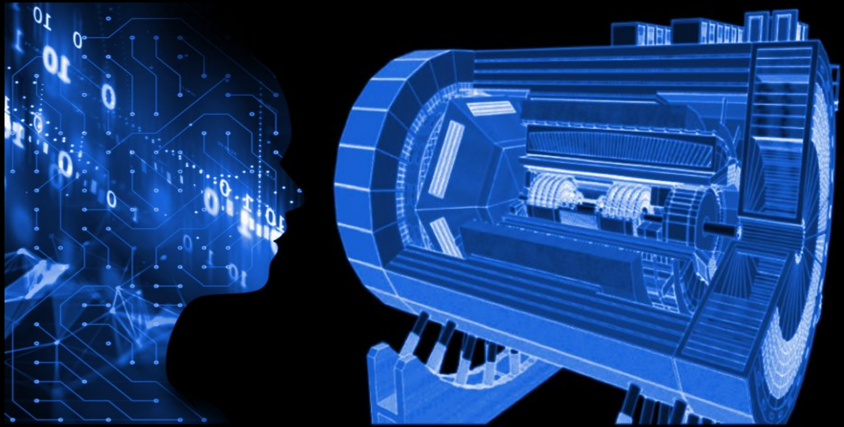


AI-assisted Detector Design: Examples from EIC

Cristiano Fanelli

WILLIAM & MARY
CHARTERED 1693

Jefferson Lab
Exploring the Nature of Matter

AI4EIC

EIC²
EIC Center at Jefferson Lab

QNP2022 - 9th International Conference on Quarks and Nuclear Physics

AI for Design

It is a relatively new but active area of research.
Many applications in, e.g., industrial material, molecular and drug design.

Guo, Kai, et al. *Materials Horizons* 8.4 (2021): 1153-1172.

Table 1 Popular ML methods in design of mechanical materials

ML method	Characteristics	Example applications in mechanical materials design
Linear regression; polynomial regression	Model the linear or polynomial relationship between input and output variables	Modulus[112] or strength[123] prediction
Support vector machine; SVR	Separate high-dimensional data space with one or a set of hyperplanes	Strength[123] or hardness[125] prediction; structural topology optimization[159]
Random forest	Construct multiple decision trees for classification or prediction	Modulus[112] or toughness[130] prediction
Feedforward neural network (FFNN); MLP	Connect nodes (neurons) with information flowing in one direction	Prediction of modulus,[97,112] strength,[93] toughness[130] or hardness;[97] prediction of hyperelastic or plastic behaviors;[143,145] identification of collision load conditions;[147] design of spinodoid metamaterials[163]
CNNs	Capture features at different hierarchical levels by calculating convolutions; operate on pixel-based or voxel-based data	Prediction of strain fields[104,105] or elastic properties[102,103] of high-contrast composites, modulus of unidirectional composites,[136] stress fields in cantilevered structures,[137] or yield strength of additive-manufactured metals;[121] prediction of fatigue crack propagation in polycrystalline alloys;[140] prediction of crystal plasticity;[120] design of tessellate composites;[107–109] design of stretchable graphene kirigami;[155] structural topology optimization[156–158]
Recurrent neural network (RNN); LSTM; GRU	Connect nodes (neurons) forming a directed graph with history information stored in hidden states; operate on sequential data	Prediction of fracture patterns in crystalline solids;[114] prediction of plastic behaviors in heterogeneous materials;[142,144] multi-scale modeling of porous media[173]
Generative adversarial networks (GANs)	Train two opponent neural networks to generate and discriminate separately until the two networks reach equilibrium; generate new data according to the distribution of training set	Prediction of modulus distribution by solving inverse elasticity problems;[138] prediction of strain or stress fields in composites;[139] composite design;[164] structural topology optimization;[165–167] architected materials design[115]
Gaussian process regression (GPR); Bayesian learning	Treat parameters as random variables and calculate the probability distribution of these variables; quantify the uncertainty of model predictions	Modulus[122] or strength[123,124] prediction; design of supercompressible and recoverable metamaterials[110]
Active learning	Interacts with a user on the fly for labeling new data; augment training data with post-hoc experiments or simulations	Strength prediction[124]
Genetic or evolutionary algorithms	Mimic evolutionary rules for optimizing objective function	Hardness prediction;[126] designs of active materials;[160,161] design of modular metamaterials[162]
Reinforcement learning	Maximize cumulative awards with agents reacting to the environments.	Deriving microstructure-based traction-separation laws[174]
Graph neural networks (GNNs)	Operate on non-Euclidean data structures; applicable tasks include link prediction, node classification and graph classification	Hardness prediction;[127] architected materials design[168]

Z. Zhou et al., *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019

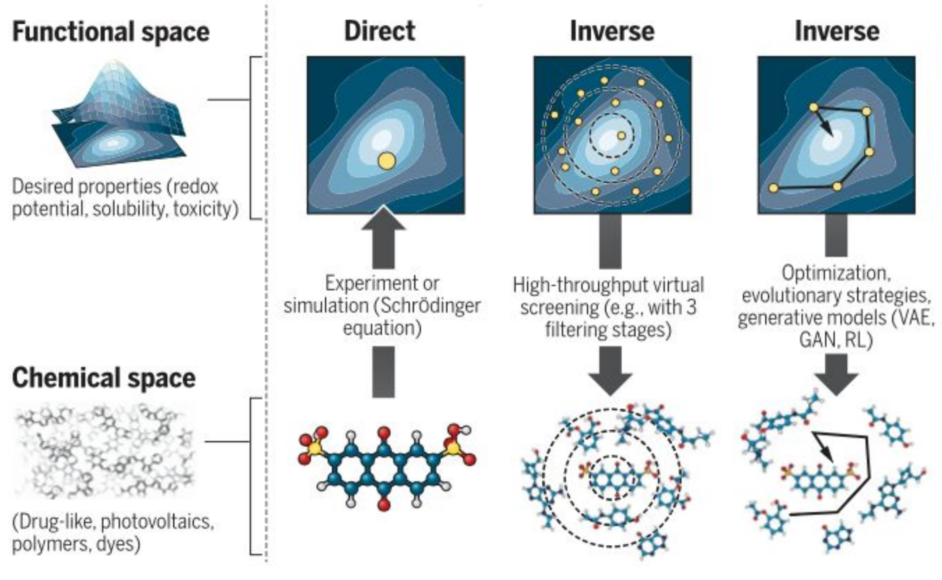


Fig. 2. Schematic of the different approaches toward molecular design. Inverse design starts from desired properties and ends in chemical space, unlike the direct approach that leads from chemical space to the properties.

B. Sanchez-Lengeling, A. Aspuru-Guzik. *Science* 361.6400 (2018): 360-365.

AI-assisted Detector Design

When it comes to designing detectors with AI this is a frontier topic with few examples in the literature.

S. Shirobokov, V. Belavin, M. Kagan, A. Ustyuzhanin, and A.G. Baydin. Black-Box Optimization with Local Generative Surrogates, 2020. arXiv: 2002.04632.

T. Dorigo. Geometry optimization of a muon-electron scattering detector. Physics Open, 4:100022, 2020.

F. Ratnikov. Using machine learning to speed up and improve calorimeter R&D. Journal of Instrumentation, 15(05):C05032, 2020.

E. Cisbani, CF, et al. AI-optimized detector design for the future Electron Ion Collider: the dual-radiator RICH case. JINST 15(05):P05009, 2020.

S. Meyer et al. Optimization and performance study of a proton CT system for pre-clinical small animal imaging. Phys. Med. Biol., 65(15):155008, 2020.

CF, et al. (ECCE), AI-assisted Optimization of the ECCE Tracking System at the Electron Ion Collider arXiv:2205.09185, 2022

- For years the full detector design has been studied after the subsystem prototypes are ready (taking into account the phase constraints from the full detector or outer layers). We need to use advanced simulations which are computationally expensive (Geant).
- Modern complex design: many parameters (and multiple objective functions) — curse of dimensionality. AI-assisted strategies can help designing more efficiently (in terms of performance and resources needed).

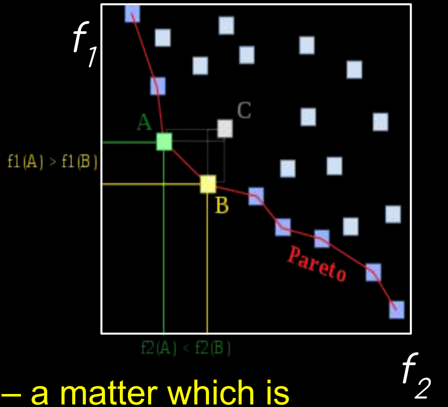
https://cfteach.github.io/nnpss/ — NNPSS, MIT 2022

https://indico.bnl.gov/event/16328/ — Meeting @AI WG Detector Design

In this talk I will use the future “EPIC” detector at EIC as an example https://wiki.bnl.gov/EPIC/

(starting from the experience of the detector proposal “ECCE” https://www.ecce-eic.org/)

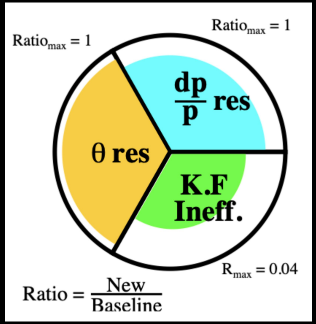
Multi-Objective Optimization



- The problem becomes challenging when the objectives are of conflict to each other, that is, the optimal solution of an objective function is different from that of the other.
- In solving such problems, with or without constraints, they give rise to a trade-off optimal solutions, popularly known as Pareto-optimal solutions.
- MO-based solutions are helping to reveal important hidden knowledge about a problem – a matter which is difficult to achieve otherwise
- In this talk we use both evolutionary (1) and bayesian approaches (2).
  - During the proposal phase we utilized (1), which relies on a population approach in its search procedure.
  - After the proposal we implemented and started utilizing (2).

The ECCE Tracker Design Optimization considers simultaneously:

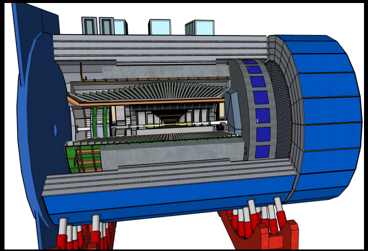
- momentum resolution
- angular resolution
- Kalman filter efficiency
- (pointing resolution)
- Mechanical constraints



The Typical Workflow

See talk at IAEA Technical Meeting on AI

IAEA
International Atomic Energy Agency

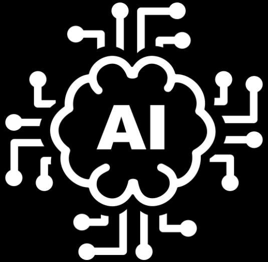


- AI can assist in designing more efficiently detectors (performance, costs).
- It helps steering the design (and eventually fine-tune it).
- It can capture hidden correlations among design parameters.

A.I.
gathers observations and suggests new points

customization

Design parameters



Physics Events

Detector Simulation

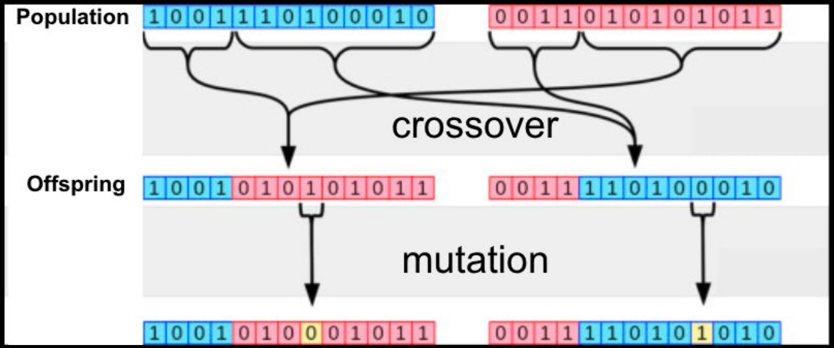
Analysis of High-level reconstruction of events

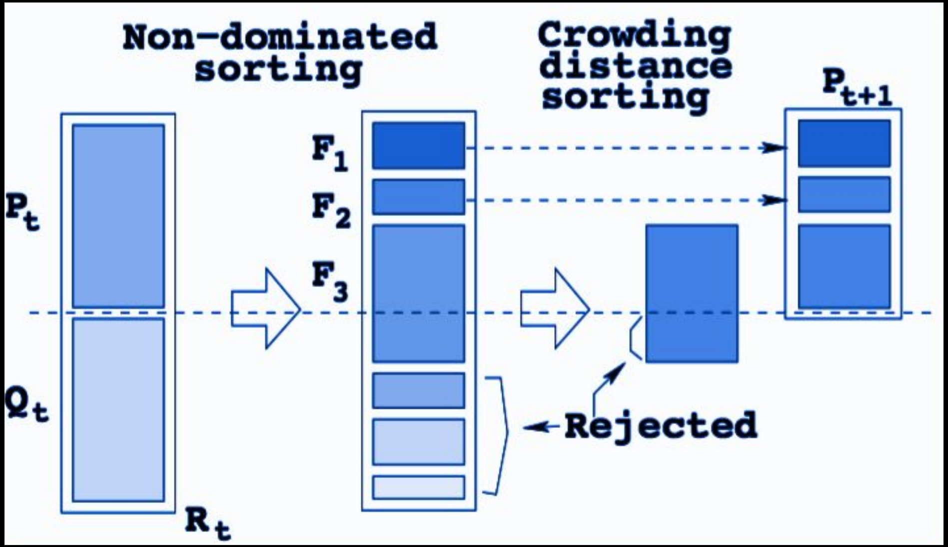
compute intensive (Geant4)

Forward simulations needed to simulate quantum phenomena (interaction of particles with matter)

Popular AI-Strategies (in a nutshell)

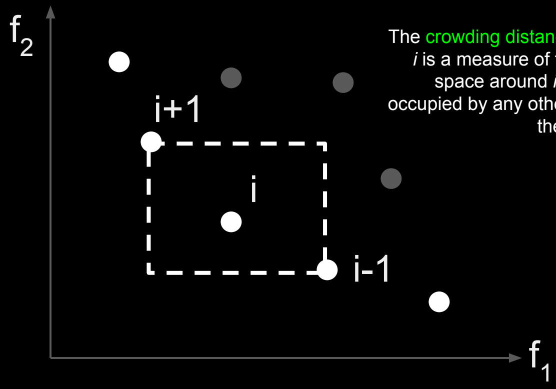
Evolutionary





This is one of the most popular approach, characterized by:

- Use of an elitist principle
- Explicit diversity preserving mechanism
- Emphasis in non-dominated solutions



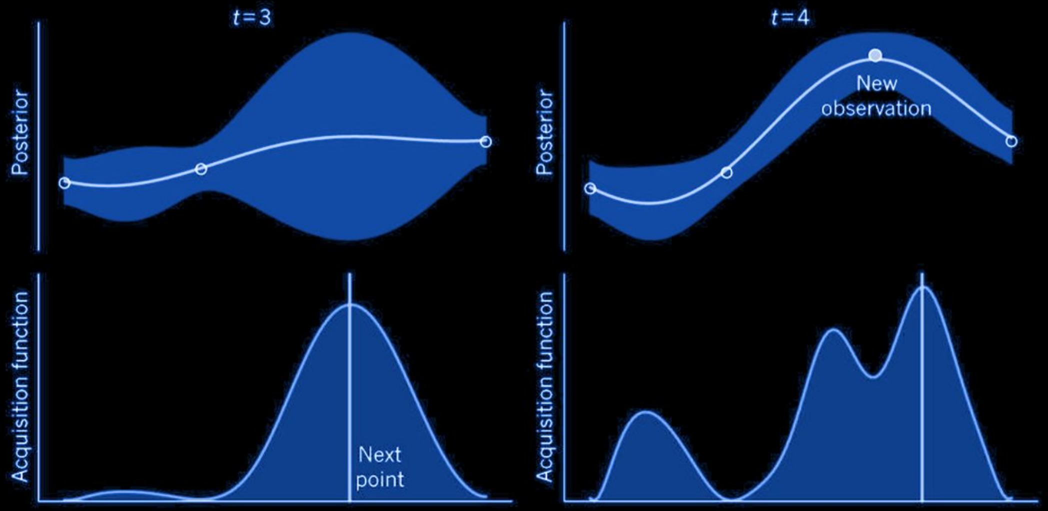
The crowding distance d_i of point *i* is a measure of the objective space around *i* which is not occupied by any other solution in the population.

The population R_t is classified in non-dominated fronts.
Not all fronts can be accommodated in the N slots of available in the new population P_{t+1}. We use crowding distance to keep those points in the last front that contribute to the highest diversity.

Popular AI-Strategies (in a nutshell)

Bayesian

- BO is a sequential strategy developed for global optimization.
- After gathering evaluations we builds a posterior distribution used to construct an **acquisition function**.
- This cheap function determines what is **next query point**.



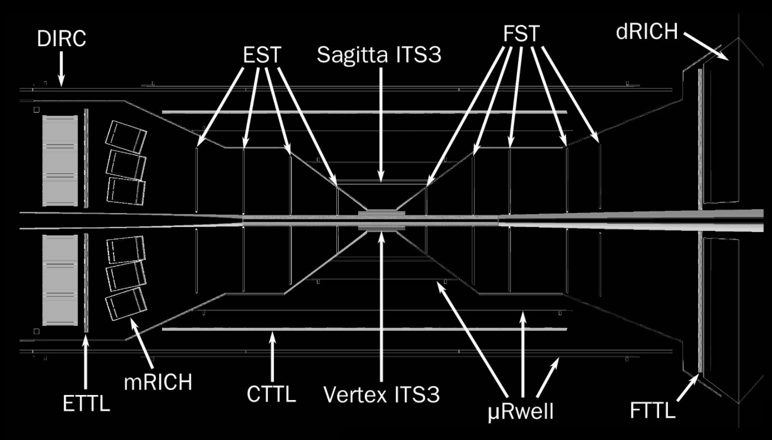
1. Select a Sample by Optimizing the Acquisition Function.
2. Evaluate the Sample With the Objective Function.
3. Update the Data and, in turn, the Surrogate Function.
4. Go To 1.

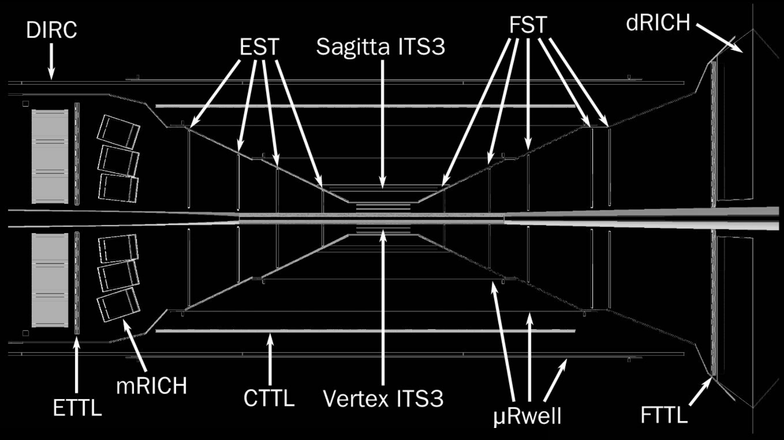
Extension to multiple objectives…

ECCE Tracker: Reference VS Projective

arXiv:2205.09185

Parametrization underlies the AI-assisted design and can explore non-projective as well as projective





non-projective

projective (ongoing R&D)

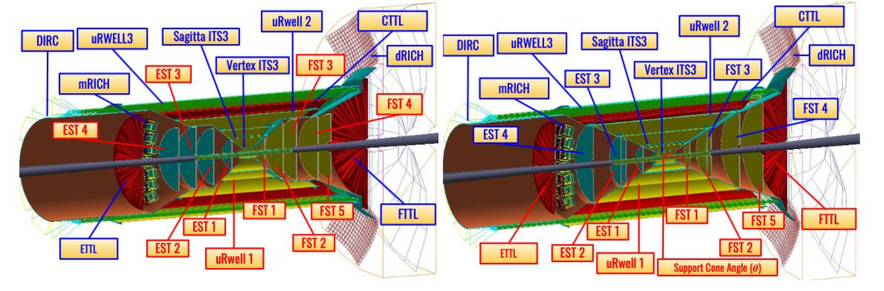


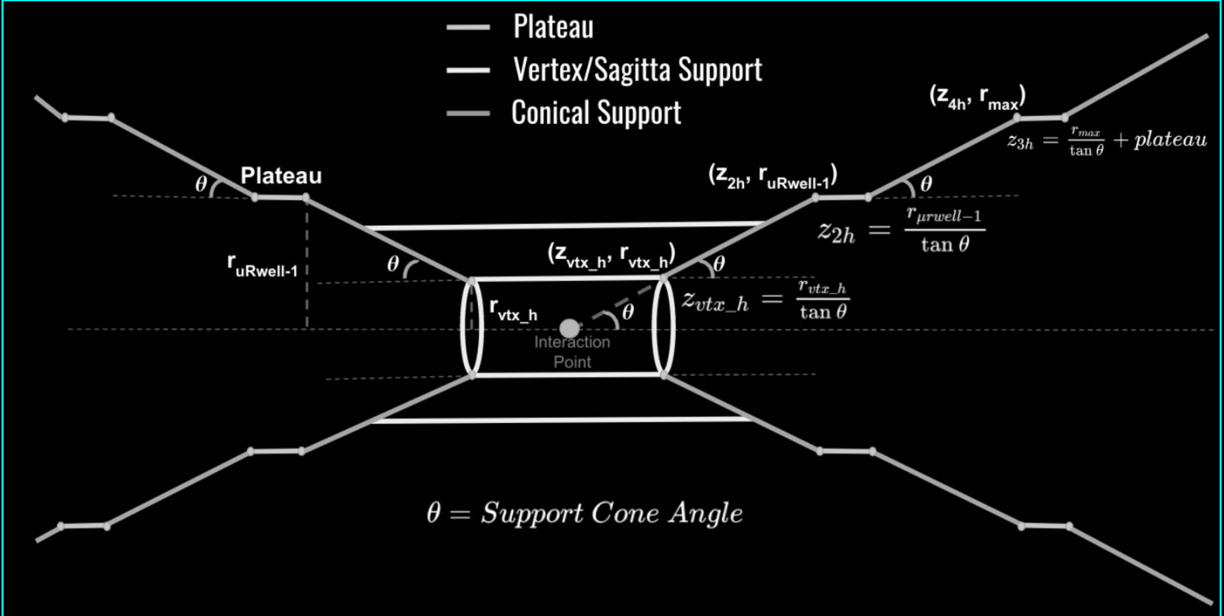
Figure 5: **Tracking and PID system in the non-projective (left) and the ongoing R&D projective (right) designs:** the two figures show the different geometry and parametrization of the ECCE non-projective design (left) and of the ongoing R&D projective design to optimize the support structure (right). Labels in red indicate the sub-detector systems that were optimized, while the labels in blue are the sub-detector systems that were kept fixed due to geometrical constraint. The non-projective geometry (left) is a result of an optimization on the inner tracker layers (labeled in red) while keeping the support structure fixed, The angle made by the support structure to the IP is fixed at about 36.5°. The projective geometry (right) is the result of an ongoing project R&D to reduce the impact of readout and services on tracking resolution.

CF, et al. (ECCE), AI-assisted Optimization of the ECCE Tracking System at the Electron Ion Collider arXiv:2205.09185, 2022

ECCE Tracker: Parametrization

arXiv:2205.09185

Parametrization of the support structure



Parametrization of
Disks, tracking layers, TTL

Geometric Constraints

Disks: r_{max} and r_{min} are calculated based on the support structure.

Sagitta: length fixed and radius changed based on the cone angle.

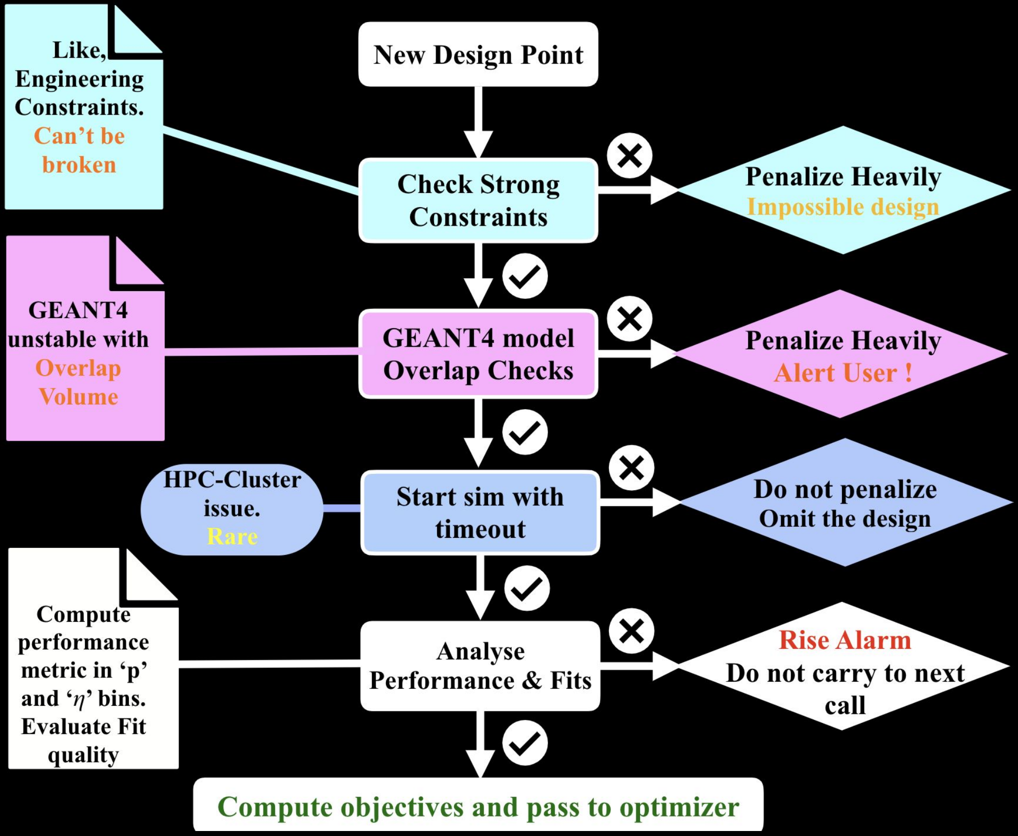
Parametrization underlies the AI-assisted design and can explore non-projective as well as projective

CF, et al. (ECCE), AI-assisted Optimization of the ECCE Tracking System at the Electron Ion Collider arXiv:2205.09185, 2022

"Soft"/"Strong" Constraints and Checks

$$
\begin{aligned}
& \min \mathbf{f}_\mathrm{m}(\mathbf{x}) && m = 1, \cdots, M \\
s.t. \quad & \mathbf{g}_\mathrm{j}(\mathbf{x}) \le 0, && j = 1, \cdots, J \\
& \mathbf{h}_\mathrm{k}(\mathbf{x}) = 0, && k = 1, \cdots, K \\
& x_i^L \le x_i \le x_i^U, && i = 1, \cdots, N
\end{aligned}
$$

sub-detector	constraint	description
EST/FST disks	$min\left\{\sum_i^{disks}\left\lvert\frac{R^i_{out}-R^i_{in}}{d}-\left\lfloor\frac{R^i_{out}-R^i_{in}}{d}\right\rfloor\right\rvert\right\}$	**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	$min\left\{\left\lvert\frac{2\pi r_{sagitta}}{w}-\left\lfloor\frac{2\pi r_{sagitta}}{w}\right\rfloor\right\rvert\right\}$	**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

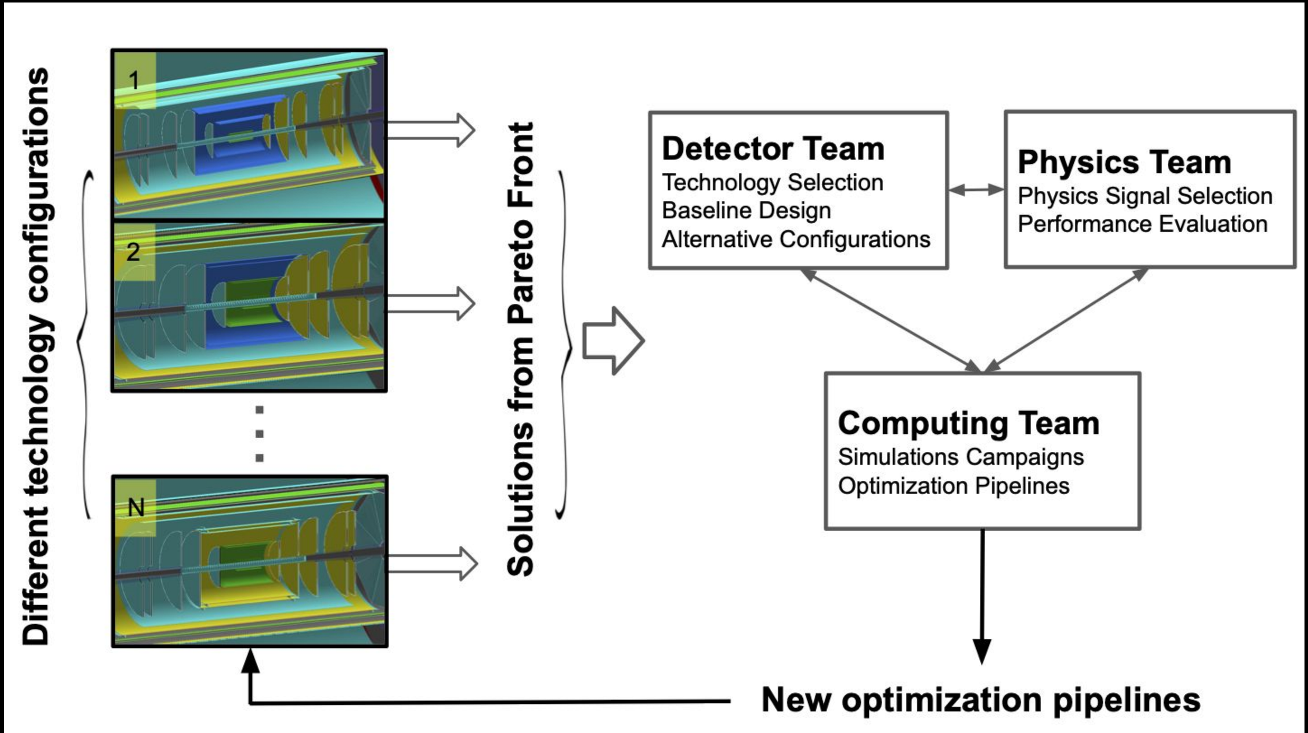


Integration during the EIC Detector Proposal

AI-"Optimization" does not necessarily mean "fine-tuning"

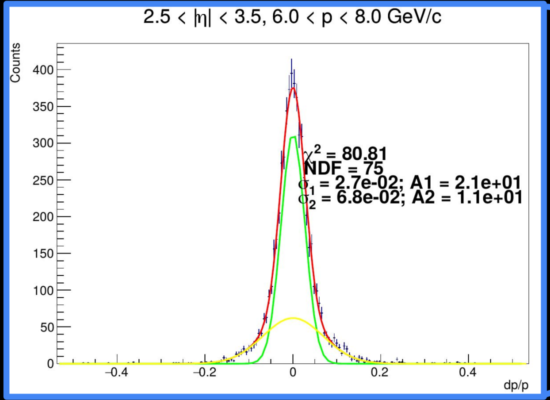
- We want to use these algorithms to: (1) steer the design and suggest parameters that a "manual"/brute-force optimization will likely miss to identify; (2) further optimize some particular detector technology (see d-RICH paper, e.g., optics properties)
- AI allows to capture hidden correlations among the design parameters.
- All "steps" (physics, detector) involved in the AI optimization, strong interplay between working groups

Light/smart optimization pipelines ran during the "explorative" phase of the detector proposal

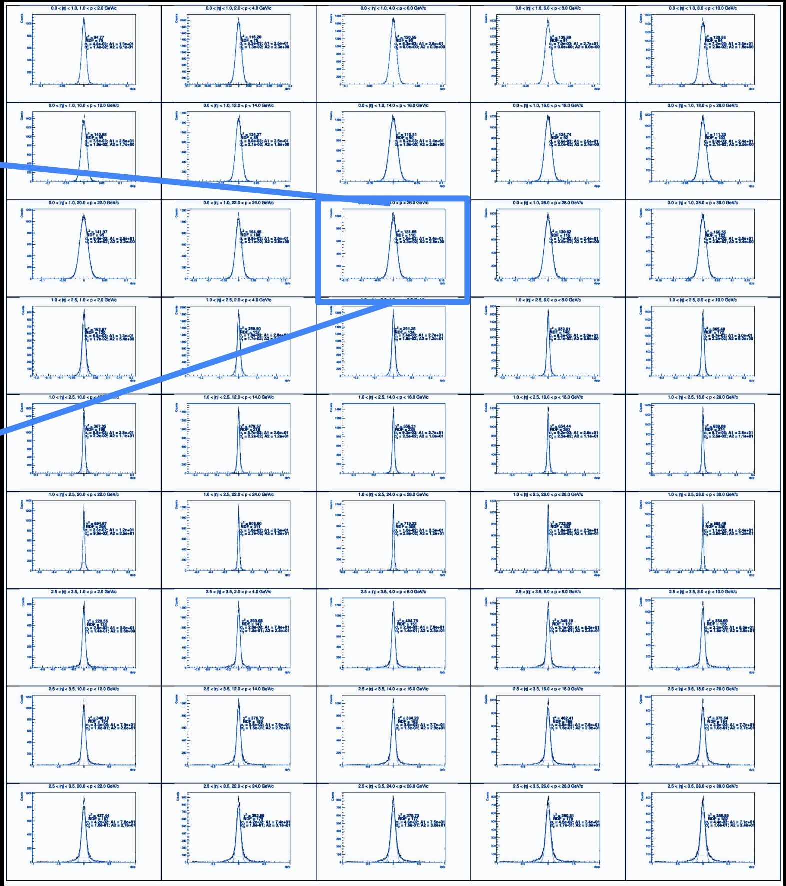


Implementation

- Objectives evaluated in fine-grained phase-space
- Propagate uncertainties from fits



Weighted sum with errors



Weighted sum with errors

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

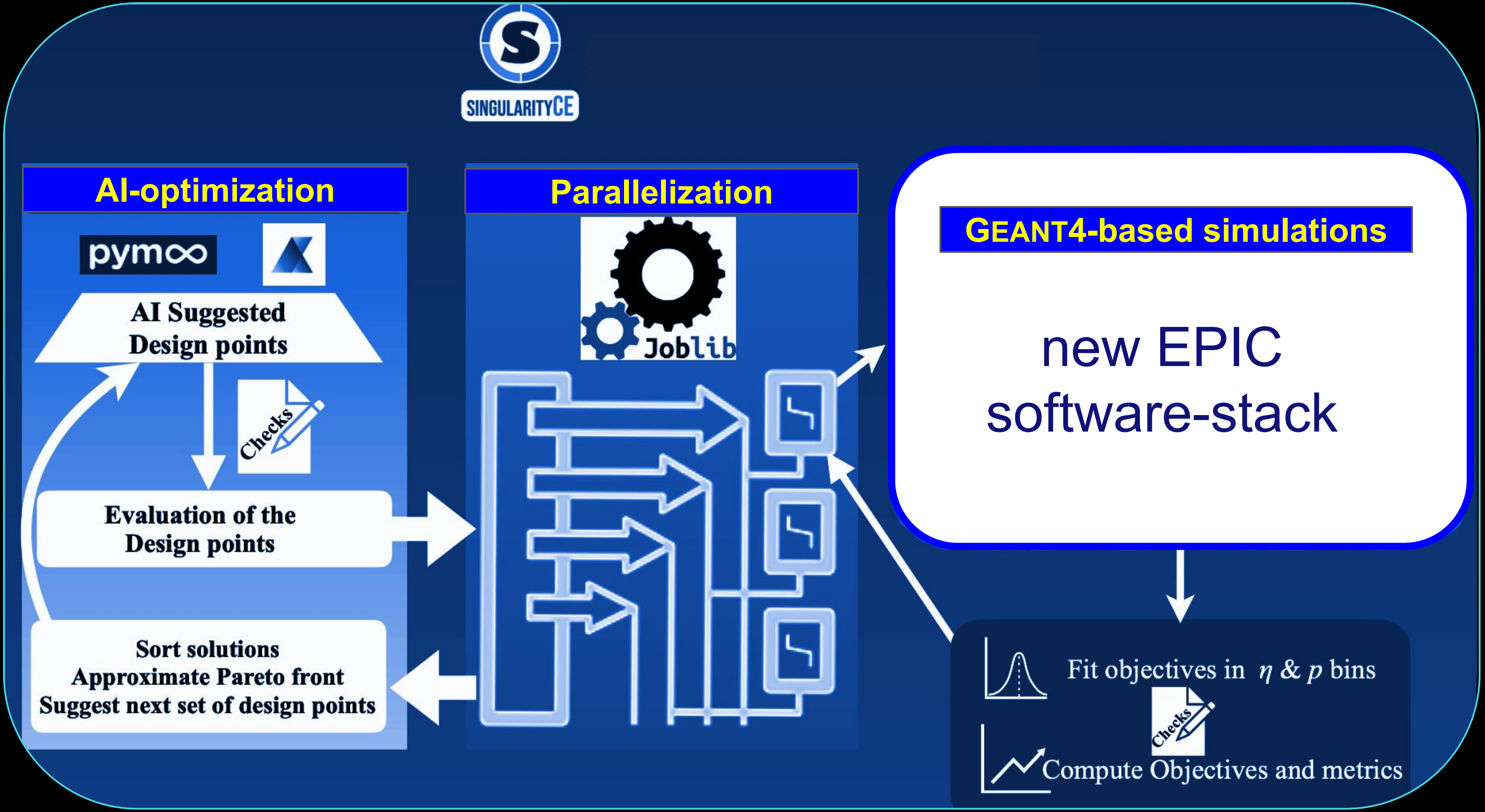
(sum in bins of 14 bins of P)

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

(Average objective in a η bin)

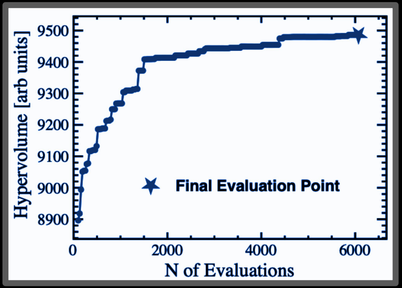
$$\Rightarrow \quad R(f) = \frac{1}{N_\eta} \sum_\eta \left(\frac{\sum_p w_{p,\eta} \cdot R(f)_{p,\eta}}{\sum_p w_{p,\eta}} \right)$$

Workflow during proposal and beyond

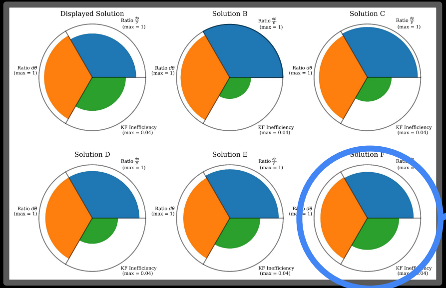


"Navigate" Pareto Front

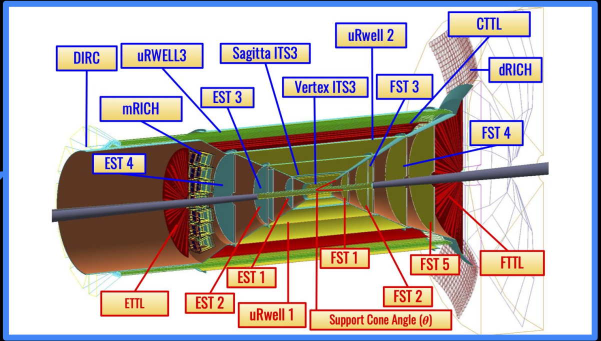
1 Can take a snapshot any time during evaluation



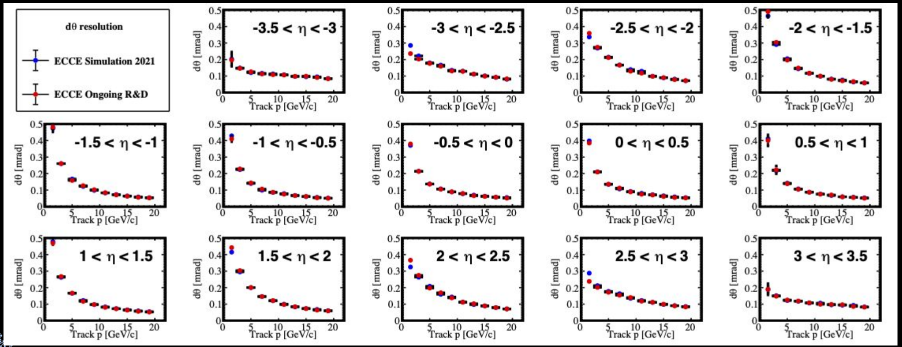
2 Updated Pareto Front at time t

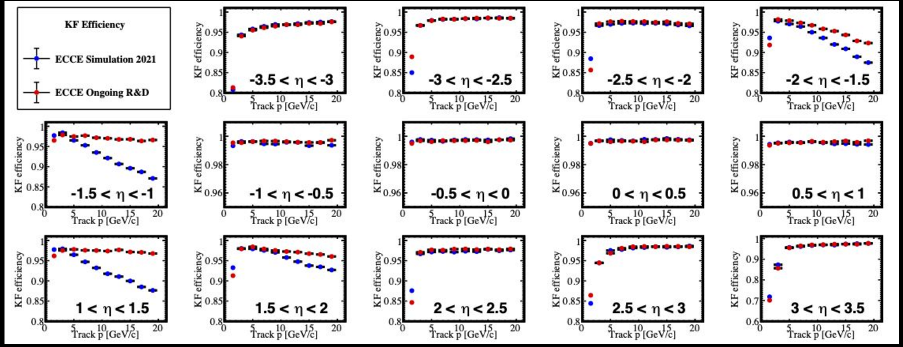


3 At each point in the Pareto front corresponds a design



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

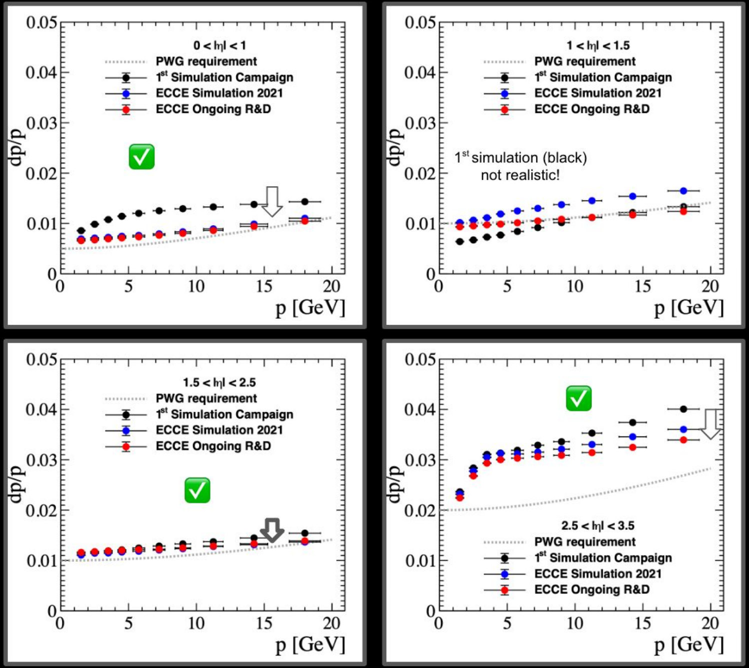




Evolution and Validation

Evolution

- Black points: first simulation campaign, a preliminary detector concept in phase-I optimization with no developed support structure;
- Blue: fully developed simulations for final ECCE detector proposal; Red: the ongoing R&D for the optimization of the support structure.
- There is an improvement in performance in all η bins with the exception of the transition region, an artifact of the fact that black do not include a realistic simulation of the material budget in the transition region!
- In the transition region, it can be also appreciated the improvement provided by the projective design



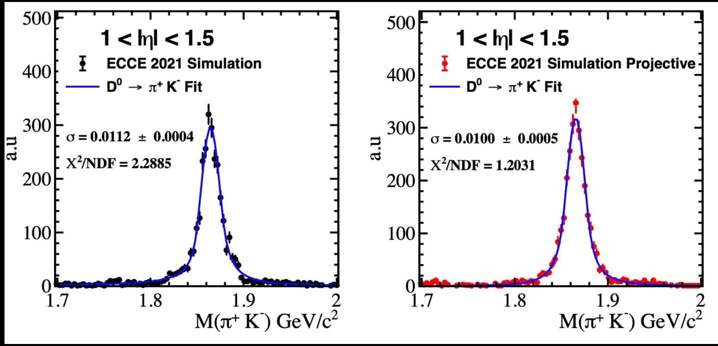
Validation

Observables not directly used in the optimization as objectives

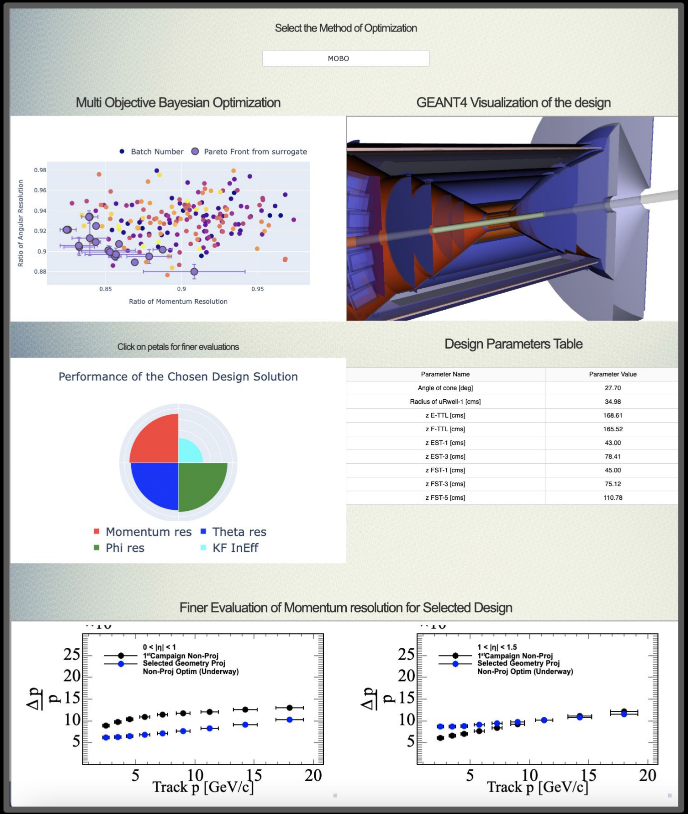
Performance evaluated after optimization process (both designs) using standard analysis procedures

Notice red points are related to an ongoing project R&D with a projective support structure for the ECCE tracker.

E.g., D0 invariant mass from semi-inclusive deep inelastic scattering



Interactive Navigation of Pareto front



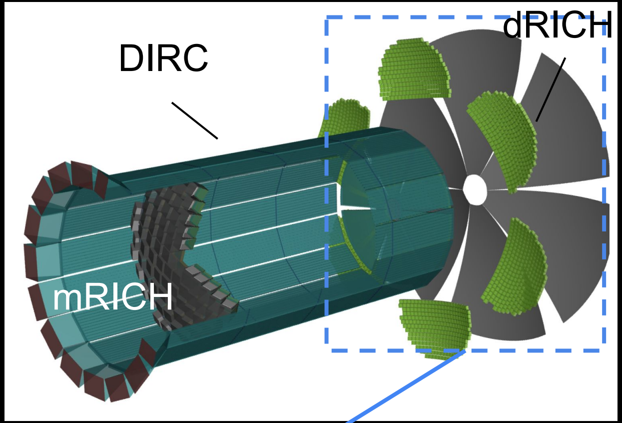
https://ai4eicdetopt.pythonanywhere.com

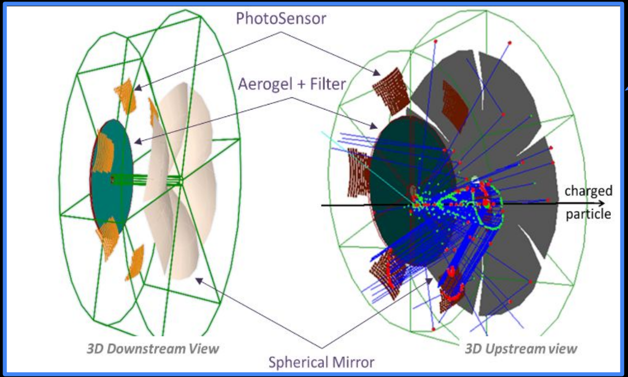
- Use cutting-edge data science tools for visualization of results from approximated Pareto front
- Exploration in a multiple objective space
- Facilitate study/comparison of trade-off solutions
- Here MOBO is used using BoTorch/Ax (benefit from strong community support — Meta/Facebook)

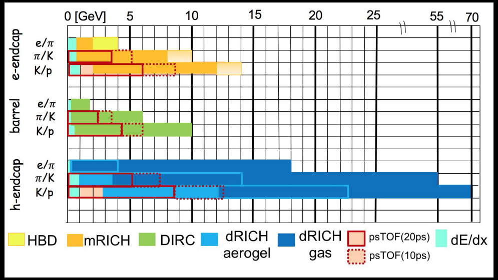
Credits: K. Suresh (U. of Regina)

Extension to larger system of sub-detectors

- Cherenkov detectors are essential part of the PID system of EPIC
  - Simulating these detectors is typically compute expensive, involving many photons that need to be tracked through complex surfaces.
  - All of them rely on pattern recognition of ring images in reconstruction, and the DIRC is the one having the more complex ring patterns!
- Extension of design optimization to tracker + PID system
  - Potential to optimize parameters of the dRICH design in the hadronic endcap
- E.g., dRICH design
  - Large momentum coverage
  - Two radiators: aerogel and gas
  - Legacy design from INFN
    - 6 Identical open sectors
    - Large focusing mirror



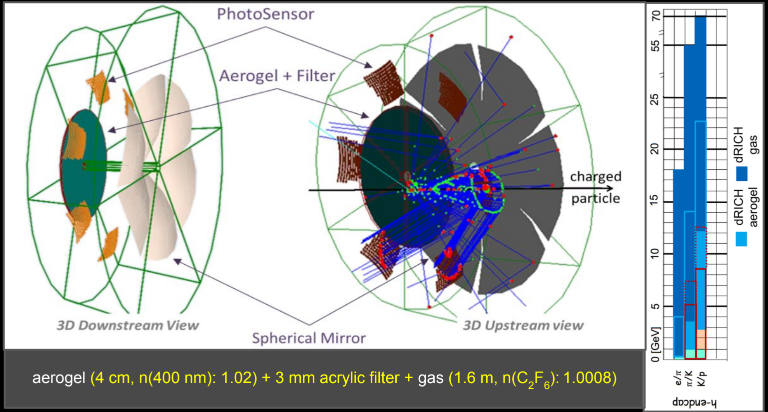




E. Cisbani, A. Del Dotto, CF*, M. Williams et al. "AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case." *JINST* 15.05 (2020): P05009.

dRICH: ante-proposal

- Two radiators with different refractive indices for continuous momentum coverage.
- Simulation of detector and processes is compute-intensive
- Legacy design from INFN (EICUG2017).



aerogel (4 cm, n(400 nm): 1.02) + 3 mm acrylic filter + gas (1.6 m, n(C_2F_6): 1.0008)

1 Define design parametrization and space: optics + geometry

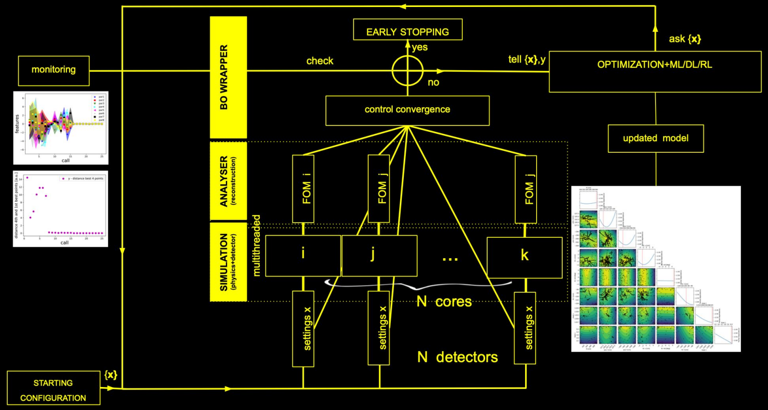
parameter	description	range [units]	tolerance [units]
R	mirror radius	[290,300] [cm]	100 [μm]
pos r	radial position of mirror center	[125,140] [cm]	100 [μm]
pos l	longitudinal position of mirror center	[-305,-295] [cm]	100 [μm]
tiles x	shift along x of tiles center	[-5,5] [cm]	100 [μm]
tiles y	shift along y of tiles center	[-5,5] [cm]	100 [μm]
tiles z	shift along z of tiles center	[-105,-95] [cm]	100 [μm]
$n_{aerogel}$	aerogel refractive index	[1.015,1.030]	0.2%
$t_{aerogel}$	aerogel thickness	[3.0,6.0] [cm]	1 [mm]

2 Come up with a smart objective; study / characterize properties (noise, stats needed etc): simulation + reconstruction

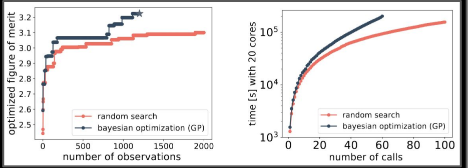
$$N\sigma = \frac{||\langle\theta_K\rangle - \langle\theta_\pi\rangle||\sqrt{N_\gamma}}{\sigma_\theta^{1p.e.}}$$

$$h = 2\cdot\left[\frac{1}{(N\sigma)|_1} + \frac{1}{(N\sigma)|_2}\right]^{-1}$$

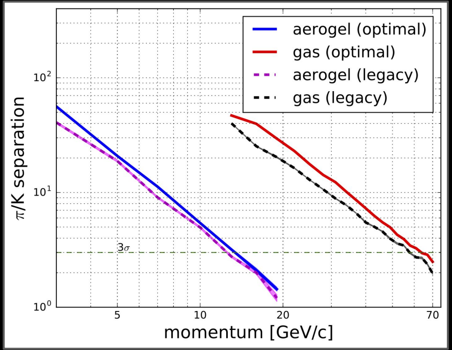
3 Optimization framework (embed convergence criteria)



4 Analysis + Validation



principled vs random



EICUG AI WG (AI4EIC)

https://eic.ai

AI4EIC

https://indico.bnl.gov/e/AI4EIC

AI4EIC workshop
on October 10-14 2022 at W&M

WILLIAM & MARY
CHARTERED 1693

AI4EIC workshop sessions

- Design
- Theory/Experiment connections
- Reconstruction and PID
- Infrastructure and Frontiers
- Streaming
- Tutorials and Hackathon

AI4EIC - October 10-14, 2022

2nd General Workshop on Artificial Intelligence for the Electron Ion Collider
https://indico.bnl.gov/e/AI4EIC
Venue: William and Mary

Registration: https://indico.bnl.gov/e/AI4EIC

Directions to campus: https://mason.wm.edu/about/visiting/directions/index.php
There is a parking lot across Ukrop Way from the business school, and I hope to reserve some parking spots there.

Special hotel rates: https://www.wm.edu/about/visiting/lodging/special-rates/

Contacts:

support@eic.ai

Conclusions

- AI can assist the design and R&D of complex experimental systems by providing more efficient design (considering multiple objectives) utilizing effectively the computing resources needed to achieve that.
- EIC is one of the first experiments to be designed with the support of AI.

  The ECCE 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.
- This workflow can be further extended for EPIC to optimize the reference detector and to include:
  - More realistic effects in the simulation and reconstruction techniques
  - A larger system of sub-detectors, e.g, detectors like the dRICH
- Design optimization pipelines of increased complexity can take advantage of distributed computing.