Cluster Counting Algorithm Based on Machine Learning

Zhefei Tian[1], Guang Zhao[2], Shuiting Xin[2], Zhenyu Zhang[1], Linghui Wu[2]

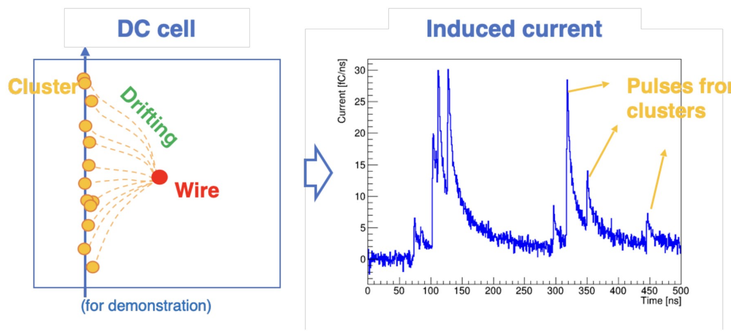
[1]Wuhan University, China

[2]Institute of High Energy Physics, China

Introduction

PID Technique

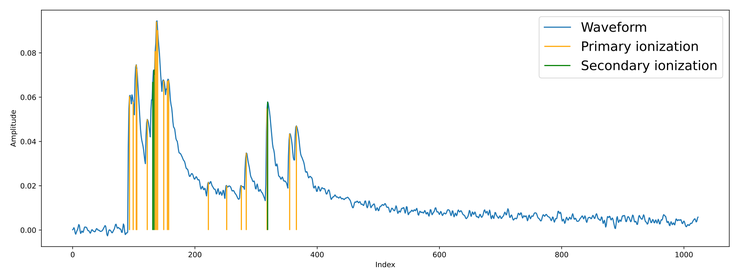
- The prior physics objective of CEPC is precise Higgs measurement, which requires powerful PID, better than 2σ π/K separation at $p < 20$ GeV[1].



- Using the $\mathrm{d}N/\mathrm{d}x$ technique instead of the $\mathrm{d}E/\mathrm{d}x$, we can have better PID[2].
  - $\mathrm{d}E/\mathrm{d}x$: Energy loss per unit length, Landau distribution, large fluctuation.
  - $\mathrm{d}N/\mathrm{d}x$: Number of primary ionization clusters per unit length, Poisson distribution, small fluctuation.
- **Measure the number of primary ionization clusters (N_p) $\Rightarrow$ Cluster counting algorithm.**

Cluster Counting Algorithm

- Cluster counting algorithm: Determine the N_p from the induced current waveform.

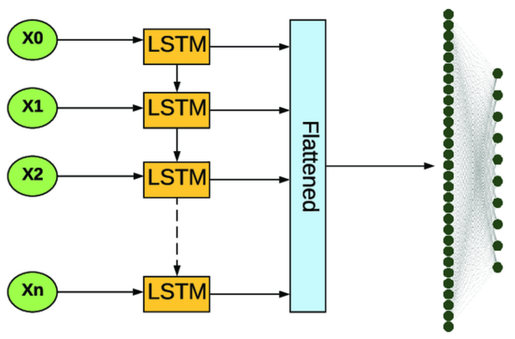


- Both primary and secondary ionizations contribute peaks on the waveform. What we need is the number of primary ionization clusters.
- Two steps: **Peak finding algorithm** and N_p **determination algorithm**.
- Machine Learning with **TensorFlow** and **Keras**.

Peak Finding Algorithm with LSTM

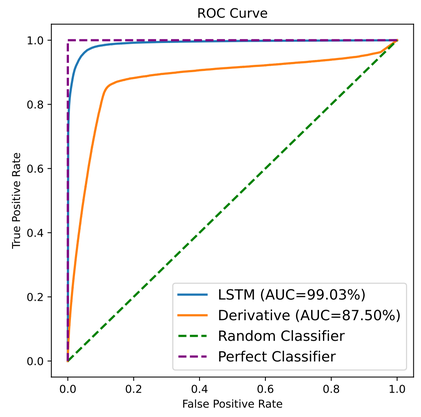
- **Peak Finding Algorithm**: Find all ionization peaks from waveforms.
- A classification problem on slicing windows (peak candidates) samples.
- Time-sequence data $\Rightarrow$ suitable for LSTM model.

Neutral Network Structure

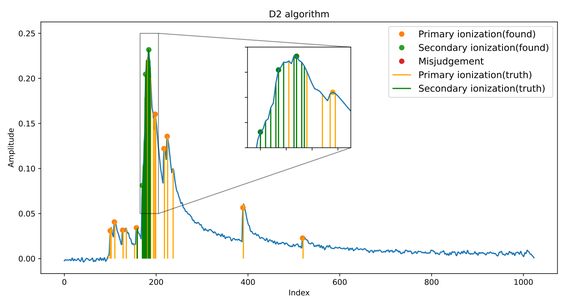


- Main block: Long Short-Term Memory (LSTM), a special kind of Recurrent Neutral Network (RNN).

Results



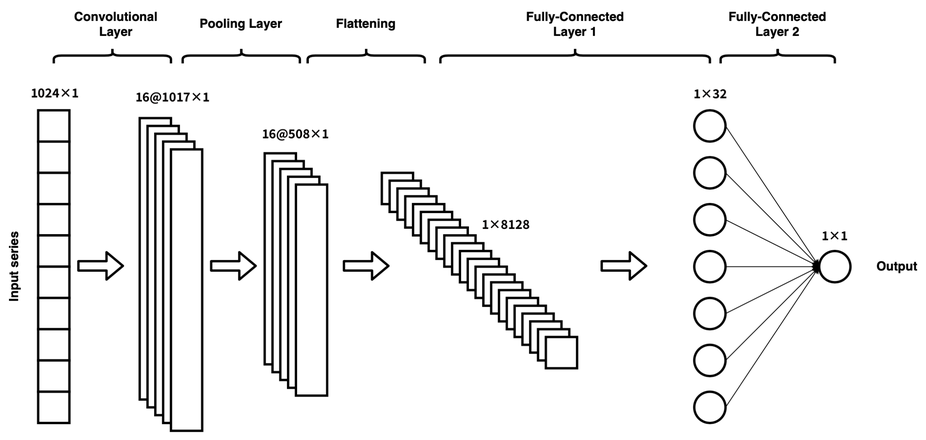
- ROC curve with larger area under-curve (AUC) is better.
- LSTM-based classifier is much better than the one with traditional derivative method, especially for pile-up situations.



N_p Determination Algorithm with CNN

- N_p **Determination Algorithm**: Determine N_p from detected times in peak finding.
- A regression problem to predict N_p.
- 1D data series $\Rightarrow$ suitable for 1D-CNN model.

Neutral Network Structure



- Main block: Convolutional Neural Network (CNN).

Results

- Single cell resolution $\sim 22.8\%$ (close to truth 22.3%).
- Very good Gaussian distribution of prediction.

Summary and conclusions

- A two-step cluster counting algorithm with Machine Learning is developed.
- The algorithm is able to achieve a resolution close to the truth level, which is better than the algorithm based on derivatives.
- NEXT TO DO: Make the full evaluation of the algorithm and apply the algorithm to the experimental data.

Ref

[1] CEPC Study Group, [arXiv:1811.10545 [hep-ex]].

[2] Walenta, A. H., *IEEE Trans. Nucl. Sci.* **26** (1979) 73-80.

Contact

Zhefei Tian (Wuhan University)

E-mail: tianzhefei@whu.edu.cn