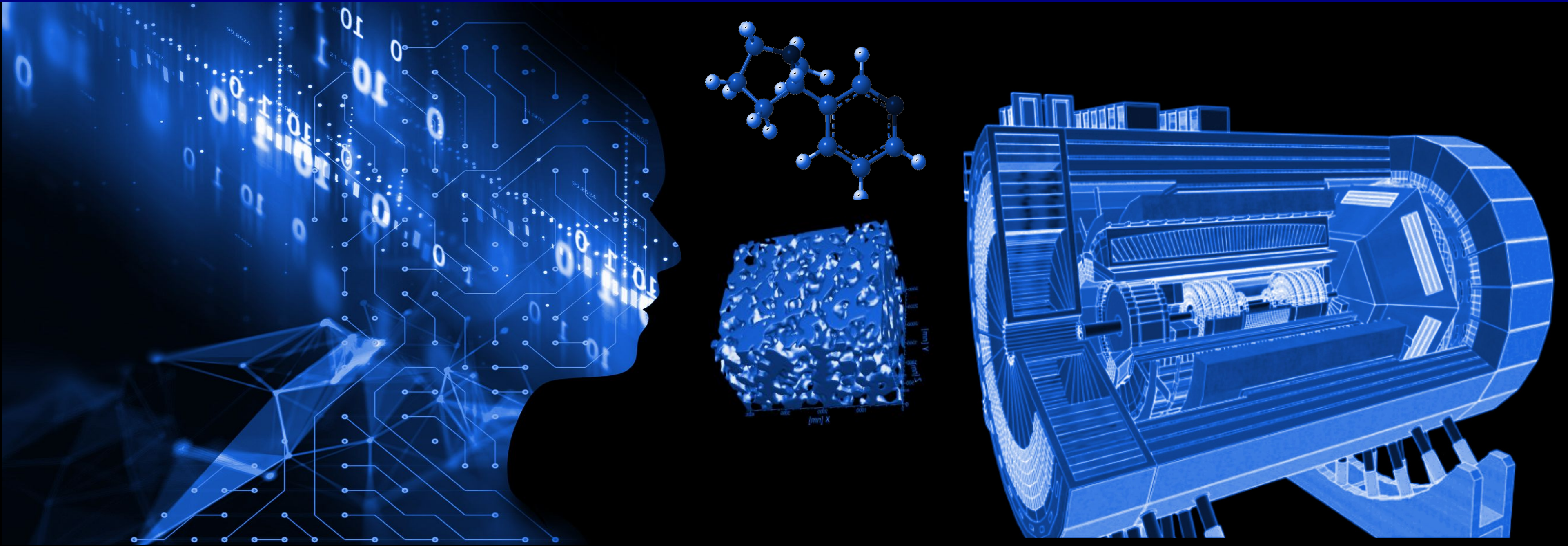
AI-assisted Design and R&D of advanced Experimental Systems



AI@INFN
Bologna, May 2-3

Cristiano Fanelli

AI for Design

It is a relatively new but active area of research.
Many applications in, e.g., industrial material, molecular and drug design. Z. Zhou et al., *Scientific Reports*, vol. 9, no. 1, pp. 1–10, 2019

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]

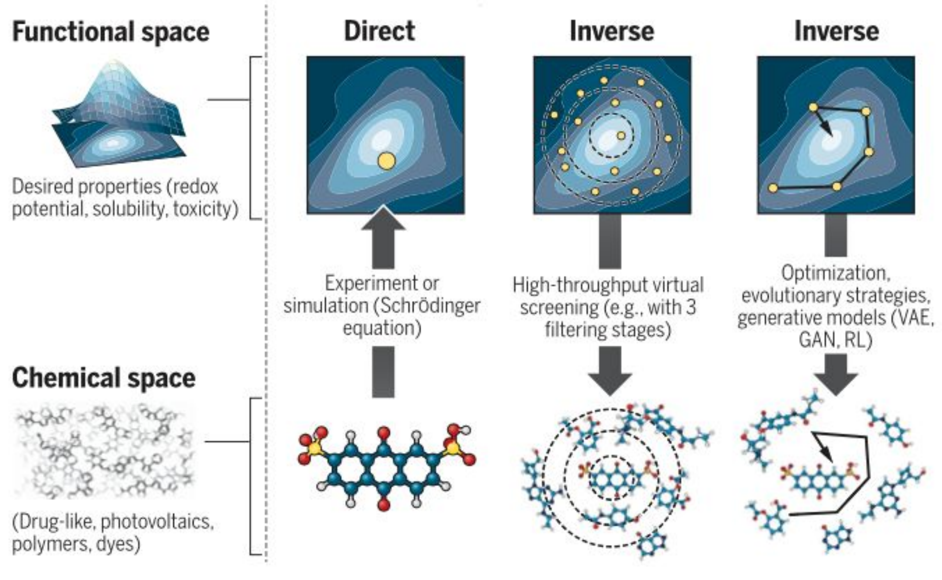


Fig. 2. Schematic of the different approaches toward molecular design. Inverse design starts from desired properties and ends in chemical space, unlike the direct approach that leads from chemical space to the properties.

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

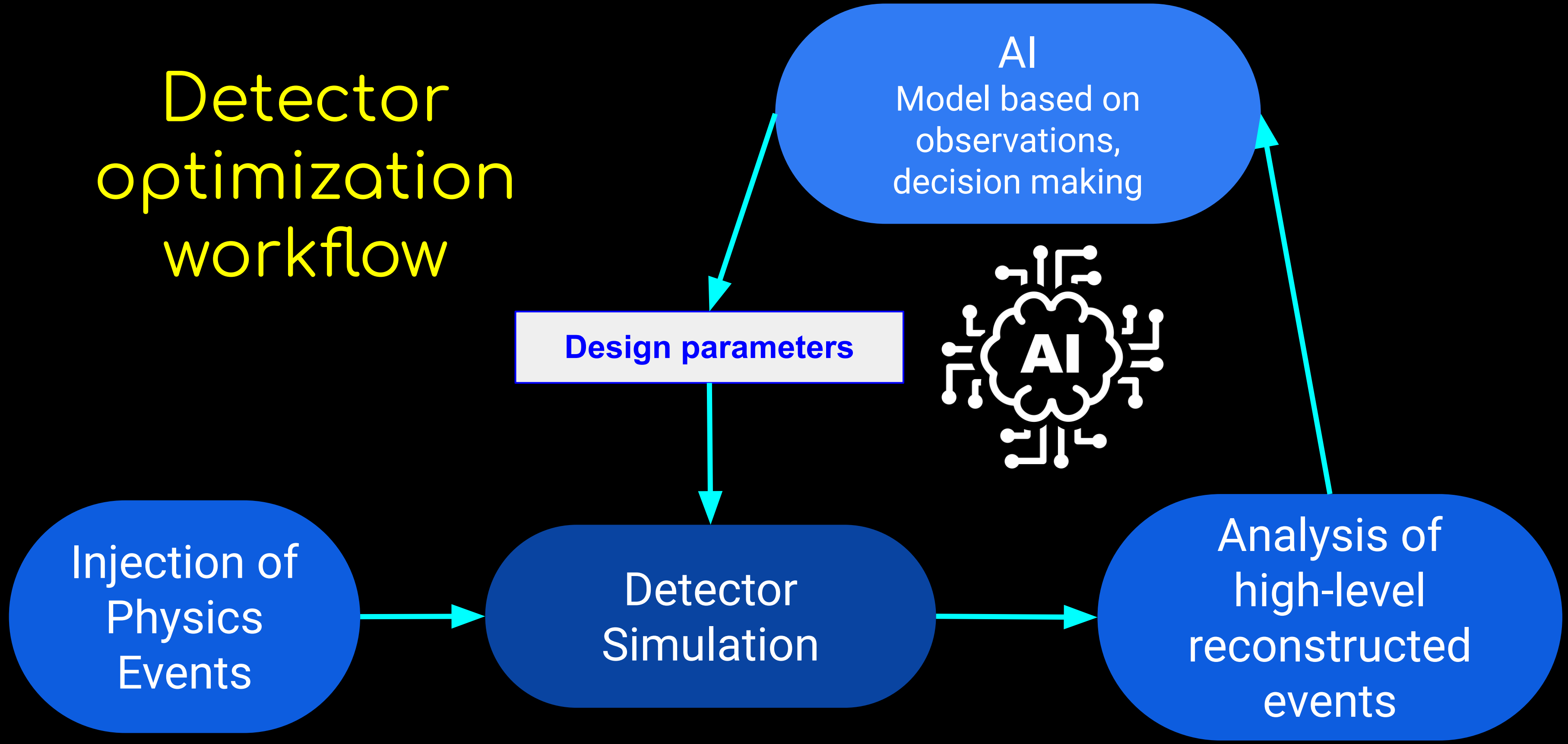
Experimental Design in NP/HEP

- When it comes to designing detectors and accelerators with AI this is an area at its “infancy”. What follows uses “detector” as example but applies to both detector and accelerator.
- Typically full detector design is studied once the subsystem prototypes are ready (phase constraints from the full detector or outer layers are taken into consideration).
- Need to use advanced simulations which are computationally expensive (Geant).
- Many parameters (and multiple objective functions): curse of dimensionality [1].
- Entails establishing a procedural 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.

AI offers SOTA solutions to solve complex optimization problems in an efficient way

[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

Detector optimization workflow

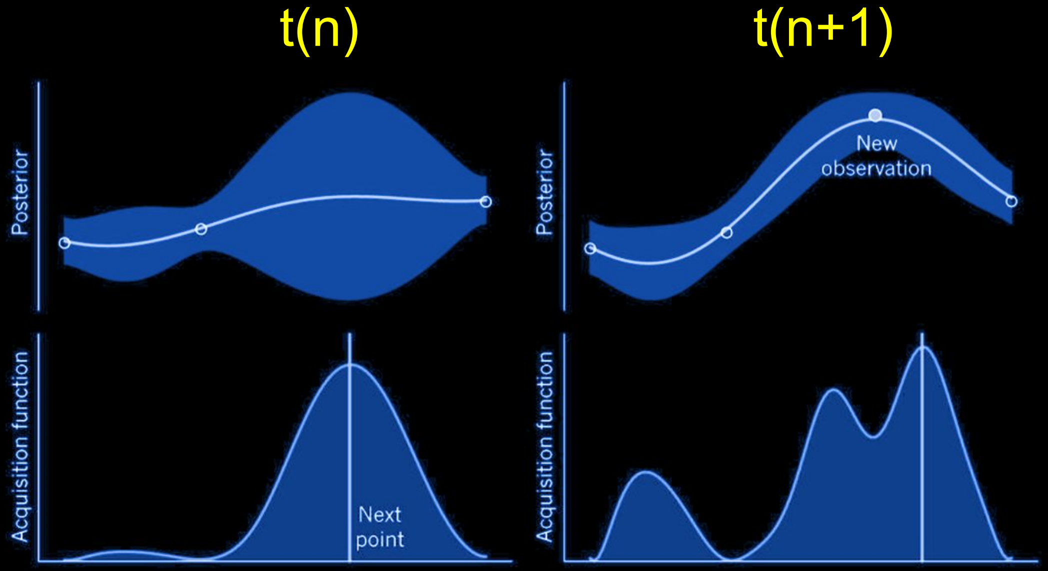


AI/ML gathers observations and suggests new design points in an efficient way

(AI/ML can also speed-up the simulation/reconstruction stack; cf. Amdahl's law)

Bayesian Optimization

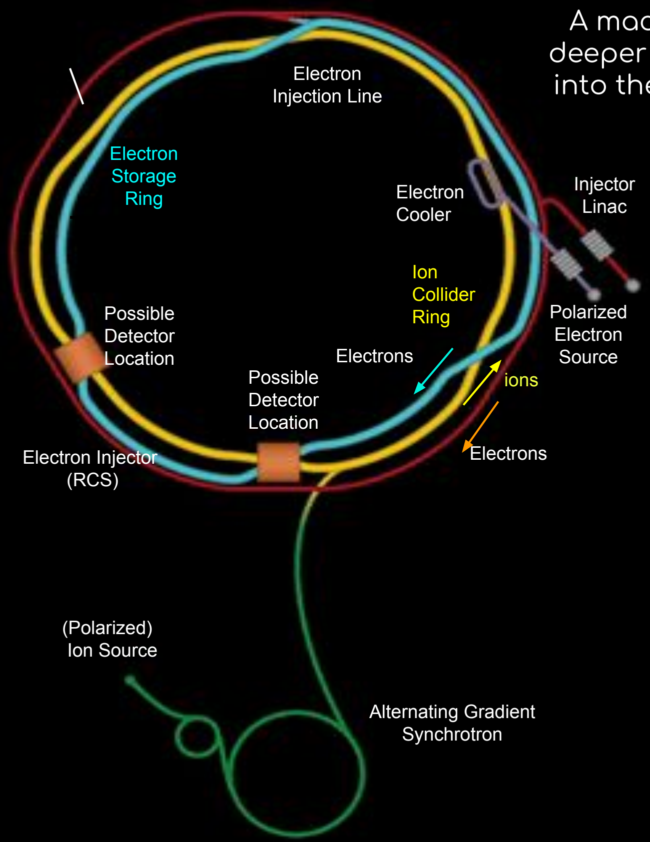
- BO is a sequential strategy developed for global optimization.
- After gathering evaluations we builds a posterior distribution used to construct an **acquisition function.**
- This cheap function determines what is **next query point.**



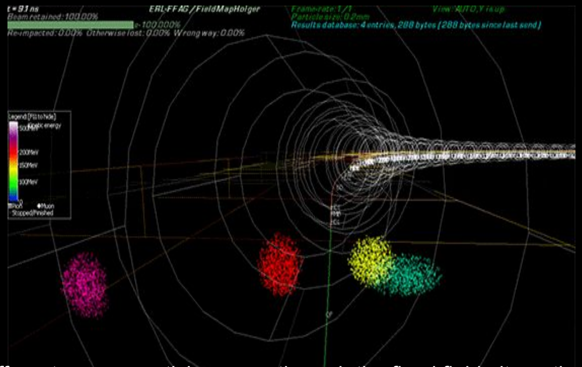
1. Select a Sample by Optimizing the Acquisition Function.
2. Evaluate the Sample With the Objective Function.
3. Update the Data and, in turn, the Surrogate Function.
4. Go To 1.

http://krasserm.github.io/2018/03/21/bayesian-optimization/
http://krasserm.github.io/2018/03/19/gaussian-processes/

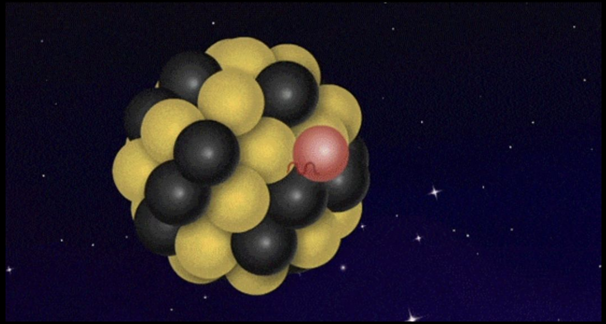
The Electron Ion Collider



A machine for delving deeper than ever before into the building blocks of matter



Different energy particles move through the fixed field alternating linear gradient accelerator.



Credits:scitechdaily

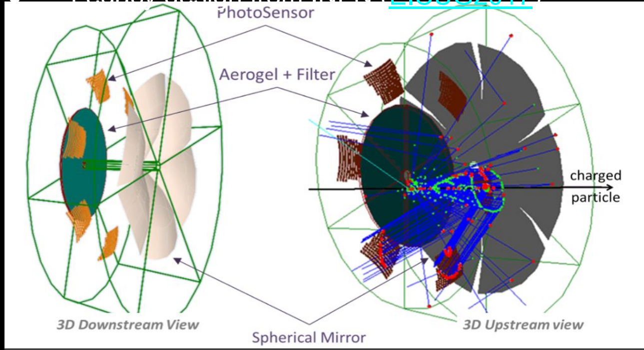
- Beams of electrons and high-energy protons or heavier atomic nuclei
- Wide coverage of CoM energy $\sqrt{s_{e\text{-}p}}$ ~ (20-140) GeV

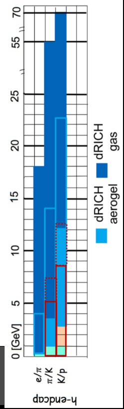
Will be constructed over ten years at an estimated cost between $1.6 and $2.6 billion

Dual RICH: ante proposal

E. Cisbani, A. Del Dotto, CF*, M. Williams et al. *JINST* 15.05 (2020): P05009

- 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

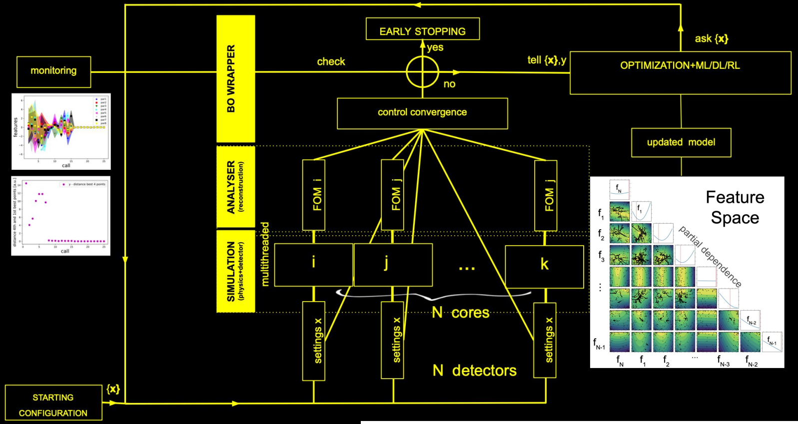
parameter	description	range [units]	tolerance [units]
R	mirror radius	[290,300] [cm]	100 [μm]
pos r	radial position of mirror center	[125,140] [cm]	100 [μm]
pos l	longitudinal position of mirror center	[-305,-295] [cm]	100 [μm]
tiles x	shift along x of tiles center	[-5,5] [cm]	100 [μm]
tiles y	shift 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

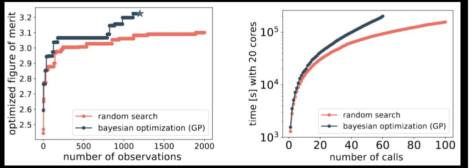
$$N\sigma = \frac{||\langle\theta_K\rangle - \langle\theta_\pi\rangle||\sqrt{N_\gamma}}{\sigma_\theta^{1p.e.}}$$

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

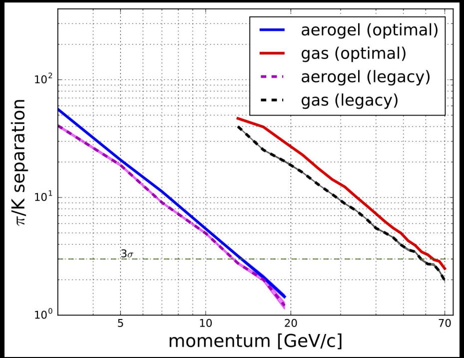
3 Optimization framework (embed convergence criteria)



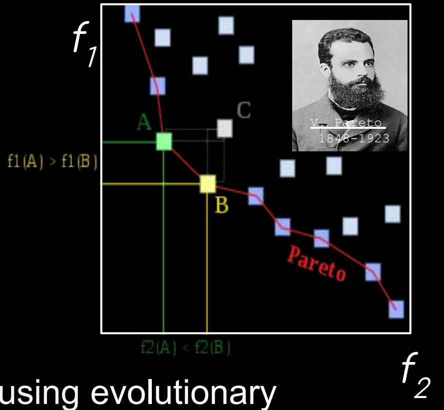
4 Analysis + Validation



principled vs random



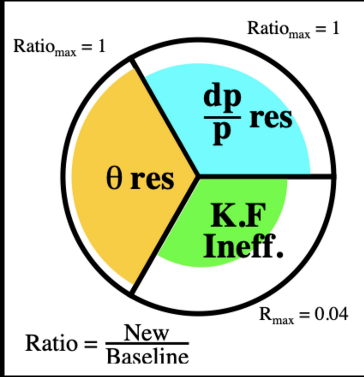
Multi-Objective Optimization



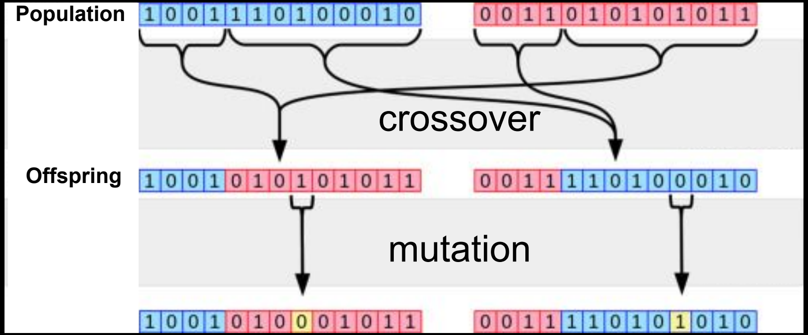
- The problem becomes challenging when the objectives are of conflict to each other, that is, the optimal solution of an objective function is different from that of the other.
- In solving such problems, with or without constraints, they give rise to a trade-off optimal solutions, popularly known as 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).

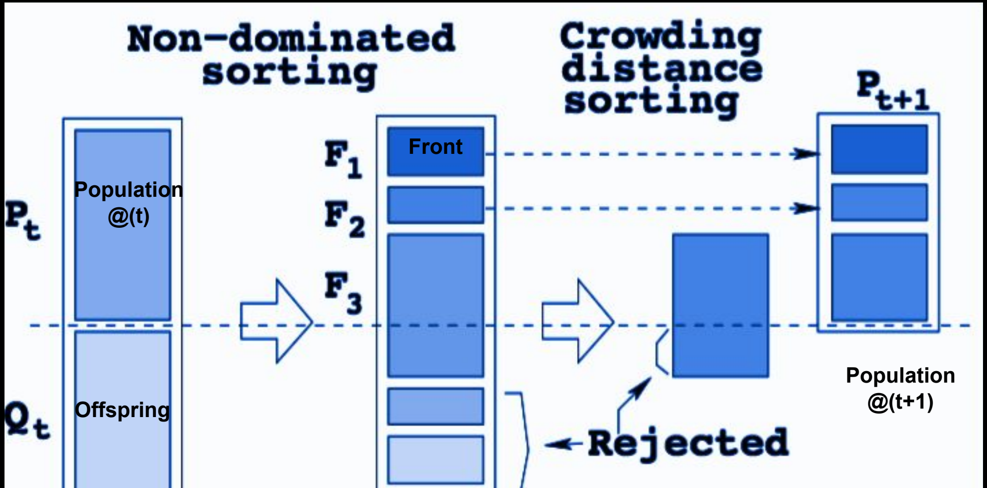
The ECCE Inner Tracker Design Optimization considers simultaneously:

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



Elitist Non-Dominated Sorting Genetic



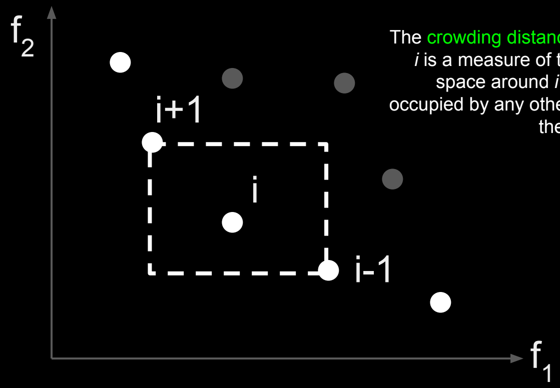


[1] Deb, K., et al. "A fast and elitist multiobjective genetic algorithm" *IEEE transactions on evolutionary computation* 6.2 (2002): 182-197.

This is one of the most popular approach (>35k citations on google scholar), characterized by:

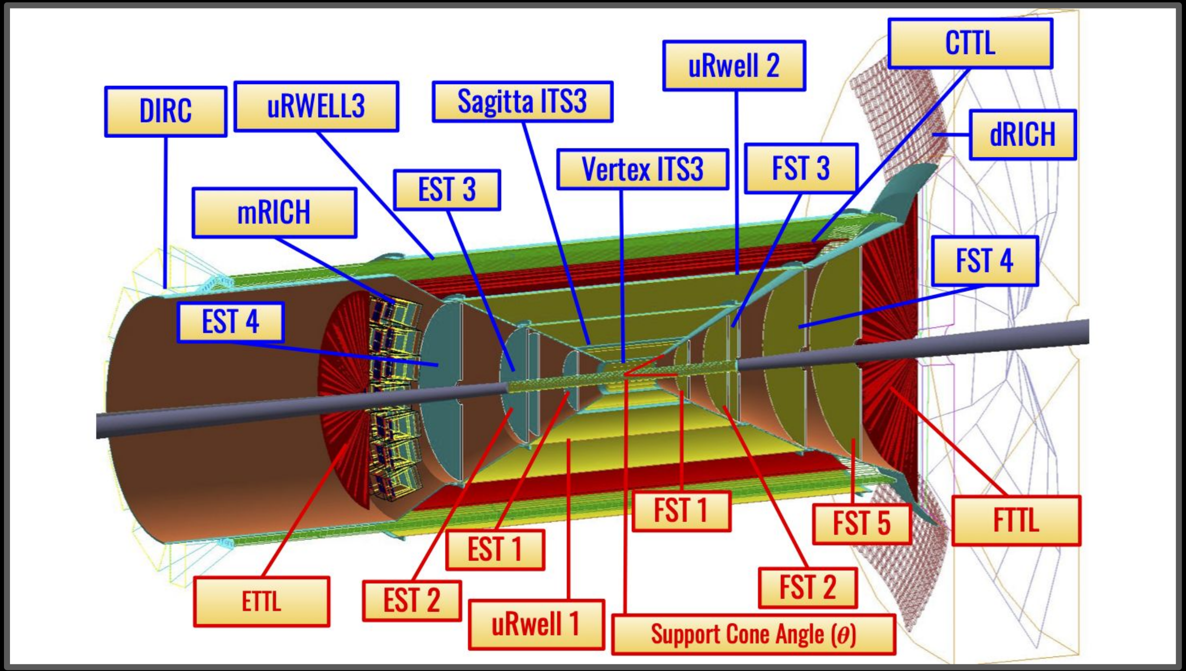
- Use of an elitist principle
- Explicit diversity preserving mechanism
- Emphasis in non-dominated solutions

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.



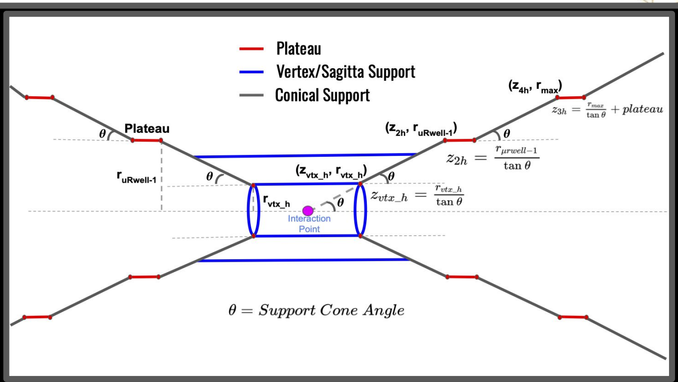
The crowding distance d_i of point *i* is a measure of the objective space around *i* which is not occupied by any other solution in the population.

The EIC Detector Inner Tracker



Ongoing R&D that includes in the optimization the support structure of the inner tracker

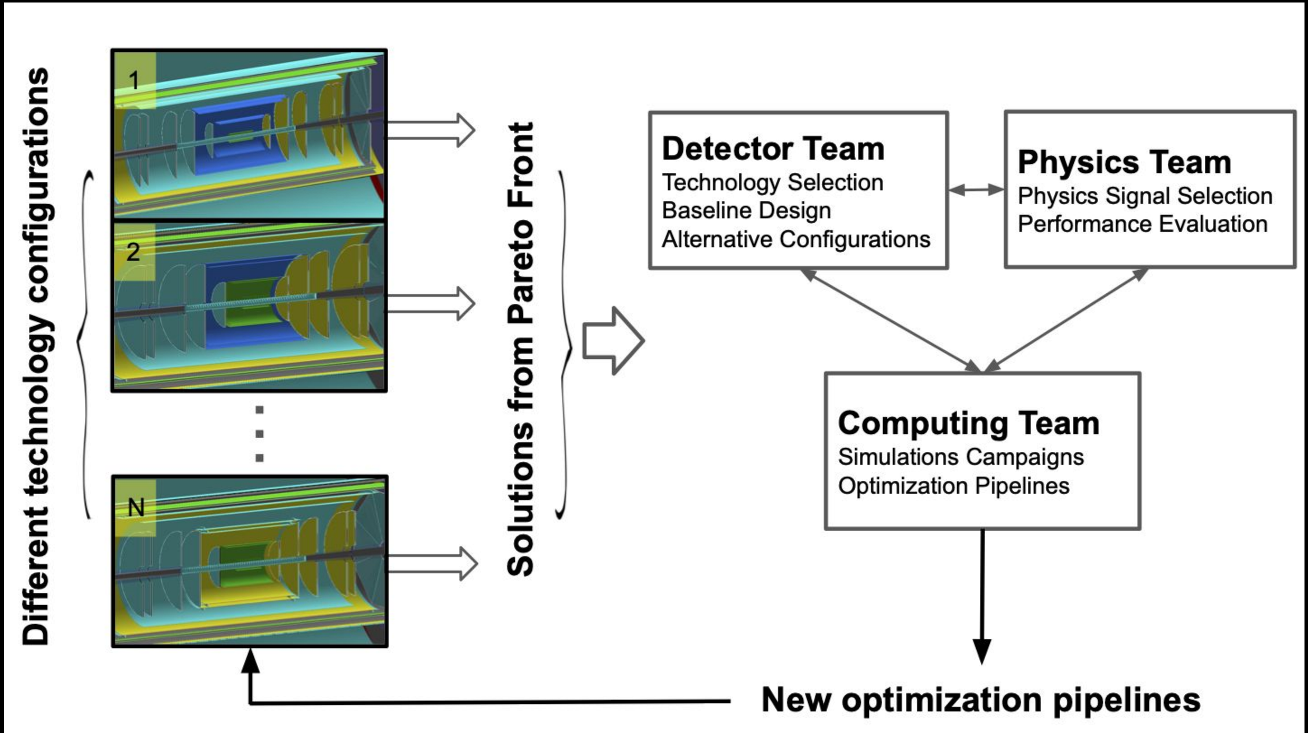
Parametrization



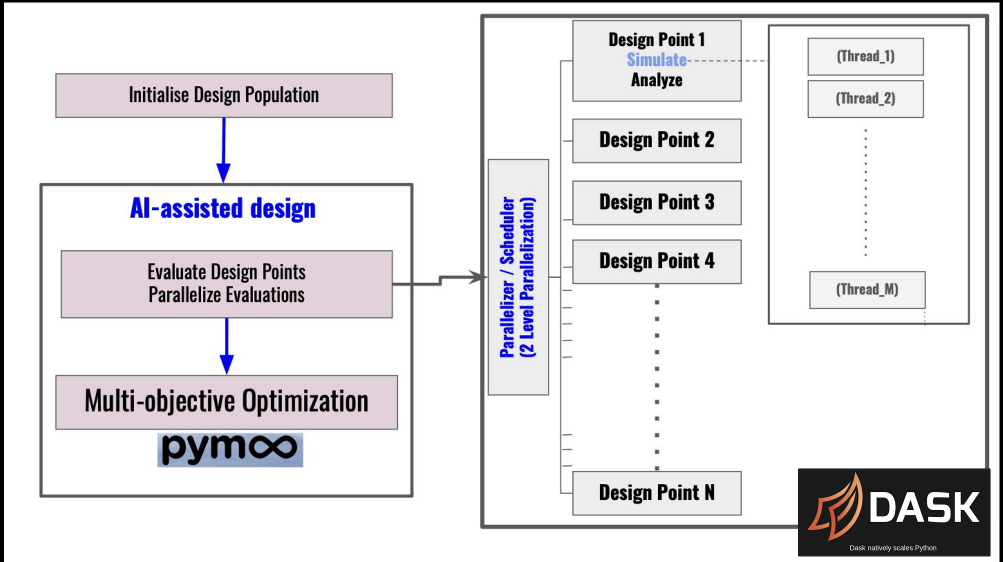
Integration during EIC Detector Proposal

Optimization does not mean necessarily "fine-tuning"

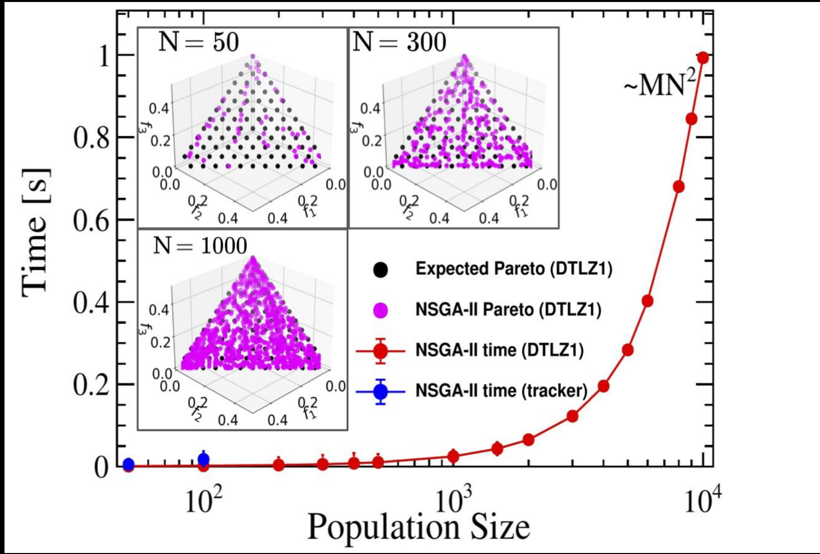
- We want to use these algorithms to: (1) steer the design and suggest parameters that a "manual"/brute-force optimization will likely miss to identify; (2) further optimize some particular detector technology (see d-RICH paper, e.g., optics properties)
- AI allows to capture hidden correlations among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



Computational Resources



time taken by GA + sorting

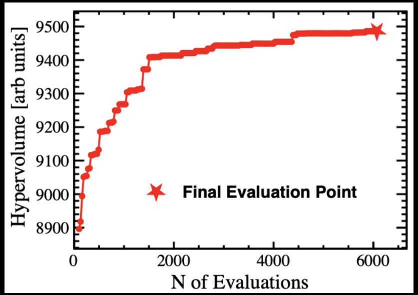


description	symbol	value
population size	N	100
# objectives	M	3
offspring	O	30
design size	D	11 (9)
# calls (tot. budget)	–	200
# cores	–	same as offspring
# charged π tracks	N_{trk}	120k
# bins in η	N_{η}	5
# bins in p	N_p	10

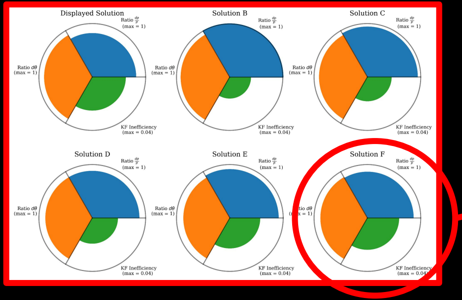
- Used a test problem DTLZ1
- Verified scaling following MN^2 and convergence to true front
- ~1s/call with 10^4 size!
- For 11 variables and 3 objectives needs ~ 10000 evaluations to converge
  ~10k CPUhours / pipeline

"Navigate" Pareto Front

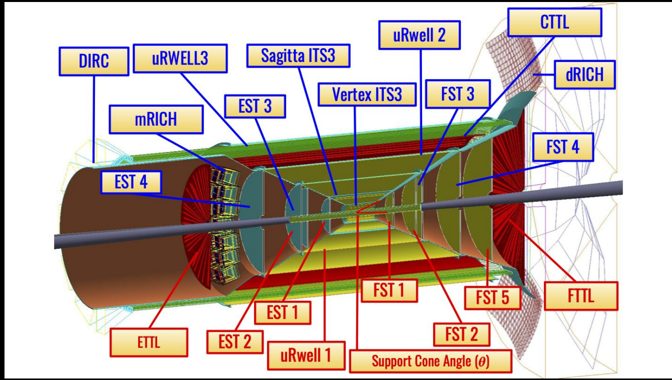
1 Can take a snapshot any time during evaluation



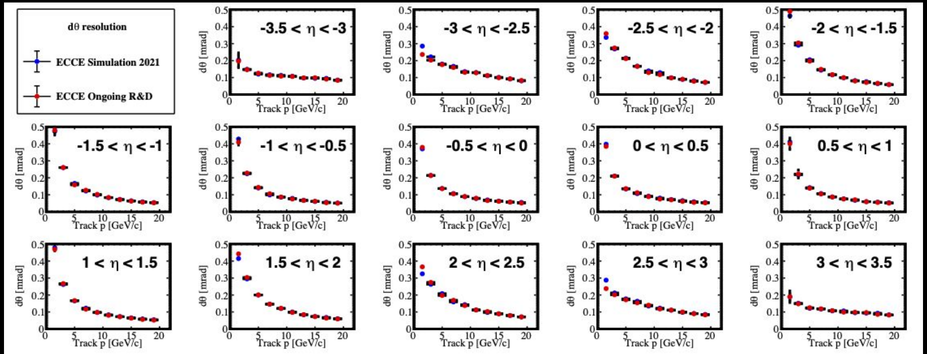
2 Updated Pareto Front at time t

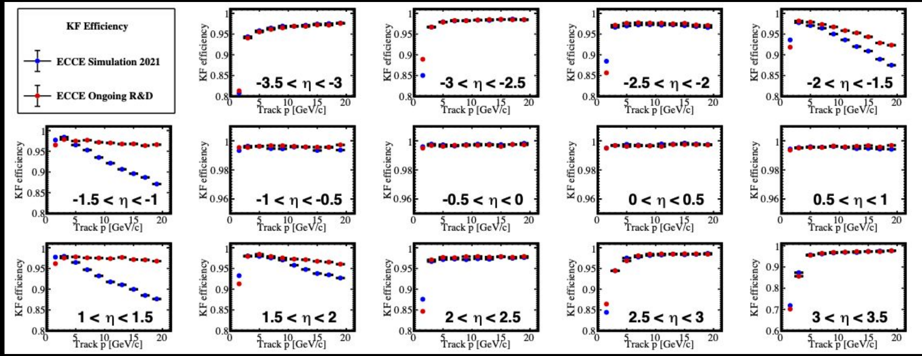


3 At each point in the Pareto front corresponds a design



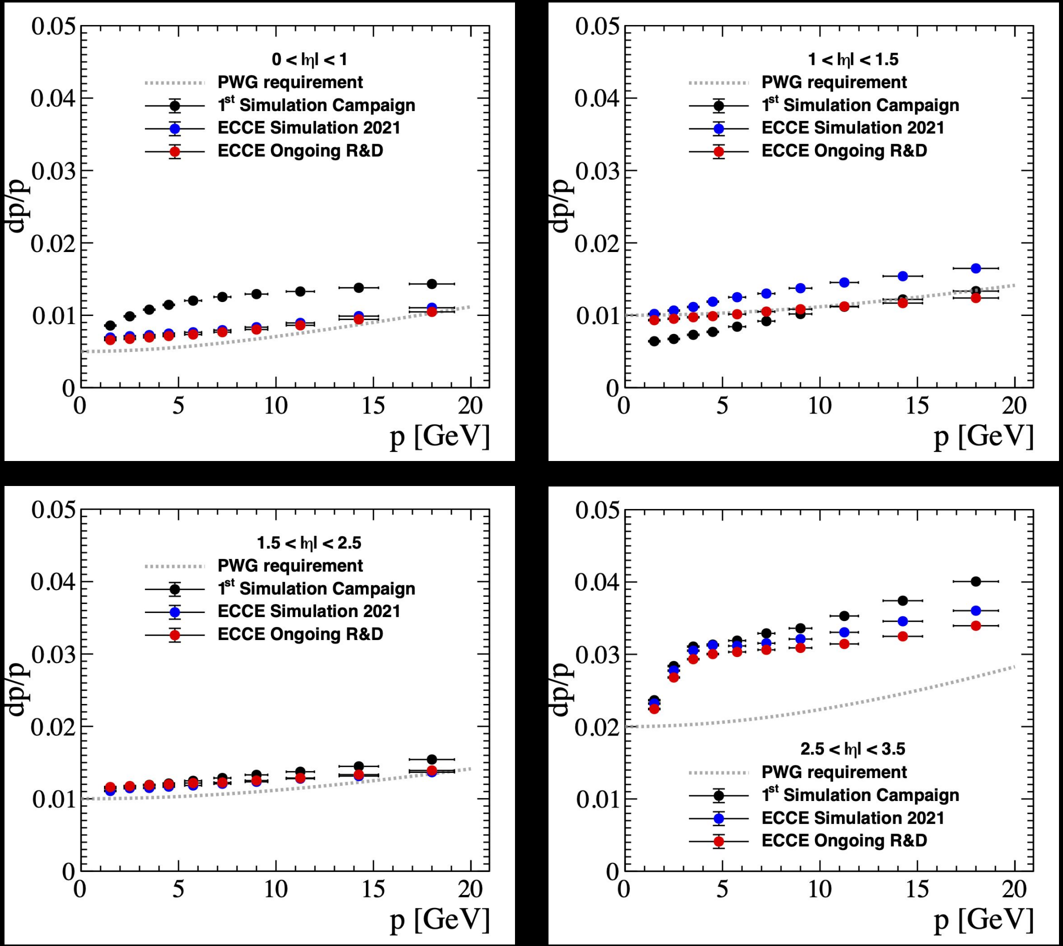
4 Analysis of Objectives (momentum resolution, angular resolution, KF efficiency)





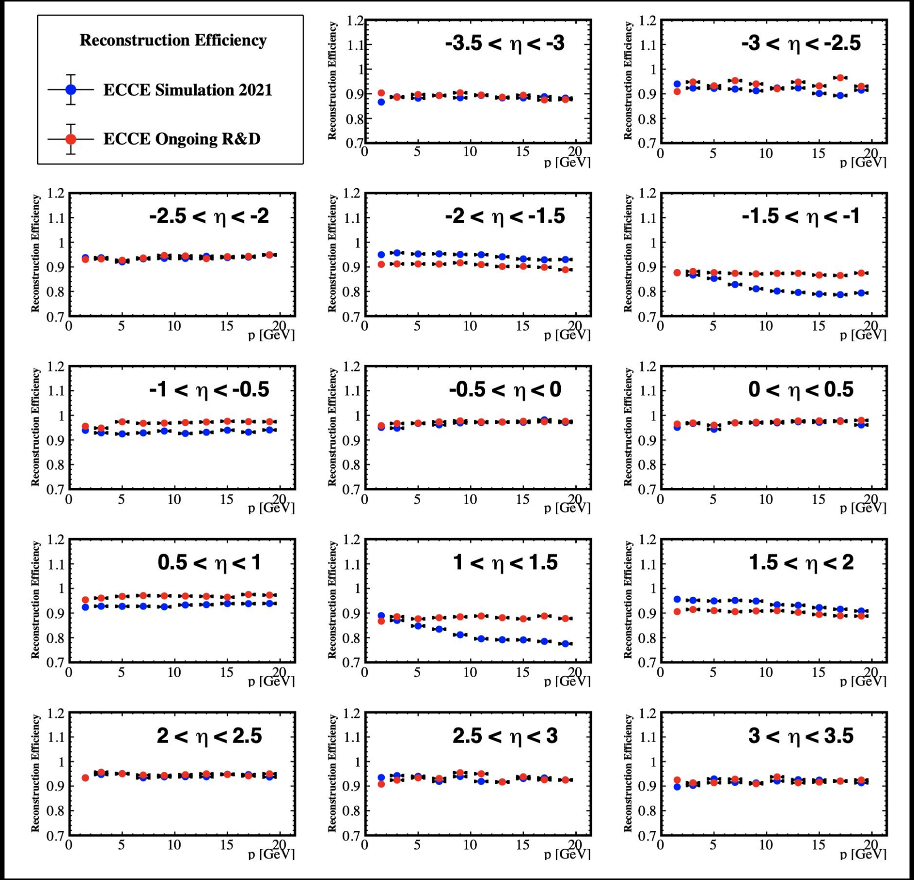
“Evolution”

- 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

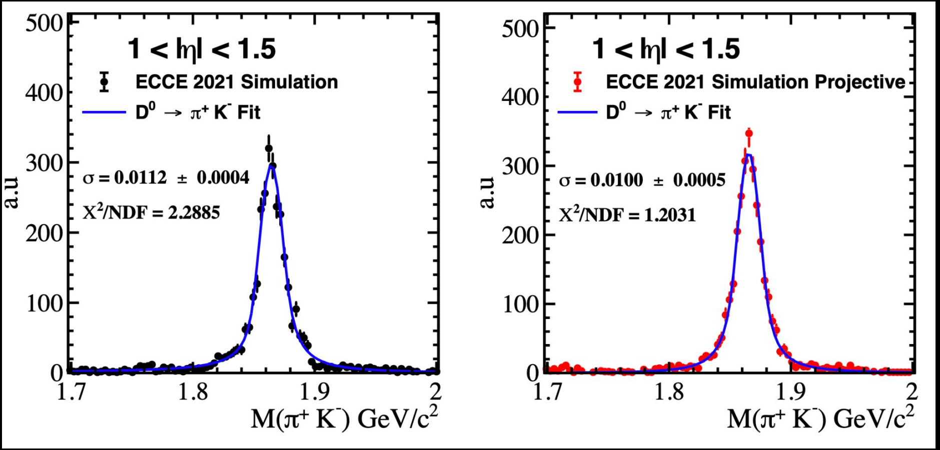
Reconstruction Efficiency



Performance evaluated after optimization process (both designs).

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

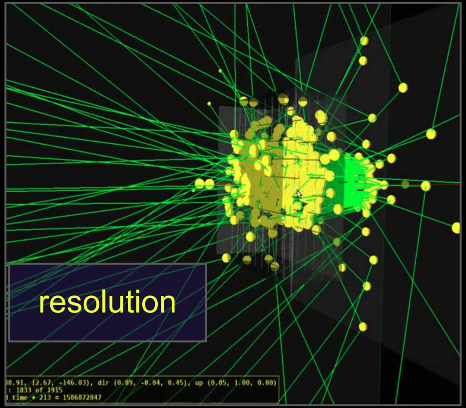
D0 invariant mass from semi-inclusive deep inelastic scattering



Novel Aerogel Material **aefib**

The team: V. Berdnikov, J. Crafts, E. Cisbani, CF, T. Horn, R. Trotta

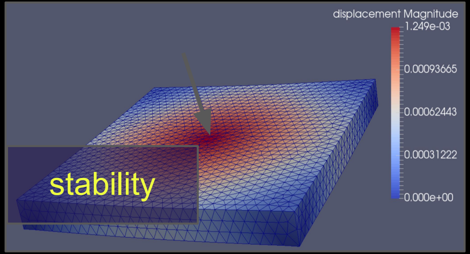
- Aerogels with low refractive indices are very fragile tiles break during production and handling, and their installation in detectors.
- To improve the mechanical strength of aerogels, Scintilex developed a reinforcement strategy. The general concept consists of introducing fibers into the aerogel that increase mechanical strength, but do not affect the optical properties of the aerogel.
- Paper in preparation.



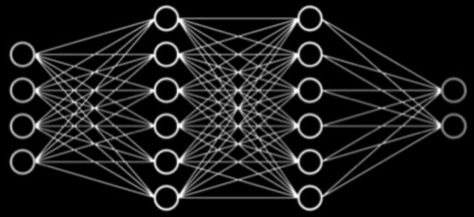
Software Stack

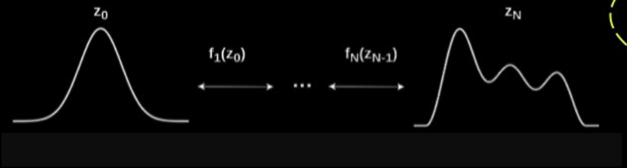
Simple Ring Imaging CHerenkov Geant4 based simulation
Aerogel + Optical Fibers

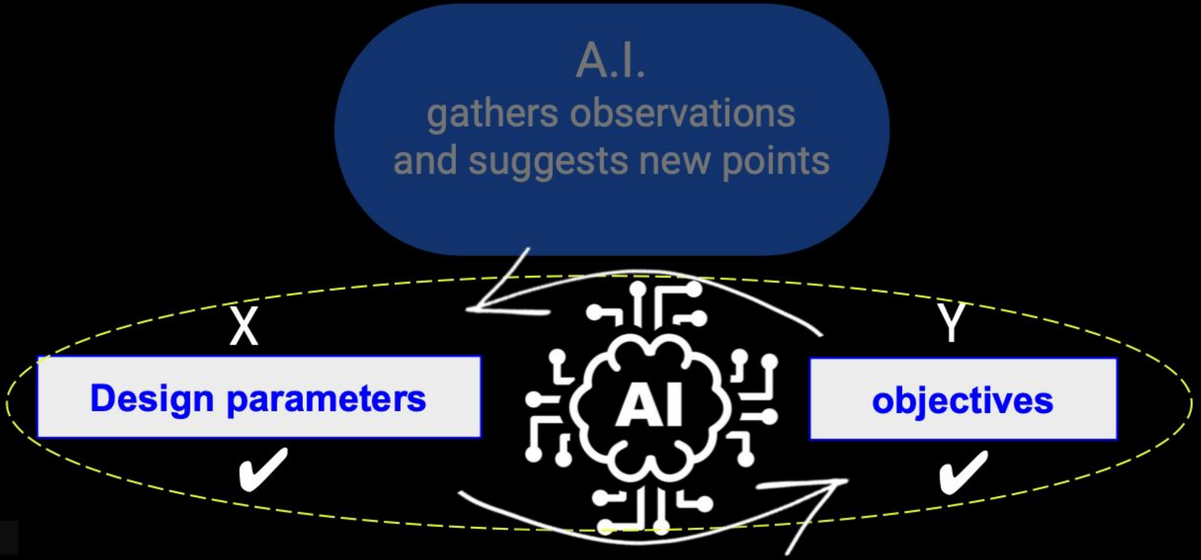
Gmsh - define geometry and produce mesh
ElmerGrid - convert the gmsh mesh to elmer compatible mesh
ElmerSolver - do modeling (solve linear and nonlinear equation)
Paraview - visualize Elmer Solver and provide a python interface to automate



… with larger datasets







Summary

- AI can assist the design and R&D of complex experimental systems by providing more efficient design (considering multiple objectives) and optimizing the computing budget needed to achieve that.
- EIC can be one of the first experiments to be designed with the support of AI.
- The ECCE consortium is leading these efforts with a multidimensional design and objective spaces.
- Additional considerations:
  - None ever accomplished a multi-dimensional / multi-objective optimization of the global design, i.e., made by many sub-detectors combined together, that can be solved with AI
  - Costs can be explicitly included during the optimization provided a reliable parametrization)
  - An intrinsic overhead regards compute expensive simulations + reconstruction/analysis. This can be speeded up embedding AI/ML in the SW stack.
  - Larger populations of design points can be simulated to improve accuracy of the Pareto front in multidimensional spaces with AI-based accelerated optimizations.

Likely future detectors will be designed with the help of AI achieving optimal performance and cost reduction. One of the conclusions from the DOE Town Halls on AI for Science on 2019 was that “*AI techniques that can optimize the design of complex, large-scale experiments have the potential to revolutionize the way experimental nuclear physics is currently done*”.