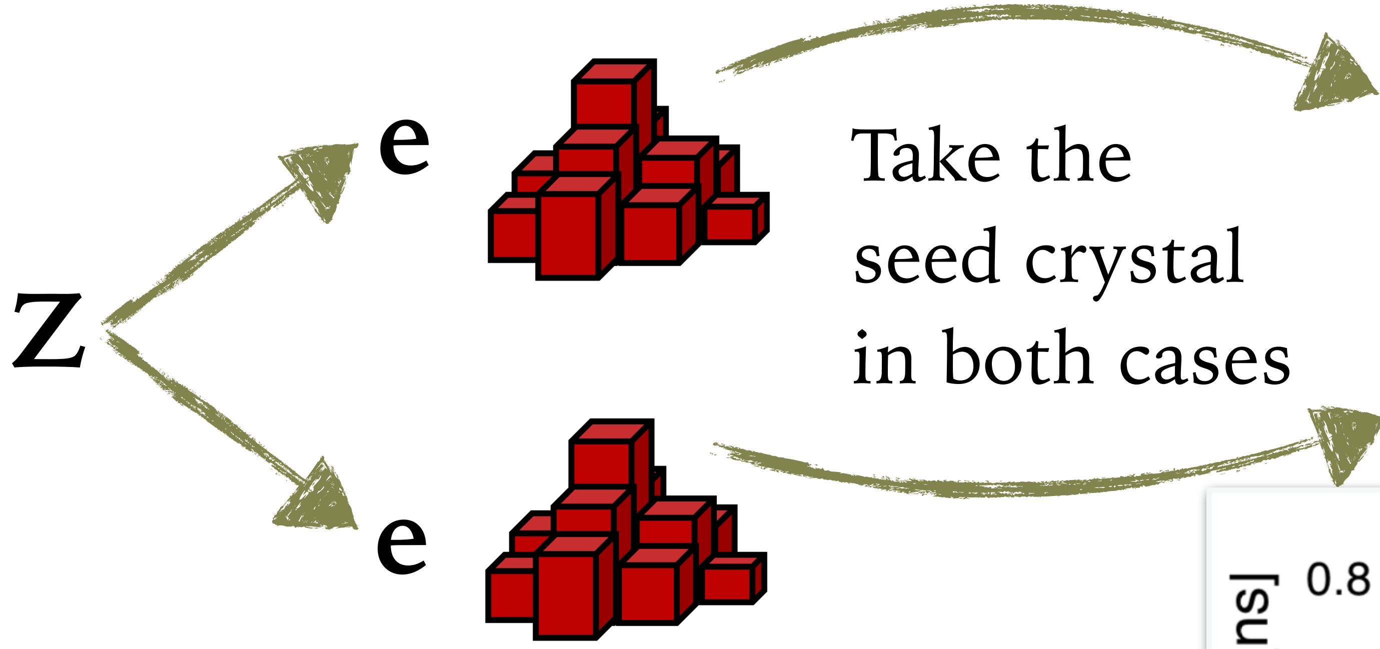


Excellent energy resolution, 1-3.4% depending on η

~40% improvement in highest η bin in legacy-2017 w.r.t EOY-2017

https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSEgammaFullRun2PerformancePlotsICHEP2020

TIME RESOLUTION

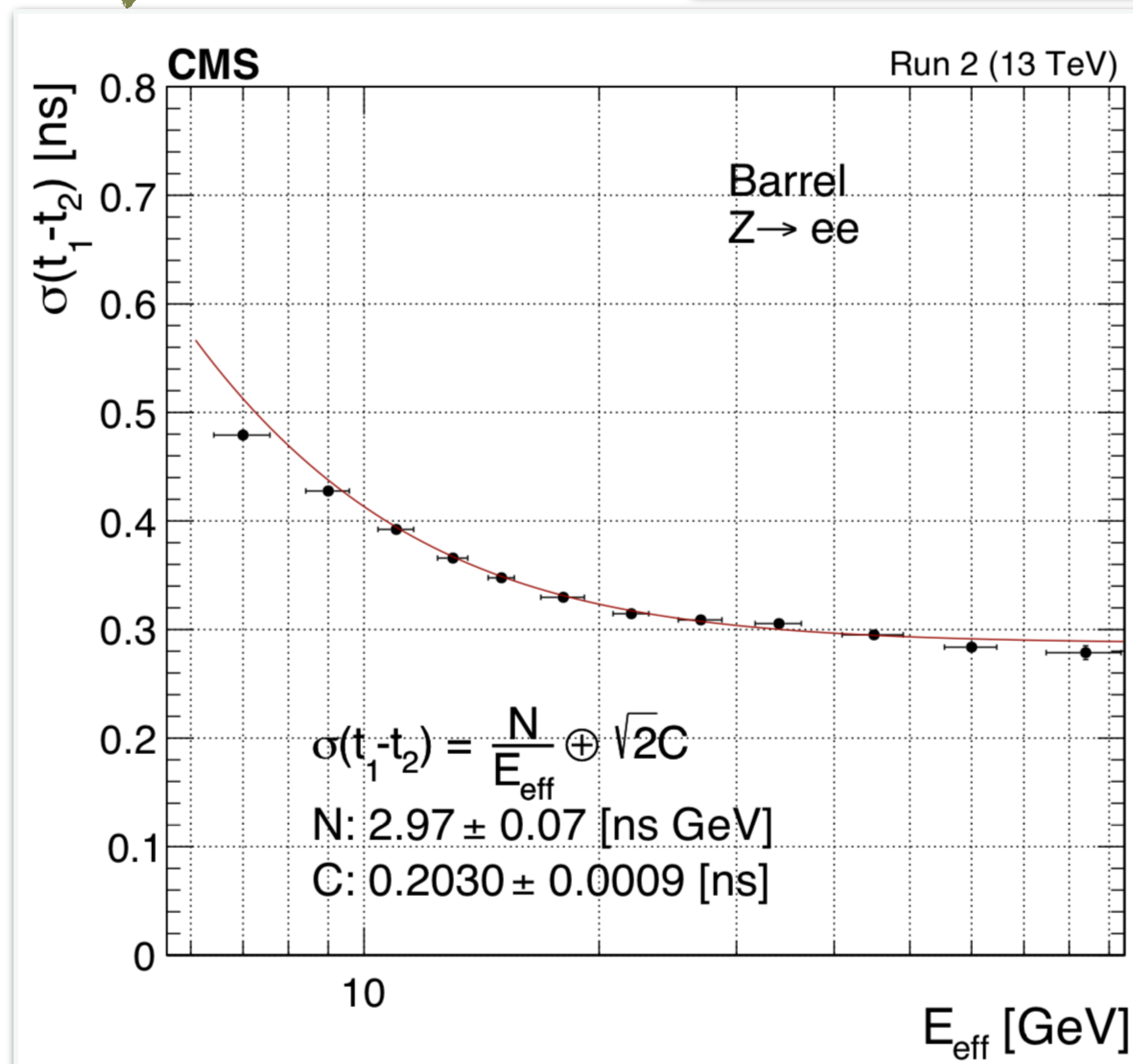


Seed energy = E1, E2

10< E1, E2 <120 GeV

Seed time = t1, t2

$$E_{\text{eff}} = \frac{E_1 E_2}{\sqrt{E_1^2 + E_2^2}}$$



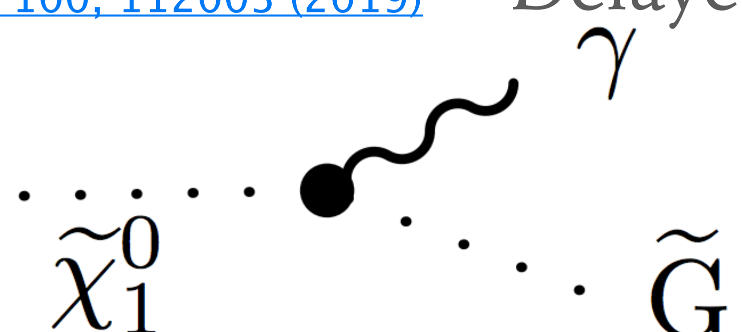
Overall resolution ~300 ps, and constant term ~200 ps.

Searches for long-lived particles profit from this

Example:

[Phys. Rev. D 100, 112003 (2019)]

Delayed photon



Long-lived

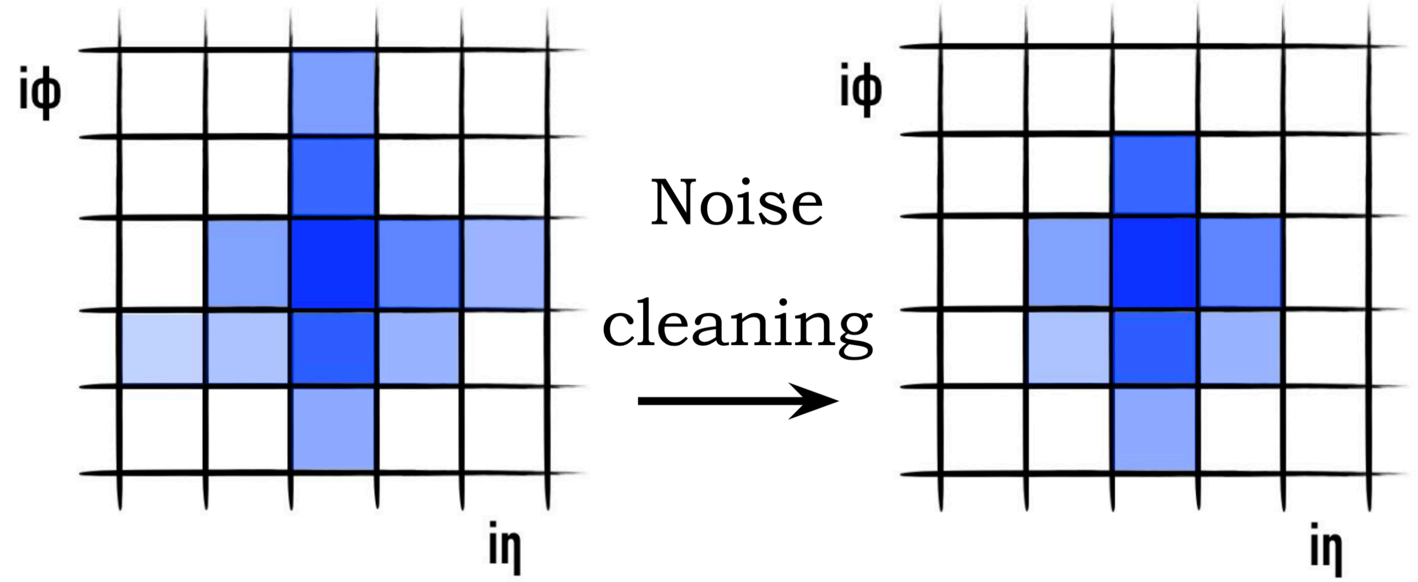
https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSEgammaFullRun2PerformancePlotsICHEP2020

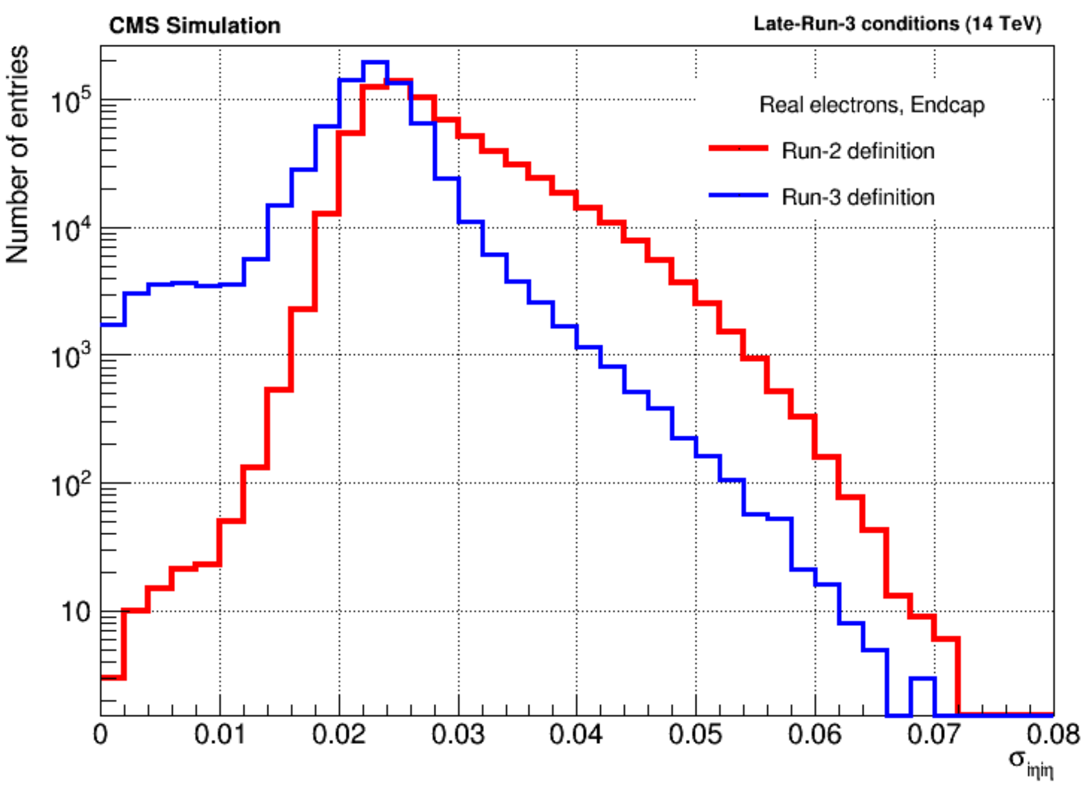
REDEFINITION OF SHOWER SHAPE VARIABLE FOR RUN-3

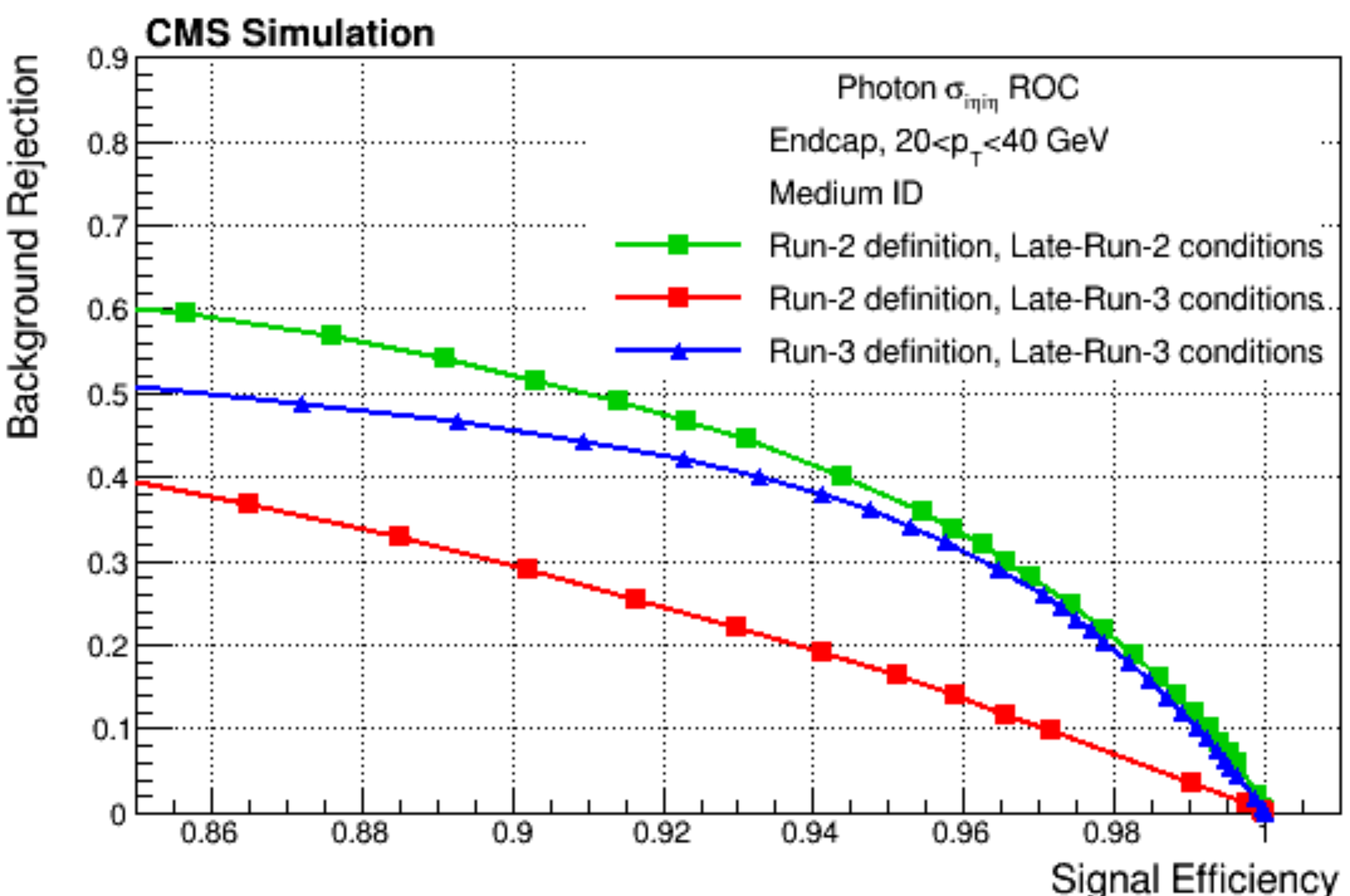
Radiation damage leads to increased electronic noise in ECAL in Run-3

Discrimination power of **σiηiη** affected by high noise

Mitigate the effect by noise cleaning







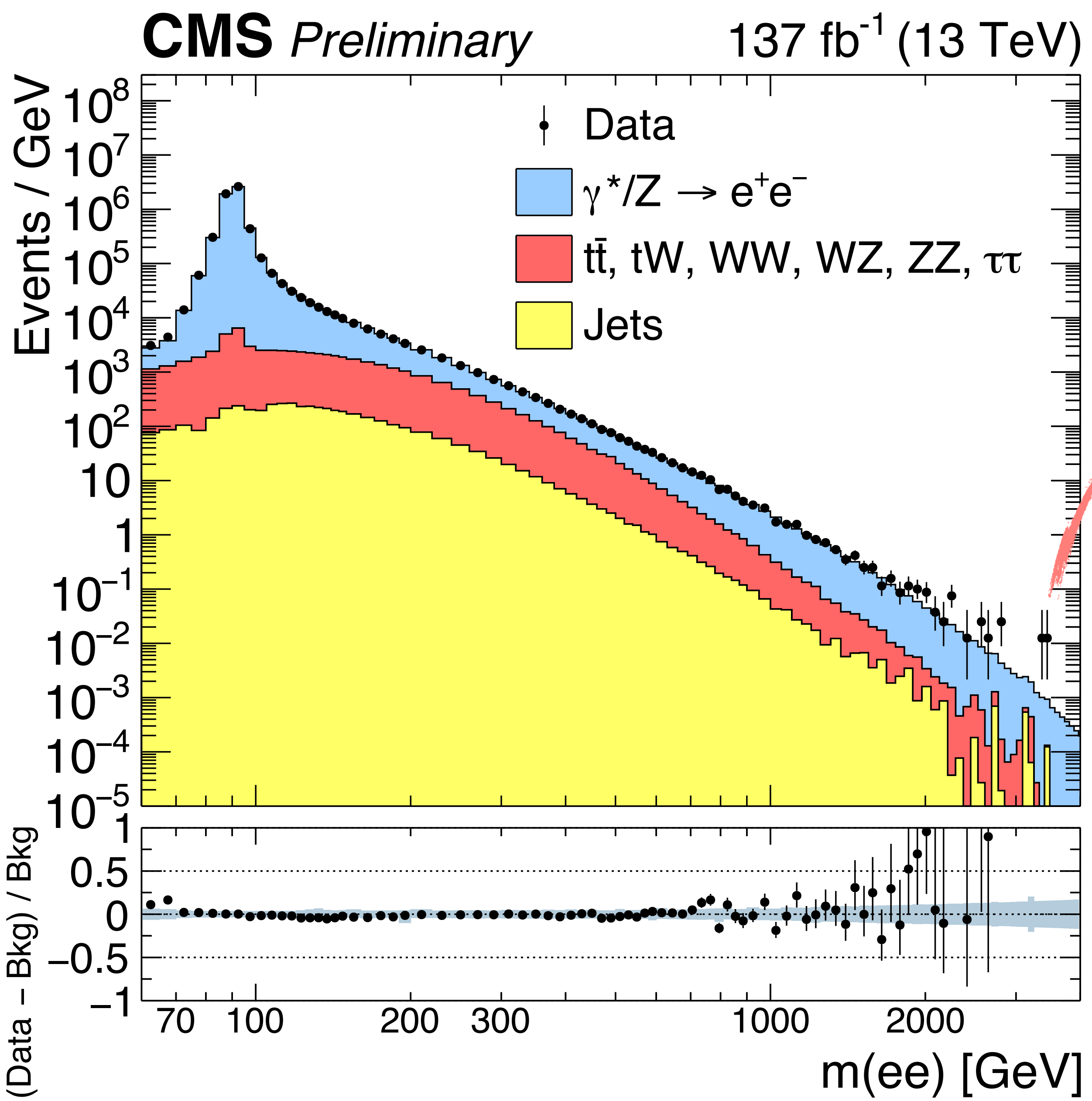
Performance partially regained with redefinition

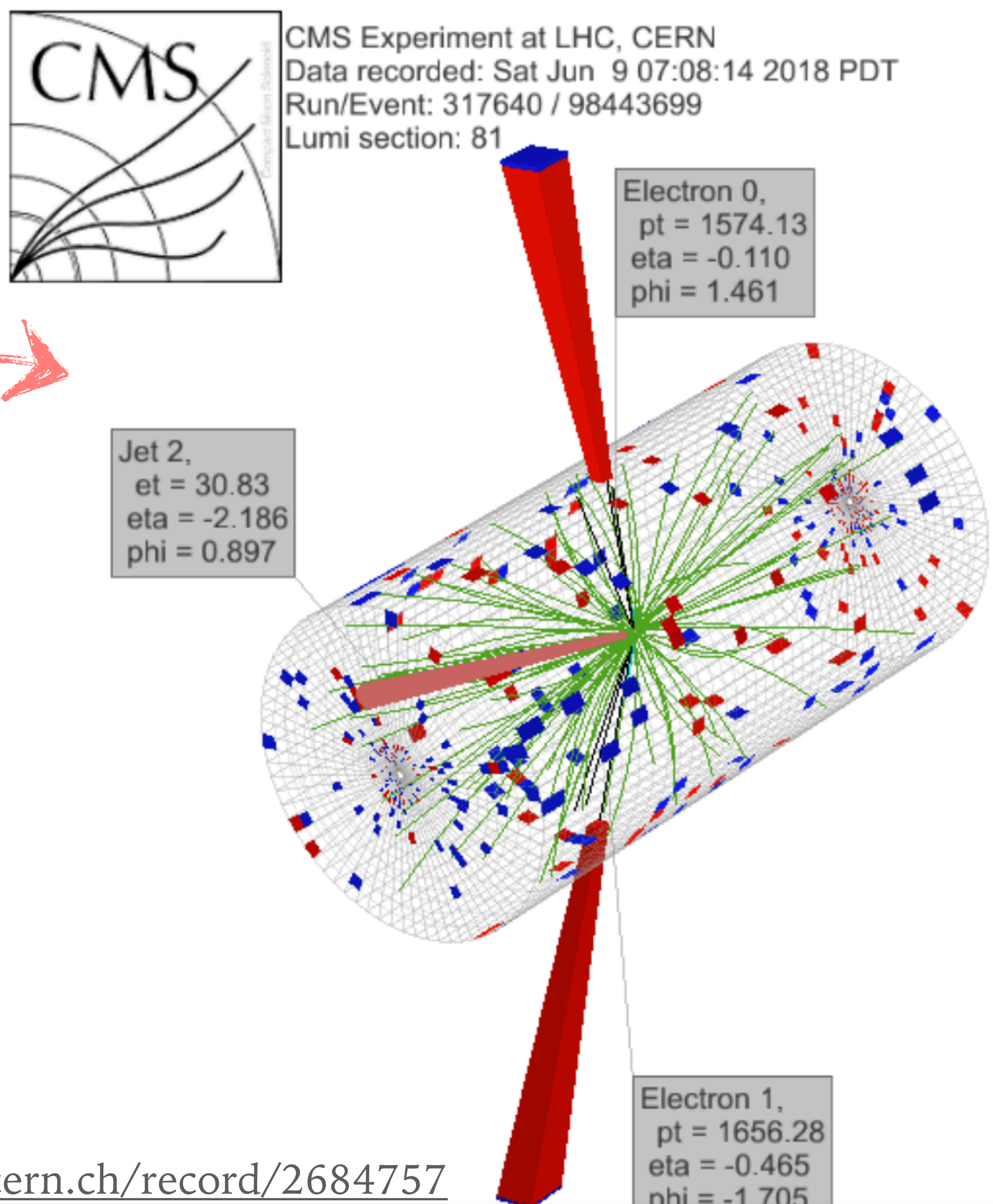
https://twiki.cern.ch/twiki/bin/view/CMSPublic/CMSEgammaRun3IDvariable

SUMMARY

Physics program of CMS profits from

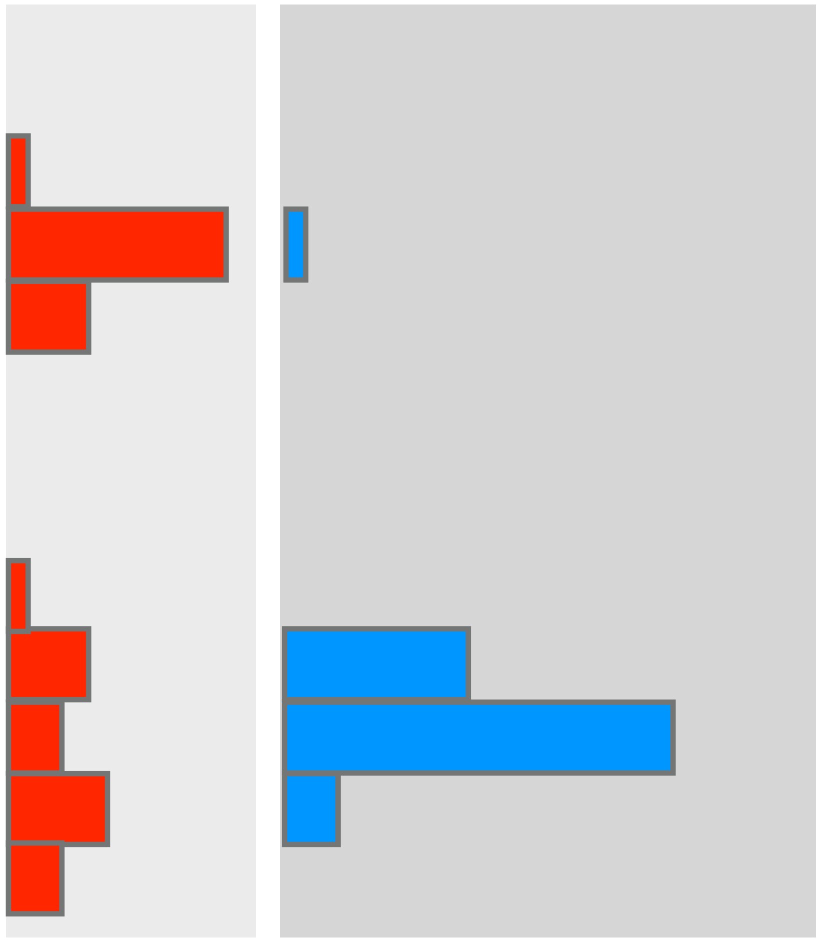
- Diverse and robust electron/photon triggers
- Highly efficient reconstruction algorithms & identification criteria of electron/photon
- Precise measurements of energy and time electron/photon

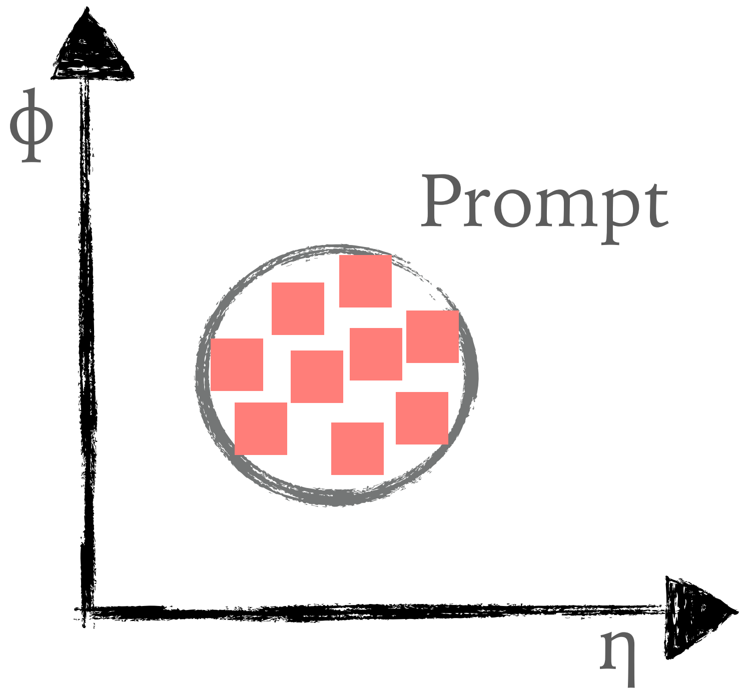


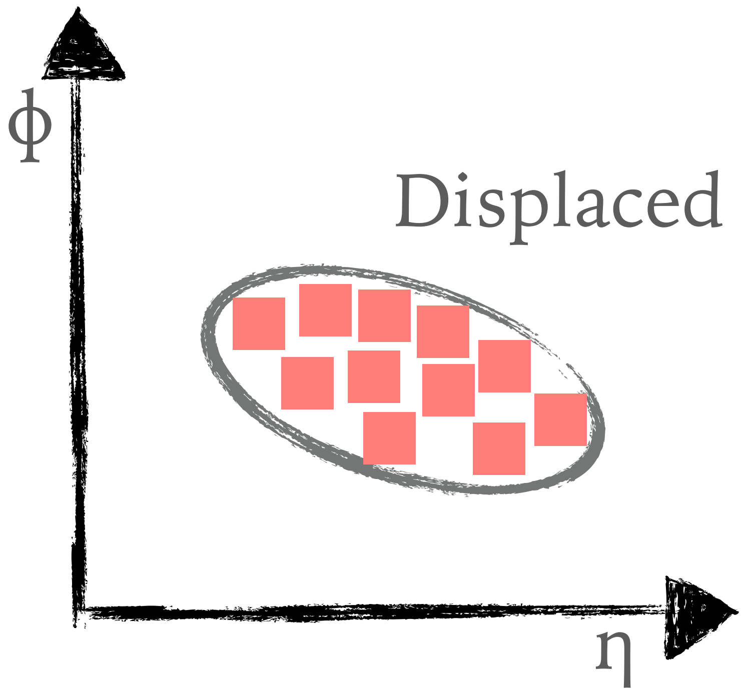


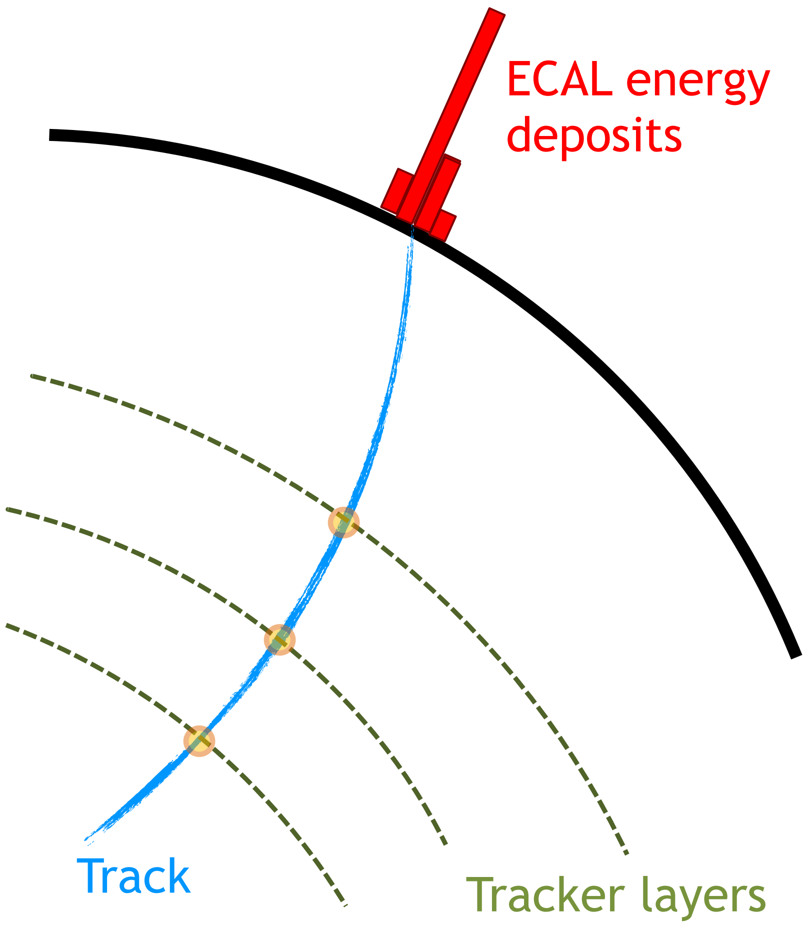
https://cds.cern.ch/record/2629365/files/DP2018_043.pdf & http://cds.cern.ch/record/2684757

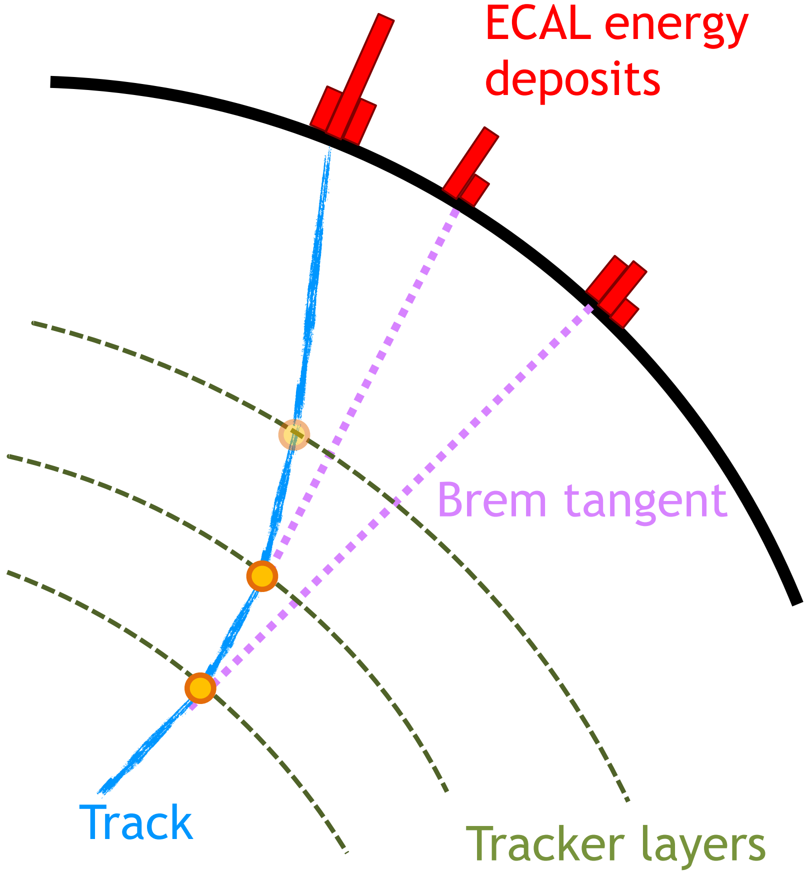
EXTRA SLIDES




ϕ
Prompt
η


ϕ
Displaced
η


ECAL energy deposits
Track
Tracker layers


ECAL energy deposits
Brem tangent
Track
Tracker layers

E/GAMMA OFFLINE RECONSTRUCTION FLOW

Clustering and SuperClustering algorithms

Clustering:

group together recHits with energy above a threshold

ECAL & ES RecHits → ECAL & ES PFClusters → Mustache SuperClusters

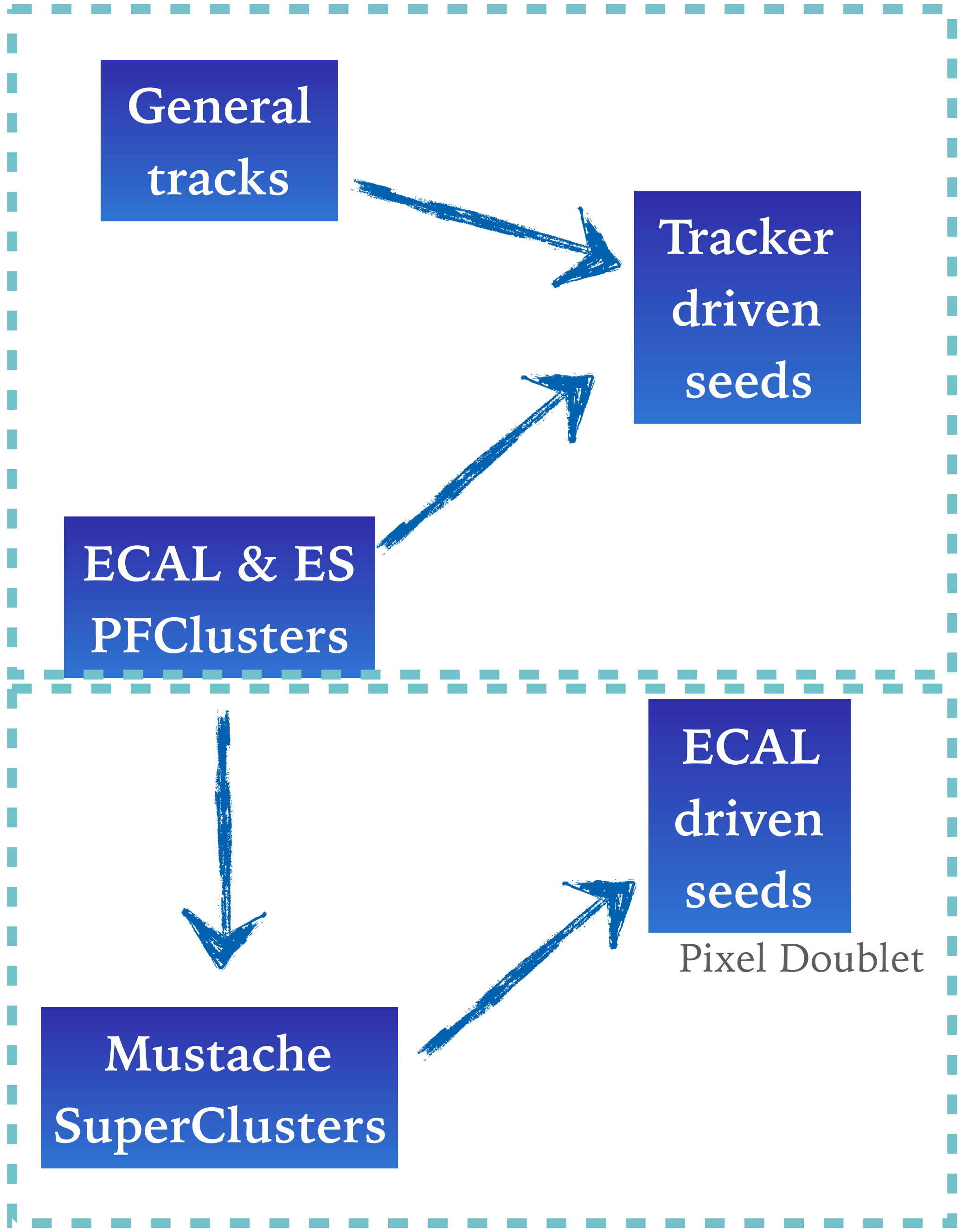
SuperClustering:

group together clusters around the seed(most-energetic) cluster, within a geometrical area.

Efficiently recovers brem / conversion

E/GAMMA OFFLINE RECONSTRUCTION FLOW

Electron track seeding



2 types of tracking seeds:

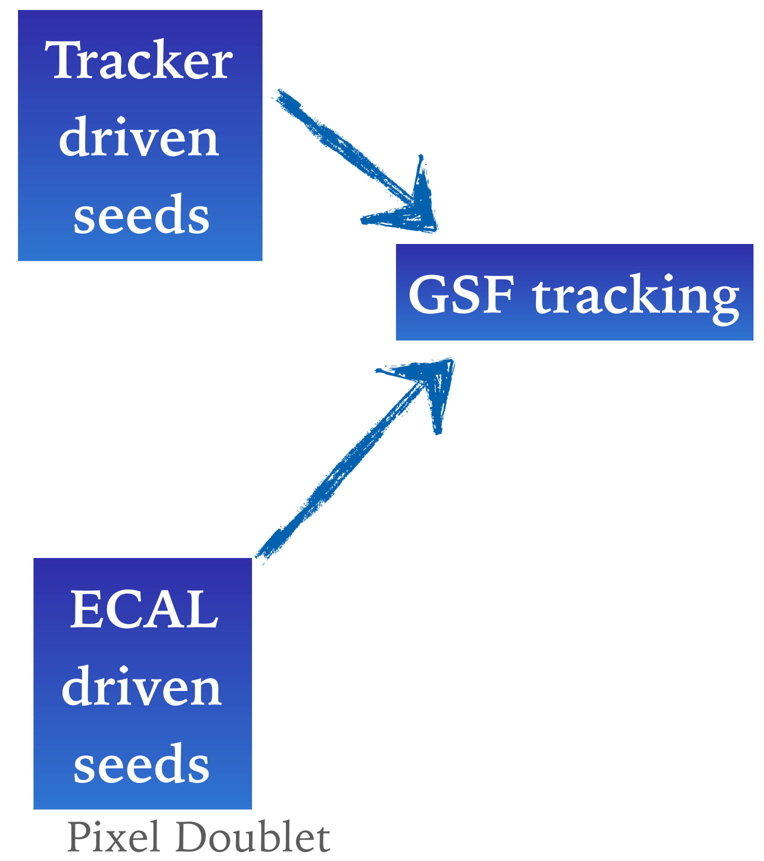
- **ECAL driven (outside-in)**
  - Start from mustache SC and look for pixel-match (doublet)
  - Useful for medium-high pT isolated electrons
- **Tracker driven (inside-out)**
  - Start from general track collection and match with PFClusters using PF techniques
  - Useful for low pT, non-isolated electrons

E/GAMMA OFFLINE RECONSTRUCTION FLOW

Electron track reconstruction

Dedicated **GSF** tracking

- Takes into account radiative loss due to brem so that we can measure p_{in} , p_{out} and p at any intermediate layer
- Brem tangents for later matching to missed ECAL clusters

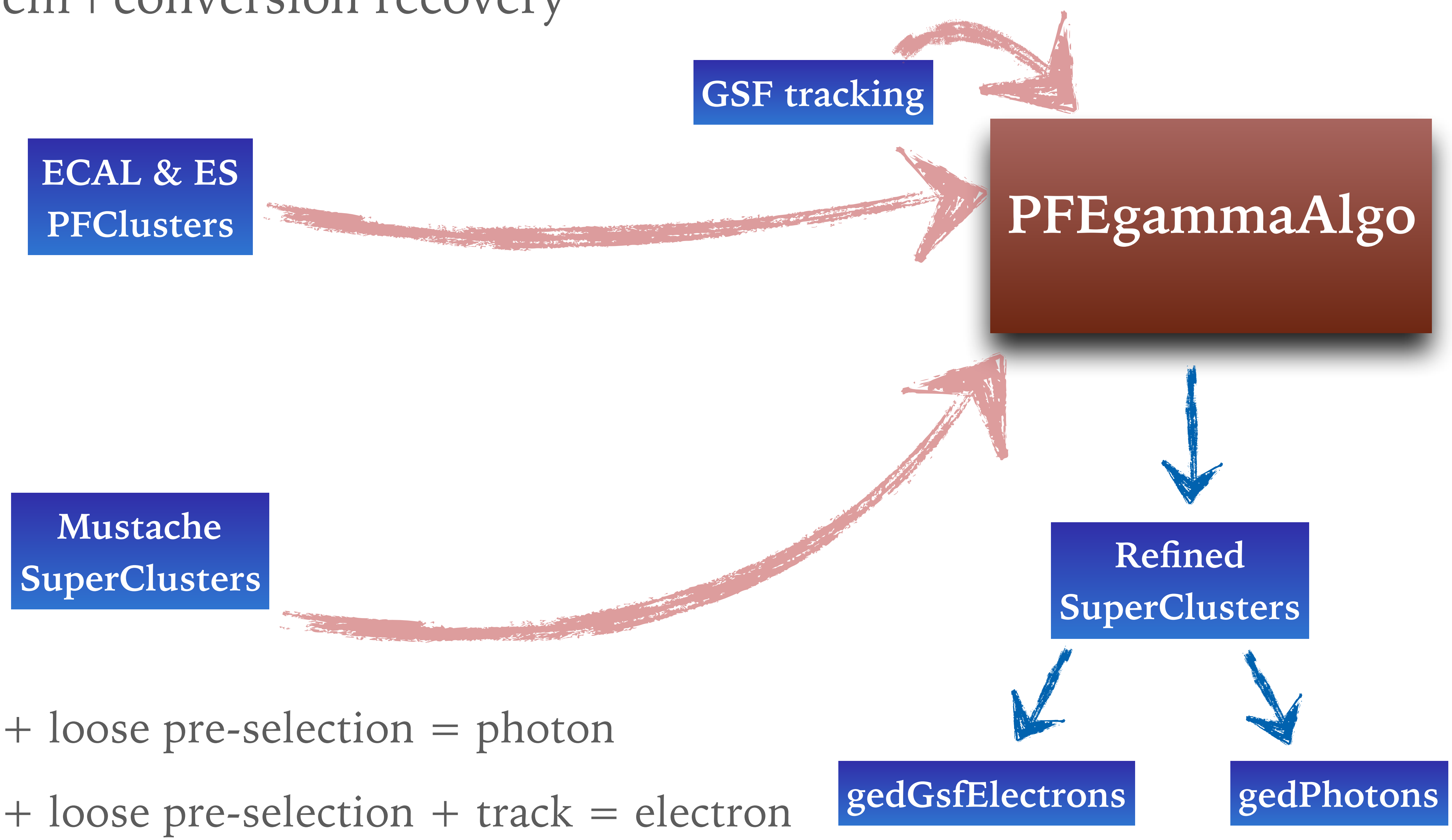


E/GAMMA OFFLINE RECONSTRUCTION FLOW

Refinement & final objects

Associate additional ECAL clusters to mustache SC by looking at GSF track tangents

additional brem+conversion recovery



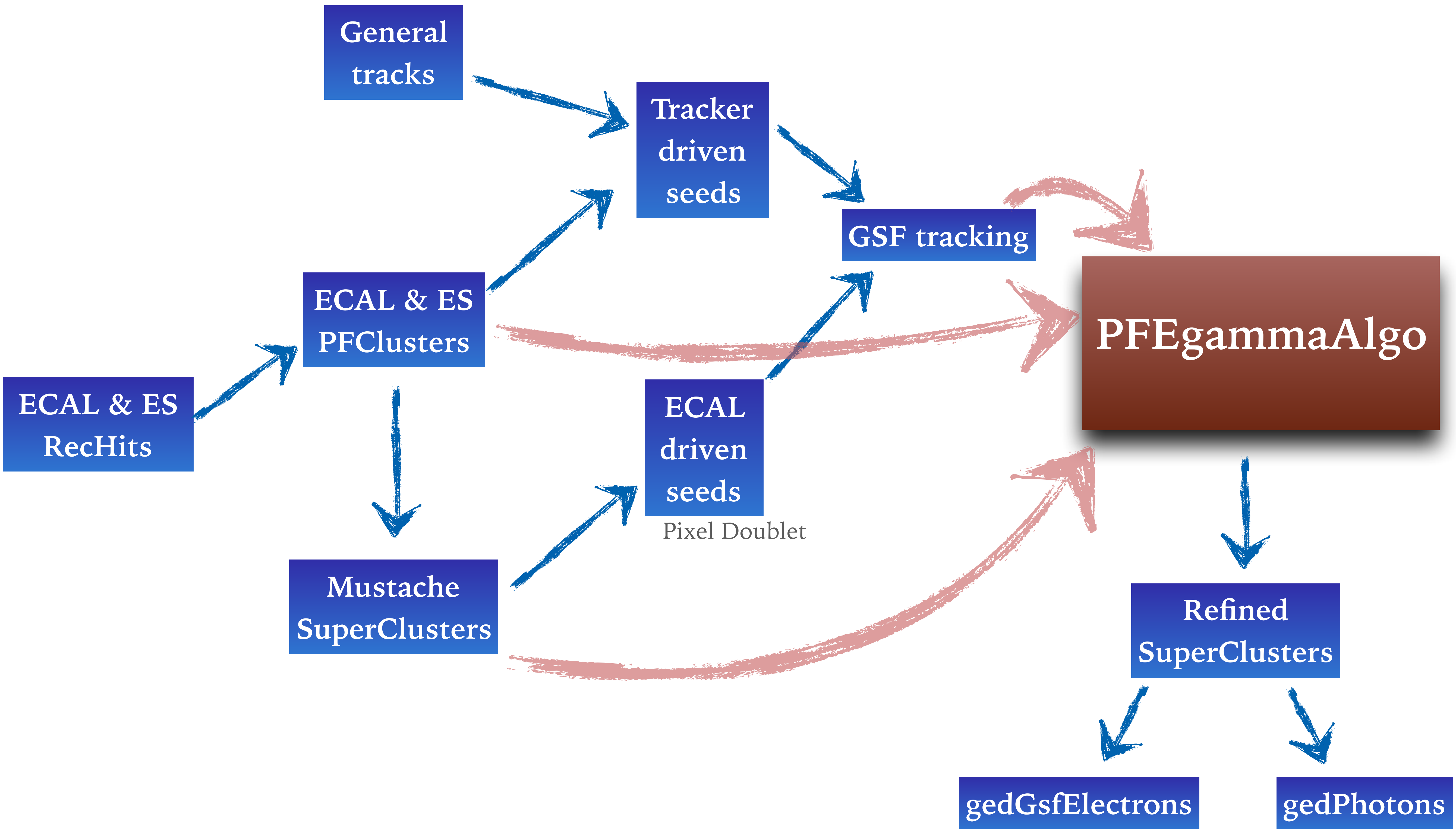
Refined SC + loose pre-selection = photon

Refined SC + loose pre-selection + track = electron

E/GAMMA OFFLINE RECONSTRUCTION FLOW

Reconstruction@Offline

Putting it all together



E/GAMMA ONLINE RECONSTRUCTION FLOW

Reconstruction@HLT

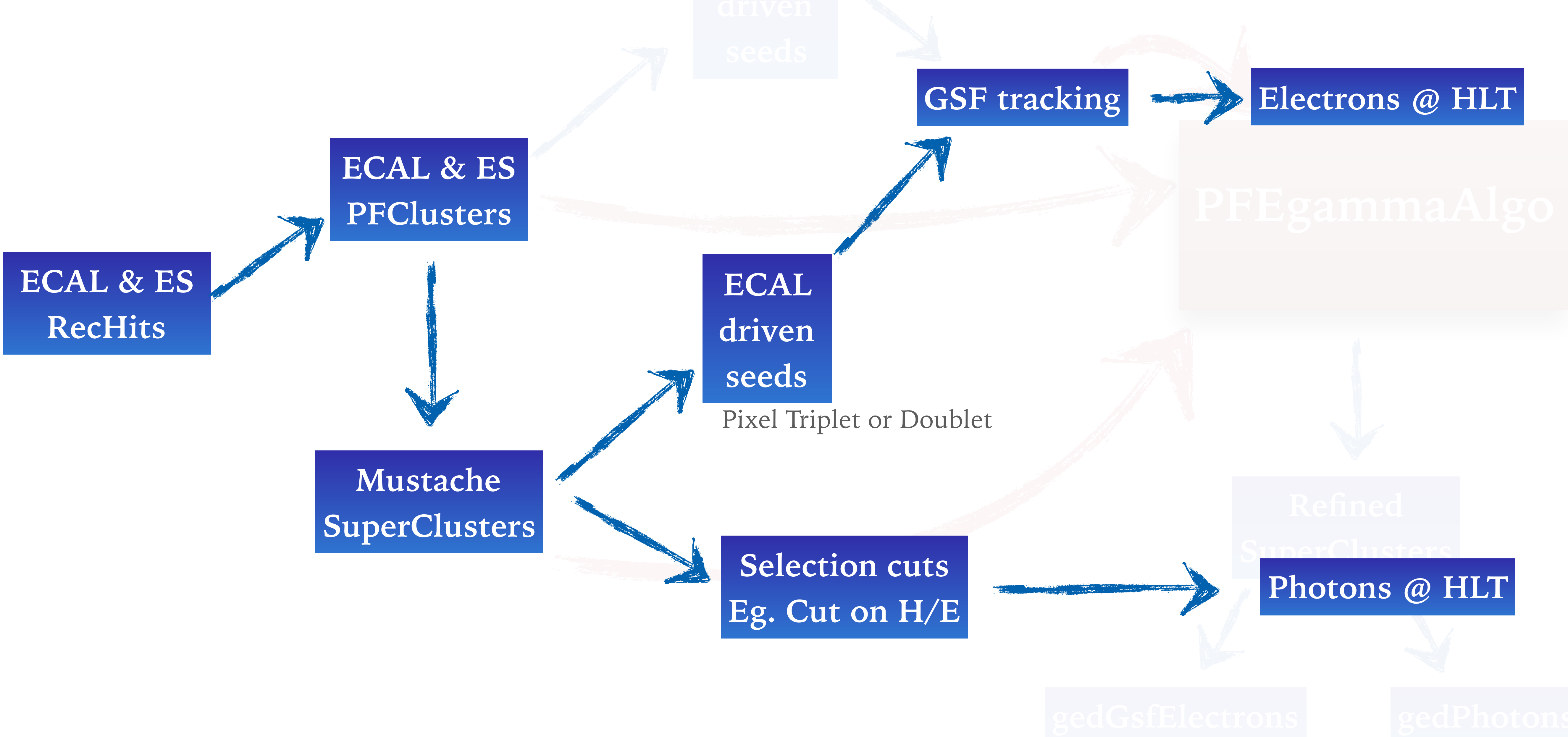
For timing constraints:

No tracker-driven seed

No refined supercluster

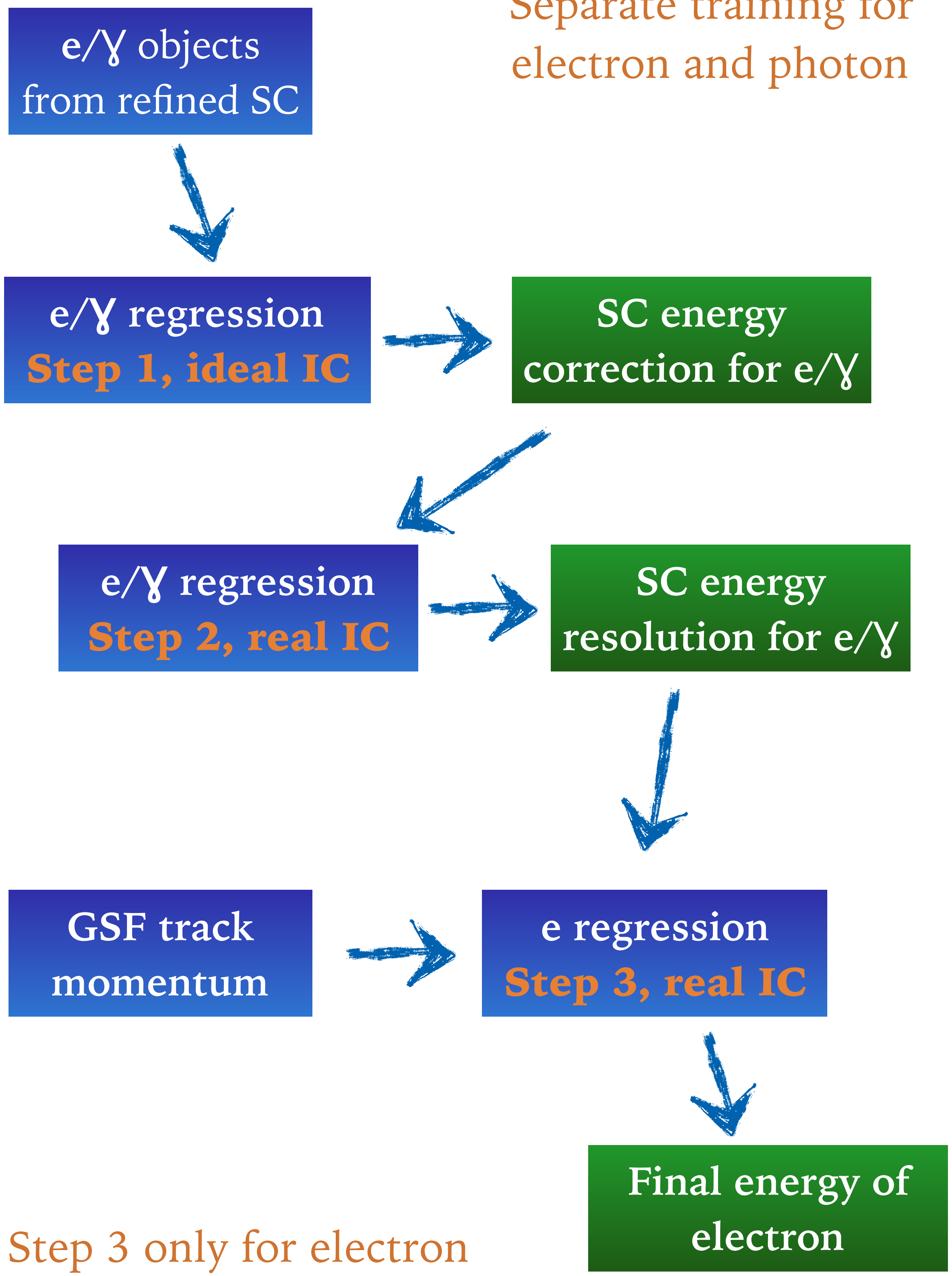
For background rate reduction:

Pixel ‘triplet’ seed whenever possible



ENERGY REGRESSION: STEPS

Separate training for electron and photon



Step 3 only for electron

Crystal-to-crystal variation in light output.

Variation effect taken care of in MC by inter-calibration constants (IC).

ICs not 100% accurate, which affects energy resolution.

There are 2 types of MC for regression.

1) **This inaccuracy is taken care of by applying a random smearing of IC (real IC)**
2) **This inaccuracy not taken care of (ideal IC)**

From past studies: Step 1 "learns" the random smearing, and energy correction factors are correlated with smearing.

So, Ideal IC used for Step 1, to remain unbiased.

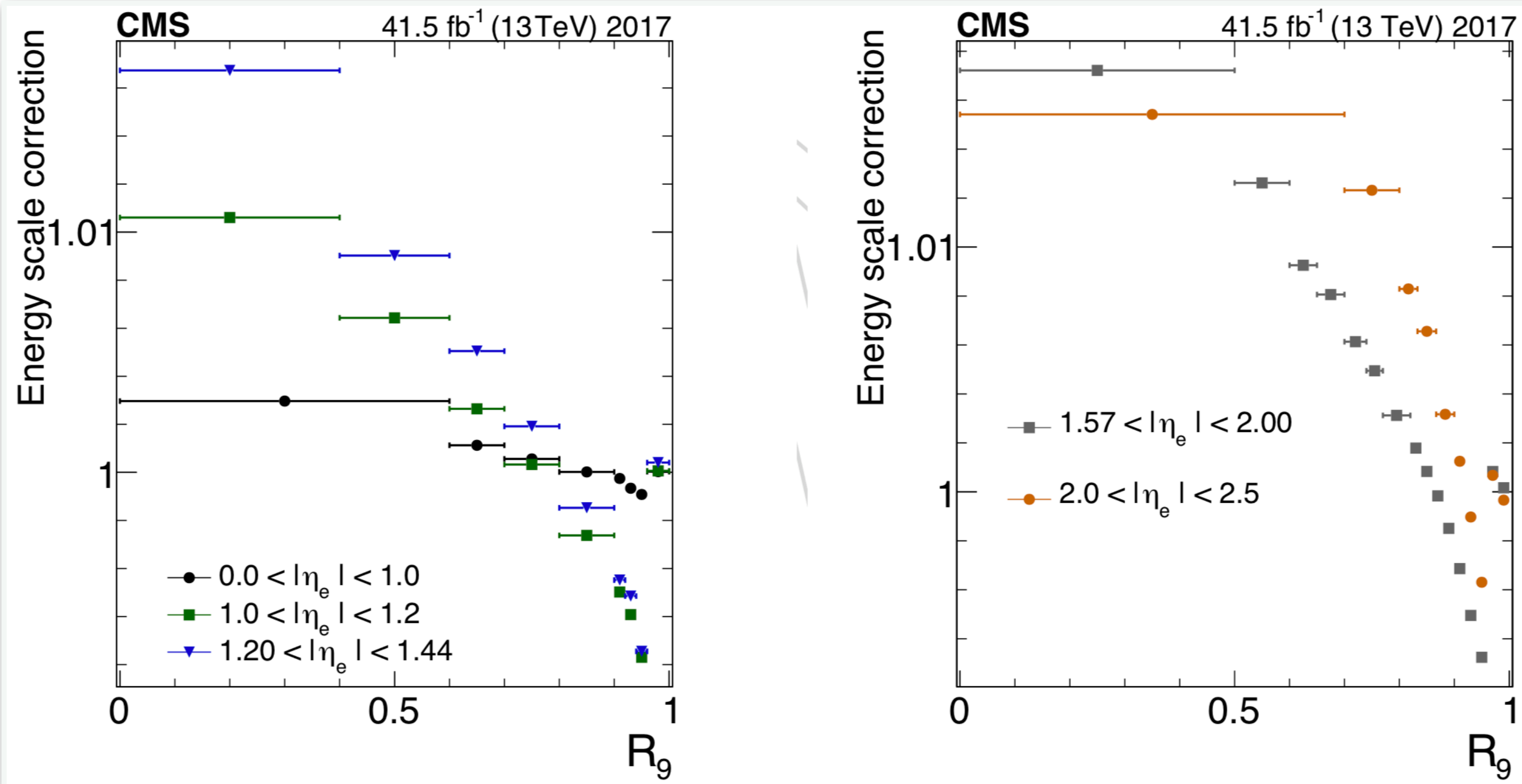
Real IC used for Step 2 and 3

RESIDUAL CORRECTIONS (SCALE & SMEARING)-I

- ✓ After applying corrections from regression, data have slightly different energy scale and slightly worse resolution compared to MC. Why?
  - ✓ Calibrations in data not perfect
  - ✓ Applying MC-based correction to data can lead to mis-calibration, as simulation is not perfect.
- ✓ Solve both effects by deriving residual energy corrections (scale & smearing)
  - ✓ Calibrate data so that scale in data matches scale in MC
  - ✓ Smear MC resolution to match that of data
- ✓ Estimate scale and smearing corrections from Z(ee) events
  - ✓ First step: Scale correction using simple and quick **Fit method**. Use **BW⊗OSCB** to fit data and MC separately, extract scale correction. **Time(LHC fill) and η dependent correction factors.**

RESIDUAL CORRECTIONS (SCALE & SMEARING)-II

- ✓ Second step: Remaining scale correction + resolution correction using more accurate but time consuming **Smearing method**. **ML fit to data, using MC Z(ee) shape** which already accounts for detector effects, reconstruction inefficiencies, kinematic properties. **η and R9 dependent correction factors.**



Energy correction factors from Smearing method
Stat-only uncertainty shown in plots

Systematics derived by changing R9, and selection cuts.

0.05–0.1% for barrel electrons, 0.1–0.3% for endcap electrons

RESIDUAL CORRECTIONS (SCALE & SMEARING)-III

Third step: Additional gain dependent scale correction factors in EB and EE, for **high pT electrons** (which has gain-switched seed).

Scale correction factors

Gain 6: 1.008 (1.011) in EB(EE)

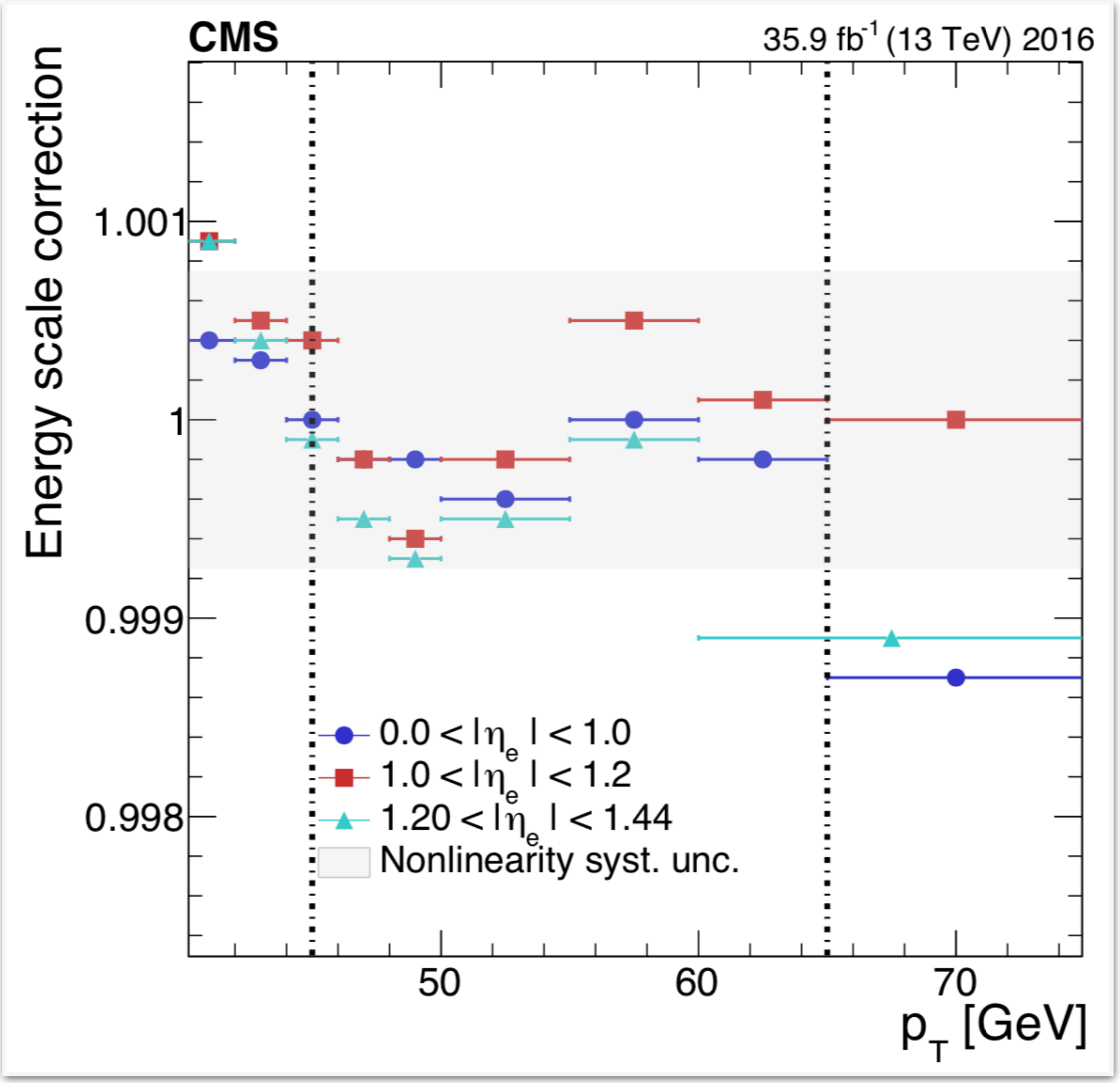
Gain 1: No correction factor (lack of statistics)

Systematics

Gain 6: Conservative systematics, assign 100% of the derived scale correction as systematic, amounts to 0.8% in barrel, 1.1 % in endcaps

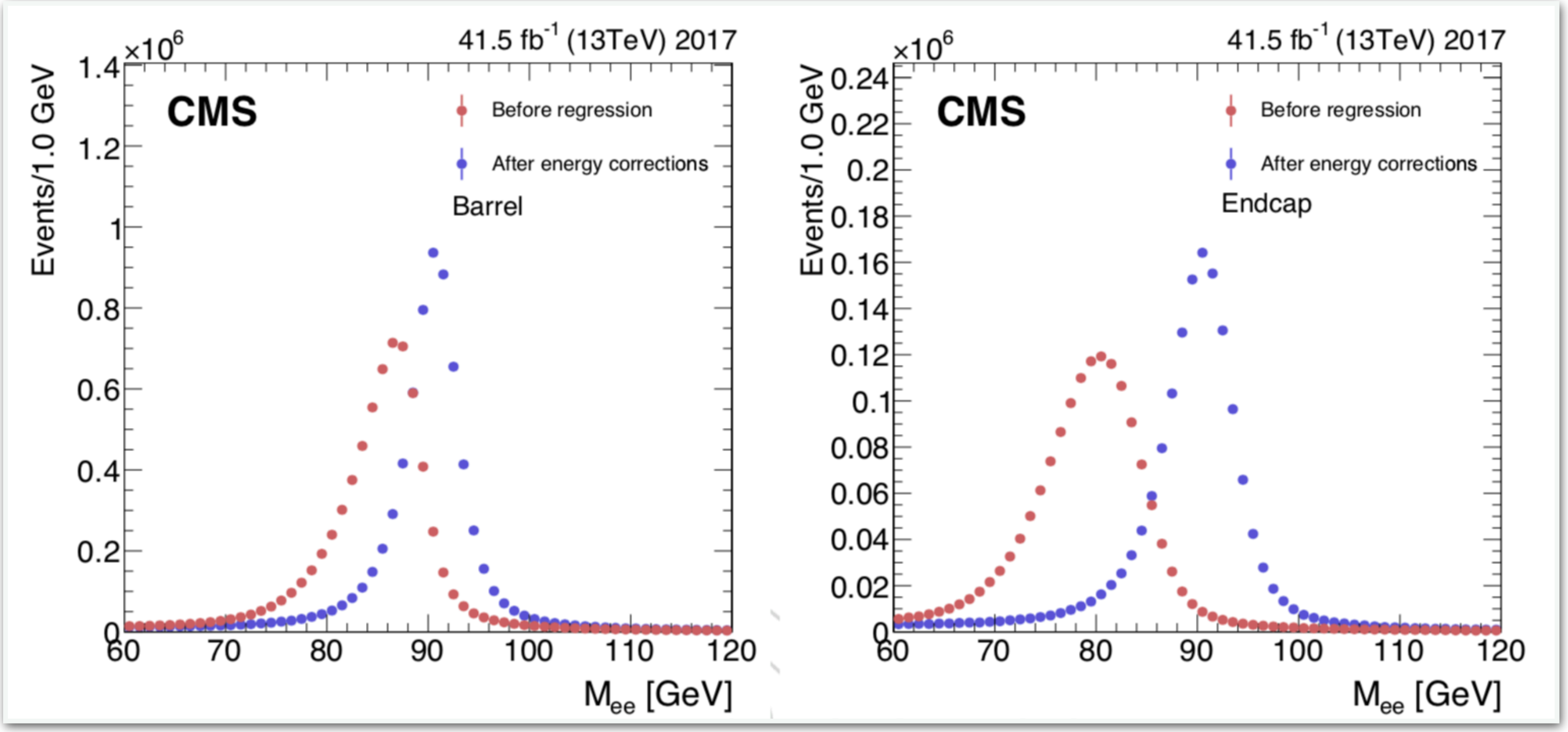
Gain 1: Take twice the Gain 6 systematics as conservative estimate of Gain 1 systematics, which is 2% in EB, 3% in EE

RESIDUAL CORRECTIONS (SCALE & SMEARING)-IV

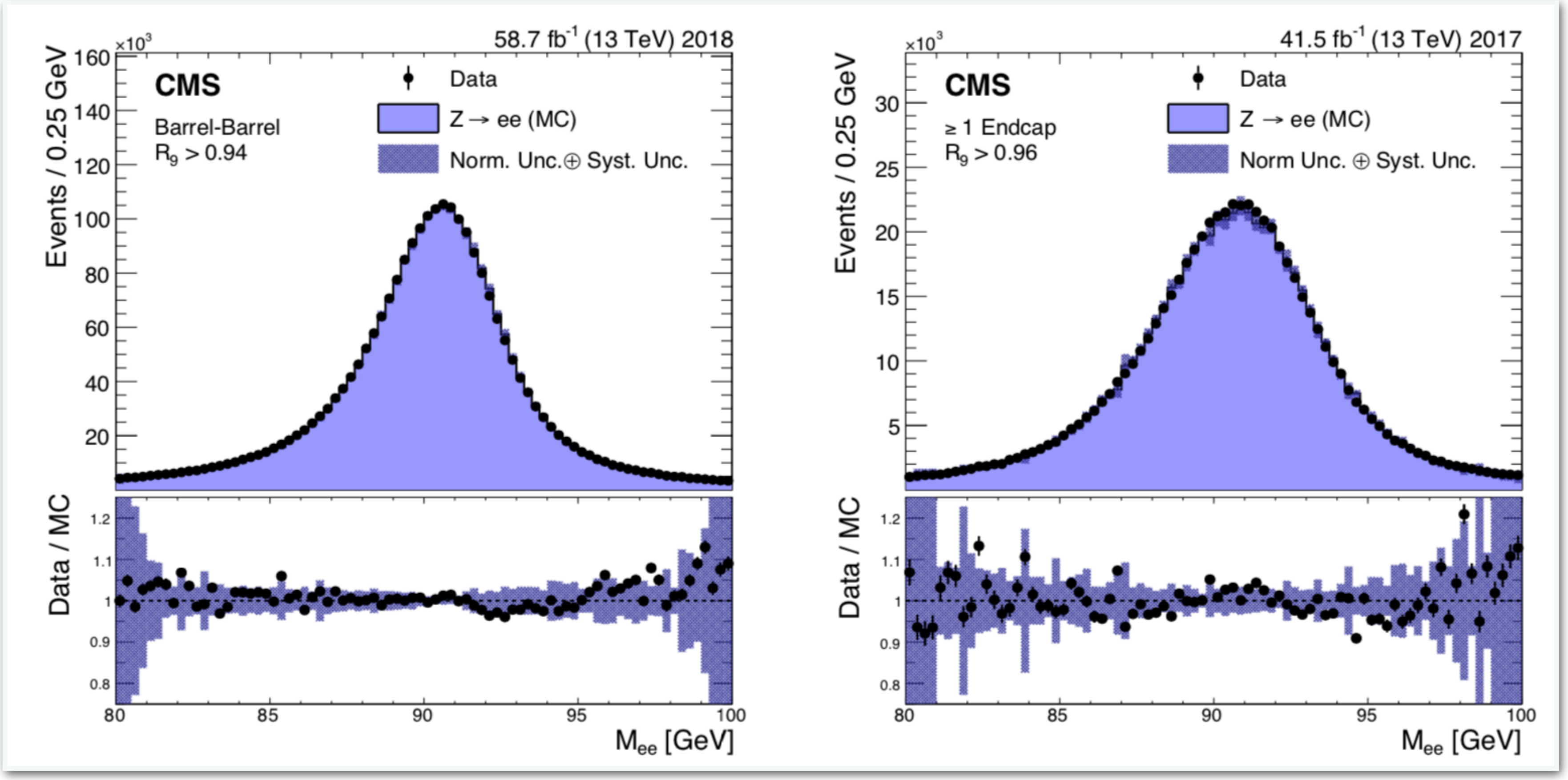


- Additional energy corrections in bins of η and pT
- Critical to precisely measure Higgs mass in diphoton channel.
- **This is a new plot added after pre-approval**

PERFORMANCE OF REGRESSION + SCALE & SMEARING IN DATA



Clear improvement in Z(ee) mass scale and resolution after applying corrections from regression, and residual scale correction.



Z(ee)

Good data/MC agreement

PERFORMANCE OF REGRESSION + SCALE & SMEARING IN DATA

Z(μμɣ) event selection cuts

Pass dimuon trigger

At least two tight ID passed muons, pT>30/10 GeV, |η| < 2.4

Oppositely charged muons

Mμμ>35 GeV

Photon |η| < 2.5 (veto EB-EE transition region)

Photon pT>20 GeV

ΔR(photon,muon)<0.8

Apply scale and smearing corrections derived Z(ee) to the photons

Estimator of energy scale difference between data corrected Z(ee) correction factors, and MC

$$s = \frac{(m^2_{\mu\mu\gamma} - m^2_{\mu\mu})}{(m^2_Z - m^2_{\mu\mu}))} - 1$$

Scale difference <0.1%

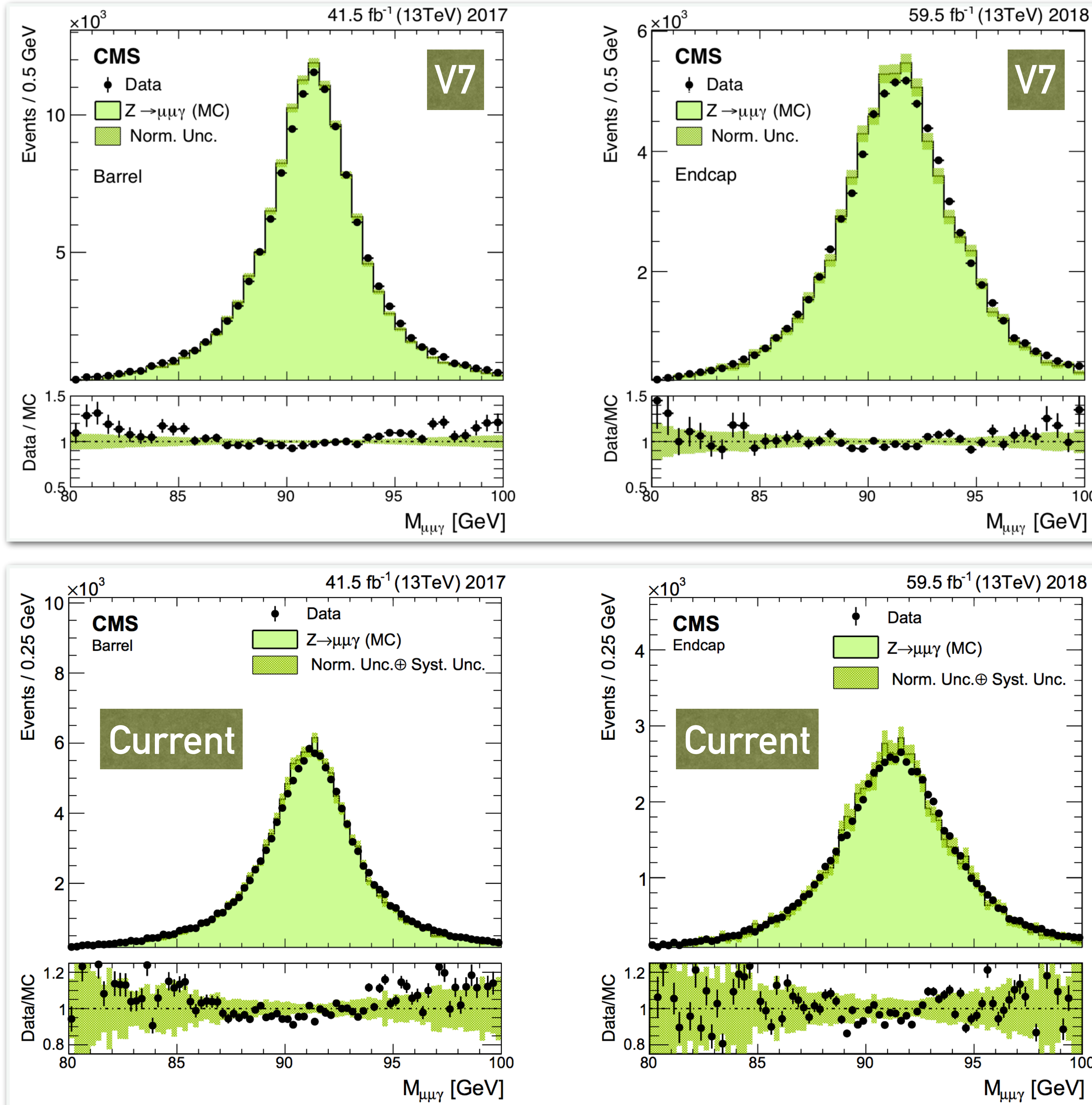
Estimation of resolution difference not done.

PERFORMANCE OF REGRESSION + SCALE & SMEARING IN DATA

<u>Z(μμɣ)</u>

Main changes: v7 vs current

- Muon Rochester correction applied in 2017 plot
- Finer binning, similar to Z(ee) plots
- Systematics now included



IDENTIFICATION EFFICIENCY AND SCALE FACTORS-I

Cut-based and BDT-based generic IDs for electron and photon. Several WPs.

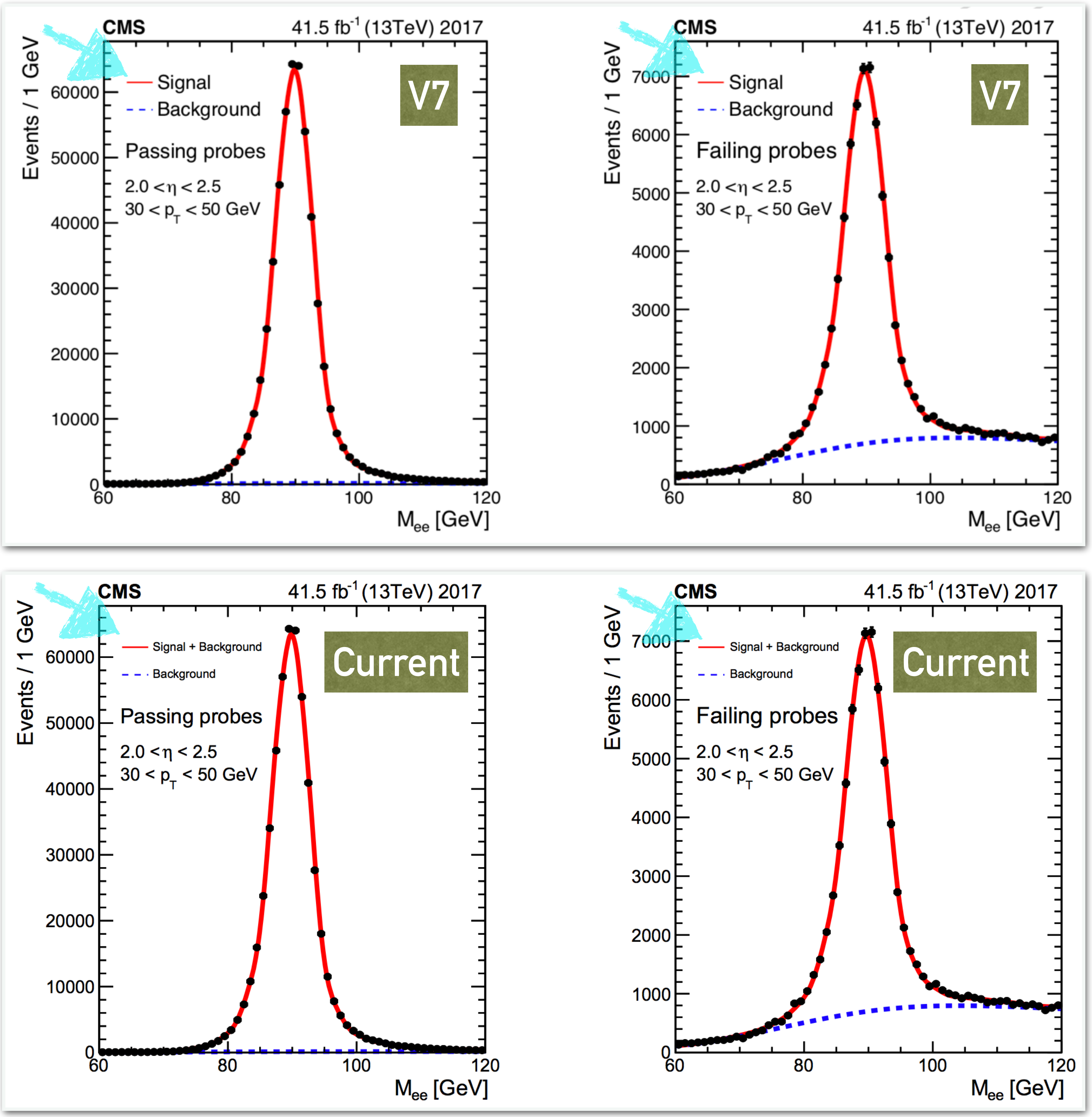
Cut-based ID for high energy electrons (HEEP ID)

ID (also reco and trigger) efficiency measured using Z(ee) tag-and-probe

Signal+background fit for (tag+passing probe) and (tag+failing probe)

New plots added after pre-approval

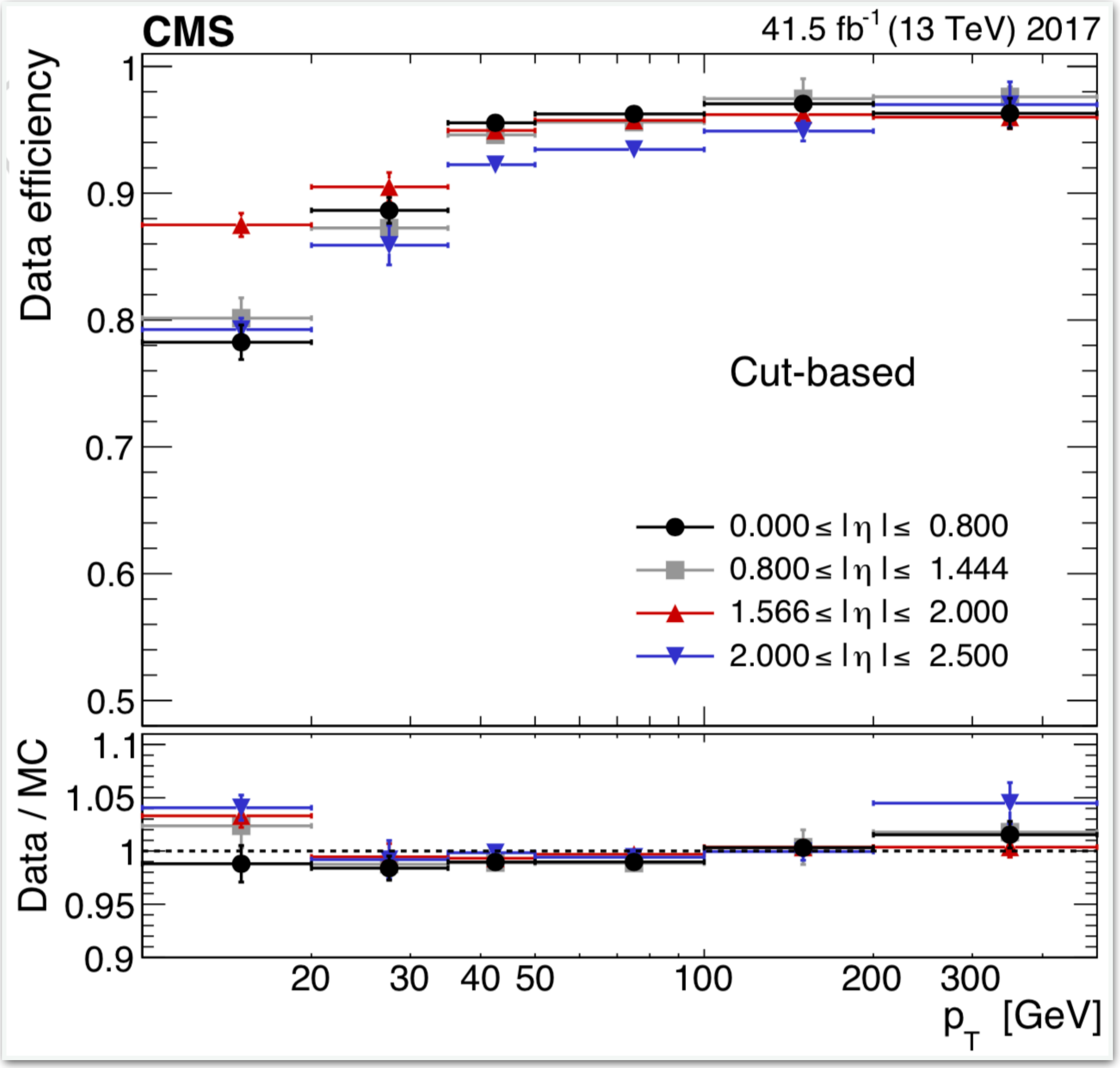
Plot label updated compared to frozen v7



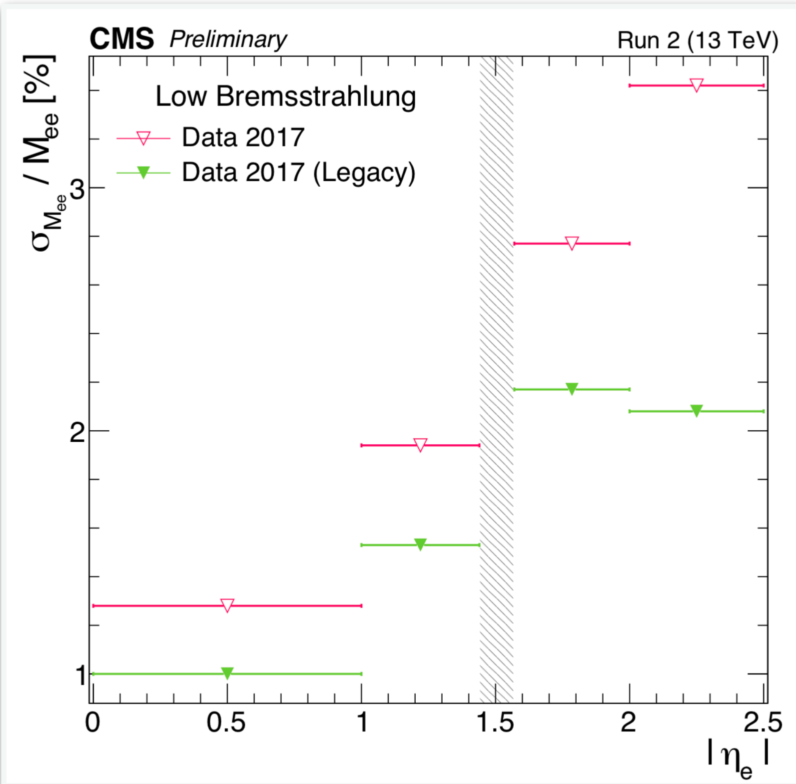
IDENTIFICATION EFFICIENCY AND SCALE FACTORS-II

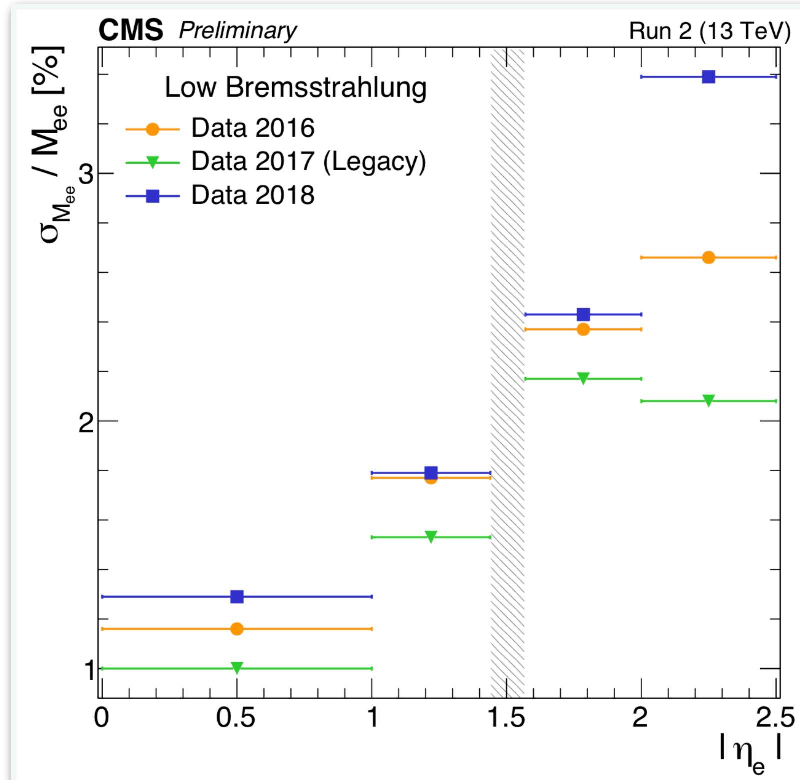
Efficiency in data and data/MC SF shown for cut based medium WP for electron.

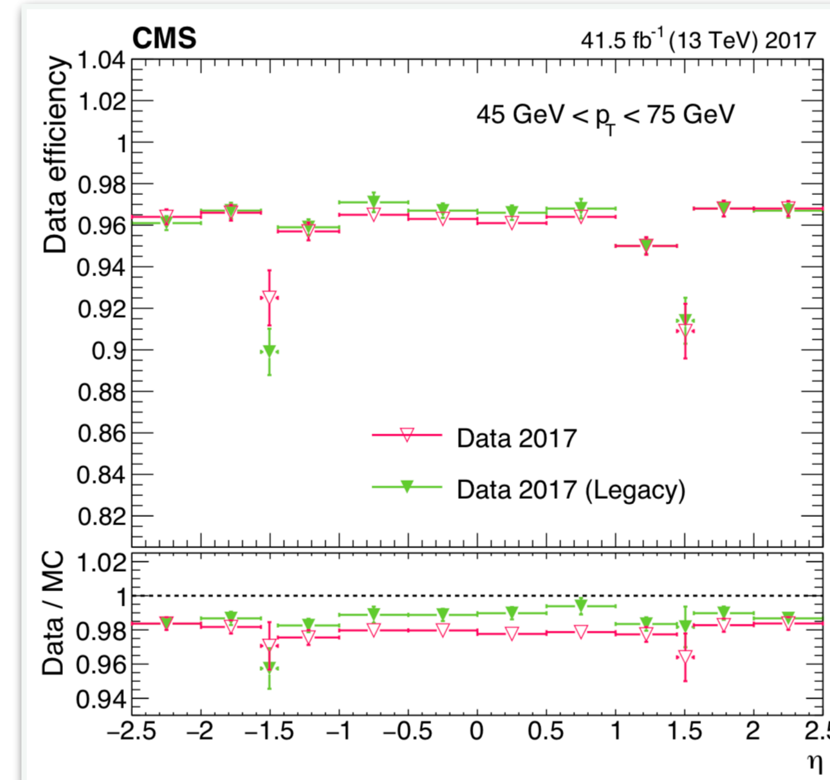
SF in this WP is within 1% in most pT, eta bins, within 5% always.



PERFORMANCE OF LEGACY 2017 DATASET







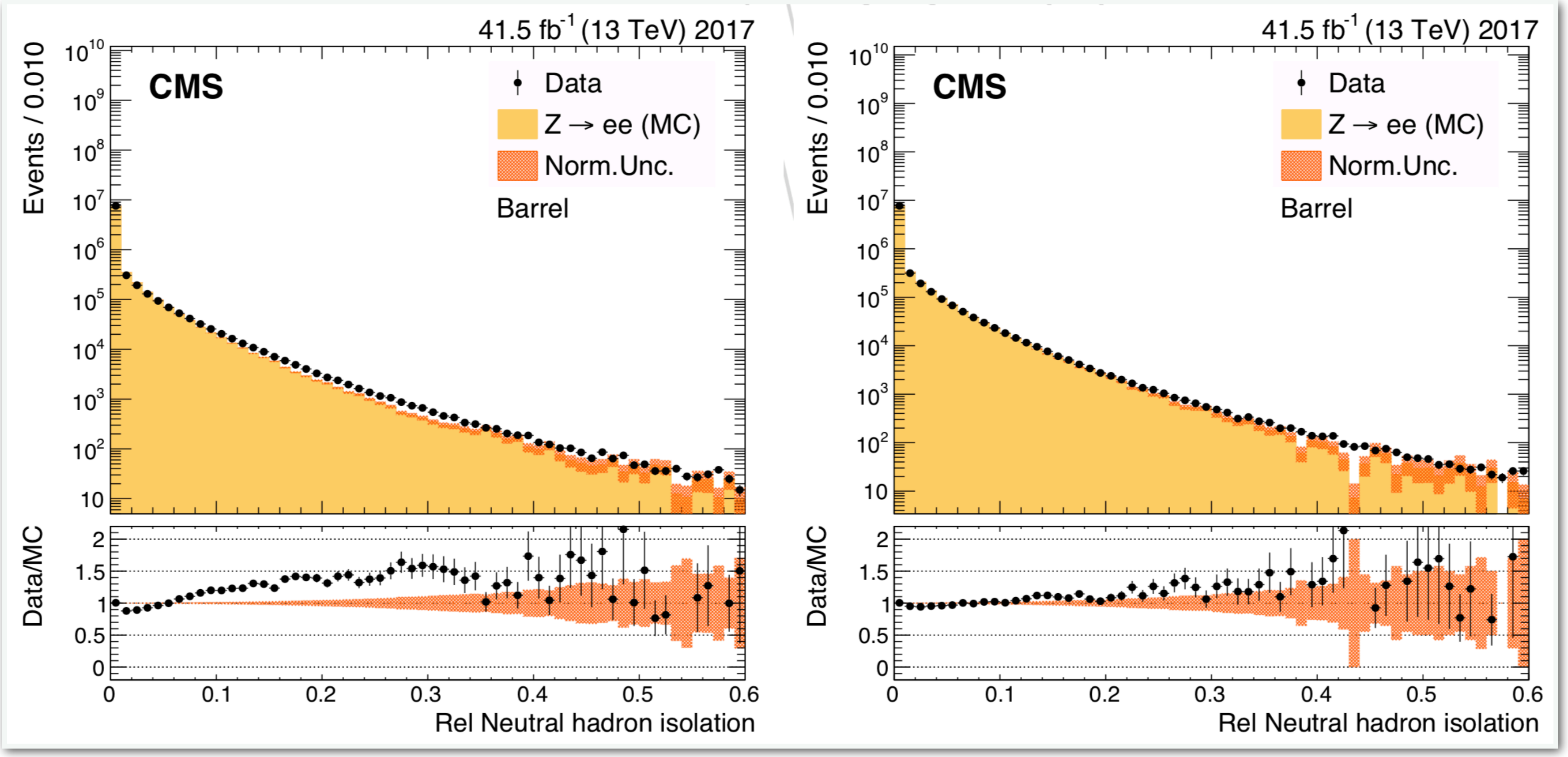
Better energy resolution in Legacy 2017 compared to EOY 2017, due to improved ECAL calibration.

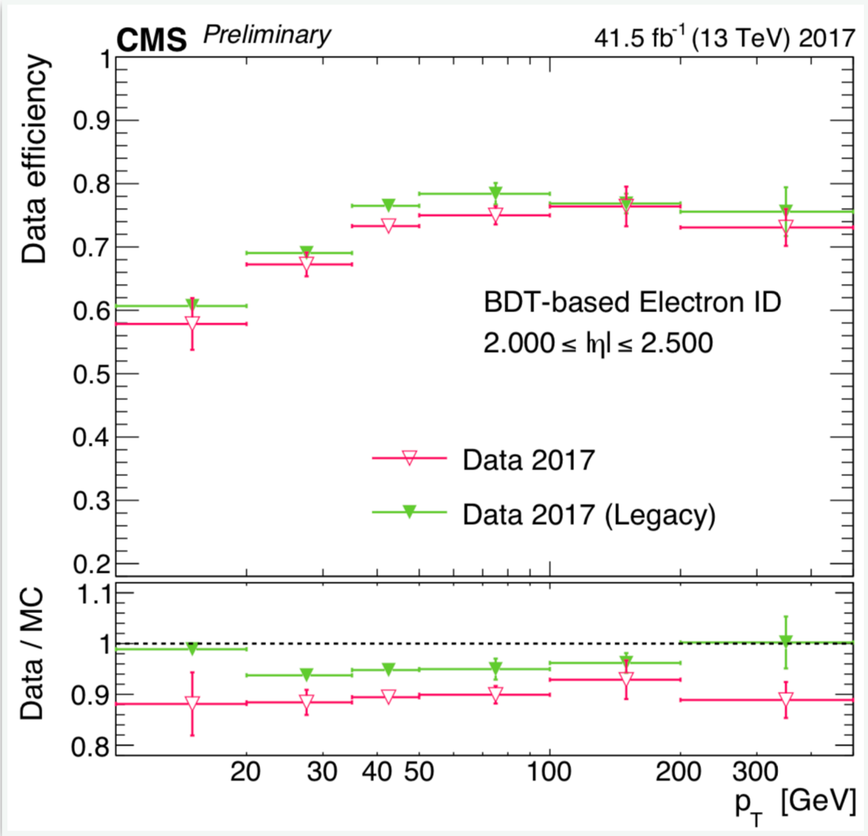
With ageing, resolution degrades, but can be recovered with improved calibration

In legacy 2017 dataset, RECO SF closer to 1

Slightly better efficiency in barrel

PERFORMANCE OF LEGACY 2017 DATASET





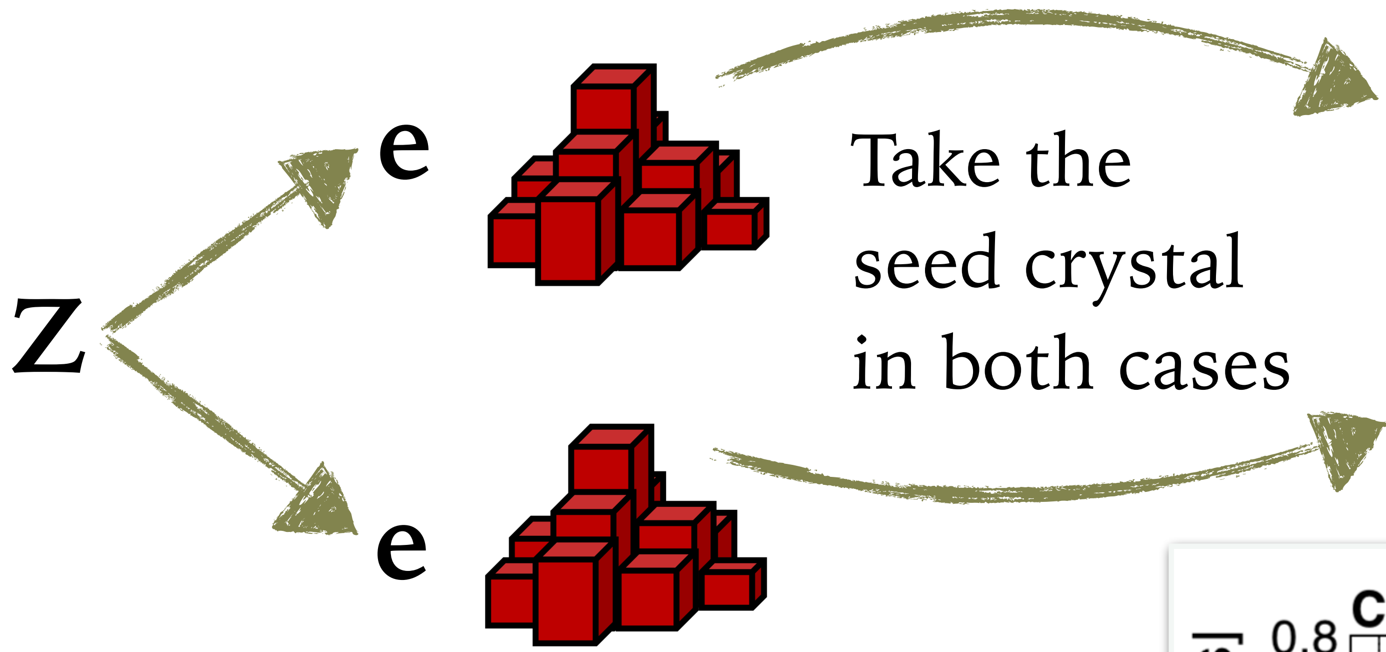
In Legacy 2017, several variables used in ID show better data/MC agreement

Example: Relative neutral hadron isolation in 2017 EOY(left) and Legacy(right)

Better data/MC agreement in ID variables, lead to data/MC SFs closer to 1

ID efficiency in data also show improvement in Legacy 2017 dataset

TIMING RESOLUTION

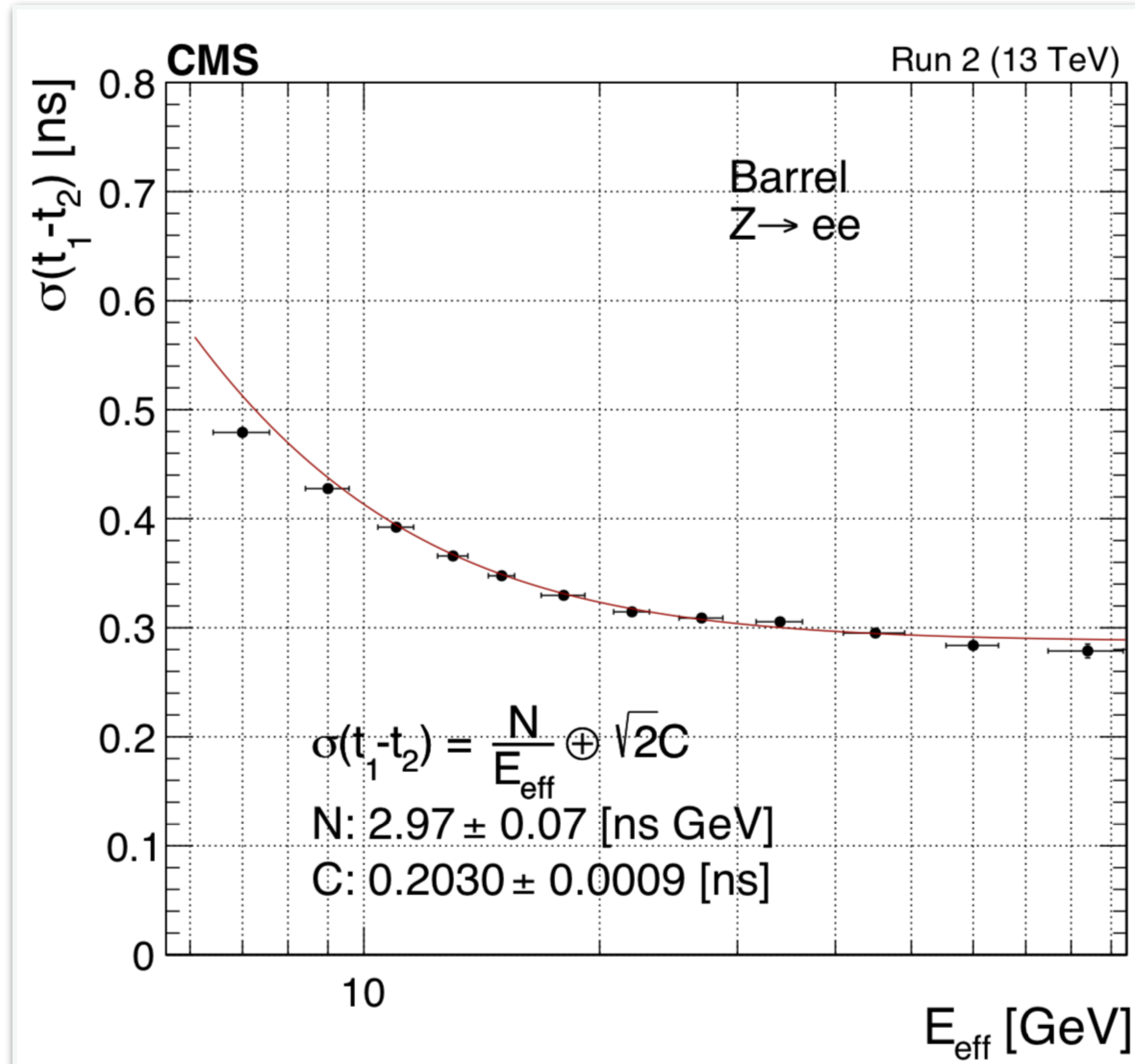


Seed energy = E1, E2

10< E1, E2 <120 GeV

Seed time = t1, t2

$$E_{\text{eff}} = \frac{E_1 E_2}{\sqrt{E_1^2 + E_2^2}}$$



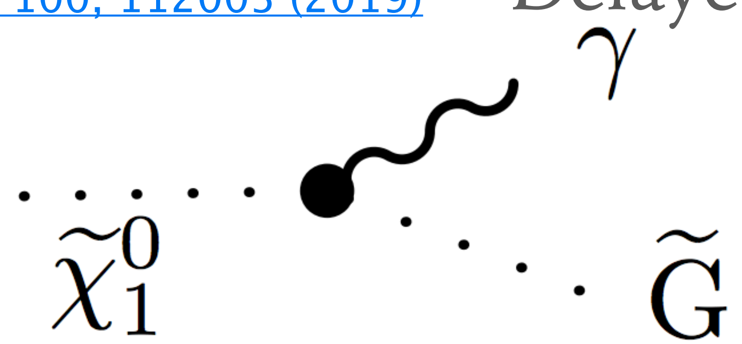
Overall resolution ~300 ps, and constant term ~200 ps.

Searches for long-lived particles profit from this

Example:

Phys. Rev. D 100, 112003 (2019)

Delayed photon



Long-lived

E/GAMMA MODIFICATIONS FOR PBPB COLLISIONS

PbPb environment is much busier than pp

- **track multiplicity ~10k in PbPb vs ~ 750 in pp**
  - Often takes excessively long reconstruction time
  - Default algorithms lead to higher fake rate, worse energy resolution

Main changes made w.r.t pp are following:

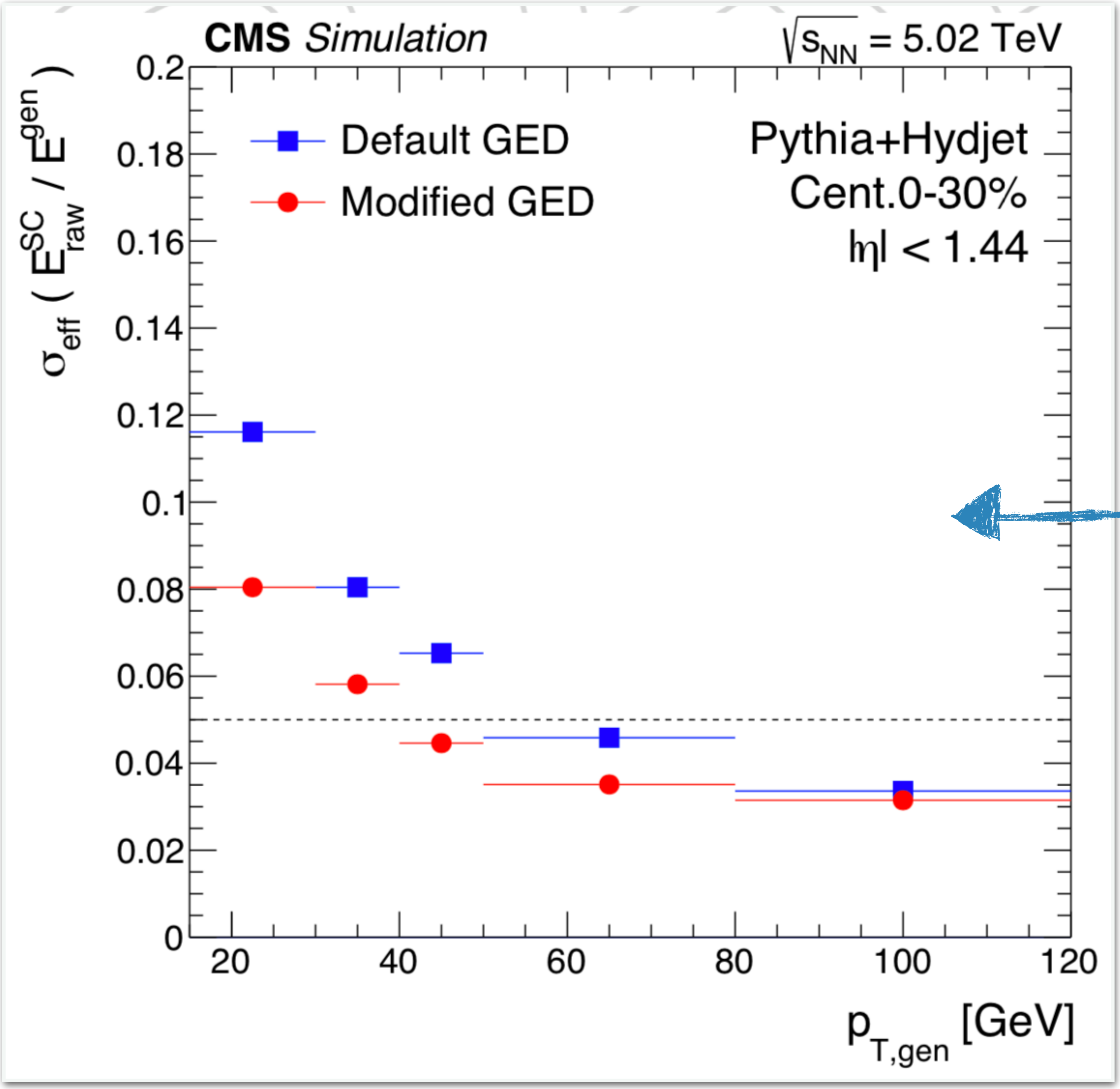
Changes made mainly to save reconstruction time

- No tracker-driven electron
- Tracking region centered around PV (instead of beamspot)
- Tighter cut on SC energy (15 GeV)
- Drop OOT recHits and photons

Changes made mainly to improve performance

- dPhi window in Mustache algorithms:
  - Dynamic → Fixed upto 0.2

IMPROVEMENT OF PERFORMANCE (PBPB)



Photon fakerate decreased

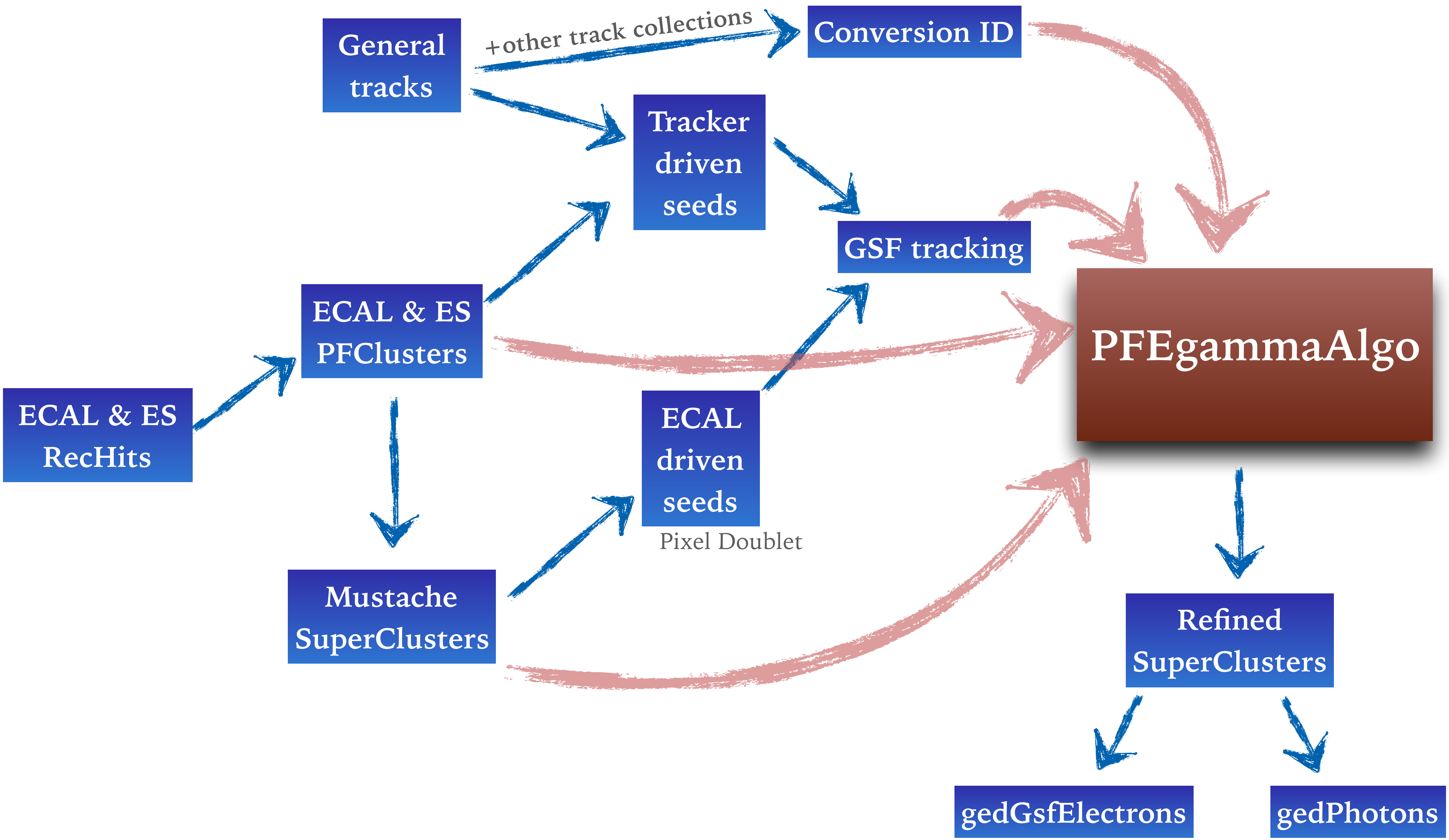
From 2.7% to 0.5% for 40<pT<60 GeV

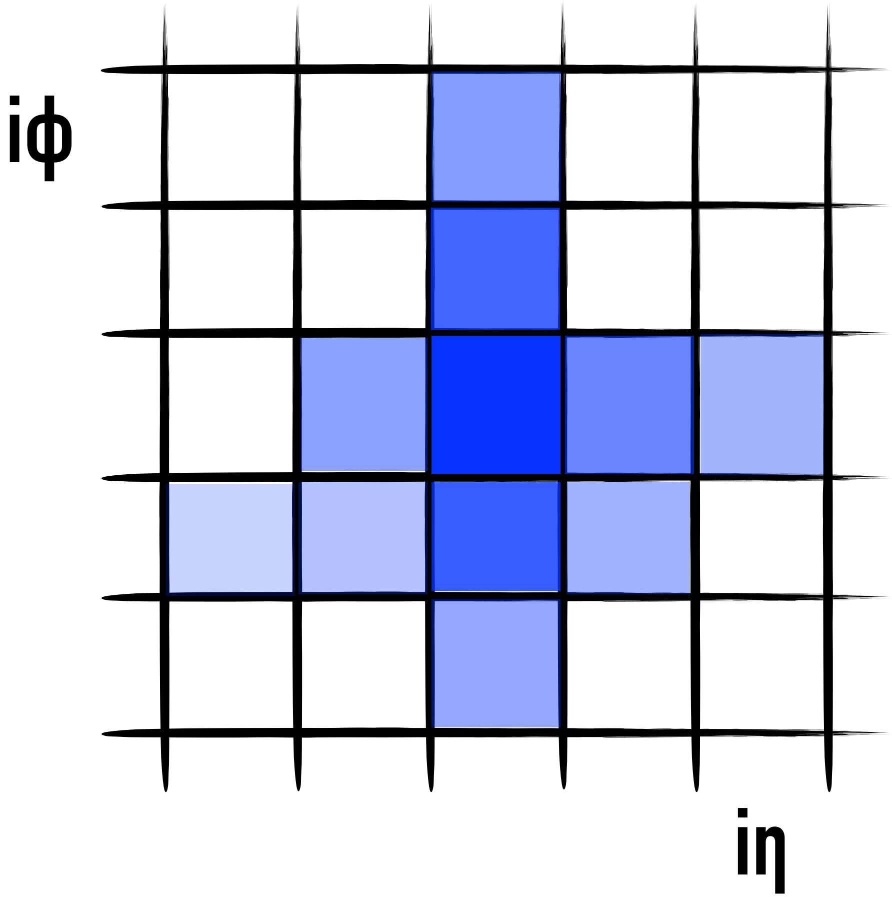
Photon energy resolution improved

E/GAMMA OFFLINE RECONSTRUCTION FLOW

Reconstruction@Offline

Putting it all together




iφ
iη