Evidence for production of single top quarks and first direct measurement of $|V_{tb}|$

The D0 Collaboration presents first evidence for the production of single top quarks at the Fermilab Tevatron pp collider. Using a 0.9 fb\(^{-1}\) dataset, we apply a multivariate analysis to separate signal from background and measure \(\sigma(pp \to t\bar{b} + X) = 4.9 \pm 1.4\) pb. The probability to measure a cross section at this value or higher in the absence of signal is 0.035\%, corresponding to a 3.4 standard deviation significance. We use the cross section measurement to directly determine the CKM matrix element that describes the \(Wtb\) coupling and find 0.68 < \(|V_{tb}|\) < 1 at 95% C.L. within the standard model.

PACS numbers: 14.65.Ha; 12.15.Ji; 13.85.Qk
Top quarks were first observed in strong $t\bar{t}$ pair production at the Tevatron collider in 1995 [1]. In the standard model (SM), $\sigma(pp \to t\bar{t} + X) = 6.8_{-0.5}^{+0.6}$ pb at $\sqrt{s} = 1.96$ TeV for a top quark mass of 175 GeV. Top quarks are also expected to be produced singly via the electroweak processes [3, 4] illustrated in Fig. 1. For brevity, we use the notation “tb” to represent the sum of $tb$ and $t\bar{b}$, and “$t\bar{b}q$” for the sum of $t\bar{b}q$ and $t\bar{b}q$. The next-to-leading order (NLO) prediction for the s-channel single top quark cross section is $\sigma(pp \to tb + X) = 0.88 \pm 0.11$ pb, and for the t-channel process, the prediction is $\sigma(pp \to t\bar{b} + X) = 1.98 \pm 0.25$ pb [5, 6].

![FIG. 1: Representative Feynman diagrams for (a) s-channel single top quark production and (b) t-channel production.](image)

Single top quark events can be used to study the $Wtb$ coupling [7], and to measure the magnitude of the element $|V_{tb}|$ of the quark mixing matrix, (the Cabibbo-Kobayashi-Maskawa (CKM) matrix [8]), without assuming only three generations of quarks [9]. The quark mixing matrix must be unitary, which for three families implies $|V_{tb}| \simeq 1$ [10]. A smaller measured value would indicate the presence of a fourth quark family to make up the difference. Single top quark production can also be used to measure the top quark partial decay width $\Gamma(t \to Wb)$ [11] and hence the top quark lifetime.

The D0 collaboration has previously published limits [12] on single top quark production. The best 95% C.L. upper limits are $\sigma(pp \to tb + X) < 6.4$ pb and $\sigma(pp \to t\bar{b}q + X) < 5.0$ pb. The CDF collaboration has also published limits on the cross sections [13].

This Letter describes a search for single top quark production using 0.9 fb$^{-1}$ of data produced at a center-of-mass energy of 1.96 TeV. The data were collected from 2002 to 2005 using the D0 detector [14] with triggers that required a jet and an electron or a muon. The search focuses on the final state consisting of one high transverse momentum ($p_T$) isolated lepton and missing transverse energy ($E_T$), together with a $b$-quark jet from the decay of the top quark ($t \to Wb \to \ell v b$). There is an additional $b$ quark in s-channel production, and an additional light quark and $b$ quark in t-channel production. The second $b$ quark in the t-channel is rarely reconstructed since it is produced in the forward direction with low transverse momentum. The main backgrounds are: W bosons produced in association with jets; top quark pairs decaying into the lepton+jets and dilepton final states, when a jet or a lepton is not reconstructed; and multijet production, where a jet is misreconstructed as an electron, or a heavy-flavor quark decays to a muon that passes the isolation criteria.

We model the signal using the SINGLETOP NLO Monte Carlo (MC) event generator [15]. The event kinematics for both s-channel and t-channel reproduce distributions found in NLO calculations [5]. The decays of the top quark and resulting W boson, with finite widths, are modeled in the SINGLETOP generator to preserve particle spin information. PYTHIA [16] is used to model the hadronization of the generated partons. For the $tb$ search, we assume SM $t\bar{b}q$ as part of the background, and vice versa. For the $tb+t\bar{b}q$ search, we assume the SM ratio between the $tb$ and $t\bar{b}q$ cross sections.

We simulate the $t\bar{t}$ and $W+jets$ backgrounds using the ALPGEN leading-order MC event generator [17] and PYTHIA to model the hadronization. A parton-jet matching algorithm [18] is used to ensure there is no double-counting of the final states. The $W+jets$ background is normalized to the integrated luminosity times the predicted $t\bar{t}$ cross section [2]. The multijet background is modeled using data that contain nonisolated leptons but which otherwise resemble the lepton+jet dataset. The $W+jets$ background, combined with the multijet background, is normalized to the lepton+jet dataset separately for each analysis channel (defined by lepton flavor and jet multiplicity) before b-jet tagging (described later). In the $W+jets$ background simulation, we scale the $Wbb$ and $Wcc$ components by a factor of $1.50 \pm 0.45$ to better represent higher-order effects [19]. This factor is determined by scaling the numbers of events in an admixture of light- and heavy-flavor $W+jets$ MC events to data that have no b tags but which otherwise pass all selection cuts. The uncertainty assigned to this factor covers the expected dependence on kinematics and the assumption that the factor is the same for $Wbb$ and $Wcc$.

We pass the MC events through a GEANT-based simulation [20] of the D0 detector. To correct differences between the simulation and data, we apply weights to the simulated events to model the effects of the triggers, lepton identification and isolation requirements, and the energy scale of the jets. The b-tagging algorithm [21] is modeled by applying weights that account for the probability for each jet to be tagged as a function of jet flavor, $p_T$, and pseudorapidity $\eta$.

We choose events with two, three, or four jets, reconstructed using a cone algorithm [22] with radius $R = \sqrt{(\Delta y)^2 + (\Delta \phi)^2} = 0.5$ (where $y$ is rapidity and $\phi$ is azimuthal angle) to cluster energy deposits in the calorimeter. The leading jet has $p_T > 25$ GeV and $|\eta| < 2.5$, the second leading jet has $p_T > 20$ GeV and $|\eta| < 3.4$, and subsequent jets have $p_T > 15$ GeV and $|\eta| < 3.4$. Events are required to have $15 < \Sigma p_T < 200$ GeV and exactly one isolated electron with $p_T > 15$ GeV and $|\eta| < 1.1$ or one isolated muon with $p_T > 18$ GeV and $|\eta| < 2.0$. Misreconstructed events are rejected by requiring that the direction of the $\Sigma p_T$ is not aligned or anti-aligned in azimuth with the lepton or a jet. To enhance the signal content of the selection, one or two of
the jets are required to be identified as originating from long-lived $b$ hadrons by a neural network $b$-jet tagging algorithm. The variables used to identify such jets rely on the presence and characteristics of a secondary vertex and tracks with high impact parameters inside the jet. For a 0.5% light-jet $b$-tag efficiency (the average mistag probability), we obtain a 50% average tag rate in data for $b$ jets with $|\eta| < 2.4$.

We select 1,398 $b$-tagged lepton+$j$ets data events, which we expect to contain $62 \pm 13$ single top quark events. To increase the search sensitivity, we divide these events into twelve independent analysis channels based on the lepton flavor ($e$ or $\mu$), jet multiplicity (2, 3, or 4), and number of identified $b$ jets (1 or 2). We do this because the signal acceptance and signal-to-background ratio differ significantly from channel to channel. Event yields are given in Table I, shown separated only by jet multiplicity for simplicity. The acceptances for single top quark signal as percentages of the total production cross sections are $(3.2 \pm 0.4)\%$ for $tb$ and $(2.1 \pm 0.3)\%$ for $tqb$.

The dominant contributions to the uncertainties on the backgrounds come from: normalization of the $tt$ background ($18\%$ of the $tt$ component), which includes a term to account for the top quark mass uncertainty; normalization of the $W$+$jets$ and multijet backgrounds to data ($17$–$27\%$), which includes the uncertainty on the heavy-flavor fraction of the model; the jet energy scale corrections ($1$–$20\%$); and the $b$-tagging probabilities ($12$–$17\%$ for double-tagged events). The uncertainty on the integrated luminosity is $6\%$; all other sources contribute at the few percent level. The uncertainties from the jet energy scale corrections and the $b$-tagging probabilities affect both the shape and normalization of the simulated distributions. Having selected the data samples, we check that the background model reproduces the data in a multitude of variables (e.g., transverse momenta, pseudorapidities, azimuthal angles, masses) for each analysis channel and find agreement within uncertainties.

Since we expect single top quark events to constitute only a small fraction of the selected event samples, and since the background uncertainty is larger than the expected signal, a counting experiment will not have sufficient sensitivity to verify their presence. We proceed instead to calculate multivariate discriminants that separate the signal from background and thus enhance the probability to observe single top quarks. We use decision trees [23] to create these discriminants. A decision tree is a machine-learning technique that applies cuts iteratively to classify events. The discrimination power is further improved by averaging over many decision trees constructed using the adaptive boosting algorithm AdaBoost [24]. We refer to this average as a boosted decision tree.

We identify 49 variables from an analysis of the signal and background Feynman diagrams [25], studies of single top quark production at NLO [26], and from other analyses [4, 27]. The variables may be classified into three categories: individual object kinematics, global event kinematics, and variables based on angular correlations. Those with the most discrimination power include the invariant mass of all the jets in the event, the invariant mass of the reconstructed $W$ boson and the highest-$p_T$ $b$-tagged jet, the angle between the highest-$p_T$ $b$-tagged jet and the lepton in the rest frame of the reconstructed top quark, and the lepton charge times the pseudorapidity of the untagged jet. We find that reducing the number of variables always reduces the sensitivity of the analysis.

We use a boosted decision tree (DT) in each of the twelve analysis channels for three searches: $tb+tqb$, $tqb$, and $tb$. These 36 DTs are trained to separate one of the signals from the sum of the $tt$ and $W$+$jets$ backgrounds. One-third of the MC signal and background events is used for training; the remaining two-thirds are used to determine the acceptances in an unbiased manner. A boosted decision tree produces a quasi-continuous output distribution $O_{\text{DT}}$ ranging from zero to one, with background peaking closer to zero and signal peaking closer to one. Figures 2(a) and 2(b) show the DT output distributions for two background-dominated data samples to demonstrate the agreement between background model and data. Figure 2(c) shows the high end of the sum of the 12 $tb+tqb$ DT output distributions to illustrate where the signal is expected, and Fig. 2(d) shows the invariant mass of the reconstructed $W$ boson with the highest-$p_T$ $b$-tagged jet (where the neutrino longitudinal momentum has been chosen to be the smaller absolute value of the two possible solutions to the mass equation), for events in a signal-enhanced region with $O_{\text{DT}} > 0.65$. The background peaks near the top quark mass because the DTs select events similar to single top quark events.

We apply a Bayesian approach [28] to measure the single top quark production cross section. We form a binned likelihood as a product over all bins and channels

<table>
<thead>
<tr>
<th>Source</th>
<th>2 jets</th>
<th>3 jets</th>
<th>4 jets</th>
</tr>
</thead>
<tbody>
<tr>
<td>$tb$</td>
<td>16±3</td>
<td>8±2</td>
<td>2±1</td>
</tr>
<tr>
<td>$tqb$</td>
<td>20±4</td>
<td>12±3</td>
<td>4±1</td>
</tr>
<tr>
<td>$tt\rightarrow t\bar{t}$</td>
<td>39±9</td>
<td>32±7</td>
<td>11±3</td>
</tr>
<tr>
<td>$tt\rightarrow t\bar{t}$+$jets$</td>
<td>20±5</td>
<td>103±25</td>
<td>143±33</td>
</tr>
<tr>
<td>$Wbb$</td>
<td>261±55</td>
<td>120±24</td>
<td>35±7</td>
</tr>
<tr>
<td>$Wcc$</td>
<td>151±31</td>
<td>85±17</td>
<td>23±5</td>
</tr>
<tr>
<td>$Wjj$</td>
<td>119±25</td>
<td>43±9</td>
<td>12±2</td>
</tr>
<tr>
<td>Multijets</td>
<td>95±19</td>
<td>77±15</td>
<td>29±6</td>
</tr>
<tr>
<td>Total background</td>
<td>686±41</td>
<td>400±39</td>
<td>253±38</td>
</tr>
<tr>
<td>Data</td>
<td>697</td>
<td>455</td>
<td>246</td>
</tr>
</tbody>
</table>
We assume a Poisson distribution for the observed counts and flat nonnegative prior probabilities for the signal cross sections. Systematic uncertainties and their correlations are taken into account by integrating over the signal acceptances, background yields, and integrated luminosity with Gaussian priors for each systematic uncertainty. The final posterior probability density is computed as a function of the production cross section. For each analysis, we measure the cross section using the position of the posterior density peak and we take the 68% asymmetric interval about the peak as the uncertainty on the measurement.

We test the validity of the cross section measurement procedure using six ensembles of pseudo-datasets selected from the full set of $tb+tb$ signal and background events weighted to represent their expected proportions. A Poisson distribution with a mean equal to the total number of selected events is randomly sampled to determine the number of events in each pseudo-dataset. Each ensemble has a different assumed $tb+tb$ cross section between 2 pb and 8 pb. No significant bias is seen in the mean of the measured cross sections for these ensembles.

The expected SM and measured posterior probability densities for $tb+tb$ are shown in Fig. 3. We use the measured posterior density distribution for $tb+tb$ as shown in Fig. 3 and similar distributions for $tb$ and $tb$ to make the following measurements: $\sigma(pp \rightarrow tb + X, tb + X) = 4.9 \pm 1.4$ pb, $\sigma(pp \rightarrow tb + X) = 4.2^{+1.8}_{-1.4}$ pb, and $\sigma(pp \rightarrow tb + X) = 1.0 \pm 0.9$ pb. These results are consistent with the SM expectations. The uncertainties include statistical and systematic components combined. The data statistics contribute 1.2 pb to the total 1.4 pb uncertainty on the $tb+tb$ cross section.

We assess how strongly this analysis rules out (or is expected to rule out) the background-only hypothesis by measuring the probability for the background to fluctuate up to give the measured (or SM) value of the $tb+tb$ cross section or greater. From an ensemble of over 68,000 background-only pseudo-datasets, with all systematic uncertainties included, we find that the background fluctuates up to give the SM cross section of 2.9 pb or greater 1.9% of the time, corresponding to an expected significance of 2.1 standard deviations (SD) for a Gaussian distribution. The probability that the background fluctuates up to produce the measured cross section of 4.9 pb or greater is 0.035%, corresponding to a significance for our result of 3.4 SD. Using a second ensemble of pseudo-datasets which includes a SM $tb+tb$ signal with 2.9 pb cross section, with all systematic uncertainties included, we find the probability to measure a cross section of at least 4.9 pb to be 11%.

We apply two alternative methods to calculate $tb+tb$ discriminants. The first technique calculates the probability for each event to be signal or background based on the leading-order matrix element description of each process for two-jet and three-jet events [29]. It takes as input the four-momenta of the reconstructed objects and incorporates the $b$-tagging information for each event. This is a powerful method to extract the small signal because it encodes the kinematic information of the signal and background processes at the parton level. The probability that the background fluctuates up to give the SM cross section or greater in the matrix element analysis is 3.7% (1.8 SD). We measure $\sigma(pp \rightarrow tb + X, tb + X) = 4.6^{+1.8}_{-1.5}$ pb. The probability for the background to fluctuate up to give a cross section of at least 4.6 pb is 0.21% (2.9 SD). The second alternative method uses Bayesian neural networks [30] to separate $tb+tb$ signal from background. We train the networks separately for each analysis channel on a sample of signal events and on an equal-sized sample of background events containing the back-
ground components in their expected proportions, using 24 input variables (a subset of the 49 used in the boosted decision tree analysis). Large numbers of networks are averaged, resulting in better separation than can be achieved with a single network. The probability that the background fluctuates up to give the SM cross section or greater in the Bayesian neural network analysis is 9.7% (1.3 SD). We measure \(\sigma(pp \rightarrow tb + X, tqb + X) = 5.0 \pm 1.9 \text{ pb}\). The probability for the background to fluctuate up to give a cross section of at least 5.0 pb is 0.89% (2.4 SD).

The three analyses are correlated since they use the same signal and background models and data, with almost the same systematic uncertainties. We take the decision tree measurement as our main result because this method has the lowest a priori probability for the background to have fluctuated up to give the SM cross section or greater. That is, we expect the decision tree analysis to rule out the background-only hypothesis with greatest significance.

We use the decision tree measurement of the \(tb + tqb\) cross section to derive a first direct measurement of the strength of the \(V-A\) coupling \(|V_{tb}|^2|f_T^4|^2\) in the \(Wtb\) vertex, where \(f_T^4\) is an arbitrary left-handed form factor \([31]\). We measure \(|V_{tb}|^2|f_T^4|^2 = 1.3 \pm 0.2\). This measurement assumes \(|V_{td}|^2 + |V_{ts}|^2 \ll |V_{tb}|^2\) and a pure \(V-A\) and CP-conserving \(Wtb\) interaction. Assuming in addition that \(f_T^4 = 1\) and using a flat prior for \(|V_{tb}|^2\) from 0 to 1, we obtain 0.68 < \(|V_{tb}| \leq 1\) at 95% C.L. These measurements make no assumptions about the number of quark families or CKM matrix unitarity.

To summarize, we have performed a search for single top quark production using 0.9 fb\(^{-1}\) of data collected by the D0 experiment at the Tevatron collider. We find an excess of events over the background prediction in the high discriminant output region and interpret it as evidence for single top quark production. The excess has a significance of 3.4 standard deviations. We use the boosted decision tree discriminant output distributions to make the first measurement of the single top quark cross section: \(\sigma(pp \rightarrow tb + X, tqb + X) = 4.9 \pm 1.4 \text{ pb}\). We use this cross section measurement to make the first direct measurement of the CKM matrix element \(|V_{tb}|\) without assuming CKM matrix unitarity, and find 0.68 < \(|V_{tb}| \leq 1\) at 95% C.L.

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA); CEA and CNRS/IN2P3 (France); FASI, Rosatom and RFBR (Russia); CAPES, CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF (Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); PPARC (United Kingdom); MSMT (Czech Republic); CRC Program, CFI, NISERg and WestGrid Project (Canada); BMF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); The Radcliffe Institute Fellowship program; Research Corporation; Alexander von Humboldt Foundation; and the Marie Curie Program.

---

[6] After this analysis was complete, the following NNNLO calculation was published: N. Kidonakis, Phys. Rev. D 74, 114012 (2006). For \(m_{op} = 175 \text{ GeV}\), \(\sigma(pp \rightarrow tb + X, tqb + X) = 3.21 \pm 0.21 \text{ pb}\).
[22] G.C. Blazey et al., in Proceedings of the Workshop on QCD and Weak Boson Physics in Run II, 1999, edited...
by U. Baur, R.K. Ellis, and D. Zeppenfeld, Fermilab-Pub-00/297, p. 47.


