Tracking the identities of boosted particles

Zhenyu Han

Tracking the Identities of Boosted Particles

Zhenyu Han

Center for the Fundamental Laws of Nature
Harvard University,
Cambridge, MA 02138, USA

We show that the tracking system in a collider detector can be used to efficiently identify boosted massive particles from their QCD backgrounds. We examine variables defined with tracking information which are sensitive to jet radiation patterns, including charged particle multiplicity and N-subjettiness. These variables are barely correlated with variables sensitive to the hard splitting scale in the jet, such as the filtered jet mass. Therefore these two kinds of variables should be combined to optimize the discriminating power. We illustrate the method with W jet tagging. It is shown that for jet $p_T = 500$ GeV, one can gain a factor of 1.6 in statistical significance by combining filtered jet mass and charged particle multiplicity, over filtered mass alone. Adding N-subjettiness increases the factor to 1.8.

PACS numbers:

1. INTRODUCTION

Highly boosted massive standard model (SM) particles are important probes to TeV scale new physics at the Large Hadron Collider (LHC). These particles include the W and Z gauge bosons, the top quark, and possibly the Higgs boson if it is light. For example, it is essential to measure how the WW scattering cross section grows with increasing center of mass energy, in which case boosted W bosons are involved [1, 2]. When these particles decay hadronically, we have to identify them from their QCD backgrounds. See Refs. [3–29] for previous studies. For this purpose, a large boost brings us both advantages and disadvantages. On the one hand, since the decay products are collimated, they are often clustered into a single jet and we are exempted from the combinatorial problem associated with unboosted particles. Sometimes it is even convenient to use a large jet radius to group as many as possible such particles to single fat jets [3]. We will call these jets W/Z/top jets and in general, *boosted massive particle jets* or simply *boosted jets*. On the other hand, since they behave as a single jet, we need to distinguish them from high p_T QCD jets, namely, jets initiated from a high p_T quark or gluon. Because of QCD radiation, the jet mass alone is not a good discriminant especially when we choose a large jet radius.

To distinguish massive particle jets from QCD jets, we can utilize two differences between them. First, compared with QCD splittings, which are usually hierarchical, the momentum of a boosted massive particle is more evenly distributed among its decay products. This results in more than one hard subjets within the fat jet for a massive particle, while only one hard subjet for most QCD jets. Several algorithms [3, 5, 30] have been invented to identify subjets and "groom" the fat jet by discarding the soft subjets. After grooming, the mass of a boosted jet remains close to the decaying particle's mass, while that of a QCD jet usually becomes very small. This allows us to use a mass window cut to eliminate most QCD jets and retain the massive particles we are interested in. To get the characteristic mass of a particular particle, it is important we measure the momenta of all stable particles, which is possible only with calorimeters. Therefore, in most previous jet substructure studies, hadronic calorimeter (HCAL) granularity is assumed.

The second difference between boosted jets and QCD jets stems from their different color structures: QCD jets are initiated from a colored particle, while W/Z/Higgs bosons are color singlets. The color flow of the top decay is also different from a generic QCD jet with three hard subjets. This difference results in different radiation patterns, which are manifest in jet shape variables such as planar flow [4], pull [19], N-subjettiness [24] and R-cores [25]. Unlike the jet mass, to define and examine these variables, we do not need to have the information of all stable particles. In particular, we can take advantage of the tracking system, which has much finer granularity as well as better momentum resolution for low to moderately high momenta. Obviously, the disadvantage of using the tracking

information is it is only available for charged particles whose fraction fluctuates. Therefore, it is complementary to calorimeter information.

In this article, taking W jets as an example, we discuss how to use tracking information to distinguish boosted massive particle jets from QCD jets. We explore variables sensitive to jets' radiation patterns. The simplest such variable is charged particle multiplicity, which nonetheless shows excellent discriminating power. All previously defined jet shape variables can also be calculated with charged particles alone. In particular, we will see N-subjettiness is very useful. These variables are not very sensitive to the mass scale of the particle decay or QCD splitting, and they should be combined with a jet grooming algorithm to optimize the discriminating power. Therefore, we examine these variables for jets containing two hard subjets (identified with the filtering algorithm) with masses close to the W mass. The performances of these variables are conveniently quantified by the significance improvement characteristic (SIC) which is defined as [31]

$$\mathrm{SIC} \equiv \varepsilon_S/\sqrt{\varepsilon_B}, \tag{1}$$

where ε_S and ε_B are respectively the signal (W jets) and background (QCD jets) efficiencies. It is shown that for the LHC, by combining charged multiplicity (or N-subjettiness) with filtering, we achieve an SIC of ~ 1.6 over the filtering method alone. This approach also gives better discriminating power than methods that combine correlated variables, such as in Ref. [10], where different jet grooming algorithms are combined. Combining filtered mass, charged multiplicity and N-subjettiness all together, we can reach an SIC of 1.8.

The article is organized as follows. In Sec. 2, we illustrate the difference between W jets and QCD jets at e^+e^- machines, which is a clean environment without contaminations from initial state radiation and the underlying event. In Sec. 3, we examine jet substructure variables at the LHC and quantify their performances. We discuss our results in Sec. 4 and conclude in Sec. 5.

2. W JETS AND QCD JETS AT e^+e^- MACHINES

We first compare W jets with QCD jets produced at e^+e^- machines, where jet substructure is not contaminated by initial state radiation and the underlying event. The lessons learned in this section can be easily adapted for the LHC, which we discuss in the next section.

A. Charged particle multiplicity

Charged particle multiplicity is a simple but powerful variable that is sensitive to the jet color structure. For example, it can be used to distinguish gluon jets from quark jets [32]. Average charged particle multiplicities ($\langle N_{\mathrm{ch}} \rangle$) in inclusive hadronic events have been measured at a number of e^+e^- machines. The results are compiled in Ref. [33] and reproduced in Fig. 1. These measurements include a variety of center of mass energies from 12 GeV at PETRA, up to LEP 2 energies. Theoretically, the absolute value of $\langle N_{\mathrm{ch}} \rangle$ cannot be predicted because it involves non-perturbative physics. However, its scaling can be described by the modified leading log approximation (MLLA) and local parton-hadron duality (LPHD) [34], see Ref. [36] for a review. The prediction from MLLA+LPHD is shown in Fig. 1, together with the prediction from Pythia 8 [41] simulations. We see that the MLLA+LPHD fit and the Pythia 8 prediction are almost identical, which agree with data excellently.

In Fig. 1, we also notice that charged particle multiplicity grows slowly especially at high energies. This is an example of the fact that the radiation patter is more sensitive to the color structure than the energy scale. This effect has important consequence as will be seen in our study of jet substructure.

For the W boson, a color singlet particle, the average charged multiplicity should be the same as in Fig. 1 at $\sqrt{s} = M_W$, which was confirmed at LEP 2 [35]. Not shown in Fig. 1 is the dispersion of the charged multiplicity (defined as $\langle (N_{\mathrm{ch}} - \langle N_{\mathrm{ch}} \rangle)^2 \rangle^{1/2}$), which is ~ 6 for the W boson. Except for experimental effects, the charged particle multiplicity distribution is invariant under the boost to the W boson, which allows us to use it to identify W jets.

The inclusive charged multiplicity is not directly applicable for identifying W jets since jets are defined for a finite spatial region. We then examine charged multiplicities for individual jets using Pythia 8 simulations. As mentioned in the introduction, a QCD jet without a hard splitting can be easily distinguished from a W jet by using the jet grooming algorithms, therefore, we focus on QCD jets with a hard splitting (2-prong QCD jets), which mimic a W jet more closely. Nonetheless, we will also include 1-prong QCD jets for comparison. For illustration, we consider $e^+e^- \to W^+W^- \to q\bar{q}l\nu$ events and $e^+e^- \to q\bar{q}g$ events in the following fixed configurations (Fig. 2): the hadronically
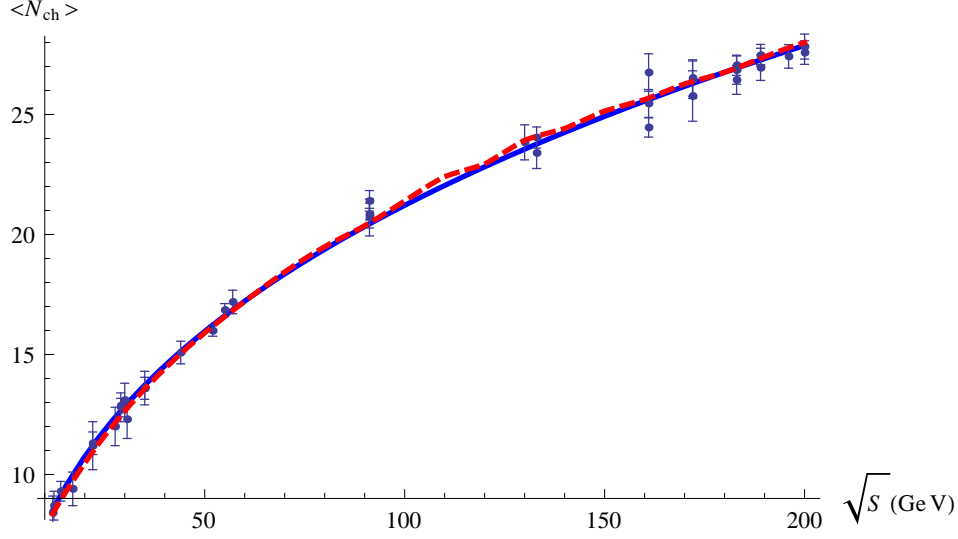
FIG. 1: Average charged particle multiplicities as a function of center of mass energy measured at e^+e^- machines (points) [33], together with the MLLA+LPHD fit (blue solid) and Pythia 8 simulations (red dashed).

decaying W moves along the x axis which decays to two quarks with symmetric four momenta,

$$p_q = \left(\frac{\sqrt{p_T^2 + M_W^2}}{2}, \frac{p_T}{2}, 0, \frac{M_W}{2} \right), \quad p_{\bar{q}} = \left(\frac{\sqrt{p_T^2 + M_W^2}}{2}, \frac{p_T}{2}, 0, -\frac{M_W}{2} \right), \tag{2}$$

where p_T is fixed to be 500 GeV. For $e^+e^- \to q\bar{q}g$, we choose the quark and the gluon to mimic the partons from a W decay, therefore, the 4-momenta are

$$p_q = \left(\frac{\sqrt{p_T^2 + M_W^2}}{2}, \frac{p_T}{2}, 0, \frac{M_W}{2} \right), \quad p_g = \left(\frac{\sqrt{p_T^2 + M_W^2}}{2}, \frac{p_T}{2}, 0, -\frac{M_W}{2} \right),$$
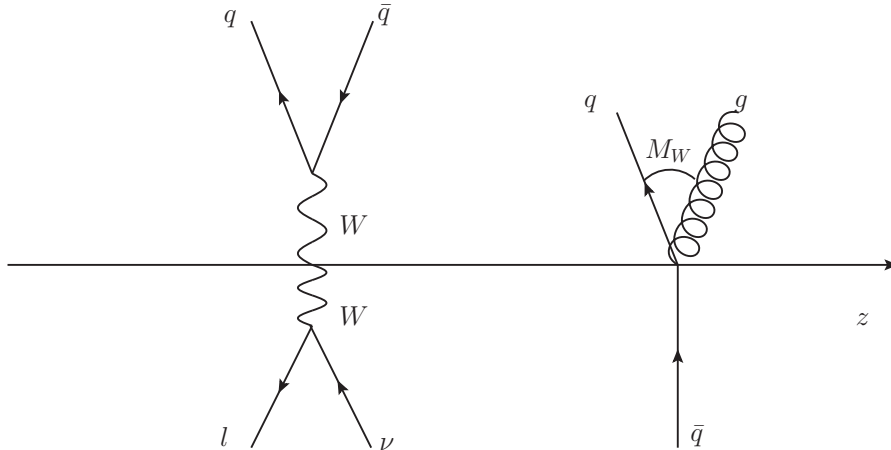
$$p_{\bar{q}} = (p_T, -p_T, 0, 0). \tag{3}$$

FIG. 2: Fixed momentum configurations for illustration. Left: $WW \to q\bar{q}l\nu$ with the hadronic W moving perpendicularly to the beam and decaying to quarks with symmetric momenta. Right: $e^+e^- \to q\bar{q}g$ with a qg pair mimicking a W boson.

There is another configuration with the above momenta in which the two quarks are close to each other and the gluon is in the opposite hemisphere. This configuration happens much less often than the one showing in Fig. 2, and

we have ignored it in our illustration. In a realistic situation at the LHC as will be discussed in Sec. 3, we should include all possible configurations contributing to a high p_T jet. This is more conveniently done with Pythia 8 or other simulations.

We keep the momentum and color flow configurations fixed as in Fig. 2, and repeatedly use Pythia 8 to simulate showering and hadronization and obtain two data samples corresponding to the two processes. We cluster stable particles to anti-k_t jets (R=1.2) with FastJet [42]. Each WW event then contains a W jet in the upper hemisphere, and each $q\bar{q}g$ event contains a 2-prong jet in the upper hemisphere and a 1-prong jet in the lower hemisphere. No cut is used in this procedure. The average number of charged particles for the W jet, the 2-prong and 1-prong QCD jets are shown in Fig. 3. From Fig. 3, we see the average charged multiplicity of a W jet is larger than that of a 1-prong QCD jet while smaller than a 2-prong QCD jet. This is due to their different color structure. In particular, the 2-prong QCD jet is color connected to the other side of the event, therefore it contains more radiation than the W jet. To distinguish a W jet from a 2-prong QCD jet, we can apply a cut $N_{ch} \leq N_{ch}^{cut}$. For example, when $N_{ch}^{cut} = 19$, we keep 63% W jets and 7.7% 2-prong QCD jets, which boosts the SIC by a factor of 2.3. Because of the large boost and the large jet radius, $R = 1.2$, almost all particles from the W decay are included in the W jet. Due to charge conservation, the W jet (almost) always contains an odd number of charged particles. If we keep only jets with odd number of charged particles, we obtain a larger SIC. However, this feature is easily lost due to experimental acceptance, and we do not pursue this possibility further.
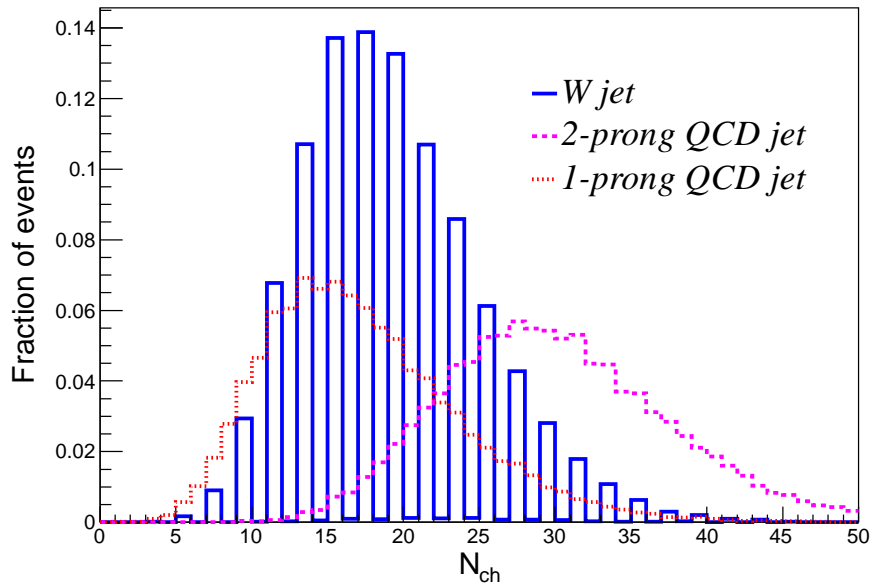

FIG. 3: Charged particle multiplicities for W jets, 2-prong and 1-prong QCD jets with $p_T = 500$ GeV, in the fixed momentum configurations of Fig. 2 (see text).

B. N-subjettiness

Other existing jet shape variables can be defined with charged particles too. Here, we take N-subjettiness as an example, which is derived from N-jettiness [43] and defined in Ref. [24] as follows. For a set of particles $\{i\}$ in a jet with radius R_0 and a set of N axes $\{J\}$, we define the distance $\Delta R_{J,i} \equiv \sqrt{\Delta\eta_{J,i}^2 + \Delta\phi_{J,i}^2}$ for each (J,i) pair. Then we define a quantity

$$\tilde{\tau}_N^{(\beta)} \equiv \frac{1}{d_0} \sum_i p_{T,i} \min\{\Delta R_{1,i}^\beta, \ldots \Delta R_{J,i}^\beta \ldots \Delta R_{N,i}^\beta\}, \tag{4}$$

where $d_0 = \sum_i p_{T,i} R_0^\beta$ and β is a pre-selected constant. We vary the directions of the N axes to find the minimum $\tilde{\tau}_N^{(\beta)}$, which is defined as N-subjettiness, $\tau_N^{(\beta)}$.

In our example with fixed momentum configuration, we can simply use the momenta in Eqs. (2) as the two axes to calculate $\tilde{\tau}_2^{(\beta)}$, which does not differ significantly from the true $\tau_2^{(\beta)}$ after minimization. The $\tau_2^{(\beta)}$ distributions for W jets and 2-prong QCD jets are shown in Fig. 4, where we have chosen $\beta = 1$. The variable τ_2 quantifies how likely a given jet contains two hard subjets, a smaller value corresponding to a larger likelihood. In our special momentum configuration, we always have two hard subjets. In this case, τ_2 becomes a measure of how diffuse the radiation is. As expected, more often τ_2 is larger for 2-prong QCD jets than W jets, which allows us to apply a cut, $\tau_2 < \tau_2^{\text{cut}}$, to suppress background QCD jets. In the general case, when the jet momentum configuration is not fixed, a better
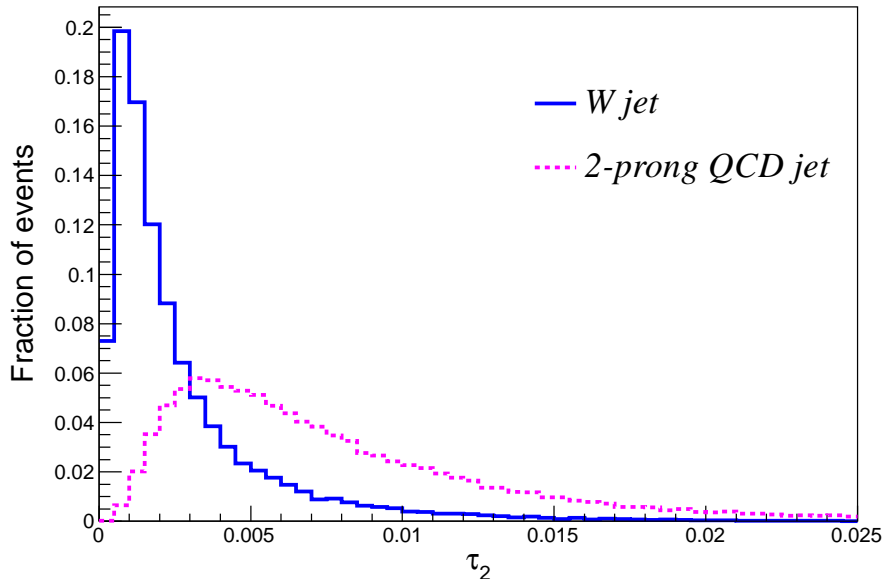

FIG. 4: 2-subjettiness with axes fixed to Eq. (2), for jets with $p_T = 500$ GeV in the fixed momentum configurations of Fig. 2.

variable is τ_2/τ_1 [24], which will be discussed in the next section for the LHC.

3. W JET TAGGING AT THE LHC

We now turn to the LHC, where W jet tagging is much more challenging due to presence of initial state radiation and the underlying event. As an example, we consider W jets and QCD jets in the p_T range $(500, 550)$ GeV. These jets are obtained as in Ref. [25]: we use Pythia 8 to generate high p_T WW pairs which decay semiletonically, and W+jet events with the W decay leptonically. Each WW event then contains a W jet, while each W+jet event contains a high p_T QCD jet. The visible stable particles in these events are grouped in 0.1×0.1 bins in the (η, ϕ) plane, corresponding to the HCAL granularity. Jets are found with the Cambridge/Aachen algorithm ($R = 1.2$). Since we are interested in jets with a hard splitting and mass close to the W mass, we apply the filtering algorithm with the mass drop method and examine further jets passing the filtered jet mass (m_{filt}) cut, $(60, 100)$ GeV. The mass drop parameters are $\mu = 0.71$ and $y_{\text{cut}} = 0.09$, which give a factor of 2.2 improvement in the SIC over the original fat jets. With this cut, we reduce the number of QCD jets by 91% and at the same time retain 66% W jets. Events passing this cut all have a hard splitting, therefore we can apply the lessons learned from the previous section.

We identify tracks with $p_T > 1$ GeV and $|\eta| < 2.5$ in a given jet, and use them to calculate the charged particle multiplicity and N-subjettiness. Note that although the jet has passed the filtered mass window cut, we include all tracks in the original fat jet, which give us more information about the jet color structure. The most efficient N-subjettiness variable for W-tagging is τ_2/τ_1, where we have set $\beta = 1$ in Eq. (4) and used the code in Ref. [44]. In addition, the filtered jet mass distributions still differ for W jets and QCD jets after imposing the $(60, 100)$ GeV mass window cuts. We therefore include the filtered mass as well. The three variables under consideration are shown in Fig. 5. From Fig. 5, we see N_{ch} and τ_2/τ_1 have similar features as in a e^+e^- machine (Fig. 3 and Fig. 4), but the distinctions between QCD jets and W jets are smaller.
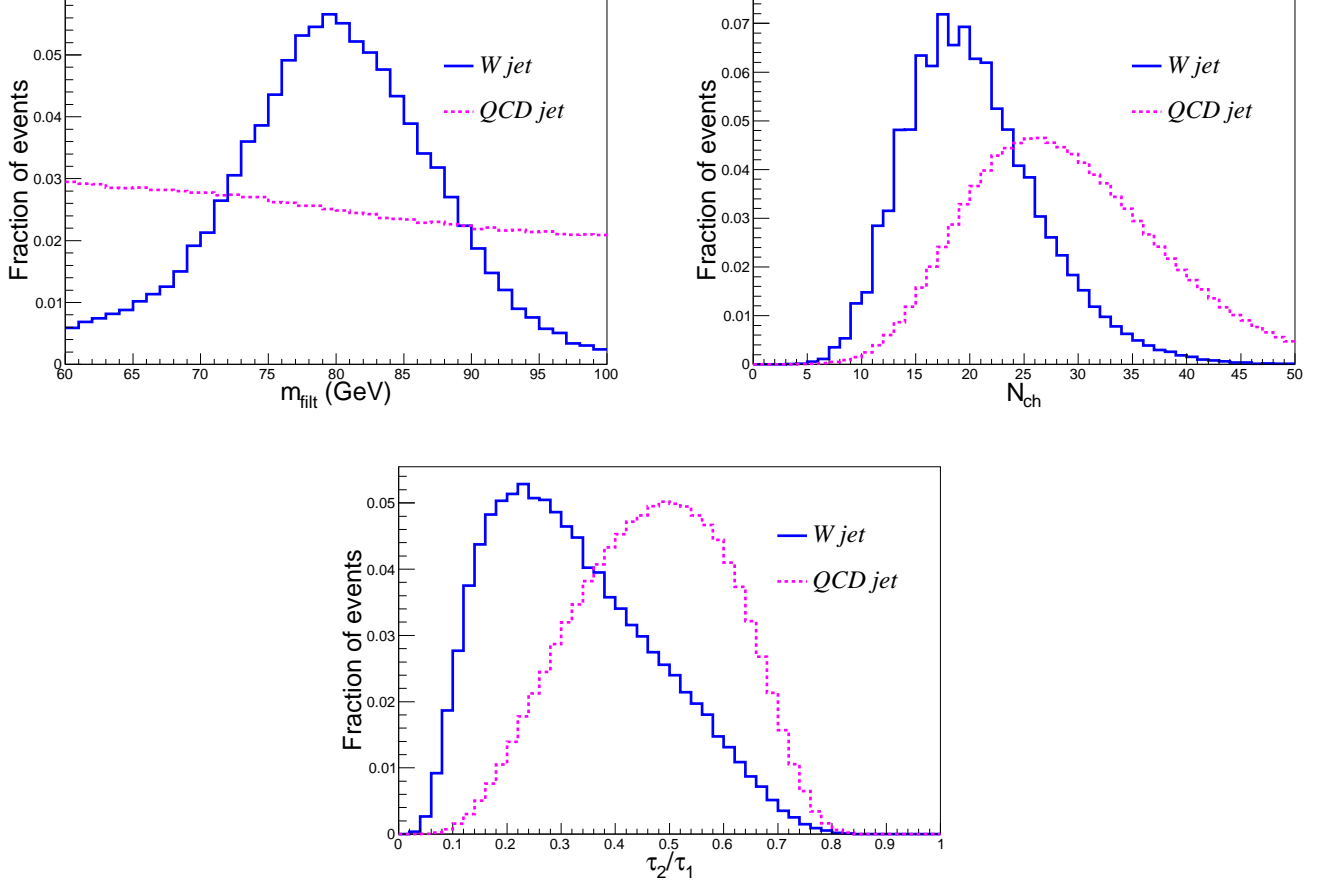
FIG. 5: Variables considered for the LHC: jet mass after filtering, charged multiplicity and τ_2/τ_1, for $p_T^{\text{jet}} \in (500, 550)$ GeV. The filtered mass are constructed assuming HCAL granularity; N_{ch} and τ_2/τ_1 are constructed using tracks with $p_T > 1$ GeV and $|\eta| < 2.5$.

	m_{filt}	N_{ch}	τ_2/τ_1
m_{filt}	1.15	1.66 (1.59)	1.67 (1.58)
N_{ch}	-	1.34	1.55 (1.50)
τ_2/τ_1	-	-	1.39
all:		1.85	

TABLE 1: Optimized improvement in the SIC. The events have passed an overall filtered mass cut $(60, 100)$ GeV. The diagonal elements of the first three rows are obtained by using individual variables with an optimized rectangular cut. The off-diagonal elements are obtained by combining a pair of variables: the numbers in the parentheses are obtained using rectangular cuts and the numbers outside are from BDT. The best improvement for combining all three variables in BDT is given in the last row.

For a single variable, we can apply a rectangular cut to improve the SIC. For two or more variables, it is better to combine them using a multivariate classifier such as the Boosted Decision Trees method [45] in the package TMVA [46]. The best performances for the three individual variables as well as all combinations are given in Tab. 1.

We emphasize here the numbers in Tab. 1 are obtained on jet samples that have passed the filtered mass cut, $(60, 100)$ GeV, which has already increased the SIC by a factor of 2.2. Therefore, the overall improvement is the number in Tab. 1 multiplied by 2.2. We see the number for using filtered mass alone is 1.15, which means $(60, 100)$ GeV is not the optimum mass window. The extra factor of 1.15 corresponds to a narrower mass window $(72, 92)$ GeV and the best improvement one can get from filtering alone is 2.5. Better improvements are obtained from N_{ch} or τ_2/τ_1

	m_{filt}	N_{ch}	τ_2/τ_1
m_{filt}	1	-0.08	-0.12
N_{ch}	-0.08	1	0.51
τ_2/τ_1	-0.12	0.51	1

	m_{filt}	N_{ch}	τ_2/τ_1
m_{filt}	1	0.07	-0.14
N_{ch}	0.07	1	0.50
τ_2/τ_1	-0.14	0.50	1

TABLE 2: Linear correlation matrices of the variables. Left: W jets; right: QCD jets.

(with the filtered mass window fixed to $(60, 100)$ GeV): we obtain 1.34 (1.39) by optimizing the cut on N_{ch} (τ_2/τ_1).

One may also be interested in the performance of a single variable sensitive to the radiation, such as τ_2/τ_1, without imposing a filtered mass cut. It turns out one can improve the significance by a factor of 1.44 using τ_2/τ_1 calculated from charged particles [1] for jets with $R = 1.2$ and $p_T = 500$ GeV. This is lower than what we get from filtering. Nevertheless, τ_2/τ_1 may be more useful if we do not know the exact mass of the boosted particle. Also at higher p_T's, the resolution for mass measurement degrades, while we expect radiation variables to work better. This is because for higher p_T's, the decay products of a color singlet particle occupy a smaller region while the radiation pattern of a QCD jet does not change significantly, which make the two cases more distinguishable. Similar observation has been made in Ref. [25]. In the extreme case, all decay products of a color singlet particle enters a single or a few adjacent calorimeter cells which makes the mass information unavailable, and we are forced into an inclusive search of color singlet particles without using their mass information.

We can obtain better discriminating power if we combine two variables and vary the cuts on both of them. From Table 1, we see a factor of ~ 1.6 is reached if we optimize rectangular cuts on both m_{filt} and N_{ch} (τ_2/τ_1), or if we combine them in BDT, the latter being slightly better. We also notice combining N_{ch} and τ_2/τ_1 gives us smaller improvement (1.55) than combining one of them with the filtered mass, despite they each alone are excellent discriminants. This is due to the larger correlation between N_{ch} and τ_2/τ_1, both of which measure the amount of radiation in the jets. On the other hand, the correlation between N_{ch} (τ_2/τ_1) and m_{filt} is small. These correlations are manifest in the two dimensional distributions for each pair of variables, shown in Fig. 6. The linear correlation matrices of the three variables are given in Tab. 2.

From the above results, we draw the conclusion that if we are to use two variables for W jet tagging, the best way is to choose one variable sensitive to the hard splitting scale and the other one sensitive to the amount of radiation. These two kinds of variables characterize two major differences between W jets and QCD jets and they are barely correlated. This is in accordance with the observation we made in Sec. 2, where we saw that charged multiplicity increase slowly with respect to the center of mass energy. By doing so, we obtain significant improvement than using each individual variable.

Finally, we can combine all three variables in BDT. Despite the sizable correlation between N_{ch} and τ_2/τ_1, they still contain different information which can improve the SIC when combined. We show the SIC as a function of signal efficiency in Fig. 7. From Fig. 7, we see that a factor of 1.85 is achieved using the optimum cut, with sizable signal efficiency of ~ 0.3. We may also add other variables as in Ref. [25], and the SIC will gradually saturate. In Ref. [25], a set of 25 variables are used which yield a factor of 2.4. It turns out by adding the two extra variables, $N_{\text{ch}}, \tau_2/\tau_1$ to the set (m_{filt} was included in Ref. [25]), we only obtain a few percent improvement, which means most of the information is redundant. However, we emphasize that if one would like to sacrifice performance for simplicity and choose to use only a few variables, the variables we have considered in this article are among the best ones.

4. DISCUSSIONS

A. Experimental considerations

In the above discussions, we have not taken into account the experimental efficiency for reconstructing tracks in a jet. In Ref. [37], jets with $p_T \sim 200$ GeV are studied where it is shown the efficiency for identifying tracks is around 90% with a fake rate about 0.1%. The performance of our tracking variables will degrade accordingly. One may be concerned about whether the efficiency decrease significantly for higher p_T's. This deserves dedicated studies using

---

[1] The performance depends on the value of β in Eq. (4). It turns out that with a filtered mass cut, $\beta = 1$ is a better choice than $\beta = 2$. Therefore we have used $\beta = 1$ all through the paper. However, without a filtered mass cut, $\beta = 2$ works better, which gives a larger SIC of 1.58 for jet $p_T = 500$ GeV. A detailed study of the β dependence is beyond the scope of the article.
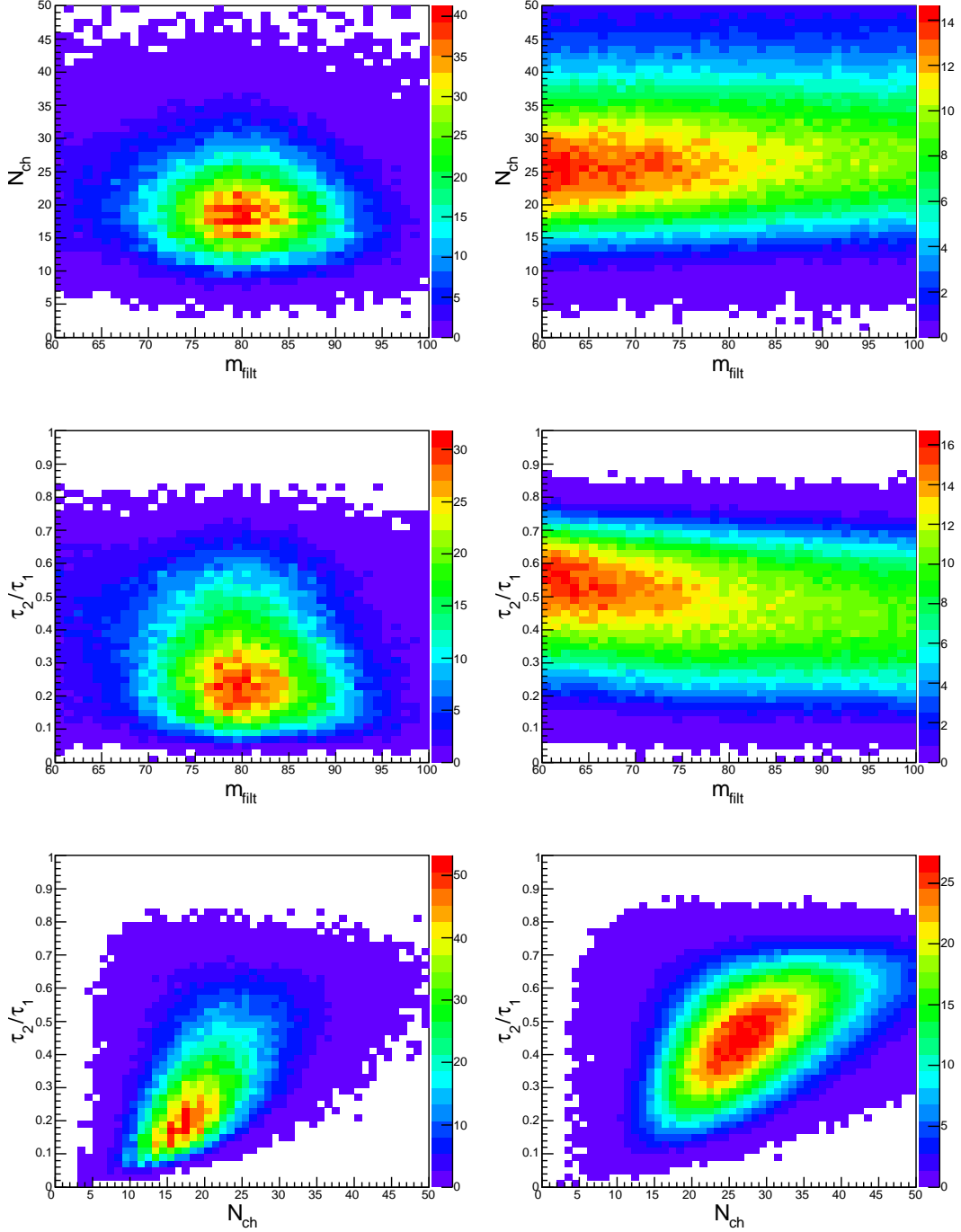
FIG. 6: Two dimensional distributions for variable pairs. Left: W jets; right: QCD jets. The number of events is normalized to 10k for each plot.

both simulations and the real data. However, we believe this is not the case for QCD jets. The reason is, the scaling of N_{ch} at a hadron collider is similar to that of e^+e^- machines [40] we see in Fig. 1. As we have noticed, the number of tracks grow very slowly at high energies. The angular distributions of these tracks will not change significantly either. Therefore, the efficiency will not change significantly for QCD jets. The W jet is a different case: when the boost is larger, all the tracks will be packed in a smaller region, which may cause the efficiency to drop significantly. However, as we have used the fact N_{ch} is smaller for W jets, a drop in efficiency will only make the number even
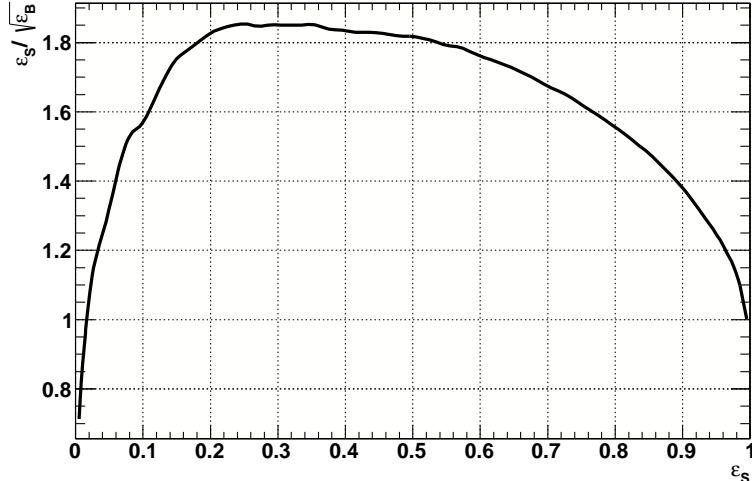
FIG. 7: The SIC as a function of signal efficiency by combining m_{filt}, N_{ch} and τ_2/τ_1 in BDT, for $p_T^{\text{jet}} \in (500, 550)$ GeV.

smaller and will not hurt the discriminating power. In fact, an observation of dense tracks (or hits if tracks are difficult to reconstruct) in a small region while scarce outside is already a very clean signal of a W jet.

Another concern about using high momentum tracks is the tracking resolution degrades for higher p_T's. This does not affect the charged multiplicity measurement, but N-subjettiness. However, since the momentum of the jet is shared by tens of particles, each charged particle does not usually have a very high p_T. For the 500 GeV jets we considered in Sec. 2, the leading track's p_T is shown in Fig. 8, from which we see the track's p_T rarely goes about 200 GeV where the resolution of the momentum measurement is still better than 10% [37, 38]. Even in the presence of very high p_T tracks, these tracks will likely dominate the directions of the subjets, and according to Eq. (4), not significantly contribute to N-subjettiness. Therefore, in this article we have ignored the experimental resolution which should not affect our results significantly unless we are interested in jets with extremely high p_T's (> 1 TeV).
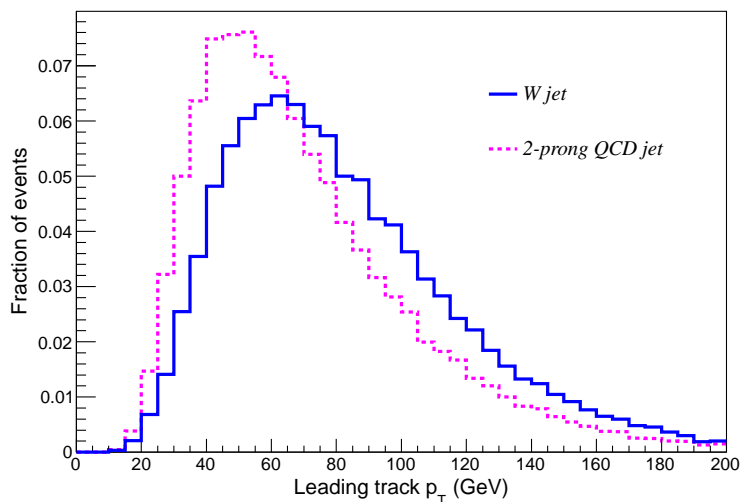

FIG. 8: The leading track's p_T in jets with $p_T^{\text{jet}} = 500$ GeV.

Particle flow [39] is an interesting and useful experimental approach, in which one combines the information from all subdetectors to reconstruct both charged and neutral particles. Very briefly, charged particles, including muons, electrons and charged hadrons, are first reconstructed and their energy removed from HCAL and ECAL clusters.

The remaining clusters are identified as neutral particles including photons and neutral hadrons. Therefore, particles reconstructed from the particle flow algorithm (which we call PF particles) are a superset of the charged particles. If PF particles are used in jet reconstruction, the variables discussed in this paper can be naturally defined without explicitly referring to whether tracking or calorimeter information is used. For example, we can count the number of PF particles in stead of charged particles in our particle multiplicity definition. Since the PF particles are a superset of the charged particles, one may obtain better results than those quoted in this articles. Of course, for jets with a higher p_T, more particles are merged in single calorimeter clusters and it remains to be studied whether significant improvement can be achieved.

B.   Other particles

The same method can be used on other color singlet particles such as the Z boson and a light Higgs boson. The major difference between these particles and the W boson is in their masses. As we have discussed, the small mass difference between Z and W does not affect significantly the charged multiplicity or N-subjettiness. Therefore, if we are only interested in one of them and treat the other one as a background, mass variables such as the filtered mass is necessary. However, as we have discussed in the previous section, for extremely high p_T jets, one cannot measure accurately the mass of the color singlet particle, although it is easy to distinguish a color singlet particle from a QCD jet.

Color singlet particles also differ in their spins, which affect the angular distributions of the decay products with respect to the moving direction of the decaying particle. For a transverse vector boson, one of the fermions from the decay tends to go along the direction of the vector boson, while the other one against it in the rest frame of the decaying particle. When boosted, this renders the momentum of one of the fermions to be much smaller than the other one. For a longitudinal vector boson, the two fermions tend to move perpendicular to the vector boson's direction which results in more balanced momentum configurations. Therefore, the decay of a transverse vector boson is more like a QCD splitting, which makes it harder to be identified. The Higgs boson, being a spin-0 particle, has its decay products evenly distributed in its rest frame, thus the distinguishing power for a spin-0 particle is in between a transverse vector boson and a longitudinal one.

Top jets are more complicated objects. On the one hand, compared with a W jet, it is easier to distinguish a top jet from a QCD jet using kinematic information because it contains three hard subjets. One may also utilize the presence of a b jet and other kinematic variables such as the helicity angle [16] to improve the top tagging efficiency. On the other hand, top quark is a colored particle, and its radiation pattern is more like a QCD jet than a boosted W, which makes it more difficult to use radiation variables to identify a top jet. Nonetheless, a top jet contains a W jet among its decay products. One may try to use the W tagging method or similar techniques to improve top tagging. This merits a detailed study. Here we only note the difference in charged multiplicity between top jets and QCD jets. Similar to Sec. 2, we examine top jets and QCD jets in the same fixed momentum configuration: we consider $e^+e^- \to t\bar{t}$ and $e^+e^- \to q\bar{q}gg$ events. Denoting the beam direction as the z axis, we let the top move in the x direction and its decay products all lie in the $y-z$ plane in the top rest frame. The momenta of the two quarks from W decay are set to be of the same size. We let the qgg from the $e^+e^- \to q\bar{q}gg$ process to have the same momentum configuration as the top decay. Then we count the numbers of charged particles in the resulting top jets and QCD jets, which are shown in Fig. 9. We see a clear distinction between top jets and 3-prong QCD jets. Note that, unlike the W jet case, the charge is not conserved in a top jet no matter how large the jet size is. This is because the top quark is color connected to the anti-top in the opposite hemisphere, and we expect soft particles in between them which can easily change the charge of the top jet from even to odd or vice versa.

5.   CONCLUSION

In this article, we have demonstrated that hadronically decaying boosted massive particles can be tagged using tracking information. Although one cannot reconstruct the masses of the decaying particle using tracks alone, the distributions of these tracks are sensitive to jet radiation patterns. In particular, the charged particle multiplicity from a color singlet particle decay is boost invariant, which serves as an excellent discriminant between $W/Z/Higgs$ jets and QCD jets. Other jet shape variables can also be calculated using charged particles alone, which is complementary to variables calculated from calorimeter information.

We have also shown jet substructure variables can be classified to those sensitive to the jet hard splitting scale and those sensitive to the radiation pattern, which have small correlations. If two or a few variables are used to distinguish massive particle jets from QCD jets, it is most efficient to combine variables from the two categories. We have used W jet tagging as an example and demonstrated that by combining the jet filtering algorithm with charged
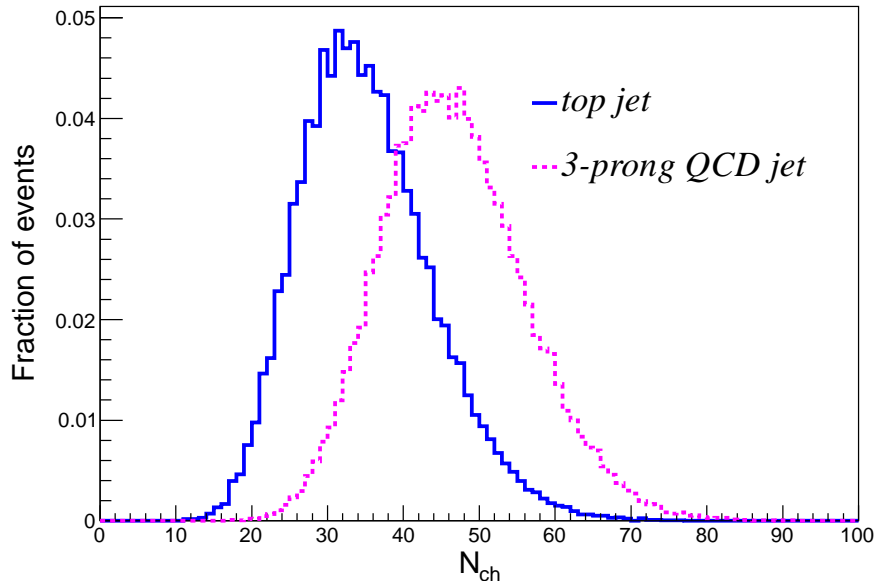
FIG. 9: Charged particle multiplicities for top jets and QCD jets with $p_T = 500$ GeV and fixed momentum configuration (see text) at a e^+e^- machine.

particle multiplicity or N-subjettiness, we can improve the statistical significance by a factor of ~ 1.6 over filtering alone. This approach simplifies the multivariate method in Ref. [25] that utilizes as many as 25 variables, and may find applications at the LHC especially at its early stages.

Acknowledgments

[1] M. S. Chanowitz and M. K. Gaillard, Nucl. Phys. B **261**, 379 (1985).
[2] J. M. Butterworth, B. E. Cox and J. R. Forshaw, Phys. Rev. D **65**, 096014 (2002) [arXiv:hep-ph/0201098].
[3] J. M. Butterworth, A. R. Davison, M. Rubin and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008) [arXiv:0802.2470 [hep-ph]].
[4] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung and J. Virzi, Phys. Rev. D **79**, 074017 (2009) [arXiv:0807.0234 [hep-ph]]; L. G. Almeida, S. J. Lee, G. Perez, G. Sterman and I. Sung, arXiv:1006.2035 [hep-ph].
[5] 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]].
[6] C. Hackstein and M. Spannowsky, Phys. Rev. D **82**, 113012 (2010) [arXiv:1008.2202 [hep-ph]].
[7] A. Katz, M. Son and B. Tweedie, JHEP **1103**, 011 (2011) [arXiv:1010.5253 [hep-ph]].
[8] T. Plehn, G. P. Salam and M. Spannowsky, Phys. Rev. Lett. **104**, 111801 (2010) [arXiv:0910.5472 [hep-ph]].
[9] G. D. Kribs, A. Martin, T. S. Roy and M. Spannowsky, Phys. Rev. D **81**, 111501 (2010) [arXiv:0912.4731 [hep-ph]].
[10] D. E. Soper and M. Spannowsky, JHEP **1008**, 029 (2010) [arXiv:1005.0417 [hep-ph]].
[11] C. -R. Chen, M. M. Nojiri and W. Sreethawong, JHEP **1011**, 012 (2010) [arXiv:1006.1151 [hep-ph]].
[12] A. Falkowski, D. Krohn, L. -T. Wang, J. Shelton and A. Thalapillil, Phys. Rev. D **84**, 074022 (2011) [arXiv:1006.1650 [hep-ph]].
[13] G. D. Kribs, A. Martin, T. S. Roy and M. Spannowsky, Phys. Rev. D **82**, 095012 (2010) [arXiv:1006.1656 [hep-ph]].
[14] A. Katz, M. Son and B. Tweedie, Phys. Rev. D **83**, 114033 (2011) [arXiv:1011.4523 [hep-ph]].
[15] J. Thaler and L. T. Wang, JHEP **0807**, 092 (2008) [arXiv:0806.0023 [hep-ph]].
[16] D. E. Kaplan, K. Rehermann, M. D. Schwartz and B. Tweedie, Phys. Rev. Lett. **101**, 142001 (2008) [arXiv:0806.0848 [hep-ph]].
[17] L. G. Almeida, S. J. Lee, G. Perez, I. Sung and J. Virzi, Phys. Rev. D **79**, 074012 (2009) [arXiv:0810.0934 [hep-ph]].
[18] D. Krohn, J. Shelton and L. T. Wang, JHEP **1007**, 041 (2010) [arXiv:0909.3855 [hep-ph]].
[19] J. Gallicchio and M. D. Schwartz, Phys. Rev. Lett. **105**, 022001 (2010) [arXiv:1001.5027 [hep-ph]].
[20] T. Plehn, M. Spannowsky, M. Takeuchi and D. Zerwas, JHEP **1010**, 078 (2010) [arXiv:1006.2833 [hep-ph]].
[21] B. Bhattacherjee, M. Guchait, S. Raychaudhuri and K. Sridhar, Phys. Rev. D **82**, 055006 (2010) [arXiv:1006.3213 [hep-ph]].
[22] K. Rehermann and B. Tweedie, JHEP **1103**, 059 (2011) [arXiv:1007.2221 [hep-ph]].
[23] J. -H. Kim, Phys. Rev. D **83**, 011502 (2011) [arXiv:1011.1493 [hep-ph]].
[24] J. Thaler and K. Van Tilburg, arXiv:1011.2268 [hep-ph]; J. Thaler and K. Van Tilburg, arXiv:1108.2701 [hep-ph].
[25] Y. Cui, Z. Han and M. D. Schwartz, Phys. Rev. D **83**, 074023 (2011) [arXiv:1012.2077 [hep-ph]].
[26] ATLAS Collaboration, Report No. ATL-PHYS-PUB-2009-088. ATLAS Collaboration, Report No. ATL-COM-PHYS-2009-345.
[27] G. Giurgiu [for the CMS collaboration], arXiv:0909.4894 [hep-ex].
[28] A. Hook, M. Jankowiak and J. G. Wacker, JHEP **1204**, 007 (2012) [arXiv:1102.1012 [hep-ph]].
[29] T. Plehn, M. Spannowsky and M. Takeuchi, Phys. Rev. D **85**, 034029 (2012) [arXiv:1111.5034 [hep-ph]].
[30] D. Krohn, J. Thaler and L. T. Wang, JHEP **1002**, 084 (2010) [arXiv:0912.1342 [hep-ph]].
[31] J. Gallicchio, J. Huth, M. Kagan, M. D. Schwartz, K. Black and B. Tweedie, JHEP **1104**, 069 (2011) [arXiv:1010.3698 [hep-ph]].
[32] J. Gallicchio and M. D. Schwartz, Phys. Rev. Lett. **107**, 172001 (2011) [arXiv:1106.3076 [hep-ph]].
[33] Table 11.2 in G. Dissertori, I. Knowles and M. Schmelling, "Quantum Chromodynamics - High Energy Experiments and Theory", page 370.
[34] B. R. Webber, Phys. Lett. B **143**, 501 (1984).
[35] N. Robertson *et al.* [ALEPH Collaboration], ALEPH-99-026; P. Abreu *et al.* [DELPHI Collaboration], Phys. Lett. B **416**, 233 (1998).
[36] V. A. Khoze and W. Ochs, Int. J. Mod. Phys. A **12**, 2949 (1997) [hep-ph/9701421].
[37] CMS Collaboration, Report No. CERN-LHCC-99-14.
[38] G. L. Bayatian *et al.* [CMS Collaboration], "CMS physics: Technical design report,"
[39] CMS Collaboration, Report No. CMS-PAS-PFT-09-001.
[40] J. F. Grosse-Oetringhaus and K. Reygers, J. Phys. G **37**, 083001 (2010) [arXiv:0912.0023 [hep-ex]].
[41] T. Sjostrand, S. Mrenna and P. Z. Skands, Comput. Phys. Commun. **178**, 852 (2008) [arXiv:0710.3820 [hep-ph]].
[42] M. Cacciari and G. P. Salam, Phys. Lett. B **641**, 57 (2006) [arXiv:hep-ph/0512210].
[43] I. W. Stewart, F. J. Tackmann and W. J. Waalewijn, Phys. Rev. Lett. **105**, 092002 (2010) [arXiv:1004.2489 [hep-ph]].
[44] J. Thaler and K. Van Tilburg, http://www.jthaler.net/doku.php?id=jets.
[45] B. P. Roe, H. J. Yang, J. Zhu, Y. Liu, I. Stancu and G. McGregor, Nucl. Instrum. Meth. A **543**, 577 (2005) [arXiv:physics/0408124].
[46] http://tmva.sourceforge.net/.