2009/07/03
Archive Id: 1.45
Archive Date: 2009/07/03 08:49:42

Prospects for the measurement of the single-top t-channel cross section in the muon channel with 200 pb^{-1} of CMS data at 10 TeV

The CMS Collaboration

Abstract

We report on a study aiming at an early observation of single-top events produced in the t channel in proton-proton collisions, at a centre-of-mass energy of $\sqrt{s} = 10$ TeV, in the decay channel $t \to bW \to b\mu\nu$. A template-fit method is proposed, that takes advantage of the spin correlations of the decay products in signal events, and appears robust against several systematic effects. This article assumes the use of 200 pb^{-1} of integrated luminosity. Under these conditions, a cross section uncertainty of $\pm$35% (statistical) $\pm$14% (systematic) $\pm$10% (luminosity) and a sensitivity of 2.7σ are expected, assuming the standard-model prediction of σ(single top, t channel) $=$ 130 pb.

1 Introduction

The theory of electroweak interactions predicts three different production mechanisms for single top quarks in hadron-hadron collisions, in addition to the more abundant pair production due to the strong interaction: t channel, s channel, and tW (or W-associated). Recently the D0 and CDF experiments reported a 5σ observation of single top at the Tevatron $p\bar{p}$ collider [1, 2]. At the Large Hadron Collider (LHC) the reobservation is expected to happen first in the t-channel mode, by far the most abundant of the three at LHC energies, and with the most striking final state topology. This article treats this production mode as signal, including the other two in the definition of background, and assumes a centre-of-mass energy of $\sqrt{s} = 10$ TeV.

The t-channel events from Monte Carlo simulation used in this study have been generated with the MadGraph event generator [3]. In order to give a fair approximation of the full next-to-leading order (NLO) properties of the signal, the $2\rightarrow3$ diagram, corresponding to the dominant NLO contribution to the t channel, is combined with the leading order (LO) $2\rightarrow2$ process by a matching procedure based on Ref. [4], giving a merged sample that describes the entire phase space while avoiding double counting.

Several standard-model processes are taken into account as background to the analysis. MadGraph is used also for top-quark pair production ($t\bar{t}$), for the other single-top modes, and for the inclusive single-boson production ($W/Z + X$, where X can indicate light or heavy partons). A procedure implemented during the event generation and based on the so called "MLM prescription" [5] avoids double counting between Matrix Element and parton shower generated jets. A very similar procedure prevents double counting of the heavy flavour content of $W/Z + X$ samples. The remaining background samples, due to di-boson production (WW, WZ, ZZ) and multi-jet QCD enriched in events with muons coming from the decay of b and c quarks or long-lived hadrons, were simulated using PYTHIA [6].

All generated events undergo a full simulation of the detector response according to the CMS implementation of GEANT4 [7]. Only one pp collision per bunch crossing is simulated.

2 Event selection

The study presented here focuses on the $t \rightarrow bW \rightarrow b\mu\nu$ decay channel. All events must pass the high-level single-muon trigger requirement which includes a 15 GeV/c transverse momentum threshold and $|\eta| < 2.1$; this trigger will be available without prescaling at instantaneous luminosities up to $\approx 10^{32}\ cm^{-2}s^{-1}$ or larger. Reconstructed muons with a transverse momentum $p_{T,\mu} > 20$ GeV/c within the trigger acceptance, passing additional quality criteria, are selected. The event is rejected if more than one such muon is present, and also if an electron candidate is present with tight quality selection and $p_{T,e} > 20$ GeV/c, $|\eta| < 2.4$.

We define the relative isolation variable as

$$\text{relIso} = \frac{p_{T,\mu}}{p_{T,\mu} + \text{tkIso} + \text{caloIso}}\,, \tag{1}$$

where tkIso (caloIso) is the scalar sum of the transverse momenta (transverse energies) of the tracks (calorimeter deposits) in a cone of size $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.3$ around the muon direction, excluding the track (calorimetric footprint) of the muon itself. Events with relIso $>$ 0.95 are selected.

Jets are defined according to the iterative cone algorithm [8] with a cone size of 0.5. We consider jets within $|\eta| < 5$ whose calibrated transverse momentum is greater than 30 GeV/c. The event

Table 1: Expected event yield with 200 pb^{-1} of data for all processes considered in the analysis, after the analysis cuts. The uncertainty was obtained from the size of the simulated samples, indicated by the integrated luminosity. The cross sections shown include branching ratios when necessary. The single-top cross section in the t channel has been calculated at 10 TeV as in Ref. [9], while the s channel and the tW channel have been rescaled from Refs. [9] and [10], respectively; the diboson (WW, WZ, ZZ) cross sections are calculated as in Ref. [11]; the $t\bar{t}$ cross section comes directly from Ref. [12]; a filter at generator level has been applied on the μ-enriched multi-jet sample, and the cross section times filter efficiency has been taken from PYTHIA; all other cross sections come from MadGraph.

Process	$\sigma \times$ BR[pb]	$L\ [fb^{-1}]$	N_{evt} in 200 pb^{-1}
single top, t channel ($W \to l\nu, l = e, \mu, \tau$)	42.9 (NLO)	6.6	102±1.8
single top, s channel ($W \to l\nu, l = e, \mu, \tau$)	1.6 (NLO)	7.5	1.8±0.2
single top, tW	29 (NLO)	5.8	22.3± 0.9
$t\bar{t}$	414 (NLO+NLL)	2.2	136.0±3.5
QCD multi-jet (μ-enriched)	121675 (LO)	0.05	12±6.7
Wc ($W \to l\nu, l = e, \mu, \tau$)	1 490 (LO)	2.0	29±1.7
$Wb\bar{b}$ ($W \to l\nu, l = e, \mu, \tau$)	54.2 (LO)	2.9	8.0±0.7
$Wc\bar{c}$ ($W \to l\nu, l = e, \mu, \tau$)	118.8 (LO)	4.5	1.2±0.2
$W+$ light partons ($W \to l\nu, l = e, \mu, \tau$)	40 000 (LO)	0.24	12±2.6
$Zb\bar{b}$ ($Z \to ll, l = e, \mu, \tau$)	44.4 (LO)	3.5	2.7±0.4
$Zc\bar{c}$ ($Z \to ll, l = e, \mu, \tau$)	71.7 (LO)	5.0	0.2±0.1
$Z+$ light partons ($Z \to ll, l = e, \mu, \tau$)	3 700 (LO)	0.33	2±1.2
WW	74 (LO)	2.8	0.9±0.3
WZ	32 (LO)	7.4	1.2±0.2
ZZ	10.5 (LO)	19.0	0.17±0.04
Total Background			229±8.4

is accepted for further analysis only if exactly two such jets were reconstructed. Furthermore, we reject events where the distance ΔR between the muon and the closest jet is less than 0.3 (*near-jet veto*).

We apply a *track counting* (TC) b-tagging algorithm that calculates the signed 3D impact-parameter significance (IP/σ_{IP}) of all the tracks passing tight quality criteria associated to the jet, orders them by decreasing values of this observable, and defines as jet discriminator the value of IP/σ_{IP} for the second (*high-efficiency* TC) or third (*high-purity* TC) track. The event is accepted for further analysis only if exactly one of the selected jets passes a tight threshold on the high-purity TC. Since we expect most of the signal events to have only one b quark inside the Tracker acceptance ($|\eta| < 2.5$), we reject the event if the remaining jet passes a loose threshold on the high-efficiency TC.

To further suppress contributions from processes where the muon does not come from a leptonically decaying W boson, we select events with $M_T > 50$ GeV/c^2, with

$$M_T = \sqrt{\left(p_{T,\mu} + p_{T,\nu}\right)^2 - \left(p_{x,\mu} + p_{x,\nu}\right)^2 - \left(p_{y,\mu} + p_{y,\nu}\right)^2}\,, \tag{2}$$

where the neutrino momentum vector is assumed equal to the calibrated transverse missing energy ($E_{\mathrm{T}}^{\mathrm{miss}}$) of the event.

The expected event yield from this selection with 200 pb^{-1} of data is shown in Table 1.

3 QCD multi-jet background estimation

Estimations of the QCD multi-jet contamination from simulated data have to be considered particularly unreliable for the purposes of our analysis, because only events from specific kinematical regions pass the selection, and tail effects are the most difficult to properly simulate. These arguments lead to the conclusion that only *in situ* data-driven estimations will give the needed confidence on the amount of this background.

We extract the size of the QCD multi-jet and signal-like contributions using the M_T shape after all other selection criteria have been applied, by parametrizing the M_T distribution as

$$F(M_T) = a \cdot S(M_T) + b \cdot B(M_T) \,, \tag{3}$$

where $S(M_T)$ and $B(M_T)$ are the expected distributions for signal-like (muons coming from W decays) and QCD multi-jet events, respectively.

$S(M_T)$ and $B(M_T)$ are extracted from high-statistics control samples. We verified with simulated events that the shape of the M_T distribution for events of a same process passing the control selections does not differ, within statistical uncertainties, from the events passing the standard selection. In order to obtain a background-enriched sample, we apply a dedicated selection that differs from the standard one by the absence of the b-tagging requirements and by an anti-isolation cut (relIso < 0.8). These requirements reject most of the signal-like events (single top, $W + X$, $t\bar{t}$, and in general any process with a charged lepton from an intermediate W boson) leaving a background-dominated sample. Different options have been explored for the extraction of $S(M_T)$, among which are

- the use of a W-enriched control sample, which differs from the standard selection only by the absence of b tagging;
- the use of a Z-enriched control sample, obtained with a dedicated selection requiring at least two muons with invariant mass in the range $76 < M_{\mu\mu} < 106$ GeV/c^2, and two jets selected as in the standard selection apart from the absence of the b tagging requirements.

In the Z-enriched case the definition of the M_T variable is modified *ad hoc*: we rescale the momenta of the two leading muons by M_W / M_Z, we treat one of them (randomly chosen) as a neutrino, and we vectorially add its transverse momentum to $E_{\mathrm{T}}^{\mathrm{miss}}$. The distribution obtained in this way has a maximum at the same position as the standard M_T definition in the case of signal or $W + jets$, with minimal QCD multi-jet contamination.

The yields obtained with either of these signal templates are compatible with the actual count of simulated QCD multi-jet events passing the full selection. Extracting the signal-like template from the Z-enriched control sample, which is the method chosen due to its higher purity with respect to the W-enriched one, yields a prediction of 22 events in the signal region. By considering the spread between the results obtained with different signal templates, including those from simulated samples containing only signal or only W+ light partons events, we assign a $\pm 40\%$ systematic uncertainty to the rate of QCD multi-jet background, for an overall uncertainty of $\pm 45\%$ when considering the expected statistical uncertainty.

4 Top quark reconstruction

The first step in the reconstruction of top-quark candidates makes use of the precise knowledge of the W mass to provide a kinematic constraint, which leads to a quadratic equation in the

longitudinal neutrino momentum. This equation has, in general, two solutions, which can have an imaginary part (this happens when M_T is larger than the W pole mass used in the constraint); here, the imaginary component is eliminated by modifying $E_{\mathrm{T}}^{\mathrm{miss}}$ such as to give $M_T = M_W$, still respecting the W mass constraint. When two real solutions are present, we choose the solution with the smallest absolute value.

A similar two-fold ambiguity presents itself when reconstructing a top-quark hypothesis, since two jets are selected. The b-tagged jet is assigned to the top-quark decay.

Figure 1 shows the mass of the reconstructed top quark ($M_{l\nu b}$) for events passing the full selection. The observation of a maximum around the known value of the top mass in real collision data will be a strong indication of the presence of top quarks.

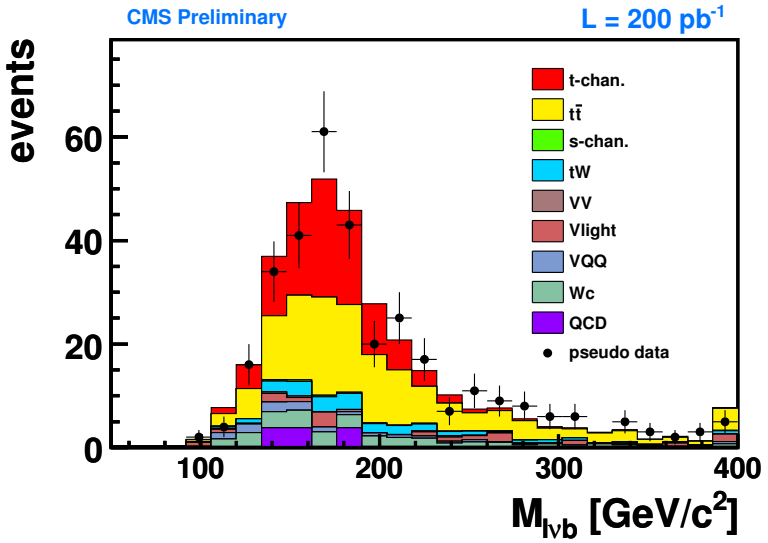

Figure 1: Reconstructed top-quark mass after the full selection. The last bin also contains events with $M_{l\nu b} > 400$ GeV/c^2. Signal events are labeled as t-chan.; single top in s channel as s-chan.; VV indicates the sum of WW, WZ, and ZZ; Vlight the sum of W and Z events in association with light partons, while in VQQ they are associated to $b\bar{b}$ or $c\bar{c}$ pairs; QCD is a short-hand notation for multi-jet QCD events.

5 Top quark polarization angle

A specific feature of the signal, stemming from the $V - A$ structure of the weak interaction, is the almost 100% left-handed polarization of the top quark with respect to the spin axis [13, 14]. The direction of the top-quark spin is visible in angular correlations of its decay products, which are distributed according to

$$\frac{1}{\Gamma}\frac{d\Gamma}{d\cos\theta^*_{lj}} = \frac{1}{2}(1 + A\cos\theta^*_{lj}) \,, \tag{4}$$

where θ^*_{lj} is the angle between the direction of the outgoing lepton and the spin axis, approximated by the direction of the untagged jet, in the top-quark rest frame. A is the coefficient of spin asymmetry, equal to $+1$ for charged leptons.

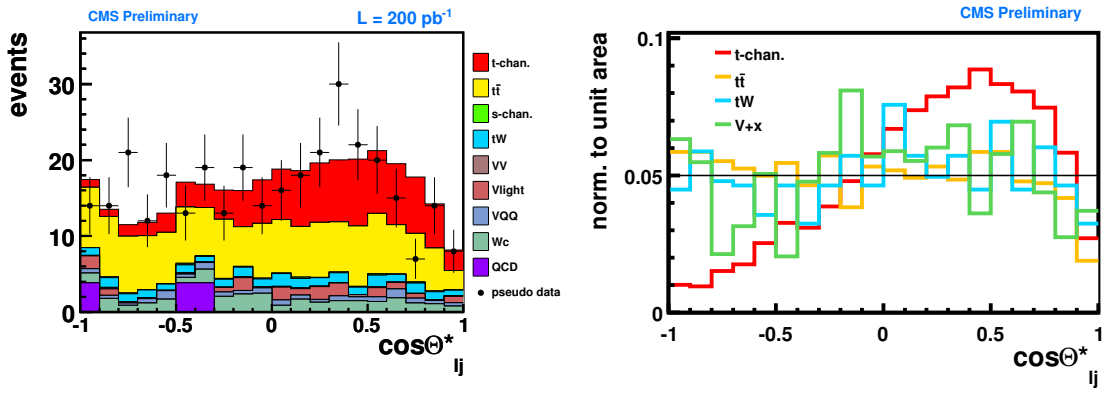

Figure 2: Cosine of the angle between charged muon and untagged jet, in the reconstructed top rest frame after the full event selection.

Figure 2 shows the distribution of the cosine of this angle, for events passing our event selection. The dip at $\cos\theta^*_{lj} \approx 1$ is due to the muon p_T and M_T cuts.

6 Signal extraction and cross section measurement

The cross section is determined by performing a binned likelihood fit to the $\cos\theta^*_{lj}$ distribution of the selected events. The inputs to the fit are the template distributions for signal and background. The signal template is taken from simulation, while the overall background is assumed to be flat. This assumption is verified with background-enriched control samples, finding distributions consistent with the flatness hypothesis within the statistical uncertainties. The fit is restricted to the $[-1, 0.75]$ interval in order to minimize the aforementioned kinematic effects.

The statistical sensitivity of the signal extraction has been determined by simulating 500 000 pseudoexperiments. This procedure yields a 35% statistical uncertainty on the cross section for 200 pb^{-1} of data at 10 TeV, assuming that the true value is the one predicted by the standard model, and an expected sensitivity of 2.8σ. The evolution of the sensitivity with the integrated luminosity is shown in Fig. 3.

7 Systematic uncertainties and robustness tests

The systematic uncertainties considered correspond to a level of understanding of the detector as foreseen to be achieved at the time when 200 pb^{-1} of data will be available.

The impact of the Parton Distribution Function (PDF) uncertainty on both the event yield and the $\cos\theta^*_{lj}$ shapes is estimated by reweighting the selected events according to each PDF eigenvalue in the CTEQ61 collection [15]. We observe that the deviations in event yield from the default PDF set are dominated, for each process, by one eigenvector in the positive and one in the negative direction; therefore, in order to simplify the estimation, only the eigenvectors giving the most extreme deviations are taken into account for each process to calculate the deviations in the extracted cross section.

In order to take into account the expected Jet Energy Scale (JES) uncertainty on both the event yield and the $\cos\theta^*_{lj}$ shape, we apply a simultaneous variation of the overall JES by $\pm 10\%$ [16]. Since the missing transverse energy is also corrected for jets, its uncertainty is correlated with

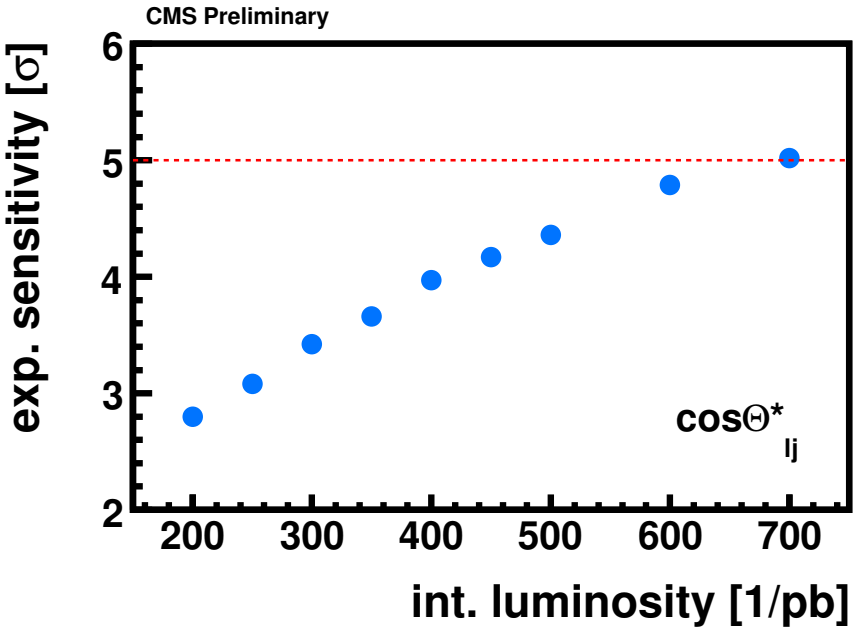

Figure 3: Evolution of the expected sensitivity with the integrated luminosity. Systematic uncertainties are not included.

the JES uncertainty. Here, two independent sources of $E_{\mathrm{T}}^{\mathrm{miss}}$ systematics are considered:

- correlated with JES: all the jets with a transverse momentum above 20 GeV/c are corrected by the same factors discussed before and $E_{\mathrm{T}}^{\mathrm{miss}}$ is recalculated accordingly;
- uncorrelated with JES: after subtracting the jet corrections, $E_{\mathrm{T}}^{\mathrm{miss}}$ is varied by 10%.

We vary the b-tagging efficiency of the track-counting algorithm at the tight (loose) working point by $\pm 8.0\%$ ($\pm 8.2\%$), while the mistagging probability is varied by $\pm 18.1\%$ ($\pm 3.4\%$), according to the expected performance of the b tagging algorithms [17, 18]. The corresponding variations in the event yield and in the $\cos\theta^*_{lj}$ shapes are taken into account as systematic uncertainties.

The effect on the cross section extraction is estimated by taking event yields and shapes corresponding to the extremes of the quoted ranges and repeating the likelihood fit under these conditions. The result is summarized in Table 2.

The expected sensitivity of the analysis is calculated by performing two ensemble tests with pseudo experiments, one including single top-quark events in t and s channels (hypothesis H_1) and one without them (hypothesis H_0). The acceptance and shape uncertainties are incorporated in the pseudo experiments by randomly drawing the strength of a systematic uncertainty according to a Gaussian distribution centered on zero and with a variance equal to its quoted extreme systematic variance. Under the assumption that H_1 is true, we expect a 50% probability to obtain a 2.8σ excess over the H_0 hypothesis when systematic effects are ignored, and 2.7σ when considering the systematic uncertainties in the fit (see last column of Table 2), with 200 pb^{-1} of data at 10 TeV.

The sensitivity of our procedure to the overall background level is tested by rescaling it by $\pm$ 50%. No bias is introduced in the cross section determination, and the statistical uncertainty

Table 2: Statistical and systematic uncertainties on the cross section measurement and on the expected sensitivity with 200 pb^{-1}. The absolute value of the maximum deviation is quoted as uncertainty on the cross section.

Source of uncertainty	$\Delta\sigma$ [%]	Expected sensitivity
statistical	$\pm$ 35	2.8σ
b tagging	$\pm$ 7.3	2.7σ
mistag	$\pm$ 0.4	2.7σ
JES	$\pm$ 5.5	2.7σ
MET	$\pm$ 9.9	2.7σ
PDF	$\pm$ 5.5	2.7σ
total	$\pm$ 39	2.7σ

becomes 40.8% and 27.8 % for upward and downward variations respectively. In these two scenarios, we obtain expected sensitivities of 2.2σ and 3.2σ over the H_0 hypothesis, respectively.

An uncertainty of 10% is assumed on the luminosity determination [19].

Since the analysis depends on the shapes assumed for signal and background, we perform the following further tests of robustness.

In the analysis the signal is modeled by matching the $2 \to 2$ and $2 \to 3$ diagrams at leading order, and normalized to the NLO cross section. In order to test the effect of parton-level signal modeling, we conservatively compare the $\cos\theta^*_{lj}$ distribution of the $2 \to 2$ and the $2 \to 3$ components separately, and we find a negligible difference, fully accountable by the size of our simulated samples.

We consider the deviations from flatness of $\cos\theta^*_{lj}$ for the main background components separately. For $t\bar{t} + tW$ events the shape is taken from simulation and its flatness is tested with a $t\bar{t}$-enriched control sample obtained from events with a second b jet (passing the loose threshold), for $W/Z + X$ events the shape is taken from the W-enriched control sample described in Sec. 3, and for multi-jet QCD from the QCD-enriched control sample with anti-isolation requirement. In the last two cases the $M_T > 50$ GeV/c^2 requirement is added and the most central jet is treated as a b jet in the top-quark reconstruction of Sec. 4. Note that *a priori* there is no reason to expect a non-flat distribution in $\cos\theta^*_{lj}$ for these backgrounds significantly below $\cos\theta^*_{lj} \approx 1$ where acceptance cuts bias the distribution, and the small observed deviations can be accounted as statistical fluctuations.

As a further test of the robustness of the analysis against $t\bar{t}$ modeling we vary the PYTHIA parameters that are responsible for the amount of initial and final state gluon radiation (ISR/FSR) according to the extreme values recommended in Ref. [20]. We also verify the compatibility between the $\cos\theta^*_{lj}$ shapes obtained from MadGraph and PYTHIA samples. In both cases we observe negligible effects, that can be accounted as statistical fluctuations.

8 Conclusions

The central result of the analysis presented here is that it is realistic to provide the first evidence of single-top production in a pp collider with ≈ 200 pb^{-1} of data at 10 TeV.

After applying a selection optimized for t-channel single top events, which leaves $t\bar{t}$ as the dominant background, we achieve the needed separation of the signal from background by

exploiting the polarization of the top quark, which is entirely transferred to the decay products, yielding a very characteristic muon angular distribution in the top-quark rest frame.

After consideration of several instrumental and theoretical uncertainties we obtain an expected relative uncertainty of 35% for the cross section measurement and a sensitivity of 2.8σ when systematic effects are ignored. The inclusion of systematic uncertainties coming from PDFs and from detector knowledge contribute an additional 14%, and the uncertainty on the luminosity is estimated as 10%, yielding an overall relative uncertainty of 39%, while the expected sensitivity is lowered to 2.7σ.

We tested the robustness of the method by applying extreme variations in the modeling of the signal and of the main backgrounds. We conclude that our results do not depend critically on the model assumptions for signal and backgrounds.

Several improvements are possible for this analysis. In particular, an important property of the signal that has not been exploited in the present study is its charge asymmetry in pp collisions (83.6 pb for top and 46.5 pb for anti-top production [9]). Preliminary studies indicate that the use of this feature can be very advantageous provided that the amount of $W + X$ events is under control.

References

[1] **D0** Collaboration, "Observation of single top quark production," arXiv:0902.4883v2.

[2] **CDF** Collaboration, "First Observation of Electroweak Single Top Quark Production," arXiv:0903.0885v2.

[3] F. Maltoni and T. Stelzer, "MadEvent: Automatic event generation with MadGraph," *JHEP* **0302** (2003) 027.

[4] E. Boos et al., "Method for simulating electroweak top-quark production events in the NLO approximation: SingleTop event generator," *Phys.Atom.Nucl.* **69, 8** (2006) 1317.

[5] J. Alwall et al., "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions," *Eur.Phys.J.* **C53** (2008) 473–500, arXiv:0706.2569.

[6] T. Sjostrand et al., "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026.

[7] J. Allison et al., "Geant4 developments and applications," *IEEE Transactions on Nuclear Science* **53, 1** (2006) 270–278.

[8] **UA1** Collaboration, "Observation of jets in high transverse energy events at the CERN proton antiproton collider," *Phys. Lett. B* **123** (1983) 115.

[9] J. Campbell, R. Ellis, and F. Tramontano, "Single top production and decay at next-to-leading order," *Phys.Rev.* **D70** (2004) 094012.

[10] J. Campbell and F. Tramontano, "Next-to-leading order corrections to Wt production and decay," *Nucl.Phys.* **B726** (2005) 109.

[11] J. Campbell and R. Ellis, "Vector boson pair production at the Tevatron, including all spin correlations of the boson decay products," *Phys.Rev.* **D60** (1999) 114012.

[12] M. Cacciari, S. Frixione, M. Mangano, P. Nason, and G. Ridolfi, "Updated predictions for the total cross sections for top and of heavier quark pairs at the Tevatron and at the LHC," *JHEP* **0809** (2008) 127.

[13] G. Mahlon and S. Parke, "Improved Spin Basis for Angular Correlation Studies in Single Top Quark Production at the Tevatron," *Phys.Rev.* **D55** (1997) 7249.

[14] P. Motylinski, "Angular correlations in t-channel single top production at the LHC," arXiv:0905.4754.

[15] J. Pumplin et al., "New Generation of Parton Distributions with Uncertainties from Global QCD Analysis," *JHEP* **0207** (2002) 012.

[16] **CMS** Collaboration, "Plans for Jet Energy Corrections at CMS," *CMS PAS JME-07-002* (2007).

[17] **CMS** Collaboration, "Performance Measurement of b tagging Algorithms Using Data containing Muons within Jets," *CMS PAS BTV-07-001* (2007).

[18] **CMS** Collaboration, "Evaluation of *udsg* Mistag Rate of b-tag Jets using Negative Tags," *CMS PAS BTV-07-002* (2007).

[19] H. Jung et al., "Proceedings of the workshop: HERA and the LHC workshop series on the implications of HERA for LHC physics," arXiv:0903.3861.

[20] **CMS** Collaboration, P. Bartalini, R. Chierici, and A. D. Roeck, "Guidelines for the estimation of theoretical uncertainties at the LHC," *CMS NOTE 2005/015* (2005).