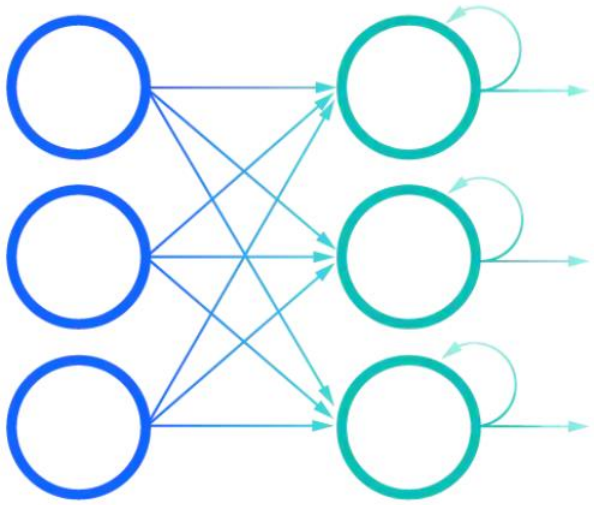
Status of Peak Finding Algorithm

Guang Zhao, Siman Liu, Shuiting Xin, Linghui Wu, Mingyi Dong

zhaog@ihep.ac.cn

March 17, 2022

Recap: Peak finding with deep learning



Machine Learning:

- "Learn" the characteristics of data automatically by the machine

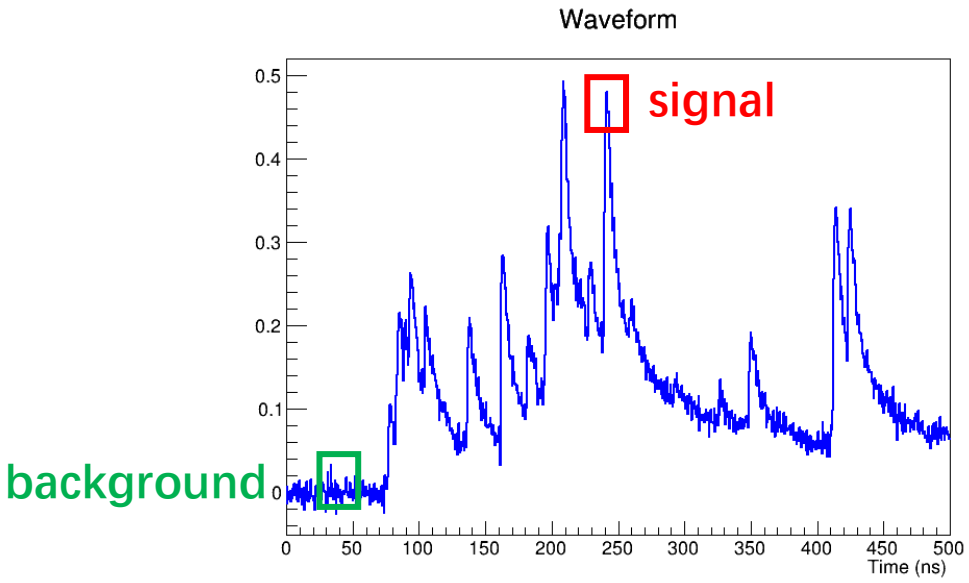
Recurrent Neural Network (RNN):

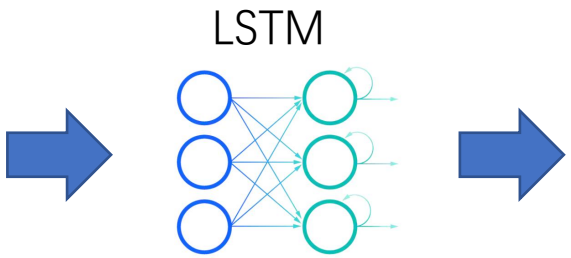
- Internal loops over sequence elements. Has "memories"
- Powerful to handle time-sequence problems

Waveform Peak Finding:

- Need to classify "peaks" and "noises": classification problem
- Time series information: appropriate for RNN

Model



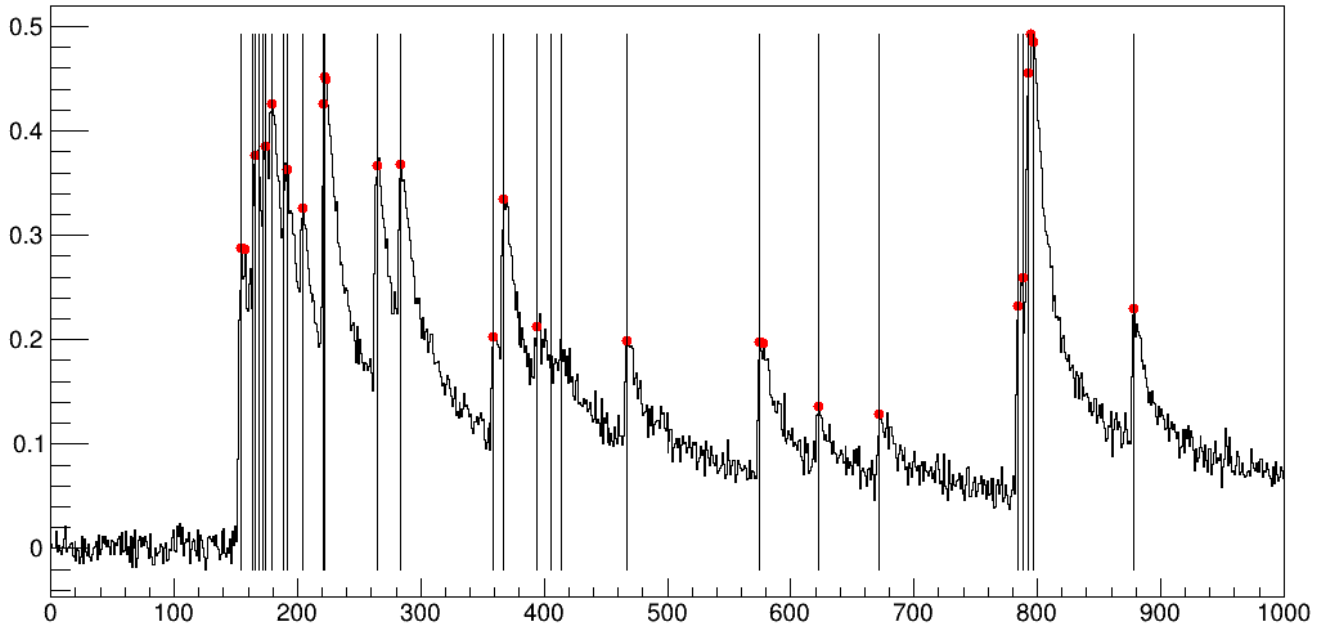


Output probability (classifier)

- Using slices of the waveform as the inputs
- Using LSTM (a variation of RNN) to train the dataset
- Preliminary results show good performances (training has not been fully optimized yet)

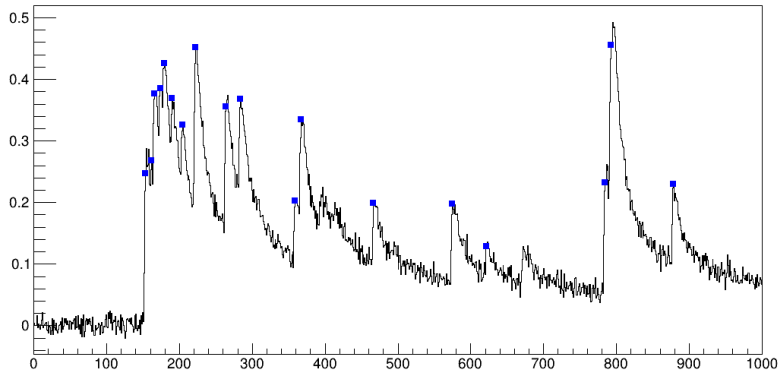
Peak finding with toy MC

RNN (LSTM)



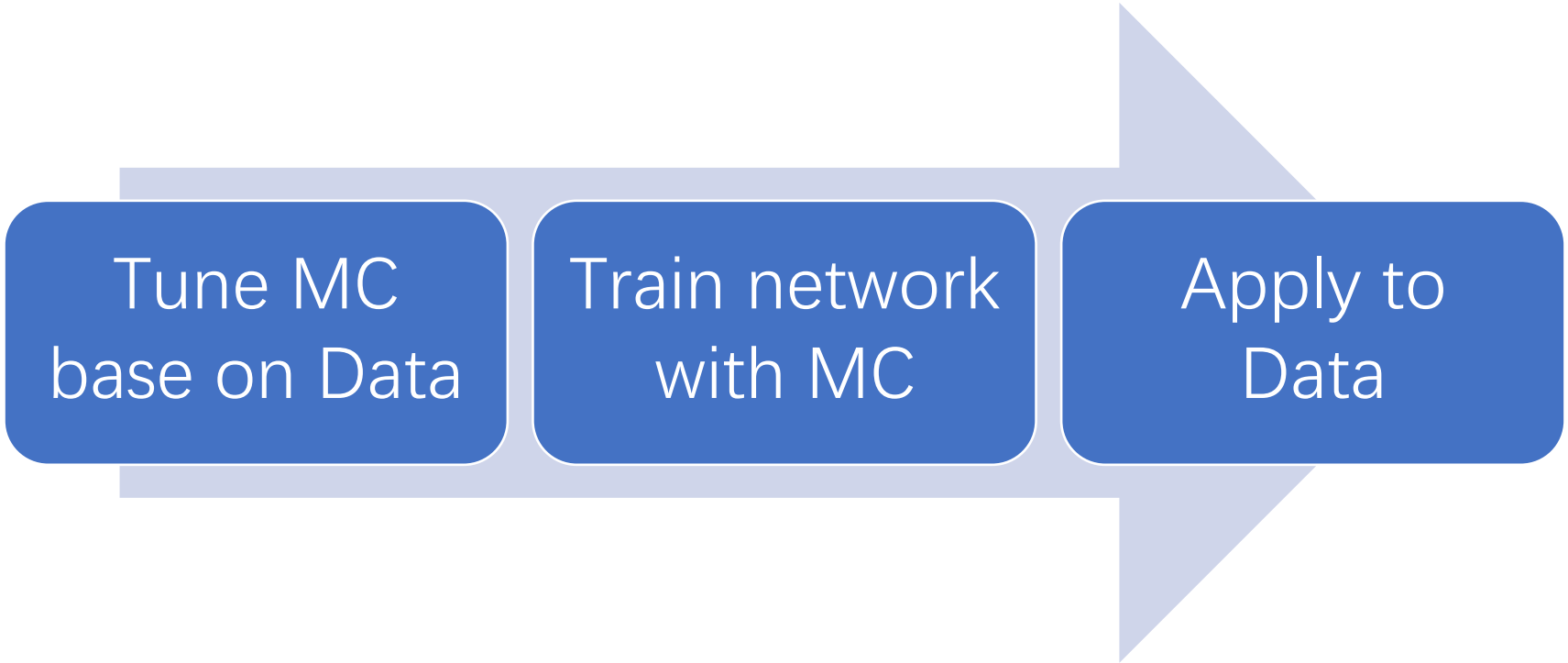
Black line: truth (primaries and secondaries)

Derivative



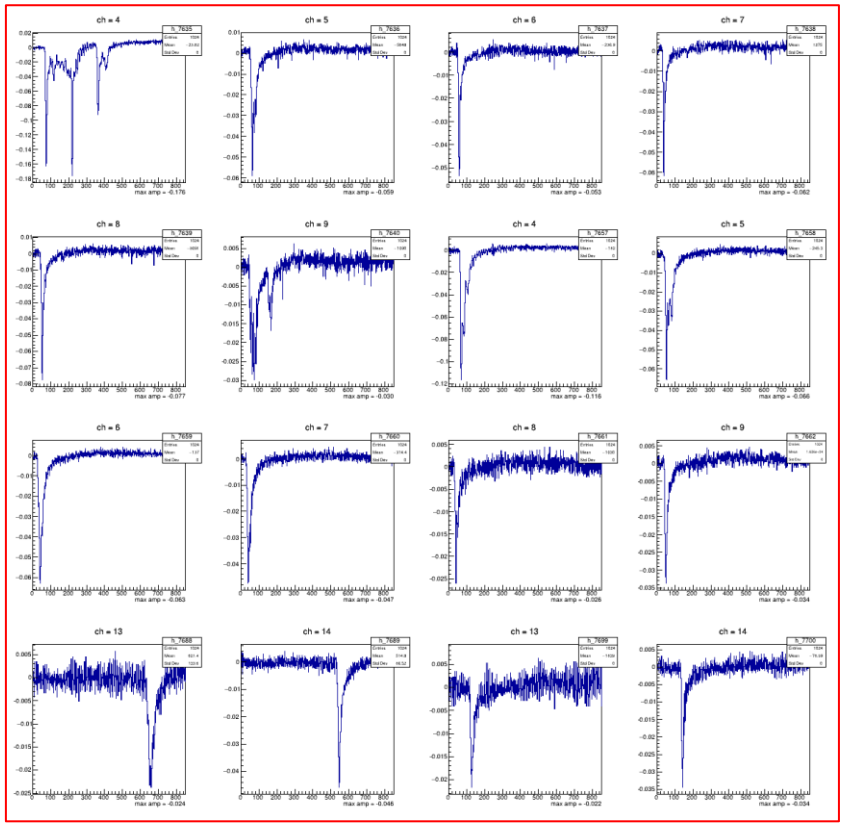
Strategy

- **How to apply to the real data?**

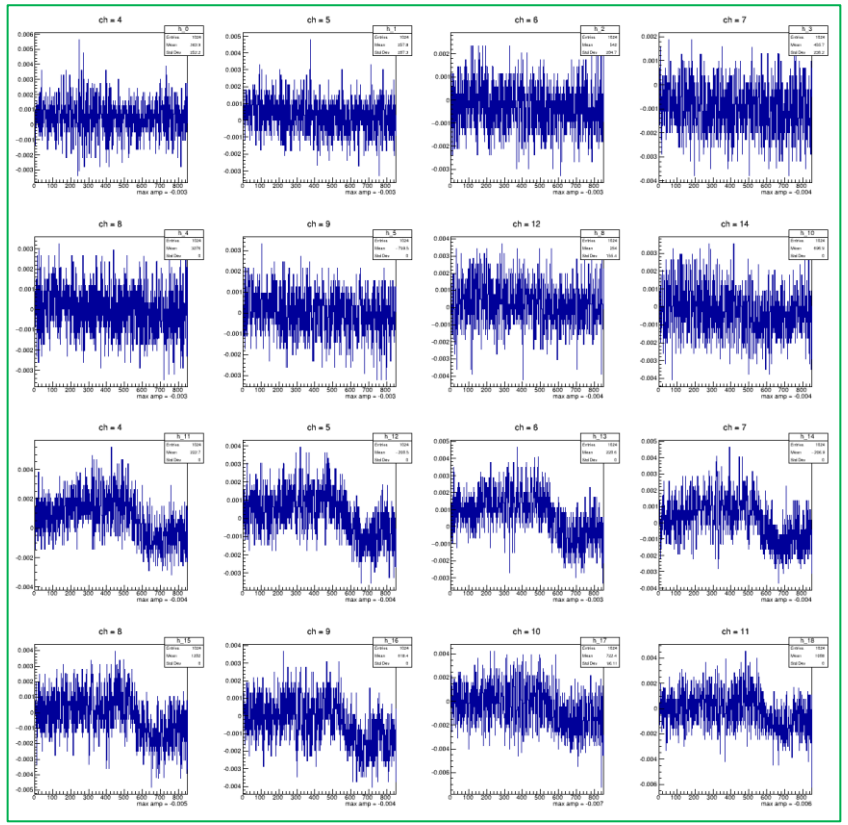


Beam test data selection

Signal-like data

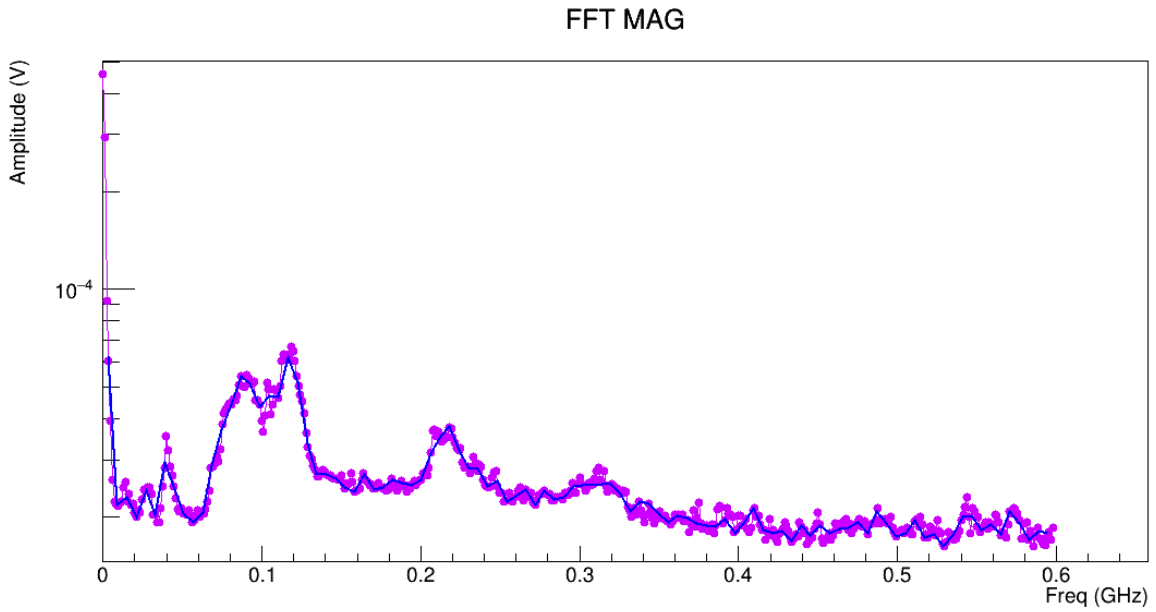


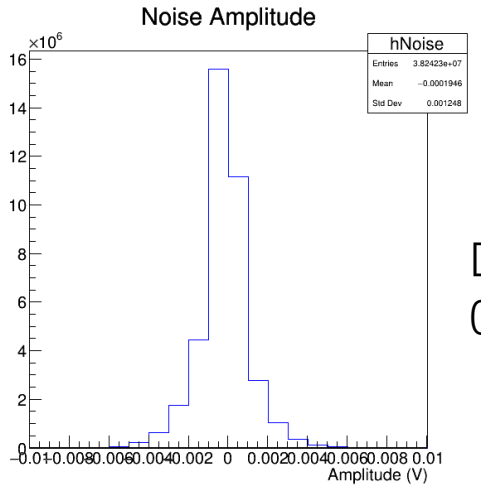
Noise-like data



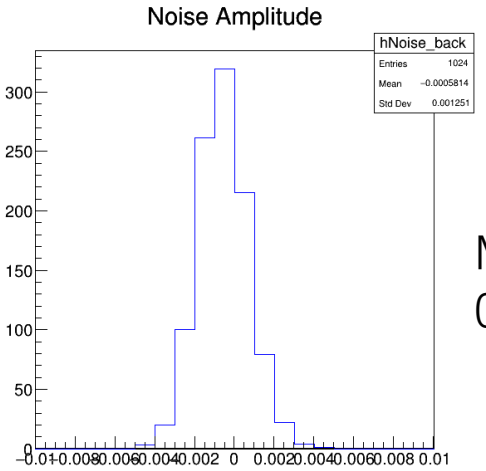
- Signal-like data: rising-time, average amplitude, peak finding
- Noise-like data: noise information

Noise extraction





Data RMS:
0.001248

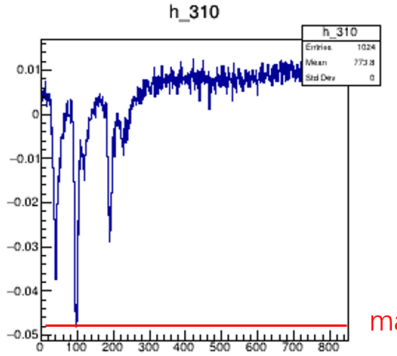


MC RMS:
0.001251

- Extract noise frequency response from FFT
- Generate noise by performing iFFT with a random phase

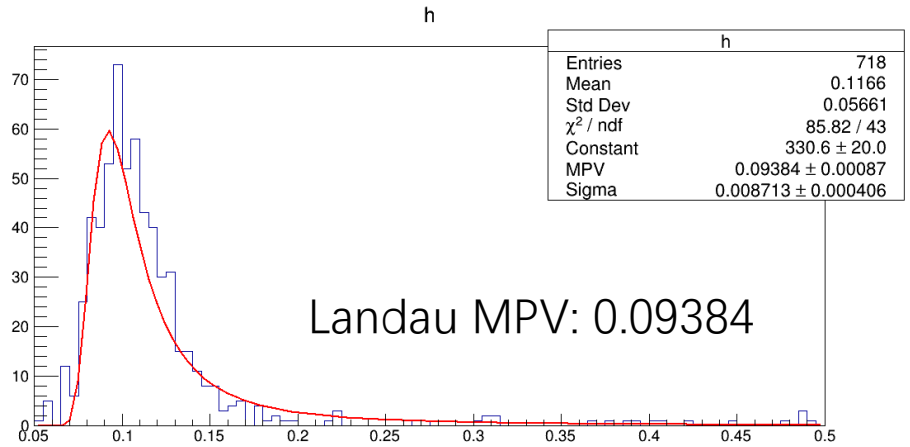
Amplitude measurement

- **Scale the amplitude of MC by comparing to data**

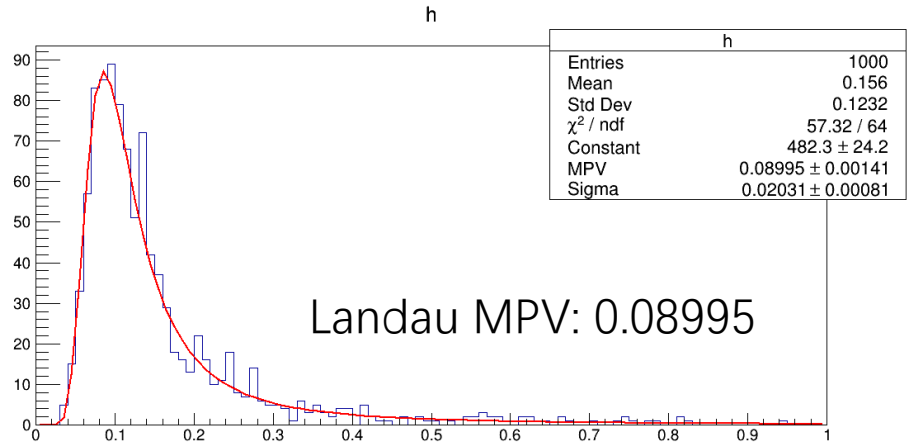


maximum height

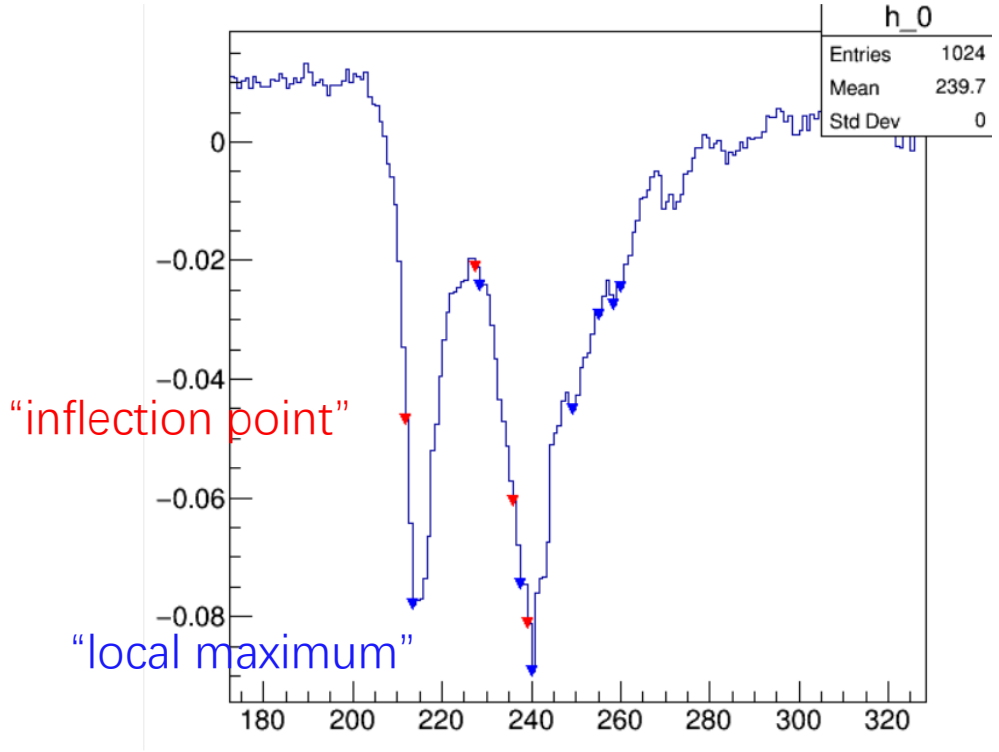
Average max. height in Data



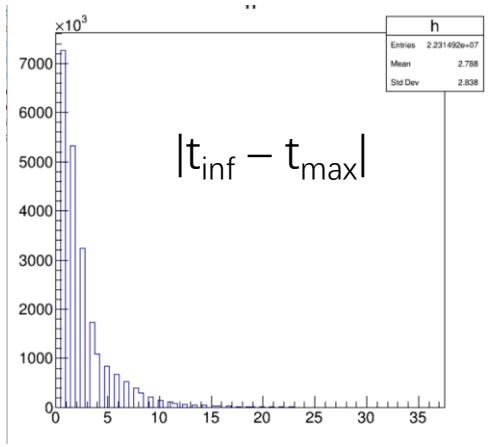
Average max. height in scaled MC



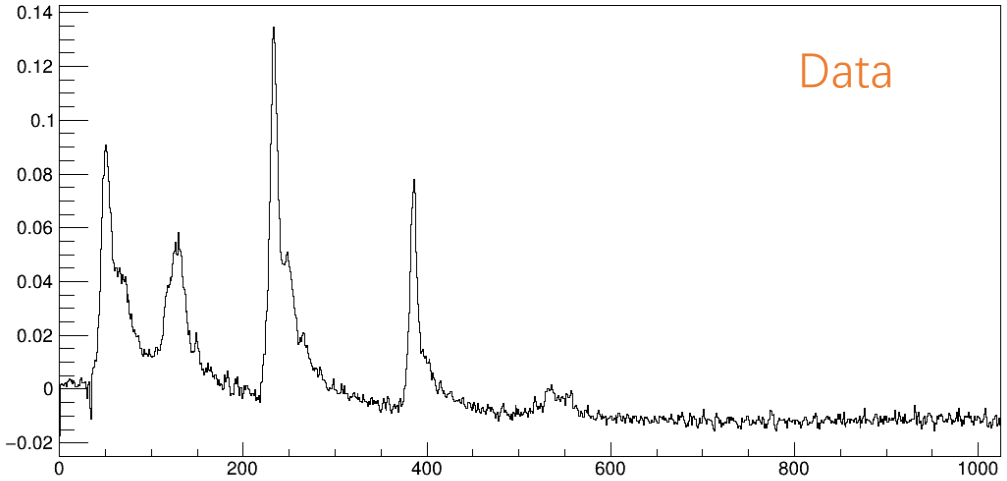
Rough rising-time estimation

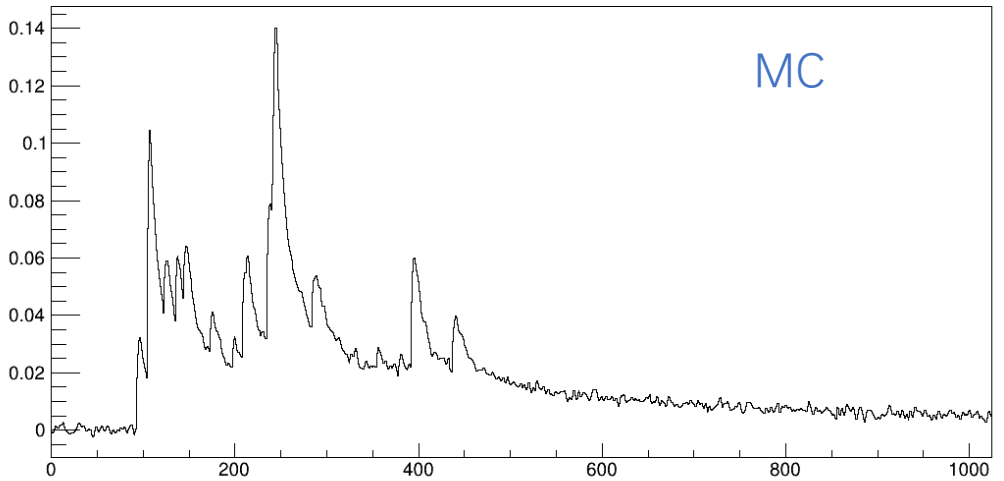


- Measure time between the “inflection” and the “local maximum”
- Rising time (very rough): ~4 ns
- Put this rising time in MC simulation



Updated MC waveform





Much better agreement with Data

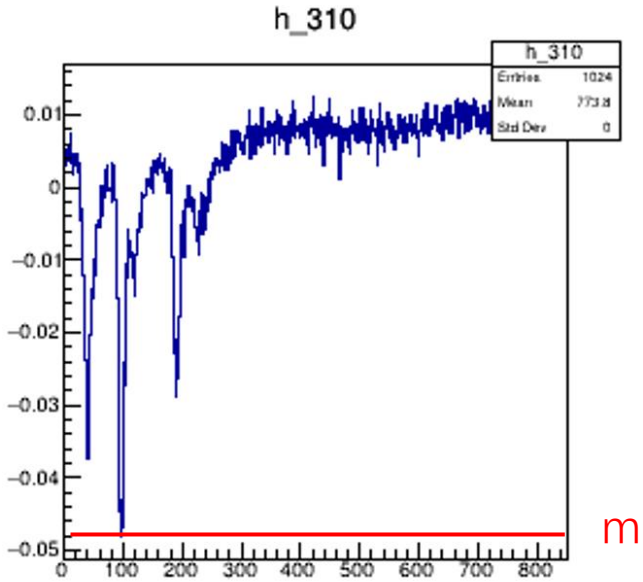
Summary

- **Updated the MC model based on beam test data**
- **Will try to train the peak finding algorithm with the new MC and apply to data**

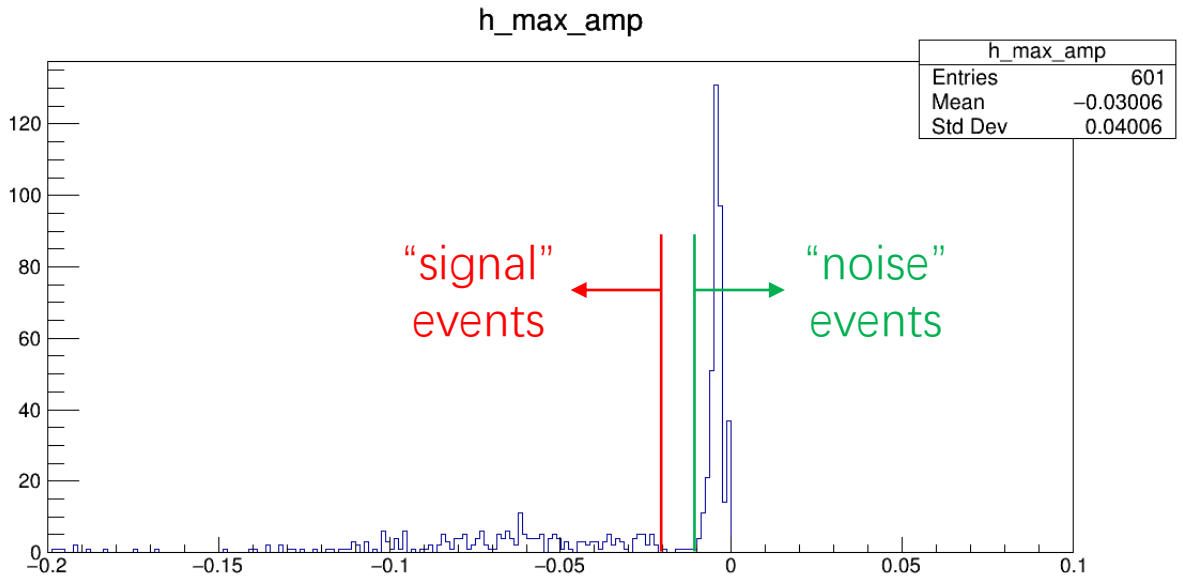
Thank you

Backup

Data preprocessing: classification



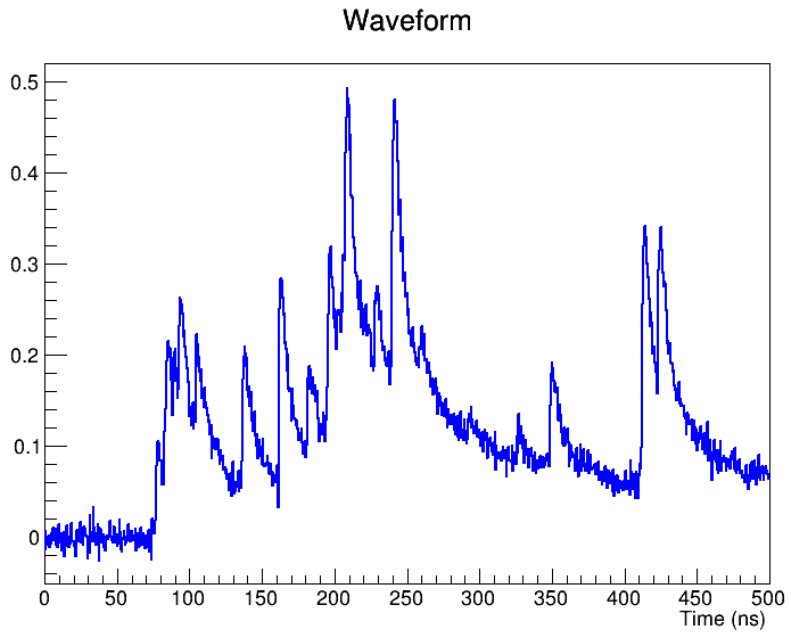
maximum height

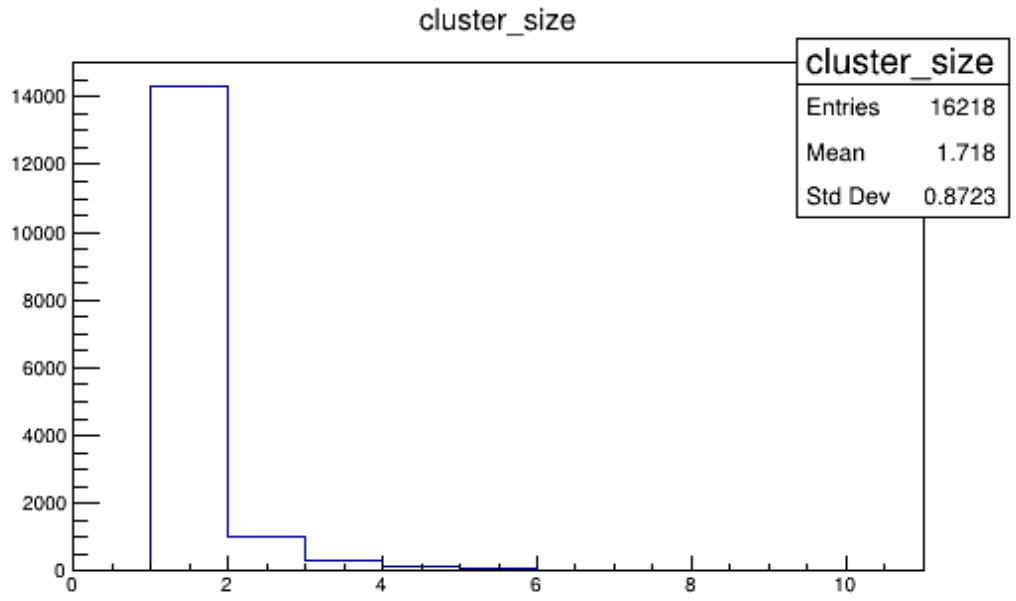


Classify waveforms to “signals” and “noises”

Dataset

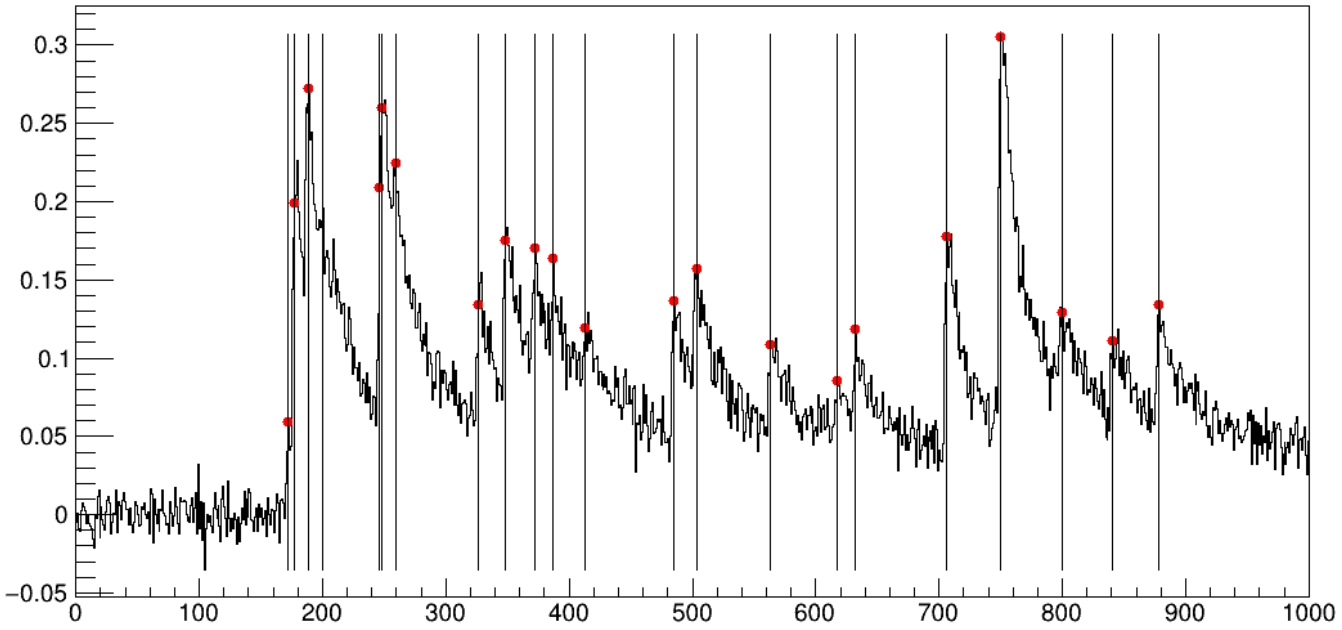
- **Toy Monte Carlo samples**
  - ~20 primary ionizations per waveform (~90% cluster size == 1)
  - ~10% noise level
  - 10000 samples



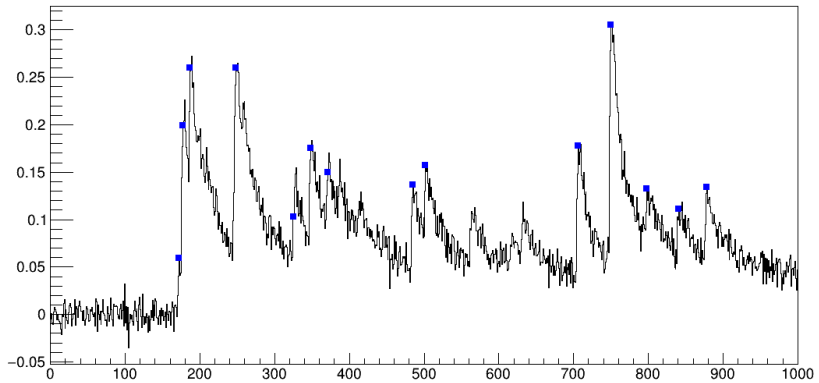


Peak finding (II)

RNN (LSTM)



Derivative



Black line: truth (primaries and secondaries)