PROCEEDINGS OF SCIENCE

Data-driven Methods for the Estimation of $t\bar{t}$ Backgrounds to Charged Higgs Searches

Trevor Vickey* **on behalf of the ATLAS Collaboration**
University of Wisconsin, Madison
E-mail: Trevor.Vickey@cern.ch

The dominant background to most charged Higgs boson analyses are $t\bar{t}$ events with one or more tau leptons in the final state. As the relative contributions from this background in different jet multiplicities are not known, and unknowns related to critical analysis cut-specific variables exist, the measurement and the subtraction of these backgrounds using a data-driven method is required. A procedure that can be used to obtain $t\bar{t}$ background control samples from ATLAS data is discussed, for both leptonically- and hadronically-decaying taus, and its effectiveness is demonstrated by using Monte Carlo events in lieu of collision data.

Prospects for Charged Higgs Discovery at Colliders
September 16-19 2008
Uppsala, Sweden

*Speaker.

PoS(CHARGED2008)034

1. Introduction

Theoretically predicted by many non-minimal Higgs scenarios, the experimental observation of charged Higgs bosons would indicate new physics beyond the Standard Model (SM). As one example, consider the Higgs sector of the Minimal Supersymmetric Standard Model (MSSM) [1] which contains five Higgs bosons: light and heavy CP-even scalars (h^0, H^0), a CP-odd scalar (A^0) and two charged scalars ($H^{\pm}$)[1]. For $m_{H^+} < m_t$ the dominant production mode at the Large Hadron Collider (LHC) at CERN takes place in $t\bar{t}$ events via the top quark decay $t \to H^+ b$ with $H^+ \to \tau^+ \nu$ dominating for values of $\tan\beta > 3$ (Fig. 1(a)).

The LHC, with a center-of-mass energy of 14 TeV and poised to begin proton-proton collisions in 2009, is expected to give experimentalists access to the majority of the MSSM parameter space. These proceedings summarize a procedure that can be used to obtain data-driven control samples for the estimation of the $t\bar{t}$ background events (Fig. 1(b)), by far the largest background contribution to charged Higgs searches at the LHC experiments [2, 3]. As the relative contributions from this background in different jet multiplicities are not known, and unknowns related to critical analysis cut-specific variables exist, the measurement and the subtraction of these backgrounds using such a data-driven method is mandatory.

Inspired by data-based Monte Carlo methods used at the Tevatron during Run I to estimate the $Z \to \tau\tau \to e\mu$, ee, and $\mu\mu$ backgrounds to top quark searches using $Z \to ee$ events collected from data [4], and methods that used $W \to e\nu$ data to model the electronic noise in the calorimeter and underlying event effects in $W \to \tau\nu$ events [5], we employ the `TAUOLA` decay package [6] to emulate leptonically- (τ_L) and hadronically-decaying taus (τ_H) by replacing muons from $t\bar{t}$ events collected in collision data with Monte Carlo tau leptons. The method presented here and published in Ref. [3] greatly parallels that used in Ref. [3] for the data-driven estimation of $Z \to \tau\tau$ backgrounds to SM and MSSM neutral Higgs searches in the $\tau_L\tau_H$, $\tau_L\tau_L$ and $\tau_H\tau_H$ final states. The approach here is data-driven in the sense that it will use

$$pp \to t\bar{t} + X \to W^+W^-b\bar{b} + X \to \mu^+\nu q\bar{q}' + X'$$

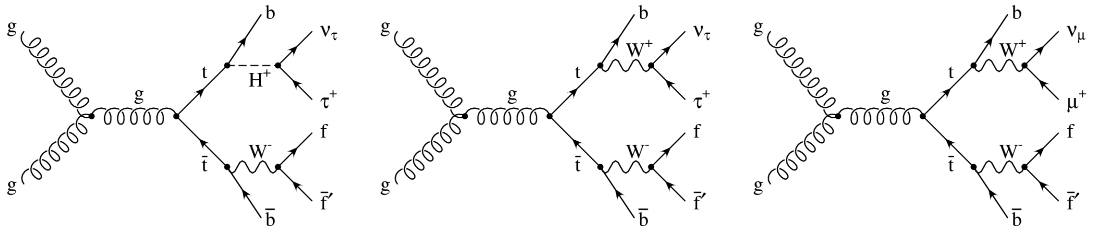
and

$$pp \to t\bar{t} + X \to W^+W^-b\bar{b} + X \to \mu^+\nu\mu^-\bar{\nu} + X'$$

events[2] (Fig. 1(c)) collected by ATLAS from pp collisions to model the $t\bar{t}$ backgrounds to charged Higgs searches in the $t\bar{t} \to H^+W^-b\bar{b} \to b\tau_L\nu\bar{b}q\bar{q}'$, $b\tau_H\nu\bar{b}q\bar{q}'$ and $b\tau_H\nu\bar{b}l\bar{\nu}$ final states. The decision to use final states containing muons was motivated by the level of efficiency and purity that could be obtained for the control samples.

After applying a very minimal set of event selection criteria on the data (optimized for both efficiency and purity), one muon from each event is fed into the `TAUOLA` Monte Carlo decay package and considered as a tau lepton. `TAUOLA` is used to decay this particle, and the decay products are passed to the ATLAS detector simulation (based on `GEANT4` [7]) and the reconstruction software. Once the Monte Carlo tau decay is merged back together with the rest of the event, the result is a data-driven control sample for each of the charged Higgs final states using events from data on

[1]Hereafter the charged Higgs bosons will be denoted H^+, with the charge conjugate processes implied.
[2]Including the charge conjugate processes.



(a) (b) (c)

Figure 1: Feynman diagrams for: (a) the production of a charged Higgs boson via $gg \rightarrow t\bar{t}$, the dominant production mode at the LHC for $m_{H^+} < m_t$. The decay $H^+ \rightarrow \tau^+\nu$ is preferred for values of $\tan\beta > 3$. (b) $t\bar{t}$ background events to H^+ searches that contain tau decays. (c) $t\bar{t}$ events with decays $t \rightarrow W^+b \rightarrow \mu^+\nu$; exploited to generate the background control sample.

which we can more easily and efficiently trigger. In lieu of collision data the effectiveness of this method is evaluated by using Monte Carlo events.

2. Obtaining a pure $t\bar{t}$ sample from data

To extract the $t\bar{t}$ control samples from data a set of selection criteria must be applied. These criteria evolved as a result of optimizing the efficiency and purity of the samples. This section reports on a study conducted to evaluate the necessary cuts.

2.1 The di-muon control sample

To obtain the $pp \rightarrow t\bar{t} + X \rightarrow W^+W^-b\bar{b} + X \rightarrow \mu^+\nu\mu^-\bar{\nu} + X'$ sample, events with at least two identified muons with transverse momenta ≥ 20 GeV are selected. These muons are required to be isolated (i.e., the energy in a cone with an opening angle $\Delta R = \sqrt{\Delta\phi^2 + \Delta\eta^2} = 0.3$ is required to be less than 20 GeV). To exclude muons coming from the decay of a Z boson, events with a di-muon invariant mass in the range 70-110 GeV are rejected. An additional requirement that the missing transverse energy in the event be > 40 GeV is imposed. Other backgrounds considered as potential contaminants to the control sample are $b\bar{b}$ and $W+$ jets events. The efficiency of the signal to survive this selection is 28% and the sample purity is estimated to be 72% [3].

2.2 The muon+jets control sample

The selection criteria for the $pp \rightarrow t\bar{t} + X \rightarrow W^+W^-b\bar{b} + X \rightarrow \mu^+\nu q\bar{q}' + X'$ sample are designed to reject the $b\bar{b} \rightarrow \mu + X$ events, which have a rather large cross-section. Events are accepted if an isolated high-p_T muon is found and two jets in the event with transverse momenta above 40 GeV have an invariant mass within 20 GeV of the nominal W mass. Events with high-p_T muons in the jets are rejected. A missing transverse energy cut of 40 GeV is imposed, and at least two more jets with transverse momenta above 40 GeV are required to be present. At least one of the jets in the event is required to have an associated b-tag. The overall transverse energy of the event is required to be greater than 250 GeV and the presence of a high-p_T isolated electron ($p_T > 20$ GeV) leads to the rejection of the event. The selection efficiency for signal events is 8.6% and the signal purity is 75% [3].

3. Turning muons into taus

Once the events have been selected in either the di-muon or muon+jets sample, they are bifurcated with one part containing the muon to be fed to TAUOLA (referred to here as the event "fragment"), and the other containing the event missing-energy, jets and any remaining charged leptons (hereafter referred to as the event "remnant"). The event remnant is set aside and the muon in the event fragment has its momentum three-vector individually re-scaled so that it appears to have the same mass as a tau lepton. The amount by which the muon momentum is scaled, ξ, is calculated as

$$\xi = \sqrt{\frac{E_\mu^2 - m_\tau^2}{|\vec{p}_\mu|^2}}.$$

Using the muon, with its re-scaled momentum, a customized event record of particle four-vectors is built in a format understood by the ATLAS offline software. This event record of the fragment, with the muon relabeled as a tau, is then pushed through TAUOLA and the tau is decayed according to the settings passed on to the decay package by the user (e.g., requesting a leptonically- or hadronically-decaying tau). After being decayed, the fragment is pushed through the GEANT4-based ATLAS detector simulation and the offline reconstruction software.

After the simulation and reconstruction of the event fragment has finished, the event remnant and the decayed tau are weaved back together at the physics-object level. It is this merged event that now contains the background control sample, over which the charged Higgs analysis event selection criteria is run. A flow diagram of the control sample selection, scaling, simulation and merging steps is depicted in Figure 2.

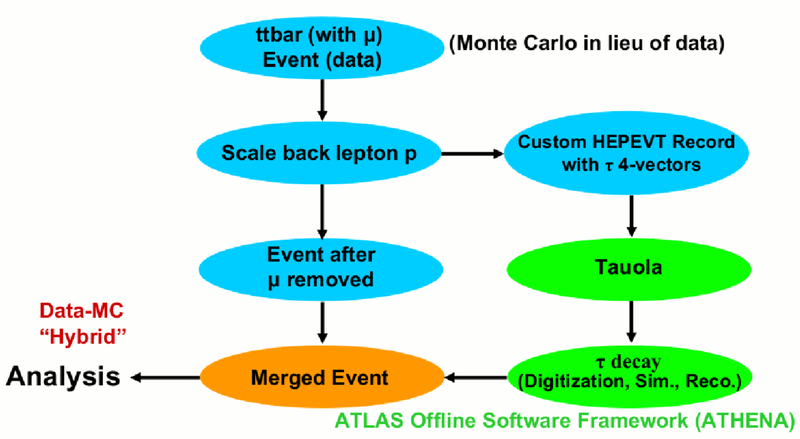


Figure 2: A flow diagram depicting the generation steps of the data-driven $t\bar{t}$ background control sample.

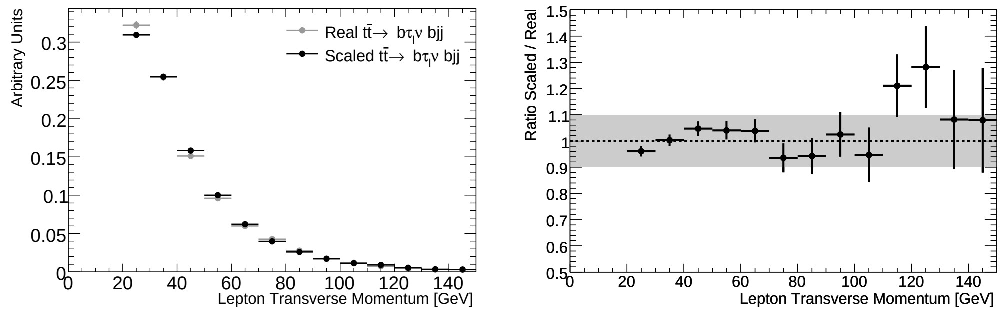


Figure 3: A validation plot for the $t\bar{t} \rightarrow b\tau_L \nu \bar{b} q \bar{q}'$ final state. The lepton transverse momentum from both the "real" and the "scaled" $t\bar{t}$ events (left) and the corresponding bin-by-bin ratio (right). The gray band represents $\pm 10\%$ around a ratio of 1.

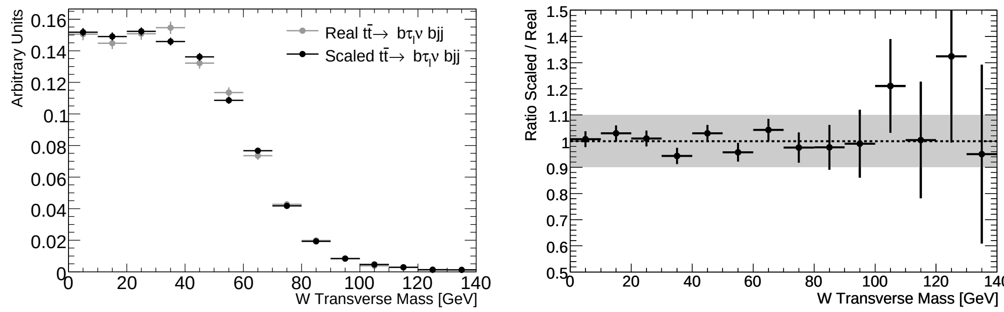


Figure 4: A validation plot for the $t\bar{t} \rightarrow b\tau_L \nu \bar{b} q \bar{q}'$ final state. The reconstructed W transverse mass distribution from both the "real" and the "scaled" $t\bar{t}$ events (left) and the corresponding bin-by-bin ratio (right). The gray band represents $\pm 10\%$ around a ratio of 1.

4. Results

To thoroughly test the $t\bar{t}$ control sample method, the produced events have to be run through each of the different analyses and the background shapes and normalizations obtained must be compared to those from $t\bar{t}$ Monte Carlo events containing real tau leptons in the final state. However, a quick global check can be done by comparing various distributions from the "real" and the "scaled" events (from the control sample), where muons have been replaced by tau leptons. This has been done separately for the three final states of interest: $b\tau_L \nu \bar{b} q \bar{q}'$ [8], $b\tau_H \nu \bar{b} q \bar{q}'$ [9] and $b\tau_H \nu \bar{b} l \bar{\nu}$ [10].

Figure 3 shows the transverse momentum of the muon from the tau decay in the $t\bar{t} \rightarrow b\tau_L \nu \bar{b} q \bar{q}'$ final state, demonstrating that the replacement of the original muon with a scaled tau (decaying to a muon) has been successful. Figure 4 shows the reconstructed transverse mass of the W boson in the $t\bar{t} \rightarrow b\tau_L \nu \bar{b} q \bar{q}'$ final state, defined as $M_T^W = \sqrt{2 p_T^l p_T^{miss} (1 - \cos(\Delta\phi))}$, which is a complex quantity based on two objects and thus gives evidence that the relevant correlations in the event are not destroyed in the emulation process.

The irreducible background for charged Higgs searches in the $t\bar{t} \rightarrow b\tau_H \nu \bar{b} q \bar{q}'$ final state can be emulated by replacing muons in $t\bar{t} \rightarrow b\mu\nu \bar{b} q \bar{q}'$ events recorded during collision data-taking with simulated taus, and forcing the tau to decay hadronically. Figure 5 shows the reconstructed top

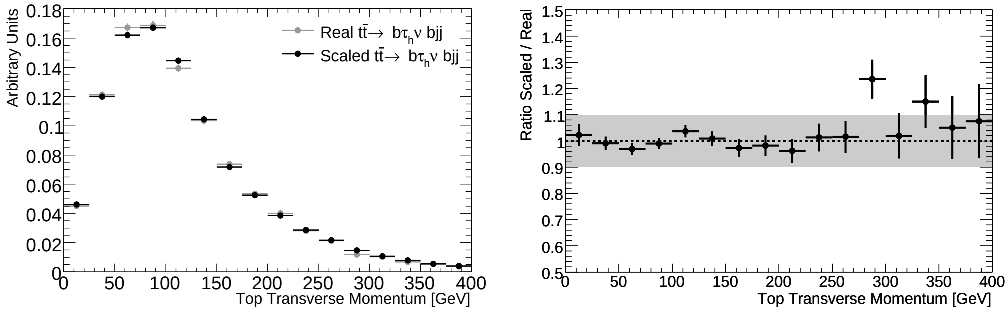


Figure 5: A validation plot for the $t\bar{t} \rightarrow b\tau_H \nu \bar{b} q \bar{q}'$ final state. The top quark transverse momentum distribution (from $t \rightarrow Wb \rightarrow b\tau_H \nu$) from both the "real" and the "scaled" $t\bar{t}$ events (left) and the corresponding bin-by-bin ratio (right). The gray band represents $\pm$10% around a ratio of 1.

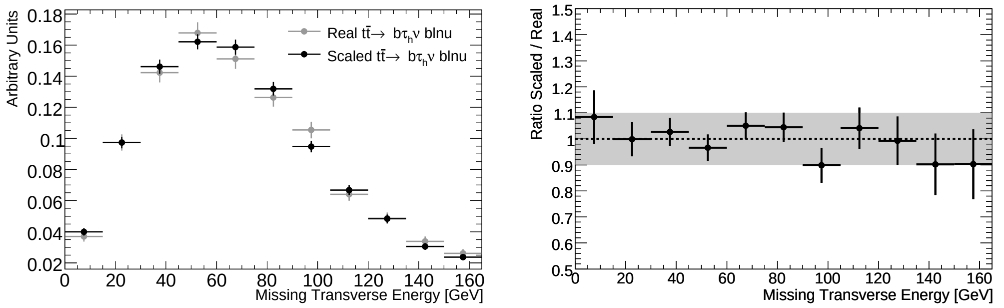


Figure 6: A validation plot for the $t\bar{t} \rightarrow b\tau_H \nu \bar{b} l \bar{\nu}$ final state. The missing transverse energy distribution from both the "real" and the "scaled" $t\bar{t}$ events (left) and the corresponding bin-by-bin ratio (right). The gray band represents $\pm$10% around a ratio of 1.

quark transverse momentum (from $t \rightarrow Wb \rightarrow b\tau_H \nu$) for both the "real" and the "scaled" $t\bar{t}$ events.

The irreducible background for the $t\bar{t} \rightarrow b\tau_H \nu \bar{b} l \bar{\nu}$ analysis can be emulated by replacing a muon with a simulated tau in a $t\bar{t} \rightarrow b\mu\nu \bar{b} l \bar{\nu}$ event, and forcing the tau to decay hadronically. Due to the presence of leptons, a hadronic tau and large missing transverse energy in the event, this channel can potentially also contribute to the background of all other charged Higgs analyses. Figure 6 shows the missing transverse energy which is a complex quantity resulting from the combination of all objects in the event from both the "real" and the "scaled" $t\bar{t}$ events.

5. Conclusions

The distributions presented in the previous section demonstrate that the dominant background for all H^+ analyses, $t\bar{t}$ events, can be modeled with the $t\bar{t}$ control sample method. In the regions of interest, quantities relevant to the H^+ analyses can be modeled within a 10% error margin. This is remarkable in particular for complex quantities, i.e., variables derived from a combination of several objects (for example, the top quark transverse momentum), and gives confidence that the $t\bar{t}$ control sample method allows one to reproduce many of the relevant correlations in the event.

PoS(CHARGED2008)034

The background normalization can be obtained either by algebraic means using number counting and measured efficiencies, or through the use of maximum likelihood fitting to data using signal and background shapes.

Early into the data-taking era, the uncertainties in the method will largely be dominated by limited statistics. The contamination of the control sample and the introduction of possible trigger biases are important systematic effects that warrant additional study.

Further refinement of this method is currently in progress. For example, it is possible for the removal and replacement scheme to be carried out at the level of the tracking hits and calorimeter depositions rather than at the higher level of reconstructed physics objects. Further investigations may reveal that electrons in $t\bar{t}$ lepton+jets and di-lepton events are also suitable for replacement with Monte Carlo taus, thereby doubling and quadrupling the statistics available for these control samples, respectively.

Acknowledgments

The author would like to thank E. Gross of the Weizmann Institute in Israel for initially suggesting this study. Additionally, the author would like to acknowledge Y. Rozen and S. Behar Harpaz of Technion-Israel Institute of Technology for their study on the purity of the background control sample, as well as the assistance and support of E. Coniavitis, M. Flechl and B. Mohn and the ATLAS Higgs Working Group conveners K. Assamagan, L. Fayard and A. Nisati. The author was supported in part by the U. S. Department of Energy through grant No. DE-FG02-95ER40896.

References

[1] A. Djouadi, Phys. Rept. **459**, 1-241 (2008), and references therein.

[2] ATLAS Collaboration, ATLAS Physics TDR, CERN-LHCC 99-14/15 (1999).

[3] ATLAS Collaboration, "Expected Performance of the ATLAS Experiment—Detector, Trigger and Physics," CERN-OPEN-2008-020 (2008) [arXiv:0901.0512v1 [hep-ex]].

[4] F. Abe *et al.*, Phys. Rev. D **50**, 2966 (1994).

[5] B. Abbott *et al.*, Phys. Rev. Lett. **84**, 5710-5715 (2000).

[6] S. Jadach *et al.*, "The tau decay library, TAUOLA: version 2.4," Comp. Phys.Commun. 76 (1993).

[7] S. Agostinelli *et al.*, Nuclear Instruments and Methods in Physics Research A **506**, 250-303 (2003); J. Allison *et al.*, IEEE Transactions on Nuclear Science **53** No. 1, 270-278 (2006).

[8] O. Vitells, "Charged Higgs searches in the leptonic tau channel," this volume.

[9] E. Coniavitis, "Light charged Higgs searches in the hadronic mode with ATLAS," this volume.

[10] T. Ehrich, "Searches for light charged Higgs decaying to a hadronic tau in the one lepton mode with ATLAS," this volume.