RAPID COMMUNICATIONS

PHYSICAL REVIEW D **83,** 011502(R) (2011)

Rest frame subjet algorithm with SISCone jet for fully hadronic decaying Higgs search

Ji-Hun Kim[1,2]

[1] *FPRD and Department of Physics, Seoul National University, Seoul, 151-747, Korea*
[2] *Theory Division, CERN, CH-1211 Geneva 23, Switzerland*

(Received 14 November 2010; published 27 January 2011)

The rest frame subjet algorithm is introduced to define the subjets for the SISCone jet; with this algorithm, an infrared and collinear safe jet shape observable N-subjettiness τ_N^j is defined to discriminate the fat jet, from a highly boosted color singlet particle decaying to N partons, from the QCD jet. Using rest frame subjets and τ_2^j on dijets from highly boosted $H/W/Z$ bosons through $pp \to HW, HZ$ with $m_H = 120$ GeV, we found that the statistical significance of the signal, from the fully hadronic channels, is about 2σ for 14 TeV collisions with $\mathcal{L} \sim 30\ \text{fb}^{-1}$.

DOI: 10.1103/PhysRevD.83.011502 PACS numbers: 13.87.Fh, 13.87.Ce, 14.80.Bn

Current electroweak precision fits of the standard model (SM) prefer a light Higgs boson [1]; experimental bounds from LEP [1] and Tevatron [2] suggest the SM Higgs mass m_H is about 120 GeV. However, the SM Higgs boson $m_H \lesssim 135$ GeV has been considered hard to discover because it dominantly decays to a b-quark pair. Unlike the signals from leptonic decay modes, those from hadronic decay channels undergo overwhelming QCD background processes. Finding the Higgs boson with $m_H \lesssim 135$ GeV requires combining many possible decay channels such as $H \to \gamma\gamma$, WW, $b\bar{b}$, and $\tau\tau$.

Recently, there has been a study on the jet substructure via $pp \to WH, ZH$ where H is highly boosted to form a single fat jet and W/Z decays leptonically [3]. According to ATLAS detector simulation [4], it is expected to also be a promising channel for the Higgs search. Other subjet techniques are also proposed for reconstructing the mass peak of heavy particles [5–7] and reducing the QCD contamination effects [8]. The jet radius also plays an important role for the mass peak reconstruction [9], and several schemes for optimizing the jet radius to improve the invariant mass distribution are proposed [10,11].

Beside jet substructure, Ref. [12] suggested a "template overlap" method to match jet energy flow with the partons directly. The property of color structure of the Higgs decay has also been studied for searching the new particle via the double diffractive process [13] and the Higgs boson decaying to the two b-tagged jets [14].

To search heavy particles, however, those techniques usually require signal processes to involve additional unusual signatures, such as a lepton or missing p_T, to suppress the QCD background; fully hadronic decay channels, such as Higgs production associated with the W/Z boson, $pp \to HW, HZ$, are still considered too hard to be used for the heavy particle study. Usually, however, cross sections of hadronic decay channels are, at least, a few times larger than (semi)leptonic decay channels; utilizing them will improve, for instance, the Higgs discovery potential.

In this paper, to identify boosted color singlet particles, we introduce the subjet definition for the SISCone jet [15] and an infrared and collinear safe jet shape observable, τ_N^j. With τ_2^j, as an illustration, we investigate the statistical significance of the signals from fully hadronic decay channels of light SM Higgs bosons.

The SISCone jet algorithm is one of the best algorithms for reconstructing mass peaks; i.e. it has low sensitivity to underlying event (UE) contamination [9,16]. In many cases including light SM Higgs searches, however, such mass peaks are ruined by huge QCD background contributions. To suppress the background, the underlying structure of the jet is essential, but the SISCone jet algorithm misses the information. Moreover, the jet substructure is also required for the boosted Higgs jet tagging. The boosted Higgs jets contain not a single, but two secondary vertices originating from two b quarks. The identification of the two b quarks is crucial for separating the signal from the large background; since gluon splitting into b quarks is a major source of QCD background, the directions of the subjets can be used to reduce the background [4]. Previously, such two-b subjet tagging algorithms were devised for sequential recombination jet algorithms, for example, the Cambridge/Aachen (C/A) [17] algorithm. With a subjet definition for the SISCone jet, however, those two-b subjet tagging algorithms can also be employed with the SISCone jet. These reasons have motivated us to devise a subjet algorithm using the SISCone jet algorithm for finding a boosted particle while reducing QCD background.

We considered several possible ways to define the subjet of the SISCone jet. One is, for a given jet, clustering its constituents with a sequential recombination jet algorithm such as C/A with mass drop and filtering (MD-F) techniques [3]. However, sequential recombination jet algorithms require a larger initial jet radius than the SISCone jet algorithm to catch perturbative radiation of the particle; applying C/A MD-F to the SISCone jet is less efficient than using C/A MD-F solely. Another one is clustering the constituents of the jet by using the SISCone algorithm with a smaller jet radius, i.e. applying the jet trimming algorithm [8]. As the Higgs boson is boosted highly, however, the cones which define subjets are likely to overlap.

Although the SISCone algorithm merges the overlapping cones or splits them according to the overlap threshold parameter, we found that this subjet definition is inefficient, too. The last method we considered is to define the subjet in the "rest frame" of the jet.

The fat jet from a boosted color singlet particle is, to a good approximation, a closed system; i.e. it does not interact with the rest of the system when it hadronizes. Thus, the rest frame of the particle is identical to the c.m. frame of its decay products. In contrast, a jet from a colored particle is not a closed system. Although colored partons are almost free at high energy due to asymptotic freedom, they must hadronize and the hadronization process involves exchanging the four-momentum. Because the four-momentum of the jet is different from that of the hard parton, the rest frame of a colored particle has ambiguities.

We define the jet rest frame as a frame where the four-momentum of the jet equals $p_\mu^{\text{rest}} \equiv (m_{\text{inv}}^{\text{jet}}, 0, 0, 0)$.[1] A jet consists of its constituent particles. The distribution of constituent particles of a fat jet, in their center of mass frame, is nearly identical to those of the color singlet particle which is produced at rest. Thus, we recluster them in the jet rest frame, i.e. by using their four-momenta at the rest frame of the jet. For a given jet, the "rest frame subjets" are defined as these reclustered jets. This procedure is illustrated in Fig. 1.

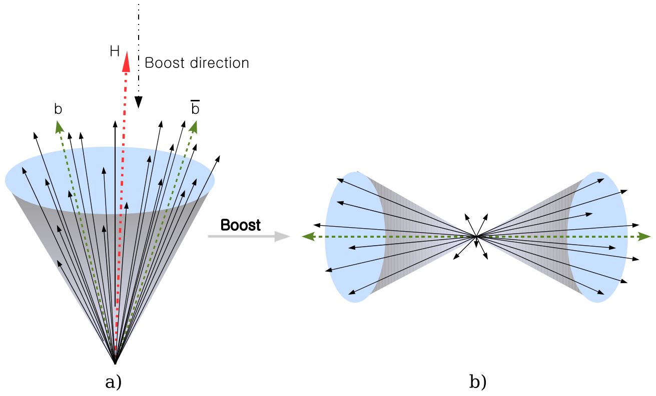

FIG. 1 (color online). Illustration of the jet and the rest frame subjets. (a) The lab frame. (b) The jet rest frame.

The UE affects the jet mass and, thus, the jet rest frame. Consider a fat jet of $p_\mu = \gamma m(1, \vec{\beta})$, defined by a cone with angle θ_c. In the jet rest frame, the energy of massless UE particles can be, at most, $\gamma(1 - \beta \cos\theta_c)$ times larger than its energy in the lab frame. For UE particles with $E < 1$ GeV, inside the SISCone jet with a jet radius 0.8 and $\gamma \sim 10$, the factor is about 3; soft massless objects in the lab frame remain soft in the jet rest frame. Moreover, they are collimated to the boost axis, $\cos\theta \sim 0.97$, and unlikely to include the leading rest frame subjets; to be identified as a single fat jet, the angles between the axis and the leading rest frame subjets are usually larger. Thus, the rest frame subjet algorithm is infrared and collinear safe if an infrared and collinear safe jet algorithm is used for the rest frame subjet clustering, and the leading rest frame subjets of fat jets are not much affected by UE particles.

In the jet rest frame, we can treat the constituent particles of the jet as the final state particles of a fictitious event, because it is the center of mass frame for the particles. For example, in the rest frame of the Higgs jet, the constituent particles look like a dijet event from the Higgs boson produced at rest. To have such a shape, the QCD jet should radiate only one hard parton which is improbable. Since the gluon and quark jets are not closed systems, their shape in the rest frame does not correspond to any physical state and is more likely to be irregular. Thus, by checking whether a shape of a jet in the rest frame looks like an "N-jet" event, i.e. by analyzing whether the jet has N rest frame subjets, we can discriminate between the fat jet, from the boosted color singlet particle decaying to N partons, and the QCD jet.

To select jets which have N rest frame subjets, we employ the N-jettiness [19]. A global event shape "N-jettiness" τ_N is devised to filter out events which have additional undesired jets beyond the required N jets. Treating the constituent particles of the jet as final state particles of a fictitious event, we can apply τ_N to the jet, i.e. calculate "N-subjettiness," to determine whether the jet has N rest frame subjets. τ_N vanishes in the limit of exactly N infinitely narrow jets; we expect the fat jets to tend to result in smaller τ_N^j than that of the QCD jet.

Except for the fact that it is defined by constituent particles of a jet instead of whole final state particles of an event, the definition of τ_N^j is identical to that of τ_N. In the jet rest frame, we tag the N most energetic subjets and define τ_N^j as

$$\tau_N^j \equiv \frac{2}{(m_{\text{inv}}^{\text{jet}})^2} \sum_{k \in J} \min\{q_1 \cdot p_k, q_2 \cdot p_k, \ldots, q_N \cdot p_k\}, \quad (1)$$

where p_k is four-momenta of the constituent particles of the jet J and q_i is four-momenta of the N energetic subjets. Definitions of τ_N and the rest frame subjet are infrared and collinear safe; and, thus, τ_N^j is also an infrared and collinear safe observable.

τ_N^j is calculated through the following steps. Given a hard jet J with p_μ^J and its constituent particles $\{i | i \in J\}$:

(1) Define the boost vector $\vec{\beta}_J$ to transform p_μ^J to $p_\mu^{J,\text{rest}} \equiv (m_{\text{inv}}^{\text{jet}}, 0, 0, 0)$.
(2) Boost p_μ^i by $\vec{\beta}_J$, and obtain $p_\mu^{i,\text{rest}}$.
(3) Cluster the subjet with $p_\mu^{i,\text{rest}}$ and R_{subjet}.
(4) Sort jets into decreasing E_i^{rest} order, and label the N most energetic subjets.

---

[1] We found that a similar concept for the top tagging exists in Ref. [18].

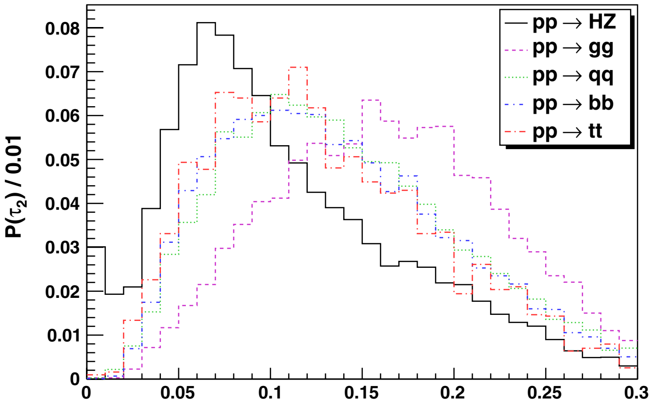

FIG. 2 (color online). τ_2^j distributions of jets with 110 GeV $\leq m_{\rm inv}^{\rm jet} \leq$ 130 GeV, $p_T > 200$ GeV, and $|y| < 2.5$.

(5) Get four-momenta of the N subjets in the lab frame q_μ^i.

(6) $\tau_N^j = \frac{2}{(m_{\rm inv}^{\rm jet})^2}\sum_{k\in J}\min\{q_1\cdot p_k,\ldots,q_N\cdot p_k\}$.

τ_2^j distributions of jets with 110 GeV $\leq m_{\rm inv}^{\rm jet} \leq$ 130 GeV, $p_T > 200$ GeV, $|y| < 2.5$ are shown in Fig. 2.

Applying τ_N^j with $N = 2$, we investigate the discovery potential of the SM Higgs boson with $m_H = 120$ GeV, through the fully hadronic $pp \to WH, ZH$ channels where both W/Z and H are highly boosted enough to be identified as single fat jets. The PYTHIA 6.4.23 [20] with the ATLAS MC09 parameter tune [21] and the modified leading-order MRST2007 parton distribution functions [22] are used to generate both the signal samples $pp \to HV$ and the background samples $pp \to VV$, Vj, tt, tV, tj, and jj events at 14 TeV. No K factors are applied to emulate higher order effects.

FASTJET 2.4.2 [23] with the SISCone plug-in [15] is used for the jet clustering, and SISCone in spherical coordinates is used for the rest frame subjet clustering. The SISCone jet algorithm has two parameters: the jet radius R and the overlap threshold f. For the fat jet tagging using τ_2^j, the jet radius should be set to maximize the mass peak of the signal, and the optimal subjet radius should be set to maximize the discrimination power of τ_2^j. With $m_H = 120$ GeV, we use $R = 0.8$, overlap threshold $f = 0.75$ for jet clustering, and $R = 0.6$, $f = 0.75$ for subjet clustering.

The b-tagging, c-jet misidentification and light-jet misidentification probabilities ϵ_b, ϵ_c, and $\epsilon_{\rm mis}$, respectively, are crucial factors to the analysis of this paper. According to Ref. [4], $\epsilon_b \sim 70\%$ (corresponding to $\approx 50\%$ signal efficiency) with $\epsilon_{\rm mis} \sim 1\%$ and $\epsilon_c \sim 10\%$ is expected to be achieved. Although these values are estimated by using the C/A MD-F algorithm, the SISCone jet with the rest frame subjet algorithm can also provide required information and, thus, is compatible with it. Therefore, we employ these values for the analysis in this paper. For estimating the effects on the double b-tag performance of the rest frame subjet, separate event samples are generated with B mesons and are set to be stable; then, a b tag is assigned to each subjet with probability ϵ_b for the b subjet, ϵ_c for the c subjet, and $\epsilon_{\rm mis}$ for the QCD jet.

To select the signal events which contain two fat jets from the boosted H and V of fully hadronic $pp \to WH, ZH$ channels while reducing QCD backgrounds, we require:

(1) no hard leptons with $p_T > 10$ GeV, except those from B mesons;
(2) two hardest jets j_1 and j_2 with $p_{T,j_{1,2}} > 200$ GeV;
(3) $|y_{j_{1,2}}| < 2.5$ and $|y_{j_1} - y_{j_2}| < 2.0$, where y_{j_i} is rapidity of j_i;
(4) third hardest jet, j_3, to be not hard: $p_{T,j_3} < 30$ GeV;
(5) b cut: Both of the two energetic rest frame subjets of the Higgs candidate jet are b-tagged;
(6) τ_2^j cut: τ_{2,j_1}^j, $\tau_{2,j_2}^j < 0.08$;
(7) $\cos\theta_s$ cut: In the jet rest frame, angles θ_s between two leading rest frame subjets and the boost axis should be large enough to be $\cos\theta_s < 0.8$.

To check whether leptons come from semileptonic decays of the B meson, we use PYTHIA's internal data; we have

TABLE I. Cross section for the signal and background events which one of the two hardest jets has $m_{\rm inv}^{\rm jet}$ in the range between 120 ± 10 GeV and the other jet has $m_{\rm inv}^{\rm jet}$ in the range between ($m_W \pm 10$ or $m_Z \pm 10$) GeV. The SISCone jet with $R = 0.8$ is used.

σ (fb)	p_T cut	y cut	$p_T^{j_3}$ cut
$\sigma(pp \to HV)$	2.5×10^1	2.4×10^1	1.4×10^1
$\sigma(pp \to VV)$	4.1×10^1	3.6×10^1	1.1×10^1
$\sigma(pp \to tt)$	3.4×10^3	3.0×10^3	1.9×10^2
$\sigma(pp \to Vj)$	1.4×10^3	1.3×10^3	4.2×10^2
$\sigma(pp \to gg)$	1.8×10^5	1.5×10^5	2.9×10^4
$\sigma(pp \to qq)$	1.7×10^4	1.1×10^4	3.6×10^3
$\sigma(pp \to qg)$	1.3×10^5	1.0×10^5	2.3×10^4
$\sigma(pp \to bb)$	4.6×10^2	4.2×10^2	1.0×10^2

TABLE II. Cross section for the signal and background events with a mass window of 20 GeV centered on the mass peak after applying the b cut, the τ_2^j cut, or both cuts.

σ (fb)	b cut	τ_2^j cut	b cut & τ_2^j cut
$\sigma(pp \to HV)$	0.4×10^1	0.2×10^1	0.9×10^0
$\sigma(pp \to VV)$	0.2×10^{-1}	0.2×10^{-1}	0.1×10^{-2}
$\sigma(pp \to tt)$	0.4×10^1	0.7×10^1	0.3×10^{-1}
$\sigma(pp \to Vj)$	0.5×10^1	0.2×10^2	0.2×10^0
$\sigma(pp \to gg)$	0.2×10^3	0.1×10^3	0.3×10^1
$\sigma(pp \to qq)$	0.2×10^2	0.8×10^2	0.4×10^0
$\sigma(pp \to qg)$	0.1×10^3	0.2×10^3	0.2×10^1
$\sigma(pp \to bb)$	0.1×10^0	0.3×10^1	0.4×10^{-1}

checked that it gives a good approximation for the lepton isolation for the purpose of this paper.

The $\cos\theta_s$ cut plays similar roles as symmetry cut $y_{\rm cut}$ introduced in Ref. [3]. To be reconstructed properly, two b quarks of the Higgs jet should be energetic; and it results in smaller $\cos\theta_s$. The $\cos\theta_s$ cut also reduces the effects of the UE. As mentioned previously, UE particles are concentrated along the boost axis in the jet rest frame. Since they are mixed with soft particles from the Higgs jet, however, removing these particles is difficult, and removing them affects the jet mass. By requiring the $\cos\theta_s$ cut, we can select the fat jets where the leading two rest frame subjets are not much contaminated by the UE, while preserving the jet mass and reducing QCD backgrounds; for SISCone jets, we found that it is more efficient than removing the soft particles near the boost axis.

We define signal and background events as which one of j_1 and j_2 is the Higgs candidate jet, i.e. has $m_{\rm inv}^{\rm jet}$ in the range between 120 ± 10 GeV and passes the b cut, and the other jet has $m_{\rm inv}^{\rm jet}$ in the range between ($m_W \pm 10$ or $m_Z \pm 10$) GeV. The result before the b-cut step of this scheme is shown in Table I. The cross sections for signal and background events after applying the b cut, the τ_2^j cut, or both cuts are shown in Table II. Finally, about 36% of QCD dijet backgrounds and less than 5% of the Higgs and other background processes are filtered out by the $\cos\theta_s$ cut. With $\mathcal{L} \sim 30\ {\rm fb}^{-1}$, expected $m_{\rm jet}$ distributions are shown in Fig. 3; note that the mass peaks at m_W and m_Z should be clearer than the Higgs signals, and we expect they can be used for the calibrations. The signal to background ratio is about 30/200, and, thus, the statistical significance of the signal is about 2σ. With a more conservative b-tagging efficiency of 60% and a light-quark jet fake rate of 2%, it is decreased to about 1.5σ.

The uncertainties of the statistical significance of signals largely come from mass resolution of jet masses and modeling of gluon splitting into $b\bar{b}$ since most backgrounds come from light-quark jets and gluon jets with the gluon splitting. Both the τ_2^j cut and b tagging use the leading subjets' information, and, thus, there is a loose correlation between them; a jet of lower τ_2^j is more likely to pass the b-tag cut. About 2% of light-quark jets and 6% of gluon jets which have $\tau_2^j < 0.08$ and $m_{\rm inv}^{\rm jet}$ of 120 ± 10 GeV are expected to pass the b cut.

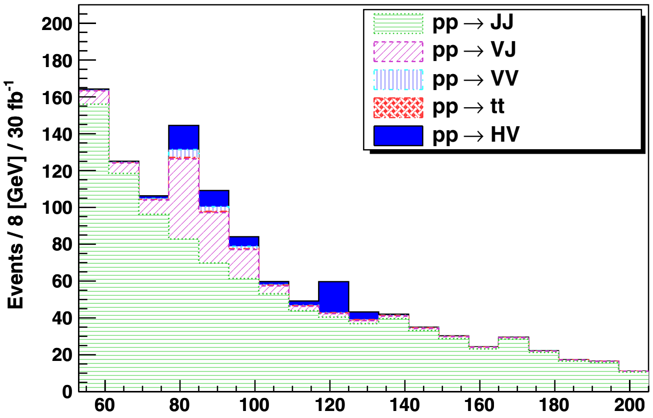

FIG. 3 (color online). Expected $m_{\rm jet}$ distributions of the signal and background events that pass all cuts with $\mathcal{L} \sim 30\ {\rm fb}^{-1}$. PYTHIA with the ATLAS MC09 parameter tune and the SISCone jet with $R = 0.8$ are used. The JJ sample includes bb. The b-tag efficiency, c-jet misidentification, and light-jet misidentification probabilities are assumed to be 70%, 10%, and 1%, respectively.

In conclusion, for the boosted color singlet particle searches, we have introduced the jet rest frame, the rest frame subjet, and the N-subjettiness. Using the SISCone jet with τ_2^j, we see that, for the fully hadronic $pp \to HV$ channels of the SM Higgs boson with $m_H = 120$ GeV, the statistical significance of the signals is about 2σ for 14 TeV collisions with $\mathcal{L} \sim 30\ {\rm fb}^{-1}$. Although it will be complementary to the known Higgs search channels, the scheme suggested in this paper is rather a proof of concept; the scheme will be improved further to increase the signal to background ratio and to make full use of the jet rest frame. It involves comprehensive studies on theoretical uncertainties of the scheme, and we left them for future study. The rest frame subjet can be defined by any jet algorithms, although the effects of the UE and pileup on the scheme depend on the jet algorithm. We also expect the scheme can also be employed for highly boosted colored particles.

The work of J.-H. K. has been supported by the Korean-CERN theory collaboration program and KRF-2008-313-C00162. J.-H. K. thanks Josh Cogan, Hyung Do Kim, and Gavin Salam for useful comments and discussions.

---

[1] J. Alcaraz, arXiv:0911.2604.
[2] CDF and D0 Collaboration, arXiv:1007.4587.
[3] J. M. Butterworth, A. R. Davison, M. Rubin, and G. P. Salam, Phys. Rev. Lett. **100**, 242001 (2008).
[4] ATLAS Collaboration, Report No. ATL-PHYS-PUB-2009-088, 2009.
[5] T. Plehn, G. P. Salam, and M. Spannowsky, Phys. Rev. Lett. **104**, 111801 (2010).

[6] S. D. Ellis, C. K. Vermilion, and J. R. Walsh, Phys. Rev. D **81**, 094023 (2010).

[7] D. E. Soper and M. Spannowsky, J. High Energy Phys. 08 (2010) 029.

[8] D. Krohn, J. Thaler, and L. T. Wang, J. High Energy Phys. 02 (2010) 084.

[9] M. Cacciari, J. Rojo, G. P. Salam, and G. Soyez, J. High Energy Phys. 12 (2008) 032.

[10] D. Krohn, J. Thaler, and L. T. Wang, J. High Energy Phys. 06 (2009) 059.

[11] G. Soyez, J. High Energy Phys. 07 (2010) 075.

[12] L. G. Almeida, S. J. Lee, G. Perez, G. Sterman, and I. Sung, Phys. Rev. D **82**, 054034 (2010).

[13] M. G. Albrow *et al.* (FP420 R and D Collaboration), JINST **4**, T10001 (2009).

[14] J. Gallicchio and M. D. Schwartz, Phys. Rev. Lett. **105**, 022001 (2010).

[15] G. P. Salam and G. Soyez, J. High Energy Phys. 05 (2007) 086.

[16] S. Sapeta, Q. C. Zhang, and Q. C. Zhang, arXiv:1009.1143.

[17] Y. L. Dokshitzer, G. D. Leder, S. Moretti, and B. R. Webber, J. High Energy Phys. 08 (1997) 001; M. Wobisch and T. Wengler, arXiv:hep-ph/9907280.

[18] G. P. Salam, Centre of Mass Top Tagger, http://www.lpthe.jussieu.fr/salam/fastjet/tools.html.

[19] I. W. Stewart, F. J. Tackmann, and W. J. Waalewijn, Phys. Rev. Lett. **105**, 092002 (2010).

[20] T. Sjostrand, S. Mrenna, and P. Z. Skands, J. High Energy Phys. 05 (2006) 026.

[21] ATLAS Report No. ATL-PHYS-PUB-2010-002, 2010.

[22] A. Sherstnev and R. S. Thorne, Eur. Phys. J. C **55**, 553 (2008).

[23] M. Cacciari and G. P. Salam, Phys. Lett. B **641**, 57 (2006).