AI4EIC Hackathon: PID with the ePIC dRICH

https://eic.ai/

AI4EIC hackathon ’22 documentation:
CF, J Giroux, D McSpadden, K Rajput, K Suresh
https://doi.org/10.5281/zenodo.7197023

5/9/2023

C. Fanelli for the AI4EIC
(credits: J. Giroux, K. Suresh)
Electron Ion Collider

A precision machine to study the “glue” that binds us all

polarized electron - polarized protons/ions

Total estimated cost ~ $1.6-2.6B

fundamental questions

How does the mass of the nucleon arise?

How does the spin of the nucleon arise?

What are the emergent properties of dense systems of gluons?

World-wide interest, thousands of users and hundreds of institutions already involved
AI/ML and the EIC schedule

- EIC will integrate AI from beginning through all project phases.
- EIC can be the first experiment for QCD to be designed with the assistance of AI.
- EIC will take advantage of intelligent decisions in all aspects of data processing.
- Streaming Readout (SRO) will enable AI/ML integration for seamless offline and online analysis convergence.

See also other CHEP2023 talks, e.g., D. Lawrence (EIC SW Overview) and CF (AI/ML for ePIC).

What is AI4EIC?

- AI will be an integral part of the EIC software and to work in this direction, a dedicated AI Working Group (AI4EIC) has been established within the EICUG (https://www.eicug.org).
- AI4EIC will serve as an entry point to AI applications and organizes workshops, tutorials, hackathons, Kaggle-like challenges, etc.
- AI4EIC is a vibrant community built through multiple events organized during the last two years and collaborates with the recently formed ePIC Collaboration.
- Workshops serve as a body of essential knowledge for AI4EIC, and produce proceedings, annual report, journal special issues. Educational activities and outreach are aimed at disseminating AI in the EIC community.
  - AI4EIC has already organized 2 workshops (200+ participants), several tutorials (https://eic.ai/workshops), monthly meetings (https://eic.ai/events), and an international hackathon (https://eic.ai/hackathons).
  - Hackathon events are built around specific challenges for EIC and help identify strategies, architectures and algorithms that will benefit the EIC physics program.
  - Additionally, AI4EIC is committed to establishing educational events (e.g., schools) designed to enhance AI and ML proficiency within the EIC community. For more information, https://eic.ai/community and https://eic.ai/ai-ml-references.
AI4EIC 2022 Workshop & Hackathon

https://eic.ai

Workshop:
- Total of 220 registered participants (also last year, >200!)
  - Very good attendance in person!
- 6 sessions (15 conveners, 40+ speakers)
  - Design
  - Theory/Exp connections (morning + afternoon sessions)
  - Recon & PID
  - Infrastructure (+ Panel Discussion)
  - Streaming

Tutorials:
- MOBO (Meta AI)
- OmniFold
- MLFlow
- GNN

Discuss from this workshop contributed to LRP WP

Hackathon:
- (10 teams from North, South America, Asia, Europe)

https://eic.ai/community

https://doi.org/10.5281/zenodo.7197023
Hackathon: Brings together communities with diverse skill sets, such as CS, DS, and physics, fostering collaboration and launching projects that align with the core objectives of the EIC.

Organized on Fri October 14, 2022 after AI4EIC workshop

**Background/Motivation:** Cherenkov detectors constitute the backbone of PID at EIC

**Challenges:**
- Simulations are typically compute-expensive
- Reconstruction relies on pattern recognition of noisy, sparse rings of photons (with complicated topologies for the DIRC);
- Events with close tracks that can easily produce mis-identification

**ML/DL ideal for a holistic event-level reconstruction and can also provide computational speed-up**

**Goal:** Can we begin to leverage ML/DL for PID based on low-level features from imaging Cherenkov detectors? This can be a first step towards ML/DL applications for PID with Cherenkov detectors.

**“Targeted” detector:** dual-RICH in ePIC is instrumental to guarantee an efficient identification of hadrons in the hadron-endcap covering a large range in momentum

https://doi.org/10.5281/zenodo.7197023
Two radiators with different refractive indices to provide PID with large momentum coverage

π/K classification is tackled, being the most challenging task chose as working point $P \sim 15$ GeV where both radiators contribute to PID
3 Problems of Increasing Complexity

- **Problem 1:** Fixed momentum and direction (threshold accuracy 94%)

- **Problem 2:** Larger phase-space (threshold accuracy 86%)

- **Problem 3:** Larger phase-space + noise (threshold accuracy 80%)

\[
\text{score} = 50.0 + 50.0 \times \frac{(\text{ACC} - \text{THRESHOLD})}{100.0 - \text{THRESHOLD}}
\]

\[
\text{ACC} = \frac{\sum_{i=1}^{N} |y_{\text{pred}} - y_{\text{true}}|}{N}
\]
### Data

The maximum number of detected optical photons is set to 60. Average detector photons are in the range of 30 - 40 with noise.

Dataset:

- Publicly available on Zenodo
- Data format csv and hdf5 — flat data, accessible to colleagues from various fields

---

**Event information (4 Cols)**

<table>
<thead>
<tr>
<th>eventID</th>
<th>PID</th>
<th>P [GeV/c]</th>
<th>θ [deg]</th>
<th>Φ [deg]</th>
<th>X0 [mm]</th>
<th>X59</th>
<th>Y0</th>
<th>Y59</th>
<th>Z0</th>
<th>Z59</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>211</td>
<td>15.10</td>
<td>15.12</td>
<td>0.45</td>
<td>1570</td>
<td>1412</td>
<td>1660</td>
<td>55.3</td>
<td>-2.5</td>
<td>23.4</td>
</tr>
<tr>
<td>1</td>
<td>321</td>
<td>16.76</td>
<td>15.48</td>
<td>1.23</td>
<td>1514</td>
<td>1525</td>
<td>1456</td>
<td>12.4</td>
<td>6.6</td>
<td>10.2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>N</td>
<td>211</td>
<td>17.89</td>
<td>15.67</td>
<td>1.26</td>
<td>1613</td>
<td>1645</td>
<td>1564</td>
<td>12.5</td>
<td>-7.5</td>
<td>151.2</td>
</tr>
</tbody>
</table>

The maximum number of detected optical photons is set to 60. Average detector photons are in the range of 30 - 40 with noise.

---

Used **ePIC software stack** for data generation [1] [2]


Dataset:

- Publicly available on Zenodo
- Data format csv and hdf5 — flat data, accessible to colleagues from various fields
AWS Resources

- Each team was allocated an AWS instance
  - 4x Nvidia A10G 24GB GPUs
  - 48 vCPUs with total RAM allocation of 192GB and ~4TB storage
- W&M prebuilt Conda environments containing the most common ML packages
  - PyTorch
  - TensorFlow
- Full technical assistance provided throughout the event by W&M Research Computing Group
  - Accessible to both remote and in-person participants
Efficient Training Workflows

- Multiple users introduce the need for efficient resource management within teams
  - Minimize memory usage during model training schemes
  - Provide full training and evaluation workflow - focus on model development
  - Instructions provided for specific GPU allocation within a group - capable of utilizing distributed training techniques across multiple GPUs

Provided Workflow

- Data-loading
- Network Instantiation
- Training Process
- Network Evaluation

User #0 Network

GPU #0

Documentation Link: https://doi.org/10.5281/zenodo.7197023
Submission & Leaderboard

- Leaderboard was made with a submission portal. Participants submit solutions to be graded in real time.
- Participants continued to submit solutions to improve their scores even after the hackathon.
- “JINR” team won the Hackathon - Used CatBoostClassifier
- “Jets” team secured 2nd place - Used Convolution Neural Nets (CNN)
- The winners are invited to give a talk at the next AI4EIC workshop
Conclusions

- **Hackathon useful to unveil the ML/DL potential for EIC:** During the hackathon, innovative ML/DL-based solutions outperformed traditional cut-based methods, signaling a promising initial step towards using ML/DL for PID with the dual-RICH detector.

- **Community building and collaboration:** The event fostered community building by gathering diverse individuals in collaboration, enhancing shared learning and strengthening relationships. This set a strong foundation for future joint AI + physics endeavors.

- **Igniting young minds:** The hackathon was a valuable and educational experience for the younger participants, who expressed great excitement about their involvement, nurturing their creativity and passion for AI in scientific research.

- **A launchpad for future innovations:** The hackathon kick-started the utilization of ML/DL for the ePIC dRICH, with participant enthusiasm underlining the potential for continued innovation and refinement of these approaches.

- **Forthcoming 3rd AI4EIC workshop + hackathon:** winter 2023, announcement will be made soon.
  - People interested in AI4EIC activities can email support@eic.ai.
  - More info on https://eic.ai/how-to-join
Backup
dRICH: ante-proposal

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

- Continuous momentum coverage.
- Simple geometry and optics, cost effective.
- Legacy design from INFN, see [EICUG2017](#).

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

**aerogel (4 cm, n(400 nm): 1.02)**
**+ 3 mm acrylic filter**
**+ gas (1.6 m, n(C$_2$F$_6$): 1.0008)**
**dRICH reconstruction**

- **Indirect Ray Tracing (IRT)**
  - The basic idea is that, given tracking information and RICH PMT hits, the Cherenkov-photon emission angle can be reconstructed.
  - The distribution of observed photon angles is compared to the expected angle for each particle type and the most likely particle type is determined.
  - Fast, non computationally intensive. Lowest accuracy compared to other methods in this slide.

- **Direct Ray Tracing (DRT)**
  - Simulates a PMT hit pattern based on the track kinematics and particle hypothesis
  - Construct likelihood by comparing “PDF” to the observed hit pattern

- **Event-level algorithm (EVT)**
  - Motivation: two close tracks can produce mis-identification
  - Builds upon DRT. Improvement by looking at each event as a whole rather than individual tracks
    - $\rightarrow$ sum over all tracks in the event

---

**Data Generation**

**eic shell**

**Generating \( \pi/K \) tracks**

- **Setup eic shell.**
- **Setup geometry using dd4hep framework.**

- Use `ndsim` to simulate tracks.

- The simulation does not include any detectors beyond the dRICH and in the electron endcap region.

**Afterburner analysis**

- **Juggler** was used to analyze the outputs from simulation.
- **eic-recon** was also explored. eic-recon was still under active development in September 2022.

**Apply quantum efficiency**

- To simulate real data, a quantum efficiency is applied to each of the optical photons based on its energy as in figure.

**Add Noise and save output**

- If needed, added random noise (Poisson with mean 5). Typical rates expected are between 5 optical photons per event.
- Save the output in the desired format. `csv` and HDF5.
Due to the shape of the ring in dRICH between the pions and kaons, one could place a cut on the total number of optical photons to distinguish between $\pi/K$ in a given kinematic setting. This works poorly over a wide kinematic setting, though.

For judging purposes, a threshold accuracy is calculated based on this cut.

- Q1 thres = 94%
- Q2 thres = 86%
- Q2 thres = 80%
Leaderboard Application

Evaluate.py

If (logged in)
True
False

/home

Submit Form
/submit

/sqlalchemy
Team info
User info
Submitted Solutions

Update solutions and retrieve score

/leaderboard

/login
Authenticate user

/submit

/allteams
This has been used during the workshop to collect questions and replies.

Conveners monitored discussion/questions in the live document.

Total of 26 pages.
A detailed survey was sent [https://forms.gle/6LADKTGaX7DeTVE46](https://forms.gle/6LADKTGaX7DeTVE46)

We want to learn more about our community, and we asked for feedback on what the needs and interests are, and what potential opportunities

Feedback and key-words:

- continual-learning
- kinematics
- data-management
- reproducibility
- cherenkov
- get-involved
- tutorials
- dis
- sessions
- rich-reconstruction
- advances-industry
- streaming-readout
- graph-neural
- networks

We organized our monthly meetings and workshop and [hackathon](https://indico.bnl.gov/event/16586/page/435-hackathon) taking into account this feedback

We have [tutorial](https://indico.bnl.gov/event/16586/page/426-tutorials) sessions every day of the workshop

For more details on the survey, see [https://indico.bnl.gov/event/15636/](https://indico.bnl.gov/event/15636/)
AI4EIC Tutorials

Multi-objective Optimization

F. Torales Acosta (LBNL)

Unfolding

V. Mikuni (NERSC)

MLflow — ML lifecycle

K. Rajput
JLab/DS

flow

Tracking
Record and query experiments; code, data, config, results

Projects
Packaging format for reproducible runs on any platform

Models
General format for sending models to diverse deploy tools

Graph Neural Network

Y. (Ray) Ren (BNL)
**AI4EIC Talks**

**Design**
- M. Balandat (Meta) Multi-objective Optimization Tutorial
- K. Suresh (Regina) Adaptive Experimentation in EIC
- B. Nachman (LBNL) AI-driven detector design
- E. Fol (CERN) ML Application for beam optics control in the LHC
- T. Satogata AI/ML overview for accelerator design activities

**The/Exp (morn.)**
- S. Liuti (UVA) ML for QCD analysis - 3D imaging
- T. Menzo (U. Cincinnati) Modeling Hadronization Using ML and the Lund String Model
- S. Andzrej (Jagiellonian U.) Modeling Hadronization Using ML and the Cluster Model
- A. Hiller Blin (U. Regensburg) A(I)DAPT
- N. Sato (JLab) Femtoscale Imaging of Nuclei using ML and Exascale Platforms
- B. Nachman (LBNL) Differentiable Simulations

**The/Exp (aftern.)**
- D. Shih (Rutgers) Fast Detector Simulations with ML
- W. Phelps (CNU/JLab) ML in Spectroscopy and Partial Wave Analysis
- C. Pecar (Duke) Reconstructing DIS and SIDIS properties
- A. Butter (LPNHE CNRS) Ideas for ML based unfolding
- F. Torrales Acosta (LBNL) and V. Mikuni (NERSC) Unfolding Tutorial

**Reco/PID**
- D. Whiteson (UC Irvine) Interpretable Networks for Identifying Leptons
- R. Kunnawalkam Elayavalli (Vanderbilt U.) Tagging heavy flavor jets @ RHIC
- W. Phelps (CNU/JLab) Muon identification with Deep Learning at EIC
- C. Allaire (IJC-Lab) Machine Learning in ACTS
- C. Peng (ANL) ML-PID with measured shower profiles from calorimetry
- M. McEneaney (Duke) A event tagging at CLAS12
- N. Branson (Messiah U.), ML for calorimetry
- J. Giroux (Regina), Data-driven learning : Flux + Mutability

**Infrastructure**
- D. McSpadden (JLab/DS) MLFlow tutorial
- S. Volkova (PNL) Foundation Model Infrastructure
- F. Liu (ORNL) AI/ML hardware co-design
- N. Tran (FNAL) Machine Learning with FPGA
- B. Joo (ORNL) AI/ML with HPC
- J. Huang (BNL), T. Miceli (FNAL), M. Williams (MIT), Panel Discussion

**SRO**
- M. Diefenthaler (JLab) INDRA-ASTRA
- S. Furlanov (JLab) FastML for FPGA
- R. Ammendola (Tor Vergata, Rome) AI for streaming readout: an architectural perspective
- J. Huang (BNL) AI-based data reduction for streaming DAQ
- M. Bondi (INFN/Catania) SRO for next generation electron scattering experiments
- C. Dean (MIT) ML for Heavy Flavor Identification
- T. Britton (JLab) AI for Experimental Controls

Talks: [https://indico.bnl.gov/event/16586/contributions/speakers](https://indico.bnl.gov/event/16586/contributions/speakers)
The whole idea of the AI-assisted design is that of determining trade-off optimal solutions in a multidimensional design driven by multiple objectives.

For an interactive visualization: https://ai4eicdetopt.pythonanywhere.com
The interactive visualization employs several Python and JavaScript libraries/packages to visualize the results from the optimization:

- Plotly-dash - click&play interface; interactive navigation; expanded dashboard
- JSRoot — JSRoot project allows reading binary and JSON ROOT files in JavaScript; drawing of different ROOT classes in web browsers; reading TTree data; using node.js used to visualize the detector geometry which is stored in GDML format
- Pandas: read source data (Pareto front solution)
- MySQL DB: most convenient DB that is used alongside Flask based applications. Meta-data like location of Geometry files, Location of parameters file are stored in the form of a database

https://ai4eicdetopt.pythonanywhere.com