

2024年度考核报告

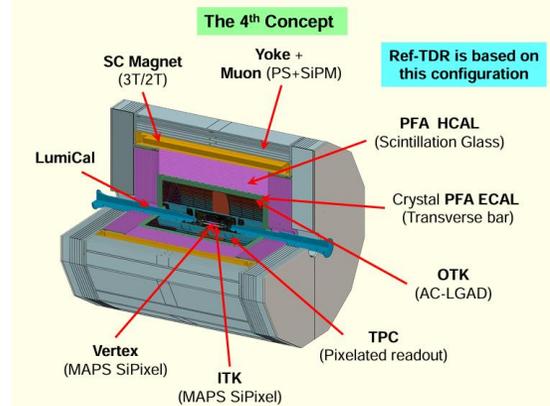
软件组：赵光

2024年11月22日

报告内容

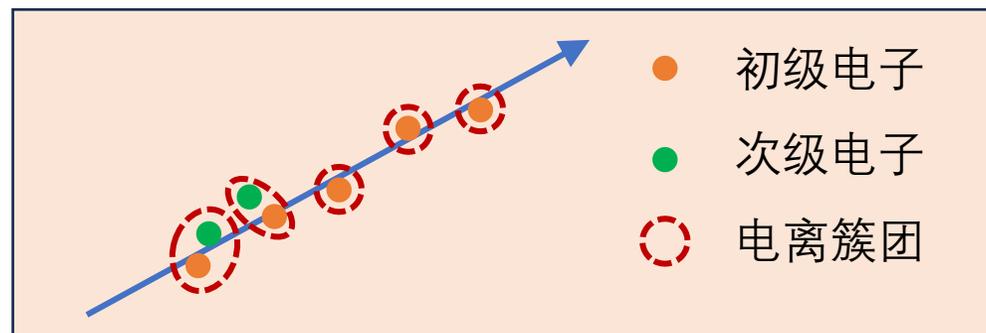
- **本年度工作情况**
 - CEPC 软件 (时间投影室和漂移室)
 - BESIII 数据分析
- **学术成果/科研项目/公共服务**
- **下年度工作计划**

CEPC 气体探测器 dN/dx 方法的研究



研究意义:

- 对新型探测器技术的探索。从时间投影室和漂移室两个技术方案入手，深入研究 dN/dx 方法。对提升 CEPC 粒子鉴别能力和实验精度意义重大。

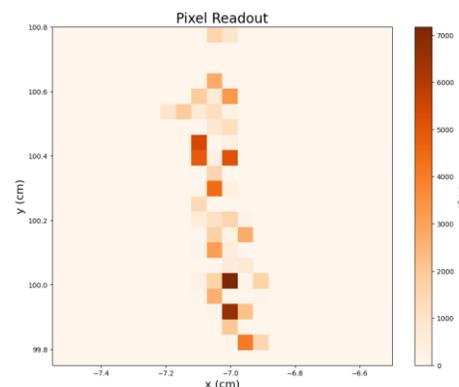


空间

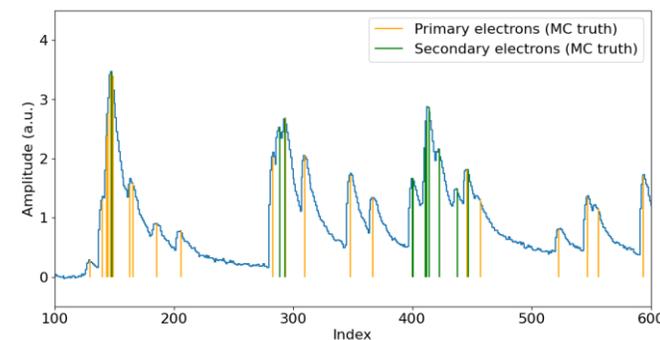
时间

主要贡献:

- 主导时间投影室和漂移室 dN/dx 方法研究的软件开发、关键算法研究、模拟性能研究和国际合作



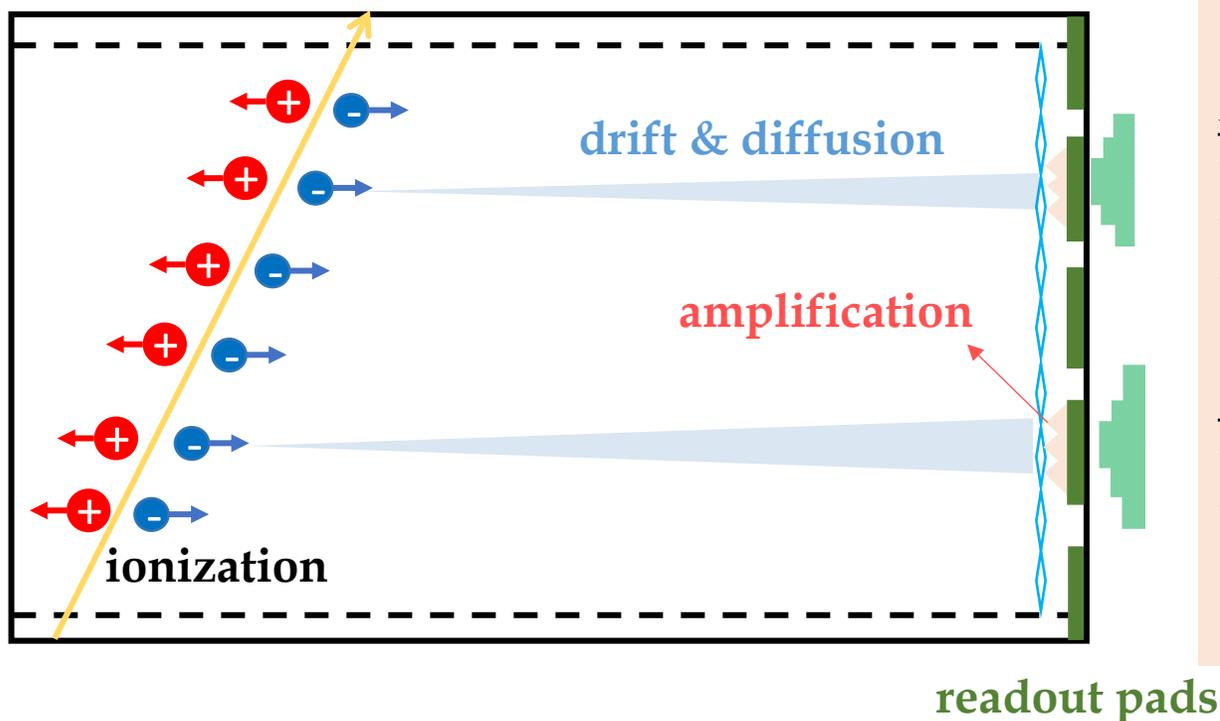
时间投影室



漂移室

时间投影室 dN/dx 方法的研究

工作重点： 新任务。完成精细模拟链条，为 CEPC 物理提供关键粒子鉴别信息



全模拟框架

基于 Garfield++ 的模拟:

- 全尺寸 TPC
- 基于 Heed 的电离模拟
- **电子漂移扩散效应**

实验驱动的数字化的:

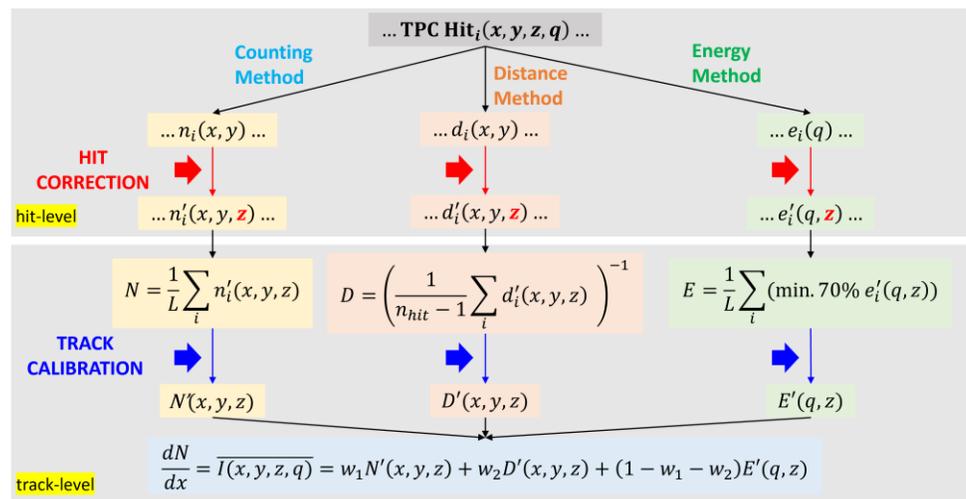
- **电子学噪声:** 100 e⁻
- **MicroMegas 放大:**
 - 电子数增益: 2000
 - 电子空间展宽: 200 um (-3~+3 sigma)

时间投影室 dN/dx 方法的研究

新算法 (考虑漂移距离、角度的修正和刻度)

重建算法:

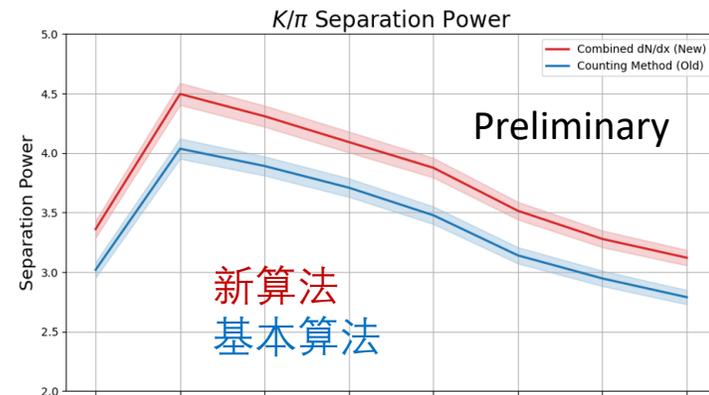
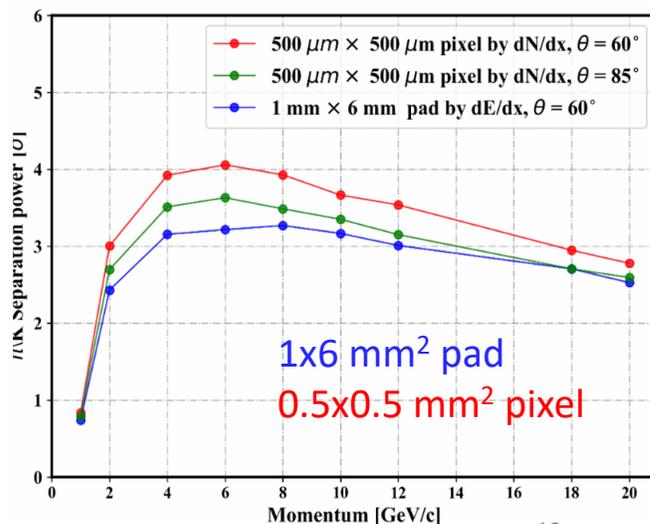
- 基于像素计数的**基本算法**
- 基于三维位置和能量的四维信息的**新算法 (研究中)**



基于全模拟的粒子鉴别性能:

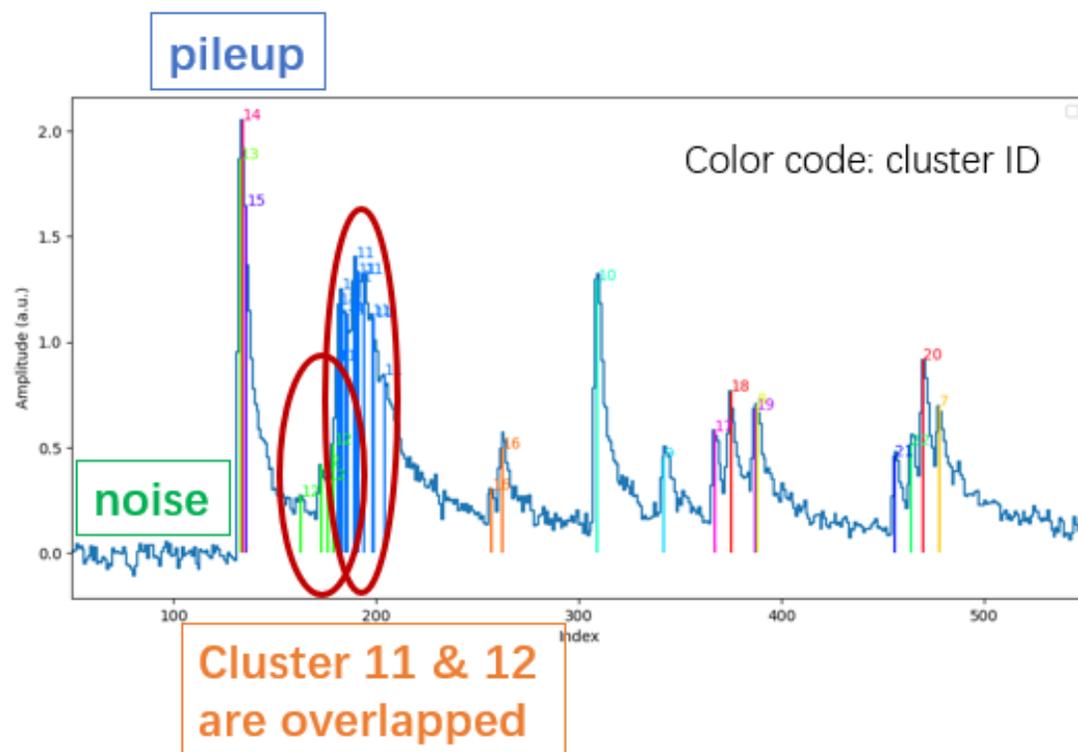
- 像素读出的粒子鉴别性能优于传统 pad 读出
 - 20 GeV/c 动量的 K/pi 粒子区分能力接近 3σ 。
- 新算法有进一步提高性能的潜力

K/pi 粒子区分能力



漂移室 dN/dx 方法的研究

工作重点： 前期已开展了完整的模拟研究。今年重点解决重建算法的挑战。



dN/dx 重建存在巨大挑战，主要体现在：信号高堆积、高噪声污染和次级电子的干扰

漂移室 dN/dx 方法的研究

关键机器学习算法的开发:

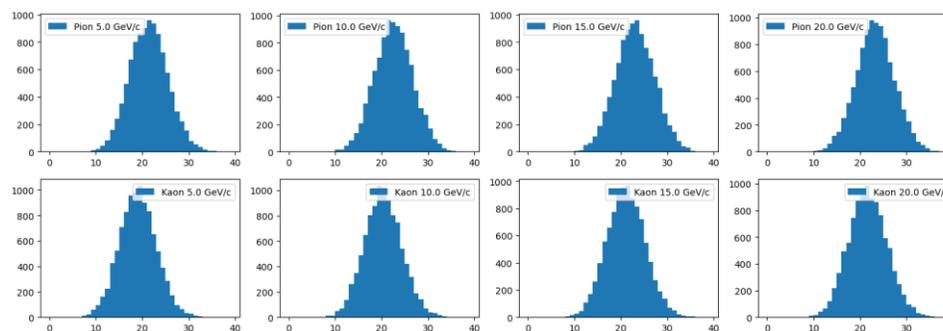
1. 针对模拟数据的监督学习算法

- 解决了信号堆积和噪声干扰的问题
- 特别地, 有效排除了次级电子的影响
- 粒子鉴别性能比传统算法提高 10%
- **文章已被 Nuclear Science and Techniques 接收**

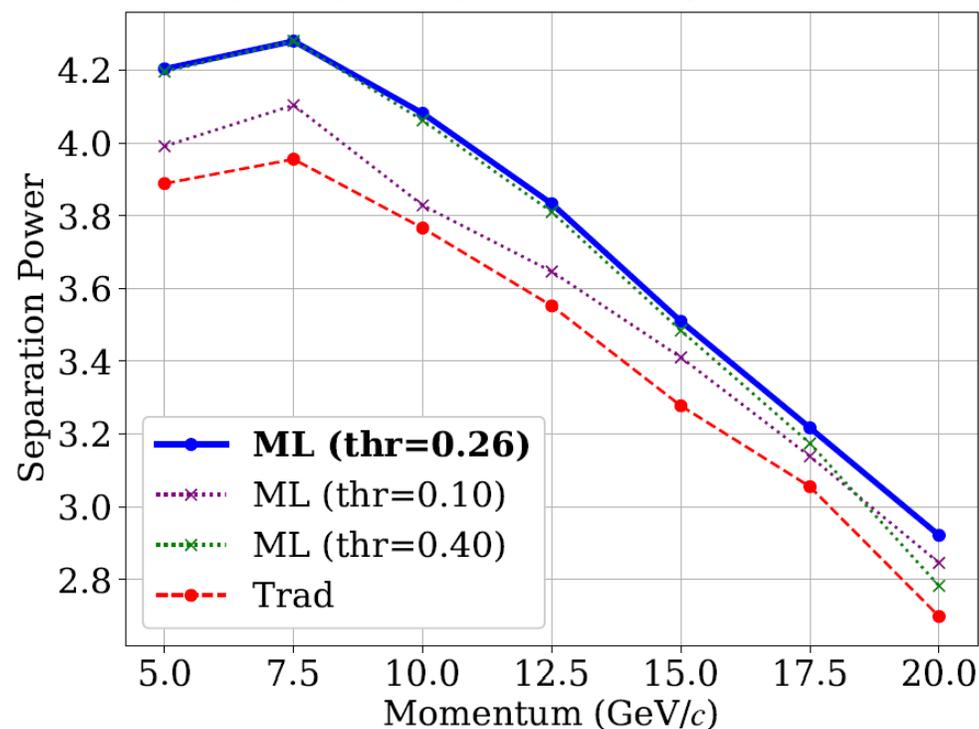
2. 针对实验数据的迁移学习算法

- 真实数据应用机器学习算法的深入思考和有效尝试
- **文章已被 Computer Physics Communications 发表**

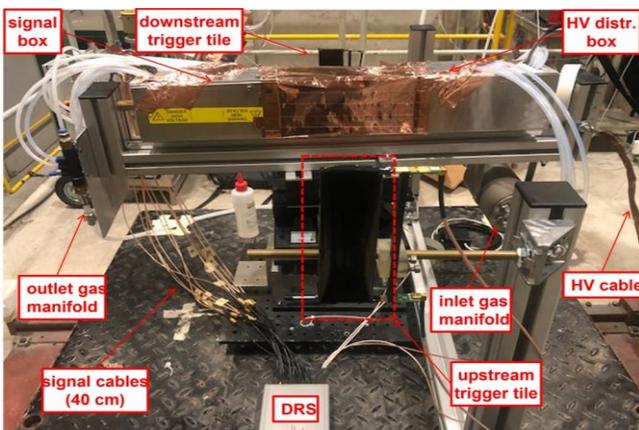
dN/dx 分布: 高斯分布 \rightarrow 有效排除了次级电离



机器学习算法提高了10%的粒子区分能力



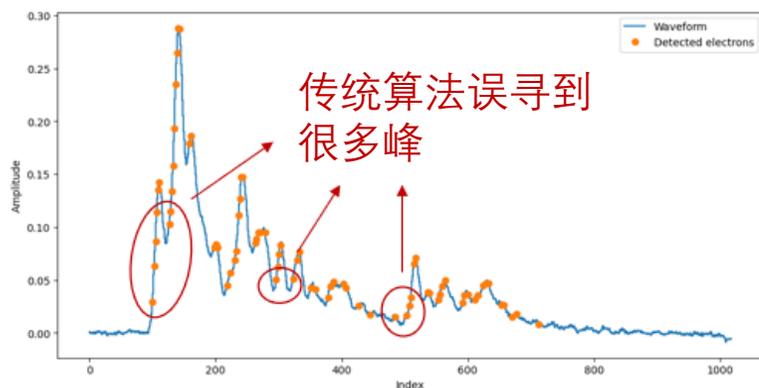
漂移室国际合作



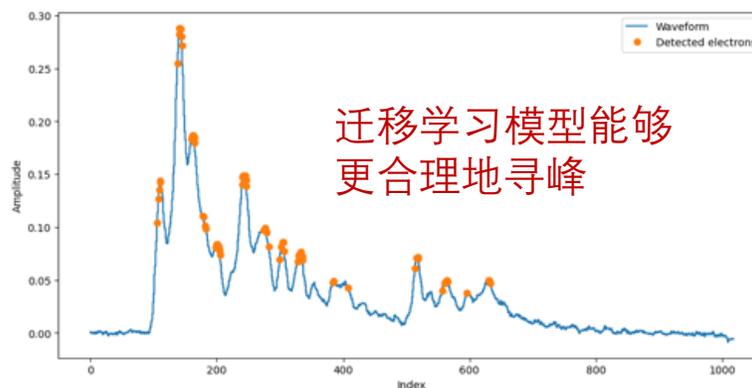
主导推动与 FCC-ee 探测器专家的合作:

- 通过在算法上的贡献，与 INFN 的 F. Grancagnolo/Nicola De Filippis 团队建立了深入合作的关系
- 共同分析在 CERN 获取的束流实验数据，并共同发表一篇文章
- 指导 INFN 博士生 Muhammad 进行机器学习算法研究

传统算法



迁移学习算法



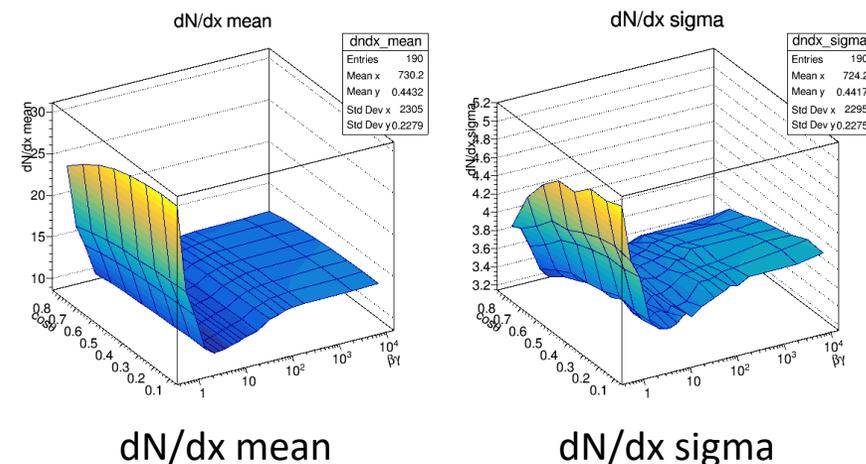
指导 INFN 博士生的国际会议报告

Poster for the 1st AI-INFN User Forum, Bologna, 12 June 2024. The poster features logos for Finanziato dall'Unione europea, Ministero dell'Università e della Ricerca, Italiadomani, and ICSC. The title is "Hyperparameter Optimization for Deep Learning Model Using High Performance Computing". The speaker is Muhammad Numan Anwar, on behalf of FCC Collaborator Nicola De Filippis, Marcello Abbrescia, Domenico Diacono, Guang Zhao, Mingyi Dong, Shengsen Sun, Linghui Wu, and Francesco Grancagnolo.

CEPCSW 实现及 Ref-TDR 撰写

- CEPCSW 中实现了粒子鉴别功能，并为物理分析提供了完整粒子鉴别信息
 - 基于Garfield++全模拟的参数化 TPC 模型
 - 基于参数化的 TOF 模型
- 正在撰写 Ref-TDR 中气体探测器和软件的章节

CEPCSW中的参数化模型

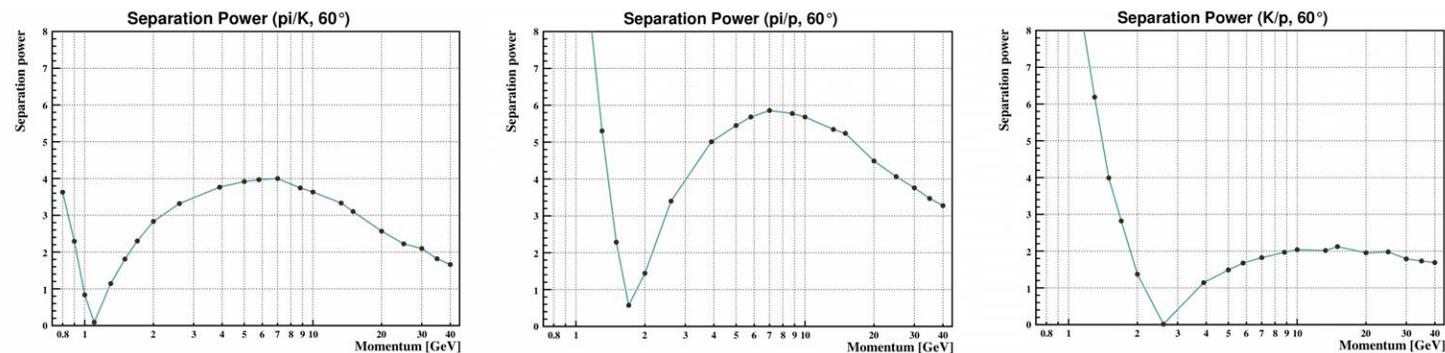


Ref-TDR 章节

Chapter 6 Gaseous Tracker	
6.1 Physics requirements	1
6.2 Gaseous tracker system overview	1
6.2.1 Technology comparison	1
6.2.2 Baseline gaseous tracker	1
6.2.3 R&D efforts and results	1
6.3 Pixelated readout Time Projection Chamber	3
6.3.1 Time Projection Chamber detector	3
6.3.2 Pixelated readout electronics	4
6.3.3 Design of mechanical and cooling	6
6.3.4 Commissioning and validation of prototype	6
6.3.5 Challenges and critical R&D	7
6.3.6 Costs	7
6.4 Performance	9
6.4.1 Overview of the simulation framework	9
6.4.2 Tracking performance	10
6.4.3 Particle identification	10
6.4.4 Improvement using the machine learning algorithm	10
6.4.5 Beam background source and estimation	11
6.4.6 Alternative the drift chamber	11
6.5 Prospects and outlook	11

Chapter 1 Software and Computing	
1.1 Introduction	1
1.2 CEPCSW structure	1
1.3 Development environment	1
1.4 Simulation	1
1.4.1 Simulation workflow and framework	1
1.4.2 Detector description in simulation	2
1.5 Reconstruction	3
1.5.1 Track reconstruction	3
1.5.2 Particle identification	4
1.5.3 Reconstruction in calorimeters	4
1.5.4 Muon reconstruction	6
1.6 Analysis software	6
1.7 Visualization software	6
1.8 Software R&D activities	7
1.8.1 Simulation on Gaussino	7
1.8.2 Fast simulation	9
1.8.3 Tracking with ACTS	9
1.9 Computing requirements	10
1.9.1 Data processing flow	10
1.9.2 Resource estimation	10
1.10 Computing model	10
1.11 Distributed computing infrastructure	10
1.12 Resource provision	11
1.12.1 Computing	11

CEPCSW中的粒子鉴别性能 (TPC)



BESIII数据分析: $e^+e^- \rightarrow KK\eta_c$

研究意义:

- 寻找 Z_c 的奇特伴随态 Z_{cs} , 理解其产生机制, 为理论提供实验信息
- 近年很多 Z_c 被实验上发现, 其中:
 - LHCb: $Z_c(4100) \rightarrow \eta_c \pi^+$ ($>3\sigma$) [Phys.Rev.D 90 (2014) 11, 112009]
- 测量 $e^+e^- \rightarrow KK\eta_c$ 过程, 并通过 $\eta_c K^+$ 寻找 Z_{cs}

研究进展:

- 未观测到显著信号, 给出了 $e^+e^- \rightarrow KK\eta_c$ 及 $Z_{cs}(3985) \rightarrow K\eta_c$ 的上限
- 回答 referee 4 轮次共 40 个问题, 正在增加 4.0-4.6 GeV 能量数据的结果

进入Memo审核阶段: BAM-704

4.680 GeV 数据的K介子反冲质量谱

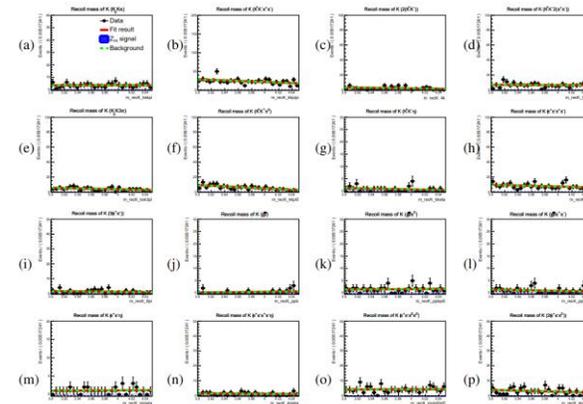


Figure 14: Nominal simultaneous fit to the recoil K mass for the decay mode (a) $K_c^0 K\pi$, (b) $K^+ K^- \pi^+ \pi^-$, (c) $4K$, (d) $K^+ K^- 2(\pi^+ \pi^-)$, (e) $K_c^0 K 3\pi$, (f) $K^+ K^- \pi^0$, (g) $K^+ K^- \eta$, (h) $2(\pi^+ \pi^-)$, (i) $3(\pi^+ \pi^-)$, (j) $p\bar{p}$, (k) $p\bar{p}\pi^0$, (l) $p\bar{p}\pi^+ \pi^-$, (m) $\pi^+ \pi^- \eta$, (n) $2(\pi^+ \pi^-) \eta$, (o) $\pi^+ \pi^- 2\pi^0$, (p) $2(\pi^+ \pi^-) 2\pi^0$ in the range of (3.9, 4.05) GeV from data samples at 4680 energy point.

$Z_{cs}(3985) \rightarrow K\eta_c$ 截面上限的初步结果

Table 17: Upper limits of signal yield at the 90% confidential level before and after considering additive systematic uncertainty, luminosity of data samples, upper limit of dressed cross section, production of ISR factor, VP factor and the summation of $\mathcal{B}_i \cdot \varepsilon_i$, upper limit of Born cross section for each energy point.

Energy point	N_{UP}	N_{UP} after considering additive systematic uncertainties	\mathcal{L} (pb ⁻¹)	$\sigma_{UP}^{dressed}$ (pb)	$(1 + \delta_{ISR}) \cdot \delta_{VP} \cdot \sum_i \mathcal{B}_i \cdot \varepsilon_i$	σ_{UP}^{Born} (pb)
4600	29.1	38.8	586.90	6.8	7.27×10^{-3}	9.1
4630	21.1	28.0	521.53	5.5	7.19×10^{-3}	7.5
4640	13.2	20.8	551.65	3.8	7.30×10^{-3}	5.2
4660	25.1	42.8	529.43	7.9	7.45×10^{-3}	10.9
4680	25.9	32.3	1667.39	1.8	7.64×10^{-3}	2.5
4700	18.7	24.3	535.54	4.3	7.06×10^{-3}	6.4
4780	25.1	35.1	511.47	6.0	8.58×10^{-3}	8.0
4840	7.8	16.6	525.16	2.8	8.39×10^{-3}	3.8

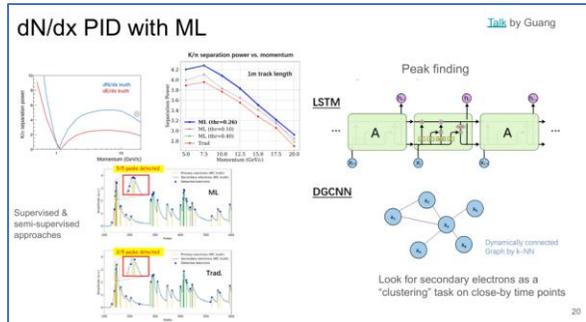
学术会议与国际合作

• 重要学术会议

- *DC for cluster counting for the CEPC 4th concept*, **IAS program 2024**, Jan 8-26, Hong Kong (talk)
- *Drift chamber developments for CEPC*, **1st DRD1 Collaboration Meeting**, Jan 29-Feb 2, online (talk)
- *dN/dx reconstruction with machine learning for drift chamber*, **CEPC Workshop (EU) 2024**, April 8-11, Marseille (talk)
- *Drift chamber option for CEPC*, **LCTPC Collaboration Meeting**, May 12-13, online (talk)
- *Drift chamber with cluster counting techniques for CEPC*, **International Workshop on Future Linear Colliders**, July 8-11, online (talk)
- *dN/dx reconstruction with machine learning for cluster counting*, **the 42nd International Conference on High Energy Physics (ICHEP 2024)**, July 18-24, Prague, Czech (poster)
- *Drift chamber with cluster counting techniques for CEPC*, **Quantum Computing and Machine Learning Workshop 2024**, Aug 2-8, Changchun (talk)
- *基于监督学习和迁移学习的电离计数重建方法研究*, **中国物理学会高能物理分会第十四届全国粒子物理学术会议 2024**, Aug 12-18, Tsingdao (talk)
- *dN/dx reconstruction with supervised learning and transfer learning*, **Conference on Computer in High Energy and Nuclear Physics (CHEP 2024)**, Oct 21-25, Krakow, Poland (talk)

• 国际合作

- **DRD1**: CERN DRD1-WG2-WP2 “Inner and central tracking with PID capabilities with Drift Chambers” 高能所联络人
- **电离计数**: F. Grancagnolo/Nicola De Filippis INFN团队，合作指导意方研究生 Muhammad Numan Anwar



Summary talk by Davide Valsecchi

学术论文与科研项目

• 期刊论文

- Peak finding algorithm for cluster counting with domain adaptation, **Computer Physics Communications 300 (2004) 109208** (第一作者, 通讯作者, **IF = 7.2**)
- Cluster counting algorithm for the CEPC drift chamber using LSTM and DGCNN, **Nuclear Science and Techniques** (通讯作者, 已接收, **IF = 3.6**)
- High granularity readout TPC R&D for tera-z at the future e+e- collider, **Proceedings of Science** (合作作者)
- Offline data processing system of the BESIII experiment, **European Physical Journal C** (合作作者)
- Time calibration of barrel TOF system at BESIII, **Radiation Detection Technology and Methods** (合作作者)

• 科研项目

- 气体探测器电离计数方法的软件研究 (国自然面上项目, 主持)
- 粲强子的强子末态衰变机制研究 (国自然重大项目, 参加)
- 利用粲强子含轻子衰变精确检验标准模型 (国家重点研发项目, 参加)

学生培养与公共服务

- **学生培养**

- 研究生 (3名):

- 田喆飞: DC电离计数方法的研究
 - 常悦: TPC电离计数方法的研究
 - 张锦闲: TPC电离计数方法的研究

- 本科生 (3名):

- 高旭: DC电离计数方法的研究
 - 王远瞻/何俊驰: TPC电离计数方法的研究

- 合作博士后 (2名):

- 曹宁: BESIII分析工作
 - 张辰光: TPC电离计数方法的研究

- **公共服务**

- BESIII分析内部审阅人: BAM-769; BAM-874; BAM-917

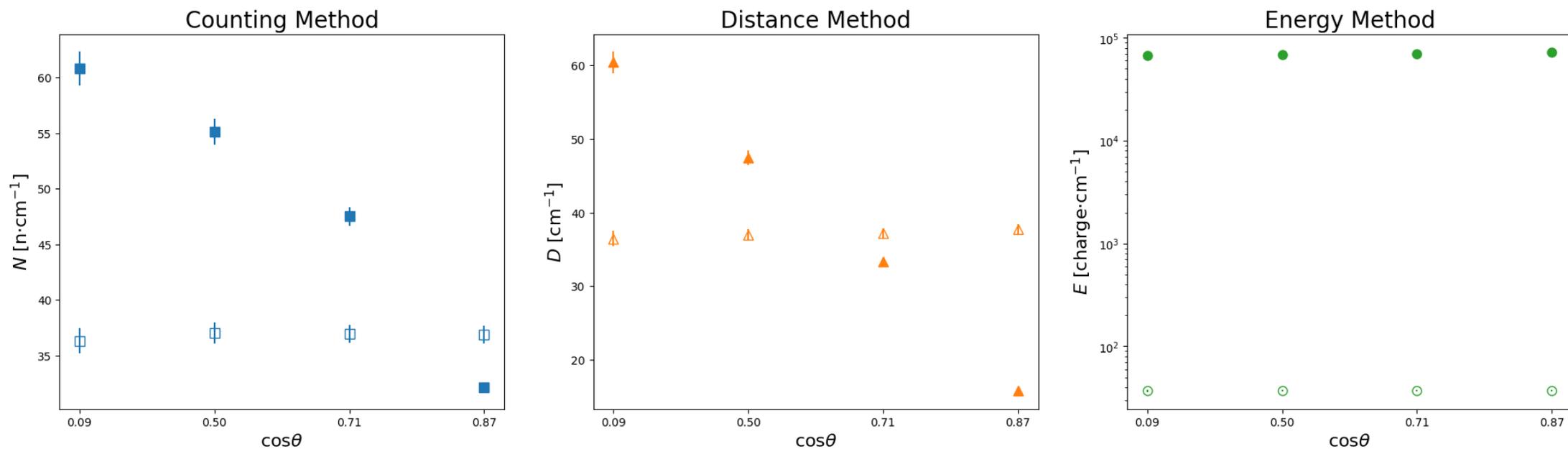
下年度工作计划

- **CEPC 气体探测器电离计数的研究**
 - 利用模拟研究优化 TPC 探测器设计
 - 开发复杂的 TPC PID 重建算法，以及 CEPCSW 下像素型读出 TPC 的实现
 - 加强国际合作，推动束流实验数据分析，验证模拟可靠性
- **BESIII 数据分析**
 - 继续完成 $KK\eta_c$ 的分析

谢谢!

Backups

电离测量中的角度刻度



The cosine theta dependencies are vanished

Peak finding results

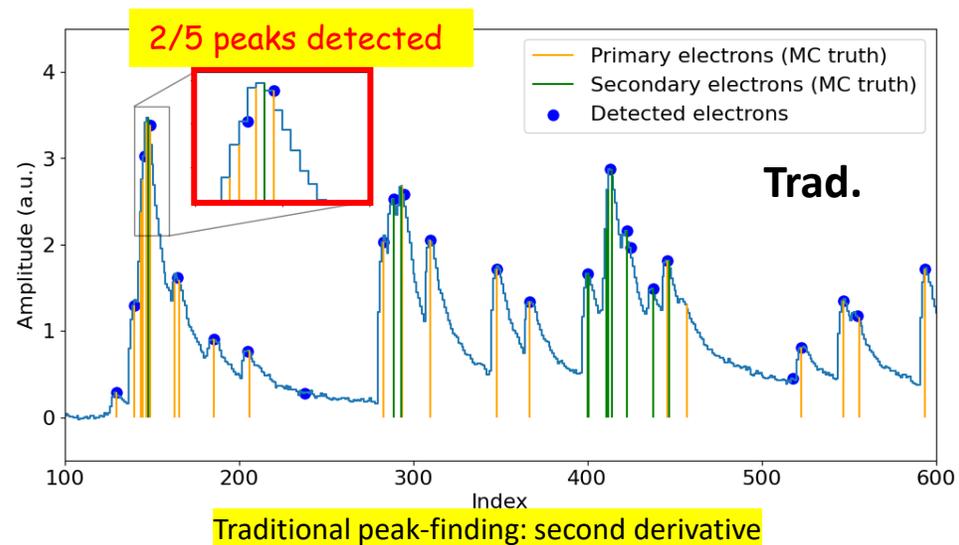
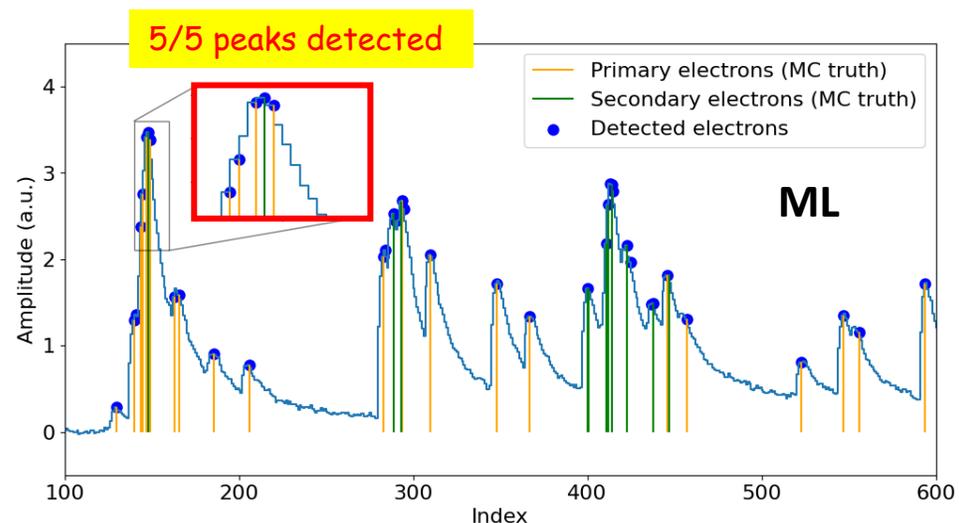
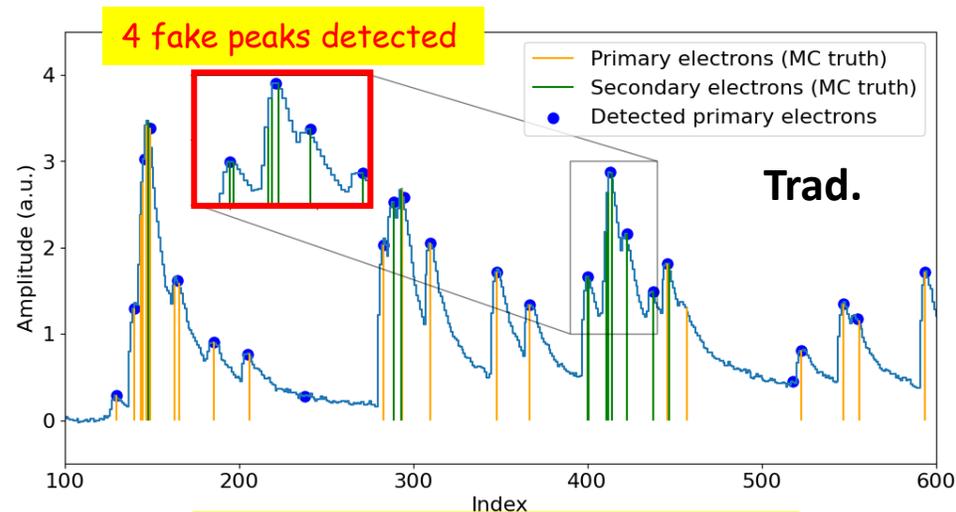
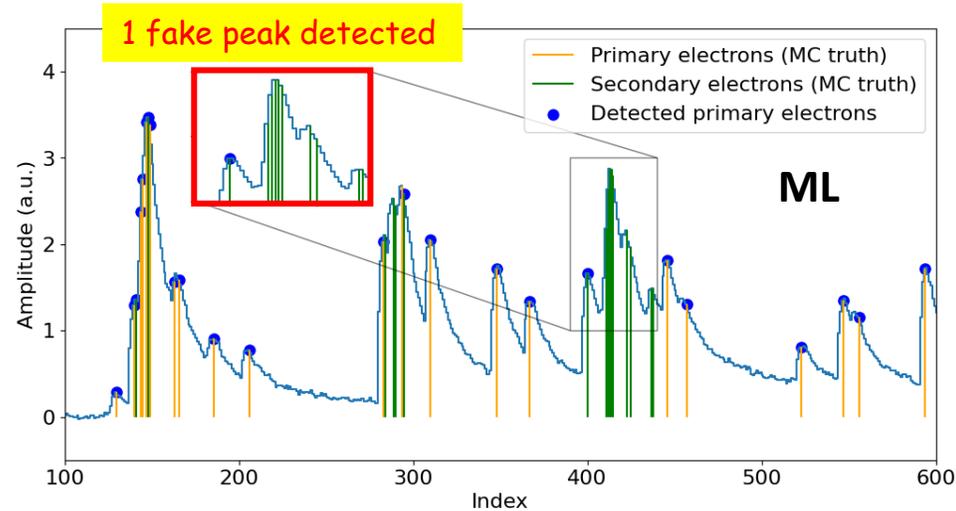


Table 2. The purity and efficiency comparison between LSTM-based algorithm and traditional D2 algorithm for peak-finding.

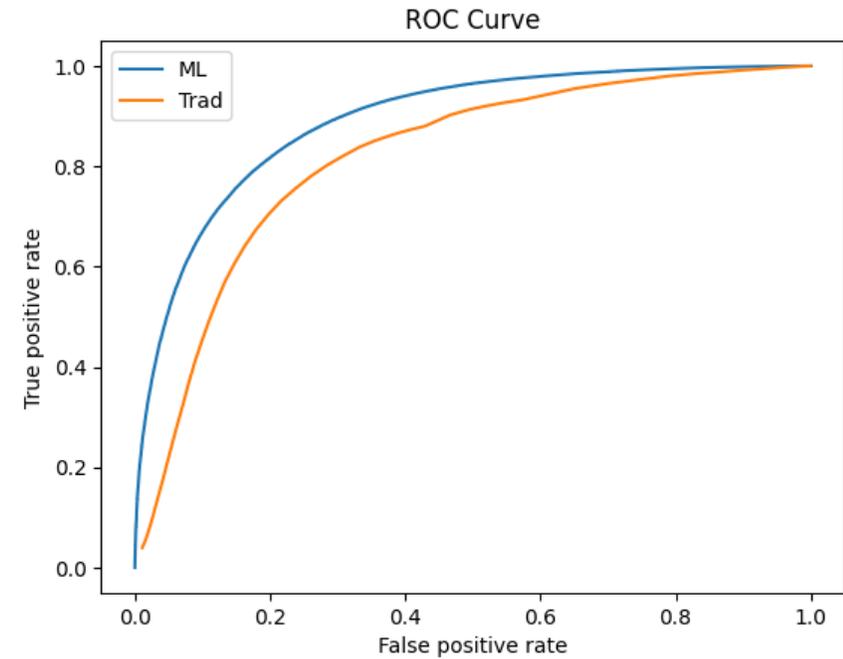
	Purity	Efficiency
LSTM algorithm	0.8986	0.8820
D2 algorithm	0.8986	0.6827

- The LSTM-based model is more powerful than the traditional derivative-based algorithm, especially for the pileup recovery

Clusterization results



Traditional clusterization: adjacent-peak merge



- The DGCNN-based model is more powerful than the traditional peak-merge algorithm, as it can remove the secondary electrons more accurate

Alg. 2: Transfer learning for real data

Challenges for real data

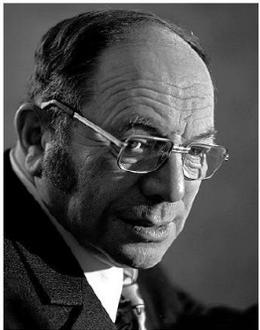
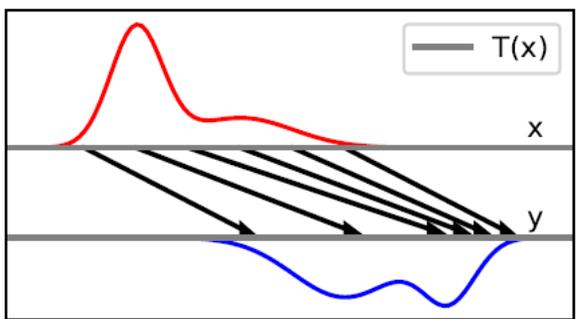
- Imperfect simulation
- Incomplete labels in real data



Solution: Domain adaptation

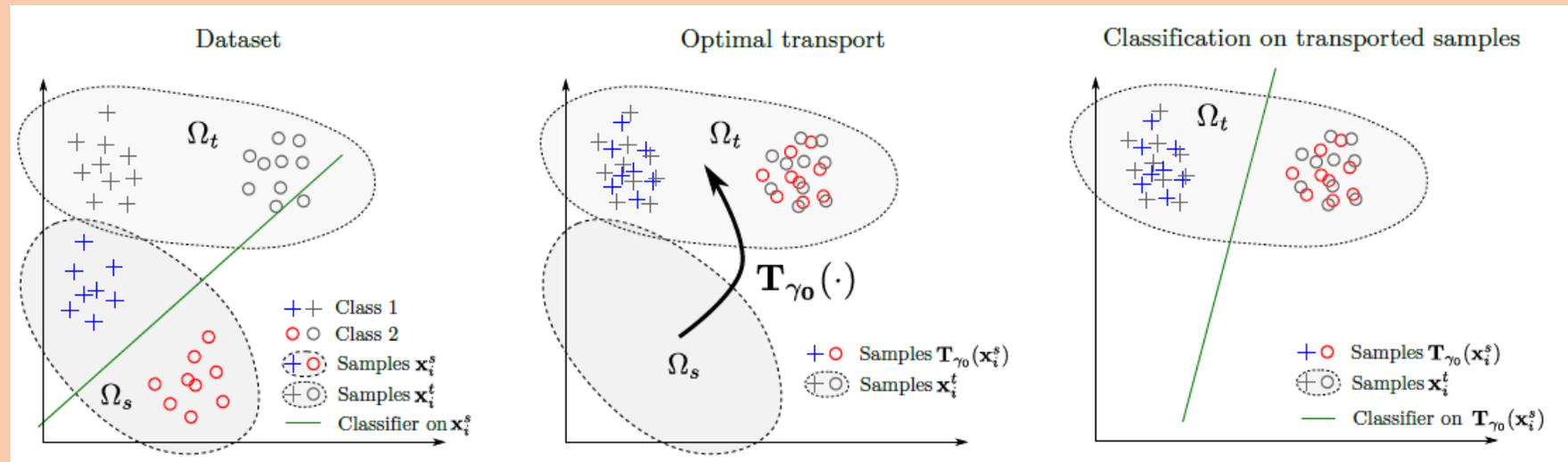
- Transfer knowledge between simulation and real data via **optimal transport**

Optimal Transport



Kantorovich
(Economic
Nobelist 1975)

Domain adaptation



Figures from Flamary's slides

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