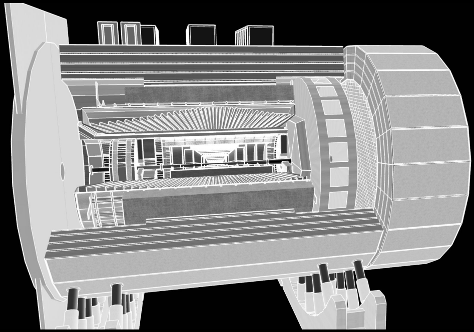
Machine Learning for High Energy (Nuclear) Physics

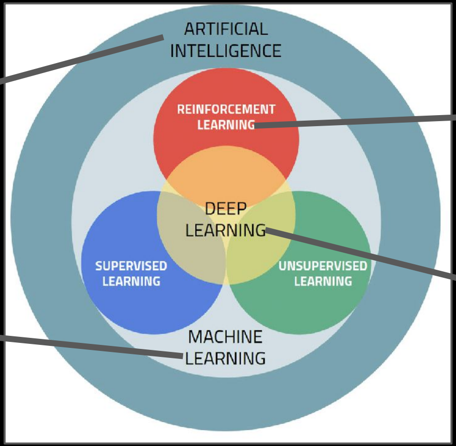


Cristiano Fanelli

ECFA-NuPECC-APPEC - Synergies between the EIC and the LHC, Dec 14-15, 2023, DESY

Taxonomy

AI: The field of computer science that focuses on creating machines or software capable of intelligent behavior, emulating human cognitive functions such as learning, reasoning, problem-solving, and perception.



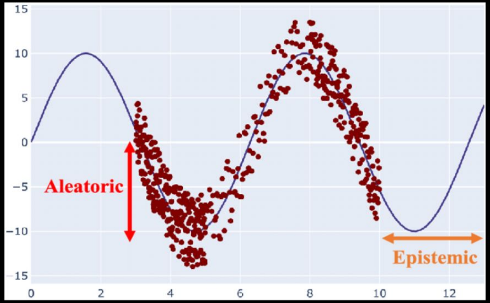
RL: Learning through trial and error, optimizing actions based on rewards

ML: A subset of AI that enables computers to learn from data without explicit programming

DL: A subset of ML that focuses on artificial neural networks with many layers

Uncertainty

- Epistemic Uncertainty: This type of uncertainty arises from a lack of knowledge which is reflected in the effectiveness of the model in describing the data. It can be reduced as more information or data becomes available, and by improving the model. It can be affected by inaccuracy.
- Aleatoric Uncertainty: This uncertainty is due to inherent variability or randomness in a process or system and cannot be reduced by collecting more data. For example, even if we know the probability of getting heads when flipping a fair coin, the outcome of each individual flip is still uncertain.

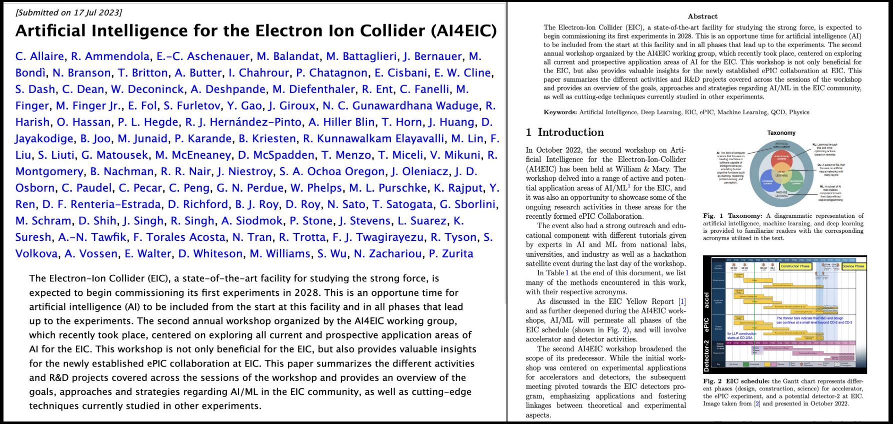


Abdar, Moloud, et al. "A review of uncertainty quantification in deep learning: Techniques, applications and challenges." Information fusion 76 (2021): 243-297.

Main References for AI/ML in HE(N)P

Several workshops identified the scientific challenges and opportunities at the intersection between AI and high energy nuclear physics research

https://eic.ai Arxiv:2307.08593 (under review on Comp. Softw. Big Sci.)



[Submitted on 17 Jul 2023]

Artificial Intelligence for the Electron Ion Collider (AI4EIC)

C. Allaire, R. Ammendola, E.-C. Aschenauer, M. Balandat, M. Battaglieri, J. Bernauer, M. Bondì, N. Branson, T. Britton, A. Butter, I. Chahrour, P. Chatagnon, E. Cisbani, E. W. Cline, S. Dash, C. Dean, W. Deconinck, A. Deshpande, M. Diefenthaler, R. Ent, C. Fanelli, M. Finger, M. Finger Jr., E. Fol, S. Furletov, Y. Gao, J. Giroux, N. C. Gunawardhana Waduge, R. Harish, O. Hassan, P. L. Hegde, R. J. Hernández-Pinto, A. Hiller Blin, T. Horn, J. Huang, D. Jayakodige, B. Joo, M. Junaid, P. Karande, B. Kriesten, R. Kunnawalkam Elayavalli, M. Lin, F. Liu, S. Liuti, G. Matousek, M. McEneaney, D. McSpadden, T. Menzo, T. Miceli, V. Mikuni, R. Montgomery, B. Nachman, R. R. Nair, J. Niestroy, S. A. Ochoa Oregon, J. Oleniacz, J. D. Osborn, C. Paudel, C. Pecar, C. Peng, G. N. Perdue, W. Phelps, M. L. Purschke, K. Rajput, Y. Ren, D. F. Renteria-Estrada, D. Richford, B. J. Roy, D. Roy, N. Sato, T. Satogata, G. Sborlini, M. Schram, D. Shih, J. Singh, R. Singh, A. Siodmok, P. Stone, J. Stevens, L. Suarez, K. Suresh, A.-N. Tawfik, F. Torales Acosta, N. Tran, R. Trotta, F. J. Twagirayezu, R. Tyson, S. Volkova, A. Vossen, E. Walter, D. Whiteson, M. Williams, S. Wu, N. Zachariou, P. Zurita

The Electron-Ion Collider (EIC), a state-of-the-art facility for studying the strong force, is expected to begin commissioning its first experiments in 2028. This is an opportune time for artificial intelligence (AI) to be included from the start at this facility and in all phases that lead up to the experiments. The second annual workshop organized by the AI4EIC working group, which recently took place, centered on exploring all current and prospective application areas of AI for the EIC. This workshop is not only beneficial for the EIC, but also provides valuable insights for the newly established ePIC collaboration at EIC. This paper summarizes the different activities and R&D projects covered across the sessions of the workshop and provides an overview of the goals, approaches and strategies regarding AI/ML in the EIC community, as well as cutting-edge techniques currently studied in other experiments.

Abstract

The Electron-Ion Collider (EIC), a state-of-the-art facility for studying the strong force, is expected to begin commissioning its first experiments in 2028. This is an opportune time for artificial intelligence (AI) to be included from the start at this facility and in all phases that lead up to the experiments. The second annual workshop organized by the AI4EIC working group, which recently took place, centered on exploring all current and prospective application areas of AI for the EIC. This workshop is not only beneficial for the EIC, but also provides valuable insights for the newly established ePIC collaboration at EIC. This paper summarizes the different activities and R&D projects covered across the sessions of the workshop and provides an overview of the goals, approaches and strategies regarding AI/ML in the EIC community, as well as cutting-edge techniques currently studied in other experiments.

Keywords: Artificial Intelligence, Deep Learning, EIC, ePIC, Machine Learning, QCD, Physics

1 Introduction

In October 2022, the second workshop on Artificial Intelligence for the Electron-Ion-Collider (AI4EIC) has been held at William & Mary. The workshop delved into a range of active and potential application areas of AI/ML[1] for the EIC, and it was also an opportunity to showcase some of the ongoing research activities in these areas for the recently formed ePIC Collaboration.

The event also had a strong outreach and educational component with different tutorials given by experts in AI and ML from national labs, universities, and industry as well as a hackathon satellite event during the last day of the workshop.

In Table 1 at the end of this document, we list many of the methods encountered in this work, with their respective acronyms.

As discussed in the EIC Yellow Report [1] and as further deepened during the AI4EIC workshops, AI/ML will permeate all phases of the EIC schedule (shown in Fig. 2), and will involve accelerator and detector activities.

The second AI4EIC workshop broadened the scope of its predecessor. While the initial workshop was centered on experimental applications for accelerators and detectors, the subsequent meeting pivoted towards the EIC detectors program, emphasizing applications and fostering linkages between theoretical and experimental aspects.

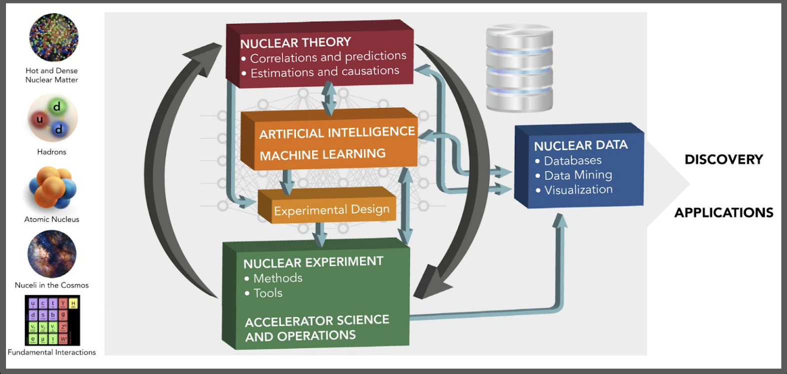
Taxonomy

Fig. 1 Taxonomy: A diagrammatic representation of artificial intelligence, machine learning, and deep learning is provided to familiarize readers with the corresponding acronyms utilized in the text.

Fig. 2 EIC schedule: the Gantt chart represents different phases (design, construction, science) for accelerator, the ePIC experiment, and a potential detector-2 at EIC. Image taken from [2] and presented in October 2022.

https://doi.org/10.1103/RevModPhys.94.031003

A. Boehnlein, M. Diefenthaler, C. Fanelli et al., Machine learning in nuclear physics, Rev. Mod. Phys. **94**, 031003 (2022) and references therein



3rd AI4EIC workshop at CUA, Washington D.C.

AI/ML for ePIC and Beyond (Nov 28, morning)

- Derek Anderson (Iowa State University), Anselm Vossen (Duke University)

Calibration, Monitoring, and Experimental Control in Streaming Environments (Nov 28, afternoon)

- Yeonju Go (Brookhaven National Lab), Torri Jeske (Jefferson Lab)

AI/ML for Accelerators (Nov 29, morning)

- Kevin Brown (Brookhaven National Lab), Elena For (CERN)

AI/ML for Data Analysis and Theory (Nov 29, afternoon)

- Brandon Kriesten (Argonne National Laboratory), Vinicius Mikuni (National Energy Research Scientific Computing Center)

Foundation Models and Trends in Data Science (Nov 30, morning)

- Yaohang Li (Old Dominion University), Daniel Murnane (Lawrence Berkeley National Laboratory)

AI/ML in Production, Distributed ML (Nov 30, afternoon)

- David Lawrence (Jefferson Lab), Rui Zhang (UW Madison)

https://iml-wg.github.io/HEPML-LivingReview/

HEPML-LivingReview

A Living Review of Machine Learning for Particle Physics

Modern machine learning techniques, including deep learning, is rapidly being applied, adapted, and developed for high energy physics. The goal of this document is to provide a nearly comprehensive list of citations for those developing and applying these approaches to experimental, phenomenological, or theoretical analyses. As a living document, it will be updated as often as possible to incorporate the latest developments. A list of proper (unchanging) reviews can be found within. Papers are grouped into a small set of topics to be as useful as possible. Suggestions are most welcome.

download review

https://indico.bnl.gov/event/19560/ Key discussion points: need for more benchmarks, uncertainty quantification

Disclaimer

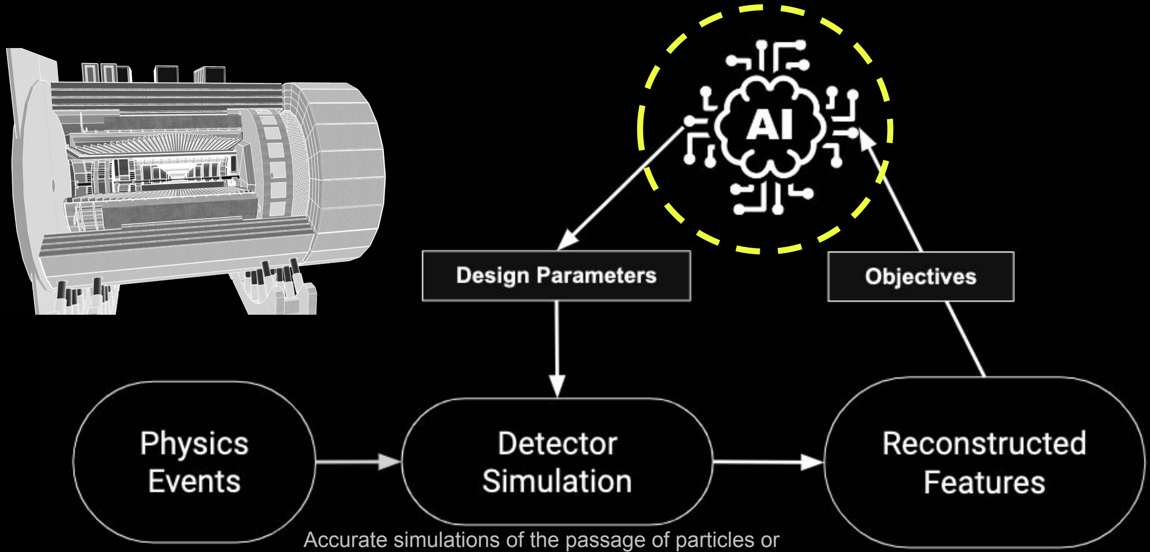
- ML/DL is ubiquitous in HEP and NP
- Hard to impossible to summarize multiple topics, and recent works, activities (e.g., recent AI4EIC workshop showed an impressive progress in the last year) and opportunities.
- Drawing from insights gained at the recent AI4EIC, this talk is approached from an EIC perspective, focusing on the potential synergies between the LHC and EIC. I will also make examples from LHC and present studies that use benchmark datasets from the LHC community (e.g., LHC Olympics).
- At AI4EIC, two cross-cutting needs have been underscored:
(i) Establishing Benchmarks and (ii) Uncertainty Quantification

The following examples have been chosen to emphasize and explore these key topics and opportunities:

- Adaptive Experimentation / AI-assisted Optimizations
- “Holistic” analysis — full event information, and real-data
- Uncertainty Quantification (event-level) and Unfolding with Uncertainty
- Towards near real-time applications (supported by Streaming Readout)

AI-Assisted Detector Design

Hot take: every optimization problem is fundamentally a multi-objective optimization problem.



Accurate simulations of the passage of particles or radiation through matter

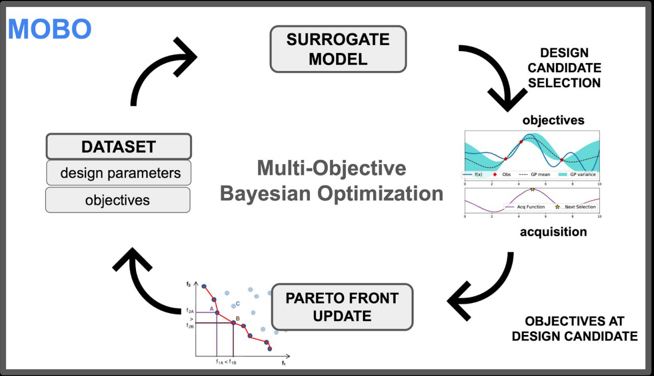
- Benefits from rapid turnaround time from simulations to analysis of high-level reconstructed observables
- The EIC SW stack offers multiple features that facilitate AI-assisted design (e.g., modularity of simulation, reconstruction, analysis, easy access to design parameters, automated checks, etc.)
- Leverages heterogeneous computing

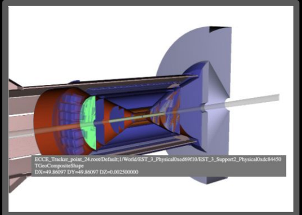
Provide a framework for an holistic optimization of the sub-detector system
A complex problem with (i) multiple design parameters, driven by (ii) multiple objectives (e.g., detector response, physics-driven, costs) subject to (iii) constraints

Those at EIC can be the first large-scale experiments ever realized with the assistance of AI

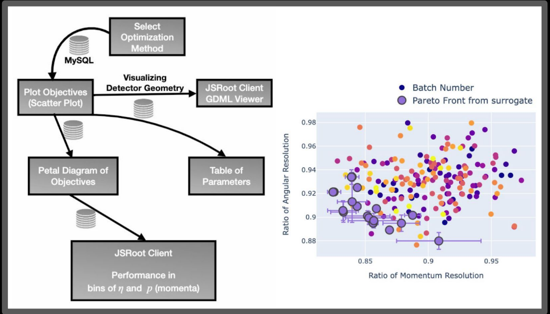
AI-Assisted Detector Design (at EIC)

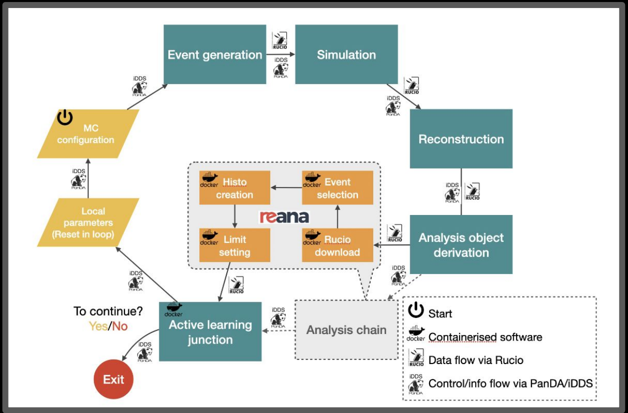
BNL, CUA, Duke, JLab, W&M*





https://ai4eicdetopt.pythonanywhere.com/





(i) Will contribute to advance state of the art MOBO complexity to accommodate a large number of objectives and will explore usage of physics-inspired approaches

(ii) Development of suite of data science tools for interactive navigation of Pareto front (multi-dim design with multiple objectives)

(iii) Will leverage cutting-edge workload management systems capable of operating at massive data and handle complex workflows

Examining solutions on the Pareto front of EIC detectors at different values of the budget can have great cost benefits

A fractional improvement in the objectives translates to a more efficient use of beam time which will make up a majority of the cost of the EIC over its lifetime

Reconstruction/Identification

Rev. Mod. Phys. 94, 031003 (2022)

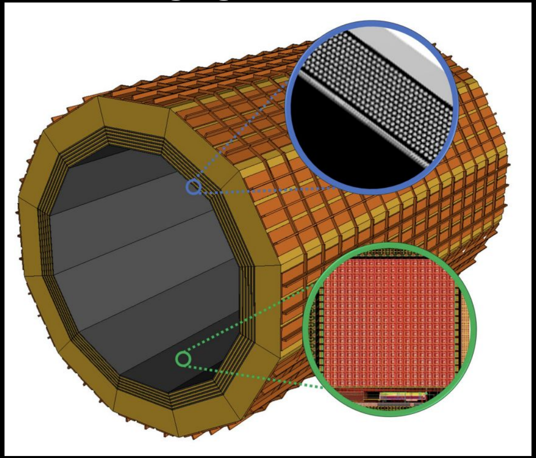
Tracking	Calorimeter
PID-detector	Jet Reco

See backup slides

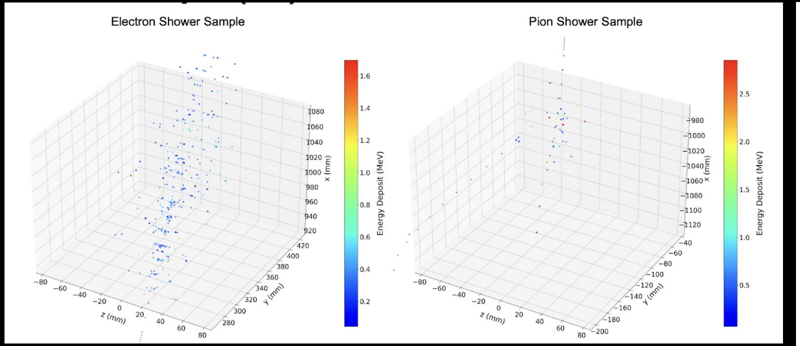
[1] G. Gavalian, et al. "Using Artificial Intelligence for Particle Track Identification in CLAS12 Detector." arXiv preprint arXiv:2008.12860 (2020).
[2] G. Gavalian. "Auto-encoders for Track Reconstruction in Drift Chambers for CLAS12." arXiv preprint arXiv:2009.05144(2020).
[3] L.-G. Gagnon, Machine learning for track reconstruction at the LHC, 2022 JINST 17 C02026 — AI4EIC workshop
[4] Exa.TrkX: HEP tracking at the exascale. A DOE CompHEP project, https://exatrkx.github.io/
[5] A. Akram, and X. Ju. "Track Reconstruction using Geometric Deep Learning in the Straw Tube Tracker (STT) at the PANDA Experiment." arXiv:2208.12178 (2022)
[6] D. Rohr "Overview of online and offline reconstruction in ALICE for LHC Run 3." arXiv:2009.07515 (2020) https://arxiv.org/abs/2009.07515

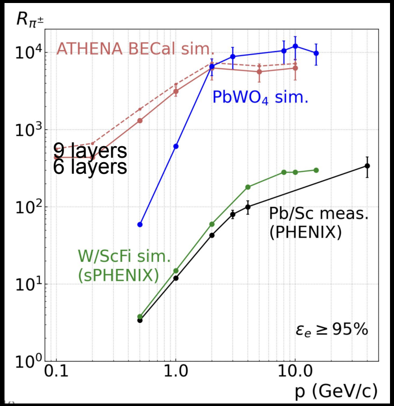
Shower Imaging

Imaging Calorimeter



shower examples





ML model: Sequential CNN + MLP

red: imaging detector and ML
blue, green and the black: other technology and traditional cut-based strategy

Hybrid Concept

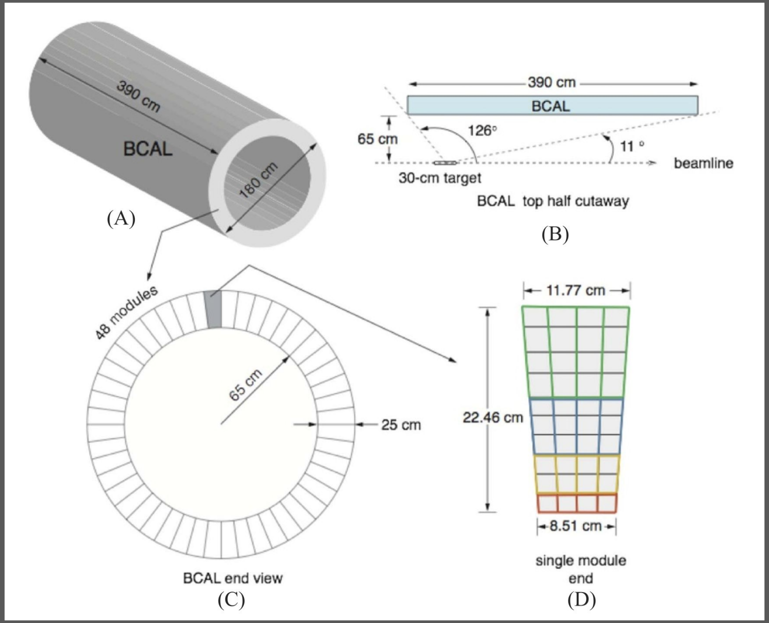
- Monolithic Silicon Sensors AstroPix
- Scintillating fibers embedded in Pb (Pb/ScFi similar to GlueX Barrel Ecal)
- “Sandwiched” 6 layers of AstroPix and 5 layers of Pb/ScFi (~1X0) followed by a large chunk of Pb/ScFi
- Total thickness ~43 cm (~21 X0)
- Large amount of data (3D shower imaging)

- ML with shower imaging significantly improves e/π rejection compared to traditional E/p cut — impact on DIS
- Separation of γ's from π^0 at high momenta (40 GeV/c) and precise position reconstruction of γ's (<1 mm at 5 GeV) — DVCS and γ physics
- Tagging final state radiative γ's from nuclear/nucleon elastic scattering at low x to benchmark QED internal corrections
- PID of low energy μ that curl in the barrel ECal — J/ψ reconstruction and TCS
- Improving PID, providing a space coordinate for DIRC reconstruction

[1] N. Apadula, et al. "Monolithic active pixel sensors on cmos technologies." arXiv preprint arXiv:2203.07626 (2022).
[2] C. Peng, ML Particle Identification with Measured Shower Profiles from Calorimetry, AI4EIC 2nd workshop (2022)

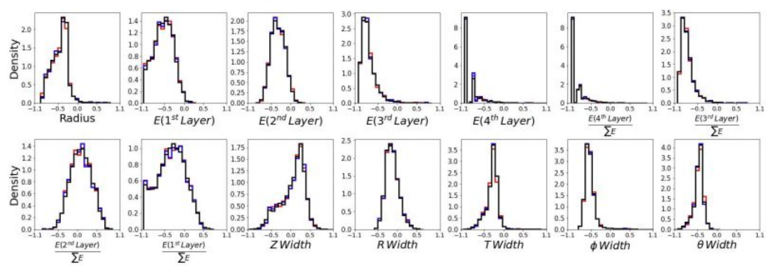
n/γ GlueX/BCAL + Large Language Model

BCAL is constructed as a lead and scintillating-fiber calorimeter and read out with 3840 large-area silicon photomultiplier arrays



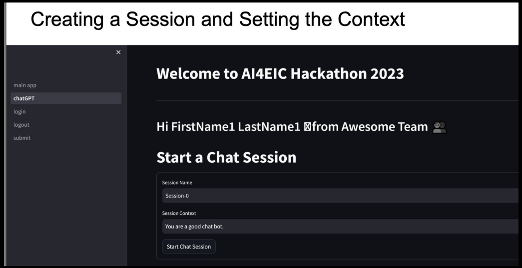
(A) BCAL schematic; (B) a BCAL module side view; (C) end view of the BCAL showing all 48 modules and (D) an end view of a single module showing readout segmentation in four rings (inner to outer) and 16 summed readout zones demarcated by colors

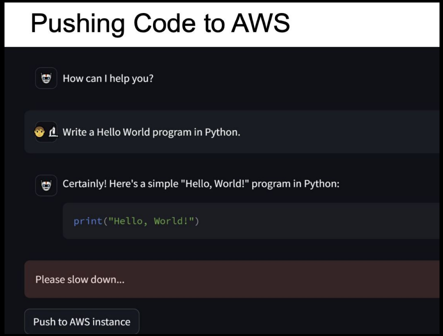
Shower Features



- Large Language Model prompt-engineering: write the whole code with ChatGPT

web-interface





- Participants had (basically) no access to the files, nor to the code editor
- ChatGPT-based solutions significantly outperformed (human-based) expected solutions! Beginner-level were able to submit excellent solutions

[1] J. Giroux, P. Moran, K. Suresh, C. Fanelli: Tutorial on 2nd AI4EIC hackathon, 2023 [indico link]

n/γ GlueX/BCAL + Large Language Model

Deploy

main app
chatGPT
login
logout
submit

AI4EIC Hackathon 2023

Welcome to the AI4EIC Hackathon 2023. Navigate to the different pages of the hackathon.

	Team#	Team Name	Username	Name of User	Q1 Score	Q2 Score	↓ Total Score
6	Team2	Jets3	hcharles	Charles Hughes	99.935	99.777	199.712
5	Team2	Jets3	aderek	Derek Anderson	99.935	99.777	199.712
4	Team2	Jets3	kdmitrii	Dmitrii Kalinkin	99.935	99.777	199.712
10	Team3	SPIN 2023 Local Organization Committee	ssimon	Simon Schneider	99.9325	99.7735	199.706
9	Team3	SPIN 2023 Local Organization Committee	mmatthew	Matthew McEneany	99.9325	99.7735	199.706
8	Team3	SPIN 2023 Local Organization Committee	mgregory	Gregory Matousek	99.9325	99.7735	199.706
7	Team3	SPIN 2023 Local Organization Committee	pconnor	Connor Pecar	99.9325	99.7735	199.706
13	Team4	Small Language Models	smanuel	Manuel Szewc	99.9285	99.77	199.6985
12	Team4	Small Language Models	tfernando	Fernando Torales	99.9285	99.77	199.6985
11	Team4	Small Language Models	mvinicius	Vinicius Mikuni	99.9285	99.77	199.6985
16	Team5	404 Brain Not Found	salex	Alex Smith	99.919	99.7565	199.6755
15	Team5	404 Brain Not Found	gsimon	Simon Gardner	99.919	99.7565	199.6755
14	Team5	404 Brain Not Found	proberto	Roberto Preghenella	99.919	99.7565	199.6755
3	Team1	Messed Ups	himran	Md. Imran Hossain	99.918	99.614	199.532
2	Team1	Messed Ups	savinsh	Avnish Singh	99.918	99.614	199.532
1	Team1	Messed Ups	sbhavya	Bhavya Singhal	99.918	99.614	199.532

⋮

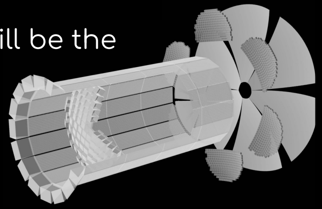
What happens when we combine physicists to LLM “assistance”:

- Great results: high performance (well beyond expectations, particularly for problem 2 with increased complexity)
- Several winners within statistical uncertainties using different approaches
- Absolute winner selected by minimum number of prompts and time of submission
- What’s next: analysis of these data/solutions

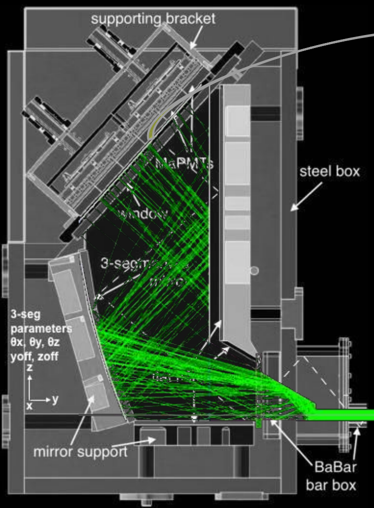
[1] J. Giroux, P. Moran, K. Suresh, C. Fanelli: Tutorial on 2nd AI4EIC hackathon, 2023 [indico link]

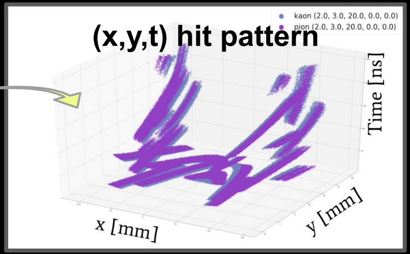
PID with Cherenkov

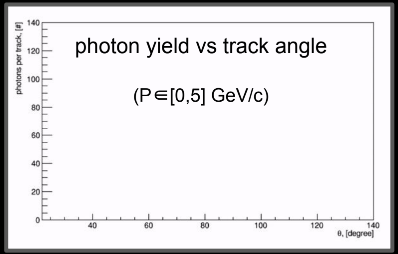
Cherenkov detectors will be the backbone of PID at EIC



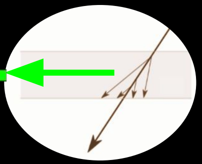
DIRC at GlueX is instrumental for PID







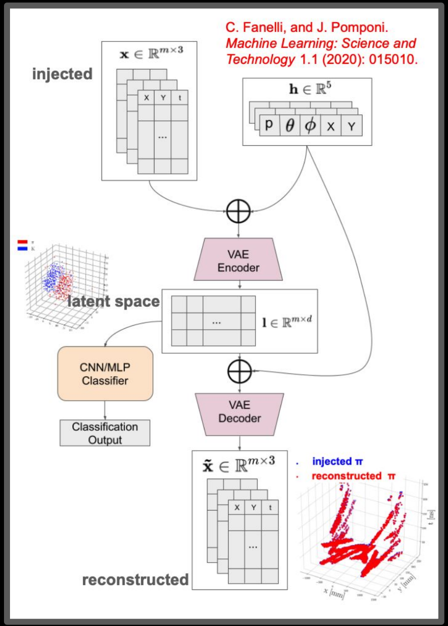
charged track



Cherenkov photons

- Need for faster and accurate simulations and reconstruction
- Complex hit patterns (DIRC is the most complex), sparse data, response vs kinematics
- DeepRICH: same reconstruction performance of best reconstruction algorithm with ~4 orders of magnitude speed-up in inference time on GPU
- Possibility to learn at the event-level (full event) rather than at the track/particle level, and using real data. E.g., two tracks with overlapping patterns in the optical box
- Bonus: fast simulation from generative models

DeepRICH



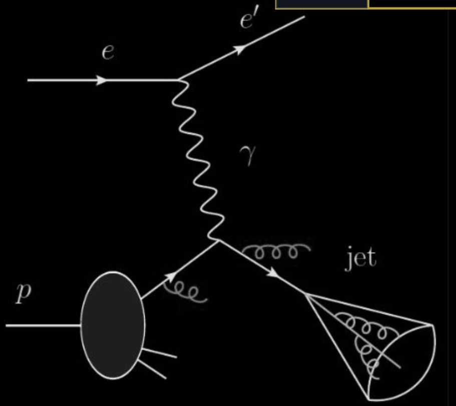
[1] C. Fanelli, J. Pomponi, “DeepRICH: learning deeply Cherenkov detectors”, Mach. Learn.: Sci. Technol., 1.1 (2020): 015010
[2] C. Fanelli, "Machine learning for imaging Cherenkov detectors." JINST 15.02 (2020): C02012.

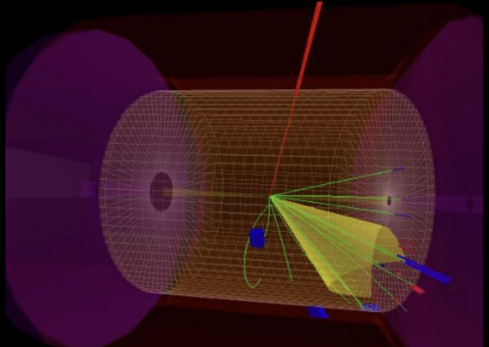
Jet Reconstruction

ML4Jets
https://indico.cern.ch/event/1159913/

A highly dynamic research area, with an abundance of ML/DL-based activities

- Jet classification in HEP has improved with ML/DL, especially DNN-based models in Run 2 at 13 TeV LHC [1,2]. Typical jet momenta range is p_T ~100-1000 GeV.
- The EIC, with up to 140 GeV CoM energy, will feature jets in its science program, with p_T ~10-30 GeV. EIC jets have lower energy and increased sparsity compared to LHC.
- In EIC context, ML/DL already “helps” determine DIS kinematic variables (see next) and extract quantum correlation functions. The key question is the improvement ML-based algorithms can offer. Due to non-perturbative modeling, biases may arise in simulated data for ML training.
- ML leverages full event information and can train on experimental data for jet flavor classification and hard process determination. This has been proposed for spin physics and nuclear matter studies [3].

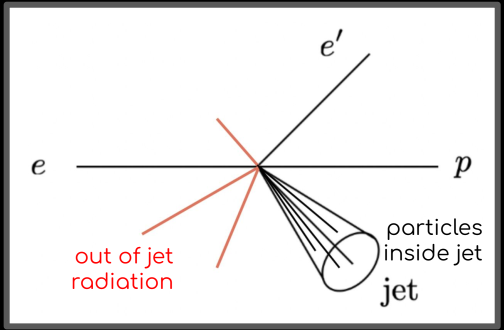




Images taken from [5]

[1] A.M. Sirunyan et al (CMS) Identification of heavy, energetic, hadronically decaying particles using machine-learning techniques, 2020 JINST 15 P06005
[2] M. Aaboud, et al., Performance of top-quark and *WW*-boson tagging with ATLAS in Run 2 of the LHC, EPJC 79, 375 (2019)
[3] K. Lee, "Machine learning-based jet and event classification at the EIC with applications to hadron structure and spin physics." JHEP 2023.3 (2023): 1-35
[4] M. Arratia, et al. "Charm jets as a probe for strangeness at the future EIC" Phys. Rev. D 103.7 (2021): 074023.; [5] M. Arratia, Jets at EIC, talk

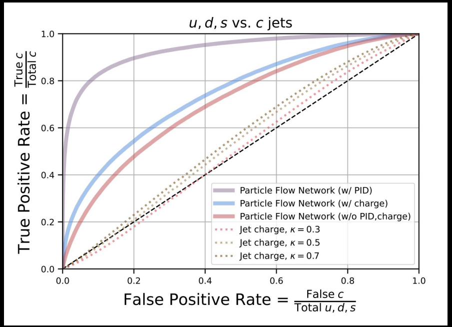
Jet Reconstruction



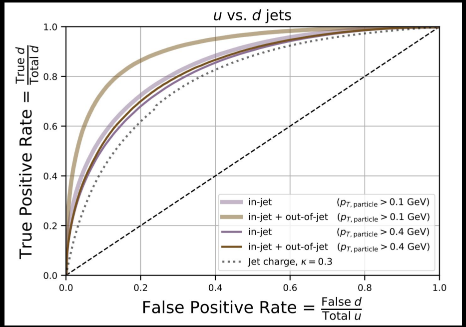
IR jet flavor definition relevant for non-perturbative QCD effects

Use real data for training

ML performance vs energy weighted jet charge



Include out-of-jet



Potential impact:

- Strengthening constraints on transverse momentum dependent PDFs
  - E.g., charm-tagged jets can increase sensitivity to collinear strange quark PDF in charged current events
  - Di-jet with charm and anti-charm can constrain the gluon TMD
  - New opportunities for gluon helicity distributions, parton-in-photon PDF
- Enhancing sensitivity to transverse single spin asymmetries (TSSA) (incoming protons have different transverse spin)

$$A_{UT} = \frac{d\sigma^{\uparrow} - d\sigma^{\downarrow}}{d\sigma^{\uparrow} + d\sigma^{\downarrow}}$$

- Elucidating cold nuclear matter effects

[1] K. Lee et al, "ML-based jet and event classification at the EIC with applications to hadron structure and spin physics." JHEP 2023.3 (2023): 1-35

Theory/Experiment Connections

Fast Sim, Data Driven Rapidly connect particle-level predictions to detector-level observables	**“Parameters” inference** Extractions preventing overfitting, parameters from entire dataset, etc…
Event-level Signatures Reco/identification of physics channels	**Unfolding** “Deconvolution” of detector effects

e.g.,

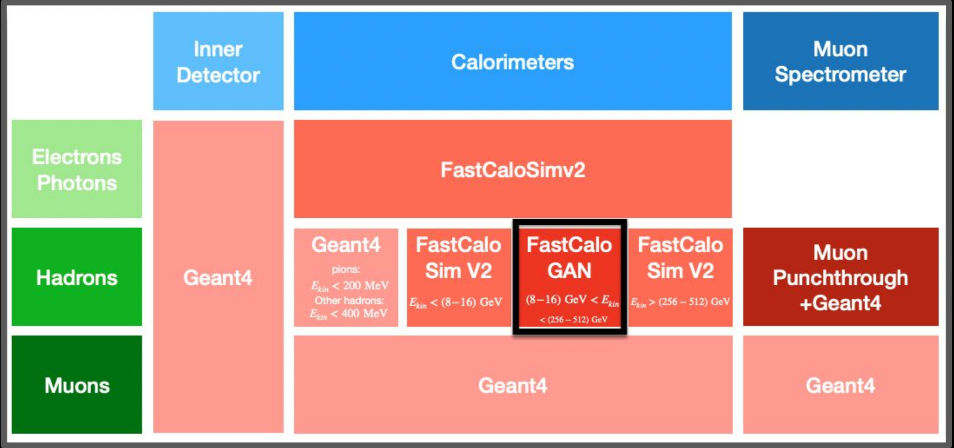
[1] M. LeBlanc, B. Nachman, and C. Sauer. "Going off topics to demix quark and gluon jets in α_S extractions." JHEP 02 (2023) 150

[2] Cao, S., et al. (JETSCAPE) "Determining the jet transport coefficient $\hat{q}$ from inclusive hadron suppression measurements using Bayesian parameter estimation." Physical Review C 104.2 (2021): 024905.

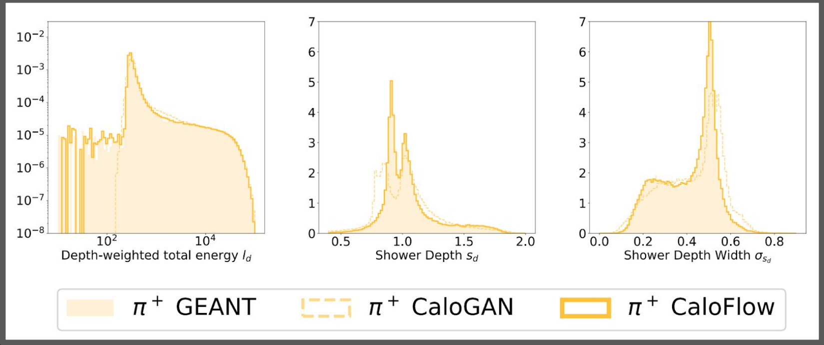
[3] K. Fraser, and M. D. Schwartz. "Jet charge and machine learning." JHEP 2018.10 (2018): 1-18.

Fast Simulations

- The ATLAS Collaboration in AtlFast3 (AF3) utilizes FastCaloGAN
- Parameterized calorimeter simulation - 500x faster than Geant4 in calorimeter



- Generative models have gotten much better, flow models particularly promising



Geant4 vs. Simulations

AUC / JSD		DNN vs. CaloGAN	DNN vs. CaloFlow
e^+	unnormalized	1.000(0) / 0.993(1)	0.847(8) / 0.345(12)
	normalized	1.000(0) / 0.997(0)	0.869(2) / 0.376(4)
γ	unnormalized	1.000(0) / 0.996(1)	0.660(6) / 0.067(4)
	normalized	1.000(0) / 0.994(1)	0.794(4) / 0.213(7)
π^+	unnormalized	1.000(0) / 0.988(1)	0.632(2) / 0.048(1)
	normalized	1.000(0) / 0.997(0)	0.751(4) / 0.148(4)

- AUC = 1 means easily distinguishable, AUC = 0.5 means not distinguishable
- JSD ~ 0 means labels are similarly distributed; JSD ~ 1 largest divergence

[1] AtlFast3: The Next Generation of Fast Simulation in ATLAS. Comput Softw Big Sci 6, 7 (2022)
[2] AtlFast3 ICHEP https://cds.cern.ch/record/2815171/files/ATL-SOFT-SLIDE-2022-253.pdf
[3] C. Kraus and D. Shih, CaloFlow, ArXiv:2106.05285 (2021)

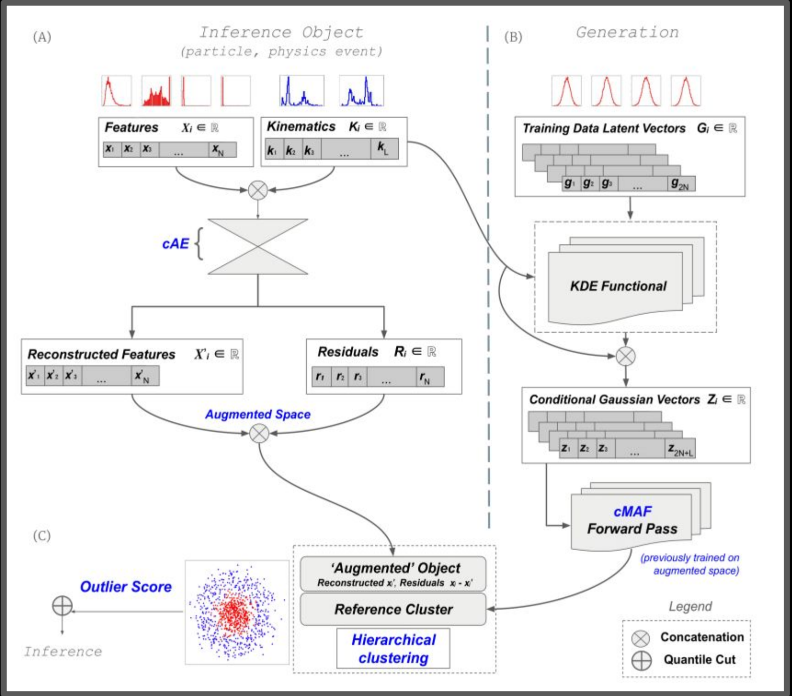
Data-driven Learning

Relies less on simulations,
e.g., one-class classification (OCC) /
anomaly-detection (AD)

Fast, Data-driven

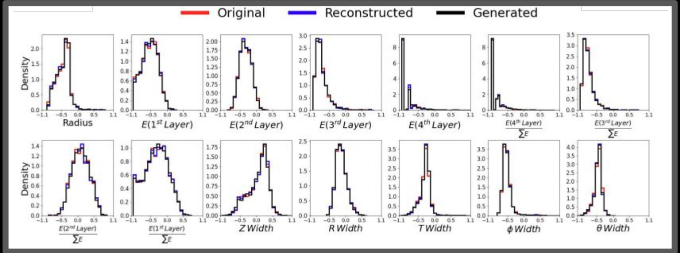
Flux+Mutability

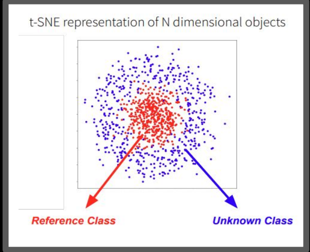
cAE + cMAF + HDBSCAN

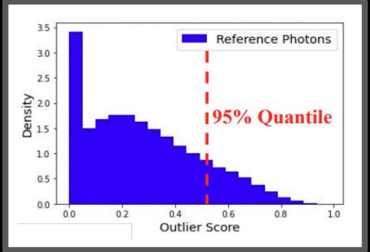


Agnostic to "anomalous" signal, requires one reference sample with high-purity

Generate "reference cluster" (30 dims) conditioned to some kinematics variable (e.g., measured shower energy or jet p_T)







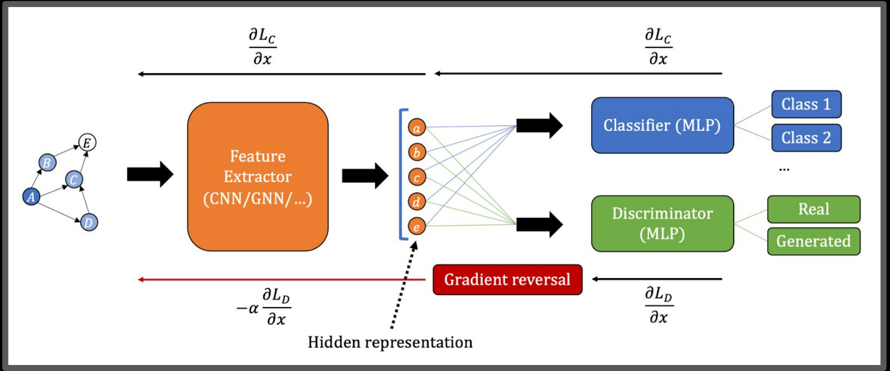
Applied to: (1) γ/n separation in BCAL (OCC);
(2) BSM/SM Di-jet separation for LHC (AD) – outperformed or on par with other methods, but with no assumption on signal

Extensions: Data Quality Control / AD

[1] C. Fanelli, J. Giroux, and Z. Papandreou. "'Flux+ Mutability: a conditional generative approach to one-class classification and anomaly detection. "IOP Mach. learn.: sci. technol. 3.4 (2022): 045012.

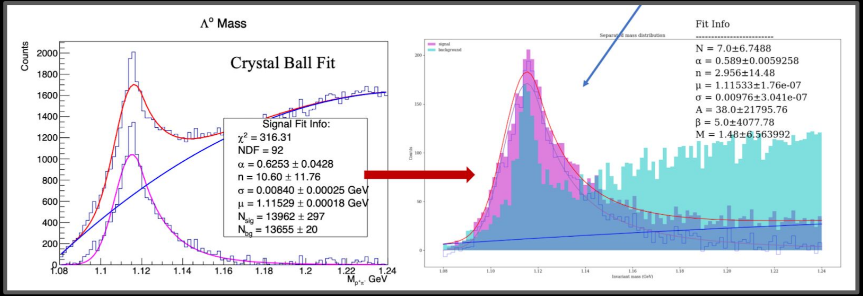
Λ-hyperon tagging in CLAS12

Domain-Adversarial GNN



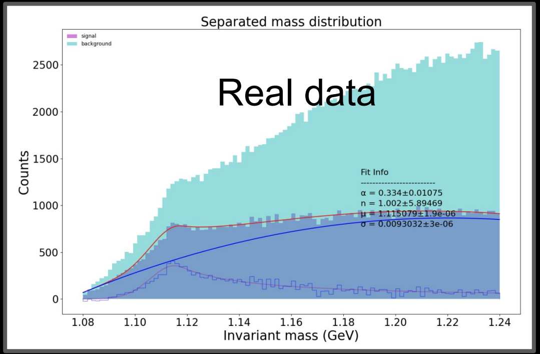
Reverse gradient from discriminator loss during backpropagation

Domain-Adversarial GIN



MC before NN

MC after NN

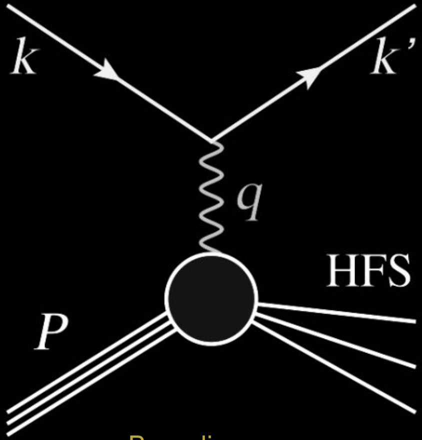


- MC: good accuracy >80%, 0.9 AUC, good signal efficiency with background significantly reduced
- Improved S/B on real data by ~30%. Purity improves by factor 1.8.
- Potential improvement from adding detector data as inputs. Similar studies can be used for EIC

[1] M. McEneaney, A. Vossen, Domain-Adversarial Graph Neural Networks for Λ Hyperon Identification with CLAS12, arXiv:2302.05481v2 (2023)

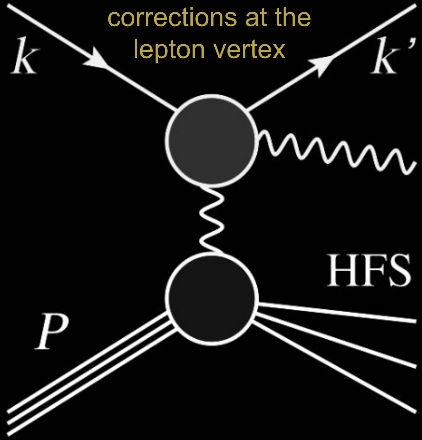
Deep Inelastic Scattering

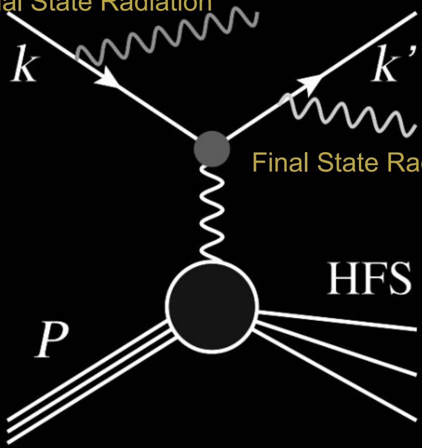
DIS is governed by the four-momentum transfer squared of the exchanged boson Q^2, the inelasticity y, and the Bjorken scaling variable x.



Born diagram

higher-order QED corrections at the lepton vertex





These kinematic variables are related via $Q^2 = s \cdot x\, y$, where s is the square of the center-of-mass energy.

$$s = (k+P)^2 , \quad Q^2 = -q^2 , \quad y = \frac{q \cdot P}{k \cdot P} , \quad \text{and} \quad x = Q^2/(sy) .$$

DIS Kinematics

DIS kinematics: Traditional Methods

Fast Sim, Data Driven Learning | "Parameters" inference | Event-level Signatures | Cross-sections inference

- Conservation of momentum and energy over constrain the DIS kinematics and leads to a freedom to calculate x, Q^2, y from measured quantities

- Each method has advantages and disadvantages, and no single approach is optimal over the entire phase space. Each method exhibits different sensitivity to QED radiative effects

- Once (real) higher-order QED effects are considered, various methods yield different results and the calculated quantities for Q^2, y and x are not representative for the γ/Z + p scattering process at the hadronic vertex.

Summary of basic reconstruction methods

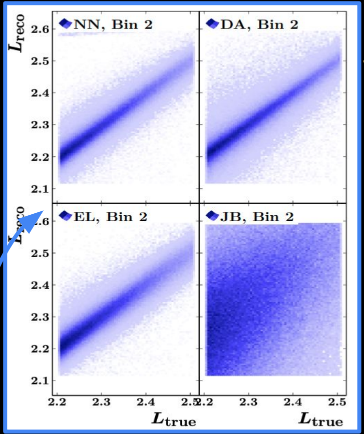
Method name	Observables	y	Q^2	$x \cdot E_p$
Electron (e)	$[E_0, E, \theta]$	$1 - \frac{\Sigma_e}{2E_0}$	$\frac{E^2 \sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
Double angle (DA) [6, 7]	$[E_0, \theta, \gamma]$	$\frac{\tan\frac{\gamma}{2}}{\tan\frac{\gamma}{2}+\tan\frac{\theta}{2}}$	$4E_0^2 \cot^2\frac{\theta}{2}(1-y)$	$\frac{Q^2}{4E_0 y}$
Hadron (h, JB) [4]	$[E_0, \Sigma, \gamma]$	$\frac{\Sigma}{2E_0}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
ISigma (IΣ) [9]	$[E, \theta, \Sigma]$	$\frac{\Sigma}{\Sigma+\Sigma_e}$	$\frac{E^2 \sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
IDA [7]	$[E, \theta, \gamma]$	y_{DA}	$\frac{E^2 \sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
$E_0 E \Sigma$	$[E_0, E, \Sigma]$	y_h	$4E_0E - 4E_0^2(1-y)$	$\frac{Q^2}{2\Sigma}$
$E_0 \theta \Sigma$	$[E_0, \theta, \Sigma]$	y_h	$4E_0^2 \cot^2\frac{\theta}{2}(1-y)$	$\frac{Q^2}{2\Sigma}$
$\theta\Sigma\gamma$ [8]	$[\theta, \Sigma, \gamma]$	y_{DA}	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
Double energy (A4) [7]	$[E_0, E, E_h]$	$\frac{E-E_0}{(xE_p)-E_0}$	$4E_0 y (xE_p)$	$E + E_h - E_0$
$E\Sigma T$	$[E, \Sigma, T]$	$\frac{\Sigma}{\Sigma+E\pm\sqrt{E^2+T^2}}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
$E_0 E T$	$[E_0, E, T]$	$\frac{2E_0 - E \mp\sqrt{E^2-T^2}}{2E_0}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{4E_0 y}$
Sigma (Σ) [9]	$[E_0, E, \Sigma, \theta]$	$y_{\mathrm{I}\Sigma}$	$Q^2_{\mathrm{I}\Sigma}$	$\frac{Q^2}{4E_0 y}$
eSigma ($e\Sigma$) [9]	$[E_0, E, \Sigma, \theta]$	$\frac{2E_0\Sigma}{(\Sigma+\Sigma_e)^2}$	$2E_0E(1+\cos\theta)$	$\frac{E(1+\cos\theta)(\Sigma+\Sigma_e)}{2\Sigma}$

Table 1. Summary of basic reconstruction methods that employ only three out of five quantities: E_0 (electron-beam energy), E and θ (scattered electron energy and polar angle), Σ and γ (longitudinal energy-momentum balance, $\Sigma = \sum_{\mathrm{HFS}}(E_i - p_{z,i})$, and the inclusive angle of the HFS). Alternatively, the A4 method makes use of the HFS total energy E_h. Shorthand notations are used

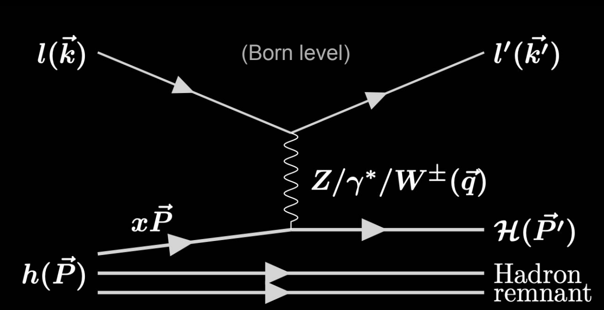
Table taken from Arratia et al., NIM-A 1025 (2022): 166164

Deeply Learning DIS

Example in one specific bin



DIS fundamental process @EIC



DIS beyond the Born approximation has a complicated structure which involve QCD and QED corrections

- Use of DNN to reconstruct the kinematic observable x, Q^2, y in the study of neutral current DIS events at ZEUS and H1 experiments at HERA.
- The performance compared to electron, Jacquet-Blondel and the double-angle methods using data-sets independent of training
- Compared to the classical reconstruction methods, the DNN-based approach enables significant improvements in the resolution of Q^2 and x

Bin	Events	Resolution of $\log x$, $\times 10^3$	Resolution of $\log Q^2/1\,\mathrm{GeV}^2$, $\times 10^3$
1	301780	**NN: 70** EL: 83 JB: 180 DA: 103	**NN: 35** EL: 35 JB: 203 DA: 62
2	350530	**NN: 69** EL: 82 JB: 167 DA: 96	**NN: 40** EL: 43 JB: 192 DA: 64
3	138456	**NN: 98** EL: 130 JB: 138 DA: 100	NN: 55 **EL: 53** JB: 150 DA: 77
4	74844	**NN: 67** EL: 84 JB: 117 DA: 77	**NN: 44** EL: 46 JB: 138 DA: 63
5	31043	**NN: 64** EL: 91 JB: 102 DA: 73	**NN: 36** EL: 41 JB: 117 DA: 53
6	11475	**NN: 53** EL: 79 JB: 83 DA: 61	**NN: 33** EL: 36 JB: 100 DA: 45
7	3454	**NN: 50** EL: 69 JB: 74 DA: 55	**NN: 36** EL: 38 JB: 93 DA: 42
8	624	**NN: 36** EL: 55 JB: 67 DA: 45	**NN: 33** EL: 37 JB: 95 DA: 41

Table 4: Resolution of the reconstructed kinematic variables in bins of x and Q^2. The resolution for x and Q^2 is defined as the RMS of the distributions $\log(x) - \log(x_{\rm true})$ and $\log(Q^2) - \log(Q^2_{\rm true})$ respectively.

First application of DL for regression of DIS kinematics

M. Diefenthaler, A. Farhat, A. Verbytskyi, Y Xu. "Deeply learning deep inelastic scattering kinematics." EPJ C 82.11 (2022): 1064.

Input Features

- Define variables to characterize the strength of QED radiation

Benchmark: input features and H1 MC dataset of paper NIM-A 1025 (2022): 166164*

$$p_T^{\mathrm{bal}} = 1 - \frac{p_{T,e}}{T} = 1 - \frac{\Sigma_e \tan\frac{\gamma}{2}}{\Sigma \tan\frac{\theta}{2}} \quad \text{and} \quad p_z^{\mathrm{bal}} = 1 - \frac{\Sigma_e + \Sigma}{2\ E_0}.$$

7 features to help indicate QED radiation in the event

- The values of p_T^{bal} and p_z^{bal}.
- The energy, η, and $\Delta\phi$ of the reconstructed photon in the event that is closest to the electron-beam direction, where $\Delta\phi$ is with respect to the scattered electron.
- The sum ECAL energy within a cone of $\Delta R < 0.4$ around the scattered electron divided by the scattered-electron track momentum.
- The number of ECAL clusters within a cone of $\Delta R < 0.4$ around the scattered electron.

+ additional 8 features

- Scattered-electron quantities $p_{T,e}$, $p_{z,e}$ and E.
- HFS four-vector quantities T, $p_{z,h}$ and E_h.
- $\Delta\phi(e, h)$ between the scattered electron and the HFS momentum vector.
- The difference $\Sigma_e - \Sigma$.

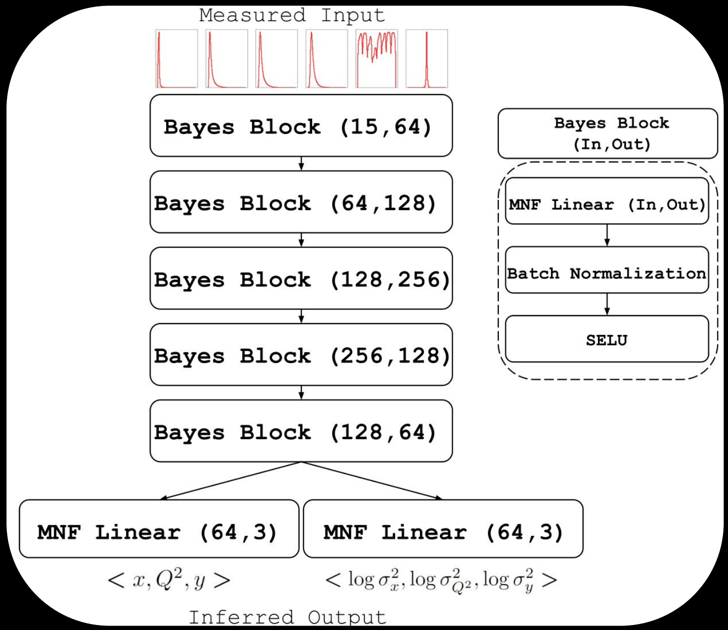
Tot. 15 input features

Dataset	Training Events	Validation Events	Testing Events	Size on Disk
H1	8.7×10^6	1.9×10^6	1.9×10^6	8 GB

*M. Arratia, D. Britzger, O. Long, B. Nachman, et al., “Reconstructing the kinematics of deep inelastic scattering with deep learning", NIM-A 1025 (2022): 166164

ELUQuant

C. Fanelli, J. Giroux, "ELUQuant: Event-Level Uncertainty Quantification" arxiv:2310.02913 [cs.LG]
(under review on Mach. Learn. Sci. Techn.)



$$\mathcal{L}_{Tot.} = \mathcal{L}_{Reg.} + \gamma\mathcal{L}_{Phys.} + \beta\mathcal{L}_{NF.}$$

Learn the Posterior over the weights

$$\mathcal{L}_{MNF.} = \mathbb{E}_{q(\mathbf{W},\mathbf{z}_T)}[-KL(q(\mathbf{W}|\mathbf{z}_{T_f})||p(\mathbf{W})) + \log r(\mathbf{z}_{T_f}|\mathbf{W}) - \log q(\mathbf{z}_{T_f})]$$

Access epistemic (systematic) uncertainty through sampling MNF [1] layers

Learn the regression transformation

$$\mathcal{L}_{Reg.} = \frac{1}{N}\sum_i\sum_j \frac{1}{2}(e^{-\mathbf{s}_j}\|\mathbf{v}_j - \hat{\mathbf{v}}_j\|^2 + \mathbf{s}_j),\ \ \mathbf{s}_j = \log\boldsymbol{\sigma}_j^2$$

Access aleatoric (statistical) as a function of regressed output [2]

Constrain the physics

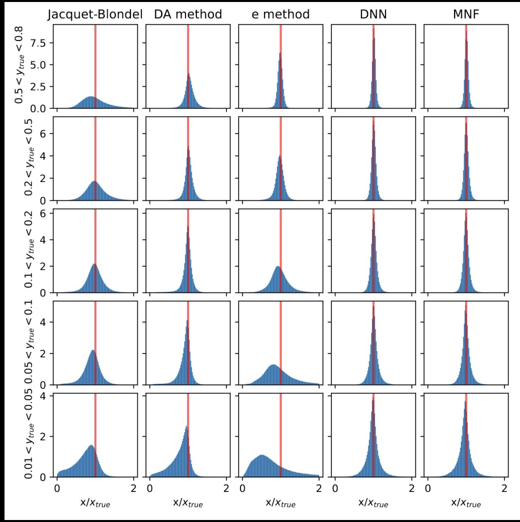
$$\mathcal{L}_{Phys.} = \frac{1}{N}\sum_i \log\hat{Q}_i^2 - (\log s_i + \log\hat{x}_i + \log\hat{y}_i)$$

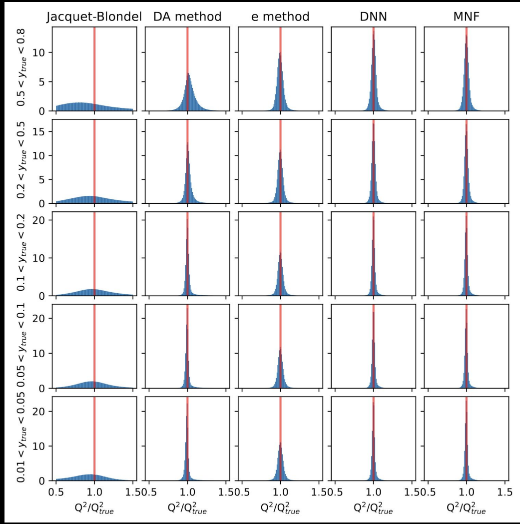
[1] C Louizos, M Welling International Conference on Machine Learning; arXiv:1703.01961 Multiplicative Normalizing Flows for Variational Bayesian Neural Networks
[2] A. Kendall and Y. Gal. "What uncertainties do we need in Bayesian deep learning for computer vision?." Adv. Neural Inf. Process. 30 (2017).

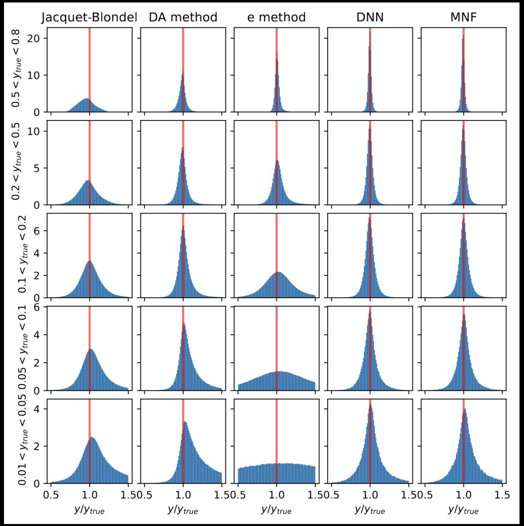
ELUQuant

C. Fanelli, J. Giroux, “ELUQuant: Event-Level Uncertainty Quantification” arxiv:2310.02913 [cs.LG]
(under review on Mach. Learn. Sci. Techn.)

Each method has advantages and disadvantages, and no single approach is optimal over the entire phase space. Each method exhibits different sensitivity to QED radiative effects







Y Bin	DA Method	DNN RMS	Aleatoric
(0.5, 0.8)	0.147955	0.061922	0.057942
(0.2, 0.5)	0.134833	0.075418	0.061706
(0.1, 0.2)	0.145530	0.097903	0.071238
(0.05, 0.1)	0.175290	0.132783	0.082945
(0.01, 0.05)	0.252723	0.184589	0.115453

Table 2: Aleatoric RMS Comparions - X

Y Bin	e Method	DNN RMS	Aleatoric
(0.5, 0.8)	0.056694	0.044052	0.041349
(0.2, 0.5)	0.055787	0.037505	0.032280
(0.1, 0.2)	0.054219	0.033230	0.029640
(0.05, 0.1)	0.053403	0.032501	0.029411
(0.01, 0.05)	0.053470	0.032139	0.029431

Table 3: Aleatoric RMS Comparison - Q2

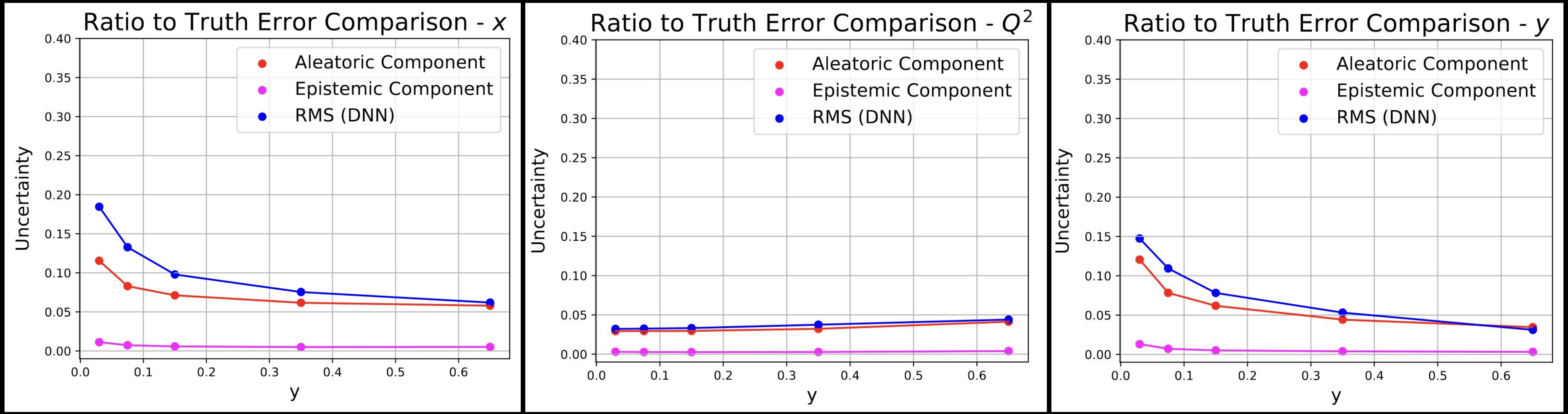
Y Bin	DA Method	DNN RMS	Aleatoric
(0.5, 0.8)	0.060537	0.031194	0.034643
(0.2, 0.5)	0.082115	0.053126	0.044249
(0.1, 0.2)	0.098631	0.078143	0.061840
(0.05, 0.1)	0.127276	0.109309	0.078276
(0.01, 0.05)	0.158493	0.147391	0.120546

Table 4: Aleatoric RMS Comparison Y

- Performance similar to DNN
- Closure test on aleatoric when epistemic is negligible and distribution is gaussian

*M. Arratia, D. Britzger, O. Long, B. Nachman, et al., “NIM-A 1025 (2022): 166164 (DNN)

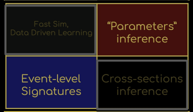
Comparison between DNN and BNN

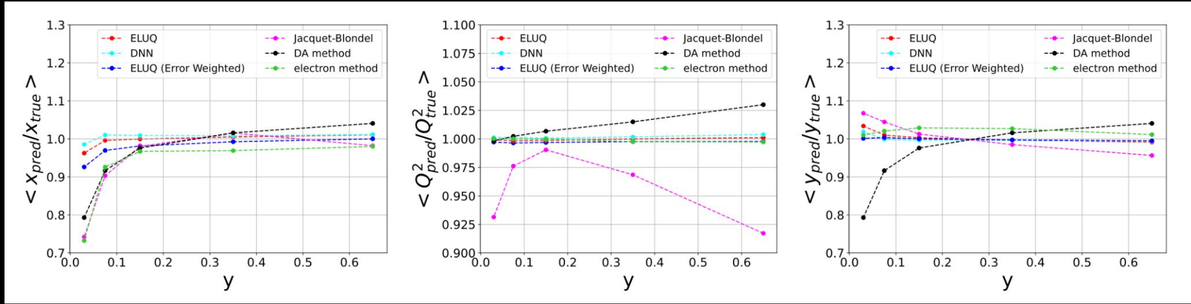


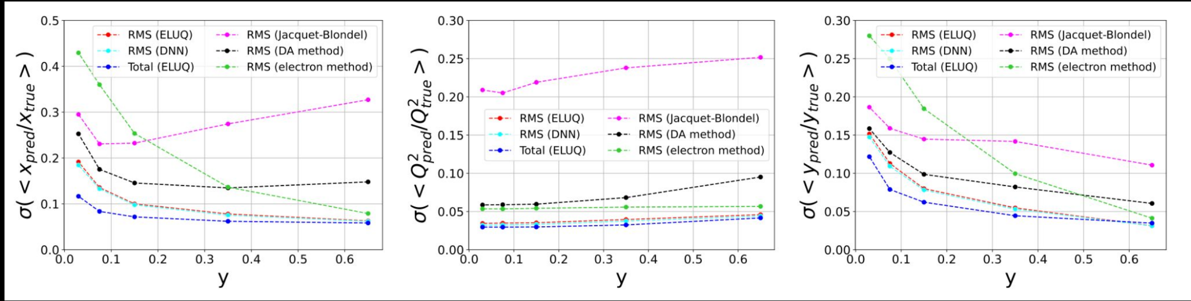
- The RMS (MNF) roughly coincide with that of DNN as seen previously
- The RMS (DNN) for x and y is larger at low y given the distributions are broader
- The epistemic is systematically smaller than aleatoric component.
- At large y, for x and y the total uncertainty (epistemic+aleatoric) close to RMS of DNN

— Reporting uncertainty at the level of the event (e.g., RMS from other methods) —

ELUQuant

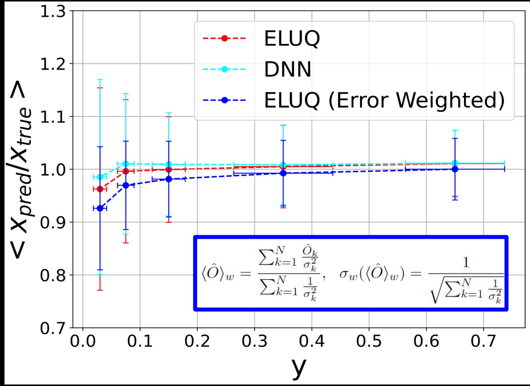


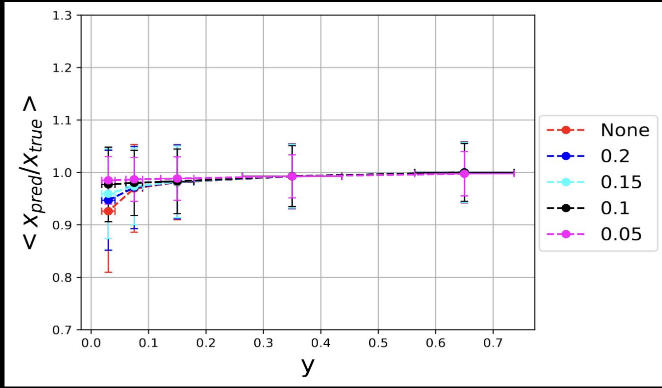




DNN and ELUQ "*outperform other methods over a wide kinematics range*"
NIM-A 1025 (2022): 166164

The RMS resolution for y and x increase at lower y, even for the DNN reconstruction. … This results … may be attributed to further acceptance, noise, or resolution effects that deteriorates the measurement of the HFS"





- A "simple" DNN does not have per se uncertainty at the event level. In the plots we use the RMS from final distributions.
- Removing events with large relative event-level uncertainty (with respect to the network prediction) improve the ratio to truth and reduce inaccuracy
- Notice these cuts do not use any information at the ground truth level
- We know that ELUQuant is sensitive to anomaly detection. Performance studies are underway.

Time performance

- This is great, but what about computing time?

Training Parameter	value
Max Epochs	100
Batch Size	1024
Decay Steps	50
Decay Factor (γ)	0.1
Physics Loss Scale (α)	1.0
KL Scale (β)	0.01
Training GPU Memory	$\sim$ 1GB
Network memory on local storage	$\sim$ 7MB
Trainable parameters	611,247
Wall Time	$\sim$ 1 Day

Inference specs of ELUQuant

Inference Parameter	value
Number of Samples (N)	10k
Batch Size	100
Inference GPU Memory	$\sim$ 24GB
Inference Time per Event	$\sim 20ms$

Inference specs of ELUQuant

- In computational terms, ELUQuant at inference showed an impressive rate of 10,000 samples/event within a 20 milliseconds on an RTX 3090.

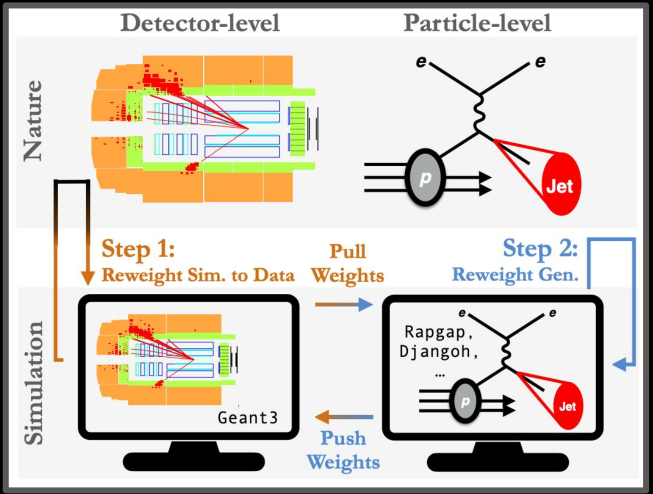
- Can we do faster than this?
  - Several ways. A rapid, streamlined approach is distilling this knowledge in a simpler but faster network (we explored a DNN with 450k parameters) called in the following “Fast UQ”, obtaining an effective inference time of 7-8us/event using batch ~0.5M events

Jefferson Lab
Exploring the Nature of Matter

Unfolding

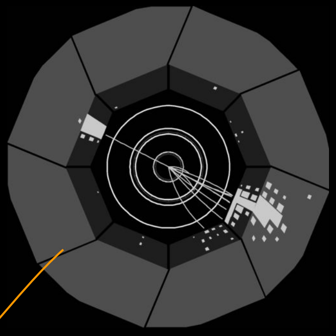
Unfolding

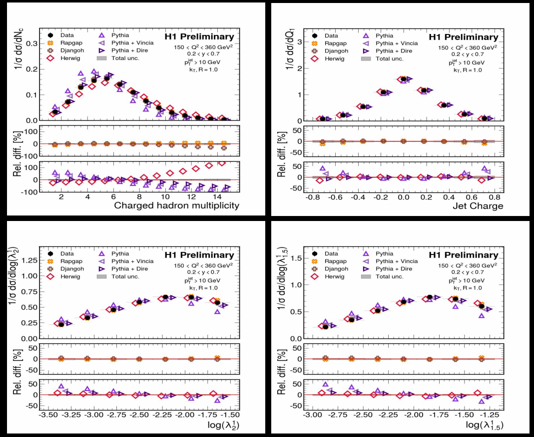
OmniFold



Using ML for differential cross section measurements (OmniFold and otherwise). These tools for recent measurements with DIS from HERA data and the same tools could be used at the EIC.

Lepton-jet correlation in DIS at H1





- First example of ML-assisted unfolding (MultiFold method): enables simultaneous and unbinned unfolding in high dimensions.
- This development will allow us to do unbinned cross-section measurements.

[1] A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler Phys. Rev. Lett. 124, 182001 2020

[2] V. Andreev et al. (H1 Collaboration), "Measurement of Lepton-Jet Correlation in Deep-Inelastic Scattering with the H1 Detector Using Machine Learning for Unfolding" Phys. Rev. Lett. 128, 132002

IBU^2: Invertible Bayesian Unfolding with Uncertainty

Benchmark: datasets of OmniFold (1911.09107) for jet physics, pp √s=14 TeV, Herwig (nature) and Pythia (synthetic) — DELPHES for CMS fast sim

- Learns an invertible mapping between injected and reconstructed events (bonus: fast and accurate generation of data)
- Utilizes Bayesian Networks (cf. ELUQuant)
- Allows to unfold a measured event and get posterior (with uncertainty)
- Further treated by MCMC

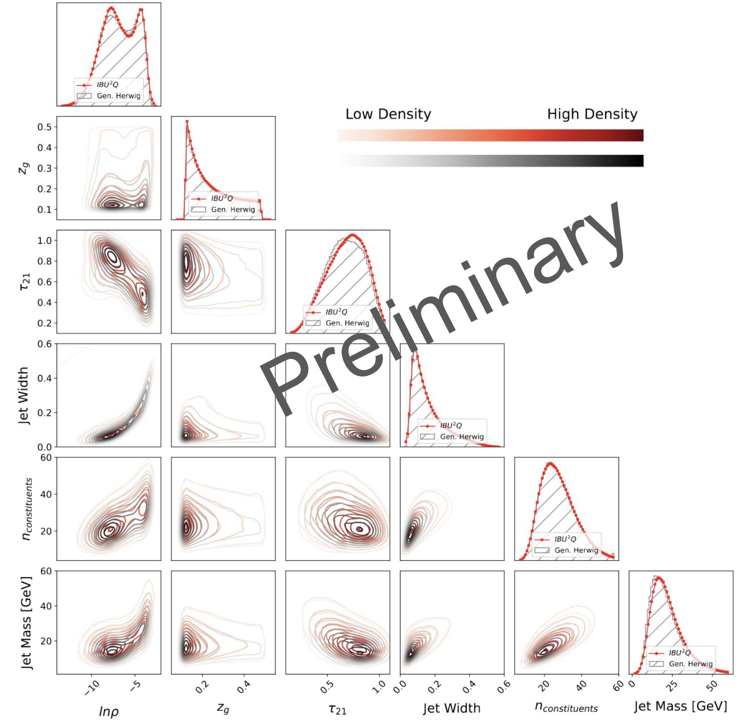
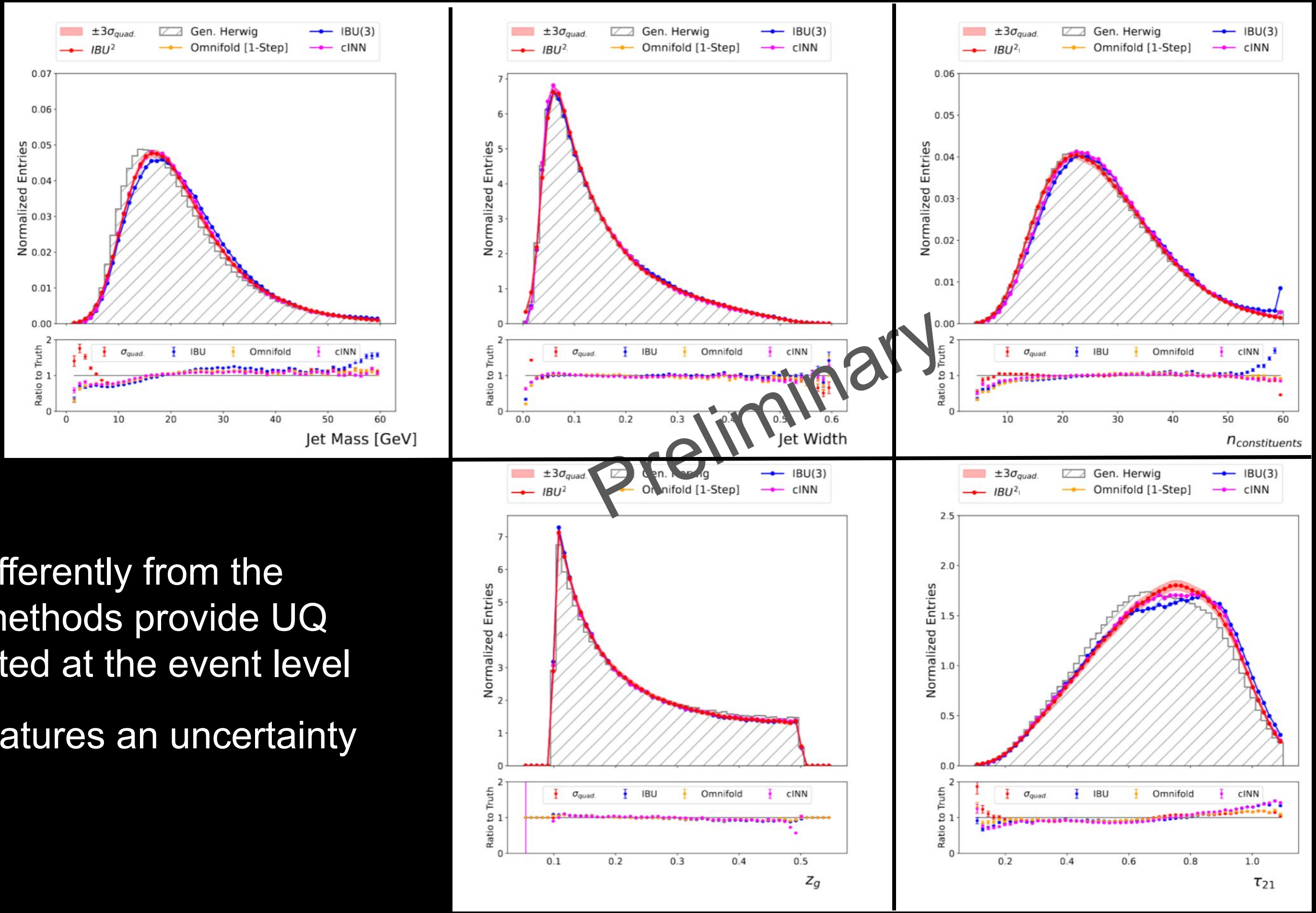


Figure 4: Full 600k events now.

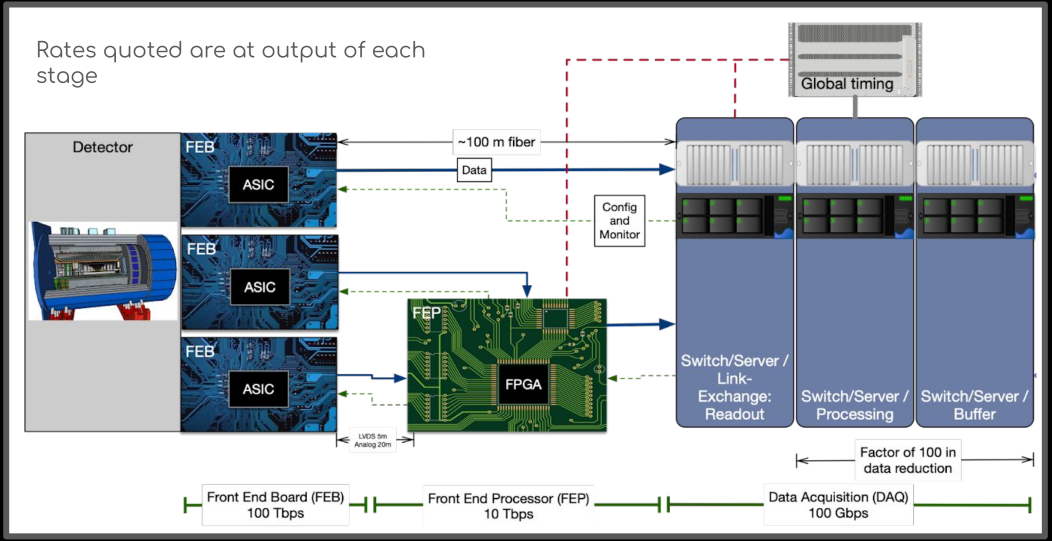
IBU^2

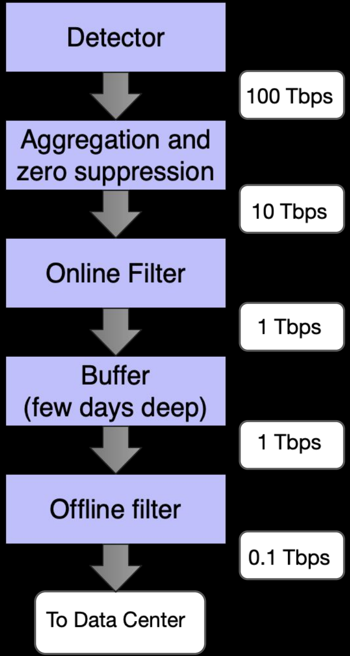


- IBU^2 differently from the other methods provide UQ calculated at the event level
- IBU^2 features an uncertainty band

AI/ML in Streaming Readout

- SRO quickly becoming the new standard readout paradigm for modern NP and HEP experiments.
- A triggerless streaming architecture gives much more flexibility to do physics (max data preservation, diverse topologies). Data flow unimpeded in parallel channels, organized in multi-dimensions and time.
- Manageable event rates at EIC (500 kHz).





SRO will further the convergence of online and offline analyses, with the possibility of incorporating AI/ML for fast reconstruction and calibrations, allowing for a rapid turnaround of physics data and results

[1] J. Bernauer, C. Dean, C. Fanelli, J. Huang, et al, NIMA 1047 (2023): 167859.

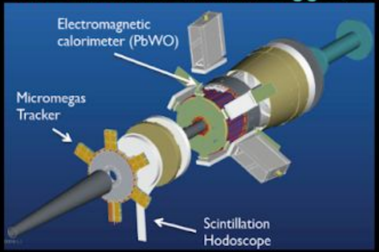
Prototype experiments for next-gen SRO

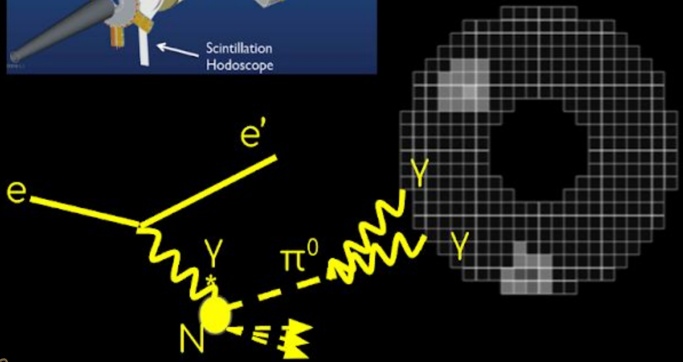
ML deployed on stream of real data

CLAS + EPSCI @JLab

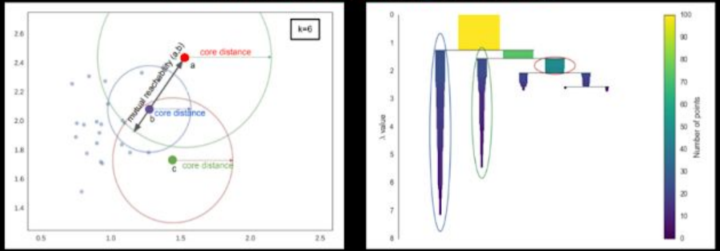
- CLAS12 SRO setup
- TriDAS SR back end
- JANA2 reconstruction framework

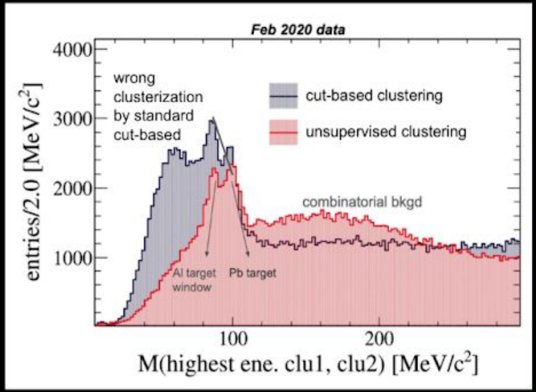
The CLAS12 Forward Tagger, JLab

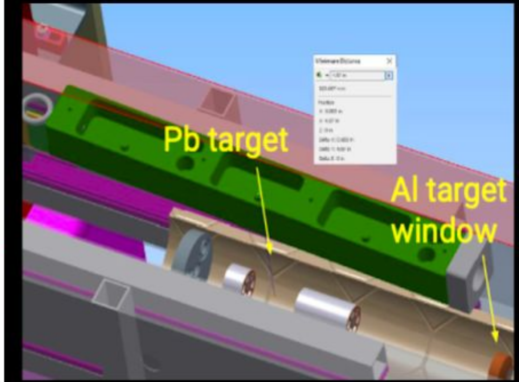




Hierarchical clustering in JANA2



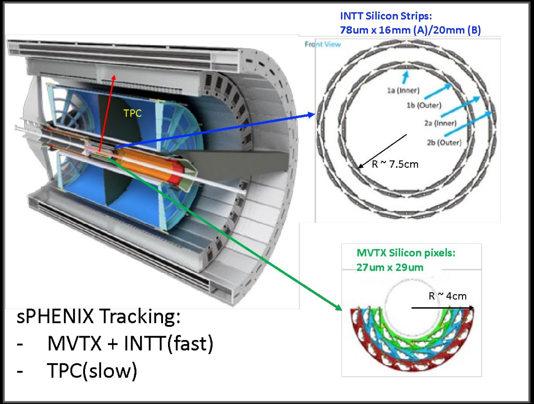




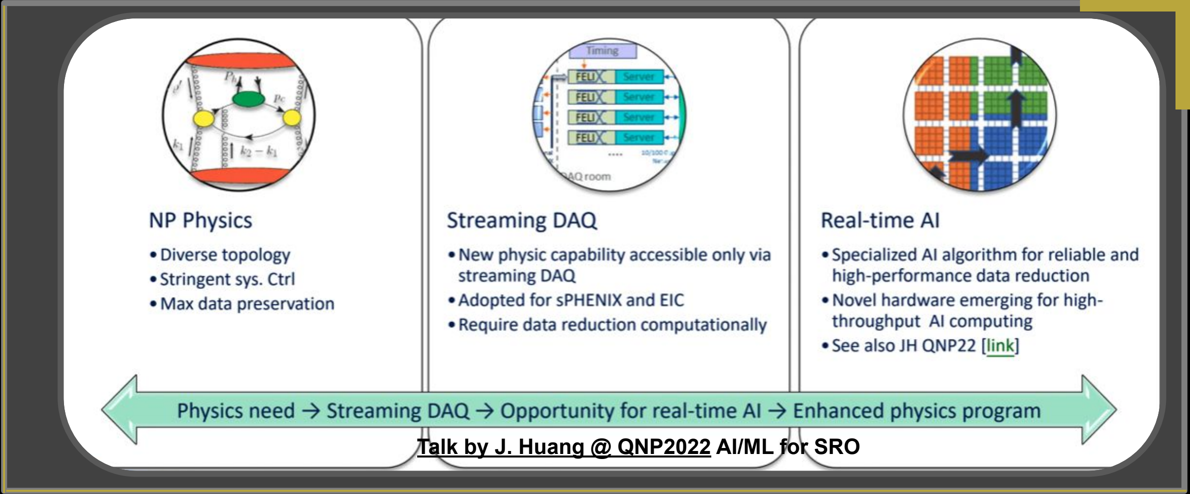
Hierarchical clustering VS traditional clustering of energy deposited by photons; AI robust against variations in experimental conditions* (uncalibrated data in SRO)

[1] F. Ameli, et al., Streaming readout for next generation electron scattering experiments, Eur. Phys. J. Plus, 2022

AI/ML in SRO



sPHENIX Tracking:
- MVTX + INTT(fast)
- TPC(slow)



Talk by J. Huang @ QNP2022 AI/ML for SRO

FastML: **Fast Data Processing and Autonomous Detector Control for sPHENIX and Future EIC Detectors**

Intelligent Experiment Through Real-Time AI
(DOE FOA funded 2022-2023)

Collaboration of NP, HEP and CS:
LANL, MIT, FNAL, NJIT, ORNL, UNT, CCNU

Identify D/B hadrons with real-time ML
- Topology of D/B decays
- Monitor collision vertex
- Feedback for improvement

The challenges:

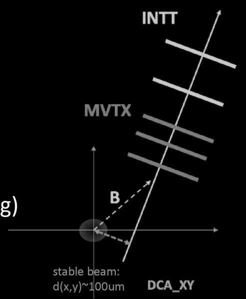
Very high p+p collision rate: ~3MHz

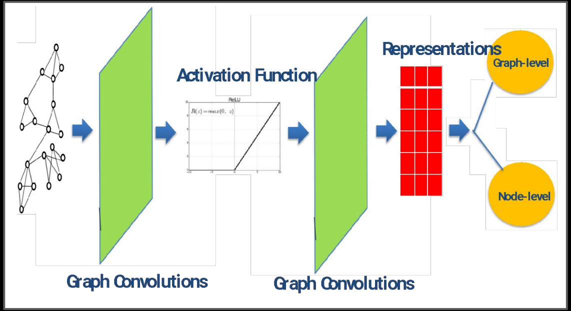
Low rate of rare signals: ~150Hz (beauty for eg)

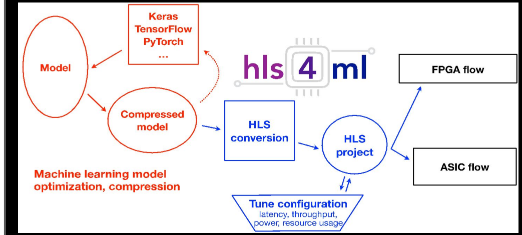
Limited DAQ trigger bandwidth: ~15 kHz

(or 0.5% of p+p collisions)

No effective conventional triggers available







Courtesy of Ming Liu (LANL)

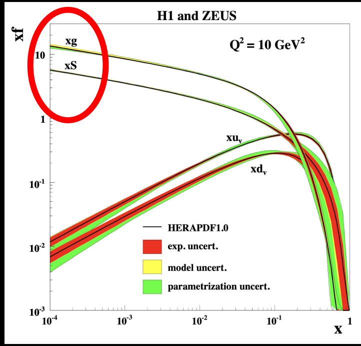
[1] Huang, Yi, et al. "Efficient Data Compression for 3D Sparse TPC via Bicephalous Convolutional Autoencoder." 2021 20th IEEE (ICMLA). IEEE, 2021.
[2] F. Fahim, et al., “HLS4ML” arXiv:2103.05579 (2021)
[3] C. Dean, Autonomous selection of physics events: A RHIC demonstrator for EIC physics — AI4EIC2023 talk

Conclusions

- AI/ML can be integrated into virtually every facet of the data processing pipelines of HEP/NP experiments
- Next generation QCD experiments like EIC are being designed during the AI revolution, and can take advantage of AI/ML since the design and R&D phase. The EIC detector(s) may the first large-scale detectors optimized with machine learning.
- Hadronic physics will increasingly benefit from ML; when it comes to study non-perturbative effects, ML allows a “holistic” approach (full event information) and can be trained on real data
- Next generation QCD experiments will take full advantage of SRO and AI using heterogeneous computing:
  - Near real-time analysis / control (e.g., intelligent / autonomous detectors). A common theme is applying AI-methods with well-understood UQ (both systematic and statistic).
  - If we understand the uncertainties and biases, near real-time analysis with SRO can result in a paradigm shift for next generation QCD experiments, with faster turnaround time to produce scientific results.

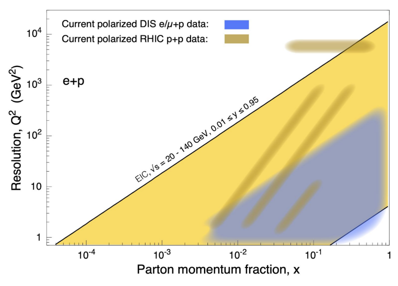
Backup

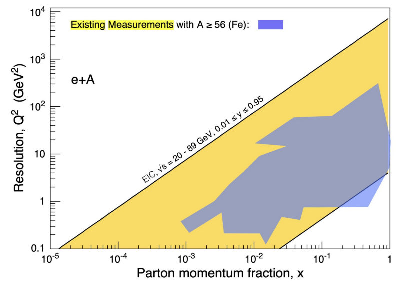
[1]R. A. Khalek, et al. "Science requirements and detector concepts for the electron-ion collider: EIC yellow report." NIMA 1026 (2022): 122447.
[2] HERA Coll. , JHEP 1001:109(2010)



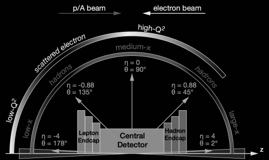
Without gluons there would be no nucleons, no atomic nuclei, …

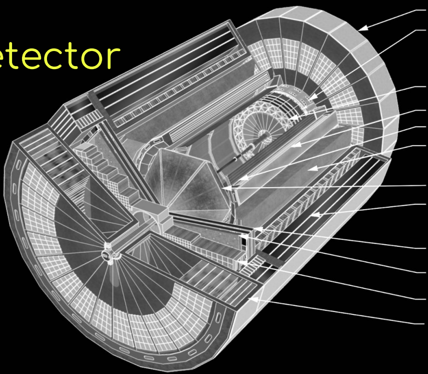
EIC Science Landscape



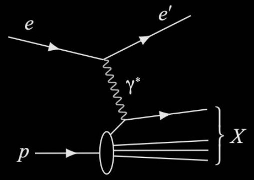


EIC detector

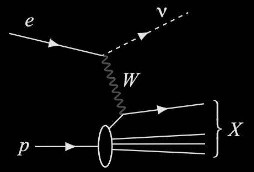




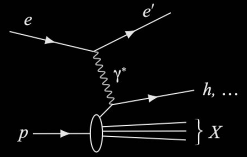
Neutral-current Inclusive DIS: $e + p/A \longrightarrow e' + X$; for this process, it is essential to detect the scattered electron, e', with high precision. All other final state particles (X) are ignored. The scattered electron is critical for all processes to determine the event kinematics.



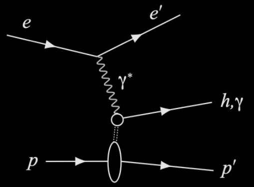
Charged-current Inclusive DIS: $e + p/A \longrightarrow \nu + X$; at high enough momentum transfer Q^2, the electron-quark interaction is mediated by the exchange of a $W^{\pm}$ gauge boson instead of the virtual photon. In this case the event kinematic cannot be reconstructed from the scattered electron, but needs to be reconstructed from the final state particles.



Semi-inclusive DIS: $e + p/A \longrightarrow e' + h^{\pm,0} + X$, which requires measurement of *at least one* identified hadron in coincidence with the scattered electron.



Exclusive DIS: $e + p/A \longrightarrow e' + p'/A' + \gamma/h^{\pm,0}/VM$, which require the measurement of *all* particles in the event with high precision.



Tracking

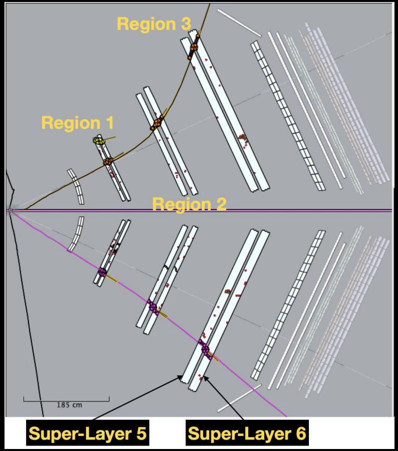
The CLAS spectrometer investigates nucleon and meson structures using a solenoid-torus magnetic field for wide acceptance, precise tracking, and efficient charged particle separation with background suppression.

Tracking in NP experiments poses unique challenges:

(1) Compared to HEP, low track multiplicities and large curvatures (lower P and relatively large, non-uniform magnetic fields)

(2) Typically represents the most substantial CPU resource usage*

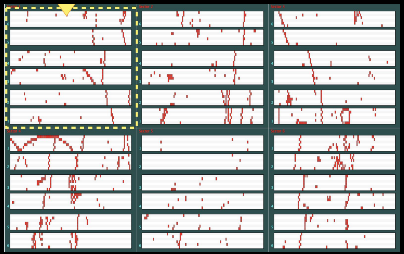
CLAS12 Tracking



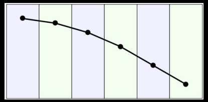
Tracks detected by DCs in toroidal field:

- Each sector has 3 regions
- Each region has 2 Super-Layers
- Super-Layer has 6 layers
- Each Layer has 112 wires

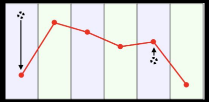
Sector 1



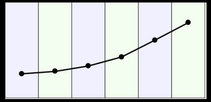
True track



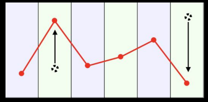
False track

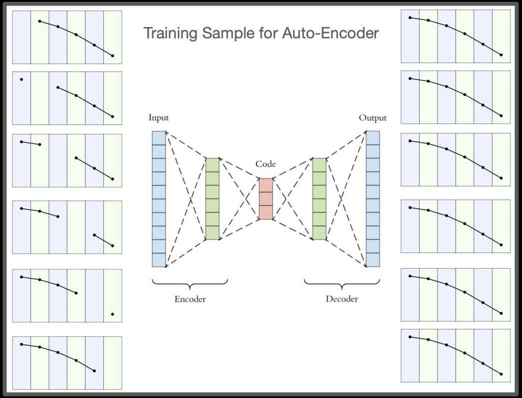


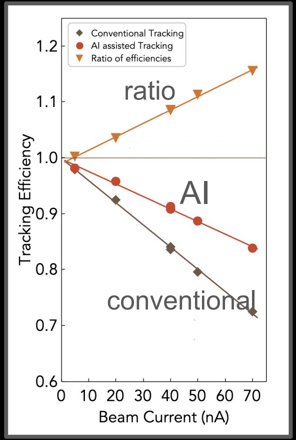
True track



False track







AI-assisted tracking in CLAS12/JLab:

- Track classification
- Missing segment generation
- Denoising drift chamber data

Implemented in the CLAS SW stack as a service. AI-assisted tracking provided a 6 times code speedup.

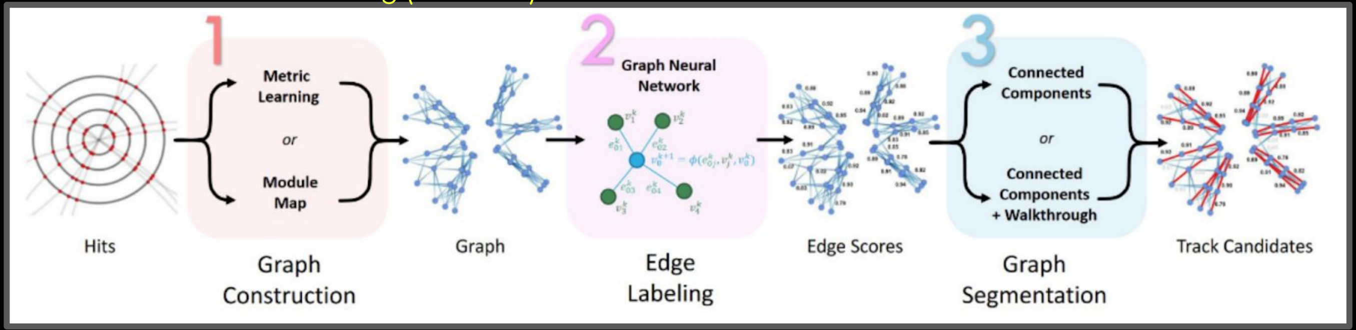
They also implement a Level 3 trigger for identifying electron candidates from raw information from DC and ECAL

[1] G. Gavalian, et al. "Using Artificial Intelligence for Particle Track Identification in CLAS12 Detector." arXiv preprint arXiv:2008.12860 (2020).
[2] G. Gavalian. "Auto-encoders for Track Reconstruction in Drift Chambers for CLAS12." arXiv preprint arXiv:2009.05144(2020).

Tracking

experiment independent ML-based tracking in HEP

Accelerated GNN tracking (IRIS-HEP)



Use Graph Neutral Network (GNN) to reconstruct tracks

Embedding : Use all the hits in the detector to build a graph

Filtering : Neural-Network predicts if nodes should be connected (can also use a connection map)

ACTS: a common tracking software

Features: (i) Tracking geometry description, (ii) simple event data model, (iii) most track reconstruction algorithms, (iv) example framework with python bindings, (v) performance evaluation algorithms

Provides a testing environment for new tracking algorithms, open detector data (based on the TrackML challenge)

https://github.com/acts-project/acts

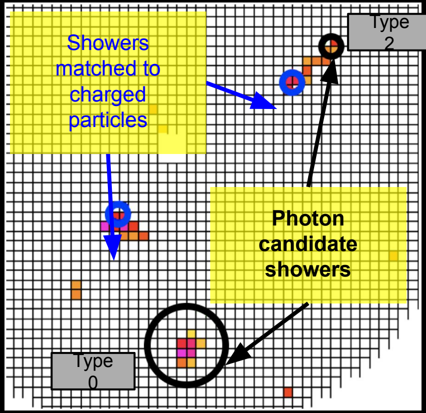
Towards end-to-end pipelines for tracking.
Kalman Filter remains a powerful tool to completely “throw” away… GPU-accelerated KF

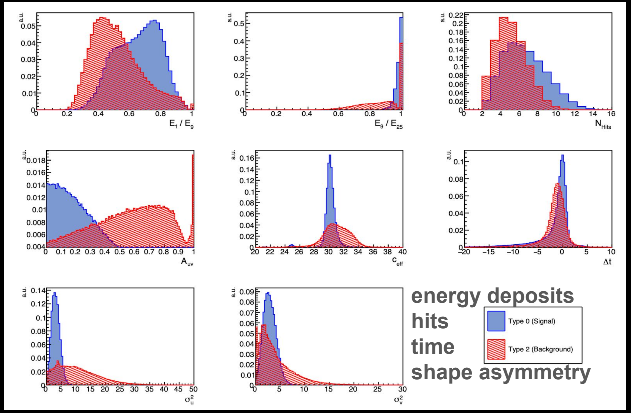
[1] L.-G. Gagnon, Machine learning for track reconstruction at the LHC, 2022 JINST 17 C02026 — AI4EIC workshop
[2] Exa.TrkX: HEP tracking at the exascale. A DOE CompHEP project, https://exatrkx.github.io/
[3] A. Akram, and X. Ju. "Track Reconstruction using Geometric Deep Learning in the Straw Tube Tracker (STT) at the PANDA Experiment." arXiv:2208.12178 (2022)
[4] D. Rohr "Overview of online and offline reconstruction in ALICE for LHC Run 3." arXiv:2009.07515 (2020) https://arxiv.org/abs/2009.07515

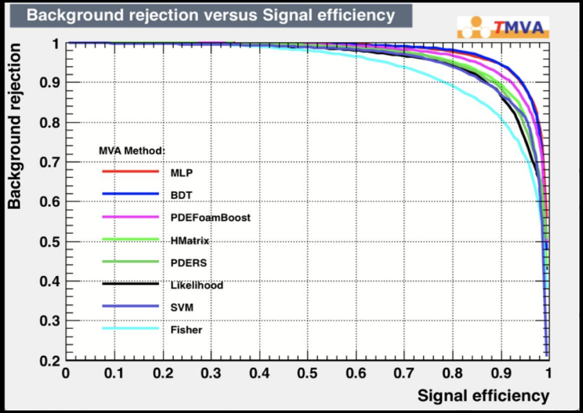
Neutral showers

Separation of electromagnetic and hadronic interactions (i.e., low energy vs split-offs) in the GlueX FCAL (2800 PbWO4 modules)

Trained on $\omega \to \pi^+\pi^-\pi^0(\gamma\gamma)$ (true photons and charged particles interacting with FCAL)







Showers classified as:
Type 0 (true photon showers from hadron decays, e.g., π^0)
Type 1 (from charged particles colliding with FCAL)
Type 2 (all other types of showers, e.g., split-offs of a Type 1 or background noise)

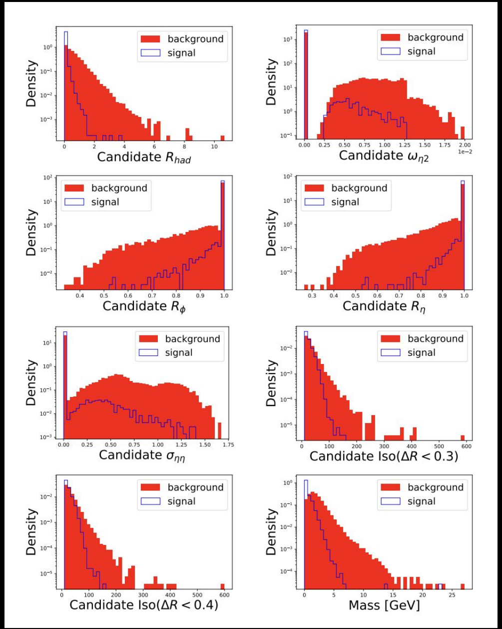
Goal: distinguish **Type 0** from **Type 2**

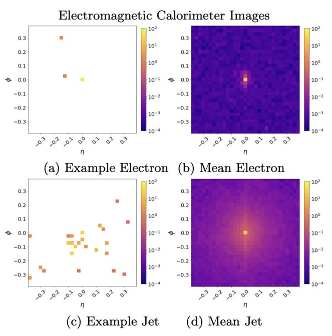
- MLP selected due to ease of implementation within GlueX SW framework
- Thorough data/MC comparison (agreement within statistical precision) — Bkgd reduction of 60% and signal retention of 85% on inclusive π^0 data.

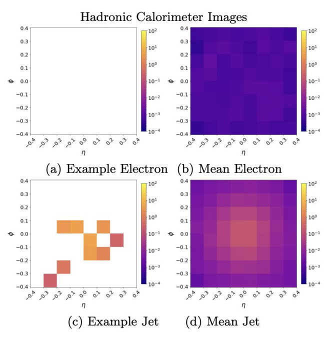
[1] R. Barsotti and M.R. Shepherd 2020 JINST 15 P05021

Electron Identification

Performance of deep networks like CNN reveal there is information in low-level image that is not captured by the suite of high-level features built by physicists



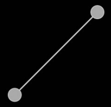




"Learning to identify electrons"
Aim is to identify new high-level features that bridge the gap between existing performance and superior performance of CNN

Search done through energy flow polynomials (EFP)

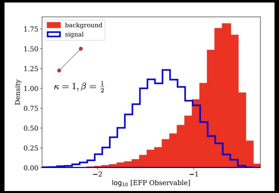
~Les Houches Angularity

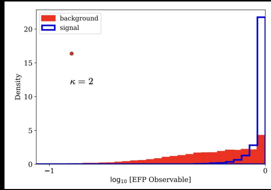


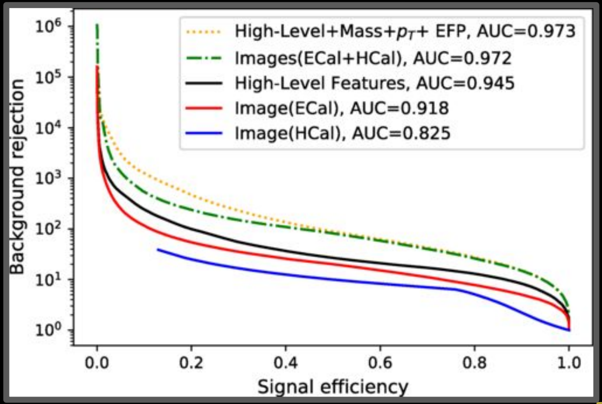
$$= \sum_{a,b=1}^{N} z_a z_b \theta_{ab}^{\frac{1}{4}}$$

~p_T^D

$$= \sum_{a=1}^{N} z_a^2$$





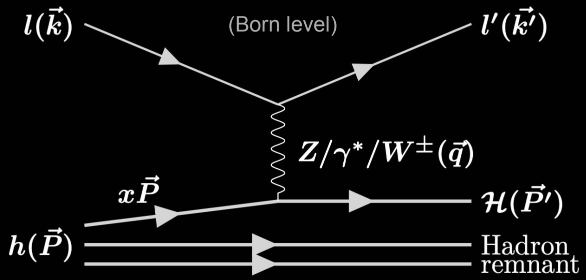


[1] J. Collado, J. N. Howard, T. Faucett, T. Tong, P. Baldi, and D. Whiteson Phys. Rev. D 103, 116028 – (2021), talk at AI4EIC (2022)

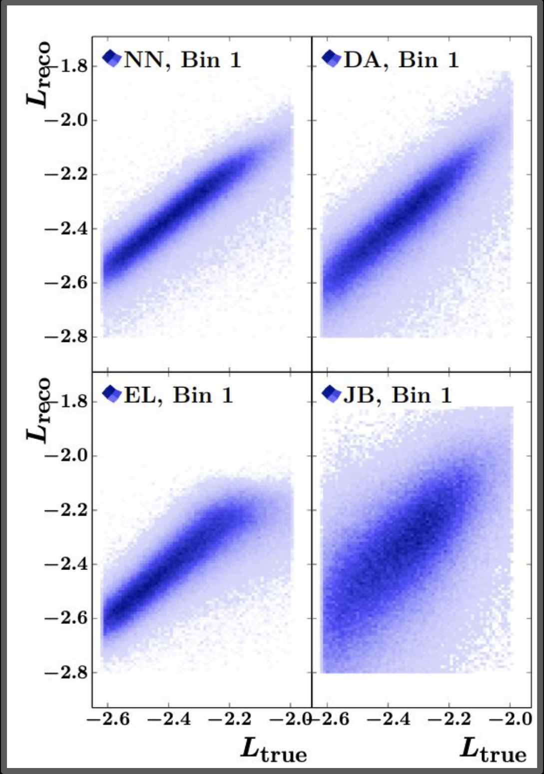
Deeply Learning DIS

Example in one specific bin

DIS fundamental process @EIC



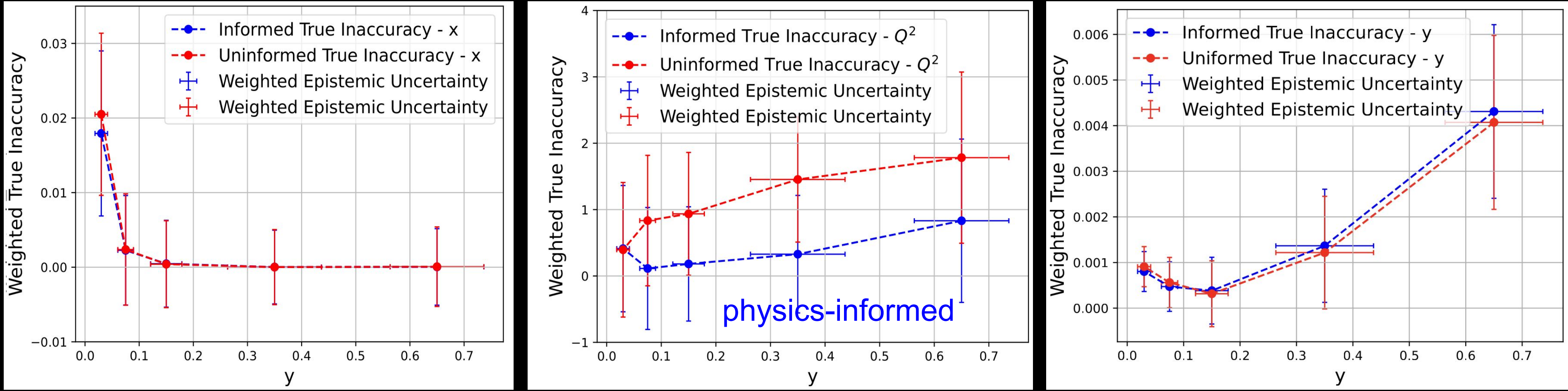
DIS beyond the Born approximation has a complicated structure which involve QCD and QED corrections



- Use of DNN to reconstruct the kinematic observable Q^2 and x in the study of neutral current DIS events at the ZEUS experiment at HERA.
- The performance compared to electron, Jacquet-Blondel and the double-angle methods using data-sets independent of training
- Compared to the classical reconstruction methods, the DNN-based approach enables significant improvements in the resolution of Q^2 and x

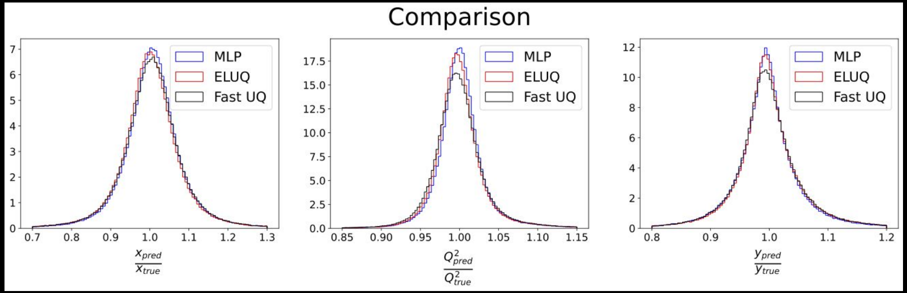
[1] M. Diefenthaler, et al. “Deeply Learning DIS Kinematics” arXiv:2108.11638
[2] M. Arratia, et al., “Reconstructing the kinematics of DIS with DL”, NIM-A 1025 (2022): 166164

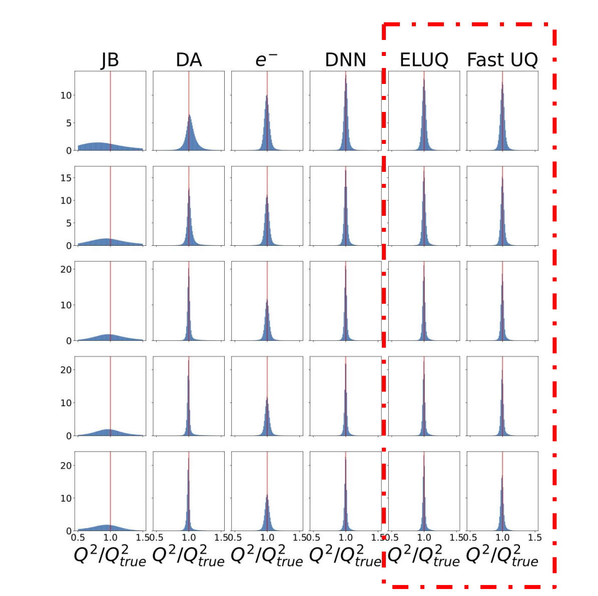
ELUQuant: Physics-informed term

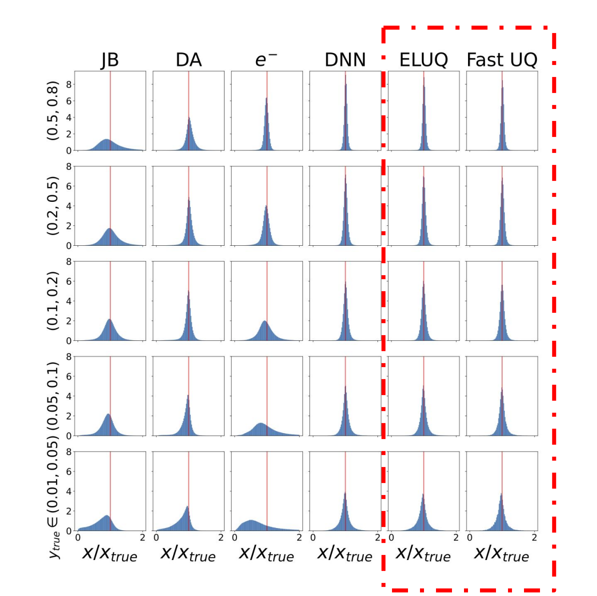


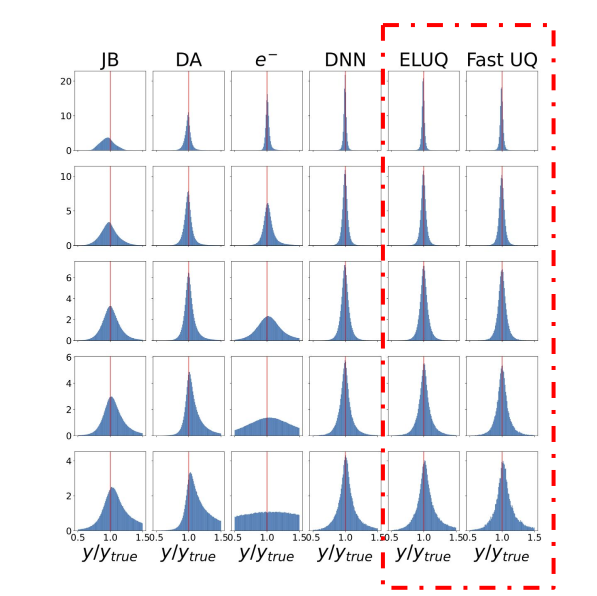
- The plots report the true inaccuracy, and the weighted epistemic uncertainty, which is larger the larger the true inaccuracy is
- The physics-informed term (blue) contributes to decrease the true inaccuracy.

ELUQuant: Towards near real-time

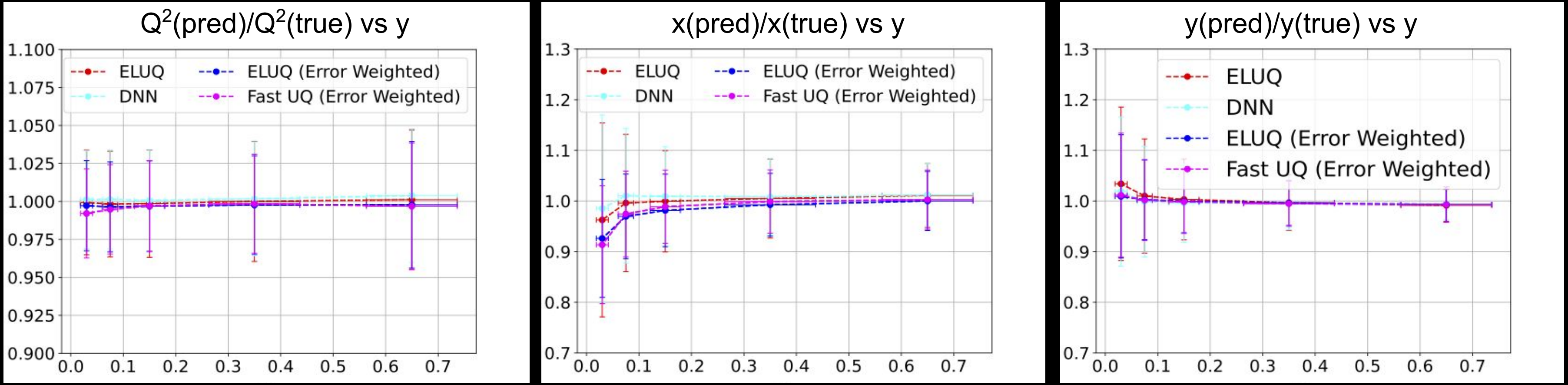


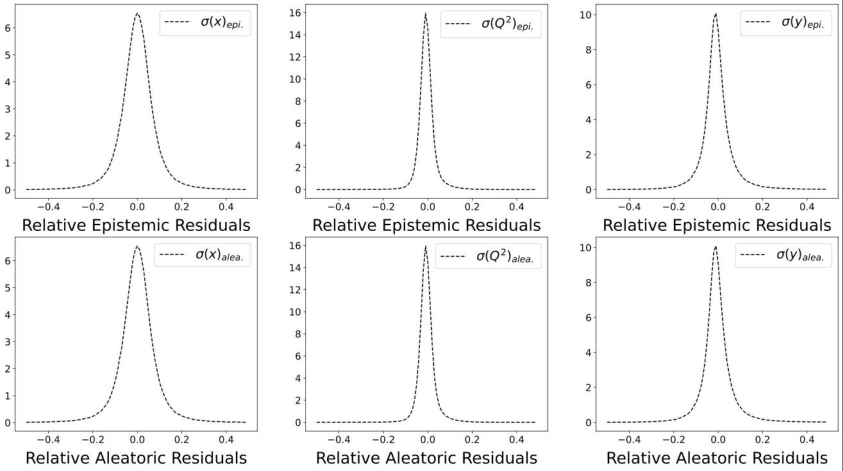






ELUQuant: Towards near real-time



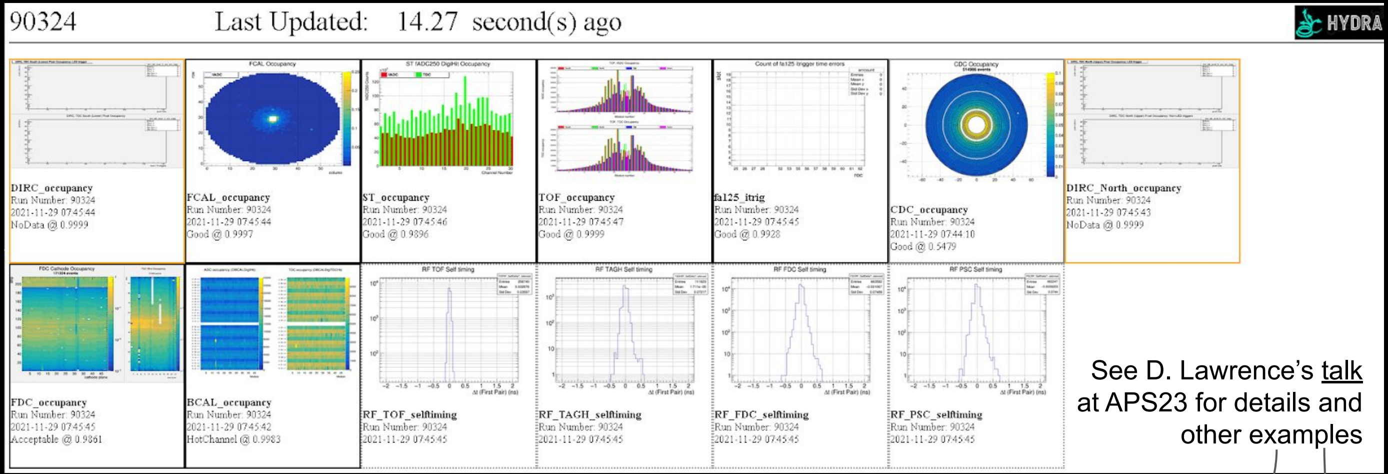


ELUQuant/Fast UQ: Very similar performance at the event level, predictions on kinematics and epistemic + aleatoric uncertainties within ~5% on average

Jefferson Lab
Exploring the Nature of Matter

Towards “autonomous” experiments

- Near real-time monitoring tasks for GlueX in Hall D
- It was the online monitoring coordinator’s job to sift through hundreds of images produced in the previous 24 hours, looking for missed anomalies. This “human-in-the-loop” method was prone to errors.
- Hydra was created to tackle these challenges. Hydra is an AI system that leverages Google’s Inception v3 for image classification. It has been shown to perform better than humans at diagnosing problems.



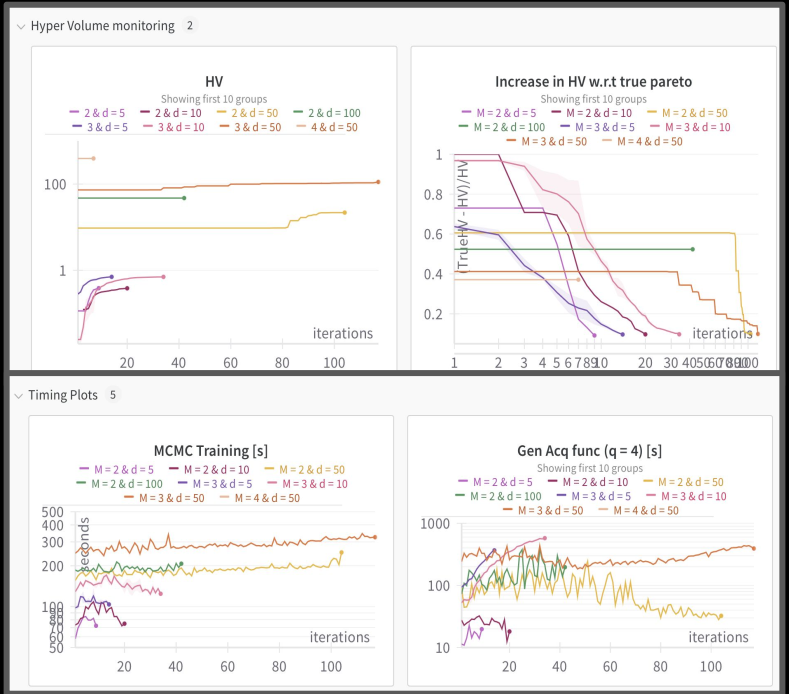
See D. Lawrence’s talk at APS23 for details and other examples

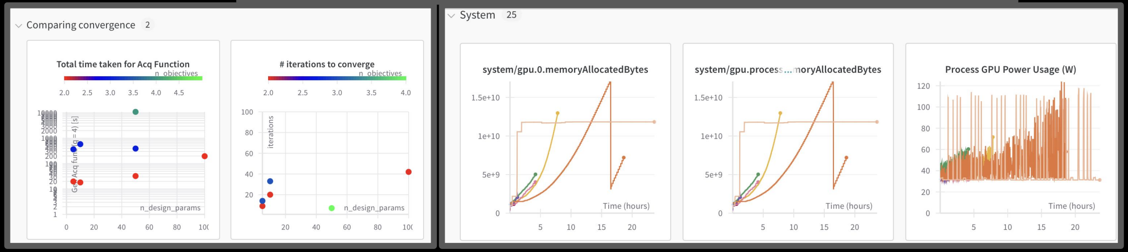
[1] T. Britton, D. Lawrence, K. Rajput, arXiv:2105.07948v1 [cs.CY]
[2] T. Jeske, et al. "AI for Experimental Controls at Jefferson Lab." JINST 17.03 (2022): C03043. — AI4EIC proceedings
[3] T. Britton, B. Nachman. "Accelerator and detector control for the EIC with machine learning." JINST 17.02 (2022): C02022. — AI4EIC proceedings

MOBO: Scaling

https://wandb.ai/phys-meets-ml/AID2E-Closure-1?workspace=user-karthik18495

- W&B dashboard for monitoring
  - MOBO stress-testing for problems with increasing complexity (design and objectives) and known Pareto
- Multiple metrics
  - Accuracy of optimization
  - Convergence properties
  - Compute resources

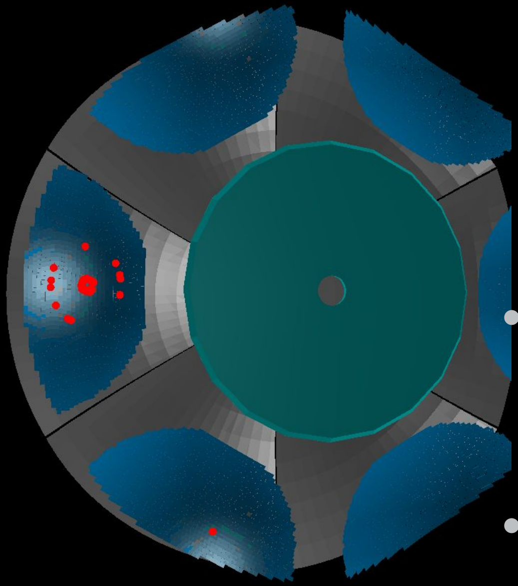




Candidates for Optimization in ePIC

Considering all the constraints as ePIC is in the process of finalizing engineering designs, we can select those sub-detectors that still have tunable parameters

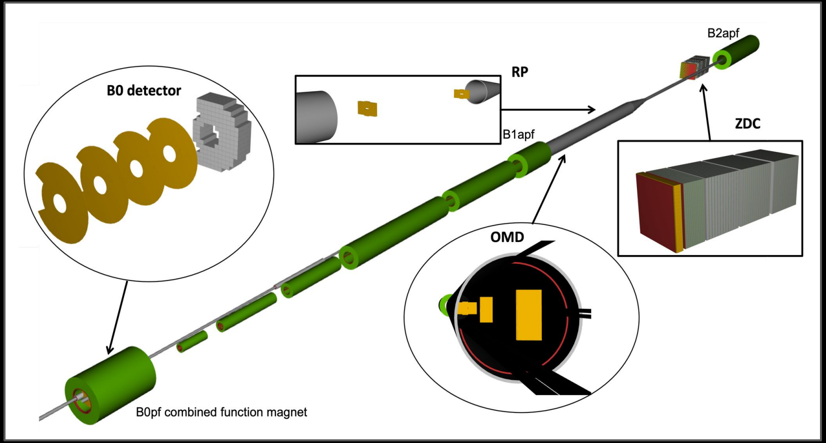
E. Cisbani *et al* 2020 *JINST* **15** P05009



- *Mirror, sensor placement, gas, mirror material (lower costs material)...*
- *PID performance, costs, …*

dual-RICH

- *B0 magnetic field map, distance between space (always considered even), central location of tracker*
- *Momentum resolution, acceptance*



Far-Forward

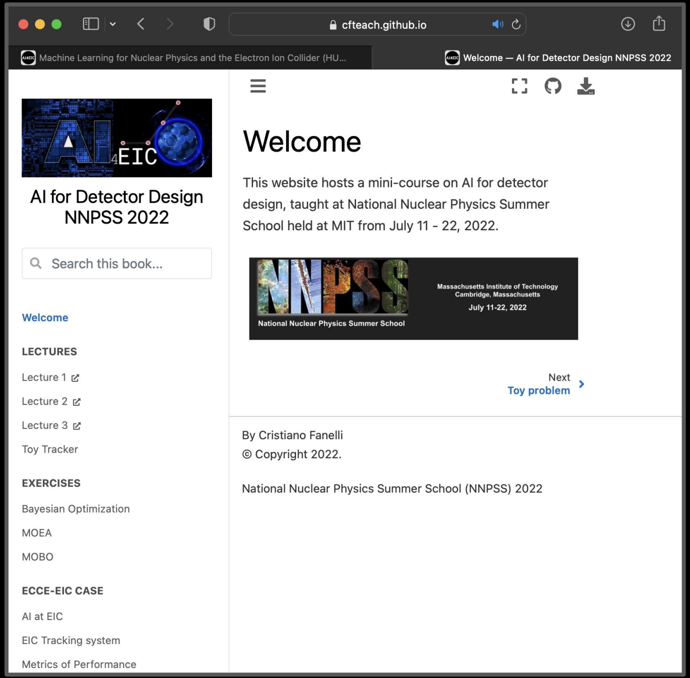
Ongoing discussion with working groups to identify potential

Documentation and Outreach

- GitBook and/or other knowledge sharing platforms will be part of the initiatives related to documentation and outreach
- Offering opportunities for experiential learning with easy access for beginners

http://cfteach.github.io/nnpss

https://cfteach.github.io/HUGS23



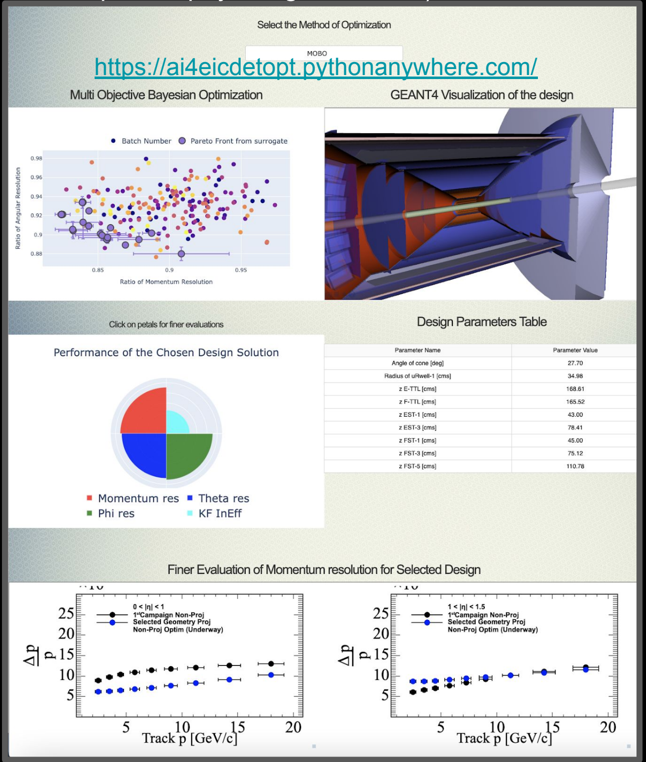
Assisted design of future QCD Experiments

Designing detectors is a multi-objective optimization problem!
(detector response, physics gains, costs)

Take full advantage of AI to learn the Pareto front

Design space: Multidimensional + Multiple Objectives!

This is a problem where with AI-assistance we can outperform more conventional strategies

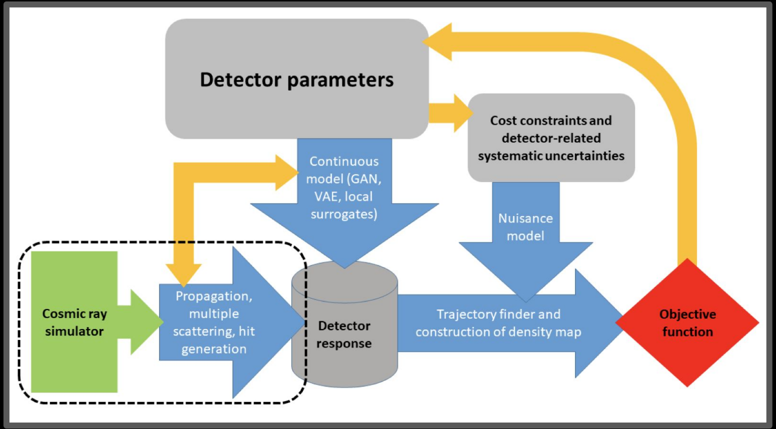


Parameter Name	Parameter Value
Angle of cone [deg]	27.70
Radius of uRwell-1 [cms]	34.98
z E-TTL [cms]	168.61
z F-TTL [cms]	165.52
z EST-1 [cms]	43.00
z EST-3 [cms]	78.41
z FST-1 [cms]	45.00
z FST-3 [cms]	75.12
z FST-5 [cms]	110.78

Differentiable surrogate model + gradient-based optimization

MODE is targeting the use of differentiable programming in design optimization of detectors for particle physics applications

End-to-end optimization pipelines with surrogate models ML require modeling of simulations, and collect reference data to train the implementations.



Conceptual layout of an optimization pipeline for a muon radiography apparatus.

[1] C. Fanelli, Z. papandreou, K. Suresh, et al. AI-assisted optimization of the ECCE tracking system at the Electron Ion Collider, NIMA 1047, 167748 (2023)
[2] A. G. Baydin et al. Nuclear Physics News 31.1 : 25-28 (2021).
[3] C. Fanelli, Design of detectors at the electron ion collider with artificial intelligence, 2022 JINST 17 C04038 (2022)
[4] F. Torales Acosta et al., "ML for Detector Optimization and Simulation", talk at AI4EIC2023 [link]