dN/dx reconstruction in CEPC

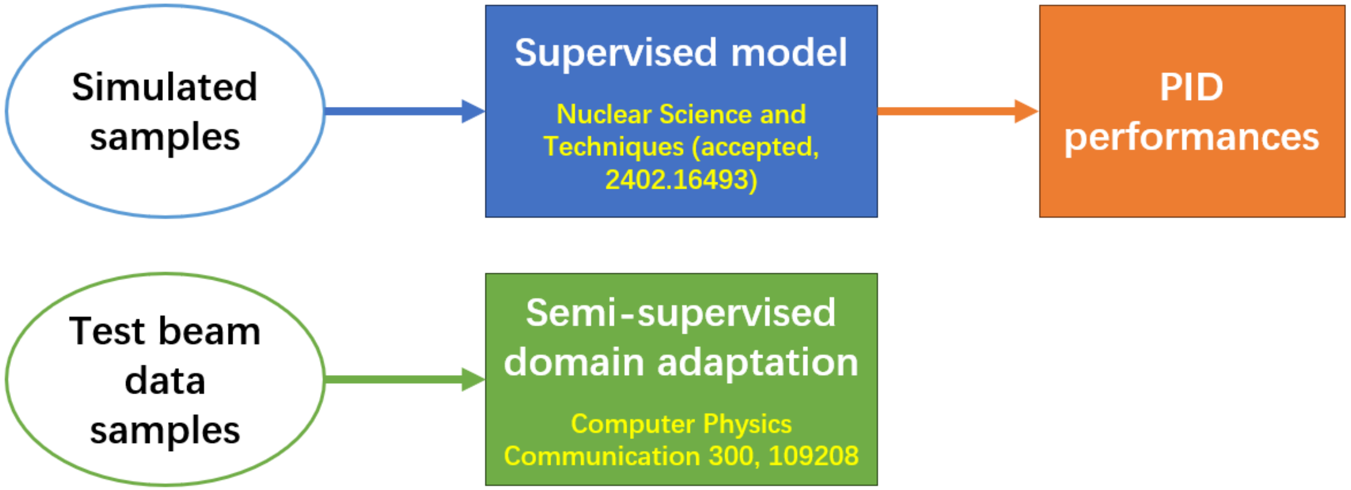
Guang Zhao (zhaog@ihep.ac.cn)

Experimental Physics Center

October 16, 2024

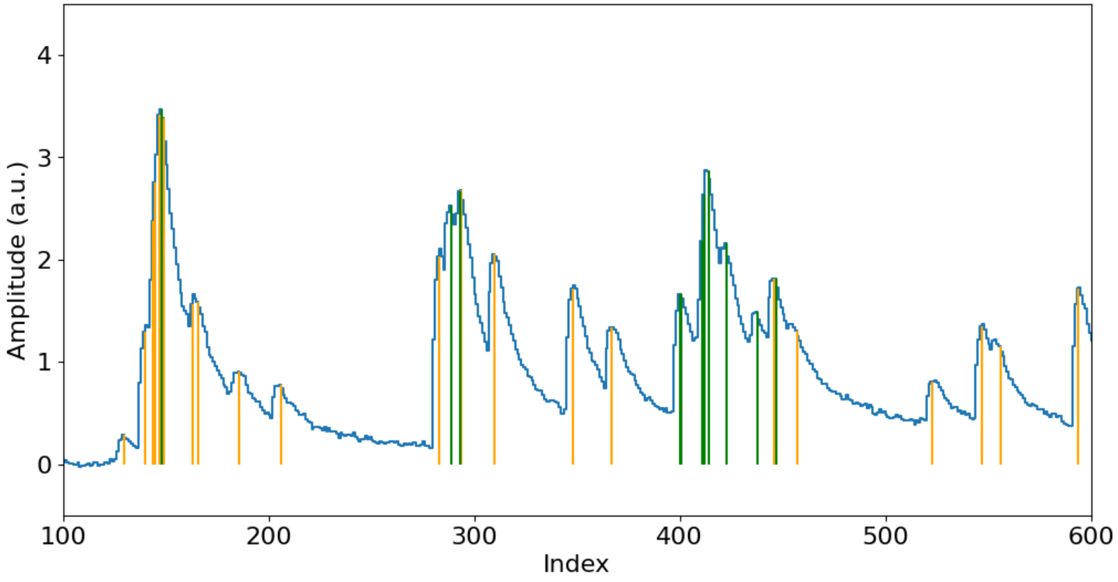
Introduction

- **Motivation:**
  - PID is essential for future collider experiments
  - dN/dx in drift chamber is a breakthrough in PID technique, which is proposed in both CEPC and FCC-ee
  - Reconstruction is a major challenge for dN/dx measurement
- **Main works:**



dN/dx reconstruction

Orange lines: Primary electrons (MC truth)
Green lines: Secondary electrons (MC truth)



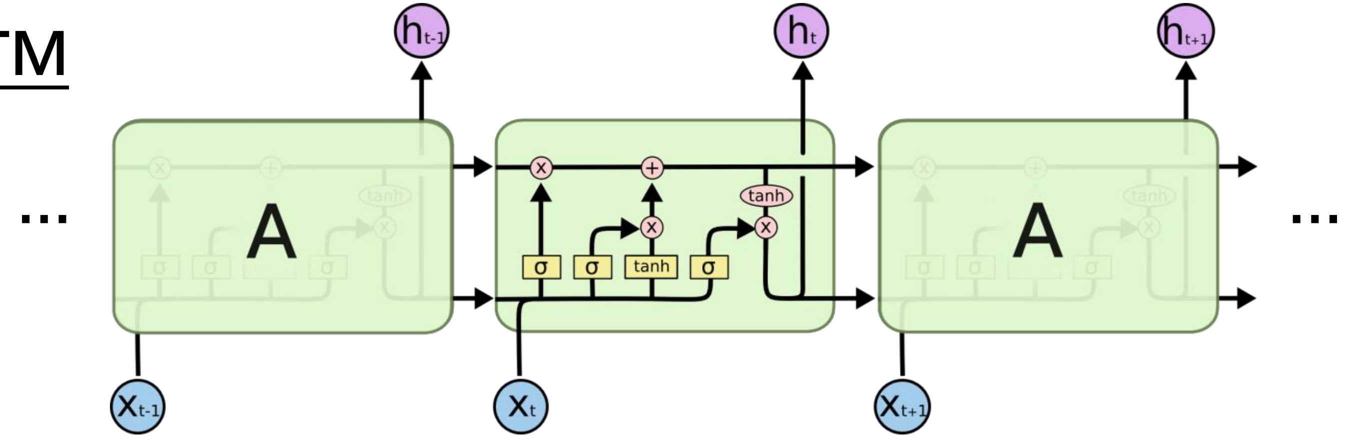
- **Goal:** Determine the number of **primary electrons** in the waveform

- **2-step algorithm:**
    - **Peak finding:** Detect peaks from both primary and secondary electrons
    - **Clusterization:** Remove secondary electrons from the detected peaks in step 1

- **Challenges:**
    - **Highly piled-up ➔** Difficult to efficiently detect pile-ups
    - **Noisy ➔** Filtering could (significantly) lose efficiency
    - **Overlapping between clusters ➔** Difficult to set a simple “cut” for clusterization

Alg. 1: Supervised model for simulation

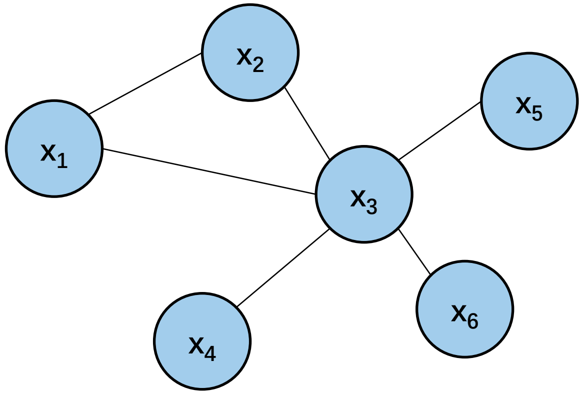
LSTM



LSTM-based peak finding:

- Can efficiently handle time-sequence
- Waveform slices as the LSTM input
- Binary classification of signals and noises

DGCNN



Dynamically connected Graph by k-NN

DGCNN-based clusterization:

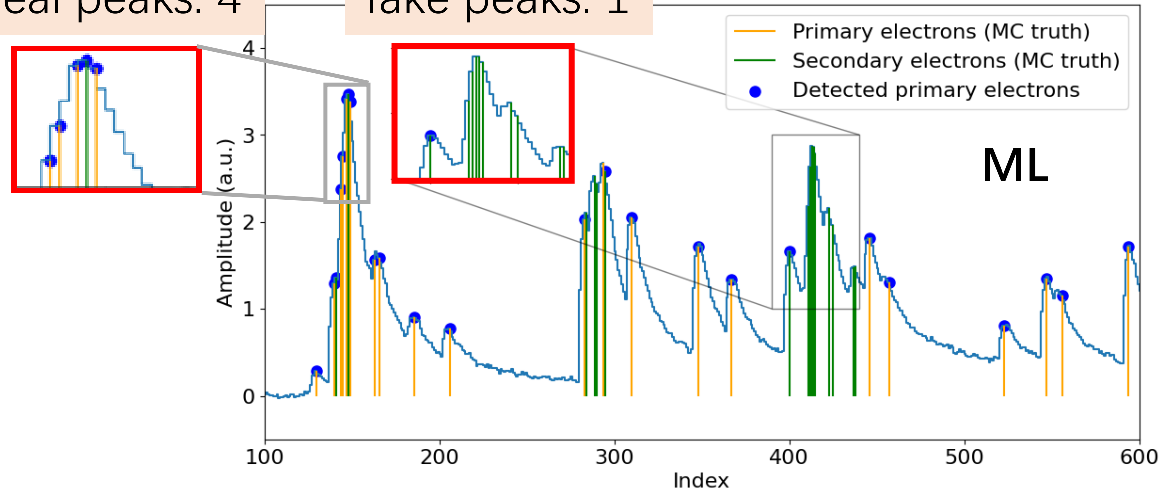
- Incorporate local information to learn global properties
- Detected timings from the peak-finding as the DGCNN input
- Binary node classification of primary and secondary electrons

Nuclear Science and Techniques (accepted, 2402.16493)

Results for supervised model

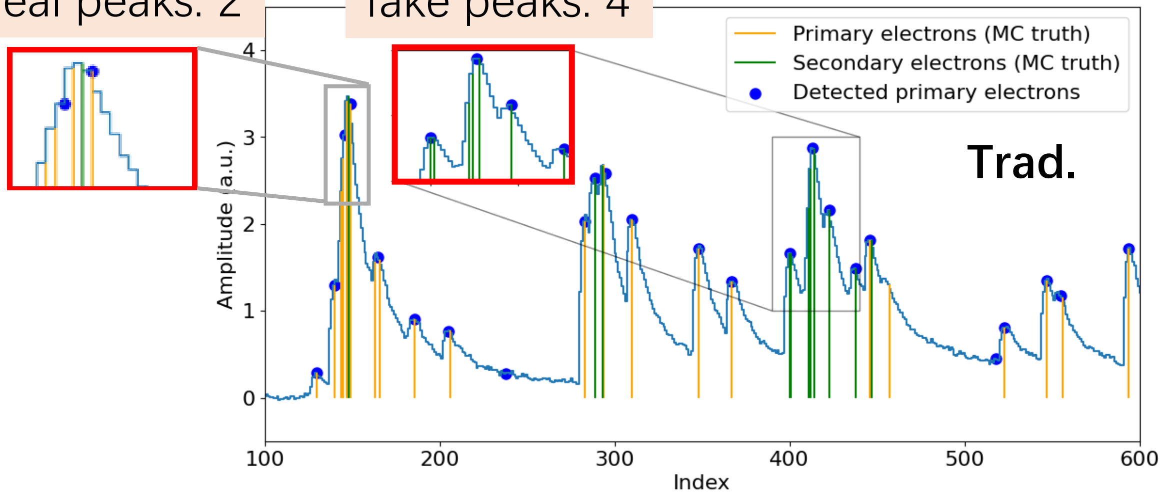
real peaks: 4

fake peaks: 1

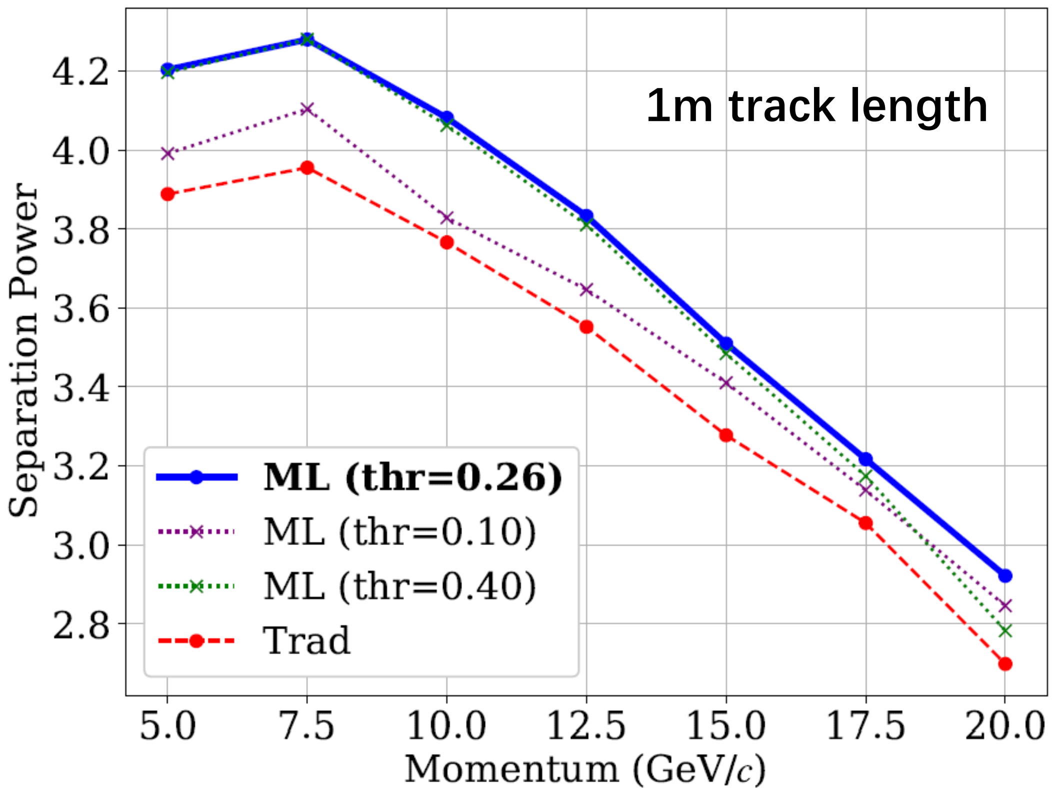


real peaks: 2

fake peaks: 4



K/π separation power vs. momentum



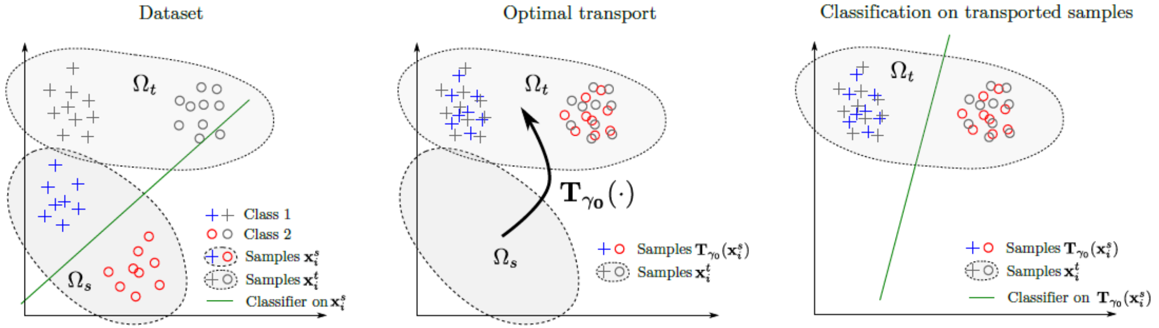
~10% improvement for ML (equivalent to a detector with 20% larger radius for trad. algorithm)

Alg. 2: Domain adaptation for real data

Challenges for real data: 1) Imperfect simulation 2) Incomplete labels in real data

Solution: Domain adaptation (DA)

- Align data/MC samples with **Optimal Transport**



Algorithm: Semi-supervised DA

Loss for labeled samples in source domain

$$\min_{f,g}\left[\sum_{i=1}^{m} L_s\left(y_i^s, f(g(x_i^s))\right) + \boxed{\frac{1}{m_l}\sum_{i=1}^{m_l} L_t\left(y_i^{t,l}, f\left(g(x_i^{t,l})\right)\right)} + \min_{\gamma\in\Delta}\sum_{i,j}^{m}\gamma_{ij}\left(\alpha\|g(x_i^s)-g(x_j^t)\|^2 + \lambda_t L_t\left(y_i^s, f\left(g(x_j^t)\right)\right)\right)\right]$$

Loss for labeled samples in target domain (THIS WORK)

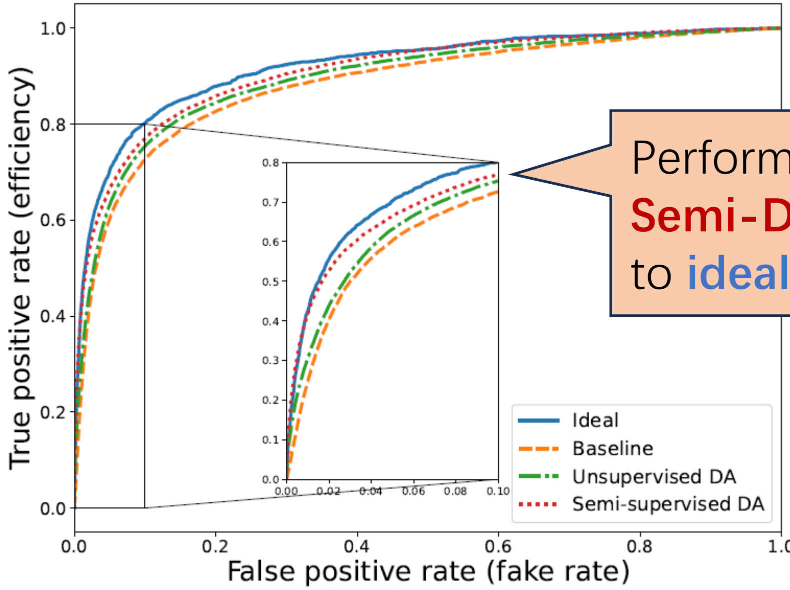
Cost of feature differences between source and target

Cost of 'label' differences between source and target

Cost of joint feature-label distribution for OT

Computer Physics Communication 300, 109208

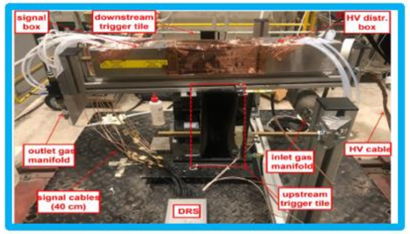
Validation with pseudo data



Performance of **Semi-DA** is closed to **ideal model**

Peak finding for test beam data

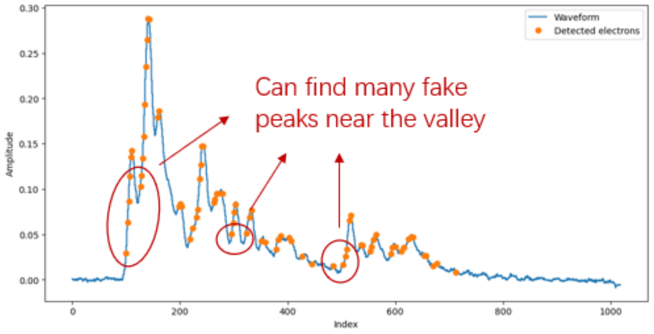
Test beam at CERN



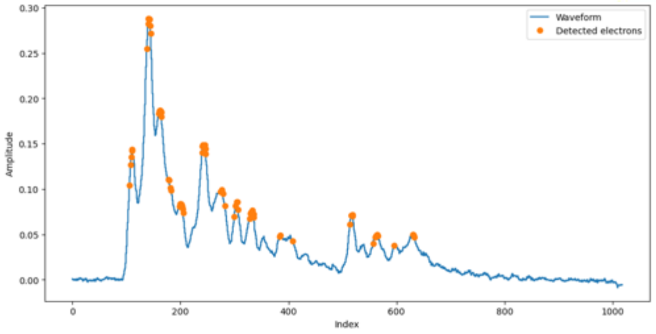
From INFN group leaded by Franco Grancagnolo and Nicola De Filippis

Single-waveform results between derivative alg. and DL alg.

Derivative



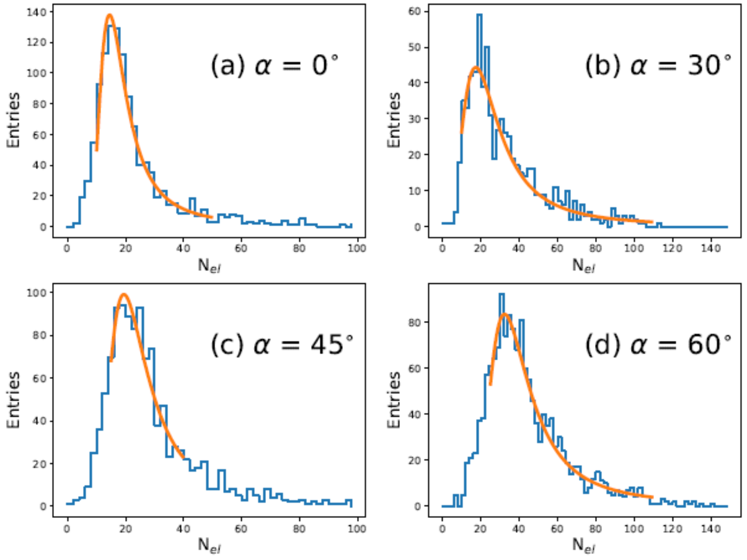
Semi-DeepJDOT

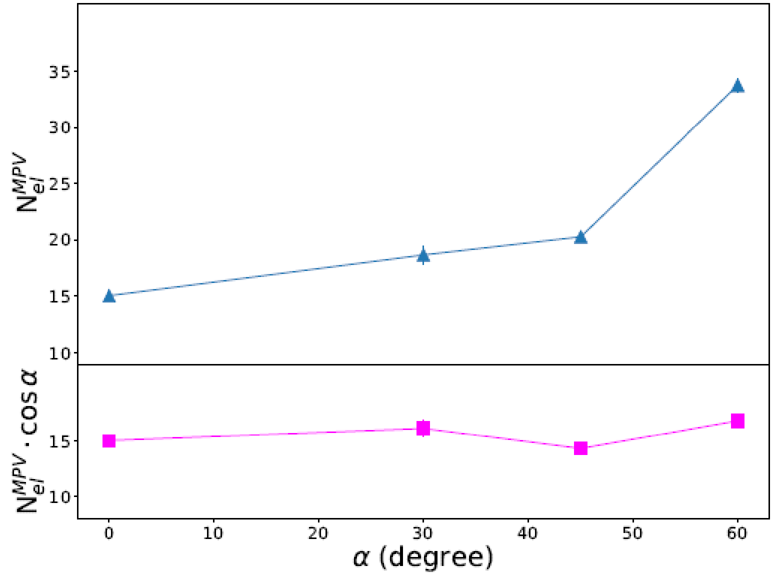


Note: Require similar efficiency for both cases

DL algorithm is more powerful to discriminate signals and noises

Multi-waveform results for samples in different angles





Scale w.r.t. track length

The algorithm is stable w.r.t. track length

Conclusion

- **dN/dx method is the breaking through for PID. Two machine learning algorithms are developed:**
  - **Supervised model:** **10% improvement** for ML methods
  - **Domain adaptation model:** Successfully **transfer information from simulation** and achieve stable performances

- **The algorithms will be utilized in test beam data analysis and can also be ap plied to similar signal processing tasks**

- **Published papers:**
  - Computer Physics Communication 300, 109208
  - Nuclear Science and Techniques (accepted, 2402.16493)