CDF Note 9462

Measurement of the Top Pair Cross Section in the Lepton Plus Jets Decay Channel with 2.7 fb^{-1}

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: August 5, 2008)

The cross-section for pair produced top quarks in the lepton plus jets channel has been measured in 2.7 fb^{-1} of collected data from the high p_t lepton triggers. To improve signal significance, the measurement utilizes a bottom "tagging" algorithm, SecVtx, to reconstruct secondary vertices as evidence of heavy flavor decay. Events are required to have at least one "tagged" jet. The result is $\sigma_{t\bar{t}} = 7.2 \pm 0.4_{stat} \pm 0.5_{sys} \pm 0.4_{lumi}$.

I. INTRODUCTION

We present a measurement of the top pair cross-section using $2.7 fb^{-1}$ of collected data from the CDF detector [1]. Data is selected using an inclusive high Pt lepton trigger requiring an electron or muon with at least 20 GeV. In addition, we require missing transverse energy $\not\!\!E_T > 25$ GeV, at least three jets present in the event with Et$>$ 20 GeV, and the scalar sum of the transverse energy (Ht) of the jets, lepton, and $\not\!\!E_T$ to be greater then 250 GeV. A "tagging" algorithm, SecVtx, is used to find a displaced secondary vertex as evidence of a bottom quark decay [3].

In general, the cross section is calculated with the formula:

$$\sigma_{t\bar{t}} = \frac{N_{data} - N_{bkg}}{A \cdot \epsilon \cdot L} \tag{1}$$

where, N_{data} is the amount of collected data in the signal region, N_{bkg} is the predicted background content, A is the acceptance of $t\bar{t}$ events before requiring a tag, ϵ is the tagging efficiency, and L is the luminosity. Monte Carlo is relied upon to estimate acceptance and tagging efficiency, though with corrections applied to account for differences in trigger efficiencies, tagging, and mis-tagging. To measure the cross-section, we construct a likelihood based upon the data, the top cross-section, and the predicted background for that cross-section. The measured value and statistical uncertainty is extracted from this likelihood. Systematic uncertainties are calculated by re-performing the measurement under $\pm 1\sigma$ deviations for a particular uncertainty.

II. BACKGROUND ESTIMATE

We take a data-driven approach to backgrounds due to inadequacies in the Monte Carlo to model heavy flavor associated with the production of a W boson, tagging of bottom jets, and difficulties associated with modeling the QCD contribution. The technique is sequential in that each step depends on the previous. The final result is a complete prediction for the process content in the lepton plus jets data sample. In the following we will go step by step through the procedure.

A. Monte Carlo Based Backgrounds

A few of the backgrounds which are considered a small contribution to the overall process content and $t\bar{t}$ (which is an important point as we will discuss later) are calculated based on Monte Carlo efficiencies. Several electroweak processes contribute to the lepton plus jets sample such as WW, WZ, ZZ, and Z $\rightarrow$ *jets* events. They exist in the sample because each process can produce a real lepton and neutrino, as well as a number of jets. The numbers in our sample are estimated using the theoretical cross section, the luminosity of the sample, trigger efficiency, and an overall selection efficiency derived from Monte Carlo simulation of the processes in question. The calculated number in our sample is given by

$$N_{p\bar{p}\rightarrow X} = \sigma_{p\bar{p}\rightarrow X} \cdot A \cdot \int dt \cdot \mathcal{L} \tag{2}$$

$$N^{tag}_{p\bar{p}\rightarrow X} = \sigma_{p\bar{p}\rightarrow X} \cdot A \cdot \epsilon \cdot \int dt \cdot \mathcal{L} \tag{3}$$

where $\sigma_{p\bar{p}\rightarrow X}$ is the theoretical cross section, $\int dt \cdot \mathcal{L}$ is the total luminosity, A is the pre-tagged selection acceptance derived from Monte Carlo, and ϵ is the tagged selection efficiency. As for top, the acceptance and tagging efficiencies are corrected for trigger efficiencies and tagging.

B. Non-W Based Background Estimate

To determine the non-W fraction in both the pretag and tagged sample, we fit the $\not\!\!E_T$ distribution of a non-W template and a MC signal template to data.

Both data and model templates are fitted to the $\not\!\!E_T$ distribution of isolated pretag data events using a binned likelihood fit. Once the fraction is calculated the normalization is simply:

$$N_{QCD}^{pretag} = F_{QCD} \cdot N_{pretag} \tag{4}$$

The same general procedure is performed for the tagged sample.

$$N_{QCD}^{tag} = F_{QCD} \cdot N_{tag} \tag{5}$$

C. W + Heavy Flavor

In the pretag data sample, W plus jets is the dumping ground for all events that are not considered QCD, electroweak, or top. The W plus jets normalization is calculated by subtracting the MC-based processes and the QCD from data as shown in equation 6.

$$N_{W\,+\,Jets}^{pretag} = N_{pretag} \cdot (1 - F_{QCD}^{pretag}) - N_{ewk}^{pretag} - N_{top}^{pretag} \tag{6}$$

For the tagged estimate, the W plus jets sample is broken down into two categories: heavy and light flavor. Each of these processes produces a tagged jet very differently and therefore requires different treatment in calculating the normalization.

The contribution of the heavy flavor background to our signal region is calculated by equation 7.

$$N_{W+hf}^{tag} = (N_{pretag}(1 - F_{QCD}) - N_{EW} - N_{singletop} - N_{t\bar{t}}) \cdot f_{HF} \cdot K \cdot \epsilon \tag{7}$$

The number of events predicted in QCD, Electroweak, singletop, and $t\bar{t}$ is subtracted from the pretag sample, leaving an estimate for the number of events with a W-boson. The fraction of these events with jets matched to heavy flavor quarks, f_{HF}, is calculated from a detailed Monte Carlo simulation Alpgen [4], which includes all possible processes contributing to the production of a single real W-boson. This fraction is corrected by a scale factor which is a correction to the Monte Carlo heavy flavor fraction. The HF correction factor is calculated in the 1 jet bin and applied to the rest of the sample. ϵ is the tagging efficiency. f_{HF} and ϵ are calculated for $Wb\bar{b}$, $Wc\bar{c}$, and Wc separately, which define the rates for each of these processes. Only the heavy flavor fraction relies on Monte Carlo, the absolute normalization is derived from the pretag sample in data. The HF correction is derived by a Neural Network fit to variables sensitive to jets matched to heavy flavor and light flavor.

D. Mistags

A secondary vertex is mistakenly reconstructed when poorly reconstructed tracks seem to cross each other near the origin. A secondary vertex that does not originate from heavy flavor quarks is called a mistag.

The negative tag rate is found to be well parametrized by five jet variables (jet E_t, number of good SVX tracks, sum of all jet E_t in the event, jet η, jet ϕ) and measured in a very high statistics sample derived from triggers on 50 GeV jets. In any subsequent analysis this parametrization then gives the probability that a jet with given values of the tag parametrization variables will be negatively tagged. The negative tag probability of an event is taken to be the sum of the probabilities of all the jets in the event. Studies in large control samples derived from jet triggers with different energy thresholds (20 GeV, 75 GeV, 100 GeV) show good agreement between the prediction and the actual number of negative tags.

This technique is applied to estimate the number of events in our sample due to mistags in W + light flavor events. The predicted number of background events from W + light flavor (W+lf) processes is:

$$N_{W+lf}^{tag} = \frac{N_{-}}{N_{pre}} \cdot (N_{pre} - N_{pre}^{t\bar{t}} - N_{pre}^{QCD} - N_{pre}^{W+hf} - N_{pre}^{EW} - N_{pre}^{singletop}) \tag{8}$$

The predicted amount of $t\bar{t}$, QCD, W+hf, Electroweak, and single top background events is subtracted from the total pretag sample leaving an estimate for the W+lf fraction. The predicted number of mistagged W+lf events is the W+lf fraction multiplied by the predicted amount of mis-tagged events from the pretag data.

E. Full Background Prediction

The following is the background estimate used in our top cross section measurement utilizing 2.7 fb^{-1} of collected data. Inclusive trigger tables for ≥ 1 Tags are shown in Table I. A histogram representing the predicted number of events and data is shown in Figure 1.

Process	1jet	2jets	3jets	4jets	5jets
Pre-tag Data	2553	3199	1988	1030	318
Wbb	18.9 ± 5.8	56.6 ± 17.5	42.5 ± 13.1	16.8 ± 5.8	5.4 ± 2.0
Wcc	8.8 ± 2.8	24.6 ± 7.8	20.7 ± 6.5	9.0 ± 3.1	3.0 ± 1.1
Wc	9.5 ± 3.0	22.5 ± 7.1	12.6 ± 4.0	4.1 ± 1.4	1.1 ± 0.4
Mistags	45.4 ± 4.9	60.7 ± 8.7	33.5 ± 5.5	10.2 ± 3.2	2.7 ± 1.3
Non-W	12.2 ± 4.4	37.3 ± 11.9	20.1 ± 6.8	5.6 ± 4.8	2.0 ± 2.3
Z+jets	1.6 ± 0.2	4.6 ± 0.6	4.3 ± 0.5	1.8 ± 0.2	0.6 ± 0.1
WW	1.2 ± 0.1	6.1 ± 0.8	5.1 ± 0.6	2.2 ± 0.3	0.8 ± 0.1
WZ	0.4 ± 0.0	1.9 ± 0.2	1.5 ± 0.2	0.7 ± 0.1	0.2 ± 0.0
ZZ	0.0 ± 0.0	0.2 ± 0.0	0.3 ± 0.0	0.2 ± 0.0	0.1 ± 0.0
Single Top (s-channel)	0.4 ± 0.0	10.1 ± 1.0	6.4 ± 0.6	2.1 ± 0.2	0.5 ± 0.1
Single Top (t-channel)	0.1 ± 0.0	8.4 ± 0.7	6.1 ± 0.5	2.0 ± 0.2	0.4 ± 0.0
$t\bar{t}$ (7.2pb)	4.2 ± 0.6	78.8 ± 10.4	271.5 ± 35.8	337.1 ± 44.3	120.5 ± 15.8
Total Prediction	102.7 ± 13.3	311.8 ± 37.1	424.6 ± 44.4	391.8 ± 46.1	137.3 ± 16.5
Observed	104	308	418	396	138

TABLE I: Background Normalizations for ≥ 1 Tag, ≥ 250 GeV, and $\not{E}_T > 25$ GeV

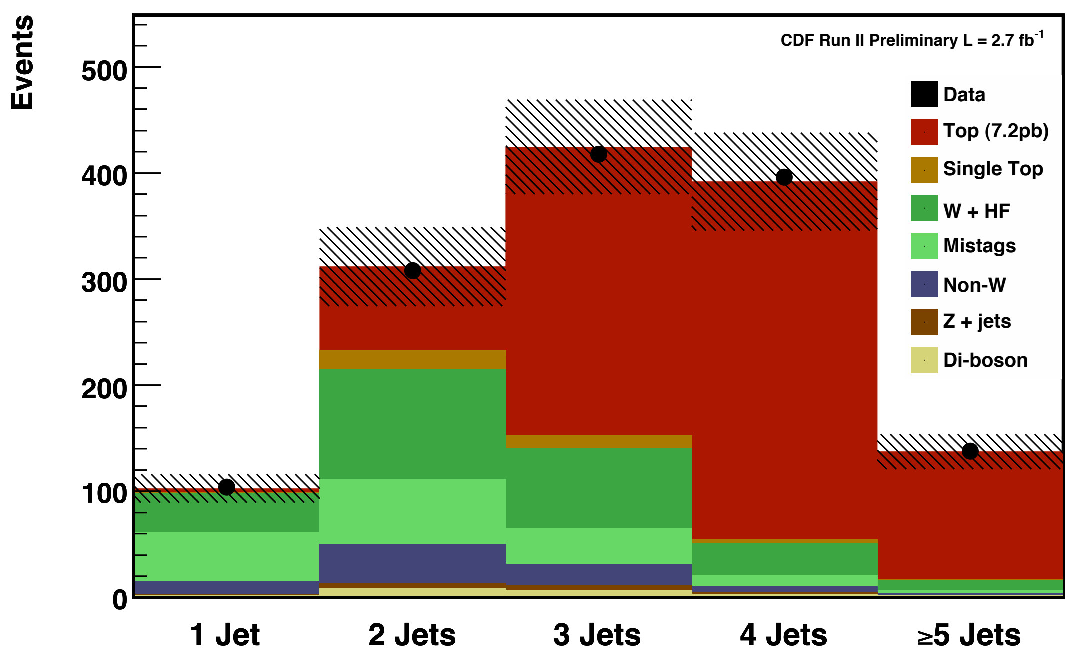

FIG. 1: Predicted vs Observed as a function of jet multiplicity

III. CALCULATING THE CROSS-SECTION

With the background estimate in hand it would appear straightforward to calculate the cross-section. Unfortunately, because the background estimate is dependent on the top cross-section, extracting the measured value is not so simple.

Instead, we construct a poisson likelihood where we take into account the background dependence. The likelihood is

$$-2 \cdot lnL = -2 \cdot (N_{data} \cdot ln(D \cdot \sigma_{t\bar{t}} + B) - ln(N_{data}!) - (D \cdot \sigma_{t\bar{t}} + B)) \quad (9)$$

where D is the denominator of equation 1 ($A \cdot \epsilon \cdot L$), N_{data} is the amount of measured data, and B is the background estimate for a given top cross-section. The likelihood is calculated for several values of the cross-section and the resulting points are fit to a second order polynomial. The minimum of this curve is taken as the measured value and the . The result for our optimized selection, $H_t \geq 250$ GeV and $\not{E}_T \geq 20$ GeV, is shown in Figure 2. The curve is not perfectly parabolic as the tagged sample is actually not gaussian, but binomial (the pretags are gaussian). To account for this we keep a somewhat narrow window along the measured value of ± 1 pb and fit only inside that window. The measured value for our optimized cuts is:

$$\sigma_{t\bar{t}} = 7.2 \pm 0.4_{stat} \ \ pb \quad (10)$$

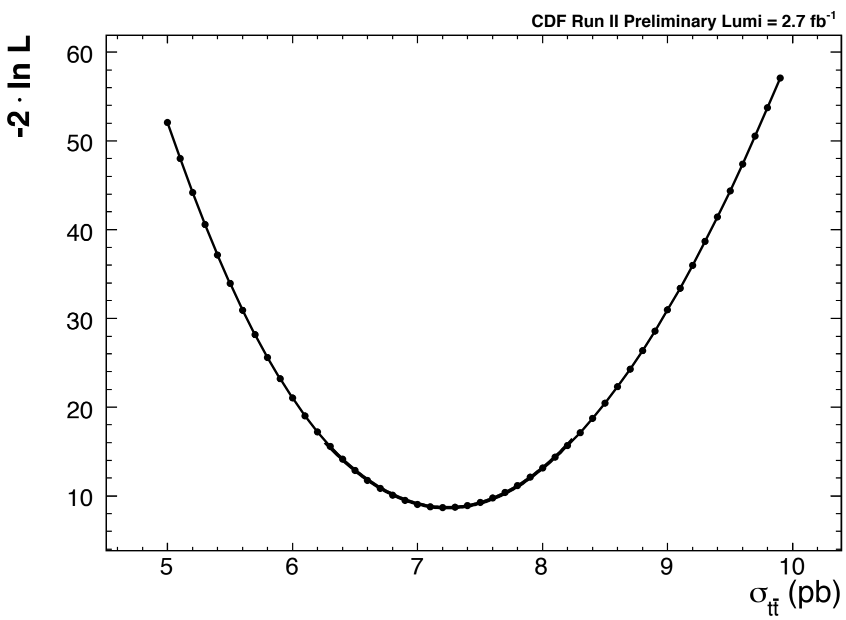

FIG. 2: Likelihood Curve For Measured Cross-Section

IV. SYSTEMATICS

Systematic uncertainties in our measure result are calculated by varying a given parameter within it's uncertainty and redoing the entire measurement. Each systematic is described below along with any relevant quantities. The individual evaluated systematic uncertainties are shown in Figure 3 at the end of the section.

A. Jet Energy Scale

The energy of jets measured by the calorimeters is subject to multiple systematic uncertainties. We study the effect on the measurement by varying the JES for our top signal Monte Carlo and background models and then re-performing the measurement. The effect of JES on this measurement is mainly through the acceptance of signal and background.

B. Initial/Final State Radiation

The measured value will be effected if we are over or under estimating the amount of initial or final state radiation present in top events. To study this effect, we replace our standard top Monte Carlo model with top Monte Carlo where the radiation has been increased or decreased and the measurement is redone.

C. Tagging

Because Monte Carlo does not model SecVtx tagging properly, a scale factor is applied to each tagged jet matched to heavy flavor, and the corresponding event then re-weighted. The scale factor is derived from data and has an uncertainty associate with it which leads to a systematic on the measurement. The effect on the measured value is calculated by fluctuating the scale factor within it's uncertainty, applying it to each appropriate jet, calculating the new event weights, and repeating the measurement.

D. Mis-tags

Mistags are so badly modeled in Monte Carlo that we scrap any mis-tagged jet and use a data-based parameterization called the mistag matrix to predict the probability that any given jet is mis-tagged. The mistag rate on any jet is fluctuated within error and the entire measurement is repeated to quantify the effect.

E. QCD Fractions

To estimate the uncertainty on the QCD fraction, fits are performed with different binning and different models. The resulting difference in the fits is 30% which is taken as a systematic uncertainty in the measurement.

F. Heavy Flavor Corrections

The correction to the heavy flavor fractions has an uncertainty derived from the Neural Network fits in the 1 and 2 jet bin as well as the fits to bottom and charm separately. A 30% uncertainty is taken on the derived correction to cover the range of fitted values.

G. Parton Shower Modeling

Differences in Monte Carlo shower models are studied simply by replacing our $t\bar{t}$ PYTHIA model with the other most popular generator, HERWIG, and repeating the measurement [5] [6]

H. Trigger Efficiency

Detector specific corrections are applied to the Monte Carlo to more correctly model the relative trigger efficiencies between CEM, CMUP, and CMX events. The corrections are data-derived from Z events and have a small uncertainty associated with them. There are two types of corrections, trigger ID and trigger efficiencies. Each are fluctuated with their uncertainty, separately, and the resulting errors are added in quadrature.

I. PDF

Uncertainty in the parton distribution function are evaluated by a re-weighting scheme at the Monte Carlo Truth level. PDF's are reweighted in our signal Monte Carlo to simulate 46 different PDF parameterizations. The measurement is performed for each different parameterization. The result is shown in Figure 3.

SYSTEMATIC	$\Delta\ \sigma$ pb	$\Delta\ \sigma / \sigma$ %
JET ENERGY SCALE	0.16	2.2
BOTTOM TAGGING	0.38	5.2
CHARM TAGGING	0.08	1.1
MIS-TAGS	0.15	2.1
HEAVY FLAVOR CORRECTION	0.23	3.2
LUMINOSITY	0.42	5.8
QCD FRACTION	0.02	0.2
PARTON SHOWER MODELING	0.13	1.8
INITIAL/FINAL STATE RADIATION	0.04	0.6
TRIGGER EFFICIENCY	0.05	0.6
PDF	0.06	1.0
TOTAL	0.67	9.3

CDF Run II Preliminary L=2.7 fb^{-1}

FIG. 3: Systematic Uncertainties

J. Luminosity

The uncertainty on our calculated luminosity is unfortunately also our largest systematic, which is derived from detector accuracy and the uncertainty on the theoretical cross section for inelastic $p\bar{p}$ collisions. The uncertainty on the luminosity is 5.8%. The luminosity used in the measurement is fluctuated within this uncertainty and the measurement redone.

V. RESULT

Extracting the result from the likelihood and adding the systematic uncertainty we find the cross-section for $2.7fb^{-1}$ using ≥ 1 Tight SecVtx Tagged events in the lepton plus jets channel is:

$$\sigma_{\mathbf{t\bar{t}}} = \mathbf{7.2 \pm 0.4_{stat} \pm 0.5_{sys} \pm 0.4_{lum}\ \ pb} \tag{11}$$

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 Comisión Interministerial de Ciencia y Tecnología, Spain; the European Community's Human Potential Programme; the Slovak R&D Agency; and the Academy of Finland.

[1] F. Abe et al., Nucl. Instrum. Methods Phys. Res A 271, 387 (1988)
D. Amidei, et al, Nucl. Imstrum. 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)
CDFII Technical Design Report, FERMILAB-PUB-96/390-E
[2] CDF Collaboration, "Measurement of the $t\bar{t}$ Production Cross Section in $p\bar{p}$ Collisions at $sqrts = 1.96$ TeV using Lepton + Jets Events with Secondary Vertex b-tagging" ,Conference Note 8795
[3] T.Affolder et al., Phys Rev D, 64, 032002 (2001)
[4] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, "ALPGEN, a generator for hard multiparton processes in hadronic collisions", JHEP 0307:001,2003, hep-ph/0206293.
[5] T.Sjostrand et al., Comput. Phys. Commun. 135, 238, (2001)
[6] G.Corcella et al., JHEP 01,10 (2001)