CDF note 9304

# Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $\sqrt{s} = 1.96$ TeV Using Lepton Plus Jets Events with Soft Muon b-Tagging

The CDF Collaboration
*URL http://www-cdf.fnal.gov*
(Dated: January 30, 2009)

We present a measurement of the $t\bar{t}$ production cross section in $\sim 2\,\mathrm{fb}^{-1}$ of Run 2 data. The muon SLT is applied to events with a lepton + 3 or more jets and missing $E_T$. Backgrounds are computed from a combination of Run 2 data and simulation. Signal acceptance is determined from Run 2 PYTHIA Monte Carlo. Based on the tags in the 3 and 4-jet bins, a production cross section of $9.1 \pm 1.6$ pb is measured.

*Preliminary Results for Spring 2008*

*Contact e-mail: grundler@uiuc.edu, tml@uiuc.edu*

## I. INTRODUCTION

This note describes a measurement of the $t\bar{t}$ production cross section in $p\bar{p}$ collisions at $\sqrt{s}$ = 1.96 TeV with the CDF detector at the Fermilab Tevatron. The standard model predicts that $t\bar{t}$ production in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV proceeds primarily through quark-antiquark annihilation with a small admixture from gluon fusion. The cross section is calculated to be $6.7^{+0.7}_{-0.9}$ pb at a top mass of 175 GeV/$c^2$ [1]. The calculated cross section decreases by approximately 0.2 pb for each 1 GeV/$c^2$ increase in the top mass over the range 170 GeV/$c^2$ $< M_{top} <$ 190 GeV/$c^2$. This measurement uses muon tagging for b-jet identification in order to reduce the background from $W$ plus multijet production. The measurement of the $t\bar{t}$ production cross section provides a test of the QCD calculations, and a significant deviation from the predicted cross section could signal beyond the standard model production mechanisms for $t\bar{t}$ pairs.

The CDF detector is described in detail in [2].

## II. DATA SAMPLE & EVENT SELECTION

This analysis is based on an integrated luminosity of 2034 pb$^{-1}$ collected with the CDFII detector between March 2002 and May 2007. The data are collected with an inclusive lepton trigger that requires an electron (muon) with $E_T > 18$ GeV ($p_T > 18$ GeV/$c$). From this inclusive lepton dataset we select events offline with a reconstructed isolated electron $E_T$ (muon $p_T$) greater than 20 GeV, missing $E_T > 30$ GeV and at least 3 jets with $E_T > 20$ GeV.

The dataset selected above, called "lepton+jets", is dominated by QCD production of W bosons with multiple jets. As a first stage of background reduction, we define a total event energy, $H_T$, as the scalar sum of the electron $E_T$, muon $p_T$, missing $E_T$ and jet $E_T$ for jets with $E_T >$ 8GeV and $|\eta| < 2.4$.

Even after the $H_T$ cut, the expected $S/B$ in $W+$ 3 or more jets events is only of order 1:1. To further improve the signal to background we identify events with one or more b-jets by searching for semileptonic decays of B hadrons into muons inside jets. This technique is called soft lepton tagging, or SLT.

In what follows, we refer to the $W$+3-or-more-jets sample after requiring $H_T$ >200 GeV, but before requiring a soft lepton tag, as the "pre-tag" sample. In 2034 pb$^{-1}$ we find 3903 pre-tag events, 1628 from $W \to \mu\nu$ and 2275 from $W \to e\nu$.

### A. Soft Lepton Tagging Algorithm

Muon identification at CDF proceeds by matching extrapolated tracks found in the central tracker to track segments reconstructed in the muon chamber. Matching is done in extrapolated position in the muon chamber drift direction and, where such information is available, in the coordinate along the chamber wires, and in the extrapolated slope compared to the slope of the reconstructed muon chamber track segment. The matching distributions, between the measured muon chamber hits and the extrapolated track, are a function of $p_T$, $\eta$ and $\phi$. The muon SLT algorithm uses a global $\chi^2$, $L$, built from the matching distributions, that separates muon candidates from background. In this analysis a jet is considered "tagged" if it contains an SLT muon with $p_T >$ 3 GeV/c, with $|L|$ <3.5 and within $\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$ <0.6 of the jet axis. The muon track is required to come within 5 cm in $z$ (the beam direction) of the primary event vertex. The SLT algorithm does not use calorimeter information and so is efficient for muon identification inside jets. Two sets of muon drift chambers are used by the SLT, "CMUP" which covers the region $|\eta| < 0.6$ and "CMX", which covers the region $0.6 < |\eta| < 1.0$.

Events are rejected if the isolated high $p_T$ lepton is a muon of opposite charge to an SLT muon tag that together with the SLT muon has an invariant mass consistent with a $J/\psi$, $\Upsilon$, or $Z^0$ decay, or a sequential, double-semi-leptonic $b \to c \to s$ decay.

### B. Total $t\bar{t}$ Acceptance

#### 1. *Geometric × Kinematic Acceptance*

The total acceptance is measured in a combination of data and Monte Carlo. The geometric times kinematic acceptance of the basic lepton+jets event selection is measured using the PYTHIA Monte Carlo program [3]. The efficiency for identifying the isolated, high $p_T$ lepton is scaled to the value measured in the data using the unbiased leg in $Z$-boson decays. The geometric times kinematic acceptance for 3-or-more-jets events is shown in Table I. The corrected acceptance includes the measured trigger efficiency and the data/MC ratio for tight lepton ID efficiencies (the scale factor) of the high $p_T$ lepton.

| | electron | muon(CMUP) | muon(CMX) |
|---|---|---|---|
| raw acceptance(%) | $4.02 \pm 0.01$ | $2.51 \pm 0.01$ | $1.090 \pm 0.005$ |
| corrected acceptance(%) | $3.71 \pm 0.01$ | $2.05 \pm 0.01$ | $0.946 \pm 0.004$ |

TABLE I: Geometric times kinematic acceptances (both raw and corrected) for $t\bar{t}$ events as a function of tight lepton type from PYTHIA Monte Carlo. The corrected acceptance includes the scale factors and trigger efficiencies of the tight lepton. The uncertainties listed are statistical only.

#### 2. *SLT Efficiency*

The muon identification efficiency of the SLT algorithm is measured in data using $J/\psi$ and $Z^0$ events. The measured efficiency vs. $p_T$ for CMUP and for CMX are shown in Figures 1 and 2

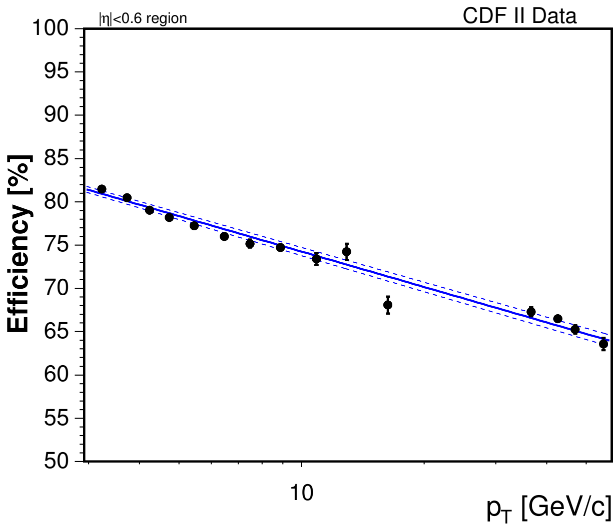


FIG. 1: The SLT efficiency for CMUP as a function of $p_T$ as measured from $J/\psi$ and $Z^0$ data for $|L|$ <3.5. The dotted lines are the 1 $\sigma$ uncertainties on the fit that are used in the evaluation of the systematic uncertainty.

The decrease in efficiency with increasing $p_T$ is a result of non-Gaussian tails in the components of the global $\chi^2$. The efficiency measurement is dominated by isolated muons whereas the muons in b-jets tend to be surrounded by other tracks. We have studied the dependence of the efficiency on

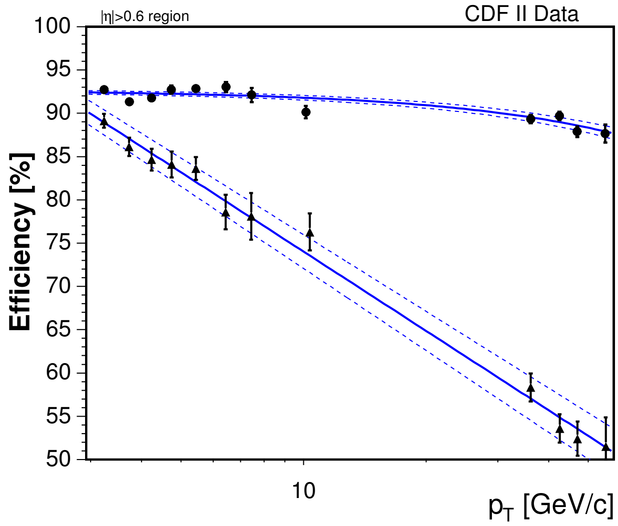


FIG. 2: The SLT efficiency for CMX as a function of $p_T$ as measured from $J/\psi$ and $Z^0$ data for $|L|$ <3.5. The upper plot shows the efficiency in the CMX arches while the lower plot shows the efficiency in the CMX miniskirt. The dotted lines are the $1\sigma$ uncertainties on the fit that are used in the evaluation of the systematic uncertainty.

$N_{trk}$, the number of tracks above 1 GeV/$c$ in a cone of $\Delta R$=0.4 around the muon track, and find no significant efficiency loss.

The measured efficiencies shown in Figures 1 and 2 are applied directly in the Monte Carlo $t\bar{t}$ samples. The efficiency for finding one or more SLT tags in a $t\bar{t}$ event ("tagging efficiency") for 3-or-more-jets events is shown in Table II. These efficiencies include mistags in $t\bar{t}$ events (i.e. tags that do not come from semileptonic-heavy-flavor decays), which contribute approximately 20% to the total tagging efficiency. The total $t\bar{t}$ detection efficiency is the product of the geometric×kinematic acceptance and the tagging efficiency, which is also shown in Table II.

| | electron | muon(CMUP) | muon(CMX) |
|---|---|---|---|
| Event Tagging Eff.(%) | $14.02 \pm 0.08$ | $13.07 \pm 0.10$ | $13.38 \pm 0.16$ |
| Total Detection Eff.(%) | $0.520 \pm 0.003$ | $0.268 \pm 0.002$ | $0.127 \pm 0.002$ |

TABLE II: $t\bar{t}$ event tagging efficiency and total detection efficiency for SLT muons in the signal region from Pythia Monte Carlo samples. Uncertainties are statistical only.

## III. BACKGROUNDS

The dominant background for this analysis is QCD production of $W$-boson plus multijet events. These events enter the signal sample when either one of the jets is a b-jet, or a light quark jet is misidentified (mistagged) as containing a semileptonic $B$-hadron decay. We measure the two components of this background separately; the mistags are measured by constructing a "mistag matrix" using samples of pions, kaons, and protons, while the $W$+heavy-flavor background is determined from Monte Carlo. The mistag matrix parameterizes the probability that a track with a given $p_T$ and $\eta$ will satisfy the SLT requirement of $|L|$ <3.5. The mistag probability is approximately 0.4% per taggable track ($p_T > 3$ GeV/$c$, $\Delta R$ <0.6 from a jet axis and fiducial to the muon chambers).

We test the accuracy of the mistag matrix for predicting SLT muon tags by using it to predict the number of tags using a variety of jet samples. We check jets in $\gamma$+jets events, events triggered on a jet with thresholds of 20, 50, 70 and 100 GeV and events triggered on the scalar sum of transverse energy in the detector. Each of these samples contains some muons from semileptonic-heavy-flavor decays that the the matrix is not intended to predict and must therefore be removed (or at least reduced) before making our test. We reduce the heavy-flavor content of these samples by requiring that there be no displaced vertex tags using the silicon tracker [4] in the event and making a cut on the maximum impact parameter significance, $d_0/\sigma_{d_0}$ allowed for any track in a jet. We test the agreement between predicted and observed tags as a function of jet $E_T$ (see Figure ??) and weight the results by jet $E_T$ spectrum of the $W$+3-or-more-jets sample (shown in Figure??) to determine an overall estimate of the accuracy of the prediction. This weighted result for predicted tags minus observed tags divided by predicted tags is $(0.1 \pm 4.4)\%$, from which we assign a systematic uncertainty of $\pm 5$ on the mistag matrix prediction.

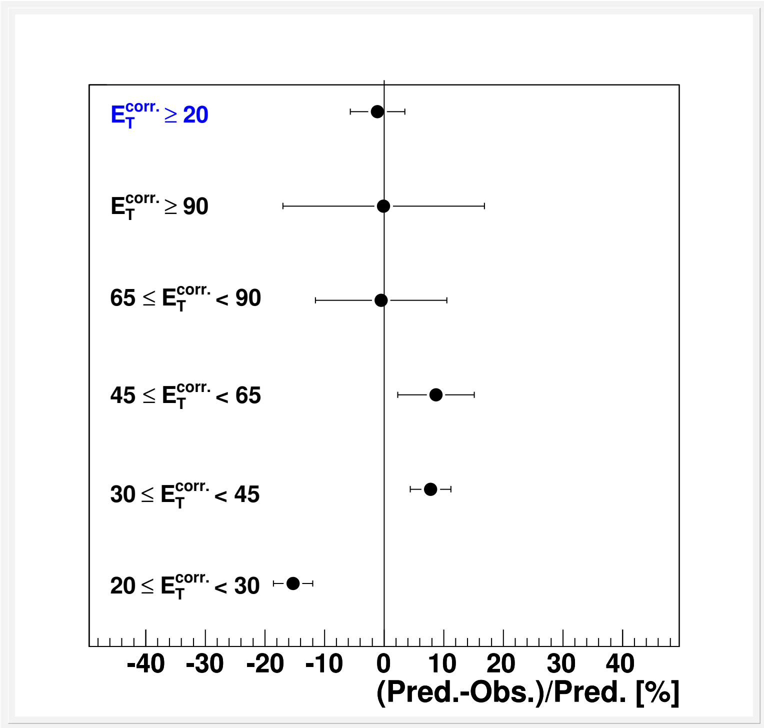


FIG. 3: The percent difference between the predicted and measured tags in different jet $E_T$ ranges.

The background contribution from $W$+heavy-flavor ($Wb\bar{b}$, $Wc\bar{c}$, and $Wc$) events is estimated using *Alpgen* Monte Carlo samples. The fraction of pre-tag $W$+jets events that include heavy-flavor-quark jets is measured in the Monte Carlo and multiplied by a scale factor that adjusts for differences between data and Monte Carlo. The *real* SLT tagging efficiency (i.e. the efficiency to tag muons from semileptonic heavy-flavor decays, not mistags) is also measured in the Monte Carlo. The results are then applied to the pre-tag data sample with the predicted number of tagged $W$+heavy-flavor events being equal to the number of pre-tag events times the (corrected) fraction of those having heavy flavor times the tagging efficiency.

The other substantial background in this analysis comes from QCD events without $W$ bosons ("QCD background"). These events are typically QCD jet events where one jet has faked a high-$p_T$ lepton and mismeasured energies produce apparent missing $E_T$. We measure this QCD background in two steps. We first estimate the fraction, $F_{QCD}$, of QCD events in the signal region by extrapolating the number

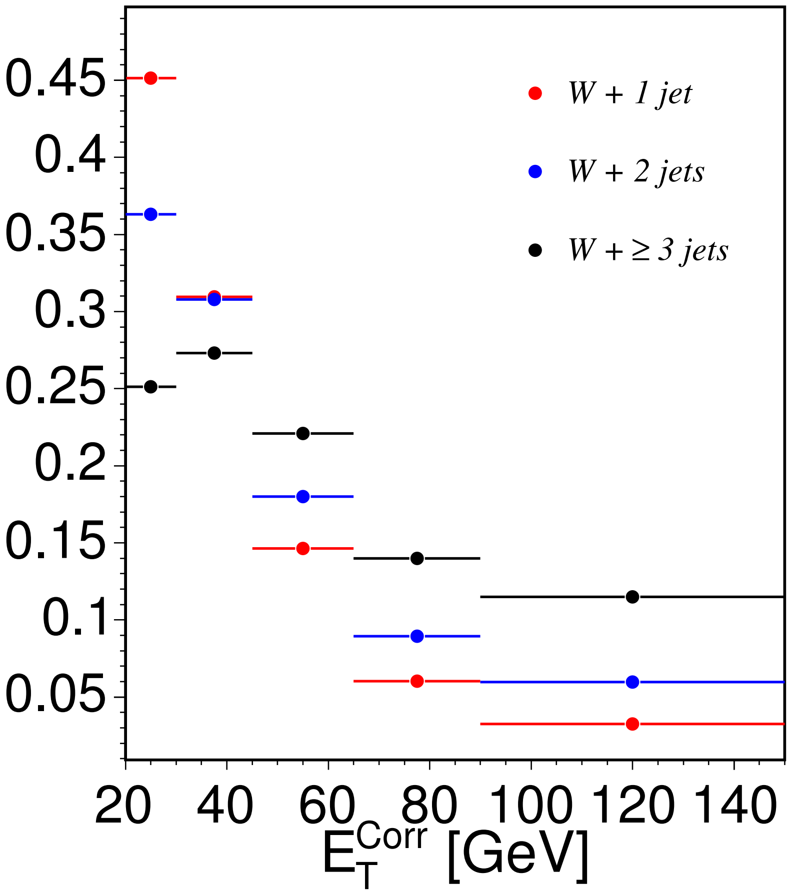


FIG. 4: The $E_T$ distribution of jets in $W$+jets events measured in ALPGEN Monte Carlo [5]. We weight the results of tests of the mistag matrix prediction (shown in Fig. 3) according to the distribution in $W$ + 3-or-more-jets events in determining the systematic uncertainty of the mistag background prediction.

of events with an isolated lepton and low missing $E_T$ into the signal region of large missing $E_T$. We then multiply the expected number of QCD events by the mistag matrix (which is designed to be applied to jets in $W$+jets events) and corrected by a factor $k$, because the missing $E_T$ in QCD events is typically generated by hadrons that are incompletely absorbed in the calorimeter and semileptonic decays of heavy flavor, both of which enhance the tag rate over generic jets. In order to get the $k$ factor, We measure the tags for QCD events using QCD enriched events with a non-isolated primary electron or muon and missing $E_T > 30$ GeV and then compare to the predicted tags from the mistag matrix.

Residual Drell-Yan background is estimated from the data by extrapolating the number of events inside the $Z$-mass window that pass the selection cuts, including the SLT tag, to events in the signal region outside the $Z$-mass window. Other, small backgrounds from a variety of sources are estimated using the Monte Carlo.

The backgrounds as a function of jet multiplicity are summarized in Table IV.

## IV. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties in this analysis come from Monte Carlo modeling of the geometrical and kinematic acceptance, knowledge of the SLT tagging efficiency, the effect on the acceptance of the uncertainty on the jet energy scale, uncertainties on the background predictions, and the uncertainty on the luminosity.

Monte Carlo modeling of geometrical and kinematic acceptance include effects of PDFs, ISR and FSR, and jet energy scale and lepton ID efficiency. These are estimated by comparing different choices for PDFs, varying ISR, FSR and the jet energy scale and lepton ID efficiency in the Monte Carlo and comparing the *Pythia* generator with *Herwig* [6]. The total systematic uncertainty due to these factors is 5.7%

There are several factors that contribute to the systematic uncertainty on the SLT tagging efficiency. The uncertainty due to the limited knowledge of the $p_T$ dependence is determined by varying the efficiency curves used in the $t\bar{t}$ Monte Carlo for the tagging efficiency measurement according to the upper and lower bands in Figures 1 and 2. We find that the tagging efficiency for $t\bar{t}$ changes by $\pm 1\%$ from its central value when varying the efficiency curves, and take this as a systematic uncertainty. An additional systematic uncertainty for the tagging efficiency comes from the fact that we implicitly use the Monte Carlo tracking efficiency for taggable tracks. As these tracks can be in dense environments in or near jets, we expect it to be less than 100%. Studies done by embedding Monte Carlo tracks in jets in data and jets in Monte Carlo events indicate that the Monte Carlo tracking efficiency in dense environments is a few percent higher than in data. We assign a 5% systematic uncertainty to the tagging efficiency for this effect. We that these uncertainties affect only the *real* part of the tagging efficiency (i.e. not the part due to tagging non-muons which is accounted for by the mistag matrix), and we only apply this uncertainty to the fraction of the tagging efficiency that comes from real heavy-flavor muons ($\sim 79\%$)). Adding the contributions in quadrature gives us an overall systematic uncertainty for the tagging efficiency of 5.1%. This uncertainty and the acceptance uncertainty are added in quadrature in Table III.

Uncertainties on the mistag matrix are determined by the level of agreement between observed tags and predictions in a variety of samples. This was described in Section III. We take 5% as the uncertainty on the mistag prediction prediction.

There are three sources of uncertainty for the $Wb\bar{b} + Wc\bar{c} + Wc$ background prediction: from the selection of ALPGEN settings (23%), errors assosiated with the data/MC scale factor (30%), and uncertainty on the tagging efficiency. The uncertainty on the tagging efficiency comes from the same sources as that for $t\bar{t}$ described above, and we roll it into the uncertainty on the $t\bar{t}$ tagging efficiency because of their correlation.

Uncertainties on the QCD background prediction are determined by the level of agreement between predicted and measured events in a background-rich region ("test region") just outside the signal region. An 11% (120%) systematic uncertainty is assigned to the measurement of the QCD fraction ($F_{QCD}$) in signal region for electrons (muons) given conservatively by the worst agreement of the test region predictions. We fold this in with the statistical uncertainty on the $F_{QCD}$ determination, the uncertainty on the correction factor $k$ and the 5% systematic uncertainty due to the application of the fake matrix. The total QCD background uncertainty is 124% and 19% for muons and electrons, respectively. To calculate the effect of the uncertainties on the QCD background on the cross section, we have accounted for the correlation in the QCD uncertainty with that of the mistag prediction.

The systematic uncertainty on the small Drell-Yan background is dominated by its statistical uncertainty. Uncertainties on the Monte Carlo background predictions come from uncertainties in the cross sections for the various processes and from the statistics of the Monte Carlo samples.

The systematic uncertainties are summarized in Table III. An additional systematic due to the uncertainty on the luminosity determination (5.9 %), is treated separately.

| Source | Fractional Sys. Uncert. | Contribution to $\sigma_{t\bar{t}}$ |
|---|---|---|
| Acceptance Modeling and SLT Tagging Efficiency | 7.7 % | +8.3 % −7.5 % |
| Fake Matrix Prediction | 5 % | 3.6 % |
| $Wbb + Wc\bar{c}$+$Wc$ Prediction | 38 % | 5.3 % |
| QCD Prediction | 19 %(e) 124 %($\mu$) | 1.1 % |
| Drell-Yan and other MC backgrounds | 11 % | 0.4 % |
| Total Systematic Uncertainty | | +10.5 % −10.0 % |

TABLE III: Summary of systematic uncertainties.

## V. RESULTS

Table IV shows a summary of the background estimates for each jet bin and the number of SLT tagged events. The total background and the $t\bar{t}$ expectation are also listed. The line labeled "Corrected Background" includes an iterative correction to the background estimate. This correction is needed because we apply the mistag matrix to the pre-tag events in order to estimate the mistags. Since the events before tagging include some $t\bar{t}$, this results in an over estimate of the background for which we correct. Similarly the $W$+heavy-flavor and QCD backgrounds are proportional to the number of pre-tag events, and we correct for that as well.

| | $H_T \geq 0$ GeV | | $H_T \geq 200$ GeV | | |
|---|---|---|---|---|---|
| Background | 1 jet | 2 jet | 3 jets | $\geq 4$ jets | $\geq 3$ jets |
| Events before tagging | 75595 | 18264 | 2587 | 1120 | 3707 |
| Fake | 622±31 | 226±12 | 53.0±2.7 | 31.4±1.6 | 84.5±4.3 |
| $Wbb$+$Wcc$+$Wc$ | 145±55 | 66.6±25.3 | 15.3±5.8 | 8.5±3.2 | 23.0±8.7 |
| QCD mutijet | 91.9±16.5 | 44.9±10.4 | 7.0±1.5 | 4.1±0.9 | 11.1±2.4 |
| $WW$+$WZ$+$ZZ$ | 3.80±0.44 | 6.98±0.66 | 1.21±0.23 | 0.64±0.14 | 1.88±0.30 |
| $Z \rightarrow \tau^+\tau^-$ | 2.65±0.57 | 1.54±0.43 | 0.65±0.28 | 0.13±0.05 | 0.65±0.27 |
| Drell-Yan | 6.02±1.25 | 4.12±0.88 | 0.82±0.44 | 0.00±0.19 | 0.82±0.48 |
| Single top | 4.36±0.39 | 9.00±0.66 | 2.14±0.18 | 0.57±0.06 | 2.71±0.23 |
| Total Background | 876.5±53.6 | 359.0±24.0 | 80.2±5.4 | 45.3±3.0 | 124.6±8.2 |
| Corrected Background | – | – | 79.5±5.3 | | 79.5±5.3 |
| $t\bar{t}$ Expectation ($\sigma$ =6.70) | 2.60±0.33 | 23.5±1.8 | 50.1±3.6 | 74.2±6.5 | 124.3±9.1 |
| Total Background + $t\bar{t}$ | 879.1±53.6 | 382.5±24.1 | 203.9±10.6 | | 203.9±10.6 |
| Tagged events | 892 | 384 | 142 | 106 | 248 |

TABLE IV: Number of tagged events and the background summary. The $H_T > 200$ GeV requirement is made only for events with at least 3 jets.

We calculate the cross section in the usual way as

$$\sigma_{t\bar{t}} = \frac{N_{obs} - N_{bckg}}{A_{t\bar{t}} \times \int \mathcal{L}dt} \tag{1}$$

where $N_{obs}$ is the number of events with $\geq$ 3 jets that are tagged with at least 1 SLT, $N_{bckg}$ is the corrected background and $A_{t\bar{t}}$ is the total acceptance (geometrical times kinematic times tagging efficiency), taken from Tables I and II For events with 3-or-more jets, the total denominator is $18.56 \pm 1.44$ pb$^{-1}$.

For $t\bar{t}$ events in 3-or-more jets, we find a cross section of

$$9.1 \pm 1.1(\text{stat})^{+1.0}_{-0.9}(\text{sys})\pm 0.6(\text{lum})\text{pb}.$$

Figure 5 shows the number of tags in $W+1, 2, \geq 3$ jet events together with a histogram representing the components of the background and $t\bar{t}$ expectation for a cross section of 9.1 pb.

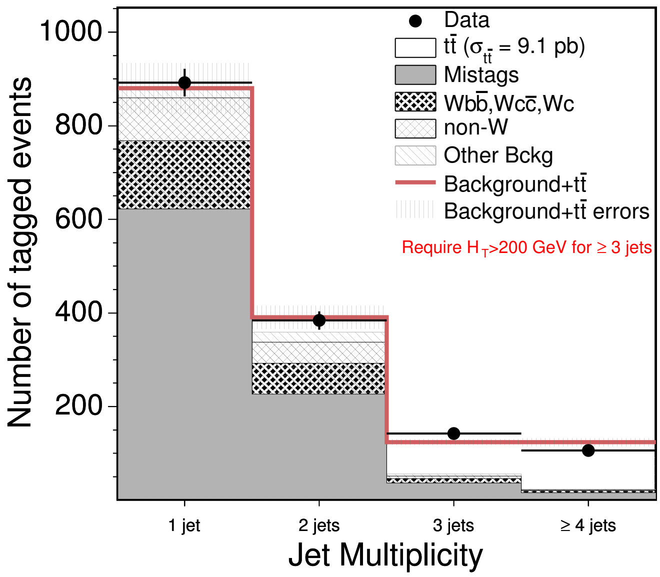


FIG. 5: The expected background and observed tags in $W+$ 1, 2, 3 and 4-or-more-jets events. The background is corrected for the $t\bar{t}$ content of the pretagged sample. The $t\bar{t}$ expectation is normalized at 9.1 pb.


### Acknowledgments

We thank the Fermilab staff and the technical staffs of the participating institutions for their vital contributions. This work was supported by the U.S. Department of Energy and National Science Foundation; the Italian Istituto Nazionale di Fisica Nucleare; the Ministry of Education, Culture, Sports, Science and Technology of Japan; the Natural Sciences and Engineering Research Council of Canada; the National Science Council of the Republic of China; the Swiss National Science Foundation; the A.P. Sloan Foundation; the Bundesministerium für Bildung und Forschung, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Science and Technology Facilities Council and the Royal Society, UK; the Institut National de Physique Nucleaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Educación y Ciencia and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.


---


[1] M. Cacciari, et al., The $t\bar{t}$ Cross-section at 1.8 TeV and 1.96 TeV: A Study of the Systematics due to Parton Densities and Scale Dependence, JHEP **404**, 68 (2004).

[2] F. Abe, et al., Nucl. Instrum. Methods Phys. Res. A **271**, 387 (1988); D. Amidei, et al., Nucl. Instum. Methods Phys. Res. A **350**, 73 (1994); F. Abe, et al., Phys. Rev. D **52**, 4784 (1995); P. Azzi, et al., Nucl. Instrum. Methods Phys. Res. A **360**, 137 (1995); The CDFII Detector Technical Design Report, Fermilab-Pub-96/390-E

[3] T. Sjostrand et al., High-Energy-Physics Event Generation with PYTHIA 6.1, Comput. Phys. Commun. **135**, 238 (2001).

[4] D. Acosta, et al., Measurement of the $t\bar{t}$ Cross Section in $p\bar{p}$ Collisions at $\sqrt{s}$=1.96 TeV Using Lepton+Jets Events with Secondary Vertex b-tagging, Phys. Rev. D71, 052003(2005)

[5] M. Mangano et al., ALPGEN, a generator for hard multiparton processes in hadronic collisions, JHEP **07**, 1 (2003).

[6] G. Corcella et al., HERWIG 6: An Event Generator for Hadron Emission Reactions with Interfering Gluons (including supersymmetric processes), JHEP **01**, 10 (2001).