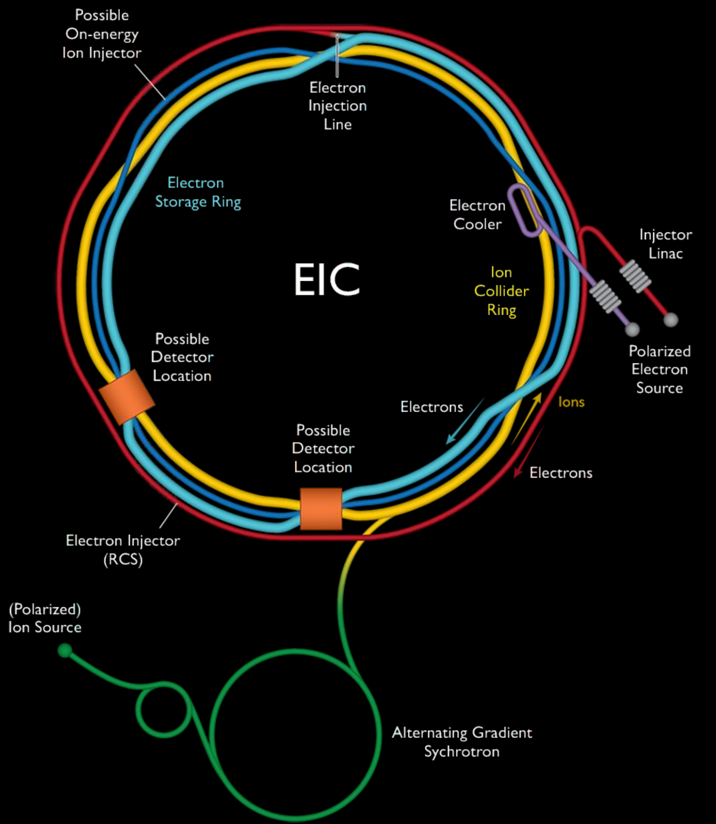
Artificial Intelligence for the ECCE Detector Design

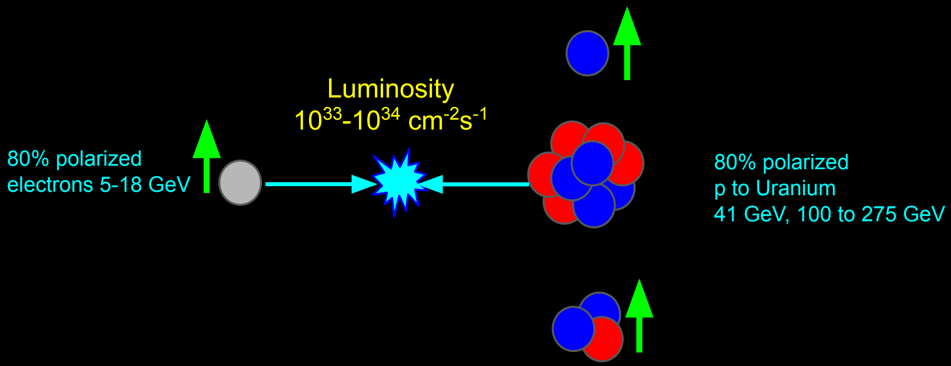
Cristiano Fanelli

The Electron Ion Collider

R. Ent, Overview of the EIC Program



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



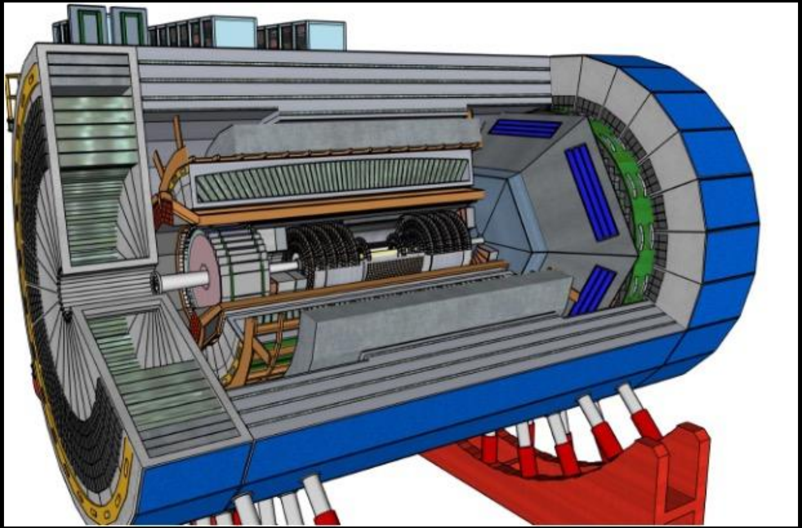
- Beams of electrons and high-energy protons or heavier atomic nuclei
- Wide coverage of CoM energy $\sqrt{s_{e\text{-}p}}$ ~ (20-140) GeV
- Two large acceptance detectors

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

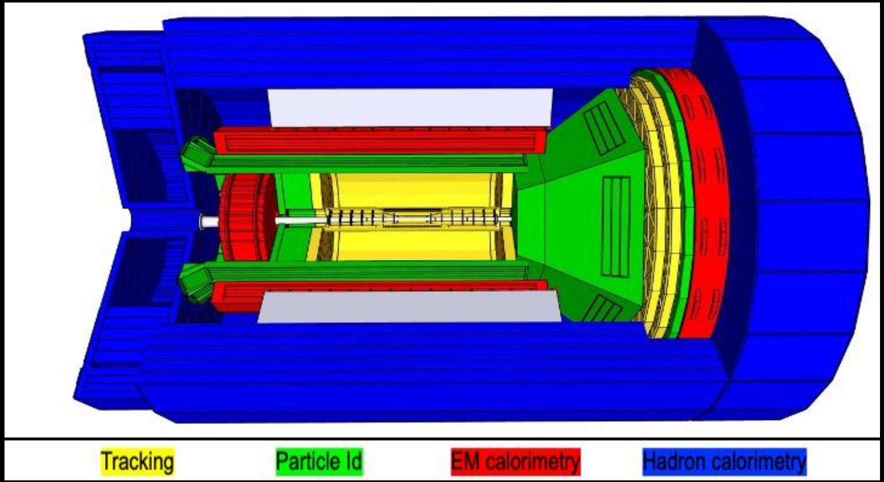
Example of EIC Central Detector

R. Ent, Overview of the EIC Program

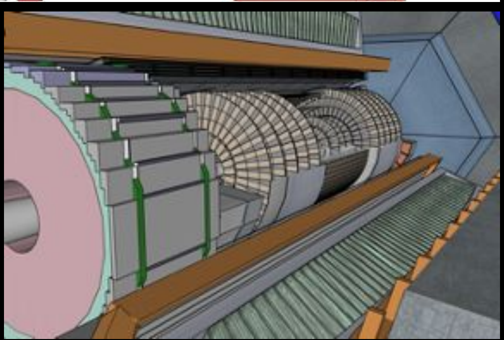
3D visualization with Sketchup



Based on new 3T Magnet



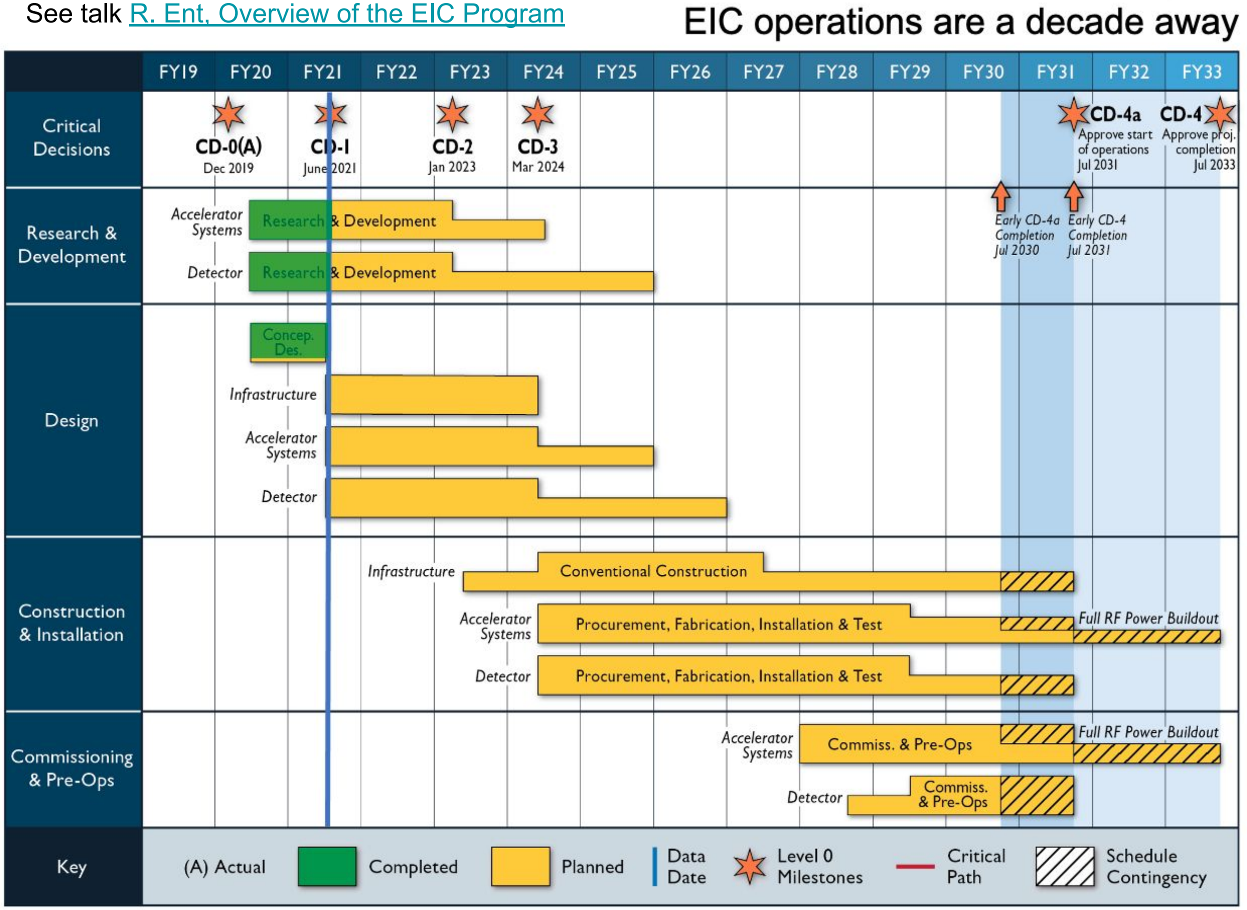
ECCE/1.5T



EIC and Timeline - How AI come into play?

- AI basically present in all phases of the EIC schedule
- The EIC R&D program can be one of the first to systematically leverage on AI during the detector design phase
- AI can advance research, design, and operation of the EIC. In the Yellow Report, Sec. 11.12 (Artificial Intelligence for the EIC detector), we individuate specific aspects that can be potentially tackled with AI.
- Supported by new approaches like Streaming RO, the EIC can become one of the first largely automated experiments (e.g., calibration)

See talk R. Ent, Overview of the EIC Program



Artificial Intelligence for the EIC Detector

From the EIC Yellow Report

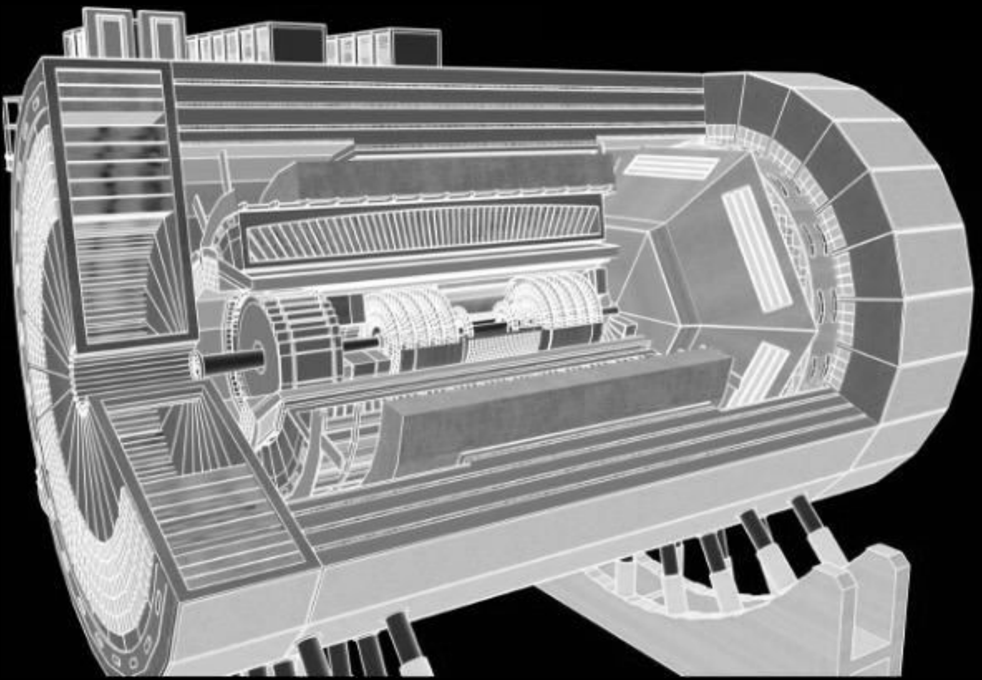
11.12 Artificial Intelligence for the EIC Detector

In the world of computing there is growing excitement for what is perceived as the revolution of the new millennium: artificial intelligence (AI). In particular the R&D program of the future EIC could be one of the first programs systematically exploiting AI. AI is becoming ubiquitous in nuclear physics [1724]. According to a standard taxonomy [1725], AI encompasses all the concepts related to the integration of human intelligence into machines; a subset of AI is machine learning (ML), which is usually grouped into supervised, unsupervised and reinforcement learning; deep learning (DL) is a particular subset of ML based on deep (*i.e.*, made by many hidden layers) neural networks, which is often considered the evolution of ML since it typically outperforms other methods when there is a large amount of data and features, provided sufficient computing resources. In the most frequent applications in our field, features are selected and a model is trained for classification or regression using signal and background examples.

Experimental particle and nuclear physics is big data [1726]: the gigantic data volumes produced in modern experiments are typically handled with “triggers”—a combination of

Partial list selected from the YR:

- Streaming Readout can further the convergence of online and offline analysis allowing for the incorporation of high level AI algorithms in the analysis pipeline: better data quality and shorter analysis cycle
- Reconstruction algorithms (e.g., tracking)
- Particle Identification/architecture for specific detectors (e.g., imaging Cherenkov)
- Event classification / Global PID
- Search for rare signatures (e.g., GlueX BDT)
- Architectures for specific physics: utilization of Jets (e.g., ML4Jets)
- Fast simulations
- Design



ECCE detector concept

Detector Design with AI

Detector Design with AI

- Designing detectors “with” AI is a new area of research at its infancy that can have a tremendous impact across many fields (NP, HEP, Astro-Phys). See lectures https://github.com/cfteach/AI4NP_detector_opt given at the AI4NP winter school https://indico.jlab.org/event/409/.
- It includes a broad range of approaches, from “optimizing” an existing expert-drawn baseline detector concept, to in principle letting AI design completely “new” and unseen configurations.
- New field, not many examples... Many applications in other fields in recent years, e.g., industrial material, molecular and drug design [1, 2].
- AI-driven design is not limited to “interfacing” AI with existing advanced simulation platforms used in our community (Geant). It also (and principally) entails establishing a procedural body of instructions to encode efficiently the optimal design requirements and validate the results in a self-consistent way [3].
- As far as optimization is concerned, the choice of a suitable algorithm is a challenge itself (no free lunch theorem [4]) and the full potential of certain algorithms always requires some degree of customization. First thing to do is to study and characterize the properties of the problem.

[1] A. Mosavi, T. Rabczuk, and A. R. Varkonyi-Koczy, “Reviewing the novel machine learning tools for materials design,” in Int. Conference on Global Research and Education, pp. 50–58, Springer, 2017
[2] Z. Zhou, S. Kearnes, L. Li, R. N. Zare, and P. Riley, “Optimization of molecules via deep reinforcement learning,” Scientific Reports, vol. 9, no. 1, pp. 1–10, 2019
[3] CF et al. "AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case." *JINST* 15.05 (2020): P05009.
[4] Wolpert, D.H., Macready, W.G., 1997. No free lunch theorems for optimization. Trans. Evol. Comp 1, 67–82

How do we design and optimize detectors?

- Typically full detector design is studied once the subsystem prototypes are ready.
- In the subsystem design phase constraints from the full detector or outer layers are taken into consideration.
- Actually many parameters (mechanics, geometry, optics) characterize the design of each sub-detector, hence the full design represents a large combinatorial problem. A well known phenomenon observed in optimization problems with high-dimensional spaces is the so-called “curse of dimensionality'” [1], introduced for the first time by Bellman when considering problems in dynamic programming.
- In addition to that, more objective functions often need to be considered at the same time in the design of each sub-detector (e.g., resolution, efficiency, cost, distinguishing power, etc).
- In this context, 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.

"Taking the Human out of the Loop"…?

B. Shahriari, et al. *Proceedings of the IEEE* 104.1 (2015): 148-175.

Human-assisted (experts knowledge)

Bayesian Optimization

DL-enhanced

Multi-objective

Autonomous

Evolutionary autoML

Meta-learning

Reinforcement Learning

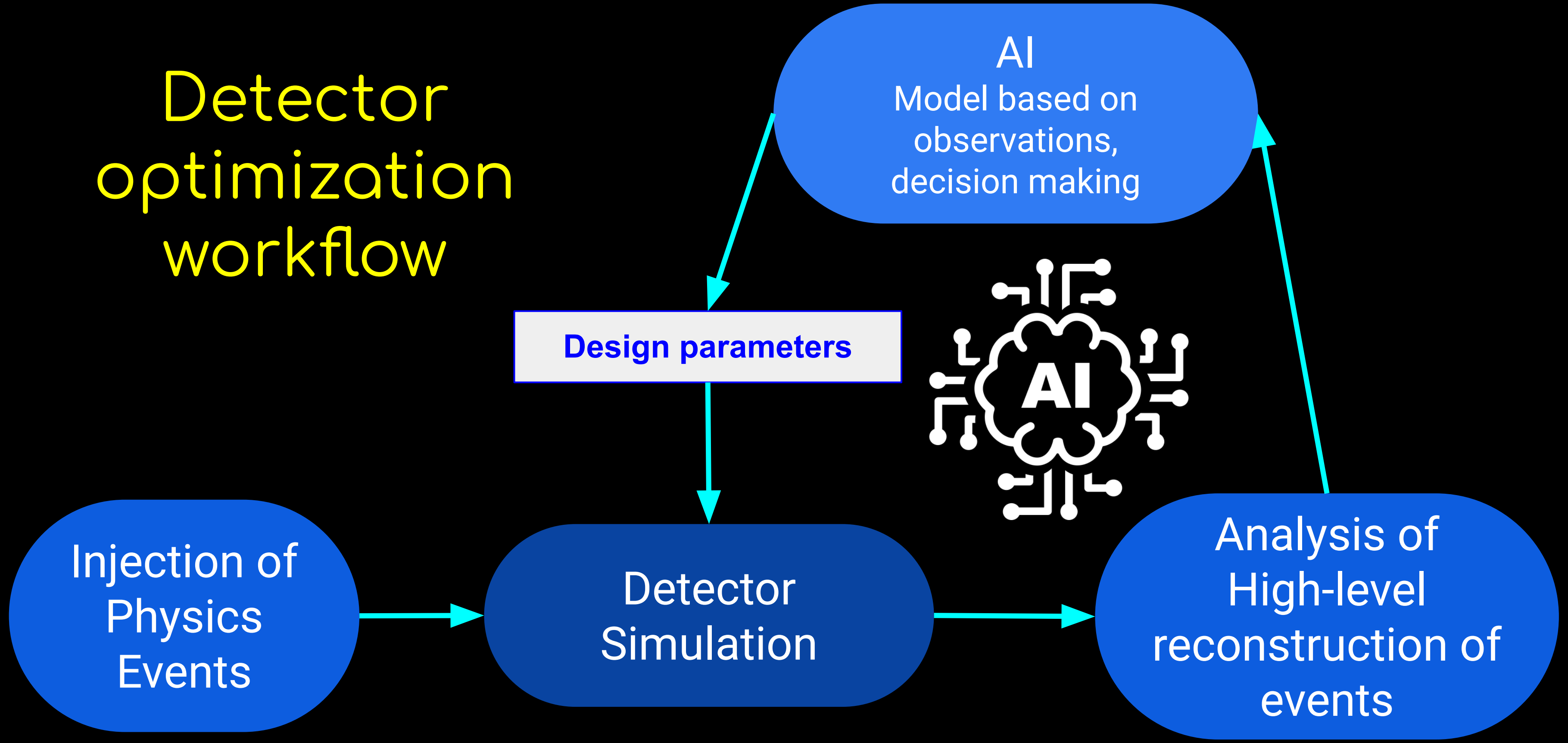
Accelerated Discovery

AI for Detector Design

Intelligent Detectors

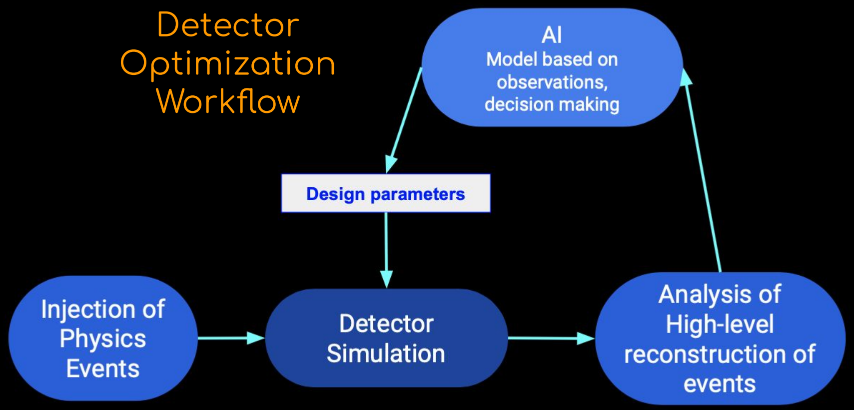
- Optimal Design
- Inverse Design
- Self-Design
- Calibration/ Alignment
- Self-Calibration
- etc.

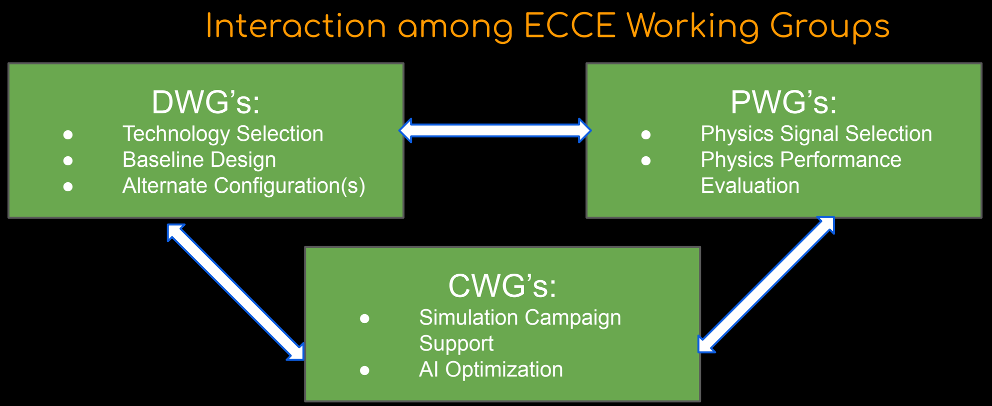
Detector optimization workflow



Why AI at this phase?

- Optimization does not mean necessarily "fine-tuning". In a complex problem with multiple design criteria (e.g., performance, cost, material) it helps identifying/approximate the best set of trade-off solutions (Pareto frontier) and decisions can be made based on that.
- We want to use these algorithms to: (1) steer the design and suggest combinations of 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)
- All "steps" (physics, detector) involved in the AI optimization, strong interplay between ECCE working groups





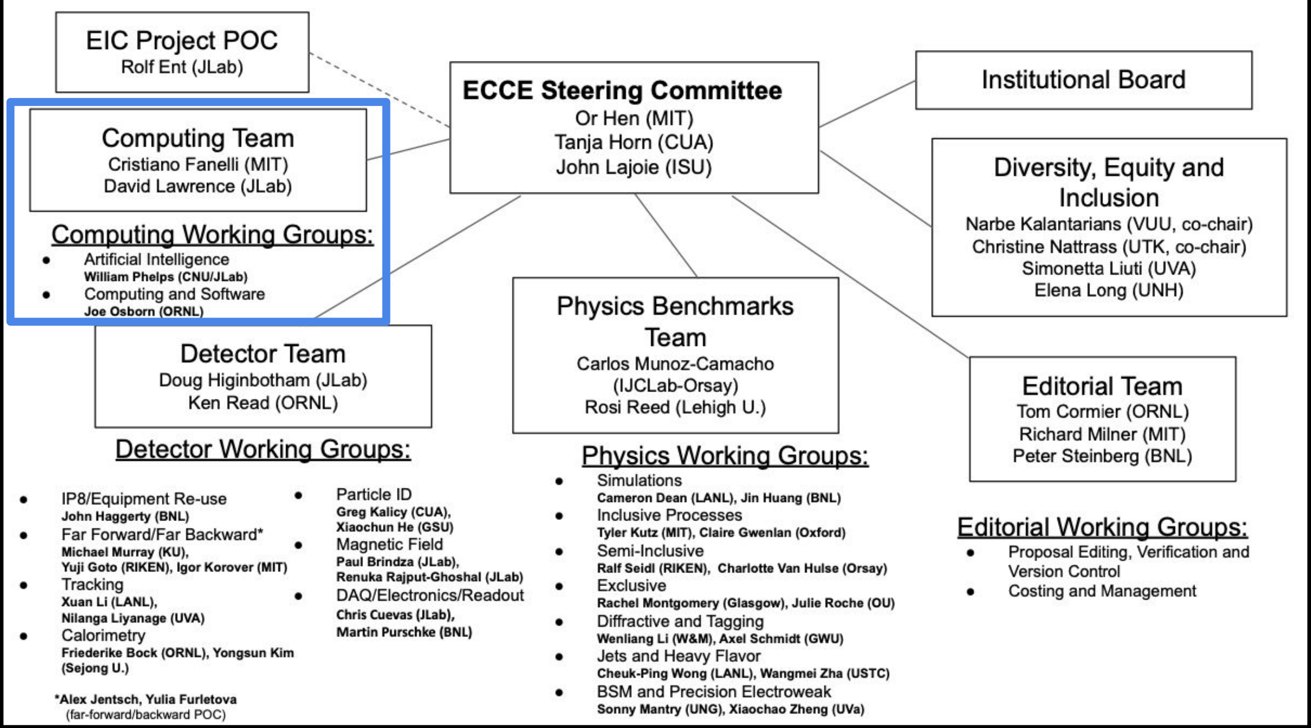
AI within ECCE

EIC Comprehensive Chromodynamics Experiment

https://www.ecce-eic.org

ECCE shares the vision of the NP community that the EIC science mission is best served by two complementary detectors, and is investigating a design based on a 1.5T solenoid in both EIC interaction regions.

ECCE recognizes the important role that AI can play in a future experiment like EIC, and includes in its structure a working group dedicated to AI (March 2021)



AI WG Activities

- Keeping in mind the “inner to outer” design process / strategy from T. Horn’s talk at the 5th ECCE IB:

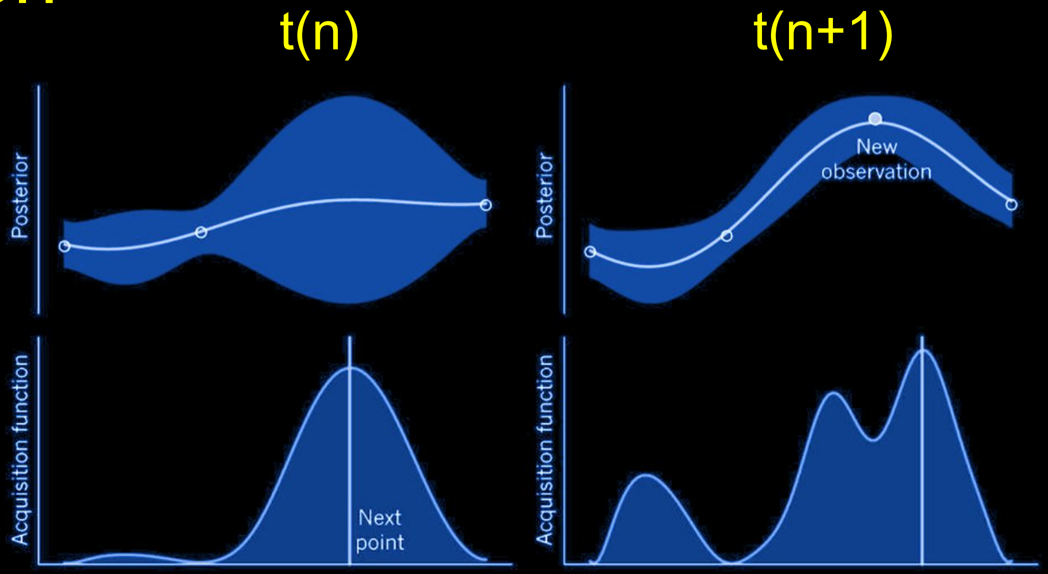
Optimize technology choices together with physics performance

- Formulate a dynamic timeline for decision making for the global simulation
- Start with the design of the inner layers, e.g., fix tracking and PID and then work outwards (radially and in polar angle), e.g., for PID it is important to have knowledge of the magnetic field and the tracking resolution and also minimizing material

- Identified activities in the AI Working Group (regarding design and reconstruction algorithms):
  - Tracking (Brunel, MIT, Regina, work in progress)
  - PID --- DIRC (CNU, MIT), d-RICH (MIT d-RICH paper)
  - Calorimetry (CUA, MIT, Regina, work started within eRD1, link to presentation)
  - (Far Forward --- ZDC (Duquense))

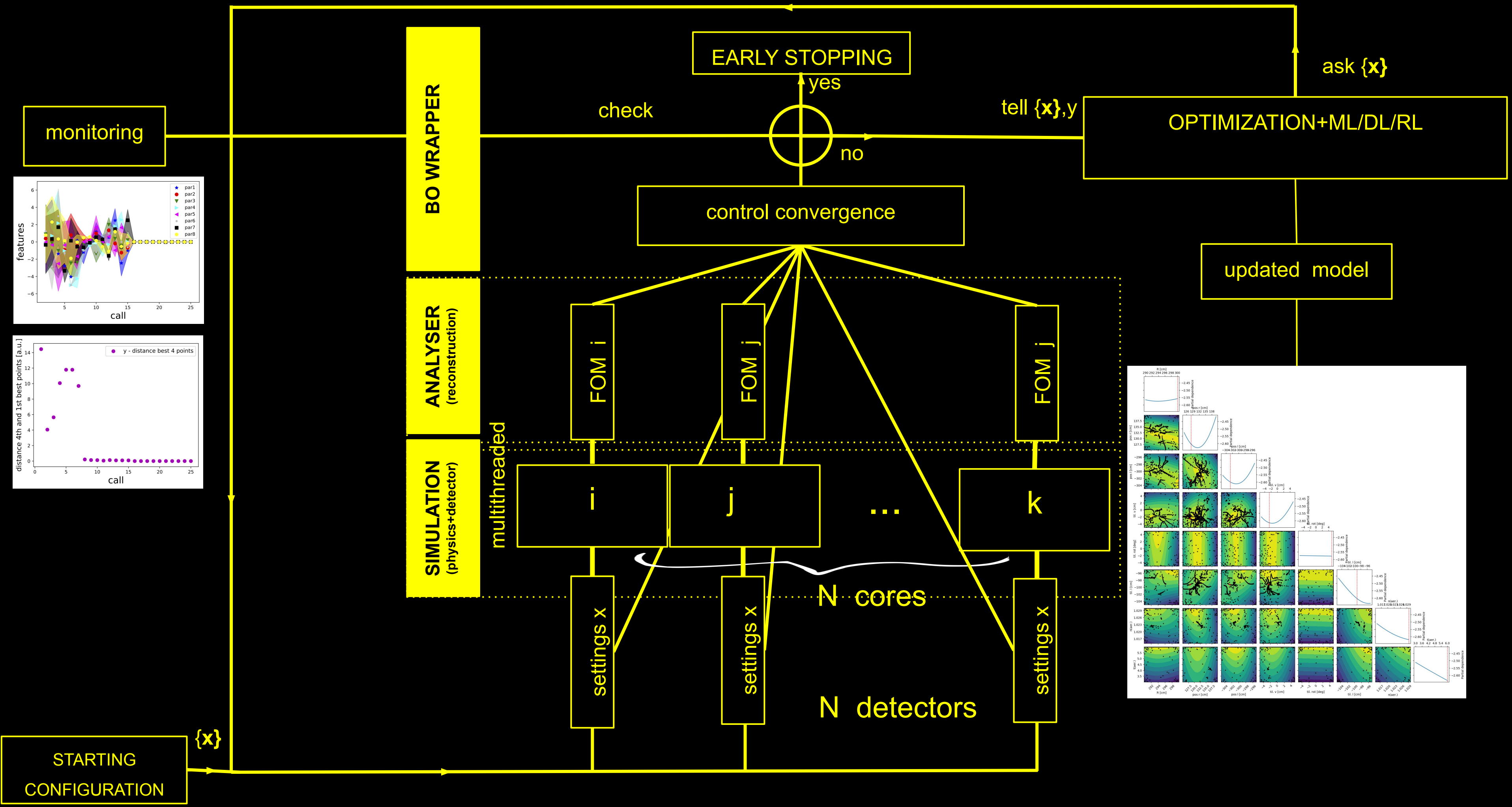
Bayesian Optimization

- 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.**



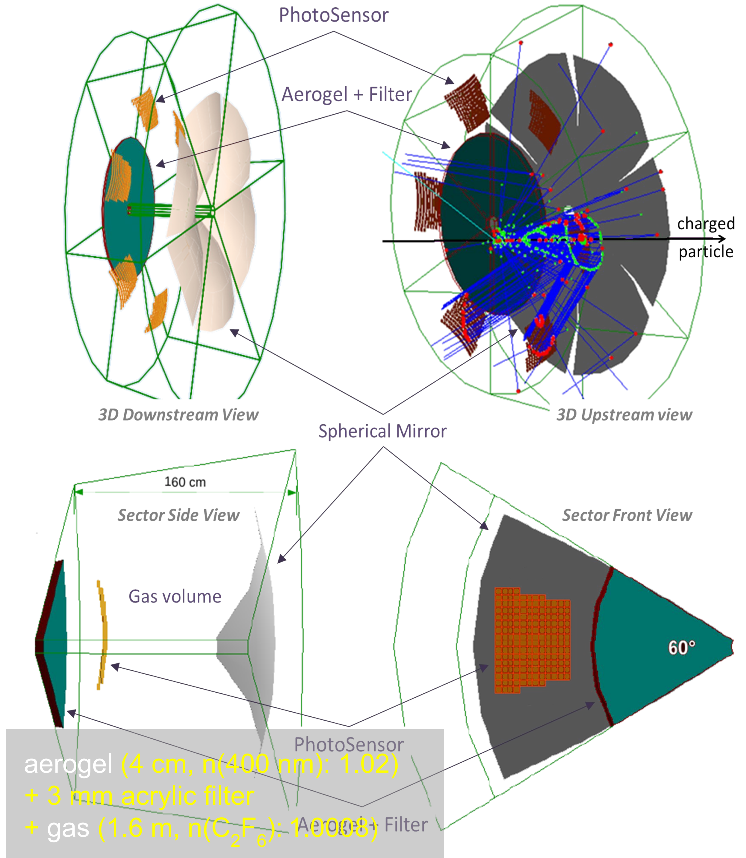
1. Select a Sample by Optimizing the Acquisition Function.
2. Evaluate the Sample With the Objective Function.
3. Update the Data and, in turn, the Surrogate Function.
4. Go To 1.

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

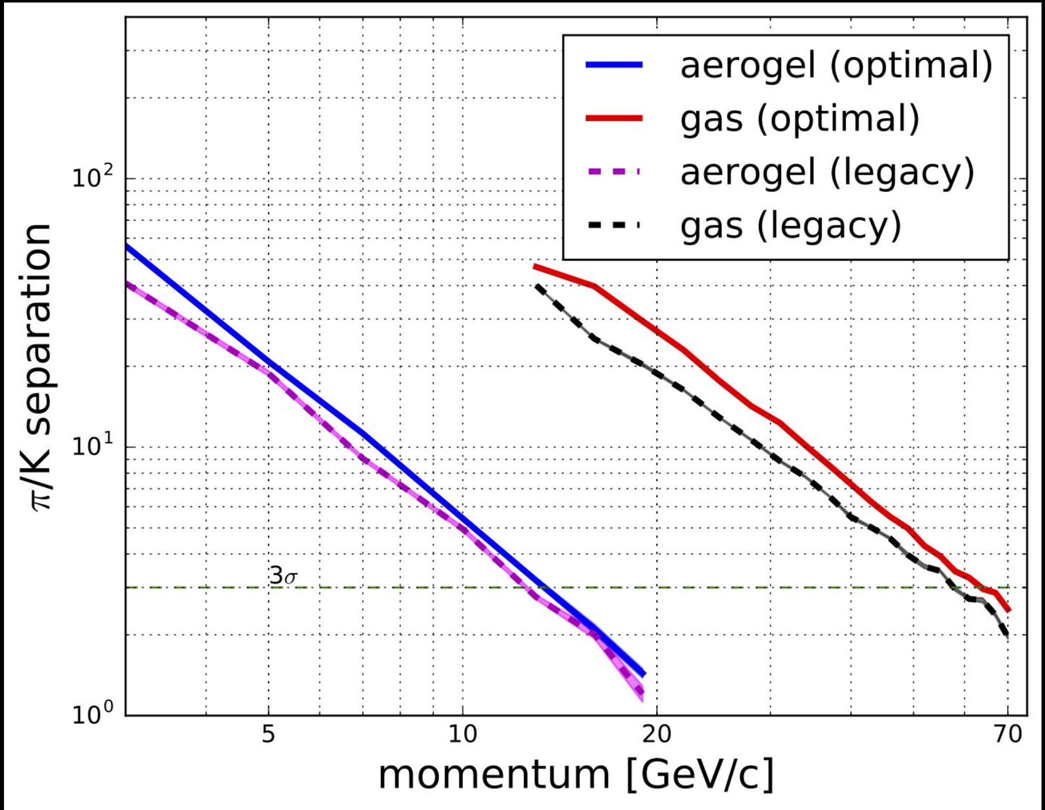

EARLY STOPPING
yes
check
no
BO WRAPPER
tell {x},y
ask {x}
OPTIMIZATION+ML/DL/RL
monitoring
control convergence
updated model
ANALYSER
(reconstruction)
SIMULATION
(physics+detector)
multithreaded
FOM i
FOM j
FOM j
i
j
...
k
N cores
settings x
settings x
settings x
N detectors
STARTING
CONFIGURATION
{x}
features
call
y - distance best 4 points
distance 4th and 1st best points [a.u.]
call

AI-Optimized dRICH

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



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

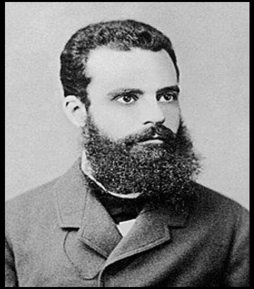


- Statistically significant Improvement in both parts.
- In particular in the gas region where the 5σ threshold shifted from 43 to 50 GeV/c and the 3σ one extended up to
- Notice that before this study we did not know “how well” the legacy design was performing.

Multiple Objectives

[1] Deb, Kalyanmoy. "Multi-objective optimisation using evolutionary algorithms: an introduction." *Multi-objective evolutionary optimisation for product design and manufacturing*. Springer, London, 2011. 3-34.

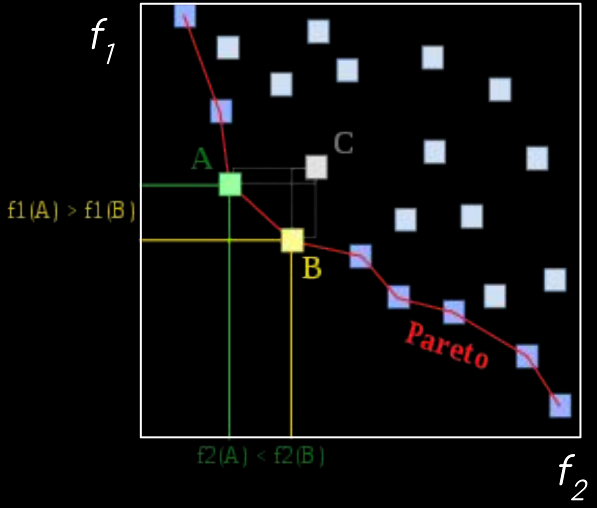
- The problem becomes challenging when the objectives are of conflict to each other, that is, the optimal solution of an objective function is different from that of the other. For example improving the resolution of a detector could imply increasing the costs for its realization.
- In solving such problems, with or without constraints, they give rise to a trade-off optimal solutions, popularly known as Pareto-optimal solutions.



V. Pareto, 1848-1923

- 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.
- Starting with parameterized procedures in early nineties, the so-called evolutionary multi-objective optimization (EMO) algorithms is now an established field of research.

Point *C* is not on the Pareto frontier because it is dominated by both point *A* and point *B*.



MO-based solutions are helping to reveal important hidden knowledge about a problem – a matter which is difficult to achieve otherwise [1].

Frameworks

- Notice that MOO with dynamic/evolutionary algorithms (see, e.g., [1-3]) are probably the most utilized approaches on github, followed by more recent developments on multi-objective bayesian optimization (see, e.g., [4-7]). Using them has the advantage of having an entire community developing those tools.

  https://github.com/topics/multi-objective-optimization

- Agent-based approaches to MOO are also possible (see, e.g., [8]), but won't be discussed here.

- Remarkably these approaches can accommodate mechanical and geometrical constraints during the optimization process.

[1] J. J. Durillo and A. J. Nebro, "jMetal: A Java framework for multi-objective optimization," Advances in Engineering Software, vol. 42, no. 10, pp. 760–771, 2011.

[2] F.-A. Fortin, F.-M. De Rainville, M.-A. G. Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary algorithms made easy," The Journal of Machine Learning Research, vol. 13, no. 1, pp. 2171–2175, 2012.

[3] J. Blank and K. Deb, "pymoo: Multi-objective Optimization in Python," IEEE Access, vol. 8, pp. 89497–89509, 2020

[4] M. Laumanns and J. Ocenasek, "Bayesian optimization algorithms for multi-objective optimization," in International Conference on Parallel Problem Solving from Nature, pp. 298–307, Springer, 2002.

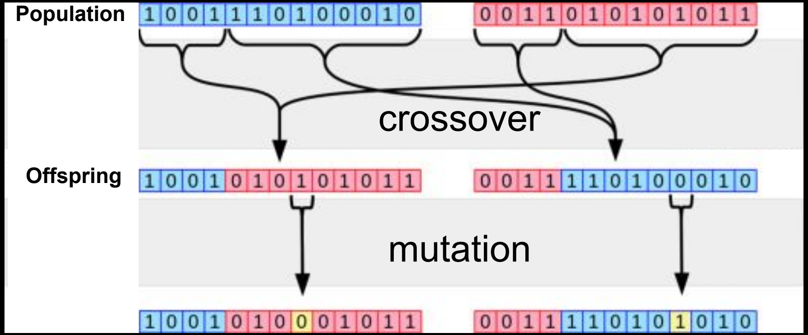
[5] M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson, and E. Bakshy, "Botorch: Programmable bayesian optimization in pytorch," arXiv preprint arXiv:1910.06403, 2019.

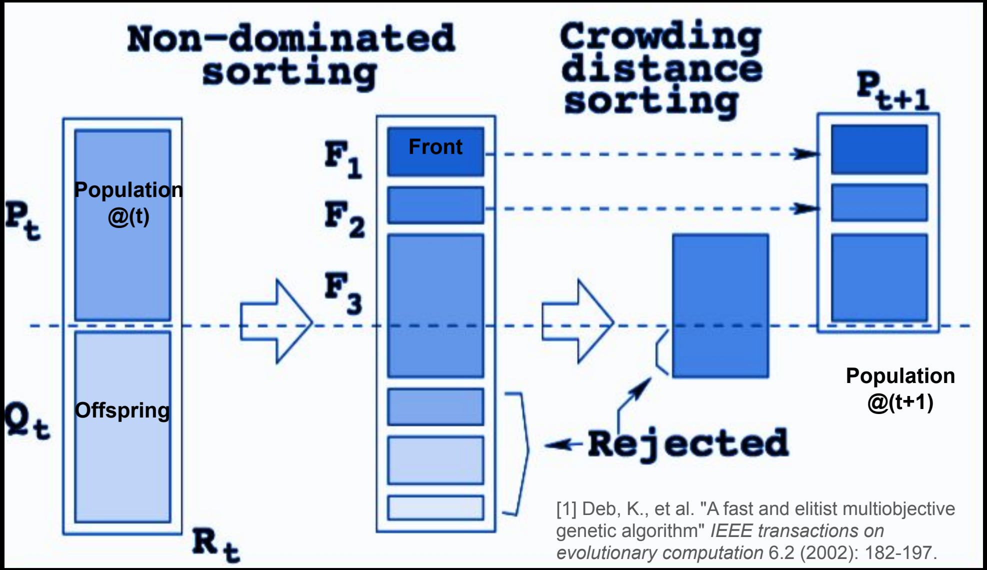
[6] P. P. Galuzio, E. H. de Vasconcelos Segundo, L. dos Santos Coelho, and V. C. Mariani, "MOBOpt—multi-objective Bayesian optimization," SoftwareX, vol. 12, p. 100520, 2020.

[7] A. Mathern, O. S. Steinholtz, A. Sjöberg, M. Önnheim, K. Ek, R. Rempling, E. Gustavsson, and M. Jirstrand, "Multi-objective constrained Bayesian optimization for structural design," Structural and Multidisciplinary Optimization, pp. 1–13, 2020.

[8] R. Yang, X. Sun, and K. Narasimhan, "A generalized algorithm for multi-objective reinforcement learning and policy adaptation," in Advances in Neural Information Processing Systems, pp. 14636–14647, 2019

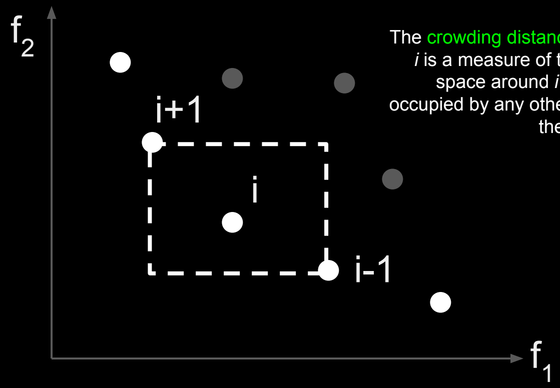
Elitist Non-Dominated Sorting Genetic





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

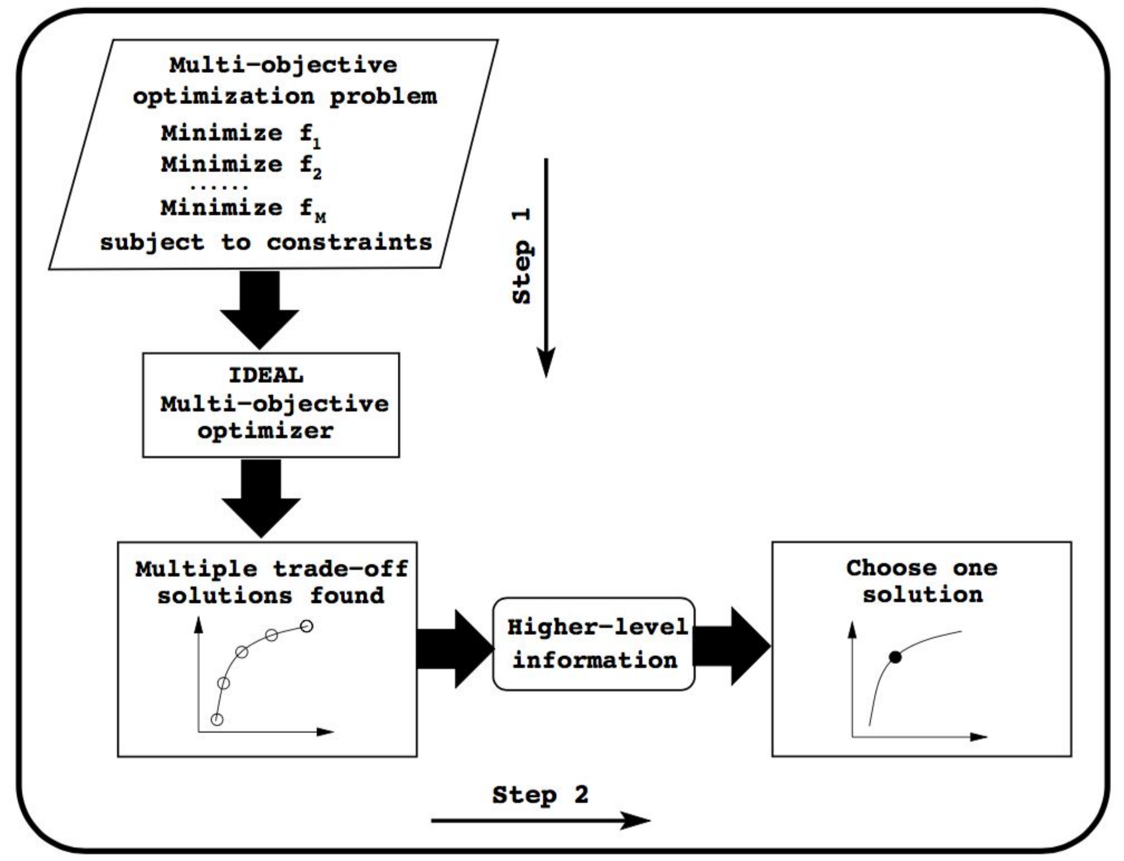


The crowding distance d_i of point *i* is a measure of the objective space around *i* which is not occupied by any other solution in the population.

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.

Workflow

[1] K. Deb, *Multi-objective evolutionary optimisation for product design and manufacturing*. Springer, London, 2011. 3-34.



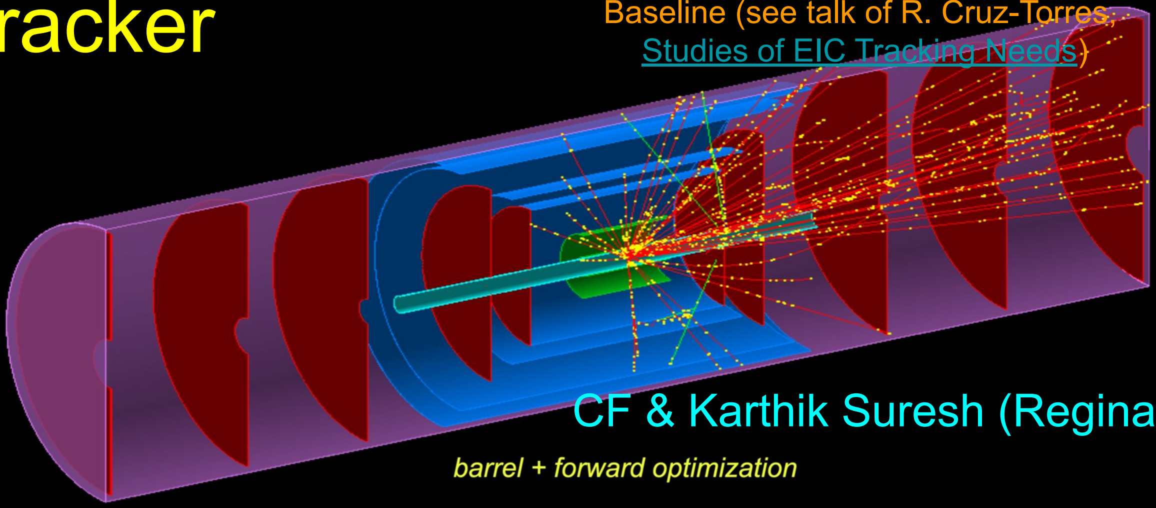
Step 1: Find multiple non-dominated points as close to the Pareto-optimal front as possible, with a wide trade-off among objectives.

Step 2 Choose one of the obtained points using higher-level information.

N.b.: This scheme is particularly useful also for single-objective optimization when multiple global optima are present

The ECCE Inner Tracker

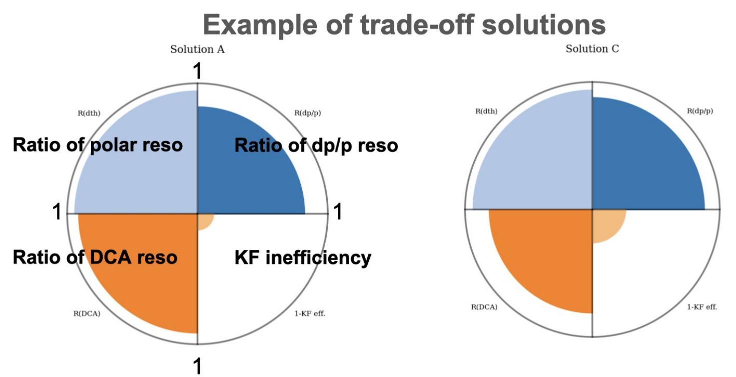
Baseline (see talk of R. Cruz-Torres; Studies of EIC Tracking Needs)



- Extended the design criteria to include simultaneously Kalman filter efficiency, pointing resolution, along with momentum and angular resolutions.
- Mechanical constraints

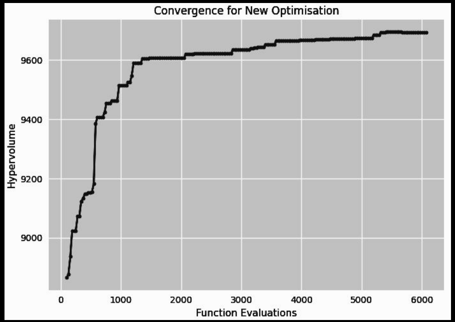
CF & Karthik Suresh (Regina)

barrel + forward optimization



11 parameters
4 objectives
Population size 100
Offspring distributed over 30 cores

Each proposed design is consistent with baseline Aluminum support shell (not displayed)



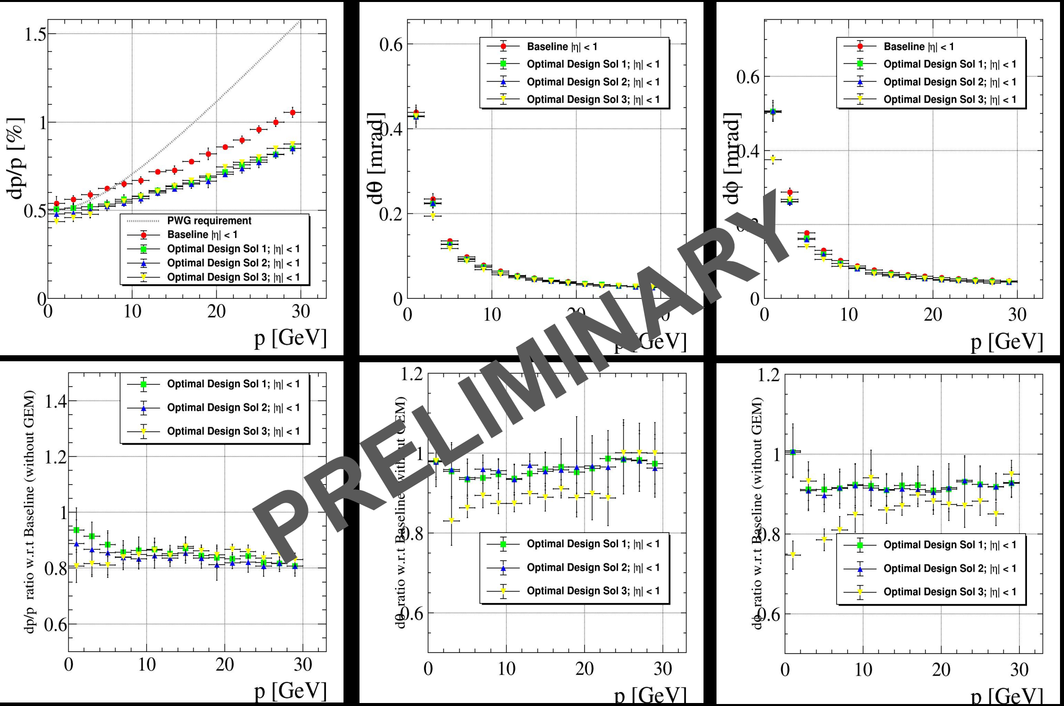
Ratios are with respect the LBNL all Si baseline

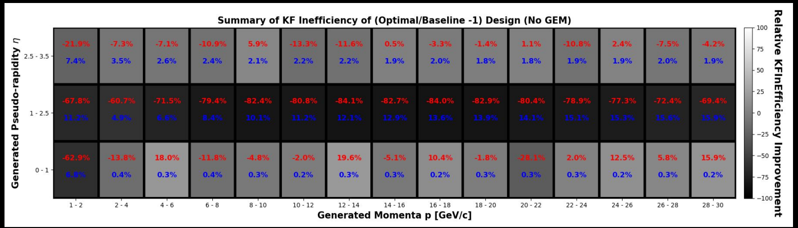
This is an unprecedented attempt in detector design for complexity

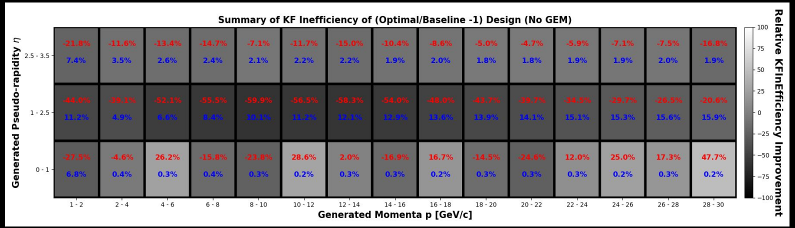
Inner Tracker

For each design solution in the front one can study the corresponding detector performance.

- The decision making process on the design can happen after the optimization, exploring the performance of the trade-off solutions.
- On left are displayed momentum, angular resolutions) for one solution. Below the Kalman Filter inefficiency.
- Performance not included as objectives can be used for validation. For example, pattern recognition and fake tracks rejection studies eventually studied to validate designs.



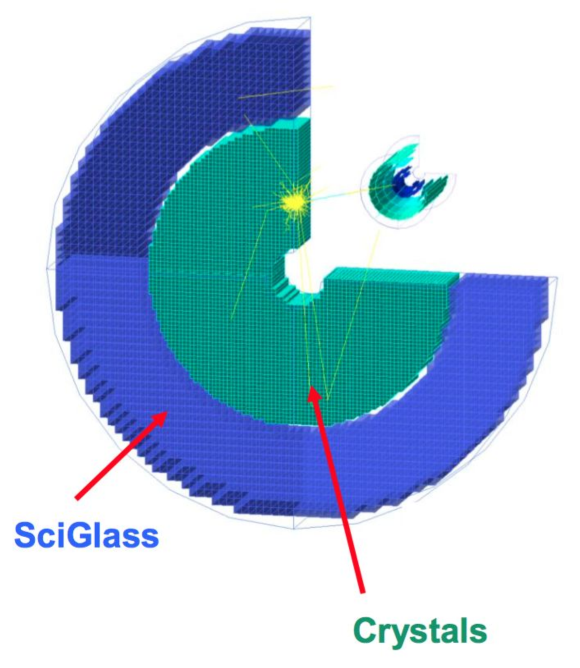




EIC Electron Endcap EMCal

The team: V. Berdnikov, M. Bondi’, CF, Y. Furletova, T. Horn, I. Larin, D. Romanov, R. Trotta

- We can use Multi-objective Optimization to optimize glass/crystal material selection in shared rapidity regions including mechanical constraints.
- Like in the Hall B project, we can explore implementation of AI for clustering/reconstruction.



EIC Electron Endcap require an inner part (crystal) with high resolution and an outer part (glass) with less stringent requirements

Crystals have been used in homogeneous calorimeters but their production is slow and expensive.

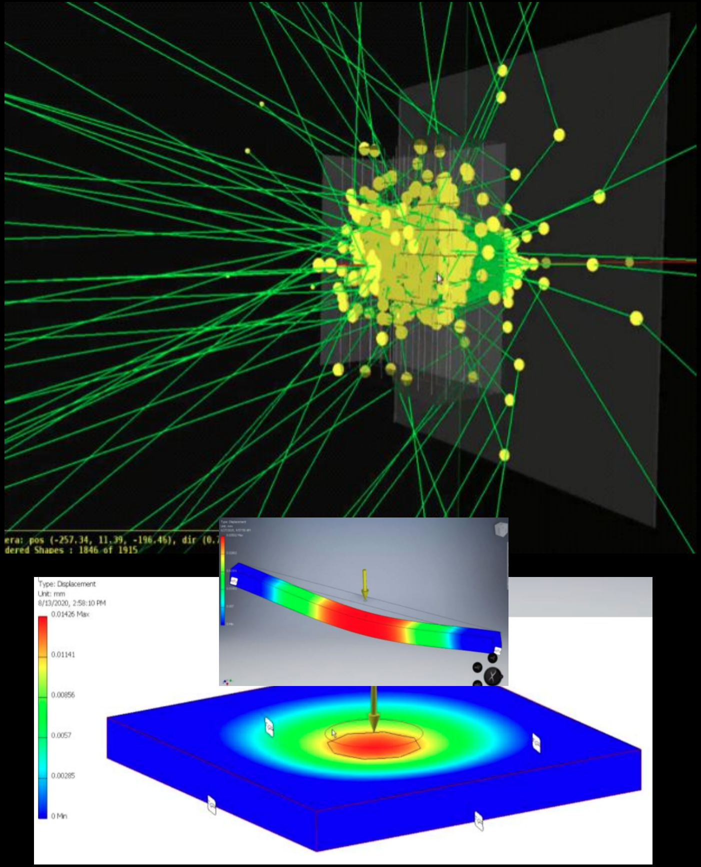
As an alternative Scintilex develops SciGlass that is much simpler and less expensive to produce and thus offers great potential for both cost reduction and wider application if competitive performance parameters can be achieved.

Goal: maintain the resolution needed by the physics processes while reducing the number of crystals/cost, taking into account constraints.

Novel aerogel material

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 is introducing fibers into the aerogel that increase mechanical strength, but do not affect the optical properties.
- We are designing the aerogel+fibers optimizing mechanical stability and resolution.
- Paper in preparation.



AI4EIC

[1] B. Bedaque, et al., Report from the AI For Nuclear Physics Workshop, arXiv:2006.05422, 2020
[2] Joint Machine Learning Workshop, GlueX Panda EIC, 2020
[3] AI4NP Winter School, https://indico.jlab.org/event/409/, 2021

- Strategic moment to discuss how to fully take advantage of the new opportunities offered by AI to advance research, design, and operation of EIC.
- Growing convergence of AI, Data, and HPC provides a once in a generation opportunity to profoundly accelerate scientific discovery, create synergies across scientific areas.
- The interest of the community evidenced by the number of contributions and attendance of workshops dedicated to AI in Nuclear Physics, e.g. the [1, 2, 3],
- The AI4EIC workshop will bring together the communities directly using AI technologies and provide a venue for discussion and identifying the specific needs and priorities for EIC.
- This will be a series of workshops. The first one will have a focus on experimental applications, therefore AI4EIC-exp

https://indico.bnl.gov/event/10699/

Experimental Design,
Simulations,
Reconstruction / Analysis,
Control of Experimental Systems,
Detector Readout,
Computing Frontiers

Summary

- AI is becoming ubiquitous in NP, and remarkable accomplishments have been recently obtained.
- AI will likely play a major role in multiple aspects of the Electron Ion Collider experiment. We will have a dedicated workshop on AI4EIC on September 7-10 2021.
- AI is at present already contributing to the EIC design.
- In NP we started exploring AI for optimal design in multidimensional space with single objectives. Most of the problems are multi-objectives though. None ever accomplished a multi-dimensional / multi-objective optimization of the performance of detectors when operating together. This is a high-dimensional combinatorial problem (with many parameters) that can be solved with AI.
- 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*".