# Reconstruction of merged electrons at CMS

**A Thesis**

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Soumya Sarkar


Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2024

Supervisor: Prof. Sourabh Dube

# Certificate

This is to certify that this dissertation entitled Reconstruction of merged electrons at CMS towards the partial fulfillment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Soumya Sarkar at Indian Institute of Science Education and Research under the supervision of Prof. Sourabh Dube, Associate Professor, Department of Physics , during the academic year 2023-2024.

Prof. Sourabh Dube

Committee:

Prof. Sourabh Dube

Dr. Anshul Kapoor

This thesis is dedicated to my father...

# Declaration

I hereby declare that the matter embodied in the report entitled Reconstruction of merged electrons at CMS are the results of the work carried out by me at the Department of Physics, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. Sourabh Dube and the same has not been submitted elsewhere for any other degree.

Soumya Sarkar

# Acknowledgments

I have always been passionate about doing research, right from my high school years but never knew how to carry out research, and more importantly didn't know what is research. I would like to thank IISER Pune for giving me this opportunity and exposing me to the world of research. I have found, met, and made friends with many people in IISER Pune that I will cherish throughout my life.

I would start by thanking Prof. Sourabh Dube. He has guided me for the last two years, and without any doubt, I would say these are the most enjoyable years of my IISER life. I joined Sourabh's group as an under-confident and introverted person, who is shy about asking questions and speaking up for himself. Sourabh has instilled in me a sense of self-confidence and without a doubt, I have improved a lot, for which I will be always grateful to him. I have always admired the fact of how nicely he explains any concept, always drawing examples and analogies from the most common parts of daily life. It's also fun to get into intense discussions with him on tea or coffee breaks over any topic, be it physics or any other random thing like geography, politics, or economy, and the best part about each discussion is that it will follow a logical path of arguments and counter-arguments just like it happens in science. Apart from academic meetings and informal discussions, I would also cherish the memories of the countless lunches, dinners, and treks with him and the group. It is because of his contributions that made me into becoming who I am today.

I would also take this moment to thank the PhDs of our group, Arnab, Prachurjya, Riya, and Yash. I would like to thank Arnab, for his help and suggestions whenever I got stuck at something. It was also very enjoyable to get into long discussions with him on every aspect of physics. I would also like to express my gratitude towards Prachurjya, Yash, and Riya for their help with all the technical difficulties I faced. I have thoroughly enjoyed all their suggestions and all the informal discussions with them.

I would also like to thank my expert Anshul for his innumerable suggestions and comments

that have helped in the successful completion of this project.
My experience in the EHEP lab wouldn't have been this great without the presence of my co-workers, Parijat and Chitrakshee, who have now turned into great friends, a friendship that I will cherish for life. I have thoroughly enjoyed our innumerable discussions on academics, research, and beyond. I am pleased to have the opportunity to meet Chitrakshee. Beyond our conversations on physics and research, I will always be thankful to her for all the care and help she has given to me. All the mature and delightful conversations with her are something I will remember throughout. I would also like to appreciate my friendship with Parijat. I have enjoyed his company a lot. The never-ending enthusiasm of both of us to discuss physics problems and drooling over it is something I have thoroughly enjoyed. Along with it the great late-night philosophical discussions with him over anything and everything are the best part of my memory that I will rejoice throughout.
Apart from my EHEP group, I have found friends in IISER, who have now turned into lifelong companions. I would like to thank two of my best friends, Ankan and Hritwik. I met them on the first day of my college life at IISER Pune, and they have stuck with me till now. They are the reason I never felt down, and it is because of their encouragement and expectations of me, that I could bear my IISER life with this much ease.
I would also like to thank one of the biggest driving forces behind all my successes, Simantini. I would like to thank her for everything she did for me. She has helped me in innumerable number of ways that I couldn't even list them all. She is one of the biggest reasons I have continued in IISER Pune and I owe almost all my successes to her.
I would also take this opportunity to thank Shalini. She is the person who has been on my side from the day she joined IISER Pune. Her encouragement, expectations, and belief in me are something that made my MS thesis days bearable.
At last but not the least I would like to thank my parents for believing in me. It is them who have ignited a spark in me, and it is because of their effort I could reach this far in life.

# Abstract

The fundamental composition of and interactions of matter in the universe is described by a collection of quantum field theories known as the Standard Model (SM). Though the SM has been thoroughly tested at colliders such as Tevatron and the Large Hadron Collider (LHC), there remains strong motivation for physics beyond the SM or BSM.

The LHC searches for BSM rely heavily on reconstructing and identifying clean and isolated electron trajectories. However, a class of BSM model predicts the very close or merged electron signatures in the detectors. A merged electron means that there is a huge overlap of clusters among the two electrons. One such model is the Right Handed Neutrino's (RHN). An SM counterpart of the above is a boosted photon or $Z$ boson giving close-by or merged electrons. All the current reconstruction algorithm fails to reconstruct the two individual electrons.

In this study reconstruction of these merged electrons is studied. Multivariate analysis (MVA) techniques like a neural network (NN) classifier have been used to tag merged electrons. The NN classifier showed good performance in tagging these objects and separating them from genuine clean and isolated electrons.

# Contents

**Abstract** **vii**

**1 The Standard Model and Beyond** **7**

1.1 Quantum Electrodynamics (QED) . . . . . 8

1.2 Quantum Chromodynamics (QCD) . . . . . 9

1.3 The Weak Interaction . . . . . 9

1.4 Limitations of Standard Model . . . . . 9

**2 The CMS experiment** **11**

2.1 Coordinate system at CMS . . . . . 11

2.2 Some useful Definitions used in CMS . . . . . 12

2.3 The CMS detector . . . . . 13

**3 Reconstruction of electrons and photons at CMS** **16**

3.1 Clustering and Superclustering . . . . . 16

3.2 Electron track reconstruction . . . . . 19

3.3 PF algorithm . . . . . 21

3.4 When does the reconstruction fail? . . . . . 22

**4 Analysis** **24**

4.1 Workflow . . . . . 24

4.2 Source of merged and single electrons . . . . . 26

4.3 ECAL cluster distribution of merged and single electrons . . . . . 28

4.4 ECAL properties of electrons . . . . . 31

4.5 Track properties of electrons . . . . . 34

4.6 Hybrid properties of electrons . . . . . 37

**5 Classification of electrons** **39**

5.1 Neural Networks (NN) . . . . . 39

5.2 Output of neural networks . . . . . 41

5.3 Training the merged electron classifier . . . . . 43

**6 Results** **45**

**7 Conclusion** **47**

# List of Figures


1 Feynman diagram of the right handed neutrino's (RHN) production and its subsequent decay. A light mass RHN, $N_l$ will be produced with high boost and hence the two leptons from the decay of the $N_l$ will be very close and give the merged lepton signatures. . . . . 6

1.1 The Standard Model of elementary particles. *image courtesy : Wikipedia* . . 8

2.1 The CMS detector follows a cylindrical coordinate system. The coordinate system along with some kinematic properties like $p_{\mathrm{T}}$ are shown in this figure. [1] . . . . . 12

2.2 Transverse section of the CMS detector. On the top the length scale is given, that shows the radial width of various detector components. [2] . . . . . 13

3.1 Flowchart showing the steps in the reconstruction of electrons and photons at CMS. . . . . 17

3.2 Pictorial depiciton of Bremsstrahlung. . . . . 18

3.3 The mustache supercluster. It shows the distribution of PF clusters around the seed cluster at $(0, 0)$ (shown in white). Most of spread is in $\phi$ direction with little spread along $\eta$ direction due to the bremsstrahlung. [3] . . . . . 19

3.4 Three scenarios depicting the limitations of CMS reconstruction algorithm as the electrons come closer and closer and finally become merged. . . . . 23

4.1 Workflow followed by me to generate the samples and analyze them. . . . . 25

4.2 The above plots (a) and (b) show $dR$ distribution of generated and reconstructed electrons respectively, between the two electrons coming from the decay of $J/\Psi$. Clearly, most of the electrons lie within $dR < 0.1$ of each other. The fraction of electrons below $dR < 0.1$ is also reduced for the reconstructed scenario than generated. This will be more clearly shown in the pie charts Fig 4.4. . . . . 27

4.3 The above plots (a) and (b) show $dR$ distribution of generated and reconstructed electrons respectively, between the two electrons coming from the Drell-Yan process. Most of the electrons are far away from each other with a peak at around $dR \approx 1$. . . . . 27

4.4 The above pie charts show the number of reconstructed electrons for two $dR$ ranges. (a) $dR < 0.1$ and (b) $0.1 < dR < 0.2$. For $dR < 0.1$, almost in 77% of the events, only one electron is reconstructed, which shows the CMS reconstruction algorithm is not able to reconstruct very close-by electrons. In the range $0.1 < dR < 0.2$, in 65% of the events both the electrons are reconstructed, hence the reconstruction algorithm is performing better for high $dR$ values. $dR$ mentioned above is between the two generated electrons coming from the decay of $J/\Psi$. . . . . 28

4.5 In figure (a) Shows the Cluster distribution for the single electron around the seed cluster at $(0, 0)$ in the $\Delta\eta$-$\Delta\phi$ plane. (b) Shows the zoomed cluster distribution around $(0, 0)$. Each cluster is weighted by a factor of $E_{cluster}/E_{seedcluster}$. 29

4.6 In figure (a) Shows the Cluster distribution for a merged electron around the seed cluster at $(0, 0)$ in the $\Delta\eta$-$\Delta\phi$ plane. (b) Shows the zoomed cluster distribution around $(0, 0)$. Each cluster is weighted by a factor of $E_{cluster}/E_{seedcluster}$. The distributions for merged electrons are thicker than single electrons. . . . 29

4.7 In figure (a) Shows the rechit distribution for the single electron around the seed crystal at $(0, 0)$ in the $\Delta\eta$-$\Delta\phi$ plane. (b) Shows the zoomed crystal distribution around $(0, 0)$. Each crystal is weighted by a factor of $E_{crystal}/E_{seedcrystal}$. 30

4.8 In figure (a) Shows the rechit distribution for a merged electron around the seed crystal at $(0, 0)$ in the $\Delta\eta$-$\Delta\phi$ plane. (b) Shows the zoomed crystal distribution around $(0, 0)$. Each crystal is weighted by a factor of $E_{crystal}/E_{seedcrystal}$. Again merged electrons have a bit thicker distribution. . . . . 31

4.9 Fig above shows the distributions of (a) $r9$ and (b)$\sigma_{\eta\eta}$. The distributions are compared for merged and single electrons. They show differences as expected from the definitions of those quantities. . . . . 32

4.10 Fig above shows the distributions of (a) $\sigma_{i\eta i\eta}$ and (b)$\sigma_{i\phi i\phi}$. Again the distributions for merged and single electrons and these quantities show some difference for merged and single electrons. . . . . 33

4.11 The above figure shows the distribution of $E_{corrected}/E_{raw}$ compared for merged and single electrons. The peak of the distribution slightly shifts to the right for merged electrons and hence the correction factor is more for merged electrons than single electrons. . . . . 34

4.12 Fig above shows the number of GSF tracks around the seed GSF track for (a) $dR < 0.1$ and (b)$dR < 0.05$, for single and merged electrons. For merged electrons, it peaks at two but for single electron, it peaks at one as expected. 35

4.13 Invariant mass of the two very close GSF tracks within $dR < 0.1$ of the seed GSF track. For the merged electrons they peak at the $J/\Psi$ mass of around 3.1GeV, but for single electrons, they are just backgrounds. . . . . 36

4.14 Figure above shows the $p_{\mathrm{T}}$ distributions of the seed GSF track and reconstructed electron for (a) merged (b) single electrons. The agreement between the two $p_{\mathrm{T}}$'s is great for single electrons, but it's worse for merged electrons. 36

4.15 The figure shows the distribution of the quantity: $\frac{p_T(seedtrk)-p_T(electron)}{p_T(electron)}$, as described in the text. The single electron peaks at 0, but for merged it doesn't peak at 0. . . . . 37

4.16 The figure above compares the (a)$\eta_{trk-in} - \eta_{SC}$ and (b)$\phi_{trk-in} - \phi_{SC}$, for merged and single electrons. The merged electrons have a wider width of the distributions than single electrons. . . . . 38

4.17 Overlayed plots of (a)$E_{corrected}/P_{seedtrk}$ and (b)$E_{raw}/P_{seedtrk}$, are shown for merged and single electrons. The merged electrons have larger widths and hence larger values of $E/P_{track}$. . . . . 38

5.1 The basic NN architecture, showing the input layer, hidden layer/layers, and the output layer, along with the edges connecting the nodes. [4] . . . . . 40

5.2 The figure shows a schematic NN score plot. The signal events have a higher NN score and hence peak to the right i.e. close to 1. The backgrounds on the other hand peak on the left i.e. close to 0. The figure also shows the true positive regions (TPR) and false positive regions (FPR). [5] . . . . . 42

5.3 An example ROC curve. The curve regions of better and worse performance of an NN. An AUC of 0.5 corresponds to a diagonal along the ROC curve. It depicts a random NN classifier, with no separation in the NN score plot. As the curve convexs upwards from the diagonal, the NN performs better in categorization signals and backgrounds. A perfect classifier has an AUC score of 1. [6] . . . . . 43

6.1 The above figure shows the NN score of the NN classifier, which classifies merged electrons as signals and single electrons as backgrounds. Hence the merged electron peaks towards 1 (right) and the single electron towards 0 (left) 45

6.2 The ROC curve for the NN classifier used to classify merged and single electrons. The ROC for the training and testing datasets are very close giving an AUC score of 0.85 for both the training and testing datasets . . . . . 46

# Introduction

Particle physics is the study of matter and their interactions at the fundamental level. The most successful model that is able to precisely predict most of the observations to date is the standard model (SM) of particle physics. The SM is a collection of quantum field theories. According to SM all of the visible matter is composed of quarks and leptons and these interact via three kinds of fundamental forces namely electromagnetic, strong, and weak forces. The fourth force gravity is not included in the standard model. Though the SM is hugely successful in explaining many of the phenomena observed, it is known that SM is not complete. From astrophysical observations, it is known that around 95 % of the universe is made up of dark matter and dark energy which is not explained in the framework of SM. Also from neutrino oscillations, it is known that neutrinos have non-zero masses. The SM predicts neutrinos as massless, hence it also can't explain neutrino oscillations [7]. To account for these gaps in SM many beyond standard model (BSM) theories have been proposed to date. One such theory is the existence of right-handed neutrinos (RHN) [8]. According to the standard model, only left-handed neutrinos exist in nature (and only right-handed anti-neutrinos). However, according to this new model, the RHN might exist in nature. If the RHN exists they will interact via the SM neutrinos and other leptons. One way in which the RHN interacts is shown via the Feynman diagram in Fig 1.

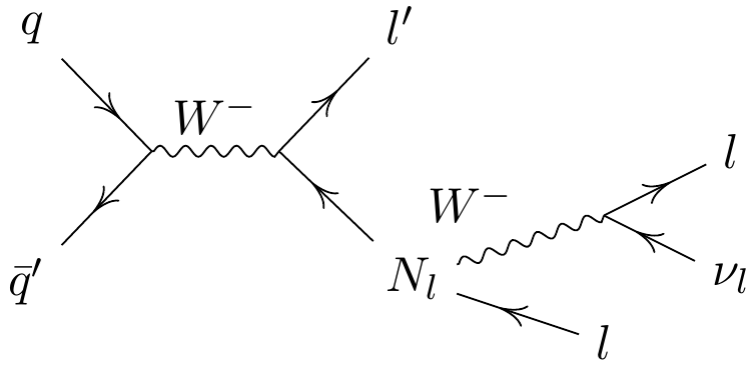



Figure 1: Feynman diagram of the right handed neutrino's (RHN) production and its subsequent decay. A light mass RHN, $N_l$ will be produced with high boost and hence the two leptons from the decay of the $N_l$ will be very close and give the merged lepton signatures.

If the RHN that exists is a low mass RHN (say around 2 GeV's) then it will be highly boosted as it decays from a W-boson (which has a mass of around 80 GeV). Hence, its decay products will be very close to each other. From the Feynman diagram Fig 1 the two leptons, will be really close, and hence act as a merged object. In the detector, these merged objects will appear as a single object, and hence the current reconstruction algorithm at CMS fails to reconstruct these as two separate objects. Previously such studies have been done for muons [9], but no such study has been done for electrons. Hence, this project aims to increase the efficiency of detecting these merged electron objects. Increasing the efficiency of reconstructing these merged electrons, could significantly improve the discovery potential of low-mass right-handed neutrinos.

Chapter 1 describes the Standard Model (SM) along with its limitations and the need for beyond standard model (BSM) theories. Chapter 2 describes the CMS detector, the coordinate systems along with its various components. Chapter 3 describes the full chain of reconstruction for electrons and photons at CMS. Also it briefly discusses the limitation of this reconstruction scheme, in the merged electron scenarios. Chapter 4 gives the overview of the work that could reconstruct these merged electrons. Chapter 6 and 7 gives the results that has been achieved along with the future prospective of this work.

# Chapter 1

# The Standard Model and Beyond

Standard Model is a non-abelian gauge theory with a group structure of $SU(3) \times SU(2) \times U(1)$. The SM Lagrangian can be broken as the sum of all the fundamental interaction Lagrangians:

$$\mathcal{L}_{SM} = \mathcal{L}_{QED} + \mathcal{L}_{QCD} + \mathcal{L}_{Weak}$$

The Standard Model (SM) predicts all the visible matter in the universe is made up of quarks and leptons. These quarks and leptons interact via fundamental forces which are mediated by the exchange of gauge bosons, which is shown in Fig 1.1. The photon ($\gamma$) is the mediator of the electromagnetic force and it is described by Quantum Electrodynamics (QED). The strong force is mediated by gluons and the theory describing it is called Quantum Chromodynamics (QCD). The weak interactions are mediated by $W^+, W^-$ and $Z$. The only scalar boson in SM, the Higgs boson was discovered in 2012 at CERN by both ATLAS and CMS experiments [10]. The Higgs interactions and mechanism are responsible for the bare masses of all the fundamental particles that couples to it. Photons and gluons don't directly couple to higgs, and hence are massless particles.
In Fig 1.1 all the fundamental particles of the SM are listed along with their mass and spins. The leptons and quarks are fermions and they are categorized into three generations based on the mass. The mediators of the forces are bosons. The Higgs boson has spin = 0 and hence it's a scalar boson. In total, the SM consists of a total of 61 particles and anti-particles.

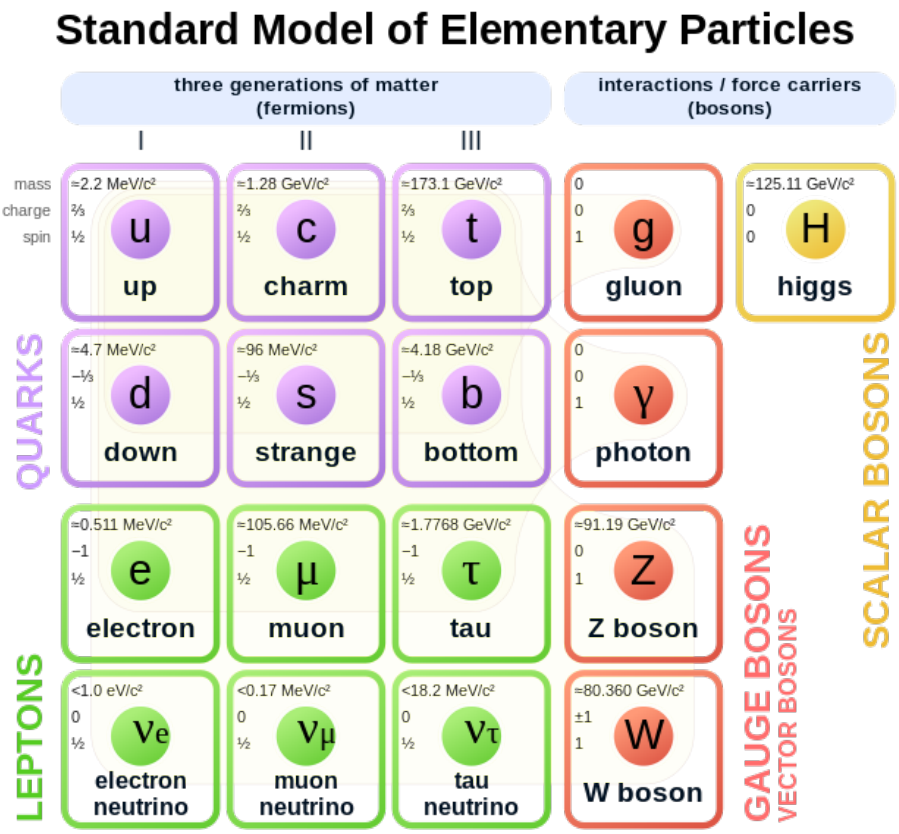



Figure 1.1: The Standard Model of elementary particles. *image courtesy : Wikipedia*

## 1.1 Quantum Electrodynamics (QED)

QED describes the electromagnetic interaction i.e. interaction of charged particles with photons and is the most complete theory to date for light-matter interactions. QED is an abelian $U(1)$ gauge theory. The mediators of QED are photons, $\gamma$ which are massless and have spin = 1. All particles with electromagnetic charge interact via photons.

## 1.2 Quantum Chromodynamics (QCD)

QCD is a non-abelian $SU(3)$ gauge theory. It describes the strong nuclear interaction i.e. the interaction of quarks and gluons. Quarks come in three color states namely red ($r$), blue ($b$), green ($g$), and the corresponding anti-colors. Quarks and gluons can form bound states called hadrons. Hadrons can only exist in the color-neutral states. The quark-antiquark ($q\bar{q}$) state is called mesons and three quark states ($qqq$) are called baryons. For example, pions ($\pi^{\pm}, \pi^{0}$) are mesons, and protons and neutrons are baryons.

## 1.3 The Weak Interaction

The weak interaction is a non-abelian $SU(2)$ gauge theory. Weak interaction violates parity as only the left (right) handed fermions (anti-fermions) take part in interactions. It also violates flavor symmetries, for example, $u$ quark converted to a $d$ quark with a weak vertex. The mediators of weak interactions are massive. $W^{\pm}$ bosons have a mass of around 80.4 GeV and $Z$ with a mass of 91.2 GeV. The neutrinos in SM only interact via weak interactions.

## 1.4 Limitations of Standard Model

Though the SM is hugely successful in explaining many of the phenomena with very high precision, it still has limitations and is unable to explain some observations. These include:

- **Neutrino Oscillations**: It is a well-established fact that neutrinos exist as mass eigenstates and hence exist as a superposition of flavor eigenstates [7]. This observation proves that neutrinos have masses, but SM predicts neutrinos as massless. Hence in this case SM is in contradiction with the experiment.
- **Dark Matter candidates**: From cosmological observations of galaxy rotation curves the existence of dark matter is proven. Dark matter constitutes 27% of the universe, and only 5% of the universe is visible matter. But the SM has no description of dark matter. There is no particle in SM that could be a dark matter candidate.

These are a few of the many observations that could not be explained by the SM. Hence many beyond standard model (BSM) models are proposed to explain these observations. These BSM models are currently being tested at many places around the globe. The Large Hadron Collider (LHC) at CERN is one such place, where new models are tested in the high mass and high energy frontier.

# Chapter 2

# The CMS experiment

The CMS experiment is located at the Large Hadron Collider. The LHC is a giant 27 km ring located at CERN, Geneva. At the LHC, proton beams are collided at a record COM energy of $\sqrt{s} = 13.6$ GeV at every 25ns, hence, a collision or event rate is around 400 MHz. At such high energies, many new particles are produced in the collision. Most of the particles are unstable and decay to other stable particles. There are four detectors placed at four points around the LHC ring. They are like cameras, detecting the particles produced in these collisions. CMS is one of the four detectors.
At the CMS experiment, searches for physics beyond the standard model are ongoing. As explained in Section 1.4, these new BSM models are tested in the energy frontier at CMS. Also at CMS precision measurements are carried out for many SM quantities.

## 2.1 Coordinate system at CMS

The CMS detector follows the cylindrical coordinate system as shown in Fig 2.1, with the beam axis along the z direction. The proton-proton collisions take place along the beam axis. The collision vertex is known as the interacting point (IP) and is the center of the coordinate system in this case. The x-axis is from the beam direction toward the center of the LHC ring and the y-axis points upwards towards the ground. The angle around the z-axis is the azimuthal angle ($\phi$), and the angle the vector makes with the x-y plane is the

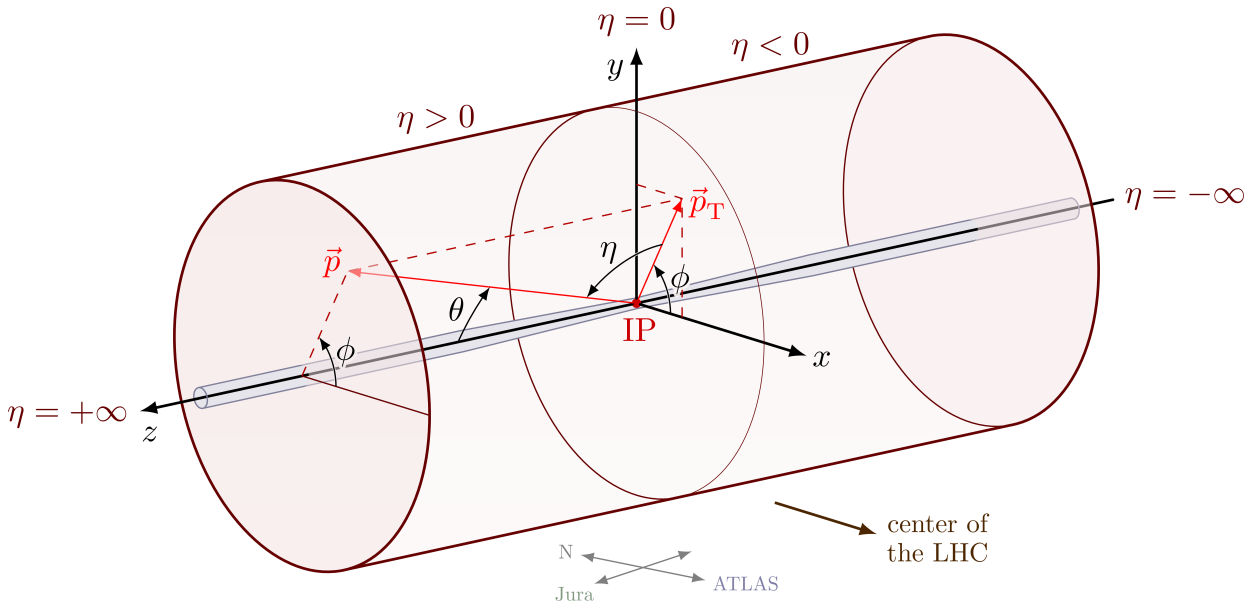



Figure 2.1: The CMS detector follows a cylindrical coordinate system. The coordinate system along with some kinematic properties like $p_{\mathrm{T}}$ are shown in this figure. [1]

polar angle $\theta$. As protons are hadrons, which means they are formed of quarks and gluons. Hence they have a substructure, therefore Lorentz invariant coordinate systems are used like rapidity, $y = \frac{1}{2} \ln \frac{E+p_z}{E-p_z}$, and pseudorapidity, $\eta = -\frac{1}{2} \ln \tan(\theta/2)$. The difference in $y$ and $\eta$ between two points i.e. $\Delta y$ and $\Delta \eta$ are Lorentz invariant quantities. Thus when moving from the proton frame to the lab frame, these quantities remain invariant. Pseudorapidity is an approximation to rapidity for very low-mass particles and exactly equals rapidity for massless particles. Pseudorapidity is easier to calculate in the transverse plane and hence widely used at CMS.

## 2.2 Some useful Definitions used in CMS

- $\overrightarrow{p_T} = p_x\hat{x} + p_y\hat{y}$, is the transverse momentum and its magnitude is defined as $p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2}$.
- $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, which shows the angular separation between two objects which are $\Delta\eta$ and $\Delta\phi$ apart in the $\eta - \phi$ plane.
- $M_{invariant} = E^2 - \overrightarrow{p}^2$, where $\overrightarrow{p} = p_x\hat{x} + p_y\hat{y} + p_z\hat{z}$. For two-particle in the final state,

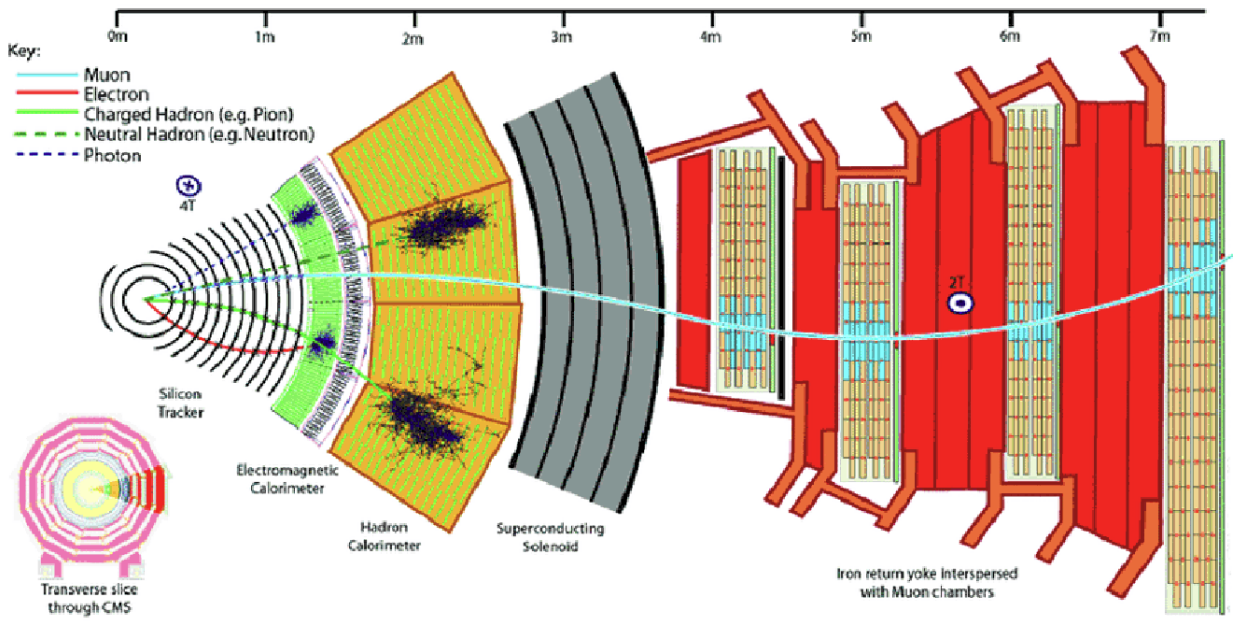



Figure 2.2: Transverse section of the CMS detector. On the top the length scale is given, that shows the radial width of various detector components. [2]

with four vectors $(E_1, \overrightarrow{p_1})$ and $(E_2, \overrightarrow{p_2})$, $M_{invariant} = (E_1 + E_2)^2 - (\overrightarrow{p_1} + \overrightarrow{p_2})^2$

## 2.3 The CMS detector

The CMS detector weighs 14,000 tonnes. It has a height of 15 meters and a length of 21 meters. The solenoid magnet produces a magnetic field of 3.8 T along the axis of the detector [11]. The CMS detector is like a cylindrical onion, with each concentric layer having components, that help to measure the position, momentum, and energy of all the stable particles produced in the collisions, and from that information reconstruct the full event information of what has happened in the collision. The transverse section of the detector is shown in Fig 2.2. Each of the components of the detector is explained below.

### 2.3.1 The Tracker

The tracker is used to reconstruct the path of charged particles [11]. The charged particles bend in the magnetic field leaving behind hits in the tracker. From the hit the track is

reconstructed. The curvature of the track in the magnetic field helps in the reconstruction of the momentum of the object. The tracker is made completely of silicon. The core few layers are silicon pixel detectors and the outer layers are silicon microstrip detectors. Silicon is chosen because of its high granularity, to accurately get the position of the hit, also silicon is radiation hard, hence it can withstand radiation damage which is required as the inner tracker receives the highest flux of particles. Also, silicon chips help in faster readout-electronics which is also needed as protons collide every 25 ns.

### 2.3.2 Electromagnetic Calorimeter or ECAL

The electromagnetic calorimeter or ECAL measures the energy of electrons and photons ([11], [12]). Any particle that interacts electromagnetically showers in the ECAL. The electrons and photons are mostly completely stopped by the ECAL and deposit all their energy. ECAL is made up of lead tungstate crystals ($PbWO_4$ crystals). The $PbWO_4$ crystals have a high density (8.28 g/$cm^3$), small Moliere radius (2.3 cm), and small radiation length (0.89 cm) which makes it perfect to make a close compact calorimeter and good resolution of close-by clusters.
The ECAL barrel covers a range of $|\eta| < 1.48$ and endcap covers a range till $|\eta| < 3$. There are 61,200 $PbWO_4$ crystals in the barrel region which are subdivided into 36 supermodules. If the barrel is flattened as a rectangle with arrays of crystals, there are 360 crystals along $\phi$ direction numbered 1 through 360 and 170 crystals along $\eta$ direction numbered from -85 through 85 except at 0. Hence on average a crystal width in $\Delta\eta \times \Delta\phi$ is $0.0175 \times 0.0175$ which is $2 \times 2cm^2$. These numbering of 1 to 360 and -85 to 85 are called the $i\eta$ and $i\phi$ indices or crystal indices and will be a useful handle as will be discussed later. There are also an additional 15,000 crystals in the endcaps. A pre-shower detector is also installed in front of endcaps to distinguish high energy photons from pairs of low energy photons.

### 2.3.3 Hadron Calorimeter or HCAL

The hadron calorimeter or HCAL is designed to capture and measure the energy of hadrons (particles made of quarks and gluons), coming out of the collision. These hadrons could be charged hadrons like protons ($p$), pions($pi^{\pm}$), Kaons ($K^{\pm}$) and also neutral hadrons like neutrons ($n$), neutral pions ($\pi^0$) etc. The charged hadron will also have some showering in

the ECAL as they also interact electromagnetically, but the majority of the showering will occur in HCAL.
HCAL barrel (HB) is made of brass scintillators and covers a range of $|\eta| < 3.0$. HCAL endcaps (HF) covers the remainign $\eta$ range i.e $3.0 < |\eta| < 5.0$. Hence, HCAL is almost hermitic so that it can detect any invisible particle, using missing energy and momentum in the transverse direction.
The HCAL is a sampling calorimeter, which means it measures the particle's energy, position and time of arrival using layers of absorbers and scintillators. The HCAL formas the last layer of detectors within the solenoidal magnetic coil, except an outer HCAL barrel (HO) exists outside the coil to absorb any hadrons that might have leaked the inner barrel i.e HB.

### 2.3.4 Muon Chambers

The last layer of detectors are dedicated for detecting muons. Muons are minimum ionizing, hence they deposit very little to almost no energy in the ECAL and HCAL. Hence the muon chambers and stations are placed as the outermost layer, outisde the magnetic coil, where muons are the only particles likely to produce signals.
The muon detector consists of alternating layers of muon stations with iron return yoke. The muons produce hits in the muon stations, and from fitting those hits to the inner hits in silicon tracker, gives the track of the muons from which muon's momentum can be reconstructed.
There are 1400 muon chambers, with 250 drift tubes (DTs) and 540 cathode strip chambers (CSCs) to track the muon's positions and provide a trigger, while 610 resistive plate chambers (RPCs) and 72 gas electron multiplier chambers (GEMs) form a redundant trigger system. These can quickly decide whether to keep or discard the acquired muon data. Because of the many layers of detector and different specialities of each type, the system is naturally robust and able to filter out background noise.

# Chapter 3

# Reconstruction of electrons and photons at CMS

The electrons and photons are reconstructed both at the online level and offline levels [12]. Online reconstructions are performed during the data taking i.e. at trigger level. Offline reconstruction is performed after the data is collected i.e. at analysis level. For this work, mostly offline reconstruction will be discussed [12]. During Run-1 CMS uses separate algorithms to reconstruct electrons and photons. But from Run-2 CMS uses a common algorithm to reconstruct both electrons and photons. It reconstructs them as a combined $e/\gamma$ object. The tracking is done via the Gaussian sum Filter (GSF) tracking algorithm and for energy measurement, the ECAL mustache superclustering algorithm is used. The complete chain of reconstruction is shown in Fig 3.1. Each step of this process is described below.

## 3.1 Clustering and Superclustering

Electrons and photons both deposit almost all their energy in the electromagnetic calorimeter (ECAL). The ECAL is completely destructive, hence electrons and photons deposit almost their complete energy by showering in the ECAL. Because of these showers, their energy deposits are not bound to one crystal but are spread over a group of crystals. These continuous groups of crystals are called **PF clusters** or **basic clusters** or **CaloClusters**. But there is another tricky thing associated with electrons, as electrons are charged they bend in the

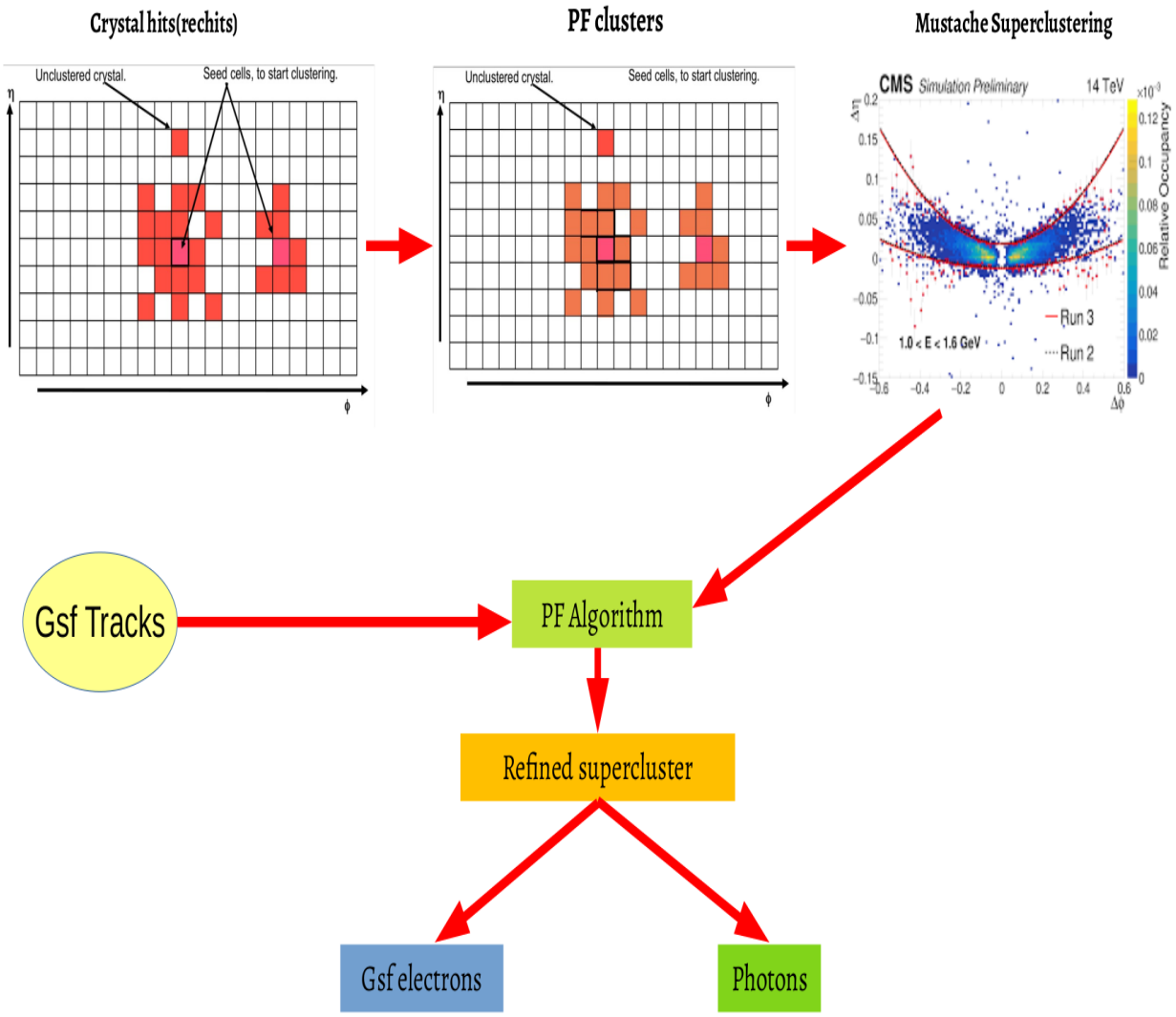
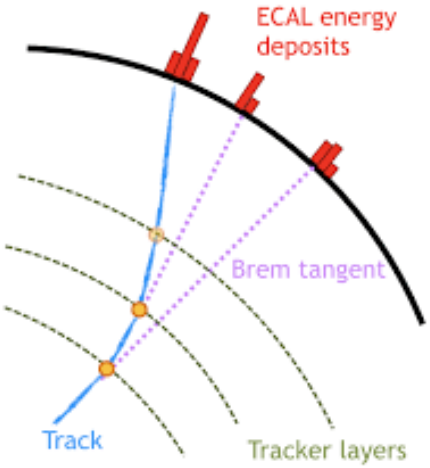
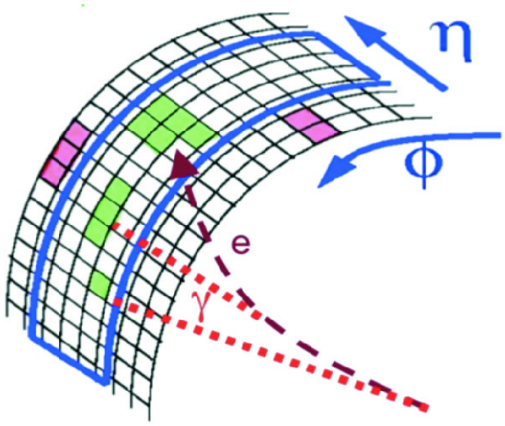



Figure 3.1: Flowchart showing the steps in the reconstruction of electrons and photons at CMS.

magnetic field. Due to this, they emit photons. This is called Bremsstrahlung, as shown in the Fig 3.2. To reconstruct the total energy of the electron with which it was produced all the bremsstrahlung photons have to be collected i.e. all the clusters that are produced due to bremsstrahlung has to be collected into one more clusters of cluster. This step is called superclustering as it is the clustering of the clusters.




(a) pictorial depiction of Bremsstrahlung. It is shown how the bremsstrahlung photons can be recovered by drawing tangents from each tracker layer of the track. [13]




(b) This figure depicts how the bremsstrahlung photons and the parent electron energy and cluster deposits will look in the ECAL. All these PF clusters has to be collected to form the super-cluster. [14]

Figure 3.2: Pictorial depiciton of Bremsstrahlung.

Currently, CMS uses the mustache superclustering algorithm in the ECAL to recover the bremsstrahlung photons [12]. As shown in Fig 3.3, it looks like a mustache, hence the name. The mustache supercluster Fig 3.3, is the distribution of PF clusters around the seed cluster (highest energy cluster). Therefore the seed cluster is always at $(0, 0)$ and $\Delta\eta$, $\Delta\phi$ are the $\eta$ and $\phi$ coordinates of the clusters with respect to the seed cluster. The large spread in the $\phi$ direction is due to the bremsstrahlung photons which are mostly distributed along $\phi$ direction, due to the magnetic field being along $z$-direction. As both electrons and photons have similar deposit patterns in the ECAL, hence mustache superclustering is performed to reconstruct both electrons and photons.

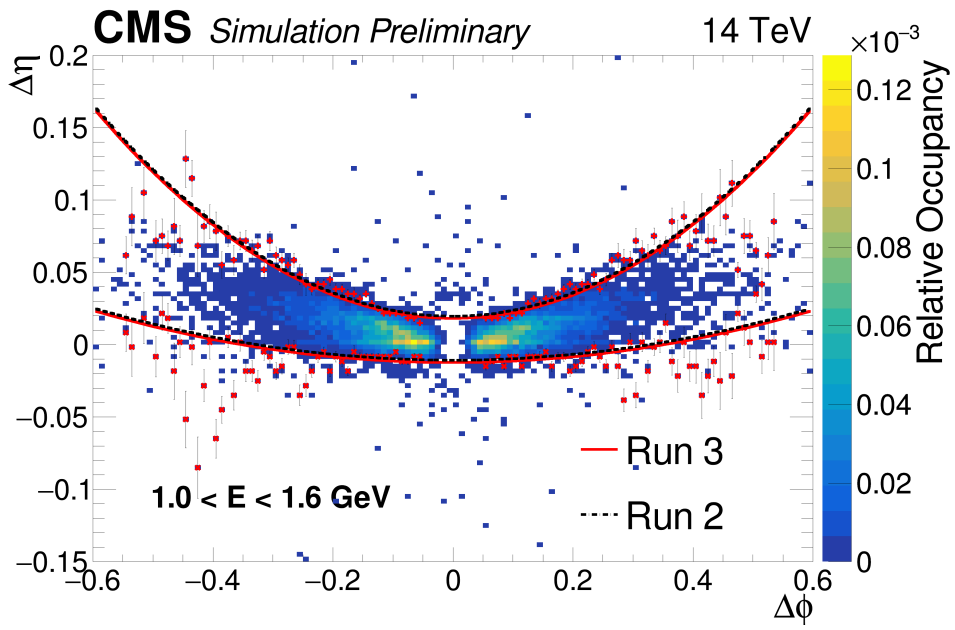



Figure 3.3: The mustache supercluster. It shows the distribution of PF clusters around the seed cluster at $(0, 0)$ (shown in white). Most of spread is in $\phi$ direction with little spread along $\eta$ direction due to the bremsstrahlung. [3]

## 3.2 Electron track reconstruction

The tracks for electrons are reconstructed using **Gaussian Sum Filter (GSF)** algorithm. GSF tracking algorithm takes into account the radiative losses due to bremsstrahlung. GSF tracking was used for both Run 1 and Run 2 by CMS. CMS also has a general track reconstruction algorithm known as **Kalman Filter (KF)** algorithm which is used for all charged particle tracks and not just specifically for electrons.

### 3.2.1 Track Seeding

The GSF track fitting algorithm is CPU intensive and, hence cannot be run on all reconstructed hits in the tracker. Therefore it starts with identifying hit patterns that might resemble an electron trajectory. This process is called seeding. The seeds can be either **ECAL-driven** or **tracker-driven** [12].

ECAL-driven seeds starts from the reconstructed mustache SCs with transverse energy, $E_{SC,T} > 4$ GeV and $H/E_{SC} < 0.15$, where $H$ is the sum of energy deposits of all HCAL

towers within a cone $\Delta R < 0.15$ centered around the SC position. For getting the seed hits in the tracker, a helical trajectory is extrapolated from the ECAL SC, towards the collision vertex. This extrapolation neglects the effect of bremsstrahlung photons. The first two hits of the tracker is then matched to the extrapolated track from SC within some charge dependent $\Delta z \times \Delta\phi$ window for barrel and $\Delta r \times \Delta\phi$ window for endcaps. If they are within that window they are selected as seeds for the GSF track.
The tracker-driven approach iterates over all general tracks i.e KF tracks. The seeds of all those KF tracks that are compatible with ECAL supercluster are used as seeds for the GSF tracking algorithm. The compatibility check is done using a BDT or cut-based method that uses track quality and track-cluster matching variables as inputs. Tracker-driven seeding is done during offline reconstruction and not at HLT level, as it is computationally expensive to reconstruct all tracks in an event.
The GSF tracks from ECAL-driven approach works better for high $E_T$ isolated electrons, with a seeding efficiency larger than 95% for $E_T > 10$ GeV for electrons from Z boson decay. The tracker-driven approach works better for low $p_T$ tracks or non-isolated electrons. It has a seeding efficiency of $\approx 50\%$ for electrons with $p_T > 3$ GeV. The tracker-driven approach also helps to recover efficiency in regions such as barrel-endcap transition region and/or in the gaps between supermodules.
The GSF tracking algorithm runs on both ECAL-driven and tracker-driven seeds.

### 3.2.2 Tracking

The final electron seeds collected (both ECAL-driven and tracker-driven) are used to initiate the reconstruction of electron tracks. Final track reconstruction is like connecting the hits so that it reconstructs a real object, in this case, an electron. Starting with a given seed, track parameters are iteratively calculated at each layer using the KF algorithm, with the electron energy loss modeled using Bethe-Heitler distribution [12]. If the KF algorithm finds multiple hits compatible with the prediction in the next layer, it forms all the tracks with the $\chi^2$ values associated with each track. The track with the lowest $\chi^2$ gets the preference. The tracks reconstructed can have at most one missing hit. If the track has one missing hit it is penalized by increasing its $\chi^2$. This penalizing helps to reduce the inclusion of tracks that arise from converted bremsstrahlung photons from the primary electron trajectory. Finally, if one hit is common to many track candidates, the track with minimum $\chi^2$ or lower number

of missing hits is considered.
Once these track candidates are reconstructed using KF algorithm, the parameters at each layer are measured using a GSF fit, where energy loss is approximated using an admixture of Gaussian distributions. The GSF tracks formed using this procedure is extrapolated to the ECAL under the homogenous magnetic field assumption, by fitting a helix. Then track-cluster association is performed on those GSF tracks.

### 3.2.3 Track-cluster association

The reconstructed GSF tracks now have to be matched to the SCs. The $\eta$ and $\phi$ of the SC are defined as the energy-weighted $\eta$ and $\phi$ of all its constituent clusters. Currently at CMS, a BDT is used to match a GSF track to a SC. The BDT combines the information from tracks like kinematical properties, track-quality parameters, track-cluster matching variables, and supercluster information. The SC information includes the shower spread in $\eta$ and $\phi$ as well as transverse shower shape variables around the seed cluster.
For electrons reconstructed from tracker-driven seeds only BDT is used to decide whether to match a GSF track to a SC. In case of ECAL-driven electrons, candidates has to pass either the BDT requirements or the following two conditions :

- $|\Delta\eta| = |\eta_{SC} - \eta_{trk-in}| < 0.02$, where $\eta_{SC}$ is the SC $\eta$, and $\eta_{trk-in}$ refers to the $\eta$ of the GSF track closest to the SC and which is extrapolated to the ECAL from the innermost track position and direction
- $|\Delta\phi| = |\phi_{SC} - \phi_{trk-in}| < 0.15$, with analogous definition of $\phi_{SC}$ and $\phi_{trk-in}$. The wider window in $\phi$ accounts for the bending of electrons in the magnetic field and also to some extent the material effect.

## 3.3 PF algorithm

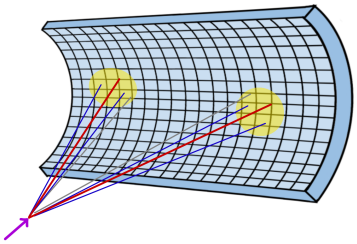
All the mustache SCs, ECAL crystals, KF tracks, GSF tracks, and all conversion flagged tracks are inputs to the PF algorithm. PF algorithm takes all this information together and gives the output as $e/\gamma$ objects. These $e/\gamma$ objects are made from refined superclusters.

### 3.3.1 Supercluster refinement in the ECAL

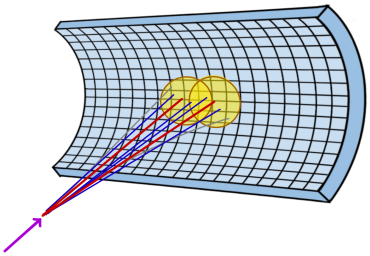
The mustache SC can then be further refined using information from GSF tracks. This further refinement tries to include additional bremsstrahlung photons that might have been missed by the mustache algorithm. One way tracks could help is, from each tracker layer, a tangent is drawn to the GSF tracks and extrapolated to the ECAL. If there is some PF cluster within some $\Delta\phi$, $\Delta\eta$ window of the track, it is considered a part of the refined supercluster. There are also conversion-finding algorithm that tries to associate tracks with photon conversions [12].

## 3.4 When does the reconstruction fail?

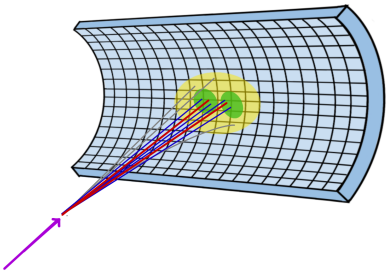
All the things explained above are the standard electron reconstruction algorithm used by CMS. But this fails when two electrons come very close to each other (typically $dR < 0.1$) or become merged. When two electrons are very close in the detector they have overlapping energy clusters as shown in Fig 3.4. As the clusters have such a large overlapping region, the SCs also overlap and from the ECAL deposits, the current reconstruct scheme reconstructs it as a single electron. These scenarios will be further discussed in Chapter 4.



(a) This is the scenario when the two electrons are widely separated and hence can be resolved as two separate electrons.



(b) In this scenario, the electrons overlap partially, so sometimes it could be resolved by the CMS reconstruction algorithm. This scenario can be referred to as semi-merged.



(c) Here the two electrons have very high degree of overlap in the ECAL and hence could not be resolved by the CMS reconstruction algorithm. Here the electrons are merged.

Figure 3.4: Three scenarios depicting the limitations of CMS reconstruction algorithm as the electrons come closer and closer and finally become merged.

# Chapter 4

# Analysis

Merged electrons refer to those objects that are reconstructed as one single electron by the PF algorithm, but in reality, there were two very close electrons. From now on they will be referred to as merged electrons. The genuine electrons, which are in reality single electrons and also reconstructed as single electrons, are referred to as single electrons only. Hence background for merged electrons are single electrons.

## 4.1 Workflow

All the analysis for this project has been performed using Analysis Object Data (AOD) file format. AOD formats give more handles and information to use which are required to reconstruct electrons. AOD file format requires a specialised CMS software environment which is called CMSSW. For this analysis, CMSSW_12_0_0 version has been used [15].
The workflow can be divided into two major parts for this analysis:

1. Generating fully simulated AOD samples as a source of single and merged electrons
2. Analysis of the simulated samples

The workflow in the pictorial form is shown in Fig 4.1.

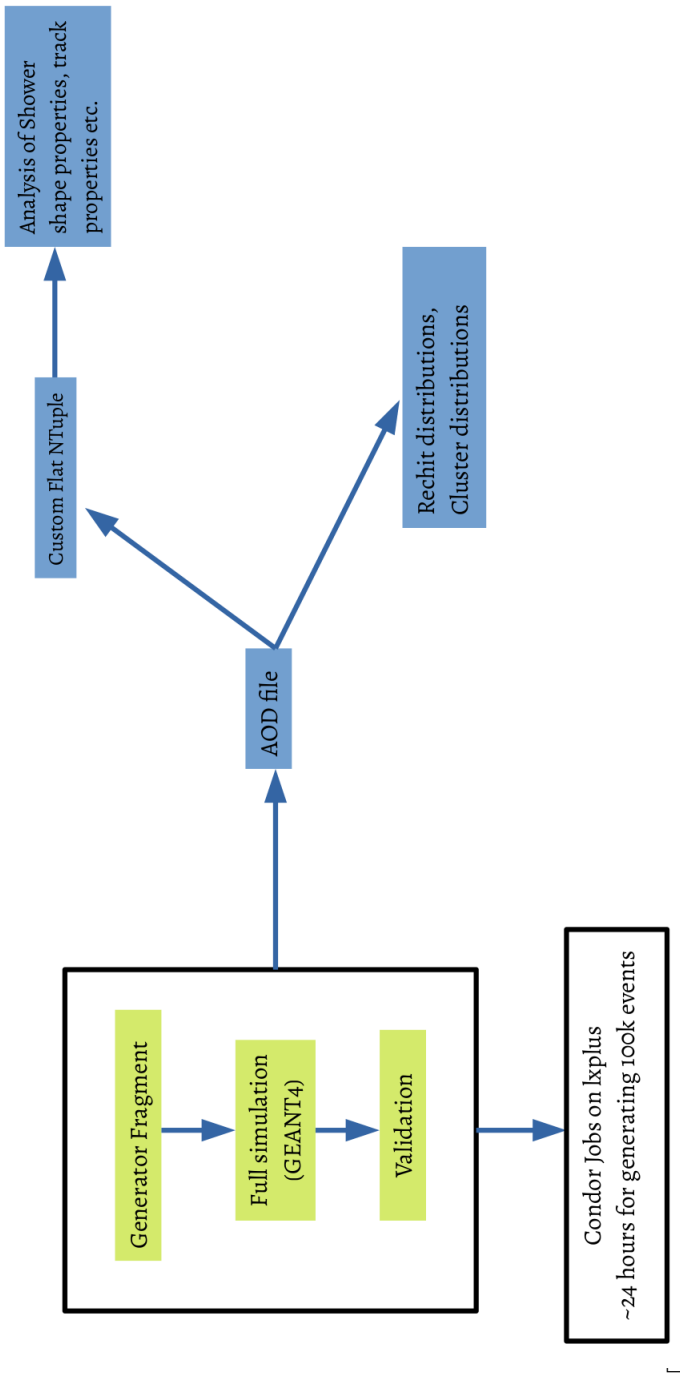



Figure 4.1: Workflow followed by me to generate the samples and analyze them.

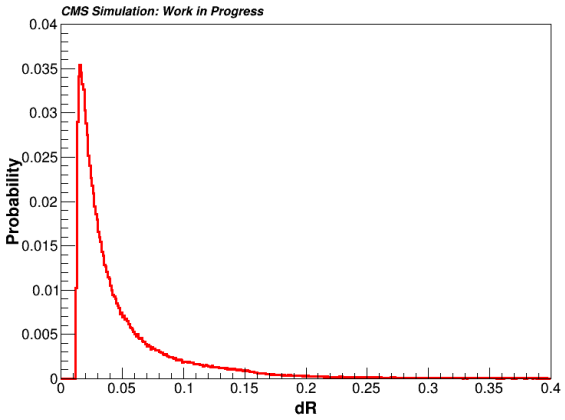
## 4.2 Source of merged and single electrons

Merged and single electrons are generated using a particle gun generator. For merged electrons boosted $J/\Psi \rightarrow e^+e^-$ sample is generated and for single electron $Z \rightarrow e^+e^-$ sample is generated. These generated samples are then run through the whole chain of full simulation in CMS to obtain reconstructed quantities. The table 4.1 gives the preliminary information about the samples.

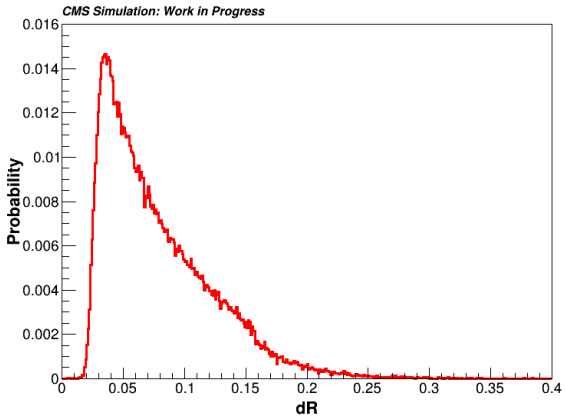
| | $J/\Psi \rightarrow e^+e^-$ | $Z/\gamma* \rightarrow e^+e^-$ |
|---|---|---|
| Number of events generated | 421000 | 444648 |
| $p_{\mathrm{T}}$-range of parent particle | 40 to 500 GeV | 40 to 300 GeV |
| events with 1 reconstructed electron | 307690 | 69458 |
| events with 2 reconstructed electrons | 101472 | 370391 |

Table 4.1: The above table gives some preliminary information on the simulated samples.

The $J/\Psi$ has a mass of around 3.1GeV, hence the $J/\Psi$'s are all highly boosted as they have a $p_{\mathrm{T}}$ of 40-500 GeV. Therefore the two electrons from the decay of $J/\Psi$ come very close to each other as shown by the $dR$ distributions Fig 4.2a and 4.2b. These plots show the $dR$ distributions between the two electrons from the decay of $J/\Psi$. All of these two very close electrons that are reconstructed as one electron by the current CMS electron reconstruction scheme will be called merged electrons and will be used in our analysis.
For a single electron source, I simulated the Drell-Yan events, $Z/\gamma* \rightarrow e^+e^-$ with a $Z$ $p_{\mathrm{T}}$ from 40 to 300 GeV. I only allowed the $Z$ boson to decay to electrons, to increase the statistics for the analysis. $Z$ boson mass is around 91GeV. The $p_{\mathrm{T}}$ window of 40 to 300 GeV allows some boost, but the electrons from the decay of $Z$ are still far apart as shown in Fig 4.3a and 4.3b. Hence in principle both the electrons could be used as single electrons.

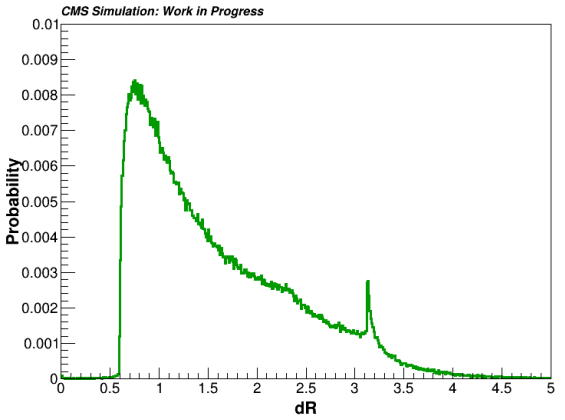



(a) $dR$ between generated electrons

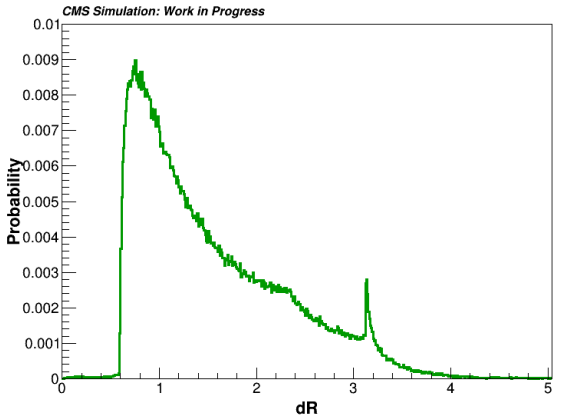



(b) $dR$ between reconstructed electrons

Figure 4.2: The above plots (a) and (b) show $dR$ distribution of generated and reconstructed electrons respectively, between the two electrons coming from the decay of $J/\Psi$. Clearly, most of the electrons lie within $dR < 0.1$ of each other. The fraction of electrons below $dR < 0.1$ is also reduced for the reconstructed scenario than generated. This will be more clearly shown in the pie charts Fig 4.4.




(a) $dR$ between generated electrons

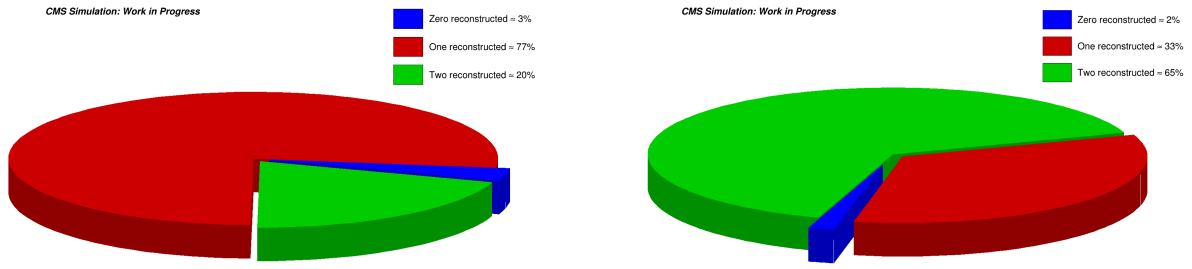



(b) $dR$ between reconstructed electrons

Figure 4.3: The above plots (a) and (b) show $dR$ distribution of generated and reconstructed electrons respectively, between the two electrons coming from the Drell-Yan process. Most of the electrons are far away from each other with a peak at around $dR \approx 1$.

For the $J/\Psi \to e^+e^-$ sample, in every event two electrons are generated in final states. The following pie charts Fig 4.4b and 4.4a, show the fraction of events where 0, 1, or 2 electrons are reconstructed for two $dR$ ranges, $dR < 0.1$ (merged) and $0.1 < dR < 0.2$

(reasonably separated). The $dR$ is between the two generated electrons.



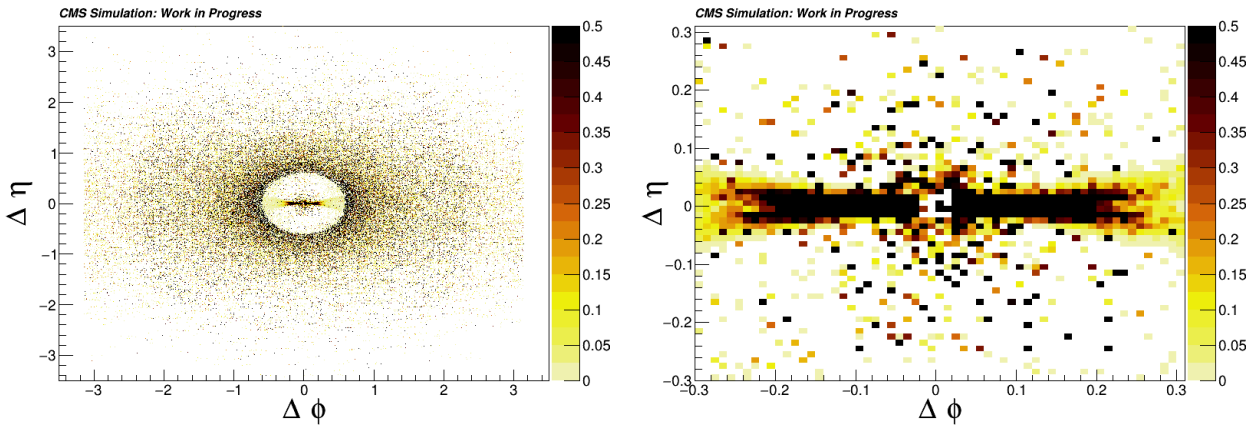

(a) $dR < 0.1$ (b) $0.1 < dR < 0.2$

Figure 4.4: The above pie charts show the number of reconstructed electrons for two $dR$ ranges. (a) $dR < 0.1$ and (b) $0.1 < dR < 0.2$. For $dR < 0.1$, almost in 77% of the events, only one electron is reconstructed, which shows the CMS reconstruction algorithm is not able to reconstruct very close-by electrons. In the range $0.1 < dR < 0.2$, in 65% of the events both the electrons are reconstructed, hence the reconstruction algorithm is performing better for high $dR$ values. $dR$ mentioned above is between the two generated electrons coming from the decay of $J/\Psi$.

## 4.3 ECAL cluster distribution of merged and single electrons

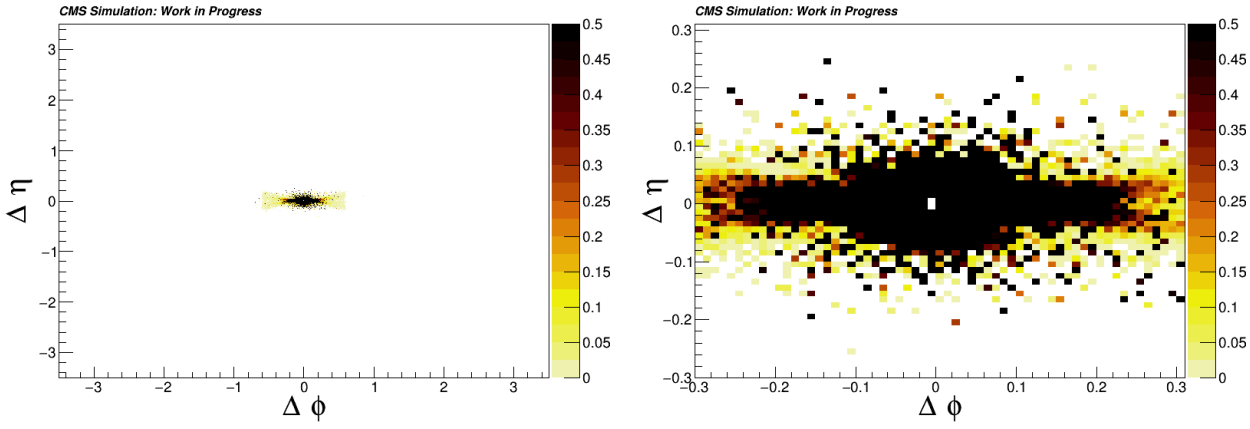
Here cluster refers to PF clusters. PF clusters as explained in 3.1, are objects made from individual ECAL crystals. ECAL cluster distribution refers to the statistical distribution of cluster hits around the seed cluster of the supercluster (SC), from which the electron is reconstructed. Hence seed cluster is always at the origin or $(0, 0)$. The distribution of other clusters around the seed cluster is plotted in the $\Delta\eta$-$\Delta\phi$ plane. The $\Delta\eta$ and $\Delta\phi$ are with respect to the seed cluster, therefore $\Delta\phi = \phi_{cluster} - \phi_{seedcluster}$ and similar definition for $\Delta\eta$. These clusters are weighted by the factor $\frac{E_{cluster}}{E_{seed}}$, where $E_{cluster}$ is the energy of the PF cluster and $E_{seed}$ is the energy of the seed cluster. In case of events with two or more electrons, the seed cluster associated with the highest energy electron is chosen as origin and the distribution of all the clusters around it is plotted.
For merged electrons, the cluster distribution looks a bit thicker than for single electrons. The plots for single and merged electrons cluster distribution are shown in Fig 4.6 and 4.5.

The circular kind of cluster distribution distribution for $Z \rightarrow e^+e^-$ is due to the $dR$ distribution peaking at around 1 Fig 4.3.




(a) (b)

Figure 4.5: In figure (a) Shows the Cluster distribution for the single electron around the seed cluster at $(0, 0)$ in the $\Delta\eta$-$\Delta\phi$ plane. (b) Shows the zoomed cluster distribution around $(0, 0)$. Each cluster is weighted by a factor of $E_{cluster}/E_{seedcluster}$.

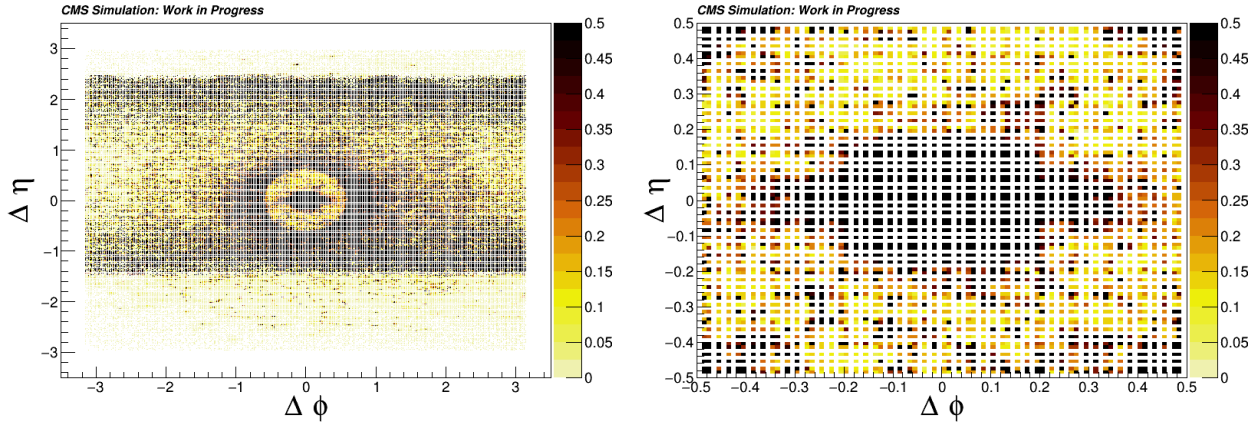



(a) (b)

Figure 4.6: In figure (a) Shows the Cluster distribution for a merged electron around the seed cluster at $(0, 0)$ in the $\Delta\eta$-$\Delta\phi$ plane. (b) Shows the zoomed cluster distribution around $(0, 0)$. Each cluster is weighted by a factor of $E_{cluster}/E_{seedcluster}$. The distributions for merged electrons are thicker than single electrons.

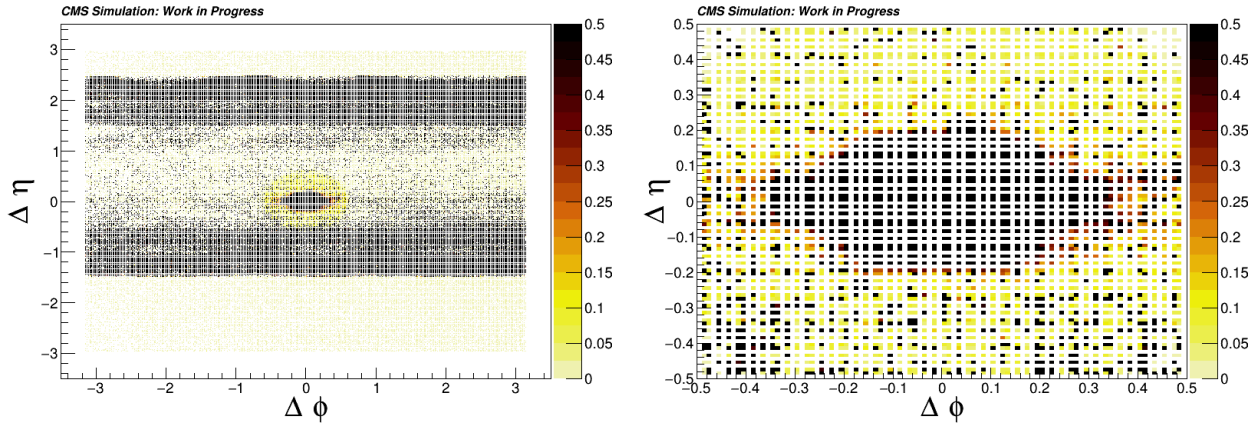
Similar distributions were made for individual ECAL crystal hits. Crystal hits give more

handle on lower-level information that could help to find limitations in the reconstruction algorithm that fails to reconstruct merged electrons. For crystal hit distributions, the origin is chosen as the seed crystal of the seed cluster of the supercluster from which the electron is reconstructed. The distribution is plotted in the $\Delta\eta$-$\Delta\phi$ plane where $\Delta\eta = \eta_{crystal} - \eta_{seedcrystal}$ and similar for $\Delta\phi$. A similar pattern has been observed like the cluster distributions, the distribution was thicker for merged electrons than single electrons. The Fig 4.8 and 4.7 shows the distribution. The circle kind of pattern is because of the same reason as was explained for cluster distributions.
The PF clustering smoothens out the effect of showering in ECAL crystals, hence the cluster distribution looks smoother and less noisy than the crystal distribution.




(a) (b)

Figure 4.7: In figure (a) Shows the rechit distribution for the single electron around the seed crystal at $(0, 0)$ in the $\Delta\eta$-$\Delta\phi$ plane. (b) Shows the zoomed crystal distribution around $(0, 0)$. Each crystal is weighted by a factor of $E_{crystal}/E_{seedcrystal}$.




(a) (b)

Figure 4.8: In figure (a) Shows the rechit distribution for a merged electron around the seed crystal at $(0, 0)$ in the $\Delta\eta$-$\Delta\phi$ plane. (b) Shows the zoomed crystal distribution around $(0, 0)$. Each crystal is weighted by a factor of $E_{crystal}/E_{seedcrystal}$. Again merged electrons have a bit thicker distribution.

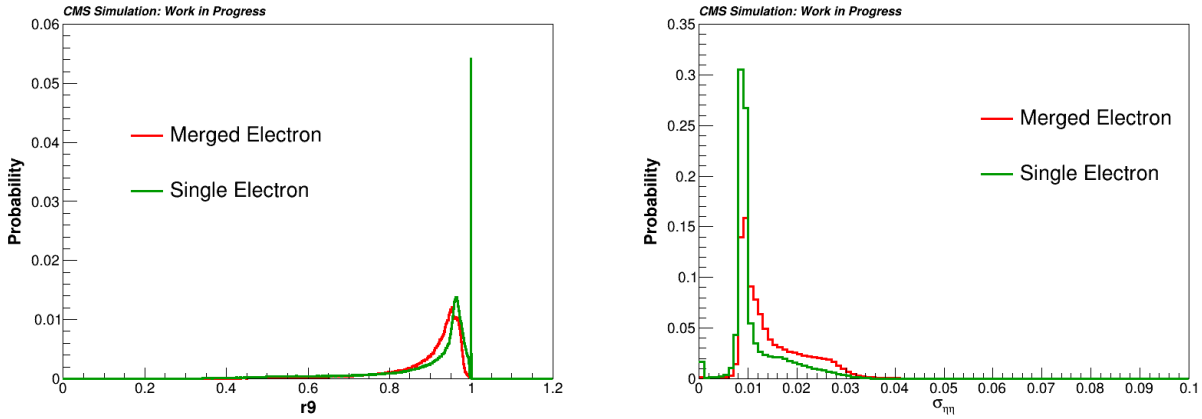
## 4.4 ECAL properties of electrons

Electrons deposit almost all their energies in the ECAL. They shower while depositing their energies. Hence, the energy is spread across several crystals, that's why clustering is done as discussed earlier. Therefore merged electrons should differ in shower shape from single electrons showers, as there are two electrons in merged electrons. Hence, from the energy distribution in the $\eta$-$\phi$ plane in the ECAL some analytic variables can be derived, and those could be used for separating merged and single electrons. Some of these variables are defined below:

- $r9 = \frac{E_{3\times3}}{E_{supercluster}}$, where $E_{3\times3}$ is the energy of the $3 \times 3$ crystal array, around the seed crystal of the seed cluster of the supercluster. $r9$ is a good variable to differentiate converted and unconverted photons. For merged and single electron it could be used also. It will work best if the two electrons are at least 3 crystals away.
- $\sigma_{\eta\eta} = \sqrt{\frac{\sum_i^{5\times5} w_i(\eta_i - \overline{\eta}_{5\times5})^2}{\sum_i^{5\times5} w_i}}$, where $w_i = 4.2 + \ln \frac{E_i}{E_{5\times5}}$. Here $i$ runs over all the crystals in the $5 \times 5$ crystal matrix. $\sigma_{\eta\eta}$ defines the shower width in the $\eta$ direction, which is largely unaffected by electrons and photons showers, and therefore its an important

variable in electron and photon identification. The definition refers to the local width of the shower, in $5 \times 5$ crystal matrix around the seed crystal of the seed cluster in the supercluster.

The overlayed plots of $r9$ and $\sigma_{\eta\eta}$ for merged and single electrons are shown in Fig 4.9.




(a) (b)

Figure 4.9: Fig above shows the distributions of (a) $r9$ and (b)$\sigma_{\eta\eta}$. The distributions are compared for merged and single electrons. They show differences as expected from the definitions of those quantities.

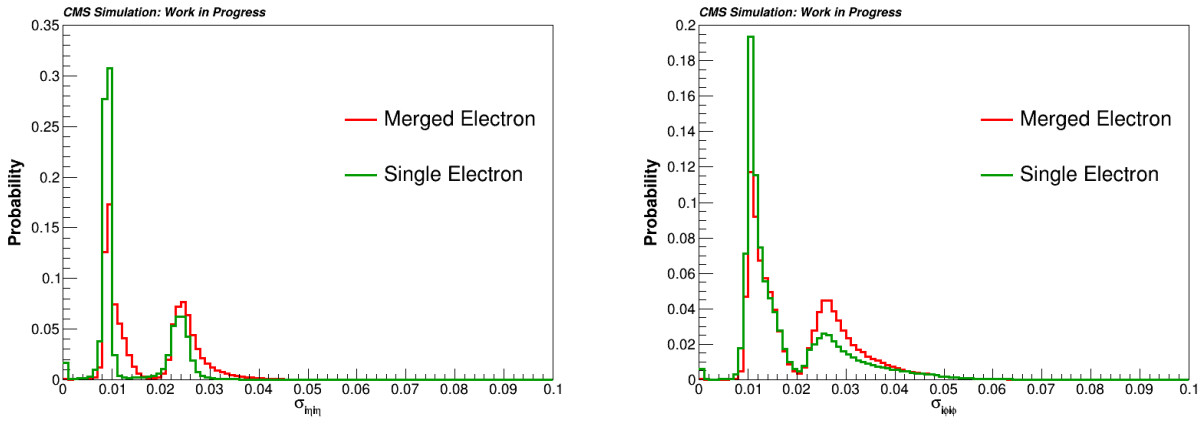
But $\sigma_{\eta\eta}$ suffers from some limitations like the crystal size in the endcap varies with cylindrical radius. Also, the intermodule spaces ( 6mm) change the crystal position more than it affects the shower shape. Hence, redefining variables in terms of crystal spacing can solve the problem. $\sigma_{\eta\eta}$ doesn't perform well near the cracks. Hence, new variables are defined which are based on units of crystal spacing rather than $\eta$ and $\phi$ spacings. These variables are $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ which are defined below:

- $\sigma_{i\eta i\eta} = \sqrt{\frac{\sum_i^{5\times5} w_i(\eta_i^{crys.nr} \times 0.0175 + \eta^{seedcrys} - \overline{\eta}_{5\times5})^2}{\sum_i^{5\times5} w_i}}$ where $w_i = 4.2 + \ln \frac{E_i}{E_{5\times5}}$, here $\eta_i^{crys.nr}$ denotes the crystal index i.e $i\eta$ with respect to the seed crystal index $i\eta_{seed}$. The number 0.0175 denotes the average crystal $\eta$ width. The average crystal $\eta - \phi$ width is $0.0175 \times 0.0175$.

- Similar definition of $\sigma_{i\phi i\phi}$ as $\sigma_{i\eta i\eta}$, with $\eta_i^{crys_{nr}}$ replaced by $\phi_i^{crys_{nr}}$ and $\eta_{seedcrys}$ with $\phi_{seedcrys}$. $\sigma_{i\phi i\phi}$ is only defined for ECAL barrel but $\sigma_{i\eta i\eta}$ is defined for both barrel and

endcap.

The plots for $\sigma_{i\eta i\eta}$ and $\sigma_{i\phi i\phi}$ are shown in Fig 4.10.




(a) (b)

Figure 4.10: Fig above shows the distributions of (a) $\sigma_{i\eta i\eta}$ and (b)$\sigma_{i\phi i\phi}$. Again the distributions for merged and single electrons and these quantities show some difference for merged and single electrons.

All these properties are based on ECAL information. As the source of single electrons are $Z \rightarrow e^+e^-$ sample and the source for merged electrons is the $J/\Psi \rightarrow e^+e^-$ sample, hence the electrons from the two samples has different $p_{\mathrm{T}}$ distribution. The variables defined above might have some dependence on $p_{\mathrm{T}}$. To remove the difference both the samples have been $p_{\mathrm{T}}$ reweighted i.e in different bins of $p_{\mathrm{T}}$, a factor is calculated to match the event numbers in that $p_{\mathrm{T}}$ bin. Once this factor is calculated in different $p_{\mathrm{T}}$ bins, it is applied as weights to all variables plotted. In this analysis, to be more sure that nothing else is affecting the distribution of these properties except the presence of two electrons in merged and one electron in a single electron object, the factor has been calculated in 2D bins of $p_{\mathrm{T}}$ and $\eta$.

The plots of these variables are given below, with merged and single electron overlayed. In all the plots $p_{\mathrm{T}} - \eta$ reweighted factor has been induced as weights while making all the Figs 4.9 and 4.10.

Another set of properties that could be used to differentiate merged and single electrons

are the energy calibrations. The sum of energies of all the constituent PF clusters of the supercluster is called the raw energy of the SC or $E_{raw}$. On this raw energy, calibrations and regressions are performed to get the corrected energy of the SC or $E_{corrected}$. For merged electrons, it is expected that the correction factor of $E_{raw}$ to $E_{corrected}$ is larger than for single electrons. hence, The quantity $\frac{E_{corrected}}{E_{raw}}$ should be larger for merged electrons than single electrons, which is as shown in Fig 4.11.

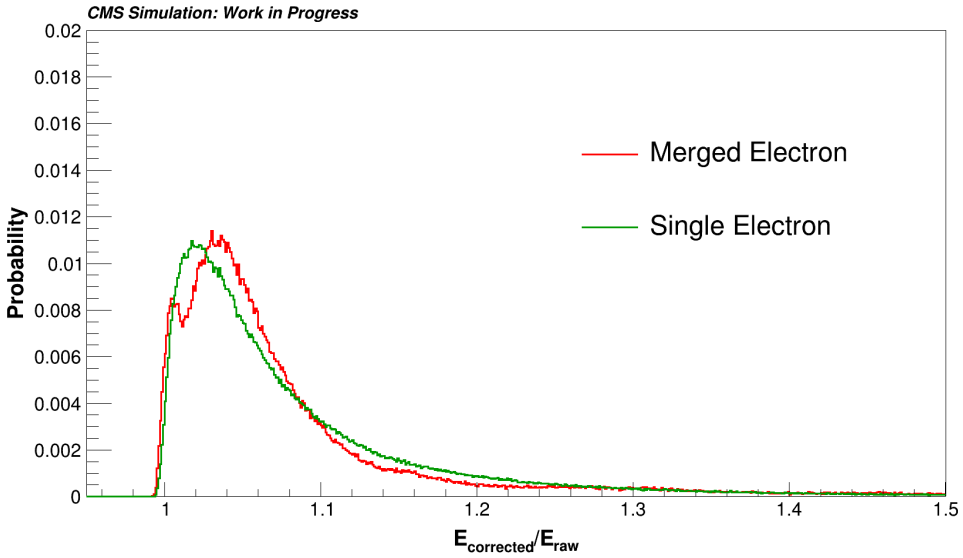
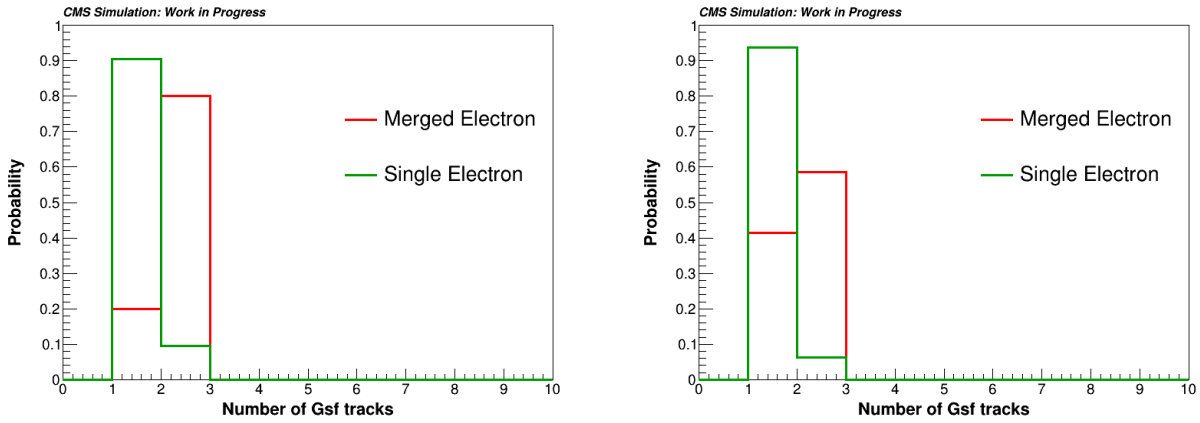



Figure 4.11: The above figure shows the distribution of $E_{corrected}/E_{raw}$ compared for merged and single electrons. The peak of the distribution slightly shifts to the right for merged electrons and hence the correction factor is more for merged electrons than single electrons.

## 4.5 Track properties of electrons

The previous section deals with only the calorimeter properties or the ECAL information of the electrons. Electrons, as explained in addition to calorimeter deposits have tracks. As described in Section 3.2.2 electrons are reconstructed using GSF tracking. For merged electrons, The seed GSF track is associated with the merged electron. The second GSF track which belonged to the second electron might have been deleted while the track was assigned to the supercluster for some reasons like, if it shares too many common hits with the seed GSF track or if it has a higher $\chi^2$ or was farther away from the SC position than the first GSF track. But for a single electron, only one GSF track is expected near the SC of the

reconstructed electron. The number of GSF tracks is plotted at $dR < 0.1$ and $dR < 0.05$ around the reconstructed seed GSF track for both merged and single electrons. The plots are given in Fig 4.12. The plot clearly shows for a single electron has a peak at 1, as only one GSF track is expected near the single electron, but for merged electrons, it peaks at 2, as there were two electrons but only one is reconstructed.




(a) (b)

Figure 4.12: Fig above shows the number of GSF tracks around the seed GSF track for (a) $dR < 0.1$ and (b)$dR < 0.05$, for single and merged electrons. For merged electrons, it peaks at two but for single electron, it peaks at one as expected.

In the sample the merged electrons are coming from $J/\Psi$, hence they should have the properties of $J/\Psi$ contained in them. The invariant mass plot of the two close GSF tracks ($dR < 0.1$) and with the fourth component of the four-vector set-up assumed as the mass of the electron, the mass peak of $J/\Psi$ is regained. The plot is shown in Fig 4.13, overlayed with the single electron plot. In single electrons also there could be some spurious GSF tracks within $dR < 0.1$, but these are just backgrounds as shown in the Fig 4.13.

Another interesting feature of these merged objects is their $p_{\mathrm{T}}$ reconstruction. The $p_{\mathrm{T}}$ of the seed GSF track i.e track from which the electron is reconstructed, and $p_{\mathrm{T}}$ of the reconstructed merged electron, have very different distributions. But the $p_{\mathrm{T}}$ distribution of a single electron matches very well with that of the reconstructed object as shown in the Fig 4.14.

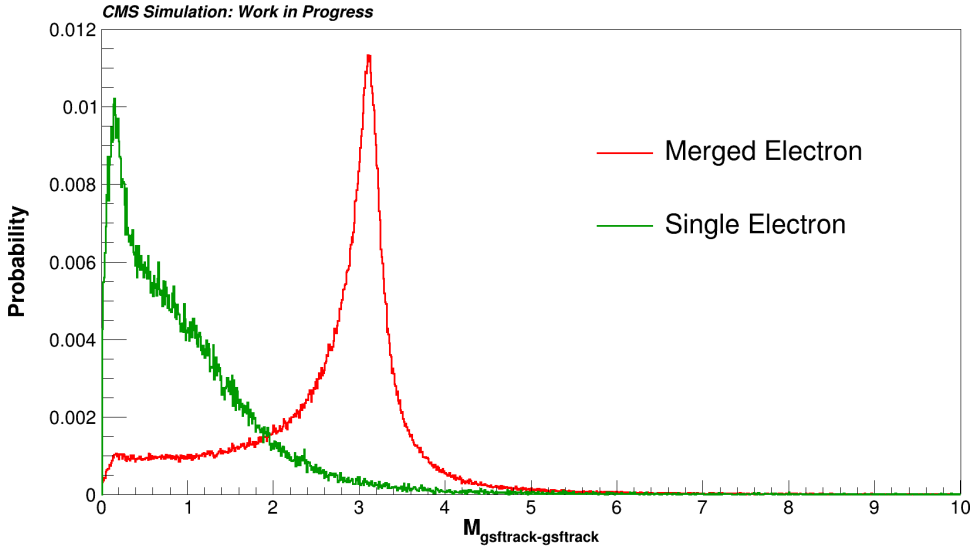
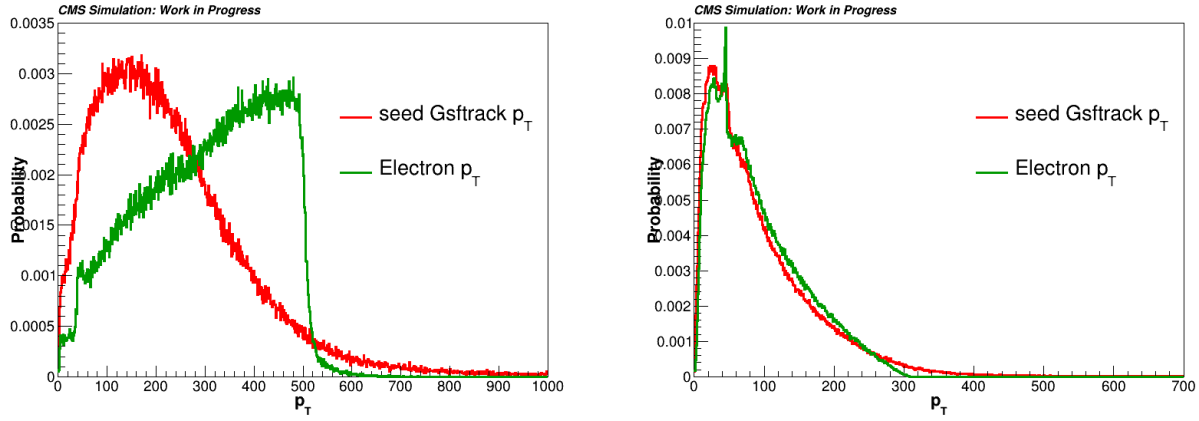



Figure 4.13: Invariant mass of the two very close GSF tracks within $dR < 0.1$ of the seed GSF track. For the merged electrons they peak at the $J/\Psi$ mass of around 3.1GeV, but for single electrons, they are just backgrounds.




(a) (b)

Figure 4.14: Figure above shows the $p_{\mathrm{T}}$ distributions of the seed GSF track and reconstructed electron for (a) merged (b) single electrons. The agreement between the two $p_{\mathrm{T}}$'s is great for single electrons, but it's worse for merged electrons.

A quantity that can be constructed out of $p_{\mathrm{T}}$ of tracks and electron is: $\frac{p_T(seedtrk)-p_T(electron)}{p_T(electron)}$, where $p_T(seedtrk)$ is the $p_{\mathrm{T}}$ of the seed GSF track or the GSF track from which electron was reconstructed and $p_T(electron)$ refers to the $p_{\mathrm{T}}$ of the reconstructed electron. For single

electrons, the agreement between the two $p_{\mathrm{T}}$ is good this quantity should peak at 0, and for merged electrons, it should not peak at 0 as the agreement is worse. The overlayed plot of merged and single electrons of this quantity is shown in Fig 4.15.

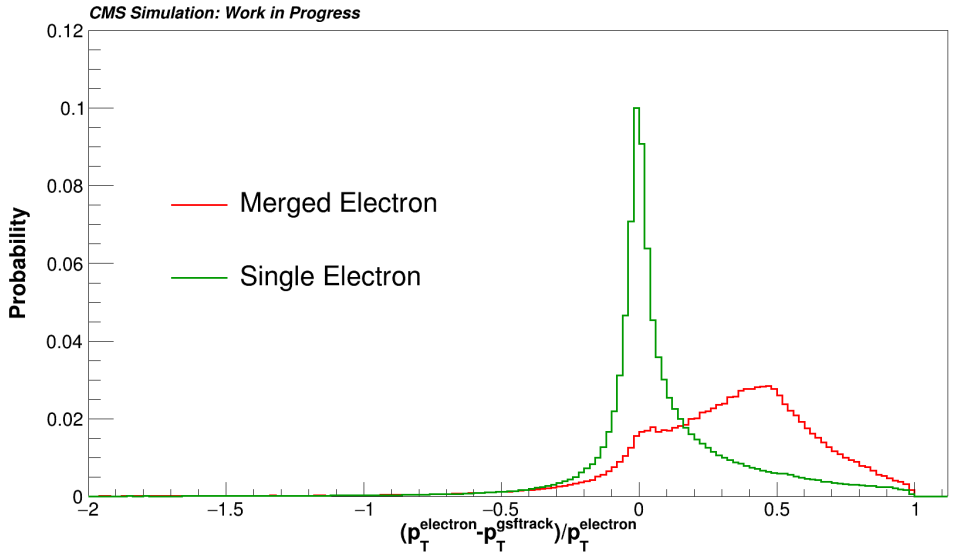



Figure 4.15: The figure shows the distribution of the quantity: $\frac{p_T(seedtrk)-p_T(electron)}{p_T(electron)}$, as described in the text. The single electron peaks at 0, but for merged it doesn't peak at 0.

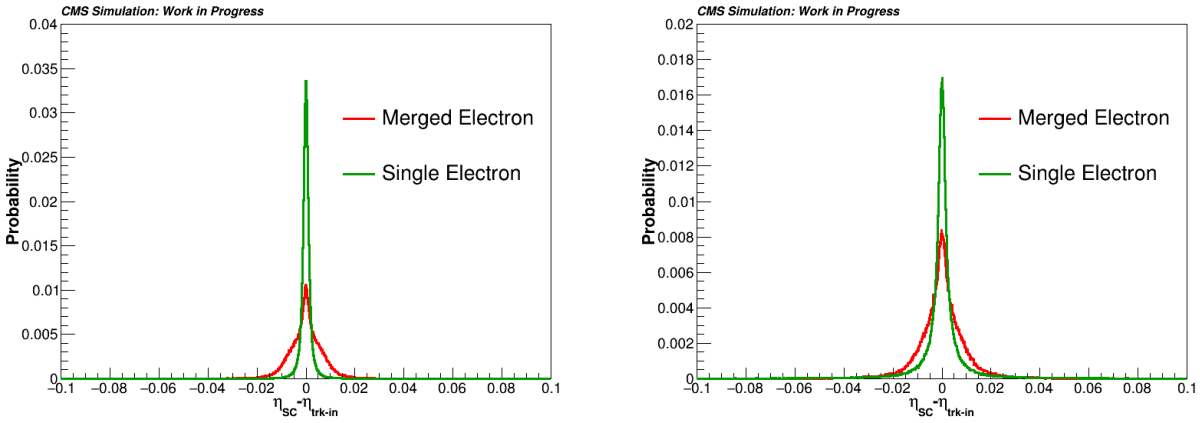
## 4.6 Hybrid properties of electrons

These hybrid properties of electrons refer to those properties that take into account both GSF tracks and ECAL SC information. These are mostly track-cluster matching variables as explained in Section 3.2.3. The important ones are:

- $\eta_{trk-in} - \eta_{SC}$, is the difference between $\eta_{trk-in}$ and $\eta_{SC}$, where $\eta_{trk-in}$ is the $\eta$ of the GSF track extrapolated to the ECAL from the inner tracker and $\eta_{SC}$ refer to the SC's $\eta$. This quantity is a track-cluster matching variable, which could be different for merged and single electrons due to the presence of two electrons for merged.
- $\phi_{trk-in} - \phi_{SC}$, similar definition as $\eta_{trk-in} - \eta_{SC}$.

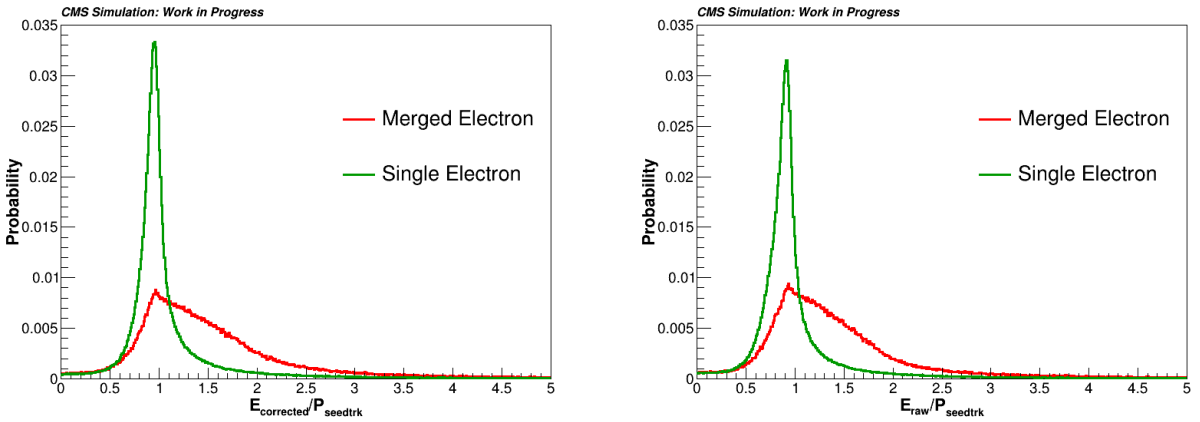
- $\frac{E_{raw}}{P_{seedtrk}}$ and $\frac{E_{corrected}}{P_{seedtrk}}$, where $P_{seedtrk}$ refers to the momentum of seed GSF track. This quantity should be approximately 1 for single electrons, but larger for merged electrons.

The overlayed plots for all the above properties are shown in Fig 4.16 and 4.17. All the figures show the result as expected.




(a) (b)

Figure 4.16: The figure above compares the (a)$\eta_{trk-in} - \eta_{SC}$ and (b)$\phi_{trk-in} - \phi_{SC}$, for merged and single electrons. The merged electrons have a wider width of the distributions than single electrons.




(a) (b)

Figure 4.17: Overlayed plots of (a)$E_{corrected}/P_{seedtrk}$ and (b)$E_{raw}/P_{seedtrk}$, are shown for merged and single electrons. The merged electrons have larger widths and hence larger values of $E/P_{track}$.

# Chapter 5

# Classification of electrons

All these properties that are discussed in Chapter 4, could potentially be used to separate merged from single electrons. In the above scenarios, it was easy to distinguish as merged electrons are compared to single electrons. As these are clean simulated samples, the difference could be captured by looking at them with the naked eye. But in data, there is no label associated with an electron. Hence by looking at the properties of this one reconstructed electron, it's very difficult to classify it as single or merged just based on observation by eye. However, from the previous analysis, it is well established that there is some correlation between the properties described earlier and whether the object is merged or a single electron. These correlations can be exploited by using multivariate analysis (MVA) techniques which take these properties as inputs and give the output as a classification between merged and single electrons. One such MVA technique is the neural networks which is explained below.

## 5.1 Neural Networks (NN)

Neural Networks (NN) are a class of Machine Learning (ML) algorithms or models, that work much like the human brains. The NN mimics the neurons, working together to process a given information, weigh the options, and finally arrive at a conclusion.
The NNs exploit the correlation and non-linear relation between inputs and outputs. The NN has a layered structure, with each layer having nodes, and the nodes of two consecutive layers are connected by an edge which may or may not have a weight associated with it. A

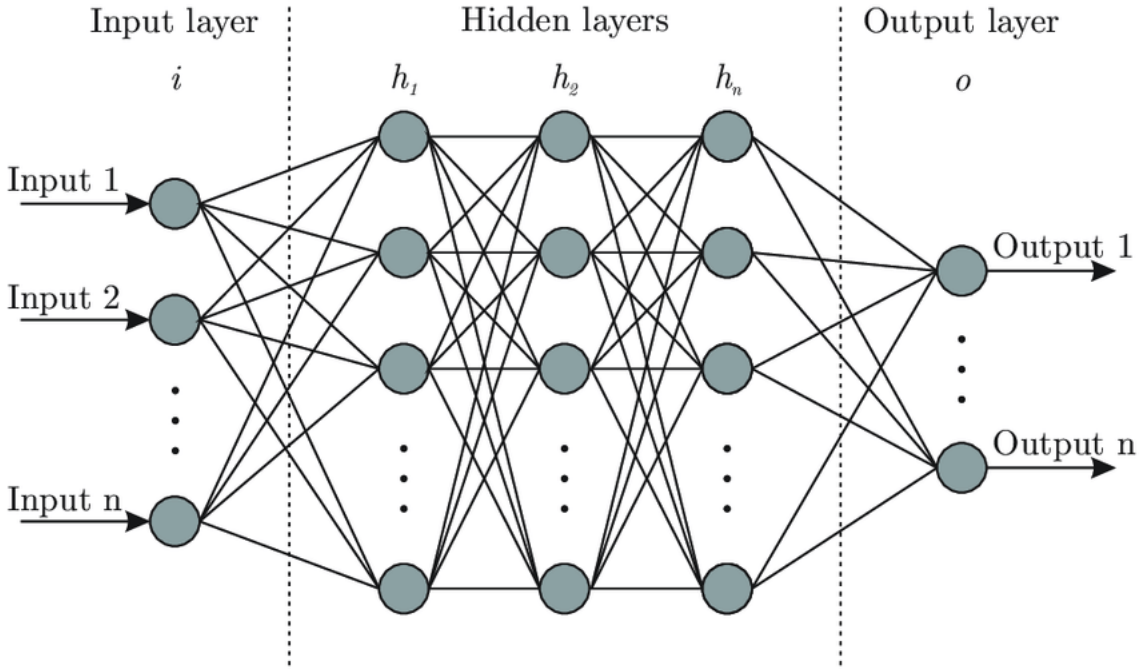



Figure 5.1: The basic NN architecture, showing the input layer, hidden layer/layers, and the output layer, along with the edges connecting the nodes. [4]

simplified NN structure is shown in Fig 5.1. From this figure, the first layer from the left is called the input layer, which takes the input information. The last layer is the output layer, where each node gives the output. In the case of classification, each output node gives the probability of the object being classified into a particular category. For a binary classifier, only one output node is present, as it's a binary decision. The other layers in the middle are called the hidden layers. As the number of hidden layers increases, it adds to the complexity of the NN. The nodes of a layer are connected to the nodes of the consecutive layer by edges which has a weight assigned to it. The weights that connect to the nodes, define the importance of the node and the weights along with the hidden layers define the correlation among different nodes or input features.

NNs could be classified into multiple categories based on the type of input it receives. The deep neural networks (DNN) are those whose inputs are analytic variables or real numbers. The convolution neural network (CNN) takes images as inputs. There are graph neural networks (GNN) whose inputs are graph-level features like vertex and edge features. These multiple categories just receive inputs in different forms but have a common aim, for example,

a classifier's job will be to classify objects into different categories. For this analysis, a binary classifier is needed to classify electrons into merged and single electrons.
The NNs fall under the category of supervised machine learning. Hence the NN is first trained on some known dataset where input and outputs are already known. By doing this NN learns the correlation between the inputs and how it affects the output. How well the NN is trained is defined by a loss function. The loss function can be any complicated expression but its main job is to calculate how different is the output of the NN from the actual output. During training, as the values of nodes in the input and output layers are fixed, the weights are tuned and varied so as to match the NN output to the actual output or reduce the loss function. These adjustments of weights are achieved using a method of back-propagation and gradient descent. After each iteration, the weights are adjusted. The performance of the NN is tested by the loss function. After many iterations, the loss function should decrease and then stabilize when the training is complete. After the NN is trained, it is validated on some datasets. In this validation step, the user knows the output, but the NN is tested on how well it can predict the output. If the NN performs well in validation, the NN could be used in the wild to predict outcomes of events that are unknown.

## 5.2 Output of neural networks

The NN could perform both classification and regression jobs. Classification as described in the previous section 5.1, classifies the objects into different categories giving probabilities of each object being how likely to fall into a particular category, whereas the regression could predict the value of a particular object based on the input feature. For example, a NN classifier to distinguish merged and single electrons will tell how likely an object is to be a merged or single electron. But a regressor would predict some property of the object, say given the energy and momentum of the object, the NN regression could try to predict the mass of the object.
For this analysis, a binary classifier is used and hence output properties of a binary classifier are explained in this section.
A binary classifier has only one output node and it gives a probability of whether an object should be classified with label A or label B (say). This the called the output NN score. The signals are always labeled 1 and the background is labeled 0. The signals are the objects that the analysis is trying to search or identify. In our case, the merged electrons are given

the label 1 and the single electrons the label 0. Hence, if the output of the NN is closer to 1 it means the object can be a potential signal and if it is closer to zero it resembles the background. Hence, an NN that could give better separation between signal and background events based on the NN score is preferable. An example is shown in Fig 5.2.

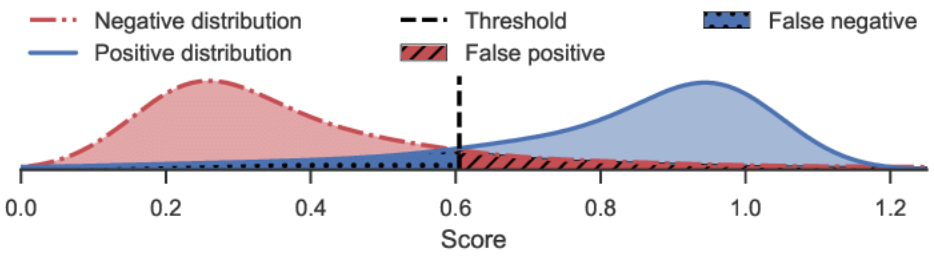



Figure 5.2: The figure shows a schematic NN score plot. The signal events have a higher NN score and hence peak to the right i.e. close to 1. The backgrounds on the other hand peak on the left i.e. close to 0. The figure also shows the true positive regions (TPR) and false positive regions (FPR). [5]

Another quantity that defines the performance of the NN output is the receiver operating characteristic (ROC) curve. The NN score for signal and background events gives a probability of how likely a particular event is to be a signal or background. For example, a particular event has a NN score of 0.6. It means it is with 60% confidence can be considered as a signal event, but still, it has a 40% chance of being a background. Therefore two quantities can be defined for a particular NN score (say $\alpha$), true positive rate (TPR) and false positive rate (FPR) as shown in Fig 5.2. These can be defined as :

- TPR $= \frac{TP}{TP+FP}$, where $TP$ is the number of signal events with NN score $> \alpha$
- FPR $= \frac{FP}{TP+FP}$, where $FP$ is the number of background events with NN score $> \alpha$

The plot of TPR vs FPR at different NN score values or working points is called the ROC curve. The area under the ROC curve is called the AUC score. For a perfect signal and background separation we expect an AUC score of 1. Hence higher the AUC score, the

better is the performance of the NN.

A sample ROC curve is shown in Fig 5.3. The figure shows different regions of ROC's performance as a classifier.

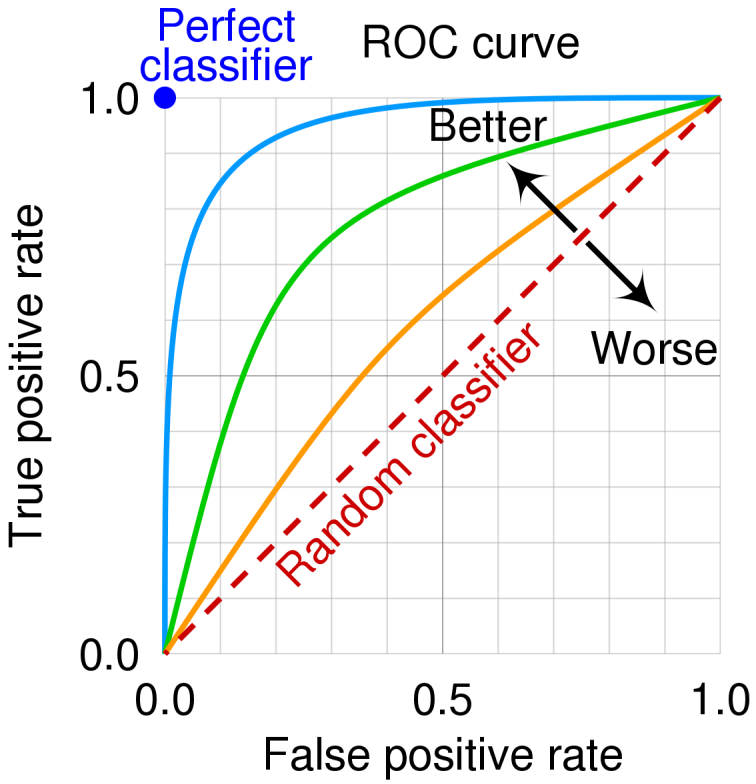



Figure 5.3: An example ROC curve. The curve regions of better and worse performance of an NN. An AUC of 0.5 corresponds to a diagonal along the ROC curve. It depicts a random NN classifier, with no separation in the NN score plot. As the curve convexs upwards from the diagonal, the NN performs better in categorization signals and backgrounds. A perfect classifier has an AUC score of 1. [6]

## 5.3 Training the merged electron classifier

As explained in the previous Section 5.2, an NN can be trained to differentiate between merged and single electrons. In this analysis, a DNN is trained with input variables being

real numbers. These input features include shower shape variables, track variables, and track-cluster matching variables as described in Chapter 4. The variables used as inputs to the NN are :

- Shower shape variables : $r9$, $\sigma_{\eta\eta}$, $\sigma_{i\eta i\eta}$, $\sigma_{i\phi i\phi}$
- Track variables : $\frac{p_T(seedtrk)-p_T(electron)}{p_T(electron)}$
- Hybrid variables : $\frac{E_{raw}}{P_{track}}$, $\frac{E_{corrected}}{P_{track}}$
- $\frac{H}{E}$, which is the ratio of HCAL energy over ECAL energy as explained in Section 3.2.1. This variable is used to differentiate QCD jets faking as electrons from real electrons and can be used as a handle for QCD backgrounds in data.

The NN is trained on merged electrons from $J/\Psi \rightarrow e^+e^-$ sample and single electrons from $Z \rightarrow e^+e^-$ sample. In this case signal is the merged electrons (label = 1) and the backgrounds are single electrons (label = 0).
Different NN architectures are tried and trained with different numbers of epochs. The NN could successfully differentiate merged and single electrons to a large extent. More on this is discussed in the next Chapter 6 on results.

# Chapter 6

# Results

Different properties of electrons like, track information, ECAL information, etc. are used as inputs to the NN. In the events, the merged electrons are labeled as signal (label = 1) and the background is the single electrons (label = 0). The NN gives good separation between signal and background events as shown in Fig 6.1.

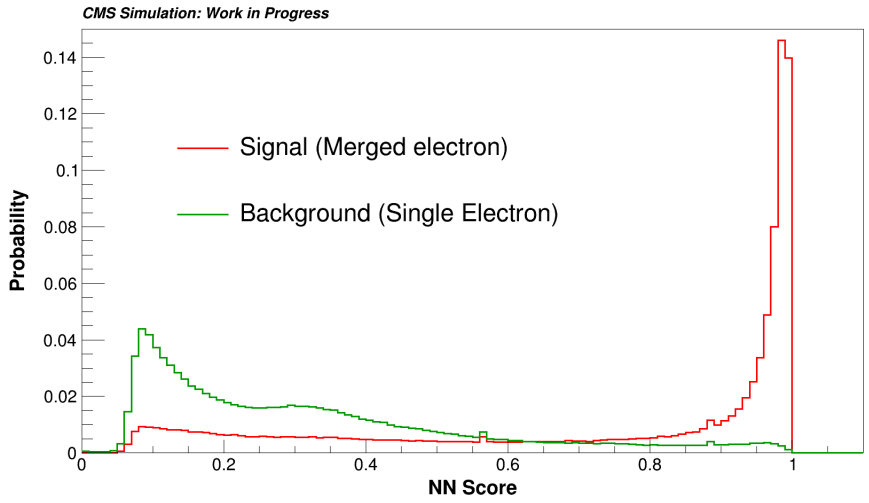



Figure 6.1: The above figure shows the NN score of the NN classifier, which classifies merged electrons as signals and single electrons as backgrounds. Hence the merged electron peaks towards 1 (right) and the single electron towards 0 (left)

The ROC curve for the classifier is shown in Fig 6.2. For both testing and training datasets, the AUC score is around 0.85. Hence, it is shown that merged and single electrons could be separated by a DNN in simulated samples.

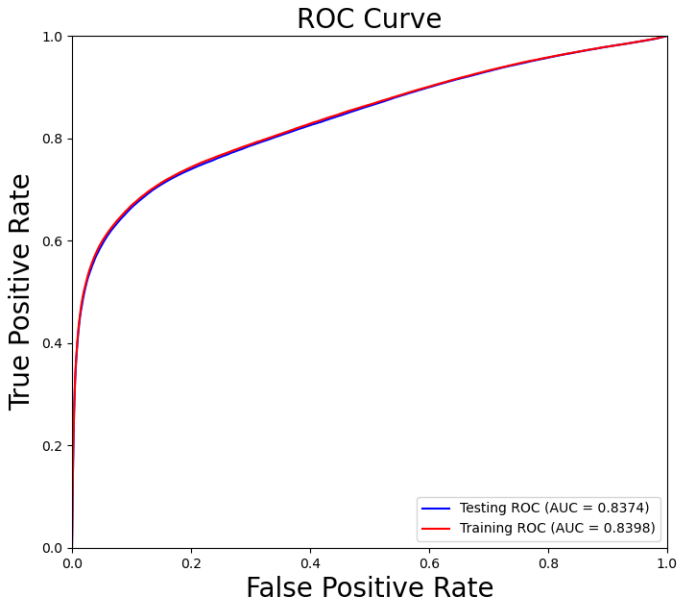



Figure 6.2: The ROC curve for the NN classifier used to classify merged and single electrons. The ROC for the training and testing datasets are very close giving an AUC score of 0.85 for both the training and testing datasets

From the above Fig 6.2 and 6.1, clearly the DNN has trained to separate the merged and single electrons. The next plan of the project is to test the DNN on some samples it hasn't been trained on. For this, the study of right-handed neutrino samples (RHN) is currently ongoing. As mentioned in Chapter , low-mass boosted RHN could be a source of these kinds of merged electron signatures. In this part of the study, a low-mass electron-philic RHN sample is chosen ($M_{RHN} = 2$ GeV), hence it will only couple to electrons. In this sample, the electrons are expected to show merged signatures in the detector. Hence testing the DNN on these merged electrons could prove to be a significant test of the DNN. That's the work in progress.

# Chapter 7

# Conclusion

The full chain of CMS reconstruction algorithm for electrons and photons is studied. Currently, the mustache superclustering algorithm used by CMS for clustering PF clusters in the ECAL to reconstruct electrons, cannot reconstruct very close-by electrons or merged electrons as separate objects but it reconstructs them as one electron. These shortcomings of the mustache algorithm could be overcome by carefully analyzing cluster deposit distribution along with the GSF track information could be used to separate merged and single electrons.

Many analytic properties of the electrons are studied, which include ECAl properties like $\sigma_{\eta\eta}$, $\sigma_{i\eta i\eta}$, etc., some track properties like the number of GSF tracks around electrons and also track and ECAL hybrid properties. These properties are carefully chosen to reflect the difference between merged and single electrons. These properties are then fed to a deep neural network (DNN).

It is seen that a DNN that is trained to differentiate single vs merged electrons, based on the above-mentioned properties shows a good separation of signal (merged) vs background (single).

The next goal of this work is to test the DNN on other sources of merged electrons like the close-by electrons from the RHN sample. Testing the performance of the DNN on the RHN sample shows how well the DNN has been trained. The work is still in progress.

Testing this NN on a $e/\gamma$ dataset could be another interesting check. If it could identify merged electrons and from that reconstruct $J/\Psi$ or $\Upsilon$ mass peaks. Similar studies have been done for $\pi^0 \to \gamma\gamma$, where the two photons were merged.

Another goal is to improve the performance of the NN. Convolution Neural Networks (CNN) gives a promising prospect by giving it input as ECAL energy deposit's images and training it to differentiate single and merged electrons. Hence training a CNN is another possible aim of this project.

# Bibliography


[1] Alexandros Tsagkaropolulos Izaak Neutelings. CMS coordinate system. https://tikz.net/axis3d_cms/.

[2] GL Bayatian, S Chatrchyan, G Hmayakyan, AM Sirunyan, W Adam, T Bergauer, M Dragicevic, J Eroe, M Friedl, R Fruehwirth, et al. CMS physics: Technical design report. 2006.

[3] Davide Valsecchi, CMS Collaboration, et al. Deep learning techniques for energy clustering in the cms ecal. In *Journal of Physics: Conference Series*, volume 2438, page 012077. IOP Publishing, 2023.

[4] Lavanya Shukla. Designing your neural networks. https://www.kdnuggets.com/2019/11/designing-neural-networks.html.

[5] Denis dos Reis, André Maletzke, Everton Cherman, and Gustavo Batista. One-class quantification. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2018, Dublin, Ireland, September 10–14, 2018, Proceedings, Part I 18*, pages 273–289. Springer, 2019.

[6] MartinThoma cmglee. Receiver operating characteristic (roc). https://commons.wikimedia.org/wiki/File:Roc_curve.svg.

[7] DV Forero, M Tortola, and JWF Valle. Global status of neutrino oscillation parameters after neutrino-2012. *Physical Review D*, 86(7):073012, 2012.

[8] Manimala Mitra, Richard Ruiz, Darren J Scott, and Michael Spannowsky. Neutrino jets from high-mass w r gauge bosons in tev-scale left-right symmetric models. *Physical Review D*, 94(9):095016, 2016.

[9] Sourabh Dube, Divya Gadkari, and Arun M Thalapillil. Lepton jets and low-mass sterile neutrinos at hadron colliders. *Physical Review D*, 96(5):055031, 2017.

[10] CMS Collaboration. Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 6:081, 2013.

[11] CMS collaboaration. CMS detector. https://cms.cern/detector.

[12] The CMS collaboration. Electron and photon reconstruction and identification with the cms experiment at the cern lhc. *Journal of Instrumentation*, 16(05):P05014, may 2021.

[13] CMS collaboration. Electron and photon performance in cms in run2 and prospects for run3. https://indico.particle.mephi.ru/event/35/contributions/2384/attachments/1109/1586/EGM_confTalk.pdf.

[14] CMS collaboration. 02/23/12 1 photon reconstruction and performance in atlas and cms. https://indico.in2p3.fr/event/6838/contributions/39722/attachments/32079/39620/EPetit_Photons_ATLAS_CMS.pdf.

[15] CMS collaboration. CMSSW application framework. https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookCMSSWFramework.