AI-assisted Optimization of the ECCE Tracking System

J. K. Adkins[35], Y. Akiba[52], A. Albataineh[68], M. Amaryan[46], I. C. Arsene[72], J. Bae[60], X. Bai[77], M. Bashkanov[86], R. Bellwied[66], F. Benmokhtar[14], J. C. Bernauer[54,55,56], F. Bock[48], W. Boeglin[16], M. Borysova[82], E. Brash[10], P. Brindza[27], W. J. Briscoe[20], M. Brooks[31], S. Bueltmann[46], M. H. S. Bukhari[26], A. Bylinkin[68], R. Capobianco[65], W.-C. Chang[2], Y. Cheon[58], K. Chen[7], K.-F. Chen[45], K.-Y. Cheng[39], M. Chiu[4], T. Chujo[75], Z. Citron[1], E. Cline[54,55], E. Cohen[43], T. Cormier[48], Y. Corrales Morales[31], C. Cotton[77], C. Crawford[69], S. Creekmore[48], C.Cuevas[27], J. Cunningham[48], G. David[4], C. T. Dean[31], M. Demarteau[48], S. Diehl[65], N. Doshita[84], R. Dupré[23], J. M. Durham[31], R. Dzhygadlo[19], R. Ehlers[48], L. El Fassi[37], A. Emmert[77], R. Ent[27], C. Fanelli[36], R. Fatemi[69], S. Fegan[86], M. Finger[8], M. Finger Jr.[8], J. Frantz[47], M. Friedman[22], I. Friscic[87], D. Gangadharan[66], S. Gardner[18], K. Gates[18], F. Geurts[51], R. Gilman[53], D. Glazier[18], E. Glimos[48], Y. Goto[52], N. Grau[3], S. V. Greene[78], A. Q. Guo[24], L. Guo[16], S. K. Ha[85], J. Haggerty[4], T. Hayward[65], X. He[17], O. Hen[36], D. W. Higinbotham[27], M. Hoballah[23], P.-h. J. Hsu[44], J. Huang[4], G. Huber[73], A. Hutson[66], K. Y. Hwang[85], C. Hyde[46], M. Inaba[63], T. Iwata[84], H.-S. Jo[30], K. Joo[65], N. Kalantarians[80], K. Kawade[59], S. Kay[73], A. Kim[65], B. Kim[60], C. Kim[50], M. Kim[52], Y. Kim[50], Y. Kim[58], E. Kistenev[4], V. Klimenko[65], S. H. Ko[57], I. Korover[36], W. Korsch[69], G. Krintiras[68], S. Kuhn[46], C.-M. Kuo[39], T. Kutz[36], J. Lajoie[25], D. Lawrence[27], S. Lebedev[25], J. S. H. Lee[57], S. W. Lee[30], Y.-J. Lee[36], W. Li[51], W. Li[54,55,83], X. Li[9], X. Li[31], Y. T. Liang[24], S. Lim[50], C.-h. Lin[2], D. X. Lin[24], K. Liu[31], M. X. Liu[31], K. Livingston[18], N. Liyanage[77], W. J. Llope[81], C. Loizides[48], E. Long[71], R.-S. Lu[45], Z. Lu[9], W. Lynch[86], D. Marchand[23], M. Marcisovsky[13], P. Markowitz[16], P. McGaughey[31], M. Mihovilovic[70], R. G. Milner[36], A. Milov[82], Y. Miyachi[84], P. Monaghan[10], R. Montgomery[18], D. Morrison[4], C. Munoz Camacho[23], M. Murray[68], K. Nagai[31], J. Nagle[64], I. Nakagawa[52], C. Nattrass[76], D. Nguyen[27], S. Niccolai[23], R. Nouicer[4], G. Nukazuka[52], M. Nycz[77], V. A. Okorokov[42], S. Orešić[73], J.D. Osborn[48], C. O'Shaughnessy[31], S. Paganis[45], Z Papandreou[73], S. Pate[41], M. Patel[25], C. Paus[36], G. Penman[18], M. G. Perdekamp[67], D. V. Perepelitsa[64], H. Periera da Costa[31], K. Peters[19], W. Phelps[10], E. Piasetzky[61], C. Pinkenburg[4], I. Prochazka[8], T. Protzman[33], M. Purschke[4], J. Putschke[81], J. R. Pybus[36], R. Rajput-Ghoshal[27], J. Rasson[48], B. Raue[16], K. Read[48], K. Røed[72], R. Reed[33], J. Reinhold[16], E. L. Renner[31], J. Richards[65], C. Riedl[67], T. Rinn[4], J. Roche[47], G. M. Roland[36], G. Ron[22], M. Rosati[25], C. Royon[68], J. Ryu[50], S. Salur[53], N. Santiesteban[36], R. Santos[65], M. Sarsour[17], J. Schambach[48], A. Schmidt[20], N. Schmidt[48], C. Schwarz[19], J. Schwiening[19], R. Seidl[52], A. Sickles[67], P. Simmerling[65], S. Sirca[70], D. Sharma[17], Z. Shi[31], T.-A. Shibata[40], C.-W. Shih[39], S. Shimizu[52], U. Shrestha[65], K. Slifer[71], K. Smith[31], R. Soltz[34], W. Sondheim[31], J. Song[9], J. Song[50], I. I. Strakovsky[20], P. Steinberg[4], J. Stevens[83], J. Strube[49], P. Sun[9], X. Sun[7], K. Suresh[73], W.-C. Tang[39], S. Tapia Araya[25], S. Tarafdar[78], L. Teodorescu[5], A. Timmins[66], L. Tomasek[13], N. Trotta[65], T. S. Tveter[72], E. Umaka[25], A. Usman[73], H. W. van Hecke[31], J. Velkovska[78], E. Voutier[23], P.K. Wang[23], Q. Wang[68], Y. Wang[7], Y. Wang[62], D. P. Watts[86], L. Weinstein[46], M. Williams[36], C.-P. Wong[31], L. Wood[49], M. H. Wood[6], C. Woody[4], B. Wyslouch[36], Z. Xiao[62], Y. Yamazaki[29], Y. Yang[38], Z. Ye[62], H. D. Yoo[85], M. Yurov[31], N. Zachariou[86], W.A. Zajc[11], J. Zhang[77], Y. Zhang[62], Y. X. Zhao[24], X. Zheng[77], P. Zhuang[62]

[1]*A. Alikhanyan National Laboratory, Yerevan, Armenia*
[2]*Institute of Physics, Academia Sinica, Taipei, Taiwan*
[3]*Augustana University, , Sioux Falls, , SD, USA*
[4]*Brookhaven National Laboratory, , Upton, 11973, NY, USA*
[5]*Brunel University London, , Uxbridge, , UK*
[6]*Canisius College, , Buffalo, , NY, USA*
[7]*Central China Normal University, Wuhan, China*
[8]*Charles University, Prague, Czech Republic*
[9]*China Institute of Atomic Energy, Fangshan, Beijing, China*
[10]*Christopher Newport University, , Newport News, , VA, USA*
[11]*Columbia University, , New York, , NY, USA*
[12]*Catholic University of America, 620 Michigan Ave., Washington DC, 20064, USA*
[13]*Czech Technical University, , Prague, Czech Republic*
[14]*Duquesne University, , Pittsburgh, , PA, USA*
[15]*Duke University, , , , NC, USA*
[16]*Florida International University, , Miami, , FL, USA*
[17]*Georgia State University, , Atlanta, , GA, USA*
[18]*University of Glasgow, , Glasgow, , UK*
[19]*GSI Helmholtzzentrum fuer Schwerionenforschung, , Darmstadt, , Germany*
[20]*The George Washington University, , Washington, DC, , USA*
[21]*Hampton University, , Hampton, , VA, USA*
[22]*Hebrew University, , Jerusalem, , Isreal*
[23]*Universite Paris-Saclay, CNRS/IN2P3, IJCLab, Orsay, France*
[24]*Chinese Academy of Sciences, , Lanzhou, , , China*
[25]*Iowa State University, , , , IA, USA*
[26]*Jazan University, Jazan, Sadui Arabia*
[27]*Thomas Jefferson National Accelerator Facility, 12000 Jefferson Ave., Newport News, 24450, VA, USA*
[28]*James Madison University, , , , VA, USA*

[29]Kobe University, Kobe, Japan
[30]Kyungpook National University, Daegu, Republic of Korea
[31]Los Alamos National Laboratory, , , , NM, USA
[32]Lawrence Berkeley National Lab, , Berkeley, , , USA
[33]Lehigh University, , Bethlehem, , PA, USA
[34]Lawrence Livermore National Laboratory, , Livermore, , CA, USA
[35]Morehead State University, Morehad, KY, USA, , Morehead, , KY, USA
[36]Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, 02139, MA, USA
[37]Mississippi State University, , Mississippi State, , MS, USA
[38] National Cheng Kung University, Tainan, Taiwan, , , , ,
[39]National Central University, Chungli, Taiwan
[40]Nihon University, Tokyo, Japan
[41]New Mexico State University, , Las Cruces, , NM, USA
[42]National Research Nuclear University MEPhI, Moscow, Russian Federation
[43]Nuclear Research Center - Negev, Beer-Sheva, Isreal
[44]National Tsing Hua University, Hsinchu, Taiwan
[45]National Taiwan University, Taipei, Taiwan
[46]Old Dominion University, , Norfolk, , VA, USA
[47]Ohio University, , Athens, , OH, USA
[48]Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, 37831, TN, USA
[49]Pacific Northwest National Laboratory, , Richland, , WA, USA
[50]Pusan National Univeristy, Busan, Republic of Korea
[51]Rice University, P.O. Box 1892, Houston, 77251, TX, USA
[52]RIKEN Nishina Center, Wako, Saitama, Japan
[53]The State University of New Jersey, , Piscataway, , NJ, USA
[54]Center for Frontiers in Nuclear Science, Stony Brook, 11794, NY, USA
[55]Stony Brook University, 100 Nicolis Rd., Stony Brook, 11794, NY, USA
[56]RIKEN BNL Research Center, , Upton, , NY, USA
[57]Seoul National University, Seoul, Republic of Korea
[58]Sejong University, Seoul, Republic of Korea
[59]Shinshu University, Matsumoto, Nagano, Japan
[60]Sungkyunkwan University, Suwon, Republic of Korea
[61]Tel Aviv University, P.O. Box 39040, Tel Aviv, 6997801, Israel
[62]Tsinghua University, Beijing, China
[63]Tsukuba University of Technology, Tsukuba, Ibaraki, Japan
[64]University of Colorado Boulder, Boulder, CO, USA, , Boulder, , CO, USA
[65]University of Connecticut, , Boulder, Storrs, , CT, USA
[66]University of Houston, , Houston, , TX, USA
[67]University of Illinois, , Urbana, , IL, USA
[68]Unviersity of Kansas, 1450 Jayhawk Blvd., Lawrence, 66045, KS, USA
[69]University of Kentucky, , Lexington, 40506, KY, USA
[70]University of Ljubljana, Ljubljana, Slovenia, , Ljubljana, , , Slovenia
[71]University of New Hampshire, , Durham, , NH, USA
[72]University of Oslo, Oslo, Norway
[73] University of Regina, , Regina, , SK, Canada
[74]University of Seoul, Seoul, Republic of Korea
[75]University of Tsukuba, Tsukuba, Japan
[76]University of Tennessee, , Knoxville, 37996, TN, USA
[77]University of Virginia, , Charlottesville, , VA, USA
[78]Vanderbilt University, , Nashville, , TN, USA
[79]Virginia Tech, , Blacksburg, , VA, USA
[80]Virginia Union University, , Richmond, , VA, USA
[81]Wayne State University, , Detroit, , MI, USA
[82]Weizmann Institute of Science, Rehovot, Israel
[83]The College of William and Mary, Williamsburg, VA, USA
[84]Yamagata University, Yamagata, Japan
[85]Yarmouk University, Irbid, Jordan
[86]Yonsei University, Seoul, Republic of Korea
[87]University of York, York, UK
[88]University of Zagreb, Zagreb, Croatia

Abstract

The Electron-Ion Collider (EIC) is a cutting-edge accelerator facility proposed to study the nature of the "glue" that binds the building blocks of the visible matter in the universe. The proposed experiment will be realized at Brookhaven National Laboratory in approximately 10 years from now, with the detector design and R&D currently ongoing. Noticeably EIC can be one of the first facilities to leverage on Artificial Intelligence (AI) during the design and R&D phases. Optimizing the design of its tracker is

of crucial importance for the EIC Comprehensive Chromodynamics Experiment (ECCE), a protocollaboration that is proposing a detector design based on a 1.5T solenoid. The optimization is an essential part of the design and R&D and ECCE includes in its structure a working group dedicated to AI-based applications for the EIC detector. In this note we describe a comprehensive study in detector design of large-scale experiments using AI that has been accomplished during the detector proposal. This work required a complex parametrization of the simulated detector system. Our approach deals with an optimization problem in a multidimensional design space driven by multiple objectives that encode the detector performance, while satisfying several mechanical constraints. We describe our strategy and show results obtained for the ECCE tracking system. The AI-assisted design is agnostic to the simulation framework and can be extended after the detector proposal to other sub-detectors or system of sub-detectors to optimize the performance of the EIC detector.

Keywords: ECCE, Electron Ion Collider, Tracking, Artificial Intelligence, Evolutionary, Bayesian Optimization.

Contents

1 Introduction **3**

2 AI-assisted Detector Design **3**

3 ECCE Tracking System Simulation **5**

4 Analysis Workflow **6**

5 Computing Resources **10**

6 Summary **11**

Appendix A Details on Parametrization **12**

Appendix B Baseline and R&D designs **13**

1. Introduction

The Electron Ion Collider (EIC) [1] is a future cutting-edge discovery machine that will unlock the secrets of the gluonic force binding the building blocks of the visibile matter in the universe. The EIC will consist of two intersecting accelerators, one producing an intense beam of electrons, the other a beam of protons or heavier atomic nuclei, and it will be the only electron-nuclei collider operating in the world. The EIC Comprehensive Chromodynamics Experiment (ECCE) [2] is an international proto-collaboration assembled around the idea of developing an EIC detector envisioned to offer full energy coverage and an optimized far forward detection region. ECCE is currently investigating a detector design based on the existing BABAR 1.5T ready for the beginning of the EIC accelerator operation. More details on the ECCE detector design and what is described in the following can be found in [3].

The EIC is an integrated detector that extends for about 40 meters which includes the central detector built around the interaction point and the far-forward and far-backward regions [1]. To fulfill the physics goals of the EIC, the central detector needs to be hermetic and provide good particle identification (PID) over a large phase space. The central detector itself consists of multiple sub-detectors: a tracking system made by inner and outer tracker stations allows to reconstruct the charged particles moving in the magnetic field; a system of PID sub-detectors will cover the barrel and the e-going and h-going directions; electromagnetic and hadronic calorimeters are used to detect electromagnetic and hadronic showers and complete information on the particle flow which is essential for certain event topologies, *e.g.*, those containing jets. The optimization of the EIC detector is of utmost importance during the design and R&D phases [4].

As outlined in [1], Artificial Intelligence (AI) can provide dedicated strategies for complex combinatorial searches and can handle multi-objective problems characterized by a multidimensional design space, allowing to capture hidden correlations among the design parameters. ECCE included these techniques in the design workflow during the detector proposal. This AI-assisted design strategy can be used at first to steer the design. After the technology choice is selected (based also on the important insights provided by AI), it can be further utilized to eventually fine-tune the detector parameters. During the ECCE detector proposal, the design of the detector has been characterized by a continued optimization process that evolved in time. In the following we describe the role played by the AI-assisted workflow and highlight this as a novel aspect introduced during the design process of an experiment.

The article is structured as follows: in Sec. 2 we provide an overview of design optimization and describe the AI-assisted strategy; in Sec. 3 we describe the software stack utilized in this work to which AI is coupled for the optimization and we will introduce the ECCE tracker; in Sec. 4 we describe the implemented pipeline that supports the design optimization which results in a sequential strategy fostering the interplay between the different working groups in a post hoc decision making process; in Sec. 5 we present perspectives and planned activities.

The ECCE detector at EIC will be one of the first examples of detectors that will be realized leveraging AI during the design and R&D phases.

2. AI-assisted Detector Design

Detector optimization with AI is an essential part of the R&D and design process and it is anticipated to continue in the months following the detector proposal towards CD-2 and CD-3. Optimizing the design of large scale detectors such as ECCE which are made of multiple sub-detector systems is a complex problem. Each sub-detector system is in fact characterized by a multi-dimensional design parameter space (describing, *e.g.*, geometry, mechanics, optics, etc). In addition

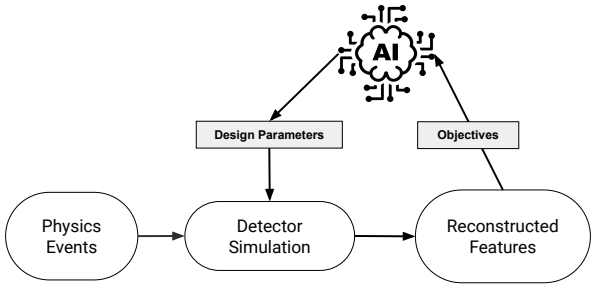

Figure 1: **Workflow of detector design assisted by AI:** physics events are injected in a detector characterized by some given design parameters. Reconstructed events are analyzed and figures of merit are quantified and passed to some AI-based strategy, which in turn suggests the next design point to observe in this sequential approach; notice that AI can also intervene in the simulation and reconstruction steps.

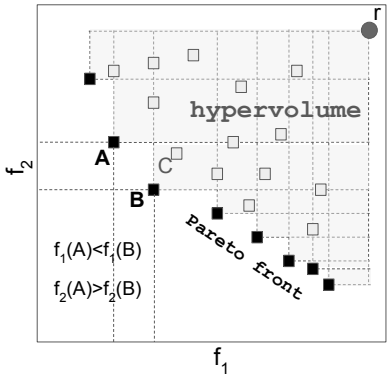

Figure 2: **Example of *Pareto frontier* in a two-dimensional objective space:** The point C is not on the frontier and is in fact dominated by both point A and point B. All the other points which are dominated by the Pareto frontier and that satisfy any constraints in the optimization problem are called *feasible* solutions; The hypervolume is used as a metric for convergence, and is calculated with respect to a reference point r.

to that, detector simulations are typically computationally intensive, and rely on advanced simulation platforms used in our community like Geant [5] to simulate the interaction of radiation with matter. On top of that, other computationally expensive steps are present along the data reconstruction and analysis pipeline. The software stack that is utilized in the detector design process involves in fact three main steps: (i) generation of events, (ii) detector simulations and (iii) reconstruction and analysis.

As pointed out in [6], the above bottlenecks make cumbersome the generation and exploration of mutliple design points. This in turn represents an obstacle for deep learning (DL)-based approaches to learn the mapping between the design space and the functional space (as in, *e.g.*, [7, 8, 9]) which could facilitate the identification of optimal design points. It should be also noticed that fast simulations with DL can in principle reduce the most CPU-intensive parts of the simulation and provide accurate results (see, *e.g.*, ATLAS [10]) still for detector design several design points need to be produced with Geant4 before injection in any DL architecture. Similar considerations can be made in deploying DL for reconstruction during the design optimization process.

In this context, a workflow for detector design that has gained popularity in recent years [11] is represented by the scheme is shown in Fig. 1. It consists in a sequential AI-based strategy that collects information associated to previously generated design points, *i.e.* in the form of figures of merit (called *objectives* in the following) that quantify the goodness of the design, and based on that suggests new promising design points to simulate at the next iteration.

During the detector proposal the AI Working Group has accomplished a continued multi-objective optimization (MOO) of the ECCE Tracker design. Our approach deals with a complex optimization in a multidimensional design space driven by multiple objectives that encode the detector performance, while satisfying several mechanical constraints. This framework has been developed in such a way to be easily extended to other sub-detector or to a system of sub-detectors, provided a viable parametrization of the simulation.

The definition of a generic MOO problem can be formulated as follows:

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

where one has M objective functions f_m to optimize (*e.g.*, detector resolution, efficiency, 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 design) with lower and upper bounds on each dimension. In solving these problems, one can come up with a set of *non-dominated* or *trade-off* solutions [12], popularly known as Pareto-optimal solutions (see also Fig. 2).

In this setting, we used a recently developed framework for MOO called pymoo [13] which supports evolutionary MOO algorithms such as Non-Dominated Sorting Genetic Algorithm (or NSGA-II, [14]). The rationale behind this choice instead of, *e.g.*, principled-approaches such as Bayesian Optimization as in [11], emanates directly from the needs we had at the time of the detector proposal, like the capability to quickly implement and run multiple parallel optimization pipelines implementing different technology choices and the possibility of dealing with non-differentiable objectives at this exploratory stage.

The NSGA workflow is described in Fig. 3. The main features of NSGA-II are (i) the usage of an elitist principle, (ii) an explicit diversity preserving mechanism, and (iii) ability of determining non-dominated solutions. The latter feature is of great importance for problems where objectives are of conflict to each other, that is: an improved performance in an objective results in worse performance in another objective.

The pymoo framework also supports other MOO approaches and a full list is documented in [13]. For our purposes, we also tested NSGA-III which is suitable for the optimization of large

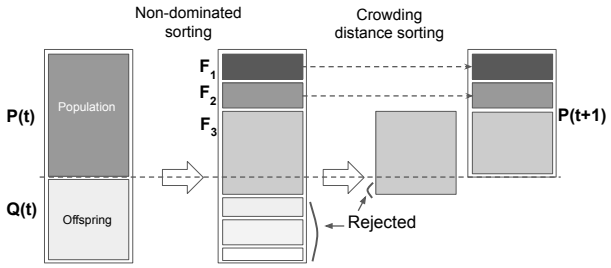

Figure 3: **The NSGA Workflow:** At time t, an offspring is created through a genetic algorithm [15] from an $N-$sized population of design points. The two populations are combined into a population R_t, which is classified into different non-dominated classes F_i, starting from the first front F_1. To restore the initial size of the population, the augmented space of solutions is trimmed. A metric called crowding distance is used to reject solutions and eventually provide an updated population of size N at time $t+1$.

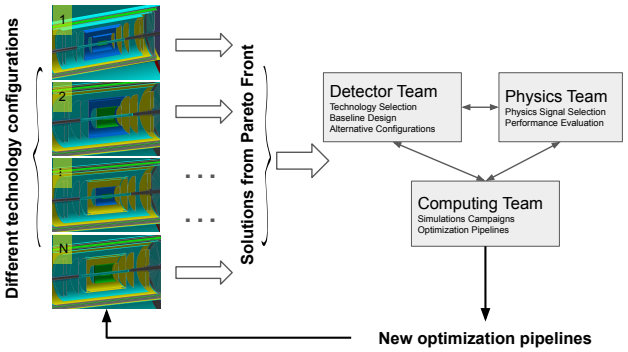

Figure 4: **Flowchart of continued optimization during proposal:** AI assists the design optimization process by providing insights and capturing hidden correlations among the design parameters. This has been used during the entire detector proposal process to steer the design. At a given instant in time, N alternative candidate configurations are studied. For each we create an optimization pipeline, which results in a Pareto front of solutions. This new information helps steering the design: some configurations are rejected, while other ones (also dubbed 'new references') are identified to potentially improve the design. New optimization pipelines are defined inspired by the new results and the process is iterated. During the design process, AI propelled the fundamental interplay between the ECCE Teams working on Physics, Detector and Computing.

number of objectives [16].[1]

As described in Sec. 3, the simulation and reconstruction rely on the Fun4All framework [17]. During the design optimization process of the tracking system, we used full Geant4 simulations of the entire ECCE detector. In Sec. 4 the reader can find a detailed description of the different parametrizations that have been utilized for several optimization pipelines of the inner tracker design optimization. Eventually the parametrization has been extended to include also the support structure in the design optimization problem. Notice that overlaps in the design are checked before and during the optimization and are excluded by the constraints and ranges of the parameters.

At least three objective functions characterizing the tracker performance (intrinsic detector response such as resolutions and efficiency) have been optimized simultaneously. The different designs have been optimized using particle gun samples of pions; after the optimization we used Pythia [18] samples to evaluate the improvement in the reconstruction of physics events (see Sec. 4.3).

AI has played a crucial role in helping to choose a combination of technologies for the inner tracker and has been used as input to multiple iterations of the ECCE tracker design, which led to the current tracker layout. This has been the result of a continued optimization process that evolved in time and required the interplay between the ECCE teams: results are validated by looking at figures of merit that do not enter as objective functions in the optimization process (more details can be found in Sec. Appendix B); the decision making is left post hoc and discussed among the Computing, Detector and Physics teams. Fig. 4 displays a flowchart describing this continued optimization process that integrates AI and the interplay between the ECCE teams that took place during the detector proposal.

AI provided insights in the design of the ECCE tracker by showing each time by how much the tracking system with different technology choices could be optimized. Ultimately this continued AI-assisted optimization led to a projective design after having extended the parametrized design to include the support structure of the inner tracker. The latter represents an ongoing R&D project that is discussed in the next sections.

[1] For ≥ 4 objectives, NSGA-III is expected to perform better than NSGA-II.

3. ECCE Tracking System Simulation

The simulation and detector response shown in this document is based on Geant4 [19] and was carried out using the Fun4All framework [17].

The optimization pipelines are based on particle gun samples of pions, where we used π^- and tested that the performance with π^+ were consistent. Performance in the electron-going direction has been also checked post-hoc with particle gun samples of electrons. The improved performance is further validated with physics analyses, using the datasets generated during the ECCE simulation campaigns; in Sec. 4 we show in particular results based on semi-inclusive deep inelastic scattering (SIDIS) events.

The ECCE tracking detector [20], consists in different layers in the barrel and the two end-caps, and is tightly integrated with the PID detectors:

(i) The silicon vertex/tracking detector is an ALICE ITS-3 type high precision cylindrical/disk vertex tracker [21, 22]) nased on the new Monolithic Active Pixel Sensor (MAPS); the barrel detector consists of 5 MAPS layers; the silicon hadron endcap consists of 5 MAPS disks; whereas the silicon electron endcap has 4 MAPS disks.

(ii) A gas tracking system is based on μRwell technology, that is a single-stage amplification Micro Pattern Gaseous Detector (MPGD) that is a derivative of the Gas Electron Multiplier (GEM) technology. In ECCE μRwell layers will form three barrel tracking layers further out from the beam-pipe than the silicon layers; namely, two inner barrel μRwell layers and a single outer barrel μRwell layer. All μRwell detectors will have

2D strip based readout. The strip pitch for all 3 layers will be 400 μm.

(iii) The tracking system is completed by AC-LGAD-based time of flight (TOF) detectors providing additional hit information for track reconstruction as well. In the central region a TOF (dubbed CTTL) is placed behind the high-perofrmance DIRC (hpDIRC); in the hadron going side a TOF (dubbed FTTL) is placed before the dual RICH (dRICH) and a μRwell placed after the dRICH; in the electron going direction a μRwell layer is placed before the modular RICH (mRICH), which is followed by a TOF later (dubbed ETTL).

The ECCE simulation of tracking system is shown in Fig. 5 (left). Given the importance of the service structure in the tracking detector, the reduction of the impact of readout and services on tracking resolution is subject of ongoing project R&D. A parallel setup for the tracking system based on the same detector technology is maintained in a branch of the ECCE software repository. This is a projective design characterized by a conical support structure with a polar angle of 25.5 degrees on both endcaps, as shown in Fig. 5 (right).

In the following sections, the reader can find more details on the implementation of the optimization pipelines and utilized computing resources.

4. Analysis Workflow

The optimization of the ECCE-tracking system has been characterized by two main phases during which the sub-detectors composing the tracker evolved into more advanced simulations.

Phase-I optimization. Phase-I corresponds to a timeline between June-2021 to Sept-2021.[2] The Geant-4 implementation of the detectors were at first simplified, *e.g.*, the support cone was modelled as a simple conical structure with simplified material made of Aluminium. The optimization pipelines consisted of symmetric arrangement of detectors in the e-going and h-going directions (5 disks on each side). The DIRC detector for PID in the barrel region was modelled in a simplified way. ACLGAD-based TOF detectors were modelled as simplified silicon disks at first; more detailed simulations of the outer trackers have been developed during this phase. The optimization pipelines have been performed including various combinations of detector technologies for the inner trackers. At the end of this phase, a decision on the choice of the barrel technology and the disk technologies was made taking advantage of results provided by AI.

Phase-II optimization. Phase-II corresponds to optimization pipelines that run from Sept-2021 to Nov-2021. These pipelines had a more realistic implementation of the support structure incorporating cabling, support carbon fiber, cooling system etc. More detailed simulation of the PID Detectors (DIRC and dRICH) have been integrated as well as fine-grained simulations of TTL layers (CTTL, ETTL, FTTL). More stringent engineering constraints have been considered like the sensor size for MAPS detector (ITS3). This phase also considers an asymmetric arrangement of the detectors in the endcap regions, with a maximum of 4 EST disks in the e-going end-cap and 5 FST disks in the h-going end-cap. Due to the asymmetric arrangement of detectors, the angle subtended by detectors in the two end cap regions could be changed. An optimization was carried out with asymmetric spatial arrangement of EST and FST disks. This eventually developed into the idea of a projective geometry in a pipeline that characterizes an ongoing R&D project for optimizing the design of the support structure.

The reader can find more details on the detector technology characterizing the tracker system in [3].

A detailed description of the most recent parametrization used for the detector proposal can be found in Appendix A. Such parametrization has been used for the optimization of the non-projective design, which had the inner tracker support structure fixed. As already mentioned, the parametrization has been modified to explore the possibility of including in the optimization the design of the support structure. This resulted in the R&D project always described in Appendix A; a schematic representation of this parametrization is displayed in Fig. A.11.

Fig. 5 shows a comparison of the ECCE reference non-projective design and the projective design from the ongoing R&D, both of which resulted from the AI-assisted procedure described in this paper.

4.1. Encoding of Design Criteria

Design criteria need to be encoded to steer the design during the optimization process. For each design point we need to compute the corresponding objectives f (*i.e.*, momentum resolution, angular resolution, and Kalman filter efficiency).

We will refer in the following only to more recent phase-II optimization. Similar considerations apply also for phase-I optimization. Phase-II has been characterized by two types of optimization pipelines: the first used a parametrization of the inner tracker during the optimization process and led to the ECCE tracker non-projective design; the second branched off the first as an independent R&D effort which included the parametrization of the support structure and led to a projective design.

More details on the two types of optimization pipelines can be found in the following tables: Table 1 describes the main hyperparameters; Table 2 describes the dimensionality of the optimization problem, in particular of the design space, the objective space and the constraints; Table 3 reports the the ranges of each design parameter[3]; Table 4 summarizes the constraints for both the non-projective and projective geometries. We also considered in our design a safe minimum distance between the disks of 10 cm and include a constraint on the difference between the outer and inner radii of each disk, namely R_{max} -

[2] Preliminary studies done between March-2021 to May-2021 are not reported here.

[3] The design points are normalized in the range [0-1], using a min-max scaler $x_i = x(x_{max} - x_{min}) + x_{min}$, where x_i is the normalized design point with a unnormalized design point x generated between the range $[x_{min}, x_{max}]$.

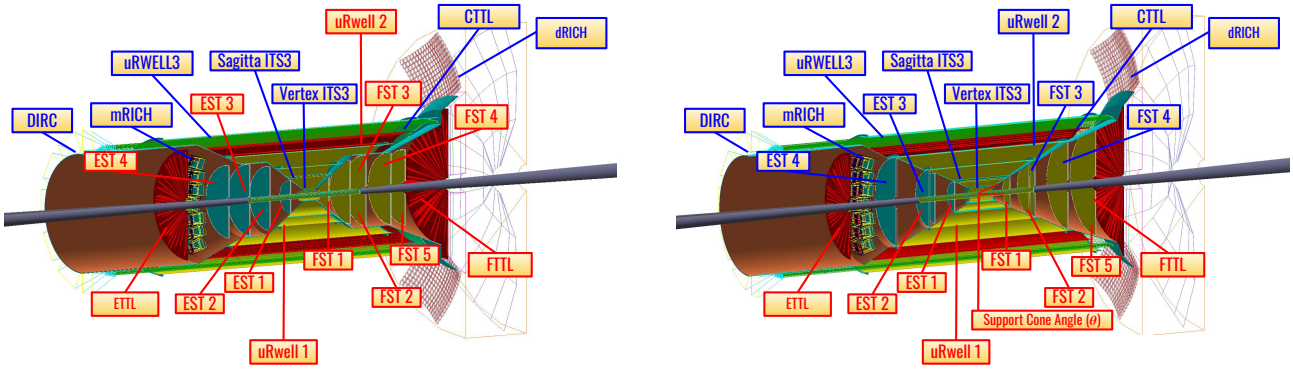

Figure 5: **Tracking and PID system in the non-projective (left) and the ongoing R&D projective (right) designs:** the two figures show the different geometry and parametrization of the ECCE non-projective design (left) and of the ongoing R&D projective design to optimize the support structure (right). Labels in red are the sub-detector systems that are optimized, while the labels in blue are the sub-detector systems that are kept fixed due to geometrical constraint. The non-projective geometry (left) is a result of an optimization done on the inner tracker layers (labeled in red) while keeping the support structure fixed, The angle made by the support structure to the IP is fixed at about 36.5°. The projective geometry (right) is the result of an ongoing project R&D to reduce the impact of readout and services on tracking resolution.

R_{min}, to be a multiple of the sensor cell size (17.8 mm × 30.0 mm), see Table 4. These constraints are common to the non-projective and the projective designs. For more details on the parametrizations the reader can refer to Appendix A and for a detailed comparison of the corresponding detector performance can refer to Appendix B. In what follows we provide a thorough description of the design criteria that have been encoded in the optimization strategy.

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

Table 1: **Summary of the hyperparameters of the design optimization:** The values reported in the table have been used during the optimization of the non-projective design of the ECCE tracker. For completeness and when they differ from the non-projective case, we also report in parentheses the values corresponding to the ongoing R&D project for the projective design of the support structure.

The objectives depend on the kinematics and are calculated in 5 main bins in pseudo-rapidity (η), *i.e.*, (i) -3.5 ≤ η < -2.0 (corresponding to electron going direction), (ii) -2.0 ≤ η < -1.0 (corresponding to transition region in electron going direction), (iii) -1 ≤ η < 1 (corresponding to the central barrel), (iv) 1 ≤ η < 2.0 (corresponding to transition region in hadron going direction) and (v) 2.0 ≤ η < 3.5 (corresponding to hadron going direction). The rationale behind this binning is a combination of different aspects: the correspondence with the binning in the EIC Yellow Report [1], the asymmetric arrangement of detectors in e-going and h-going directions and the

	design pars	objectives	constraints
#	4(11)	3 (4)	4(5)
ECCE tracker design (non-projective)	(Phase-I symmetric design) 5 disks z-location (hadron beam) and 6 barrel layers radii; (Phase-II asymmertic design) 3 disks z location (endcap region) 1 barrel layer	momentum resolution; angular resolution; Kalman filter efficiency; (pointing resolution)	(i) inner 2^{nd} vtx barrel radius ≤ 15 cm; (ii) outermost barrel radius < 51 cm; (iii) outermost z disk < 125 cm; (iv) 4^{th} barrel radius ≤ 45 cm; (v) Minimum distance between any 2 layers >= 1 cm (Additional constraints for Phase-II)
ECCE ongoing R&D design (projective)	Angle Tracking Support (θ), 2 TTL z-location, 1 uRwell-1 Radius, 3 FST disks, 2 EST disks (9 parameters);	momentum resolution; angular resolution; Kalman filter efficiency; angular resolution at dRICH; (pointing resolution)	(i) FST/EST Disks $(R_{Max} - R_{Min})\%(30.0\ mm) = 0$; (ii) FST/EST Disks $(R_{Max} - R_{Min})\%(17.8\ mm) = 0$; (iii) FST/EST $z_{n+1} - z_n \geq 10$; (iv) Length of vertex support structure ≤ 30 cm; (v) distance between any 2 μRwell barrel layers have to be ≥ 5 cm

Table 2: **Dimensionality of the optimization pipelines:** non-projective design and objective space. Following Eq. (1), constraints are also included. For each category, a description has been added. Values in parentheses are the largest ones utilized in the optimization pipelines. The Phase-II optimization had less number of parameters to optimize due to the fixed support cone structure. Notice that two additional constraints are implicitly taken into account, that is: (i) the inner radii of the disks in the endcap and the inner vertex layer in the barrel have to be compatible with the beam envelope dimensions; (ii) the barrel layers lengths and the outer radii of the disks have to be compatible with an Al-shell support structure provided by the LBNL design [23]. Conditions (i) and (ii) are provided by internal parametrization and reflected in the Geant4 design. Potential overlaps among modules are checked before and during the optimization.

ECCE design (non-projective)		
Design Parameter	**Reference value**	**Range**
μRWELL 1 (Inner) (r) Radius	44.2 cm	[17.0, 51.0 cm]
μRWELL 2 (Inner) (r) Radius	47.4 cm	[18.0, 51.0 cm]
EST 4 z position	-107.1 cm	[-110.0, -50.0 cm]
EST 3 z position	-80.0 cm	[-110.0, -40.0 cm]
EST 2 z position	-57.5 cm	[-80.0, -30.0 cm]
EST 1 z position	-35.0 cm	[-50.0, -20.0 cm]
FST 1 z position	35.0 cm	[20.0, 50.0 cm]
FST 2 z position	57.5 cm	[30.0, 80.0 cm]
FST 3 z position	80.0 cm	[40.0, 110.0 cm]
FST 4 z position	115.0 cm	[50.0, 125.0 cm]
FST 5 z position	125.0 cm	[60.0, 125.0 cm]
ECCE ongoing R&D (projective)		
Design Parameter	**Reference value**	**Range**
Angle (Support Cone)	25.5 °	[25.0°, 30.0°]
μRWELL 1 (Inner) Radius	33.12 cm	[25.0, 45.0 cm]
ETTL z position	-169.0 cm	[-171.0, -161.0 cm]
EST 2 z position	-80.05 cm	[45, 100 cm]
EST 1 z position	-33.2 cm	[35, 50 cm]
FST 1 z position	33.2 cm	[35, 50 cm]
FST 2 z position	79.85 cm	[45, 100 cm]
FST 5 z position	144 cm	[100, 150 cm]
FTTL z postion	182	[156, 183 cm]

Table 3: **Ranges of the design parameters:** the table summarises the design points that are optimized for the Non-projective and the projective case. The optimization range is also given for each parameter. The parameters corresponding to the Non-projective case were optimized sequentially over atleast three iterations with each iteration having a set of parameters fixed. For instance, The EST/FST disks were optimized during the first iteration assuming symmetric design, and for the subsequent iterations the EST disks were optimized further in the electron endcap region. Along with the design parameters the design comes with geometrical constraints too. These are implemented as strong and soft constraints in the Table 4.

	Constraint	**Description**
FST/EST Disks	$\sum_i \frac{(RMax_i - RMin_i)}{3.0} = \mathbb{N}$	MAPS sensor length is 3.0cm. Soft Constraint
FST/EST Disks	$\sum_i \frac{(RMax_i - RMin_i)}{1.8} = \mathbb{N}$	MAPS sensor width is 1.78cm. Soft Constraint
FST/EST Disks	$z_{n+1} - z_n >= 10.0$	Minimum distance between 2 consecutive disk. Strong constraint

Table 4: **Constraints in the design optimization:** the table summarises constraints for both for projective and non projective design problem. The constraints are from the FST/EST disks since they are made by tiling up MAPS pixels. The pixels will have fixed dimensions. Soft constraints are constraints that can be violated to a certain degree by the MOO and then one can quantify the degree of unfeasibility (see [13]). Also, minor adjustments on tiling up the pixels can be done post hoc optimization such that the detector geometry is realisable. Strong constraints penalises the MOO with high penalties since these constraints cannot be violated: *e.g.*, the distance between the disks cannot be closer than 10 cm for servicing requirement. The constraints in this table are common to all designs.

division in pseudo-rapidity between the barrel region and the endcap. Particular attention during the optimization and post hoc is given to the transition region between barrel and endcaps as well as at large $|\eta| \sim 3.5$ close to the beamline. Charged pions are generated uniformly in the phase-space that covers P $\in$ [0,20] GeV/c $\times$ $\eta \in$ (-3.5,3.5). Each bin in η is further subdivided in 15 bins in momentum P.. Studies have been performed to determine the number of tracks (N_{trk}). For each design point we utilize 120k charged pions.[4] This number ensures large enough statistics over the entire phase space and the stability of the fits in all of the bins of Eqs. (4).

In order to calculate the relative momentum (cf. Fig. B.12) and absolute angular resolution (cf. Fig B.13) we do fit the following objectives:

$$\frac{\Delta P}{P} = \frac{P_{reco} - P_{true}}{P_{true}} \qquad \Delta\Theta = \Theta_{reco} - \Theta_{true} \tag{2}$$

Following the definitions of Eq. (2), histograms of the relative momentum resolution and the absolute angular resolution are produced for each bin in η and P and the corresponding fits are calculated. Using single-gaussian (SG) fits (also utilized in the Yellow Report [1]) implies systematically better resolutions but worse reduced χ^2, therefore we decided to utilize during the optimization procedure double-gassian (DG) fits, as shown in Fig. 6. This provided a more robust fit strategy. The reduced χ^2 range with DG fits ranges from 1.2 to 2.8 at most, with the majority of the fits stable at lower values. The largest numbers correspond either to the transition between the barrel and endcaps where tracks cross more material in the non-projective design, or to large pseudo-rapidities, particularly close to the inner radii of the disks. By using SG fits, the reduced χ^2 values can be as large as 10-20 in the transition region. A detailed study comparing SG to DG fits has been done in Fig. B.12.

The final DG resolution has been defined as an average of the two σ's weighted by the relative areas of the two gaussians:[5]

$$\sigma_{DG} \sim \frac{A_1\sigma_1 + A_2\sigma_2}{A_1 + A_2} \tag{3}$$

The results obtained for the resolutions in each bin corresponding to each new design point are divided by the values corresponding to the baseline design, so that in each bin a ratio R is provided. Finally a weighted sum of these ratios is performed to build a global figure of merit (for both the relative momentum and the angular resolutions):

$$R(f) = \frac{1}{N_\eta} \sum_\eta \left(\frac{\sum_p w_{p,\eta} \cdot R(f)_{p,\eta}}{\sum_p w_{p,\eta}} \right) \tag{4}$$

[4] From phase I to phase II, the design became asymmetric in the two endcaps, therefore we needed to extend the η-coverage and increase the statistics. The momentum range has been reduced to [0,20] GeV/c to optimize the computing budget.

[5] A different definition could be based on the weighted average of the variances $\sigma^2_{1,2}$ to obtain the final variance σ^2_{DG}. This typically implied a few % relative difference on the final value of σ_{DG} which has been considered a negligible effect.

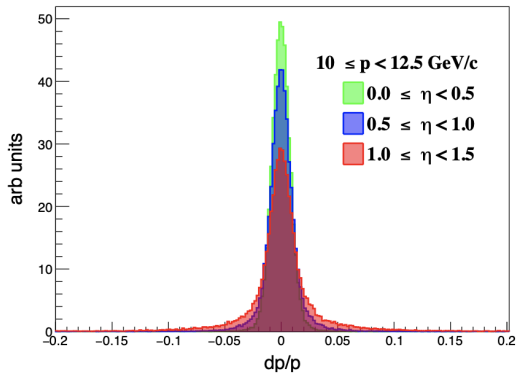

Figure 6: **Fit strategy:** a double-gaussian fit function is utilized to extract the resolutions. Such a fit function provided good reduced χ^2 and more stable extractions compared to single-gaussian fits. The resolution is obtained as an average of the two σ's weighted by the relative areas of the two gaussians according to Eq. (3). The figure represents the results corresponding to a particular bin in η and P.

where the objective function f is either the momentum or the angular resolution described by Eq. (2), and the weight $w_{p,\eta} = 1/\sigma^2(f_{p,\eta})$ is calculated in each η, p bin and it is proportional to the inverse of the variance corresponding to the objective functions f.

In addition to these objectives, we also included another objective function in the optimization problem. This is another global objective function corresponding to the fraction of tracks that are not reconstructed by the Kalman filter (KF), or equivalently the KF inefficiency:

$$R(\mathrm{KF}) = \frac{\mathrm{N(not\ \ reconstructed)}}{\mathrm{N(tot.\ tracks)}} \tag{5}$$

Following the above definitions, the design problem corresponds to minimizing the objective functions defined in Eqs. (4), (5): a weighted average ratio smaller than one corresponds to an improvement in performance compared to the baseline design; as for the KF inefficiency, we ideally want to find a design point that pushes this value as close as possible to zero.

4.2. Convergence and Performance at Pareto Front

We remind the reader that the Pareto front is the set of trade-off solutions of our problem. Fig. 7 shows the convergence plot obtained utilizing the hypervolume as metric in the objective space.[6] A petal diagram is used to visualize the values of three objectives corresponding to one of the solutions extracted from the Pareto front.

Checkpoints are created to store the NSGA-updated population of design points. A survey of the detector performance is created after each call to monitor potential anomaly behavior of the fits. The fitting procedure is quite stable, if an exception occurs the analysis has been automated to adjust the fitting parameters and ranges. In case of persistent anomalous behavior a flag is raised, the critical design point purged from the population and examined.

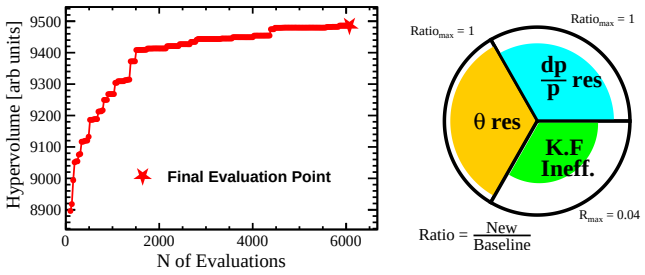

Figure 7: **Metric for convergence and performance at Pareto front:** (left) The hypervolume can be used as a metric for convergence. Checkpoints are created during the optimization and snapshots of the evolving designs are saved. (right) A petal diagram with the three objectives corresponding to one solution in the Pareto front. The momentum and angular resolutions are expressed as ratios with respect to a baseline design to improve; the KF inefficiency is taken as an absolute value. An optimal design minimizes all of the above defined objectives.

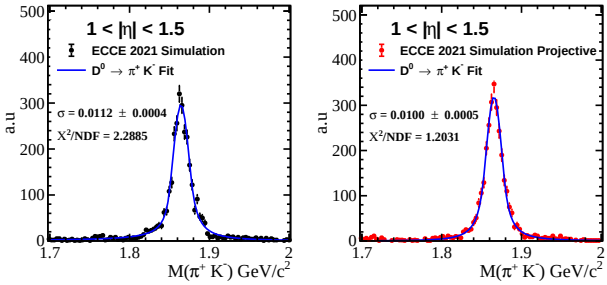

Figure 8: **Analysis of SIDIS events with AI-assisted detector designs:** The π^+K^- invariant mass obtained from the SIDIS events produced in the third simulation campaign of ECCE. A comparison of the resolution on the D meson obtained with the non-projective and the projective design of the ECCE tracker is shown. The mass distribution of π^+K^- is modelled by a double-sided Crystal Ball function (see, *e.g.*, [25, 26]). In order to emphasize the improvement in the region between $1.0 < |\eta| < 1.5$, a selection was made such that both the tracks π^+ or the K^- are within the region of interest. The resolution with the projective design is improved by more than 10% relative to the resolution obtained with the non-projective design.

4.3. Physics Analysis

To show a comparison in performance between the non-projective and projective designs, we analysed D^0 meson decay into π^+K^-. Data have been produced during the ECCE third simulation campaign and utilize SIDIS events generated with Pythia6 [18], corresponding to ep events with 18 GeV × 275 GeV and high Q^2.[7] More info can be found in [24].

In Fig. 8, the π^+K^- invariant mass is fitted with a double-sided Crystal Ball function.

The decay events are selected in such a way to have at least one particle (either π or K, or both) in the pseudo-rapidity bin

[6] Early stopping can occur if no change in the hypervolume is observed after a certain number of evaluations.

[7] production: prop.5/prop.5.1; generator: pythia6; kinematics: ep-18x275-q2-high.

$1.0<|\eta|<1.5$, where the projective design is expected to improve the performance by concentrating all the material in a smaller dead area compared to the non-projective design.

The analysis shows that the resolution obtained with the projective design is improved by more than 10% relative to that obtained with the non-projective design. We also calculate the efficiency, defined as the number of reconstructed D-mesons divided by the number of true D-mesons. It turns out that the efficiency obtained with the two designs is consistent within the statistical uncertainties.

5. Computing Resources

Parallelization. A two-level parallelization has been implemented in the MOO framework: the first level creates the parallel simulations of design points, the second level parallelizes each design point (see Fig. 9). The evaluation itself can be distributed to several workers or a whole cluster with libraries like Dask [27].

Computing Budget. Computing time studies have been carried out to evaluate the simulation time of each single design point as a function of the number of tracks generated. We made this study with simulations that included the tracking system and the PID system and estimated an effective simulation time of 0.2 s/track after removing an initial latency time. Similarly we made studies of the computing time taken by the AI-based algorithm in generating a new population of design points. Results of these studies are summarized in Fig. 10.

A larger population allows to approximate the Pareto front with larger accuracy. Extension of the design parameter space and the objective space to larger dimensionality implies a larger amount of CPU time which is mainly dominated by simulations if the population size remains smaller than 10^4-10^5, see Fig. 10.

For our goals during the detector proposal the optimization pipelines of the ECCE tracking system have been parametrized with 10-20 design parameters and 3-4 objectives; this allowed us to achieve good convergence with evolutionary MOO using a 2-level parallelization strategy, and deployment on single nodes of 128 CPU cores available on the sci-comp farm at Jefferson Lab [28].[8].

Planned Activities. As described in this document, detector optimization with AI is an essential part of the R&D and design process and it is anticipated to continue after the detector proposal. The AI-assisted design optimization of the ECCE inner tracker has been based on evolutionary algorithms; during the detector proposal multiple optimization pipelines have been run each with a population size of 100, representing different detector design configurations. At each iteration, AI updates the population. The total computing budget for an individual pipeline amounted to approximately 10k CPU-core hours. This number depends on the dimensionality of the problem.

[8] Work is in progress to efficiently distribute the optimization pipeline to multiple nodes

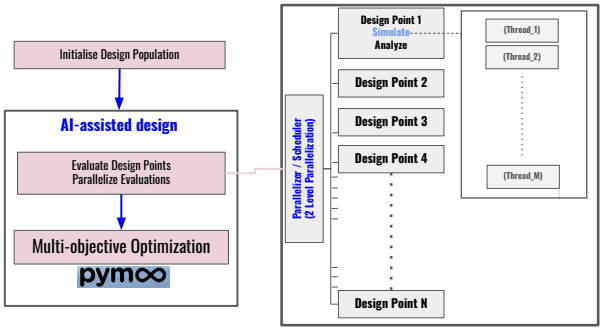

Figure 9: **Flowchart of the two-level parallelization implemented for detector design:** Multiple design points are produced in parallel, and each design point is parallelized. The design is assisted by an AI-strategy, in this case MOO with evolutionary algorithms supported by pymoo [13].

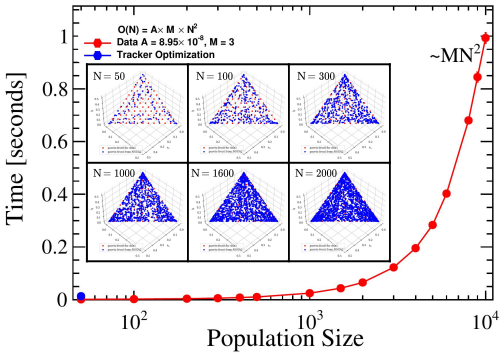

Figure 10: **The computing time taken by the genetic algorithm and the sorting in NSGA-II:** Performance have been benchmark with test problems like DTLZ1 [16] and the scaling $\sim MN^2$ has been verified with convergence to the Pareto front. For the complexity of our problems, the simulation time dominates the AI times. A two-level parallelization has been introduced in the framework to reduce this bottleneck. The AI contribution typically becomes dominant when very large population size is needed to accurately approximate the Pareto front (cf.[29]).

Activities are planned to continue the detector optimization: new optimization pipelines can deal with larger parameter space to include a system of sub-detectors like already did in the case of the whole ECCE tracker or to combine tracking and PID in the optimization process; we also plan to optimize other sub-detectors like, *e.g.*, the dRICH, leveraging on the expertise internal to the ECCE collaboration regarding specifically the design of the dRICH with AI-based techniques [11].

Larger populations may need to be simulated to cope with the increased complexity in order to improve the accuracy of the approximated Pareto front. Different AI-based strategies will be compared.

For the tracker we used as driving criteria figures of merit that embody the performance of the tracking detector: momentum and angular resolution, Kalman filter efficiency. After the detector optimization, we validate the performance on other figures of merit not utilized during the optimization process. We also completed these studies by analysing the performance of the optimized detector on physics channels, particularly we made an analysis of the $D^0 \to \pi^+ K^-$ that showed an improved resolution by at least 10% with the projective tracker design as

compared to the non-projective geometry. As a future activity we aim to encode the result physics analyses as objectives in the MOO problem.

We anticipate for 2022 roughly 1M CPU-core hours for these activities.

6. Summary

Large scale experiments in high energy nuclear physics entail unprecedented computational challenges and the optimization of such complex systems can benefit from state of the art AI-based strategies [6].

In this paper we described the successful implementation of a multi-objective optimization approach to steer the multi-dimensional design of the ECCE tracking system, taking into account the constraints from the global detector design. This work has been accomplished during the EIC detector proposal, and has been characterized by a continued optimization process where multiple optimization pipelines integrating different configurations of sub-detectors have been compared using full Geant-4 simulations. The insights provided by AI in such a multi-dimensional objective space characterizing the detector performance (*e.g.*, tracking efficiency, momentum and angular resolutions), combined to other aspects like risk mitigation and costs reduction, helps selecting the candidate technology of the ECCE tracker. This approach is also used in an ongoing R&D project where the design parametrization has been extended to include the support structure of the tracking system.

The design optimization can be also extended to tune the parameters of a larger system of sub-detectors. Physics analyses are at the moment done after the optimization for a given detector design solution candidate, but they can be encoded during the optimization process as physics-driven objectives in addition to objectives representing the detector performance. A thorough comparison of results obtained with different AI-based strategies (e.g., multi-objective optimization based on genetic algorithms and bayesian multi-objective optimization) can be also studied.

Detector optimization with AI is anticipated to continue after the detector proposal, and activities are planned to further optimize the tracking system, including PID sub-detectors, particularly the dual-RICH [11].

Acknowledgements

We thank the many individuals and funding agencies which provided important advice and support for this work.

References

[1] R. A. Khalek, A. Accardi, J. Adam, D. Adamiak, W. Akers, M. Albaladejo, A. Al-Bataineh, M. Alexeev, F. Ameli, P. Antonioli, et al., Science requirements and detector concepts for the electron-ion collider: EIC yellow report, arXiv preprint arXiv:2103.05419 (2021).
URL https://arxiv.org/abs/2103.05419

[2] The ECCE consortium (2021).
URL https://www.ecce-eic.org

[3] First Authors, et al., Design of the ECCE Detector for the Electron Ion Collider, to be published in Nucl. Instrum. Methods A (in this issue) (2022).

[4] R. Ent, EIC Overview and Schedule, the AI4EIC Workshop - First workshop on Artificial Intelligence for the Electron Ion Collider, http://eic.ai (2021).
URL https://indico.bnl.gov/event/10699/contributions/53658/attachments/36955/60868/AI4EIC_EIC_Overview_and_Schedule_090721_post.pptx

[5] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. . Barrand, et al., GEANT4—a simulation toolkit, Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (3) (2003) 250–303.

[6] C. Fanelli, Design of Detectors at the Electron Ion Collider with Artificial Intelligence, arXiv preprint arXiv:2203.04530 (2022).

[7] Z. Zhou, S. Kearnes, L. Li, R. N. Zare, P. Riley, Optimization of molecules via deep reinforcement learning, Scientific reports 9 (1) (2019) 1–10. doi:https://doi.org/10.1038/s41598-019-47148-x.

[8] K. Guo, Z. Yang, C.-H. Yu, M. J. Buehler, Artificial intelligence and machine learning in design of mechanical materials, Materials Horizons 8 (4) (2021) 1153–1172. doi:https://doi.org/10.1039/D0MH01451F.

[9] B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering, Science 361 (6400) (2018) 360–365. doi:https://doi.org/10.1126/science.aat2663.

[10] A. collaboration, et al., AtlFast3: the next generation of fast simulation in ATLAS, arXiv preprint arXiv:2109.02551 (2021).

[11] E. Cisbani, A. Del Dotto, C. Fanelli, M. Williams, , et al., AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case, Journal of Instrumentation 15 (05) (2020) P05009.

[12] G. Debreu, Valuation equilibrium and Pareto optimum, Proceedings of the National Academy of Sciences of the United States of America 40 (7) (1954) 588.

[13] J. Blank, K. Deb, pymoo: Multi-objective Optimization in Python, IEEE Access 8 (2020) 89497–89509.

[14] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, IEEE transactions on evolutionary computation 6 (2) (2002) 182–197.

[15] D. Whitley, A genetic algorithm tutorial, Statistics and computing 4 (2) (1994) 65–85.

[16] H. Ishibuchi, R. Imada, Y. Setoguchi, Y. Nojima, Performance comparison of NSGA-II and NSGA-III on various many-objective test problems, in: 2016 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2016, pp. 3045–3052.

[17] Eic, fun4all_coresoftware.
URL https://github.com/eic/fun4all_coresoftware

[18] T. Sjöstrand, S. Mrenna, P. Skands, PYTHIA 6.4 physics and manual, JHEP 05 (2006) 026. arXiv:hep-ph/0603175, doi:10.1088/1126-6708/2006/05/026.

[19] S. Agostinelli, et al., GEANT4: A Simulation toolkit, Nucl. Instrum. Meth. A506 (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.

[20] First Authors, et al., Design and Simulated Performance of Tracking Systems for the ECCE Detector at the Electron Ion Collider, to be published in Nucl. Instrum. Methods A (in this issue) (2022).

[21] G. A. Rinella, et al., First demonstration of in-beam performance of bent Monolithic Active Pixel Sensors (5 2021). arXiv:2105.13000.

[22] D. Colella, ALICE ITS 3: the first truly cylindrical inner tracker (2021). arXiv:2111.09689.

[23] J. Arrington, R. Cruz-Torres, W. DeGraw, X. Dong, L. Greiner, S. Heppelmann, B. Jacak, Y. Ji, M. Kelsey, S. R. Klein, et al., EIC Physics from An All-Silicon Tracking Detector, arXiv preprint arXiv:2102.08337 (2021).

[24] ECCE Simulation Working Group page.
URL https://wiki.bnl.gov/eicug/index.php/ECCE_Simulations_Working_Group

[25] M. Aaboud, et al., Search for resonances in diphoton events at $\sqrt{s}$= 13 TeV with the ATLAS detector, Journal of High Energy Physics 2016 (9) (2016) 1–50.

[26] L. Collaboration, et al., First observation of the decay $B^0 \to D^0 \bar{D}^0 K^+ \pi^-$,

Physical Review D 102 (5) (2020) 051102, supplemental Material.
[27] D. D. Team, Dask: Library for dynamic task scheduling (2016). URL https://dask.org
[28] Jlab scientific computing. URL https://scicomp.jlab.org/scicomp/home
[29] X. Liu, J. Sun, L. Zheng, S. Wang, Y. Liu, T. Wei, Parallelization and optimization of NSGA-II on sunway TaihuLight system, IEEE Transactions on Parallel and Distributed Systems 32 (4) (2020) 975–987.

Appendix A. Details on Parametrization

Tracking System Parametrization

Vertex layers. There are 3 vertex barrel layers in the ECCE tracking system made of MAPS technology. The vertex cylinder consists of strips which are made of pixels, where the individual sensor unit cell size is 17.8 mm × 30.0 mm. The length of the vertex layers is fixed at 27 cm; the radii of the three vertex layers are fixed to 3.4, 5.67, 7.93 cm, respectively. For the non-projective design, the angle of the support structure to the interaction points ($\theta = 36.5°$) is fixed and the radius of the support structure is at 6.3 cm, while the length of the support structure is 17 cm. For the projective design parameterization, the radius of the support structure is the same, while the length is calculated based on the angle of projection and the radius as shown in Fig. A.11.

Sagitta layers. There are 2 sagitta barrel layers in the ECCE tracking system. The sagitta barrel layers are also made of MAPS technology. The sagitta barrel layer has a fixed length of 54 cm. For the non-projective design the radius of the sagitta layers are 21.0, 22.68 cm, respectively. For the projective parameterization, The radius of the sagitta barrel is calculated such that there are no gaps in the acceptance of the region enclosed by the barrels. The radius of the barrel layer is calculated as:

$$r_{sagitta} = \frac{l_{sagitta}}{2} \tan\theta.$$

The radius of the sagitta layers are also constrained since the strips have fixed width of 1.78 cm; therefore the radius of the constraint has to satisfy:

$$\frac{2\pi r_{sagitta}}{1.78} = n; n \in 0, 1, 2, ..., N.$$

µRwell layers. In the ECCE tracking system there are 3 cylindrical µRwell layers, each endowed with a support ring. An extended supporting plateau is included at either ends of the µRwell to rest the entire cylindrical detector on this platform. This results in a constant shift of the support cone by the plateau length (5 cm) as shown in Fig. A.11. Therefore, for each µRwell layer the support cone is shifted constantly by the length of the plateau. For both the non-projective design and the projective design the µRwell-1 radius is a free parameter. The length of the µRwell-1 is calculated based on the angle of the conical support structure. In the case of non-projective design we have the conical support structure angle fixed ($\theta = 36.5°$) and therefore the length of µRwell-1 depends only on its radius. In the non-projective design µRwell-2 has its radius as a free parameter. Since the angle of the conical support structure is fixed the length of µRwell-2 depends on its radius. The optimization always preferred to maximize the radius and therefore was fixed in the projective design. For the projective design the µRwell-2 has a fixed radius of 51 cm (*i.e.*, R_{max}+ 1 cm). The length of the µRwell-2 is calculated based on the angle of the conical support structure. The length of the µRwell takes into account the constant shift due to the plateau. µRwell-3 is outside of the inner tracking system. It has radius of 77 cm. The total length of the µRwell-3 is 290 cm. The dimensions of µRwell-3 is fixed in both non-projective and projective designs.

FST/EST Disks. For both the non-projective and projective designs, R_{min} of the EST/FST disks must be compatible with the beam pipe envelope which increases in radius as a function of z; R_{max} of the EST/FST disks is parametrized to be compatible with the support cone structure shown in figure A.11 which has an angle θ (Fixed for non-projective case).

For the non-projective design which has the fixed angle for conical support structure, z positions of EST1, EST2, EST3, EST4, and FST1, FST2, FST3, FST4, FST5 are all free parameters during initial sets of optimizations. However, it can be realised that, to maximize the hit efficiency, disks are placed within the support cone at the start of every plateau (figure A.11 with fixed angle $\theta = 36.5°$). Therefore, 2 disks in the electron going direction and 2 disk in the forward direction are not free in z position. For instance, consider Fig. 5 (right), where EST3, EST4, FST3, FST4 are placed at the begin of the pleateau. Whereas the disks EST1, EST2, FST1, FST2, FST5 are free to vary in z position.

The same parameterization is extended to projective design too such that it is compatible with varying conical support structure.

As the disks are tiled up using MAPS pixels, the $R_{max} - R_{min}$ [mm] are constrained in such a way that the following has to be satisfied:

$$(R_{max} - R_{min})/d = n; n \in \mathbb{Z}^+,$$

where d =17.8 and 30.0 mm. To account for the constraint in all the FST and EST disks in the end caps region, two constraint function summing up all the violations are considered, namely:

$$\sum_i \left| \frac{R^i_{max} - R^i_{min}}{d} - \left\lfloor \frac{R^i_{max} - R^i_{min}}{d} \right\rfloor \right|,$$

where $\lfloor x \rfloor$ represents the ceiling of x. This ensures the amount of violation made by a design solution.

TOF system. The central barrel TOF (CTTL) is an AC-LGAD based TOF detector with a fixed radius of 64 cm and a fixed length of 280 cm. The TOFs at the e-going endcap (ETTL) and the h-going endcap (FTLL) are AC-LGAD-based TOF disks. For the non-projective design the TOF detectors have fixed dimensions. For the projective design the TOF detectors in the end cap regions have their z positions as free parameters. R_{min} and R_{max} of the ETTL/FTTL disks depends on the position of the disk z. The R_{min} of the disk should be compatible with the radius of the beam envelope which increases linearly as a function of z; R_{max} of the disks varies as a function of z such that the acceptance coverage by the ETTL/FTTL is roughly unaltered.

PID Detectors. The Detection for Internally Reflected Cherenkov light (DIRC) is a detector for PID in the barrel region. DIRC system has fixed dimensions and occupies a radial space from 71.5 cm to 76.6 cm. The modular RICH (mRICH) is a Ring Imaging Cherenkov detector system in the e-going direction with fixed dimensions. mRICH has a z position starting at -135 cm extending in z to -161 cm. The dual Radiator Imaging CHerenkov (dRICH) detector is a detector system in the forward direction with fixed dimensions. dRICH has a z position starting at 180 cm and extends up to 280 cm.

After all the parametrization, the thickness of the detectors and support structures are also taken into account to avoid overlaps between the detectors. The most recent optimization pipelines during the detector proposal have been extended to include in the parametrization also the outer tracking layers in the two endcaps, as explained in Sec. 4.1. An overlap check is performed each time a new design point is evaluated during the optimizaton process.

Support Structure Parametrization

The implementation of the projective geometry of the inner tracker is described in Fig. A.11 which shows the parametrization used for the support cone structure of the inner tracker.

Some parameters have been considered fixed and other free to vary within their ranges. Parameters that are fixed typically do not have much room for optimization considering the constraints of the design and potential overlaps.

The non-projective design can be realised by fixing the support structure angle to ($\theta = 36.5°$) shown in A.11. Therefore, the non-projective design solutions are a subset of solutions that can be achieved by this parameterization.

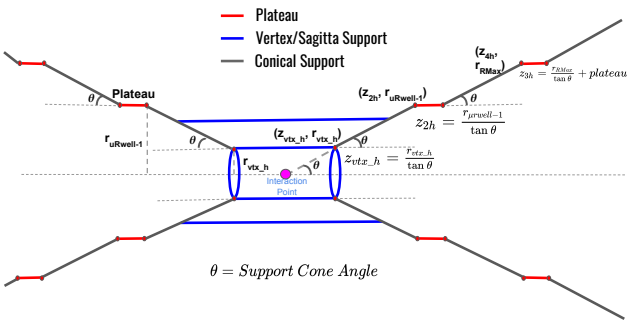

Figure A.11: **Parametrization of the support cone for the inner tracker:** The inner tracker support is characterised by 5 variables. θ (the angle of projection of the support cone structure), r_{vtx} (radius of vertex support structure), $r_{\mu rwell-1}$ μRwell-1 radius, plateau length, r_{max} maximum allowed radius of inner tracker).

Appendix B. Baseline and R&D designs

Resolutions

We report here a thorough comparison between the non-projective ECCE simulation and the ongoing R&D to optimize the support structure through a projective design.

Fig. B.12 shows a comparison of the obtained the momentum resolution in bins of pseudo-rapidity. Similarly Fig. B.13 shows the angular resolution while Fig. B.14 shows the Kalman filter efficiency as defined in Eq. (5).

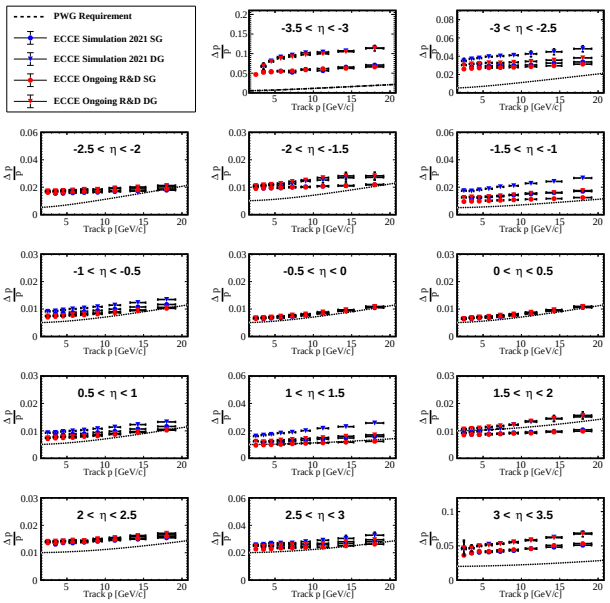

Figure B.12: **Momentum resolution for the ECCE tracker reference design and the projective design (ongoing R&D):** The $\Delta p/p$ are extracted with single-gaussian (SG) fits (circles) to be consistent with the YR PWG and compared to double-gaussian (DG) fits (triangles). DG (see Eq. (3)) is used in the AI-assisted pipelines as it provides more stable fits. DG results are systematically larger than the SG ones as they capture more realistic distributions (see Fig. 6). This effect is larger in $1 < |\eta| < 1.5$ for the non-projective design (blue points) since more material is traversed by the tracks; the same effect is moderately present also in $0.5 < |\eta| < 1$ for the same reason. In the same bins, the ongoing R&D design shows a significant reduction of the impact of the readout and services and SG (red circle) and DG (red triangles) are close to each other. DG points are systematically large in $3 < \eta < 3.5$ and in $-3.5 < \eta < 3.0$, and this corresponds to edge effects at large $|\eta|$ values close to the inner radii of the disks in the endcaps. In the region $-1 < \eta < 3.0$ $\Delta p/p$ is consistent with the YR physics requirements with SG. Note that physics in the backward region relies on the EM calorimeter also, and in the proposal it has been demonstrated that larger resolutions are acceptable.

Validation

Validation is performed by looking at figures of merit that are not used during the optimization process. In Sec. 4.3 we already described a physics analysis with SIDIS events that further consolidates our conclusions. We show here additional examples of validation: Fig. B.15 and Fig. B.16 display the azimuthal angular resolution and the reconstruction efficiency obtained for both the non-projective and the projective ECCE tracker. The azimuthal resolution looks consistent within the uncertainty for the two designs, while the reconstruction efficiency looks in general better for the projective design, particularly in the $1 < |\eta| < 1.5$ region where the non-projective has a larger dead area that affect the reconstruction of tracks.

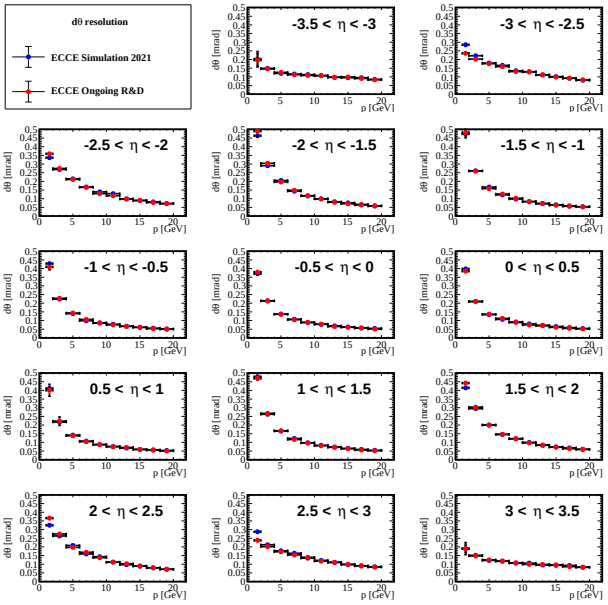

Figure B.13: **Angular resolution ($d\theta$) for ECCE projective geometry and ECCE simulation:** The resolution shows an improvement in the transition region between $1.0 < |\eta| < 1.5$. These plots have been produced using double-gaussian fits.

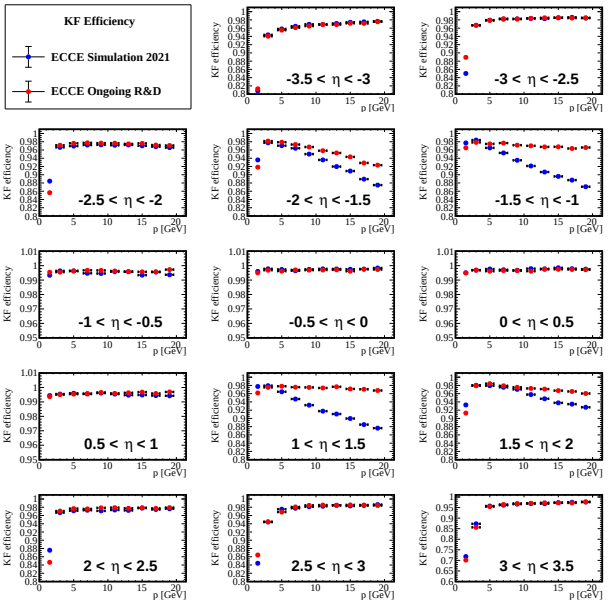

Figure B.14: **KF Efficiency for ECCE projective geometry and ECCE simulation:** Reconstruction efficiency shows an improvement in the transition region between $1.0 < |\eta| < 1.5$. There is a significant drop in the transition region caused due to multiple scattering of the support structure.

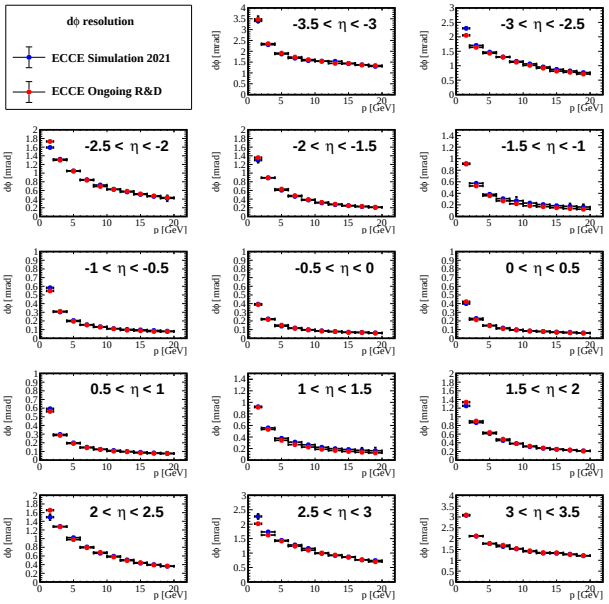

Figure B.15: **Angular resolution ($d\phi$) for ECCE projective geometry and ECCE simulation:** The resolution shows an improvement in the transition region between $1.0 < |\eta| < 1.5$. These plots have been produced using double-gaussian fits.

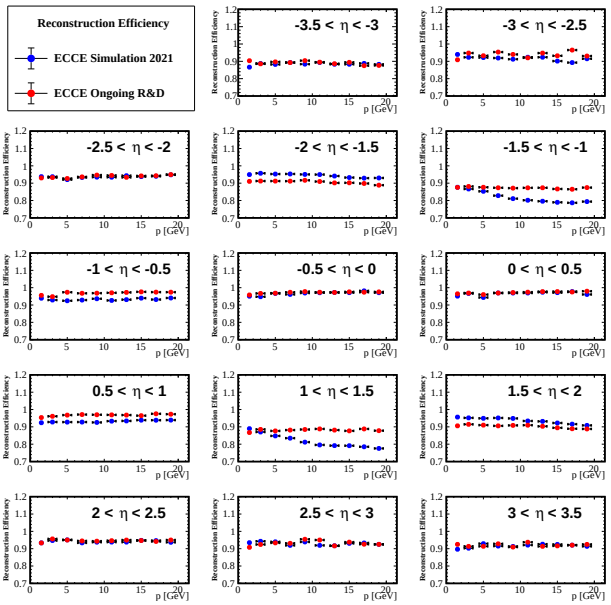

Figure B.16: **Reconstruction efficiency for ECCE projective geometry and ECCE simulation:** Reconstruction efficiency shows an improvement in the transition region between $1.0 < |\eta| < 1.5$

A comparison between the non-projective and projective designs of the inner tracker is also shown in Fig. B.17, where the projective design concentrates the material in a smaller dead area resulting in better resolution on a wider range of the pseudo-rapidity.

Such improvement in resolution is the result of a continued multi-objective optimization process as described in Sec. 4 and summarized in Fig. B.18, where the momentum resolution ob-

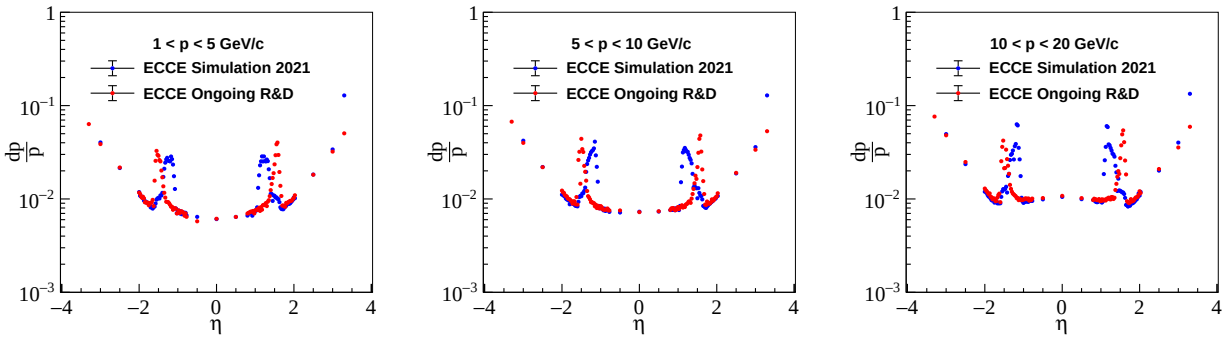

Figure B.17: **Comparison between non-projective and projective inner tracker:** a comparison in momentum resolution as a function of the pseudo-rapidity η between the non-projective and the projective designs in three bins of momentum. The projective design concentrates the material in a smaller dead area resulting in better resolution on a wider range of the pseudo-rapidity.

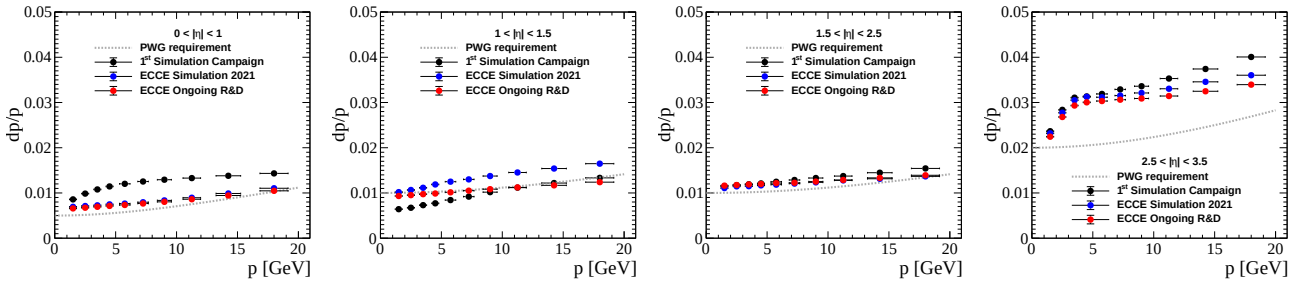

Figure B.18: **Evolution of momentum resolution driven by the continued MOO:** Momentum resolutions in four bins of η. From left to right, respectively: $|\eta| <$ 1.0 corresponding to the barrel, 1.0 $| \leq \eta <$ | 1.5 corresponding to the transition region, and $1.5 \leq |\eta| < 2.5$ and $2.5 \leq \eta < 3.5$ corresponding to the two end-caps. Black points represent the first simulation campaign, and a preliminary detector concept in phase-I optimization which did not have a developed support structure; blue points represent the fully developed simulations for the final ECCE detector proposal concept; red points the ongoing R&D for the optimization of the support structure. Compared to black, there is an improvement in performance in all η bins with the exception of the transition region, an artifact that depends on the fact that black points do not include a realistic simulation of the material budget in the transition region. In the transition region, it can be also appreciate the improvement provided by the projective design.

tained during phase-I optimization using a preliminary detector concept is compared to both the non-projective and the projective R&D designs which are instead derived from fully developed simulations.