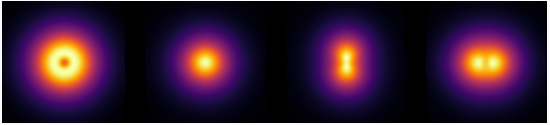
Femtoscale Imaging of Nuclei using Exascale Platforms



Ian Cloët

Argonne National Laboratory

Argonne NATIONAL LABORATORY

Jefferson Lab

VIRGINIA TECH.

NREL NATIONAL RENEWABLE ENERGY LABORATORY

OLD DOMINION UNIVERSITY

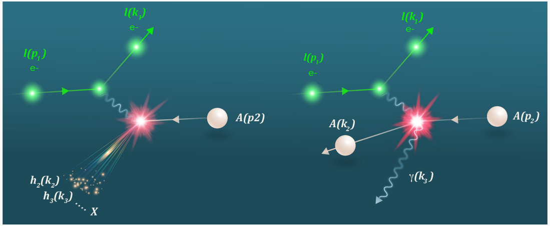
THE OHIO STATE UNIVERSITY

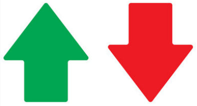
Workshop on Software Infrastructure
for Advanced Nuclear Physics Computing (SANPC 2024)
20—22 June 2024

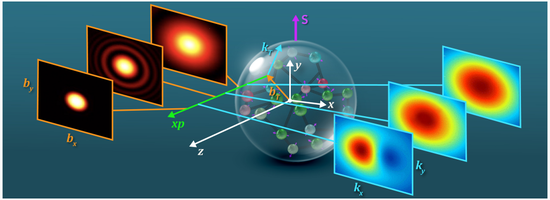
QuantOm Collaboration

Motivation for this SciDAC

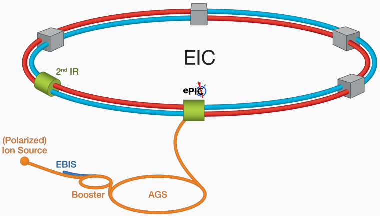
- **Goal:** To develop a new paradigm for the interface between theory and experiment for the analysis of data to infer femtoscale *images* of proton's and nuclei to reveal their 3D quark and gluon structure
- **Science Motivation:** To make optimal use of the petabytes of data from JLab, EIC, etc. to shed light on some of the key questions in nuclear physics:
  - What is the 3D confined motion and spatial distribution of quarks and gluons in nucleons and nuclei?
  - How do quark-gluon dynamics produce proton mass and thereby vast bulk of mass in the visible universe?
- To deliver these goals need a diverse team: domain experts in QCD theory and experiment, in collab with applied math, AI/ML, data science, and high-performance computing expertise

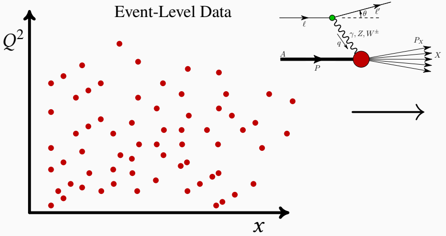


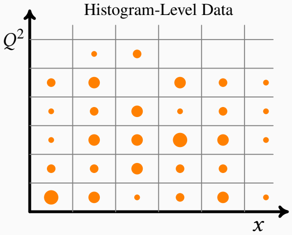




Current Paradigm



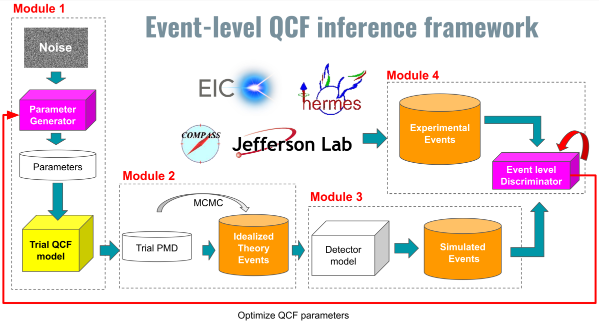


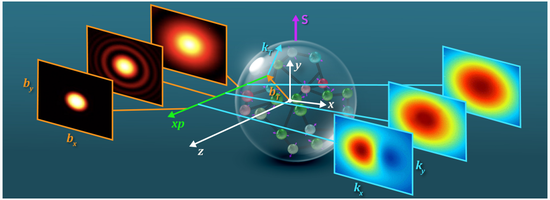


- Events are the basic *quantum of information* for our SciDAC – EIC will produce PBs of event data
- Current approach takes measured events and puts them in "bins" to obtain an average result over the phase space of the bin (histogram) – several shortcomings to this process, including:
  - Information is lost in this process
  - Limited resolution on events can cause bin migration effects
  - Detector effects need to be unfolded which is much more difficult the folding in the detector effects
- Histograming events works well enough in low dimensions with a sufficient amount of data, however, taking 3D pictures of the proton requires events in 5 or more dimensions
  - Loss of correlations/information in the data which could greatly impact the experimental program

An Event-Level Approach and Framework

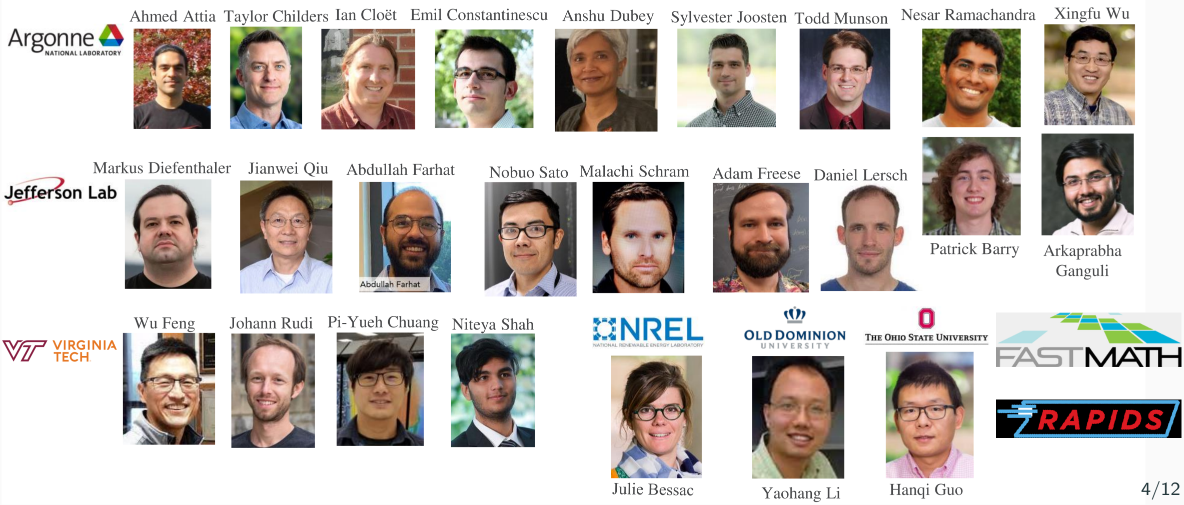
- In general, our approach is to represent the pictures of the proton in some manner e.g., as images, using piecewise polynomials, etc. that are governed by a large number of parameters (up to millions)
- Use these pictures, together with QCD theory, sampling, detector models, etc., to create set of simulated events
- We then use some approach to adjust the parameters until the simulated events and experimental events can be attributed to the same theory
- Workflow requires numerous methods from applied math, AI/ML, HPC, etc.
  - Statistical methods, Generative Adversarial Networks, event-level loss functions, distributed learning, . . .



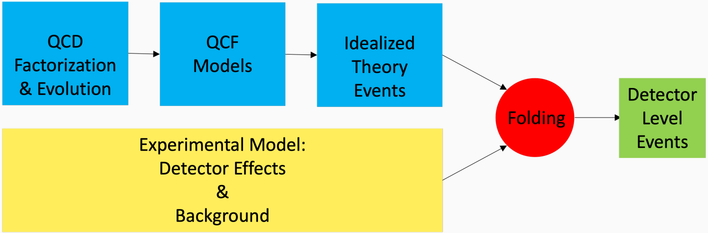


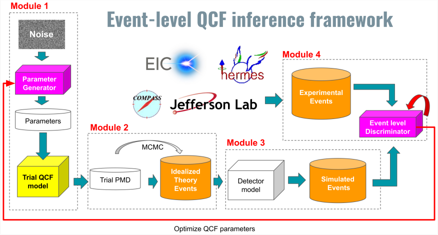
QuantOm Collaboration

QuantOm (QUAntum chromodynamics Nuclear TOMography) **Collaboration** is the team that will deliver the *"Femtoscale Imaging of Nuclei using Exascale Platforms"* SciDAC Project



Unifying Theory and Experiment via Folding

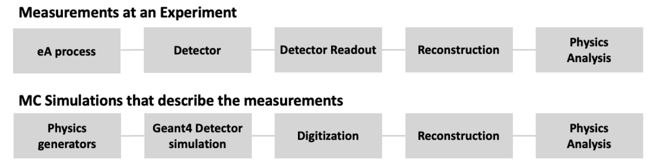




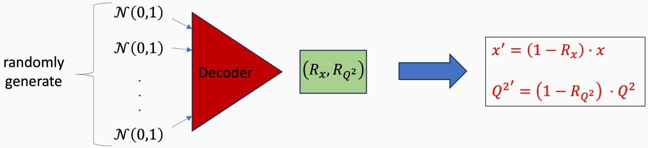
- Theory and Experiment usually meet at the differential cross-section level
  - Requires the *unfolding* of detector effects, backgrounds, etc.
- Folding in detector effects, backgrounds, etc. much more robust
  - Folding is not an invertible transformation, so reduces systematic uncertainties associated with unfolding
- Folding enables theory and experiment to be treated in an equal and unified manner, and variations in the theory can be much more rigorously studied

Experimental Modeling

- To develop an AI/ML enabled QuantOm workflow for event-level analysis need a differentiable detector module
  - Need to develop surrogate models for detectors



- Developed an event-level approach to model experimental effects from detailed simulations of the experiment, including background, e.g., Variational Autoencoder (VAE)



- Found that VAE demonstrates better performance over Deep Neural Networks for the cases studied

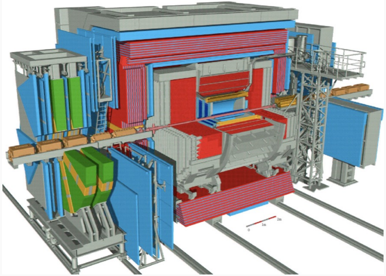
ZEUS Example

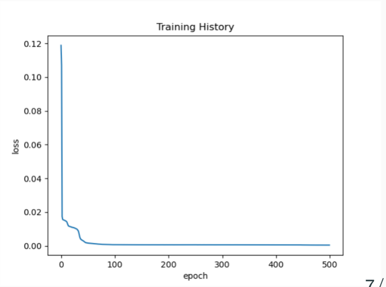
- Selected inclusive DIS events from detailed simulations of the ZEUS experiment at HERA.

- Used electron method for (x, Q^2) reconstruction

- AE detector surrogate specifications:
  - Encoder hidden layers and units: [50,50,50,100,100]
  - Decoder Hidden layers and units: [100,100,50,50,50]
  - Latent Dimension 128, RELU activation function

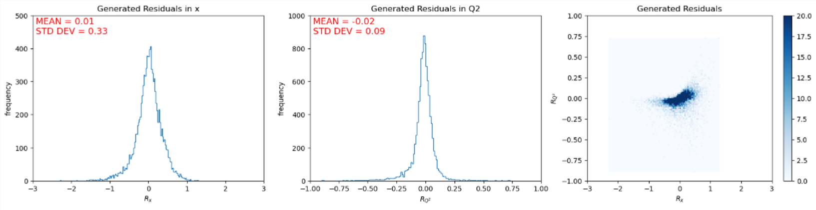
- Training: 20k events, 80/20 train/test split, outliers removed



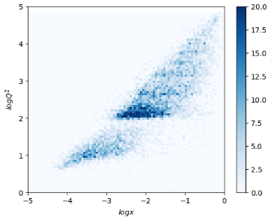


ZEUS Example: Residual and (x, Q^2) Distributions

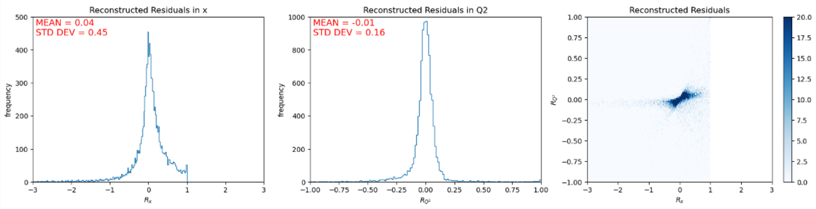
VAE Detector Surrogate



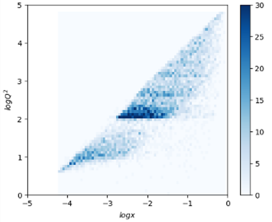
VAE Detector Surrogate



ZEUS Simulation (electron method only)



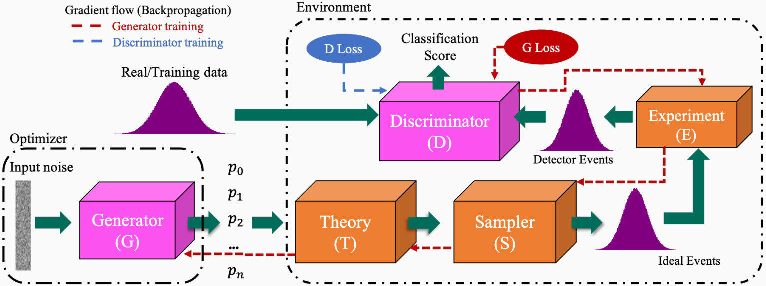
Ground Truth



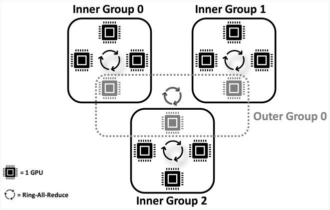
Developed a detector surrogate and training procedure to model various eA experiments

Scaling QuantOm Workflow using GANs

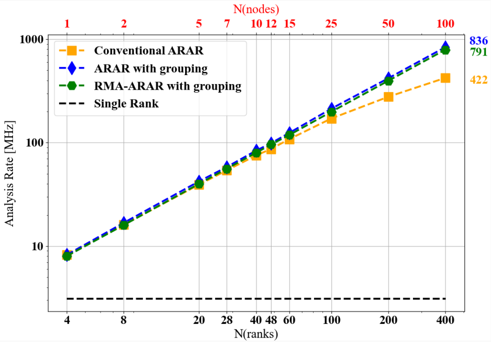
- **Goal:** Want to run workflow across multiple GPUs
  - Handle data size
  - Distribute computational load, e.g. sampler module

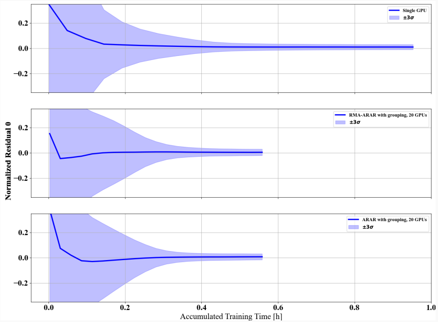


- **Approach:** Asynchronous Ring All-Reduce (ARAR)
  - Data is shared across GPUs
  - Each GPU trains discriminator locally
  - Generator gradients are transferred between GPUs
  - GPUs are bundled into groups
  - Enabled usage of Remote Memory access (RMA-ARAR)



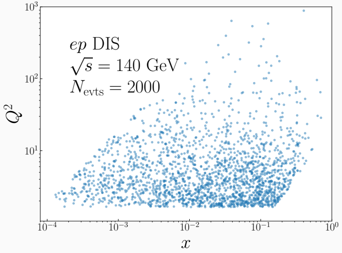
Results on Distributed Learning Approaches using Polaris

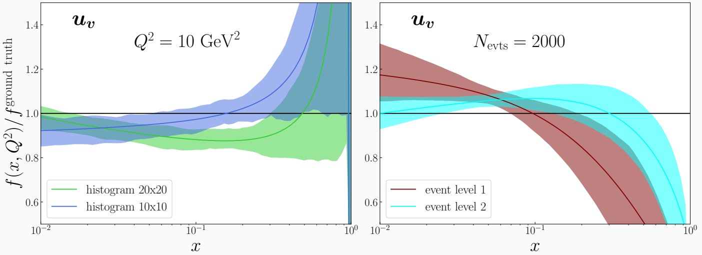




- Test distributed learning approaches on for event-level PDF analysis using Polaris
  - Used ensemble technique to determine convergence quality
  - ARAR/RMA-ARAR with grouping allow for earlier convergence
  - Observe weak nearly linear scaling
- Method will be further tested and developed on Aurora at Argonne
- Demonstration that we can develop an event-level analysis framework at scale

First Event-Level Analysis for DIS





- Generated 2000 simulated DIS events, which were sampled from a differential cross-section generated from ground-truth PDFs
- Analyzed these events using the traditional histogram approach and two event-level approaches
  - The histogram approach and event-level 2 perform about the same
  - However, a different binning produces different results
- Event-level approach removes a key systematic uncertainty: *How does different binning schemes impact the extraction of quantum correlation functions?*

Conclusion and Outlook

- Performing an analysis of scattering data at the event level requires significantly more upfront computing resources
  - Real-world deployment will require exascale resources, however, this will compress the time scales from measurement to discovery, which is often years to up to more than a decade
  - Real-time data analysis becomes a possibility, and when combined with autonomous optimization, could lead to autonomous discovery at facilities like the EIC
- The success of this SciDAC project should represent paradigm shift in the way science is conducted at high-energy accelerator facilities
  - Will remove the artificial wall between theory and experiment and seamlessly connect them into a single analysis framework