

Novel Deep Learning Techniques for Reconstruction of Particles Decaying to Merged Photons and Energy Calibration of Charged Hadrons

A thesis
Submitted towards the partial fulfillment of
BS-MS dual degree programme
by

CHIRAYU GUPTA

Registration Number: 20191060



DATE:

under the guidance of

DR. SEEMA SHARMA

IISER PUNE

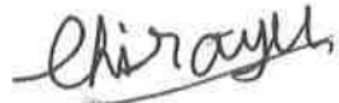
INDIAN INSTITUTE OF SCIENCE EDUCATION AND RESEARCH PUNE

Certificate

This is to certify that this dissertation entitled "Novel Deep Learning Techniques for Reconstruction of Particles Decaying to Merged Photons and Energy Calibration of Hadrons" submitted towards the partial fulfillment of the BS-MS degree at the Indian Institute of Science Education and Research, Pune represents original research carried out by "Chirayu Gupta" at "IISER Pune", under the supervision of "Dr. Seema Sharma" during academic year May 2023 to March 2024.



Supervisor:
DR. SEEMA SHARMA
ASSOCIATE
PROFESSOR
IISER PUNE



CHIRAYU GUPTA
20191060
BS-MS
IISER PUNE

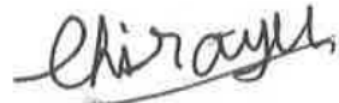
DATE: 26/03/2024

Declaration

I, hereby declare that the matter embodied in the report titled “Novel Deep Learning Techniques for Reconstruction of Particles Decaying to Merged Photons and Energy Calibration of Charged Hadrons” is the results of the investigations carried out by me at the “IISER Pune” from the period 01-05-2023 to 15-03-2024 under the supervision of Dr. Seema Sharma and the same has not been submitted elsewhere for any other degree.



Supervisor:
DR. SEEMA SHARMA
ASSOCIATE
PROFESSOR
IISER PUNE



CHIRAYU GUPTA
20191060
BS-MS
IISER PUNE

DATE: 26/03/2024

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Dr. Seema Sharma, for her invaluable guidance, unwavering support, and encouragement throughout the completion of this thesis. Her expertise, patience, and constructive feedback have been instrumental in shaping this work. All the necessary resources needed for the project were generously provided by my supervisor.

I extend my appreciation to Prof. Rajdeep Mohan Chatterjee who is also a member of my thesis committee for his guidance and continuous support throughout the duration of the project. His mentorship and assistance, including the invitation to the Tata Institute of Fundamental Research Mumbai to discuss my work and providing GPU resources, have been invaluable in gaining exposure and advancing my research. I am also very grateful that I got the opportunity to discuss my work and to learn the details of some of the theoretical motivations of my project from Prof. Monoranjan Guchait. I would also like to thank Anirban who is a PhD student at TIFR, for his assistance in generating data samples and generously allowing the use of his storage area to store my data.

I am thankful to my senior colleague, Bhumika, for helping me get started on the hadron calibration project. She promptly answered to all my queries related to the technical details on the project. Special thanks to Prof. Kenichi Hatakeyama for helping me understand the particle flow elements for the inputs used for hadron calibration.

I am also thankful to my senior colleague, Alpana, for her support in getting started on this project and providing valuable feedback whenever needed. I also thank other lab members like Nitish, Chetan and Tanay for their contributions and discussions during our regular group meetings. I would like to thank the National Supercomputing Mission for providing the necessary computing resources at IISER, and Nisha ma'am for her assistance with any technical issues I faced. I would like to thank the Department of Science and Technology, Government of India for providing me financial assistance through the KVPY fellowship during the course of this project.

Lastly, but certainly not the least, I extend my heartfelt gratitude to my parents, Rajesh Gupta and Neelam Gupta, my brother, Gauransh Gupta, and my close friends at IISER for their unwavering emotional support throughout this journey.

Contents

1	Introduction	2
2	Particle interactions with matter	5
2.1	Energy loss by charged particles	5
2.1.1	Energy loss by ionization process	5
2.1.2	Energy loss by scintillation	6
2.2	Interaction of photons with matter	8
2.3	Energy deposition by high energy particles	8
2.3.1	Electromagnetic showers	9
2.3.2	Hadronic showers	11
2.4	Calorimeters Detectors	13
2.4.1	Energy resolution and response	14
3	CMS detector and the particle flow algorithm	16
3.1	CMS detector	17
3.1.1	Tracker	17
3.1.2	Electromagnetic Calorimeter (ECAL)	18
3.1.3	Hadronic Calorimeter (HCAL)	19
3.1.4	Muon Subsystems	20
3.2	The particle flow algorithm	20
3.2.1	Reconstruction of charged objects	21
3.2.2	Electromagnetic clusters	23
3.2.3	Hadronic Clusters	24
4	Deep learning	26
4.1	Stochastic gradient descent	27
4.2	Graph convolution	28
4.2.1	Graph clustering	30
4.3	Dynamic Reduction Network(DRN)	31
5	Reconstruction of mass of merged photon candidates in the CMS ECAL	33
5.1	Sample production and processing	34
5.1.1	Generation and simulation	36
5.1.2	Digitisation	36
5.1.3	Reconstruction (Reco)	36
5.2	Event selection and model inputs	37
5.2.1	Event selection and data cleaning	37
5.2.2	Inputs to the model	37
5.3	Model training and results	37

5.3.1 Using SupClu rechits	39
5.3.2 Using SupClu+dR rechits	42
5.3.3 Using SupClu+dR rechits using ResNet	44
6 Particle flow hadron calibration	47
6.1 Hadron Calibration	47
6.2 Sample details	50
6.3 DRN model details	53
6.4 Results	54
7 Summary	61
References	61
A Angle between daughter particles in $a \rightarrow \gamma\gamma$ decay process	64
A.1 Notation	64
A.2 Calculation	64
B Comparisons of inputs to χ^2 method and DRN method	66
C Raw energy in calorimeters for different energy bins	70
D Aggregation step in Dynamic Reduction Network	74

Abstract

The Standard Model (SM) of particle physics, continued to be validated by measurements from the CMS and ATLAS experiments at the CERN LHC, is an excellent description of known elementary particles. However, its inadequacy is evident in the divergent values obtained for the quantum corrections to the Higgs boson mass calculated in the SM framework and its failure to explain phenomena such as dark matter and matter-antimatter asymmetry. Null results from searches for physics beyond the standard model (BSM) have prompted investigations into more exotic decay topologies.

This study focuses on measuring the invariant mass of highly boosted low mass particles which decay to two photons and are reconstructed as a single photon object in the CMS detector. Such signatures can be used to identify unconventional Higgs boson decays, specifically $H \rightarrow aa$, where "a" is a hypothetical particle resembling a light axion-like pseudoscalar and it can decay to two merged photons. To measure the invariant mass of these merged $a \rightarrow \gamma\gamma$ candidates, graph neural network (GNNs) methods are explored, leveraging their capability to handle irregular geometries of electromagnetic showers produced in the electromagnetic calorimeter of the CMS detector. We use low level input features to train the machine learning model, necessitating an understanding of CMS event reconstruction machinery and access to relevant underlying information. We also studied the calibration of hadron showers using the GNNs and demonstrate that resolution could be significantly improved as compared to the conventional methods currently used by the CMS collaboration.

Chapter 1

Introduction

Particle physics deals with understanding the fundamental laws that govern fundamental particles and forces among them, which ultimately lead to the universe we see. Decades of relentless pursuit have led to the formulation of the standard model (SM) of physics of elementary particles, a collaborative endeavor uniting minds from diverse theoretical and experimental backgrounds. The SM is a theoretical framework which describes the interactions between various fundamental particles like quarks, leptons, gauge bosons, and lastly the Higgs boson which was experimentally discovered in 2012. It can be stated with the help of quantum field theory using a lagrangian having a group symmetry of $SU(3) \times SU(2) \times U(1)$. The elementary particles in the SM model are illustrated in figure 1.1. There are 6 quarks in the SM namely the up, down, charm, strange, top and bottom quarks. There are also 6 leptons in the SM which are the electron, muon and tau and their neutral partners, the electron neutrino, muon neutrino and tau neutrino, respectively.

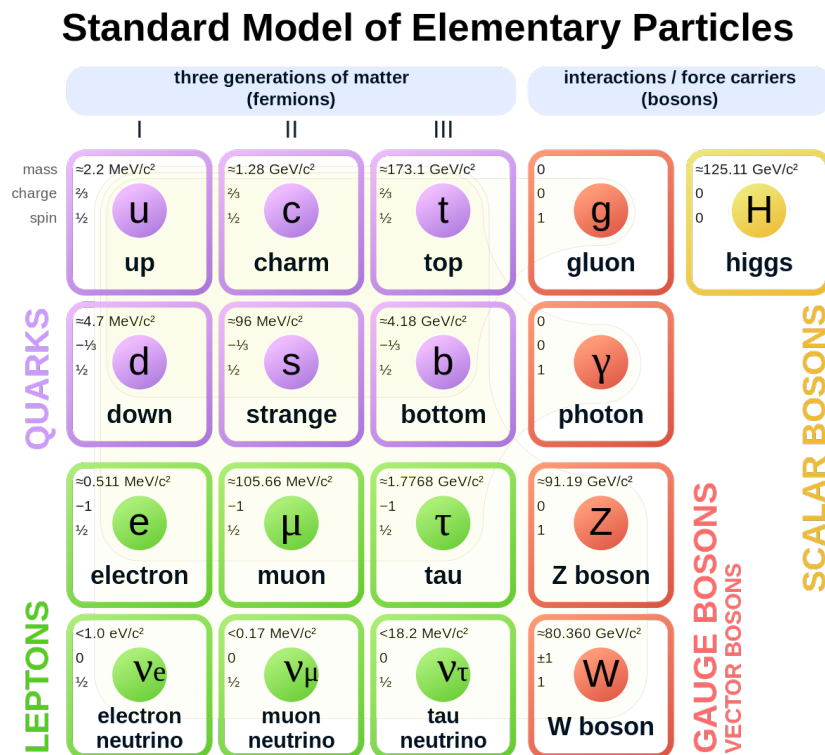


Figure 1.1: The SM of particle physics [17]

Gauge bosons are responsible for mediating the fundamental interactions like photons for electromagnetism, gluons for strong interactions and W^+ , W^- and Z bosons for weak interactions. Gravity is not included in the SM framework.

Quarks and leptons are fermions with a spin of $1/2$. Both quarks and leptons can interact electromagnetically. Quarks also possess a color charge which makes them susceptible to interactions with gluons and subsequently the formation of hadrons like protons, neutrons, pions, kaons, etc. Quarks and gluons are detected as jets in collider based experiments. In fact, it is these complex Quantum Chromodynamics(QCD) interactions that are responsible for more than 90 percent of the rest mass of a proton as well as a neutron(which are hadrons) and consequently the mass of all matter around us.

However, the fundamental particles described by the SM do not get their mass using this mechanism. Their mass is governed by their interactions with the Higgs boson studying the properties of which is currently the central focus of the Compact Muon Solenoid (CMS) experiment at the European Organization for Nuclear Research(CERN).

The Higgs field is a doublet with 2 real and 2 imaginary components having four degrees of freedom in total. This field has a non zero vacuum expectation value of around 246 GeV. This causes the $SU(2) \times SU(1)$ symmetry of the SM lagrangian to be spontaneously broken to $U(1)$ giving rise to 3 goldstone bosons and one massive boson which is the Higgs boson. The 3 goldstone bosons are observed as components of the W and Z bosons that are made massive by their inclusion. This mechanism is known as electroweak symmetry breaking. In SM, the Higgs boson is a massive scalar particle with no electric or color charge and having a lifetime of the order of 10^{-22} seconds.

In the current state of experimental particle physics, the SM has been very successful in explaining physics observed by colliders [14]. However it has its drawbacks like its inability to explain phenomenon such as matter-antimatter symmetry, gravity and the finite masses of the neutrinos. Since the quantum corrections to the Higgs mass are large, the Higgs mass is expected to approach the plank mass of the order of $10^{19} GeV/c^2$, however the Higgs mass is observed to be around $125 GeV/c^2$. New models with an expanded Higgs sector were proposed particularly to keep the corrections to the Higgs mass finite. A striking feature of these models like supersymmetry(SUSY), is the existence of 2 Higgs doublets instead of a single doublet. In such a model, there are 8 degrees of freedom within the Higgs sector, 3 of which are eaten up by the W and Z bosons. The remaining 5 degrees of freedom appear as 5 physical Higgs bosons: a charged pair (H^+ , H^-), two neutral CP even scalars (H , h), and a neutral CP-odd scalar (a), also known as a pseudoscalar [19].

Motivated by the search of these exotic Higgs bosons, particularly the axion like pseudoscalar (a), which is theoretically constrained to have a mass lower than a few GeV/c^2 . I take on the experimental challenge of reconstructing the mass of a which decays to a pair of merged photons. Furthermore, in the production of the Higgs boson like the Vector Boson Fusion (VBF), the Higgs is accompanied by 2 jets, which if reconstructed correctly can help us in isolating events where we can further investigate coupling of the Higgs boson to W and Z gauge bosons. Charged hadrons and neutral hadrons(π_0), being one of the main constituents of the jets, are essential to be reconstructed with maximum precision but the nature of the hadronic showers makes it difficult. Motivated by this, I also take on the challenge of the energy calibration of hadrons specifically charged pions.

This thesis is divided into 6 chapters, the second chapter goes into details of how particles deposit their energy in matter. Third chapter covers the details of the CMS detector, its layout and the particle flow algorithm used for the reconstruction of particles produced in pp collisions. The fourth chapter describes the deep learning methods used in these analyses. The fifth chapter

goes into details about the reconstruction of this pseudoscalar, a and the sixth chapter is about improving the energy calibration and resolution of charged hadrons in the CMS calorimeters.

Chapter 2

Particle interactions with matter

To unravel the secrets of nature at subatomic and subnuclear levels, we rely on instruments called particle detectors which are used to measure the four-momenta and positions of these particles. Most of the SM particles are short-lived and decay into lighter particles, e.g. a Z boson decays primarily into a quark antiquark pair (with probability of about 0.7), but it may also decay to a neutrino-antineutrino pair (with probability of about 0.2) and rarely (with probability less than 0.1) to charged lepton-antileptons. The quarks and gluons hadronize into hadrons like π^\pm , π^0 , p , n , etc. So, most of the time, we measure e^\pm , μ^\pm , γ , charged and neutral hadrons in our detectors and infer kinematic properties of all other particles using them.

As a particle passes through a material, it can interact with its electrons and nuclei and lose energy in the process. To better grasp and enhance detector design, it is crucial to explore how different particles interact with various materials that are used in making detection systems. This understanding is crucial for improving detector performance and extracting meaningful physics information from the data collected by collider experiments. This section briefly summarizes energy losses via ionization and scintillation processes followed by electromagnetic and hadron showers generated by high energy e^\pm/γ and hadrons, which are the main focus of the studies presented in this thesis. This is followed by a discussion on calorimeter detectors and sources of uncertainties involved in calorimeter measurements. For a detailed description of particle interactions with matter in a more general context of high energy physics experiments, please refer to [21].

2.1 Energy loss by charged particles

As the charged particles traverse the material, they can lose energy via ionizing or exciting the atoms. Key features of the two processes are summarized in this section.

2.1.1 Energy loss by ionization process

Charged particles undergo inelastic collisions with the atoms and lose energy along the path they traverse in a material. These inelastic collisions lead to the ionization of atoms i.e. releasing loosely bound electrons from the outer shells. This mechanism is referred to as energy loss by ionization or simply ionization losses. The average energy loss per unit path length by charged particles heavier than electrons via ionization process is given by the *Bethe-Bloch*

formula (equation 2.1 [24]).

$$-\frac{dE}{dx} = 2\pi N_A r_e^2 m_e c^2 \frac{Z z^2}{A \beta^2} \left[\ln \frac{2m_e c^2 \beta^2 \gamma^2}{I} - 2\beta^2 \right] \quad (2.1)$$

where,

- N_A is Avogadro's number,
- r_e is the classical electron radius,
- m_e is the electron rest mass,
- c is the speed of light,
- Z is the atomic number of the material,
- A is the atomic mass of the material,
- z is the charge of the incident particle,
- β is the velocity of the incident particle relative to the speed of light,
- γ is the Lorentz factor, and
- I is the mean excitation potential.

The equation 2.1 needs to include a few modifications to describe the ionization energy loss for electrons [21].

Ionization energy loss per unit pathlength as a function of incident particle energy is shown in Fig 2.1 for a few particles. As also indicated by equation 2.1, the ionization energy loss per unit pathlength (dE/dx) is larger for low energies of incident particles (small β).

As the incident energy of particles increase, the dE/dx reaches to its minimum value and rises only very slowly at higher energies. Particles with energy range in which the dE/dx is minimum are called minimum ionizing particles (MIPs). In this energy range, the Bethe Bloch formula predicts a nearly linear relationship between the energy loss and the path length of the particle, resulting in a characteristic plateau in the dE/dx versus energy curve. This behaviour is particularly relevant for particles such as muons, for which electromagnetic showers are highly suppressed given its large mass ($m_\mu = 200m_e$), resulting in their highly penetrating nature.

2.1.2 Energy loss by scintillation

When a charged particle traverses a material, it may also cause the atoms or molecules to go into an excited state. Excited states are unstable and the atom or molecule quickly returns to its ground state releasing one or more photons in the process. When the frequency of these photons lie in the range that can be measured easily by photodetectors, this phenomenon is called fluorescence or *scintillation*.

Using the process of scintillation, we can detect energy deposited by charged particles in a scintillator detector. Neutral particles can also be detected when they produce secondary charged particles, e.g. following a nuclear interaction and photons which can produce a electron using processes discussed below. A scintillating material converts the energy of one photon or

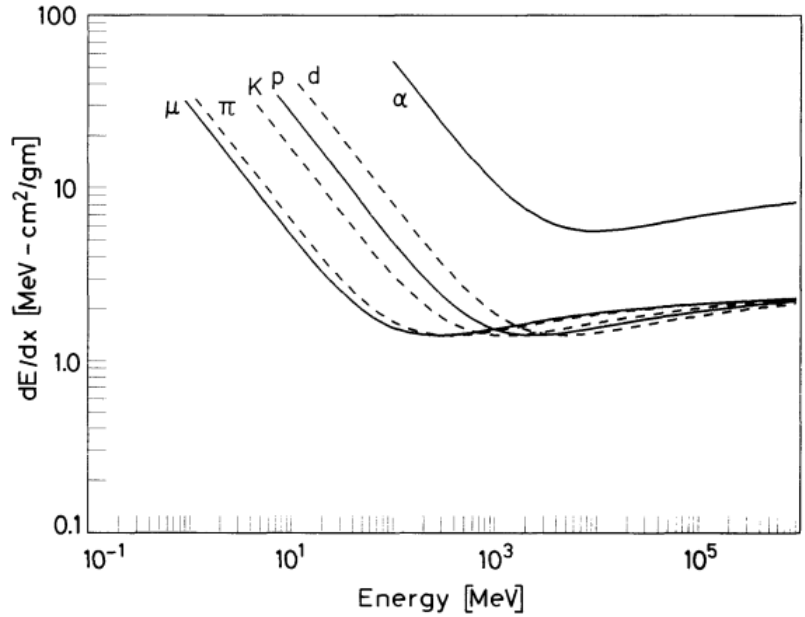


Figure 2.1: The average energy loss dE/dx as function of energy for different particles. Solid lines are loss predicted by eq 2.1. Dashed lines represent corrections to this formula 2.1

particle into a number of photons of different wavelength. Scintillation detectors can be made of organic (e.g. plastic scintillators used in the CMS hadron calorimeter) or inorganic material (e.g. $PbWO_4$ used in the CMS electromagnetic calorimeter (ECAL)).

The scintillation photons produced in the detector have to be collected and converted to an electronic signal. Photomultiplier tubes (PMTs) are widely used for this purpose. Scintillation photons interact with a photocathode at the face of a PMT and liberate electrons through the photoelectric effect. These ejected electrons are then accelerated and focused by an electric field within the detector, resulting in a cascade of secondary electrons from a series of anodes. The cascade amplifies the initial signal, producing a measurable electrical current that corresponds to the intensity of the incident photons. The process of converting scintillation photons to an electric signal is illustrated in Figure 2.2, where the sequential steps of photon interaction, electron ejection, and signal amplification are visually represented.

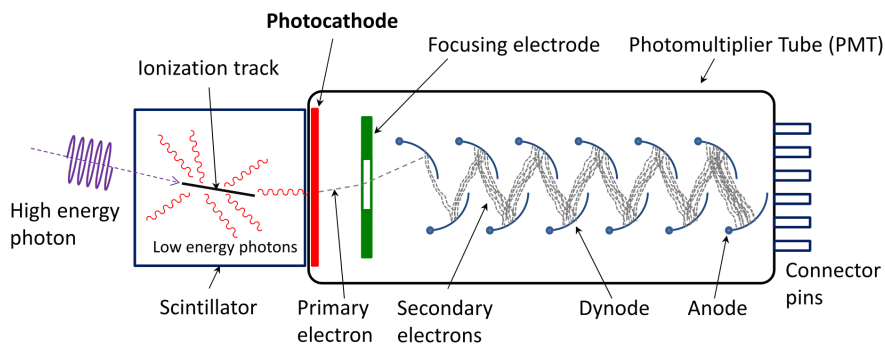


Figure 2.2: Scintillator detector coupled to a photo multiplier tube (PMT) 4

2.2 Interaction of photons with matter

Photons traversing a material can lose its energy mainly via pair production, Compton effect, and photoelectric effect. A high energy photon, under the influence of an external electromagnetic field, convert to an electron-positron pair. For pair production to occur, the incident photon must possess a minimum energy of 1.02 MeV, corresponding to the rest mass of the e^+e^- pair.

When the energy of the photon falls below the pair production threshold, it may lose their energy by scattering off bound electrons from the atoms, thereby changing its direction. This phenomenon is called *Compton scattering*. Finally, at further lower energies, the photon is simply absorbed by the atoms leading to photo-emission of the electrons, called photoelectric effect. The dominant energy loss mechanisms of photons at various energies is illustrated in Figure 2.3.

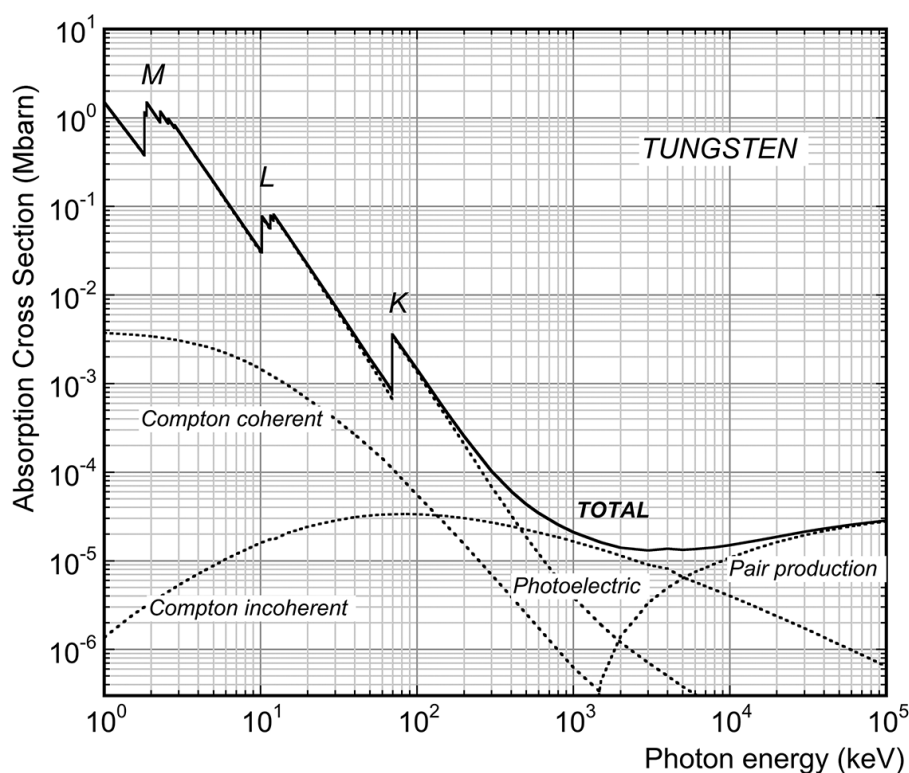


Figure 2.3: Cross section (quantity proportional to the probability of interaction) of various energy loss processes by photons in tungsten as a function of photon energy [22]

2.3 Energy deposition by high energy particles

The energy deposition of high-energy particles, such as e^\pm , γ , charged and neutral hadrons, often occurs via the formation of particle showers in the detector. In the process of shower formation, the energy of the incident particle is successively distributed among various secondary particles at each branching step. Such particle showers can be broadly classified into two types depending on the nature of interactions involved: 1) electromagnetic showers, produced by e^\pm and γ , and 2) hadron showers, produced by charged or neutral hadrons such as π^\pm , K , p , n , etc.

It is noteworthy to mention that neutral π^0 is a hadron but its experimental signature is an electromagnetic shower. It decays into two photons with a characteristic lifetime of 10^{-16} seconds [5]. The resulting photons from π^0 decays initiate electromagnetic showers.

2.3.1 Electromagnetic showers

When a high-energy photon traverses an absorber material, such as the scintillation crystals used in the ECAL, it undergoes an e^+e^- pair-production under the influence of the coulomb field of the nuclei of the absorber atoms, hence initiating an electromagnetic shower. Following the pair-production, each electron or positron may engage in further electromagnetic interactions with atomic nuclei, leading to the emission of *bremstrahlung* photons. These emitted photons may, in turn, undergo pair-production, perpetuating the cycle and forming an electromagnetic shower containing a large number of e^+ , e^- and γ . This cascading phenomenon is illustrated in Figure 2.4

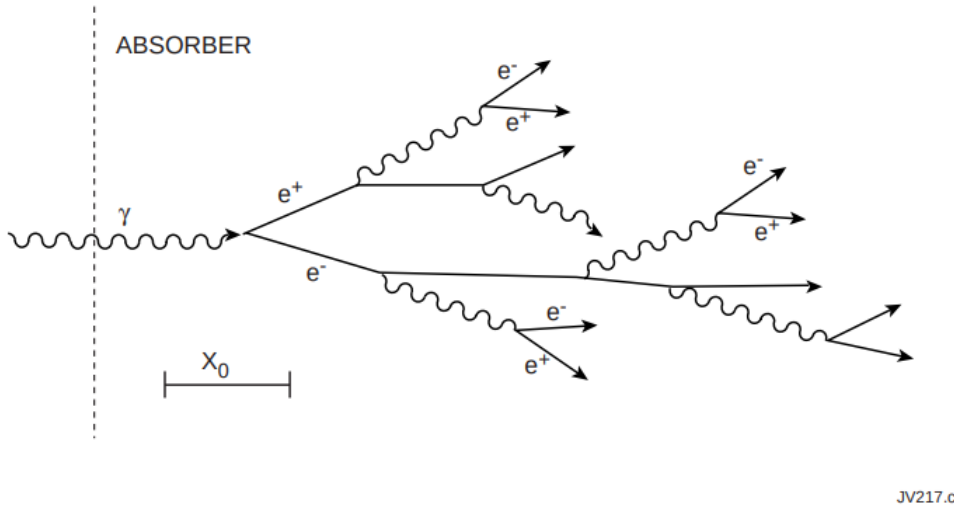


Figure 2.4: Illustration of an electromagnetic shower produced by a high energy photon incident on an absorber material. [24]

The initiation of the electromagnetic shower can also occur through light charged particles moving at relativistic speeds, such as electrons or positrons, which emit photons via the bremsstrahlung process under the influence of the coulomb field of the nuclei. Notably, the probability of a charged particle undergoing bremsstrahlung is inversely proportional to the square of its mass. Consequently, heavier charged particles, like charged hadrons or muons, typically do not initiate an electromagnetic shower.

In contrast, electrons can lose energy through both ionization and radiation processes. The total energy loss by electrons is given by:

$$\left(\frac{dE}{dx}\right)_{total} = \left(\frac{dE}{dx}\right)_{radiation} + \left(\frac{dE}{dx}\right)_{ionisation} \quad (2.2)$$

Radiative losses dominate at higher energies. The energy at which radiative losses become equal to ionisation losses is known as **critical energy** (See Figure 2.5).

To study the lateral and longitudinal development of electromagnetic showers initiated by particles of different energies in various absorber materials, we define a few quantities:

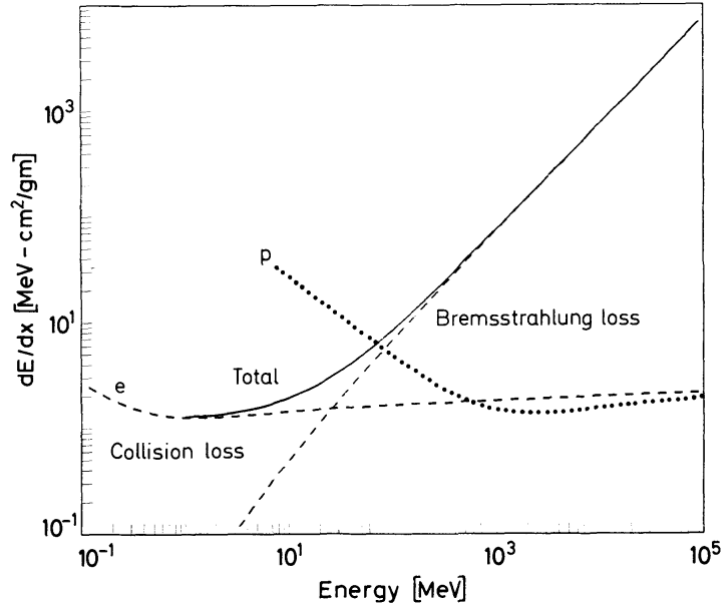


Figure 2.5: Radiation loss vs ionisation/collision loss for electrons in copper. For comparison proton loss is also shown [21]

- **Radiation length(X_0):** It is defined as the mean length in which the energy of the electron is reduced by a factor of $1/e$ through radiative losses.
- **Moliere radius(R_m):** It is defined as the radius of an infinite cylinder which contains 90% of the shower energy

Longitudinal energy deposition by electromagnetic showers first increases, reaches a maximum (called *shower maxima*), and then decreases exponentially as illustrated in Figure 2.6. It can be seen that higher Z material have shower maxima at a larger depth defined in units of X_0 . This can be attributed to the fact that higher Z material have lower critical energies allow the shower multiplicity to grow at relatively lower energies.

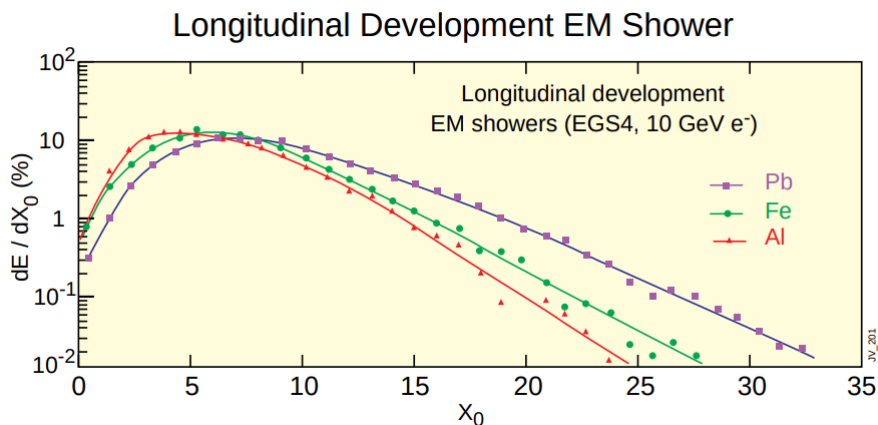


Figure 2.6: Longitudinal development of 10 GeV electron showers in Al ($Z=13$), Fe ($Z=26$) and Pb ($Z=82$) [24]

Electromagnetic showers are narrow in the beginning which makes them distinguishable from pions decaying to two photons. They continue to grow laterally as the shower develops,

with high energy electromagnetic cascade at its core, surrounded by a halo of soft photons and electrons. This halo may be associated to the multiple scattering of electrons away from the shower axis.

2.3.2 Hadronic showers

While traversing the detector, high energy hadrons interact strongly with the nuclei of the detector medium and causes multiparticle production. These interactions can lead to the breakup of nuclei, generating spallation products. Spallation is a two stage process: a fast intranuclear cascade followed by a slow evaporation of the target nucleus (see Fig 2.7).

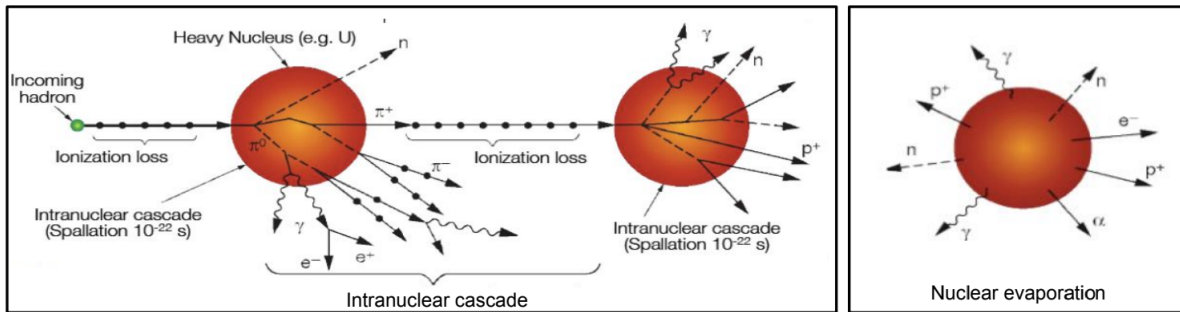


Figure 2.7: Pictorial representation of intranuclear cascade(left) and slow evaporation(right). Courtesy of H. C. Schoultz Coulon

The incoming hadron undergoes quasi-free collisions with nucleons within the struck nucleus, initiating a cascade of fast-moving nucleons as the affected nucleons travel and collide with others. This process may lead to the creation of pions and other unstable hadrons if the transferred energy is significant, and some of the cascade particles escape the nucleus, while others redistribute their kinetic energy among the remaining nucleons. The particles that leave the nucleus have to overcome the nuclear binding energy. This energy is lost and is not measured by the detectors.

During the evaporation stage, the nucleons within the target nucleus dissipate absorbed kinetic energy through the evaporation of free nucleons, continuing until the excitation energy drops below the binding energy of an individual nucleon.

The secondary particles that leave the nucleus, in turn, may engage with other nuclei in the detector material, setting off a cascading effect known as a hadronic shower. The cascade contains two distinct components namely the electromagnetic one (comprising primarily of neutral pions which almost instantaneously decay to two photons) and a hadronic components (illustrated in Figure 2.8).

Owing to the intricate dynamics of hadronic showers in comparison to electromagnetic showers, coupled with the substantial uncertainties inherent in quantifying the fraction of energy lost as binding energy during hadron interactions, the energy measured for an e^\pm or a γ and a π^\pm or p/n of the same incident energy is generally very different in calorimeters (also known as non-compensating calorimeters). The complexity introduced by these processes contributes to significant fluctuations in the energy distribution between electromagnetic and hadronic components of the shower. Detectors respond differently to these shower components, presenting challenges in accurately predicting or measuring the energy of incident hadrons. Figure 2.9 shows the shower-to-shower fluctuations in energy deposited along the hadron showers at different depths of the detector. The peaks in the figure correspond to local deposits of energy by

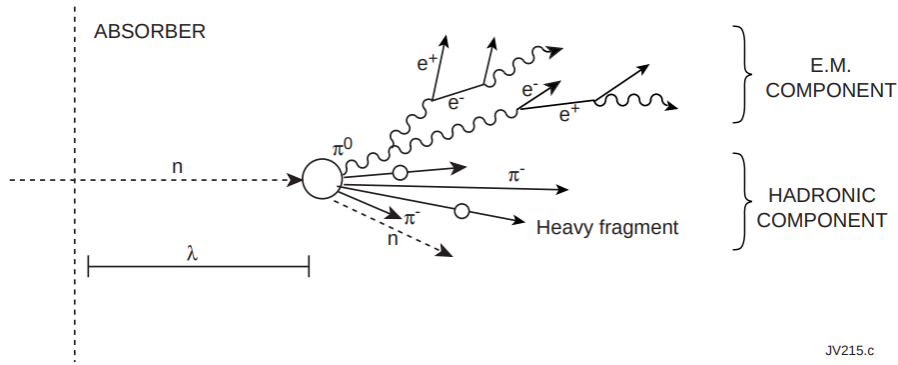


Figure 2.8: Pictorial representation of components of a hadronic shower [24]

neutral pions (electromagnetic component). The depth at which this interaction occurs and the energy carried by neutral pions varies largely from one shower to the other for the hadrons of given type and energy.

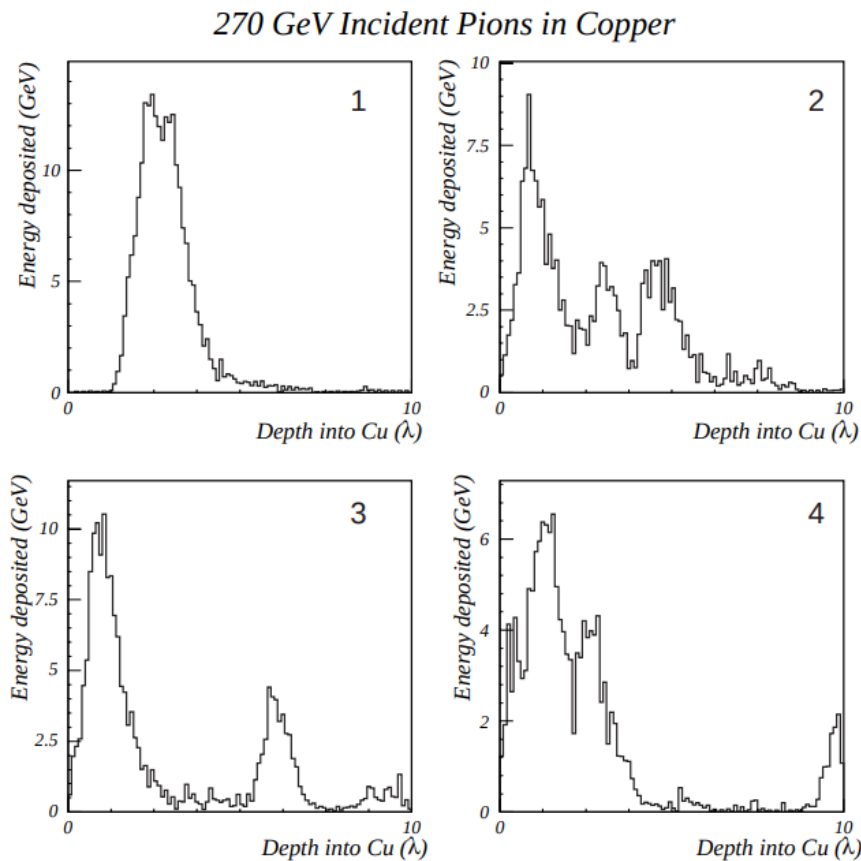


Figure 2.9: A simulation of the development of four representative pion showers in a block of copper [24] illustrating the shower to shower fluctuations in hadronic showers

To describe the longitudinal and lateral development of hadron showers, we define nuclear interaction length (λ_{Int}) as the mean distance travelled by a hadron before undergoing a nuclear interaction. Similar to the electromagnetic showers, the average longitudinal energy deposited by hadrons increase as the number of shower particles increase resulting in a broad peak followed by a slow exponential drop-off with the depth. Longitudinal energy profile of charged

pions of different energies as a function of depth in units of λ_{INT} is illustrated in figure 2.10.

Similar to the electromagnetic showers, the average longitudinal energy deposited by hadrons experience a peak at the first interaction point (from neutral pions). We define λ_{INT} , the nuclear interaction length as the mean distance travelled by a hadron before undergoing a nuclear interaction.

Another key difference between electromagnetic and hadronic showers is the number of particles present at shower maxima. The average number of particles present at shower maxima can be approximated by

$$n_{max} = E/E_{th} \quad (2.3)$$

where E is the energy of the incident particle and E_{th} is the threshold energy required for initiating a shower. For electromagnetic showers this corresponds to the critical energy whereas for hadronic showers, $E_{th} = 2m_{\pi} = 280 \text{ MeV}$.

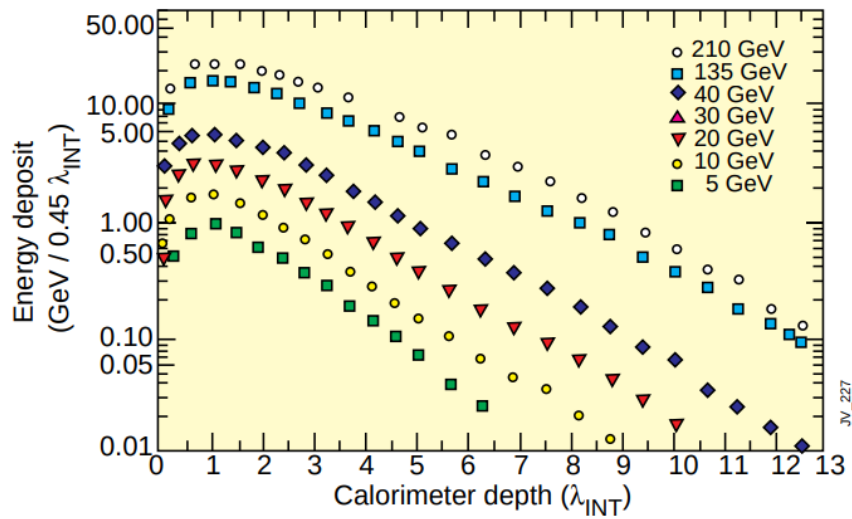


Figure 2.10: Longitudinal energy desposition profile of pions at different energies [24]

Since, the number of independent particles in hadronic showers is smaller than electromagnetic showers by a factor of E_{th}/ϵ , where ϵ is the critical energy. This causes the intrinsic energy resolution of hadronic showers to be worse than that of electromagnetic showers by a factor of $\sqrt{E_{th}/\epsilon} = 6$. The realistically achievable resolution is further worsened by the shower-to-shower fluctuations explained earlier.

2.4 Calorimeters Detectors

Calorimeter detectors or calorimeters are the essential instruments utilized in particle physics experiments to measure the energy and position of particles by fully absorbing them. Often covering almost the full solid angle of 4π , these detectors are optimized to have high energy particles induce showers of secondary particles whose energies can be precisely measured. Calorimeters can be classified as sampling or homogeneous, depending on whether they consist of alternating layers of active and passive materials or a continuous medium throughout, respectively. The choice of materials, segmentation, and calibration methods are crucial factors influencing the performance of calorimeters and their resolution in determining particle energies over a wide range.

Homogeneous Calorimeter A homogeneous calorimeter is characterized by a active medium throughout its entire volume. In this design, the absorbing material is responsible for initiating particle showers as well as for converting the energy deposited in a measurable signal. The uniformity of the material allows for a predictable response to incoming particles. The ECAL of the CMS experiment uses scintillating crystals (lead-tungstate or $PbWO_4$) as absorbers. Electromagnetic showers are relatively compact and 90 percent of its energy can be mostly contained in a 3x3 array of crystals in the CMS ECAL. A generic homogeneous calorimeter is illustrated in figure 2.11.



Figure 2.11: Design principle of a homogeneous calorimeter [10]

Sampling Calorimeter In contrast, a sampling calorimeter is characterized by alternating layers of absorbing or passive and active materials, as shown in figure 2.12. The absorbing layers are responsible for initiating and developing particle showers while the active layers serve as measurement regions in which energy deposited is converted to a measurable signal. This design allows for a cost-effective approach, as the active layers can be less expensive material designed primarily for detecting the particle's energy. Sampling calorimeters are widely used to make hadron calorimeters. For instance, hadron calorimeter (HCAL) of the CMS experiment is a sampling calorimeter with brass absorbers and scintillator tiles as active material. The advantage of sampling calorimeters lies in their flexibility to tailor the balance between accuracy and cost-effectiveness by adjusting the thickness and composition of the sampling layers. However, they require sophisticated calibration techniques to translate the observed signals into accurate energy measurements.

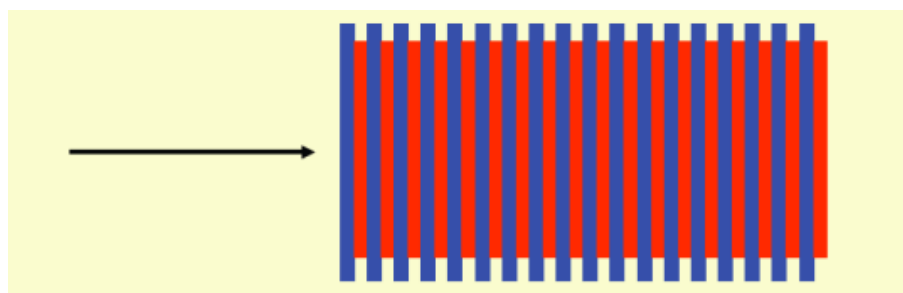


Figure 2.12: Sampling calorimeter [10]

2.4.1 Energy resolution and response

The energy resolution of calorimeters can be parameterised as

$$\frac{\sigma}{E} = \frac{a}{\sqrt{E}} \otimes \frac{b}{E} \otimes c \quad (2.4)$$

The stochastic or sampling term (a) accounts for statistical fluctuations in the number of primary and independent signal generating processes. It includes effects such as statistical fluctuations in particle showers and energy conversion processes.

The noise term (b) incorporates contributions from electronics noise and fluctuations in energy carried by particles other than those of interest that enter the measurement area (known as pileup).

The constant term (c) takes into account imperfections in construction, non-uniformity of signal generation and collection, inter-calibration errors between cells, fluctuation in energy leakage, and fluctuations in contributions from electromagnetic components in hadronic showers

The tolerable size of these terms depends on the energy range involved in an experiment and the physics being targeted. For optimal performance, it is desirable to minimize these terms as much as possible. However, achieving a good energy resolution is often challenging due to various factors such as detector limitations and the nature of particle interactions.

In the context of CMS, different types of calorimeters are used to measure energy deposition. Intrinsic electromagnetic energy resolution is studied in homogeneous calorimeters where all the energy is deposited in the active medium. The improvement in resolution is characterized by the Fano factor, which quantifies the fluctuation in producing electron-ion pairs or photons

Sampling electromagnetic calorimeters, on the other hand, use passive absorber layers along with active layers to measure only a fraction of shower energy. The energy resolution in these calorimeters depends on the fluctuation in this fraction and can be improved by optimizing the thickness of absorber layers.

Hadronic calorimeters, necessary for measuring hadron energies, are typically sampling calorimeters with large depth requirements. The response of these calorimeters depends on various components such as electromagnetic and charge hadrons, low-energy neutrons, and nuclear breakup energies. Achieving compensation between different components involves carefully choosing absorber materials and optimizing their thickness ratios.

The energy resolution plays a critical role in determining jet energies in high-energy physics experiments. In CMS, jets are often reconstructed by clustering the energy contained around a seed cell using dedicated algorithms. The jet energy resolution is affected by algorithms used for jet definition, fluctuations in particle content within jets, underlying event contributions, pileup effects, and magnetic field influences.

Understanding and improving the energy resolution of calorimeters is essential for accurate measurements in high-energy physics experiments. Efforts are continuously made to optimize detector designs and calibration techniques to minimize these resolution limitations.

Chapter 3

CMS detector and the particle flow algorithm

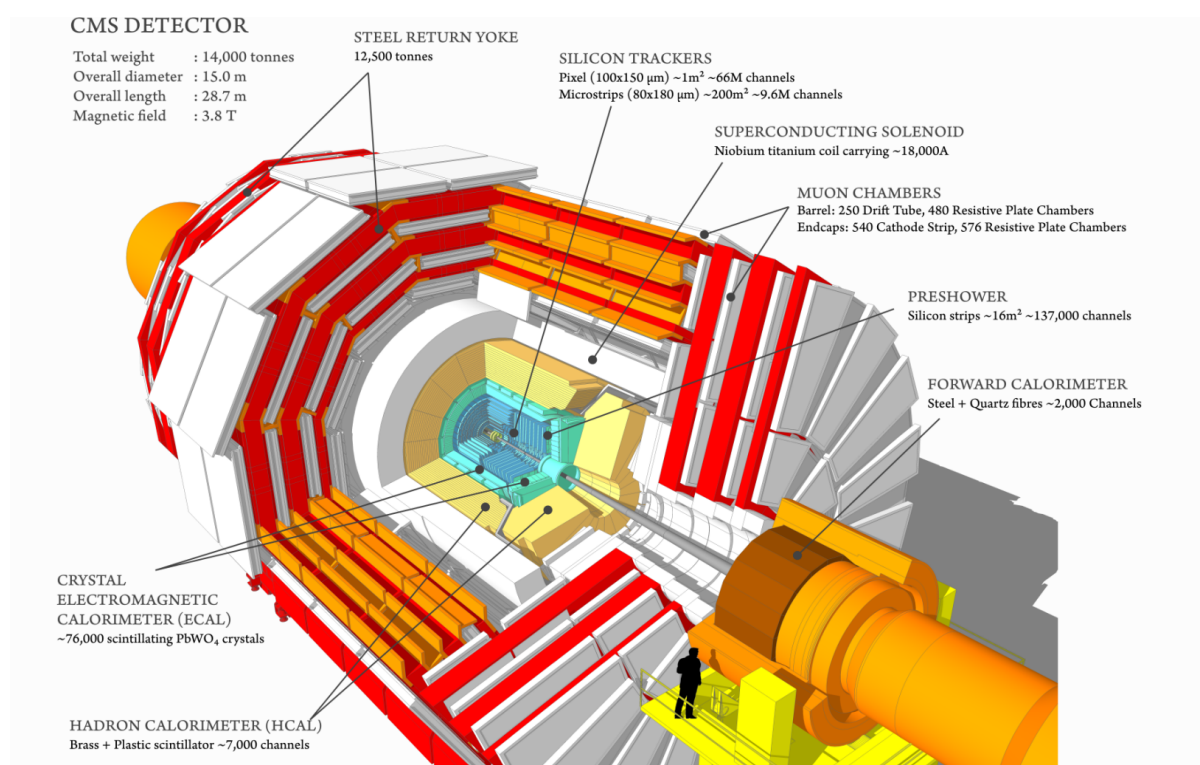


Figure 3.1: The CMS detector [13]

The Large Hadron Collider (LHC) is the world's largest and most powerful particle accelerator, located at the European Organization for Nuclear Research (CERN) at the borders of Switzerland France. The LHC consists of superconducting magnets and accelerating structures installed along a 27-kilometer ring to guide bunches of protons around the accelerator. It also accelerates and collides heavy ions (lead and gold) but these are not discussed in this thesis. The LHC is designed to collide protons at a centre-of-mass energy of 14 TeV every 25 ns and to operate at a design luminosity of $10^{34}\text{cm}^{-2}\text{s}^{-1}$. There are 10^{11} protons in each bunch. The beams are made to cross at four intersection points. The CMS detector is situated at one of the collision points of the LHC and it is designed to study outcome of both proton-proton (pp) and heavy ion collisions.

3.1 CMS detector

The CMS detector is installed around the pp collision point covering a substantial portion of the 4π solid angle. It is of cylindrical geometry with a length of 21.6 m and a diameter of 14.6 m. It comprises several essential components, such as advanced tracking systems, electromagnetic and hadron calorimeters, and muon detectors along with a large-bore superconducting solenoid to provide 4 Tesla magnetic field.

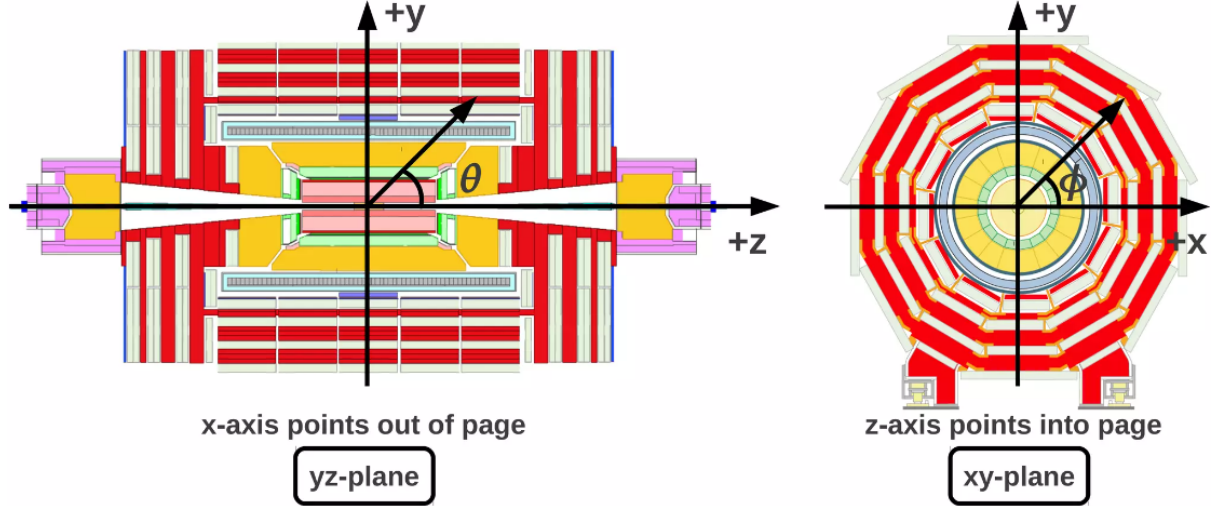


Figure 3.2: θ and ϕ in the CMS coordinate system [13]

A right handed coordinate system is used for the CMS detector, where the z direction is the beam direction and x and y directions lie in the plane perpendicular to the beam axis as illustrated in figure 3.2. At hadron colliders, particle density is higher in forward direction and not uniform in equally spaced θ ranges. Hence instead of using the polar coordinate θ , we define another quantity, pseudorapidity(η).

$$\eta = \frac{1}{2} \ln \left(\frac{p + p_z}{p - p_z} \right) \quad (3.1)$$

In terms of angle from the beam axis(z axis) pseudorapidity becomes:

$$\eta = -\ln \tan \frac{\theta}{2} \quad (3.2)$$

The other variable used to cover the coordinate space is ϕ which is the azimuthal angle as shown in figure 3.2. It ranges from $-\pi$ to π . The pseudorapidity coverage provided by the CMS detector is from -5 to 5 which corresponds to -179.2 degrees to 179.2 degrees. Unlike polar coordinate θ , differences in pseudorapidity ($\Delta\eta$) are preserved under Lorentz transformations along the beam axis for highly relativistic particles $E \approx p$.

The CMS detector is like peeling an onion, made up of different layers, each serving a specific purpose. These layers or subdetectors, work together to capture and understand the interactions of particles in a precise and detailed manner.

3.1.1 Tracker

As we go away from the p-p collision point, we first encounter the **tracker** at a radius of 4.4 cm. The tracker is composed of 2 subsystems, namely the silicon pixel detectors which extend

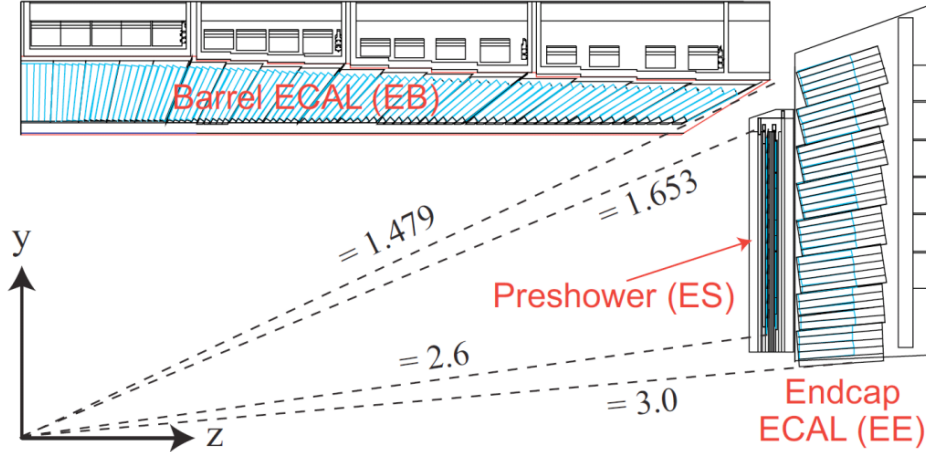


Figure 3.3: Schematic view of one quarter of ECAL [7]

upto a radius of 10.2 cm and the silicon strip detectors which extend outwards upto a radius of 110 cm. The tracker is supplemented by endcaps which extend the pseudorapidity coverage upto $|\eta| < 2.5$. Being so close to the collision point, the tracker is subjected to high levels of radiation and the detector material has been carefully chosen to make this system compact and radiation hard. This tracker system is constructed to provide a precise measurement of the trajectories of charged particles and, unlike the calorimeters, it does so in a non destructive way. Ideally, only charged particles, register hits in the tracker and neutral particles pass through unaffected.

3.1.2 Electromagnetic Calorimeter (ECAL)

Moving further away from the collision point, we encounter the ECAL at a radius of about 129 cm which is responsible for precise energy measurements of electrons and photons. Its layout comprises a barrel and two endcaps, each equipped with lead tungstate ($PbWO_4$) scintillating crystals. These crystals have a high density (8.28 g/cm^3), a short radiation length (0.89 cm) and a small moliere radius (2.2 cm). This results in compact detector with fine granularity. These crystals are radiation tolerant and have a fast decay time (about 80% of the scintillation light decays in 25 ns) to account for the high frequency of p-p collisions.

The barrel section consists of 61,200 crystals arranged in a fine-grained matrix, covering an η range of approximately $|\eta| < 1.48$. Crystals are projective and positioned slightly off-pointing (≈ 3 degrees) relative to the interaction point to avoid cracks aligned with particle trajectories. Each crystal has a lateral cross section of $22 \times 22 \text{ mm}^2$ on the front face and $26 \times 26 \text{ mm}^2$ at the rear face (0.0174×0.0174 in the $\eta - \phi$ plane) with the length of the crystal being 230 mm or about $25.8 X_0$. The longitudinal dimensions of the crystals ensure that approximately 95 percent of the electromagnetic showers generated by particles of up to a TeV energy range are contained within a single crystal

The endcaps, extending the coverage to $|\eta| \approx 3$, employ 7,324 crystals in each disk. Each endcap crystal has a lateral cross section of $28.62 \times 28.62 \text{ mm}^2$ on the front face and $30 \times 30 \text{ mm}^2$ at the rear face with the length of the crystal being 220 mm or about $24.7 X_0$.

The photodetector system employed in the barrel are avalanche photodiode (APD) which can be regarded as the functional analogue of photomultiplier tubes which were discussed earlier. For the endcaps, where high radiation tolerant photodetectors are required, we use vacuum

phototriodes (VPT) .

3.1.3 Hadronic Calorimeter (HCAL)

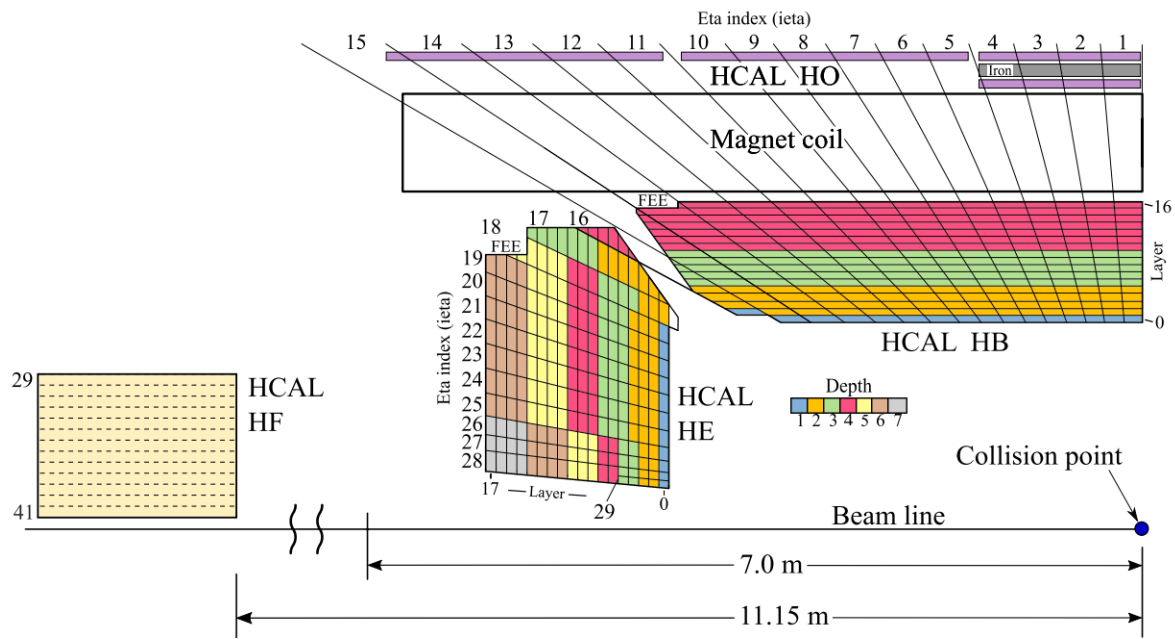


Figure 3.4: Schematic view of one quarter of HCAL [12] showing HE, HB, HO and HF. Each color in the barrel denotes a new layer used in Run 3 data taking (ongoing LHC run). The signals from these layers of a given tower are optically summed and read out by a single photodetector.

Going further away from the collision point, we encounter the HCAL at a radius of about 1.7 meters. The CMS Hadron Calorimeter (HCAL) is a crucial component in the CMS detector, designed to measure the energy and position of hadrons. The HCAL is composed of four major subdetectors: the Hadron Barrel (HB), Hadron Endcap (HE), Hadron Forward (HF), and Hadron Outer (HO) calorimeters, each serving a specific purpose in covering different pseudorapidity regions.

The Hadron Barrel (HB) covers the region with a pseudorapidity $|\eta| < 1.392$, while the Hadron Endcap (HE) extends the coverage to $1.305 < |\eta| < 3.0$. The Hadron Outer (HO) provides measurements of shower tails in the region $|\eta| < 1.26$, and the Hadron Forward (HF) covers the forward region with $3.0 < |\eta| < 5.2$. This comprehensive coverage allows the HCAL to capture and measure the energy deposition of hadrons across a wide pseudorapidity range.

The HB and HE primarily utilize brass as the absorber material, with exceptions for the inner and outer layers of the HB, constructed from steel. The recordable signals are generated in plastic scintillating tiles, where the resulting blue light is shifted to green through embedded wavelength-shifting fibers. Each tower in HB (HE) comprises up to 17 (18) scintillator layers, grouped into "depth" segments. The optical summing of signals within a depth segment is facilitated by clear plastic fibers, transmitting the signal to silicon photomultipliers (SiPMs).

3.1.4 Muon Subsystems

Going further away from the collision point, we encounter the Muon System. The CMS Muon System is a vital component in the CMS detector, designed for the precise identification and measurement of muons (along with providing signals for triggering on them), which are crucial particles in many physics analyses.

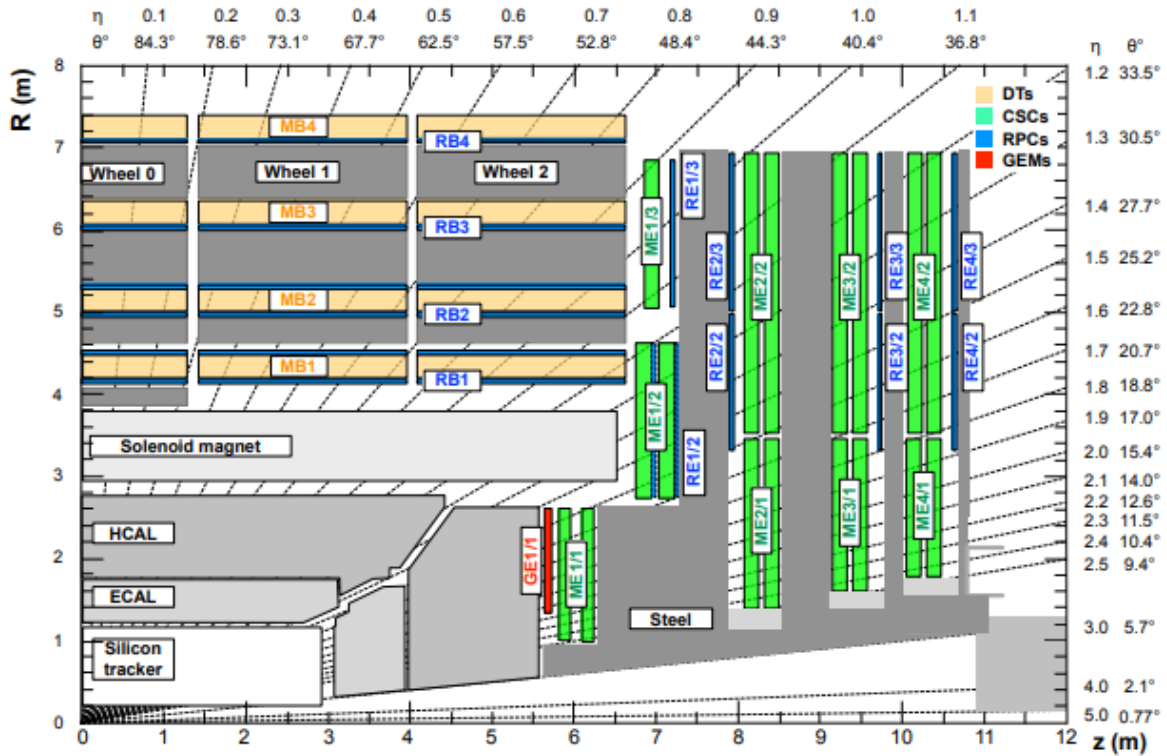


Figure 3.5: Schematic view of the r-Z plane of the CMS detector showing the muon systems. The interaction point is in the lower left corner [23]

The Muon System comprises three main subdetectors: the Drift Tube (DT) system, the Cathode Strip Chambers (CSC) system, and the Resistive Plate Chambers (RPC) system. These subdetectors are strategically placed in the barrel and endcap regions to provide complete coverage.

Following proton-proton collisions, a diverse array of particles is generated, each depositing its energy in specialized subdetectors as discussed above. The process of quark hadronization results in the formation of jets, observed as clusters of energy in calorimeters. Isolated electrons and photons, characterized by specific interaction patterns, are confined within the crystals of the ECAL. Muons, known for their minimum ionizing nature, traverse the calorimeter and are subsequently detected by muon subsystems.

3.2 The particle flow algorithm

While the deployment of complex detector systems is a prerequisite, it alone does not provide a comprehensive event description. Several sophisticated algorithms are used to identify various particles and measure their four-momenta and positions with the best precision possible. Hence, the particle flow reconstruction approach is employed. This methodology involves the systematic correlation of basic detector elements across all layers, enabling the precise identification

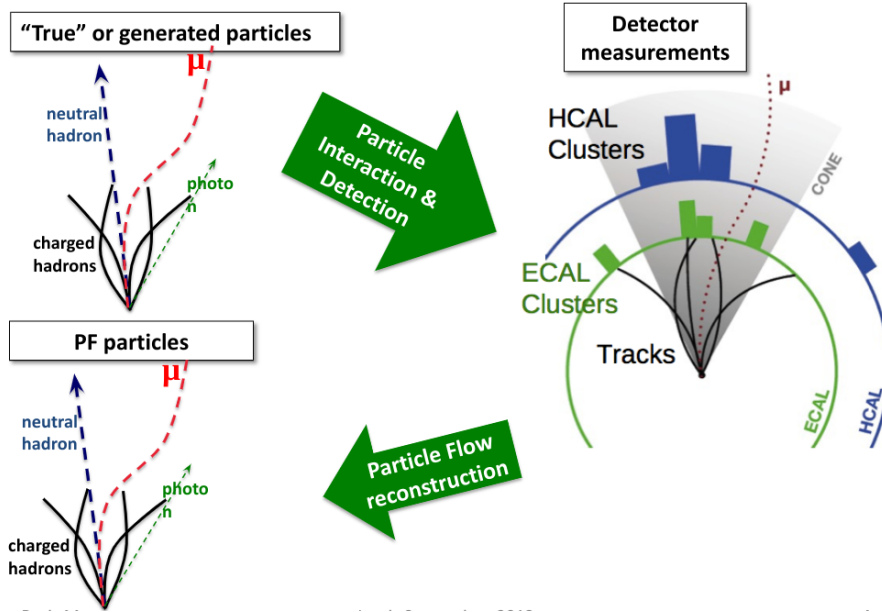


Figure 3.6: Particle flow used to reconstruct the event. Courtesy of Boris Mangano

of each final state particle and facilitating the comprehensive reconstruction of their properties. The fundamental elements constitute tracks from the tracker, clusters of energy deposits from the calorimeters and muon subsystems which are collectively referred to as *PFElements*.

The particle flow algorithm considers a spectrum of final state products, including charged and neutral hadrons, electrons, photons, and muons, however, it can not differentiate between different species of hadrons. These constitute the *PFCandidates*. The *PFCandidates* are further examined to be qualified as various particle candidates like electrons, photons or muons or more complex objects like tau-leptons or missing transverse momentum. These are also inputs to reconstruct jets, as shown in figure 3.6, using dedicated clustering algorithms.

Jets are the result of the hadronization process involving gluons and quarks and typically consist of multiple particles, as shown in figure 3.7 (right), due to the complex interactions governed by Quantum Chromodynamics (QCD). Although few charged particles within a jet may bend in the magnetic field, these appear conical in shape and are characterised by energy deposits both in ECAL and HCAL, shown in figure 3.7 (left). For studying the compositions of jets, it is important for us to calibrate hadrons which are a major component of jets, this is done using various techniques, such as the χ^2 method discussed later. In this thesis, a new calibration method for hadrons based on GNNs will also be discussed.

Notably, this algorithm remains effective even in scenarios involving coarse-grained detections, such as in the HCAL. In this context, where signals from different particles, especially within jets, may merge, the algorithm leverages its superior track resolution to ensure accurate reconstruction.

The particle flow algorithm consists of various steps involving different algorithms, some of them are discussed below

3.2.1 Reconstruction of charged objects

The reconstruction of charged particles is a fundamental aspect of unraveling the complex tapestry of high-energy collisions within the CMS experiment. At the core of this process lies the tracking system, comprising detectors such as the pixel detector and silicon strip tracker

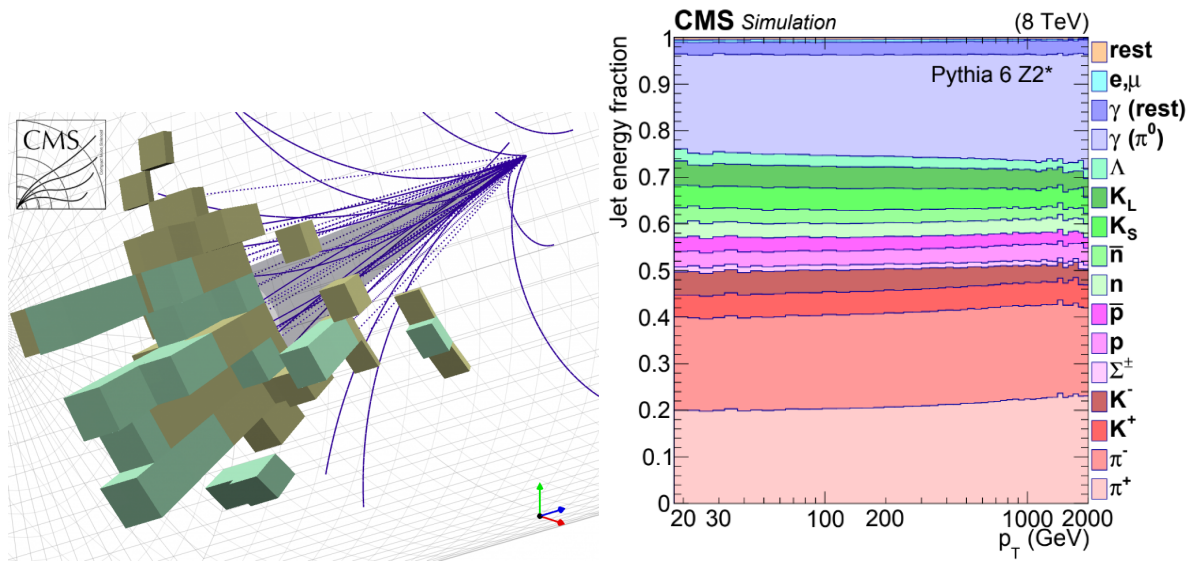


Figure 3.7: Left: Jets are shown using a dark cone and its energy deposition in ECAL and HCAL [15]. Right: Jet energy composition as a function of jet p_T at $|\eta| < 1.3$ in simulated samples [16]. The fraction of various particles are stacked on top of each other in the same order as the legend.

immersed in a magnetic field of approximately 4 Tesla. This is a highly granular system with almost one million independently readout detector cells, also called channels. Energy deposited by charged particles in a channel along with its position coordinates in the detector is called a "hit" or "rechit". The hits created by charged particles provide essential information for trajectory reconstruction. The momentum measured using the reconstructed trajectory of the charged particle in the magnetic field combined with the energy deposits in calorimeter provide an accurate estimation of properties of the charged particle.

One of the challenges in reconstructing tracks of charged particles is increasing the efficiency of track reconstruction without deteriorating the misreconstruction rate. The efficiency of track reconstruction is defined as fraction of simulated tracks reconstructed with at least half of the associated simulated hits, and with less than half of unassociated simulated hits. The misreconstruction rate is the fraction of reconstructed tracks that cannot be associated with a simulated track. If a track of a charged hadron is missed by the tracking algorithm, it will be reconstructed as a neutral hadron with biased direction and a degraded energy resolution thus affecting the reconstruction of jets. The track efficiency can be substantially improved by accepting tracks with smaller p_T and with less number of hits, however, this will cause the misreconstruction rate to spike up.

To overcome this challenge, an iterative tracking algorithm is employed. The iterative tracking process begins with the generation of seed tracks using hits in the pixel detector. These seed tracks are refined through multiple iterations, adjusting parameters to optimize agreement with the measured hits. The algorithm considers factors such as energy loss, multiple scattering, and the magnetic field within the tracker. After each iteration, low-quality or outlier tracks are filtered out based on predefined criteria.

In the realm of electron identification, the iterative tracking approach plays a pivotal role. Electrons, characterized by energy-spreading phenomena like bremsstrahlung, pose a challenge in associating hits in the ECAL with their tracks. To tackle the substantial uncertainty arising from this, the iterative tracking approach is employed. Efficiency of this approach is illustrated

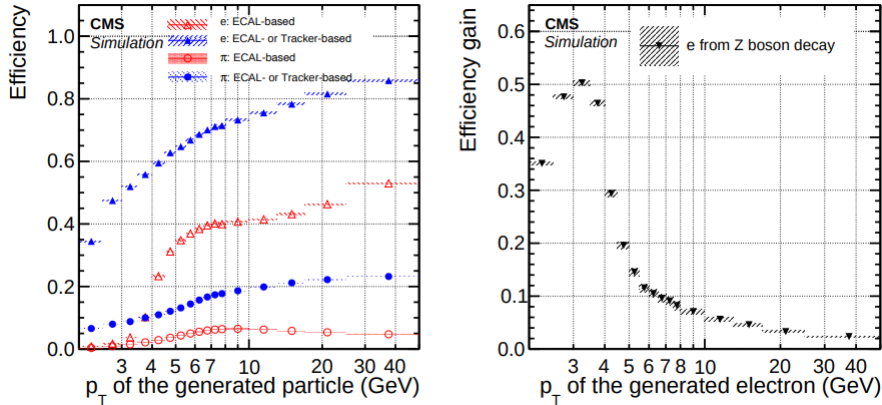


Figure 3.8: The left side illustrates electron seeding efficiency (triangles for electrons, circles for pions) versus p_T from a simulated event sample enriched in b quark jets (p_T : 80-170 GeV) with at least one semileptonic b hadron decay [8]. Both ECAL-based seeding only (hollow symbols) and combined with tracker-based seeding (solid symbols) are depicted. On the right, the tracker-based seeding’s absolute efficiency gain for electrons from Z boson decays is shown against p_T , with shaded bands denoting p_T bin size and statistical uncertainties on efficiency.

in figure 3.8. Additionally, since electrons can emit high-energy photons, causing significant trajectory shifts, a Gaussian Sum Filter (GSF) fitting method is employed. This sophisticated technique aids in mitigating the challenges posed by the radiative nature of electrons, resulting in a robust electron track reconstruction.

To accurately reconstruct the energy of the primary electron or photon, a dedicated algorithm is employed to combine clusters of hits in the ECAL into a single supercluster.

3.2.2 Electromagnetic clusters

Calorimeter clusters are pivotal in the PF algorithm, detecting and quantifying the energy and direction of particles. The ECAL is responsible for the reconstruction of electromagnetic objects like electrons and photons. Objects associated with a track in the tracker are classified as electrons, while those without a track are identified as photons.

In ECAL, the clustering algorithm starts by identifying cluster seeds, cells exceeding a defined threshold and having greater energy than neighboring cells. Topological clusters are then formed by aggregating cells with a common corner and energy exceeding a specified threshold. A subsequent expectation-maximization algorithm, based on a Gaussian mixture model, refines the clusters within a topological cluster. However, due to the showering nature of electrons and photons, neighboring clusters around the seed cluster, accounting for processes like pair production and bremsstrahlung loss, also need to be included. This is done using the mustache algorithm. The mustache algorithm solely relies on information from the ECAL and preshower detector. Since the low p_T constituents of electromagnetic showers bend in the azimuthal(ϕ) direction, giving showers a mustache shape in the $\delta\eta$ - $\delta\phi$ plane with respect to seed as shown in figure 3.9. Multiple clusters in the mustache are combined to form a supercluster, this process is referred to as superclustering.

The superclusters may be refined using a refined algorithm which incorporates tracking information from converted photons or brems from electrons in addition to ECAL data. At each tracker layer, the trajectory of the reconstructed track is extrapolated to form a "bremsstrahlung tangent," which is then linked to a compatible ECAL supercluster. Clusters associated with this

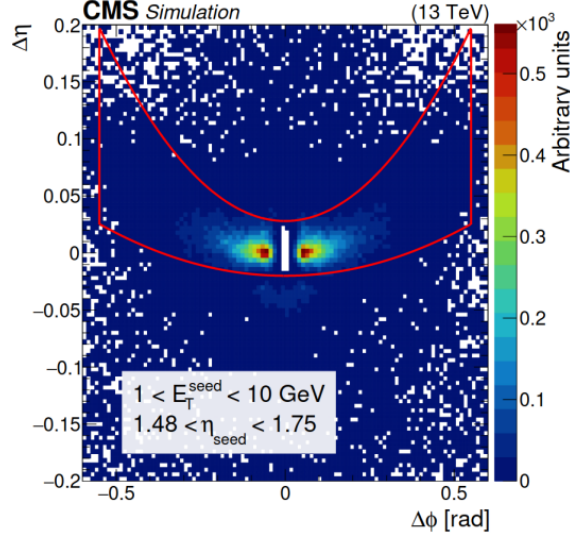


Figure 3.9: Distribution of $\Delta\eta$ versus $\Delta\phi$ with respect to seed for simulated electrons with $1 < E_T^{\text{seed}} < 10\text{GeV}$ and $1.48 < \eta_{\text{seed}} < 1.75$ [11]. The red line contains approximately the set of clusters selected by the Mustache algorithm. The white region at the centre of the plot represents the η - ϕ footprint of the seed cluster.

"bremsstrahlung tangent" are then incorporated into the refined Supercluster.

These algorithms play crucial roles in accurately reconstructing the energy and position of electrons and photons within the CMS detector at CERN. Calibrating calorimeter clusters is crucial for accurate energy measurements. The ECAL calibration involves both absolute and relative calibrations. Absolute calibration is refined using collision data collected for resonances of different masses, while residual energy calibration accounts for clustering algorithm thresholds. Calibration parameters, including α , β , and γ , are determined to minimize the χ^2 function.

The calibration process extends to ECAL endcaps, where calibration is expressed as a function of ECAL and preshower layer energies. This ensures accurate measurements even in regions partially shadowed by the preshower. Calibration parameters are determined through a similar χ^2 minimization process.

The calibration of electromagnetic deposits in ECAL is mathematically represented by the following equation [8]:

$$E_{\text{calib}} = \alpha(E_{\text{true}}, \eta_{\text{true}}) \cdot E_{\text{ECAL}} + \beta(E_{\text{true}}, \eta_{\text{true}}) \cdot [E_{\text{PS1}} + \gamma(E_{\text{true}}, \eta_{\text{true}}) \cdot E_{\text{PS2}}] \quad (3.3)$$

Here, α , β , and γ depend on the true energy (E_{true}) and pseudorapidity (η_{true}) of the generated photon.

3.2.3 Hadronic Clusters

Hadron Calorimeter (HCAL) follows a similar clustering procedure as the electromagnetic clusters, involving seed identification and topological cluster formation. However, these clusters may be across multiple layers of the HCAL.

A charged hadron is identified by a geometric link in the $\eta - \phi$ plane between a reconstructed track and one or more calorimeter clusters. A neutral hadron is identified by a geomet-

ric link between ECAL and HCAL calorimeter clusters or just clusters in HCAL without any track link.

Calibration parameters in HCAL are determined using a comparable χ^2 minimization process to ensure accurate energy measurements for hadronic clusters. As stated above, calibration of hadrons is difficult since they may start depositing their energy in ECAL. In such cases, different calibration coefficients are obtained for energy deposited in ECAL and HCAL. For details, please refer to [8].

Chapter 4

Deep learning

In 1943, the McCulloch-Pitts neuron model, which served as a theoretical foundation for artificial neural networks, was proposed by Warren McCulloch and Walter Pitts. Subsequently, in 1957, Frank Rosenblatt introduced the perceptron algorithm, an extension of the McCulloch-Pitts model, which included weighted inputs and a training procedure for binary classification of linearly separable data. In modern deep learning networks, a perceptron or neuron consists of two main components: first, a weighted sum on the inputs, and second, an activation function that introduces non-linearity. Additionally, a perceptron may be provided with an extra unity input known as bias. It is important to note that this learning process is a form of supervised learning, where the perceptron is trained on labeled data to determine its weights. Multiple such neurons can be interconnected to form a network which known as a Multi Layered Perceptron (MLP), also known as a dense layer in deep learning networks.

An MLP is versatile, capable of both regression and classification tasks. In regression, it predicts numerical values based on inputs, while in classification, it assigns inputs to specific classes by calculating probabilities for each class and selecting the one with the highest probability. However a simple MLP takes high level features as input.

Deep learning refers to a class of machine learning techniques that leverage complex neural network architectures with many layers to model data. The training procedure for these networks consists of two processes: forward propagation, which produces an output based on the current weights, and backward propagation, which updates the weights through a process known as stochastic gradient descent, as described below. This is done iteratively until a good prediction rate is achieved.

Since the advent of modern GPU technologies like CUDA (Compute Unified Device Ar-

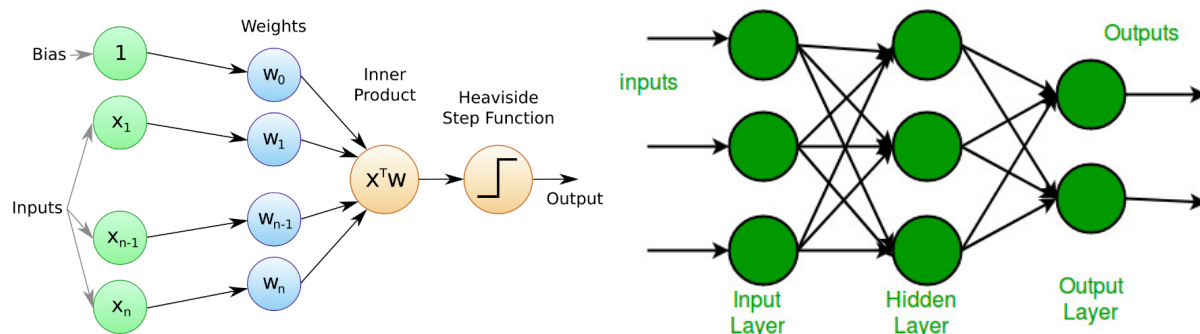


Figure 4.1: The perceptron (left) [3]. A multi layered perceptron (MLP) showing input and output layers (right). The number of nodes in input(output) layer determines the dimensionality of the input(output).

chitecture) , deep learning has undergone substantial improvements in performance by parallelizing the forward propagation step to occur in batches and doing a single backward pass for multiple forward passes. This greatly reduces the training time and allows for more complex architectures to be trained.

4.1 Stochastic gradient descent

The training procedure for neural networks commences with the initialization of the weights of each neuron to random values. Subsequently, the objective is to optimize these weights to minimize the error in prediction or loss. This optimization process relies on a loss function(L), and in our case, we employ the mean squared error (MSE) defined as:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

During the backpropagation step, the following operations are performed:

- Calculate the gradient of the loss function with respect to the weights $\frac{\partial L}{\partial w}$.
- Update the weights using a specified optimization algorithm. This iterative process involves adjusting the weights in the opposite direction of the gradient to minimize the loss. In basic Stochastic Gradient Descent (SGD), the update formula is:

$$w_{t+1} = w_t - \eta \cdot \frac{\partial L}{\partial w}$$

Where η is defined as the learning rate and can be chosen to be small or large depending on the problem.

- Iterate through the dataset multiple times (epochs) to improve the model's performance. Each iteration updates the weights incrementally, refining the model's ability to generalize to new data.
- Monitor the learning process by evaluating the model on a separate validation set, preventing overfitting and ensuring good generalization.

While this technique ensures convergence to a minimum in the loss function, it does not guarantee a global minimum. A modification to the standard weight update formula is introduced with the Adam optimization algorithm which uses:

$$w_{t+1} = w_t - \frac{\eta}{\sqrt{v_t} + \epsilon} \cdot m_t$$

Here, w_{t+1} is the updated weight, w_t is the current weight, η is the learning rate, m_t is the first-order momentum, v_t is the second-order momentum, and ϵ is a small constant for numerical stability. First order momentum refers is a running average of the gradients and is defined as:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot \frac{\partial L}{\partial w_t}$$

An analogy can be drawn to a rolling ball on a hilly terrain, where the momentum of the ball aids in navigating and reaching the deepest valley. This momentum helps the optimization process overcome local minima and escape saddle points, contributing to more effective weight updates and faster convergence.

Along with optimizers, learning rate schedulers are employed to adapt the learning rate during training.

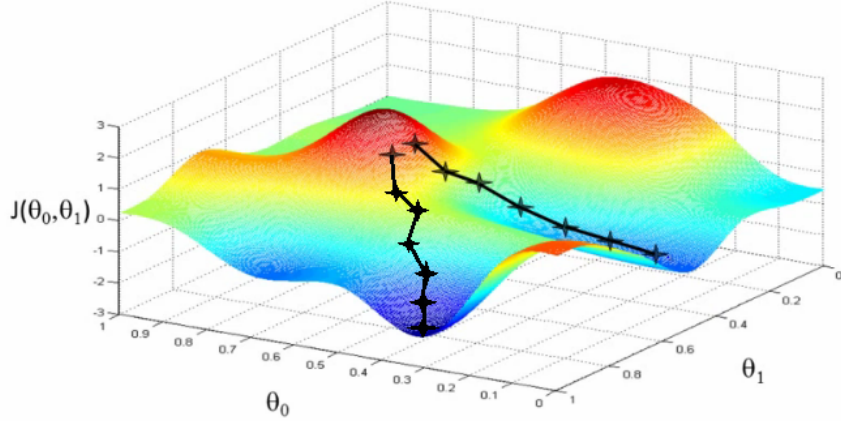


Figure 4.2: An example of loss function (J) on z axis and its values for different values of the weights θ_0 and θ_1 . The black line denoted the path taken by gradient descent to reach global minima. Courtesy of Howie Choset

4.2 Graph convolution

A network based on only MLP can only take high level data as input. To harness the capabilities of neural networks for training on low level features like image-based data, characterized by 2D matrices of pixel values across multiple channels (e.g., RGB), convolutional neural networks (CNNs) are employed. In this context, each image undergoes convolution with a kernel matrix containing trainable parameters. This operation not only captures information at each pixel but also encodes details about its neighboring context, enabling the network to discern spatial patterns effectively. In case of our calorimeters ECAL and HCAL have different granularity and the particle showers are stochastic with irregular boundaries, leading to sparse images. Since CNNs require inputs to be zero-padded to make uniform input grids, thereby increasing the size of the inputs, which proves to be challenging and ineffective.

In response to these challenges posed by irregular geometries, graph-based neural networks have been developed. Unlike traditional CNNs, which operate on regular grids, graph-based models are well-suited for scenarios where irregularities are present. In graph-based neural networks, the input data is represented as a graph, where in our case nodes correspond to individual hits in the detector, and edges capture relationships between them. The model dynamically learns the connectivity patterns, making it adaptable to irregular structures.

In response to these challenges, EdgeConv, a specific graph convolution operation, has proven effective. This operation is analogous to the convolution operation done in CNNs. We start by constructing a graph joining k nearest neighbours to each node in a higher dimensional feature space. Formally, if we have a F dimensional point cloud with n points, denoted by $X = x_1, x_2, \dots, x_n \in \mathbb{R}^F$, we compute a directed graph $G = (V, E)$ representing the local point cloud structure, where $V = \{1, \dots, n\}$ and $E \subseteq V \times V$ are the vertices and edges, respectively. In the simplest case, we construct G as the undirected k -nearest neighbor (k -NN) graph of X in \mathbb{R}^F . The graph includes a self-loop, meaning each node also points to itself. Following this operation we get a list of the nearest neighbours to each node.

Now that we have our graph constructed, we are ready to perform EdgeConv, for which we need to define our edge features (e_{ij}) between each node pair.

$$e_{ij} = h_{\theta}(x_i, x_j) \quad (4.1)$$

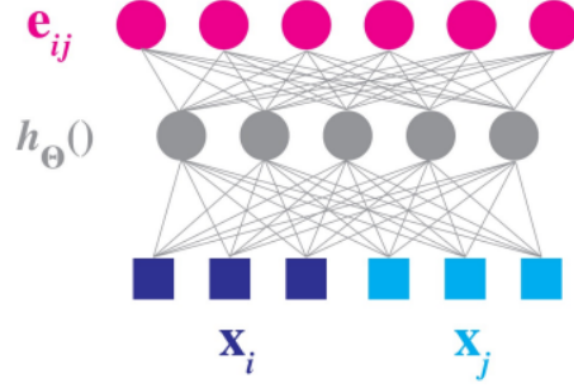


Figure 4.3: An example of how the nonlinear function operates on the features of two nodes x_i and x_j . In this illustration, the dimensionality of edge features (6) is double than that of the feature space ($F = 3$). [9]

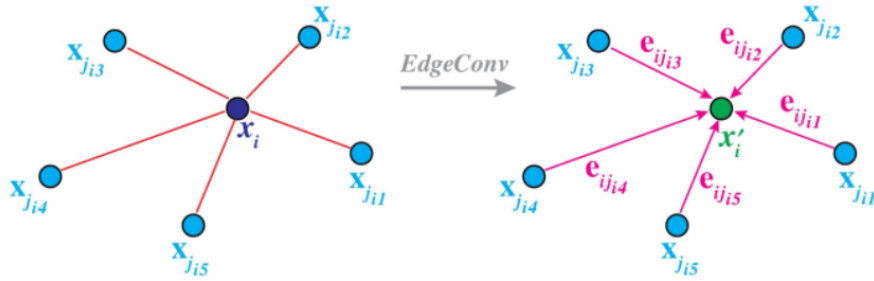


Figure 4.4: An illustration of the edgeconv operation. $x_{j,i1}, x_{j,i2}, \dots, x_{j,i5}$ represent the $k=5$ nearest neighbours of x_i and $e_{ij,i1}, e_{ij,i2}, \dots, e_{ij,i5}$ represent the corresponding edge features [9]

where h is a non linear function with a set of learnable parameters (θ)

$$h : R^F \times R^F \rightarrow R^F \quad (4.2)$$

To mimic a nonlinear function, we use a FCNN with a set of learnable parameters. The goal is to have a suitable function capable of capturing all the information from the connected nodes and finally do ‘EdgeConv’ to aggregate all this information and pass it into the central node along with its own information such that the node learns about its surrounding nodes (Illustrated in Fig 4.3).

We define the EdgeConv operation by applying a feature-wise symmetric aggregation operation \square (e.g., sum or max) on the edge features associated with all the edges emanating from each vertex.

$$x'_i = \square_{j:(i,j) \in E} h_{\theta}(x_i, x_j) \quad (4.3)$$

Making analogy to convolution along images, we regard x_i as the central pixel and $\{x_j : (i, j) \in E\}$ as a patch around it. Consequently, the i 'th node gets updated with information about its k nearest neighbours (Illustrated in 4.4). This exchange of information between nodes is known as message passing and the fully connected layers in the non linear function h are known as message passing layers.

Furthermore, for our use case , we define h as follows. For one message passing layer, the

edge features are defined as:

$$e'_{ijm} = \text{ReLU}(\theta_m \cdot (x_j - x_i) + \phi_m \cdot x_i), \quad (4.4)$$

Where ReLU is the rectified linear unit activation function. $\Theta = (\theta_1, \dots, \theta_M, \phi_1, \dots, \phi_M)$, where each $\theta_m \in R^F$ is the set of trainable parameters. M corresponds to the dimensions of the output which is analogous to the number of kernels in CNNs. Furthermore, the edge features can become more complex functions of $(x_j - x_i, x_i)$ depending on the number of message passing layers, similar to how a simple FCNN can become more complex and robust by adding more layers.

Finally, we find the updated features of a node by aggregating the edge features of its neighbouring nodes.

$$x'_{im} = \max_{j:(i,j) \in E} e'_{ijm} \quad (4.5)$$

The output is a M dimensional point cloud with same number of nodes as the initial graph.

Our definition of the edge features preserves the property of translational invariance but only partially. If a constant T is added to the features of each of the nodes, our edge features become

$$e'_{ijm} = \text{ReLU}(\theta_m \cdot (x_j + T - x_i - T) + \phi_m \cdot (x_i + T)), \quad (4.6)$$

$$e'_{ijm} = \text{ReLU}(\theta_m \cdot (x_j x_i) + \phi_m \cdot (x_i + T)), \quad (4.7)$$

Hence model is partially translationally invariant in the sense that the model takes account into the local geometry of patches while keeping global shape information.

4.2.1 Graph clustering

Following graph convolution, we cluster the graph and pool nodes in individual clusters, allowing relevant local information to be propagated to global intermediate feature sets. Graph clustering can be treated as an indexing of nodes. A variety of clustering algorithms have recently been proposed to handle data that is not linearly separable, for our case we will use graclus, a multilevel graph clustering algorithm. This algorithm performs graph clustering without the need of any eigenvector computation, which can be prohibitive for very large graphs.

The algorithm consists of three main phases. First is the Coarsening Phase, where the graph is iteratively transformed into smaller graphs, starting from the initial graph G_0 . Sets of nodes in each level G_i are combined to create supernodes in the next level G_{i+1} , with the edge weights out of the supernode being the sum of the original nodes' edge weights. Starting with all nodes unmarked, it selects vertices randomly and merges each with the unmarked vertex exhibiting the highest edge weight. The process continues until all vertices are marked, emphasizing significant edges and preserving overall graph structure during coarsening. This is known as heavy edge coarsening.

The second phase is the Base Clustering Phase, which efficiently clusters the base graph by selecting random vertices and expanding regions around them in a breadth-first manner to form clusters. This phase has the shortest runtime among all three phases, and various alternative algorithms can be utilized for this step. The third phase is refinement, where given a graph G_i , the algorithm forms the graph G_{i-1} from the previous coarsening level $G_i - 1$. The clustering in G_i induces a clustering in G_{i-1} , where if a supernode in G_i is in a cluster, all nodes in G_{i-1} formed from that supernode are also in the same cluster. This initial clustering for G_{i-1} is then enhanced using a refinement algorithm to improve the overall clustering quality.

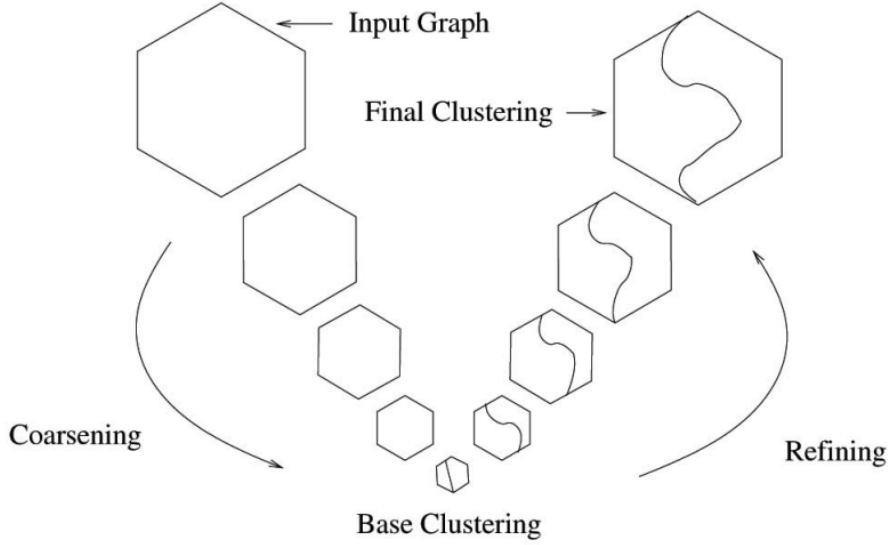


Figure 4.5: An overview of the graclus clustering algorithm [6]

4.3 Dynamic Reduction Network(DRN)

The deep learning architecture employed for this analysis is the dynamic reduction network based on GNN, specifically designed to process data from the ECAL. The input to this network consists of the spatial coordinates (X, Y, Z) of the hits, along with their corresponding energies. These input values are normalized to the range of 0 to 1. Each data point in the input corresponds to a unique event, and the dimensions of these data points are $(Number\ of\ rechits) \times 4$ where the 4 corresponds to the X, Y, Z and energy coordinate of each rechit.

Subsequently, the input undergoes processing through a Multilayer Perceptron (MLP) with multiple layers, effectively increasing the dimensionality from 4 to 128. This transformed space serves as the feature set for the subsequent graph generation step. This is illustrated as step 1 in figure 4.6.

A graph is generated using the k-nearest neighbors (kNN) approach, where each hit is represented as a node in the graph shown as step 2.1 in figure 4.6. The EdgeConv operation is then applied to capture information about neighboring hits in step 2.2. Following this, normalized edge weights are computed in step 2.3 according to the formula.

$$E_{ij} = e_{ij} \cdot \left(\frac{1}{deg(i)} + \frac{1}{deg(j)} \right)$$

Here, E_{ij} represents the calculated edge weight between an edge connecting node i and j, e_{ij} represents the distance in feature space between the nodes and $deg(i)$ and $deg(j)$ represents the number of connections to node i and j respectively. Using the calculated edge weights, nodes are clustered using the graclus algorithm in step 2.4. Nodes within a cluster are subjected to max-pooling in step 2.5.

As the number of nodes is reduced through this process, a new graph is generated iteratively, repeating the aforementioned steps based on the specified number of aggregation layers as illustrated in figure 4.6. Eventually, a global max-pooling operation is executed on the resulting graph, yielding a single node in step 3. The 128 features from this node are then fed through another MLP which is the outputnet in step 4, culminating in a single-node output, effectively providing a final numerical result for the analysis.

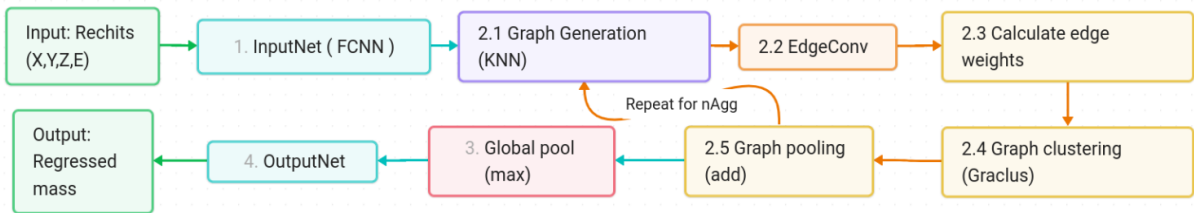


Figure 4.6: DRN architecture overview

Chapter 5

Reconstruction of mass of merged photon candidates in the CMS ECAL

As discussed earlier, in the 2HDM + singlet extension of the SM, we have a CP odd higgs a which can decay to 2 photons. This particle is a pseudoscalar and a axion-like particle which arises from the decay of the SM like Higgs (H) boson. (as shown in [5.1](#)).

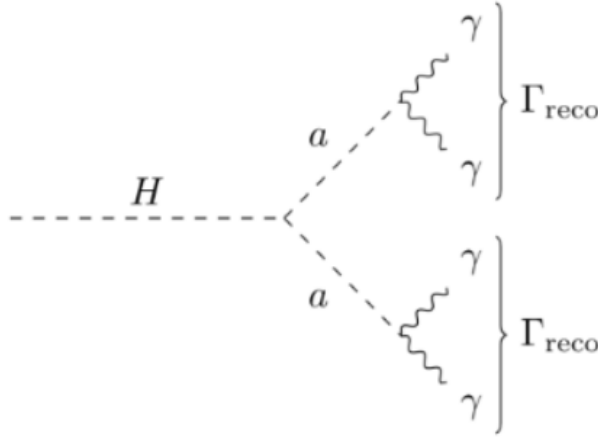


Figure 5.1: Feynman diagram for the decay $H \rightarrow aa \rightarrow \gamma\gamma \gamma\gamma$ with the photons from each a decay merged into a single reconstructed photon Γ .

Theoretically, this particle can have a mass below 1 GeV. As we go to lower masses of a , the photons tend to be merged. The angle between the 2 photons is governed by the equation

$$\cos(\theta) = 1 - \frac{m_a^2}{2E_{\gamma 1}E_{\gamma 2}} \quad (5.1)$$

We can also get a lower bound on the angle (as shown in appendix)

$$\cos(\theta) < 1 - \frac{2}{\gamma_a^2} \quad (5.2)$$

where γ_a is the boost of a defined as $\frac{E_a}{m_a}$. Hence, if we plot the angle between the two photons, we will get a peak around $\theta \approx 4/\gamma_a^2$ with some events at higher θ values as shown in as shown in figure [5.2](#)

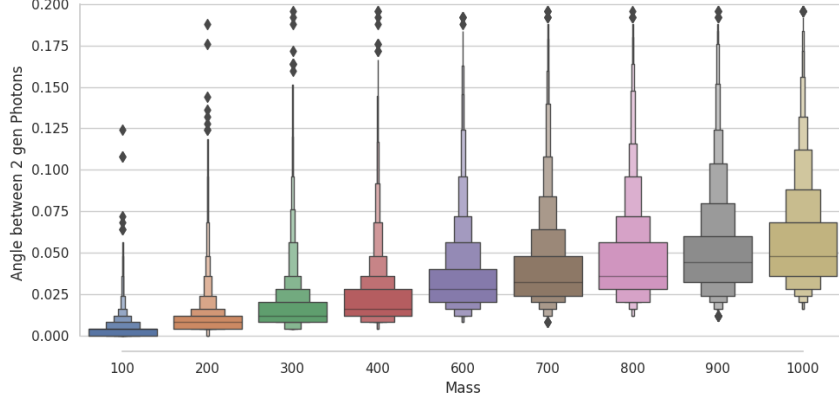


Figure 5.2: Boxen plot showing the distribution of decay angles (Y-axis) of a as a function of its mass (X-axis). Each box represents the interquartile range (IQR) of angles, with the median angle indicated by the horizontal line within the box.

The ECAL lacks the necessary precision with its granularity ($\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$) to identify merged photons resulting from the decay of highly boosted a . These photons primarily deposit their entire energy in few ECAL crystals. Figure 5.3 illustrates the angular separation between two photons for different boosts in terms of number of ECAL crystals in the η and ϕ directions. As the boost of the mother particles increases, the photons tend to deposit most of their energy in just a single ecal crystal. Because of this the ECAL superclustering algorithm is unable to isolate and detect these photons individually and reconstructs them as a single photon object. We use low level detector features as inputs to train a DRN model. To train our model, we use a sample of “a” bosons generated uniformly in the mass, transverse momentum, η and ϕ in the following kinematic ranges: .

- Mass of a : $0.6\text{GeV} < m_a < 2\text{GeV}$
- Transverse momentum of a : $20\text{GeV} < p_T < 100\text{GeV}$
- Pseudorapidity of a : $|\eta_a| < 1.4442$ (Barrel region)
- Phi of a : $|\phi_a| < \pi$

The 2D histograms for mass vs transverse momentum and η vs ϕ are shown in figure 5.4 showing that distribution of these quantities is uniform. The reason for choosing a flat mass distribution as a training sample is to ensure the model learns effectively and the prediction is not biased to preferentially predict certain mass points. The lower range of the mass starts from 0.6 GeV since this was being attempted for the first time and not a fundamental limitation of the network.

5.1 Sample production and processing

A sample of single $a \rightarrow \gamma\gamma$ events are generated in the above kinematic regimes in the following steps. We use CMS software framework (CMSSW) to generate, simulate and reconstruct our event samples. These events are overlaid with realistic pileup simulation expected in the pp collisions.

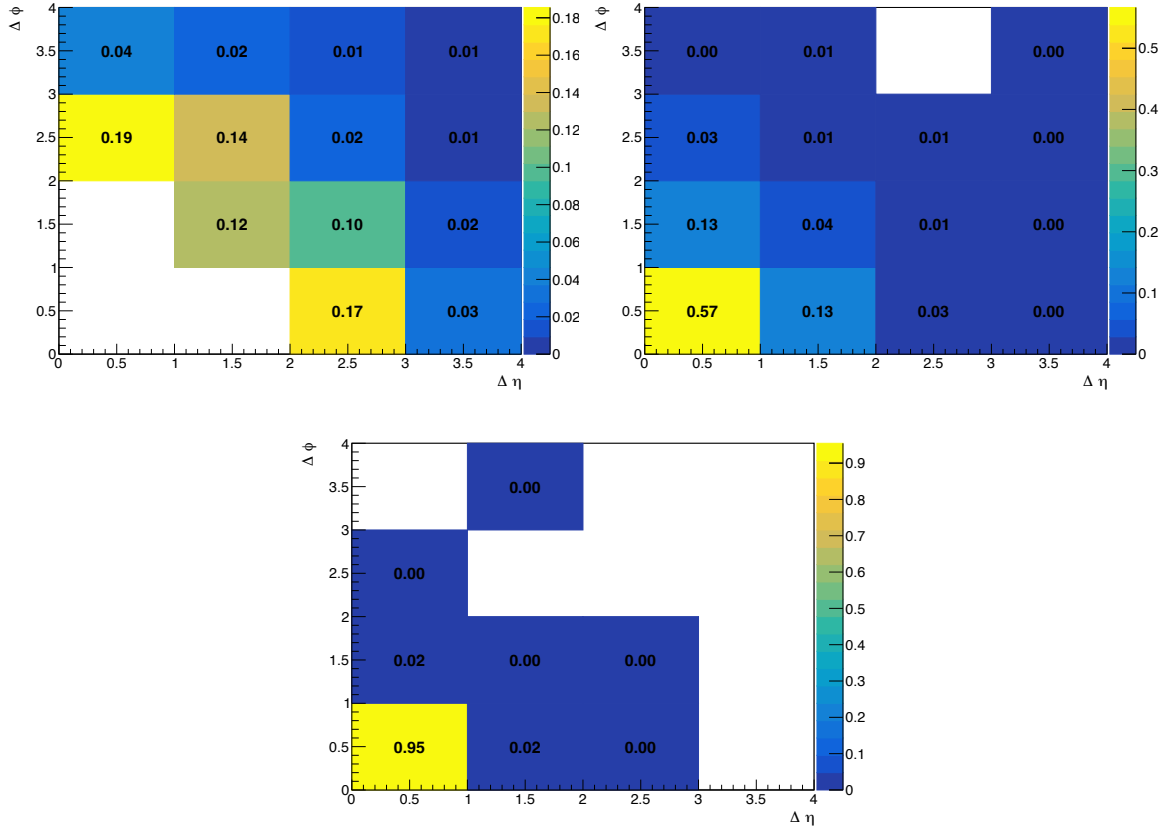


Figure 5.3: Generator level distribution of the separation between two photons in $a \rightarrow \gamma\gamma$ decay, as a function (η) and (ϕ), expressed in terms of the number of ECAL crystals. This plot illustrates the angular separation between the two photons in the decay process, providing insights into their spatial distribution within the detector, where η and ϕ are discretized in units of ECAL crystals. ($\Delta\eta \times \Delta\phi = 0.0174 \times 0.0174$). Top left plot is for a with a boost of 50 and mass 1 GeV. Top right plot is for a boost between 50 to 250 and mass 0.4 GeV. Bottom plot is for a boost between 250 to 500 and mass 0.1 GeV.

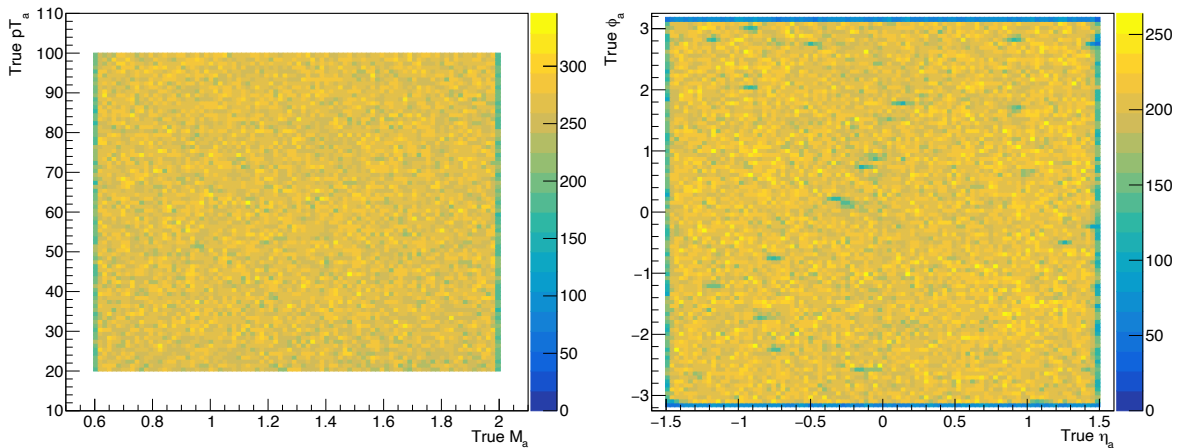


Figure 5.4: 2D histograms for generator level quantities showing that the kinematic variables were sampled from a uniform distribution in the above mentioned ranges. No underflow or overflow of events are in these figures.

5.1.1 Generation and simulation

In this step, a PYTHIA p_T gun is used to simulated $a \rightarrow \gamma\gamma$ events in the kinematic ranges defined above. The pileup samples are also generated using Monte Carlo (MC) to generate multiple overlapping pp collisions as explained later. We use the pileup samples centrally available for all CMS analyses. The particles are passed through a GEANT4 [2] detailed simulation of the CMS detector which simulates the response of various sub-detectors.

5.1.2 Digitisation

The output from the simulation step is in the form of energy deposits in different detector components. Digitization simulates the response of the detector to these energy deposits, converting them into digital signals that mimic the signals produced by real detector components. It takes into account factors such as electronic noise due to the effects of the front end electronics. Furthermore, neutrino samples are overlaid with existing samples to simulate pileup. The following dataset is used for this:

```
/Neutrino_E-10_gun/RunIISummer20ULPrePremix-UL18_106X_upgrade2018_realistic_v11_L1v1-v2/PREMIX
```

5.1.3 Reconstruction (Reco)

The reconstruction step starts with inferring the signal amplitude from the digitized samples for each detector cell or detector channel of various subdetectors. A detector cell along with its position coordinates and signal amplitude is referred to as a “rechit”. The next step is to reconstruct physics objects such as tracks, vertices, photons, electrons, muons, jets, and missing transverse energy (MET) using the rechits. As explained in Chapter 3, the CMS experiment uses a particle flow algorithm to optimally combine the information from various subdetectors to achieve the best resolution and efficiency for particle reconstruction [8]. Detector level calibration and physics object level energy corrections are applied to the reconstructed objects to account for detector effects and improve the accuracy of the measurements.

5.2 Event selection and model inputs

5.2.1 Event selection and data cleaning

To ensure a reliable sample of $a \rightarrow \gamma\gamma$ for training the model, we require the following criteria for selecting the events:

- Each event requires exactly 2 generator level photons which result from the decay of a
- Each event requires at least 1 rechit

These selections only reject about 4 percent of the generated events. Finally, the X,Y,Z and E information of rechits from relevant photon/s are stored in a data structure which can easily be accessed by our model.

5.2.2 Inputs to the model

We use rechit attributes, namely their energy and position coordinates, as inputs to our models. As explained in Chapter 3, electrons are reconstructed using Moustache algorithm. Here two approaches are employed to select the rechits for inputs to the model and performance is compared in the results section.

- **SupClu rechits:** In this approach, reconstructed photons are spatially matched with the simulated photons using $dR < 0.1$ criteria. We expect to get either one or two reconstructed photons depending on if the photons are merged or is one of them is too low energy to be clustered. The rechits belonging to these photon objects, superclusters, are then selected and processed for training the model.
- **SupClu+dR rechits:** In this method, we select only the reconstructed photon object with the highest energy. This is done because the highest energy rechit will most likely be a signal event and not one arising from pileup. We explicitly checked this. In addition to the rechits belonging to this photon super cluster, all additional rechits within $dR < 0.3$ from the seed rechit are also included. Since, the distance between the two photons in the η - ϕ plane is less than 0.3, this procedure includes all clustered and unclustered rechits from both photons.

It is important to note that *SupClu rechits* includes only those rechits that pass the ECAL clustering thresholds which may not be suitable for our analysis. A comparison of clustered and unclustered rechits is illustrated in figure [5.5](#). It can be seen that clustered rechits are of a higher energy than unclustered rechits.

5.3 Model training and results

From the above simulated samples, roughly 80% are used for training and the remaining 20% are used for validation. The training runs over several epochs and validation loss is calculated for each epoch. Finally the epoch with minimum validation loss is used for prediction. As explained in Chapter 4, our model is the Dynamic Reduction Network with the following hyperparameters:

- Input layers: 3

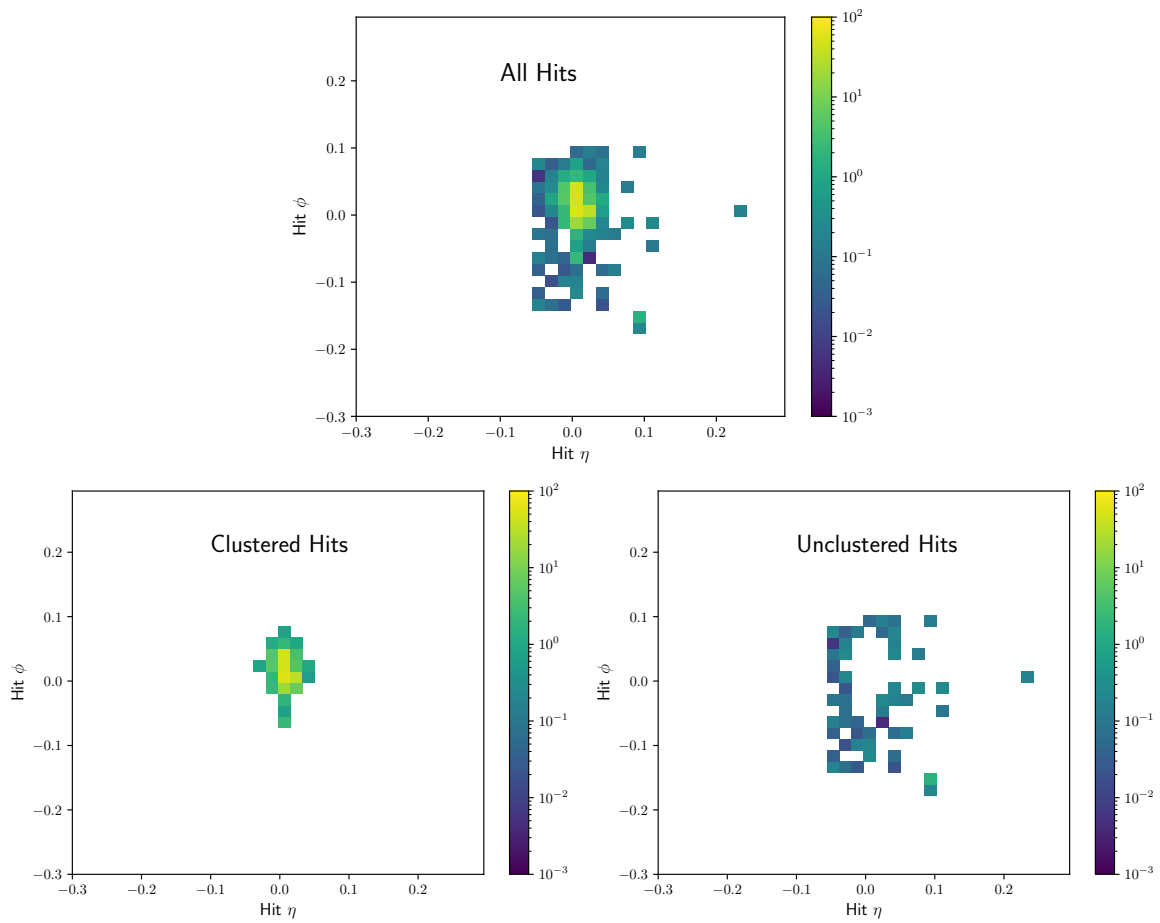


Figure 5.5: Plots showing the energy deposition by photons in the ecal in the $\eta - \phi$ plane. All rechits within a dR cone of 0.3 are shown on the top. The bottom left side of the figure shows the clustered rechits and bottom right side shows the unclustered rechits

- Aggregation layers (nAgg) : 2
- Message passing layers: 3
- Output layers: 2
- Hidden Dimension: 64
- Batch size (Train and valid) : 100
- AdamW optimizer with constant learning rate of 0.001
- Model Parameters: 62405
- Max pooling

5.3.1 Using SupClu rechits

Training was conducted on a dataset comprising of nearly 1 million events. The results of this training are depicted in Figure 5.6. The model demonstrates accurate predictions for the majority of events. However, a notable deviation occurs in the predicted versus true plot around the midpoint of the training range (approximately 1.3 GeV), indicating the model’s inability to correctly predict the mass for these events.

The model predicts these events to lie at 1.3 GeV, likely due to insufficient information from the rechits to precisely reconstruct the mass of the particle. This tendency to predict the midpoint of the sample range serves to minimize loss, a standard objective in deep learning networks.

To investigate this loss of information, particularly its dependency on the transverse momentum (p_T) of the simulated particle a , we have checked predicted versus true distributions in the bins of p_T of a , for both inclusive training and validation samples, shown in Figure 5.7. Notably, the flat line is more pronounced for lower p_T values, suggesting a potential deficiency in capturing low-energy rechits resulting from the energy deposition of soft photons originating from the decay of the parent particle a . These rechits remain uncaptured by the clustering algorithm.

Further reinforcing this understanding, the total rechit energy in the unclustered rechits as explained in 5.2.2 increases with the mass of particle a , as depicted in Figure 5.8. It can be seen that the second peak shifts towards higher energy ranges with increasing mass of a . This observation underscores the possibility of missing crucial information in certain events by excluding unclustered rechits from consideration.

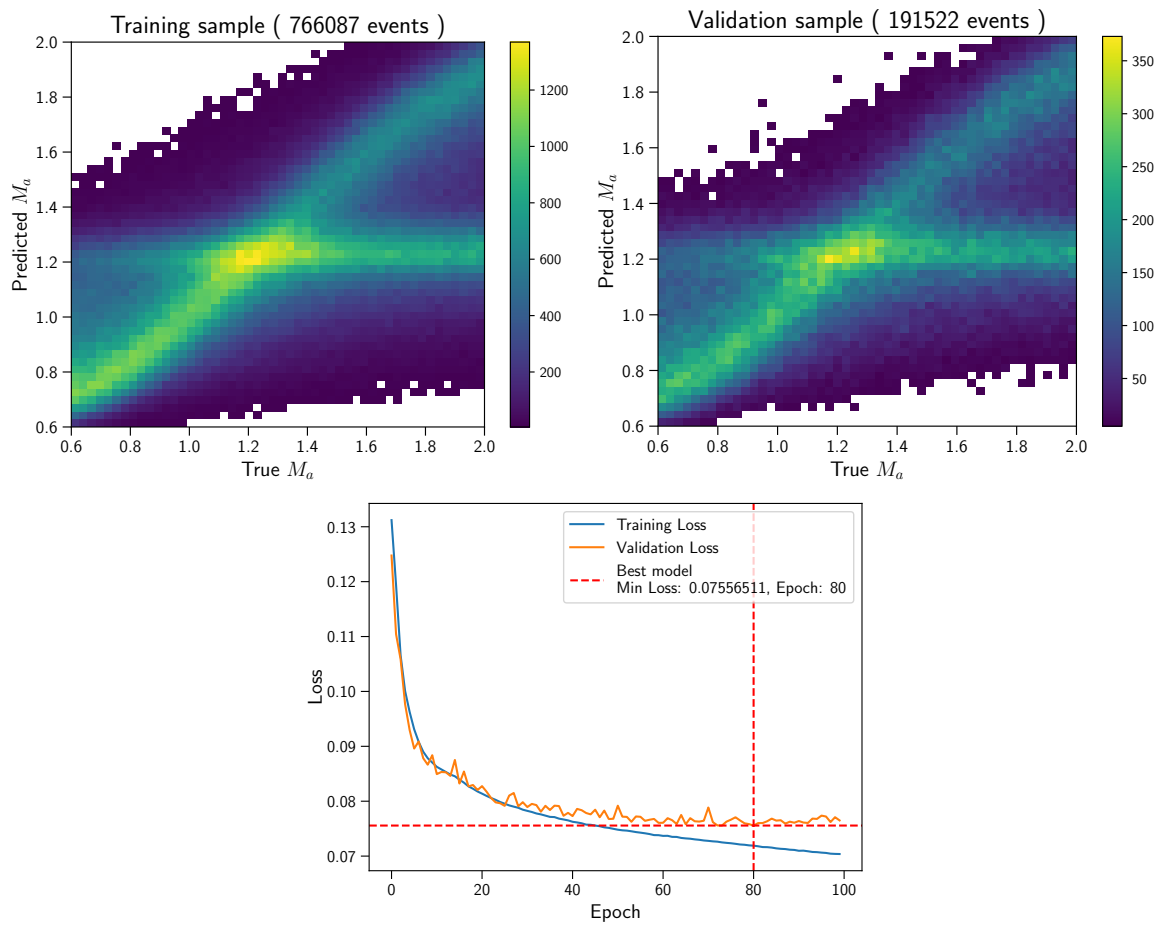


Figure 5.6: Predicted mass vs true mass of a for training dataset (top left) using **SupClu re-chits**. Predicted mass vs true mass of a for validation dataset (top right). Loss vs epoch showing the model at the best epoch with lowest validation loss

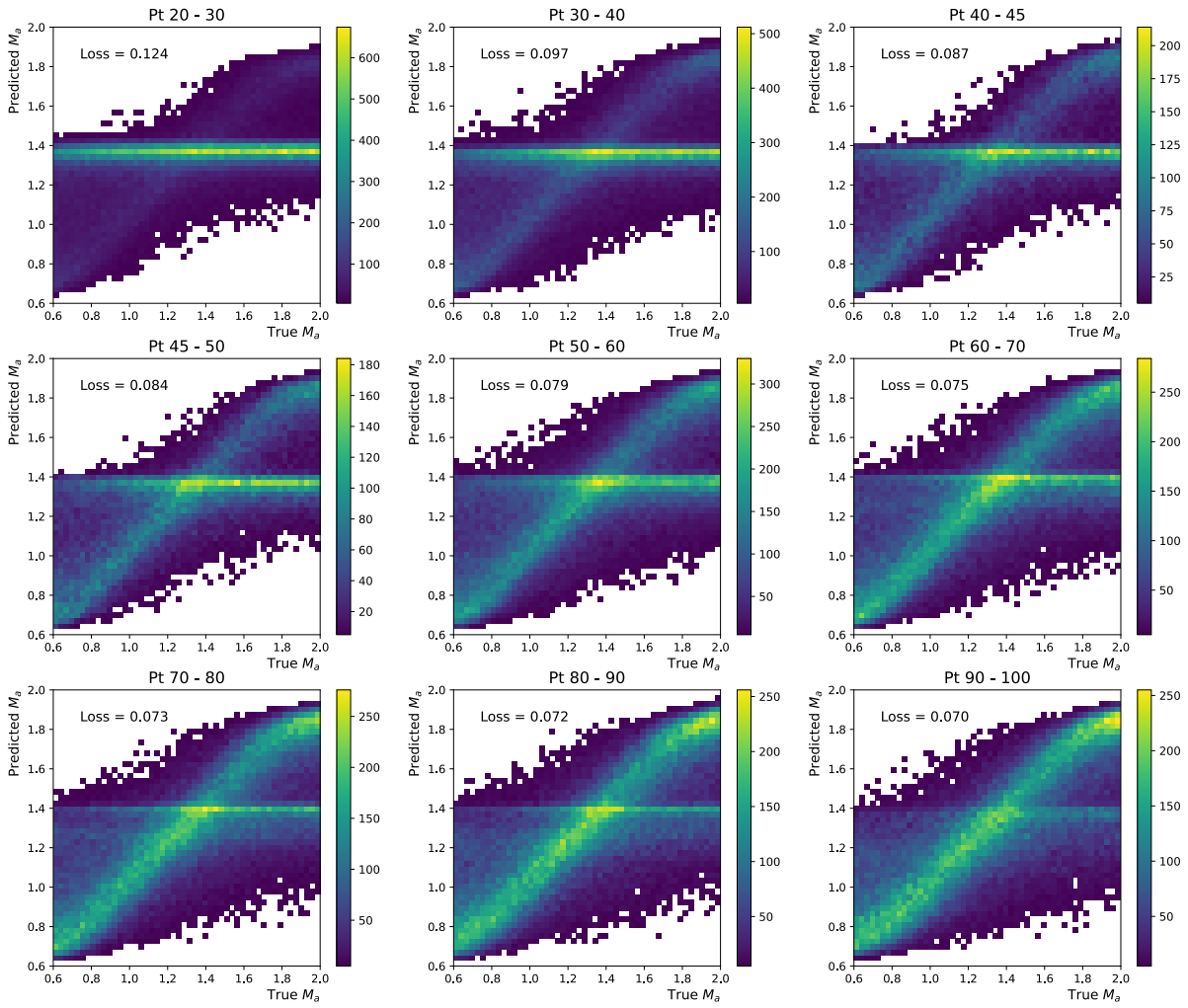


Figure 5.7: Predicted mass vs true mass in different bins of p_T of a using SupClu rechits

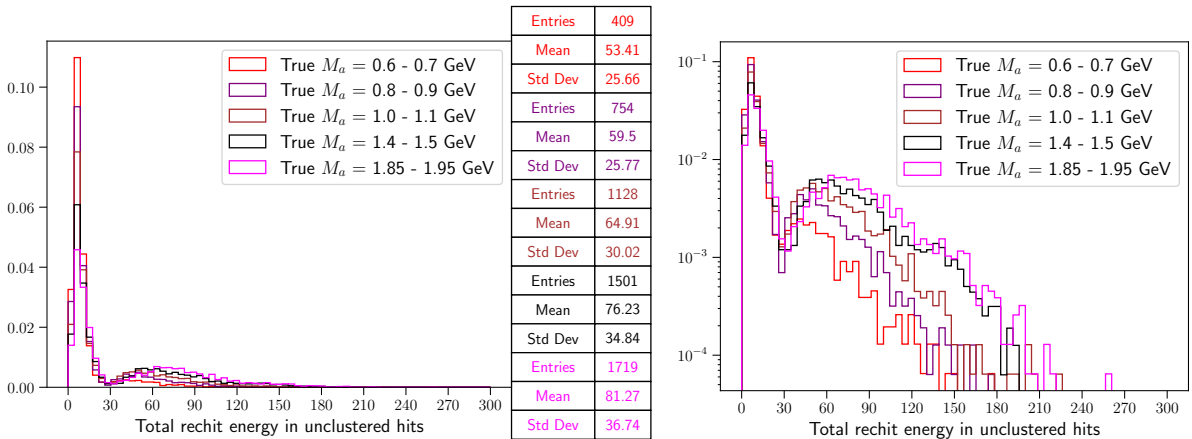


Figure 5.8: Total energy in unclustered rechits for different masses of a on linear scale(left), and log scale(right). This shows that the unclustered rechit energy increases as a function of mass of a

5.3.2 Using SupClu+dR rechits

Another training is performed, with 3 million events, this time including these unclustered rechits [5.2.2](#). Results are shown in [5.9](#). It can be seen that the performance of this training is significantly better than previous training with only clustered hits [5.2.2](#), with reduced effects of the flat line around the mean of the sample range and there is good agreement between predicted and true values. The p_T dependence of the flat line is also reduced as shown in figure [5.10](#). This means that unclustered rechits are crucial for the reconstruction of a .

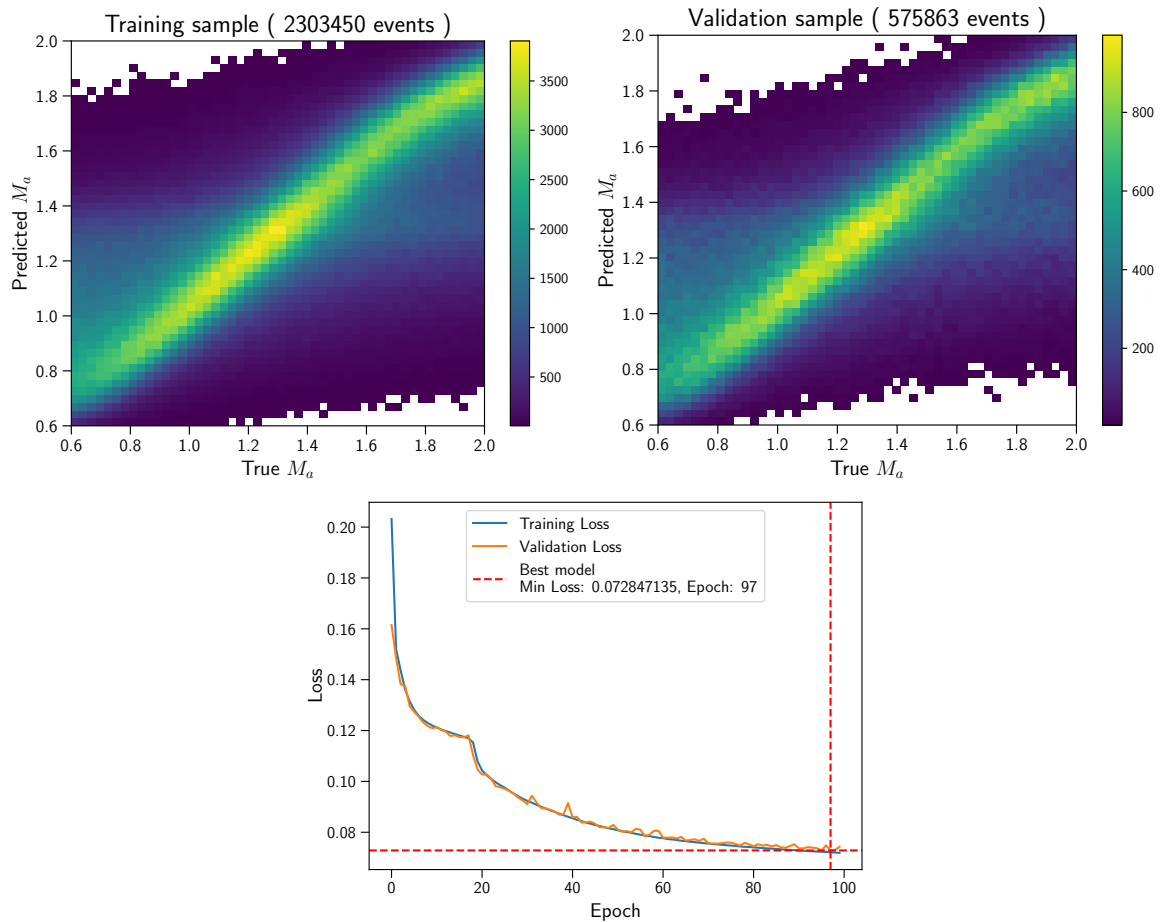


Figure 5.9: Predicted mass vs true mass of a for training dataset(top left) **using SupClu+dR rechits**. Predicted mass vs true mass of a for validation dataset(top right). Loss vs epoch showing the model at the best epoch with lowest validation loss

The mass of a would in reality be a single value. To validate our results, we have taken 40 MeV bins around few mass points and their 1D predictions are depicted in figure [5.11](#). The peak coincides with the true value for samples not too close to the boundary of the range of the training sample. For predictions close to the boundary effects such as a additional peak in the middle of the sample range is observed which can be treated as edge effects.

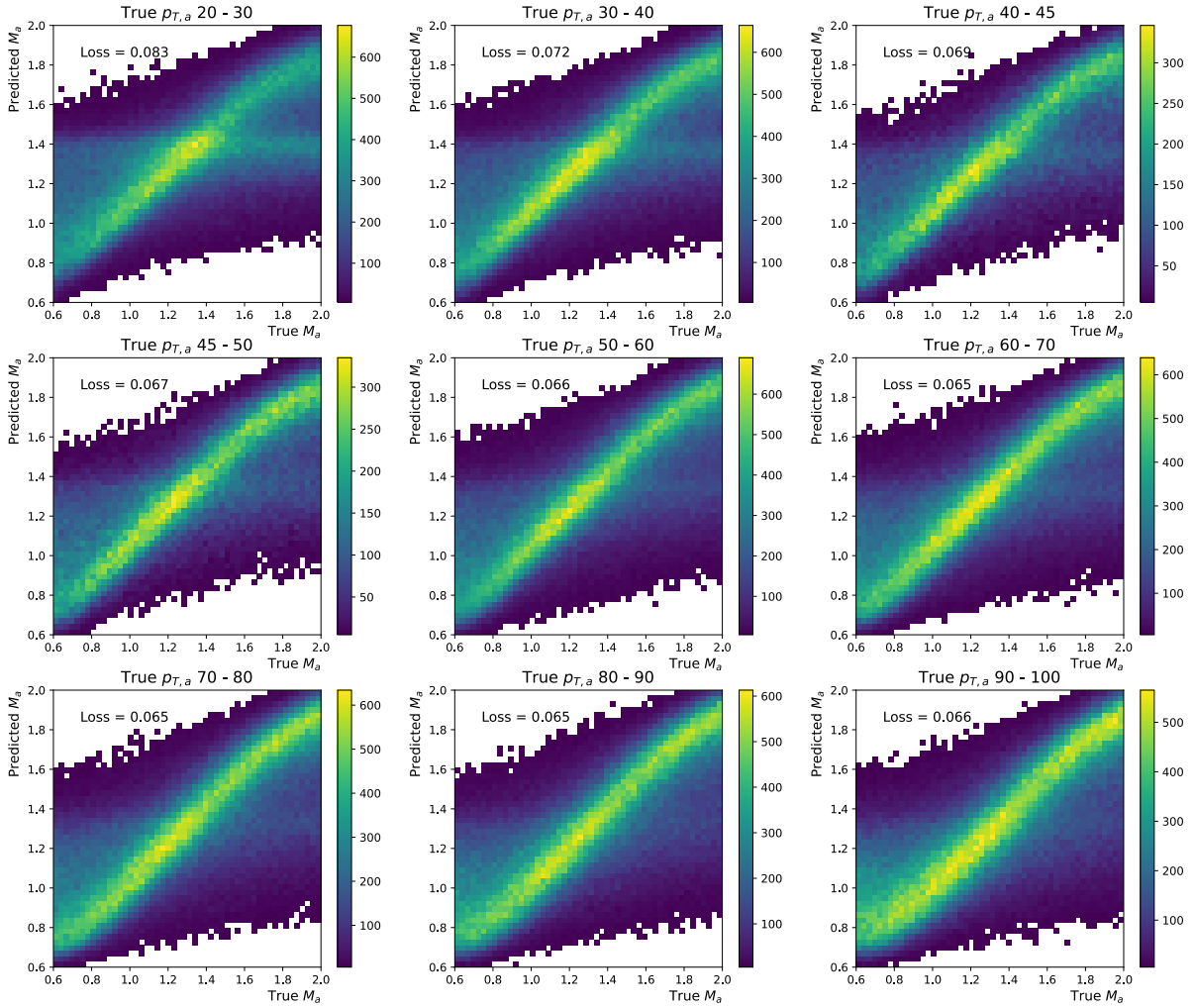


Figure 5.10: Predicted mass vs true mass in different bins of p_T of a using SuperClu+dR rechits

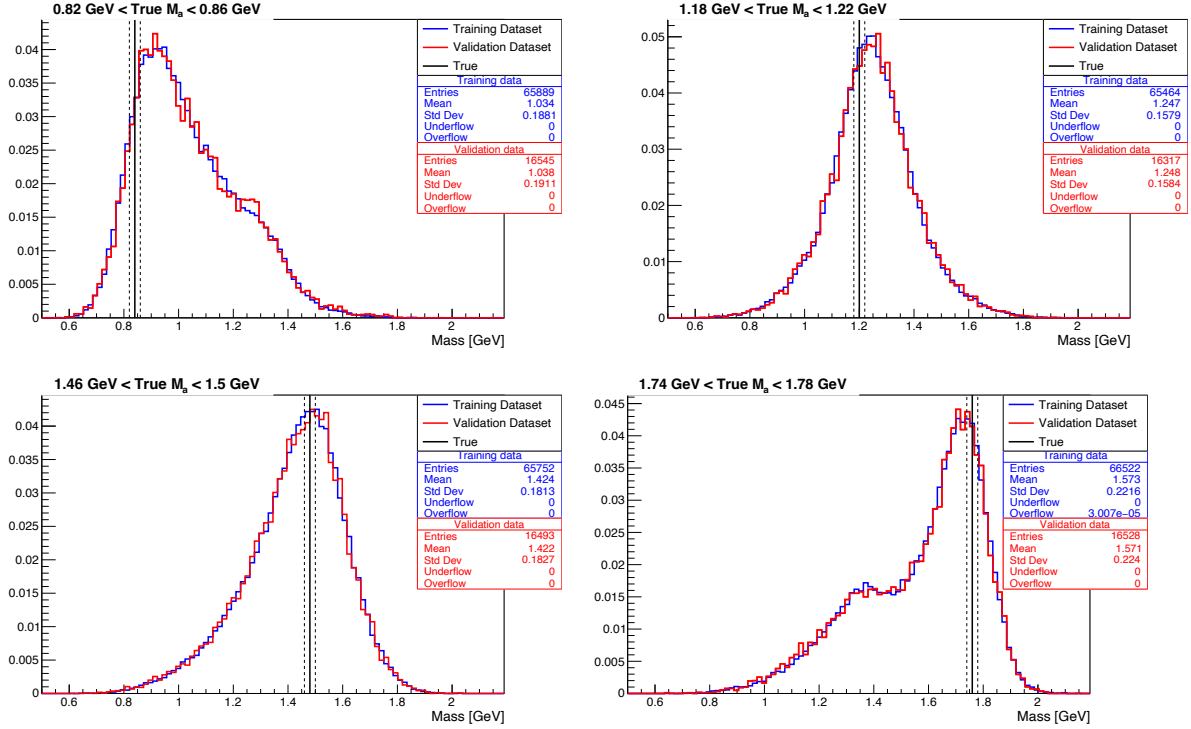


Figure 5.11: Predictions for subsamples from the training sample in different mass ranges of width 40 MeV using SupClu+dR rechits

5.3.3 Using SupClu+dR rechits using ResNet

We have validated our results using ResNet, a CNN-based model specifically designed to accommodate deep layers without significant overtraining. Here we have used the same loss function as DRN which is the MSE loss. More details of ResNet can be found here [20]. ResNet achieves this by learning the residuals between blocks, each comprising several layers. The effectiveness of ResNet architecture is demonstrated in Figure 5.12. It is to be noted that one can't directly compare DRN to ResNET since ResNET has about 15 million parameters as compared to less than half a million for DRN. Also being a CNN based network, ResNET can't be generalized into the endcaps due to the difficulties in giving complex geometries as inputs while the DRN can easily be extended to the endcaps.

The results of this training is slightly better than using DRN as the minimum validation loss is lower than that of DRN. Additionally, the flat line feature is suppressed which can also be seen in different bins of p_T in figure 5.13 displays the predictions versus true values across various p_T bins.

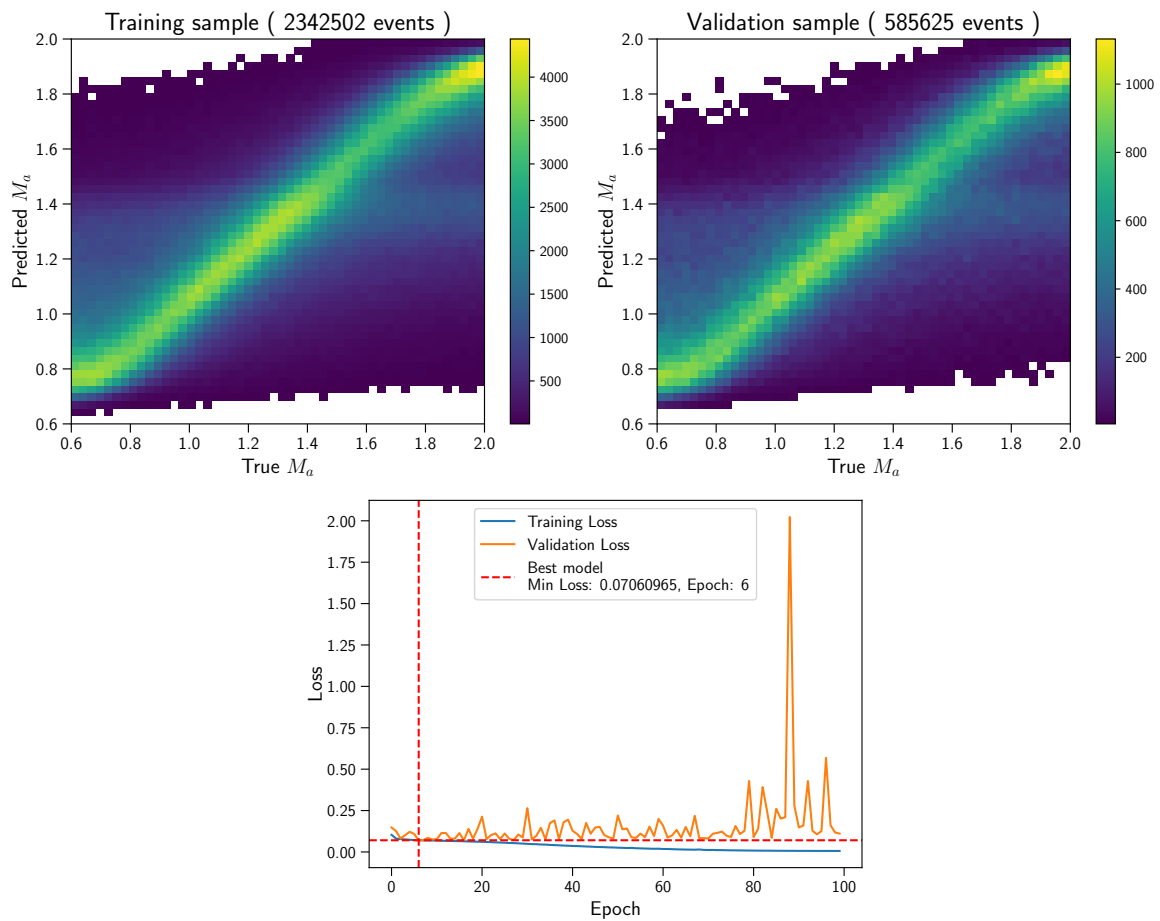


Figure 5.12: Predicted mass vs true mass of a for training dataset(top left) **using SupClu+dR rechits with ResNet**. Predicted mass vs true mass of a for validation dataset(top right). Loss vs epoch showing the model at the best epoch with lowest validation loss

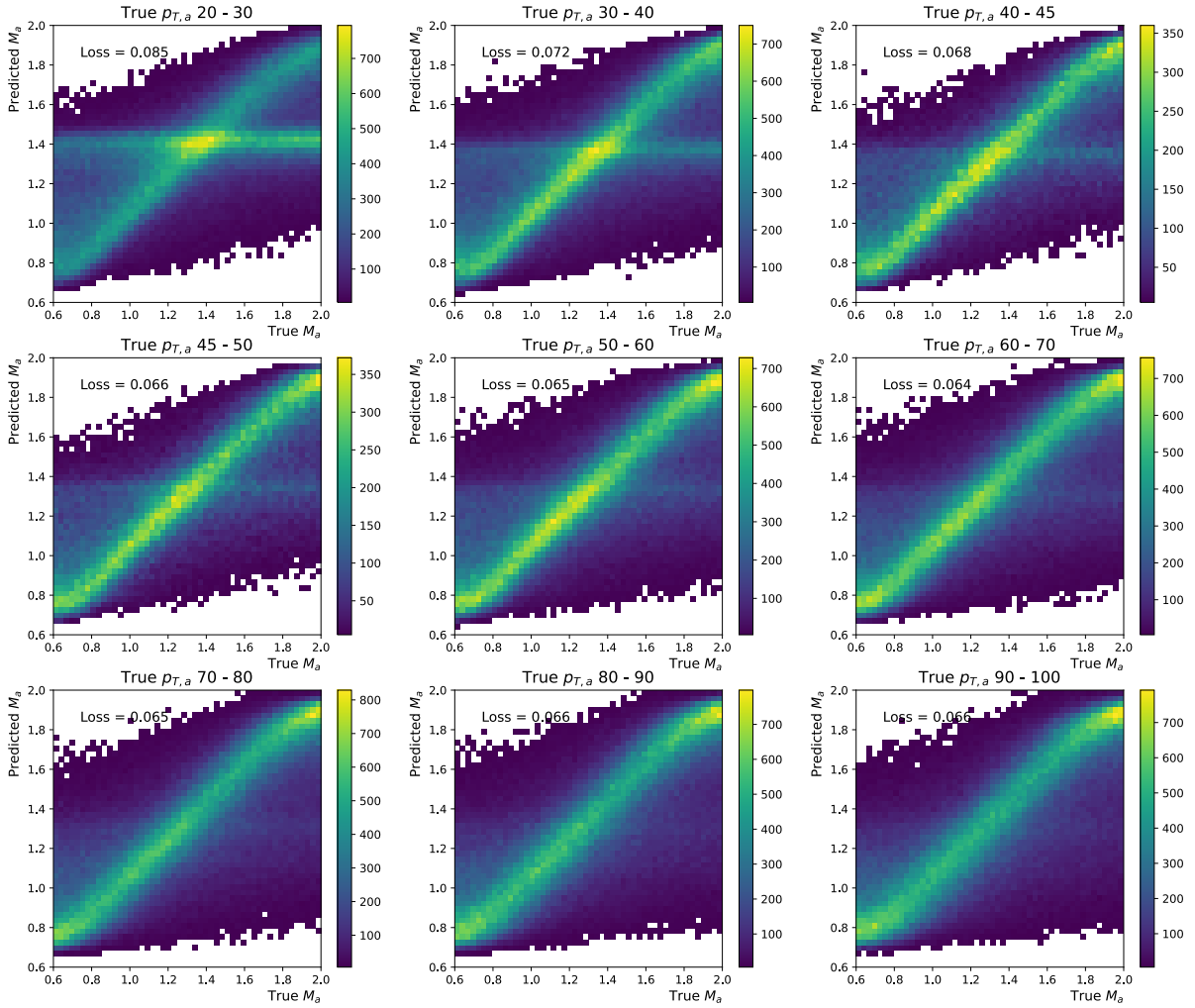


Figure 5.13: Predicted mass vs true mass in different bins of p_T of a using using SupClu+dR rechits with ResNet

Chapter 6

Particle flow hadron calibration

6.1 Hadron Calibration

The energy calibration of hadronic showers is essential given they are copiously produced in pp collisions at the LHC and make the key constituents of jets and MET. However, their inherent complex nature accompanied with non-linear response of detectors necessitate dedicated algorithms. In our study, we focus on calibrating pions, a crucial component in particle flow reconstruction within the CMS detector. Pions initiate showers primarily in the ECAL, with the majority of their energy deposited in the HCAL. To address this calibration challenge, we employ a simulated pion sample divided into two categories: those initiating showers in ECAL (**EH Hadrons**) and those not (**H Hadrons**). This differentiation is pivotal, as the response of the CMS ECAL and HCAL are very different for hadrons and electromagnetic components of hadron showers. This is the reasons that the calibration coefficients are derived for EH hadrons and H hadrons separately using the χ^2 minimization technique currently used in the event reconstruction [1].

Figure 6.1 illustrates the distributions of deposited energy in the ECAL and HCAL for these pions between 100 and 104 GeV in the barrel. The ECAL and HCAL energy scales are separately fixed using e γ and charged hadrons using dedicated methodologies, respectively. Energy reconstructed by summing up individual channels i.e. rechits is referred to as "raw" energy in the following. Top left and top right figures illustrate the energy deposited in ECAL (ECAL raw) and HCAL (HCAL raw) respectively. Bottom right plot illustrates the total energy deposited in ECAL and hcal. Bottom left plot is a 2D plot between ECAL raw (x axis) and HCAL raw (y axis) which illustrates the sharing of deposited energy in ECAL and HCAL. For H hadrons, the energy deposited in ECAL is zero. All other events where there is a non-zero amount of energy deposited in the ECAL can be identified as EH Hadrons.

To measure of the accuracy of the calorimeters in estimating the true energy of the particle, we define raw response of the calorimeters as:

$$response = \frac{E_{raw} - E_{true}}{E_{true}} \quad (6.1)$$

Where E_{raw} is the sum of raw energies deposited in ECAL and HCAL.

To obtain response as a function of true energy of pions, we divide our sample into different energy bins and η regions and fit a Gaussian function to the distribution of the of raw response obtained in each energy bin. This procedure is repeated for each η region. This is illustrated for energy bin of 100 GeV to 104 GeV for the barrel region ($|\eta| < 1.55$) in Fig 6.2. Thus we obtained the fitted parameters of the gaussian: mean (μ) and standard deviation (σ). The raw

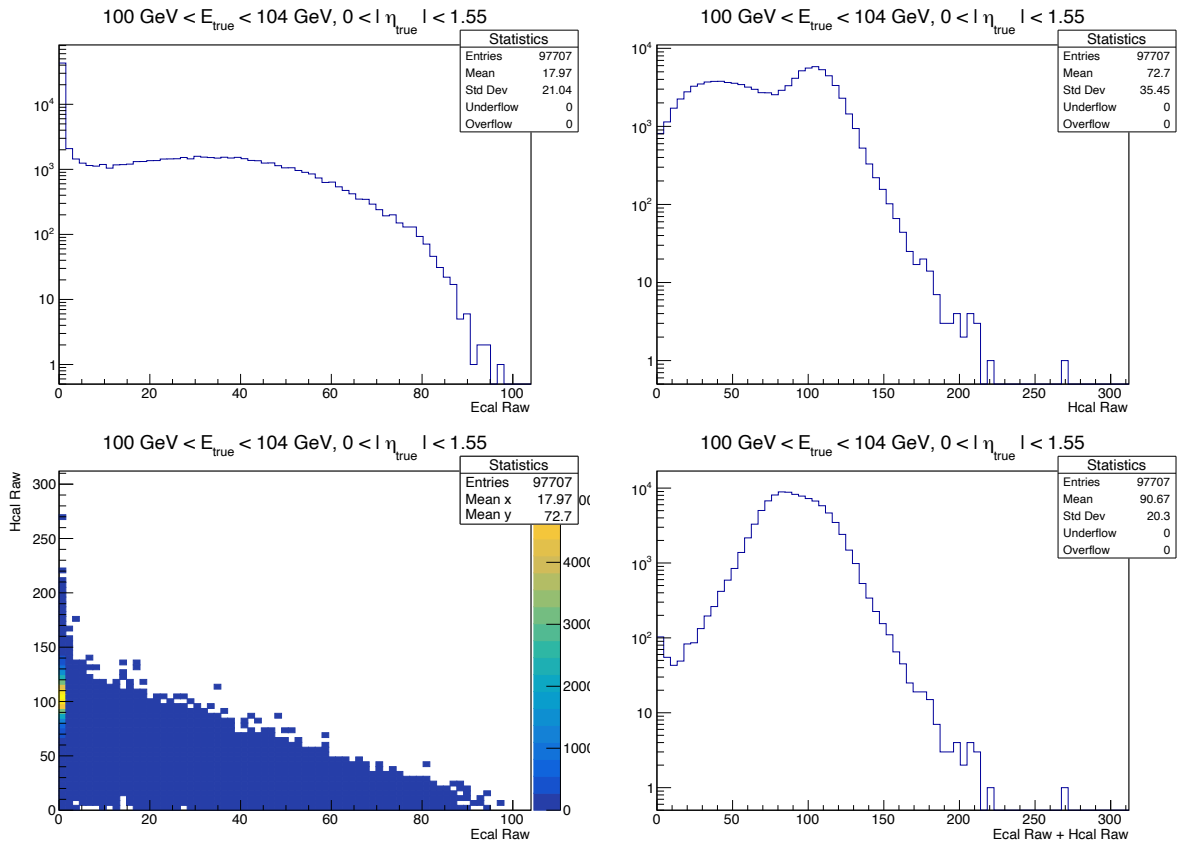


Figure 6.1: Ecal raw distribution (top left). HCAL raw distribution (top right). ECAL+HCAL raw distribution (bottom right). ECAL raw vs HCAL raw (bottom left). Plots are for pions between energy range of 100 GeV and 104 GeV and with $|\eta| < 1.55$

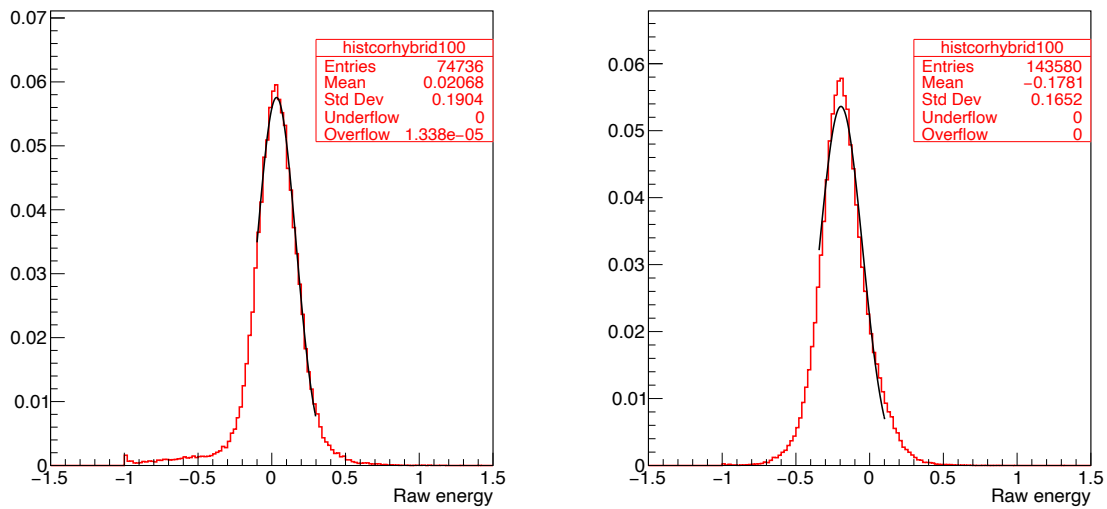


Figure 6.2: Raw response for H hadrons (left) and EH hadrons (right) in the energy bin of 100 GeV to 104 GeV for the barrel region ($|\eta| < 1.55$).

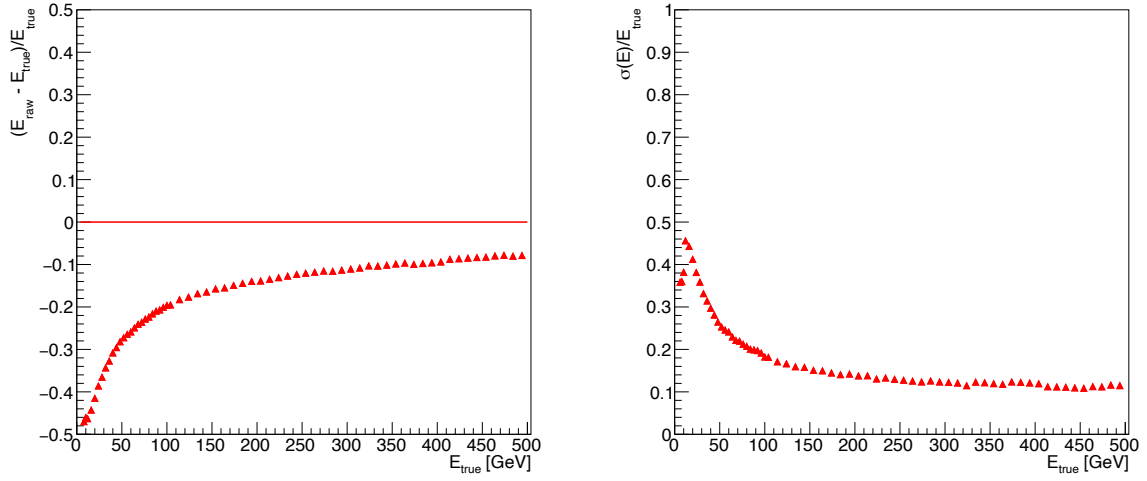


Figure 6.3: Response(left) and resolution(right) as a function of true energy of pion for **EH Hadrons**

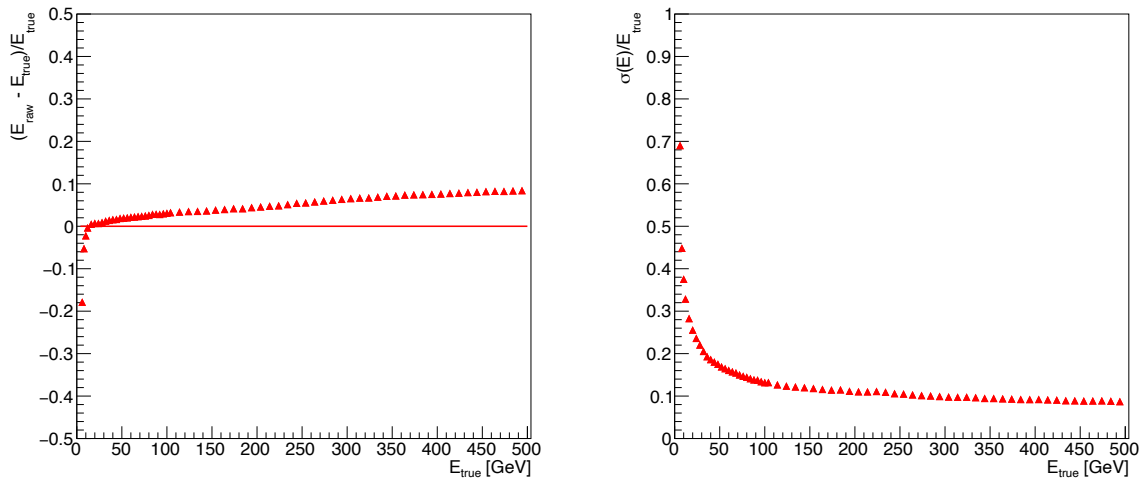


Figure 6.4: Response(left) and resolution(right) as a function of true energy of pion for **H Hadrons**

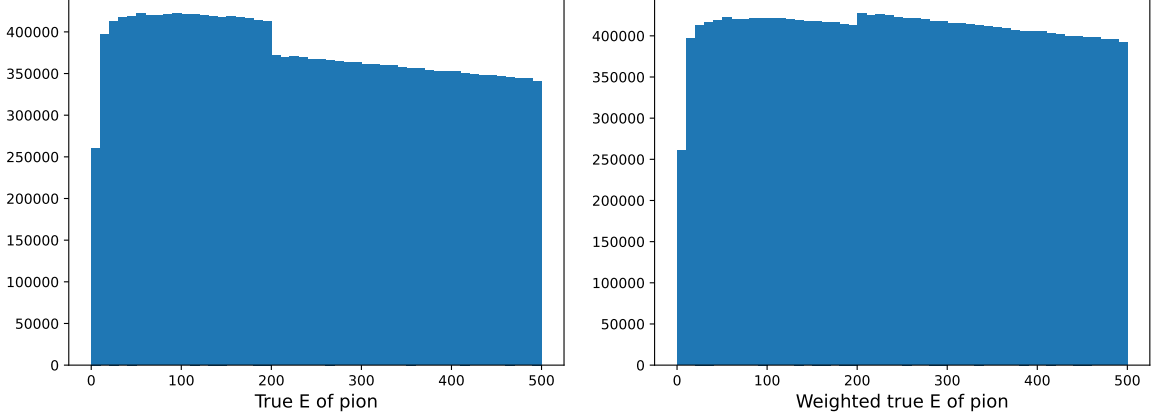


Figure 6.5: True energy of pions (left) and weighted energy of pions used as regression target(right). Pions with true energy less than 200 GeV are given a weight of 1 and pions with true energy greater than 200 GeV are given a weight of 1.15

response obtained for that energy bin would then be the mean of the fitted gaussian (μ). To measure the precision of this energy measurement, we define resolution as:

$$resolution = \frac{\sigma}{1 + \min(0, \mu)} \quad (6.2)$$

where $\min(a, b)$ denotes the minimum value between a and b .

Figure 6.3 and Figure 6.4 illustrate the raw response and resolution obtained in different energy bins as a function of the true energy for EH and H hadrons respectively in the barrel. In an idealistic scenario for perfectly calibrated calorimeters, the raw response, as defined above, should be zero for all energy points. However, we can see that the response is non-zero and non-linear with respect to true energy of pion for both H and EH hadrons as discussed above. The response and resolution for EH hadrons is worse than H hadrons for lower energies. In the following, we propose a new methodology based on GNN to reconstruct the energy of hadron showers and demonstrate much improved performance as compared to the χ^2 method.

6.2 Sample details

Sample used for calibration is centrally generated sample of 20 M charged pions with the following kinematic ranges:

- True momentum(p): $2 \text{ GeV} < p < 500 \text{ GeV}$
- True pseudorapidity(η): $|\eta| < 3$
- True azimuthal angle(ϕ): $|\phi| < \pi$

The following GEN-SIM-RAW datasets were used:

- /SinglePionGun_E0p2to200/Run3Winter23Digi-NoPUGTv4_126X_mcRun3_2023_forPU65_v4-v2/GEN-SIM-RAW (10M events)
- /SinglePionGun_E200to500/Run3Winter23Digi-NoPUGTv4_126X_mcRun3_2023_forPU65_v4-v2/GEN-SIM-RAW (10M events)

$ \eta $ range	Minimum number of hits
0.4 - 1.4	14
1.4 - 1.6	17
1.6 - 2.0	20
2.0 - 2.4	17
2.4 - 2.6	10

Table 6.1: η dependents selections for PF charged hadron

These samples were then privately reconstructed in *CMSSW_12_6_4* with global tag: *126X_mcRun3_2023_forPU65_v4*.

The following event selections were made to reject clusters originating from noise or fake hadrons. When the PF Candidate is a charged hadron:

- Transverse momentum of PF Candidate, $p_T \geq 1.0 \text{ GeV}$
- ECAL raw energy + HCAL raw energy $\geq 0.5 \text{ GeV}$
- Number of tracks associated with the PFBlock should be exactly 1
- Total momentum of the track, $p_{track} \geq 1.0 \text{ GeV}$ and its transverse momentum, $p_{T,track} \geq 1.0 \text{ GeV}$
- Number of pixel hits associated to the track should be greater than 2
- Number of total tracker hits for each η region are summarised in table [6.1](#)

If the PF candidate is a neutral hadron, then the spatial separation in the η - ϕ plane, between the PF candidate and true particle (dR) should be less than 0.4. If the PF candidate is anything other than this, the event is rejected.

Finally, our inputs to model are the rechits which were used to reconstruct our PF candidate. The following procedure is applied to get the relevant rechits:

- For each event we get the desired PF candidate
- We then get the corresponding elements in its PFBlock
- We then collect all PFElements of the PF candidate which are from either ECAL or HCAL
- For each of the PFElements, we get the corresponding cluster reference and associated *hitsAndFractions*.
- These *hitsAndFractions* are the rechits which were used by the PF algorithm to reconstruct our pion. These are referred to as **PF rechits**

To validate that these PF rechits capture the relevant information needed for the calibration, the total rechit energy of these rechits is compared with the raw energy of the reconstructed PF Candidate as well as the true energy. Since these rechits capture all the information about the PF candidate, the total rechit energy is expected to match with the raw energy which is what is observed in figure [6.6](#). Due to detector effects discussed above, total rechit energy is not expected to match exactly with the true energy of the pion, however, on average is expected to scale with the true energy.

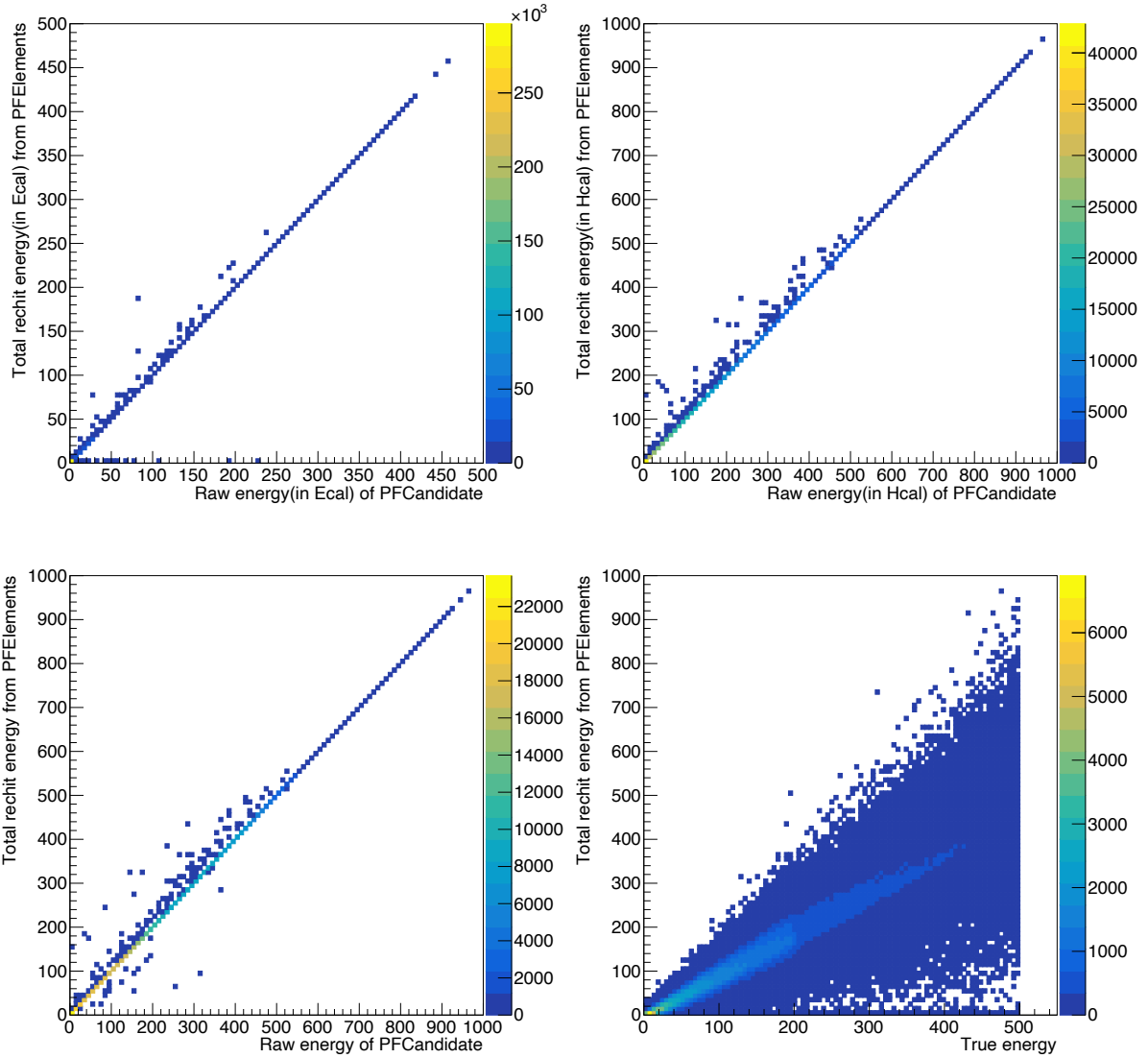


Figure 6.6: Total rechit energy using PF Rechits vs true energy in ECAL (top left), HCAL (top right), ECAL+HCAL (bottom left). Total rechit energy vs true(gen level) energy (bottom right)

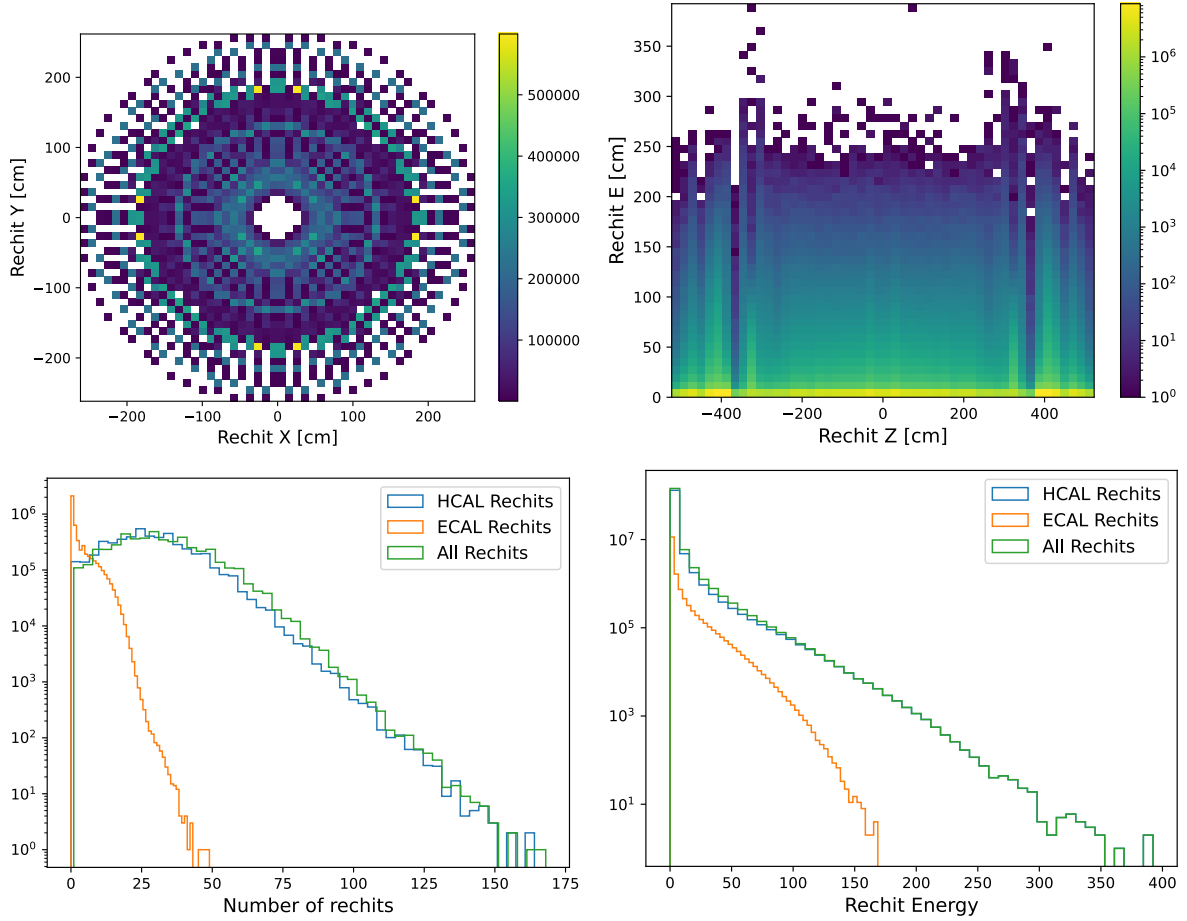


Figure 6.7: X vs Y coordinates of rechits(top left), Z coordinate vs rechit energy(top right). Energy of rechits distribution in ECAL and HCAL with multiple entries for each event(bottom left). Number of rechits per event in ECAL, HCAL and ECAL+HCAL

6.3 DRN model details

To get the corrected energy for pions, we train our model on PF Rechits from each event with true energy of the pion as the regression target. We use the cartesian coordinates of the rechits along with their energies. The various distribution of these rechits are illustrated in figure [6.7](#). Since our sample is not uniformly distributed in true energy of the pion, we use a weighted sample for training as illustrated in figure [6.5](#) which shows the unweighted distribution on the left and the weighted distribution on the right. Following are the model parameters used with the Dynamic Reduction Network (DRN) model architecture.

- **Inputs to the model:** Cartesian coordinates (X, Y, Z) of each rechit and their corresponding energies (E).
- **Input layers:** 3
- **Aggregation layers (nAgg) :** 2
- **Message passing layers:** 3
- **Output layers:** 2

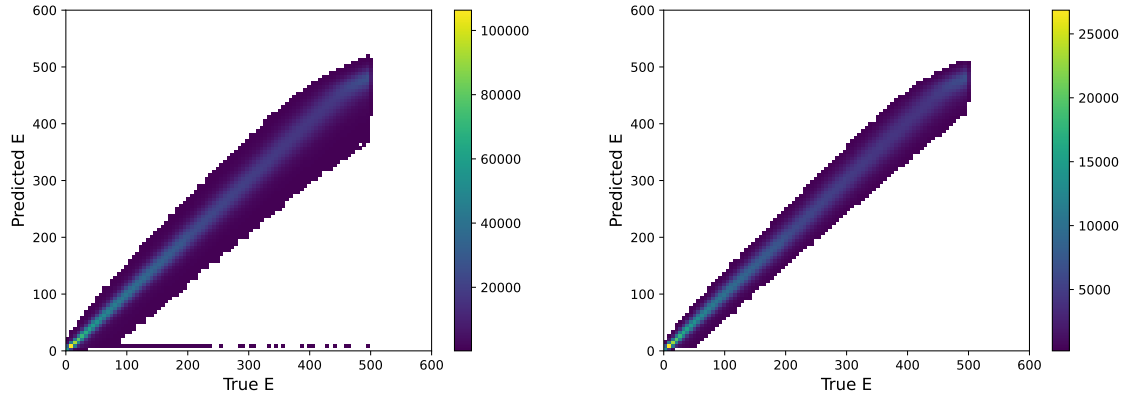


Figure 6.8: Predicted energy of pion vs true energy of pion for training dataset(top left) and validation dataset(top right)

- **Hidden Dimension:** 64
- **Batch size (Train and valid) :** 100
- **AdamW optimizer** with constant learning rate of 0.001
- **Model Parameters:** 62405
- **Total Events:** 790k training , 200k validation
- **Regression target:** True E of pion
- **Loss function:** $\frac{(E_{predicted}-E_{true})^2}{E_{true}}$

6.4 Results

The model was trained was 100 epochs for which 80% of our sample was used for training and remaining 20% for model validation after each epoch. The best model was selected which had the lowest validation loss and predictions were performed. Figure 6.8 shows the results of training. This shows that the model is able to predict energies correctly except for pions with energy near the boundary of the training sample range, which is due to edge effects as discussed below. The loss versus epoch for both the training and validation is illustrated in Figure 6.9. The loss decreases smoothly with each epoch and saturates after around 60 epochs. There is no evidence of overtraining as the training and validation losses do not diverge. As done for raw responses, the predicted energy is divided into different bins of E_{true} and response and resolution are obtained by fitting the predicted distributions with a Gaussian function. The responses are shown in figure 6.10 for EH hadrons and in figure 6.11 for H hadrons. Comparisons are also made with the response obtained using χ^2 method. The 1D predictions by DRN are narrower than that using χ^2 method which shows that there is an improvement in each energy bin. Response and resolution obtained as a function of energy of pion as well as the improvements with respect to χ^2 method is illustrated in figure 6.12 for EH hadrons and 6.13 for H hadrons. An improvement of around 60 percent is observed in the resolution using the DRN method.

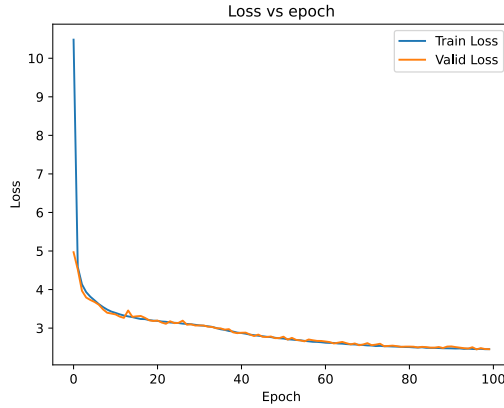


Figure 6.9: Loss vs epoch during training

A characteristic feature of machine learning models is the development of edge effects which is the region beyond the dotted red line on these plots. Edge effects refer to the phenomenon where the performance of a model deteriorates near the boundaries of the dataset, often resulting in inaccuracies or anomalies in predictions. In our case, these start developing for pions with energy greater than around 400 GeV. Therefore our model can accurately calibrate pions of energy upto this energy range which is the case in most particle jets which have a high abundance of pions.

In figure 6.14, the response as a function of η is obtained using the DRN method is plotted and compared with the χ^2 method. A great amount of improvement is observed especially in the barrel to endcap transition region in the η range of 1.2 to 1.5 . We observe a significant improvement in the resolution obtained using DRN as compared to the conventionally using χ^2 method by upto 30 percent for EH hadrons and around 2 percent for H hadrons. The DRN is able to learn the event-by-event fluctuations in the hadron showers, especially in the EH hadrons where ECAL also plays a major role. This has also been observed in HGICAL studies performed by Alpana et. al. as presented in [1].

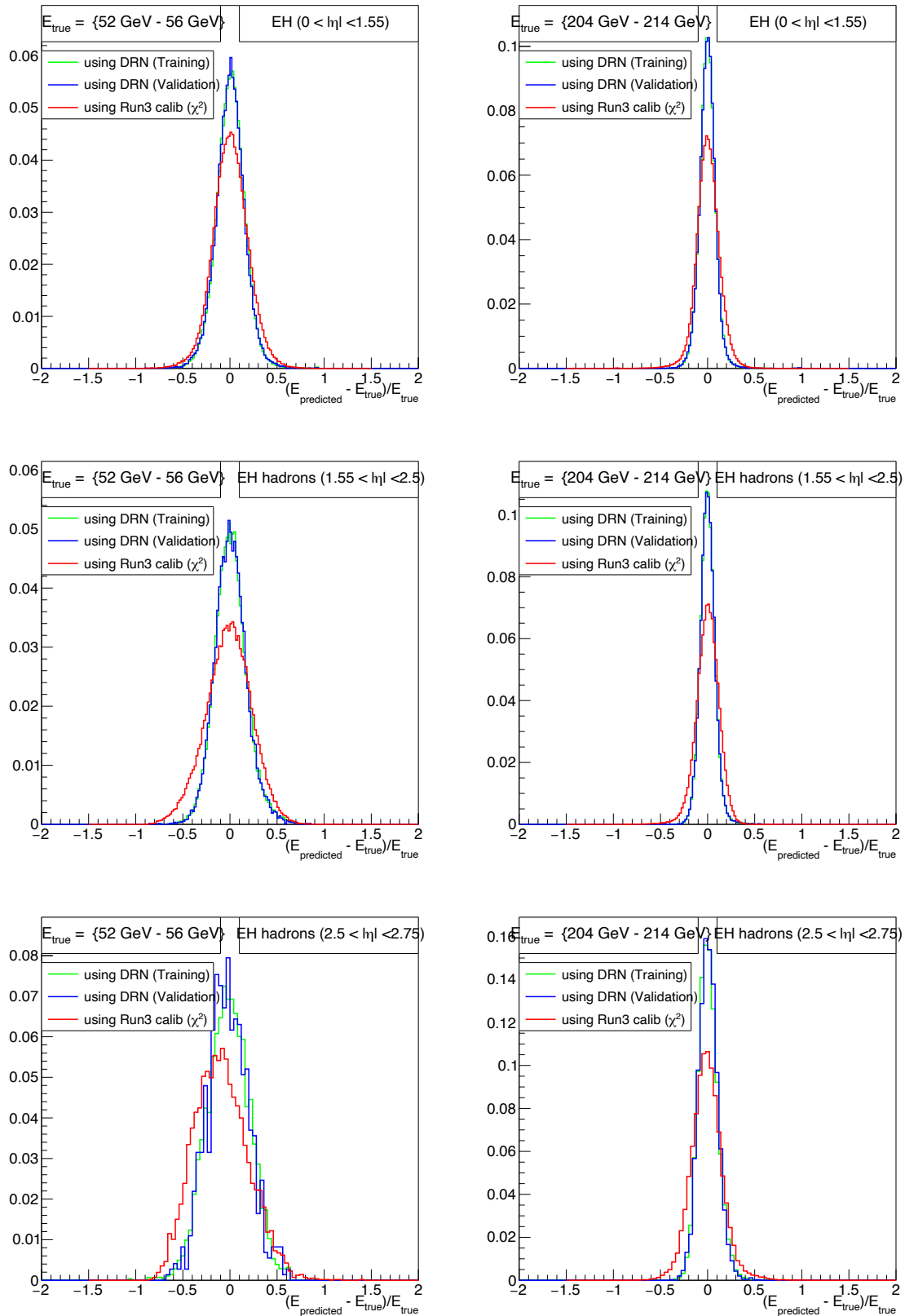


Figure 6.10: Distribution of 1D responses obtained from prediction from model in training and validation datasets, along with comparison with χ^2 method for **EH Hadrons**. Top plots are for responses in barrel for pions of energy 52 GeV to 56 GeV (left) and 204 to 214 GeV (right). Responses for same energy ranges are shown in the middle and bottom plots for endcap within tracker and endcap outside tracker regions respectively.

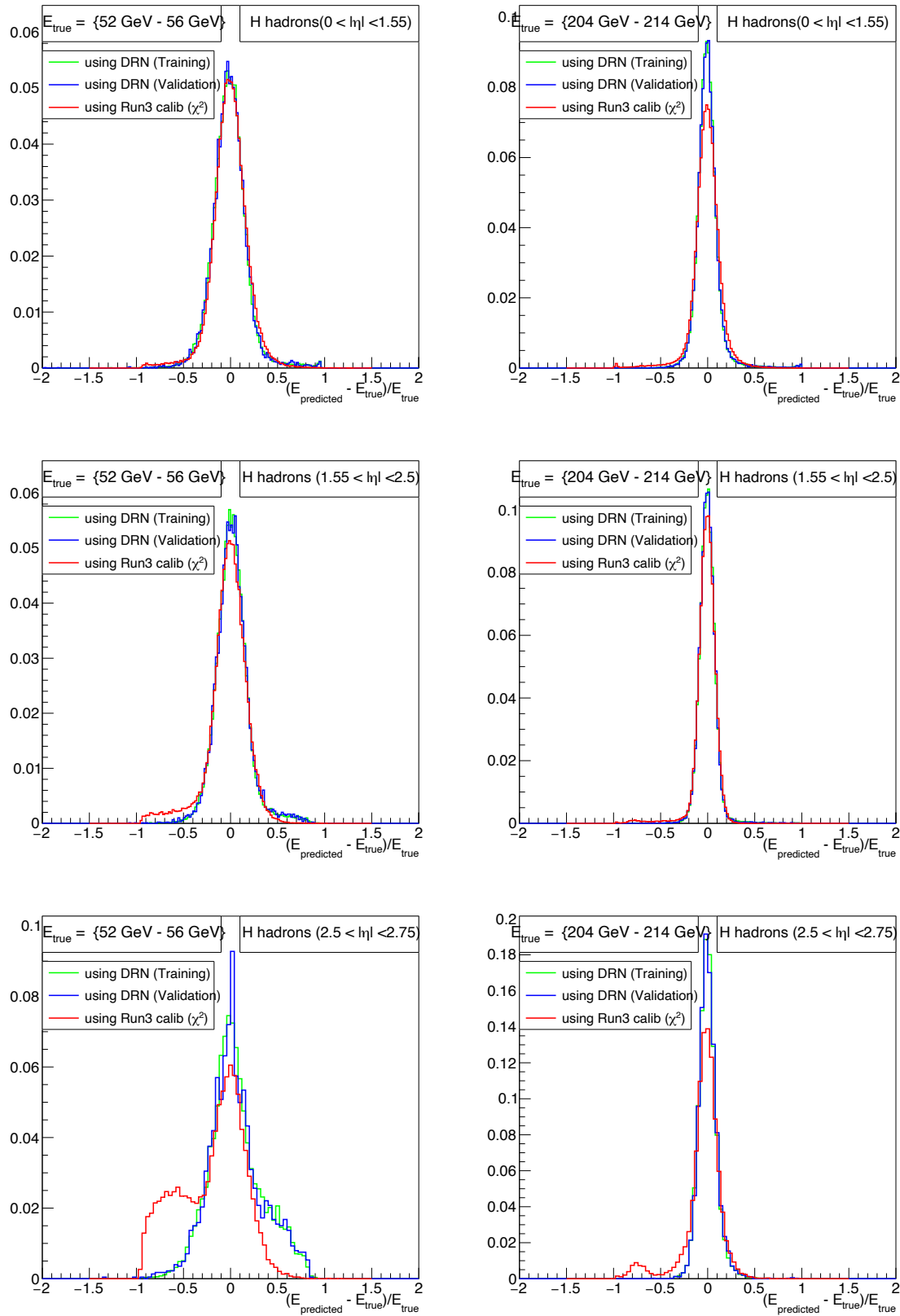


Figure 6.11: Distribution of 1D responses obtained from prediction from model in training and validation datasets, along with comparison with χ^2 method for **H Hadrons**. Top plots are for responses in barrel for pions of energy between 52 GeV to 56 GeV (left) and 204 to 214 GeV (right). Responses for same energy ranges are shown in the middle and bottom plots for endcap within tracker and endcap outside tracker regions respectively.

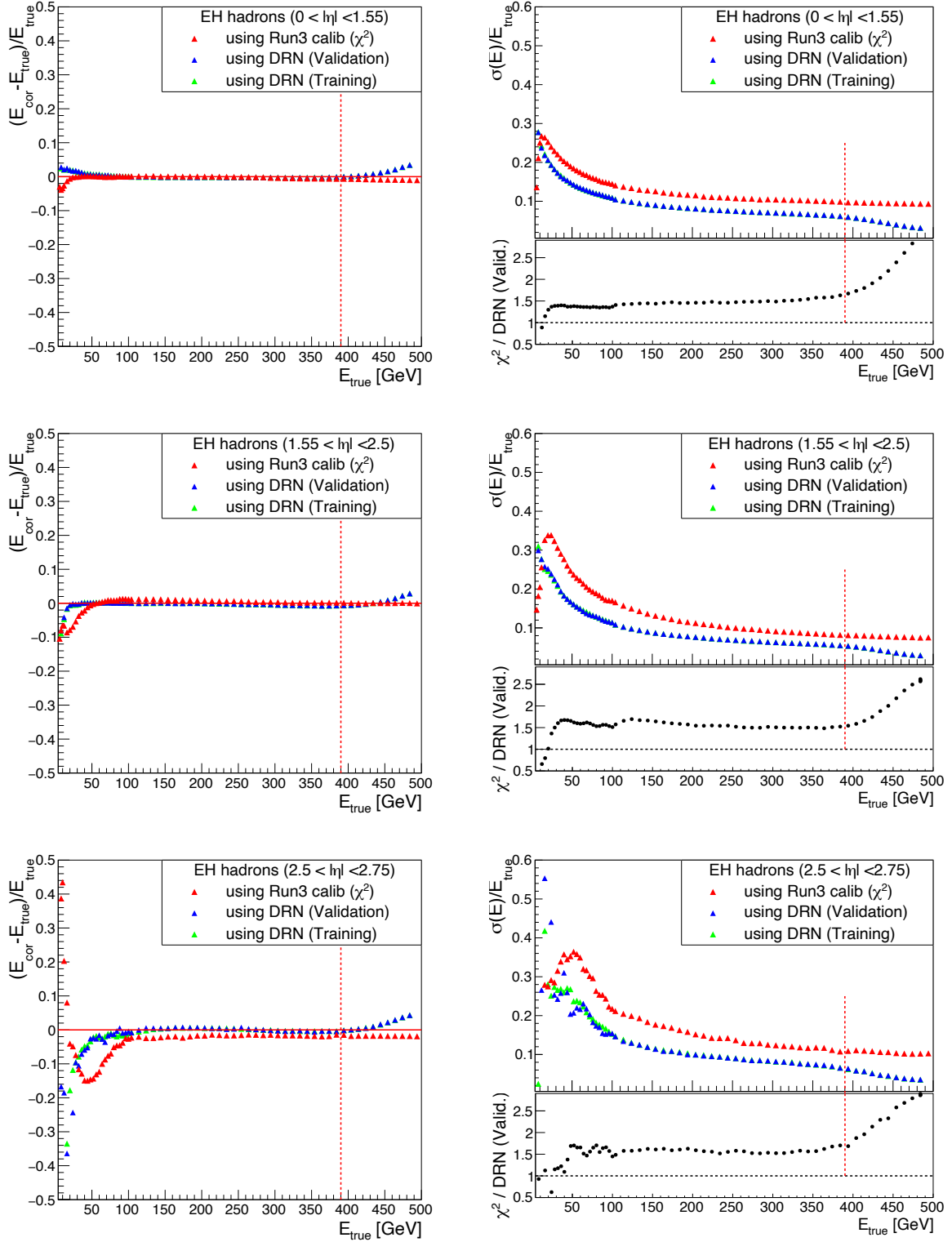


Figure 6.12: DRN training, DRN validation and χ^2 calibration results for **EH Hadrons**. Response vs energy is shown on left for different η regions: barrel, endcap within tracker region and endcap outside tracker region respectively. The plots on the right illustrate the resolution vs energy on the top panel and the improvement with respect to χ^2 is shown on the bottom panel. This is shown for the 3 different η regions. Edge effects occur beyond the dotted red line.

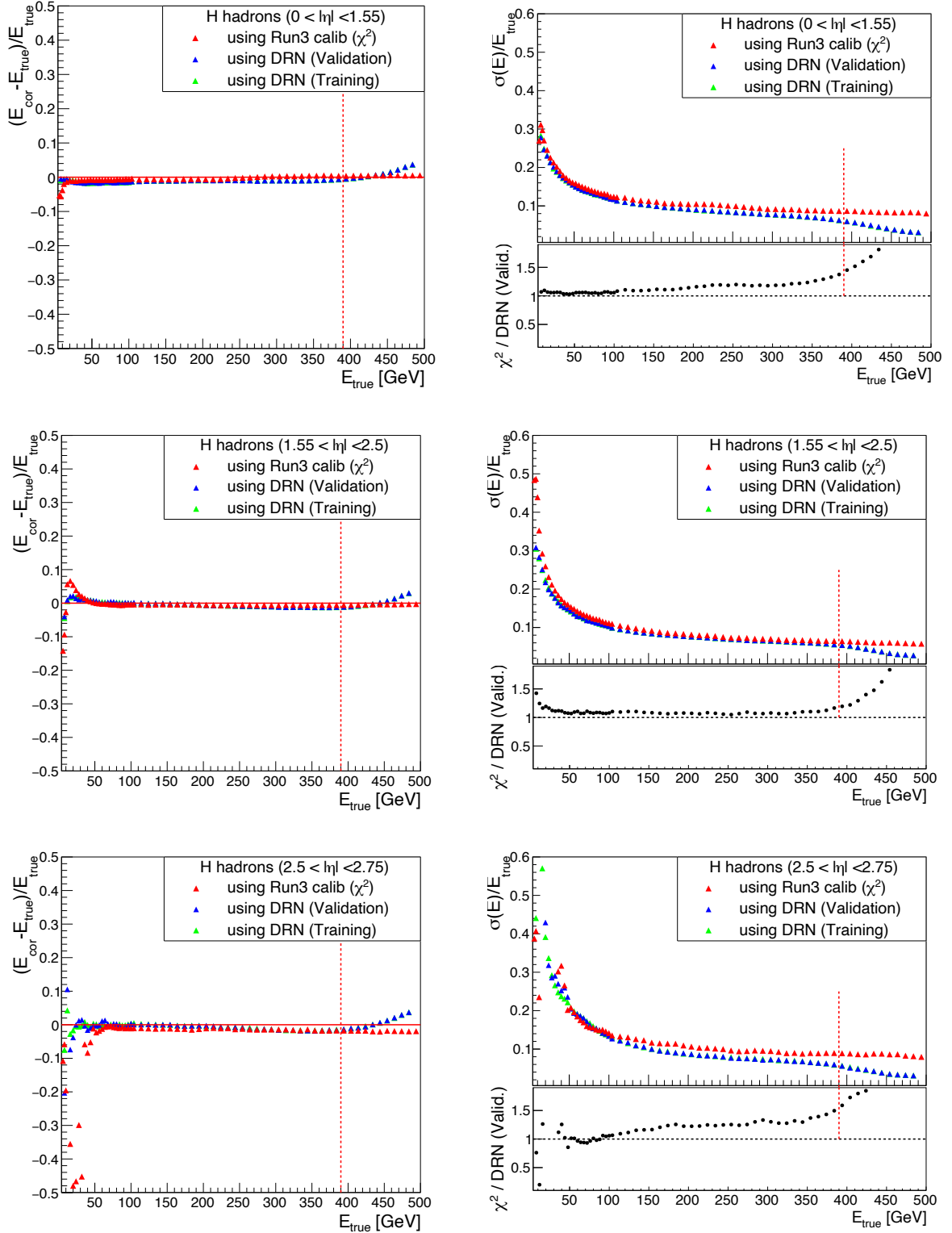


Figure 6.13: DRN training, DRN validation and χ^2 calibration results for **H Hadrons**. Response vs energy is shown on left for different η regions: barrel, endcap within tracker region and endcap outside tracker region respectively. The plots on the right illustrate the resolution vs energy on the top panel and the improvement with respect to χ^2 is shown on the bottom panel. This is shown for the 3 different η regions. Edge effects occur beyond the dotted red line.

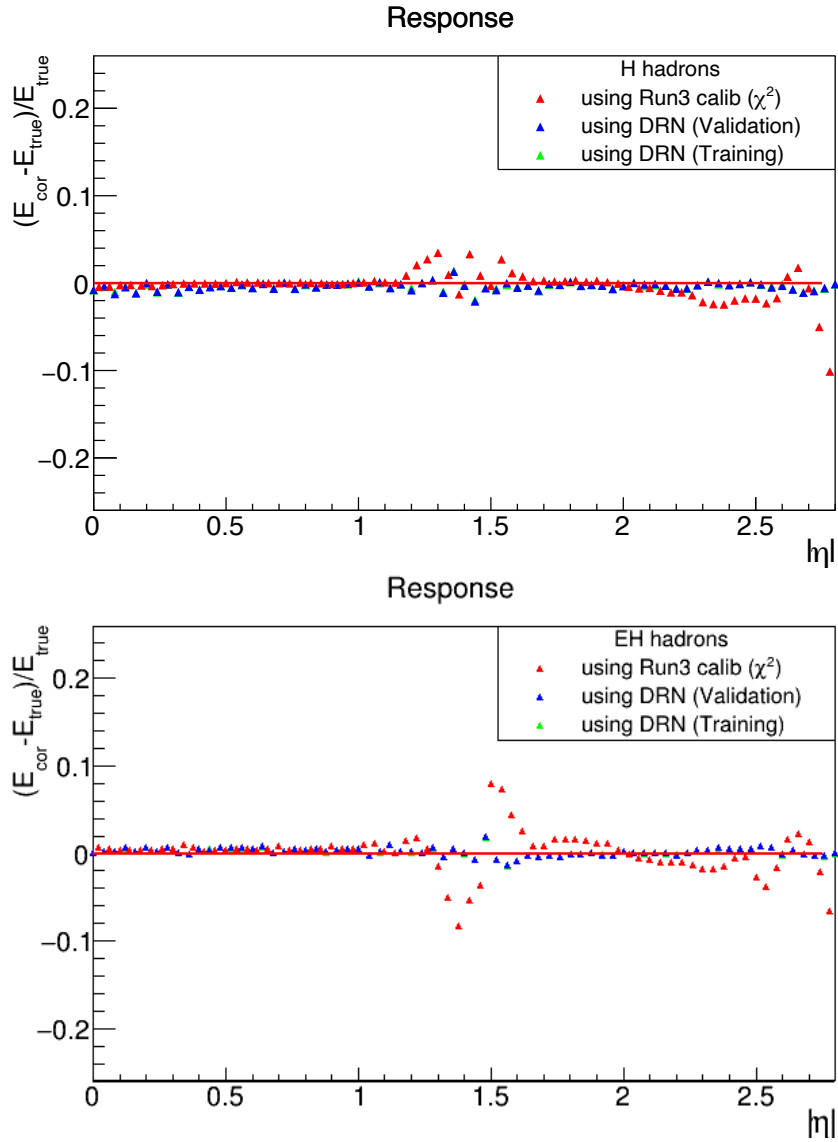


Figure 6.14: Response obtained using DRN and χ^2 method as a function of η for H hadrons(left) and EH hadrons(right)

Chapter 7

Summary

In this thesis, we have developed an end-to-end framework for reconstructing the mass of boosted particles decaying to a pair of photons which are not identified as individual photon objects in the CMS detector. This reconstruction can not only help in the reconstruction of the pseudoscalar a but also will help in reconstruction of particles of low masses which were previously inaccessible owing to the limited granularity of the ECAL. Since graph neural networks can work with irregular geometries, this framework will be even more effective in cases of a finer granularity like the upgrade of the endcaps with the high granularity calorimeter. It can also find out hidden peaks in the resonance of the SM like Higgs decaying to two photons. If the decay of Higgs to the pseudoscalar a occurs, the two pairs of merged photons may have been misidentified as just two photons.

To improve the efficiency of our method, additional tracking information may be given to the model to give additional information for converted photons. Since the information available about the decay is limited owing to the few number of rechits and no clear separation between the clusters of the two photons, any additional information can greatly improve the model. The next steps include validating our model with real data from the detector which involves estimating the invariant mass of other known particles which decay to a pair of photons like neutral pions ($m_\pi = 140\text{MeV}$) and the η meson ($m_\eta = 550\text{MeV}$).

In addition to the above, we have also explored the nuances of the hadronic showers and calibrated hadrons where we have used optimised rechit information from the PF algorithm to be fed into our machine learning framework. Since machine learning frameworks have the ability to capture key patterns in data, it is very effective in a tasks such as calibration where it is able to provide software compensation for event to event fluctuations in the deposition of energy by hadrons.

References

- [1] CERN document server. Online.
- [2] GEANT4 website. <https://geant4.web.cern.ch/>.
- [3] Introduction to Artificial Neural Networks and the Perceptron. <https://www.quantstart.com/articles/introduction-to-artificial-neural-networks-and-the-perceptron/>.
- [4] Scintillation Counter. https://en.wikipedia.org/wiki/Scintillation_counter.
- [5] Lifetime Measurement of Pion. *Physical Review Letters*, 1985.
- [6] GRAclus: GRAph CLUStering and Partitioning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.
- [7] Performance of the ECAL in High Energy Physics Experiments. *Journal of Instrumentation*, 2014.
- [8] Particle Flow Algorithm in High Energy Physics. *European Physical Journal C*, 2017.
- [9] Dynamic Graph Convolutional Neural Networks for Graph Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.
- [10] Analysis of Particle Collisions at CFNS School. *Proceedings of CFNS School*, 2019.
- [11] Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC. *JINST*, 16(05):P05014, 2021.
- [12] Development of the HCAL for Particle Detection. *Physical Review D*, 2023.
- [13] CERN. CERN Document Server. <https://cms.cern/book/export/html/1618>.
- [14] CMS Collaboration. CMS Publications Repository. *CERN Open Data*.
- [15] CMS Collaboration. Anatomy of a Jet in CMS. *Quantum Diaries*, 2011.
- [16] CMS Collaboration. Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV. *Journal of Instrumentation*, 12:P02014, 2017.
- [17] Wikipedia contributors. Standard model. https://en.wikipedia.org/wiki/Standard_Model.
- [18] David J. Griffiths. *Introduction to Elementary Particles*. Wiley, 1998.

- [19] John F. Gunion, Howard E. Haber, Gordon L. Kane, and Sally Dawson. *The Higgs Hunter's Guide*, volume 80. 2000.
- [20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [21] William R. Leo. *Techniques for Nuclear and Particle Physics Experiments: A How-to Approach*. Springer Berlin, Heidelberg, 1994.
- [22] Fabio Sauli. *Interaction of photons and neutrons with matter*, page 43–75. Cambridge Monographs on Particle Physics, Nuclear Physics and Cosmology. Cambridge University Press, 2014.
- [23] The JINST Collaboration. JINST. *Journal of Instrumentation*, Volume 3:08004, 2008.
- [24] Tejinder S Virdee. *Experimental techniques*. 1999.

Appendix A

Angle between daughter particles in $a \rightarrow \gamma\gamma$ decay process

In this section, we derive a formula for the angle between two photons in $a \rightarrow \gamma\gamma$ decay in terms of the kinematic quantities of the mother particle a .

A.1 Notation

p denotes four momentum, P denotes the 3D spatial momentum, E denotes energy, m denotes rest mass.

a refers to the mother particle, γ_1 refers to the first photon, γ_2 refers to second photon.

θ denotes the angle between the two photons The calculation is done in natural units and relation using the following Minkowski metrix is used:

$$\eta_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad (\text{A.1})$$

Refer to [\[18\]](#) for details.

A.2 Calculation

Using four momentum conservation

$$p_a = p_{\gamma_1} + p_{\gamma_2} \quad (\text{A.2})$$

Multiplying both sides by p_{γ_1}

$$p_a p_{\gamma_1} = p_{\gamma_1} p_{\gamma_1} + p_{\gamma_2} p_{\gamma_1}$$

Using $m_{\gamma_1} = 0$ and thus $|P_{\gamma_1}| = E_{\gamma_1}$,

$$E_a E_{\gamma_1} - |P_a| E_{\gamma_1} \cos(\theta_1) = 0 + E_{\gamma_1} E_{\gamma_2} (1 - \cos(\theta))$$

Dividing by E_{γ_1}

$$E_a - P_a \cos(\theta_1) = E_{\gamma_2} (1 - \cos(\theta))$$

Thus,

$$\cos(\theta_1) = \frac{E_a - E_{\gamma 2}(1 - \cos(\theta))}{P_a} \quad (\text{A.3})$$

Similarly,

$$\cos(\theta_2) = \frac{E_a - E_{\gamma 1}(1 - \cos(\theta))}{P_a} \quad (\text{A.4})$$

Also multiplying [A.2](#) with p_a

$$\begin{aligned} p_a p_a &= p_{\gamma 1} p_a + p_{\gamma 2} p_a \\ m_a^2 &= E_a E_{\gamma 1} - E_{\gamma 1} P_a \cos(\theta_1) + E_a E_{\gamma 2} - E_{\gamma 2} P_a \cos(\theta_2) \\ m_a^2 &= E_a(E_{\gamma 1} + E_{\gamma 2}) - E_{\gamma 1}(E_a - E_{\gamma 2}(1 - \cos(\theta))) - E_{\gamma 2}(E_a - E_{\gamma 1}(1 - \cos(\theta))) \\ m_a^2 &= 2E_{\gamma 1} E_{\gamma 2}(1 - \cos(\theta)) \\ \cos(\theta) &= 1 - \frac{m_a^2}{2E_{\gamma 1} E_{\gamma 2}} \end{aligned} \quad (\text{A.5})$$

To get a lower bound on the angle:

$$(E_{\gamma 1} - E_{\gamma 2})^2 \geq 0$$

Rearranging terms,

$$((E_{\gamma 1} + E_{\gamma 2})/2)^2 \geq E_{\gamma 1} E_{\gamma 2}$$

Using $E_{\gamma 1} + E_{\gamma 2} = E_a$ and equation [A.5](#) we get:

$$\cos(\theta) \leq 1 - \frac{2m_a^2}{E_a^2} \quad (\text{A.6})$$

Since $\cos(\theta) \geq 1 - \theta^2/2$

$$1 - \frac{\theta^2}{2} \leq 1 - \frac{2m_a^2}{E_a^2}$$

Therefore

$$\frac{2m_a}{E_a} \leq \theta \quad (\text{A.7})$$

Or in terms of boost of a (γ_a)

$$\frac{2}{\gamma_a} \leq \theta \quad (\text{A.8})$$

This shows that the lower bound of the angle between the two photons is inversely proportional to the boost of a . Furthermore it is the boost of the mother particle(a) alone that dictates this lower bound.

Appendix B

Comparisons of inputs to χ^2 method and DRN method

The inputs to the DRN method and the χ^2 method were slightly different due to differences in the samples used. Given below are the raw responses and resolution as a function of energy of pion. As done above, pion samples were divided into energy bins, fitted with a gaussian in each energy bin and then response and resolution for each energy bin were calculated using the fitted parameters. These plots are shown for barrel region (η range of 0 to 1.55) in figure [B.1](#), endcap inside tracker region (η range of 1.55 to 2.75) in figure [B.2](#) and endcap outside tracker region (η range of 2.75 to 3) in figure [B.3](#).

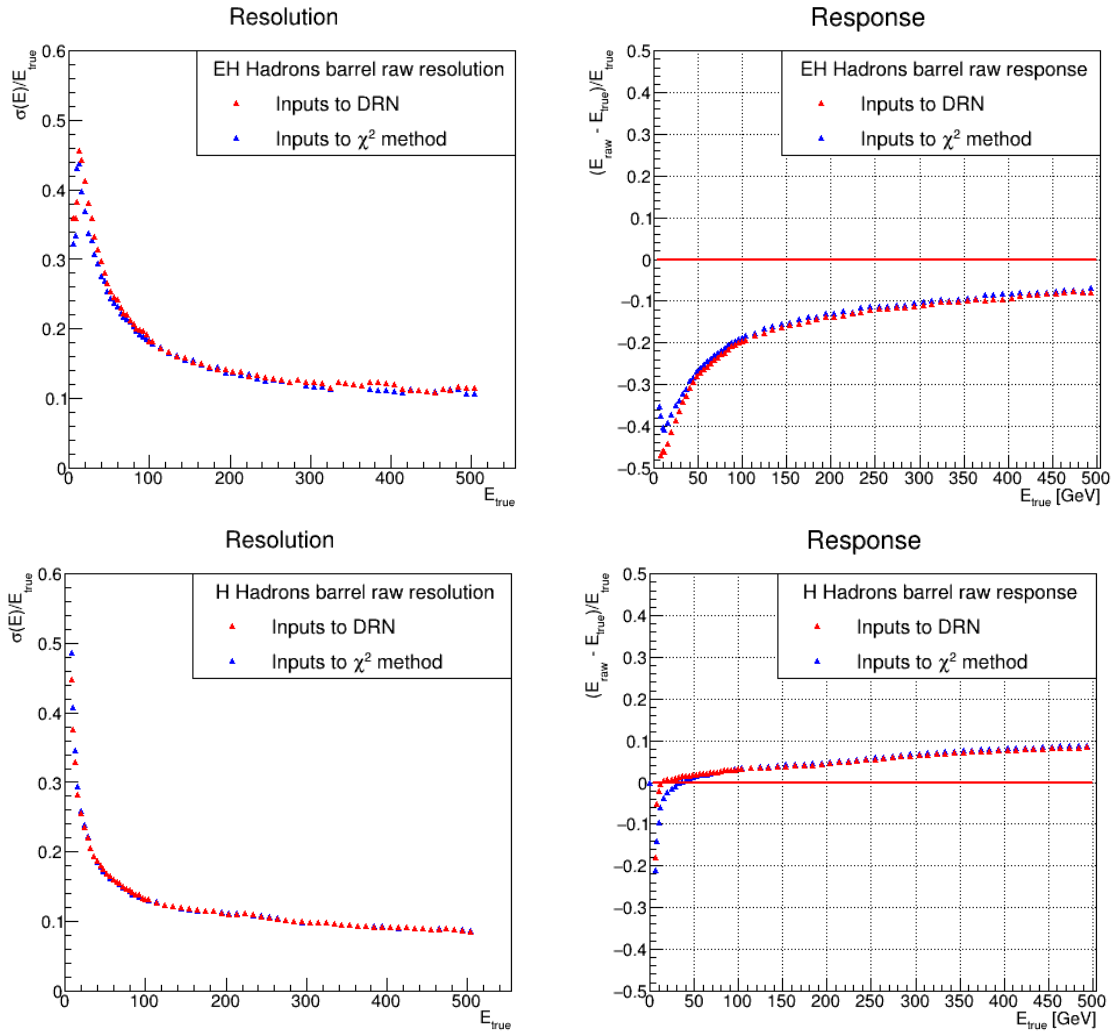


Figure B.1: Comparison of the raw responses and resolution between the inputs used by the two methods: DRN (In red) and χ^2 method (In blue). Raw resolution vs true energy (left), Raw response vs true energy (right) for EH Hadrons (top) and H Hadrons (bottom) in barrel.

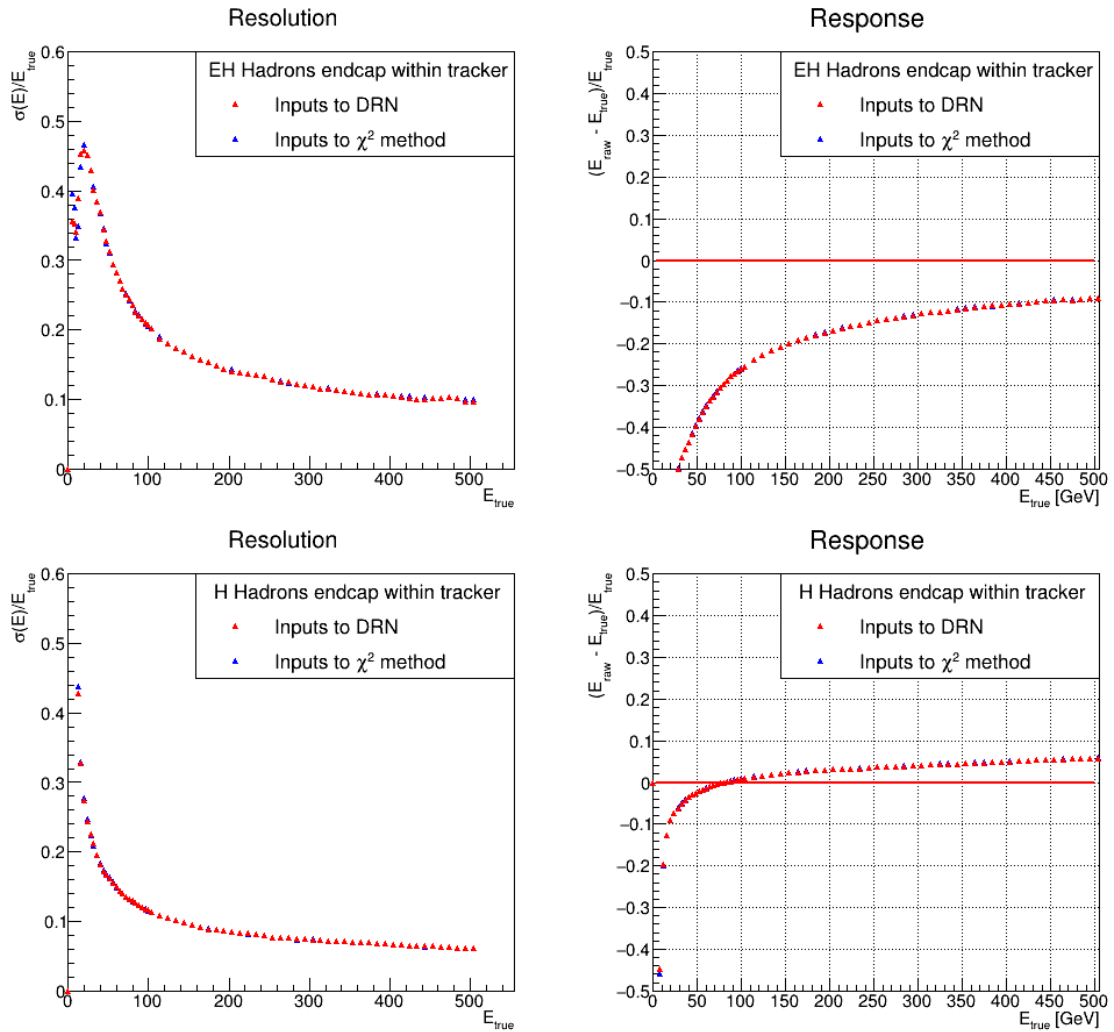


Figure B.2: Comparison of the raw responses and resolution between the inputs used by the two methods: DRN (In red) and χ^2 method (In blue). Raw resolution vs true energy (left), Raw response vs true energy (right) for EH Hadrons (top) and H Hadrons (bottom) in endcap inside tracker region.

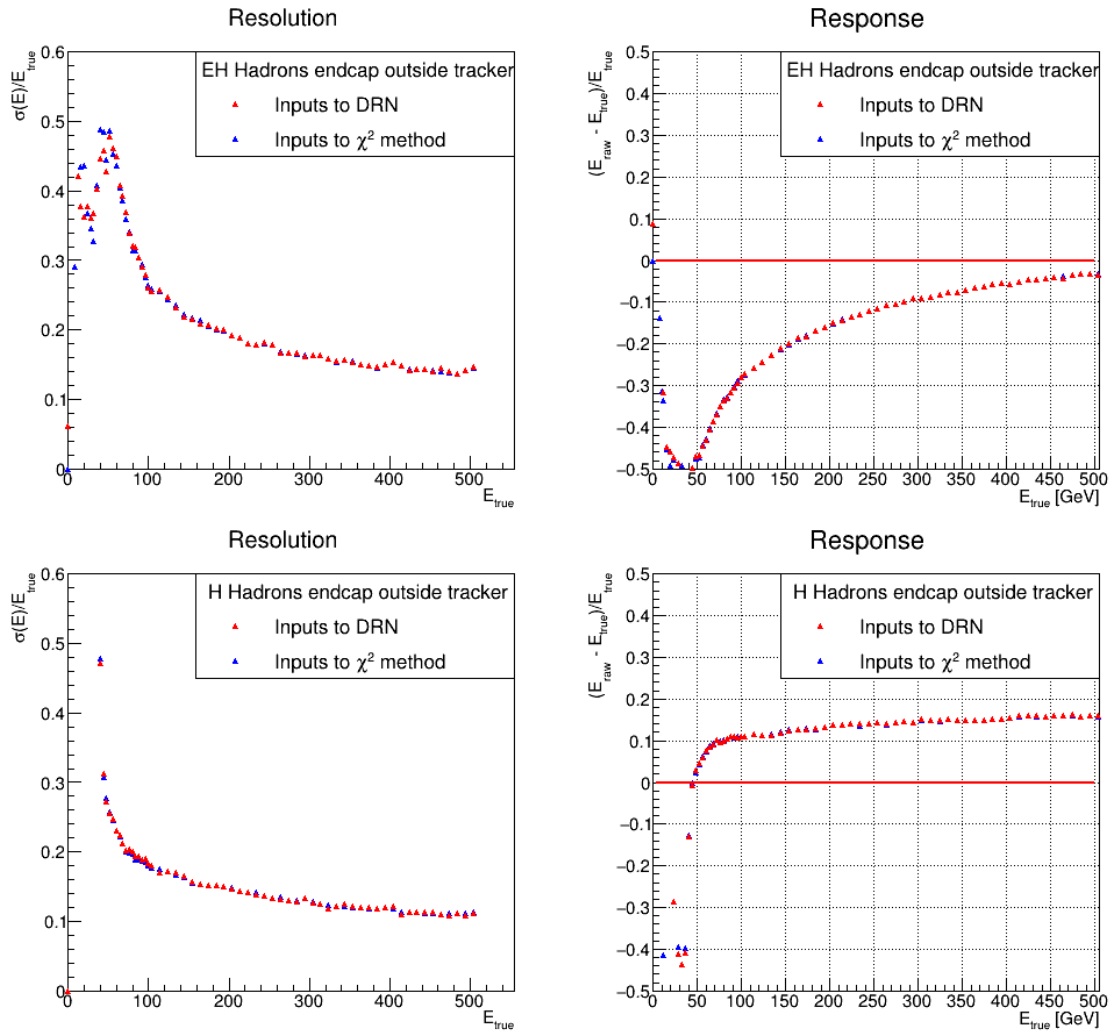


Figure B.3: Comparison of the raw responses and resolution between the inputs used by the two methods: DRN (In red) and χ^2 method (In blue). Raw resolution vs true energy (left), Raw response vs true energy (right) for EH Hadrons (top) and H Hadrons (bottom) in endcap outside tracker region.

Appendix C

Raw energy in calorimeters for different energy bins

As done above for energy range of 100 to 104 GeV in the barrel, we show here the distributions of the raw energy deposited in ECAL(plots on the left), HCAL(plots on the right) and ECAL vs HCAL(plots in the middle).

Figures [C.1](#), [C.2](#),[C.3](#) shows the distributions for pions of energy between 20 GeV to 24 GeV , 144 GeV to 154 GeV , 244 to 254 GeV respectively.

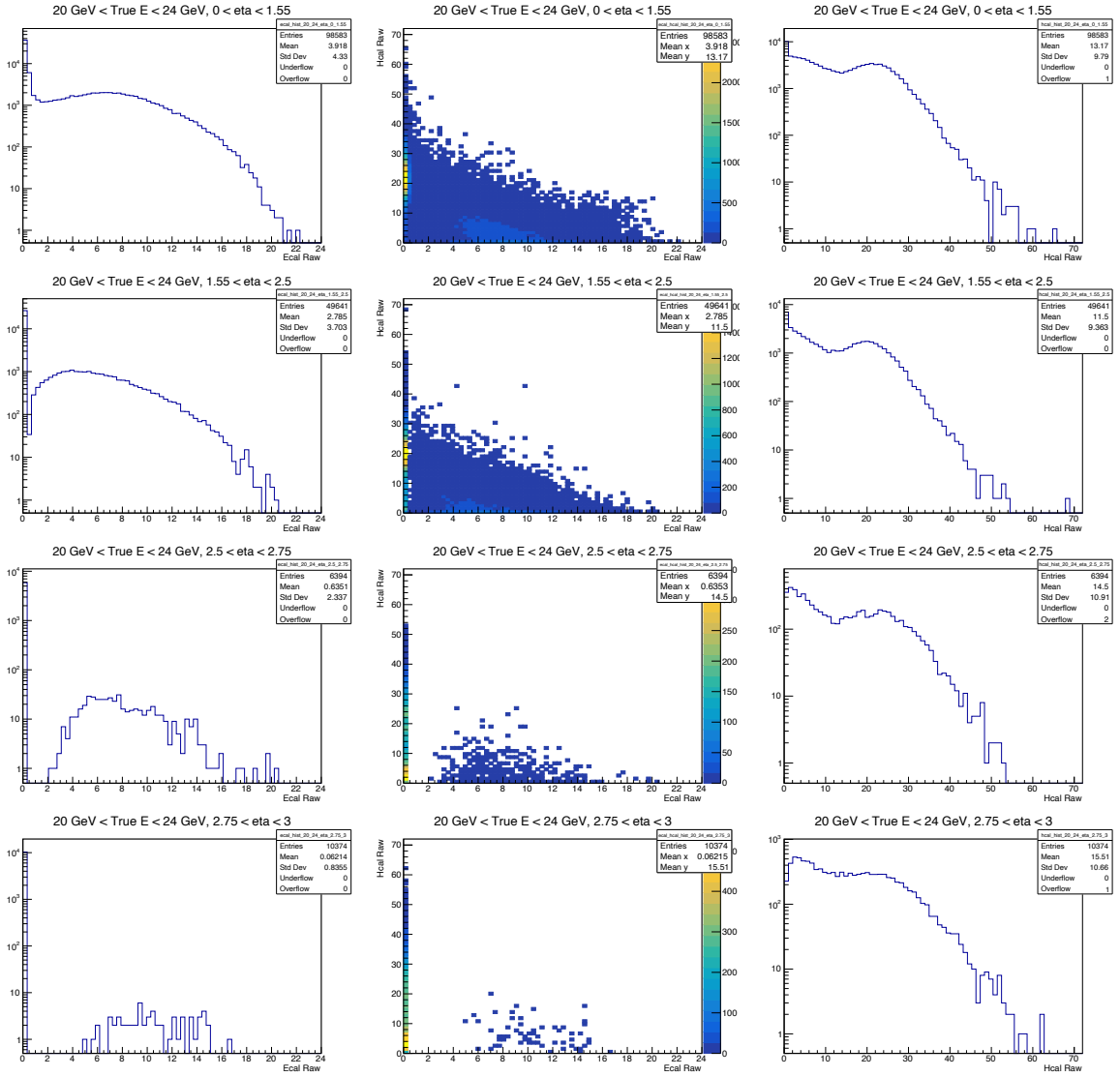


Figure C.1: ECAL raw(left), HCAL raw(right), ECAL raw vs HCAL raw(middle) for pions of energy between 20 to 24 GeV. Top plots are for pions within barrel ($0 < \eta < 1.55$), below them are plots for pions in endcap region with $1.55 < \eta < 2.5$, below them are plots for pions in endcap region with $2.5 < \eta < 2.75$, and finally plots for pions in endcap region with $2.75 < \eta < 3$.

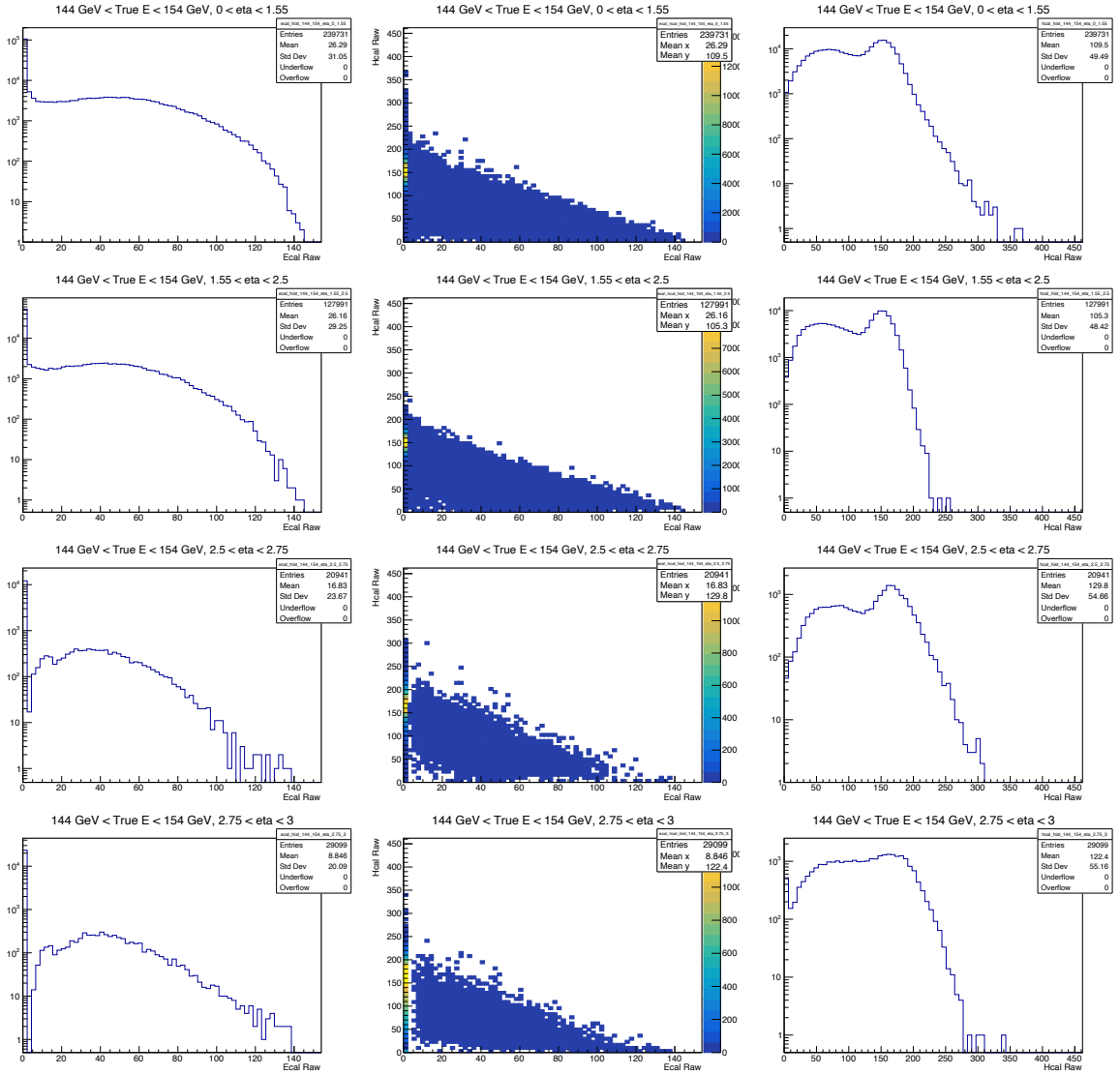


Figure C.2: ECAL raw(left), HCAL raw(right), ECAL raw vs HCAL raw(middle) for pions of energy between 144 to 154 GeV. Top plots are for pions within barrel($0 < \eta < 1.55$), below them are plots for pions in endcap region with $1.55 < \eta < 2.5$, below them are plots for pions in endcap region with $2.5 < \eta < 2.75$, and finally plots for pions in endcap region with $2.75 < \eta < 3$.

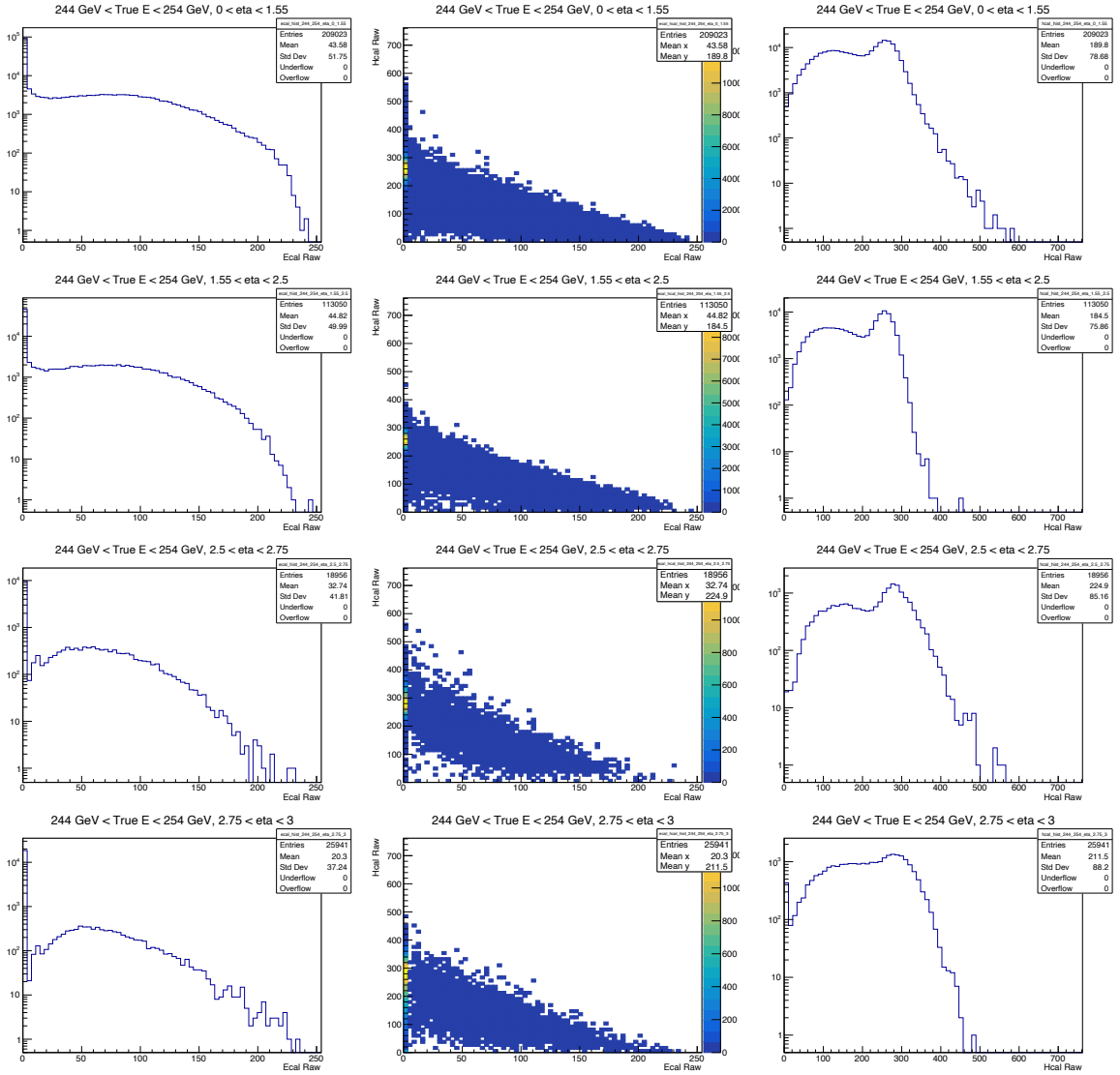


Figure C.3: ECAL raw(left), HCAL raw(right), ECAL raw vs HCAL raw(middle) for pions of energy between 244 to 254 GeV. Top plots are for pions within barrel ($0 < \eta < 1.55$), below them are plots for pions in endcap region with $1.55 < \eta < 2.5$, below them are plots for pions in endcap region with $2.5 < \eta < 2.75$, and finally plots for pions in endcap region with $2.75 < \eta < 3$.

Appendix D

Aggregation step in Dynamic Reduction Network

The following figures illustrate the working of the aggregation step in the Dynamic Reduction Network. The aggregation step is responsible for graph convolution and combining the information from neighbouring rechits which is a crucial step in the reconstruction of mass of pseudoscalar a .

Figure [D.1](#) shows the distribution of rechits or equivalently nodes of the graph in the η - ϕ plane with energy of each rechit on the Z axis. Graph is generated for $k = 9$ in the kNN graph generation step. Figure [D.2](#) represents the graph after applying edge weights. The width and opacity of each edge denotes its edge weight. Finally, the clustered graph is shown in figure [D.3](#). The nodes with the same color or number belong to the same cluster

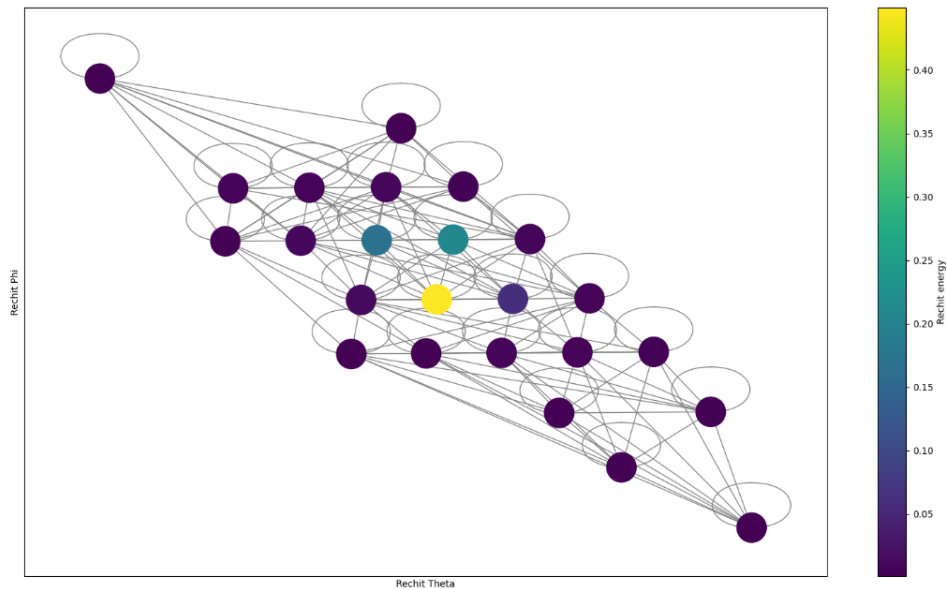


Figure D.1: Rechit θ (angle in η direction) vs Rechit ϕ vs Rechit Energy (on z axis). Graph generated with $k = 9$

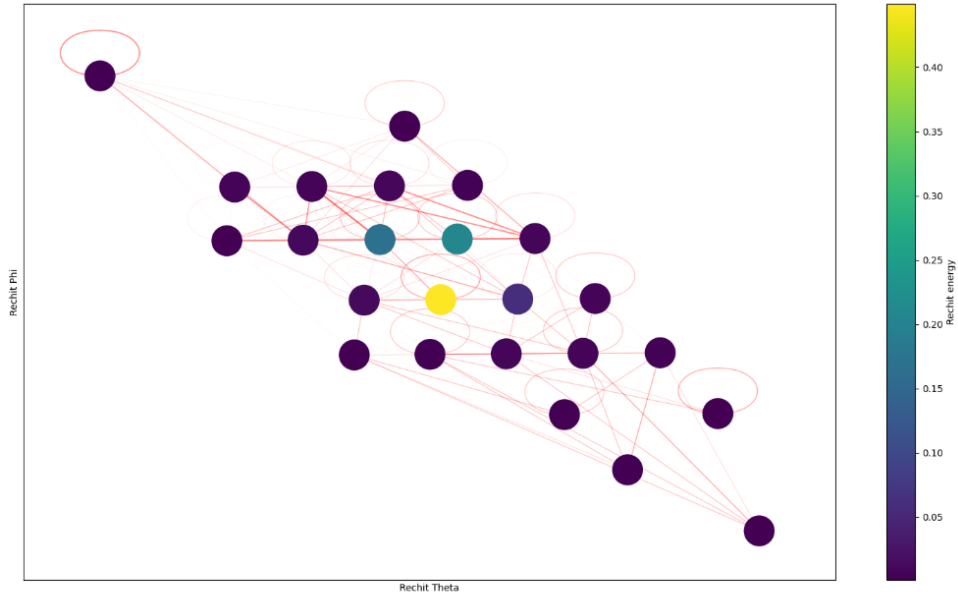


Figure D.2: Rechit θ (angle in η direction) vs Rechit ϕ vs Rechit Energy (on z axis). Width of edge represents edge weight

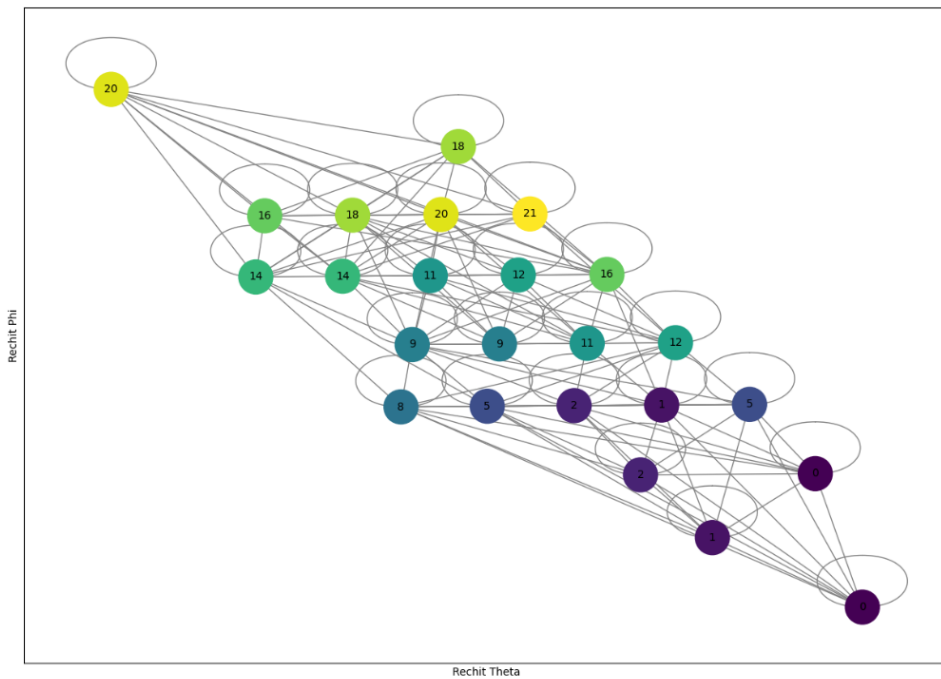


Figure D.3: Rechit θ (angle in η direction) vs Rechit ϕ vs Rechit Energy (on z axis). Same colors / same number on node represent same cluster.