Published for SISSA by Springer

Received: *January 30, 2015*
Revised: *March 5, 2015*
Accepted: *March 12, 2015*
Published: *April 23, 2015*

Associated jet and subjet rates in light-quark and gluon jet discrimination

Biplob Bhattacherjee,[a] Satyanarayan Mukhopadhyay,[b] Mihoko M. Nojiri,[b,c] Yasuhito Sakaki[c] and Bryan R. Webber[d]

[a] *Centre for High Energy Physics, Indian Institute of Science,*
Bangalore, India
[b] *Kavli IPMU (WPI), The University of Tokyo,*
Kashiwa, Chiba 277-8583, Japan
[c] *KEK Theory Center and Sokendai,*
Tsukuba, Ibaraki 305-0801, Japan
[d] *Cavendish Laboratory, J.J. Thomson Avenue,*
Cambridge, U.K.

E-mail: biplob@cts.iisc.ernet.in, satya.mukho@ipmu.jp, nojiri@post.kek.jp, sakakiy@post.kek.jp, webber@hep.phy.cam.ac.uk

Abstract: We show that in studies of light quark- and gluon-initiated jet discrimination, it is important to include the information on softer reconstructed jets (associated jets) around a primary hard jet. This is particularly relevant while adopting a small radius parameter for reconstructing hadronic jets. The probability of having an associated jet as a function of the primary jet transverse momentum (p_T) and radius, the minimum associated jet p_T and the association radius is computed up to next-to-double logarithmic accuracy (NDLA), and the predictions are compared with results from `Herwig++`, `Pythia6` and `Pythia8` Monte Carlos (MC). We demonstrate the improvement in quark-gluon discrimination on using the associated jet rate variable with the help of a multivariate analysis. The associated jet rates are found to be only mildly sensitive to the choice of parton shower and hadronization algorithms, as well as to the effects of initial state radiation and underlying event. In addition, the number of k_t subjets of an anti-k_t jet is found to be an observable that leads to a rather uniform prediction across different MC's, broadly being in agreement with predictions in NDLA, as compared to the often used number of charged tracks observable.

Keywords: QCD Phenomenology, Jets

ArXiv ePrint: 1501.04794

doi:10.1007/JHEP04(2015)131

Contents

1 Introduction

Hadronic jets are the most abundant objects at a proton-proton collider like the LHC, and it is a major challenge to separate the signals being looked for from standard model (SM) backgrounds in multijet final states. One promising direction that has recently received attention in both theoretical and experimental studies is that the separation of light quark-initiated jets from gluon-initiated ones can be viable in these search channels. Quarks are often encountered in the decays of new particles predicted by scenarios beyond the standard model, as well as in the decay of the weak bosons, Higgs and top quark in the SM itself. On the other hand, in the corresponding SM backgrounds involving multiple hard jets, there is a larger fraction of gluon-initiated jets from QCD radiation. Here, quark- or gluon-initiated jets (henceforth simply referred to as quark and gluon jets) refer to the parton in the hard process at leading order in perturbation theory that initiates the parton shower. Based on the difference in the radiation pattern of quarks and gluons, a likelihood based discriminant can be built to separate decay jets from QCD radiation jets with a certain efficiency [1].

Several variables have been proposed to separate quark and gluon jets, mostly relying on the fact that a gluon of similar energy leads to more soft emissions compared to a quark. This includes both discrete variables like the number of charged tracks inside the jet cone, as well as continuous ones like the width of a jet and energy-energy-correlation (EEC) angularity [1–5]. ATLAS and CMS collaborations have also studied the discrimination of light quarks from gluons along these lines with the 7 and 8 TeV LHC data respectively [6, 7].

Using data samples with "enriched quark and gluon content", data-based taggers were also developed, and compared to the predictions from Monte Carlo (MC) simulations. While there are differences between the predictions of different MC's, as well as between the data-based tagger and the MC results, they are consistent with each other within the large systematic uncertainties at present.

An important question in this regard is the proper choice of a jet algorithm and radius parameter. In the LHC environment, in order to keep the contribution of the underlying event and multiple proton-proton collisions at a minimum, for multijet processes the standard choice is an anti-k_t algorithm with radius parameter $R = 0.4$. In addition, in the ATLAS study mentioned above, jets are required to satisfy an isolation criterion: a jet is considered isolated if there is no other reconstructed jet within a cone of size $\Delta R < 0.7$ (where $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ is the standard distance measure in the pseudorapidity-azimuthal angle plane). An optimum choice for the jet radius parameter was discussed in refs. [8, 9] for quark and gluon jets as a function of their transverse momenta (p_T), and it was observed that one usually requires a larger radius for a gluon jet in order for the parton p_T to be close to the jet p_T. However, for experimental purposes it is advantageous to use a fixed and small radius parameter for the jets, for reasons mentioned above. Therefore, we propose to recover the missed information on radiation from the parent parton outside the chosen jet radius by including softer reconstructed jets that can be present (with a calculable probability) around a certain radius of a primary hard jet. These softer jets are referred to as "associated jets" in this study. It is important to note here that imposing an isolation criterion as above while studying quark and gluon jet properties might not be appropriate, since it leads to rejecting a fraction of the jet candidates beforehand, and thus biasing the sample to ones where the initial quark or gluon has not radiated outside the adopted jet radius.

We first compute the associated jet rates in QCD to next-to-double logarithmic accuracy in section 2, and then compare the analytical results with those from different parton shower MC's in section 3. Using the information on the presence (or absence) of associated jets can improve the discrimination of quarks and gluons. We demonstrate this through a multivariate analysis in section 4. Several combinations of jet discrimination variables are tried out, and an attempt is made to determine an optimum choice. Even though we include standard discrimination variables like the number of charged tracks as inputs to our multivariate analysis, it should be emphasized that they are subject to MC ambiguities stemming from parton shower algorithms and their associated parameters, and tunings of hadronization and underlying event (UE) models. However, in order to judge the improvement in tagger performance on using the associated jet rates, we compare the performance of different sets of variables within the same MC.

In sections 5 and 6 we study the use of the number of subjets of a jet (defined with an exclusive k_t algorithm) in place of the number of charged tracks, since the different MC prediction tend to be similar for the former observable. We compute the subjet rates upto NDLA as well, and compare the NDLA results with predictions from different MC's. Our results on both associated jets and subjets are summarized in section 7. We discuss the 2-dimensional joint distributions of the three discrimination variables used as inputs in the

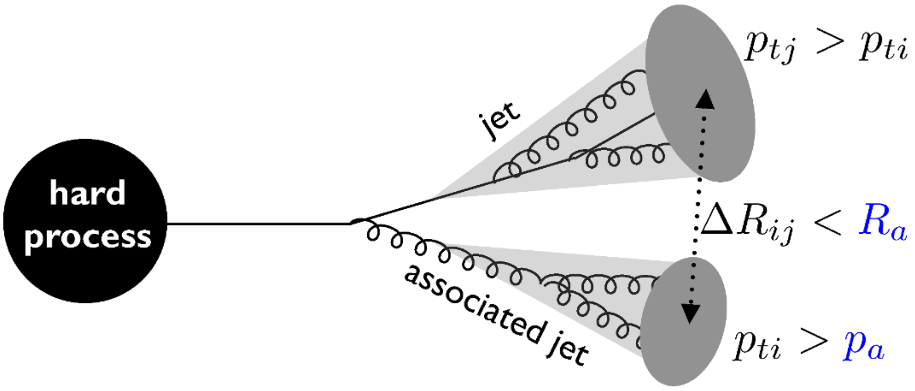

Figure 1. A schematic illustration of associated jets, and the relevant variables which determine the associated jet rate (see text for details).

multivariate analysis in an appendix.

2 Associated jet rates: analytical calculations

To begin with, let us define the longitudinally invariant jet algorithms [10–13] adopted in this study. The distance measures between each pair of objects i and j (d_{ij}), and between an object and the beam (d_{iB}) are given by

$$\begin{aligned} d_{ij} &= \min\{p_{ti}^{2p}, p_{tj}^{2p}\}\frac{\Delta R_{ij}^2}{R^2}\,, \\ d_{iB} &= p_{ti}^{2p}\,, \end{aligned} \tag{2.1}$$

where p_{ti}, y_i and ϕ_i are the transverse momentum, rapidity and azimuth of object i, respectively, $\Delta R_{ij}^2 \equiv (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, and R is the jet radius parameter. The jet algorithm in use is fixed by the parameter p, with $p = 1, 0, -1$ for the k_t [11], Cambridge/Aachen [14–16] and anti-k_t [13] algorithms, respectively. At any stage of clustering, if a d_{ij} is the smallest measure we combine objects i and j. If d_{iB} is the smallest we call i a jet and remove it from the clustering list. This procedure is continued until there are no more objects left to cluster.

Once a primary jet j has been defined, say using the anti-k_t algorithm with radius parameter R, we define a nearby jet i with $p_{tj} > p_{ti} > p_a$ and $R < \Delta R_{ij} < R_a$ as an *associated jet*. Thus the associated jet rates are functions of the primary jet $p_t = p_j$, its radius R, the association radius R_a and the minimum associated jet $p_t = p_a$. In figure 1 we illustrate the idea of an associated jet schematically, and show the relevant variables that determine the associated jet rate.

In perturbative QCD, the rate of n-jet production from a primary object of type i ($i = q, g$ in this case), R_n^i, can be obtained from the associated generating function [17–21]

$$\Phi_i(u) = \sum_n R_n^i u^n\,. \tag{2.2}$$

We can recover the jet rates by differentiating at $u = 0$,

$$R_n^i = \frac{1}{n!} \left. \frac{d^n \Phi_i}{du^n} \right|_{u=0} . \tag{2.3}$$

The jet rates $R_n^i = R_n^i(p_j, \xi)$ are functions of the trigger jet transverse momentum p_j, and the evolution scale for parton showering, which, for hadron-hadron collisions is taken as $\xi = \Delta R^2/2$. This is equivalent to the evolution scale for coherent parton showering, $\xi \equiv 1 - \cos\theta$, with θ being the emission angle ($\Delta R^2/2 \approx \theta^2/2 \approx 1 - \cos\theta$). To be resolved, an emission must have $\xi > \xi_j = R^2/2$ and $p_t > p_a$. Since the jet rates R_n^i include the trigger jet j, the probability of n *associated jets* for a jet of type i with transverse momentum p_j is

$$P_n^i = R_{n+1}^i(p_j, \xi_a) . \tag{2.4}$$

Here, $\xi_a = R_a^2/2$, with R_a being the association radius defined above.

The generating functions $\Phi_i(u)$ were computed in the context of e^+e^- collisions in ref. [17], upto next-to-double logarithmic accuracy (NDLA). Here, leading double and next-to-double logarithms refer to $\alpha_{\rm S}^n \log^{2n}$ and $\alpha_{\rm S}^n \log^{2n-1}$, where the logarithms are those of R_a/R and/or p_j/p_a. For p_a sufficiently large, these terms are determined by the timelike showering of final-state partons, while contributions from initial-state showers and the underlying event can be avoided. Following the same methods as in ref. [17] for hadron hadron collisions, for $\xi > \xi_j$ and $p_j > p_a$, we have the quark and gluon generating functions to NDLA

$$\begin{aligned}
\Phi_q(u, p_j, \xi) &= u + \int_{\xi_j}^{\xi} \frac{d\xi'}{\xi'} \int_{p_a/p_j}^{1} dz \frac{\alpha_{\rm S}(k_t^2)}{2\pi} P_{gq}(z) \Phi_q(u, p_j, \xi') \left[\Phi_g(u, zp_j, \xi') - 1\right] , \\
\Phi_g(u, p_j, \xi) &= u + \int_{\xi_j}^{\xi} \frac{d\xi'}{\xi'} \int_{p_a/p_j}^{1} dz \frac{\alpha_{\rm S}(k_t^2)}{2\pi} \Big\{ P_{gg}(z) \Phi_g(u, p_j, \xi') \left[\Phi_g(u, zp_j, \xi') - 1\right] \\
&\qquad + P_{qg}(z) \left[\{\Phi_q(u, p_j, \xi')\}^2 - \Phi_g(u, p_j, \xi')\right] \Big\} .
\end{aligned} \tag{2.5}$$

Here, the running coupling is evaluated at the transverse momentum scale of the emission, $k_t^2 = z^2 p_j^2 \xi'$. Defining $\overline{\alpha}_{\rm S} = \alpha_{\rm S}(p_j^2 \xi)/\pi$, i.e. in terms of the coupling at the hard scale, we have to NDLA

$$\frac{\alpha_{\rm S}(k_t^2)}{\pi} = \overline{\alpha}_{\rm S} - b_0 \overline{\alpha}_{\rm S}^2 \left[2 \ln z + \ln\left(\frac{\xi'}{\xi}\right) \right] , \tag{2.6}$$

with $b_0 = (11C_A - 2n_f)/12$.

The solution for the quark generating function is easily seen to be

$$\Phi_q(u, p_j, \xi) = u \exp\left\{ \int_{\xi_j}^{\xi} \frac{d\xi'}{\xi'} \int_{p_a/p_j}^{1} dz \frac{\alpha_{\rm S}(k_t^2)}{2\pi} P_{gq}(z) \left[\Phi_g(u, zp_j, \xi') - 1\right] \right\} . \tag{2.7}$$

We can solve for the gluon generating function by iteration, and then substitute in this equation to get the complete solution. For brevity we define the following logarithms:

$$\begin{aligned}
\kappa &= \ln(p_j/p_a) , & \kappa' &= \ln(zp_j/p_a) , \\
\lambda &= \ln(\xi_a/\xi_j) = 2\ln(R_a/R) , & \lambda' &= \ln(\xi'/\xi_j) .
\end{aligned} \tag{2.8}$$

In terms of these variables the NDLA quark generating function is

$$\Phi_q(u,\kappa,\lambda) = u\exp\left\{\int_0^\lambda d\lambda' \int_0^\kappa d\kappa'\, \Gamma_q\left(\kappa',\lambda',\kappa,\lambda\right)\left[\Phi_g(u,\kappa',\lambda')-1\right]\right\} \tag{2.9}$$

where, including the full P_{gq} splitting function,[1]

$$\Gamma_q\left(\kappa',\lambda',\kappa,\lambda\right) = C_F\overline{\alpha}_{\rm S}\left[1-{\rm e}^{\kappa'-\kappa}+\frac{1}{2}{\rm e}^{2(\kappa'-\kappa)}\right] - C_F b_0 \overline{\alpha}_{\rm S}^2\left[2(\kappa'-\kappa)+\lambda'-\lambda\right]\,. \tag{2.10}$$

Defining similarly[2]

$$\begin{aligned}\Gamma_g(\kappa',\lambda',\kappa,\lambda) &= C_A\overline{\alpha}_{\rm S}\left[1-{\rm e}^{\kappa'-\kappa}+\frac{1}{2}{\rm e}^{2(\kappa'-\kappa)}-\frac{1}{2}{\rm e}^{3(\kappa'-\kappa)}\right] - C_A b_0\overline{\alpha}_{\rm S}^2\left[2(\kappa'-\kappa)+\lambda'-\lambda\right]\,,\\ \Gamma_f(\kappa',\kappa) &= \frac{n_f}{4}\overline{\alpha}_{\rm S}\left[{\rm e}^{\kappa'-\kappa}-2{\rm e}^{2(\kappa'-\kappa)}+2{\rm e}^{3(\kappa'-\kappa)}\right]\,,\end{aligned} \tag{2.11}$$

we solve the gluon generating function by iteration to second order in u, which gives the probabilities for 0 or 1 associated jets:

$$\begin{aligned}\Phi_g(u,\kappa,\lambda) = u\Delta_g(\kappa,\lambda)\Big\{1+u\int_0^\lambda d\lambda'\int_0^\kappa d\kappa'\,&\left[\Gamma_g(\kappa',\lambda',\kappa,\lambda)\,\Delta_g(\kappa',\lambda')\right.\\ &\left.+\Gamma_f(\kappa',\kappa)\Delta_f(\kappa,\lambda')\right]+\mathcal{O}(u^2)\Big\}\,,\end{aligned} \tag{2.12}$$

where $\Delta_q(\kappa,\lambda)$ and $\Delta_g(\kappa,\lambda)$ are the quark and gluon Sudakov factors (the probabilities for no associated jets) and we have defined $\Delta_f(\kappa,\lambda)=\Delta_q^2(\kappa,\lambda)/\Delta_g(\kappa,\lambda)$. Hence

$$\begin{aligned}P_0^q &= \Delta_q(\kappa,\lambda) = \exp\left\{-\int_0^\lambda d\lambda'\int_0^\kappa d\kappa'\,\Gamma_q(\kappa',\lambda',\kappa,\lambda)\right\}\\ &= \exp\Big\{-C_F\overline{\alpha}_{\rm S}\lambda\left[\kappa-\frac{3}{4}+{\rm e}^{-\kappa}-\frac{1}{4}{\rm e}^{-2\kappa}\right]-C_F b_0\overline{\alpha}_{\rm S}^2\kappa\lambda\left[\kappa+\frac{1}{2}\lambda\right]\Big\}\,, \qquad (2.13)\\ P_0^g &= \Delta_g(\kappa,\lambda) = \exp\left\{-\int_0^\lambda d\lambda'\int_0^\kappa d\kappa'\left[\Gamma_g(\kappa',\lambda',\kappa,\lambda)+\Gamma_f(\kappa',\kappa)\right]\right\}\\ &= \exp\Big\{-C_A\overline{\alpha}_{\rm S}\lambda\left[\kappa-\frac{11}{12}+{\rm e}^{-\kappa}-\frac{1}{4}{\rm e}^{-2\kappa}+\frac{1}{6}{\rm e}^{-3\kappa}\right]\\ &\qquad -\frac{n_f}{4}\overline{\alpha}_{\rm S}\lambda\left[\frac{2}{3}-{\rm e}^{-\kappa}+{\rm e}^{-2\kappa}-\frac{2}{3}{\rm e}^{-3\kappa}\right]-C_A b_0\overline{\alpha}_{\rm S}^2\kappa\lambda\left[\kappa+\frac{1}{2}\lambda\right]\Big\}\,, \qquad (2.14)\\ P_1^q &= \Delta_q(\kappa,\lambda)\int_0^\lambda d\lambda'\int_0^\kappa d\kappa'\,\Gamma_q(\kappa',\lambda',\kappa,\lambda)\,\Delta_g(\kappa',\lambda') \qquad (2.15)\\ P_1^g &= \Delta_g(\kappa,\lambda)\int_0^\lambda d\lambda'\int_0^\kappa d\kappa'\big[\Gamma_g(\kappa',\lambda',\kappa,\lambda)\Delta_g(\kappa',\lambda')+\Gamma_f(\kappa',\kappa)\Delta_f(\kappa,\lambda')\big]. \qquad (2.16)\end{aligned}$$

[1] We keep terms that are formally power-suppressed in order to satisfy the boundary condition $P_0 = 1$ when $p_a = p_j$.

[2] We drop the $\overline{\alpha}_{\rm S}^2$ term in Γ_f as it is beyond NDLA and does not affect the boundary condition.

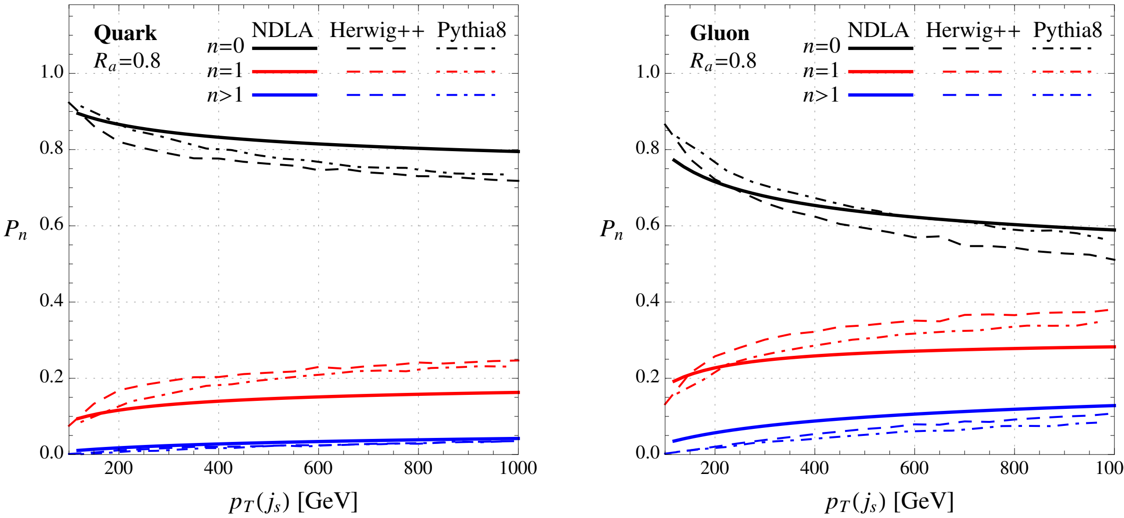

Figure 2. Comparison of the `Herwig++` and `Pythia8` MC predictions for associated jet rates with the NDLA results, as a function of $p_T(j_s)$: for quark jets (left), and gluon jets (right), with $R_a = 0.8$ and $p_a = 20\,\text{GeV}$. Here, $p_T(j_s)$ is the vector sum of the leading jet and associated jet p_T's.

3 Associated jet rates: comparison with Monte Carlo

We are now in a position to compare the NDLA predictions for associated jet rates discussed in the previous section with the results obtained using the `Herwig++` [22] and `Pythia8` [23, 24] event generators,[3] where the quark- and gluon-initiated jets are simulated using the $Z+q$ and $Z+g$ processes at leading order in QCD (with the Z boson subsequently decayed to $\nu\bar{\nu}$). The event samples were generated for proton-proton collisions at the 13 TeV LHC, using the `CTEQ6L1` [25] parton distribution functions (PDF) for the `Pythia` generators and the default `MRST LO`** [26] PDF and UE model for `Herwig++`. Subsequently, we used a modified version of `DELPHES2` [27] for including detector effects. For observables based on charged tracks to be discussed in the following, we use a minimum p_T threshold of 1 GeV for each track. All jets are reconstructed with an anti-k_t algorithm [13, 28, 29] with radius parameter $R = 0.4$, and are required to have $p_T > 20\,\text{GeV}$. In addition, the leading jet is required to be central with $|\eta| < 2$.

In figure 2 we show the probability of obtaining n associated jets P_n as a function of the jet p_T for $n = 0, 1$ and $n > 1$, for quark- and gluon-initiated jets, in the left and right columns respectively. The association radius is set to be $R_a = 0.8$ and the minimum associated jet transverse momentum is $p_a = 20\,\text{GeV}$. In the MC simulations, P_n has been computed as a function of $p_T(j_s)$, which is the vector sum of the leading jet and associated jet p_T's. The jet rates are studied as a function of $p_T(j_s)$, as it is closer to the transverse momentum of the parton that initiates the final state shower.

We see that the functional behaviour with respect to the jet p_T in the MC computation[4]

[3] To be specific, we use `Herwig++ 2.7.0` and `Pythia 8.201` (tune `4C`) for all our calculations.

[4] For the associated jet rate calculations, we generated MC event samples with a statistics of 20,000 events each fixing the threshold for the minimum leading jet p_T at $50 \times (i+1)\,\text{GeV}$, for $i \in [0, 19]$. Only

and the NDLA calculation are similar, although there are some differences in the values of P_n. In particular, the MC prediction of P_1 for quark and gluon jets is higher than the NDLA result, especially at higher $p_T(j_s)$, with Herwig++ giving rise to a slightly larger P_1 compared to Pythia8. For a quark jet, the probability of having at least one associated jet ranges from around 15% to 25% as we go from $p_T(j_s) = 200\,\text{GeV}$ to $p_T(j_s) = 500\,\text{GeV}$ and at higher $p_T(j_s)$ the probability essentially remains the same. For gluon jets, the corresponding probability ranges from around 30% to 40% as we go from $p_T(j_s) = 200\,\text{GeV}$ to $p_T(j_s) = 500\,\text{GeV}$. The larger probability to have an associated jet around a gluon can thus be utilized to better discriminate it from quarks, as we shall see in the next section.

The NDLA computation includes only the time-like showering of the final state partons, and ignores some power-suppressed effects due to momentum conservation and hadronization. On the other hand, the MC results shown above include momentum conservation and hadronization as well as the effects of initial state radiation (ISR) and multiple interaction (MPI). In order to quantify the effect of ISR and MPI, we compare the predictions for P_n with and without ISR and MPI in Herwig++, Pythia8 as well as in Pythia6 [30] (we use the version Pythia 6.4.28 with the AUET2B-CT6L tune) in figure 3. It is clear from this figure that the impact of ISR and MPI is rather small for our choice of the association radius $R_a = 0.8$, thereby making the predictions stable against such effects. For this choice of R_a, we can see that Pythia8 shows the highest variation against such effects, followed by Pythia6, while the effects are indeed negligible for the case of Herwig++.[5] Furthermore, the MC results become closer to the NDLA ones when ISR and MPI effects are switched off.

We also investigated the effects of momentum conservation, by changing the recombination scheme in the anti-k_t jet algorithm from the default E-scheme to the "winner-take-all" scheme introduced in [31], which is less sensitive to recoils in the parton shower [32]. Such a change increases the MC associated jet rates very slightly. We believe this is because the axis of the leading jet is moved away from the overall momentum vector of the system. The effects are roughly proportional for quark and gluon jets, so they would not affect discrimination significantly.

4 Quark-gluon separation: multivariate analysis

4.1 Variables for quark-gluon separation

A large number of variables have been surveyed in the context of quark-gluon discrimination, constructed out of either track based observables or calorimeter based ones [1–5]. While the former category has the practical advantage of being more accurate due to better track momentum resolution as well as being less prone to pile-up contamination, the latter category can be used for jets with larger rapidities outside the tracker coverage. The most widely studied variables include the number of charged tracks inside the jet cone

events with the leading jet $p_T(j_s)$ above the generation threshold are used in the analysis. This ensures uniform MC statistics in the whole range of $p_T(j_s)$.

[5] However, we have checked that if we take a larger association radius, $R_a > 1.2$, the ISR effects become appreciable in Herwig++.

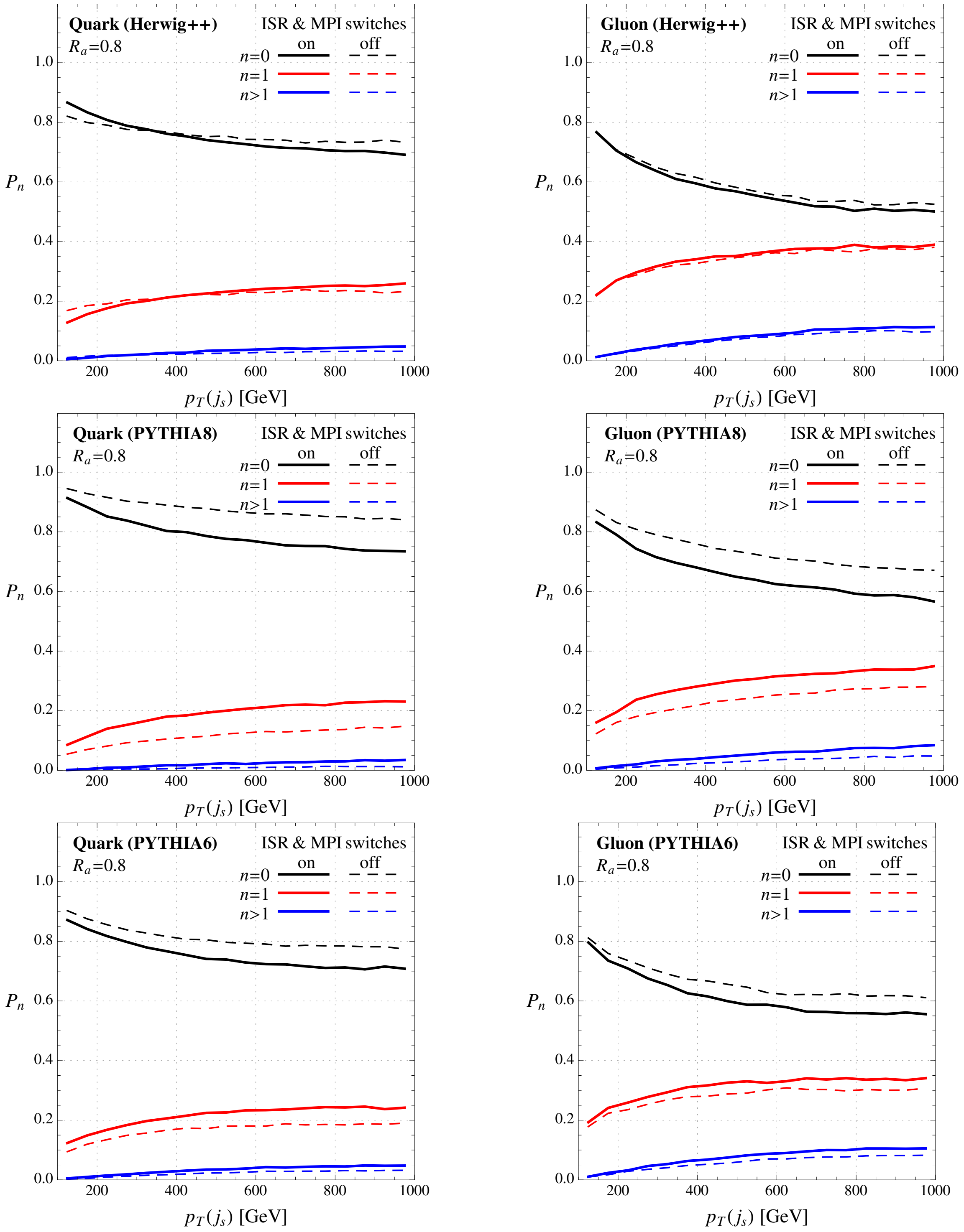

Figure 3. Comparison of the `Herwig++`, `Pythia8` and `Pythia6` predictions for associated jet rates with and without ISR and MPI, as a function of $p_T(j_s)$: for quark jets (left), and gluon jets (right). Here, $p_T(j_s)$ is the vector sum of the leading jet and associated jet p_T's.

(n_{ch}), the jet width [1] and energy-energy-correlation (EEC) angularity [4]. The jet width is defined as

$$w = \frac{\sum_i p_{T,i} \times \Delta R(i,\ \text{jet})}{\sum_i p_{T,i}} \tag{4.1}$$

where the sum goes over all the tracks associated to the jet. A similar track-based EEC variable, denoted by $C_1^{(\beta)}$ can be defined as

$$C_1^{(\beta)} = \frac{\sum_i \sum_j p_{T,i} \times p_{T,j} \times (\Delta R(i,j))^\beta}{(\sum_i p_{T,i})^2}. \tag{4.2}$$

Here again the sum over i and j run over all the tracks associated to the jet with $j > i$, while β is a tunable parameter. It has been demonstrated in ref. [3, 4] that smaller values of the exponent β leads to a better quark-gluon separation, and $\beta = 0.2$ is found to be optimal from perturbative calculations and MC studies based on `Herwig++` and `Pythia8` generators. We have compared the performance of the jet width variable w and the EEC variable $C_1^{(\beta=0.2)}$ in the multivariate analyses (MVA) to be discussed below, and find that in all cases $C_1^{(\beta=0.2)}$ leads to a better separation of gluons from quarks. Therefore, in the following, we only show results based on n_{ch} (with each charged track having $p_T > 1\,\text{GeV}$) and $C_1^{(\beta=0.2)}$. In addition, we shall include the associated jet information as well as the jet mass variable and compare the performance of the different MVA methods. As seen in the previous section, for $n = 1$ or $n > 1$, the probability of finding n associated jets, P_n, is significantly larger for gluon jets compared to quark-initiated ones across the whole p_T range of interest. Therefore, the presence (or absence) of an associated jet within a certain distance R_a of a high-p_T jet can be used to further improve the separation.

As the boundary between the signal and background regions in the hyper-surface spanned by the variables is non-linear, it is beneficial to adopt a multivariate analysis strategy as compared to a cut-based one. For this purpose, we employed a Boosted Decision Tree (BDT) algorithm with the help of the `TMVA-Toolkit` [33–35] in the `ROOT` framework. The training of the classifier was performed with $Z+q-$jet and $Z+g-$jet samples, and we generated the above MC samples uniformly distributed in jet-p_T.[6] The input variables for the two variable training are taken to be n_{ch} and $C_1^{(\beta=0.2)}$, while for three-variable trainings we further include the variable $m_J/p_{T,J}$, where m_J is the jet mass and $p_{T,J}$ is the transverse momentum of the leading jet. The information on the number of associated jets is included in the form of two categories ($n = 0$ or $n \geq 1$) in the MVA.

It should be emphasized that the MC prediction of the discrimination variables, especially the number of charged tracks n_{ch} is quite sensitive not only to the parton shower (PS) algorithm adopted and the related parameters, but also to the tuning of the hadronization and underlying event models. This is expected, since n_{ch} is not an infrared safe quantity, and only the ratio $n_{\text{ch}}^{\text{gluon}}/n_{\text{ch}}^{\text{quark}}$ converges rather slowly to the ratio of the colour factors C_A/C_F for high jet p_T [36, 37]. The disagreement between different MC's can therefore be reduced only by appropriate tuning at the LHC energies. With this limitation of the MC

[6] The MC event samples for the training of the classifier were generated in the same manner as for the associated jet rate computation in the previous section, but with a smaller step size of 10 GeV for the minimum $p_T(j_s)$ thresholds.

predictions in view, in this study, we compare the performance of different MVA methods within the same MC generator to estimate the improvement in adding associated jet related observables. We also show the quark-gluon separation as predicted by the different MC's for comparison. In appendix A we present details of the distributions of the discrimination variables and the differences between the MC predictions for them.

4.2 Performance in MVA

Based on the BDT analysis, we obtain the efficiencies of tagging quark (ϵ_q) and gluon jets (ϵ_g) as a function of the cut on the BDT score. It is more useful to compare the ratio of the tagging efficiencies as a function of ϵ_q, in order to judge the separation power of a "quark-rich signal" from a "gluon-rich" background. In figures 4–6 (left column) we show the ratio of the quark and gluon tagging efficiencies, ϵ_q/ϵ_g as a function of ϵ_q, for $400 < p_T(j_s) < 500\,\text{GeV}$, with the event samples generated with all the three MC codes. Four different MVA methods are shown corresponding to different choices for the discrimination variables:

- **Method-1:** Two variables, n_{ch} and C_1 with $\beta = 0.2$.
- **Method-2:** Two variables, n_{ch} and C_1 with $\beta = 0.2$, with two categories determined in terms the number of associated jets ($n = 0$ or $n \geq 1$).
- **Method-3:** Three variables, n_{ch}, C_1 with $\beta = 0.2$ and $m_J/p_{T,J}$.
- **Method-4:** Three variables, n_{ch}, C_1 with $\beta = 0.2$ and $m_J/p_{T,J}$, with two categories determined in terms the number of associated jets ($n = 0$ or $n \geq 1$).

We can quantify the improvement in quark-gluon separation using ϵ_g(Method-1)/ϵ_g(Method-{2,3,4}) as a function of ϵ_q, as shown in figures 4–6 (right). For example, for an operating point of $\epsilon_q = 0.4$, we can obtain an improvement of around $10\%, 15\%$ and 20% using Methods-2,3 and 4 respectively, when compared to Method-1. The differences between the improvement factors obtained using the three MC's are found to be small.

In order to estimate the change in tagger performance as we consider lower p_T jets, we show in figure 7 the same results as in figure 4, but now with $150 < p_T(j_s) < 200\,\text{GeV}$. The improvement on adding associated jet rates is still appreciable, although it is somewhat reduced compared to the higher p_T range. The fluctuations in the ϵ_g ratio for lower values of ϵ_q in figure 7 are due to low MC statistics.

We can see in figures 4–6 that there is an improvement in going from a two variable analysis to a three variable one by including the variable $m_J/p_{T,J}$. This can be understood as follows. The jet mass variable is related to $C_1^{(\beta=2)}$, as can be seen by writing both of them in terms of the z, θ variables for the hardest emission inside the jet cone: $m_J^2 \simeq z(1-z)\theta^2 p_{TJ}^2$. Furthermore, $C_1^{(\beta=2)}$ and $C_1^{(\beta=0.2)}$ are two independent variables belonging to the C_1 class which carry all the information on this hardest emission, and including both of them improves the tagger performance. For this reason, further addition of a third variable in the C_1 class does not change the performance appreciably, a fact that we explicitly checked by a separate MVA analysis. There is a further improvement in the

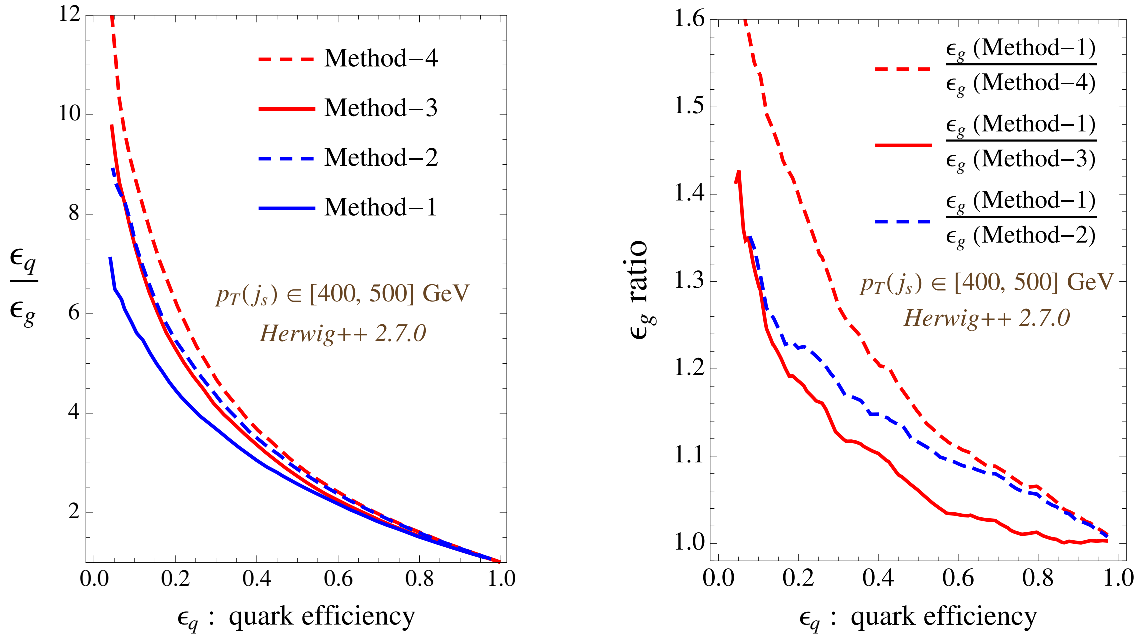

Figure 4. The ratio of the quark and gluon tagging efficiencies, ϵ_q/ϵ_g as a function of ϵ_q, for $400 < p_T(j_s) < 500\,\text{GeV}$, as determined by MC simulations with `Herwig++` (left column). The different MVA methods, determined in terms of the input variables are explained in the text. To quantify the improvement in quark gluon separation as we go to Methods 2,3 and 4, we show ϵ_g(Method-1)/ϵ_g(Method-{2,3,4}) as a function of ϵ_q as well (right column).

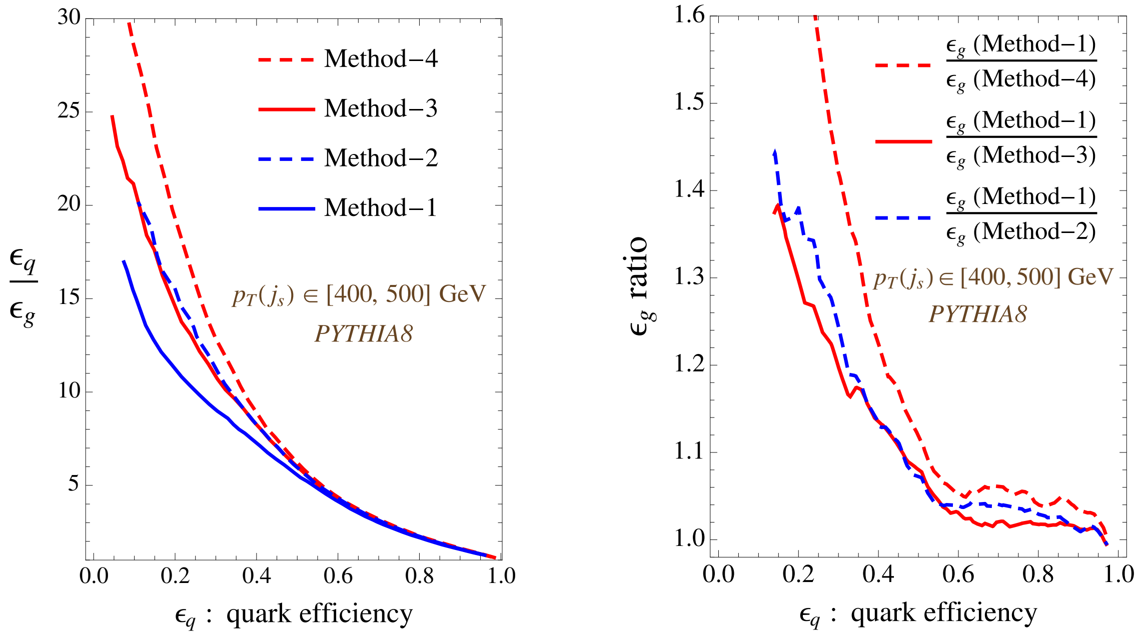

Figure 5. Same as figure 4, with MC simulations using `Pythia8`.

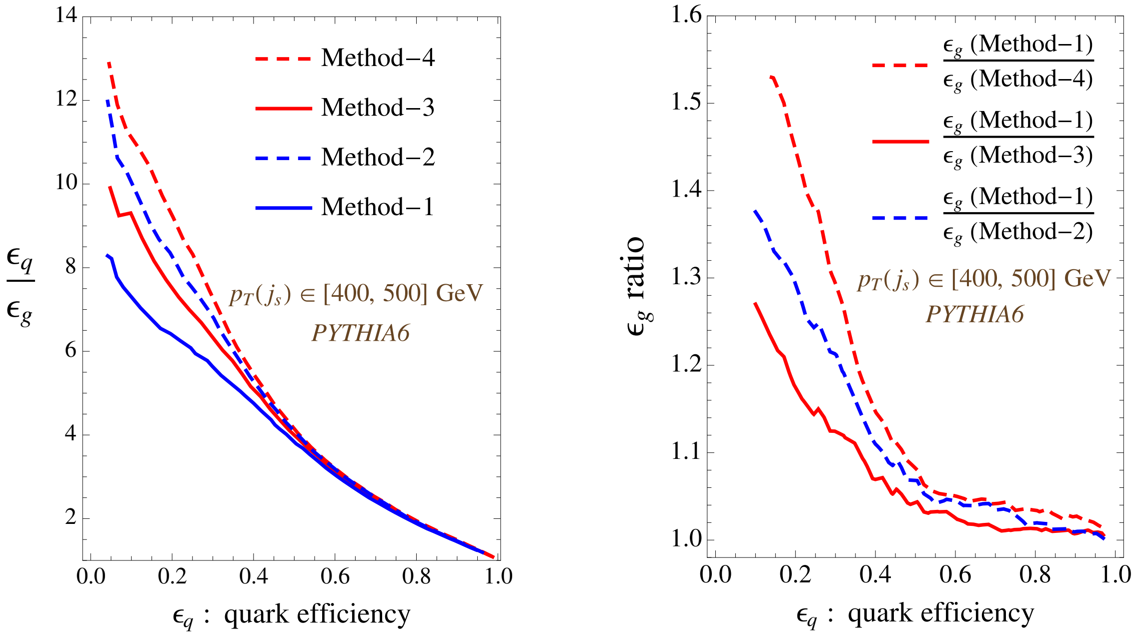

Figure 6. Same as figure 4, with MC simulations using `Pythia6`.

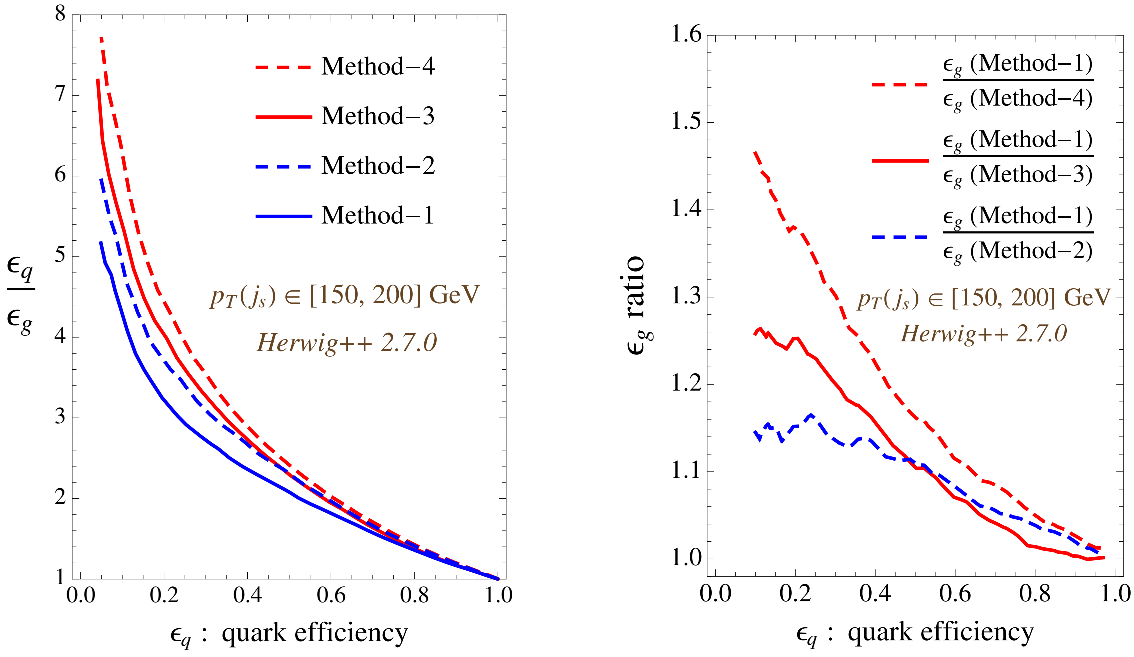

Figure 7. Same as figure 4, for a lower range of jet p_T, $150 < p_T(j_s) < 200\,\text{GeV}$. Results using only `Herwig++` are shown.

quark-gluon separation when the number of associated jets information is included at the level of categories in both the two and three variable MVA analyses. Since the associated jet rates carry the additional information of radiation outside the jet cone, Methods 2 and 4 lead to further improvements as compared to Methods 1 and 3, respectively.

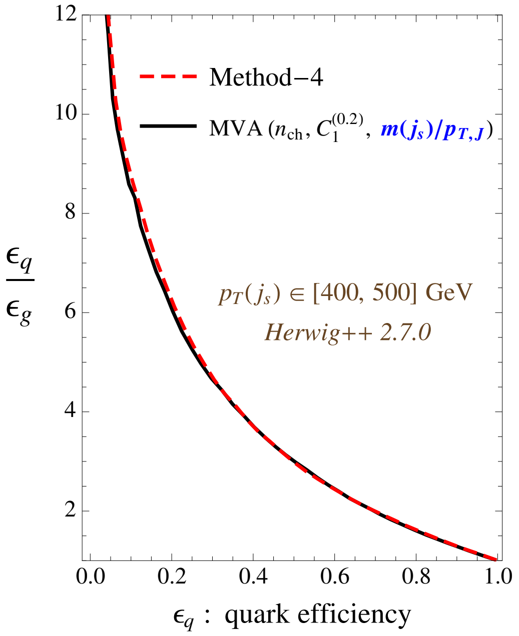

Figure 8. Comparison of Method 4 which includes $m_J/p_{T,J}$ and the associated jet rates as categories in the MVA, and the alternative method of including the associated jet rate information by using the modified jet mass variable $m(j_s)/p_{T,J}$. Both methods lead to the same MVA performance.

Method 4 leads to the best performance out of the four different MVA's considered. In fact, we find that there is an alternative way to include the associated jet rates information in Method 4 by using the modified jet mass variable $m(j_s)/p_{T,J}$ in Method 3. Here, $m(j_s)$ is the jet mass computed by adding the leading jet and associated jet four momenta. Because of a larger associated jet rate, for the same $p_T(j_s)$, $m(j_s)$ is higher for a gluon jet compared to a quark, while $p_{T,J}$ is lower. Therefore, using either associated jet rate categories and $m_J/p_{T,J}$, or using only the variable $m(j_s)/p_{T,J}$ leads to the same MVA performance, as shown in figure 8.

5 Subjet rates in jets: analytical calculations

The number of charged tracks inside a jet cone, $n_{\rm ch}$ (with each track having transverse momentum above a threshold, usually taken to be around 1 GeV) is often used as a good discriminating variable. However, as mentioned earlier, the MC predictions for this observable are quite sensitive not only to the parton shower (PS) algorithm and the related parameters, but also to the tuning of the hadronization and underlying event models. On the other hand, we find that the number of subjets of a primary jet leads to a more uniform prediction across the MC's, and thus can be better suited in quark gluon separation studies. The number of subjets as a quark-gluon separation variable was considered earlier in ref. [1]. In this study, we compute the subjet rates to NDLA accuracy, and show a detailed comparison with different MC's.

We find the *subjets* of jet j with the *exclusive k_t algorithm*, which applies the dimensionless distance measure

$$y_{ik} = \min\{p_{ti}^2, p_{tk}^2\} \frac{\Delta R_{ik}^2}{R^2 p_j^2} \,, \tag{5.1}$$

to its constituent objects and clusters them as discussed for a generalized k_t algorithm in section 2, until the smallest y_{ik} is above $y_{\rm cut}$. Thus the subjet rates are functions of the jet $p_t = p_j$, the jet radius R, and $y_{\rm cut}$.

In this section, we compute the subjet rates to NDLA, i.e. considering double and next-to-double logarithms, $\alpha_{\rm S}^n L^{2n}$ and $\alpha_{\rm S}^n L^{2n-1}$, where now $L = \ln(1/y_{\rm cut})$. The relevant generating functions in this case are those given in refs. [10, 21]:

$$\phi_q(u,Q) = u\,\Delta_q(Q) \exp\left(\int_{Q_0}^{Q} dq\,\Gamma_q(Q,q)\phi_g(u,q)\right) \,, \tag{5.2}$$

$$\phi_g(u,Q) = u\,\Delta_g(Q) \exp\left(\int_{Q_0}^{Q} dq\left[\Gamma_g(Q,q)\phi_g(u,q) + \Gamma_f(q)\frac{\phi_q(u,q)^2}{\phi_g(u,q)}\right]\right) \tag{5.3}$$

where $Q = R\,p_j$ is the jet scale, $Q_0 = R\,p_j\sqrt{y}_{\rm cut}$ is the resolution scale,[7]

$$\Gamma_q(Q,q) = \frac{2C_F}{\pi}\frac{\alpha_{\rm S}(q^2)}{q}\left(\ln\frac{Q}{q} - \frac{3}{4} + \frac{q}{Q} - \frac{1}{4}\frac{q^2}{Q^2}\right) \,, \tag{5.4}$$

$$\Gamma_g(Q,q) = \frac{2C_A}{\pi}\frac{\alpha_{\rm S}(q^2)}{q}\left(\ln\frac{Q}{q} - \frac{11}{12} + \frac{q}{Q} - \frac{1}{4}\frac{q^2}{Q^2} + \frac{1}{6}\frac{q^3}{Q^3}\right) \,, \tag{5.5}$$

$$\Gamma_f(q) = \frac{n_f}{3\pi}\frac{\alpha_{\rm S}(q^2)}{q}\left(1 - \frac{3}{2}\frac{Q_0}{q} + \frac{3}{2}\frac{Q_0^2}{q^2} - \frac{Q_0^3}{q^3}\right) \,. \tag{5.6}$$

The Sudakov factors for no resolvable emission are now

$$\Delta_q(Q) = \exp\left(-\int_{Q_0}^{Q} dq\,\Gamma_q(Q,q)\right) \,, \tag{5.7}$$

$$\Delta_q(Q) = \exp\left(-\int_{Q_0}^{Q} dq\left[\Gamma_g(Q,q) + \Gamma_f(q)\right]\right) \,. \tag{5.8}$$

Hence the rates for 1, 2 or 3 subjets in a quark jet are:

$$\begin{aligned}
R_1^q &= \Delta_q(Q) \,, \\
R_2^q &= \Delta_q(Q)\int_{Q_0}^{Q} dq\,\Gamma_q(Q,q)\Delta_g(q) \,, \\
R_3^q &= \Delta_q(Q)\int_{Q_0}^{Q} dq\int_{Q_0}^{q} dq'\,\Gamma_q(Q,q)\Delta_g(q)\,\times \\
&\quad \left\{\left[\Gamma_q(Q,q') + \Gamma_g(q,q')\right]\Delta_g(q') + \Gamma_f(q')\Delta_f(q')\right\} \,,
\end{aligned} \tag{5.9}$$

[7] Here again we keep power-suppressed corrections in order to satisfy boundary conditions.

where $\Delta_f = \Delta_q^2/\Delta_g$, and for a gluon jet we have

$$\begin{aligned}
R_1^g &= \Delta_g(Q)\,, \\
R_2^g &= \Delta_g(Q)\int_{Q_0}^{Q} dq\,[\Gamma_g(Q,q)\Delta_g(q) + \Gamma_f(q)\Delta_f(q)]\,, \\
R_3^g &= \Delta_g(Q)\int_{Q_0}^{Q} dq \int_{Q_0}^{q} dq'\Big[\Gamma_g(Q,q)\Delta_g(q)\times \\
&\quad \big\{[\Gamma_g(Q,q') + \Gamma_g(q,q')]\,\Delta_g(q') + \Gamma_f(q')\Delta_f(q')\big\} + \Gamma_f(q)\Delta_f(q)\times \\
&\quad \big\{[\Gamma_g(Q,q') - \Gamma_g(q,q')]\,\Delta_g(q') + 2\Gamma_q(q,q')\Delta_q(q')\big\}\Big]\,. \qquad (5.10)
\end{aligned}$$

6 Subjet rates in jets: comparison with Monte Carlo

We now compare the above results with Monte Carlo predictions. MC samples of quark and gluon jets were prepared for the subjet analysis using the same setup as in the associated jet study in section 2, however, detector effects and minimum p_T cuts for the charged and neutral hadrons were not included for this analysis. In this sense, our study of the subjet rates should be taken as illustrative, and we do not include the subjet rates in an MVA analysis in this paper. As we shall see in the following, one needs to go down to at least $L = 4$ to have some discrimination power. This corresponds to going down to 0.1 for ΔR resolution, which is the typical size of calorimeter cells, although the ΔR separation of subjets would be larger when the subjet p_T is smaller compared to the primary jet p_T. Therefore, in a proper analysis, combining track and calorimeter information is essential, and a detailed experimental study is necessary, which is beyond the scope of this paper.

Figure 9 shows comparisons between the resummed results of eqs. (5.9), (5.10) and the MC results for jets with $p_{T,J} \in [500, 600]\,\mathrm{GeV}$ and $R = 0.4$. For quark jets the different MC's agree quite well with each other and with the resummed calculations, the MC predictions being somewhat below the resummed 1-subjet rate for $L > 4$, and vice-versa for 2 subjets. Hadronization effects are small for $L < 7$, after which the 1- and 2-subjet rates are suppressed and the higher subjet rates are therefore enhanced. At this value of $R\,p_{Tj}$, $L = 7$ corresponds to resolving subjets with $\min\{p_{ti}, p_{tj}\}\Delta R_{ij} \sim 6\,\mathrm{GeV}$.

For gluon jets the agreement between the resummed results and the Monte Carlos is still quite close for 1 subjet. For 2 and 3 subjets the peak rates are in roughly the same place but have higher values than the resummed ones, with the effect that the rate for 4 or more subjets is substantially suppressed. Once again the hadronization effects are small for $L < 7$, after which the 1- and 2-subjet rates are suppressed and the higher subjet rates are enhanced, actually bringing the latter into close agreement with the analytical calculations.

In conclusion, the fairly good agreement between the Monte Carlos and the resummed 1-, 2- and 3-subjet rates for $R = 0.4$ and L not too large ($L < 5$, subjet resolution above about 15 GeV) suggests that in this range those subjet rates can be used for quark-gluon discrimination. At larger jet radii, the agreement remains similar, as we have checked using $R = 0.8$.

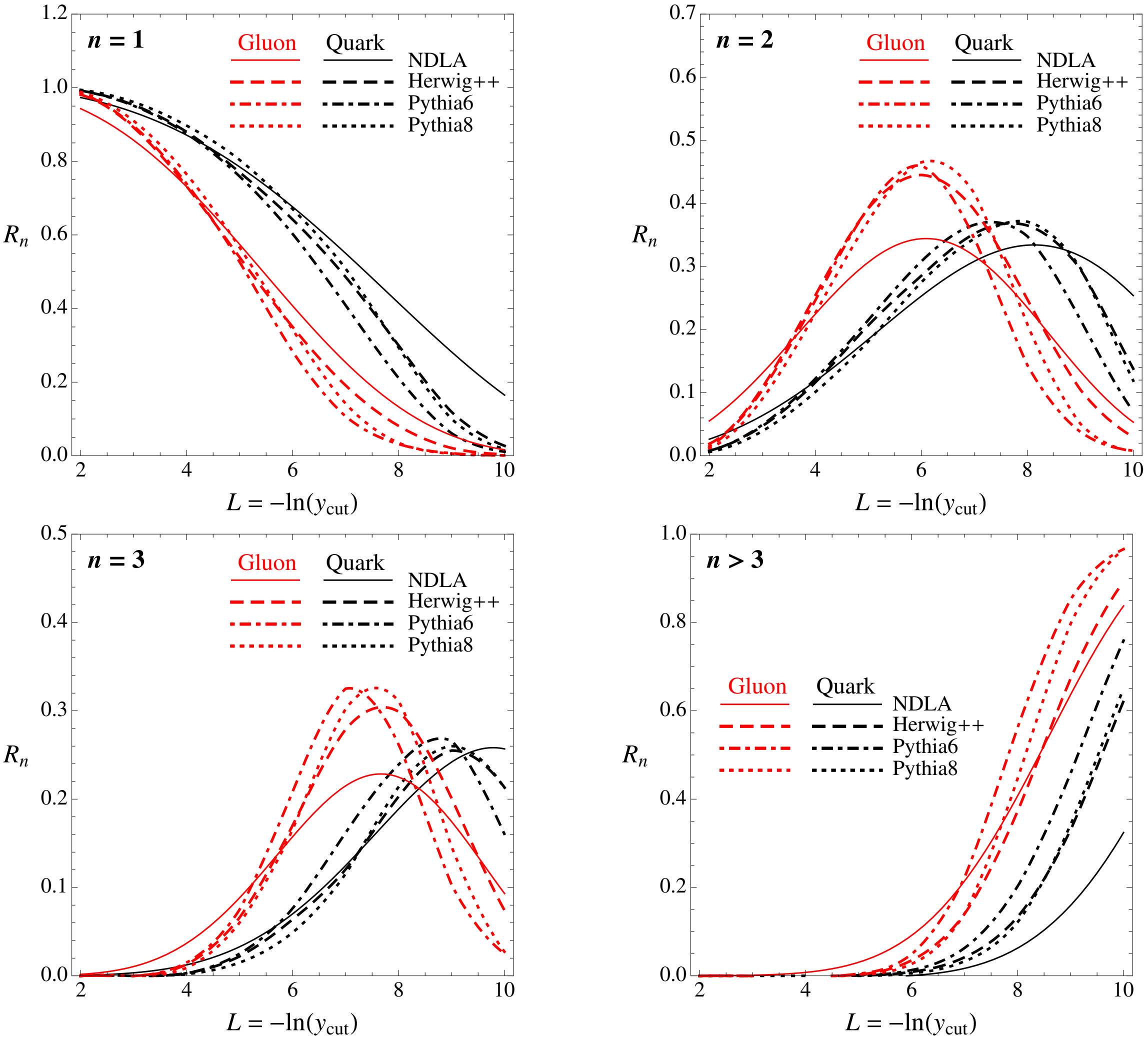

Figure 9. Subjet rates R_n with $n = 1, 2, 3$ and $n > 3$ as a function of $L = -\ln(y_{\text{cut}})$, for quark jets (black) and gluon jets (red), with $p_{T,J} \in [500, 600]$ GeV, $R = 0.4$. Curves are `Herwig++` (dashed), `Pythia6` (dot-dashed), `Pythia8` (dotted) and NDLA resummed (solid).

7 Summary

To summarize our findings, we show that in studies of light quark and gluon jet separation at the LHC, it is important to include the information on associated jet rates around a primary hard jet. Associated jet rates are defined as the probability of finding at least one softer reconstructed jet around the primary hard jet under consideration. This probability is found to be substantially higher for a gluon-initiated jet compared to a quark-initiated one. Since commonly a small jet radius parameter is adopted in LHC studies of hadronic jets, the associated jet rates carry the information on the radiation outside the chosen jet radius.

We compute the associated jet rates up to NDLA accuracy in perturbative QCD, as a function of the primary jet and minimum associated jet p_T's, as well as the jet radius and association radius parameters. The NDLA results are thereafter compared with predictions from different parton shower MC's. Since the NDLA predictions include only the time-like showering of the final state partons, we demonstrate the effects of ISR and MPI in the MC predictions as well, and it is observed that the NDLA predictions are closer to the MC's when ISR and MPI are switched off. Overall, the associated jet rates are not very sensitive to these effects as long as the association radius is not too large.

The probability of having at least one associated jet for a primary gluon jet is roughly a factor of two larger than for a quark jet, with a small variation in this number as a function of the jet p_T. This fact makes the presence or absence of associated jets a good variable for quark-gluon discrimination studies. We demonstrate the impact of including the associated jet rate information by including this variable in an MVA analysis, along with the well-studied variables of number of charged tracks, energy-energy-correlation angularities and jet mass. Comparing different two and three variable MVA's with and without the associated jet information, we find that including the associated jets leads to an improvement of around 10% in rejecting gluons, for a fixed quark selection efficiency of 0.4. We also show that using a three variable MVA with associated jet categories leads to the best performance, with an improvement of 20% in rejecting gluons, for the same quark efficiency as above.

Since for the number of charged tracks variable the MC predictions tend to differ, and are dependent on the parton shower and underlying event parameter tunes, we explore the number of k_t subjets of an anti-k_t jet as a quark-gluon separation variable. We compute the number of subjets to NDLA accuracy, and compare the resummed predictions with different MC's. The different MC predictions are found to be rather uniform, with the resummed predictions being broadly in agreement with them. However, for gluon jets the peak rates for 2 and 3 subjets are found to be lower in the resummed computation, which might arise due to higher-order effects that are in general bigger for gluons.

Acknowledgments

This work is supported by JSPS KAKENHI No. 26287039 and Grant- in-Aid for Scientific research from the Ministry of Education, Science, Sports, and Culture (MEXT), Japan, No. 23104006, and also by the World Premier International Research Center Initiative (WPI Initiative), MEXT, Japan. The work of B. Bhattacherjee is supported by the Department of Science and Technology, Government of INDIA, under the Grant Agreement numbers IFA13-PH-75 (INSPIRE Faculty Award). B.R. Webber thanks the Kavli IPMU for hospitality during part of this work.

A Distributions of discrimination variables

In figures 10–12 we show 2-dimensional plots of the joint distributions of the three discrimination variables used in the MVA presented in section 4, for the two Monte Carlo event generators `Herwig++` and `Pythia8`. The following features may be observed:

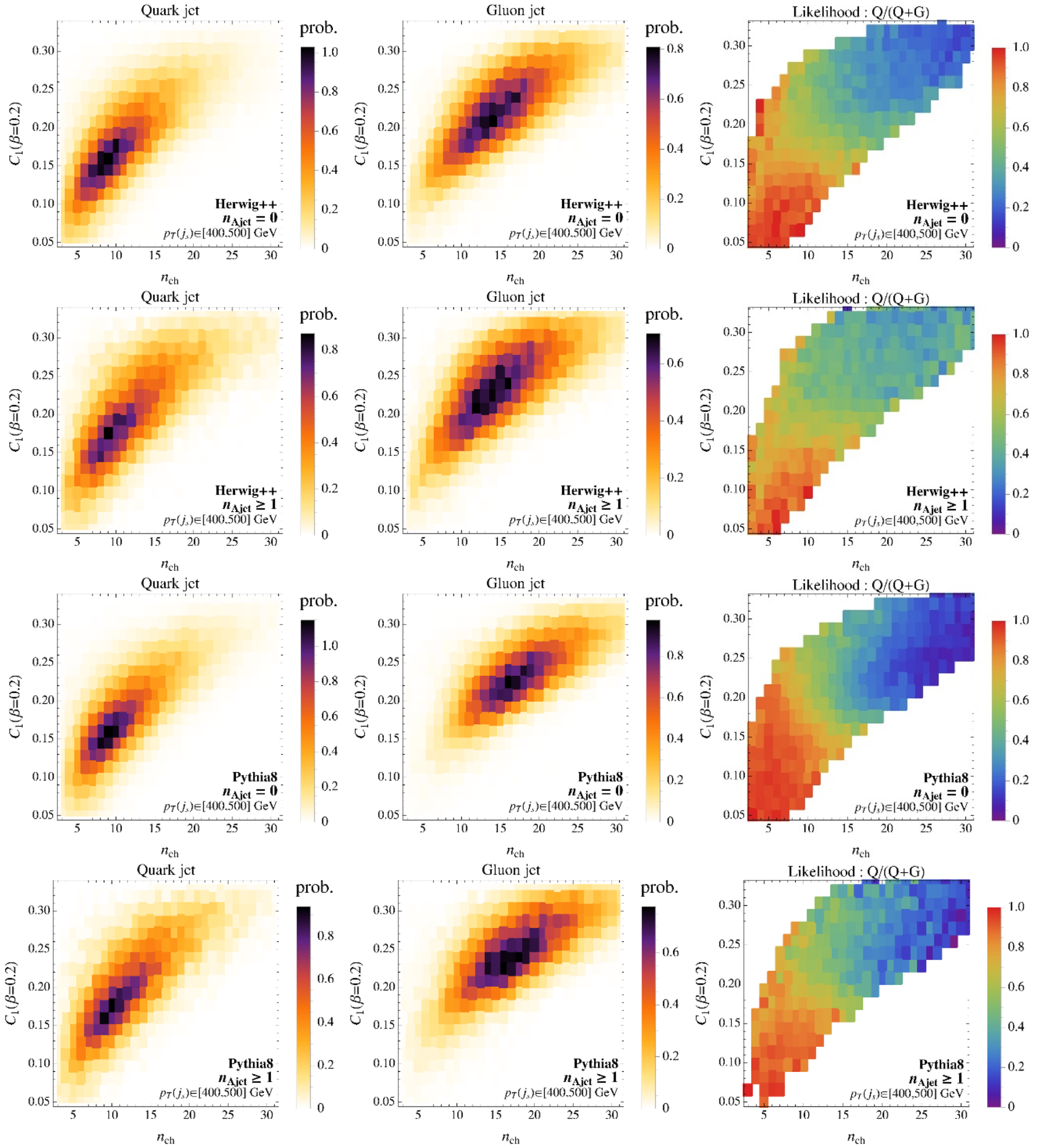

Figure 10. Joint distributions of $n_{\rm ch}$ and $C_1^{(\beta=0.2)}$ in Herwig++ and Pythia8, for quark and gluon jets with $p_T(j_s) \in [400, 500]$ GeV having $n_{\rm Ajet} = 0$ and ≥ 1 associated jets.

- There are differences between the distributions predicted by the two Monte Carlos, those of Pythia8 being somewhat narrower for quark jets and substantially narrower for gluon jets.

- The distributions of the infrared-unsafe variable $n_{\rm ch}$ show the greatest differences, with those of Pythia8 being larger at high $n_{\rm ch}$. This could be due to differences in tuning of the non-perturbative parameters of the generators.

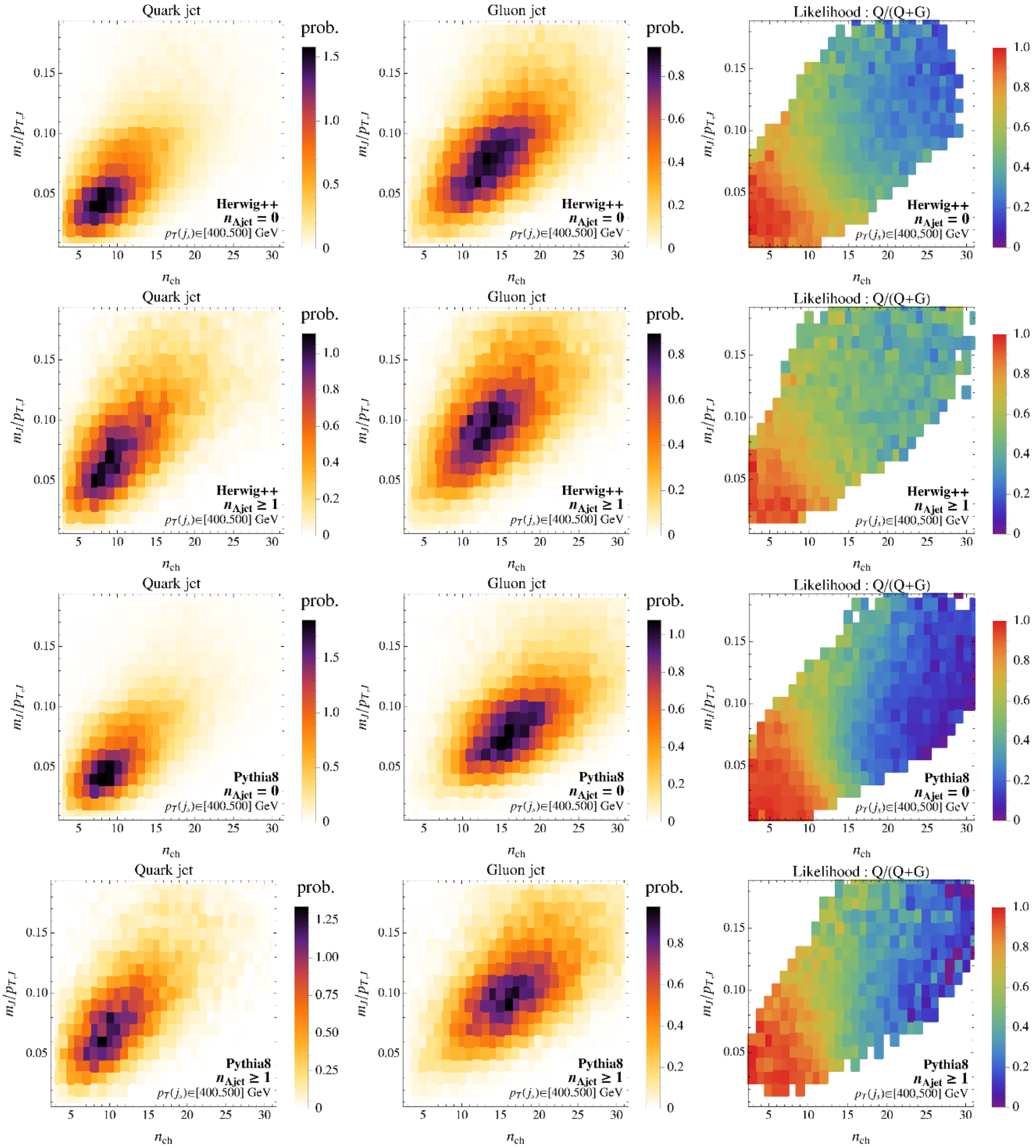

Figure 11. Joint distributions of $n_{\rm ch}$ and $m_J/p_{T,J}$ in `Herwig++` and `Pythia8`, for quark and gluon jets with $p_T(j_s) \in [400, 500]$ GeV having $n_{\rm Ajet} = 0$ and ≥ 1 associated jets.

- The above features are reflected in the likelihood plots, showing the probability ratio $P_q/(P_q + P_g)$, and account for the higher discrimination efficiency predicted by `Pythia8` (figure 5 vs figure 4).

- The quark-gluon discrimination in the events with associated jets is weaker than that for $n_{\rm Ajet} = 0$. This is expected because the events are selected according to $p_T(j_s)$, the sum of leading and associated jet p_T's. Therefore those with associated jets have leading jets with lower p_T's, which have lower discriminating power.

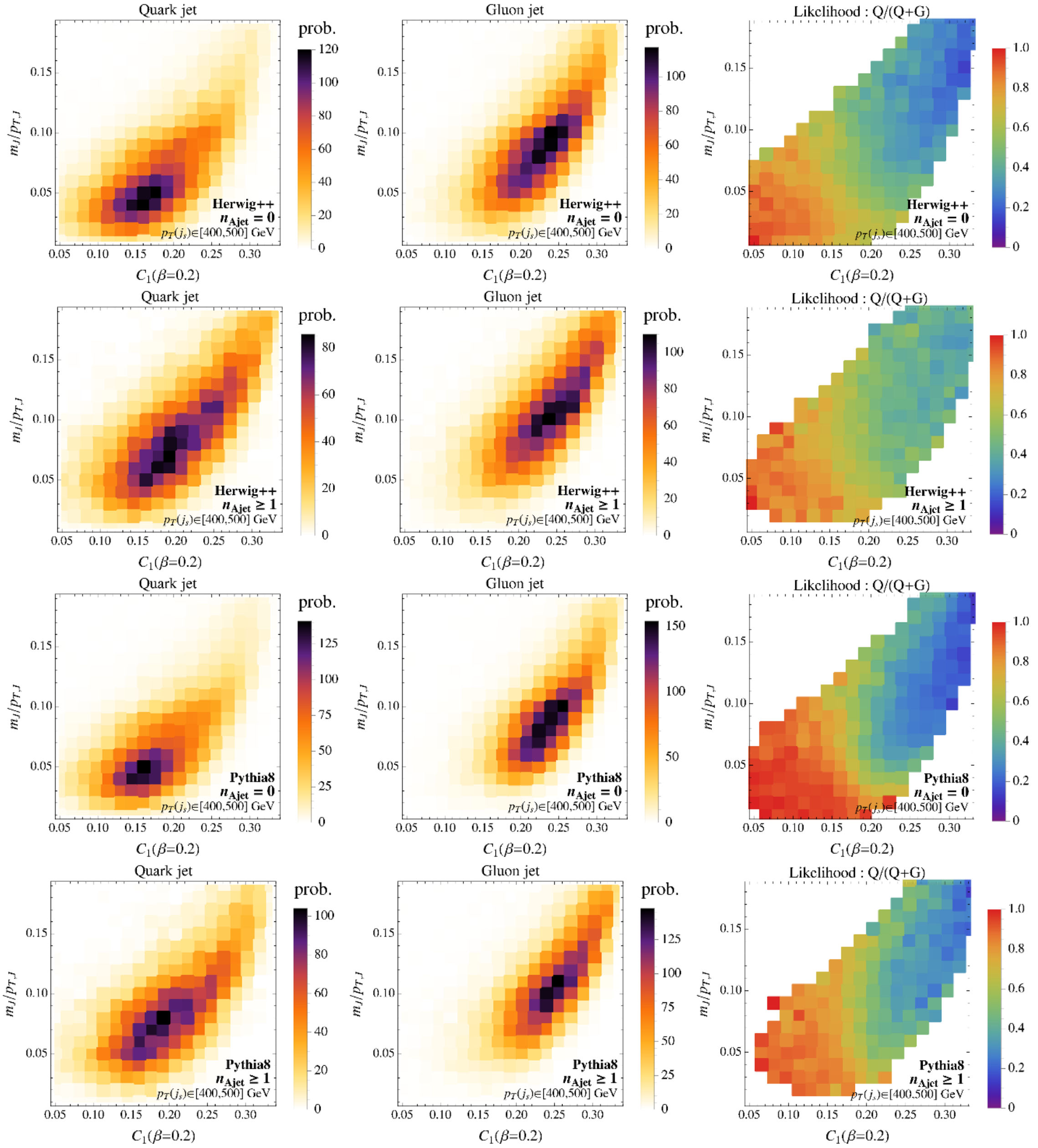

Figure 12. Joint distributions of $C_1^{(\beta=0.2)}$ and $m_J/p_{T,J}$ in Herwig++ and Pythia8, for quark and gluon jets with $p_T(j_s) \in [400, 500]$ GeV having $n_{\rm Ajet} = 0$ and ≥ 1 associated jets.

- Nevertheless the inclusion of the associated jet category improves the MVA performance, because the probability of an associated jet is lower for quark jets.

References

[1] J. Gallicchio and M.D. Schwartz, *Quark and Gluon Tagging at the LHC*, *Phys. Rev. Lett.* **107** (2011) 172001 [arXiv:1106.3076] [INSPIRE].

[2] J. Gallicchio et al., *Multivariate discrimination and the Higgs + W/Z search*, *JHEP* **04** (2011) 069 [arXiv:1010.3698] [INSPIRE].

[3] J. Gallicchio and M.D. Schwartz, *Quark and Gluon Jet Substructure*, *JHEP* **04** (2013) 090 [arXiv:1211.7038] [INSPIRE].

[4] A.J. Larkoski, G.P. Salam and J. Thaler, *Energy Correlation Functions for Jet Substructure*, *JHEP* **06** (2013) 108 [arXiv:1305.0007] [INSPIRE].

[5] A.J. Larkoski, J. Thaler and W.J. Waalewijn, *Gaining (Mutual) Information about Quark/Gluon Discrimination*, *JHEP* **11** (2014) 129 [arXiv:1408.3122] [INSPIRE].

[6] ATLAS collaboration, *Light-quark and gluon jet discrimination in pp collisions at* $\sqrt{s} = 7$ TeV *with the ATLAS detector*, *Eur. Phys. J.* **C 74** (2014) 3023 [arXiv:1405.6583] [INSPIRE].

[7] CMS collaboration, *Performance of quark/gluon discrimination in 8 TeV pp data*, CMS-PAS-JME-13-002 (2013).

[8] M. Dasgupta, L. Magnea and G.P. Salam, *Non-perturbative QCD effects in jets at hadron colliders*, *JHEP* **02** (2008) 055 [arXiv:0712.3014] [INSPIRE].

[9] G.P. Salam, *Towards Jetography*, *Eur. Phys. J.* **C 67** (2010) 637 [arXiv:0906.1833] [INSPIRE].

[10] S. Catani, Y.L. Dokshitzer, M. Olsson, G. Turnock and B.R. Webber, *New clustering algorithm for multi - jet cross-sections in* e^+e^- *annihilation*, *Phys. Lett.* **B 269** (1991) 432 [INSPIRE].

[11] S. Catani, Y.L. Dokshitzer, M.H. Seymour and B.R. Webber, *Longitudinally invariant* K_t *clustering algorithms for hadron hadron collisions*, *Nucl. Phys.* **B 406** (1993) 187 [INSPIRE].

[12] S.D. Ellis and D.E. Soper, *Successive combination jet algorithm for hadron collisions*, *Phys. Rev.* **D 48** (1993) 3160 [hep-ph/9305266] [INSPIRE].

[13] M. Cacciari, G.P. Salam and G. Soyez, *The Anti-k(t) jet clustering algorithm*, *JHEP* **04** (2008) 063 [arXiv:0802.1189] [INSPIRE].

[14] Y.L. Dokshitzer, G.D. Leder, S. Moretti and B.R. Webber, *Better jet clustering algorithms*, *JHEP* **08** (1997) 001 [hep-ph/9707323] [INSPIRE].

[15] M. Wobisch and T. Wengler, *Hadronization corrections to jet cross-sections in deep inelastic scattering*, hep-ph/9907280 [INSPIRE].

[16] M. Wobisch, *Measurement and QCD analysis of jet cross sections in deep-inelastic positron proton collisions at* $\sqrt{s} = 300$ *GeV*, DESY-THESIS-2000-049 [INSPIRE].

[17] E. Gerwick, S. Schumann, B. Gripaios and B. Webber, *QCD Jet Rates with the Inclusive Generalized kt Algorithms*, *JHEP* **04** (2013) 089 [arXiv:1212.5235] [INSPIRE].

[18] E. Gerwick and P. Schichtel, *Jet properties at high-multiplicity*, arXiv:1412.1806 [INSPIRE].

[19] K. Konishi, A. Ukawa and G. Veneziano, *Jet Calculus: A Simple Algorithm for Resolving QCD Jets*, Nucl. Phys. **B 157** (1979) 45 [INSPIRE].

[20] Y.L. Dokshitzer, V.A. Khoze, A.H. Mueller and S.I. Troian, *Basics of perturbative QCD*, Gif-sur-Yvette, France: Ed. Frontieres (1991).

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

[22] M. Bahr et al., *HERWIG++ Physics and Manual*, Eur. Phys. J. **C 58** (2008) 639 [arXiv:0803.0883] [INSPIRE].

[23] T. Sjöstrand, S. Mrenna and P.Z. Skands, *A Brief Introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852 [arXiv:0710.3820] [INSPIRE].

[24] T. Sjöstrand et al., *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159 [arXiv:1410.3012] [INSPIRE].

[25] J. Pumplin et al., *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **07** (2002) 012 [hep-ph/0201195] [INSPIRE].

[26] A. Sherstnev and R.S. Thorne, *Different PDF approximations useful for LO Monte Carlo generators*, arXiv:0807.2132 [INSPIRE].

[27] S. Ovyn, X. Rouby and V. Lemaitre, *DELPHES, a framework for fast simulation of a generic collider experiment*, arXiv:0903.2225 [INSPIRE].

[28] M. Cacciari, G.P. Salam and G. Soyez, *FastJet User Manual*, Eur. Phys. J. **C 72** (2012) 1896 [arXiv:1111.6097] [INSPIRE].

[29] M. Cacciari and G.P. Salam, *Dispelling the N^3 myth for the k_t jet-finder*, Phys. Lett. **B 641** (2006) 57 [hep-ph/0512210] [INSPIRE].

[30] T. Sjöstrand, S. Mrenna and P.Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **05** (2006) 026 [hep-ph/0603175] [INSPIRE].

[31] D. Bertolini, T. Chan and J. Thaler, *Jet Observables Without Jet Algorithms*, JHEP **04** (2014) 013 [arXiv:1310.7584] [INSPIRE].

[32] A.J. Larkoski, D. Neill and J. Thaler, *Jet Shapes with the Broadening Axis*, JHEP **04** (2014) 017 [arXiv:1401.2158] [INSPIRE].

[33] A. Hocker et al., *TMVA - Toolkit for Multivariate Data Analysis*, *PoS* **ACAT** (2007) 040 [physics/0703039] [INSPIRE].

[34] P. Speckmayer, A. Hocker, J. Stelzer and H. Voss, *The toolkit for multivariate data analysis, TMVA 4*, J. Phys. Conf. Ser. **219** (2010) 032057 [INSPIRE].

[35] http://tmva.sourceforge.net.

[36] P. Bolzoni, B.A. Kniehl and A.V. Kotikov, *Gluon and quark jet multiplicities at N^3LO+NNLL*, Phys. Rev. Lett. **109** (2012) 242002 [arXiv:1209.5914] [INSPIRE].

[37] P. Bolzoni, B.A. Kniehl and A.V. Kotikov, *Average gluon and quark jet multiplicities at higher orders*, Nucl. Phys. **B 875** (2013) 18 [arXiv:1305.6017] [INSPIRE].