# Electron identification in HF sub-detector of CMS

Sai Krishna

Mentor : Saranya Samik Ghosh

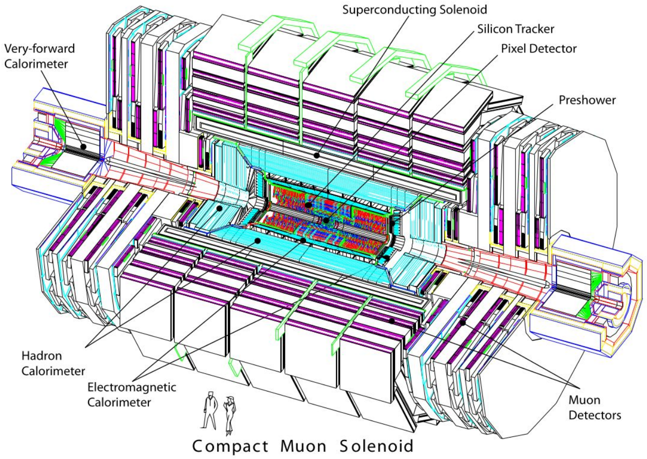
भारतीय प्रौद्योगिकी संस्थान है
Indian Institute of Technology I

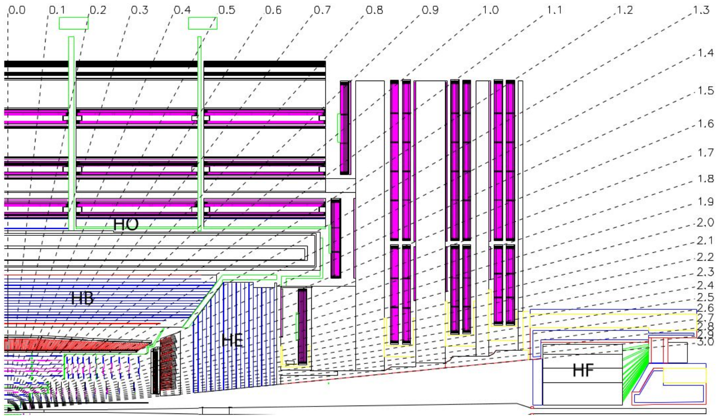
HSF-INDIA

# Introduction

- The Compact Muon Solenoid (CMS) is a general-purpose detector at the Large Hadron Collider (LHC).
- Conventional analyses focus on electrons within $-2.5 < \eta < 2.5$, while our study extends electron reconstruction in Hadronic Forward (HF) region, covering $-5 < \eta < 5$.

- By including HF coverage into electron detection, physics measurements can utilize a wider range of accepted electrons.



# The Forward Hadron Calorimeter (HF)

- The HF detector covers the region $2.8 < |\eta| < 5.2$. of the CMS experiment.
- The HF detector consists of quartz fibers of two lengths, arranged longitudinally along the detector.
- Long fibers (1.65 m) originate at the detector's face, while short fibers (1.43 m) begin 22 cm from the face.
- The HF detector primarily detects Cerenkov radiation emitted from particle showers to determine energy levels.

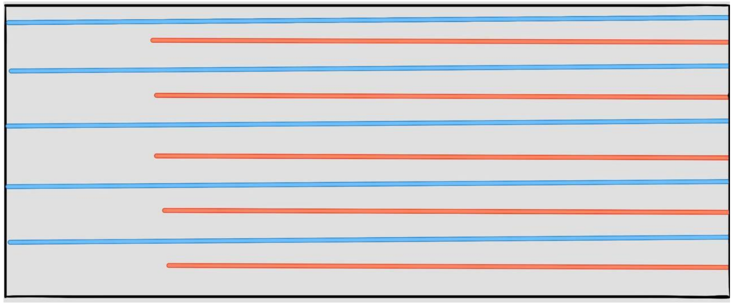


Fig : visual representation of long and short fibers in HF

भारतीय प्रौद्योगिकी संस्थान है
Indian Institute of Technology I

HSF-INDIA

# HF Cluster variables

- eS/eL is ratio of energy in the short fiber to energy in the long fiber of the HF Cluster.
- E9e25 is the ratio of long fiber energy of the 3x3 cluster and that of the 5x5 cluster (e9/e25).
- eCe9 is ratio of long fiber energy in core to energy within the 3x3 cluster (EC/E9).
- These shower shape variables tells us the characteristics of cluster like how spread the cluster or how narrow the cluster is.
- Other HF_Cluster variables:
  - HFEMClust_e1x1 HFEMClust_eLong1x1 HFEMClust_eShort1x1
  - HFEMClust_e3x3 HFEMClust_eLong3x3 HFEMClust_eShort3x3
  - HFEMClust_e5x5 HFEMClust_eLong5x5 HFEMClust_eShort5x5

Fig : Cross-section of HF

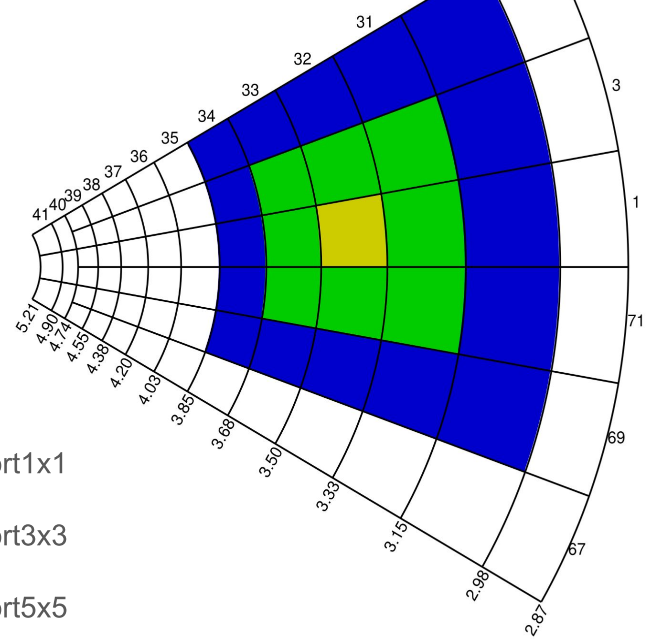
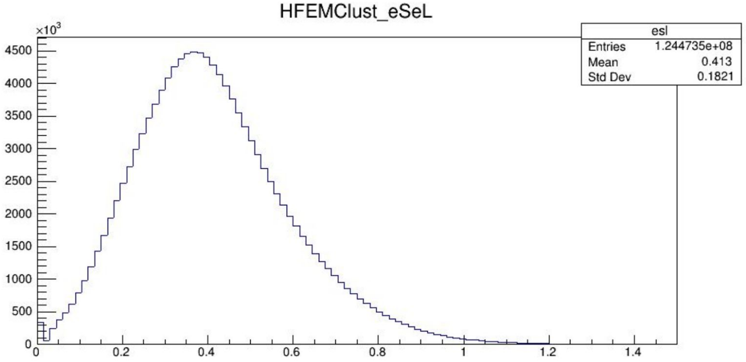
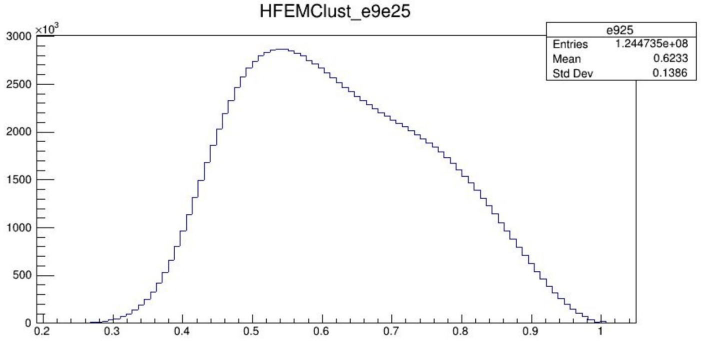
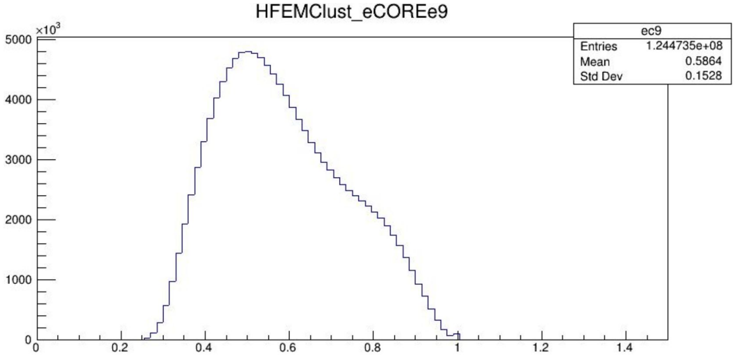



Image: Kevin H. Klapoetke, PhD thesis, 2014.

# HF Cluster shower shape variables



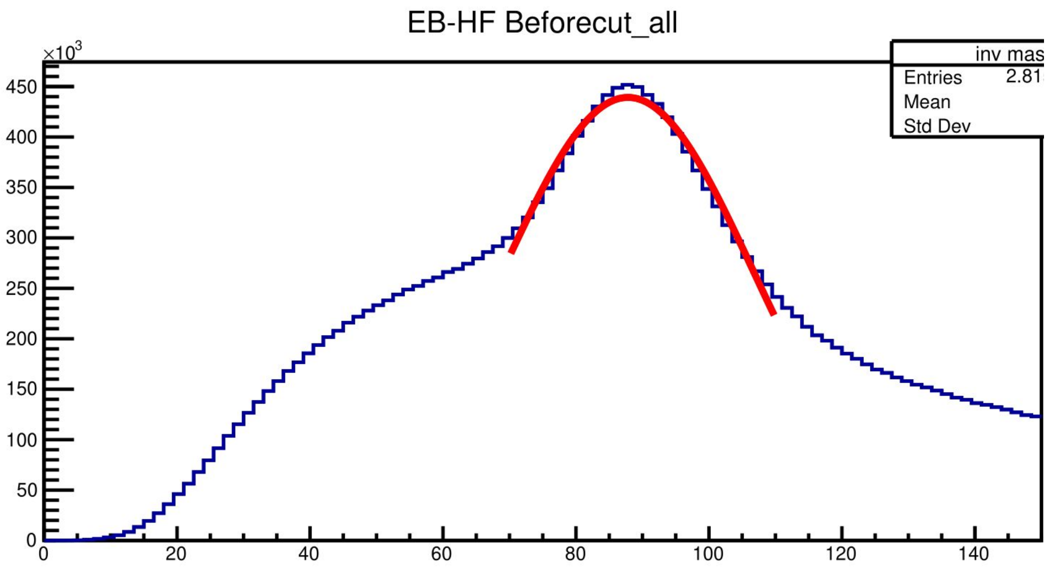

- We are working with Z ->ee and jets MC events.

भारतीय प्रौद्योगिकी संस्थान है
Indian Institute of Technology I

HSF-INDIA

# EB-HF :

- We are worked on Z ->ee and jets events.
- We did invariant mass with one EB (ECAL Barrel electron) and HF Cluster.
- The mean of gaussian fit is nearer to Z Mass (~91)
- It is basically a tag and probe method.




| mean | sigma |
|---|---|
| 87.8167 | 18.9102 |

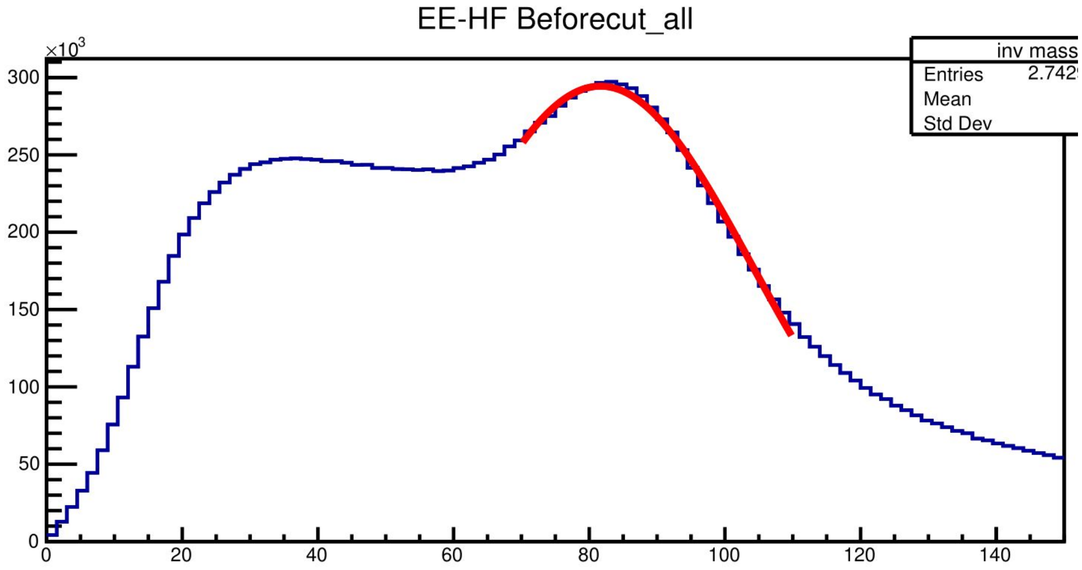
Indian Institute of Technology I

HSF-INDIA

# EE-HF:

- Similarly, We did invariant mass with one EE (ECAL Endcap electron) and HF Cluster.

- To further improve it, we apply cuts on HF variables .

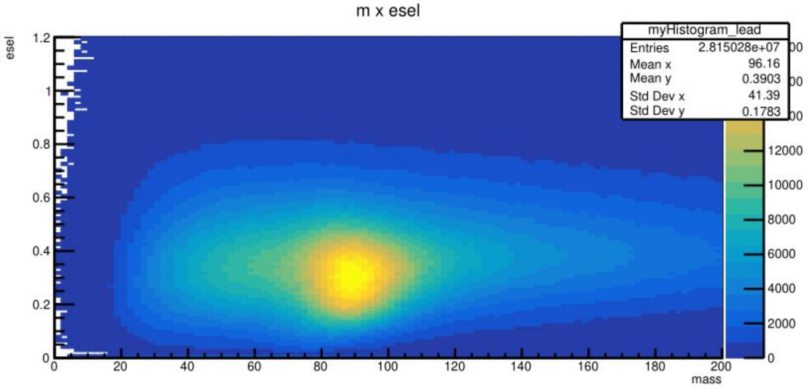


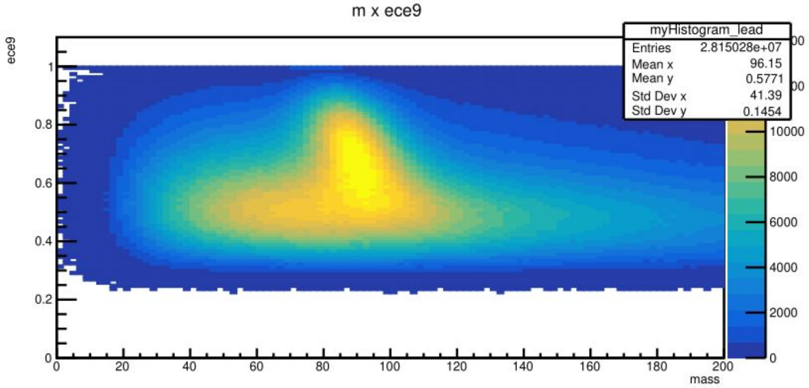

| mean | sigma |
|---|---|
| 81.6610 | 22.3494 |

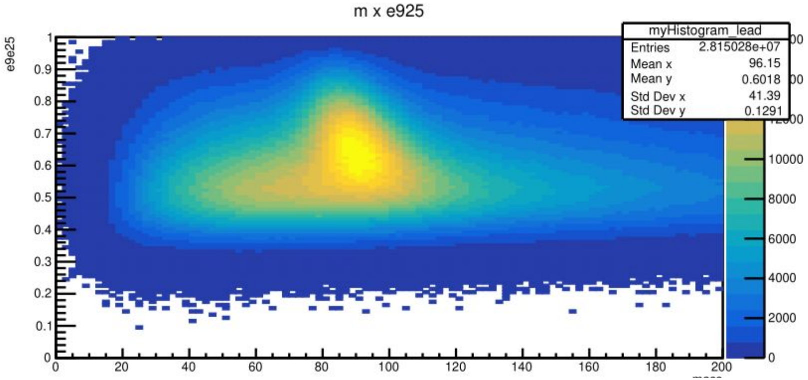
Indian Institute of Technology

HSF-INDIA

# EB-HF 2D Histograms



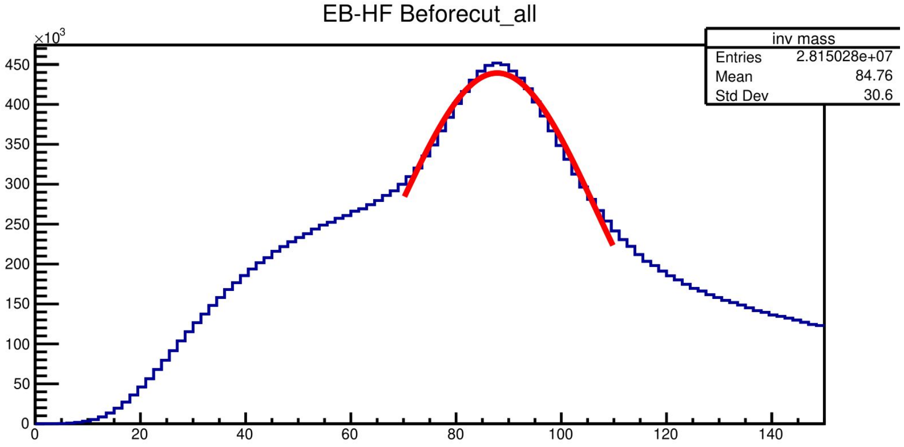



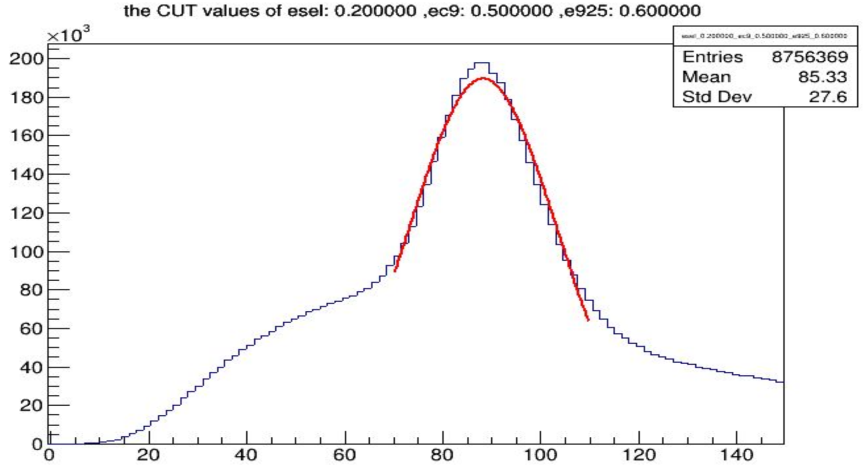



# EB-HF cuts

| esel | ece9 | e9e25 | mean | sigma |
|---|---|---|---|---|
| 0.3 | 0.5 | 0.6 | 87.71 | 16.11 |
| 0.2 | 0.5 | 0.6 | 88.19 | 14.66 |
| 0.3 | 0.35 | 0.4 | 88.19 | 21.74 |
| 0.2 | 0.45 | 0.6 | 88.23 | 15.21 |
| 0.2 | 0.5 | 0.55 | 88.52 | 15.34 |
| 0.2 | 0.45 | 0.5 | 88.70 | 16.96 |

# Before and After Applying Cuts



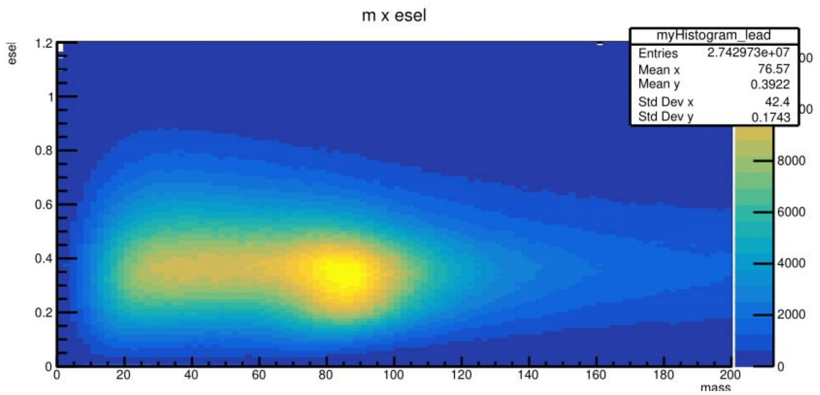



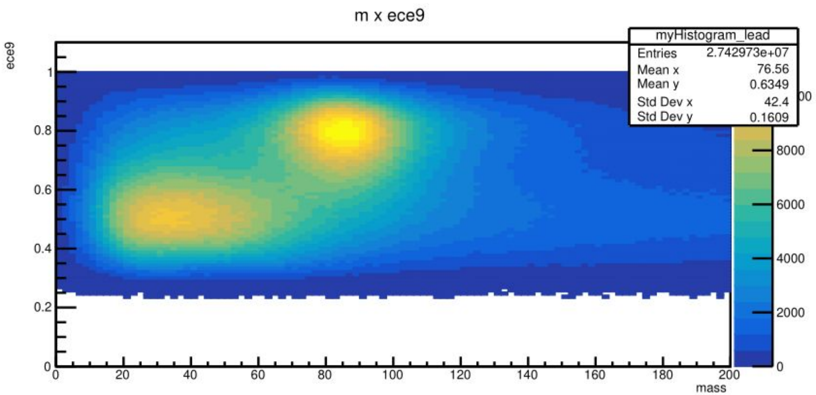

| mean | sigma |
|---|---|
| 87.8167 | 18.9102 |

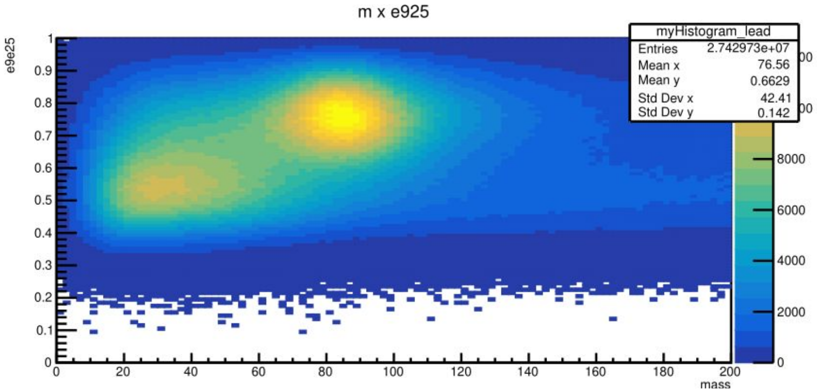
| mean | sigma |
|---|---|
| 88.19 | 14.66 |

HSF-INDIA

# EE-HF 2D Histograms




भारतीय प्रौद्योगिकी संस्थान है
Indian Institute of Technology I

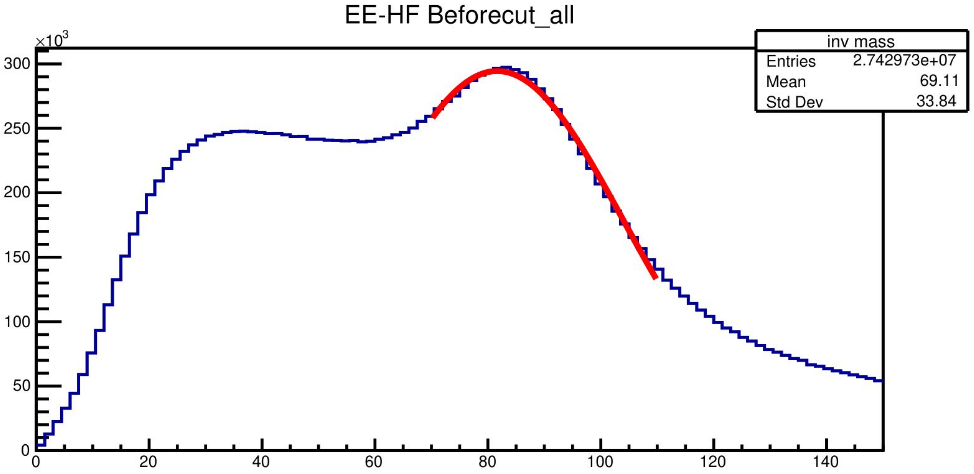
HSF-INDIA

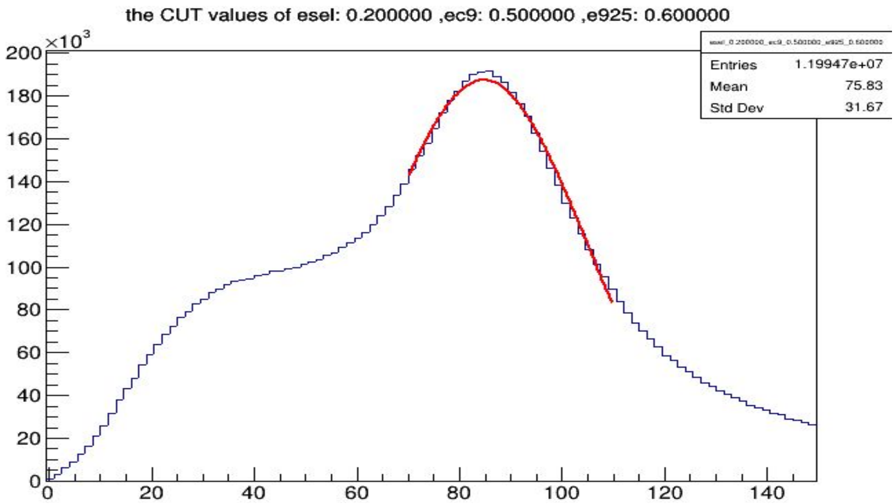
# EE-HF

- We got cuts at same combinations for EEHF and EBHF

| esel | ece9 | e9e25 | mean | sigma |
|---|---|---|---|---|
| 0.3 | 0.3 | 0.4 | 79.96 | 24.65 |
| 0.3 | 0.35 | 0.4 | 80.22 | 24.45 |
| 0.3 | 0.4 | 0.6 | 83.20 | 21.60 |
| 0.25 | 0.45 | 0.55 | 83.64 | 21.10 |
| 0.2 | 0.4 | 0.6 | 84.23 | 20.28 |
| 0.2 | 0.5 | 0.6 | 84.67 | 19.71 |

# Before and After Applying Cuts



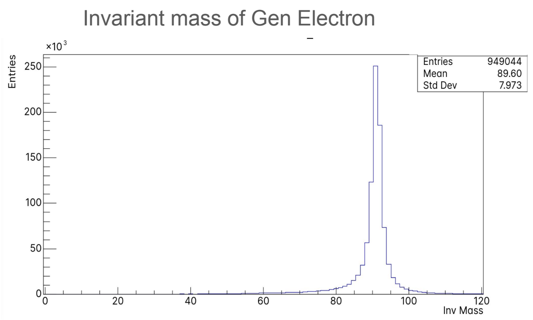




| mean | sigma |
|---|---|
| 81.6610 | 22.3494 |

| mean | sigma |
|---|---|
| 84.67 | 19.71 |

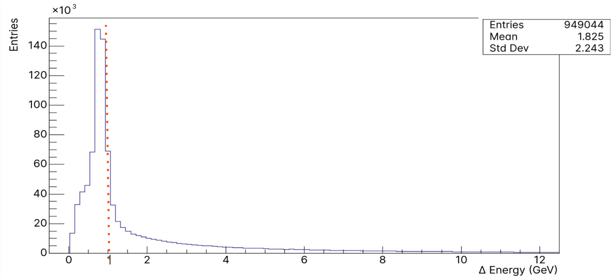
# HF Cluster energy

- Reason for the deviation in invariant mass is that HF's energy is not addingup energy loss it had undergone by the time it reach HF region.
- So, now we use Gen_Electron energy, which refers to the generator-level energy of the electron in simulations, representing its ideal energy before any detector effects are applied.



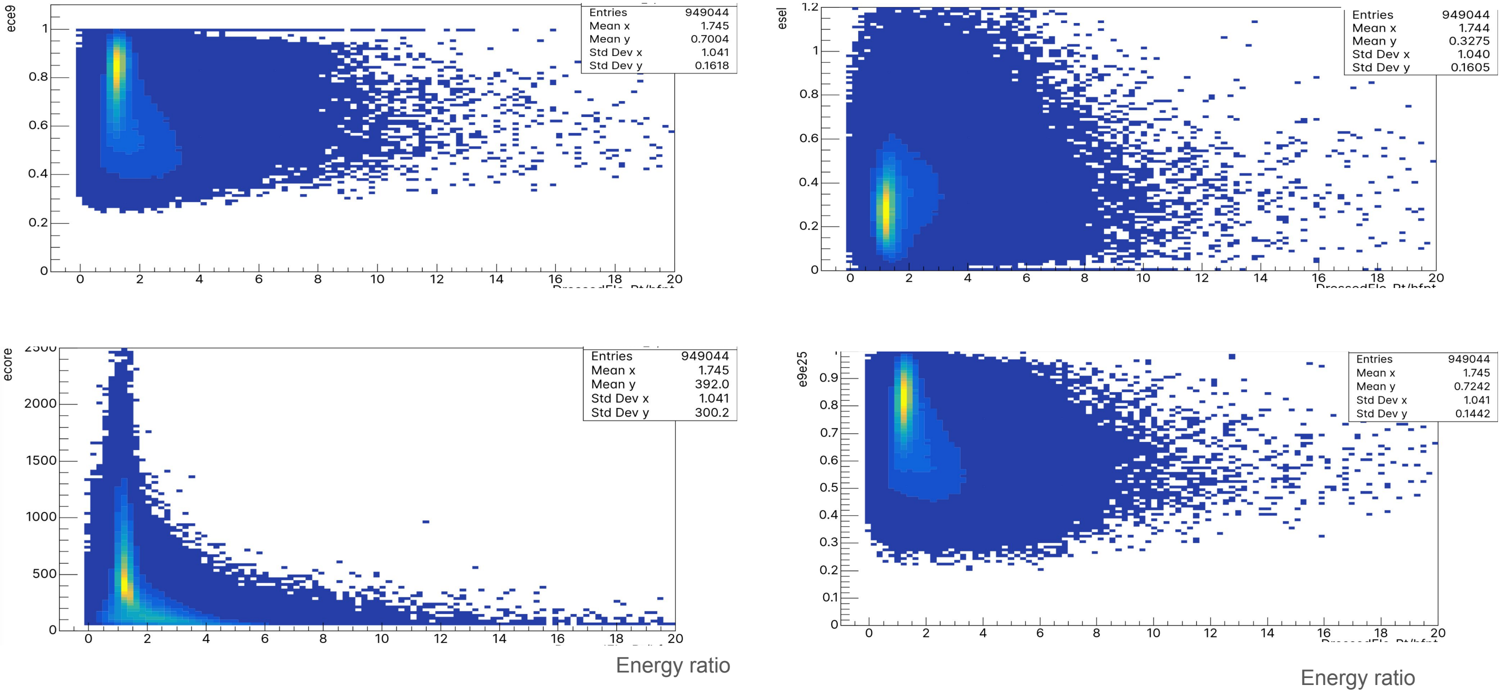

- If the HF_Cluster truly corresponds to an Gen Electron, the ratio of their energies should be around 1.

$$\text{Energy Ratio} = \frac{\text{HF cluster energy}}{\text{Gen electron energy}}$$



# Energy Ratio v/s HF variables

- These 2d histograms show the relationship between energy and hf cluster variables



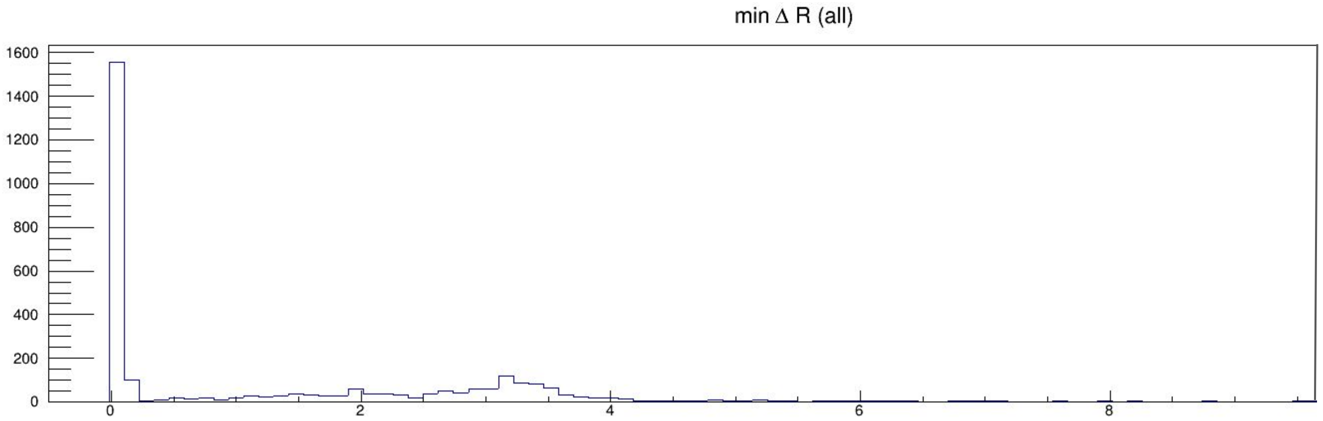

भारतीय प्रौद्योगिकी संस्थान है
Indian Institute of Technology I

HSF-INDIA

# ΔR(DeltaR)

- Another variable, that we used to apply cut is ΔR(Delta R) is a measure of distance between HF_clusters and Gen_electron in the $\eta$-$\varphi$.
- It is equivalent of distance between of 2 point in 2-d plane.

$$\Delta R = \mathrm{sqrt}((\eta_{Gen} - \eta_{HF})^2 + (\phi_{Gen} - \phi_{HF})^2)$$

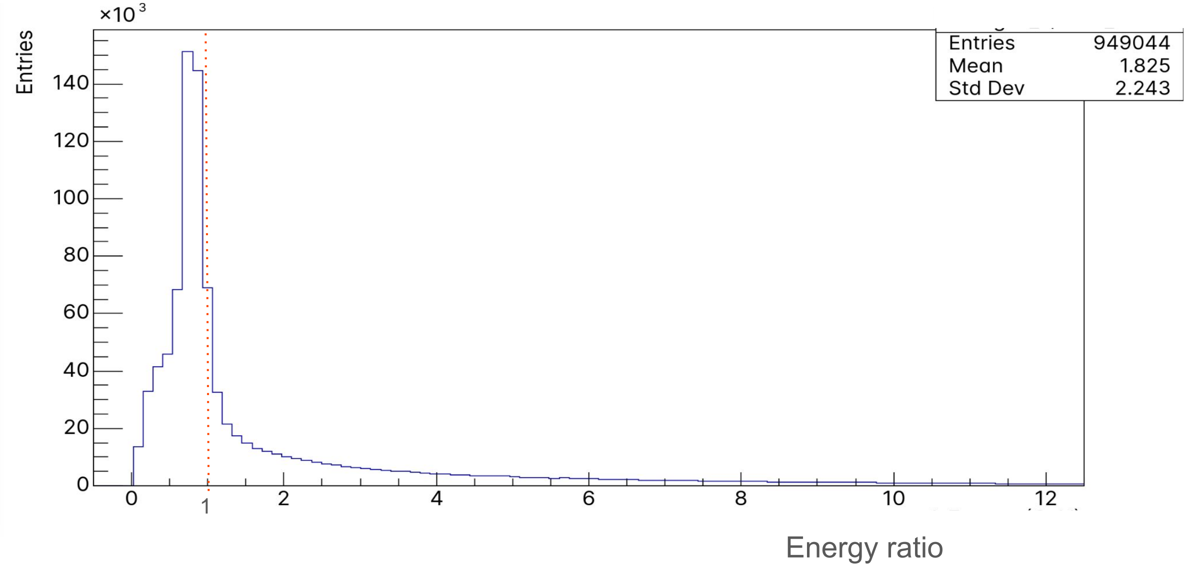




HSF-INDIA

# Applying Cuts and DeltaR

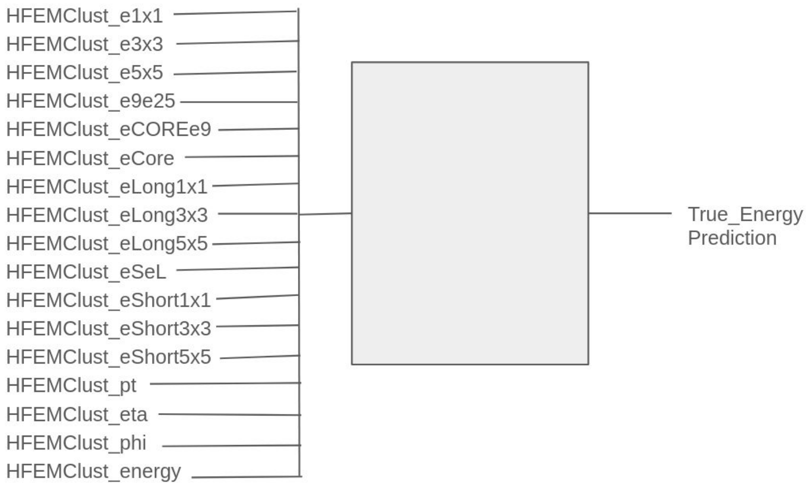
- Energy Ratio after applying the cuts on HF cluster variables and ΔR <0.4



# Using ML Model

- We can use ml model to train it with hf_cluster variables as input label and true electron energy as target label. It's a regression problem. We used xgboost decision trees

- Hyperparameter given are :
  - n_estimators=2000,
  - learning_rate=0.1,
  - subsample=1,
  - colsample_bytree=1,
  - max_depth=7



# Result

- The energy ratio with applied cuts is improved towards 1.
- The energy ratio with predicted energy is peaked around 1.
- We have trained it with
    - Train sample size = 367k events
    - Test sample size = 157k events

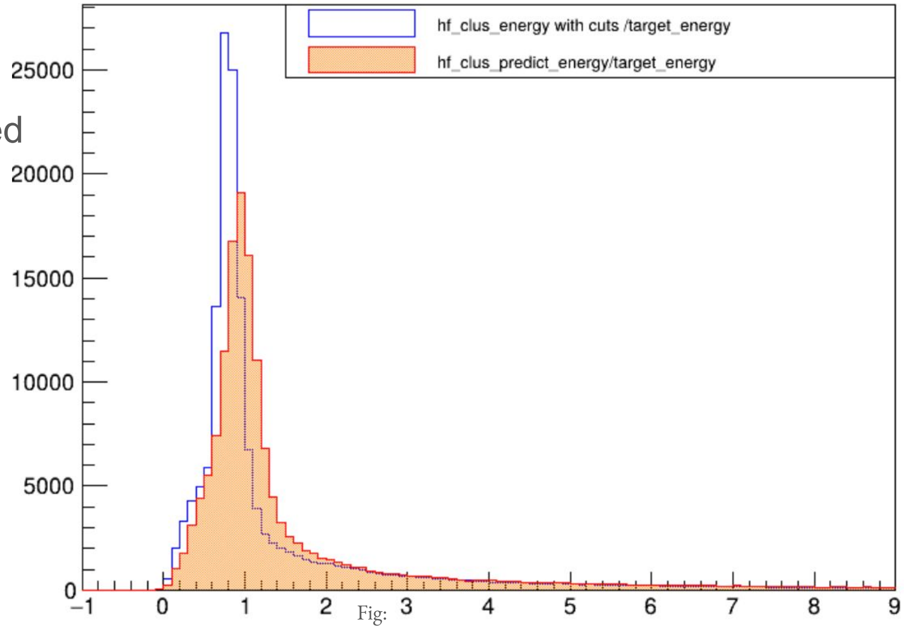



Fig: Overlay of histograms HF_cluster energy with True_ele_energy and Predicted_energy by xgboost

भारतीय प्रौद्योगिकी संस्थान है
Indian Institute of Technology I

HSF-INDIA

# Tools used

CMSSW ( CMS Softawre Framework) is the software used by the CMS experiment for processing and analyzing data.

It provides tools for event simulation, reconstruction and custom analysis.

- Features:
    - Tools for simulation, reconstruction, and custom analysis.
    - Python-based configuration for flexibility.
    - C++ modules (e.g., EDAnalyzer) for dataset reduction.
- Applications:
    - Analyze particle collision events.
    - Retrieve large datasets via CRAB.
    - Prepare data for ML models and apply event-specific cuts.

# Future Work

- We can explore more advanced machine learning models, such as Neural Networks, to evaluate how effectively they can predict true energy using HF_Cluster variables as input.

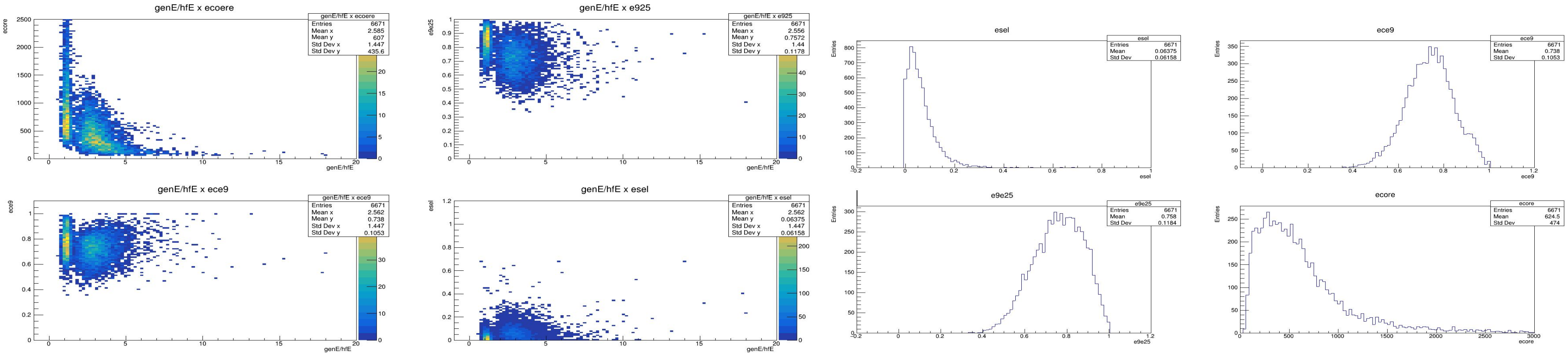
- Also, this study is limited for electrons of certain energy range  within Zee electrons. We can expand this range by generating  with entire range of energy with Pythiagun.

# Thank You

भारतीय प्रौद्योगिकी संस्थान है
Indian Institute of Technology I

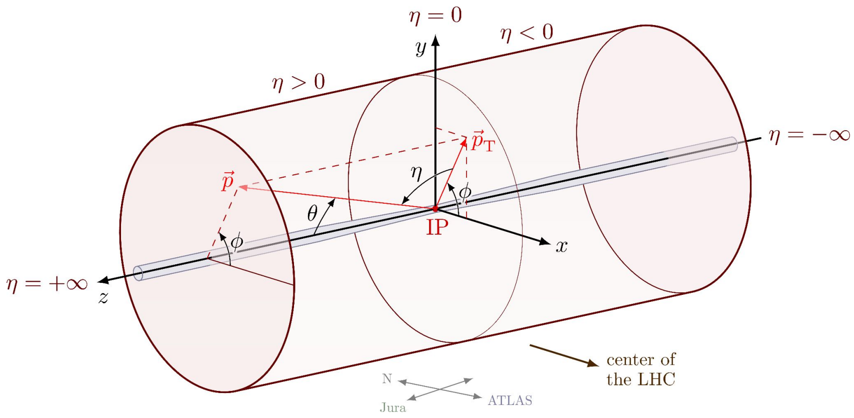
HSF-INDIA

# Future Work


genE/hfE x ecoere
genE/hfE x e925
esel
ece9
genE/hfE x ece9
genE/hfE x esel
e9e25
ecore

# CMS Geometry

- The origin of the coordinates is at the center of the beam line, where the particle collisions occur, with a right-handed coordinate system.
- The z-axis aligns with the beam line, the y-axis points vertically upward, and the x-axis points toward the LHC's center of curvature.
- Positions are described using azimuthal angle ($\varphi$) in the x-y plane and pseudorapidity ($\eta$) along the z-axis, where $\eta = -\ln(\tan(\theta/2))$




Img source : https://tikz.net/axis3d_cms/

# Conclusion

Although the invariant mass still falls slightly below the expected value of 90 GeV,we have observed improvements in both EE-HF and EB-HF conditions, as indicatedby the histograms and Gaussian fit results discussed earlier.

**EB-HF:**

• We explored different cut combinations, adjusting them incrementally by 0.05 around the initial estimate derived from the 2D-histogram on HF variables:eSeL, e9e25, and ece9.

• Subsequently, we created histograms of their invariant mass distributions and applied Gaussian fitting.

• Among all these combinations, we sought the cut combination fit result with the minimum sigma (resolution) and a mean nearest to the mass of the Z-boson.

• Based on the data presented earlier, it is evident that the minimum standard deviation (sigma) occurs around the cuts of 0.2, 0.5, and 0.6 for the variables esel, ece9, and e9e25, respectively. At these thresholds, we observe a mean value of 88.19 and a standard deviation of 14.66

EE-HF:

- We Repeated steps followed for EB-HF for EE-HF.
- Based on the data presented earlier, it is evident that the minimum standard deviation (sigma) occurs around the cuts of 0.2, 0.5, and 0.6 for the variables esel, ece9, and e9e25, respectively.
- At these thresholds, we observe a mean value of 84.67 and a standard deviation of 19.71

# Appendix

- There are three steps to reconstructing electrons in LHC collision events using HF.
- The first step searches through the signals in HF to find potential candidates for electrons by searching for clusters of energy in the calorimeter.
- The second applies corrections to adjust the energy and position of the candidates to account for detector response.
- The last step applies identification requirements that compare the characteristics of the candidates to the profile of an electron shower, and accept or reject the candidate accordingly in order to remove non-electron responses.
- The quartz fibers alternate in type and are transversely spaced 55 mm apart within the detector to ensure comprehensive particle detection.
-

- We explored different cut combinations, adjusting them incrementally by 0.05 around the initial estimate derived from the 2D-histogram on HF variables: eSeL, e9e25, and ece9.

- Subsequently, we created histograms of their invariant mass distributions and applied Gaussian fitting.

- Among all these combinations, we sought the cut combination fit result with the minimum sigma (resolution) and a mean nearest to the mass of the Z-boson.

# EB-HF 2d histograms

- I have plotted 2d histograms for HF variables vs Invariant Mass.
- From these histograms, we Took Cuts HF variables for along the 90 Gev .
- Invariant mass is calculated with one Electron from EB(eta<1.479) and other from HF.

# EB-HF after cut

- Based on the data presented earlier, it is evident that the minimum standard deviation (sigma) occurs around the cuts of 0.2, 0.5, and 0.6 for the variables esel, ece9, and e9e25, respectively. At these thresholds, we observe a mean value of 88.19 and a standard deviation of 14.66

# EE-HF 2d histograms

- I have plotted 2d histograms for HF variables vs Invariant Mass.
- From these histograms, we Took Cuts HF variables for along the 90 Gev .
- Invariant mass is calculated with one Electron from EB(eta>=1.479) and other from  HF.

# EE-HF after cut

- Based on the data presented earlier, it is evident that the minimum standard deviation (sigma) occurs around the cuts of 0.2, 0.5, and 0.6 for the variables esel, ece9, and e9e25, respectively. At these thresholds, we observe a mean value of 84.67 and a standard deviation of 19.71

भारतीय प्रौद्योगिकी संस्थान है
Indian Institute of Technology I

HSF-INDIA