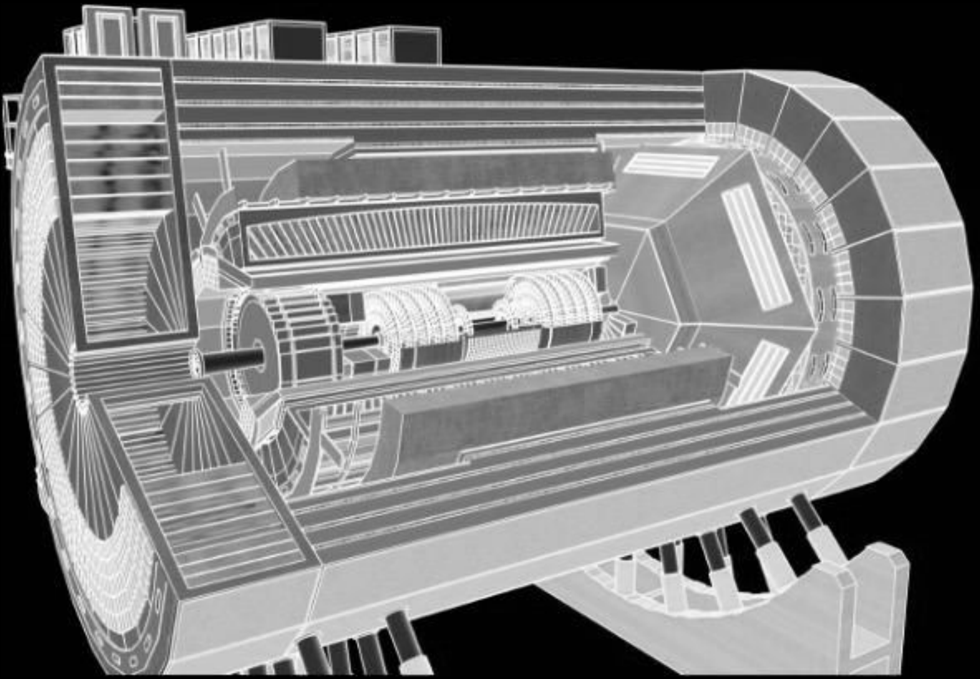
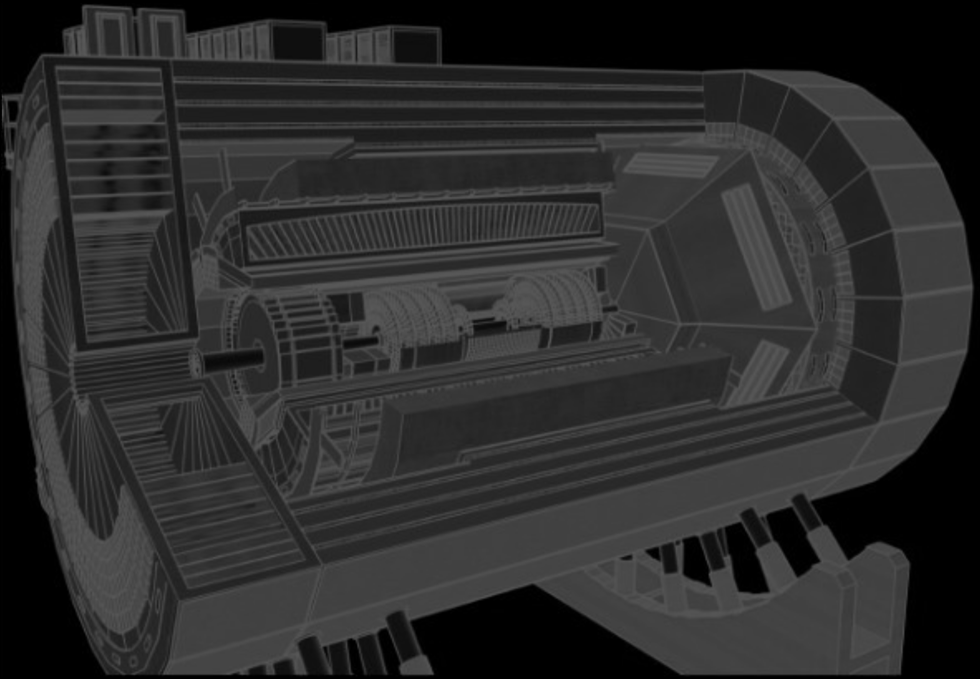
Detector Design with AI @EIC



Cristiano Fanelli

Detector Design with AI

What follows is largely based on a series of lectures I gave on Detector Design with AI at the AI4NP Winter School



Detector Design with AI

What follows is largely based on a series of lectures I gave on Detector Design with AI at the AI4NP Winter School

- New area of research at its infancy. Many applications in, e.g., industrial material, molecular and drug design [1, 2].
- 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).
- But actually many parameters and multiple objective functions: curse of dimensionality [3].
- Entails establishing a procedural body of instructions [4].
- The choice of a suitable algorithm is a challenge itself (no free lunch theorem [5]) and always requires some degree of customization.

[1] A. Mosavi, T. Rabczuk, and A. R. Varkonyi-Koczy, 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, Scientific Reports, vol. 9, no. 1, pp. 1–10, 2019

[3] Bellman, Richard. *Dynamic programming*. Vol. 295. RAND CORP SANTA MONICA CA, 1956.

[4] CF et al. *JINST* 15.05 (2020): P05009.

[5] Wolpert, D.H., Macready, W.G., 1997. Trans. Evol. Comp 1, 67–82

Detector Design with AI

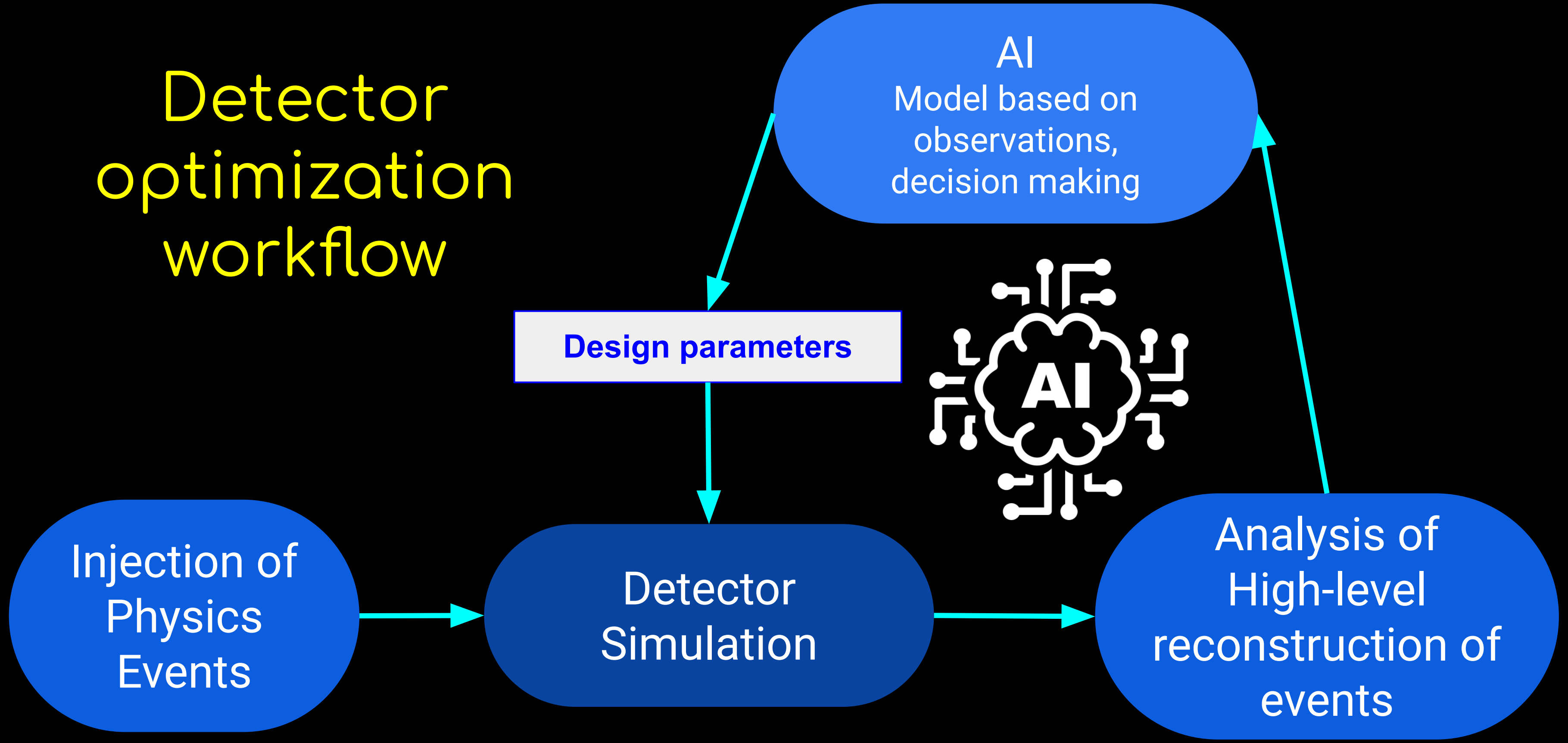
What follows is largely based on a series of lectures I gave on Detector Design with AI at the AI4NP Winter School

- New area of research at its infancy. Many applications in, e.g., industrial material, molecular and drug design [1, 2].
- 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).
- But actually many parameters and multiple objective functions: curse of dimensionality [3].
- Entails establishing a procedural body of instructions [4].
- The choice of a suitable algorithm is a challenge itself (no free lunch theorem [5]) and always requires some degree of customization.

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

[1] A. Mosavi, T. Rabczuk, and A. R. Varkonyi-Koczy, 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, Scientific Reports, vol. 9, no. 1, pp. 1–10, 2019
[3] Bellman, Richard. *Dynamic programming*. Vol. 295. RAND CORP SANTA MONICA CA, 1956.
[4] CF et al. *JINST* 15.05 (2020): P05009.
[5] Wolpert, D.H., Macready, W.G., 1997. Trans. Evol. Comp 1, 67–82

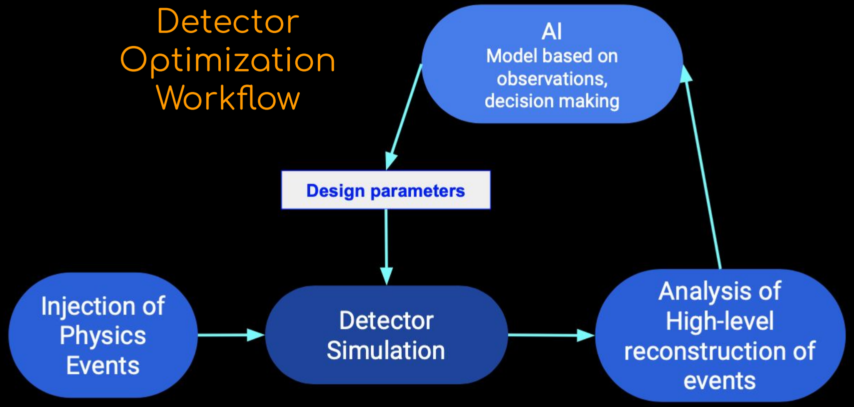
Detector optimization workflow



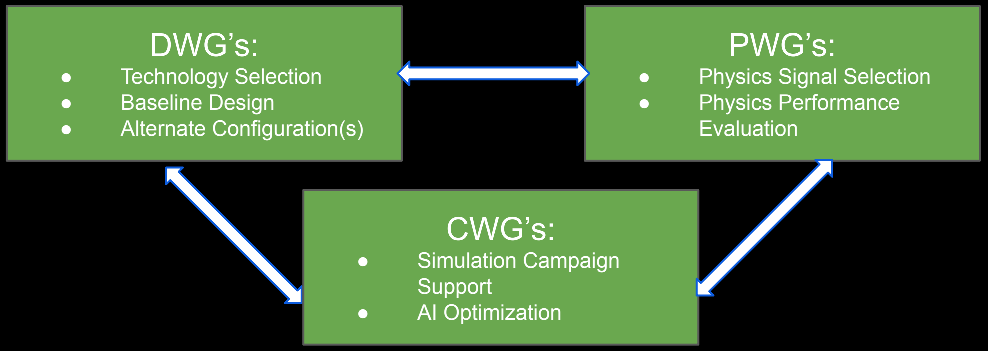
Why Design with AI now?

- Optimization does not mean necessarily “fine-tuning”.
- 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 <u>d-RICH paper</u>, e.g., optics properties)
- All “steps” (physics, detector) involved in the AI optimization, strong interplay between working groups

Detector Optimization Workflow



AI promotes Interaction among Working Groups



AI/ML in EIC Detector Design

CORE a more traditional approach, **ATHENA** and **ECCE** working in the AI supported design

ATHENA AI/ML [from Software & Computing Meeting, as of 7/28/21]

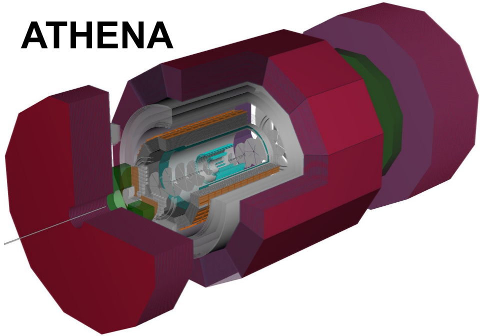
- Current use of AI/ML
  - ✅ e/π PID with 3D shower profiles from imaging calorimeter in center barrel region.
- Near-term anticipated use:
  - ACTS: Track finding (& evaluate ML expertise)
  - PID: Pattern recognition in RICH, DIRC0
  - ✅ Calorimetry clustering (2D, 2+1D and 3D clustering)
  - 🚧 DNN-based fast simulation
  - 🚧 DNN-based detector optimization (Bayesian Optimization)
  - 🚧 DNN-based reconstruction
- Implications on computing infrastructure:
  - 🚧 May exascale GPU accelerators, but lack of support in current software tools limited by IO/memory bandwidth

✅ = working, 🚧 = in progress

ECCE AI/ML [provided to J. Lauret, AI/ML in NP @ BNL, 7/6/21]

- Detector Design with AI/ML (currently on slurm)
  - ✅ Evolutionary approach with multiple objectives (Kalman Filter, momentum, angular and pointing resolutions simultaneously) for the design of the tracker (Brunel, MIT, Regina)
  - ✅ ML/DL + Bayesian optimization for DIRC design (CNU, MIT)
  - ✅ Calorimetry (MIT, Regina)
  - dual-RICH (MIT paper) and ZDC
- Near-term anticipated use:
  - ✅ Calorimetry clustering (4D clustering originally developed for streaming RO)
  - Tracking: track finding and ghost
  - PID: Pattern recognition, e.g., for imaging Cherenkov detectors
- AI/ML/DL in the ECCE computing plan:
  - Online, Offline, HPC, streaming, autonomous control

ATHENA

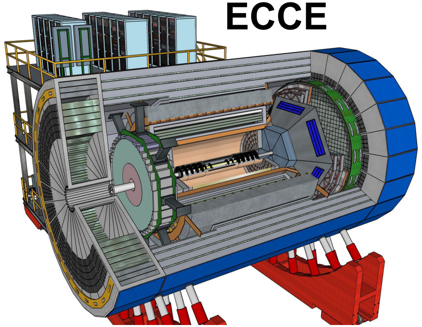


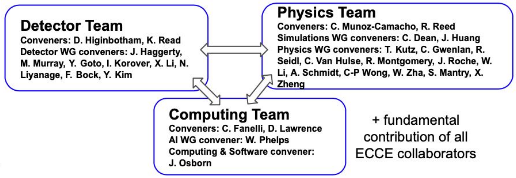
Full simulation/reconstruction team
Whitney Armstrong, Miguel Arratia, Wouter Deconinck, Sylvester Joosten, Jihee Kim, Chao Peng, Tomas Polakovic, Dmitry Romanov, Marshall Scott, Zhenyu Ye, Ziyue Zhang, Maria Żurek

...and a rapidly growing amount ATHENA collaborators!

Software & Computing Conveners:
Whitney Armstrong, Andrea Bressan, Wouter Deconinck, Sylvester Joosten, Dmitry Romanov

ECCE





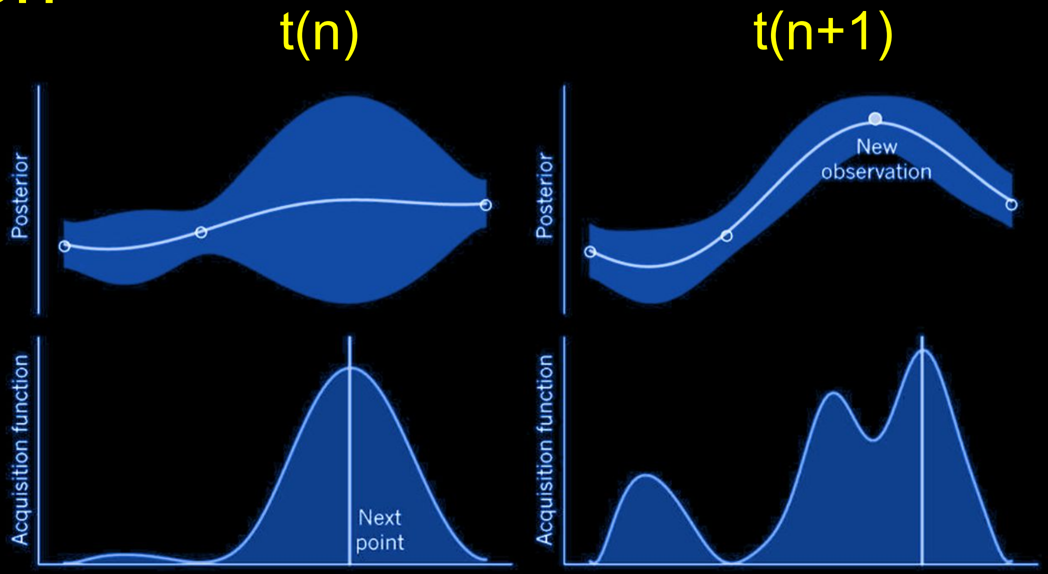
+ fundamental contribution of all ECCE collaborators

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 (since March 2021).

Cristiano Fanelli (MIT) and William Phelps (CNU/JLab) leading AI Working Group

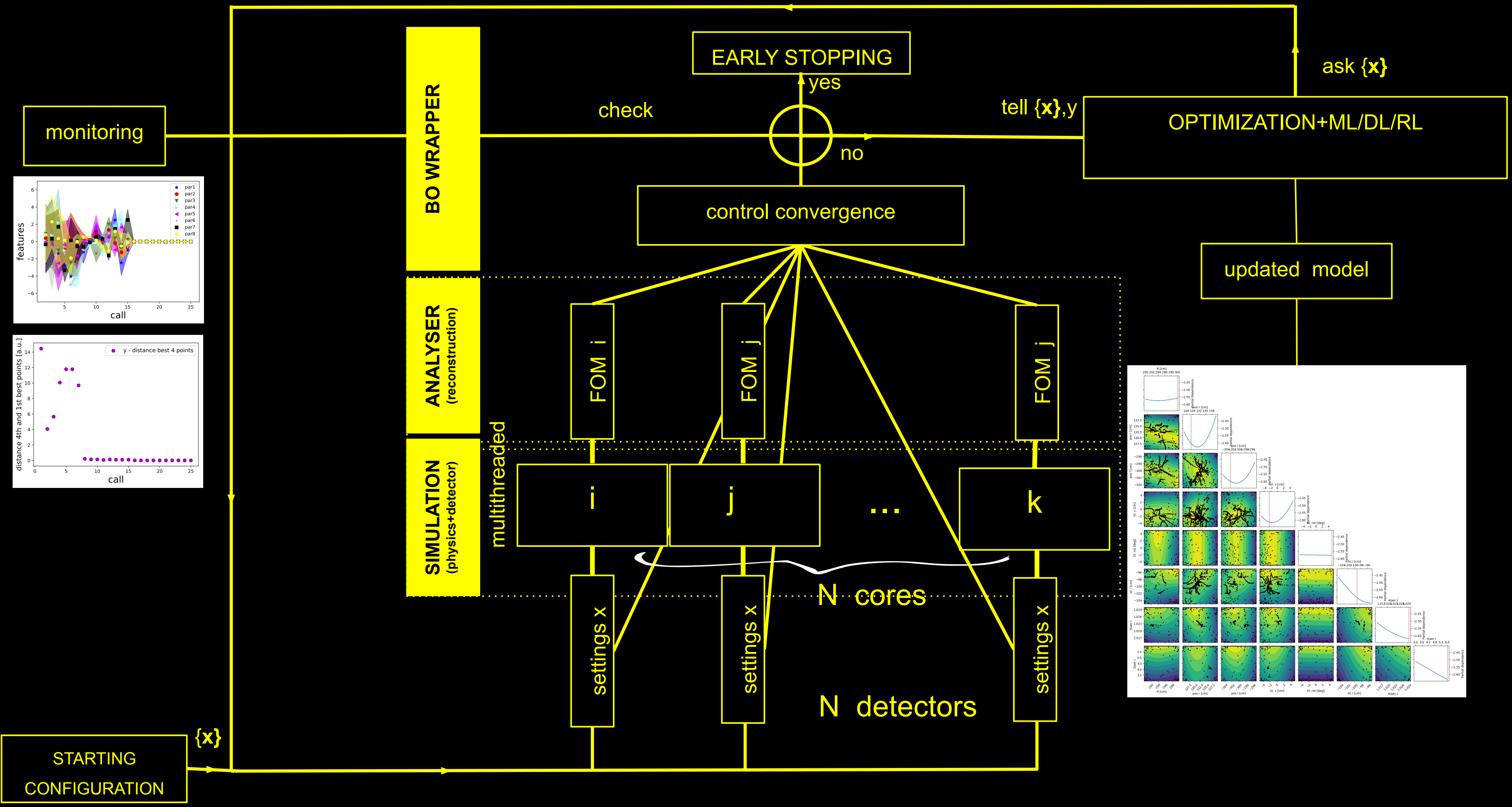
Bayesian Optimization

- 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.**



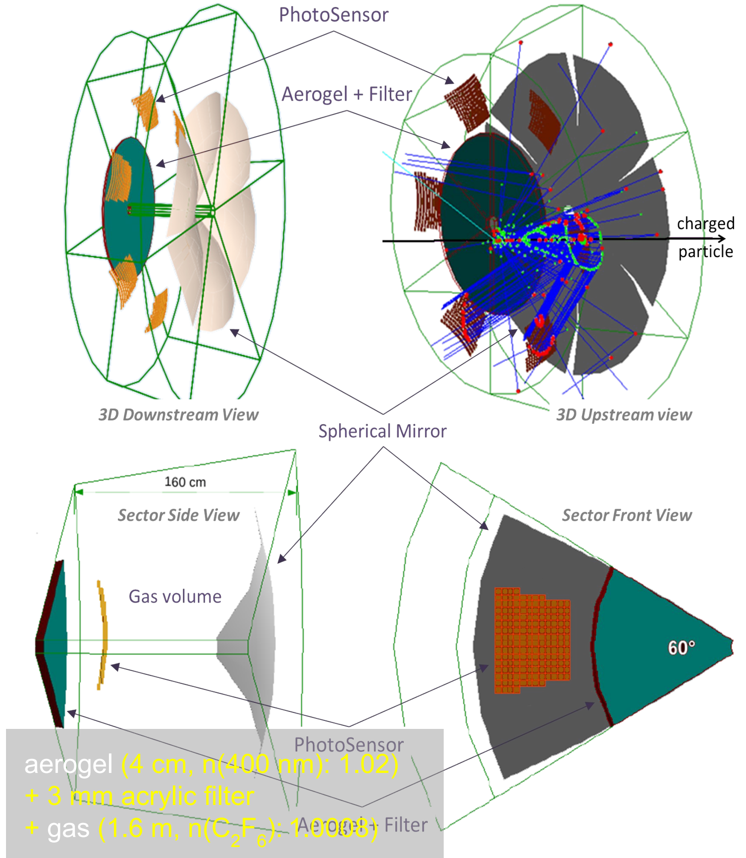
1. Select a Sample by Optimizing the Acquisition Function.
2. Evaluate the Sample With the Objective Function.
3. Update the Data and, in turn, the Surrogate Function.
4. Go To 1.

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

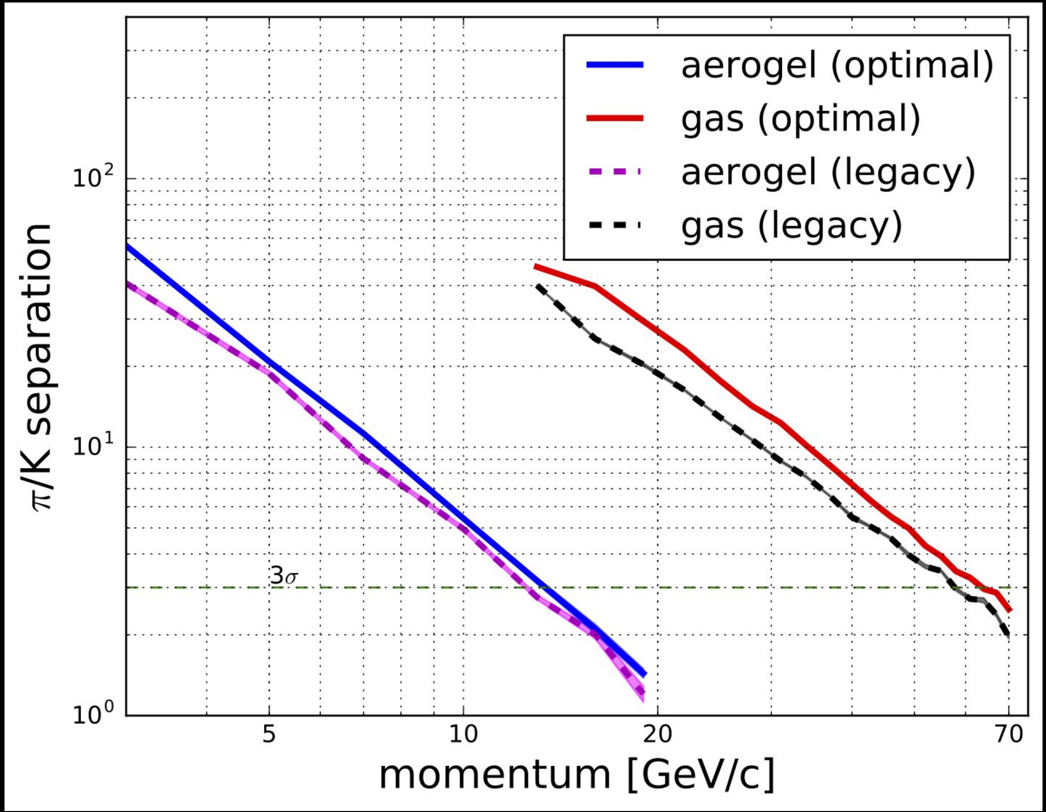

EARLY STOPPING
yes
check
no
tell {x},y
ask {x}
OPTIMIZATION+ML/DL/RL
monitoring
BO WRAPPER
control convergence
updated model
ANALYSER
(reconstruction)
FOM i
FOM j
FOM j
SIMULATION
(physics+detector)
multithreaded
i
j
...
k
N cores
settings x
settings x
settings x
N detectors
{x}
STARTING
CONFIGURATION
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)

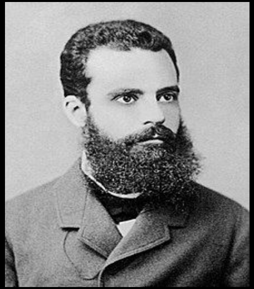


- Statistically significant Improvement in both parts.
- In particular in the gas region where the 5σ threshold shifted from 43 to 50 GeV/c and the 3σ one extended up to
- Notice that before this study we did not know "how well" the legacy design was performing.

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.

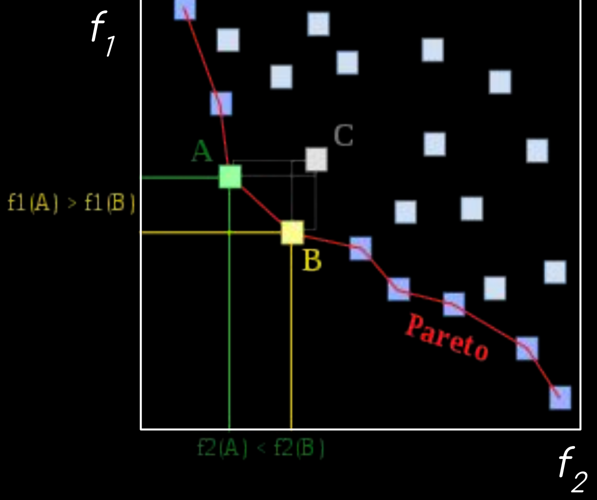
- The problem becomes challenging when the objectives are of conflict to each other, that is, the optimal solution of an objective function is different from that of the other. For example improving the resolution of a detector could imply increasing the costs for its realization.
- In solving such problems, 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.

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, 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.

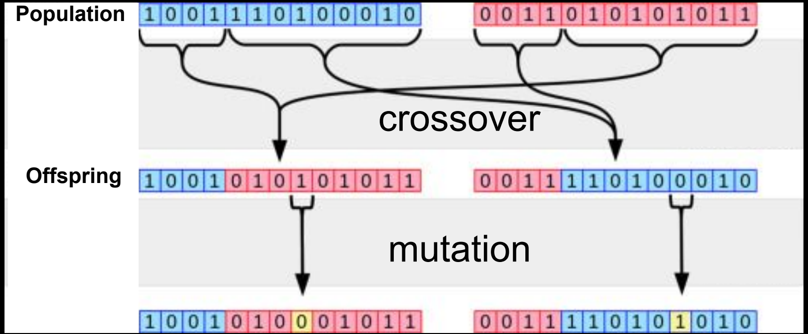
[8] R. Yang, X. Sun, and K. Narasimhan, "A generalized algorithm for multi-objective reinforcement learning and policy adaptation," in Advances in Neural Information Processing Systems, pp. 14636–14647, 2019

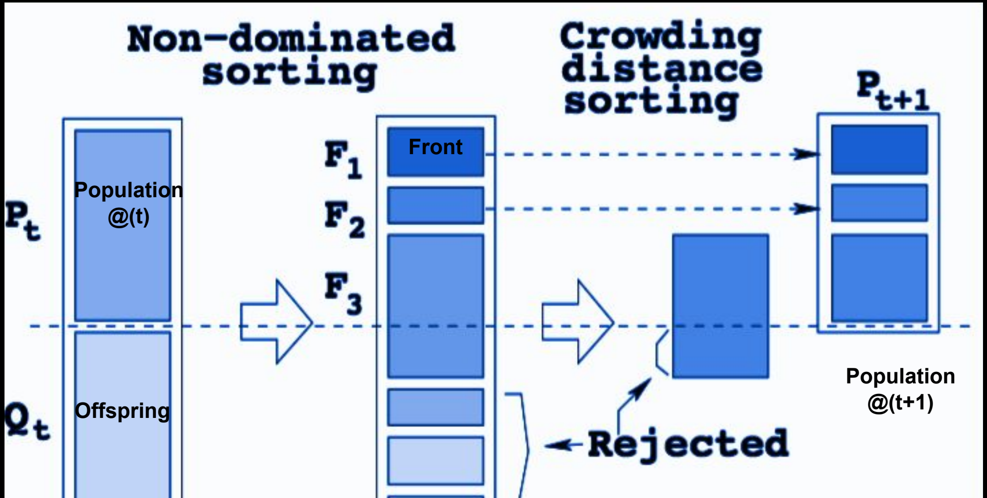
BoTorch

pymoo

DEAP

Elitist Non-Dominated Sorting Genetic

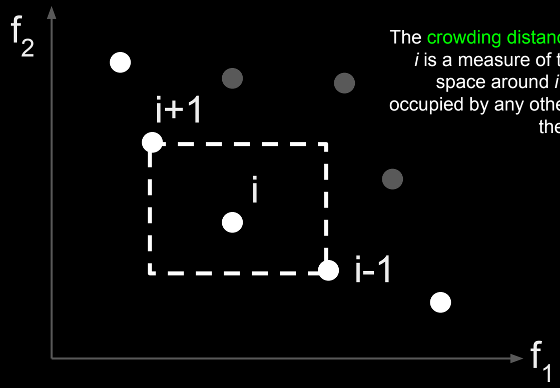




[1] Deb, K., et al. "A fast and elitist multiobjective genetic algorithm" *IEEE transactions on evolutionary computation* 6.2 (2002): 182-197.

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

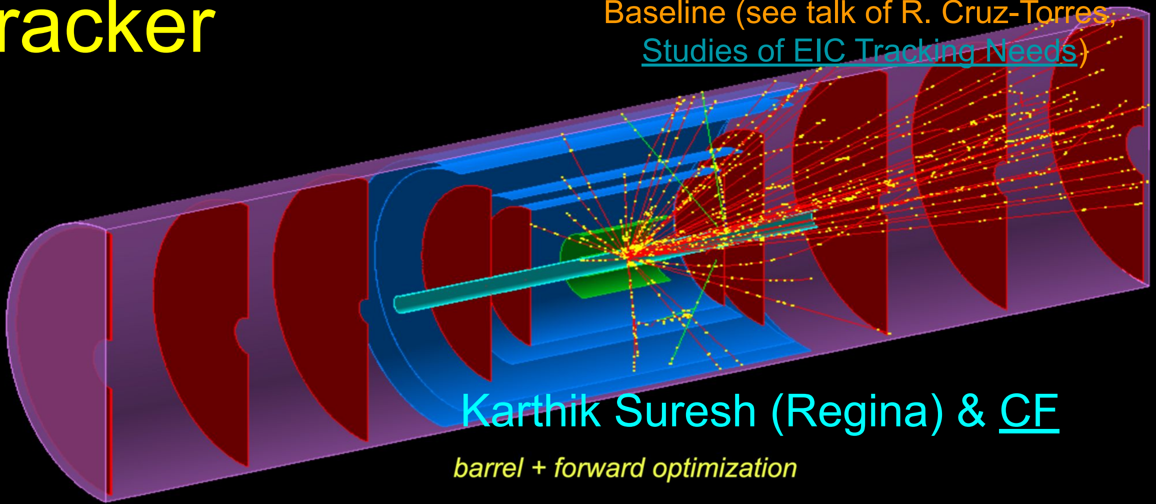


The crowding distance d_i of point *i* is a measure of the objective space around *i* which is not occupied by any other solution in the population.

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.

The ECCE Inner Tracker

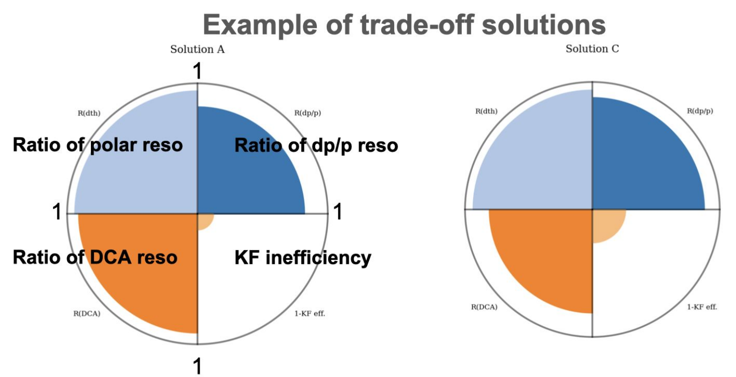
Baseline (see talk of R. Cruz-Torres; Studies of EIC Tracking Needs)



Karthik Suresh (Regina) & CF

barrel + forward optimization

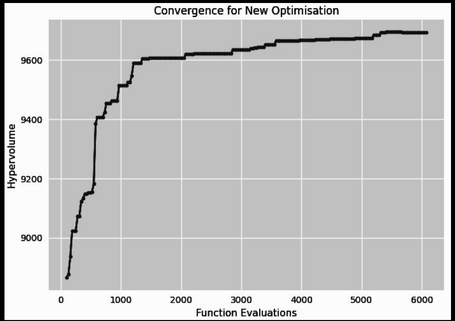
- Extended the design criteria to include simultaneously Kalman filter efficiency, pointing resolution, along with momentum and angular resolutions.
- Mechanical constraints



Ratios are with respect the LBNL all Si baseline

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

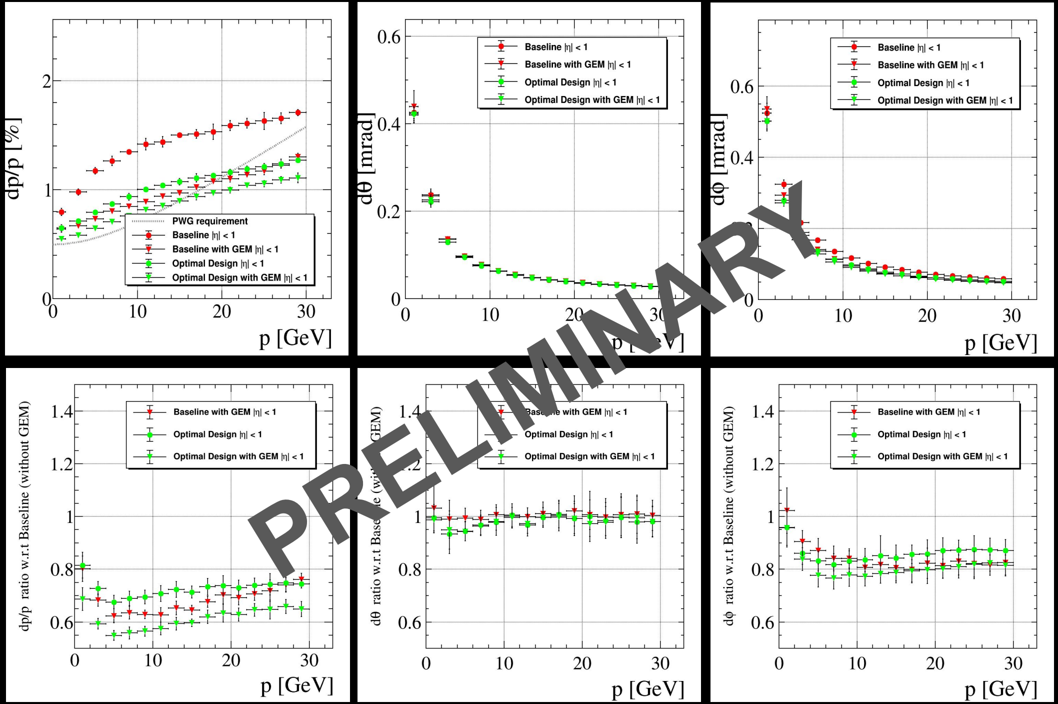
Each proposed design is consistent with baseline Aluminum support shell (not displayed)



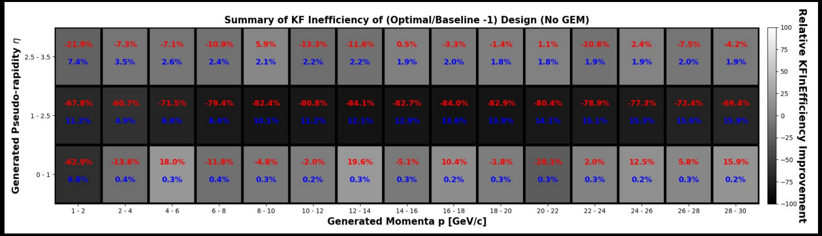
This is an unprecedented attempt in detector design for complexity!

The ECCE Inner Tracker

The decision making process done after optimization.
For each design solution in the Pareto Front one can study the corresponding detector performance.
Shown here tracking with the inner tracker points!



- 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.



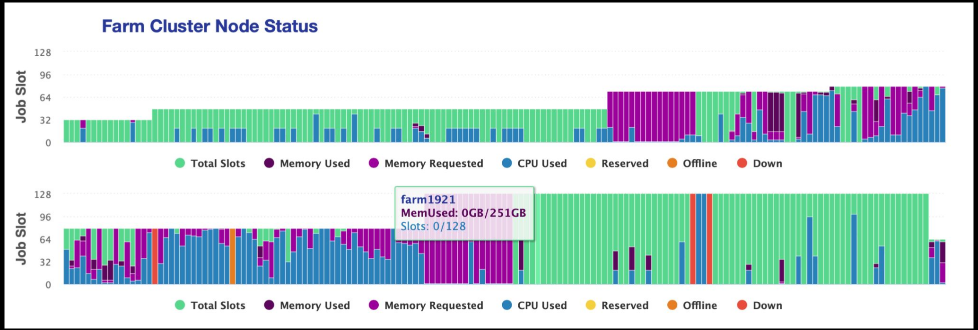
Computing Resources

[1] Kung HT, Luccio F, Preparata FP. On finding the maxima of a set of vectors. Journal of the Association for Computing Machinery. 1975;22(4):469–476
[2] Jensen, Mikkel T. "Reducing the run-time complexity of multiobjective EAs: The NSGA-II and other algorithms." *IEEE Transactions on Evolutionary Computation* 7.5 (2003): 503-515.
[3] Liu, Xin, et al. "Parallelization and optimization of NSGA-II on sunway TaihuLight system." *IEEE Transactions on Parallel and Distributed Systems* 32.4 (2020): 975-987.

- Stress testing + simulations currently running on scicomp @ JLab *

* The scientific computing cluster has:

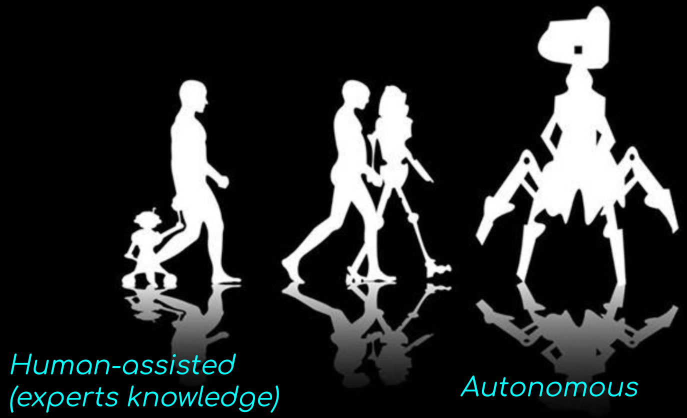
- 25k cores EIC Projects are allocated 10%
- 1PB for EIC use
- Batch use as well as interactive use supported with
    - Nodes with up to two 32 core AMD Epyc Processors (128 threads), 256GB Ram, 1TB SSD local storage
    - 3 Nodes with 4 Titan RTX Cards (24 GB Memory)
    - GPU nodes also available through jupyterhub.jlab.org



The "Farm" status on an unusually quiet day (scicomp.jlab.org)

- Optimization(s) running now with full ECCE detector simulation!
- The computational effort needed to select the points of the non-domination front from a population of size N is O(N log N) for 2 and 3 objectives, and $O(N (\log N)^{M-2})$ for M > 3 objectives [1,2].
- MOO can take advantage of supercomputing facilities, see, e.g., [3]: on the specific topic of global Detector Design with AI, White Paper submitted to DOE (CF, T. Horn, on Jan 2021).

Summary



- AI is becoming ubiquitous in NP, and remarkable accomplishments have been recently obtained.
- AI is at present already contributing to the EIC detector design: can be one of the first experiment to be designed with the support of AI.
- In NP we started exploring AI for optimal design in multidimensional space with single objectives. Most of the problems are multi-objectives though. ECCE is leading these efforts with an unprecedented attempt in detector design for complexity.
- None ever accomplished a multi-dimensional / multi-objective optimization of the global design, i.e., made by many detectors combined together, 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"*.

AI4EIC: Workshop on AI for the EIC

AI4EIC-*exp*: 1st Workshop focus on Experimental Applications of AI for the EIC

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

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