PROCEEDINGS OF SCIENCE

$t\bar{t}$ backgrounds in charged Higgs boson searches

Martin Flechl*
Universität Freiburg
E-mail: martin.flechl@cern.ch

Standard-Model top-quark pair production is the main background for searches in most charged Higgs boson channels, in particular if the charged Higgs boson itself is produced in $t\bar{t}$ decays, i.e. $t\bar{t} \to bWbH^+$ with e.g. $H^+ \to \tau\nu$. The characteristics of this background are discussed, and the results from ATLAS top quark searches with early LHC data are presented. In the light of charged Higgs boson searches, the different $t\bar{t}$ decay modes can be separated into irreducible $t\bar{t}$ modes with the same final state as the signal, and reducible modes which contribute e.g. if an electron is misreconstructed as a tau. Techniques to suppress and estimate the $t\bar{t}$ background contribution to the charged Higgs boson searches are introduced.

Third International Workshop on Prospects for Charged Higgs Discovery at Colliders - CHARGED2010,
September 27-30, 2010
Uppsala Sweden

*Speaker.

PoS(Charged 2010)014

1. Introduction

Top quark pair production ($t\bar{t}$) is the main background for most charged Higgs boson (H^+) searches as the final state can be similar or even identical (see Figure 1). To establish (or exclude) a signal, it is thus essential to have a detailed knowledge of this background: total and differential cross sections, kinematics, topology, as well as trigger and reconstruction efficiencies in the high-multiplicity environment characteristic of $t\bar{t}$ events. An important step towards understanding these properties are the efforts towards top quark observation, described below.

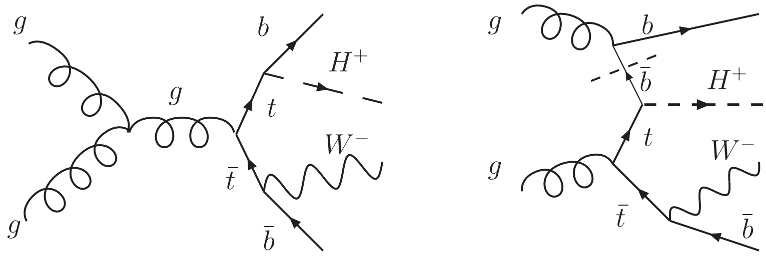

Figure 1: Charged Higgs boson production modes. Left: Top quark decays. Right: gg/gb fusion.

To estimate the $t\bar{t}$ background to H^+ boson searches, different methods can be applied depending on whether the background is irreducible (same final state as the signal) or reducible (some objects are misreconstructed, e.g. a quark-initiated jet reconstrued as a τ jet). This note introduces a selection of such methods. The studies described in the following have been performed with events using a simulated ATLAS detector [1] response or collision data events collected by the ATLAS experiment. The ATLAS detector is a general-purpose detector covering nearly the entire solid angle around the collision point, consisting of a tracking detector, calorimeter, and muon system.

2. Characteristics of top quark events

The expected $t\bar{t}$ production cross section at the LHC is about 160 (870) pb at $\sqrt{s} = 7$ (14) TeV [2]. This means that in one year at low luminosity (collecting 10 fb^{-1}), the LHC experiments will see the creation of about 20 million top quark pairs, three orders of magnitude more than the Tevatron. The top quark mass has been measured by the Tevatron experiments to be $m_t = 173.3 \pm 1.1$ GeV [3]. The dominant production mode at the LHC is through $gg \to t\bar{t}$ while the channel dominant at the Tevatron, $q\bar{q} \to t\bar{t}$, only contributes a few percent of the total cross section. In the Standard Model (SM), the top quark decays almost exclusively to Wb and thus top quark pair events are characterized by the decay modes of the W bosons:

- dilepton (e/μ + e/μ), 7%,
- semi-leptonic (e/μ + qq'), 35%,
- fully hadronic (qq' + $q''q'''$), 44%,
- hadronic tau, (τ + X), 14%.

PoS(Charged 2010)014

3. ATLAS $t\bar{t}$ searches with early data

The ATLAS early-data searches focus on the dilepton and the lepton+jets channel, as the cleanest detector signal (largest signal-to-background ratio) is expected for these decay modes.

3.1 $t\bar{t}$ dilepton searches

$t\bar{t}$ candidate events in the dilepton mode must pass the following requirements [4][5]:

- pass lepton trigger,
- have exactly 2 leptons with $p_T > 20$ GeV and opposite charge,
- have at least 2 jets with $p_T > 20$ GeV,
- in the ee ($\mu\mu$) channel, $E_T^{miss} > 40$ (> 30) GeV,
- in the ee ($\mu\mu$) channel, $|m_{ll} - m_Z| > 5$ (> 10) GeV,
- in the $e\mu$ channel, E_T^{sum}(leptons, jets)> 150 GeV.

Figure 2 shows the distributions of the number of jets and b-tagged jets after the $e\mu$ event selection. An excellent signal-to-background ratio can be observed, in particular when requiring one or more b-tagged jets.

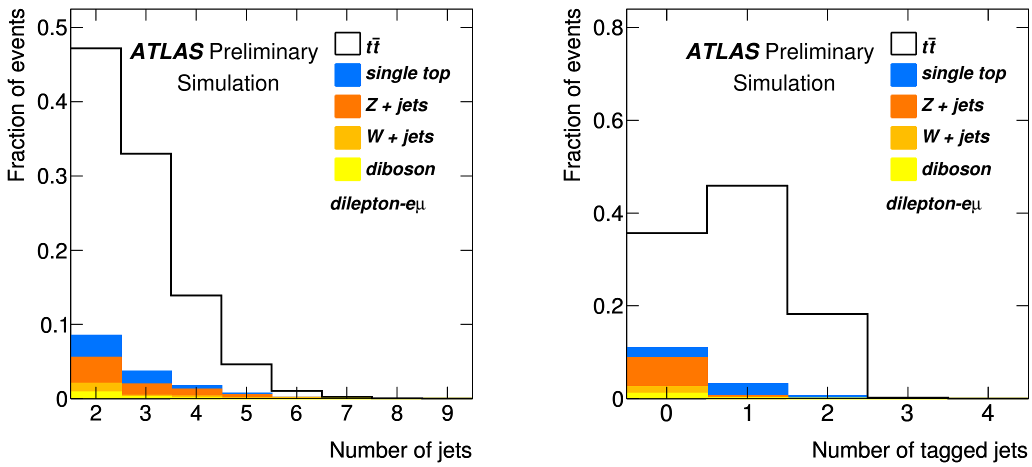

Figure 2: Left: Expected distribution of the number of jets for events passing the $e\mu$ dilepton event selection. Right: Expected distribution of the number of b-tagged jets for events passing the $e\mu$ dilepton event selection.

3.2 $t\bar{t}$ lepton+jets searches

$t\bar{t}$ candidate events in the lepton+jets mode must pass the following requirements [4][5]:

- pass lepton trigger,
- have exactly one lepton with $p_T > 20$ GeV,
- have at least 4 jets with $p_T > 20$ GeV,
- have at least one b-tag (secondary vertex, 50% efficiency),
- have $E_T^{miss} > 20$ GeV.

In Figure 3, collision data results and expectation are compared in number-of-jets bins, both with and without b-tagging requirements. The $t\bar{t}$, single top, W+jets, and Z+jets backgrounds are taken from simulation while the QCD multi-jet background is estimated using data [5]. Even in this first data set (295 nb^{-1}) a clear excess over the non-$t\bar{t}$ expectation (albeit with limited statistical significance) can be observed for high jet multiplicity and when requiring b-tags, hinting at what can be expected from the full ATLAS 2010 data set which is more than two orders of magnitude larger. Expected and observed data agree within uncertainties.

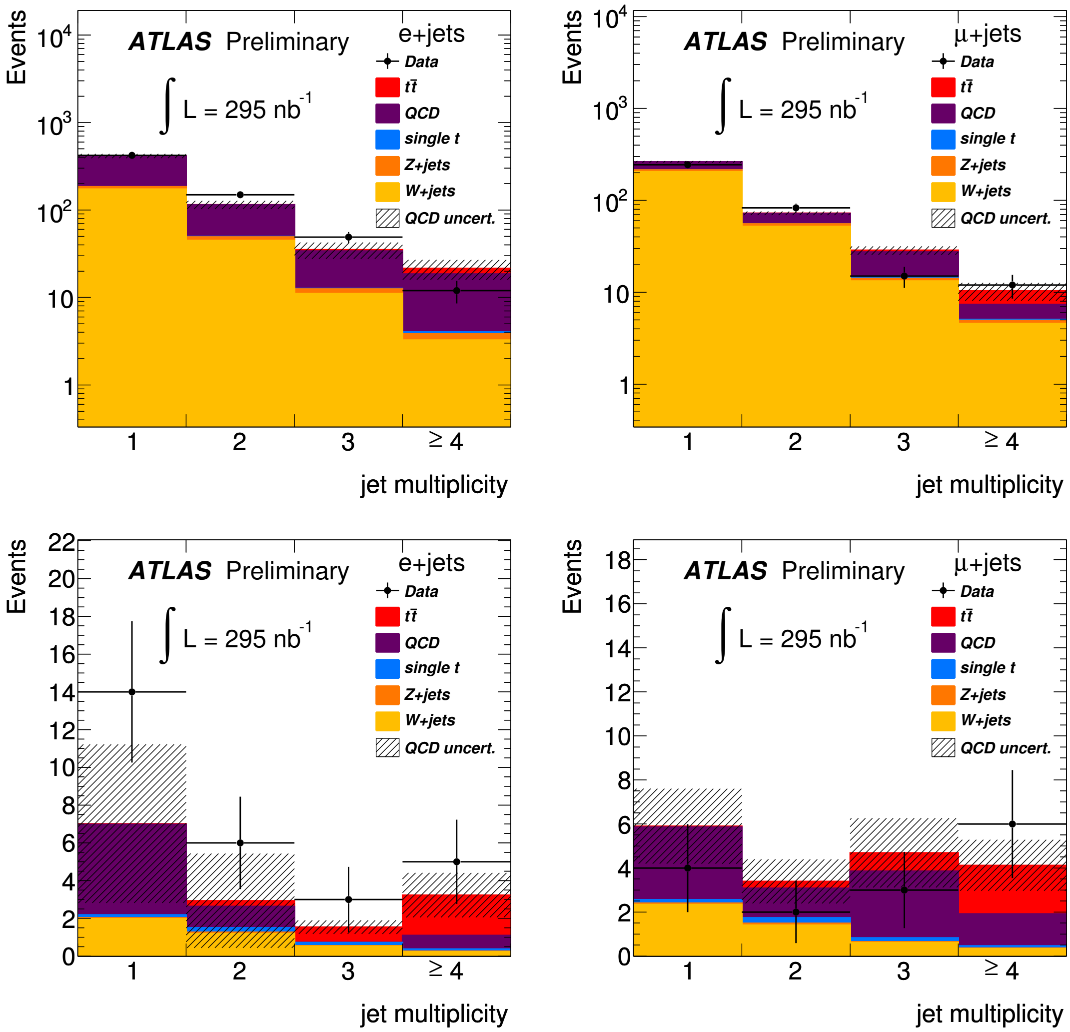

Figure 3: Jet multiplicity distribution for the lepton+jets selection. The data are shown by the points with error bars, compared to the sum of all expected contributions ($t\bar{t}$, single top, W and Z+jets, QCD multi-jet). The hatched area shows the uncertainty on the total expectation due to the statistical error on the QCD background estimate. Left: e+jets, right: μ+jets. Top: no b-tag requirements, bottom: at least one b-tag.

4. Data-driven $t\bar{t}$ background estimation

Previous studies have shown that SM $t\bar{t}$ production is the dominant background for most H^+

searches. This applies to channels with hadronically decaying tau leptons [6] as well as decay modes with leptonic and cs final states [7]. Typically, the SM $t\bar{t}$ background is at least 1-2 orders of magnitude larger than all other SM backgrounds. It has been shown [8] that not only the irreducible but also the reducible $t\bar{t}$ background contributes significantly, in particular via events in which QCD jets are misreconstructed as τ jets. This is important as many data-driven background estimation techniques are only sensitive to one of these two.

The SM $t\bar{t}$ background estimation for H^+ searches has large uncertainties, in total about 15-40% (depending on the search channel) [6]. The dominant systematic uncertainties on the theoretical side come from the $t\bar{t}$ production cross section and the Monte Carlo generator and detector models, and on the experimental side from the jet and τ energy scale, the b-tagging efficiency, and the luminosity measurement. In order to retain any H^+ sensitivity, it is thus necessary to estimate the SM $t\bar{t}$ background from data. Two methods to do so are described below.

4.1 Embedding

The basic principle of the embedding method is to select events with muons in collision data, replace the muon with a simulated tau lepton and then use these events for background estimation. This approach is taken as muons can be identified in a much more efficient and purer way than τ jets, and because typically it is easier to select a signal-free control sample with muons.

To illustrate the method, the procedure to estimate the background from $t\bar{t} \rightarrow \tau$+jets is detailed - however, the method can be used for any kind of events with τ final states (including leptonic τ decays) as long as a sufficiently pure muon sample can be selected from data. Other examples are e.g. $Z \rightarrow \tau\tau$ and $W \rightarrow \tau\nu$ events.

1. Select a pure sample of $t\bar{t} \rightarrow b\mu\nu bqq$ from data.
2. Remove the muon signature from the event (tracks, calorimeter cell depositions).
3. Embed a *simulated* τ lepton carrying the (rescaled) muon four-momentum.
4. Use this hybrid $t\bar{t} \rightarrow b\tau\nu bqq$ for background estimation instead of simulation.

To this end, one can either only take the shape of distributions like $m_T(H^+_{cand})$ from the embedded events, or use them also for normalization, after correcting for trigger and reconstruction efficiencies. The advantage of using these events is that everything except for the τ decay products is taken from collision data, e.g. the underlying event, pile-up, jets, E_T^{miss}, and b-tagging. However, the method is technically complex and can only model one $t\bar{t}$ decay mode at a time. Figure 4 investigates the performance of this method in simulated events, by comparing shapes of embedded and reference events. A good agreement within an error of about 5-10% can be observed.

4.2 Sideband-based methods

Here, the idea is to obtain the data-to-simulation event ratio for a given background from a sideband in which this background is enhanced; and to then apply it to correct the background as expected from simulation in the signal region. In ATLAS H^+ dilepton sensitivity studies [7], requiring $\cos\theta_l^* > -0.4$ is used to enhance the background in the SM $t\bar{t}$ sideband. This quantity is related to the angle between the b quark and the lepton coming from the top quark on the H^+ candidate side (for details, see Reference [7]). As is shown in Figure 5, the method performs well in closure tests and successfully estimates the background scale factors within statistical uncertainties.

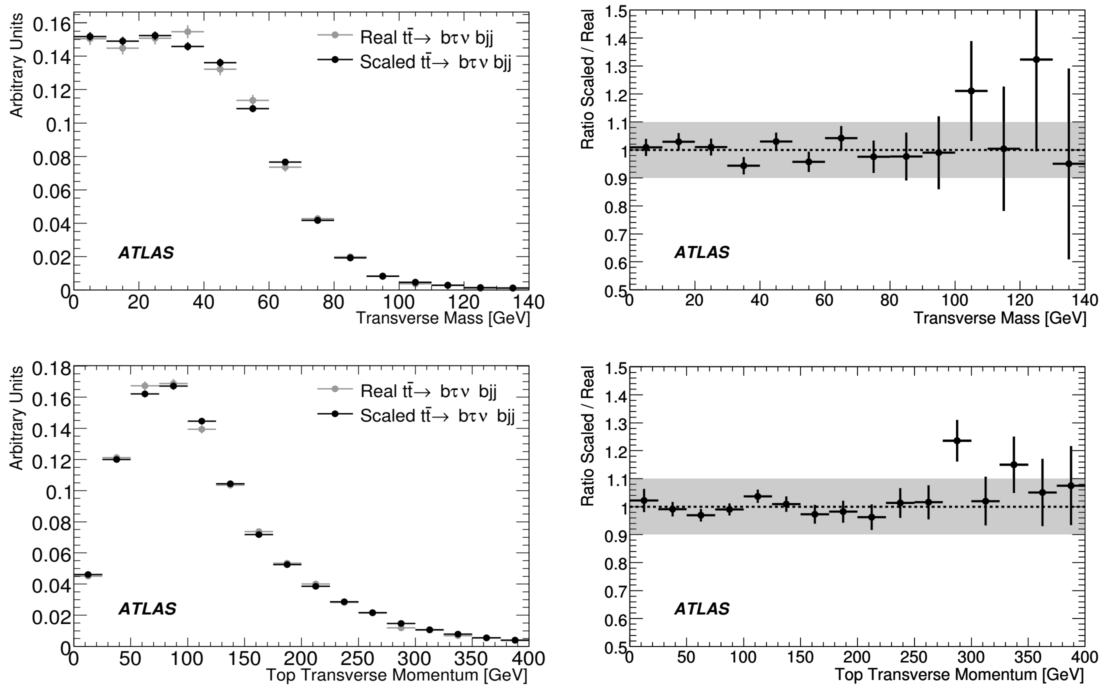

Figure 4: Top: $t\bar{t} \to b\tau(\text{lep})\nu bqq$, $W \to \tau\nu$ transverse mass. Bottom: $t\bar{t} \to b\tau(\text{had})\nu bqq$, top quark $t \to b\tau(\text{had})\nu$ transverse momentum. The plots at the left show the distribution for reference and embedded events, while the plots at the right show the ratio of both, with statistical uncertainties and a gray band highlighting the $\pm10\%$ area.

Compared to embedding, this method is technically simple. It requires non-trivial simulation input by taking shapes of distributions from simulation, however, testing the agreement of these shapes between data and simulation is quite straightforward. The method only works if the assumption that the data-to-simulation ratio is identical in sidebands and the signal region is fulfilled, which is difficult to establish in data.

5. Conclusions

Data-taking is advancing well at the LHC, and ATLAS has already collected a large number of top quark event candidates. Top quark pair events can be selected with high purity and so far, a good agreement between the properties of expected and observed events has been seen.

As previous studies have shown, $t\bar{t}$ events are the main background for nearly all H^+ searches, dominating by at least 1-2 orders of magnitude over other SM backgrounds. The systematic uncertainties of predicting the $t\bar{t}$ background contribution in relevant distributions from simulated events is large, and thus data-driven $t\bar{t}$ background estimation methods are essential. Two such methods have been introduced, namely the embedding method and a sideband-based method. The embedding method is technically complex, but as only input from simulation the τ decay products are needed. The sideband-based method, on the other hand, is technically simple and robust, but requires some assumptions to be fulfilled which need to be carefully tested in collision data.

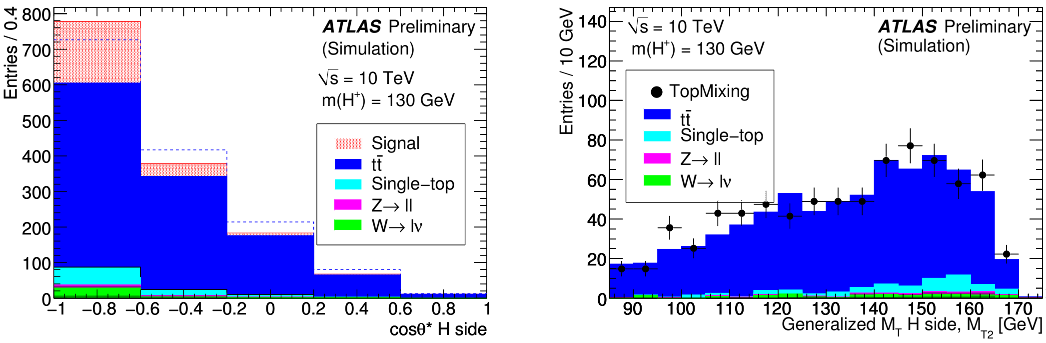

Figure 5: Left: The $\cos\theta_l^*$ distribution as expected from simulation. The dashed line is the SM expectation, the filled histogram as obtained when assuming $m_{H^+} = 130$ GeV (which reduces the SM $t\bar{t}$ production cross section). Right: Comparison of the generalized transverse mass [7] distributions as obtained from scaled simulated SM events and pseudo-data (called TopMixing in the legend) in the $t\bar{t}$ sideband.

References

[1] G. Aad et al., ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST 3 S08003 (2008).

[2] U. Langenfeld, S. Moch and P. Uwer, *New results for t anti-t production at hadron colliders*, Proc. of DIS2009

[3] Tevatron Electroweak Working Group, CDF, D0 Collaborations, *Combination of CDF and D0 Results on the Mass of the Top Quark*, arXiv:1007.3178v1

[4] The ATLAS Collaboration, *Search for top pair candidate events in ATLAS at sqrt(s)=7 TeV*, ATLAS-CONF-2010-063

[5] The ATLAS Collaboration, *Background studies for top pair production in lepton plus jets final states in sqrt(s)=7 TeV ATLAS data*, ATLAS-CONF-2010-087

[6] The ATLAS Collaboration, *Expected performance of the ATLAS experiment : detector, trigger and physics*, CERN-OPEN-2008-020

[7] The ATLAS Collaboration, *Expected sensitivity in light charged Higgs boson searches for H+ to taunu and H+ to csbar with early LHC data at the ATLAS Experiment*, ATL-PHYS-PUB-2010-006

[8] B. Mohn, M. Flechl, and J. Alwall, *ATLAS discovery potential for the charged Higgs boson in H+ to taunu decays*, ATL-PHYS-PUB-2007-006