EPJ Web of Conferences **49**, 17007 (2013)
DOI: 10.1051/epjconf/20134917007

Top Polarization from Boosted Jet Substructure

Biplob Bhattacherjee[1,], Sourav K. Mandal[1,], and Mihoko Nojiri[1,2,3,]

[1] *Kavli IPMU (WPI), University of Tokyo, Kashiwa, Chiba 277-8583, Japan*
[2] *Theory Group, KEK, Tsukuba, Ibaraki 305-0801, Japan*
[3] *The Graduate University for Advanced Studies (SOKENDAI) Tsukuba, Ibaraki 305-0801, Japan*

Abstract. Top polarization is an important probe of new physics that couples to the top sector, and which may be discovered at the 14 TeV LHC. Taking the example of the MSSM, we develop a *detector level* analysis methodology for extracting polarization information from hadronic tops using boosted jet substructure. We show that with 100 fb^{-1} of data, left and right 600 GeV stops can be distinguished to 4.5σ, and 800 GeV stops can be distinguished to 3σ.

1 Introduction

1.1 Motivation

Top physics is an important probe of theories of new physics at the TeV scale, as many of these theories posit TeV-scale partners to the top quark in order to resolve the Higgs hierarchy problem.

In the case of supersymmetry with R-parity, the composition of the stop $\tilde{t}$ in terms of the weak eigenstates $\tilde{t}_R$ and $\tilde{t}_L$ can be constrained by observing the polarization of tops in the decay $\tilde{t} \to t\tilde{\chi}_1^0$. In the MSSM, the stop mixing matrix is

$$\begin{pmatrix} \tilde{t}_1 \\ \tilde{t}_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \tilde{t}_L \\ \tilde{t}_R \end{pmatrix} \tag{1}$$

where

$$\tan 2\theta = \frac{2m_t(A_t - \mu/\tan\beta)}{m_{Q_3}^2 - m_{\tilde{u}_3}^2 + \Delta_{\tilde{u}_L} - \Delta_{\tilde{u}_R}} \quad . \tag{2}$$

The fermionic top partners in extra-dimensions theories with KK parity and little Higgs theories with T-parity have the analogous decays $t^{(1)} \to tB^{(1)}$ and $T' \to tB_H$. There are also numerous theories with and without top partners containing extra massive gauge bosons Z' with decays such as $Z' \to t\bar{t}$. Top polarization measurements can constrain the chiral structure of their couplings to the top quark.

In this proceeding (based on Ref. [1]) we study top polarization at *detector level* including all contamination sources (e.g., ISR/FSR and MPI), relevant detector effects (e.g., magnetic field) and backgrounds. We focus on pair production of 600 GeV and 800 GeV $\tilde{t}_1$ at the 14 TeV LHC under the simplified model in which they decay entirely to $t\chi_1^0$, where $\chi_1^0 \simeq \tilde{B}$ and $m_{\chi_1^0} = 100$ GeV. We choose to focus on supersymmetry also because it may be the most well-motivated of this class of theories, solving the hierarchy problem up to Planck scale as well as enhancing gauge coupling unification at high scale. We employ the techniques of boosted jet substructure to access the jet clustering information, allowing us to reduce the effects of contamination while also avoiding the combinatoric problem posed by having numerous monolithic jets.

1.2 Top polarization

The decay products of the top quark have the angular distributions

$$\frac{1}{\Gamma}\frac{d\Gamma}{d(\cos\theta_{tf})} \propto 1 + \mathcal{P}_t k_f \cos\theta_{tf} \tag{3}$$

where $\cos\theta_{tf}$ is the angle between the daughter momentum and the top spin axis in the top rest frame; we can take the latter as the direction of top momentum in the lab frame. $\mathcal{P}_t = \pm 1$ is the polarization of the top quark, and k_f is the "spin analyzing power" of the daughter flavor. For the b-quark $k_b = -0.4$ whereas for the lepton daughter $k_l = 1$. Consequently we can measure $\mathcal{P}_t$ by observing the distribution of $\cos\theta_{tf}$. For the case of hadronic tops, this can be done directly by reconstructing the top and resolving the b-quark daughter. However, for leptonic tops, one cannot fully reconstruct the top momentum due to the neutrino from the W decay. It has been proposed to define $\cos\theta_{tl}$ in an "approximate rest frame" in semileptonic events with reconstruction of the accompanying hadronic top [2], or require the leptonic top in a semileptonic event to be highly boosted such that one can use alternative measurables which are insensitive to top momentum in the limit $\beta \to 1$ [3, 4].

In our analysis we measure the polarization of hadronic tops, which has not only the advantage of being more simple than leptonic methods, but also that of greater statistics. Moreover, leptonic top analysis entails

all the difficulties of identifying isolated leptons in a real hadron collider environment.

2 Simulation and analysis

2.1 Event generation and detector simulation

Using `Herwig++ 2.5.0` with all physics effects (hadronization, ISR/FSR, MPI) included, we generated left ($\sin\theta = 0$), mixed ($\tan\theta = 0.25$) and right ($\sin\theta = 1$) $\tilde{t}_1\tilde{t}_1^*$ samples with masses 600 GeV and 800 GeV for the 14 TeV LHC under our simplified model. We calculated NLO production cross sections using `Prospino2.1`.

We consider the backgrounds $t\bar{t}$+jets, Z+jets, W+jets and $t\bar{t} + Z$, which we generated using `MadEvent/MadGraph 5.1.3` and `PYTHIA 6.4.25` (also with all physics effects), taking their leading order cross sections. All extra jets are five-flavor ($g + u, d, c, s, b$).

For our detector simulation we used `Delphes 2.0.2` [5]. We modified the `Delphes` codebase to use `FastJet 3.0.3` [6] instead of the bundled version, as the newer version has an interface to manipulate subjets at specific clustering scales or steps. This allows us to "prune" the clustering tree to remove contamination, then store the resulting subjets, all within the analysis pipeline.

The `Delphes` detector settings are tuned to ATLAS, with the hadronic calorimeter grid set to match that in the ATLAS TDR. The magnetic field is turned on in the simulation.

2.2 Cuts

We implement the following event cuts, designed to increase the significance $S/\sqrt{S+B}$ for our characteristic $t + \not{E}_T$ signature:

1. *$\not{E}_T$ >300 GeV.*
2. *Leading fat jet p_T >400 GeV.*
3. *If there are no leptons w/ $p_T > 5$ GeV, require subleading fat jet p_T >100 GeV.* This cut suppresses processes like W/Z+jets in which the second fat jet is from QCD, as these jets are likely to be soft. Since the majority of signal events have leptons due to $W \to l\nu$ and the decay of b-flavored mesons, we require that there are no leptons for this cut. Thus, this cut is most effective against Z+jets.
4. *Lepton is not collimated with $\not{E}_T$.* For every lepton with $p_T > 5$ GeV, require

$$\frac{\cos(\phi_{\not{E}_T} - \phi_l)}{(\not{E}_T + p_{T,l})/(350\ \text{GeV})} < 0.4 \quad . \tag{4}$$

   This selects against high $\not{E}_T$ arising from boosted leptonic W decays in $t\bar{t}$+jets and W+jets, as the opening angle between the lepton and $\vec{\not{E}}_T$ is likely to be much smaller in these processes than from a top partner decay.
5. *Hard subjet is not collimated with $\not{E}_T$.* For every subjet with $p_T > 50$ GeV, require the same as above. This works against hadronic τ from W decays in $t\bar{t}$+jets and W+jets, as well as highly collimated b-subjets from top decays in $t\bar{t}$+jets.
6. *Require at least one top tagged jet*, using the procedure described in the next section.
7. $225 < M_{T2} < 650$ *for 600 GeV stops, and* $325 < M_{T2} < 850$ *for 800 GeV stops.* M_{T2} is calculated for the leading and subleading jet, with $m_\chi = 0$. For the leading jet we used the reconstructed top jet if top tagged, otherwise we used the trimmed jet; similarly for the subleading jet.

2.3 Top jet reconstruction

2.3.1 Jet clustering and grooming

Hadrons are clustered as fat jets using the Cambridge/Aachen algorithm with cone size $R = 1.2$ and subjet cone size $\Delta R = 0.2$. These numbers are chosen such that the distribution in the number of subjets per fat jet peaks at ~ 3 in our signal samples after the following grooming procedures:

1. The jet clustering trees are "pruned" [7] using the mass-drop condition

$$m_{j_{n-1}} < 0.8 \times m_{j_n} \tag{5}$$

   where $m_{j_{n-1}}$ is the invariant mass of the hardest parent jet, and m_{j_n} is the invariant mass of the child jet at clustering step n. We also require the subjet separation condition

$$d_{k_T}(j_{n-1,1}, j_{n-1,2}) > (\Delta R)^2 \cdot m_{j_n}^2 \tag{6}$$

   where d_{k_T} is the k_T distance between the two parents jets. This removes contamination from MPI and ISR, improving top reconstruction quality. We take the fat jet (and its subjets) at the clustering step where both conditions are satisfied.
2. This jet is then "trimmed", [8], removing subjets with $p_T < 10$ GeV. This further reduces contamination.

2.3.2 Reconstruction and tagging

The trimmed fat jets are then fed to our top reconstruction algorithm. The following steps are attempted for each of these jets:

1. Require $p_T > 400$ GeV for the *untrimmed* jet.
2. Require at least three subjets.
3. Require that one b-subjet j_b is reconstructed in the jet.

Table 1. Utilized b-tagging efficiencies.

Kinematic region	Efficiency
$p_T < 30$ GeV	0%
30 GeV $< p_T <$ 60 GeV	60%
60 GeV $< p_T <$ 450 GeV	70%
450 GeV $< p_T <$ 1000 GeV	60%
1000 GeV $< p_T$	0%

4. Find the subjet combination $j_1 j_2 j_b$, of which no pair of subjets are within ΔR of each other, and which gives the closest invariant mass to m_t. Require also that this invariant mass be in the window (150, 200) GeV.

5. If there is no successful tag, require at least four subjets and retry the step above with four-subjet combinations $j_1 j_2 j_3 j_b$.

6. Optionally require one two-subjet combination to have an invariant mass in the loose W mass window (50, 110) GeV. We present our main results *with* this requirement.

We make these choices to enhance top tagging efficiency, presuming that the top parent particle has already been discovered. Nonetheless, top mistagging does not overwhelm the signal as will be apparent in our results.

2.3.3 *b-tagging subjets*

We require b-tagging in our top reconstruction algorithm to reconstruct the observable $\cos\theta_{tb}$ with high fidelity, but also to suppress top mistagging from background processes. Utilizing recent advances in b-tagging by the LHC detector collaborations, we employ the b-tagging efficiencies (shown in Table 1) recently validated at 7 TeV LHC by CMS for their CSVM tagger [9]. We impose the upper limit $p_T = 1000$ GeV to be conservative, though it is not indicated by the CMS study. We choose to use the CMS efficiencies since they are validated up to $p_T = 670$ GeV, though a recent ATLAS study [10] shows similar efficiencies up to 200 GeV.

However, we *do not* implement mistagging, as this depends on various factors that are best implemented by experimenters — mistagging rates vary rapidly with tagging efficiency and so are sensitive to systematic uncertainties that we cannot model. We expect charm jets to contribute to most of the mistags.

We apply these b-tagging efficiencies at parton level. To reconstruct a b-subjet we sum all the subjets within $\Delta R = 0.2$ of the b-parton, as this captures some hard FSR. If more than one b-parton yields a matching subjet inside a given fat jet, that fat jet is rejected and is not top tagged.

3 Results

3.1 Reconstruction quality

Our measurable is $\cos\theta_{tb}$, where θ_{tb} is the angle between the b daughter and the top spin axis in the top rest frame. The usefulness of this measurable depends on the fidelity of the top jet reconstruction. Here we illustrate the effectiveness of our reconstruction method.

First, we show the jet invariant mass distribution at different stages in Figure 1. For signal processes and $t\bar{t}$+jets, few jets are lost in the pruning phase; however, without jet cluster pruning, we find that the $\cos\theta_{tb}$ distribution loses fidelity compared to the parton-level expectation. The subsequent trimming step is essential for removing soft contamination, resulting in a shift of the invariant mass lower towards the correct top mass. Then, requiring at least three subjets and a b-tag narrows the distribution further. Finally, top reconstruction assembles the subjets with correct invariant mass. Modulo b-tagging efficiency, the top tagging efficiency for hadronic tops in our samples which pass the p_T cut is $\sim 60\%$. Conversely, top reconstruction suppresses Z+jets and W+jets by a factor of $O(100)$.

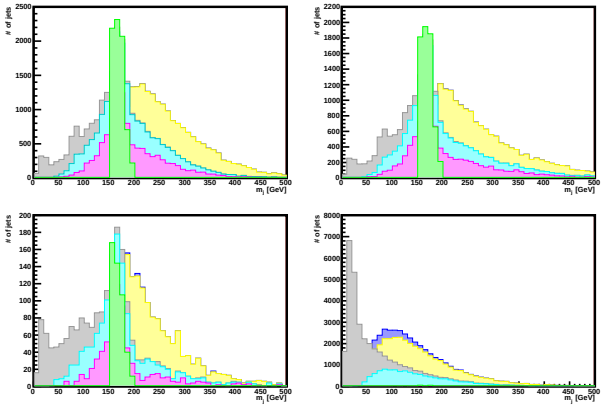

Figure 1. Jet invariant mass sequence for 600 GeV mixed stops (upper left), 800 GeV mixed stops (upper right), $t\bar{t}$+jets (lower left) and Z+jets (lower right). In each panel, from back to front, shown is the jet mass distribution after $p_T > 400$ GeV cut (blue), requiring successful pruning (yellow), trimming (gray), requiring three or more subjets (cyan), b-tagging (magenta), and finally top tagging (green).

The quality of reconstruction is also apparent in the signal-only $\cos\theta_{tb}$ distributions shown in Figure 2. One sees that the parton-level and reconstructed distributions coincide up to statistical fluctuations. Near $\cos\theta_{tb} = -1$ there is depletion due to poor b-tagging efficiency, whereas near $\cos\theta_{tb} = +1$ there is depletion due to the W subjets being too soft to pass the trimming threshold. One sees for this reason that there is less contrast between left and mixed stops with mass 600 GeV than with mass 800 GeV. The distributions at pre-cut/tag parton-level match Ref. [3], except they have a downward left-to-right tilt due to the harder b-partons losing energy to FSR.

Finally, we show sample M_{T2} distributions for 600 GeV and 800 GeV mixed stops in Figure 3 at reconstruction level, signal only. One can clearly see the expected inflection points at $M_{T2} = 600$ GeV and $M_{T2} = 800$ GeV.

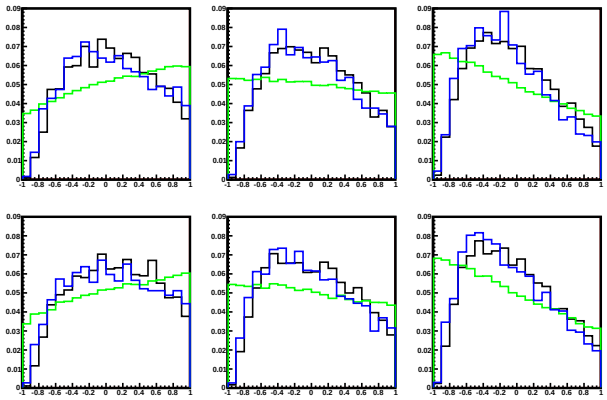

Figure 2. $\cos\theta_{tb}$ distributions for stops only, normalized to one. Left to right are left, mixed, and right stop samples; upper row is for 600 GeV stops, the lower for 800 GeV stops. Shown in each panel are the reconstructed distribution (black), parton-level distribution (blue), and parton-level distribution before cuts and tagging (green).

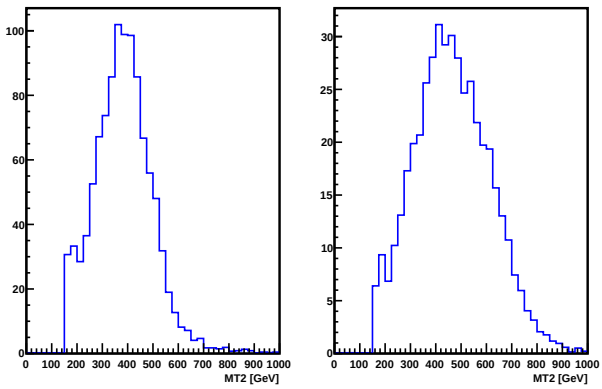

Figure 3. Sample M_{T2} distributions for mixed stops at reconstruction level, signal only. Left panel is for stop mass 600 GeV, and the right panel is for mass 800 GeV.

3.2 $t\bar{t}$ + jets control region

As experimenters are likely to rely on data rather than Monte Carlo for the $t\bar{t}$+jets background, we demonstrate a control region by inverting cuts #4 and #5, *requiring* that there be at least one lepton or hard subjet highly collimated with $\vec{\not{E}}_T$. An array of $\cos\theta_{tb}$ distributions for this region is shown in Figure 4. They evince little contamination from signal and other backgrounds.

It may be of concern that the collimation cuts #4 and #5 might introduce distortions to the $\cos\theta_{tb}$ distributions obtained from top tagging. In Figure 5 we show that the polarization distribution for $t\bar{t}$+jets is invariant under these cuts up to statistical fluctuations. We show also in Figure 6 that the M_{T2} cut does not distort the polarization distribution of $t\bar{t}$+jets in the control region. Thus, it is safe to use a data-driven $t\bar{t}$+jets background from this region.

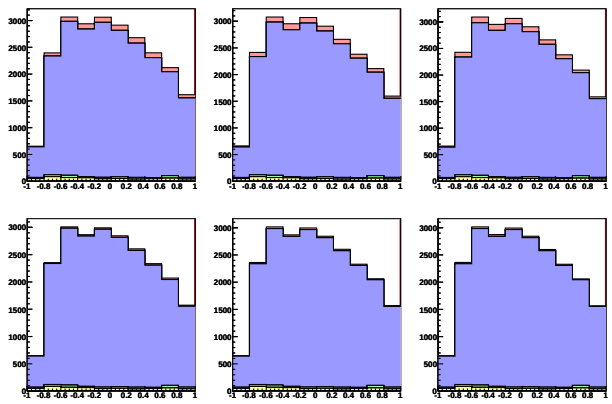

Figure 4. Event distributions of $\cos\theta_{tb}$ for $t\bar{t}$+jets in the control region at 14 TeV LHC @ 100 fb^{-1}. Left-to-right are left, mixed and right stop samples; upper row is for 600 GeV stops, the lower row for 800 GeV stops. Shown in each panel top to bottom is signal (red), $t\bar{t}$+jets (blue), Z+jets (green), W+jets (yellow), and $t\bar{t}+Z$ (gray; hardly visible).

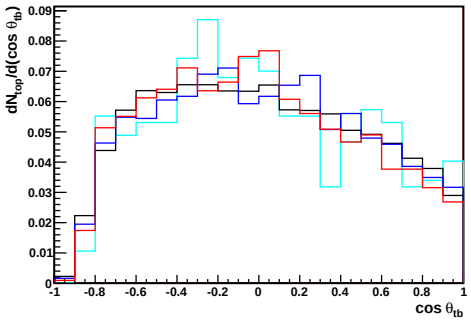

Figure 5. Distributions of $\cos\theta_{tb}$ for $t\bar{t}$+jets, normalized to one. Shown are the distribution without collimation cuts (black), with the lepton collimation cut only (blue), with the subjet collimation cut only (red), and with both collimation cuts applied (cyan).

3.3 Sensitivity to stop mixing

Table 2. p-values for distinguishing 600 GeV stop mixing hypotheses at 14 TeV LHC @ 100 fb^{-1}.

Truth	*Hypothesis ($m_{\tilde{t}_1}$ =600 GeV)*		
	left	mixed	right
left	1	0.018	1.7×10^{-6}
mixed	0.025	1	0.017
right	9.0×10^{-7}	0.018	1

To calculate sensitivity, we sum the $\cos\theta_{tb}$ distributions from signal and background processes. For the $t\bar{t}$+jets contribution, we use the control region distribution normalized to the total of the signal region distribution. This sum is shown in Figure 7, where we have rebinned the data to use only two bins in order to minimize the trials penalty. The corresponding p-values for distinguishing stop mixing hypotheses are shown in Tables 2, 3. For 600 GeV stops at 14 TeV LHC @ 100 fb^{-1}, left and

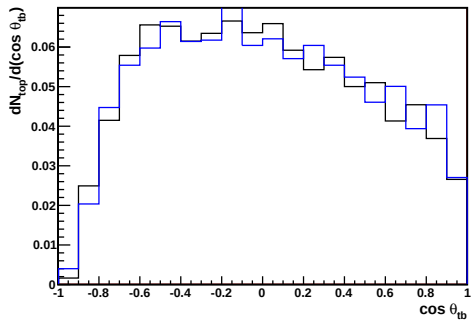

Figure 6. Distributions of $\cos\theta_{tb}$ for the $t\bar{t}$+jets control sample, normalized to one. Shown are the distribution for $M_{T2} < 225$ GeV (black), and for $M_{T2} > 225$ GeV.

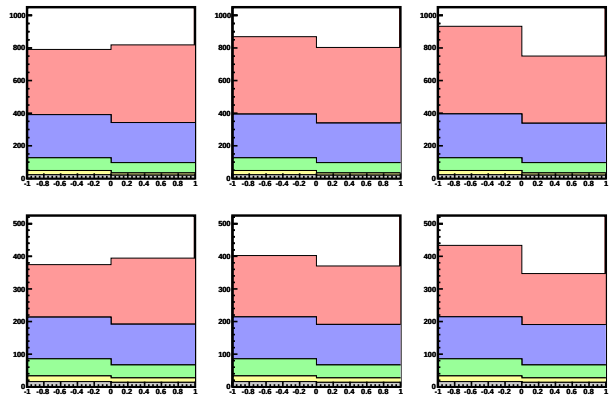

Figure 7. Sum of $\cos\theta_{tb}$ distributions from different processes for 14 TeV LHC @ 100 fb^{-1}. Left to right are left, mixed and right stop samples; upper row is for 600 GeV stops, and the lower is for 800 GeV stops. Shown are contributions from signal (red), $t\bar{t}$+jets (blue), Z+jets (green), W+jets (yellow), $t\bar{t}+Z$ (gray).

Table 3. p-values for distinguishing 800 GeV stop mixing hypotheses at 14 TeV LHC @ 100 fb^{-1}.

Truth	*Hypothesis ($m_{\tilde{t}_1}$ =800 GeV)*		
	left	mixed	right
left	1	0.17	5.6×10^{-4}
mixed	0.17	1	0.14
right	7×10^{-4}	0.15	1

right mixtures can be distinguished to better than 4.5σ, and left/right can be distinguished from the mixed state to better than 2σ. For 800 GeV stops, left and right can be distinguished to better than 3σ.

4 Conclusion and outlook

In summary, we described a simulation and analysis methodology for the $t + \not{E}_T$ collider signature that can distinguish stop mixing hypotheses at 14 TeV LHC with ~ 100 fb^{-1} of data using the substructure of boosted jets.

There are several possible improvements to the methodology. The first and foremost is to include polarization information from leptonic decays. Also if, in consultation with experimenters, the trimming threshold of $p_T = 10$ GeV or the subjet cone size $\Delta R = 0.2$ can be reduced, this would enhance the performance of the top reconstruction algorithm. Finally, it may be useful to implement the top reconstruction technique of Ref. [11] to supplement b-tagging, especially at large boosts. Other possible improvements are to implement charm mistagging, use spin-correlated backgrounds, and to calculate backgrounds at NLO.

Acknowledgments

This work is supported by the World Premier International Research Center Initiative (WPI Initiative) as well as Grant-in-Aid for Scientific Research Nos. 22540300, 23104005 from the Ministry of Education, Science, Sports, and Culture (MEXT), Japan. The authors would like to thank T. T. Yanagida and S. Matsumoto for preliminary discussions on this topic.

References

[1] B. Bhattacherjee, S. K. Mandal and M. Nojiri, [arXiv:1211.7261 [hep-ph]].
[2] M. Perelstein and A. Weiler, JHEP **0903**, 141 (2009) [arXiv:0811.1024 [hep-ph]].
[3] J. Shelton, Phys. Rev. D **79**, 014032 (2009) [arXiv:0811.0569 [hep-ph]].
[4] E. L. Berger, Q. -H. Cao, J. -H. Yu and H. Zhang, Phys. Rev. Lett. **109**, 152004 (2012) [arXiv:1207.1101 [hep-ph]].
[5] S. Ovyn, X. Rouby and V. Lemaitre, [arXiv:0903.2225 [hep-ph]].
[6] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **72**, 1896 (2012) [arXiv:1111.6097 [hep-ph]].
[7] S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **80**, 051501 (2009) [arXiv:0903.5081 [hep-ph]]; S. D. Ellis, C. K. Vermilion and J. R. Walsh, Phys. Rev. D **81**, 094023 (2010) [arXiv:0912.0033 [hep-ph]].
[8] D. Krohn, J. Thaler and L. -T. Wang, JHEP **1002**, 084 (2010) [arXiv:0912.1342 [hep-ph]].
[9] [CMS Collaboration], CMS-PAS-BTV-11-004.
[10] [ATLAS Collaboration], ATLAS-CONF-2012-043.
[11] D. Krohn, J. Shelton and L. -T. Wang, JHEP **1007**, 041 (2010) [arXiv:0909.3855 [hep-ph]].