Flavor tagging TeV jets for physics beyond the Standard Model

Zack Sullivan[1,a] and Keith Pedersen[1]

[1] *Department of Physics, Illinois Institute of Technology, Chicago, Illinois 60616-3793, USA*

Abstract. We present a new scheme for tagging boosted heavy flavor jets called "μ_x tagging." At the LHC, the primary method to tag b-jets relies on tracking their charged constituents. However, when highly boosted, track-based b-tags lose efficiency, and the probability to mistag light jets rises dramatically. Using muons from B hadron decay and defining a particular combination "x" of angular information and boost estimation, we find fairly flat efficiencies to tag b-jets, c-jets, light-quark jets, and light-heavy jets (containing B hadrons from gluon splitting) of $\epsilon_b = 14\%$, $\epsilon_c = 6.5\%$, $\epsilon_{\text{light-light}} = 0.1\%$, and $\epsilon_{\text{light-heavy}} = 0.5\%$, respectively. We demonstrate the usefulness of this new scheme by showing the reach for discovery of a leptophobic $Z' \to b\bar{b}$ in the dijet channel.

1 Introduction

As searches for W' and Z' bosons at the CERN Large Hadron Collider (LHC) shift to TeV-scale energies, observation of their decay products becomes challenging. Observation of dijet resonances above QCD background is hampered by falling b-tagging efficiencies (28–15% around 1–2 TeV) and large light-jet fake rates of 1–2% [1]. In addition to the low purity ($\epsilon_{\text{fake}}/\epsilon_b \sim 1/10$), large uncertainties in the tagging efficiencies affect the mass limits; e.g., the ATLAS b-tag uncertainty is 35% for $p_T > 500$ GeV [2]. In order to discover multi-TeV physics beyond the Standard Model (BSM), we need a better b tag with good efficiency and purity.

At this conference, we presented a new method for flavor tagging at TeV-scale energies called "μ_x boosted-bottom-jet tagging" [3]. This method is derived from kinematic first principles, and provides both a well-determined 14% efficiency for b-tagging, and a factor of 10 improvement in fake rejection over existing tags ($\epsilon_{\text{fake}}/\epsilon_b \sim 1/100$). In Sec. 2 we summarize the algorithm and cuts for the μ_x tag, show why it works, and plot its transverse momentum p_T- and pseudorapidity η-dependent efficiencies. In Sec. 3 we briefly describe the application of μ_x boosted-b tagging to an analysis for discovery of a leptophobic $Z' \to b\bar{b}$. We summarize our results in Sec. 4.

2 μ_x boosted-b tag

Consider a jet containing a semi-muonic decay of a B hadron. In the center-of-momentum (CM) frame, the muon is emitted with a speed $\beta_{\mu,\text{cm}}$ and at an angle θ_{cm} with respect to the beam axis (see

[a] e-mail: Zack.Sullivan@IIT.edu

Fig. 1). In the lab frame, the boost γ_B of the B hadron compresses its decay products into a narrow subjet at high energy. We define a lab frame observable

$$x \equiv \gamma_B \tan\theta_{\rm lab} = \frac{\sin\theta_{\rm cm}}{\kappa + \cos\theta_{\rm cm}}, \tag{1}$$

where $\kappa \equiv \beta_B/\beta_{\mu,{\rm cm}}$.

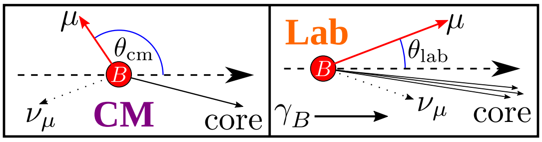

Figure 1. Nomenclature for the center-of-momentum frame and boosted lab frame.

While κ is unobservable, for sufficiently boosted B hadrons ($\gamma_B \gg \gamma_{\mu,{\rm cm}} \geq 3$) the lab frame distribution of the muon count N vs. x is effectively independent of κ,

$$\frac{dN}{dx} \approx \frac{2x}{(x^2+1)^2}. \tag{2}$$

This leads to a universal shape in x for highly boosted jets containing B hadrons. Using this shape we define the μ_x boosted-b tag as a cut on two variables: We capture 90% of muons from B decay by demanding $x < 3$. To further isolate b decays, we note the hard fragmentation function for b quarks leads to the B hadron subjet carrying a large fraction $f_{\rm subjet}$ of the total jet momentum. Hence, we demand

$$f_{\rm subjet} \equiv \frac{p_{T\,\rm subjet}}{p_{T\,\rm jet}} \geq 0.5. \tag{3}$$

There are two challenges in applying the μ_x tag to real events: we must identify the correct decay remnant of the B hadron to reconstruct its four-vector $p_{\rm subjet}$, and we must deal with the missing muon neutrino. Most of the neutrino energy in the lab frame comes from the boost, so we use the measured four-vector of the muon as a proxy $p_\nu = p_\mu > 10$ GeV. In order to find the non-leptonic remnant "core" of the subjet, we need a more sophisticated algorithm.

In order to reconstruct the boosted subjet we first cluster the jet using the anti-k_T algorithm with a $R = 0.4$. We then search for the core (generally the charm hadron remnant) by reclustering the muon and calorimeter towers with total jet energy fraction $f_{\rm tower}^{\rm min} > 0.05$ using a smaller $R_{\rm core} = 0.04$. We assume $m_{\rm core} = 2$ GeV (a typical charm hadron mass), and identify the "correct" core as the one which comes closest to $\sqrt{p_{\rm subjet}^2} = 5.3$ GeV. Since mismeasurements smear out the reconstructed energy of the subjet, if $m_{\rm subjet} > 12$ GeV we constrain the subjet mass to be 12 GeV. The parameters of the μ_x tag are summarized in Table 1.

Table 1. A summary of parameters chosen for μ_x boosted bottom jet tagging.

$\boldsymbol{R}$	0.4	$\boldsymbol{m_{\rm core}}$	2 GeV	$\boldsymbol{p_{T\mu}^{\rm min}}$	10 GeV
$\boldsymbol{R_{\rm core}}$	0.04	$\boldsymbol{m_B}$	5.3 GeV	$\boldsymbol{x_{\rm max}}$	3 ($x_{90\%}$)
$\boldsymbol{f_{\rm tower}^{\rm min}}$	0.05	$\boldsymbol{m_{\rm subjet}^{\rm max}}$	12 GeV	$\boldsymbol{f_{\rm subjet}^{\rm min}}$	0.5

In spite of its non-trivial reconstruction, x is effectively a dynamic angular cut on the muon. Defining ξ, the lab frame angle between the muon and the core, it is possible to calculate $\xi_{\rm max}$, the maximum μ-to-core angle which produces $x \leq 3$. For "soft" muons ($E_\mu \ll E_{\text{core}}/18$), this angular cut is relatively tight

$$\xi_{\text{max}}^{\text{soft}} \approx 3\frac{m_{\text{core}}}{E_{\text{core}}}. \tag{4}$$

Once the muons become "hard" ($E_\mu \geq E_{\text{core}}/18$), the cut loosens significantly

$$\xi_{\text{max}}^{\text{hard}} \approx 3\frac{m_{\text{subjet}}^{\text{max}}}{E_{\text{core}}}. \tag{5}$$

While the transition between these limits depends explicitly on the muon's p_T, this dependence is small until just below the hard threshold. Thus, not only is x a *smart* angular cut — scaling with the energy of the core — it is a *dual* angular cut; tight for soft muons, looser for hard muons, and sensitive to the p_T resolution of the muon system only within the narrow transition region.

The separation of reconstructed b jets from light-quark-initiated jets can be seen in Fig. 2. Bottom jets (b-quarks hadronized as B hadrons) above 500 GeV produce large f_{subjet} and $x \sim 0.8$. Light jets (mostly π and K) produce either incompatible values of $x > 3$, or random subjet recombinations that lead to small f_{subjet}. A small fraction of b jets is not well-reconstructed (represented by the low-f_{subjet} tail), but it has little effect on the total efficiency.

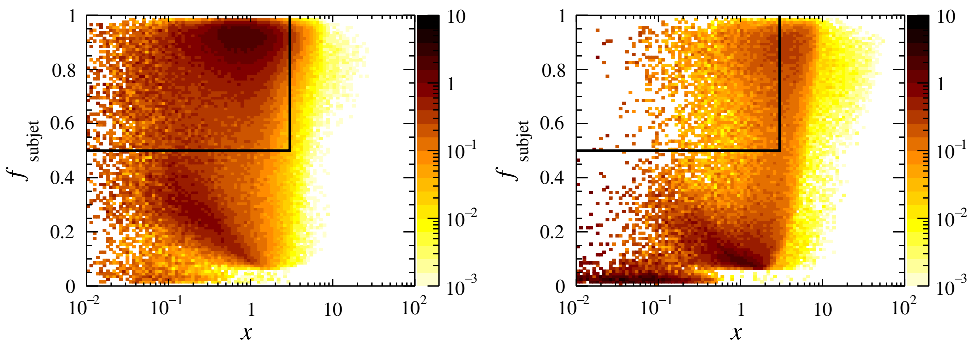

Figure 2. Density of reconstructed candidate tags with $\mu = 40$ pileup events as a function of f_{subjet} vs. x for (left) bottom and (right) light-quark-initiated jets.

We extract the standalone μ_x tagging efficiencies using PYTHIA 8.210 [4, 5] fed into an ATLAS-like version of DELPHES 3.2 [1], and a custom μ_x tagging module `MuXboostedBTagging` (available on GitHub [6]). In Fig. 3 we show separate efficiencies as a function of p_T and η for bottom jets, charm jets, light-light jets (where the muon came from a light-flavor hadron), and light-heavy jets (where a gluon split to $b\bar{b}/c\bar{c}$ — producing heavy-flavor hadrons in the final state). The kinematic nature of the tagging variables leads to fairly flat efficiencies in pseudorapidity, and when $p_T > 500$ GeV. The exception is the η distribution for B hadrons from gluon splitting. This leads to the intriguing possibility that the $g \to b\bar{b}$ contribution to jets in the Monte Carlo could be calibrated using the rapidity dependence of these highly-boosted jets.

3 A search for leptophobic $Z' \to b\bar{b}$

Very massive Z' bosons are expected to exist in many BSM models. We test the μ_x boosted-bottom tag by examining the reach at a 13 TeV LHC for a leptophobic Z' decaying to $b\bar{b}$ or $c\bar{c}$. For this study

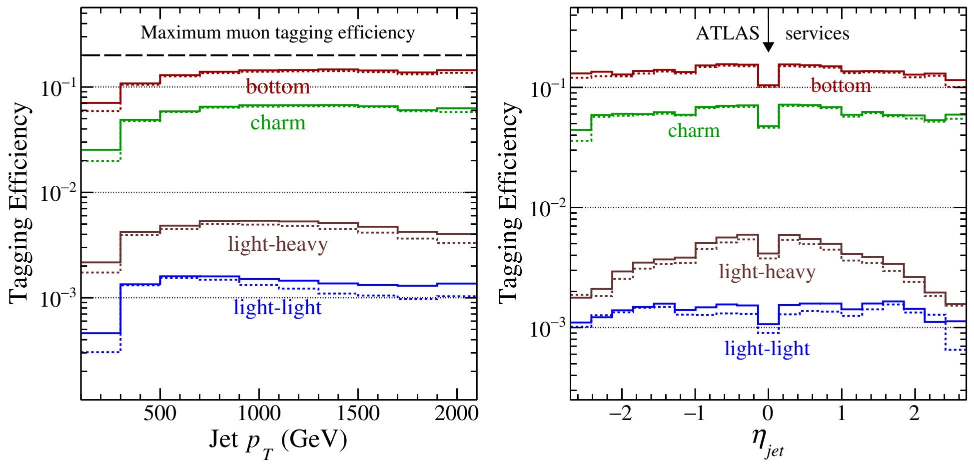

Figure 3. μ_x tagging efficiency vs. (left) jet p_T and (right) $\eta_{\rm jet}$. Solid (dashed) lines include $\mu = 0$ (40) pileup events.

we choose a $U(1)'_B$ Lagrange density

$$\mathscr{L} = \frac{g_B}{6} Z'_{B\mu} \bar{q} \gamma^\mu q, \tag{6}$$

with a flavor-independent coupling to quarks [7, 8].

We simulate the signal and backgrounds using a MLM-matched MadEvent sample [9] and CT14llo PDFs [10] fed through PYTHIA into DELPHES. In addition to demanding one or two μ_x tags (as defined in Sec. 2), we require $|\eta_j| < 2.7$, and $\Delta\eta_{jj} < 1.5$. We reconstruct a dijet mass out of the two leading-p_T jets, and look for a resonance in the mass window $[0.85, 1.25] \times M_{Z'_B}$.

The results for 5σ discovery of this leptophobic Z' are shown in Fig. 4 for a two-tag, and one-tag inclusive sample, compared to current exclusion limits from Ref. [7]. In 100 fb^{-1} of integrated luminosity at 13 TeV, a two b-tag analysis could discover a Z' of 3 TeV if the universal coupling $g_B \sim 2.5$. For this particular model, the single-tag inclusive search would be more effective — allowing for discovery up to nearly 1 TeV above current mass limits. Should a discovery not be made, the two-tag search (not shown) would set a 95% C.L. exclusion comparable to the one-tag discovery reach; while the one-tag search would set a 95% C.L. exclusion that can access g_B couplings a factor of 2 smaller than current limits, and masses up to 2 TeV higher.

4 Conclusions

In this paper we discuss the new μ_x boosted-bottom-jet tag. Combining angular information x from B hadron decay with jet substructure $f_{\rm subjet}$ in TeV-scale jets allows for tagging efficiencies of $\epsilon_b = 14\%$, $\epsilon_c = 6.5\%$, $\epsilon_{\rm light-light} = 0.1\%$, and $\epsilon_{\rm light-heavy} = 0.5\%$, respectively. The results here focused on ATLAS because their standalone non-isolated muon tagging efficiency is publicly available. We expect that if CMS has similar non-isolated muon tagging capability this tag will be just as effective, since it is kinematically driven and not sensitive to fine details of the detector.

When applying the μ_x tag to a search for leptophobic Z' bosons, we find that the reach for discovery at a 13 TeV LHC is about 1 TeV higher than current limits. If a Z' is not found, 95% C.L.

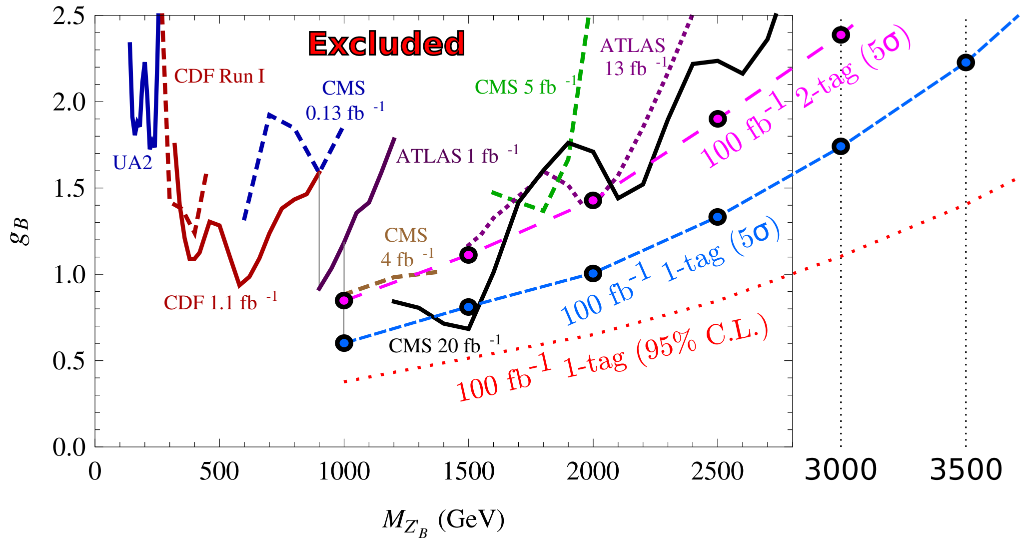

Figure 4. 5σ discovery reach for a leptophobic Z' with universal coupling in the with one or two boosted-b tags at a 13 TeV LHC compared to exclusion limits from Ref. [7]. Also shown is the 95% C.L. exclusion reach of the one-tag analysis.

exclusion limits can be set up to 2 TeV higher, or for g_B couplings a factor of 2 smaller, than the current limits. In addition to $Z' \to b\bar{b}$, the μ_x tag should be of immediate use in the search for $W' \to t\bar{b}$ in the boosted-top and boosted-bottom channel [11] conducted by the ATLAS Collaboration [2].

Acknowledgments

Z.S. would like to thank the ISMD organizers for an exceptional symposium and environment for lively discussions. This work was supported by the U.S. Department of Energy under award No. DE-SC0008347.

References

[1] J. de Favereau *et al.* (DELPHES 3 Collaboration), J. High Energy Phys. **02**, 057 (2014) [arXiv:1307.6346 [hep-ex]].
[2] G. Aad *et al.* (ATLAS Collaboration), Phys. Lett. B **743**, 235 (2015) [arXiv:1410.4103 [hep-ex]].
[3] K. Pedersen and Z. Sullivan, arXiv:1511.05990 [hep-ph].
[4] T. Sjostrand, S. Mrenna, and P.Z. Skands, J. High Energy Phys. **05**, 026 (2006) [hep-ph/0603175].
[5] T. Sjostrand, S. Mrenna, and P.Z. Skands, Comput. Phys. Commun. **178**, 852 (2008) [arXiv:0710.3820 [hep-ph]].
[6] K. Pedersen, [https://github.com/keith-pedersen/delphes/tree/MuXboostedBTagging].
[7] B.A. Dobrescu and F. Yu, Phys. Rev. D **88**, 035021 (2013); Erratum: Phys. Rev. D **90**, 079901 (2014) [arXiv:1306.2629 [hep-ph]].

[8] B.A. Dobrescu, arXiv:1506.04435 [hep-ph].
[9] J. Alwall *et al.*, J. High Energy Phys. **07**, 079 (2014) [arXiv:1405.0301 [hep-ph]].
[10] S. Dulat *et al.*, arXiv:1506.07443 [hep-ph].
[11] D. Duffty and Z. Sullivan, Phys. Rev. D **90**, 015031 (2014) [arXiv:1307.1820 [hep-ph]].