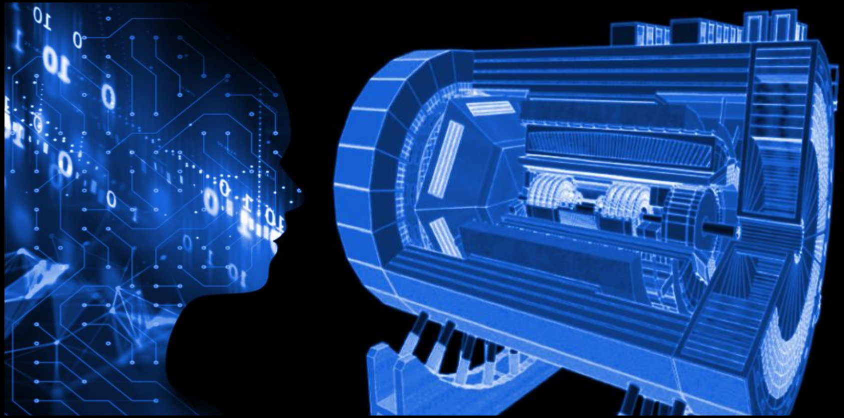
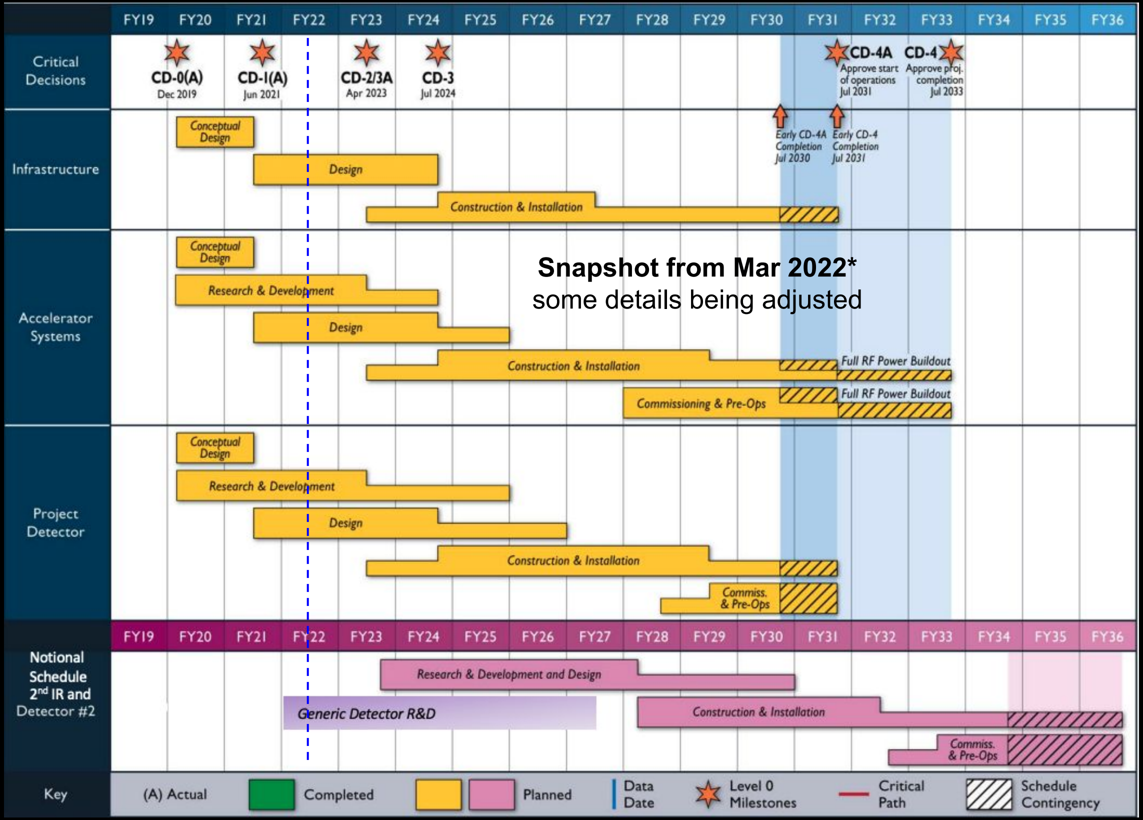


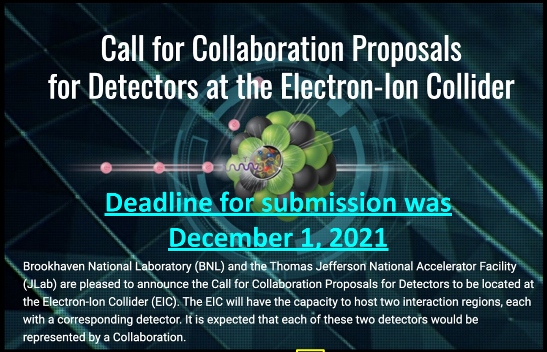
AI-assisted optimization of the EIC detector-1: Tracking System and extension to PID

Cristiano Fanelli

EIC Schedule and Milestones



proto-collaborations



EIC Detector Proposal Advisory Panel Meeting

Process completed on March 21, 2022

Panel Report

6. Recommendations: ECCE Reference Detector

The panel unanimously recommends ECCE as Detector 1. The proto-collaboration is urged to openly accept additional collaborators and quickly consolidate its design so that the Project Detector can advance to CD2/3a in a timely way.

EIC DETECTOR 1 GENERAL MEETING

Following the DPAP process , the EIC Community is moving towards the formation of a scientific collaboration to support the realization of the EIC project detector - temporarily referred to as "Detector-1".

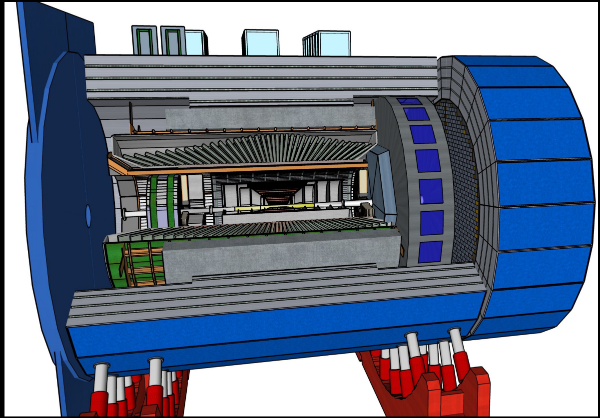
Towards Collaboration

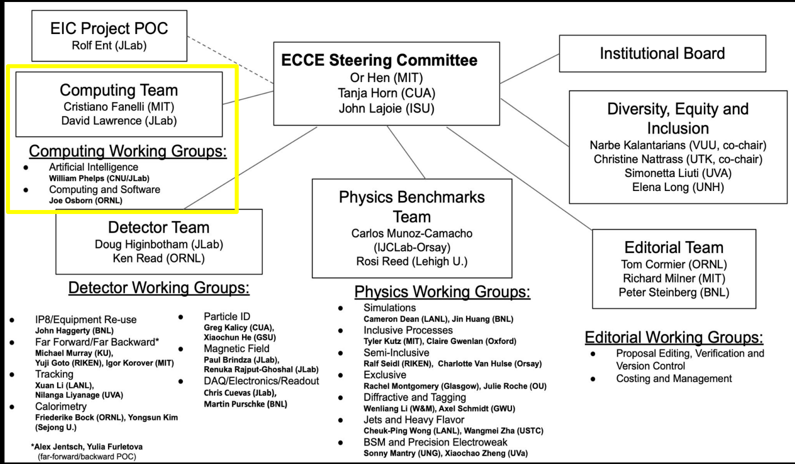
*EIC Schedule, J. Yeck, Mar 2022

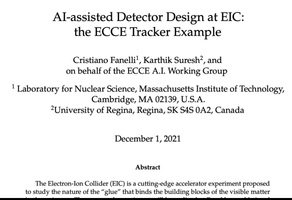
EIC Comprehensive Chromodynamics Experiment

- Proto-collaboration that comprised scientists from 98 institutions
- Develop low-risk, cost-effective, flexible and optimized EIC detector
- Detector concept based on a 1.5 T solenoidal magnet

https://www.ecce-eic.org







AI-assisted Detector Design at EIC:
the ECCE Tracker Example

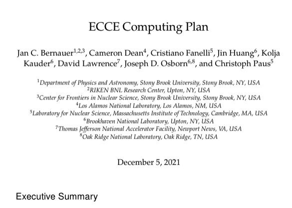
Cristiano Fanelli[1], Karthik Suresh[2], and
on behalf of the ECCE A.I. Working Group

[1] Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.
[2]University of Regina, Regina, SK S4S 0A2, Canada

December 1, 2021

Abstract

The Electron-Ion Collider (EIC) is a cutting-edge accelerator experiment proposed to study the nature of the "glue" that binds the building blocks of the visible matter



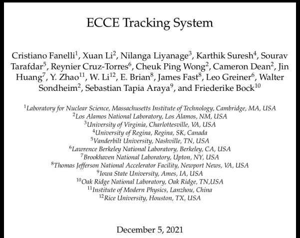
ECCE Computing Plan

Jan C. Bernauer[1,2,3], Cameron Dean[4], Cristiano Fanelli[5], Jin Huang[6], Kolja Kauder[6], David Lawrence[7], Joseph D. Osborn[6,8], and Christoph Paus[5]

[1]Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA
[2]RIKEN BNL Research Center, Upton, NY, USA
[3]Center for Frontiers in Nuclear Science, Stony Brook University, Stony Brook, NY, USA
[4]Los Alamos National Laboratory, Los Alamos, NM, USA
[5]Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[6]Brookhaven National Laboratory, Upton, NY, USA
[7]Thomas Jefferson National Accelerator Facility, Newport News, VA, USA
[8]Oak Ridge National Laboratory, Oak Ridge, TN, USA

December 5, 2021

Executive Summary



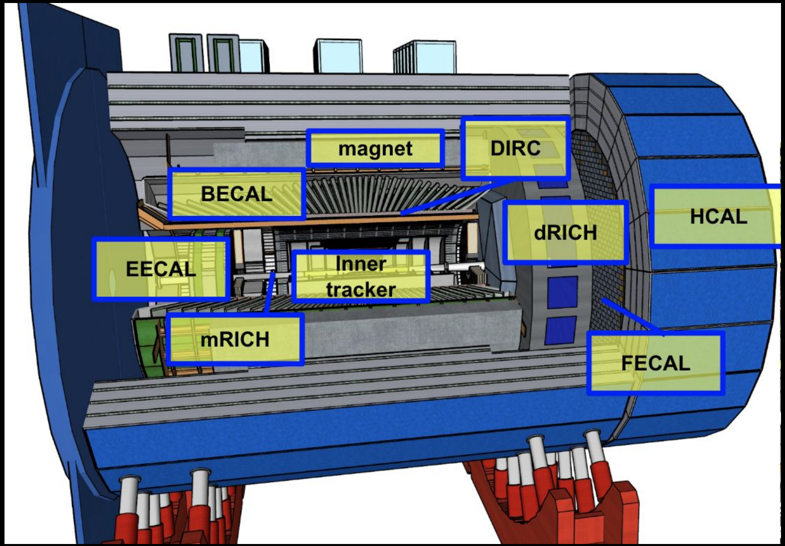
ECCE Tracking System

Cristiano Fanelli[1], Xuan Li[2], Nilanga Liyanage[3], Karthik Suresh[4], Sourav Tarafdar[5], Reynier Cruz-Torres[6], Cheuk Ping Wong[2], Cameron Dean[2], Jin Huang[7], Y. Zhao[11], W. Li[12], E. Brian[8], James Fast[8], Leo Greiner[6], Walter Sondheim[2], Sebastian Tapia Araya[9], and Friederike Bock[10]

[1]Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Los Alamos National Laboratory, Los Alamos, NM, USA
[3]University of Virginia, Charlottesville, VA, USA
[4]University of Regina, Regina, SK, Canada
[5]Vanderbilt University, Nashville, TN, USA
[6]Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[7]Brookhaven National Laboratory, Upton, NY, USA
[8]Thomas Jefferson National Accelerator Facility, Newport News, VA, USA
[9]Iowa State University, Ames, IA, USA
[10]Oak Ridge National Laboratory, Oak Ridge, TN,USA
[11]Institute of Modern Physics, Lanzhou, China
[12]Rice University, Houston, TX, USA

December 5, 2021

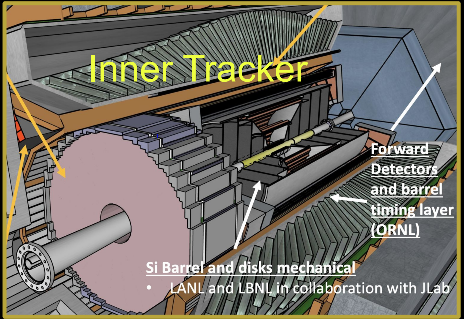
The Reference Detector



*Tracker

Combines:

- ITS-3 Si technology
- Gaseous detectors
- AC-LGAD ToFs



*Particle Identification (PID) with Cherenkov detectors

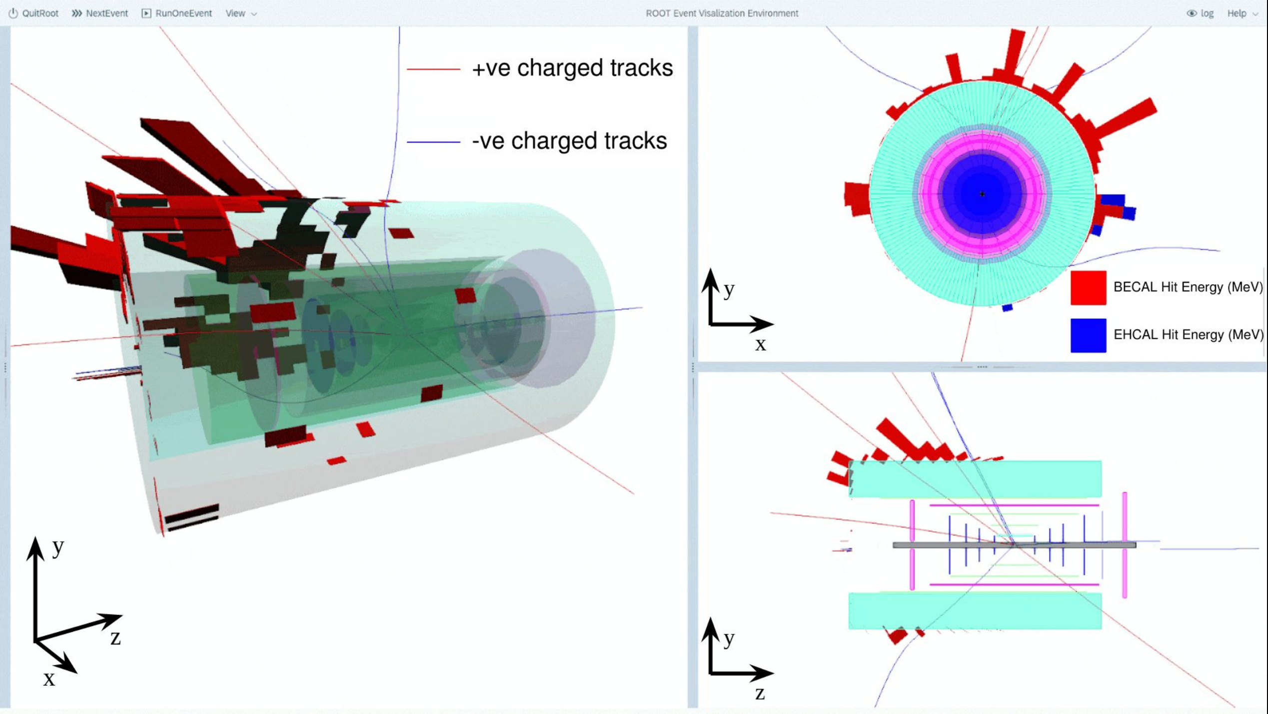
- dual radiator ring-imaging Cherenkov detector (RICH) in the hadron direction
- DIRC (detection of internally reflected Cherenkov light) in the barrel
- modular RICH in the electron direction.

Simulating these detectors is typically compute expensive, involving many photons that need to be tracked through complex surfaces.

All three rely on pattern recognition of ring images in reconstruction, and the DIRC is the one having the more complex ring patterns!

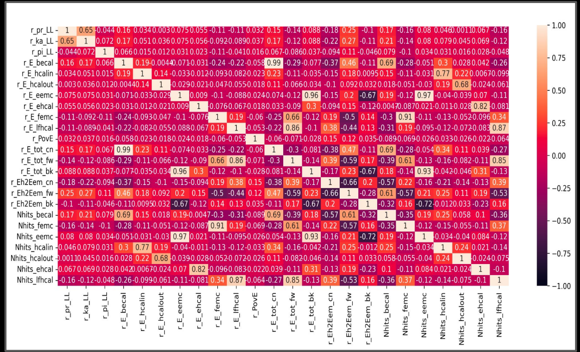
*Highlighting parts that will be discussed in this talk

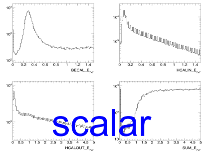
Event Display and Reconstructed Features

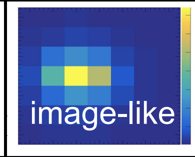


Reconstruction typically deals with relatively large feature space (low and high-level features) combining sub-detectors

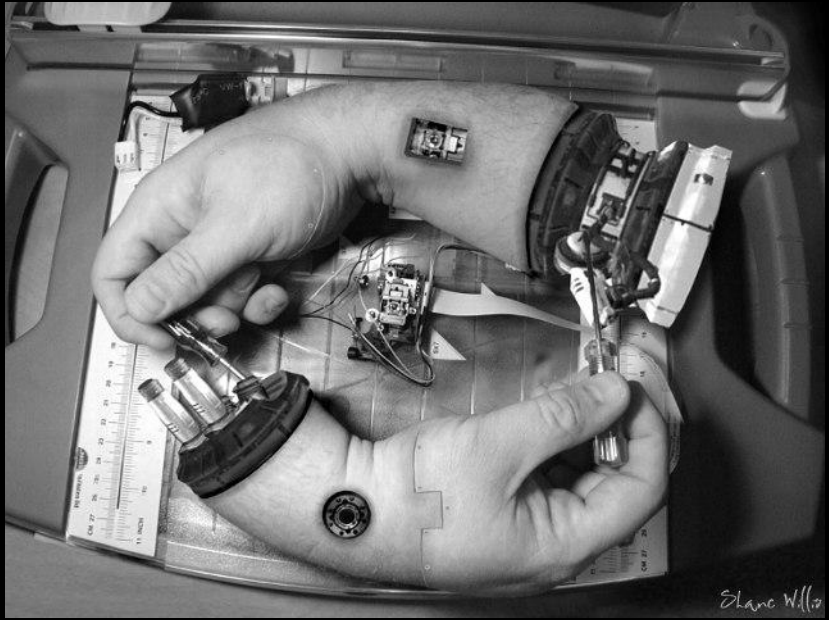
For illustrative purposes, showing example of calorimetry (outer layers)







How do we design and optimize Detectors?



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

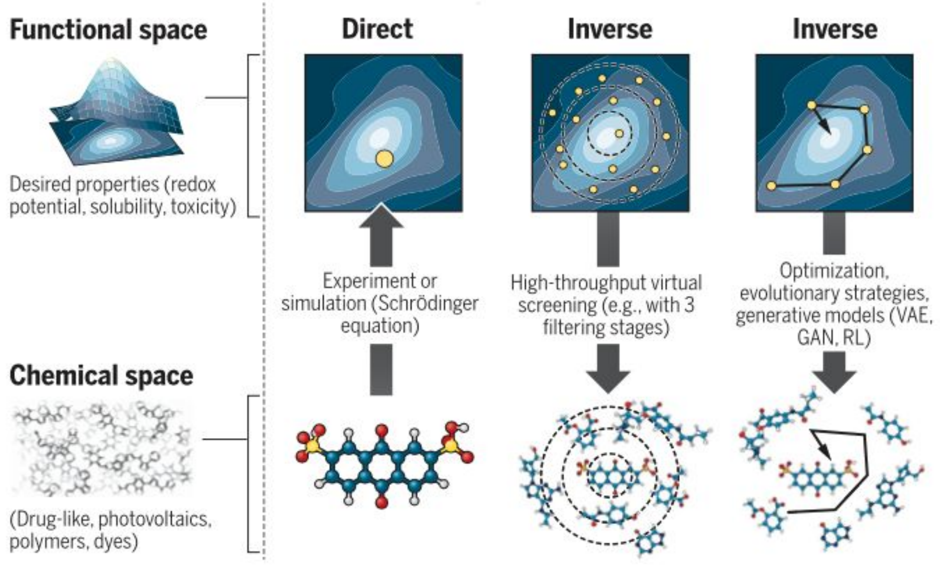


Fig. 2. Schematic of the different approaches toward molecular design. Inverse design starts from desired properties and ends in chemical space, unlike the direct approach that leads from chemical space to the properties.

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

Optimization of Detectors 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. doi:10.1088/1361-6560/ab8afc.

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

Full Optimization of Detectors/Accelerators

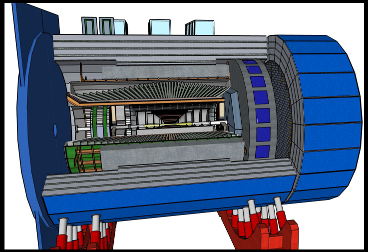
- When it comes to designing detectors and accelerators with AI this is a frontier topic with few examples in the literature.
  - What follows uses “detector” as example but applies to both detector and accelerator — and can be extended to many other applications
- For years the full detector design 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 [1].
- AI-assisted strategies can help designing more efficiently (in terms of performance and resources needed).
  - Need establishing a full body of instructions [2].
  - The choice of a suitable algorithm is a challenge itself (no free lunch theorem [3]) and always requires some degree of customization.

[1] Bellman, Richard. *Dynamic programming*. Vol. 295. RAND CORP SANTA MONICA CA, 1956.
[2] CF et al. *JINST* 15.05 (2020): P05009.
[3] Wolpert, D.H., Macready, W.G., 1997. Trans. Evol. Comp 1, 67–82

The Typical Workflow

See invited talk at IAEA Technical Meeting on AI

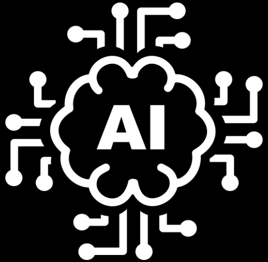
IAEA
International Atomic Energy Agency



- 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

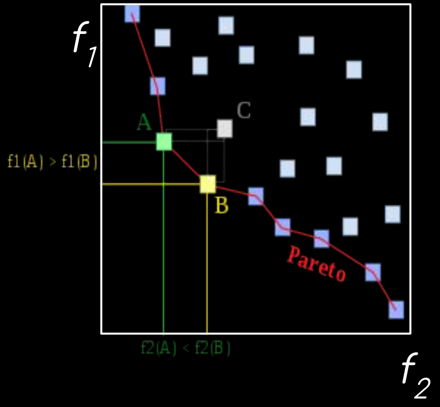
Detector Simulation

Analysis of High-level reconstruction of events

compute intensive (Geant4)

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

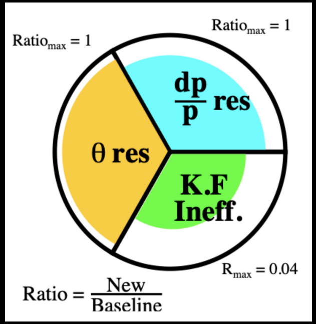
Multi-Objective Optimization



- 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.
- Due to the multiplicity in solutions, these problems were proposed to be solved suitably using evolutionary algorithms which use a population approach in its search procedure.
- MO-based solutions are helping to reveal important hidden knowledge about a problem – a matter which is difficult to achieve otherwise
- During the proposal we used both evolutionary (1) and bayesian approaches (2). I will describe now (1).

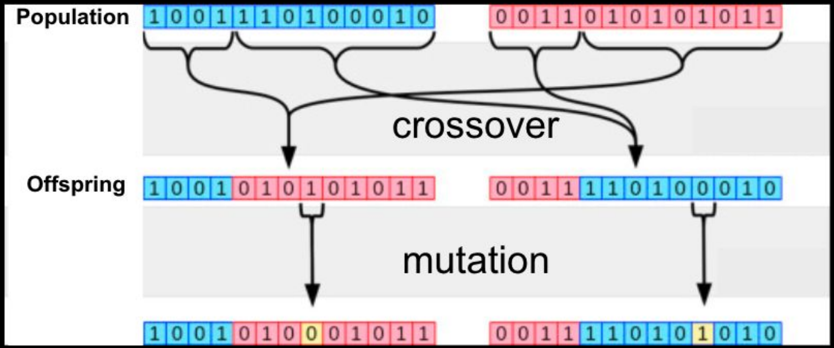
The ECCE Inner Tracker Design Optimization considers simultaneously:

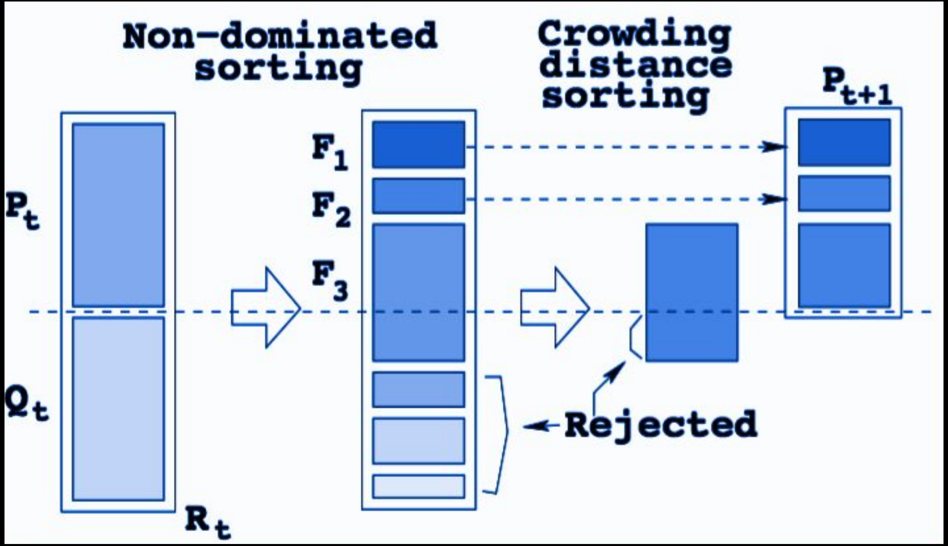
- momentum resolution
- angular resolution
- Kalman filter efficiency
- (pointing resolution)
- Mechanical constraints



Popular AI-Strategies (in a nutshell)

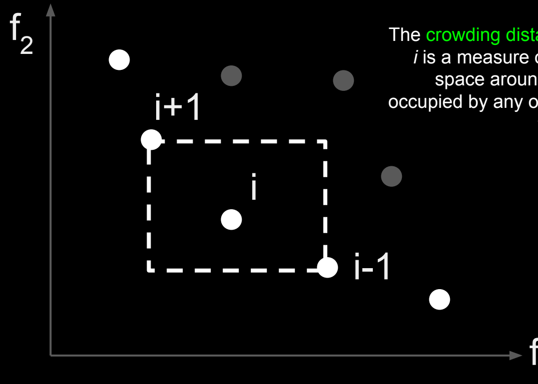
Evolutionary





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

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



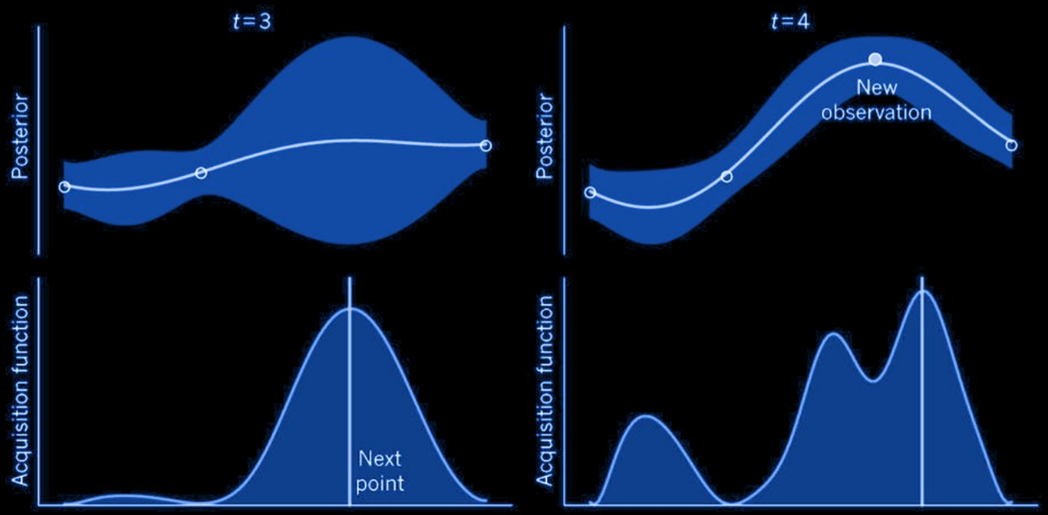
The crowding distance d_i of point *i* is a measure of the objective space around *i* which is not occupied by any other solution in the population.

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

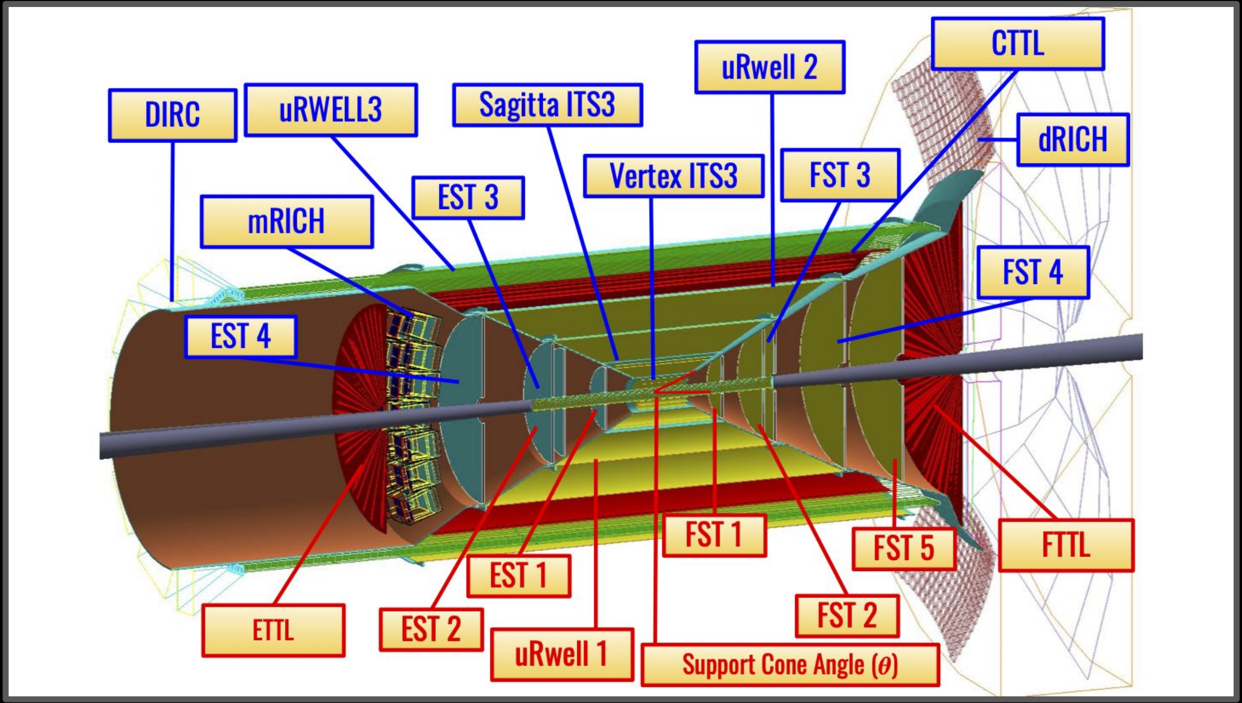
- BO is a sequential strategy developed for global optimization.
- After gathering evaluations we builds a posterior distribution used to construct an **acquisition function**.
- This cheap function determines what is **next query point**.



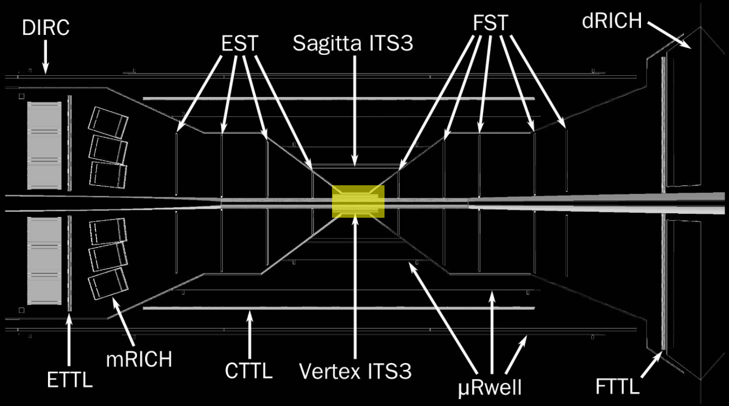
1. Select a Sample by Optimizing the Acquisition Function.
2. Evaluate the Sample With the Objective Function.
3. Update the Data and, in turn, the Surrogate Function.
4. Go To 1.

Extension to multiple objectives

AI-Assisted Optimization of the ECCE Tracking System at the Electron Ion Collider



https://ai4eicdetopt.pythonanywhere.com

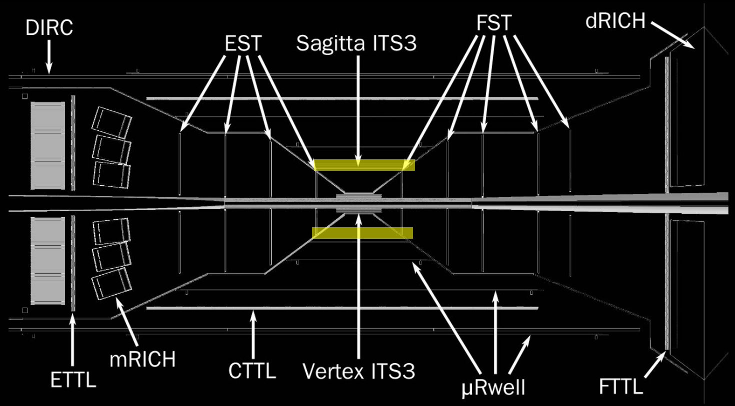


Click on hyperlinks (Fun4All)

Vertex Si Barrel

			Reference		Ongoing R&D	
Barrel	**X/X0 [%]**	**Pitch [um]**	**Radii [cm]**	**Length [cm]**	**Radii [cm]**	**Length [cm]**
Layer 1	0.05	10	3.3	27	3.3	27
Layer 2	0.05	10	4.35	27	4.35	27
Layer 3	0.05	10	5.4	27	5.4	27

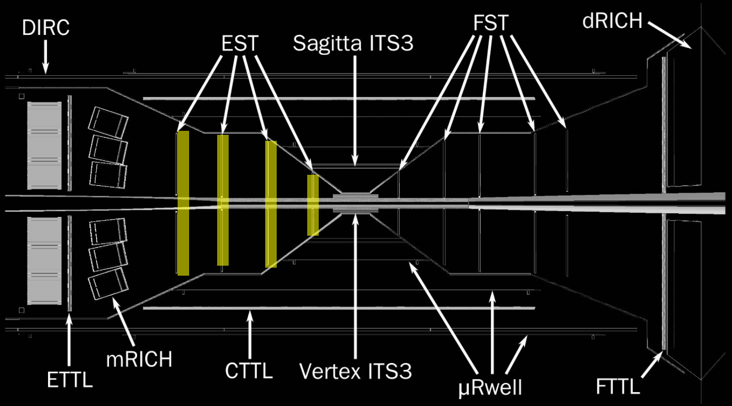
Values being used in these slides



Sagitta Si Barrel

			Reference		Ongoing R&D	
Barrel	X/X0 [%]	Pitch [um]	Radii [cm]	Length [cm]	Radii [cm]	Length [cm]
Layer 1	0.05	10	21	54	14.0	54
Layer 2	0.05	10	22.68	54	15.5	54

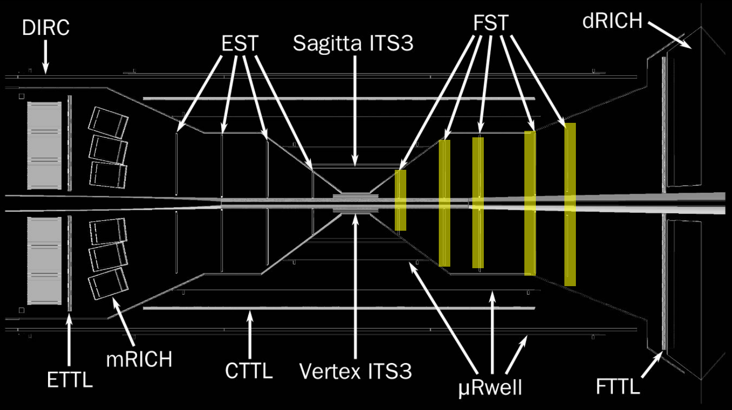
another potential parameter to optimize?



Additional thickness for services, cooling is given here

EST Disks

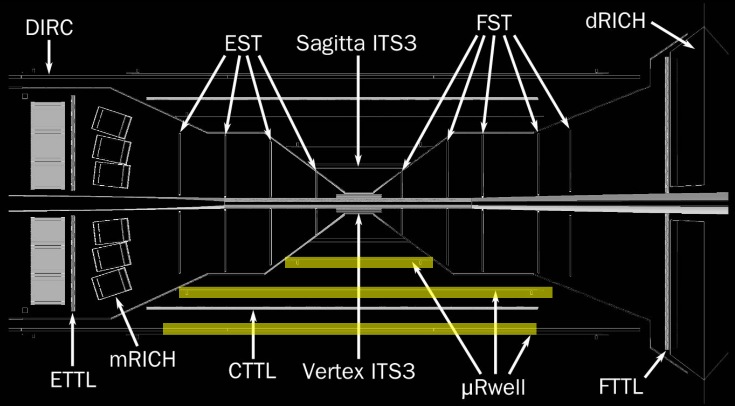
			Reference			Ongoing R&D		
Disk	**Si Thickness[um]**	**Pitch[um]**	**RMin [cm]**	**RMax[cm]**	**ZPos[cm]**	**RMin [cm]**	**RMax [cm]**	**ZPos[cm]**
EST 4	35	10	5.5	41.5	-106	6.0	48.0	-107.4
EST 3	35	10	4.5	40.5	-79	4.8	35.25	-80.05
EST 2	35	10	3.5	36.5	-52	3.3	27.3	-58.29
EST 1	35	10	3.5	18.5	-25	3.3	15.3	-33.2



Additional thickness for services, cooling is given here

FST Disks

			Reference			Ongoing R&D		
Disk	**Si Thickness [um]**	**Pitch [um]**	**RMin [cm]**	**RMax [cm]**	**ZPos [cm]**	**RMin [cm]**	**RMax [cm]**	**ZPos [cm]**
FST 5	35	10	7.5	43.5	125	8.2	62.2	144
FST 4	35	10	5.5	41.5	106	5.8	49.8	115
FST 3	35	10	4.5	40.5	73	4.8	34.8	79.85
FST 2	35	10	3.5	36.5	49	3.5	27.5	58.29
FST 1	35	10	3.5	18.5	25	3.5	15.5	33.2

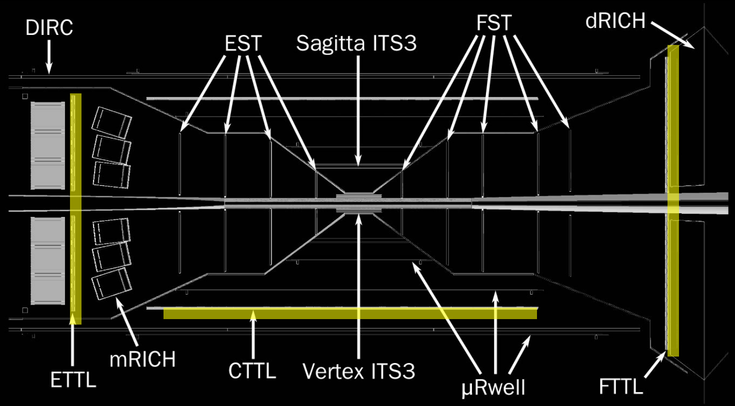


μRwell Cylinder

Additional thickness for services, cooling is given here

Barrel	Res [um]	Thickness [cm]	Reference Radii [cm]	Reference Length [cm]	Ongoing R&D Radii [cm]	Ongoing R&D Length [cm]
Layer 1	55	0.03	33.14	80	33.14	140
Layer 2	55	0.03	51.00	212	51.00	230
Layer 3	55	0.03	77.02	342	77.02	342

another potential parameter to optimize?



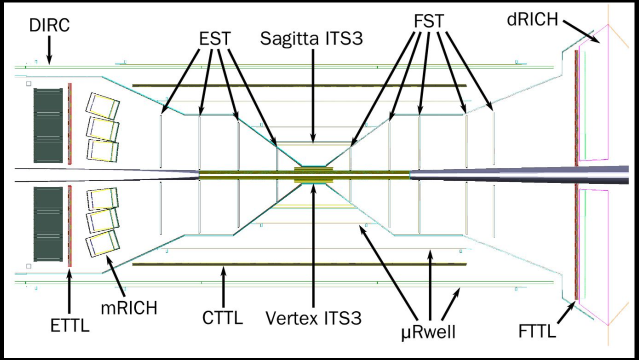
TOF Detectors

			Reference			Ongoing R&D		
TOF TTL	Si Thickness [um]	Pitch [um]	RMin [cm]	RMax [cm]	L [cm]	RMin [cm]	RMax [cm]	L [cm]
CTTL	85	30	64	-	140	64	-	140
ETTL	85	30	8	64	-155.5	8	64	169
FTTL	85	30	7	87	182	7	87	182

Additional thickness for services, cooling is given here

EIC Detector Tracker

Sub Detector System	No Of Layers	Technology	Pitch/res [µm]	Thickness [X/X0]	Description
Vertex Barrel	3	MAPS-ITS3	10	0.05	Monolithic Active Pixel Sensor; EIC R&D eRD111. High precision tracking.
Sagitta Barrel	2	MAPS-ITS3	10	0.05	Monolithic Active Pixel Sensor; EIC R&D eRD11. High precision tracking.
Outer Barrel	3	µRwell	55	0.2	µRwell is a gaseous based tracker. EIC R&D ERD6. Low Cost tracking solution
CTTL (TOF)	1	AC-LGAD	30	~0.1	Low Gain Avalanche Detectors (ACLGAD): EIC R&D ERD112. High precision tracking and Timing.
EST	4	MAPS-ITS3	10	0.3	Monolithic Active Pixel Sensor; EIC R&D eRD11. High precision tracking.
FST	5	MAPS-ITS3	10	0.3	Monolithic Active Pixel Sensor; EIC R&D eRD111. High precision tracking.
ETTL	1	AC-LGAD	30	~0.1	Low Gain Avalanche Detectors (ACLGAD): EIC R&D ERD112. High precision tracking and timing
FTTL	1	AC-LGAD	30	~0.1	Low Gain Avalanche Detectors (ACLGAD): EIC R&D ERD112. High precision tracking and timing

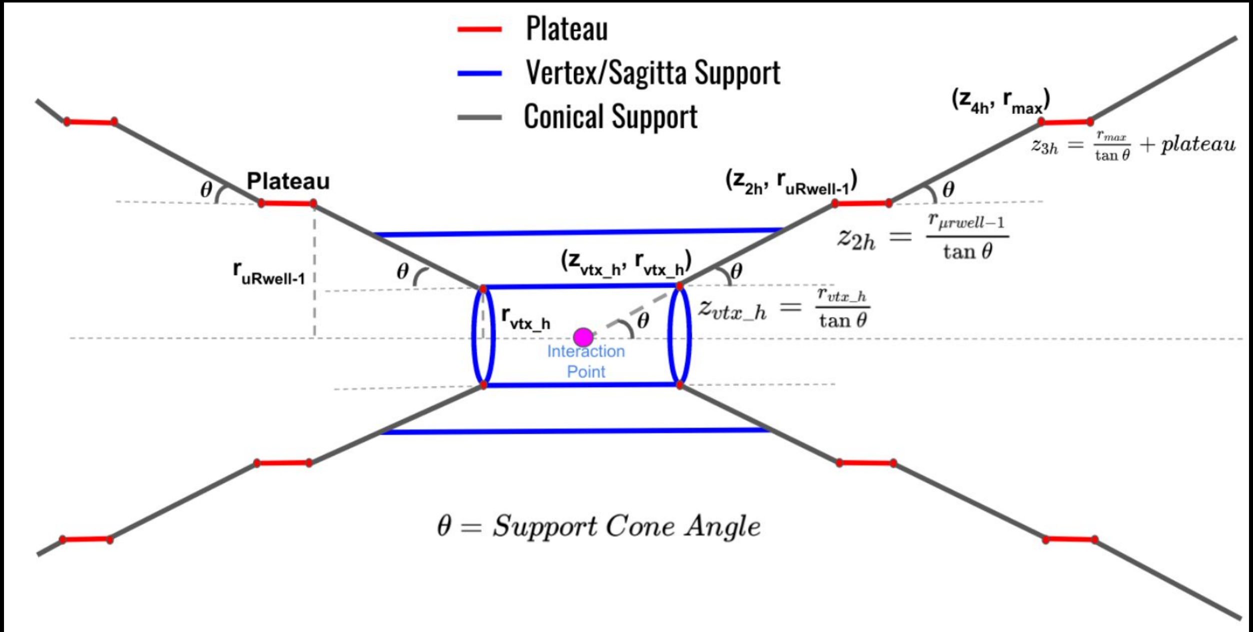


ECCE design (non-projective)	
Design Parameter	**Range**
μRWELL 1 (Inner) (r) Radius	[17.0, 51.0 cm]
μRWELL 2 (Inner) (r) Radius	[18.0, 51.0 cm]
EST 4 z position	[-110.0, -50.0 cm]
EST 3 z position	[-110.0, -40.0 cm]
EST 2 z position	[-80.0, -30.0 cm]
EST 1 z position	[-50.0, -20.0 cm]
FST 1 z position	[20.0, 50.0 cm]
FST 2 z position	[30.0, 80.0 cm]
FST 3 z position	[40.0, 110.0 cm]
FST 4 z position	[50.0, 125.0 cm]
FST 5 z position	[60.0, 125.0 cm]
ECCE ongoing R&D (projective)	
Design Parameter	**Range**
Angle (Support Cone)	[25.0°, 30.0°]
μRWELL 1 (Inner) Radius	[25.0, 45.0 cm]
ETTL z position	[-171.0, -161.0 cm]
EST 2 z position	[45, 100 cm]
EST 1 z position	[35, 50 cm]
FST 1 z position	[35, 50 cm]
FST 2 z position	[45, 100 cm]
FST 5 z position	[100, 150 cm]
FTTL z postion	[156, 183 cm]

Parametrization

arXiv:2205.09185

Parametrization of the support structure



Parametrization of disks radii and TTL

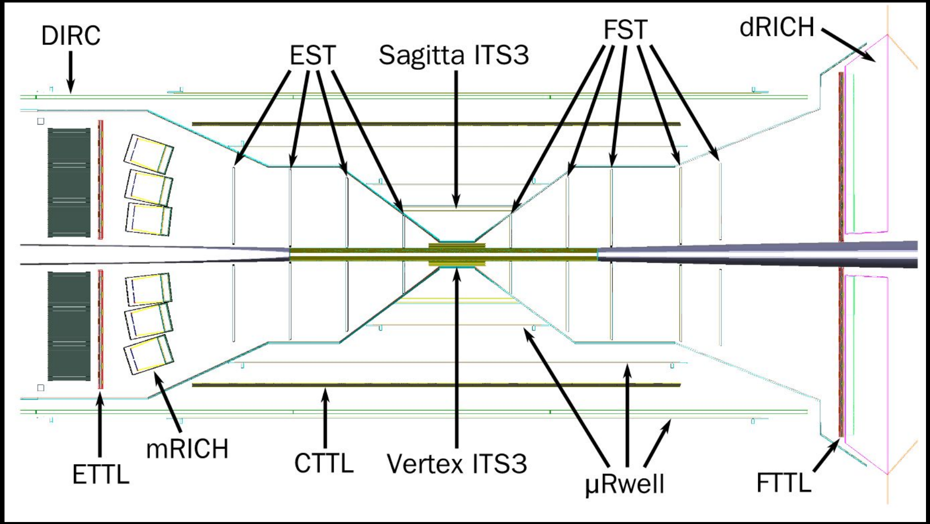
Implementation of Geometric Constraints

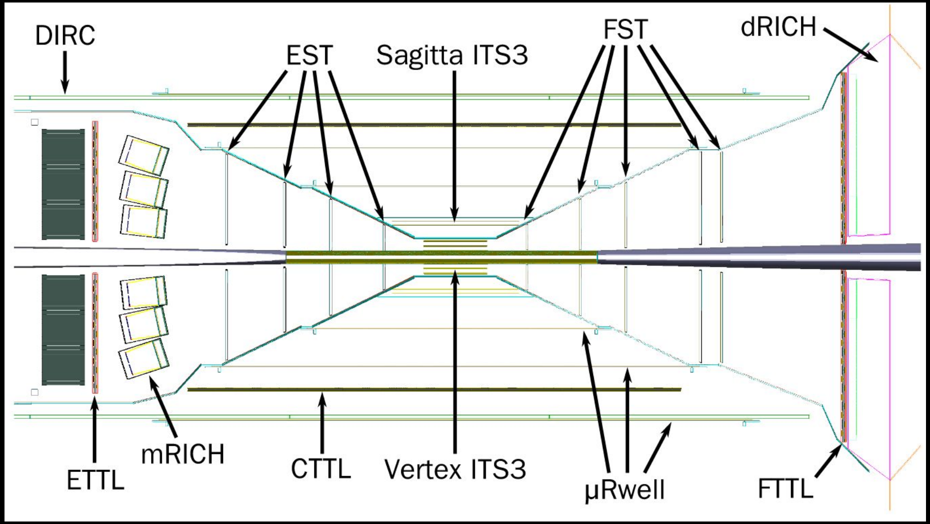
RMax and RMin of the disks are then calculated based on the support structure.

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

Reference VS Projective (R&D)





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

Reference VS Projective (R&D)

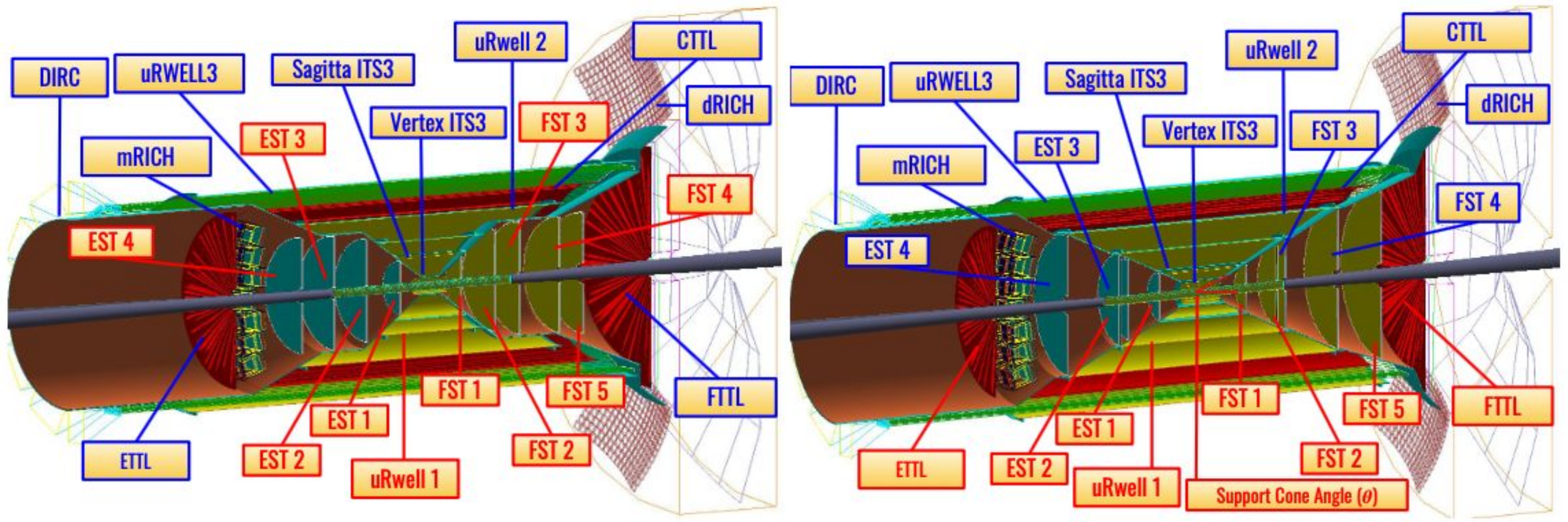
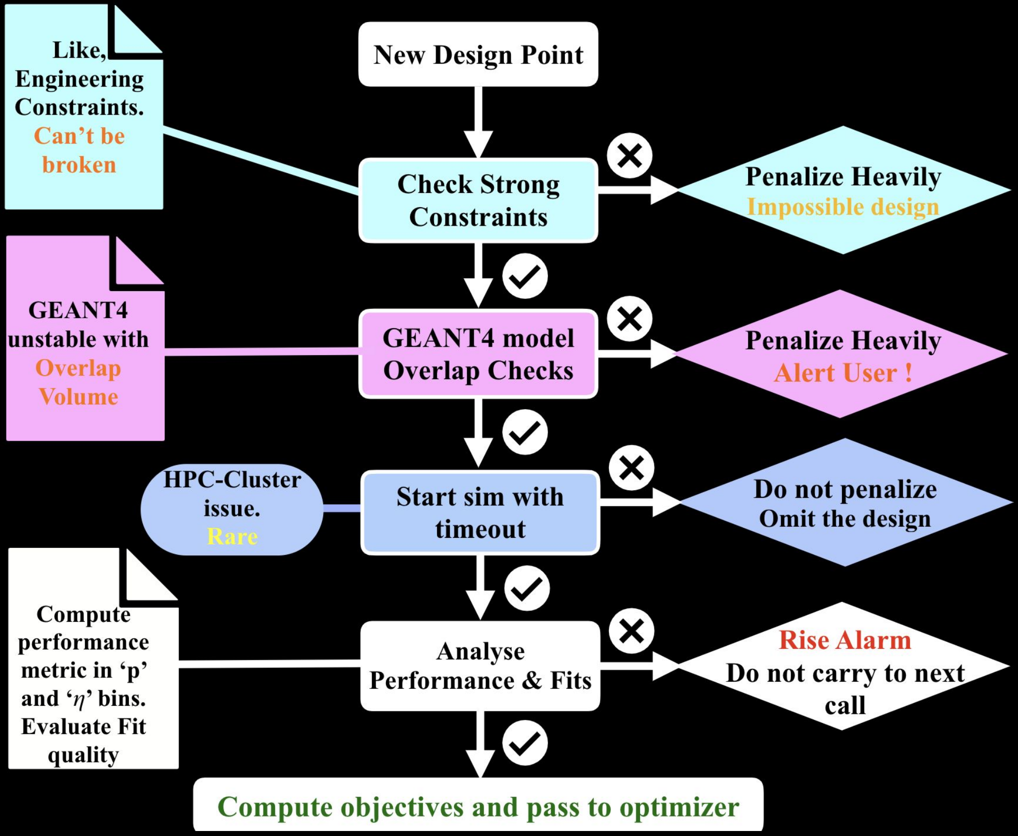


Figure 5: **Tracking and PID system in the non-projective (left) and the ongoing R&D projective (right) designs:** the two figures show the different geometry and parametrization of the ECCE non-projective design (left) and of the ongoing R&D projective design to optimize the support structure (right). 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.

“Soft”/“Hard” Constraints

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

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

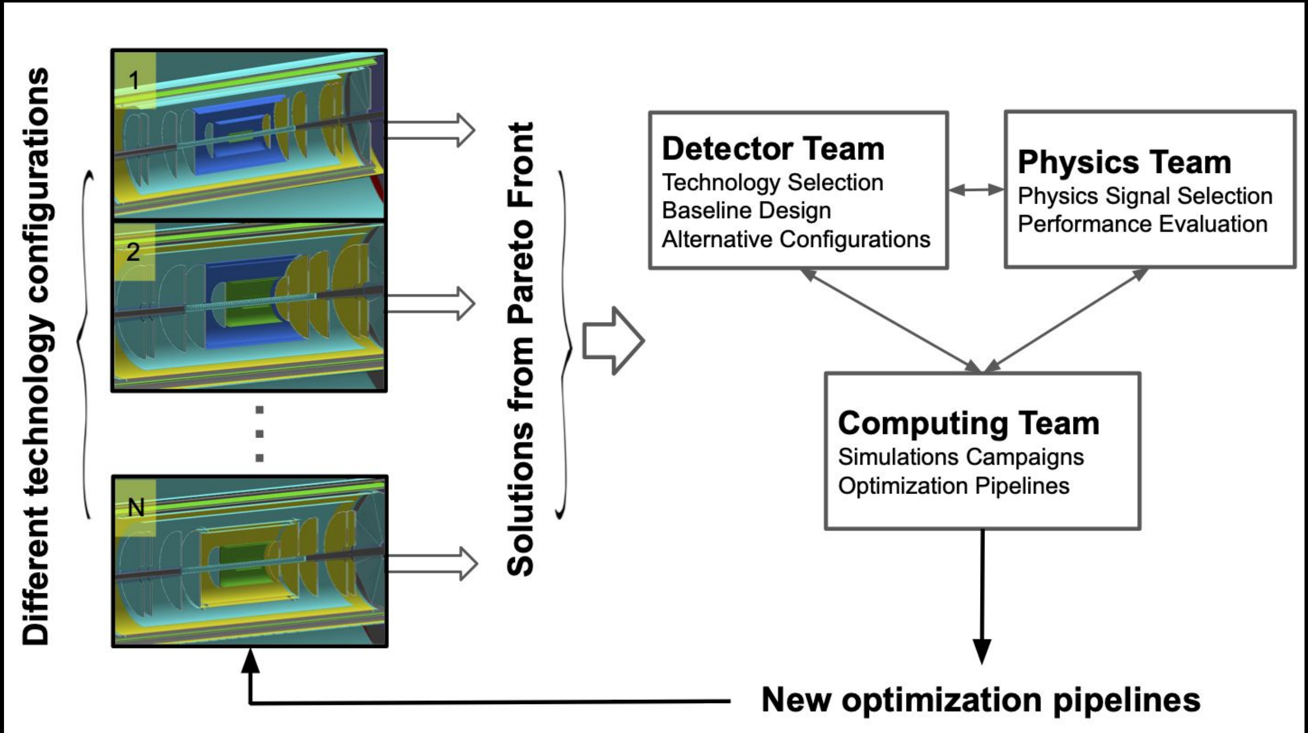


Integration during the EIC Detector Proposal

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



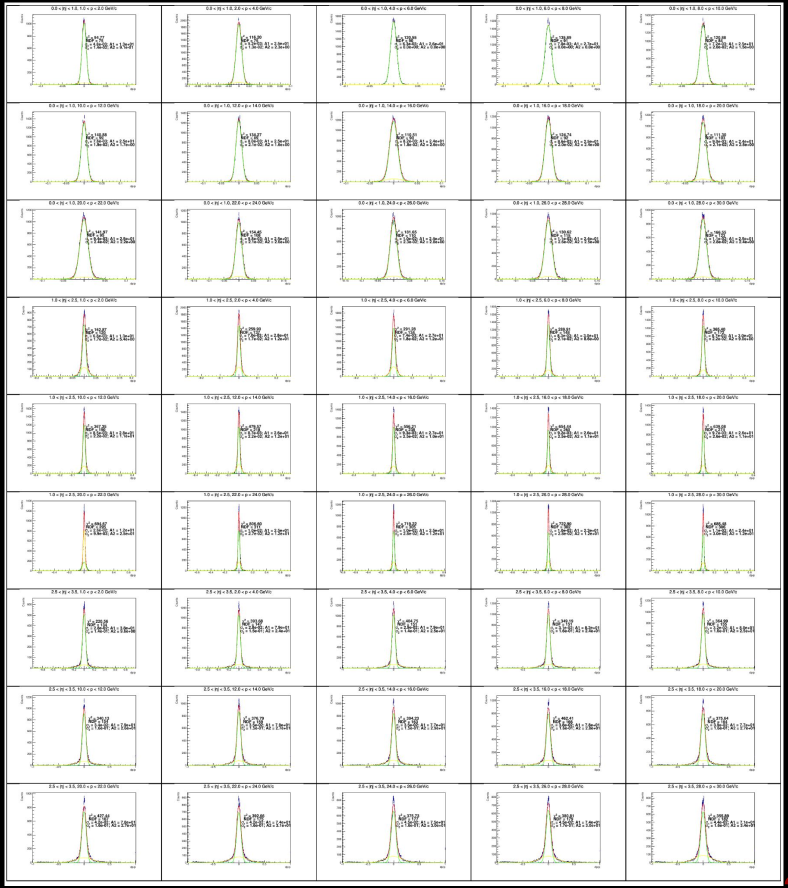
Implementation

- **Objective functions** Average of Weighted Averages (n_obj ≥ 3)
  - **Momentum resolution dp/p**
  - **Theta resolution dθ/θ**
  - **Projected dθ/θ at PID location.**
  - **Kalman Filtering inefficiency** (improving the tracking reconstruction ability of the algorithm)

- **Validation** of the solutions
  - Validate by comparing optimal vs baseline dφ resolution, vertex resolution and reconstruction efficiency

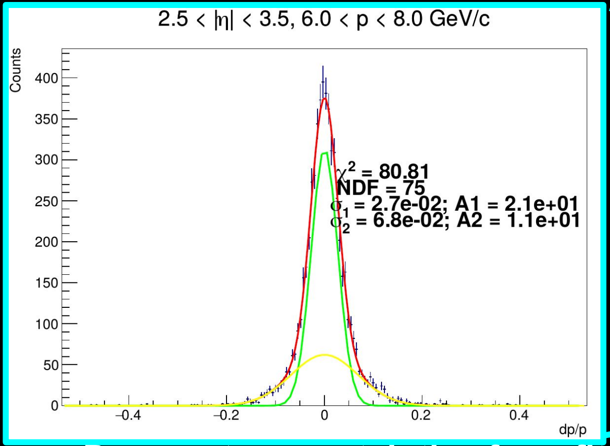
Weighted sum with errors

Weighted sum with errors



Implementation

Weighted sum with errors



Propagate uncertainties from fits

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

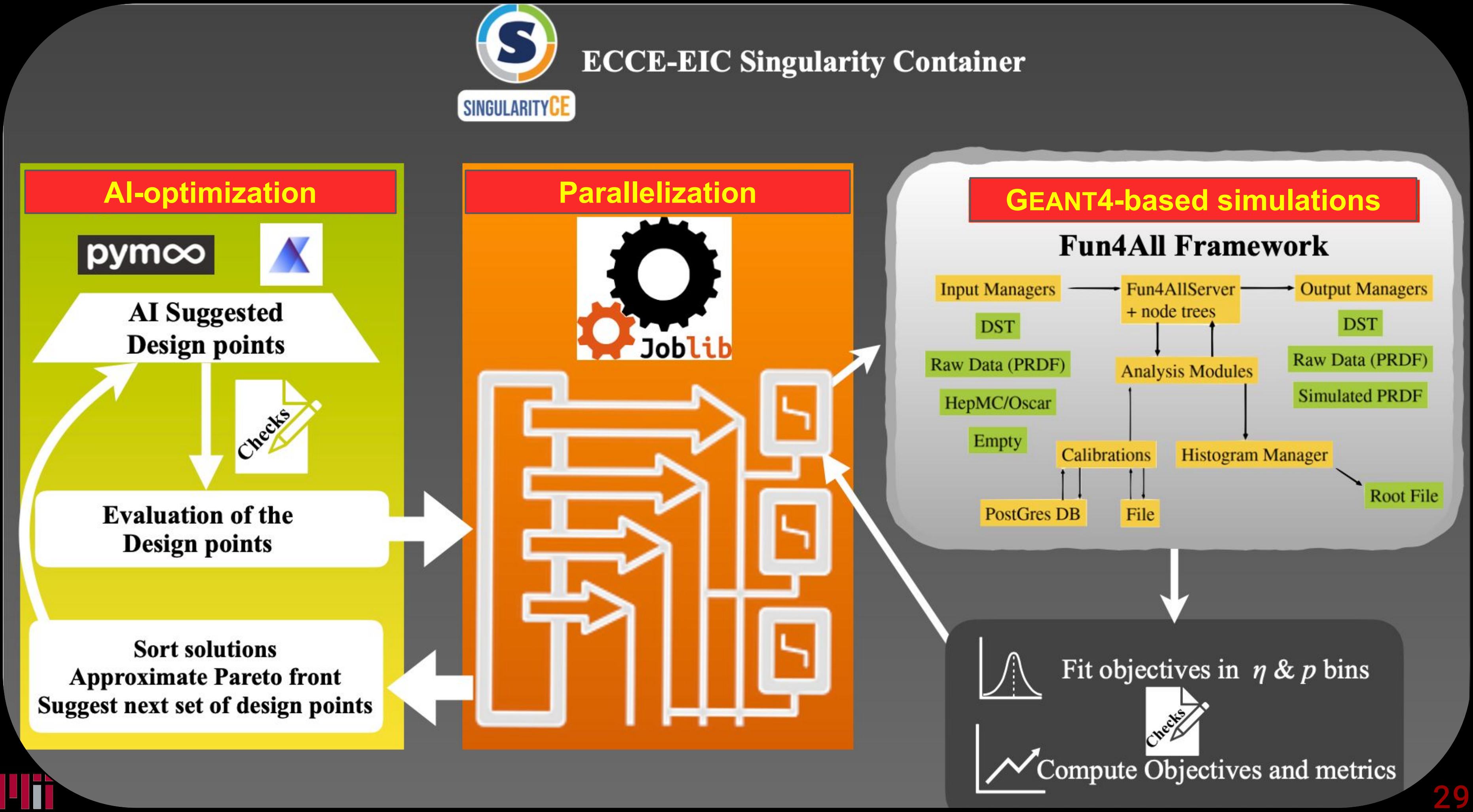
Average objective in a η bin

Sum in bins of P 14 bins

Weighted sum with errors

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

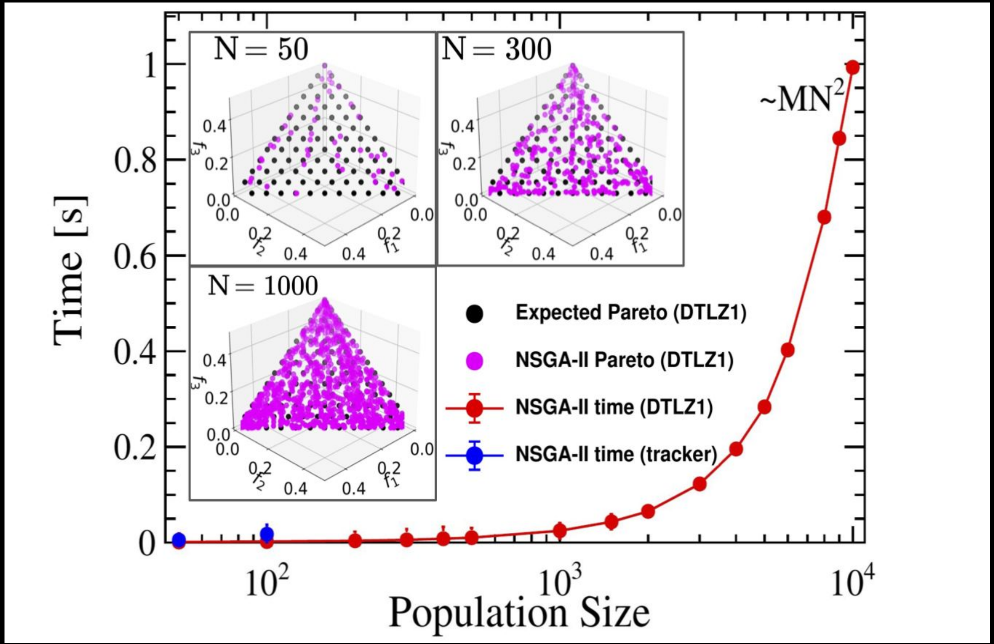
$$R(f) = \frac{1}{N_\eta}\sum_\eta \left(\frac{\sum_p w_{p,\eta} \cdot R(f)_{p,\eta}}{\sum_p w_{p,\eta}}\right)$$



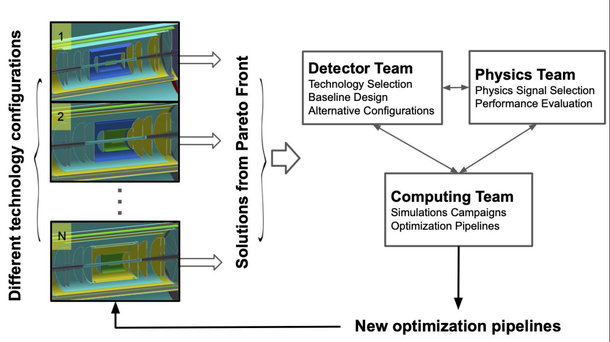
ECCE-EIC Singularity Container
SINGULARITYCE
AI-optimization
pymoo
AI Suggested Design points
Checks
Evaluation of the Design points
Sort solutions
Approximate Pareto front
Suggest next set of design points
Parallelization
Joblib
GEANT4-based simulations
Fun4All Framework
Input Managers
DST
Raw Data (PRDF)
HepMC/Oscar
Empty
Fun4AllServer + node trees
Analysis Modules
Output Managers
DST
Raw Data (PRDF)
Simulated PRDF
Calibrations
Histogram Manager
Root File
PostGres DB
File
Fit objectives in η & p bins
Checks
Compute Objectives and metrics

Computational Resources

time taken by GA + sorting



- For the complexity of the problem and the chosen population size, the computing time is dominated by simulations and not by the AI part

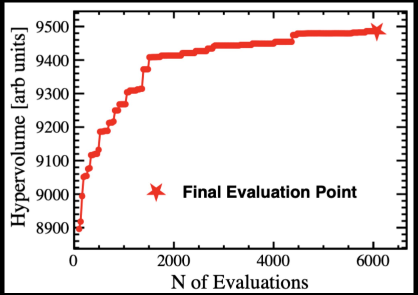


description	symbol	value
population size	N	100
# objectives	M	3
offspring	O	30
design size	D	11 (9)
# calls (tot. budget)	–	200
# cores	–	same as offspring
# charged π tracks	N_{trk}	120k
# bins in η	N_η	5
# bins in p	N_p	10

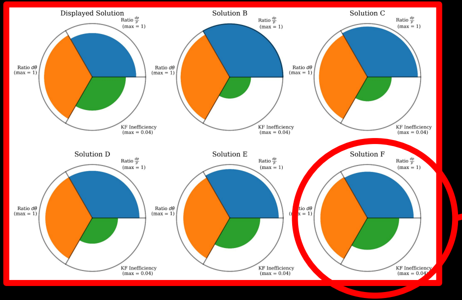
- Used a test problem DTLZ1
- Verified scaling following MN^2 and convergence to true front
- ~1s/call with 10^4 size!
- Smart pipelines of 11 variables and 3 objectives needs ~ 10000 evaluations to converge
  ~10k CPUhours / pipeline

"Navigate" Pareto Front

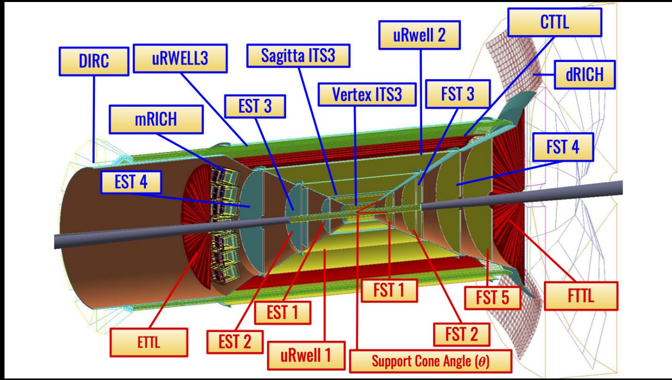
1 Can take a snapshot any time during evaluation



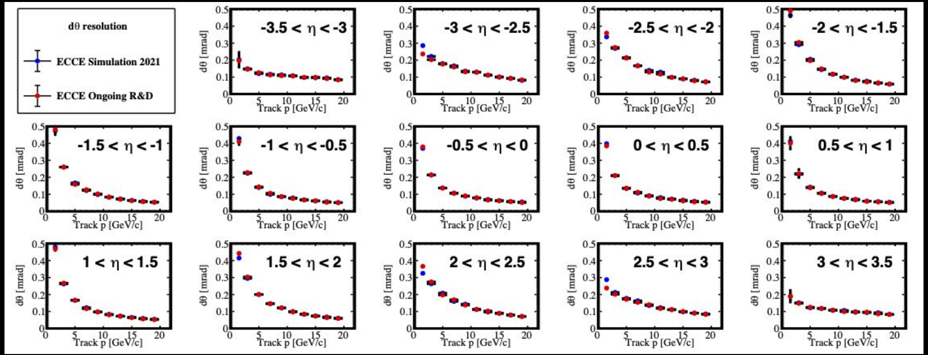
2 Updated Pareto Front at time t

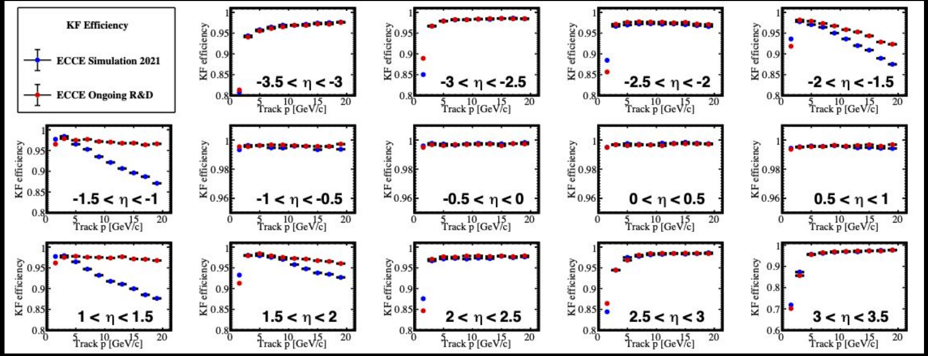


3 At each point in the Pareto front corresponds a design

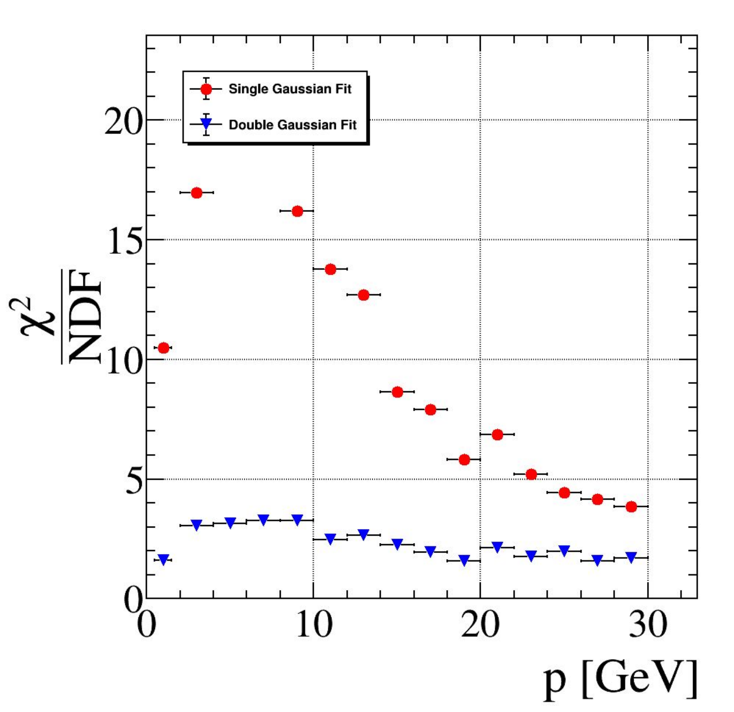


4 Analysis of Objectives (momentum resolution, angular resolution, KF efficiency)





Single VS Double Gaussian



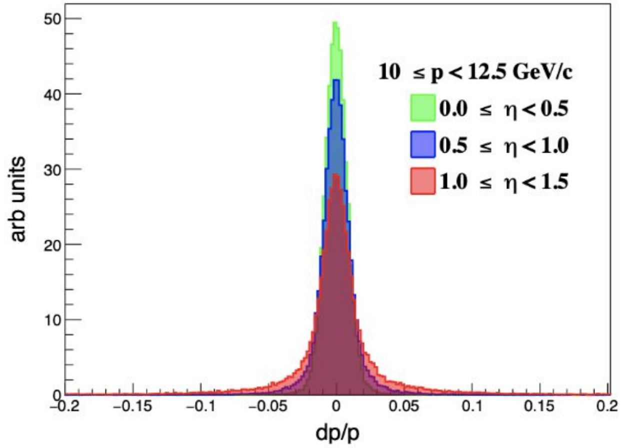
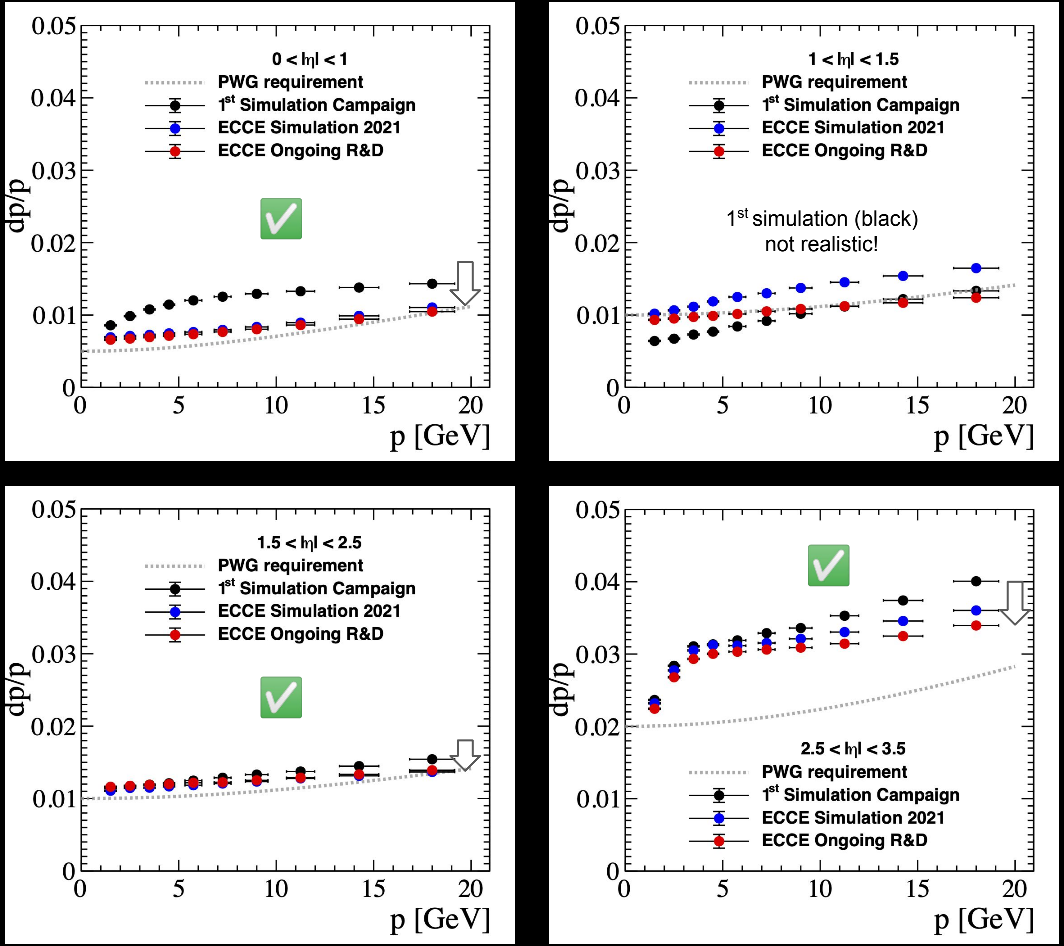


Figure 6: **Fit strategy:** a double-Gaussian fit function is utilized to extract the resolutions. Such a fit function provided good reduced χ^2 and more stable extractions compared to single-Gaussian fits. The resolution is obtained as an average of the two σ's weighted by the relative areas of the two Gaussians according to Eq. (3). The figure represents the results corresponding to a particular bin in η and p.

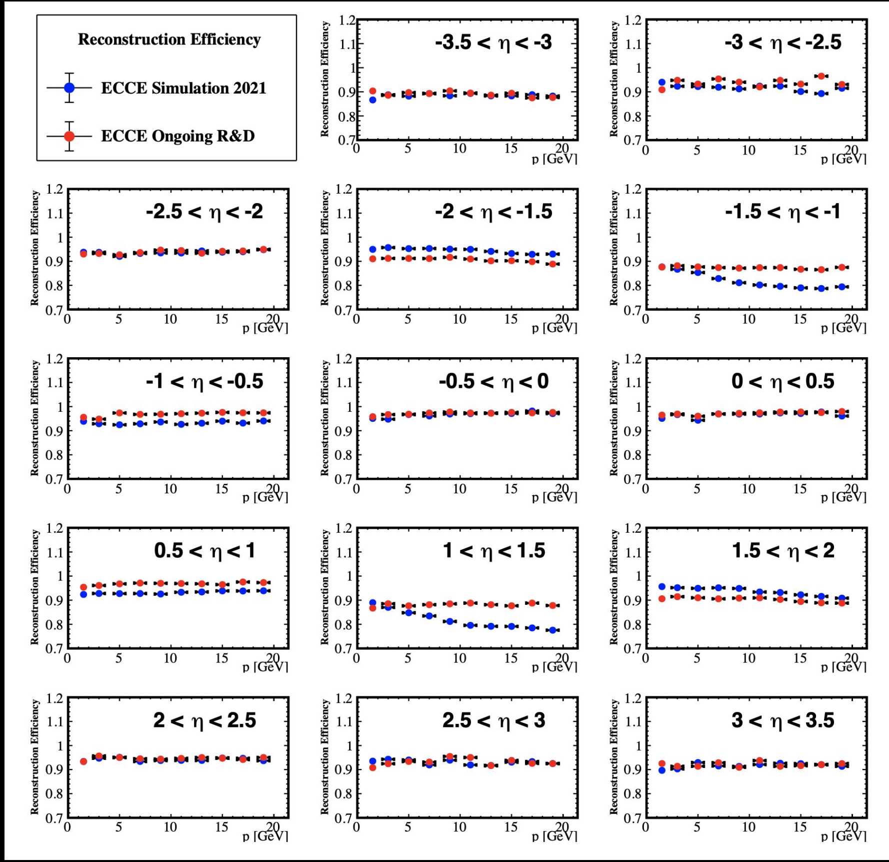
Evolution

- Black points represent the first simulation campaign, and a preliminary detector concept in phase-I optimization which did not have a developed support structure;
- Blue points represent the fully developed simulations for the final ECCE detector proposal concept; red points the ongoing R&D for the optimization of the support structure.
- Compared to black, there is an improvement in performance in all η bins with the exception of the transition region, an artifact that depends on the fact that black points 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

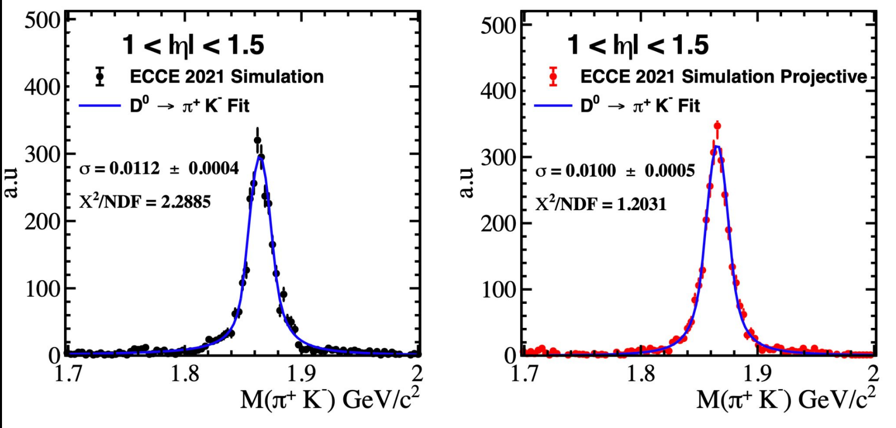
Reconstruction Efficiency



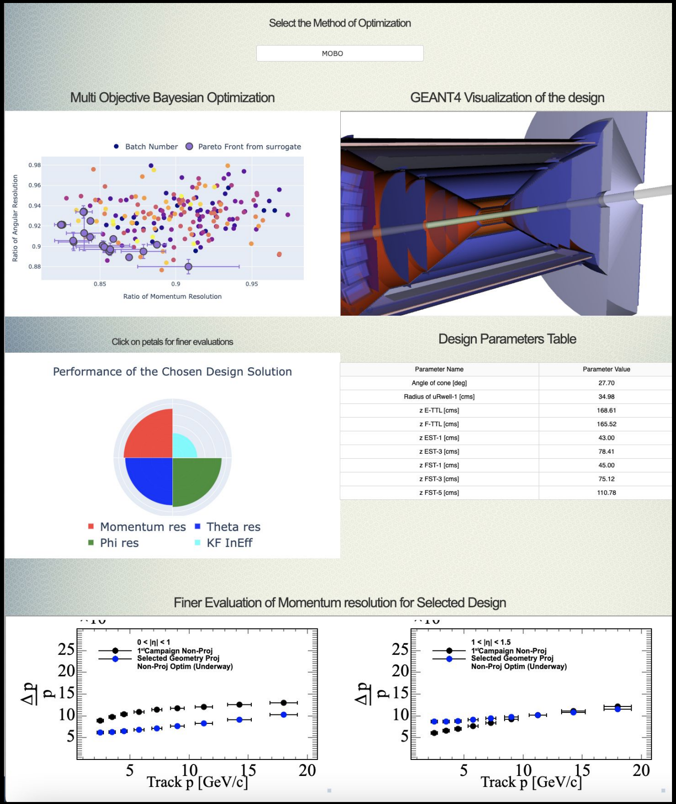
Performance evaluated after optimization process (both designs) using standard analysis procedures

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

D0 invariant mass from semi-inclusive deep inelastic scattering



Interactive Navigation of Pareto front



- 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 — Facebook)

K. Suresh (U. of Regina) https://ai4eicdetopt.pythonanywhere.com

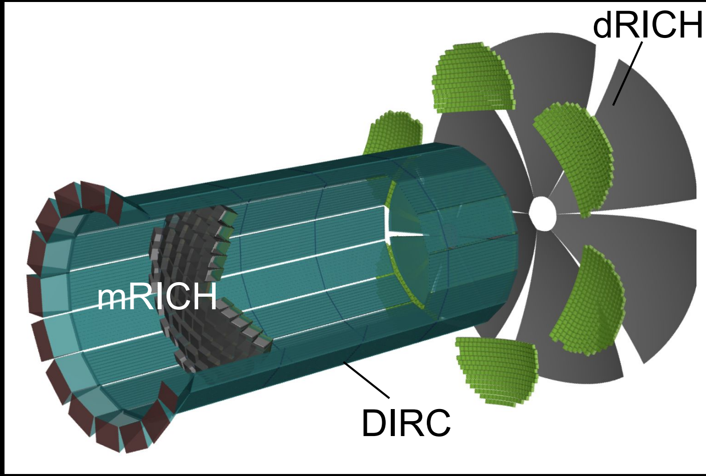
CF, Z. Papandreou, K. Suresh, *Designing EIC with the assistance of AI: strategies and perspectives* (in progress)

Plans

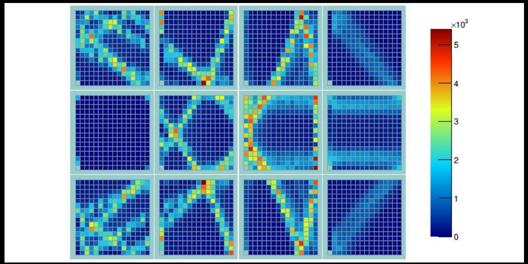
- This work was accomplished during the detector proposal and provided valuable insights in a multi-dimensional design space with multiple objective characterizing the detector performance (e.g., KF efficiency, momentum and angular resolution)
- This combined with other aspects like risk mitigation and costs reduction helped designing the ECCE reference detector. This reference is the new baseline for a new optimization phase as we are also moving towards the collaboration formation
- Consolidation of technology choice and optimization of design will be supported by:
    - Always more realistic effects integrated in the simulations, e.g., beam background
    - Integration of reconstruction algorithms and utilization without truth information (e.g. track finding for tracking) — N.b., reconstruction should be “flexible” against changes in design
    - Explore physics-driven optimization — include physics observables/full analysis as objectives
    - Extension of the design optimization to a larger system of sub-detectors, e.g., tracker + PID
        - Previous studies of dRICH show how this detector critical for PID in the hadronic endcap can benefit from AI-assisted design

Particle Identification with Cherenkov

		electrons/photons		π/K/p	
eta	Nomenclature	PID	Min E Photon	P-range [GeV/c]	Separation
-3.5 to -2.0	Backward	π suppression up to 1:1E-4	20 MeV	≤ 10 GeV/c	≤ 3σ
-2.0 to -1.0	Backward	π suppression up to 1:1E-3 - 1:1E-2	50 MeV		
-1.0 to 1.0	Barrel	π suppression up to 1:1E-2	100 MeV	≤ 6 GeV/c	
1.0 to 3.5	Forward	3σ e/π up to 15 GeV/c	50 MeV	≤ 50 GeV/c	

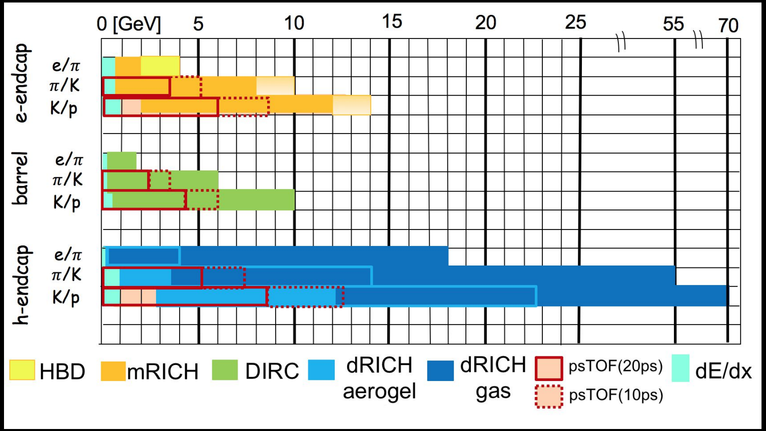


- Cherenkov detectors form the backbone of PID at EIC
  - Currently, all EIC detector designs use a dual radiator ring-imaging Cherenkov detector (RICH) in the hadron direction, a DIRC (detection of internally reflected Cherenkov light) in the barrel, and a modular RICH in the electron direction.
  - Simulating these detectors is typically compute expensive, involving many photons that need to be tracked through complex surfaces.
  - All three rely on pattern recognition of ring images in reconstruction, and the DIRC is the one having the more complex ring patterns!



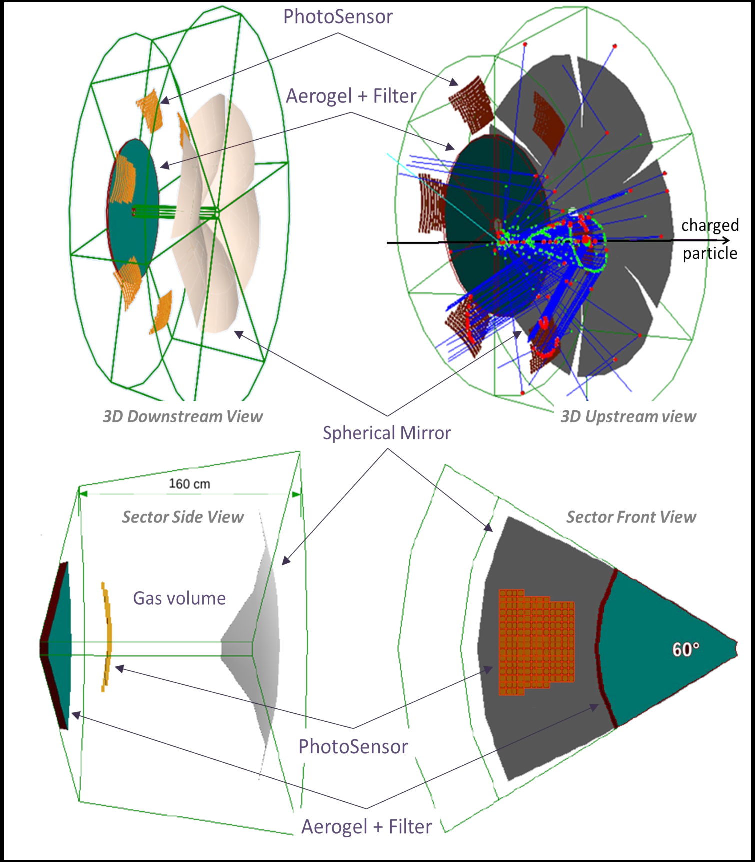
dRICH: ante-proposal

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

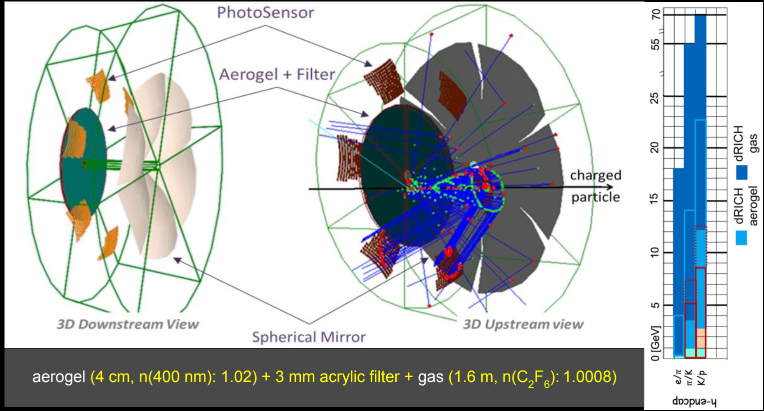
aerogel (4 cm, n(400 nm): 1.02)
+ 3 mm acrylic filter
+ gas (1.6 m, n(C_2F_6): 1.0008)



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

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

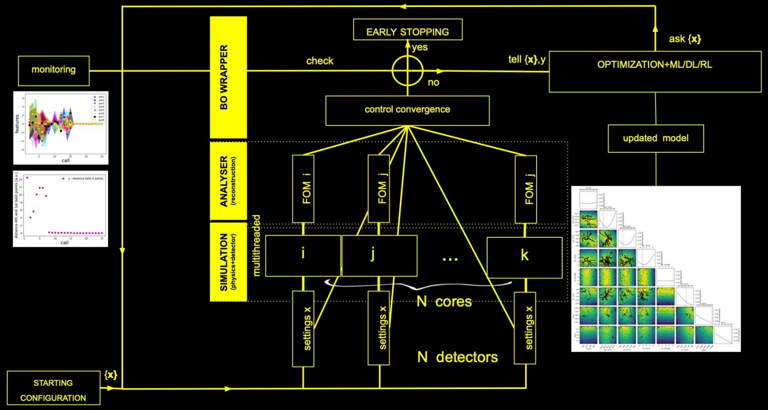
Come up with a smart objective; study / characterize properties (noise, stats needed etc): simulation + reconstruction

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

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

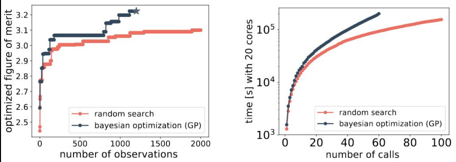
3

Optimization framework (embed convergence criteria)

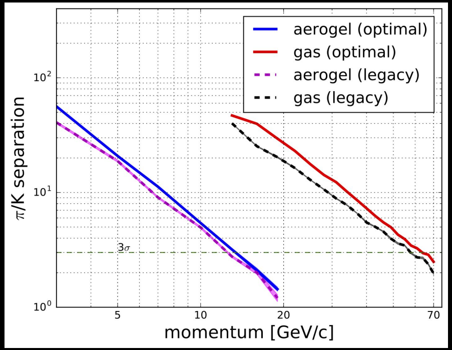


4

Analysis + Validation

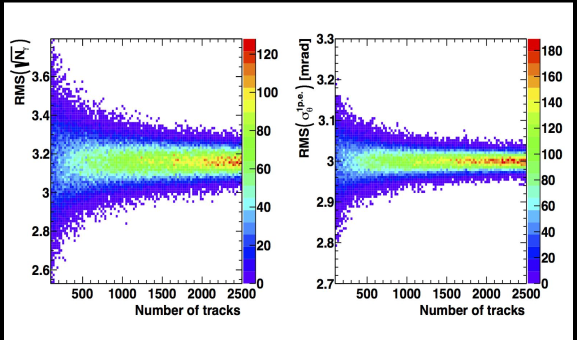


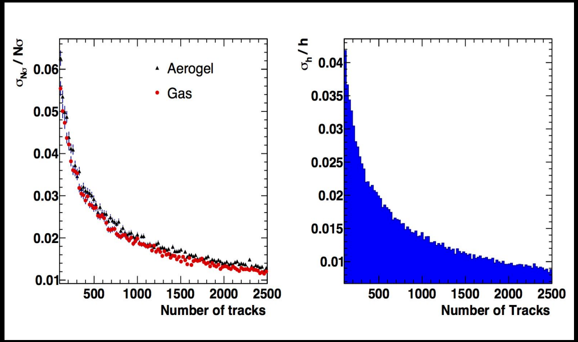
principled vs random



dRICH: ante-proposal

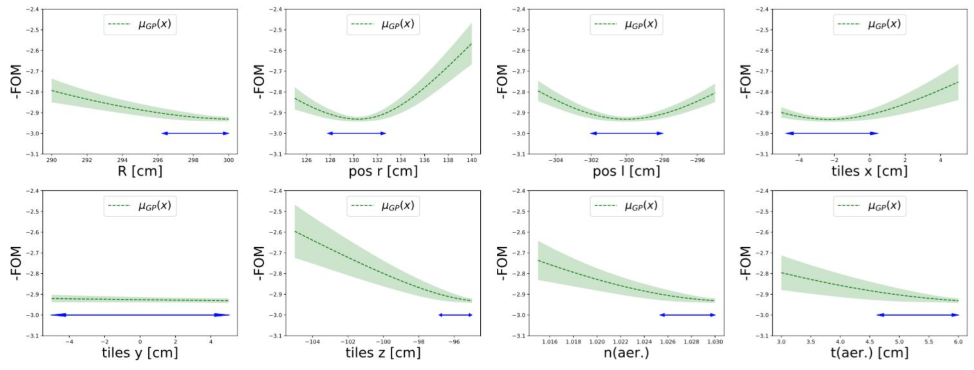
- Dedicated studies to characterize the noise as this is an optimization of a noisy function





- Ranges depend mainly on mechanical constraints and optics requirements. These requirements can change in the next future based on inputs from prototyping.

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]



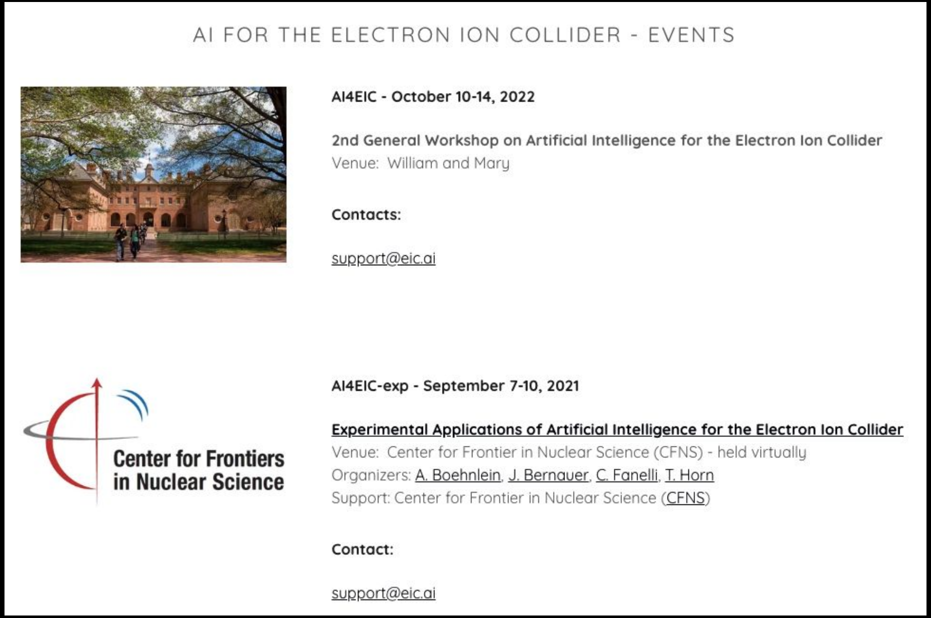
- Larger than the construction tolerances on each parameter.

EICUG AI WG (AI4EIC)

First Workshop on September 2021 at CFNS

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

Workshops


AI FOR THE ELECTRON ION COLLIDER - EVENTS

AI4EIC - October 10-14, 2022

2nd General Workshop on Artificial Intelligence for the Electron Ion Collider
Venue: William and Mary

Contacts:

support@eic.ai

Center for Frontiers in Nuclear Science

AI4EIC-exp - September 7-10, 2021

Experimental Applications of Artificial Intelligence for the Electron Ion Collider
Venue: Center for Frontier in Nuclear Science (CFNS) - held virtually
Organizers: A. Boehnlein, J. Bernauer, C. Fanelli, T. Horn
Support: Center for Frontier in Nuclear Science (CFNS)

Contact:

support@eic.ai

https://eic.ai

Meetings

AI4EIC Meeting on Detector Design:
https://indico.bnl.gov/event/16328/

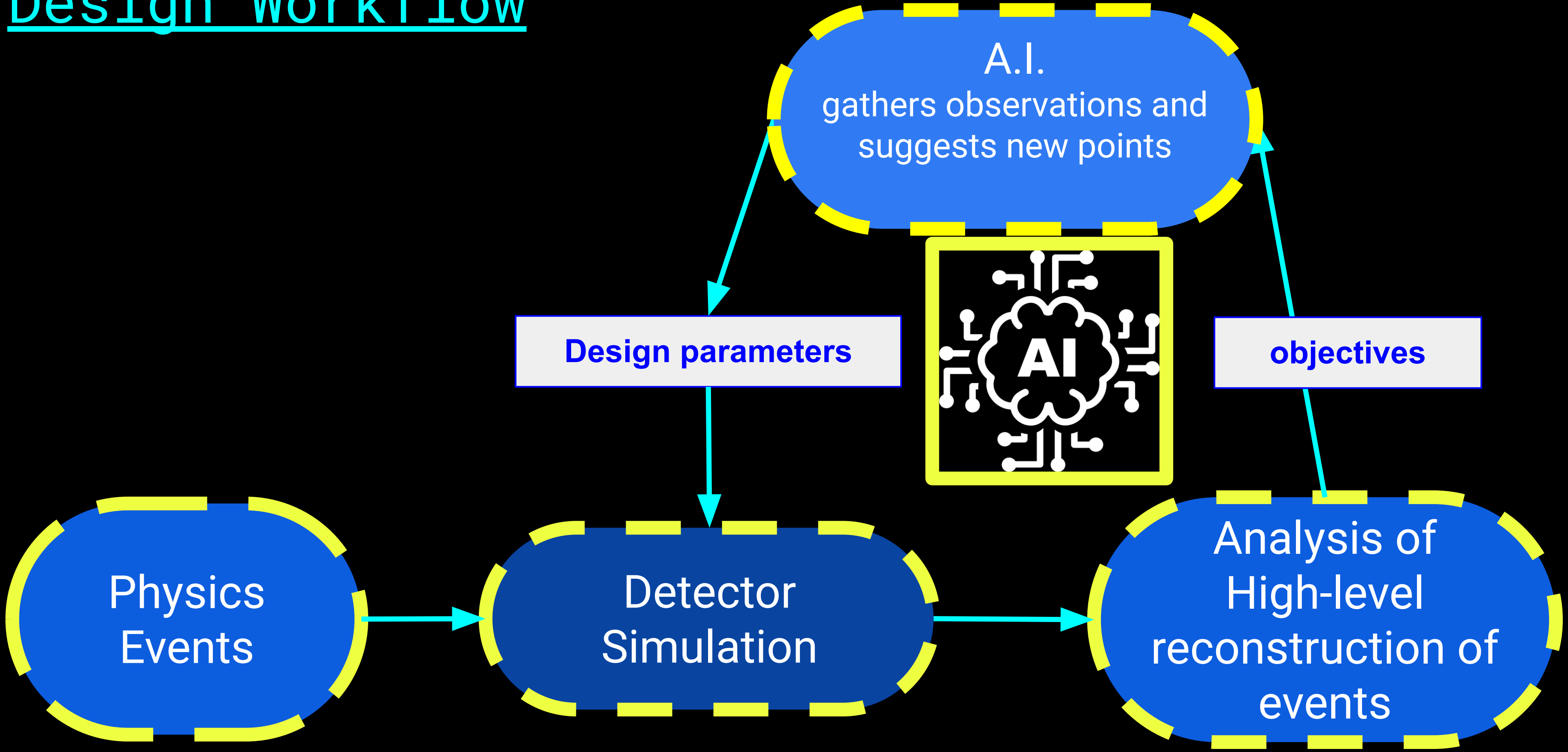
July 20, 9-11am ET

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 can be one of the first experiments to be designed with the support of AI and 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 utilized to optimize the reference detector; we anticipate roughly 1M CPU-core hours/year for these studies which will be extended to include
  - More realistic effects in the simulation and reconstruction techniques
  - A larger system of sub-detectors to include, e.g, detectors like the dRICH, in addition to the tracker system
- Design optimization pipelines of increased complexity can take advantage of distributed computing.

Backup

Design Workflow



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

Speed-up

- In general speed-up is reached by:
    - Hardware-based solutions to accelerate traditional algorithms
    - Hybrid ML/traditional techniques
    - End-to-end ML methods

- Of course the overall performance improvement gained by optimizing a single part of a system is limited by the fraction of time that the improved part is actually used (Amdhal's law)

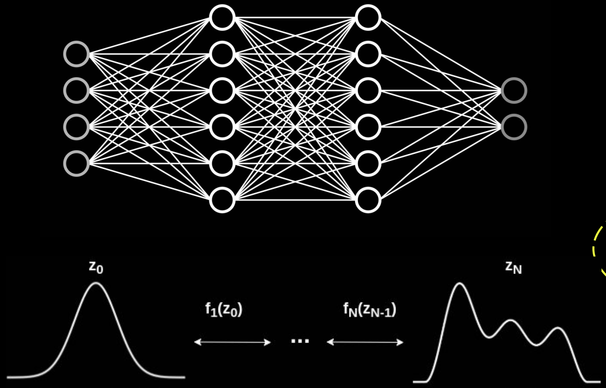
$$\frac{1}{(1-P)+\frac{P}{S}}$$

P: fraction of execution time that the part benefiting from improved resources originally occupied

s: speed-up of the part benefiting

- What follows will show some ML/DL example and is not meant to be exhaustive — argument for another talk; see P. Harris' lectures

Design Workflow



A.I.
gathers observations and suggests new points

✔ **Design parameters**
X

AI

✔ **objectives**
Y

Physics Events → Detector Simulation → Analysis of High-level reconstruction of events

…with large datasets…

Projective vs non-projective

arXiv:2205.09185