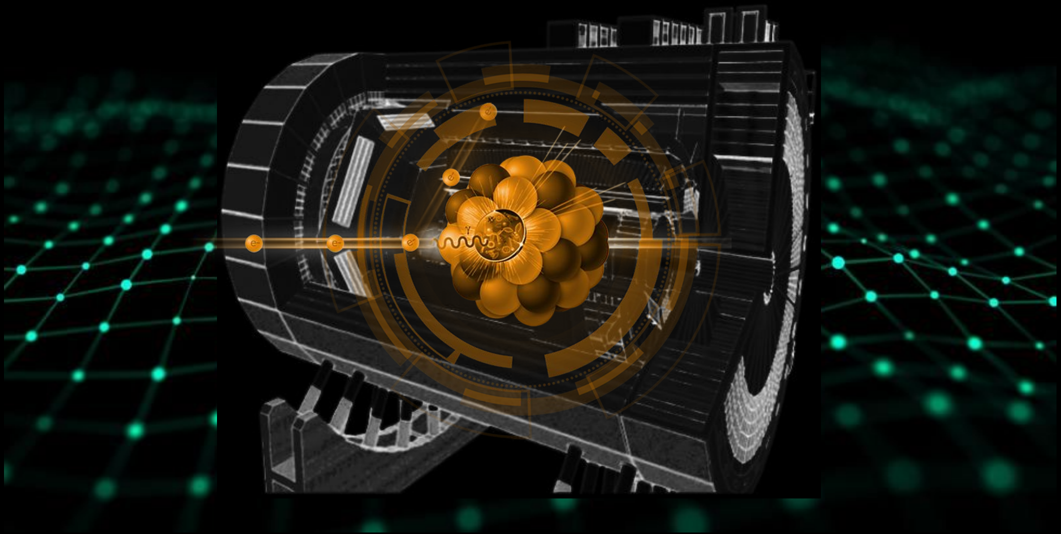
AI/ML + Data Science tools for detector design at the EIC



5/26/2023

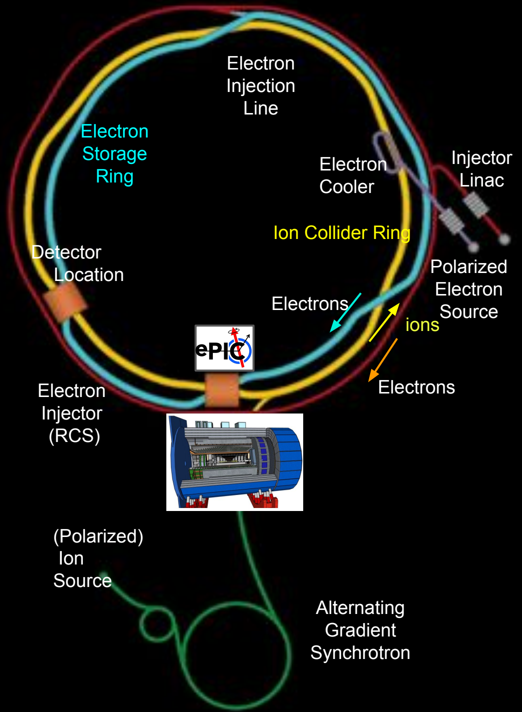
Cristiano Fanelli (and Authors in the slides)

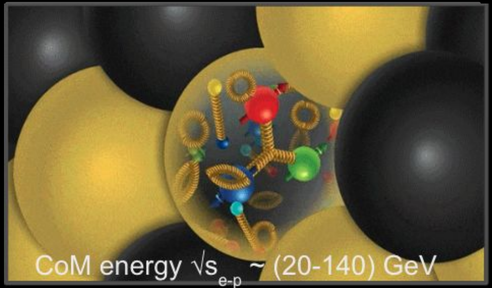
ISNET-9 INFORMATION AND STATISTICS IN NUCLEAR EXPERIMENT AND THEORY

Electron Ion Collider (EIC)

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

polarized electron - polarized protons/ions



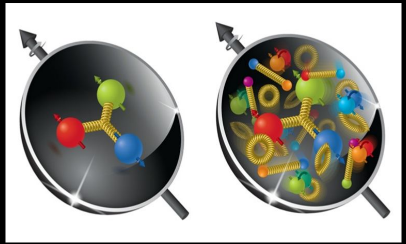


World-wide interest in EIC, thousands of users and hundreds of institutions involved

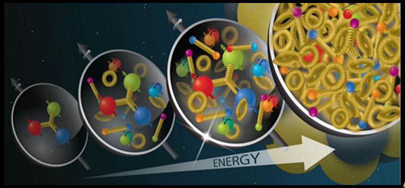
Total estimated cost ~ $1.6-2.6B

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?



Jefferson Lab
Exploring the Nature of Matter

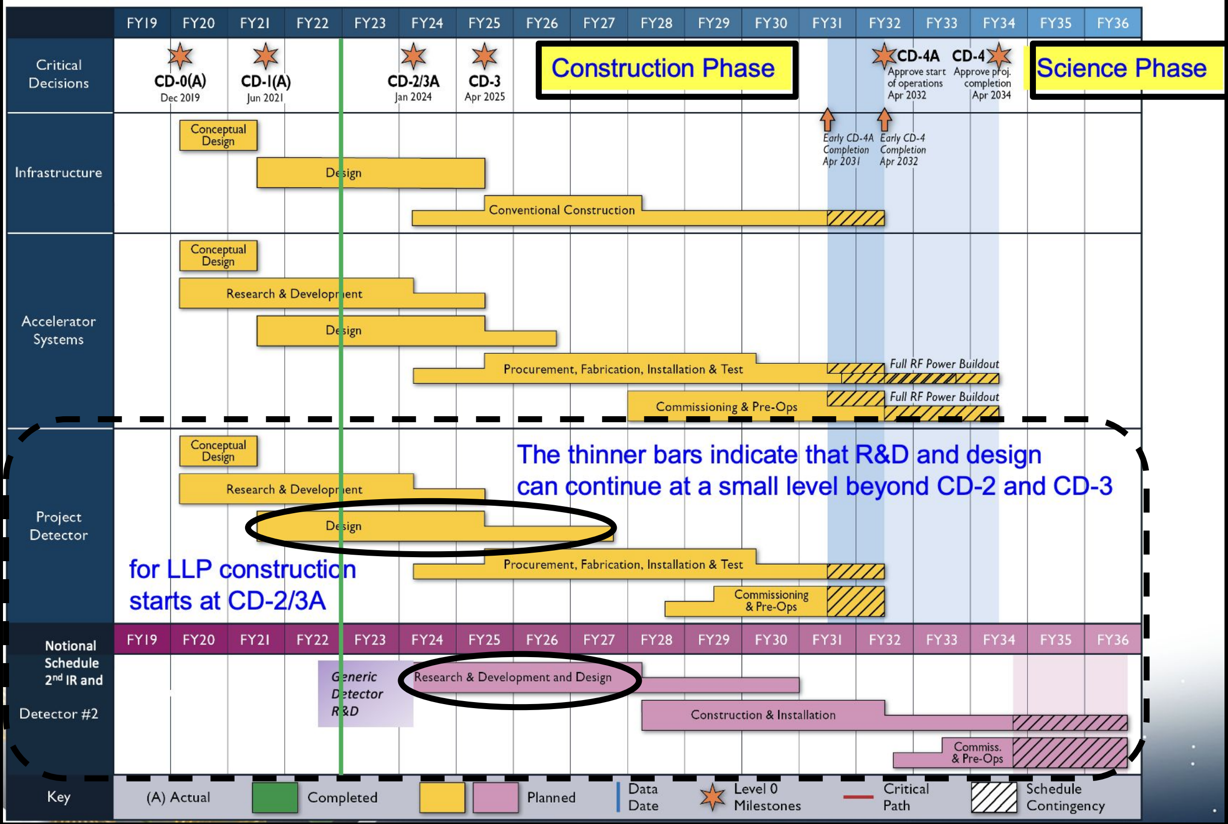
EIC Schedule

R. Ent, AI4EIC workshop, October 2022

Magnet landed in the partially assembled flux return and outer HCAL (in 2021)

ePIC Collaboration recently formed (previously ECCE proto-collaboration selected as reference detector)

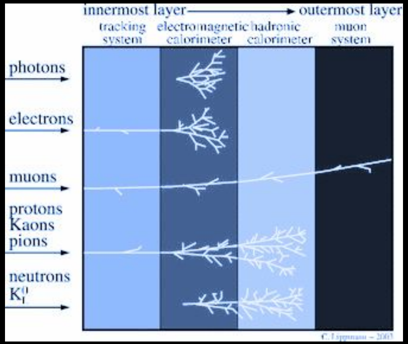
Detector-2 ePIC accelerator

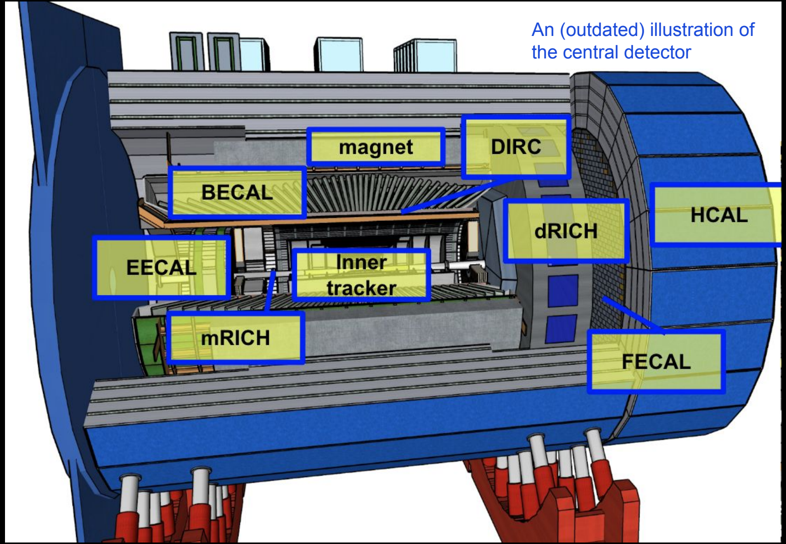


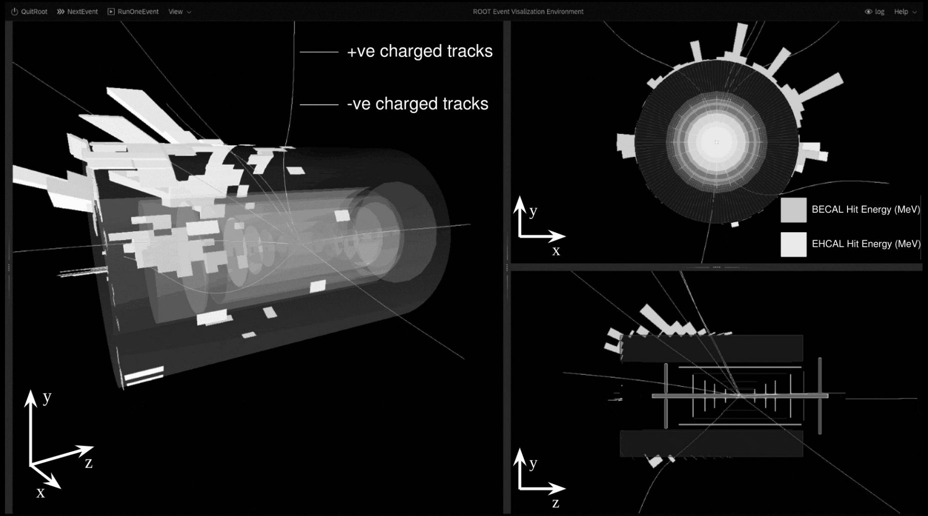
Opportunities for (Detector Design) Optimization!

The ePIC Detector

- A large-scale experiment with an integrated detector that extends for ~ ± 35 m to include the central, far-forward, and far-backward regions.
- To enable the EIC physics we need a central detector that is: hermetic and asymmetric
- From fundamental physics, we know that different types of particles interact differently with matter and we need to develop specific devices to identify them

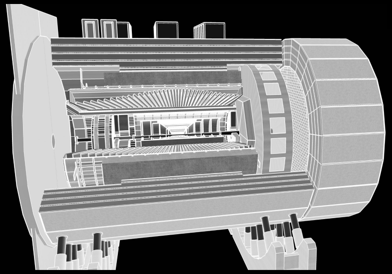


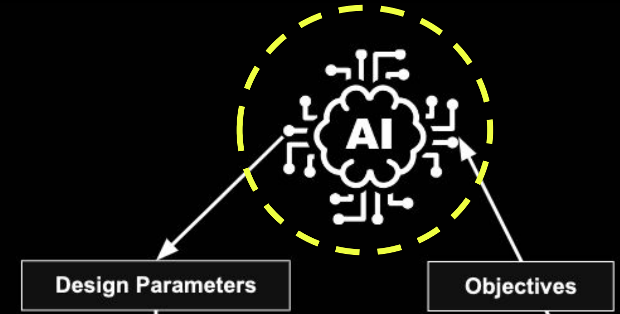


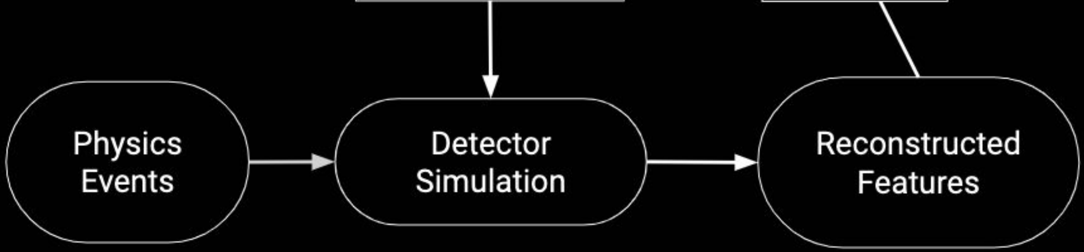


AI-assisted design

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)

helps steering the design (and eventually fine-tune it)

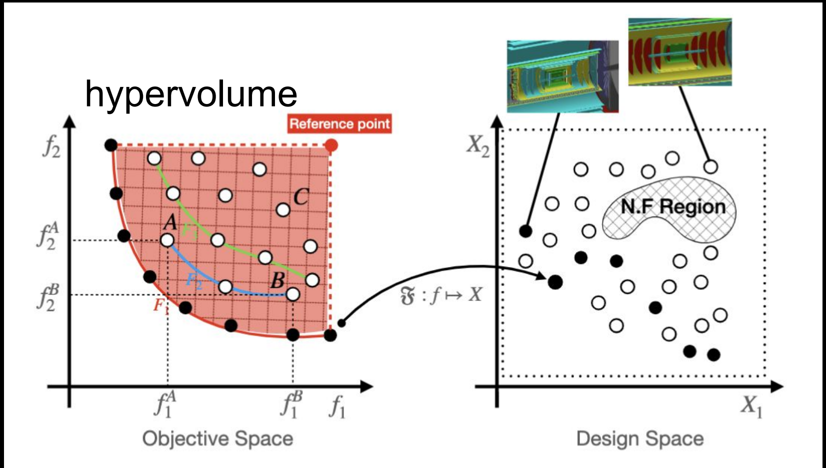
can capture hidden correlations among design parameters

Jefferson Lab
Exploring the Nature of Matter

Why Multi-Objective Problem?

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

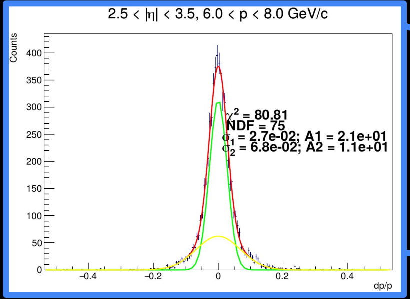
- 3 Types of Objectives
  - **Intrinsic detector performance** (resolutions, efficiencies) for each sub-detector — Tracking, calorimetry, PID — noisy
  - **Physics-performance** — Multiple physics channels, equally important in the EIC physics program
  - **Costs** (e.g., material costs, provided a reliable parametrization)
- Objectives can be competing with each other
  - E.g. Better detector response come with higher costs; better resolutions may imply lower efficiencies; etc.

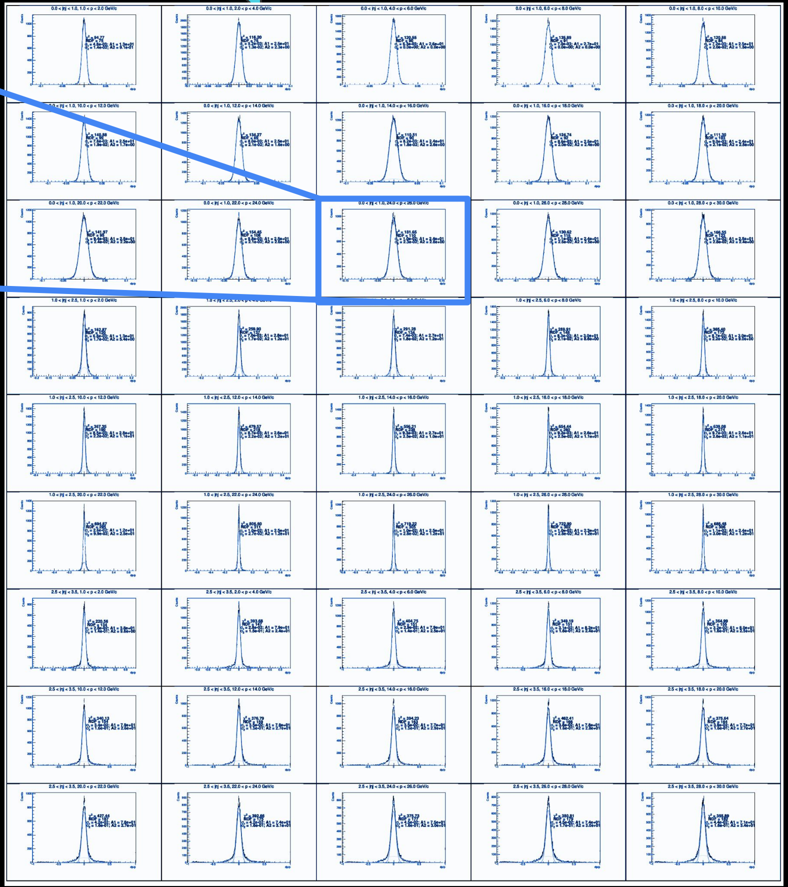


For illustrative purposes

Fine-grained analysis

Weighted sum with errors





Weighted sum with errors

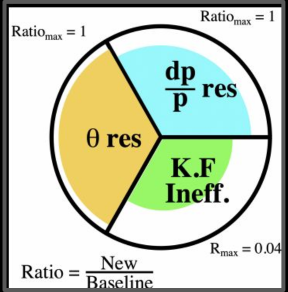
1. Robust fitting procedure in fine-grained phase-space
2. Propagate uncertainties

$$\bar{x_\eta} = \frac{\sum_p x_p w_p}{\sum_p w_p} \qquad \bar{x} = \frac{\sum_\eta^{N_\eta} \bar{x_\eta}}{N_\eta}$$

(sum in bins of P) (Average objective in a η bin)

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

3. Do this for several objectives

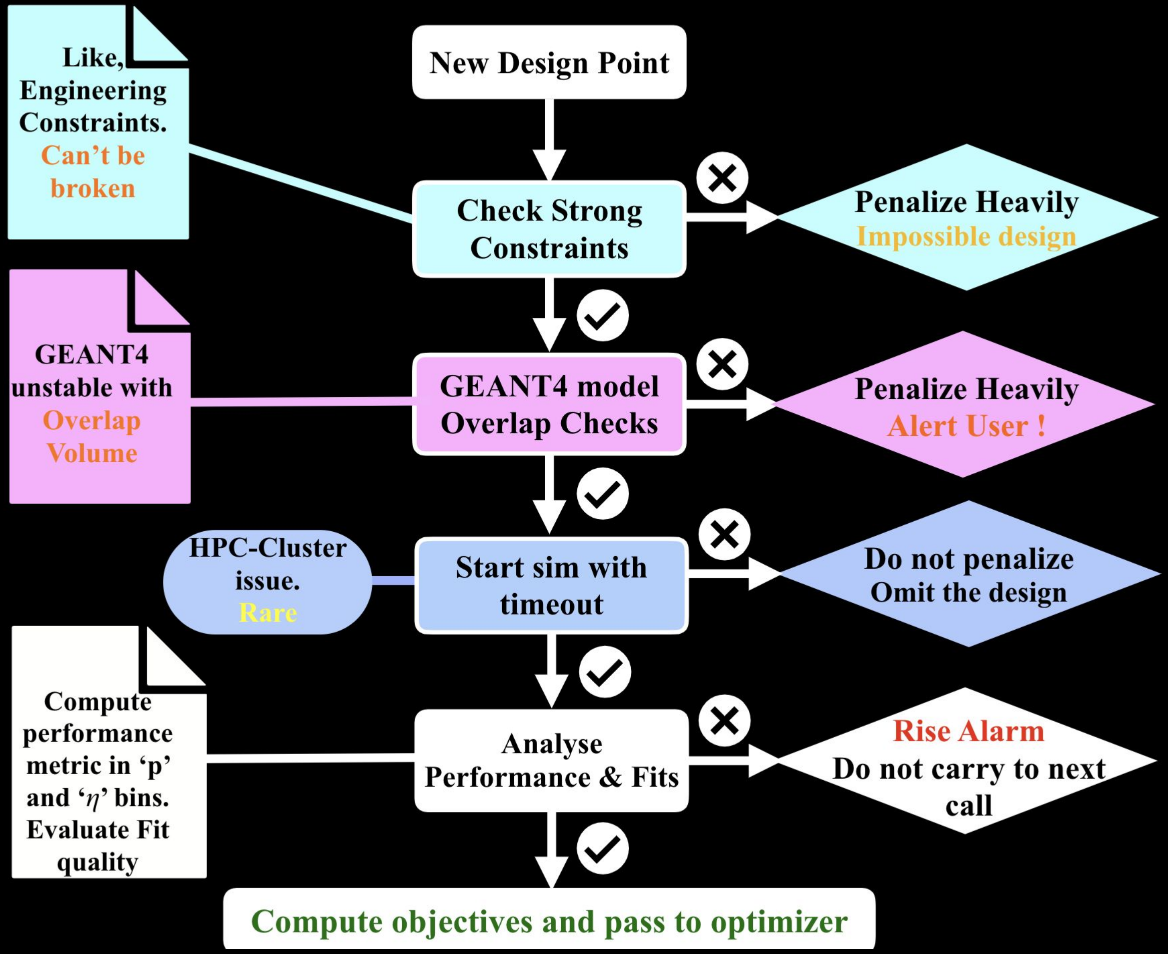


Example for tracking system

(more details in C. Fanelli et al (ECCE Coll.), NIMA Vol 1047, Feb 2023, 167748)

Jefferson Lab
Exploring the Nature of Matter

Checks performed



(more details in C. Fanelli et al (ECCE Coll.), NIMA Vol 1047, Feb 2023, 167748)

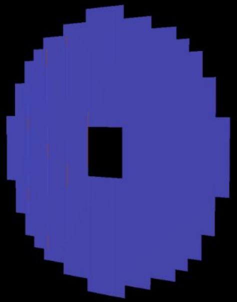
Constraints

$$
\begin{aligned}
& \min \mathbf{f_m}(\mathbf{x}) && m = 1, \cdots, M \\
s.t.\quad & \mathbf{g_j}(\mathbf{x}) \leq 0, && j = 1, \cdots, J \\
& \mathbf{h_k}(\mathbf{x}) = 0, && k = 1, \cdots, K \\
& x_i^L \leq x_i \leq x_i^U, && i = 1, \cdots, N
\end{aligned}
$$

sub-detector	constraint	description
EST/FST disks	$min\left\{\sum_{i}^{disks}\left\|\frac{R^i_{out}-R^i_{in}}{d}-\left\lceil\frac{R^i_{out}-R^i_{in}}{d}\right\rceil\right\|\right\}$	**soft constraint**: sum of residuals in sensor coverage for disks; sensor dimensions: $d = 17.8$ (30.0) mm
EST/FST disks	$z_{n+1} - z_n >= 10.0$ cm	**strong constraint**: minimum distance between 2 consecutive disks
sagitta layers	$min\left\{\left\|\frac{2\pi r_{sagitta}}{w}-\left\lceil\frac{2\pi r_{sagitta}}{w}\right\rceil\right\|\right\}$	**soft constraint**: residual in sensor coverage for every layer; sensor strip width: $w = 17.8$ mm
μRWELL	$r_{n+1} - r_n >= 5.0$ cm	**strong constraint**: minimum distance between μRwell barrel layers

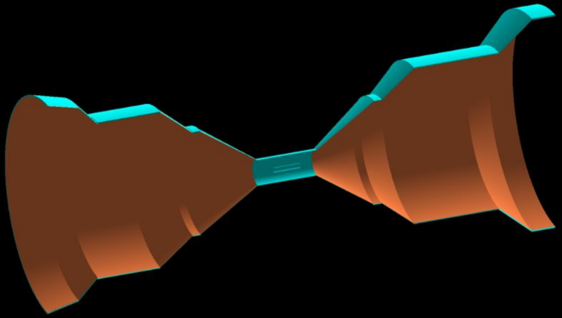
Example of constraints implemented for the tracking system

(more details in C. Fanelli et al (ECCE Coll.), NIMA Volume 1047, Feb 2023, 167748)



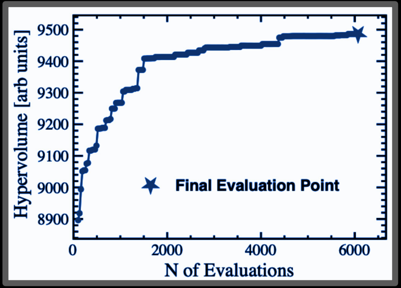
FST/EST Disks

Barrel Si Layer

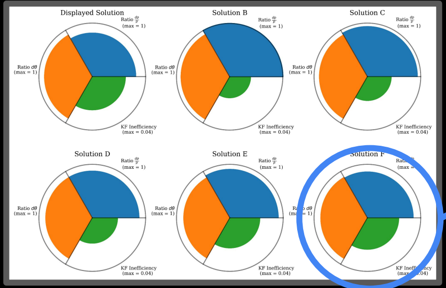


Analyzing the results

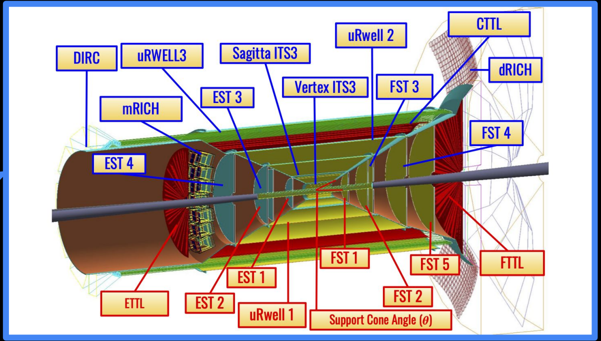
1 Can take a snapshot any time during evaluation



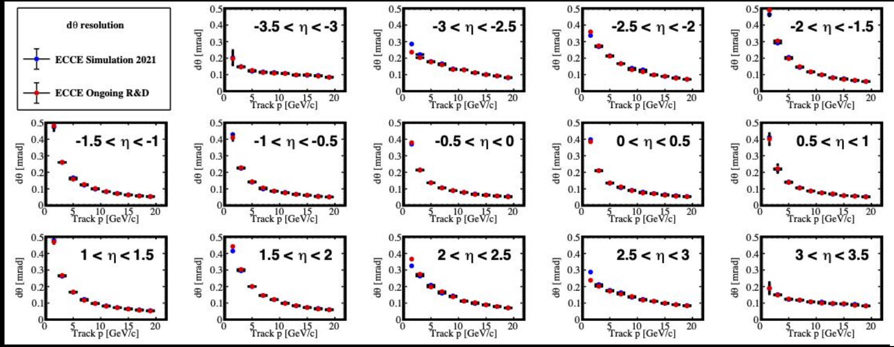
2 Updated Pareto Front at time t

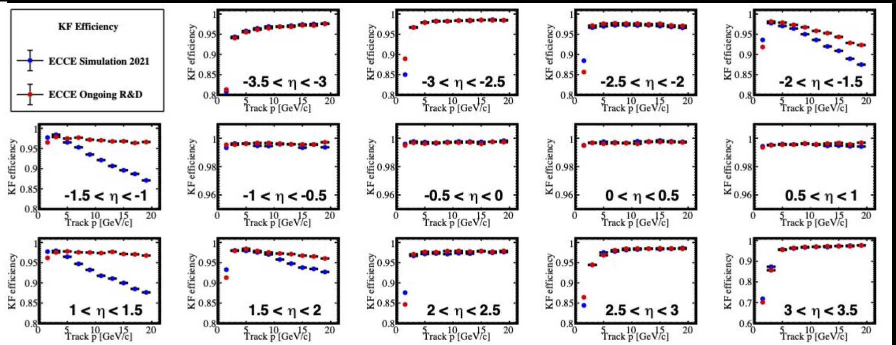


3 At each point in the Pareto front corresponds a design



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





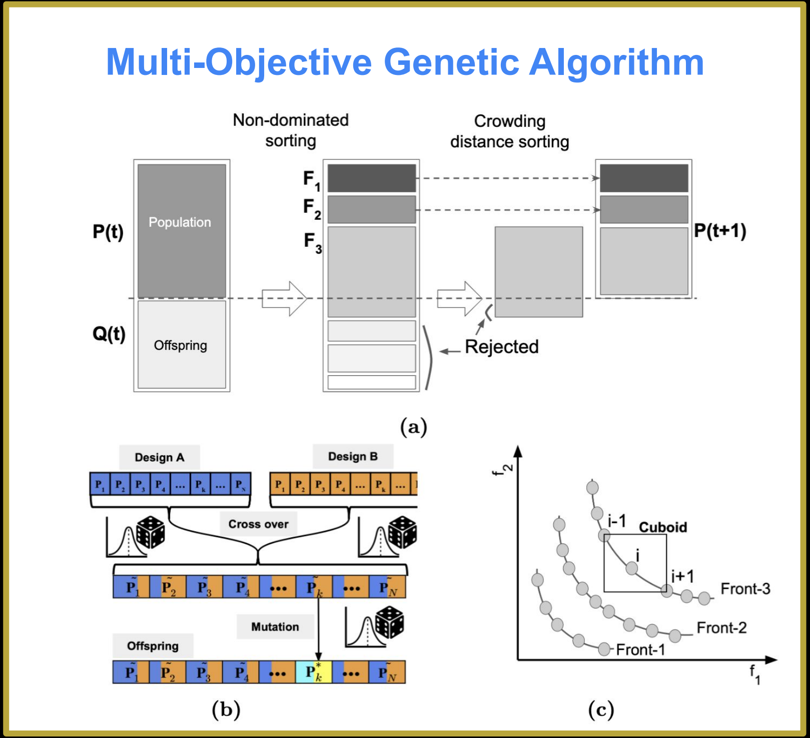
MOO Pipelines: MOGA

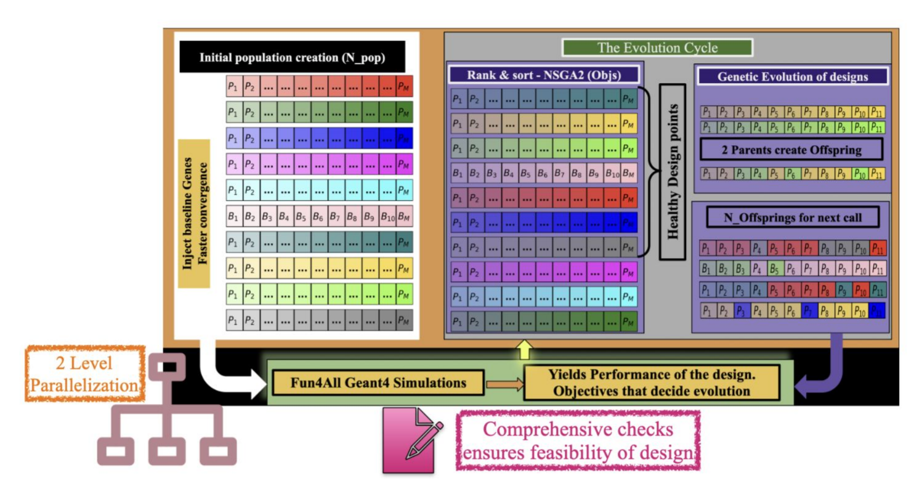
CF et al., NIM-A Vol 1047, Feb 2023, 167748

arXiv:2205.09185

- Engage with Open Source projects
- In a first exploratory phase*, we used MOGA (NSGA-II)

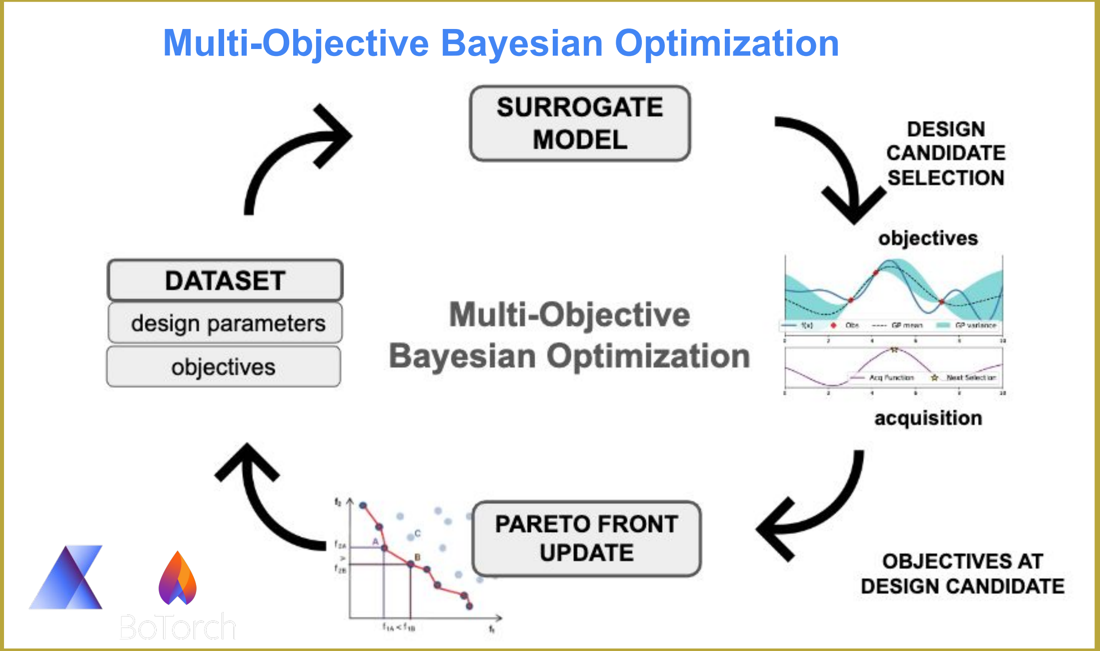
Description	Symbol	Value
Algorithm used	MOEA	NSGA-II
# Objectives	M	3
# Offspring	O	30(50)
# design Parameters	D	11
# calls (tot. budget)	-	200
# Cores	-	same as offspring
# Charged π^- tracks	N_{trk}	80k
Population Size	N	100





MOO Pipelines: MOBO

- As the project evolved, so did our understanding of the design space and the possible ranges for each design parameter.
- With MOBO, we aim to determine a more accurate approximation of the Pareto front



- Using Ax/BoTorch and novel qNEHVI acq. function with improved computational performance arXiv:2105.08195
- Currently in the process of generalizing the design problem and increase its complexity

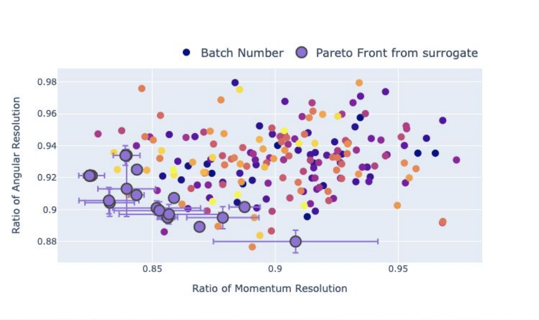
(e.g., Ax: adaptive experimentation platform supported by Meta AI)

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

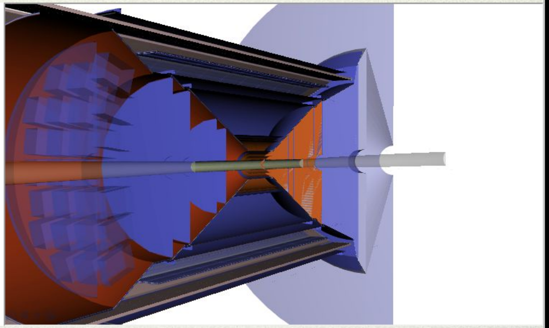
Navigate the Pareto front

C.Fanelli et al, NIM A, 2023, 167748, arXiv:2205.09185

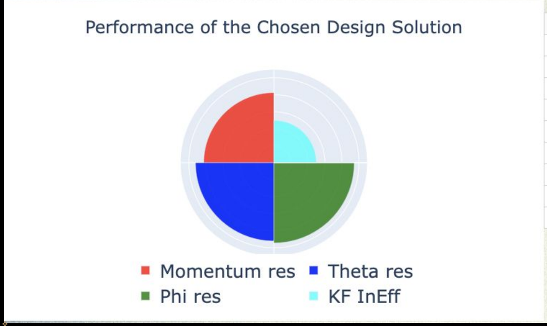
Multi Objective Bayesian Optimization



GEANT4 Visualization of the design



Click on petals for finer evaluations

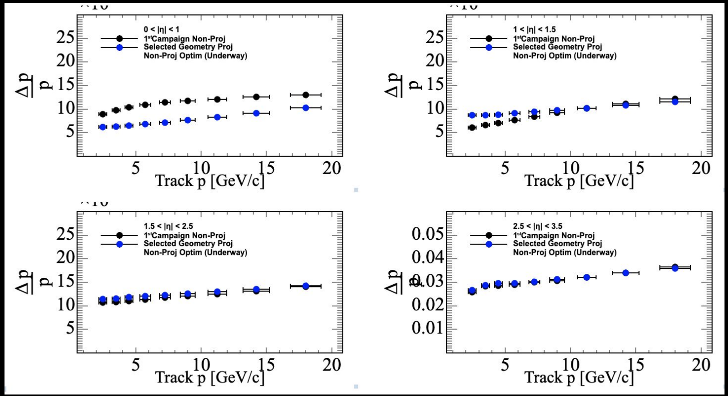


Design Parameters Table

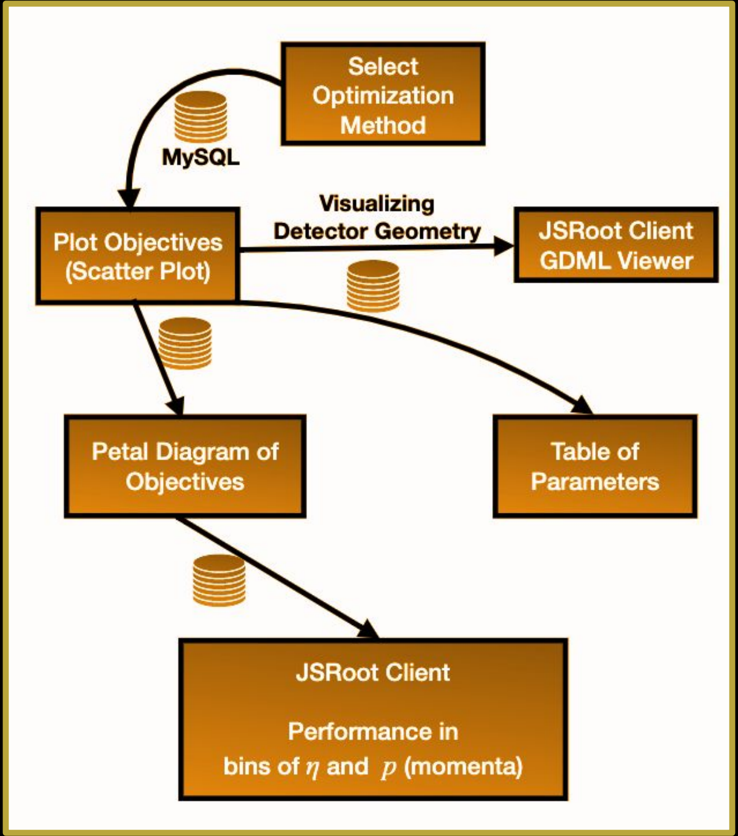
Parameter Name	Parameter Value
Angle of cone [deg]	25.00
Radius of uRwell-1 [cms]	32.47
z E-TTL [cms]	171.00
z F-TTL [cms]	157.60
z EST-1 [cms]	40.39
z EST-3 [cms]	85.09
z FST-1 [cms]	35.03
z FST-3 [cms]	83.78
z FST-5 [cms]	131.27

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

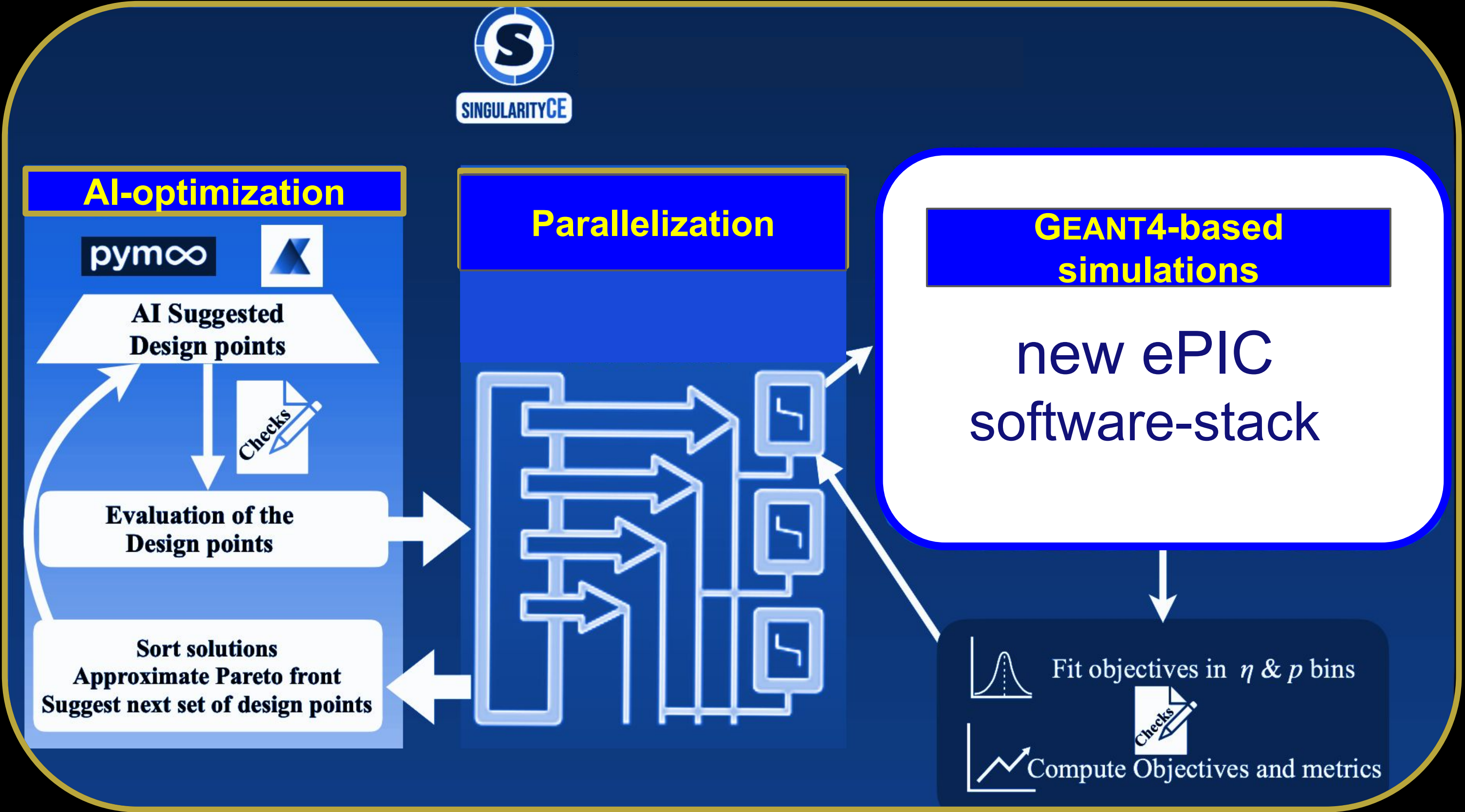


Visualization



- 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

Workflow utilized in the ECCE proposal (2022)

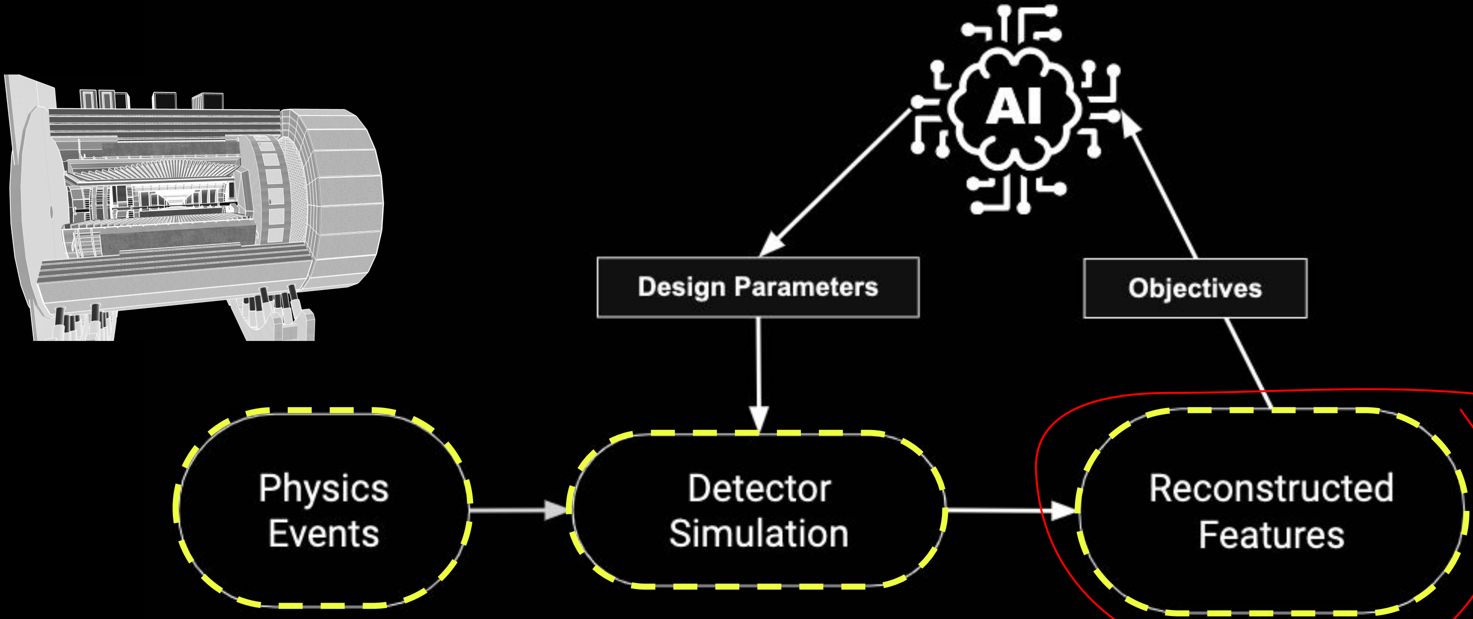


*Future implementations will explore a scalable and distributed AI-assisted design framework

ePIC SW Features Streamlining AI-Assisted Design

- Design:
  - Geometry implementation via data source makes transparent the coupling of AI to the software stack design parameters
  - Modularity of geometry description
  - Automated features (checking overlaps)
- General Properties:
  - Code repository, continuous integration, containerization
  - Open, simple, self-descriptive data formats (flat data model in general allows flexibility for AI/ML applications)
  - Support for truth information
  - Use of ML-supported packages (e.g., ACTS, includes ONNX plugin)
  - JANA2 with integrated Python interface
  - etc.

AI-assisted design



See example (3)

Event level

⇑

Particle level

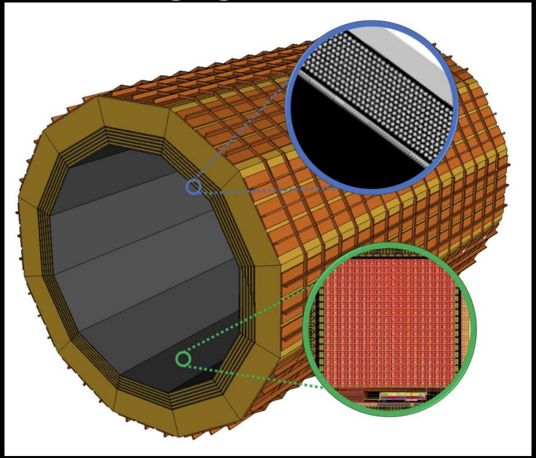
See examples (1) and (2)

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

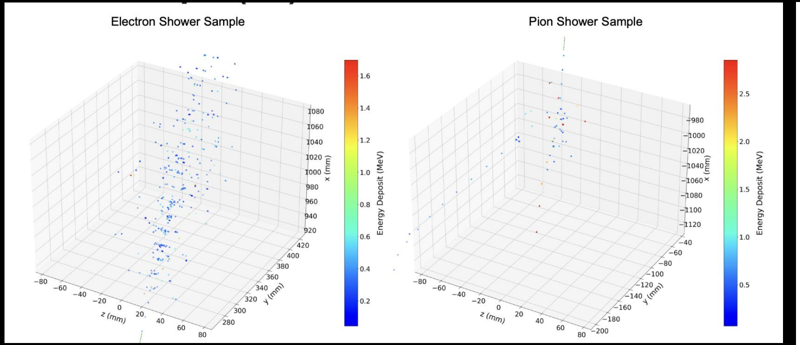
Jefferson Lab
Exploring the Nature of Matter

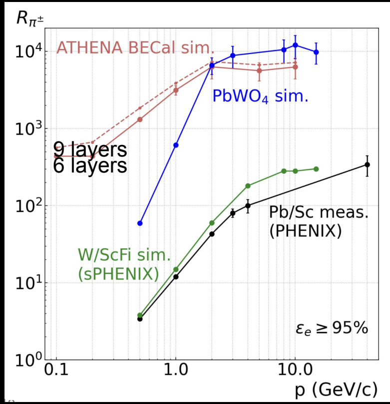
(1) ML-based PID for Shower Imaging

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

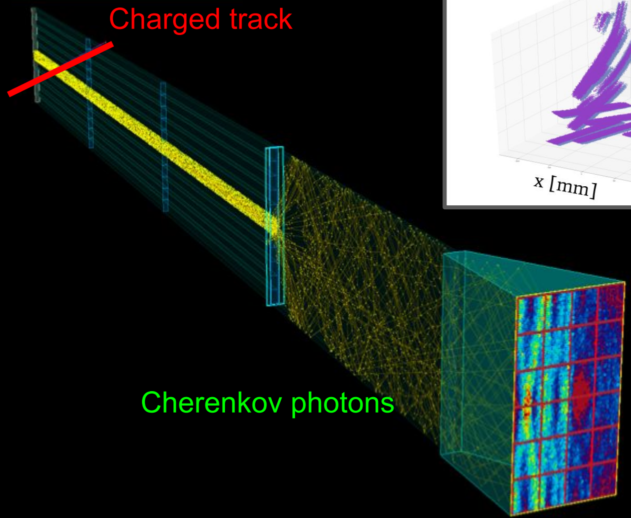
- Monolithic Silicon Sensors AstroPix
- Scintillating fibers embedded in Pb (Pb/ScFi similar to GlueX Barrel Ecal)
- “Sandwiched” 6 layers of AstroPix and 5 layers of Pb/ScFi (~1X0) followed by a large chunk of Pb/ScFi
- Total thickness ~43 cm (~21 X0)
- Large amount of data (3D shower imaging)

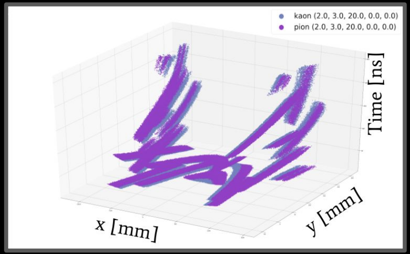
- ML with shower imaging significantly improves e/π rejection compared to traditional E/p cut — impact on DIS
- Separation of γ’s from π^0 at high momenta (40 GeV/c) and precise position reconstruction of γ’s (<1 mm at 5 GeV) — DVCS and γ physics
- Tagging final state radiative γ’s from nuclear/nucleon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

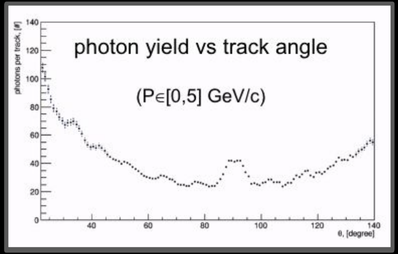
[1] N. Apadula, et al. "Monolithic active pixel sensors on cmos technologies." arXiv preprint arXiv:2203.07626 (2022).
[2] C. Peng, ML Particle Identification with Measured Shower Profiles from Calorimetry, AI4EIC 2nd workshop (2022)

(2) ML-based PID for Cherenkov

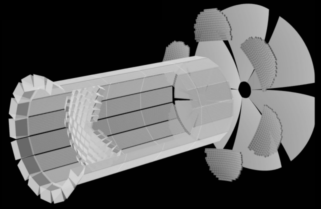
DIRC at GlueX is instrumental for PID



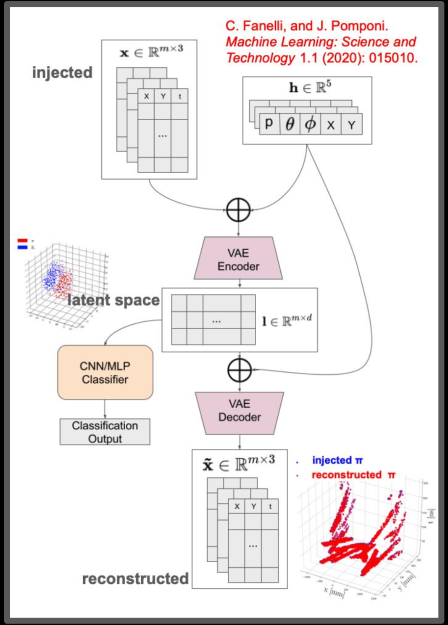




Cherenkov detectors will be the backbone of PID at EIC



DeepRICH



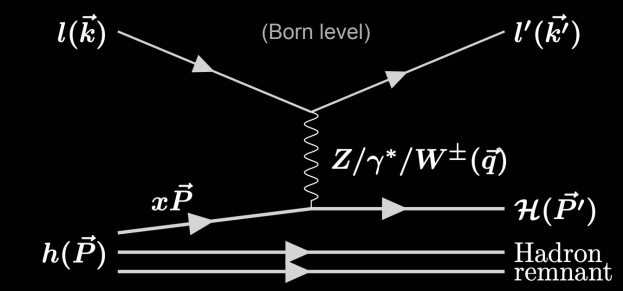
- Need for faster and accurate simulations and reconstruction
- Complex hit patterns (DIRC is the most complex), sparse data, response vs kinematics
- DeepRICH: same reconstruction performance of best reconstruction algorithm with ~4 orders of magnitude speed-up in inference time on GPU
- Possibility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

Jefferson Lab
Exploring the Nature of Matter

[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.

(3) Deeply Learning DIS

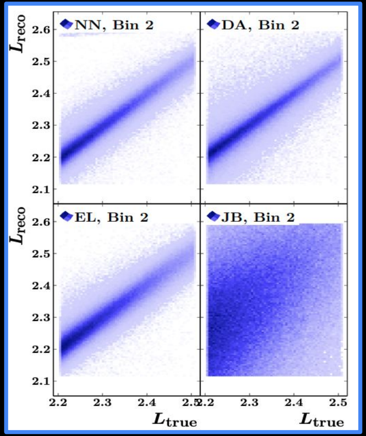
DIS fundamental process @EIC



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 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

Example in one specific bin



Bin	Events	Resolution of $\log x$, $\times 10^3$		Resolution of $\log Q^2/1\,\text{GeV}^2$, $\times 10^3$	
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 $\log(x) - \log(x_{\text{true}})$ and $\log(Q^2) - \log(Q^2_{\text{true}})$ respectively.

[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

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

Jefferson Lab
Exploring the Nature of Matter

Forthcoming AI4EIC workshop from Nov 28 to Dec 1, 2023 — CUA, Washington, D.C.

Conclusions

- **Next generation QCD experiments like ePIC are being designed during the AI revolution:**
  - AI can assist the design and R&D (two phases, slightly different needs) of complex experimental systems
    - providing more efficient design (considering multiple objectives)
    - utilizing effectively the computing resources needed to achieve that.
- **EIC will be one of the first large-scale experiments (involved hundreds of institutions world-wide) to be designed with the support of AI**
  - The reference detector has been already designed taking advantage of a multi-objective optimization approach and a complex parametrization of its design which takes into account constraints.
  - The optimization framework utilizes accurate full simulations based on Geant4, identify the tradeoffs in the Pareto front (and reduce the total computational budget to converge to Pareto).
- **This workflow can be further extended for ePIC (and Detector-2)**
  - More realistic effects in the simulation and reconstruction techniques
  - A larger system of sub-detectors, e.g, detectors like the dRICH — increased complexity, e.g.
    - $\mathcal{O}(100)$ pars, 4-8 objectives, $\mathcal{O}(10)$ constraints
    - Room for ideas, e.g., explore physics-inspired acquisition functions — exciting work ahead!!