# Analysis of the Real Test Beam Data 2022 by Using ML Models

Muhammad Numan Anwar

Bari-LECCE Meeting

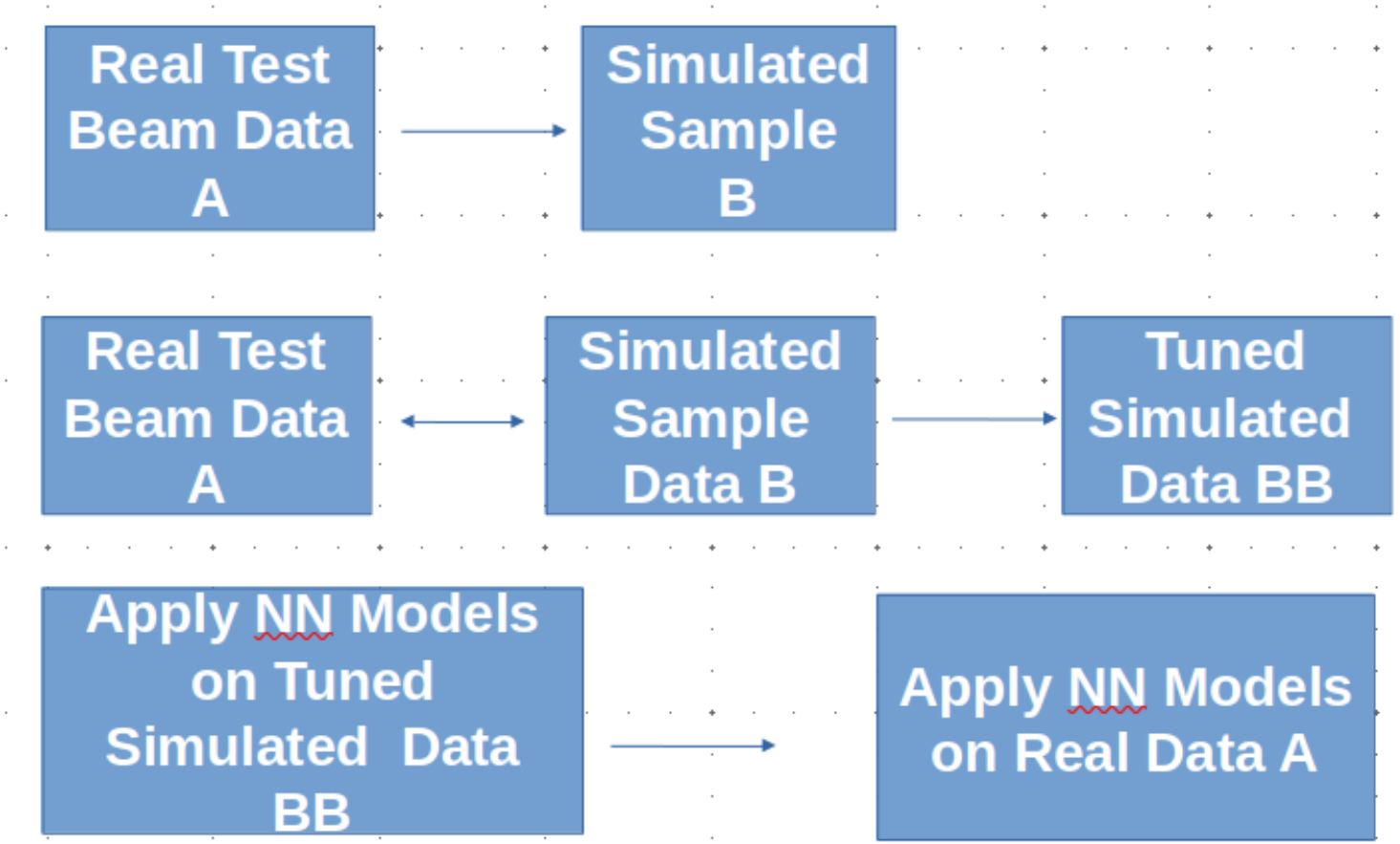
Nov 16, 2025

# Introduction

- How to Apply NN on the data
- Waveform selections in Data 2022
- Tune MC compare to Data 2022
  - Maximum Amplitude
  - RMS Noise
- Very Preliminary  LSTM Peak Finding and CNN regression Model

# How to Apply Neural Network Models on Data

- **Now, we need real test beam data (A) as reference**
- **we would simulate waveform in simulation data (B) with the same parameters as were used in real data (A)**
- **Then we would compare real data (A) and simulated data (B)**
- **Onward, we would tune simulated data (BB) according to real data (A)**
- **After that we should Train Neural Network Models on tune simulated data (BB)**
- **At the end, we would apply Neural Network Models on the real beam test data (A)**



# Configuration Setup for the Real Test Beam data 2022

## Confiurations and setup information

| DRS16 channels | HV channels | Tubes |
|---|---|---|
| 0 | 0 | 1.0cm-20µm |
| 1 | 1 | 1.0cm-20µm |
| 2 | 2 | 1.0cm-20µm |
| 3 | 3 | 1.0cm-20µm |
| 4 | 4 | 1.0cm-20µm |
| 5 | 5 | 1.0cm-20µm |
| 6 | 12 | 1.5cm-20µm |
| 7 | 13 | 1.5cm-20µm |
| 8 | 14 | 1.5cm-20µm |
| 9 | 15 | 1.5cm-20µm |
| 10 | - | - |
| 11 | - | - |
| 12 | - | - |
| 13 | - | - |
| 14 | - | Sipm Scintillator upstream |
| 15 | - | Sipm Scintillator downstream |

Tubes setup with DRS

| Oscilloscope | HV channels | Tubes |
|---|---|---|
| 1 | 16 | 1.5cm-20µm |
| 2 | 17 | 1.5cm-20µm |
| 3 | 18 | 1.5cm-20µm |
| 4 | 19 | 1.5cm-20µm |
| 5 | 8 | 1.0cm-20µm |
| 6 | 6 | 1.0cm-20µm |
| 7 | 9 | 1.0cm-20µm |
| 8 | 10 | 1.0cm-20µm |

Tubes setup with OSC

| | | | HV tag | HV tag |
|---|---|---|---|---|
| | | | 1 | 2 |
| Tubes | Channels | HV channels | Volt (V) | Volt (V) |
| OSC 1cm-20um | 5,6,7,8 | 8,6,9,10 | 1450 | |
| DRS16 1cm-20um | 0,1,2,3,4,5 | 0,1,2,3,4,5 | 1450 | |
| OSC 1.5cm-20um | 1,2,3,4 | 16,17,18,19 | 1550 | |
| DRS16 1.5cm-20um | 6,7,8,9 | 12,13,14,15 | 1550 | |

HV tags with 90/10 gas mix runs

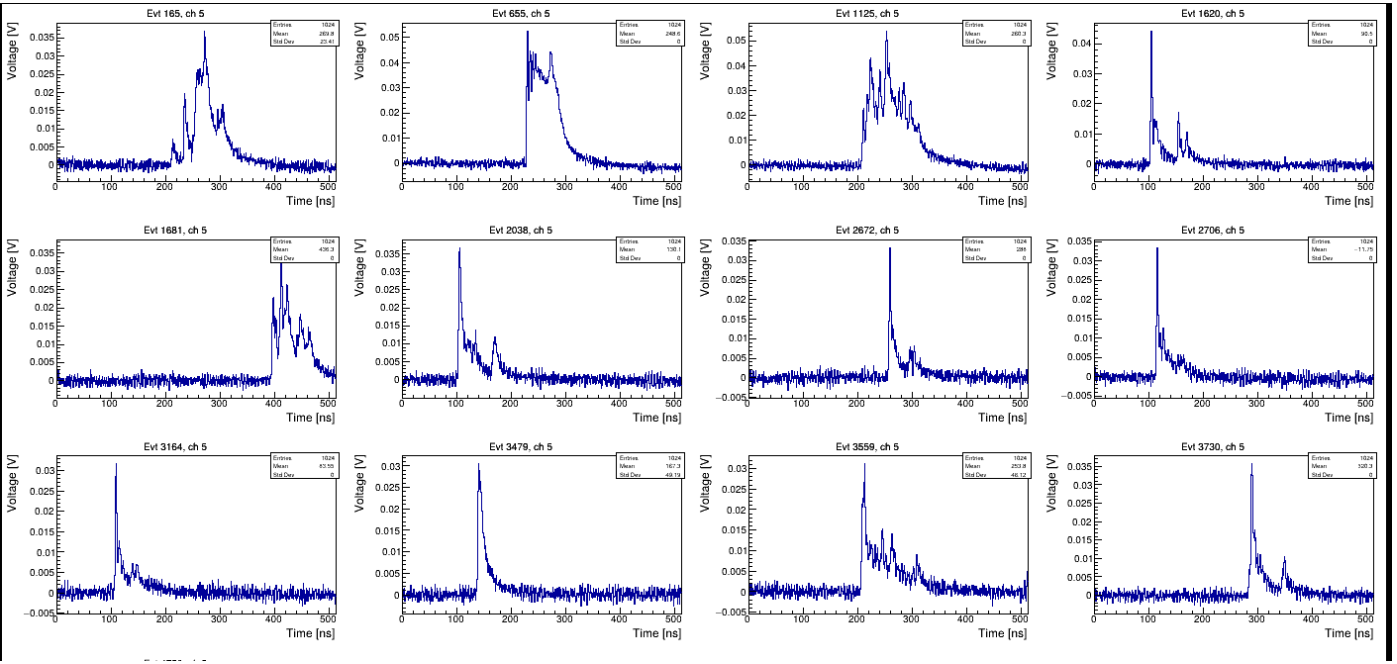
| | | | HV tag | HV tag |
|---|---|---|---|---|
| | | | 1 | 2 |
| Tubes | Channels | HV channels | Volt (V) | Volt (V) |
| OSC 1cm-20um | 5,6,7,8 | 8,6,9,10 | 1450 | 1630 |
| DRS16 1cm-20um | 0,1,2,3,4,5 | 0,1,2,3,4,5 | 1450 | 1630 |
| OSC 1.5cm-20um | 1,2,3,4 | 16,17,18,19 | 1550 | 1730 |
| DRS16 1.5cm-20um | 6,7,8,9 | 12,13,14,15 | 1550 | 1730 |

HV tags with 85/15 gas mix runs

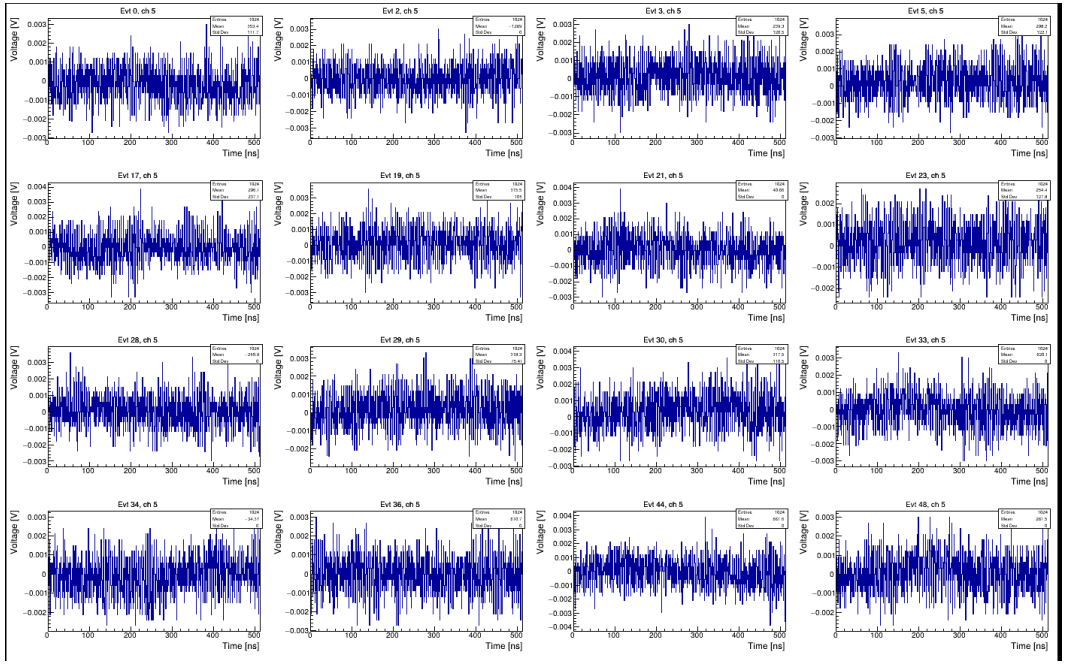
| | |
|---|---|
| **Particle** | **muon** |
| **Gas Mixture** | **He 90% isobutane 10%** |
| **Cell size** | **1 cm** |
| **Momentum** | **180 GeV** |
| **Angle** | **45** |
| **Voltage at sense wire** | **1450** |

# Beam test data selection for Channel 5

## Signal-like data




## Noise-like data

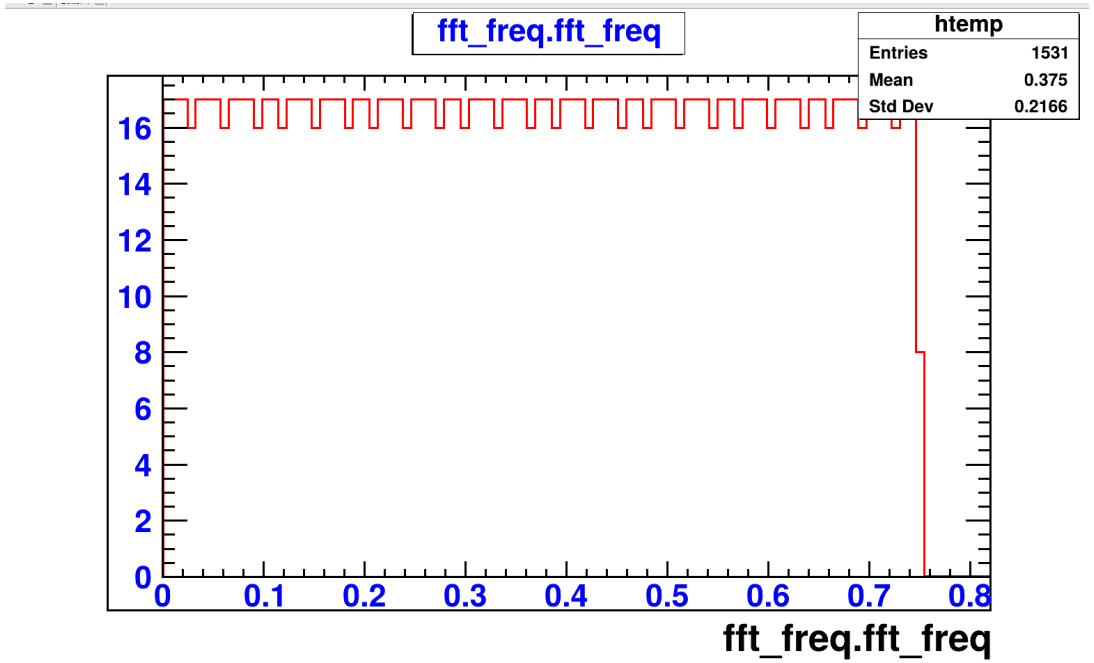


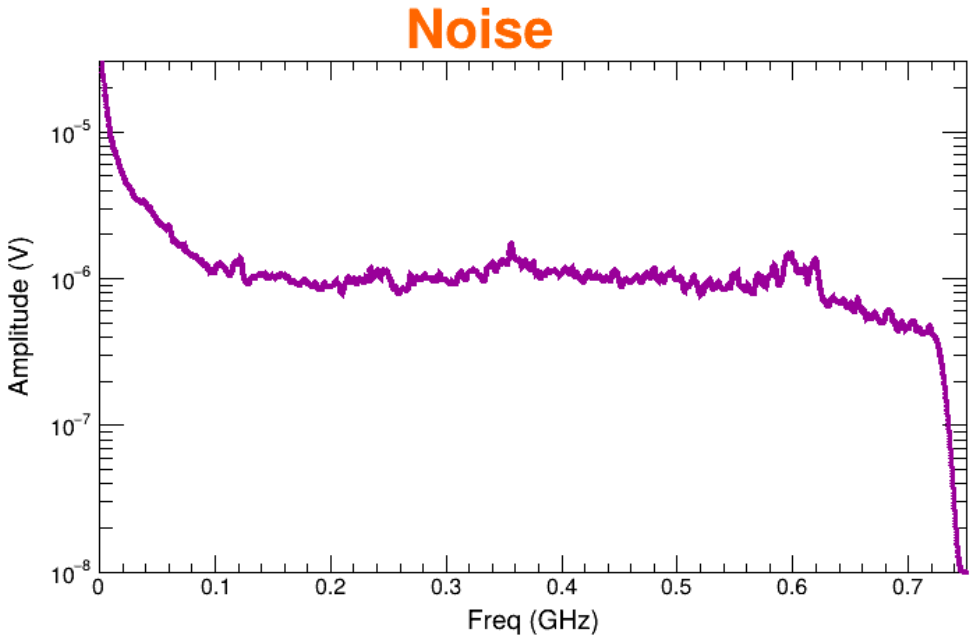

- **Signal-like data: peak finding**
- **Noise-like data: noise information**

# Waveform Selection Logic

| **Quantity** | **Value** |
|---|---|
| Total entries in TChain (all channels) | 58,002 |
| Entries for channel 5 (within range) | 4,832 |
| **Noise RMS cut** | $\sigma_{\text{noise}} < 2$ mV |
| **Amplitude cut (VALID)** | $A_{\text{max}} > 0.03$ V (30 mV) |

# Noise Extraction

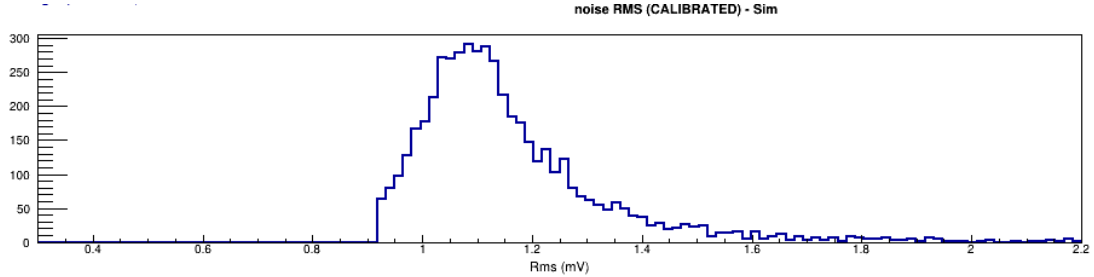



- **This plot shows the full range of frequencies analysed from all processed waveforms combined, covering from 0 up to the Nyquist frequency based on the sampling rate.**

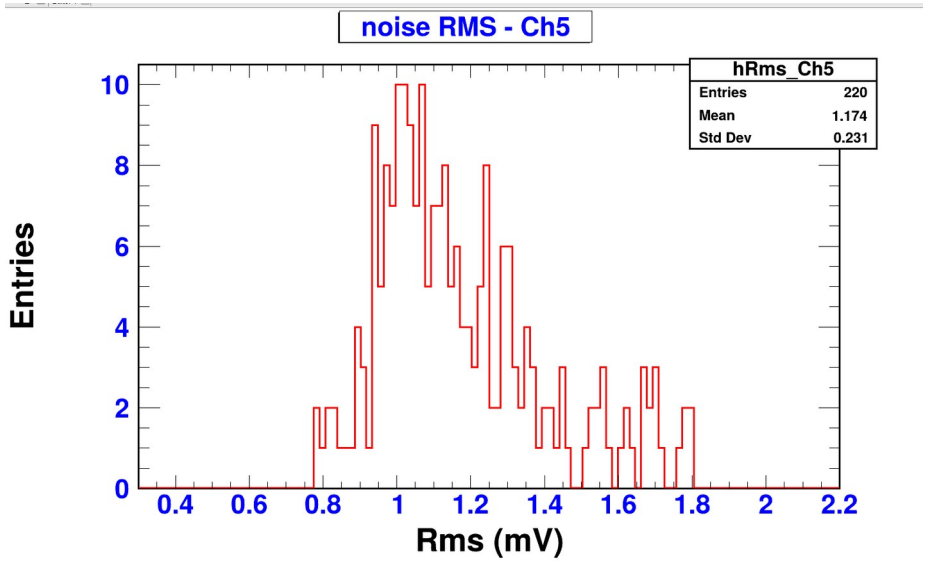
# Noise extraction (Sim VS Data)

**Simulation**




- The above plot shows the distribution rms noise for simulated waveforms(5000 with mean value 1.1mV)

**Data**




- This plot shows the noise RMS(mV) for 220 waveforms with mean value 1.1mV

Summary of Noise Analysis Results (Channel 5)

| Distribution | Entries | Mean | Std. Dev. | Description |
|---|---|---|---|---|
| **FFT** $f_{\text{freq}}$ [GHz] | 1531 | 0.3750 | 0.2166 | Frequency bins of FFT up to the Nyquist limit. |
| **FFT** $f_{\text{mag}}$ [V] | 1531 | 1.4430 | 1.7500 | Average spectral amplitude per frequency. |
| **Noise RMS (Data)** [mV] | 220 | 1.1740 | 0.2310 | RMS noise level in the waveform (Channel 5 data). |
| **Noise RMS (Simulation)** [mV] | 5000 | 1.10 | — | RMS noise level in simulated waveforms (from top distribution). |

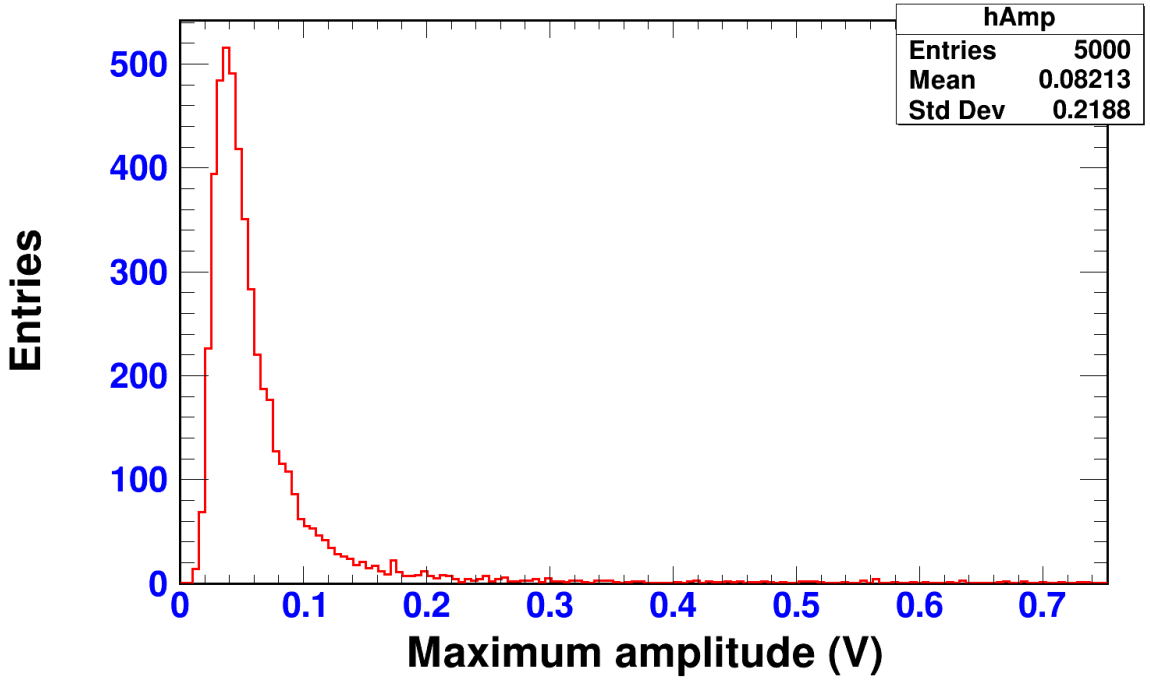
**Noise extracted from the real test beam 2022 data and simulation with mean value 1.170mV by using FFT**

# Signal Amplitude Scale

| Step | $\tau$ | ScaleFactor | Mean (V) | p95 (V) | Match Quality |
|---|---|---|---|---|---|
| 1 | 2.0 | 1.00000 | 1.98 | 6.68 | Too high |
| 2 | 2.0 | 0.01940 | 0.061 | 0.128 | Good amplitude |
| 3 | 3.0 | 0.02161 | 0.0827 | 0.178 | High again |
| 4 | 3.0 | **0.01578** | **0.0607** | **0.1404** | Excellent p95 match |
| 5 | 6.2 | 0.01290 | 0.0768 | 0.167 | Start of detailed study |
| 6 | 7.0 | 0.01400 | 0.0906 | 0.183 | Mean too high |
| 7 | 8.0 | 0.01320 | 0.0934 | 0.201 | Still high |
| 8 | 4.0 | 0.01300 | 0.0588 | 0.136 | Mean too low |
| 9 | 4.0 | 0.01584 | 0.0795 | 0.163 | Slightly low |
| 10 | 4.0 | 0.01660 | 0.0766 | 0.169 | Low |
| 11 | 4.0 | 0.01806 | 0.0850 | 0.182 | Slightly high (+2%) |
| 12 | 4.0 | 0.01769 | 0.0817 | 0.184 | Slightly low (–2%) |
| 13 | 4.0 | 0.01803 | 0.0878 | 0.187 | High again |
| **14 (Final)** | **4.0** | **0.01710** | **0.0821** | **0.176** | **Excellent data match** |

Table 1: Comprehensive tuning evolution of simulated amplitude distribution. The final configuration ($\tau = 4.0$, ScaleFactor= 0.01710) achieves excellent agreement with the measured test-beam mean amplitude ($\approx 0.083$ V).

MC




**Real test beam data 2022**

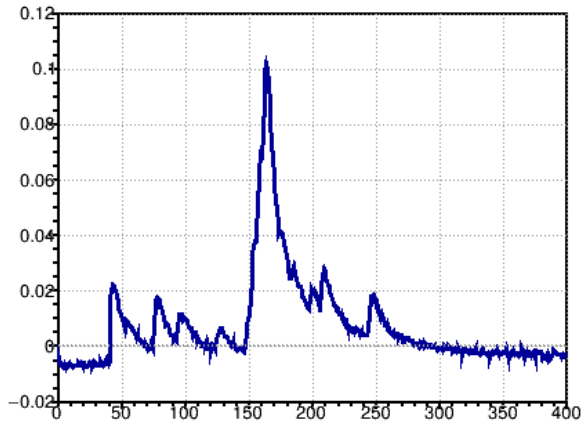
| Quantity | Test-Beam Value (V) |
|---|---|
| Mean amplitude | 0.0833 |
| 95th percentile | 0.1476 |

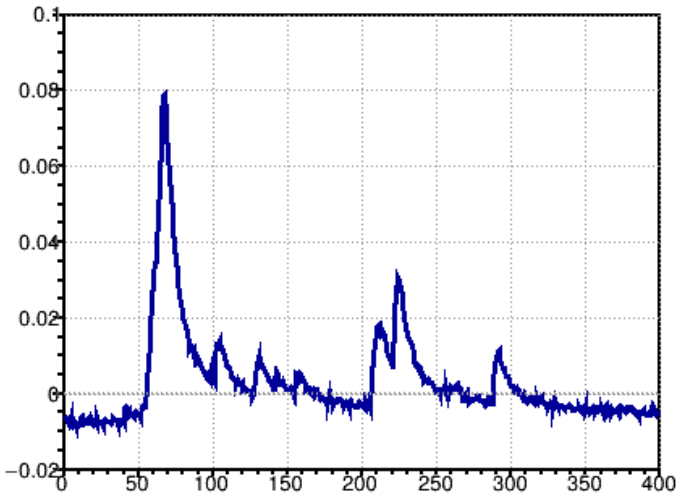
Data

$$\text{ScaleFactor}_{\text{new}} = \text{ScaleFactor}_{\text{current}} \times \frac{\text{mean}_{\text{data}}}{\text{mean}_{\text{MC}}}$$

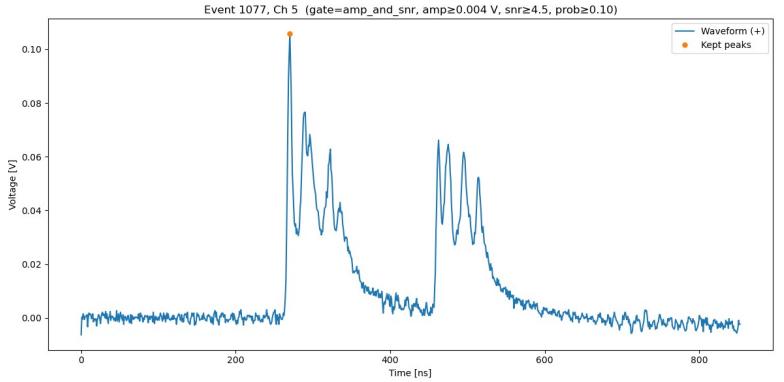
# Comparison of Tuned Simulated and Real Test Beam Data Waveforms

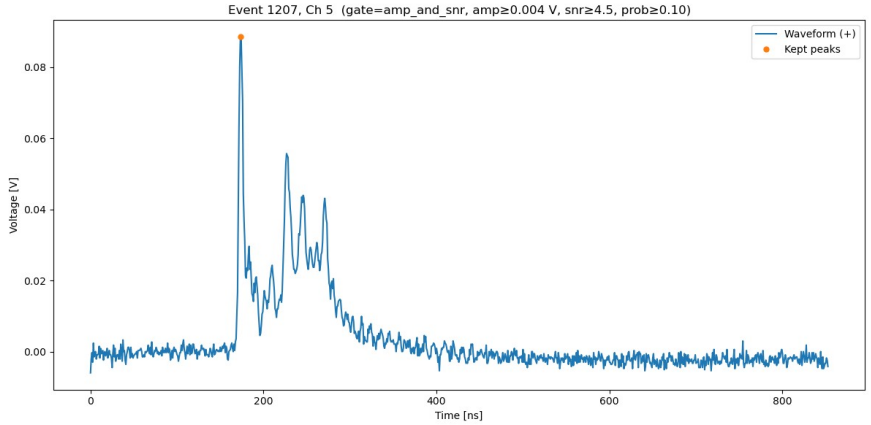
## A. Tuned Simulated Digitized Wave-forms




## B. Real Test Beam Data wave-forms

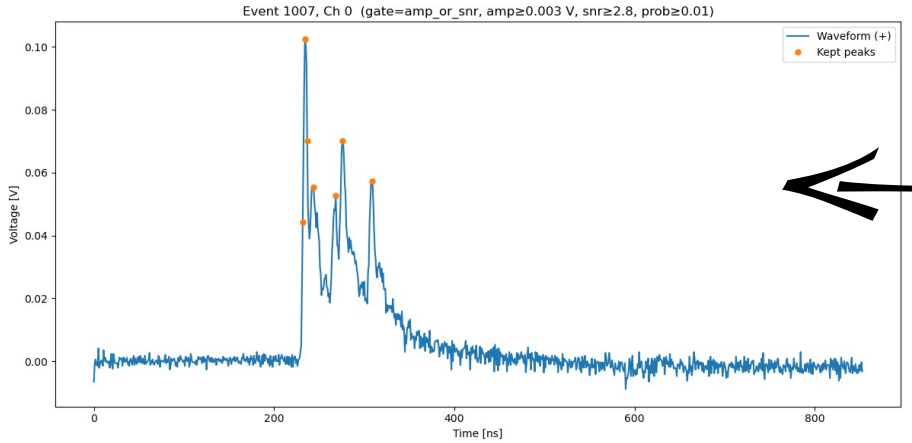


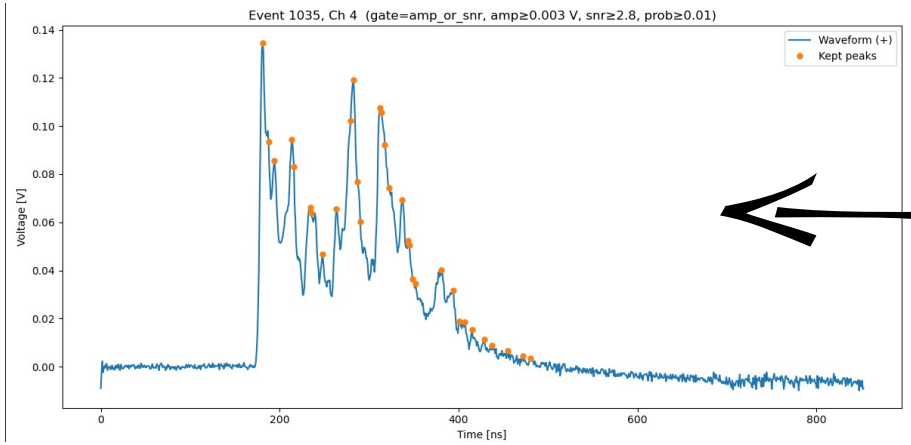
# Applying Untrained NN on beam test data 2022 (180 GeV) Just for Checking

## Step1: Peak Finding Algorithm

### A. Very Preliminary Results of Peak Finding (LSTM)

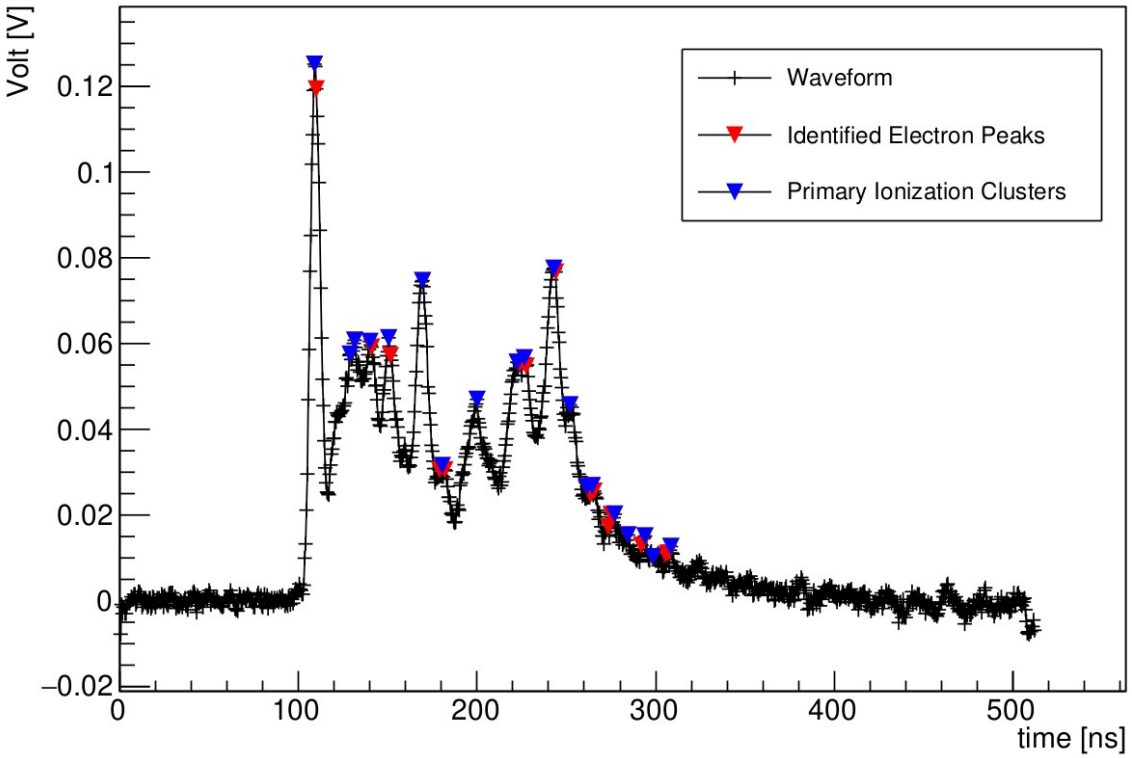



**Tight CUT on Amplitude =0.03 V**



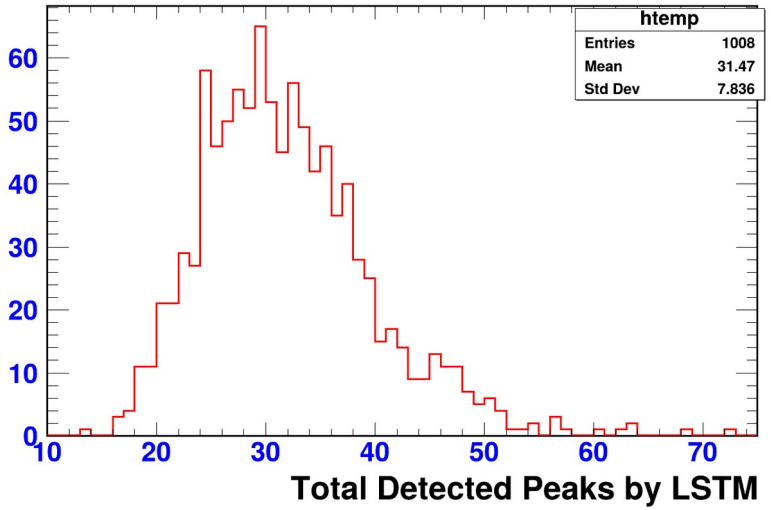

**LOSE CUT on Amplitude=0.01V**

### B. RTA Algorithm



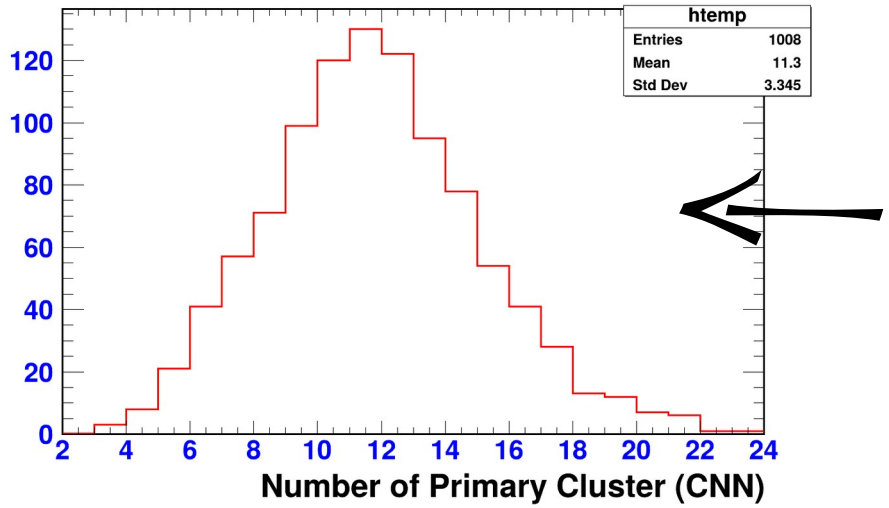
# Applying Untrained NN on beam test data 2022 (180 GeV) Just for Checking

## Preliminary Results of Peak Finding distribution

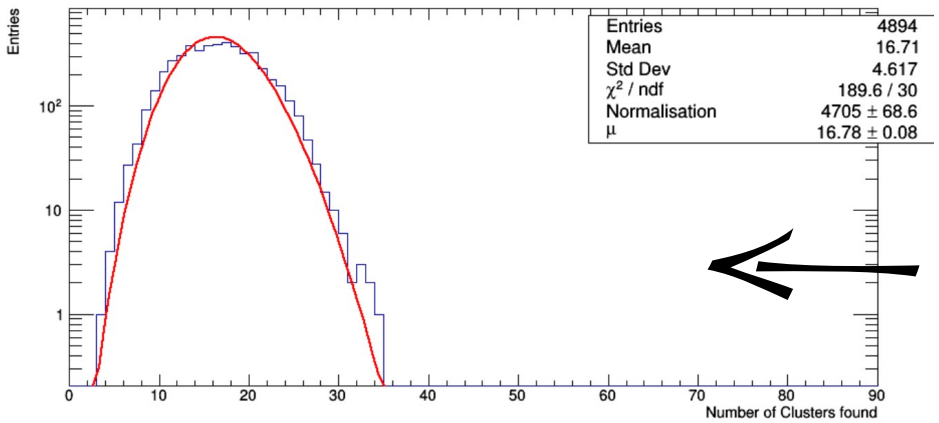



- The above distribution shows us total number detected peaks (primary and secondary electrons) tl channel

## Clusterization




**Preliminary CNN Clusterization**




**RTA Algorithm**

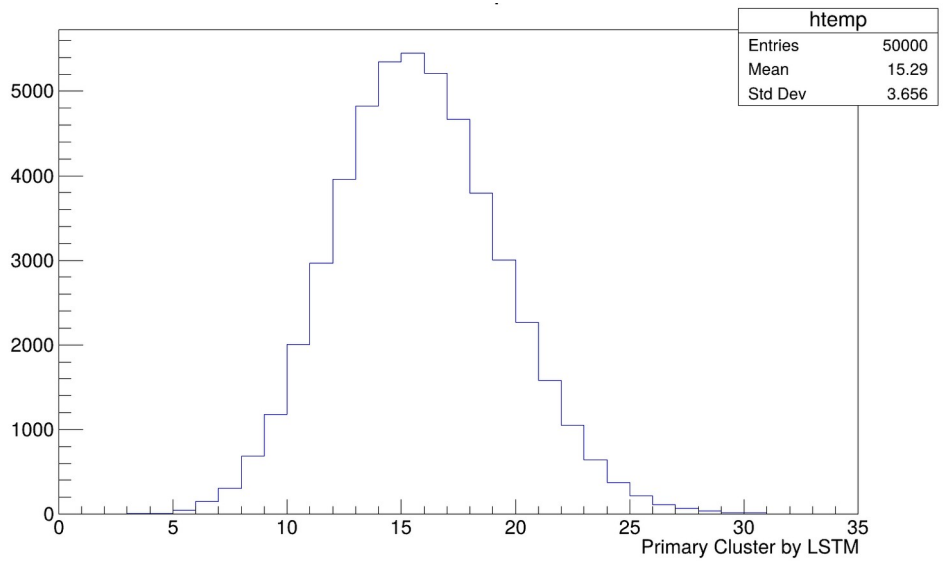
# What to do Next

- Training of LSTM on tune simulated 50000 waveforms are still running
- Highest AUC LSTM model will apply on tune simulated waveforms and data
- All the results should be compare with RTA algorithm

Thank you

# Background Slides

TUNE SIMULATED RESULT
just for checking



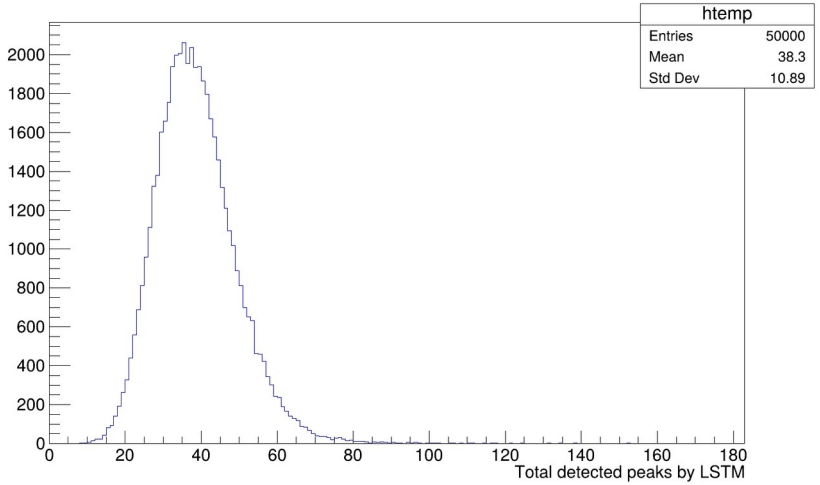




Table 1: **Summary of Measured Medians and Tuned Parameters**

| Quantity | Data (real test beam) | MC @ 180 GeV | Comments |
|---|---|---|---|
| Median amplitude (V) | 0.0749 | 0.0527834 | Used for gain (`ScaleFactor`) tuning |
| Median RMS (V) | 0.001118 | 0.012474 | Used for noise (`NoiseAmplitude`) tunin |

### 1) Preamplifier Gain (ScaleFactor) Tuning

We adjust the simulation gain so that its *median amplitude* matches data:

$$\boxed{\text{ScaleFactor}_{\text{new}} = \text{ScaleFactor}_{\text{current}} \times \frac{\text{median}_{\text{data}}}{\text{median}_{\text{MC}}}}$$

*Numerical evaluation*

$$\begin{aligned}\text{ScaleFactor}_{\text{new}} &= 0.0276 \times \frac{0.0749}{0.0527834} \\ &= 0.0276 \times 1.35885 \\ &= \mathbf{0.03916} \approx \mathbf{0.0392}.\end{aligned}$$

### 2) Noise Amplitude (Baseline RMS) Tuning

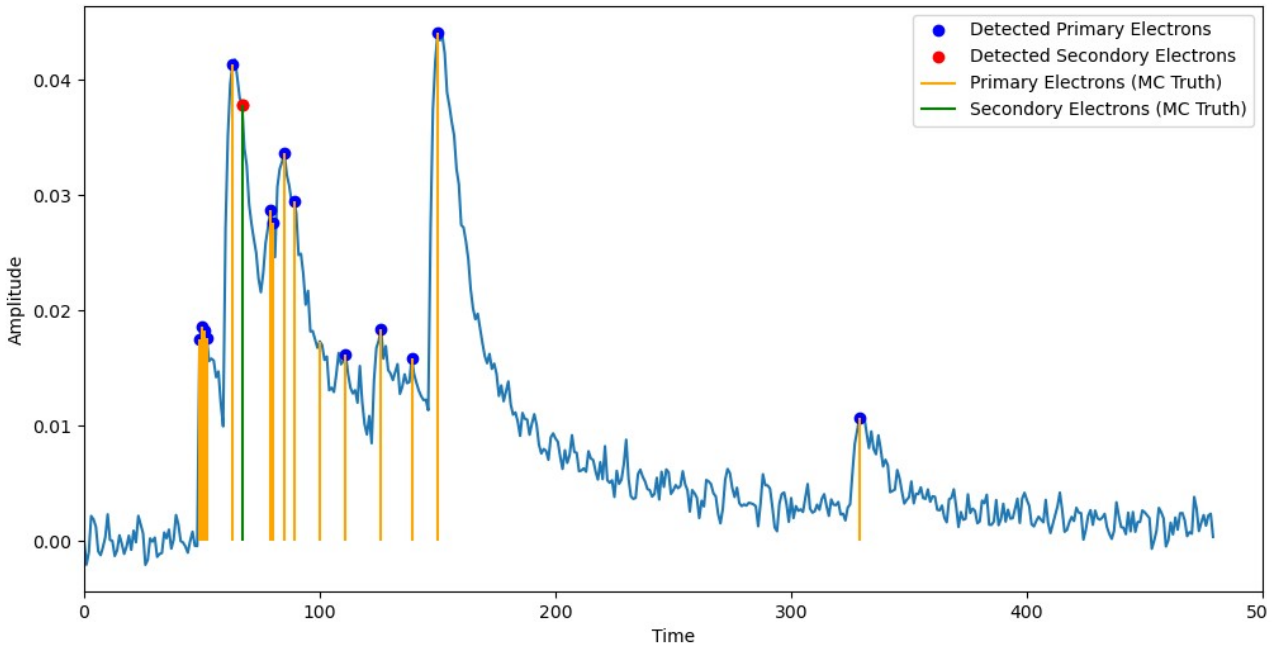
We scale the simulated noise so the *median RMS* matches data:

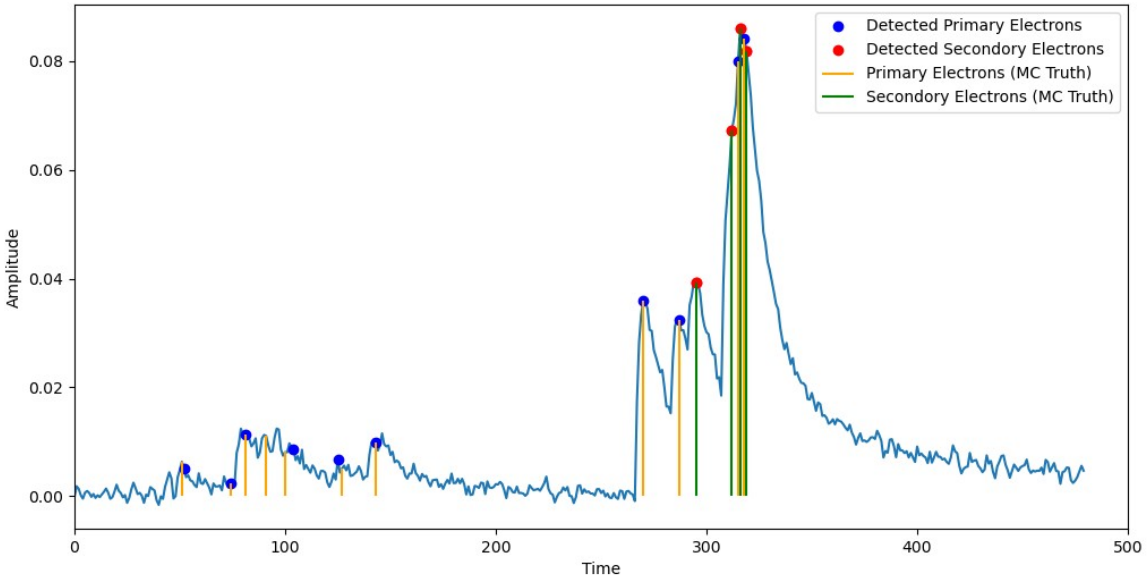
$$\boxed{\text{NoiseAmplitude}_{\text{new}} = \text{NoiseAmplitude}_{\text{current}} \times \frac{\text{RMS}_{\text{data}}}{\text{RMS}_{\text{MC}}}}$$

*Numerical evaluation*

$$\begin{aligned}\text{NoiseAmplitude}_{\text{new}} &= 0.00112 \times \frac{0.001118}{0.012474} \\ &= 0.00112 \times 0.089625 \\ &= 1.0038 \times 10^{-4} \approx \mathbf{0.00010}.\end{aligned}$$

# TUNE SIMULATED Peak finding waveform result by using LSTM Model (AUC =0.97810)



**1. Peak Detection Based on Probability Cut:**
The detection process involves looping over all events and applying a probability cut to decide whether a peak is considered a valid detection:

**Looping Over Events:** The script iterates over all entries (events) in the probability file.

**Applying the Cut:** For each event, it checks if the predicted probability (prob_ml) exceeds the cut threshold (cut), which is set to 0.95/0.65.

**Storing Detected Peaks:** If the probability exceeds the threshold, the corresponding peak time is stored in the detected_time dictionary, keyed by event number (evtno)

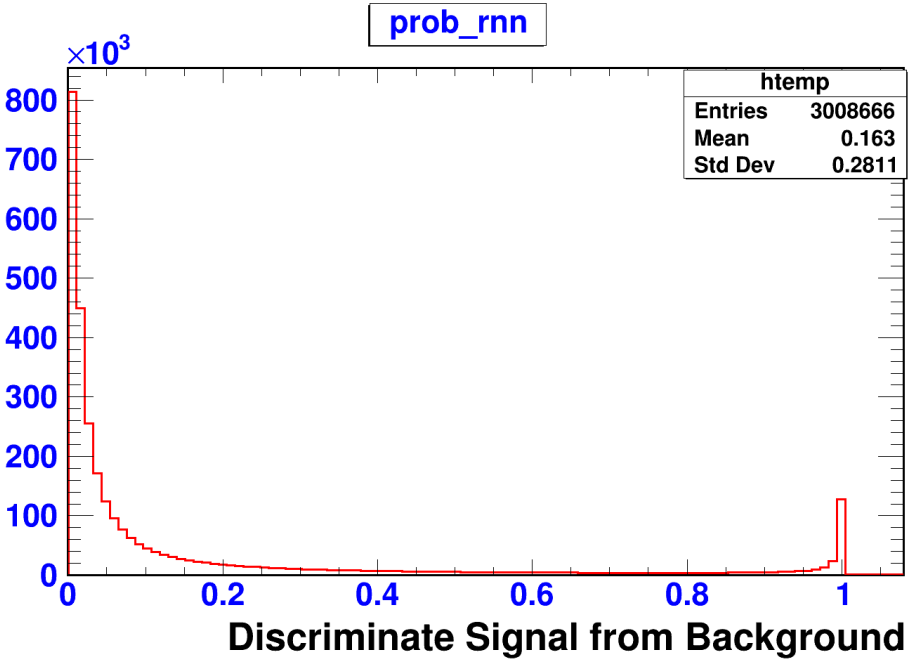
**2. Matching Detected Peaks with Truth Data**
After detecting peaks, the script matches these peaks with the Monte Carlo (MC) truth data to classify them as primary or secondary:

**Truth Data:** The truth data (truth_time, truth_tag) contains the actual times of primary and secondary peaks, labeled by truth_tag (1 for primary, 2 for secondary).

**Matching Function:** The match function compares detected peak times with the truth peak times. For each detected peak, it finds the closest truth peak and assigns the corresponding tag (primary or secondary) based on the truth data.

**ID Assignment:** The id_list array stores the classification of each detected peak as primary (1) or secondary (2)

**3. Counting Primary Peaks:** After classifying the detected peaks, the script counts how many of them are primary peaks:
**Counting:** The script iterates over the id_list and increments ncount_pri for every primary peak (tag 1)



```
muanwar ID: 1216426  11/16 15:27      _      _      1      1 1216426.0
muanwar ID: 1216427  11/16 15:28      _      _      1      1 1216427.0
muanwar ID: 1216428  11/16 15:28      _      _      1      1 1216428.0
muanwar ID: 1216429  11/16 15:28      _      _      1      1 1216429.0
muanwar ID: 1216430  11/16 15:28      _      _      1      1 1216430.0
muanwar ID: 1216431  11/16 15:28      _      _      1      1 1216431.0
muanwar ID: 1216432  11/16 15:29      _      _      1      1 1216432.0
muanwar ID: 1216433  11/16 15:29      _      _      1      1 1216433.0
muanwar ID: 1216434  11/16 15:29      _      _      1      1 1216434.0
muanwar ID: 1216435  11/16 15:29      _      _      1      1 1216435.0
muanwar ID: 1216436  11/16 15:29      _      _      1      1 1216436.0
muanwar ID: 1216437  11/16 15:29      _      _      1      1 1216437.0
muanwar ID: 1216441  11/16 15:30      _      _      1      1 1216441.0
muanwar ID: 1216442  11/16 15:30      _      _      1      1 1216442.0
muanwar ID: 1216443  11/16 15:30      _      _      1      1 1216443.0
muanwar ID: 1216444  11/16 15:30      _      _      1      1 1216444.0
muanwar ID: 1216445  11/16 15:30      _      _      1      1 1216445.0
muanwar ID: 1216446  11/16 15:31      _      _      1      1 1216446.0
muanwar ID: 1216447  11/16 15:31      _      _      1      1 1216447.0
muanwar ID: 1216448  11/16 15:31      _      _      1      1 1216448.0
muanwar ID: 1216449  11/16 15:31      _      _      1      1 1216449.0
muanwar ID: 1216450  11/16 15:31      _      _      1      1 1216450.0
muanwar ID: 1216451  11/16 15:32      _      _      1      1 1216451.0
muanwar ID: 1216452  11/16 15:32      _      _      1      1 1216452.0
muanwar ID: 1216453  11/16 15:32      _      _      1      1 1216453.0
muanwar ID: 1216454  11/16 15:32      _      _      1      1 1216454.0
muanwar ID: 1216455  11/16 15:32      _      _      1      1 1216455.0
muanwar ID: 1216456  11/16 15:33      _      _      1      1 1216456.0
muanwar ID: 1216457  11/16 15:33      _      _      1      1 1216457.0
muanwar ID: 1216458  11/16 15:33      _      _      1      1 1216458.0
muanwar ID: 1216459  11/16 15:33      _      _      1      1 1216459.0
muanwar ID: 1216460  11/16 15:33      _      _      1      1 1216460.0
muanwar ID: 1216461  11/16 15:33      _      _      1      1 1216461.0
muanwar ID: 1216462  11/16 15:34      _      _      1      1 1216462.0
muanwar ID: 1216463  11/16 15:34      _      _      1      1 1216463.0
muanwar ID: 1216464  11/16 15:34      _      _      1      1 1216464.0
muanwar ID: 1216465  11/16 15:34      _      _      1      1 1216465.0
muanwar ID: 1216466  11/16 15:34      _      _      1      1 1216466.0
muanwar ID: 1216467  11/16 15:35      _      _      1      1 1216467.0
muanwar ID: 1216468  11/16 15:35      _      _      1      1 1216468.0
muanwar ID: 1216469  11/16 15:35      _      _      1      1 1216469.0
muanwar ID: 1216470  11/16 15:35      _      _      1      1 1216470.0
muanwar ID: 1216471  11/16 15:35      _      _      1      1 1216471.0
muanwar ID: 1216472  11/16 15:36      _      _      1      1 1216472.0
muanwar ID: 1216473  11/16 15:36      _      _      1      1 1216473.0
muanwar ID: 1216474  11/16 15:36      _      _      1      1 1216474.0
muanwar ID: 1216475  11/16 15:36      _      _      1      1 1216475.0
muanwar ID: 1216476  11/16 15:36      _      _      1      1 1216476.0
muanwar ID: 1216477  11/16 15:37      _      _      1      1 1216477.0
muanwar ID: 1216478  11/16 15:37      _      _      1      1 1216478.0
muanwar ID: 1216479  11/16 15:37      _      _      1      1 1216479.0
muanwar ID: 1216480  11/16 15:37      _      _      1      1 1216480.0
muanwar ID: 1216481  11/16 15:37      _      _      1      1 1216481.0
muanwar ID: 1216482  11/16 15:37      _      _      1      1 1216482.0
muanwar ID: 1216483  11/16 15:38      _      _      1      1 1216483.0
muanwar ID: 1216484  11/16 15:38      _      _      1      1 1216484.0

Total for query: 2475 jobs; 20 completed, 0 removed, 1677 idle, 778 running, 0 held, 0 suspended
Total for muanwar: 2475 jobs; 20 completed, 0 removed, 1677 idle, 778 running, 0 held, 0 suspended
Total for all users: 7481 jobs; 573 completed, 1176 removed, 2199 idle, 3533 running, 0 held, 0 suspended
```

```
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model955/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model956/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model957/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model958/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model959/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model96/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model960/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model961/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model962/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model963/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model964/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model965/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model966/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model967/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model968/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model969/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model97/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model970/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model971/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model972/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model973/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model974/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model975/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model976/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model977/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model978/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model979/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model98/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model980/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model981/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model982/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model983/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model984/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model985/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model986/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model987/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model988/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model989/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model99/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model990/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model991/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model992/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model993/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model994/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model995/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model996/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model997/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model998/model_metrics.txt
Processing File: /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model999/model_metrics.txt

=============== BEST MODEL ===============
Model: Model1840
AUC:   0.978100
File:  /lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/Model1840/model_metrics.txt
==========================================

Saved full table to: all_model_aucs.csv
/lustre/home/muanwar/NEW/NN/ClusterCounting/New_Opt_LSTM/find_best_auc.py:107: UserWarning: This figure includes Axes that are not compatible with tight_layout, so resu
  plt.tight_layout()
Saved plot: auc_scatter_with_zoom.png
Saved zoom figure: auc_zoom.png
```