AI for BSM Physics Searches

Patrick Moran, The College of William & Mary

JLab 22 GeV, LNF-INFN, 9-13 Dec. 2024

WILLIAM & MARY
CHARTERED 1693

Overview

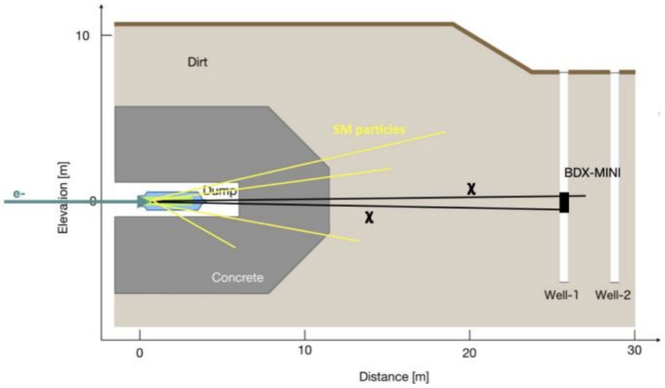
1. Dark Matter Detection

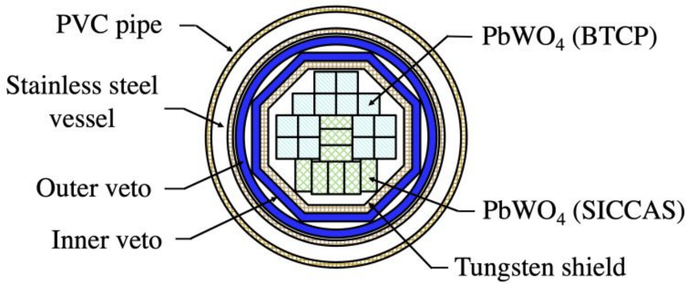
2. Anomaly Detection using Generative Models

3. Precision Measurements using Deep Learning

Mini-BDX

- Pilot version of BDX Experiment
- 25m downstream of Hall A beam dump
- Collected 6 months of data in 2019-20
- Detector consists of two layers of 22 $PbWO_4$ calorimeters each

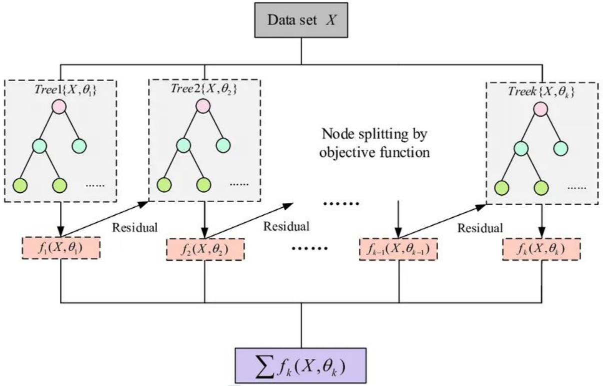




M. Battaglieri et al., Phys. Rev. D 106 (2022) 7, 072011

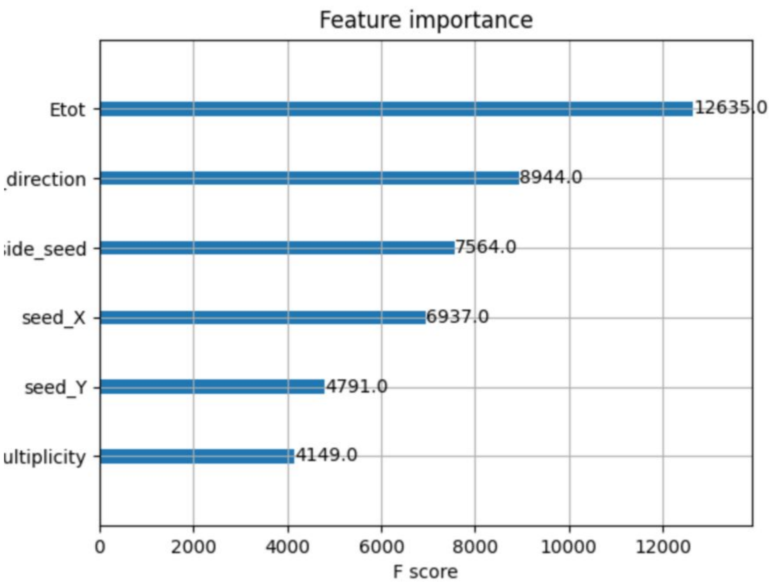
Boosted Decision Trees

- Boosted Decision Trees (BDTs) are machine learning models that combine multiple decision trees sequentially, where each tree is trained to correct the errors of the previous ones, creating a strong ensemble model
- Highly effective for classification and regression tasks
- XGBoost is an open-source library that uses gradient boosting
- Want to use BDT to discriminate dark matter signal from background (cosmics and neutrinos coming from beam)

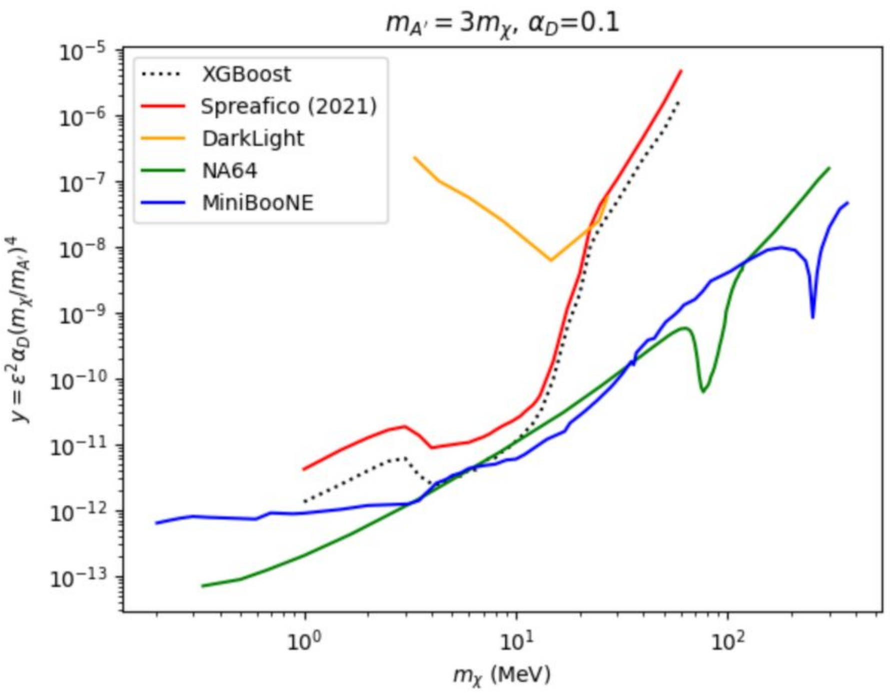


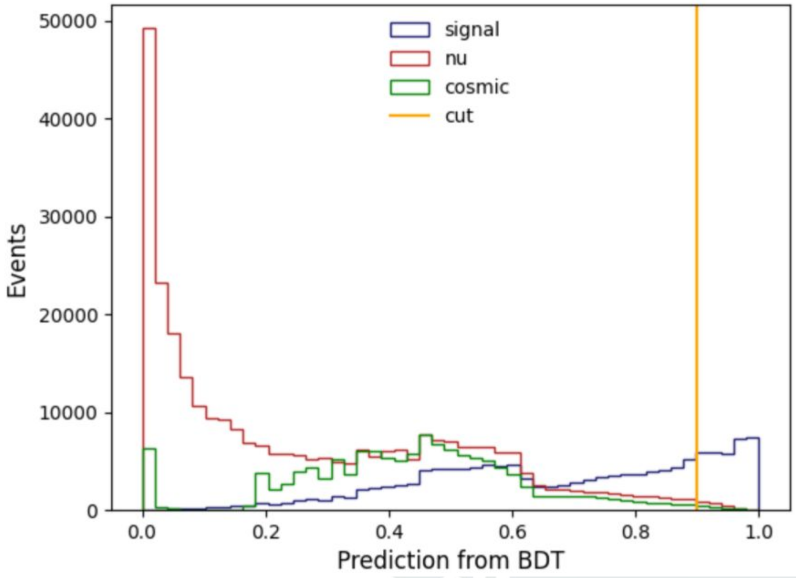
Input Features

1. Total energy
2. Shower direction
3. Energy outside the seed (i.e. outside the highest energy crystal)
4. x-y position of the seed
5. Multiplicity (number of crystals above the threshold)



Experimental reach improved by BDT cut

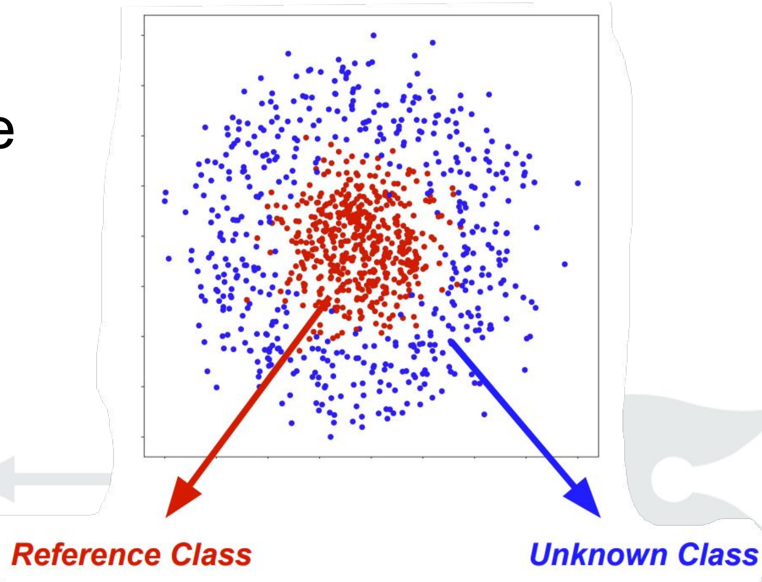




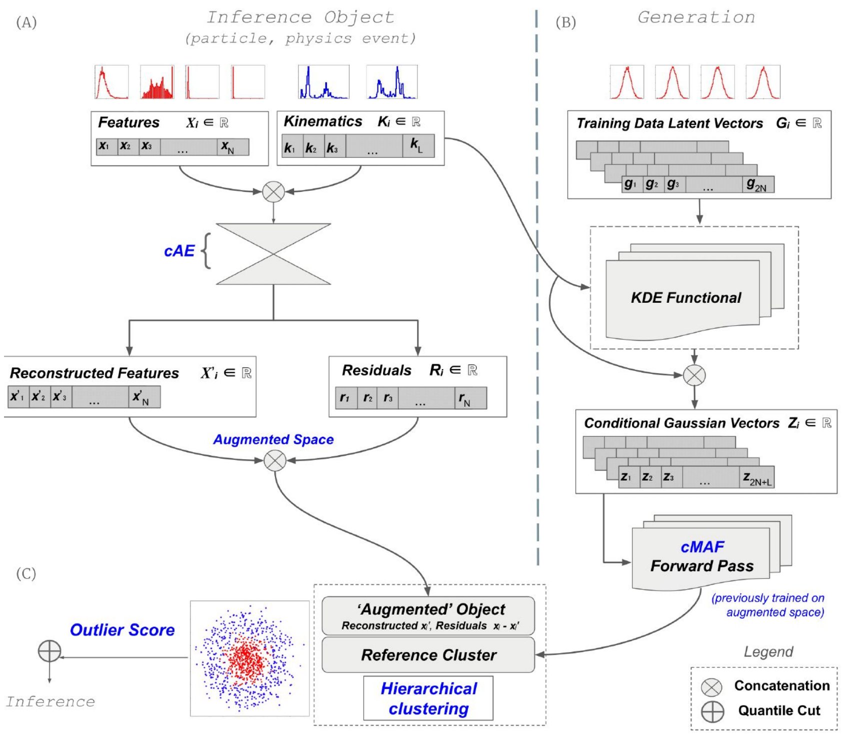
Flux + Mutability

- A conditional generative approach to One-Class Classification (OCC) and Anomaly Detection (AD)
- Can we use deep learning to separate two classes more efficiently than rectangular cuts?
- While remaining agnostic towards the unknown class?

t-SNE representation of
N-dimensional objects



Flux + Mutability: Architecture



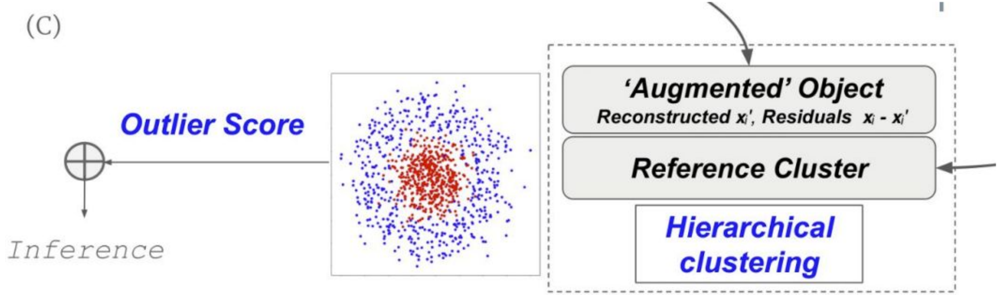
A. Inference Object fed through cAE
   a. Features ⊗ Kinematics
   b. Features ⊗ Residuals (**x' - x**)

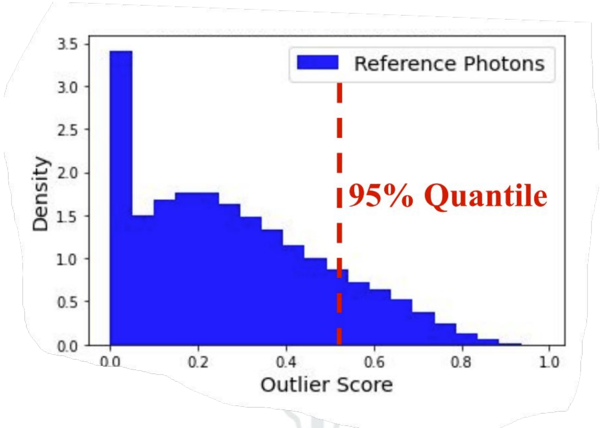
B. Continuous Conditional Generation
   a. Pre-fit KDE Objects in kinematic bins
   b. Map inference kinematics to KDE object
   c. Sample new Gaussian vectors from restricted domain
   d. Gaussian Vectors ⊗ Inference Kinematics
   e. **Conditionally generate reference population** via cMAF

C. Compare inference object to **reference population** via Hierarchical clustering and quantile cuts

https://arxiv.org/abs/2204.08609

HDBScan and Quantile Cuts

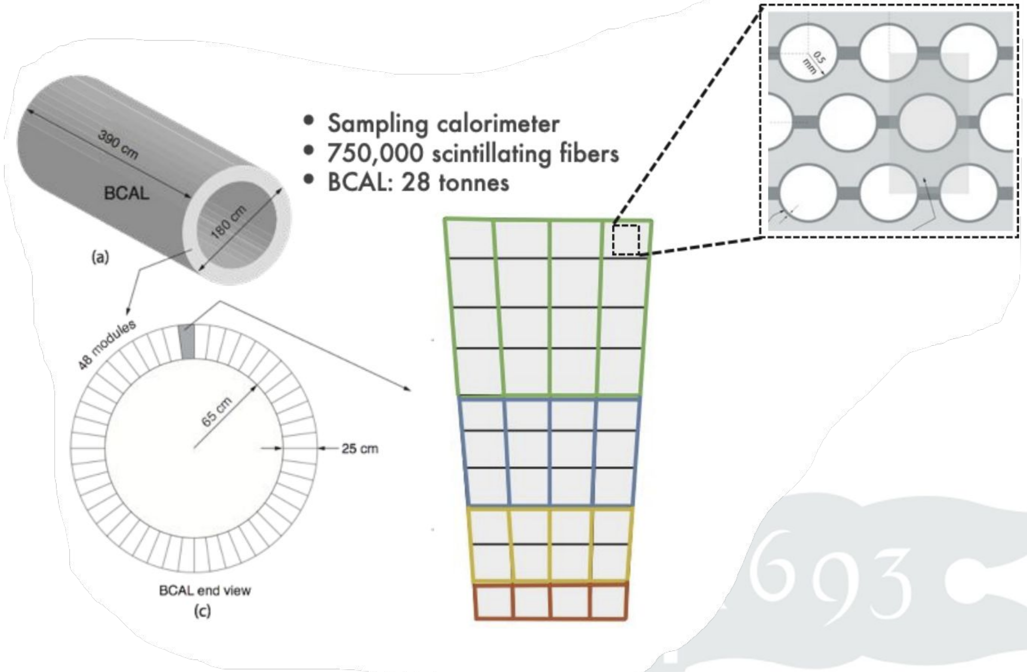




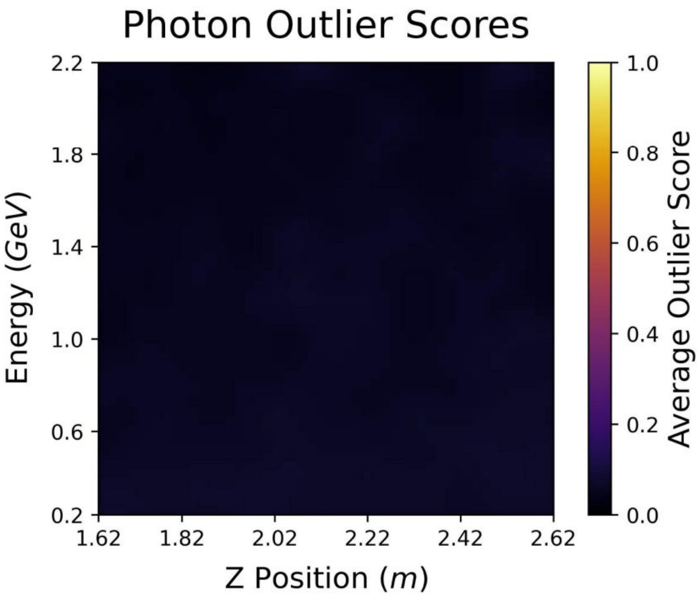
- Augment the inference particle into the **reference cluster** space
  - Two notions of membership: density-based & distance-based
- Combine the two PMFs and extract a probability of membership (P_{in})
- Define Outlier Score as complementary probability $P_{out} = 1 - P_{in}$
- Extract **reference population** outlier score corresponding to a desired quantile

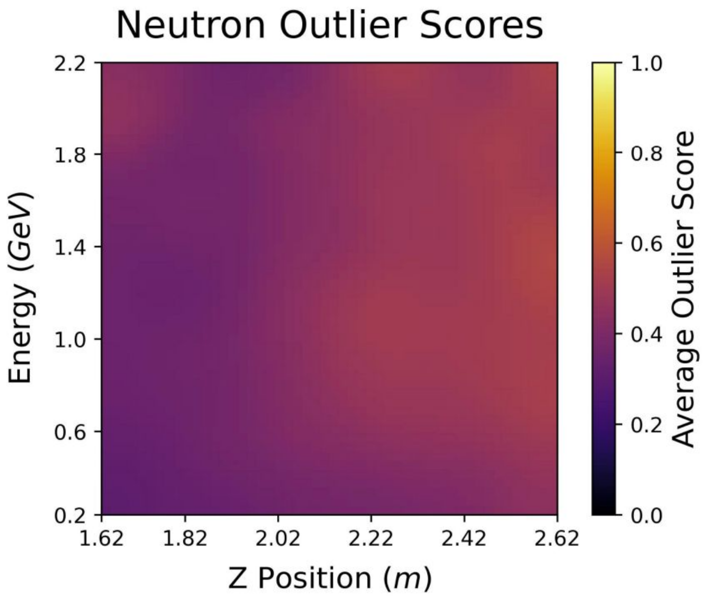
Case 1: γ/n Separation at GlueX (OCC)

- High confidence on one class
- Isolate highly active area of BCAL
- Reconstructed energy and z-position as kinematic conditions
- Simulated showers of photons (**inference**) and neutrons (**reference**)
- Strict preselection cuts
- Deploy fiducial cuts to extract only neutron showers which highly resemble photons
- 14 input features comprising of detector response variables



OCC: γ/n Separation at GlueX



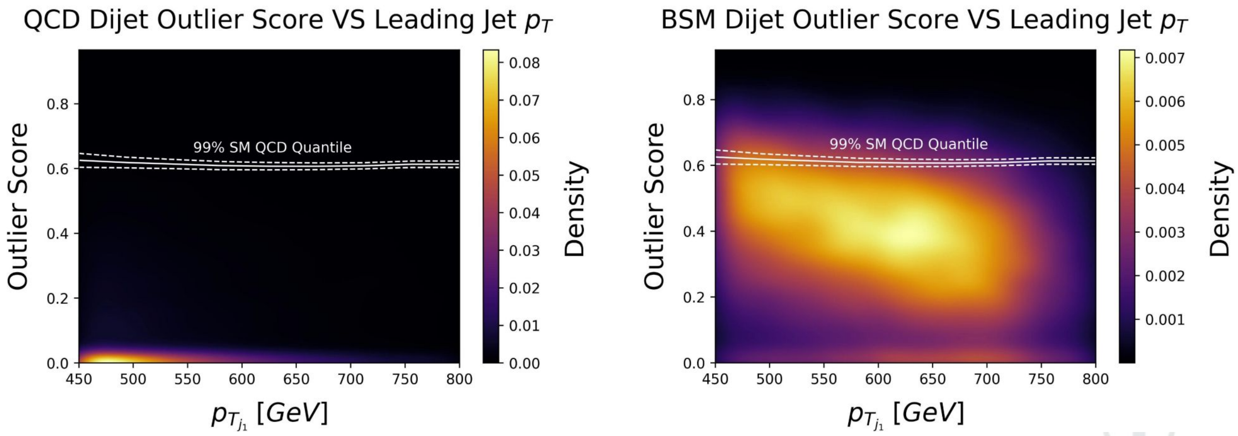


Quantile	True Positive Rate	True Negative Rate
68%	**68.15 ± 0.18 %**	**87.48 ± 0.13 %**
95%	**95.03 ± 0.08 %**	**52.70 ± 0.19 %**
99%	**98.96 ± 0.04 %**	**35.44 ± 0.18 %**

Case 2: BSM Dijet Separation at LHC (AD)

- Consider QCD dijet events as **reference**
- Isolate $Z' \rightarrow tt$ dijets as **unknown**
- Publicly available datasets generated via MADGRAPH and Pythia8 using the DELPHES framework for fast detector simulation
- Require leading jet transverse momenta 450 GeV < pT < 800 GeV and sub-leading jet pT > 200 GeV
- Consider leading jet pT as single kinematic condition
- 15 input features
  - Remaining 4 vector properties of the leading jet and n-subjettiness variables
  - Sub-leading jet 4 vector and n-subjettiness variables

Anomaly Detection: BSM Dijet Separation at LHC

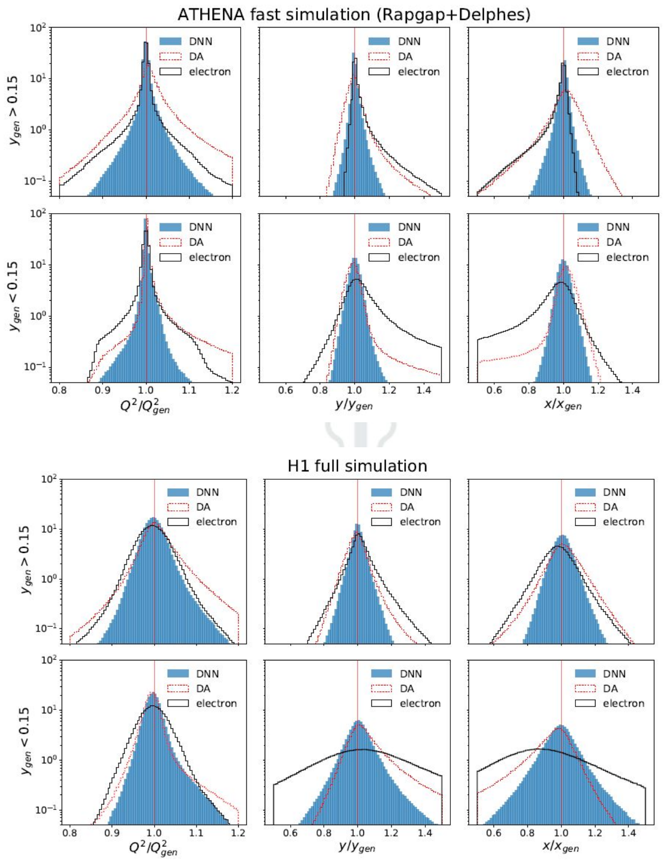


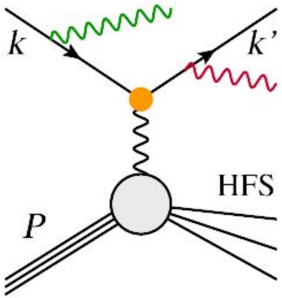
Quantile	True Positive Rate	True Negative Rate
68%	**68.48 ± 0.22 %**	**93.05 ± 0.06 %**
95%	**95.27 ± 0.10 %**	**43.07 ± 0.22 %**
99%	**99.04 ± 0.05 %**	**12.74 ± 0.15 %**
Fiducial cuts (99%)	**98.92 ± 0.05 %**	**2.35 ± 0.06 %**

	F+M	Fraser et al.	Cheng et al.
AUC	**0.891 ± 0.005**	0.87	0.89

Kinematical Reconstruction with DNN

- Arratia et al (2022) previously showed improved kinematical reconstruction of DIS variables using DNN over standard reconstruction techniques
- Exploited full kinematical information and accounting for the presence of QED radiation
- Did not consider event-level uncertainty quantification





$$s = (k+P)^2\,, \qquad Q^2 = -q^2\,,$$

$$y = \frac{q\cdot P}{k\cdot P}\,, \qquad \text{and} \qquad x = Q^2/(sy)$$

Event-Level Uncertainty Quantification (ELUQuant)

Total loss function is the sum of components

$$\mathcal{L}_{Tot.} = \mathcal{L}_{Reg.} + \alpha \mathcal{L}_{Phys.} + \beta \mathcal{L}_{MNF.}$$

Learn the posterior over the weights

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

Access epistemic (systematic) uncertainty through sampling MNF layers

Learn the regression transformation

aleatoric

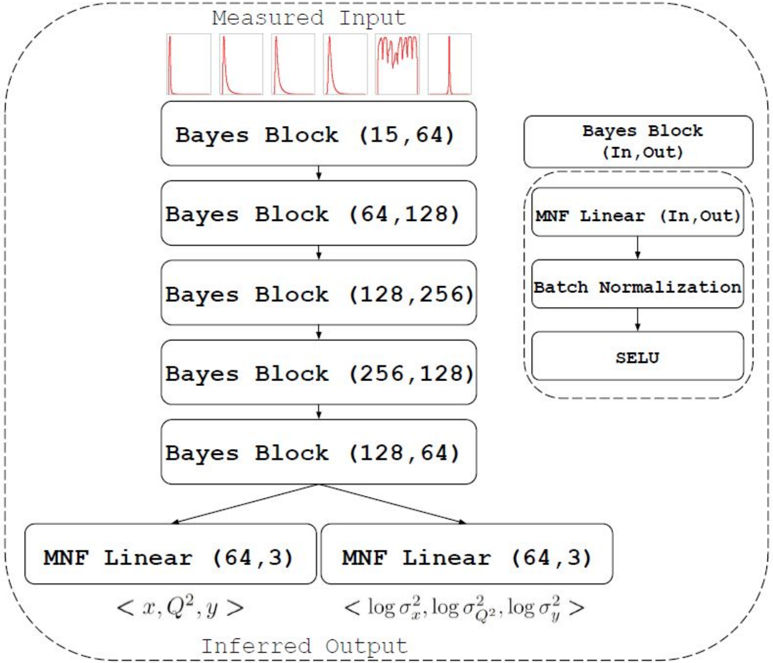
$$\mathcal{L}_{Reg.} = \frac{1}{N} \sum_i \sum_j \frac{1}{2} (e^{-\mathbf{s}_j} \| \mathbf{v}_j - \hat{\mathbf{v}}_j \|^2 + \mathbf{s}_j), \; \mathbf{s}_j = \log \boldsymbol{\sigma}_j^2$$

epistemic

Access aleatoric (statistical) as a function of regressed output

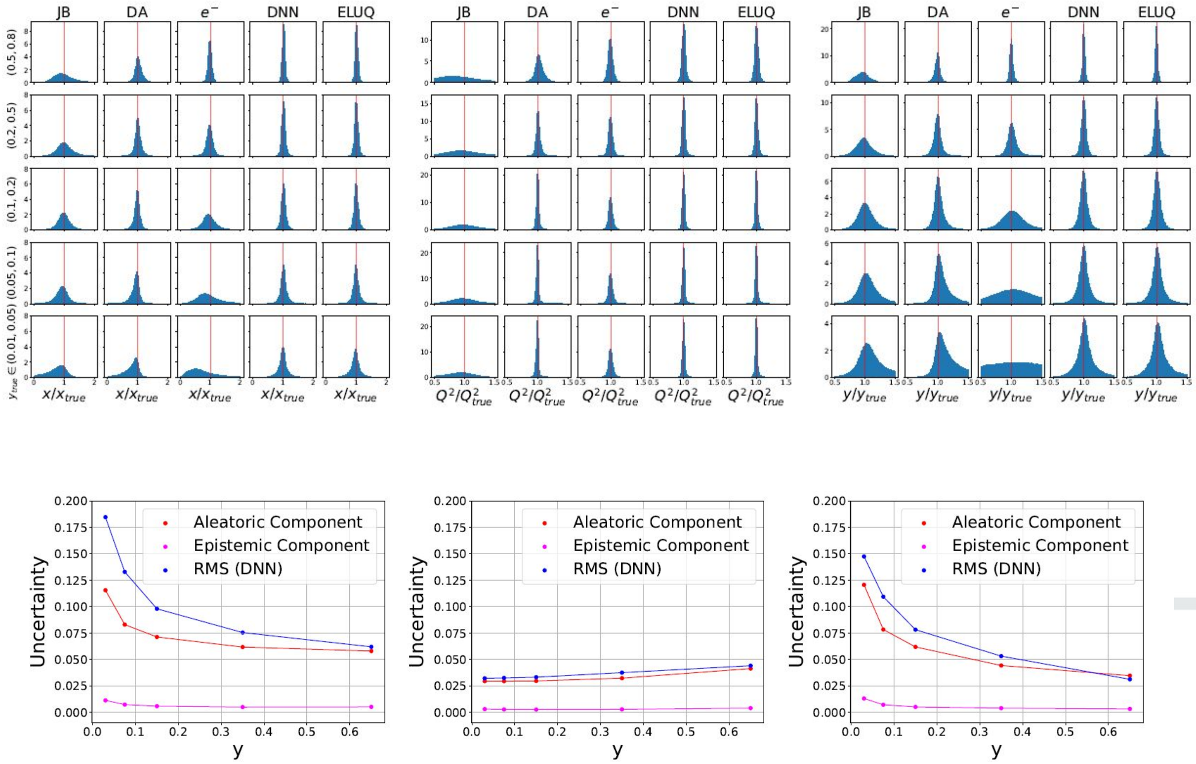
Constrain the physics

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



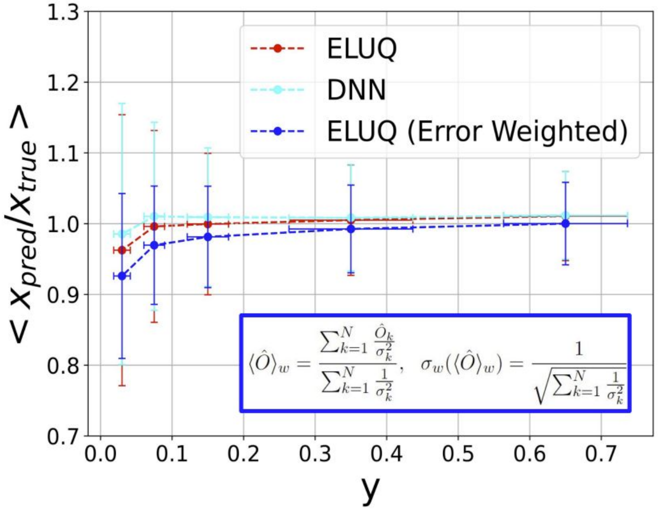
C. Fanelli, and J. Giroux. Machine Learning: Science and Technology 5.1 (2024): 015017.

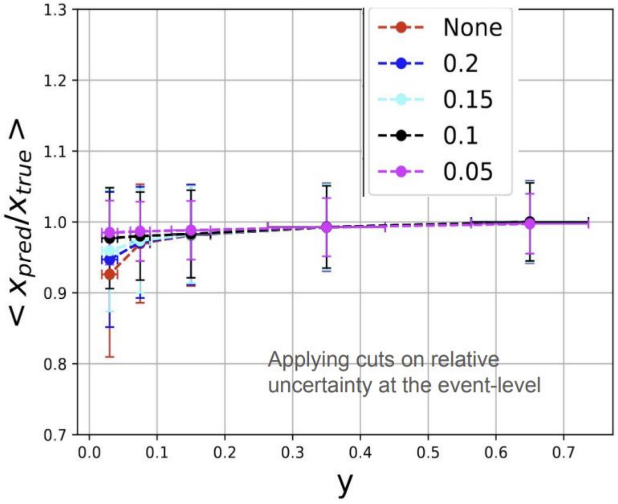
ELUQuant Performance Similar to DNN



- Reconstruction of NC DIS kinematics from H1 comparable to Arratia (2022)
- Comparing aleatoric/epistemic components from ELUQuant to RMS from DNN

Leveraging the Event-Level UQ

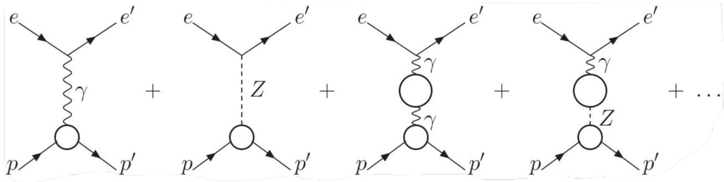


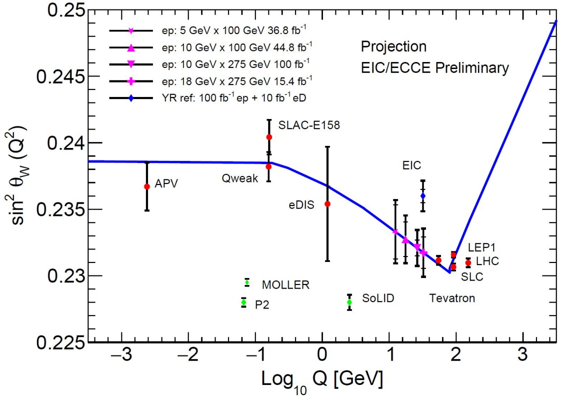


- The ability to remove events with large event-level uncertainty allows us to improve the ratio to truth
- Can be exploited for anomaly detection

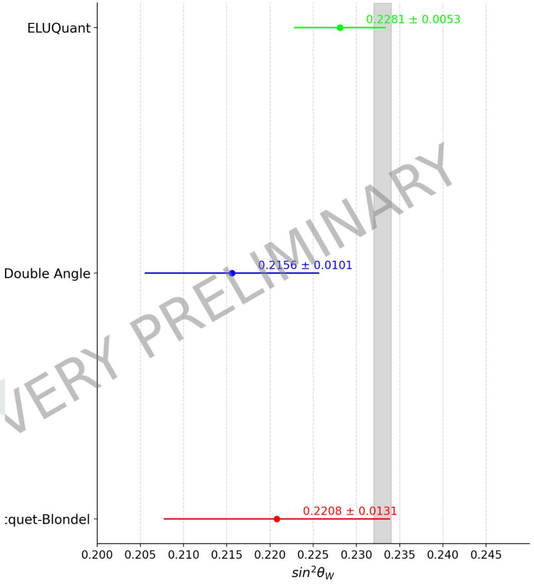
Precision Measurement of $\sin^2\theta_W$

- Deviations from the SM prediction of the running of the weak mixing angle would be evidence of BSM
- Currently in progress: measuring $\sin^2\theta_W$ at EIC kinematics





Boughezal et al. (2022) Phys. Rev. D 106, 016006



Summary

- Boosted Decision Trees demonstrates improved signal discrimination for BDX-MINI
- Flux + Mutability uses generative models in an unsupervised way to identity anomalies with respect to a reference class
- Event-level uncertainty quantification and kinematical reconstruction using BNN can allow for anomaly detection
- **Thank you!**

Backup Slides

Input Features for GlueX OCC

- **LayerM_E** $= \sum_i^N E_i$
  $M \in \{1,2,3,4\}$ is the layer number and E_i is the energy of the i^{th} reconstructed point in the layer.

- **Layer$M$bySumLayers_E** $= \frac{1}{E_{total}} \sum_i^N E_i$
  $M \in \{1,2,3,4\}$ is the layer number and E_i is the energy of the i^{th} reconstructed point in the layer.

- **Z Width** $= \sqrt{\frac{1}{E_{total}} \sum_i^N E_i(\Delta z_i)^2}$, $\Delta z_i = (z_i + T_z)$ - S_z
  E_i and z_i are the energy and z position of the i^{th} point in the shower.

- **R Width** $= \sqrt{\frac{1}{E_{total}} \sum_i^N E_i(\Delta r_i)^2}$, $\Delta r_i = (R$ - $r_i)$
  E_i and r_i are energy and radial position of the i^{th} point.

- **T Width** $= \sqrt{\frac{1}{E_{total}} \sum_i^N E_i(\Delta t_i)^2}$, $\Delta t_i = t_i - S_t$
  E_i and t_i are the energy and timing information of the i^{th} point.

- **θ Width** $= \sqrt{\frac{1}{E_{total}} \sum_i^N E_i(\Delta \theta_i)^2}$, $\Delta \theta_i = \theta_i - S_\theta$
  E_i and θ_i are the energy and polar angle (from the target center) of the i^{th} point.

- **ϕ Width** $= \sqrt{\frac{1}{E_{total}} \sum_i^N E_i(\Delta \phi_i)^2}$, $\Delta \phi_i = \phi_i - S_\phi$
  E_i and ϕ_i are the energy and azimuthal angle of the i^{th} point.

- **z Entry** $= (S_z - T_z)\frac{R}{S_r} + T_z$
  The position at which the particle hits the inner radius of the BCAL.

Summary of Basic DIS Kinematic 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]	$[E_0,\theta,\gamma]$	$\frac{\tan\frac{\gamma}{2}}{\tan\frac{\gamma}{2}+\tan\frac{\theta}{2}}$	$4E_0^2\cot^2\frac{\theta}{2}(1-y)$	$\frac{Q^2}{4E_0y}$
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}	$\frac{E^2\sin^2\theta}{1-y}$	$\frac{E(1+\cos\theta)}{2y}$
$E_0E\Sigma$	$[E_0,E,\Sigma]$	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]	$[E_0,E,E_h]$	$\frac{E-E_0}{(xE_p)-E_0}$	$4E_0y(xE_p)$	$E+E_h-E_0$
$E\Sigma T$	$[E,\Sigma,T]$	$\frac{\Sigma}{\Sigma+E\pm\sqrt{E^2+T^2}}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{2\Sigma}$
E_0ET	$[E_0,E,T]$	$\frac{2E_0-E\mp\sqrt{E^2-T^2}}{2E_0}$	$\frac{T^2}{1-y}$	$\frac{Q^2}{4E_0y}$
Sigma (Σ) [9]	$[E_0,E,\Sigma,\theta]$	$y_{\mathrm{I}\Sigma}$	$Q^2_{\mathrm{I}\Sigma}$	$\frac{Q^2}{4E_0y}$
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

Input features of ELUQuant

Utilized input features and H1 MC dataset of paper NIM-A 1025 (2022): 1661

- Define variables to characterize the strength of QED radiation

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

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

Slide borrowed from C. Fanelli

*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