Summer Student Programme
DESY Hamburg 2014

Jet shapes in top quark pair events in the di-lepton channel

Name: Harrison Schreeck

EMail: harrison.schreeck@stud.uni-goettingen.de

Supervisor: James Howarth

Contents

Figure 1: Top quark pair production at the LHC. Via Quark-Antiquark (top) or gluon fusion (bottom). [3]

1 Introduction

Since its discovery in 1995 at the Tevatron by the CDF[1] and D0[2] the Top quark has been studied in great detail. The Top quark is very special compared to other quarks as it does not hadronize due toits high mass and hence very short lifetime. At hadron colliders Top quarks are usually produced in pairs (details of the production and decay of Top quark pairs are given in Sec.(2.1)). Top quark pair events are particularly interesting as they feature a high energy component, the hard-scattering parton interaction which leads to the creation of the top pair, but also a soft energy part because of the interacting proton remnants, called underlying event (UE). The UE consists of partons which have not interacted in the main hard-scattering and any additional hard-scattering processes in the same collision, which is known as multiple parton interaction (MPI). There are also contributions from other effects like initial and final state radiation (ISR/FSR) and the interaction of the top decay products with the rest of the event. This is discussed in more detail in Sec.(2.3). Because soft interactions cannot be calculated by perturbative quantum chromodynamics (QCD), they are modelled phenomenologically in Monte Carlo generators. Top-pair events can be used to check if the effects of UE are modelled properly. In this analysis jet shapes in Top pair events are used to check tuning of the Monte Carlo models used to simulate Top-pair events.

2 Theory

In this section an overview is given of the used event observables, the concept of Monte Carlo generators, and the selection criteria for Top-Antitop events.

2.1 Top Pair Production

Top quarks are usually produced in pairs at hadron collider via QCD interaction, although a single top production via electroweak interaction is also possible but with a significantly lower cross section. The various production channels for top-antitop pairs, at leading order, can be

seen in Fig.(1). Because the Top quark has a much higher mass ($173 \pm 0.52 \pm 0.72$ GeV [4]) than any other quark, it decays before hadronization and nearly exclusively into a W boson and a b quark. The W boson then decays into either a quark/antiquark pair (hadronically) or into a lepton and the corresponding flavour neutrino (leptonically). Top-quark events are therefore characterized by the decay of the two W bosons. If both W bosons decay hadronically, the event has 6 jets in total with two of them being b jets, which can experimentally be separated from light jets. This decay channel has by far the largest branching ratio, but has the disadvantage of significant backgrounds from multi-jet events. If one W boson decays hadronically and the other leptonically, which is called the lepton + jets channel, there are 4 jets with two b-jets and a high p_T lepton together with missing transverse energy ($E\!\!\!/_T$) from the neutrino. The branching ratio

Top Pair Branching Fractions
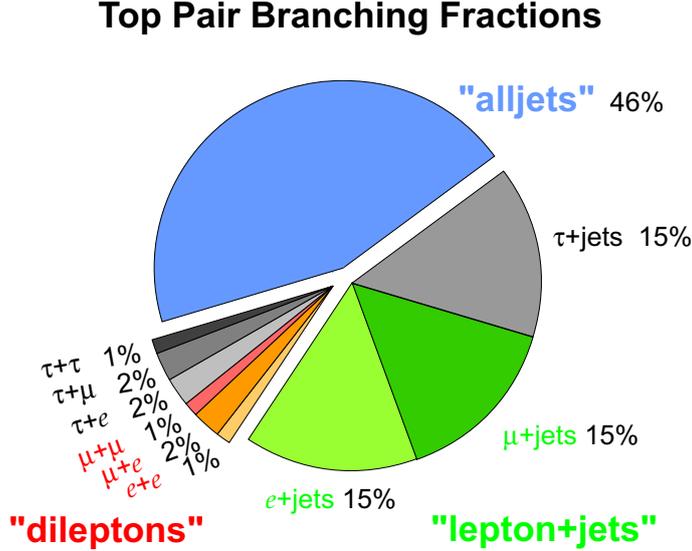
Figure 2: Branching fraction of the different possible decay channels for Top Quark pairs. [5]

for this decays channel is roughly 15% per lepton, which gives 30% in total (taus are usually not considered because they are much more difficult to identify). This decay channels suffers less from background contributions and has a reasonably high branching ratio, which is why this channel is often called 'golden channel'. The last decay channel is the dilepton channel in which both W bosons decay leptonically. This gives just 2 b-jets, two isolated, high p_T leptons and missing transverse energy from the two neutrinos. This channel has by far the lowest branching ratio ($\approx 5\%$), but features a very unique signature which makes it easy to identify and distinguish from background events. This channel is used in the following analysis because of its high purity. The main backgrounds in this channel are Drell-Yan events, single top and WW/WZ/ZZ+jets events. By requiring two b-tags (inclusive) and other selection criteria, such as exactly two leptons with opposite charge, these backgrounds are suppressed. The b-tagging is achieved using the MV1 algorithm [6]. This algorithm combines the results of multiple variables to identify a bjet at an efficiency of 70 %. To remove the Drell-Yan background, only the $e\mu$ channel is used, which decreases the statistics but leaves, in combination with the other criteria, only single-top events as a relevant background. A list of all selection criteria used can be seen in Tab.(1).

2.2 ATLAS detector & LHC

The ATLAS detector [7] is a general-purpose particle detector at the Large Hadron Collider (LHC) at CERN in Geneva. The LHC [8] is the worlds largest particle accelerator with a circumference of 27 km where protons are collided with protons. In its first run the center of mass energy was

Event selection criteria
$n_{\text{Jets}} \geq 2$
$n_{\text{Leptons}} = 2$
Opposite sign leptons
Cosmic muon removal
$n_{\text{B-Jets}} \geq 2$
$H_T > 130\,\text{GeV}$

Table 1: List of applied event selection criteria.

7 TeV (2010-2011) and 8 TeV (2012). The LHC was built to search for the Higgs boson and other new particles. Because of the large center of mass energy it also serves as a Top-quark factory. The cross section at the LHC is two orders of magnitude higher than at the Tevatron where the Top-quark was discovered and studied for the first time.

The ATLAS detector has a solid angle coverage of nearly 4π and features an inner detector responsible for tracking that covers the pseudorapidity range $|\eta| < 2.5$. The azimuthal angle ϕ is covered completely. The inner detector is built of silicon pixel detectors, a semiconductor tracker (SCT) and a straw-tube transition radiation tracker(TRT). The inner detector is surrounded by a 2 T magentic field created by a superconducting solenoid. The calorimeters used in the detector are sampling calorimeters. The electromagnetic calorimeter uses lead and liquid argon while the hadronic calorimeter uses scintillating tiles.Muons are measured in the muon spectrometer (MS) which has a toroidal magnet system to measure the muon track momenta separately. At ATLAS a trigger system with three levels is implemented, where the first one is a hardware trigger and the other two are software triggers.

2.3   Hadron Colliders and MC Generators

At hadron colliders, the colliding particles are not elementary particles but composite particles, which results in some experimental challenges. When two protons collide, the hard interaction occurs between the partons, which are the valence quarks, gluons and sea quarks.
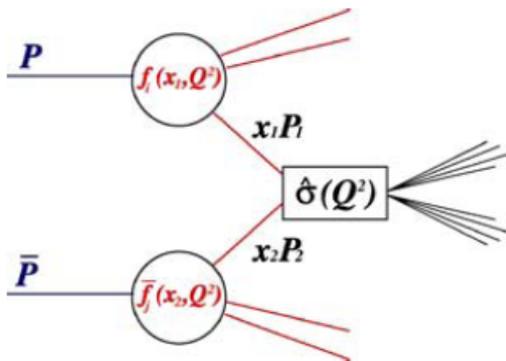
Figure 3: Parton interaction at a hadron collider. [3]

The collision is divided into three parts (see Fig.(3)), which is called factorization theorem. The first part is the short-distance partonic cross section and the second part consists of long-range effects, described by Parton Density Functions (PDFs), fragmentation functions and form factors.

The colliding partons carry only a fraction x of the momentum from the parent particle. The probability of a certain parton (valence quark, gluon, sea quark) to carry a specific fraction x is described by the PDFs. The PDFs are not calculated directly by perturbative QCD but are extracted from fits to data from deep inelastic scattering experiments like HERA at DESY.

The hard-scattering process between the two partons from the protons can be calculated with perturbative QCD, as the energy is high and the strong coupling constant α_s is small. The

matrix elements of the process can be calculated using feynman rules in leading-order (LO), next-to-leading-order (NLO) or even higher orders, where more and more virtual corrections to the tree level process are considered.

After the hard-scattering process the outgoing quarks and gluons start begin to shower and hadronize, where quarks radiate additional gluons and gluons split into quark/antiquark pairs. In this process the energy is lowered to a point where perturbative QCD is no longer applicable, because α_s is no longer small. Eventually quarks are confined into hadrons. To describe the hadronization different models are used. The most popular are String fragmentation (also called Lund model) and Cluster fragmentation shown in Fig.(4). The motivation behind the String model is that the QCD potential is between two coloured objects and the the force between them rises linearly for large distances, like a spring, which is usually visualised by cylindric shaped field lines that look like strings, see Fig.(4(b)). The cluster model on the other hand is motivated by the observation that partons in a shower tend to be clustered in colorless groups. This is represented by ellipses shown in Fig.(4(a)). Both models are used today in Monte Carlo generators. Pythia6 [9], for example uses the string model, while Herwig [10] uses the Cluster model. Both programs use, as an input, the perturbative calculations for a given process and then model the hadronization. The perturbative calculation is achieved in this analysis with POWHEG [11] and MC@NLO [12] for Pythia and Herwig, respectively. Both of these generators use next-to-leading order calculations for the matrix elements.

As mentioned in Sec.(1) the UE has to be considered as well. For the individual contributions special Monte Carlo tunes are created, usually from data. The ATLAS group has developed a special MPI tune from dijet events at the LHC. It is now interesting to see if this tune is also sensible for Top-Antitop events, which can be at higher scales than dijet-data.

There are also other effects that have to be taken into account, such as color reconnection. Color reconnection describes the interaction of coloured partons during the showering and hadronization. Gluons are emitted and absorbed by the partons which modifies, for example, jet shapes. A summary of all the different effects that play a role can be seen in Fig.(5).

Tuning is needed in the first place because perturbative QCD calculations can only be made in the high energy regime. The MC tunes used in this analysis are the Perugia P2011C tunes [13]. In one of them the Multi Parton Interaction is increased and in the other one the color reconnection is lowered, but not completely turned off.

2.4   Jets at ATLAS

The reconstruction of jets in a given event is a rather difficult procedure and there are multiple algorithms to achieve this. The ATLAS collaboration uses the so called anti-k_t algorithm [15] to reconstruct jets. One parameter of this algorithm is the jet radius. The radius of a jet is defined as:

$$R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}, \tag{2.1}$$

where $\eta = -\ln\left[\tan(\frac{\theta}{2})\right]$ is the pseudorapidity[1] and ϕ the azimuth. At ATLAS, $R = 0.4$ or $R = 0.6$ is used to identify jets. This value is chosen arbitrarily and other experiments like CMS use $R = 0.5$ or $R = 0.7$ for example. In this analysis, only jets of size $R = 0.4$ are used.

---

[1] In the coordinate system used at ATLAS, the beam axis lies along the z-axis and θ is the angle between the particle/jet and the z-axis.

(a) Cluster model　　　　　　　　(b) String model
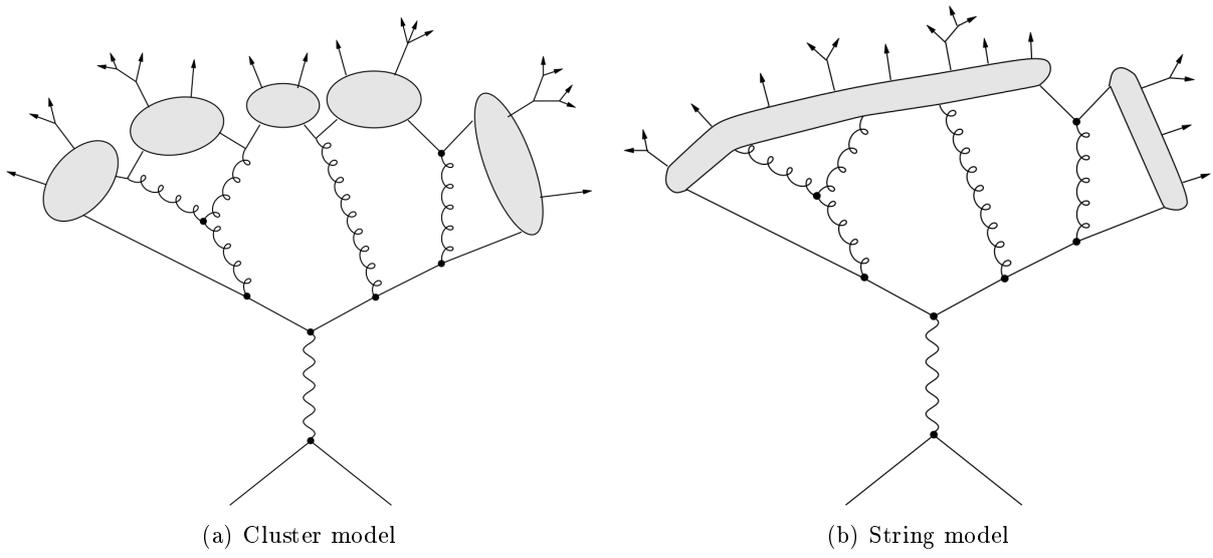
Figure 4: Different hadronization models. [14]
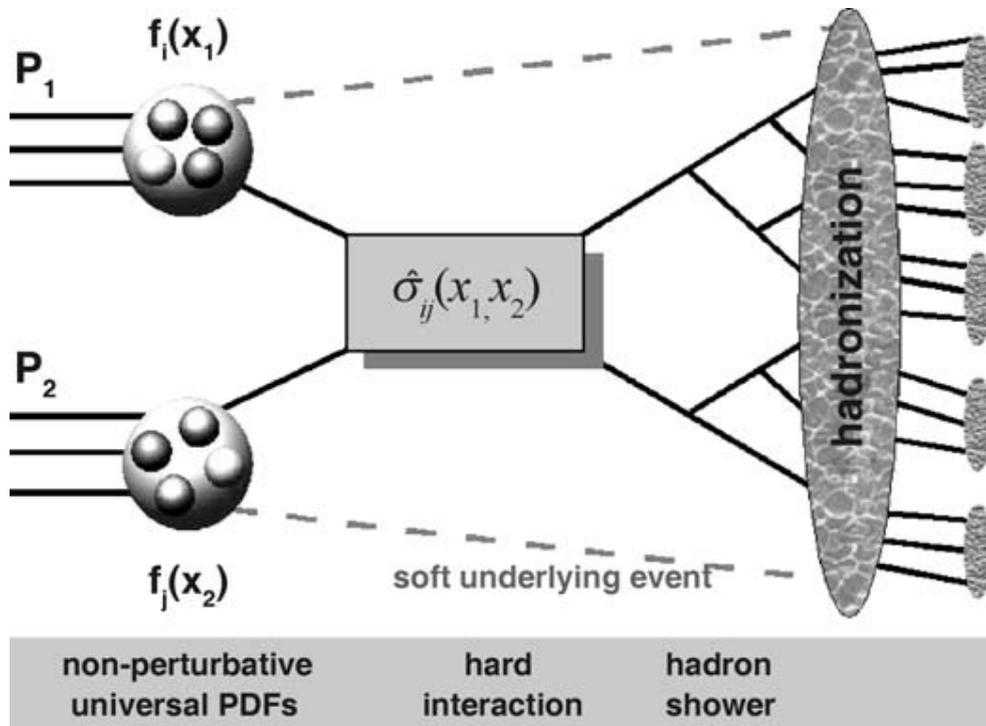
Figure 5: Illustration of parton interaction in a $p\bar{p}$ collision. [3]

Track selection criteria
≥ 6 hits in the SCT
≥ 1 hit in the pixel detector
$p_T \geq 1$
$\eta \neq 0$

Table 2: List of applied track selection criteria.

In this analysis, tracks in the inner detector are used to probe jet shapes. Tracks are required to be well measured, which means that there have to be at least 6 hits in the Semiconductor Tracker (SCT) and at least one hit in the pixel detector. Tracks with a p_T higher than $100\,\text{GeV}$ are required to have a goodness of fit probability > 0.01 to exclude tracks that have not been measured correctly. Additionally a cut on $p_T \geq 1$ is used and tracks below this threshold are not considered. A list of all track selection criteria can be seen in Tab.(2). In this analysis tracks are only considered if they are closer than $\Delta R = 0.6$ to one of the two b-jets.

2.5 Observables

The main observable that is used in this analysis is the averaged sum of the transverse momenta of the tracks inside the jets:

$$\sum_{\text{Tracks}} p_T. \tag{2.2}$$

This observable is used to probe the energy of the jet in two different ways. The first is to define a cone inside the jet-cone and consider all tracks that are inside. The size of this cone is then increased until it matches the jet-cone and even a little further, since jets do not stop at the arbitrary cut of $R = 0.4$. This probes the jet energy as a function of jet radii. A schematic visualization of this procedure can be seen in Fig.(6). The other way that the $\sum_{\text{Tracks}} p_T$ is used is not to increase the size of the cone but to look at the differential jet shape with an annulus starting at the center and then going outwards. Using this approach it is possible to look at
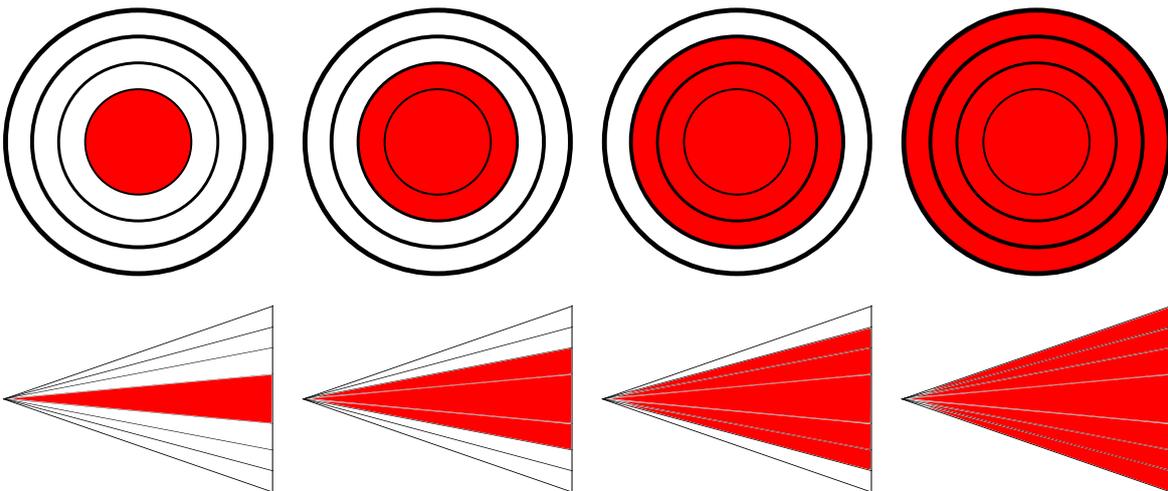
Figure 6: Visualization of the growing cone inside the jet.

the energy distribution inside the jet in more detail. A schematic of this technique is shown in Fig.(7). In addition to these observables based on the individual tracks, the angular distribution
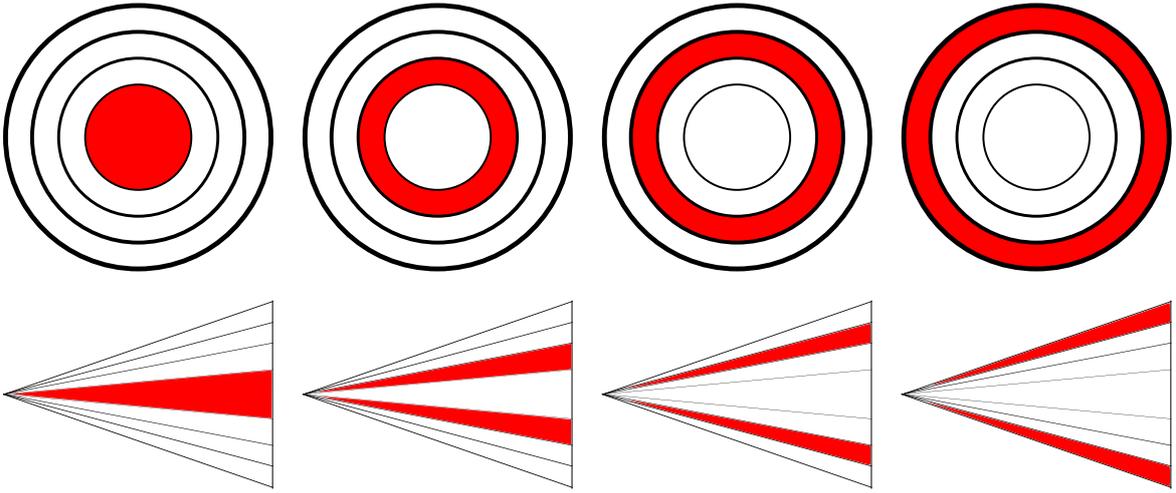
Figure 7: Visualization of the growing annulus inside the jet.

in ϕ-space of the jets is also investigated. This observable is of a particular interest for events with an additional third jet, originating from initial or final state radiation, see Fig.(8). This observable should also be sensitive to the underlying color reconnection in the event as described in [16]. The concrete observables used are $\Delta\phi(b, \bar{b})$, the azimuthal difference between the two b-
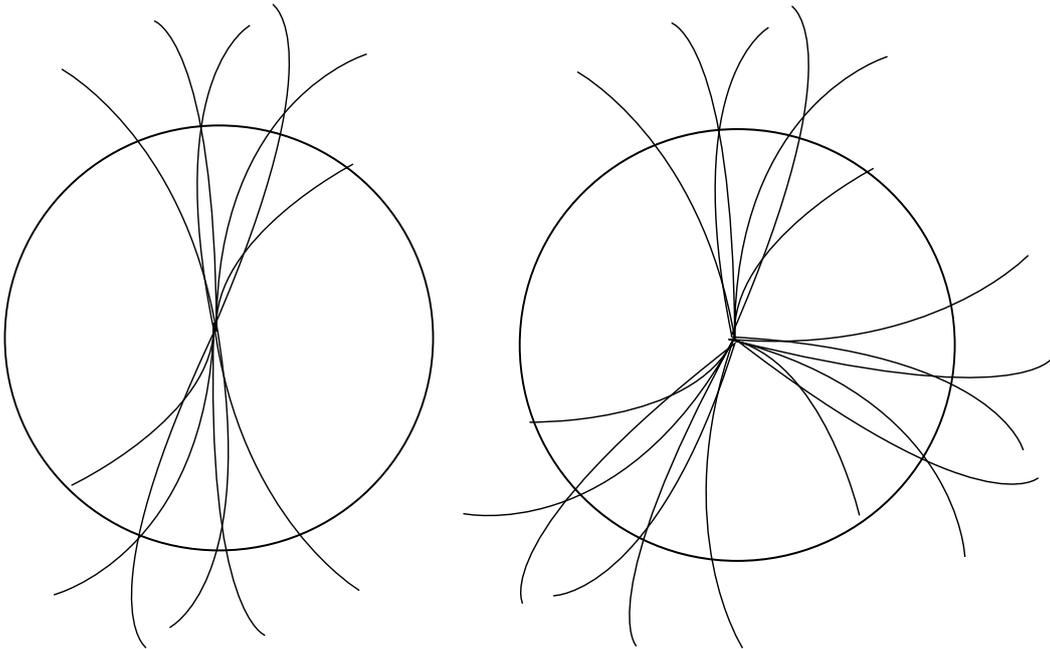
Figure 8: Schematic view on Top-Antitop events with and without an additional third jet.

jets, $\Delta\phi(b/\bar{b}, t\bar{t})$, the azimuthal difference between the leading/sub-leading b-jet and the transverse $t\bar{t}$ system and $\Delta\phi(\text{3rd jet}, t\bar{t})$, the azimuthal difference between the 3rd jet and the transverse $t\bar{t}$ system. In Addition to this definition, a variation with modified jet axes is used as suggested in [16]. The modified jet axis $\bar{\phi}_{\text{Jet}}$ is given as:

$$\bar{\phi}_{\text{Jet}} = \phi_{\text{Jet}} + \frac{1}{n}\sum_{i=1}^{n}\Delta\phi_i \quad \text{with} \quad \Delta\phi_i = \phi_i - \phi_{\text{Jet}}, \tag{2.3}$$

where n is the number of tracks inside the Jet. This weighting of the jet axis by the tracks might be affected by the color reconnection tune and this modification should affect the pull between

Backgrounds
WW + 0/1/2/3 jets
ZZ + 0/1/2/3 jet
ZW + 0/1/2/3 jets
Single top

Table 3: Considered Backgrounds for the analysis

two jets and therefore the distribution in ϕ-space.

3 Analysis

In this section the single steps of the analysis will be explained in detail. The data used here is the $20\,\text{fb}^{-1}$ sample from the ATLAS detector, collected at a center of mass energy of 8 TeV.

3.1 Background Events and Systematic Uncertainties

Although the amount of background is very small because of the chosen selection criteria, separate Monte Carlo samples were used to subtract any remaining background contributions from the data. The backgrounds considered are listed in Tab.(3). These samples were scaled and then subtracted from the data. For the POWHEG + Pythia MC sample systematic uncertainties were looked at. These include jet energy scale (JES), b-jet energy scale (bJES) and trigger efficiencies as well as misstagging. The plots shown in Sec.(3.3) are created with the background subtraction and include the systematic uncertainties.

3.2 MC Tunes

In the first step the different MC tunes are compared with the data. The tunes however are generated using a simple beam-spot model and are not compared directly to data. To make the comparison possible the difference between the tunes and a nominal sample which uses the same simpler beam-spot model is calculated. This difference is then added to a nominal sample with a better beam-spot model (POWHEG+Pythia), which gives two new tunes. These tunes are then used for the comparison. The result for the leading Jet can be seen in Fig.(11). The first thing to notice is that the MC tunes are in good agreement with the data. Looking at the shape of the distribution itself, one can see in the inclusive Plot (Fig.(9(a))) the sum of the momenta is rising, which is expected since more tracks are included in a bigger cone. From the differential plot (Fig.(9(b))) one can see that most of the energy, is around the center of the jet and if one goes further outside the average energy of the tracks decreases. For the sub-leading b-jet and the extra gluon jet the shape is similar but shows small differences. In case of the sub-leading jet one can see that the maximum in the differential distribution is slightly shifted to the right, which means that the high p_T tracks are not as centred as in the leading jet. Additionally one can see that for values $\Delta R > 0.4$ the differential distribution is not flat but starts to rise. The same behaviour can also be seen for the extra jet. Because this rise starts at $\Delta R = 0.4$, which is the ATLAS jet radius, it is likely that the rise is caused by the analysis itself. One explanation is that other jets are too close and tracks of these jets are accidentally counted. To check this hypothesis, an analysis was made with an additional jet removal. All events in which there was a jet closer than $\Delta R = 0.6$ to the sub-leading jet were ignored and only those considered where
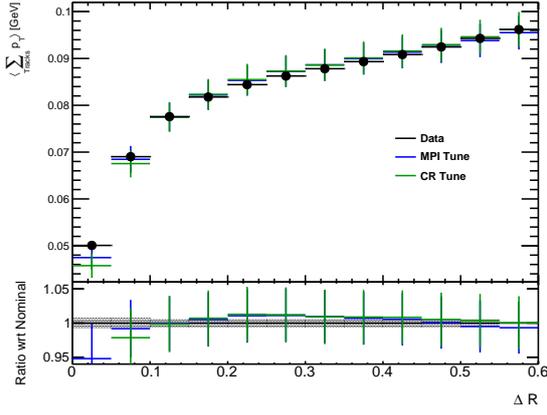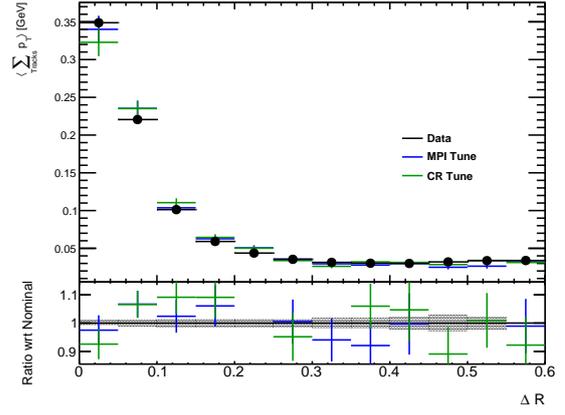
Figure 9: Comparison of Monte Carlo tunes with data for the leading B-Jet.
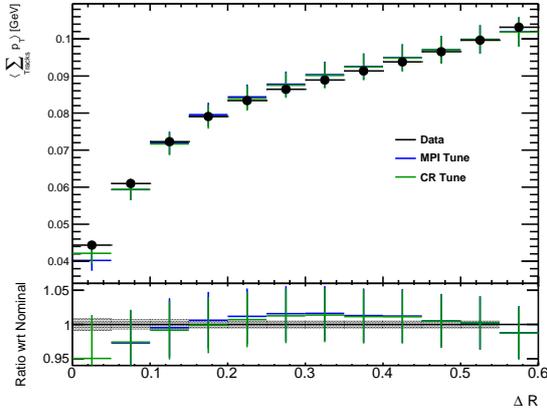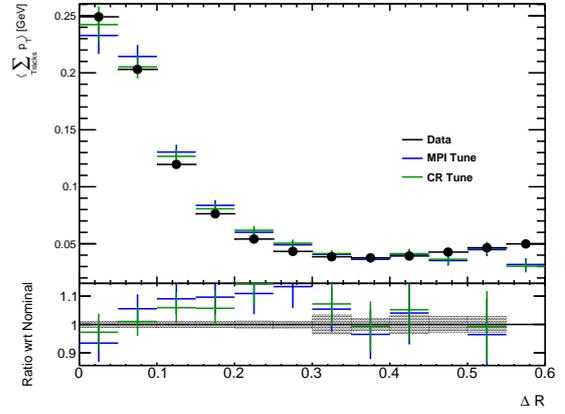
Figure 10: Comparison of Monte Carlo tunes with data for the sub-leading B-Jet.
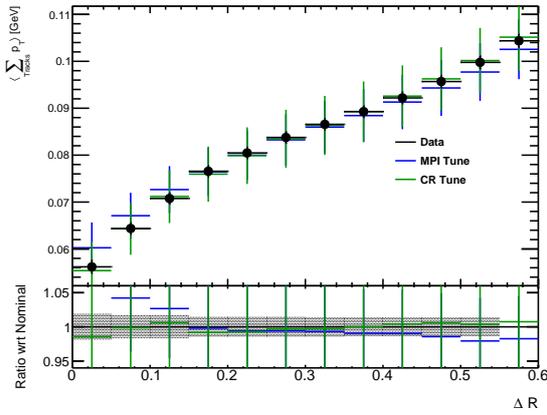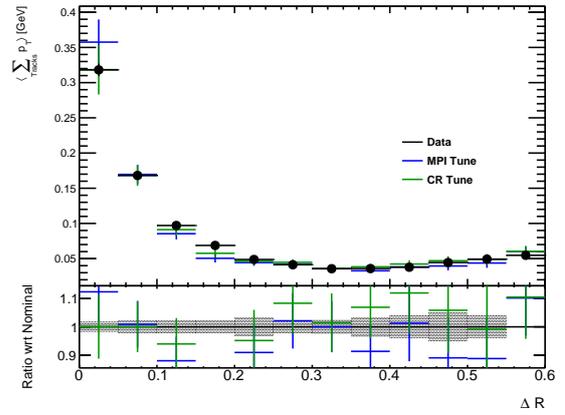
Figure 11: Comparison of Monte Carlo tunes with data for extra jets from gluons (3rd or 4th jet).
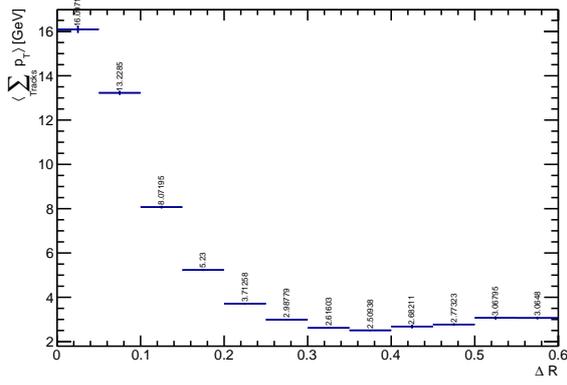
(a) Without jet removal.
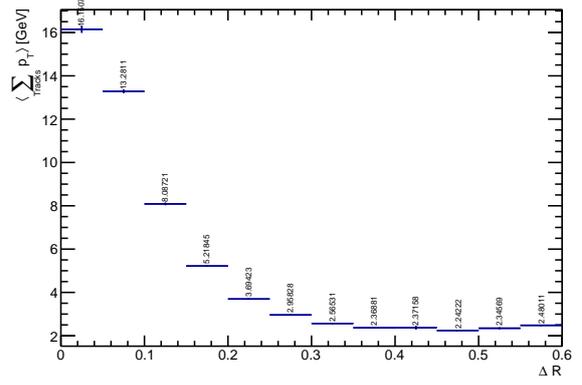
(b) With jet removal.

Figure 12: Comparison of differential $\sum_{\text{Tracks}} p_T$ for the sub-leading b-jet with and without additional jet removal for the region $\Delta R \leq 0.6$ (nominal MC sample).
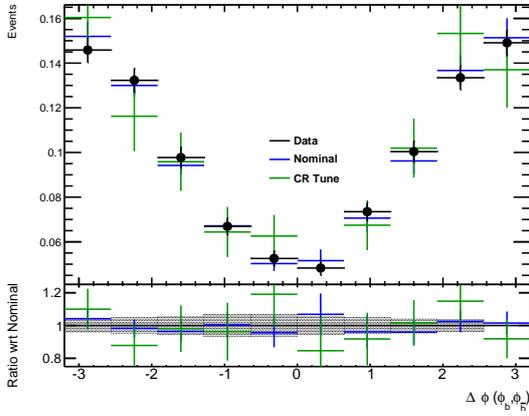
(a) 2-Jet events
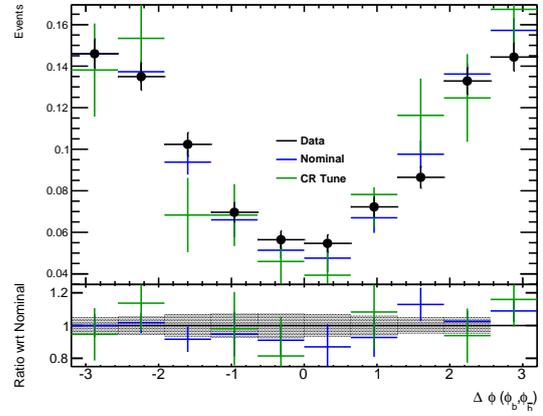
(b) 3-Jet events

Figure 13: $\Delta\phi(b,\bar{b})$ distribution for events with exactly 2 or 3 jets.

the jet was isolated. The result of this analysis can be seen in Fig.(12). The removal of the jets does indeed flatten the shape of the histogram, which supports the theory that tracks from these jets are responsible for the rise seen before.

3.2.1 Phi distribution

The second major variable used in this analysis, the ϕ difference between the two b-jets, is also tested with the MC tunes. The differences between the two b-jets in ϕ-space should be affected especially by the low color reconnection tune. Therefore the nominal MC sample is compared with the low colour reconnection tune and the data. Fig.(13) shows the result when the normal jet axes are used and Fig.(14) shows the result with the weighted jet axes. The rather large uncertainties for the low colour reconnection tune make a comparison more difficult but neither of the plots show a significant difference between the two MC samples. From the plots one can also see that a third jet from a gluon does not change the shape of the distribution significantly, most likely because gluons jets tend to be soft and do not carry enough momentum to change the ϕ-distribution of the two high-p_T b-jets.

(a) 2-Jet events

(b) 3-Jet events
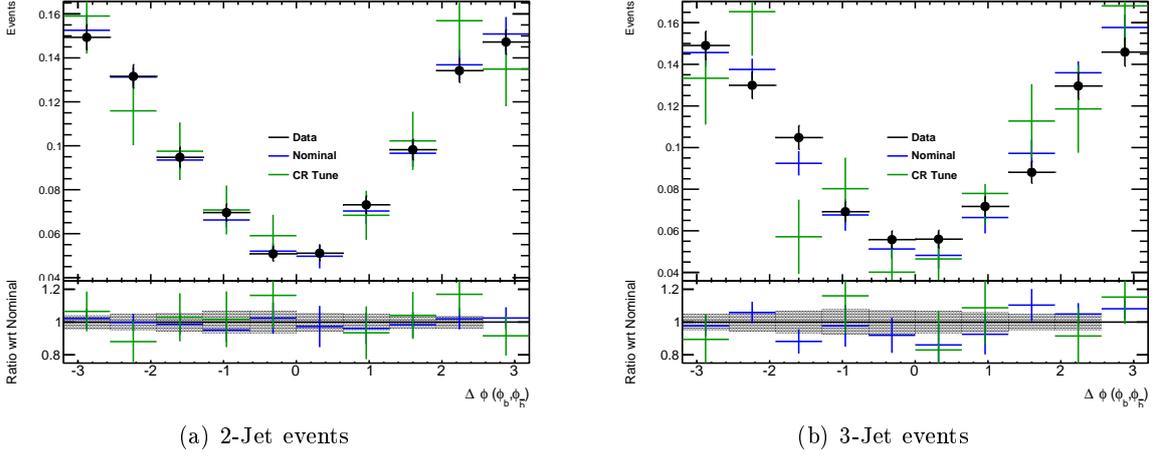
Figure 14: $\Delta\phi(b, \bar{b})$ distribution for events with exactly 2 or 3 jets.

(a) 2-Jet events

(b) 3-Jet events
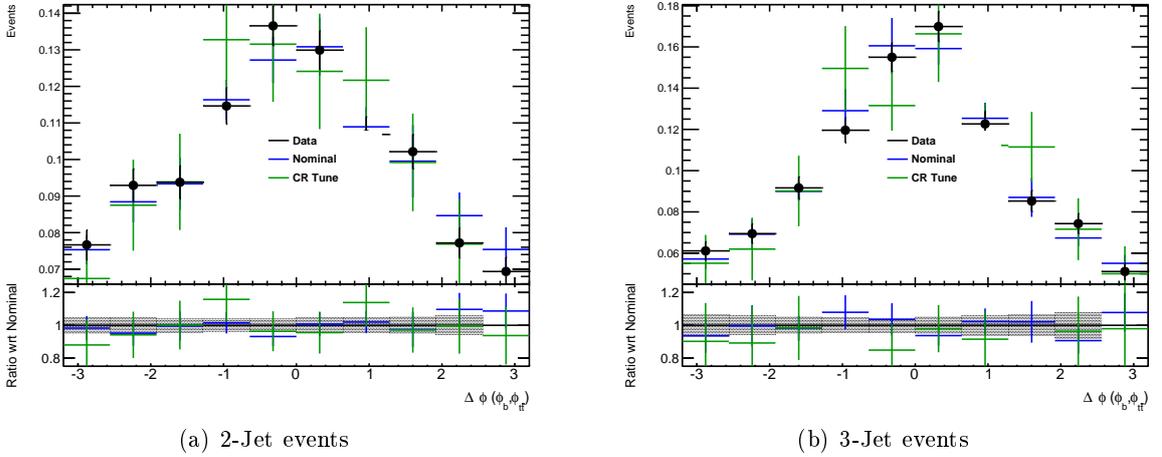
Figure 15: $\Delta\phi(b, t\bar{t})$ distribution for events with exactly 2 or 3 jets.

Aside from the azimuthal difference between the 2 b-jets, also the difference between the leading jet and the transverse $t\bar{t}$ system $\Delta\phi(b, t\bar{t})$ has been investigated. The result is shown in Fig.(15) and Fig.(16). The first thing to notice here is that the leading b-jet prefers to go into the direction of the transverse $t\bar{t}$-system. Again there are no significant changes to see for the low colour reconnection tune. The 2/3 jet distinction also has no visible effect.

3.3 Data/MC agreement

The final part of this analyis focuses on the Data/Monte Carlo agreement of the previously used observables. In addition to the POWHEG+Pythia sample used in the previous steps, a second Monte Carlo sample MC@NLO+Herwig is used here. The motivation behind that being that Herwig seems to have problems modelling the fragmentation in certain event types. The question is if the description of Top-Antitop events is correct. Looking at the results in Fig.(17), Fig.(18) and Fig.(19) one can see that the agreement between the Monte Carlo samples and the data is good, considering the uncertainties. There seem to be some differences in the shapes, but the uncertainties do not allow to say whether this effect is a statistical fluctuation or a problem with the modelling. On the other hand it is very clear that there is no significant difference between POWHEG+Pythia and MC@NLO+Herwig. Neither of them describes the data better than the
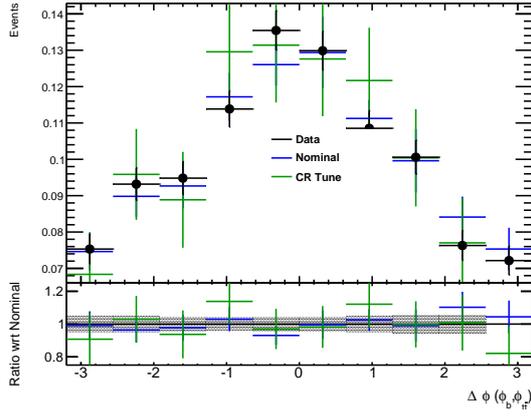
(a) 2-Jet events
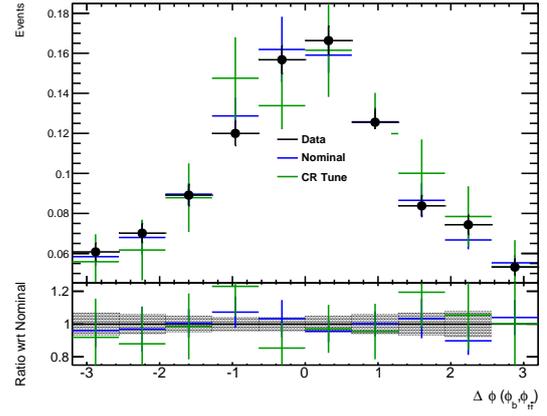
(b) 3-Jet events

Figure 16: $\Delta\phi(b,t\bar{t})$ distribution for events with exactly 2 or 3 jets.
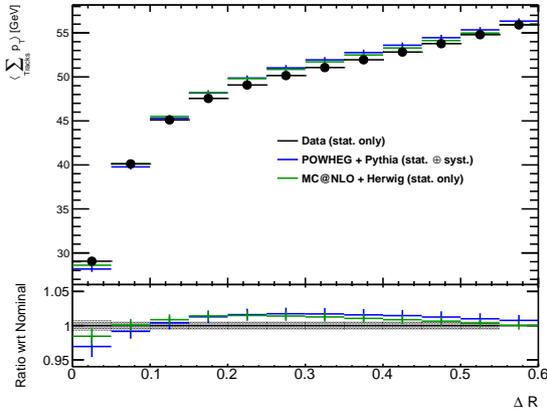
(a) inclusive
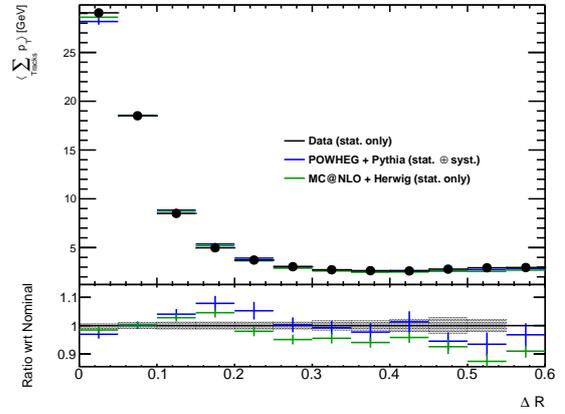
(b) differential

Figure 17: Comparison of Pythia and Herwig samples with ATLAS data for leading b-jet.
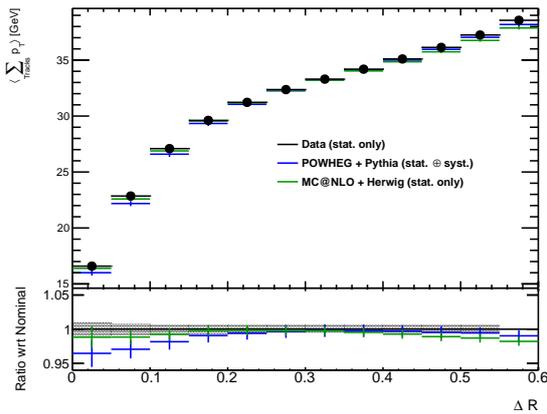
(a) inclusive
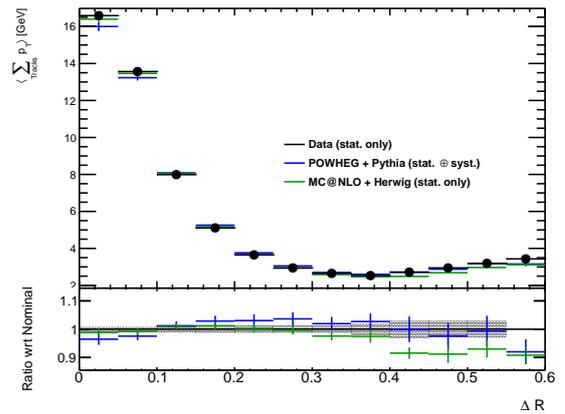
(b) differential

Figure 18: Comparison of Pythia and Herwig samples with ATLAS data for sub-leading b-jet.

other. For all three jets they show a very similar behaviour.
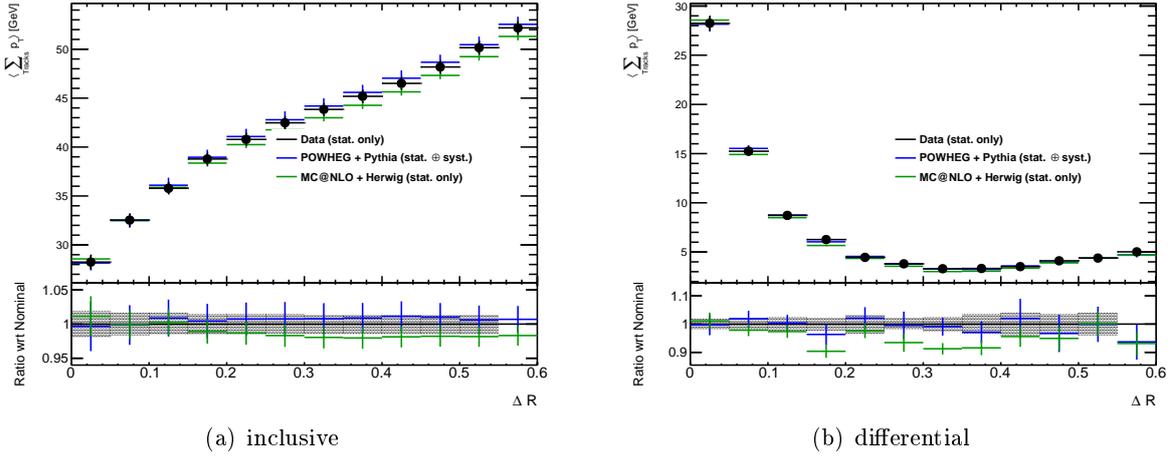
Figure 19: Comparison of Pythia and Herwig samples with ATLAS data for extra jet.

4   Summary & Conclusion

This analysis has shown that, although hadronization is very complex process with many different variables which are not easy to model correctly, the description of it by current Monte Carlo generators is indeed sufficient to describe jet shapes, at least in $t\bar{t}$ events. Further it was shown that the Perugia P2011C tunes are in fact sensible for top pair events, which might not be that surprising but is still a good cross-check.

The ϕ-distribution between the two b-jets, especially with the weighted jet-axes, is expected to have shown some differences in the shapes, but these differences were not visible within the statistic. This was not the main part of this project and was not looked at in more detail. Future analyses could investigate this to see if more statistics reveal a shape difference. It is also interesting that a third jet does not change the distributions significantly.

The comparison of POWHEG+Pythia and MC@NLO+Herwig showed that although there might be cases in which Herwig has problems to model jet parameters correctly, $t\bar{t}$ events are modelled quite well. A very similar analysis has already been performed with the 7 TeV ATLAS data [17]. Fig.(20) shows the result from this analysis. They also compared Pythia and Herwig samples and received similar results. Together with the results from the analysis performed here there seems to be no reason to prefer one generator.
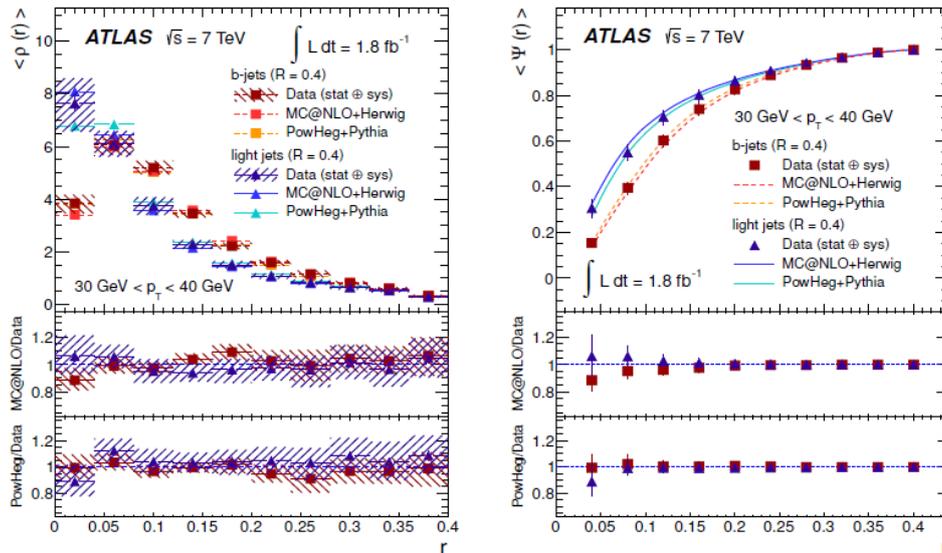
Figure 20: Results of similar analysis by Georges Aad et al. with the 7 TeV dataset. Differential (left) and inclusive (right) analysis. [17]

References

[1] F. Abe et al. Observation of top quark production in $\bar{p}p$ collisions. Phys.Rev.Lett., 74: 2626–2631, 1995. doi: 10.1103/PhysRevLett.74.2626.

[2] S. Abachi et al. Observation of the top quark. Phys.Rev.Lett., 74:2632–2637, 1995. doi: 10.1103/PhysRevLett.74.2632.

[3] Arnulf Quadt. Top quark physics at hadron colliders. The European Physical Journal C-Particles and Fields, 48(3):835–1000, 2006.

[4] J. Beringer et al. (Particle Data Group). The review of particle physics. Phys. Rev. D, 86: 010001, (2012) and 2013 partial update for the 2014 edition.

[5] Ann Heinson. Useful diagrams of top signals and backgrounds, October 2011. URL http://www-d0.fnal.gov/Run2Physics/top/top_public_web_pages/top_feynman_diagrams.html.

[6] 1204180. Commissioning of the ATLAS high-performance b-tagging algorithms in the 7 TeV collision data. 2011.

[7] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. JINST, 3: S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003.

[8] Lyndon Evans and Philip Bryant. Lhc machine. Journal of Instrumentation, 3(08):S08001, 2008.

[9] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. PYTHIA 6.4 Physics and Manual. JHEP, 0605:026, 2006. doi: 10.1088/1126-6708/2006/05/026.

[10] G. Corcella, I.G. Knowles, G. Marchesini, S. Moretti, K. Odagiri, et al. HERWIG 6: An Event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). JHEP, 0101:010, 2001. doi: 10.1088/1126-6708/2001/01/010.

[11] Stefano Frixione, Paolo Nason, and Carlo Oleari. Matching NLO QCD computations with Parton Shower simulations: the POWHEG method. JHEP, 0711:070, 2007. doi: 10.1088/1126-6708/2007/11/070.

[12] Stefano Frixione and Bryan R. Webber. Matching NLO QCD computations and parton shower simulations. JHEP, 0206:029, 2002. doi: 10.1088/1126-6708/2002/06/029.

[13] Peter Zeiler Skands. Tuning Monte Carlo Generators: The Perugia Tunes. Phys.Rev., D82: 074018, 2010. doi: 10.1103/PhysRevD.82.074018.

[14] B. Webber. Parton shower Monte Carlo event generators. Scholarpedia, 6(12):10662, 2011. revision #128236.

[15] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The Anti-k(t) jet clustering algorithm. JHEP, 0804:063, 2008. doi: 10.1088/1126-6708/2008/04/063.

[16] Spyros Argyropoulos and Torbjörn Sjöstrand. Effects of color reconnection on $t\bar{t}$ final states at the lhc. arXiv preprint arXiv:1407.6653, 2014.

[17] Georges Aad et al. Measurement of jet shapes in top-quark pair events at $\sqrt{s} = 7$ tev using the atlas detector. Eur.Phys.J., C73:2676, 2013. doi: 10.1140/epjc/s10052-013-2676-3.