Generic R&D Proposal: BeAGLE, a tool to refine IR and detector requirements for the EIC

Mark D. Baker*[1], Wan Chang[2], Markus Diefenthaler[3], Florian Hauenstein[3], Or Hen[4], Douglas Higinbotham[3], Alex Jentsch[5], Jeong-Hung Lee[5], Wenliang Li[6,7], Pawel Nadel-Turonski[6,7], Dien Nguyen[3], Jackson R. Pybus[4], Zhoudunming Tu[5], Jan Vanek[5], Natalie Wright[4], and Liang Zheng[8]

[1]MDBPADS LLC, Miller Place, NY 11764
[2]Nanyang Normal University, Nanyang 473061, Henan, China
[3]Jefferson Lab, Newport News, VA 23606
[4]Massachusetts Institute of Technology, Cambridge, MA 01239
[5]Brookhaven National Laboratory, Upton, NY 11973
[6]Department of Physics, Stony Brook University, Stony Brook, NY 11794
[7]Center for Frontiers in Nuclear Science at SBU, Stony Brook, NY 11794
[8]China University of Geosciences, Wuhan, Hubei, China

July 25, 2022

Abstract

The BeAGLE Monte Carlo package was used to help develop and refine the IR (Interaction Region) and forward detector requirements for e+A collisions at the EIC. Further development of BeAGLE into a robust, modern, maintainable package tuned to the latest data from JLAB is essential for the EIC in at least two ways. First, BeAGLE will be needed for the ongoing design of the second IR and the forward elements of the second detector. Second, it will ensure that the simulations being used for the project detector are reliable and reduce the risk of having the detector fall short of physics expectations. In this proposal, we outline a detailed plan to address the most urgent shortcomings of BeAGLE in FY2023 and we sketch out a plan for placing BeAGLE on a solid long-term footing by the end of FY2025.

*PI & Contact: mdbaker@mdbpads.com

1 Introduction

BeAGLE stands for **B**enchmark **eA G**enerator for **LE**ptoproduction. It is a general-purpose Monte Carlo model for simulating inelastic scattering in eA collisions, including the nuclear response [4]. It was supported by the Generic EIC R&D program and was used to help develop and refine the IR (Interaction Region) and forward detector requirements for e+A collisions at the EIC. The Yellow report itself [12] explicitly mentions BeAGLE when talking about the purpose and success of the first EIC Generic R&D program:

> Besides hardware R&D, the program supports various vital projects such as machine background studies and simulation software developments to enable more accurate definition of the physics' requirements. Sartre and Beagle are two examples of Monte-Carlo event generators whose development was substantially boosted by the program. Both were intensively used in the context of this report.

This proposal, to further develop BeAGLE as a tool for the community, relates to the charge of the EIC Generic R&D program:

> [The] focus of this EIC-related generic detector R&D program is to evaluate opportunities to achieve new, cost-effective detector capabilities that reduce risk. This program will support advanced R&D on innovative detector concepts that either the one detector in the project scope or a second detector could incorporate.

BeAGLE simulations have played a vital role in the design of the machine-detector interface and the far-forward hadron instrumentation and the BeAGLE team is leading the MC simulations for eA. Reliable simulations are essential for cost-effective detector solutions and additional studies are needed for the integrated IR and detector design, in particular for the far-forward hadron instrumentation. So the simulation effort perfectly aligns with the charge. In particular, the goal of this proposal is not really just software (although that is a deliverable), but rather reliable physics and detector simulations, mandatory for any detector R&D program.

Given the importance of BeAGLE to the EIC forward detector and IR design process so far, the Snowmass 2022 Summer Study [3] raised a concern.

> The future development of this code is uncertain primarily due to lack of manpower and reliable funding.

This proposal will help address that concern and provide a common and generic simulation tool for the EIC community which is unique in that it is the only general-purpose eA MCEG available to the community.

1.1 Status of BeAGLE

BeAGLE is a hybrid model that uses the DPMJet [15], PYTHIA6 [17], PyQM [5], FLUKA [2, 7] and LHAPDF5 [20] codes to describe high-energy lepto-nuclear scattering. Overall steering and optional multi-nucleon scattering (shadowing) is provided in BeAGLE as well as an improved description of Fermi momentum distributions of nucleons in the nuclei (compared to DPMJet). DPMJet is not designed for light nuclei, so substantial changes were made for the case when the nucleus is a deuteron [19], and it does not work properly for the case of $A = 3$. The geometric density distribution of nucleon in the nucleus is provided primarily by PyQM while the quark distributions within that geometry are taken from the EPS09 nPDF [6]. BeAGLE allows the user to provide "Woods-Saxon" parameters, including non-spherical terms, to override the default geometric density description. The parton-level interactions and subsequent fragmentation is carried out by PYTHIA6. The optional PyQM module implements the Salgado-Wiedemann quenching weights to describe partonic energy loss [16]. Hadronic formation and interactions with the nucleus through an intranuclear cascade is described by DPMJet. The decay of the excited nuclear remnant is described by FLUKA, including neutron and proton evaporation, nuclear fission, Fermi breakup of the decay fragments and finally de-excitation by photon emission.

The process that led to the EIC Project IR and detector design, including the production of the Yellow Report and the three Detector Proposals, was fast-paced and, unfortunately, BeAGLE was left in a non-ideal state. The known problems include:

1. Due to limitations in DPMJET, it is not possible to get reliable results when the incoming ion has an A of 3 (e.g. ^3He or ^3H).

2. The intranuclear cascade used in e+A (from DPMJET) has not been tuned to any recent results from JLAB. An attempt was made to tune using Fermilab E665 data, but there is a lot of uncertainty remaining.

3. There are instances of modest 4-momentum non-conservation.

4. Due to the merger of multiple FORTRAN-based codes into BeAGLE (also FORTRAN), the logic is confusing in places and some code is initialized, but not used, adding to the running time, complexity of debugging, and difficulty of migration to a more modern language (C++).

5. The code includes packages, such as PYTHIA6, which are no longer maintained.

1.2 Development of BeAGLE

The current proposal outlines a plan to address items 1–4 on the above list during FY2023. We anticipate addressing item 5 during FY2024-2025, but the detailed plan will be the subject of a future proposal cycle round.

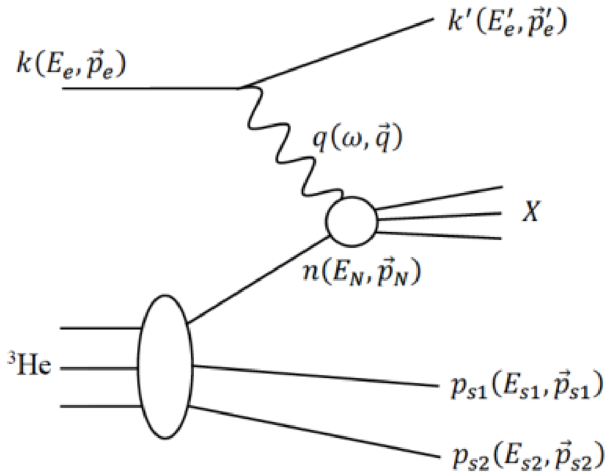

Figure 1: A diagram of Deep Inelastic e+^{3}He scattering with double spectator tagging. The channel shown here is electron scattering off a neutron in^3He; the two spectator nucleons are the protons in the process $^3He(e, e'p_s1)p_{s2})X$. Figure taken from [8]

2 Proposed work

2.1 FY2023

We plan to address the first four issues on the list above during FY2023, as well as making a more detailed plan for addressing issue five in the outyears.

2.1.1 A=3 nuclei

The BeAGLE description of the nucleus is based on DPMJET which is designed for incoming nuclei large enough that there is meaningful nuclear remnant after the hard scatter off of one nucleon to absorb any inelasticity. This causes serious problems for the deuteron and the A=3 nuclei (^3H and ^3He). The deuteron handling has been significantly improved already (see, for instance [10, 19]), but $A = 3$ does not really work properly.

A description of an improved procedure for handling A=3 nuclei, using the example of ^3He can be found in Ref. [8]. Figure 1, taken from that publication, is a diagram of the physicially significant case where the struck nucleon is a neutron and there are two spectator protons which can potentially be tagged, allowing us to control the kinematics of the neutron and extrapolate back to a free neutron.

2.1.2 Comparison to JLAB data and tuning

Short-range nucleon-nucleon correlations (SRCs) within the nucleus can be explored at the EIC by tagging the partner nucleon in a correlated pair when one of the two nucleons is struck in the hard collision. Depending on the kinematics of the collision, there may be more or less final state interaction (FSI) between the outgoing nucleons and the nuclear remnant. Natalie Wright has already compared data on light nuclei from JLAB to eGENIE, a program which simulates e+A collisions with a simplified FSI based on a single-scatter and attenuation model.

The GCF+BeAGLE code can also simulate SRC+FSI events [9], and the model should be appropriate to both light nuclei (as long as $A > 4$) and heavy nuclei including Pb. Natalie has joined our effort and can help compare JLAB data to BeAGLE. Furthermore, this data should help us potentially tune the intranuclear cascade in BeAGLE, in particular the formation time parameter τ_0. So far, it has been difficult to pin this parameter down precisely [4].

This could be extremely valuable as validating the BeAGLE intranuclear cascade model and tuning τ_0 could reduce the uncertainty in many of the $e + A$ simulations for the EIC and therefore reduce the risk of the detector under-performing in it's physics goals.

2.1.3 Four-momentum non-conservation

A small percentage of events have small errors in the total 4-momentum of the final state. There are two main sources of this four-momentum non-conservation.

First, the DPMJET intranuclear cascade code is fairly complicated and uses multiple reference frames. There are some unusual cases where particles are inserted in the event in the wrong reference frame, leading to an error. Second, different parts of the code (Pythia, DPMJET, Fluka) can use slightly different masses for different nuclei.

The larger, more frequent, bugs of this type have already been fixed, but more need to be tracked down.

2.1.4 Cleanup of the code

BeAGLE is an agglomeration of DPMJET and PYTHIA6 along with steering routines which handle Glauber modeling of the nucleus and possible multiple scattering etc. Many aspects of the DPMJET code are not actually used, most notably the Glauber routines. Nevertheless, they are fully initialized which wastes time and also makes the code very confusing and difficult to maintain and difficult to port to C++.

This code needs to be cleaned up and the logic streamlined.

2.1.5 Planning for an upgrade

It is widely recognized that a general-purpose generator for e+A collisions is needed for the EIC. Furthermore, it is recognized that the current form of BeAGLE is not as robust or maintainable as desired and will not automatically take into account new improvements to the description of electroproduction events (such as in Pythia 8). This has been discussed, for instance at the Snowmass 2021 Summer Study [3]

One deliverable of this effort for FY2023 will be a concrete and more detailed plan for addressing these concerns in a timely fashion. In particular, we expect to have a specific proposal for FY2024-2025 with some additional key collaborators by the time of the next cycle of the Generic EIC R&D proposal process (April 2023?). If there is an interim status report (December 2022?), we will have a tentative plan with any open questions clearly stated.

2.2 Outyear plans (FY2024-25)

The ability to model the nuclear response is a particularly unique feature of this code, especially the modeling of the excited nuclear remnant and it's decay, including photonic de-excitation. We want to preserve the best features of this code/model into the future. Some aspects of BeAGLE will therefore need to be ported to C++.

Pythia6 [17] is no longer being maintained and all new physics development is occuring in Pythia8 [18]. This includes up to date theoretical handling of hard diffraction and parton showering. Complicating this situation is the fact that the Pythia6, particularly the EIC version, has a better description of leptoproduction in the $Q^2 < 10$ GeV^2 region which will need to be included in Pythia8.

A big open question is how to handle the nuclear response. We can move in one of two directions (or both): 1) engage more completely with FLUKA (PEANUT) using it for the intranuclear cascade as well as the nuclear response or 2) model the nuclear remnant decay with an open-source code such as ABLA [11] or GEMINI^{++} [13]. The open-source approach may require further development, for instance implementing photonic de-excitation.

It should be noted, that medium-term, the current version of BeAGLE can be used as an afterburner to another primary model (e.g. Angantyr) to apply the nuclear breakup model to the spectator nucleons. A similar approach has already been used to study the physics of tagged Short-Range Correlations [9]. The Generalized-Contact-Formalism (GCF) generator [14] simulates the hard interaction between the electron and a pair of nucleons and leaves the rest to BeAGLE. In any case, we will need to keep the current version of BeAGLE around for comparisons to the new version (BeAGLE++?) for some time.

2.3 Deliverables

If the proposal receives full support, the deliverables would be as follows:

- Year 1 (FY2023)

  A version of BeAGLE will be released with an improved handling of A=3 nuclei, including the possibilities of both 2- and 3-body breakup as well as the correct kinematics.

  GCF+BeAGLE (with Intranuclear Cascade) will be compared to JLAB data, and if needed and possible, the τ_0 parameter in BeAGLE will be tuned to improve agreement.

  A version of BeAGLE will be released with 4-momentum conserved.

  A version of BeAGLE will be released with the code cleaned up and streamlined.

On the time table of the next proposal cycle (April 2023?), a more detailed plan for migration to C++ will be available.

Preliminary reports of our results will be presented at meetings.

- Year 2–3 (FY2024–2025)

  Publish key findings from Year 1 (arXiv or journal), if not already done.

  Implement a “BeAGLE++” including Pythia8 with one or more options for handling the nuclear breakup (FLUKA and/or ABLA).

3 Personnel

Mark D. Baker and the MIT students (Jackson Pybus and Natalie Wright) are directly funded under this proposal, but advice and work will be provided by all the collaborators. Baker has been leading the development of BeAGLE for the last seven years and has extensive experience in nuclear physics. Pybus has been involved in the simulation of $A = 3$ and SRCs. Wright has worked on comparisons between SRC simulations with FSI to data for light nuclei which will be valuable as we extend the simulation to heavier nuclei where an intranuclear cascade can form as opposed to a single reinteraction.

Baker, Chang, Jentsch, Lee, Tu and Zheng have been part of the BeAGLE development team and the eRD17 project for many years. Baker, Tu, and Zheng have all participated in modifying the BeAGLE code itself. Jentsch and Tu, in particular, have experience in implementing light nuclei (specifically the $A = 2$ deuteron) in BeAGLE which will be helpful in extending to $A = 3$. Vanek is a new BNL postdoc who will work on the project, but will be supported by BNL.

Baker, Hauenstein, Hen, Higinbotham, Nadel-Turonski, Nguyen, Pybus and Tu were part of JLAB LDRD projects which used BeAGLE and will provide the basis for items 1 and 2 on the list above. In particular, Nguyen is an expert on ^3He experiments and she and Pybus have worked on simulating $A = 3$ nuclei [8]. Similarly, Hauenstein is an expert on SRC simulations using BeAGLE [9]. Finally Hen and Pybus are experts on SRCs and Wright has worked on comparisons between SRC simulations with FSI to data for light nuclei.

Diefenthaler is working with developers of MC event generators, including the Pythia collaboration, on simulation needs for the EIC. In his role as convener of the event generator working group of the HEP Software Foundation, he is coordinating common efforts on event generators. He will advise on the validation of BeAGLE and the upgrade to Pythia8.

Finally Li has been heavily involved in using BeAGLE to simulate a detector response, most notably for the ECCE proposal and notes [1].

4 Funding Request

One component of the funding is for generator development and expert guidance on nuclear simulations, provided by M.D. Baker whose work at 0.25 FTE (under the full funding scenario) will be contracted from MDBPADS LLC (a small business). Another component of the funding is to support graduate students at MIT who have experience in items 1 and 2 of the list above. Finally a modest travel budget is included to allow presentation of results and gathering of feedback from the community.

The large number of collaborators and the large user base for BeAGLE make this proposal very cost effective. As mentioned above, Vanek and his supervisor Tu will provide effort along with Hauenstein and Nguyen.

4.1 Budget Table

Item	100%	80%	60%
MDBPADS LLC cost	\$70.4k	\$68.3k	\$63.2k
MIT student support	\$30.0k	\$15.0k	\$0k
Travel	\$5k	\$1k	\$0k
Total	\$105.4k	\$84.3k	\$63.2k

Table 1: Budget for Year 1. All items include institutional overheads.

4.2 Impact of Reduced Funding Scenarios

The impact of reduced funding will be to reduce the available effort.

At 100%, the goals and deliverables will be achieved.

Under the 80% funding scenario, one or more of the goals will be at risk. Mostly likely the basic code cleanup and debugging (items 3 & 4 a) may be incomplete and need to be finished in year 2, possibly delaying the migration to C++.

The 60% funding scenario will lead to the postponement of the basic cleanup and debugging and will likely delay the overall timetable for development of a robust, modern, maintainable software package which includes the latest known physics.

5 Diversity, Equity, and Inclusion

The proposed program provides excellent opportunities for training and career development through engagement in research of undergraduate and graduate students from underrepresented groups in physics as well as from economically disadvantaged communities and first-generation college students. Software, in particular, is an area where students typically have strong skills and can participate on a more equal footing with their mentors. During their research on the project, these students will be integrated in the research groups of the senior personnel committed to the project, where they will receive training in nuclear instrumentation, simulations, and data reduction techniques. Students will be also trained in how to communicate their research to an audience of peers and will be encouraged to present at collaborative meetings and workshops where they will be able to benefit from networking with other senior and junior researchers. The program also provides training and career opportunities for qualified postdoctoral fellows. The program, thus, supports retention of highly-skilled individuals in nuclear physics. Overall, the proposed program supports the commitment of the participants to educate and mentor a diverse work force through research.

References

[1] A. Bylinkin, et al. *Exclusive and Diffractive Processes and Tagging with the ECCE Detector at the Electron Ion Collider*. to be published in Nucl. Instrum. Methods A. 2022.

[2] T.T. Böhlen et al. "The FLUKA Code: Developments and Challenges for High Energy and Medical Applications". In: *Nucl. Data Sheets* 120 (2014), pp. 211–214. DOI: 10.1016/j.nds.2014.07.049.

[3] J. M. Campbell et al. "Event Generators for High-Energy Physics Experiments". In: *2022 Snowmass Summer Study*. Mar. 2022. arXiv: 2203.11110 [hep-ph].

[4] Wan Chang et al. "BeAGLE: Benchmark *e*A Generator for LEptoproduction in high energy lepton-nucleus collisions". In: (Apr. 2022). arXiv: 2204.11998 [physics.comp-ph].

[5] Raphaël Dupré. "Quark Fragmentation and Hadron Formation in Nuclear Matter". PhD thesis. Lyon, IPN, 2011.

[6] K.J. Eskola, H. Paukkunen, and C.A. Salgado. "EPS09: A New Generation of NLO and LO Nuclear Parton Distribution Functions". In: *JHEP* 04 (2009), p. 065. DOI: 10.1088/1126-6708/2009/04/065. arXiv: 0902.4154 [hep-ph].

[7] Alfredo Ferrari et al. "FLUKA: A multi-particle transport code (Program version 2005)". In: (Oct. 2005). DOI: 10.2172/877507.

[8] Ivica Friscic et al. "Neutron spin structure from e-3He scattering with double spectator tagging at the electron-ion collider". In: *Phys. Lett. B* 823 (2021), p. 136726. DOI: 10.1016/j.physletb.2021.136726. arXiv: 2106.08805 [nucl-ex].

[9] F. Hauenstein et al. "Measuring recoiling nucleons from the nucleus with the future Electron Ion Collider". In: *Phys. Rev. C* 105.3 (2022), p. 034001. DOI: 10.1103/PhysRevC.105.034001. arXiv: 2109.09509 [physics.ins-det].

[10] Alexander Jentsch, Zhoudunming Tu, and Christian Weiss. "Deep-inelastic electron-deuteron scattering with spectator nucleon tagging at the future Electron Ion Collider: Extracting free nucleon structure". In: *Phys. Rev. C* 104.6 (2021), p. 065205. DOI: 10.1103/PhysRevC.104.065205. arXiv: 2108.08314 [hep-ph].

[11] P. Kaitaniemi et al. "INCL Intra-Nuclear Cascade and ABLA De-Excitation Models in Geant4". In: *Prog. Nucl. Sci. Tech.* 2 (2011). Ed. by Toshikazu Takeda, pp. 788–793. DOI: 10.15669/pnst.2.788.

[12] R. Abdul Khalek et al. *Science Requirements and Detector Concepts for the Electron-Ion Collider: EIC Yellow Report.* 2021. arXiv: 2103.05419 [physics.ins-det].

[13] Davide Mancusi, Robert J. Charity, and Joseph Cugnon. "Unified description of fission in fusion and spallation reactions". In: *Phys. Rev. C* 82 (2010), p. 044610. DOI: 10.1103/PhysRevC.82.044610. arXiv: 1007.0963 [nucl-th].

[14] J. R. Pybus et al. "Generalized contact formalism analysis of the $^4\mathrm{He}(e, e'pN)$ reaction". In: *Phys. Lett. B* 805 (2020), p. 135429. DOI: 10.1016/j.physletb.2020.135429. arXiv: 2003.02318 [nucl-th].

[15] Stefan Roesler, Ralph Engel, and Johannes Ranft. "The Monte Carlo event generator DPMJET-III". In: *Advanced Monte Carlo for radiation physics, particle transport simulation and applications. Proceedings, Conference, MC2000, Lisbon, Portugal, October 23-26, 2000.* Dec. 2000, pp. 1033–1038. DOI: 10.1007/978-3-642-18211-2_166. arXiv: hep-ph/0012252.

[16] C. Salgado and U. A. Wiedemann. "Calculating quenching weights". In: *Phys. Rev. D* 68 (2003), p. 014008. DOI: 10.1103/PhysRevD.68.014008. arXiv: hep-ph/0302184.

[17] Torbjorn Sjöstrand, Stephen Mrenna, and Peter Skands. "PYTHIA 6.4 Physics and Manual". In: *JHEP* 05 (2006), p. 026. DOI: 10.1088/1126-6708/2006/05/026. arXiv: hep-ph/0603175 [hep-ph].

[18] Torbjörn Sjöstrand et al. "An Introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[19] Zhoudunming Tu et al. "Probing short-range correlations in the deuteron via incoherent diffractive J/ψ production with spectator tagging at the EIC". In: *Phys. Lett. B* 811 (2020), p. 135877. DOI: 10.1016/j.physletb.2020.135877. arXiv: 2005.14706 [nucl-ex].

[20] M. R. Whalley, D. Bourilkov, and R. C. Group. "The Les Houches accord PDFs (LHAPDF) and LHAGLUE". In: *HERA and the LHC: A Workshop on the Implications of HERA and LHC Physics (Startup Meeting, CERN, 26-27 March 2004; Midterm Meeting, CERN, 11-13 October 2004).* Aug. 2005, pp. 575–581. arXiv: hep-ph/0508110.