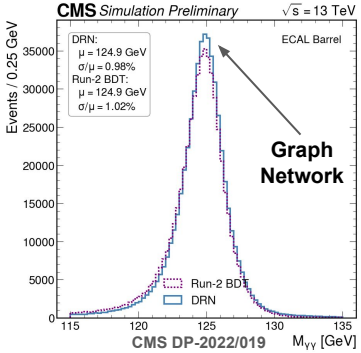
Calorimetry with Graph Neural Networks in CMS

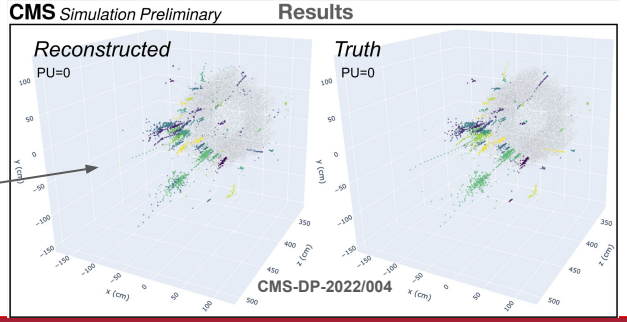
Simon Rothman (srothman@mit.edu)
On behalf of the CMS Collaboration,

Massachusetts Institute of Technology

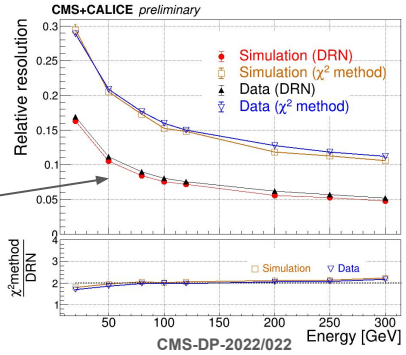
Outline

1. Calorimetery in CMS
2. Why use graph networks?
3. How do graph networks work?
4. Three case studies in CMS:
   a. Electron/photon energy reconstruction in the CMS crystal ECAL
   b. Pion energy reconstruction in the prototype HGCAL testbeam
   c. Hit-to-particle reconstruction in the CMS HGCAL
5. Conclusions





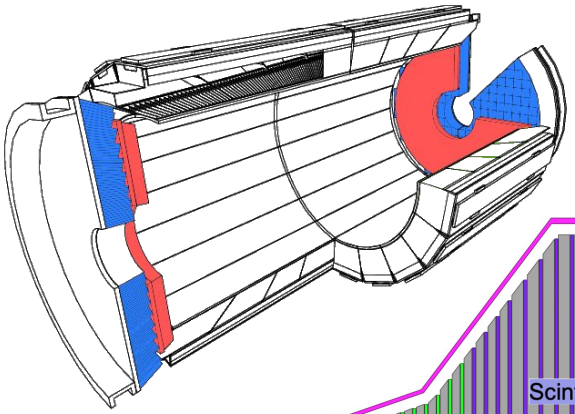
Graph Network

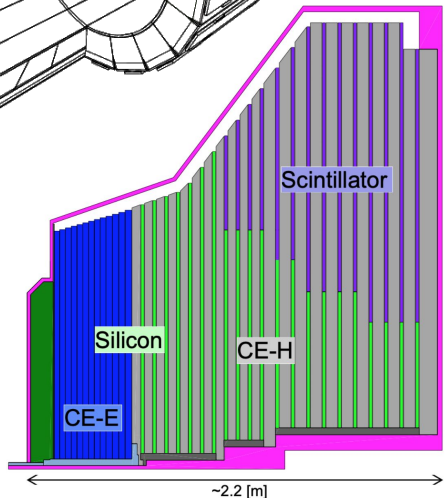


Graph Network

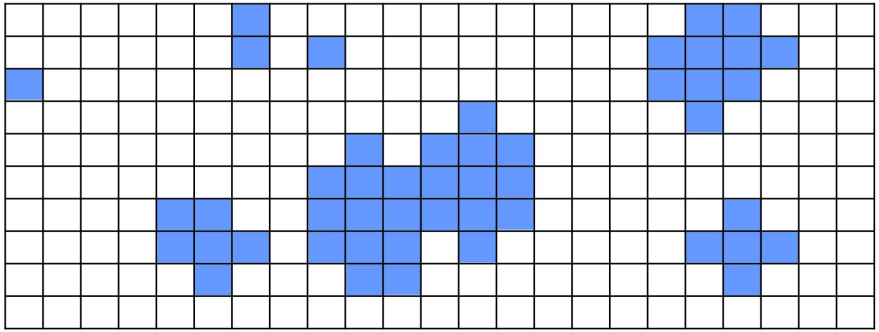
Calorimetery in CMS

- Phase 1 detector:
  - **Crystal ECAL:** homogeneous calorimeter composed of ~75k scintillating $PbWO_4$ crystals
  - **Preshower**: sampling calorimeter, in front of ECAL endcaps
  - **HCAL:** Sampling hadronic calorimeter situated just outside ECAL
- Phase 2 detector:
  - Add **HGCAL:** Replaces endcap calorimeters (both ECAL and HCAL). Silicon and scintillator tile sampling calorimeter with unprecedented granularity (~6 million readout channels)
- Today focus on ECAL and HGCAL reconstruction
  - Similar work also in progress in HCAL
- Performance of calorimeter energy measurements is critical for nearly all physics in CMS



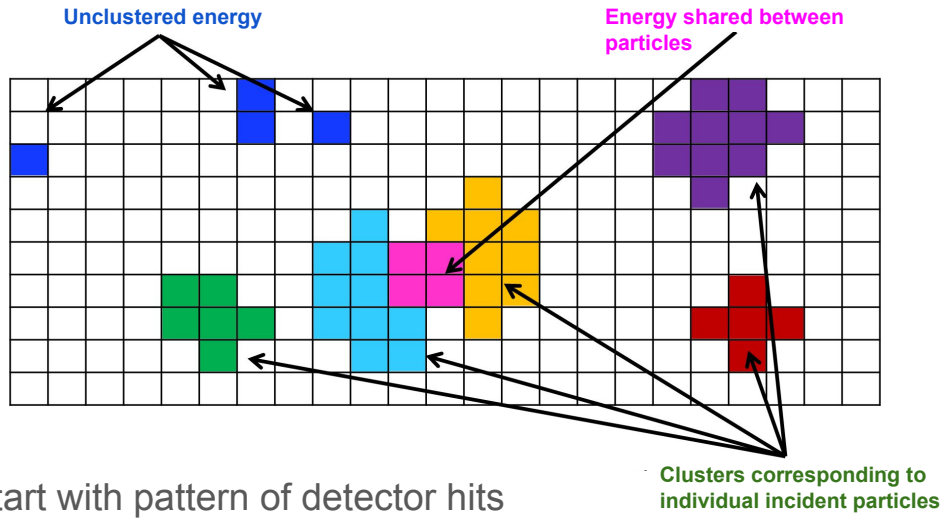


Calorimeter reconstruction



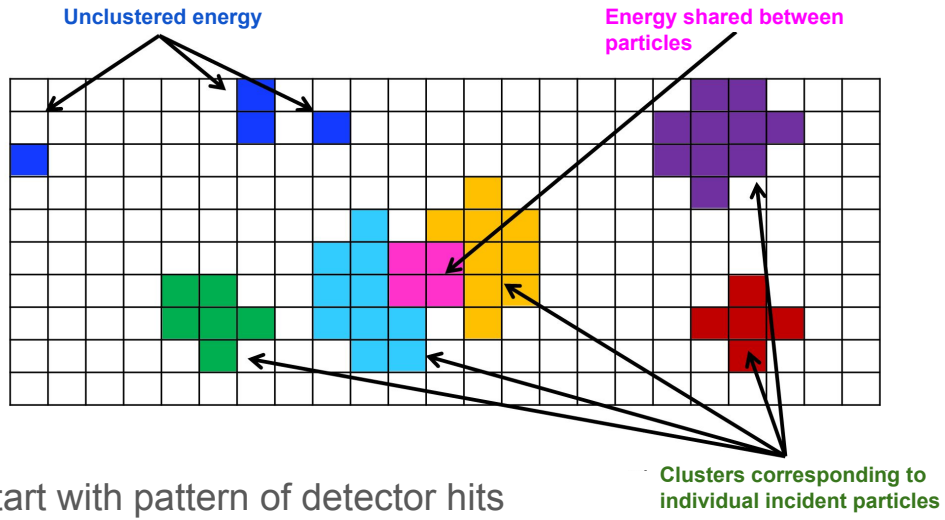
- Start with pattern of detector hits
- Two main tasks:

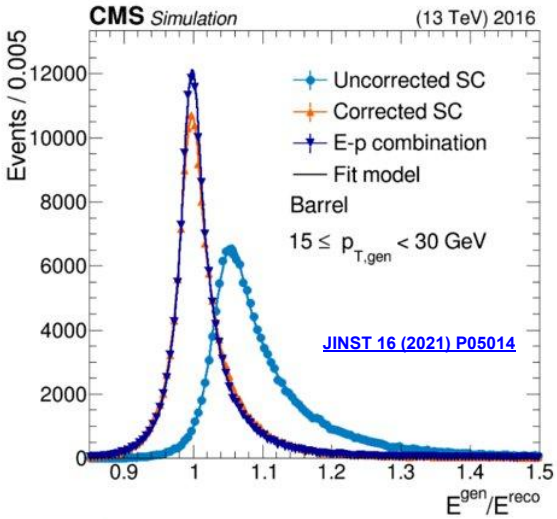
Calorimeter reconstruction



- Start with pattern of detector hits
- Two main tasks:
  - Associate each incident particle with a collection of hits

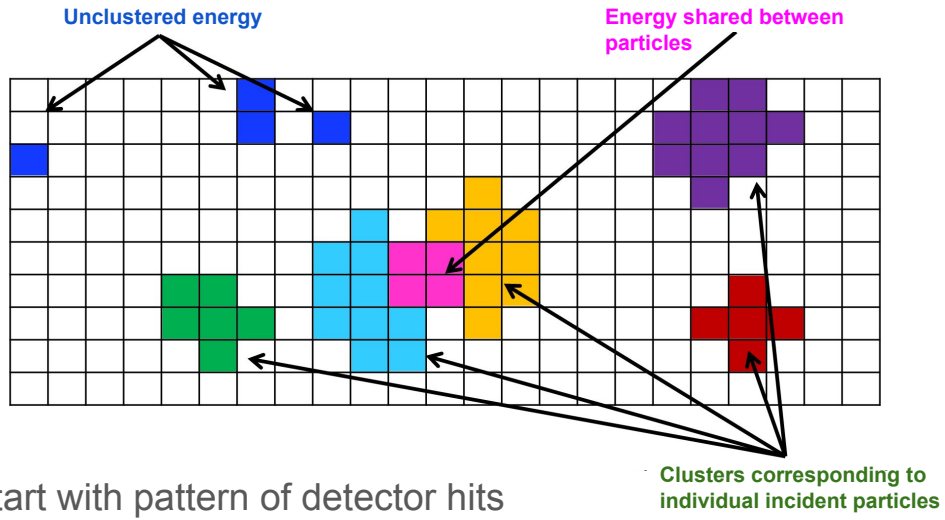
Calorimeter reconstruction

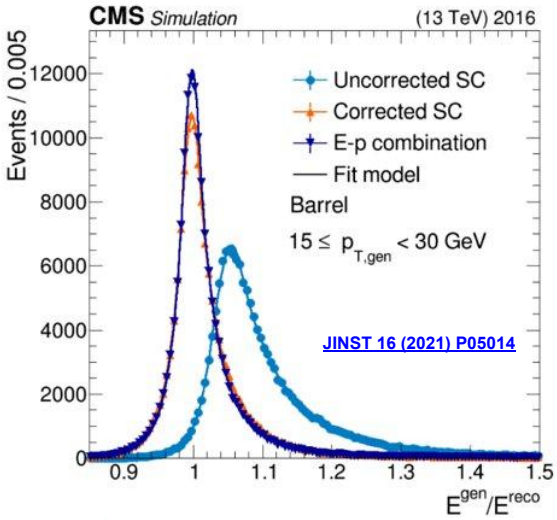




- Start with pattern of detector hits
- Two main tasks:
  - Associate each incident particle with a collection of hits
  - Compensate for any energy mismeasurement

Calorimeter reconstruction

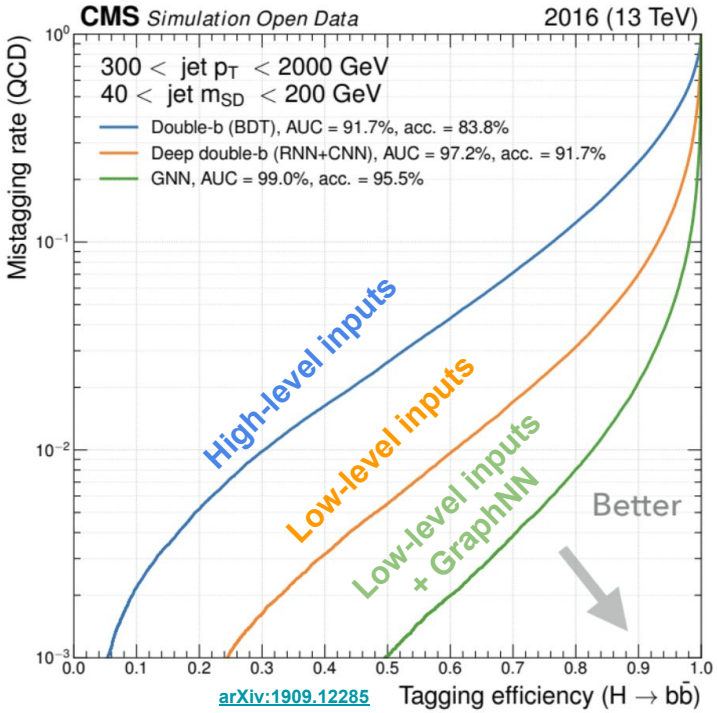




- Start with pattern of detector hits
- Two main tasks:
  - Associate each incident particle with a collection of hits
  - Compensate for any energy mismeasurement
- Both typically done by rule-based algorithms or boosted decision trees (BDTs)
  - Rely on high-level human-engineered inputs

New machine learning strategy

- ML is most powerful when applied on low-level inputs
  - Gives model access to full information content of every event
  - Avoids potential for biases from human feature engineering
- This has been seen in e.g. jet tagging*
  - Train on jet constituents rather than high-level variables
- We would like to also train on low-level calorimeter inputs: the detector hits



*First with deepJet: arXiv:2008.10519

arXiv:1909.12285

What architecture to use?

- Inputs can be challenging for most architectures
  - There can be any number of hits (detector is zero-suppressed, so not all channels are active)
  - They can be distributed across multiple very different detector components
  - They are naturally represented in at least 4 dimensions (x, y, z, energy)
  - They are in no particular order

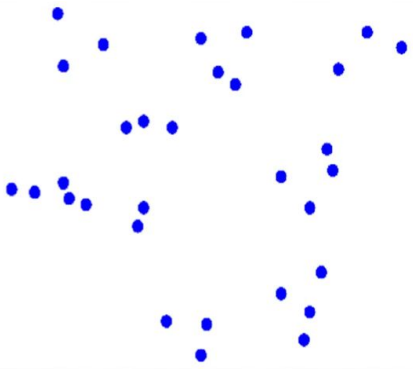
Can it (easily) handle…	BDT	MLP	CNN	RNN	GNN
Variable-size input	❌	❌	✅	✅	✅
Complicated geometries	✅	✅	❌*	✅	✅
4D inputs	✅	✅	❌	✅	✅
Unordered inputs	❌	❌	✅	❌	✅

- **Graph networks are the best option for these types of inputs**

***CNNs work best with rectangular input spaces**

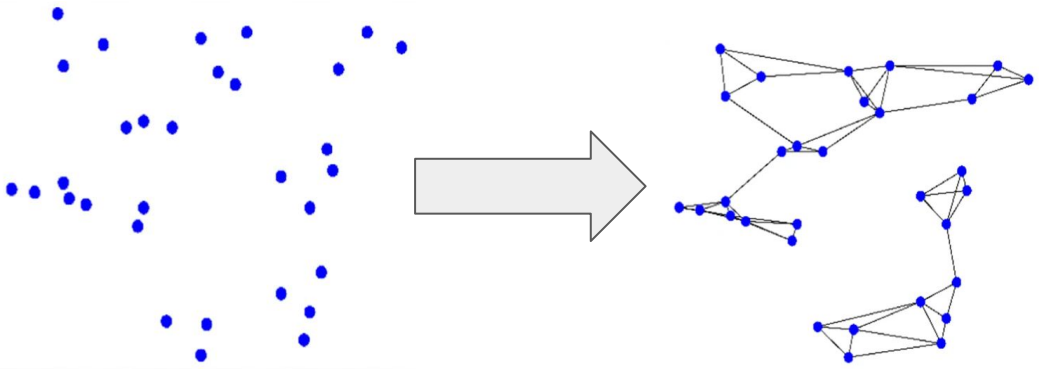
How do graph networks work?

1. Collection of hits is represented as a point cloud



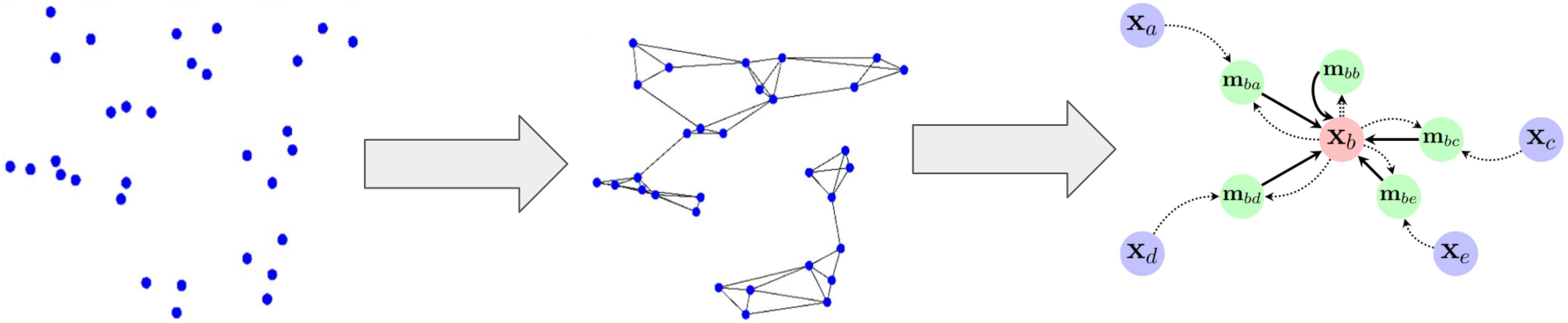
How do graph networks work?

1. Collection of hits is represented as a point cloud
2. Generate graph by drawing edges between k nearest neighbors of each hit

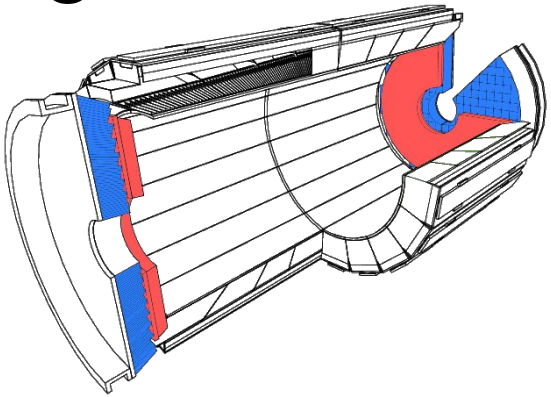


How do graph networks work?

1. Collection of hits is represented as a point cloud
2. Generate graph by drawing edges between k nearest neighbors of each hit
3. Perform “message passing” to allow information to flow along graph edges (analogous to image convolutions in CNNs)



Case Study: ECAL energy regression

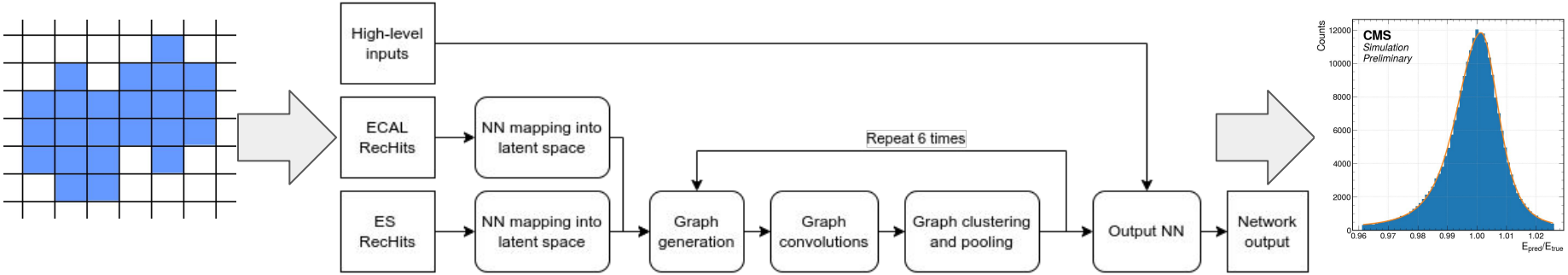


Problem statement

- Electron/photon energies measured in ECAL subject to losses, including
  - Energy lost in gaps and upstream material
  - Longitudinal energy leakage
  - Finite thresholds to suppress noise
- Compensated per-particle by ML regression
- Run-2 corrections implemented as a BDT
  - Uses ~30 high-level input features to describe EM shower
- BDT corrections first developed for Higgs discovery in 2012
- BDT energy corrections have supported all physics analyses using electrons/photons in CMS during LHC Run 2
- Can we do better using a graph network?

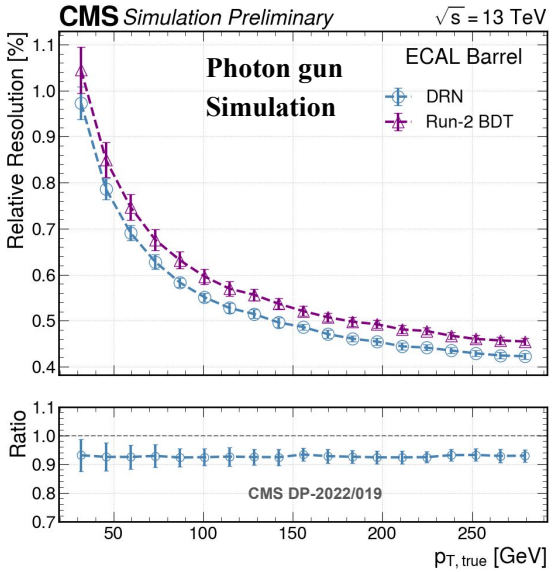
The Dynamic Reduction Network

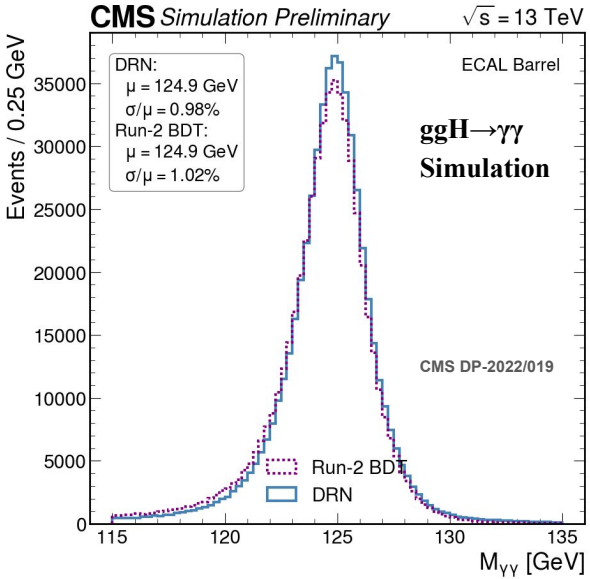
- New architecture: Dynamic Reduction Network (DRN)[1], based on dynamic graph NNs
  - Graph operations take place in a high-dimensional latent space
  - Added clustering and pooling step to aggregate information across the graph
- Input includes hits from both ECAL and ECAL preshower, as well as additional features to describe information not encoded in the hit collection (pileup, leakage into HCAL)
- Outputs parametrization of double-sided crystal ball probability density for energy correction factor

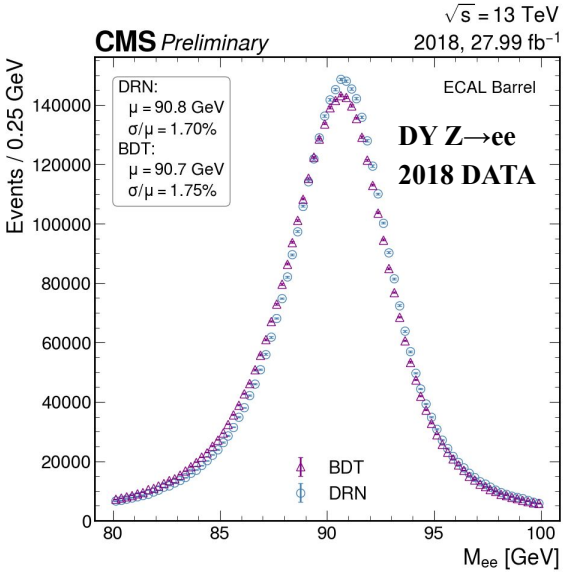


[1]: arXiv:2003.08013

Performance

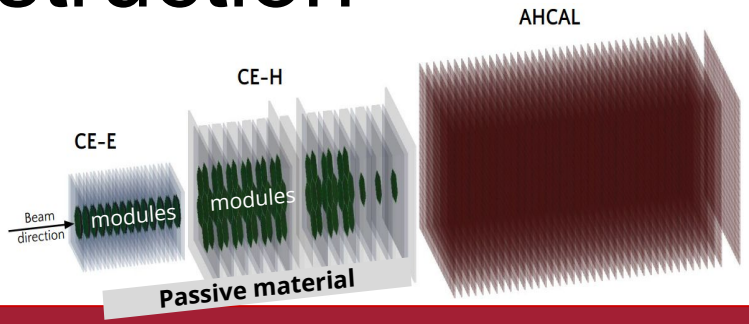






- Improved per-object energy resolution by factor of ~10%
- Translates to improved invariant mass resolution by factor of ~5%
- First major change to energy regression algorithm since 2012
  - In process of being deployed for Run 3

Case Study: HGCAL prototype testbeam pion energy reconstruction



Baseline energy reconstruction

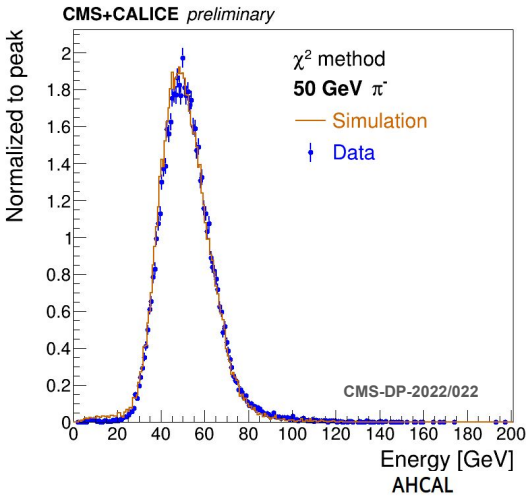
- Prototype detector with 3 different sections
  - Si CE-E, CE-H, scintillator-tile CALICE AHCAL
- **Detector-level calibration:**
  - CE-E (CE-H+AHCAL) calibrated with 50 GeV e^+ (π^-) beam

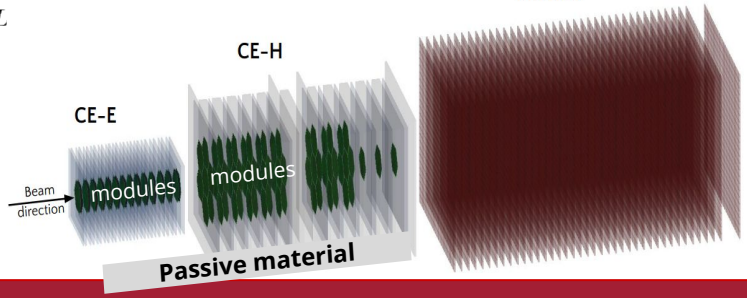
$$E^{fix} = \alpha^{fix} E_{CE-E} + \beta^{fix} E_{CE-H} + \gamma^{fix} E_{AHCAL}$$

- **χ^2-method :**
  - CE-E, CE-H, AHCAL combined with energy-dependent weights

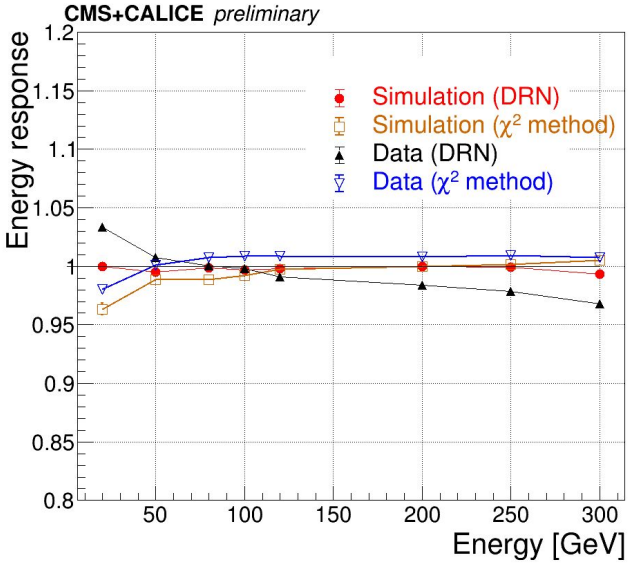
$$E_{\chi^2} = \alpha(E) E^{fix}_{CE-E} + \beta(E) E^{fix}_{CE-H} + \gamma(E) E^{fix}_{AHCAL}$$

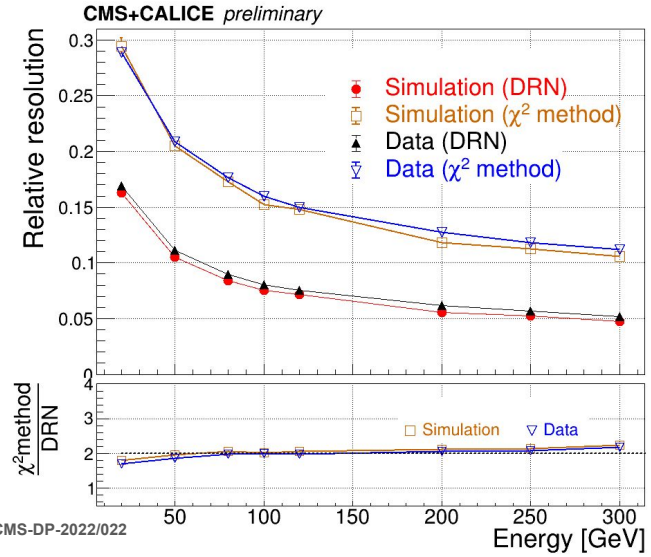
    - MIP-like in CE-E optimized separately
  - Very simple baseline
    - Doesn't know about event-by-event fluctuations
    - Ignores high granularity of detector

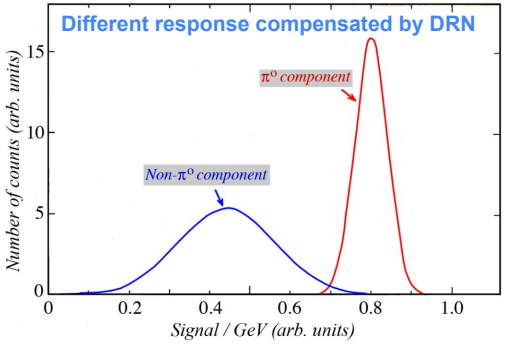




With Dynamic Reduction Network

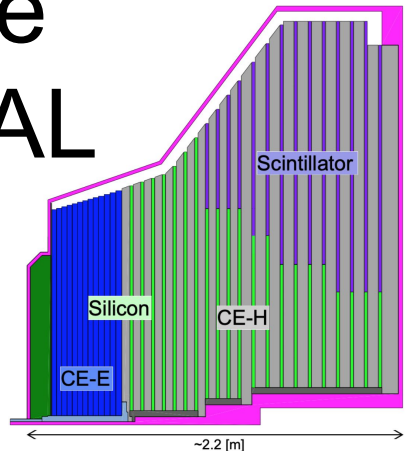






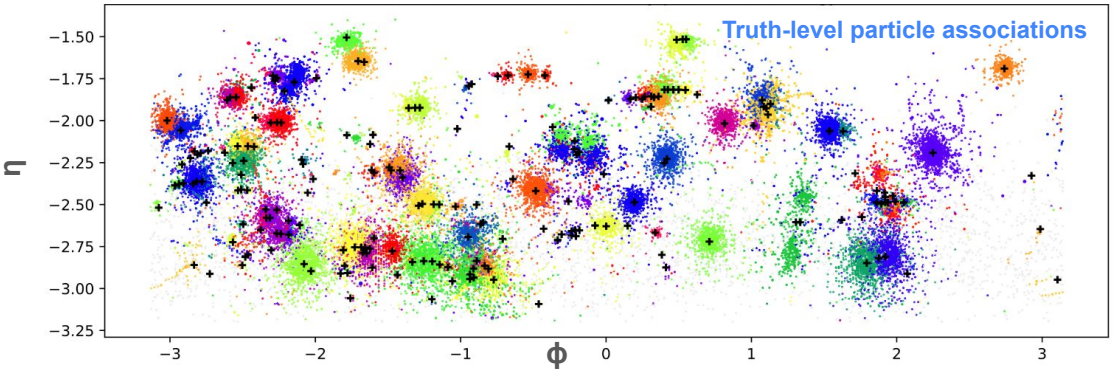
- Use same Dynamic Reduction Network as in ECAL regressions
- Dramatic improvement in energy resolution w.r.t. χ^2 method
  - DRN able to compensate for different response in EM and Hadronic shower components
  - Can correct for both longitudinal and transverse leakage

Case Study: Hit-to-particle reconstruction in the HGCAL



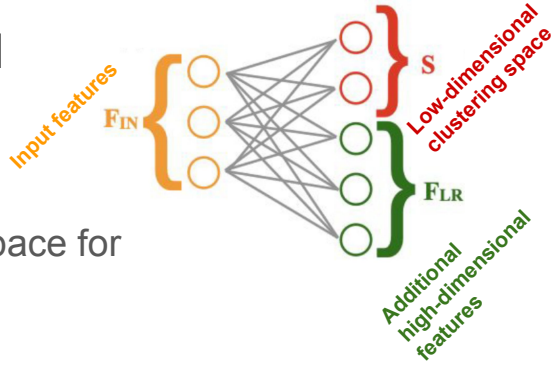
Problem statement

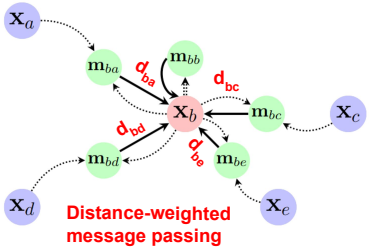
- HGCAL detector poses novel problems for event reconstruction
  - ~3 million readout channels per endcap
  - ~200,000 hits per event in HL-LHC conditions
- Need to perform “particle tracking” for showers
  - Not simple helical trajectories
- How do we go from >100,000 hits to collection of particles and their properties?
  - No viable pre-existing algorithm for this task



Graph-based approach

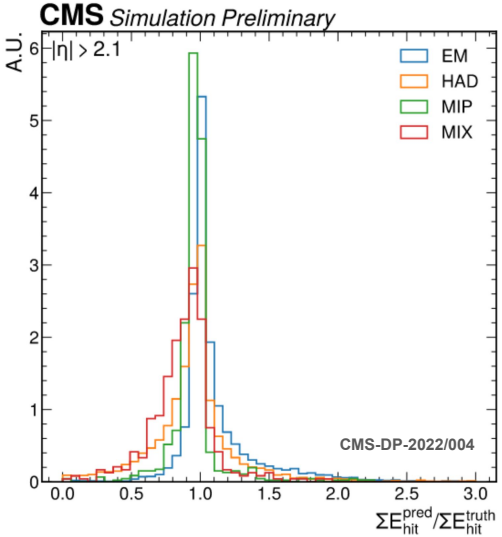
- Huge number of channels requires novel computational techniques
- Modified graph architecture[1]
  - High-dimensional information is projected into low-dimensional space for graph generation
  - Add distance weighting to graph message passing
- Novel loss function allows identification of arbitrary number of particles and reconstruction of their energies[2]
- Input: all HGCAL hits in a given event
- Output: clustering of hits into particles with corrected energies
  - Alternatively could apply separate dedicated corrections with e.g. DRN

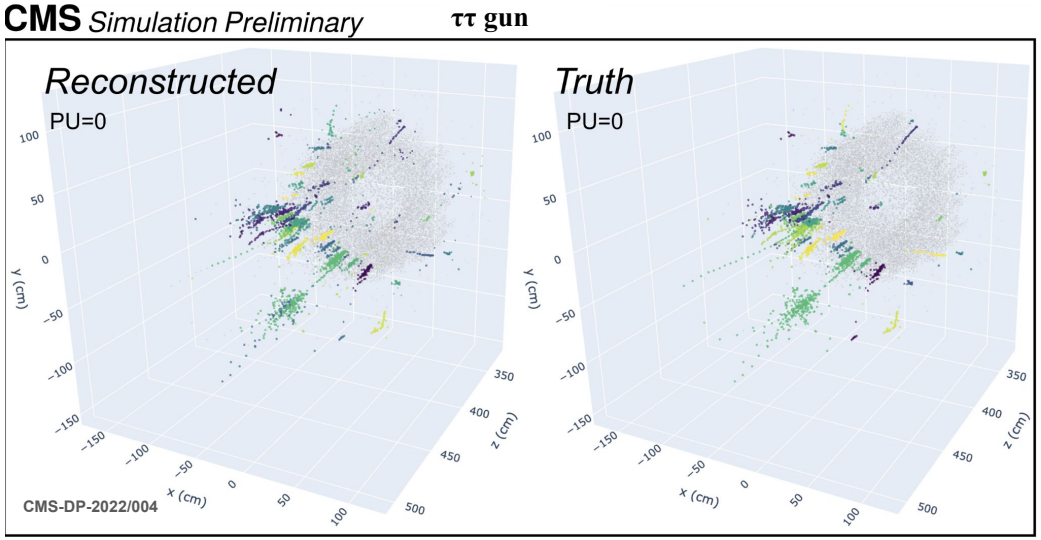




[1]: arXiv:2204.01681
[2]: arXiv:2002.03605

Performance





Gray hits are noise
Colored hits are due to individual incident particles

- Efficiently recovers hadronic and EM energy deposits
- Clear (qualitative) separation between particles
- Performance very good even in dense areas;
  - Expected to work well for pileup, substructure, etc

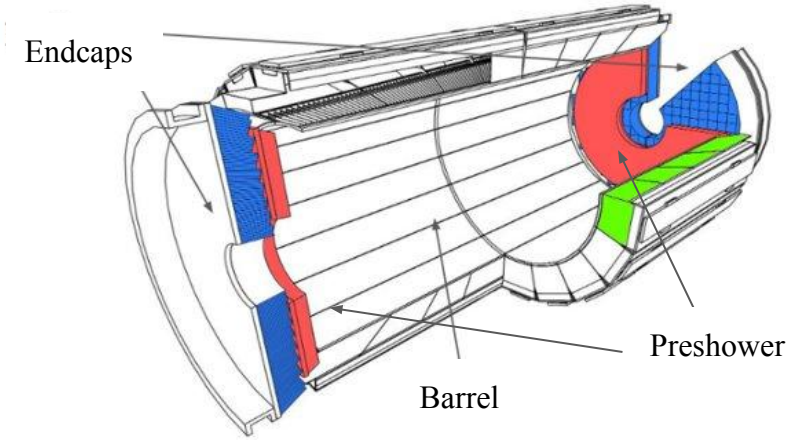
Conclusions

- Even the best analyses can only be as precise as their inputs
- Transition to training on low-level inputs allows new strategies and more precisely reconstructed objects
- The most effective way to incorporate low-level calorimeter hits into a machine learning model is with graph neural networks
- The three examples shown today are just some of many being used across CMS, in applications such as
  - Energy reconstruction
  - Hit clustering (see also Badder's talk from yesterday)
  - Particle identification
  - Jet substructure
- Methods are general and applicable to any detector

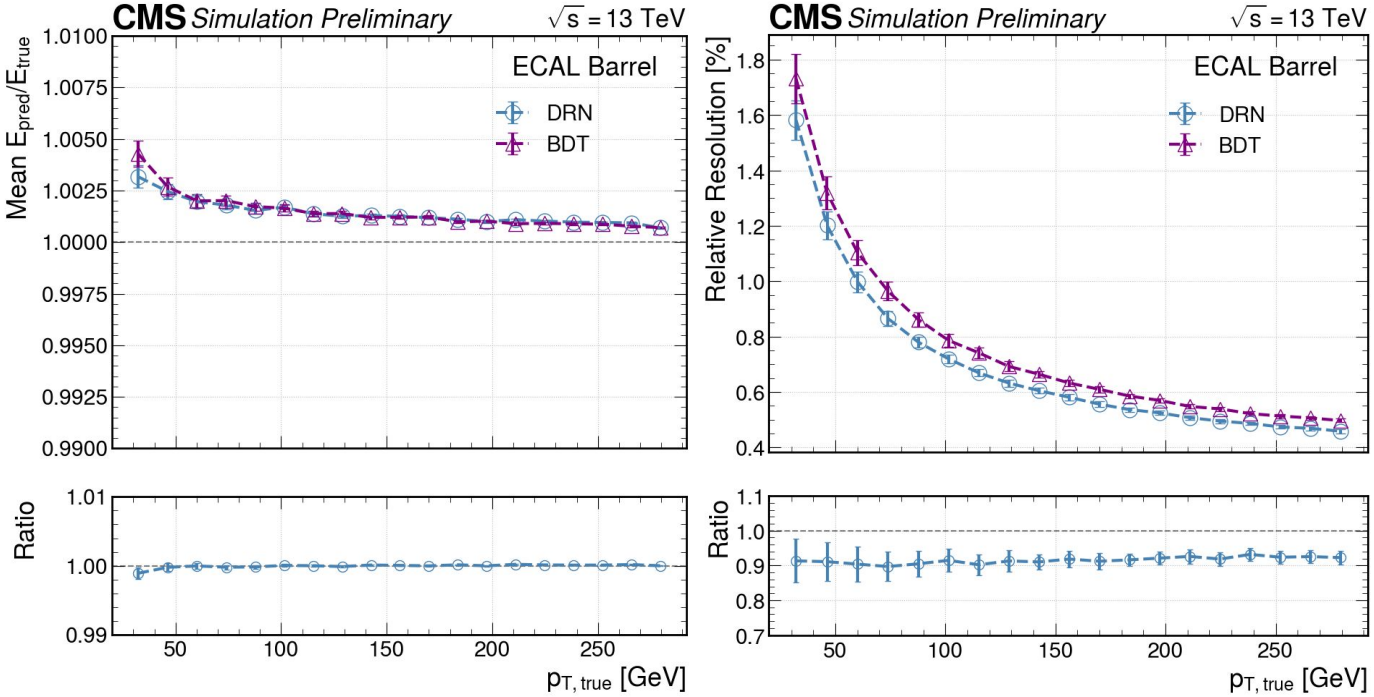
Backup

The CMS Electromagnetic Calorimeter

- Electron/photon energies measured in electromagnetic calorimeter (ECAL)
- Homogeneous calorimeter made of 75,848 scintillating $PbWO_4$ crystals
  - Divided into barrel (61,200 crystals) and two endcaps (7,324 crystals each)
  - Endcaps are preceded by sampling preshower calorimeter (ES)

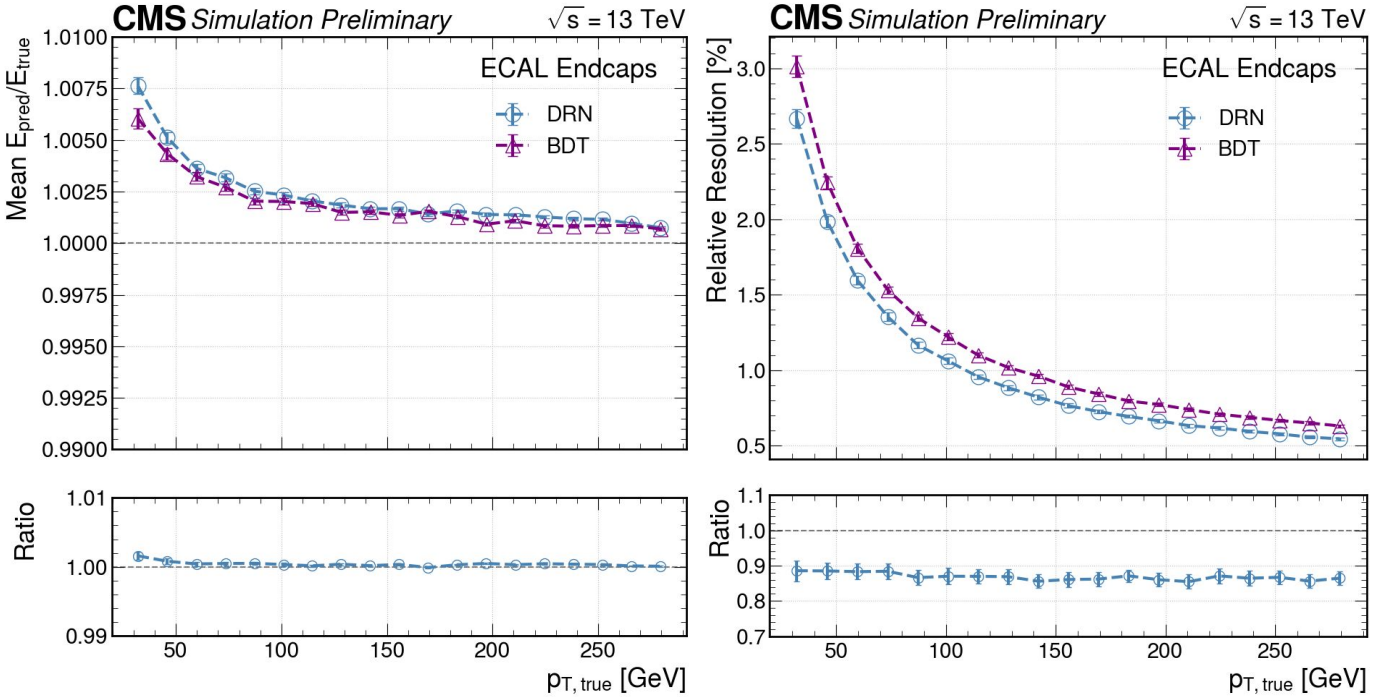


DRN electron regression (barrel)



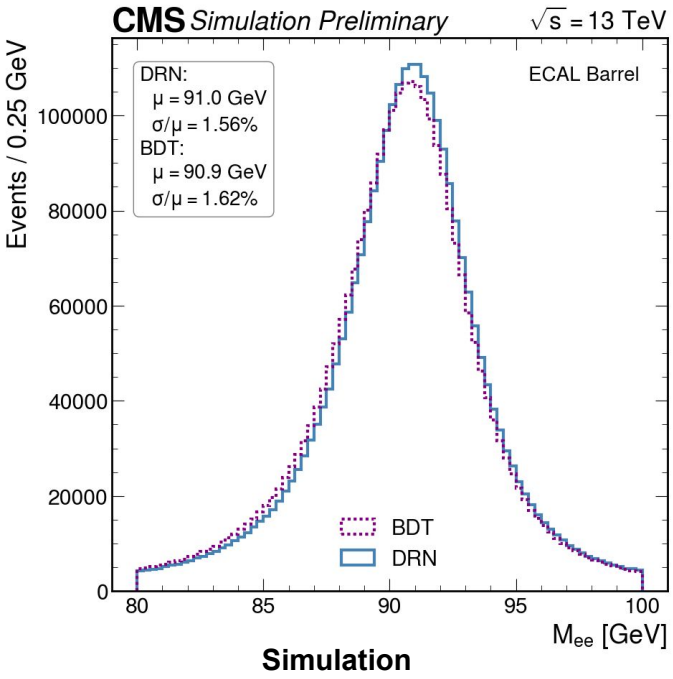
Dynamic Reduction Network (DRN) and Boosted Decision Tree (BDT) performance in the ECAL barrel as a function of generated transverse momentum. Performance is evaluated on electron gun simulation with ideal detector calibration. Error bars represent fitting uncertainties. **Left:** Mean response E_{Pred}/E_{True}. Regression response is stable to within better than 0.5%. **Right:** Relative resolution. The DRN obtains a better resolution than the BDT by a factor of ≈ 10% at all values of p_T.

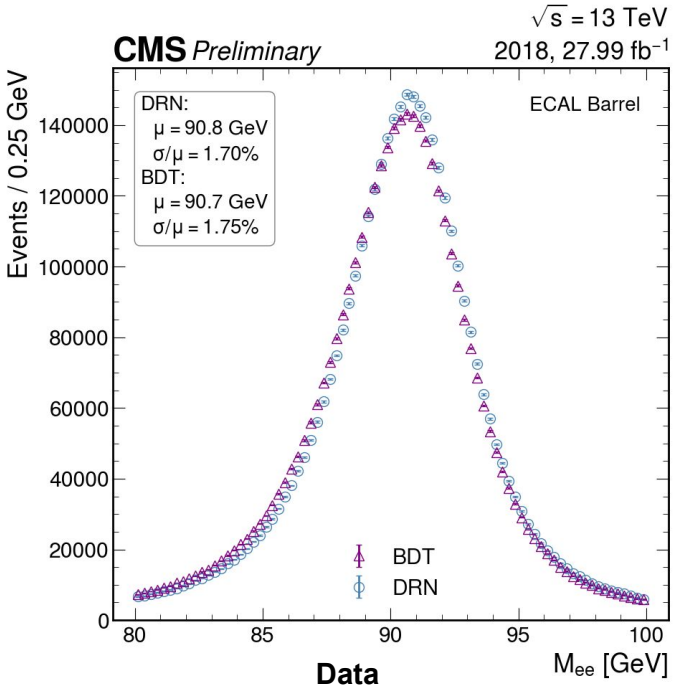
DRN electron regression (endcaps)



Dynamic Reduction Network (DRN) and Boosted Decision Tree (BDT) performance in the ECAL endcaps as a function of generated transverse momentum. Performance is evaluated on electron gun simulation with ideal detector calibration. Error bars represent fitting uncertainties. **Left:** Mean response E_{Pred}/E_{True}. Regression response is stable to within better than 0.75%. **Right:** Relative resolution. The DRN obtains a better resolution than the BDT by a factor of ≈ 10% at all values of p_T.

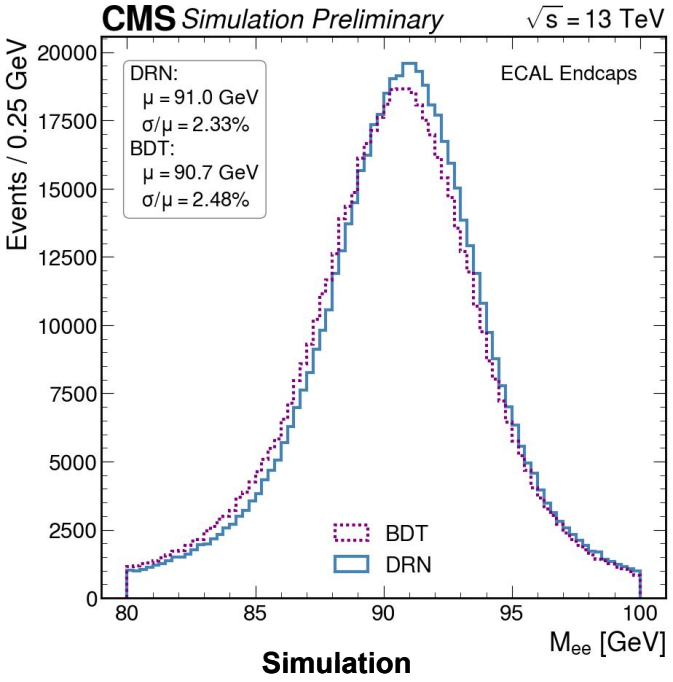
DRN electron regression (Z peak, barrel)

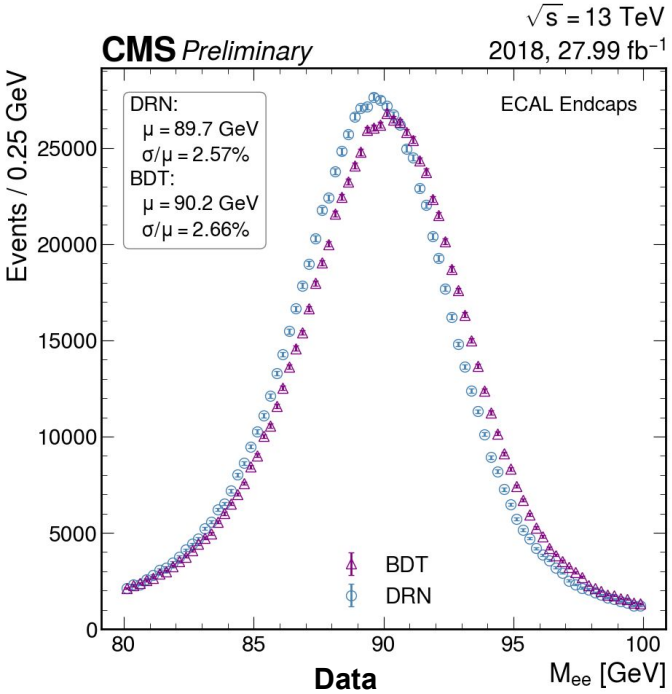




Di-electron invariant mass distributions of Z→ee events in simulation (**left**) and 2018 Legacy data (**right**) for both the Dynamic Reduction Network (DRN) and Boosted Decision Tree (BDT) architectures. Events are selected in which both electrons pass a tight ID requirement and are detected in the ECAL barrel. The DRN obtains an improved resolution with respect to the BDT by a factor of about 5% in both data and simulation. Note that in these results the residual corrections which match the energy scale in data to that in simulation have not been applied. However, we do see that in simulation the DRN obtains a Z peak closer to the known Z mass of 91.1 GeV.

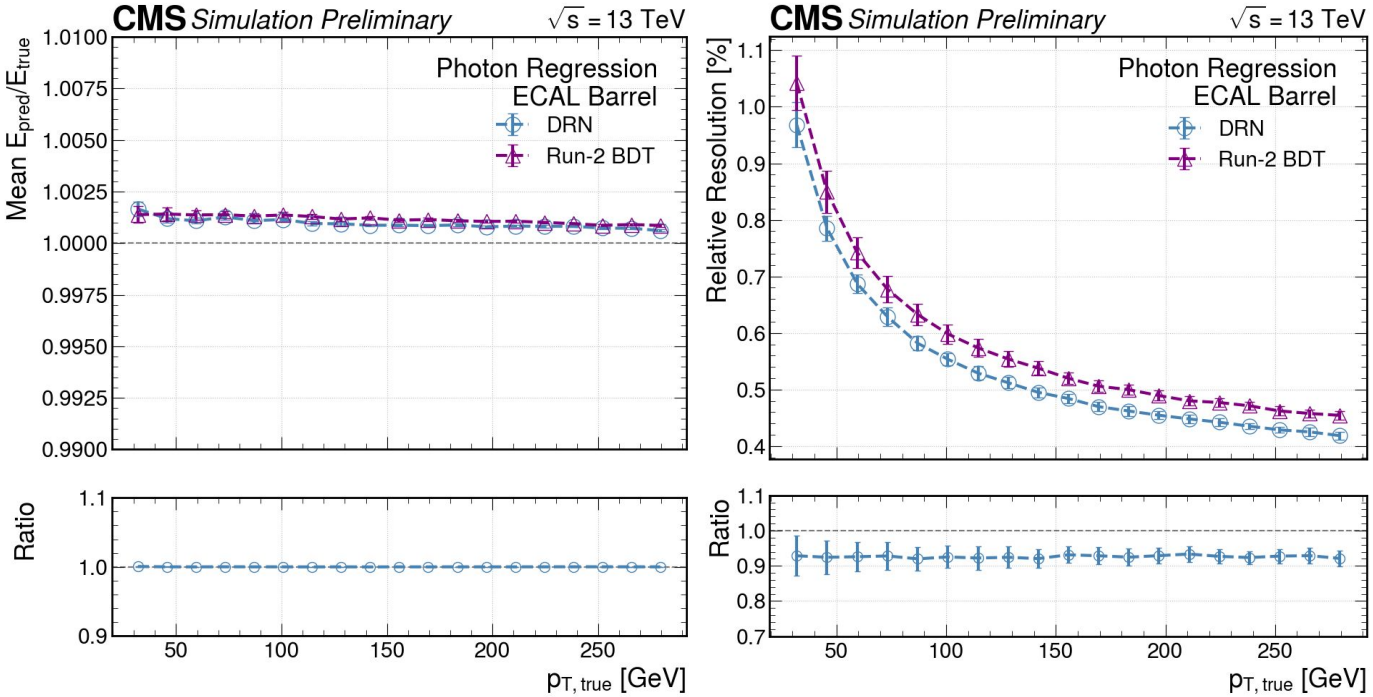
DRN electron regression (Z, endcaps)





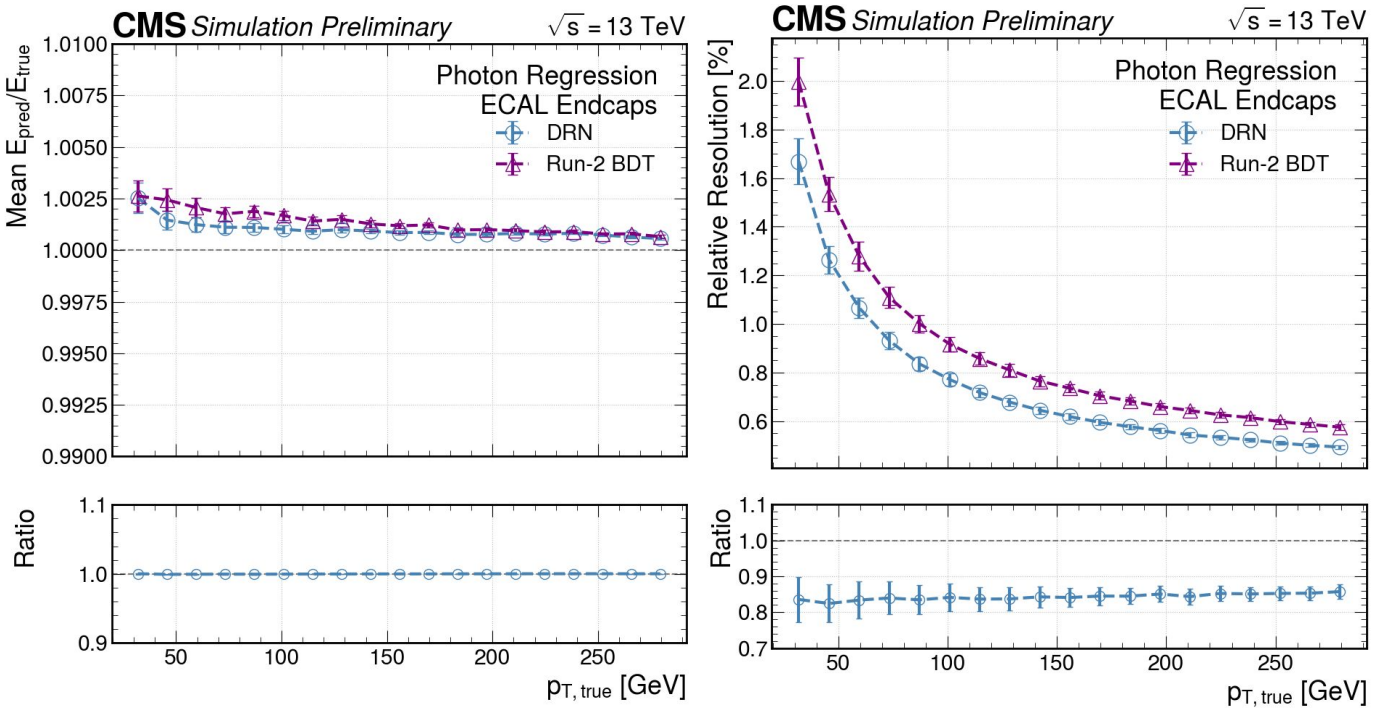
Di-electron invariant mass distributions of Z→ee events in 2018 Legacy simulation (**left**) and 2018 Legacy data (**right**) for both the Dynamic Reduction Network (DRN) and Boosted Decision Tree (BDT) architectures. Events are selected in which both electrons pass a tight ID requirement and are detected in the ECAL endcaps. The DRN obtains an improved resolution with respect to the BDT by a factor of about 5% in both data and simulation. Note that in these results the residual corrections which match the energy scale in data to that in simulation have not been applied. However, we do see that in simulation the DRN obtains a Z peak closer to the known Z mass of 91.1 GeV. The difference in energy scale between the data and simulation is likely related to differences between the detector conditions in data and simulation. This difference is larger in the endcaps than in the barrel.

DRN photon regression (barrel)



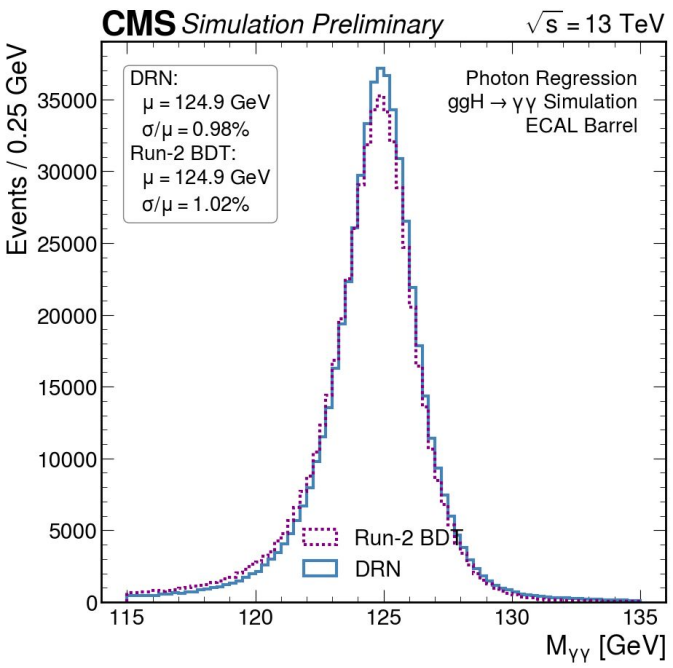
Dynamic Reduction Network (DRN) and Boosted Decision Tree (BDT) performance in the ECAL barrel as a function of generated transverse momentum. Performance is evaluated on photon gun simulation with ideal detector calibration. Error bars represent fitting uncertainties. **Left:** Mean response E_{Pred}/E_{True}. Regression response is stable to within better than 0.2%. **Right:** Relative resolution. The DRN obtains a better resolution than the BDT by a factor of ≈ 10% at all values of p_T.

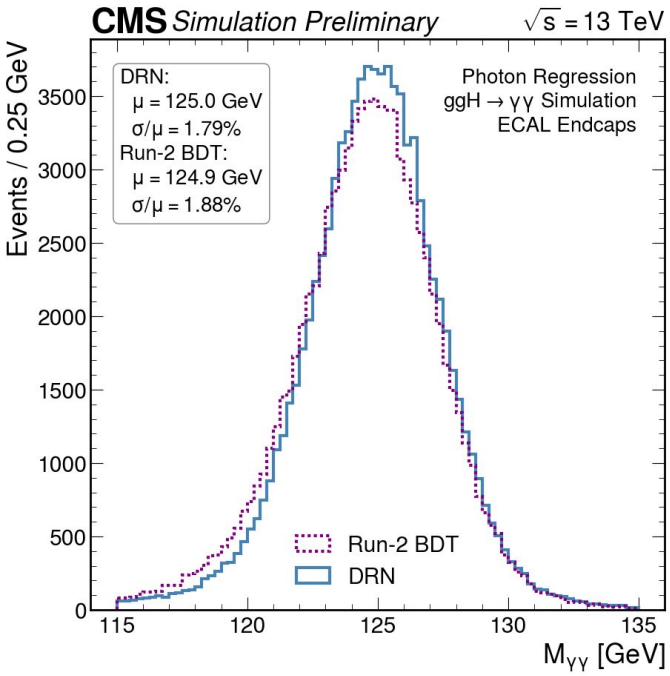
DRN photon regression (endcaps)



Dynamic Reduction Network (DRN) and Boosted Decision Tree (BDT) performance in the ECAL endcaps as a function of generated transverse momentum. Performance is evaluated on photon gun simulation with ideal detector calibration. Error bars represent fitting uncertainties. **Left:** Mean response E_{Pred}/E_{True}. Regression response is stable to within better than 0.4%. **Right:** Relative resolution. The DRN obtains a better resolution than the BDT by a factor of ≈ 15% at all values of p_T.

DRN photon regression (ggH→γγ)





Di-photon invariant mass distributions of H→γγ events in 2018 Legacy simulation for both the Dynamic Reduction Network (DRN) and Boosted Decision Tree (BDT) architectures. The Higgs peak is fit with a Cruijff function to parameterize the detector response and resolution. **Left:** events in which both photons are detected in the barrel region of the ECAL. **Right:** events in which both photons are detected in the ECAL endcaps. The DRN obtains an improved resolution with respect to the BDT by a factor of about 5% in both detector regions.

HGCAL testbeam

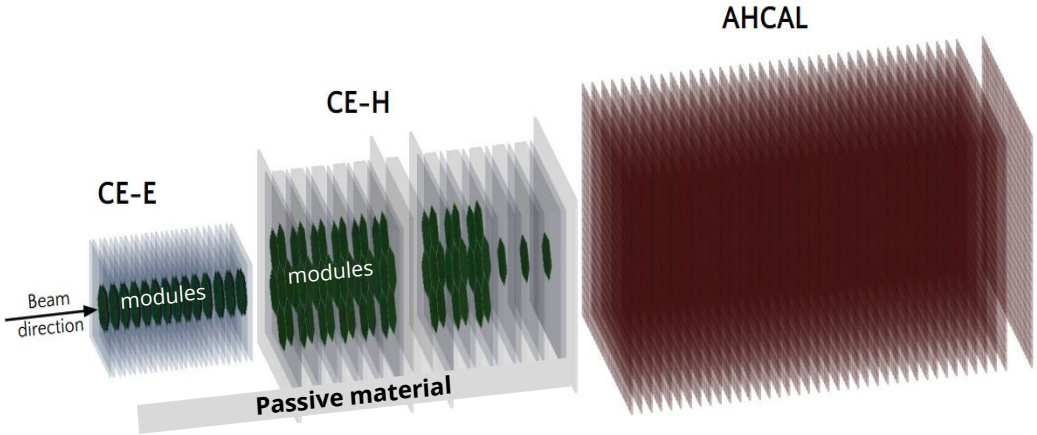
CE-E
- Si sensors
- Cu/CuW & Pb absorbers
- 28 sampling layers

CE-H
- Si sensors
- Cu/CuW & Steel absorbers
- 12 sampling layers

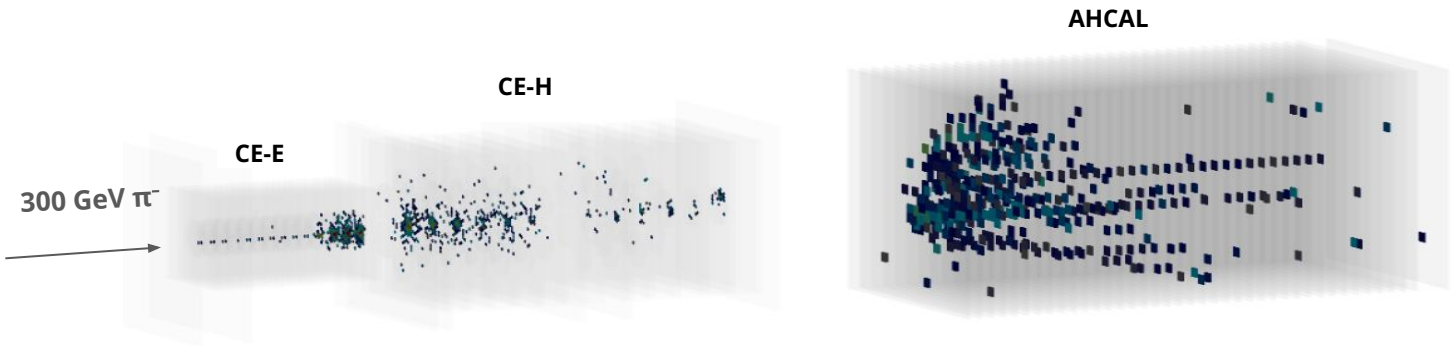
CALICE AHCAL
- Scintillator tiles on SIPMs
- Steel absorbers
- 39 sampling layers
  - Downsample to only every fourth layer to match final HGCAL geom

- Total of ~17k readout channels (after downsampling AHCAL)

- Testbeam Oct. 2018
  - At H2 Beamline, CERN
- Exposed to e^+ and π^- beams with energies ranging from 20–300 GeV



Example pion shower event display



Pion shower in HGCAL prototype testbeam. Each square is a detector hit

Effect of different DRN input feature sets

- HGCAL testbeam DRN trained with different hit coordinates
  - DRN(E) = trained only with hit energies
  - DRN(E, z) = trained with energy and z coordinate (can learn longitudinal shower development)
  - DRN(E, x, y, z) = trained with energy and full 3D hit coordinates (full view of shower)
- Major improvement comes from DRN(E) compensating EM/hadronic response