PRINCIPAL INVESTIGATOR	**Kolja Kauder (PO/NPPS)**	**5935**
DEPARTMENT/DIVISION	**Physics/NPP**	**7/21/23**
OTHER INVESTIGATORS	**Maxim Potekhin (PO/NPPS), Torre Wenaus (PO/NPPS)**	
TITLE OF PROPOSAL	**EIC simulation infrastructure**	
TYPE A		
PROPOSAL TERM (month/year)	From **10/2023**	Through **9/2026**

SUMMARY OF PROPOSAL

Description of Project:

The Electron Ion Collider (EIC) has been identified by the nuclear physics (NP) community as its highest priority new facility, providing in the 2030's unprecedented insights into the innermost structure of protons and nuclei [1]. It is the largest new user facility in development at BNL. The approved funding includes a single detector, ePIC (Electron-Proton/Ion Collider experiment). The consensus in the NP community is that fully exploiting the science potential of this ca. $2B scale facility requires a second detector ("Detector 2"), and the community is mobilizing to secure its funding from national and international NP parties.

EIC software and computing (S&C) development has advanced rapidly since the formation of the ePIC Collaboration a year ago, with a full modern software stack now in place. JLab and ANL have been the principal drivers in achieving this, thanks to their substantial investments in ePIC software and computing. BNL has invested considerably less, primarily because of the budgetary weight of building the EIC machine. As a consequence, BNL's contributions to and influence on ePIC software are less than one would expect of the lab hosting the facility. Strength in software directly supports the science, at both the design/development and datataking phases, so correcting this is a high priority for the BNL EIC community.

The newly established ePIC S&C organization presents opportunities to rectify this. A critical software activity supporting EIC detector design and physics performance studies in the coming years is detector simulation development. The understanding of detector performance and physics reach that drives detector design comes from highly detailed modeling of the detector and its physics and readout response. Principals on this proposal are Physics and Detector Simulation Co-Convener (Kauder) and Deputy Software and Computing Coordinator (Wenaus). If we can establish at BNL a strong simulation development project, we have the levers to ensure it aligns with EIC needs. Another important lever is that we have a long standing – but under-resourced – collaboration with JLab on leading edge detector simulation R&D for EIC, the eAST [2] project (e-A Simulation Toolkit, where "e-A" stands for electron-ion), which builds on the NP and HEP standard Geant4 [3] simulation software. eAST is led by Makoto Asai, our collaborator at JLab, one of Geant4's principal architects and experts for over 20 years. eAST is a development testbed for leading edge simulation technologies important first and foremost to EIC but also to the wider NP and HEP communities.

We propose to leverage our EIC software leadership roles, simulation expertise and nascent collaboration with JLab in an R&D program that would transform our current low-level, best-effort work into the EIC community's principal simulation development activity. We initially target Detector 2; its longer timeline is a better fit to our 3-year R&D program, with ePIC to follow. At its conclusion, the project delivering community-level contributions at the leading edge of detector simulation.

This proposal addresses the topic area from the LDRD call: *Research and Development towards the Second Detector and Computing at the Electron-Ion Collider*.

Expected Results:

We propose three targeted high priority areas of simulation infrastructure development:

Integrated fast and full simulation Detailed simulations are computationally expensive, which can be eased by leveraging new and rapidly evolving technologies for fast simulation; speed-ups by a factor of 400 have been demonstrated. Fast simulations require approximations that impact physics fidelity. We will implement an integrated fast and full simulation capability in Geant4 that allows for simulation fidelity tunable to the physics needs of the study in question.

Python interface to Geant4 and eAST We will develop a Python interface to Geant4 and eAST that provides flexible, user-friendly access to the rich functionality of the simulation software in a much more accessible and configurable form than the software's native C++. We will extend the Python interface to the full EIC software stack, to provide for control and configuration of EIC software across the simulation workflow, and integrating with Python's powerful machine learning and scientific software ecosystems.

Sub-event level parallelism Geant4 newly supports the ability to compute processes in parallel inside a single simulated collision event. We propose to incorporate this sub-event level parallelism as an enabler for the use of highly parallel computing architectures including GPUs (graphics processing units), which are increasingly prevalent, and for which the simulation needs of EIC detectors are particularly well suited because of the importance of Cherenkov [4] particle ID detectors with their high processing demands for optical photon simulation, well suited to leveraging GPUs.

PROPOSAL

1 Detector simulation for the Electron Ion Collider

This proposal addresses the topic area from the LDRD call: *Research and Development towards the Second Detector and Computing at the Electron-Ion Collider*.

The Electron Ion Collider (EIC) [1] and its funded detector ePIC [5] (Electron-Proton/Ion Collider experiment) are slated to be constructed at BNL within the next decade. Should the funding be found for a second detector, its scientific desirability to complement the project detector is undisputed; in a recent Whitepaper [6] regarding Long Range Planning for the EIC, the EIC User Group recommended to the Nuclear Science Advisory Committee that provides official advice to the Department of Energy (DOE) and the National Science Foundation "targeted efforts to enable the timely realization of a second, complementary detector".

The EIC detector development process to date has in its software and computing aspects advanced very rapidly to establish the basis for a modern software stack and infrastructure directed at effectively meeting near term needs and providing a smooth development path towards EIC datataking. The needs of the ePIC detector's development program naturally prescribe priorities, design choices and work plans for software development. Detector 2 software and computing is expected to leverage ePIC's, reflecting a universal emphasis on common software in the EIC community. The longer Detector 2 timeline presents opportunities to advance EIC S&C computing R&D in the near term without conflicting with the detector development timeline. For example, in the crucial simulation domain, much research and development is still needed to effectively leverage both new capabilities in our simulation tools and the rapidly evolving ecosystem in common scientific software.

We propose a three year R&D program on EIC detector simulation infrastructure, targeting initially the Detector 2 development program and timeline, with deliverables defined also in collaboration with ePIC for eventual adoption there as well. The work would be a collaborative effort with JLab where in recent years a strong capability in the most advanced detector simulation tools and techniques has been established. This project would closely couple BNL and JLab efforts in a combined program, with BNL taking leadership roles in its targeted activity areas. The proposal builds on an existing collaborative BNL-JLab EIC simulation project, eAST [2] (e-A Simulation Toolkit), which extends the functionality of Geant4 [3], the NP and HEP universal standard for detailed detector simulation. Led by Makoto Asai at JLab, one of Geant4's principal architects and experts, eAST is being used as a development testbed for new tools and technologies that are important to EIC simulation, but are on a longer developmental timescale than the tight timeline of ePIC (ePIC is selectively integrating eAST components such as new EIC-specific physics optimizations). Neither BNL nor JLab currently has the effort to turn eAST's planned development program into an active one.

Without the resources this project would provide, BNL is unable to fully draw on our simulation expertise and EIC software leadership roles towards advancing EIC simulation. Meanwhile JLab plans to increase further their simulation effort. Having recently hired Makoto Asai, they plan to hire a software physicist to work with him. This is the moment for BNL to ramp its effort and establish itself as a leader in the EIC's most vital software activity.

Our three targeted activities, described in the following section, are 1) integrated fast and full simulation, to provide fast, tunable, physics-optimized simulation capability incorporating the latest fast simulation techniques such as machine learning; 2) Python interface to Geant4, eAST and the EIC software stack to provide ease of use, access to eAST's extended functionality, and integration with Python's powerful machine learning and scientific software ecosystems; and 3) sub-event level parallelism, leveraging Geant4's new capability in this area and its applicability to EIC simulation in particular (via the importance of optical photon propagation, the original design purpose of GPUs, graphical processing units) with the prospect to be the first demonstrated success in Geant4 parallelization and GPU utilization in an experiment simulation.

A letter of support from JLab is appended to this proposal.

2 Proposed R&D

The proposed R&D encompasses three technical activity (TA) areas as follows, chosen to establish active efforts in eAST's highest priority areas for research and development towards the needs of EIC detector simulation.

2.1 TA1: Integrated fast and full simulation

Detailed simulation of detector subsystems, interaction region, and support structures are time consuming, both in terms of people- and CPU-hours. The hermeticity and particle identification (PID) requirements for any EIC detector are unparalleled. Any second general purpose detector as recommended by the Yellow Report [7] therefore needs electromagnetic and hadronic calorimetry, as well as Cherenkov detectors around and at both ends of the interaction point. Both of these technologies are especially computationally expensive due to long and intricate showers inside the respective material, and they often dominate the simulation time. In recent times, technologies are rapidly developing to speed these up using approximate simulation, for example with parameterization or with deep learning methods, see e. g. Refs. [8–11]

But the industry standard full simulation package Geant4 [3] does have built-in fast simulation capabilities, and they are expanding. Figure 1 shows a speed-up by a factor of 400, demonstrated using project, eAST (e-A Simulation Toolkit, where "e-A" stands for electron-ion). This is an existing collaborative BNL-JLab EIC simulation project, led by Makoto Asai, one of Geant4's principal architects and experts, and a development test bed for tools and technologies.

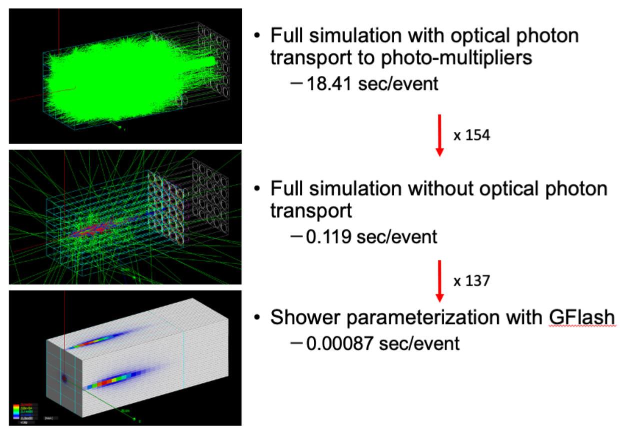

Figure 1: Demonstration of the simulation speed-up of a 5 GeV electron in an Electromagnetic calorimeter. Created by Makoto Asai using Geant4 tools with eAST.

The holistic approach of using fast simulation within Geant4 not only allows to switch between the two option, but also to speed up only certain time-consuming regions while leaving others that need more fine-grained full simulation untouched.

Furthermore, Geant4 supports swapping in and out any simulation technology for a given subdetector. Thus, in addition to Geant4's built-in capabilities, external methods can be used, giving entry to existing and quickly developing acceleration techniques. At present, detector simulation is projected to be the primary computational bottleneck for experiments at the high luminosity Large Hadron Collider (LHC) runs; accelerating these simulations is a prime concern and is evolving very rapidly. It is worth noting that most LHC experiments focus on calorimetry, whereas the EIC could take lead and give back to the community with fast photon propagation. New methods are mostly based on advanced Machine Learning but could also synergize with hardware acceleration from TA3.

While the project detector team must use what is currently at hand and focus on creating simulations, the longer lead time of Detector 2 opens up an opening to fundamentally improve on the interface with Geant4 and take full advantage of all its capabilities, thus minimizing resource needs and turnaround time for Detector 2 while maximizing statistics. It enables easier and faster Research and Development for any second detector, while also cross-fertilizing technologies into ePIC in keeping with the EIC User Group's stated goal of an as-much-as-possible common software stack. We will implement a common and integrated approach for fast and full simulations in Geant4 with a plug and play modular approach, leveraging and building on Geant4's fast simulation capabilities.

2.2 TA2: Python interface to Geant4 and its extensions (eAST)

We propose to develop a Python interface to major components of the EIC simulation software stack (including the ePIC framework) geared toward immediate integration of Detector 2 software into the common EIC software stack. The motivation is to provide flexible, user-friendly access to the rich and complex functionality of the simulation software, while incorporating new features and capabilities of Geant4 [3] and eAST.

In addition, such a Python interface will facilitate integration with a variety of Machine Learning applications and packages (which are commonly equipped with Python API), and facilitate experimentation with different detector configurations and parameters.

There is more than one method of building Python interfaces to complex C++ software. For example, experience of the ATLAS Collaboration with the *cppyy* software package [12] was leveraged by one of us (M.Potekhin) to create a highly functional Python interface to the C++ simulation suite developed for the LuSEE-Night project [13]. This achieves the goal of combining the speed and complexity of C++ with the user friendly front end, opening the possibilities of integration with powerful Python packages and popular analysis tools such as Jupyter Notebooks [14]. Based on this experience, a prototype Python interface for a simple subset of eAST commands was created and tested.

At the same time, there are alternative methods based on software such as *pybind11* [15], recently utilized in the *geant4-pybind* Python package [16]. Initial evaluation of this package was done in both LuSEE-Night and eAST contexts. Our approach will be to continue to evaluate advantages and disadvantages of each method and choose one most appropriate for the implementation. This Python interface will be implemented as a package, and will be usable by the EIC researchers as well as the wider Geant4 community. Our plan for this item is as follows:

- Perform preparatory work on the details of the Geant4 software build procedure.
- Create and evaluate a prototype of the *cppyy*-based API for Geant4 and eAST.
- Investigate the feasibility of adding the eAST component to the *geant4-pybind* package.
- Make the technology choice for the final product. Implement the Python interface for Geant4 and its eAST extension.

The milestones of this plan are itemized in Table 2 (Section 4).

2.3 TA3: Sub-event level parallelism

Geant4 has a new capability for sub-event level parallelism as an enabler for high concurrency and use of accelerators such as GPUs. While previous projects to re-engineer detailed detector simulation for high concurrency, such as GeantV [17], have not delivered to expectation, Geant4's new Task subsystem builds the capability directly into Geant4 in a manner compatible with Geant4's proven architecture. The principal developer of the Task subsystem is our JLab collaborator, Makoto Asai. This R&D project would support a postdoc to draw on this capability, in collaboration with our JLab colleagues, to create a high concurrency, GPU-capable EIC simulation, targeting specifically the Cherenkov-based particle ID (PID) systems. PID is of utmost importance for EIC physics, and in the EIC's energy domain, Cherenkov technology is particularly well suited. The ePIC detector alone has three such detectors. Cherenkov detectors operate by sensing optical photons triggered by the passage of charged particles. While in general it has proven difficult to effectively use GPUs in detailed detector simulation because of the complexity of the physics processes particles are subject to, optical photons are a well suited special case, involving the optical ray tracing that GPUs were originally designed to perform in graphical processing. The Opticks [18] software has been developed to leverage GPUs for optical photon propagation in Geant4 simulation. We will integrate use of the Task subsystem and Opticks in EIC particle ID detector simulation with the prospect to be the first demonstrated success in Geant4 parallelization and GPU utilization in an experiment simulation.

This activity has the longest timeline of the three and can ramp up at a slower pace, and so our expectation is that the second postdoc hired will work on this activity. The timeline is reflected in the milestones shown in Table 2.

3 Personnel

Personnel include three staff members in the Physics Department's Nuclear and Particle Physics Software (NPPS) Group, all with deep experience across the project's technical areas, together with two postdoc hires. All personnel have backgrounds in simulation leadership: Wenaus (BaBar), Kauder (ePIC), Potekhin (STAR, PHENIX). Wenaus at BNL and Diefenthaler & Asai at JLab founded the eAST project.

Kolja Kauder has led EIC fast simulation development since 2019 [19]. He has responsibility for ePIC simulation as Physics and Detector Simulation Co-Convener. He is lead PI and will lead TA1 on fast and full simulation, and co-lead TA3 on sub-event parallelism.

Maxim Potekhin is a long time expert on detector simulation as the Simulation Coordinator on both the major RHIC experiments, currently PHENIX and previously STAR. Maxim has been substantially involved in the testing and documentation work for eAST, and is the BNL lead on the EIC Collaborative Tools infrastructure. He is Co-PI and will lead TA2 on python integration.

Torre Wenaus is Deputy Software and Computing Coordinator for the ePIC experiment, co-creator of the eAST project, and NPPS Group Leader. He was Simulation Manager on BaBar, the first experiment to use Geant4. He co-manages ATLAS High Luminosity LHC computing for the US, where fast simulation and parallelization are major activities. He is Co-PI and will co-lead TA3.

The primary workforce on the project will be two full-time postdocs. Planning conservatively we do not assume that we will have the postdocs on board in year 1. Accordingly, in year one the three principals will spend a larger fraction of their time on the project than in years 2 and 3, and in year 1 we will engage the help of another NPPS member and nuclear physics software developer, Dmitri Smirnov. Kauder, Potekhin and Smirnov will be at 40% in year one, Wenaus at 20%. To the extent we are able to bring postdocs on board during year one, these fractions will go down. The project activities are aligned closely enough with existing staff responsibilities that this level of participation early in the project is feasible.

Once the postdoc effort is ramped, the three staff will work at about 40% FTE level in aggregate. Their senior roles in EIC software and substantial participation in the work will ensure strong engagement with and mentoring of the postdocs, as well as preservation and continuity of the work beyond the term of the LDRD. Project participation (once postdoc effort is ramped) is summarized in Table 1.

Participant	*FTE level*	*Roles*
Kolja Kauder	0.13	Lead PI. Technical lead for TA1, co-lead for TA3
Maxim Potekhin	0.13	Co-PI. Technical lead for TA2
Torre Wenaus	0.10	Co-PI. Technical co-lead for TA3
Postdoc 1 - first hire	1.0	TA1, TA2 development. Supervision by KK, MP
Postdoc 2 - second hire	1.0	TA1, TA3 development. Supervision by KK, TW

Table 1: Project participation once postdocs are in place

4 Deliverables, milestones and metrics

The project deliverables in the form of milestones and metrics are summarized below. In broad outline, FY24 establishes the project team (initially all staff pending postdoc hires) and begins the technical work in a generic standalone way, outside the EIC detector context, establishing feasible approaches and laying the roadmap. In FY25 the target becomes EIC Detector 2, integrating with its software stack and having specific subdetector simulation objectives, culminating in performance studies. In FY25 the Detector 2 integrations and associated performance studies are completed, R&D evaluations are made on simulation approaches including ML integration and GPU utilization, and the developments are generalized and packaged/documented for the future use of ePIC and others in the community.

Date	*Activity*	*Milestone/Metric*
		FY24 - Standalone development and feasibility demonstrations
Q1	All	Hires initiated and interim ramp-up effort plan established
Q1	TA2	Geant4 build procedure refactored to facilitate integrating the Python interface
Q2	All	First postdoc hire completed
Q2	TA1	Existing detector geometry imported with Region envelopes
Q2	TA3	Geant4 parallelization Task subsystem and Opticks integrated in eAST
Q3	All	Second postdoc hire completed
Q4	TA1	Integrated fast and full demonstrated with at least one subdetector and a parameterized simulation drawing on eic-smear experience
Q4	TA2	eAST/Geant4 *cppyy* interface prototyping and evaluation
Q4	TA3	Feasibility demonstration of parallelized Cherenkov toy detector simulation
		FY25 - EIC detector integration, validation, performance studies
Q1	TA2	Creation and testing of realistic use cases for evaluation of the *cppyy* API
Q2	TA1	Fast simulation tuned and validated against full with at least one subdetector
Q2	TA3	Demonstration with performance study of parallelized Cherenkov toy detector simulation using GPUs
Q3	TA1	Create well documented examples to support others implementing hybrid simulations
Q3	TA2	Evaluation of the *pybind11* Python interface for use with eAST
Q4	TA1	Fast simulation for a second subdetector and approach (e.g. ML)
Q4	TA2	*cppyy* vs *pybind11* technology downselect, for building the Python API for eAST
Q4	TA3	Demonstration with performance study of parallelized ePIC Cherenkov detector simulation
		FY26 - Complete EIC integration, evaluations, community generalization, documentation
Q1	TA2	Design of a comprehensive Python interface optimal for both Geant4 and its eAST extension
Q2	TA1	Refine and compare methods, document, establish best practice recommendations
Q2	TA1	Implementation of the eAST Python interface
Q4	TA2	Demonstrate integration with the EIC software eco-system and potentially key4hep or other HEP frameworks
Q4	TA3	Complete parallelized implementation of ePIC Cherenkov detectors operational on CPUs and GPUs
Q4	TA1, TA3	R&D evaluations completed on simulation approaches including ML applied to fast simulation and GPU utilization in Cherenkov detector simulation
Q4	All	Developments generalized and packaged/documented for the future use of ePIC and others in the community

Table 2: Timeline with milestones

5 Final points

BNL's EIC S&C effort is under-resourced relative to BNL's EIC role and the well-resourced efforts elsewhere, particularly JLab. This project would contribute to redressing this imbalance to the benefit of the scale and scope of BNL's EIC S&C program in the future, and to the benefit of BNL's detector development program now (highly dependent on simulation) and EIC physics program later (highly dependent on S&C). The unique opportunity of collaborating with, and being mentored by, a world-leading expert in Makoto Asai will further boost the post-doctoral hires' potential.

It is a particularly opportune moment for this project. The 1st International Workshop on a 2nd Detector for the Electron-Ion Collider [20] was held only in May of this year. Drawing from experience with the Yellow Report and the original proposals, efforts will for a while rely on fast simulation with external packages only, allowing for a window in time to improve the simulation infrastructure and then to synergistically grow as Detector 2's R&D needs change to Geant4 simulations.

References

[1] A. Accardi, et al., Electron Ion Collider: The Next QCD Frontier: Understanding the glue that binds us all, Eur. Phys. J. A 52 (9) (2016) 268. arXiv:1212.1701, doi:10.1140/epja/i2016-16268-9.

[2] eAST Working Group, e-A Simulation Toolkit), https://eic.github.io/east/, [Online; accessed 18-July-2023] (2023).

[3] S. Agostinelli, et al., GEANT4 – a simulation toolkit, Nucl. Instrum. Meth. A 506 (2003) 250–303. doi:10.1016/S0168-9002(03)01368-8.

[4] E. Nappi, J. Seguinot, Ring imaging Cherenkov detectors: The state of the art and perspectives, Riv. Nuovo Cim. 28 (8-9) (2005) 1–130. doi:10.1393/ncr/i2006-10004-6.

[5] ePIC Collaboration, electron-Proton/Ion-Experiment, https://wiki.bnl.gov/EPIC/index.php?title=Main_Page, [Online; accessed 19-July-2023] (2023).

[6] EIC User Group, White Paper on the Electron-Ion Collider in Preparation for the NSAC Long Range Plan, https://nuclearsciencefuture.org/wp-content/uploads/2023/03/LRP_EIC_White_Paper_Final.pdf, [Online; accessed 19-July-2023] (2023).

[7] R. Abdul Khalek, et al., Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report, Nucl. Phys. A 1026 (2022) 122447. arXiv:2103.05419, doi:10.1016/j.nuclphysa.2022.122447.

[8] G. Aad, et al., AtlFast3: The Next Generation of Fast Simulation in ATLAS, Comput. Softw. Big Sci. 6 (1) (2022) 7. arXiv:2109.02551, doi:10.1007/s41781-021-00079-7.

[9] A. Collaboration, Deep generative models for fast photon shower simulation in atlas (2022). arXiv:2210.06204.

[10] M. P. Whitehead, A palette of fast simulations in LHCb, PoS ICHEP2018 (2019) 271. doi:10.22323/1.340.0271.

[11] G. R. Khattak, S. Vallecorsa, F. Carminati, G. M. Khan, Fast simulation of a high granularity calorimeter by generative adversarial networks, Eur. Phys. J. C 82 (4) (2022) 386. arXiv:2109.07388, doi:10.1140/epjc/s10052-022-10258-4.

[12] W. Lavrijsen, cppyy: Python-C++ bindings interface based on Cling/LLVM, https://pypi.org/project/cppyy/, [Online; accessed 18-July-2023] (2023).

[13] S. D. Bale, et al., Lusee 'night': The lunar surface electromagnetics experiment (2023). arXiv:2301.10345.

[14] JupyterLab: A Next-Generation Notebook Interface, https://jupyter.org/, [Online; accessed 18-July-2023] (2023).

[15] W. Jakob, Seamless operability between C++11 and Python, https://pypi.org/project/pybind11/, [Online; accessed 18-July-2023] (2023).

[16] H. Haariger, Python bindings for Geant4, https://pypi.org/project/geant4-pybind/, [Online; accessed 18-July-2023] (2023).

[17] T. Wenaus, et al., HSF GeantV R&D Panel Report, https://hepsoftwarefoundation.org/assets/GeantVPanelReport20161107.pdf (2016).

[18] S. C. Blyth, Opticks : Gpu optical photon simulation for particle physics using nvidia® optix™, Journal of Physics: Conference Series 898 (4) (2017) 042001. doi:10.1088/1742-6596/898/4/042001.
URL https://dx.doi.org/10.1088/1742-6596/898/4/042001

[19] K. Kauder, T. Burton, et al., eic-smear, a Monte Carlo analysis package originally developed by the BNL EIC task force, https://github.com/eic/eic-smear, [Online; accessed 19-July-2023] (2010-2023).

[20] EIC User Group, 1st International Workshop on a 2nd Detector for the Electron-Ion Collider, https://indico.bnl.gov/event/18414/, [Online; accessed 20-July-2023] (2023).