Calibrating Electrons and Photons in the CMS ECAL using Graph Neural Networks

Simon Rothman (srothman@wesleyan.edu)
On behalf of the CMS Collaboration

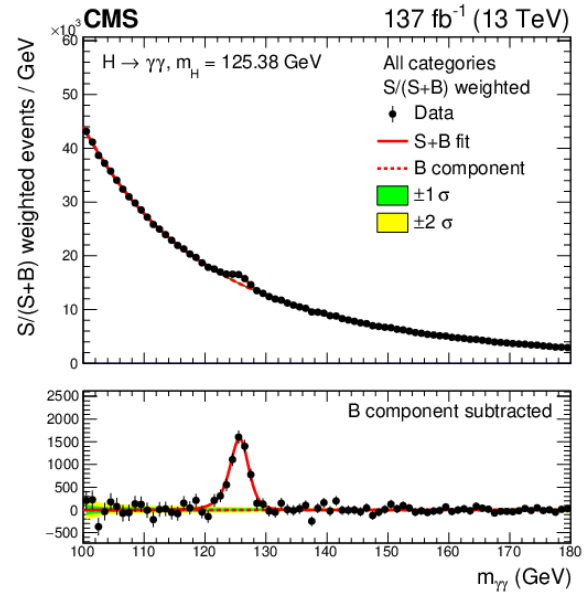
University of Minnesota

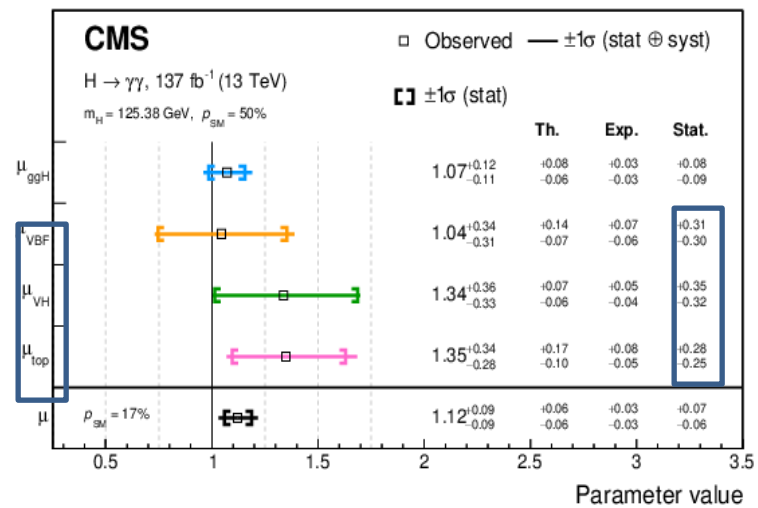
2021 Meeting of the APS Division of Particles and Fields
12-14 July 2021

Overview

- Electrons and photons are essential for all physics results in CMS
  - Higgs physics in the H→ γγ and H→ ZZ → 4ℓ channels.
  - SM measurements : W mass, top physics, multibosons
  - BSM physics : SUSY and heavy resonance searches
- Some of the key properties of the Higgs boson in the diphoton channel are still limited by statistics
  - New precision reconstruction techniques are needed to improve the sensitivity of these measurements.
  - An N% improvement in photon energy resolutions would yield a $\sqrt{N}\%$ and N% improvement in the statistical precision of the signal strength and mass of the Higgs boson, respectively
- We have developed a novel machine learning architecture for the energy calibration of electron and photon objects

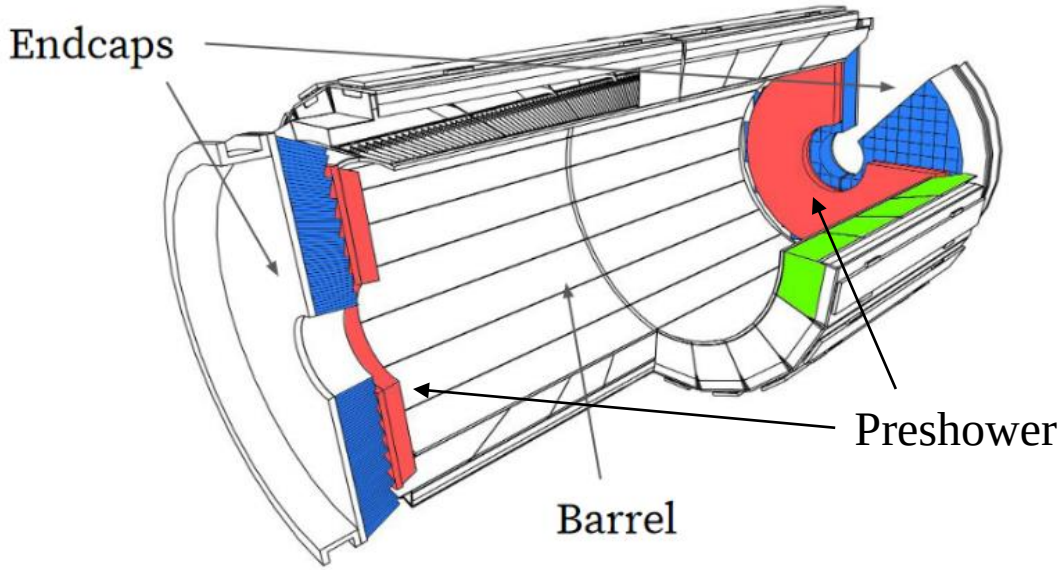
JHEP 07 (2021) 027





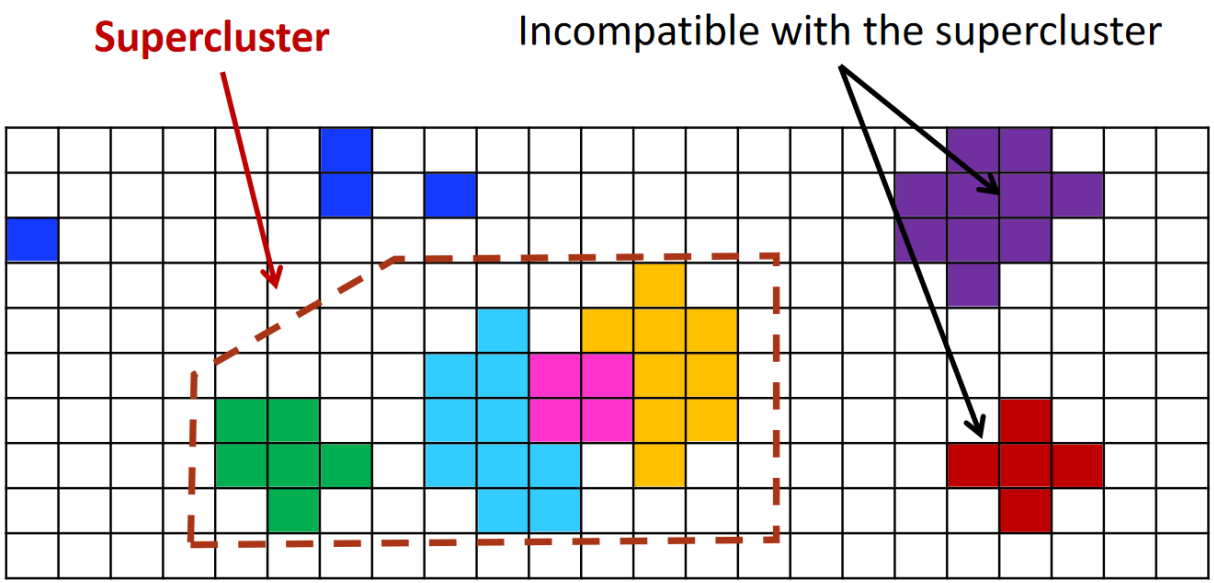
The CMS Electromagnetic Calorimeter (ECAL)

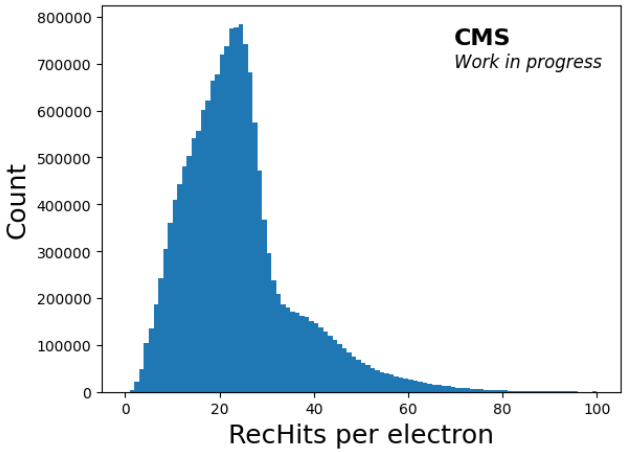
- CMS ECAL is a homogeneous, hermetic crystal calorimeter made of 75,848 scintillating $PbWO_4$ crystals**
- Divided into barrel (61,200 crystals) up to $|\eta|<1.48$ and endcaps (7,324 crystals each) up to $|\eta|<3$
- Endcaps are preceded by sampling preshower detector



**Refer to Jelena's talk on ECAL performance and calibration for more details [link]

Electron and photon object reconstruction at CMS

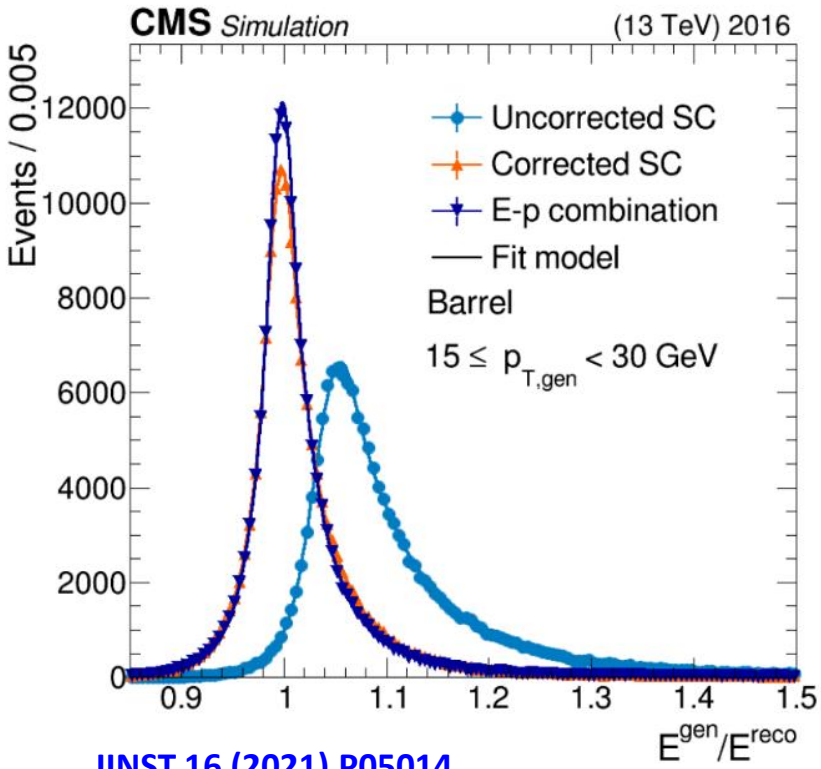




- Per-crystal energy deposits are calibrated and recorded as “RecHits”
- Individual particles are built by clustering nearby RecHits
- Local clusters corresponding to a single EM shower (eg. additional brem. photons) are grouped into “superclusters”
- Superclusters can contain up to >100 RecHits across the ECAL
  - Median ≈25 hits per supercluster

Electron and photon energy corrections

- Supercluster energies are subject to losses, including
  - Energy lost in gaps and in upstream material
  - Finite thresholds to suppress noise
  - Unclustered energy
- Losses are compensated by energy corrections, currently implemented as a Boosted Decision Tree (BDT) based regression
  - Uses ≈30 high-level input features to describe EM shower
- BDT energy corrections have been fine-tuned over several years and have supported all physics analyses in CMS during LHC Run 2
  - E.g. Higgs mass with 0.1% precision



JINST 16 (2021) P05014

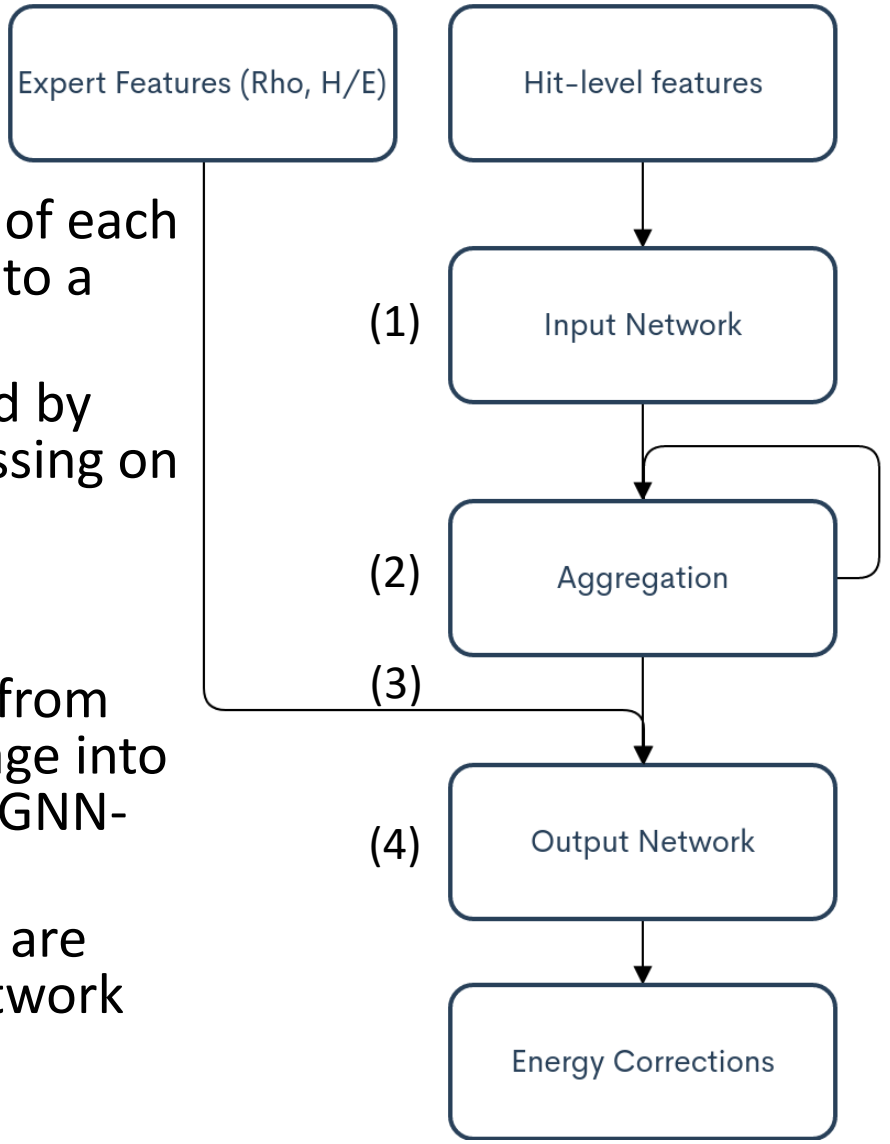
New machine learning approach (1)

- We have developed a new architecture for these energy corrections, the Dynamic Reduction Network (DRN)
- Built on graph neural network (GNN) techniques to allow regression on low-level detector signals, instead of reconstructed parameters
- Electrons and photons are represented as graphs, where vertices are per-crystal energy deposits and the edges are dynamically formed by clustering in a high-dimensional latent space
- Advantages of GNN approach:
  - Use of low-level input features ensures access to full information content of every event
  - Arbitrary-geometry events can be handled with no padding
  - Easy incorporation of multiple subdetectors (e.g. ECAL preshower, tracker), with layer alignment guaranteed by geometric input features (in progress)

New machine learning approach (2)

DRN proceeds as follows:

1. The (position, energy) coordinates of each hit in a supercluster are mapped into a high-dimensional latent space
2. High-level information is developed by iteratively performing message passing on dynamically-generated graphs and aggregating similar vertices
3. Additional event- and object-level features which cannot be inferred from the detector hits (e.g. pileup, leakage into the HCAL) can be concatenated to GNN-produced features
4. Resulting set of high-level features are passed through another neural network to produce energy corrections



Technical Details

We show today the performance of the electron regression using ECAL-only information

- Used for absolute calibration of the ECAL
- Not the final e/γ objects used in physics analyses

Training

- Training on particle gun simulation with ideal ECAL calibration (≈17M events, 80/20 train/test split, uniform p_T distribution)
- Performed at University of Minnesota Supercomputing Institute (MSI)
- Parameters of a double-sided crystal ball pdf are the regression target

Computational Performance

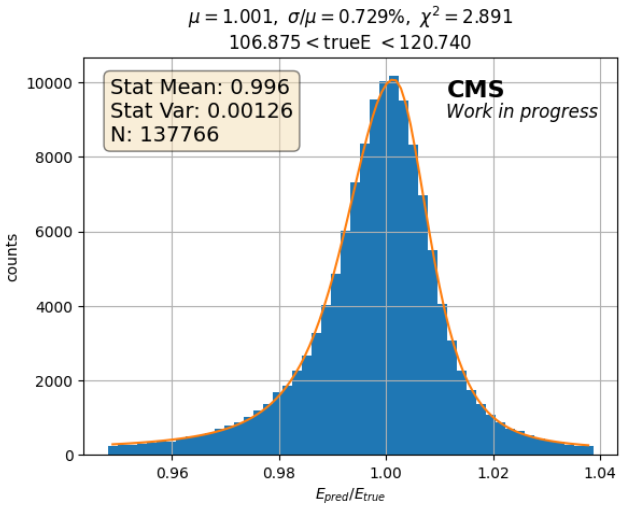
- Total training time ≈20 hours
- 50 epochs
- 10,000 stochastic gradient descent batches per epoch

Performance validation

- Form histogram of E_{Pred}/ E_{True} and fit with Cruijff function

$$f(x) = \begin{cases} N \exp\left[-\frac{(x-\mu)^2}{2\sigma_L^2+\alpha_L(x-\mu)^2}\right], & \text{if } (x-\mu) < 0 \\ N \exp\left[-\frac{(x-\mu)^2}{2\sigma_R^2+\alpha_R(x-\mu)^2}\right], & \text{if } (x-\mu) \geq 0 \end{cases}$$

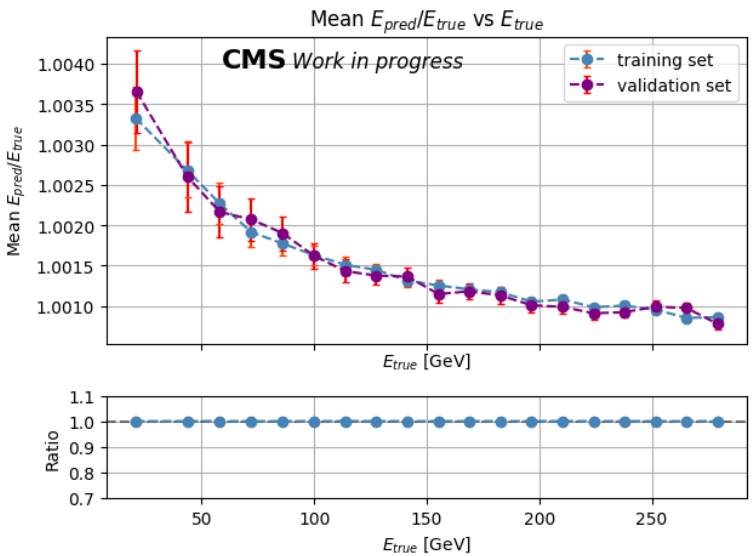
$$\sigma = \frac{\sigma_L + \sigma_R}{2}$$

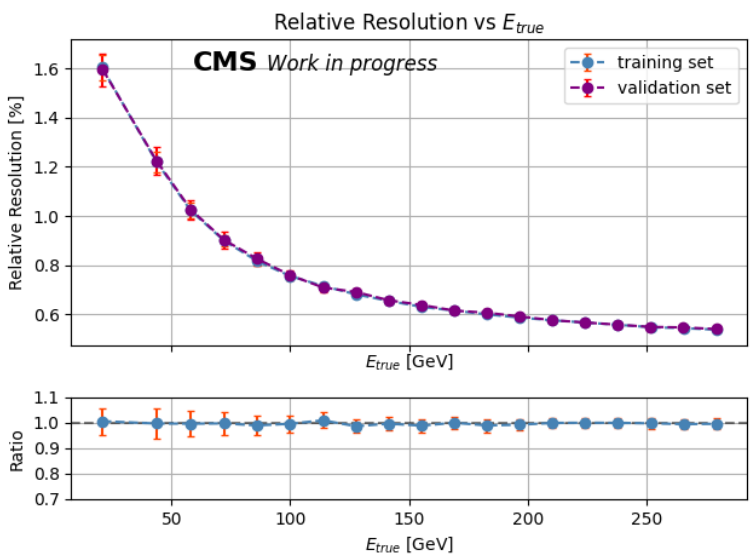


- Extract mean (μ) and relative resolution (σ/μ) as key performance metrics
- Performance also evaluated in Z→ee decays, in both simulation and collision data
- **New architecture (DRN) and previous state-of-the-art (BDT) are evaluated and validated on the same events**

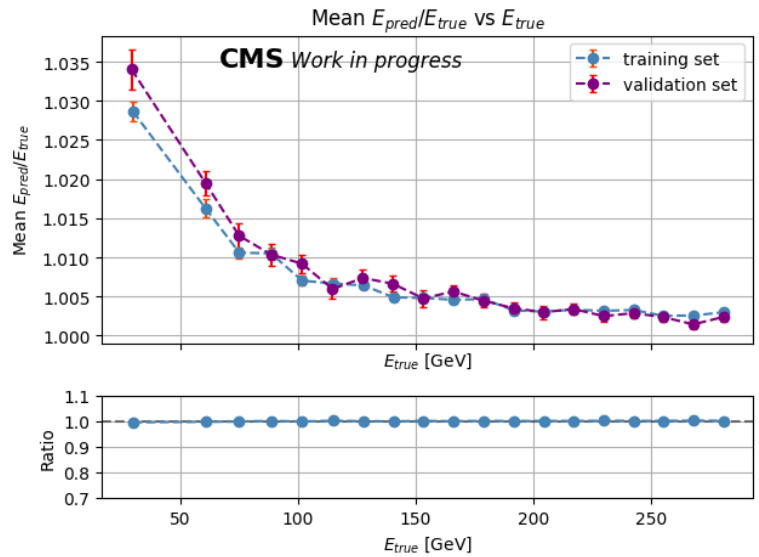
Overtraining checks

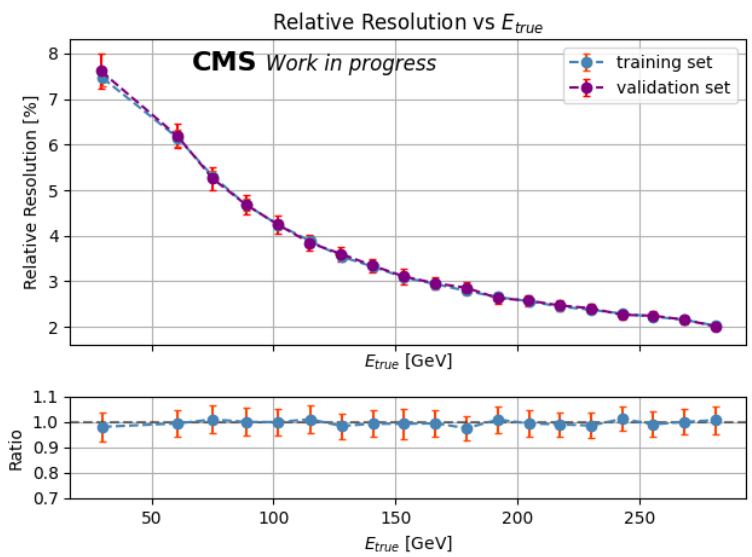
Barrel





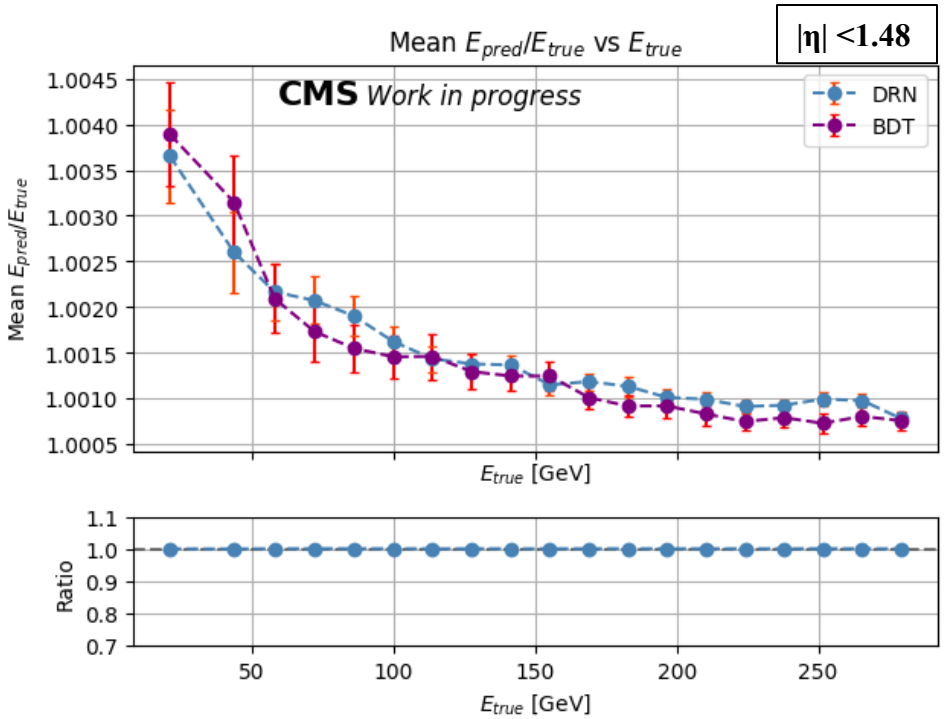
Endcaps

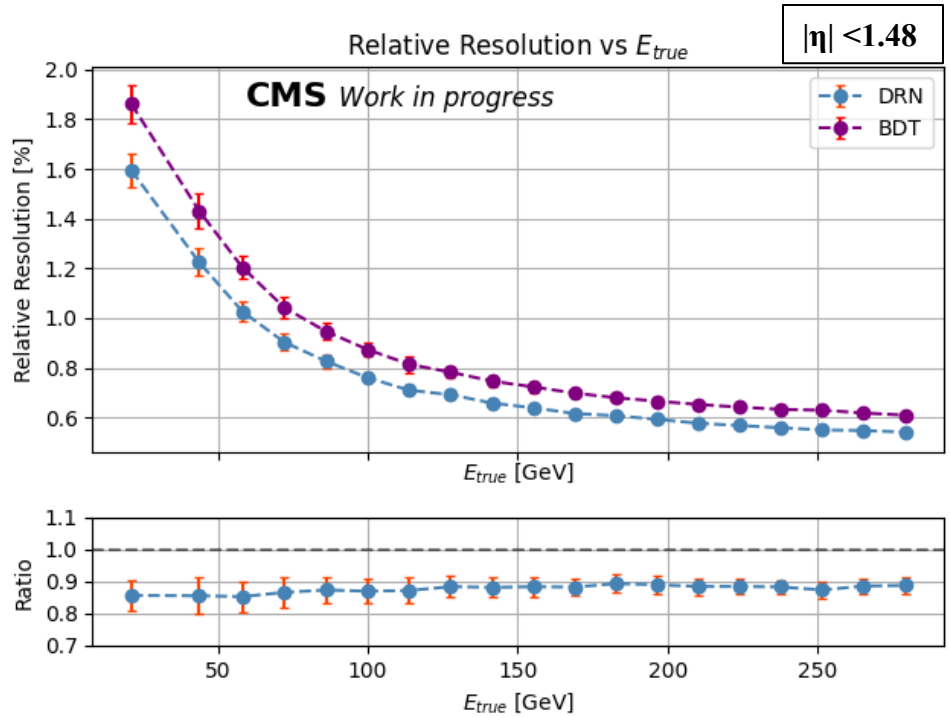




PERFORMANCE IN PARTICLE GUN SIMULATION

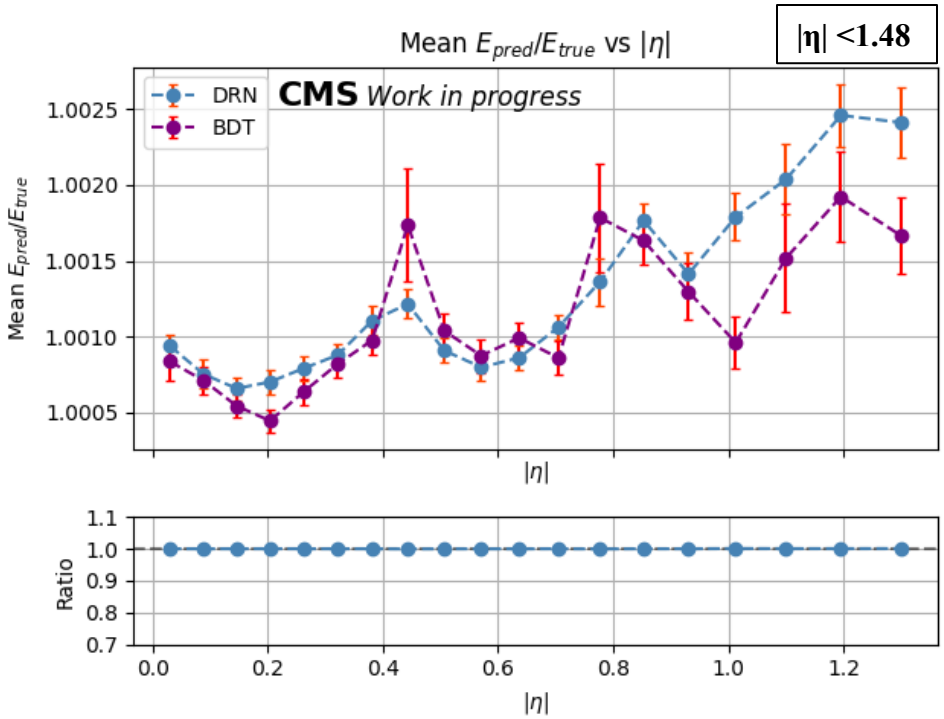
Performance vs energy

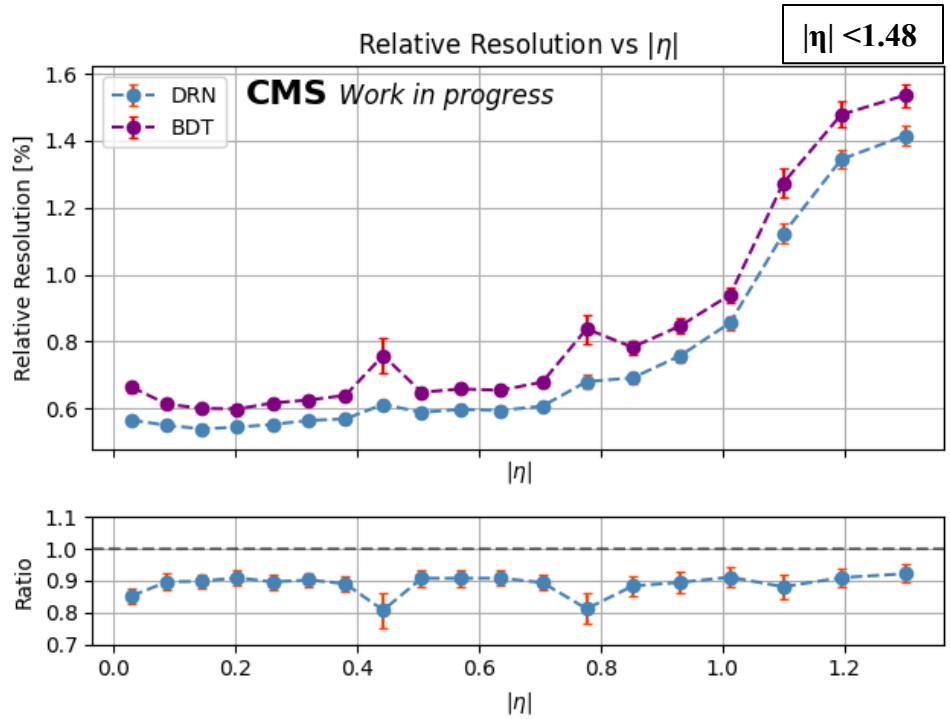




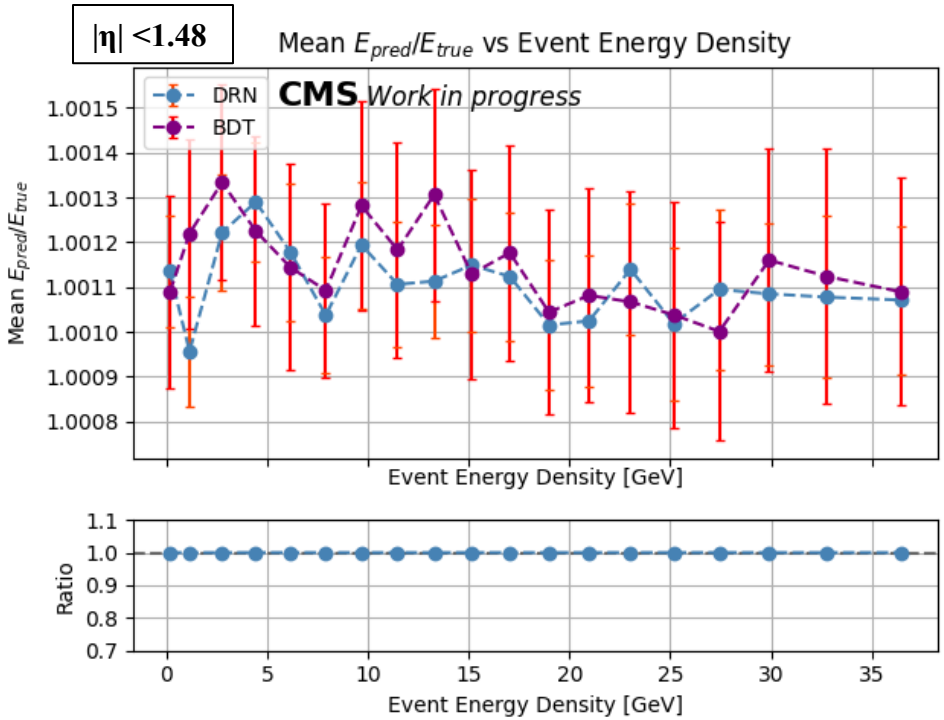
- Performance comparison similar for 1.48<|η|<3 (figures in backup)
- Resolution improved by 10% with respect to previous state-of-the-art

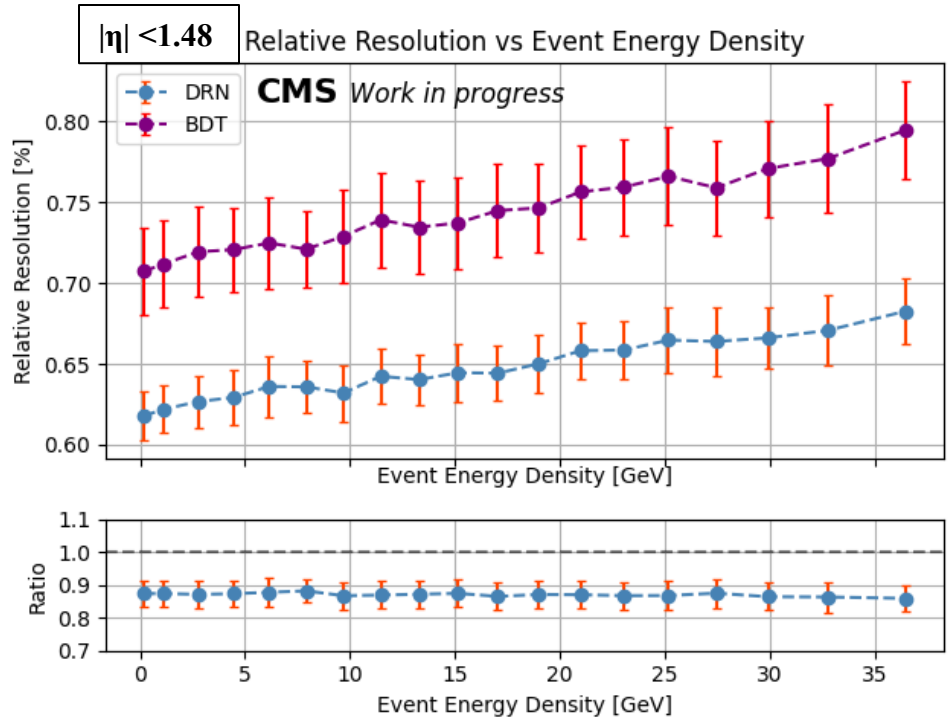
Performance vs pseudorapidity





Performance vs pileup



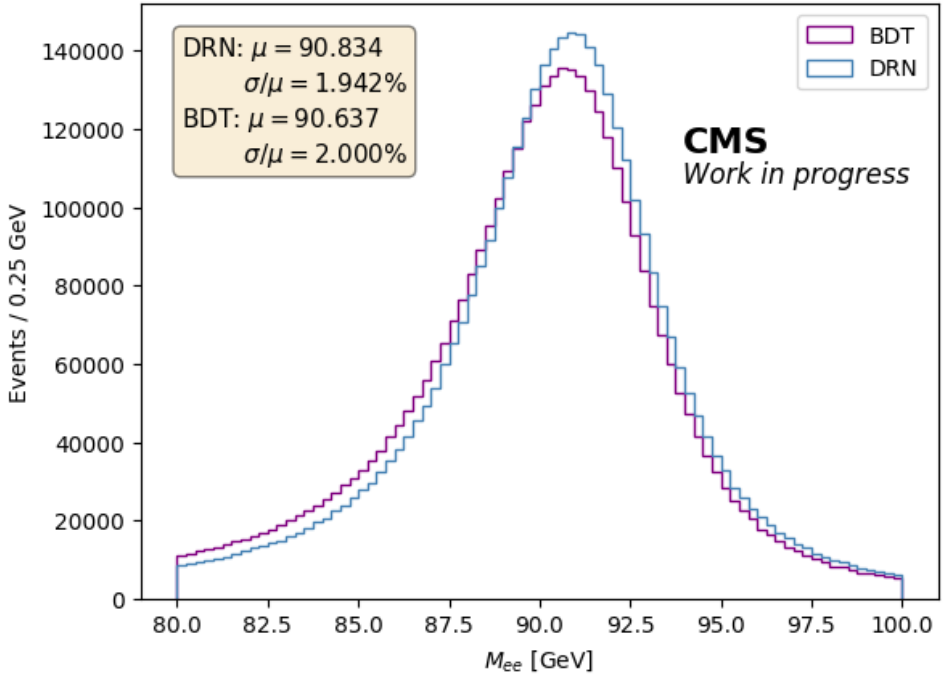


- Pileup measured as energy density of entire event
- Incorporation of high-level information on pileup in addition to low-level hits ensures highly-stable response as a function of pileup

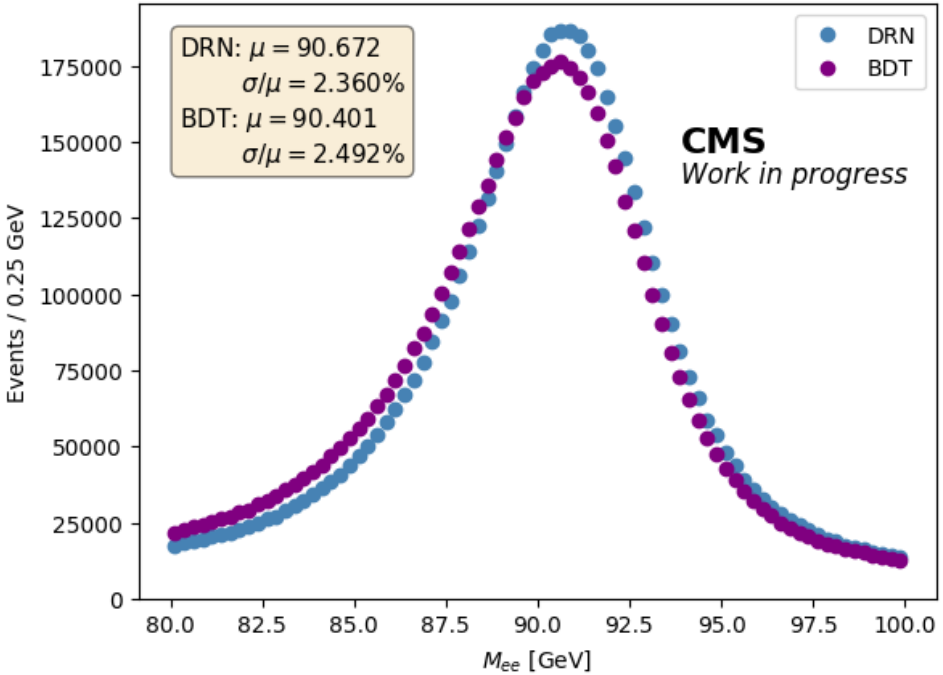
PERFORMANCE IN Z DECAYS

Z→ee performance

Simulation



Collision Data



- Both electrons satisfy $|\eta|<1.48$; this shows the best performance gains ($1.57 <|\eta|$ in backup)
- Select electron pairs with $p_T > 20$ GeV
- Peak fit with crystal ball function, convolved with relativistic Breit Wigner function to account for natural width of Z boson (≈2.5 GeV)
- Resolution is improved with respect to BDT by ≈5%
- Work is ongoing to bring performance gain in Z decays up to 10% improvement seen in ideal particle gun simulation

Summary

- We have developed a novel machine learning architecture based on dynamic graph neural networks for HEP problems with variable and/or complicated geometries
- We have applied this novel architecture to electrons used for the calibration of the CMS ECAL
  - Preliminary results yield an improvement in the energy resolution of up to 10% in ideal simulation and 5% in realistic simulation and collision data
  - Work is ongoing to bring real-world performance up to 10% improvement seen in ideal simulation
- Work is ongoing to derive similar corrections for the electron and photon objects used in physics analyses.
- We are in the process of deploying these models for use in CMS during Run 3 of the LHC using heterogeneous computing resources

BACKUP

Training Details

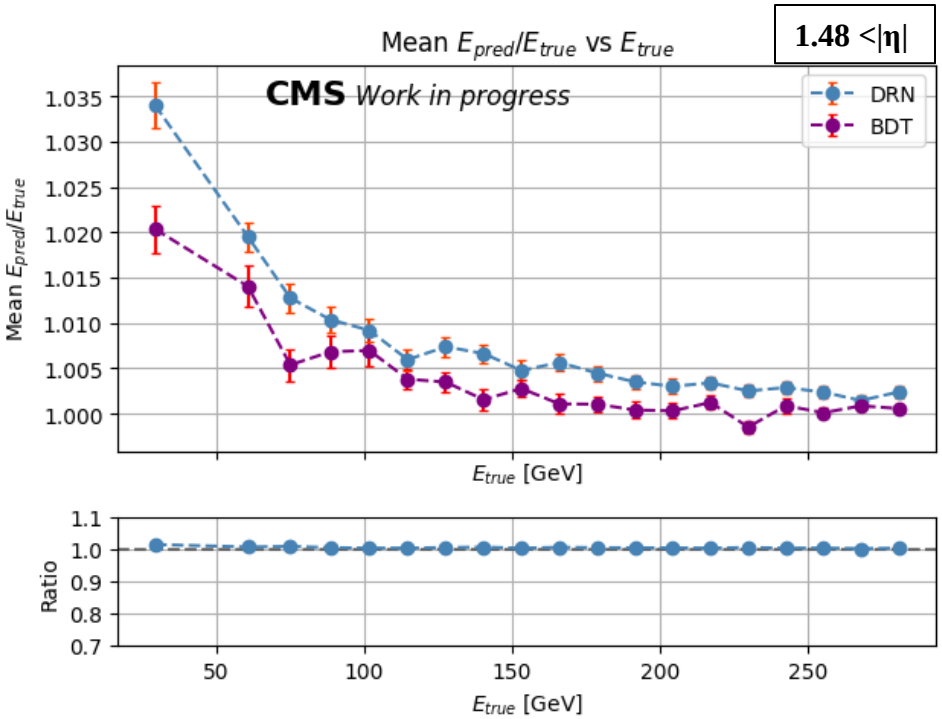
- Training sample is flat Pt gun 1-300 GeV with ideal intercalibration
  - Cut 5 < Gen Energy [GeV] < 300
  - Perform gen matching with a dR <= 0.05
- Barrel and Endcaps trained separately
- Barrel: 13,776,520 total electrons
- Endcaps: 3,436,566 total electrons
- 80/20 training/validation split applied independently in barrel and endcaps
- Training done on Nvidia Tesla v100 GPUs
- Training target is energy correct factor ERaw/ETrue, log-compressed to ensure symmetry between over- and under-measurement
- DRN output is 6-dimensional parameterization of double-sided crystal ball pdf; loss function is negative log-likelihood
- Predicted energy correction taken as pdf central value

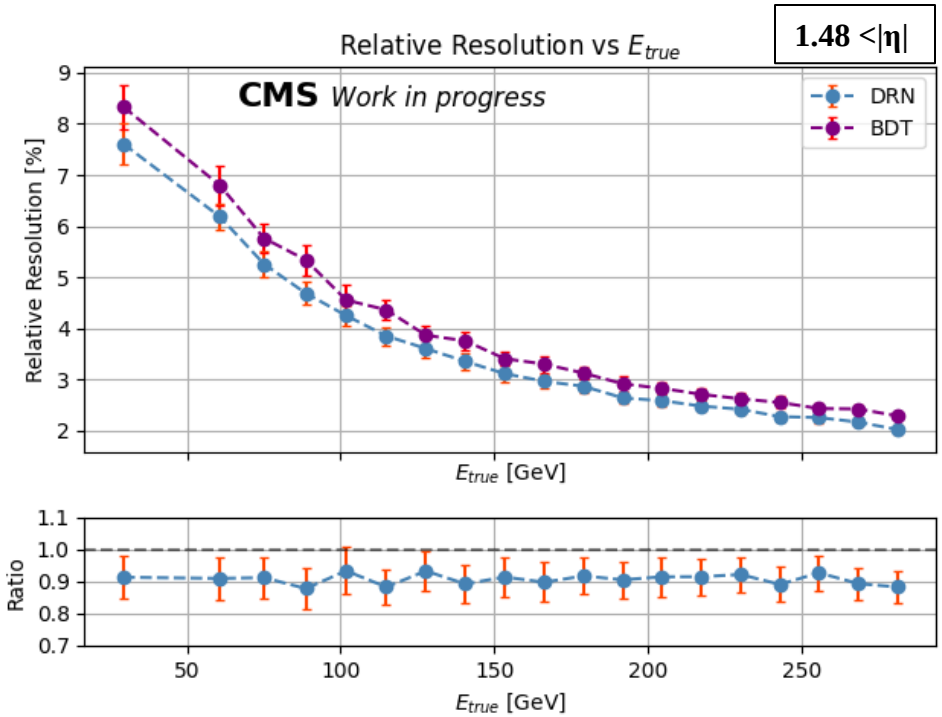
Throughput performance

- Training and inference run on Nvidia Tesla v100 GPU
- A number of performance optimizations remain to be applied

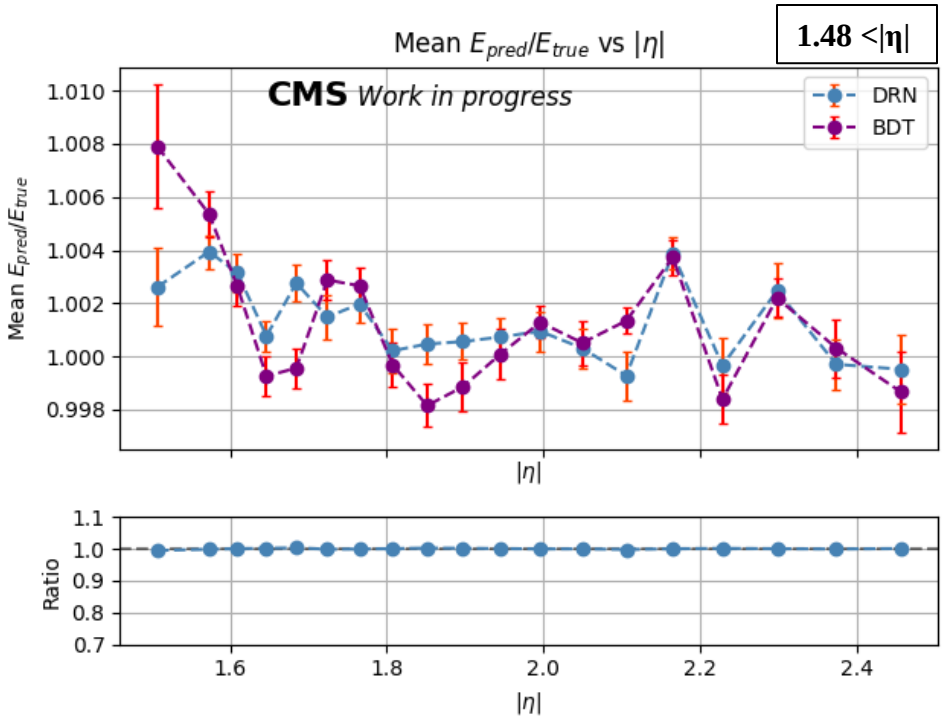
Batch Size	CPU Inference Speed (e/sec)	GPU Inference Speed (e/sec)
1	90	50
100	500	3400
50000	70	18000

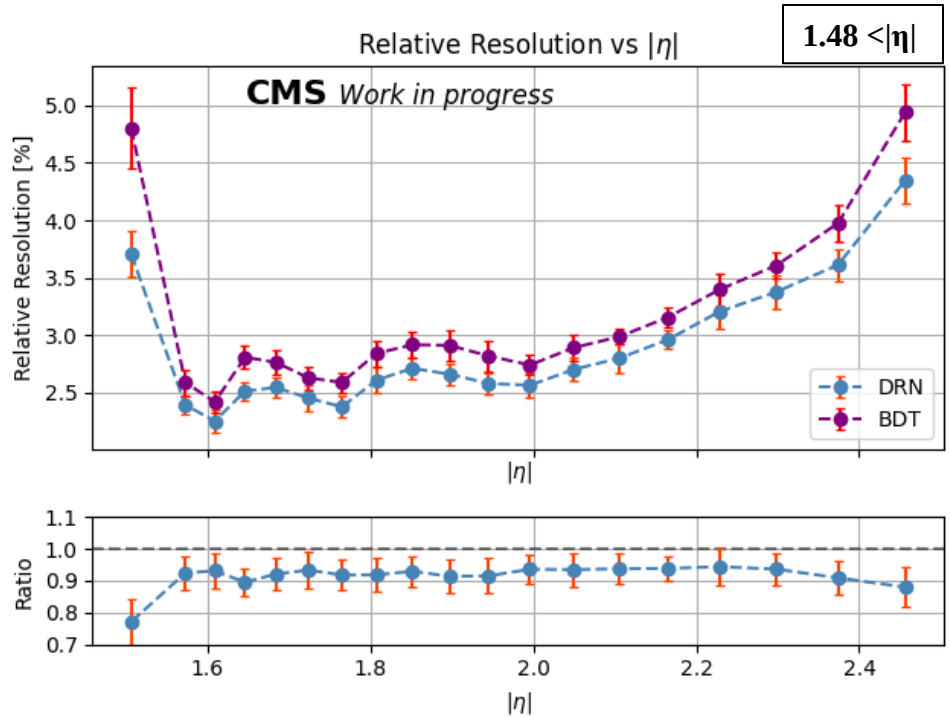
Performance vs energy (Endcaps)



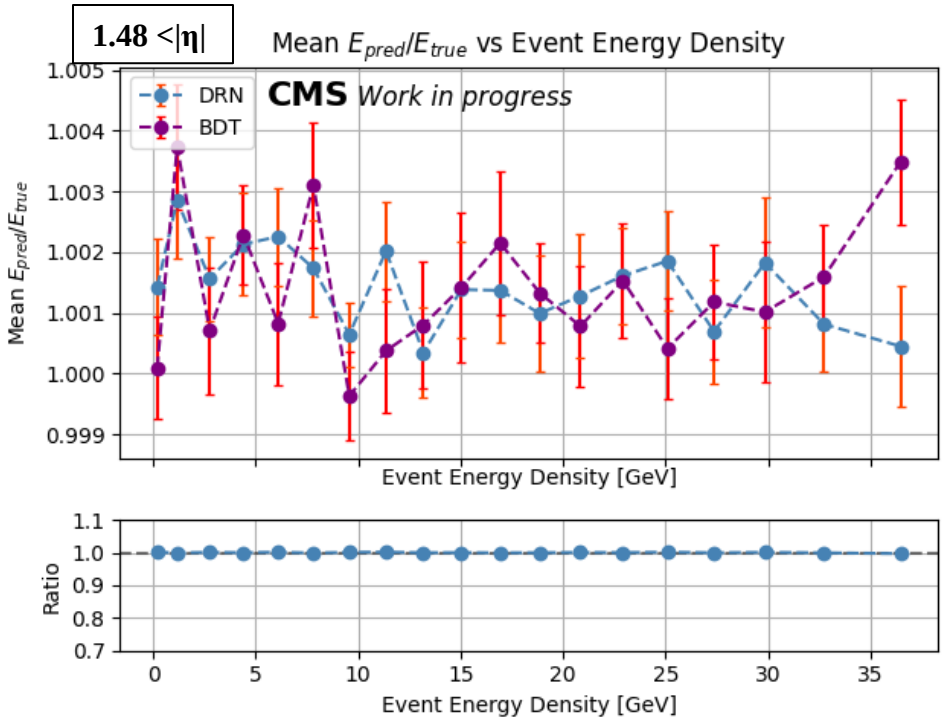


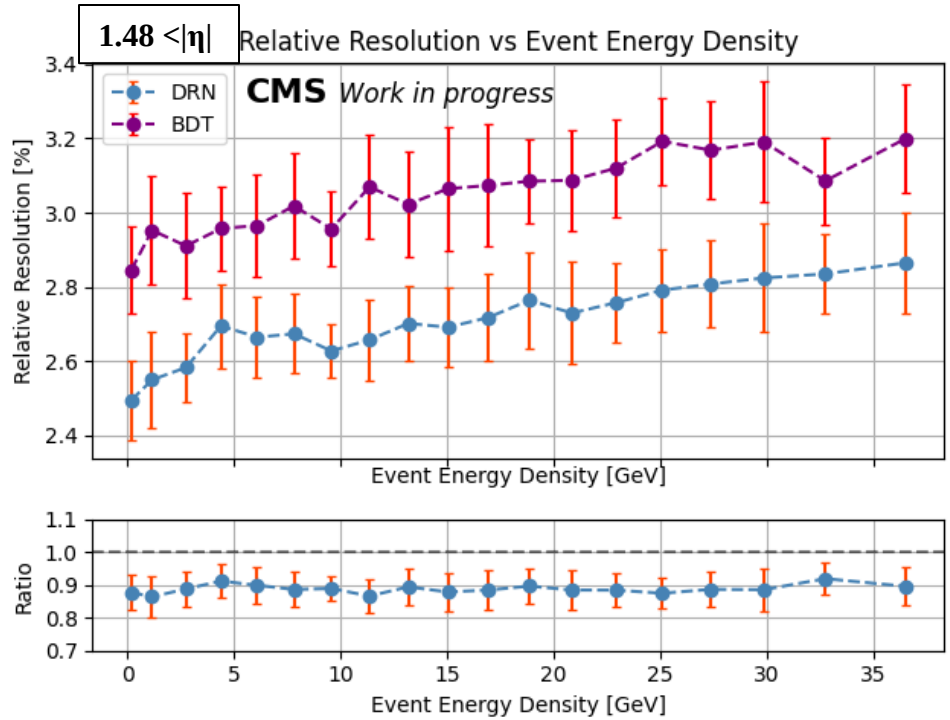
Performance vs pseudorapidity (Endcaps)





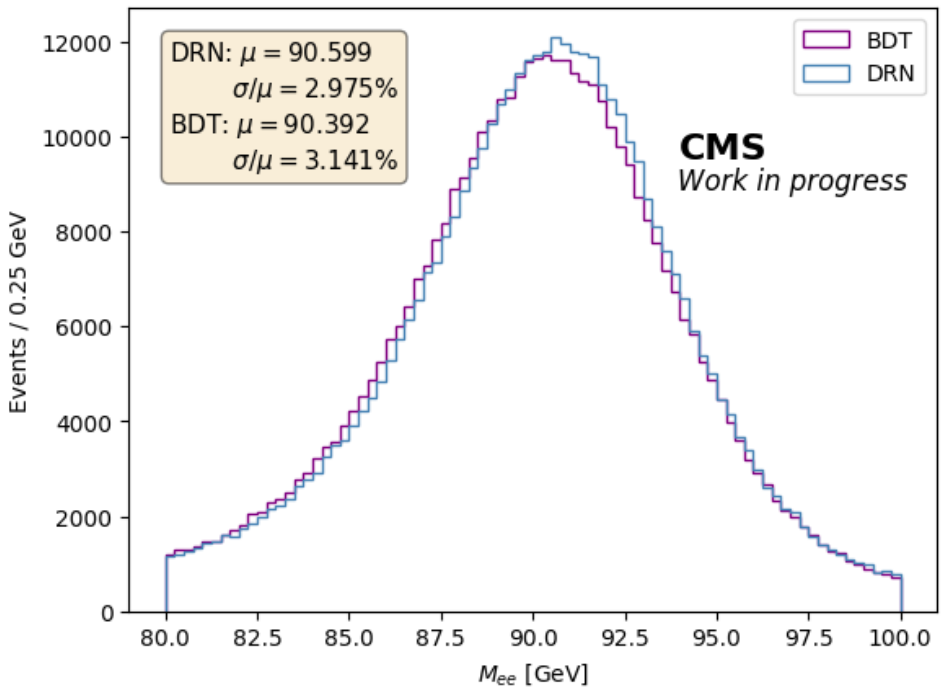
Performance vs pileup (Endcaps)



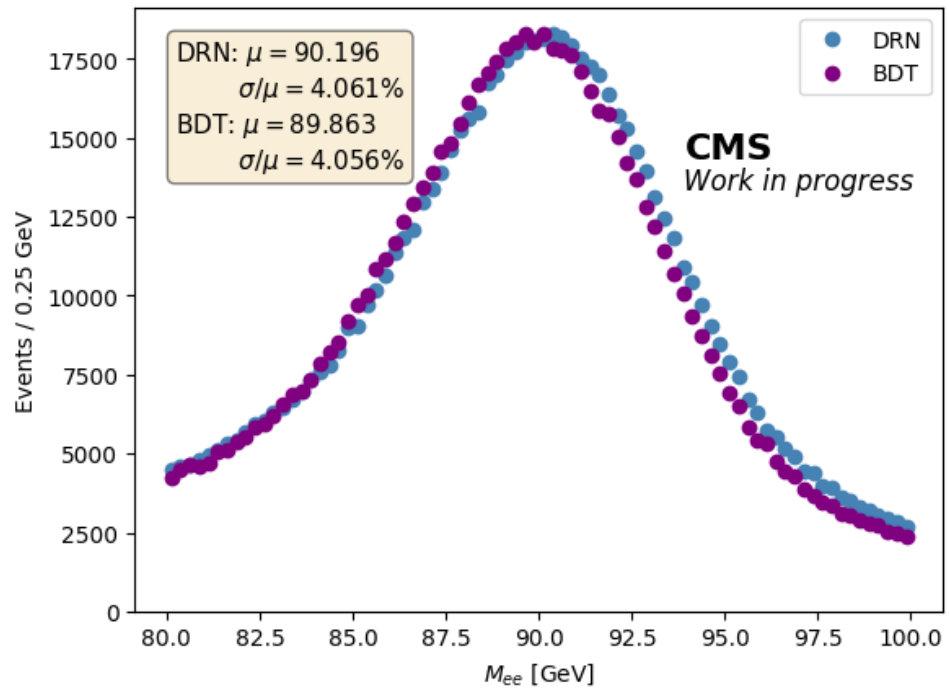


Z→ee Performance (Endcap-Endcap)

Simulation



Collision Data



- Here both e $|\eta| > 1.48$
- Select electron pairs with $p_T > 20$ GeV
- Peak fit with crystal ball function, convolved with relativistic Breit Wigner function to account for natural width of Z boson (≈2.5 GeV)
- Resolution unchanged with respect to BDT
- Lack of performance gain likely due to limited training stats in endcaps. Work is in progress to remedy this