Double-charming Higgs identification using machine-learning assisted jet shapes

Alexander Lenz,† Michael Spannowsky,‡ and Gilberto Teltalmatzi-Xolocotzi†

† IPPP, Department of Physics, Durham University, Durham DH1 3LE, UK
‡ Nikhef, Science Park 105, 1098 XG Amsterdam

(Dated: August 14, 2017)

We study the possibility of identifying a boosted resonance that decays into a charm pair against different sources of background using QCD event shapes, which are promoted to jet shapes. Using a set of jet shapes as input to a boosted decision tree, we find that observables utilizing the simultaneous presence of two charm quarks can access complementary information compared to approaches relying on two independent charm tags. Focusing on Higgs associated production with subsequent $H \rightarrow c\bar{c}$ decay and on a CP-odd scalar $A$ with $m_A \leq 10$ GeV we obtain the limits $Br(H \rightarrow c\bar{c}) \leq 6.09\%$ and $Br(H \rightarrow A(\rightarrow c\bar{c})Z) \leq 0.01\%$ at 95\% C. L..

I. INTRODUCTION

After the discovery of the Higgs boson [1, 2] a precise determination of its couplings is now of fundamental importance. The couplings of the Higgs boson to the $W$ and $Z$ bosons are already known to be in good agreement with the standard model (SM) expectation, as can be inferred from the measurements of the Higgs decay and production rates by ATLAS and CMS [3–5].

In the SM the Yukawa interaction describes the coupling of the Higgs boson to a fermion $f$ with a strength given by the Yukawa-coupling $y_f^{SM}$. Deviations from the SM expectation can be parametrised by $\kappa_f = y_f/y_f^{SM}$, which can be deduced from a measurement of the signal strength $\mu_f$ defined as $\mu_f = \sigma_H^{SM} Br_{Hf}/(\sigma_H^{SM} Br_{H\bar{f}f}^{SM})$. Here $\sigma_H$ is the Higgs boson production cross section and $Br_{Hf}$ is the branching ratio of the decay process $H \rightarrow f\bar{f}$. Currently the couplings between the Higgs boson and the third generation fermions are consistent with the SM expectations, one gets: \( \mu_t = 2.2 \pm 0.6 \) [6] (see [4] for slightly older values), $\mu_b = 0.90 \pm 0.18 \pm 0.25$ and $\mu_\tau = 0.98 \pm 0.18$ [7]. However, much less is known about the couplings of the Higgs boson to fermions of the first two families: the current bounds found by ATLAS [10] and CMS [11] are $\mu_e \leq 4 \times 10^4$ and $\mu_\mu \leq 7$. During LHC’s High-Luminosity run $\mu_\mu \simeq 1$ might be achievable [12], while the electron coupling to the Higgs is far below the experimental sensitivity. Here a future $e^+e^-$ collider could get close to the SM value [13, 14].

In this paper we focus on the coupling of the charm quark to the Higgs boson. Besides a measurement of the exclusive branching ratio $H \rightarrow J/\psi \gamma$ [15], yielding $\kappa_c \leq 220$ [16, 19] inclusive $H \rightarrow c\bar{c}$ decays were studied e.g. in [7, 18, 20–22].

A global fit to Higgs signal strengths gives the strongest bound of $\kappa_c \leq 6.2$ [7]. Modifications of the charm Yukawa coupling can occur in different new physics models [23–28], it can even be zero [6]. Our aim is to develop a strategy that allows to set a direct upper limit on the charm Yukawa coupling.

The improvement in our bounds on $\mu_c$, derived from inclusive analyses, depends strongly on the $c$-tagging efficiency at the LHC. While dedicated charm tagging algorithms are relatively new [29], flavour tagging has been used in the identification of jets derived from the hadronization of $b$ quarks for more than 20 years, and were employed at the Tevatron for the discovery of the top quark [30, 31]. Two features of the $b$-mesons are exploited to achieve a good $b$-tagging performance: 1) the dominance of semileptonic rates when a $b$-hadron decays and 2) the long life-time of $b$-hadrons. For the latter one can search for displaced secondary vertices (decay vertex) of $b$-hadrons with respect to the primary vertex (interaction point) in a given event. This distance, known as impact parameter, is normally larger for $b$-hadrons in comparison with that associated with states obtained from the hadronization of light quarks ($u,d,s$) and gluons. A similar approach can be followed for $c$-jets. However, as tagging procedures for $b$-jets and $c$-jets are quite similar, their mutual mis-identification rates are consequently quite large.

In general, bottom- or charm-taggers are designed to find jets initiated by individual $b$ or $c$ quarks, allowing for a generic use of these algorithms in a wide range of applications. However, in searches for light or boosted resonances that decay into a charm or bottom pair, such algorithms might not be ideal, as they neglect correlations between the decay products. For example, if the decaying resonance is a colour singlet particle, its decay products are colour connected and soft gluon emissions of either decay product have a preference to be emitted into the cone between the quark pair [32]. Thus to increase the sensitivity in searches for new physics or Higgs boson measurements it can be beneficial to design dedicated 2-prong reconstruction algorithms that allow to utilise more information about the decaying quark.
resonances. Observables that are particularly sensitive to the radiation profile of the event are so-called event shape observables [33, 34], which have been proposed as hypothesis-tester in the study of Higgs boson properties [35, 36]. By promoting those well-studied observables to jet shape observables, applied to a fat jet, they can be used as input to a machine-learning algorithms to separate signal from large QCD backgrounds.

In this letter we present a procedure to identify jets initiated by $c\bar{c}$ pairs from Higgs boson decays based on the application of different event shapes and the transverse momenta of leptons ($e^\pm$ and $\mu^\pm$). It is expected that high-$p_T$ jets arising from highly-boosted Higgs bosons have a different energy flow in comparison to jets arising from pure QCD backgrounds. We would like to emphasize that in the double tagging strategy presented in this work, we study the energy distribution of the full jet associated with the boosted Higgs bosons when detector effects are also included.

The structure of the paper is as follows: In Sec. II we describe the event generation and the selection criteria. Then in Sec. III we present the performance of our approach for the selection of the SM Higgs boson $H$ against different sources of background. Using the tagging efficiencies derived from the optimization against QCD $c$-jets, we present an upper bound for our sensitivity to $Br(H \to c\bar{c})$. In order to evaluate the efficiency of the simultaneous double $c$-tagging identification with strategies based on the double application of a single $c$-tagger, we compare our results with those obtained applying the Atlas single charm tagging algorithm JetFit-T-charm. Sec. IV is devoted to the study of the decay channel $H(\to A(\to c\bar{c})Z) +$ jets, with $A$ the THDM CP odd scalar. Finally in Sec. V we conclude. The discussion is complemented with the correlation matrices among the event shapes used as well as with the distributions for the leading ones in each one of our studies. A brief description of most of the observables considered is included in the appendix.

II. EVENT GENERATION AND EVENT SELECTION

The signal channels are $pp \to H(\to c\bar{c})Z$ and $pp \to H(\to A(\to c\bar{c})Z) +$ jets. Here $H$ is the SM Higgs boson and $A$ denotes the CP odd THDM scalar. As background channel we include $pp \to Z +$ jets. In all cases we consider $Z \to l^+l^−$, for $l = e, \mu$. We take into account two possible values for the mass of the scalar $A$, $m_A = 4$ GeV and $m_A = 10$ GeV. We generate our samples with SHERPA 2.2.1 [37] at $\sqrt{s} = 13.0$ TeV, and include parton shower, hadronization and underlying event contributions. For the jet reconstruction we use the jet finding package FastJet 3.2.1 [38]. The event selection is performed with the version 2.4.2 of the RIVET analysis framework [39].

Our selection strategy is based on the identification of the Higgs and a $Z$ boson in the highly boosted regime, when both particles have a large transverse momentum and are back-to-back. In order to reconstruct the $Z$ boson we require two isolated leptons $l^+l^−$ (for $l = e, \mu$) with a combined mass satisfying $80.0$ GeV $< m_{ll} < 100.0$ GeV. A lepton $l$ will be considered isolated if the following inequality is satisfied $E_l/E_R < 0.1$, where $E_l$ is the energy of $l$ and $E_R$ is the total energy inside a cone of radius $R = 0.3$ around $l$. The identification of the $Z$ boson concludes by imposing a cut $p_T > 200.0$ GeV over the combined transverse momentum of the pair $l^+l^-$. We proceed with the next steps only if the $Z$ boson has been successfully reconstructed as described in this paragraph.

A boosted Higgs decaying into a pair of quarks $q\bar{q}$ produces a jet with a relatively large active area $R_{q\bar{q}}$ and thus is commonly referred to as fat jet. As a matter of fact, in the boosted regime, the radius of the jet depends on the mass and the transverse momentum of the Higgs ($m_H$ and $p_T,H$) as well as on the momentum fractions of the quark and the antiquark ($z$ and $1−z$) according to $R_{q\bar{q}} = m_H/(p_T,H\sqrt{1−z})$. Thus, for a Higgs boson of mass $m_H \simeq 125$ GeV and a transverse momentum $p_T \simeq 200$ GeV decaying symmetrically into a pair charm-anticharm, we expect an angular separation of the Higgs decay products of $\Delta \phi \simeq 1.25$. In practice we demand jets with radius $R = 1.2$ and a transverse momentum $p_T > 200$ GeV reconstructed, with the anti-$k_T$ algorithm and select the Jet with the highest $p_T$. We translate all the constituents of this jet to the plane $\eta = 0$ by taking $p_\eta = 0$ and replacing their total energy by their corresponding transverse energy [40].

From NLO-QCD calculations, in the SM the decay fractions of $c$-quarks into leptons obey with good approximation $Br(c \to l\nu X) = (21.74 \pm 3.90)\%$ and $Br(c \to X') = 100\% − Br(c \to l\nu X)$ where $l = e, \mu$ and $X,X'$ denote quark final states. Hence, if we consider jets originated from the hadronization process of $c\bar{c}$ pairs, we can expect to find 0, 1 and 2 leptons with the following probabilities $61.24\%$, $34.03\%$ and $4.73\%$ respectively. For each one of our analyses we perform three independent studies: non-leptonic, single leptonic and double leptonic, if zero, one and two non-isolated leptons are found inside the fat jet respectively. A cut in the transverse momentum of the leptons of $p_{T,i} \geq 2.0$ GeV allow us to reproduce these numbers with good approximation. Nevertheless, we consider this to be a relatively soft cut, hence in practice we impose
III. STANDARD MODEL HIGGS \( c\bar{c} \)-TAGGING USING EVENT SHAPES

A. Performance

We begin by obtaining the performance of our strategy when selecting the signal channel \( pp \to H(\to c\bar{c})Z \) against \( pp \to Z + \) jets. The set of observables that give us the best performance are presented in Table I and the correlations among them are shown in Figs. 6, additionally we provide the distributions for the top two discriminating observables in each one of the leptonic categories in Fig. 10.

Our curves for the signal selection efficiencies as well as our background fake rates in each one of the leptonic categories are shown in Fig. 11. We can combine the three leptonic studies to obtain a single selection efficiency for signal (S) and for background (B) according to the formula

\[
\varepsilon_{S/B}^{Tot} = \varepsilon_{S/B}^{(0)} f^{(0)}_{S/B} + \varepsilon_{S/B}^{(1)} f^{(1)}_{S/B} + \varepsilon_{S/B}^{(2)} f^{(2)}_{S/B}.\]

Our optimal point after the combination of the different leptonic categories corresponds to

\[
\varepsilon_{c\bar{c}} = 0.40 \quad \varepsilon_{QCD,jets} = 0.03
\]

obtained from the following partial efficiencies

\[
\begin{align*}
\varepsilon_{S}^{(0)} &= 0.37, \quad \varepsilon_{S}^{(1)} = 0.49, \quad \varepsilon_{S}^{(2)} = 0.19 \\
\varepsilon_{B}^{(0)} &= 0.03, \quad \varepsilon_{B}^{(1)} = 0.06, \quad \varepsilon_{B}^{(2)} = 0.04
\end{align*}
\]

and the leptonic fractions shown in Table II. Our branching ratio for the process \( H \to c\bar{c} \) is then \( Br(H \to c\bar{c}) = 6.1\% \) leading to the cross section \( \sigma_{pp \to H(\to c\bar{c})Z} = 0.08 \text{ fb} \). In the case of background the corresponding cross section is \( \sigma_{pp \to Z + \text{jets}} = 23.56 \text{ fb} \). Based on these results and considering the integrated luminosity \( \int L dt = 3000 \text{ fb}^{-1} \) we can verify the 2 sigma condition for the significance \( S/\sqrt{B} = 2.0 \).

![FIG. 1: Double c-jets selection efficiency (from \( H \to c\bar{c} \)) against the QCD-jets rejection achieved by the Event-Shapes tagger. The curves are obtained from independent optimizations considering different subsamples with zero (non leptonic), one (single leptonic) and two (double leptonic) leptons inside the highest transverse momentum fat jet.](image-url)
B. Comparison against the ATLAS JetFitterCharm algorithm

In order to evaluate the performance of our double charm identification approach against more conventional life-time based single charm tagging procedures, we provide a "naive" comparison with the ATLAS JetFitterCharm algorithm [29]. In the ATLAS study two main sources of backgrounds are considered, the first one are light-flavor jets, i.e. jets arising from the hadronization of $g, u, d, s, \bar{u}, \bar{d}, \bar{s}$; the second background is heavy-flavor jets, in this context $b$-jets.

From [29] we extract the JetFitterCharm single selection efficiencies $\epsilon_c, \epsilon_b$ and $\epsilon_{light}$ for the charm-jets, $b$-jets and light-jets respectively. The double tagging coefficients for each category are calculated as $\tilde{\epsilon}_2 = \epsilon_c^2$, $\tilde{\epsilon}_b = \epsilon_b^2$ and $\tilde{\epsilon}_{light} = \epsilon_{light}^2$.

For the comparison of the different tagging strategies, we used the boosted Higgs search described in Section II where the dominant backgrounds are light-flavor-jets $+Z$ and $b\bar{b}$ jets $+Z$. As the JetFitterCharm efficiencies are not provided in terms of separate analyses for the different leptonic categories introduced in Sec II we combine the selection efficiencies achieved in our approach for the non-leptonic, single-leptonic and double-leptonic studies for a given background according to Eq. (1).

We find the best results in rejecting light-flavor jets, which have in this analysis a cross section at least an order of magnitude bigger than the $bb$ background. In Fig. 2 we show the ROC curves for the different leptonic analyses and in Fig. 3 we present the performance obtained from the combination of the leptonic categories. Without access to the ATLAS detector simulation a direct comparison between the two approaches is not feasible. However, it can be inferred from Fig. 3 that for $0.16 < \tilde{\epsilon}_2$, the jet-shapes strategy shows a strong performance and is likely to add to the tagging strategy employed by ATLAS. Consequently, using event shapes it is possible to outperform the double application of a charm-tagger by a single application of a double-charm tagger. This is achieved by looking at the full radiation profile inside a fat jet; without disentangling the radiation signatures of the $c$-quark and the $\bar{c}$-quark independently.

IV. CP ODD THDM SCALAR

The coupling between the CP odd THDM scalar $A$ and the pair $c\bar{c}$ is directly proportional to the charm quark mass $m_c$ and inversely proportional to the THDM vacuum ratio $\tan \beta$. As shown in [32], the decay channel $A \rightarrow c\bar{c}$ is expected to be dominant for $4.0 \text{ GeV} \lesssim m_A \lesssim 10.0 \text{ GeV}$ and low values of $\tan \beta$. Here we determine a 95% C.L. upper bound for the branching ratio $Br(H \rightarrow A\rightarrow c\bar{c})Z$ in this mass range. Our signal is the process $pp \rightarrow H(\rightarrow c\bar{c})Z + \text{jets}$ and our background is given by $pp \rightarrow Z + \text{jets}$. For $m_A = 4.0 \text{ GeV}$ the combination of observables that give the best performance are presented in Table IV, from here the ROC’s corresponding to the different leptonic categories are determined, see Fig. 4. The optimal selection efficiency point is

$$\epsilon_{c\bar{c},m_A=4 \text{ GeV}} = 0.81 \quad \epsilon_{QCD,jets} = 0.01 \quad (4)$$

resulting from the efficiencies

$$\epsilon_S^{(0)} = 0.83, \quad \epsilon_S^{(1)} = 0.69, \quad \epsilon_S^{(2)} = 0.39$$
$$\epsilon_B^{(0)} = 0.01, \quad \epsilon_B^{(1)} = 0.01, \quad \epsilon_B^{(2)} = 0.01 \quad (5)$$

combined with the leptonic fractions presented in Table V as given in Eq. (1). Thus, we get the following 95% C.L. upper limit for the branching ratio $Br(H \rightarrow A\rightarrow c\bar{c})Z < 0.01\%$, leading to the cross section for the signal process $\sigma_{pp \rightarrow H(\rightarrow c\bar{c})Z + \text{jets}} = 0.02 \text{ fb}$. For comparison, using track-based substructure observables and considering $m_A = 4.0 \text{ GeV}$, the 95% C.L. bound
TABLE III: Top observables determined by the Multivariate Analysis to discriminate the process $pp \rightarrow H(-\rightarrow c\bar{c})Z$ against $pp \rightarrow Z +$ light jets.

\[ Br(H \rightarrow A(-\rightarrow c\bar{c})Z) \leq 2.1\% \] has been previously determined in [19].

For $m_A = 10$ GeV the observables per leptonic category that yield the best selection efficiency curves, shown in Fig. [5] are presented in Table (VII). Our optimal result corresponds to

\[ \varepsilon_{c\bar{c}, m_A=10 \text{ GeV}} = 0.38 \quad \varepsilon_{QCD,jets} = 0.0004 \] (6)

calculated from the individual efficiencies per leptonic category

\[ \varepsilon_{S.}^{(0)} = 0.39, \quad \varepsilon_{S.}^{(1)} = 0.29, \quad \varepsilon_{S.}^{(2)} = 0.19 \]
\[ \varepsilon_{B.}^{(0)} = 0.9 \times 10^{-4}, \quad \varepsilon_{B.}^{(1)} = 14.9 \times 10^{-4}, \quad \varepsilon_{B.}^{(2)} = 177.0 \times 10^{-4} \] (7)

and the partial fractions of leptonic events presented in Table (VII). The 95% C.L. limit on the branching ratio is $Br(H \rightarrow A(-\rightarrow c\bar{c})Z) \leq 0.003\%$, leading to the signal cross section $\sigma_{pp \rightarrow H(-\rightarrow A(-\rightarrow c\bar{c})Z)+jets} = 0.01$ fb. The correlation matrices for the analyses of this section and the histograms for the main discriminating observables are shown in Figs. [8] and Figs. [12] respectively.

V. CONCLUSIONS

We have studied the efficiency of event shapes for tagging jets resulting from $c\bar{c}$ pairs originated in the

FIG. 4: Double c-jets selection efficiency (from $A \rightarrow c\bar{c}$) against the QCD-jets rejection achieved by the Event-Shapes tagger per leptonic study. Here we are considering $m_A = 4$ GeV.

TABLE V: Fraction of events in each one of the leptonic categories for the samples $pp \rightarrow H(-\rightarrow A(-\rightarrow c\bar{c})Z) +$ jets and $pp \rightarrow Z +$ jets after the selection cuts for $m_A = 10$ GeV.
decay $H \to c\bar{c}$. The results obtained can be considered as an optimal limit for the performance of our selection strategy as we have not included detector effects. We have optimized our analysis depending on the main backgrounds in the selected processes and have taken into account the following possibilities $pp \to q\bar{q}Z$ for $q = \{u, d, s, c\}$ and $pp \to ggZ$. Our signal channel is $pp \to H(\to c\bar{c})Z$ and we select highly boosted Higgs bosons. Using jet shape observables as input to a BDT, we find a good performance to separate the $c\bar{c}$ signal from $b\bar{b}$ and light-flavor fat jets.

Thus, with this approach we can project an upper limit on $Br(H \to c\bar{c}) \lesssim 6.1\%$ with SM production rates for $\int \mathcal{L} dt = 3000.0 \text{ fb}^{-1}$ and $\sqrt{s} = 13.0$ TeV.

Following an analogous strategy we have studied the CP-odd THDM scalar $A$ decaying into pairs $c\bar{c}$. In particular we have determined $Br(H \to A(\to c\bar{c})Z) \leq 0.01\%$ by considering masses for $A$ inside the range $4.0$ GeV $\lesssim m_A \lesssim 10.0$ GeV where the channel $A \to c\bar{c}$ is particularly dominant for low values of $\tan \beta$.

### Acknowledgments

MS is supported in part by the European Commission through the 'HiggsTools' Initial Training Network PITN-GA-2012-316704. We thank Silvan Kuttimalai and Nathan Hartland for useful discussions on SHERPA and multivariate analyses respectively.

### Appendix: Event Shapes

This section summarizes most of the observables considered during our analysis. For a more extensive discussion see [33, 44, 45] and the references cited therein.

We begin by introducing the definition of Thrust [46, 47]

$$T = 1 - \max \left( \sum_i |\vec{p}_i \cdot \hat{n}_T| \right)$$

(A.1)

where $\hat{n}_T$ is the direction that maximizes the numerator. To avoid confusion, in the subsequent discussion the symbol “$\perp$” will be used to denote the transverse contribution of different kinematical variables. Then, the Thrust major is determined according to [48]

$$T_M = \max_{\hat{n} \perp \hat{n}_T = 0} \left( \sum_i |\vec{p}_i \cdot \hat{n}| \right)$$

(A.2)
FIG. 6: Correlation coefficients for the observables used in the analysis $pp \rightarrow H(\rightarrow c\bar{c})Z$ (signal) vs $pp \rightarrow Z + jets$ (background).
FIG. 7: Correlation coefficients for the observables used in the analysis $pp \rightarrow H(\rightarrow c\bar{c})Z$ (signal) vs $pp \rightarrow Z + \text{light jets}$ (background).
FIG. 8: Correlation coefficients for the observables used in the analysis of the CP odd THDM scalar $A$, for $m_A = 4$ GeV.
FIG. 9: Correlation coefficients for the observables used in the analysis of the CP odd THDM scalar $A$, for $m_A = 10$ GeV
where it should be understood that $\vec{n}$ is perpendicular to $\vec{n}_T$.

We use the Thrust of $e^{-\eta}$ momenta calculated according to Eq. (A.1) but with the three-momenta of each one of the subjets in the event modified according to

$$\vec{p}_i \to \vec{p}_i e^{-|\eta_i|}.$$  \hspace{1cm} (A.3)

We include the Fox Wolfram moment inspired observable

$$H_n = \sum_{i,j} \frac{\vec{p}_i \cdot \vec{p}_j}{E_{Ttot}^2} \sin^{2n} \theta_{ij},$$  \hspace{1cm} (A.4)

with $E_{Ttot}$ being the total energy of the jet constituents, thus $E_{Ttot} = \sum_i E_i$. The sum in the numerator of Eq. (A.4) considers only pairs of particles within the same hemisphere, i.e. those particles satisfying $\vec{p}_i \cdot \vec{p}_j > 0$, and $n$ is a rational number. In what follows we will refer to $H_n$ as the Fox Wolfram-like $n$ moment, and we will consider the value $n = 1/4$.

The Transverse spherocity is given by

$$S_{ph} = \min \left( \sqrt{\frac{2}{\pi}} \left( \frac{\sum_i |\vec{p}_{\perp i} \times \hat{n}''|}{\sum_i |\vec{p}_{\perp i}|} \right)^2 \right),$$  \hspace{1cm} (A.5)

where $\hat{n}''$ is the direction that minimizes the sum in the numerator.

FIG. 10: Histograms for the main discrimination event shapes when selecting $pp \to H(\to c\bar{c})Z$ against $pp \to Z + jets$. 

Non Leptonic Analysis

Single Leptonic Analysis

Double Leptonic Analysis
The 3-jet resolution $y_{ij}$ defines the lower bound for the jet recombination parameter $y_{ij}$ in order to have a 3-jet event. Before presenting the determination algorithm for $y_{ij}$, according to different schemes, let us first introduce the possible definitions for the parameter $y_{ij}$

$$y_{ij} = \begin{cases} 
2 \min(E^2_{ij}, E^2_j)(1 - \cos \theta_{ij})/E^2_{vis} & \text{Durham} \\
\frac{\sum_i E_i (1 - \cos \theta_{ij})}{\sqrt{E_i + E_j}} & \text{Geneva} \\
(p_i + p_j)^2/E^2_{vis} & \text{Jade}.
\end{cases}$$

with $E_{vis}$ is the sum of the energies for the different final state subjets before the recombinations.

In addition, the recombination schemes between the $i$-th and $j$-th subjets are

$$\begin{align*}
P: & \tilde{p} = \tilde{p}_i + \tilde{p}_j, E_p = |\tilde{p}| \\
\text{Schemes:} & \begin{cases} 
P: & \tilde{p} = \tilde{p}_i + \tilde{p}_j, E_p = |\tilde{p}| \\
E0: & \tilde{p} = \frac{E_i + E_j}{|\tilde{p}_i + \tilde{p}_j|} (\tilde{p}_i + \tilde{p}_j), E_p = E_{p,i} + E_{p,j}.
\end{cases}
\end{align*}$$

Then for example, in order to calculate the Resolution $y_{ij}$ Durham (P-scheme) [53 [51], we start by assigning an arbitrary high value to $y_{ij}$. Next, we calculate the parameter $y_{ij}$ between all the subjets inside a given fat jet using the Durham recombination rule shown before. We then determine the pair of elements whose $y_{ij}$ is minimum, $y_{i,j}^{\text{vis}}$, and recombine them applying
the P-scheme presented above. Finally, if $y_3 < y_{ij}^{\text{min}}$, we do the substitution $y_3 = y_{ij}^{\text{min}}$ and repeat the entire process, starting with the re-calculation of the values $y_{ij}$ over the set of subjets determined in the last iteration. The algorithm stops when the total number of subjets left after all the recombinations is equal to 3. The value of $y_3$ obtained in the final iteration is the number we are aiming for. The determination of the Resolution $y_3$ Jade (E-scheme) and the Resolution $y_3$ Jade (E0-scheme) proceed in an analogous way; however the Durham parameter $y_{ij}$ should be substituted by the Jade distance parameter; and the P recombination scheme should be replaced by the E-scheme (E0-scheme).

The Directly global $y_3$ [34] is constructed using the $k_t$ jet algorithm. To begin with, for all $n$ final state particles we define the beam-distance measure

$$d_{k,B} = p_{\perp k}^2,$$

(A.6)

and for constituent pairs we calculate

$$d_{kl} = \min\{p_{\perp k}^2, p_{\perp l}^2\} \frac{(y_k - y_l)^2 + (\phi_k - \phi_l)^2}{R^2},$$

(A.7)

in terms of their corresponding pseudo rapidity $y$ and azimuthal angle $\phi$.

In our analysis we use $R = 0.7$. Let $d^{(n)} = \min\{d_{k,B}, d_{kl}\}$, where the entire set of distances calculated at a given stage is considered. If $d^{(n)}$ is one of the values $d_{ij}$, then the pseudojets $i$ and $j$ are recombined using the E-scheme defined above. If $d^{(n)}$ is one of the
individual coefficients $d_{kR}$ then the pseudojet is removed and included in the beam. These steps are repeated until only 3-pseudo jets are left. At this stage we determine

$$y_{23} = \frac{1}{P_{\perp}^2} \max_{n \geq 3} \{d^{(n)}\},$$  \hspace{1cm} (A.8)

with

$$P_{\perp} = p_{\perp,1} + p_{\perp,2},$$  \hspace{1cm} (A.9)

being $p_{\perp,1}$ and $p_{\perp,2}$ the transverse momenta of the jets obtained by continuing reclustering the event up to 2-pseudojets.

The observable $\tau_x$ can be modified to give the Fractional energy correlation [15]

$$FC_x = \sum_{i \neq j} E_i E_j |\sin \theta_{ij}| \left(1 - |\cos \theta_{ij}|\right)^{1-x} \times$$

$$\Theta(\vec{p}_i \cdot \vec{n}_T)(\vec{p}_j \cdot \vec{n}_T).$$  \hspace{1cm} (A.10)

Here $x$ is a continuous parameter. During the analysis we use $x = 1.5$ that makes the observable particularly sensitive to collinear emissions for fixed transverse momentum.

To define the Transverse sphericity let us first introduce the transverse momentum tensor
Then the transverse sphericity can be determined in terms of the eigenvalues $\lambda_1$ and $\lambda_2$ of $M_{xy}$ (for $\lambda_1 \geq \lambda_2$) as

$$S^{\text{spheri}}_{\perp,g} = \frac{2\lambda_2}{\lambda_1 + \lambda_2}, \quad (A.12)$$

for circular events in the transverse plane we have $S^{\text{spheri}}_{\perp,g} \to 1$, whereas for pencil like events $S^{\text{spheri}}_{\perp,g} \to 0$.

To describe the cone jet mass let us start by introducing some definitions. The components of the highest $p_T$ fat jet selected in our studies are first reclustered using the $k_t$ algorithm. Then, the region $C$ results from the union of the cones around the two new highest transverse momentum subjets (with coordinates $\eta_{j,j},\phi_{j,j}$, for $j = 1, 2$) according to

$$\sqrt{(\eta_i - \eta_{j,j})^2 + (\phi_i - \phi_{j,j})^2} \leq R. \quad (A.13)$$

Where the subindex $i$ runs over the rest of the newly generated subjets. During our implementation we considered $R = 1$. The central transverse thrust axis $\vec{n}_{T,C}$ is then defined as the vector that maximizes

$$\sum_{i \in C} |\vec{p}_{\mid i} \cdot \vec{n}_{T,C}| \quad (A.14)$$

where

$$Q_{\perp,C} = \sum_{i \in C} |\vec{p}_{\mid i}|. \quad (A.15)$$

The vector $\vec{n}_{T,C}$ allow us to divide the region $C$ into the subregions $C_U$ and $C_D$, defined in terms of the conditions $0 < \vec{p}_{\mid i} \cdot \vec{n}_{T,C}$ and $\vec{p}_{\mid i} \cdot \vec{n}_{T,C} < 0$ respectively. The partial masses in each one of these regions are

$$M_{xy} = \sum_{i} \left( \begin{array}{ccc} p_{x,i}^2 & p_{x,i}p_{y,i} & p_{x,i}p_{y,i} \end{array} \right). \quad (A.11)$$

Then, the Cone total jet mass is

$$\rho_{U,C} = \left( \frac{\sum_{i \in C_U} p_i}{Q_{\perp,C}} \right)^2, \quad \rho_{D,C} = \left( \frac{\sum_{i \in C_D} p_i}{Q_{\perp,C}} \right)^2. \quad (A.16)$$

Finally, we consider the following $C$ parameter-like observable [53-55]

$$C = 3 - \frac{3}{E_{Ttot}^2} \sum_{i<j} \frac{(p_i \cdot p_j)^2}{E_i E_j}. \quad (A.22)$$

with $E_{Ttot} = \sum_i E_i$. 

---


