Measuring b-quark jet structure at the LHC

PHYS4022P

Author Student ID: 2547473m
Supervisor: Dr. Andy Buckley

2023-10-11

Contents

Introduction

While the Standard Model remains highly successful in particle physics, theories beyond it, known as Beyond the Standard Model (BSM), often predict an abundance of b-quarks [1]. Thus, understanding the composition and behavior of these b-quarks is crucial, forming the foundational framework for a comprehensive understanding of particle physics.

Jets, characterized as collimated sprays of hadrons resulting from high-energy collisions [1], serve as a primary focus in this project. The primary source of bottom quarks in this project stems from the decay pattern of top and antitop quark pairs ($t\bar{t}$). This choice is deliberate, given its distinct decay channel of b-jets plus dileptonic [12], offering a complex yet distinguishable experiment.

This project delves into the in-depth measurement of bottom quark jet structures in two parts: simulation data and real-life ATLAS data. The analysis encompasses various key characteristics of these jets, including Les Houches Angularity (LHA), N-Subjettiness ratio (τ_{21}), Energy Correlation Functions (ECF), C2, and D2 correlations [1]. A significant emphasis lies in establishing solid selection criteria to identify valid b-jets. The "Tag and Probe" method will be implemented for the analysis of real data, evaluating the efficiency and purity of these events. The objective is to showcase the b-jet structures from collision events at the LHC, providing insights into particle interactions at high-energy scales.

The simulation employs Rivet and Pythia systems (C++) and is informed by prior ATLAS analyses, specifically "ATLAS_2022_I2152933" [10] and "ATLAS_2019_I1724098" [4]. Analysis code for real data is crafted anew using Python. All simulation models and analysis code are available on GitHub: (https://github.com/HowaiMak/ATLAS_2023_BJETS). Detailed instructions on implementations and startup guidelines are noted in the accompanying "README" file.

Monte-Carlo Collision Model

Monte Carlo (MC) event generation, fundamental in particle physics, utilizes iterative random sampling and statistical techniques to simulate complex collision events [1].

Our simulation employs the Rivet and Pythia systems. During our experimentation phase involving 200,000 events, optimal outcomes were achieved by sampling events with either 2 or 3 jets for ttbar events, termed the signal (SIG) mode. This ensures the presence of at least 2 jets, crucial for $t\bar{t}$ events, while accommodating random jet production at LHC with a third jet [12]. While this additional jet may influence reconstruction, it doesn't impact b-tagging accuracy. Conversely, the background (BKG) mode requires events to possess at least 1 jet, ensuring the presence of light and gluon jets for noise considerations. Mode toggling and decay pathway adjustments are facilitated using a separate C++ file, "ttbar-dilep.cmnd".

The "ATLAS_2023_BJETS.cc" serves as the primary analysis simulation model, allowing convenient exploration of diverse structures and filtering conditions. This includes constraining jet production within $|\eta| < 3.5$, with a slightly extended detection range of $|\eta| < 4.0$. Lepton production is limited within $|\eta| < 2.5$. The minimum p_T value for particles to be recognized as either a lepton or jet is set at $25GeV$. The code structure draws inspiration from two previous ATLAS analyses, "ATLAS_2022_I2152933" [10] and "ATLAS_2019_I1724098" [4].

Sources of Bottom Jets

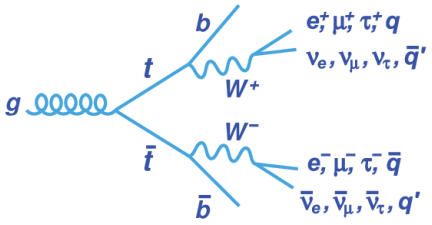


Figure 1: Top and anti-Top decay channel [2]

Bottom quarks (b quarks) can be produced through various radiational processes at the LHC, each with its own set of advantages and disadvantages:

- $t\bar{t}$ decay: The $t\bar{t}$ channel exhibits a significant incidence rate, characterized by a natural double b-jet purity, thus enabling the generation of purer samples for analysis[12]. This channel is notably prominent in decay processes observed at high-energy colliders such as the LHC[1]. Nevertheless, the presence of random accompanying jets can pose challenges in accurately identifying and distinguishing b-jets. The reconstruction process is further complicated by fragmentation and hadronisation phenomena.

- Higgs bosons: Higgs bosons are directly associated with bottom quarks, enabling investigations into b-quark interactions during Higgs decay processes [3]. They provide valuable insights into the properties of the Higgs boson and its interactions with other particles. However, their production rate is lower compared to ttbar events, and they may involve additional decay channels, complicating the identification of b-jets.

- Gluon jets: Gluon jets provide valuable insights into gluon splitting and fragmentation processes, offering a diverse range of energies and multiplicities for studying b-jet characteristics. However, they generally exhibit a lower occurrence rate of b-jets compared to ttbar events. Analysis and interpretation are complicated by background contamination from non-b-jets.

Among these, the top and anti-top $t\bar{t}$ decay channel stands out as a dominant means of probing bottom quarks at the LHC. Despite challenges posed by accompanying random jets, event simulation leveraging double b-jet purity within the $t\bar{t}$ channel enables efficient and targeted investigations into the structural characteristics of b-jets.

W^+W^- Boson Decay

The $t\bar{t}$ decay channel inherently produces two b-quarks and W^+, W^- bosons. However, diverse decay modes exist for the W bosons, with a 9% probability of decaying into a lepton pair (comprising electrons, muons, or tau in any two-part combination), a 45% likelihood of decay into one lepton plus a jet (originating from a light quark), and a 46% chance of complete hadronic decay (involving jets only) [9]. To exclusively attain a final state in the $t\bar{t}$ events featuring jets of bottom quarks (dijets case), it is imperative to configure the simulation such that the W bosons are constrained to decay solely into dileptonic pairs. This framework establishes a nuanced yet distinguishable context for studying bottom quarks within a sufficiently complex yet distinguishable channel.

Background Contamination

Several decay channels, such as double light jet emissions from Quantum Chromodynamics (QCD) and Weak Boson Fusion (WBF), among others, can produce excessive light and gluon jets, contributing to background noise [1]. However, these channels are intricate, involving complex particle interactions that are not thoroughly investigated and pose challenges for accurate modeling and simulation. In our model, the light and gluon jets are generated through the decay of double bosons W^+, W^-, Z, or the jet plus single Z boson decay channel. This focused selection ensures a more manageable and targeted investigation of the desired b-jet characteristics.

Simulation Results

Bottom Jet Identification

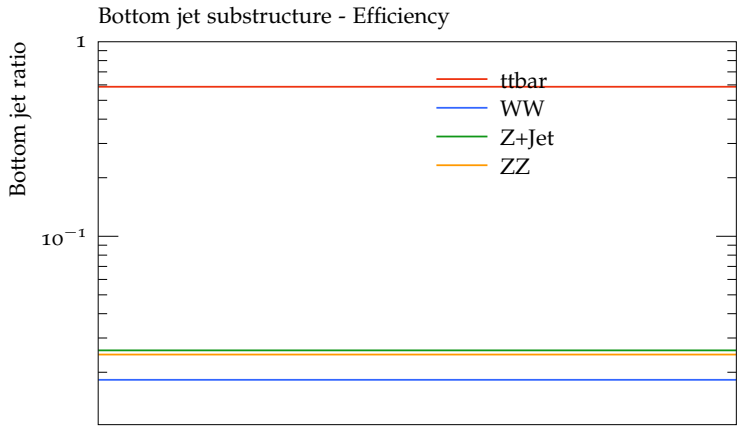


Figure 2: Bottom Jet Natural Occurance Rate

Particle collision experiments encounter challenges in accurately identifying b-jets and implementing effective quark flavor tagging, primarily due to mechanical constraints and the complex experimental environment, including limitations in the detection angle $|\eta|$ [1]. These factors contribute to suboptimal jet reconstruction from final states. Computational simulations offer a crucial solution to address these challenges by providing a controlled environment with perfect flavor tagging. Within this controlled setting, parameters can be adjusted, enabling the exploration of scenarios that may present difficulties in replication under real experimental conditions.

Figure 2 illustrates the natural occurrence rate of b-jets in the collision model. Within the $t\bar{t}$ decay channel, 58.7% of the total jets are identified as b-jets. In contrast, within the background, 1.83% originate from the W^+W^- channel, 2.59% from Z+Jet, and 2.47% from ZZ. These results align with expectations, demonstrating a higher detection rate of bottom jets in $t\bar{t}$ events compared to background events, which typically involve numerous light and gluon jets.

Jet Variables

Les Houches Angularity (LHA)

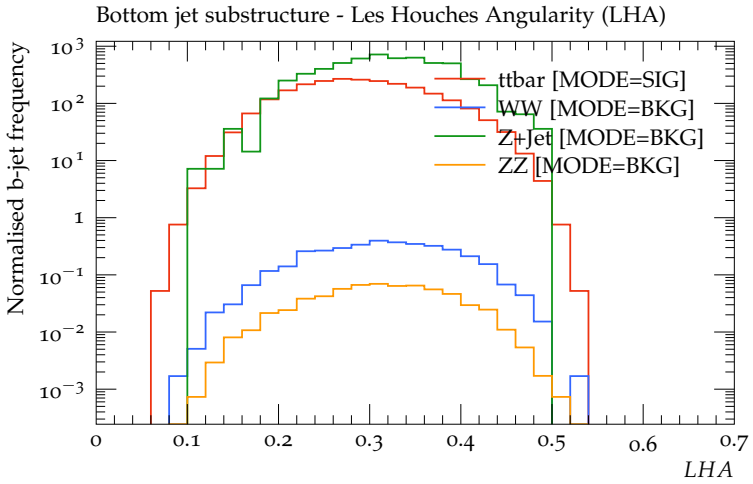


Figure 3: Scaled angular and momentum distribution of jet

The Les Houches Angularity (LHA) is a type of observable known as "generalised angularities" [1]. The entire family of angularities is determined by the variation of parameters κ and $\underline{(}eta$, which are restricted to values greater than or equal to 0.

$$\begin{aligned} \lambda_\beta^\kappa &= \sum_i z_i^\kappa \theta_i^\kappa \\ &= \sum_i \left(\frac{(p_T)_i}{(p_T)_{\text{jet}}}\right)^\kappa \left(\frac{\Delta R_i}{R}\right)^\beta \end{aligned} \tag{1}$$

Here, the index i iterates over the jet constituents, where $z_i \in [0,1]$ represents the i-th subjet's transverse momentum (p_T) relative to the total p_T of the containing jet, and $\theta_i \in [0,1]$ denotes the i-th subjet's radius to the jet central axis relative to the total jet radius $R = 1$.

$\lambda_{0.5}^1 = LHA$ stands out for its precise characterisation of the interplay between angular and momentum scalings within jets. The condition $\kappa = 1$ is pivotal in confining angularities to the infrared and collinear (IRC) safe domain, maintaining stability in the presence of infrared divergence enabling calculable within perturbation theory [5]. Simultaneously, we specify $\beta = 0.5$ to introduce a non-zero angular dependence, which has less emphasis than $\kappa = 1$ [1].

Figure 3 illustrates the normalized count of detected b-jets corresponding to various LHA values across distinct decay channels in our simulation. The LHA value is scaled down by a factor of 10^3. A noticeable increase in the detected b-jet count is observed, notably in the $t\bar{t}$ and Z+Jet channels. This observation aligns with expectations, given the well-established tendency of both decay modes to produce an abundance of bottom quarks.

The peak LHA value across all channels converges at approximately 320, indicating that the optimal LHA value for b-jets should center around this numerical value. The distribution of b-jet LHA values spans a range from 100 to 500.

N-subjettiness Ratio (τ_{21})

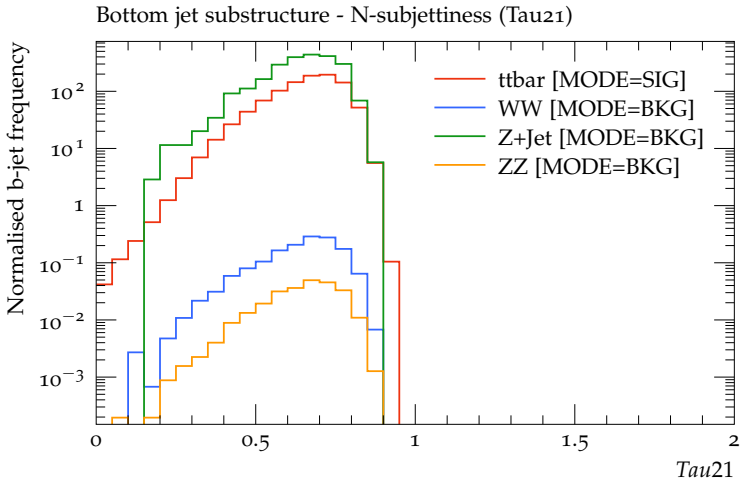


Figure 4: 2-prong bjet against 1-prong bjet ratio frequency

The N-subjettiness (τ_N) quantifies the extent to which a jet can be considered composed of N subjets. It is calculated by hypothesising N subjets from a reconstructed jet using jet algorithm, and is evaluated using the formula:

$$\tau_N = \frac{1}{\sum_i (p_T)_i R_0} \sum_i (p_T)_i \min(\Delta R_{1,i}, \Delta R_{2,i}, ..., \Delta R_{N,i}) \tag{2}$$

Here, the index i iterates over the constituent particles within a given jet, pT_i represents their transverse momenta, and $\Delta R_{J,i} = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ denotes the distance in the rapidity-azimuth plane between a candidate subjet J and a constituent particle i. R_0 signifies the characteristic jet radius ($R = 1$) utilised in the original jet clustering algorithm [11]. The N-subjettiness Ratio τ_{NM} gains additional power when employed, akin to the likelihood-ratio discriminant between hypotheses. It takes the form of an N-subjettiness ratio between different hypothesized N values,

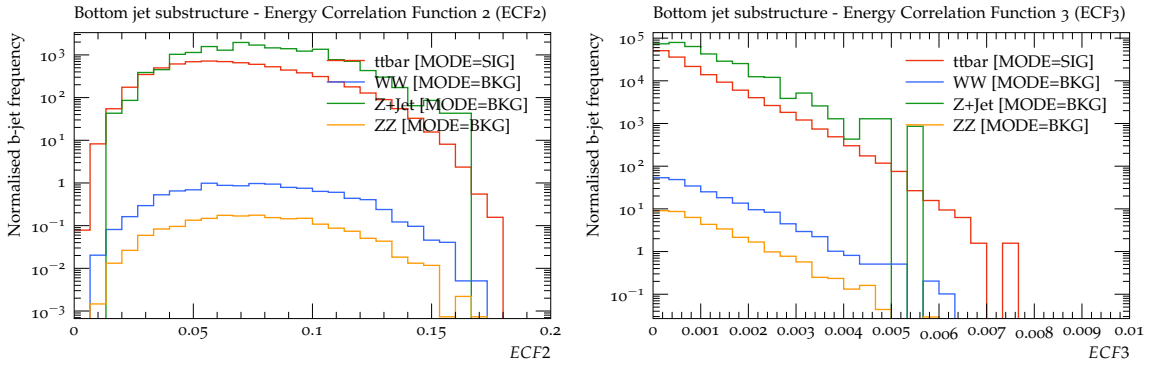
$$\tau_{nm} = \frac{\tau_n}{\tau_m} \tag{3}$$

In our simulation, the specific value of $\tau_{21} = \tau_2/\tau_1$ is selected for evaluation to ascertain whether our events exhibits a 1-jet-like or 2-jet-like configuration. Figure 4 displays the distribution of τ_{21} values across all events. As $\tau_N = 0$ indicates a definite N-jet-like event, so a higher τ_{21} value suggests a greater likelihood of the bottom jet event having a 1-subjet-like structure, whereas a lower value suggests a 2-subjet-like configuration.

Across all decay channels, the τ_{21} value does not exceed 1.0, indicating a high propensity towards 2-subjettiness. With an approximate peak at 0.67, this suggests that events are roughly 30% more

likely to be characterized by 2 jets. Particularly noteworthy is the distinctive 2-subjettiness feature evident ($\tau_{21} = 0$) in the $t\bar{t}$ channel, absent in the background channels. This distinction is attributed to the heightened production of light quarks and gluons in the background channels.

Energy Correlation Coefficient (ECF)



(a) ECF2 (b) ECF3

Figure 5: Likely-hood of 2 or 3 particles event comparison

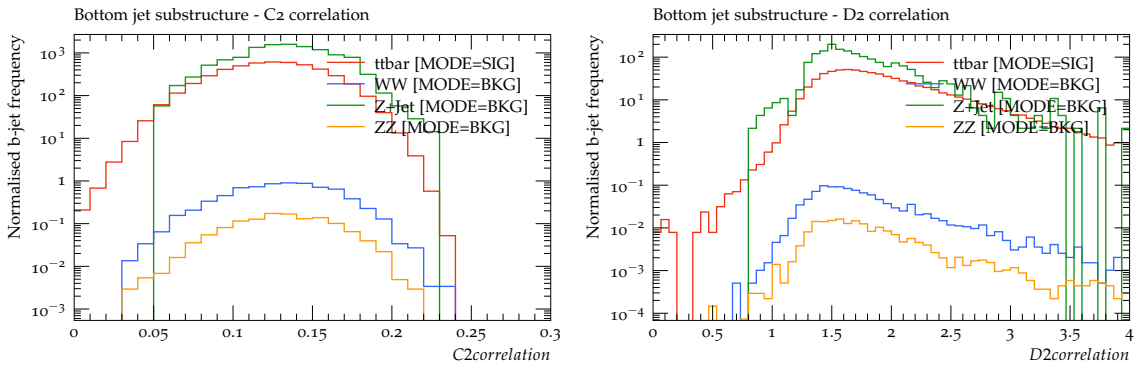
The energy correlation function is expressed as the summation of the transverse momentum products of each particle pair, denoted as i, j, and k within the jet J. This summation is multiplied by the β-weighted distance ΔR in the rapidity-azimuth plane, as previously defined. Unlike several prior jet substructure methodologies, these correlation functions do not necessitate the explicit delineation of subjet regions [8].

$$ECF(2,\beta) = \sum_{i<j\in J} (p_T)_i (p_T)_j (\Delta R_{ij})^\beta \tag{4}$$

$$ECF(3,\beta) = \sum_{i<j<k\in J} (p_T)_i (p_T)_j (p_T)_k (\Delta R_{ij}\Delta R_{ik}\Delta R_{jk})^\beta \tag{5}$$

Figure 5 presents the Energy Correlation function for both 2-particle and 3-particle likelihoods of bottom quark events. Notably, the $ECF2$ values showcase a relatively uniform distribution across their range, whereas the $ECF3$ values display a peak at 0, gradually decreasing as the range increases. This suggests that the majority of events are highly likely to involve 2 particles, with approximately 10^1 events (at the halfway point of ECF3) potentially involving 3 particles. Once again, the $t\bar{t}$ channel exhibits a higher inclination than the background channels towards generating 3-particle events, as indicated by its ECF3 values extending to 0.0075.

C2 & D2 correlation



(a) C2 correlation

(b) D2 correlation

Figure 6: Subjet-number sensitivity

Equation 4 and 5 demonstrate that if a jet contains fewer than N particles, the corresponding Energy Correlation Function (ECF) returns a value of 0. Thus, in a system with N subjets, the magnitude of $ECF(N+1,\beta)$ should be notably smaller than that of $ECF(N,\beta)$ [1]. This prompts an investigation into the energy correlation function ratio, the $C2$ and $D2$ correlation.

$$C_2 = \frac{ECF3 \times ECF1}{(ECF2)^2} \tag{6}$$

$$D_2 = \frac{ECF3 \times (ECF1)^3}{(ECF2)^3} \tag{7}$$

One immediate point of contrast between N-subjettiness and the energy correlation function ratio is that $C2$ and $D2$ do not require a separate procedure (such as minimization, see equation 2) to determine. This reduces complexity when investigating subjet structures with reconstruction algorithms.

The value of $C2$ in Figure 6a is scaled down by 10^2, while the values of $D2$ in Figure 6b are scaled up by 10^1. Both ratios effectively identify 2-body structures [7]; their values increase as the number of subjets inside the large-R jet increases. Notably, the C2 value peaks around 125 and maintains a relatively even distribution ranging from 0 to 250. On the other hand, the value of D2 exhibits a rapid increase within the range 0 to 0.1, followed by a gradual decrease up to 0.4, where heavy noises are observed, disturbing the data.

Real Data from ATLAS at LHC

Event Selections

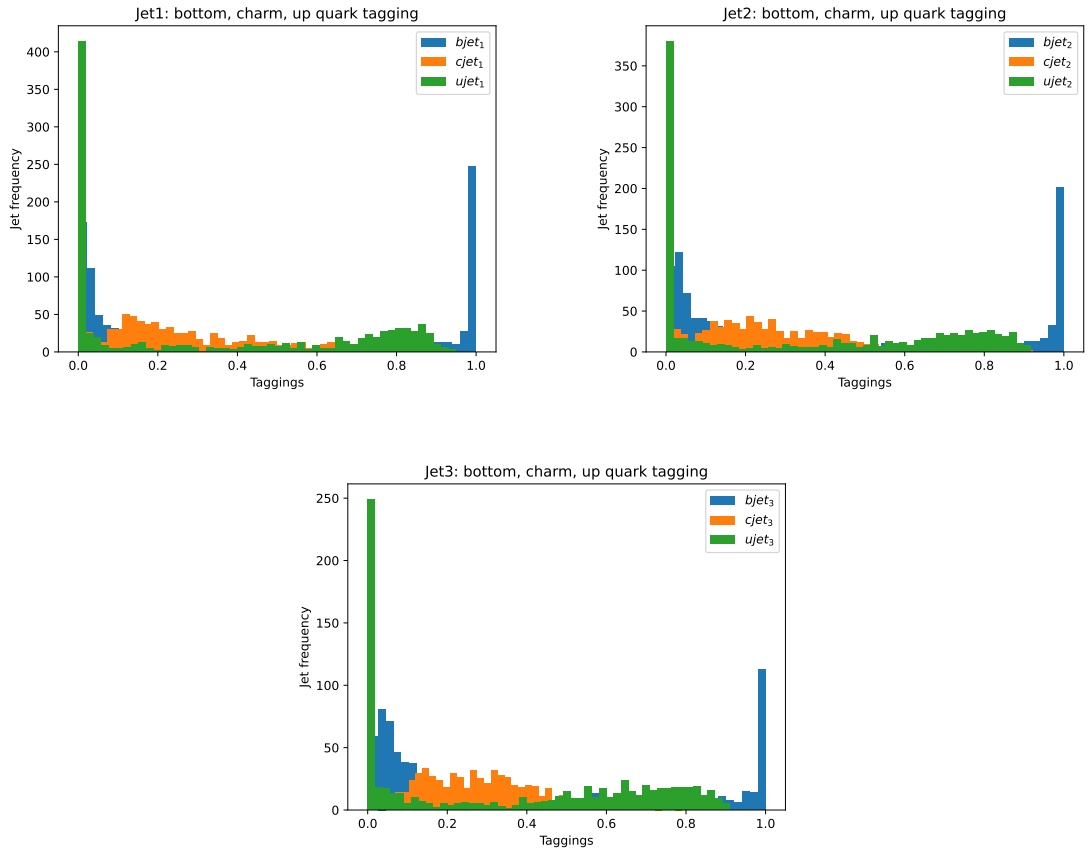


Figure 7: Probabilities of jets being a bottom, charm, or up quark in real collision events

From simulation studies, insights into the characteristic features of b-jets have been gained. Utilizing this knowledge, the tagging algorithm can be applied to real data to estimate the probabilities

of each quark. To adhere to the constraints established in our initial investigations, which focus on studying b-quarks through the $t\bar{t}$ channel, real collision events containing 1 electron and 1 muon, each with a transverse momentum (p_T) of 20,000 GeV, are selected. Additionally, jets occurring in numbers of 2 or 3, with $p_T > 30,000$ GeV, are chosen.

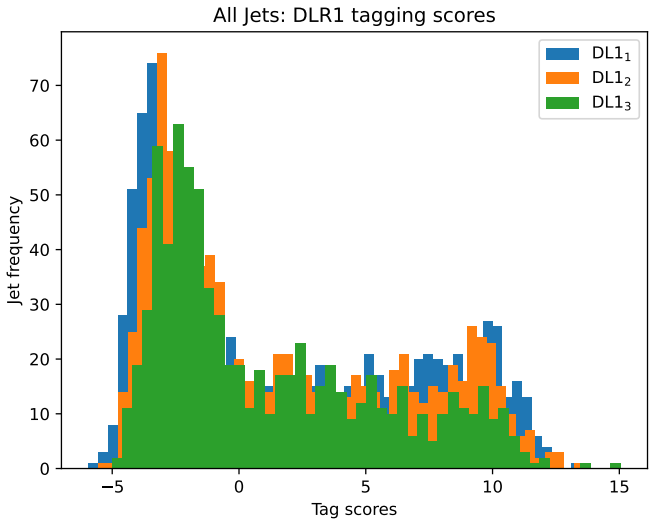


Figure 8: DL1 tag scores for all jets

In general, a balance must be struck between the efficiency and purity of particle detection. Achieving high efficiency is straightforward if all detections are deemed valid (i.e., considering everything as a bottom quark). However, this approach is impractical, especially in collider physics, where jet structures carry various particles.

The DL1 score is an equation utilizing the probability of quark detection in a jet to assess the detection efficiency standards for a desired quark. Equation 8 displays the DL1 score for b-tagging [6].

$$DL1_b = \ln\left(\frac{p_b}{f_c p_c + (1 - f_c) p_u}\right) \tag{8}$$

Here, p_b, p_c, and p_u represent the bottom, charm, and up quark tagging probabilities respectively. f_c is the fraction of charm quarks within each event, typically defined as a fixed value, which in our experiment will be 0.03.

Figure 7 illustrates the frequency of jets against the flavor tagging probabilities (p_b, p_c, p_u). Jet1 represents the leading pT in each event, with Jet2 and Jet3 following. It's evident that the b-taggers exhibit relatively high certainty. The identification of b-jets peaks only around 1.0 or 0.0, indicating a 100% and 0% b-tagging probability, respectively. The identification of u-jets is mainly centered around the lower end, which also demonstrates a good representation, as fewer light jets are expected after the initial selections. The production of c-jets is minimal, as indicated by the f_c value previously, rendering them less effective.

With all the above p values processed through equation 8, Figure 8 displays the DL1 score for the jets. The general efficiency for b-tagging is chosen around 77%, corresponding to a DL1 score of 2.23 [6]. In our investigation, we aim for a value larger than 2.96, which corresponds to efficiencies of 70% or below. The reason for choosing jet events with lower efficiency is to leverage for investigating the purity of bottom jet structure measurements through a special method (Tag & Probe) to mitigate tagging bias.

Tag and Probe

The Tag and Probe method (T&P) is employed as a technique to mitigate bias in detection. This strategy utilizes the properties of the dobule b-jet from $t\bar{t}$ channel to scrutinize the tagging bias of bottom quarks [1], following this procedure:

- Tag: Define a well-identified bottom quark, known as the "b-tag", which passes the DL1 test score.
- Probe: Instead of measuring the "b-tag", assess the substructure of its paired b-jet from the $t\bar{t}$ decay channel.

The analysis of bottom quarks will encompass both the "b-tag" and its accompanying jet, known as the direct and TNP methods, respectively. To compare variations between these methods, the χ^2 test will be employed, and the data will be normalized in equal binning. A χ^2 value of 0 indicates perfect symmetry between measurements.

Bottom Jet Structures

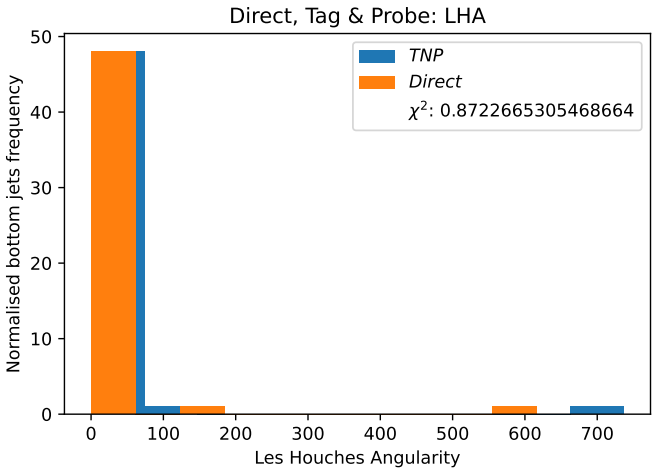


Figure 9: Les Houches Angularity

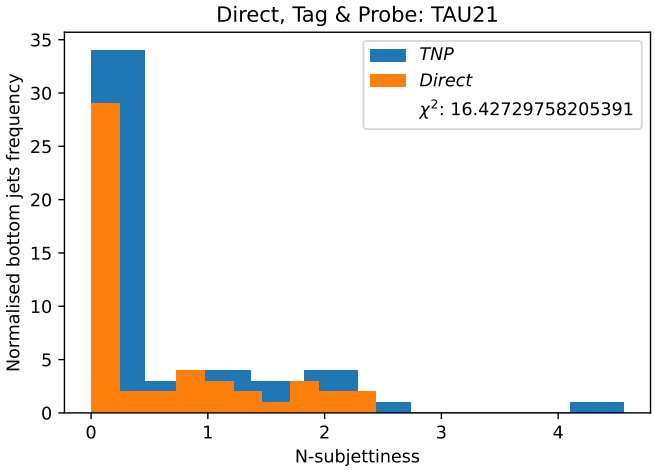


Figure 10: N-subjettiness ratio

Figure 9 presents the measured LHA value from the selected passing b-jets. Both methods exhibit close similarity in the measurements, as indicated by the $\chi^2 = 0.87$ value. The distribution of LHA is more condensed around 0 to 100, in contrast to the wide spread of values predicted in Figure 3.

Figure 10 illustrates the measured τ_{21} value from the selected passing b-jets. Although the plots peak between 0 to 1, consistent with simulations (see Figure 4), the two methods show disagreement, as suggested by the $\chi^2 = 16.43$ value. The TNP method identified overlooked high τ_{21} b-jets, which appear to be four times more likely to be 1-prong than 2-prong jets in this region.

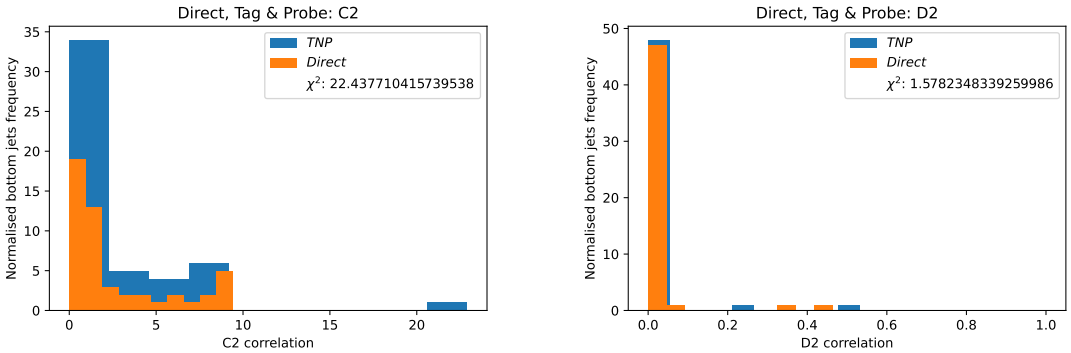


Figure 11: C2 & D2 correlation

Comparing Figure 11 to Figure 6, we observe a significant shift towards lower values in both energy correlation ratios. This suggests that the model overestimated the number of subjet structures for the b-jets compared to the actual results [7].

Bibliography

[1] Andy Buckley, Christopher White, and Martin White. *Practical Collider Physics*. 2053-2563. IOP Publishing, 2021. ISBN: 978-0-7503-2444-1. DOI: 10.1088/978-0-7503-2444-1. URL: https://dx.doi.org/10.1088/978-0-7503-2444-1.

[2] The CDF et al. "Combination of the top-quark mass measurements from the Tevatron collider". In: *Physical Review D* 86 (July 2012). DOI: 10.1103/PhysRevD.86.092003.

[3] ATLAS Collaboration. *Higgs boson observed decaying to b quarks*. URL: https://atlas.cern/updates/briefing/higgs-observed-decaying-b-quarks.

[4] Christian Gutschow Deepak Kar Amal Vaidya. *Jet substructure at 13 TeV Experiment: ATLAS (LHC)*. 2019. URL: https://rivet.hepforge.org/analyses/ATLAS_2019_I1724098.html.

[5] Philippe Gras et al. "Systematics of quark/gluon tagging". In: *Journal of High Energy Physics* 2017.7 (July 2017). ISSN: 1029-8479. DOI: 10.1007/jhep07(2017)091. URL: http://dx.doi.org/10.1007/JHEP07(2017)091.

[6] Marie Lanfermann. "Deep Learning in Flavour Tagging at the ATLAS experiment". In: Jan. 2018, p. 764. DOI: 10.22323/1.314.0764.

[7] Andrew J. Larkoski, Ian Moult, and Duff Neill. "Power counting to better jet observables". In: *Journal of High Energy Physics* 2014.12 (Dec. 2014). ISSN: 1029-8479. DOI: 10.1007/jhep12(2014)009. URL: http://dx.doi.org/10.1007/JHEP12(2014)009.

[8] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. "Energy correlation functions for jet substructure". In: *Journal of High Energy Physics* 2013.6 (June 2013). ISSN: 1029-8479. DOI: 10.1007/jhep06(2013)108. URL: http://dx.doi.org/10.1007/JHEP06(2013)108.

[9] Izaak Neutelings. *Branching fraction matrix of pair decays*. 2021. URL: https://tikz.net/sm_decay_matrix/.

[10] Andrea Helen Knue Shayma Wahdan Dominic Hirschbuhl. *Measurement of observables sensitive to colour reconnection in ttbar dileptonic emu channel at 13 TeV Experiment: ATLAS (LHC)*. 2022. URL: https://rivet.hepforge.org/analyses/ATLAS_2022_I2152933.html.

[11] Jesse Thaler and Ken Van Tilburg. "Identifying boosted objects with N-subjettiness". In: *Journal of High Energy Physics* 2011.3 (Mar. 2011). ISSN: 1029-8479. DOI: 10.1007/jhep03(2011)015. URL: http://dx.doi.org/10.1007/JHEP03(2011)015.

[12] Timothee Theveneaux-Pelzer. "Measurement of ttbar with additional jets with the ATLAS detector". In: (2019). URL: https://cds.cern.ch/record/2687376.