Benchmark math xmlns="http://www.w3.org/1998/Math/MathML" display="inline">mi>e/mi>mi>A/mi>/math> generator for leptoproduction in high-energy lepton-nucleus collisions

Wan Chang, Elke-Caroline Aschenauer, Mark D. Baker, Alexander Jentsch, Jeong-Hun Lee, Zhoudunming Tu, Zhongbao Yin, and Liang Zheng

Phys. Rev. D **106**, 012007 — Published 26 July 2022

DOI: 10.1103/PhysRevD.106.012007

BeAGLE: Benchmark eA Generator for LEptoproduction in high energy lepton-nucleus collisions

Wan Chang,[1, 2, *] Elke-Caroline Aschenauer,[2, †] Mark D. Baker,[3, ‡] Alexander Jentsch,[2] Jeong-Hun Lee,[2] Zhoudunming Tu,[2, §] Zhongbao Yin,[1] and Liang Zheng[4, ¶]

[1] *Key Laboratory of Quark and Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan 430079,China*

[2] *Department of Physics, Brookhaven National Laboratory, Upton, New York 11973, USA*

[3] *Mark D. Baker Physics and Detector Simulations LLC, Miller Place, New York 11764, USA*

[4] *School of Mathematics and Physics, China University of Geosciences (Wuhan), Wuhan 430074, China*

(Dated: April 14, 2022)

The upcoming Electron-Ion Collider (EIC) will address several outstanding puzzles in modern nuclear physics. Topics such as the partonic structure of nucleons and nuclei, the origin of their mass and spin, among others, can be understood via the study of high energy electron-proton (ep) and electron-nucleus (eA) collisions. Achieving the scientific goals of the EIC will require a novel electron-hadron collider and detectors capable to perform high-precision measurements, but also dedicated tools to model and interpret the data. To aid in the latter, we present a general-purpose eA Monte Carlo (MC) generator - BeAGLE. In this paper, we provide a general description of the models integrated into BeAGLE, applications of BeAGLE in eA physics, implications for detector requirements at the EIC, and the tuning of the parameters in BeAGLE based on available experimental data. Specifically, we focus on a selection of model and data comparisons in particle production in both ep and eA collisions, where baseline particle distributions provide essential information to characterize the event. In addition, we investigate the collision geometry determination in eA collisions, which could be used as an experimental tool for varying the nuclear density.

Keywords: EIC, BeAGLE, Monte Carlo Event Generator, Collision Geometry

I. INTRODUCTION

One of the pillars of the Standard Model [1, 2] is the theory of Quantum Chromodynamics (QCD), which describes the mechanism for the interactions between quarks and gluons [3]. It is a self-contained fundamental theory of quark and gluon fields that is rich in symmetries [4, 5]. However, despite the successes of QCD, many fundamental questions remain open to-date, some of which will have to be addressed by a highly anticipated new machine - the Electron-Ion Collider (EIC) [6, 7].

The upcoming U.S.-based EIC is being designed to achieve a wide range of center-of-mass energies from 20 to 140 GeV, ion beams from deuteron to heavy nuclei (e.g. lead), high luminosities of $10^{33-34}\mathrm{cm}^{-2}\mathrm{s}^{-1}$, and highly polarized (70%) electron, proton, and light-ion beams [8]. The EIC will be the world's first dedicated electron-nucleus collider and the first collider to scatter polarized electrons off polarized light ions. The EIC science covers a broad range of topics from detailed investigations of hadronic structure with unprecedented precision to exploring new regimes of strongly interacting matter [9, 10]. The EIC will allow us to investigate the full three-dimensional dynamics of the proton, going well beyond the information about the longitudinal momentum nuclear structure contained in colinear parton distributions. With the unique capability to reach a wide range of momentum transfer Q^2 and Bjorken-x (x_{Bj}) values, the EIC can offer the most powerful tool to precisely quantify how the spin of gluons and that of quarks of various flavors contribute to the proton spin. Another frontier of the EIC science is to understand the formation of nuclei and their partonic structure. Particularly, the nucleus itself is an unprecedented QCD laboratory, where novel nuclear phenomena can be systematically studied by colliding electrons with different nuclear species [6].

However, the challenge of achieving the entire EIC science program via a single machine and a general-purpose detector is also unprecedented. The design of the Interaction Region (IR) and integration of a general purpose collider detector, along with its ancillary detectors over ± 40 meters along the beam-lines requires careful planning. This design has to be guided and optimized via simulations of the physics processes and their kinematics to achieve the optimal placement of the detectors to maximize geometric acceptance, and to aid in identification of the best technologies. Therefore, a general-purpose eA Monte Carlo (MC) model suitable for both investigating the physics and the impact of the machine design is sorely needed.

The BeAGLE (**B**enchmark **eA** **G**enerator for **LE**ptoproduction) general-purpose MC generator simulates eA collisions with the production of exclusive final-state particles, including the fragments from the nuclear remnant breakup process [11]. Prior to the present paper, it has already been used extensively for

* changwan@mails.ccnu.edu.cn
† elke@bnl.gov
‡ mdbaker@mdbpads.com
§ zhoudunming@bnl.gov
¶ zhengliang@cug.edu.cn

exploring physics with final-state particles at pseudo-rapidities > 4.5, e.g., diffractive and spectator-tagging physics, and the associated detector/IR integration requirements for the "far-forward" region (ion-going direction) at the EIC [7]. Key physics topics at the EIC, which are very demanding on far-forward detection, include tagging and vetoing of incoherent Vector Meson (VM) production in ePb collisions to enable studies of gluon imaging in nuclei [12] and tagging of the spectator nucleon in eD scattering to allow for the extraction of free nucleon structure [13], as well as to study Short-Range Correlations [14, 15] in the deuteron [16].

The design of the far-forward detectors and subsequent IR integration issues are urgent at this time because the EIC accelerator design will soon be settled, and the detector technology choices are happening in parallel, with both efforts requiring input from the other. Therefore, in order to maximize the EIC physics output and design the interaction region that is optimized for the aforementioned scientific goals, a reliable MC generator that can describe a wide range of final-states with different kinematic regions is needed. In this paper, we will significantly extend our focus from studies on exclusive observables [12] in the far-forward region to inclusive particle production for both forward and central regions based on BeAGLE simulations. Moreover, we will compare BeAGLE simulations with available fixed-target eA data to further improve the model, and systematically study the model parameter dependence on various observables.

The outline of this paper is as follows. A detailed introduction of BeAGLE is given in Sec. II. In Sec. III, we discuss the validation process on the PYTHIA-6 MC model [17] using the HERA leading proton data [18]. Based on the established PYTHIA parameters, we compare the BeAGLE simulations with fixed target μA data from the E665 experiment [19]. In Sec. IV, we present results from a systematic investigation of the collision geometry, determined via the detection of neutrons from the nuclear breakup. In Sec. V, we describe the future opportunities and challenges of the BeAGLE model. Finally, a summary is provided in Sec. VI.

II. BEAGLE

BeAGLE is a hybrid model that uses modules from DPMJet [20], PYTHIA-6 [17], PyQM [21], FLUKA [22, 23] and LHAPDF5 [24] 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 had to be made for the case when the nucleus is a deuteron; details are described below. The geometric density distribution of nucleons in the nucleus is provided primarily by PyQM while the parton distributions within that geometry are taken from the EPS09

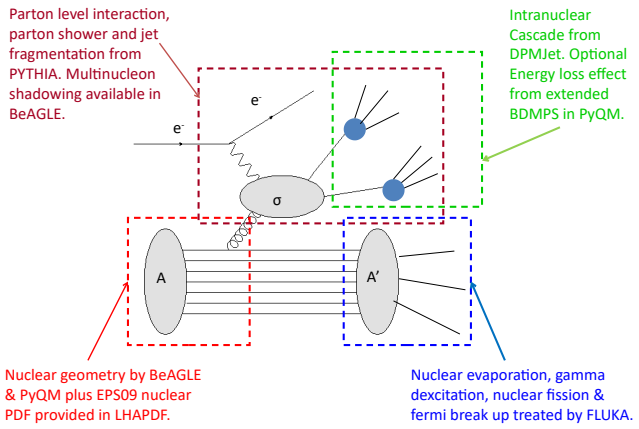

FIG. 1. The BeAGLE MC event generator with its main components, e.g., PYTHIA-6, DPMJet, PyQM, and FLUKA.

nuclear parton distribution functions (nPDFs) [25]. BeAGLE also allows the user to provide "Woods-Saxon" parameters, including non-spherical terms, to override the default geometric density description. The partonic interactions and subsequent fragmentation process is carried out by PYTHIA-6. The optional PyQM module implements the Salgado-Wiedemann quenching weights to describe partonic energy loss [26]. Hadron formation and interactions with the nucleus through an intra-nuclear cascade are described by DPMJet. The decay of the excited nuclear remnant is described by FLUKA, including nucleon and light ion evaporation, nuclear fission, Fermi breakup of the decay fragments, and finally de-excitation by photon emission. See Fig. 1 for an illustration and the User's Guide here: `https://eic.github.io/software/beagle.html`.

Due to the structure of the BeAGLE generator coherent diffraction is currently not included. Since the primary interaction is modeled by PYTHIA-6 at the nucleon level, for any nuclear beam, the target nucleus will break up or at least be excited in the final state. Furthermore, for diffractive interactions, the lepton-nucleus cross section is assumed to be A times the lepton-nucleon cross-section, rather than calculated from first principles. As observed in the data in ep collisions at HERA, coherent diffraction in DIS was found to be 15% of the total inclusive DIS cross section [27], while in the nucleus, it has been predicted that coherent diffractive processes can be enhanced due to possible gluon saturation effects at high energy [28]. Measurements of coherent diffraction in nuclei are expected to be one of the golden channels to study non-linear QCD effects[29] at the EIC.

In this framework, the lepton-nucleus collision can be illustrated in several steps as follows:

A. The collision is simulated by selecting a struck nucleon in the nucleus according to a Glauber-type model, where the nucleon level cross sec-

tion is weighted by the EPS09 nPDFs leading to an event at the partonic level; optional gluon radiation by PyQM [30], accounting for nuclear medium effects, is available; finally, the fragmentation/hadronization is performed with the Lund string model provided by PYTHIA-6;

B. Hadrons produced during the previous stage participate in a "formation-zone" Intra-Nuclear Cascade (INC) [31], which produces secondary particles;

C. The breakup of the excited nuclear remnants will be treated by the FLUKA model.

A. Hard interactions and Fermi momentum

Initial nucleons are placed in coordinate space according to the Woods-Saxon distribution [32] with intrinsic Fermi momentum, some of which will be struck off the nucleus by the exchanged virtual photon emitted by the lepton. The corresponding nucleon level cross section, $\sigma_{\gamma^* N}(x, Q^2)$, is obtained from PYTHIA-6, where the magnitude of $\sigma_{\gamma^* N}(x, Q^2)$ is parametrized such that the $\sigma_{\gamma^* A}/(A\sigma_{\gamma^* N})$ follows the EPS09 nuclear modification factor $R(x, Q^2) = f^A(x, Q^2)/f^p(x, Q^2)$ [25]. This scaling feature based on the nPDFs at the cross section level enables general studies of nuclear effects.

For the hard scattering between a virtual photon and the struck nucleon discussed above, three different options of an initial collision geometry, including multiple nucleon scattering and shadowing effects [33–35], are available. The BeAGLE framework also allows a user-defined parameter, *genShd*, to switch between the different modes: i) $genShd = 1$, only one nucleon is probed by the virtual photon and participates in the primary scattering simulated by PYTHIA-6; ii) $genShd = 2$, if the impact parameter between the virtual photon and any nucleon is less than a distance, $D = \sqrt{\sigma_{\gamma^* N}/\pi}$, one random selected nucleon will be simulated by PYTHIA-6 for an inelastic interaction, while the other nucleons will undergo elastic interactions; iii) $genShd = 3$ is the same as $genShd = 2$ except the order is fixed such that the first struck nucleon always undergoes inelastic scattering, while the rest scatter elastically.

Due to nuclear binding, nucleons inside of a nucleus have internal momentum, commonly known as Fermi motion [36]. In BeAGLE, we adopt a non-relativistic model of the nucleon spectral function, provided by Ref. [37]. This parametrization applies to all nuclei, ranging from deuterons to heavy nuclei, e.g., lead (Pb). For the case of the deuteron, the parametrization has been extended with the Light-Front formalism by Strikman & Weiss [38]. Details from recent BeAGLE deuteron studies can be found in Refs. [13, 16].

The nucleon momentum distribution in the ion rest frame is parametrized as follows,

$$n(k) = n_0(k) + n_1(k). \tag{1}$$

For nucleus $A = 2, 3, 4$,

$$n_0(k) = \sum_{i=1}^{m_0} A_i^{(0)} \frac{e^{-B_i^{(0)} k^2}}{(1 + C_i^{(0)} k^2)^2}. \tag{2}$$

For nucleus $A > 4$,

$$n_0(k) = A^{(0)} e^{-B^{(0)} k^2} [1 + C^{(0)} k^2 + D^{(0)} k^4 + E^{(0)} k^6 + F^{(0)} k^8]. \tag{3}$$

Here k is the internal nucleon momentum and $A^{(0)}$ to $F^{(0)}$ are parameters given in Table A1-A2 in Ref. [37]. Note that $n_0(k)$ describes the low momentum part of the wave function, or the Mean-Field region, while $n_1(k)$ describes the high momentum tail, known as the Short-Range Correlation region. Currently, only $n_0(k)$ has been implemented in BeAGLE. However, for the deuteron case, $n(k) = n_0(k)$, where Short-Range Correlations in the high momentum tail have been studied in Ref. [16]. In addition, for $A > 4$, the parametrizations are based on a few typical nuclei A_{typical}, e.g., Carbon-12, Oxygen-16, Calcium-40, Iron-56, Lead-208, and above. Any nucleus between them, A_{select}, will use one of the nearest typical nuclei for the mass number, such that $A_{\text{select}} < A_{\text{typical}}$. Differences in $n_0(k)$ for various mass numbers A are generally small for $A > 12$.

BeAGLE currently does not account for the Fermi motion in the DIS cross section calculations, where kinematic distributions, e.g., x and Q^2, are unmodified from the PYTHIA-6 generator. Accounting for the Fermi momentum would violate energy-momentum conservation because the primary interaction simulated by PYTHIA-6 assumes an on-shell nucleon mass. The higher the off-shell mass as determined from the nuclear wave function, the more violation in energy and momentum it will cause. In order to correct for this artifact, the excess energy and momentum are absorbed by the remnant nucleus.

However, in the case of deuteron (or light nuclei in general) this correction will not be reasonable because there is only one spectator nucleon (or a few spectator nucleons) in the system. The correction would artificially distort the spectator momentum distribution. Therefore, for the deuteron, we leave the spectator unmodified, where the energy and momentum are corrected by the outgoing particles from the current fragmentation[1]. By using this approach, the spectator tagging and related physics topics can be studied with the genuine information from the wave function. No Final-State Interactions (FSI) are present in the BeAGLE generator for lepton-deuteron collisions.

[1] This correction can be switched on and off in the BeAGLE control card.

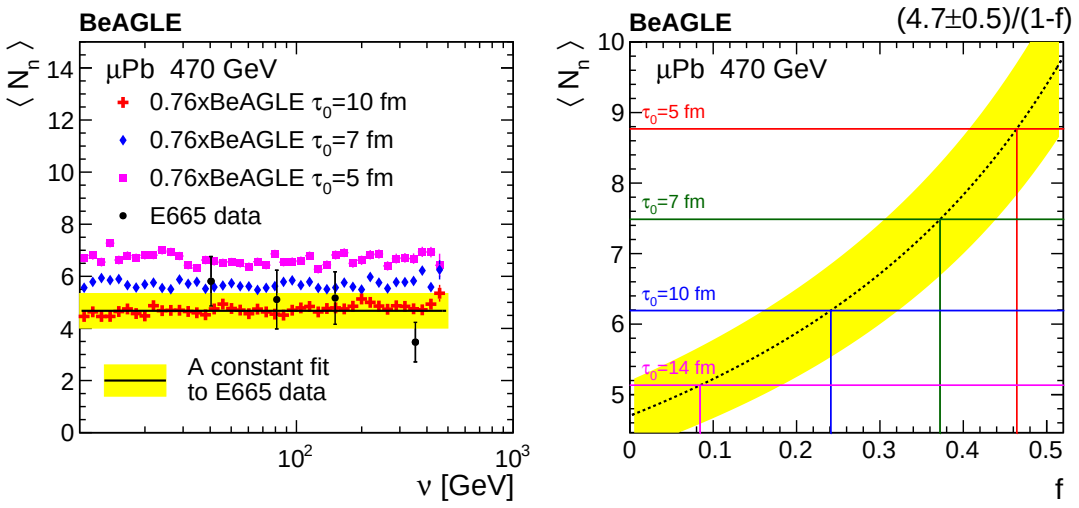

FIG. 2. (Left) Neutron multiplicity as a function of ν for μPb collisions with 470 GeV muons. The results with different values of τ_0 from the BeAGLE generator with a weight ($f = N_{\rm coherent}/N_{\rm total}$, see text for details) of 0.76 represented by different color markers. The neutron data from Ref. [39] are represented by the black points, the black solid line is the result of a fit to a constant function for the data result, the band shows the statistical uncertainty from the fit. (Right) The average neutron multiplicity $\langle N_n \rangle$ vs. f for a variety of different values of τ_0 in the BeAGLE model.

B. Intra-nuclear Cascade

The generated particles from the primary scattering will be placed at the struck nucleon position and transported through the INC following the formation zone formalism implemented in DPMJet [20]. Each primary particle is assigned a formation time sampled from an exponential distribution with the characteristic time scale τ [40, 41] defined in the lab frame as follows:

$$\tau = \tau_0 \frac{E}{m} \frac{m^2}{m^2 + p_\perp^2}, \tag{4}$$

where E, m and $p_\perp$ are the energy, mass, and transverse momentum of the produced particle, respectively. The parameter τ_0 is treated as a free parameter to be determined/tuned by the experimental data.

These produced particles can induce secondary interactions (a cascade process) if they are formed inside the nucleus. Particles with higher energy and smaller transverse mass are more likely to be formed outside and leave the nucleus without a secondary interaction. The value of the τ_0 parameter has been systematically extracted from the experimental data in our previous publication [12], and in the present work.

Since forward neutron production from the evaporation process is sensitive to the INC, we use the multiplicity data of neutron emission in μPb collisions from the E665 experiment at Fermilab [39] to tune the τ_0 parameter. BeAGLE does not simulate coherent diffractive events, which do not produce neutrons in the final state. However, the E665 data do include contributions from the coherent diffractive process. In order to properly use the neutron multiplicity to determine the τ_0 parameter, a weight ($f = N_{\rm coherent}/N_{\rm total}$) is needed for the BeAGLE model to account for the coherent contribution in the cross section data.

$$N_n(\text{E665}) = 0 * f + N_n(\text{BeAGLE}) * (1 - f). \tag{5}$$

Note that the coherent event fraction f was not explicitly determined in the E665 experiment [39], thus a few assumptions on f were made in order to determine the value of τ_0.

Figure 2 (left) shows the average neutron multiplicity $\langle N_n \rangle$ as a function of photon energy ν as measured by the E665 experiment and simulated by BeAGLE with $f = 24\%$, where different values of τ_0 are presented. A constant fit is performed to the E665 data, where the yellow band shows a statistical uncertainty corresponding to one standard deviation. With the assumption of $f = 24\%$, the best value of τ_0 is found to be 10 fm. The $\langle N_n \rangle$ from E665 is found to be 4.7 ± 0.5. The choice of f is inspired by HERA measurements [27], where one finds a large fraction of diffractive events, contributing about 15% to the total deep inelastic cross-section for ep collisions [42, 43]. However, theoretical studies, e.g., in Ref [28], indicate that the ratio of diffractive events to the total cross section in eA could be larger than what is observed in ep collisions, due to non-linear QCD effects.

In Fig. 2 (right), the average neutron multiplicity as a function of f is presented, where the dotted line with the yellow band represents a match to the E665 data with different assumptions for f, given by $(4.7 \pm 0.5)/(1 - f)$. The straight colored lines show a few selected τ_0 values with their corresponding $\langle N_n \rangle$ in BeAGLE. As shown in Fig. 2 (right), for $\tau_0 = 14$ fm, the corresponding f value is

less than that in ep collisions at HERA, while for $\tau_0 = 5$ fm or 7 fm, the corresponding f needs to be larger than 0.3, which exceeds current theoretical predictions [28, 44]. Therefore, we use $\tau_0 = 10$ fm as the default setting in the BeAGLE model, while in the following analysis we perform systematic studies using other τ_0 values.

C. Nuclear remnant breakup

After all possible secondary interactions are exhausted, excitation energies of the nuclear remnant can be calculated by summing up the recoil momenta transferred to the remnant by the particles leaving the nuclear potential. The breakup of the nuclear remnant is modeled using fission, the evaporation of nucleons and light nuclei, and photon emissions within the FLUKA machinery [22, 23] for a given excitation energy. Since FLUKA is not an open-source program, the BeAGLE event generator has no handle on changing the evaporation process and can only adjust the INC in the previous step.

III. DATA AND MC COMPARISON

In this section, comparisons of experimental data and the BeAGLE MC will be presented. We start with the case of ep by using the PYTHIA-6 model, which is independent of the eA modeling in BeAGLE. The target fragmentation of the leading proton distribution has been investigated, and a good set of baseline parameters regarding the nucleon target fragmentation are established. The leading proton data are based on the ZEUS experiment at HERA [45]. These improvements on the PYTHIA-6 parameters will be used later in the BeAGLE event generator. After that, we will show a comparison of BeAGLE and E665 data [19] for inclusive charged particle rapidity distributions in both the forward and backward regions for μXe collisions.

A. Comparison between PYTHIA-6 and ep data at ZEUS

Since PYTHIA-6 is used to model the primary interaction in BeAGLE, it is crucial to optimize the parameters used in this stage of the framework. Leading proton data collected by the ZEUS experiment at HERA [45] were examined in order to optimize the PYTHIA parameters for the fragmentation $p_{\rm T}$ and intrinsic $k_{\rm T}$. Three parameters were investigated in PYTHIA and are detailed in Table I. MSTP(94) controls the energy partitioning in the beam remnant cluster decay. The default value of 3 uses the regular fragmentation function, while MSTP(94)=2 uses the function $P(\chi) = (k+1)(1-\chi)^k$, where χ is the light cone energy fraction taken by the hadron or diquark. The fragmentation function corresponding to MSTP(94)=2 and PARP(97)=6 is, $P(\chi) = 7(1-\chi)^6$. PARJ(170) is a parameter which we added to PYTHIA to allow separate control of the Gaussian rms $p_{\rm T}$ for hadrons in the recoil, which in standard PYTHIA is set to be the same as that for the string fragmentation: PARJ(21).

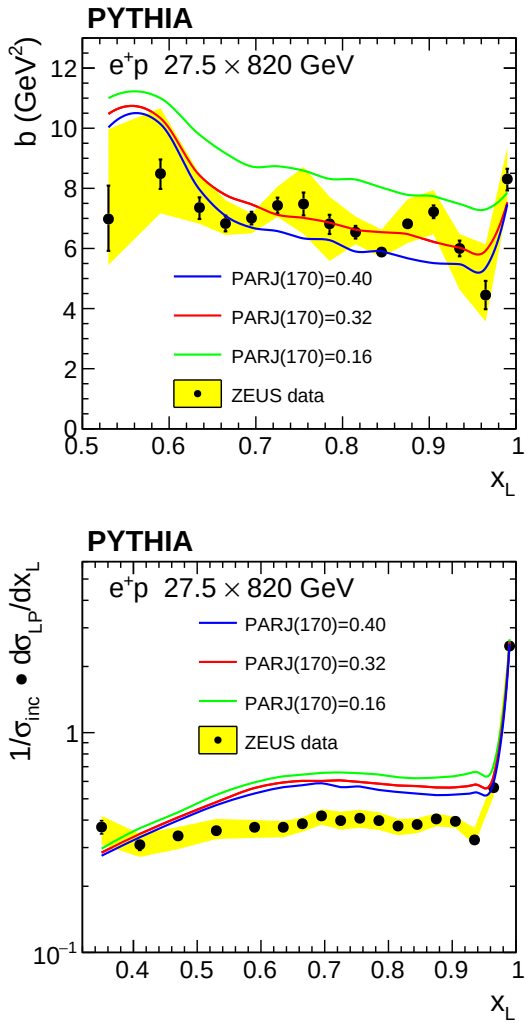

FIG. 3. (Top) The $p_{\rm T}^2$-slope, b, of the cross-section $d^2\sigma_{\rm LP}/dx_L dp_{\rm T}^2$ of leading proton, as defined by the parameterisation $A \cdot e^{-b \cdot p_{\rm T}^2}$ and obtained from a fit to the data in bins of x_L. (Bottom) Single differential cross section of the leading proton normalized to the total DIS cross section $1/\sigma_{\rm inc} \cdot d\sigma_{\rm LP}/dx_L$. The MC results represented by lines are compared to data from Ref. [45].

Figure 3 shows comparisons of different calculated distributions with the measurements of leading protons from the ZEUS experiment [45] for positron-proton scattering with beam energies of $E_e = 27.5$ GeV and $E_p = 820$ GeV. In Ref. [45], the semi-inclusive reaction $e^+p \to e^+Xp$ was studied with the ZEUS detector with an integrated luminosity of 12.8 pb^{-1}. The final-state proton, carrying a large fraction of the incoming proton energy, but a small transverse momentum, was detected by the ZEUS leading proton spectrometer (LPS). The selection of the LPS

TABLE I. The parameters of PYTHIA tuned by ZEUS leading proton data at HERA.

Parameter	Default	Tuned
MSTP(94)	3	2
PARP(97)	-	6
PARJ(170)	-	0.32

proton sample requires a dedicated LPS trigger and acceptance cuts to omit tracks very close to beam line or the edge of LPS detector.

These measurements were carried out in the kinematic range $Q^2 > 3$ GeV$^2, 45 < W < 225$ GeV$, y > 0.03$ and the leading proton is measured with $p_{\rm T}^2 < 0.5$ GeV$^2, x_L > 0.32$, where x_L is the longitudinal momentum fraction of the measured proton and the incoming proton beam momentum. Figure 3 (top) shows the $p_{\rm T}^2$-slope, b, of the cross-section $d^2\sigma_{\rm LP}/dx_L dp_{\rm T}^2$ for leading protons, as defined by the parametrization $A\cdot e^{-b\cdot p_{\rm T}^2}$ and obtained from a fit to the data in bins of x_L. The black points represent the data, while the yellow band is the experimental systematic uncertainty [45]. The other colored lines represent the distributions for different values of PARJ(170) in PYTHIA, corrected for the leading proton spectrometer (LPS) acceptance effects. The distribution of the single differential cross section normalized to the total DIS cross section $1/\sigma_{\rm inc}\cdot d\sigma_{\rm LP}/dx_L$ is shown in Fig. 3 (bottom), and the $\sigma_{\rm inc} = 223.9$ nb. It is found that the optimal parameter in this comparison is PARJ(170) = 0.32.

There are two additional parameters that are sensitive to the leading proton distribution: i) PARJ(21), the width of the transverse momentum distribution in the fragmentation, and ii) PARP(91), the Gaussian width of the intrinsic $k_{\rm T}$ distribution. We find that the result with PARJ(21) = PARP(91) = 0.32 agrees best with the ZEUS leading proton data. However, a value of 0.4, tuned to data collected by the HERMES experiment at HERA [46, 47], does a better job in the current fragmentation region. It is noted that to describe the European Muon Collaboration (EMC) data [48], higher values of the fragmentation $p_{\rm T}$ and intrinsic $k_{\rm T}$ are preferred. We use PARJ(21) = PARP(91)= 0.4 as the default settings for results presented in this study. The summary of the PYTHIA parameters used in this paper which are different from the default values are listed in the Appendix in Table II.

B. Comparison between BeAGLE and the E665 μXe data

A challenge in validating the BeAGLE generator is that there are only limited eA collision data available to compare with. The best available data are measurements of particle production from the E665 experiment [19] at Fermilab. In Ref. [19], the data were collected with the E665 spectrometer, which used the 490 GeV muon beam from the Tevatron at Fermilab. The experiment used a streamer chamber as a vertex detector, providing nearly 4π acceptance for charged particles. Results of charged hadron production in muon-xenon (μXe) and muon-deuteron (μD) collisions [19] are used to compare with the BeAGLE model. The general picture of the interaction is the virtual photon, emitted by the incoming muon, interacts with a parton of a nucleon in the target nucleus. The hadronic center-of-mass frame (cms) is defined as the system formed by the virtual photon and the target nucleon: the struck parton is scattered into the forward direction, while the target remnant travels into the backward direction.[2] Given a limited particle identification capability in the E665 experiment [19], all positively charged hadrons in the data with $x_F(m_\pi)$ less than -0.2 are assigned the proton mass, while all other positively and negatively charged hadrons are treated as pions. The variable, $x_F(m_\pi)$, is defined as $x_F = 2p_L^*/W$, with p_L^* being the longitudinal momentum of the hadron in the cms frame, assuming all particles are pions. In order to properly compare BeAGLE simulations with the results from Ref. [19], the partial identification of particles is performed in the same way on the BeAGLE simulated events. The version of the BeAGLE event generator used here is 1.01.03. The BeAGLE control card is shown in Table III in the Appendix.

Figure 4 shows the distributions of $W, Q^2, \nu, x_{\rm Bj}$ for BeAGLE compared with the E665 μXe data [19]. The positive muon beam with an energy of 490 GeV is scattered off a Xe target. A set of kinematic cuts had to be applied to select events: $\theta > 3.5$ mrad, $Q^2 > 1$ GeV^2, $8 < W < 30$ GeV, and $0.1 < \nu/E_\mu < 0.85$. The red solid lines represent the generated MC events from BeAGLE, while the E665 data [19] after correcting for detector acceptance effects are shown in black open circles. The ratio between the BeAGLE data and the corrected E665 results are shown in the bottom of each plot. The comparison shows that BeAGLE can do a reasonable job of describing the kinematics of the E665 data, while large deviations can be seen at the small $x_{\rm Bj}$ and high Q^2.

Figure 5(a) and (b) show the average multiplicity, $\bar{n}$, of positively and negatively charged hadrons produced in μXe collisions with 490 GeV muons. BeAGLE simulations with different values of τ_0 are compared with the E665 data. It is found that the distribution for negatively charged hadrons does not show any τ_0 dependence in the BeAGLE model and the E665 data are underestimated. For positively charged hadrons, the average multiplicity increases with decreasing τ_0. A lower τ_0 value, e.g., 2–3 fm, reproduces the data better, indicating a contradiction

[2] Note that in collider physics, the terminology of forward and backward is reversed with respect to the fixed target experiments. For comparison to the E665 data, we adopt the convention of fixed target experiments.

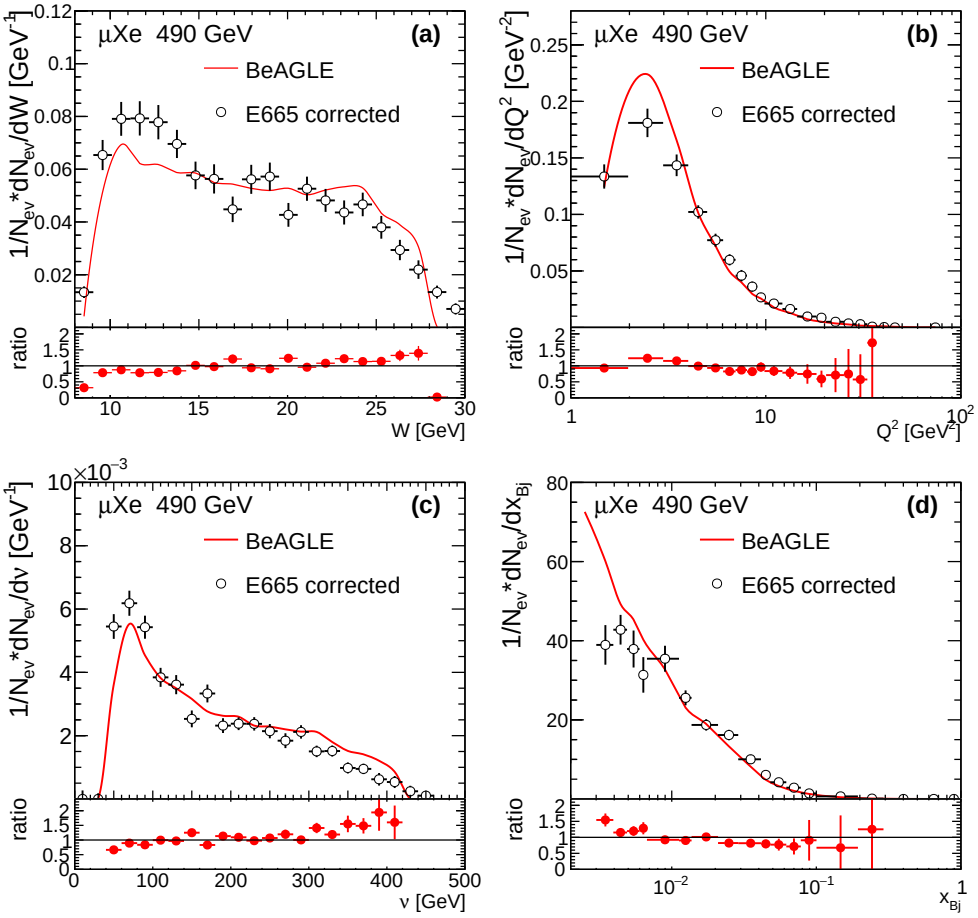

FIG. 4. Distributions of $W, Q^2, \nu, x_{\rm Bj}$ for BeAGLE compared to the E665 data [19] for μXe events.

with respect to our default value of $\tau_0 = 10$ fm, which was determined from the evaporated neutron multiplicity data (see Fig. 2 (right)). Lower τ_0 values in Fig. 2 (right) would suggest a very large fraction of diffractive events in μA collisions. However, the discrepancy in negatively charged particle production needs to be considered, for a clear understanding of the τ_0 dependence.

In Fig. 5(c) and (d), the average charged particle multiplicity for positive and negative y^* are shown based on BeAGLE simulations. Different τ_0 parameters in the BeAGLE model and E665 data are also presented for comparison. The distribution for charged hadrons from the BeAGLE simulations underestimates the E665 data. However, with a lower τ_0 value, the average multiplicity distributions for charged hadrons in the target fragmentation region are improved. The quantitative dependence on τ_0 is similar to that of the positively charged hadrons shown in Fig. 5(a), as positively charged hadrons dominate in the target fragmentation region.

We find that the BeAGLE model underestimates the multiplicity everywhere, especially for negatively charged particles, and in the current fragmentation region, where neither of them has a τ_0 dependence. Although the data may suggest a lower τ_0 parameter in μA collisions, this comparison also implies that something other than τ_0 plays an important role in the particle production. Recent results from the H1 experiment at HERA have reported a measurement of the charged particle multiplicity distribution [49] in a wide range of DIS kinematics, where the PYTHIA-8 model [50, 51] also underestimates the data almost everywhere. While a separate analysis on this subject in ep DIS is highly important, e.g., a Rivet analysis [52], the analysis in this paper is only focused on parameters sensitive to nuclear effects.

In order to quantitatively understand the difference between the small amount of available experimental data and the BeAGLE model, we investigate the particle production in a differential way. In Fig. 6, the normalized cms-rapidity y^* distributions for positively and negatively charged hadrons are shown in μXe (top row (a) and (b)) and μD (bottom row (c) and (d)) with 490 GeV muon beams. The selected kinematic phase space is within $14 < W < 20$ GeV. In μXe events, for positively charged hadrons, there is no τ_0 dependence found at forward rapidities, while a strong dependence is observed in the backward region. In the E665 y^* distri-

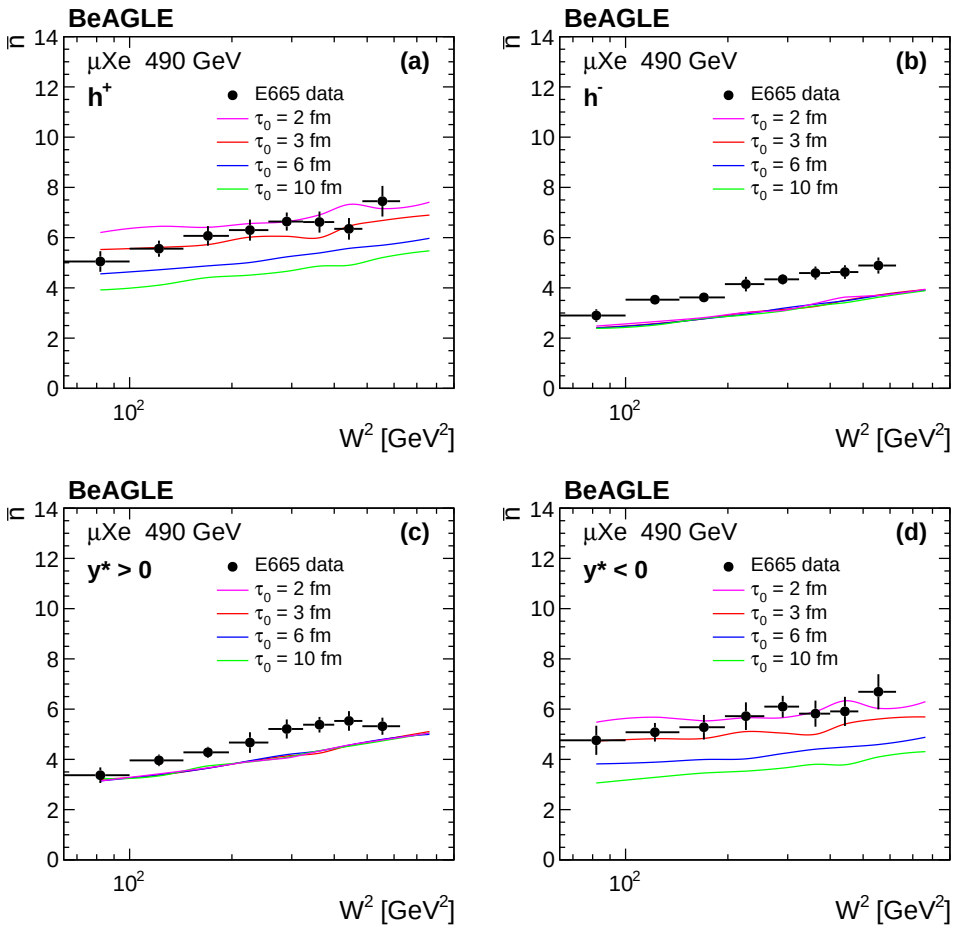

FIG. 5. Multiplicity distributions as a function of W^2 for positively and negatively charged hadrons (top row (a) and (b)) and for positive and negative y^* (bottom row (c) and (d)) in μXe interactions at 490 GeV. The results are shown with different values of the parameter τ_0 in the BeAGLE generator, represented by different colors, and are compared to data from Ref. [19].

bution comparisons with BeAGLE simulations shown in Fig. 6, BeAGLE underestimates the forward particle production, and predicts a different peak position of the backward production. Additionally, BeAGLE underestimates the negatively charged particles and all charged particles in μD almost everywhere in rapidity except for the very forward and backward regions. For both μXe and μD systems, similar observations are found in other W ranges, thus it is observed that the discrepancies between the data and BeAGLE are not dependent on the kinematics.

To further isolate the contributions from the primary interaction and the nuclear remnant fragmentation, we study the difference between positively and negatively charged particles, which is more sensitive to effects like INC. Therefore, the normalized cms-rapidity y^* distributions of the net charge, $\rho^+(y^*) - \rho^-(y^*)$, are shown in Fig. 7 for both μXe and μD collisions, where $\rho^\pm$ is defined as follows,

$$\rho^\pm(y^*) = \frac{1}{N_{ev}} \cdot \frac{dN^\pm}{dy^*}. \quad (6)$$

Here N_{ev} is the number of selected events and $N^\pm$ is the number of positively or negatively charged hadrons, respectively. In μXe events, for charged hadrons, there is no τ_0 dependence of the $\rho^\pm$ distribution at forward rapidity or in the current fragmentation region, similar to what has been found in Fig. 6. However, in the backward region, despite the large τ_0 dependence in the BeAGLE model, the peak position of the distribution is found to be stable for all τ_0 values, and is different compared to the E665 data by about half a unit of rapidity. In μD events, τ_0 dependence is hardly visible, while the shift in the peak position in the backward region is even larger than that in μXe events.

Comparisons of the normalized y^* distributions between μXe and μD collisions for positively charged hadrons are presented in Fig. 8 for both E665 data and BeAGLE simulations. Here, BeAGLE uses a τ_0 of 3 fm but shows different assumptions for final-state nuclei. The discrepancy still exits in the backward region, where the peak positions from BeAGLE sit at larger negative values of y^* compared to the E665 data. Since there is no clear description of remnant nuclei detection in Ref. [19],

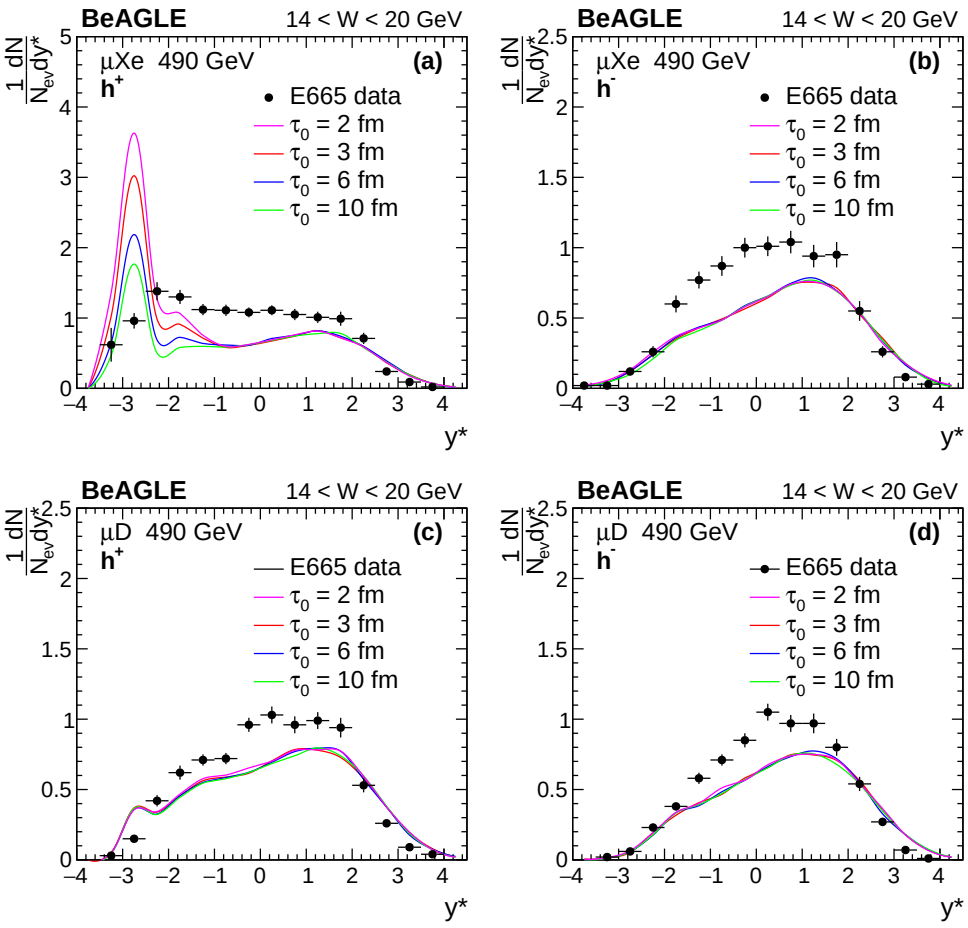

FIG. 6. Rapidity distribution in cms frame y^* for positively and negatively charged hadrons in W bin: $14 < W < 20$ GeV in μXe (top row (a) and (b)) and μD (bottom row (c) and (d)) interactions at 490 GeV. The results with different value of τ_0 from BeAGLE generator represented by different colors are compared to data from Ref. [19]. The comparison results in $8 < W < 14$ GeV and $20 < W < 30$ GeV are the same as that in $14 < W < 20$ GeV.

we try a few different ways to treat the remnant nuclei in the BeAGLE simulation. The red line is the result with all remnant nuclei included. The magenta line denotes a randomly selected 50% of all nuclei. The blue line, green line, and orange line represent only nuclei whose mass number A is smaller than 4, a random selection of 25% of all nuclei, and no nuclei, respectively. With different fractions of nuclei included, the net charge density in the region of $-4 < y^* < -2$ changes, while the peak position remains the same.

For the comparisons presented above, specifically from Fig. 5 to 8, a few things should be noted. First, the particle identification in the E665 experiment assigned either a proton or pion mass in the reconstruction, based on the x_F value. In the absence of a more precise particle identification (PID) methodm such as those currently in use (e.g. dE/dx or time-of-flight measurements), this approach might be problematic. Figure 9 shows the y^* distributions of π^+ mesons (upper) and protons (lower) in μXe collisions using a 490 GeV muon beam from the BeAGLE generator. The red points, which are labeled as "E665 selection", represent the same method as the E665 data from Ref. [19], while the blue curves represent the result based on the true mass from the MC PID, and the magenta curves assume a wrong mass assignment, e.g., proton mass for pions (top), and pion mass for protons (bottom). Pions are mostly produced in the hard scattering and dominate the current fragmentation region. A large proportion of protons are generated during the INC process and its y^* distribution is dominated in the region of $-3 < y^* < -2.5$. If the protons were mis-identified as pions, the cms-rapidity y^* would be shifted toward more-central values of rapidity. Although the data were fully corrected at the particle level, residual mis-identification in the data and a subsequent discrepancy between the data and BeAGLE in particle compositions are possible. Secondly, the E665 measurement from Ref. [19] did not explicitly describe the details of experimental detection of remnant nuclei. Although the peak position of μXe $-$ μD shown in Fig. 8 remains the same, the details of remnant nuclei detection together with a different particle composition as described above may cause the peak

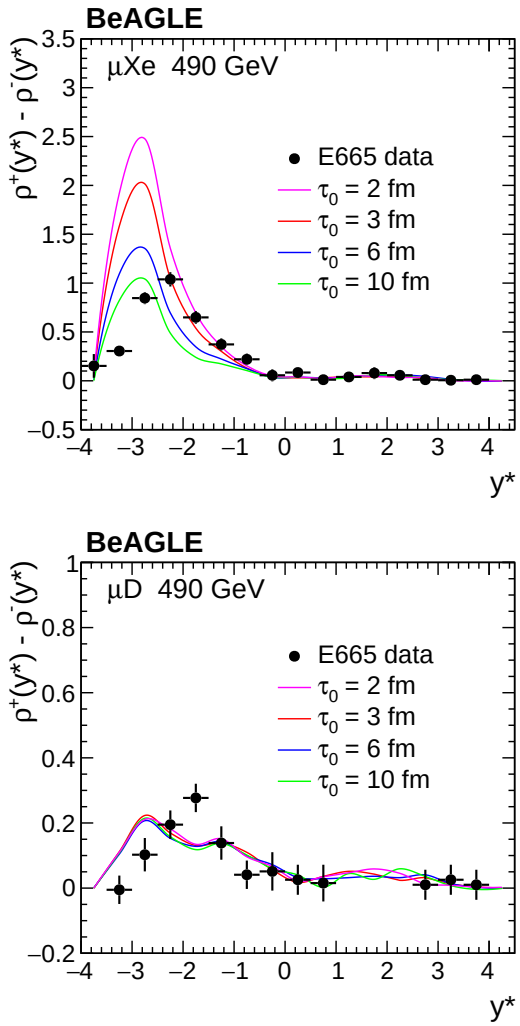

FIG. 7. Normalized cms-rapidity y^* distribution of the hadronic net charge in μXe (top) and μD (bottom) collisions using 490 GeV muon beams. The results with different values of τ_0 from the BeAGLE generator represented by lines with different colors are compared to data from Ref. [19].

position of the distribution change. Finally, the missing coherent diffractive events in the BeAGLE model could be another reason for the observed discrepancy. Naively, the diffractive DIS events would have a rapidity gap, and the y^* distribution would be expected to be shifted more towards the forward than the backward direction.

In addition to Ref. [19], a similar result was reported by the E665 Collaboration in Ref. [53]. In this study, it employed so-called "gray tracks" to enhance proton identification. "Gray tracks" are particles whose momenta are between 200-600 MeV/c, and the streamer density as observed in the streamer chamber picture is clearly higher than that of a minimum ionizing particle. Unfortunately the data reported in Ref. [53] were not corrected for experimental inefficiencies and there is no reliable method to study the impact of such gray tracks

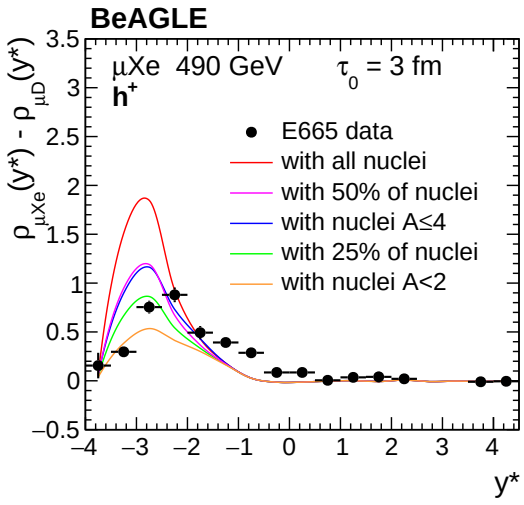

FIG. 8. The difference of normalized y^* distributions for positively charged hadrons between μXe and μD. The results from BeAGLE generator are with $\tau_0 = 3$ fm. The lines with different color represent the results with different way treated to nucleus measurement. Black points represent the data from Ref. [19].

in our simulations. In light of these challenges, a truly equivalent comparison between existing data and BeAGLE cannot be made. Therefore, in order to further understand particle production over a wide range of rapidity, only the EIC can provide more information about the target fragmentation in lepton-nucleus collisions.

Figure 10 shows the normalized distribution of positively charged particles as a function of pseudorapidity (η) at the top EIC energy, simulated by BeAGLE. The total distribution includes all particle species, depicted in the black curve. Other colors indicate distributions for different particle species. Almost all pions and kaons are produced during the hard collision, and their pseudorapidities range from -4 to 4, which falls into the acceptance of the expected general purpose detector of the EIC. Protons are distributed across a wide range of pseudorapidity, from -4 to 10, where three different far-forward proton detectors (B0 tracker, Off-momentum detector, and Roman Pots) can cover a large fraction of the phase space of pseudorapidity > 4.5 [7]. Nuclei are produced in the last step of the BeAGLE proceeses via evaporation, but they are separated into two kinematic regions. The nuclei distributed within $7 < \eta < 10$ are light nuclei, e.g., deuterons and alpha particles. The large remnant nuclei are distributed within $10 < \eta < 15$. Detecting these nuclei is a major experimental challenge, and is one of the on-going efforts at the future EIC, hopefully achieved through optimizing the far-forward instrumentation and the 2nd IR design [7].

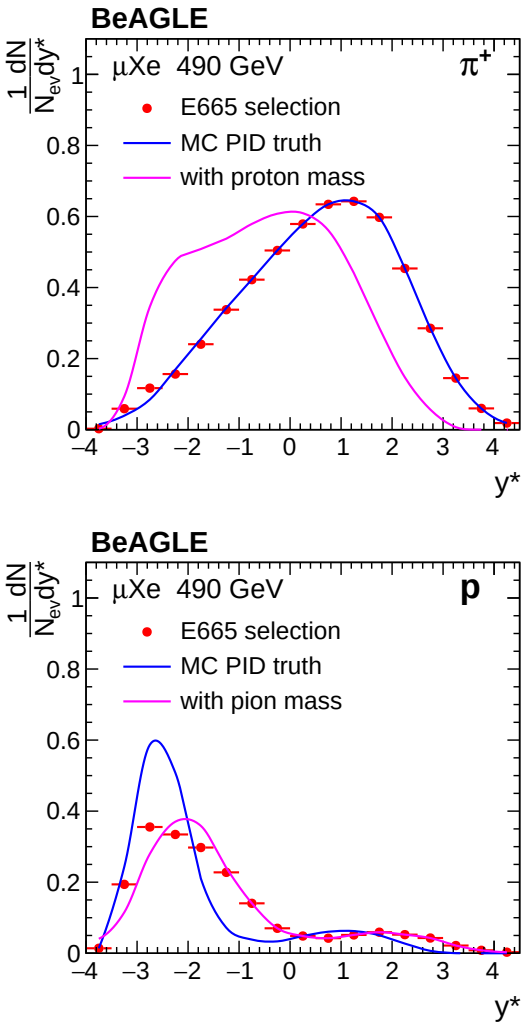

FIG. 9. Normalized cms-rapidity y^* distribution of π^+ (top) and proton (bottom) in μXe collisions using 490 GeV muon beams. Different colors represent the results with different methods of calculating cms-rapidity y^* from the BeAGLE generator.

IV. COLLISION GEOMETRY DETERMINATION IN LEPTON-NUCLEUS INTERACTIONS

In this section, we show an example of how BeAGLE can help optimize measurements with different collision geometry in lepton-nucleus interactions at the future EIC. Precise quantification of the nuclear effects in eA collisions requires knowledge of the underlying collision geometry. In fixed target DIS experiments of nuclei up to now, the collision geometry has only been qualitatively investigated by varying the target nucleus. However, at the EIC, it is possible to characterize an event-by-event collision geometry by studying the nuclear breakup, an idea initially introduced in Ref. [41]. The collision geometry in each event can be linked to the multiplicity

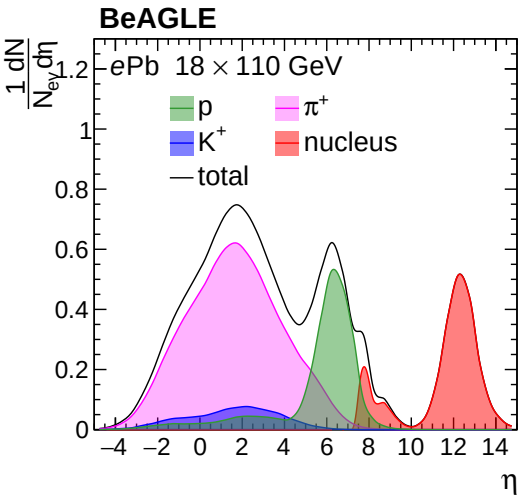

FIG. 10. Distribution of normalized pseudo-rapidity of positive charged particles for ePb collisions with 18 GeV on 110 GeV with BeAGLE. Different colors indicate results for different particle species.

of evaporation neutrons at very forward rapidities ($>$ 4.5), measured by the Zero-Degree Calorimeter (ZDC) (see Ref. [7] for details). In the following, we will introduce variables that are sensitive to the collision geometry and their correlation with experimental observables, e.g., evaporated neutrons and protons. Compared to Ref. [41], we provide more systematic studies by varying model parameters using the BeAGLE generator to demonstrate the robustness of this measurement. This result provides an important experimental handle to all inclusive and semi-inclusive DIS measurements at the EIC.

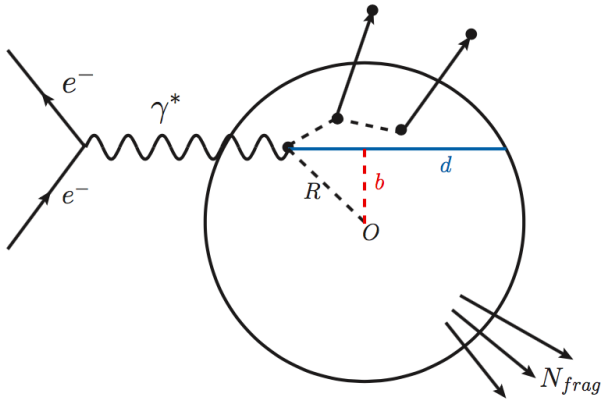

FIG. 11. Relevant quantities to describe the collision geometry. The effective interaction length, d, is the distance between the photon-nucleon interaction point and the edge of the nucleus in the direction of the virtual photon, weighted by the nuclear density. The variable b is the impact parameter between d and the center of the nucleus.

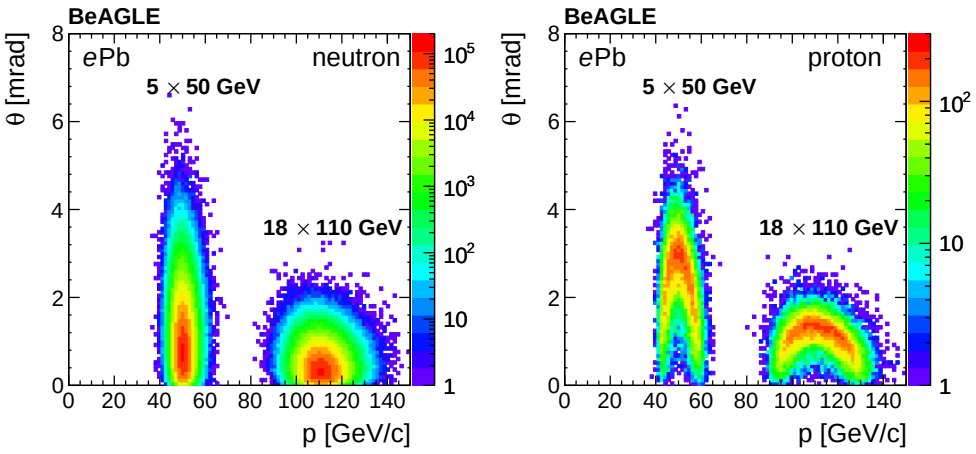

FIG. 12. The correlation of momentum and scattering angle for two beam energies for neutrons and protons.

A. Definition of eA collision geometry

The collision geometry in DIS reveals the underlying spatial information of the nuclear matter probed by the exchanged virtual photon with respect to the rest of the nuclear target. Depending on the physics process under study, different collision geometry quantities can be defined. In Ref. [41], the fiducial traveling length and the impact parameter were used as important controls to quantify the effect of parton energy loss and gluon saturation. In this paper, we define the effective interaction length, d, as the distance between the photon-nucleon interaction point and the edge of the nucleus in the direction of the virtual photon, weighted by the nuclear density ρ_0,

$$d(b, z_0) = \int_{z_0}^{+\infty} \mathrm{d}z \rho(b, z)/\rho_0. \tag{7}$$

Here, z_0 is the position of the nucleon involved in the scattering along the photon moving direction, and b is the impact parameter. If multiple nucleons are participating in the interaction, we use the effective interaction length averaged over all the involved nucleons. This definition avoids the possible negative region[3] of fiducial d used in Ref. [41] and is more directly connected to the amount of nuclear material. These geometric variables are depicted in Fig. 11. In addition, we use the scaled thickness function $T(b)$ as an alternative to characterize the collision geometry as follows,

$$T(b)/\rho_0 = \int_{-\infty}^{+\infty} \mathrm{d}z \rho(b, z)/\rho_0, \tag{8}$$

in units of fm. This quantity can be explicitly studied together with the gluon saturation physics in eA collisions.

[3] If one scattered nucleon is outside of the geometric nuclear radius due to fluctuation, the fiducial d becomes negative.

B. Measuring forward nuclear fragments

The event sample used for the collision geometry study is generated from the BeAGLE model for ePb collisions at 18 × 110 GeV with τ_0 = 10 fm, shadowing model $genShd = 3$, 1 GeV^2 < Q^2 < 100 GeV^2, and $0.01 < y < 0.95$.

The most abundant final-state products produced during the nuclear breakup are evaporated protons and neutrons. The left and right panel in Fig. 12 show the distribution of the neutrons and protons, respectively, as a function of momentum and scattering angle at two collision energies (18 × 110 GeV and 5 × 50 GeV). The evaporation momenta are close to the beam momentum and their scattering angles are small (∼ few milliradians). At a beam energy of 50 GeV, the largest scattering angle is about 6 mrad, while at 110 GeV, the maximum scattering angle is about 3 times smaller. In contrast to neutrons, there are only a few protons emitted at very small angles because the protons need to overcome the Coulomb barrier to leave the nucleus. As the number of emitted protons during the nuclear evaporation is significantly lower than that of neutrons, it is best to study the properties of the nuclear evaporation process by measuring neutrons.

The neutron multiplicity has been demonstrated to be a tool to access the collision geometry variables [41], inspired by how centrality is determined in heavy-ion collisions [54]. As it is difficult to directly measure a large number of neutrons, we use the energy deposition in the ZDC, similar to Ref. [41]. Higher energy deposited in the ZDC (large multiplicity of evaporation neutrons) is expected to correspond to more-central events. The distribution of energy deposition in the ZDC $E_{\mathrm{n}}^{\mathrm{ZDC}}$ (in generated level) is shown in Fig. 13(a), where the blue area corresponds to the events with a centrality of 0-1%, representing the top 1% events with the highest energy deposition being greater than 2.82 TeV. The red area corresponds to the events with a centrality of 60-100 % with

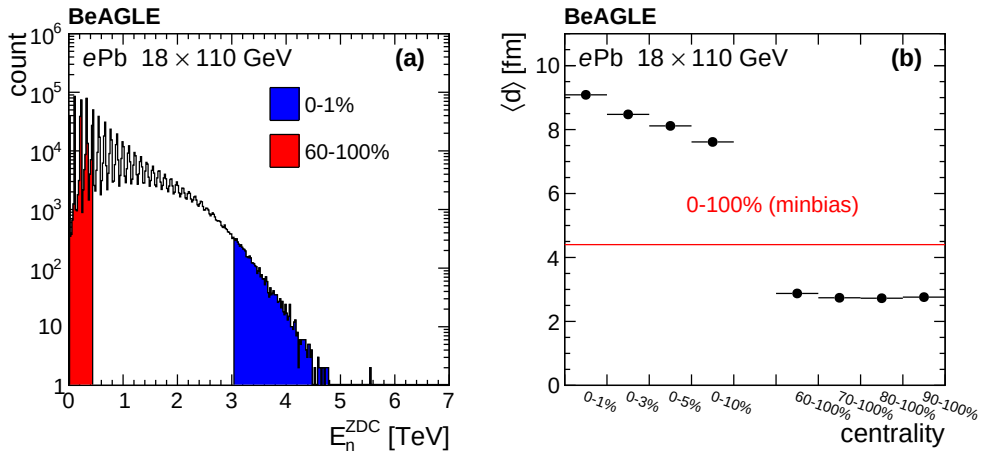

FIG. 13. (a) the centrality selection by the energy deposition in the ZDC. (b) the distribution of $\langle d \rangle$ for different choices of central: 0-1%, 0-3%, 0-5%, 0-10%, as well as for the peripheral centrality: 60-100%, 70-100%, 80-100%, 90-100%. Note that both of these two distributions are in MC generate level, without detector smearing.

the energy being less than 0.44 TeV. In Fig. 13(b), the average traveling distance $\langle d \rangle$ is shown as a function of the ZDC energy percentage class. The value of $\langle d \rangle$ for minimum-bias (0–100%) events is 4.402. $\langle d \rangle$ decreases clearly going from a centrality of 0-1% to 0-10%, but one loses a factor of 10 in statistics, this decreasing trend is not obvious in peripheral collisions. For the following analysis, we choose 0-1% as a central collision, and 60-100% as the most-peripheral collision.

The correlation between the deposited energy in the ZDC and impact parameter b, the nuclear thickness $T(b)/\rho_0$ are shown in Fig. 14(a) and (c), separately. With increasing energy, b decreases while $T(b)/\rho_0$ becomes larger. Figure. 14(b) and (d) show the b and $T(b)/\rho_0$ distributions in central (0-1%) and peripheral (60-100%) collisions, separately. They are normalized by the number of total events. A clear difference between central and peripheral collisions in both b and $T(b)/\rho_0$ can be seen, and by selecting different centrality classes, we obtain an experimental handle on the collision geometry.

C. Systematic study of collision geometry

In this subsection, we perform four different systematic tests, shown in Fig. 15 and Fig. 16, respectively: (a) detector effect; (b) parameter τ_0 dependence; (c) energy dependence; (d) shadowing effect. In these figures, black solid circles and red solid squares represent the results of central and peripheral collisions with the default event sample, respectively. The open markers show the results with the change as labeled in the legend.

First, as the centrality is selected via the energy deposition in the ZDC, we take the ZDC energy resolution and the angular acceptance ($\theta < 5.5$ mrad) into account. We assume a ZDC energy resolution of $\frac{\sigma}{E} = \frac{100\%}{\sqrt{E}} + 10\%$ to smear the energy of each individual neutron with a Gaussian distribution. Figure 15(a) and Fig. 16(a) illustrate the change in the b and $T(b)/\rho_0$ distributions after detector smearing, respectively. The black points represent central collisions, while the red points depict peripheral collisions. The solid markers show the generated distribution without smearing, while the open markers include detector smearing. One can conclude that the results at generator level and after detector smearing are almost identical. The small impact of the ZDC energy resolution on centrality does not put stringent requirements on the ZDC performance.

Secondly, in this analysis, the default option is $\tau_0 = 10$ fm and $genShd = 3$. In order to study the impact of τ_0 on centrality, it was lowered to 3 fm. A smaller τ_0 means more particles can be formed in the nucleus, which results in more emitted neutrons from the nuclear break up, and consequently a larger energy deposition in the ZDC. Figure 15(b) and Fig. 16(b) shows the b and $T(b)/\rho_0$ comparison for $\tau_0 = 10$ and 3 fm in both central and peripheral collisions for the $genShd = 3$ case, respectively. There is no significant difference between the distributions of τ_0 = 10 and 3 fm observed for peripheral events, while some differences for central events. However, the difference between peripheral and central events is small, showing a weak dependence on τ_0.

Thirdly, the energy of the emitted particles scales with the beam energy. However, for the b distribution, there is no significant difference between central and peripheral collisions for the various beam energies, as shown in Fig. 15(c). The same behavior is observed for $T(b)/\rho_0$, and summarized in Fig. 16(c). This indicates that there is no beam energy dependence for the centrality definition. Therefore, although some model parameters are not precisely determined in BeAGLE, we find the correlation between ZDC energy and collision geometry is very stable.

To model nuclear shadowing effects, BeAGLE has 3

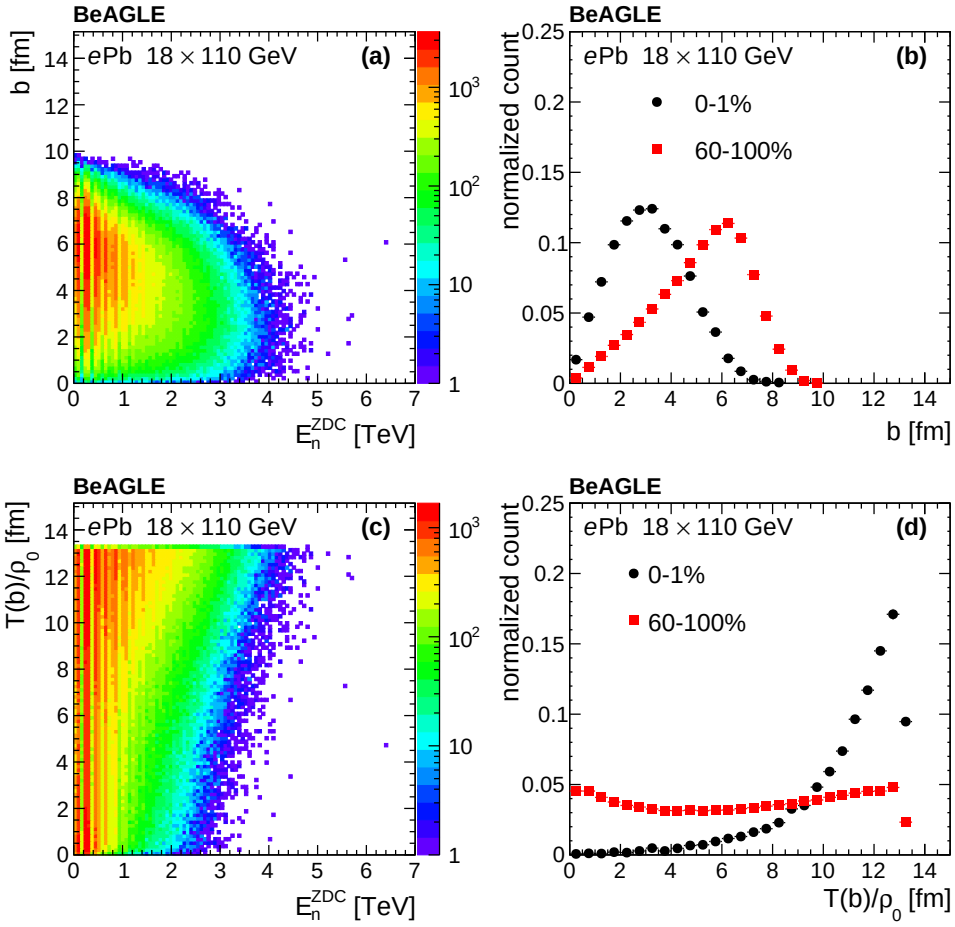

FIG. 14. (a) the correlation between the deposited energy in the ZDC and impact parameter b. (b) b distributions for both central and peripheral collisions. (c) the correlation between the deposited energy in the ZDC and the nuclear thickness $T(b)/\rho_0$. (d) $T(b)/\rho_0$ distributions for both central and peripheral collisions. Note that all of these distributions are in MC generated level, without detector smearing.

different models implemented, as described in Sec. II A. Studies indicate a very small effect of shadowing on the energy deposition in the ZDC in the BeAGLE framework. Predictions for b and $T(b)/\rho_0$ with the different shadowing models are also studied. Fig. 15(d) and Fig. 16(d) show the comparison of b and $T(b)/\rho_0$ between $genShd = 3$ and $genShd = 1$, respectively. In both distributions, no difference is observed between the two shadowing options in central collisions, but some differences are seen in peripheral collisions. The observed differences arise from the low Q^2 region. This can be understood from the fact that $Q^2 \propto \frac{1}{\lambda}$, where λ is the wavelength of the photon. At low Q^2, the photon has a large wavelength, and can interact with many nucleons at once. However, for high Q^2, the wavelength of the photon is small, and therefore less nucleons particpate in the interaction. No difference as a function of $x_{\rm bj}$ for these two shadowing models is found.

V. DISCUSSION

In the previous sections, we compared ep DIS events from the PYTHIA event generator to data from the ZEUS experiment at HERA, as well as μXe and μD collision results from the BeAGLE event generator and E665 data at FermiLab. The results show that we can tune the PYTHIA model to describe target fragmentation in ep collisions, while BeAGLE can not fully describe the target fragmentation region in eA at E665. Model uncertainties, e.g., τ_0, and insufficient knowledge of the experimental selection in E665 might be responsible for the observed discrepancy. In order to further improve our understanding on the way toward the EIC, currently available Ultra-Peripheral Collisions (UPC) data at the Relativistic Heavy-Ion Collider, e.g., the recent data of J/ψ photoproduction in the deuteron-gold UPC [55], and UPC data from the Large Hadron Collider, will be extremely valuable, along with tagged target fragmentation studies at the Continuous Electron Beam Accelerator Fa-

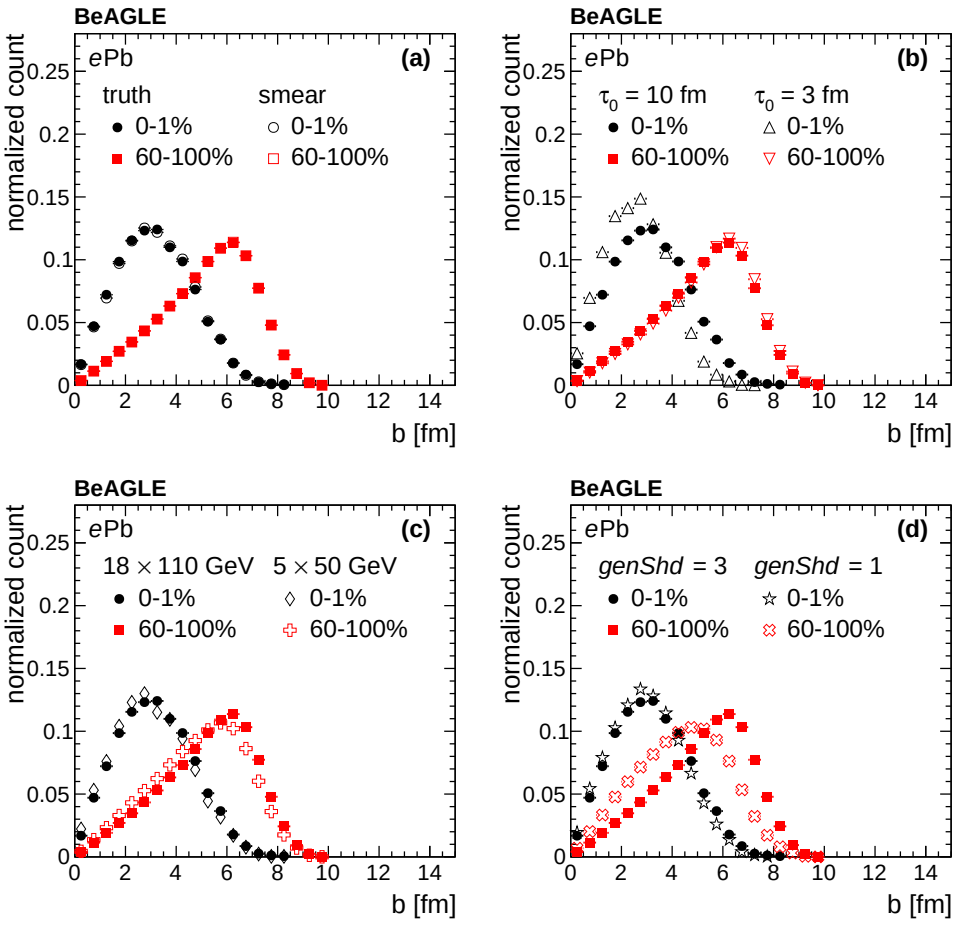

FIG. 15. The comparison of b distributions in central and peripheral collisions for: (a) before and after accounting for detector smearing, the assumed resolution is $\frac{\sigma}{E} = \frac{100\%}{\sqrt{E}} + 10\%$, (b) $\tau_0 = 10$ fm and $\tau_0 = 3$ fm, (c) collision beam energy is 18 × 110 GeV and 5 × 50 GeV, (d) shadowing model 1 and 3. In all the plots, the black solid points and the red solid squares represent the results of central and peripheral collisions with the default event sample, separately. The default event sample is at 18 × 110 GeV for ePb collisions with $\tau_0 = 10$ fm, shadowing model $genShd = 3$, 1 GeV^2 < Q^2 < 100 GeV^2, and $0.01 < y < 0.95$. The open markers show the results with the only change as labeled in the legend.

cility at Jefferson Lab. These data provide a new pathway for study and validatation and improvement of the BeAGLE generator.

In addition, BeAGLE currently cannot simulate coherent diffraction in eA due to the construction of the model. This is closely related to the determination of the formation time parameter, e.g., τ_0. Another future plan for the BeAGLE development will be in this area, where coherent diffraction will provide important insights into the underlying gluon dynamics in the nucleus.

In parallel to this work, there are recent efforts in improving the parton energy loss model PyQM in a different study [30], modification of the DIS kinematics in light nuclei to account for Fermi momentum, implementation of the EMC effect[56–60], and short-range correlations using a generalized-contact formalism [61–63] for lower energy scattering. All past studies, the current work, and future studies have positioned BeAGLE as the prime MC tool for studying lepton-nucleus collisions at high energy, particularly towards the upcoming EIC.

VI. SUMMARY

In this work, we provide a comprehensive description of a high energy lepton-nucleus collision MC event generator - BeAGLE. We validate the model by comparing simulated observables from BeAGLE to available experimental data. The comparison of the PYTHIA-6 model calculations with the ZEUS experimental data in ep collisions shows that we have a good PYTHIA model with refined tunes for target fragmentation in lepton-proton collisions. The BeAGLE event generator describes the E665 lepton-nucleus data for various kinematic variables. However, it only gives a fair description of the charged particle production spectra as measured by the E665 experiment. In order to obtain a full understanding of particle production in the current and target fragmentation

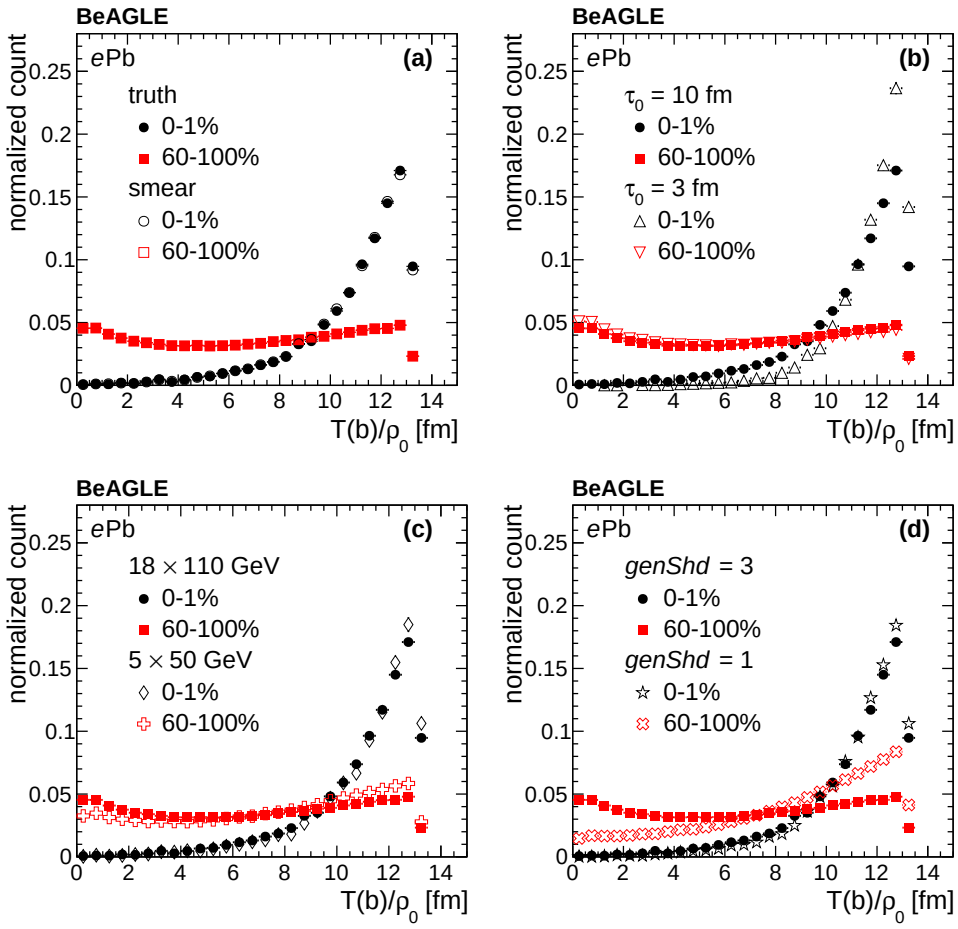

FIG. 16. The comparison of $T(b)/\rho_0$ distributions in central and peripheral collisions for: (a) before and after accounting for detector smearing, the assumed resolution is $\frac{\sigma}{E} = \frac{100\%}{\sqrt{E}} + 10\%$, (b) $\tau_0 = 10$ fm and $\tau_0 = 3$ fm, (c) collision beam energy is 18 × 110 GeV and 5 × 50 GeV, (d) shadowing model 1 and 3. In all the plots, the black solid points and the red solid squares represent the results of central and peripheral collisions with the default event sample, separately. The default event sample is at 18 × 110 GeV for ePb collisions with $\tau_0 = 10$ fm, shadowing model $genShd = 3$, 1 GeV^2 < Q^2 < 100 GeV^2, and $0.01 < y < 0.95$. The open markers show the results with the only change as labeled in the legend.

region, a future facility of high-energy lepton-nucleus collisions, e.g., the EIC, is required.

Based on the BeAGLE event generator, a systematic investigation of collision geometry determination using the detection of neutrons from the nuclear breakup is presented. We find the forward neutron production can provide a good experimental handle on the effective interaction length and nuclear thickness. These parameters will be important for the quantitative study of partonic energy loss in a nuclear medium, and for studies of nonlinear gluon dynamics. Detector requirements for a ZDC are discussed, where the energy resolution has a small impact on the centrality determination and thus does not put stringent requirements on the detector, in contrast to studies of spectator tagging [13, 16]. In addition, we present the dependence of the collision geometry on shadowing effects, the formation time parameter τ_0, and the beam energy. All systematic variations are found to have small impact on the determination of the collision geometry, showing that this robust experimental measurement has minimal model dependence. The study reported in this paper provides an important baseline for developing a general-purpose MC event generator for high energy lepton-nucleus collisions.

ACKNOWLEDGMENTS

We thank Vasiliy Morozov, Pawel Nadel-Turonski, Charles Hyde, and Douglas Higinbotham for fruitful discussions on EIC sciences and experiments. We thank Alberto Accardi, Raphaël Dupré, and Nestor Armesto for their insights on the nuclear geometry and DPMJET used in BeAGLE. The work of W. Chang is supported by the U.S. Department of Energy under Contract No. desc0012704 and the National Natural Science Foundation of China with Grant No. 11875143. The work of E.C. Aschenauer, A. Jentsch, JH. Lee, and Z. Tu is supported

by the U.S. Department of Energy under Contract No. de-sc0012704, and A. Jentsch is also supported by the Program Development program at Brookhaven National Laboratory. The work of M. D. Baker is supported by DOE contracts de-sc0012704, DE-AC05-06OR23177, and by Jefferson Lab LDRD project LDRD1706. The work of L. Zheng is supported by National Natural Science Foundation of China under Grant No. 11905188. The work of Z. Yin is supported by the National Natural Science Foundation of China with Grant No. 11875143.

Appendix A: PYTHIA parameters and BeAGLE control card

Table II summarizes the PYTHIA parameters used in this paper which are not the same as the default value. The meaning of each parameter and the default value can be found in Ref. [17]. Except the parameters that introduced in Sec. III A, others were tuned by HERMES data [46, 47]. Note that the parameters of MSTP(17)=6 and PARP(166)=0.67597 are not PYTHIA-6 standard parameters. They are defined as a different parameterization of $R_{\rm VMD}$ with respect to the default, where $R_{\rm VMD}$ is the ratio of the hadronic cross sections of longitudinally to transversely polarized vector mesons and defined as,

$$R_{\rm VMD} = C(\frac{Q^2}{M_\rho^2})^B, \tag{A1}$$

with C =PARP(165) and B =PARP(166). See Refs. [17, 47] for details.

Table III shows the BeAGLE input control card, including the meaning of each parameter and different value.

[1] M. K. Gaillard, P. D. Grannis, and F. J. Sciulli, Rev. Mod. Phys. **71**, S96 (1999), arXiv:hep-ph/9812285.
[2] R. Oerter, *The theory of almost everything: The standard model, the unsung triumph of modern physics* (2006).
[3] H. D. Politzer, Phys. Rept. **14**, 129 (1974).
[4] D. J. Gross and F. Wilczek, Phys. Rev. Lett. **30**, 1343 (1973).
[5] D. J. Gross and F. Wilczek, Phys. Rev. D **8**, 3633 (1973).
[6] A. Accardi *et al.*, Eur. Phys. J. A **52**, 268 (2016), arXiv:1212.1701 [nucl-ex].
[7] R. Abdul Khalek *et al.*, (2021), arXiv:2103.05419 [physics.ins-det].
[8] J. Adam *et al.*, "Electron ion collider conceptual design report," `https://www.bnl.gov/ec/files/EIC_CDR_Final.pdf` (2021).
[9] F. Gelis, E. Iancu, J. Jalilian-Marian, and R. Venugopalan, Ann. Rev. Nucl. Part. Sci. **60**, 463 (2010), arXiv:1002.0333 [hep-ph].
[10] J. Jalilian-Marian, EPJ Web Conf. **66**, 04012 (2014).
[11] E. Aschenauer, M. Baker, W. Chang, J. Lee, Z. Tu, and L. Zheng, "BeAGLE: A Tool to Refine Detector Requirements for eA Collisions EIC R&D Project eRD17: Progress Report (January-June 2019) and Proposal," `https://wiki.bnl.gov/eic/index.php/BeAGLE` (2019).
[12] W. Chang, E.-C. Aschenauer, M. D. Baker, A. Jentsch, J.-H. Lee, Z. Tu, Z. Yin, and L. Zheng, Phys. Rev. D **104**, 114030 (2021), arXiv:2108.01694 [nucl-ex].
[13] A. Jentsch, Z. Tu, and C. Weiss, Phys. Rev. C **104**, 065205 (2021), arXiv:2108.08314 [hep-ph].
[14] O. Hen, G. Miller, E. Piasetzky, and L. Weinstein, Rev. Mod. Phys. **89**, 045002 (2017), arXiv:1611.09748 [nucl-ex].
[15] C. C. degli Atti, Physics Reports **590**, 1 (2015).
[16] Z. Tu, A. Jentsch, M. Baker, L. Zheng, J.-H. Lee, R. Venugopalan, O. Hen, D. Higinbotham, E.-C. Aschenauer, and T. Ullrich, Phys. Lett. B **811**, 135877 (2020), arXiv:2005.14706 [nucl-ex].
[17] T. Sjöstrand, S. Mrenna, and P. Skands, JHEP **05**, 026 (2006), arXiv:hep-ph/0603175 [hep-ph].
[18] M. Basile *et al.*, Il Nuovo Cimento A (1965-1970) **66**, 129 (1981).
[19] M. R. Adams *et al.* (E665), Z. Phys. C **61**, 179 (1994).
[20] S. Roesler, R. Engel, and J. Ranft, in *Advanced Monte Carlo for radiation physics, particle transport simulation and applications. Proceedings, Conference, MC2000, Lisbon, Portugal, October 23-26, 2000* (2000) pp. 1033–1038, arXiv:hep-ph/0012252.
[21] R. Dupré, *Quark Fragmentation and Hadron Formation in Nuclear Matter*, Ph.D. thesis, Lyon, IPN (2011).
[22] T. Bhlen, F. Cerutti, M. Chin, A. Fass, A. Ferrari, P. Ortega, A. Mairani, P. Sala, G. Smirnov, and V. Vlachoudis, Nucl. Data Sheets **120**, 211 (2014).
[23] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, (2005), 10.2172/877507.
[24] M. R. Whalley, D. Bourilkov, and R. C. Group, 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)* (2005) pp. 575–581, arXiv:hep-ph/0508110.
[25] K. Eskola, H. Paukkunen, and C. Salgado, JHEP **04**, 065 (2009), arXiv:0902.4154 [hep-ph].
[26] C. Salgado and U. A. Wiedemann, Phys. Rev. D **68**, 014008 (2003), arXiv:hep-ph/0302184.
[27] H. Abramowicz and A. Caldwell, Rev. Mod. Phys. **71**, 1275 (1999), arXiv:hep-ex/9903037.
[28] T. Toll and T. Ullrich, Phys. Rev. C **87**, 024913 (2013), arXiv:1211.3048 [hep-ph].
[29] E. Iancu, A. Leonidov, and L. McLerran, in *Cargese Summer School on QCD Perspectives on Hot and Dense Matter* (2002) pp. 73–145, arXiv:hep-ph/0202270.
[30] C. M. Robles Gajardo, A. Accardi, M. D. Baker, W. K. Brooks, R. Dupré, M. Ehrhart, J. A. López, and Z. Tu, (2022), arXiv:2203.16665 [hep-ph].
[31] H. W. Bertini, Phys. Rev. **131**, 1801 (1963).
[32] M. L. Miller, K. Reygers, S. J. Sanders, and P. Steinberg, Ann. Rev. Nucl. Part. Sci. **57**, 205 (2007), arXiv:nucl-ex/0701025.
[33] L. L. Frankfurt and M. I. Strikman, Phys. Rept. **160**, 235 (1988).
[34] L. Frankfurt, V. Guzey, and M. Strikman, Phys. Rev.

Lett. **91**, 202001 (2003).
[35] L. Frankfurt, V. Guzey, and M. Strikman, Mod. Phys. Lett. A **21**, 23 (2006).
[36] A. Bodek and J. L. Ritchie, Phys. Rev. D **23**, 1070 (1981).
[37] C. Ciofi degli Atti and S. Simula, Phys. Rev. C **53**, 1689 (1996), arXiv:nucl-th/9507024.
[38] M. Strikman and C. Weiss, Phys. Rev. C **97**, 035209 (2018), arXiv:1706.02244 [hep-ph].
[39] M. R. Adams *et al.* (E665), Phys. Rev. Lett. **74**, 5198 (1995), [Erratum: Phys.Rev.Lett. 80, 2020–2021 (1998)].
[40] A. Ferrari, P. R. Sala, J. Ranft, and S. Roesler, Z. Phys. C **70**, 413 (1996), arXiv:nucl-th/9509039.
[41] L. Zheng, E. Aschenauer, and J. Lee, Eur. Phys. J. A **50**, 189 (2014), arXiv:1407.8055 [hep-ex].
[42] G. Wolf, Rept. Prog. Phys. **73**, 116202 (2010), arXiv:0907.1217 [hep-ex].
[43] N. Armesto, P. R. Newman, W. Słomiński, and A. M. Staśto, Phys. Rev. D **100**, 074022 (2019), arXiv:1901.09076 [hep-ph].
[44] T. Toll and T. Ullrich, Comput. Phys. Commun. **185**, 1835 (2014), arXiv:1307.8059 [hep-ph].
[45] S. Chekanov *et al.* (ZEUS), JHEP **06**, 074 (2009), arXiv:0812.2416 [hep-ex].
[46] A. Airapetian *et al.* (HERMES), JHEP **08**, 130 (2010), arXiv:1002.3921 [hep-ex].
[47] P. Liebing, *Can the gluon polarization in the nucleon be extracted from HERMES data on single high-p(T) hadrons?*, Ph.D. thesis, Hamburg U. (2004).
[48] T. Sloan, *History of the European Muon Collaboration (EMC)*, CERN Yellow Reports: Monographs (CERN, Geneva, 2019).
[49] V. Andreev *et al.* (H1), Eur. Phys. J. C **81**, 212 (2021), arXiv:2011.01812 [hep-ex].
[50] T. Sjstrand, S. Mrenna, and P. Z. Skands, Comput. Phys. Commun. **178**, 852 (2008), arXiv:0710.3820 [hep-ph].
[51] T. Sjstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, Comput. Phys. Commun. **191**, 159 (2015), arXiv:1410.3012 [hep-ph].
[52] C. Bierlich *et al.*, SciPost Phys. **8**, 026 (2020), arXiv:1912.05451 [hep-ph].
[53] M. R. Adams *et al.* (E665), Zeitschrift für Physik C Particles and Fields **65**, 225 (1995).
[54] B. Alver, M. Baker, C. Loizides, and P. Steinberg, (2008), arXiv:0805.4411 [nucl-ex].
[55] M. Abdallah *et al.* (STAR), Phys. Rev. Lett. **128**, 122303 (2022), arXiv:2109.07625 [nucl-ex].
[56] O. Hen, D. W. Higinbotham, G. A. Miller, E. Piasetzky, and L. B. Weinstein, Int. J. Mod. Phys. **E22**, 1330017 (2013), arXiv:1304.2813 [nucl-th].
[57] L. B. Weinstein, E. Piasetzky, D. W. Higinbotham, J. Gomez, O. Hen, and R. Shneor, Phys. Rev. Lett. **106**, 052301 (2011).
[58] J. Seely *et al.*, Phys. Rev. Lett. **103**, 202301 (2009), arXiv:0904.4448 [nucl-ex].
[59] P. R. Norton, Rept. Prog. Phys. **66**, 1253 (2003).
[60] S. Malace, D. Gaskell, D. W. Higinbotham, and I. Cloet, Int. J. Mod. Phys. E **23**, 1430013 (2014), arXiv:1405.1270 [nucl-ex].
[61] R. Weiss, B. Bazak, and N. Barnea, Phys. Rev. C **92**, 054311 (2015).
[62] R. Cruz-Torres *et al.*, Nature Phys. **17**, 306 (2021), arXiv:1907.03658 [nucl-th].
[63] F. Hauenstein *et al.*, (2021), arXiv:2109.09509 [physics.ins-det].

TABLE II. Summary of PYTHIA parameters used in this paper that are different to the default tune.

parameter	value
MSEL	2
MSTP(14)	30
MSTP(17)	6
MSTP(19)	1
MSTP(20)	4
MSTP(38)	4
MSTP(51)	10042
MSTP(52)	2
MSTP(81)	0
MSTP(82)	1
MSTP(94)	2
MSTP(101)	1
PARP(18)	0.17
PARP(91)	0.40
PARP(97)	6.0
PARP(99)	0.40
PARP(161)	3.00
PARP(162)	24.6
PARP(163)	18.8
PARP(165)	0.47679
PARP(166)	0.67597
PARJ(21)	0.40
PARJ(170)	0.32
MSTJ(12)	1
MSTU(113)	5
CKIN(1)	1.0
CKIN(65)	1.e-09

TABLE III. BeAGLE input control card

Parameter	Descriptions
PROJPAR	Lepton beam can be "ELECTRON" (or "MUON+").
TARPAR	The first number is nucleus A number and the second number is charge Z for the target. The third number is the n/p handling mode for the Pythia subevent: • 0 = sequential n, p. The first events are en, the remaining ep (binomial prob.) Useful for getting en and ep cross-sections from Pythia. • 1 = en only test mode (not really for physics). • 2 = ep only test mode (not really for physics). • 3 = random mix. The events are randomly en or ep. Useful because you can analyze just a subset of the data without bias.
TAUFOR	The first number is the formation time parameter (τ_0) in fm/c for the intra-nuclear cascade, where the second number is the number of generations followed. Default=25, 0 means no cascade.
MOMENTUM	The first number is for the lepton beam (GeV/c), the second for the ion (or p) beam. Both numbers should be entered as positive, but the lepton beam will be multiplied by -1.
L-TAG	These numbers are cuts: yMin, yMax, Q^2Min, Q^2Max, thetaMin, thetaMax, where y and Q^2 (GeV^2) are the leptoproduction kinematics and theta (radians) refers to the lepton scattering angle in the laboratory frame.
PY-INPUT	Specifies the file (with an eight-character maximum name!) used as a pythia input file. See instructions at `https://eic.github.io/software/beagle.html`.
FERMI	First number: • -1 = no Fermi motion at all. • 1 = DPMJET Fermi motion, but Pythia subevent neglects it (DPMJetHybrid mode). • 2 = Fermi motion added to Pythia subevent after the fact. • 3 = Pythia subevent used correct Fermi motion (not yet implemented). Second number: "Scale factor" for Fermi momentum distribution in GeV (0.62 is default, and experts-only parameter). Third number: Fermi momentum distribution 0 (D) = most recommended distribution. Fourth number: Post-processing flag (not yet implemented except deuteron). 0 (D) = no post-processing. 1 = fix energy non-conservation in ion frame (for small nuclei).
FSEED	Leave it as it is and only change the FSEED in the PY-INPUT for production or debugging.
OUTLEVEL	First 4 numbers are verbosity flags: -1=quiet, $\geqslant 1$ increasing verbosity. Fifth number is the number of events to print out and in some cases to be verbose about.
PYVECTORS	Allowed Pythia vector mesons for diffraction: 0(D)=all, 1=ρ, 2=ω, 3=ϕ, 4=J/ψ.
USERSET	First number specifies the meaning of the variables User1,User2,User3. Second number specifies the maximum excitation energy in GeV handed to FLUKA (D=9.0). Note: This should not come into play, but it protects against infinite loops.
PHOINPUT	Any options explained in the PHOJET-manual can be used in between the "PHOINPUT" and "ENDINPUT" cards.
PROCESS	1 1 1 1 1 1 1 1
ENDINPUT	
START	First number specifies the number of events to run. Second number should be 0 (or missing).
STOP	