First Observation of Electroweak Single Top Quark Production

We report the first observation of single top quark production using 3.2 fb$^{-1}$ of $p\bar{p}$ collision data with $\sqrt{s} = 1.96$ TeV collected by the Collider Detector at Fermilab. The significance of the observed data is 5.0 standard deviations, and the expected sensitivity for standard model production and decay is in excess of 5.9 standard deviations. Assuming $m_t = 175$ GeV/$c^2$, we measure a cross section of $2.2^{+0.6}_{-0.5}$ (stat + syst) pb, extract the CKM matrix element value $|V_{tb}| = 0.91 \pm 0.11$ (stat + syst) $\pm 0.07$ (theory), and set the limit $|V_{tb}| > 0.71$ at the 95% C.L.

PACS numbers: 14.65.Ha, 13.85.Qk, 12.15.Hh, 12.15.Ji

---

*Deceased
†With visitors from aUniversity of Massachusetts Amherst, Amherst, Massachusetts 01003, bUniversiteit Antwerpen, B-2610
In the standard model (SM), top quarks are expected to be produced singly in $p\bar{p}$ collisions through $s$-channel or $t$-channel exchange of a virtual $W$ boson [1]. The reasons for studying single top quarks are compelling: the production cross section is directly proportional to the square of the CKM matrix element $|V_{tb}|$, and thus a measurement of the rate constrains fourth-generation models, models with flavor-changing neutral currents, and other new phenomena [3]. Electroweak production of single top quarks is a difficult process to measure because the expected production cross section is too small compared with those of competing background processes, and it is also smaller than the uncertainty on the total background rate. The presence of only one top quark in the event provides fewer features to use in separating the signal from background, compared with measurements of top pair production ($tt\bar{t}$), which was first observed in 1995 [1].

To overcome these challenges, a variety of multivariate techniques for separating single top events from the backgrounds have been developed. Using different combinations of techniques, both the CDF and D0 collaborations have published evidence for single top quark production at significance levels of 3.7 and 3.6 standard deviations, respectively [2,8]. The analysis described in this Letter supersedes that of Ref. [7] and achieves a significantly improved sensitivity by including a larger data sample and by adding three new analyses. We report a signal significance of 5.0 standard deviations, thus conclusively observing electroweak production of single top quarks, and we make the most precise measurement of $|V_{tb}|$ to date.

We assume that single top quarks are produced in the $s$- and $t$-channel modes with the SM ratio, and that the branching ratio of the top quark to $Wb$ is 100%.

Antwerp, Belgium, University of Bristol, Bristol BS8 1TL, United Kingdom, Chinese Academy of Sciences, Beijing 100864, China, Instituto Nazionale di Fisica Nucleare, Sezione di Cagliari, 09042 Monserrato (Cagliari), Italy, University of California Irvine, Irvine, CA 92697, University of California Santa Cruz, Santa Cruz, CA 95064, Cornell University, Ithaca, NY 14853, University of Cyprus, Nicosia CY-1678, Cyprus, University College Dublin, Dublin 4, Ireland, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom, University of Fukuoka, Fukuoka City, Fukuoka Prefecture, Japan 901-0017, Kinki University, Higashi-Osaka City, Japan 577-8502, Universidad Iberoamericana, Mexico D.F., Mexico, Queen Mary, University of London, London, E1 4NS, England, University of Manchester, Manchester M13 9PL, England, Nagasaki Institute of Applied Science, Nagasaki, Japan, University of Notre Dame, Notre Dame, IN 46556, University de Oviedo, E-33007 Oviedo, Spain, Texas Tech University, Lubbock, TX 79469, IFIC(CSIC-Universitat de Valencia), 46071 Valencia, Spain, University of Virginia, Charlottesville, VA 22904, Bergische Universit"at Wuppertal, 42097 Wuppertal, Germany, On leave from J. Stefan Institute, Ljubljana, Slovenia.

We seek events in which the $W$ boson decays leptonically in order to improve the signal-to-background ratio $s/b$. We simulate single top events using the tree-level matrix-element generator madevent [4]. The $t$-channel signal is modeled by the two processes $qb \rightarrow q't$ and $gg \rightarrow q'tb$, which are combined to match the event kinematics predicted by a fully differential NLO calculation [6,10].

A total of six analyses are combined to yield the final results reported here. The likelihood function (LF), matrix element (ME), and neural network (NN) analyses of [2] are re-used with an additional 1 fb$^{-1}$ of integrated luminosity; their methods remain unchanged. The three new analyses introduced here are: a boosted decision tree (BDT), a likelihood function optimized for $s$-channel single top production (LFS), and a neural-network-based analysis of events with missing transverse energy $E_T$ [11] and jets (MJ). The BDT and LFS analyses use events that overlap with the LF, ME, and NN analyses, while the MJ analysis uses an orthogonal event selection that adds about 30% to the signal acceptance. This paper concentrates on the three new analyses and their combination with the analyses of [2] using 3.2 fb$^{-1}$ of integrated luminosity collected with the CDF II detector [12].

For the LF, ME, NN, BDT, and LFS analyses we select $\ell + E_T + \text{jets}$ events as described in [7], where $\ell$ is an explicitly reconstructed electron or muon from the $W$ boson decay and at least one jet is identified as containing a $B$ hadron. The background has contributions from events in which a $W$ boson is produced in association with one or more heavy-flavor jets ($W+HF$), events with mistakenly $b$-tagged light-flavor jets (mistags), multijet events (QCD), $tt\bar{t}$ and diboson processes, as well as $Z$-jet events. The expected event yields in Table I are estimated as in [7] where the signal, $tt\bar{t}$, and diboson categories are Monte Carlo (MC) predictions scaled to the total integrated luminosity while the remaining categories use predictions derived from data control samples. The uncertainties quoted in Table I include theoretical uncertainties, the luminosity uncertainty for the MC predictions, and experimental uncertainties for the data-driven background normalizations.

The MJ analysis is designed to select events with $E_T$ and jets and to veto events selected by the $\ell + E_T + \text{jets}$ analyses. It accepts events in which the $W$ boson decays into two leptons and those in which the electron or muon fails the lepton identification criteria. We use data corresponding to 2.1 fb$^{-1}$ of integrated luminosity for the MJ analysis and select events that have $E_T > 50$ GeV and two jets with $|\eta| < 2.0$, at least one of which has $|\eta| < 0.9$. The jet energy measurements include information from both the calorimeter and the charged-particle spectrometer. Events must have one jet with transverse energy $E_T$ greater than 35 GeV, and a second jet with $E_T$ greater than 25 GeV. The angular separation between the two jets, $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, is required to exceed 1.0. We reject events with four or more jets with
to reduce the multijet (QCD) and $t\bar{t}$ backgrounds. We identify $b$ jets with the same algorithm used in [7], supplemented with a jet probability algorithm [13].

The primary background in the MJ analysis is QCD events in which mismeasured jet energies produce large $E_T$ aligned in the same direction as jets. To reduce this background, we use the transverse momentum imbalance ($\vec{p}_T$) as measured in the spectrometer. This variable is more correlated to the neutrino energy and its direction than $E_T$ in this class of events. The absolute amount of $E_T$ and $\vec{p}_T$, the angle between them, the azimuthal angles between $E_T$ or $\vec{p}_T$ and the jet directions, and several other less powerful variables are used as inputs to a neural network (NNQCD). The NNQCD output is required to pass a threshold, removing 77% of the QCD background while keeping 91% of the signal acceptance.

The backgrounds in the MJ analysis due to QCD events and events with light-flavor jets produced in association with $W$ and $Z$ bosons are estimated using data in a control region composed of events in which the $E_T$ is aligned with one of the jets. The observed and expected event counts for the MJ analysis are given in the $E_T + \text{jets}$ column of Table I.

<table>
<thead>
<tr>
<th>Process</th>
<th>$\ell + E_T + \text{jets}$</th>
<th>$E_T + \text{jets}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s$-channel signal</td>
<td>77.3 $\pm$ 11.2</td>
<td>29.6 $\pm$ 3.7</td>
</tr>
<tr>
<td>$t$-channel signal</td>
<td>113.8 $\pm$ 16.9</td>
<td>34.5 $\pm$ 6.1</td>
</tr>
<tr>
<td>$W + HF$</td>
<td>1551.0 $\pm$ 472.3</td>
<td>304.4 $\pm$ 115.5</td>
</tr>
<tr>
<td>$t\bar{t}$</td>
<td>686.1 $\pm$ 99.4</td>
<td>184.5 $\pm$ 30.2</td>
</tr>
<tr>
<td>$Z + \text{jets}$</td>
<td>52.1 $\pm$ 8.0</td>
<td>128.6 $\pm$ 53.7</td>
</tr>
<tr>
<td>Diboson</td>
<td>118.4 $\pm$ 12.2</td>
<td>42.1 $\pm$ 6.7</td>
</tr>
<tr>
<td>QCD$+$mistags</td>
<td>777.9 $\pm$ 103.7</td>
<td>679.4 $\pm$ 27.9</td>
</tr>
<tr>
<td><strong>Total prediction</strong></td>
<td>3376.5 $\pm$ 504.9</td>
<td>1404 $\pm$ 172</td>
</tr>
<tr>
<td><strong>Observed</strong></td>
<td>3315</td>
<td>1411</td>
</tr>
</tbody>
</table>

After event selection, the samples are dominated by background. We further discriminate the signal with multivariate techniques. Each multivariate technique defines a function which reduces several reconstructed quantities for each event into a single output variable whose distribution can be studied and fit to extract signal and background contributions. Validation of the background modeling for the input variables and output distributions is a crucial step in the use of multivariate techniques. We first describe the construction of our multivariate tools and then the checks we used to prove the validity of our background model. The LF, ME, and NN discriminants are described in [7]. The BDT discriminant uses a decision tree method that applies binary cuts iteratively to classify events [14]. The discrimination is further improved using a boosting algorithm [15, 16]. The BDT discriminant uses over 20 input variables. Some of the most sensitive are the neural-network jet-flavor separator [14], 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 $H_T$, $Q \times \eta$ [15], the dijet mass $M_{jj}$, and the transverse mass of the $W$ boson.

The LFS discriminant uses projective likelihood functions [19] to combine the separation power of several variables and is optimized to be sensitive to the $s$-channel process. The subset of the $\ell + E_T + \text{jets}$ sample with two $b$-tagged jets is used and consists of 609 events. The dominant backgrounds are $W + HF$ and $t\bar{t}$ production. A kinematic fitter is used to find the most likely resolution of two ambiguities: the $z$-component of the neutrino momentum and the $b$ jet that most likely came from the top quark decay. In addition to the outputs of the kinematic fitter, other important inputs to the likelihood are the invariant mass of the two $b$-tagged jets $M_{bb}$, the transverse momentum of the $b\bar{b}$ system, the leading jet transverse momentum, $M_{t\bar{b}b}$, $H_T$, and $E_T$.

The MJ discriminant uses a neural network to combine information from several input variables. The most important variables are the invariant mass of the $E_T$ and the second leading jet, the scalar sum of the jet energies, the $E_T$, and the azimuthal angle between the $E_T$ and the jets.

We combine the LF, ME, NN, BDT, and LFS channels using a super-discriminant (SD) technique similar to that which was applied in [7]. The SD method uses a neural network trained with neuro-evolution [20] to separate the signal from the background taking as inputs the discriminant outputs of the five analyses for each event. With the super-discriminant analysis we improve the sensitivity (defined below) by 13% over the best individual analysis. We perform a simultaneous fit over the two exclusive channels, MJ and SD, to obtain the final combined results.

Before investigating the sample of selected events, we used background-dominated data control samples to check the modeling of each input variable as well as the output distributions of each multivariate discriminant. For the $\ell + E_T + \text{jets}$ analyses the control samples used 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 there is no $b$-tagged jet. The latter are enriched in $W + \text{jets}$ and QCD events with kinematics similar to the $b$-tagged signal samples and have high statistics, making it possible to observe that the background model describes the data well over three orders of magnitude in our output discriminants. For the MJ analysis, three control samples are used: in the first sample, the $E_T$ is required to be aligned along one of the jets, and in the second, the events are required to fail the NNQCD requirement, and
in the third, a lepton is required to be present. The data distributions in all control samples are described well by our models for each of the analysis input variables and a large set of other variables not used as inputs. More than two thousand distributions were checked for evidence of mismodeling. Small discrepancies were found in the distributions of the angles between two jets in the untagged lepton + two-jet sample and the modeling of jets with rapidity greater than 2.4. These effects are included as systematic uncertainties on the shape of the background models.

Figure 1 shows the distributions of the five $\ell + \not{E}_T +$ jets discriminants. These are combined to give the SD distribution shown in Fig. 2 together with the MJ distribution. In the rightmost bins, assuming SM production and decay, the SD has an $s/b$ that exceeds 5.0. This large $s/b$ significantly reduces our sensitivity to systematic uncertainties affecting the background. We use the distributions of the SD and MJ discriminants to extract the measured cross section and the signal significance.

We measure the single top cross section using a Bayesian binned likelihood technique [21] assuming a flat prior in the cross section and integrating the posterior over all sources of systematic uncertainty. We quote the measured cross section as the value that maximizes the posterior over all sources of systematic uncertainty. The systematic uncertainties on the shape of the background are largely constrained by the data in the background-enriched portions of the SD and MJ discriminant distributions. Uncertainties on the shapes of these distributions degrade the extrapolation of these constraints to more signal-like regions. The sources of systematic uncertainties affecting these shapes are discussed below and are also included in all calculations. The uncertainties assigned were conservatively chosen to cover the full range of variations studied. We quote the measured cross section as the value that maximizes the posterior likelihood, and use the shortest interval containing 68% of the integral of the posterior to set the uncertainties.

We calculate the significance as a $p$-value [21], which is the probability, assuming single top quark production is absent, that $-2 \ln Q = -2 \ln (p(\text{data}|s + b)/p(\text{data}|b))$ is less than that observed in the data. Figure 2(c) shows the distributions of $-2 \ln Q$ in pseudoexperiments that assume SM single top ($S + B$) and also those that assume single top production is absent ($B$), along with the value observed in data. The effects of the systematic uncertainties are included in the pseudoexperiments. We convert the observed $p$-value into a number of standard deviations using the integral of one side of a Gaussian function.

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, parton distribution functions, factorization and renormalization scale, and background modelings have been explored and incorporated in all individual analyses and the combination. We include uncorrelated MC statistical uncertainties in each bin of each discriminant distribution. A $\pm 2.5 \text{ GeV}/c^2$ uncertainty on the top quark mass $m_t$ is included in the significance and $|V_{tb}|$ results but quote the dependence on $m_t$ separately in the cross section.

Table II lists the measured cross sections and significances for each of the component analyses and the combination. The measured cross sections for the five correlated analyses and the SD are close to each other even though the analyses choose different input variables and are optimized differently. We interpret the excess of signal-like events over the expected background as observation of single top production with a $p$-value of $3.10 \times 10^{-7}$, corresponding to a signal significance of 5.0 standard deviations. The sensitivity is defined to be the median expected significance and is in excess of 5.9 standard deviations, assuming the SM signal cross section. The most probable value of the combined $s$-channel and $t$-channel cross section is $2.3^{+0.6}_{-0.5}$ pb assuming a top quark mass of 175 GeV/$c^2$. The dependence on the top quark mass is $+0.02 \text{ pb}/(\text{GeV}/c^2)$. From the cross section measurement at $m_t = 175 \text{ GeV}/c^2$, we obtain $|V_{tb}| = 0.91 \pm 0.11(\text{stat + syst}) \pm 0.07(\text{theory})$ and limit $|V_{tb}| > 0.71$ 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.

In summary, we combine six multivariate analysis techniques to precisely measure the electroweak single top production cross section and the CKM matrix element $|V_{tb}|$. We have carefully cross-checked our analysis techniques with data control samples and we assign generous rate and shape uncertainties to all predictions we use. Our combined discriminant allows us to purify a signal...
FIG. 1: Discriminant distributions for the $\ell+E_T+\text{jets}$ analyses. The data are indicated with points, and the predictions are shown separately for each contribution with stacked histograms. The signal expectations shown are the SM predictions. The insets show the distributions of the candidate events in the high-discriminant region.

FIG. 2: Discriminant distributions for the (a) SD, and (b) MJ analyses (see Fig. 1 for their caption and legend). Figure (c) shows the distribution of the likelihood ratio test statistic $-2\ln Q$.

sample with $s/b > 5.0$ in the most sensitive region, allowing for a significant outcome in the presence of these conservative systematic uncertainties. We observe single top quark production for the first time with a significance of 5.0 standard deviations.

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 Humboldt Foundation, 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 Nucléaire et Physique des Particules/CNRS; the Russian Foundation for Basic Research; the Ministerio de Ciencia e Innovación, and Programa Consolider-Ingenio 2010, Spain; the Slovak R&D Agency; and the Academy of Finland.
[11] 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 $\vec{E}_T = -\sum_i E_T^i \vec{n}_i$, $i =$ calorimeter tower number, where $\vec{n}_i$ is a unit vector perpendicular to the beam axis and pointing at the $i^{th}$ calorimeter tower. $E_T$ is corrected for high-energy muons and also jet energy corrections. We define $E_T = |\vec{E}_T|$. The transverse momentum $p_T$ is defined to be $p_T \sin \theta$.
[18] $Q \times \eta$ is the charge of the lepton times the pseudorapidity of the jet not assigned to be the b from top quark decay; see C. P. Yuan, Phys. Rev. D 41, 42 (1990).