

Calibrating two jets at once

Rikab Gambhir^{1,2,*} and Benjamin Nachman^{3,4,†}

¹*Center for Theoretical Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA*

²*The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, USA*

³*Physics Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA*

⁴*Berkeley Institute for Data Science, University of California, Berkeley, California 94720, USA*



(Received 11 March 2024; accepted 11 September 2024; published 15 October 2024)

Jet-energy calibration is an important aspect of many measurements and searches at the LHC. Currently, these calibrations are performed on a per-jet basis, i.e., agnostic to the properties of other jets in the same event. In this work, we propose taking advantage of the correlations induced by momentum conservation between jets in order to improve their jet-energy calibration. By fitting the p_T asymmetry of dijet events in simulation, while remaining agnostic to the p_T spectra themselves, we are able to obtain correlation-improved maximum likelihood estimates. This approach is demonstrated with simulated jets from the CMS detector, yielding a 3%–5% relative improvement in the jet-energy resolution, corresponding to a quadrature improvement of approximately 35%.

DOI: [10.1103/PhysRevD.110.076015](https://doi.org/10.1103/PhysRevD.110.076015)

I. INTRODUCTION

Jet-energy calibration [1–6] is a crucial ingredient of most measurements and searches at the Large Hadron Collider (LHC). Given a spray of dozens to hundreds of detected particles, the goal is to estimate the underlying predetector total energy or p_T of the jet, along with a resolution on this estimate. This is a complex and high-dimensional problem, and there have been a large number of proposals, many using machine learning (ML), to improve jet-energy calibration [7–27] and uncertainty estimates [28–33]. Many of these approaches involve better or more sophisticated modeling of detector interactions or extracting more information from the high-dimensional particle phase space. A majority of these proposals directly regress the particle-level energy from detector-level quantities, which is prior dependent [21]. Some approaches achieve prior independence by avoiding direct regression [14,16,20]. As prior independence is a desirable feature for a universal calibration, we focus on this setting (in particular, using maximum likelihood estimation [20]).

In this paper, we present an alternate method for improving jet-energy calibrations and resolutions by considering

information *between* jets, rather than correlations *within* jets. In particular, we can obtain “correlation-improved” maximum likelihood estimates (MLEs) and resolutions for jets in dijet events by using the fact that their momenta add to approximately zero in the c.m. frame. This can be done in a p_T spectrum prior-independent way, that is, without any assumptions on the overall p_T distribution of the individual jets. We only assume that the momentum asymmetry between the two is either a Gaussian or symmetrized exponential of width Λ that can be extracted from simulation. For jets with original resolution σ , this leads to calibrations with resolutions improved by an average of $\frac{\sigma^2}{2\Lambda^2} \sim 3\text{--}5\%$ (representing a 35% improvement in quadrature¹) as estimated using simulations of the early Run 1 CMS detector [34], all while remaining unbiased.

To apply this procedure, all one has to do is fit the distribution of $z_1 - z_2$ (obtained, for example, via a truth-level Monte Carlo simulation) to either a Gaussian or exponential to extract Λ , then apply simple corrections to the estimates given by Eq. (10) (for a Gaussian) or Eq. (14) (for an exponential) to obtain the improved estimates. The choice of functional form is not strongly important, and either will result in an improvement.

While multijet information is not currently used to constrain the detector response of individual jets,

*Contact author: rikab@mit.edu

†Contact author: bpnachman@lbl.gov

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.

¹The “relative improvement” is simply the ratio of the new-to-old resolutions (minus 1), while the quadrature improvement is the relative difference between the new and old variances, as given precisely by Eq. (24). This is a measure of the amount of “new” information being used to constrain the resolutions.

constraints from multijet systems do play a role in jet calibrations. In particular, momentum conservation is used to determine how well the jet calibration is modeled by simulation, resulting in residual calibrations [1–6,35,36]. This information is the same as what we use in this paper, only instead of constraining all jets at once, we constrain each jet individually. For example, there are existing studies to use dijet balances to fit a p_T - and rapidity-dependent scale factor for the jet energy, but a given jet is not adjusted based on the energy of the balancing jet.

In Sec. II we go over the statistical methodology of the correlation-improved MLEs. In Sec. III we demonstrate our method on a simplified toy model to show its expected behavior. In Sec. IV we show how correlation-improved MLEs may be used to improve the p_T calibration of dijet events using simulated jets from the CMS detector. Finally, in Sec. V we briefly discuss how this technique may be applied to dijets sourced from boosted resonances. We present our conclusions and outlook in Sec. VI.

II. CORRELATION-IMPROVED MAXIMUM LIKELIHOOD ESTIMATES

In the following, x represents what we can measure and z represents the true (“latent”) values that we are trying to predict.² Given a single measurement $x \in \mathbb{R}^M$, we can construct an MLE $\hat{z}$ of a latent parameter $z \in \mathbb{R}^N$ and its resolution in the Gaussian approximation σ_z as follows (see Ref. [21] for a review):

$$\hat{z}(x) = \underset{z}{\operatorname{argmax}} L(x|z), \quad (1)$$

$$[\sigma_z^2(x)]_{ab} = - \left(\frac{\partial^2 L(x|z)}{\partial z_a \partial z_b} \right)^{-1} \Big|_{z=\hat{z}(x)}, \quad (2)$$

where $L(x|z)$ is shorthand for $\log(p(x|z))$. For example, $z \in \mathbb{R}$ will be our truth-level jet p_T and $x \in \mathbb{R}^{\sim 100}$ are the measured detector-level particles. The log-likelihood L can be estimated from data or simulation, using ML techniques if either x or z is high dimensional. The prototypical example is to think of $L(x|z)$ as a detector model or noise model, where noise or information loss is applied to a “true” parameter z to produce the measured value x . These estimates are manifestly independent of the true latent parameter prior $p(z)$ and produce independent estimates $\hat{z}$ for independent measurements x , which is a desirable feature for calibration.

Now, suppose that we measure a set of measurements all at once, with $x_1, \dots, x_n$ (denoted $\vec{x}$) corresponding to

$z_1, \dots, z_n$ (denoted $\vec{z}$). As a concrete example, we measure n different sets of particles x_i corresponding to n different jets, and we wish to determine all n truth-level jet p_T ’s, z_i . When each of the n objects is isolated,³ $L(x_i|z_i)$ is independent of i , i.e., the detector adds noise to each latent parameter independently. If the z_i are uncorrelated, then the total likelihood for the entire measurement factorizes and we can obtain independent MLEs and resolutions for each z_i . However, suppose we had prior knowledge of correlations $p(z_1, \dots, z_n)$. Then, by Bayes’ rule,

$$\begin{aligned} L(\vec{z}|\vec{x}) &= L(\vec{x}|\vec{z}) + L(\vec{z}) - L(\vec{x}) \\ &= \sum_i L(x_i|z_i) + L(\vec{z}) + \text{Norm}, \end{aligned} \quad (3)$$

where “Norm” is an unimportant constant to ensure that the distribution is normalized. The left-hand side of Eq. (3) is dependent on the prior, $L(\vec{z})$, which is undesirable for calibration tasks. However, it is sufficient for the multi-measurement calibrations we aim to perform that we are only independent with respect to each individual $L(z_i)$, that is, while we want to maintain our prior knowledge of the correlations between each z_i (e.g., “How is object 1 related to object 2?”), we do not want our inference to depend on the distribution of z_i (e.g. “How many of object 1 are there?”).

One way this can be accomplished is by decomposing $L(\vec{z}) = L(z_n|z_{n-1} \dots z_1) + L(z_{n-1}|z_{n-2} \dots z_1) + \dots + L(z_2|z_1) + L(z_1)$. The entire prior dependence is placed on $L(z_1)$, and so we can isolate and remove it to construct a likelihood.

Having removed explicit per-parameter prior dependence, we can construct a likelihood $L'(\vec{x}|\vec{z})$ (ignoring unimportant normalizations),

$$L'(\vec{x}|\vec{z}) = \sum_i (L(x_i|z_i) + L(z_i|z_{i-1} \dots z_1)), \quad (4)$$

which can be thought of as the ordinary likelihood $L(\vec{x}|\vec{z})$ plus a constraint on the relationship between the z_i ’s. The prime on $L'(\vec{x}|\vec{z})$ is to indicate that this likelihood is *not* just $L(\vec{x}|\vec{z})$, as it contains additional information about the constraints. We emphasize that this is not completely prior independent: the likelihood L' does depend on $p(\vec{z})$, but only through the correlations between the z_i ’s and not directly from any individual $p(z_i)$.

From Eq. (4), it is now possible to achieve a correlation-improved MLE estimate and resolution using Eqs. (1) and (2). Typically, as in the case in jet physics applications, x is very high dimensional [e.g., $M \sim \mathcal{O}(10^3)$], whereas z is often relatively low dimensional (e.g., for only the jet energy or p_T , $N = 1$). We can attempt to simplify the calculation of Eq. (4) by recasting it in terms of the

²To simplify notation, we use lowercase letters x and z to refer to both the random variable and the realization of the random variable.

³The general calibration problem would allow for the dimensionality of $\vec{x}$ to not necessarily equal that of $\vec{z}$. In that case, the calibration is not universal and other approaches are required.

low-dimensional $\hat{z}(x)$, as calculated in Eq. (1) if $\hat{z}(x)$ is a (nearly) sufficient estimator of z , then we may write⁴

$$L(\hat{z}(x)|z) \approx L(x|z), \quad (5)$$

$$\approx \log(\mathcal{N}(z, \sigma_z^2)), \quad (6)$$

where $\mathcal{N}$ is the probability density corresponding to the normal distribution with mean z and variance σ_z^2 , and σ_z is the resolution of the estimator $z(x)$ (in most cases, σ_z is the added Gaussian noise). Therefore, we can write

$$L'(\vec{x}|\vec{z}) = \sum_i (L(\hat{z}_i(x_i)|z_i) + L(z_i|z_{i-1} \dots z_1)), \quad (7)$$

which allows us to use our previously determined per-object estimators as inputs to the correlation-improved estimator.

A. The $n = 2$ Gaussian case

We are primarily interested in the $n = 2$ case of Eq. (7), where we have two correlated objects z_1 and z_2 . The correlation is completely determined by $L(z_2|z_1)$, which we assume takes the form

$$L(z_1|z_2) = L(z_1 - z_2). \quad (8)$$

We assume that $z_1 - z_2$ follows a Gaussian distribution with mean 0, though much of what we say here qualitatively holds for most other nonpathological distributions. In other words, z_1 equals z_2 , up to Gaussian noise Λ :

$$L(z_1 - z_2) = -\frac{1}{2\Lambda^2}(z_1 - z_2)^2 + \text{Norm}. \quad (9)$$

Then, given a set of measurements x_1 and x_2 , we can find correlation-improved estimates of the latent parameters z_1 and z_2 [which we denote $\hat{z}'_1(x)$ and $\hat{z}'_2(x)$] in terms of the original MLEs and resolutions (unprimed) by applying Eq. (1) to Eq. (7) and taking the maximum likelihood with respect to z_1 and z_2 :

$$\begin{aligned} \hat{z}'_1(x_1, x_2) &= \frac{1}{\sigma_\Lambda^2} (\hat{z}_1(\Lambda^2 + \sigma_{z_2}^2) + \hat{z}_2 \sigma_{z_1}^2), \\ \hat{z}'_2(x_1, x_2) &= \frac{1}{\sigma_\Lambda^2} (\hat{z}_2(\Lambda^2 + \sigma_{z_1}^2) + \hat{z}_1 \sigma_{z_2}^2), \end{aligned} \quad (10)$$

where for convenience we have defined $\sigma_\Lambda^2 \equiv \Lambda^2 + \sigma_{z_1}^2 + \sigma_{z_2}^2$. We have also suppressed the dependence of $\hat{z}_1$ and σ_{z_1} on x_1 , and the dependence of $\hat{z}_2$ and σ_{z_2} on x_2 for notational cleanliness.

⁴This is not guaranteed; prior independence is necessary, but not sufficient to be unbiased, as discussed in Ref. [21]. However, if the noise model $L(x|z)$ is Gaussian, $\hat{z}$ will indeed be unbiased.

These improved estimates are completely unbiased, *as long as x_1 and x_2 are unordered*. More precisely, as long as we are agnostic to whether or not $x_1 > x_2$, then the bias $b(z)$ is

$$\begin{aligned} b(z_{1,2}) &= \mathbb{E}_{p_{\text{Test}}} [\hat{z}'_{1,2}(x_1, x_2) - z_{1,2}], \\ &= 0, \end{aligned} \quad (11)$$

where p_{Test} is any test set such that $p(z_2|z_1)$ is correlated with the same Gaussian Λ .⁵ This is no longer true if we know which x is “leading” and “subleading”; the estimates $\hat{z}'$ will tend towards the resolution-weighted average of x_1 and x_2 , so the leading estimate always decreases and the subleading estimate always increases.

We can also compute the corresponding resolutions of each estimator using Eq. (2), $\sigma'_{z_1}(x)$ and $\sigma'_{z_2}(x)$, as the diagonal elements of the following covariance matrix:

$$\begin{aligned} \Sigma_z'^2(x_1, x_2) &= \frac{1}{\sigma_\Lambda^2} \begin{bmatrix} \sigma_{z_1}^2(\sigma_\Lambda^2 - \sigma_{z_1}^2) & \sigma_{z_1}^2 \sigma_{z_2}^2 \\ \sigma_{z_1}^2 \sigma_{z_2}^2 & \sigma_{z_2}^2(\sigma_\Lambda^2 - \sigma_{z_2}^2) \end{bmatrix} \\ &\downarrow \\ \sigma_{z_1}^2(x_1, x_2) &= \sigma_{z_1}^2 \left(1 - \frac{\sigma_{z_1}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2 + \Lambda^2} \right), \\ \sigma_{z_2}^2(x_1, x_2) &= \sigma_{z_2}^2 \left(1 - \frac{\sigma_{z_2}^2}{\sigma_{z_1}^2 + \sigma_{z_2}^2 + \Lambda^2} \right). \end{aligned} \quad (12)$$

In the case that z_1 or z_2 are multidimensional, this covariance matrix should be thought of as block diagonal in z_1 - z_2 space.

For any $\Lambda \geq 0$, Eq. (12) always improves the resolutions of both measurements. In the limit $\Lambda \rightarrow \infty$, we expect the correlation structure of $L(z_1 - z_2)$ to vanish, and indeed we see there is no improvement to the resolution and the off diagonals of the covariance matrix disappear. In this case, the improved estimates $\hat{z}'$ are completely unchanged from the original estimates $\hat{z}$. On the other hand, in the limit that $\Lambda \rightarrow 0$, we expect that $z_1 = z_2$ exactly, and indeed $\hat{z}'_1 = \hat{z}'_2$ in this limit. Both estimates converge to the resolution-weighted average of $\hat{z}_1$ and $\hat{z}_2$, with the improved resolution approaching the usual inverse-quadrature sum of σ_{z_1} and σ_{z_2} . In this limit, the two estimates become perfectly correlated with a Pearson correlation coefficient of $\rho = 1$. In the case where $\sigma_{z_1} \sim \sigma_{z_2}$, this leads to improved resolutions of $\sigma'_{z_i} \sim \sigma_{z_i}/\sqrt{2}$, an improvement of nearly 30%.

If we relax the assumption that $L(z_1 - z_2)$ is a Gaussian and instead consider generic unimodal and smooth distributions, it is still possible to obtain calibration-improved estimates. So long as $L(z_1 - z_2)$ has a sufficiently simple

⁵This is not strictly necessary; many other choices of the distribution of $|z_1 - z_2|$, including Gaussians and symmetrized exponentials for *any* width Λ , also lead to zero bias.

parametric form (which can be obtained from a simple fit) and is convex, it is still straightforward to calculate the correlation-improved MLEs and improved Gaussian resolutions, albeit only numerically.

B. The $n = 2$ exponential case

The case where $L(z_1 - z_2)$ is a symmetrized exponential is pathological and qualitatively different than the Gaussian case. If we take z_1 to be equal to z_2 up to *exponential* noise Λ_e of the form

$$L(z_1 - z_2) = -\frac{1}{\Lambda_e} |z_1 - z_2| + \text{Norm}, \quad (13)$$

then the singular behavior of the distribution at $z_1 = z_2$ causes our MLEs to be nonanalytic. Indeed, we find that there are three possible critical values, corresponding to being to the left of (Case 1), to the right of (Case 2), or directly on (Case 3) the critical point $z_1 = z_2$:

$$\begin{aligned} \text{Cases 1 and 2: } \hat{z}'_1(x_1, x_2) &= \hat{z}_1(x_1) \mp \frac{1}{2} \frac{\sigma_{\hat{z}_1}^2(x_1)}{\Lambda_e}, \\ \hat{z}'_2(x_1, x_2) &= \hat{z}_2(x_2) \pm \frac{1}{2} \frac{\sigma_{\hat{z}_2}^2(x_2)}{\Lambda_e}, \\ \text{Case 3: } \hat{z}'_1(x_1, x_2) &= \hat{z}'_2(x_1, x_2) = \frac{\sigma_{\hat{z}_1}^2 \hat{z}_2 + \sigma_{\hat{z}_2}^2 \hat{z}_1}{\sigma_{\hat{z}_1}^2 + \sigma_{\hat{z}_2}^2}. \end{aligned} \quad (14)$$

The only way to disambiguate the three cases is to evaluate the full likelihood [Eq. (7)], and then determine which of the three gives the maximum likelihood, which in general depends on the values of $\hat{z}_1$, $\hat{z}_2$, $\sigma_{\hat{z}_1}$, $\sigma_{\hat{z}_2}$, and Λ_e . However, we can glean some insight by considering the simpler case where $\sigma_{\hat{z}_1} = \sigma_{\hat{z}_2} = \sigma$, in which case the full likelihoods become, up to unimportant normalization factors,

$$L_{1,2} = -\left| \frac{\hat{z}_1 - \hat{z}_2}{\Lambda_e} \mp \frac{2\sigma^2}{\Lambda_e^2} \right| - \frac{\sigma^2}{\Lambda_e^2}, \quad (15)$$

$$L_3 = -\frac{(\hat{z}_1 - \hat{z}_2)^2}{4\sigma^2}. \quad (16)$$

If $\frac{(\hat{z}_1 - \hat{z}_2)^2}{4\sigma^2} < \frac{\sigma^2}{\Lambda_e^2}$, then Case 3 is preferred; otherwise, Case 1 or 2 is preferred, depending on the sign of $x_1 - x_2$. This can be understood as the estimators “snapping” to the resolution-weighted average if the measured values are sufficiently close, as controlled by Λ_e . This behavior is reminiscent of the similar “snapping” or “lasso” behavior of L1 regulators [37].⁶ Otherwise, the improved estimators

⁶In fact, it is exactly the same: L1 regulators can be thought of as placing a symmetrized exponential prior on parameters.

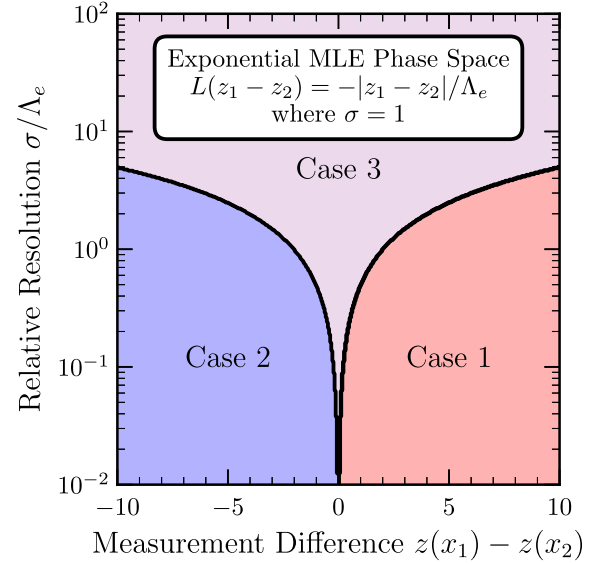


FIG. 1. Phase-space plot showing which of the three solutions presented in Eq. (14) is preferred as a function of $x_1 - x_2$ and σ/Λ_e , where $\sigma_{\hat{z}_1} = \sigma_{\hat{z}_2} = \sigma$ in this plot. Cases 1, 2, and 3 are shown in red, blue, and purple, respectively.

are just a slight shift of the original estimators. If the difference between x_1 and x_2 is large compared to the difference between z_1 and z_2 (given by Λ_e), there is a large correction; otherwise, there is only a small correction, and the jump is discontinuous. In Fig. 1, we show which of the three solutions is preferred as a function of $x_1 - x_2$ and σ/Λ_e .

Cases 1 and 2 merely shift the estimators, and thus the resolution is unchanged.⁷ Like the Gaussian case, as long as the measurements are unordered, these shifts do not produce bias since the sign of the shift will average to zero. On the other hand, Case 3 produces the ordinary resolution-weighted average of the two measurements and thus the improved resolution is the ordinary inverse-quadrature sum of the original resolutions. That is,

$$\sigma_{\hat{z}_i}^2 = \begin{cases} \sigma_{\hat{z}_i}^2 & \text{Cases 1 and 2,} \\ (\sigma_{\hat{z}_1}^{-2} + \sigma_{\hat{z}_2}^{-2})^{-1} & \text{Case 3.} \end{cases} \quad (17)$$

Note that, while it appears that Case 3 of Eq. (14) and all of Eq. (17) are explicitly independent of Λ_e , there is still a strong implicit dependence on Λ_e through the selection of the three cases. In the limits $\Lambda_e \rightarrow \infty$ and $\Lambda_e \rightarrow 0$, these estimates and resolutions agree exactly with the Gaussian version. This is to be expected as the former limit represents the no-information case and the latter limit represents the perfect-correlation case. However, unlike the Gaussian case, where Eq. (12) does *not* depend directly on $\hat{z}_1$ or

⁷There is a slight change due to the dependence of $\sigma_{\hat{z}_i}$ on x , making the shift not entirely constant, but we assume that σ varies slowly with x and ignore it.

$\hat{z}_2$, Eq. (17) does depend on $\hat{z}_1$ and $\hat{z}_2$ through the case choice. In the case where $\sigma_{\hat{z}_1} = \sigma_{\hat{z}_2} = \sigma$ for simplicity, we can use the exponential prior on $|z_1 - z_2|$ to directly obtain the average improvement in resolution:

$$\begin{aligned} \sigma_{\text{avg}}^2 &= \sigma^2 \left[\frac{1}{2} + \frac{1}{2} e^{-2\sigma^2/\Lambda_e^2} \right] \\ &\approx \sigma^2 \left(1 - \frac{\sigma^2}{\Lambda_e^2} \right). \end{aligned} \quad (18)$$

Note that this lines up with Eq. (12) in the limit that $\Lambda \gg \sigma$. That is, even though the Gaussian and exponential priors have very different behavior, we expect that they should both have the same *average* correlation-improved resolution when $\Lambda = \Lambda_e \gg \sigma$.

III. EXAMPLE 1: TOY MODEL DECAY

We begin with an example of applying the correlation-improved MLE using known distributions where we have full analytic and numeric control of the analysis.

Suppose we have a simplistic toy model where events have a “final state” consisting of three (distinguishable) particles, two of which are visible and one invisible. Focusing only on the x component of momentum, we measure the momenta of the first two particles to be $k_{1,\text{reco}}$ and $k_{2,\text{reco}}$, with unknown true values k_1 and k_2 that we would like to estimate. Assume that our detector adds a known amount of Gaussian noise σ_1 and σ_2 to each of the two measured particles, and that we cannot measure the third particle. We also assume that we know that the true

momentum k_3 follows a Gaussian with mean 0 and variance Λ^2 . This cartoon model resembles a $\tau \rightarrow \pi^- \pi^0 \nu_\tau$ decay in the rest frame of the τ , where $\sigma_- \neq \sigma_0$, and enables us to demonstrate the features of the correlation-improved MLE. We will revisit the assumption of knowing that the parent particle is perfectly at rest in Sec. V in the case of boosted resonance decays.

We can independently apply maximum likelihood estimation to obtain estimates for the values and resolutions of each measurement:

$$\begin{aligned} \hat{k}_1 &= k_{1,\text{reco}}, & \hat{\sigma}_1 &= \sigma_1, \\ \hat{k}_2 &= k_{2,\text{reco}}, & \hat{\sigma}_2 &= \sigma_2. \end{aligned} \quad (19)$$

Momentum conservation will induce correlations between k_1 and k_2 , via k_3 . To match the sign conventions of Sec. II, we choose k_2 to be in the negative direction. Since $k_1 - k_2 + k_3 = 0$, these correlations take the form

$$p(k_1 - k_2) = \mathcal{N}(0, \Lambda^2), \quad (20)$$

which is precisely the setup required to apply Eqs. (10) and (12) to improve the estimates and resolutions obtained in Eq. (19), with k_{reco} corresponding to the measurement x and k corresponding to the parameter z .

In Fig. 2, we show the results of the correlation-improved MLE procedure [Eqs. (10) and (12)] for various choices of missing momentum scale Λ , resolutions $\sigma_{1,2}$, and measurements $x_{1,2}$. We pick two regimes: $\sigma_1 \gg \sigma_2$ (Fig. 2a) and $\sigma_1 \sim \sigma_2$ (Fig. 2b); between these two plots, and the varying

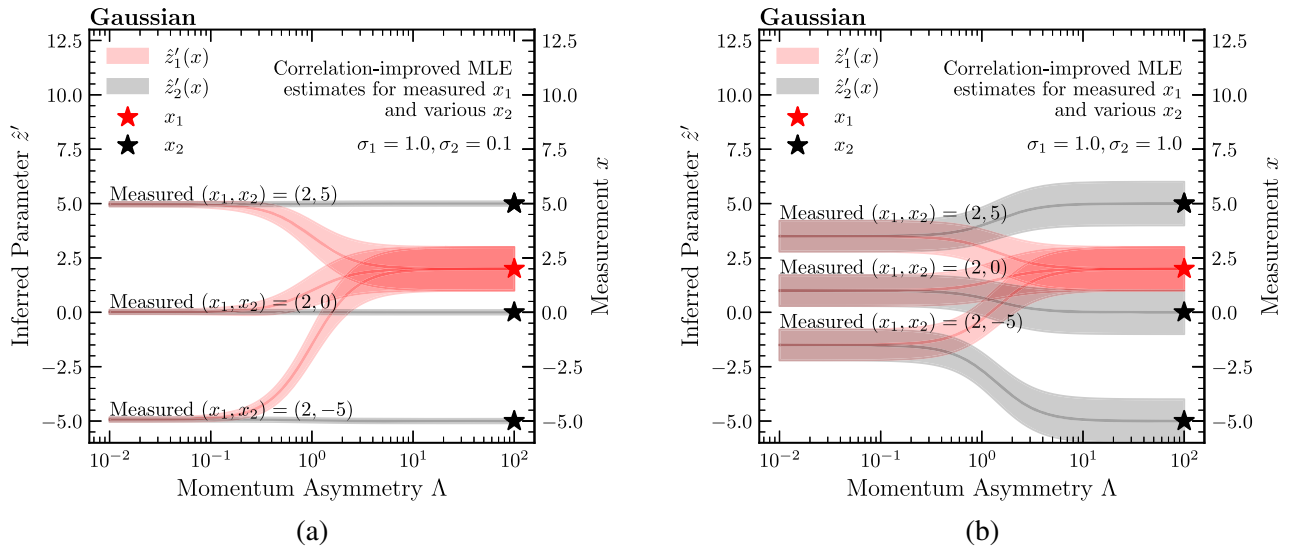


FIG. 2. Gaussian correlation-improved MLE parameters $\hat{z}'_1$ and $\hat{z}'_2$ as functions of Λ , for $\sigma_1 = 1.0$ and $\sigma_2 = 0.1$ (a), and $\sigma_1 = 0.1$ and $\sigma_2 = 1.0$ (b). A single x_1 measurement is shown with a red star ($x_1 = 2$), and several possible corresponding x_2 measurements are shown with black stars ($x_2 = -5, 0$, or 5). The bands correspond to the inferred $\hat{z}_1$ (red) and $\hat{z}_2$ (black), and the widths of the bands correspond to $\sigma_{\hat{z}_1}$ and $\sigma_{\hat{z}_2}$. As Λ decreases, going from right to left, the inferred estimates go from matching x_1 and x_2 with large bands on the right to matching the resolution-weighted average of x_1 and x_2 with smaller bands on the left.

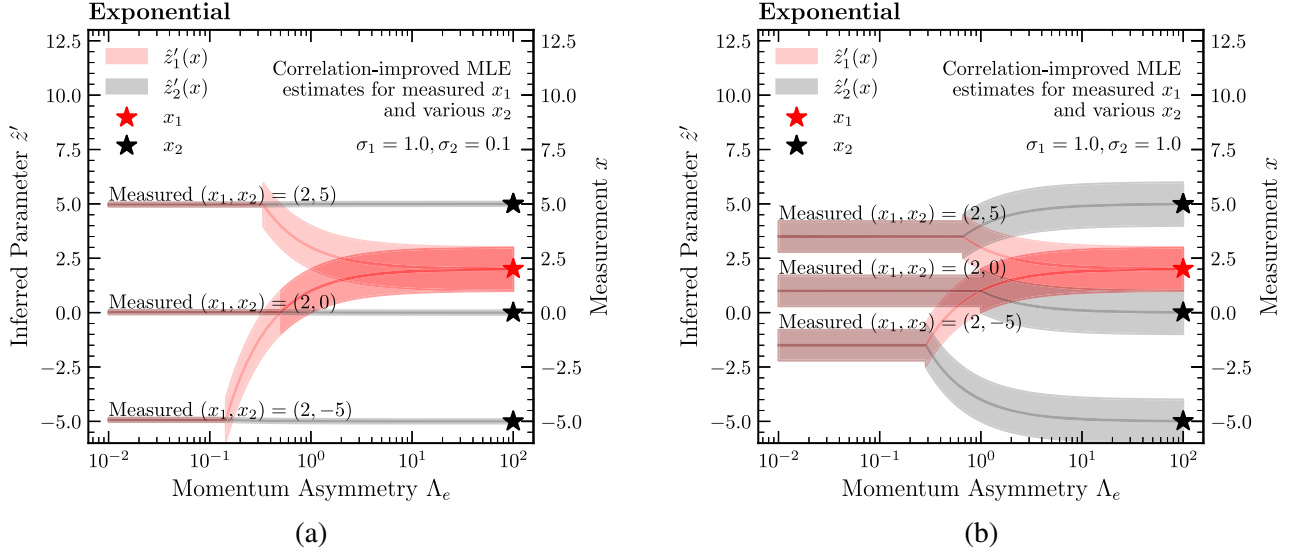


FIG. 3. The same as Fig. 2, but with an exponential prior rather than a Gaussian prior. Note the nonsmooth “snapping” behavior.

Λ on the x axis, all possible regimes in the $(\Lambda, \sigma_1, \sigma_2)$ parameter are represented. We pick one specific value for the x_1 measurement (red) and several values of a possible x_2 measurement (black), where the values of the measurements are represented by stars. Then, for each (x_1, x_2) pair, we draw bands corresponding to $\hat{z}'(x_1, x_2) \pm \sigma'_z(x_1, x_2)$ as a function of Λ . We can see that as $\Lambda \rightarrow 0$, the estimates converge to the same resolution-weighted average of the two measurements. On the other hand, as $\Lambda \rightarrow \infty$, the improved resolutions are no different than those expected by Eq. (19).

Similarly, we can also analyze the case of an exponential prior for $k_1 - k_2$,

$$p(k_1 - k_2) = \frac{1}{2\Lambda_e} e^{-\frac{|k_1 - k_2|}{\Lambda_e}}, \quad (21)$$

using Eqs. (14) and (17). In this scenario, the momentum of our cartoon ν particle is exponentially close to zero. The results of this calibration are shown in Fig. 3. Note that, while the limiting behaviors of these curves are similar to the Gaussian in Fig. 2, we can observe the characteristic “snapping” behavior of the exponential as $\Lambda \rightarrow \mathcal{O}(\sigma)$.

IV. EXAMPLE 2: QCD DIJETS

We now apply our technique to a more realistic scenario: the calibration of jet energies at the LHC, using the CMS detector [38] as an example. Our approach is deployed on simulated dijet events from CMS using early Run 1 conditions. In particular, in dijet events we expect the momentum of the two jets to be correlated due to momentum conservation, which is precisely the scenario in which we can apply correlation-improved calibration.

A. Data set

We use the same CMS Open Simulation [34,38] data set as described in Ref. [20], namely, the CMS2011AJets simulation data set. This data set consists of simulated dijets generated using Pythia 6 [39] with a Geant4-based [40] simulation of the CMS detector. This data set is formatted in the MIT Open Data format; more details about this format and data set can be found in Ref. [41].

The data consists of both detector-level “measured” jets, SIM, and corresponding generator-level “truth” jets, GEN. Each SIM event consists of a list of particle flow candidates (PFCs) [42], which are the reconstructed four-momentum and particle identification for each measured particle. The PFCs are clustered into anti- k_t jets with a radius parameter of $R = 0.5$ [43–45]. The data set has an implicit SIM p_T cut of 375 GeV on all jets, which was applied to avoid trigger turn-on affects. We additionally require that all jets are within $|\eta| < 2.4$ and that the jets are of at least “medium” quality [46]. Within this data set, almost all (99.4%) jets are paired up into dijet events. The remaining nonpaired jets are discarded. For each dijet event, we keep track of which is the leading and subleading jet at the SIM level. This information is *only* used to bin the jets and to differentiate the jets in the forthcoming plots; this is important for our calibrations to remain unbiased.

To mitigate the effect of the global 375 GeV trigger cut on all jets within this data set, we only look at events whose leading jet p_T is greater than 600 GeV. In Fig. 4 we show the distribution of the leading and subleading p_T ’s for each dijet event satisfying our cuts at both the SIM and GEN levels. For the following studies, the latent variable of interest z is the GEN jet p_T , while the corresponding measured quantities x are either the SIM jet p_T or the SIM PFCs. We bin the dijet events according to their leading SIM p_T . The bins we consider are [600, 700] GeV,

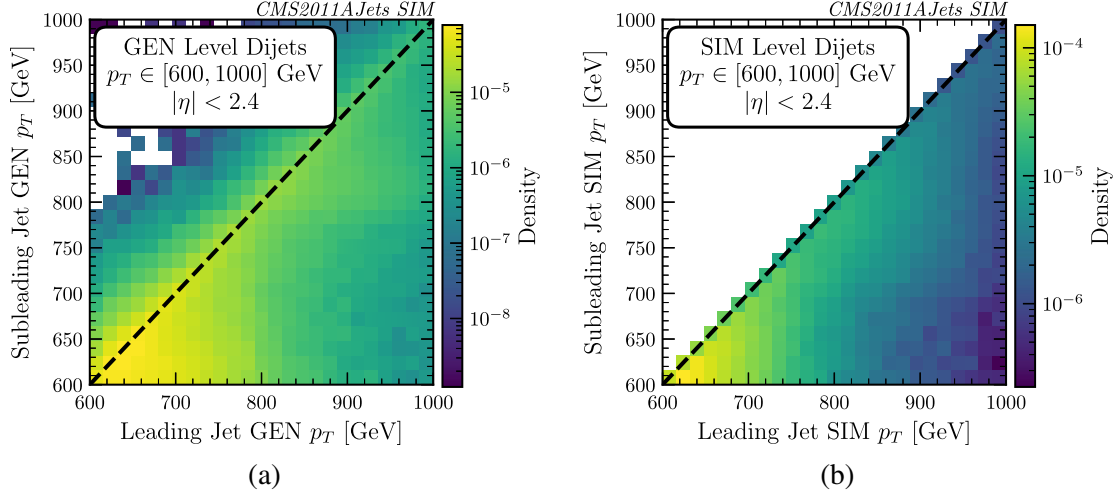


FIG. 4. Distribution of the (a) GEN p_T and (b) SIM p_T of the leading and subleading jets (as determined by the SIM p_T) of each event in the data set.

[700, 800] GeV, [800, 900] GeV, and [900, 1000] GeV. We perform our analysis separately on each bin, as well as one additional time on the entire inclusive [600, 1000] GeV set.

B. Momentum asymmetry fits

Due to momentum conservation, the p_T 's of dijets are correlated. In particular, at the truth level, $p_{T,1} \approx p_{T,2}$; however, a variety of factors make this inexact, such as initial-/final-state radiation, the underlying event, neutrinos/electroweak radiation, and inherent transverse momentum

imbalances (see transverse-momentum-dependent effects in [47]).

As in the case of the toy example presented in Sec. III, as long as the distribution $L(p_{T,1} - p_{T,2})$ is known, it is possible to use Eqs. (10)–(12) to obtain improved calibrations beyond what is possible in the single-jet case.

We perform two different fits and two independent analyses: one assuming an exponential fit with width Λ_e with the same functional form as Eq. (21), and one assuming a Gaussian fit with width Λ with the same functional form as Eq. (20). We find that the exponential

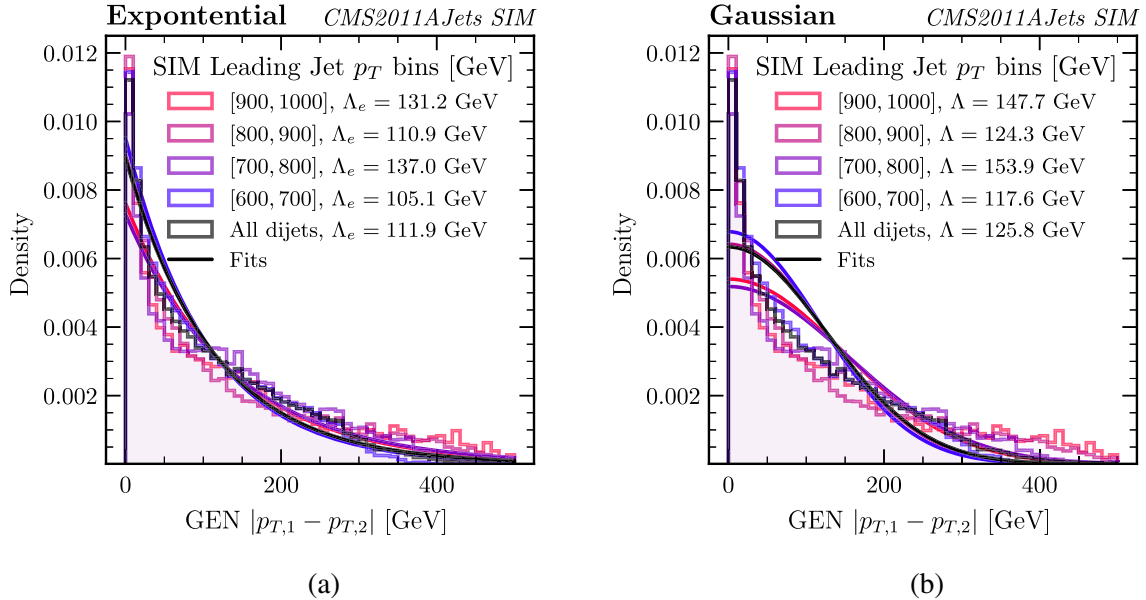


FIG. 5. Distribution of the difference in GEN p_T between the leading and subleading jets in each event, for several leading p_T bins. The absolute value is taken for plotting convenience; the labels 1 and 2 are random and the curves are symmetrized about the vertical axis. An (a) exponential and a (b) Gaussian distribution is fit to each distribution, and the parameters Λ_e and Λ (respectively) are extracted to describe the extent to which momentum is conserved. The different p_T bins are shown in blue through red in increasing p_T , and the p_T -inclusive sample is shown in black.

TABLE I. Mean squared errors and fitted Λ values for the exponential and Gaussian fits to the $|p_{T,1} - p_{T,2}|$ distributions for different leading jet p_T bins.

Bin [GeV]	MSE (exponential)	Λ_e [GeV]	$\Lambda_e/\text{Central } p_T$	MSE (Gaussian)	Λ [GeV]	$\Lambda/\text{Central } p_T$
[600, 700]	2.8×10^{-7}	105.1	0.162	6.7×10^{-7}	117.6	0.181
[700, 800]	2.8×10^{-7}	137.0	0.183	8.7×10^{-7}	153.9	0.205
[800, 900]	2.9×10^{-7}	110.9	0.130	5.6×10^{-7}	124.3	0.146
[900, 1000]	2.2×10^{-7}	131.2	0.138	7.2×10^{-7}	147.7	0.155
Inclusive	2.7×10^{-7}	111.9	0.140	7.5×10^{-7}	125.8	0.157

fits more closely match the distribution, but since exponentials have pathological behavior (as discussed in Sec. II B), we do the analysis a second time with Gaussian fits. Consistent improvements in the resolution are still possible even if the particle-level fit is not exact.

In Fig. 5 we show the distribution of $|p_{T,1} - p_{T,2}|$ in dijets at the GEN level, with the absolute value taken for plotting convenience. Note that the labels 1 and 2 do *not* correspond to leading or subleading, but rather are randomly assigned so that the distribution $L(p_{T,1} - p_{T,2})$ is symmetric and we apply the same initial calibration to both jets agnostic to whether it was the leading or subleading jet. We emphasize that this is crucial to avoid introducing bias. We show both distributions binned in the leading SIM p_T value (blue through red) and the total inclusive distribution (black). For each distribution, we fit an exponential (Fig. 5) and a Gaussian (Fig. 5) to extract the momentum asymmetry parameters Λ_e and Λ , respectively, which are shown in the respective plots. The fits are accomplished by minimizing the mean-squared error (MSE) loss of the distribution. We also summarize the results of these fits in Table I, where we show the extracted momentum

asymmetry values, fractional asymmetry, and fit MSE for the exponential and Gaussian.

Both by eye and by looking at the MSEs in Table I, we can see that the exponential is a much better fit than the Gaussian, but we consider both approaches to highlight their similarities and differences. The extracted Λ_e and Λ values range from ~ 100 – 150 GeV across the different p_T bins.

C. Correlation-improved jet calibration

We now show how Eqs. (10)–(12) can be used to improve jet-energy resolutions. As a baseline, we use CMS-prescribed resolutions, as computed in Ref. [2]:

$$\frac{\hat{\sigma}_{p_T}}{\hat{p}_T} = \sqrt{\frac{\text{sgn}(N)N^2}{\hat{p}_T^2} + \frac{S^2}{\hat{p}_T} + C^2}, \quad (22)$$

where $N \approx 30$ GeV, $S \approx 0.81$ GeV $^{1/2}$, and $C \approx 0.04$. This leads to a jet-energy resolution of approximately 3%–5% for jets in the p_T ranges we consider.

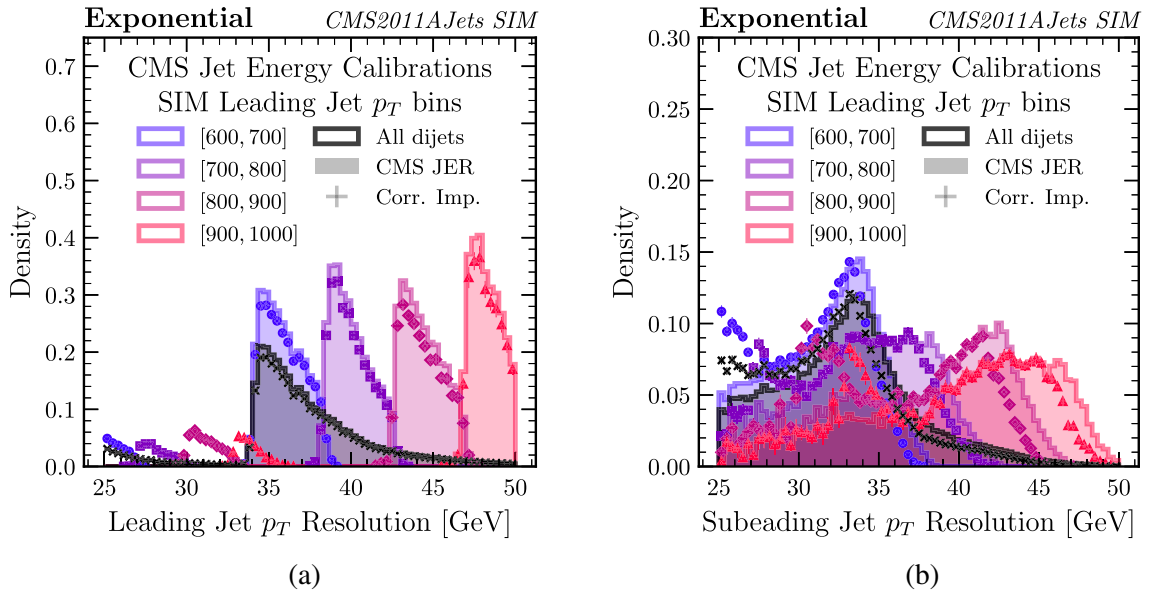


FIG. 6. Distributions of (a) leading and (b) subleading jet resolutions for each p_T bin, with the original CMS resolution (solid) and the correlation-improved (Corr. Imp.) resolution (points), for the exponential fits. The different p_T bins are shown in blue through red in increasing p_T , and the p_T -inclusive sample is shown in black.

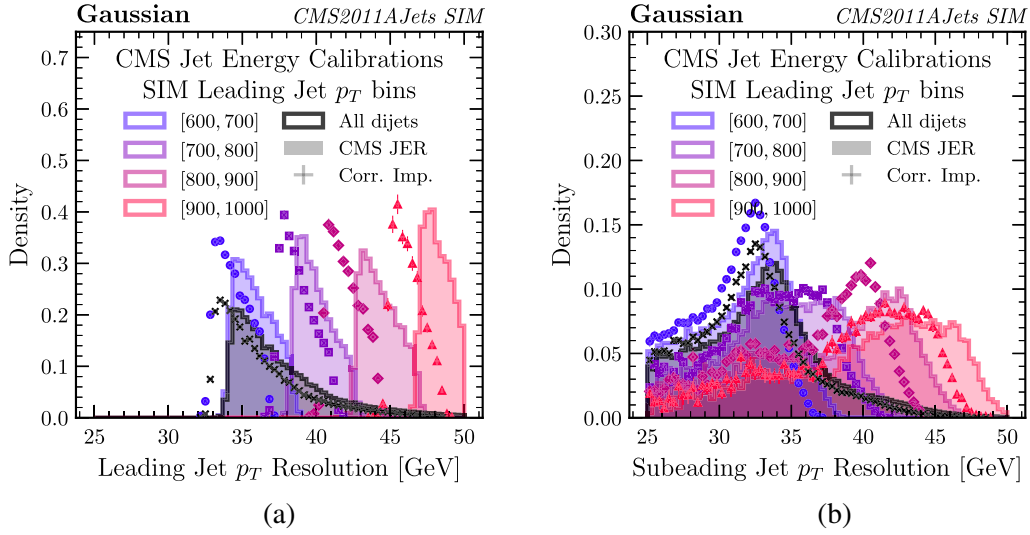


FIG. 7. Distributions of (a) leading and (b) subleading jet resolutions for each p_T bin, with the original CMS resolution (solid) and the correlation-improved (Corr. Imp.) resolution (points), for the Gaussian fits. The different p_T bins are shown in blue through red in increasing p_T , and the p_T -inclusive sample is shown in black.

In Figs. 6 and 7 we show both the original CMS jet-energy resolution [Eq. (22)] and the correlation-improved resolutions [Eqs. (17) and (12), respectively] for both leading and subleading jets. We show the correlation-improved resolutions for both the exponential fit in Fig. 6 and the Gaussian fit in Fig. 7. As with Fig. 5, events are split into bins by their leading SIM p_T , and we also include the inclusive distribution. In all cases, the

correlation-improved resolution distributions are shifted to the left of the original resolution distributions by a few GeV, indicating that the resolutions are indeed slightly improved. However, in the exponential case (Fig. 6), the correlation-improved distribution is highly bimodal; a majority of events have an unchanged resolution, whereas a small number of events (those close to the zero of the exponential, controlled by Λ_e) “snap” to the improved resolution. On the other hand,

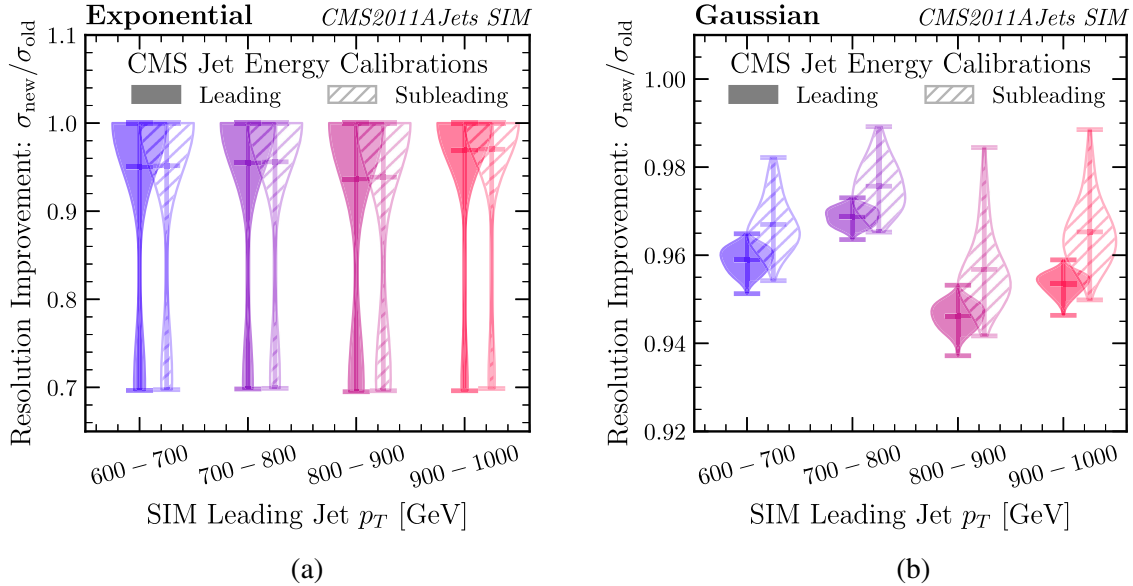


FIG. 8. Violin plots showing the distribution of ratios of the correlation-improved resolution to the CMS-based resolution for each p_T bin, for (a) the exponential fit and (b) the Gaussian fit. The leading jet ratio is shown with solid shading, and the subleading jet ratio is shown with lighter hatched shading. Note that the vertical axis scale is much larger for the exponential than the Gaussian, though the means of all distributions (indicated by the horizontal lines) are roughly the same at about 0.95. The different p_T bins are shown in blue through red in increasing p_T .

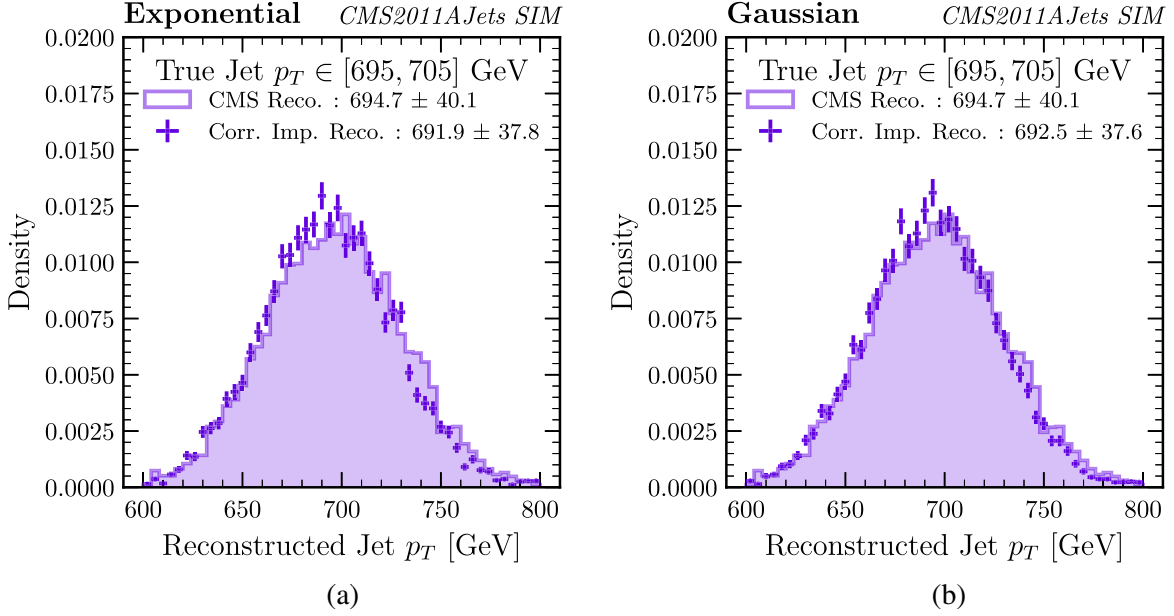


FIG. 9. Distribution of the reconstructed (Reco.) jet p_T for jets with a true (GEN) $p_T \in [695, 705]$ GeV, using CMS jet-energy corrections (solid, corresponding to $\hat{z}$) and the correlation-improved (Corr. Imp.) MLE (points, corresponding to $\hat{z}'$) for the (a) exponential and (b) Gaussian fits using Eqs. (14) and (10), respectively. Jets are included regardless of their leading or subleading label. For each distribution, we write its mean and standard deviation, the latter of which should correspond to the average resolution.

in the Gaussian case, the correlation-improved distribution is a slight shift of the original distribution, reflecting the smooth behavior of the estimates in Λ .

Next, in Fig. 8 we quantify this improvement by plotting the distribution of ratios of the new-to-old resolution distributions. We can see that the resolution is improved by an average of about 3%–5% for both subleading and leading jets in the exponential and Gaussian fits, though the distribution of improvements around the average is wildly different between the two fits. The exponential fit consists of many events with no improvement at all and few events with drastic improvement, consistent with the bimodal picture. In contrast, the Gaussian fit is far more consistent, improving most events by about the same amount. This is in accordance with the discussion in Sec. II A, where the improvement in the resolution can be approximated using Eqs. (12) and (18) for large Λ :

$$\frac{\sigma'_{\hat{z}_i}}{\sigma_{\hat{z}_i}} \approx 1 - \frac{\sigma_{\hat{z}_i}^2}{2\Lambda^2} \approx 0.95 \text{ for } \sigma_{\hat{z}_1} = \sigma_{\hat{z}_2} = 35 \text{ GeV}, \Lambda = 100 \text{ GeV}, \quad (23)$$

where the factor of $\frac{1}{2}$ comes from taking the square root of σ^2 .

It is also common to report the quadrature improvement, which is given by

$$\frac{\sqrt{\sigma_{\hat{z}_i}^{\prime 2} - \sigma_{\hat{z}_i}^2}}{\sigma_{\hat{z}_i}}. \quad (24)$$

For large Λ , this reduces to $\frac{\sigma_{\hat{z}_i}}{\Lambda} \approx 35\%$, which contains the same information as Eq. (23).

Finally, to check whether these estimates indeed reflect better resolutions for the p_T estimators, in Fig. 9 we plot the original CMS-estimated and correlation-improved reconstructed jet p_T (corresponding to $\hat{z}$ and $\hat{z}'$, respectively) for a narrow window of true $p_T \in [695, 705]$. For both the exponential and Gaussian fits, we see that the correlation-improved MLE does not induce any significant extra bias over the original estimates. We also see that the average resolution improves from 40.1 GeV to approximately 37.5 GeV, an improvement of approximately 6% of exactly the type expected by Eq. (23). Note that the highly bimodal snapping behavior of the exponential has been washed out when looking at the full distribution of jets and on average behaves qualitatively similar to the Gaussian.

We end this analysis by noting that, while the exponential is a better fit to the momentum asymmetry than the Gaussians in Fig. 5, both fits lead to consistent improved calibrations, as validated by Fig. 9. Moreover, both methods give similar final average resolutions, in spite of the very different way z' is calculated. As discussed in Sec. II B, this is expected; both fits give the same average resolution improvement for $\Lambda \approx \Lambda_e \gg \sigma$.

V. DISCUSSION: BOOSTED RESONANCES

The above discussions all revolved around the fact that the momenta of the two jets add to zero up to noise. However, this is only true in the c.m. frame of the jets, which we always assumed we were in. If the frame is

known, one can always boost to the c.m. frame, where the same analysis can be performed. However, many analyses involve the production of dijets from a boosted narrow resonance, such as a top, a Higgs boson, a Z, a $W^\pm$, or a new particle. In this case, the sum of the two jet momenta is no longer zero, but rather the momentum of the resonance, which is largely unconstrained.

If we know that the two jets originated from a boosted source of mass M ,⁸ then it is still possible to partially constrain the momenta of the two jets. Rather than summing to zero, the components of the two momenta are constrained such that

$$M^2 = m_1^2 + m_2^2 + 2p_{T,1}p_{T,2}(\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2)). \quad (25)$$

This information can be used to improve the calibration on the p_T and m of both jets, assuming η and ϕ are well known. For simplicity, suppose that we are only interested in improving just the p_T calibration and absorb the original resolution of m_1^2 and m_2^2 into M^2 . Define, for convenience, the quantity

$$\mu \equiv \frac{M^2 - m_1^2 - m_2^2}{2(\cosh(\eta_1 - \eta_2) - \cos(\phi_1 - \phi_2))}, \quad (26)$$

which is approximately the mass squared over 2, for heavy resonances. Then, the product $p_{T,1}p_{T,2}$ is approximately given by

$$L(p_{T,1}p_{T,2}) = \log(\mathcal{N}(\mu, \Lambda_\mu^2)), \quad (27)$$

where Λ_μ is the resolution on μ , including any effects due to resolutions on M^2 , $m_{1,2}^2$, $\phi_{1,2}$, and $\eta_{1,2}$. Note that, unlike the cases considered previously, this is a constraint on the *product* of the p_T 's, not the sum of the p_T 's. It is also possible, as before, to consider exponentials rather than Gaussians here.

We can now take derivatives of Eq. (7) to solve for the correlation-improved estimates and resolutions for $z_1 = p_{T,1}$ and $z_2 = p_{T,2}$. In general, this requires solving a system of coupled nonlinear equations and lacks a closed-form solution. We perform this numerically and present the resolution improvement, $\sigma'_{z_1}/\sigma_{z_1}$, in Fig. 10 for example parameter choices. We see that the resolution indeed improves; the improvement is more pronounced if μ/Λ_μ is small, with the typical $1/\sqrt{2}$ improvement factor that occurs whenever $\sigma_{z_1} \sim \sigma_{z_2} \gg \Lambda_\mu$. Note that the results in Fig. 10 are just purely numeric for example parameter choices, and there are no jets (real or simulated) here. We leave the study of these results in the setting of actual or simulated jets to potential future work.

⁸This is possible if the resonances are produced in pairs and the partner is independently tagged, e.g., a semileptonic $t\bar{t}$ decay.

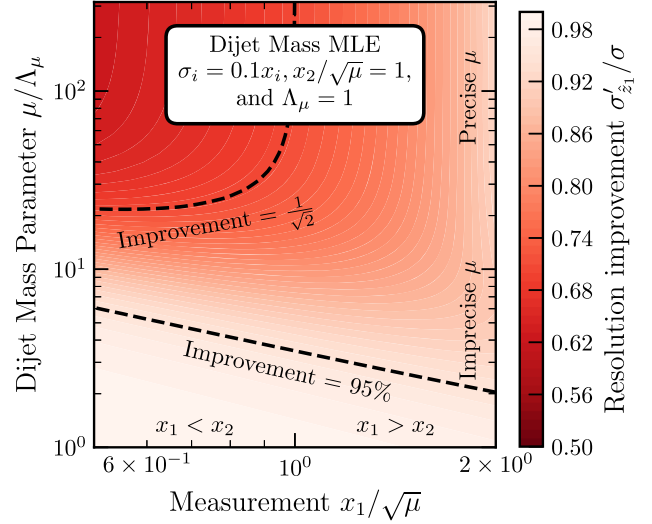


FIG. 10. Ratio of the correlation-improved resolution to the original resolution for one of the two jets, as a function of the measured p_T , $x_1/\sqrt{\mu}$, and the dijet mass parameter μ/Λ_μ as defined in Eq. (26). We fix the original jet energy resolutions σ_{z_1} and σ_{z_2} to be 10% of the nominal energy, and we also fix $x_2/\sqrt{\mu} = 1$ and Λ_μ for simplicity.

As an aside, constraining the jet p_T using momentum conservation, as we have done in Sec. IV, does *not* help to improve the dijet resolution. This is because the dijet mass is calculated using the sum of the jet momenta, which already contains all of the information about momentum conservation.

VI. CONCLUSION

In this paper, we have shown how one can take advantage of correlations between jets, namely, those induced by momentum conservation in dijet events, to obtain jet-by-jet correlation-improved MLEs and resolutions. Using this, we were able to achieve improvements in resolution over the baseline of 3%–5% (35% quadrature improvement) for simulated dijet events from the CMS detector. This is possible by only making assumptions on the relative momenta between the jets, without making assumptions on the overall prior of jet production, and does not introduce any bias as long as the jets remain unordered. This is in contrast to other proposals for improving jet-energy calibrations, which focus on the high-dimensional substructure and detector interactions of individual jets, or proposals that use momentum conservation to constrain the statistical properties of jets.

While we primarily focused on the application of correlation-improved MLE to dijet events in the c.m. frame, the procedure of applying Eqs. (1) and (2) to Eq. (7) can be expanded. For instance, we briefly discussed how one can loosen the constraint of momentum conservation in the c.m. frame to simply knowing the dijet mass in Sec. V. Other types of correlations between jets,

such as those induced by color connections [48–52], can also be used here. Additionally, we used CMS-prescribed jet energy resolutions and calibrations as our baseline, though in principle one could use any (especially machine-learned) unbiased calibration for $L(x_i|z_i)$ to source $\hat{z}(x)$ and $\sigma_{\hat{z}}(x)$ for the correlation-improved MLE, such as the Gaussian ansatz of Ref. [20]. Since we are only using out-of-jet information, correlation-improved MLE can be used on top of any type of single-jet calibration method. There is also no reason to restrict to jets; this general procedure can apply to any multiobject final state. We focused on jets because of the largest potential gains from their complex structure and relatively poor resolution; however, a similar analysis could apply, e.g., to $Z/h_1 \rightarrow \gamma\gamma$, e^+e^- , $\mu^+\mu^-$, and h_2^+, h_3^- , for identified hadrons h_i .

An important aspect of this technique is fitting the distribution of the momentum asymmetry $z_1 - z_2$ at the truth level. This requires that the simulation of the relevant effects causing the momentum asymmetry is reliable. However, because of the way prior dependence enters Eq. (7), *only* the momentum asymmetry needs to be reliable; the distribution of z_1 or z_2 themselves are not important, which is a desirable feature of calibration. Moreover, as we have shown with the Gaussian fit, it is possible to still get a consistent and valid improvement, even if the fit is suboptimal. Global event features are

therefore a robust way to add calibration information, and in the future these features can offer an approach complementary to modern ML methods to continue improving jet-energy calibrations.

The CMS Open Data used in Sec. IV was accessed in the MIT Open Data format and can be found at [53–57]. The code used for the analyses presented in this paper is located at [58], from which all of the results and plots shown here may be reproduced.

ACKNOWLEDGMENTS

We would like to thank Aviv Cukierman for useful discussions, as well as Sean Benevides and Jesse Thaler for feedback on the manuscript. B.N. is supported by the U.S. Department of Energy (DOE), Office of Science under contract DE-AC02-05CH11231. R.G. is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, [59]), and by the U.S. DOE Office of High Energy Physics under Grant No. DE-SC0012567. R.G. thanks the Galileo Galilei Institute for Theoretical Physics for their hospitality and the Istituto Nazionale di Fisica Nucleare (INFN) for support during the completion of this work.

-
- [1] The CMS Collaboration, Determination of jet energy calibration and transverse momentum resolution in CMS, *J. Instrum.* **6**, P11002 (2011).
 - [2] Vardan Khachatryan *et al.* (CMS Collaboration), Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV, *J. Instrum.* **12**, P02014 (2017).
 - [3] Morad Aaboud *et al.* (ATLAS Collaboration), Determination of jet calibration and energy resolution in proton-proton collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector, *Eur. Phys. J. C* **80**, 1104 (2020).
 - [4] Georges Aad *et al.* (ATLAS Collaboration), Jet energy scale and resolution measured in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector, *Eur. Phys. J. C* **81**, 689 (2021).
 - [5] Jet energy scale and resolution performance with 13 TeV data collected by CMS in 2016–2018 (2020), <https://cds.cern.ch/record/2715872?ln=en>.
 - [6] Georges Aad *et al.* (ATLAS Collaboration), New techniques for jet calibration with the ATLAS detector, *Eur. Phys. J. C* **83**, 761 (2023).
 - [7] Patrick T. Komiske, Eric M. Metodiev, Benjamin Nachman, and Matthew D. Schwartz, Pileup Mitigation with Machine Learning (PUMML). *J. High Energy Phys.* **12** (2017) 051.
 - [8] J. Arjona Martínez, Olmo Cerri, Maurizio Pierini, Maria Spiropulu, and Jean-Roch Vlimant, Pileup mitigation at the Large Hadron Collider with graph neural networks, *Eur. Phys. J. Plus* **134**, 333 (2019).
 - [9] ATLAS Collaboration, Generalized numerical inversion: A neural network approach to jet calibration, Report No. ATL-PHYS-PUB-2018-013, 2018.
 - [10] J. Arjona Martínez, Olmo Cerri, Maurizio Pierini, Maria Spiropulu, and Jean-Roch Vlimant, Pileup mitigation at the Large Hadron Collider with graph neural networks, *Eur. Phys. J. Plus* **134**, 333 (2019).
 - [11] Rüdiger Haake and Constantin Loizides, Machine Learning based jet momentum reconstruction in heavy-ion collisions, *Phys. Rev. C* **99**, 064904 (2019).
 - [12] Rüdiger Haake (ALICE Collaboration), Machine Learning based jet momentum reconstruction in Pb-Pb collisions measured with the ALICE detector, *Proc. Sci. EPS-HEP2019* (2020) 312 [arXiv:1909.01639].
 - [13] Albert M. Sirunyan *et al.* (CMS Collaboration), A deep neural network for simultaneous estimation of b jet energy and resolution, *Comput. Software Big Sci.* **4**, 10 (2020).
 - [14] Convolutional neural networks with event images for pileup mitigation with the ATLAS detector, Technical Report, CERN, Geneva, 2019.

- [15] Stefano Carrazza and Frédéric A. Dreyer, Jet grooming through reinforcement learning, *Phys. Rev. D* **100**, 014014 (2019).
- [16] ATLAS Collaboration, Simultaneous jet energy and mass calibrations with neural networks, Report No. ATL-PHYS-PUB-2020-001, 2020.
- [17] Baldi, Pierre, and Blecher, Lukas and Butter, Anja and Collado, Julian and Howard, Jessica N. and Keilbach, Fabian and Plehn, Tilman and Kasieczka, Gregor and Whiteson, Daniel, How to GAN higher jet resolution, *SciPost Phys.* **13**, 064 (2022).
- [18] Gregor Kasieczka, Michel Luchmann, Florian Otterpohl, and Tilman Plehn, Per-object systematics using deep-learned calibration, *SciPost Phys.* **9**, 089 (2020).
- [19] Benedikt Maier, Siddharth M. Narayanan, Gianfranco de Castro, Maxim Goncharov, Christoph Paus, and Matthias Schott, Pile-up mitigation using attention, *Mach. Learn. Sci. Tech.* **3**, 025012 (2021).
- [20] Rikab Gambhir, Benjamin Nachman, and Jesse Thaler, Learning uncertainties the frequentist way: Calibration and correlation in high energy physics, *Phys. Rev. Lett.* **129**, 082001 (2022).
- [21] Rikab Gambhir, Benjamin Nachman, and Jesse Thaler, Bias and priors in machine learning calibrations for high energy physics, *Phys. Rev. D* **106**, 036011 (2022).
- [22] Tianchun Li, Shikun Liu, Yongbin Feng, Garyfallia Paspalaki, Nhan V. Tran, Miaoyuan Liu, and Pan Li, Semi-supervised graph neural networks for pileup noise removal, *Eur. Phys. J. C* **83**, 99 (2023).
- [23] G. Angloher *et al.* (CRESST Collaboration), Towards an automated data cleaning with deep learning in CRESST, *Eur. Phys. J. Plus* **138**, 100 (2023).
- [24] Loukas Gouskos, Fabio Iemmi, Sascha Liechti, Benedikt Maier, Vinicius Mikuni, and Huilin Qu, Optimal transport for a novel event description at hadron colliders, *Phys. Rev. D* **108**, 096003 (2023).
- [25] C. H. Kim, S. Ahn, K. Y. Chae, J. Hooker, and G. V. Rogachev, Restoring original signals from pile-up using deep learning, *Nucl. Instrum. Methods Phys. Res., Sect. A* **1055**, 168492 (2023).
- [26] Kilian Lieret, Gage DeZoort, Devdoot Chatterjee, Jian Park, Siqi Miao, and Pan Li, High pileup particle tracking with object condensation, *arXiv:2312.03823*.
- [27] Daniel Holmberg, Dejan Golubovic, and Henning Kirschenmann, Jet energy calibration with deep learning as a Kubeflow pipeline, *Comput. Software Big Sci.* **7**, 9 (2023).
- [28] Albert M Sirunyan *et al.* (CMS Collaboration), A deep neural network for simultaneous estimation of b jet energy and resolution, *Comput. Software Big Sci.* **4**, 10 (2020).
- [29] Sanha Cheong, Aviv Cukierman, Benjamin Nachman, Murtaza Safdari, and Ariel Schwartzman, Parametrizing the detector response with neural networks, *J. Instrum.* **15**, P01030 (2020).
- [30] Sven Bollweg, Manuel Haußmann, Gregor Kasieczka, Michel Luchmann, Tilman Plehn, and Jennifer Thompson, Deep-learning jets with uncertainties and more, *SciPost Phys.* **8**, 006 (2020).
- [31] Marco Bellagente, Manuel Haußmann, Michel Luchmann, and Tilman Plehn, Understanding event-generation networks via uncertainties, *SciPost Phys.* **13**, 003 (2022).
- [32] Jack Y. Araz and Michael Spannowsky, Combine and conquer: Event reconstruction with Bayesian ensemble neural networks, *J. High Energy Phys.* **04** (2021) 296.
- [33] Braden Kronheim, Michelle P. Kuchera, Harrison B. Prosper, and Raghuram Ramanujan, Implicit quantile neural networks for jet simulation and correction, *arXiv:2111.11415*.
- [34] Cern open data portal, <https://opendata.cern.ch>.
- [35] Morad Aaboud *et al.* (ATLAS Collaboration), Measurement of the top quark mass in the $t\bar{t} \rightarrow \text{lepton} + \text{jets}$ channel from $\sqrt{s} = 8$ TeV ATLAS data and combination with previous results, *Eur. Phys. J. C* **79**, 290 (2019).
- [36] Armen Tumasyan *et al.* (CMS Collaboration), Measurement of the top quark mass using a profile likelihood approach with the lepton + jets final states in proton–proton collisions at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **83**, 963 (2023).
- [37] Robert Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. Ser. B* **58**, 267 (1996).
- [38] S. Chatrchyan *et al.* (CMS Collaboration), The CMS experiment at the CERN LHC, *J. Instrum.* **3**, S08004 (2008).
- [39] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands, Pythia 6.4 physics and manual, *J. High Energy Phys.* **05** (2006) 026.
- [40] S. Agostinelli *et al.*, Geant4—a simulation toolkit, *Nucl. Instrum. Methods Phys. Res. Sect. A* **506**, 250 (2003).
- [41] Patrick Komiske, Radha Mastandrea, Eric Metodiev, Preksha Naik, and Jesse Thaler, CMS 2011A Open Data | Jet Primary Dataset | $p_T > 375$ GeV | MOD HDF5 Format (2019).
- [42] Florian Beaudette (CMS Collaboration), The CMS particle flow algorithm, in *International Conference on Calorimetry for the High Energy Frontier* (2013), pp. 295–304, *arXiv:1401.8155*.
- [43] Matteo Cacciari and Gavin P. Salam, Dispelling the N^3 myth for the k_t jet-finder, *Phys. Lett. B* **641**, 57 (2006).
- [44] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, The anti- k_t jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [45] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez, FastJet user manual, *Eur. Phys. J. C* **72**, 1896 (2012).
- [46] Jet Performance in pp Collisions at 7 TeV (2010), *10.5281/zenodo.3340205*.
- [47] Renaud Boussarie *et al.*, TMD handbook, *arXiv:2304.03302*.
- [48] Jason Gallicchio and Matthew D. Schwartz, Seeing in color: Jet superstructure, *Phys. Rev. Lett.* **105**, 022001 (2010).
- [49] Victor Mukhamedovich Abazov *et al.* (D0 Collaboration), Measurement of color flow in $t\bar{t}$ events from $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV, *Phys. Rev. D* **83**, 092002 (2011).
- [50] Georges Aad *et al.* (ATLAS Collaboration), Measurement of colour flow with the jet pull angle in $t\bar{t}$ events using the ATLAS detector at $\sqrt{s} = 8$ TeV, *Phys. Lett. B* **750**, 475 (2015).
- [51] Morad Aaboud *et al.* (ATLAS Collaboration), Measurement of colour flow using jet-pull observables in $t\bar{t}$ events with the ATLAS experiment at $\sqrt{s} = 13$ TeV, *Eur. Phys. J. C* **78**, 847 (2018).

- [52] Luca Cavallini, Andrea Coccaro, Charanjit K. Khosa, Giulia Manco, Simone Marzani, Fabrizio Parodi, Daniela Rebuzzi, Alberto Rescia, and Giovanni Stagnitto, Tagging the Higgs boson decay to bottom quarks with colour-sensitive observables and the Lund jet plane, *Eur. Phys. J. C* **82**, 493 (2022).
- [53] Patrick Komiske, Radha Mastandrea, Eric Metodiev, Preksha Naik, and Jesse Thaler, CMS 2011A Simulation | Pythia 6 QCD 1000-1400 | $p_T > 375$ GeV | MOD HDF5 Format (2019), [10.5281/zenodo.3341502](https://zenodo.org/record/3341502).
- [54] Patrick Komiske, Radha Mastandrea, Eric Metodiev, Preksha Naik, and Jesse Thaler, CMS 2011A Simulation | Pythia 6 QCD 1400-1800 | $p_T > 375$ GeV | MOD HDF5 Format (2019), [10.5281/zenodo.3341770](https://zenodo.org/record/3341770).
- [55] Patrick Komiske, Radha Mastandrea, Eric Metodiev, Preksha Naik, and Jesse Thaler, CMS 2011A Simulation | Pythia 6 QCD1800-inf | $p_T > 375$ GeV | MOD HDF5 Format (2019), [10.5281/zenodo.3341772](https://zenodo.org/record/3341772).
- [56] Benjamin Nachman and Jesse Thaler, Delphes dijet dataset (2021), [10.5281/zenodo.5108967](https://zenodo.org/record/5108967).
- [57] <https://energyflow.network/docs/datasets/>.
- [58] <https://github.com/rikab/SeeingDouble>.
- [59] <http://iaifi.org/>.