

ANALYSIS NOTE: MEASUREMENT OF COLLINEARDROP JET MASS AND ITS CORRELATION WITH SUBSTRUCTURE OBSERVABLES IN pp COLLISIONS

YOUQI SONG

CONTENTS

1. Introduction	1
1.1. Motivation for the paper	1
1.2. Motivation for the initial study	3
1.3. Shortcuts to paper plots and the relevant sections	3
2. Datasets, embedding and jet reconstruction	3
2.1. Datasets and event selection	3
2.2. Embedding	4
2.3. Jet reconstruction and matching	4
2.4. MC event generation	6
3. Unfolding	6
3.1. Method	6
3.2. Closure test	8
3.3. Application on data and obtaining the fully corrected distributions	9
4. Statistical uncertainties	13
5. Systematic uncertainties	13
5.1. Variation of detector effects	13
5.2. Variation of the random seed	13
5.3. Variation of the number of iterations	13
5.4. Prior variation	14
6. Results	15
6.1. Jet mass	15
6.2. Jet charge vs mass	16
6.3. Correlations between each pair of substructure observables	19
6.4. CollinearDrop jet mass	19
6.5. CollinearDrop jet mass vs R_g or z_g	23
6.6. CollinearDrop jet mass vs R_g and z_g	30
7. Conclusions	32
References	33

1. INTRODUCTION

High-energy particle collisions provide opportunities to study experimentally quarks and gluons (partons), the fundamental degree of freedom in the theory of Quantum Chromodynamics (QCD). In some of these collisions, incoming quarks and gluons (partons) interact with each other through the exchange of a high-momentum virtual particle, producing outgoing partons with high transverse momentum (p_T). Such outgoing partons are highly virtual and will undergo a series of splitting processes as they come on mass shell. This process is called the parton shower, and can be described perturbatively in terms of the Dokshitzer–Gribov–Lipatov–Altarelli–Parisi (DGLAP) evolution equations [1–3]. When the virtuality of the partons is comparable to the confinement scale Λ_{QCD} , the non-perturbative transition to baryons and mesons (hadrons), known as hadronization, begins. Experimentally, the spray of the final-state hadrons can be measured and clustered into jets. Jets reconstructed with a clustering algorithm [4] can serve as a proxy for the kinematics of the outgoing partons.

1.1. Motivation for the paper. While the interaction among partons can be well understood with the principles of perturbative QCD (pQCD), the non-perturbative components of the parton shower and hadronization remain challenging for theoretical calculations and rely mostly on phenomenological models in Monte Carlo event generators. Measurements of observables sensitive to such non-perturbative QCD (npQCD) effects will provide important tests for the theories and constraints on the models.

E-mail address: youqi.song@yale.edu.

Date: March 11, 2024.

Together with studies of observables calculable from pQCD, investigation of those sensitive to npQCD effects offers an avenue for a comprehensive understanding of the full parton-to-hadron evolution picture.

Beyond the jet p_T , or other combinations of the jet four-momentum observables, jet substructure observables [5] are useful tools that can provide insight into the parton shower and hadronization processes. To enhance perturbative contributions, SoftDrop [6] grooming is often used to remove wide-angle soft radiation within the jet. The procedure, detailed in Ref. [6], starts by re-clustering the jet with an angular-ordered sequential recombination algorithm called Cambridge/Aachen [7, 8]. Then the last step of the clustering is undone and the softer prong is removed based on the SoftDrop condition:

$$(1) \quad z_g = \frac{\min(p_{T,1}, p_{T,2})}{p_{T,1} + p_{T,2}} > z_{\text{cut}}(R_g/R_{\text{jet}})^\beta,$$

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

$$(2) \quad R_g = \sqrt{(y_1 - y_2)^2 + (\phi_1 - \phi_2)^2},$$

where $y_{1,2}$ and $\phi_{1,2}$ are, respectively, the rapidities and azimuthal angles of the two subjects. z_g and R_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 [9–14] and theoretically [15], the wide-angle and soft radiation which are dominated by npQCD processes, have not yet been explored in detail.

One set of observables that are sensitive to the soft wide-angle radiation are known as CollinearDrop [16]. The general case involves the difference of two different SoftDrop selections $\text{SD}_1 = (z_{\text{cut},1}, \beta_1)$ and $\text{SD}_2 = (z_{\text{cut},2}, \beta_2)$ on the same jet. For nonzero values of SD_1 and SD_2 parameters with $z_{\text{cut},1} \leq z_{\text{cut},2}$ and $\beta_1 \geq \beta_2$, SD_2 aims to reduce the collinear contributions from fragmentation, and SD_1 aims to reduce the wide-angle contributions from initial-state radiation (ISR), underlying event (UE) and pileup.

As the QCD parton shower is angular ordered [17], the soft wide-angle radiation captured by the CollinearDrop jet observables happens on average at an early stage of the shower. Unlike CollinearDrop, SoftDrop then captures the late stage collinear and perturbative splittings. Therefore, a simultaneous measurement of CollinearDrop jet and SoftDrop jet observables can help illustrate the hard-soft dynamics in the parton shower.

The CollinearDrop jet mass is defined in terms of the ungroomed jet mass M and the SoftDrop groomed jet mass M_g :

$$(3) \quad M_{(g)} = \left| \sum_{i \in (\text{groomed}) \text{ jet}} p_i \right| = \sqrt{E_{(g)}^2 - |\vec{\mathbf{p}}_{(g)}|^2},$$

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

$$(4) \quad a = \frac{M^2 - M_g^2}{p_T^2}.$$

a is calculable in Soft Collinear Effective Field Theory (SCET) at the parton level [16]. As shown in Figure 1, such an observable is sensitive to soft radiation within the ring, where the contribution of underlying event and pileup is suppressed.

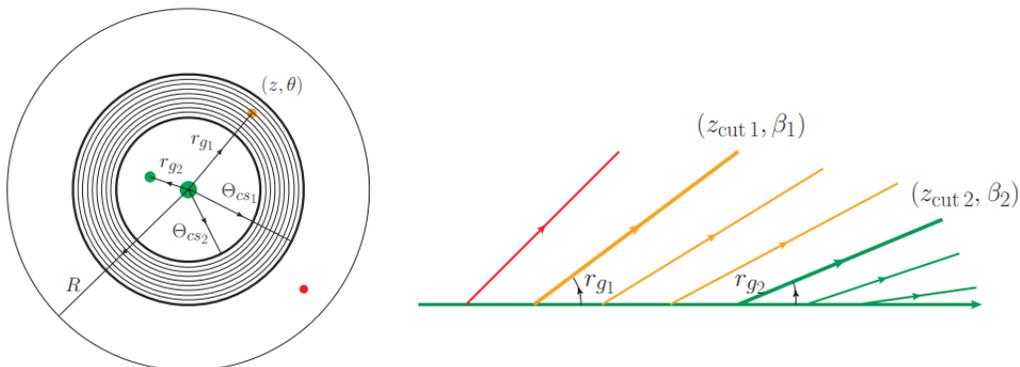


FIGURE 1. CollinearDrop schematic.

1.2. **Motivation for the initial study.** Jet substructure measurements can also be used to tag the flavor of the parton that initiated the shower. On an ensemble level, knowledge of whether a jet population is quark- or gluon-dominated, together with measurement of other jet observables, can be used in pp collisions as a precision QCD test, and in heavy-ion collisions as a handle on jet energy loss mechanism in the quark-gluon plasma. In this analysis, we explore the possibility of separating a jet population by its quark/gluon jet fraction through a correlation measurement between jet mass observable and jet charge.

Jet charge is defined as

$$(5) \quad Q^\kappa = \frac{1}{(p_{T\text{jet}})^\kappa} \sum_{i \in \text{jet}} q_i \cdot (p_{Ti})^\kappa$$

where q_i and p_{T_i} are the electric charge and p_T of the i th jet constituent, respectively, and κ is a tuning parameter that determines the relative importance of soft/hard charged particles within the jet. This is designed to be a proxy for the electric charge of the parton that initiates the parton shower.

Jet mass is related to the QCD coupling while jet charge is related to the QED coupling, both of which are different for a quark vs. for a gluon. In particular, in the soft and collinear limit and in the absence of hadronization,

$$(6) \quad \langle M_{\text{jet}}^2 \rangle \approx \frac{\alpha_S C_{F/A}}{\pi} p_{T\text{jet}}^2 R_{\text{jet}}^2$$

where α_S is the strong coupling constant, R_{jet} is the jet radius, and $C_{F/A}$ is the quadratic Casimir operator of the $SU(3)$ gauge group in the fundamental/adjoint representation, where quarks are in the fundamental representation and gluons are in the adjoint representation. Since $C_A/C_F = 9/4$, we expect gluon jets to have larger masses on average. Since quarks are electrically charged while gluons are not, we also expect gluon jets to have a smaller $|Q|$ on average.

To compare with theory or model predictions or similar measurements done with data taken at other detectors, jet measurements need to be corrected for detector effects to uncover the “truth” information. This procedure is called unfolding. Traditionally this is done with an iterative Bayesian unfolding (IBU) method [18]. In this analysis, we explore a novel machine-learning based unfolding method, MultiFold, which can correct for detector effects in multiple dimensions in an un-binned fashion, therefore retaining the correlation between various substructure observables by design and suitable for a multi-dimensional jet substructure correlation measurement.

We measure the following jet observables: p_T , z_g (defined in Eq. 1), R_g (defined in Eq. 2), M (defined in Eq. 3), M_g (defined in Eq. 3), and $Q^{\kappa=2}$ (defined in Eq. 5). We explain the choice of $\kappa = 2$ for the jet charge in 6.2.1. We will suppress the $\kappa = 2$ superscript unless specified otherwise.

1.3. **Shortcuts to paper plots and the relevant sections.** Our paper plots include Figs 27, 50 and 51, discussed in Sections 6.4 and 6.6.

2. DATASETS, EMBEDDING AND JET RECONSTRUCTION

2.1. **Datasets and event selection.** The STAR experiment [19] recorded data from $\sqrt{s} = 200$ GeV pp collisions during the 2012 RHIC run. As energetic charged particles travel from the interaction point to the perimeter of the Time Projection Chamber (TPC), they ionize the gas atoms in the TPC and leave hits, from which we reconstruct tracks. Neutral particles do not interact with the gas in the TPC and instead deposit their energy through the development of electromagnetic showers in Barrel Electro-Magnetic Calorimeter (BEMC) towers. Events are required to pass the “Jet Path 2” trigger, which requires that a minimum transverse energy $E_T > 7.3$ GeV be deposited in a 1×1 patch in $\eta \times \phi$ in the BEMC. Before any run selections, 65M events pass this trigger selection, corresponding to an integrated luminosity of 23 pb^{-1} . In addition, events are required to have primary vertices within ± 30 cm from the center of the detector along the beam axis. We apply a 100% hadronic correction to tower energy measurement: if a charged track extrapolates to a tower, then the whole track’s p_T is removed from the tower E_T . If the track p_T is greater than the tower E_T , then the tower is removed completely. The same track and tower selections are applied as in Ref. [11] and [14], and summarized in Table 2.

The bad run list is the same as [20]: 13040003, 13040016, 13040017, 13040018, 13040037, 13040040, 13040041, 13040042, 13040043, 13040044, 13040045, 13040046, 13041002, 13041004, 13041006, 13041007, 13041008, 13041009, 13041010, 13041082, 13041101, 13041102, 13041108, 13041109, 13041110, 13041115, 13041116, 13041119, 13042001, 13042002, 13042003, 13042006, 13042007, 13042008, 13042013, 13042015, 13042016, 13042017, 13042018, 13042019, 13042020, 13042022, 13042023, 13042024, 13042025, 13042026, 13042027, 13042046, 13042047, 13042048, 13042049, 13042050, 13043005, 13043006, 13043011, 13043012, 13043013, 13043017, 13043023, 13043030, 13043031, 13043032, 13043035, 13043044, 13043047, 13043048, 13043049, 13043050, 13043051, 13043052, 13043053, 13043054, 13043055, 13043056, 13043057, 13043058, 13043059, 13043060, 13043063, 13044017, 13044018, 13044019, 13044020, 13044021, 13044022, 13044023, 13044024, 13044025, 13044026, 13044029, 13044030, 13044031, 13044034, 13044118, 13044119, 13044121, 13044122, 13044123, 13044124, 13044125, 13044126, 13045001, 13045003, 13045005, 13045006, 13045007, 13045012, 13045029, 13045056, 13045133, 13045134, 13045135, 13045138, 13045145, 13045146, 13045164, 13046001, 13046002, 13046003, 13046004, 13046008, 13046010, 13046011, 13046012, 13046013, 13046014, 13046015, 13046017, 13046028, 13046029, 13046118, 13046119, 13046120, 13046121, 13047002, 13047003, 13047004, 13047014, 13047018, 13047022, 13047023, 13047024, 13047026, 13047027, 13047028, 13047029, 13047030, 13047031, 13047032, 13047033, 13047034, 13047035, 13047036, 13047037, 13047039, 13047040, 13047041, 13047042, 13047043, 13047044, 13047045, 13047046, 13047047, 13047048, 13047049, 13047050, 13047051, 13047052, 13047053, 13047054, 13047055, 13047122, 13047123, 13047124, 13047126, 13048007, 13048022, 13048046, 13049004, 13049005, 13049050, 13049052, 13049075, 13049086, 13049087, 13049088, 13049089, 13050007, 13050025, 13050026, 13050027, 13050033, 13050039, 13050042, 13050043, 13050044, 13050046, 13050047, 13050049, 13050050, 13051068, 13051080, 13051088, 13051095, 13051102, 13052021, 13052022, 13052054, 13052063, 13052068, 13053010, 13053021, 13054004, 13054005, 13054006, 13054007, 13054008, 13054009, 13054011, 13054012, 13054013, 13054014, 13054015, 13054016, 13054017, 13054018, 13054019, 13054020, 13054022, 13054042, 13054045, 13054046, 13054057, 13055015, 13055072, 13055081, 13055082, 13055086, 13055087, 13055088, 13055089, 13055090, 13056011, 13056012, 13056034, 13056035, 13056037, 13056038, 13056039, 13057038, 13057039,

129 13058019, 13058030, 13058047, 13058048, 13059003, 13059004, 13059005, 13059006, 13059007, 13059008, 13059009, 13059010, 13059011, 13059012, 13059013, 13059014,
 130 13059015, 13059016, 13059017, 13059018, 13059019, 13059020, 13059021, 13059022, 13059023, 13059025, 13059026, 13059027, 13059035, 13059037, 13059038, 13059039,
 131 13059076, 13059077, 13059078, 13059079, 13059080, 13059082, 13059083, 13059084, 13059085, 13059086, 13059087, 13060001, 13060002, 13060003, 13060009, 13060012,
 132 13061026, 13063033, 13064030, 13064057, 13064059, 13064074, 13065048, 13066035, 13066036, 13066101, 13066102, 13066104, 13066109, 13066110, 13067001, 13067002,
 133 13067003, 13067004, 13067005, 13067006, 13067007, 13067008, 13067009, 13067010, 13067011, 13067012, 13067013, 13067014, 13067015, 13067017, 13068017, 13068022,
 134 13068027, 13068029, 13068034, 13068036, 13068037, 13069006, 13069009, 13069029, 13070030, 13070056, 13071034, 13071037, 13071038, 13071040.

135 The bad tower list is also the same as [20]: 34, 95, 106, 113, 160, 182, 266, 267, 275, 280, 282, 286, 287, 293, 308, 410, 504, 533, 541, 555, 561, 562, 594,
 136 609, 615, 616, 629, 633, 637, 638, 647, 650, 653, 657, 671, 673, 743, 789, 790, 791, 792, 806, 809, 810, 811, 812, 813, 814, 821, 822, 823, 824, 829, 830, 831, 832, 837, 841, 842,
 137 843, 844, 846, 849, 850, 851, 852, 857, 873, 875, 882, 897, 899, 903, 939, 953, 954, 956, 993, 1026, 1028, 1044, 1046, 1048, 1080, 1081, 1100, 1125, 1130, 1132, 1180, 1197,
 138 1198, 1199, 1200, 1204, 1207, 1217, 1218, 1219, 1220, 1221, 1222, 1223, 1224, 1233, 1237, 1238, 1240, 1241, 1242, 1243, 1244, 1257, 1258, 1259, 1260, 1294, 1312, 1348, 1353,
 139 1354, 1375, 1382, 1388, 1407, 1409, 1434, 1448, 1537, 1567, 1574, 1575, 1597, 1612, 1654, 1668, 1709, 1713, 1762, 1765, 1766, 1877, 1878, 1984, 2032, 2043, 2054, 2073, 2077,
 140 2092, 2093, 2097, 2107, 2162, 2168, 2214, 2305, 2392, 2409, 2415, 2439, 2445, 2459, 2589, 2590, 2633, 2652, 2749, 2834, 2961, 2969, 3005, 3017, 3070, 3071, 3186, 3220, 3289,
 141 3360, 3473, 3493, 3494, 3495, 3508, 3588, 3604, 3611, 3668, 3678, 3679, 3690, 3692, 3718, 3732, 3738, 3821, 3838, 3840, 3927, 3945, 4005, 4006, 4013, 4017, 4018, 4019, 4053,
 142 4057, 4059, 4124, 4331, 4343, 4355, 4357, 4458, 4464, 4500, 4677, 4678, 4684, 4768, 360, 493, 779, 1284, 1306, 1337, 1438, 1709, 2027, 2445, 3407, 3720, 4217, 4288. **Note**
 143 **briefly for the record how the bad run and tower lists were created.**

144 **2.2. Embedding.** Embedding sample contains particle-level and detector-level jets used for correction of detector effects. The
 145 particle-level events are generated with PYTHIA6 [21] with the STAR tune [22]. This is a single-parameter modification to the
 146 Perugia 2012 tune [23] to better match STAR data. The selections on these PYTHIA events and particles are summarized in
 147 Table 1.

148 The PYTHIA events are then run through GEANT3 [24] simulation of the STAR detector, and embedded into data from
 149 zero-bias events from the same run period as the analyzed data. The selections on the embedding events, tracks and towers are
 150 the same as the ones on data, and are summarized in Table 2.

151 Three versions of embedding samples are relevant to this analysis.

- 152 • 2015 embedding This is the embedding used by previous analyses [11] and [14], but has relatively low statistics. Most
 153 of the studies shown in the note as of now are based on this embedding, unless specified otherwise. The \hat{p}_T bins and
 154 weights for event generation can be found in the embedding Drupal page.
- 155 • 2021 embedding This is the embedding used by an ongoing analysis [**Dmitry’s analysis**] and has a higher statistics,
 156 but has π^0 , η and Σ^0 decayed at the particle-level. In the 2015 embedding, these particles were kept stable at the
 157 particle-level and only decayed at the detector-level in GEANT. The distributions of some jet substructure observables
 158 are no longer comparable with analyses published with the 2015 embedding. In addition, there is an increase in fake
 159 rate for jets and high- p_T tracks seen in this embedding, which we do not understand yet. The \hat{p}_T bins and weights for
 160 event generation can be found in the embedding Drupal page.
- 161 • 2024 embedding This embedding will have high statistics, the same decay treatment as the 2015 embedding, and rea-
 162 sonable fake rate for jets, but has not been produced yet.

163 Comparisons of jet observable distributions between the 2015 embedding (“old embedding”), 2021 embedding (“new embed-
 164 ding”) and data are shown in Figure 2. The agreement among them demonstrates the validity of the embedding procedure.
 165 The definitions for these observables are given in the end of Section 1. The details of jet reconstruction are given in the next
 166 subsection. **How do you mock up the high Tower trigger in the embedding?**

167 **2.3. Jet reconstruction and matching.** We reconstruct jets from TPC tracks ($0.2 < p_T < 30$ GeV/ c , with a charged pion
 168 mass assignment) and BEMC towers ($0.2 < E_T < 30$ GeV, assuming massless) using the anti- k_T sequential recombination
 169 clustering algorithm [4] with a resolution parameter of $R = 0.4$. We apply the selections of $p_T > 15$ GeV/ c , $|\eta| < 0.6$, transverse
 170 energy fraction of all neutral components < 0.9 , and $M > 1$ GeV/ c^2 on reconstructed jets, consistent with the selections in Ref.
 171 [14]. Similar to Ref. [11] and [14], no background subtraction is done, because the UE contribution to jets is low for STAR
 172 kinematics and unfolding can correct for any fluctuation in it. In addition, we select jets that pass SoftDrop grooming with
 173 the standard cuts of $(z_{\text{cut}}, \beta) = (z_{\text{cut},2}, \beta_2) = (0.1, 0)$. For this analysis, the less aggressive SoftDrop grooming criteria is set to
 174 no grooming, $(z_{\text{cut},1}, \beta_1) = (0, 0)$, so the CollinearDrop groomed observables are the difference in the ungroomed and SoftDrop
 175 groomed observables. This simplification can be made since the wide-angle contributions from ISR, UE and pileup are not
 176 significant for the dataset used in this analysis. Specifically, the contribution of UE to jet p_T for a jet with $20 < p_T < 25$ GeV/ c
 177 is less than 1% [25].

178 The method for jet reconstruction (anti- k_T , $R = 0.4$, full jets reconstructed with charged particle tracks measured in the
 179 Time Projection Chamber and neutral towers measured in the Barrel Electromagnetic Calorimeter), and event/track/tower/jet
 180 selections used in this analysis are summarized in Tables 1 and 2.

181 For the events that pass the selections in embedding, we reconstruct detector-level jets, and geometrically match a detector-
 182 level jet to a particle-level jet with the following procedure:

- 183 • If embedding event does not pass the event cut, but PYTHIA event does, then all the particle-level jets in the event are
 184 **missed**
- 185 • If both the embedding and PYTHIA events pass the event cut, sort the detector-level and particle-level jets by p_T into
 186 2 lists
- 187 • Starting from the highest p_T particle-level jet in the list, calculate $\Delta R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2}$ with the highest p_T detector-
 188 level jet.

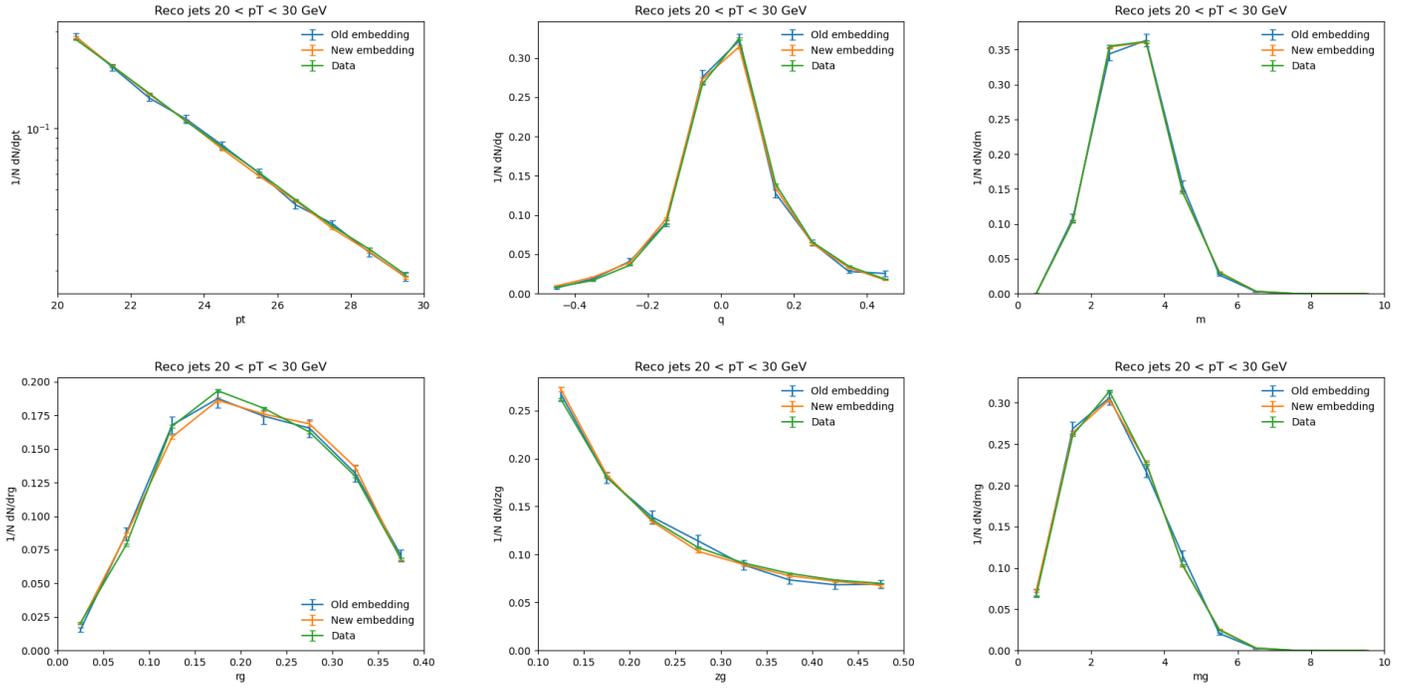


FIGURE 2. Jet observable distributions for reconstructed jets in embeddings and data.

	Selections
Events	events with a jet having a p_T value greater than twice the upper value of the \hat{p}_T bin removed, events with $ v_z > 30$ cm removed, events whose embedding is in bad run list removed
Charged particles	$0.2 < p_T < 30$ GeV/c, $ \eta < 1$
Neutral particles	$0.2 < E_T < 30$ GeV, $ \eta < 1$
Jets	anti- k_T $R = 0.4$, $p_T > 5$ GeV/c, $ \eta < 0.6$, pass $(z_{\text{cut}}, \beta) = (0.1, 0)$

TABLE 1. PYTHIA selections

	Selections
Events	events with a jet having a p_T value greater than twice the upper value of the \hat{p}_T bin removed, events with $ v_z > 30$ cm removed, events whose embedding is in bad run list removed, events that do not pass JP2 trigger ($E_T > 7.3$ GeV) removed, events that have a track $p_T > 30$ GeV/c or a tower $E_T > 30$ GeV removed
Tracks	$0.2 < p_T < 30$ GeV/c, $ \eta < 1$, DCA < 1 cm, Nhits,fit ≥ 20 , Nhits,fit/Nhits,max ≥ 0.52
Towers	$0.2 < E_T < 30$ GeV, $ \eta < 1$ not in bad tower list
Jets	anti- k_T $R = 0.4$, $p_T > 15$ GeV/c, $ \eta < 0.6$, $M > 1$ GeV/c ² , neutral p_T scalar sum fraction < 0.9 , pass $(z_{\text{cut}}, \beta) = (0.1, 0)$

TABLE 2. Embedding and data selections

- If $\Delta R < 0.4$, then these 2 jets are **matched**. Remove them from the lists
- If $\Delta R > 0.4$, then repeat this comparison for the highest p_T PYTHIA jet with the second highest p_T embedding jet in the list
- If no detector-level jet has $\Delta R < 0.4$ with the particle-level jet, then the PYTHIA jet is **missed**
- If after all the particle-level jets are looped through, there are still jets in the detector-level jet list left unmatched, then they are **fakes**

The effect of the selections and jet matching on the efficiency as a function of particle-level jet p_T is shown in Figure 3.

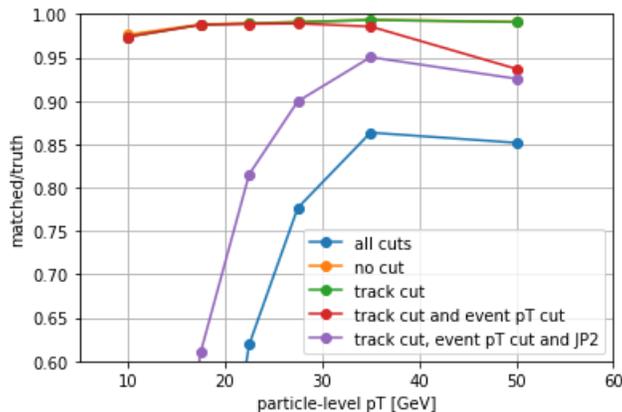


FIGURE 3. Number of matched reconstructed jets in the embedding sample dividing by the number of truth jets, as a function of truth jet p_T .

\hat{p}_T boundaries	weights from PYTHIA	weights from HERWIG
15-20	1.9	1.7
20-25	0.25	0.24
25-30	4.3E-2	4.3E-2
30-35	8.9E-3	8.9E-3
35-40	2.0E-3	2.0E-3
40-45	4.7E-4	4.5E-4
45-50	1.1E-4	1.0E-4
50-2000	3.6E-5	2.9E-5

TABLE 3. Event weights for PYTHIA8 and HERWIG simulations

24. **MC event generation.** For comparison with Monte Carlo event generator predictions and estimate of systematic uncertainties due to prior variation, we also generate events with PYTHIA 8.303 Detroit tune [26] and HERWIG 7.2.2 Default tune [27]. To ensure that the generated events contain hard scattering processes, we set “HardQCD:all = on” in PYTHIA and “insert SubProcess:MatrixElements[0] MEQCD2to2” in HERWIG. 200,000 events are generated for each \hat{p}_T bin, and weights (not normalized) are determined from cross sections given by PYTHIA and HERWIG. Table 3 lists the cuts on \hat{p}_T bin and the weights in PYTHIA and HERWIG. Similar selections for particles and jets are made as in Table 1.

3. UNFOLDING

3.1. **Method.** The goal of unfolding is to “undo” the detector effects on raw data and extract the truth distributions of jet observables, i.e., the distributions of what we would measure if our detector were perfect. Jet substructure observables are usually unfolded with a binned, one- or two-dimensional IBU method. Potentially, it is more desirable to unfold in higher dimensions which can account for the possible correlation in the multi-dimensional observable phase space. This can be done with MultiFold, a machine learning method that simultaneously corrects for multiple observables in an un-binned fashion.

It has been shown that MultiFold and IBU both converge to the maximum likelihood estimate of the true particle-level distribution [28]. However, there are several advantages of MultiFold over IBU, as was also detailed in [28]. First, IBU requires binning of the data beforehand, while MultiFold is unbinned. While experimental data is usually presented in a binned fashion, with MultiFold one can choose such a binning after unfolding, possibly taking into account the uncertainties that might have arisen from the unfolding procedure, but with IBU the choice would be more arbitrary, and one needs to redo the unfolding if the binning needs to be changed. Second, IBU cannot unfold for more than three observables simultaneously while MultiFold can. In this analysis, we unfold for six jet observables, p_T , charge, mass, groomed jet radius, subjet shared momentum fraction, and groomed jet mass. This is not feasible with IBU. (In principle, MultiFold can be generalized to OmniFold, which can unfold for the full phase space information). Third, since the multi-dimensional phase space at the raw data level is used as input for the unfolding procedure, the output of MultiFold retains the correlation information between the observables, something that would be hard to access with IBU.

We fully corrected six jet observables simultaneously for detector effects using MultiFold. In addition to jets from data, matched pairs of jets from simulations with (detector-level) and without (particle-level) detector effects are input for MultiFold. See Section 2.2 for the procedures for creating these matched jet pairs.

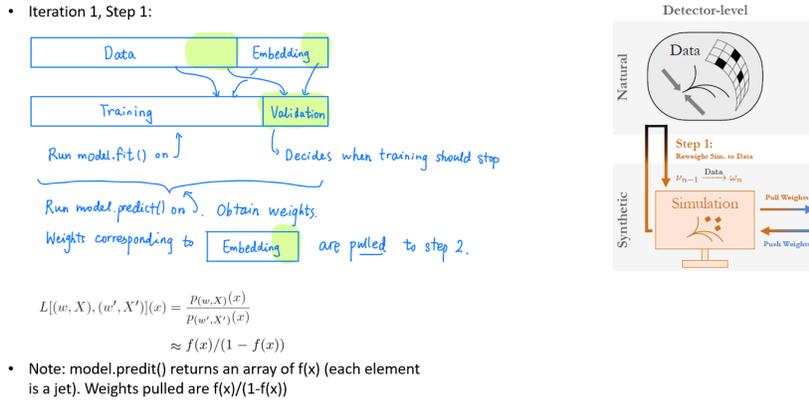


FIGURE 4. Iteration 1 Step 1

MultiFold achieves likelihood-free inference through iterative reweighting. For an overview of the MultiFold method, see [28]. A key concept for this reweighting problem is the likelihood ratio, defined in [28] as,

$$(7) \quad L[(w, X), (w', X')](x) = \frac{p(w, X)(x)}{p(w', X')(x)}$$

“where $p(w, X)$ is the probability density of x estimated from empirical weights w and sample X , which can be approximated using a classifier to distinguish (w, X) from (w', X') ” [29]:

$$(8) \quad L[(w, X), (w', X')](x) \approx \frac{f(x)}{1 - f(x)}$$

where $f(x)$ is “a neural network and trained with the binary cross-entropy loss” (loss function for a categorization problem). “Here, we use neural network classifiers to iteratively reweight the particle-and detector-level Monte Carlo weights,” turning the reweighting problem into a classification problem that can be accomplished through machine learning.

Each iteration of MultiFold has two steps. Intuitively, in step 1, detector-level jets are reweighted to match data jets, and in step 2, particle-level jets with the new weights of the detector-level jets are further reweighted to estimate the truth. In each of the two steps, reweighting is done with a dense neural network architecture. There are three hidden layers, each with 100 nodes. The activation function for the dense layer is rectified linear unit. The activation function for output layer is sigmoid. The loss function is binary cross-entropy. The optimization algorithm is Adam. The input dimension is the number of jet observables (6 in our case). The output dimension is 2 (corresponding to either sample (w, X) or (w', X')). We keep these parameters and architecture setup as default.

The details of the method are as followed. We use the notation from [28] where m denotes a jet in the detector-level jet phase space (i.e. from data or embedding (emb)), and t denotes a jet in the particle-level phase space (i.e. from PYTHIA). The schematic of step 1 of the first iteration is shown in Figure 4. Here we first separate the data jets (each with weight 1) and embedding jets (each with the corresponding event cross section weight $\nu_0(m \in \text{emb})$) into a training set and a validation set. We use the default values of training fraction of 80% and validation fraction of 20%. A dense neural network model is trained to distinguish the data jets coming from the training set from the embedding jets coming from the training set. The validation set is used to decide when the training should stop. By default, the number of epochs per step is 100, and if the loss value for the validation set stops improving after 10 epochs (patience), then the training stops. We use a patience value of 50 since it gives us better closure than a patience value of 10 (show plots comparison here). The trained model $f(x)$ is then applied to do prediction (in batches, with batch size determined by the sample size) on the entire sample of data and embedding jets (both training and validation included). In machine learning, prediction is typically done on samples that are not trained on, but this is acceptable here since the procedure is repeated for multiple iterations. (Hannah said something like this, need to confirm). The prediction for the detector-level jets $f(m \in \text{emb})$ are used to calculate new weights:

$$(9) \quad \omega_1(m \in \text{emb}) = \nu_0(m \in \text{emb}) \frac{p(1, \text{data})(m)}{p(\nu_0, \text{emb})(m)} \approx \nu_0(m \in \text{emb}) \frac{f(m \in \text{emb})}{1 - f(m \in \text{emb})}$$

The weights $\omega_1(m \in \text{emb})$ are “pulled” to step 2.

In step 2 of the first iteration (see Figure 5), the datasets of interest here are two sets of PYTHIA jets, with the same observable values but different weights. The first one is with weights “pulled” from step 1, $\omega_1(t) = \omega_1(m \in \text{emb})$. The second set is with weights as the original event cross section weights $\nu_0(t) = \nu_0(m \in \text{emb})$. Naively, we can reweight the second set to the first set by simply taking the ratios of $\omega_1(t)/\nu_0(t)$. However, due to the stochastic nature of detector response, for two identical jets t_1, t_2 , we might not have $\omega_1(t_1) = \omega_1(t_2)$. This is why the second step is needed to convert these “per-instance” [28] (where each instance is a detector-level and particle-level pair) weights to a function $\nu_1(t)$ that gives a unique prescription to any particle-level jet, since intuitively, $\nu_1(t)$ is obtained from training from a population of jets and outputs a value representative of the ensemble (as opposed to “per-instance”). We use the same neural network setup as step 1 and the same parameters of validation fraction being 20%, maximum number of epochs being 100, and patience being 50. Prediction is again done on both

- Iteration 1, Step 2:

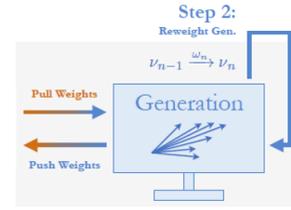
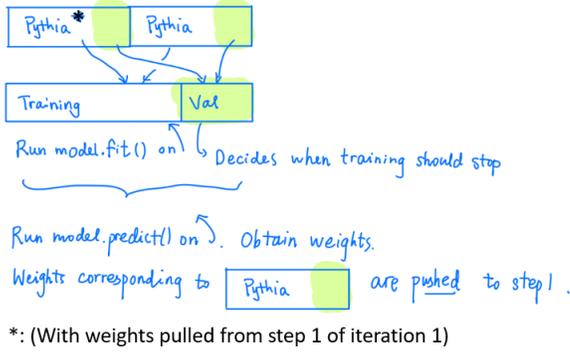


FIGURE 5. Iteration 1 Step 2

- Iteration n, Step 1:

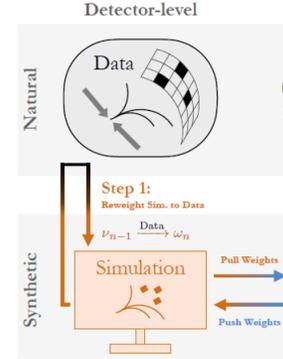
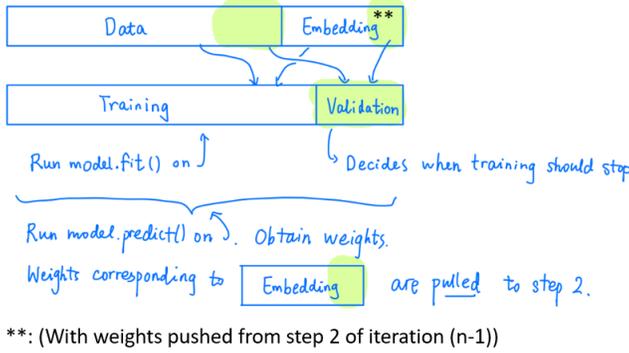


FIGURE 6. Iteration n Step 1

the training and validation samples, and weights $\nu_1(t)$ corresponding to the second set of particle-level jets are “pushed” to step 1 of next iteration.

$$(10) \quad \nu_1(t) = \nu_0(t) \frac{p(\omega_1, \text{PYTHIA})(t)}{p(\nu_0, \text{PYTHIA})(t)} \approx \nu_0(t) \frac{g(t)}{1 - g(t)}$$

where $g(x)$ is the model trained in step 2.

To clarify, training and validation are done with weights, prediction is done without weights, but the weights calculated from the prediction take into account the initial weights assigned at the beginning of the step.

For iteration $n > 1$, the procedure is similar. In step 1 (see Figure 6), instead of original cross section weights, the embedding jets get assigned weights “pushed” from step 2 of iteration $(n - 1)$, $\nu_{n-1}(m \in \text{emb}) = \nu_{n-1}(t)$. In step 2 (see Figure 7), the initial weights for the two sets of particle-level jets are weights “pulled” from step 1 of iteration n , $\omega_n(t) = \omega_n(m \in \text{emb})$, and weights “pushed” from step 2 of iteration $(n - 1)$, $\nu_{n-1}(t)$. Putting this together, we have:

$$(11) \quad \omega_n(m \in \text{emb}) = \nu_{n-1}(m \in \text{emb}) \frac{p(1, \text{data})(m)}{p(\nu_{n-1}, \text{emb})(m)}$$

$$\nu_n(t) = \nu_{n-1}(t) \frac{p(\omega_n, \text{PYTHIA})(t)}{p(\nu_{n-1}, \text{PYTHIA})(t)}$$

In step 2 of the last iteration N , instead of getting “pushed” to step 1 of the next iteration, the weights $\nu_N(t)$ are taken as the output of MultiFold, giving the unfolded jet population:

$$(12) \quad p(1, \text{unfolded})(t) = \nu_N(t) \cdot p(1, \text{PYTHIA})(t)$$

See Figure 8 for a summary of the correction procedure with MultiFold.

3.2. Closure test. We perform a closure test to verify the validity of the MultiFold method and investigate its dependence of the prior and number of iterations. The matched jet pairs (matched between PYTHIA and embedding) are separated into two statistically independent sets according to whether the event IDs assigned in the simulation is even or odd. These two sets are denoted “sample 1” and “sample 2”. Sample 2 is used to mimic data and truth jets. Sample 1 detector-level and particle-level and sample 2 detector-level jets are input to MultiFold. See Fig. 9 for a closure test procedure illustration.

Due to the random nature of machine learning methods’ weight initialization process, individual result of MultiFold is not reproducible, especially when run on GPUs. However, the average result over many random seeds is reproducible. The results

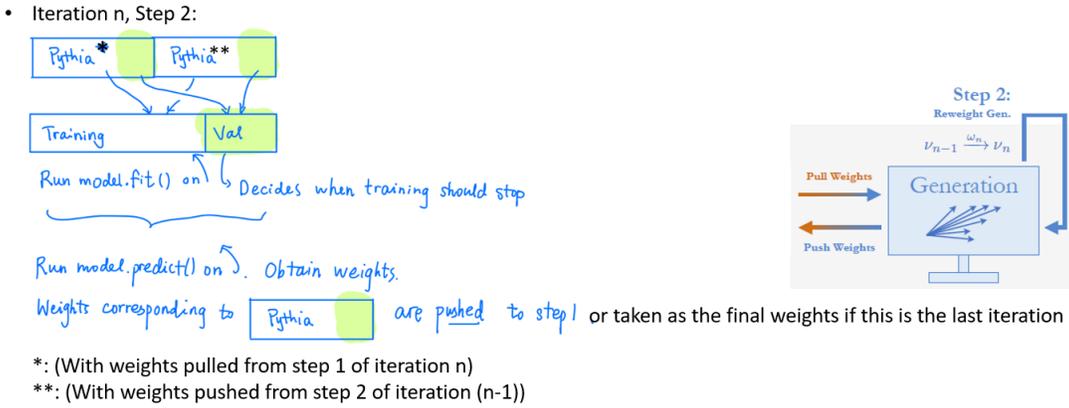


FIGURE 7. Iteration n Step 2

278 shown below are averages over 100 seeds. We show in Section 3.3 that the choice of 100 does not affect the unfolding result on
 279 data.

280 Figure 10 shows MultiFold achieves decent closure at 3 iterations, as can be seen by the agreement between MultiFolded
 281 spectra (red) and sample 2 particle-level spectra (green) for all observables. In addition, the significant difference for most
 282 observables between sample 2 detector-level spectra (grey shaded) and the generator (green) shows the need for unfolding, and
 283 the difference between the grey detector-level and red unfolded shows the effect of unfolding. Note that the binning is chosen
 284 after MultiFold to make histograms; the unfolding procedure itself is unbinned. Note also that the particle-level and unfolded
 285 substructure observable distributions shown include all jets with $p_T > 5$ GeV/c; additional cuts can be applied later if needed.
 286 Error bars indicate the statistical uncertainties. Specifically, the statistical uncertainties on the MultiFolded spectra are estimated
 287 based on sample 1 particle-level statistics for convenience. We treat the statistical uncertainties more carefully in Section 4 for
 288 unfolding on data.

289 Figure 11 shows MultiFold closure from 2 to 8 iterations. Ratios are taken after each distribution is normalized per jet.
 290 Statistical uncertainties (estimated from prior) are only shown in error bars for 4 iterations. The variation in the unfolded
 291 spectra due to the variation in number of iterations is small.

292 We check that the robustness against the variation in the number of iteration persists even when samples 1 and 2 are
 293 significantly different. We modify sample 1 weights by multiplying the original cross section weights for sample 1 jets (both
 294 generator- and detector-levels) by the ratios of HERWIG over PYTHIA p_T spectrum. We rerun the closure test and take the
 295 average of the results from 100 seeds. Figure 12 shows that the unfolded result becomes stable after 2 iterations as well.

296 Figure 13 shows MultiFold closure at 3 iterations, with statistical uncertainties on samples 1 and 2 both included, and
 297 unfolding systematic uncertainty on data included as well (See Section 5). This means that any non-closure will be covered by
 298 the unfolding systematics.

299 **3.3. Application on data and obtaining the fully corrected distributions.** Particle-level, detector-level and data jets
 300 are used for MultiFold on data. The initial weights for particle-level and detector-level jets are determined by the cross section
 301 weights from PYTHIA. The initial weights for data need to have fake rates taken care of properly. We use matched jets from
 302 particle-level and detector-level as the prior, so we need to effectively “remove” the fake jets from the data before unfolding
 303 with MultiFold. To do this, we put in $(1 - \text{fake rate})$ instead of 1, as weights for data. We estimate fake rates by taking the
 304 number of embedding jets that are not matched to particle-level jets (weighted by event cross sections) and dividing it by the
 305 total number of embedding jets (weighted by event cross sections). Figure 14 shows that fake rate has a weak dependence on
 306 any jet observable of our interest. We choose to assign a value of fake rate to each data jet based on its p_T .

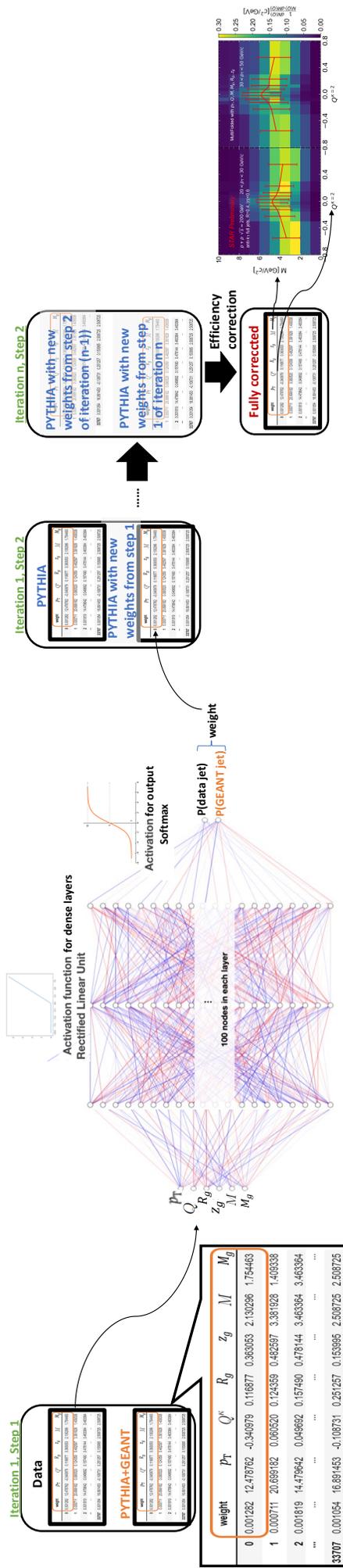
307 After running MultiFold, we first select a p_T range, and for a jet observable x , take the unfolded distribution of x averaged
 308 from 100 seeds at 3 iterations $y_{unf}(p_T, x)$. (Note: we experimented with different criteria to determine the number of iterations.
 309 3 iterations was chosen for Method 1 of Section 5.3. See discussion in 5.3 for more details.) To obtain the fully corrected jet
 310 observable distributions, we also need to include the contribution from missed jets, so we obtain

$$(13) \quad y_{full}(p_T, x) = \frac{y_{unf}(p_T, x)}{e(p_T, x)}$$

311 To obtain the per-jet normalized distribution of $\tilde{y}_{full}(p_T, x)$, where \tilde{y} indicates that the distribution is normalized, we use:

$$(14) \quad \tilde{y}_{full}(p_T, x) \doteq \frac{y_{full}(p_T, x)}{\int dx y_{full}(p_T, x)} = \frac{y_{unf}(p_T, x)}{e(p_T, x) \int dx \frac{y_{unf}(p_T, x)}{e(p_T, x)}}$$

312 Figure 15 shows the ratio of the normalized fully corrected jet mass distribution ($\tilde{y}_{full}(p_T \in (20, 25) \text{ GeV}/c, m)$) over normal-
 313 ized PYTHIA jet mass distribution for jets with $20 < p_T < 25$ GeV/c, with the MultiFolded distribution being an average of
 314 over 100 or 200 seeds. Error bars indicate the standard deviations from varying the seeds, which will be considered as a source



Closure test for unfolding

- Step 1: Separate matched jets from PYTHIA and PYTHIA+GEANT into 2 samples

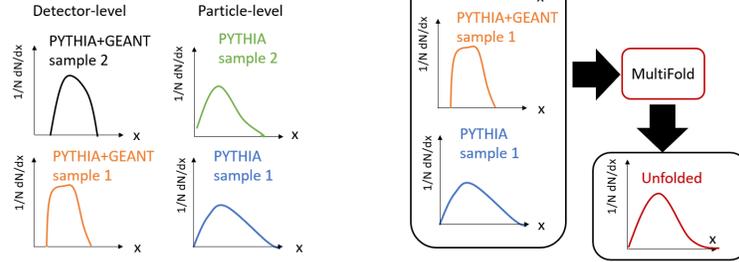


FIGURE 9. Unfolding closure test procedure illustration

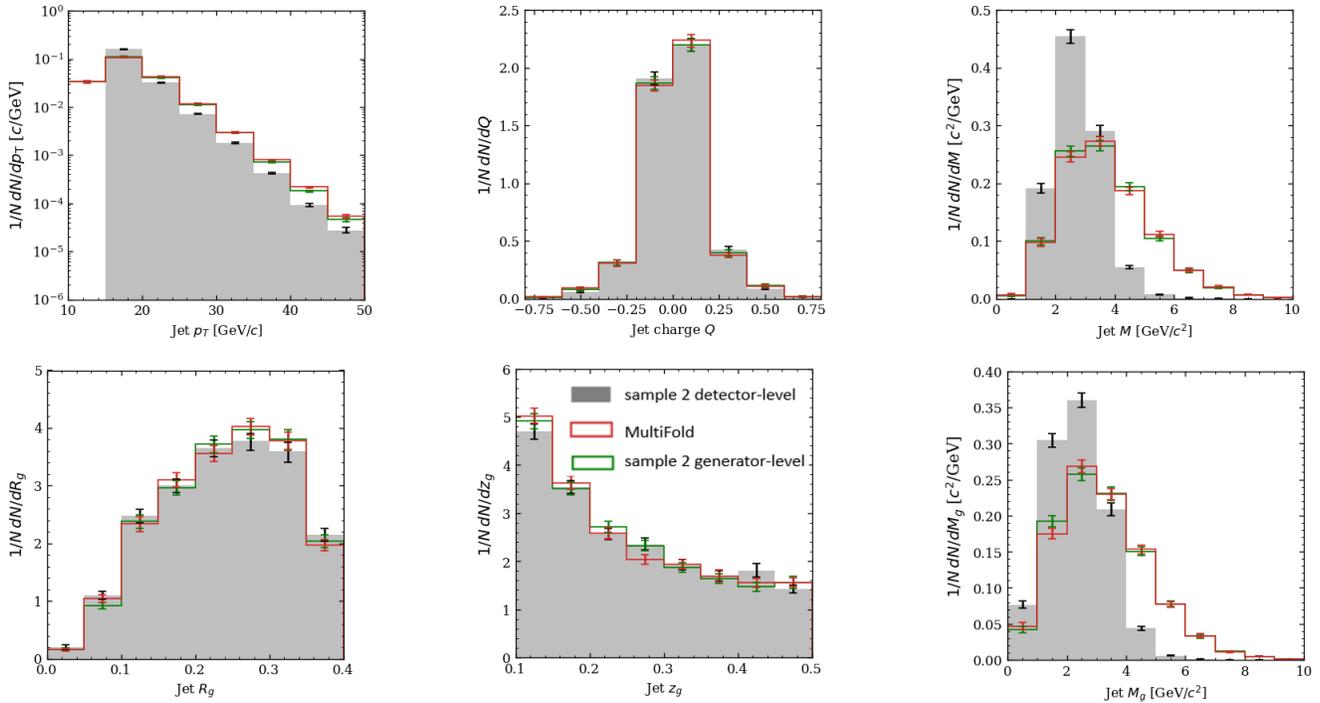


FIGURE 10. Sample 2 jet observable distributions at detector-level (grey) and particle-level (green) and MultiFolded distributions (red). Error bars indicate only statistical uncertainty.

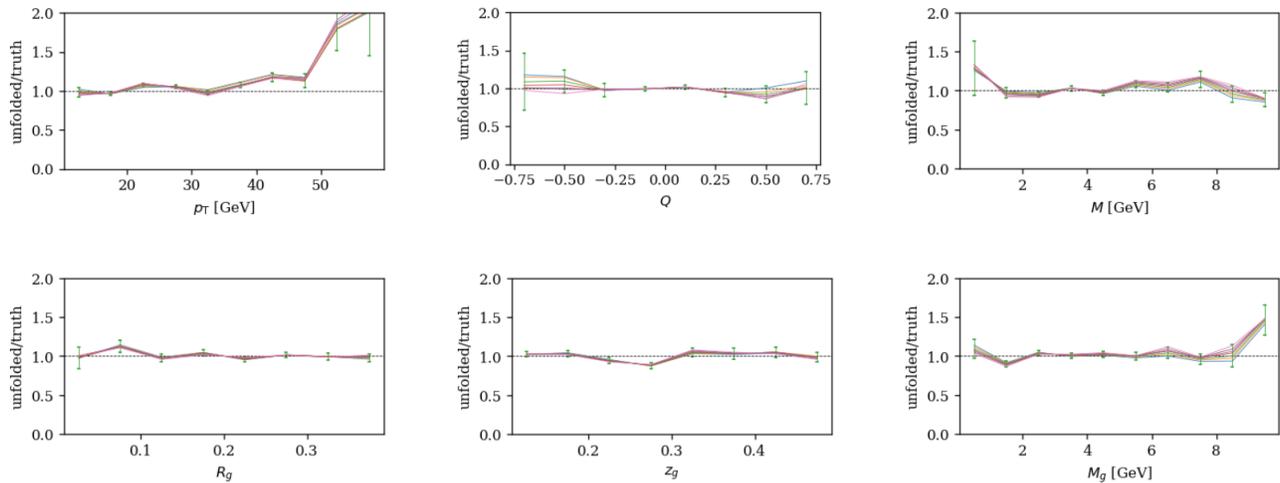


FIGURE 11. Ratios of MultiFolded normalized jet observable distributions over normalized truth distributions for iterations 2 to 8. Each color is for a different number of iteration parameter. Statistical uncertainty is shown for iteration 4.

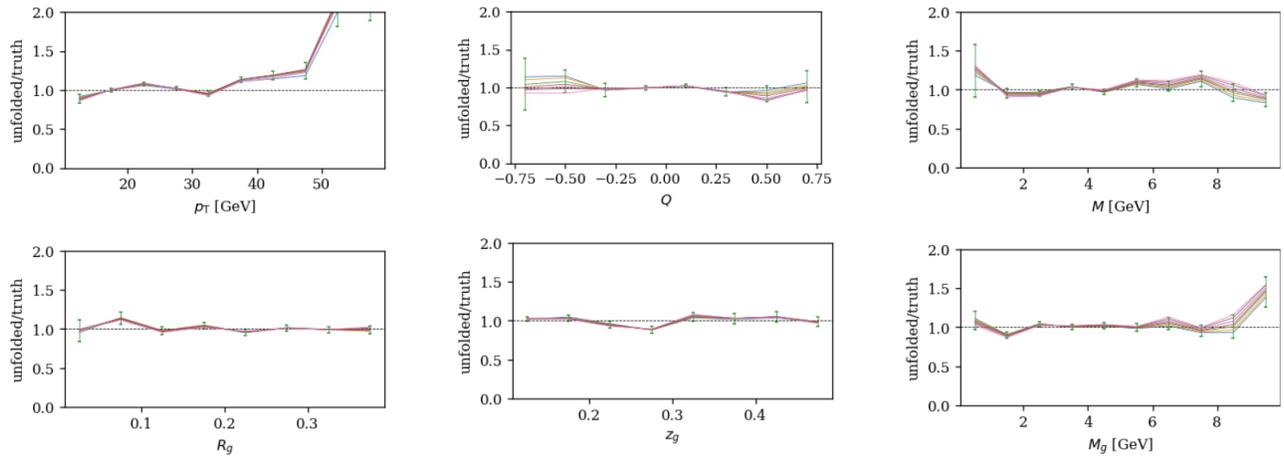


FIGURE 12. Same as Fig 11, except the weights of the prior jets are scaled by the ratio of HERWIG p_T spectrum over PYTHIA p_T spectrum.

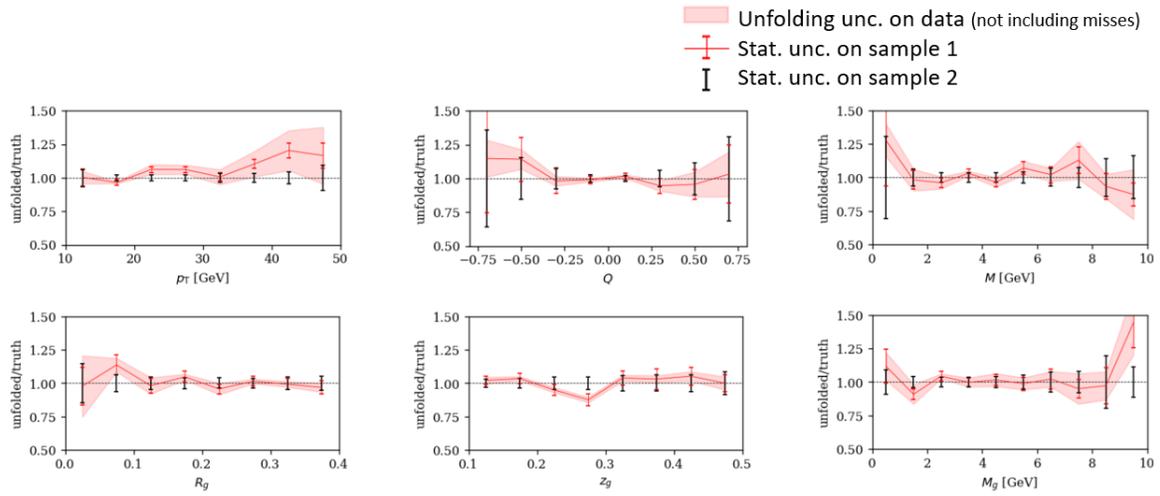


FIGURE 13. Ratios of MultiFolded normalized jet observable distributions over normalized truth distributions at 3 iterations averaged over 100 seeds. Red error bars indicate the statistical uncertainties on sample 1, black error bars indicate the statistical uncertainties on sample 2, and red error bands indicate the unfolding uncertainties on data.

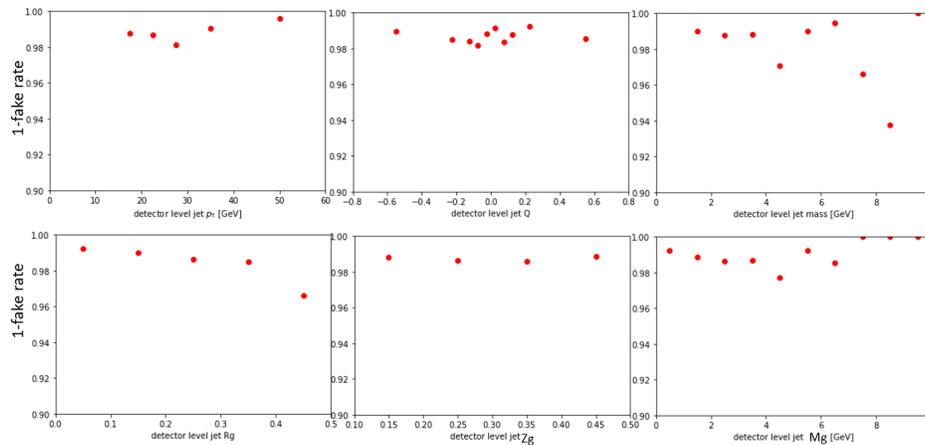


FIGURE 14. (1-fake rate) as a function of embedding jet observable values

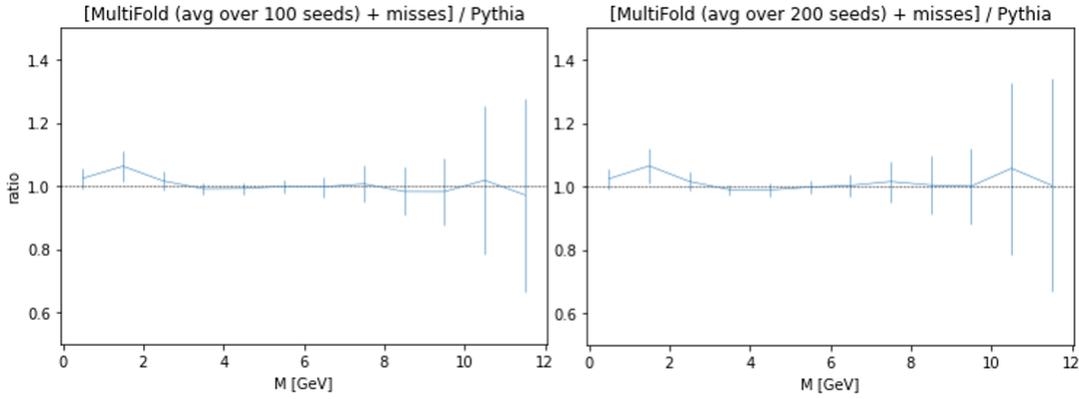


FIGURE 15. Ratios of fully corrected and normalized jet mass distributions over normalized PYTHIA distributions. Error bars indicate the standard deviation due to variation from seed to seed.

317

4. STATISTICAL UNCERTAINTIES

318 Statistical uncertainty is estimated with a bootstrap method. In the nominal case, each jet in the raw data has an initial
 319 weight of $(1 - \text{fake rate})$. For each bootstrap trial, we vary the initial weight of each jet to $(1 - \text{fake rate}) \cdot \text{Pois}(1)$, where $\text{Pois}(1)$
 320 denotes a Poisson distribution with a mean at 1. This method is analogous to the “kCovToy2” option in RooUnfold. We run
 321 100 bootstrap trials, obtain the unfolded distribution of a jet observable x for each trial, and take the standard deviation of the
 322 100 distributions as $\sigma_{unf,stat}(p_T, x)$. Then the statistical uncertainty that we report is

$$(15) \quad \tilde{\sigma}_{full,stat}(p_T, x) = \frac{\sigma_{full,stat}(p_T, x)}{\int dxy_{full}(p_T, x)} = \frac{\sigma_{unf,stat}(p_T, x)}{e(p_T, x) \int dxy_{full}(p_T, x)}$$

323 where $\tilde{\sigma}$ indicates that the uncertainty is associated with a normalized distribution.

324 Since $\sigma_{unf,stat}(p_T, x)$ has contribution due to limited statistics in both raw data and in the unfolding procedure, but the
 325 statistical error from the misses does not have contribution from the unfolding procedure, scaling $\sigma_{unf,stat}(p_T, x)$ by efficiency
 326 as in the previous equation might have resulted in a slight overestimation of the statistical uncertainty.

327

5. SYSTEMATIC UNCERTAINTIES

328 We list the sources of systematic uncertainties for this analysis:

- 329 • Hadronic correction: Nominally, 100% of track energy from matched towers is subtracted. We vary this fraction to 50%.
 330 See details in [30].
- 331 • Tower scale: We increase this by 3.8%. See details in [31].
- 332 • Tracking efficiency: We decrease this by 4%. See details in [31].
- 333 • Unfolding procedure, treated as correlated uncertainties:

$$(16) \quad \tilde{\sigma}_{unf}(p_T, x) = \max(\tilde{\sigma}_{Herwig}(p_T, x), \tilde{\sigma}_{Pythias}(p_T, x), \tilde{\sigma}_{seed}(p_T, x), \tilde{\sigma}_{it}(p_T, x))$$

- 334 – Variation of the random seed.
- 335 – Variation of the number of iterations.
- 336 – Variation of the prior.

337 **5.1. Variation of detector effects.** For **hadronic correction, tower scale and tracking efficiency uncertainty**, the
 338 variation is done in the embedding (but not in data, following the convention in previous analyses [11] and [14]), which affects
 339 the prior for unfolding, PYTHIA misses/matches distribution and efficiency. The analysis procedure is carried out with these
 340 variations (including averaging over 100 random seeds, etc). The fully corrected spectrum is compared with the nominal full
 341 spectrum, and the difference is taken as one side of the systematic uncertainty.

342 **5.2. Variation of the random seed.** We run MultiFold with 100 different random seeds that affect the initialization of the
 343 weights of the neural networks and the train-test split, obtain the unfolded distribution of a jet observable x for each seed, take
 344 the error on the mean over the 100 distributions (i.e., the standard error), and denote it as $\sigma_{seed}(p_T, x)$.

345 We choose to characterize this uncertainty with the error on the mean instead of standard deviation, because increasing the
 346 number of measurements by sampling with different seeds should decrease the randomness of the measurement.

347 The uncertainty associated with the normalized distribution after misses are taken into account is

$$(17) \quad \tilde{\sigma}_{seed}(p_T, x) = \frac{\sigma_{seed}(p_T, x)}{\int dxy_{full}(p_T, x)}$$

348 **5.3. Variation of the number of iterations.** Two different treatments of this source of systematic uncertainty were applied
 349 to different results.

5.3.1. *Method 1.* We take the maximum deviation from 2-8 iterations with the nominal (3 iterations),

$\sigma_{it}(p_T, x) = \max_{i \in [2, 8]} (|y_{unf, i}(p_T, x) - y_{unf}(p_T, x)|)$, then the uncertainty associated with the normalized distribution is

$$(18) \quad \tilde{\sigma}_{it}(p_T, x) = \frac{\sigma_{it}(p_T, x)}{\int dx y_{full}(p_T, x)}$$

5.3.2. *Method 2.* We do not explicitly account for the variation of the number of iterations as a separate source of uncertainty. Going to a higher number of iterations reduces the prior dependence bias. However, the limited statistics (in both the prior sample and data) introduces unwanted fluctuations at high number of iterations. The limited statistics in embedding can manifest through a large uncertainty from the variation of initial seeds, and the limited statistics in data can manifest through a large statistical fluctuation from variation of the statistical bootstrap trials. Therefore, the number of iterations can be selected by considering when a) the prior dependence uncertainty, b) seed uncertainty, and c) statistical uncertainty are low. When unfolding with the 2021 embedding, we select an iteration number of 12, low enough such that the uncertainty due to seed variation and statistical uncertainty are both reasonable, at the cost of a non-negligible prior dependence uncertainty.

Mathematically, the most correct number of iterations is infinity. The deviation of the result due to not able to perform an infinite number of iterations shows up as the prior dependence. Therefore, the prior variation uncertainty effectively accounts for the uncertainty due to the number of iterations not being ideal.

5.4. **Prior variation.** Prior variation uncertainty is the dominant source for both [14] and this analysis. As we try to come up with the most appropriate approach for accounting for this uncertainty in MultiFold, we have tried several different approaches as detailed below.

5.4.1. *Method 0. Reweight prior to HERWIG or PYTHIA8 weights in 2D, and then unfold data with it.* This is the same method as used in previous analyses such as [14]. Uncertainties in Sections 6.1 and 6.2 are accounted for with this method. We generate HERWIG7 (LHC-UE-EE-4-CTEQ6L1 tune) [32] and PYTHIA8 (Detroit tune) [26] events and apply the same particle and jet selections. Then we select on $20 < p_T < 30$ GeV/ c and $30 < p_T < 50$ GeV/ c jets, take the ratio of jet mass distribution (at truth level) of HERWIG7 (PYTHIA8) over PYTHIA6, for the two p_T bins respectively, and then reweight the matched PYTHIA6 embedding and generator jets by the mass ratios for unfolding.

5.4.2. *Method 1. Reweight prior to HERWIG or PYTHIA8 weights in 6D, and then unfold data with it.* For this method and the ones following, HERWIG7 events are generated with the default 7.2 tune [27]. Even though only a in of the observables that we unfolded for is reported for any given measurement, all 6 are used for unfolding, so prior variation should take into account of all of them.

Binning chosen: 10 GeV for p_T , $[-0.8, -0.15, -0.1, -0.05, 0, 0.05, 0.1, 0.15, 0.8]$ for Q , 1 GeV for M and M_g , and 0.1 for R_g and z_g .

However, the binning for the reweighting is significantly restricted by the statistics of our PYTHIA6 sample and its embedding. In particular, this binning might not capture the difference in shape of a between different MC models, especially in the small but nonzero a region, so the apparent small prior variation uncertainty in a measurement could be attributed to underestimation of prior variation itself. This leads us to move to Method 2.

5.4.3. *Method 2. Reweight prior to HERWIG or PYTHIA8 weights in 7D, and then unfold data with it.* We additionally include a -reweighting with binning of $[-0.01, 0.000001, 0.005, 0.01, 0.015, 0.02, 0.03, 0.04, 1]$. The first bin contains only jets with $a = 0$. We also relax the p_T to two bins only: $20 < p_T < 30$ GeV/ c and $p_T < 30$ GeV/ c . p_T is not included as a dimension to reweight, since we do not aim to correct the p_T spectrum per se, but rather aim to study the dependence of the substructure observables on p_T . The total number of bins for the reweighting is 432640. This shows again that having a high-statistics embedding sample is important for this analysis.

Note that all the methods above rely on reweighting the prior in a binned manner. Although this might seem a sacrifice of MultiFold's full potential of correcting data in a completely unbinned way, we benefit from not having to run the full embedding procedure on two other other truth-level samples for HERWIG and PYTHIA8.

6. RESULTS

In this section, we present results obtained using the methods described above. **Sections 6.1, 6.4 and 6.6 are most relevant to the paper, while Sections 6.2, 6.3 and 6.5 contain discussions of preliminary results that are not shown in the paper.**

6.1. Jet mass. Figure 16 shows the fully corrected normalized jet mass distributions for $20 < p_T < 25$ GeV/c, $25 < p_T < 30$ GeV/c and $30 < p_T < 40$ GeV/c. Red markers are data fully corrected with MultiFold while black markers are with RooUnfold [14]. The error bands in the top panels include both statistical and systematic uncertainties. The bottom panels show the ratio of the MultiFolded distributions over RooUnfolded distributions, where the error bands include only the uncorrelated systematic uncertainties, calculated by taking the quadrature of the unfolding systematics from MultiFold and RooUnfold. For the systematic uncertainties for the MultiFolded result, Method 1 of 5.3 and Method 0 of 5.4 are applied, since they are most similar to how uncertainties are calculated in the RooUnfolded result. The ratios are consistent with unity within the uncertainty, which shows that the MultiFolded result agrees with the RooUnfolded result as expected and serves as another proof that MultiFold works.

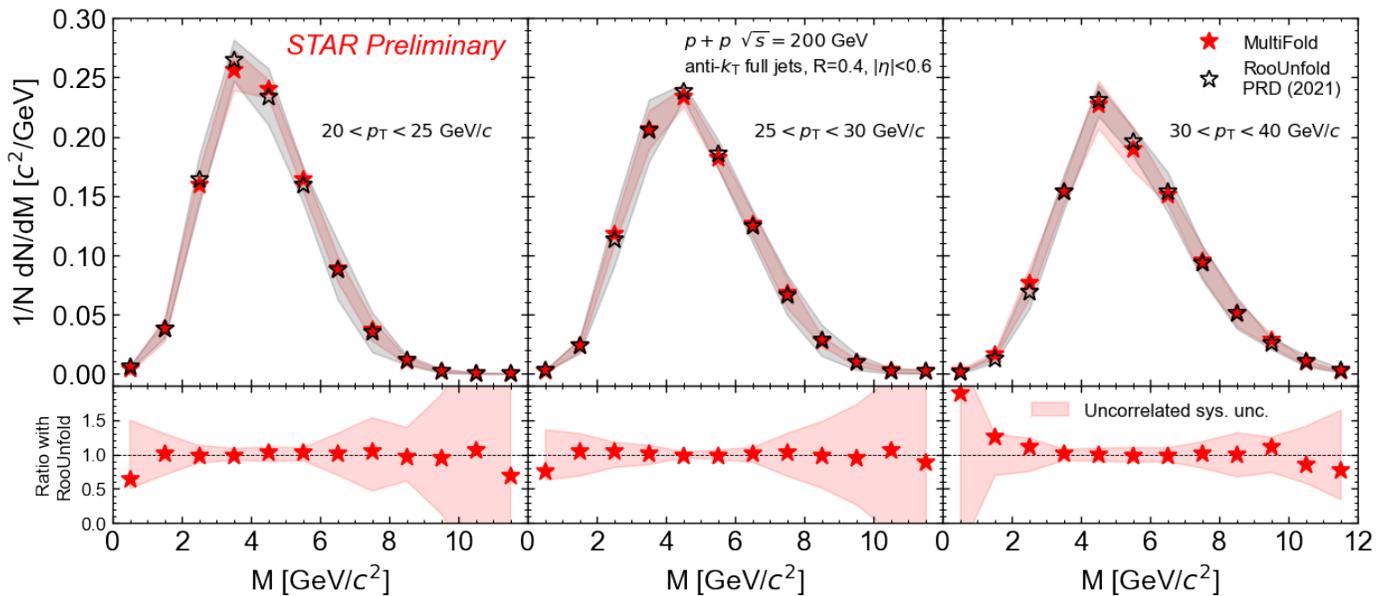


FIGURE 16. (Top) Jet mass distributions unfolded with MultiFold (red), compared with previously published STAR measurement unfolded with RooUnfold (black), for three different jet p_T ranges. The error bands include both statistical and systematic uncertainties. (Bottom) MultiFolded distributions dividing by RooUnfolded distributions. The error bands include only the uncorrelated systematic uncertainties.

Figure 17 shows the uncertainty breakdown of the fully corrected jet mass for $20 < p_T < 25$ GeV/c, $25 < p_T < 30$ GeV/c and $30 < p_T < 40$ GeV/c.

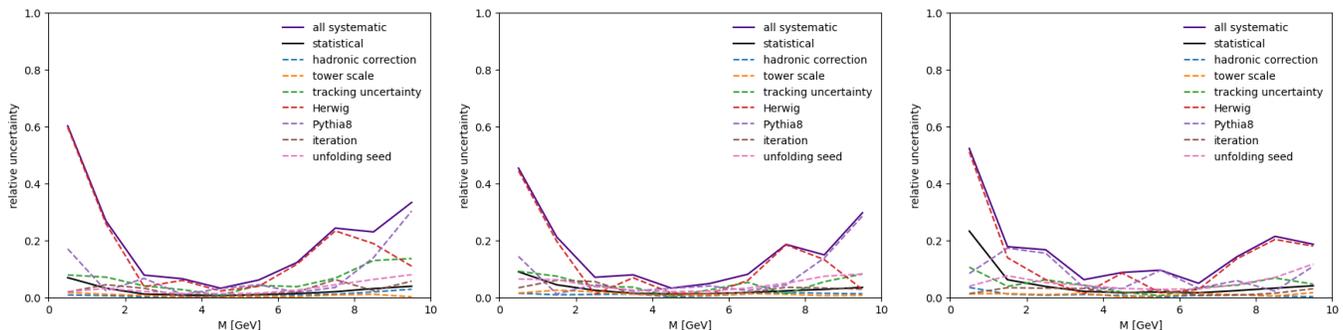


FIGURE 17. Uncertainty breakdown of the fully corrected jet mass for $20 < p_T < 25$ GeV/c, $25 < p_T < 30$ GeV/c and $30 < p_T < 40$ GeV/c, from the left to the right

6.2. **Jet charge vs mass.** As promised earlier, MultiFold naturally keeps the correlation between the jet observables from the raw data level. Figure 18 shows the correlation between jet mass and charge $Q^{\kappa=2}$ for $20 < p_T < 30$ GeV/ c and $30 < p_T < 50$ GeV/ c . The normalization of the 2D histograms is done per column, namely, the color of each bin indicates the number of jets that fall into that bin dividing by the total number of jets that fall into the entire column. The red error bars are widths of jet mass distributions for each charge bin. Again, for the systematic uncertainties for the MultiFolded result, Method 1 of 5.3 and Method 0 of 5.4 are applied. We observe that jet M increases with decreasing jet $|Q|$, and a few factors can contribute to this trend.

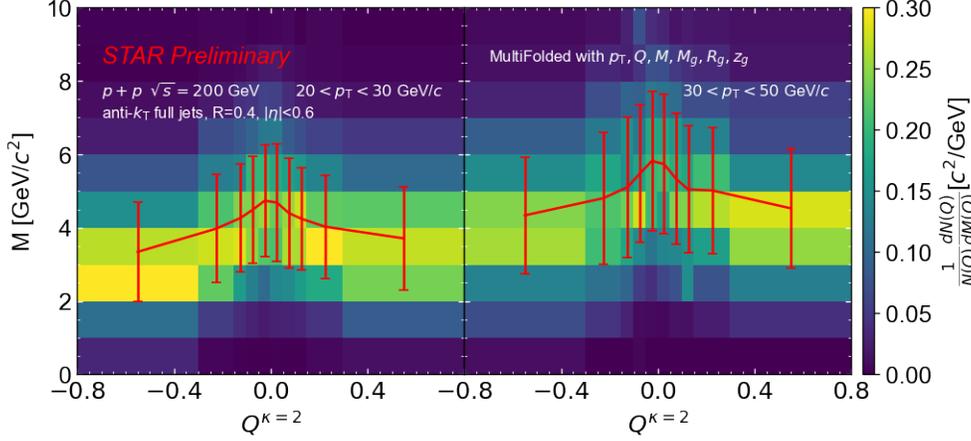


FIGURE 18. Fully corrected jet mass vs charge correlation for two different jet p_T ranges. The normalization of the 2D histograms is done per column.

- Since quarks are electrically charged while gluons are not, lower $|Q|$ jets are slightly more likely to be gluon-dominated and thus have on average a larger M . However, in the particle-level PYTHIA jet sample, we are also able to see this trend (although weaker) in pure quark jets, and that a cut on Q does not change the quark-gluon fraction of the jet sample by much. We discuss our findings with MC studies in 6.2.1.
- The peak in jet Q distribution around 0 could be related to local charge conservation during the hadronization process. This can be further explored in other observables, such as the charge correlator ratio r_c .
- There is correlation between M and Q due to their common dependence on jet and constituent p_T by definition.

Both M and Q are closely related to the generalized angularity, a class of jet substructure observables defined as

$$(19) \quad \lambda_{\beta}^{\kappa} = \frac{1}{(p_{T\text{jet}})^{\kappa}} \sum_{i \in \text{jet}} \left(\frac{\Delta R_i}{R} \right)^{\beta} \cdot (p_{T_i})^{\kappa}$$

Specifically, λ_2^1 is closely related to the jet mass M .

$$(20) \quad M = p_{T\text{jet}} \sqrt{\lambda_2^1} = \sqrt{p_{T\text{jet}} \sum_{i \in \text{jet}} \left(\frac{\Delta R_i}{R} \right)^2 \cdot p_{T_i}}$$

and λ_1^2 has a similar form as Q .

$$(21) \quad Q^{\kappa=2} = \frac{1}{p_{T\text{jet}}^2} \sum_{i \in \text{jet}} q_i \cdot p_{T_i}^2$$

Figure 19 shows the projection of the 2D M vs Q correlation plot onto the M axis for $|Q| < 0.15$ (green) and $|Q| > 0.15$ (blue), compared with descriptions from event generators, PYTHIA8 (Detroit tune, shown in solid lines) and HERWIG7 (LHC-UE-EE-4-CTEQ6L1 tune, shown in dashed lines). The error bands on data include both systematic and statistical uncertainties. The error bars on the generator curves include only statistical uncertainties and can be reduced in the future. We observe the mean and spread of the jet mass distributions both increase with decreasing jet $|Q|$, and that our data is well-described by PYTHIA8 while the M distribution for large $|Q|$ is significantly under-predicted by HERWIG7.

Figure 20 shows the uncertainty breakdown of the fully corrected jet mass vs charge for $20 < p_T < 30$ GeV/ c and $30 < p_T < 50$ GeV/ c .

6.2.1. *MC studies.* Our very initial MC study of the M vs Q correlation is from PYTHIA8 pp 500 GeV, anti- k_T $R = 0.4$ charged jets, with $p_T > 20$ GeV/ c , shown in Figure 21. The normalization is again done per column, as in Fig. 18. Although it has been shown that jet charge observables with different κ -values are strongly correlated with one another [paper that we read for Alba's journal club], we observe that $\kappa = 2$ gives us an enhanced correlation between M and Q . We investigate this correlation further to see if this can facilitate separation of quark vs gluon jets.

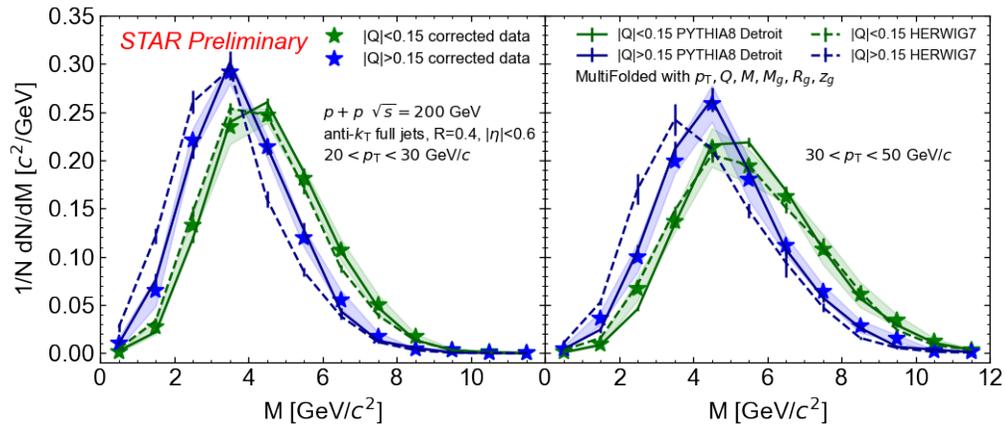


FIGURE 19. Projection of the 2D M vs Q correlation plot onto the M axis for $|Q| < 0.15$ (green) and $|Q| > 0.15$ (blue), compared with descriptions from event generators, PYTHIA8 (Detroit tune, shown in solid lines) and HERWIG7 (LHC-UE-EE-4-CTEQ6L1 tune, shown in dashed lines)

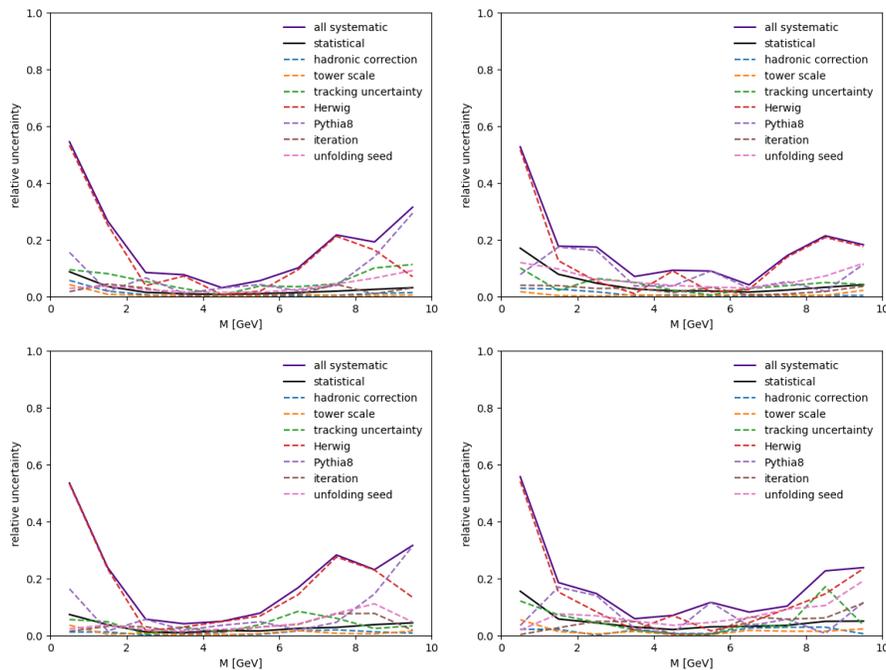


FIGURE 20. Uncertainty breakdown of the fully corrected jet mass for $20 < p_T < 30$ GeV/ c and $30 < p_T < 50$ GeV/ c , from the left to the right. Top row is for $|Q| < 0.15$ and bottom row is for $|Q| > 0.15$.

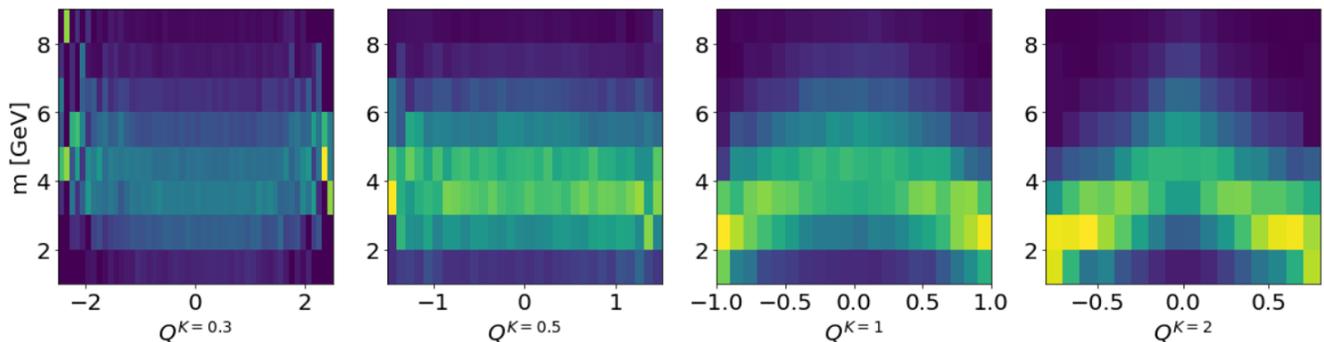


FIGURE 21. Jet mass vs charge correlation for various κ values, from PYTHIA8 pp 500 GeV, anti- k_T $R = 0.4$ charged jets, with $p_T > 20$ GeV/ c . The normalization of the 2D histograms is done per column.

- obtain a list of outgoing partons
- obtain a list of jets, sorted from the highest p_T
- for each outgoing parton, find the ΔR between the parton and a jet from the jet list
 - If $\Delta R < 0.4$, declare that jet to come from that outgoing parton, and remove the jet from the jet list. Move onto the next parton in the outgoing parton list
 - If $\Delta R > 0.4$, move onto the next jet in the jet list

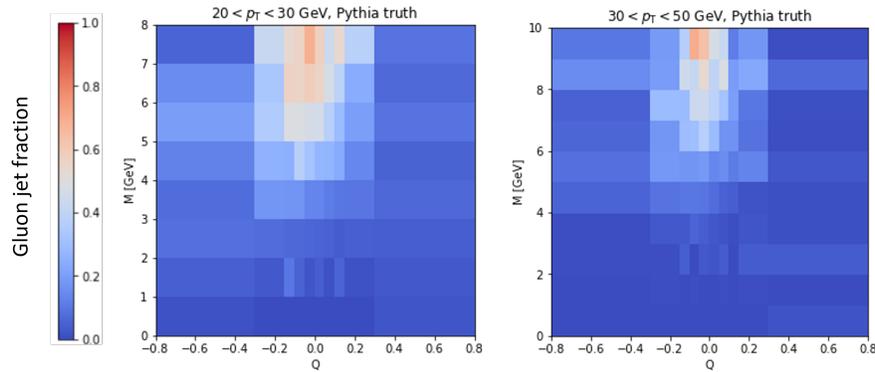


FIGURE 22. Gluon jet fraction for different selections of jet M , Q , and p_T , from PYTHIA8 pp 200 GeV, anti- k_T $R = 0.4$ full jets.

From the simulation, we observe that the quark fraction for jets in $20 < p_T < 30$ GeV/ c is about 65%. To select a jet population with a quark fraction of 91%, we can apply a cut of $M < 4$ GeV/ c^2 , and still have a lot of statistics. However, this is more challenging with selecting a population with a large gluon fraction, and a simultaneous cut on M and Q shows a better prospect than a cut on either variable.

To obtain a population with gluon fraction of about 67% for jets in $20 < p_T < 30$ GeV/ c , we can:

- select on $M > 8.6$ GeV/ c^2 , and arrive at 0.8% of the statistics;
- or select on $M > 7$ GeV/ c^2 and $-0.025 < Q < 0$, and arrive at 1.1% of the statistics.

To obtain a population with gluon fraction of about 65% for jets in $30 < p_T < 50$ GeV/ c , we can:

- select on $M > 9$ GeV/ c^2 and $-0.08 < Q < -0.01$, and also arrive at 1.1% of the statistics.
- No cut on M or Q alone can achieve such purity.

Similar exercise is done on pp 500 GeV simulations and we find that to obtain a population with gluon fraction of about 72% for jets in $20 < p_T < 30$ GeV/ c , we can:

- select on $M > 8$ GeV/ c^2 , and arrive at 4% of the statistics;
- or select on $M > 7$ GeV/ c^2 and $|Q| < 0.05$, and arrive at 8% of the statistics.

These studies show the potential of achieving a relatively high gluon purity with a simultaneous selection of M and Q . However, as these selections inherently introduce fragmentation bias, they cannot be used to study the differences between quark and gluon fragmentation in an unbiased manner.

To investigate if the trend we observe in data in Fig. 18 is related to the flavor of the jet, we study the correlation with quark jets in PYTHIA, as shown in Figure 23. Since we are also able to observe the same trend that jet M increases with decreasing jet $|Q|$, although weaker, in pure quark jets, this pattern cannot be entirely attributed to the quark-gluon ratio argument.

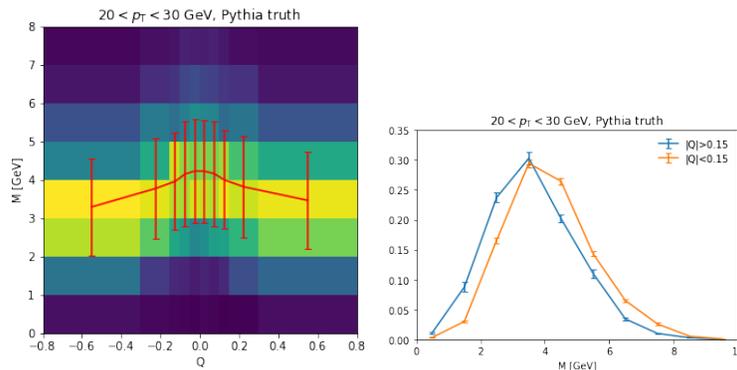


FIGURE 23. (Left) jet mass vs charge correlation for pure quark jets in PYTHIA. (Right) projection of the 2D M vs Q correlation plot onto the M axis for $|Q| < 0.15$ and $|Q| > 0.15$ for pure quark jets in PYTHIA.

465 **6.3. Correlations between each pair of substructure observables.** In addition to the correlation shown in Section 6.2,
 466 Figure 24 shows the fully corrected correlations between each pair of unfolded substructure observables for jets with $20 < p_T < 30$
 467 GeV/ c . Similar to Fig. 18, the normalization of each 2D histogram is done per column, and the red curves indicate the mean
 468 values with the error bars indicating the spread of the distributions. Perhaps some of the correlations shown here would be
 469 interesting to explore in a future study.

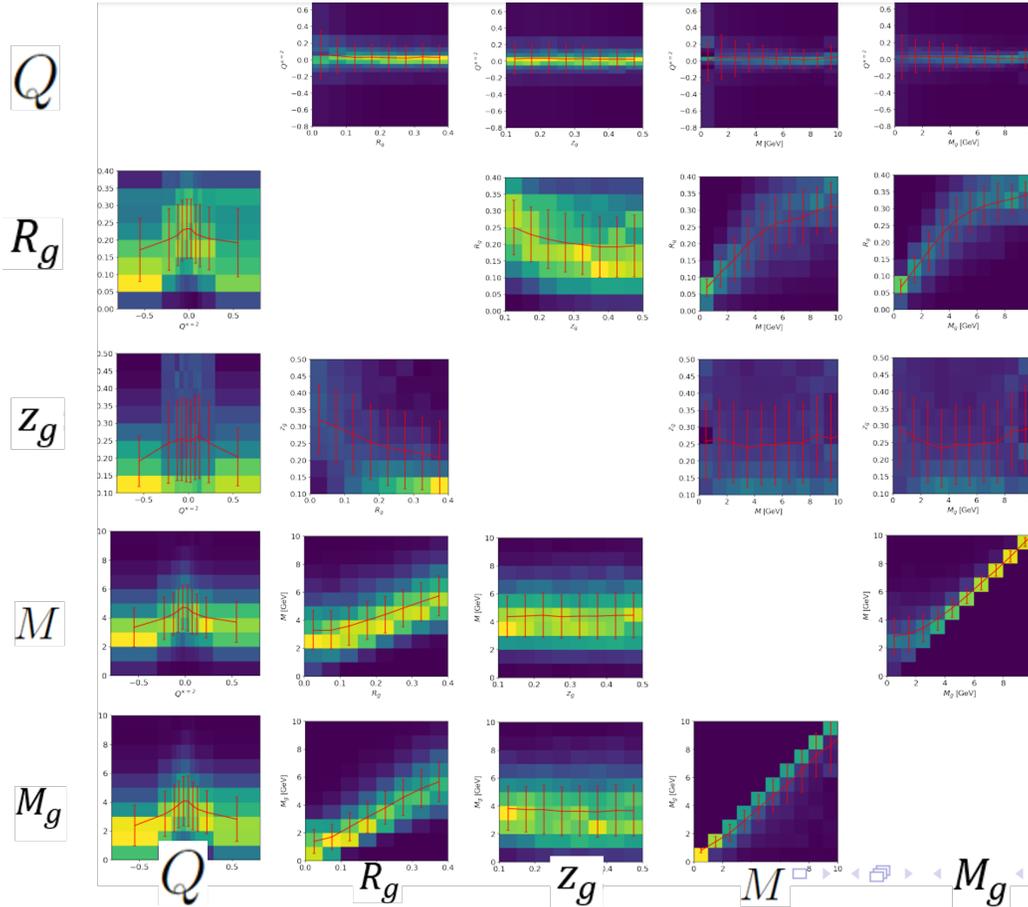


FIGURE 24. Fully corrected correlations between each pair of unfolded substructure observables for jets with $20 < p_T < 30$ GeV/ c . The normalization of each 2D histogram is done per column.

470 **6.4. CollinearDrop jet mass.** Figure 25 shows the distribution of fully corrected CollinearDrop groomed jet masses for jets
 471 within $20 < p_T < 30$ GeV/ c . This measurement excludes jets with $M = M_g$ (46% of jets in this p_T range) so that the peak in
 472 the small but nonzero ΔM region is visible. The $M = M_g$ case corresponds to the jets whose first splittings pass the criterion
 473 of $(z_{\text{cut}}, \beta) = (0.1, 0)$ without the need of SoftDrop grooming, because the lower- p_T prong of the splitting carries at least 10%
 474 of the total jet p_T . Both PYTHIA8 Detroit tune [26] and HERWIG7 7.2-Default tune [27] capture the qualitative trend of data,
 475 although there is some tension with HERWIG in the small ΔM region.

476 Figure 26 shows the breakdown of systematic uncertainties of the measurement. Method 1 of 5.4 is used for estimation of the
 477 uncertainty.

478 Figure 27 shows a different representation of the fully corrected CollinearDrop groomed jet mass for jets with $20 < p_T < 30$
 479 GeV/ c and $30 < p_T < 50$ GeV/ c . This measurement is corrected with the 2021 embedding. Systematic uncertainties are treated
 480 with Methods 2 of Sections 5.3 and 5.4. (See Figure 28 for the breakdown of these uncertainties for jets with $20 < p_T < 30$ GeV/ c .)
 481 Analytic calculation with NLL SCET performed at the parton level shows deviation from both event generator predictions and
 482 data, indicating that the CollinearDrop groomed mass is sensitive to hadronization effects. This is verified by the comparison
 483 in Figure 29, where better agreement is seen between the analytic calculation with PYTHIA with hadronization off. The error
 484 band on the theoretical curve indicates typical scale variations in theoretical calculation.

485 Figure 30 is the same as Fig. 27 except that the data are unfolded with the 2015 embedding. The uncertainties are larger
 486 compared to the results in Sections 6.1 and 6.2 because of different treatment of systematics. (Methods 2 of Sections 5.3 and 5.4
 487 are used in this case). They are also higher compared to Fig. 27 because the 2015 embedding has much lower statistics for jets.
 488 See Figure 28 for the breakdown of these uncertainties for jets with $20 < p_T < 30$ GeV/ c .

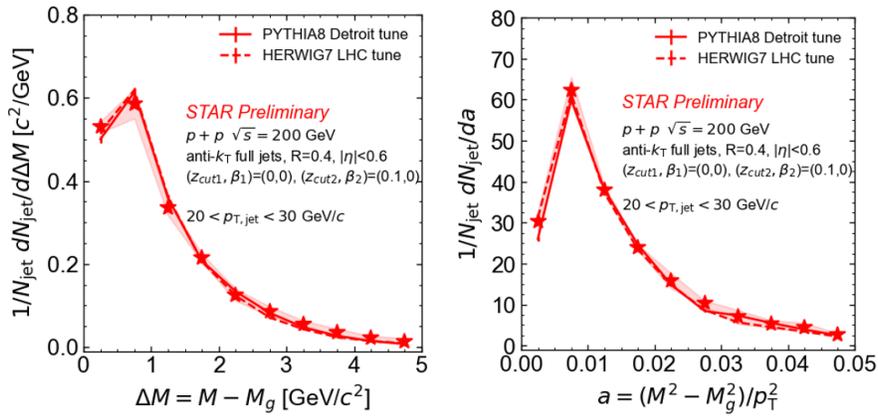
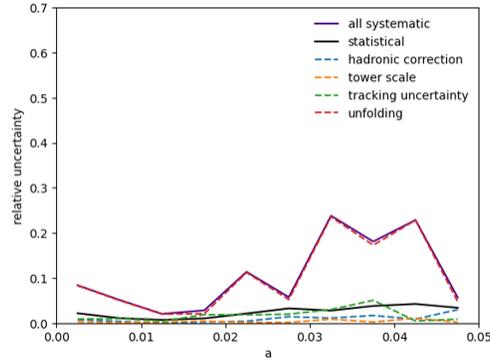
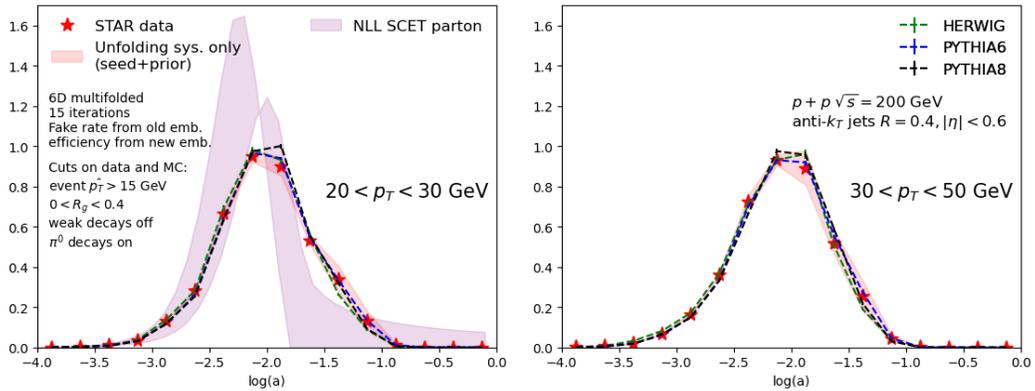
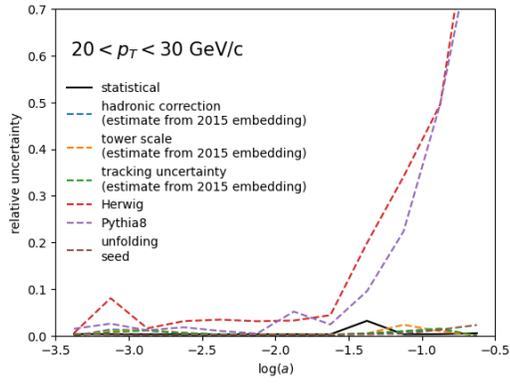
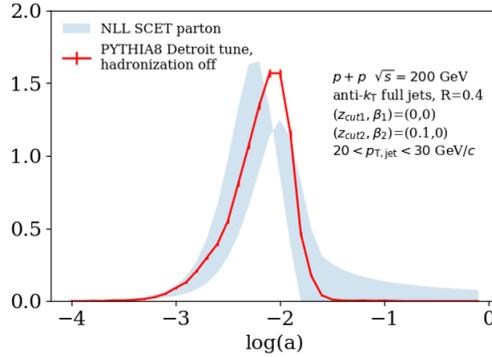
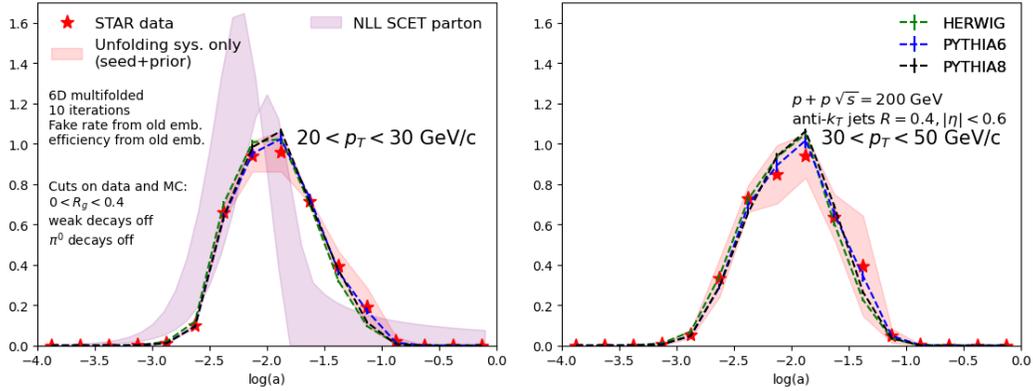
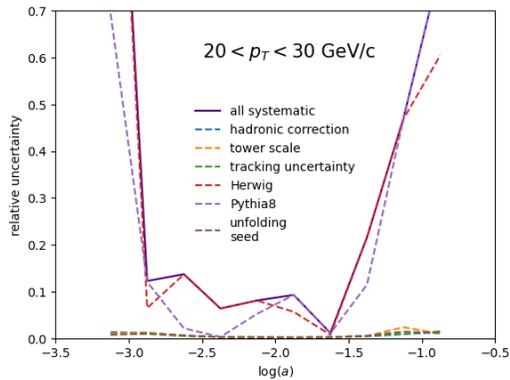


FIGURE 25. CollinearDrop jet measurement, compared with PYTHIA and HERWIG predictions.

FIGURE 26. Relative uncertainty for the measurement of a .FIGURE 27. CollinearDrop jet $\log(a)$ measurement, corrected with 2021 embedding, compared with PYTHIA and HERWIG predictions and SCET calculation

FIGURE 28. Uncertainty estimate for the CollinearDrop jet $\log(a)$ measurement, corrected with 2021 embeddingFIGURE 29. CollinearDrop jet $\log(a)$ at the parton level.FIGURE 30. CollinearDrop jet $\log(a)$ measurement, corrected with 2015 embedding, compared with PYTHIA and HERWIG predictions and SCET calculation.FIGURE 31. Uncertainty estimate for the CollinearDrop jet $\log(a)$ measurement, corrected with 2015 embedding.

489 6.4.1. *MC studies.* Figure 32 shows the a distribution for PYTHIA8 quark jets and gluon jets separately, compared with the
 490 measurement in Fig. 25. We observe that gluon jets have a higher value of a on average, which indicates that they tend to have
 491 more nonperturbative soft and wide-angle radiation at the early stage of the parton shower. The quark jet distribution is closer
 492 to our inclusive measurement, which can be explained by the high quark jet fraction for this kinematic selection of jets.

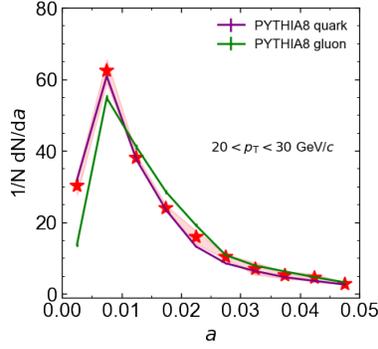


FIGURE 32. CollinearDrop a distribution for quark vs gluon jets in PYTHIA (without π^0 decays), compared with the fully corrected measurement

493 Figure 33 shows the $\log(a)$ distributions for different z_{cut} selections, from HERWIG simulations. This shows that the
 494 CollinearDrop groomed mass can be influenced by the grooming criteria used.

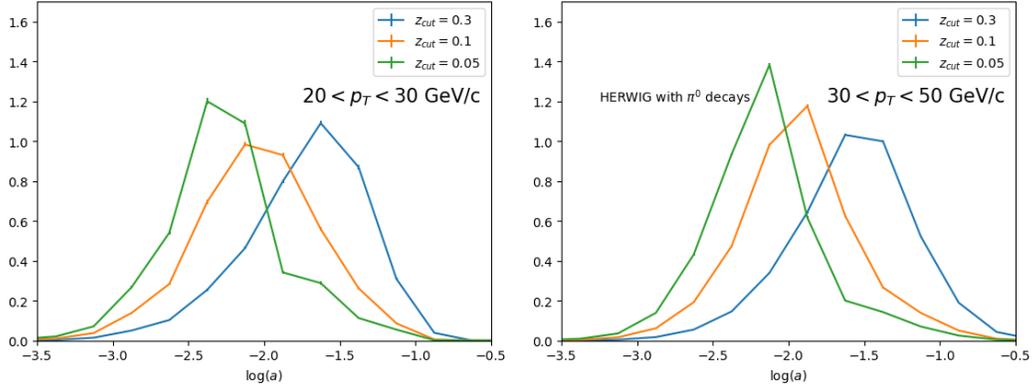


FIGURE 33. $\log(a)$ distributions for z_{cut} of 0.3, 0.1 and 0.05, from PYTHIA8 simulations.

495 Figure 34 shows the $\log(a)$ distributions for two p_T ranges with different \sqrt{s} for collision energies, simulated with PYTHIA8.
 496 We see that $\log(a)$ has a little dependence on \sqrt{s} .

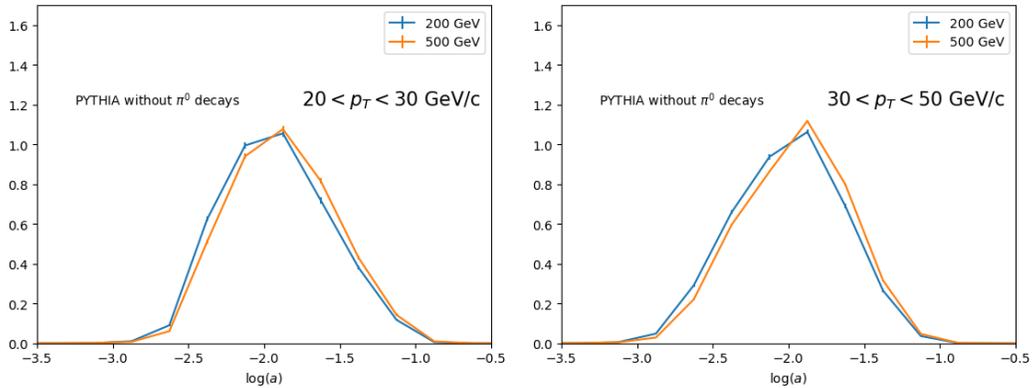


FIGURE 34. $\log(a)$ distributions for collision energies of 200 and 500 GeV, from PYTHIA8 simulations.

497 **6.5. CollinearDrop jet mass vs R_g or z_g .** Figure 35 (left) shows the correlation between the CollinearDrop groomed mass fraction $\Delta M/M$ and the SoftDrop groomed jet opening angle R_g . The jet population with $M = M_g$ has been separated out into the leftmost column. For the $M > M_g$ 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 35 (right) shows the projection of $\Delta M/M$ for different selections of R_g . Method 1 of 5.4 is used for systematic uncertainty estimation. We observe that, as shown in the blue data points, a selection on small R_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_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.

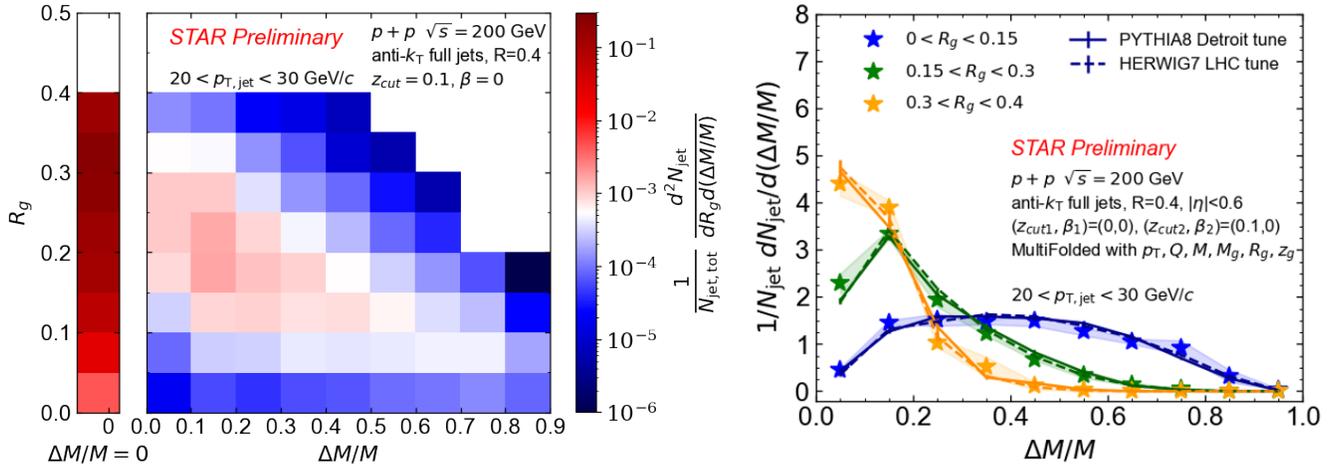


FIGURE 35. Correlation between the CollinearDrop groomed mass fraction $\Delta M/M$ and the SoftDrop groomed jet opening angle R_g . Correction is done with the 2015 embedding.

508 Figure 36 shows the correlation between $\Delta M/M$ and the SoftDrop groomed shared momentum fraction z_g . We observe that the
 509 more fractional mass that is groomed away by SoftDrop, the flatter the z_g distribution is. Since the perturbative DGLAP splitting
 510 function follows the $1/z$ behavior [1–3], a more steeply falling z_g with small ΔM indicates a larger perturbative contribution.
 511 This measurement demonstrates how an early-stage emission constrains the momentum imbalance of a later splitting. Again,
 512 PYTHIA and HERWIG are able to describe the data.

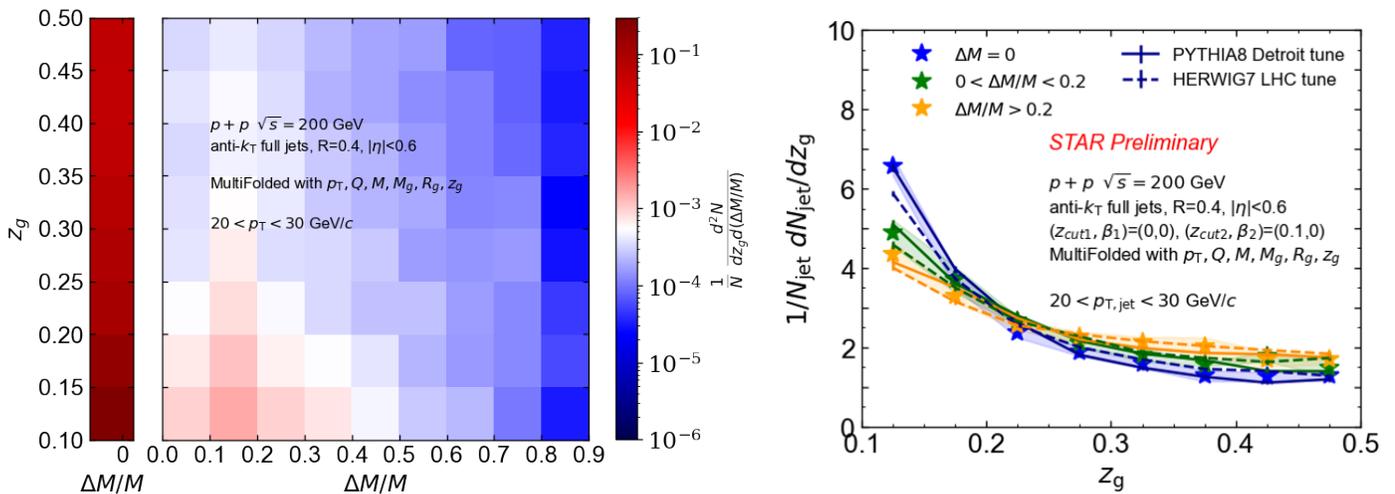


FIGURE 36. Correlation between the CollinearDrop groomed mass fraction $\Delta M/M$ and the SoftDrop groomed jet opening angle z_g . Correction is done with the 2015 embedding.

6.5.1. *Hadronization effect.* To make inference between early time and late time radiation in the parton shower, we confirm that effects of hadronization is suppressed in our correlation measurements. With PYTHIA8 Detroit tune, we compare the correlations with and without hadronization turned on. Figure 37 shows that hadronization only smears or shifts the distributions, but the soft-hard correlation with and without hadronization is the same. The different cuts for $\Delta M/M$ for the z_g take into account that a much larger fraction of jets have identically zero CollinearDrop mass without hadronization, as shown in Fig. 38.

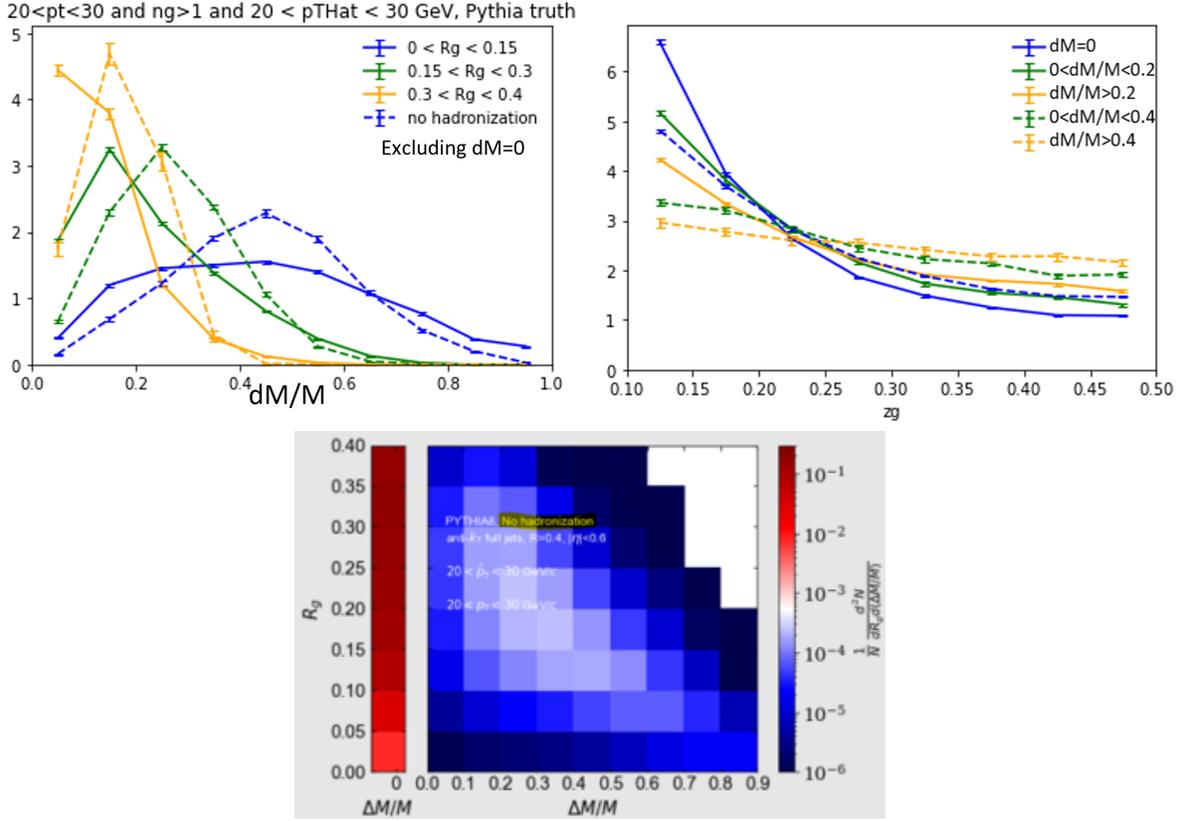


FIGURE 37. Correlation between the CollinearDrop groomed mass fraction $\Delta M/M$ and R_g or z_g , with and without hadronization, from PYTHIA simulations.

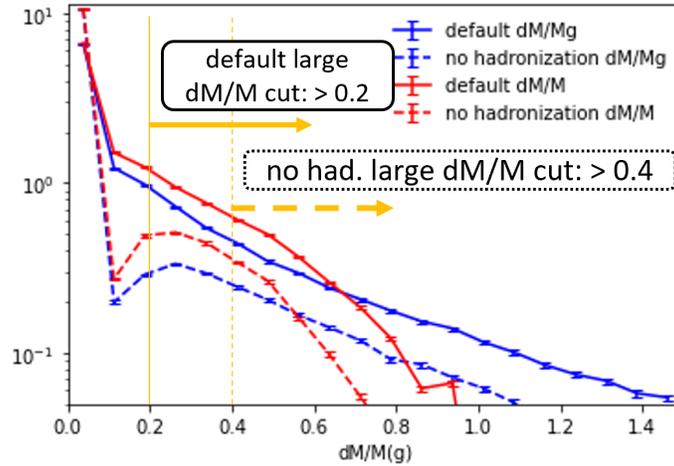


FIGURE 38. CollinearDrop $\Delta M/M$ distributions, with and without hadronization, from PYTHIA simulations.

518 6.5.2. *Switching from $\Delta M/M$ to a in 2D plots.* In this subsection, we use the 2021 embedding for corrections, and systematic
 519 uncertainties are treated with Methods 2 of Sections 5.3 and 5.4. Just as in Fig. 35, we see that when the π^0 decays at the
 520 truth-level are enabled, as in Figure 39, there is also a significant anti-correlation between $\Delta M/M$ and R_g . However, it is less
 521 significant between (1) a and R_g or (2) $\log(a)$ and R_g or (3) $\log(a \cdot p_T^2)$ and R_g .

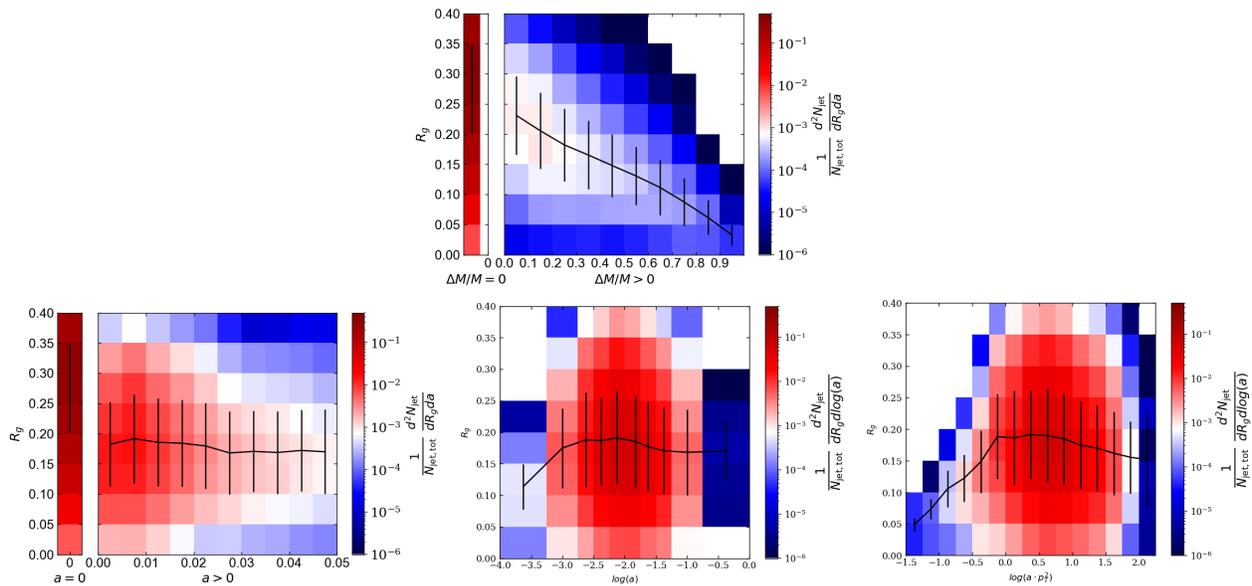


FIGURE 39. Fully corrected correlations between $\Delta M/M$ and R_g and between a -related observables and R_g . Correction is done with the 2021 embedding.

522 First, we confirmed that the correlation between the mean values of $\Delta M/M$ and $\log(a)$ is significant. See Figure 40. Note, for
 523 a given value of $\Delta M/M$ ($\log(a)$), the distribution of $\log(a)$ ($\Delta M/M$) is spread out, meaning that the strong correlation between
 524 $\Delta M/M$ and R_g might not carry on to a and R_g . More concretely, for $\Delta M/M > 0.7$, we have $R_g < 0.1$ on average according to
 525 to Figure 39. In Figure 40, the cells with $\Delta M/M > 0.7$ that are reddest have $-2 < \log(a) < -1.25$. However, if we select on
 526 $-2 < \log(a) < -1.25$, the average $\Delta M/M \approx 0.4$, corresponding to $R_g \approx 0.15$.

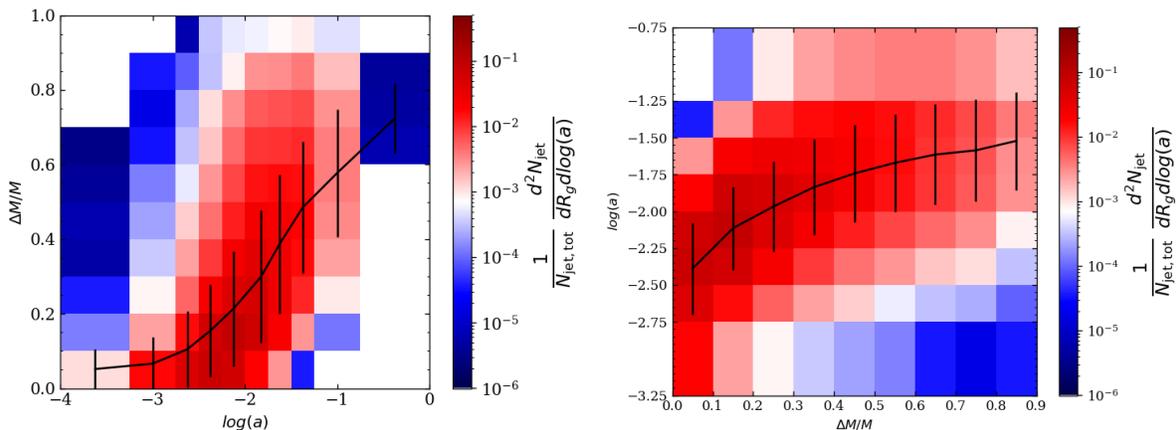


FIGURE 40. Fully corrected correlation between $\Delta M/M$ and $\log(a)$. Correction is done with the 2021 embedding.

527 We try to understand the turnover behavior of mean R_g value at around $\log(a) \approx -2.75$. Below $\log(a) \approx -2.75$, R_g increases
 528 with $\log(a)$, as shown in Fig. 39. This region is dominated by jets with small values of $M \approx M_g$. See Figure 41. The low R_g
 529 values and their relative narrow widths (indicated by the size of the vertical error bars) arise from the phase space restriction
 530 posed by the low-virtuality shower-initiating parton.

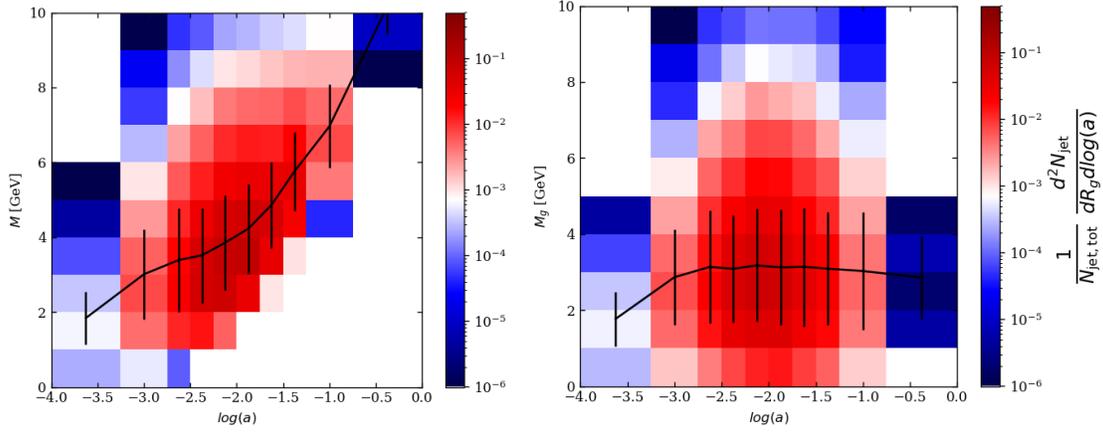


FIGURE 41. Fully corrected correlation between $M_{(g)}$ and $\log(a)$. Correction is done with the 2021 embedding.

531 6.5.3. *Effect of different z_{cut} values.* To further understand the turnover behavior of mean R_g value at around $\log(a) \approx -2.75$
 532 and what affects it, we study the correlations with various the z_{cut} values. With HERWIG simulations (decays of π^0 turned on),
 533 we select jets with $20 < p_T < 30$ GeV/c, and plot their correlation of R_g vs $\log(a)$, the same way as in the bottom middle plot of
 534 Fig. 39. Figure 42 shows the correlation for $z_{cut} = 0.05, 0.1, 0.3$. We observe that z_{cut} affects both the location of the turnover
 535 and the mean values of the R_g plateau after the turnover.

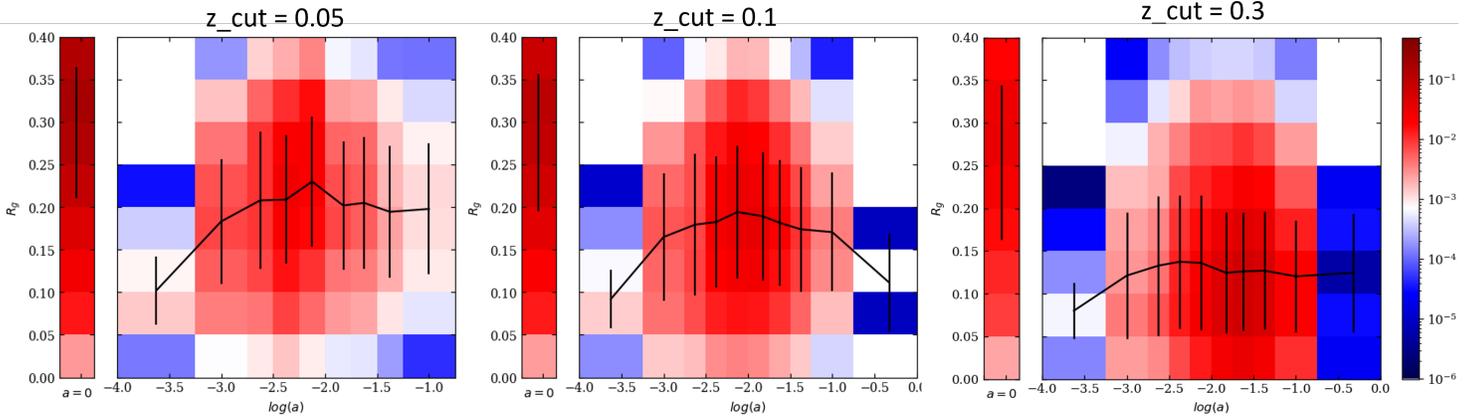


FIGURE 42. Correlation between R_g and $\log(a)$, with various z_{cut} values for SoftDrop criteria, from HERWIG jets with $20 < p_T < 30$ GeV/c (with decays of π^0 enabled, same as the 2021 embedding).

536 6.5.4. *Testing the angular ordering statement.* In this subsection, we investigate if our observation of the anti-correlation between
 537 $\Delta M/M$ and R_g in Fig. 35 can prove the angular ordering of the parton shower. With simulations, we first create a sample of
 538 jets where angular ordering (AO) is expected. Then we manually break the AO in the simulations, and study if the behavior of
 539 jet substructure observable correlations changes.

540 For the **baseline**, we first generate 200 GeV pp collisions with HERWIG event generator (with decays of π^0 turned off), run
 541 the usual anti- k_T algorithm with $R = 0.4$ to cluster the particles into jets, and run SoftDrop with the usual grooming criteria of
 542 $(z_{cut}, \beta) = (0.1, 0)$. We expect AO to hold in this case, since HERWIG uses an angular ordered parton shower.

543 For the **random ordering** toy model, the setup is as follows. We modify the HERWIG baseline such that, in each jet,

- 544 • for each jet constituent, keep (η, ϕ) the same, but randomize its p_T by reassigning it the p_T of a constituent in the same
 545 jet;
- 546 • recluster the modified constituents into the same jet;
- 547 • and run the usual SoftDrop criteria.

548 With this procedure, for each jet, p_T is roughly unchanged. Jet constituent multiplicity and spectrum are also identical to the
 549 usual case.

550 For the **thermal toy** model, we first select the baseline events that contain at least one jet above p_T of 15 GeV/c, and fit a
 551 piece-wise function for the spectrum of all the particles in such events, as shown in Figure 43. Then we

- 552 • generate particles from a “thermal event” by sampling from the fit;
- 553 • determine the event multiplicity for each event by sampling from a Gaussian with a mean of 100 and width of 10;
- 554 • cluster the particles into jets as usual;
- 555 • and run the usual SoftDrop criteria.

556 Figure 44 shows that with this procedure, we are able to replicate a roughly similar jet p_T spectrum and jet multiplicity
 557 distribution (left and middle plots, both in green), compared to the baselines (left and middle plots, blue). The jet constituent
 558 spectrum from the thermal toy events (right plot, blue) is less similar to the baseline case (right plot, black). Also shown in Fig.
 559 44 is the jet p_T and multiplicity distributions (left and middle plots, orange) where the procedure of generating the thermal events
 560 is the same, except that the event multiplicity is sampled from a Gaussian with a mean of 80 and width of 10. Although this
 561 variation does not affect the distributions significantly, it does not replicate the baseline distributions as well as the configuration
 562 with the mean of 100 does, so we will use the latter for further studies in the following.

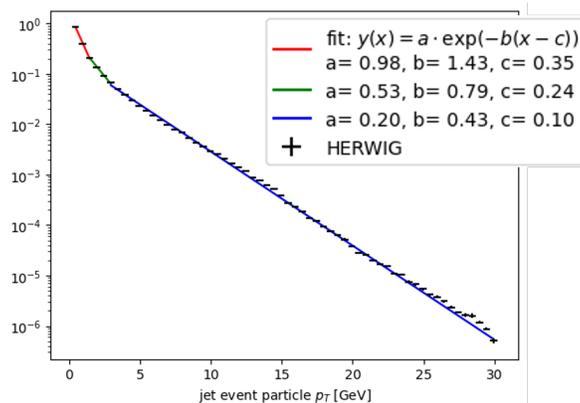


FIGURE 43. Spectrum and a fit of particles from HERWIG events that contain a jet.

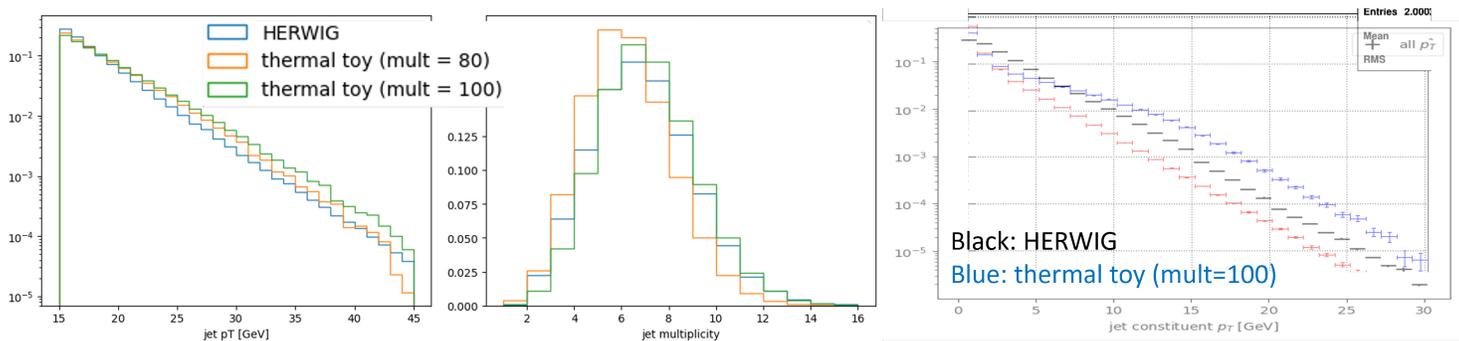


FIGURE 44. Jet p_T , multiplicity, and jet constituent p_T distributions in the thermal toy model, compared with the baseline.

563 We observe that in general, less grooming happened in the toy models. Specifically, the fraction of jets in $20 < p_T < 30$ GeV/c
 564 that has nothing removed by SoftDrop ($M = M_g$ case) is 73% in the random ordering model, 65% in the thermal toy model,

565 as compared to 54% in the baseline case. This makes sense because it is less likely for jets to have wide-angle soft radiation by
 566 setup. However, when grooming does happen, as seen in the distributions of the CollinearDrop groomed $\log(a)$ for jets from the
 567 baseline, random ordering model, and thermal toy model, in Figure 45, we observe that there is more mass groomed away in the
 568 random ordering case, and less mass groomed away in the thermal toy case. The former is likely due to the removal of wide-angle
 569 particles that have an intermediate p_T (about 2 GeV/c) and almost pass the z_{cut} of 0.1, in the random ordering model; the
 570 latter is likely due to the excess of very soft (less than 1 GeV/c) particles, in the thermal model, as seen in the leftmost bin of
 571 the right plot in Fig. 44.

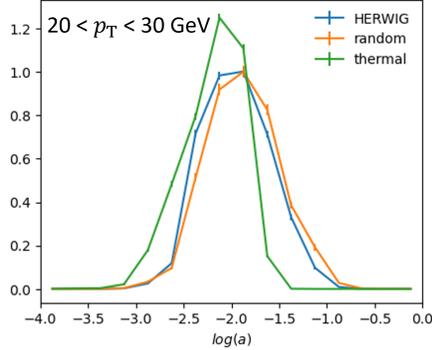


FIGURE 45. CollinearDrop groomed $\log(a)$ for jets from the baseline, random ordering model, and thermal toy model.

572 Figures 46 and 47 show respectively the analogs of Figs. 39 (lower middle panel) and 35. The similar qualitative behaviors in
 573 these toy models suggest that the anti-correlation between R_g and $\Delta M/M$ (or $\log(a)$ in the large $\log(a)$ region)
 574 due to AO. We consider the possibility that the anti-correlation arises from our choice of jet finding algorithm.

• R_g vs $\log(a)$

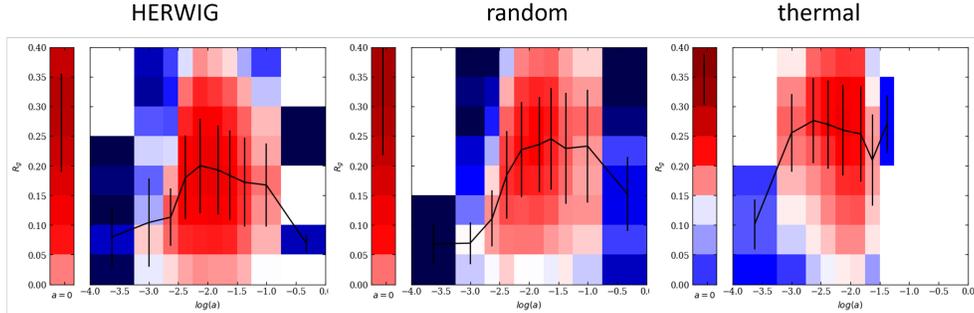


FIGURE 46. Correlations between the CollinearDrop groomed mass $\log(a)$ and R_g , in toy models compared with the HERWIG baseline.

$\Delta M/M$ for various R_g

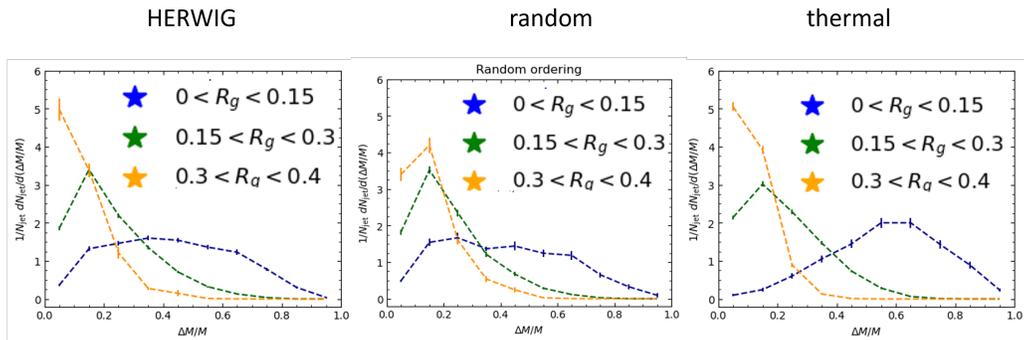


FIGURE 47. Correlations between the CollinearDrop groomed mass fraction $\Delta M/M$ and R_g , in toy models compared with the HERWIG baseline.

575 6.5.5. *Effect of different jet finding algorithms.* In this subsection, we study if (and how) the choice of jet finding algorithms
 576 affect the distributions and correlations of jet substructure observables.

577 Figure 48 shows the distributions of $\log(a)$, M , R_g , z_g and M_g for jets within $20 < p_T < 30$ GeV/ c , in HERWIG or thermal
 578 toy model events, clustered with the anti- k_T or k_T algorithm. We observe that the choice of jet finding algorithm affects the
 579 substructure observables (especially R_g) more, compared to whether the jets are from a physical event or a thermal toy model
 580 event.

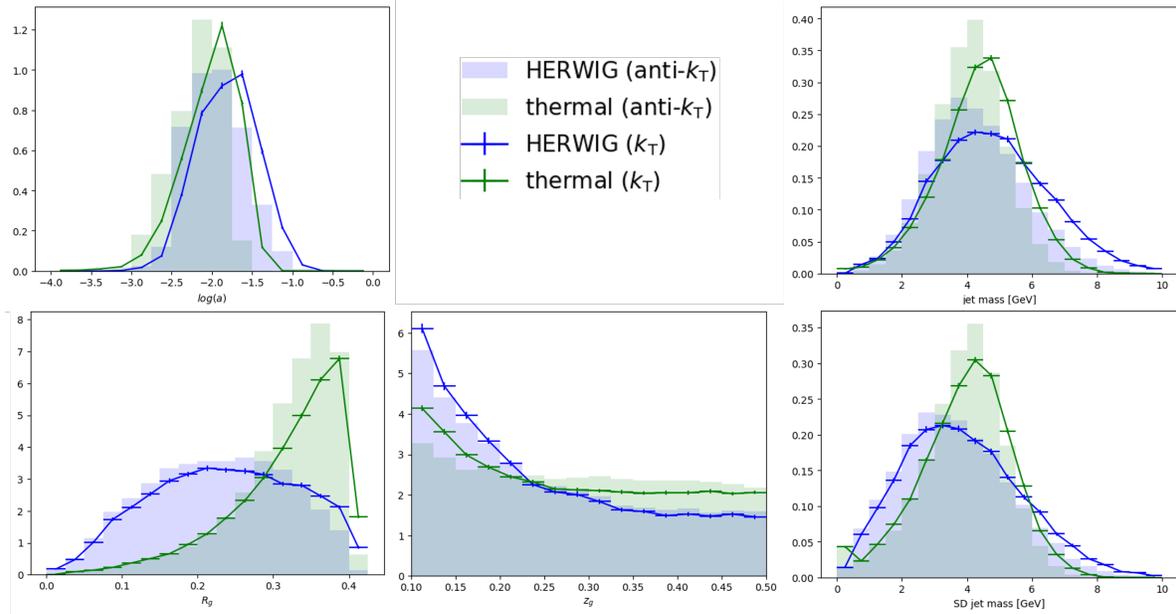


FIGURE 48. Distributions of $\log(a)$, M , R_g , z_g and M_g for jets within $20 < p_T < 30$ GeV/ c , in HERWIG or thermal toy model events, clustered with the anti- k_T or k_T algorithm.

581 Figure 49 shows the correlation between R_g and $\log(a)$ for the k_T jets. We observe that the anti-correlation between the two
 582 observables in the large $\log(a)$ region is broken significantly, as compared to the correlations in Fig. 46, where anti- k_T algorithm
 583 is used in either HERWIG or the toy models.

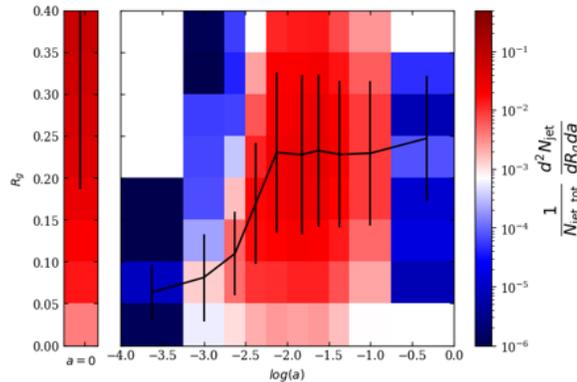


FIGURE 49. Correlations between $\log(a)$ and R_g , in HERWIG events clustered with the k_T algorithm.

584 These qualitative differences between HERWIG jets clustered with anti- k_T and k_T algorithms as seen in Figs 48 and 49
 585 indicates that jet substructure observables built using algorithmic procedures including might be susceptible to reconstruction
 586 biases in certain segments of the observable phase-space.

6.6. **CollinearDrop jet mass vs R_g and z_g .** Figure 50 shows the correlation between a and the SoftDrop groomed shared momentum fraction z_g and the SoftDrop groomed jet radius R_g in $20 < p_T < 30$ GeV/ c , where the average value of the CollinearDrop groomed jet mass is indicated by the color of each bin in the $z_g - R_g$ plane. The $M = M_g$ jets are included in this plot. This plane captures the Lund Plane of the first groomed splitting. We see that a is strongly correlated with R_g while weakly correlated with z_g .

Also shown in Fig. 50 is curves of constant formation time t , which approximates the time it takes for a parton to radiate a gluon [17]. It is related to other parton kinematic variables by:

$$(22) \quad t = \frac{1}{2Ez(1-z)(1-\cos(\theta))},$$

where E is the energy of the parent parton, z is the momentum fraction carried by the lower- p_T daughter parton, and θ is the opening angle between the two daughter. E can be approximated by the jet p_T ; for a perturbative hard splitting, z and θ can be approximated by the SoftDrop z_g and R_g , respectively [11]. We obtain the curves shown by replacing the parton variables in Eq. 22 with their (SoftDrop) jet counterparts, so t can be interpreted as the time that the first hard splitting to pass the SoftDrop criterion takes to develop. The strong correlation between a and R_g can therefore be understood as how the amount of early-stage radiation affects when the hard splitting happens. Specifically, to shed a significant amount of mass at the early stage of the parton shower, which is predominantly done via soft gluon radiation, the hard splitting needs to happen relatively late on average.

It is worth emphasizing that the measurement shown Fig. 50 showcases the power of MultiFold, which enabled us to make selections in three variables, p_T , z_g and R_g , and study a fourth one a which itself is composite of a few variables; all of these observables have been fully corrected for detector effects.

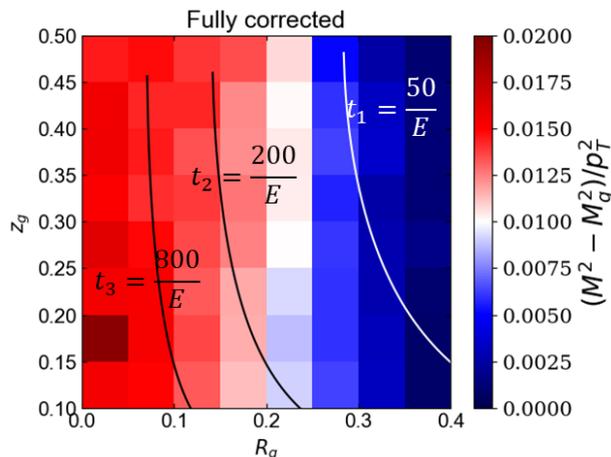


FIGURE 50. CollinearDrop groomed mass as a function of $z_g - R_g$. Correction is done with the 2021 embedding.

Figure 51 shows the $\log(a)$ distributions for specific regions of the $z_g - R_g$ plane. The leftmost bin includes the $a = 0$ jets, which do not have anything removed by SoftDrop and are therefore possibly dominated by jets whose first splittings in the parton shower are already perturbative. Region 3 ($0.15 < R_g < 0.25$ and $0.1 < z_g < 0.2$) includes asymmetric and intermediate-angle splittings while Region 2 ($0.15 < R_g < 0.25$ and $0.4 < z_g < 0.5$) includes symmetric and intermediate-angle splittings. Despite the different z_g selections, the fraction of $a = 0$ jets and the distributions in $a > 0$ are similar. The weak dependence of a on z_g is consistent with our observation made for Fig. 50.

However, as we continue to scan across the plane, we notice drastic changes in the fraction of jets with $a = 0$ as well as differences in shape in the $a > 0$ region. We first move onto Region 1 ($0 < R_g < 0.1$ and $0.4 < z_g < 0.5$), which includes symmetric and collinear radiation. Fig. 51 also shows that, compared to Regions 2 and 3, Region 1 is more likely to have soft radiation groomed away by SoftDrop as indicated by the decreased count for $a = 0$, and has a broader tail for the small but nonzero a region. On the other hand, we observe from Fig. 50 that we have on average higher values of a in this region, which can be understood as mostly affected by the slightly higher values in $\log(a) > -1.5$. The distribution of $\log(a)$ is wider in both directions arises from that a selection of narrow hard splitting opens up a large phase space for the amount of radiation preceding the splitting.

Region 4 ($0.3 < R_g < 0.4$ and $0.1 < z_g < 0.2$) includes asymmetric and wide-angle splittings, characteristic of perturbative early emissions. Again compared to Regions 2 and 3, in Region 4, the significant fraction of $a = 0$ jets indicates that it is highly probable that no non-perturbative early emission has happened before the perturbative emission. This is likely the explanation for why the z -axis values are also close to 0 in this region in Fig. 50.

Figure 52 shows the similar distribution as Fig. 50, from jets in PYTHIA8 (left) and HERWIG7 (right). Note that the decays of π^0 are not enabled in these simulations.

Figure 53 shows the similar distributions as Fig. 51, from jets in PYTHIA8 (left) and HERWIG7 (right). Note that the decays of π^0 are not enabled in these simulations.

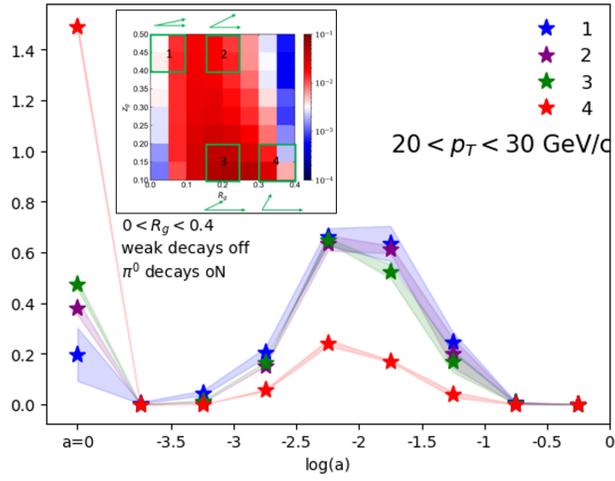


FIGURE 51. Distribution of $\log(a)$ with various selections of R_g and z_g . Correction is done with the 2021 embedding.

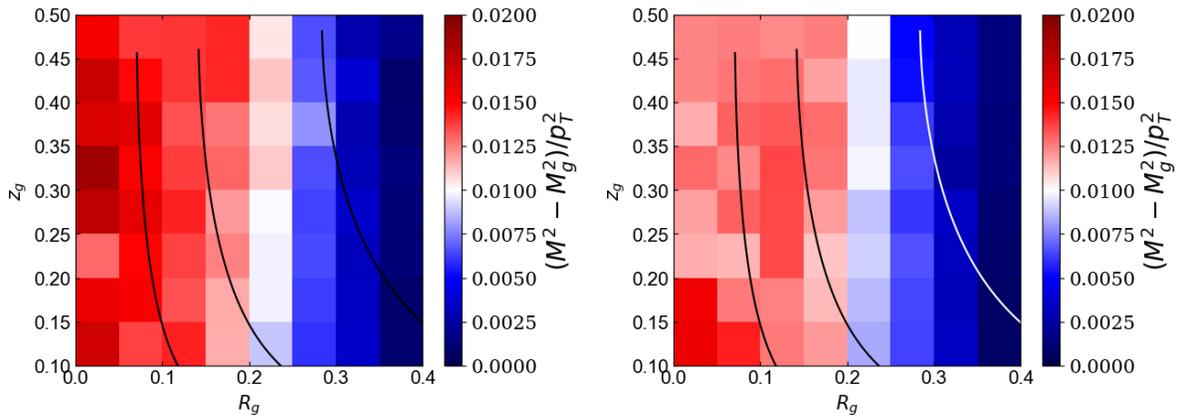


FIGURE 52. CollinearDrop groomed mass as a function of $z_g - R_g$, in PYTHIA (left) and HERWIG (right). The decays of π^0 are not enabled.

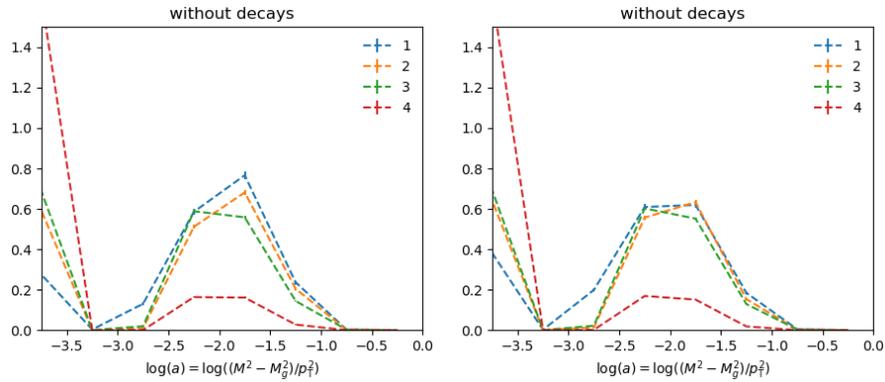


FIGURE 53. CollinearDrop groomed mass as a function of $z_g - R_g$, in PYTHIA (left) and HERWIG (right). The decays of π^0 are not enabled.

627 Figure 54 shows the similar distributions as Fig. 51, from jets in PYTHIA8 (left), HERWIG7 (middle), and PYTHIA6 (right).
 628 Note that the decays of π^0 are enabled in these simulations. PYTHIA6 here is from the truth distribution of the 2021 embedding.

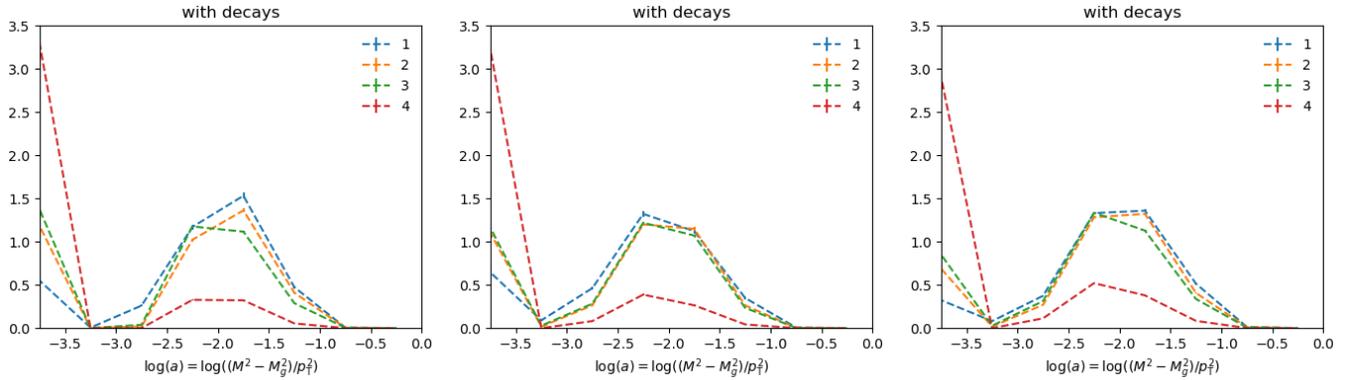


FIGURE 54. CollinearDrop groomed mass as a function of $z_g - R_g$, in PYTHIA8 (left) and HERWIG (middle) and PYTHIA6 (right). The decays of π^0 are enabled.

629

7. CONCLUSIONS

630 **Our paper plots include Figs 27, 50 and 51, discussed in Sections 6.4 and 6.6.** In this analysis note, we complement
 631 the paper titled by providing more details on MultiFold, a machine learning driven method used for unfolding. We discuss the
 632 mechanism and the specific application of it to our analyses, including the procedures to estimate the systematic uncertainties.
 633 In addition, the closure test performances and comparison with previously published results on jet mass corrected with an
 634 alternative method establish confidence for further application on substructure correlation studies.

635 We also present results that are not shown in the paper. These results are fully corrected for detector effects with MultiFold.
 636 The correlation between the jet mass and jet charge shows that jets with a higher mass tend to have a lower $|Q|$. This study
 637 also inspires us to consider fragmentation biases when trying to identify jets initiated by a parton of a specific flavor, as well
 638 as how it can be affected by local charge conservation arising from the string breaking mechanism. These provide avenues for
 639 future studies.

640 The correlations between the CollinearDrop groomed jet mass with SoftDrop groomed observables demonstrate how the
 641 early-time soft and wide-angle radiation affects the kinematics of the later-stage perturbative splitting. In particular, from the
 642 investigation of the correlation between the CollinearDrop groomed mass a and the SoftDrop groomed observables z_g and R_g , we
 643 observe that on average, a large early-time radiation biases the perturbative splitting to happen late. We also observed a strong
 644 correlation between the amount of early-stage radiation and the angular scale of a later-stage splitting. We carry out studies
 645 in simulations varying the effects of hadronization, angular ordering and the choice of jet finding algorithm, to investigate the
 646 origin of the anti-correlation between the CollinearDrop groomed mass and R_g .

REFERENCES

647

- 648 ¹V. N. Gribov and L. N. Lipatov, “Deep inelastic ep scattering in perturbation theory”, *Sov. J. Nucl. Phys.* **15**, 438–450 (1972).
- 649 ²G. Altarelli and G. Parisi, “Asymptotic freedom in parton language”, *Nuclear Physics B* **126**, 298–318 (1977).
- 650 ³Y. L. Dokshitzer, “Calculation of the structure functions for Deep Inelastic Scattering and e^+e^- annihilation by perturbation
- 651 theory in Quantum Chromodynamics”, *Sov. Phys. JETP* **46**, 641–653 (1977).
- 652 ⁴M. Cacciari, G. P. Salam, and G. Soyez, “The anti- k_t jet clustering algorithm”, *Journal of High Energy Physics* **2008**, 063
- 653 (2008).
- 654 ⁵S. Marzani, G. Soyez, and M. Spannowsky, *Looking inside jets* (Springer International Publishing, 2019).
- 655 ⁶A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, “Soft Drop”, *JHEP* **05**, 146 (2014).
- 656 ⁷Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, “Better jet clustering algorithms”, *JHEP* **08**, 001 (1997).
- 657 ⁸M. Wobisch and T. Wengler, “Hadronization corrections to jet cross-sections in deep inelastic scattering”, in *Workshop on*
- 658 *Monte Carlo Generators for HERA Physics (Plenary Starting Meeting)* (1998), pp. 270–279.
- 659 ⁹A. M. Sirunyan et al. (CMS), “Measurement of the splitting function in pp and Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV”, *Phys.*
- 660 *Rev. Lett.* **120**, 142302 (2018).
- 661 ¹⁰G. Aad et al. (ATLAS), “Measurement of soft-drop jet observables in pp collisions with the ATLAS detector at $\sqrt{s} = 13$ TeV”,
- 662 *Phys. Rev. D* **101**, 052007 (2020).
- 663 ¹¹J. Adam, L. Adamczyk, J. Adams, J. Adkins, G. Agakishiev, M. Aggarwal, Z. Ahammed, I. Alekseev, D. Anderson, A. Aparin,
- 664 et al. (STAR Collaboration), “Measurement of groomed jet substructure observables in p+p collisions at $\sqrt{s} = 200$ GeV with
- 665 STAR”, *Physics Letters B* **811**, 135846 (2020).
- 666 ¹²M. Aaboud et al. (ATLAS), “Measurement of the Soft-Drop jet mass in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS
- 667 Detector”, *Phys. Rev. Lett.* **121**, 092001 (2018).
- 668 ¹³A. M. Sirunyan et al. (CMS), “Measurement of the groomed jet mass in PbPb and pp collisions at $\sqrt{s_{NN}} = 5.02$ TeV”, *JHEP*
- 669 **10**, 161 (2018).
- 670 ¹⁴M. Abdallah et al. (STAR), “Invariant jet mass measurements in pp collisions at $\sqrt{s} = 200$ GeV at RHIC”, *Phys. Rev. D* **104**,
- 671 052007 (2021).
- 672 ¹⁵A. J. Larkoski, I. Moulton, and B. Nachman, “Jet substructure at the Large Hadron Collider: a review of recent advances in
- 673 theory and machine learning”, *Phys. Rept.* **841**, 1–63 (2020).
- 674 ¹⁶Y.-T. Chien and I. W. Stewart, “Collinear drop”, *Journal of High Energy Physics* **2020**, 1–59 (2020).
- 675 ¹⁷Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller, and S. I. Troian, *Basics of perturbative QCD* (1991).
- 676 ¹⁸G. D’Agostini, *Improved iterative bayesian unfolding*, 2010.
- 677 ¹⁹K. H. Ackermann et al. (STAR), “STAR detector overview”, *Nucl. Instrum. Meth. A* **499**, 624–632 (2003).
- 678 ²⁰I. Mooney, J. Putschke, and R. Kunnawalkam Elayavalli, *Analysis Note: Invariant jet mass measurements in pp collisions at*
- 679 *$\sqrt{s} = 200$ GeV at RHIC*, https://drupal.star.bnl.gov/STAR/system/files/ppJetMass_AN_GPC_v4.pdf.
- 680 ²¹T. Sjostrand, S. Mrenna, and P. Z. Skands, “PYTHIA 6.4 Physics and Manual”, *JHEP* **05**, 026 (2006).
- 681 ²²J. K. Adkins, “Studying transverse momentum dependent distributions in polarized proton collisions via azimuthal single spin
- 682 asymmetries of charged pions in jets”, PhD thesis (Kentucky U., 2015).
- 683 ²³P. Z. Skands, “Tuning Monte Carlo generators: the Perugia tunes”, *Phys. Rev. D* **82**, 074018 (2010).
- 684 ²⁴R. Brun, F. Bruyant, F. Carminati, S. Giani, M. Maire, A. McPherson, G. Patrick, and L. Urban, “GEANT detector description
- 685 and simulation tool”, 10.17181/CERN.MUHF.DMJ1 (1994).
- 686 ²⁵J. Adam, L. Adamczyk, J. Adams, J. Adkins, G. Agakishiev, M. Aggarwal, Z. Ahammed, I. Alekseev, D. Anderson, A. Aparin,
- 687 et al. (STAR Collaboration), “Underlying event measurements in $p+p$ collisions at $\sqrt{s} = 200$ GeV at RHIC”, *Physical Review*
- 688 *D* **101**, 052004 (2020).
- 689 ²⁶M. R. Aguilar, Z. Chang, R. K. Elayavalli, R. Fatemi, Y. He, Y. Ji, D. Kalinkin, M. Kelsey, I. Mooney, and V. Verkest,
- 690 “PYTHIA 8 underlying event tune for RHIC energies”, *Phys. Rev. D* **105**, 016011 (2022).
- 691 ²⁷J. Bellm, G. Bewick, S. Ferrario Ravasio, et al., “Herwig 7.2 release note”, *The European Physical Journal C* **80**, 452 (2020).
- 692 ²⁸A. Andreassen, P. T. Komiske, E. M. Metodiev, B. Nachman, and J. Thaler, “OmniFold: a method to simultaneously unfold
- 693 all observables”, *Phys. Rev. Lett.* **124**, 182001 (2020).
- 694 ²⁹A. Andreassen and B. Nachman, “Neural networks for full phase-space reweighting and parameter tuning”, *Phys. Rev. D* **101**,
- 695 091901 (2020).
- 696 ³⁰R. Kunnawalkam Elayavalli, J. Putschke, and K. Kauder, *Analysis Note: First fully unfolded measurements of jet splitting in*
- 697 *$p + p$ $\sqrt{s} = 200$ GeV at STAR*, https://drupal.star.bnl.gov/STAR/system/files/pp_zg_rg_analysisNote_v10.pdf.
- 698 ³¹Z. Chang and C. Gagliardi, *Analysis Note: Inclusive jet longitudinal double-spin asymmetry A_{LL} measurements in 510 GeV*
- 699 *polarized pp collisions at STAR*, <https://drupal.star.bnl.gov/STAR/system/files/AnalysisNoteV1.pdf>, 2018.
- 700 ³²J. Bellm et al., “Herwig 7.0/Herwig++ 3.0 release note”, *Eur. Phys. J. C* **76**, 1–8 (2016).