ML based reconstruction techniques for CMS HGCAL

Alpana[1] and Rajdeep M Chatterjee[2]
On behalf of the CMS Collaboration

[1]Indian Institute of Science Education and Research, Pune
[2]Tata Institute of Fundamental Research, Mumbai

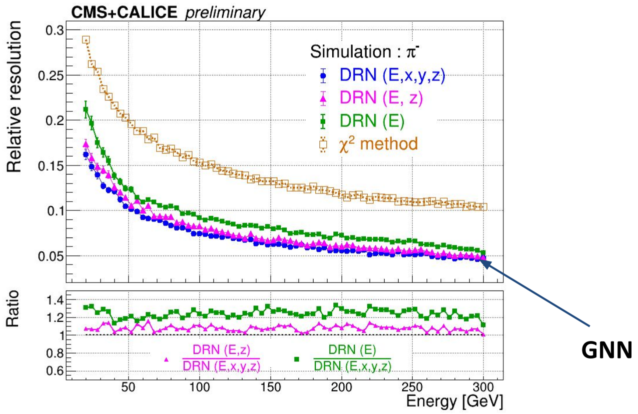
Technology in Instrumentation and Particle Physics Conference
Cape Town, South Africa
4 - 8 Sept 2023

Outline

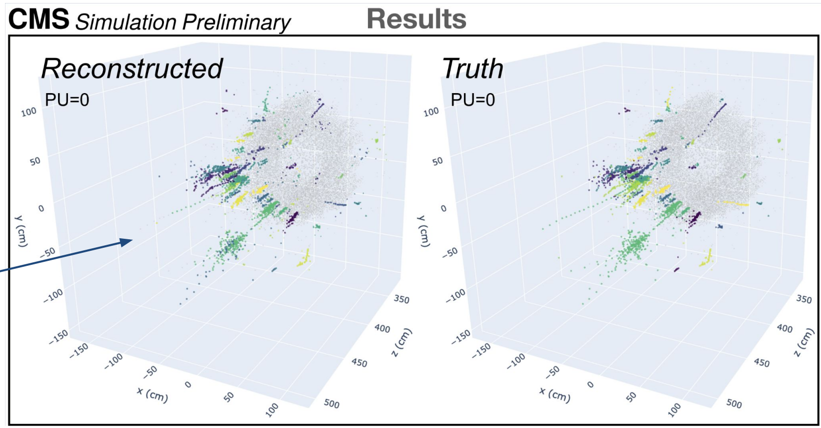
- The High Luminosity LHC and the CMS High Granularity Calorimeter (HGCAL)

- Two examples of Graph Neural Network based reconstruction for HGCAL:
  - Pion energy reconstruction in the prototype HGCAL testbeam (CMS-DP-2022/022)
  - Hit-to-particle reconstruction in the CMS HGCAL (CMS-DP-2022/004)
  - See Polina's talk tomorrow for additional applications in CMS

- Summary



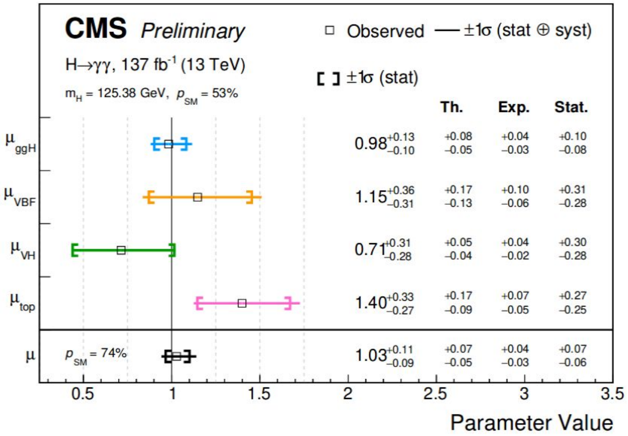
Graph Network



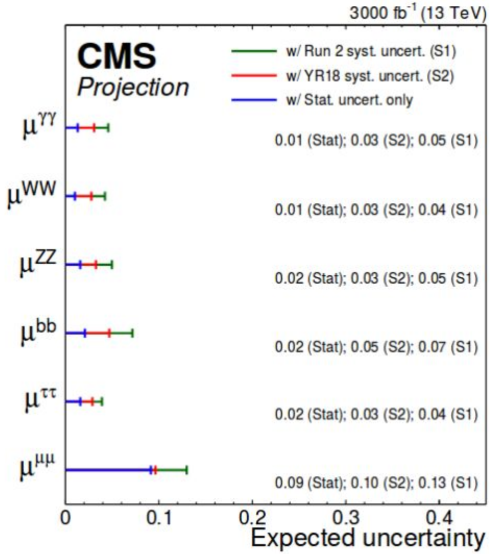
The HL-LHC: Physics motivation

- Runs 1 and 2 of the LHC have yielded a rich harvest of physics results (> 1000 papers) :
  - From the discovery of the Higgs boson to a detailed study of its properties with high precision.
  - Observation of rare decays like $B^0_s \rightarrow \mu^+\mu^-$ and rare processes like heavy triple boson production

- However a decisive increase of the LHC luminosity in order to meaningfully improve on the current results in a reasonable timeframe:
  - O(1%) precision on SM Higgs couplings
  - Rare Higgs Decays (e.g $H\rightarrow\mu\mu$) and production (e.g. HH)
  - Extending reach of BSM searches.

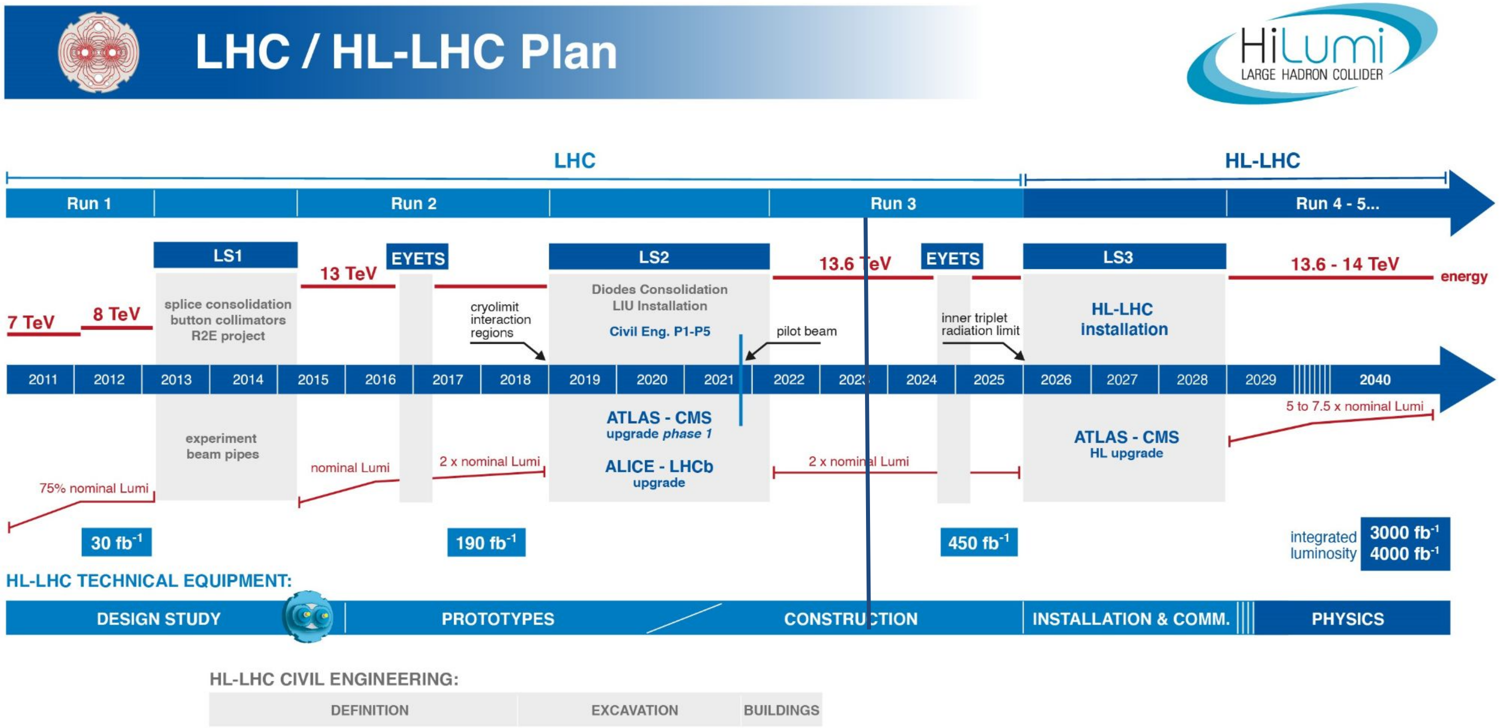


JHEP 07 (2021) 027



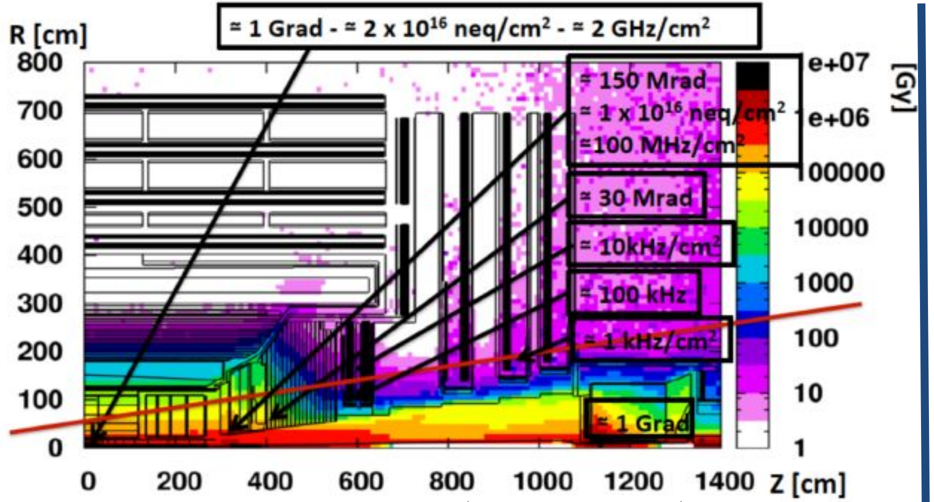
CMS-PAS-FTR-18-011

The High Luminosity LHC



The HL-LHC challenges

The HL-LHC environment presents unprecedented challenges to the CMS physics program of searching for New Physics through Precision Measurements and Direct Searches for Rare Processes.

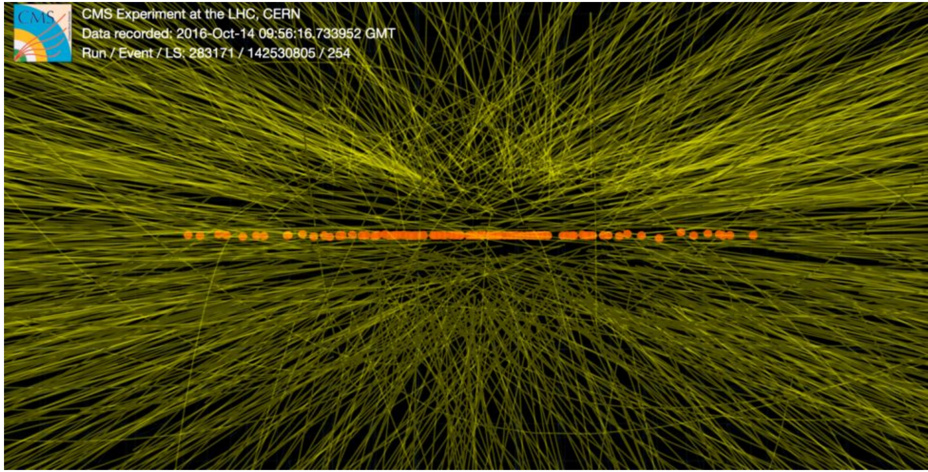


Simulation: Total of 3 ab^{-1} with 300 fb^{-1}/year

HIGH RADIATION(due to high integrated lumi.)

- Radiation levels up to $2x10^{16}$ n_{eq}/cm^2 or 1 Grad in the forward region or close to the collision point

A typical event from a 2016 High PU (<μ> = 100) run



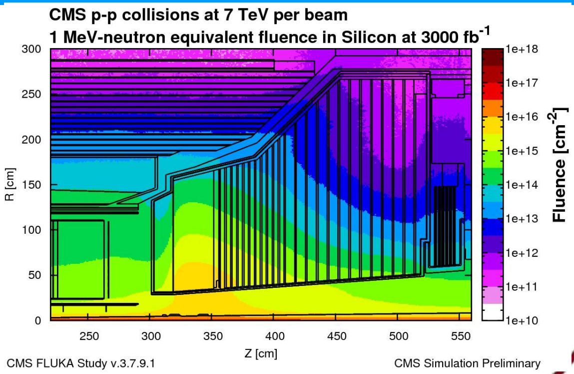
HIGH PILEUP(due to high instant. lumi.)

- Multiple collision per event: 140--200

The CMS High Granularity Calorimeter

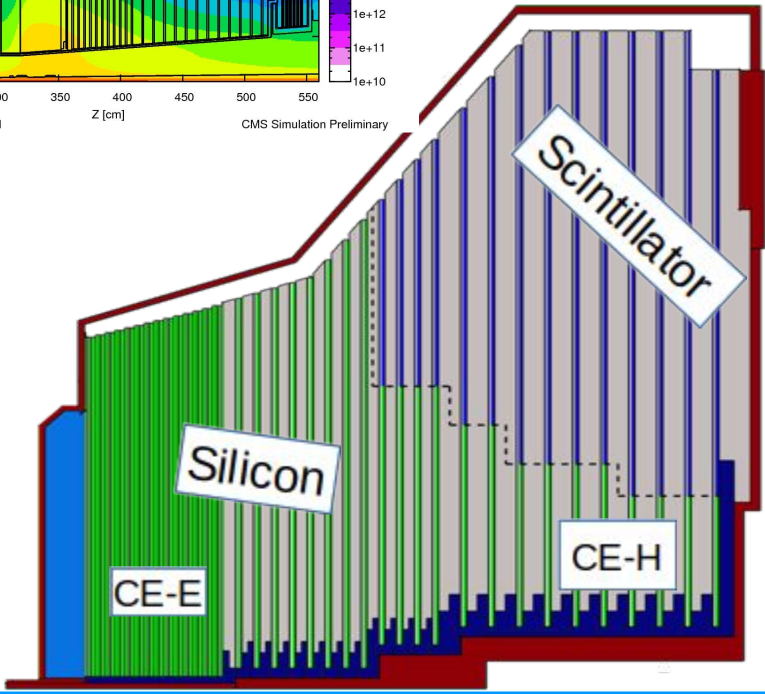
Key Parameters:

- The HGCAL covers $1.5 < |\eta| < 3.0$
- 215 ton/endcap, full system at -30 C
- **620 m^2 of silicon sensors** in ~27k modules with **6M Si channels**
- **400 m^2 of scintillators** in 4k boards with **240k scintillator channels**
- **Enables 5D reconstruction of particle showers**



Active Elements:

- Hexagonal modules based on Si sensors in CE-E and high-radiation regions of CE-H
- Scintillating tiles with SiPM readout in low-radiation regions of CE-H



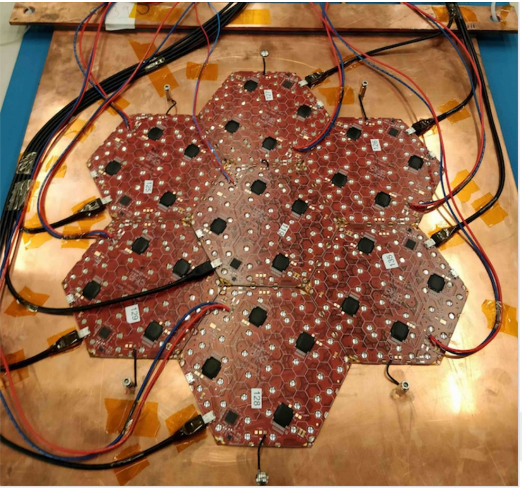
Detector Configuration:

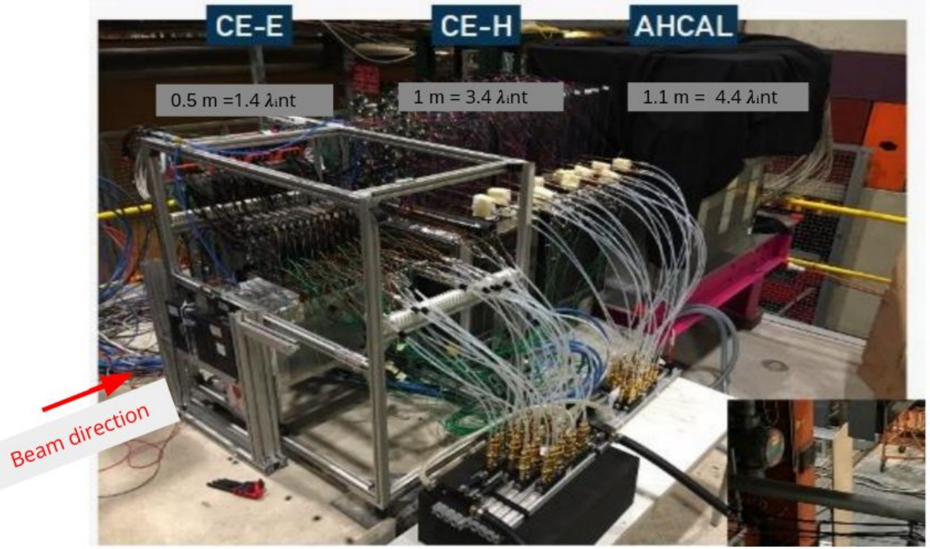
- Electromagnetic calorimeter (CE-E) : Si, Cu/CuW/Pb absorbers; **26 layers, ~28 Xo and ~1.5 λ**
- Hadronic calorimeter (CE-H) : Si & scintillator, steel absorbers; **21 layers and ~8.5 λ** (including CE-E)

Pion reconstruction with the HGCAL beam test prototype

Experimental setup of HGCAL detector prototype

- Beam test experiments of Oct 2018 at H2 Beamline, CERN
- Prototype HGCAL detector setup comprised of Si-based electromagnetic (CE-E) and hadronic (CE-H) sections followed by scintillator tile-based CALICE AHCAL
  - Exposed to e^+ and π^- beams of energies ranging from 20 – 300 GeV





For more details about instrumentation, DAQ, calibration, and simulation, please refer to 2021 JINST 16 T04001, 2021 JINST 16 T04002 and 2022 JINST 17 P05022.

Detector setup and simulation

The HGCAL beam test prototype, Si HGCAL & Scint. AHCAL, comprised of a total of ~12k and ~22k readout channels respectively. A detailed simulation of the detector prototype including the key beamline elements was developed in GEANT4.10.4.p03

CE-E

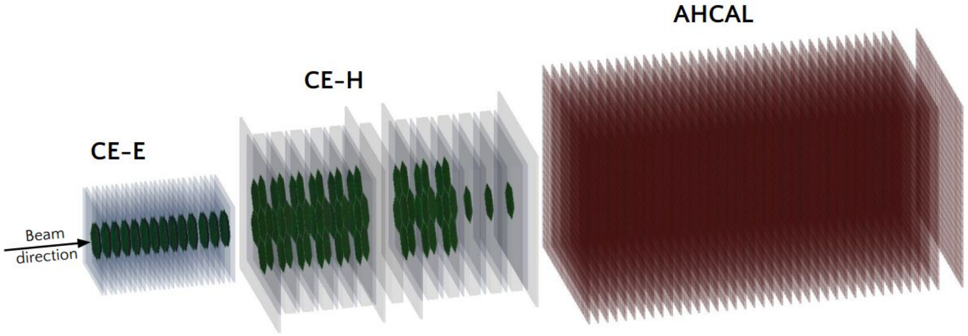
- Si sensors (1.1 cm²) + Cu/CuW & Pb absorbers. One 6' hexagonal module (128 cells) per layer, in a total of 28 layers

CE-H

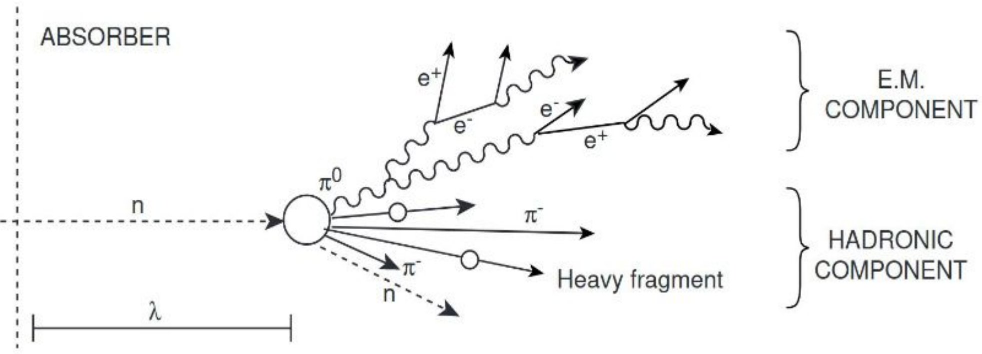
- Si sensors (1.1 cm²) + Cu/CuW & Steel absorbers. 12 sampling layers, 7 modules per layer, arranged in a daisy structure

CALICE AHCAL

- Scintillators on SiPMs ($3 \times 3 \times 0.3$ cm³) + Steel absorbers. 39 sampling layers



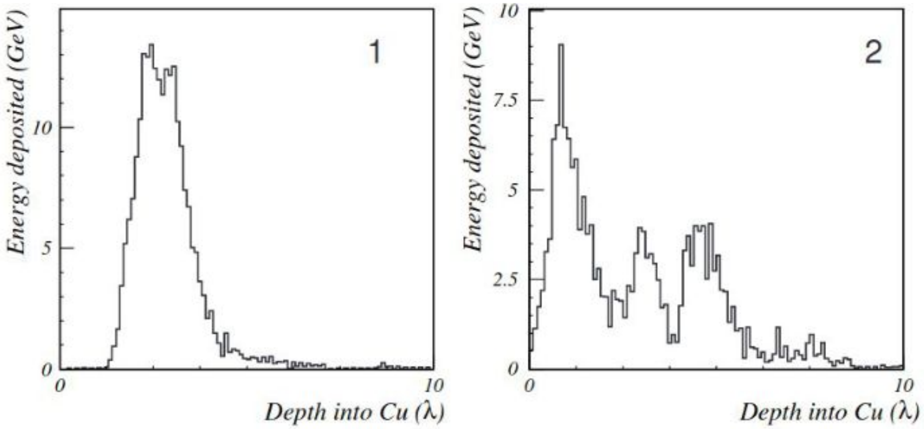
Complex nature of hadronic showers



- Hadronic showers have electromagnetic (EM) ($\pi_0/\eta \rightarrow \gamma\gamma$) & hadronic component.
- The fraction of energy carried by EM component depends on energy of incident hadrons.
  - results in a nonlinear response.
- The hadronic component has some definite contribution from invisible energy (breaking up of nuclei etc.) .

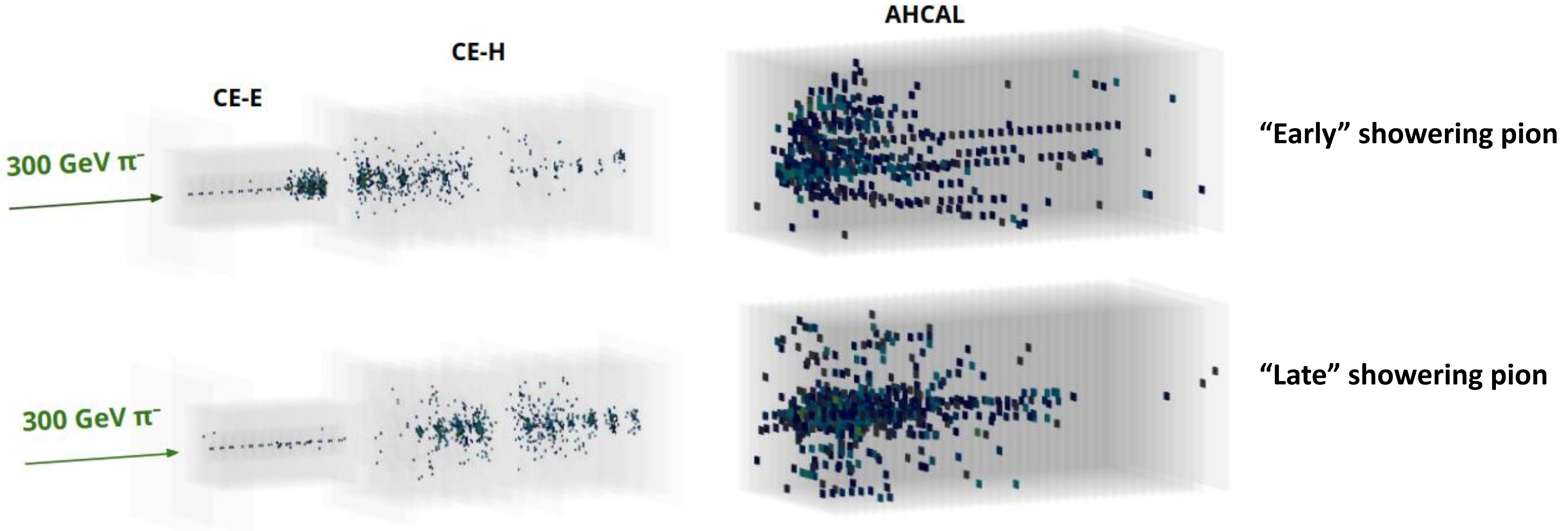
- The right figure shows the simulated development of two showers induced by 270 GeV pions in a copper block.

- Fluctuations in energy deposited across the detector are due to fluctuations in the production & the energy carried by the π_0s



Paper by T.S.Virdee

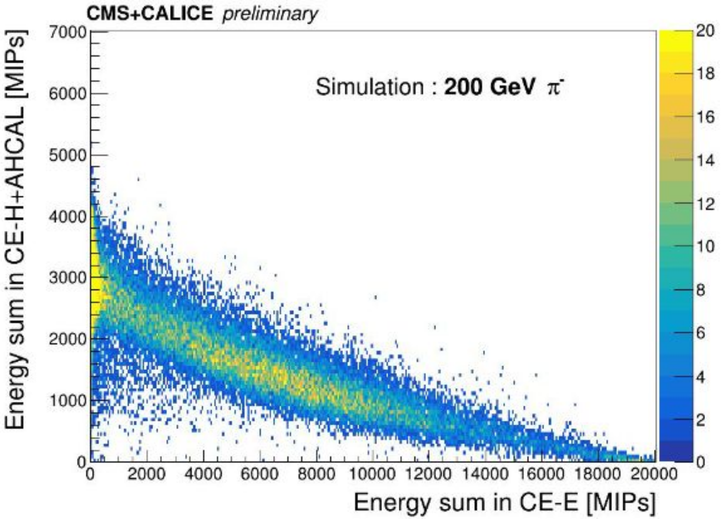
Event displays of pion showers in HGCAL beam test prototype



In addition to intrinsic fluctuations, transverse and longitudinal leakage of energy also contribute to fluctuations of the measured energy of pions.

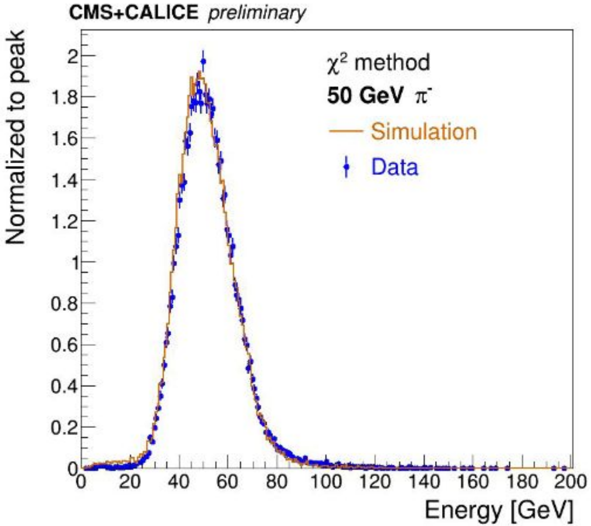
– Transverse leakage is mostly due to single modules used in CE-E and last three layers of CE-H.

Energy reconstruction of pions - classical method



The signal generated in each cell by the traversing particles is converted into energy in units of MIPs

- MIP scale is not uniform across the CE-E and CE-H/AHCAL because of different absorbers & sampling fractions.
- Detector level calibration: calibrate CE-E (CE-H+AHCAL) using a 50 GeV e^+ (a 50 GeV π^-) beam.



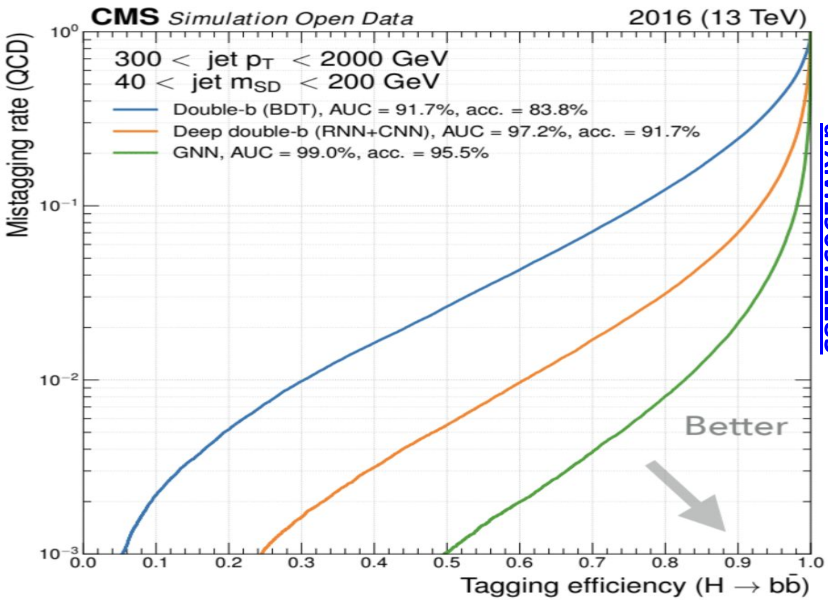
- χ^2 method : energies deposited in CE-E, CE-H and AHCAL are combined using energy dependent weight factors extracted after minimizing an estimator defined as,

$$\chi^2 = \sum_{pions} \frac{(E_{beam} - E_{corr})^2}{\sigma^2(E_{fix})}$$

- The χ^2 method in this form does not take into account the event-by-event fluctuations of showers and also it does not make use of the high granularity of the detector.

The machine learning approach

- ML is most powerful when applied on low-level inputs and has recently evolved in efficiently using them
  - Gives model access to full information content of every event
  - Avoids potential for biases from human feature engineering
- This has been demonstrated already in CMS e.g. jet tagging*
  - Train on jet constituents rather than high-level variables
- Hence we would like to use low-level calorimeter inputs, the detector hits, as input features



arXiv:1909.12285

*First with deepJet: arXiv:2008.10519

Different ML architectures : pros and cons

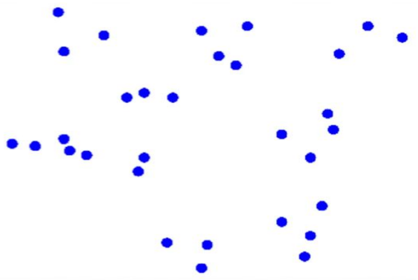
- Inputs can be challenging for most architectures
  - There can be any number of hits (Due to zero-suppression, not all channels are active)
  - They can be distributed across multiple very different detector components
  - They are naturally represented in at least 4 dimensions (x, y, z, energy), more recently 5 dimensional hits (x, y, z, energy, **time**) are in use
  - They are in no particular order

Can it (easily) handle…	BDT	MLP	CNN	RNN	GNN
Variable-size input	X	X	✅	✅	✅
Complicated geometries	✅	✅	X*	✅	✅
4D or 5D inputs	✅	✅	X*	✅	✅
Unordered inputs	X	X	✅	X	✅

- <u>GNNs are the best option for all input types</u>

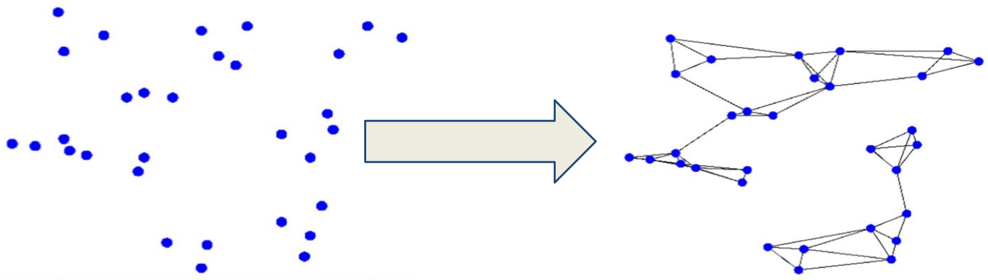
***CNNs work best with rectangular input spaces**

1. Collection of hits is represented as a point cloud



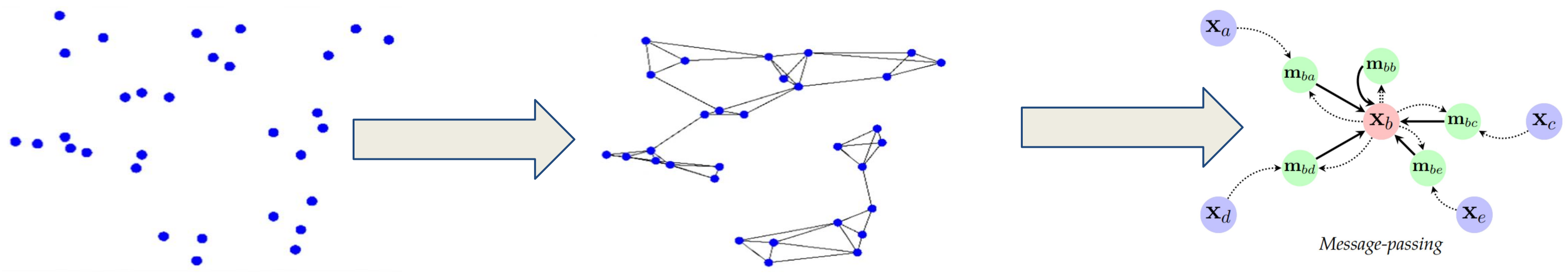
How do GNNs work

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

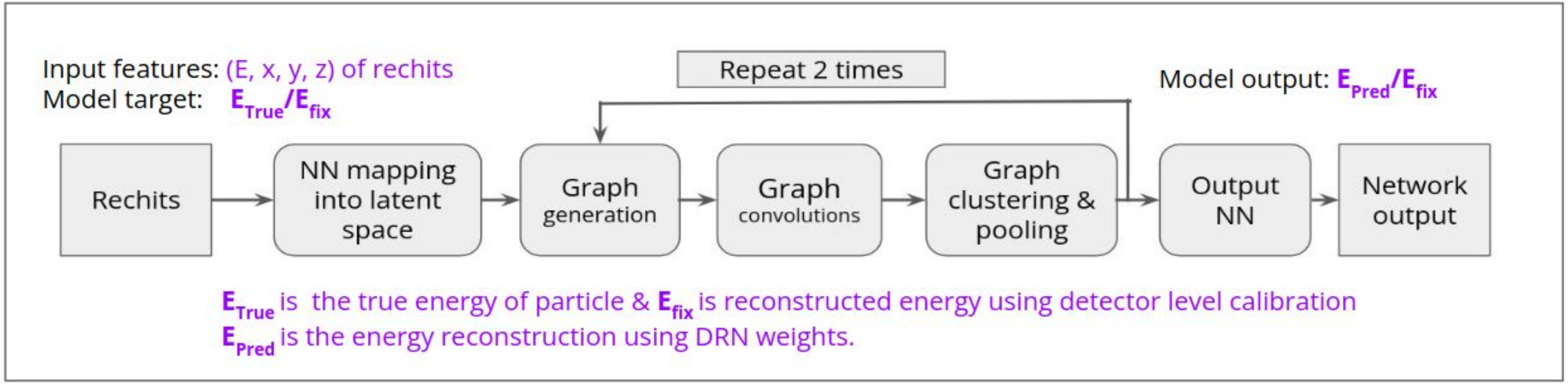


How do GNNs work

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)



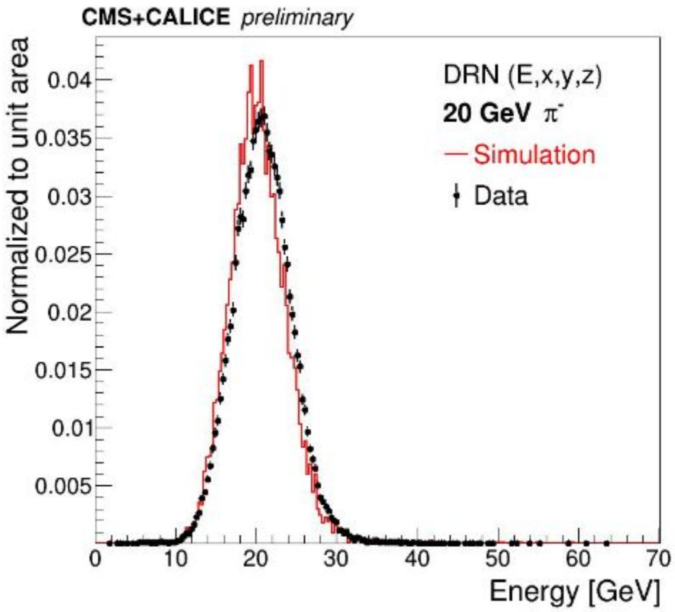
The Dynamic Reduction Network

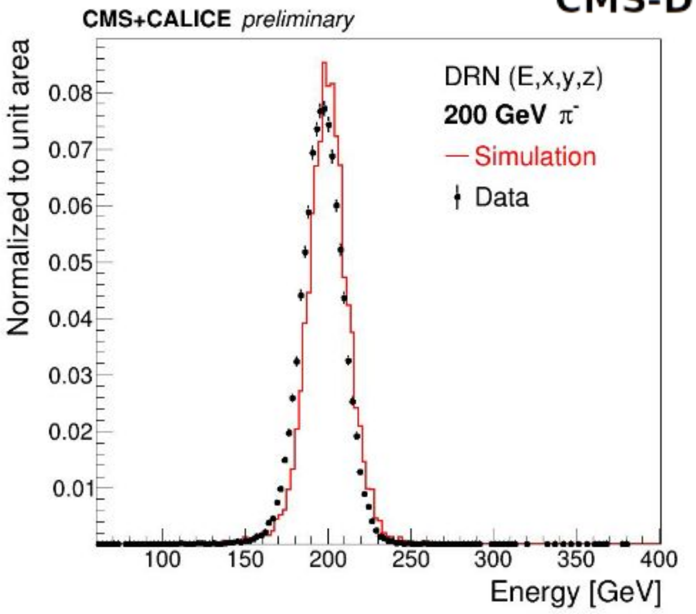


- Based on Dynamic Graph Neural Networks (1801.07829), a model is defined with the following differences(2003.08013v1)
  - Input features are mapped onto a higher dimensional latent space
  - Add clustering & pooling step to learn high level information iteratively
- The model is **trained on a flat energy sample of 10-350 GeV with a total of 4.1M events** simulated using GEANT4.10.4.p03 and FTFP_BERT_EMN hadronic physics lists.
- AdamW optimizer with a constant learning rate of 10^{-4} is used while training the model & **a total of 63k parameters** to learn in the model.
- The most time consuming step is the graph convolutions, followed by construction of the NN graph

Energy reconstruction of charged pions using the DRN

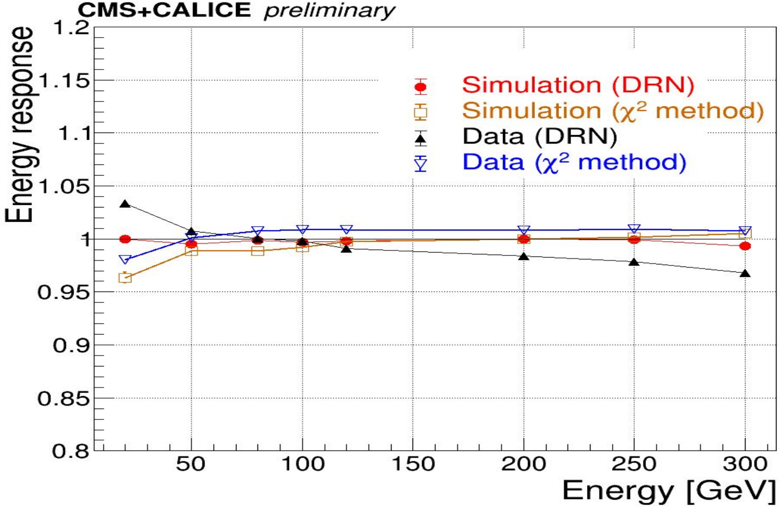
CMS-DP-2022/022

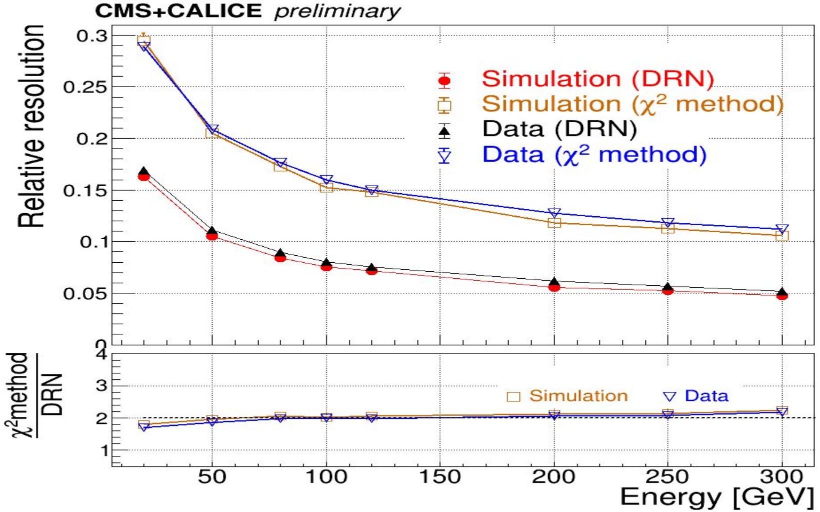




- In data, rechit energy in CE-E is scaled by 3.5% and in CE-H/AHCAL by 9.5% to account for the difference in energy scales in data and simulation (More details in this paper)
- The bulk of distribution between simulation & data are in fair agreement

Energy resolution & response

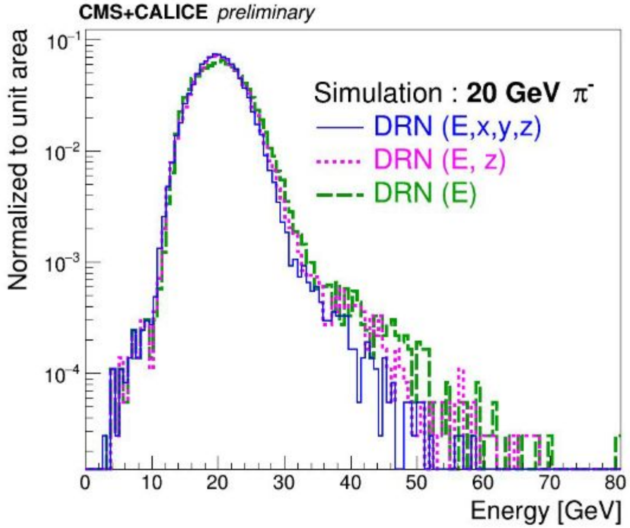


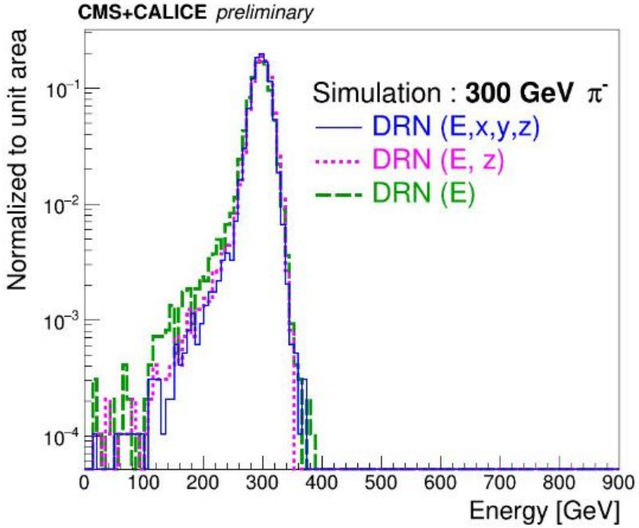


- **Dramatic improvement in energy resolution w.r.t. χ^2 method**
  - The scale factors in the χ^2 method are calculated on average for a given pion energy and applied to all events with the same true energy irrespective of the shower fluctuations

- Good agreement between the DRN predictions of the energy response and the relative resolution between beam test data and simulation

Training the DRN with different input features

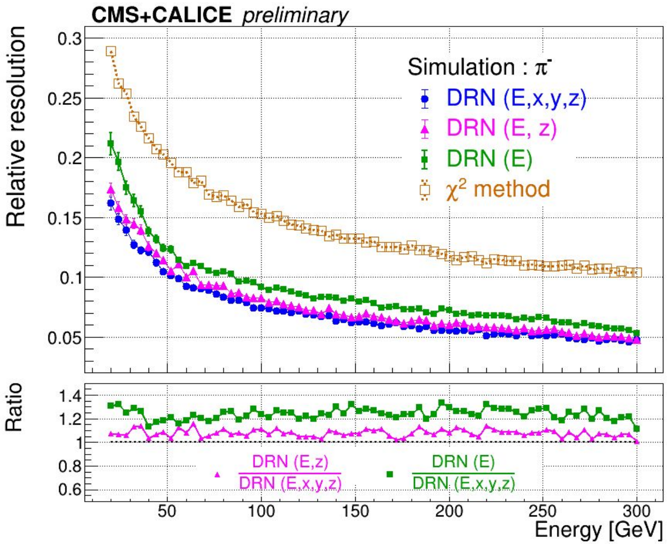


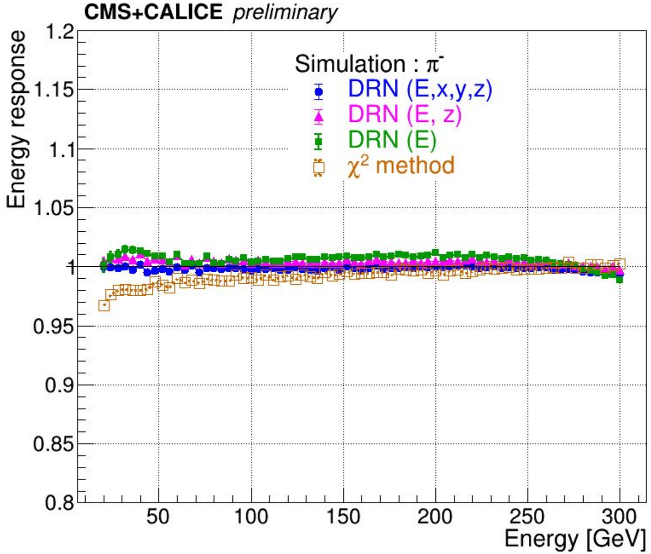


The role of input features used to train the DRN were investigated by considering three combinations of these input information:

- DRN (E) uses the energy values of all rechits as input feature
- DRN (E, z) uses in addition the z coordinate of the rechits to allow the DRN to track the longitudinal development of each shower
- DRN (E, x, y, z) using 4 input features (E, x, y, z) has full information about the longitudinal and transverse development of the showers.

Improvement in energy resolution with different DRN trainings

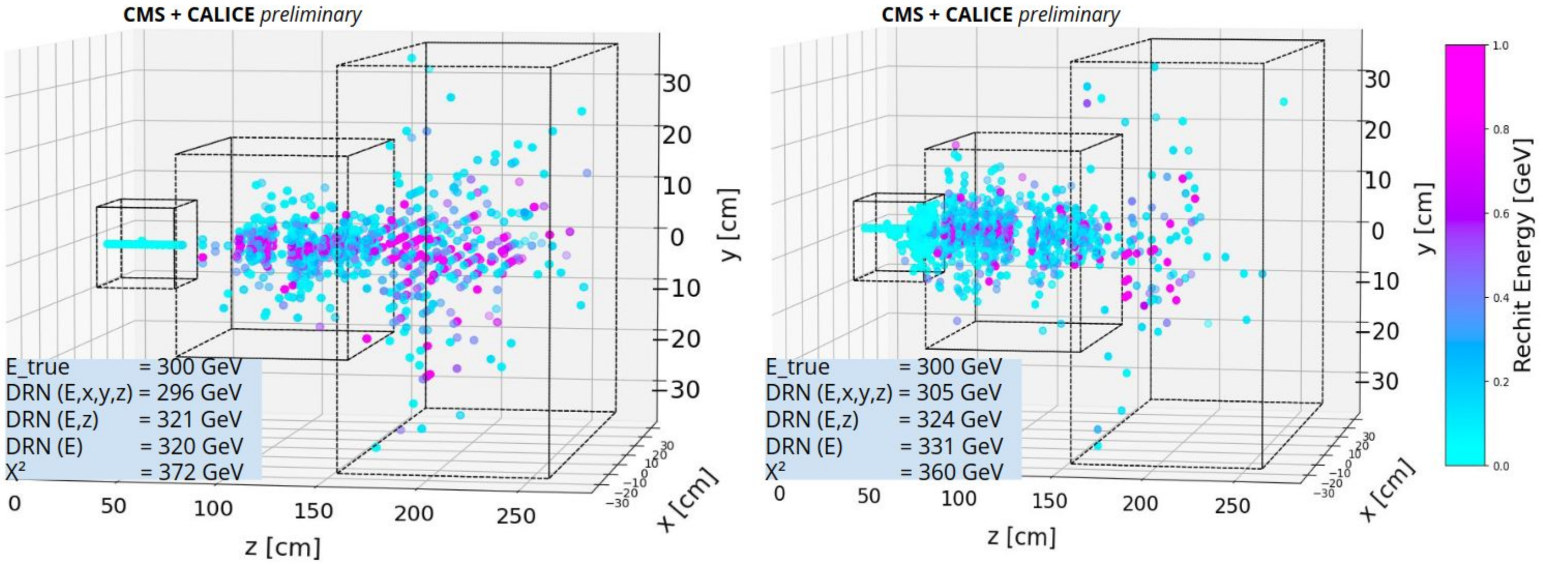




To understand the massive improvement in energy resolution with the DRN additional trainings were performed with different input features.

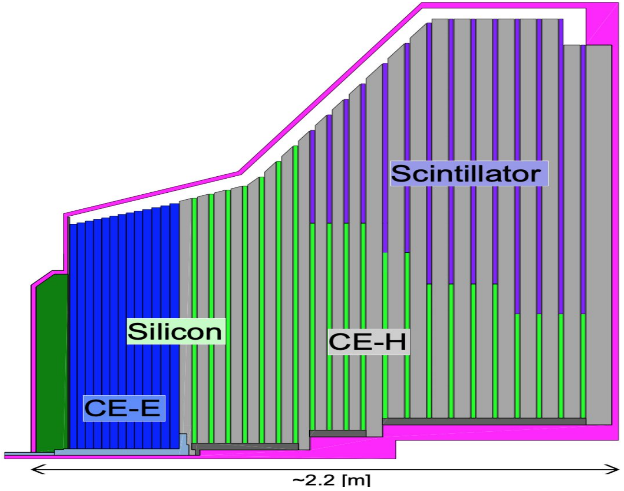
- The improvement from χ^2 method to DRN (E) comes from learning event-by-event fluctuations in EM fraction.
- Adding the spatial coordinates gives the DRN information about the spatial development of the shower → **Helps the DRN to better compensate for per event fluctuations and also learn about transverse leakage along with longitudinal leakage.**

Illustration of what the DRN learns with event displays



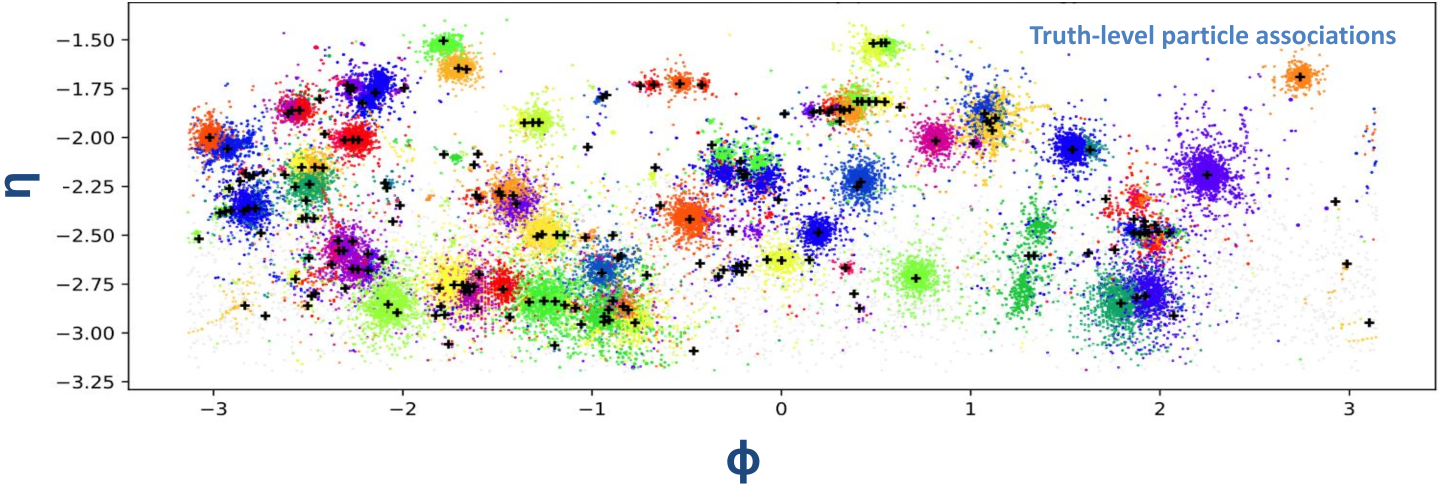
- Event displays for two representative simulated showers with the hits plotted according to their (x,y,z) position.
  - The color represents the rechit energy in units of GeV using the detector-level calibration. Hits with >1 GeV are represented with the same color as 1 GeV.

CMS HGCAL hits → particle reconstruction



Complexity of the problem

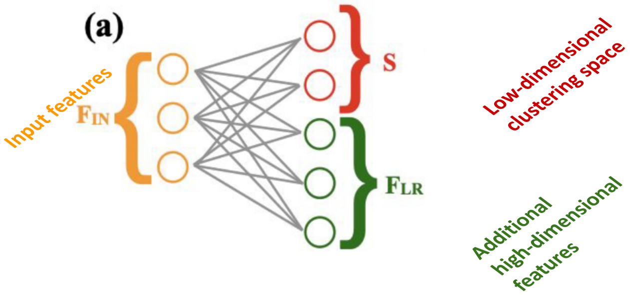
- HGCAL detector poses challenging problems for event reconstruction
  - ~3 million readout channels per endcap
  - ~200,000 hits per event in HL-LHC conditions
- Need to perform “particle tracking” for showers
  - Not simple helical trajectories
- How do we go from these >100,000 hits to collection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

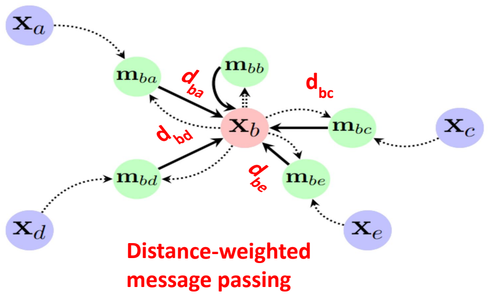


Graph-based approach

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 their energy reconstruction[2]
- Input: all 5D HGCAL hits (x, y, z, energy, **time**) in a given event
- Output: clustering of hits into particles with corrected energies
  - Alternatively could apply separate dedicated corrections with e.g. DRN

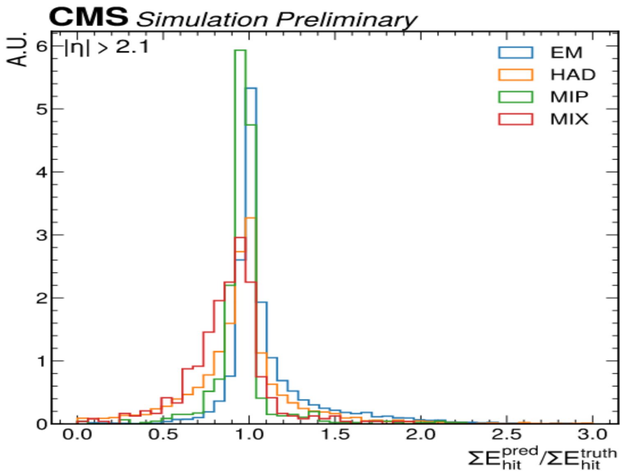


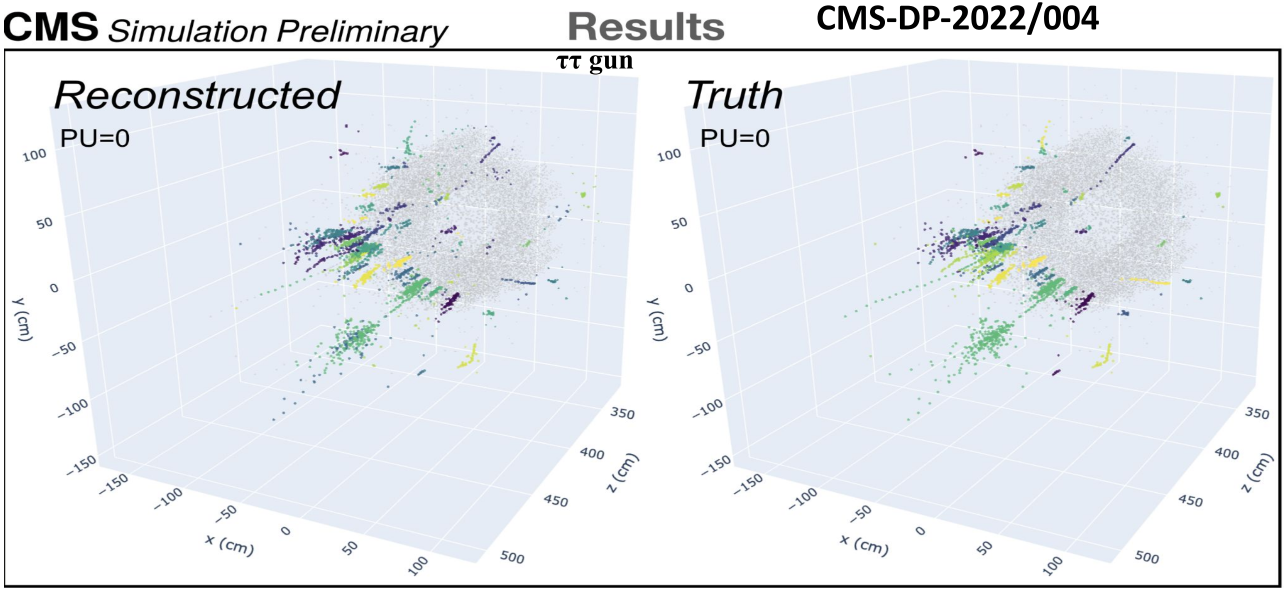


[1]: arXiv:2204.01681
[2]: arXiv:2002.03605

Performance

CMS-DP-2022/004





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

Summary

- The performance of physics analyses in any experiment is intimately linked to the performance of the underlying detectors and the reconstruction of physics objects.
- In order to extract more from a given amount of data it is imperative that we update our detector reconstruction using state-of-the-art tools.
- One of the promising approaches underway in CMS in this regard is to incorporate low level calorimeter hits as input features into GNNs, that are uniquely suited for high energy physics problems.
- Today we have seen a couple of examples amongst the many being used CMS wide.
- These methods being developed are general and applicable to any detector.