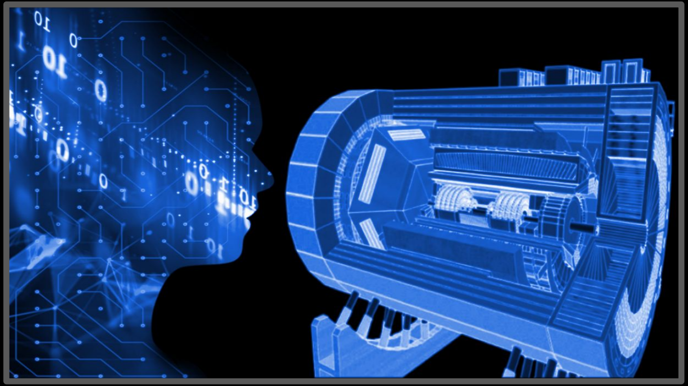
AI-assisted EIC SW framework for Detector R&D / Design



Application: Combined design of a projective tracker and PID system for ePIC

Potential extension: Design of Detector-2

E. Cisbani[1], C. Fanelli*[2,3], T. Horn[4,3], S. Oresic[5], Z. Papandreou[5], M. Schram[3], K. Suresh[5]

1 INFN
2 William & Mary, Chartered 1693
3 Jefferson Lab, Exploring the Nature of Matter
4 The Catholic University of America
5 University of Regina

Generic EIC related Detector R&D Program

Overview

- This project provides an AI-assisted framework for the ongoing R&D and design optimization of the EIC detectors. The project utilizes the EIC software infrastructure (for ePIC and potentially detector-2).
- This project leverages on existing advanced activities and a proven track record of achievements:
    - [1] C. Fanelli, Z. Papandreou, K. Suresh, et al. "AI-assisted Optimization of the ECCE Tracking System at the Electron Ion Collider." arXiv:2205.09185 (2022). — accepted on Nucl. Instr. and Meth. A
    - [2] E. Cisbani, CF, A. Del Dotto, M. Williams, et al. "AI-optimized detector design for the future Electron-Ion Collider: the dual-radiator RICH case." J. Instrum 15.05 (2020): P05009.
- The EIC is leading the efforts on AI-assisted design of large-scale experiments. Uniquely positioned to design—during the AI revolution—the ultimate machine to study the strong force

➢ Main goal of FY23: parametrization of tracker and PID systems; coupling with MOGA and MOBO and embedding of objectives (intrinsic sub-detector system response and begin work on physics-driven quantities); parallelization

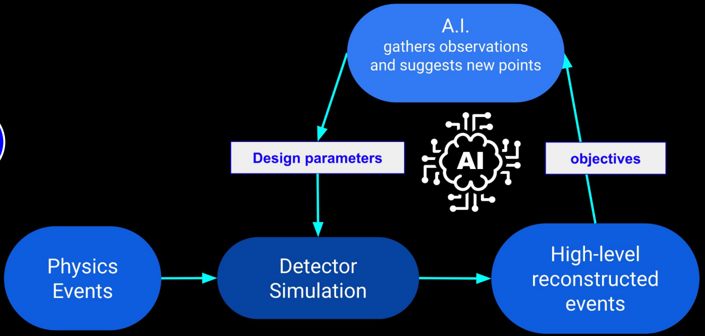
[1] CF, Z. Papandreou, K. Suresh, et al. arXiv:2205.09185 (2022). — accepted on Nucl. Instr. and Meth. A

[2] E. Cisbani, CF, A. Del Dotto, M. Williams, et al. J. Instrum 15.05 (2020): P05009.

See Gantt chart for FY23, FY24

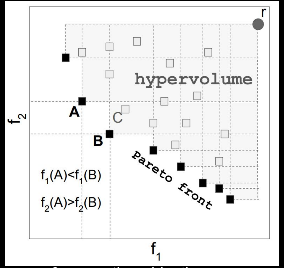
Assisted R&D and Design Optimization

1

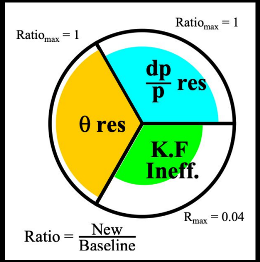


"Every optimization problem is fundamentally a multi-objective optimization problem." — M. Balandat, Meta AI [1]

2

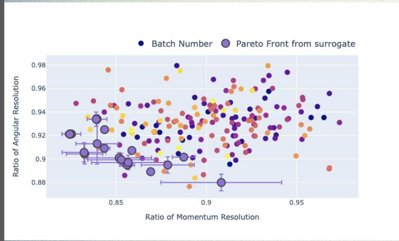


2 competing objectives

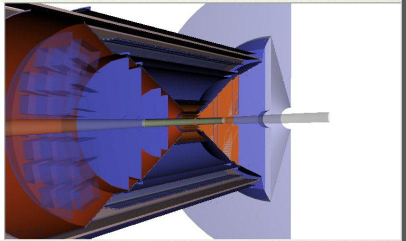


multiple objectives as in [2]

Multi Objective Bayesian Optimization

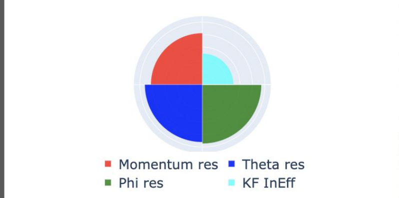


GEANT4 Visualization of the design



Click on petals for finer evaluations

Performance of the Chosen Design Solution



Design Parameters Table

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

Click here for an example of interactive navigation of the Pareto design solutions

4

3

- **We deal with coupled systems of sub-detectors**: an optimization of one sub-detector (e.g., tracker barrel), can impact another sub-detector (tracker end-cap), which in turn can affect PID sub-system (dRICH in the hadronic end-cap).
- **Most of optimizations are MOO** (e.g.,momentum resolution, angular resolution, efficiency, etc.). In addition to that, there are physics-driven objectives. Optimize one objective at a time can lead to suboptimal solutions.
- **Set of tradeoff solutions** in a multi-dimensional design space, and for a multi objective space, we have a set of Pareto solutions
- **Holistic approach to detector optimization** is possible only with modern techniques that rely on Artificial Intelligence.

[1] M. Balandat, AI/ML for Design session, AI4EIC – Ax/BoTorch [link]
[2] CF, et al. arXiv:2205.09185 (2022).

Jefferson Lab
Exploring the Nature of Matter

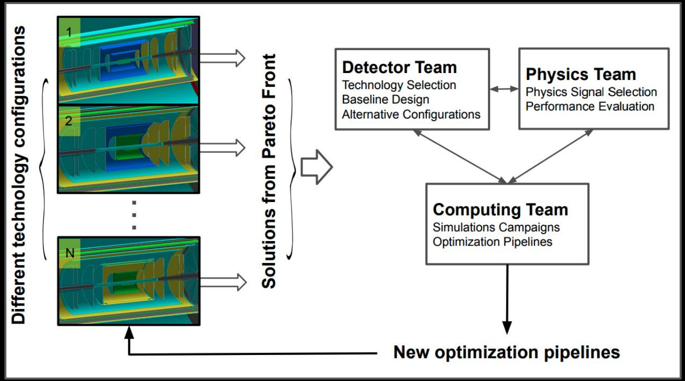
AI4EIC

Advantages from EIC Software stack

- We have already tested the coupling of the ePIC software stack with the AI framework.
- There are multiple forward-looking aspects in the ePIC SW infrastructure which are favorable for AI/ML implementation [1]:

- Support for parameterization: geometry implementation via data source makes transparent the coupling of AI to the software stack design parameters
- Modularity of geometry reduces complexity of parametrization
- CI/CD is mostly about keeping up-to-date with the ePIC simulation framework
- Containerization is being used in ePIC and previously in the proto-collaborations, make it easy to bundle AI/ML packages
- Parallelizable framework and other automated features desirable for AI-assisted design

Continuous optimization process: integrates the most up to date version of the simulations with updated technology (resulting from the R&D efforts of each group) — AI fosters interplay among groups working on sub-detectors. Taken from [2].



see Question #3

[1] C. Fanelli, EPIC Software Infrastructure Review, AI/ML Synergy, https://indico.bnl.gov/event/16676/
[2] CF, Z. Papandreou, K. Suresh, et al. arXiv:2205.09185 (2022). — accepted on Nucl. Instr. and Meth. A

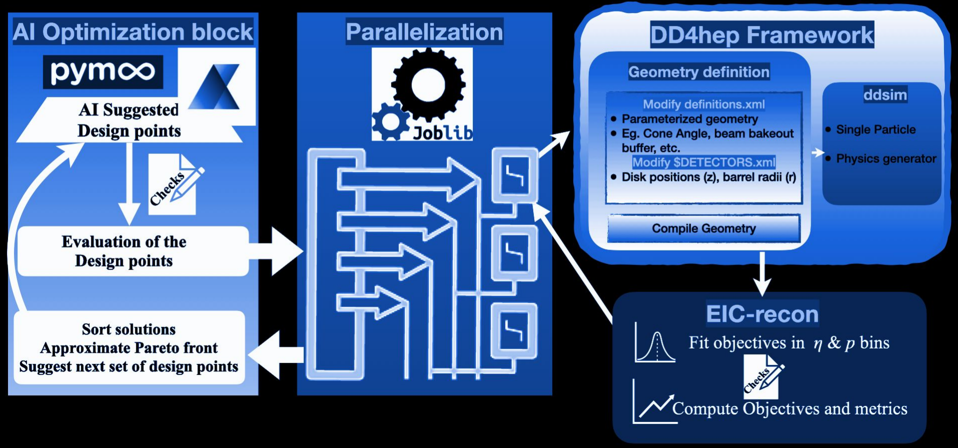
AI-assisted Framework in 1 slide

2 different AI-approaches already available:
Multi-objective Bayesian Optimization (MOBO)
Multi-objective Genetic Algorithm (MOGA)

SINGULARITYCE

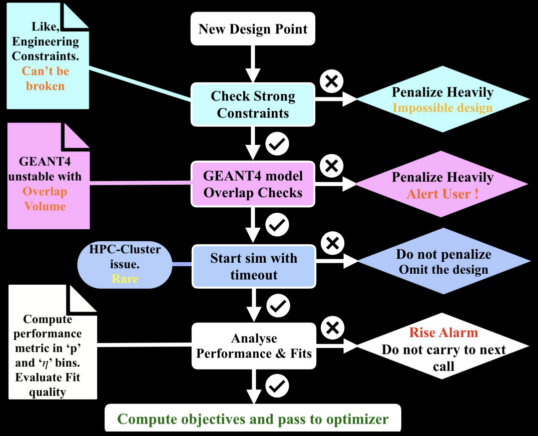
EIC Detector optimization Pipeline

1

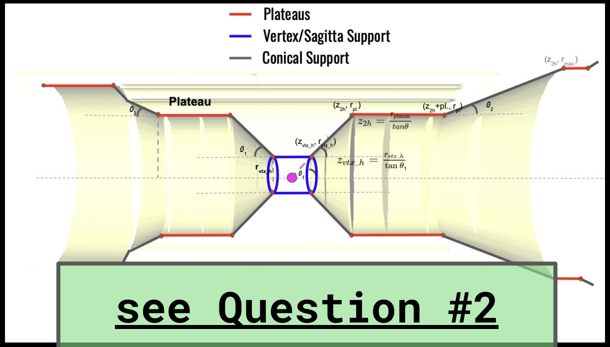


Inclusion of soft and hard constraints

2



Flexible parametrization to explore multiple design configurations simultaneously (example for tracking)

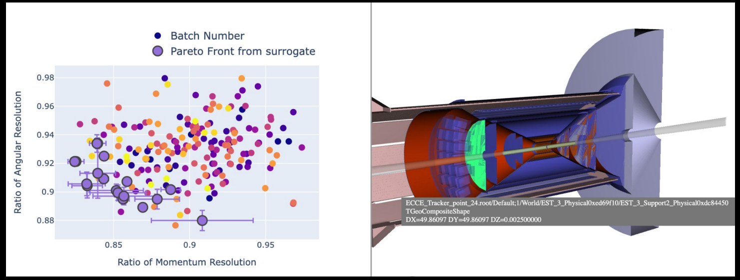


see Question #2

3

Cutting-edge data science tools for interactive browsing of the Pareto front and a posteriori selection of optimal design point

4



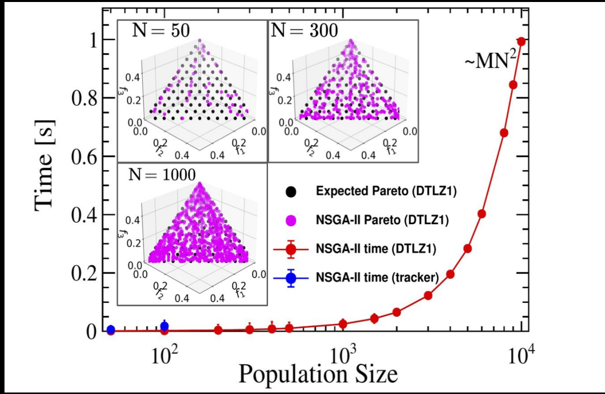
Jefferson Lab
Exploring the Nature of Matter

Computing Resources

Numbers from proto-collaboration phase [2]

- Continue to utilize JLab computing farm for this work and when needed resources made available by the institutions collaborating in the EIC project (see computing model [1]).
- For the AI-assisted design it has been anticipated 1M CPU-core hours [1].
- The implementation of MOGA in [2] was entirely based on CPUs. For MOBO, we can use GPU for the realization of surrogate models and autodifferentiation:
  - We can leverage on heterogeneous computing in the machine learning nodes available at JLab.
  - Depending on needs we could also get access to Compute Canada, which recently has been rebuilt into the Digital Research Alliance of Canada.

For the complexity of the problem and the chosen population size, the computing time is dominated by simulations and not by the AI part



- Used a test problem DTLZ1
- Verified scaling following MN^2 and convergence to true front
- ~1s/call with 10^4 size!
- Smart pipelines of 11 variables and 3 objectives needs ~ 10000 evaluations to converge

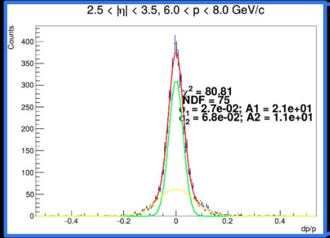
~10k CPUhours / pipeline

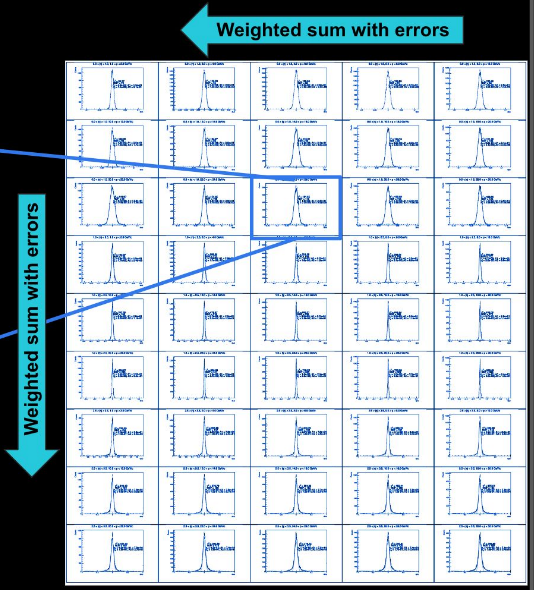
[1] C Bernauer, et al. Scientific Computing Plan for the ECCE Detector at the Electron Ion Collider. arXiv preprint arXiv:2205.08607, 2022.
[2] CF, Z. Papandreou, K. Suresh, et al. arXiv:2205.09185 (2022). — accepted on Nucl. Instr. and Meth. A

Tracking

- **Parameters that can be included**
  - Cone angles and thickness of services
  - Simultaneously barrel and endcap regions (radii of barrels and z-position of disks to maximize acceptance, minimize dead material, optimize resolutions)
  - Dimensions of layers and disks with constraints — that comes with a cost
- **Costs**
  - either as FOMs or constraints during optimization
- **Possibility to explore multiple pipelines**
- **Subsidiary optimization**
  - - e.g., efficient arrangement of MAPS readout on the disks

- Objectives evaluated in fine-grained phase-space
- Propagate uncertainties from fits





$$\bar{x}_\eta = \frac{\sum_p x_p w_p}{\sum_p w_p}$$

(sum in bins of 14 bins of P)

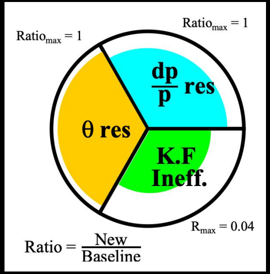
$$\bar{x} = \frac{\sum_\eta^{N_\eta} \bar{x}_\eta}{N_\eta}$$

(Average objective in a η bin)

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

Multiple Objective Optimization already explored for tracking [1]

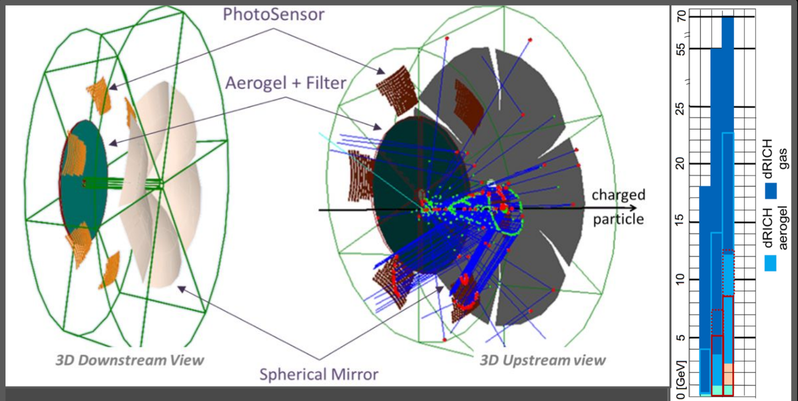
Momentum resolution
Angular resolution
Tracking Efficiency



[1] CF, Z. Papandreou, K. Suresh, et al. arXiv:2205.09185 (2022). — accepted on Nucl. Instr. and Meth. A

dRICH

- **Cherenkov detectors essential part of ePIC PID**
  - Simulation typically compute expensive, photons tracked through complex surfaces.
  - Rely on pattern recognition of ring images in reconstruction, and the DIRC is the one having the more complex ring patterns!
- **Extension to tracker + PID system**
  - Potential to optimize parameters of the dRICH design in the hadronic endcap
- **E.g., dRICH design**
  - Large momentum coverage
  - Two radiators: aerogel and gas
  - Legacy design from INFN: 6 Identical open sectors and large focusing mirror

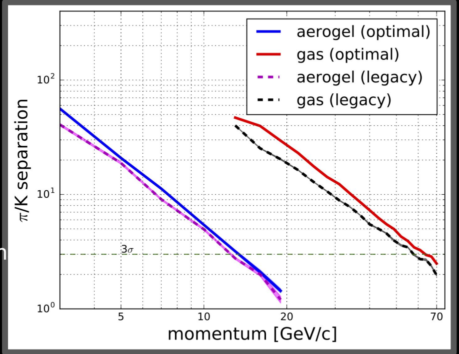


aerogel (4 cm, n(400 nm): 1.02) + 3 mm acrylic filter + gas (1.6 m, n(C_2F_6): 1.0008)

$$N\sigma = \frac{||\langle\theta_K\rangle - \langle\theta_\pi\rangle||\sqrt{N_\gamma}}{\sigma_\theta^{1p.e.}}$$

Design parametrization: optics + geometry [1]

parameter	description
R	mirror radius
pos r	radial position of mirror center
pos l	longitudinal position of mirror center
tiles x	shift along x of tiles center
tiles y	shift along y of tiles center
tiles z	shift along z of tiles center
$n_{aerogel}$	aerogel refractive index
$t_{aerogel}$	aerogel thickness



At present different R&D options are considered th could affect the parametrization.
FOMs / objectives should be studied in terms of implemented reconstruction algorithm.

[1] E. Cisbani, CF, A. Del Dotto, M. Williams, et al. J. Instrum 15.05 (2020): P05009.

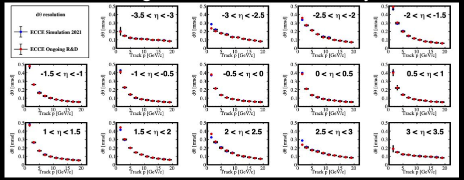
Plans – Gantt Chart

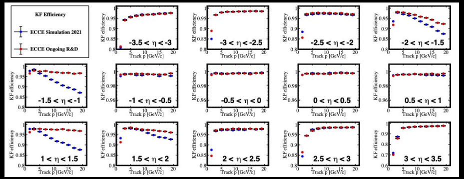
Item	Tasks	Fiscal Quarter After Award							
		FY23			FY24				
		Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
Software	Tracker parametrization (DD4Hep)	✅							
	dRICH parametrization (DD4Hep)								
	integration of realistic (beam) backgrounds†								
	realistic reconstruction algorithms (no ground truth information)†								
	extension to other sub-detectors (e-endcap PID and calorimetry)						with baseline budget		
Coupling with AI	MOGA for tracker+dRICH								
	MOBO for tracker+dRICH								
Embedding of objectives	Tracker performance (res./eff.,etc)								
	dRICH performance								
	Physics-driven								
	Costs						with budget -20%		
	extension to other sub-detectors						with baseline budget		
Infrastructure	Parallelization of simulations								
	Optimization of interactive resources (sim.+AI)								
	Distributed computing						with budget -20%‡		
Analysis	Detector intrinsic response								
	Detector + physics-driven								
	Detector + physics-driven + costs							with budget -20%	
	extension to other sub-detectors						with baseline budget		

†: contingent on progress in the software stack; provided anticipated timeline. Realistic reconstruction algorithms may include: the Acts Common Tracking Software; Inverse Ray Tracking for the dRICH. These (and similar other solutions) are ongoing efforts within the EIC Detector-1 collaboration and will be available in the central software stack repository.
‡: even in the case of reduced budget scenario by 40%, we will leverage on the consulting role of JLab/DS for this task (and possibly on additional workforce/collaboration) to speed up the design optimization workflow.

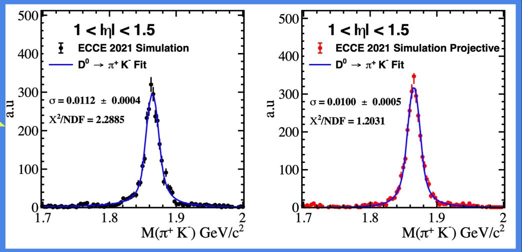
✅Already started
Work already started
along with coupling of AI with tracker

Detector performance
E.g., resolutions, efficiency





Physics driven objectives
Combine information from tracker and PID,
e.g., D0 reconstruction



Detector-2 and Beyond

see Question #1

- There are benefits for ePIC and Detector-2
    - The AI-assisted EIC SW framework can be naturally extended to Detector-2. This is supported by:
        - The fact that the approach is agnostic to what is being optimized
        - Modularity of the framework (both on the AI side and the software side as explained and stated in the EIC SW statement of principles) [1]
        - The fact that Detector-2 can potentially adopt the same EIC SW stack.
    - Spinoff: R&D of design material
- Community outreach:
    - Design optimization, lectures and tutorials at AI4EIC https://eic.ai/community

[1] EIC SW Statement of Principles, https://eic.github.io/activities/principles.html

Jefferson Lab
Exploring the Nature of Matter

Question #1

- This proposal appears to be highly optimized to finalize details of project Detector 1. How would you respond to the statement that, from the point of view of EIC-related generic R&D, the usefulness of this technology has been demonstrated and the review committee should expect a proposal for its application to the design of Detector 2 in the coming years?

- As just discussed, both ePIC and Detector-2 will benefit from this project.
  - In fact, the AI framework is by construction agnostic to what is currently optimized
  - We leverage on the modern features of the EIC software stack built for ePIC (e.g., inherent support for parametrization, parallelizability) and on the fact that potentially the same SW and the coupling between AI/SW can be utilized for Detector-2
  - The AI-assisted approach is also utilized for the R&D of new material (reinforced aerogel) that can find applications in EIC

Jefferson Lab
Exploring the Nature of Matter

Question #2

- The close connection to the detector working group is emphasized. Is there convergence/support for the projective tracker layout?

- The (quasi-)projective jargon came from an R&D project during the ECCE proposal; nowadays 45 deg projective is used in simulations where the first slope of the support structure is at 45 deg and is projective to the nominal interaction point.
- Parametrization is key in what we are proposing, in that it can include and reproduce multiple designs (and projective designs) by exploring different angles of the support structure simultaneously and with the most up-to-date constraints to satisfy; the optimization also includes the geometry and location of sub-detectors for the tracking and PID sub-systems.
- We are proposing a framework for multi-objective optimization of the EIC detector(s). The design is optimized holistically looking at the detector as a whole.

Question #3

- How will the work be affected by migration of the ePIC software framework?

- We are already working with the ePIC software stack [1]
- There are indeed multiple advantages in this migration as we already highlighted [1] thanks to the forward-looking aspects inherent to the ePIC SW infrastructure which are favorable for AI/ML:
    - Support for parameterization: geometry implementation via data source makes transparent the coupling of AI to the software stack design parameters;
    - Modularity of geometry reduces complexity of parametrization
    - CI/CD is mostly about keeping up-to-date with the ePIC simulation framework:
    - Containerization is being used in ePIC and previously in the proto-collaborations, make it easy to bundle AI/ML packages.
    - Parallelizable framework
    - Other automated features desirable for AI-assisted design

Jefferson Lab
Exploring the Nature of Matter

[1] C. Fanelli, EPIC Software Infrastructure Review, AI/ML Synergy, https://indico.bnl.gov/event/16676/

Question #4

- The entire budget in Table 2 is for the fully loaded cost of a postdoc. As stated in the proposal guidelines, “Limited support for postdoctoral fellows will be considered. There is tension between the desire for proponents to support postdocs with the hope of renewal, and the review committee’s desire to flexibly channel each year’s limited funds to the most promising new proposals.” If only 50% of the fully loaded cost of a postdoc were available from this program, is there some cost-sharing arrangement with another group with a different funding source (or faculty start-up funds) that could be quickly arranged?

- Yes. The remaining 50% can be covered by William & Mary.

Jefferson Lab
Exploring the Nature of Matter

Thank you

Questions?

> *"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"*. [1]

[1] R. Stevens et al, DOE Town Halls on AI for Science, 2019