

## Tagging top in leptonic decay

---

### Suman Chatterjee\*

Department of High Energy Physics, Tata Institute of Fundamental Research,  
1 Homi Bhabha Road, Mumbai, 400005, India

E-mail: [suman.chatterjee@cern.ch](mailto:suman.chatterjee@cern.ch), [suman.chatterjee@tifr.res.in](mailto:suman.chatterjee@tifr.res.in)

### Rohini Godbole

Center for High Energy Physics, Indian Institute of Science,  
Bangalore 560012, India

E-mail: [rohini@iisc.ac.in](mailto:rohini@iisc.ac.in)

### Tuhin S. Roy

Department of Theoretical Physics, Tata Institute of Fundamental Research,  
1 Homi Bhabha Road, Mumbai, 400005, India

E-mail: [tuhin@theory.tifr.res.in](mailto:tuhin@theory.tifr.res.in)

We propose a method to identify jets consisting of all the visible remnants of a boosted top quark decaying semileptonically with an electron in the final state. An overlap of electron shower with the  $b$  quark initiated shower, and the large nontrivial energy-momentum carried by the invisible neutrino in the top quark decay are the two obstacles to achieving this aim. Our method uses the distribution of energy in different parts of the detector to identify a jet containing an energetic electron, involves use of substructure of the jet to determine the momentum associated with the electron and then completes the identification of top jet with the construction of new variables which would reflect the top quark decay kinematics. The last part involves an ansatz of the existence of a massless, invisible 4-momentum *roughly* collimated with the electron, whose 4-momentum when combined with that of the the electron and the full jet, reconstructs to the  $W$  boson and the top quark respectively. We demonstrate the efficacy of this proposal using simulated data and show that our method not only reduces the backgrounds from light flavor jets,  $b$  jets from QCD, and hadronic top jets, it can also tell apart jets rich in electrons but not due to top quark decays.

European Physical Society Conference on High Energy Physics - EPS-HEP2019 -  
10-17 July, 2019  
Ghent, Belgium

---

\*Speaker.

## 1. Introduction

Identifying events containing remnants of top particles produced at the Large Hadron Collider (LHC) is of paramount interest, since they provide profound insights into the standard model (SM) of particle physics as well as allow us to look for the hints of new physics. In fact, models which address the naturalness problem in the SM, or attempt to solve the issue of flavor hierarchies in SM fermions, often contain new heavy states which couple to top quark. Consequently, traces of new physics, if it exists, are very likely to show up in events with top quarks with transverse momenta  $p_T \gg m_t$ , where  $m_t$  denotes the top quark mass. Because of high boost factors, these top quarks decay into collimated collections of particles that often look like single jets of large size in the detector. A plethora of tagging algorithms to identify jets containing all the decay products of hadronically decaying top quarks have been proposed which range from the substructure analyses to methods taking full advantage of recent advances in the machine learning.

In case of top quark decays to muons, these “muonic top jets” get characterized in terms of non-isolated muons which leave tracks in the muon spectrometer portion of the detector. Matching the spectrometer track to a track seen at the tracker portion of the detector may allow us to reduce the background due to QCD heavy flavor jets. Note, however, that tagging “electronic top jets” where the boosted top quark decays to an electron is rather hard, since identifying an electron when its shower partially overlaps with the shower from the fragmentation of a  $b$  quark is difficult. Also, neutrinos associated with these semileptonic decays carry away substantial fraction of top quark energy. This makes tagging these semileptonic top jets highly nontrivial.

In this work (detailed in [1]) we propose a dedicated tagger for electronic top jets. The procedure of identification relies on two different kinds of observables and an algorithm. The first set of these variables is a collection of jet substructure based observables which attempt to decide whether the observed jet is consistent with a jet containing an energetic electron or not, by drawing information from various parts of the calorimeters (namely, the electromagnetic portion or the ECal and the hadronic portion or the HCal) and the tracker. The second set of observables are calculated using an ansatz that there exists a non-zero momentum 4-vector (massless and approximately collimated to the candidate 4-momentum for the electron) which reconstructs  $W$  boson mass when combined with the electron, and reconstructs top quark mass when combined with the full jet. These observables therefore have physical interpretations only when the jet is an electronic top jet. Clearly, the efficacy of the second set relies on finding the 4-momentum for the electron from top quark decay. The purpose of our proposed algorithm is to find a candidate for the electron given a jet by combining the tracker and the calorimeter information. One of the novel aspects of our proposal is that since we utilize more kinetic information associated with top quark decays, we are able to tell apart the electronic top jets from those jets which contain showers due to a  $b$  quark and an energetic electron but are not due to top quark decays. As a proof of principle, we consider jets containing decay products of the stop quark of supersymmetry [2], where the stop quark decays to a  $b$  quark, an electron, a neutrino, and an invisible neutralino. Even though we do not use any information pertaining to these stop jets for constructing the tagger, we show that our method can separate electronic top jets from stop jets quite efficiently. Based on these observations we propose a simple extension of the tagger, which apart from finding electronic top jets, can also identify jets due to new physics processes as anomalous jets.

## 2. Simulations Details

To demonstrate the potency of the methodology proposed, we generate all the Monte Carlo (MC) samples at partonic level using MadGraph5\_aMC@NLO V5 2.6.2 [3] with NNPDF2.3LO [4] parton distribution function at centre-of-mass energy of 13 TeV. We use DELPHES 3.4.1 software package [5] with CMS geometry [6], where stable particles from PYTHIA [7] are converted into detector objects such as energy deposits in calorimeter cells and tracks. The particle-flow (PF) algorithm [8], as implemented in DELPHES uses these detector elements to construct particle-flow candidates, namely muons, electrons, photons, charged and neutral hadrons. Additionally, we use 4C tune [9] to simulate the relevant busy hadronic collider environment.

We denote the jets from  $t\bar{t}$  sample, where both the top quarks decay to electrons, by  $t(e)$ , the jets from  $t\bar{t}$  sample, where both the top quarks decay hadronically, by  $t(h)$ , the jets from  $b\bar{b}$  production in QCD by  $b$ , the light quark or gluon jets from multijet production by  $j$ , and the jets from stop pair production with dielectronic final state by  $\tilde{t}(e)$ .

We use anti- $k_T$  jet clustering algorithm [10] in FASTJET [11] to cluster the particle-flow candidates into jets. In particular, we use  $R = 0.8$  with  $p_{T,\min} = 500$  GeV. The hardest jet (in  $p_T$ ) in the event is used in our analysis.

## 3. Variables for tagging

The first stage of this procedure is to groom the jet using “soft drop” (SD) method [12], which allows us to identify the last stage of clustering with hard splitting, and to remove soft radiation making it somewhat robust under underlying event and pileup. We use only the groomed jet for further analysis.

We utilize the distribution of energy of the groomed jet, as well as of its subjets, over ECal and HCal, and the signature left at the tracker to develop a number of variables which allow us to determine whether the jet contains an energetic electron within itself or not. We use the notation  $\mathcal{V}_e$  to denote the this of variables.

### 3.1 Variables in $\mathcal{V}_e$

- We use the mass of the groomed jet, denoted by  $m_{\text{SD}}$ .
- We use the nonhadronic energy fraction of the jet  $J$ , defined by

$$f_{1-h} \equiv 1 - f_h \quad (3.1)$$

where  $f_h$  is the ratio of the energy deposit by the jet in HCal to the total energy of the jet.

- We propose a new variable which measures the asymmetry between hadronic energy deposits of two SD subjets  $j_1$  and  $j_2$ .

$$A_h \equiv \frac{(f_h^1 - f_h^2)^2}{(f_h^1 + f_h^2)^2}, \quad (3.2)$$

where  $f_h^1$  and  $f_h^2$  are the hadronic energy fractions of two subjets respectively.

- We denote the charge radius of the jet  $J$  by  $r_C$ , and define it by

$$r_C \equiv \frac{1}{d_0} \sum_{k \in \text{tracks}} q^{(k)} p_T^{(k)} \Delta R_{kJ}, \quad \text{where} \quad d_0 = \sum_k p_T^{(k)}, \quad (3.3)$$

$q^{(k)}$  and  $p_T^{(k)}$  are the charge and the transverse momentum of the  $k$ -th track inside the jet  $J$  respectively,  $\Delta R_{kJ}$  is the angular distance of the track from the jet axis.

- We estimate the neutral fraction of nonhadronic energy of the jet  $J$ , denoted by  $f_{1-h}^N$  and defined as

$$f_{1-h}^N \equiv \frac{1}{E_J \times f_{1-h}} \sum_{k \in \text{ECal}} \delta_{q^{(k)},0} E_J^{(k)}, \quad (3.4)$$

where  $\delta_{q^{(k)},0}$  ensures that we use only those ECal cells that don't correspond to any track matched to it.

- We also use ratio of  $N$ -subjettiness variables [13],  $\tau_2$ , and  $\tau_1$ , denoted by  $\tau_{21}$  ( $\equiv \tau_2/\tau_1$ ), which measures the compatibility of the jet with a 2-prong structure.

We devise a set of steps, which, as described below, starts with the constituents of the groomed jet and determines the momentum four-vectors, which may correspond to the momenta of the  $b$  quark and the electron from top quark decay.

- The subjet (referred as  $\tilde{j}$ ) with lower  $f_{\text{h}}^j$ , is identified to be the one most likely to contain an electron<sup>1</sup>.
- The hardest track in  $\tilde{j}$  is denoted by  $\mathcal{T}$ . In particular, the pseudorapidity and the azimuthal angle of  $\mathcal{T}$  are referred to as  $\eta_e$  and  $\phi_e$  respectively.
- Constituents of  $\tilde{j}$  are clustered further to find two exclusive  $k_T$  subjets. Among these, the energy of the subjet containing  $\mathcal{T}$  is recorded as  $E_e$ .
- The four-vector of  $e$  candidate, denoted by  $p_e$ , is defined as

$$p_e \equiv \{E_e, E_e \sin \eta_e \cos \phi_e, E_e \sin \eta_e \sin \phi_e, E_e \cos \eta_e\}, \quad (3.5)$$

- The four-vector of the  $b$  candidate (denoted by  $p_b$ ) is defined using the four-vector of the full jet ( $p_J$ ) and  $p_e$  as follows :

$$p_b \equiv p_J - p_e. \quad (3.6)$$

The ansatz that the jet contains the remnants of a boosted top quark implies that if a momentum 4-vector (representing the invisible neutrino), *massless and roughly collimated* to the electron, is added to the electron or the full jet, the resultant should reconstruct the momentum of a  $W$  or a top particle respectively. This ansatz allows us to determine the energy carried away by the neutrino and some aspects of its direction inside the jet; we develop two measurable variables,  $Z_b$  and  $\Theta_{b/e}$ , using this ansatz along with kinematic constraints, and refer this set of two variables by  $\mathcal{V}_V$ .

<sup>1</sup>This assignment has been checked, using generator-level information, to be correct for more than 85% of times.

### 3.2 Variables in $\mathcal{V}_\nu$

Variables in  $\mathcal{V}_\nu$  are defined by -

$$Z_b \equiv \frac{E_b}{E_e + E_\nu} \quad (3.7)$$

$$\Theta_{b/e} \equiv \frac{E_e (m_t^2 - \Delta^2)}{E_b m_W^2} \quad (3.8)$$

, where  $\Delta^2 \equiv (m_W^2 + m_b^2 + 2p_b \cdot p_l)$ ,  $E_\nu \simeq \frac{1}{2} \frac{m_t^2 - \Delta^2}{(E_b - p_b \cdot \hat{e})}$ , and  $m_W$  ( $m_b$ ) is the mass of  $W$  boson ( $b$  quark).

$Z_b$  is the ratio of the energies of  $b$  and  $W$  candidates and  $\Theta_{b/e}$  reduces to  $\frac{1 - \cos \theta_{vb}}{1 - \cos \theta_{ve}}$  at  $p_b \gg m_b$ , where  $\theta_{vb}$  ( $\theta_{ve}$ ) is the opening angle between the neutrino and  $b$  ( $e$ ) candidates.

We construct two BDT based multivariate discriminators, which separate electronic top jets (to be treated as signal jets) from QCD  $b$  jets (to be treated as background jets). Following the notation defined in Eq. (3.9), we construct a plane of responses, where any jet is represented by a point.

$$\begin{aligned} r_e &\equiv \text{response of the BDT to discriminate } t \text{ from } b \text{ using variables in } \mathcal{V}_e \text{ in the range } \{-1, +1\} \\ r_\nu &\equiv \text{response of the BDT to discriminate } t \text{ from } b \text{ using variables in } \mathcal{V}_\nu \text{ in the range } \{-1, +1\} \end{aligned} \quad (3.9)$$

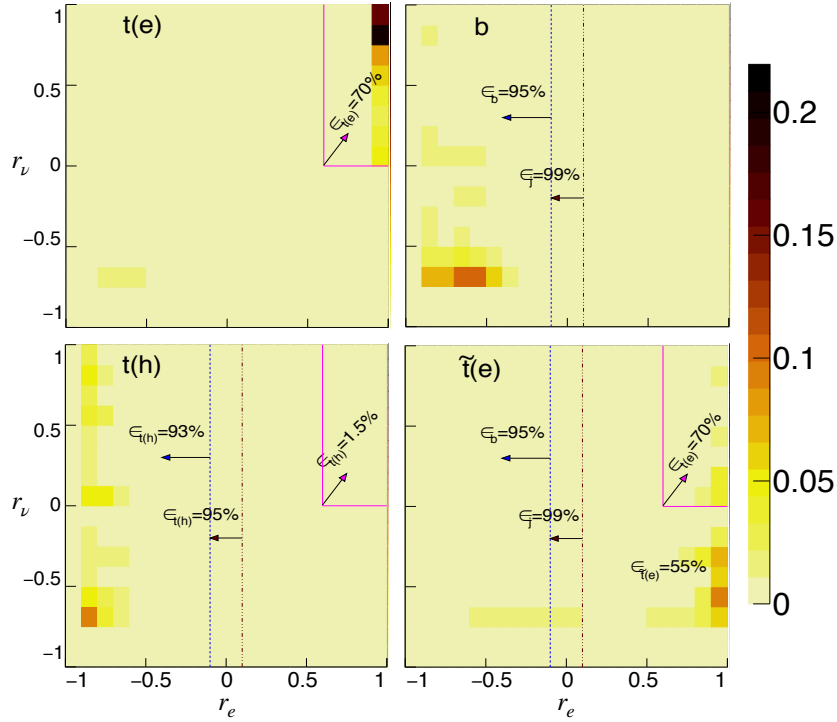
In Fig. 1, we show the probability distributions of different types of jets considered. By construction, we find  $t(e)$  jets around the corner  $(+1, +1)$  and  $b$  jets around the corner  $(-1, -1)$ . As far as  $t(h)$  jets are concerned, these are characterized by small  $r_e$ . Note that this is a bonus feature since the BDT using  $\mathcal{V}_e$  is optimized to separate  $t(e)$  jets from  $b$  jets only. In the  $r_\nu$  direction, however, there is no clear separation of  $t(e)$  jets from  $t(h)$  jets. This is understandable, since we expect  $t(h)$  jets to satisfy the critical ansatz used for neutrino reconstruction.

Note that the true benefit of using variables in  $\mathcal{V}_\nu$  (consequently  $r_\nu$ ) is that, in this direction  $\tilde{t}(e)$  jets get completely separated from  $t(e)$  jets, where  $r_e$  fails to set them apart. To separate electronic top jets from SM backgrounds, we can simply use a cut on  $r_e$ . In Fig. 2, we show the Receiver Operator Characteristic curves for signal ( $t(e)$  jets) and various backgrounds. Note again, in order to produce this plot we only use  $r_e$ , the response of the BDT trained to separate  $t(e)$  jets from  $b$  jets using only the observables in  $\mathcal{V}_e$ . As expected, we see for a top quark decaying to an electron, our tagger is able to provide fantastic separation:  $\gtrsim 75\%$  signal efficiency at 1% mistag rate from  $b$  jets as well as light flavor jets and less than 2% mistag rate from hadronic top jets.

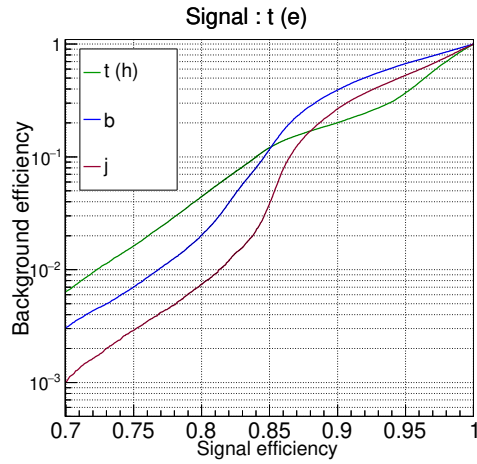
In the second approach, we construct different zones in  $r_e$ - $r_\nu$  plane, which are enriched in jets from different samples (depicted with quantitative details in Fig. 1), namely, a rectangular box near  $(+1, +1)$  containing  $t(e)$  jets (with small fakes from other samples), zones with low  $r_e$  containing light and heavy quark jets from QCD, we term the remaining space in the plane as ‘‘anomalous zone’’, which contains mostly stop jets with small contamination from jets arising in SM processes.

## 4. Conclusion

In this letter, we presented a new method to identify jets consisting of all the visible remnants of a boosted top quark which decays to an electron. The first part of this method uses observables



**Figure 1:** Probability distribution of BDT responses over a 2-D plane for various event samples. Selection efficiencies for different samples are also overlaid in regions enriched by jets from particular samples.



**Figure 2:** Receiver operator characteristic (ROC) curves quantifying the performance of the BDT classifier to identify top jets, where the top quark decays to a final state with electron, from various backgrounds.

computed for a large sized jet using information from different parts of the detector, with which one tries to determine whether the jet is consistent with a jet containing an energetic electron. In the second part of this methodology, we propose a way to identify the 4-momentum associated with the candidate for electron inside the jet and employ an extra ansatz that there exists a massless invisible 4-momentum, roughly collimated with the electron, which when combined with the electron and the full jet, reconstructs the 4-momentum of an on-shell  $W$  boson and top quark respectively. This allows us to determine tell-tale features of the top quark decay kinematics given entirely in terms of visible objects. Combining both these parts we construct a tagger which identifies electronic top jets at high efficiency, with small mistag rate from background jets from SM processes, and even jets containing electrons inside (but not due to top quark decay).

## References

- [1] S. Chatterjee, R. Godbole and T. S. Roy, *Jets with electrons from boosted top quarks*, [[arXiv:1909.11041](#)].
- [2] S. P. Martin, *A Supersymmetry primer*, [hep-ph/9709356](#). [Adv. Ser. Direct. High Energy Phys.18,1(1998)].
- [3] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, *JHEP* **07** (2014) 079, [[arXiv:1405.0301](#)].
- [4] **NNPDF** Collaboration, R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP* **04** (2015) 040, [[arXiv:1410.8849](#)].
- [5] **DELPHES 3** Collaboration, J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi, *DELPHES 3, A modular framework for fast simulation of a generic collider experiment*, *JHEP* **02** (2014) 057, [[arXiv:1307.6346](#)].
- [6] **CMS** Collaboration, S. Chatrchyan et al., *The CMS experiment at the CERN LHC*, *JINST* **3** (2008) S08004.
- [7] T. Sjöstrand, S. Ask, J. R. Christiansen, R. Corke, N. Desai, P. Ilten, S. Mrenna, S. Prestel, C. O. Rasmussen, and P. Z. Skands, *An Introduction to PYTHIA 8.2*, *Comput. Phys. Commun.* **191** (2015) 159–177, [[arXiv:1410.3012](#)].
- [8] **CMS** Collaboration, A. M. Sirunyan et al., *Particle-flow reconstruction and global event description with the CMS detector*, *JINST* **12** (2017), no. 10 P10003, [[arXiv:1706.04965](#)].
- [9] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, *JHEP* **05** (2006) 026, [[hep-ph/0603175](#)].
- [10] M. Cacciari, G. P. Salam, and G. Soyez, *The Anti- $k(t)$  jet clustering algorithm*, *JHEP* **04** (2008) 063, [[arXiv:0802.1189](#)].
- [11] M. Cacciari, G. P. Salam, and G. Soyez, *FastJet User Manual*, *Eur. Phys. J.* **C72** (2012) 1896, [[arXiv:1111.6097](#)].
- [12] A. J. Larkoski, S. Marzani, G. Soyez, and J. Thaler, *Soft Drop*, *JHEP* **05** (2014) 146, [[arXiv:1402.2657](#)].
- [13] J. Thaler and K. Van Tilburg, *Identifying Boosted Objects with  $N$ -subjettiness*, *JHEP* **03** (2011) 015, [[arXiv:1011.2268](#)].