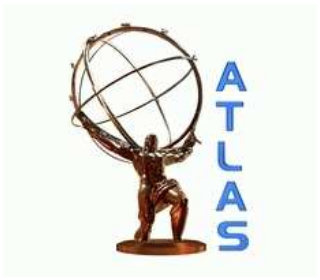
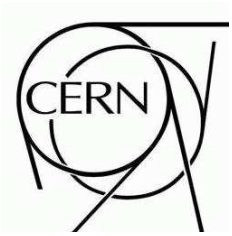


ATLAS NOTE

ATL-COM-PHYS-2008-001

February 4, 2008



High p^T Hadronic Top Quark Identification
Part I: Jet Mass and YSplitter

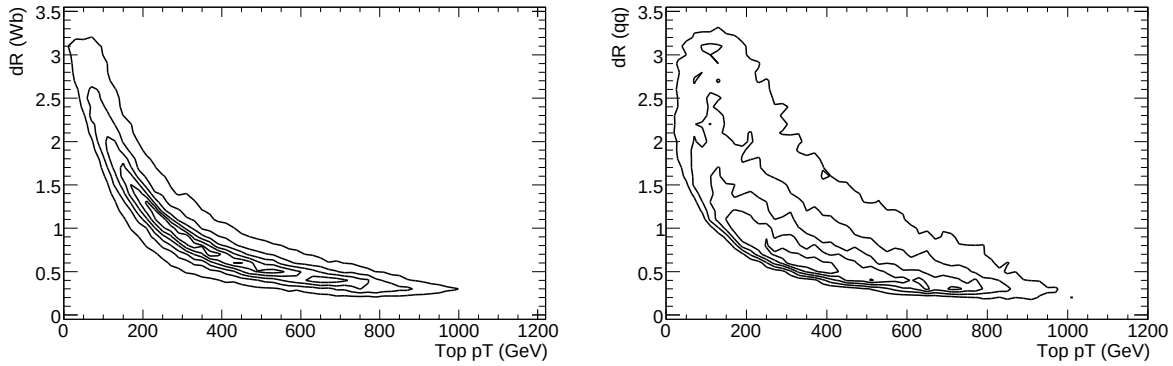
Gustaaf Brooijmans
Columbia University

Abstract

At the LHC objects with masses at the electroweak scale will for the first time be produced with very large transverse momenta. In many cases, these objects decay hadronically, producing a set of collimated jets. This interesting new experimental phenomenology requires the development and tuning of new tools, since the usual reconstruction methods would simply reconstruct a single jet. This note describes the application of the YSplitter algorithm in conjunction with the jet mass to identify high transverse momentum top quarks decaying hadronically.

1 Introduction

At the LHC, top quarks, W and Z bosons are relatively *light* and can be produced with very high transverse momenta with respect to their masses. In the case of hadronic decays, the quarks can be so close together in the detector that they are in principle reconstructed as a single jet. This is illustrated in Figure 1 which shows generator-level distributions of angular distance between the b quark and W boson, and quarks from W boson decays in $t\bar{t}$ events. Here $dR = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ with ϕ the azimuthal angle and η the rapidity. For top quark transverse momenta larger than about 200 GeV, the distance between the decay products is often smaller than twice the typical jet radius. Since these events were generated using PYTHIA, no top polarization effects are included. Such effects are very model dependent and therefore beyond the scope of this generic study.



(a) (b)

Figure 1: Angular distances between decay products in top quark decays as a function of top quark transverse momentum: (a) between the b quark and W boson, and (b) between quarks from W boson decays.

Identifying top quarks at high transverse momentum with high efficiency is of particular interest in searches for new physics. In addition to a number of recent theoretical models specifically proposing the existence of high mass resonances decaying dominantly to top quarks (see for example [1] [2] [3]), the large top quark mass suggests it might be closely linked to forms of new physics that would manifest themselves at very high energies. It is therefore quite probable that new heavy objects decay to top quarks at least some fraction of the time, if not exclusively.

The samples used for this study are presented in section 2, the tools in section 3 and the results in section 4.

2 Monte Carlo Data Sets

Five datasets are used in this study, all at ESD level for full access to calorimeter cells. They are:

- `trig1_misal_mc12.006233.pythia_Zprime2000_tt.recon.ESD.v12000601` (20k events),
- `trig1_misal1_mc12.006234.pythia_Zprime3000_tt.recon.ESD.v12000601` (15k events),
- `trig1_misal1_csc11_V1.005014.J5_pythia_jetjet.recon.ESD.v12000604` (221k events),
- `trig1_misal1_csc11_V2.005015.J6_pythia_jetjet.recon.ESD.v12000604` (239k events), and

- `trig1_misal1_csc11.005016.J7_pythia_jetjet.recon.ESD.v12000604` (162k events).

The first two are signal samples consisting of standard model-like M = 2, 3 TeV Z' bosons forced to decay to top-antitop pairs, with one of the resulting W bosons forced to decay leptonically (to an electron or muon) and the other hadronically. The three background samples are multijet samples with 280 GeV $< \hat{p}_T <$ 560 GeV ($\sigma = 12.5$ nb), 560 GeV $< \hat{p}_T <$ 1120 GeV ($\sigma = 0.35$ nb) and 1120 GeV $< \hat{p}_T <$ 2240 GeV ($\sigma = 6$ pb) respectively. Samples for the next (and last) bin, 2240 GeV $< \hat{p}_T$, were not available at the time of this writing, and are of very limited use for this study which has top quarks with p_T values up to approximately 1500 GeV. All these samples were generated with PYTHIA [4], then (fully) simulated and reconstructed as part of the official CSC effort. The cross-sections are those given by PYTHIA.

3 Tools

The scope of this first note is limited to the usage of the jet mass, which is the invariant mass of all the jet's constituents (typically cells or towers), and YSplitter [5], which determines the scales at which jets can be resolved into two or more subjets. The jet mass is already stored with the rest of the information produced during the standard reconstruction, but this is not true for the results of YSplitter.

The jet constituents are therefore accessed from the ESD to run the YSplitter algorithm. Since the jet constituents are not calibrated, it is necessary to first go back to cell-level, apply the calibration[1)], and rebuild the towers. This gives acceptable results: the obtained towers are reclustered using the kt algorithm and the energy ratio between the "old" and "new" jet is measured. This ratio is shown as a function of jet transverse momentum in Figure 2. The deviation from 1 can probably mainly be assigned

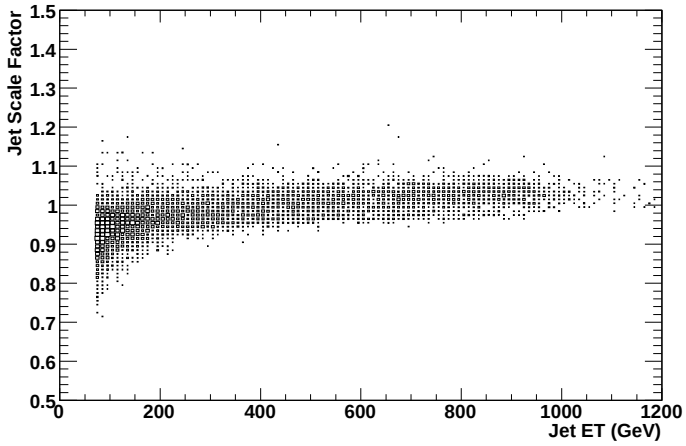


Figure 2: Ratio of original jet energy over jet energy after reclustering calibrated towers as a function of jet transverse momentum.

to the non-application of cryostat correction factors. To mitigate this, all energy-related measurements used are re-scaled with this scale factor. It should be noted that in the future the YSplitter information will be stored along with the jet so that these "gymnastics" are not necessary.

[1)] The *H1WeightToolCSC12Kt6Tower* cellcalibrator option was used for the Kt6TowerJets considered

4 Results

4.1 Signal Characteristics

The purpose of this study is discrimination between hadronically decaying top quarks and jets originating from a single hard parton, and the focus is therefore on the jets in the event. Figure 3 shows the angular distance of each jet with $p_T(jet) > 30$ GeV to the closest top quark and the hadronically decaying W boson. Since the top quark is highly boosted, its direction is highly correlated with the W boson direction as expected. To consider the case where all three jets from a hadronic top decay are merged, events are

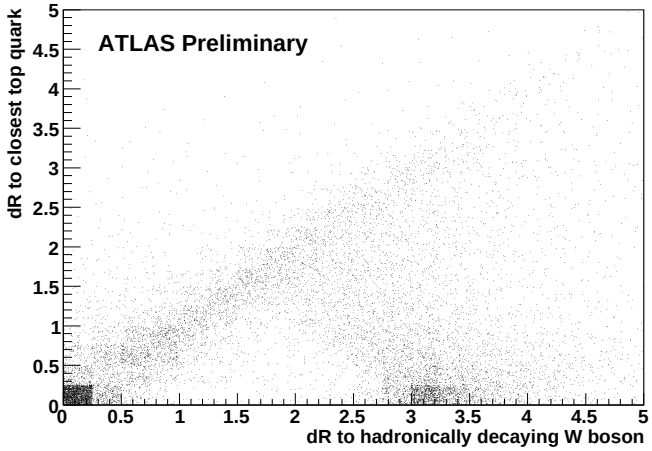
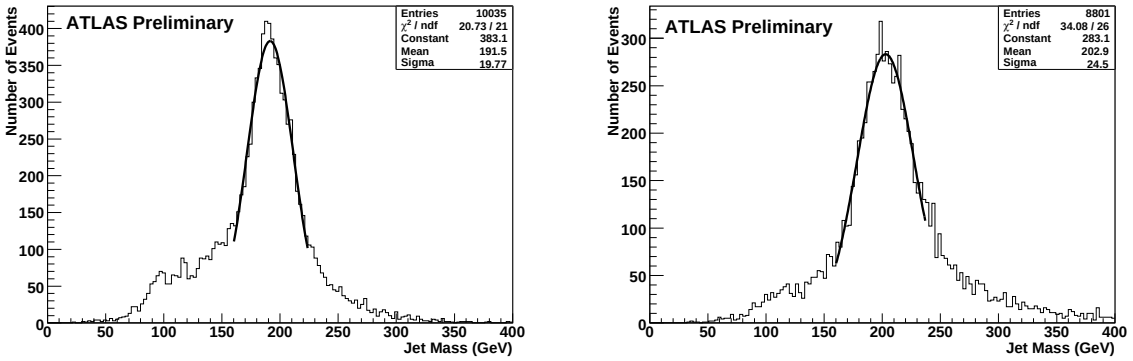


Figure 3: Angular distance between each jet with $p_T(jet) > 30$ GeV and the closest top quark and hadronically decaying W boson.

selected in which only a single reconstructed jet with $p_T(jet) > x$ GeV has $dR < 1.0$ from the closest top quark and $dR < 2.0$ from the hadronically decaying W boson. The number of events surviving this cut is sensitive to the choice of p_T cut (i.e. x), but the peak in the jet mass distribution for the single jet, shown in Figure 4 isn't. The low mass tail is suppressed when the threshold is lowered, and enhanced as it is raised. Between 20 and 30 GeV the ratio between the number of entries at the peak and around 100 GeV is relatively stable, suggesting that this is the transition region between jets from soft radiation and direct top quark/W boson decay quarks. In the following, $x = 20$ is chosen. The jet p_T distribution for the selected "top monojets" is shown in Figure 5. It displays the expected jacobian-like behavior with an endpoint around 1 and 1.5 TeV for the two signal samples respectively. We require $p_T(jet) > 300$ GeV since below that the decay products are rarely sufficiently collimated to be reconstructed as a single jet, and analysis cuts would suppress the few remaining cases where this happens. For completeness, Figure 6 shows the jet mass as a function of the jet transverse momentum for the top "monojets" in both samples.

The "YScale" values at which the top monojet splits into two, three and four jets are illustrated in Figure 7 for jets with $p_T(jet) > 300$ GeV which pass the dR cuts. Overall the distributions are very similar, with a somewhat slower turn-on for the higher p_T jets. This shows that as for jet mass, these variables can be used to identify hadronic top quark decays over a large transverse momentum spectrum. For the first two YScale values, the distributions are centered around 90 and 40 GeV respectively, corresponding to approximately $m(top)/2$ and $m(W)/2$ as expected. The correlation between the scales for splitting off a second and a third, and a third and fourth jet are also shown and seen to be relatively weak, implying additional rejection can be obtained.



(a) (b)

Figure 4: Jet mass distribution for reconstructed jets with no other jet close to a top quark and hadronically decaying W boson (see text). In this example, jets were required to have $p_T(jet) > 20$ GeV to be considered for the veto. (a) Jets in the M = 2 TeV Z' sample and (b) jets in the M = 3 TeV Z' sample.

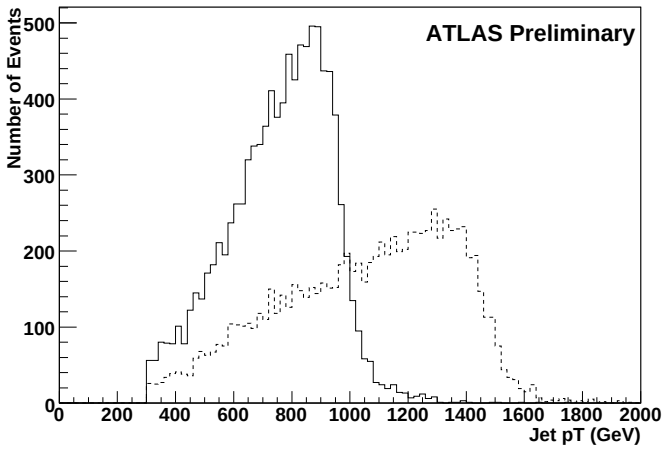


Figure 5: Jet p_T distribution for reconstructed jets with no other jet close to a top quark and hadronically decaying W boson (see text). The full (dashed) line corresponds to jets in the M = 2 (3) TeV Z' sample.

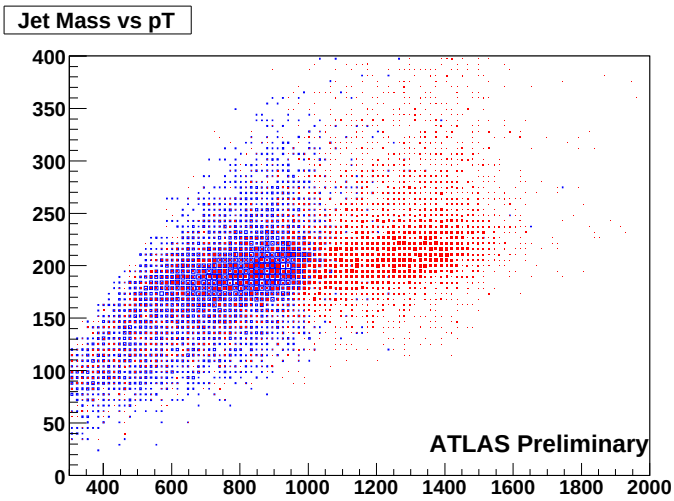


Figure 6: Jet mass as a function of transverse momentum for jets passing the top monojet selection: jets from the M = 2 (3) TeV Z' sample in blue/open (red/filled).

The correlation between YScale values and jet mass for the first three splitting values is shown in Figure 8.

4.2 Background Characteristics

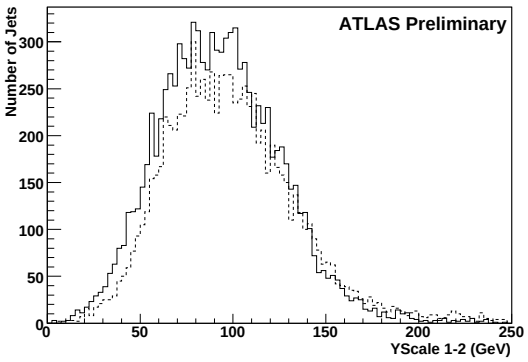
The 5014, 5015 and 5016 samples are combined according to their relative cross-sections to obtain a smooth distribution in jet transverse momentum distribution as shown in Figure 9. Events from sample 5015 get unit weight, so that the equivalent integrated luminosity of the sample is approximately 700 pb^{-1}. To ensure that generator-level selection cuts do not bias the results, only jets with 300 GeV $< p_T(jet) <$ 2200 GeV will be used at the analysis stage. This covers the onset of the monojet signature and includes the region where it is dominant.

The distributions of jet mass as well as jet mass as a function of jet transverse momentum for the multijet background are shown in Figure 10. The YScale values for these jets, the correlation between the splitting values for two and three, and three and four jets, and the correlation between the splitting values and jet mass are shown in Figure 11. From these plots it is clear that the jet mass and YScale distributions drop exponentially for jets originated by light quarks and gluons. The combination of these variables should therefore yield a very effective discrimination between jets originating from top quarks and those from light quarks or gluons. We quantify this in the next section.

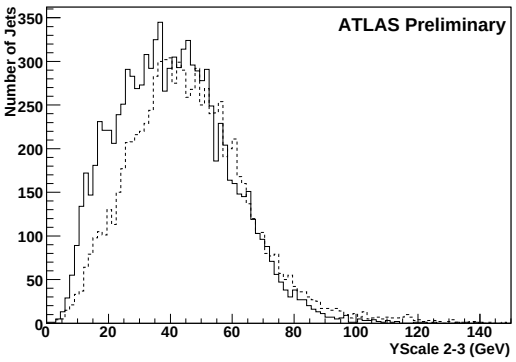
4.3 Quantitative Analysis

From this point on, jets are required to have 300 GeV $< p_T(jet) <$ 2200 GeV to avoid any biases originating from phase space selection at the event generation level. The aim of this study is not to study the particular signal used but rather to provide an **estimate** of the discriminating power between top and light jets of the variables described above. For this purpose this kinematic region is adequate given the relatively small variations in the signal distributions between the two signal samples.

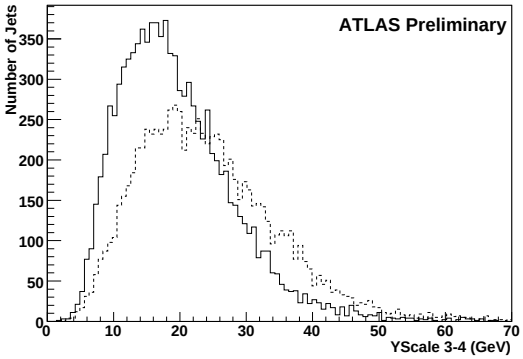
Given the number of variables and the correlations, a multivariate method is likely to obtain the best results. For clarity, however, we choose to use two methods: a set of "square" cuts, i.e. cuts applied to individual variables, and explicit two-variable cuts. We keep in mind that an additional rejection of a



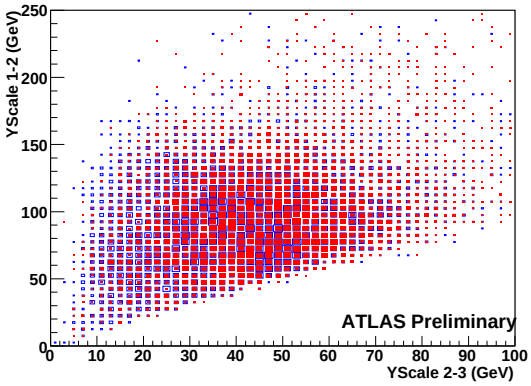
(a)



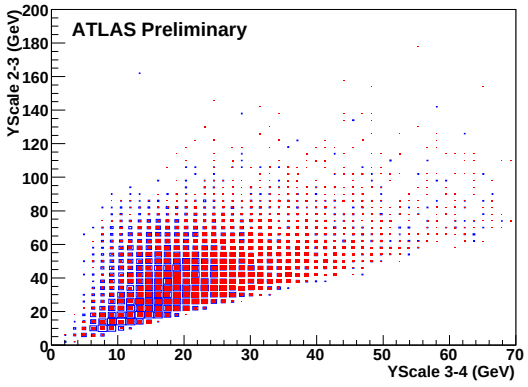
(b)



(c)

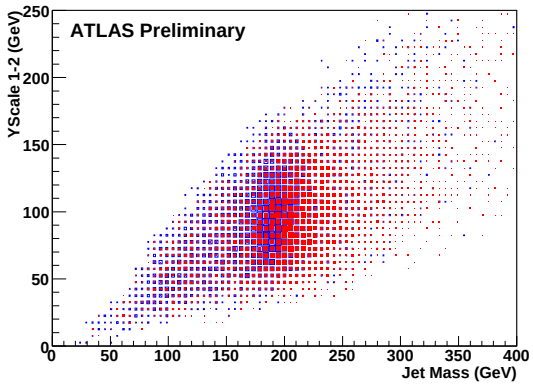


(d)

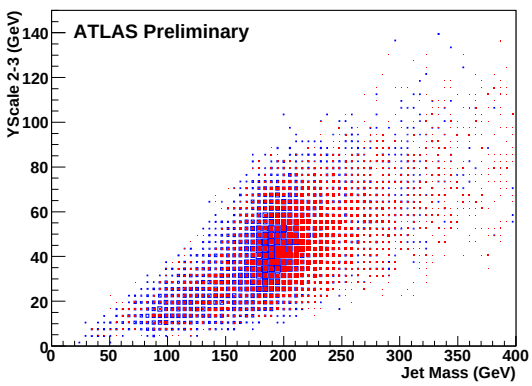


(e)

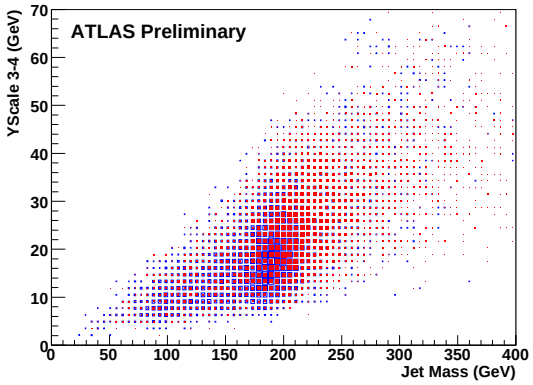
Figure 7: YScale values at which the selected top monojets split into (a) two, (b) three, and (c) four jets. Jets from the M = 2 (3) TeV Z' samples are drawn as a solid (dashed) line. The correlation between the splitting scale into two and three jets is shown in (d), and three and four jets in (e). Here the jets from the M = 2 (3) TeV Z' samples are drawn in blue/open (red/filled).



(a)



(b)



(c)

Figure 8: Correlation between YScale values and jet mass for splitting into (a) two jets, (b) three jets and (c) four jets. The jets from the M = 2 (3) TeV Z' samples are drawn in blue/open (red/filled).

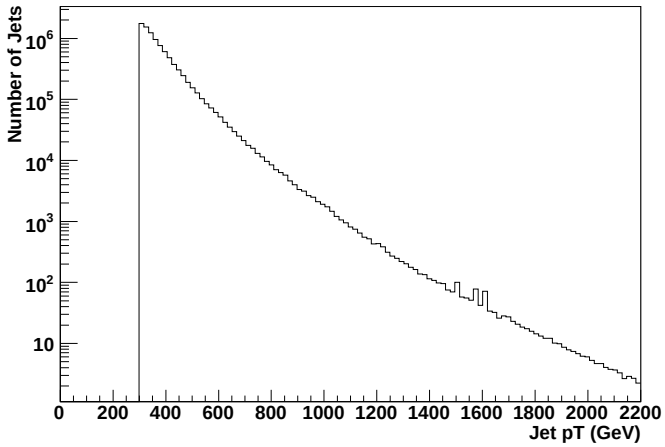
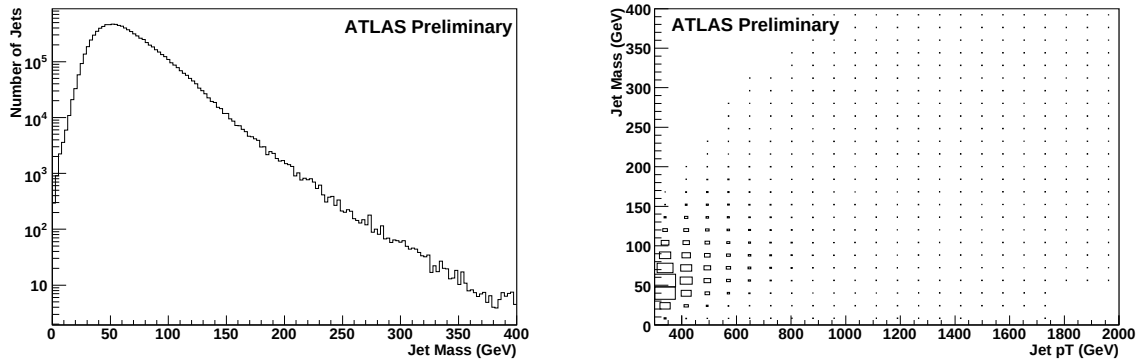


Figure 9: Jet p_T distribution for reconstructed jets in the combined background samples.



(a) (b)

Figure 10: (a) Jet mass and (b) jet mass versus p_T distributions for reconstructed jets in the multijet sample.

Jet p_T (GeV)	500	600	700	800	900	1000	1100	1200	1300	1400	1500
Top (%)	20.4	46	58	64	69	63	70	64	64	60	52
Background (%)	1.2	4.0	7.2	8.7	10.3	13.8	13.4	12.9	14.9	19.1	15.9

Table 1: Selection efficiencies for various jet transverse momenta for jets reconstructed close to top quarks in the signal samples and in the background samples using square cuts.

factor close to 1.5 is typical when using neural networks or similar methods. Results will be presented in terms of efficiency as a function of jet transverse momentum.

4.3.1 One-Dimensional Cuts

For this simple approach, cuts are made on jet mass, $YScale_{12}$, $YScale_{23}$ and $YScale_{34}$. The requirements imposed are a) 170 GeV $< mass(jet <$ 250 GeV, b) 50 GeV $< YScale_{12}(jet) <$ 150 GeV, c) 10 GeV $< YScale_{23}(jet) <$ 70 GeV, and d) 6 GeV $< YScale_{34}(jet)$. The resulting efficiencies are given in table 1.

4.3.2 Two-Dimensional Cuts

Here, the jet mass is required to be $>$ 170 GeV, and then cuts are made in the two-dimensional planes defined by jet mass as a function of jet transverse momentum, different YScale values, and different YScale values as a function of jet mass. The cuts are described in table 2 and illustrated in Figures 12-18.

The resulting selection efficiency for top monojets (as defined in section 4.1) in the signal samples and jets in the background samples as a function of jet transverse momentum are shown in Figure 19. Numerical values at various transverse momenta are given in Table 3. It is important to state that the cuts were not optimized for a particular region in phase space, but rather "by eye" to achieve a maximal efficiency for top monojets while keeping the light jet efficiency below 10%. A factor of approximately three improvement is obtained with respect to the "square cuts" case.

Note that a specific physics analysis will optimize the cuts for the relevant region in phase space after consideration of the other objects in the event useful in the rejection of backgrounds. This could be a

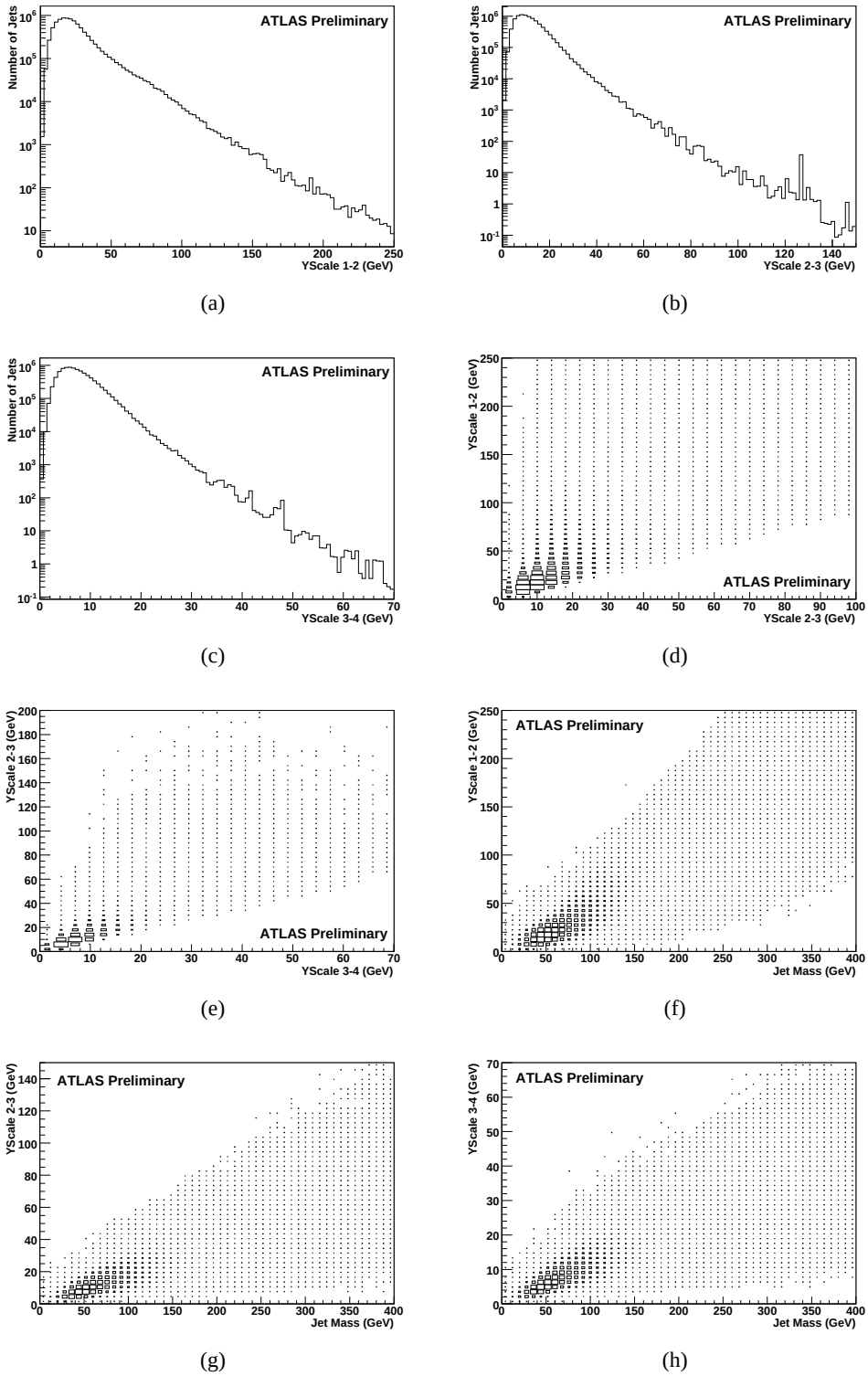


Figure 11: YScale values at which reconstructed jets in the multijet sample split into (a) two, (b) three, and (c) four jets. The correlation between the splitting scale into two and three, and three and four jets is shown in (d) and (e) respectively. (f), (g) and (h) depict the correlation between YScale values and jet mass.

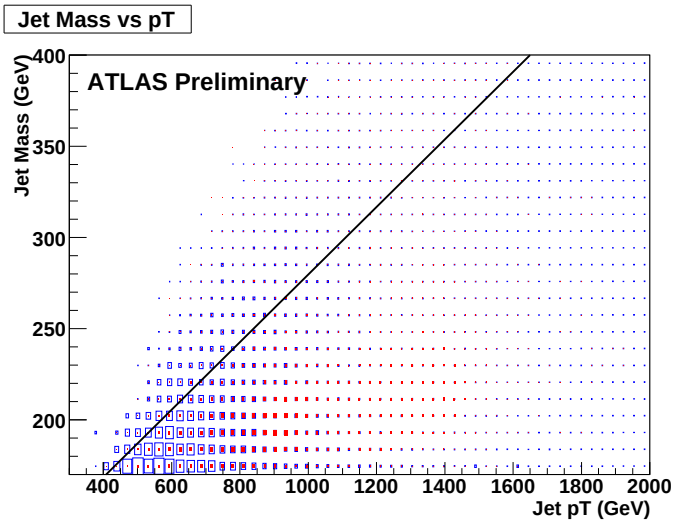


Figure 12: Distribution of jet mass as a function of jet transverse momentum for the background (blue/open) and signal (red/filled) samples. Events are required to lie below the line.

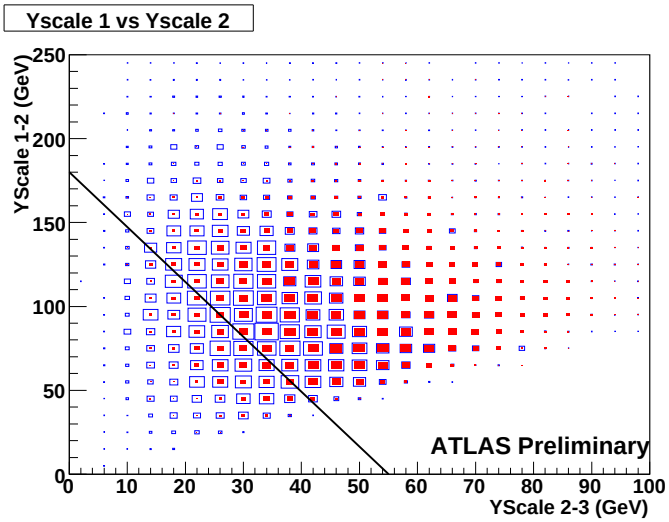


Figure 13: Distribution of $YScale_{12}$ as a function of $YScale_{23}$ for the background (blue/open) and signal (red/filled) samples. Events are required to lie above the line.

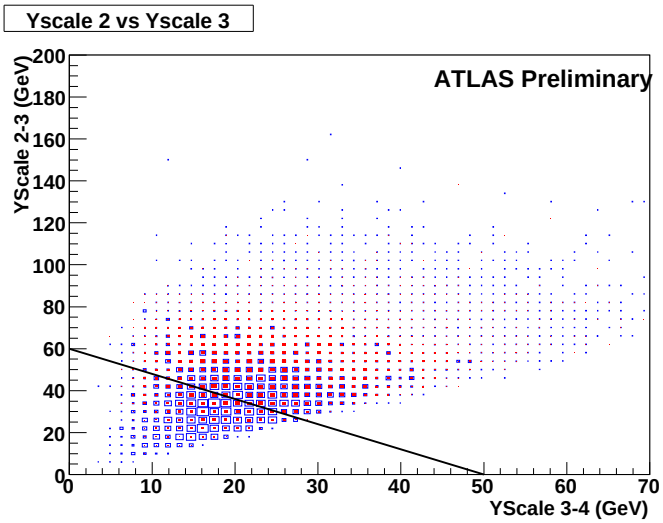


Figure 14: Distribution of $YScale_{23}$ as a function of $YScale_{34}$ for the background (blue/open) and signal (red/filled) samples. Events are required to lie above the line.

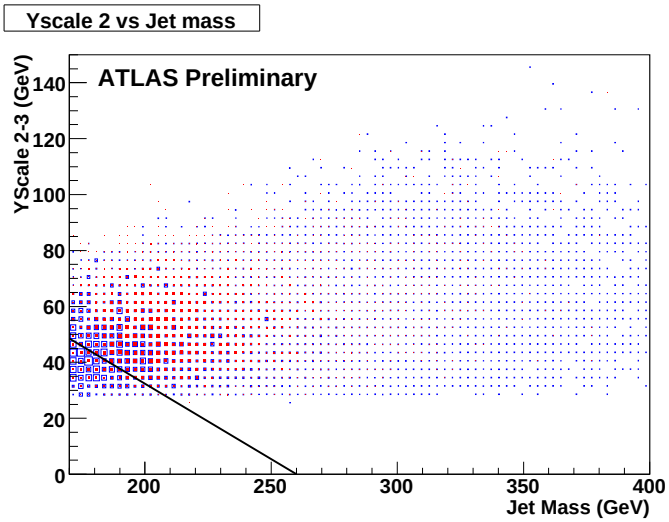


Figure 15: Distribution of $YScale_{23}$ as a function of jet mass for the background (blue/open) and signal (red/filled) samples. Events are required to lie above the line.

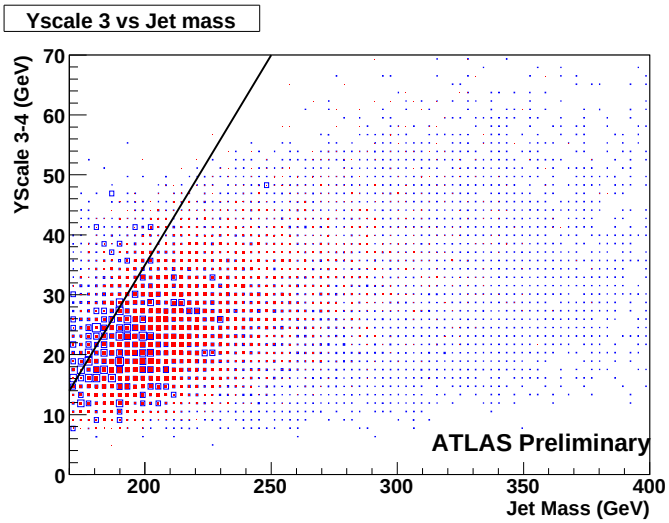


Figure 16: Distribution of $YScale_{34}$ as a function of jet mass for the background (blue/open) and signal (red/filled) samples. Events are required to lie below the line.

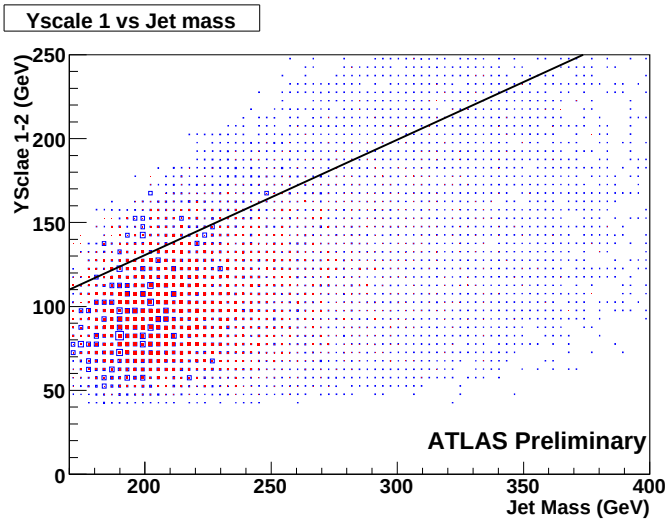


Figure 17: Distribution of $YScale_{12}$ as a function of jet mass for the background (blue/open) and signal (red/filled) samples. Events are required to lie below the line.

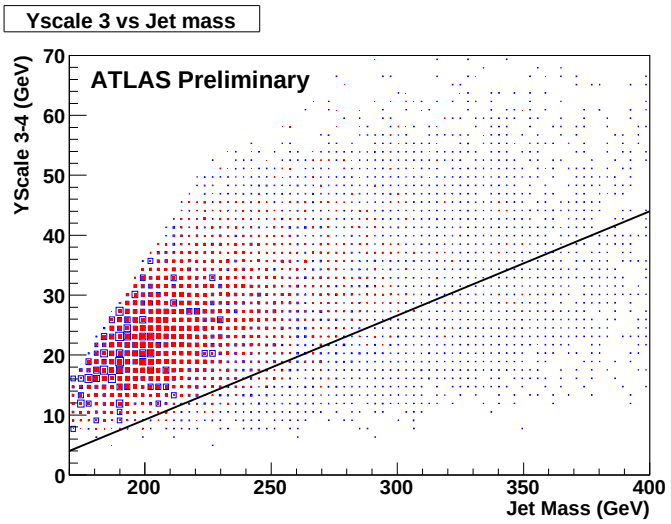


Figure 18: Distribution of $YScale_{34}$ as a function of jet mass for the background (blue/open) and signal (red/filled) samples. Events are required to lie above the line.

Cut variables	Cut	Figure
Jet mass vs jet p^T	$mass(jet) < (5/27)p_T(jet) + (850/9)$	12
YScale 1-2 vs YScale 2-3	$YScale_{12}(jet) > (-180/55)YScale_{23}(jet) + 180$	13
YScale 2-3 vs YScale 3-4	$YScale_{23}(jet) > (-6/5)YScale_{34}(jet) + 60$	14
YScale 2-3 vs jet mass	$YScale_{23}(jet) > (-7/13)mass(jet) + 140$	15
YScale 3-4 vs jet mass	$YScale_{23}(jet) < 0.7mass(jet) - 105$	16
YScale 1-2 vs jet mass	$YScale_{12}(jet) < (11/16)mass(jet) - (55/8)$	17
YScale 3-4 vs jet mass	$YScale_{34}(jet) > (4/23)mass(jet) - (588/23)$	18

Table 2: Two dimensional cuts applied.

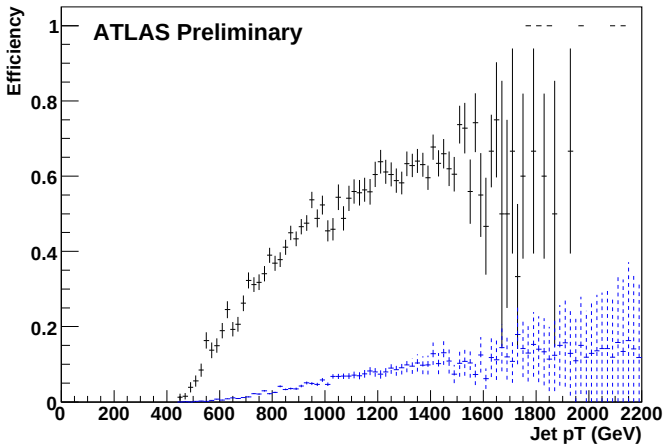


Figure 19: Selection efficiency for jets close to top quarks (solid, black) and jets in the background samples (blue, dashed) as a function of reconstructed jet transverse momentum.

Jet p_T (GeV)	500	600	700	800	900	1000	1100	1200	1300	1400	1500
Top (%)	5.6	19	32	37	47	45	56	64	63	68	74
Background (%)	0.1	0.5	1.3	2.5	4.2	4.7	7.1	7.4	9.8	12.8	10.2

Table 3: Selection efficiencies for various jet transverse momenta for jets reconstructed close to top quarks in the signal samples and in the background samples using two-dimensional cuts.

lepton from the semileptonic decay of the other top quarks in $Z' \to t\bar{t}$ events for example. The usage of a multivariate tool will presumably lead to further improvement in the result.

5 Modeling Uncertainties

The study described in this note is completely based on events generated and hadronized with PYTHIA, run through the ATLAS GEANT-based full simulation program and reconstructed with the ATLAS reconstruction. The variables used could be sensitive to the choices made. Specific issues are the choice of hadronization model and differences between data and simulation in jet mass and YSplitter variables.

- The effect of the hadronization model on many jet variables has been studied extensively at the Tevatron by comparing PYTHIA and HERWIG [6] to real data. The ensuing systematic uncertainties are typically of the order of a few percent at most.
- Jet mass is known to be sensitive to cell-level thresholds for inclusion and noise. This is partially due to the fact that lower energy cells on the outer edge of a jet can still make a significant contribution to the jet mass.
- The YSplitter output could be sensitive to these as well, although to a lesser extent since it really involves the core of the jet. The sensitivity will presumably be worse as the splitting number increases.

Further simulation studies can and will be made in the near future to investigate these issues. More important, however, is the development of a strategy to calibrate and/or verify the performance of these variables in real data. The hadronization parameters in PYTHIA and HERWIG are tuned to LEP and Tevatron data, and we expect further work in this direction at the LHC. Comparisons with data involve variables like the track density and the production rate of various light mesonic and baryonic resonances, and this should be feasible at the LHC as well.

For the jet mass and YSplitter variables, the main difficulty is in the identification of a relatively clean calibration sample. Luckily, it will be possible in ATLAS to isolate a relatively pure sample of top quarks pair events in the lepton plus jets channel without the need to impose drastic cuts [7]. By using a jet algorithm with a very large radius, we will be able to artificially create "top (and W) monojets" and understand the performance and sensitivity of these algorithms. A simulation study of this procedure will be performed as well.

6 Other Systematic Effects

While YSplitter is an intrinsic aspect of kT jet algorithms, it is of course possible to do the initial jet selection with cone jets. In both cases, the effect of algorithmic parameters like the jet radius need to be studied. Jet energy resolution and scale will also potentially lead to large systematic uncertainties since the distributions for light jets are all exponential. These will all be studied in the near future.

Other systematic effects related for example to acceptance effects are model-dependent and therefore analysis-specific.

7 Conclusion

In this study, the measured jet mass and jet splitting scales have been used to distinguish high transverse momentum "top monojets" from jets originating from light quarks. The combination of algorithms allows for good separation of signal and background, with the ratio of selection efficiencies for signal and background evolving from approximately 30 for jets with $p_T = 600$ GeV to 10 for 1000 GeV and 7 for 1500 GeV. Further work using subjets and tracking information is underway.

8 Acknowledgements

This work has been performed within the ATLAS Collaboration, and we thank collaboration members for helpful discussions. We have also made use of the physics analysis framework and tools which are the result of collaboration-wide efforts.

References

[1] K. Agashe, A. Belyaev, T. Krupovnickas, G. Perez, and J. Virzi, `hep-ph/0612015`.

[2] S. Matsumoto, M. M. Nojiri, and D. Nomura, *Phys. Rev.* **D75** (2007) 055006, [`hep-ph/0612249`].

[3] A. L. Fitzpatrick, J. Kaplan, L. Randall, and L.-T. Wang, *JHEP* **09** (2007) 013, [`hep-ph/0701150`].

[4] T. Sjostrand *et. al.*, *Comput. Phys. Commun.* **135** (2001) 238–259, [`hep-ph/0010017`].

[5] J. Butterworth *et. al.*, `ATL-COM-PHYS-2007-077`.

[6] G. Corcella *et. al.*, `hep-ph/0210213`.

[7] S. Bentvelsen, M. Cobal, *et. al.*, `ATL-CSC-T6, to be published in the ATLAS CSC Book`.