FUTURE CIRCULAR COLLIDER

Hyperparameter Optimization of Deep Learning Models for Cluster Counting

Asif Ali
Politecnico University
& INFN di Bari

Tutors
Prof. Nicola De Filippis
Dr. Walaa Elmetenawee

Outline

1. Motivation & Signal Generation in a Drift Tube

2. Peak finding with Hybrid (CNN + Transformer) model

3. Clusterization approaches with Hybrid CNN + Transformer models
   - Regression-based method
   - Classification-based method

4. Primary Cluster Counting

5. Comparison with MC truth

6. Brief Higgs Analysis at FCC-ee

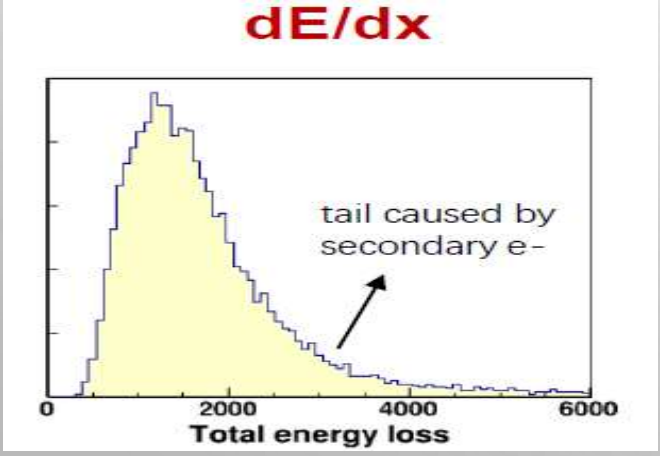
Motivation: Cluster Counting $\left(\frac{dN}{dx}\right)$ in Drift Chamber

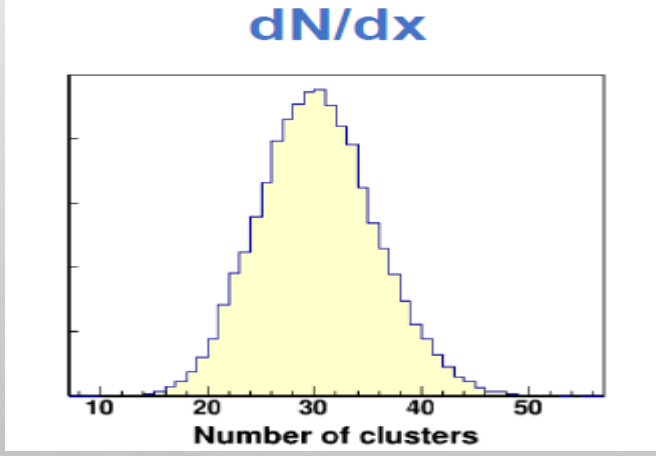
$\left(\frac{dE}{dx}\right)$: **Measure the ionization by total energy loss**

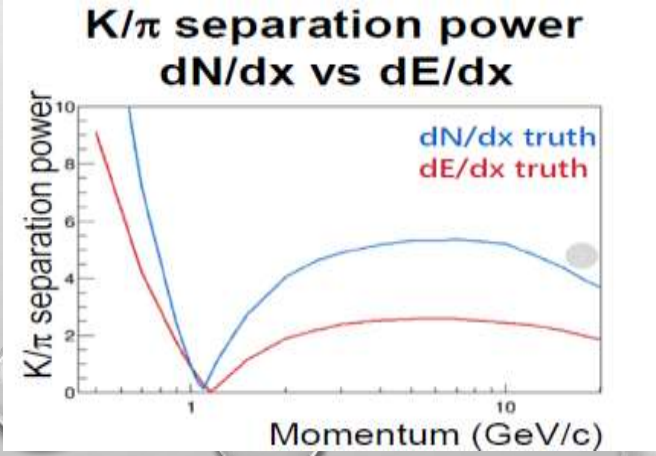
- Landau distribution
- Large fluctuation from many sources

$\left(\frac{dN}{dx}\right)$: **Measure the ionization by number of primary ionizations**

- Poisson distribution
- Small fluctuation; Potentially improve the resolution by a factor of 2

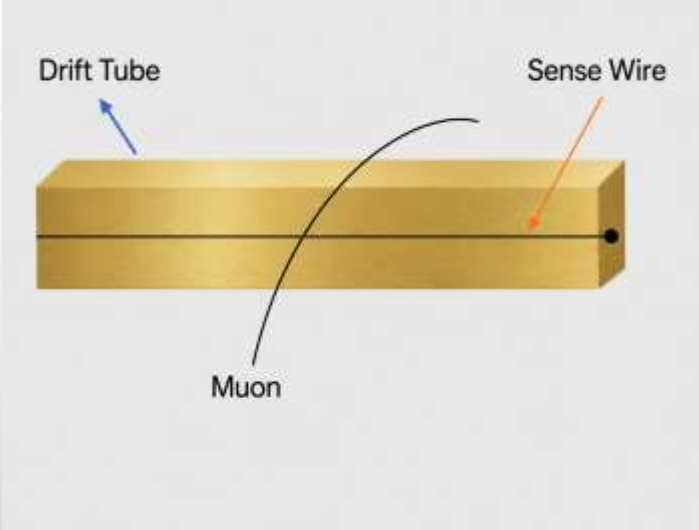


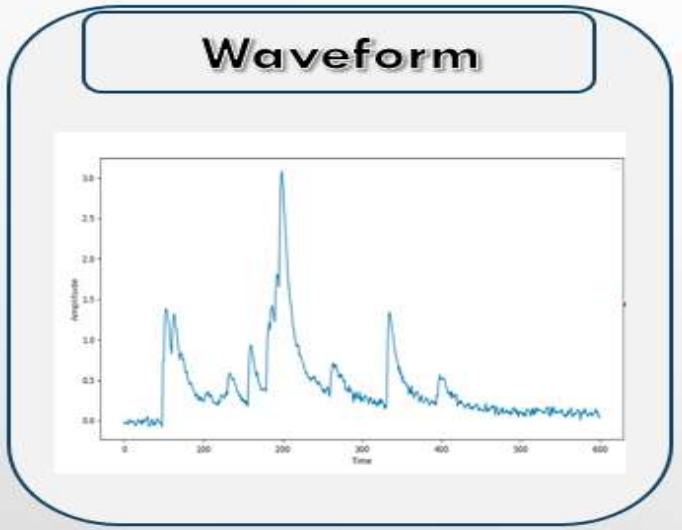




Signal Generation in a Drift Tube

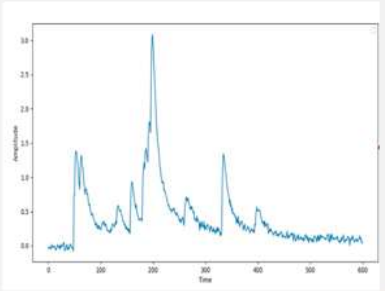
A muon particle is passed through mixture of gases (90% He and 10% C4H10) generate electron-ion pairs drift toward thier opposite polarity and generate signal.





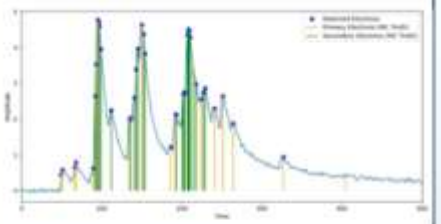
Workflow for Cluster Counting

Waveform



Step 1: Peak Finding

Classification
Signal(Primary +Secondary) vs Noise



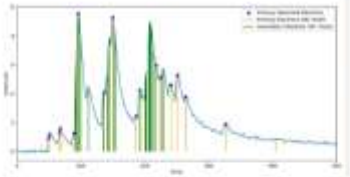
Output: Detected peaks

Step 2 : Clusterization

Primary clusters only

Method A: Regression

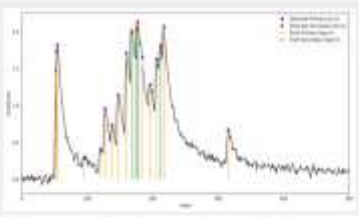
Directly predict the number of primary clusters



Issue: Worse standard deviation

Method B: Classification

Classify detected peaks as Primary or Secondary



Count only the predicted Primary peaks

Final approach

Output

Primary ionization cluster count

$$\left(\frac{dN}{dx}\right)$$

What is the Cluster Counting $\left(\frac{dN}{dx}\right)$ Reconstruction:

The $\left(\frac{dN}{dx}\right)$ reconstruction is to determine **the number of primary clusters** in the waveform.

Two Steps of Reconstruction Algorithm in Cluster Counting

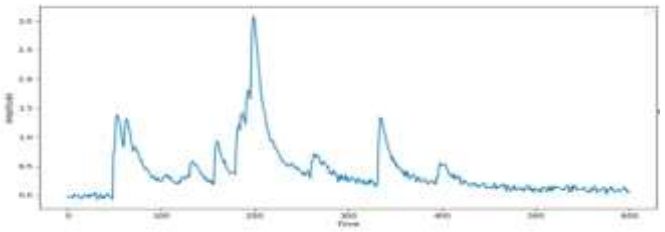
1. **Peak finding:** Find all peaks (primary and secondary) in the waveform.
2. **Clusterization:** From the detected peaks, identify only the true primary ionization clusters

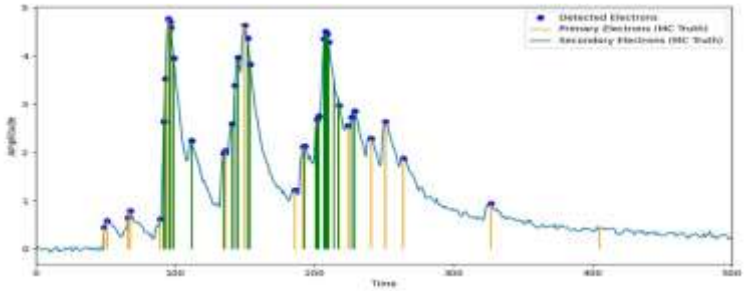
Hybrid Models (CNN + Transformer) for peak finding

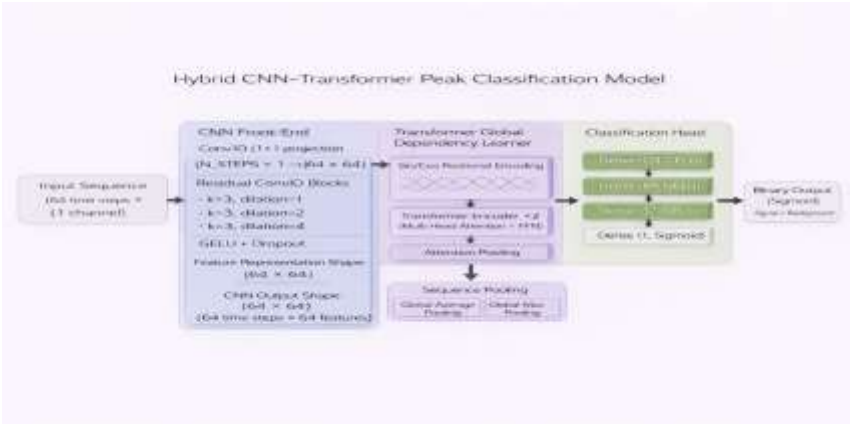
Machine Learning based algorithm
Hybrid(CNN + Transformer)

- inputs: 64-time windows/peak
- Optimization: Grid Search
- Metric: AUC

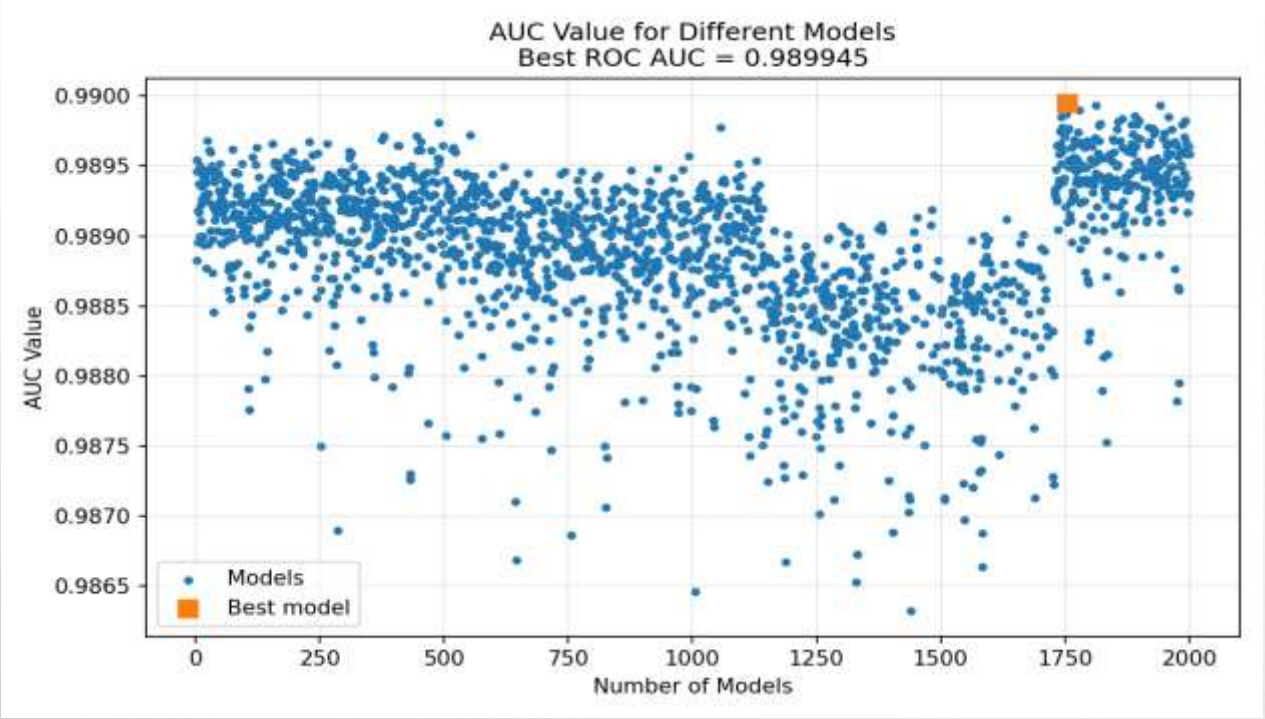
No. of Heads	4	Output_activation	Sigmoid
Key_Dim	16	Optimizer	AdamW
Learning_rate	8e-5	Batch_Size	64
Epochs	120	Hidden Activation	GeLu

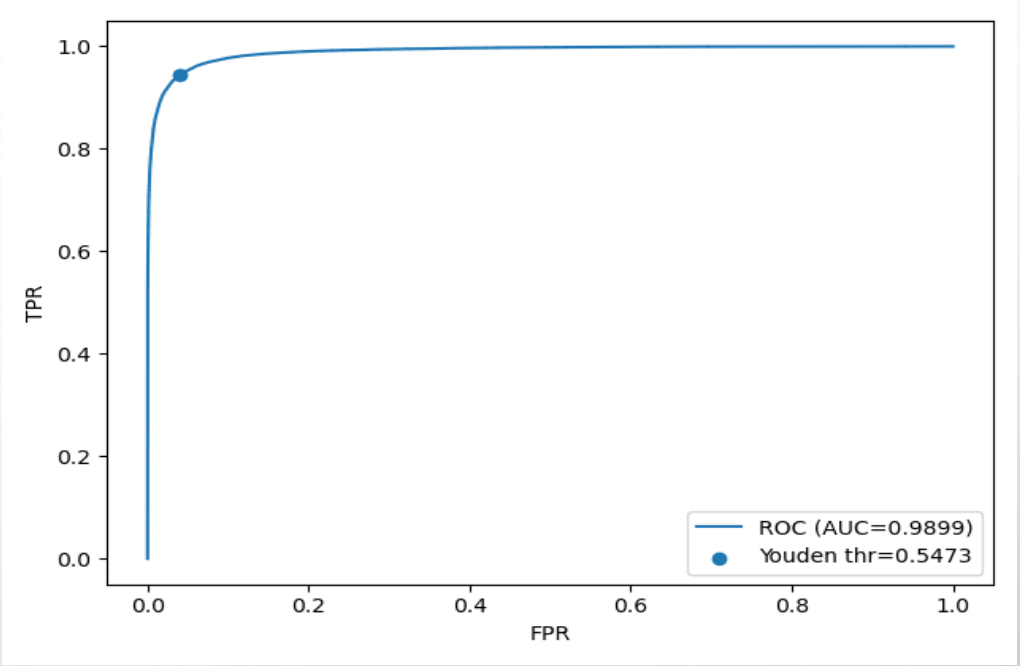






Performance of the Best Hybrid Models (CNN + Transformer) for peak finding





The above plot shows us different configuration models with Area under the curve value and the orange box shows us the best model among all.

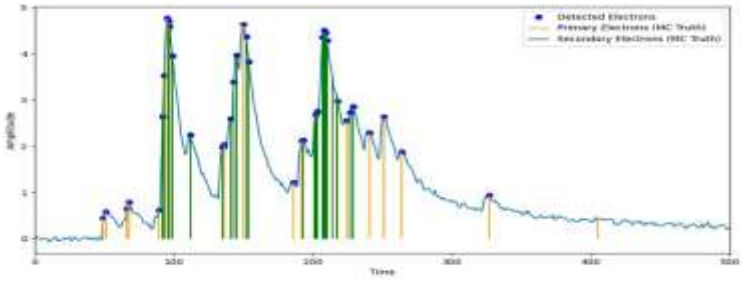
The above plot shows ROC curve for the best Transformer model with Area under the curve value 0.9899 ~ 0.99 which show a best classification to discriminate signal from background.

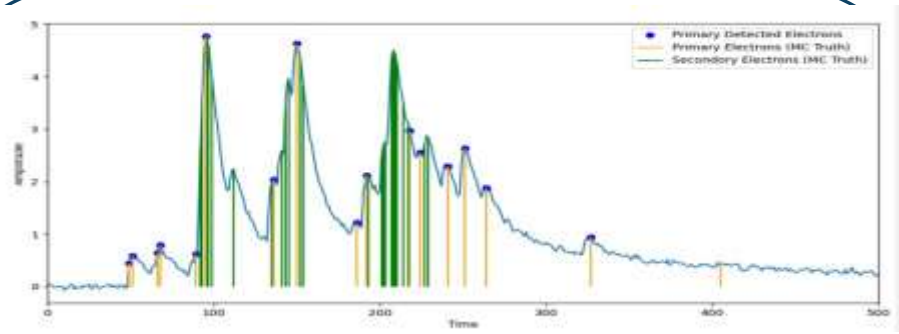
Hybrid Regression Models (CNN + Transformer) for Clusterization

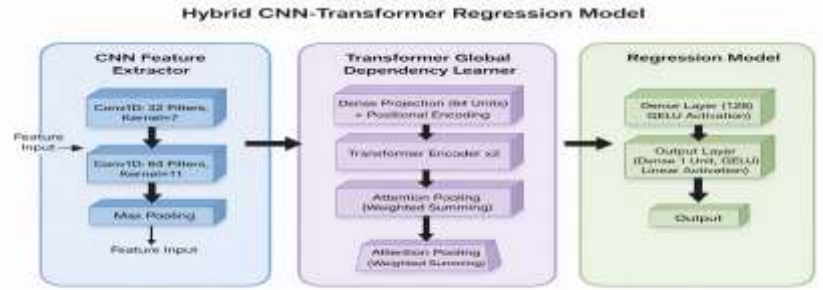
Machine Learning based algorithm
Hybrid(CNN + Transformer)

- inputs: Peaks from peak finding step 1
- Metric: MSE
- Data split: 67% training 33% testing

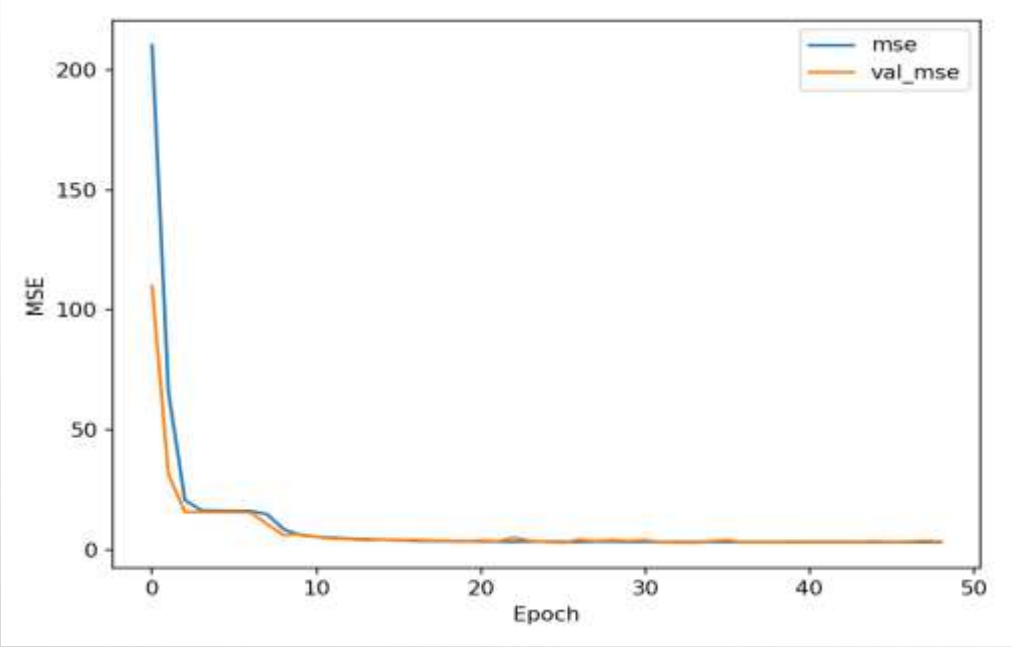
Parameter	Value	Parameter	Value
Hidden neuron	CNN: 32, 64 Transformer FFN: 128 (per Block)	Optimizer	Adam
Learning_rate	10^{-3}	epochs	60
Activation Function CNN/Transformer	ReLU/GeLU	Attention_Head	4







Performance of the Best Hybrid Models (CNN + Transformer) for Clusterization




```
HYBRID CNN + TRANSFORMER REGRESSION (MSE)
==========================================

TRAIN_MSE    2.785482
TRAIN_MAE    1.316739
TRAIN_RMSE   1.668976

VAL_MSE      2.905806
VAL_MAE      1.339935
VAL_RMSE     1.704642

VAL_MAE_ROUNDED    1.318485
VAL_RMSE_ROUNDED   1.727408
```

- Training and validation curves overlap ⇒no overfitting.
- Stable MSE indicate strong generalisation.

$$\text{MAE} = \frac{1}{N}\sum |y_i - \hat{y}_i|, \qquad \text{MSE} = \frac{1}{N}\sum (y_i - \hat{y}_i)^2$$

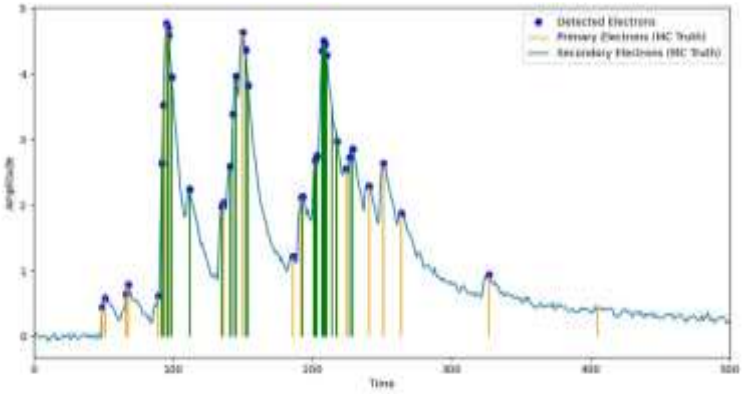
Notation: N = number of samples, y_i = true value, $\hat{y}_i$ = predicted value

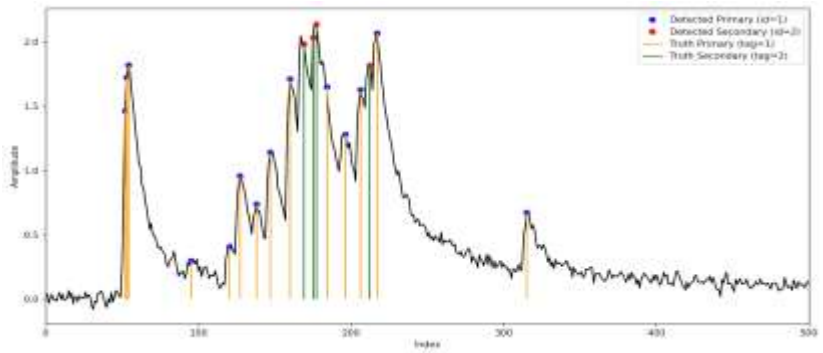
Hybrid classification Models (CNN + Transformer) for Clusterization

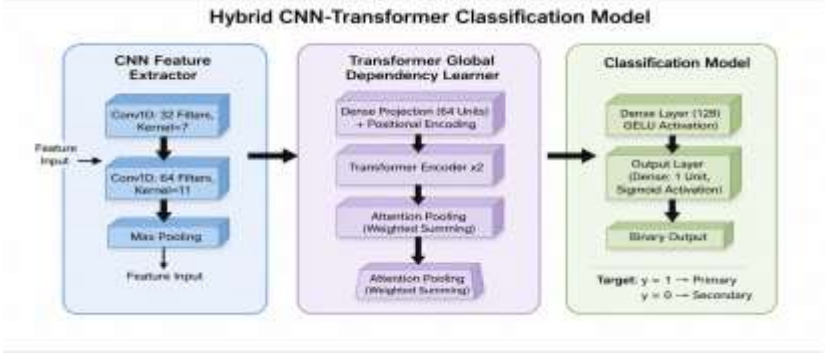
Machine Learning based algorithm

Hybrid(CNN + Transformer)

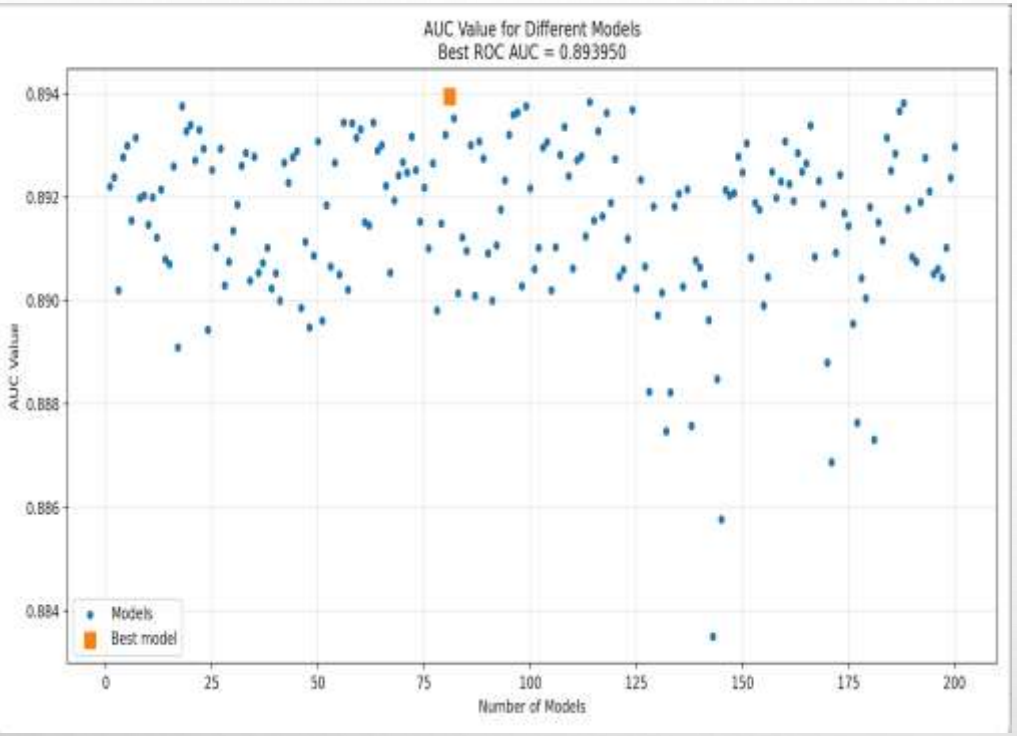
- inputs: Peaks from peak finding step 1
- - Optimization: Grid search
- Metric: AUC
- Data split: 67% training 33% testing

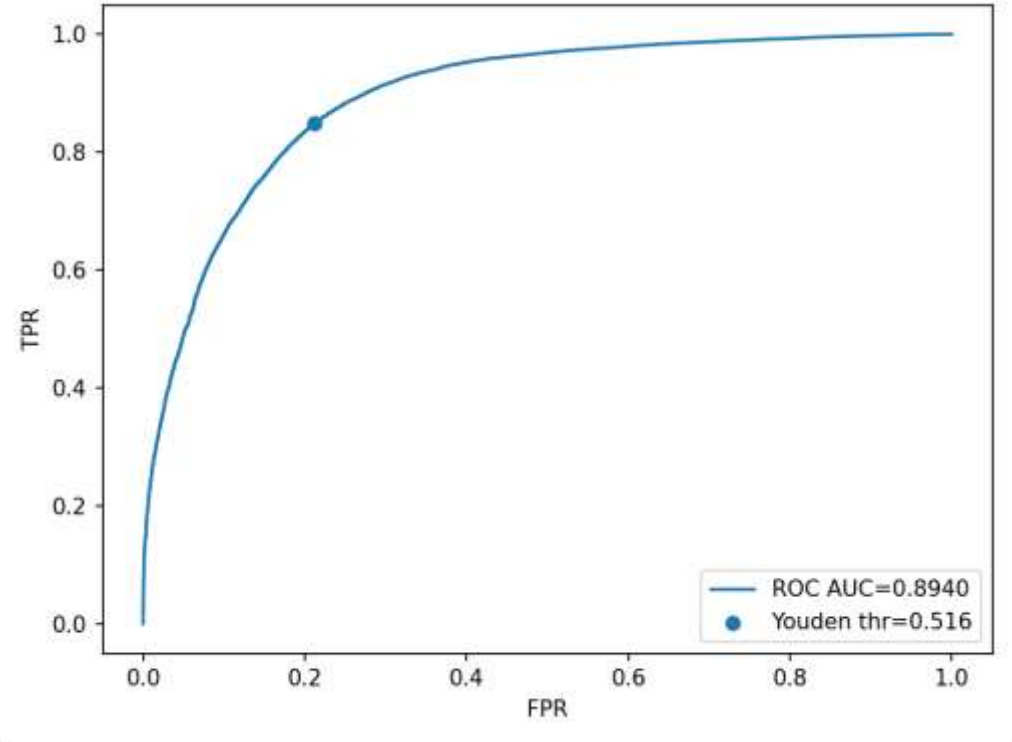






Performance of the Best Hybrid Models (CNN + Transformer) for Clusterization



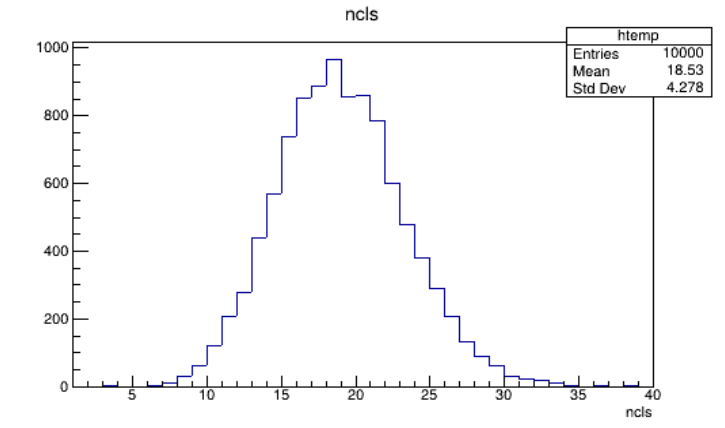


The above plot shows us different configuration models with Area under the curve value and the orange box shows us the best model among all.

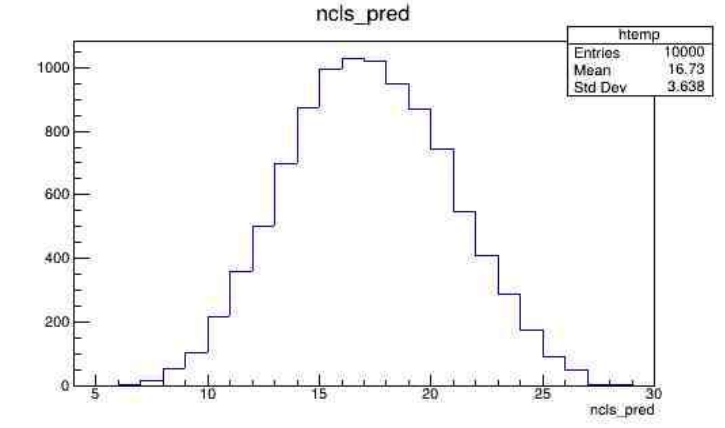
The above plot shows ROC curve for the best Transformer model with Area under the curve value 0.8940

Results

Reconstruction Results for 10 GeV Momentum of Muon

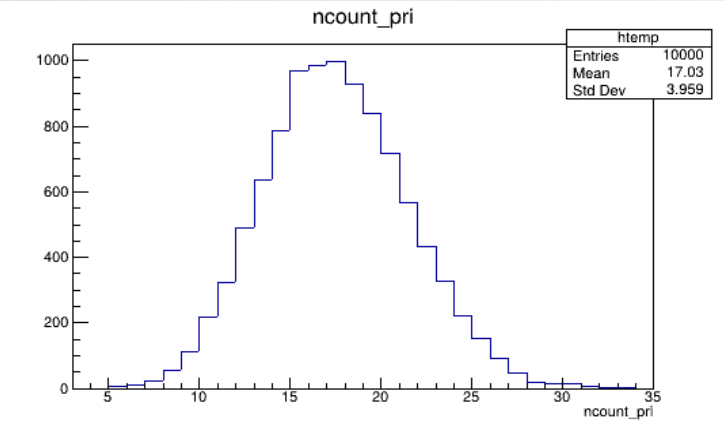


Number of Primary Clusters(MC)

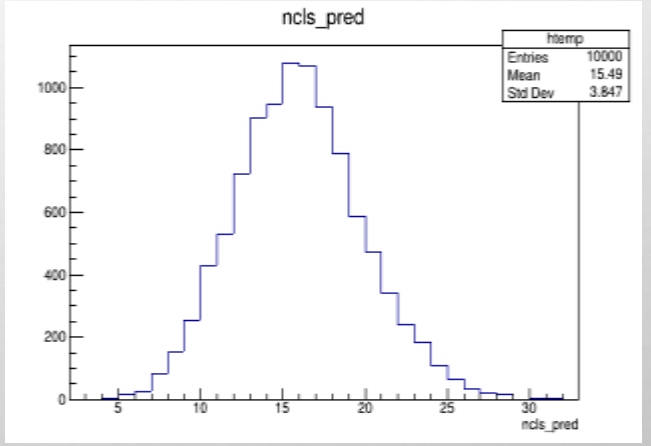


Mean: 16.73
Std: 3.63

Predicted Number of Primary Clusters from Regression



Number of Primary Clusters from Peak Finding (CNN + Transformer) — Target



Mean: 15.49
Std: 3.84

Predicted Number of Primary Clusters from Classification

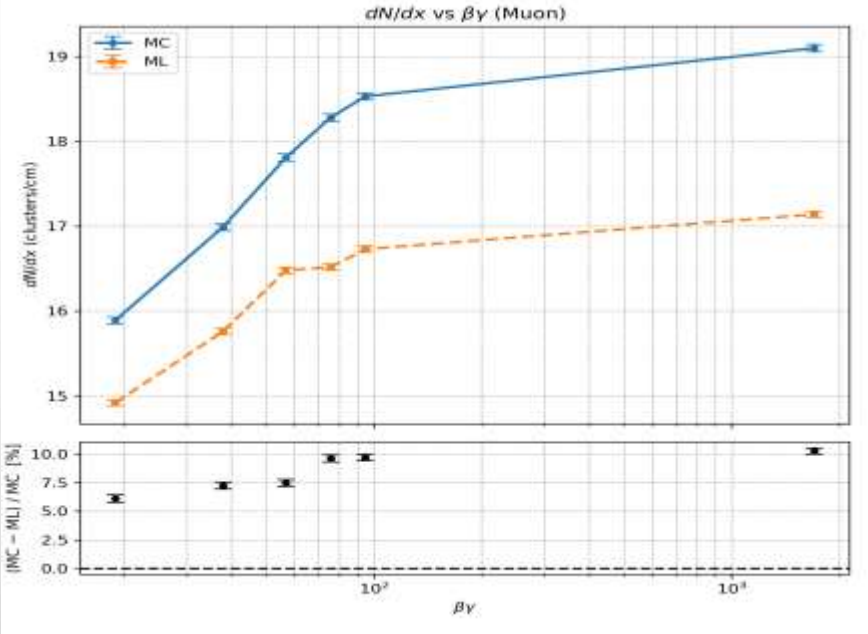
Table: Regression-Based Cluster Counting Results

Different Momenta of Muon(GeV)	Primary Cluster (MC)	Standard Deviation (MC)	Primary Cluster (Peak Finding)	Standard Deviation (Peak Finding)	Primary Cluster (Cluster number regression)	Standard Deviation (Cluster number regression)
2GeV	15.89	4.01	15.10	3.83	14.92	3.48
4GeV	16.99	4.10	15.92	3.87	15.76	3.55
6GeV	17.81	4.10	16.49	3.91	16.48	3.57
8GeV	18.28	4.30	16.70	3.92	16.52	3.61
10GeV	18.53	4.20	17.03	3.95	16.73	3.63
180GeV	19.1	4.30	17.49	3.98	17.14	3.51

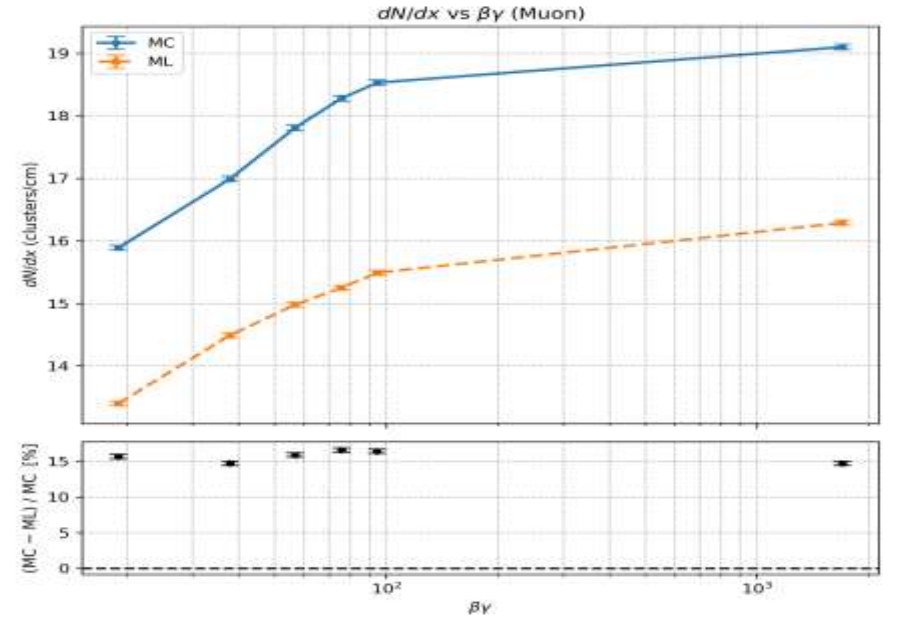
Table: Classification-Based Cluster Counting Results

Different Momenta of Muon (GeV)	Primary Cluster (MC)	Standard Deviation (MC)	Primary Cluster (Peak Finding)	Standard Deviation (Peak Finding)	Primary Cluster (Cluster number classification)	Standard Deviation (Cluster number classification)
2GeV	15.89	4.01	15.10	3.83	13.40	3.53
4GeV	16.99	4.10	15.92	3.87	14.49	3.72
6GeV	17.81	4.10	16.49	3.91	14.98	3.78
8GeV	18.28	4.30	16.70	3.92	15.25	3.80
10GeV	18.53	4.20	17.03	3.95	15.49	3.84
180GeV	19.1	4.30	17.49	3.98	16.29	3.97

$\frac{dN}{dX}$ vs $\beta\gamma$ for Muon



Regression-based clusterization for $\frac{dN}{dx}$ vs $\beta\gamma$



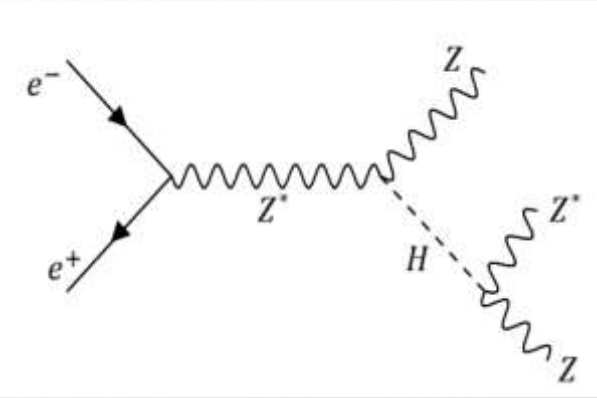
Classification-based clusterization for $\frac{dN}{dx}$ vs $\beta\gamma$

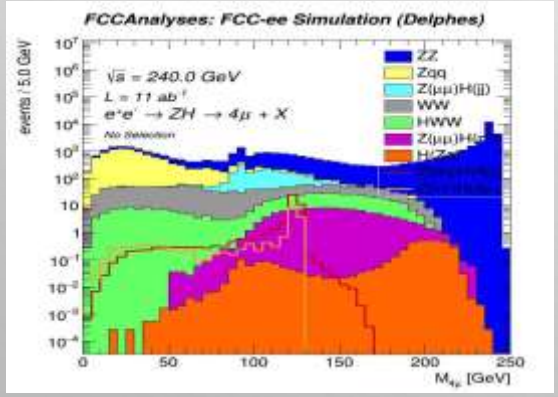
- $\frac{dN}{dx}$ increases with βγ, showing that relativistic muons produce more ionization clusters in the gas.
- At high $\beta\gamma$, the increase in $\frac{dN}{dx}$ becomes slower, showing the beginning of saturation.
- MC and ML show a similar trend of $\frac{dN}{dx}$ versus $\beta\gamma$, indicating that the momentum dependence of cluster production is preserved.

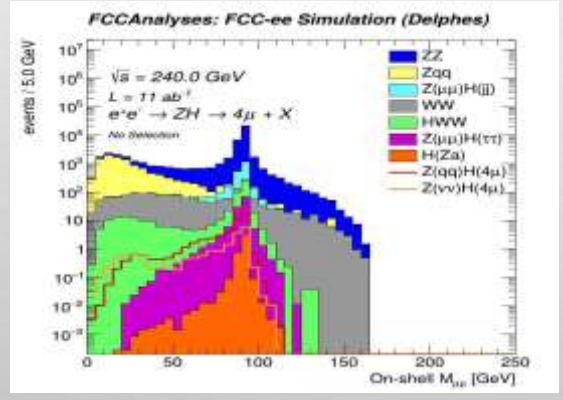
Higgs Analysis at FCC-ee

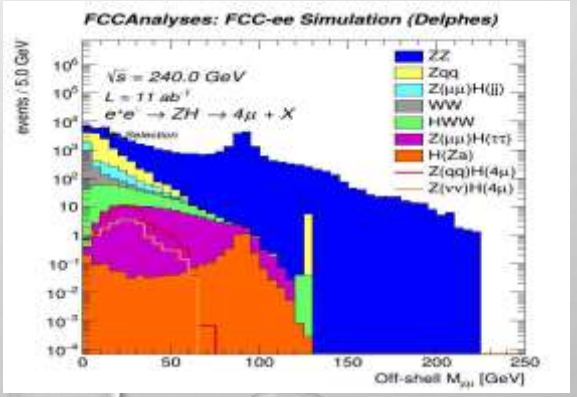
- FCC-ee provides a clean environment for precision Higgs studies
- At $\sqrt{s} = 240\text{GeV}$, the dominant Higgs production mode is Higgsstrahlung
- Process:

  e+ e- → Z H
- The Higgs boson is produced in association with a Z boson
- Z boson studied in $Z \to jj$ and $Z \to \nu\nu$ channels
- Target decay channel: $H \to ZZ^* \to 4\ell$
- Final states: $4\mu,\ 4e,\ 2e2\mu,\ 2\mu2e$









Conclusion

1. ML-based cluster counting chain developed for drift tube signals

2. Hybrid CNN + Transformer used for peak finding and clusterization

3. Regression and classification approaches studied for clusterization

4. Reconstructed $\frac{dN}{dx}$ follows the expected $\beta\gamma$ trend

5. ML-based cluster counting shows potential for future PID studies

6. FCC-ee Higgs analysis was briefly introduced

Future Plan

1. Apply the trained NN model to real test-beam data

2. Reconstruct primary clusters in real waveforms

3. Extend cluster counting toward particle identification

4. Evaluate muon /pion/kaon/ separation power

5. Evaluation of ML-based cluster counting for PID

6. Extend the FCC-ee Higgs analysis to $\sqrt{s} = 365\text{GeV}$

Thank you for your attention!

Backup

Transformer Encoder: Input & Attention

$$x_i \in \mathbb{R}, \qquad i = 1, \ldots, T$$

$$z_i = \underbrace{W_e x_i + b_e}_{\text{Amplitude Embedding}} + \underbrace{PE(i)}_{\text{Positional Encoding}}$$

$$PE(pos, 2j) = \sin\left(\frac{pos}{10000^{2j/d_{\text{model}}}}\right)$$

$$PE(pos, 2j+1) = \cos\left(\frac{pos}{10000^{2j/d_{\text{model}}}}\right)$$

$$Z = [z_1, z_2, \ldots, z_T] \in \mathbb{R}^{T \times d_{\text{model}}}$$

$$Q = ZW^Q, \qquad K = ZW^K, \qquad V = ZW^V$$

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Transformer Encoder: Block & Output

$$\text{head}_h = \text{Attention}\left(QW_h^Q, KW_h^K, VW_h^V\right)$$

$$\text{MHA}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_H)W^O$$

$$Z' = \text{LayerNorm}\left(Z + \text{MHA}(Q, K, V)\right)$$

$$\text{FFN}(Z') = \sigma(Z'W_1 + b_1)W_2 + b_2$$

$$Z_{\text{out}} = \text{LayerNorm}\left(Z' + \text{FFN}(Z')\right)$$

$$p_i = \sigma\left(W_o Z_{\text{out},i} + b_o\right)$$

$$\hat{y}_i = \begin{cases} 1, & p_i \geq \theta \\ 0, & p_i < \theta \end{cases}$$

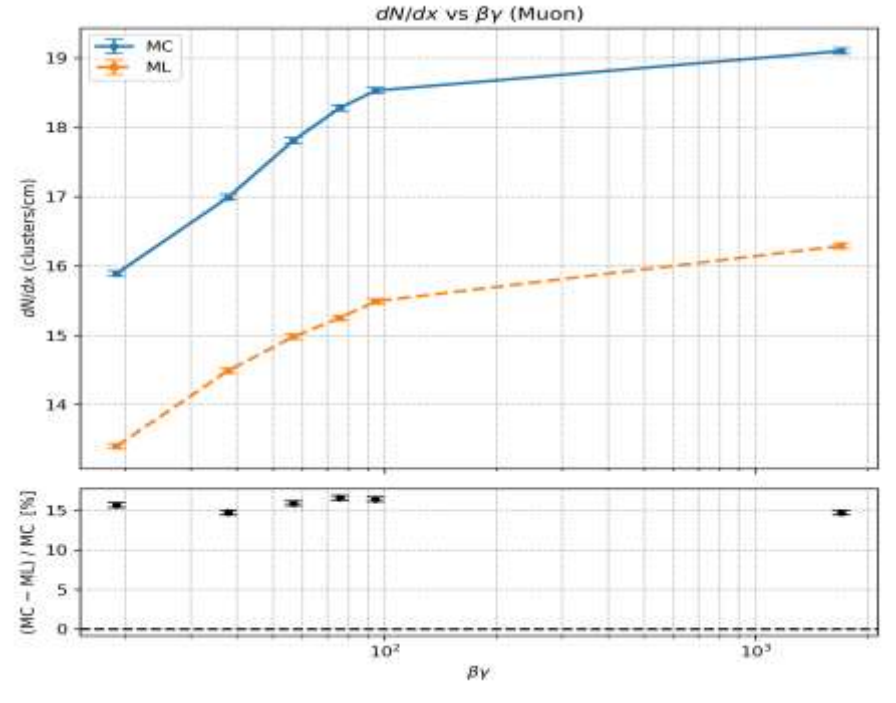
- Regression directly predicts the total primary clusters per event, giving a good average count, but it pulls predictions toward the mean, reducing the spread.

Mean: 16.73
Std: 3.63

- Classification counts primary clusters peak by peak, so the spread is better preserved, but threshold cuts can reject some true primaries, lowering the average count.

Mean: 15.49
Std: 3.84

$\frac{dN}{dX}$ vs $\beta\gamma$ for Muon



$$\sigma_{mean} = \frac{\sigma}{\sqrt{N}}$$

$$\sigma = event\ by\ event\ standard\ deviation$$

$$N = number\ of\ events$$

$$\sigma_{mean} = statistical\ uncertainity\ on\ the\ mean\ dN/dx$$

- Relative Difference $= \Delta(\%) = \frac{MC-ML}{MC} \times 100$
- Gaussian error propagation is used because the relative difference is a function of statistically independent mean values, whose uncertainties are Gaussian.

- Gaussian Error propagation:

$$\sigma_{\Delta(\%)} = \sqrt{\left(\frac{\sigma_{ML}}{MC}\right)^2 + \left(\frac{ML.\sigma_{MC}}{MC^2}\right)^2}$$