

# Jet substructure in neutral current deep inelastic $e^+p$ scattering at the upcoming Electron-Ion Collider

Siddharth Jain *Department of Physics, Jamia Millia Islamia University, Delhi 110025, India*R. Aggarwal *USAR, Guru Gobind Singh Indraprastha University, East-Delhi 110092, India*M. Kaur *Department of Physics, Panjab University, Chandigarh 160014, India  
and Department of Physics, Amity University, Punjab, Mohali 140306, India*

(Received 16 February 2023; accepted 23 April 2023; published 2 June 2023)

Predictions are made for the jet substructure of one-jet events produced in positron-proton neutral current deep inelastic scattering at the future Electron-Ion Collider for exchanged four-momentum squared,  $Q^2 > 125 \text{ GeV}^2$ . Data are simulated using Monte Carlo event generators PYTHIA 8.304 and RAPGAP 3.308 at the center of mass energies  $\sqrt{s} = 63.2, 104.9, \text{ and } 141 \text{ GeV}$ . Jets and subjets are produced by longitudinally invariant  $k_T$  cluster algorithm. The subjet multiplicity distributions and differential jet shapes are measured for different jet sizes and varying jet-resolution parameter. A comparison is presented between the  $k_T$  and anti- $k_T$  cluster algorithms for the study of jets and subjets using the simulated data and the HERA data.

DOI: [10.1103/PhysRevD.107.116002](https://doi.org/10.1103/PhysRevD.107.116002)

## I. INTRODUCTION

Partons produced in high energy collisions fragment into hadrons. The process called hadronization corresponds to the transition from partons-to-hadrons. The hadrons are observed in particle detectors as collimated sprays of particles called jets [1]. These jets retain information on the underlying partonic interactions which can be used to study partons undergoing hadronization. The properties of jets are affected by perturbative radiation as well as by low- $p_T$  nonperturbative effects. The description of jet production in high energy collisions is described by quantum chromodynamics (QCD). As the quantum field theory is not yet fully defined, QCD continues to be guided by experimental observations. Whether or not the particle jets of perturbative QCD would exhibit the known properties at high energies at the future colliders with enhanced luminosity needs investigation. The evolution of parton radiation within a jet is dictated by QCD by the splitting

functions which are calculable as power series in the coupling constant  $\alpha_s$ . These can be used to exploit the perturbative QCD description of the jet structure in deep inelastic scattering (DIS) by allowing the measurement of  $\alpha_s$ . Thus the study of jet production in neutral current (NC) DIS [2] events provides a rich testing ground for pQCD and parametrization of the proton parton distribution functions (PDFs) [3].

In addition the analysis of jet substructure [4,5] also provides information on the transition between a parton produced in a hard process and the experimentally observable jet of hadrons [6]. The jetlike substructures within a jet are known as subjets and their multiplicity is measured as the number of clusters resolved in a jet. Characteristics and topology of substructure provide direct access to the QCD splitting functions. For an inclusive sample of jets produced in NC DIS with high jet transverse energy, the fragmentation effects become negligible, making the subjet multiplicity and its mean, calculable in pQCD. Further, the dependence of the pQCD calculations on the PDFs is reduced for this variable. The lowest order contribution to this variable is of the order of first power in  $\alpha_s$ , thus making the stringent test of pQCD beyond leading order possible. In DIS, the jet production is governed by leading order QCD. But to describe internal structure of jets, information at the next-to-leading-order (NLO) is required.

\*ritu.aggarwal1@gmail.com

†manjit@pu.ac.in

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

Internal structure of jets has been studied in  $e^+e^-$  collisions at the Large Electron-Positron Collider (LEP) [7–9],  $p\bar{p}$  collisions at Tevatron [10–12],  $pp$  collisions at LHC [13–15] and  $ep$  scattering at HERA [16]. Subjet multiplicity and jet shape variables were used to study jet substructure in DIS. The subjet multiplicity has been used to calculate the strong coupling constant  $\alpha_s$  [17–19], one of the fundamental parameters of the standard model.

Previous studies revealed that jets in DIS events at HERA were similar to those in  $e^+e^-$  events and narrower than in  $p\bar{p}$  interactions [20,21]. In NC DIS, the jet became narrower [16] and resolved to fewer subjets [17] as the jet transverse energy  $E_{T,\text{jet}}$  increased. In photo-production, the mean subjet multiplicity increased and jet shape broadened as the jet pseudorapidity  $\eta_{T,\text{jet}}$  increased [22].

Electron-Ion Collider (EIC) is scheduled to be built at Brookhaven National Laboratory (BNL) in the US. It aims to accelerate polarized nucleon beams which are crucial to study the properties of quarks and gluons in nuclear medium through deep inelastic scattering process. The lepton probe would provide high precision data to explore the hadron structure. The EIC aims to provide a high luminosity of order  $10^{32}$ – $10^{34}$   $\text{cm}^{-2}\text{s}^{-1}$  for  $ep$  collisions and  $10^{30}$ – $10^{32}$   $\text{cm}^{-2}\text{s}^{-1}$  for  $eAu$  collisions [23]. A minimum center of mass energy (cms) of about 10 GeV is needed for close interaction with quarks and cms energy of order 100 GeV is required to avail higher  $Q^2$  [24]. At EIC, beam energies would range from 5 to 20 GeV for  $e^+$  (or  $e^-$ ) beam and 25 to 275 GeV for beams of nuclei such as gold, lead, or uranium. The corresponding center of mass energies would vary from 20 to 141 GeV [25]. The energy variations would increase the sensitivity to gluon distributions and spin-polarized beams would also facilitate the study of the spin structure of proton [26]. Utilizing the proposed variability in the center-of-mass energy in the EIC and particle identification capabilities of its detectors, one can study hadronization within jets in a wide kinematic regime by characterizing the  $x$  and  $Q^2$  scales of the process.

Recently jet studies at the EIC energies have gained momentum and we present here a brief review of different interesting studies performed. Jets have been studied in photo-production events [27] for the  $e^+p$  and  $e^+A$  simulated data from PYTHIA 6 [28]. Here it was concluded that in spite of low transverse momentum and small particle multiplicities, the jet substructure study would be feasible at the EIC in the photo-production. Photon-proton processes are generally categorized by the virtuality  $Q^2$  of the photon. The value of  $Q^2$  is close to zero for quasi-real photo-production interactions whereas, scattering which involves a  $Q^2$  larger than a few  $\text{GeV}^2$  is called DIS. The study by B. S. Page *et al.* [29] emphasized the importance of jets as probes and gives an example of studying the gluon helicity contributing to the spin of proton through di-jet events in DIS at the future EIC. In another recent

publication by Z.-B. Kang *et al.* [30], the jet charge has been studied to tag the jet in terms of initiating parton flavors. This information can be used to explore the nucleon flavor and spin structure. The study [31] has explored the feasibility of charm-jet cross sections for charged current DIS interactions at the future EIC. The charm-jet cross sections are known to be sensitive to the strange sea content of the proton. Jets would be used as the precision probes at the future EIC which will be collecting the first ever data from lepton-nucleus collisions. The data is expected to be cleaner than the ones collected using hadron beam probes. The properties of DIS jets, such as momentum balance and azimuthal correlation with electrons and substructure in terms of number of constituents are presented in detail in [32]. This paper emphasized the need to revisit the HERA data, the only lepton-hadron collider, in the light of future EIC.

The perturbative QCD calculations done by using jet shape and the jet substructure observables known as jet angularities have been compared with the data from HERA, for the process of photoproduction [27]. Overall, a good agreement is shown and it is concluded that jet substructure studies are feasible at EIC despite the relatively low jet transverse momentum and particle multiplicities. Precision jet substructure measurements at the EIC can be used for the tuning of Monte Carlo event generators and the extraction of nonperturbative corrections.

Recently some other techniques have also been developed to improve the understanding of jet substructure. These include the techniques such as jet grooming which is designed to remove soft wide-angle radiation from the identified jets. It allows for a more direct comparison of perturbative QCD calculations and experimental data. The details can be found in [33,34]. In another work [35] have studied jet-broadening in heavy ion collisions at EIC. They have shown that jet grooming provides a new opportunity to investigate jet broadening effects. This approach is orthogonal to the jet variables such as the azimuthal angle and jet-momentum used at HERA, which is more traditional. Probing the same physics with independent observables offers an important cross-check to ensure consistency and predictive power of theoretical calculations.

In the present work, we study the jets and jet-substructure in terms of the observables, the subjet multiplicity and differential jet shape in NC DIS events for different EIC energies. These variables have been well studied at HERA, the first and only  $ep$  collider, serves as a reference for future EIC jet measurements. Such an analysis has not been investigated so far. Also, we compare two Monte Carlo event generators for this analysis. The results are useful for comparison with the data and measuring nonperturbative effects by further improving the event generator-tuning. Subjet multiplicity is the number of jetlike substructures resolved within jets by reapplying the jet algorithm at a smaller resolution scale  $y_{\text{cut}}$ . Jet shape is defined as the

average fraction of the jet's transverse energy  $E_{T,\text{jet}}$  contained inside an annulus of radius  $r$  [16].

The subjet multiplicity distributions and their average values are studied for different jet sizes and resolution parameter. Validity of  $k_T$  and anti- $k_T$  algorithms for the production of jets and subjets in DIS  $e p$  is also studied. Subjet multiplicity is studied as a function of number of jets  $N_{\text{jet}}$ , resolution parameter  $y_{\text{cut}}$  and jet transverse energy  $E_{T,\text{jet}}$ . Predictions for subjet multiplicity and differential jet shape are made for the upcoming EIC. A detailed introduction to the subject of this paper is presented in Sec. I. For the present study, kinematics of the neutral current deep inelastic scattering is discussed in Sec. II. Details of data generation from different Monte Carlo event generators are given in Sec. III. Methods of jet and subjet formation and jet-shape study are detailed in Sec. IV. Results are presented in Sec. V, followed by conclusion in Sec. VI.

## II. KINEMATICS OF DEEP INELASTIC SCATTERING

In deep inelastic scattering, a high-energy lepton scatters off a hadron with large momentum transfer as shown in Fig. 1. For the present work, neutral current events are simulated which are mediated by the exchange of  $\gamma$  or  $Z$  boson. Alternatively, if the scattering occurs via  $W$  boson then it is called charged current. The kinematics of deep inelastic scattering can be described by the following Lorentz invariant variables [36]. The negative of the invariant mass squared of the exchanged virtual boson is represented by  $Q^2$ . It can be interpreted as the power with which the exchanged boson can resolve the proton structure:

$$Q^2 = -\mathbf{q}\cdot\mathbf{q} = -(\mathbf{k} - \mathbf{k}')^2 \quad (1)$$

$$y = \mathbf{P}\cdot\mathbf{q}/\mathbf{P}\cdot\mathbf{k} \quad (2)$$

where the inelasticity  $y$ , represents the fractional energy loss from lepton to the hadronic system. Bjorken variable ( $x$ ) is associated with the fraction of momentum of the proton carried by the struck parton:

$$x = Q^2/2\mathbf{P}\cdot\mathbf{q} \quad (3)$$

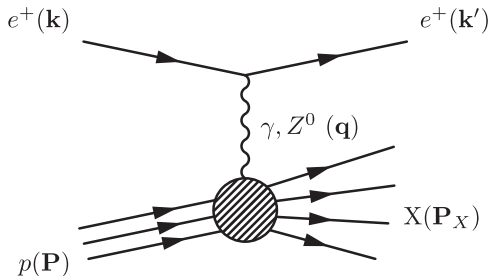


FIG. 1. Neutral current  $e^+ p$  deep inelastic scattering.

The variables  $Q^2$ ,  $y$  and  $x$  follow the relation:

$$Q^2 = sxy \quad (4)$$

where  $s = (\mathbf{P} + \mathbf{k})^2$  is square of the center of mass energy.  $W$  is the energy of the  $\gamma^* p$  system, which is equal to the total mass of the final state hadronic system  $X$ :

$$W^2 = (\mathbf{P} + \mathbf{k})^2 (\equiv \mathbf{P}_X^2) \quad (5)$$

## III. DATA GENERATION

Monte Carlo event generators [37] PYTHIA 8.304 [38] and RAPGAP 3.308 [39] are used to simulate  $10^7$  NC DIS  $e^+ p$  events at different cms energies as shown in Table I. To validate the present work, we also simulated NC events at HERA energy of  $\sqrt{s} = 300$  GeV. Jets are formed by implementing the  $k_T$  sequential recombination algorithm with different jet radii  $R = 0.4, 0.6, 0.8$  and  $1.0$  within the kinematic region  $Q^2 > 125$  GeV<sup>2</sup>, where  $Q^2$  is the virtuality of the exchanged boson and jets are identified with  $E_{T,\text{jet}} > 10$  GeV. Subjets are resolved by re-running the  $k_T$  algorithm on the jets and using a resolution parameter,  $y_{\text{cut}}$ . For both the event generators, QCD cascade is taken to be  $p_t$  ordered parton showers with all QED radiations turned off. The parton distribution function (PDF) set NNPDF 2.3 [40] with  $\alpha_s = 0.130$  is implemented in both the generators through LHAPDF package 6.4.0 [41]. PYTHIA 6.4 [28] is used for fragmentation in RAPGAP. For jet finding and implementation of sequential recombination clustering algorithms, FastJet 3.4.0 [42] package is used on data samples from each event generator. Figure 2 shows the range of Bjorken variable  $x$  versus virtuality  $Q^2$  for center of mass energies at EIC and HERA. EIC aims to provide an enhanced range to investigate quarks and gluons with small momentum fraction ( $x$ ) and analyze their properties over a wide range of momentum transfers  $Q^2$  [24].

## IV. JETS

Quarks and gluons are the fundamental entities produced in high-energy interactions, which instantly fragment and hadronize producing collimated sprays of energetic hadrons. An algorithm is used to cluster these hadrons into

TABLE I. The beam energies for colliding particles and the center of mass energies at EIC and HERA colliders.

Positron energy (GeV)	Proton energy (GeV)	$\sqrt{s}$ (GeV)	Collider
10	100	63.2	EIC
10	275	104.9	EIC
20	250	141	EIC
27.5	820	300	HERA

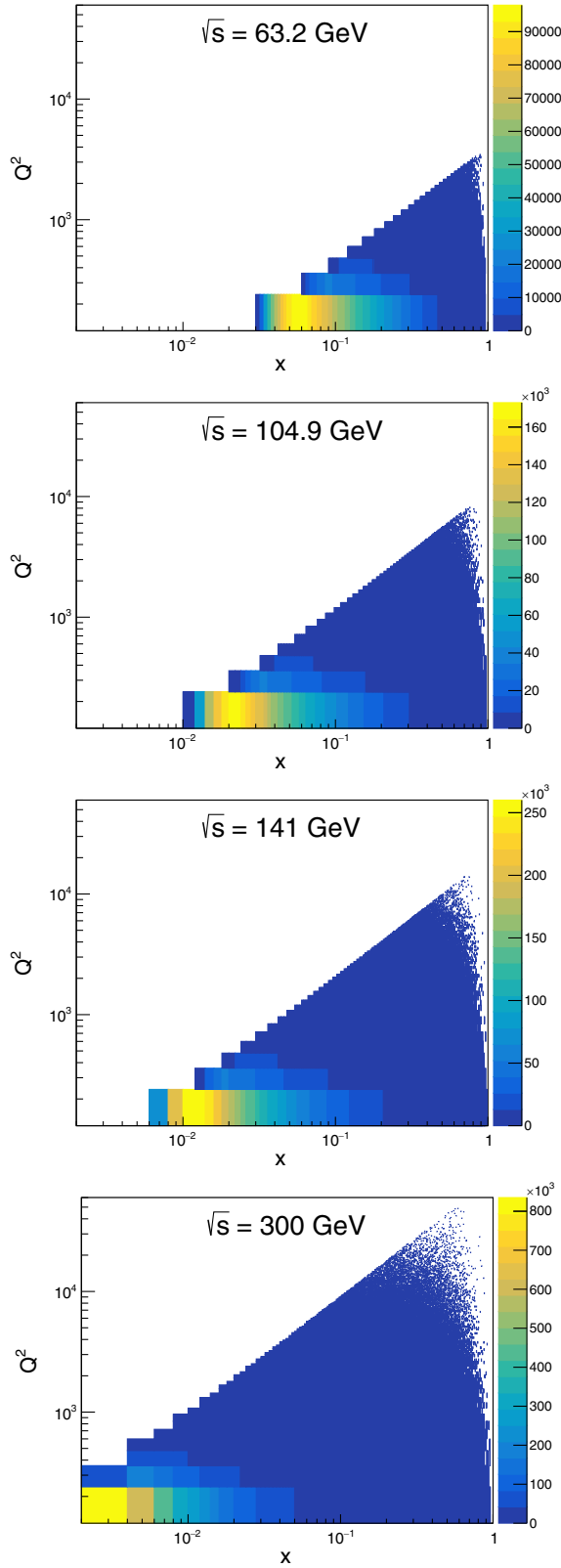


FIG. 2. The range of Bjorken variable  $x$  versus virtuality  $Q^2$  for interactions at  $\sqrt{s} = 63.2, 104.9, 141$  GeV at EIC and  $300$  GeV at HERA.

jets and several clustering algorithms have been developed. A review of the jet algorithms can be found in the paper [43]. Number of jet forming algorithms [43] are available with  $k_T$  [44,45] and anti- $k_T$  [46] being the most utilized to identify the jets and subjets. Sequential clustering algorithms, particularly  $k_T$  and anti- $k_T$  algorithms, are used to cluster the final-state hadrons for the production of jets. Subjets are obtained by reapplying the jet algorithms. The procedure for generalized  $k_T$  jet algorithm is:

- (i) For every pair of  $i$ th and  $j$ th particles, interparticle and particle-beam distance is defined as:

$$d_{ij} = \min(p_{t_i}^{2\rho}, p_{t_j}^{2\rho}) \frac{(y_i - y_j)^2 + (\phi_i - \phi_j)^2}{R^2} \quad (6)$$

$$d_{iB} = p_{t_i}^{2\rho} \quad (7)$$

$p_{t_i}$ ,  $y_i$ , and  $\phi_i$  are the transverse momentum with respect to the beam direction, rapidity, and azimuth respectively of the particle  $i$ .  $R$  is a free parameter usually called the jet radius.

- (ii) The smallest distance among all the  $d_{ij}$  and  $d_{iB}$  are found by iteration. If the minimum distance is  $d_{ij}$ , then the  $i$ th and  $j$ th particles are merged into single cluster and removed from the list. And if it is  $d_{iB}$ , then  $i$ th particle is called a jet and removed from the list.

Generalized  $k_T$  algorithm reduces to  $k_T$  jet algorithm for  $\rho = 1$ , Cambridge/Aachen jet algorithm [47] for  $\rho = 0$ , and anti- $k_T$  jet algorithm for  $\rho = -1$ . All of these algorithms are infrared and collinear (IRC) safe. By equation (6),  $k_T$  algorithm [44,45] is dominated by low  $p_t$  and clusters soft particles first which results in a jet area that differs significantly and makes it more susceptible to underlying events (UE) and pile-up (PU). Whereas, anti- $k_T$  algorithm [46] prefers to cluster hard particles first, and the jet area only differs marginally making the algorithm only slightly susceptible to the UE and PU [48]. In general any algorithm can produce jets with one particle, as discussed in [43], but the implementation of kinematic selections minimises such events. Figures 3 and 4 show the distributions of jet transverse energy  $E_{T,\text{jet}}$  and jet angle  $\theta_{\text{jet}}$  respectively for jets produced by  $k_T$  algorithm at each energy.  $\theta_{\text{jet}}$  is the polar angle of the jet relative to the beam axis and is measured from jet pseudorapidity  $\eta_{\text{jet}}$  as follows:

$$\theta_{\text{jet}} = 2 \arctan(e^{-\eta_{\text{jet}}}) \quad (8)$$

### A. Subjets

Subjets are used to probe the internal structure of jets, and are produced by reapplying procedure for jet algorithm on constituents of jets as input particles. The procedure

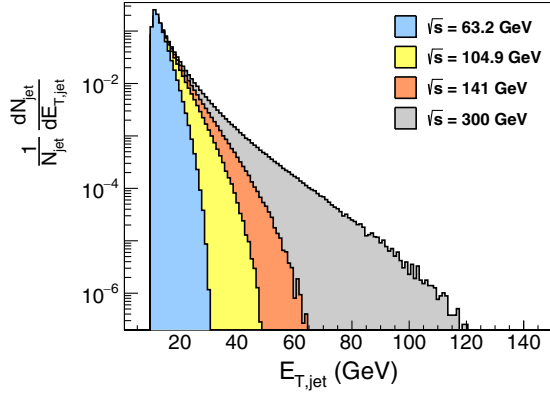


FIG. 3. Jet transverse energy  $E_{T,jet}$  for each center of mass energy with  $R = 1$ .

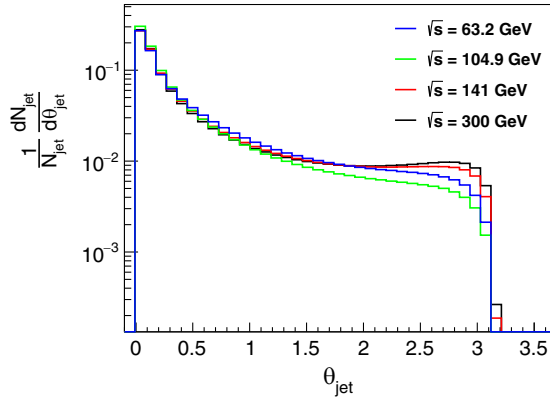


FIG. 4. Jet angle  $\theta_{jet}$  defined in Eq. (8), for jets with  $R = 1$  at different  $\sqrt{s}$ .

from jet algorithm is repeated until all particles satisfy the following condition:

$$d_{cut} = y_{cut}(E_{T,jet}^2). \quad (9)$$

Figure 5 shows an example dijet event with constituents, subjects and scattered positron simulated by PYTHIA 8.304. The subject multiplicity ( $n_{sbj}$ ) is dependent on the selected values of resolution parameter  $y_{cut}$ . The average number of subjects resolved within a jet at a specific  $y_{cut}$  value is defined as mean subject multiplicity  $\langle n_{sbj} \rangle$ : The perturbative QCD value of  $\langle n_{sbj} \rangle$  can be calculated as the ratio of the cross section for subject production to that for inclusive jet production ( $\sigma_{jet}$ ).

$$\langle n_{sbj}(y_{cut}) \rangle = 1 + \frac{1}{\sigma_{jet}} \sum_{j=2}^{\infty} (j-1) \sigma_{sbj,j}(y_{cut}), \quad (10)$$

where,  $\sigma_{sbj,j}(y_{cut})$  is the cross section for producing jets with  $j$  subjects at a resolution scale of  $y_{cut}$ . The mean subject multiplicity  $\langle n_{sbj}(y_{cut}) \rangle$  is measured by [17]:

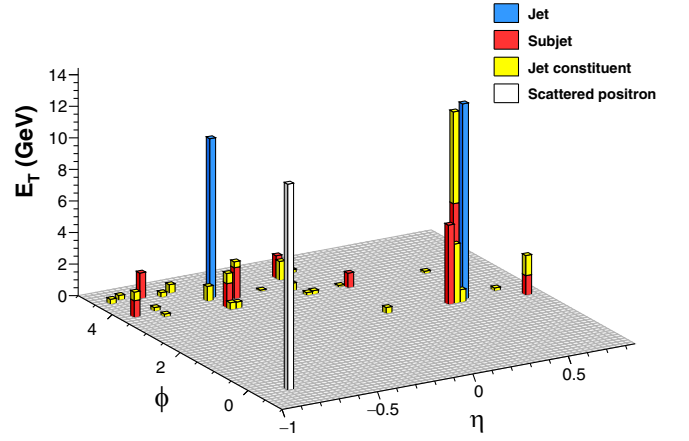


FIG. 5. An example dijet event with constituents, subjects, and scattered positron simulated from PYTHIA 8.304. The number of subjects are three and seven in respective jets.

$$\langle n_{sbj}(y_{cut}) \rangle = \frac{1}{N_{jets}} \sum_{i=1}^{N_{jets}} n_{sbj}^i(y_{cut}), \quad (11)$$

where,  $N_{jets}$  is the total number of jets in the sample and  $n_{sbj}^i(y_{cut})$  are total resolved subjects in the  $i$ th jet.

## B. Differential jet shape

The differential jet shape is defined as the average fraction of the jet's transverse energy  $E_{T,jet}$  contained inside an annulus of inner radius  $r_a(r - \Delta r/2)$  and outer radius  $r_b(r + \Delta r/2)$  concentric to the jet-axis in  $\eta$ - $\phi$  plane.

$$\rho(r) = \frac{1}{N_{jets}} \frac{1}{\Delta r} \sum_{jets} \frac{E_T(r - \Delta r/2, r + \Delta r/2)}{E_T(0, R)} \quad (12)$$

where  $N_{jets}$  is the total number of jets with jet radius  $R = 1$ . Differential jet shape is studied as a function of distance  $r = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$  from the jet-axis, with  $\Delta r = 0.1$  increments. Figure 6 shows the representation of differential jet shape at any radius  $r$ .

## V. RESULTS AND DISCUSSION

Samples of  $10^7$  NC DIS  $e^+p$  collisions are simulated by two Monte Carlo Event generators PYTHIA 8.304 and RAPGAP 3.308 for center of mass energies  $\sqrt{s} = 63.2, 104.9, 141$  GeV at the EIC and  $\sqrt{s} = 300$  GeV at HERA for comparison. The kinematic region is defined by  $Q^2 > 125$  GeV<sup>2</sup>. Final state hadrons in each event are clustered by using longitudinally invariant  $k_T$  jet algorithm for different jet radii. Only the jets with transverse energy  $E_{T,jet} > 10$  GeV are studied. Jet finding and algorithm implementation is done using FastJet package [42]. For each cms energy, one-jet events are found to be dominant in NC DIS  $e^+p$ . The  $E_{T,jet}$  distributions in Fig. 3 show that at

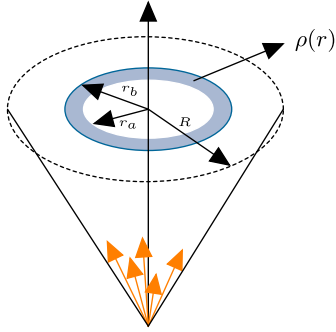


FIG. 6. The differential jet shape is the average fraction of the jet’s transverse energy  $E_{T,\text{jet}}$  contained inside an annulus of radius  $r$ , centered around the jet-axis.

$\sqrt{s} = 300$  GeV, jets with the highest  $E_{T,\text{jet}}$  can be found up to 130 GeV. While, for  $\sqrt{s} = 63.2, 104.9$  and 141 GeV, the highest  $E_{T,\text{jet}}$  is observed to decrease with values around 30, 50 and 60 GeV respectively. It is also observed that most of the jets are produced in the forward direction as shown in Fig. 4, making very small angles to the initial hadron beam direction. They can however, be produced significantly in the barrel region.

### A. Subjet multiplicity

Subjets are formed by reapplying longitudinally invariant  $k_T$  jet algorithm on hadronic jets at smaller resolution parameter  $y_{\text{cut}}$ . Events with jet transverse energy  $E_{T,\text{jet}} > 15$  GeV and jet pseudorapidity  $-1 < \eta_{\text{jet}} < 2$  are used for forming the subjets. Figure 7 shows the distribution of constituents of subjets formed using  $y_{\text{cut}} = 0.0005$  at  $\sqrt{s} = 63.2, 104.9$  and 141 at EIC and 300 GeV at

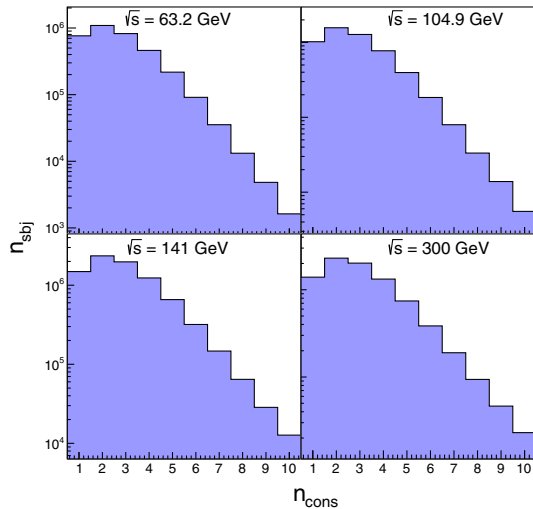


FIG. 7. Number of constituents ( $n_{\text{cons}}$ ) present in subjets for  $10^7$  events simulated by PYTHIA 8. The mean is 2.65, 2.83, 2.92, 3.06 and the variance is 2.00, 2.31, 2.48, 2.79 for  $\sqrt{s} = 63.2, 104.9, 141,$  and 300 GeV respectively.

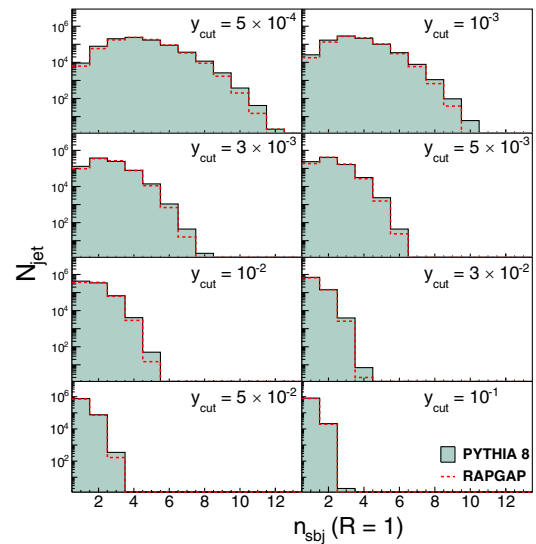
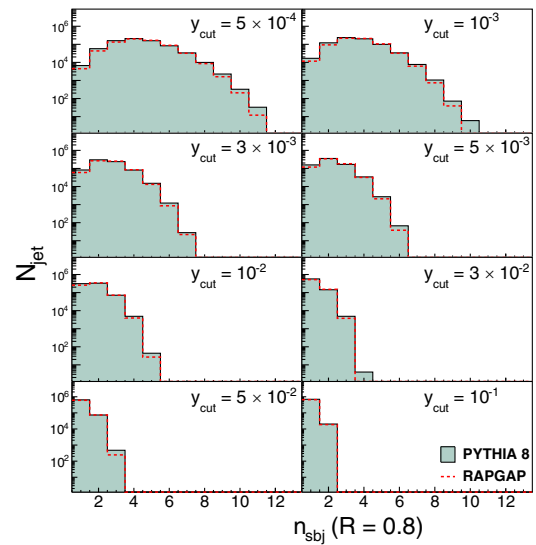
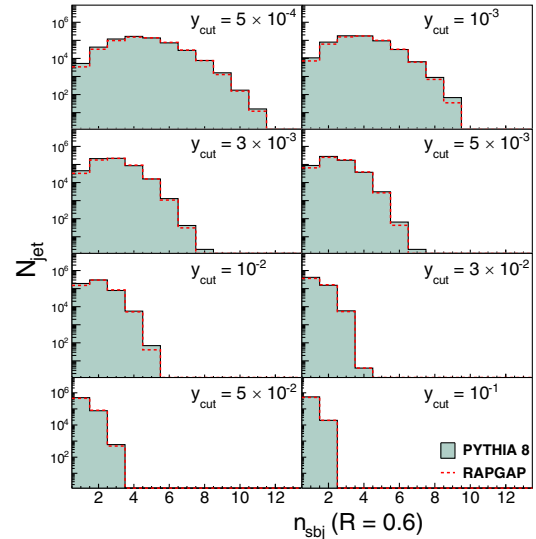


FIG. 8. Subjet multiplicity for EIC at  $\sqrt{s} = 63.2$  GeV for  $R = 0.6, 0.8,$  and 1 from top to bottom.

HERA. It may be observed that the mean value of subjet constituents increases slowly with collision energy and the results at 300 GeV are consistent with the observations from HERA. The subjets with one constituent are also observed which is probable from any jet algorithm [43]. The mean value is greater than 2.5 for every cms energy. The subjet multiplicity measurements at  $\sqrt{s} = 63.2, 104.9$  and 141 GeV, at various resolution parameter  $y_{\text{cut}}$  are shown in Figs. 8–10 for radius 0.6, 0.8, and 1, respectively. It is observed that the subjet multiplicity decreases as  $y_{\text{cut}}$  increases from 0.0005 to 0.1 for all jet radii. Table II shows mean subjet multiplicity for each value of  $y_{\text{cut}}$  and jet radius R for both event generators. It can also be seen that for a given jet radius R,  $y_{\text{cut}}$  and minimum  $E_{T,\text{jet}}$  values of subjet multiplicities are similar for each center of mass energy, inferring that production of subjets in a jet of given transverse energy is independent of center of mass energy. RAPGAP gives slightly higher subjet multiplicity than PYTHIA 8. For each generator, and for each  $y_{\text{cut}}$  value, the subjet multiplicity decreases very marginally for increasing R. However for the largest  $y_{\text{cut}}$  of 0.1, it becomes roughly constant.

For each center of mass energy, Fig. 11 shows mean subjet multiplicity  $\langle n_{\text{sbj}} \rangle$  as a function of  $y_{\text{cut}}$  at each radius. It is seen that mean subjet multiplicity decreases rapidly with increase in value of  $y_{\text{cut}}$ . It can be coherently observed that for a given  $y_{\text{cut}}$  value, mean subjet multiplicity  $\langle n_{\text{sbj}} \rangle$  reduces slightly as the jet radius R increases. The figure also shows a comparison of the data at  $\sqrt{s} = 300$  GeV from the ZEUS experiment [17] with the data simulated from PYTHIA 8 and RAPGAP. A good agreement is observed. Figure 12 presents mean subjet multiplicity as a function of jet transverse energy  $E_{T,\text{jet}}$  at  $y_{\text{cut}} = 10^{-2}$ . Average subjet multiplicity is observed to decrease as  $E_{T,\text{jet}}$  increases for each center of mass energy. Similar observation was also made with data at  $\sqrt{s} = 300$  GeV which agrees very well with the simulated data with jet of radius = 1 in the paper [16,49]. In this paper, jet shape was used to study jet substructure with conclusion that jets become narrower as  $E_{T,\text{jet}}$  increases, resulting in lower values of subjet multiplicity.

### B. Comparison of subjets from $k_T$ and anti- $k_T$ algorithms

Comparison of  $k_T$  and anti- $k_T$  algorithms is studied on the production of jets and subjets for DIS  $e^+p$  events with  $E_{T,\text{min}} > 10$  GeV and  $Q^2 > 125$  GeV<sup>2</sup>. In both the cases when jets are produced by  $k_T$  or anti- $k_T$  algorithm for R = 1, subjets are produced through the  $k_T$  algorithm for  $E_{T,\text{jet}} > 15$  GeV and  $-1 < \eta_{\text{jet}} < 2$  and the subjet multiplicity is measured. The results are shown in Fig. 13 for  $\sqrt{s} = 141$  GeV and it is found that in both the cases for the chosen jet radius and  $y_{\text{cut}}$ , values are nearly the same for

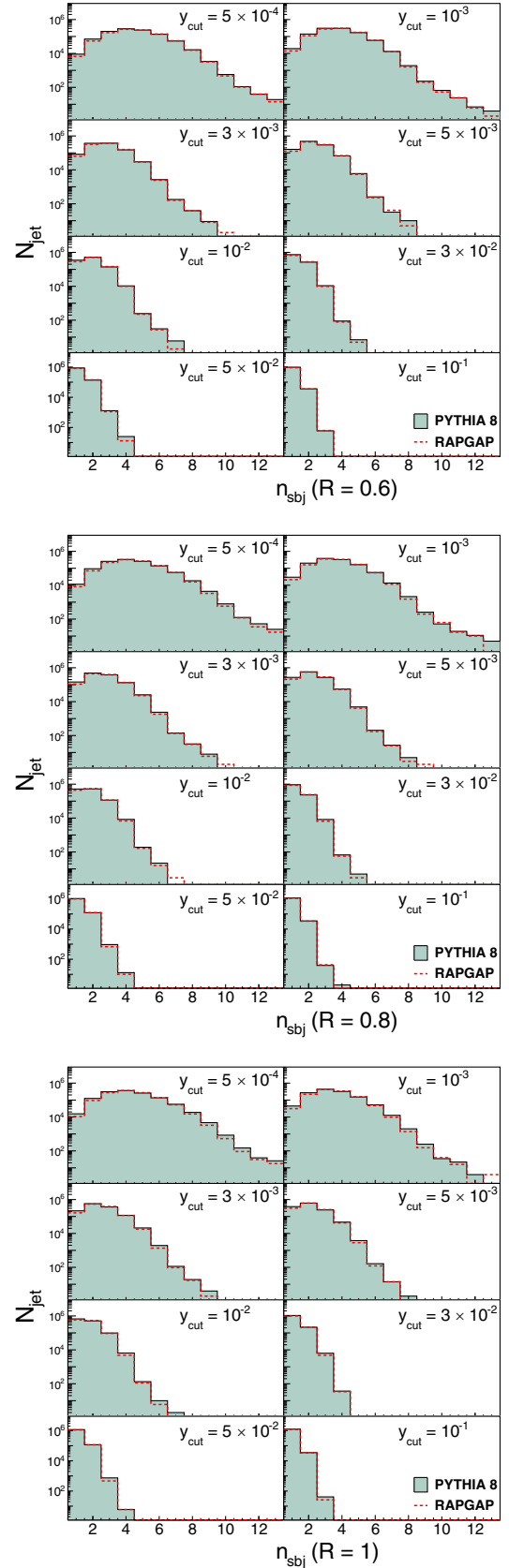


FIG. 9. Subjet multiplicity for EIC at  $\sqrt{s} = 104.9$  GeV for R = 0.6, 0.8, and 1 from top to bottom.

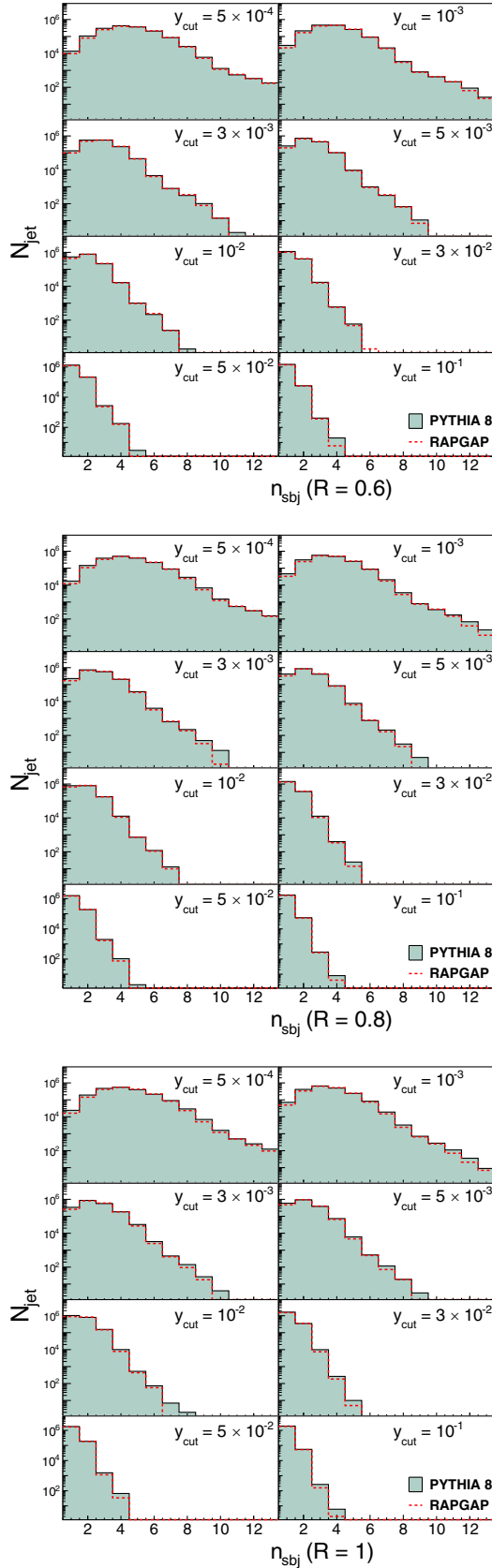


FIG. 10. Subject multiplicity for EIC at  $\sqrt{s} = 141$  GeV for  $R = 0.6, 0.8,$  and  $1$  from top to bottom.

TABLE II. Number of subjects within a jet with a minimum  $E_{T,\text{jet}}$  cut for multiple values of  $y_{\text{cut}}$  and jet radius  $R$ .  $10^7$  events are generated at each  $\sqrt{s}$ . Statistical errors are found to be negligible and excluded from the table.

$y_{\text{cut}}$	$\langle n_{\text{subj}} \rangle$ for $\sqrt{s} = 63.2$ GeV					
	$R = 0.6$		$R = 0.8$		$R = 1$	
	PYTHIA	RAPGAP	PYTHIA	RAPGAP	PYTHIA	RAPGAP
0.0005	4.32	4.46	4.26	4.37	4.16	4.25
0.001	3.67	3.78	3.53	3.64	3.40	3.49
0.003	2.69	2.79	2.51	2.60	2.38	2.45
0.005	2.29	2.38	2.12	2.20	2.00	2.06
0.01	1.82	1.90	1.68	1.74	1.58	1.63
0.03	1.28	1.32	1.21	1.23	1.18	1.19
0.05	1.14	1.16	1.10	1.11	1.09	1.09
0.1	1.03	1.04	1.03	1.03	1.03	1.02

$y_{\text{cut}}$	$\langle n_{\text{subj}} \rangle$ for $\sqrt{s} = 104.9$ GeV					
	$R = 0.6$		$R = 0.8$		$R = 1$	
	PYTHIA	RAPGAP	PYTHIA	RAPGAP	PYTHIA	RAPGAP
0.0005	4.39	4.50	4.28	4.38	4.15	4.24
0.001	3.70	3.79	3.53	3.62	3.38	3.46
0.003	2.69	2.77	2.50	2.58	2.35	2.42
0.005	2.28	2.36	2.11	2.17	1.98	2.03
0.01	1.82	1.88	1.67	1.72	1.57	1.61
0.03	1.28	1.31	1.21	1.23	1.18	1.19
0.05	1.14	1.15	1.11	1.11	1.09	1.09
0.1	1.04	1.04	1.03	1.03	1.03	1.03

$y_{\text{cut}}$	$\langle n_{\text{subj}} \rangle$ for $\sqrt{s} = 141$ GeV					
	$R = 0.6$		$R = 0.8$		$R = 1$	
	PYTHIA	RAPGAP	PYTHIA	RAPGAP	PYTHIA	RAPGAP
0.0005	4.41	4.52	4.30	4.40	4.17	4.26
0.001	3.70	3.80	3.54	3.63	3.39	3.47
0.003	2.68	2.76	2.50	2.57	2.36	2.42
0.005	2.28	2.35	2.11	2.17	1.98	2.04
0.01	1.81	1.87	1.67	1.72	1.57	1.61
0.03	1.29	1.31	1.22	1.23	1.19	1.20
0.05	1.14	1.15	1.11	1.11	1.10	1.10
0.1	1.04	1.04	1.03	1.03	1.03	1.03

both PYTHIA and RAPGAP. A similar study is also presented in Fig. 14 for  $\sqrt{s} = 300$  GeV and compared to the data from the ZEUS experiment [17]. It is observed that the data also agrees well with the predicted values from PYTHIA and RAPGAP.

In a second study, jets are produced from anti- $k_T$  algorithm and subjects are again produced using anti- $k_T$  algorithm with smaller subject radius of  $R_{\text{sub}} = 0.4$  [50] at both the cms energies. It is observed that the  $\langle n_{\text{subj}} \rangle$  dependence on  $y_{\text{cut}}$  does not follow the trend as in the case when subjects are formed from  $k_T$  algorithm, as evident from Figs. 13 and 14. It is observed that the  $\langle n_{\text{subj}} \rangle$  decreases



TABLE III. Number of subjets within a jet with a minimum  $E_{T,\text{jet}}$  cut for different values of  $y_{\text{cut}}$  and jet radius  $R = 0.6, 0.8, 1.0$  for HERA data at  $\sqrt{s} = 300$  GeV.  $10^7$  events are generated for each set. Statistical errors are found to be  $< 0.05\%$  for each value. Results from the ZEUS experiment are shown in the last column.

$y_{\text{cut}}$	$\langle n_{\text{subj}} \rangle$ for $\sqrt{s} = 300$ GeV						Experiment [17]
	R = 0.6		R = 0.8		R = 1		
	PYTHIA	RAPGAP	PYTHIA	RAPGAP	PYTHIA	RAPGAP	
0.0005	4.41	4.52	4.30	4.40	4.17	4.26	$4.30 \pm 0.0068$
0.001	3.70	3.80	3.54	3.63	3.39	3.47	$3.49 \pm 0.0057$
0.003	2.68	2.76	2.50	2.57	2.36	2.42	$2.44 \pm 0.0043$
0.005	2.28	2.35	2.11	2.17	1.98	2.04	$2.06 \pm 0.0038$
0.01	1.81	1.87	1.67	1.72	1.57	1.61	$1.63 \pm 0.0033$
0.03	1.29	1.31	1.22	1.23	1.19	1.20	$1.20 \pm 0.0024$
0.05	1.14	1.15	1.11	1.11	1.10	1.10	$1.10 \pm 0.0018$
0.1	1.04	1.04	1.03	1.03	1.03	1.03	$1.03 \pm 0.0010$

less steeply with  $y_{\text{cut}}$  in this case. A further investigation into this observation reveals that equation (6) is dominated by high  $p_T$  for anti- $k_T$  algorithm ( $\rho = -1$ ) and tends to cluster hard radiation first. This assigns majority of energy to one subjet and the weaker subjet is most likely to get discarded. This leads to the production of imbalanced subjets. But  $k_T$  algorithm prefers to cluster soft particles first, which results in even distribution of energy between subjets that contain final state radiation [51]. The deviation observed is also reported in [48] which states that anti- $k_T$  algorithm prefers to cluster hard particles first, which makes it best at resolving jets but is ineffective for producing jet substructure due to its inability to decluster.

### C. Differential jet shape measurements

For differential jet shape measurements, jets are produced through longitudinally invariant  $k_T$  algorithm for jet

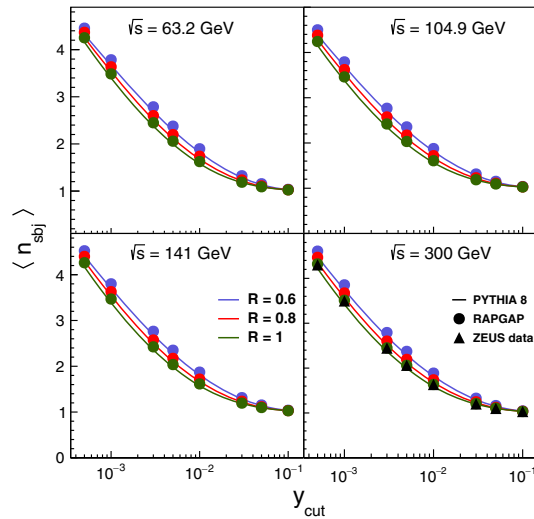


FIG. 11. Mean subjet multiplicity as a function of  $y_{\text{cut}}$ . The errors on the simulated values are less than 0.05% and the errors on the ZEUS data are very small as shown in Table III.

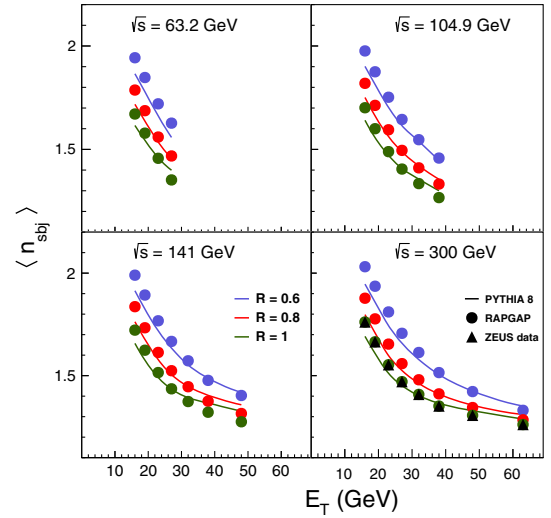


FIG. 12. Mean subjet multiplicity as a function of  $E_{T,\text{jet}}$ . Solid lines represent the predictions from PYTHIA and points represent predictions from RAPGAP for different values of  $R$ .

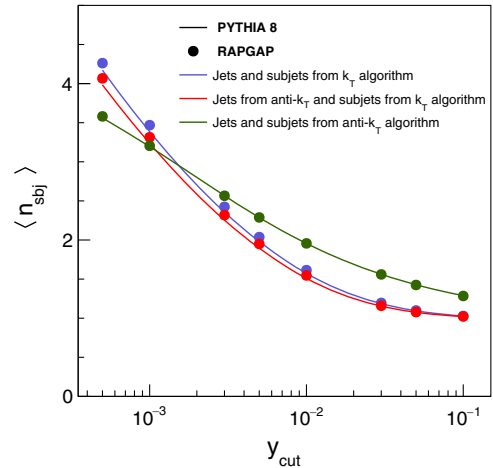


FIG. 13. The mean subjet multiplicity as a function of  $y_{\text{cut}}$  for  $\sqrt{s} = 141$  GeV for  $R = 1$  and  $R_{\text{sub}} = 0.4$ .

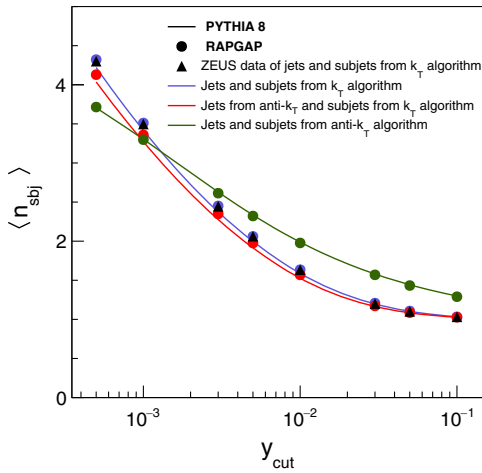


FIG. 14. The mean subjet multiplicity as a function of  $y_{\text{cut}}$  for  $\sqrt{s} = 300$  GeV for  $R = 1$  and  $R_{\text{sub}} = 0.4$ . The errors shown in Table III for ZEUS data are negligibly small.

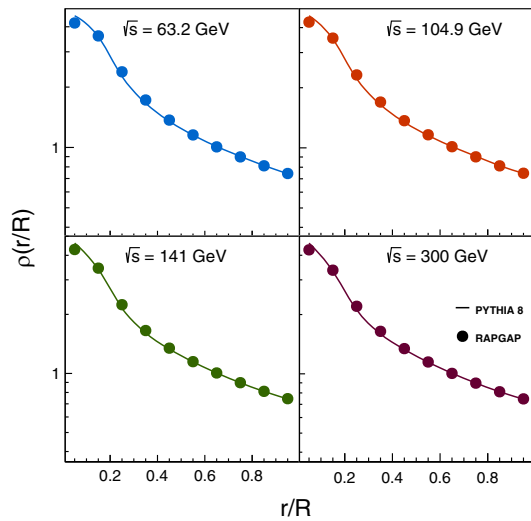


FIG. 15. Differential jet shape as the function of annulus radii  $r$ .

radius  $R = 1$ . Figure 15 shows measurement of differential jet shape as defined in Eq. (12) at various annulus radii  $r$  for each energy. An apparent peak is observed in differential jet shape nearby the jet-axis, indicating the presence of high  $E_T$  region around center of the jet. The differential jet shape variable  $\rho$  decreases as  $r$  increases. For a specific annulus radius, jet shape is observed to be similar at each center of mass energy for both PYTHIA and RAPGAP. This leads to the observation that both the subjet multiplicity and jet shape are independent of center of mass energy.

#### D. Experimental outlook

The jet and jet-substructure studies at EIC in  $ep$  collisions would serve as one of the most important calibration tools for performing  $eA$  studies. The jets in the  $eA$  collisions traverse through the nuclear medium. Differences between the jet-substructure in  $ep$  and  $eA$  collisions provide information of how partonic energy is dissipated in the cold nuclear medium. High  $P_T$  jets are expected to be present mostly in the region  $-1 < \eta < 2$ , where the scattered positron is also present as seen in Fig. 16 which shows the positron momentum as a function of its pseudorapidity. It can be seen that for  $\sqrt{s} = 63.2, 104.9$ , and  $141$  GeV the highest positron momentum reaches around 90, 200, and  $180$  GeV/c and jet transverse momentum is found up to 30, 50, 65 GeV/c respectively. The positron position in the detector is  $Q^2$  dependent and for the  $Q^2 > 125$  GeV $^2$ , it is found in the barrel and forward endcap region. Higher center-of-mass energies would be vital to enable the jet studies at EIC. The precision of jet energy and positron energy measurements should be better than the earlier experiments at HERA. As the luminosity increases manifold at EIC, there is a need to improve the detector related systematic errors. The increased luminosity and hence the increased event rate are an asset at the EIC which dictates a high data rate of  $\approx 500$  kHz [52]. For the informative jet substructure studies, the lepton-hadron identification is

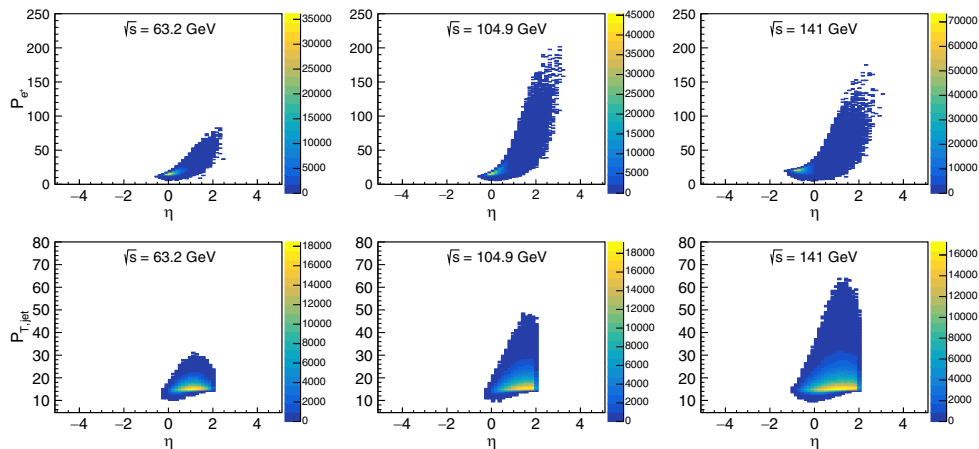


FIG. 16. Positron momentum ( $P_{e^+}$ ) and jet transverse momentum ( $P_{T,\text{jet}}$ ) as function of their respective pseudorapidity  $\eta$ , in the events with  $E_{T,\text{jet}} > 15$  GeV and  $-1 < \eta_{\text{jet}} < 2$ .

crucial in the barrel and forward endcap regions. The positron measurements are designed using cylindrical symmetric electromagnetic detectors in the barrel region and block shaped electromagnetic detectors in the forward region.

The hadron physics can be handled using high granularity electromagnetic and hadronic calorimeters. A detailed discussion on the optional technologies and the region covered by them can be found in [52].

## VI. CONCLUSION

Predictions for the properties of subjects within jets produced in neutral current deep inelastic  $e^+p$  scattering at the future EIC are presented using simulated data. The mean subjet multiplicity at each of the center-of-mass energies of 63.2, 104.9 and 141 GeV has been found using jets in the pseudorapidity region,  $-1 < \eta_{\text{jet}} < 2$  and with each jet having  $E_{T,\text{jet}} > 15$  GeV. Various combinations of jet clustering algorithms are used and compared. It is observed that the average number of subjects within a jet decreases as  $E_{T,\text{jet}}$  increases. The observation agrees with the results from HERA.

It has been shown that for jets, both  $k_T$  and anti- $k_T$  algorithms give similar results. Multiple values of the jet radius are used for selecting jets and subjects. It is observed that the subjet multiplicity decreases as the jet radius increases. For each radius, subjet multiplicity is measured for various values of  $y_{\text{cut}}$ , the resolution parameter. Subjet multiplicity is observed to decrease with the increase

in  $y_{\text{cut}}$ . The jet shape is studied in terms of differential jet shape variable  $\rho(r)$ , which is observed to decrease as the annulus radius is increased. It has been shown that both subjet multiplicity and differential jet shape in a jet of given transverse energy  $E_T$ , are independent of the center of mass energy. When jets and subjects both are produced by  $k_T$  algorithm, mean subjet multiplicity tends to decrease as  $E_{T,\text{jet}}$  is increased.

Comparison of  $k_T$  algorithm and anti- $k_T$  algorithm has been studied for the production of subjects. It is observed that for the subjects,  $k_T$  algorithm is much better suited than anti- $k_T$  algorithm as can be well observed from the Fig. 14 for the data at  $\sqrt{s} = 300$  GeV. It is due to the fact that anti- $k_T$  algorithm prefers to cluster hard particles first, which makes it best at resolving jets but due to its inability to decluster, it is ineffective for producing substructure. It is observed that characteristics of jets and subjects produced in PYTHIA and RAPGAP are in close agreement.

The upcoming EIC detector can increase the accessible range in  $x$  for a given  $Q^2$  due to its reduced cms energy as compared to the HERA detector. This region is interesting and not much explored. The improved precision of the proton parton distributions is foreseen at the EIC detector due to the promising increase in the luminosity. The jet and jets' substructure studies at the EIC would have very small statistical uncertainties and a deep knowledge of the detector would be required to pin down the systematic uncertainties.

- 
- [1] K. Rabbertz, *Jet Physics at the LHC* (Springer, New York, 2018).
  - [2] R. G. Roberts, *The Structure of the Proton* (Cambridge University Press, Cambridge, England, 1993).
  - [3] M. Klasen, K. Kovařík, and J. Potthoff, *Phys. Rev. D* **95**, 094013 (2017).
  - [4] S. Marzani, G. Soyez, and M. Spannowsky, *Looking Inside Jets* (Springer, New York, 2019).
  - [5] R. Kogler, *Advances in Jet Substructure at the LHC*, (Springer, New York, 2021); Report No. PUBDB-2021-04184.
  - [6] S. D. Ellis, Z. Kunszt, and D. E. Soper, *Phys. Rev. Lett.* **69**, 3615 (1992).
  - [7] D. Buskulic, D. Casper, I. De Bonis, D. Decamp, P. Ghez, C. Goy, J.-P. Lees, M.-N. Minard, P. Odier, B. Pietrzyk *et al.*, *Phys. Lett. B* **346**, 389 (1995).
  - [8] R. Barate (ALEPH Collaboration), *Eur. Phys. J. C* **17**, 1 (2000).
  - [9] P. Abreu (DELPHI Collaboration), *Eur. Phys. J. C* **4**, 1 (1998).
  - [10] F. Abe, D. Amidei, C. Anway-Weiss, G. Apollinari, M. Atac, P. Auchincloss, A. Baden, N. Bacchetta, W. Badgett, M. W. Bailey *et al.*, *Phys. Rev. Lett.* **70**, 713 (1993).
  - [11] S. Abachi, B. Abbott, M. Abolins, B. Acharya, I. Adam, D. Adams, M. Adams, S. Ahn, H. Aihara, J. Alitti *et al.*, *Phys. Lett. B* **357**, 500 (1995).
  - [12] V. Abazov, B. Abbott, A. Abdesselam, M. Abolins, V. Abramov, B. Acharya, D. Adams, M. Adams, S. Ahmed, G. Alexeev *et al.*, *Phys. Rev. D* **65**, 052008 (2002).
  - [13] M. Kaur and A. P. Kaur, *Adv. High Energy Phys.* **2013**, 1 (2013).
  - [14] G. Aad, B. Abbott, J. Abdallah, A. Abdelalim, A. Abdesselam, B. Abi, M. Abolins, H. Abramowicz, H. Abreu, E. Acerbi *et al.*, *Phys. Rev. D* **83**, 052003 (2011).
  - [15] A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, M. Campanelli, B. Chapleau, A. Cholakian, J. Chou, M. Dasgupta *et al.*, *J. Phys. G* **39**, 063001 (2012).
  - [16] ZEUS Collaboration and J. Breitweg, *Eur. Phys. J. C* **8**, 367 (1999).
  - [17] S. Chekanov, D. Krakauer, J. Loizides, S. Magill, B. Musgrave, J. Repond, R. Yoshida, M. C. Mattingly, P. Antonioli, G. Bari *et al.*, *Phys. Lett. B* **558**, 41 (2003).
  - [18] M. Derrick, D. Krakauer, S. Magill, D. Mikunas, B. Musgrave, J. Repond, R. Stanek, R. Talaga, H. Zhang, R. Ayad *et al.*, *Phys. Lett. B* **363**, 201 (1995).

- [19] J. Breitweg, S. Chekanov, M. Derrick, D. Krakauer, S. Magill, B. Musgrave, A. Pellegrino, J. Repond, R. Stanek, R. Yoshida *et al.*, *Phys. Lett. B* **507**, 70 (2001).
- [20] K. Ackerstaff (OPAL Collaboration), *Eur. Phys. J. C* **1**, 479 (1998).
- [21] R. Akers, G. Alexander, J. Allison, K. Anderson, S. Arcelli, S. Asai, A. Astbury, D. Axen, G. Azuelos, A. Ball *et al.*, *Z. Phys. C* **63**, 197 (1994).
- [22] C. Glasman (ZEUS Collaboration), in *AIP Conference Proceedings* (American Institute of Physics, Ambleside, England, 2001) Vol. 571, pp. 203–208.
- [23] V. Ptitsyn and eRHIC Collaboration, in *AIP Conference Proceedings* (American Institute of Physics, Santa Fe, New Mexico (USA), 2006) Vol. 842, pp. 1046–1048.
- [24] A. Deshpande, R. Milner, R. Venugopalan, and W. Vogelsang, *Annu. Rev. Nucl. Part. Sci.* **55**, 165 (2005).
- [25] R. A. Khalek, U. d’Alesio, M. Arratia, A. Bacchetta, M. Battaglieri, M. Begel, M. Boglione, R. Boughezal, R. Boussarie, G. Bozzi *et al.*, [arXiv:2203.13199](https://arxiv.org/abs/2203.13199).
- [26] A. Accardi, J. Albacete, M. Anselmino, N. Armesto, E. Aschenauer, A. Bacchetta, D. Boer, W. Brooks, T. Burton, N.-B. Chang *et al.*, *Eur. Phys. J. A* **52**, 268 (2016).
- [27] E.-C. Aschenauer, K. Lee, B. Page, and F. Ringer, *Phys. Rev. D* **101**, 054028 (2020).
- [28] T. Sjöstrand, S. Mrenna, and P. Skands, *J. High Energy Phys.* **05** (2006) 026.
- [29] B. S. Page, X. Chu, and E. Aschenauer, *Phys. Rev. D* **101**, 072003 (2020).
- [30] Z.-B. Kang, X. Liu, S. Mantry, and D. Y. Shao, *Phys. Rev. Lett.* **125**, 242003 (2020).
- [31] M. Arratia, Y. Furlletova, T. Hobbs, F. Olness, and S. J. Sekula, *Phys. Rev. D* **103**, 074023 (2021).
- [32] M. Arratia, Y. Song, F. Ringer, and B. V. Jacak, *Phys. Rev. C* **101**, 065204 (2020).
- [33] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *J. High Energy Phys.* **05** (2014) 146.
- [34] Z.-B. Kang, in *Proceedings of the EPJ Web of Conferences* (EDP Sciences, Les Ulis, France, 2020), Vol. 235, p. 05001, [https://www.epj-conferences.org/articles/epjconf/abs/2020/11/epjconf\\_ismd2019\\_05001/epjconf\\_ismd2019\\_05001.html](https://www.epj-conferences.org/articles/epjconf/abs/2020/11/epjconf_ismd2019_05001/epjconf_ismd2019_05001.html).
- [35] F. Ringer, B.-W. Xiao, and F. Yuan, *Phys. Lett. B* **808**, 135634 (2020).
- [36] R. Devenish and A. Cooper-Sarkar, *Deep Inelastic Scattering* (Oxford University Press, Oxford, 2011).
- [37] M. A. Dobbs, S. Frixione, E. Laenen, K. Tollefson, H. Baer, E. Boos, B. Cox, R. Engel, W. Giele, J. Huston *et al.*, [arXiv:hep-ph/0403045](https://arxiv.org/abs/hep-ph/0403045).
- [38] C. Bierlich, S. Chakraborty, N. Desai, L. Gellersen, I. Helenius, P. Ilten, L. Lönnblad, S. Mrenna, S. Prestel, C. T. Preuss *et al.*, [arXiv:2203.11601](https://arxiv.org/abs/2203.11601).
- [39] H. Jung, *Comput. Phys. Commun.* **86**, 147 (1995).
- [40] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo *et al.*, *Nucl. Phys.* **B867**, 244 (2013).
- [41] A. Buckley, J. Ferrando, S. Lloyd, K. Nordström, B. Page, M. Rüfenacht, M. Schönherr, and G. Watt, *Eur. Phys. J. C* **75**, 1 (2015).
- [42] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **72**, 1896 (2012).
- [43] G. P. Salam, *Eur. Phys. J. C* **67**, 637 (2010).
- [44] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Nucl. Phys.* **B406**, 187 (1993).
- [45] S. D. Ellis and D. E. Soper, *Phys. Rev. D* **48**, 3160 (1993).
- [46] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [47] Y. L. Dokshitzer, G. Leder, S. Moretti, and B. Webber, *J. High Energy Phys.* **08** (1997) 001.
- [48] R. Atkin, *J. Phys. Conf. Ser.* **645**, 012008 (2015).
- [49] O. Gonzalez and J. Terron, Internal ZEUS-Note 02, 2002.
- [50] N. F. Bell, Y. Cai, and R. K. Leane, *J. Cosmol. Astropart. Phys.* **01** (2016) 051.
- [51] D. Krohn, J. Thaler, and L.-T. Wang, *J. High Energy Phys.* **02** (2010) 084.
- [52] F. Willeke and J. Beebe-Wang, Electron Ion Collider Conceptual Design Report 2021. No. BNL-221006-2021-FORE. Brookhaven National Lab. (BNL), Upton, NY (United States); Thomas Jefferson National Accelerator Facility (TJNAF), Newport News, VA (United States) (2021), [10.2172/1765663](https://arxiv.org/abs/10.2172/1765663).