ELSEVIER

The importance of jet shapes for tagging purposes.

Sergio Calvente, Javier Llorente, Fernando Barreiro

Universidad Autónoma de Madrid

Abstract

Recent measurements by the ATLAS Collaboration of b- and light jet-shapes in $t\bar{t}$ final states at the LHC, show clear evidence that b-jets are wider. There are also indications that b-jets in $t\bar{t}$ final states are narrower than those produced inclusively. We apply multivariate analysis techniques on MC generated samples to illustrate possible applications of these observations in two study cases: 1) measurement of $|V_{ts}|$ and 2) separation of $pp \rightarrow ZH \rightarrow l^+l^-b\bar{b}$ from the continuum background i.e. $pp \rightarrow Z + b\bar{b} \rightarrow l^+l^- + b\bar{b}$. The lesson to be learned is that jet shape measurements show more potential in the former case than in the latter.

Keywords: jet shapes , $|V_{ts}|$, Higgs associated production

1. Introduction

The $t\bar{t}$ final states in the dileptonic mode at the LHC are a copious and clean source of b jets, while the single lepton mode offers the possibility to study light jets through $W \rightarrow u\bar{d}, c\bar{s}$ decays. Integrated jet shapes, in a cone of radius $r \leq R = 0.4$ around the jet axis, defined as: $\Psi(r) = p_T(0, r)/p_T(0, R)$ where $p_T(0, r)$ is the scalar sum of the transverse momenta within that cone, have been recently measured by the ATLAS Collaboration [1]. These measurements show clear evidence that b-jets are wider than light jets, see Fig.1 (left). This can be understood as due to the b-quark being heavier than the light quarks and decaying sequentially via $b \rightarrow c \rightarrow s$.

Furthermore, by comparing the energy flow outside half of the jet cone for these b-jets in ATLAS $t\bar{t}$ final states with those inclusively measured by CDF [2] the former seem narrower than the latter, as shown in Fig.1 (right).

The purpose of this analysis is to investigate possible applications of jet shape measurements in two study cases: 1) the feasibility to measure $|V_{ts}|$ in $t\bar{t}$ final states, and 2) the possibility to discriminate $pp \rightarrow ZH(b\bar{b})$, from the main QCD background $pp \rightarrow Z + b\bar{b}$
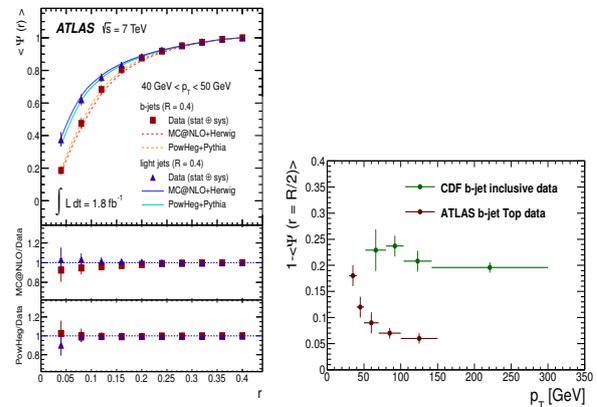

Figure 1: Average integrated b- and light-jet shapes measured by AT-LAS, left, and energy flow outside half of the jet cone, ATLAS and CDF comparison, right

2. Prospects of measuring the CKM element $|V_{ts}|$

Measurements of single top production cross sections at the Tevatron and the LHC have provided first direct measurement of the dominant CKM matrix element $|V_{tb}|$. A very recent determination by the CMS Collaboration reads $|V_{tb}| = 0.998 \pm 0.038(exp) \pm 0.015(syst)$ yielding $|V_{tb}| \geq 0.92$ at $95\%C.L.$ [7]. This is to be

compared with values obtained from overall fits of the CKM unitarity yielding $|V_{tb}| = 0.999133(44)$ and $|V_{ts}| = 0.0407 \pm 0.001$ so that $|V_{ts}|^2/|V_{tb}|^2 = 1.7 \times 10^{-3}$.

In Ref.[3] the prospects of measuring the CKM matrix element $|V_{ts}|$ have been discussed. The goal is to discern between $pp \to t\bar{t} \to WbWs$, which will be our signal process, from $pp \to t\bar{t} \to WbWb$, which will be the dominant background. They used the PYTHIA.6 Monte Carlo [4] to generate 10^6 events for the process $pp \to t\bar{t}X$ at $\sqrt{s} = 14\ TeV$, followed by the decay chains $t \to W^+b, W^+s$ and $\bar{t} \to W^-\bar{b}, W^-\bar{s}$. The $W^\pm$ are forced to decay leptonically to reduce the jet activity associated to hadronic W decays. This corresponds to an integrated luminosity of $10\ fb^{-1}$. The top branching ratios to Wb and Ws were artificially set equal to 50%. With this generated sample they obtained the following distributions: 1) the energy and transverse momentum distributions of neutral strange particles, K_s^0 and Λ^0's, which are expected to be leading (non-leading) in s- (b-) initiated jets, 2) the presence (absence) of secondary vertices in b- (s-) initiated jets, and 3) the spectrum of soft leptons. They were fed into a BDT classifier using TMVA [6]. For a rejection of the background of 0.1%, the tagging efficiency of the signal was found to lie in the range $5 - 10\%$. This means that we expect $0.05 \times 2 \times 1.7 \times 10^{-3} \times 10^6 = 170$ signal events, with a background of $10^{-3} \times 10^6 = 10^3$ events, giving a significance of $170/\sqrt{1000}$ i.e. more than 5σ.
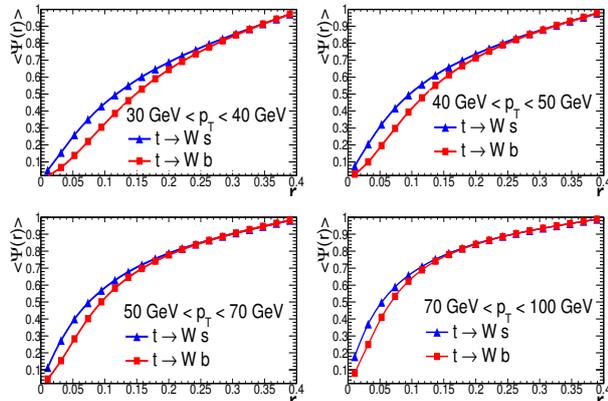
Figure 2: Average integrated b- (red squares) and s-jet (blue triangles) shapes with PYTHIA.

We have investigated the benefit of including b- and s-jet shapes in our MVA. Fig.2 shows the average integrated shapes in several p_T bins. The differences between b- and s-jets are more significant at low values of r and diminish with increasing p_T. This is supported by recent measurements by the ATLAS Collaboration [1].

We carry out a two step multivariate analysis based on a BDT classifier. To begin with, we feed in the BDT classifier the jet shapes data alone. The corresponding ROC curve is shown in Fig. 3 (green). Notice that one can achieve a non-negligible 90% rejection of the background process, with an efficiency of the signal, of roughly 40%. In a second step, we compare the ROC curves obtained using the variables in [3] with and without adding the jet shape information. Their inclusion improves the discrimination power of the BDT algorithm by a factor of two for high background rejection, Fig.3.
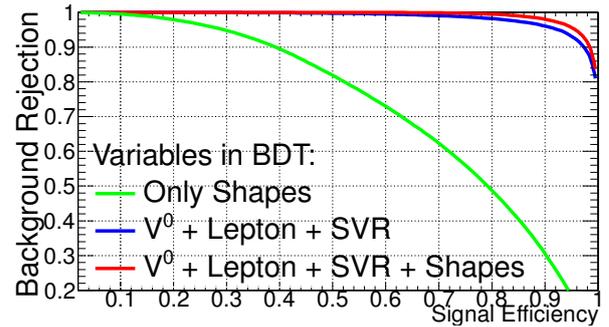
Figure 3: ROC curves obtained from shapes data only for $pp \to WbWs$, signal and $pp \to t\bar{t} \to WbWb$, background, left, and in addition with all the distributions described in the text [3], right

3. Associated Higgs production $pp \to Z(l^+l^-)H(b\bar{b})$

The observation of inclusive $H \to b\bar{b}$ is precluded by the overwhelming QCD background. A big effort is currently devoted to associated production with a gauge boson, Z or W. In the case of $pp \to ZH(b\bar{b})$, from now on our signal process, the main background is due to $pp \to Z + b\bar{b}$.

As b-jets in the signal (background) process are coming from a colour singlet (triplet or octet), one could also expect b-jet shapes in the former to be somewhat narrower than those in the latter. Therefore, our aim is to test the b-jet shapes power to improve on the separation between these signal and background.

We begin by looking at relevant *kinematic* quantities associated to the b-jets or leptons in these two processes, like transverse momenta of the leading (subleading) b-jet i.e. p_{T1} (resp.p_{T2}) as well as the $b\bar{b}$ pair and Z boson; their separation in rapidity, azimuth or (η, ϕ) plane i.e. $\Delta\eta, \Delta\phi, \Delta R$, and between the Z boson and the $b\bar{b}$, see Fig. 4. We also considered the polar angle of the leading b-jet in the rest frame of the $b\bar{b}$ pair, which is expected to be flat in the case of Higgs boson decays.

These distributions have been obtained with the MADGRAPH Monte Carlo generator [5] coupled to PYTHIA [4]. The Z boson has been forced to decay to $\mu^+\mu^-$ pairs and the b-jets, defined with the anti-k_T algorithm with radius R=0.4, are required to have $p_T \geq 25\ GeV$ to allow an efficient b-tagging.
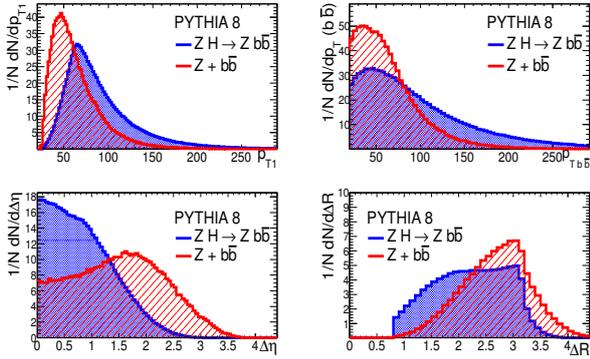
Figure 4: Relevant kinematic variables used to discriminate $pp \rightarrow ZH(b\bar{b})$ from $pp \rightarrow Z + b\bar{b}$.

The main observation in Fig. 4 is that b-jets coming from the signal process are harder, in transverse spectra, and better separated in (η, ϕ), than those coming from the background process. These differences were exploited by the authors of Ref. [8] who performed a multivariate analysis for this problem based on a BDT classifier. In fact, they also studied the relevance of two new variables which they called pull angles for the leading and subleading b-jets. The pull vector, defined as $\vec{t} = \sum_{i \in Jet}(p_T^i |r_i| / p_T^{jet}) \vec{r}_i$ with $\vec{r}_i = (\Delta y_i, \Delta \phi_i)$ being the distance in the (rapidity,azimuth) plane between a given cell in a jet and the jet axis. This variable can also be written in polar coordinates in that plane in terms of a pull angle θ_t. The discrimination power added by both the pull angle and event shape variables such as sphericity, see Fig. 5 is not that strong as already noted in [8].
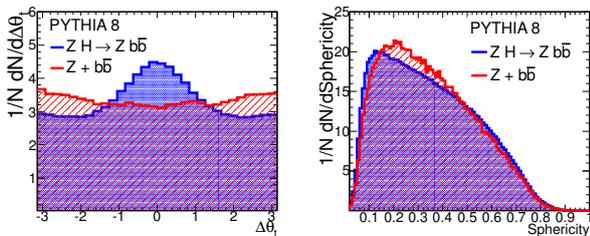

Figure 5: Pull angle for the leading b-jets (left), and sphericity of the two lepton and two b-jets (right), also used to discriminate $pp \rightarrow ZH(b\bar{b})$ from $pp \rightarrow Z + b\bar{b}$.

Next, we included in the BDT, distributions on differential jet shapes, Fig. 6. These are defined as $\rho(r) = p_T(r - \Delta r, r) / p_T(0, R) / \Delta r$. The value $\Delta r = 0.02$ has been used in Fig.6 and only differential shapes up to $r = 0.12$ have been fed into a TMVA classifier.

The MVA proceeds in three steps. Firstly we feed only the kinematic variables discussed above, in a second step we add pull information, and finally we also include the b-jet shapes. The resulting ROC curves for signal and background are displayed in Fig. 7. All three ROC curves are very similar with marginal improve-

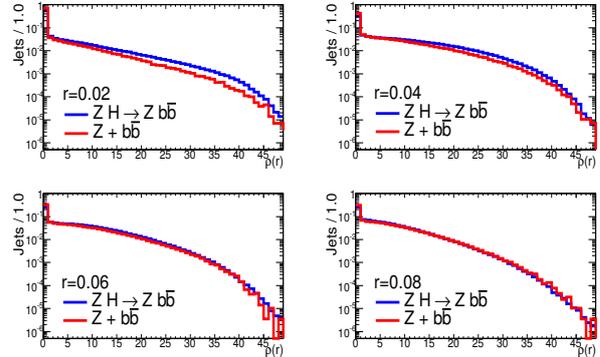ment after inclusion of pull and shapes information.

Figure 6: Differential b-jet shapes in the signal, $pp \rightarrow ZH(b\bar{b})$, and background, $pp \rightarrow Z + b\bar{b}$, processes.
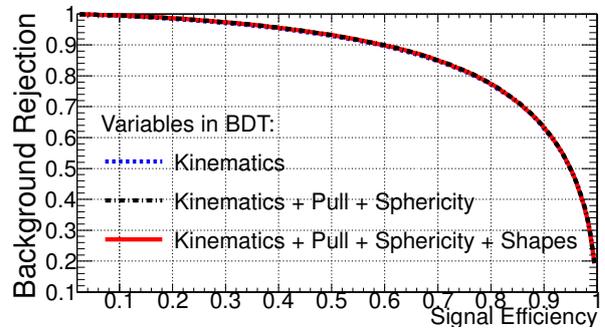

Figure 7: ROC curves for the separation between $pp \rightarrow ZH(b\bar{b})$ and $pp \rightarrow Z + b\bar{b}$.

4. Summary and Conclusion

We have shown that jet shapes data have a potential to discriminate b- from light-jets. This can be of help for improving b-tagging algorithms which are mainly based on secondary vertices and charged particle impact parameter measurements. In the case of the prospects for measuring $|V_{ts}|$, the inclusion of jet shape variables in a MVA leads to significant improvements of the signal efficiency for high background rejection figures. This is not the case for the separation of $pp \rightarrow Z(l^+l^-)H(b\bar{b})$ from $pp \rightarrow Z(l^+l^-) + b\bar{b}$.

References

[1] ATLAS Coll., EPJ C73 (2013) 2676
[2] CDF Coll., Phys. Rev. D 78 (2008) 072005 (2010) 022001
[3] A. Ali, F. Barreiro and Th. Lagouri, Phys. Lett. B693 (2010) 44-51
[4] T. Sjostrand, S. Mrenna and P.Z. Skands, Comp. Phys. Comm. 178 (2008) 852
[5] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni et al., JHEP 1407 (2014) 079
[6] A. Hocker et al., PoS A CAT (2007) 040, arXiv:physics/0703039
[7] CMS Coll., arXiv:1403.7366.
[8] J. Gallicchio et al., JHEP 1104 (2011) 069