Deep learning techniques for energy clustering in the CMS Electromagnetic Calorimeter

Jin Wang[1]
On behalf of the CMS collaboration

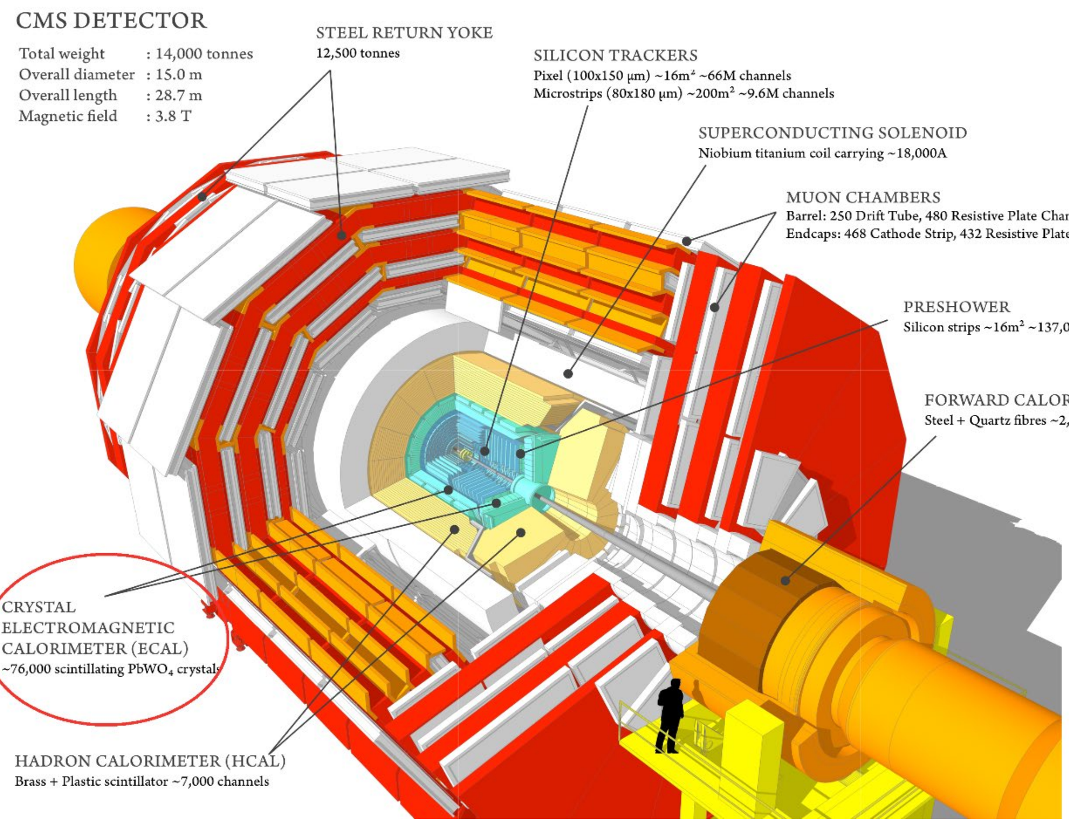
1: Institute of High Energy Physics, CAS

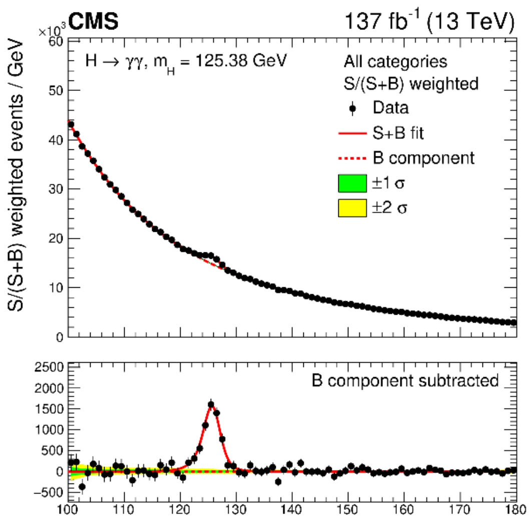
7th September 2023

IHEP ML Meeting

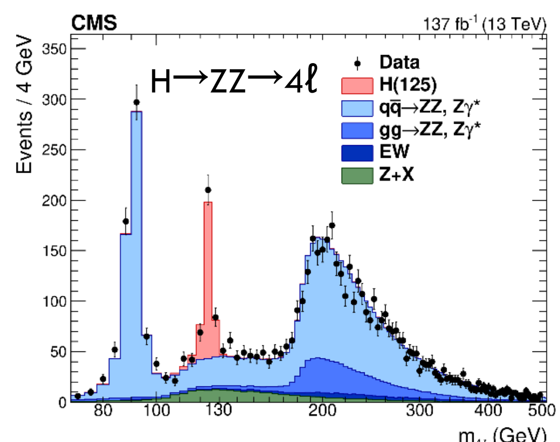
CMS Electromagnetic Calorimeter (ECAL)

- CMS is a general-purpose detector at the LHC
  - test Standard Model (SM) predictions
  - search for new physics beyond the SM





JHEP 07 (2021) 027



Eur. Phys. J. C 81 (2021) 488

- The electromagnetic calorimeter plays a crucial role in many CMS physics analyses that involve electrons/photons/jets

CMS Electromagnetic Calorimeter (ECAL)

- **ECAL: compact, homogeneous, hermetic and fine-grain crystal calorimeter**
  - designed to provide highly efficient and accurate reconstruction of photons and electrons

- 75848 lead tungstate crystals $PbWO_4$
- high density of 8.3 g/cm^3
- short radiation length 0.89 cm
- small Moliere radius 2.2 cm
- fast light emission : ~80% in ~25 ns

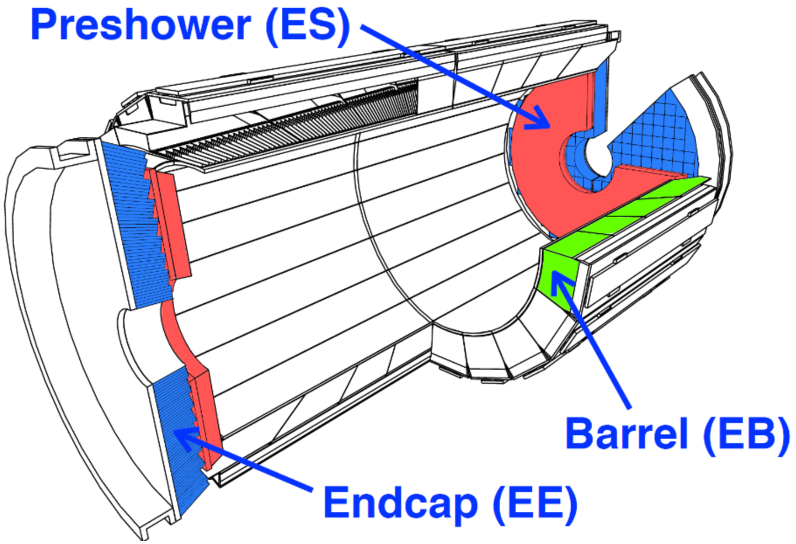
Coverage:

Barrel (EB): $|\eta| < 1.48$

Endcap (EE): $1.48 < |\eta| < 3.0$

Preshower (ES): $1.65 < |\eta| < 2.6$

(ES: discriminate between prompt photons and photons from π_0 decay)



ECAL challenges in recent LHC Run3:

- higher pileup and noise, increased exposure to radiation
  - main factors affecting the resolution of electrons and photons

ECAL electrons and photons reconstruction

- Electromagnetic particles deposit their energy over several ECAL crystals.
  - dynamic clustering algorithms used to collect the energy deposits in ECAL

- The reconstructed energy of electrons and photons is estimated by:

Cluster corrections from a regression method

the reconstructed signal amplitude

preshower energy

$$E_{e,\gamma} = F_{e,\gamma} \times [G \times \sum_i (A_i \times LC_i \times IC_i) + E_{ES}]$$

global scale factor for the ADC-to-GeV conversion

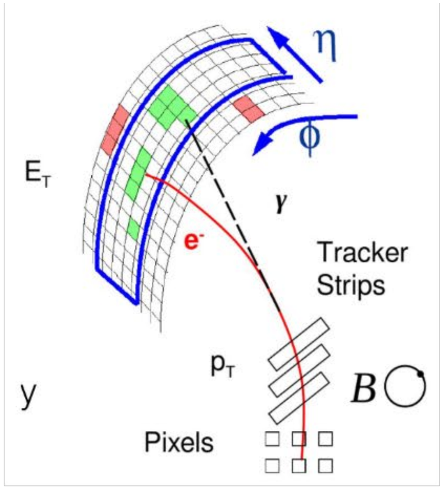
laser correction: correct for crystal transparency loss

intercalibration: equalize the channel response at same η

SuperClustering in ECAL

Rechits → Clusters → SuperClusters

Reconstructed energy deposits in the $PbWO_4$ crystals of the calorimeter (rechits) left by particles.



Rechits are gathered together around the crystal with highest deposited energy to form **clusters**.

Each cluster represents a **single particle**.
➔ Or several overlapping particles.

Due to bremsstrahlung and photon conversion before the ECAL, the individual clusters have to be combined together to form a **SuperCluster**.

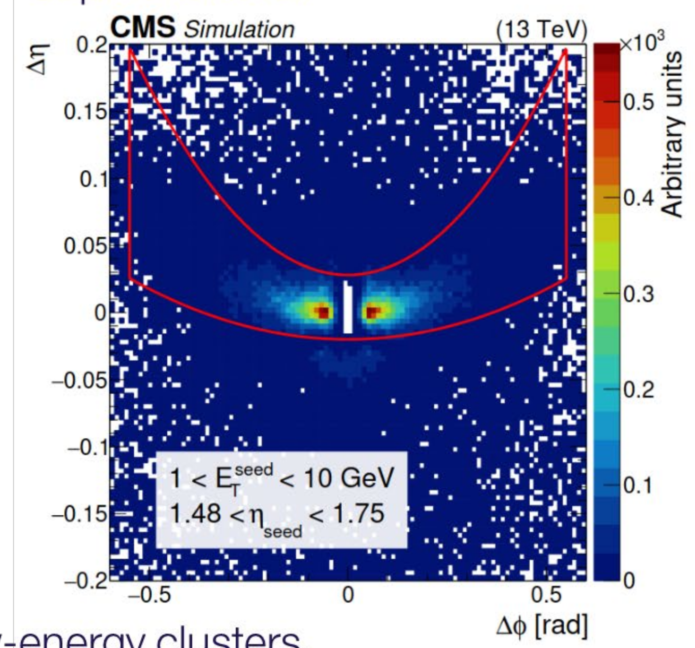
The energy of the initial particle can be reconstructed from the SuperCluster.

“Mustache” algorithm

The algorithm currently used in CMS for reconstruction of SuperClusters.

Purely **geometrical** approach:

- ➔ All the clusters falling into the specified “mustache” shape are considered as part of the SuperCluster. The size of the area depends on energy and position of the seed.
- ➔ “Mustache” shape due to the CMS magnetic field (spread along ϕ).



High efficiency: the algorithm is able to gather even low-energy clusters.

Downside: **suffers from pileup (PU) and noise contamination.**

Energy regression is further applied that can **correct PU and noise on average**.

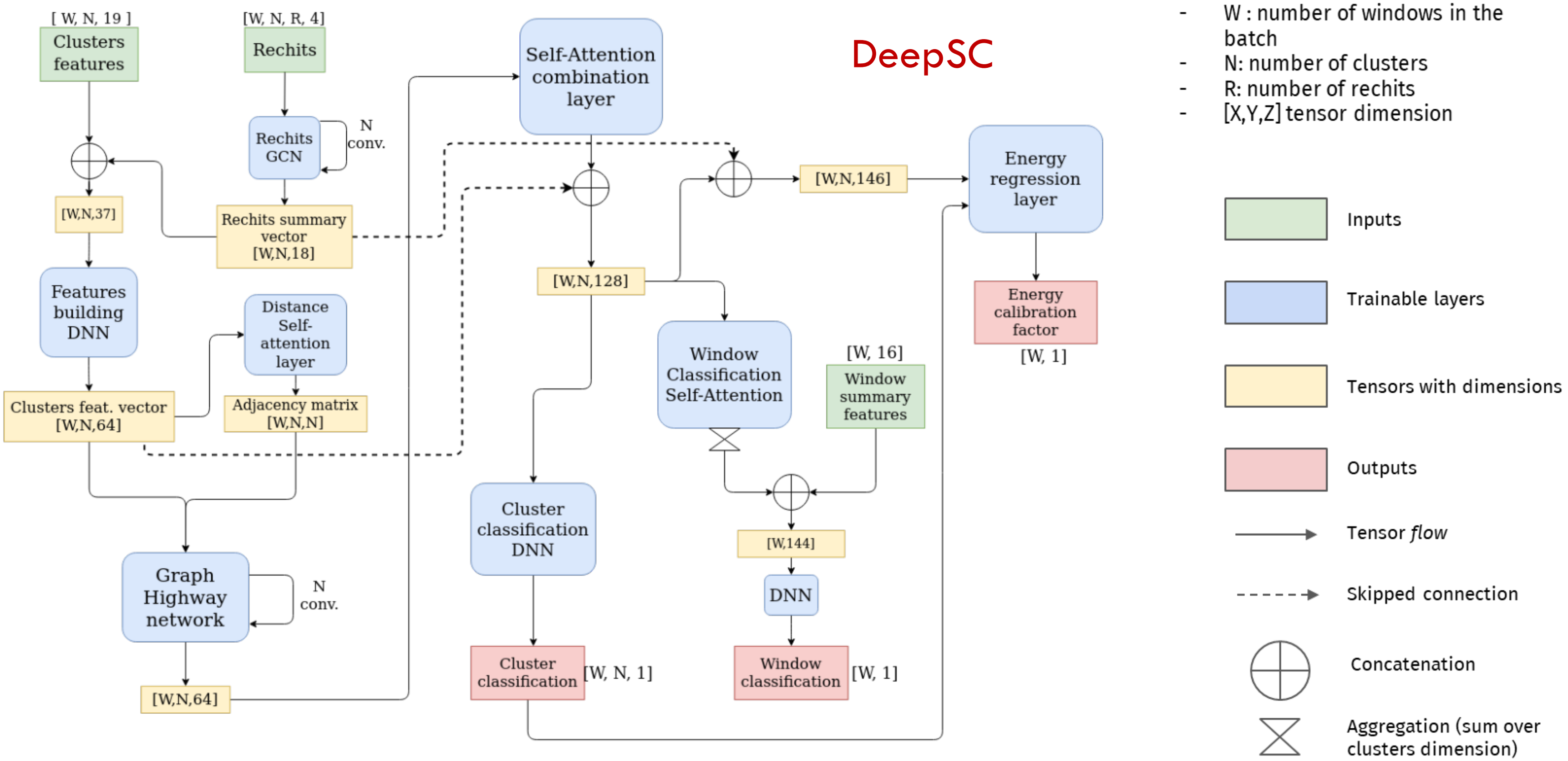
https://iopscience.iop.org/article/10.1088/1748-0221/16/05/P05014

GNN for ECAL SuperClustering

New **graph-based Machine Learning** algorithm for SuperClustering.

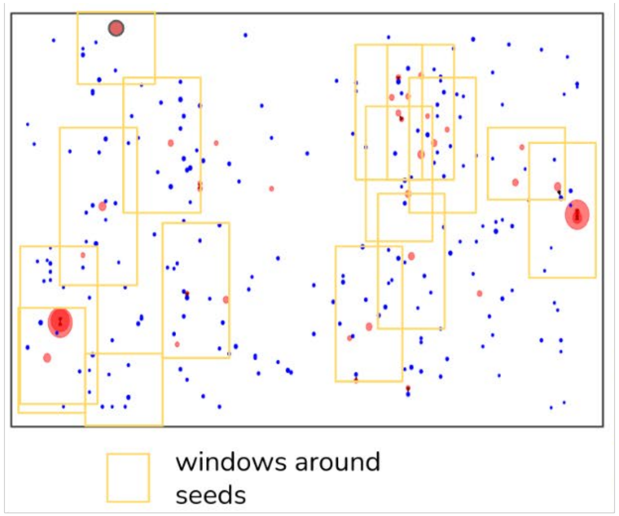
➔ **Maintains the efficiency while improving PU and noise rejection.**

➔ Graph NN are able to **aggregate the information between the neighbors.**



arxiv 2204.10277

Dataset, inputs and outputs



Dataset for the training:

- Electrons and photons generated uniformly in p_T = [1,100] GeV.
- PU uniformly distributed between [55,75] interactions.
- Windows opened around all the clusters with E_T > 1 GeV (**seeds**).

Model Input: **Cluster information** *(E, E_T, η, ϕ, z, number of crystals, …)*, **list of rechits**, **summary window features** *(max, min, mean of the crystal variables)*.

Model Output: cluster classification (in/out of SC), particle classification, energy regression.

➔ Same network to identify the **flavor of the particle**.

➔ Extra dataset: sample containing **jets**.

➔ Goal: classify jets/electrons/photons.

➔ **Transfer Learning** was used to re-train only the ID part of the network to avoid performance degradation for electrons/photons.

arxiv 2204.10277

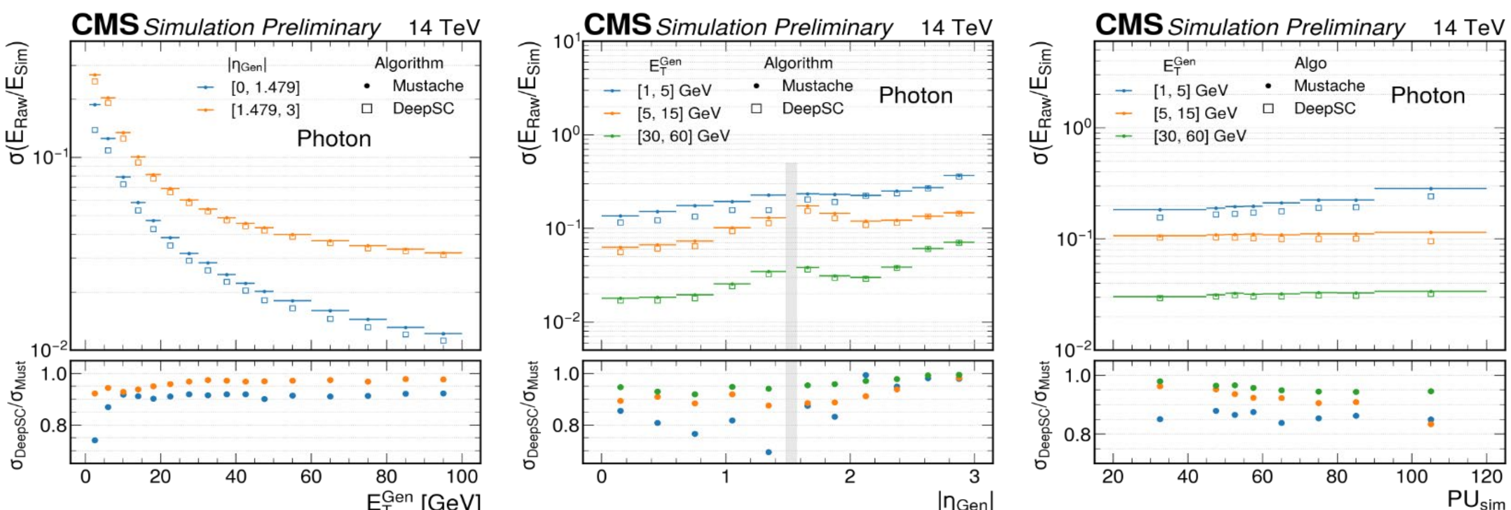
Results: Energy Resolution

Resolution of the reconstructed uncorrected **SuperCluster energy** (E_{Raw}) divided by the **true energy deposits** in ECAL (E_{Sim}) versus:

- the transverse energy of the gen-level particle E_T^{Gen} (left)
- the gen-level particle position $|\eta_{Gen}|$ (center)
- the number of simulated PU interactions (right)

The resolution is computed as half of the difference between the 84% quantile and the 16% quantile (one σ) of the E_{Raw}/E_{Sim} distribution in each bin.
The lower panel shows the ratio of the resolution of the two algorithms:

$$\sigma_{DeepSC} / \sigma_{Mustache}$$

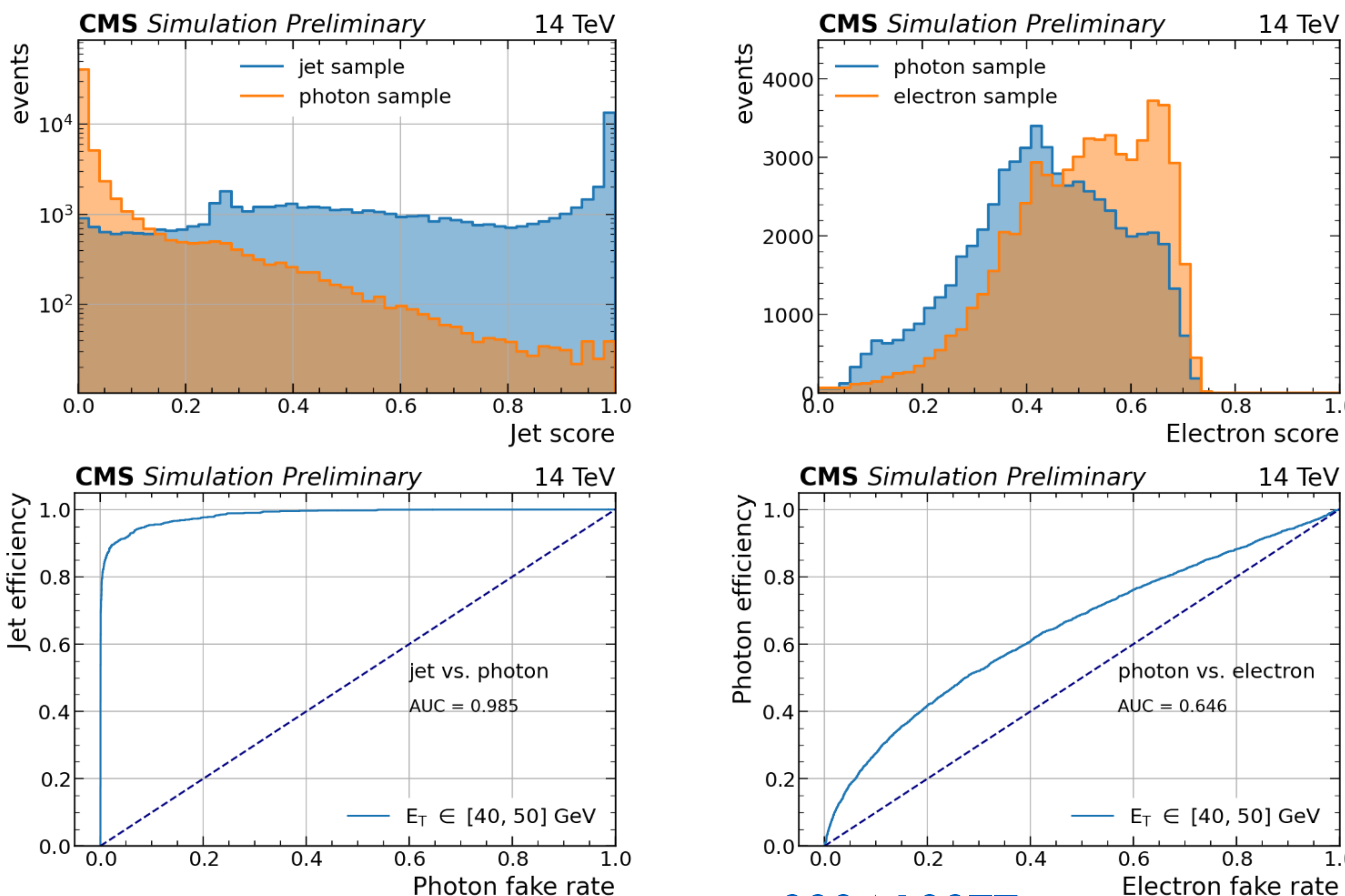


arxiv 2204.10277

Significantly improved resolution, particularly for low E_T signals and at high PU.

Results: Particle Classification

- ➔ Particle classification performance (DeepSC model) for jet vs. photon (left) and photon vs. electron samples (right).
- ➔ Only ECAL variables are used.
- ➔ **High performance for jet vs. photon discrimination.**



[arxiv 2204.10277](https://arxiv.org/abs/2204.10277)

Summary

- New calorimeter reconstruction algorithm based on Graph Neural Networks, DeepSC model.
  - Outperforms the traditional approach in terms of energy resolution.
- DeepSC model is also able to perform particle identification based solely on the information from the ECAL.
  - Shows promising results for photon vs. jet discrimination.
  - Electron vs. photon discrimination can be additionally used for the cases where the track information is lost or not reconstructed.

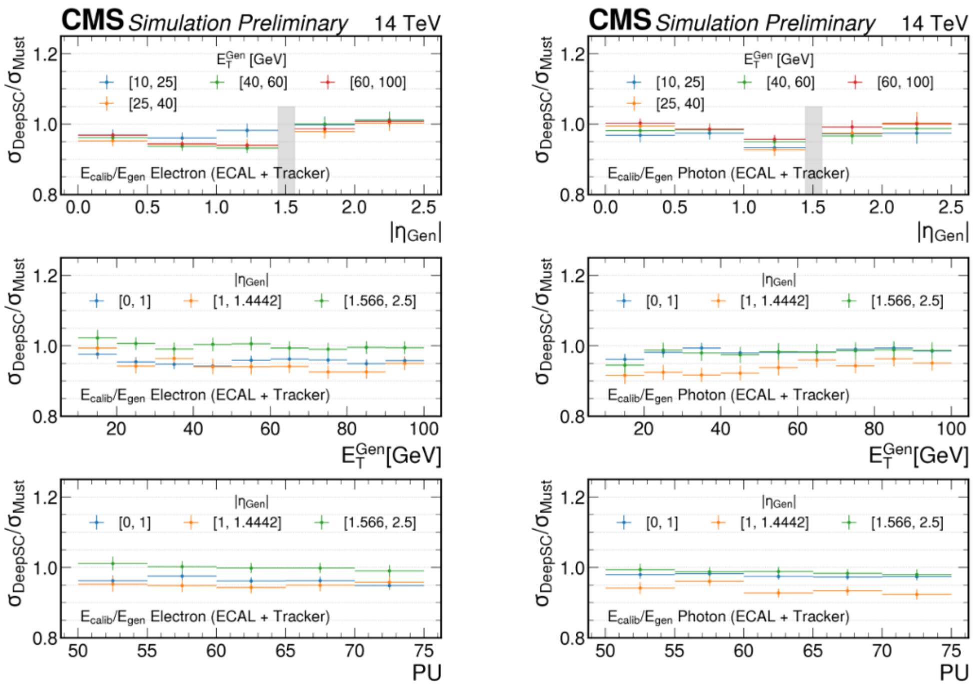
Back Up

Results: Energy Resolution

Resolution of the reconstructed uncorrected **SuperCluster energy** (E_{Raw}) divided by the **true energy deposits** in ECAL (E_{Sim}) versus:

- the gen-level particle position $|\eta_{Gen}|$ (top)
- the transverse energy of the gen-level particle E_T^{Gen} (center)
- the number of simulated PU interactions (bottom)

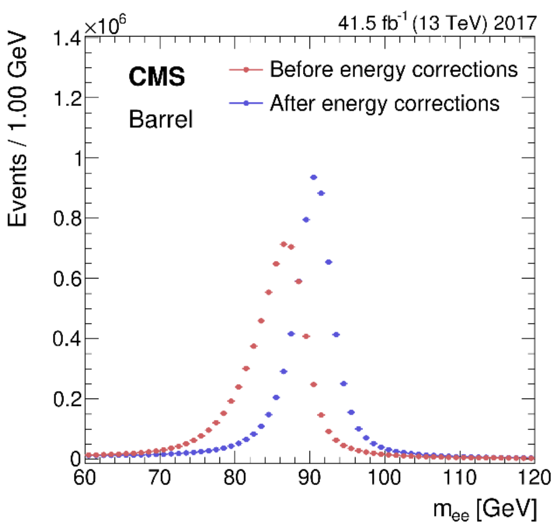
More results in CMS-DP-2022-032

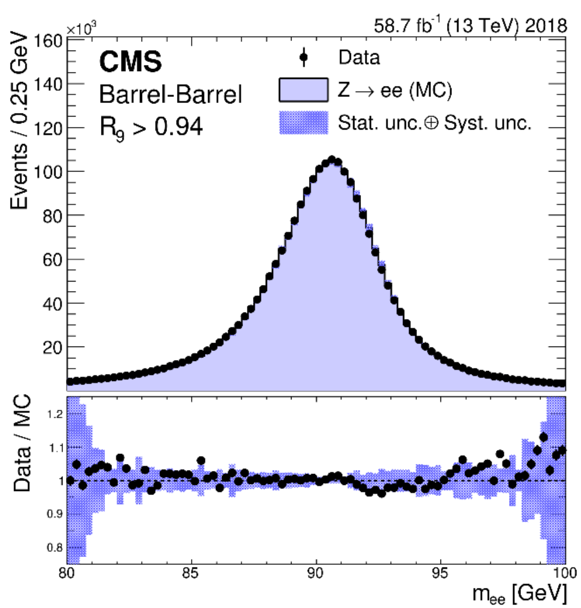


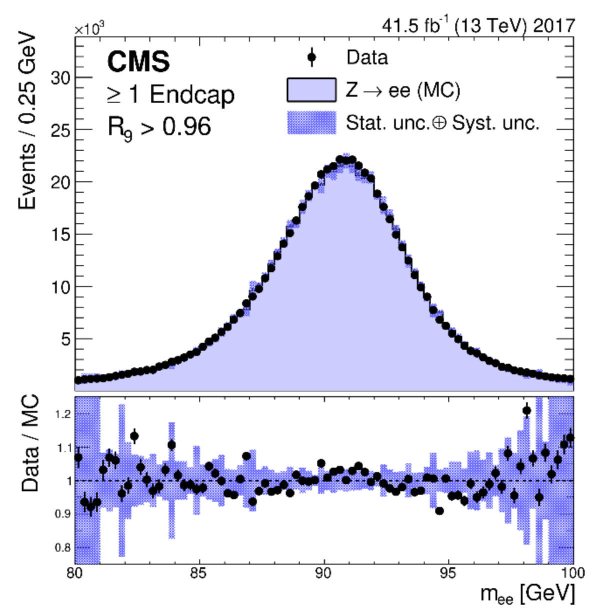
Clustering corrections

- Multivariate corrections applied to reconstruct the original deposited energy
- Energy thresholds for hits clustering re-tuned to mitigate pile-up and noise contamination
- Energy scale uncertainty smaller than 0.1 (0.3)% in the barrel (endcap) region in proton-proton collisions

JINST 16 (2021) P05014



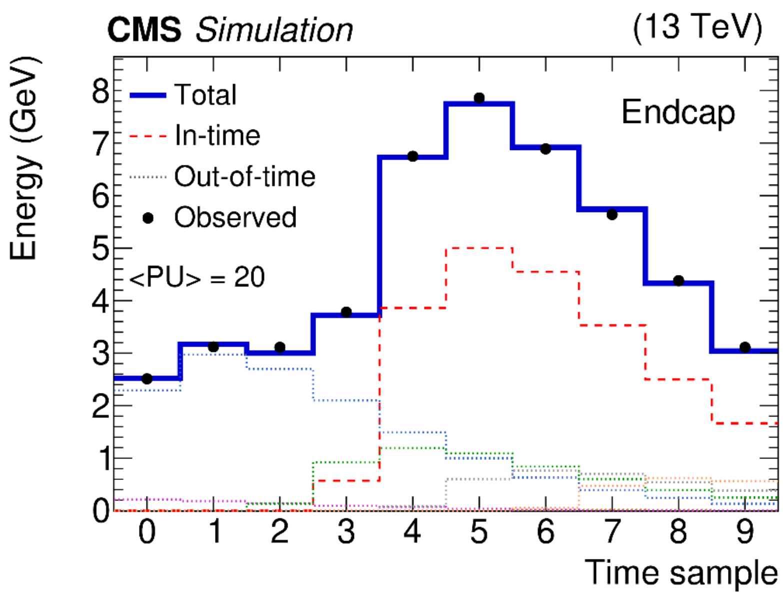


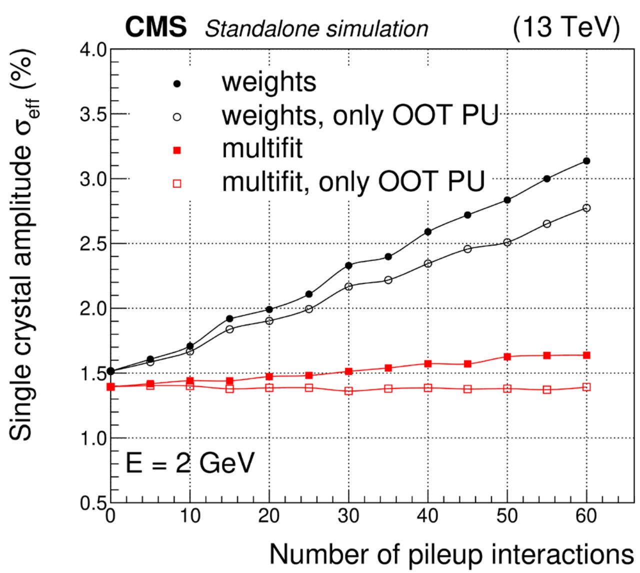


Signal amplitude reconstruction (A_i)

- 10 digitized ECAL pulse samples recorded for signal amplitude reconstruction
  - one in-time pulse and up to 9 out-of-time (OOT) pulses
  - Run 1: amplitude reconstructed from a weighted sum of samples
  - Run 2: ’multifit’ reconstruction method used to mitigate higher pileup

JINST 15 (2020) P10002



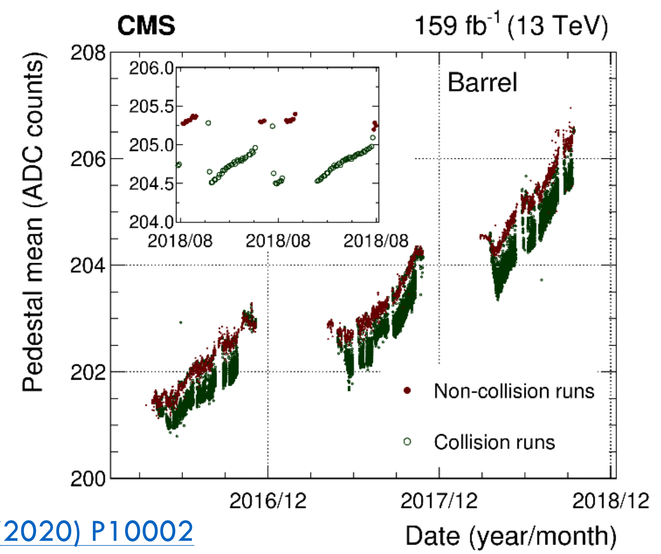


- The ’multifit’ reconstruction method is robust against pile-up increase.

Pedestal condition and timing calibration

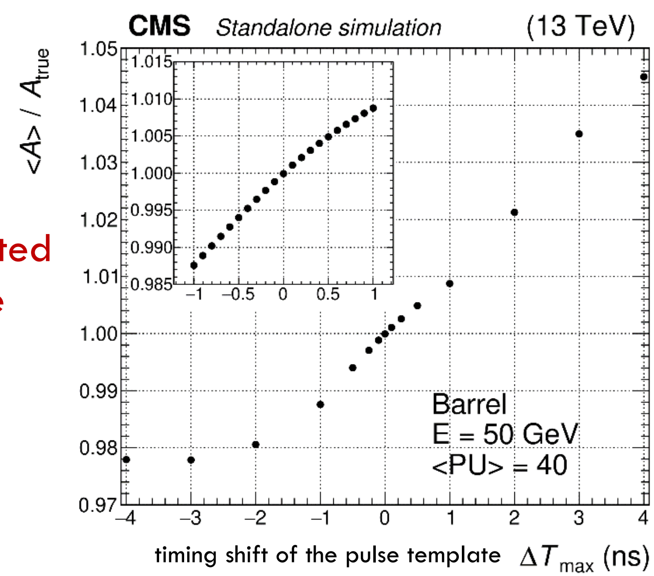
- The signal pedestal and pulse shape are inputs to the amplitude reconstruction algorithm
  - Pedestal measured from laser events every 40 minutes.
  - Time shift due to irradiation corrected every year
    - towards negative times during collisions and towards positive times during recovery

Pedestal mean over time for ECAL barrel

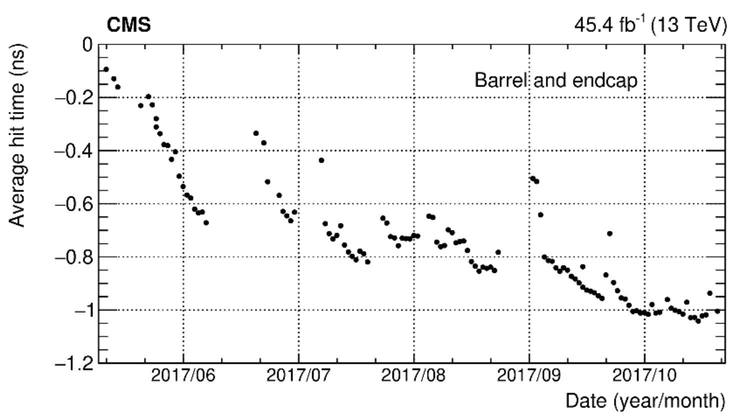


JINST 15 (2020) P10002

Good agreement between reconstructed amplitude over true amplitude



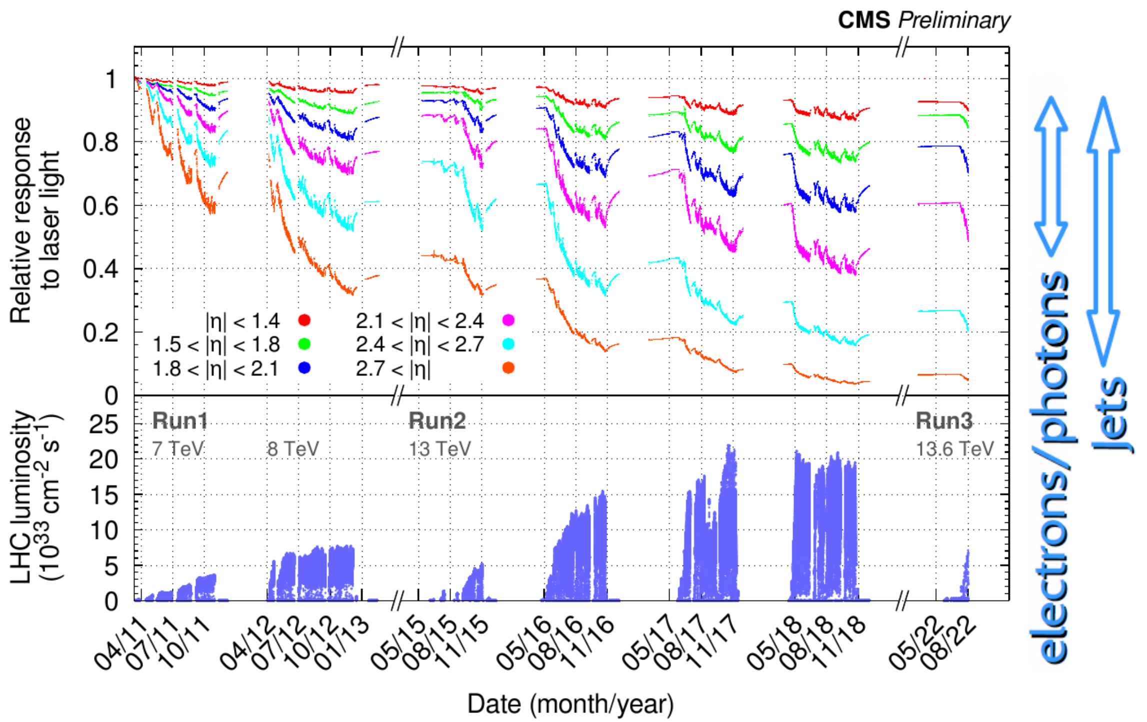
Average ECAL pulse timing in 2017



ECAL transparency loss

- ECAL channel response varies with time due to radiation-induced effects
  - crystal transparency changes over time
  - photocathode aging with accumulated charge

CMS-DP-2022-042



Transparency loss correction is crucial to maintain stable ECAL energy scale and resolution over time

Laser Correction (LC_i)

- A dedicated laser monitoring system is designed to provide corrections for transparency changes.
  - injects laser light with a wavelength of 447nm into each crystal
  - relates ECAL channel response variation to changes in the scintillation signal
  - measures the calibration point per crystal every 40 minutes
  - obtains and applies corrections within 48 hours for the prompt reconstruction

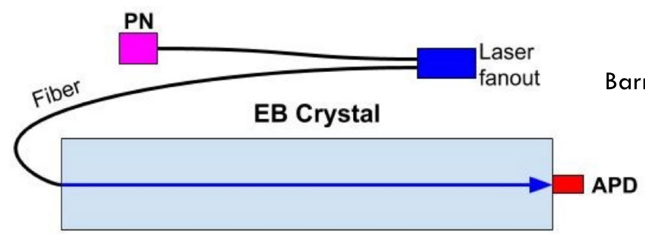
Correction for e/y scintillation

Response to injected laser

$$\frac{S(t)}{S_0} = \left(\frac{R(t)}{R_0}\right)^{\alpha}$$

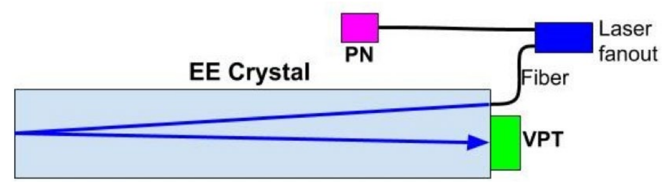
α parameter

Barrel: Avalanche Photo-Diodes (APDs)



Endcaps: Vacuum Photo-Triodes (VPTs)

Relative response = APD(VPT) / PN

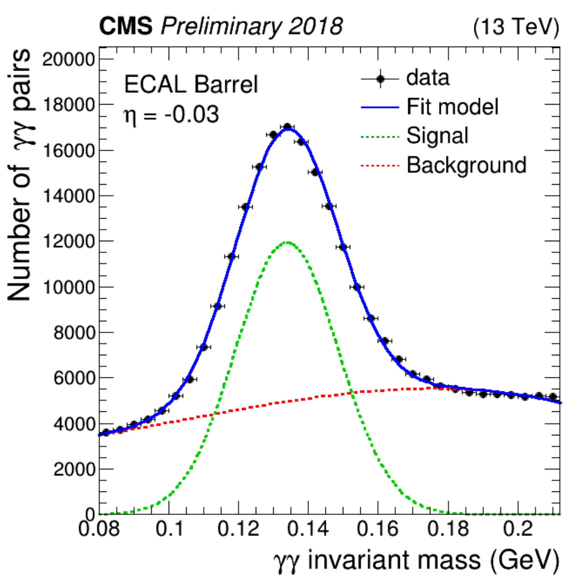


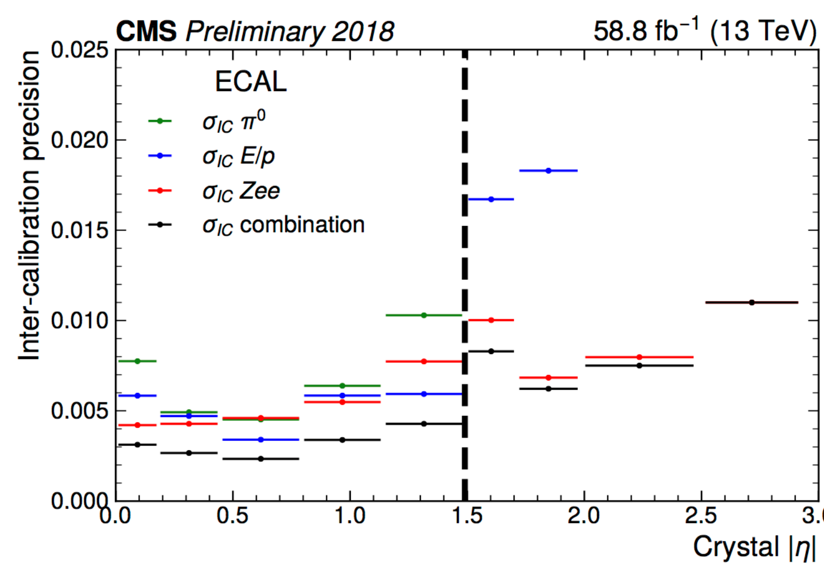
- α parameter depends on η and evolves with integrated luminosity
  - periodically re-computed to ensure energy scale stability and high resolution

Intercalibration (IC_i)

- IC: equalize the ECAL response for different crystals at the same η coordinate.
- A combination of several methods based on different physics signals
  - π^0 mass: exploit reconstructed π^0 mass with its decay of photon pairs
  - E/p: comparison of the ECAL energy to the tracker momentum for isolated electrons from W/Z boson decay
  - Zee: exploit the invariant mass reconstructed with electron pairs from Z decays

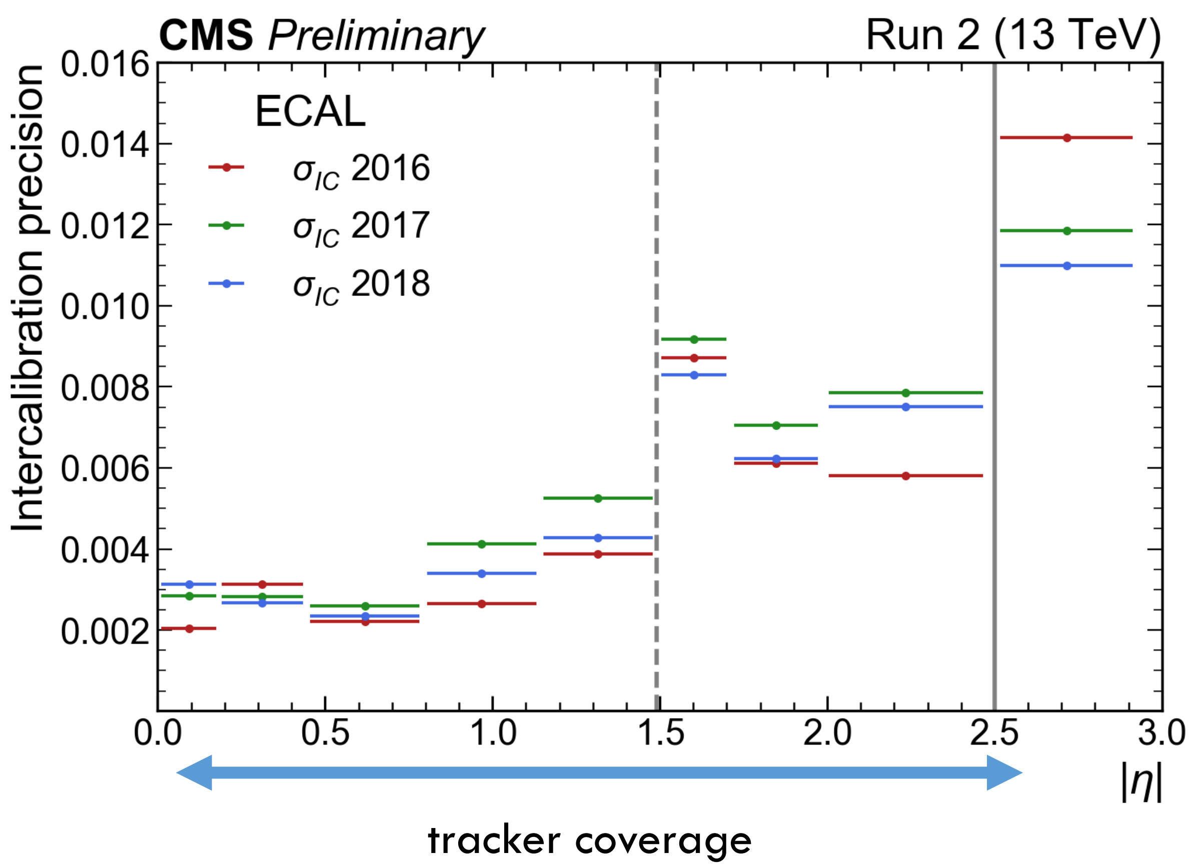
CMS-DP2019-038





Intercalibration precision

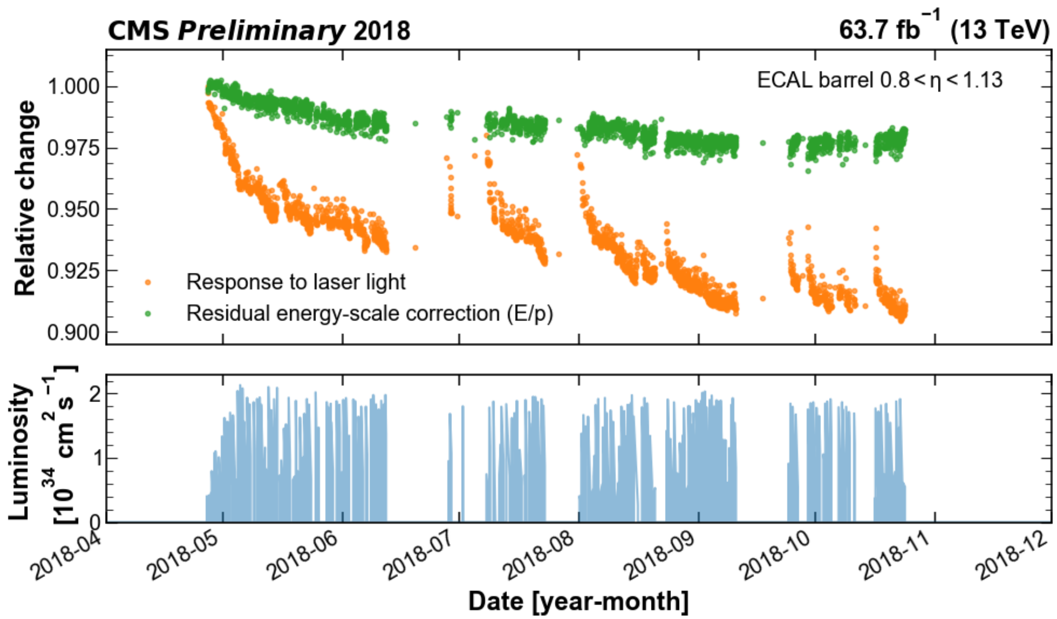
- Final intercalibration combines different methods by weighting their respective precision
  - precision evaluated with the relative energy resolution of Zee



CMS-DP-2020-021

IC reaches very good precision
- <0.5% at barrel region
- <1% at endcap region

Laser correction with E/p residual correction



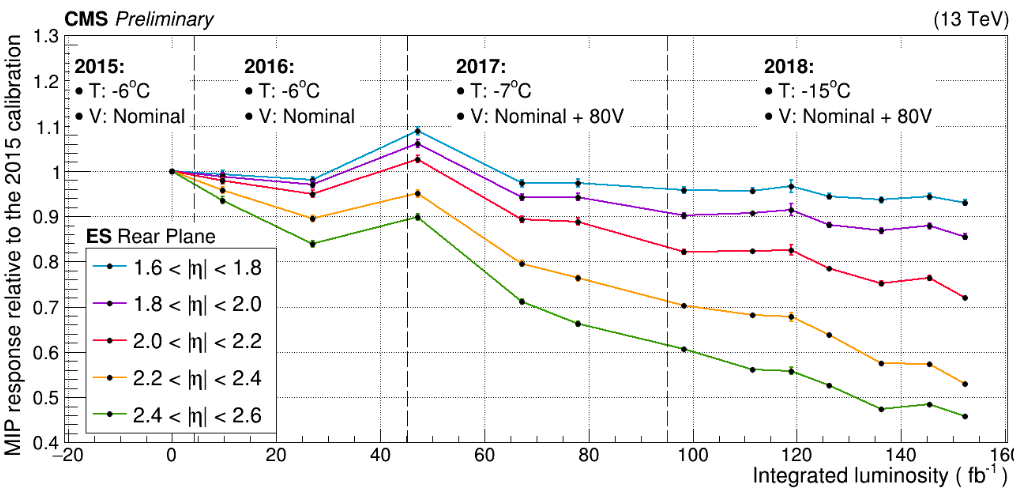
CMS-DP-2019/030

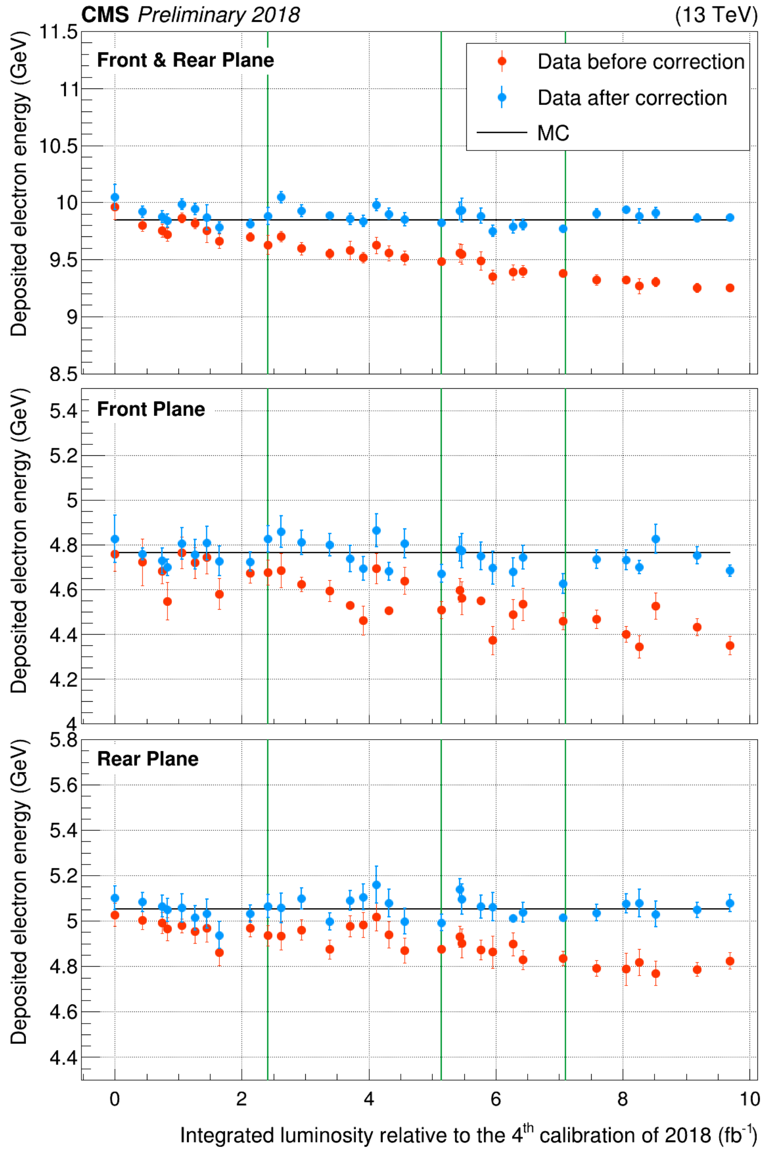
- Orange: relative response variations to laser light injected in the ECAL crystals
- Green: the residual energy-scale correction after the application of the laser corrections
  - correction needed due to a drift of the response of the PN diode used in the laser-based calibration system, determined by comparison with the tracker-measured momentum of electrons from W/Z bosons (E/p ratio)
  - a few percent variation the whole year and independent of instantaneous luminosity

Preshower (E_{ES}) Calibration

- Preshower calibrated using minimum ionizing particles (MIPs)
  - channel by channel calibration
  - special runs taken for calibration every 10 fb^{-1}

[CMS-DP-2019-038](CMS-DP-2019-038)

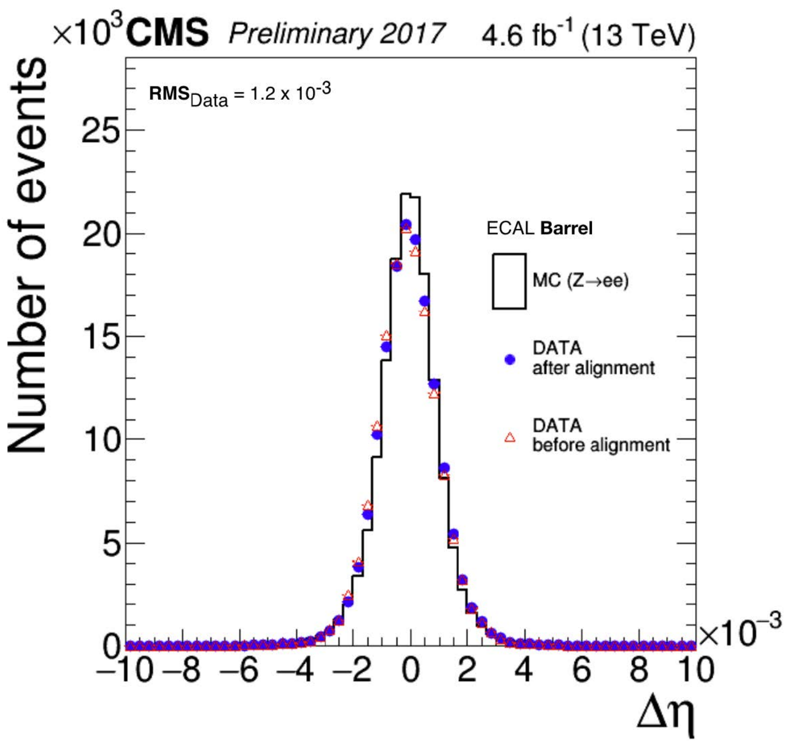


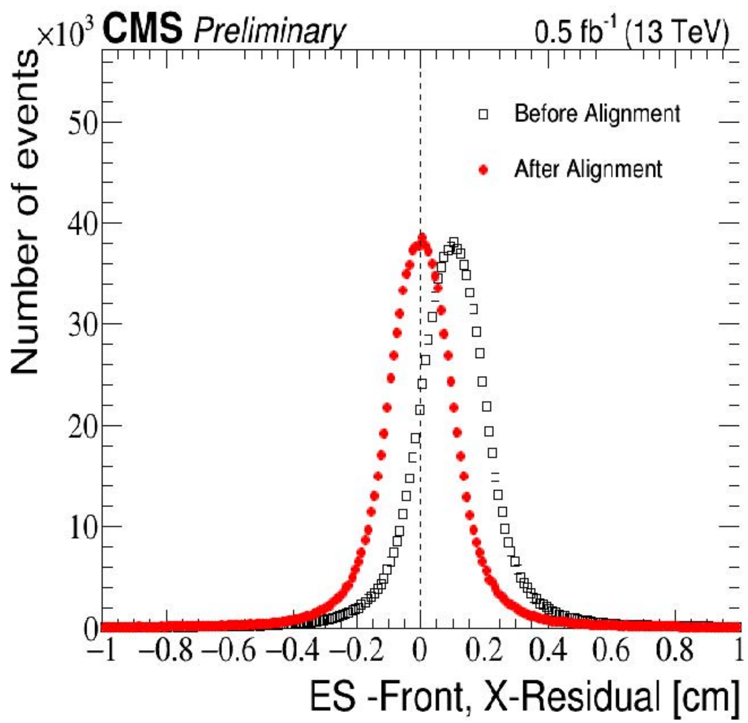


  - correction computed by minimizing the χ^2 value between the energy distribution of data and MC using Z→ee events

- Measured energy of ES cluster is stabilized by applying the correction.

ECAL and Preshower (ES) alignment w.r.t the tracker in Run 2

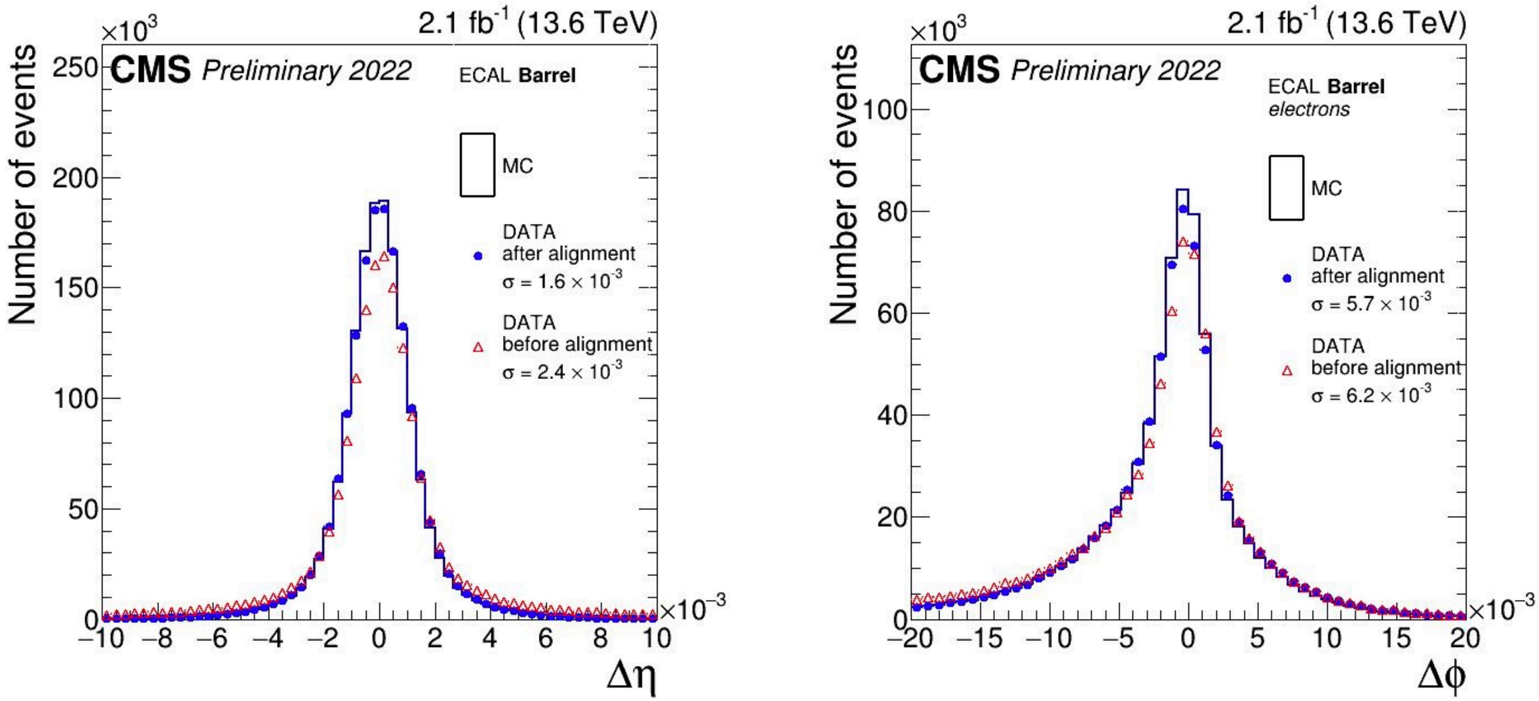




CMS-DP2018-015

- ECAL-tracker alignment: minimizing the difference in the η/φ between the ECAL super-cluster and the extrapolated track position
- ES-tracker: a minimization of the expected hit in the ES and the extrapolated track

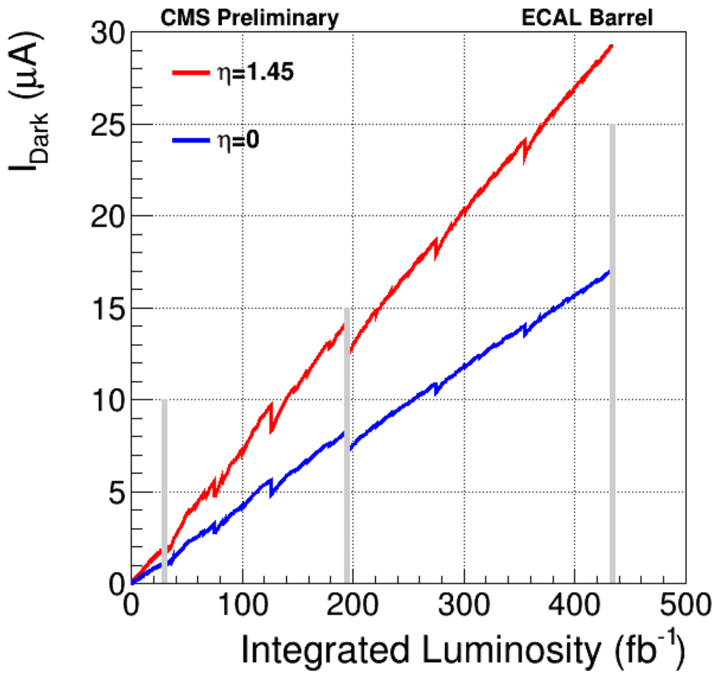
ECAL alignment w.r.t. tracker in Run 3

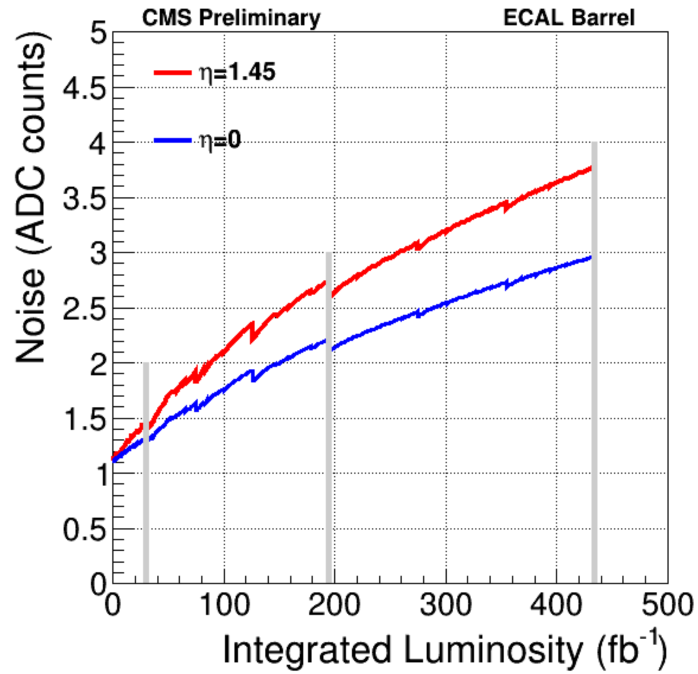


CMS-DP2023-001

- Relative alignment of ECAL crystals with the tracker detector using $Z \rightarrow e^+e^-$ events
  - For each e$^+$ and e$^-$, the distance between its track extrapolated from the tracker and its ECAL supercluster (SC) position is minimized along η and φ directions

Evolving noise in ECAL

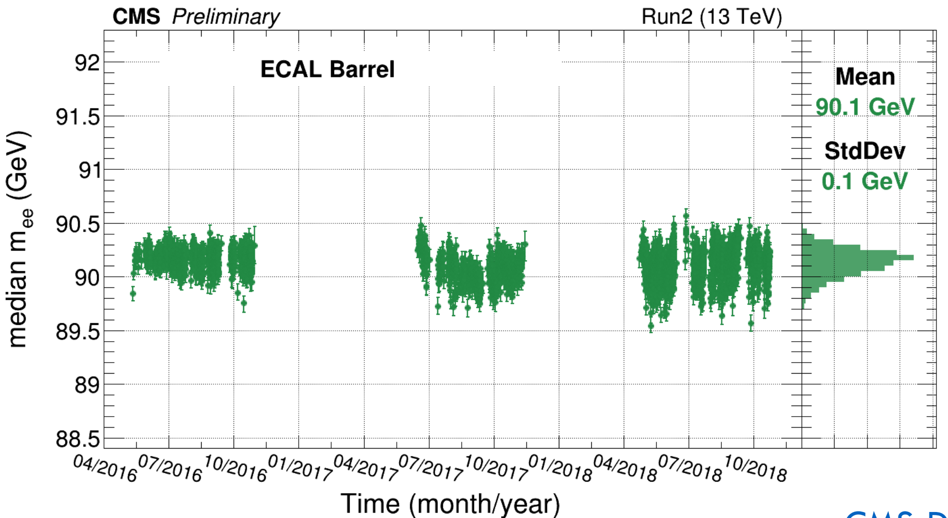


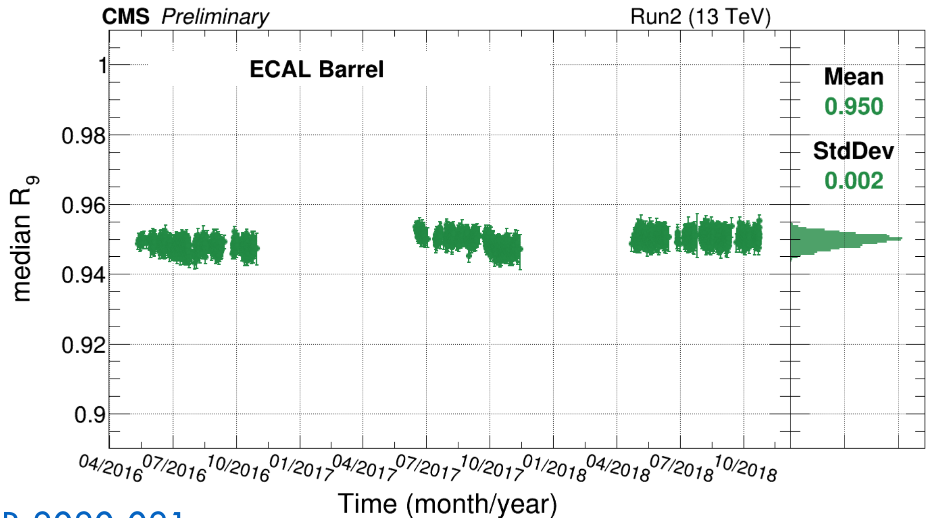


- The leakage current in the ECAL Barrel APDs increases due to radiation-induced hadron fluence.
- The noise increases due to the increase of the APD leakage current.

ECAL performance in Run 2

- **ECAL response is stable over time after corrections**
  - validated with Z→ee events

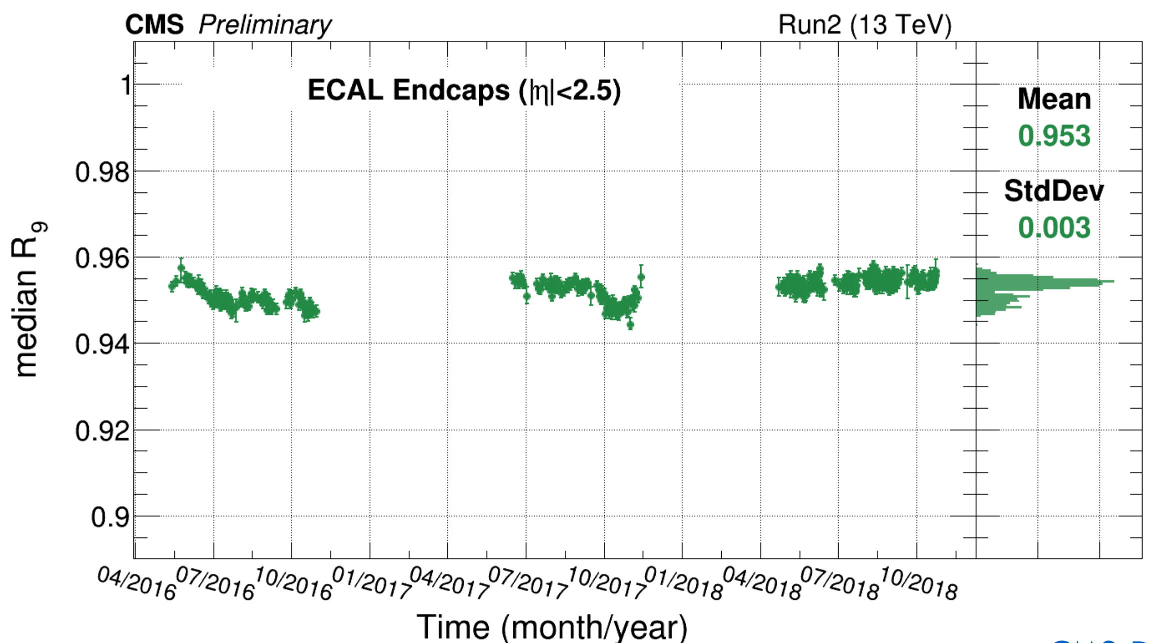




CMS-DP-2020-021

- energy scale stable at ~1% level across 3 years
- shower shape variable (R_9) also stable over time with spread <<1%
  - R_9: ratio of the energy deposit in the 3x3 crystal matrix around the seed crystal to that in the supercluster
  - important variable for the electron and photon identification

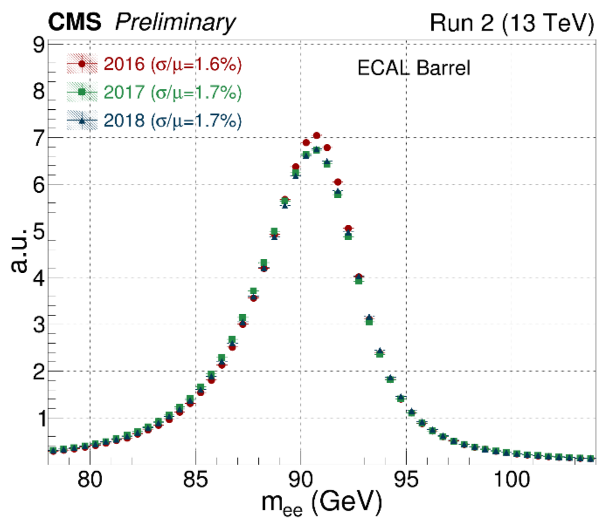
ECAL time stability in endcaps

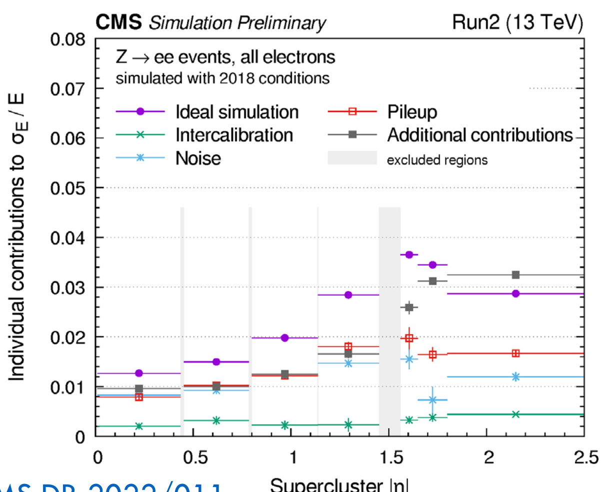


CMS-DP-2020-021

The shower shape is measured by the variable R_9, defined as the ratio of the energy deposit in the 3x3 crystal matrix around the seed crystal to that in the supercluster. R9 is responsive to changes in pedestal and noise.

ECAL performance in Run 2



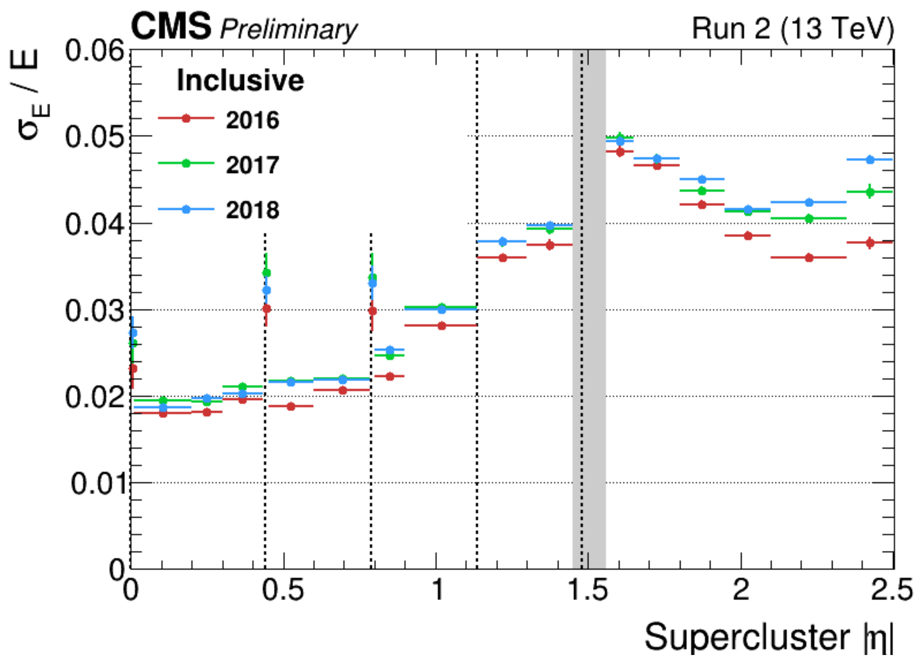


CMS DP-2022/011

Large impacts on resolution from pile-up and noise related effects

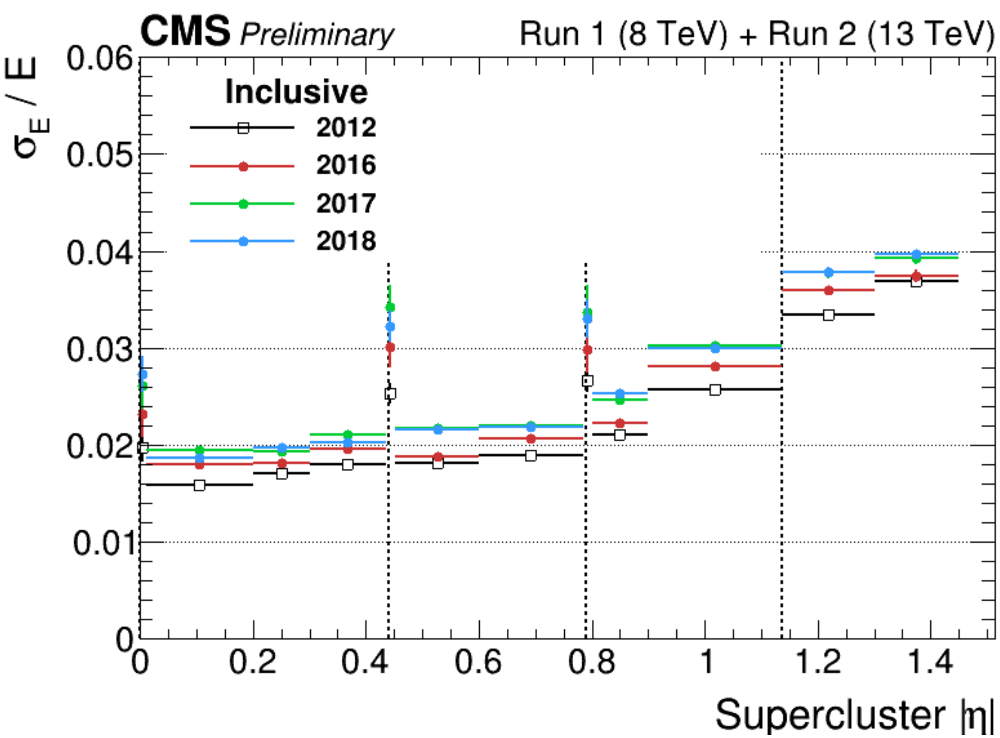
- Energy and mass resolution with ECAL calibration

CMS-DP-2020-021



- Excellent ECAL performance throughout Run 2
  - resolution at ~2% in the central, <5% elsewhere
  - stable in different years in Run 2

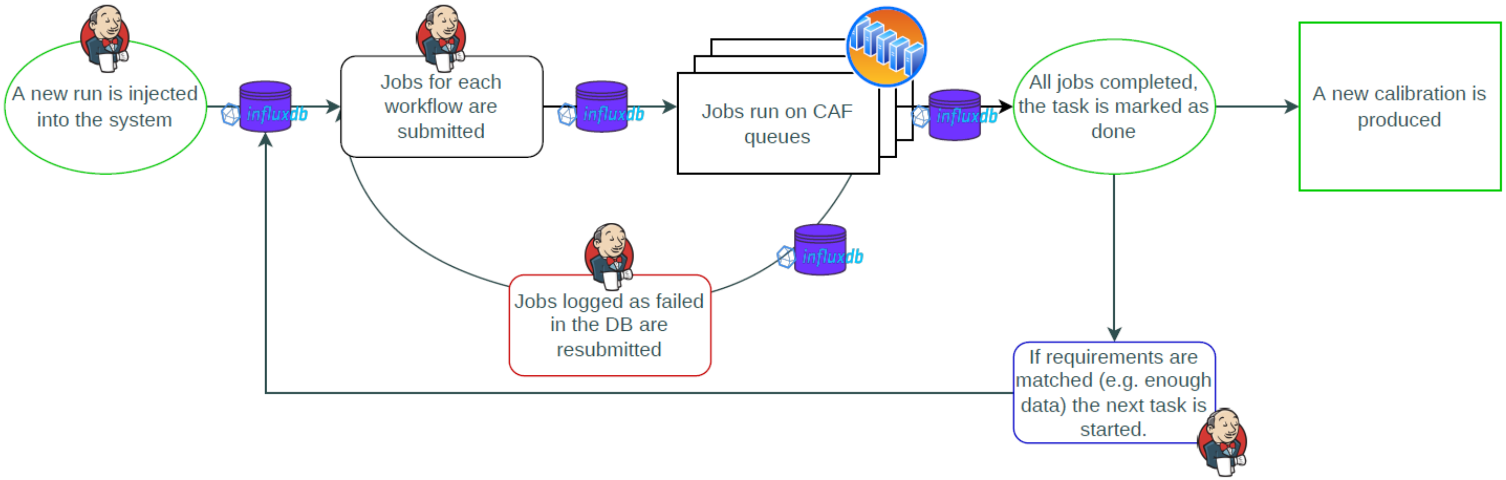
ECAL performance in Run 2

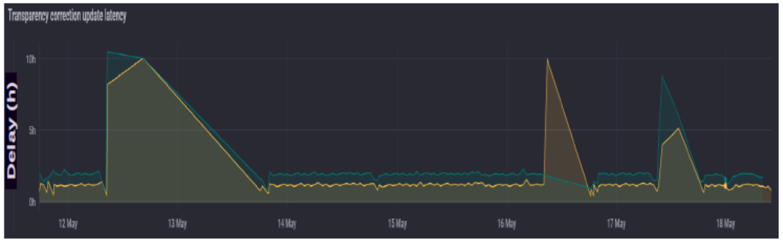


- Similar performance in Run 2 and Run 1

Calibration of prompt reconstruction in Run3

- Implement each calibration workflow as a finate state machine
- Execute jobs regularly updating conditions with predefined conditions
- Constant monitoring and update calibration with fine time granularity





System successfully deployed in Run 3