Measurement of the Single-Top-Quark Production Cross Section at CDF

Aaltonen, Timo Antero

2008


http://hdl.handle.net/10138/24286
https://doi.org/10.1103/PhysRevLett.101.252001

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.
Measurement of the Single Top Quark Production Cross Section at CDF

We report a measurement of the single top quark production cross section in 2.2 fb$^{-1}$ of $p\bar{p}$ collision data collected by the Collider Detector at Fermilab at $\sqrt{s} = 1.96$ TeV. Candidate events are classified as signal-like by three parallel analyses which use likelihood, matrix element, and neural network discriminants. These results are combined in order to improve the sensitivity. We observe a signal consistent with the standard model prediction, but inconsistent with the background-only model by 3.7 standard deviations with a median expected sensitivity of 4.9 standard deviations. We measure a cross section of $2.2^{+0.7}_{-0.6}$ (stat + sys) pb, extract the CKM matrix element value $|V_{tb}| = 0.88^{+0.13}_{-0.12}$ (stat + sys) $\pm 0.07$ (theory), and set the limit $|V_{tb}| > 0.66$ at the 95% C.L.

PACS numbers: 14.65.Ha, 13.85.Qk, 12.15.Hh, 12.15.Ji
The top quark was discovered by the CDF and D0 collaborations in 1995 \cite{1} in the strong interaction $p\bar{p} \rightarrow t\bar{t} + X$. Since then, a comprehensive program of measurements has brought more precise knowledge of the top quark’s mass, pair-production cross section, and a number of its decay properties \cite{2}. The evidence strongly suggests that the particle observed in 1995 is the $SU(2)$ partner of the bottom quark and that it decays nearly 100% of the time into $Wb$ with a very short lifetime. The weak couplings of the top quark are less well constrained, except that $|V_{tb}|^2 \gg |V_{td}|^2 + |V_{ts}|^2$ \cite{3}. Requiring that the $3 \times 3$ Cabibbo-Kobayashi-Maskawa (CKM) matrix is unitary implies that $|V_{tb}| \simeq 1$ \cite{2}. With a matrix of higher rank, though, $|V_{tb}|$ could be small without measurably changing the $t \rightarrow Wb$ branching ratio. Production of single top quarks provides a direct measurement of $|V_{tb}|$ and a test of the $b$-quark content of the proton.

Top quarks are expected to be produced singly, as shown in Fig. 1. The combined $s + t$-channel cross section is predicted at next-to-leading order (NLO) to be $\sigma_{st} = 2.86 \pm 0.36 \text{ pb}$ \cite{4}. The small signal cross section and the presence of only one top quark in the final state make the separation of the signal from the large background challenging. Since the signal has very similar final states to the standard model Higgs boson production process $WH \rightarrow t\ell b\bar{b}$, the methods of this analysis can be used to search for the Higgs boson.

Recently, the D0 collaboration has reported evidence for single top quark production using 0.9 fb$^{-1}$ of data \cite{5} while measuring a cross section of $\sigma_{st} = 4.7 \pm 1.3 \text{ pb}$. This Letter reports a significantly more precise measurement of $\sigma_{st}$ in 2.2 fb$^{-1}$ of $p\bar{p}$ collisions at $\sqrt{s} = 1.96 \text{ TeV}$ using the CDF II detector.

The CDF II detector \cite{6} is a general purpose apparatus located at the Tevatron collider at Fermilab. The detector consists of a solenoidal charged particle spectrometer which includes a silicon microstrip detector array surrounded by a cylindrical drift chamber in a 1.4 T axial magnetic field. The energies of electrons and jets are measured with segmented sampling calorimeters. Surrounding the calorimeters are layers of steel instrumented with planar drift chambers and scintillators used for muon identification.

Three distinct trigger algorithms are employed to se-

![FIG. 1: Representative Feynman diagrams of single top quark production. Figures (a) and (b) are $t$-channel processes, and figure (c) is the $s$-channel process.](image)

lect the data used in this analysis: a high $p_T$ electron trigger, a high $p_T$ muon trigger, and a trigger that requires large missing transverse energy with either an energetic electromagnetic cluster or two separated jets \cite{7,8}.

Events are further selected by requiring the presence of an isolated electron or muon candidate with $p_T > 20 \text{ GeV}/c$, large missing transverse energy $E_T > 25 \text{ GeV}$, and either two or three jets each with $E_T > 20 \text{ GeV}$ and $|\eta| < 2.8$. The jets are identified by a fixed-cone algorithm with radius $\Delta R \equiv \sqrt{(\Delta \eta)^2 + (\Delta \phi)^2} = 0.4$, and their energies are corrected for instrumental effects \cite{9}. At least one of the jets is required to have a displaced vertex (tag) as identified by the secvtx algorithm \cite{10}. This $b$ tag preferentially selects jets containing $B$ hadrons.

In order to reduce the $Z+jets$, $t\bar{t}$, and diboson backgrounds, candidate events with a second charged lepton are rejected. Cosmic ray and photon candidates are identified and removed. Multijet background events without a leptonic $W$ decay (“non-$W$”) are rejected with specific selection requirements \cite{11,12}.

The diboson ($WW$, $WZ$, $ZZ$) and $t\bar{t}$ event yields are predicted using PYTHIA \cite{13} Monte Carlo (MC) samples normalized to the theoretical cross section \cite{14,15}. The processes in which a vector boson is produced in association with one or more jets ($Z$ or $W+jets$) are generated with ALPGEN \cite{16} using PYTHIA’s parton showering and underlying event model. The $W+jets$ samples are normalized to the measured data using events with exactly one, two, or three jets. A normalization factor of $1.4 \pm 0.4$ is applied to ALPGEN’s prediction for the fraction of $Wb\bar{b}$ and $Wc\bar{c}$ events. This factor is estimated by comparing the flavor content in $b$-tagged $W+1$ jet events in the data to the prediction from simulation. The background from events with mistakenly $b$-tagged light-flavor jets ($Wjj$) is estimated by measuring the rate of such mistags in multijet data \cite{17}. The mistag rate is then applied to the $W+jets$ data samples before $b$ tagging. The contributions to the data samples from non-$Wjj$ sources are subtracted from the prediction \cite{17}. Multijet non-$W$ events typically have less $E_T$ than events containing $W$ bosons.

By using templates for non-$W$ and $W+jets$, we fit the $E_T$ distribution and extract the non-$W$ fraction in the high $E_T$ signal region. The kinematic properties of the non-$W$ events are modeled using data events and $W+jets$ is
modeled using MC simulated events. The observed event yields and corresponding predictions are given in Table I.

<table>
<thead>
<tr>
<th>Process</th>
<th>$W + 2$ jets</th>
<th>$W + 3$ jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$-channel signal</td>
<td>40.3 ± 5.8</td>
<td>13.1 ± 1.9</td>
</tr>
<tr>
<td>$t$-channel signal</td>
<td>60.8 ± 8.9</td>
<td>17.9 ± 2.6</td>
</tr>
<tr>
<td>$Wb\bar{b}$</td>
<td>451.1 ± 136.0</td>
<td>138.0 ± 41.7</td>
</tr>
<tr>
<td>$Wc\bar{c} + Wc\bar{t}$</td>
<td>372.5 ± 114.8</td>
<td>103.2 ± 31.8</td>
</tr>
<tr>
<td>$Wjj$</td>
<td>337.1 ± 41.9</td>
<td>101.6 ± 12.8</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>142.0 ± 20.3</td>
<td>327.8 ± 46.6</td>
</tr>
<tr>
<td>Non-$W$</td>
<td>60.5 ± 24.2</td>
<td>21.0 ± 8.4</td>
</tr>
<tr>
<td>Diboson</td>
<td>61.1 ± 6.2</td>
<td>20.4 ± 2.1</td>
</tr>
<tr>
<td>$Z+$jets</td>
<td>25.5 ± 3.8</td>
<td>10.5 ± 1.5</td>
</tr>
<tr>
<td><strong>Total prediction</strong></td>
<td>1550.9 ± 256.6</td>
<td>753.5 ± 87.6</td>
</tr>
</tbody>
</table>

Observed 1546 719

Table I: Background composition and predicted number of single top events in 2.2 fb$^{-1}$ of CDF Run II data with at least one $b$-tagged jet.

Single top events are simulated using the tree-level matrix-element generator **MADEVENT** [18]. The two $t$-channel processes of Figs. 1(a) and 1(b) are combined to match the event kinematics as predicted by a fully differential NLO calculation [4, 19].

The expected standard model signal-to-background ratio for selected events is $\sim 7\%$ in the two-jet sample and $\sim 5\%$ in the three-jet sample. The uncertainties on the background predictions are larger than the expected signals; therefore, we have developed three powerful discriminants to distinguish signal from background events. The predicted distributions of each discriminant are fit to the data to extract the single top production cross section. All analyses use the same event selection and were optimized with the signal region blinded.

$Wjj$, $Wc\bar{c}$ and $Wc\bar{t}$ events do not contain $b$-quark jets, but constitute $\sim 40\%$ of the estimated background after imposing a $b$ tag requirement. As part of all three discriminants we employ a jet-flavor separating variable, $b_{nn}$, constructed using the neural network tool **NEUROBAYES** [20], which is trained to distinguish $b$ jets from charm and light-flavor jets based on secondary vertex tracking information [11]. The usage of $b_{nn}$ leads to an improvement in sensitivity of 15 to 20% in each analysis.

**Likelihood Function Discriminant (LF):** A projective likelihood technique [17, 21] is used to combine information from several input variables to optimize the separation of the single top signal from the backgrounds. Two likelihood functions are created, one for two-jet events, $L_{2j}$, and one for three-jet events, $L_{3j}$. The input variables used for $L_{2j}$ are $b_{nn}$, $Q \times \eta$ [22], the invariant mass of the $t\bar{b}b$ system $M_{t\bar{b}b}$, the total scalar sum of transverse energy in the event $H_T$, $\cos \theta_T$ [23], the dijet mass $M_{jj}$, and the $t$-channel matrix element.

The matrix element used here is computed using four-vectors from the event after kinematically constraining $M_{\ell\nu} = M_W$ and $M_{\ell\bar{b}} = m_t$, where $M_W$ and $m_t$ are the $W$ and top quark pole masses in the matrix element. The $M_W$ constraint introduces a quadratic ambiguity in the $z$ component of the neutrino momentum; we choose the solution with the smaller $|p_Z|$. For events with three jets, ten input variables are used to construct $L_{3j}$: $b_{nn}$, $Q \times \eta$, $M_{t\ell\nu}$, $\cos \theta_T$, $M_{jj}$, of the two jets not selected as the $b$ from top decay, the number of $b$ tags, the smallest $\Delta R$ between any two jets, the smallest $p_T$ of the three jets, $p_T(W)$, and $E_T$ of the jet selected as coming from the $b$ from the top quark decay. The $b$-quark jet is chosen using a linear combination of the jet $E_T$ and the $\chi^2$ resulting from the comparison of the kinematically constrained jet energy and the measured jet energy.

**Matrix Element Discriminant (ME):** The matrix element method relies on the evaluation of event probability densities for signal and background processes based on calculations of the standard model differential cross sections [12, 24]. We construct these probability densities for each process for each event given their measured quantities $x$ by integrating the appropriate differential cross section $\frac{d\sigma}{dy}$ over the underlying partonic quantities $y$, convolved with the parton distribution functions (PDFs) and detector resolution effects:

$$P(x) = \sum_{\text{perm.}} \int \frac{d\sigma(y)}{dy} f(q_1) f(q_2) dq_1 dq_2 W(x, y) dy.$$ (1)

The PDFs ($f(q_1)$ and $f(q_2)$) take into account the flavors of the colliding quark and anti-quark. We use the CTEQ PDF parameterization [25]. The detector resolution effects are described by a transfer function $W(x, y)$ relating $x$ to $y$. The momenta of electrons, muons, and the angles of jets are assumed to be measured exactly. $W(x, y)$ maps parton energies to measured jet energies after correction for instrumental effects [8]. This mapping is obtained by parameterizing the jet response in fully simulated MC events. The definition of the probability densities includes possible permutations of matching jets with partons. The integration is performed over the energy of the partons and $p_Z^\nu$. We calculate the matrix element for the event probability at tree-level using **MADGRAPH** [26]. Event probability densities are computed for the $s$-channel and $t$-channel signal as well as $Wb\bar{b}$, $Wc\bar{c}$, $Wc\bar{t}$, $Wjj$, and $t\bar{t}$ background hypotheses. In the specific case of the $t\bar{t}$ matrix element, additional integrations are performed over the momenta of particles not detected.

The event probability densities are combined into an event probability discriminant: $EPD = \frac{P_{\text{signal}}}{P_{\text{signal}} + P_{\text{background}}}$. To better classify signal events that contain $b$ jets, we incorporate the output $b_{nn}$ of the neural net-
work jet-flavor separator into the final discriminant:

$$EPD = \frac{b_{nn} \cdot P_{st}}{b_{nn} \cdot (P_{st} + P_{tt} + P_{Wb}) + (1 - b_{nn}) (P_{Wc} + P_{Wc} + P_{Wj})}.$$ (2)

Both signal channels are combined to one single top probability density $P_{st} = P_{t\text{-channel}} + P_{t\text{-channel}}$. 

**Neural Network Discriminant (NN):** The third multivariate approach [11] employs neural networks, which have the general advantage that correlations between the discriminating input variables are identified and utilized to optimize the separation power between signal and background. The networks are developed using the NEUROBAYES analysis package [20], which combines a three-layer feed-forward neural network with a complex and robust preprocessing of the input variables. Bayesian regularization techniques are utilized to avoid over-training.

Four separate networks are trained to identify different signals in distinct samples using simulated events from the common samples described previously. An $s$-channel signal is used for training on events with two $b$-tagged jets. A $t$-channel signal is used for the two-jet sample with a single $b$ tag and for the three-jet samples with one or two $b$ tags. The networks use 11 to 18 input variables. The most important ones are $M_{t\bar{t}b}$, $b_{nn}$, $M_{jj}$, $Q \times \eta$, $\cos \theta_{jj}$, the transverse mass of the $W$ boson, and $H_T$. The input variables are selected from a large list using an automated evaluation during the preprocessing step before the network training. In an iterative process, we determine those variables whose removal would cause a significant loss in separation power between signal and background and use them for network training.

**Combination:** We studied two methods to combine the cross section fit results. The best linear unbiased estimator (BLUE) [27] technique optimizes the coefficients of a linear combination using the uncertainties and correlations of the three individual analyses: LF, ME, and NN. The correlation coefficients between the analyses are: LF-ME: 59%; LF-NN: 74%; ME-NN: 61%. In another combination approach, a “super analysis” is built based on the outcomes for each event in each of the three individual analyses. The super-discriminant method uses a neuro-evolution network [28] trained to separate the signal from the background based on the discriminant outputs of the three analyses. With the super-discriminant analysis we improve the sensitivity by 10% over the best individual analysis, and we use it to quote our final results. As a cross-check, BLUE yields a 7% sensitivity improvement.

Before unblinding the data, the MC simulation of each input variable and the discriminant outputs were checked in data control samples depleted in signal. These are the lepton + $b$-tagged four-jet sample, which is enriched in $t\bar{t}$ events, and the two- and three-jet samples in which no jet is $b$ tagged. The latter are high-statistics samples with similar kinematics to the $b$-tagged signal samples. The data distributions in the control samples are described well by the MC simulation.

Figure 2 shows the distributions of the individual analyses’ discriminants and the super-discriminant. We calculate the probability ($p$-value) [2] of the background-only discriminant distribution to fluctuate to the observed data or more which is then converted into signal significance under a Gaussian assumption. All sources of systematic uncertainty are included and correlations between normalization and discriminant shape changes are considered. Uncertainties in the jet energy scale, $b$-tagging efficiencies, lepton identification and trigger efficiencies, the amount of initial and final state radiation, PDFs, factorization and renormalization scale, and MC modeling have been explored and incorporated in this combination and all individual analyses.

We interpret the excess of signal-like events over the expected background as strong evidence for single top production with a signal significance of 3.7 standard deviations, with a sensitivity, defined to be the median expected significance, of 4.9 standard deviations. The most probable value of the combined $s$-channel and $t$-channel cross sections is $\sigma_{st} = 2.2^{+0.7}_{-0.6}$ pb for a top quark mass of 175 GeV/c$^2$ which is consistent with

![FIG. 2: Discriminant distribution for all channels combined of the (a) LF analysis, (b) ME analysis, (c) NN analysis and (d) combined super-discriminant analysis. Points with error bars indicate the data. The predicted signal and background distributions are shown as stacked histograms. The insets show the candidate events in the signal regions. A summary of all results is shown in (e).](image-url)
the cross check result, obtained from BLUE, $\sigma_{st} = 2.3 \pm 0.7 \text{ pb}$. The dependence on the top quark mass is $+0.02 \text{ pb/(GeV/$c^2$)}$. From the cross section measurement at $m_t = 175 \text{ GeV/$c^2$}$, we obtain $|V_{tb}| = 0.88^{+0.13}_{-0.12} \text{ (stat. + sys.)} \pm 0.07 \text{ (theory)}$ and limit $|V_{tb}| > 0.06$ at the 95% C.L. assuming a flat prior in $|V_{tb}|^2$ from 0 to 1. This is the most precise direct measurement of $|V_{tb}|$ to date.

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 f"ur Bildung und Forschung and the Alexander von Humboldt Foundation, Germany; the Korean Science and Engineering Foundation and the Korean Research Foundation; the Academy of Finland; and the Academy of the Royal Society, UK; the Institut National de Physique theorie; 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 Educaci\'on y Ciencia and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.

[7] We use a cylindrical coordinate system with its origin in the center of the detector, where $\theta$ and $\phi$ are the polar and azimuthal angles, respectively, and pseudorapidity is $\eta = -\ln(|\tan(\theta/2)|)$. The missing $E_T (\vec{E}_T)$ is defined by $\vec{E}_T = -\sum_i E_{Tn_i}, i = \text{calorimeter tower number, where } n_i \text{ is a unit vector perpendicular to the beam axis and pointing at the } i^{th} \text{ calorimeter tower. } \vec{E}_T \text{ is corrected for high-energy muons and also jet energy corrections. We define } E_T = |\vec{E}_T|. \text{ The transverse momentum } p_T \text{ is defined to be } p \sin \theta.$