This is the accepted manuscript made available via CHORUS. The article has been published as:

Jet radiation radius

Zhenyu Han

Phys. Rev. D **90**, 074033 — Published 30 October 2014

DOI: 10.1103/PhysRevD.90.074033

Jet Radiation Radius

Zhenyu Han

Institute for Theoretical Science, University of Oregon, Eugene, OR 97403, USA

Abstract

Jet radiation patterns are indispensable for the purpose of discriminating partons' with different quantum numbers. However, they are also vulnerable to various contaminations from the underlying event, pileup, and radiation of adjacent jets. In order to maximize the discrimination power, it is essential to optimize the jet radius used when analyzing the radiation patterns. We introduce the concept of jet radiation radius which quantifies how the jet radiation is distributed around the jet axis. We study the color and momentum dependence of the jet radiation radius, and discuss two applications: quark-gluon discrimination and W jet tagging. In both cases, smaller (sub)jet radii are preferred for jets with higher p_T's, albeit due to different mechanisms: the running of the QCD coupling constant and the boost to a color singlet system. A shrinking cone W jet tagging algorithm is proposed to achieve better discrimination than previous methods.

I. INTRODUCTION

It often happened in high energy physics that previously discovered particles later became the "standard candles", and led to the discoveries of new particles. For example, the W and Z gauge bosons were discovered by identifying the leptons they decay to; they (especially the Z boson) in turn played important roles in the discovery of the Higgs boson at the Large Hadron Collider (LHC) [1]. It is our hope that at the LHC and future colliders, we will be able to find new physics beyond the standard model (SM), again by first identifying particles that we are already familiar with. Therefore, it is crucial that we are able to efficiently identify all known particles, or, equivalently, their masses and quantum numbers. This is a difficult task when the particle is a QCD parton, namely a quark or a gluon, or when it decays to QCD partons. In these cases, what we see in a collider detector is one or more energetic jets that contain multiple hadrons, and their identities are concealed.

There are several handles one can utilize to distinguish jets from different origins: the jet mass or multi-jet invariant mass is determined by the original parton's mass (for convenience, in this article we use the term "parton" to refer to either a QCD parton or a massive particle that decays to QCD partons); a massive hadronically decaying particle, such as a W or Z boson, results in more than one jets with similar momenta, which is different from a usual QCD jet originated from a single quark or gluon; finally, the QCD quantum number determines the amount of jet radiation and how it is distributed. The first two handles have been used routinely in high energy experiments, while the last, though receiving significant amount of studies recently[1], has not been used as commonly. Part of the reason is, compared with the hard kinematics, the radiation information of a jet is often associated with soft particles, making it more vulnerable to various contaminations, including the underlying event, pileup, as well as radiation from nearby jets. We will focus on tackling this difficulty in this article.

Variables sensitive to jet radiation patterns are calculated from jet constituents. On the one hand, we would like to choose a larger jet radius such that more radiation from the initial parton is included to give us more color information; on the other hand, we would like to avoid introducing excessive contamination, and a smaller radius is preferred. Therefore,

[1] See Refs. [2–4] for reviews and more references.

a careful choice of the jet radius is important. Similar considerations have been made in Refs. [5–8]. In Refs. [5, 6], it is shown that the optimum jet radius is around $R = 0.7$ for jet $p_T \sim O(100$ GeV) at a hadron collider, where the effects of the underlying event and hadronization are minimized, and it yields the optimum W/Z mass measurement for the UA2 experiment. The study has been extended to the LHC in Refs. [7, 8] where the criterion for optimizing the jet radius is to obtain the best mass resolution for new massive hadronically decaying resonances. The purpose of those studies is to properly reconstruct the kinematics, *i.e.*, the momenta of the partons initiating the observed jets. Our goal is different: we want to choose the jet radius such that the discrimination power, including that from the jet radiation patterns, is maximized. For this purpose, it is important to study the distribution of the "intrinsic" radiation, *i.e.*, the radiation originated from the initial parton instead of the contaminations. In particular, it is helpful to know what is the amount of the intrinsic radiation contained in a cone of a particular size around the jet axis. Nonetheless, there is not a unique definition of the "amount" of radiation, and the radiation distribution depends on various factors such as the color configuration, the boost to the system, *etc.*.

In this article, we give a rigorous definition of the jet *radiation radius*, taking the simplest case as the start point: a dijet color singlet system in its rest frame. For such a system, the jet radiation radius, $R_F(x)$, is the jet radius one needs to include on average a fraction of x of the total amount of radiation in the leading two jets. The subscript, F, denotes a variable that quantifies the amount of radiation, such as thrust [9], girth [10, 11], or N-subjettiness [12]. This definition is similar to that of the jet shape (also known as the jet profile) [13] which measures the average fraction of the jet transverse energy inside a cone around the jet axis. Here, we have replaced the transverse momentum by a direct measure of the radiation. We study the dependence of $R_F(x)$ on the jet QCD quantum number and the jet momentum. We examine two applications, quark-gluon discrimination and W jet tagging, and show that knowledge of the jet radiation radius can help us achieve the optimum discrimination power. Interestingly, in both cases, the jet radiation radius, and thus the optimum (sub)jet radius decreases with increasing jet momentum, albeit for different reasons. For a quark or gluon jet produced from a color singlet system in its (nearly) rest frame, the jet momentum is roughly proportional to the center of mass energy, which determines the amount of radiation and how it is distributed. It is well known that the QCD coupling constant decreases when the energy scale increases, resulting in smaller radiation radius. We will see that the radiation

radius approximately follows a power law dependence on the jet momentum, with the power being around $-0.5 \sim -0.2$. Therefore, to reduce the contaminations, we are able to use a smaller cone size to evaluate jet radiation variables for higher jet momentum.

For the case of a boosted W boson (or other massive color singlet particles), the radiation scale is fixed to the W mass, while the boost makes the radiated particles collimated. This results in a radiation radius inversely proportional to the jet momentum, or subjet momentum if the W is identified as a single jet. In light of that, we propose a new W tagging algorithm to deal with high pileup. In this algorithm, we start as usual with a fat jet that contains a W boson or its QCD counterpart, and use a jet grooming method [14–16] to reconstruct the kinematics of the W decay. Then, when calculating a jet radiation variable, we use two much smaller cones around the two leading subjets' axes, with the cone sizes inversely proportional to the subjets' p_T's. With this "shrinking cone" algorithm, we find a 30% (60%) improvement in the statistical significance for jet $p_T = 300$ GeV (150 GeV), when the average number of pileup events is 60, compared with methods not using the shrinking cones.

The rest of the article is organized as follows. In Section II, we give the definition of the jet radiation radius. In Section III, we discuss the difference in the radiation radius between a quark jet and a gluon jet in a dijet color singlet system, and demonstrate that the optimum jet radius for quark-gluon discrimination decreases for increasing jet p_T. Section IV is devoted to the discussion of the radiation radius in a boosted system, and in particular its role in W jet tagging. Section V contains some discussions.

II. DEFINITION

A jet has a finite radius precisely because of the QCD radiation and hadronization. Therefore, it seems redundant to have a definition of a *radiation radius*. However, the usual purpose of using a finite radius for jet clustering is to ensure most of the jet radiation is included in the jet such that one can infer the initial partons' momenta correctly. For this purpose, the radiation gives us difficulties rather than provides us useful information. Moreover, due to collinear singularity, the momentum of a jet is concentrated in the 'core', that is, the center of the jet. Therefore, it is not essential to control precisely the jet radius, and multiple choices of jet radius coexist in high energy experiments. For the purpose

of determining the quantum number of a jet, the information from the QCD radiation is essential and one should be more careful about choosing the optimum jet radius. As mentioned in the introduction, this is particularly important when the jet is contaminated by other hadronic activities in the event, including the underlying event, pileup and radiation from nearby jets. Therefore, it is useful to define a jet radiation radius, which is a measure of how the radiation is distributed. Knowing the "intrinsic", *i.e.,* uncontaminated, jet radiation radii corresponding to partons with different quantum numbers will allow us to have a generic understanding of how to choose a jet radius to better deal with the contaminations. As we will show later, it also allows us to engineer new algorithms that provide better discrimination powers.

The first jet definition was given by Sterman and Weinberg [17]. In Ref. [17], a hadronic event from an e^+e^- collision is classified as a two jet event if at least a fraction of $1-\epsilon$ of the event's total energy is contained in two cones of opening half-angle δ, where ϵ is a small fraction. Obviously, for fixed ϵ, when we increase δ, the fraction of hadronic events that are classified as two jets, f_2, will also increase. At the next leading order, f_2 is given by [18]

$$f_2 = 1 - 8C_F\frac{\alpha_S}{2\pi}\left\{\ln\frac{1}{\delta}\left[\ln\left(\frac{1}{2\epsilon}-1\right)-\frac{3}{4}+3\epsilon\right]+\frac{\pi^2}{12}-\frac{7}{16}-\epsilon+\frac{3}{2}\epsilon^2+O(\delta^2\ln\epsilon)\right\}. \quad (1)$$

When ϵ is small, we keep only the leading term and find

$$\delta_q \sim \exp\left[-\frac{\pi(1-f_2)}{4C_F\alpha_S(s)\ln(1-\epsilon)}\right]. \quad (2)$$

In Eq. (2), we have added a subscript "q" because hadronic events in an e^+e^- machine are dominated by quark-antiquark pairs. In such a machine, the event is not contaminated by the underlying event and initial state QCD radiation[2], and two-jet events are dominantly initiated from a pair of back-to-back high energy quarks. Therefore, δ_q can be viewed as the "intrinsic" size of a quark jet[3]. From Eq. (2), we see the definition of the size δ_q depends on two parameters, ϵ and f_2. Similarly, if the system under study is initiated from two hard gluons (in the color singlet configuration), we obtain the angular size, δ_g, of a gluon jet. At the leading order, $\delta_g \sim \delta_q^{C_F/C_A}$ [18]. Therefore, δ_q or δ_g is a measure of how the radiation

[2] It is still possible to have initial state electroweak radiation.

[3] Of course, this is incorrect for special configurations. For example, a hard gluon, containing about half of the total energy, may be occasionally emitted at a large angle from both of the two quarks and classified as one of the two jets. However, in practice, this rarely happens and does not affect our studies of the properties of the quark jet.

is distributed, which is sensitive to the partons' color structure. This motivates us to give a general definition of a jet *radiation* radius, by replacing the energy fraction $1-\epsilon$ with a measure of the amount of radiation. The definition of the amount of radiation is not unique, and suitable choices include thrust (denoted T), charged particle multiplicity [4] (denoted $N_{\rm ch}$) , N-subjettiness, *etc.*

While a possible definition of the jet radiation radius could be given following the definition of the variable δ, in practice it is perhaps more convenient to use the following definition.

For an ensemble of events in a 2-jet color singlet system in the center of mass frame, we choose a variable, F, to quantify the amount of radiation in an event. Then the jet radiation radius, $R_F(x)$, is defined as the jet radius that the amount of radiation contained in the 2 leading jets, averaged over the ensemble, is a fraction of x of the average total radiation, i.e.,

$$\frac{\langle F_{R=R_F(x)}\rangle_{ensemble}}{\langle F_0\rangle_{ensemble}} = x, \tag{3}$$

where F_0 quantifies the "total" radiation including all particles in the color singlet system. Note that the "amount of radiation" is averaged over all events in the ensemble. Therefore, the radiation radius is not defined event by event. Rather, it is a measure of how the radiation is distributed statistically for a *type* of events. We have chosen to define the jet radiation radius for a color singlet system, which guarantees that the radiation pattern is not affected by the rest of the event. Otherwise the definition is ambiguous. For two partons to form a color singlet system, they are either $q\bar{q}$ or gg. Then it is reasonable to use a common jet radius for the two jets, which can be regarded as the radiation radius of a quark jet or a gluon jet[5]. It is possible to define the jet radiation radius in an N-jet ($N \geq 3$) system, which however would require a prescription on how to choose the jet radii – this is not a concern in this article because we will only discuss dijet systems. Moreover, the property of a jet can be approximately described by the QCD splitting functions [20], which is independent of the process in which the initial parton is produced. Therefore, it is our hope that the

[4] The number of particles is a more direct measure of the amount of radiation. However only the charged particle leaves a track in a collider detector, making it experimentally measurable. Due to fluctuations in the number of charged particles, charged particle multiplicity could be zero or small even when there is a large amount of radiation.

[5] Note, however, the definition of a quark or a gluon jet is ambiguous when a hard splitting occurs [10, 19].

lessons learned on quark or gluon jets from the simple dijet systems can to some extent be applied to a more sophisticated case.

As mentioned above, there are multiple choices of the variables that can quantify the amount of radiation. Event shape variables such as thrust can be directly used to quantize the radiation of an event (or the radiation in the two leading jets); while to quantify the radiation of an event using a jet shape variable, we need to first cluster the event by using a common jet radius, and then sum the jet shape variable over the two leading jets. To be a good candidate, the variable should satisfy: *a*. it is (near) zero when there is no radiation. *b*. it increases monotonously when the jet radius is increased, and approaches a finite value when the jet radius is taken to be infinity, *i.e.*, when all particles in the event (or the part of the event being studied) are included. We easily see that among the most commonly used variables, jet mass is a good radiation variable, while jet momentum is not. The aforementioned variables including $1-T$, charged particle multiplicity, N-subjettiness are also good radiation variables.

In this article, we will rely on Pythia 8 [21] simulations to calculate the jet radiation radius, leaving analytical calculations to future studies.

III. DIJETS: QUARK AND GLUON DISCRIMINATION

We start from the simplest case: dijet color-singlet systems. There are two possibilities for the initial partons, $q\bar{q}$ or gg. Given the large number of variables that have been proposed in the literature, we will only consider a few of them that are particular useful for our later discussions: charged particle multiplicity, girth and N-subjettiness. Charged particle multiplicity is simply the number of charged particles passing a certain p_T threshold, or the number of tracks in the jet. Girth is defined as [10]

$$g \equiv \sum_{i\in \text{jet}} \frac{p_T^i}{p_T^{\text{jet}}} r_i, \tag{4}$$

where r_i is the distance in the (y,ϕ) plane between particle (or jet constituent) i and the jet axis. Here, y is the rapidity. Note the variable contains a normalization factor, p_T^{jet}, which is the jet p_T. As described in the previous section, we will need to calculate the "total amount of radiation" starting from the fat jets (or, for the e^+e^- case, starting from all particles in the event), as well as the amount of radiation for a variety of smaller jet radius.

Then it is a question whether we should use the p_T of the fat jet (or half of the center of mass energy for an e^+e^- machine) as a common normalization factor for all jet radii, or use the jet p_T of the corresponding radius. The former is more natural for our definition of the jet radiation radius. However, the latter is used in the usual definition of girth, and we have found in quark-gluon discrimination, it works slightly better. Therefore we will stick to the latter definition in the rest of the paper. One of the concerns of this definition is: both the denominator and the numerator in Eq. (4) will increase as a function of the jet radius, then girth is not necessarily a monotonous function of the jet radius. However, it is straightforward to prove that if the jet axis does not change when more particles are included due to a larger radius, girth is increasing. Sometimes the jet axis does change significantly when increasing the radius, but this does not often happen and since the jet radiation radius is defined by averaging the amount of radiation over all events, this does not in practice result in a problem. The same argument applies when thrust is chosen as the radiation variable.

N-subjettiness is derived from N-jettiness [22], and defined as follows [12]. We define a distance measure between a particle k and an axis J as $\Delta R_{J,k}^{\beta}$, where $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, and β is a constant. Then for N-axes we define

$$\tilde{\tau}_N = \frac{1}{d_0} \sum_k p_{T,k} \min \left\{ \Delta R_{1,k}^{\beta}, \Delta R_{2,k}^{\beta}, \cdots, \Delta R_{N,k}^{\beta} \right\}. \tag{5}$$

Namely, for each particle in the jet, we find the nearest axis and calculate the p_T weighted distance. Then $\tilde{\tau}_N$ is a sum of the distances over all particles in the jet, normalized by a factor

$$d_0 = \sum_i p_{T,i} (R_0)^{\beta}, \tag{6}$$

where R_0 is the jet radius. Note that, to make N-subjettiness a monotonously increasing function as the jet radius increases, we should use a common R_0 for all jet radius. Finally, we find the N-axes such that $\tilde{\tau}_N$ is minimized, and the minimum value is called N-subjettiness and denoted τ_N.

When $N = 1$ and $\beta = 1$, we see that 1-subjettiness, τ_1, is defined very similar to girth, except for the normalization factor. Therefore, the discussion on girth applies to τ_1 as well. In general, these variables are examples of a set of variables called *radial moments* [11] that

are all valid radiation variables:

$$M_f = \sum_{i \in \text{jet}} \frac{p_{T,i}}{p_{T,\text{jet}}} f(r_i), \tag{7}$$

where $f(r)$ is a function of r, such as r, r^2, $r^3 \ldots$. Although it will be infrared/collinear (IRC) unsafe, we may also generalize the definition to allow a non-linear dependence on $p_{T,i}$, for example, by making the sum over $p_{T,i}^{\alpha} r_i^{\beta}$. Then the particle multiplicity can be viewed as a special case where $\alpha = \beta = 0$ (without the normalization factor of $1/p_{T,\text{jet}}$). IRC unsafe variables will be more difficult to calculate because they involve non-perturbative effects. Experimentally, they can be problematic to measure because they are sensitive to the geometry and the granularity of the detector, especially the calorimeters. An exclusion is variables defined from the tracking system, since each track corresponds to a physical particle and the variables are defined unambiguously. Sometimes IRC unsafe variables provide more information than IRC safe quantities. The particle multiplicity is such an example: it is sensitive to collinear splittings, which is important in quark-gluon discrimination.

As mentioned in the previous section, it is convenient to study events from e^+e^- collisions, where we only have final state QCD radiation emitted from the partons being studied. Therefore, we study the processes $e^+e^- \to q\bar{q}$ and $e^+e^- \to gg$. We fix the pseudorapidity of the two outgoing partons to $\eta = 0$, *i.e.*, perpendicular to the beam, and use Pythia 8 for showering and hadronization. For an e^+e^- machine, it is better to use the spherical coordinates. Nevertheless, we eventually would like to extend our results to a hadron collider, so we will use the (η, ϕ) coordinates for convenience. We then apply the anti-k_t algorithm with various R's for jet clustering, using FastJet [23]. It is known that the anti-k_t algorithm gives us circular jets, which mimic cones of size R around the jet axes. Since all contaminations are proportional to the area of the jet, we draw the radiation variables as a function of R^2 in Fig. 1 for better illustration. It is clear from Fig. 1 that the radiation variables grow much faster at smaller R^2's than larger R^2's, and eventually saturate at very large R^2's. This indicates that increasing R beyond a certain value will only introduce more contamination rather than provide more information, therefore, an optimum radius should exist.

Given Fig. 1, it is easy to read the radiation radius $R_F(x)$ by locating the radius R where $\langle N_{\text{ch}} \rangle$ or $\langle g \rangle$ is x times the saturation value. We then turn to the study of the p_T dependence of the jet radiation radius. It is well known that the jet angular size δ decreases with the jet momentum because of the RGE running of the strong coupling constant g_S. This can

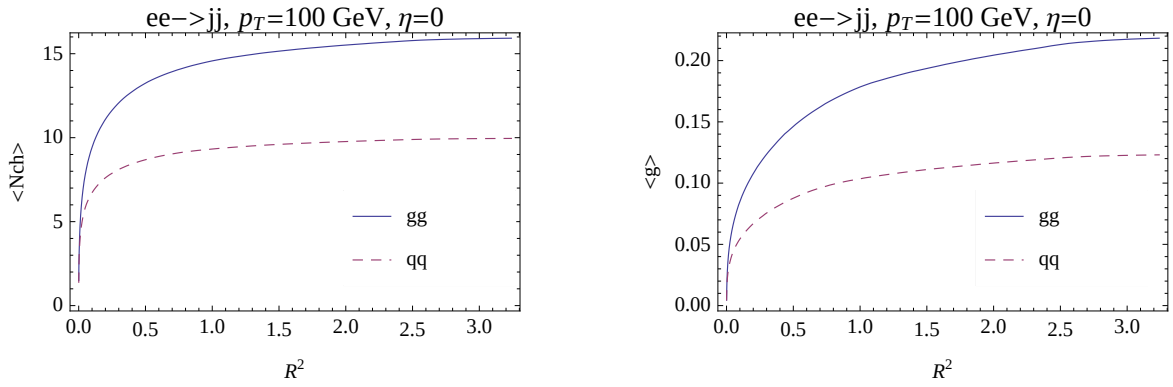

FIG. 1: The average values of charged particle multiplicity ($N_{\rm ch}$) and girth (g) as a function of R^2.

be easily seen from Eq. (2): the exponentiation combined with the logarithmic running of α_S results in a power law decrease in p_T for δ_q. Similarly, the jet radius as defined in Section II also scales with the jet momentum. In Fig. 2, we show the p_T dependence of the jet radiation radius for $F = N_{\rm ch}$ and $F = girth$. For convenience we have chosen to evaluate the saturation values at $R = 1.8$, which differs by less than 1% from the true saturation values obtained using all particles in the event. We see that the radiation radius always decreases

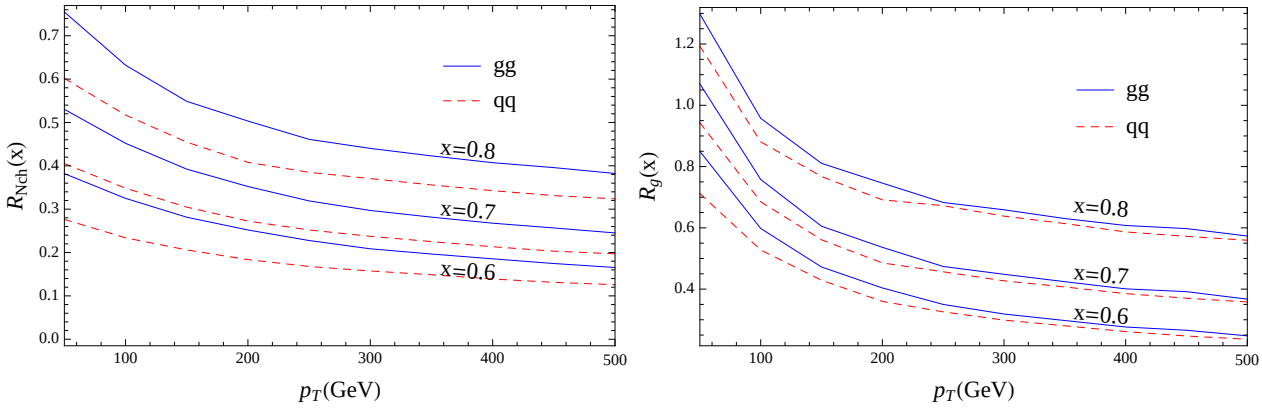

FIG. 2: The jet radiation radius $R_{N_{\rm ch}}(x)$ and $R_g(x)$ as a function of jet p_T, for x=0.6, 0.7, 0.8.

as the jet p_T increases. The dependence on p_T abides by an approximate power law with the power varying between -0.5 and -0.2 for the cases in Fig. 2. The immediate conclusion is, in the presence of contamination, if we would like to obtain the best performance for quark-gluon discrimination, we should use a shrinking radius for increasing p_T.

To study quark-gluon discrimination, we consider dijet events at the LHC[6]. We use Pythia

[6] We refer readers to Refs. [10, 11] for a comprehensive study of quark-gluon discrimination. In this article,

8 to generate $pp \to qq$ and $pp \to gg$ events, with the underlying event and initial/final state radiation turned on. We generate events with parton level p_T cut in 50 GeV p_T bins, with nominal $p_T = 100$ GeV corresponding to a $(100, 150)$ GeV parton level cut, etc. Pileup is not included. For each choice of the dijet p_T, we cluster the events with varying R using the anti-k_t algorithm and calculate the corresponding $N_{\rm ch}$ and girth for the two leading jets. Only tracks with $p_T > 0.5$ GeV are counted in $N_{\rm ch}$. The two variables are used as two separate discriminators. From Fig. 1, we see a gluon jet usually have more radiation than a quark jet, therefore, we can apply an upper cut on $N_{\rm ch}$ or girth to keep more quark jets than gluon jets. Denoting the efficiencies of quark jets and gluon jets by ε_q and ε_g respectively, we fix $\varepsilon_q = 50\%$ (for $N_{\rm ch}$, whose value is discrete, we choose ε_q to be the closest to 50% we can get), and evaluate the corresponding $\varepsilon_q/\sqrt{\varepsilon_g}$, which quantifies the improvement in the significance when quark jets are taken to be the signal and gluon jets the background. For illustration, in Fig. 3 we plot $\varepsilon_q/\sqrt{\varepsilon_g}$ corresponding to the two variables for jet $p_T = 200$ GeV. For comparison, we have also ploted the same quantity for an e^+e^- machine.

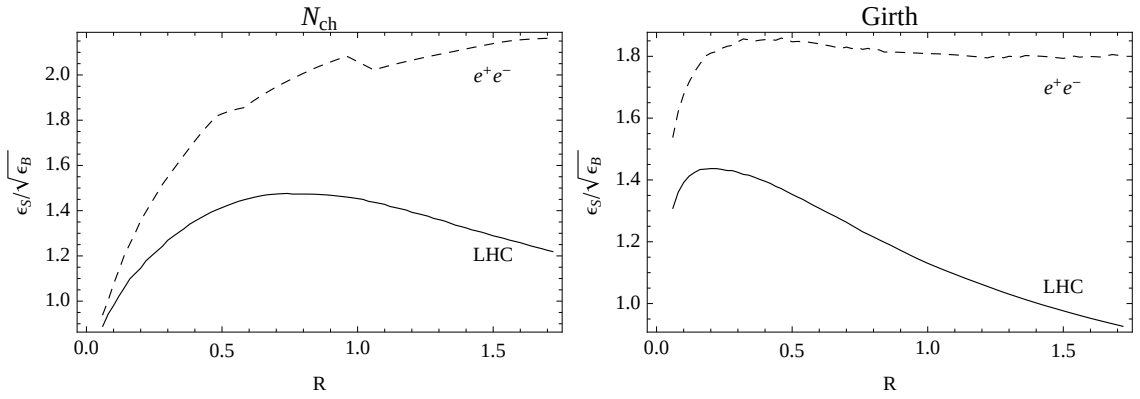

FIG. 3: The significance improvement, $\varepsilon_q/\sqrt{\varepsilon_g}$ (treating quark jets as the signal and gluon jets as the backgroud), for jet $p_T = 200$ GeV, as a function of the jet radius. The quark jet efficiency is fixed to 50% (or the value closest to 50% for $N_{\rm ch}$).

From Fig. 3, we see that without contamination, that is, at an e^+e- machine, the two variables behave similarly but also have difference: for increasing R, the significances for both variables increase rapidly at $R \lesssim 0.4$, and change slowly for $R \gtrsim 0.4$. Therefore, the distinguishing power is mostly coming from the 'core' of the jet[7]. The difference between

we focus on studying how the discrimination power depends on the jet size.

[7] A similar observation has been made in Ref. [24]

the two variable is, the significance keeps growing for $N_{\rm ch}$, and decreases slightly for girth. At a hadron collier, due to the contamination, both variables perform poorly at large R's, resulting in an optimum R where the significance is maximized. The difference between $N_{\rm ch}$ and girth remains, which causes the optimum R to be larger for $N_{\rm ch}$. More study is needed to understand the subtle difference between the two variables.

We vary the radius between 0.1 and 1.2 and find the one that maximizes $\varepsilon_q/\sqrt{\varepsilon_g}$ for jet p_T between 50 GeV and 500 GeV, which is plotted in Fig. 4. Comparing Fig. 2 and Fig. 4, we

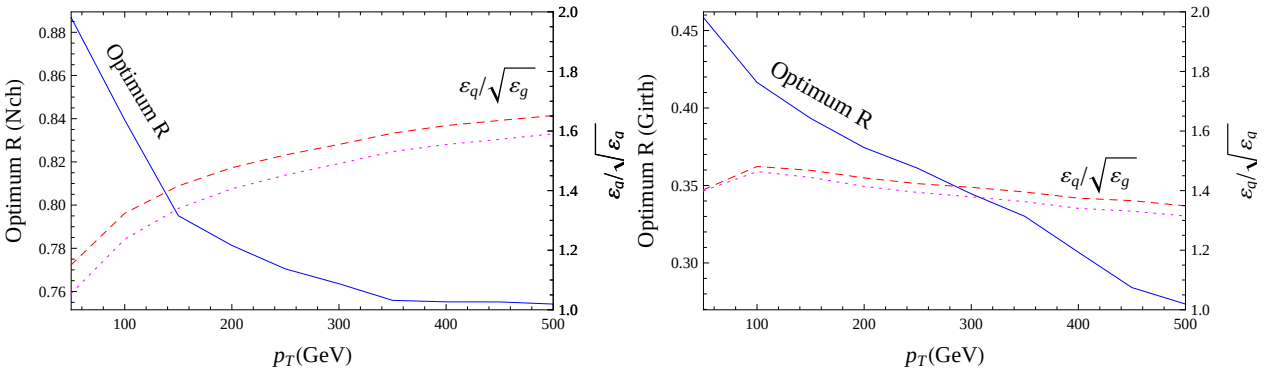

FIG. 4: The optimum jet radius for quark-gluon discrimination at the LHC (blue solid) and the corresponding significance improvement, $\varepsilon_q/\sqrt{\varepsilon_g}$ (red dashed, treating quark jets as the signal and gluon jets as the backgroud). For comparison, we also draw $\varepsilon_q/\sqrt{\varepsilon_g}$ for fixed radius, $R = 0.5$ (pink dashed). The quark jet tagging efficiency is fixed to 50% for all cases (or the value closest to 50% for $N_{\rm ch}$). Left: N_{ch}; right: girth.

see that the optimum jet radius does not correpsond to the radiation radius with a fixed x. This is because it is also affected by how the contamination is distributed. Nonetheless, the optimum radius does decrease for increasing p_T as expected. In Fig. 4, we have also given $\varepsilon_q/\sqrt{\varepsilon_g}$ for a fixed jet radius, $R = 0.5$, which is not significantly lower than the optimum values. Nevertheless, we learn that a smaller than usual (< 0.4) jet radius is preferred for a large range of jet p_T when using girth as the discriminator. This may turn out to be essential once the contamination is larger than what we have assumed, for example, when pileup is included or when we consider a busier event topology. Moreover, we also see the optimum radii are quite different for $N_{\rm ch}$ and girth, which indicates if we would like to combine the two variables to maximize the distinguishing power, we should use two (or more) jet radii.

In the above discussion, we have included all charged particles (with a p_T threshold) in the calculation of $N_{\rm ch}$, and all charged and neutral particles when calculating girth. In a collider detector, the momenta of the charged particles can be measured with the tracking system, while those of the neutral particles have to be measured with the calorimeters. The former is less sensitive to pileup and also has a better resolution than the latter, and the two pieces of information are complementary. We have seen that even without pileup, we obtain a better discrimination power by choosing the optimum jet radius. As we will show, in the presence of pileup, a proper consideration of the radiation radius is more important, especially when we are using the information from the neutral constituents of the jet. In order to see the effects of pileup, instead of expanding upon the discussion on quark-gluon discrimination, in the next section, we consider another important application – W jet tagging in a high pileup environment.

IV. W JET TAGGING

A. Jet radiation radius in a boosted system

In order to avoid ambiguities, we defined the jet radiation radius for a dijet system in its center of mass frame, such that it is natural to use a common radius for the two jets. When the system is boosted, we need a prescription to decide how to choose the jet radii for jets with different momenta. As we discussed, the radiation is mostly concentrated around the jet axis. Therefore, we see that the jet radiation radius roughly scales as $1/|\mathbf{p}|$ from the following argument.

Consider two lightlike 4-vectors, p_1 and p_2 close to the jet axis, and therefore close to each other. They are boosted to p_1' and p_2' under a common boost, and their opening angle θ ($\ll 1$) becomes θ' ($\ll 1$). Unless the boost is well aligned with the opposite direction of the jet momentum, it changes the magnitudes of the two momenta by roughly the same fraction, *i.e.*,

$$|\mathbf{p}_1'|/|\mathbf{p}_1| \approx |\mathbf{p}_2'|/|\mathbf{p}_2|.$$

Then we have

$$\begin{aligned} & p_1 \cdot p_2 = p_1' \cdot p_2' \\ \Rightarrow\ & |\mathbf{p}_1||\mathbf{p}_2|(1-\cos\theta) = |\mathbf{p}_1'||\mathbf{p}_2'|(1-\cos\theta') \\ \Rightarrow\ & \theta/\theta' \approx |\mathbf{p}_1'|/|\mathbf{p}_1|. \end{aligned}$$

Therefore, the angle between the two lightlike vectors is inversely proportional to their momenta. We extend this observation to a cone with multiple lightlike particles and draw the conclusion that the cone size shrinks when the jet is boosted[8]. Of course, the inverse proportionality is only approximate, especially for radiation at large angles and for soft hadrons whose masses cannot be ignored. However, as we will show, this does not prevent us from applying the scaling relation in a realistic situation such as the W tagging, and engineer new algorithms to achieve improvement.

For illustration, we consider the case of a hadronically decaying W. Again, we first consider an e^+e^- machine to avoid contaminations. We fix the kinematics of $e^+e^- \to W^+W^-$ such that the W's are produced at $\eta = 0$, each of which subsequently decays to two quarks with equal energy, also at $\eta = 0$. Choosing the W's momentum to be 400 GeV, we have the two partons from the W decay each carrying about 200 GeV momentum. Note that due to the large boost, the W decay products are collimated and often clustered as a single jet. In this case, the two jets from the W decay become two subjets. We then repeat the showering and hadronization and jet clustering procedure as we did for $e^+e^- \to$ dijet in Section III. Comparing with a W decaying in its center of mass frame, we should be able to obtain the same amount of radiation in the leading two jets for the boosted case, by using a (sub)jet radius that is 1/5 of the center-of-mass case. In Fig. 5, we plot $\langle N_{\rm ch}\rangle$ as a function of the (sub)jet radius for the two cases, which to a good approximation confirms our expectation. Thus, we can extend our definition of the jet radiation radius to a boosted color singlet system, which should be inversely proportional to the boost.

Consider the radiation radius for two variables, N_{ch} and τ_2. We again fix the momentum configuration such that the two partons from the W decay are located at $\eta = 0$ and have

[8] We note that this observation is also the motivation of the variable R algorithm proposed in Ref. [25], where jets originated from new (colored) resonance decays are studied. We emphasize that the inverse proportionality is a good approximation only for color singlet resonance decays because a colored resonance will also radiate *before* it decays.

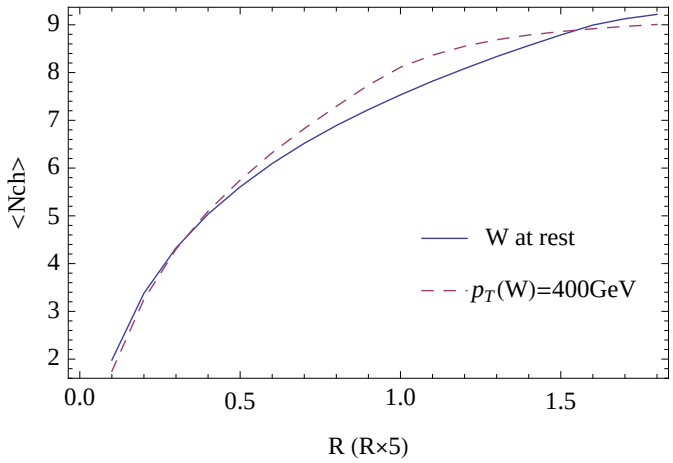

FIG. 5: The average number of tracks as a function of the (sub)jet radius for a W at rest and a boosted W. The x axis is the jet radius for a W at rest, and 5 times the subjet radius for a boosted W. We set the two subjets' momenta to be equal in magnitude such that we can use the same subjet radius. The number shown is summed over the two leading (sub)jets.

the same p_T, which allows us to use the same subjet radius for the two subjets. Since the kinematics is fixed at the parton level, we avoid a jet finding procedure by choosing the subjet axes to be those given by the two partons from the W decay. Then for a given subjet radius R, we simply count the number of charged particles in two cones of size R around the two axes to obtain N_{ch}; similarly, we use Eq. (5) to calculate τ_2 using the particles within the two cones, without applying a minimization procedure for the subjet axes. The (sub)jet radidation radius is plotted in Fig. 6 as a function of the W momentum. From Fig. 6, we see that the dependence of the radiation radius on $1/p_{T,W}$ is almost strictly linear when $p_{T,W} > 100$ GeV, and the two variables give us very similar results: averagely 70% of the radiation is contained in two cones of size $R \sim 0.5$ when $p_{T,W} = 100$ GeV, and the same amount of radiation is contained in much smaller cones of size $R \sim 0.1$ when $p_{T,W} = 500$ GeV.

We then turn to discuss the QCD background. In W jet tagging, after applying a jet grooming algorithm and a W mass window cut, the remaining background will be QCD jets that kinematically resemble a boosted W decay. Besides the case that two unrelated jets accidentally merge to one fat jet, most of the background jets come from relatively hard QCD splittings. For these QCD jets, the remaining handle is the difference in the radiation patterns. We have just shown that for a boosted W, a smaller jet radius is needed

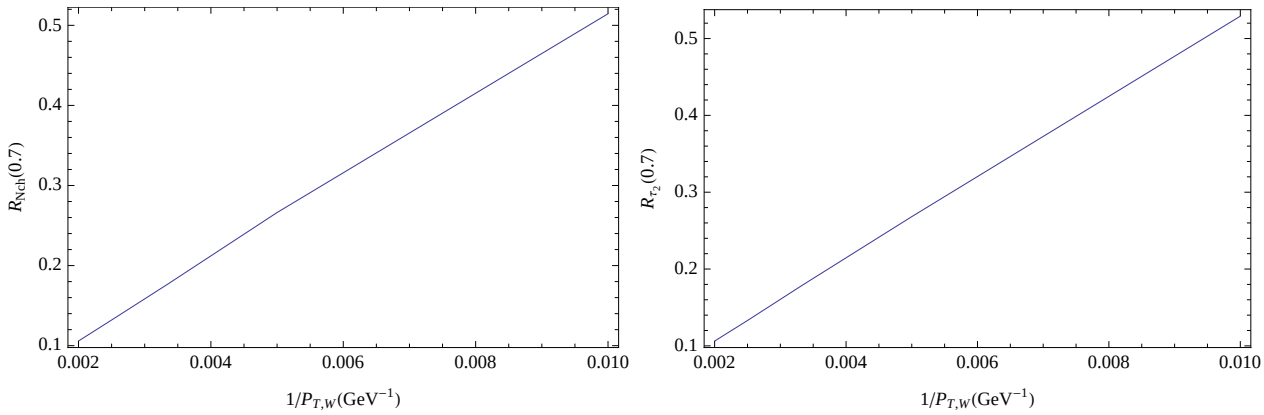

FIG. 6: The (sub)jet radiation radius for a boosted W as a function of $1/p_{T,W}$. The momenta of the two partons from the W decay are set to be equal (at $\eta = 0$), such that the two (sub)jets share the same radiation radius.

to include most of the radiation in the jet. Then we will need to know if it is the case for a QCD jet with a similar kinematic configuration. For this purpose, we consider the process $e^+e^- \to \bar{q}qg$, with q and g having exactly the same kinematic configuration as the boosted W decay discussed above. Therefore, if we use a large R to cluster the event, the qg pair will be clustered as a single fat jet with two hard subjets. We will call this jet a 2-prong QCD jet. Again, we shower and hadronize the event, and use various smaller R's to cluster the events. We count the number of charged particles in the two subjets that correspond to the qg pair, which is compared to the W jet (Fig. 7). From Fig. 7, we clearly see that a 2-prong QCD jet tends to have more radiation than a W jet. Also, we see the number of tracks in the QCD jet does not, but nearly saturates at $R = 0.35$, where the W jet is almost saturated. It is conceivable that we do not gain more discrimination power by going to larger R's, because of the increasing contaminations. While it is as expected that the radiation is mostly contained in small cones of $R \sim 0.3$ for a boosted W, it is not obvious for a 2-prong QCD jet – we see in Fig. 1 that N_{ch} does not saturate for a generic QCD jet even at $R \sim 1.0$. Qualitatively, this can be understood using the dipole language: the QCD splitting $q \to qg$ creates a color singlet dipole. Since we have fixed the kinematics to be the same as a boosted W, this dipole also has a small energy scale and it behaves very similar to a boosted W. Therefore, the radiation from this dipole is also confined in small cones. Another color dipole exists from the initial $e^+e^- \to \bar{q}q$ production, which connects the 2-prong QCD jet to the jet in the opposite hemisphere. The radiation from this dipole

is not confined, but only contributes to a small fraction of the radiation of the 2-prong jet. Thus, we find most of the radiation is contained in two small cones. Of course the above explanation is crude and theoretically it is very interesting to study the jet radiation distribution for special kinematic configurations.

In the above example, we have taken the two partons from the boosted W decay to have the same momenta, which allows us to use the same cone size for jet clustering. It is of course not the case for a generic W decay, for which we should use two different cone sizes for the two subjets, inversely proportional to their p_T's. As we will show, this motivates us to design a new, improved W tagging method, using different cone sizes when evaluating the radiation variables.

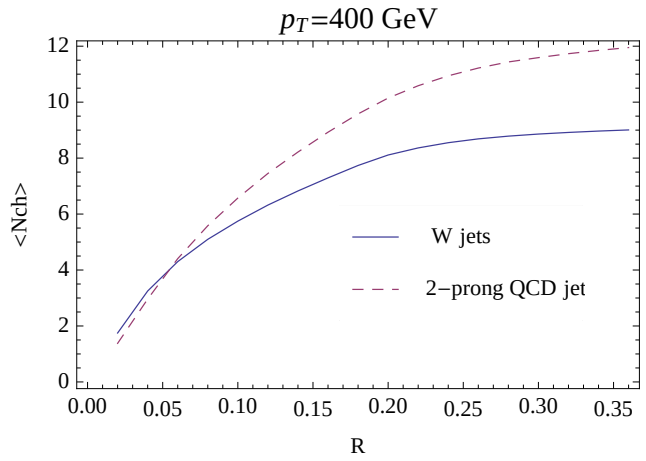

FIG. 7: The average number of tracks for a boosted W and a 2-prong QCD jet as a function of the subjet radius. We have fixed the kinematic configurations such that the two partons from the W decay have the same p_T and the QCD splitting has exactly the same kinematics as the W decay. The number shown is the sum of the two leading subjets.

B. W tagging with pileup

Besides the fact that W has a fixed mass, a boosted W differs from a QCD (quark or gluon) jet in two other aspects. First, a W jet contains two hard subjets with balanced momenta, while a QCD jet more often has only one hard subjet. Second, the W boson is a color singlet particle which has a different radiation pattern from a QCD jet. A jet grooming method [14–16] is efficient for exploiting the first difference, while after grooming we can

further study variables sensitive to the radiation patterns. In Refs. [26, 27], we showed that the two pieces of information should be combined to obtain the optimum discrimination power. In those studies, the contamination from initial state radiation and the underlying event is included and they do not significantly affect the discrimination power. On the other hand, pileup, by which we mean multiple collisions in a beam crossing, may become the main obstacle to W tagging. In particular, it has significant impact on the efficiency of the radiation variables. We illustrate this by considering two variables: the jet mass after the filtering/mass drop (MD) procedure, $m_{\rm filt}$, defined in Ref. [14], and the N-subjettiness ratio, $\tau_{21} \equiv \tau_2/\tau_1$ [12].

As shown in Ref. [27], after the filtering/MD procedure, the variable τ_{21} becomes an efficient variable for measuring the amount of radiation, and it has small correlation with $m_{\rm filt}$. Therefore, we adopt a two step cut-and-count method for W-tagging, cutting on $m_{\rm filt}$ first and then on τ_{21}. The signal (W-jets) efficiencies and background (QCD jets) mistag rates in the two steps are denoted $\varepsilon_S(m_{\rm filt}), \varepsilon_B(m_{\rm filt})$ and $\varepsilon_S(\tau_{21}), \varepsilon_B(\tau_{21})$. The final efficiencies are the products of those in the two steps, for example, $\varepsilon_S({\rm final}) = \varepsilon_S(m_{\rm filt}) \cdot \varepsilon_S(\tau_{21})$. Given the efficiencies, we can quantify the change in the significance by $\varepsilon_S/\sqrt{\varepsilon_B}$, *i.e.*, we achieve an improvement when $\varepsilon_S/\sqrt{\varepsilon_B} > 1$.

We use Pythia 8 to simulate all-hadronic WW's as our signal events and QCD dijets as the background. To simulate pileup, we turn on all soft QCD processes in Pythia 8, and add them on top of each signal/background event. The number of pileup events follows a Poisson distribution with an expectation value of $\langle N_{\rm pu}\rangle$. We then find jets using FastJet [23] with the anti-k_t algorithm ($R = 1.0$). The two leading jets in each event are included in the following analysis.

In Fig. 8, we show the effect of pileup events when $\langle N_{\rm pu}\rangle = 60$, for jet $p_T = 300$ GeV[9]. We see that without pileup, the filtering/MD method is efficient to reconstruct the W mass. We apply a mass window cut $(60, 100)$ GeV [10] on $m_{\rm filt}$ and obtain a gain in the significance, $\varepsilon_S(m_{\rm filt})/\sqrt{\varepsilon_B(m_{\rm filt})} = 1.78$ (see Table I). The τ_{21} distribution after the $m_{\rm filt}$ cut is shown

[9] We apply a cut $p_T > 300$ GeV at parton level in Pythia 8. Due to PDF suppressions, the events are dominated by jets with p_T close to 300 GeV. We take the leading two jets after jet clustering and do not further apply any p_T cut at the jet level.

[10] This is not the best mass window based on Fig. 8. We have used a relatively wider mass window to take into account possible experimental smearing to the reconstructed mass.

in Fig. 8. Since all jets passing the filtering/MD procedure are required to have two hard subjets, we see Fig. 8 confirms our observation in the previous subsection that a 2-prong QCD jet tends to have more radiation than a W jet. We then impose an upper cut on τ_{21}, and obtain $\varepsilon_B(\tau_{21}) = 0.10$ at $\varepsilon_S(\tau_{21}) = 0.5$ – an improvement in the significance of 1.58. This number is very close to the maximum $\varepsilon_S(\tau_{21})/\sqrt{\varepsilon_B(\tau_{21})}$ we can get, 1.61, by scanning the τ_{21} cut, which occurs at $\varepsilon_S(\tau_{21}) = 0.38$ and $\varepsilon_B(\tau_{21}) = 0.055$.

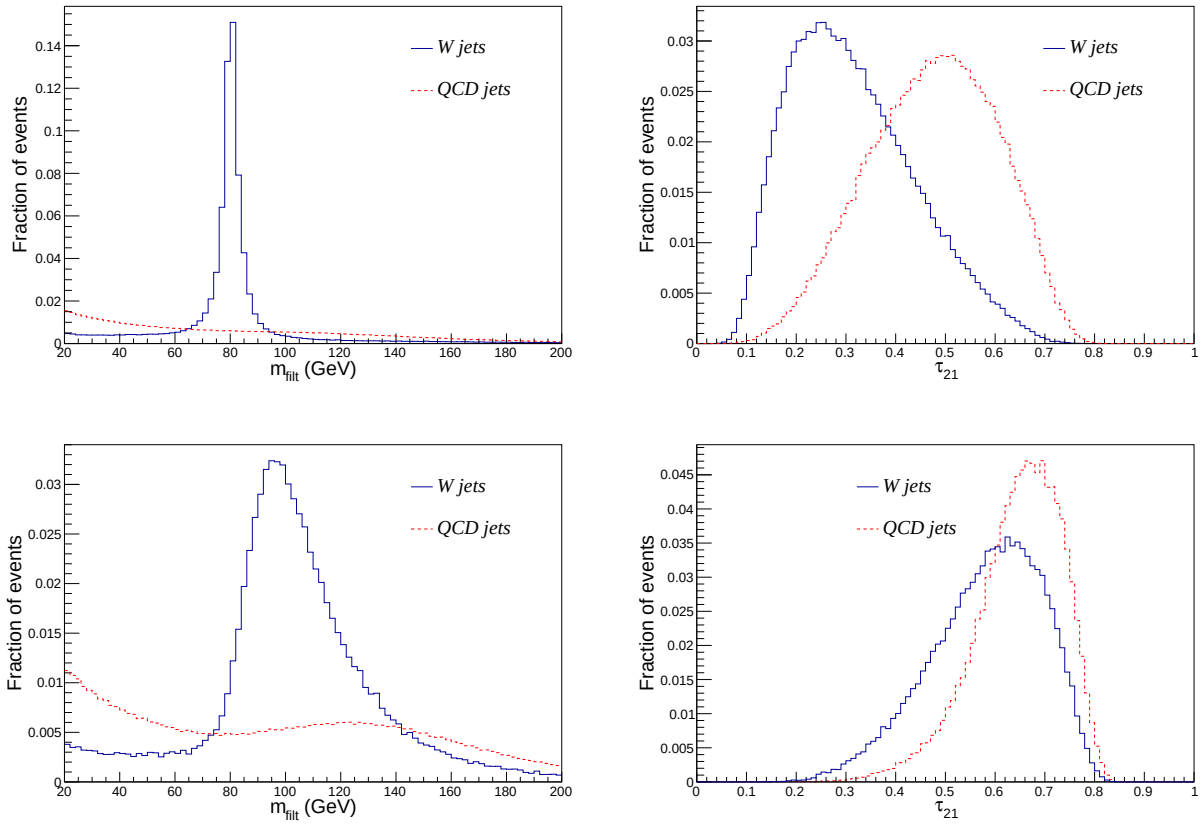

FIG. 8: The variables m_{filt} and τ_{21} before (top) and after (bottom) including the pileup events.

After including the pileup, we see the filtered mass is shifted to larger values and the mass peak is broader. Choosing the mass window to be $(80, 120)$ GeV, we obtain a signal efficiency of 0.48, and an increase in $S/\sqrt{B}$ of 1.47 which is smaller than the case without pileup. Moreover, we obtain almost no improvement by further using τ_{21}: by scanning the cut on τ_{21}, the biggest significance improvement is 1.02 for $\varepsilon_S(\tau_{21}) = 0.89$ and $\varepsilon_B(\tau_{21}) = 0.76$.

In Ref. [28], a method for pileup subtraction is proposed. In this method, one first obtains the pileup p_T and mass densities for a given event by dividing the event to patches and taking the medians. Then for a generic infrared and collinear safe jet shape variable, one finds its

sensitivity to pileup and extrapolate to its zero pileup value using the obtained densities. Applying the method on $m_{\rm filt}$ for the W jets and the QCD dijets, we find the mass peak is largely restored to its original position (Fig. 9). Using the mass window $(60, 100)$ GeV, we increase the significance by a factor of 1.51. On the other hand, applying the same subtraction to τ_{21}, we see limited improvement. The best $\varepsilon_S(\tau_{21}^{\rm subtr})/\sqrt{\varepsilon_B(\tau_{21}^{\rm subtr})}$ we can get is 1.08. This manifests that, compared with the kinematics, the radiation information is much more vulnerable to soft contaminations. Therefore, a careful consideration of the jet radiation radius is essential, which is the subject of the next subsection.

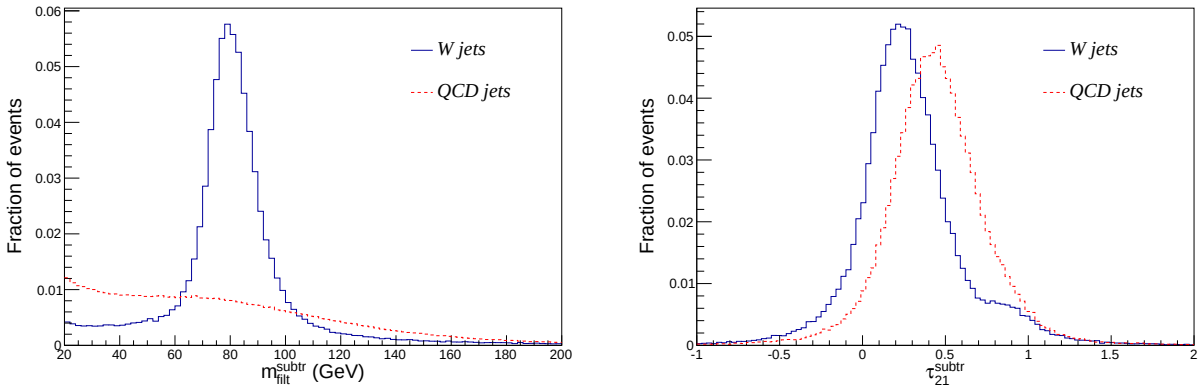

FIG. 9: The variables $m_{\rm filt}$ and τ_{21} after the pileup subtraction method described in Ref. [28].

C. The shrinking cone algorithm for W jet tagging

In Section III, we found that to optimize quark-gluon discrimination, we should use a smaller radius for a larger p_T. There, the reason for using a shrinking radius is the decrease of α_S for increasing momentum, and the dependence on p_T abides by a soft power law. A boosted W is similar except that the radiation radius is $\propto p_T^{-1}$, *i.e.*, a much stronger dependence on p_T. Moreover, W tagging is slightly more complicated because two partons with usually different momenta are produced from the W decay: first, we need to use two different radii for the two (sub)jets from the W decay. Second, in practice, in order to cluster as many as W's to a single jet, we often start with a large R and later apply the filtering/MD procedure to identify the two subjets. This may be unnecessary because one can always use a smaller R from the beginning, but it does provide us a universal and convenient procedure for a large range of p_T's. For these reasons, we are motivated to adopt

the following procedure:

1. Use a large jet radius to cluster as many as W's to single jets.

2. Apply the filtering/MD procedure to identify the two leading subjets.

3. Collect jet constituents that are within two cones, each of which around one of the two subjet axes. The cone size is determined by the subjet's p_T,

$$R_{\rm sub} = R_{\rm ref}(100\ {\rm GeV})\frac{100\ {\rm GeV}}{p_{T,{\rm sub}}} \tag{8}$$

   where $R_{\rm ref}(100\ {\rm GeV})$ is a reference radius at $p_T = 100$ GeV. In order to avoid excessively large cone sizes when one of the subjets has a small p_T, $R_{\rm sub}$ is capped at 0.7.

4. Use the jet constituents obtained in Step 3 to calculate jet radiation variables, and combine it with the (subtracted) filtered mass to get better discrimination.

Since the key ingredient in this procedure is in Step 3, we will call this method the "shrinking cone" (SC) algorithm. Note that the filtering/mass drop method also provides us subjets with a radius depending on the W momentum. However, unlike our shrinking cone algorithm, the subjets share the same radius, and there is not a tunable parameter that one can adjust according to the level of pileup. The N-subjettiness ratio, τ_2/τ_1, calculated using the shrinking cone algorithm is denoted $\tau_{21}^{\rm SC}$. Choosing $R_{\rm ref}(100\ {\rm GeV}) = 0.2$, we show the $\tau_{21}^{\rm SC}$ distributions for our W jets and QCD jets in Fig. 10. Similar to the previous subsection, we have applied a fixed mass window cut, $(60, 100)$ GeV, on the subtracted filtered mass, and only included jets passing this cut in Fig. 10. Note that we have not included a lower bound on the cone size in Step 3. For W $p_T = 300$ GeV as we are considering, the cone size can go below the angular resolutions (~ 0.1) of the hadronic calorimeter of the CMS or the ATLAS detectors, which may be remedied by including the information from the tracking system and the electromagnetic calorimeter as in the particle flow approach [29]. For W's of higher momemta, a lower bound may be necessary. On the other hand, the results from our particle level simulations may shed light on the design goals of future collider detectors.

Comparing with the $\tau_{21}^{\rm subtr}$ distributions in Fig. 9, we see a better distinction between W jets and QCD jets. Applying a cut on $\tau_{21}^{\rm SC}$ such that $\varepsilon_S(\tau_{21}^{\rm SC}) = 0.5$, we obtain $\varepsilon_B(\tau_{21}^{\rm SC}) = 0.10$ and $\varepsilon_S(\tau_{21}^{\rm SC})/\sqrt{\varepsilon_B(\tau_{21}^{\rm SC})} = 1.45$. As expected, we are unable to achieve the original

performance of this variable in the zero pileup case. However, it does contribute to the discrimination power almost as much as the filtering/MD procedure. One may also wonder whether applying the pileup subtraction procedure will improve further the performance. We have tested it and found no improvement.

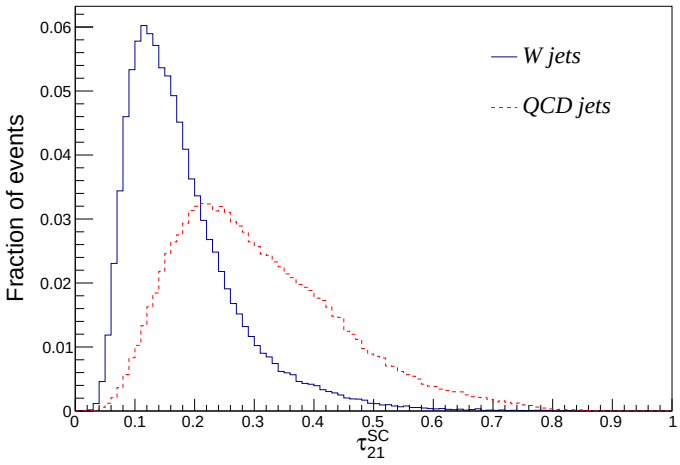

FIG. 10: The variable τ_{21} using the shrinking cone algorithm, $R_{\rm ref}(100\ {\rm GeV}) = 0.2$.

The signal and background efficiencies for various combinations of mass variables and N-subjettiness variables are shown in Table I.

A complete comparison between $\tau_{21}^{\rm SC}$ and $\tau_{21}^{\rm subtr}$ is given in Fig. 11, where we plot the background fake rate as a function of the signal efficiency. When making the plots, we have again fixed the mass window cut as 60 GeV $< m_{\rm filt}^{\rm subtr} <$ 100 GeV, which gives us $(\varepsilon_S, \varepsilon_B) = (0.60, 0.16)$ as the maximum values at the top-right conner. Then we scan the cut on $\tau_{21}^{\rm SC}$ and $\tau_{21}^{\rm subtr}$ to produce the curves. It is seen that $\tau_{21}^{\rm SC}$ is a better variable for all ε_S's.

In Fig. 11, we have also compared the performances of different choices of the reference radius: $R_{\rm ref}(100\ {\rm GeV}) = 0.1, 0.2, 0.3, 0.4$. It turns out for most of the signal efficiencies, $R_{\rm ref}(100\ {\rm GeV}) = 0.2$ is preferred. Nonetheless, as long as $R_{\rm ref}$ is not too large, the performance does not degrade significantly. This leaves room for practical cases where a very small radius is not viable, for example, when information from the hadronic calorimeter alone is used. It is also interesting to see if shrinking cones are better than cones of a fixed size. To see that, after obtaining the two subjets from the filtering/MD procedure, we use two cones of the same size to evaluate τ_{21}, independent of the subjets' p_T. It turns out

		$\langle N_{\rm pu}\rangle = 0$	$\langle N_{\rm pu}\rangle = 60$		
Filtering/MD		$m_{\rm filt} \in (60, 100)$	$m_{\rm filt} \in (80, 120)$	$m_{\rm filt}^{\rm subtr} \in (60, 100)$	$m_{\rm filt}^{\rm subtr} \in (60, 100)$
	ε_S	0.64	0.48	0.60	0.60
	ε_B	0.12	0.11	0.16	0.16
	$\varepsilon_S/\sqrt{\varepsilon_B}$	1.84	1.47	1.51	1.51
τ_{21}		τ_{21}	τ_{21}	$\tau_{21}^{\rm subtr}$	$\tau_{21}^{\rm SC}(R_{\rm ref} = 0.2)$
	ε_S	0.50 (0.38)	0.50 (0.89)	0.50 (0.66)	0.50 (0.47)
	ε_B	0.10 (0.054)8	0.25 (0.76)	0.23 (0.37)	0.12 (0.10)
	$\varepsilon_S/\sqrt{\varepsilon_B}$	1.58 (1.61)	1.00 (1.02)	1.04 (1.08)	1.45 (1.45)
Total	ε_S	0.32 (0.24)	0.24 (0.43)	0.30 (0.39)	0.30 (0.28)
	ε_B	0.012 (0.0067)	0.027 (0.082)	0.037 (0.058)	0.019 (0.016)
	$\varepsilon_S/\sqrt{\varepsilon_B}$	2.92 (2.93)	1.46 (1.49)	1.57 (1.63)	2.18 (2.18)

TABLE I: Signal and background efficiencies and the improvement in $S/\sqrt{B}$. In the first step, "filtering/MD", we use fixed mass window cuts as shown in the second row. In the second step, we further cut on the N-subjettiness ratio for events within the mass window, and present the results for two choices of ε_S: 1. fixed $\varepsilon_S = 0.5$; 2. (in the parentheses) ε_S maximizing $\varepsilon_S/\sqrt{\varepsilon_B}$. The last group of numbers denoted "Total" are the products of the two steps, for example, $\varepsilon_S(\text{total}) = \varepsilon_S(m_{\rm filt}) \cdot \varepsilon_S(\tau_{21})$.

by using a cone size of 0.4 (0.7), we get $\varepsilon_S(\tau_{21})/\sqrt{\varepsilon_B(\tau_{21})} = 1.39$ (1.21) at $\varepsilon_S(\tau_{21}) = 0.5$. Comparing with the number from shrinking cones, 1.45, we see $R = 0.4$ is almost as good, while the performance degrades significantly for $R = 0.7$. If the pileup level is higher than that assumed in this article, the preferred jet radius will fall below those adopted by the LHC collaborations.

In the above discussions, we have used $R = 1.0$ to obtain the initial fat jet. If the W p_T is smaller, we will need a larger jet radius to cluster the W decay products to a single jet[11]. The contamination is even bigger because the jet area scales as R^2. In the following, we consider W's with $p_T = 150$ GeV, clustered with $R = 1.5$. The filtered mass after subtraction is given

[11] Alternatively, one may start with slim jets and find W's by pairing jets with invariant masses close to the W mass. Our method can be easily adapted accordingly.

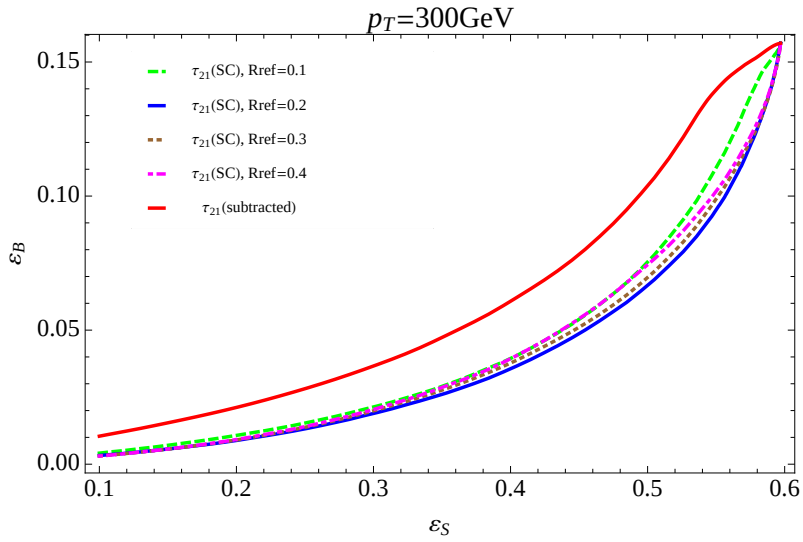

FIG. 11: Signal efficiency versus background fake rate for jet $p_T = 300$ GeV.

in the left panel of Fig. 12, for both W jets and QCD jets. We see that, due to the larger jet radius, the filtering/MD procedure leaves a much larger portion of the background jets in the W mass window: by choosing 60 GeV $< m_{\rm filt}^{\rm subtr} <$ 100 GeV, we obtain $\varepsilon_S(m_{\rm filt}^{\rm subtr}) = 0.54$ and $\varepsilon_B(m_{\rm filt}^{\rm subtr}) = 0.31$, which gives us no increase in $S/\sqrt{B}$. Similarly, applying the pileup subtraction method on $\tau_{21}^{\rm subtr}$ does not provide any improvement either. On the other hand, the τ_{21}^{SC} variable are still efficient for separating the signal from the background, as shown in Fig 12 (b). Choosing $R_{\rm ref}(100\ {\rm GeV}) = 0.2$, we have $\varepsilon_B(\tau_{21}^{\rm SC}) = 0.096$ at $\varepsilon_S(\tau_{21}^{\rm SC}) = 0.5$, yielding $\varepsilon_S(\tau_{21}^{\rm SC})/\sqrt{\varepsilon_B(\tau_{21}^{\rm SC})} = 1.62$. This has made jet radiation patterns the most important handle for tagging semi-boosted W's.

Similar to Fig. 11, we plot the $\varepsilon_S \sim \varepsilon_B$ curves for several choices of $R_{\rm ref}$ in Fig.13, for jet $p_T = 150$ GeV. It turns out $R_{\rm ref}(100\ {\rm GeV}) = 0.2$ is still the best choice among the values being considered. The difference between $p_T = 150$ GeV and $p_T = 300$ GeV is, the performance for $p_T = 150$ GeV is getting worse faster when we increase $R_{\rm ref}$. This is because for the same $R_{\rm ref}$, the actual cone size is larger for lower p_T, and thus more contamination from pileup is included.

V. DISCUSSIONS

In this article, we have given a definition of the jet radiation radius, which quantifies the size of a jet due to its QCD radiation. This definition is closely related to the jet shape

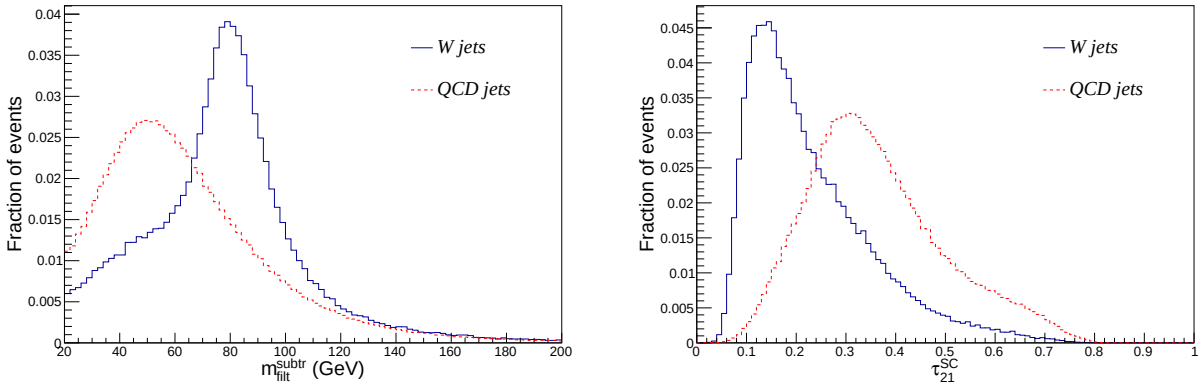

FIG. 12: The filtered mass after subtraction and τ_{21} using the the shrinking cone algorithm, for jet $p_T = 150$ GeV.

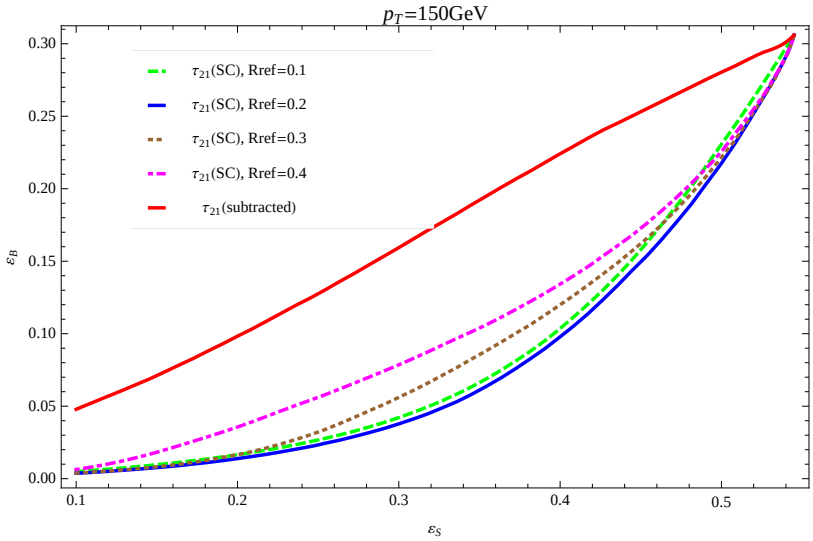

FIG. 13: Signal efficiency versus background fake rate for jet $p_T = 150$ GeV.

[13] variable which measures, on average, the fraction of transverse momentum that is included in a cone of size R around the jet axis. For the purpose of studying the jet radiation distribution, momentum is not a good measure because a large (small) momentum does not correspond to large (small) amount of radiation. Therefore, in our definition, we have replaced it with variables that directly measure the amount of radiation. Moreover, we have emphasized the nature of the radiation radius being *intrinsic*, *i.e.*, it is a characteristic of a QCD parton in a particular color configuration. In particular, it is defined before various contaminations and experimental effects are included. It is our hope that by "factorizing" the contributions to jet shape variables into intrinsic and environmental ones, we can sim-

plify jet substructure studies and use jet radiation patterns more efficiently to distinguish jets with different quantum numbers. To avoid ambiguity, we have defined jet radiation radius for the simplest system, dijet in the color singlet configuration, which already finds applications in quark/gluon discrimination and W jet tagging. Quantitatively the radiation radius corresponding to a parton depends on the color configuration of the whole event. Nevertheless, the main features of quark or gluon jets discussed in this paper, namely, the concentration of radiation around the jet axis and its p_T dependence, can be well described by the QCD splitting functions [20]. Therefore, we only expect small differences in jet radiation radius for a more sophisticated color configurations and the methods discussed in this article can be applied with suitable modifications.

A key observation in this article is: in order to efficiently use jet radiation variables, a smaller than usual jet radius is often preferred. This is particularly true for the W jet tagging method we proposed, where shrinking cones are used when calculating jet radiation variables. In a high energy experiment, either a simpler or a more complicated method may be adopted. On the one hand, due to the limitation in granularities, especially those of the hadronic calorimeter, one may not be able to use a jet radius smaller than $O(0.1)$, or/and may not be able to use continuous jet radii in the calculation. In that case, we may choose to simplify the method by choosing a few typical, but small, cone sizes. It is shown in Fig. 11 and Fig. 13 that increasing the jet radius within a sizable region will not significantly hurt the discrimination power. On the other hand, one does see the advantages of using small cones, for example, $R_{\rm sub} = 0.2$ is preferred for subjet $p_T \sim 100$ GeV when the average number of pileup events is 60. An even smaller radius might be preferred if the pileup level is higher. Therefore, ideally we would want to use the finest granularity for jet constituents, including the information from the electromagnetic calorimeter and the tracking system as in the particle flow approach [29]. Moreover, we have sticked to a single choice of (sub)jet radius for each (sub)jet in this article. Similar to Ref. [26], one may benefit from using two or more radii for each (sub)jet, which not only gives us the information of how much the radiation is, but also captures how it grows with increasing R.

The shrinking cone algorithm we proposed for W tagging is parallel and complementary to other pileup reduction methods and may be combined to obtain the optimum results. We have already used it with the pileup subtraction method proposed in Ref. [28], where we see the subtraction method is convenient to extract the kinematic information while

the shrinking cone method is more useful to obtain the radiation information. Another set of useful techniques utilize the fact that a charged particle from a pileup event leaves a track not originated from the primary vertex, thus it can be subtracted from the jet. These methods include charged hadron subtraction [30], using a jet vertex fraction [31] cut and jet cleansing [32, 33]. One may even use the charged particles from the primary vertex alone when calculation radiation variables, which still provides us a lot of information for the color structure [27]. To improve over these method, one may simply apply them for cones with sizes determined by the (sub)jet p_T's, as discussed in this article. Here, we emphasize that even if we only use tracks from the primary vertex to avoid most of the contamination from pileup events, it is still useful to optimize the jet cone sizes. We have seen that it is the case for quark-gluon discrimination when pileup is turned off. We expect this consideration to be more important when dealing with events with many hard partons, where jets can be easily contaminated by nearby radiation and a large jet radius should be avoided. This happens in, for example, SUSY cascades with long decay chains.

In conclusion, we have shown that the knowledge of the intrinsic jet radiation radius will lead us towards the optimum discriminations for jets with different quantum numbers.

Acknowledgments

The author thanks Dave Soper for numerous useful discussions and comments on the manuscript. The author is in part supported by US Department of Energy under grant numbers DE-FG02-96ER40969 and DE-FG02-13ER41986.

---

[1] S. Chatrchyan *et al.* [CMS Collaboration], Phys. Lett. B **716**, 30 (2012) [arXiv:1207.7235 [hep-ex]]. G. Aad *et al.* [ATLAS Collaboration], Phys. Lett. B **716**, 1 (2012) [arXiv:1207.7214 [hep-ex]].

[2] J. Shelton, arXiv:1302.0260 [hep-ph].

[3] A. Altheimer, S. Arora, L. Asquith, G. Brooijmans, J. Butterworth, M. Campanelli, B. Chapleau and A. E. Cholakian *et al.*, J. Phys. G **39**, 063001 (2012) [arXiv:1201.0008 [hep-ph]].

[4] A. Altheimer, A. Arce, L. Asquith, J. Backus Mayes, E. Bergeaas Kuutmann, J. Berger, D. Bjergaard and L. Bryngemark *et al.*, Eur. Phys. J. C **74**, 2792 (2014) [arXiv:1311.2708

[hep-ex]].

[5] J. Alitti *et al.* [UA2 Collaboration], Z. Phys. C **49**, 17 (1991).

[6] J. Huth *et al.* in E. L. Berger, "Research directions for the decade. Proceedings, 1990 Summer Study on High-Energy Physics, Snowmass, USA, June 25 - July 13, 1990," Singapore, Singapore: World Scientific (1992).

[7] M. Cacciari, J. Rojo, G. P. Salam and G. Soyez, JHEP **0812**, 032 (2008) [arXiv:0810.1304 [hep-ph]].

[8] G. Soyez, JHEP **1007**, 075 (2010) [arXiv:1006.3634 [hep-ph]].

[9] S. Brandt, C. Peyrou, R. Sosnowski and A. Wroblewski, Phys. Lett. **12**, 57 (1964); E. Farhi, Phys. Rev. Lett. **39**, 1587 (1977).

[10] J. Gallicchio and M. D. Schwartz, JHEP **1110**, 103 (2011) [arXiv:1104.1175 [hep-ph]]; J. Gallicchio and M. D. Schwartz, Phys. Rev. Lett. **107**, 172001 (2011) [arXiv:1106.3076 [hep-ph]];

[11] J. Gallicchio and M. D. Schwartz, JHEP **1304**, 090 (2013) [arXiv:1211.7038 [hep-ph]].

[12] J. Thaler and K. Van Tilburg, arXiv:1011.2268 [hep-ph]; J. Thaler and K. Van Tilburg, arXiv:1108.2701 [hep-ph].

[13] S. D. Ellis, Z. Kunszt and D. E. Soper, Phys. Rev. Lett. **69**, 3615 (1992) [hep-ph/9208249].

[14] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008) [arXiv:0802.2470 [hep-ph]].

[15] S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **80**, 051501 (2009) [arXiv:0903.5081 [hep-ph]]; S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **81**, 094023 (2010) [arXiv:0912.0033 [hep-ph]].

[16] D. Krohn, J. Thaler and L. T. Wang, JHEP **1002**, 084 (2010) [arXiv:0912.1342 [hep-ph]].

[17] G. F. Sterman and S. Weinberg, Phys. Rev. Lett. **39**, 1436 (1977).

[18] R. K. Ellis, W. J. Stirling and B. R. Webber, Camb. Monogr. Part. Phys. Nucl. Phys. Cosmol. **8**, 1 (1996).

[19] A. Banfi, G. P. Salam and G. Zanderighi, Eur. Phys. J. C **47**, 113 (2006) [hep-ph/0601139].

[20] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15**, 438 (1972) [Yad. Fiz. **15**, 781 (1972)]; G. Altarelli and G. Parisi, *Asymptotic freedom in parton language*, Nucl. Phys. B **126**, 298 (1977); Y. L. Dokshitzer, *Calculation of the structure functions for deep inelastic scattering and e^+e^- annihilation by perturbation theory in quantum chromodynamics*, Sov. Phys. JETP **46**, 641 (1977) [Zh. Eksp. Teor. Fiz.

73, 1216 (1977)].

[21] T. Sjostrand, S. Mrenna and P. Z. Skands, Comput. Phys. Commun. **178**, 852 (2008) [arXiv:0710.3820 [hep-ph]].

[22] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, Phys. Rev. Lett. **105**, 092002 (2010) [arXiv:1004.2489 [hep-ph]].

[23] M. Cacciari and G. P. Salam, Phys. Lett. B **641**, 57 (2006) [arXiv:hep-ph/0512210].

[24] A. J. Larkoski, G. P. Salam and J. Thaler, JHEP **1306**, 108 (2013) [arXiv:1305.0007 [hep-ph]].

[25] D. Krohn, J. Thaler and L. T. Wang, JHEP **0906**, 059 (2009) [arXiv:0903.0392 [hep-ph]].

[26] Y. Cui, Z. Han and M. D. Schwartz, Phys. Rev. D **83**, 074023 (2011) [arXiv:1012.2077 [hep-ph]].

[27] Z. Han, Phys. Rev. D **86**, 014026 (2012) [arXiv:1112.3378 [hep-ph]].

[28] G. Soyez, G. P. Salam, J. Kim, S. Dutta and M. Cacciari, Phys. Rev. Lett. **110**, 162001 (2013) [arXiv:1211.2811 [hep-ph]].

[29] [CMS Collaboration], CMS-PAS-PFT-09-001.

[30] For example, A. Perloff [CMS Collaboration], J. Phys. Conf. Ser. **404**, 012045 (2012).

[31] For example, The ATLAS collaboration, ATLAS-CONF-2013-083.

[32] D. Krohn, M. Low, M. D. Schwartz and L. -T. Wang, arXiv:1309.4777 [hep-ph].

[33] M. Cacciari, G. P. Salam and G. Soyez, arXiv:1404.7353 [hep-ph].