

Background estimator for jet studies in $p + p$ and $A + A$ collisions

Yacine Mehtar-Tani,^{1,*} Alba Soto-Ontoso,^{1,†} and Marta Verweij^{2,3,4,‡}

¹*Physics Department, Brookhaven National Laboratory, Upton, New York 11973, USA*

²*Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA*

³*RIKEN BNL Research Center, Brookhaven National Laboratory, Upton, New York 11973, USA*

⁴*Institute for Subatomic Physics, Utrecht University, Utrecht, Netherlands*



(Received 3 May 2019; revised manuscript received 15 November 2019; published 13 December 2019)

Experimentally, jet physics studies face an unavoidable task: distinguishing, at the detector level, the particles produced in the hard partonic scattering from the ones created in unrelated soft processes such as pileup interactions in high-luminosity proton-proton scattering or the underlying event in heavy-ion collisions. The fluctuating nature of the background constitutes the main source of uncertainty for any subtraction algorithm. Aiming at mitigating the effect of such fluctuations, we present a new method to estimate the background contribution to the transverse momentum on a jet-by-jet basis. Our approach is based on estimating the median background momentum density stored above a p_T -cut applied at the constituent level and an experimentally accessible correction term related to the signal contribution below the cut. This allows to trade part of the uncertainty due to background contamination for that of the signal below the cut, similarly to the SoftKiller method. We propose to reduce the fluctuations of the latter by exploiting intrinsic correlations among the soft and hard sectors generated in the branching process of QCD jets. Our data-driven approach is tested against PYTHIA8 and JEWEL dijet events embedded in minimum bias events and thermal background, respectively, and compared to the area-median and SoftKiller methods. The main result of this study is a $\sim 5\%$ – 35% improvement on the resolution of the reconstructed jet p_T compared to previous methods in both high-luminosity proton-proton and heavy-ion collisions.

DOI: [10.1103/PhysRevD.100.114023](https://doi.org/10.1103/PhysRevD.100.114023)

I. INTRODUCTION

The study of jets in ultrarelativistic hadronic collisions has decisively contributed to our understanding of the perturbative and nonperturbative aspects of quantum chromodynamics (QCD). Currently, at the Large Hadron Collider (LHC), events with jets in the final state are being extensively exploited to search for signatures of new physics. In the context of heavy-ion collisions, jet physics is immersed into the precision era both from a theoretical and an experimental perspective. Traditional observables, sensitive to the formation of a quark-gluon plasma (QGP), such as the nuclear modification factor [1,2] or dijet asymmetry [3], are being complemented with substructure analysis techniques inherited from the $p + p$ community both at RHIC and the LHC [4–6].

To perform meaningful theory-to-data comparisons, jet physics relies on two main ingredients: jet reconstruction algorithms and background subtraction methods. The former provide an operational definition of a jet by clustering the particles in the event following an infrared and collinear-safe method based on the notion of a distance between them either in momentum (k_T) or angular space [7–9]. The latter deals with the fact that the situation in which an isolated hard partonic scattering is responsible for all the clustered particles is hardly realized in current collider experiments. In high-energy hadronic collisions two competing mechanisms are relevant: pileup interactions, ubiquitous in high-luminosity facilities and the underlying event whose importance grows with the collision system size, i.e., from $p + p$ to $A + A$ [10]. Both of them populate the low- p_T sector of the event and are commonly referred to as *background*. The aforementioned clustering algorithms do not distinguish signal from background particles. Consequently, any reconstructed jet property X has two independent contributions, i.e., $X_{\text{jet}}^{\text{reco}} = X_{\text{jet}}^{\text{sig}} + X_{\text{jet}}^{\text{bkg}}$. Background subtraction methods are developed to subtract an estimated value for the $X_{\text{jet}}^{\text{bkg}}$ term. We shall focus in the present work on jet transverse momentum reconstruction so that $X = p_T$. To this aim, two steps have to be addressed. First, the value of $p_{T,\text{jet}}^{\text{bkg}}$ has

*mehtartani@bnl.gov

†ontoso@bnl.gov

‡m.verweij@uu.nl

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

to be estimated. This is an intricate task due to the fact that the background contribution fluctuates on an event-by-event basis and locally within the same event. Next, the question on how to remove this estimated background at the particle level arises. These two elements are strongly correlated as an improvement on the $p_{T,\text{jet}}^{\text{bkg}}$ estimate largely impacts the reconstruction of other features such as the jet mass [11–13]. In what follows we focus on background estimation. The subtraction part will be addressed in a future work.

Several methods are used to estimate the background transverse momentum contribution on a jet-by-jet basis in $p + p$ and $A + A$ collisions [14–17]. In this work we would like to draw the attention on the subtle interplay between signal and background in the low- p_T , few GeV, region. We pinpoint the underlying mechanisms driving the mean and standard deviation of the reconstructed jet momentum distribution. This systematic study leads to the design of a new background estimator that improves the jet p_T reconstruction resolution. Furthermore, we propose an experimentally measurable quantity, to be described in the following, to correct for signal contamination. Our method, dubbed ρ -correction as it corrects for the background energy density, aims at mitigating both background and signal fluctuations on a jet-by-jet basis.

This paper is organized as follows. In the next section we provide a detailed explanation of the generator samples together with a review of existing background estimators and emphasize the importance of signal contamination. The new approach and its application to $p + p$ and $A + A$ collisions are presented in Sec. III. To conclude, future lines of work are outlined in Sec. IV.

II. BACKGROUND ESTIMATION

In this section, the simulation framework together with the key ingredients of existing methods that aim to provide an accurate estimate of the background transverse momentum contribution to a jet p_T are exposed. In the last part of this section we analyze the impact of signal contamination below a given soft p_T^{cut} .

A. Simulation setup

First, we introduce the toy data samples that have been used to compute all the results presented in this manuscript.

The situation of a $p + p$ collision at 13 TeV with large pileup is mimicked by superimposing a PYTHIA8 [18] dijet event into a number n_{PU} of minimum bias events. The virtuality of the hard partonic scattering is set to $\hat{p}_T = 100$ GeV/ c and the number of pileup interactions is set to $n_{\text{PU}} = 200$ for the high-luminosity LHC [19]. For completeness, we also consider current LHC running conditions with $n_{\text{PU}} = 60$.

In the case of a central heavy-ion collision, which will be addressed in Sec. III B, we make use of the JEWEL [20]

event generator without recoil (JEWEL NR) embedded into a thermal background, as a proxy for the QGP. Dijet events are produced at a center-of-mass energy per nucleus-nucleus collisions $\sqrt{s} = 5.02$ TeV. Then, the parton shower is modified by collisions with the background constituents and medium-induced radiative processes. Background particles are obtained by sampling a Boltzmann distribution with a fixed multiplicity $\langle N \rangle = 7000$ and an average momentum $\mu = 1.2$ GeV/ c . This corresponds to an average momentum density of $\langle \rho \rangle = 250$ GeV, a realistic value for central Pb + Pb collisions. In this case, background particles are massless and uniformly distributed in (η, ϕ) -space with $|\eta| < 3$.

For each event, background and signal particles are clustered into anti- k_T [9] jets with $R = 0.4$ using FastJet3.1 [21]. The true, $p_{T,\text{truth}}^{\text{jet}}$, and the reconstructed, $p_{T,\text{reco}}^{\text{jet}}$, jet momenta are defined as follows:

$$p_{T,\text{truth}}^{\text{jet}} = \sum_i p_{T,i \in \text{sig}} \quad (1)$$

$$p_{T,\text{reco}}^{\text{jet}} = p_{T,\text{raw}}^{\text{jet}} - p_{T,\text{bkg}}^{\text{jet}}, \quad (2)$$

where $p_{T,\text{raw}}^{\text{jet}}$ refers to the jet momentum resulting from the clustering and the subscript sig labels the particles from PYTHIA/JEWEL. The goal of this paper is to provide an estimate of $p_{T,\text{bkg}}^{\text{jet}}$ that leads to a $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution with a close to zero mean and minimal standard deviation. Finally, the analysis is performed on jets with $p_{T,\text{truth}}^{\text{jet}} > 120$ GeV/ c .

B. Previous methods

The most largely used background estimation method by the jet experimental community is the area median [14] whose biggest strength is its data-driven and unbiased nature. On an event-by-event basis, the particles are clustered with the k_T algorithm [7]. After discarding the two hardest clusters (patches), under the assumption that they are predominantly composed of signal particles, ρ defined as

$$\rho = \frac{p_T}{A} \quad (3)$$

is computed for each cluster, where p_T and A are the total transverse momentum and area in the (η, ϕ) plane of the corresponding patch, respectively. The background contribution to the reconstructed jet p_T is then given by

$$p_{T,\text{bkg}}^{\text{jet}}|_{\text{area-median}} = \text{med}(\rho)A, \quad (4)$$

where $\text{med}(\rho)$ is the median p_T density of the set of patches.

Although highly successful at estimating $p_{T,\text{bkg}}^{\text{jet}}$ on average, this method lacks control over background

fluctuations. This translates into a $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution centered at zero but, as shown in Sec. III, with a sizable standard deviation [10].

An effective way of improving the resolution of the area-median method is to introduce a p_T^{cut} at the constituent level. If signal and background do not overlap in momentum space, removing all particles in the event below the separation scale would result into an exact estimation of the background in each patch (jet). In this ideal scenario the $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution would be a Dirac Delta. This is the underlying idea behind SoftKiller (SK) [15].

In this case, the event is clustered using a grid in (η, ϕ) -space of size a that is a free parameter of the method whose role is discussed below.¹ For every cell in the grid, the maximum constituent's momentum, $p_{T,i}^{\text{max}}$, is found in order to compute the p_T^{cut} defined as

$$p_T^{\text{cut,SK}} = \text{med}(p_{T,i}^{\text{max}}). \quad (5)$$

Then, the jet-by-jet background estimate of this method is expressed like

$$p_{T,\text{bkg}}^{\text{jet}}|_{\text{SK}} = \sum_i^{p_{T,i} < p_T^{\text{cut,SK}}} p_{T,i}. \quad (6)$$

Behind its apparent simplicity, this algorithm hides a few remarkable subtleties that we would like to comment on. Clearly, SoftKiller's performance depends on the scale separation between background and signal. In general, soft signal particles would be removed if they fall below the p_T^{cut} leading to a narrower but shifted to negative values $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution. An effective solution to overcome this bias is to tune the size of the patches where the p_T^{cut} is computed. That is, by construction there is a value of the parameter a or equivalently R_{kT} (the resolution parameter in the k_T clustering) for which the shift of the $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution is zero. This is so because a nontrivial interplay exists between the grid size/clustering radius and the p_T^{cut} : modifying R_{kT} is equivalent to change the value of p_T^{cut} and, consequently, the grid size can be adjusted to balance the signal contamination and background contribution below and above the cut. Therefore, in the following, we tune the value of R_{kT} such that the absolute mean value of the p_T reconstruction distribution is smaller than 0.5 GeV/c. We shall refer to this method as "SoftKiller R^{adj} ." Notice that this choice does not constrain, in principle, the standard deviation of the momentum reconstruction distribution. Finally, we would like to comment on the fact that, in this work, we use SoftKiller as a

background estimator and not as a subtraction method. In our opinion, the resulting $p_{T,\text{bkg}}^{\text{jet}}$ could be used as an input to different subtraction algorithms, e.g., [16], to avoid distortions in the jet substructure due to the removal of all particles below p_T^{cut} .

C. Impact of signal contamination

In the previous section, we remarked that any background estimator that relies on a soft momentum cut is sensitive to QCD radiation below this threshold. To quantify the signal contamination, we have computed the average contribution of the total jet p_T accounted by constituents whose momentum is below a certain p_T threshold:

$$p_{T,<}^{\text{sig}} = \sum_{i \in \text{sig}}^{p_{T,i} < p_T^{\text{cut}}} p_{T,i}. \quad (7)$$

The event-averaged value of this quantity, for the PYTHIA8 sample, as a function of multiples of the momentum cut at the constituent level, p_T^{cut} , is shown in Fig. 1. For completeness, we include the values of SoftKiller p_T^{cut} obtained by clustering the event containing both signal and background particles from $n_{\text{PU}} = 60$ minimum bias events into k_T -patches of $R_{kT} = 0.2, 0.4$, removing the two hardest ones and then applying Eq. (5). As expected, the signal contribution is a monotonous increasing function of p_T^{cut} . It is worthwhile to mention that even for considerably low thresholds such as 2 GeV/c the contribution of signal particles amounts to almost 15 GeV/c, which is more than 10% of the jet p_T selection threshold. This indicates a strong overlap between signal and background. Further, the aforementioned direct relationship between R and the SoftKiller p_T^{cut} [see Eq. (5)] is explicitly shown,

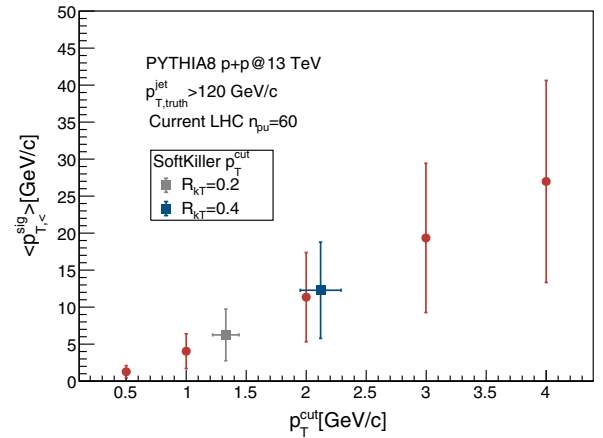


FIG. 1. Average p_T contribution from signal particles [see Eq. (7)] as a function of p_T^{cut} (red squares). The SoftKiller cut as given by Eq. (5) is represented by gray ($R_{kT} = 0.2$) and blue ($R_{kT} = 0.4$) squares. Error bars reflect the standard deviation.

¹In this work we use k_T -clustering instead of a rectangular grid to ensure an apples-to-apples comparison with the ρ -correction approach in what follows.

i.e., p_T^{cut} grows when increasing R_{kT} . More concretely, increasing the p_T^{cut} by 1 GeV leads to a signal contamination growth of ~ 5 GeV. This is an important aspect given the fact that SoftKiller R^{adj} is based on canceling out the signal contamination with the background contribution above the cut. That is, minimizing the shift requires a precise fine-tuning of R_{kT} and small deviations from the optimal value would result into a sizable bias in the p_T reconstruction.

In Fig. 1 we have presented an inclusive computation of $\langle p_{T,<}^{\text{sig}} \rangle$, i.e., integrating out the information about other jet features. However, QCD dynamics generates intrinsic correlations between the soft and hard p_T sectors within a jet owing to the tree structure of QCD parton cascades. The constituents whose momenta is below a given $p_T^{\text{cut,soft}}$, responsible for the $\langle p_{T,<}^{\text{sig}} \rangle$ -term, originate from the branching of hard p_T partons. Each of these multiple soft radiations is not independent from the rest of the shower as its phase space is restricted by mechanisms such as energy conservation or color coherence. This correlated radiation pattern is expected to manifest itself in jet substructure observables. For instance, an enhancement of soft radiation for broader jets is expected. We test this hypothesis within PYTHIA8 and the result is shown in Fig. 2 where the width is defined as [22]

$$\text{width}_{>,\text{hard}} = \sum_i^{p_{T,i} > p_T^{\text{cut,hard}}} \frac{p_{T,i}}{p_{T,\text{jet}}} \frac{\Delta R_i}{R_{\text{jet}}} \quad (8)$$

and $p_T^{\text{cut,hard}}$ is a hard momentum threshold above which the width is computed, R refers to the jet radius and ΔR_i is the cylinder (η, ϕ) -distance from each constituent to the jet axis. A clear correlation between $\langle p_{T,<}^{\text{sig}} \rangle$ and $\text{width}_{>,\text{hard}}$ is observed. Further, the standard deviation of $\langle p_{T,<}^{\text{sig}} \rangle$ is

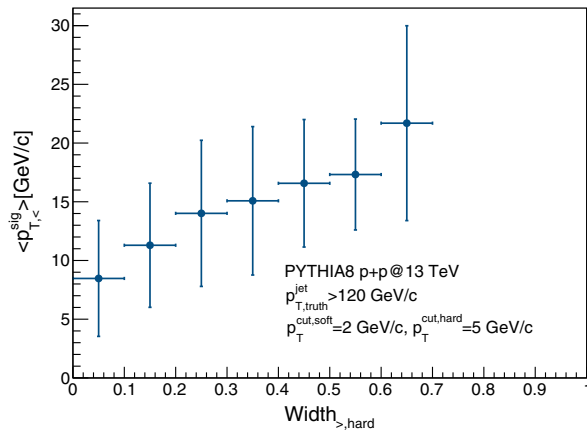


FIG. 2. Average p_T contribution from signal particles [see Eq. (7)] as a function of the width above the hard cut [see Eq. (8)]. The vertical error bars reflect the standard deviation while the horizontal ones represent the bin size.

reduced when fixing $\text{width}_{>,\text{hard}}$ as compared to the inclusive case (see Fig. 1). Therefore, signal fluctuations in the low- p_T regime can be mitigated by using hard-soft correlations within a jet. For simplicity, we restrict the results of this manuscript to the width and leave a more systematic study of other metrics and combinations of them for future work.

All the elements discussed in this section, i.e., a momentum cut, the median ρ and the average signal in the soft sector, are combined into a new background estimator to be described below.

III. THE ρ -CORRECTION METHOD

On an event-by-event basis, the ρ -correction algorithm proceeds as follows:

- (1) Divide the event into patches with $R_{kT} = 0.4$ via the k_T algorithm and remove the two hardest ones.
- (2) For each patch, compute ρ with the particles whose p_T is above a $p_T^{\text{cut,soft}}$ (which will be discussed below)

$$\rho_{>} = \frac{1}{A} \sum_i^{p_{T,i} > p_T^{\text{cut,soft}}} p_{T,i}, \quad (9)$$

where A denotes the area of the patch.

- (3) Find the median of these ρ values, $\text{med}(\rho_{>})$.
- (4) Recluster the whole event via the anti- k_T algorithm with $R = 0.4$.
- (5) For each anti- k_T cluster, add up the p_T 's of those particles that are below $p_T^{\text{cut,soft}}$:

$$p_{T,<} = \sum_i^{p_{T,i} < p_T^{\text{cut,soft}}} p_{T,i}. \quad (10)$$

- (6) The final estimate for the background p_T of a given anti- k_T jet is

$$p_{T,\text{bkg}}^{\text{jet}}|_{\rho\text{-correction}} = \text{med}(\rho_{>})A + p_{T,<} - p_{T,<}^{\text{shift}}, \quad (11)$$

where the subscripts “<”, “>” refer to the region where these quantities are computed, i.e., below or above the $p_T^{\text{cut,soft}}$.

The pocket formula of the ρ -correction method, Eq. (11) neglecting the last term, interpolates between the area median and SoftKiller for which $p_T^{\text{cut,soft}} = 0$ and $p_T^{\text{cut,SK}}$ [see Eq. (5)], respectively.

The role of the last element, $p_{T,<}^{\text{shift}}$, is to correct for the offset induced by signal contamination (see Figs. 1 and 2). To address the definition of $p_{T,<}^{\text{shift}}$, an explicit distinction between jets arising from $p + p$ and $A + A$ collisions is needed. This is due to the impact of the well established phenomenon of energy loss i.e., parton showers are

modified in the presence of a QGP leading to a degradation on the jet p_T with respect to its parent parton [23].

- (i) $p + p$.—We explore two possibilities for $p_{T,<}^{\text{shift}}$. First, using an inclusive determination of $\langle p_{T,<}^{\text{sig}} \rangle$ (ρ -correction) i.e., $p_{T,<}^{\text{shift}} = \langle p_{T,<}^{\text{sig}} \rangle$. Second, the background p_T estimate of each jet is corrected for signal contamination with a fluctuating value of $p_{T,<}^{\text{sig}}$ depending on its $\text{width}_{>,\text{hard}}$ such that

$$p_{T,<}^{\text{shift}} = \langle p_{T,<}^{\text{sig}} | \text{width}_{>,\text{hard}} \rangle. \quad (12)$$

Given the fact that we exploit hard-soft correlations we name this option “ ρ -correction HS_{corr}.” This choice requires a value of $p_T^{\text{cut,hard}}$ so that the width is computed in the background-free region. We fix it to be 5μ . We have checked that the results are robust against modifications of this value. Also, when $p_T^{\text{cut,hard}} \gg 5\mu$ the soft and hard sectors become decorrelated and the inclusive result is recovered. At this point the natural question on how to extract this information, outside a Monte Carlo environment, arises. We propose to measure it at low pileup proton-proton collisions. The idea of using low pileup events as relatively background-free samples was also explored in [24]. We have studied the behavior of $\langle p_{T,<}^{\text{sig}} \rangle$ with the jet p_T finding a logarithmic dependence. This is an important aspect in terms of applicability of the method to experimental data as inaccuracies in the determination of the true jet p_T due to detector effects or background contamination would not substantially affect the value of $\langle p_{T,<}^{\text{sig}} \rangle$. Therefore, the ρ -correction method ensures a $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution that is centered around zero thanks to the $p_{T,<}^{\text{shift}}$ -term.

- (ii) $A + A$.—The impossibility of experimentally accessing a background-free environment in a heavy-ion collision in which $p_{T,<}^{\text{shift}}$ could be determined in an unbiased way demands an alternative definition of it. For the ρ -correction case we propose

$$p_{T,<}^{\text{shift}} = \langle p_{T,<} \rangle - \langle \text{med}(\rho_{<}) A \rangle, \quad (13)$$

where $\langle p_{T,<} \rangle$ is obtained by performing an ensemble average of Eq. (10) over a set of unsubtracted jets i.e., containing signal and background particles. The second term in Eq. (13), $\text{med}(\rho_{<})$, is extracted from the same sample of k_T jets that we used to compute Eq. (9). Again, $\langle \cdot \rangle$ represents an ensemble average over the unsubtracted jets with area A . Remarkably, every quantity in Eq. (13) can be extracted from data making the proposed method directly applicable to experimental analyses in heavy-ion collisions.

If hard-soft correlations were to be exploited, the correction reads as follows:

$$p_{T,<}^{\text{shift}} = \langle p_{T,<}^{\text{sig}} | \text{width}_{>,\text{hard}} \rangle_{p+p} + \langle p_{T,<} \rangle - \langle \text{med}(\rho_{<}) A \rangle - \langle p_{T,<}^{\text{sig}} \rangle_{p+p}. \quad (14)$$

The two extra terms in Eq. (14), as compared to Eq. (13), have a transparent interpretation. As previously discussed, $\langle p_{T,<}^{\text{sig}} | \text{width}_{>,\text{hard}} \rangle_{p+p}$ cannot be determined from heavy-ion data. Therefore, one has to use the same correction as in the $p + p$ case i.e., obtained from low pileup events. Clearly, this introduces an additional shift that can be compensated with $\langle p_{T,<}^{\text{sig}} \rangle_{p+p}$, defined in Eq. (7).

Finally, the determination of $p_T^{\text{cut,soft}}$ proceeds as follows. We select a fixed value of $p_T^{\text{cut,soft}}$ requiring that it minimizes the standard deviation of the $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution. We have checked that the optimal value of $p_T^{\text{cut,soft}}$ does not strongly depend on whether the inclusive or exclusive version of $\langle p_{T,<}^{\text{sig}} \rangle$ is used to correct for the signal contamination. Hence, we believe that the most direct way to apply our method to experimental data is to use Eq. (11) with the inclusive value of $\langle p_{T,<}^{\text{sig}} \rangle$ to find the optimal $p_T^{\text{cut,soft}}$. Once this value is known one can use the width-dependent version of $\langle p_{T,<}^{\text{sig}} \rangle$ to get a further reduction on the p_T resolution. The results of the method in two different contexts are presented in the next section.

A. Proton-proton collisions

We first test the performance of the method on $p + p$ collisions with different pileup settings: $n_{\text{PU}} = 60$, the actual running conditions of the LHC and $n_{\text{PU}} = 200$, the foreseen scenario at the high-luminosity phase. In Fig. 3, the resolution of the reconstructed jet p_T as a function of $p_T^{\text{cut,soft}}$ for both cases is displayed. The generic shape of this curve can be understood using Eq. (11). For low values of $p_T^{\text{cut,soft}}$, $\text{med}(\rho_{>})$ dominates in Eq. (11) and, as expected, no significant improvement is achieved with respect to the traditional area-median method. In the opposite regime, when $p_T^{\text{cut,soft}} > p_T^{\text{cut,SK}}$, the contribution of $\text{med}(\rho_{>})$ is identically zero and σ monotonically increases with $p_T^{\text{cut,soft}}$ due to an increased number of signal particles being misidentified as background, as shown in Fig. 1.

The minimal resolution is achieved for $p_T^{\text{cut,soft}} = 1.4 \text{ GeV}/c$ in the $n_{\text{PU}} = 60$ case while in the high-luminosity scenario the p_T -resolution is optimized for $p_T^{\text{cut,soft}} = 2.2 \text{ GeV}/c$. In the case of SoftKiller, the optimal values of R_{kT} are $R_{kT} = 0.23$ for $n_{\text{PU}} = 60$ and $R_{kT} = 0.22$ for $n_{\text{PU}} = 200$. We observe a weak dependence of the standard deviation on $p_T^{\text{cut,soft}}$ in the region near the

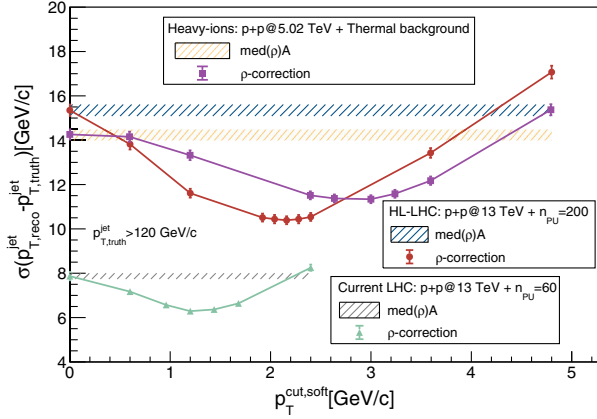


FIG. 3. Standard deviation for the jet p_T reconstruction distributions as a function of $p_T^{\text{cut,soft}}$ for the area-median approach and ρ -correction for three different scenarios: HL-LHC ($n_{\text{PU}} = 200$) (dashed blue, red dots), ($n_{\text{PU}} = 60$) (dashed grey, green triangles), JEWEL vacuum embedded in thermal background ($\langle N \rangle = 7000$, $\mu = 1.2$ GeV/c) (dashed orange, purple squares). Error bars reflect statistical fluctuations.

minimum as shown in Fig. 3. This is a clear advantage with respect to SoftKiller where a precise fine-tuning was needed in order to minimize the shift.

The precise values for the mean and standard deviation of the p_T -reconstruction distributions for both background settings can be found in Tables I and II for the four explored methods: area median, SoftKiller R^{adj} , ρ -correction and ρ -correction HS_{corr} where the width-dependent $\langle p_{T,<}^{\text{sig}} \rangle$ correction is obtained for the optimal value of $p_T^{\text{cut,soft}}$. First of all, any method that includes a momentum threshold highly reduces the standard deviation as compared to the area median. We find that the p_T resolution of ρ -correction and

TABLE I. Mean and standard deviation of the p_T signal reconstruction distributions for PYTHIA8 events embedded into $n_{\text{PU}} = 200$ minimum bias.

[GeV/c]	Mean	Standard deviation
med(ρ)A	-0.87 ± 0.37	15.35 ± 0.26
SoftKiller R^{adj}	-0.29 ± 0.25	10.37 ± 0.18
ρ -correction	1.20 ± 0.25	10.39 ± 0.18
ρ -correction HS _{corr}	1.26 ± 0.25	10.31 ± 0.18

TABLE II. Mean and standard deviation of the p_T signal reconstruction distributions for PYTHIA8 events embedded into $n_{\text{PU}} = 60$ minimum bias.

[GeV/c]	Mean	Standard deviation
med(ρ)A	-1.18 ± 0.19	7.87 ± 0.13
SoftKiller R^{adj}	-0.49 ± 0.15	6.36 ± 0.11
ρ -correction	0.13 ± 0.15	6.29 ± 0.11
ρ -correction HS _{corr}	0.14 ± 0.15	6.21 ± 0.11

TABLE III. Mean and standard deviation of the p_T signal reconstruction distributions for JEWEL events embedded into the ($\langle N \rangle = 7000$, $\mu = 1.2$ GeV/c) background.

[GeV/c]	Mean	Standard deviation
med(ρ)A	-0.61 ± 0.17	14.73 ± 0.12
SoftKiller R^{adj}	2.90 ± 0.12	10.91 ± 0.09
ρ -correction	0.18 ± 0.12	10.92 ± 0.09
ρ -correction HS _{corr}	-0.74 ± 0.12	10.34 ± 0.08

SoftKiller R^{adj} are comparable. These two methods, although approaching the problem from different angles, are in fact similar and, as noted in [15], the value of R_{kT} for which the shift is minimized results into a close-to-minimal resolution. In the high-luminosity scenario, the ρ -correction HS_{corr} method results in a (35%, 5%) improvement on the standard deviation with respect to area median and SoftKiller, respectively. Remarkably, our method is resilient against the particular density of the background i.e., the general trends persist when reducing the number of minimum bias events as observed in the bottom part of Fig. 3. As in the previous case, the best performance is achieved by ρ -correction HS_{corr}, i.e., when $\langle p_{T,<}^{\text{sig}} \rangle$ is computed for bins of width $_{>,\text{hard}}$. We conclude that the use of a fundamental property of QCD such as hard-soft correlations improves the background momentum estimation in jet experimental analyses. This is one of the main results of this paper.

B. Heavy-ion collisions: Impact of energy loss

The application of the ρ -correction algorithm to a heavy-ion context starts by obtaining the value of $p_T^{\text{cut,soft}}$ by minimizing $\sigma(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ within the JEWEL sample. The main source of uncertainty with respect to the $p + p$ case is the impossibility of determining $p_{T,\text{truth}}^{\text{jet}}$ due to energy loss. Therefore, to avoid Monte Carlo dependence, the adjustment of free parameters in any method has to be done with $p + p$ data. The standard deviation in vacuum is minimized, as shown in Fig. 3, for a value of $p_T^{\text{cut,soft}} \sim 2.5 \mu$ for which we compute $p_{T,<}^{\text{shift}}$ (both in the inclusive and the width-dependent case) in JEWEL without energy loss. In the case of SoftKiller R^{adj} we follow a similar procedure: first we optimize the value of R_{kT} in JEWEL without energy loss $R_{kT} = 0.21$ ($\langle p_T^{\text{cut}} \rangle = 3$, $\sigma(p_T^{\text{cut}}) = 0.05$ GeV/c) and then use it in the heavy-ion context.

These are all the ingredients that we need to apply Eq. (11). The two first statistical moments of the momentum reconstruction distributions in the JEWEL NR case are displayed in Table III. Regarding the standard deviation we reach a similar conclusion to that in the $p + p$ case: the ρ -correction HS_{corr} method reduces it by 0.6 and 4.4 GeV/c when compared to SoftKiller R^{adj} and area median, respectively.

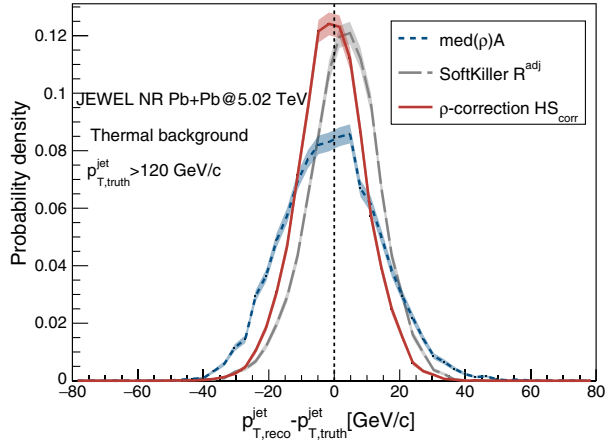


FIG. 4. Normalized p_T signal reconstruction distributions for the median approach (dotted, blue), SoftKiller R^{adj} (dashed, gray) and ρ -correction HS_{corr} (solid, red). The colored error bands account for statistical fluctuations.

The fact that the mean value within SoftKiller deviates from zero, as compared to the $p + p$ case, reflects in-medium jet modification absent in the vacuum baseline. In the case of both ρ -correction and ρ -correction HS_{corr} the mean value of the $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution is close to zero thanks to the $p_{T,<}^{\text{shift}}$ -term as given in Eqs. (13) and (14), respectively. This is a remarkable and nontrivial feature as the applicability of our unbiased method to heavy-ions does not lie on the common trade-off between resolution improvement and the shift on the mean due to energy loss. The ρ -correction approach leads to simultaneously obtain a momentum reconstruction distribution centered at zero alike the area median with a substantial reduction of its standard deviation along the lines of SoftKiller.

For completeness, the momentum reconstruction distributions are shown in Fig. 4. In there, the positive shift of SoftKiller together with the broadness of the area-median distribution are clearly visible. We conclude that ρ -correction HS_{corr} is an unbiased method ready to be used as background estimator in jet experimental analyses of heavy-ion collisions.

IV. SUMMARY AND OUTLOOK

In this work we propose a new background estimation method dubbed ρ -correction that aims at improving the resolution on jet p_T reconstruction analyses in hadronic collisions. To that end, we analyze the role of key elements of existing algorithms, such as the area median and SoftKiller, in terms of their impact on the mean and

standard deviation of the $(p_{T,\text{reco}}^{\text{jet}} - p_{T,\text{truth}}^{\text{jet}})$ -distribution. In the spirit of SoftKiller, we introduce a p_T^{cut} at the constituent level that we determine by minimizing the standard deviation of the jet momentum reconstruction distribution, in contrast to their mean value optimization. The price to pay for having a reduction on the standard deviation thanks to a p_T^{cut} is an overestimation of the background component due to abundant soft QCD radiation. Our method, summarized in a compact way by Eq. (11), overcomes this offset with an experimental accessible quantity: the average momentum stored below a given p_T^{cut} inside a jet to be measured in a clean environment like low pileup $p + p$ collisions. We show how the average of this distribution is enough to correct for the shift in high-luminosity proton-proton collisions while a couple of additional measurable terms are needed in a heavy-ion context due to in-medium jet modification. The method results into a comparable performance to SoftKiller in the high-luminosity scenario while it outperforms both area-median and SoftKiller in heavy-ion collisions.

Finally, we show that exploiting correlations between the soft and hard sectors within a jet, signal fluctuations are mitigated and, consequently, the jet momentum reconstruction resolution is improved with respect to the area-median and SoftKiller methods. The use of QCD correlations generated in the branching process of the shower to reduce the resolution in jet momentum reconstruction is the main novelty of this work. The natural continuation would be to implement a machine learning setup, in the spirit of [17], to systematically identify the set of substructure observables that better correlate with $\langle p_{T,<}^{\text{sig}} \rangle$. Work in this direction is on-going. Finally, the combination of this background estimator with a novel subtraction method will be the subject of a follow-up publication [13].

ACKNOWLEDGMENTS

We would like to express our gratitude to Antonio Bueno, Megan Connors, Kolja Kauder and Brian Page for helpful discussions during the realization of this work. We thank Gavin Salam for a careful reading of the manuscript and sharing his insights into the SoftKiller method. Y. M. T. and A. S. O.'s work was supported by the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, under Contract No. DE-SC0012704, and by Laboratory Directed Research and Development (LDRD) funds from Brookhaven Science Associates.

- [1] K. Adcox *et al.* (PHENIX Collaboration), *Phys. Rev. Lett.* **88**, 022301 (2001).
- [2] K. Aamodt *et al.* (ALICE Collaboration), *Phys. Lett. B* **696**, 30 (2011).
- [3] G. Aad *et al.* (ATLAS Collaboration), *Phys. Rev. Lett.* **105**, 252303 (2010).
- [4] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **120**, 142302 (2018).
- [5] K. Kauder (STAR Collaboration), *Nucl. Phys. A* **967**, 516 (2017).
- [6] H. A. Andrews *et al.*, [arXiv:1808.03689](#).
- [7] S. Catani, Y. L. Dokshitzer, M. H. Seymour, and B. R. Webber, *Nucl. Phys. B* **406**, 187 (1993).
- [8] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, *J. High Energy Phys.* **08** (1997) 001.
- [9] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [10] B. Abelev *et al.* (ALICE Collaboration), *J. High Energy Phys.* **03** (2012) 053.
- [11] S. Acharya *et al.* (ALICE Collaboration), *Phys. Lett. B* **776**, 249 (2018).
- [12] ATLAS Collaboration, ATLAS Report No. 2018/014, 2018.
- [13] Y. Mehtar-Tani, A. Soto-Ontoso, and M. Verweij (to be published).
- [14] M. Cacciari and G. P. Salam, *Phys. Lett. B* **659**, 119 (2008).
- [15] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **75**, 59 (2015).
- [16] P. Berta, M. Spousta, D. W. Miller, and R. Leitner, *J. High Energy Phys.* **06** (2014) 092.
- [17] R. Haake and C. Loizides, *Phys. Rev. C* **99**, 064904 (2019).
- [18] T. Sjostrand, S. Mrenna, and P. Z. Skands, *Comput. Phys. Commun.* **178**, 852 (2008).
- [19] ATLAS and CMS Collaborations, [arXiv:1902.10229](#).
- [20] K. C. Zapp, J. Stachel, and U. A. Wiedemann, *J. High Energy Phys.* **07** (2011) 118.
- [21] M. Cacciari, G. P. Salam, and G. Soyez, *Eur. Phys. J. C* **72**, 1896 (2012).
- [22] A. J. Larkoski, J. Thaler, and W. J. Waalewijn, *J. High Energy Phys.* **11** (2014) 129.
- [23] Y. Mehtar-Tani, J. G. Milhano, and K. Tywoniuk, *Int. J. Mod. Phys. A* **28**, 1340013 (2013).
- [24] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, *J. High Energy Phys.* **12** (2017) 051.