

OPEN ACCESS

AI-assisted detector design for the EIC (AID(2)E)

To cite this article: M. Diefenthaler *et al* 2024 *JINST* **19** C07001

View the [article online](#) for updates and enhancements.

You may also like

- [Analysis of the MHD stability and energetic particles effects on EIC events in LHD plasma using a Landau-closure model](#)
J. Varela, D.A. Spong, L. Garcia et al.
- [Recent results from deuterium experiments on the large helical device and their contribution to fusion reactor development](#)
Masaki Osakabe, Hiromi Takahashi, Hiroshi Yamada et al.
- [ON THE IMPLICATIONS OF LATE INTERNAL DISSIPATION FOR SHALLOW-DECAY AFTERGLOW EMISSION AND ASSOCIATED HIGH-ENERGY GAMMA-RAY SIGNALS](#)
Kohta Murase, Kenji Toma, Ryo Yamazaki et al.

**HONOLULU, HI**
October 6-11, 2024

Joint International Meeting of

The Electrochemical Society of Japan (ECSJ)

The Korean Electrochemical Society (KECS)

The Electrochemical Society (ECS)



Early Registration Deadline:

September 3, 2024

MAKE YOUR PLANS NOW!



ARTIFICIAL INTELLIGENCE FOR THE ELECTRON-ION COLLIDER

WASHINGTON, D.C., U.S.A.

28 NOVEMBER – 1 DECEMBER 2023

AI-assisted detector design for the EIC (AID(2)E)

M. Diefenthaler,^d C. Fanelli^{Ⓜ,e,*} L.O. Gerlach,^c W. Guan,^a T. Horn,^b A. Jentsch,^a M. Lin,^a
K. Nagai,^c H. Nayak,^e C. Pecar,^c K. Suresh,^e A. Vossen,^{c,d} T. Wang^a and T. Wenaus^a
on behalf of the AID(2)E collaboration

^aBrookhaven National Laboratory, Upton, NY 11973, U.S.A.

^bCatholic University of America, Washington, DC 20064, U.S.A.

^cDuke University, Durham, NC 27708, U.S.A.

^dJefferson Lab, Newport News, VA 23606, U.S.A.

^eWilliam & Mary, Williamsburg, VA 23185, U.S.A.

E-mail: cfanelli@wm.edu

ABSTRACT: Artificial Intelligence is poised to transform the design of complex, large-scale detectors like ePIC at the future Electron Ion Collider. Featuring a central detector with additional detecting systems in the far forward and far backward regions, the ePIC experiment incorporates numerous design parameters and objectives, including performance, physics reach, and cost, constrained by mechanical and geometric limits. This project aims to develop a scalable, distributed AI-assisted detector design for the EIC (AID(2)E), employing state-of-the-art multiobjective optimization to tackle complex designs. Supported by the ePIC software stack and using GEANT4 simulations, our approach benefits from transparent parameterization and advanced AI features. The workflow leverages the PanDA and iDDS systems, used in major experiments such as ATLAS at CERN LHC, the Rubin Observatory, and sPHENIX at RHIC, to manage the compute intensive demands of ePIC detector simulations. Tailored enhancements to the PanDA system focus on usability, scalability, automation, and monitoring. Ultimately, this project aims to establish a robust design capability, apply a distributed AI-assisted workflow to the ePIC detector, and extend its applications to the design of the second detector (Detector-2) in the EIC, as well as to calibration and alignment tasks. Additionally, we are developing advanced data science tools to efficiently navigate the complex, multidimensional trade-offs identified through this optimization process.

KEYWORDS: Detector design and construction technologies and materials; Computing (architecture, farms, GRID for recording, storage, archiving, and distribution of data); Software architectures (event data models, frameworks and databases); Detector alignment and calibration methods (lasers, sources, particle-beams)

*Corresponding author.



Contents

1	Introduction	1
2	Methodology and strategy	2
3	Broader impact and future work	7
4	Documentation and outreach	8
5	Conclusions	8

1 Introduction

The Electron Ion Collider (EIC) stands out as one of the pioneering large-scale experiments to incorporate artificial intelligence (AI) right from its design phase. As highlighted in the 2019 DOE Town Halls on AI for Science: “*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.*” This paper presents AID(2)E, a DOE-funded initiative aimed at developing a scalable and distributed framework for AI-assisted detector design at the EIC.

The project was initiated to address the limitations of current detector optimization methods. Traditionally, each subsystem is optimized individually within global design constraints before integration. This conventional approach, often relying on ad hoc or brute-force methods like grid search, is sub-optimal. It struggles with handling multiple design parameters, fails to optimize multiple objectives simultaneously (such as efficiency, resolution, and costs), and lacks holistic optimization of the entire detector. Furthermore, it does not fully exploit the benefits of multi-objective optimization, which can reveal optimal trade-offs between competing objectives. The rationale for employing a multi-objective optimization approach in detector design is that there is no single optimal solution; instead, there exists a spectrum of viable solutions that balance factors such as detector response, physics design requirements, and costs. This is especially significant given the high costs associated with large-scale detectors, exemplified by the estimated \$300M budget for the ePIC detector [1]. As demonstrated in previous studies (e.g., [2, 3]), these methods provide new insights into internal correlations among design parameters and enable trade-offs that are not achievable with traditional methods. AI-assisted optimization complements the expertise of detector specialists and, as shown in [3], can improve objectives compared to a baseline design across a large phase space. A holistic optimization of the entire detector, considering multiple objectives simultaneously, represents a paradigm shift in detector design. AI-assisted methods can significantly impact future large nuclear physics projects, such as those at the EIC. Using realistic simulation pipelines and minimizing the computational budget required, the AID(2)E Collaboration aims to streamline the design process while ensuring high-fidelity results. The primary goal of AID(2)E is to develop infrastructure that optimizes detector design in a scalable and distributed manner. The project begins with two closure tests designed to demonstrate the convergence of the Pareto front [4] in multi-objective optimization problems and to validate the effectiveness of distributing the compute-intensive simulations of design

points suggested by the optimizer. AID(2)E has already made contributions towards optimizing the ePIC detector subsystems. Currently, the collaboration is optimizing two detector subsystems, which are discussed in this proceeding. This optimization approach can potentially be extended to other subsystems of ePIC, making it one of the first large-scale detector designed with the assistance of AI. Evaluating different solutions on the Pareto front of ePIC at various budget levels can yield significant cost benefits. Even fractional improvements in objectives, such as resolution and efficiency, can lead to a more efficient use of beam time, a major cost factor for the EIC over its lifetime. Characterizing the Pareto front can inform the design of a complementary Detector-2 at EIC. Our team is also developing a suite of data science tools to facilitate effective exploration of the determined Pareto front [4]. It is worth mentioning that the AID(2)E framework also holds potential for other applications, such as calibration and distributed machine learning inference engines. Furthermore, AID(2)E aims to engage the community through bootcamps and short-term schools for students.

2 Methodology and strategy

The primary goal of the AID(2)E project is to develop a scalable and distributed AI-assisted Detector Design framework for the EIC, specifically focusing on the ePIC detector. Central to AID(2)E is Multi-Objective Optimization (MOO), which evaluates various design objectives such as sub-detector intrinsic responses (e.g., resolution and efficiency), performance for key EIC physics channels, and material cost proxies. This project is pioneering in its application of MOO at such a scale, facilitated by a developing distributed workflow infrastructure, enhanced with PanDA for simulation workflows [5]. Our strategy includes two closure tests: first, we want to test and characterize the performance of approaches like Multi-Objective Bayesian Optimization (MOBO) [6] and Multi-Objective Genetic Algorithms (MOGA), specifically NSGA-II [7], using test functions. The second one proves the effectiveness of distributing the simulations of different design points, coordinating data collection for objective updates, and refining the design iteratively. The framework supports continuous and distributed AI-driven optimization. Initial efforts are concentrated on optimizing specific sub-detector systems within ePIC, paving the way for broader application. Detailed discussions on the optimization approach, distributed workflow, and specific applications to the ePIC detector are presented in what follows.

Multi-objective optimization. The definition of a generic detector MOO problem can be formulated as follows:

$$\begin{aligned}
 & \min \mathbf{f}_m(\mathbf{x}) & m = 1, \dots, M & \quad \text{(objectives)} \\
 \text{s.t. } & \mathbf{g}_j(\mathbf{x}) \leq 0, & j = 1, \dots, J & \quad \text{(inequality constraints)} \\
 & \mathbf{h}_k(\mathbf{x}) = 0, & k = 1, \dots, K & \quad \text{(equality constraints)} \\
 & x_i^L \leq x_i \leq x_i^U, & i = 1, \dots, N & \quad \text{(parameters)}
 \end{aligned} \tag{2.1}$$

where one has M objective functions f_m to optimize (e.g., detector resolution, efficiency, signal/background for a physics channel, costs), subject to J inequalities $g_j(x)$ and K equality constraints $h_k(z)$ (e.g., mechanical constraints), in a design space of N dimensions (e.g., geometry parameters that change the GEANT4 [8] design) with lower (x_i^L) and upper (x_i^U) bounds on each dimension. We are utilizing Multi-Objective Bayesian Optimization (MOBO) and Multi-Objective Optimization using Genetic Algorithms (MOGA) to tackle detector design optimization. These techniques assess

multiple objectives such as sub-detector responses, physics channel performance, and material costs, and are detailed in our workflow illustrated in figure 1.

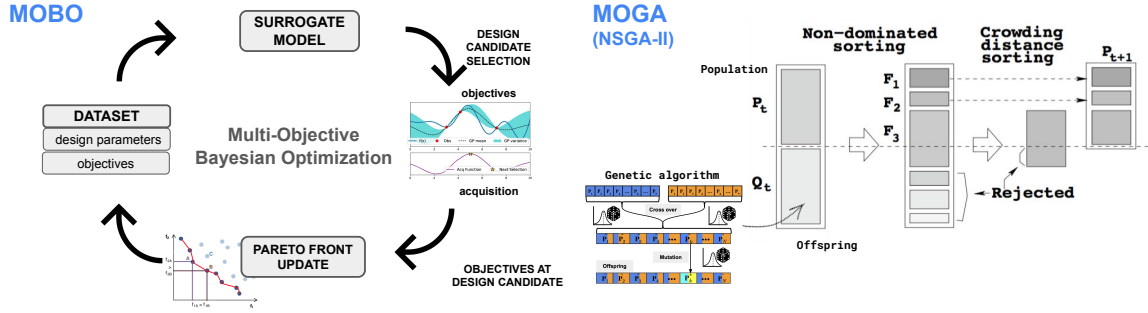


Figure 1. *Multi-objective optimization workflows:* the left panel shows the Bayesian workflow MOBO, which is the primary choice in this project (generally capable of providing accurate determination Pareto fronts). The right panel shows MOGA (NSGA-II, based on genetic algorithms [7]), which is an alternative approach (more suitable to handle, e.g., possible discontinuities or high-dimensional spaces).

MOBO is our primary choice for the ePIC detector, driven by our improved understanding of technology and parameter ranges, the ability to incorporate uncertainties and physics-inspired acquisition functions, and a more accurate approximation of the Pareto front compared to MOGA. This approach has proven efficient in requiring fewer evaluations to converge and allows for custom acquisition functions, supported by tools like `ax-platform` and `BoTorch` [9, 10]. We are currently testing MOBO with a `qNEHVI` acquisition function [11], optimizing for scalability across multiple design points and dimensions. Scalability challenges for MOBO include computational constraints as the number of observations grows, which we can address using techniques to manage the increasing complexity of the surrogate model, such as `SAASBO` and `TURBO` [12, 13]. Both MOBO and MOGA offer distinct computational and scalability properties that dictate their application and workload distribution, crucial for the efficient design of the ePIC detector. One advantage of MOGA compared to MOBO is the possibility of handling nondifferentiable terms, and higher dimensionality problems, leveraging the inherent parallelization of genetic algorithms.

Distributed workflow. The AID(2)E team is committed to advancing AI-based detector design optimization at the EIC by developing a scalable and distributed framework. This framework is refined through its application to the ePIC detector design and is delivered as a robust, experiment-agnostic solution ready for broader detector design applications. The project leverages existing, scalable distributed workflow and workload management tools like `PanDA` [14], developed for the ATLAS experiment and now supporting complex AI/ML workflows including Bayesian Optimization. Historically, these systems have supported large-scale data handling and processing needs of the ATLAS experiment and other projects like the Rubin Observatory and sPHENIX, demonstrating the versatility and scalability of the `PanDA` and the intelligent Data Delivery Service (iDDS) [14]. These tools have been integral in managing workflows across diverse computing environments — grids, clouds, and supercomputers — enhancing their stability and adaptability.

Recent developments have focused on integrating complex AI/ML workflows, particularly those utilizing GPUs, into this system. This includes hyperparameter optimization and active learning based on Bayesian optimization, extending the capabilities of ATLAS analyses. Additionally, full-chain

integrated production-analysis workflows, including the use of REANA [15] in the analysis phase, have been supported. The PanDA/iDDS development program emphasizes ease of use and accessibility, supporting a range of AI/ML applications through a unified platform that integrates multiple computing facilities. This capability is critical for processing-intensive, distributed AI-based detector design, offering: (i) seamless operation across distributed resources; (ii) support for comprehensive AI/ML tools and full-chain workflows from simulation to iterative refinement; (iii) efficient authentication and user-friendly interfaces, including support for modern and legacy security mechanisms. Ultimately, this framework not only supports current project needs but is also adaptable to similar future applications, demonstrating its potential through an implementation of Bayesian Optimization-based active learning now enhancing the physics reach of ATLAS analyses.

Closure tests. The primary objective of the closure tests within the AID(2)E framework is to rigorously evaluate system constraints and identify potential bottlenecks. This framework enhances the distribution of simulations using the ePIC software stack [16] and advances MOO techniques for detector design. The first closure test analyzes these bottlenecks within a test optimization problem, employing MOBO to assess its scalability and efficiency relative to problem complexity. To meet the computational demands of evaluating multiple detector design points, the ePIC detector design optimization requires a large volumes of simulated data. The PanDA System [14] is utilized to manage these tasks across varied geographic locations, coordinated by the iDDS. The second closure test integrates and stress-tests the PanDA framework, initially with test functions and subsequently with actual ePIC simulation payloads. This process validates the framework’s capability to efficiently distribute simulations and synchronize them with an optimization strategy, ensuring robust performance across diverse operational scenarios. A high-level workflow of the AID(2)E framework, based on the mentioned closure tests, is displayed in figure 2.

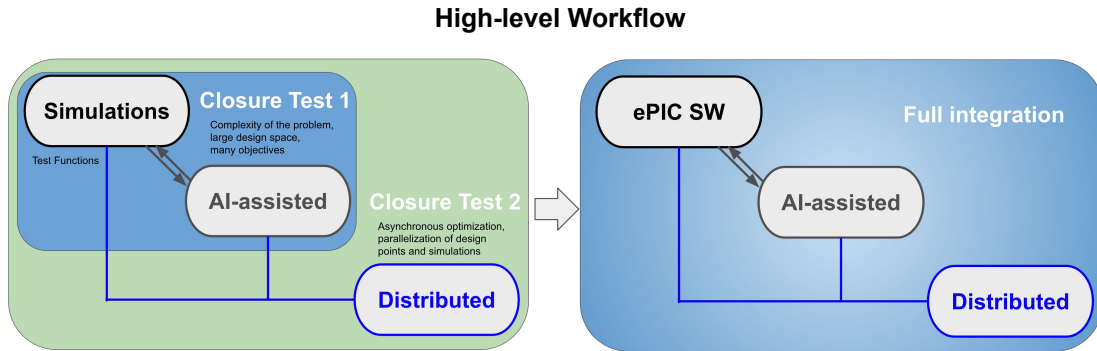


Figure 2. *High-level workflow of AID(2)E:* we first demonstrate an AI-assisted framework handling multiple test objectives (“simulations”) with a known Pareto front (*closure test 1*). We also show the feasibility of distributing these simulations across various HPC/HTC clusters using the PanDA framework [14] (*closure test 2*). Following closure test 1, we replace the test functions with the ePIC software (compute-intensive simulations and reconstruction). The final step is a *full integration*, where the ePIC software is coupled with a distributed AI-assisted framework that asynchronously orchestrates the design points suggested by AI.

The primary goal of Closure Test 1 is to evaluate the complexity of the optimization algorithm for convergence with up to 5 objectives and $O(100)$ design parameters. This test identifies bottlenecks in multi-objective Bayesian optimization algorithms and helps deploy the AID(2)E framework for complex

optimization problems. A realistic test function is chosen to pinpoint choke points. Improvements are achieved by reducing computational time and space complexity or the number of iterations needed for convergence. Closure Test 1 uses a test function with a known Pareto front to evaluate convergence. For this purpose, the DTLZ-2 benchmark problem for multi-objective optimization [17] is selected and defined as follows. Given n decision variables and M objectives, where $M \geq 2$, the decision vector $x = [x_1, x_2, \dots, x_n]$ is subject to the constraints $0 \leq x_i \leq 1$ for $i = 1, 2, \dots, n$, and $g(x) = \sum_{i=M}^n (x_i - 0.5)^2 - \cos(20\pi(x_i - 0.5)) \leq 0$. The objective functions are defined as:

$$\begin{aligned} f_1(x) &= (1 + g(x)) \cos\left(x_1 \frac{\pi}{2}\right) \cos\left(x_2 \frac{\pi}{2}\right) \cdots \cos\left(x_{M-1} \frac{\pi}{2}\right) \\ f_2(x) &= (1 + g(x)) \cos\left(x_1 \frac{\pi}{2}\right) \cos\left(x_2 \frac{\pi}{2}\right) \cdots \sin\left(x_{M-1} \frac{\pi}{2}\right) \\ f_3(x) &= (1 + g(x)) \cos\left(x_1 \frac{\pi}{2}\right) \cos\left(x_2 \frac{\pi}{2}\right) \cdots \sin\left(x_{M-2} \frac{\pi}{2}\right) \\ &\vdots \\ f_{M-1}(x) &= (1 + g(x)) \cos\left(x_1 \frac{\pi}{2}\right) \sin\left(x_2 \frac{\pi}{2}\right) \\ f_M(x) &= (1 + g(x)) \sin\left(x_1 \frac{\pi}{2}\right) \end{aligned}$$

Here, $g(x)$ is the sum constraint defined earlier. The solution to this multi-objective function is satisfied when at least M design variables meet the condition $x_i = 0.5$ for $i = 1, 2, \dots, M$. This function is chosen because, like real-world scenarios, not all design variables influence every objective. The true Pareto front for the DTLZ-2 problem is known for given n design variables and M objectives. The convergence criterion is based on the computed hyper-volume, with a tolerance of 5% to the true hyper-volume:

$$\frac{|\text{HV}_{\text{true}} - \text{HV}_i|}{\text{HV}_{\text{true}}} \leq 5\%$$

To understand the optimization algorithm's complexity and bottlenecks, we study the time and space complexity as functions of M objectives and n design variables. The overall complexity is the convolution of the complexities in building the surrogate model and using the acquisition function. Current studies evaluate these complexities at each step of the Bayesian process. Preliminary results are shown in figure 3.

Future work will extend this to more combinations of surrogate models and acquisition functions to verify the complexities reported in the literature.

Closure Test 2 aims to efficiently distribute simulations for various optimization tasks through three stages of parallelization. First, multi-processing within a single node is used when simulation time is negligible compared to queue time. This is parallelized with Joblib [18], as in Closure Test 1. Second, parallelization with HPC workload managers like SLURM [19] and PBS [20, 21] is applied when simulation time is comparable to queue time. Third, parallelization with PanDA [14] — a data-driven workload management system designed for the LHC data processing scale — is used when the simulation time is much longer than the queue wait time for job submission. Closure Test 2 objectives are to integrate the optimization pipeline to use PanDA for evaluating design points and to integrate the ePIC software stack with PanDA for detector design evaluation. PanDA offers advantages such as OpenID Connect (OIDC) authentication for cross-site submissions and PanDA

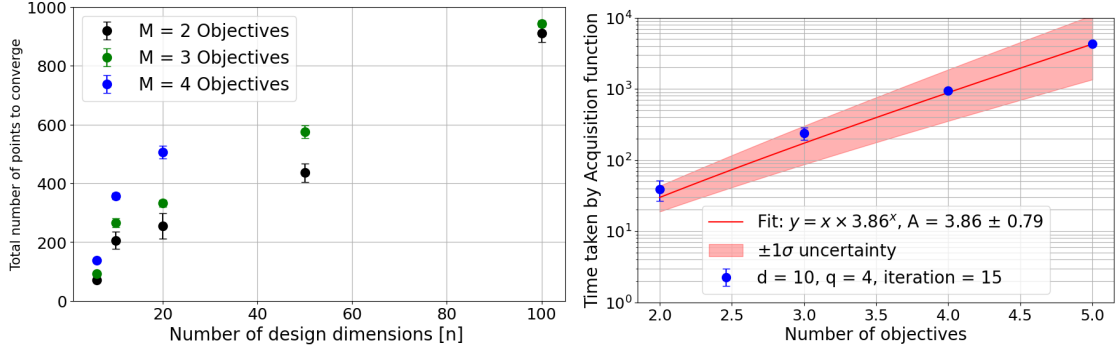


Figure 3. *Results from closure test* (Left) The figure depicts the total explored points before DTLZ-2 convergence as a function of design parameters (n). Each trial was repeated for statistical robustness, with varying colors indicating a different number of objective. The process continued until either 1000 points were explored or convergence was reached, except for cases with $n = [50, 100]$ and $m = 4$ where convergence was not achieved. A comparison with expected computational complexity is underway. (Right) The figure illustrates the time complexity of the qNEHVI acquisition function. According to [11], it is $O(m \cdot (N + I)^m)$, where N is the observed points and I is partitions in the cached box decomposition. Further studies on varying design parameters and standard latency are in progress to elaborate on the problem’s complexity.

Monitor for job status tracking [14]. Ongoing efforts include creating authentication pipelines and retrieving information from various sites.

Applications to the ePIC detector. We are currently optimizing two detection systems in ePIC: the dual-RICH (dRICH) in the central region and the B0 detector in the forward region.

The dual-radiator Ring Imaging Cherenkov detector (dRICH) will provide charged-hadron particle ID in the forward rapidity region of ePIC for a wide momentum range of $\sim 2 \text{ GeV}/c < p < 50 \text{ GeV}/c$ [2, 22]. The objective of this optimization is to improve pion-kaon and kaon-proton separation, as well as acceptance, at both low and high momentum across the full η range covered by the dRICH. An initial study applying MOBO to the design of the ePIC dRICH has been conducted, with design points evaluated using the ePIC full simulation. The design parameters used in this optimization are the aerogel disk radius, the spherical mirror position (z, x) and radius (with the z position constrained such that the mirror backplane is within 4 cm of the dRICH back wall), and the photosensor sphere position (z, x) and radius (with the z position constrained such that sensors are within the sensor box required for services). This totals seven design parameters with two hard constraints on the search space. Overlap checks are conducted for each sampled design point before simulation. For each design point, 1000 K^+ and 1000 π^+ particles are simulated in six p/η ranges. The momentum values are 15 GeV/ c and 40 GeV/ c , at the upper edge of the momentum range covered by each radiator. The η ranges sampled are $[1.3, 2.0]$, $[2.0, 2.5]$, $[2.5, 3.5]$. Following [2], the objectives chosen for this test optimization are the $N\sigma$ pion-kaon separation for the low and high η ranges, computed as

$$N\sigma_{\pi-K} = \frac{|\langle \theta_K \rangle - \langle \theta_\pi \rangle| \sqrt{\langle N_{\text{photons}} \rangle}}{\sigma_\theta},$$

where $\langle \theta \rangle$ is the average reconstructed Cherenkov angle for pions/kaons, $\langle N_{\text{photons}} \rangle$ is the average number of reconstructed photons per track, and σ_θ is the Cherenkov angle resolution from a Gaussian fit to the Cherenkov angle residual, averaged over pions and kaons. $N\sigma_{\pi-K}$ is averaged over the two

momentum points for each η range. This optimization was carried out with 50 initial Sobol points, a SAASBO surrogate model, and a qNEHVI acquisition function, with the optimization executed for 200 total evaluations in batches of 10.¹ Convergence was achieved after 150 trials. Preliminary Pareto-optimal points found after these trials include potential designs that outperform the nominal dRICH geometry. Initial results shown at the dRICH working group meetings of ePIC suggest that while the nominal dRICH geometry is well-optimized for the large η region, the MOBO algorithm identifies opportunities for improvement within the available design space.

Far-forward detectors [23] measure very forward neutral and charged particles [24]. The EIC far-forward detectors consist of four independent B0 subsystems: Tracking and Photon Detection, Roman Pots, Off-Momentum Detectors, and the Zero-Degree Calorimeter. The B0 system measures charged particles in the forward direction and tags neutral particles. Off-momentum detectors measure charged particles from decays and fission, Roman pot detectors measure charged particles near the beam, and the zero-degree calorimeter measures neutral particles at small angles. The optimization objective for the B0 ECAL calorimeter is to improve momentum resolution under a non-homogeneous magnetic field while simultaneously increasing the B0 ECAL acceptance.² Acceptance is defined as the ratio of the number of tracks before the first tracking disk to the number of showers detected by the B0 ECAL. This is challenging because the magnetic field peaks within a limited range in z , and moving it to extremes can disrupt acceptance. There are four main design parameters, specifically the disk z positions. Some constraints have also been applied to reflect the real scenario, such as all trackers should be placed within 6.65 m, and the sum of the z -positions of two subsequent trackers should be ≥ 10 cm.

A visualization of the sub-detecting systems described in this section is shown in figure 4.

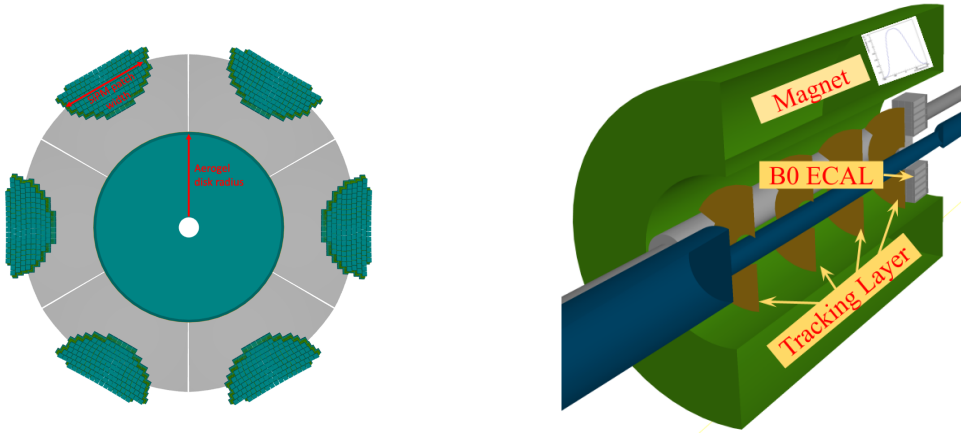


Figure 4. Sub-detector system in ePIC optimized with AID(2)E: (left) the dRICH design is characterized by two radiators (aerogel and gas) and 6 identical sectors with photosensors [2]; (right) The B0 Tracker and the B0 ECAL detector. The magnetic field in the B0 Region is inhomogeneous making the optimization more challenging.

3 Broader impact and future work

Our work with the dRICH and B0 sub-detectors at the ePIC experiment demonstrates the potential for extending our methodologies to other sub-detector systems, tailoring to the ePIC detector needs.

¹Definitions of Sobol points, SAASBO and qNEHVI can be found in the tutorial [12].

²A schematic representation of the far-forward detectors can be found in [25].

Following the approach implemented in [3], the tracking system preceding the dRICH is a prime candidate for further optimization in combination with the dRICH. This method is largely experiment-agnostic and applicable to various detection systems, although MOBO currently handles up to four objective functions effectively. For more complex scenarios, the MOGA approach currently available within the AID(2)E framework offers a viable alternative.

Large-scale, compute-intensive applications often face resource-based limitations that hinder iterative development, training, and optimization. Our framework facilitates AI/ML workflows across geographically distributed, multi-platform sites, significantly enhancing the scalability and accessibility for researchers. This reduces barriers from small-scale development to large-scale deployment, enabling deeper and more complex model processing that can yield substantial scientific benefits. The potential applications of this framework include: (i) *Design of future large-scale experiments* in nuclear physics and high-energy physics, such as the early-stage second EIC detector, optimizing for complementarity with the ePIC detector across their full physics reach; (ii) *Alignments and calibrations*, addressing compute-intensive and multi-objective needs in detector operations; (iii) *Multi-objective hyperparameter optimization* for deep neural networks, aiming to balance network performance against complexity, as shown in [26].

4 Documentation and outreach

We will employ GitBook [27] or Jupyter Book [28] along with other knowledge-sharing platforms to support our documentation and outreach efforts, ensuring that these resources are easily accessible to beginners. Automated documentation will be implemented using a Retrieval-Augmented Generation (RAG)-based approach, as described in [29], with an AI4EIC bot available to the public via a specific URL. The AID(2)E project employs a comprehensive suite of data science tools to analyze the results of various optimizations. Interested readers can refer to [30] for a dashboard that displays the performance of closure tests, and to [4] for an example of interactive navigation of Pareto front solutions from the ECCE tracker design. This project will also feature summer bootcamps on AI-assisted design optimization at William & Mary, incorporating lectures cited in [31] and [32], along with hands-on research activities with the AID(2)E group. Participants will engage in cutting-edge research, learning about AI applications in detector design and data science for physics. Additional information about the project is available in the main repository, detailed in [33].

5 Conclusions

We are actively working to integrate Multi-Objective Optimization with the EIC shell software, and initial results have already been obtained for two detector systems: the dRICH in the central region and the B0 in the far-forward region of ePIC. This development represents a significant milestone, testing the capabilities of our approach in a real-world scenario. Ultimately, our goal is to develop a scalable and distributed framework capable of holistically optimizing large-scale detectors. Ultimately, our goal is to develop a scalable and distributed framework capable of holistically optimizing large-scale detectors, integrating and simultaneously optimizing multiple subdetector systems. Given its extended development timeline compared to ePIC at the EIC, Detector-2 is an ideal candidate for a holistic optimization due to its early stage of development. It is also worth noting that during the construction phase of detector projects, numerous changes are often necessary, such

as adjustments due to material availability or budget constraints. The AID(2)E framework is the ideal tool to optimize these design modifications efficiently, utilizing objectives like cost reduction. Moreover, this framework is poised to have broader impacts beyond its immediate application. It is versatile enough to be adapted for various experimental setups and is particularly well-suited for compute-intensive tasks that require Multi-Objective Optimization, such as calibrations and alignments or optimization of complex machine learning pipelines.

Acknowledgments

C.F., K.S. and H.N. were supported by the Office of Nuclear Physics of the U.S. Department of Energy under Grant Contract No. DE-SC0024625. C.P. and K.N. were supported by the Office of Nuclear Physics of the U.S. Department of Energy under Grant Contract No. DE-SC0024478. M.D. was supported by the U.S. Department of Energy Office of Science, Office of Nuclear Physics contract number DE-AC05-06OR23177, under which Jefferson Science Associates, LLC operates Jefferson Lab. T.H. was supported by DOE U.S. Department of Energy Office of Science, Office of Nuclear Physics contract number DE-SC-0024691. This work was supported by the U.S. Department of Energy, Office of Science, under Brookhaven Science Associates (BSA) contract number DE-SC0012704.

References

- [1] R. Ent, *Electron Ion Collider*, in the proceedings of the *JLUO Meeting on NSAC Long Range Plan*, 8 September 2022, <https://indico.jlab.org/event/589/contributions/10644/attachments/8403/12002/EICTJLLUOLRP2022.pptx>.
- [2] E. Cisbani et al., *AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case*, *2020 JINST* **15** P05009 [[arXiv:1911.05797](https://arxiv.org/abs/1911.05797)].
- [3] C. Fanelli et al., *AI-assisted optimization of the ECCE tracking system at the Electron Ion Collider*, *Nucl. Instrum. Meth. A* **1047** (2023) 167748 [[arXiv:2205.09185](https://arxiv.org/abs/2205.09185)].
- [4] *Interactive visualization of the Pareto front in a GEANT4-based design for ECCE*, <https://ai4eicdetopt.pythonanywhere.com/>.
- [5] J.C. Bernauer et al., *Scientific computing plan for the ECCE detector at the Electron Ion Collider*, *Nucl. Instrum. Meth. A* **1047** (2023) 167859 [[arXiv:2205.08607](https://arxiv.org/abs/2205.08607)].
- [6] P.I. Frazier, *A Tutorial on Bayesian Optimization*, [arXiv:1807.02811](https://arxiv.org/abs/1807.02811).
- [7] K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, *A fast and elitist multiobjective genetic algorithm: NSGA-II*, *IEEE Trans. Evol. Comput.* **6** (2002) 182.
- [8] GEANT4 collaboration, *Geant4 — a simulation toolkit*, *Nucl. Instrum. Meth. A* **506** (2003) 250.
- [9] M. Balandat et al., *BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization*, [arXiv:1910.06403](https://arxiv.org/abs/1910.06403).
- [10] E. Bakshy, M. Balandat and K. Kashin, *Open-sourcing Ax and BoTorch: New AI tools for Adaptive Experimentation*, <https://ai.facebook.com/blog/open-sourcing-ax-and-botorch-new-ai-tools-for-adaptive-experimentation/>.
- [11] S. Daulton, M. Balandat and E. Bakshy, *Parallel Bayesian Optimization of Multiple Noisy Objectives with Expected Hypervolume Improvement*, [arXiv:2105.08195](https://arxiv.org/abs/2105.08195).

- [12] *High-Dimensional Bayesian Optimization with SAASBO*, Ax tutorial, <https://ax.dev/versions/0.2.9/tutorials/saasbo.html>.
- [13] D. Eriksson et al., *Scalable Global Optimization via Local Bayesian Optimization*, in the proceedings of the 33rd International Conference on Neural Information Processing Systems (NIPS'19), Red Hook, NY, U.S.A. (2019), p. 5496–5507 [PDF] [arXiv:1910.01739].
- [14] T. Maeno et al., *PanDA: Production and Distributed Analysis System*, *Comput. Softw. Big Sci.* **8** (2024) 4.
- [15] T. Šimko et al., *REANA: A System for Reusable Research Data Analyses*, *EPJ Web Conf.* **214** (2019) 06034.
- [16] D. Lawrence, *EIC Software Overview*, *EPJ Web Conf.* **295** (2024) 03011.
- [17] K. Deb, A. Sinha and S. Kukkonen, *Multi-objective test problems, linkages, and evolutionary methodologies*, in the proceedings of the 8th annual conference on Genetic and evolutionary computation, Seattle, WA, U.S.A. (2006), p. 1141–1148, <https://www.egr.msu.edu/~kdeb/papers/k2006001.pdf>.
- [18] *Joblib: running Python functions as pipeline jobs*, documentation, <https://joblib.readthedocs.io/en/stable/>.
- [19] A.B. Yoo, M.A. Jette and M. Grondona, *SLURM: Simple Linux Utility for Resource Management*, in the proceedings of the Workshop on job scheduling strategies for parallel processing, Lecture Notes in Computer Science, vol. 2862, Springer (2003), p. 44–60 [DOI:10.1007/10968987_3].
- [20] D.K. Brown, T.M. Musyoka, D.L. Penkler and T. Bishop, *JMS: A workflow management system and web-based cluster front-end for the Torque resource manager*, arXiv:1501.06907.
- [21] R.L. Henderson, *Job scheduling under the portable batch system*, in the proceedings of the Workshop on Job Scheduling Strategies for Parallel Processing, Lecture Notes in Computer Science, vol. 949, Springer (1995), p. 279–294 [DOI:10.1007/3-540-60153-8_34].
- [22] S. Vallarino et al., *Prototype of a dual-radiator RICH detector for the Electron-Ion Collider*, *Nucl. Instrum. Meth. A* **1058** (2024) 168834.
- [23] A. Jentsch, *EIC Far Forward Detector Design*, APS April Meeting Abstracts **2022** (2022) G04.003.
- [24] R. Abdul Khalek et al., *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report*, *Nucl. Phys. A* **1026** (2022) 122447 [arXiv:2103.05419].
- [25] A. Jentsch, *Physics Opportunities in the Far-forward Region at the Future Electron-Ion Collider*, *Acta Phys. Polon. Supp.* **16** (2023) 7.
- [26] *Efficient Multi-Objective Neural Architecture Search with Ax*, (2022), <https://pytorch.org/blog/effective-multi-objective-nueral-architecture/#fully-automated-multi-objective-nas-with-ax>.
- [27] *The open source frontend for GitBook doc sites*, <https://github.com/GitbookIO/gitbook>.
- [28] *Jupyter Book*, <https://jupyterbook.org/en/stable/intro.html>.
- [29] K. Suresh, N. Kackar, L. Schleck and C. Fanelli, *Towards a RAG-based Summarization Agent for the Electron-Ion Collider*, arXiv:2403.15729.
- [30] *Multi-Objective Optimization Dashboard — AID(2)E Closure Tests Performance*, <https://wandb.ai/phys-meets-ml/AID2E-Closure-1?workspace=user-karthik18495>.
- [31] C. Fanelli, *AI for Detector Design NNPSS-2022*, 11–22 July 2022, <https://cfteach.github.io/nnpss/intro.html>.
- [32] C. Fanelli, *Machine Learning for Nuclear Physics and the Electron Ion Collider*, in the proceedings of the 38th Annual Hampton University Graduate Summer Program at Jefferson Lab Newport News, VA, U.S.A., 30 May–16 June 2023, <https://cfteach.github.io/HUGS23/intro.html>.
- [33] AID(2)E collaboration, *AID(2)E Repository*, <https://github.com/aid2e>.