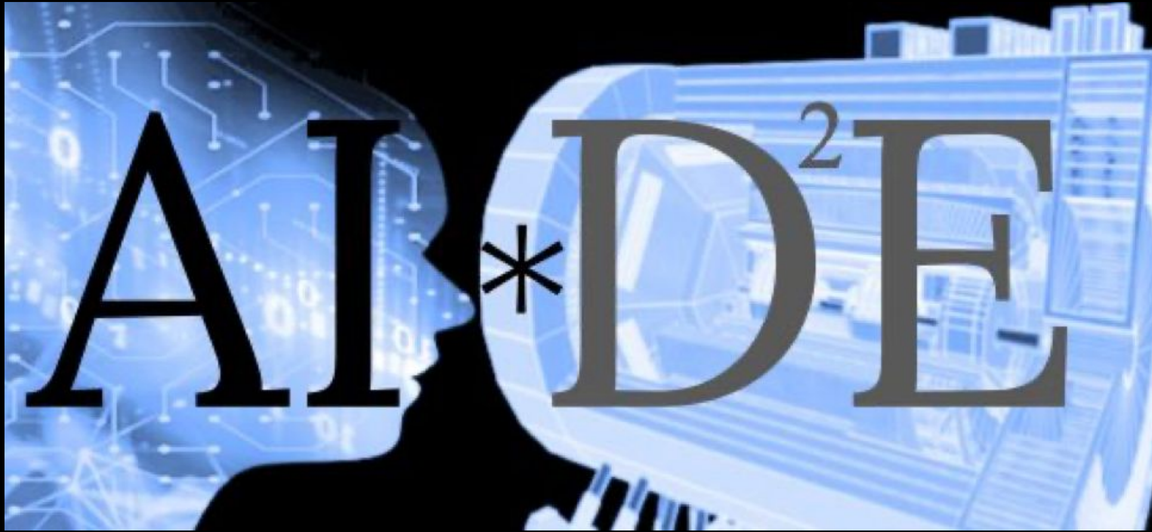
AID(2)E:
AI-Assisted Detector Design at EIC

DE-FOA-0002785



BNL, T. Wenaus
CUA, T. Horn
Duke, A. Vossen
JLab, M. Diefenthaler
W&M, CF

Cristiano Fanelli on behalf of AID(2)E

Brookhaven National Laboratory · The Catholic University of America · Duke University · Jefferson Lab · William & Mary

3rd AI4EIC Workshop, Nov 28 - Dec 1, 2023, Catholic University of America, Washington D.C.

Hot take: every optimization problem is fundamentally a multi-objective optimization problem.

Outline

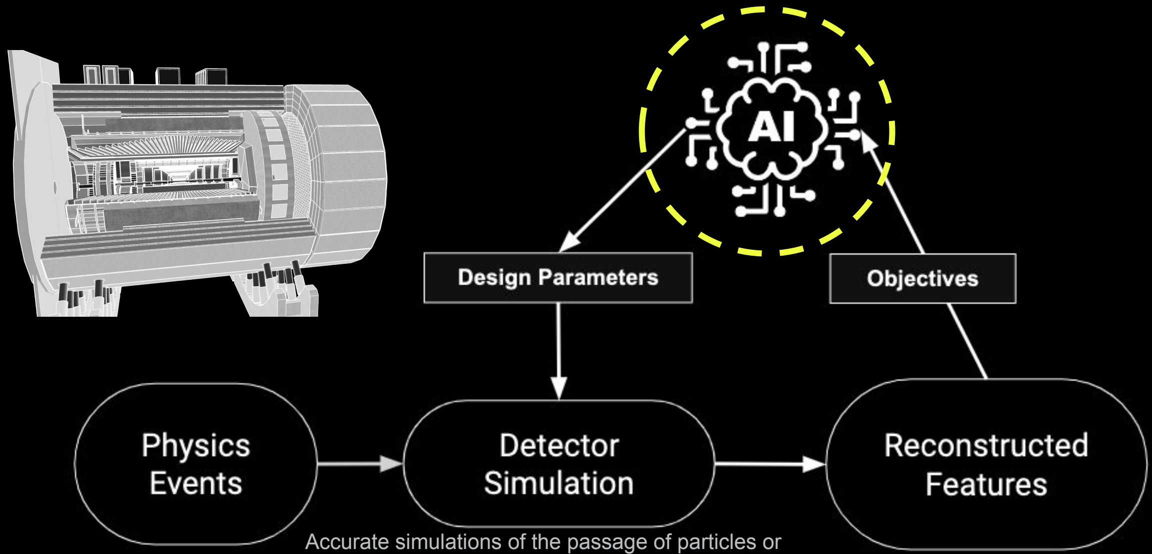
- Rationale of a MOO Approach
- Detector Design and AI assistance
- Closure Tests and Project Workflow
- Timeline and Future Applications

Rationale

1. There is not one optimal design, rather multiple tradeoff solutions
2. We want to utilize the most realistic simulation pipelines to accurately identify this set of optimal solutions
3. We want to minimize the number of design points to generate—the total computational budget—that are needed to achieve the above.

AI-Assisted Detector Design

The AI-assisted design embraces all the main steps of the sim/reco/analysis pipeline…



Accurate simulations of the passage of particles or radiation through matter

- Benefits from rapid turnaround time from simulations to analysis of high-level reconstructed observables
- The EIC SW stack offers multiple features that facilitate AI-assisted design (e.g., modularity of simulation, reconstruction, analysis, easy access to design parameters, automated checks, etc.)
- Leverages heterogeneous computing

Provide a framework for an holistic optimization of the sub-detector system
A complex problem with (i) multiple design parameters, driven by (ii) multiple objectives (e.g., detector response, physics-driven, costs) subject to (iii) constraints

Those at EIC can be the first large-scale experiments ever realized with the assistance of AI

Simulation Campaigns

1 design point

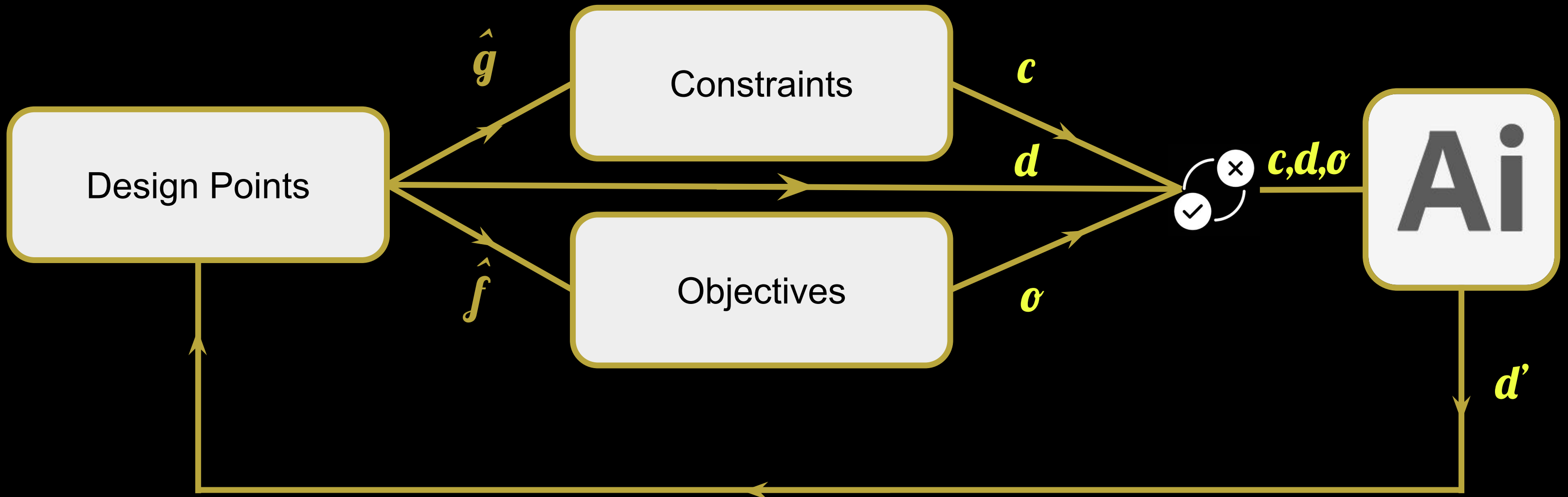
- Large simulation campaigns needed since proto-collaboration phase, where we adopted solutions with containerized software with distribution over the OSG
- This typically entails a large volume of events which are simulated for any given design of the detector ("design point")
- And more simulations need to be generated to explore multiple design points

Year	Number of Events [$\times 10^6$]	Storage [TB]	CPU-core hours [Mcore-hrs]
2022	200	50	45
2023 - 2024	100	25	22.5
2025 - 2028	50	12.5	11
2029 - 2030	500	125	110
Total	1600	400	354

Estimated simulation requirements based on observed performance in 2021. Include only large scale productions. Productions will then decrease as focus moves into hardware development before increasing significantly before initial data taking as "Mock Data Challenges" — from NIM-A: 1047 (2023): 167859 https://arxiv.org/pdf/2205.08607.pdf

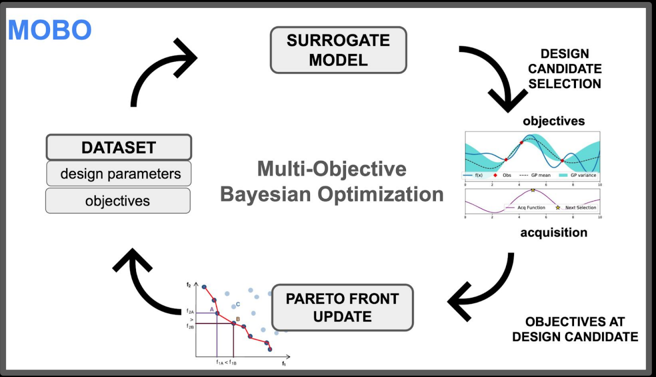
From an AI perspective

- The AI-assisted framework utilizes heterogeneous resources and waits to collect the results of the analysis from the simulated events (for a given design) and suggests next design points to query
- Main ingredients

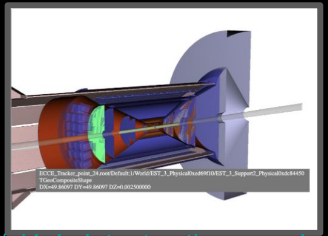


Contributions

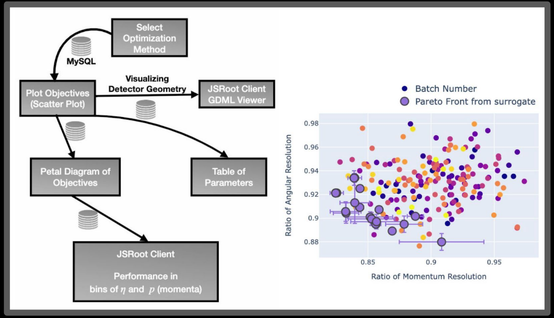
CF, Z. Papandreou, K. Suresh, et al. "AI-assisted optimization of the ECCE tracking system at the Electron Ion Collider." NIMA: 1047 (2023): 167748.
CF "Design of detectors at the electron ion collider with artificial intelligence." JINST 17.04 (2022): C04038.



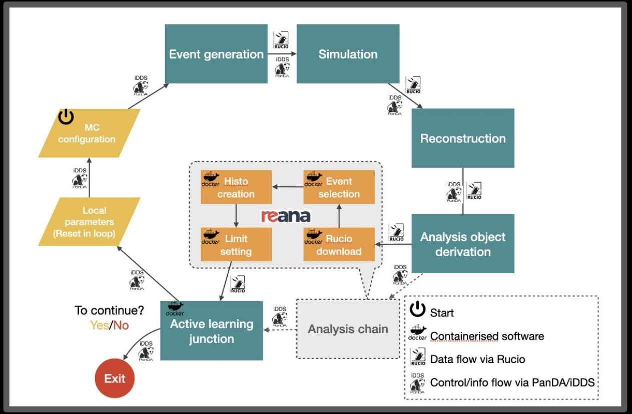
(i) Will contribute to advance state of the art MOBO complexity to accommodate a large number of objectives and will explore usage of physics-inspired approaches



https://ai4eicdetopt.pythonanywhere.com/



(ii) Development of suite of data science tools for interactive navigation of Pareto front (multi-dim design with multiple objectives)



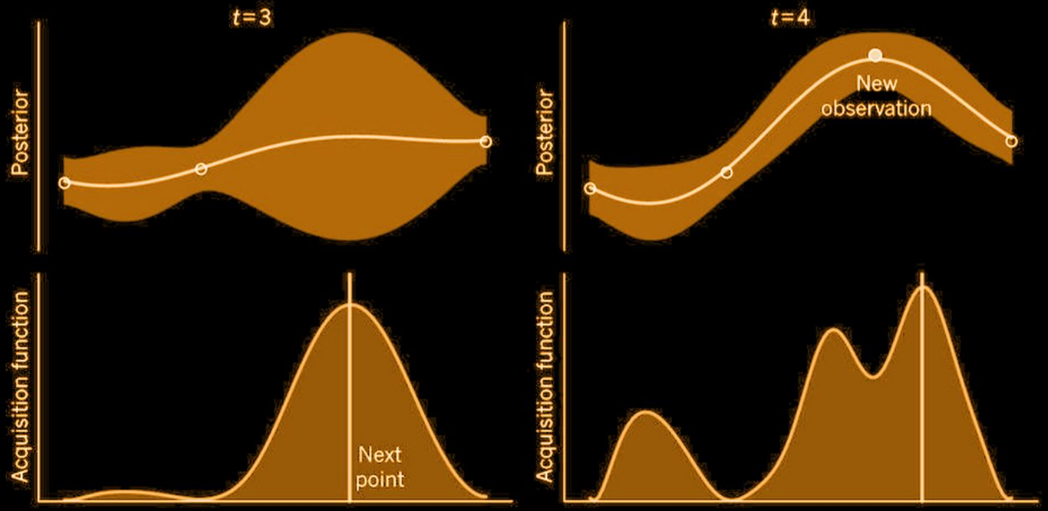
(iii) Will leverage cutting-edge workload management systems capable of operating at massive data and handle complex workflows

Examining solutions on the Pareto front of EIC detectors at different values of the budget can have great cost benefits

A fractional improvement in the objectives translates to a more efficient use of beam time which will make up a majority of the cost of the EIC over its lifetime

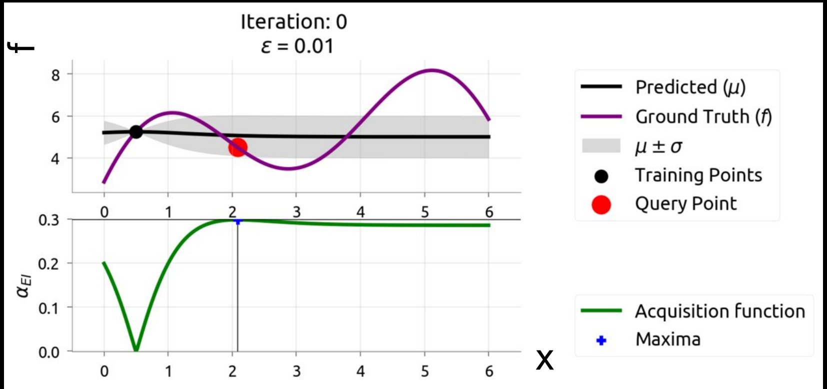
Bayesian Optimization In a nutshell

- 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.

Acquisition Functions



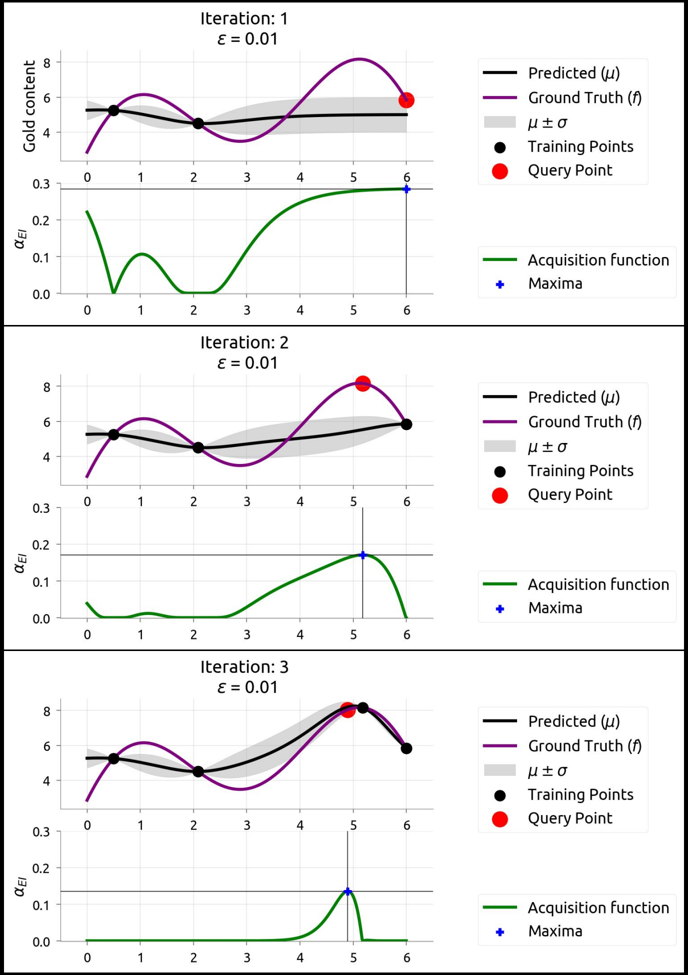
Utilization **Exploration**

$$EI(x) = \begin{cases} (\mu_t(x) - f(x^+) - \epsilon)\Phi(Z) + \sigma_t(x)\phi(Z), & \text{if } \sigma_t(x) > 0 \\ 0, & \text{if } \sigma_t(x) = 0 \end{cases}$$

Best found so far

$$Z = \frac{\mu_t(x) - f(x^+) - \epsilon}{\sigma_t(x)}$$

We are sampling x

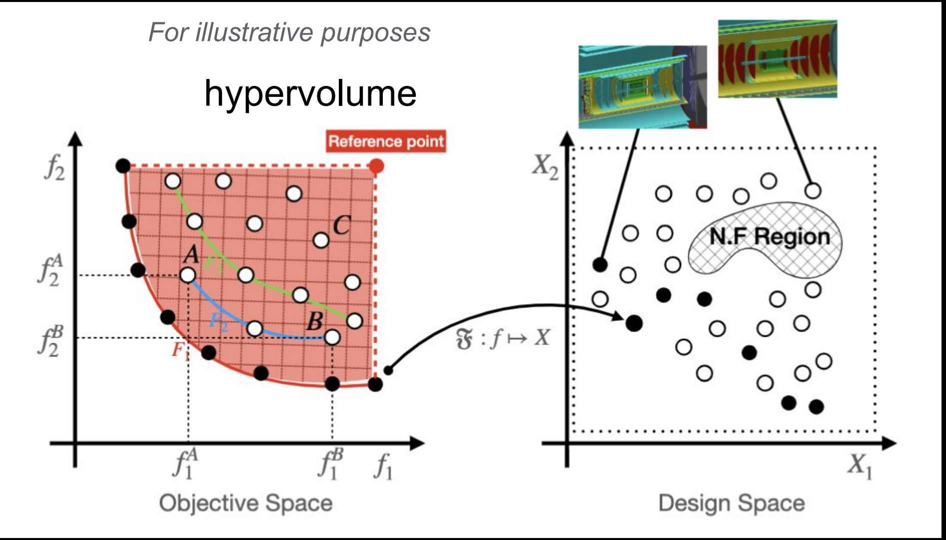


- **“Exploitation”**: search where μ is high
- **“Exploration”**: search where σ is high

Multi-Objective Optimization

MOO is needed to optimize a system of sub-detectors

- 3 Types of Objectives
  - **Intrinsic detector performance** (resolutions, efficiencies) for each sub-detector — Tracking, calorimetry, PID — noisy
  - **Physics-performance** — Multiple physics channels, equally important in the EIC physics program
  - **Costs** (e.g., material costs, provided a reliable parametrization)
- Objectives can be competing with each other
  - E.g. Better detector response come with higher costs; better resolutions may imply lower efficiencies; etc.



MOO

- In the following we will refer to the multi-objective optimization based on evolutionary algorithms [1], and in particular pymoo [2], written in Python, which also includes visualization and decision making tools.
- The definition of a generic MOO problem can be formulated as:

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

- M objective functions *f(x)* to optimize. By construction, pymoo performs minimization so a function to maximize needs a minus sign.
- There can be J inequalities *g(x)*
- There can be K equality constraints *h(x)*
- There are N variables x_i with lower and upper boundaries.

[1] Deb, Kalyanmoy. *Multi-objective optimization using evolutionary algorithms*. Vol. 16. John Wiley & Sons, 2001.
[2] Blank, Julian, and Kalyanmoy Deb. "pymoo: Multi-objective Optimization in Python." *IEEE Access* 8 (2020): 89497-89509

MOBO

n: number of design points
d: design dimensionality (each point)
M: objectives

Gaussian Process $O(n^3)$	Bayesian Sampling from posteriors NUTS – $O(Md^{5/4})$[NUTS]	Acquisition function qNEHVI – $O(Md(n+i)^M)$[2]

Gaussian Process $O(n^3)$

- Surrogate model.
- SAAS[1] priors have been proven to be successful up to 388 design dimensions
- Assumes several design variables has increased importance compared to others
- Computational expensive as iteration increases
- Benefit from GPU hardware acceleration

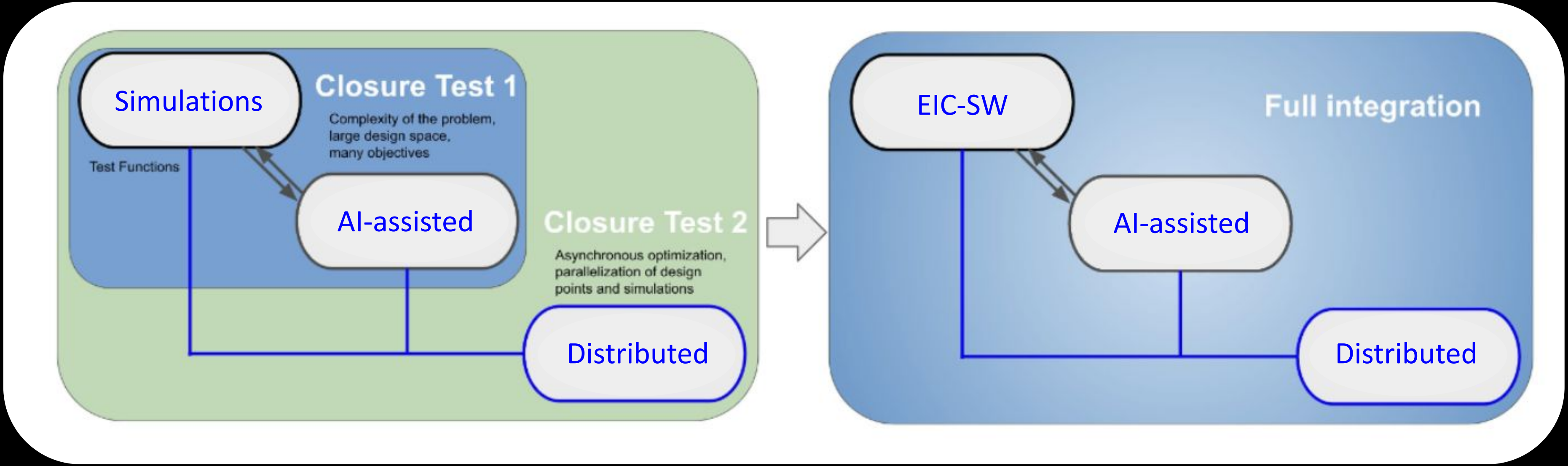
Bayesian Sampling from posteriors NUTS – $O(Md^{5/4})$[NUTS]

- Sample L points from the posterior distribution
- HMC is a popular algorithm, NUTS is a variant
- Mainly depends on the number of objectives and design space dimensions
- Has minimal dependence on iteration.
- GPU acceleration through JAX backend.

Acquisition function qNEHVI – $O(Md(n+i)^M)$[2]

- Captures HV improvement
- A “cheaper” function to evaluate as a proxy for the black box function
- Scales nonlinearly with iteration, total points explored, design space and objective space.
- Partially benefitted by GPU acceleration.

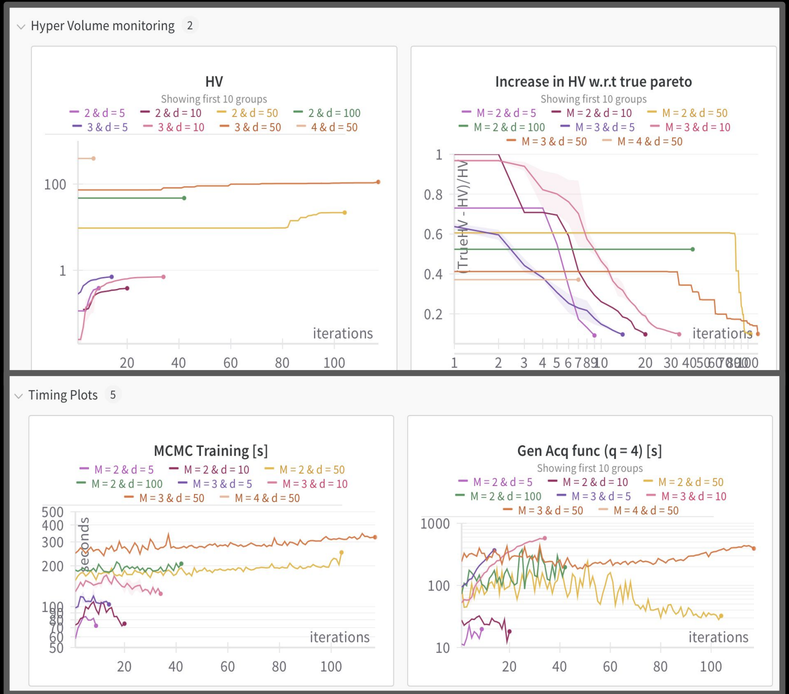
Project Workflow

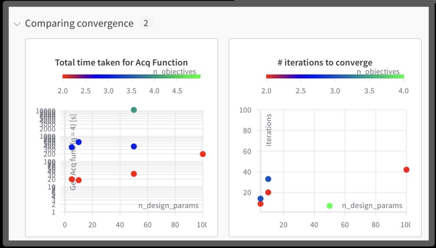


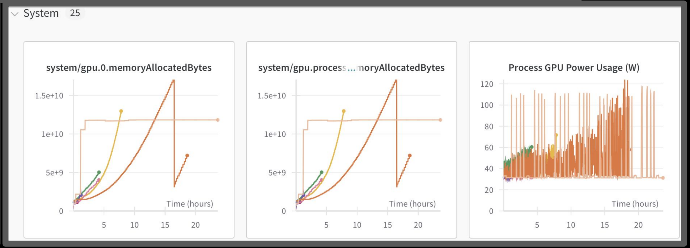
Closure Test 1

https://wandb.ai/phys-meets-ml/AID2E-Closure-1?workspace=user-karthik18495

- W&B dashboard for monitoring
  - MOBO stress-testing for problems with increasing complexity (design and objectives) and known Pareto
- Multiple metrics
  - Accuracy of optimization
  - Convergence properties
  - Compute resources

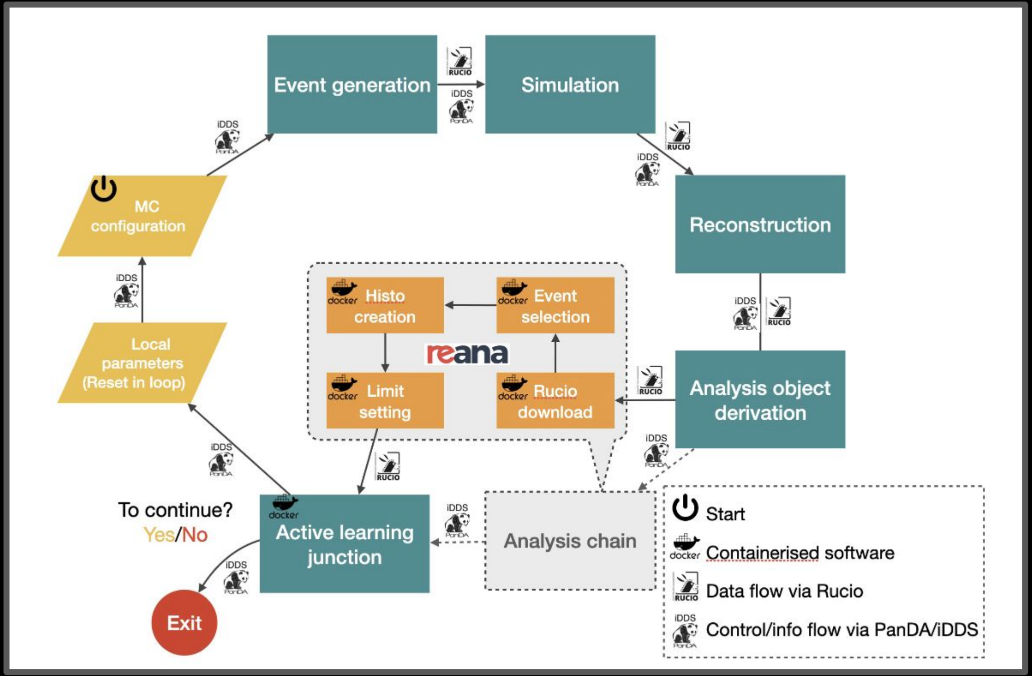






Closure Test 2: PaNDA/iDDS

- Stress-testing scalability, robustness across distributed resources
- Adapt [PanDA/iDDS]() AI/ML services to support MOBO workflow for design optimization

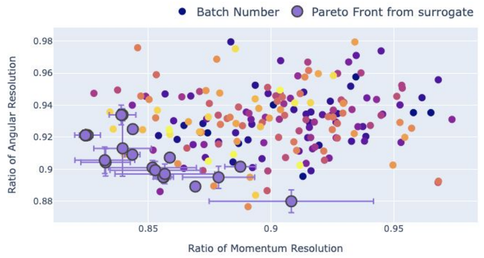


Bayesian optimisation based active learning with Panda/iDDS and Rucio

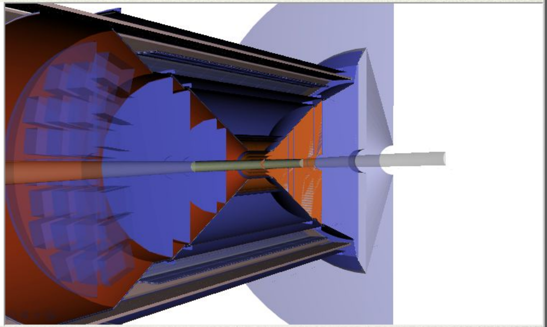
Interactive navigation of Pareto front

C.Fanelli et al, NIM A, 2023, 167748, arXiv:2205.09185

Multi Objective Bayesian Optimization



GEANT4 Visualization of the design



Click on petals for finer evaluations

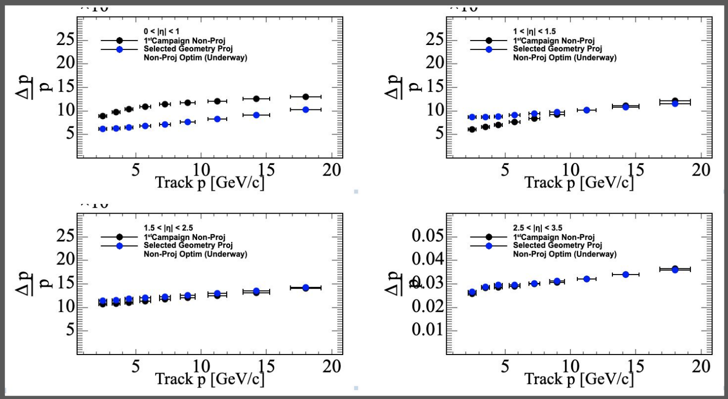
Performance of the Chosen Design Solution

Design Parameters Table

Parameter Name	Parameter Value
Angle of cone [deg]	25.00
Radius of uRwell-1 [cms]	32.47
z E-TTL [cms]	171.00
z F-TTL [cms]	157.60
z EST-1 [cms]	40.39
z EST-3 [cms]	85.09
z FST-1 [cms]	35.03
z FST-3 [cms]	83.78
z FST-5 [cms]	131.27

The whole idea of the AI-assisted design is that of determining trade-off optimal solutions in a multidimensional design space driven by multiple objectives

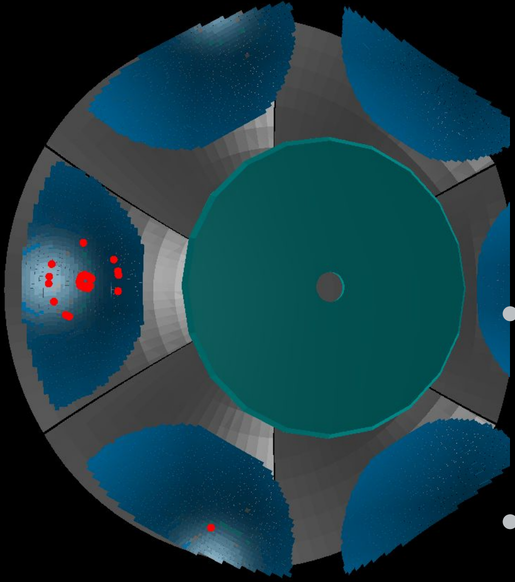
For an interactive visualization:
https://ai4eicdetopt.pythonanywhere.com



Candidates for Optimization in ePIC

Considering all the constraints as ePIC is in the process of finalizing engineering designs, we can select those sub-detectors that still have tunable parameters

E. Cisbani *et al* 2020 *JINST* **15** P05009



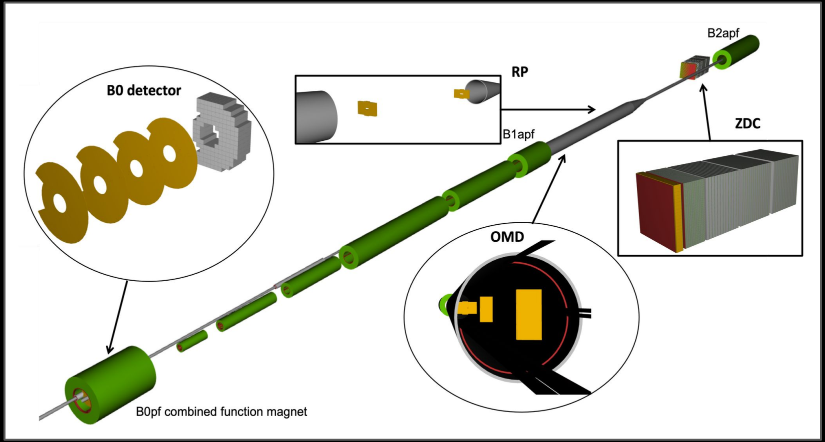
- *Mirror, sensor placement, gas, mirror material (lower costs material)...*

- *PID performance, costs, …*

dual-RICH

- *B0 magnetic field map, distance between space (always considered even), central location of tracker*

- *Momentum resolution, acceptance*



Far-Forward

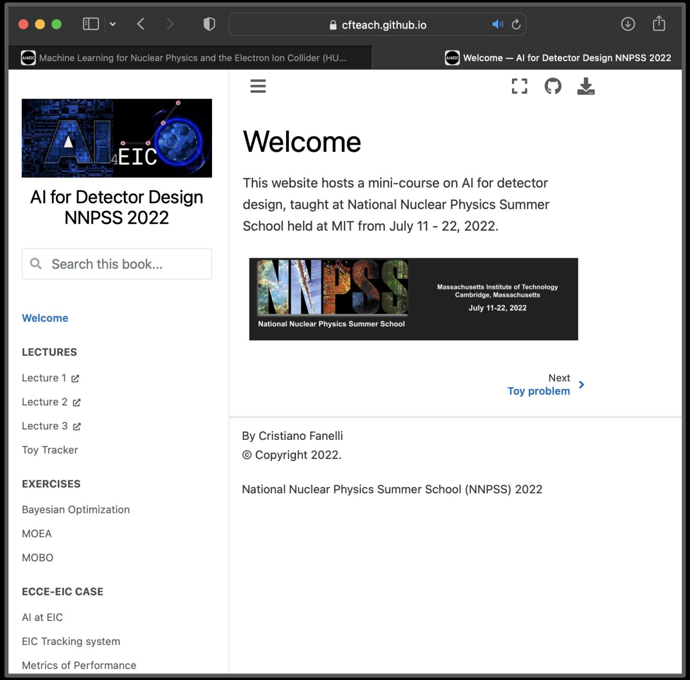
Ongoing discussion with working groups to identify potential

Documentation and Outreach

- GitBook and/or other knowledge sharing platforms will be part of the initiatives related to documentation and outreach
- Offering opportunities for experiential learning with easy access for beginners

http://cfteach.github.io/nnpss

https://cfteach.github.io/HUGS23



Conclusions

- *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* — DOE Town Halls on AI for Science in 2019
- We are aiming to complete a first coupling of MOBO with the EIC shell in the next few months (targeting Spring) to exercise the machinery
- Ultimately, we can realize a framework that can optimize holistically a large-scale detector, and that is scalable and distributed. The Detector-2 at EIC seems to be an ideal candidate for this at this stage.
- In detector projects, most changes happen during the construction phases (e.g., changes in the available material or budget). AID2E will be an ideal tool to optimize design changes with objectives (e.g., reduce cost).
- This framework inherently offers broader impacts, as it can be adapted for use in various experiments and is suitable for a wide range of compute-intensive applications that necessitate MOO (e.g., calibrations, alignments, etc)

Backup