



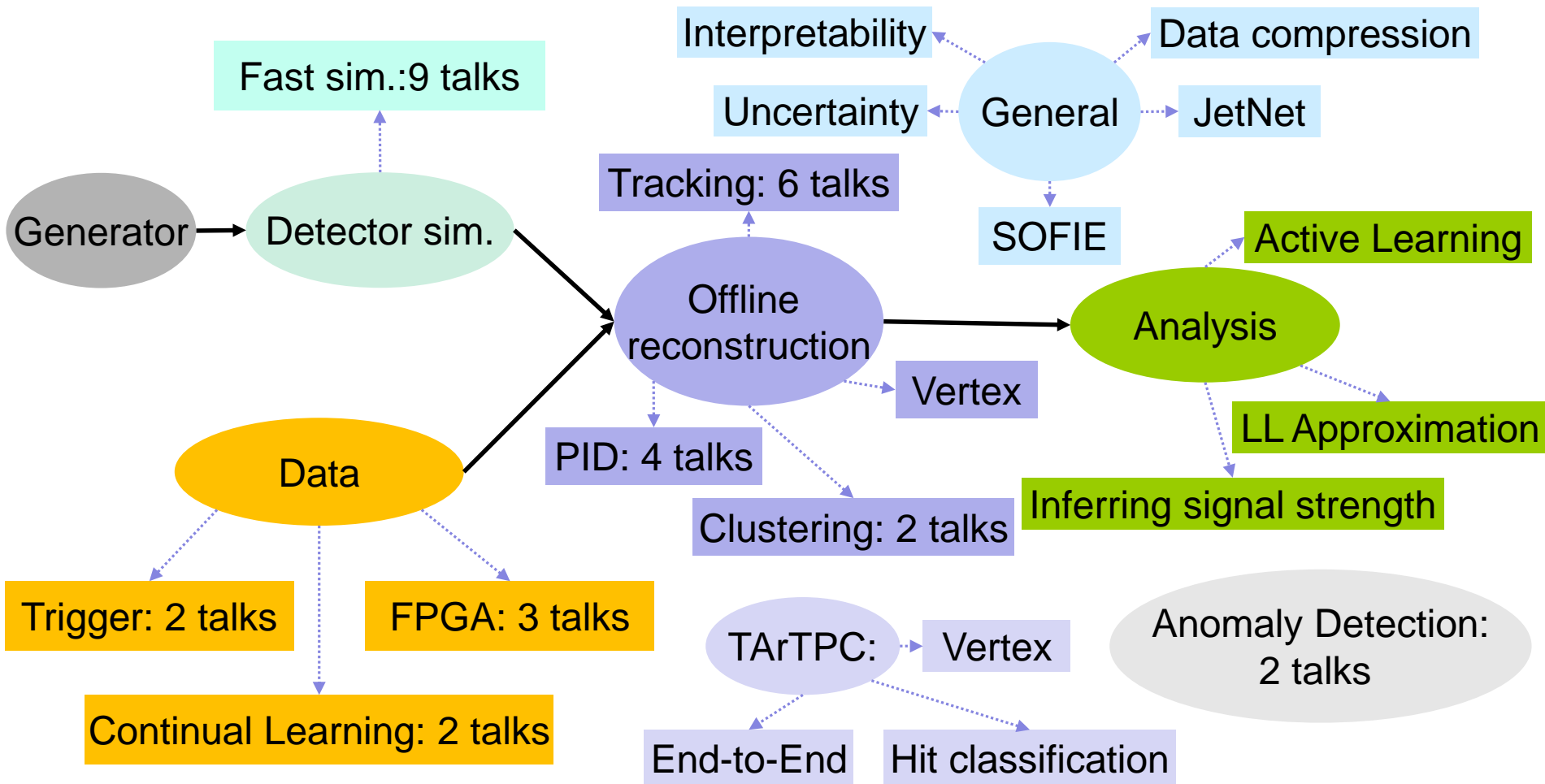
中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

CHEP2023 机器学习 进展分享

Wenxing Fang (IHEP)

**2023年粒子物理实验计算软件与技术研讨会, 2023.06.09-12,
山东大学青岛校区**

机器学习口头报告内容



共 42 个口头报告

快速模拟

- ❖ 背景：未来的 HL-LHC 实验将产生海量实验数据，数据处理、分析、模拟数据产生等都需要大量计算资源。如果不进行 R&D 研究将面临计算资源不足的问题
- ❖ 其中实现 MC 快速模拟能够大幅降低计算资源的需求，具有重要意义
 - 传统方法: shower 参数化、Frozen shower、快径迹模拟（简化几何和物理过程）、模拟部分粒子、全参数化 (Delphes) 模拟
 - 机器学习方法：主要集中在量能器的快速模拟和 Ultra-Fast Simulation（跳过 Geant4）的 End-to-end 模拟

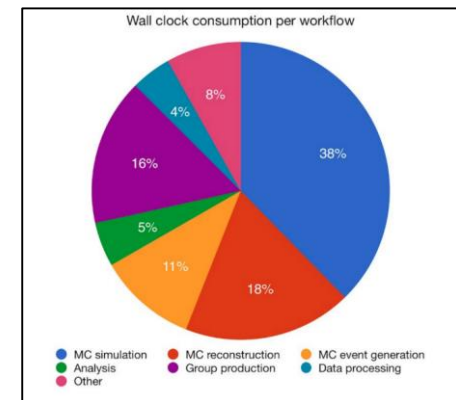
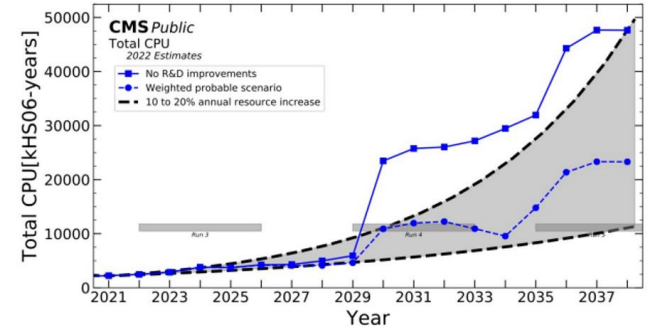
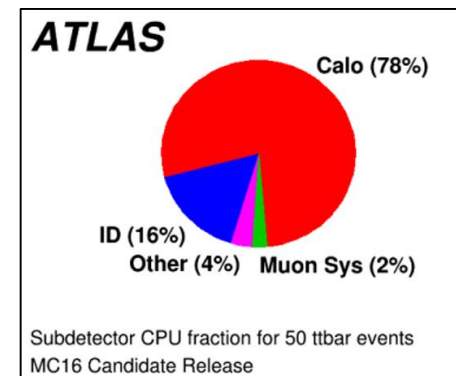
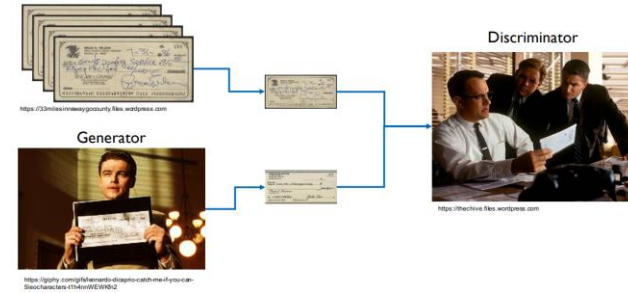


Figure 1: ATLAS CPU hours used by various activities in 2018

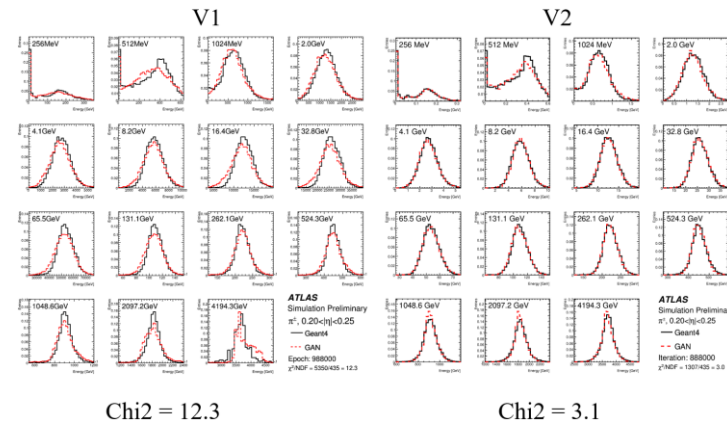
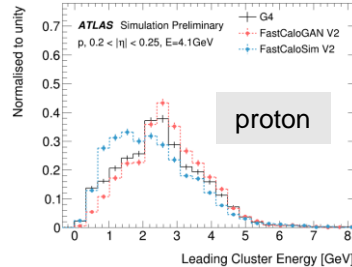
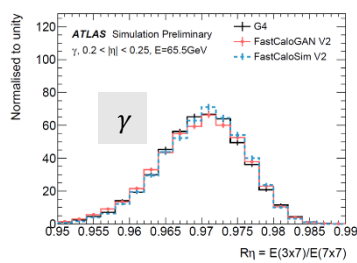
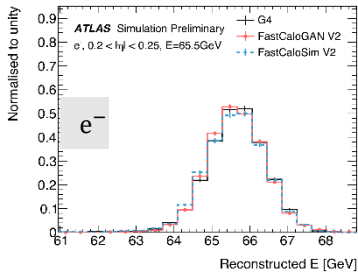


量能器快速模拟 (1)

- ❖ 2017 年 [CaloGAN](#) 被提出来, 使得基于 GAN 的量能器快速模拟成为研究热点
- ❖ 目前 ATLAS 的 AltFast3 快速模拟软件已经将 GAN 整合其中, 用于 Run2 的模拟数据产生 (2B)
- ❖ 最新的 FastCaloGAN v2 对实现了更加精确的模拟:
 - 不同种类粒子、能量、方向 (η)。从原先 1 个 GAN 到 500 个 GAN
 - 将被广泛用于 Run3 的模拟样本产生



AtI Fast3	Inner Detector	Calorimeters				Muon Spectrometer
Electrons Photons	Geant4	FastCaloSimV2				
Hadrons		Geant4 <small>point $E_{min} < 200$ MeV Other hadrons $E_{min} < 400$ MeV</small>	FastCalo Sim V2 <small>$E_{min} < (8-16)$ GeV</small>	FastCalo GAN <small>$(8-16)$ GeV < E_{min} < $126-192$ GeV</small>	FastCalo Sim V2 <small>$E_{min} > (258-312)$ GeV</small>	Muon Punchthrough +Geant4
Muons		Geant4				Geant4

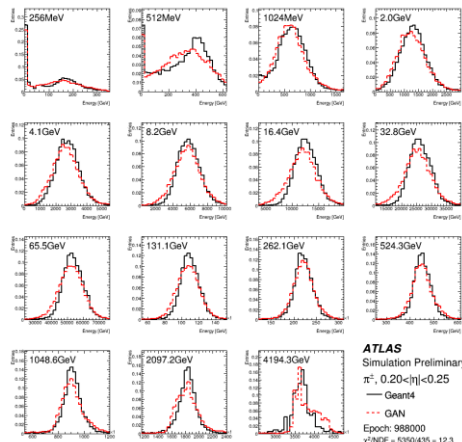
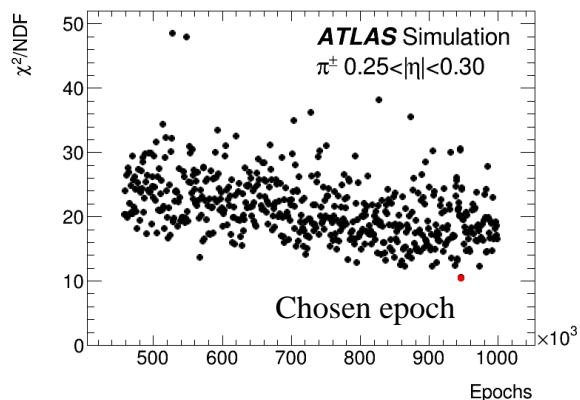
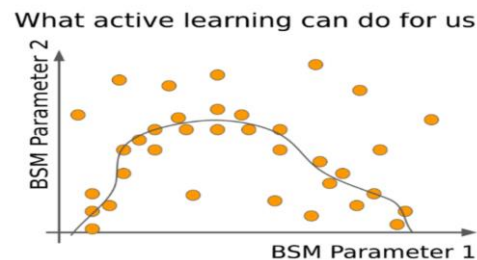
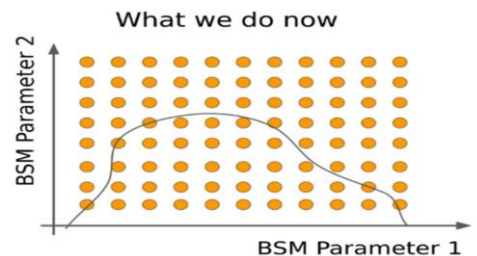
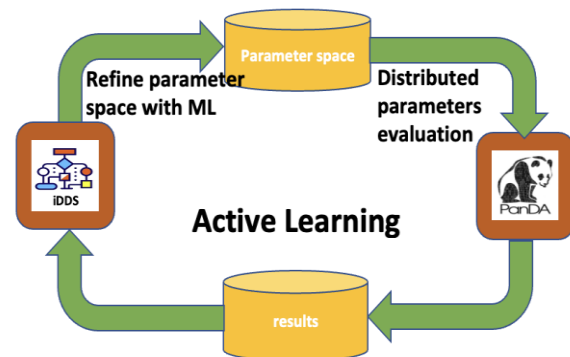


Chi2 = 12.3

Chi2 = 3.1

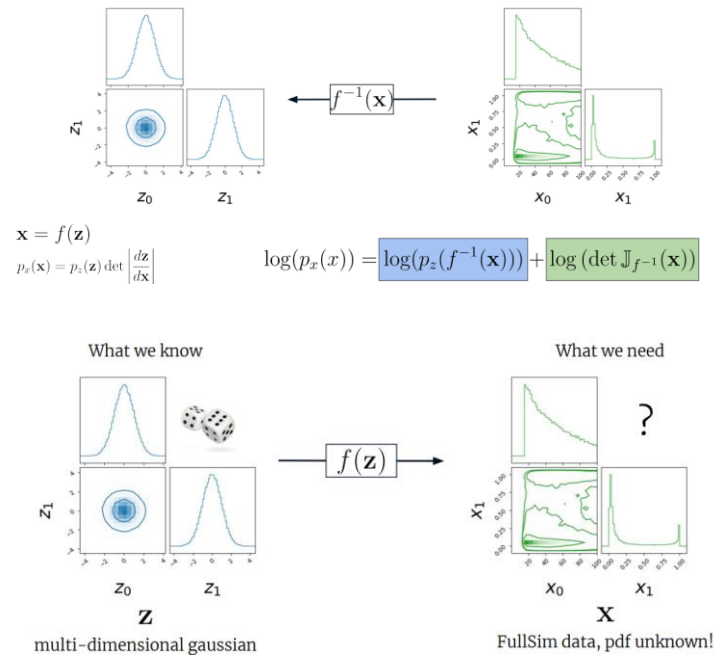
量能器快速模拟 (1)

- ❖ 由于要训练的 GAN 模型个数多 (500 个), 且每个 GAN 模型的训练也不容易 (训练过程不稳定、需要优化超参数)。因此, 需要解决 GAN 模型训练的问题
- ❖ 为此 ATLAS 实验利用 [Active Learning](#) 的技术, 实现网络模型的自动训练和超参数的优化:
 - iDDS (intelligent Data Delivery Service) 负责根据当前模型训练的结果产生下一批模型训练作业 (如利用贝叶斯算法、GP 等算法缩小超参数范围)
 - [PanDA](#) 系统将作业调度到分布式的异构计算资源上进行模型的训练, 返回训练结果

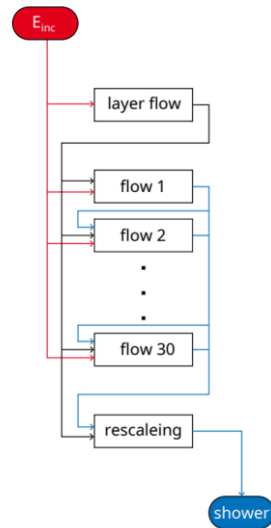


量能器快速模拟 (2)

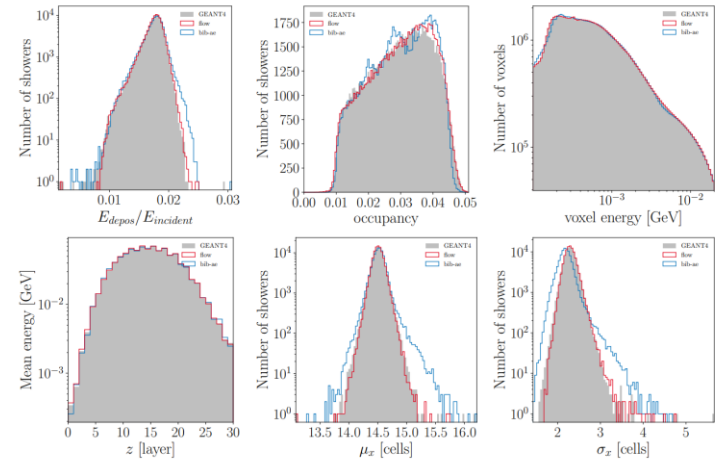
- ❖ 近些年，基于其它机器学习技术的量能器快速模拟也在快速发展
- ❖ 2021年提出的 CaloFlow 基于 Normalizing Flows 技术，能够得到比 CaloGAN 更精确的模拟。相比 GAN，其训练更加稳定，能够收敛。在推理速度上要比 GAN 慢一个量级左右



Architecture



Preliminary Results



Simulator	Hardware	Batch size	time [ms]	Speedup
GEANT4	CPU	1	4081.53 ± 169.92	×1.0
BIB-AE	CPU	1	102.25 ± 0.64	×40.0
		10	37.81 ± 0.13	×108.0
		100	48.51 ± 0.01	×84.1
		1000	48.19 ± 0.01	×84.7
Flow	CPU	1	1746.61 ± 64.50	×2.3
		10	392.61 ± 0.34	×10.4
		100	228.86 ± 7.09	×17.8
		1000	275.55 ± 3.01	×14.8
BIB-AE	GPU	1	74.22 ± 3.18	×42.5
		1000	0.249 ± 0.002	×16326.1
Flow	GPU	1	2471.07 ± 70.20	×1.7
		1000	3.39 ± 0.09	×1202.3

量能器快速模拟 (3)

❖ 2020年提出 [Diffusion model](#)，其能生成非常逼真的图像，目前基于 Diffusion model 的图像生成算法已经成为主流

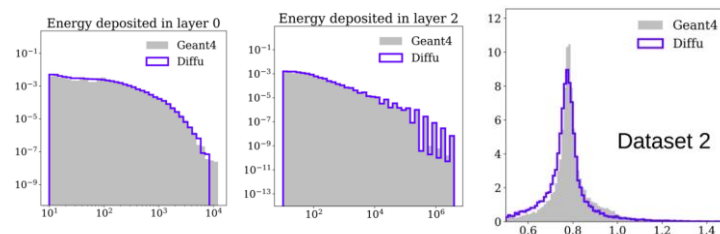
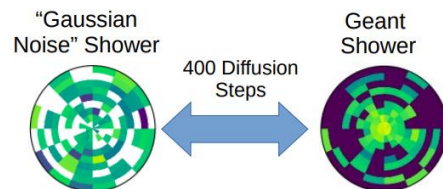
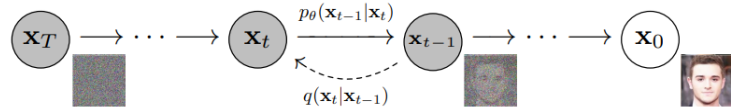
❖ CaloDiffusion (CHEP2023)：基于 Diffusion 的能量器快速模拟

- 大部分模拟结果能和 Geant4 一致，有些全局变量符合的还不是很好
- 速度上比其它机器学习方法慢 (比 Geant4 快)

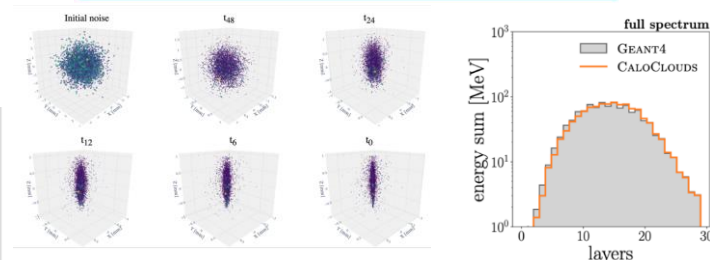
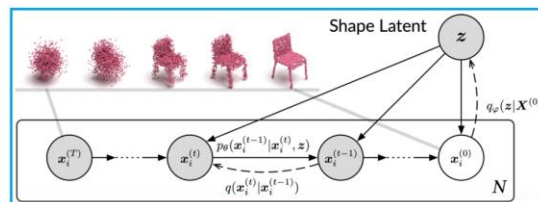
❖ CaloCloud: 基于 Diffusion 和 [Point cloud](#) 的量能器快速模拟

- 克服几何不规则问题
- 得到 promising 结果，证明方法可行
- 模拟速度需要提高

Hardware	Simulator	Time / Shower [ms]	Speed-up
CPU	GEANT4	4082 ± 170	×1
	CALOCLOUDS	3509 ± 220	×1.2
GPU	CALOCLOUDS	38 ± 3	×107



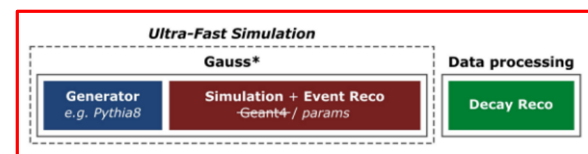
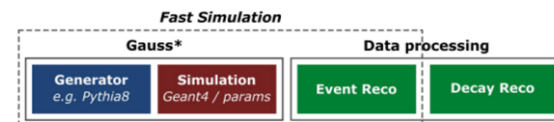
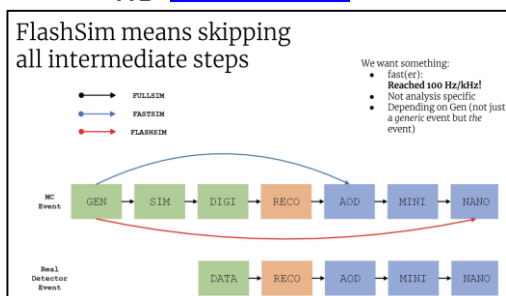
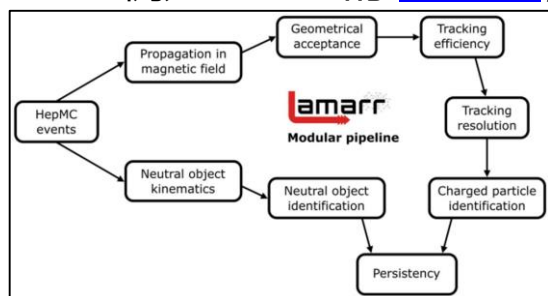
	Dataset 1 (ATLAS-like)	Dataset 2	Dataset 3
Classifier AUC *	~0.65	~0.6	~0.7



End-to-end 快速模拟

- ❖ Ultra-Fast Simulation. 跳过 Geant4 模拟, 实现从 MC Particle 到用于物理分析的高级对象的模拟

- 例如 LHCb 的 [Lamarr](#), CMS 的 [FlashSim](#)



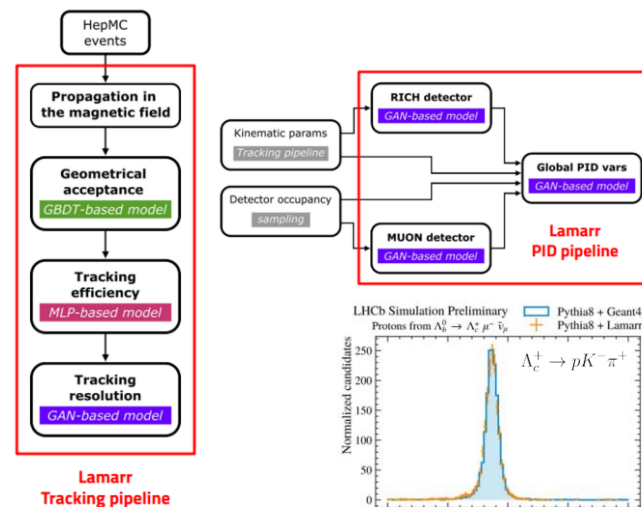
- ❖ 大部分的模拟是基于机器学习的:

- 使用 GBDT 模拟几何的 acceptance
- 使用 MLP 模拟 tracking 效率
- 使用 GAN 模拟 tracking 分辨率、PID

- ❖ 对于 fake object 或者粒子之间关联的问题需要额外处理:

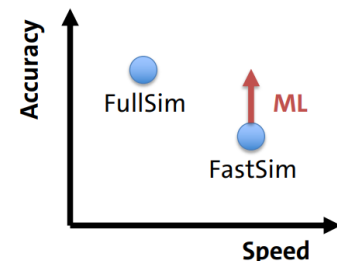
- 例如, n 个产生的光子到 m 个重建的 cluster 的模拟。可以类比语言翻译问题

- ❖ 快速模拟可以在探测器设计、机器学习训练、系统误差研究等方面减轻 CPU 的压力。其不会完全取代 Geant4

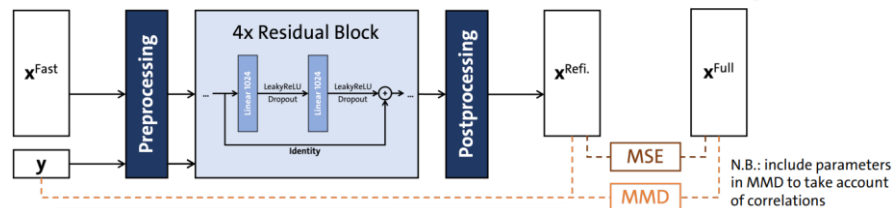
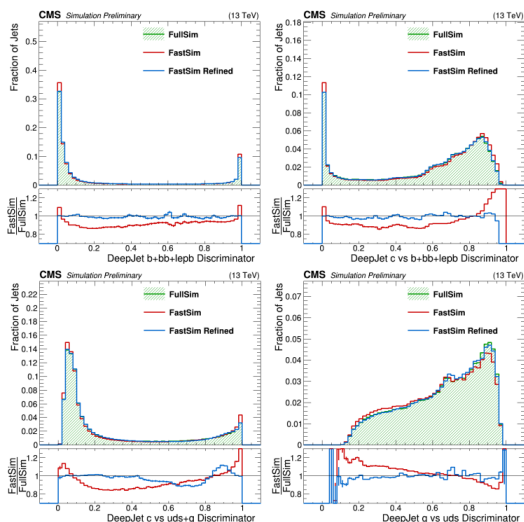


快速模拟物理性能的优化

- ❖ 虽然快速模拟能够节约计算资源，但模拟的精度也会有所下降
- ❖ 可以采用 Reweighting 或 Refining 的方法进一步提高模拟的精度
- ❖ CMS 实验展示了 **Refining** 方法在 Jet 模拟方面的应用：



- 网络输入: x^{Fast} (4 个快速模拟得到的 Jet flavor 鉴别量) 以及 y (Gen Jet 的 p_T , η , flavor)
- 网络输出: 经过 refined 的 x 值

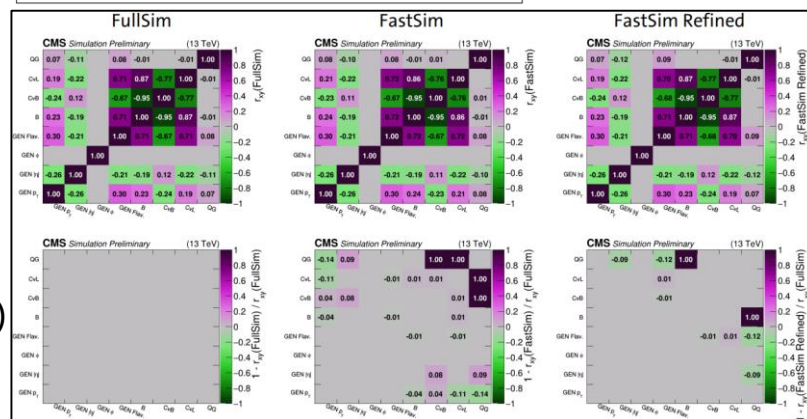


given two samples from $P(X)$ and $Q(Y)$:

$$\text{MMD}(P, Q) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n k(x_i, x_j) + \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m k(y_i, y_j) - \frac{2}{nm} \sum_{i=1}^n \sum_{j=1}^m k(x_i, y_j)$$

N.B.: include parameters in MMD to take account of correlations

- ❖ 扩展用于其它量的 refining (e.g. jet substructure)
- ❖ 直接 refining 到真实数据？

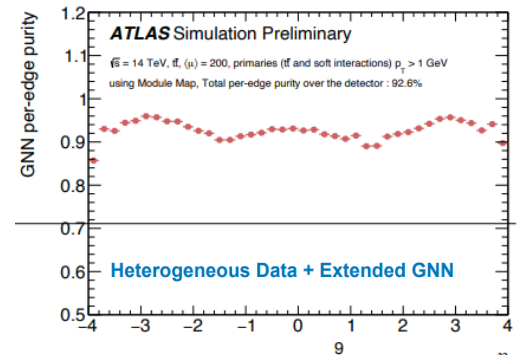
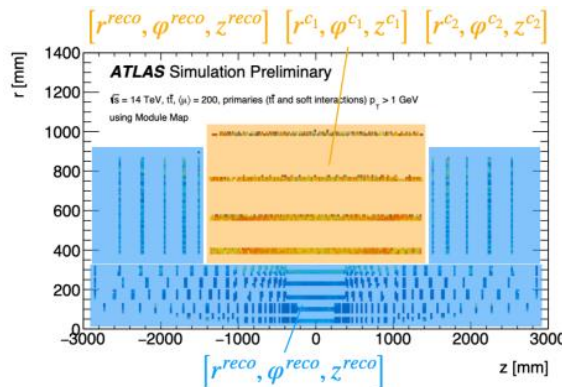
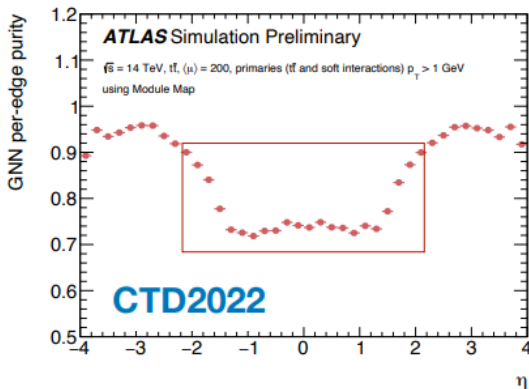
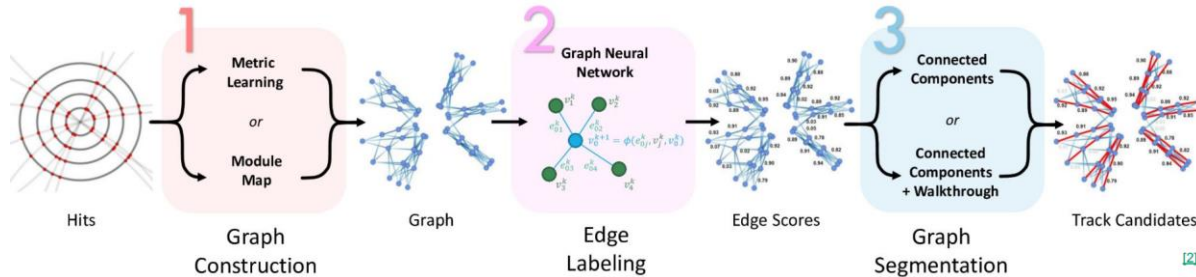
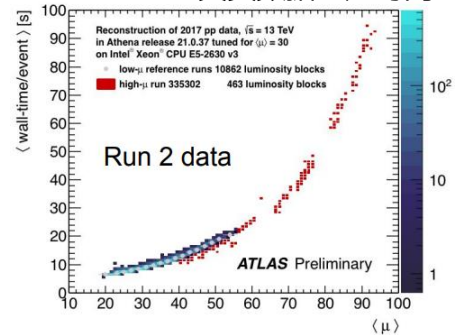


径迹重建

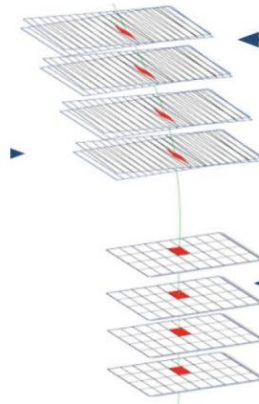
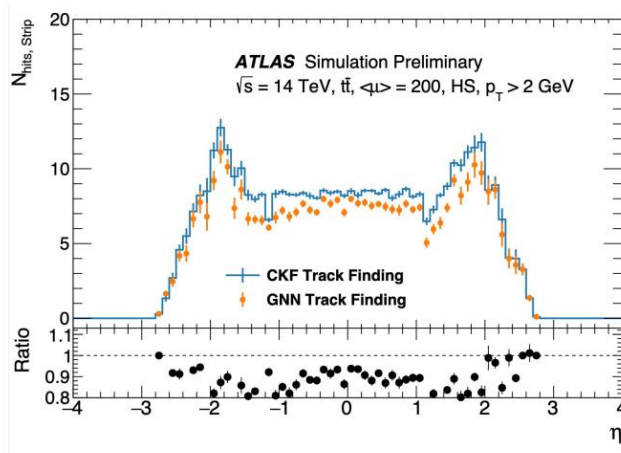
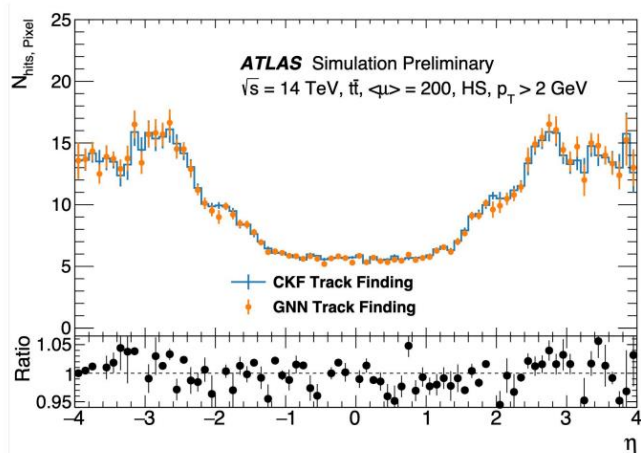
- ❖ 由于带电径迹重建的重要性，其一直是机器学习研究的热点之一
- ❖ 其主要目的是提升重建速度（例如在线重建）、物理性能
- ❖ ATLAS 实验研究了基于 GNN 的 ITK 径迹重建：

- **在线**：将模型部署到 FPGA 上，加快径迹重建速度
- **离线**：节约计算资源，径迹重建时间占总重建时间~40%

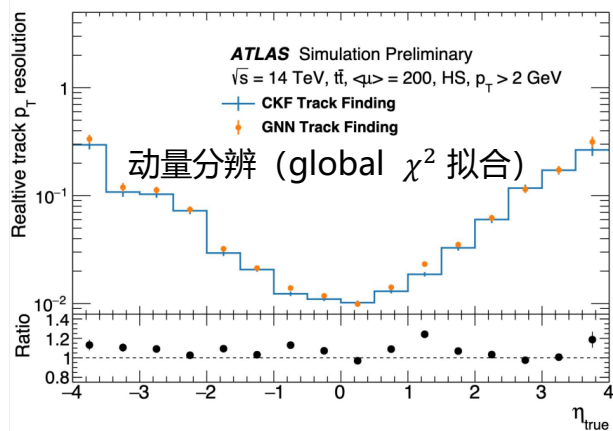
真实数据重建时间



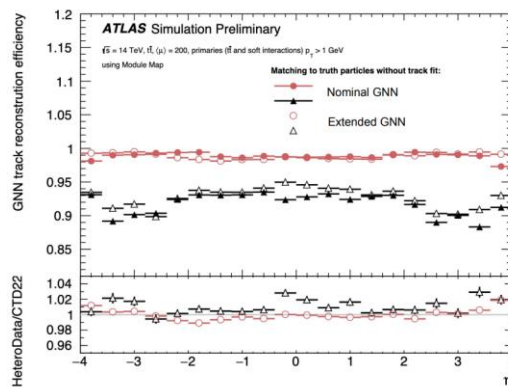
径迹重建



- ❖ 比较 GNN 和 CKF 方法的径迹上击中数目, pixel 基本一致, strip 偏少:
 - 可能有的 strip missing 的击中, 导致被 GNN 重建忽略

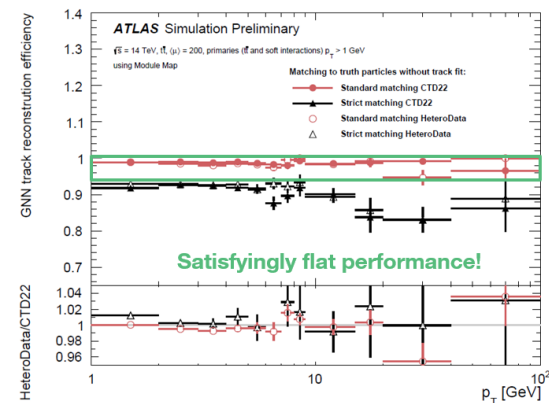


Relative track p_T resolution is measured as the multiplication of $p_{T,true}$ and the RMS of the pull distribution of $(q/p_{T, reco} - q/p_{T, true}) / q/p_{T, true}$.



GNN track reconstruction efficiency calculated with two matching schemes:

- Red circle ("standard matching"): > 50% of the spacepoints in the reconstructed track are matched to a true track
- Black triangle ("strict matching"): 100% of the spacepoints in the reconstructed track are matched to a true track

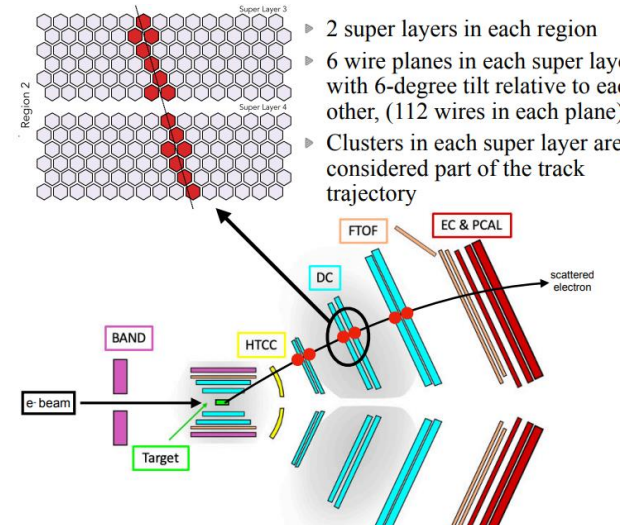


- Fake rate is $O(10^{-3})$ using standard truth matching

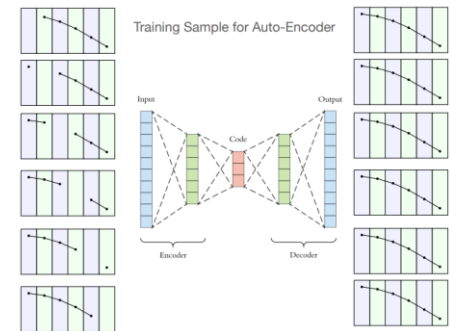
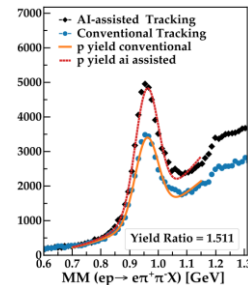
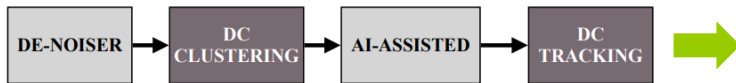
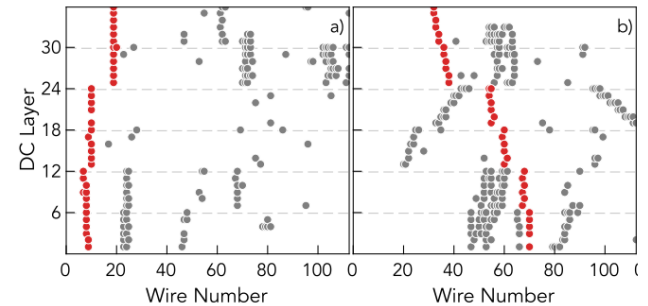
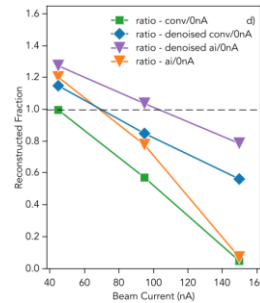
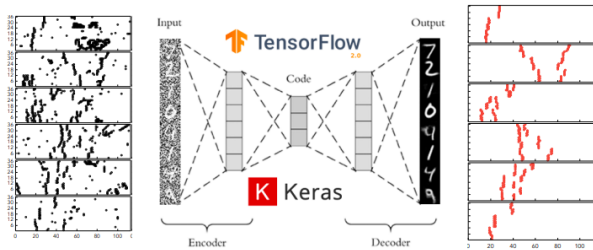
径迹重建

❖ Jlab 的 [CLAS12](#) 实验 (打靶实验) 展示了在机器学习的辅助下径迹重建性能的提升

- 1, 利用 MLP 去判断不同的超层径迹段是否来自同一条径迹, 提升径迹重建速度~30%
- 2, 利用 MLP Auto-Encoder 去补充缺失的径迹段, 单径迹重建效率可以提升~10% (双径迹~20%)
- 3, 随束流强度的增加 (噪声增加) 径迹重建效率会明显下降。在降噪网络 (CNN Auto-Encoder) 的帮助下, 可以大幅缓解噪声带来的影响



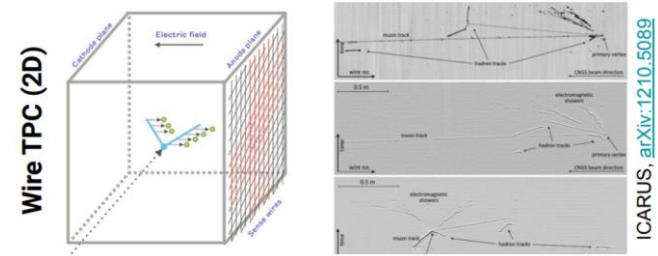
- 2 super layers in each region
- 6 wire planes in each super layer with 6-degree tilt relative to each other, (112 wires in each plane)
- Clusters in each super layer are considered part of the track trajectory



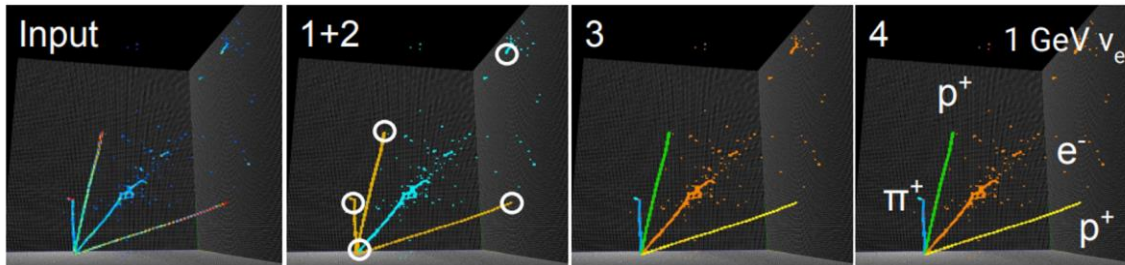
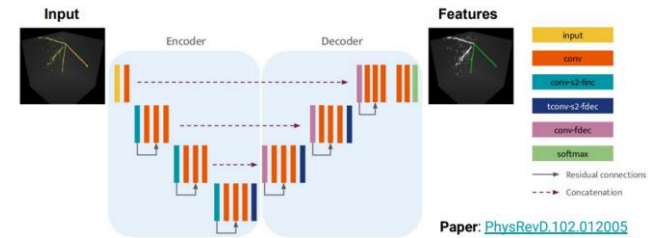
The Liquid Argon TPC (LArTPC) 重建

- ❖ 研究基于机器学习的 End-to-end 重建
 - 利用探测器测量得到的图像信息，实现 pixel level 和 particle level 的重建
- ❖ Pixel level: 利用 UResNet 网络实现对 pixel 的分类，例如是否为反应顶点，是否来自shower 或 track 等
- ❖ Particle level: 利用 GNN 网络实现对 pixel 的聚类以及最终粒子的重建和 PID

The modern Particle Imaging Detector

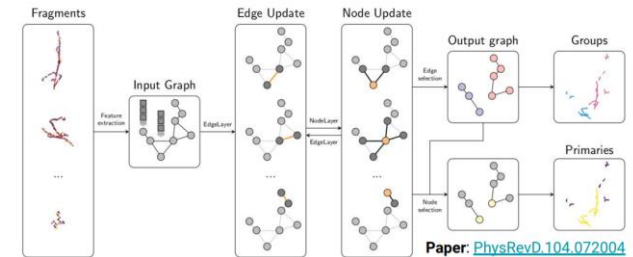


UResNet (UNet + ResNet + Sparse Conv.) as the backbone feature extractor



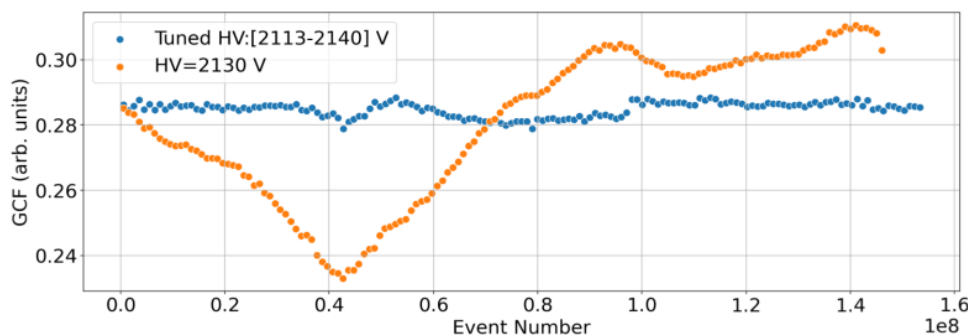
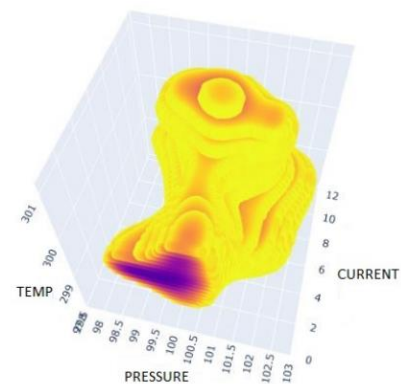
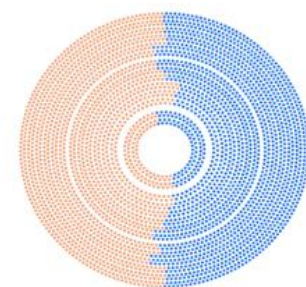
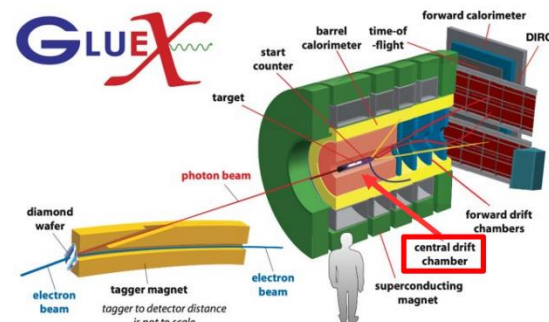
Graph Particle Aggregator (GrapPA)

Graph Neural Network: fragments/particles (nodes), correlations (edges)



在线刻度

- ❖ 刻度对获得高质量的分析数据至关重要，一般是在离线环境进行
- ❖ 如果能实现在线环境的刻度，则能得到更高质量的实验数据
- ❖ JLab 的 [GlueX](#) (光子打靶实验) 利用机器学习方法 (Gaussian Process)，根据漂移室气压、温度以及束流情况对漂移室的工作高压进行在线调整
 - 将漂移室分成 2 左右部分，左边是固定高压，右边是机器学习控制的高压 (每 5 分钟自动调整一次)
 - 该机器学习方法还能给出需调整高压值的不确定程度，确保高压的调整是合理、可控的



其它

- ❖ 异常检测：基于 VAE 网络实现对 SNS 加速器相关模块异常的提前预测，减少加速器维修时间
- ❖ 数据压缩：为了解决未来 HL-LHC 实验存储空间不足的问题，研究了基于 Auto-Encoder 的数据压缩方法，结果 promising
- ❖ 模型部署：ROOT/SOFIE 可以根据用户提供的网络模型 (onnx, Pytorch, Keras) 产生相应的 C++ 代码，将模型部署在 C++ 环境中，实现高性能的推理 (性能和 ONNXRuntime 相当)
- ❖ JetNet: 提供公共数据集、数据分析工具、模型性能评价指标，让用户将精力集中在模型的构建和训练

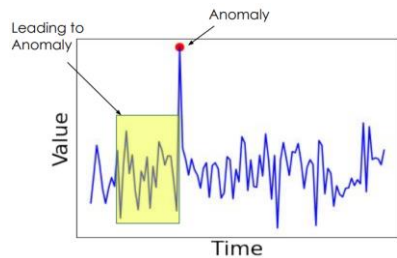
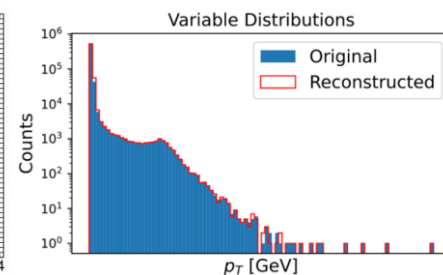
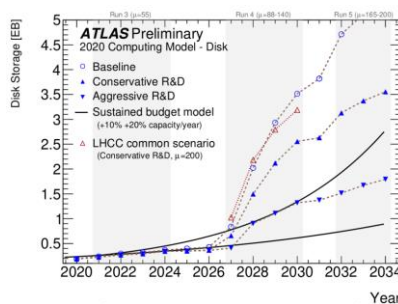
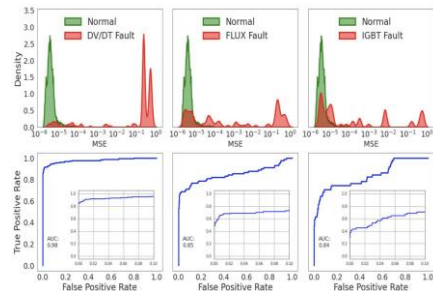
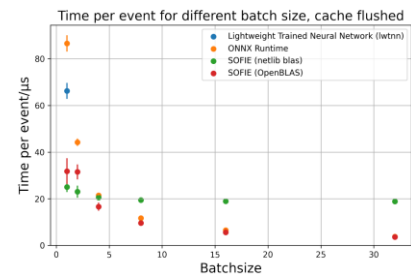
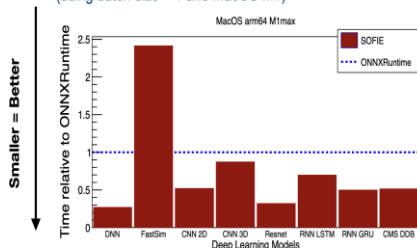


Figure 2: Toy example shows a visual representation of anomaly (red dot) and data leading to anomaly (yellow shaded area).



Test event performance of SOFIE vs ONNXRuntime (using batch size = 1 and MacOS M1)



Impact of JetNet

30k downloads

JetNet has enabled a lot of exciting research in ML for jets!

- 2024: Diffusion for Particle Cloud Generation in High Energy Physics. Use Case: accelerated jets via JetNet for diffusion models, and metrics via JetNet for evaluating the jets. Paper: 2303.05376, 2304.01295
- 2024: Equivariant Point Cloud Generation for Particle Jets. Use Case: accelerated jets via JetNet to train a Lorentz-equivariant auto-encoder for compression and anomaly detection. Paper: 2301.08128
- 2024: Equivariant Point Cloud Generation for Particle Jets. Use Case: accelerated jets via JetNet to train a Lorentz-equivariant auto-encoder for compression and anomaly detection. Paper: 2212.02362
- 2024: Equivariant Point Cloud Generation for Particle Jets. Use Case: accelerated jets via JetNet and distort them to test evaluation metrics. Paper: 2211.10090
- 2024: Equivariant Point Cloud Generation for Particle Jets. Use Case: Uses TopTagging defined and observed in the ParticleNet model's classification score. Paper: 2211.09912

总结

- ❖ 机器学习已经广泛应用于各个实验的各个方面，具有广阔的前景
- ❖ 基于机器学习的模拟是目前研究的热点之一
 - 量能器快速模拟，End-to-end 快速模拟
 - 实验数据驱动的模拟可以减小系统误差（ BES3 dE/dx 、量能器模拟等 ）
- ❖ 面对未来高噪声水平的情况（例如 BEPCII 升级，STCF 实验），机器学习在降低漂移室噪声、提高寻迹效率方面应该能发挥比较重要作用
- ❖ 在线刻度能够提高实验数据的质量，应该是未来发展的方向之一

谢谢！

Backup

Detector simulation

❖ Calorimeter Fast simulation:

- [FastCaloGAN: a fast simulation of the ATLAS Calorimeter with GANs](#)
- [gaede_chep23_caloml_v01 \(jlab.org\)](#)
- [Generating Accurate Showers in Highly Granular Calorimeters Using Normalizing Flows](#)
- [Fast and Accurate Calorimeter Simulation with Diffusion Models](#)
- [Transformers for Generalized Fast Shower Simulation](#)

❖ Ultra-fast simulation

- [THE LHCb ULTRA-FAST SIMULATION OPTION, LAMARR](#)
- [Flashsim: an ML simulation framework](#)

❖ [Refining fast simulation using machine learning](#)

❖ [Hadronic Simulation with conditional Masked Autoregressive Flow](#)

Reconstruction

❖ PID:

- [Particle identification with machine learning in ALICE Run 3](#)
- [Pion/Kaon Identification at STCF DTOF Based on Classical/Quantum Convolutional Neural Network](#)
- [Fast Inclusive Flavor Tagging at LHCb](#)
- [A deep-learning reconstruction algorithm for cluster counting](#)

❖ Vertex:

- [Advances in developing deep neural networks for finding primary vertices in proton-proton collisions at LHC](#)

❖ Tracking:

- [An Object Condensation Pipeline for Charged Particle Tracking](#)
- [End-to-End Geometric Representation Learning for Track Reconstruction](#)
- [BESIII track reconstruction algorithm based on machine learning](#)
- [Track Identification for CLAS12 using Artificial Intelligence](#)
- [Novel fully-heterogeneous GNN designs for track reconstruction at the HL-LHC](#)
- [HyperTrack: neural combinatorics for high energy physics](#)

❖ Cluster:

- [Improved Clustering in the Belle II Electromagnetic Calorimeter with Graph Neural Networks](#)
- [Development of particle flow algorithms based on Neural Network techniques for the IDEA calorimeter at future colliders](#)

❖ LArTPC:

- [Neutrino interaction vertex-finding in a DUNE far-detector using Pandora deep-learning](#)
- [Scalable, End-to-End, Machine-Learning-Based Data Reconstruction Chain for Particle Imaging Detectors](#)
- [Graph Neural Network for 3D Reconstruction in Liquid Argon Time Projection Chambers](#)

Online

❖ Trigger:

- [Applications of Lipschitz neural networks to the Run 3 LHCb trigger system](#)
- [Development of a Deep-learning based Full Event Interpretation \(DFEI\) algorithm for the future LHCb trigger](#)

❖ Continual Learning :

- [The Deployment of Realtime ML in Changing Environments](#)
- [Embedded Continual Learning for HEP](#)

❖ FPGA:

- [Symbolic Regression on FPGAs for Fast Machine Learning Inference](#)
- [Machine Learning for Real-Time Processing of ATLAS Liquid Argon Calorimeter Signals with FPGAs](#)
- [Acceleration of a CMS DNN based Algorithm](#)

Analysis

- ❖ Efficient search for new physics using Active Learning in the ATLAS Experiment
- ❖ Using a Neural Network to Approximate the Negative Log Likelihood Distribution
- ❖ A method for inferring signal strength modifiers by conditional invertible neural networks

Anomaly Detection

- ❖ Multi-Module based VAE to predict HVCM faults in the SNS accelerator
- ❖ Resilient Variational Autencoder for Unsupervised Anomaly Detection at the SLAC Linac Coherent Light Source

General

- ❖ [Uncertainty Aware Machine Learning Models for Particle Physics Applications](#)
- ❖ [Exploring Interpretability of Deep Neural Networks in Top Tagging](#)
- ❖ [JetNet library for machine learning in high energy physics](#)
- ❖ [New developments of TMVA/SOFIE: Code Generation and Fast Inference for Graph Neural Networks](#)
- ❖ [Machine learning based compression for scientific data](#)