Artificial Intelligence + Machine Learning for

5/11/2023

Cristiano Fanelli (and Authors in the slides) for ePIC

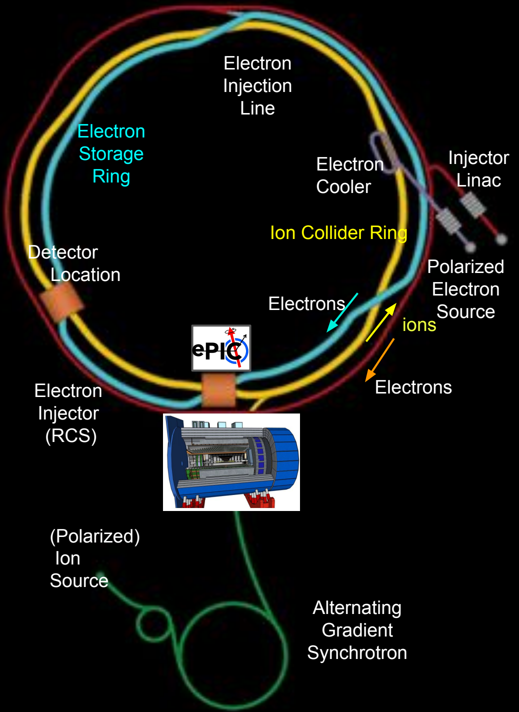
26TH INTERNATIONAL CONFERENCE ON COMPUTING IN HIGH ENERGY & NUCLEAR PHYSICS

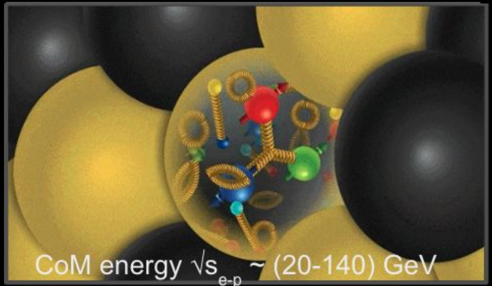
Jefferson Lab
Exploring the Nature of Matter

ePIC detector at EIC

EIC @BNL:
a precision machine to study the "glue" that binds us all

polarized electron - polarized protons/ions



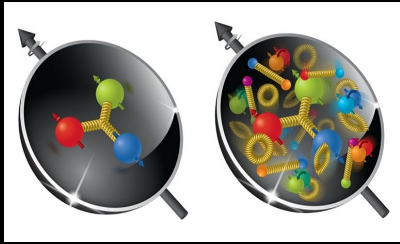


World-wide interest in EIC, thousands of users and hundreds of institutions involved

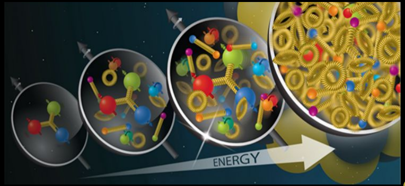
Total estimated cost ~ $1.6-2.6B

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?



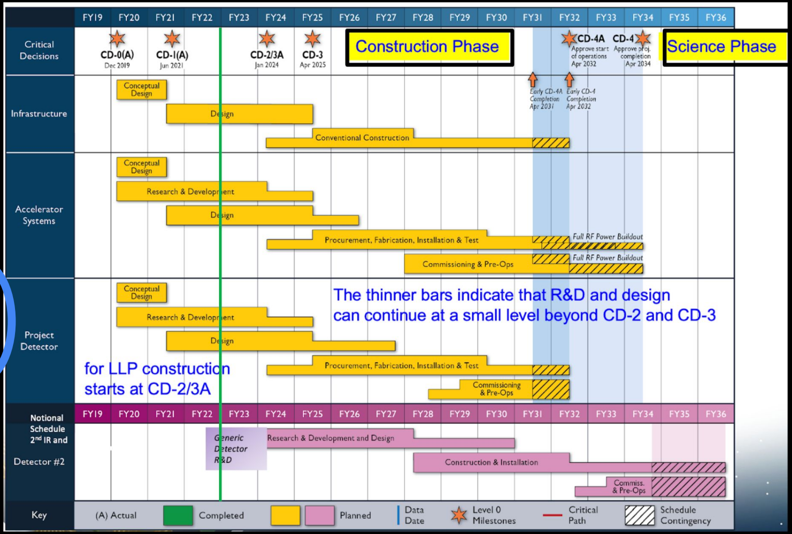
AI/ML at ePIC

(see also talk by [D. Lawrence, EIC SW Overview](#))

- **ePIC is one of the first experiments whose design phase is unfolding during the AI revolution**
- **AI/ML is also anticipated to contribute to multiple aspects of ePIC for near real-time analysis, autonomous calibration, alignments etc.**

R. Ent, AI4EIC workshop, October 2022

Detector-2 ePIC accel



https://eic.ai/workshops

Some AI/ML references for EIC (collaborative efforts):

R. Abdul Khalek et al arXiv:2103.05419, Yellow Report, Chap 11

AI-optimized detector design for the future EIC: the dual-radiator RICH case
- E. Cisbani *et al JINST* 15 P05009 (2020)

AI-assisted Optimization of the ECCE Tracking System at the EIC - C. Fanelli et al, *arXiv:2205.09185*, NIM-A 1047 (2023): 167748.

Ongoing activities in ePIC

AI-assisted design, Fast ML for SRO, ML/DL for PID (e.g., muon-ID, low photons in ZDC, etc), DIS event-level analysis with DL, etc.

AI4EIC Proceedings

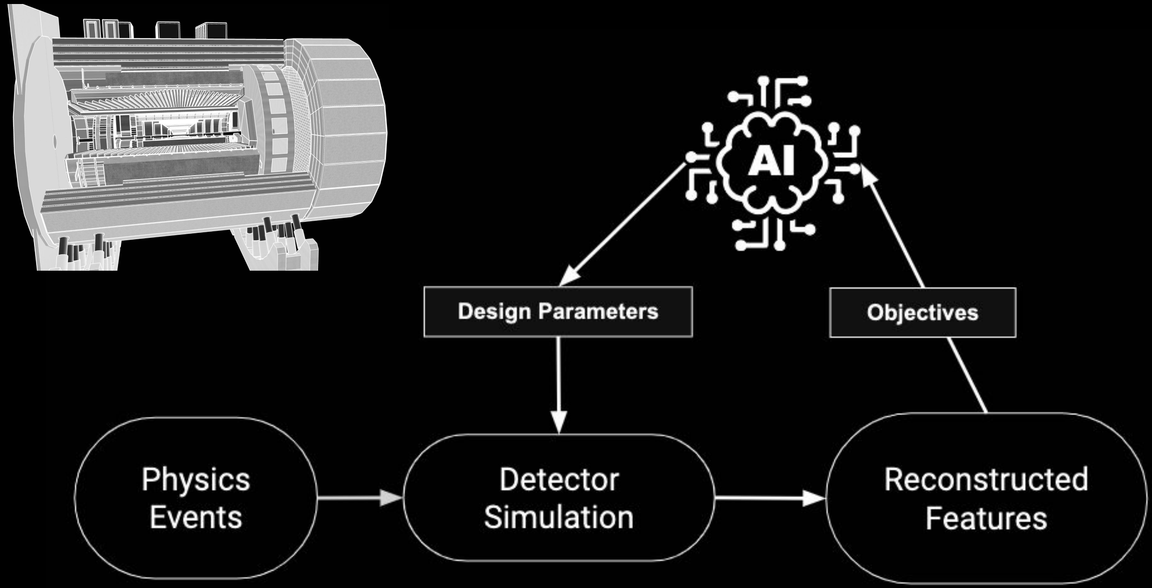
https://eic.ai/ai-ml-references

AI/ML activities are ramping up, and this trend will continue to grow in the next few years.

Jefferson Lab
Exploring the Nature of Matter

AI-assisted design

The AI-assisted design is a good example of how AI can be folded into the SW planning as it embraces all the main steps of the simulation/reconstruction/analysis pipeline



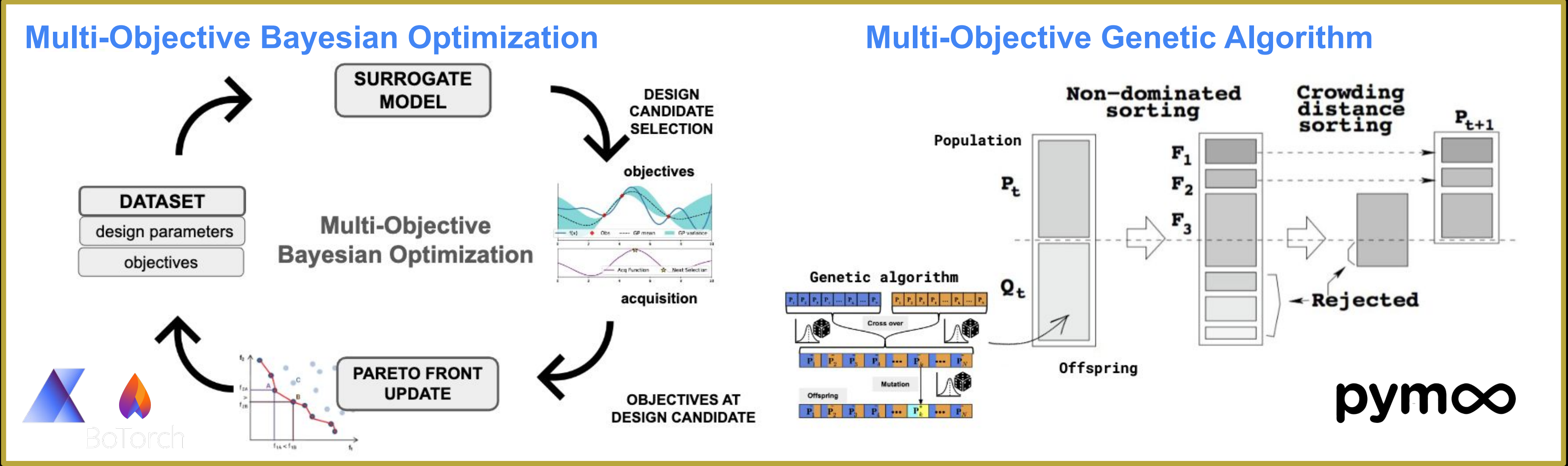
- Leverages heterogeneous computing
- Benefits from rapid turnaround time from simulations to analysis of high-level reconstructed observables
- The ePIC SW stack offers multiple features that facilitate AI-assisted design (e.g., modularity of simulation, reconstruction, analysis, easy access to design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

(see also talk by [D. Lawrence, EIC SW Overview](#) at CHEP2023)

*AI/ML can potentially enter in all the steps of the design pipeline

AI-assisted design

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



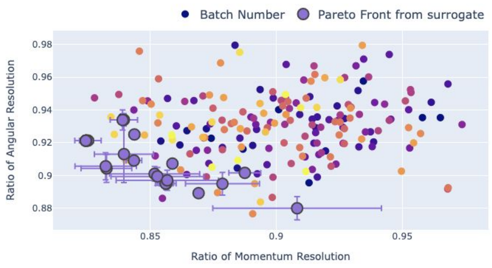
- Engage with Open Source projects (e.g., Ax: adaptive experimentation platform supported by Meta AI)
- Integrate cutting-edge data science built-in features for database backend to store experiments, visualization/interpretation, and presentation of results.

— See 2nd AI4EIC workshop, https://indico.bnl.gov/e/AI4EIC —

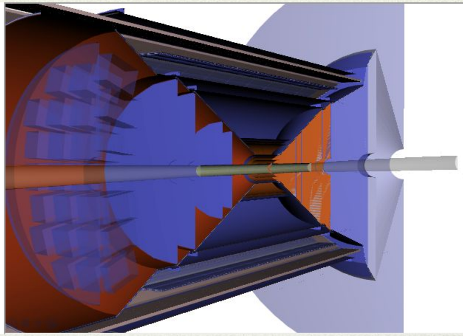
Navigate the Pareto front

CF et al, NIM A, 2023, 167748

Multi Objective Bayesian Optimization

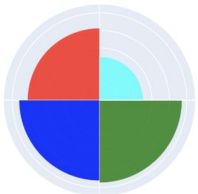


GEANT4 Visualization of the design



Click on petals for finer evaluations

Performance of the Chosen Design Solution

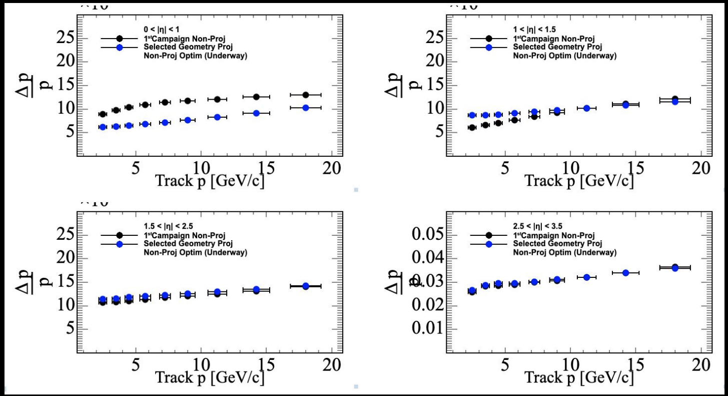


Design Parameters Table

Parameter Name	Parameter Value
Angle of cone [deg]	25.00
Radius of uRwell-1 [cms]	32.47
z E-TTL [cms]	171.00
z F-TTL [cms]	157.60
z EST-1 [cms]	40.39
z EST-3 [cms]	85.09
z FST-1 [cms]	35.03
z FST-3 [cms]	83.78
z FST-5 [cms]	131.27

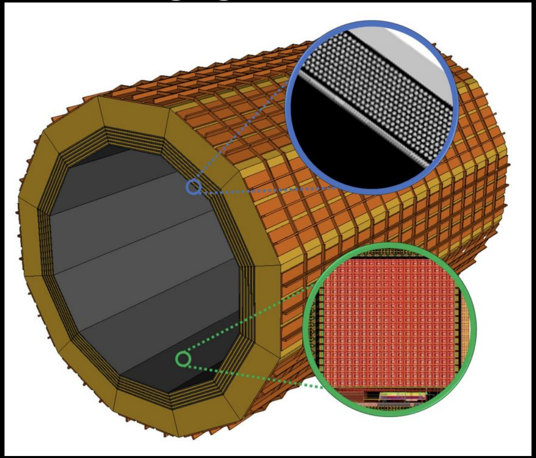
The whole idea of the AI-assisted design is that of determining trade-off optimal solutions in a multidimensional design space driven by multiple objectives

For an interactive visualization:
https://ai4eicdetopt.pythonanywhere.com

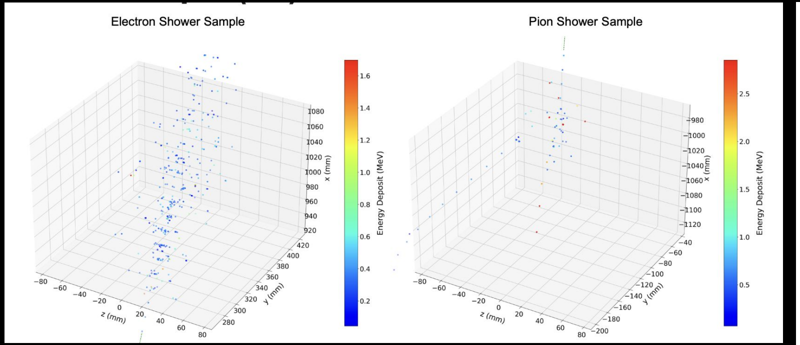


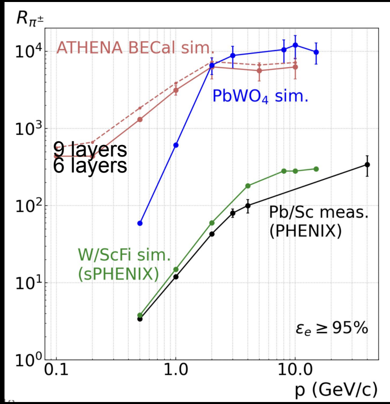
AI/ML for PID: Shower Imaging

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

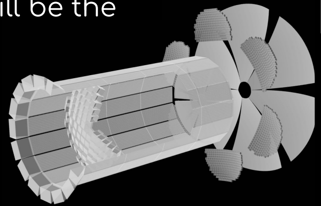
PID of low energy μ that curl in the barrel ECal — J/ψ reconstruction and TCS

Improving PID, providing a space coordinate for DIRC reconstruction

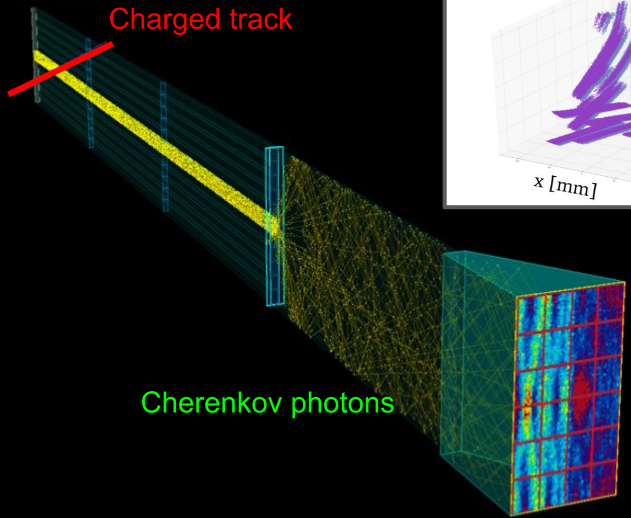
[1] N. Apadula, et al. "Monolithic active pixel sensors on cmos technologies." arXiv preprint arXiv:2203.07626 (2022).
[2] C. Peng, ML Particle Identification with Measured Shower Profiles from Calorimetry, AI4EIC 2nd workshop (2022).

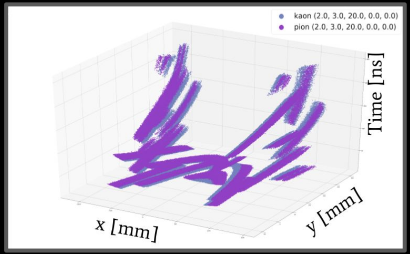
AI/ML for PID: Cherenkov

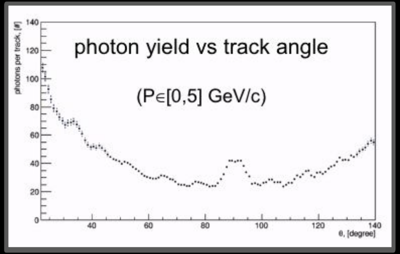
Cherenkov detectors will be the backbone of PID at EIC



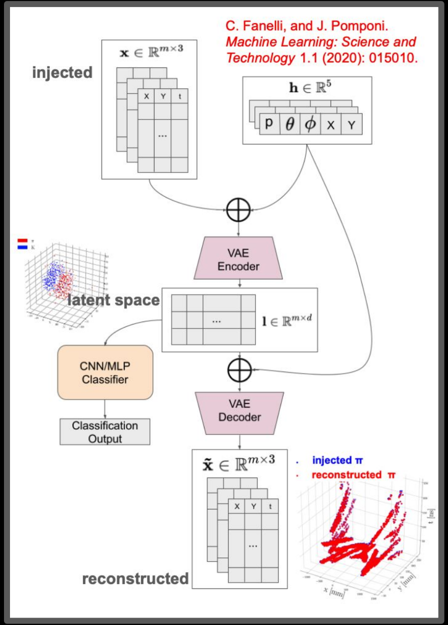
DIRC at GlueX is instrumental for PID







DeepRICH



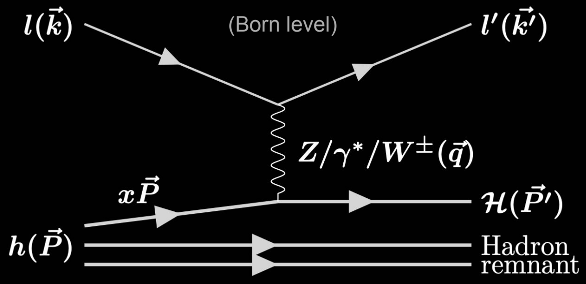
- Need for faster and accurate simulations and reconstruction
- Complex hit patterns (DIRC is the most complex), sparse data, response vs kinematics
- DeepRICH: same reconstruction performance of best reconstruction algorithm with ~4 orders of magnitude speed-up in inference time on GPU
- Possibility to learn at the event-level rather than at the track/particle level. Can generate hit pattern.

[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.

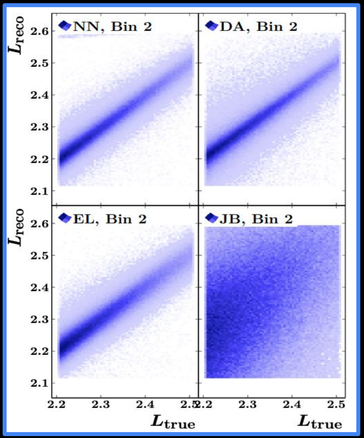
Deeply Learning DIS

Example in one specific bin

DIS fundamental process @EIC



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



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	**NN: 35**	EL: 35
		JB: 180	DA: 103	JB: 203	DA: 62
2	350530	**NN: 69**	EL: 82	**NN: 40**	EL: 43
		JB: 167	DA: 96	JB: 192	DA: 64
3	138456	**NN: 98**	EL: 130	NN: 55	**EL: 53**
		JB: 138	DA: 100	JB: 150	DA: 77
4	74844	**NN: 67**	EL: 84	**NN: 44**	EL: 46
		JB: 117	DA: 77	JB: 138	DA: 63
5	31043	**NN: 64**	EL: 91	**NN: 36**	EL: 41
		JB: 102	DA: 73	JB: 117	DA: 53
6	11475	**NN: 53**	EL: 79	**NN: 33**	EL: 36
		JB: 83	DA: 61	JB: 100	DA: 45
7	3454	**NN: 50**	EL: 69	**NN: 36**	EL: 38
		JB: 74	DA: 55	JB: 93	DA: 42
8	624	**NN: 36**	EL: 55	**NN: 33**	EL: 37
		JB: 67	DA: 45	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 $\log(x) - \log(x_{\mathrm{true}})$ and $\log(Q^2) - \log(Q^2_{\mathrm{true}})$ respectively.

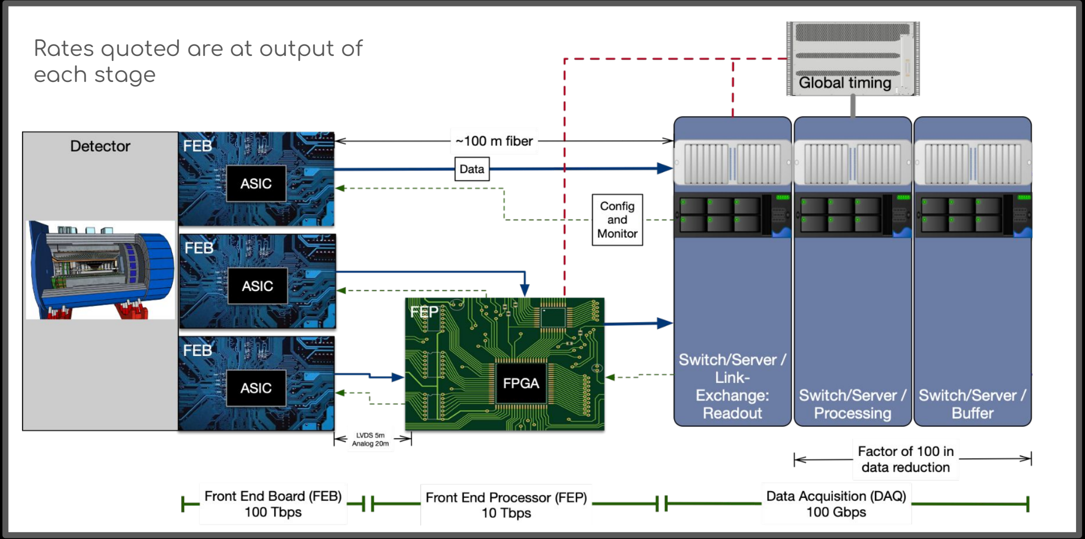
- Use of DNN to reconstruct the kinematic observable Q^2 and x in the study of neutral current DIS events at ZEUS and H1 experiments at HERA.
- The performance compared to electron, Jacquet-Blondel and the double-angle methods using data-sets independent of training
- Compared to the classical reconstruction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

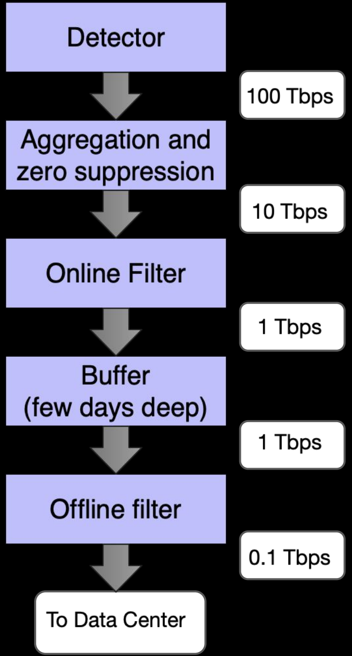
Jefferson Lab
Exploring the Nature of Matter

[1] M. Diefenthaler, et al. “Deeply Learning DIS Kinematics” arXiv:2108.11638, EPJC 82, 1064 (2022)
[2] M. Arratia, et al., “Reconstructing the kinematics of DIS with DL”, NIM-A 1025 (2022): 166164

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. Data flow unimpeded in parallel channels, organized in multi-dimensions and time. (max data preservation, diverse topologies, bkgd control)
- 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.
[2] F. Ameli, et al., Streaming readout for next generation electron scattering experiments, Eur. Phys. J. Plus, 2022

AI4EIC community

https://eic.ai

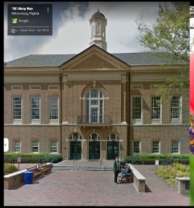
- AI will be an integral part of the EIC science and to work in this direction, a dedicated AI Working Group (AI4EIC) has been established 2 years ago within the EICUG (https://www.eicug.org)
- AI4EIC serves as an entry point to AI applications and organizes workshops, tutorials, hackathons, challenges, etc.
- AI4EIC fosters connections between ePIC and the Data Science / Computer Science community

- Workshops —2 workshops, 200+ participants each (https://eic.ai/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
    - Several tutorials (https://eic.ai/community)
    - Hackathon events are built around specific challenges for EIC and help identify strategies, architectures and algorithms that will benefit the EIC physics program (https://eic.ai/hackathons)
    - Additionally, AI4EIC is committed to establishing educational events (e.g., schools) designed to enhance AI and ML proficiency within the EIC community (https://eic.ai/community), (https://eic.ai/ai-ml-references)

AI4EIC

https://eic.ai

https://eic.ai/workshops




- Workshop: (2022)

  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

- **Discussion from this workshop contributed to NSAC LRP**
- **Paper in preparation**

- Tutorials:
  - MOBO
  - OmniFold
  - MLFlow
  - GNN

https://eic.ai/community

- Hackathon:

  (10 teams from North, South America, Asia, Europe)

https://doi.org/10.5281/zenodo.7197023

— Forthcoming AI4EIC workshop in winter 2023, announcement will be made soon —

Jefferson Lab

Conclusions

- **Next generation QCD experiments like ePIC are being designed during the AI revolution:**
  - The ePIC detector could be the first large-scale experiment to have its design optimized using AI/ML
- **Activities in AI/ML for ePIC are ramping up:**
  - Many examples have not been covered in these slides.
  - Distinct challenges posed by the EIC science and the ePIC experiment provide a fertile ground for fostering the development of innovative AI/ML-based solutions.
  - 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.
- **ePIC will take full advantage of SRO and AI using heterogeneous computing**:
  - Near real-time data processing with SRO and autonomous detector control can result in a paradigm shift for NP with faster turnaround time to produce scientific results.
  - A common theme is applying AI-methods with well-understood UQ (both systematic and statistic).
- **Transition from prototyping to deployment in production environments:**
  - This includes integrating AI/ML into our workflows, working closely with the DS community in collaboration with AI4EIC.

Backup

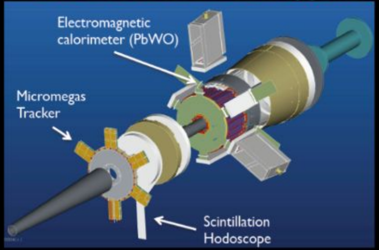
Prototype experiments for next-gen SRO

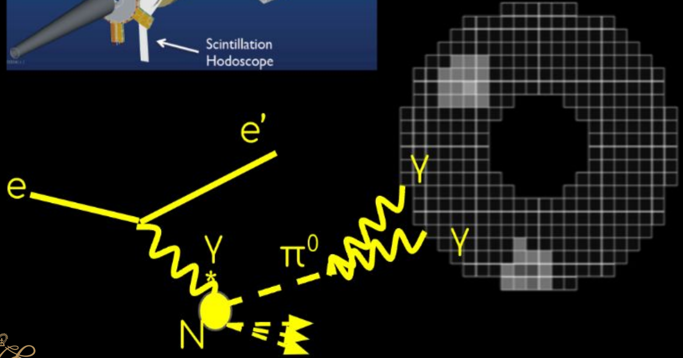
ML deployed on stream of real data

CLAS + EPSCI @JLab

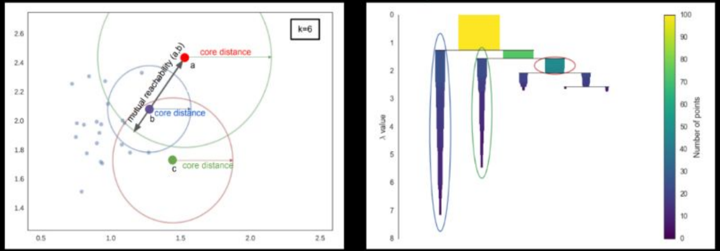
- CLAS12 SRO setup
- TriDAS SR back end
- JANA2 reconstruction framework

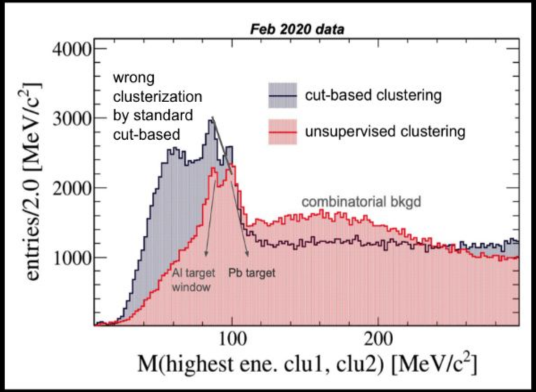
The CLAS12 Forward Tagger, JLab

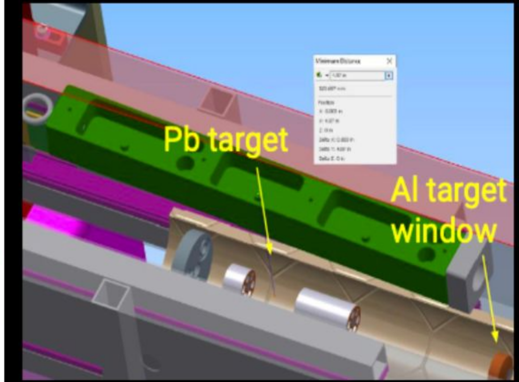




Hierarchical clustering in JANA2



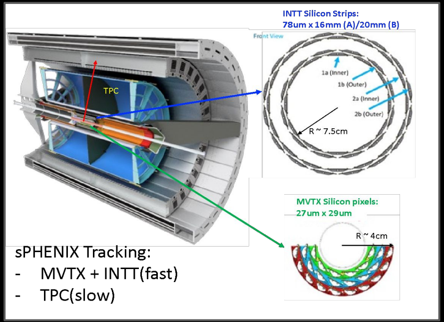


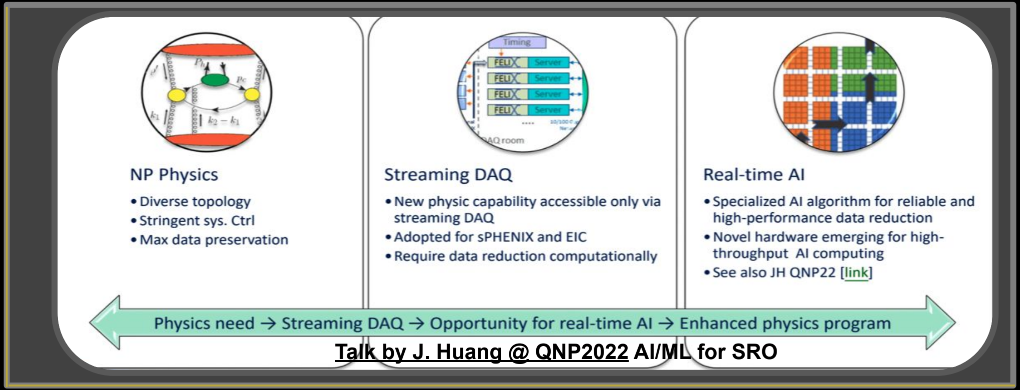


Hierarchical clustering VS traditional clustering of energy deposited by photons; AI robust against variations in experimental conditions* (uncalibrated data in SRO)

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

Many activities towards SRO at EIC





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:

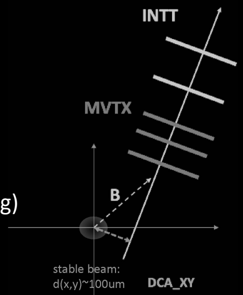
Very high p+p collision rate: ~3MHz

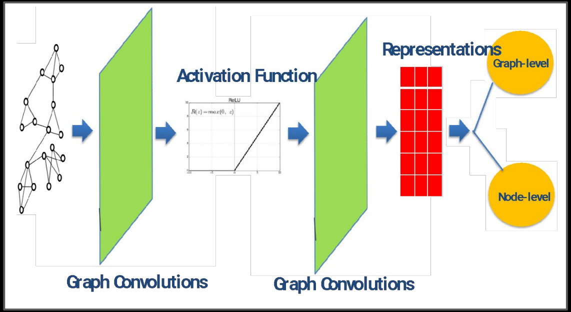
Low rate of rare signals: ~150Hz (beauty for eg)

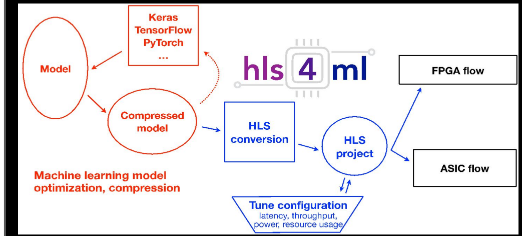
Limited DAQ trigger bandwidth: ~15 kHz

(or 0.5% of p+p collisions)

No effective conventional triggers available



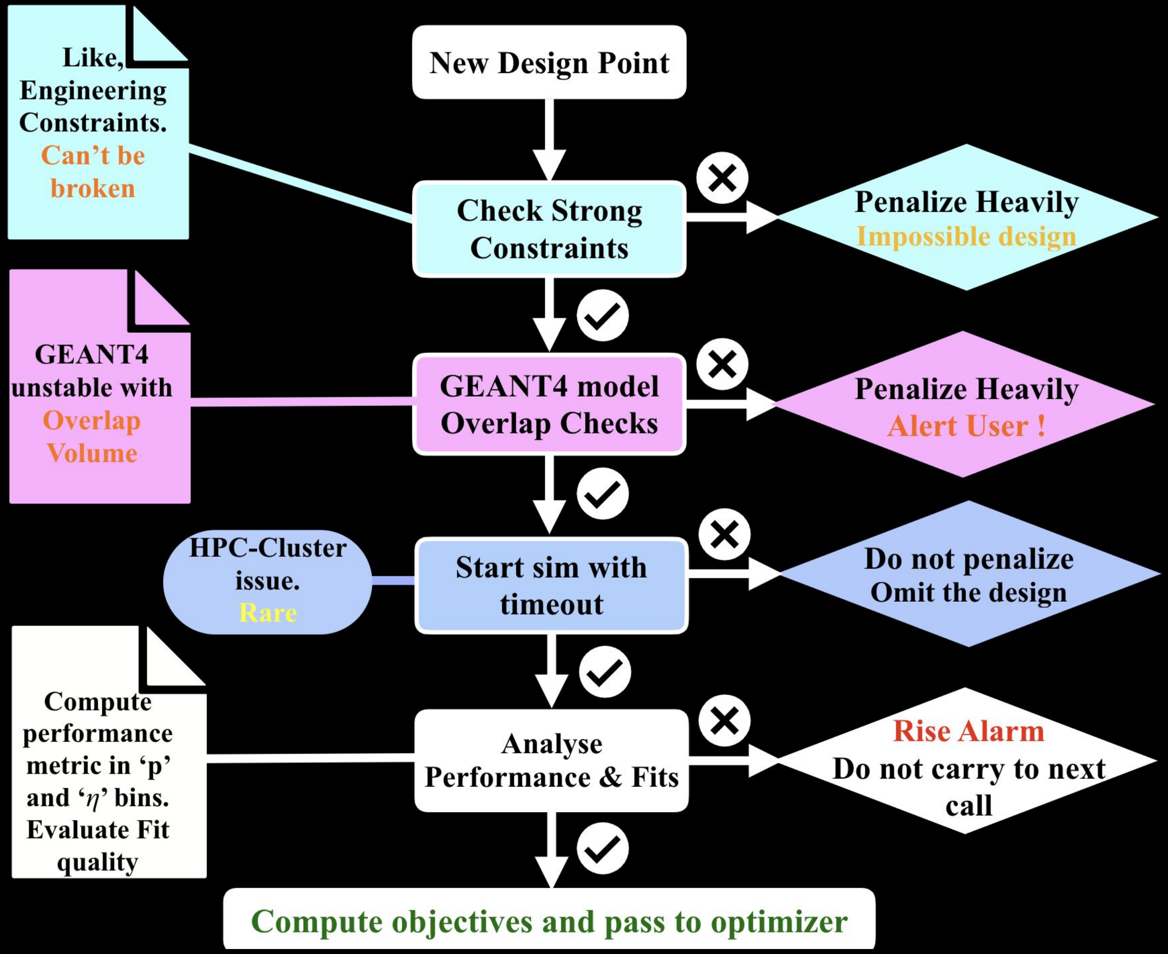




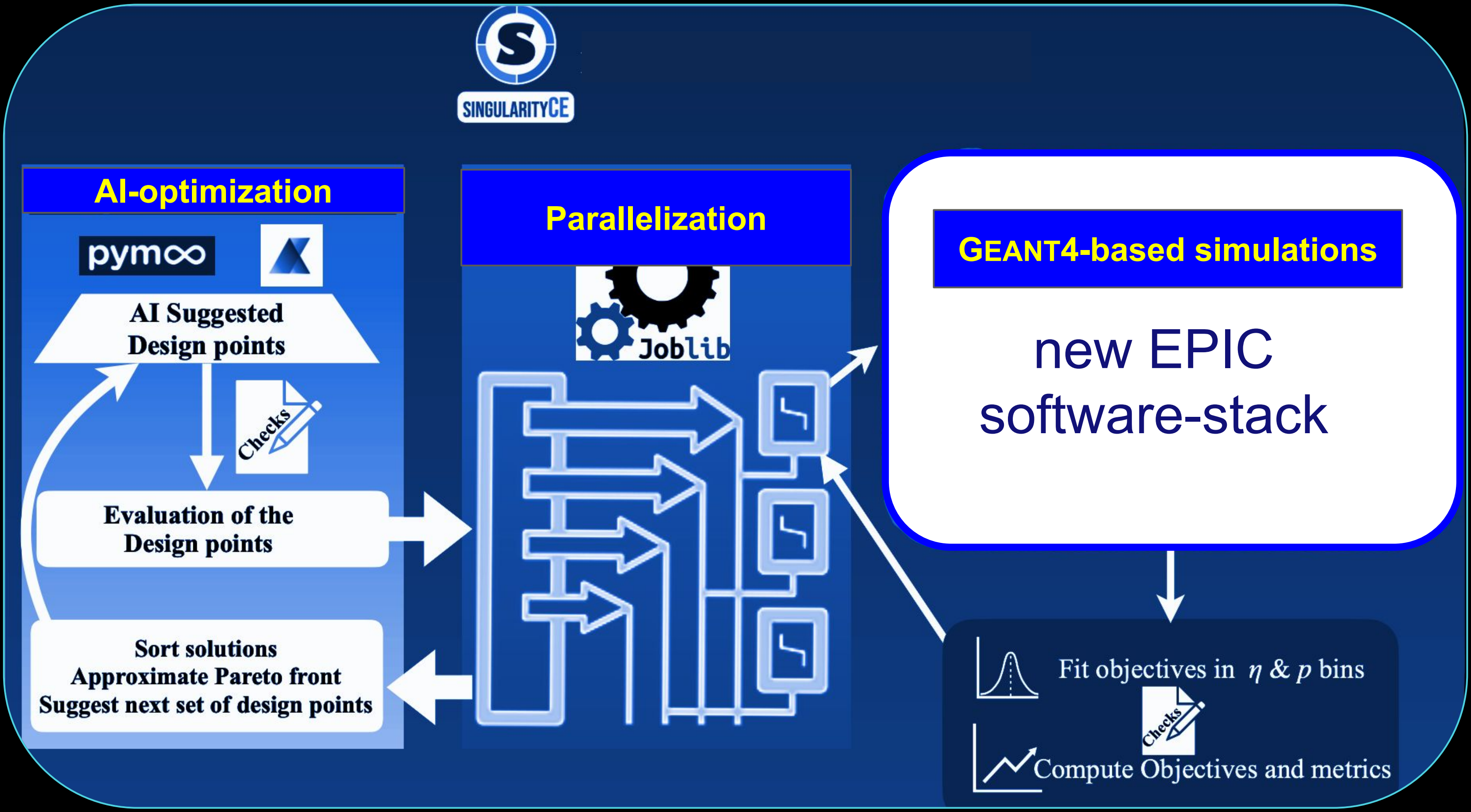
Courtesy of Ming Liu (LANL)

[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)

Internal Checks



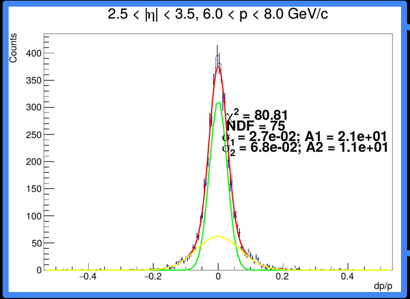
Workflow used in 2205.09185



Fine-grained analysis

Weighted sum with errors

1. Robust fitting procedure in fine-grained phase-space
2. Propagate uncertainties



$$\bar{x_\eta} = \frac{\sum_p x_p w_p}{\sum_p w_p}$$

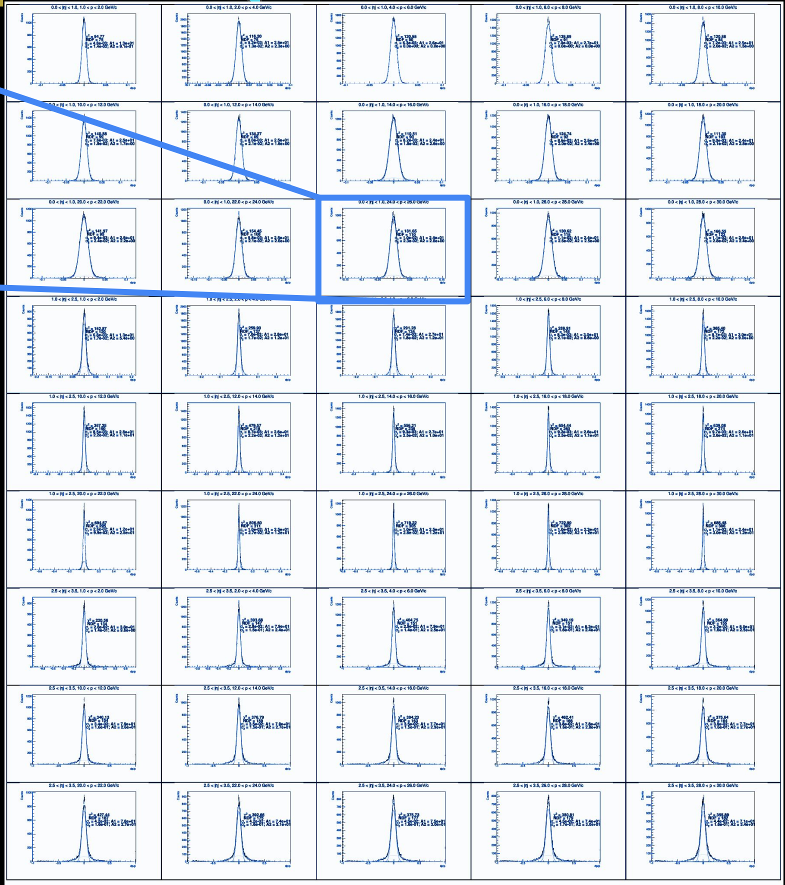
(sum in bins of P)

$$\bar{x} = \frac{\sum_\eta^{N_\eta} \bar{x_\eta}}{N_\eta}$$

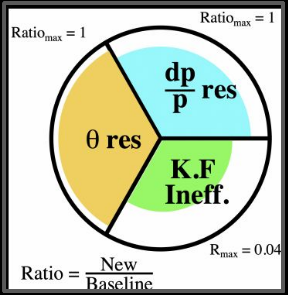
(Average objective in a η bin)

$$\Rightarrow R(f) = \frac{1}{N_\eta} \sum_\eta \left(\frac{\sum_p w_{p,\eta} \cdot R(f)_{p,\eta}}{\sum_p w_{p,\eta}} \right)$$

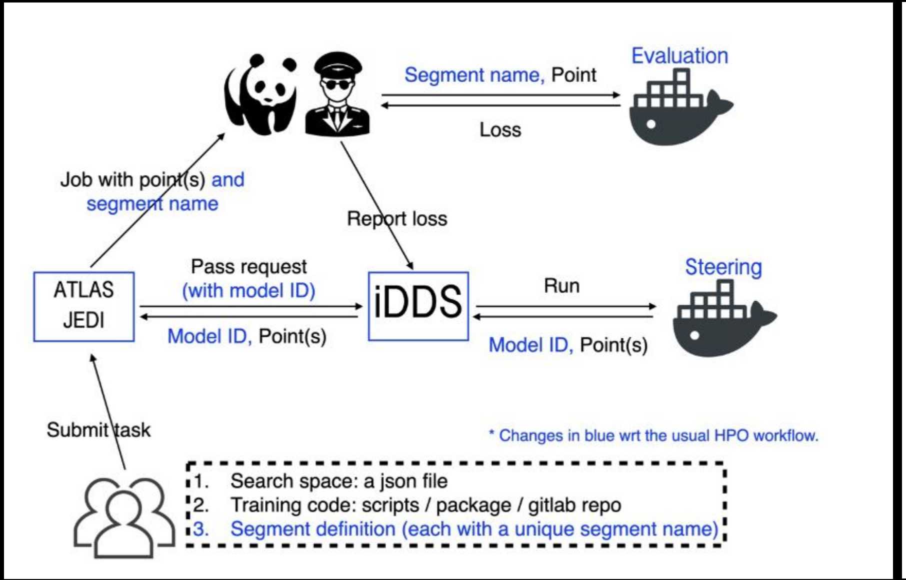
Weighted sum with errors

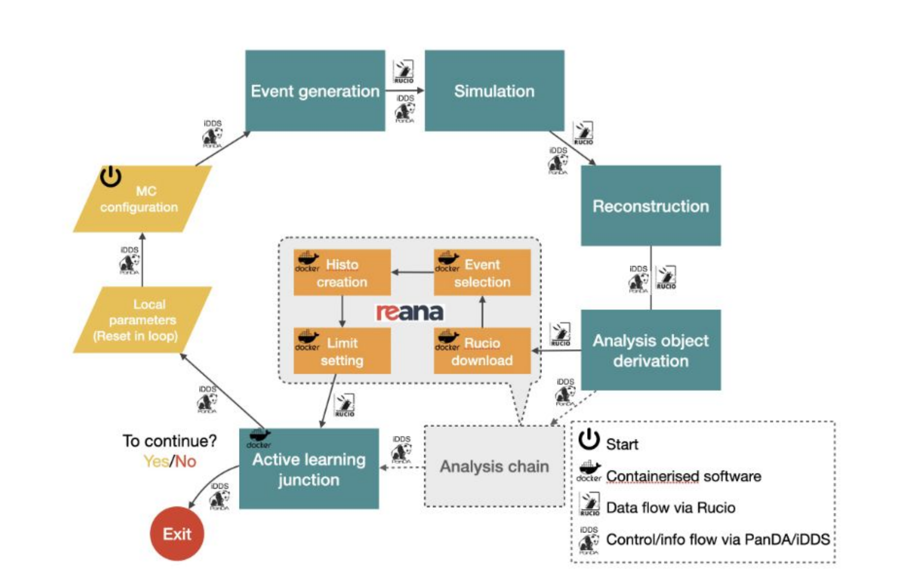


3. Do this for several objectives



Complex workflow

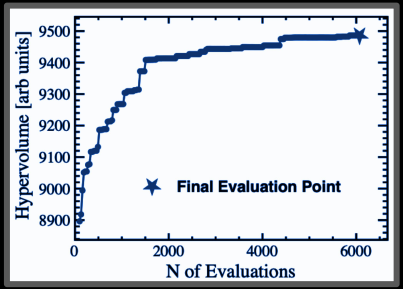




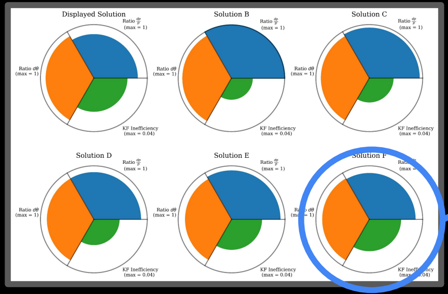
Example of HPO service in use in ATLAS, and BO workflow

Analyzing the results

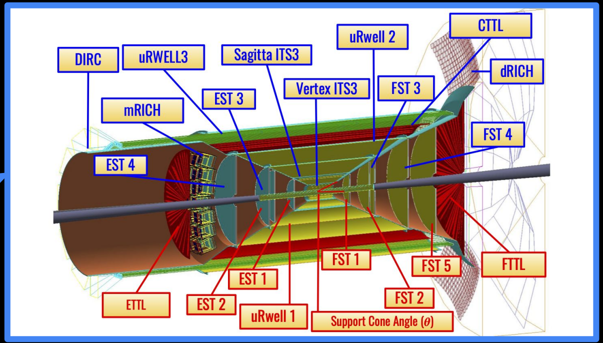
1 Can take a snapshot any time during evaluation



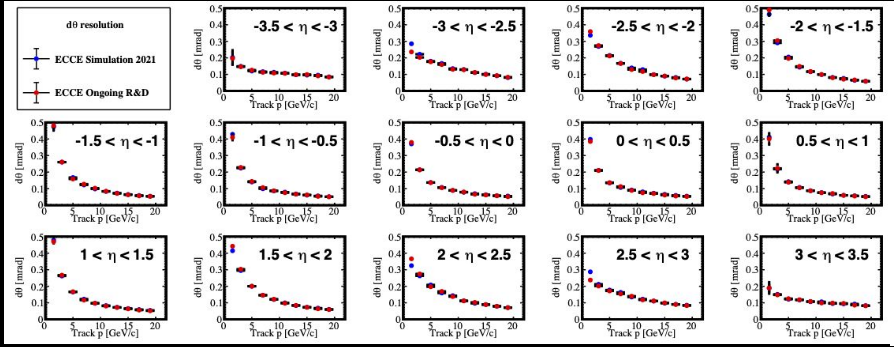
2 Updated Pareto Front at time t

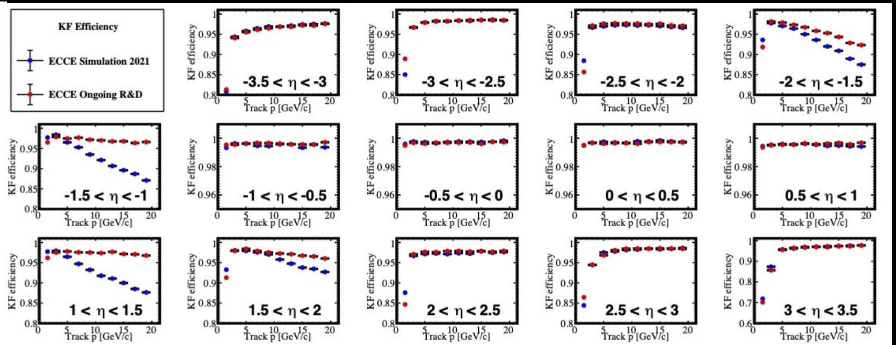


3 At each point in the Pareto front corresponds a design

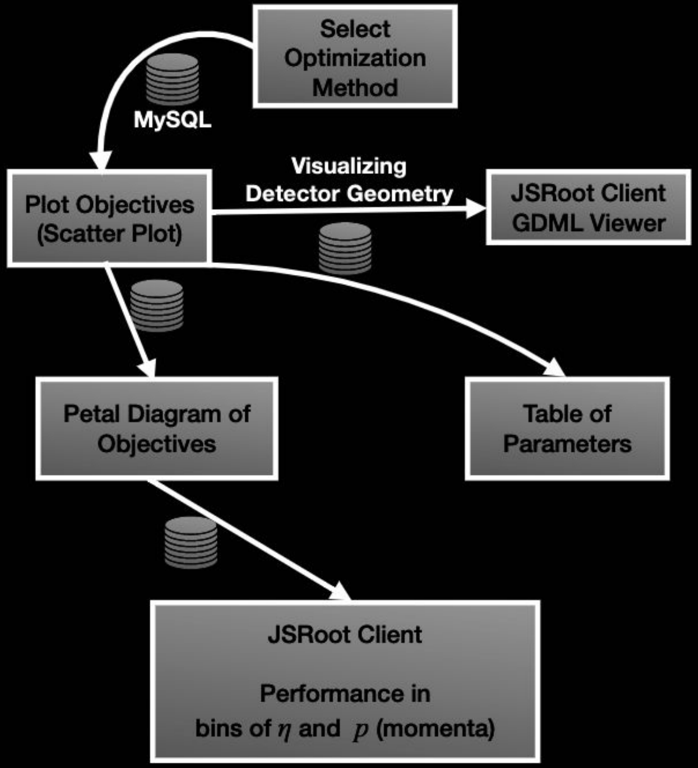


4 Analysis of Objectives (momentum resolution, angular resolution, KF efficiency)





Visualization



- 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

Geometry

D. Lawrence, EIC SW Overview

Geometry defined using **DD4HEP**
https://github.com/AIDASoft/DD4hep

DD4hep provides a geometry description that is used for both ***simulation*** and ***reconstruction***

XML

```xml
<define>
  <comment>
    ---------------------------------------
    EM Calorimeter Parameters with AstroPix
    ---------------------------------------
  </comment>
  <comment>
    To change the number of imaging layers from, e.g., 9 to 6,
    change EcalBarrelImagingLayers_nMax variable
  </comment>
  <constant name="EcalBarrelImagingLayers_nMax"   value="6"/>
  <constant name="EcalBarrel_Calorimeter_zmin"
    value="min(257*cm, EcalBarrelBackward_zmax)"/>
```

XML

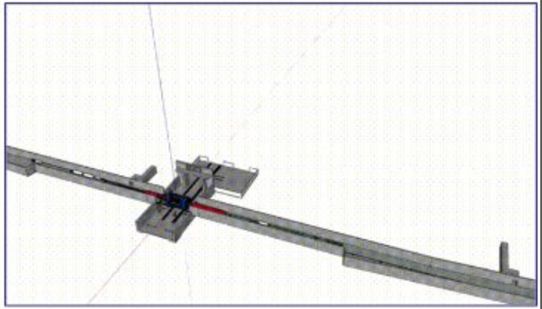
```xml
      ---------------------------------------
      Imaging Layers of Barrel EM Calorimeter
      ---------------------------------------
      Silicon (Astropix) readout layers for imaging 3d showers
    </comment>
    <detector
      id="ECalBarrel_ID"
      name="EcalBarrelImaging"
      type="epic_EcalBarrelInterlayers"
      readout="EcalBarrelImagingHits"
```

C++

```cpp
// barrel ecal layers contained in an assembly
static Ref_t create_detector(Detector& desc, xml_h e, SensitiveDetector sens)
{
  Layering   layering(e);
  xml_det_t  x_det    = e;
  Material   air      = desc.air();
  int        det_id   = x_det.id();
  string     det_name = x_det.nameStr();
  double     offset   = x_det.attr<double>(_Unicode(offset));
  xml_comp_t x_dim    = x_det.dimensions();
  int        nsides   = x_dim.numsides();
  double     inner_r  = x_dim.rmin();
  double     dphi     = (2 * M_PI / nsides);
  double     hphi     = dphi / 2;

  DetElement sdet(det_name, det_id);
  Volume     motherVol = desc.pickMotherVolume(sdet);

  Assembly    envelope(det_name);
  Transform3D  tr_global = Translation3D(0, 0, offset) * RotationZ(hphi);
```



ePIC specific: ***npsim*** (python)
https://github.com/eic/npsim

- EIC physics
- Optical photon settings
- etc.

- Integrated PODIO *(see next slide)*
- Parameters defined in XML that can be accessed at runtime
- Detector types implemented in C++ using ***root TGeo***

Data Model and I/O

D. Lawrence, EIC SW Overview

- “Plain Old Data” I/O
  - helpful for ***exchange*** with outside software (e.g. ***AI/ML***)
- Data model defined in **yaml**
- Maps objects to ROOT ***TTrees***
- Core Data model is ***EDM4hep*** (https://github.com/key4hep/EDM4hep)
  - *“A generic event data model for future HEP collider experiments”*
  - Supports 1-1, 1-many, and many-many associations
    - e.g. reconstructed particle can be associated back to generated particle
- EIC-specific Data model is ***EDM4eic*** (https://github.com/eic/EDM4eic)
  - Augments EDM4hep
  - Changes propagated upstream to EDM4hep where appropriate

example structure defined in ***edm4eic.yaml***

```
edm4eic::CalorimeterHit:
  Description: "Calorimeter hit"
  Author: "W. Armstrong, S. Joosten"
  Members:
    - uint64_t          cellID          // The detector specific (geometrical) cell id.
    - float             energy          // The energy for this hit in [GeV].
    - float             energyError     // Error on energy [GeV].
    - float             time            // The time of the hit in [ns].
    - float             timeError       // Error on the time
    - edm4hep::Vector3f position        // The global position of the hit in world coordinates [mm].
    - edm4hep::Vector3f dimension       // The dimension information of the cell [mm].
    - int32_t           sector          // Sector that this hit occurred in
    - int32_t           layer           // Layer that the hit occurred in
    - edm4hep::Vector3f local           // The local coordinates of the hit in the detector segment [mm].
```