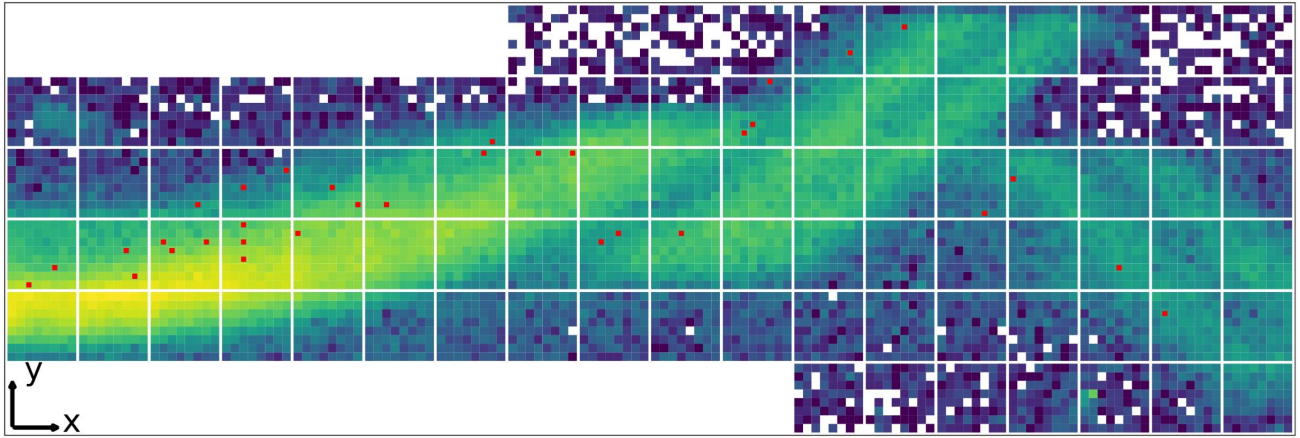
AI/ML for Imaging Cherenkov Reconstruction and Simulation



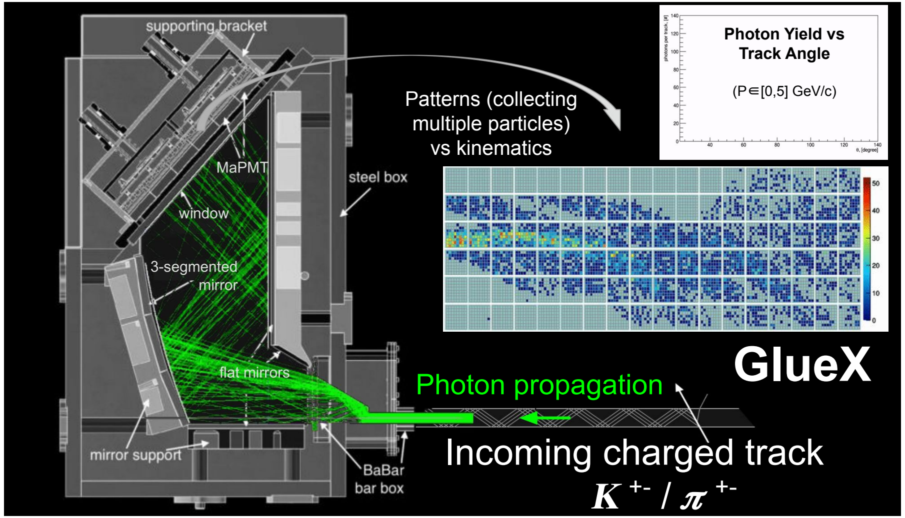
James Giroux

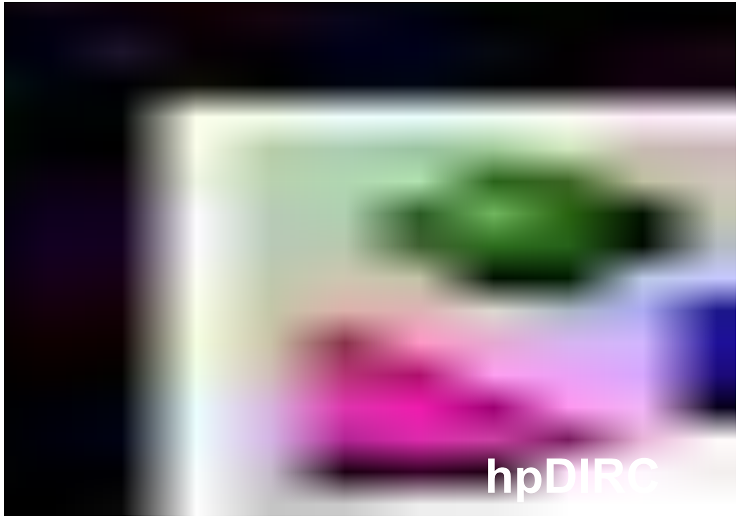
July 2026 Joint ePIC/EICUG Meeting

7/16/2026

WILLIAM & MARY
CHARTERED 1693
DATA SCIENCE

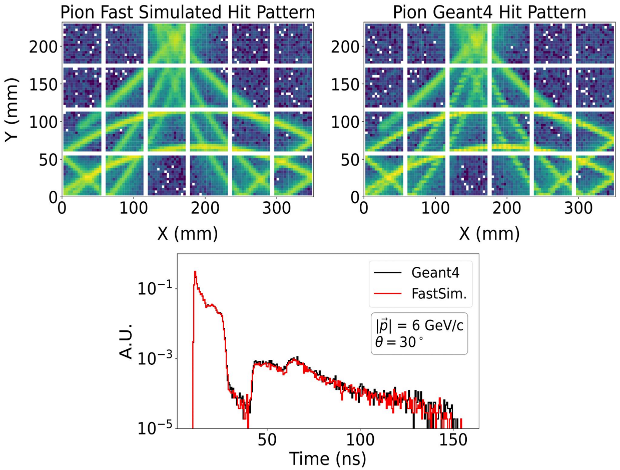
Generative Modeling for Cherenkov Detectors

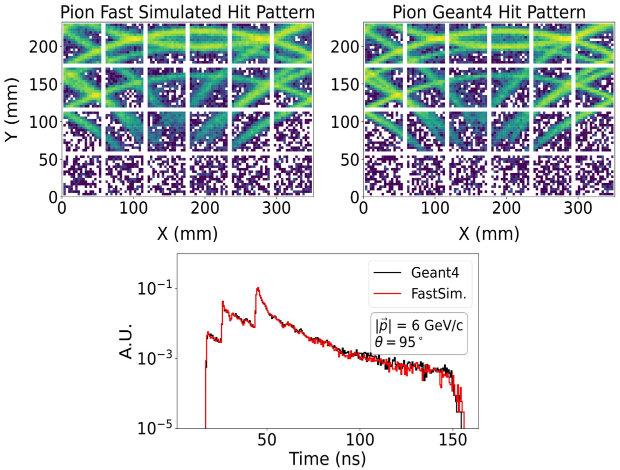


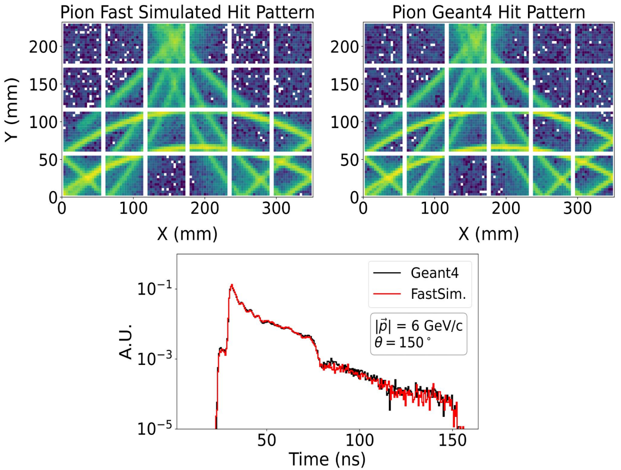


- Develop generalizable approach for *standard* generative algorithms
  - Models can be reused for different detectors
- Only data changes with varying detector geometries
- Generating Cherenkov photons one-by-one
  - Assume weak correlation at readout system

A Collection of Generative Models





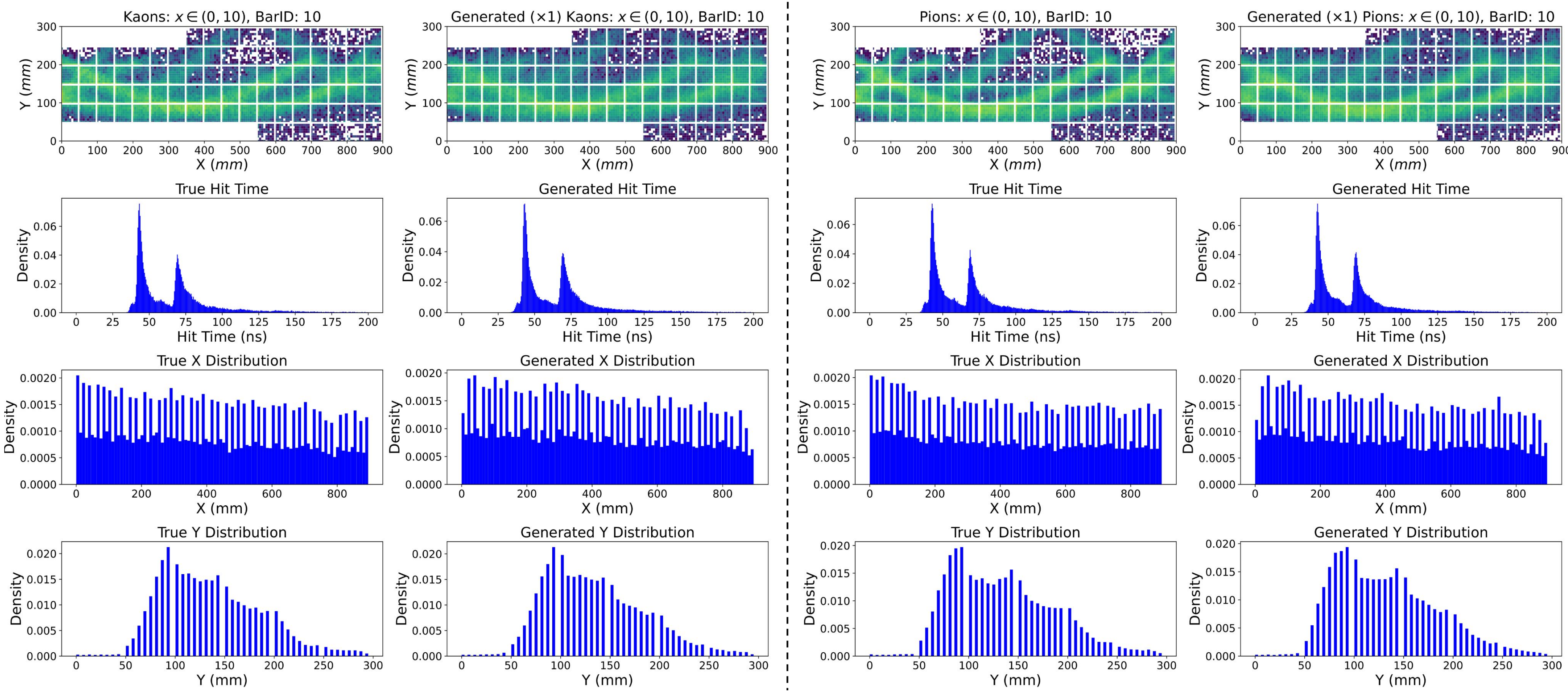


Normalizing Flows

hpDIRC Visualization of Generations at 6 GeV/c and various polar angle

A Collection of Generative Models

Normalizing Flows



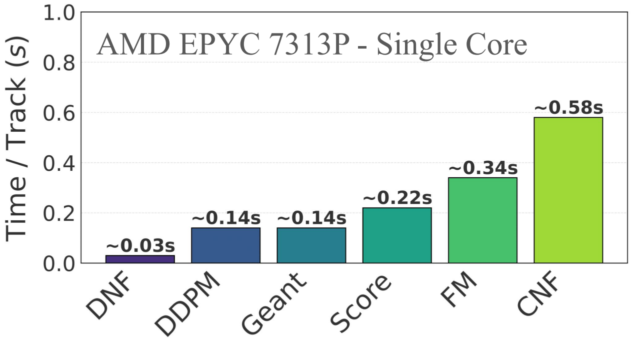
GlueX Visualization of Generations along specific bar regions

WILLIAM & MARY CHARTERED 1693 | DATA SCIENCE

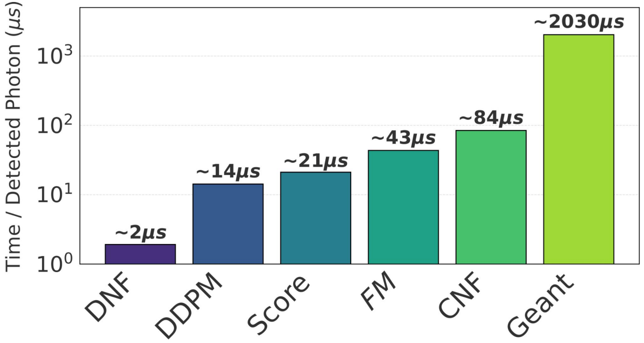
Takeaways

Github

- All Cherenkov Rings (underlying PDF's) generated by our models are "correct"
  - Ring and time structures follow correct kinematic dependencies for both PIDs
  - We incur a smoothing effect - can cause different PIDs to appear more similar
- Beyond usage in Physics environments
  - We have created an open source suite of SOTA algorithms (easily adapted to other detectors)
  - Our fast simulation is self contained, fast and capable of being run on CPU or GPU



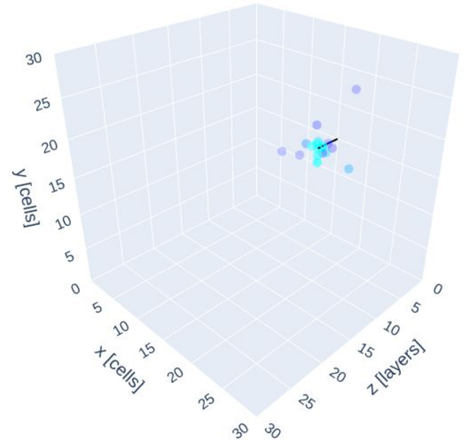
Track Generation (CPU)

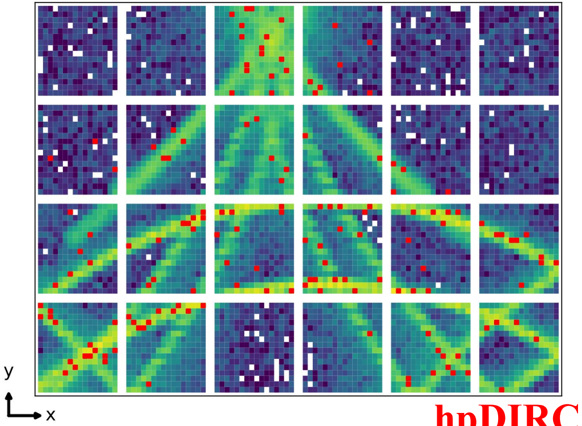


PDF Generation (GPU)

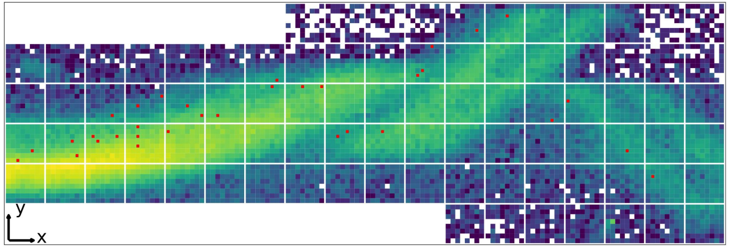
Paper Links: hpDIRC; GlueX

Generalizable Foundation Models for Readout Systems combining Discrete and Continuous Data





hpDIRC



GlueX

Calorimetry

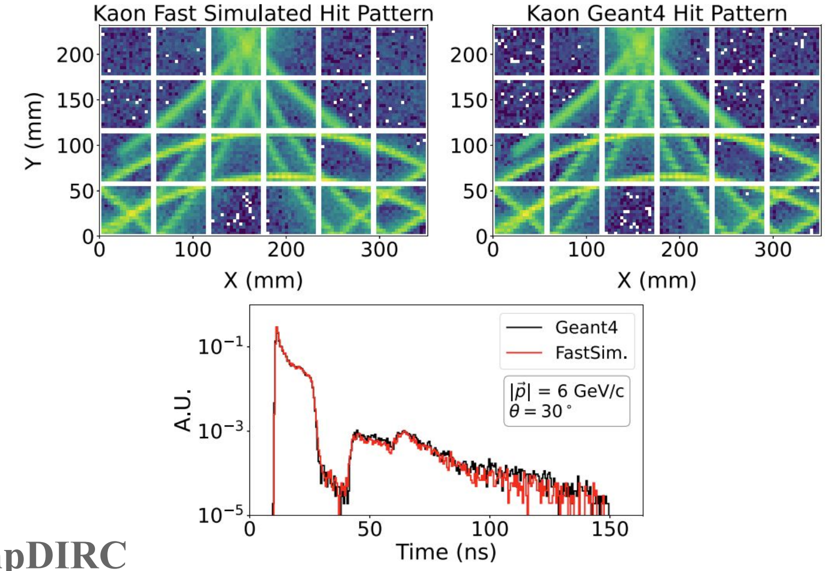
DIRC

WILLIAM & MARY
CHARTERED 1693
DATA SCIENCE

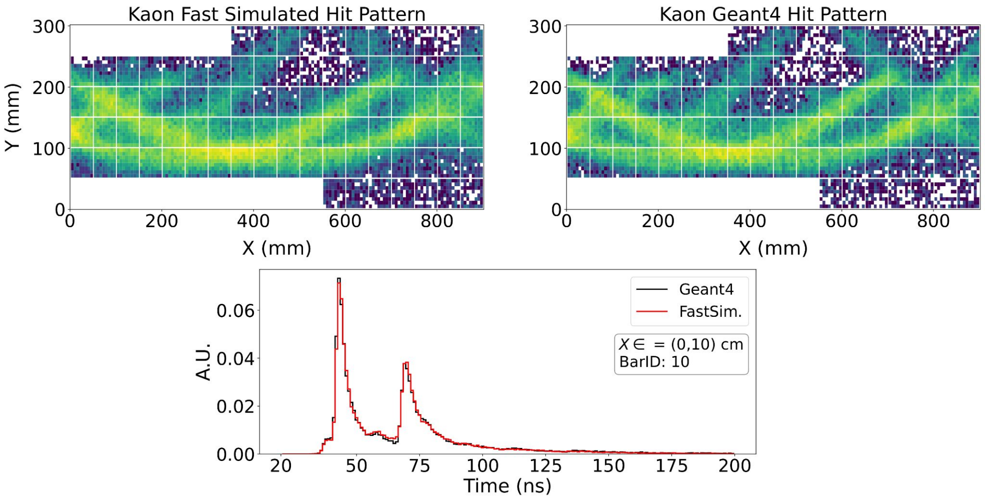
Foundation Models in Physics

- Foundation Models (FM) are becoming increasingly popular in the Physics community
  - Relatively large pre-trained models
- Capable of supporting multiple downstream tasks - e.g., fast simulation, reconstruction, noise filtering, etc.
  - Leverage multi-task learning paradigms
  - Fine-tuning across related tasks

Fast Simulation



hpDIRC

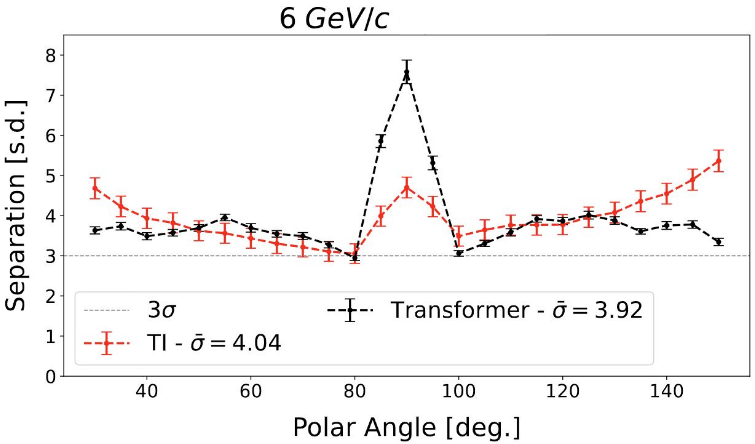


GlueX

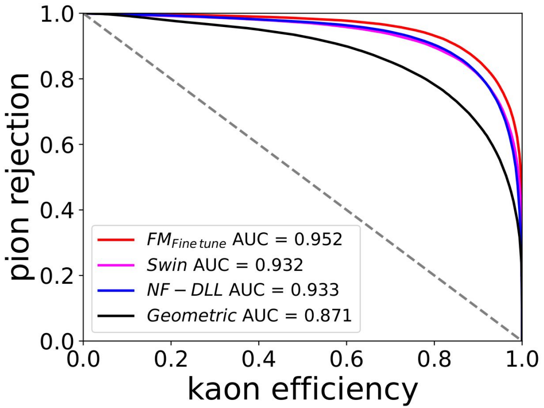
Both models trained over fully continuous phase-space

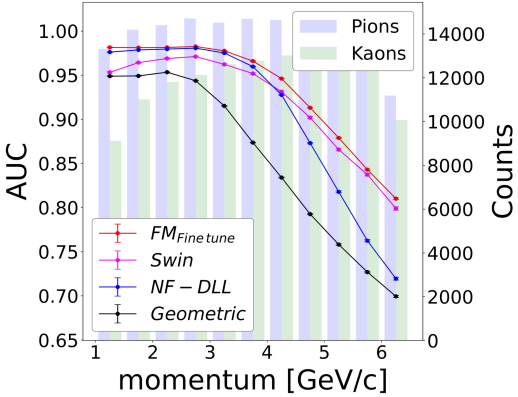
WILLIAM & MARY CHARTERED 1693 | DATA SCIENCE

Particle Identification (π/K)



hpDIRC

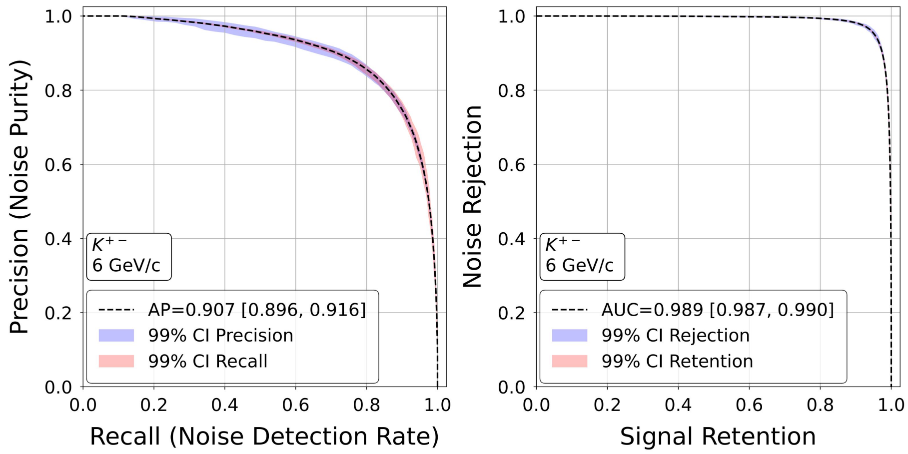




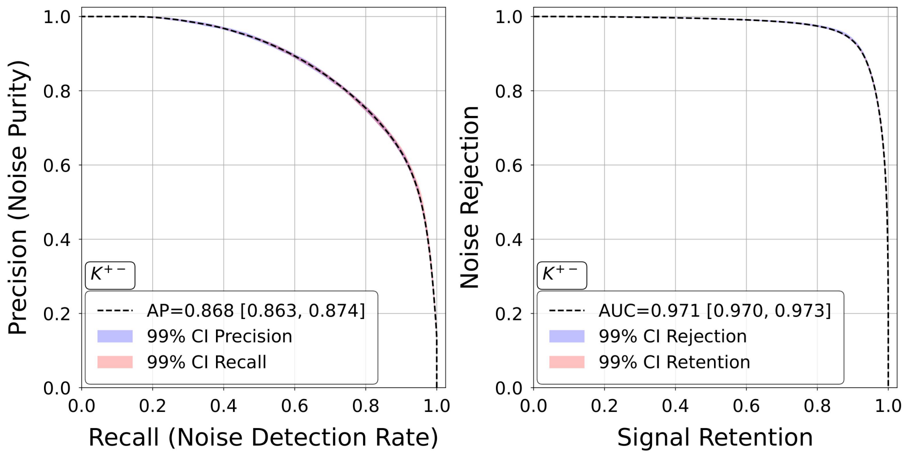
GlueX

Both models trained over fully continuous phase-space

Noise Filtering



hpDIRC ~ 100khz/cm^2



GlueX ~ 25khz/cm^2

Both models trained over fully continuous phase-space

Key Takeaways

- **Foundation Model** operating directly on low level detector signals
  - Generalizable to other detectors with discrete + continuous data streams
- Four core innovations
  - **Dual Vocabularies** for spatial and temporal data - fused through **Cross Attention**
  - **Scaleable**, **higher resolution tokenization** with joint vocab inflation - our resolutions correspond to 36M tokens in a joint vocabulary
  - **Continuous conditioning** using prepending kinematic embeddings
  - **Class conditional generation** through a Mixture of Experts
- Capable of multiple downstream tasks
  - Generation
  - Particle Identification
  - Noise Filtering

hpDIRC
GlueX
Calorimetry

Paper Links: hpDIRC, GlueX, Calorimetry