July 15, 2023

# Measurement of CollinearDrop jet mass and its correlation with SoftDrop groomed jet substructure observables in $\sqrt{s} = 200$ GeV $pp$ collisions by STAR

Youqi Song (Wright Laboratory, Yale University)

*on behalf of the STAR Collaboration*

Jet substructure variables aim to reveal details of the parton fragmentation and hadronization processes that create a jet. By removing collinear radiation while maintaining the soft radiation components, one can construct CollinearDrop jet observables, which have enhanced sensitivity to the soft phase space within jets. We present a CollinearDrop jet measurement, corrected for detector effects with a machine learning method – MultiFold – and its correlation with groomed jet observables, in $pp$ collisions at $\sqrt{s} = 200$ GeV at STAR. We demonstrate that the population of jets with a large non-perturbative contribution can be significantly enhanced by selecting on higher CollinearDrop jet mass fractions. In addition, we observe an anti-correlation between the amount of grooming and the angular scale of the first hard splitting of the jet.

PRESENTED AT

DIS2023: XXX International Workshop on Deep-Inelastic Scattering and Related Subjects,
Michigan State University, USA, 27-31 March 2023

# 1 Introduction

Jets are collimated sprays of final-state hadrons produced from initial high-momentum transfer partonic scatterings in particle collisions. One class of jet clustering algorithms, sequential recombination algorithms, relies on recursively recombining two particles/objects into one object by minimizing a distance metric between the two, mimicking the QCD parton shower pattern [1]. Since jets are multi-scale objects that connect asymptotically free partons to confined hadrons, jet substructure measurements can provide insight into the parton evolution and the ensuing hadronization processes.

To enhance perturbative contributions, SoftDrop grooming [2] is often used to remove wide-angle soft radiation from jets. The procedure, detailed in Ref. [2], starts by re-clustering the jet with an angular-ordered sequential recombination algorithm called Cambridge/Aachen [3, 4]. Then the last step of the clustering is undone and the softer prong of the two is removed until the SoftDrop condition specified by $(z_{\rm cut}, \beta)$ is met:

$$z_{\rm g} = \frac{\min(p_{\rm T,1}, p_{\rm T,2})}{p_{\rm T,1} + p_{\rm T,2}} > z_{\rm cut}(R_{\rm g}/R_{\rm jet})^{\beta}, \tag{1}$$

where $z_{\rm cut}$ is the SoftDrop momentum fraction threshold, $\beta$ is an angular exponent, $R_{\rm jet}$ is the jet resolution parameter, $p_{\rm T,1,2}$ are the transverse momenta of the two subjets, and $R_{\rm g}$ is defined as:

$$R_{\rm g} = \sqrt{(y_1 - y_2)^2 + (\phi_1 - \phi_2)^2}, \tag{2}$$

where $y_{1,2}$ and $\phi_{1,2}$ are, respectively, the rapidity values and azimuthal angles of the two subjets. $z_{\rm g}$ and $R_{\rm g}$ describe the momentum imbalance and the opening angle of the SoftDrop groomed jet, respectively.

Although the SoftDrop groomed jet substructure observables have been extensively studied both experimentally [5–10] and theoretically [11], the effects of wide-angle and soft radiation which are suppressed by SoftDrop measurements, have not yet been explored as much. One set of observables that are sensitive to the soft wide-angle radiation are known as the CollinearDrop jet observables [12]. The general case involves the difference of two different SoftDrop selections $(z_{\rm cut,1}, \beta_1)$ and $(z_{\rm cut,2}, \beta_2)$, which reduces the collinear contributions from fragmentation, and the wide-angle contributions from initial-state radiation (ISR), underlying event (UE) and pileup. For this analysis, the less aggressive SoftDrop grooming criteria is set to no grooming, $(z_{\rm cut,1}, \beta_1) = (0, 0)$, so the CollinearDrop groomed observables are the difference in the ungroomed and SoftDrop groomed observables. This simplification can be made since the wide-angle contributions from ISR, UE and pileup are not significant for the dataset used in this analysis. Specifically, the contribution of UE to jet $p_{\rm T}$ for a jet with $20 < p_{\rm T} < 25$ GeV/$c$ is less than 1% [13].

As the QCD parton shower is angular ordered [14], the soft wide-angle radiation captured by the CollinearDrop jet observables on average happens at an early stage of the shower. Unlike the opposite approach of CollinearDrop, SoftDrop will then capture the late stage collinear splittings. Therefore, a simultaneous measurement of a CollinearDrop jet observable and a SoftDrop jet observable can help illustrate the dynamics of the parton shower.

The CollinearDrop jet observable that we focus on is $\Delta M$ or $\Delta M/M$, defined as:

$$\Delta M/M = \frac{M - M_{\rm g}}{M}, \tag{3}$$

where $M$ is the jet mass and $M_{\rm g}$ is the SoftDrop groomed jet mass, defined as:

$$M_{\rm (g)} = \left| \sum\nolimits_{i \in \text{(groomed) jet}} p_i \right| = \sqrt{E_{\rm (g)}^2 - |\vec{\mathbf{p}}_{\rm (g)}|^2}, \tag{4}$$

where $p_i$ is the four-momentum of the $i$th constituent in a (groomed) jet, and $E_{\rm (g)}$ and $\vec{\mathbf{p}}_{\rm (g)}$ are the energy and three-momentum vector of the (groomed) jet, respectively.

The CollinearDrop groomed mass is also expressed as [12]:

$$a = \frac{M^2 - M_{\rm g}^2}{p_{\rm T}^2}. \tag{5}$$

where $p_{\mathrm{T}}^2$ is the jet transverse momentum squared. While $\Delta M$ is a more straightforward measure of the amount of wide-angle soft radiation, $a$ is calculable in Soft Collinear Effective Field Theory (SCET) at the parton level [12].

In these proceedings, we present measurements of the CollinearDrop groomed jet mass, to study the less-explored phase space of soft and wide-angle radiation; we also measure the correlation between the CollinearDrop groomed mass with $R_{\mathrm{g}}$ and $z_{\mathrm{g}}$, in $pp$ collisions at $\sqrt{s} = 200$ GeV at STAR. The measurements are fully corrected for detector effects with MultiFold, a novel machine learning method which preserves the correlations in the multi-dimensional observable phase space. We then compare our fully corrected measurements with predictions from event generators.

## 2 Analysis details

The STAR experiment [15] recorded data from $\sqrt{s} = 200$ GeV $pp$ collisions during the 2012 RHIC run. Tracks are reconstructed from the Time Projection Chamber (TPC), and neutral energy deposits are measured from the Barrel Electro-Magnetic Calorimeter (BEMC) towers. Events are selected within $\pm 30$ cm from the center of the detector along the beam axis, and to pass the jet patch trigger which requires a minimum transverse energy $E_{\mathrm{T}} > 7.3$ GeV deposited in a $1 \times 1$ patch in $\eta \times \phi$ in the BEMC. The same track and tower selections are applied as in Ref. [7] and [10]. We reconstruct jets from TPC tracks ($0.2 < p_{\mathrm{T}} < 30$ GeV/$c$, with a charged pion mass assignment) and BEMC towers ($0.2 < E_{\mathrm{T}} < 30$ GeV, assuming massless) using the anti-$k_{\mathrm{T}}$ sequential recombination clustering algorithm [16] with a resolution parameter of $R = 0.4$. We apply the selections of $p_{\mathrm{T}} > 15$ GeV/$c$, $|\eta| < 0.6$, transverse energy fraction of all neutral components $< 0.9$, and $M > 1$ GeV/$c^2$ on reconstructed jets, consistent with the selections in Ref. [10]. We select jets that pass SoftDrop grooming with the standard cuts of $(z_{\mathrm{cut}}, \beta) = (z_{\mathrm{cut},2}, \beta_2) = (0.1, 0)$.

We measure the following jet observables: $p_{\mathrm{T}}$, $z_{\mathrm{g}}$ (defined in Eq. 1), $R_{\mathrm{g}}$ (defined in Eq. 2), $M$ (defined in Eq. 4), $M_{\mathrm{g}}$ (defined in Eq. 4), and jet charge $Q^{\kappa=2}$. $Q^{\kappa=2}$ is defined as:

$$Q^{\kappa=2} = \frac{1}{p_{\mathrm{Tjet}}^2} \sum_{i \in \mathrm{jet}} q_i \cdot p_{\mathrm{T}_i}^2, \tag{6}$$

where $q_i$ and $p_{\mathrm{T}_i}$ are the electric charge and $p_{\mathrm{T}}$ of the $i$th jet constituent, respectively.

Experimentally, jet measurements need to be corrected for detector effects to compare with theoretical calculations and model predictions. The traditional correction procedure uses the Bayesian inference in as many as three dimensions and requires the observables to be binned [17]. A novel correction procedure, MultiFold [18], uses a machine learning technique to correct with higher dimensionality in an un-binned fashion. Because it preserves the correlation between observables with high dimensionality, MultiFold is potentially more desirable for this study.

We fully corrected these six jet observables simultaneously for detector effects using MultiFold. The particle-level prior used for unfolding is jets from events generated with PYTHIA6 [19] with the STAR tune [20]. This is a single-parameter modification to the Perugia 2012 tune [21] to better match STAR data. The PYTHIA events are run through GEANT3 [22] simulation of the STAR detector, and embedded into data from zero-bias events from the same run period as the analyzed data. The detector-level prior is the jets reconstructed after this embedding procedure and geometrically matched to the particle-level truth jets. The correction was validated using a Monte Carlo closure test, which showed good performance of the unfolding among all observables for jets with $20 < p_{\mathrm{T}} < 50$ GeV/$c$. To correct for fake jets, i.e., detector-level jets arising from background, fake rates were obtained from simulations and used as initial weights for the data as an input to MultiFold. For particle-level jets that are missed at detector level due to effects such as tracking inefficiency, an efficiency correction was done post-unfolding in a multi-dimensional fashion.

The fully corrected jet mass distributions for three different $p_{\mathrm{T}}$ bins are shown in Fig. 1, using both MultiFold and RooUnfold [10]. In the top panels, the distributions obtained with MultiFold are shown to agree with the previously published RooUnfold result; in the bottom panels, the ratios of MultiFold distributions over RooUnfold distributions are confirmed to be consistent with unity. These establish further confidence in application of MultiFold to the data.

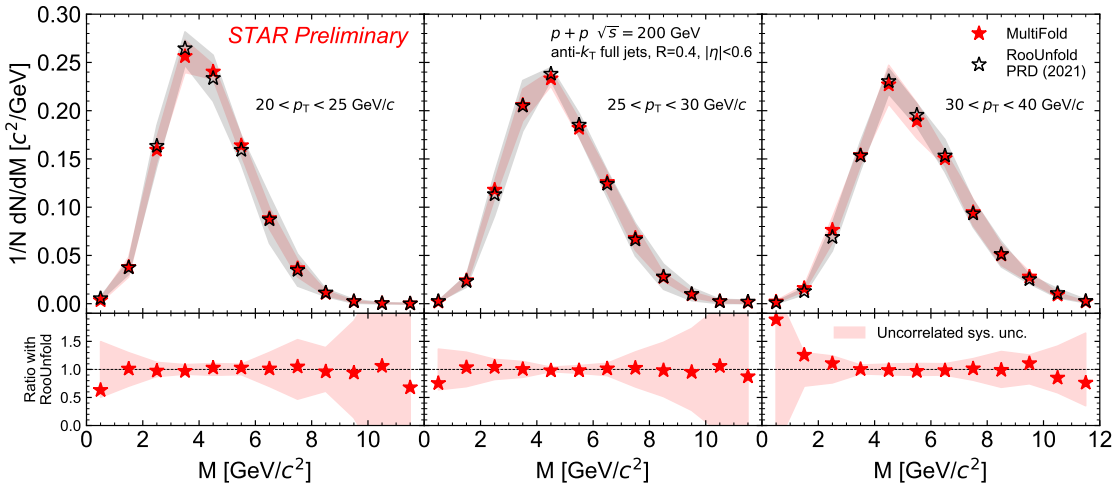


Figure 1: Jet mass distributions corrected with MultiFold and RooUnfold.

The dominant source for systematic uncertainty is from unfolding prior variation. For the MultiFold measurement in Fig. 1, it is accounted for by varying the prior shape of $p_{\mathrm{T}}$ and $M$, as in Ref. [10]. For the following MultiFold measurements, it is accounted for through simultaneous reweighting of all six observables based on prior distributions from PYTHIA [23] and HERWIG [24].

# 3 Results

Figure 2 shows the distribution of fully corrected CollinearDrop groomed jet masses for jets within $20 < p_{\mathrm{T}} < 30$ GeV/$c$. This measurement excludes jets with $M = M_{\mathrm{g}}$ (46% of jets in this $p_{\mathrm{T}}$ range) so that the peak in the small but nonzero $\Delta M$ region is visible. Both PYTHIA8 Detroit tune (tuned to RHIC kinematics) [23] and HERWIG7 H7.1-Default tune (tuned to LHC kinematics) [24] capture the qualitative trend of data, although there is some tension with HERWIG in the small $\Delta M$ region.

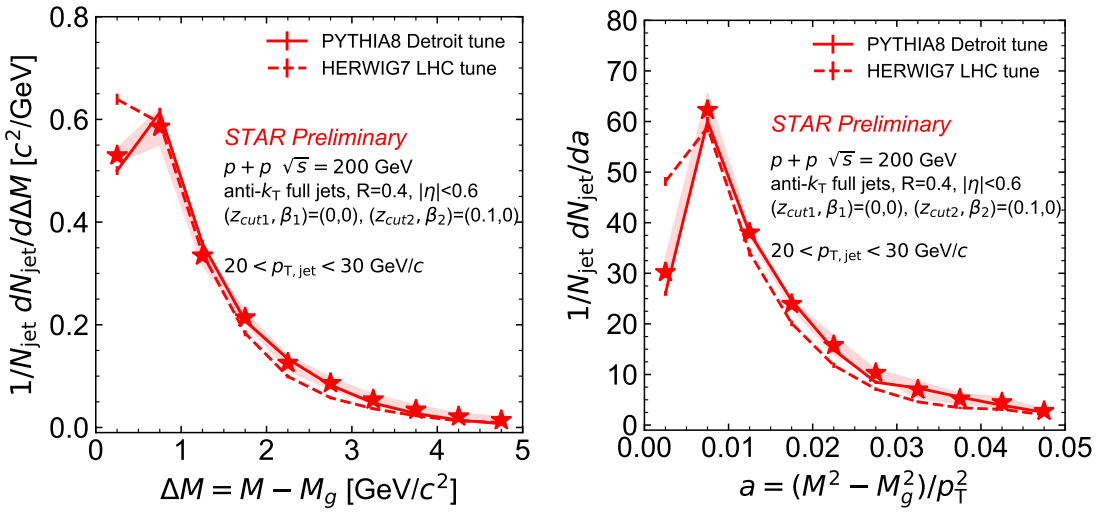


Figure 2: CollinearDrop jet mass distributions. The error bands indicate systematic uncertainties.

Figure 3 (left) shows the correlation between the CollinearDrop groomed mass fraction $\Delta M/M$ and the SoftDrop groomed jet opening angle $R_{\mathrm{g}}$. The jet population with $M = M_{\mathrm{g}}$ has been

separated out into the leftmost column. For the $M > M_{\rm g}$ panel, a diagonal trend that indicates an anti-correlation between the amount of soft radiation and the hard splitting angle is observed, consistent with the expectation of angular ordering of the parton shower. Figure 3 (right) shows the projection of $\Delta M/M$ for different selections of $R_{\rm g}$. We observe that, as shown in the blue data points, a selection on small $R_{\rm g}$ results in a relatively wide $\Delta M/M$, suggesting that a small SoftDrop groomed jet radius appears with a wide range of SoftDrop grooming. On the other hand, as shown in the orange data points, a selection on large $R_{\rm g}$ results in a sharper $\Delta M/M$ peaked towards small $\Delta M/M$ values, suggesting that a large SoftDrop groomed jet radius leaves space for little or no SoftDrop grooming. This measurement demonstrates how early soft wide-angle radiation constrains the angular phase space of later splittings. PYTHIA and HERWIG predictions, as indicated by the solid and dashed lines, describe the trends of the data.

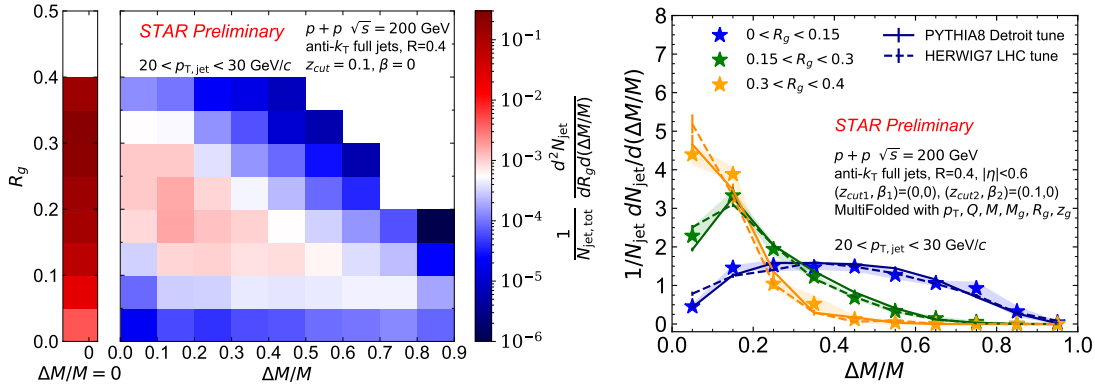


Figure 3: Correlation between the CollinearDrop groomed mass fraction $\Delta M/M$ and the SoftDrop groomed jet opening angle $R_{\rm g}$.

Figure 4 shows the correlation between $\Delta M/M$ and the SoftDrop groomed shared momentum fraction $z_{\rm g}$. We observe that the more fractional mass that is groomed away by SoftDrop, the flatter the $z_{\rm g}$ distribution is. Since the perturbative DGLAP splitting function follows the $1/z$ behavior [25–27], a more steeply falling $z_{\rm g}$ with small $\Delta M$ indicates a larger perturbative contribution. This measurement demonstrates how an early-stage emission constrains the momentum imbalance of a later splitting. Again, PYTHIA and HERWIG are able to describe the data.

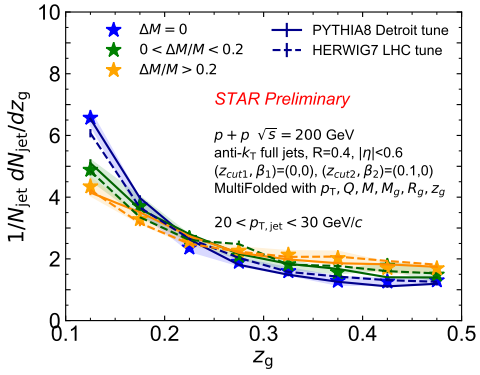


Figure 4: Correlation between the CollinearDrop groomed mass fraction $\Delta M/M$ and the SoftDrop groomed jet opening angle $z_{\rm g}$.

To provide further insight into the relation between the angular and mass scales of a jet, we turn to another new STAR result which reports the measurement of $\mu$ [28], an observable that captures the mass sharing of the hard splitting and is defined as:

$$\mu = \frac{\max(M_1, M_2)}{M_{\rm g}}, \tag{7}$$

where $M_{1,2}$ are the masses of the two subjets. Note that this measurement is corrected for detector effects with a (2+1)D Bayesian unfolding method [29], and does not require $M > 1$ GeV/$c^2$, but imposes otherwise identical selection criteria as described in Sec. 2. Figure 5 shows the distributions of $\mu$ for various ranges of $R_{\rm g}$. We observe that the trend in data is well described by PYTHIA6 STAR tune [20] and PYTHIA8 Monash tune [30], and that $\mu$ has a weaker dependence on $R_{\rm g}$ compared to $\Delta M/M$. This measurement offers complementary information to the measurement in Fig. 3, revealing the correlation between the hard radiation and angle of the hard splitting. The shift of $\mu$ to smaller values at smaller $R_{\rm g}$ indicates that a narrower splitting leads to a smaller transfer of virtuality.

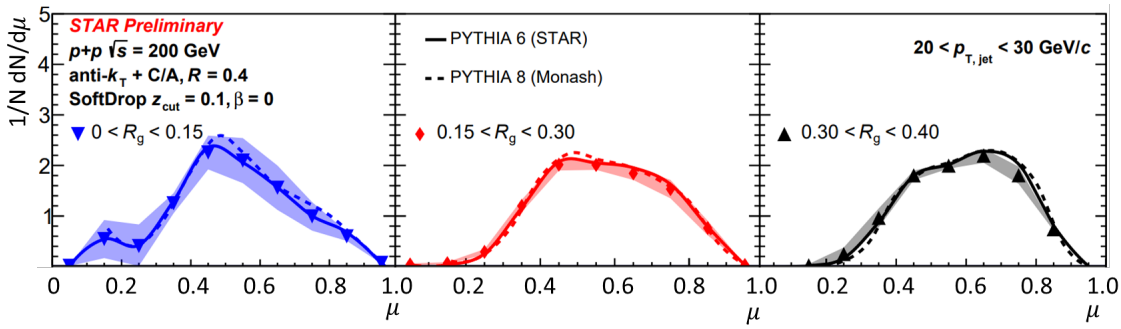


Figure 5: The $\mu$ distributions for various $R_{\rm g}$ selections.

# 4 Conclusions

We have presented measurements of the CollinearDrop groomed jet mass and its correlation with the SoftDrop groomed observables $R_{\rm g}$ and $z_{\rm g}$, in $pp$ collisions at $\sqrt{s} = 200$ GeV at STAR. MultiFold, a machine learning based unfolding method that enables access of multi-dimensional correlations on a jet-by-jet basis, has been shown to offer consistent results as RooUnfold in one-dimensional distributions, and has been applied to correct for the correlation measurements in higher dimensions. These measurements allow for a better understanding of the soft phase space within jets, and highlight how early-stage radiation affects the angular scale and momentum imbalance of a later splitting.

# References


[1] S. Marzani, G. Soyez, and M. Spannowsky, *Looking inside jets* (Springer International Publishing, 2019).

[2] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, "Soft Drop", JHEP **05**, 146 (2014).

[3] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, "Better jet clustering algorithms", JHEP **08**, 001 (1997).

[4] M. Wobisch and T. Wengler, "Hadronization corrections to jet cross-sections in deep inelastic scattering", in Workshop on Monte Carlo Generators for HERA Physics (Plenary Starting Meeting) (1998), pp. 270–279.

[5] A. M. Sirunyan et al. (CMS), "Measurement of the splitting function in $pp$ and Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV", Phys. Rev. Lett. **120**, 142302 (2018).

[6] G. Aad et al. (ATLAS), "Measurement of soft-drop jet observables in $pp$ collisions with the ATLAS detector at $\sqrt{s} = 13$ TeV", Phys. Rev. D **101**, 052007 (2020).

[7]J. Adam et al. (STAR), "Measurement of groomed jet substructure observables in p+p collisions at $\sqrt{s} = 200$ GeV with STAR", Phys. Lett. B **811**, 135846 (2020).

[8]M. Aaboud et al. (ATLAS), "Measurement of the Soft-Drop jet mass in *pp* collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector", Phys. Rev. Lett. **121**, 092001 (2018).

[9]A. M. Sirunyan et al. (CMS), "Measurement of the groomed jet mass in PbPb and pp collisions at $\sqrt{s_{\mathrm{NN}}} = 5.02$ TeV", JHEP **10**, 161 (2018).

[10]M. Abdallah et al. (STAR), "Invariant jet mass measurements in *pp* collisions at $\sqrt{s} = 200$ GeV at RHIC", Phys. Rev. D **104**, 052007 (2021).

[11]A. J. Larkoski, I. Moult, and B. Nachman, "Jet substructure at the Large Hadron Collider: a review of recent advances in theory and machine learning", Phys. Rept. **841**, 1–63 (2020).

[12]Y.-T. Chien and I. W. Stewart, "Collinear Drop", JHEP **06**, 064 (2020).

[13]J. Adam et al. (STAR), "Underlying event measurements in $p$+$p$ collisions at $\sqrt{s} = 200$ GeV at RHIC", Phys. Rev. D **101**, 052004 (2020).

[14]Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller, and S. I. Troian, *Basics of perturbative QCD* (1991).

[15]K. H. Ackermann et al. (STAR), "STAR detector overview", Nucl. Instrum. Meth. A **499**, 624–632 (2003).

[16]M. Cacciari, G. P. Salam, and G. Soyez, "The anti-$k_t$ jet clustering algorithm", JHEP **04**, 063 (2008).

[17]G. D'Agostini, "Improved iterative Bayesian unfolding", `10.48550/arxiv.1010.0632` (2010).

[18]A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler, "OmniFold: a method to simultaneously unfold all observables", Phys. Rev. Lett. **124**, 182001 (2020).

[19]T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 physics and manual", JHEP **05**, 026 (2006).

[20]J. K. Adkins, "Studying transverse momentum dependent distributions in polarized proton collisions via azimuthal single spin asymmetries of charged pions in jets", PhD thesis (Kentucky U., 2015).

[21]P. Z. Skands, "Tuning Monte Carlo generators: the Perugia tunes", Phys. Rev. D **82**, 074018 (2010).

[22]R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban, "GEANT detector description and simulation tool", `10.17181/CERN.MUHF.DMJ1` (1994).

[23]M. R. Aguilar, Z. Chang, R. K. Elayavalli, R. Fatemi, Y. He, Y. Ji, D. Kalinkin, M. Kelsey, I. Mooney, and V. Verkest, "PYTHIA 8 underlying event tune for RHIC energies", Phys. Rev. D **105**, 016011 (2022).

[24]J. Bellm et al., "Herwig 7.0/Herwig++ 3.0 release note", Eur. Phys. J. C **76**, 1–8 (2016).

[25]V. N. Gribov and L. N. Lipatov, "Deep inelastic ep scattering in perturbation theory", Sov. J. Nucl. Phys. **15**, 438–450 (1972).

[26]G. Altarelli and G. Parisi, "Asymptotic freedom in parton language", Nucl. Phys. B **126**, 298–318 (1977).

[27]Y. L. Dokshitzer, "Calculation of the structure functions for Deep Inelastic Scattering and $e^+e^-$ annihilation by perturbation theory in Quantum Chromodynamics", Sov. Phys. JETP **46**, 641–653 (1977).

[28]M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, "Towards an understanding of jet substructure", JHEP **09**, 029 (2013).

[29]M. Robotková (STAR), "Multi-dimensional measurements of parton shower in $p + p$ collisions at RHIC", PoS **ICHEP2022**, 1183 (2022).

[30]P. Skands, S. Carrazza, and J. Rojo, "Tuning PYTHIA 8.1: the Monash 2013 tune", Eur. Phys. J. C **74**, 3024 (2014).