

## Quark jet rates and quark-gluon discrimination in multijet final states

Yasuhito Sakaki\*

*Department of Physics, Korea Advanced Institute of Science and Technology, 291 Daehak-ro,  
Yuseong-gu, Daejeon 34141, Republic of Korea*



(Received 23 September 2018; published 18 June 2019)

We estimate the number of quark jets in QCD multijet final states at hadron colliders. In the estimation, we develop the calculation of jet rates into that of quark jet rates. From the calculation, we estimate the improvement on the signal-to-background ratio for a signal semianalytically by applying quark-gluon discrimination, where the signal predicts many quark jets. We introduce a variable related to jet flavors in multijet final states and propose a data-driven method using the variable. As with the semianalytical result, the improvements in the signal-to-background ratio using the variable in Monte Carlo analysis are estimated.

DOI: [10.1103/PhysRevD.99.114012](https://doi.org/10.1103/PhysRevD.99.114012)

### I. INTRODUCTION

So far we have not caught a clear sign of physics beyond the standard model at the LHC. We should maximize the discoverability of new physics at the LHC by using information of final states more precisely. In conventional analyses, we categorize events by inclusive variables like the number of jets, the scalar sum of the transverse momentum of jets and so on, then find signal regions using exclusive variables like the transverse momentum of objects, the distance between objects etc. We are able to access more specific features of events using jet substructure, and the related studies have developed dramatically in the last 10 years [1–6].

In recent studies, many methods and variables have been proposed to identify the origin of a jet using information about the jet’s substructure [7–21] including those based on resummation techniques with the assumption of soft-collinear factorization [22–42]. Applications of the jet substructure techniques to new physics searches at the LHC have also been considered (see e.g., Refs. [43–47]). In particular, numerous studies related to identifications of boosted jets originated from the top quark, Higgs and  $Z/W$  bosons have been performed. Some prominent variables were measured experimentally and the performance of tagging techniques has been tested. The tagging techniques are being used for new physics searches at the LHC [48–51]. These boosted jets have a multiprong structure inside the jet and we can distinguish these from QCD jets

using this feature, where “QCD jets” means quark- and gluon-initiated jets. There has also been progress in the studies of the separation between the quark jets and gluon jets. The performance of this separation and the shapes of variables have been measured [52–59]. In recent years, machine learning techniques have been used to improve the separation performance [60–72].

One of the differences between the boosted jet tagging technique and the quark-gluon discrimination is the size of the jet radius used in the analysis. The multiprong structure of boosted jets is formed by decays stemming from the electroweak interaction, and a large jet radius is basically required to catch most of the decay products. The QCD jets have one-prong structures such that there is a core parton carrying most of the energy of the jet, and the core is dressed in soft gluons radiated from itself. The main difference between quark jets and gluon jets stems from the difference between the color factors for the gluon radiation. Gluon jets emit more partons and wider radiation due to this difference. Neglecting the logarithmic scaling on the strong coupling and masses of the active quarks, the QCD radiation is scale invariant. That is, if one zooms in on a QCD jet, one will find a repeated self-similar pattern of jets within jets within jets, reminiscent of fractals. The difference exists even in the neighborhood of the jet core, and therefore the quark-gluon discrimination works out even if the jet radius is small.

Due to these properties, the quark-gluon discrimination is maximally utilized in multijet final states. In the case that a signal has  $n$  more quark jets compared to backgrounds, we naively expect that the signal-to-background ratio will increase  $(\epsilon_q/\epsilon_g)^n$  times using the quark-gluon tagging, where  $\epsilon_q$  and  $\epsilon_g$  are the quark and gluon jet efficiencies and  $\epsilon_q/\epsilon_g > 1$  under the assumption that quarks come from a signal.

\*yasuhito.sakaki@kaist.ac.kr

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI. Funded by SCOAP<sup>3</sup>.*

Many new physics models form multijet final states. For example, heavy colored resonances, like the gluino or squarks in supersymmetry (SUSY), could emit many partons via their cascade decays, and there are several studies that address these possibilities at the LHC [73–79]. Another example is searches of low-scale gravity which deals with the hierarchy problem concerned with the difference between the electroweak and Planck scale. These models can predict microscopic black holes or highly excited string states at the TeV scale. The objects emit a large number of energetic particles which are mostly quarks and gluons, and the phenomenon is constrained experimentally [80–84]. Moreover, the multijet final state is a good probe for the higher-dimensional operators which are caused by new color interactions at a high energy scale. There are two types of dimension-six pure QCD operators,  $g_s f_{abc} G_{av}^\mu G_{b\lambda}^\nu G_{c\mu}^\lambda$  and  $g_s^2 (\bar{q}\gamma_\mu T_a q)(\bar{q}\gamma^\mu T_a q)$ , and in particular the triple gluon field strength gets a large enhancement in high energy and large jet multiplicity regions [85,86]. The operator also predicts a characteristic quark-gluon jet fraction such that the  $G^3$  operator forms leading and subleading gluon jets although the leading and subleading jets tend to be valence quark jets in the standard model backgrounds.

We develop the calculation of jet rates into *quark jet rates*, and estimate quark-gluon jet fractions in the QCD multijet final states. Also, we consider a data-driven analysis for new physics searches in multijet final states. In the analysis, we introduce a variable defined in events having jets greater than or equal to  $n$ ,

$$d = \sqrt{\frac{1}{n} \sum_{i=1}^n Q_i^2}, \quad (1)$$

where  $Q_i (> 0)$  is assigned to the  $i$ th jet and it becomes larger when the jet substructure seems to be a quark jet. So,  $d$  gets larger if events have more quark jets. In conventional analyses of multijet final states, we fit a distribution of an inclusive variable like the scalar sum of the transverse momenta  $H_T$  in a control region and predict the number of background events in a signal region using the fit functions. We also fit the remaining rates of the number of events after imposing a  $d$  cut for each  $H_T$  bin as will be shown in Fig. 7 and find a big improvement over conventional analyses by using the new information.

This paper is organized as follows. In Sec. II, we calculate quark jet rates at hadron colliders based on the generating functional method. In Sec. III, we estimate how many quark jets are contained in the QCD multijet background using the formulas derived in the last section. Improvements of the signal-to-background ratio for the analysis in multijet final states by using the quark-gluon discrimination are also estimated semianalytically. In Sec. IV, the improvements using the variable  $d$  are

estimated in a Monte Carlo (MC) analysis. In Sec. V, we summarize our results and state our conclusions.

## II. QUARK JET RATES IN MULTIJET FINAL STATES

We first estimate how many quark jets are contained in the QCD multijet background at hadron colliders. The estimation is useful to know the impact of quark-gluon discrimination on new physics searches and helps to understand the results of the analysis. Assuming infinite computational resources, we can add any number of additional partons to parton showers using the matching schemes [87–91] in the simulation of multijet final states. However, we do not have such enormous computational resources. So, we use the generating functional method [92–95] based on the Dokshitzer-Gribov-Lipatov-Altarelli-Parisi (DGLAP) equations. All of the leading-logarithmic (LL) terms and some of the next-to-leading-logarithmic terms are taken into account in the calculation. Matrix element corrections for additional partons are absent in the calculations and those effects are examined in the Appendix A 2.

The definitions of quark and gluon jets are typically given by using parton-level information [96–98]. We also use the parton-level definition in this study. Although the definition is unphysical since jets are observed at the hadron level, we assume that the number of quark jets defined at the parton level is close to the number at the hadron level. It should be noted that a well-defined definition of quark-gluon jets at the hadron level was proposed recently in Ref. [99]. They determined that the quark jet fraction defined at the parton level can be extracted using hadron-level information. It has been demonstrated for the hardest jet in the  $Z + \text{jet}$  process and for the two hardest jets in the dijet process with a Monte Carlo event generator.

### A. Generating functionals

Conventionally, a generating functional for a final-state parton  $i$  is defined as,

$$\Phi_i(u, p, t) = \sum_{n=1}^{\infty} u^n \mathcal{R}_n^{(i,\text{out})}(p, t), \quad (2)$$

where  $p$  and  $t$  are the transverse momentum and energy scale for the parton  $i$ . We call a parton whose transverse momentum is larger than  $p_0$  and which is separated from other partons by  $R$  or more in the  $\eta$ - $\phi$  plane a jet, where the nonglobal logarithmic effect [100] is ignored. The jet rate  $\mathcal{R}_n^{(i,\text{out})}$  represents the probability that the parton  $i$  forms an  $n$ -jet configuration by final-state radiation [101–105]. We develop the definition,

$$\Phi_i(u, v, p, t) = \sum_{n=1}^{\infty} u^n \mathcal{R}_n^{(i,\text{out})}(v, p, t), \quad (3)$$

where the modified jet rates are given by a polynomial expression,

$$\mathcal{R}_n^{(i,\text{out})}(v, p, t) = \sum_{m=0}^n v^m R_{n,m}^{(i,\text{out})}(p, t). \quad (4)$$

We call  $R_{n,m}^{(i,\text{out})}$  the *quark jet rate* and it represents the probability that  $i$  forms an  $n$ -jet configuration in which  $m$  quark jets are contained.<sup>1</sup> The number of jet  $n$  starts from 1 since the final state itself becomes a jet even if it does not emit any resolved emissions. The jet rate  $R_n^{(i,\text{out})}$  in Eq. (2) is simply given as,

$$R_n^{(i,\text{out})}(p, t) = \mathcal{R}_n^{(i,\text{out})}(1, p, t). \quad (5)$$

We can acquire the quark jet rates by differentiating the functional as,

$$R_{n,m}^{(i,\text{out})}(p, t) = \frac{1}{n!m!} \frac{\partial^n}{\partial u^n} \frac{\partial^m}{\partial v^m} \Phi_i(u, v, p, t) \Big|_{u=v=0}. \quad (6)$$

Similarly, we introduce a generating functional for an initial-state parton  $i$ ,

$$\Psi_i(u, v, x, t) = \sum_{n=0}^{\infty} u^n \mathcal{R}_n^{(i,\text{in})}(v, x, t), \quad (7)$$

where

$$\mathcal{R}_n^{(i,\text{in})}(v, x, t) = \sum_{m=0}^n v^m R_{n,m}^{(i,\text{in})}(x, t). \quad (8)$$

Quark jet rates for the initial-state  $R_{n,m}^{(i,\text{in})}$  represent the probability that  $i$  emits  $n$  jets in which  $m$  quark jets are contained. The argument  $x$  is the energy fraction for  $i$ , and therefore, the parton carries an energy  $x p_{\text{beam}}$ , where  $p_{\text{beam}}$  is the hadron beam energy. The number of jets  $n$  starts from 0 since the initial state does not generate any jet if it does not emit any resolved emissions.

A generating functional for a hard process is given by a product of functionals for initial and final states. For example, a generating functional which has initial states  $i_1, i_2$  and final states  $f_1, f_2$  is given as  $\Phi = \Psi_{i_1} \Psi_{i_2} \Phi_{f_1} \Phi_{f_2}$ , and we can derive the quark jet rates for the hard process by differentiating the functional  $\Phi$ <sup>2</sup>.

For brevity, we omit the arguments  $u$  and  $v$  in the generating functionals below.

<sup>1</sup>You may prefer the definition of a functional such that  $\hat{\Phi}(u_q, u_g) = \sum_{n_q} \sum_{n_g} u_q^{n_q} u_g^{n_g} \hat{R}_{n_q, n_g}$ , where  $n_q$  and  $n_g$  are the number of quark and gluon jets, and  $\hat{R}_{n_q, n_g}$  is the probability that an event has  $n_q$  quark jets and  $n_g$  gluon jets. The two functionals are just related by  $\Phi(u, v) = \hat{\Phi}(u, uv)|_{n_q=n-n_g, n_g=m}$ .

<sup>2</sup>See in Eqs. (84) and (85).

## I. Evolution equations

We derive evolution equations for generating functionals of final and initial states. We first start with the final state. We get the following equations in the case that no resolved emission occurs, namely for  $n = 1$ :

$$\begin{aligned} \mathcal{R}_1^{(q,\text{out})}(v, p, t) &= v \tilde{\Delta}_q(p, t), \\ \mathcal{R}_1^{(g,\text{out})}(v, p, t) &= \tilde{\Delta}_g(p, t), \end{aligned} \quad (9)$$

where  $\tilde{\Delta}_i(p, t)$  are the Sudakov form factors which show the probability that any resolved emission does not happen between the starting scale  $t$  and a minimum resolved scale. We define the form factors in Sec. II A 2. If a quark does not emit any resolved emission it forms one quark jet, so we need  $v$  in front of  $\tilde{\Delta}_q$ . In the case that the resolved emission happens at least one time, namely for  $n > 1$ , the modified jet rates have the following equation:

$$\begin{aligned} \mathcal{R}_n^{(i,\text{out})}(v, p, t) &= \sum_k \int_{p_0/p}^1 \frac{dz}{z} \int_{t_0}^t \frac{dt'}{t'} \frac{\tilde{\Delta}_i(p, t)}{\tilde{\Delta}_i(p, t')} \tilde{\Gamma}_{i \rightarrow jk}(z, t') \\ &\times \sum_{n_1+n_2=n} \mathcal{R}_{n_1}^{(j,\text{out})}(v, p, t') \mathcal{R}_{n_2}^{(k,\text{out})}(v, zp, t'), \end{aligned} \quad (10)$$

where  $\tilde{\Gamma}_i(z, t) = \alpha_s(z, t) P_i(z) / \pi$ ,  $\alpha_s(z, t)$  and  $P_i(z)$  are the running strong coupling and the splitting functions,  $t_0$  is a given minimal scale, and  $p_0$  is the minimum resolved transverse momentum which corresponds to the minimum  $p_T$  cut for jets. We use the emission angle as the scale such that  $t = \sqrt{1 - \cos \theta} (\simeq \theta / \sqrt{2})$ , and thus  $t_0 = \sqrt{1 - \cos R}$ , where  $R$  corresponds to the jet radius. As said above, we simply define radiation whose transverse momentum and angle are greater than  $p_0$  and  $R$  as a jet in our calculation, and the jet algorithm dependence is beyond the scope of this paper. The ratio of Sudakov form factors shows the probability that the parton  $i$  does not emit any resolved emissions between the scales  $t$  and  $t'$ . From these equations, we get evolution equations for the generating functionals of final states [19,95] as,

$$\begin{aligned} \Phi_q(p, t) &= uv \tilde{\Delta}_q(p, t) + \int_{p_0/p}^1 \frac{dz}{z} \int_{t_0}^t \frac{dt'}{t'} \frac{\tilde{\Delta}_q(p, t)}{\tilde{\Delta}_q(p, t')} \\ &\times \tilde{\Gamma}_q(z, t') \Phi_q(p, t') \Phi_g(zp, t'), \end{aligned} \quad (11)$$

$$\begin{aligned} \Phi_g(p, t) &= u \tilde{\Delta}_G(p, t) + \int_{p_0/p}^1 \frac{dz}{z} \int_{t_0}^t \frac{dt'}{t'} \frac{\tilde{\Delta}_G(p, t)}{\tilde{\Delta}_G(p, t')} \\ &\times [\tilde{\Gamma}_g(z, t') \Phi_g(p, t') \Phi_g(zp, t') \\ &+ n_f \tilde{\Gamma}_{gg}(z, t') \Phi_q(p, t') \Phi_q(zp, t')], \end{aligned} \quad (12)$$

where  $\Phi_q$  and  $\Phi_g$  are the generating functionals for quarks and gluons. For brevity we define the following logarithms:

$$\kappa = \ln(p/p_0), \quad \kappa' = \ln(zp/p_0), \quad (13)$$

$$\lambda = \ln(t/t_0), \quad \lambda' = \ln(t'/t_0). \quad (14)$$

With these variables, the equations in Eqs. (11) and (12) are rewritten as,

$$\Phi_q(\kappa, \lambda) = uv\tilde{\Delta}_q(\kappa, \lambda) \times \exp\left[\int_0^\lambda d\lambda' \int_0^\kappa d\kappa' \tilde{\Gamma}_q(z, t') \Phi_g(\kappa', \lambda')\right], \quad (15)$$

$$\Phi_g(\kappa, \lambda) = u\tilde{\Delta}_G(\kappa, \lambda) \exp\left\{\int_0^\lambda d\lambda' \times \int_0^\kappa d\kappa' \left[\tilde{\Gamma}_g(z, t') \Phi_g(\kappa', \lambda') + n_f \tilde{\Gamma}_{q\bar{q}}(z, t') \frac{\Phi_q(\kappa, \lambda')}{\Phi_g(\kappa, \lambda')} \Phi_q(\kappa', \lambda')\right]\right\}. \quad (16)$$

We next derive generating functionals for initial states. For an initial state  $i$ , from the DGLAP equation, the normalized change of a parton density for  $i$ , in other words, the probability that  $i$  emits initial-state radiation between  $t'$  and  $t' + dt'$  is,

$$d\mathcal{P}(x, t') = \frac{df_i(x, t')}{f_i(x, t')} = \sum_k \frac{dt'}{t'} \int_x^1 \frac{dx'}{x'} \frac{\alpha_s}{\pi} P_{k \rightarrow ij}(z) \times \frac{f_k(x', t')}{f_i(x, t')}, \quad (17)$$

$$= \sum_k \frac{dt'}{t'} \mathcal{P}_{k \rightarrow ij}(z, t'), \quad (18)$$

where  $x$  and  $x'$  are the momentum fractions for  $i$  and its parent parton as shown in Fig. 1,  $f_i$  is the parton distribution function (PDF) for  $i$ , and  $z = (x' - x)/x'$ . In the case of initial-state radiation,  $\mathcal{R}_0^{(i, \text{in})}$  is given by the probability that  $i$  does not emit any resolved initial-state radiation,

$$\mathcal{R}_0^{(i, \text{in})}(v, x, t) = \exp\left(-\int_{t_0}^t d\mathcal{P}(x, t')\right) = \tilde{\Pi}_i(x, t), \quad (19)$$

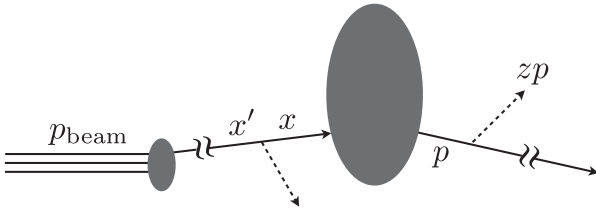


FIG. 1. Schematic illustration of initial- and final-state radiation. The central blob shows a hard process,  $p$  is the transverse momentum for a final state,  $z$  is the energy fraction for the final state, and  $x$  and  $x'$  are the momentum fractions for an initial state and its parent parton.

where  $\tilde{\Pi}_i$  is the Sudakov form factor for initial states. In the case that a resolved emission happens at least one time, namely for  $n > 0$ , the modified jet rates have the following relation:

$$\mathcal{R}_n^{(i, \text{in})}(x, t) = \sum_k \int_{t_0}^t \frac{dt'}{t'} \frac{\tilde{\Pi}_i(x, t)}{\tilde{\Pi}_i(x, t')} \mathcal{P}_{k \rightarrow ij}(z, t') \quad (20)$$

$$\times \sum_{n_1+n_2=n} \mathcal{R}_{n_1}^{(j, \text{in})}(x', t') \mathcal{R}_{n_2}^{(k, \text{out})}((x' - x)p_{\text{beam}}, t'). \quad (21)$$

From the above equations and the definitions of the generating functionals, the evolution equation for the functional of initial states is given as,

$$\Psi_i(x, t) = \Pi_i(x, t) + \sum_k \int_{t_0}^t \frac{dt'}{t'} \int_x^1 \frac{dx'}{zx'} \frac{\Pi_i(x, t)}{\Pi_i(x, t')} \frac{f_k(x', t)}{f_i(x, t)} \times \Gamma_{k \rightarrow ij}(z, t') \Psi_k(x', t') \Phi_j((x' - x)E_{\text{beam}}, t'). \quad (22)$$

The logarithms  $\kappa$  and  $\kappa'$  are modified for initial states as,

$$\bar{\kappa} = \ln((1 - x)p_{\text{beam}}/p_0), \quad \bar{\kappa}' = \ln((x' - x)p_{\text{beam}}/p_0). \quad (23)$$

In terms of these variables, the equations in Eq. (22) for quarks and gluons are,

$$\Psi_q(\bar{\kappa}, \lambda) = \tilde{\Pi}_q(\bar{\kappa}, \lambda) \exp\left[\int_0^\lambda d\lambda' \times \int_0^{\bar{\kappa}} d\bar{\kappa}' \left\{ \tilde{\Gamma}_q(z, t') \frac{f_q(x')}{f_q(x)} \Phi_g(\bar{\kappa}', \lambda') + \tilde{\Gamma}_{q\bar{q}}(z, t') \frac{f_g(x')}{f_q(x)} \frac{\Psi_g(x', t')}{\Psi_q(x, t')} \Phi_q(\bar{\kappa}', \lambda') \right\}\right], \quad (24)$$

$$\Psi_g(\bar{\kappa}, \lambda) = \tilde{\Pi}_g(\bar{\kappa}, \lambda) \exp\left[\int_0^\lambda d\lambda' \times \int_0^{\bar{\kappa}} d\bar{\kappa}' \left\{ \tilde{\Gamma}_g(z, t') \frac{f_g(x')}{f_g(x)} \Phi_g(\bar{\kappa}', \lambda') + \sum_q \tilde{\Gamma}_{gq}(z, t') \frac{f_q(x')}{f_g(x)} \frac{\Psi_q(x', t')}{\Psi_g(x, t')} \Phi_q(\bar{\kappa}', \lambda') \right\}\right], \quad (25)$$

where we neglect the scale dependence on the ratio of the PDF since the effect is negligible.<sup>3</sup>

<sup>3</sup>We fix the factorization scale for the PDF ratios to the hard scale, namely  $t' = t$ , in numerical calculations below.



The splitting kernels are summarized as follows:

$$\int \frac{dz}{z} \tilde{\Gamma}_i(z, t') = \int dz \frac{\alpha_s(k_t^2)}{\pi} P_i(z) \quad (26)$$

$$\simeq \int \frac{dz}{z} (1-D) \times \begin{cases} a_q, & i = q, q \rightarrow qg, \\ a_g, & i = g, g \rightarrow gg, \\ \frac{a_{q\bar{q}}}{n_f} z[z^2 + (1-z)^2], & i = q\bar{q}, g \rightarrow q\bar{q}, \\ a_q \frac{z}{2} \frac{1+z^2}{1-z}, & i = gq, q \rightarrow gq, \end{cases} \quad (27)$$

$$= \int \frac{dz}{z} (1-D) \times \Gamma_i(z). \quad (28)$$

We use the relative transverse momentum  $k_t$  as the scale of the strong coupling, and expand the coupling at a minimal  $k_t$  with the one-loop beta function,

$$\alpha_s(k_t^2) = \bar{\alpha}_s(1-D), \quad \bar{\alpha}_s = \alpha_s(2p_0^2 t_0^2), \quad a = 2\bar{\alpha}_s b_0, \quad (29)$$

$$D = \begin{cases} (a\kappa' + a\lambda')/(1 + a\kappa' + a\lambda'), & \text{for final states,} \\ (a\bar{\kappa}' + a\lambda')/(1 + a\bar{\kappa}' + a\lambda'), & \text{for initial states,} \end{cases} \quad (30)$$

where we employed the following expressions for the transverse momentum:

$$k_t^2 = \begin{cases} 2z^2 p^2 t'^2 & \text{for final states,} \\ 2(x' - x)^2 p_{\text{beam}}^2 t'^2 & \text{for initial states.} \end{cases} \quad (31)$$

The coefficients are  $a_{q,g} = 2C_{F,A}\bar{\alpha}_s/\pi$  and  $a_{q\bar{q}} = n_f T_R \bar{\alpha}_s/\pi$  for final states, and we remove the factor 2 in  $a_g$  for initial states because the soft singularity for  $z \rightarrow 1$  in the gluon splitting function  $P_g(z)$  is suppressed by the gluon PDF  $f_g(x')$ . The number of active flavors is given by  $n_f$  and it is set to 5 in the numerical calculations below. The nontilde splitting kernel  $\Gamma_i(z)$  is given by removing the running effect of  $\alpha_s$  from  $\tilde{\Gamma}_i(z, t')$ .

## 2. Sudakov form factors

The Sudakov form factors for final states are defined as,

$$\tilde{\Delta}_i(\kappa, \lambda) = \exp\left(-\int_0^\kappa d\kappa' \int_0^\lambda d\lambda' \tilde{\Gamma}_i(z, t')\right), \quad i \in \{q, g\} \quad (32)$$

$$\tilde{\Delta}_{q\bar{q}}(\kappa, \lambda) = \exp\left(-\int_0^\kappa d\kappa' \int_0^\lambda d\lambda' n_f \tilde{\Gamma}_{q\bar{q}}(z, t')\right), \quad (33)$$

$$\tilde{\Delta}_G(\kappa, \lambda) = \tilde{\Delta}_g(\kappa, \lambda) \tilde{\Delta}_{q\bar{q}}(\kappa, \lambda). \quad (34)$$

The Sudakov form factors which are evaluated by neglecting the running effect of  $\alpha_s$  are given as,

$$\Delta_i(\kappa, \lambda) = \exp(-a_i \kappa \lambda), \quad i \in \{q, g\}, \quad (35)$$

$$\Delta_{q\bar{q}}(\kappa, \lambda) = \exp(-a_{q\bar{q}} c_{q\bar{q}} \kappa \lambda), \quad (36)$$

$$\Delta_G(\kappa, \lambda) = \Delta_g(\kappa, \lambda) \Delta_{q\bar{q}}(\kappa, \lambda), \quad (37)$$

where  $c_{q\bar{q}} = \frac{2}{3}(1 - e^{-3\kappa}) + e^{-2\kappa} - e^{-\kappa} \sim \frac{2}{3}$ . We can see the structure of leading (or double) logarithms in  $\Delta_q$  and  $\Delta_g$ , and single logarithms in  $\Delta_{q\bar{q}}$ .

For initial states, the Sudakov factors are defined as,

$$\tilde{\Pi}_i(\bar{\kappa}, \lambda) = \tilde{\Pi}_{i,1}(\bar{\kappa}, \lambda) \tilde{\Pi}_{i,2}(\bar{\kappa}, \lambda), \quad i \in \{q, g\}, \quad (38)$$

$$\tilde{\Pi}_{i,1}(\bar{\kappa}, \lambda) = \exp\left(-\int_0^{\bar{\kappa}} d\bar{\kappa}' \int_0^\lambda d\lambda' \frac{f_i(x')}{f_i(x)} \tilde{\Gamma}_i(z, t')\right), \quad (39)$$

$$\tilde{\Pi}_{q,2}(\bar{\kappa}, \lambda) = \exp\left(-\int_0^{\bar{\kappa}} d\bar{\kappa}' \int_0^\lambda d\lambda' \frac{f_g(x')}{f_q(x)} \tilde{\Gamma}_{q\bar{q}}(z, t')\right), \quad (40)$$

$$\tilde{\Pi}_{g,2}(\bar{\kappa}, \lambda) = \exp\left(-\int_0^{\bar{\kappa}} d\bar{\kappa}' \int_0^\lambda d\lambda' \sum_q \frac{f_q(x')}{f_g(x)} \tilde{\Gamma}_{gq}(z, t')\right). \quad (41)$$

Neglecting the running of  $\alpha_s$ , we get

$$\Pi_i(\bar{\kappa}, \lambda) = \Pi_{i,1}(\bar{\kappa}, \lambda) \Pi_{i,2}(\bar{\kappa}, \lambda), \quad i \in \{q, g\}, \quad (42)$$

$$\Pi_{i,1}(\bar{\kappa}, \lambda) = \exp(-a_i \bar{\kappa}_f^{(1)} \lambda), \quad (43)$$

$$\Pi_{q,2}(\bar{\kappa}, \lambda) = \exp\left(-c_{g/q}^{(1)} \frac{a_{q\bar{q}}}{n_f} \lambda\right), \quad (44)$$

$$\Pi_{g,2}(\bar{\kappa}, \lambda) = \exp(-c_{Q/g}^{(1)} a_q \lambda). \quad (45)$$

We define a functionalized  $\bar{\kappa}$  with a function  $f$  as,

$$\bar{\kappa}_f^{(n)} = n \int_0^{\bar{\kappa}} d\bar{\kappa}' \bar{\kappa}'^{n-1} f(\bar{\kappa}'), \quad (46)$$

where  $f_{i/i}(\bar{\kappa}) = f_i(x')/f_i(x)$ . For an identity function  $I$ , we can find a simple relation,  $\bar{\kappa}_I^{(n)} = \bar{\kappa}^n$ . The two coefficients are given as,

$$c_{g/q}^{(n)} = \frac{n_f}{a_{q\bar{q}}} \int_0^{\bar{\kappa}} d\bar{\kappa}' \frac{f_g(x')}{f_q(x)} \Gamma_{q\bar{q}}(z) \times \bar{\kappa}'^{n-1}, \quad (47)$$

$$c_{Q/g}^{(n)} = \frac{1}{a_q} \int_0^{\bar{\kappa}} d\bar{\kappa}' \sum_q \frac{f_q(x')}{f_g(x)} \Gamma_{gq}(z) \times \bar{\kappa}'^{n-1}, \quad (48)$$

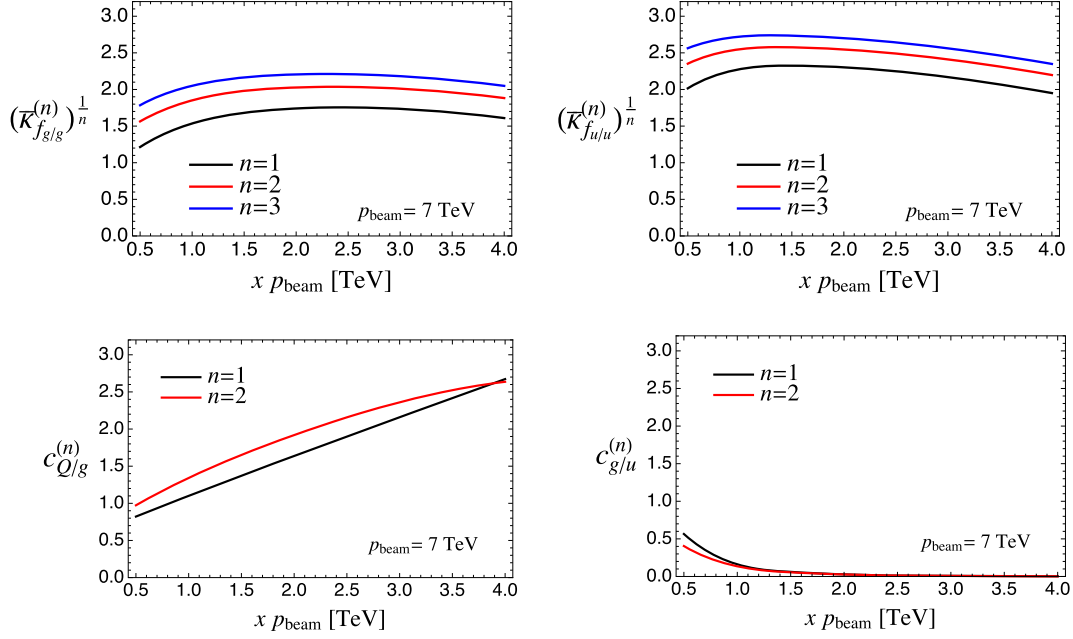


FIG. 2. The  $x$  dependence of the coefficients in Eqs. (46)–(48). The vertical axis shows  $x p_{\text{beam}}$  which is the energy of an initial state. CTEQ6L1 is used in the calculations.

where  $\sum_q$  runs over all active quarks and antiquarks. Figure 2 shows the  $x$  dependence of the coefficients in Eqs. (46)–(48). The vertical axis shows  $x p_{\text{beam}}$  which is the energy of an initial state. Valence quarks become dominant at large  $x$ ; therefore,  $c_{Q/g}^{(n)}$  becomes bigger and  $c_{g/u}^{(n)}$  becomes smaller as  $x p_{\text{beam}}$  increases. We adopt the CTEQ6L1 PDF set [106] in the calculations with the help of a PDF parser package, MANEPARSE 2.0 [107].

### B. Primary structure of functionals

Since the largest contribution to the  $t$  integration in Eqs. (11) and (12) comes from the region  $t' \sim t$ , we use the approximations  $\Phi_q(\lambda') \sim \Phi_q(t_0) = uv$  and  $\Phi_g(\lambda') \sim \Phi_g(t_0) = u$  to see the primary structures of the functionals [108].<sup>4</sup> In these approximations, subsequent emissions from a low-scale parton are prohibited. We correct the absence of the effect of subsequent emissions in the next section. Applying the approximations in the evolution equation (15)–(16) and neglecting the running of  $\alpha_s$ , we get

$$\Phi_q^{(\text{LL}+q\bar{q})} = uv \Delta_q \exp \left( \int_0^\lambda d\lambda' \int_0^\kappa d\kappa' \Gamma_q(z) u \right) \quad (49)$$

<sup>4</sup>In the approximations, the functional ratio in Eq. (16) takes the form  $\Phi_q(\kappa, \lambda') / \Phi_g(\kappa, \lambda') \sim uv/u$ , which causes unphysical terms,  $u^n v^m (n < m)$ . We remove such terms by hand, which causes a unitarity violation, namely  $\Phi|_{u=1, v=1} \neq 1$ . However the violation is very tiny, so we keep using the approximations. Note that the unitarity violation rate is  $1 - \Phi|_{u=1, v=1} \simeq (0.1\text{--}0.4)\%$  for numerical results in this paper.

$$= uv \Delta_q^{1-u}, \quad (50)$$

$$\Phi_g^{(\text{LL}+q\bar{q})} = u \Delta_G \exp \left\{ \int_0^\lambda d\lambda' \int_0^\kappa d\kappa' [\Gamma_g(z) u + n_f \Gamma_{q\bar{q}}(z) uv^2] \right\} \quad (51)$$

$$= u \Delta_g^{1-u} \Delta_{q\bar{q}}^{1-uv^2}. \quad (52)$$

The terms  $\Delta_{q,g}^{1-u} (\propto \sum_n u^n \cdot \bar{\alpha}_s^n L^{2n} / n!)$  which come from the integration of the leading splitting kernels  $\Gamma_{q,g}$  are involved in the increment of the number of gluon jets with the double logarithmic coefficients  $\bar{\alpha}_s^n L^{2n}$ , where  $L$  is proportional to  $\kappa$  or  $\lambda$ . The term  $\Delta_{q\bar{q}}^{1-uv^2} (\propto \sum_n u^n v^{2n} \cdot \bar{\alpha}_s^n L^n / n!)$  contains  $v$ , so it is involved in the increment of the number of quark jets with the single logarithmic coefficients  $\bar{\alpha}_s^n L^n$ . Since the enhancement term of gluon jets has more logarithmic enhancement, the increment of gluon jets is larger than that of quark jets. For  $\Phi_q^{(\text{LL}+q\bar{q})}$ , the functional does not contain the  $\Delta_{q\bar{q}}$  term, so the number of quark jets does not increase in this approximation.

Regarding the evolution equation for initial states in Eqs. (24) and (25), we adopt the approximation  $\Psi_i(\lambda') \sim \Psi_i(t_0) = 1$  as with the case of final states. The primary structures of the functionals are represented by,

$$\Psi_i^{(\text{LL}+q\bar{q})} = \Pi_{i,1}^{1-u} \Pi_{i,2}^{1-uv}, \quad i \in \{q, g\}. \quad (53)$$

The structures of the Sudakov form factors are,

$$\Pi_{i,1}^{1-u} \propto \sum_n u^n \cdot (a_i \bar{\kappa}_{f_{ij}}^{(1)} \lambda)^n / n!, \quad (54)$$

$$\Pi_{q,2}^{1-u} \propto \sum_n u^n v^n \cdot \left( \frac{a_{q\bar{q}}}{n_f} c_{g/q}^{(1)} \lambda \right)^n / n!, \quad (55)$$

$$\Pi_{g,2}^{1-u} \propto \sum_n u^n v^n \cdot (a_q c_{Q/g}^{(1)} \lambda)^n / n!. \quad (56)$$

Since the leading Sudakov form factor  $\Pi_{i,1}^{1-u}$  does not contain  $v$ , it is not involved in the increment of the number of quark jets. On the other hand, the subleading Sudakov form factor  $\Pi_{i,2}^{1-u}$  increases the number of quark jets since it contains  $v$ . In Fig. 2, we notice that the coefficient  $\bar{\kappa}_{f_{ij}}^{(1)}$  is basically larger than  $c_{ij}^{(1)}$  since the former is given by the integral of the splitting kernels which have soft-singularity terms. Therefore the increment of gluon jets is basically larger than that of quark jets as with the case of final-state radiation. However, it is easier for quark jets to be radiated from gluon initial states at high energy since  $c_{Q/g}^{(1)}$  gets bigger as the hard scale increases.

### C. Corrections from subsequent emissions

We add a correction to the generating functionals evaluated in the last section. In the previous section, we used the approximation  $\Phi_g(\kappa', \lambda') = u$ . This means that emissions by soft gluons which carry the energy fraction  $z$  are neglected. Consequently, a quark in a final state does not make more than one quark jet due to the absence of  $g \rightarrow q\bar{q}$ . We modify the approximation in order to take into account the radiation by the soft gluons as  $\Phi_g(\kappa', \lambda') \simeq \Phi_g^{(\text{LL}+q\bar{q})}(\kappa', \lambda')$ . We also modify the approximation for the ratio of functionals. We find that the primary structure of the functionals has the exponential suppression  $\exp(-c\kappa)$ , and thus the precision of the approximation employed in the previous section, i.e.,  $\Phi_q(\kappa, \lambda')/\Phi_g(\kappa, \lambda') = 1$  gets worse as the energy scale  $\kappa$  increases. In order to take into account the energy scaling of the ratio, we employ a new approximation  $\Phi_q(\kappa, \lambda')/\Phi_g(\kappa, \lambda') \simeq \Phi_q^{(\text{LL}+q\bar{q})}(\kappa, \lambda')/\Phi_g^{(\text{LL}+q\bar{q})}(\kappa, \lambda')$ . Applying these approximations in the evolution equation (15)–(16) and neglecting the running of  $\alpha_s$ , we get

$$\Phi_q^{(\text{LL}+q\bar{q}+\text{sub})} = uv \Delta_q \exp \left[ \int_0^\lambda d\lambda' \int_0^\kappa dk' \Gamma_q(z, t') u \Delta_g^{1-u}(\kappa', \lambda') \Delta_{q\bar{q}}^{1-uv^2}(\kappa', \lambda') \right] \quad (57)$$

$$= \Phi_q^{(\text{LL}+q\bar{q})} \times \exp(S_q), \quad (58)$$

$$\begin{aligned} \Phi_g^{(\text{LL}+q\bar{q}+\text{sub})} &= u \Delta_G \exp \left\{ \int_0^\lambda d\lambda' \int_0^\kappa dk' \left[ \Gamma_g(z, t') u \Delta_g^{1-u}(\kappa', \lambda') \Delta_{q\bar{q}}^{1-uv^2}(\kappa', \lambda') \right. \right. \\ &\quad \left. \left. + n_f \Gamma_{q\bar{q}}(z, t') \frac{\Delta_q^{1-u}(\kappa, \lambda')}{\Delta_g^{1-u}(\kappa, \lambda') \Delta_{q\bar{q}}^{1-uv^2}(\kappa, \lambda')} uv^2 \right] \right\} \quad (59) \end{aligned}$$

$$= \Phi_g^{(\text{LL}+q\bar{q})} \times \exp(S_g) \exp(S'). \quad (60)$$

The exponential factor  $\exp(S_i)$  stems from the modification for the soft-gluon generating functional. In other words, the term arises from activating subsequent emissions by soft gluons. The term  $\exp(S')$  stems from the modification for the functional ratio. Their full formulas are shown in Appendix A 1. The leading terms for the exponents are as follows:

$$S_i \sim -u \ln \Delta_i \left[ -\frac{(1-u)a_g \kappa \lambda}{4} - \frac{(1-uv^2)a_{q\bar{q}} \lambda}{2} \right], \quad (61)$$

$$S' \sim -uv^2 \ln \Delta_{q\bar{q}} \frac{w}{2}, \quad (62)$$

$$w = (1-u)(a_g - a_q) \kappa \lambda + (1-uv^2) c_{q\bar{q}} a_{q\bar{q}} \lambda. \quad (63)$$

In Eq. (61), the double logarithmic term  $-u \ln \Delta_i$  comes from the soft-gluon emission  $i \rightarrow i+g$ . The double and

single logarithms in the square brackets come from the subsequent emissions by the soft gluon, and their fractional factors arise from the integrals of the ordering variables  $\kappa'$  and/or  $\lambda'$ . In Eq. (62),  $w/2$  comes from the correction to the functional ratio.

As with the case of final states, we adopt the approximations  $\Phi_i(\lambda') \sim \Phi_i^{(\text{LL}+q\bar{q})}(\lambda')$  and  $\Psi_i(t') \sim \Psi_i^{(\text{LL}+q\bar{q})}(t')$  for the evolution equations for initial states in Eqs. (24) and (25); then we get

$$\Psi_i^{(\text{LL}+q\bar{q}+\text{sub})} = \Psi_i^{(\text{LL}+q\bar{q})} \times \exp(S_i[f_{i/i}]) \exp(S'_i). \quad (64)$$

For an analytic function  $G = \sum_n c_n \kappa^n$ , we define a functionalized one,  $G[f]$ , as,

$$G[f] = \sum_n c_n \bar{\kappa}_f^{(n)}. \quad (65)$$

The full formula of  $S'_i$  is shown in Appendix A 1, and the leading terms are,

$$S'_q \sim -uv \ln \Pi_{q,2} \frac{-\bar{w}}{2}, \quad (66)$$

$$S'_g \sim -uv \ln \Pi_{g,2} \frac{\bar{w}}{2}, \quad (67)$$

$$\begin{aligned} \bar{w} = & (1-u)(a_g \bar{\kappa}_{g/g}^{(1)} - a_q \bar{\kappa}_{q/q}^{(1)})\lambda \\ & + (1-uv) \left( a_q c_{Q/g}^{(1)} - \frac{a_{q\bar{q}}}{n_f} c_{g/q}^{(1)} \right) \lambda. \end{aligned} \quad (68)$$

Regarding  $S'_q$  and  $S'_g$ ,  $-uv \ln \Pi_{i,2}$  comes from the subleading splitting kernels for  $g \rightarrow q\bar{q}$  and  $g \rightarrow gq$  in Eqs. (24) and (25). The factors  $\pm\bar{w}/2$  come from the corrections to the functional ratios. The sign of  $\bar{w}/2$  is opposite since the numerators and denominators for the functional ratios have opposite sign.

#### D. $\alpha_s$ running correction

Finally, we consider the running effect of  $\alpha_s$ . Taking into account the effect of subsequent emission discussed in the previous section and using the Sudakov factors and the splitting functions with a tilde, i.e.,  $\tilde{\Delta}_i$  and  $\tilde{\Gamma}_i$ , we get the following equations for the final states:

$$\Phi_q^{(\text{LL}+q\bar{q}+\text{sub}+\delta\alpha_s)} = uv \tilde{\Delta}_q \exp \left[ \int_0^\lambda d\lambda' \int_0^\kappa d\kappa' \tilde{\Gamma}_q(z, t') u \Delta_g^{1-u}(\kappa', \lambda') \Delta_{q\bar{q}}^{1-uv^2}(\kappa', \lambda') \right] \quad (69)$$

$$= \Phi_q^{(\text{LL}+q\bar{q}+\text{sub})} \times \exp(\tilde{S}_q) \exp(T_q), \quad (70)$$

$$\begin{aligned} \Phi_g^{(\text{LL}+q\bar{q}+\text{sub}+\delta\alpha_s)} = & u \tilde{\Delta}_G \exp \left\{ \int_0^\lambda d\lambda' \int_0^\kappa d\kappa' \left[ \tilde{\Gamma}_g(z, t') u \Delta_g^{1-u}(\kappa', \lambda') \Delta_{q\bar{q}}^{1-uv^2}(\kappa', \lambda') \right. \right. \\ & \left. \left. + n_f \tilde{\Gamma}_{q\bar{q}}(z, t') \frac{\Delta_q^{1-u}(\kappa, \lambda')}{\Delta_g^{1-u}(\kappa, \lambda') \Delta_{q\bar{q}}^{1-uv^2}(\kappa, \lambda')} uv^2 \right] \right\} \end{aligned} \quad (71)$$

$$= \Phi_g^{(\text{LL}+q\bar{q}+\text{sub})} \times \exp(\tilde{S}_g) \exp(T_g) \exp(\tilde{S}') \exp(T'). \quad (72)$$

The full formulas for the exponential factors are shown in Appendix A 1 and their leading terms are,

$$\tilde{S}_i \sim u \ln \Delta_i \left[ -\frac{1}{6} (1-u) a_g \kappa \lambda \cdot a(\kappa + \lambda) - (1-uv^2) c_{q\bar{q}} a_{q\bar{q}} \lambda \cdot a \left( \frac{\kappa}{4} + \frac{\lambda}{3} \right) \right], \quad (73)$$

$$\tilde{S}' \sim uv^2 \ln \Delta_{q\bar{q}} \cdot w \cdot a \left( \frac{2\kappa}{3} + \frac{\lambda}{3} \right), \quad (74)$$

$$T_i = (1-u) \ln(\tilde{\Delta}_i/\Delta_i) \sim -(1-u) \ln \Delta_i \cdot a \frac{\kappa + \lambda}{2}, \quad (75)$$

$$T' = (1-uv^2) \ln(\tilde{\Delta}_{q\bar{q}}/\Delta_{q\bar{q}}) \sim -(1-uv^2) \ln \Delta_{q\bar{q}} \cdot a \left( \kappa + \frac{\lambda}{2} \right). \quad (76)$$

The two exponents  $\tilde{S}_i$  and  $\tilde{S}'$  are the  $\alpha_s$  corrections for  $S_i$  and  $S'$ , and  $e^{T_i}$  and  $e^{T'}$  are the corrections for  $\Delta_i$  and  $\Delta_{q\bar{q}}$ .

The  $\alpha_s$  corrections for the generating functionals of initial states are

$$\begin{aligned} \Psi_i^{(\text{LL}+q\bar{q}+\text{sub}+\delta\alpha_s)} = & \Psi_i^{(\text{LL}+q\bar{q}+\text{sub})} \times \exp(\tilde{S}_i[f_{i/i}]) \\ & \times \exp(T_i[f_{i/i}]) \exp(\tilde{S}'_i) \exp(T'_i), \end{aligned} \quad (77)$$

where  $\tilde{S}_i[f_{i/i}]$  and  $T_i[f_{i/i}]$  are the functionalized  $\tilde{S}_i$  and  $T_i$  [see the definition of “functionalized” in Eq. (65)]. The full

formulas for the exponents  $\tilde{S}'_i$  and  $T'_i$  are shown in Appendix A 1, and their leading terms are as follows:

$$\tilde{S}'_q \sim uv \ln \Pi_{q,2} \cdot w \cdot a \left( c_{g/q}^{(1)} + \frac{\lambda}{3} \right), \quad (78)$$

$$\tilde{S}'_g \sim uv \ln \Pi_{g,2} \cdot w \cdot a \left( c_{Q/g}^{(1)} + \frac{\lambda}{3} \right), \quad (79)$$

$$T'_q \sim -(1-uv) \ln \Pi_{q,2} \cdot ac_{g/q}^{(2)}/c_{g/q}^{(1)}, \quad (80)$$

$$T'_g \sim -(1-uv) \ln \Pi_{g,2} \cdot ac_{Q/g}^{(2)}/c_{Q/g}^{(1)}. \quad (81)$$



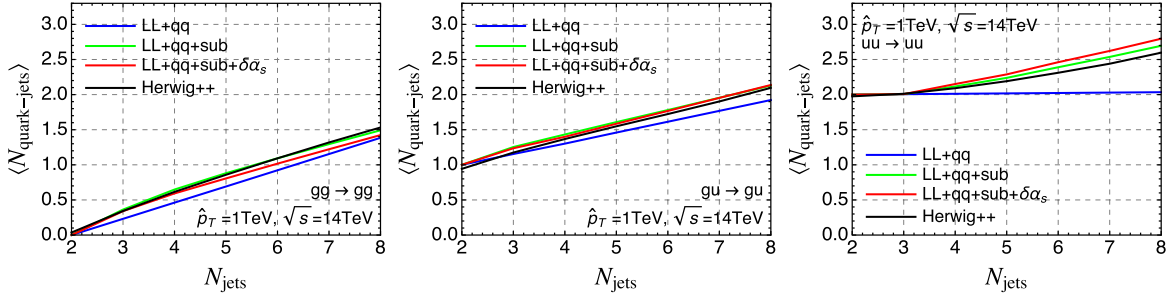


FIG. 3. Expected values of the number of quark jets for events in which  $N_{\text{jets}}$  jets are contained. The results for  $gg \rightarrow gg$  (left),  $gu \rightarrow gu$  (center) and  $uu \rightarrow uu$  (right) are shown. In the calculation,  $p_0 = 50 \text{ GeV}$  and  $R = 0.4$  are used.

### III. NUMERICAL RESULTS

#### A. Number of quark jets

We evaluate the quark jet rate for a given Born configuration,  $i_1 i_2 \rightarrow f_1 f_2$ . The generating functional for the configuration is given by,

$$\Phi_{i_1 i_2 \rightarrow f_1 f_2} = \Psi_{i_1}(x_1, t_{i_1}) \Psi_{i_2}(x_2, t_{i_2}) \Phi_{f_1}(p_{f_1}, t_{f_1}) \times \Phi_{f_2}(p_{f_2}, t_{f_2}). \quad (82)$$

We assume that the two final states scatter in the central region, which tends to occur at high energy, and set  $\hat{p}_T = p_{f_1} = p_{f_2} = x_1 p_{\text{beam}} = x_2 p_{\text{beam}}$ , where the proton beam energy is set to  $p_{\text{beam}} = 7 \text{ TeV}$ . The starting scale is set to the maximal one allowed kinematically, namely  $t_{\text{max}} = \sqrt{2}$ .

We calculate the number of quark jets for events in which  $N_{\text{jets}}$  jets are contained. The expected value for the number is given by,

$$\langle N_{\text{quark-jets}} \rangle = \frac{\sum_{m=0}^n m R_{n,m}^{(i_1 i_2 \rightarrow f_1 f_2)}}{R_n^{(i_1 i_2 \rightarrow f_1 f_2)}}, \quad n = N_{\text{jets}}, \quad (83)$$

where the jet rates and quark jet rates for  $i_1 i_2 \rightarrow f_1 f_2$  are given as,

$$R_n^{(i_1 i_2 \rightarrow f_1 f_2)} = \frac{1}{n!} \frac{\partial^n}{\partial u^n} \Phi_{i_1 i_2 \rightarrow f_1 f_2} \Big|_{u=0, v=1}, \quad (84)$$

$$R_{n,m}^{(i_1 i_2 \rightarrow f_1 f_2)} = \frac{1}{n! m!} \frac{\partial^n}{\partial u^n} \frac{\partial^m}{\partial v^m} \Phi_{i_1 i_2 \rightarrow f_1 f_2} \Big|_{u=v=0}. \quad (85)$$

In Fig. 3, the results for  $gg \rightarrow gg$  (left),  $gu \rightarrow gu$  (center) and  $uu \rightarrow uu$  (right) are shown. In the calculation,  $p_0 = 50 \text{ GeV}$  and  $R = 0.4$  are used. We set the parton transverse momentum as  $\hat{p}_T = 1 \text{ TeV}$ . The blue, green and red curves are analytical results using the functionals labeled by  $(\text{LL} + q\bar{q})$ ,  $(\text{LL} + q\bar{q} + \text{sub})$  and  $(\text{LL} + q\bar{q} + \text{sub} + \delta\alpha_s)$ . The black curves show

Monte Carlo predictions given by HERWIG++ [109].<sup>5</sup> Hadronization is turned off and the generated partons are clustered by the anti- $k_T$  algorithm [110].<sup>6</sup> We define a jet flavor from jet constituents. In the definition, it is necessary to consider the IR-unsafety of the jet flavor caused by soft gluon decaying into  $q\bar{q}$ , as noted in Ref. [96]. We take into account the decay effect and the details of estimating the number of quark jets in Monte Carlo samples are summarized in Appendix A 3. We can see that the analytic results including the terms of the subsequent emissions agree with the Monte Carlo results.

The primary structure of functionals has the form

$$\Phi \propto u^2 v^{m_0} e^{Au\lambda} e^{Buv\lambda} e^{Cuv^2\lambda}, \quad (86)$$

where  $A$ ,  $B$  and  $C$  are constants for a Born configuration. The expected value for the functional is estimated as,

$$\langle N_{\text{quark-jets}} \rangle \simeq \frac{B + 2C}{A + B + C} (N_{\text{jets}} - 2) + m_0, \quad (87)$$

where  $m_0$  is the number of quarks in final states for a targeted Born configuration. As the coefficients  $B$  and  $C$  related to  $v$  increase, the number of quark jets increases. The three coefficients and the initial number of quarks for the three configurations  $gg \rightarrow gg$ ,  $gu \rightarrow gu$ , and  $uu \rightarrow uu$  are shown in Table I. When we neglect subsequent emissions, the increase in the number of quark jets for  $uu \rightarrow uu$  is tiny because it is caused by only  $B$  and the coefficient is much smaller than the other coefficients as shown in Fig. 2. The main cause of the increase in the

<sup>5</sup>Our calculation employs a traditional angular-ordered parton shower formalism with  $1 \rightarrow 2$  splitting kernels. We compare our calculation with HERWIG++ by implementing the same formalism and splitting kernels to check the consistency of our analytical results.

<sup>6</sup>As mentioned in Sec. II, we will be able to measure the  $N_{\text{jets}}$  dependence on  $\langle N_{\text{quark-jets}} \rangle$  using the hadron-level definition of quark jets as in Ref. [99].

TABLE I. Coefficients in Eq. (86) for the three configurations  $gg \rightarrow gg$ ,  $gu \rightarrow gu$ , and  $uu \rightarrow uu$ .

$i_1 i_2 \rightarrow f_1 f_2$	$m_0$	$A$	$B$	$C$
$gg \rightarrow gg$	0	$2a_g(\kappa + \bar{\kappa}_{f_{g/g}}^{(1)})$	$2c_{Q/g}^{(1)} a_q$	$2c_{q\bar{q}} a_{q\bar{q}}$
$gu \rightarrow gu$	1	$a_g(\kappa + \bar{\kappa}_{f_{g/g}}^{(1)}) + a_q(\kappa + \bar{\kappa}_{f_{q/q}}^{(1)})$	$c_{Q/g}^{(1)} a_q + c_{g/u}^{(1)} \frac{a_{q\bar{q}}}{n_f}$	$c_{q\bar{q}} a_{q\bar{q}}$
$uu \rightarrow uu$	2	$2a_q(\kappa + \bar{\kappa}_{f_{u/u}}^{(1)})$	$2c_{g/u}^{(1)} \frac{a_{q\bar{q}}}{n_f}$	0

number of quark jets for  $uu \rightarrow uu$  stems from  $\exp(S_q)$  which is related to subsequent emissions, and the lowest order at which  $v$  appears is  $\mathcal{O}(u^4 v^4)$ ; therefore, the number of quark jets begins to increase clearly from  $N_{\text{jets}} = 4$ .

You can also see auxiliary plots in Appendix A 4 where only initial- or final-state radiation are taken into account.

### B. Expected improvement by the quark-gluon discrimination

In this subsection, we connect the knowledge obtained so far with beyond-the-standard-model (BSM) searches at the LHC. In Sec. III A, we can see that the coefficient  $A$  in Table I which is related to the increment of gluon jets is basically bigger than  $B$  and  $C$ . This means that much of the QCD multijet background is composed of gluon jets and a few quark jets stemming from the valence quarks. So, we expect to get a large improvement in the separation between QCD multijet backgrounds and signals containing many quark jets by using the quark-gluon tagging.

We estimate the improvement of the signal-to-background ratio ( $S/B$ ) for such signals using the analytical results in Sec. II and Monte Carlo results. We introduce an improvement factor  $\epsilon_S/\epsilon_B$ , where  $\epsilon_S$  and  $\epsilon_B$  are the signal and background efficiencies after applying the quark-gluon discrimination in multijet final states. Therefore the ratio factor represents how many times  $S/B$  increases after the application. We assume that all jets in signals are quark jets, i.e.,  $N_{q,S} = N_{\text{jets}}$ , where  $N_{q,S}$  is the number of quark jets in signals. Such signals will also be considered in the next section. In this assumption, the signal efficiency is naively

estimated as  $\epsilon_S \sim \epsilon_q^{N_{q,S}}$ . If a signal contains one more quark jet than backgrounds,  $S/B$  increases by  $\epsilon_q/\epsilon_g$  times by applying the quark-gluon tagging, i.e.,  $\epsilon_S/\epsilon_B = \epsilon_q/\epsilon_g$ , where  $\epsilon_q$  and  $\epsilon_g$  are the quark and gluon jet efficiencies. In the case that the expected number of quark jets in background is  $N_{q,B}$ , we can expect that the improvement factor maximally increases up to  $\epsilon_S/\epsilon_B \sim (\epsilon_q/\epsilon_g)^{N_{q,S}-N_{q,B}}$ . Below, we estimate  $N_{q,B}$  using the generating functionals from Sec. II D. Although the efficiency ratio  $\epsilon_q/\epsilon_g$  is calculable for IRC-safe and Sudakov safe variables, we calculate the ratio using a Monte Carlo generator.

For the estimation of  $N_{q,B}$ , we first define the generating functional for proton collisions as,

$$\Phi_{pp \rightarrow \text{jets}} \propto \sum_{i_1, i_2} f_{i_1}(x_{i_1}, \mu_F) f_{i_2}(x_{i_2}, \mu_F) \Phi_{i_1 i_2 \rightarrow f_1 f_2}. \quad (88)$$

The starting scale,  $\hat{p}_T$  and  $p_{\text{beam}}$  in  $\Phi_{i_1 i_2 \rightarrow f_1 f_2}$  are set as in Sec. III A. The hard scale of collisions is set to the invariant mass of the initial partons  $\sqrt{\hat{s}}$ . We are interested in the case where the hard scale is a given new physics scale  $\Lambda_{\text{new}}$ , and therefore we set  $\sqrt{\hat{s}} = \Lambda_{\text{new}}$ . The transverse momentum of the final states is set to half of the invariant mass  $\hat{p}_T = \sqrt{\hat{s}}/2 = \Lambda_{\text{new}}/2$ . The function  $f_i$  is the proton PDF for an initial parton  $i$ , and the factorization scale is set to  $\mu_F = \hat{p}_T$ . The four-flavor scheme is used in the calculation. With these settings, we calculate the expected number of quark jets  $N_{q,B}$  as in Sec. III A,

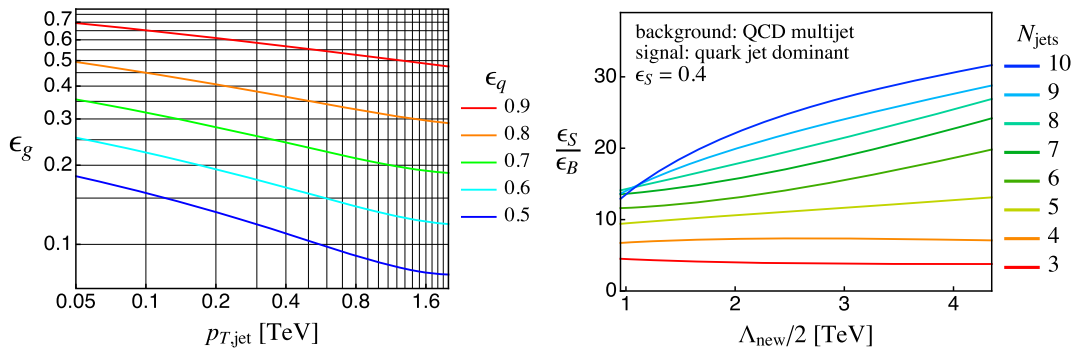


FIG. 4. Jet  $p_T$  dependence on gluon efficiencies for each quark efficiency (left panel). Dependence of the physics scale  $\Lambda_{\text{new}}$  on the improvement factor  $\epsilon_S/\epsilon_B$  for each  $N_{\text{jets}}$  category (right panel).

$$N_{q,B} = \frac{\sum_{m=0}^n m R_{n,m}^{(pp \rightarrow \text{jets})}}{R_n^{(pp \rightarrow \text{jets})}}, \quad n = N_{\text{jets}}, \quad (89)$$

$$R_n^{(pp \rightarrow \text{jets})} = \frac{1}{n!} \frac{\partial^n}{\partial u^n} \Phi_{pp \rightarrow \text{jets}} \Big|_{u=0, v=1}, \quad (90)$$

$$R_{n,m}^{(pp \rightarrow \text{jets})} = \frac{1}{n! m!} \frac{\partial^n}{\partial u^n} \frac{\partial^m}{\partial v^m} \Phi_{pp \rightarrow \text{jets}} \Big|_{u=v=0}. \quad (91)$$

Next, we show Monte Carlo results for the quark and gluon jet efficiencies. In Fig. 4, the left panel shows the jet- $p_T$  dependence on the gluon efficiency for several quark efficiencies. We use VINCIA [111–116] in the calculation.<sup>7</sup> We use an output evaluated by the boosted decision trees (BDT) algorithm as a discrimination variable. The output is trained using four variables, namely the number of charged tracks, energy correlation functions [16] with  $\beta = 0.2$  and 1.0, and  $p_T$ -normalized jet mass ( $m_{\text{jet}}/p_T$ ). Only charged track information is used for the calculation of the BDT inputs. When we calculate  $\epsilon_q$  and  $\epsilon_g$  in the improvement factor, we set the jet  $p_T$  in the estimation of the quark and gluon jet efficiencies to  $\Lambda_{\text{new}}/N_{\text{jets}}$ .

Finally, we estimate the improvement factor  $\epsilon_S/\epsilon_B = (\epsilon_q/\epsilon_g)^{N_{q,S}-N_{q,B}}$  with the above settings. The right plot in Fig. 4 shows the dependence of a new physics scale  $\Lambda_{\text{new}}$  on the factor for each  $N_{\text{jets}}$  category. The signal efficiency is fixed at 0.4. We can see that the improvement factor increases as the number of jets increases since the difference between the number of quark jets in the signal and background, namely  $N_{q,S} - N_{q,B}$ , gets larger. Also, the factor improves as the new physics scale gets larger because the discrimination power for the quark-gluon separation increases as the jet  $p_T$  increases. The effect is clear in large  $N_{\text{jets}}$  categories. The probability that valence quarks flow into final states becomes larger as  $\Lambda_{\text{new}}$  increases. This makes the difference between the number of quark jets in the signal and background small, and therefore the factor decreases. This effect is noticeable in small  $N_{\text{jets}}$  categories.

#### IV. BSM SEARCHES IN MULTIJET FINAL STATES

In this section, we calculate the improvement factor estimated in Sec. III B in a realistic data-driven way using a Monte Carlo generator. The data-driven method is often used for the analysis of multijet final states. A typical analysis is the search for micro black holes [73–79]. In the analysis, the phase space is divided by a variable related to the hard scale, e.g., the scalar sum of jet transverse

momenta  $H_T$ , the scalar sum of the masses of large- $R$  jets,  $M_J^\Sigma$  [119–121] etc. We fit the distribution of the variable in phase space at a low-energy scale referred to as the control region (CR), and estimate the number of backgrounds in phase space at a high-energy scale referred to as the signal region (SR) using the fit function. We interpret any excess from the estimated background as a sign of new physics.

One of the problems for such analyses is that we simplify the high-jet-multiplicity events too much. In the analysis explained above, only one or two inclusive variables are mainly used.<sup>8</sup> We can also utilize robust jet substructure variables for quark-gluon tagging in the analysis. It is difficult to predict multidimensional distributions for the jet substructure variables in multijet final states precisely; therefore, the data-driven approach is preferred to incorporate the jet flavor information into the analysis. We introduce a variable containing the information for the data-driven analysis,<sup>9</sup>

$$d = \sqrt{\frac{1}{n} \sum_{i=1}^n Q_i^2}, \quad \text{for } N_{\text{jets}} \geq n, \quad (92)$$

where  $Q_i (> 0)$  shows a kind of *quark-jet-ness* for the  $i$ th jet. If the jet substructure for the  $i$ th jet looks like a quark jet rather than a gluon jet,  $Q_i$  takes a larger value. In this paper, we use the BDT output used in Sec. III B as  $Q_i$ , which is trained such that the quark and gluon jets are assigned to 1 and 0. The variable  $d$  takes a large value for events which contain many quark jets.

We consider the following toy-signal topologies:

$$(gg \text{ or } u\bar{u}) \rightarrow XX, \quad X \rightarrow n_X \text{ quarks}. \quad (93)$$

The pair production of a hypothetical heavy resonance  $X$  has initial states  $gg$  or  $u\bar{u}$  in proton collisions and  $X$  decays into  $n_X$  quarks. For example, the pair production of gluinos and squarks in SUSY with  $R$ -parity violation have the same decay topology. We generate hard processes using MADGRAPH5 [122] with the CTEQ6L PDF set and have  $X$  decay in phase space flatly. When  $n_X$  is odd or even,  $X$  is assigned to the color octet or triplet respectively, and the color indices of  $X$  are connected to those of quarks in the large- $N_c$  limit. We use VINCIA for the parton showering and the hadronization.

For the simulation of the QCD multijet background, we use VINCIA with the default setting.

<sup>7</sup>The one problem in quark-gluon discrimination is Monte Carlo uncertainties of the predictions, and it is known that experimental data on certain observables related to quark/gluon tagging lie in between the predictions of the two MC generators PYTHIA and HERWIG [54,56,57]. Although VINCIA's results are close to PYTHIA's, those lie in between the predictions of the two MC generators, and the uncertainties were studied in Refs. [5,117,118].

<sup>8</sup>The sum of fat jet masses  $M_J^\Sigma$  also contains some information on exclusive variables like the jet  $p_T$  and the distance between subjets.

<sup>9</sup>We found that the performance of the discrimination between the signal and background discussed below can increase slightly with a more complicated definition of  $d$ . The optimization of the variable  $d$  is beyond the scope of this paper.

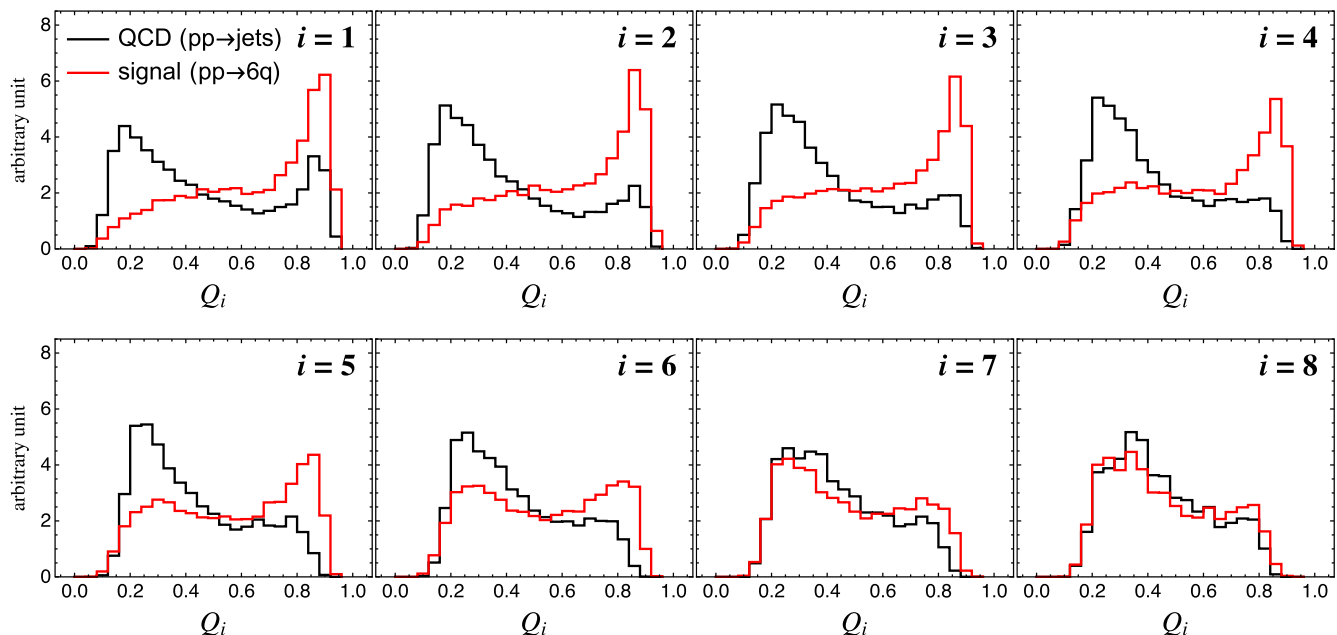


FIG. 5. The black and red curves show the distribution of  $Q_i$  for the background and signal, where  $H_T > 2$  TeV,  $N_{\text{jets}} \geq 8$  are imposed. We use the BDT outputs used in Sec. III B as  $Q_i$ , which is trained such that the quark and gluon jet samples are assigned to 1 and 0.

The generated signal and background are clustered with the anti- $k_r$  algorithm and the jet radius is set to  $R = 0.4$ . As a selection cut, the minimum transverse momentum ( $p_T > 50$  GeV) and rapidity cut ( $|\eta| < 2.8$ ) are imposed to all jets. The invariant mass of the collision system is set to  $\sqrt{s} = 14$  TeV.

In Fig. 5, the black and red curves show the distribution of  $Q_i$  for the background and signal, where  $H_T > 2$  TeV,  $N_{\text{jets}} \geq 8$  are imposed. For the background, the  $Q_i$ 's tend to be distributed in the region close to 0 since the gluon jets are dominant in QCD multijet final states; however,  $Q_1$  has a clear peak near 1 due to the effect of valence quark jets. For the signal, we set the mass of  $X$  ( $M_X$ ) to 2 TeV, and  $n_X = 3$ . The signal has six quarks in the hard process, so

the  $Q_i$ 's are distributed in the region close to 1 up to the sixth jet. The seventh and eighth jets would stem from QCD radiation, so the differences between the signal and background become small.

In Fig. 6, the distribution of  $d$  for the QCD multijets (left) and the signal (right) for each  $N_{\text{jets}}$  category are shown. The signal is set to  $M_X = 2$  TeV and  $n_X = 5$ . We can remove the background by imposing the cut  $d > d_{\text{cut}}$ , since the signal is distributed in the large- $d$  region.

In Fig. 7, the remaining rates of the number of events after imposing the  $d$  cut for each  $H_T$  bin are shown. The left and right panels show the results for the QCD multijets and the signal at  $N_{\text{jets}} \geq 6$ . The signal parameters are  $M_X = 2$  TeV and  $n_X = 5$ . The background decreases rapidly after

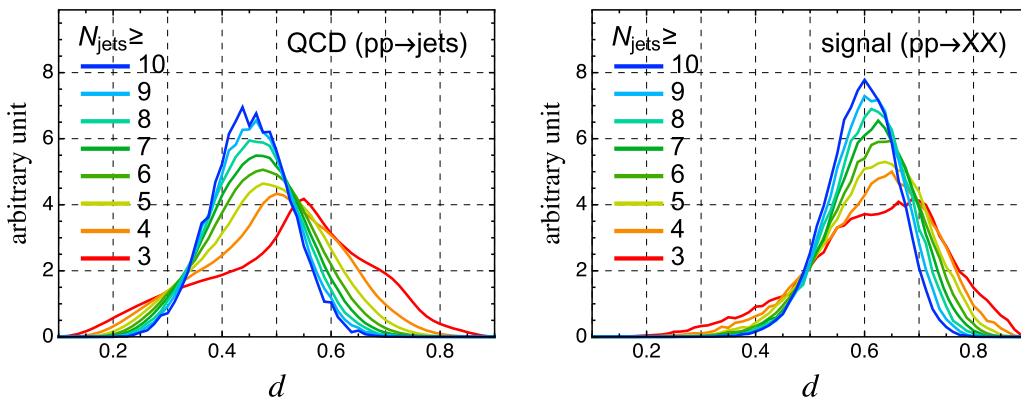


FIG. 6. The distribution of  $d$  for the QCD multijets (left) and the signal (right) for each  $N_{\text{jets}}$  category. The signal is set to  $M_X = 2$  TeV and  $n_X = 5$ .

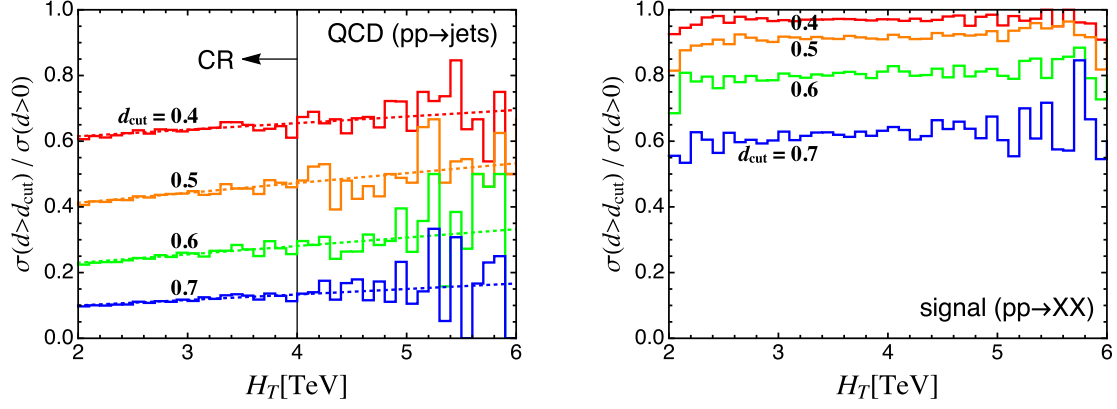


FIG. 7. The remaining rates of the number of events after imposing the  $d$  cut for each  $H_T$  bin. The left and right panels show the results for the QCD multijets and the signal at  $N_{\text{jets}} \geq 6$ . The signal parameters are  $M_X = 2$  TeV and  $n_X = 5$ . The dotted curves show an example of interpolation curves which are fitted by using data in a control region, i.e.,  $H_T < 4$  TeV.

imposing the larger  $d_{\text{cut}}$ . What we want to know is the number of backgrounds after imposing the  $d$  cut in the high-energy region. Due to the complexity of large-jet-multiplicity events, such a number should be estimated by the data-driven method. The dotted curves in Fig. 7 show an example of interpolation curves which are fitted by using data in a control region, i.e.,  $H_T < 4$  TeV in the figure. In the practical analysis, we can know the ratio of background in signal regions using such interpolation curves, and can obtain the upper bound on the cross-section imposed  $d$  cut in QCD multijet final states.

We estimate how much  $S/B$  improves by applying the quark-gluon discrimination in multijet final states. The ratio is given by,

$$\frac{S}{B} \simeq \frac{\sigma_S(\text{selection cuts \& } H_T\text{-cut})}{\sigma_B(\text{selection cuts \& } H_T\text{-cut})} \times \frac{\epsilon_S(d_{\text{cut}})}{\epsilon_B(d_{\text{cut}})}, \quad (94)$$

$$\epsilon_X(d_{\text{cut}}) = \frac{\sigma_X(\text{selection cuts \& } H_T\text{-cut \& } d > d_{\text{cut}})}{\sigma_X(\text{selection cuts \& } H_T\text{-cut \& } d > 0)}, \quad (95)$$

where  $\sigma_X$  is the cross section for the signal ( $X = S$ ) and the background ( $X = B$ ) after imposing the condition in the brackets. The impact coming from quark-gluon discrimination is included in the second ratio  $\epsilon_S(d_{\text{cut}})/\epsilon_B(d_{\text{cut}})$ .

We employ an  $H_T$  cut  $H_T > 1.8M_X$ , which almost makes the significance of signal maximum in the case that systematic uncertainties and  $d$  cut are neglected.

In Fig. 8, the  $M_X$  dependence on the efficiency ratio is shown, where the dependence on  $\epsilon_S/\sqrt{\epsilon_B}$  is also shown on the right axis. We can see how the ratio changes as we increase the lower bound of  $N_{\text{jets}}$  from 3 to 10, and  $n_X$  from 2 (leftmost) to 5 (rightmost). We choose  $d_{\text{cut}}$  which gives the signal efficiency  $\epsilon_S = 0.4$ . These are the results in the case that the initial state is  $gg$ . The ratio clearly keeps increasing until the lower bound of  $N_{\text{jets}}$  reaches  $2n_X$  since  $2n_X$  quarks are contained in the hard processes of the signal. A quark emitted from  $X$  could be softer than partons arising from initial- and/or final-state radiation. In that case, the quark from  $X$  could be the  $(2n_X + 1)$ th jet, so we can see some improvement in  $S/B$  even if the lower bound of  $N_{\text{jets}}$  is greater than  $2n_X$ . We can understand the behavior of the  $M_X$  dependence on the improvement factor from the result in Sec. III B. The ratio improves as  $M_X$  get larger because the discrimination power of the quark-gluon separation increases as the jet  $p_T$  increases. The effect is clear in the large  $N_{\text{jets}}$  categories. The probability that valence quarks are in final states becomes larger as the masses increase. This makes the difference between the number of quark jets in the signal and background small,

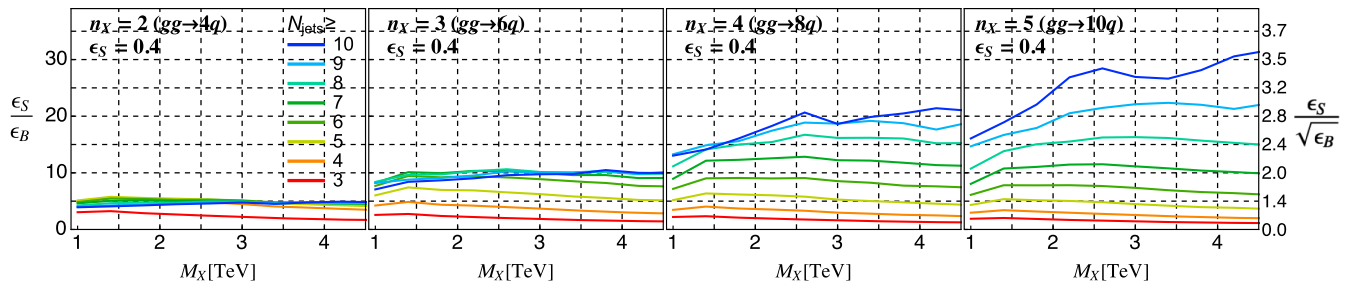
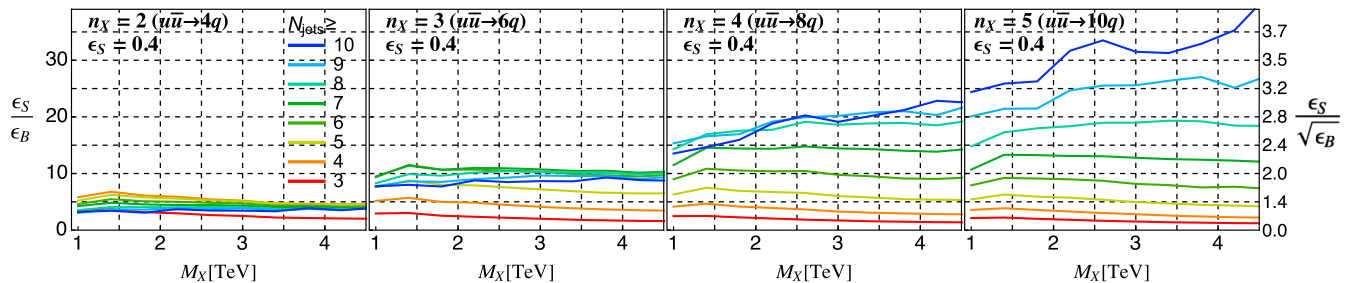


FIG. 8.  $M_X$  dependence on the efficiency ratio. We can see how the ratio changes as we increase the lower bound of  $N_{\text{jets}}$  from 3 to 10, and  $n_X$  from 2 (leftmost) to 5 (rightmost).



FIG. 9. Same as Fig. 8, with the initial states  $u\bar{u}$ .

and makes the ratio decrease. The effect looks important in the small  $N_{\text{jets}}$  categories. We also see a good agreement between the rightmost panel in Fig. 8 and the semianalytic result in Fig. 4. In both cases, the signals are quite quark-jet dominant. Note that the hard scale or the invariant mass for the pair production  $X$  is about  $2M_X$  therefore, the label  $\Lambda_{\text{new}}/2$  on the  $x$  axis in Fig. 4 almost corresponds to  $M_X$  on the  $x$  axis in Fig. 8.

Figure 9 is the same as Fig. 8, but for the initial states  $u\bar{u}$ . The Born configuration of the  $gg$  initial state tends to emit valence quark jets at high energy, but it also emits more gluon jets compared to the case of the  $u\bar{u}$  initial state since the color factor for  $g \rightarrow gg$  is  $9/4$  times larger than  $q \rightarrow qq$ . Consequently, initial-state radiation from  $gg$  reduce the quark jet fraction, and thus the improvement factors for  $u\bar{u}$  are slightly better than those for  $gg$ .

## V. CONCLUSIONS

The quark-gluon discrimination is maximally utilized for searches of new physics that predict quark and gluon jet fractions that are different from the QCD background. To determine the jet flavor structure in QCD multijet final states at hadron colliders, we have introduced the quark jet rate  $R_{n,m}$  which is the probability that a parton or a matrix element produces  $n$  jets in which  $m$  quark jets are contained. We have calculated generating functionals, which contain the quark jet rates as coefficients, for initial and final states by using the QCD resummation technique.

The exponential structures of the functionals were evaluated and obtained the quark jet rates  $R_{n,m}$  from the expansion coefficients. The increment of gluon jets mainly arises from leading-logarithmic terms in the coefficients, and that of quark jets comes from next-to-leading-logarithmic terms. More details of the logarithmic structure were also shown. In order to know the rate of increase of the number of quark jets, we have shown the expected value of the number of quark jets in  $N_{\text{jets}}$  categories for the matrix element configurations  $gg \rightarrow gg$ ,  $gu \rightarrow gu$  and  $uu \rightarrow uu$ . For example, when we set the jet radius, jet  $p_T$  cut, and the parton  $p_T$  cut to  $R = 0.4$ ,  $p_0 = 50$  TeV and  $\hat{p}_T = 1$  TeV, the number of quark jets increases by about 0.25, 0.18 and 0.12 for the three configurations while the number of jets

increases by 1. We have also checked the consistency between the analytical results and Monte Carlo predictions.

Since the QCD multijets are basically composed of a few valence quark jets and many gluon jets, we expect a big improvement in  $S/B$  for a signal which predicts many quark jets, by using the quark-gluon discrimination. We have estimated the improvement semianalytically using the above results and have shown that the improvement gets larger as the number of quark jets in signals increases. For example,  $S/B$  increases by about 20 times in the case that a new physics scale is  $\Lambda_{\text{new}} = 4$  TeV and the number of quark jets is ten.

We have introduced a variable  $d$  that takes a large value for events in which many quark jets are contained, and have suggested a data-driven analysis using the variable. Assuming the pair production of a hypothetical heavy resonance  $X$  which decays into  $n_X$  quarks as a signal, we have evaluated the large improvement in  $S/B$  for each mass of  $X$ ,  $n_X$  and initial state in a Monte Carlo analysis, and have shown the usability of the quark-gluon discrimination in multijet final states.

## ACKNOWLEDGMENTS

I thank Mihoko M. Nojiri, Hideki Okawa, Bryan R. Webber and Jesse Thaler for helpful discussions and useful comments. This work is supported by the Samsung Science and Technology Foundation under Project No. SSTF-BA1602-04.

## APPENDIX A: DETAILS ON QUARK JET RATES

### 1. Formulas

In Sec. II B, the generating functionals which contain the effects of emissions coming from only progenitor partons are evaluated, and those for final states are given as,

$$\Phi_q^{(\text{LL}+q\bar{q})} = uv\Delta_q^{1-u}, \quad (\text{A1})$$

$$\Phi_g^{(\text{LL}+q\bar{q})} = u\Delta_g^{1-u}\Delta_{q\bar{q}}^{1-uv^2}. \quad (\text{A2})$$

For initial states, we get

$$\Psi_i^{(\text{LL}+q\bar{q})} = \Pi_{i,1}^{1-u} \Pi_{i,2}^{1-uv}, \quad i \in \{q, g\}. \quad (\text{A3})$$

In Sec. II C, we also consider the effects of subsequent emissions. The functionals are factorized into the primary terms and exponential terms related to the effect as,

$$\Phi_q^{(\text{LL}+q\bar{q}+\text{sub})} = \Phi_q^{(\text{LL}+q\bar{q})} \times \exp(S_q), \quad (\text{A4})$$

$$\Phi_g^{(\text{LL}+q\bar{q}+\text{sub})} = \Phi_g^{(\text{LL}+q\bar{q})} \times \exp(S_g) \exp(S'), \quad (\text{A5})$$

where

$$S_i = -u \ln \Delta_i I_1(p\kappa\lambda, q\lambda), \quad (\text{A6})$$

$$S' = -uv^2 \ln \Delta_{q\bar{q}} I_2(w), \quad (\text{A7})$$

and

$$I_1(x, y) = \frac{\text{Ein}(x+y) - \text{Ein}(y)}{x} - 1, \quad (\text{A8})$$

$$I_2(z) = \frac{e^z - 1}{z} - 1, \quad (\text{A9})$$

$$p = (1-u)a_g, \quad q \simeq (2/3)(1-uv^2)a_{q\bar{q}}, \quad (\text{A10})$$

$$\text{Ein}(z) = \sum_{n=1}^{\infty} \frac{-(-z)^n}{nn!}. \quad (\text{A11})$$

For  $q$  in Eq. (A10), we used the approximation  $c_{q\bar{q}} \simeq 2/3$ . For initial states, the generating functionals containing effects of subsequent emissions are

$$\Psi_i^{(\text{LL}+q\bar{q}+\text{sub})} = \Psi_i^{(\text{LL}+q\bar{q})} \times \exp(S_i[f_{i/i}]) \exp(S'_i), \quad (\text{A12})$$

where

$$S'_q = -uv \ln \Pi_{q,2} I_2(-\bar{w}), \quad (\text{A13})$$

$$S'_g = -uv \ln \Pi_{g,2} I_2(+\bar{w}). \quad (\text{A14})$$

You can see the definition of  $\bar{w}$  in Eq. (66).

In Sec. II D, the running effects of  $\alpha_s$  are also considered. In this case, the functionals for final states are written as,

$$\Phi_q^{(\text{LL}+q\bar{q}+\text{sub}+\delta\alpha_s)} = \Phi_q^{(\text{LL}+q\bar{q}+\text{sub})} \times \exp(\tilde{S}_q) \exp(T_q), \quad (\text{A15})$$

$$\Phi_g^{(\text{LL}+q\bar{q}+\text{sub}+\delta\alpha_s)} = \Phi_g^{(\text{LL}+q\bar{q}+\text{sub})} \times \exp(\tilde{S}_g) \exp(T_g) \times \exp(\tilde{S}') \exp(T'), \quad (\text{A16})$$

where

$$\tilde{S}_i = u \ln \Delta_i \left( -\frac{\kappa + \lambda}{2} + I_3(s) - I_3(q) \right), \quad (\text{A17})$$

$$T_i = -(1-u) \ln \Delta_i I_4(a\kappa, a\lambda), \quad (\text{A18})$$

$$\tilde{S}' \simeq -uv^2 a_{q\bar{q}} \frac{a}{1+a\kappa+a\lambda} \left[ \left( \kappa - \frac{13}{12} \right) \left( \frac{e^w - 1}{w} - 1 \right) + \lambda \left( \frac{1 - e^w}{w^2} + \frac{e^w}{w} - \frac{1}{2} \right) \right], \quad (\text{A19})$$

$$T' \simeq (1-uv^2) \frac{2}{3} a_{q\bar{q}} \lambda \frac{a}{1+a\kappa+a\lambda} \left( \kappa + \frac{\lambda}{2} - \frac{13}{12} \right), \quad (\text{A20})$$

and

$$I_3(z) = -\frac{1}{p\kappa\lambda} \left[ \left( \frac{z}{p} + \lambda \right) I_2(-z\lambda) + \frac{q}{p} \text{Ein}(z\lambda) \right], \quad (\text{A21})$$

$$I_4(x, y) = 1 - \frac{(1+x+y) \ln(1+x+y) - (1+x) \ln(1+x) - (1+y) \ln(1+y)}{xy}. \quad (\text{A22})$$

In the calculations of  $\tilde{S}'$  and  $T'$  in Eqs. (A19) and (A20), we use the following approximation:

$$\int_0^\kappa d\kappa' \int_0^\lambda d\lambda' F(\kappa', \lambda') D = \int_0^\kappa d\kappa' \int_0^\lambda d\lambda' F(\kappa', \lambda') \frac{a(\kappa' + \lambda')}{1 + a(\kappa' + \lambda')}. \quad (\text{A23})$$

This approximation has good accuracy because the integrands for  $\tilde{S}'$  and  $T'$  are localized around  $\kappa' = \kappa$  and  $\lambda' = \lambda$ . For initial states, we get

$$\Psi_i^{(\text{LL}+q\bar{q}+\text{sub}+\delta\alpha_s)} = \Psi_i^{(\text{LL}+q\bar{q}+\text{sub})} \times \exp(\tilde{S}_i[f_{i/i}]) \exp(T_i[f_{i/i}]) \exp(\tilde{S}'_i) \exp(T'_i), \quad (\text{A24})$$

where

$$\tilde{S}'_q \simeq uv \ln \Pi_{q,2} a I_5(c_{g/q}^{(1)}, c_{g/q}^{(2)}), \quad (\text{A25})$$

$$\tilde{S}'_g \simeq uv \ln \Pi_{g,2} a I_5(c_{Q/g}^{(1)}, c_{Q/g}^{(2)}), \quad (\text{A26})$$

$$T'_q = (1-uv) \frac{a_{q\bar{q}}}{n_f} \delta c_{g/q}^{(1)} \lambda, \quad (\text{A27})$$

$$T'_g = (1-uv) a_q \delta c_{Q/g}^{(1)} \lambda, \quad (\text{A28})$$

and

$$I_5(c_1, c_2) = \left( \frac{1 - e^{\bar{w}} + \bar{w} e^{\bar{w}}}{\bar{w}^2} - \frac{1}{2} \right) \lambda + \frac{e^{\bar{w}} - 1 - \bar{w}}{\bar{w}} \cdot \frac{c_2}{c_1}, \quad (\text{A29})$$

$$\delta c_{g/q}^{(1)} = \frac{n_f}{a_{q\bar{q}}} \int_0^{\bar{\kappa}} d\bar{\kappa}' \Gamma_{g \rightarrow q\bar{q}}(\bar{\kappa}') \frac{f_g(x')}{f_q(x')} \left[ 1 - \frac{1}{a\lambda} \ln \left( 1 + \frac{a\lambda}{1 + a\bar{\kappa}'} \right) \right], \quad (\text{A30})$$

$$\delta c_{Q/g}^{(1)} = \frac{1}{a_q} \int_0^{\bar{\kappa}} d\bar{\kappa}' \sum_q \Gamma_{q \rightarrow gq}(\bar{\kappa}') \frac{f_q(x')}{f_g(x')} \left[ 1 - \frac{1}{a\lambda} \ln \left( 1 + \frac{a\lambda}{1 + a\bar{\kappa}'} \right) \right]. \quad (\text{A31})$$

In the calculations of  $\tilde{S}'_{q,g}$  in Eqs. (A25) and (A26), we expand  $D$  in logarithms and take into account only the leading term, namely  $D \simeq a(\kappa' + \lambda')$ .

## 2. Matrix element corrections

In Sec. III, the number of quark jets for each  $N_{\text{jets}}$  category are evaluated by applying parton showers to Born configurations. In the calculation, matrix element corrections are absent for more than two jets. In Fig. 10, we show the matrix element correction to the number of quark jets with Catani-Krauss-Kuhn-Webber matching using SHERPA [123,124]. The black curve is the result for  $pp \rightarrow jj + \text{parton showers}$ . One and two partons are matched in the Born configuration and the results are shown by the red and blue curves. We impose  $H_T > 2 \text{ TeV}$  and set  $\sqrt{s} = 14 \text{ TeV}$ .

For the black curve,  $\langle N_{\text{quark-jets}} \rangle$  is about 1.5 for  $N_{\text{jets}} = 2$ . This means that final states tend to become two valence quarks. In this case, the curve has an artificial kink at  $N_{\text{jets}} = 3$  as discussed in Sec. III. We can see that the kink disappears with matching. For the red and blue curves, the matrix element corrections are contained up to  $N_{\text{jets}} = 3$  and 4, and we find that the configuration containing gluons in final states increases. As we see in Sec. III, the rate of increase of the number of quark jets for the gluon final state is larger than that for the quark final state. Therefore, the rate slightly increases after the matching.

## 3. Estimating the number of quark jets in Monte Carlo samples

In Fig. 3 of Sec. III A, we compare the number of quark jets calculated analytically with that using Monte Carlo events at the parton level. We define a jet flavor for each jet in multijet final states and estimate the expected value of

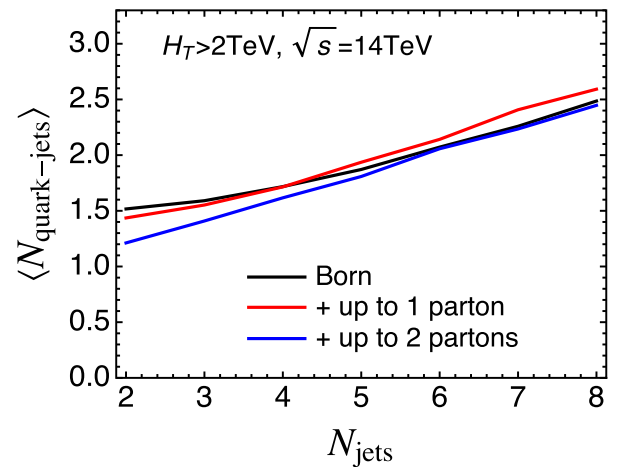


FIG. 10. Matrix element correction to the number of quark jets with Catani-Krauss-Kuhn-Webber matching. The black curve is the result for  $pp \rightarrow jj + \text{parton showers}$ . One and two partons are matched in the Born configuration and the results are shown by the red and blue curves, where we impose  $H_T > 2 \text{ TeV}$  and set  $\sqrt{s} = 14 \text{ TeV}$ .

the number of quark jets  $\langle N_{\text{quark-jets}} \rangle$  for each  $N_{\text{jets}}$  category. One can define a quark or gluon jet by its jet constituents. First, we look for a  $q\bar{q}$  pair in the constituents and convert the pair to a gluon, and add the gluon to the constituent list. We continue this until no pairs are found. After the conversion, if only gluons are in the list we call the jet a gluon jet; otherwise we call it a quark jet. If a quark jet contains only one quark in the list after the conversion we call it a well-defined quark jet; otherwise we call it an ill-defined quark jet. In the definition, the number of quark jets is IR unsafe because the number can change due to a quark of  $q\bar{q}$  that a soft gluon decays into. We temporarily call the effect from the soft gluons  $g_{\text{soft}} \rightarrow q\bar{q}$  pollution. This pollution turns a gluon jet or a well-defined quark jet into a well-defined quark jet or an ill-defined quark jet.<sup>10</sup>

In consideration of the effect, we increase the IR safety by adding a correction to the number of quark jets calculated by the above simple algorithm. We first define two probabilities  $P_{n,m}$  and  $\hat{P}_{n,m}$  which show the quark jet rate calculated in a system where we can ignore  $g_{\text{soft}} \rightarrow q\bar{q}$  pollution and one where we cannot. Here, the quark jet rate shows the probability that events contain  $m$  quark jets and  $n$  total jets.  $\hat{P}_{n,m}$  can be expressed by the following sum:

$$\hat{P}_{n,m} = \hat{P}_{n,m}^{(\text{well})} + \hat{P}_{n,m}^{(\text{ill})}. \quad (\text{A32})$$

$\hat{P}_{n,m}^{(\text{well})}$  shows the probability that events contain  $m$  quark jets and  $n$  total jets, where the quark jets are all well-defined jets.  $\hat{P}_{n,m}^{(\text{ill})}$  is the probability that the ill-defined jet is contained. Using Monte Carlo samples, we can calculate the two probabilities with the jet flavor definition explained in the above paragraph.

The ill-defined probability is approximately given as,

$$\hat{P}_{n,m}^{(\text{ill})} \simeq m\epsilon P_{n,m}, \quad (\text{A33})$$

where  $\epsilon$  shows the probability that  $g_{\text{soft}} \rightarrow q\bar{q}$  pollution gets into a jet. This shows the contribution coming from the process where a well-defined quark jet turns into an ill-defined quark jet due to the pollution. There is another relation,

$$\hat{P}_{n,m}^{(\text{well})} \simeq (1 - n\epsilon)P_{n,m} + [n - (m - 1)]P_{n,m-1}. \quad (\text{A34})$$

The first term is the contribution for the case that  $g_{\text{soft}} \rightarrow q\bar{q}$  pollution does not affect the jet flavor of any jets. The second term is the contribution for the case that a gluon jet turns into well-defined quark jets due to the pollution. Then, the ratio of the measurable probabilities is,

<sup>10</sup>We ignore the fact that more than one quark contamination gets into a single jet and also that  $g_{\text{soft}} \rightarrow q\bar{q}$  pollution turns a well-defined quark jet into a gluon jet.

$$\frac{\hat{P}_{n,m}^{(\text{ill})}}{\hat{P}_{n,m}} = m\epsilon + \mathcal{O}(\epsilon^2). \quad (\text{A35})$$

We denote the expected value of the number of quark jets in a system where  $g_{\text{soft}} \rightarrow q\bar{q}$  pollution can be neglected or cannot be neglected as  $\langle N_{\text{quark-jets}} \rangle$  or  $\langle \hat{N}_{\text{quark-jets}} \rangle$ , and the values can be related to the quark jet rates as,

$$\langle N_{\text{quark-jets}} \rangle = \frac{\sum_{m=0}^n m P_{n,m}}{\sum_{m=0}^n P_{n,m}}, \quad (N, P) \leftrightarrow (\hat{N}, \hat{P}). \quad (\text{A36})$$

Finally, we obtain the following expression:

$$\langle N_{\text{quark-jets}} \rangle = \langle \hat{N}_{\text{quark-jets}} \rangle - (n - \langle \hat{N}_{\text{quark-jets}} \rangle)\epsilon + \mathcal{O}(\epsilon^2). \quad (\text{A37})$$

Ignoring the  $\mathcal{O}(\epsilon^2)$  term, we estimate  $\epsilon$  from Eq. (A35) and correct  $\langle \hat{N}_{\text{quark-jets}} \rangle$  to  $\langle N_{\text{quark-jets}} \rangle$  with Eq. (A37). We employ  $\langle N_{\text{quark-jets}} \rangle$  as the Monte Carlo results, which is more appropriate to compare with analytical results in terms of IR safety.

One may come up with a way to reduce IR unsafety from  $g_{\text{soft}} \rightarrow q\bar{q}$  pollution with the flavor- $k_t$  algorithm [96], but it is not simple to apply it to multijet final states. In our analytical calculation a jet is defined by its jet radius  $R$  and  $p_{T,\text{cut}}$ , and in order to cluster jets based on this definition, it is necessary to partially modify the algorithm. First, in order to cluster unresolved emission, where the distance between a jet core and the unresolved emission is less than  $R$ , it is necessary to modify the distance factor in the measure of the flavor- $k_t$  algorithm for hadron colliders  $d_{ij}^{(F)}$  as  $\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2 \rightarrow (\Delta\eta_{ij}^2 + \Delta\phi_{ij}^2)/R$ . Also, the beam measure  $d_{iB}^{(F)}$  for a quark (or *flavored* particle) in this algorithm is larger than the jet  $p_T$  employed in commonly used algorithms. As a result, when a jet is clustered with the inclusive variant of the algorithm without introducing  $d_{\text{cut}}$ , hard particles separated by  $R$  or more can be clustered. This does not match our jet definition. If we introduce  $d_{\text{cut}}$  and cluster jets with the exclusive variant, it would be appropriate to choose  $d_{\text{cut}} = p_{T,\text{cut}}$ . If  $d_{\text{cut}}$  is set to greater or less than  $p_{T,\text{cut}}$ , the minimum value of the jet  $p_T$  will be greater or less than  $p_{T,\text{cut}}$ . However, even if we set  $d_{\text{cut}} = p_{T,\text{cut}}$ ,  $d_{ij}^{(F)}$  and  $d_{iB}^{(F)}$  become easily larger than  $d_{\text{cut}}$  as the hard process scale increases. In the case that  $d_{ij}^{(F)}$  and  $d_{iB}^{(F)}$  are larger than  $d_{\text{cut}}$  for  $i$  and  $j$ , the two partons are declared as jets even if the distance between  $i$  and  $j$  is less than  $R$ . This also does not match our jet definition. While avoiding problems such as those noted above, we might be able to devise a good way to define IR-safe inclusive jets in multijet final states, but this time we counted the number of quark jets in the way written in the above paragraph.

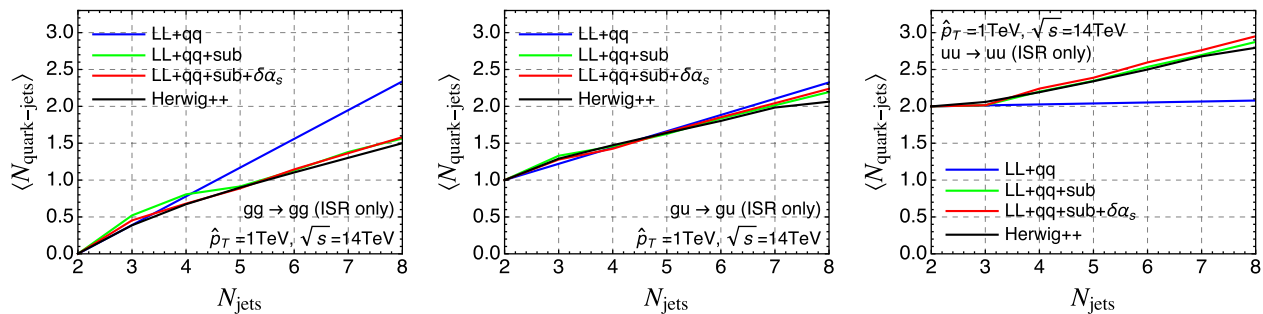


FIG. 11. Same as Fig. 3, considering only the initial-state radiation.

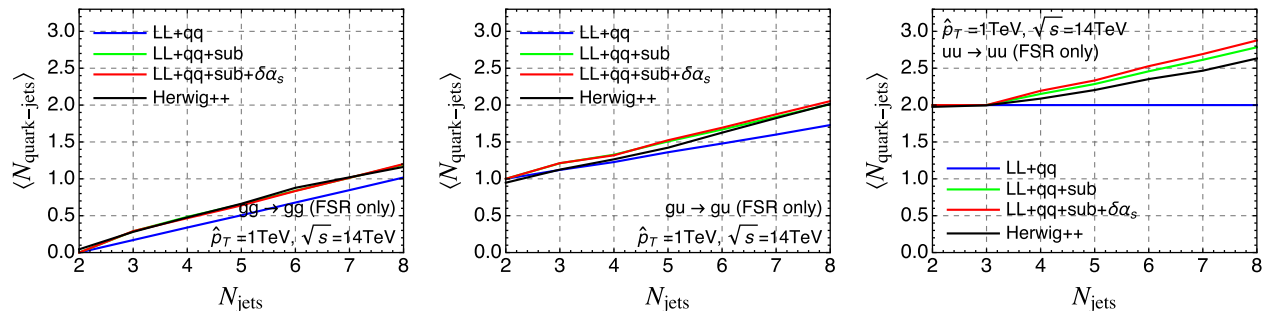


FIG. 12. Same as Fig. 3, considering only the final-state radiation.

#### 4. Initial- and final-state radiation

Generating functionals for initial-state radiation (ISR) and final-state radiation (FSR) are calculated in Sec. II. The number of quark jets is evaluated using the functionals for three Born configurations in Sec. III. In the calculation, the contributions from ISR and FSR to the number are combined. In Figs. 11 and 12, we show results in which only ISR and FSR are taken into account. The results for  $gg \rightarrow gg$  (left),  $gu \rightarrow gu$  (center) and  $uu \rightarrow uu$  (right) are shown. In the calculation,  $p_0 = 50$  GeV, and  $R = 0.4$  are used. We set  $\hat{p}_T$  to 1 TeV. The blue, green and red curves are the analytical calculations using the functionals labeled by  $(LL + q\bar{q})$ ,  $(LL + q\bar{q} + \text{sub})$  and  $(LL + q\bar{q} + \text{sub} + \delta\alpha_s)$ . The black curves show the Monte Carlo prediction given by HERWIG++.

When we neglect subsequent emissions, the increase in the number of quark jets for  $uu \rightarrow uu$  in the ISR-only case is tiny for the same reason as discussed in Fig. 3. In the FSR-only case, the generating functional does not contain  $v$ , so the number of quark jets does not increase at all. The main cause of the increase in the number of quark jets for  $uu \rightarrow uu$  stems from  $\exp(S_q)$  which is related to subsequent emissions, and the lowest order at which  $v$  appears is  $\mathcal{O}(u^4 v^4)$ ; therefore, the number of quark jets begins to increase clearly from  $N_{\text{jets}} = 4$  as discussed in Sec. III. For the ISR-only case and  $gg \rightarrow gg$ , the number of quark jets decreases a lot when we take into account the subsequent emissions in our analytic calculation. This is mainly because the coefficient for  $uv$  in Eq. (68) takes a large negative value, which stems from the improvement of the approximation to the generating functional ratio in Eq. (25).

- [1] A. Abdesselam *et al.*, *Eur. Phys. J. C* **71**, 1661 (2011).
- [2] A. Altheimer *et al.*, *J. Phys. G* **39**, 063001 (2012).
- [3] A. Altheimer *et al.*, *Eur. Phys. J. C* **74**, 2792 (2014).
- [4] D. Adams *et al.*, *Eur. Phys. J. C* **75**, 409 (2015).
- [5] P. Gras, S. Höche, D. Kar, A. Larkoski, L. Lönnblad, S. Plätzer, A. Siódmok, P. Skands, G. Soyez, and J. Thaler, *J. High Energy Phys.* **07** (2017) 091.

- [6] A. J. Larkoski, I. Moulton, and B. Nachman, [arXiv:1709.04464](https://arxiv.org/abs/1709.04464).
- [7] C. F. Berger, T. Kucs, and G. F. Sterman, *Phys. Rev. D* **68**, 014012 (2003).
- [8] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, *Phys. Rev. Lett.* **100**, 242001 (2008).
- [9] L. G. Almeida, S. J. Lee, G. Perez, G. F. Sterman, I. Sung, and J. Virzi, *Phys. Rev. D* **79**, 074017 (2009).



- [10] J. Thaler and K. Van Tilburg, *J. High Energy Phys.* **03** (2011) 015.
- [11] M. Jankowiak and A. J. Larkoski, *J. High Energy Phys.* **06** (2011) 057.
- [12] J. Gallicchio and M. D. Schwartz, *Phys. Rev. Lett.* **107**, 172001 (2011).
- [13] J. Thaler and K. Van Tilburg, *J. High Energy Phys.* **02** (2012) 093.
- [14] D. Krohn, M. D. Schwartz, T. Lin, and W. J. Waalewijn, *Phys. Rev. Lett.* **110**, 212001 (2013).
- [15] J. Gallicchio and M. D. Schwartz, *J. High Energy Phys.* **04** (2013) 090.
- [16] A. J. Larkoski, G. P. Salam, and J. Thaler, *J. High Energy Phys.* **06** (2013) 108.
- [17] A. J. Larkoski, I. Moulton, and D. Neill, *J. High Energy Phys.* **12** (2014) 009.
- [18] A. J. Larkoski, I. Moulton, and D. Neill, *Phys. Rev. D* **91**, 034035 (2015).
- [19] B. Bhattacharjee, S. Mukhopadhyay, M. M. Nojiri, Y. Sakaki, and B. R. Webber, *J. High Energy Phys.* **04** (2015) 131.
- [20] I. Moulton, L. Necib, and J. Thaler, *J. High Energy Phys.* **12** (2016) 153.
- [21] J. Davighi and P. Harris, *Eur. Phys. J. C* **78**, 334 (2018).
- [22] H. n. Li, Z. Li, and C.-P. Yuan, *Phys. Rev. Lett.* **107**, 152001 (2011).
- [23] I. Feige, M. D. Schwartz, I. W. Stewart, and J. Thaler, *Phys. Rev. Lett.* **109**, 092001 (2012).
- [24] W. J. Waalewijn, *Phys. Rev. D* **86**, 094030 (2012).
- [25] P. Bolzoni, B. A. Kniehl, and A. V. Kotikov, *Phys. Rev. Lett.* **109**, 242002 (2012).
- [26] M. Dasgupta, A. Fregoso, S. Marzani, and G. P. Salam, *J. High Energy Phys.* **09** (2013) 029.
- [27] A. J. Larkoski and J. Thaler, *J. High Energy Phys.* **09** (2013) 137.
- [28] A. J. Larkoski, D. Neill, and J. Thaler, *J. High Energy Phys.* **04** (2014) 017.
- [29] A. J. Larkoski, I. Moulton, and D. Neill, *J. High Energy Phys.* **09** (2014) 046.
- [30] A. J. Larkoski, J. Thaler, and W. J. Waalewijn, *J. High Energy Phys.* **11** (2014) 129.
- [31] M. Procura, W. J. Waalewijn, and L. Zeune, *J. High Energy Phys.* **02** (2015) 117.
- [32] M. Dasgupta, A. Powling, and A. Siodmok, *J. High Energy Phys.* **08** (2015) 079.
- [33] A. J. Larkoski, I. Moulton, and D. Neill, *J. High Energy Phys.* **05** (2016) 117.
- [34] M. Dasgupta, L. Schunk, and G. Soyez, *J. High Energy Phys.* **04** (2016) 166.
- [35] P. Pietrulewicz, F. J. Tackmann, and W. J. Waalewijn, *J. High Energy Phys.* **08** (2016) 002.
- [36] C. Frye, A. J. Larkoski, M. D. Schwartz, and K. Yan, *J. High Energy Phys.* **07** (2016) 064.
- [37] M. Dasgupta, A. Powling, L. Schunk, and G. Soyez, *J. High Energy Phys.* **12** (2016) 079.
- [38] G. P. Salam, L. Schunk, and G. Soyez, *J. High Energy Phys.* **03** (2017) 022.
- [39] Z. B. Kang, F. Ringer, and W. J. Waalewijn, *J. High Energy Phys.* **07** (2017) 064.
- [40] X. Liu, S. O. Moch, and F. Ringer, *Phys. Rev. Lett.* **119**, 212001 (2017).
- [41] X. Liu, S. O. Moch, and F. Ringer, *Phys. Rev. D* **97**, 056026 (2018).
- [42] Z. B. Kang, K. Lee, X. Liu, and F. Ringer, *J. High Energy Phys.* **10** (2018) 137.
- [43] CMS Collaboration, Report No. CMS-PAS-JME-13-002.
- [44] D. F. de Lima, P. Petrov, D. Soper, and M. Spannowsky, *Phys. Rev. D* **95**, 034001 (2017).
- [45] B. Bhattacharjee, S. Mukhopadhyay, M. M. Nojiri, Y. Sakaki, and B. R. Webber, *J. High Energy Phys.* **01** (2017) 044.
- [46] A. Chakraborty, A. M. Iyer, and T. S. Roy, *Nucl. Phys.* **B932**, 439 (2018).
- [47] M. Park and M. Zhang, [arXiv:1712.09279](https://arxiv.org/abs/1712.09279).
- [48] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **09** (2013) 076.
- [49] G. Aad *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **76**, 154 (2016).
- [50] ATLAS Collaboration, Report No. ATLAS-CONF-2015-035.
- [51] ATLAS Collaboration, Report No. ATLAS-CONF-2016-039.
- [52] G. Aad *et al.* (ATLAS Collaboration), *Phys. Rev. D* **84**, 054001 (2011).
- [53] G. Aad *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **71**, 1795 (2011).
- [54] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **05** (2012) 128.
- [55] S. Chatrchyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **06** (2012) 160.
- [56] V. Khachatryan *et al.* (CMS Collaboration), *J. High Energy Phys.* **08** (2014) 173.
- [57] G. Aad *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **74**, 3023 (2014).
- [58] G. Aad *et al.* (ATLAS Collaboration), *Phys. Rev. D* **93**, 052003 (2016).
- [59] G. Aad *et al.* (ATLAS Collaboration), *Eur. Phys. J. C* **76**, 322 (2016).
- [60] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, *J. High Energy Phys.* **07** (2016) 069.
- [61] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, *J. High Energy Phys.* **01** (2017) 110.
- [62] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, *J. High Energy Phys.* **05** (2017) 006.
- [63] G. Louppe, K. Cho, C. Becot, and K. Cranmer, *J. High Energy Phys.* **01** (2019) 057.
- [64] T. Cohen, M. Freytsis, and B. Ostdiek, *J. High Energy Phys.* **02** (2018) 034.
- [65] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, *J. High Energy Phys.* **12** (2017) 051.
- [66] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, *SciPost Phys.* **5**, 028 (2018). [arXiv:1707.08966](https://arxiv.org/abs/1707.08966).
- [67] E. M. Metodiev, B. Nachman, and J. Thaler, *J. High Energy Phys.* **10** (2017) 174.
- [68] T. Cheng, *Comput. Software Big Sci.* **2**, 3 (2018).
- [69] H. Luo, M. x. Luo, K. Wang, T. Xu, and G. Zhu, [arXiv:1712.03634](https://arxiv.org/abs/1712.03634).
- [70] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, *Phys. Rev. D* **98**, 011502 (2018).

- [71] S. Macaluso and D. Shih, *J. High Energy Phys.* **10** (2018) 121.
- [72] K. Fraser and M. D. Schwartz, *J. High Energy Phys.* **10** (2018) 093.
- [73] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **07** (2012) 167.
- [74] S. Chatrchyan *et al.* (CMS Collaboration), *Phys. Rev. Lett.* **109**, 171803 (2012).
- [75] V. Khachatryan *et al.* (CMS Collaboration), *Phys. Lett. B* **743**, 503 (2015).
- [76] G. Aad *et al.* (ATLAS Collaboration), *Phys. Lett. B* **757**, 334 (2016).
- [77] V. Khachatryan *et al.* (CMS Collaboration), *Phys. Lett. B* **770**, 257 (2017).
- [78] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Rev. D* **96**, 032003 (2017).
- [79] M. Aaboud *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **09** (2017) 088.
- [80] S. Chatrchyan *et al.* (CMS Collaboration), *J. High Energy Phys.* **07** (2013) 178.
- [81] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **08** (2014) 103.
- [82] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **07** (2015) 032.
- [83] G. Aad *et al.* (ATLAS Collaboration), *J. High Energy Phys.* **03** (2016) 026.
- [84] A. M. Sirunyan *et al.* (CMS Collaboration), *Phys. Lett. B* **774**, 279 (2017).
- [85] P. L. Cho and E. H. Simmons, *Phys. Rev. D* **51**, 2360 (1995).
- [86] F. Krauss, S. Kuttimalai, and T. Plehn, *Phys. Rev. D* **95**, 035024 (2017).
- [87] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *J. High Energy Phys.* **11** (2001) 063.
- [88] F. Krauss, *J. High Energy Phys.* **08** (2002) 015.
- [89] M. L. Mangano, M. Moretti, and R. Pittau, *Nucl. Phys.* **B632**, 343 (2002).
- [90] L. Lonnblad, *J. High Energy Phys.* **05** (2002) 046.
- [91] L. Lonnblad and S. Prestel, *J. High Energy Phys.* **03** (2012) 019.
- [92] K. Konishi, A. Ukawa, and G. Veneziano, *Nucl. Phys.* **B157**, 45 (1979).
- [93] Y. L. Dokshitzer, V. A. Khoze, A. H. Mueller, and S. I. Troian, *Basics of Perturbative QCD* (Ed. Frontieres, Gif-sur-Yvette, France, 1991), p. 274.
- [94] R. K. Ellis, W. J. Stirling, and B. R. Webber, *Nucl. Phys., Cosmol.* **8**, 1 (1996).
- [95] E. Gerwick, S. Schumann, B. Gripaios, and B. Webber, *J. High Energy Phys.* **04** (2013) 089.
- [96] A. Banfi, G. P. Salam, and G. Zanderighi, *Eur. Phys. J. C* **47**, 113 (2006).
- [97] J. Gallicchio and M. D. Schwartz, *J. High Energy Phys.* **10** (2011) 103.
- [98] A. Buckley and C. Pollard, *Eur. Phys. J. C* **76**, 71 (2016).
- [99] P. T. Komiske, E. M. Metodiev, and J. Thaler, *J. High Energy Phys.* **11** (2018) 059.
- [100] M. Dasgupta and G. P. Salam, *Phys. Lett. B* **512**, 323 (2001).
- [101] S. D. Ellis, R. Kleiss, and W. J. Stirling, *Phys. Lett.* **154B**, 435 (1985).
- [102] F. A. Berends, W. T. Giele, H. Kuijff, R. Kleiss, and W. J. Stirling, *Phys. Lett. B* **224**, 237 (1989).
- [103] F. A. Berends, H. Kuijff, B. Tausk, and W. T. Giele, *Nucl. Phys.* **B357**, 32 (1991).
- [104] E. Gerwick, T. Plehn, and S. Schumann, *Phys. Rev. Lett.* **108**, 032003 (2012).
- [105] C. Englert, T. Plehn, P. Schichtel, and S. Schumann, *Phys. Rev. D* **83**, 095009 (2011).
- [106] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky, and W. K. Tung, *J. High Energy Phys.* **07** (2002) 012.
- [107] D. B. Clark, E. Godat, and F. I. Olness, *Comput. Phys. Commun.* **216**, 126 (2017).
- [108] E. Gerwick, T. Plehn, S. Schumann, and P. Schichtel, *J. High Energy Phys.* **10** (2012) 162.
- [109] M. Bahr, S. Gieseke, M. A. Gigg, D. Grellscheid, K. Hamilton, O. Latunde-Dada, S. Platzer, and P. Richardson *et al.*, *Eur. Phys. J. C* **58**, 639 (2008).
- [110] M. Cacciari, G. P. Salam, and G. Soyez, *J. High Energy Phys.* **04** (2008) 063.
- [111] W. T. Giele, D. A. Kosower, and P. Z. Skands, *Phys. Rev. D* **78**, 014026 (2008).
- [112] W. T. Giele, D. A. Kosower, and P. Z. Skands, *Phys. Rev. D* **84**, 054003 (2011).
- [113] A. Gehrmann-De Ridder, M. Ritzmann, and P. Z. Skands, *Phys. Rev. D* **85**, 014013 (2012).
- [114] M. Ritzmann, D. A. Kosower, and P. Skands, *Phys. Lett. B* **718**, 1345 (2013).
- [115] L. Hartgring, E. Laenen, and P. Skands, *J. High Energy Phys.* **10** (2013) 127.
- [116] N. Fischer, S. Prestel, M. Ritzmann, and P. Skands, *Eur. Phys. J. C* **76**, 589 (2016).
- [117] Y. Sakaki, *J. High Energy Phys.* **08** (2015) 100.
- [118] D. Reichelt, P. Richardson, and A. Siodmok, *Eur. Phys. J. C* **77**, 876 (2017).
- [119] A. Hook, E. Izaguirre, M. Lisanti, and J. G. Wacker, *Phys. Rev. D* **85**, 055029 (2012).
- [120] S. El Hedri, A. Hook, M. Jankowiak, and J. G. Wacker, *J. High Energy Phys.* **08** (2013) 136.
- [121] T. Cohen, M. Jankowiak, M. Lisanti, H. K. Lou, and J. G. Wacker, *J. High Energy Phys.* **05** (2014) 005.
- [122] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *J. High Energy Phys.* **07** (2014) 079.
- [123] T. Gleisberg, S. Hoeche, F. Krauss, A. Schlicke, S. Schumann, and J. C. Winter, *J. High Energy Phys.* **02** (2004) 056.
- [124] T. Gleisberg, S. Hoeche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert, and J. Winter, *J. High Energy Phys.* **02** (2009) 007.