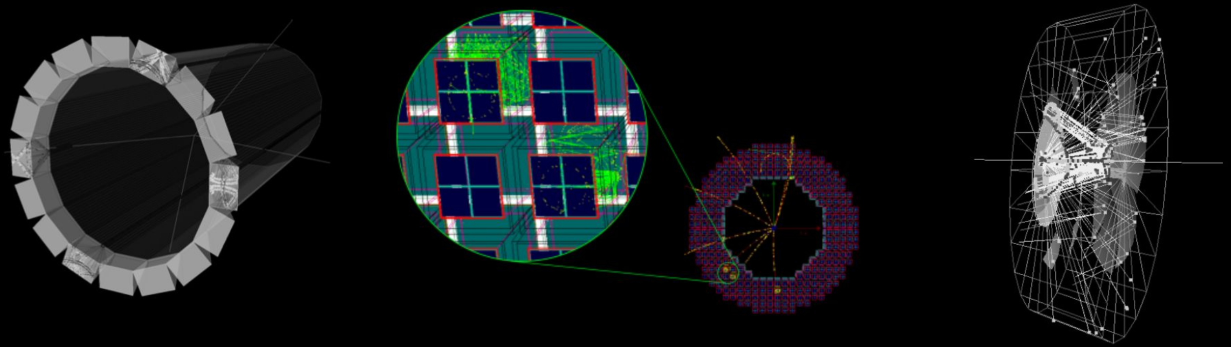
ML/DL for Cherenkov Detectors



Cristiano Fanelli

Workshop on kaons at CLAS12 — 12/13-16/2022, LNF

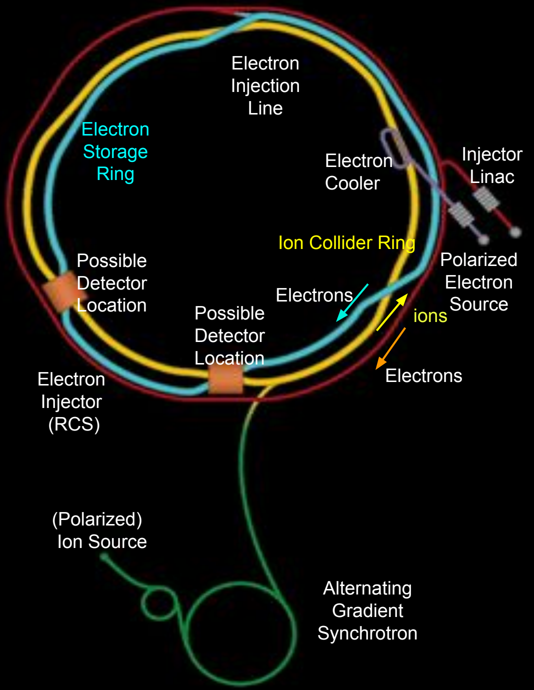
Jefferson Lab
Exploring the Nature of Matter

Outline

- Imaging Cherenkov for PID
  - EIC as an example
    - DIRC
      - Reco, alignment
    - dRICH
      - Reco / hackathon
- Conclusions

Electron Ion Collider

A precision tool to study the glue that binds visible matter



polarized electron - polarized protons/ions

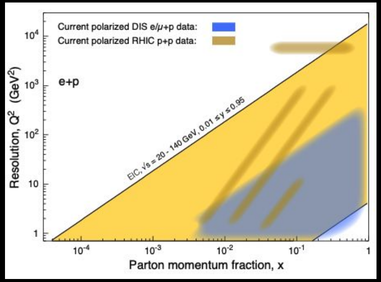
e: 5 to 18 GeV

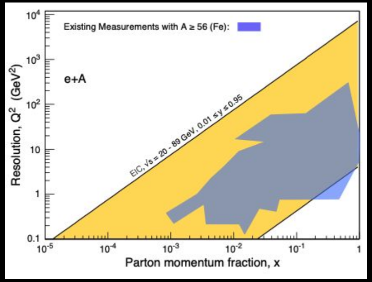
p: 41, 100 to 275 GeV

CoM energy $\sqrt{s}_{e\text{-}p}$ ~ (20-140) GeV

Center of Mass Energies	20 GeV – 140 GeV
Maximum Luminosity	10^{34} cm^{-2}s^{-1}
Hadron Beam Polarization	80%
Electron Beam Polarization	80%
Ion Species Range	p to Uranium
Number of interaction regions	up to two

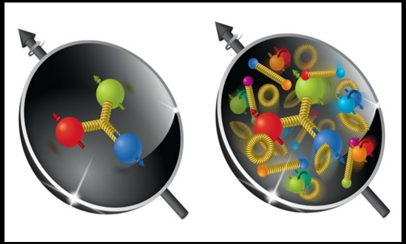
uncovered x-Q^2 range



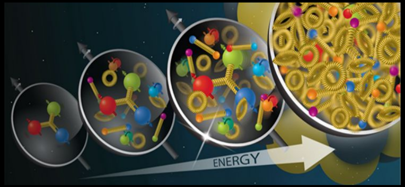


How does the mass of the nucleon arise?

How does the spin of the nucleon arise?



What are the emergent properties of dense systems of gluons?



Jefferson Lab

PID with Cherenkov: EIC example

		electrons/photons		π/K/p	
eta	Nomenclature	PID	Min E Photon	P-range [GeV/c]	Separation
-3.5 to -2.0	Backward	π suppression up to 1:1E-4	20 MeV	≤ 10 GeV/c	≤ 3σ
-2.0 to -1.0	Backward	π suppression up to 1:1E-3 - 1:1E-2	50 MeV		
-1.0 to 1.0	Barrel	π suppression up to 1:1E-2	100 MeV	≤ 6 GeV/c	
1.0 to 3.5	Forward	3σ e/π up to 15 GeV/c	50 MeV	≤ 50 GeV/c	

"**Simulations show that in order to satisfy the physics goals of the EIC, it is desirable to provide π/K identification** in the central barrel up to 5-7 GeV/c, in the electron-going endcap up ~10 GeV/c, and **in the hadron-going endcap one would need to reach ~50 GeV/c**.", from the "Electron-Ion Collider Detector Requirements and R&D Handbook", January 10, 2019

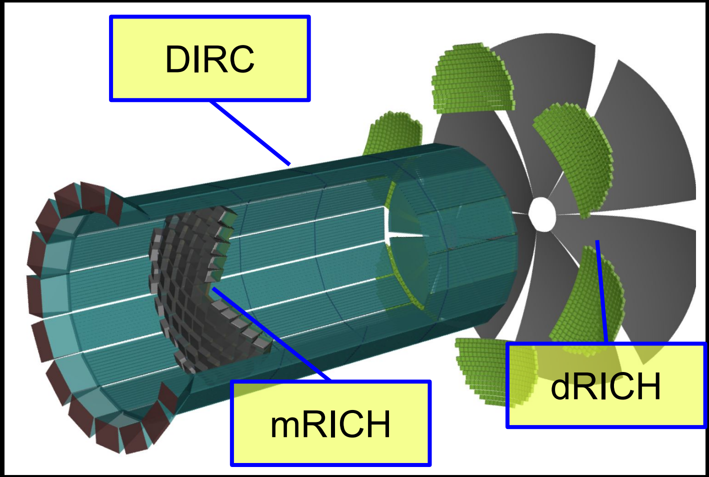
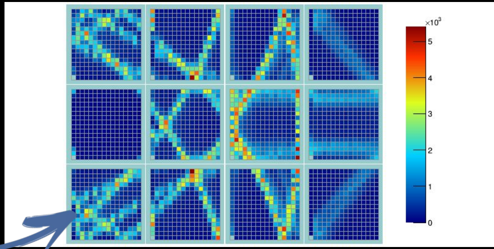
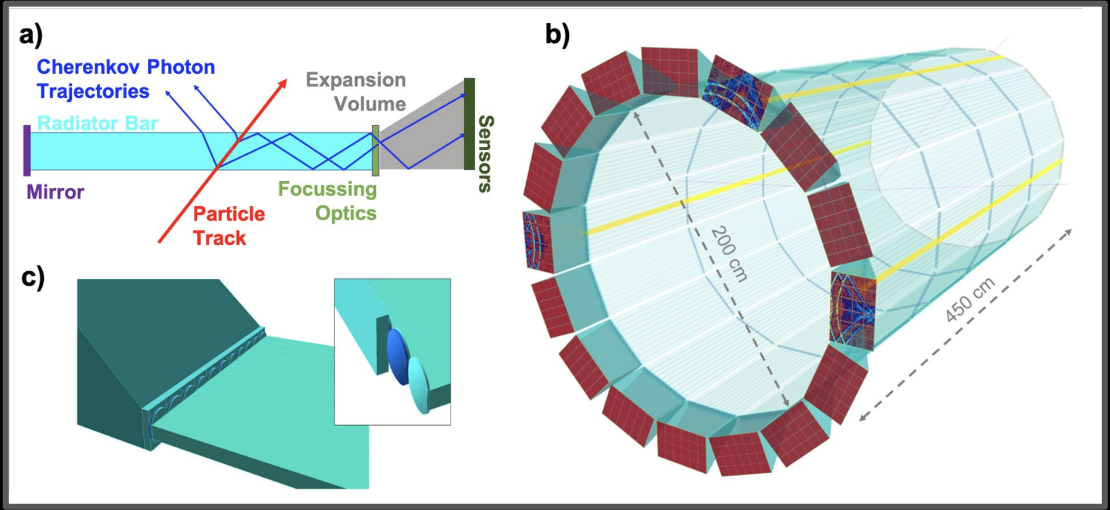


Image does not reflect latest design version of ePIC PID systems.

- Cherenkov detectors form the backbone of PID at EIC
  - Currently the ePIC detector uses a dual radiator ring-imaging Cherenkov detector (RICH) in the hadron direction, a DIRC (detection of internally reflected Cherenkov light) in the barrel, and a modular RICH in the electron direction.
  - Simulating these detectors is typically compute expensive, involving many photons that need to be tracked through complex surfaces.
  - All three rely on pattern recognition of ring images in reconstruction, and the DIRC is the one having the more complex ring patterns.

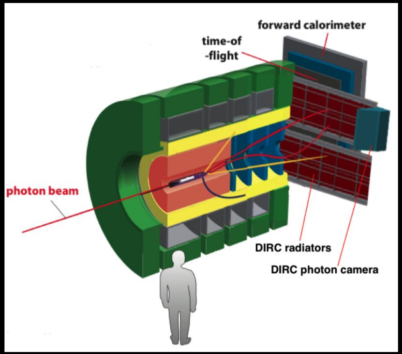


hpDIRC at EIC



- 16-sided polygonal barrel at a radius of 1 m
- Each bar box includes a set of eleven radiator bars, made of synthetic fused silica bars, each 4200 mm long, with a cross section of 17 mm × 32.7 mm
- Sensor options under consideration are MCP-PMTs and SiPM.

DIRC utilized in GlueX experiment



G. Kalicy, "Developing high-performance DIRC detector for the future Electron Ion Collider experiment." *Journal of Instrumentation* 15.11 (2020): C11006.

Image of expansion volume taken from GlueX DIRC,
Ali et al., JINST 15 (2020) 04, C04054

DIRC at GlueX (JLab)

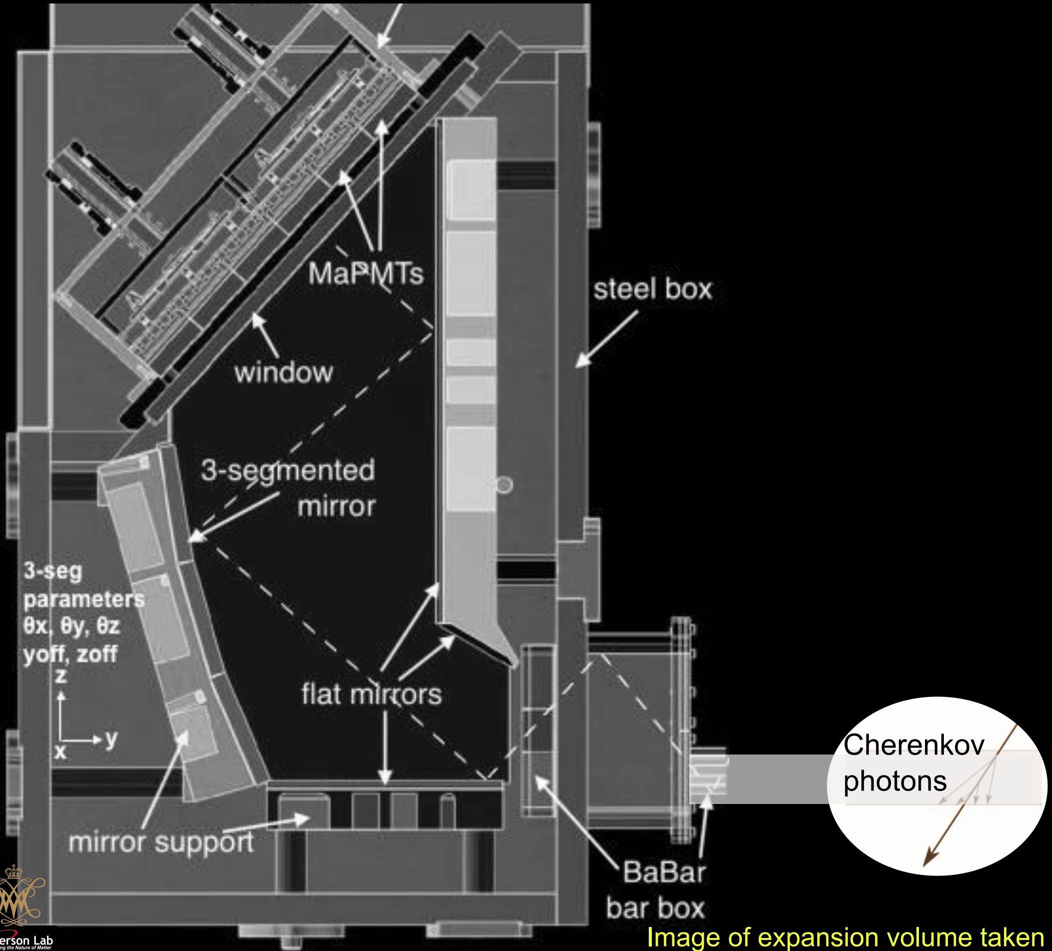
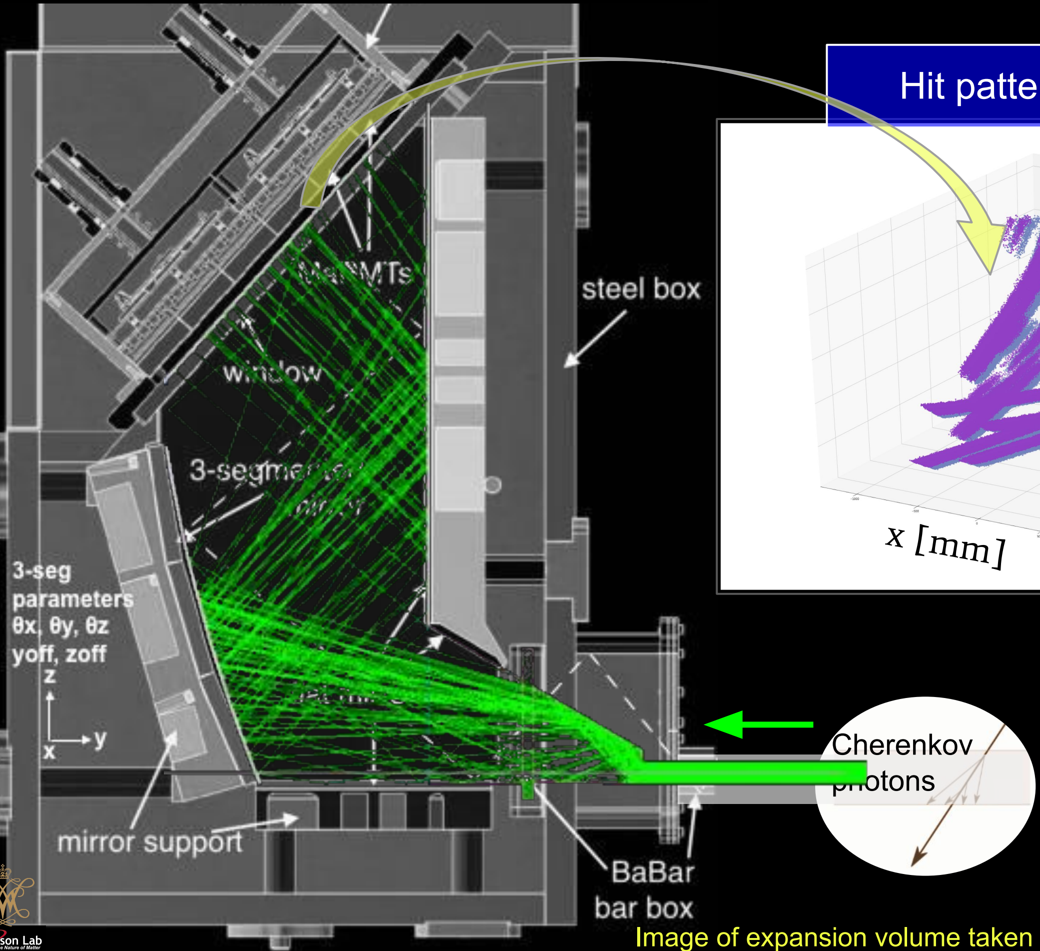


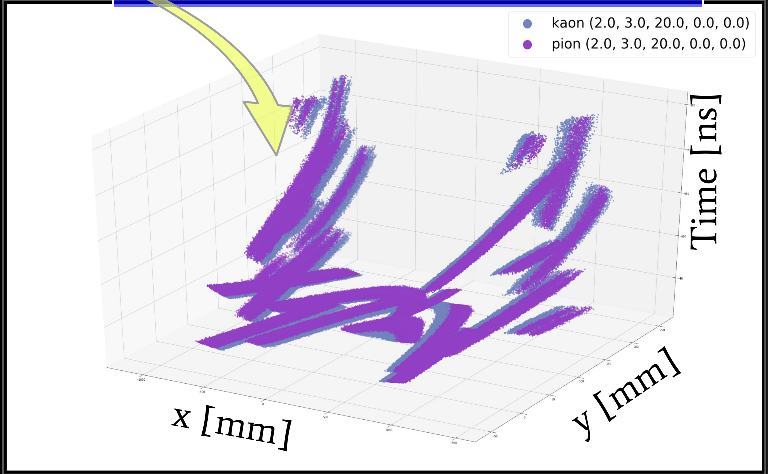
Image of expansion volume taken from GlueX DIRC, Ali et al., JINST 15 (2020) 04, C04054

Image of expansion volume taken from GlueX DIRC,
Ali et al., JINST 15 (2020) 04, C04054

DIRC at GlueX (JLab)



Hit pattern defined in (x,y,t)



3D (x,y,t) readout allows to separate spatial overlaps.

Patterns take up significant fractions of the PMT in x,y and are read out over 50-100 ns due to propagation time in bars.

H12700 PMTs have a time resolution of O(200 ps) and read-out electronics giving time information in 1 ns buckets.

1PMT made by 64 pixels, each pixel is 6mm x 6mm size

Displayed PDF. Patterns are sparse with variable photon yield

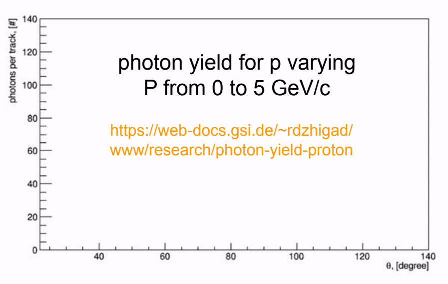
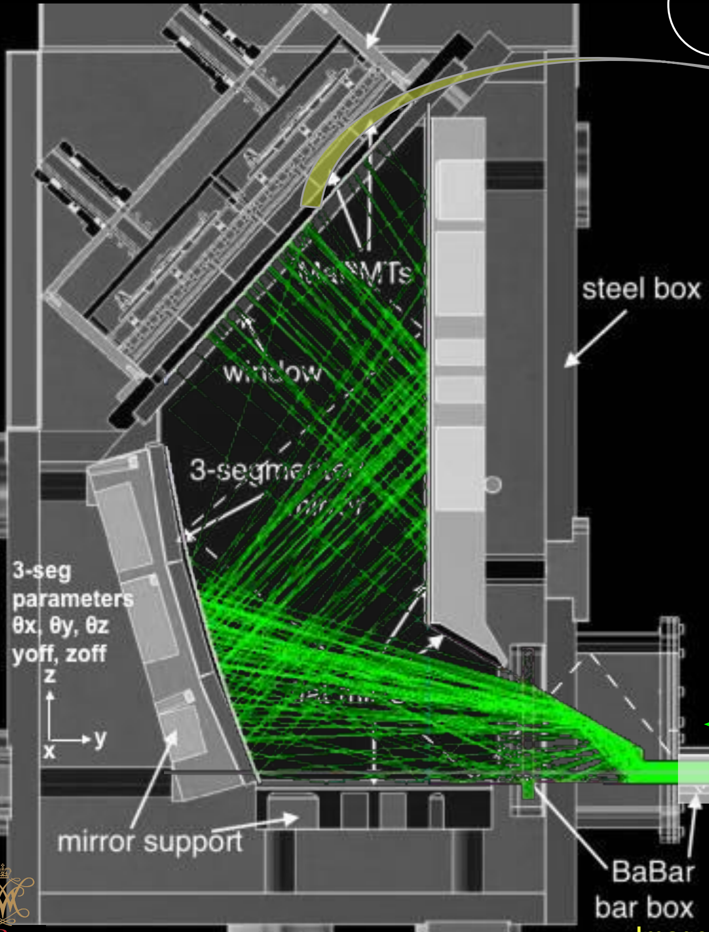


Image of expansion volume taken from GlueX DIRC, Ali et al., JINST 15 (2020) 04, C04054

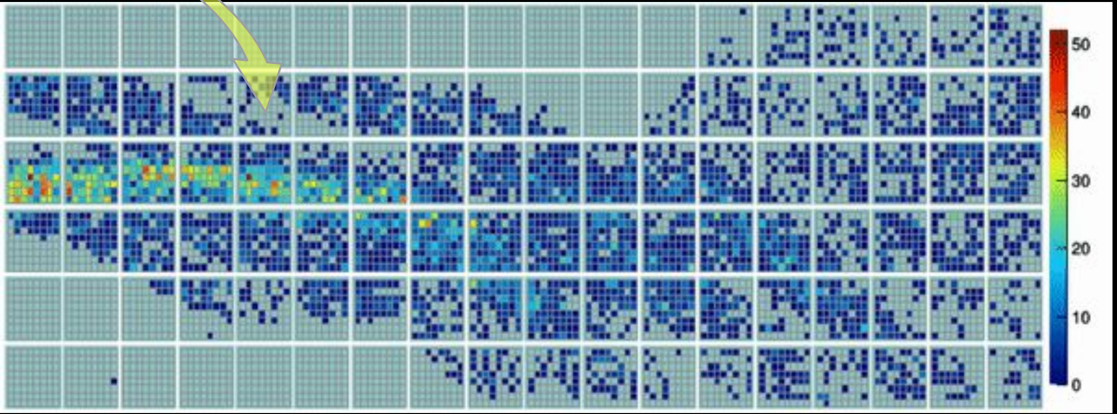
DIRC at GlueX (JLab)

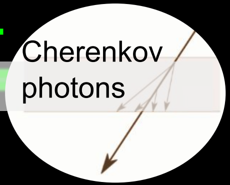
Image of expansion volume taken from GlueX DIRC,
Ali et al., JINST 15 (2020) 04, C04054

2



Kaons @ 4 GeV/c for different polar and azimuthal angle



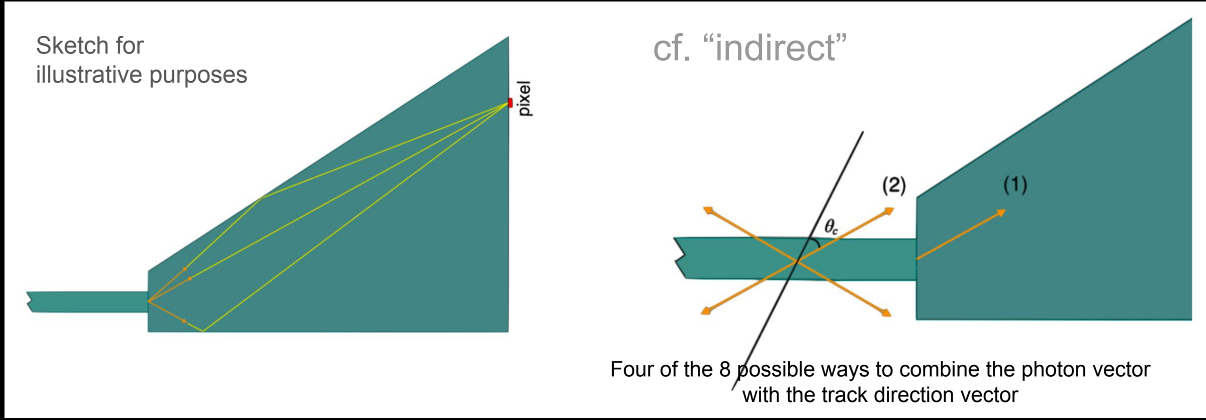


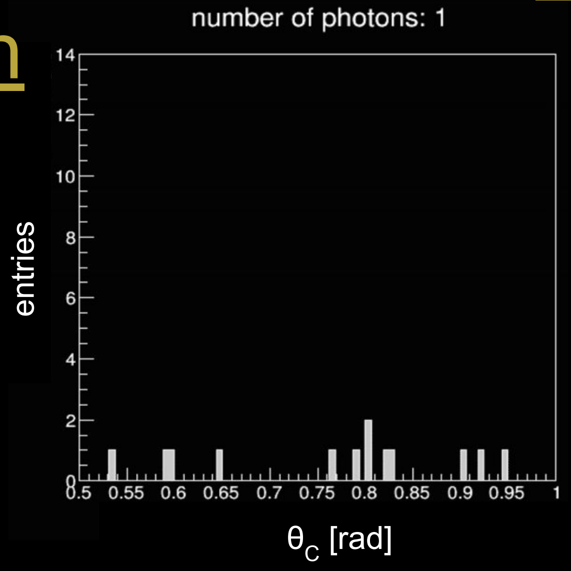
Dependence on charged particle kinematics
(p,(θ,φ)*,X,Y)

1

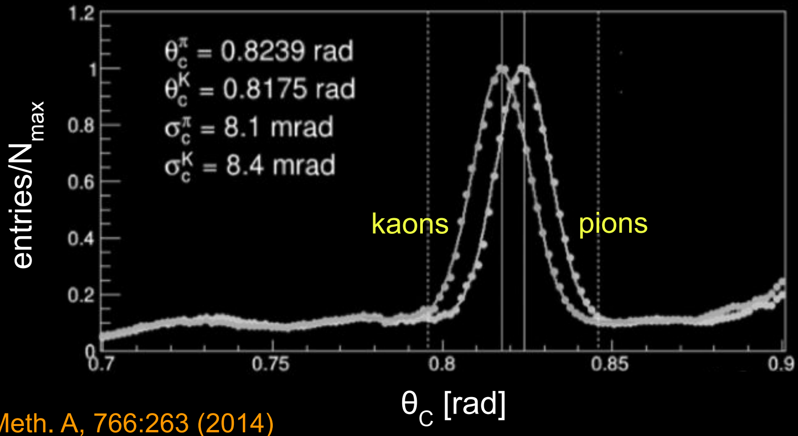
Image of expansion volume taken from GlueX DIRC, Ali et al., JINST 15 (2020) 04, C04054

Geometrical Reconstruction





- All possible photon paths from the bar to each pixel are stored in look-up tables
- The Cherenkov angle is determined by calculating the angle between the photon direction from the LUT and the charged track direction from the tracking system
- Fast reconstruction/hit pattern
- Other approaches possible (e.g., time-based imaging utilizes detection time per pixel; superior reconstruction but memory typically hungry)



R. Dzhygadlo et al. Nucl. Instr. And Meth. A, 766:263 (2014)

Jefferson Lab

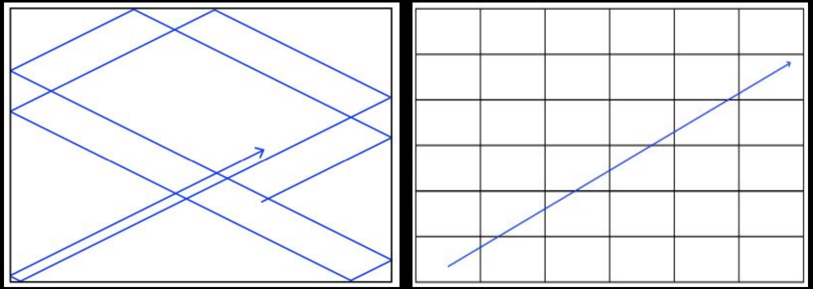
FastDIRC

cf. “direct”

https://github.com/jmhardin/FastDIRC

- Framework for Fast Monte Carlo and reconstruction.
- Simulations: fast tracing mapping straight lines through a tiled plane:
  1. Generation
  2. Traces through bars
  3. Traces through expansion volume

Each photon bounces O(100) times on average. Developed a billiard method that maps the bounces onto a straight-line trajectory through a tiled plane



10000 faster than Geant4:
This facilitates reconstruction of Cherenkov angle with an improvement of 30% as compared to the Geometrical approach.
Slower than LUT.

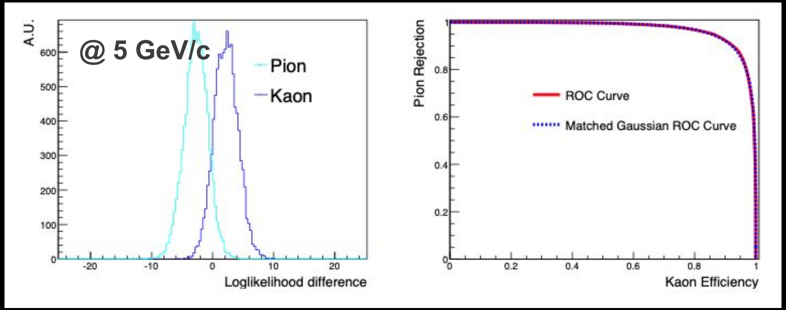
- PID strategy is likelihood based:
  - N_g photons are generated to produce the expected PDF.
  - N_g and λ chosen to provide best performance.

$$f(\vec{x}) \propto \sum_{i}^{n_\gamma^{\mathrm{sim}}} \mathscr{G}(|\vec{x}-\vec{x}_i|)$$

PDF

$$\log \mathcal{L}_{\pi(K)} = \sum_{j=1}^{N_d} \ln\left(\sum_{i=1}^{N_g^{\pi(K)}} g\left(\frac{|\mathbf{x}_i^{\pi(K)} - \mathbf{x}_j|}{\lambda}\right)\right)$$

likelihood



J. Hardin and M. Williams, FastDIRC, JINST 11.10 (2016)

Deep Reconstruction Imaging CHerenkov

- Cherenkov detectors are relatively slow to simulate with full simulations like Geant
    - for the DIRC case, each Cherenkov photon reflects on average O(10^2) times within a bar and this makes the simulation CPU intensive.
- Not many AI-based applications:
    - Some work on fast simulation with Cherenkov detectors [1].
    - Lack of ML/DL applications for reconstruction/identification:
        - Most of them use high-level features from Cherenkov detectors and combine them to other features from other sub-detectors for global PID [2].
- Can we build an AI-based architecture with the following desired properties?

- It is fast and provides accurate reconstruction
- Can be extended to multiple particle types
- Generalizes to fast simulation
- Can utilize (x,y,t) patterns if time is measured
- Can deal with different topologies and detectors
- Deeply learns the detector response (real data can be injected)

DeepRICH [3] is the first attempt in this direction.
I'll show prototype and discuss path forward.

DeepRICH: learning deeply Cherenkov detectors

Cristiano Fanelli[1,3] and Jary Pomponi[2]

Published 27 April 2020 • © 2020 The Author(s). Published by IOP Publishing Ltd

Machine Learning: Science and Technology, Volume 1, Number 1

Citation Cristiano Fanelli and Jary Pomponi 2020 *Mach. Learn.: Sci. Technol.* **1** 015010

DOI 10.1088/2632-2153/ab845a

1915 Total downloads

[1] D. Derkach et al., arXiv:1903.11788v1, 2019
[2] D. Derkach et al., J. Phys.: Conf. Ser. **1085** 042038, 2018
[3] C. Fanelli and J. Pomponi, Mach. Learn.: Sci. Technol. 1 015010, 2020

DeepRICH Architecture

- DeepRICH is a custom architecture that combines
  - VAE for reconstruction
  - CNN + MLP for classification
- The model is trained by minimizing the total loss function:

$$\mathcal{L}(\mathbf{x},\tilde{\mathbf{x}},\mathbf{y},\tilde{\mathbf{y}},\mathbf{l}) = \lambda_r\mathcal{L}_r(\mathbf{x},\tilde{\mathbf{x}}) + \lambda_c\mathcal{L}_c(\mathbf{y},\tilde{\mathbf{y}}) + \lambda_v\mathcal{L}_v(\mathbf{l})$$

Reconstruction loss (injected vs reco)

Classification loss

VAE MMD (latent)

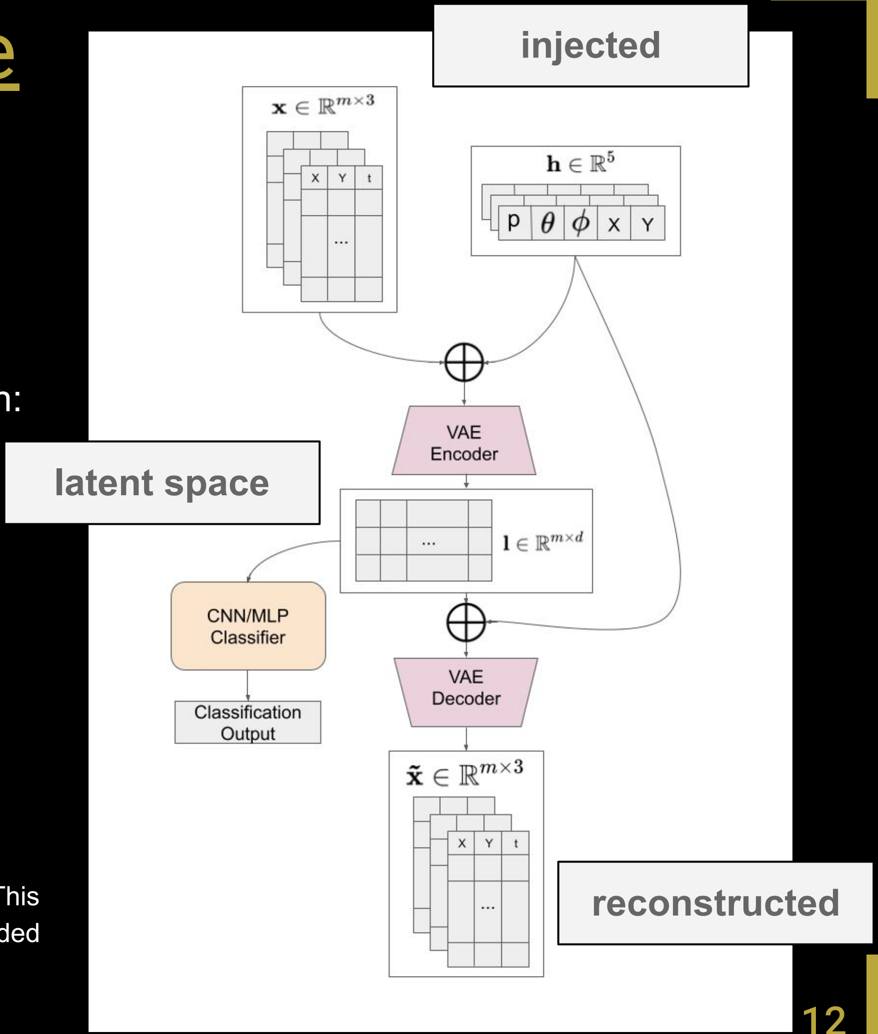
$$\mathcal{L}_r(x,\tilde{x}) = \frac{1}{3}\sum_i^3 z_i$$

$$z_i = \begin{cases} 0.5(x_i-\tilde{x}_i)^2, & \text{if } |x_i-\tilde{x}_i|<1 \\ |x_i-\tilde{x}_i|-0.5 & \text{otherwise,} \end{cases}$$

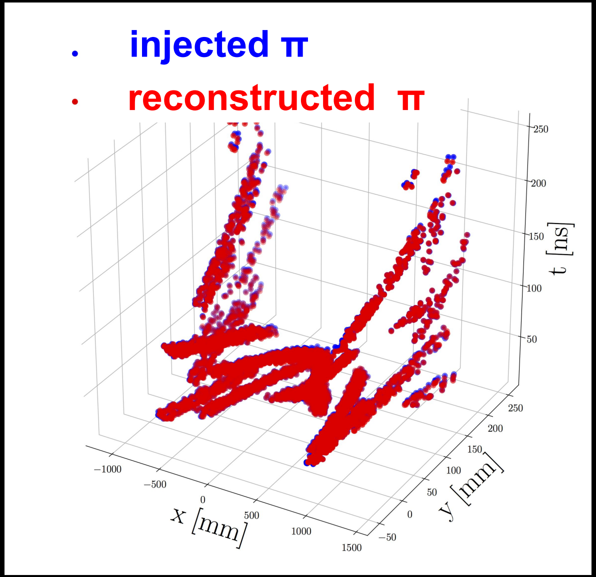
$$\mathcal{L}_c = -(\mathbf{y}\log(\tilde{\mathbf{y}}_0) + (1-\mathbf{y})\log(\tilde{\mathbf{y}}_1))$$

tackled π, K separation. This can be extended to, e.g., p

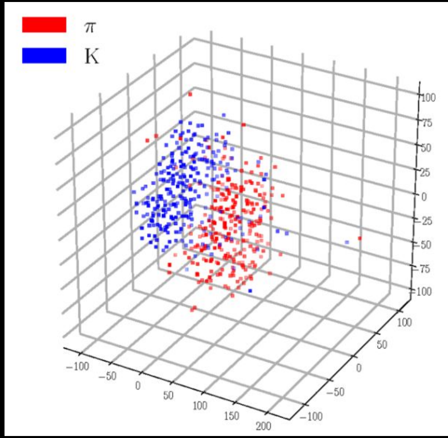
$$\mathcal{L}_v = MMD(\boldsymbol{p}(z),\boldsymbol{q}(z)) = \mathbb{E}_{\boldsymbol{p}(z),\boldsymbol{p}(z')}[\kappa(z,z')] + \mathbb{E}_{\boldsymbol{q}(z),\boldsymbol{q}(z')}[\kappa(z,z')] - 2\mathbb{E}_{\boldsymbol{p}(z),\boldsymbol{q}(z')}[\kappa(z,z')]$$



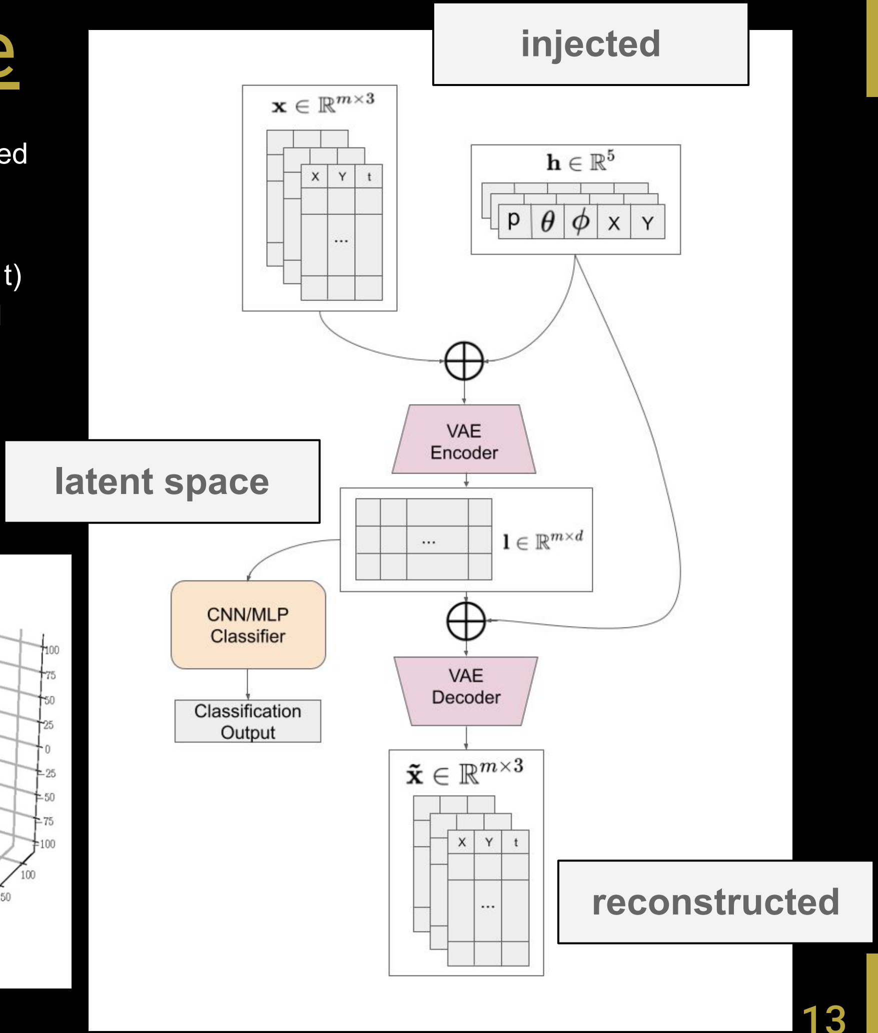
DeepRICH Architecture



Example of hit pattern detected in the PMT plane (spatial coordinates are dubbed x,y, while the time is indicated as t) simulated with FastDIRC and reconstructed with VAE

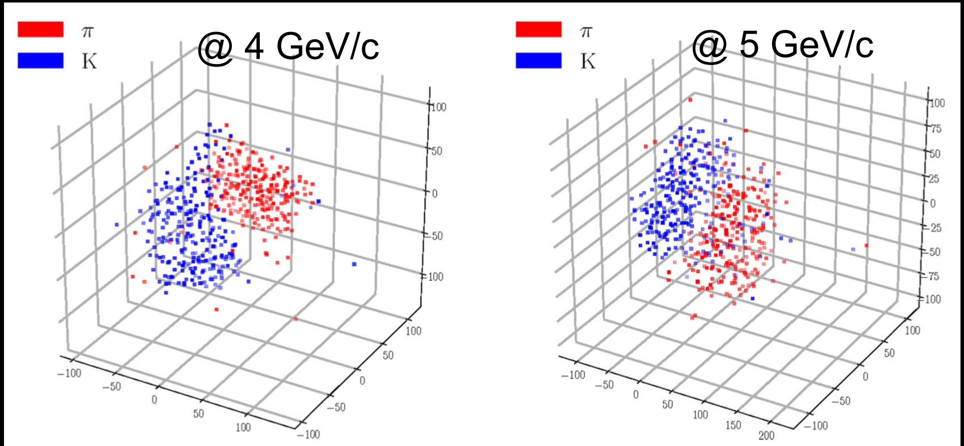


Example of features extracted by the CNN from π and K at 5 GeV. The plot shows separation power. The 3D visualization is obtained with t-SNE.

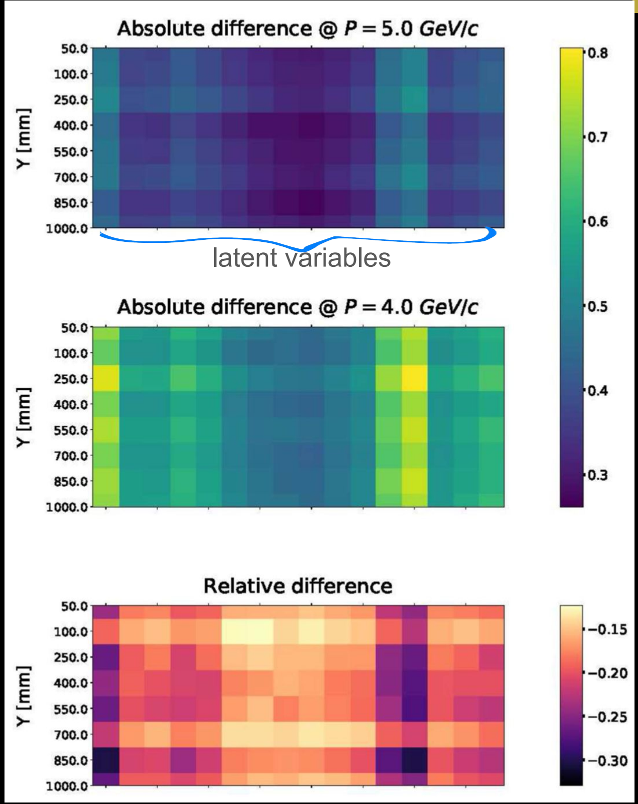


Performance

π/K distinguishing power:
visualizations



Example of features extracted by the CNN module from π's and K's at 4 GeV/c (left) and 5 GeV/c (right). These features are then used to classify the particle. The plot shows a better separation between π/*K* at 4 GeV/c, which means that the network has good distinguishing power. As expected the points become less separated at larger momentum. The 3D visualization is obtained with t-SNE.



2D plot of the absolute difference on each latent variable between π's and K's, obtained for 5 GeV/c and 4 GeV/c, respectively. The color indicates the absolute difference, the larger the difference the larger is the distinguishing power. As expected the separation becomes less clear at 5 GeV. Also there is no appreciable dependence on the position on the bar resulting in patterns with vertical bands. (Bottom) The relative difference showing negative values in the majority of the bins.

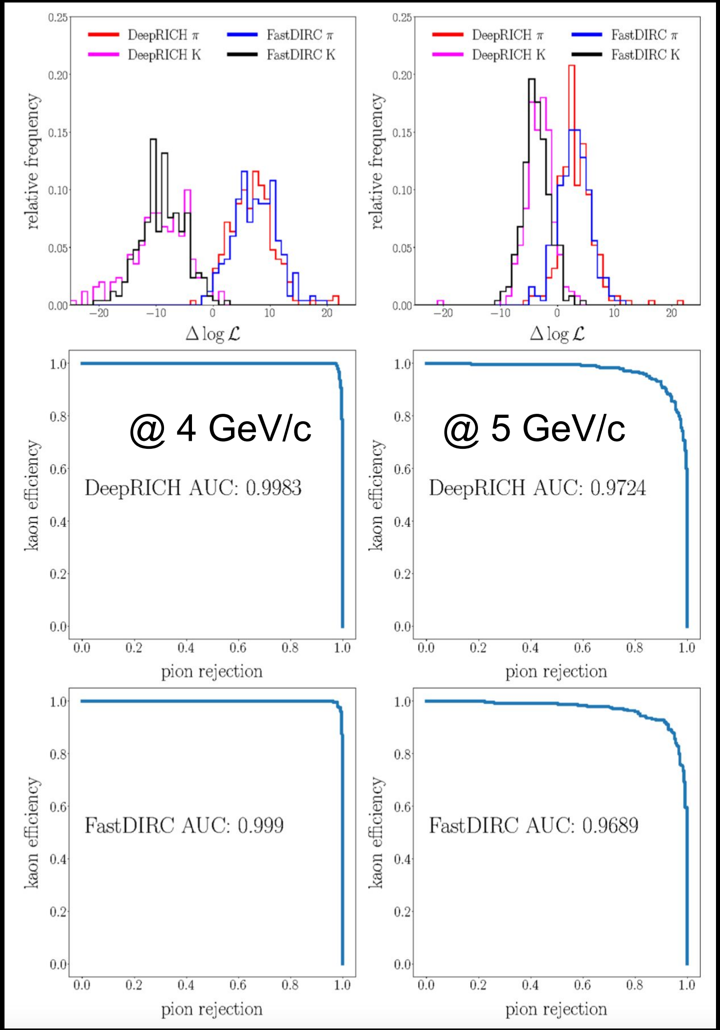
Performance

$$\log \mathcal{L}_{\pi(K)} = \sum_{j=1}^{N_d} \ln\Big(\sum_{i=1}^{N_g^{\pi(K)}} g\Big(\frac{|\mathbf{x}_i^{\pi(K)} - \mathbf{x}_j|}{\lambda}\Big)\Big)$$

In DeepRICH the output of the classifier is two-dimensional (π/K) and $\in R^2$.

- These values are utilized to build the DLL between π/K
- ROC is obtained by changing the threshold on the DLL. ROC curves are produced generating 350 particles for each kinematics
- The AUC is used as a metric to compare DeepRICH to FastDIRC

AUC(DeepRICH) ≳ 0.99 AUC(FastDIRC)



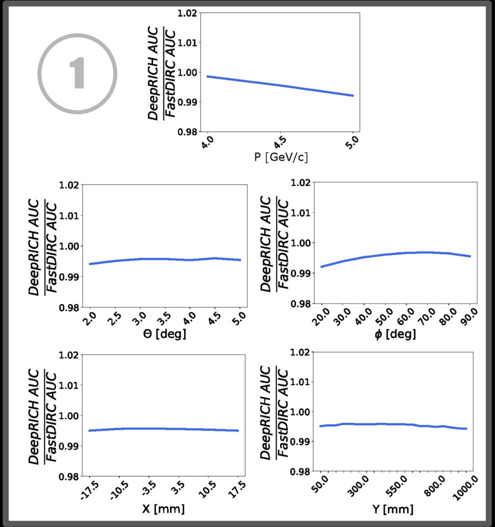
Performance

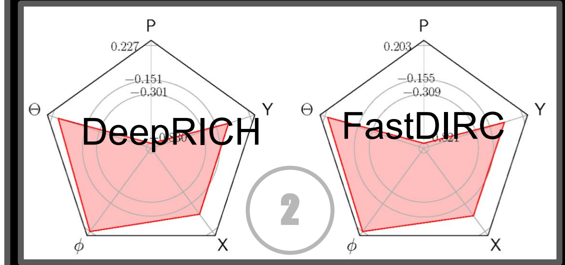
(1) The ratio between DeepRICH and FastDIRC AUCs: Each AUC is calculated to show the partial dependence on one kinematic parameter by marginalizing on all other parameters. Notice at 4 GeV/c that the two reconstruction methods perform almost identically.

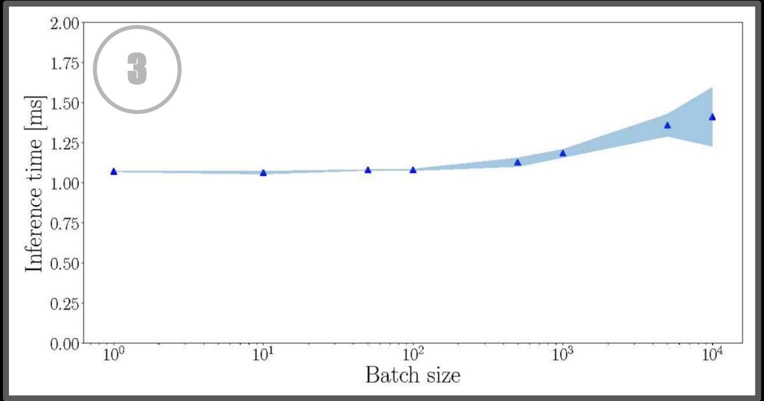
(2) Radar plots of correlation between the AUC and each kinematics parameter for DeepRICH and FastDIRC. The two reconstruction algorithms perform similarly as a function of the kinematic parameters. AUC depends on momentum, as distinguishing power gets lower at larger momentum.

	DeepRICH			FastDIRC		
Kinematics	AUC	ε_S	ε_B	AUC	ε_S	ε_B
4 GeV/c	99.74	98.18	98.16	99.88	98.98	98.85
4.5 GeV/c	98.78	95.21	95.21	99.22	96.33	96.32
5 GeV/c	96.64	91.13	91.23	97.41	92.40	92.47

specs	value
inference time per batch	$\mathcal{O}(1)$ ms
inference network memory	$\mathcal{O}(1)$ GB
training network memory	$\mathcal{O}(4)$ GB
network memory on local storage	$\sim$ 6 MB
network trainable parameters	458 592



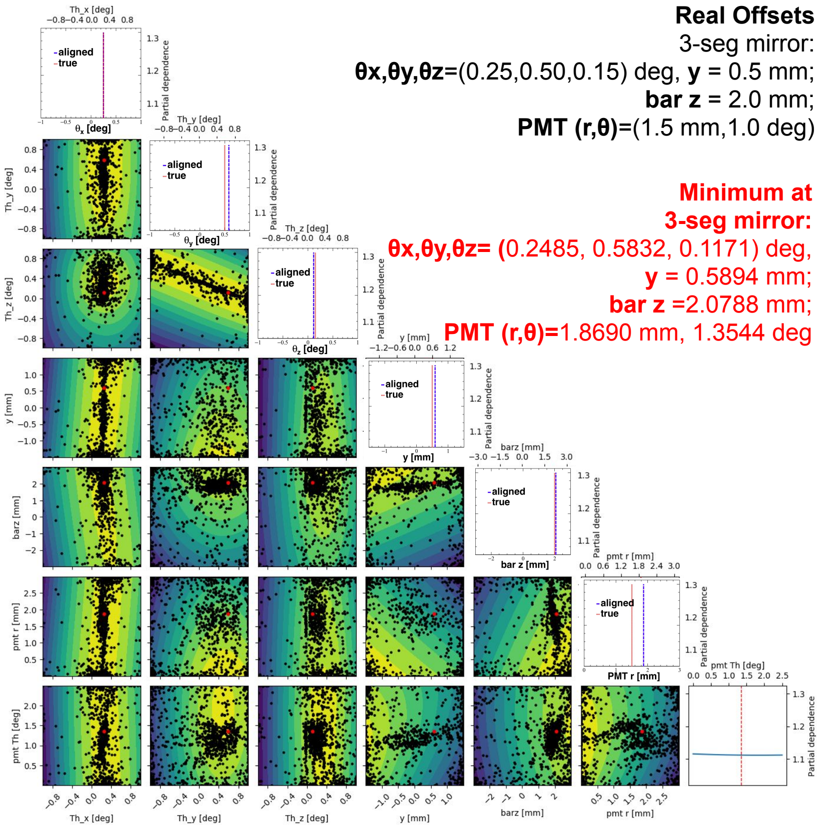




(3) Inference time: inference time is almost constant as a function of the batch size, meaning that the effective inference time—*i.e.*, the reconstruction time per particle—can be lower than a μs able to handle 10^4 particles in about 1.4 *ms* in the inference phase. Notice that the corresponding memory size in the inference phase is approximately equal to the value reported in the table.

Jefferson Lab
Exploring the Nature of Matter

DIRC alignment



Bayesian Optimization

Recipe: For each call of the optimizer, M offset points are explored using N different particles (for each call). The total number of calls is T

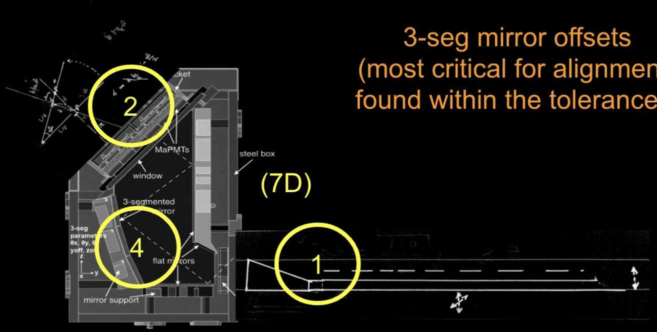
T=120 M=10 N=125

Particles used = 15000

Points explored = 1200

use high purity samples of π for alignment

FoM = LogL normalized to a default alignment



3-seg mirror offsets (most critical for alignment) found within the tolerances.

(7D)

correct

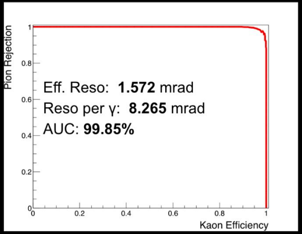
3-seg mirror:
θx,θy,θz=(0.25,0.50,0.15) deg,
y = 0.5 mm;
bar z = 2.0 mm;
PMT (r,θ)=(1.5 mm,1.0 deg)

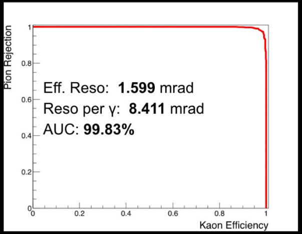
calibrated

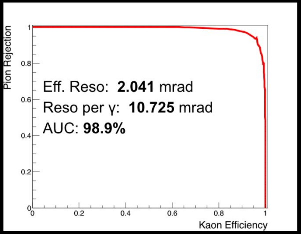
3-seg mirror:
θx,θy,θz=(0.2485, 0.5832, 0.1171) deg,
y = 0.5894 mm;
bar z = 2.0788 mm;
PMT (r,θ)=(1.8690 mm, 1.3544 deg)

nominal

3-seg mirror:
θx,θy,θz=(0., 0., 0.) deg,
y = 0. mm;
bar z = 0. mm;
PMT (r,θ)=(0. mm, 0. deg)







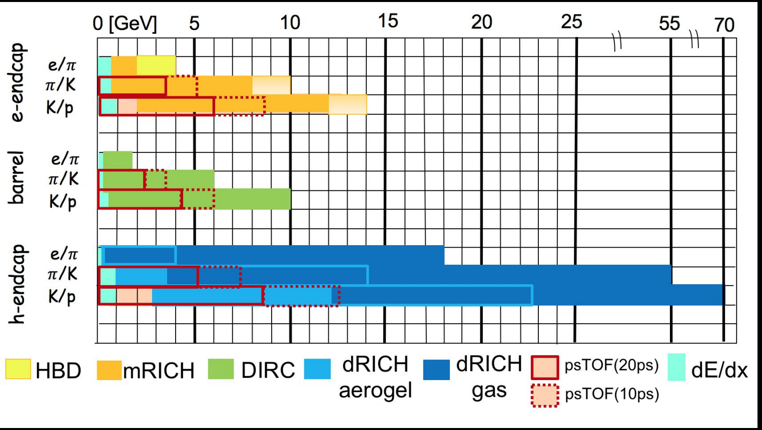
Jefferson Lab

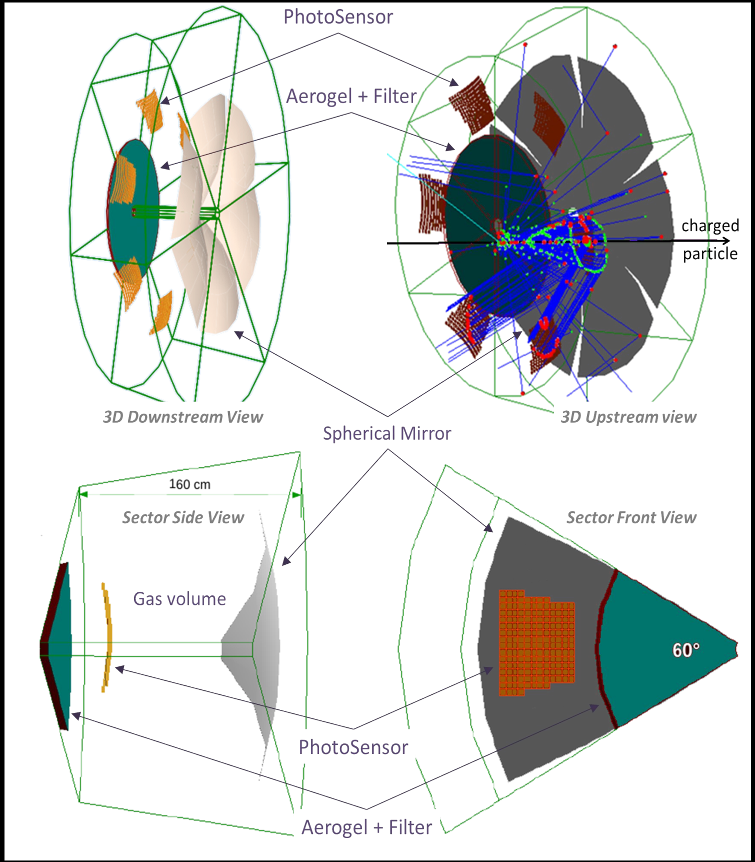
C. Fanelli 2020, Machine learning for imaging Cherenkov detectors, JINST 15 C02012

dRICH: ante-proposal

E. Cisbani, A. Del Dotto, CF*, M. Williams et al.
"AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case."
Journal of Instrumentation 15.05 (2020): P05009.

aerogel (4 cm, n(400 nm): 1.02)
+ 3 mm acrylic filter
+ gas (1.6 m, n(C_2F_6): 1.0008)





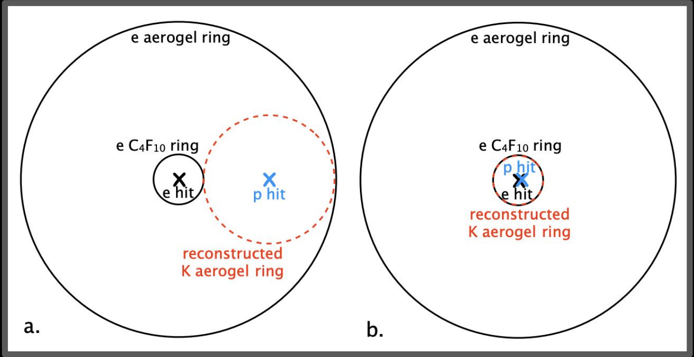
- Continuous momentum coverage.
- Simple geometry and optics, cost effective.
- Legacy design from INFN, see EICUG2017

- 6 Identical open sectors (petals)
- Optical sensor elements: 8500 cm^2/sector, 3 mm pixel
- Large focusing mirror

dRICH reconstruction

* analogy with different approaches discussed for DIRC

- Indirect Ray Tracing (IRT)
  - The basic idea is that, given tracking information and RICH PMT hits, the Cherenkov-photon emission angle can be reconstructed.
  - The distribution of observed photon angles is compared to the expected angle for each particle type and the most likely particle type is determined.
  - Fast, non computationally intensive. Lowest accuracy compared to other methods in this slide.
- Direct Ray Tracing (DRT)
  - Simulates a PMT hit pattern based on the track kinematics and particle hypothesis
  - Construct likelihood by comparing "PDF" to the observed hit pattern
- Event-level algorithm (EVT)
  - Motivation: two close tracks can produce misidentification
  - Builds upon DRT. Improvement by looking at each event as a whole rather than individual tracks
    - → sum over all tracks in the event



R. M. Lamb, PhD thesis, 2010, The Boer-Mulders and Cahn effects: Azimuthal modulations in the spin-independent SIDIS cross section at HERMES, https://inspirehep.net/literature/872161

ML/DL for dRICH reco?

- Just moved first steps in this direction…
  - https://eic.ai/hackathons (10/14/2022, W&M)

hackathon supported by AWS —
4 GPUs / instance on cloud computing;
1 instance / team;
10 teams total —
prize supported by W&M

powered by aws

- We started from particle level
- We will move to event level reconstruction

10 teams, 30+ people, both in person and virtual; participation from America, Europe, Asia

- Documentation (problem description and dataset): https://doi.org/10.5281/zenodo.7197023
- Solutions accepted only above a certain threshold for the score
  - Hackathon winning team declared by 5pm ET of 10/14/2022
    - Team JINR!
  - Solutions accepted for an additional week:
    - Best solution overall by Team Jets

https://ai4eichackathon.pythonanywhere.com/leaderboard

AI4EIC Hackathon

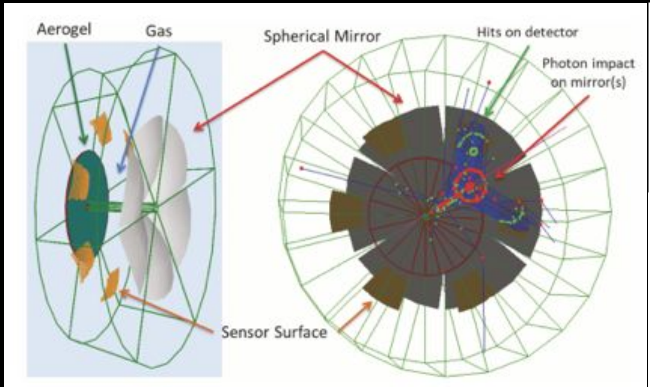
Congrats Team JINR!!!!!!!! (submission on 10-14-2022)

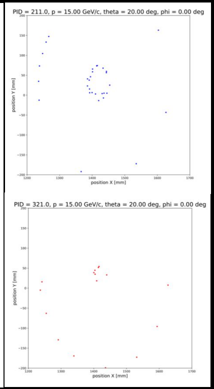
Hackathon Leaderboard

RANK	TEAM	SCORE	QUESTIONS ATTEMPED
1	Jets	295.502	Q 1, Q 3, Q 2
2	JINR	294.508	Q 1, Q 3, Q 2
3	JB and EC	262.313	Q 1, Q 3, Q 2

Jefferson Lab
Exploring the Nature of Matter

AI4EIC Hackathon





Organizers: Cris Fanelli (William & Mary/JLab), Diana McSpadden (JLab/Data Science), Kishan Rajput (JLab/Data Science)

Advisory and problem definition: Evaristo Cisbani (INFN), Wouter Deconinck (U. Manitoba)

Computing resources: Eric Walter (William & Mary, IT)

Data generation, Documentation, Validation: James Giroux (U. Regina), Karthik Suresh (U. Regina)

Technical Assistance: Eric Walter (William & Mary, IT), James Giroux (U. Regina), Karthik Suresh (U. Regina)

CF, DMcS,KR, EC,WD,EW,KS,JG

Problem Number	Threshold Accuracy
Problem 1	94%
Problem 2	86%
Problem 3	80%

π, K datasets

Problem 1

Training Events	1.5 Million Events	With Magnetic Field ($\sim$ 1.5T)
Momentum	15 GeV/c	at Interaction Point (0, 0, 0)
Theta θ	$20°$	at Interaction Point (0, 0, 0)
Phi ϕ	$0°$	at Interaction Point (0, 0, 0)

Problem 2,3*

Training Events	3 Million Events	With Magnetic Field ($\sim$ 1.5T)
Momentum	$15 - 20 GeV/c$	at Interaction Point (0, 0, 0)
Theta θ	$15 - 16°$	at Interaction Point (0, 0, 0)
Phi ϕ	$0 - 5°$	at Interaction Point (0, 0, 0)

*Problem 3: addition of noisy hits

Solutions:

- JINR: CatBoost (Yandex), https://catboost.ai/
- Jets: 2D CNN

The best solutions were all Machine Learning/Deep Learning-based, they were quite original, and they outperformed solutions based on classical approaches (followed by some teams). While this is only a first step towards deeply learning the identification of particles reconstructed with the dual-RICH, these exploratory studies clearly indicates the potential of ML/DL approaches for reconstruction and PID.

Jefferson Lab
Exploring the Nature of Matter

https://eic.ai/hackathons

Outlook

- EIC shifting towards streaming data, near real time analysis and automated alignment/calibration (see streaming readout [1,2])
- Existing ML/DL solutions for imaging Cherenkov reconstruction looks promising
    - Allows for faster reconstruction and high accuracy — streaming
    - ML/DL suitable for reconstruction at the event level (not only at the particle level) combining multiple tracks
    - Bonus: fast simulation…
- AI/ML approaches provide solutions for automated tasks (e.g., calibration/alignment)
    - Ongoing work to extend current algorithms, e.g., to accommodate multiple classes and cover larger phase-space; deploy this on ePIC data

[1] Streaming Readout Workshop X, 2022
[2] AI4EIC Workshop, 2022, Streaming Readout Session

Jefferson Lab
Exploring the Nature of Matter